



POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Informatica

Tesi di Laurea Magistrale

**Analisi della qualità in tazza e  
manutenzione predittiva di macchine  
da caffè professionali**

**Relatori**

Prof. Elena Baralis

Prof. Daniele Apiletti

**Candidato**

Lentini Marco

Dicembre 2019

# Sommario

La capacità di rilevare e immagazzinare dati è aumentata a dismisura negli ultimi anni, consentendo il monitoraggio e la modellazione di processi molto complessi. Molte aziende investono in questo ambito, tra queste Lavazza ha installato un dispositivo di telemetria nelle proprie macchine da caffè presso alcuni bar, per analizzare il processo di erogazione del caffè. In particolare, gli obiettivi dell'azienda sono: monitorare il comportamento della macchina da caffè, ottimizzare gli interventi tecnici di assistenza, e valutare la qualità del caffè basandosi su alcuni dati rilevati in telemetria, come il tempo di erogazione e il flusso.

Il lavoro di tesi si divide in due parti. Nella prima si analizzano dei dati provenienti da test svolti in laboratorio, per i quali l'azienda ha fornito sia i dati rilevati in tazza, sia quelli da telemetria. I test sono svolti solo su erogazioni doppie e con l'uso di un'unica miscela, così da ridurre il numero di variabili esterne che possano influenzare il processo di erogazione. Nella seconda parte, invece, si prendono in considerazione i dati raccolti dalle macchine da caffè installate presso alcuni bar, rilevati internamente grazie alla telemetria, senza conoscenza del tipo di erogazione e della miscela utilizzata.

Nello specifico, si è posta l'attenzione sulle tre variabili caratteristiche dell'erogazione del caffè: il tempo di erogazione, la quantità di erogato in tazza e la velocità del flusso. Per ognuna di esse, sono state fornite dagli esperti del settore delle soglie di qualità.

Le analisi in laboratorio si sono dimostrate da subito interessanti: si è difatti notato come la variazione di alcune variabili esterne, quali dose di macinato, macinatura dei grani e pressione, influenzi in maniera consistente soprattutto il tempo di erogazione e il flusso, mentre l'erogato in tazza si mantenga pressoché costante, in quanto stabilito durante la taratura della macchina. A supporto di questi risultati sono state impiegate le regole di associazione, le quali permettono di estrarre automaticamente e in modo esaustivo i comportamenti frequenti delle erogazioni in presenza di variazioni dei parametri esterni. Di grande rilevanza anche il fatto che si trovino regole diverse in gruppi macchina diversi, a causa di differenze intrinseche e diverso grado di manutenzione dovuto all'assenza di lavaggi.

Dopo aver garantito la correlazione tra dati rilevati in tazza e quelli provenienti dalla telemetria, si è passati all'analisi dei dati in produzione: nonostante la sfida dovuta allo scarso numero di informazioni di contesto rilevate (es. tipo e miscela di erogazione) rispetto ai tanti fattori esterni in gioco, è stato progettato e applicato un approccio per individuare e modellare gli interventi tecnici a partire dai soli dati di telemetria. L'analisi dei dati in

produzione ha permesso di considerare alcuni clienti di riferimento con dati reali, dai quali calcolare nuove soglie di qualità sulle grandezze caratteristiche di una erogazione.

Lo studio relativo agli esperimenti in laboratorio ha permesso di definire un processo di analisi in grado di modellare nel miglior modo possibile i vari fenomeni di interesse su diversi tipi di macchine, con la possibilità di considerare ulteriori tipologie di miscele.

# Ringraziamenti

Vorrei cogliere l'occasione di ringraziare tutte le persone che hanno permesso la realizzazione di questa tesi. In primis ringrazio la Lavazza e la professoressa Elena Baralis per la collaborazione che mi ha permesso di seguire un progetto così interessante, facendomi così conoscere un mondo, quello del caffè, che vedevo solo da lontano.

Il ringraziamento più importante va al mio corelatore Daniele Apiletti, sempre pronto a venire incontro nei momenti più critici attraversati in questi sei mesi, a fornire supporto nonostante la sua agenda pienissima e a dare spunti molto interessanti, anche fuori dall'ambito della tesi. Mi ritengo davvero fortunato ad aver potuto lavorare con lui. Inoltre, un grande grazie anche a Alessandro Chiotti, che mi ha accompagnato durante questi sei mesi di tesi, e con cui ho lavorato molto bene, aiutandoci l'un l'altro, sempre pronti a collaborare per svolgere l'analisi nel migliore dei modi. Ringrazio anche Franco Vitare, punto di riferimento in Lavazza, sempre disponibile ad accogliere le nostre richieste e a fornire chiarimenti, e Andrea Virgilio che ci ha fornito numerosi spunti per l'approfondimento dell'analisi.

Ringrazio infine la mia famiglia che mi è stata accanto in questi ultimi mesi molto densi di impegni e ai miei amici sempre presenti nelle lunghe giornate di studio.

Un ultimo ringraziamento al lettore di questo documento, sperando di fornire una descrizione chiara e interessante sugli studi effettuati.

# Indice

<b>Elenco delle tabelle</b>	7
<b>Elenco delle figure</b>	9
<b>1 Introduzione</b>	11
<b>2 Stato dell'arte</b>	15
<b>3 Contesto applicativo</b>	19
3.1 Struttura di una macchina del caffè . . . . .	19
3.2 Processo di macinatura della macchina e della macinatura . . . . .	20
3.3 Device di telemetria . . . . .	21
<b>4 Collezione dei dati e struttura dei dataset</b>	23
4.1 Collezione dei dati . . . . .	23
4.2 Prima sessione sperimentale . . . . .	24
4.2.1 File iniziale . . . . .	25
4.2.2 File modificato . . . . .	28
4.2.3 File per regole di associazione . . . . .	28
4.3 Seconda sessione sperimentale . . . . .	29
4.4 Sessione sperimentale proposta . . . . .	30
4.5 Dati provenienti dalla telemetria . . . . .	31
4.6 File degli interventi . . . . .	32
<b>5 Approccio sperimentale</b>	33
5.1 Python e Jupyter Notebook . . . . .	33
5.2 Tecniche di pulizia e allineamento dei dati . . . . .	34
5.2.1 Pulizia dei dati sul singolo test in laboratorio . . . . .	34
5.2.2 Pulizia e allineamento dei dati di telemetria . . . . .	34
5.2.3 Pulizia delle erogazioni delle macchine in produzione . . . . .	35
5.3 Caratteristiche di un insieme di dati . . . . .	36
5.3.1 Relazione tra due insiemi di dati . . . . .	37
5.3.2 Grafici in Matplotlib . . . . .	38
5.4 Regressione lineare . . . . .	39
5.5 Regole di associazione . . . . .	40
5.5.1 Algoritmo Apriori . . . . .	42

5.5.2	Implementazione in Python dell'algoritmo Apriori . . . . .	43
<b>6</b>	<b>Risultati sperimentali</b>	<b>45</b>
6.1	Analisi esplorativa della prima sessione sperimentale . . . . .	46
6.1.1	Selezione delle variabili principali . . . . .	47
6.1.2	Analisi di correlazione . . . . .	48
6.1.3	Andamento delle variabili e confronto tra gruppi . . . . .	49
6.2	Relazione tra rilevazioni da fonti differenti . . . . .	50
6.2.1	Correlazione tra scheda PC e rilevazione in tazza . . . . .	50
6.2.2	Correlazione tra dato di telemetria e rilevazione in tazza . . . . .	55
6.3	Calcolo delle nuove soglie . . . . .	58
6.3.1	Proposta di nuove soglie a partire dai dati di laboratorio . . . . .	59
6.3.2	Proposta di soglie di flusso a tre livelli . . . . .	60
6.4	Regole di associazione . . . . .	61
6.4.1	Tipi di regole . . . . .	62
6.4.2	Dati a disposizione . . . . .	63
6.4.3	Pulizia dei dati . . . . .	63
6.4.4	Soglie di qualità . . . . .	64
6.4.5	Regole di associazione per il tempo di erogazione . . . . .	65
6.4.6	Regole di associazione per il flusso . . . . .	70
6.4.7	Regole di associazione per il flusso con tre livelli di soglie . . . . .	75
6.4.8	Conclusioni . . . . .	79
6.5	Proposta sessione sperimentale . . . . .	80
6.5.1	Differenza intrinseca dei gruppi . . . . .	80
6.5.2	Relazione tra colpi di ventolino e erogato in tazza . . . . .	81
6.5.3	Modellazione dei lavaggi anomali . . . . .	81
6.5.4	Degrado dovuto alla mancanza di lavaggi . . . . .	82
6.5.5	Influenza delle variabili esterne . . . . .	83
6.6	Analisi dei dati provenienti dal mercato . . . . .	83
6.6.1	Modellazione di clienti fidati . . . . .	85
6.6.2	Intervento tecnico . . . . .	99
<b>7</b>	<b>Conclusioni</b>	<b>109</b>
7.1	Limiti attuali dell'analisi . . . . .	110
7.2	Possibili studi futuri . . . . .	111
	<b>Bibliografia</b>	<b>113</b>

# Elenco delle tabelle

5.1	Esempio di dataset transazionale . . . . .	41
6.1	Funzione di trasferimento da flusso scheda PC a flusso in tazza . . . . .	52
6.2	Statistiche delle quantità in tazza e della scheda PC . . . . .	53
6.3	Funzione di trasferimento da quantità scheda PC a in tazza . . . . .	54
6.4	Funzione di trasferimento da flusso scheda PC a flusso in tazza . . . . .	54
6.5	Statistiche della quantità rilevata in tazza e dalla scheda PC . . . . .	55
6.6	Funzione di trasferimento da flusso di telemetria a flusso in tazza . . . . .	56
6.7	Funzione di trasferimento da quantità della telemetria a quantità in tazza. . . . .	57
6.8	Statistiche sulle quantità in tazza e rilevate dalla telemetria . . . . .	57
6.9	Soglie in tazza degli esperti di dominio . . . . .	59
6.10	Statistiche e soglie in tazza ricavate da erogazioni ottimali . . . . .	59
6.11	Soglie in tazza proposte a partire dai dati. . . . .	60
6.12	Soglie di flusso a tre livelli. . . . .	61
6.13	Supporto delle etichette di qualità del flusso con le diverse soglie . . . . .	64
6.14	Regole di associazione per il tempo. Studio dell'influenza delle variabili esterne . . . . .	66
6.15	Regole di associazione per il tempo. Studio dell'influenza dei lavaggi . . . . .	68
6.16	Regole di associazione per il flusso. Studio dell'influenza delle variabili esterne . . . . .	71
6.17	Regole di associazione per il flusso. Studio dell'influenza dei lavaggi . . . . .	74
6.18	Supporti con soglie di flusso a tre livelli. . . . .	76
6.19	Regole di associazione per il flusso con soglie a 3 livelli. Studio dell'influenza delle variabili esterne. . . . .	77
6.20	Regole di associazione per il flusso con soglie a 3 livelli. Studio dell'influenza dei lavaggi. . . . .	78
6.21	Soglie in tazza e in telemetria . . . . .	86
6.22	Dettagli dei periodi analizzati . . . . .	87
6.23	Dettagli della pulizia dei dati per gruppo nella Nuvola Lavazza nel periodo 27/06/2019–11/07/2019. . . . .	89
6.24	Conteggio delle etichette di qualità per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019. . . . .	90
6.25	Calcolo statistiche per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019. . . . .	91
6.26	Calcolo soglie per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019. . . . .	91
6.27	Dettagli della pulizia dei dati per gruppo per Nuvola Lavazza nel periodo 17/05/2019–30/05/2019. . . . .	93
6.28	Conteggio delle etichette di qualità per la Nuvola Lavazza nel periodo 17/05/2019–30/05/2019. . . . .	93
6.29	Calcolo statistiche per la Nuvola Lavazza nel periodo 17/05/2019–30/05/2019. . . . .	94

6.30	Calcolo statistiche per la Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.	94
6.31	Dettagli della pulizia dei dati per gruppo per Top shop nel periodo 08/07/2019 – 27/07/2019 . . . . .	95
6.32	Conteggio delle etichette di qualità per Top shop nel periodo 08/07/2019 – 27/07/2019. . . . .	95
6.33	Calcolo statistiche per la Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.	96
6.34	Dettagli della pulizia dei dati per gruppo per Gruppo Spes SCS Etiko nel periodo 06/06/2019–16/06/2019. . . . .	97
6.35	Conteggio delle etichette di qualità per Gruppo Spes SCS Etiko nel periodo 06/06/2019–16/06/2019. . . . .	97
6.36	Calcolo statistiche per Gruppo Spes SCS Etiko nel periodo 06/06/2019–16/06/2019.	98
6.37	Statistiche dei periodi a cavallo dell'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso. . . . .	101
6.38	Statistiche dei periodi a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas. . . . .	104
6.39	Statistiche dei periodi a cavallo del 17/06/2019 al cliente Spes SCS Etiko. .	106

# Elenco delle figure

3.1	Schema dei componenti di una macchina del caffè[8]. . . . .	20
4.1	Descrizione del raccoglimento dei dati . . . . .	24
5.1	Esempio di istogramma . . . . .	38
5.2	Esempio di boxplot e le sue caratteristiche . . . . .	39
5.3	Esempio di grafico con <i>plot</i> . . . . .	39
5.4	Esempio di regressione lineare . . . . .	40
6.1	Matrice di correlazione . . . . .	48
6.2	Distribuzione delle tre variabili principali nella prima sessione sperimentale.	49
6.3	Andamento dei flussi rilevati in tazza e da scheda PC . . . . .	51
6.4	Retta di regressione da flusso scheda PC a flusso in tazza . . . . .	51
6.5	Andamento delle quantità della scheda PC e in tazza . . . . .	52
6.6	Retta di regressione da quantità della scheda PC a in tazza . . . . .	53
6.7	Andamento dei flussi in tazza e rilevati in telemetria . . . . .	55
6.8	Retta di regressione da flusso di telemetria a flusso in tazza . . . . .	56
6.9	Andamento della quantità in tazza e in telemetria . . . . .	57
6.10	Andamento quantità d'acqua scheda PC e telemetria . . . . .	58
6.11	Regole estratte per 'ok_alto' con soglie a tre livelli per il flusso 0,73-1,1-1,23	76
6.12	Tracciato del test sull'influenza delle variabili esterne . . . . .	84
6.13	Dati di vendita della Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.	88
6.14	Differenza tra le distribuzioni dell'acqua erogata dei gruppi 2 e 3 . . . . .	89
6.15	Distribuzione delle tre variabili principali per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019. . . . .	90
6.16	Statistiche degli interventi analizzati . . . . .	100
6.17	Andamento dell'acqua erogata nel periodo a cavallo dell'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso. . . . .	102
6.18	Andamento dell'acqua erogata nel periodo precedente all'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso . . . . .	102
6.19	Andamento di flusso e tempo nel periodo a cavallo dell'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso . . . . .	103
6.20	Andamento dell'acqua erogata nel periodo a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas. . . . .	104
6.21	Andamento di flusso e tempo nel periodo a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas. . . . .	105
6.22	Andamento del flusso nel periodo a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas. . . . .	105

6.23 Andamento della quantità di acqua erogata nel periodo a cavallo dell'intervento del 17/06/2019 al cliente Spes SCS Etiko. . . . .	107
--	-----

# Capitolo 1

## Introduzione

In un periodo in cui la capacità di immagazzinare e trasferire informazioni è cresciuta a dismisura, i dati sono diventati importanti perché in grado di monitorare e di predire processi e attività molto complesse. In quest'ottica, la Lavazza ha deciso di investire in un device di telemetria da inserire nelle macchine da bar per monitorare e analizzare il processo di erogazione del caffè, un'attività che, in Italia, è un'arte. Da leader incontrastato per quanto riguarda il mercato del caffè in Italia, l'obiettivo della Lavazza è offrire un'esperienza di massima qualità. In particolar modo, l'azienda vuole approfondire alcuni temi grazie all'installazione della telemetria

- Qualità in tazza del caffè: grazie ad alcuni parametri rilevati durante l'erogazione, si cerca di capire la qualità del caffè, se buono oppure meno, e di individuare la causa di un caffè venuto male. L'obiettivo è quindi modellare sotto forma numerica le caratteristiche principali di un caffè ottimo, che contiene tutto l'aroma della miscela e si presenta alla vista nel migliore dei modi;
- Efficienza del servizio: monitorare una macchina per verificarne il corretto funzionamento e per programmare un piano di manutenzione preventiva e ottimizzare l'intervento del tecnico;
- Monitoraggio delle vendite: analizzare i dati delle vendite per individuare i clienti che utilizzano dosi di caffè ridotte e quelli, invece, che utilizzano il caffè di terzi, al fine di garantire l'esperienza Lavazza a tutti i suoi consumatori.

La maggior parte degli italiani beve almeno una tazza di caffè al giorno in momenti della giornata quali la colazione, dopo i pasti, oppure durante una pausa; in altre parole il caffè in Italia è una vera e propria abitudine. Un numero considerevole di clienti di un bar, però, non si è mai soffermato a pensare a come i grani del caffè vengano trasformati in questa bevanda adatta ad ogni momento della giornata. L'erogazione di un caffè è, infatti, influenzata da molteplici fattori:

- Lo stato della macchina: la macchina necessita di lavaggi giornalieri e lavaggi brevi, detti purge, dopo ogni erogazione;
- La miscela di caffè utilizzata: il sapore, l'aroma, il colore e tutte le caratteristiche principali del caffè variano a seconda del tipo di miscela utilizzata;

- La macinatura del caffè: a seconda del tipo di miscela, i grani sono più o meno macinati affinché l'acqua acquisisca appieno l'aroma di caffè. Inoltre la macinatura può variare a seconda dello strumento utilizzato per macinare;
- Il tipo di erogazione desiderato: oltre all'espresso normale, si possono avere caffè lunghi e corti. Si hanno erogatori da singola e doppia dose, che permettono di erogare rispettivamente uno o due caffè contemporaneamente;
- La dose di macinato utilizzata: al variare della miscela e della dose di caffè in tazza desiderata cambia la dose ottimale di macinato da inserire nell'erogatore;
- I parametri ambientali: la conservazione del macinato varia a seconda della temperatura e dell'umidità, influenzandone l'aroma e la compattezza.

A queste numerose variabili tecniche, si aggiungono altri elementi determinanti che influenzano l'erogazione e la qualità del caffè in tazza:

- Le usanze culturali: ogni regione ha la sua cultura del caffè. Ad esempio, Napoli presenta un caffè molto ristretto e in cui la dose di erogato è minore rispetto a quella standard erogata nel nord Italia;
- L'esperienza e le abitudini del barista: ogni barista tende a "personalizzare" il proprio caffè: ad esempio, un nuovo dipendente potrebbe non sapere quali sono gli standard di un buon caffè, scegliendo di effettuare un'erogazione più lunga del dovuto; al contrario un barista che eroga caffè da anni potrebbe scegliere di modificare la macinatura in modo tale da ottenere un caffè con una bella schiuma, oppure interrompere l'erogazione, perché l'aroma del caffè sia più deciso.

Come punto iniziale iniziale dello studio, si è cercato di ridurre le variabili effettuando in un primo momento delle sessioni di test controllate, in cui si svolgono solo erogazioni doppie normali, le più diffuse sul mercato, usando un solo tipo di miscela conservata nel migliore dei modi. In particolare è stata posta l'attenzione alle variabili che caratterizzano il processo di erogazione del caffè, e per cui il Training Center della Lavazza ha definito delle soglie di qualità:

- Tempo di erogazione: il processo di erogazione dovrebbe durare 24-25 secondi circa e si considerano ottimali caffè la cui erogazione dura tra i 20 e i 27 secondi;
- Erogato in tazza: la quantità di erogato in tazza dovrebbe essere tra i 20 e i 30 ml. L'erogato in tazza è notevolmente influenzato dai processi di taratura della macinatura e della macchina e che verranno approfonditi in un secondo momento;
- Velocità del flusso: il flusso, calcolato come erogato in tazza diviso tempo di erogazione, deve rispettare la regola dei 30 cc in 30 secondi, e deve quindi essere centrato su 1 ml/s. Le soglie iniziali imposte sono di 0.49 e 1.33 ml/s, messe in discussione durante l'analisi.

In questa prima parte si è cercato di individuare come queste variabili di interesse cambino modificando alcuni parametri esterni, come la pressione della pompa della macchina, la macinatura dei grani e la dose di macinato utilizzato, per capire come essi influenzano

la qualità del caffè. In aggiunta, uno dei focus è stato rivolto al confronto tra erogazioni, svolte in presenza e in assenza di lavaggi, al fine di scoprire l'impatto di una mancata manutenzione della macchina.

In un secondo momento, l'attenzione è stata rivolta all'analisi di dati provenienti dal mercato e, quindi, in un ambiente non controllato: oltre a controllare la consistenza dei dati provenienti dal device di telemetria ancora nelle sue versioni iniziali, le osservazioni di periodi di alcuni clienti fidati sono state analizzate con l'obiettivo di stabilire alcune soglie di riferimento per le grandezze di interesse dell'analisi mantenendo, però, alcune approssimazioni dovute alla mancanza di informazioni su alcuni elementi fondamentali come il tipo di miscela usata e il tipo di erogazione considerata. Inoltre, sono stati analizzati alcuni interventi tecnici, confrontando il periodo precedente e successivo e cercando di modellare il tipo di intervento effettuato.

Durante lo studio iniziale, si sono riscontrate alcune problematiche: in primo luogo, il grande ammontare di tempo necessario per collezionare i dati in laboratorio, a causa del numero elevato di erogazioni necessario, oltre al lungo lavoro di allineamento e pulizia dei dati provenienti da diverse fonti. Essendo il device di telemetria nelle sue versioni iniziali, sono sorti alcuni problemi per quanto riguarda la ricezione e la consistenza dei dati con notevoli differenze a seconda del modello di macchina su cui il device è installato.

Questo documento sarà strutturato come segue. Nel secondo capitolo, si studia lo stato dell'arte del problema, gli studi già effettuati per quanto riguarda le soglie di qualità del caffè e quali conclusioni sono note. Nel terzo capitolo, si spiega il contesto in cui si è svolta la tesi, il processo di erogazione del caffè e le caratteristiche che è necessario conoscere per avere una visione generale del problema affrontato. Nel quarto capitolo, invece, si affronta la modalità secondo la quale i dati sono stati collezionati e, in particolar modo, da quali fonti provengono e come sono stati strutturati per essere analizzati. Il quinto capitolo approfondisce le tecniche utilizzate durante l'analisi, così da fornire una visione più ampia al lettore. Il sesto capitolo presenta i risultati ottenuti e ricostruisce il percorso seguito durante lo studio del problema proposto. Il settimo capitolo conclude la presente tesi e propone i risultati ottenuti, evidenziando aspetti quali le limitazioni tecniche affrontate, il limite dello studio effettuato e le possibili estensioni della ricerca.



## Capitolo 2

# Stato dell'arte

Questo capitolo descrive gli studi effettuati nell'ambito delle macchine da caffè e per quanto riguarda la qualità del caffè. Sono stati effettuati diversi tipi di studi, basati su sensori e tecniche diverse, ma mai con l'approccio usato in questa tesi. In particolare, mai si è tentato in real time di analizzare i dati di ogni erogazione per predire la qualità del caffè. Riguardo alla qualità del caffè in tazza, sono stati effettuati numerosi studi che confermano il fatto che le caratteristiche del caffè dipendano da numerosi fattori. Per esempio, nello studio [4], si cerca una correlazione tra la qualità dell'espresso e l'acqua usata per produrlo: anche le proprietà dell'acqua possono cambiare il processo di realizzazione del caffè. Infatti, è stato rilevato che la schiuma sopra il caffè dipende dalla composizione dell'acqua e, analizzando il volume, la composizione e la durata della schiuma sono state notate notevoli differenze. Lo studio in questione è stato effettuato analizzando caffè ottimali: secondo gli autori dell'articolo, il caffè in Oitalia ha una durata tra i 25 e i 30 secondi e una quantità tra i 25 e 30 ml. Quindi si nota che le soglie di qualità sono compatibili con quelle forniteci dalla lavazza, seppur leggermente più strette. Infatti, nell'analisi della schiuma è molto importante che l'erogato e il tempo siano molto precisi e costanti così da avere un termine di paragone più preciso. Nel nostro caso invece, i caffè presi in considerazione sono quelli effettuati quotidianamente nei bar, e quindi più variabili.

Un'altro studio interessante si focalizza sul rapporto acqua/caffè nella determinazione della qualità in tazza ([1]). K'autore ha deciso di analizzare il comportamento psico-chimico dei parametri del caffè, eseguendo erogazioni con una quantità di dose differente. In particolare sono state testate tre dosi diverse (6.5 g, 7.5 g, 8.5 g) per tre diverse tipologie di grani. Gli altri parametri sono mantenuti costanti: in particolare il tempo di erogazione (tra 18 e 24 secondi, ovvero leggermente inferiore a quello considerato nel nostro studio), e la quantità ( $40 \pm 2ml$ ). Questo parametro così alto risulta molto strano in quanto l'articolo è pubblicato in Italia, da un competitore di Lavazza (Illy). Il valore così elevato si può spiegare con lo scopo dell'analisi, che risulta incentrata a capire come l'acqua attraversi il panetto. I risultati sperimentali mostrano come la differenza non sia apprezzabile in quanto si presentano dei fattori di compensazione dovuti a una regolazione di pressione, temperatura e macinatura, che minimizzano l'effetto della variazione del rapporto acqua/caffè. Un'analisi più approfondita è stata effettuata in questo studio con le regole di associazione, alla sezione 6.4.

Il concetto di analisi della composizione del caffè è utilizzata anche in altri studi, come in

[4]: in questo caso, un sistema di simulazione di un naso elettronico studia i cambiamenti nel profilo aromatico dell'espresso. In particolare, esso mette in risalto l'importanza dei primi 8 secondi dell'erogazione del caffè, perchè in questo range sono estratte la maggior parte degli composti organici presenti nel caffè, e assieme ad essi, la caffeina stessa. I risultati dimostrano che sia il tempo di erogazione che il grado di macinatura influenzano significativamente il profilo aromatico dell'espresso.

Uno studio più recente([2]) si avvicina in parte al modello utilizzato nel nostro studio: si parla infatti dell'introduzione di un sensore IoT all'interno della macchinetta: il suo scopo è quello di rilevare dei dati per automatizzare il processo di vendita del caffè e per monitorare lo stato della macchina tramite una applicazione. Proprio in questa seconda parte essa si avvicina al nostro studio, in quanto il sensore è utilizzato per la predictive maintenance della macchina: la differenza sostanziale, e innovativa, è che il nostro studio si concentra anche sui dati delle singole erogazioni(tempo e flusso), oltre che sui parametri tecnici della macchina(pressione, temperatura, quantità d'acqua totale).

Nell'articolo [11] si affronta proprio il tema della predictive maintenance: l'obiettivo è di installare un sensore di telemetria su macchine vecchie, e attraverso delle tecniche di machine learning, predire quando sarà effettuato l'ultimo caffè sulla base delle vibrazioni del contenitore dei grani, che saranno maggiori quando è vuoto. Sebbene lo scopo dello studio non sia così rilevante, è importante l'approccio posto al problema: esso infatti costituisce la prova tangente che attraverso il machine learning si possa effettuare una predictive maintenance delle macchine da caffè, così da predire gli interventi tecnici e limitare molto il costo. Su questa linea di pensiero, la sezione 6.6.2, analizza alcuni interventi, e rileva alcune oscillazioni nei parametri dell'erogazione che potrebbero indicare il malfunzionamento della macchina.

Riassumendo il discorso sulla qualità in tazza, abbiamo visto come siano presenti alcune proposte di soglie per individuare il caffè ottimale: gli autori dell'articolo [13] propongono le soglie 20-25s e 20-25 ml per tempo e quantità. Gli autori del [1] invece usano un tempo di erogazione tra i 25 e i 30 secondi, ma un erogato molto elevato. Un'altra proposta si può trovare nel libro [15], che, sempre prodotto attraverso degli studi di Illy, propone delle soglie di tempo di  $30 \pm 5s$  e un erogato compreso tra i 25 e 50 ml in base alla richiesta del cliente, considerato ottimale però tra 25 e 30 ml. Si può quindi notare come le soglie proposte da Lavazza, e analizzate maggiormente nella sezione 6.3, siano una buona via di mezzo tra queste proposte di letteratura. Non si trovano invece particolari riferimenti al flusso: ciò è comprensibile in quanto date le soglie di tempo e erogato, è possibile calcolarle. Si può vedere infatti che i valori delle due grandezze rispetta la regola dei 30cc in 30 secondi indicata da Lavazza, che porta a un flusso ottimale di 1 ml/s.

Il libro [15] fornisce anche delle soglie per quanto riguarda alcuni dei parametri esterni: pressione nella fascia  $9 \pm 2bar$ , che concordano con gli esperimenti effettuati; temperatura nella fascia  $90 \pm 5^{\circ}C$  e dose di macinato centrata sui 6,5 grammi, come indicato da Lavazza. In letteratura, non sono presenti studi simili a quelli di questa tesi: in particolare, non sono mai prese in considerazione le caratteristiche dell'erogazione, ma solo quelle della macchina. Proprio nella relazione tra razza e telemetria sta la sfida che Lavazza si propone di affrontare. In questo modo, oltre alla predictive maintenance è possibile effettuare una modellazione dei clienti più precisa, individuare anomalie nei comportamenti e ottimizzare i guadagni.

Per capire l'importanza di questo studio, e quanto si stia investendo nell'ambito dei

sensori sui dispositivi e sull'analisi dei dati di telemetria, negli ultimi mesi[16], Costa Coffee, competitor di Lavazza, ha annunciato l'installazione di un device simile a quello della Lavazza sui suoi dispositivi, con gli stessi obiettivi della Lavazza, ovvero migliorare la catena di produzione e vendita del caffè, monitorare lo stato della macchina e poter valutare la qualità del prodotto offerto in tempo reale.



## Capitolo 3

# Contesto applicativo

In questo capitolo è descritto il contesto in cui si è svolto lo studio, approfondendo la struttura della macchina del caffè e i vari step di una erogazione. Inoltre viene fornita anche una descrizione di come opera il device di telemetria all'interno di essa, e quali sono i dati che esso raccoglie.

### 3.1 Struttura di una macchina del caffè

Sono numerosi i costruttori di macchine per il caffè. La Lavazza ha scelto in particolare tre produttori per le macchine da installare nei bar dei suoi clienti:

- Cimbali;
- Wega;
- Rancilio.

Ognuno di essi inoltre produce più tipi di modelli di dispositivi, ognuno con certe specifiche tecniche e implementazioni diverse. La struttura e il funzionamento di una macchina del caffè rimane però standard[8]: il dispositivo viene collegato alla rete idrica e l'acqua, dopo aver attraversato alcuni filtri depurativi, viene immagazzinata in una caldaia, dove viene riscaldata e tenuta a temperatura costante. Una volta premuto il tasto per l'erogazione, l'acqua arriva all'erogatore, e grazie a una pompa attraversa il panetto di caffè impregnandosi di aroma e terminando in tazza col suo classico colore scuro. Possiamo quindi individuare tre elementi principali:

- la caldaia: responsabile di portare e mantenere l'acqua a una temperatura corrispondente a 1,3 bar;
- la pompa: essa si trova a monte della caldaia ed è responsabile dell'erogazione. Quando l'erogazione viene avviata, imprime all'acqua una pressione di 9 bar così che essa riesca ad attraversare il panetto di macinato, dopo aver attraversato il diffusore e la doccetta, così che l'impatto col macinato sia il più omogeneo possibile. Il valore della pressione imposta può essere modificato tramite regolazione manuale della pompa stessa;



- elettronici: i grani sono macinati sul momento, in base alla quantità di macinato richiesto per il caffè da erogare. In questo modo si conserva la fragranza della miscela, oltre a evitare possibili sgrammature, in quanto la dose di macinato può essere impostata solamente tramite un tasto.

Per ottenere un caffè di qualità ottima è importante effettuare la taratura della macchina[5]. Essa è un processo in cui vengono settati i parametri corretti per avere 46 grammi di caffè in tazza in 25 secondi, per un caffè doppio.

Per raggiungere questo risultato, la macinatura del caffè è modificata attraverso una ghiera sul macinadosatore. È un processo a tentativi: impostata una certa macinatura, si preme il tasto di erogazione doppia e si eroga finché non si raggiunge la quantità in tazza desiderata. A questo punto si schiaccia nuovamente il tasto della dose doppia per interrompere l'erogazione e si valuta il tempo di erogazione ottenuto: se è maggiore di 25 secondi, la macinatura è troppo fine, e viceversa. Si cambia allora il grado di macinatura del caffè e si procede nuovamente con il processo descritto, fin quando non si raggiunge l'erogazione di 46 grammi di caffè in 25 secondi. Una volta raggiunto il risultato desiderato, la macchina salva il numero di colpi di ventolino utilizzati per ottenere quel caffè, e lo associa all'erogazione doppia. Così, ogni volta che il tasto è selezionato, la quantità di acqua è erogata. Questa è la procedura standard di taratura, ma i baristi esperti sono in grado di effettuare una taratura perfetta osservando le caratteristiche del caffè, la velocità con cui scorre, il suo colore. Infatti spesso la dose in tazza non corrisponde a quella consigliata dagli esperti, in quanto può variare a seconda delle abitudini del barista e agli usi regionali.

I parametri descritti si riferiscono a una dose di macinato di 14 grammi per un caffè doppio: alcuni baristi potrebbero inserire una dose minore di macinato per avere un risparmio economico, e, modificando la macinatura, ottenere comunque un caffè ottimale, almeno secondo i parametri forniti dagli esperti.

### 3.3 Device di telemetria

La Lavazza ha creato un device di telemetria da installare sulle macchine da bar. In particolare, esso comunica con la scheda elettronica della macchina, che rileva tramite dei sensori alcuni parametri caratteristici dell'erogazione e dello stato della macchina. A seconda della macchina, il dato rilevato dai sensori può essere o meno elaborato, così da filtrare alcune erogazioni anomale o togliere alcuni picchi nella rilevazione della stessa. Una volta preso il dato dalla scheda, il device si collega tramite connessione 2G a un cloud che raccoglie e consolida i dati ricevuti, elaborandoli per ottenere indici interessanti. A questo punto, i dati sono disponibili attraverso un portale web, in cui si possono analizzare i dati di vendita e qualità dell'erogazione, oltre a rilevazioni sullo stato della macchina.

Il device rileva, in particolare, tre tipi diversi di dato:

- dati sullo stato della macchina: sono rilevati ogni minuto, e rappresentano lo stato della macchina in quel momento. Contiene rilevazioni quali la temperatura, la pressione, il voltaggio del device, il numero di lavaggi svolti fino a quel momento. Esse sono molto utili per monitorare il corretto funzionamento della macchina;

- dati sulle erogazioni: per ogni erogazione effettuata, vengono rilevate il tempo di erogazione e il flusso medio. In questo modo é possibile vedere le caratteristiche delle varie erogazioni e analizzare la loro qualità;
- dati di vendita: riportano il numero di erogazioni effettuate nei vari giorni, divise in base al tasto di selezione. Permettono l'analisi delle vendite del cliente, la modellazione delle sue abitudini e un confronto con la quantità di caffè comprato da esso.

Il numero di parametri rilecati varia a seconda del modello di macchina su cui il device é installato, in base al numero di sensori a sua disposizione.

Attraverso il portale, è possibile scaricare i dati di un oeriodo di massimo 41 giorni in un file Excel. La struttura del file scaricato è descritto nella sezione 4.5

## Capitolo 4

# Collezione dei dati e struttura dei dataset

In questa sezione verrà analizzato il processo di collezione dei dati su cui sono state svolte le analisi e come essi siano stati strutturati nei vari casi per poter essere studiati. In questo modo viene anche brevemente descritta la struttura degli esperimenti svolti e il loro obiettivo

### 4.1 Collezione dei dati

Durante gli esperimenti in laboratorio, sono disponibili diversi tipi di misurazione per una erogazione:

- Rilevazioni esterne: sono caratteristiche rilevate in maniera manuale al di fuori della macchina. In primis, viene misurata la quantità di caffè erogato in tazza attraverso una bilancia. Inoltre sono numerosi i fattori caratterizzanti un'erogazione che possono essere misurati esternamente: tra questi, tramite display della macchina, si possono avere informazioni sulla temperatura della caldaia e sulla pressione della pompa, oltre che sul numero di colpi di ventolino impostati durante la taratura della macchina. Inoltre, si hanno informazioni sulla dose di macinato inserito nell'erogatore, e sul grado di macinatura dei grani.
- Rilevazioni della scheda PC: i dati sono rilevati direttamente dalla scheda elettronica della macchina, e sono quindi calcolati a monte del panetto di caffè. Sono dati grezzi, calcolati a partire da dei sensori che rilevano il dato ogni 200 millisecondi per ogni erogazione. A seconda del modello di macchina, i dati rilevati possono variare: la Cimbali M100 rileva solo tempo di erogazione e colpi di ventolino, mentre le macchine Wega mettono a disposizione anche i dati di pressione e temperatura. Al momento, essendo le analisi in laboratorio svolte su una Cimbali M100 ci si concentra su tempo di erogazioni e colpi di ventolino, che rappresentano la quantità di acqua erogata.
- Rilevazioni della telemetria: il device di telemetria prodotto dalla Lavazza viene installato sulla macchina e messo in comunicazione con la scheda della macchina, così

da reperire i dati rilevati, elaborarli e mandarli tramite connessione internet al portale della compagnia. Sono forniti, oltre a alcune informazioni riguardanti lo stato della macchina rilevate ogni minuto, anche il tempo di erogazione e il flusso medio di ogni erogazione. Quest'ultimo è un dato rielaborato a partire da tempo di erogazione e colpi di ventolino, escludendo però alcuni picchi e fasi dell'erogazione dal calcolo.

La figura 4.1 descrive quali sono i dati di interesse dell'analisi e come essi sono rilevati a partire dalle fonti e poi combinati per ottenere alcune grandezze interessanti come il flusso. Si può notare come le informazioni su macinatura, dose e pressione siano impostate esternamente e quindi non rilevate dalla telemetria durante l'erogazione. Inoltre il dato disponibile in telemetria è riferito alla quantità di acqua erogata dalla macchina, che comprende, oltre alla quantità che arriva in tazza, anche l'acqua risucchiata al termine del processo.

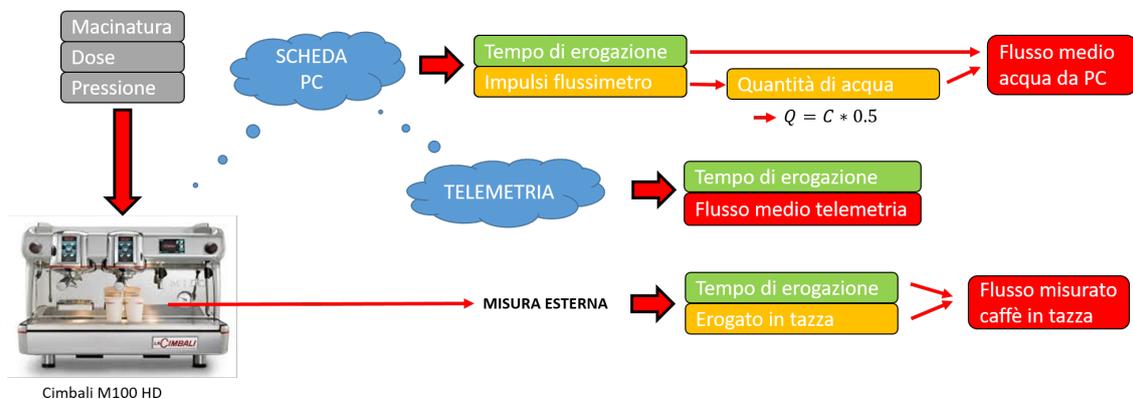


Figura 4.1: Descrizione del raccoglimento dei dati

Per gli esperimenti in laboratorio si è scelto l'utilizzo della Cimbali M100 in quanto permette la rilevazione contemporanea sia del dato di telemetria che del dato della scheda PC, che permette di monitorare tutto l'andamento dell'erogazione. Inoltre essi possono essere confrontati con il dato in tazza e analizzati tenendo conto delle condizioni di macinatura, dose e pressione che caratterizzano ogni caffè prodotto.

Le macchine monitorate sul mercato, però, presentano solo l'informazione della telemetria e non quella rilevata esternamente, e quindi non si hanno informazioni sulle caratteristiche del macinato utilizzato per l'erogazione. Confrontando i dati delle diverse fonti rilevati in laboratorio si cerca allora di trovare la correlazione tra le diverse variabili, così da poter applicare le conoscenze di dominio riferite al dato in tazza anche ai valori in telemetria.

## 4.2 Prima sessione sperimentale

L'obiettivo della raccolta dei dati nella prima sessione sperimentale è osservare la variazione delle caratteristiche del caffè (tempo di erogazione, flusso e erogato in tazza) a causa di variazioni estreme dei parametri esterni (macinatura, dose e pressione) rispetto all'ottimale. Inoltre si effettuano i lavaggi solo su uno dei due gruppi, così da valutare il degrado dovuto

alla mancanza di pulizia. Le rilevazioni sono state effettuate nel mese di marzo sulla macchina La Cimbali M100.

In particolare si effettuano 27 test per gruppo, così da coprire tutte le possibili combinazioni delle variazioni dei parametri esterni, ognuno composto da 20 erogazioni. Nel gruppo 1 inoltre non sono stati effettuati i lavaggi e le purge, mentre il gruppo 2 viene mantenuto pulito.

I dati e le informazioni messe a disposizione sono state proposte dagli esperti della Lavazza: sono presenti molte informazioni e molti indici derivati. A partire da questa proposta, è stata effettuata un'analisi (vedi sezione 6.1.1) per capire quali informazioni fossero più utili, così da ridurre i parametri rilevati e velocizzare l'operazione di collezione dati.

### 4.2.1 File iniziale

Questa sezione descrive la struttura del file Excel iniziale fornitoci dalla Lavazza, con la proposta di attributi da parte degli esperti. Il dataset presenta 1080 righe e 40 colonne.

Ogni riga corrisponde a una erogazione doppia mentre le colonne descrivono le sue caratteristiche. Le colonne sono state raggruppate in base alla tipologia di informazione contenuta. Si procede con la descrizione del significato delle varie colonne, e dei possibili valori che esse assumono.

- *Informazioni generali*: forniscono le informazioni tecniche dell'erogazione.
  - Test: identificativo del test a cui appartiene l'erogazione. Ogni test corrisponde a una certa combinazione di Macinatura, Dose e Pressione Erogazione (Es. "T\_01", "T\_02",...);
  - ID: numero identificativo unico dell'erogazione. Intero incrementale a partire da 0;
  - Giorno: data in cui è avvenuta l'erogazione;
  - Orario: ora a cui è avvenuta l'erogazione;
- *Dati della macchina*: Forniscono informazioni sul modello della macchina usata e sulla programmazione dell'erogazione.
  - Macchina: indica il modello di macchina su cui è stata effettuata l'erogazione. L'unico valore assunto è "CIMBALI M100 Dosatron".
  - Gruppo usato: indica il gruppo macchina da cui è stata effettuata l'erogazione. Assume solo due valori: "GRUPPO 1" e "GRUPPO 2";
  - Dose(singola o doppia): indica il tasto premuto per compiere l'erogazione. L'unico valore presente è "TASTO 2", corrispondente a dose doppia.
  - Dose programmata (colpi di ventolino) CdV=0,5ml: indica il numero di colpi di ventolino impostati dalla macchina. Ogni colpo di ventolino corrisponde a 0,5ml di acqua erogata. È il parametro legato alla taratura della macchina. L'unico valore assunto è 177.
- *textitInformazioni sulla miscela*: Forniscono informazioni sulla miscela utilizzata e su come essa sia stata pressata.

- Pressatura: indica come il macinato è stato pressato. L'unico valore assunto è "Tarata".
- Miscela: indica la miscela utilizzata per svolgere le erogazioni. L'unico valore assunto è "TOP CLASS". Viene quindi utilizzata un'unica miscela.
- *Rilevazioni esterne*: sono dati misurati esternamente grazie a una bilancia oppure a una lettura dal display della macchina.
  - Peso dose macinato(gr): peso del macinato utilizzato per effettuare l'erogazione, misurato con un bilancino;
  - Temperatura caldaia(Bar) da display: temperatura della caldaia espressa in bar, rilevata dal display della macchina;
  - Pressione(bar) a vista: pressione della caldaia, espressa in bar;
  - Tempo erogazione (sec): tempo di erogazione misurato attraverso un cronometro.
  - Peso erogato in tazza(gr): peso del caffè erogato, riferito alla dose doppia e misurato con una bilancia;
  - Erogato in tazza singola(ml): millilitri di caffè erogato per una singola tazza. Calcolato a partire da "Peso erogato in tazza(gr)", diviso per due e moltiplicato per 1.02, il peso specifico del caffè.
- *Indici calcolati dalle rilevazioni manuali*: mettono in relazione le varie rilevazioni esterne effettuate, con l'obiettivo di fornire maggiore conoscenza.
  - Ratio(peso dose/peso erogato): indica la relazione tra il peso della dose di macinato e il peso della quantità erogata in tazza.
  - Flusso medio macchina (dose programmata ml / tempo erogazione sec): indica il flusso medio calcolato a partire dalla quantità di acqua erogata della macchina.
  - Velocità del flusso in tazza (peso erogato/tempo erogazione): flusso medio in tazza misurato come g/ml riferito alla quantità di erogato per la doppia dose.
  - Velocità flusso (ml/Sec): indica il flusso in tazza calcolato in ml/s e riferito alla singola tazza.
- *Confronto con le soglie*: confronto dei valori di flusso, tempo di erogazione e erogato in tazza rispetto alle soglie definite
  - SOGLIA FLUSSO ml/sec (su dose singola): flusso in tazza calcolato come "Erogato in tazza singola(ml)" diviso "Tempo erogazione";
  - SOGLIA FLUSSO (per calcolo): indica l'etichetta assunta dall'erogazione rispetto alle soglie di flusso definite. Può assumere solo tre valori: MINORE, OK e MAGGIORE.
  - TEMPO DI EROGAZIONE 20/27 sec: tempo di erogazione calcolato dalla scheda PC della macchina, calcolato in secondi. Corrisponde all'attributo "Tempo di erogazione";

- TEMPO DI EROGAZIONE (per calcolo): indica l’etichetta assunta dall’erogazione rispetto alle soglie di tempo di erogazione definite. Può assumere solo tre valori: MINORE, OK e MAGGIORE.
- Erogato in tazza 18ml /30 ml: millilitri di erogato in tazza misurati sulla singola dose. Corrisponde a “Erogato in tazza singolo(ml)”;
- EROGATO IN TAZZA (per calcolo): indica l’etichetta assunta dall’erogazione rispetto alle soglie di tempo di erogazione definite. Come per le etichette precedenti assume solo i valori MINORE, OK e MAGGIORE.
- *Dati Scheda PC*: misurazioni rilevate dalla scheda PC, ovvero direttamente dalla centralina della macchina.
  - Tempo erogazione: misura del tempo di erogazione in secondi;
  - Impulsi flussimetro: numero di colpi di ventolino effettuati durante l’erogazione. Ogni colpo di ventolino corrisponde a 0,5 ml di acqua. Dipende dalla taratura della macchina ed ha un range di variazione molto ridotto;
  - Quantità erogata(ml): quantità di acqua erogata riferita alla singola dose: è calcolata come “Impulsi flussimetro” per 0,5 ml, diviso 2, così da avere un dato relativo alla singola dose in tazza;
  - Flusso acqua (ml): flusso riferito all’acqua erogata dalla macchina. È calcolato come “Quantità erogata(ml)” diviso “Tempo di erogazione”;
  - Calcolo variazione tempo estrazione (display Vs. macchina): differenza tra il tempo rilevato esternamente e quello della scheda PC. Utile per l’allineamento dei dati delle due fonti.
  - Calcolo della variazione su quantità erogata (ml/dose in tazza singola): rapporto tra la quantità di acqua totale erogata per la scheda PC e la quantità in tazza singola rilevata esternamente.
- *Diagnostica macchina* informazioni relative ai lavaggi
  - Lavaggi: indica se è stato effettuato o meno il lavaggio giornaliero sul gruppo. Può assumere solo i valori "Si a fine giornata" e "No";
  - Purge (prog. 4sec): indica se è stato effettuato un lavaggio breve dopo l’erogazione. Può assumere solo i valori "Si" e "No";
  - Lavaggi Flusso: indica un tipo di lavaggio con coppetta ceca. Non riporta alcun valore.
- *Messaggi*: riporta eventuali messaggi di errore della macchina o del device. Non assumono mai valori validi.
- *Indici calcolati dalla scheda PC*: indici calcolati partendo dai dati della scheda PC che cercano di aumentare le informazioni disponibili.
  - Flusso acqua dai dati macchina: flusso calcolato a partire dai dati della scheda PC come "Quantità erogata(ml)" diviso "Tempo erogazione". Ripetizione del campo "Flusso acqua (ml);
  - Soglia flusso dati macchina: applicazione delle soglie del flusso in tazza al dato della scheda PC. Assume valori "MINORE", "OK" e "MAGGIORE".

## 4.2.2 File modificato

Nella sezione 6.1.1, viene approfondito lo studio delle variabili in gioco e le relazioni tra esse. Durante l'analisi si è svolta anche un'azione di pulizia degli attributi considerati poco importanti, poichè aventi sempre stesso valore oppure calcolati a partire da altri dati. In questo modo si ottimizza sia le analisi successive, che riescono a elaborare meglio i dati, sia le procedure di collezione dei dati, che ne risultano velocizzate.

In particolare sono stati eliminate diverse colonne:

- "Tempo erogazione (sec)": una volta vista la coerenza del dato proveniente dalla scheda Pc con quello rilevato esternamente si è optato per l'eliminazione della rilevazione del tempo attraverso cronometro.
- "Indici calcolati dalle rilevazioni manuali": eliminati in quanto calcolati a partite da altri dati rilevati;
- "Calcolo variazione tempo estrazione (display Vs. macchina)" e "Calcolo della variazione su quantità erogata (ml/dose in tazza singola)", in quanto non portano informazioni utili.
- "Lavaggi flusso" e "Messaggi" poichè non riportano alcun valore.
- "Indici calcolati dalla scheda PC", in quanto ripetizioni di altri attributi.

Sono stati mantenuti invece le informazioni sulla macchina utilizzata e sulle caratteristiche dell'erogazione nonostante assumano sempre lo stesso valore in quanto sono utili nel caso di esperimenti con macchine, tarature e miscele diverse.

## 4.2.3 File per regole di associazione

Per l'utilizzo delle regole di associazione è necessario conoscere le caratteristiche di dose, macinatura e pressione con cui sono state svolte le varie erogazioni. Infatti ogni erogazione appartiene a uno dei 27 test che rappresentano tutte le possibili combinazioni dei tre parametri esterni.

Il dataset usato per l'analisi con le regole di associazione presenta allora tre ulteriori colonne che specificano, per ogni erogazione, il valore delle 3 variabili esterne. In particolare:

- Macinatura: indica il tipo di macinatura usata. Può assumere 3 valori:
  - -2 tacche (macinatura fine);
  - ottimale (macinatura usata durante la taratura);
  - +2 tacche (macinatura grossolana).
- Dose: indica il peso del macinato utilizzato per svolgere l'erogazione. Può assumere solo tre valori:
  - -2 grammi: rispetto all'ottimale è utilizzata una dose bassa di macinato;
  - ottimale, ovvero 14 grammi per la dose doppia;

- +2 grammi: rispetto all’ottimale è utilizzata una dose bassa di macinato.
- Pressione erogazione: indica la pressione della caldaia della macchina. Può assumere solo tre valori:
  - -3 bar: pressione bassa;
  - ottimale, ovvero 9 bar;
  - + 2 bar, pressione alta.

## 4.3 Seconda sessione sperimentale

La seconda sessione sperimentale ha la stessa impostazione della prima, con la differenza che mira a modellare delle variazioni più lievi dei fattori esterni. Essa presenta nuovamente 27 test che modellano le varie combinazioni dei fattori esterni: ognuno è composto da 20 erogazioni, ad eccezione del primo test che presenta 60 osservazioni, in quanto è svolto con i valori ottimali dei fattori esterni e permette così uno studio della situazione ottimale iniziale più approfondito. Inoltre, i test sono effettuati su entrambi i gruppi che presentano però condizioni di pulizia diverse. Infatti solo su uno dei due vengono effettuate i lavaggi giornalieri e le purge. In questo modo si vuole studiare una possibile differenza dovuta al diverso stato di manutenzione della macchina.

Un ulteriore obiettivo di questa sessione è studiare la correlazione tra dati di telemetria, dati della scheda PC e valori rilevati in tazza. Infatti, grazie alla macchina Cimbali M100 su cui è stato svolto l’esperimento, è possibile rilevare i dati di entrambe le fonti che prendono dati da sensori. La sezione 6.2 affronta nello specifico questo tema.

La struttura del dataset è molto simile a quella analizzata per la precedente sessione sperimentale, con qualche lieve differenza:

- Temperatura e pressione da display non vengono rilevate da display in quanto costanti oppure modificate volontariamente a causa delle specifiche del test;
- Sono presenti le rilevazioni della telemetria, per cui sono state predisposte le seguenti colonne:
  - tempo erogazione telemetria: tempo di erogazione rilevato dal sensore di telemetria, espresso in secondi;
  - calcolo variazione tempo estrazione: differenza tra i tempi di erogazione rilevati da telemetria e scheda PC. È molto utile per effettuare l’allineamento tra dati di scheda PC e telemetria;
  - Impulsi flussimetro telemetria: numero di colpi di ventolino calcolati a partire dalla quantità di acqua erogata per dose singola. Viene calcolato come “Quantità erogata telemetria(ml)” per 4, poiché riferita a dose doppia e ogni colpo di ventolino vale 0,5 ml;
  - Quantità erogata telemetria(ml): quantità di acqua in millilitri riferita alla dose singola, calcolata a partire dal flusso della telemetria per la dose doppia. Viene calcolata come “Flusso acqua telemetria(ml/s)” /2 per “Tempo erogazione telemetria”, così da essere riferita alla singola dose;

- Calcolo della variazione sulla quantità erogata: rapporto tra “Quantità erogata telemetria(ml)” e “Quantità erogata(ml)” calcolata dalla scheda PC, espresso in percentuale;
- Flusso acqua telemetria(ml/s): flusso medio rilevato dalla telemetria riferito alla dose doppia, espresso in millilitri al secondo;
- Flusso acqua telemetria per singola dose(ml/s): flusso medio rilevato dalla telemetria riferito alla dose singola. Calcolato come  $\text{”Flusso acqua telemetria(ml/s)”}/2$  ;
- Calcolo della variazione su flusso acqua: Rapporto tra “Flusso acqua telemetria per singola dose(ml/s)” e “Flusso acqua (ml/s)”, espresso in percentuale.

Il dataset presenta quindi 1160 righe e 30 colonne. É da notare però come manchino 437 valori rilevati dalla telemetria. Come confermato dall’analisi dei dati dal mercato della sezione 6.6, la mancanza di alcuni dati da parte del device di telemetria è dovuta alla pulizia effettuata dalla telemetria del costruttore Cimbali sui dati delle erogazioni.

## 4.4 Sessione sperimentale proposta

Dopo le prime due sessioni sperimentali, sono stati definiti alcuni obiettivi da modellare in laboratorio. In particolare, si vuole creare una sessione di test da poter effettuare su macchine diverse così da rilevare eventuali differenze da tenere conto una volta iniziata l’analisi dei clienti sul mercato. Sono definiti 5 tipi di test mirati a modellare comportamenti diversi, la cui composizione e ordine di esecuzione è stata decisa sulla base delle osservazioni fatte sui primi esperimenti.

Le specifiche e le motivazioni di ogni test sono descritte nella sezione 6.5. Di seguito si riportano le cardinalità previste per ogni sessione di test. Le colonne rilevate corrispondono a quelle della seconda sessione sperimentale. Si ricorda però che in certi modelli di macchina non è possibile rilevare contemporaneamente i dati della scheda PC e della telemetria. In questo caso la preferenza cade sui dati di telemetria, in quanto più vicini al caso reale delle macchine sul mercato.

Il numero di erogazioni attese per ogni test, considerando che il test si svolga su macchine a due gruppi, è:

- Differenza intrinseca tra i gruppi: 60 erogazioni ottimali per ogni gruppo, per un totale di 120 erogazioni.
- Relazione tra colpi di ventolino e erogato in tazza: 10 test da 10 erogazioni per gruppo, per un totale di 200 osservazioni. Il numero totale di osservazioni potrebbe essere ridotto se il numero di colpi di ventolino si allontana troppo da quello ottimale;
- Modellazione dei lavaggi anomali: 60 erogazioni ottimali senza lavaggi, più 6 test da 10 rilevazioni ciascuno, effettuati per ogni gruppo, per un totale di 240 osservazioni;
- Degrado dovuto alla mancanza di lavaggi: 60 erogazioni ottimali per avere il confronto iniziale, 500 erogazioni per accumulare un degrado e altrettante per vedere le differenze in presenza di lavaggi. In totale, la somma delle erogazioni dei due gruppi è 2120.

Inoltre, le prime 500 erogazioni potrebbero non essere sufficienti per accumulare un degrado delle prestazioni, e potrebbe essere incrementate le erogazioni senza lavaggi fino al rilevarsi di un peggioramento delle prestazioni. Inoltre solo per una piccola parte della totalità delle erogazioni è rilevato il dato in tazza, così da velocizzare gli esperimenti;

- Variazione dei parametri esterni: 19 test composti da 20 erogazioni per modellare le combinazioni delle variazioni dei parametri esterni (dose e macinatura), più 60 erogazioni iniziali ottimali, per un totale di 440 erogazioni per gruppo, ovvero 880 complessive.

## 4.5 Dati provenienti dalla telemetria

Nella seconda parte dello studio, si è effettuata l'analisi di dati provenienti dal mercato, ovvero rilevati tramite telemetria. Essi sono scaricati dal portale della telemetria della Lavazza che permette la selezione di un periodo di massimo 41 giorni su una delle macchine supervisionate.

Attraverso il portale, è possibile selezionare quali informazioni scaricare in un file excel. Nella nostra analisi sono state prese in considerazione solo le rilevazioni di flusso e tempo per ogni erogazione. Il file scaricato però non è in un formato comodo per proseguire l'analisi. Infatti esso presenta un foglio per ognuna delle caratteristiche selezionate, divise per gruppo di erogazione. In pratica, per una macchina con 2 gruppi di cui si è interessati ad analizzare tempo di erogazione e flusso, si otterrà un file avente 4 fogli, due con i tempi di erogazione dei due gruppi e due con i flussi nei due casi.

In particolare, ogni foglio presenta 3 colonne di interesse: la data di inizio della rilevazione, data di fine della rilevazione e il dato dell'erogazione, che varia a seconda che si tratti di flusso o tempo.

Teoricamente, per ogni erogazione dovrebbero essere rilevati sia il tempo di erogazione che il flusso medio. In alcuni casi però solo una delle due è rilevata, in genere in casi in cui l'erogazione sia molto corta o molto lunga. Una descrizione più approfondita è fornita alla sezione 5.2.2.

Per procedere con l'analisi è molto avere tempo e flusso associate allo stesso orario di rilevazione. È stato allora scritto un programma che ricevuto il file scaricato dal portale, restituisce un nuovo file con tutte le erogazioni svolte sulla macchina, descritte con il seguente formato:

- Data: data e ora di rilevazione dell'erogazione da parte della telemetria;
- Gruppo: gruppo della macchina su cui è stata rilevata l'erogazione;
- Tempo erogazione telemetria: tempo di erogazione misurato dalla telemetria in decimi di secondo (Es. 154= 15,4 secondi)
- Flusso telemetria: flusso medio calcolato dalla telemetria in ml/s.

In questo modo si hanno a disposizione immediatamente tutte le informazioni, e si risolve il problema dell'allineamento manuale che richiede molto tempo e attenzione. Inoltre vengono scartate automaticamente quelle erogazioni aventi solo il tempo o il flusso rilevati.

## 4.6 File degli interventi

Tutti i dati degli interventi tecnici sono contenuti in un file Excel, che ha in ogni riga varie informazioni riguardo a un intervento. Ogni intervento da uno stesso cliente, può essere composto da più righe, in base al numero di sostituzioni effettuate e ai costi dei vari servizi offerti con l'intervento.

Le informazioni contenute in questo dataset sono davvero molte, alcune di scarsa importanza al fine della nostra analisi. Di seguito, si riportano le informazioni di maggior interesse:

- Name: indica il nome del cliente da cui è stato effettuato l'intervento;
- Creation name: data in cui la richiesta di intervento è stata avviata. In genere essa è fatta dal cliente;
- Product: indica il modello della macchina su cui è stato effettuato l'intervento;
- Serial number e Manufactural Serial Number: numeri che identificano il dispositivo. Attraverso uno di questi due valori si può individuare il dispositivo sul portale di telemetria. In molti casi la macchina non è dotata del device di telemetria ed è quindi impossibile effettuare il matching;
- Category: classe dell'intervento, ad esempio "Manutenzione Programmata", "Installazione";
- TS Component: fino a un massimo di tre colonne contenenti i componenti su cui si è intervenuti, ad esempio "Doccette", "Guarnizione",...;
- TS Activity: indica l'attività svolta durante l'intervento in relazione al componente prima descritto;
- Intervention Day: data in cui è avvenuto l'intervento.
- Total Cost: costo dell'intervento descritto dalle colonne sopra citate.

L'analisi esplorativa degli interventi tecnici parte dagli interventi più costosi: il file viene quindi compattato in base al cliente del cliente, al numero identificativo della macchine e alla data dell'intervento, così da ottenere il costo totale degli interventi.

# Capitolo 5

## Approccio sperimentale

In questo capitolo saranno approfondite le tecniche di analisi utilizzate. In particolare nella sezione 5.1 viene introdotto Python, che ha permesso l'implementazione di tutte le successive tecniche attraverso le sue numerose librerie. Nella sezione successiva si elencano le tecniche di pulizia utilizzate per lavorare con dei dati consistenti. Nella sezione 5.3 si parla invece degli indici statistici che caratterizzano un insieme di dati, e di come possono essere visualizzati in python. In seguito si forniscono alcune informazioni sulla regressione lineare, tecnica utilizzata per verificare la consistenza dei dati di telemetria e per calcolare la loro relazione con quella in tazza. Il capitolo si conclude, infine, con l'approfondimento delle regole di associazione, utili per estrarre dei comportamenti tipici dai dati.

### 5.1 Python e Jupyter Notebook

Per svolgere le analisi, si è deciso di usare l'applicazione web Jupyter Notebook[17]. È un ambiente interattivo basato su web-services per creare dei documenti JSON detti Jupyter Notebook. Ogni documento è composto da una serie ordinata di celle di input e output che possono contenere codice, testo, grafici e formule matematiche. Esso risulta molto utile in quanto permette l'esecuzione di singole celle senza dover rieseguire tutto il programma. Inoltre offre un servizio di conversione e download in diversi formati standard come HTML, PDF, LaTeX.

Si è usato Python come linguaggio di programmazione perchè permette una gestione e analisi molto efficiente di dataset, in particolar modo se in formato Excel. Inoltre presenta numerosissime librerie per quanto riguarda la data analysis. Tra queste librerie, sono state usate le seguenti:

- Pandas[14]: una libreria open source molto semplice da usare che mette a disposizione strutture dati molto efficienti dette dataframe, oltre a tool di analisi molto utili. In particolar modo, permette una gestione semplice di file Excel e tabelle.
- Numpy: pacchetto fondamentale per l'elaborazione di dati con Python. Utile per la gestione di vettori e matrici multidimensionali, con funzioni matematiche di alto livello da applicare su di loro;

- Matplotlib[12]: libreria grafica di Python che permette la creazione di grafici. Il suo utilizzo è descritto nella sezione 5.3.2.
- Sklearn[18]: mette a disposizione numerosi algoritmi di classificazione, regressione e clustering, come support vector machines, random forest e k-means, e permette l'interazione con le librerie scientifiche Numpy e Scipy.

## 5.2 Tecniche di pulizia e allineamento dei dati

La pulizia dei dati è uno degli aspetti più importanti per l'analisi di un dataset: essa infatti si occupa di eliminare valori spuri, al di fuori del dominio e non accettabili, che comprometterebbero l'analisi inserendo elemento di rumore nei dati. Esso è sempre necessario, così da eliminare quei valori che si allontanano dall'andamento standard.

Inoltre esso è necessario anche per modificare la struttura dei dati così che possano essere analizzati in maniera più efficace.

All'interno della nostra analisi, sono state rilevate numerose rilevazioni anomale, oltre a mancanza di osservazioni e disallineamento tra esse. Infatti i dati sono raccolti da risorse umane e device di telemetria in fase di test, e quindi gli errori possono verificarsi spesso.

In particolare sono state individuate tre aree in cui è necessaria una pulizia o un allineamento dei dati:

- Pulizia dei dati sul singolo test in laboratorio;
- Pulizia e allineamento preliminare dei dati di telemetria
- Pulizia delle erogazioni delle macchine in produzione.

### 5.2.1 Pulizia dei dati sul singolo test in laboratorio

Durante le sessioni sperimentali in laboratorio, sono stati svolti diversi test con condizioni di macinatura, dose e pressioni diverse. Analizzando l'andamento della quantità di erogato in tazza, si è notato che spesso la prima erogazione di ogni test presenta una quantità di erogato molto ridotta.

Le cause di questa rilevazione anomala potrebbe avere due cause:

- Errore di trascrizione del dato rilevato da parte dell'operatore;
- Adattamento della macchina alle nuove condizioni di dose, macinatura e pressione.

Allora si è deciso, per ogni test, di eliminare l'erogazione con erogato in tazza minore, e per omogeneità, anche quella con erogato maggiore all'interno del test.

### 5.2.2 Pulizia e allineamento dei dati di telemetria

Il device di telemetria di Lavazza è stato prodotto da poco ed è nei primi mesi di test. Esso quindi è soggetto a errori di rilevazione. In particolare, prima di questi test, non si aveva testato in maniera approfondita la sua interazione con le diverse macchine.

Analizzando i dati ricavati dal portale della telemetria, si sono riscontrati alcuni problemi di allineamento:

- Non tutte le erogazioni effettuate dalla macchina sono rilevate dal sensore di telemetria: ciò emerge sia dalla mancanza di alcune erogazioni svolte in laboratorio, sia confrontando i dati di vendita e quelli di telemetria su macchine sul mercato. Inoltre, questo fenomeno non si verifica su tutte le macchine, ma sembra essere predominante sulle macchine del gruppo Cimbali, che hanno un device di telemetria del costruttore che filtra alcune erogazioni aventi parametri di rilevazioni anomali o comunque fuori soglia.
- Alcune erogazioni risultano mancanti poichè "accorpate": negli esperimenti in laboratorio è stato rilevato che in alcuni casi i tempi di erogazione e i flussi di due erogazioni sono sommate, forse a causa di un intervallo di tempo molto ridotto tra le due. In questo modo, il numero totale di erogazioni rilevate risulta ridotto;
- In alcuni casi particolari, per una singola erogazione è rilevata solo una delle due caratteristiche principale(tempo e flusso). Ciò crea una disparità di numero nelle rilevazioni di tempo e flusso per uno stesso gruppo, complicando le operazioni di allineamento;
- Su alcune macchine in produzione, i tempi di rilevazione di tempo e flusso di una stessa erogazione risultano differenti di alcuni secondi, rendendo quindi impossibile un match tra le due. Il problema sembra però essere stato risolto con l'aggiornamento del software di telemetria di settembre 2019.

Come descritto nella sezione 4.5, i dati scaricati dal portale di telemetria sono forniti in un file excel avente due fogli per ogni gruppo, uno per tempo di erogazione e l'altro per il flusso. Il dato, in questo formato, è difficilmente utilizzabile. Il dato diventa utile quando alla stessa ora di rilevazione vengono associate tempo e flusso dell'erogazione, oltre all'informazione del gruppo su cui essa è stata svolta. L'allineamento del dato non è però così banale a causa delle incoerenze appena descritte. È stato quindi ideato uno script che, basandosi sull'orario di rilevazione di una erogazione in un gruppo, associa i dati di tempo di erogazione e flusso. In questo modo, oltre all'allineamento delle informazioni di telemetria, si eliminano anche quelle erogazioni di cui una delle due rilevazioni manca.

### 5.2.3 Pulizia delle erogazioni delle macchine in produzione

I dati di telemetria, purtroppo, non presentano l'etichetta relativa al tasto premuto dal cliente per effettuare la singola erogazione. Per superare questa mancanza e procedere comunque con le analisi è stato necessario effettuare un processo di pulizia dei dati.

I dati di telemetria, sono in genere molto variabili. Ciò è dovuto alla varietà di erogazioni possibili: normale, corta, lunga e continua. Ognuna di esse poi può essere effettuata in dose singola o doppia, in base al numero di caffè necessari al momento. Solitamente, i tipi di erogazione più utilizzati sono due: l'erogazione normale singola e quella normale doppia. La pulizia è stata effettuata grazie ad una analisi dei dati di telemetria: il flusso medio e il tempo di erogazione.

Sono stati proposti due approcci: il primo si basa sull'analisi dell'andamento delle due variabili dell'erogazione, in modo da definire delle soglie per ciascuna di esse. Le soglie sono decise in maniera manuale, caso per caso, in modo da includere il picco più alto nella distribuzione delle frequenze, che rappresenta le erogazioni più numerose: in questo modo i

dati, dopo la pulizia, dovrebbero corrispondere quasi solamente a erogazioni normali singole o normali doppie. Si presentano quindi due possibilità:

- Eliminare tutte le erogazioni che presentano almeno una delle due variabili fuori soglia. In questo modo potrebbero però essere eliminate anche erogazioni valide che però presentano fattori esterni non ottimali, come una macinatura grossolana o na dose di macinato inferiore;
- Eliminare solo le erogazioni che presentano entrambe le variabili fuori soglia. In questo modo però il numero di erogazioni eliminato è molto ridotto.

Si è allora deciso di seguire un secondo approccio basato sull'analisi della quantità di acqua erogata secondo la telemetria, ottenuta moltiplicando le due variabili rilevate. Infatti, l'obiettivo della macchina è quello di erogare sempre la stessa quantità di acqua impostata in fase di taratura. Di conseguenza, i diversi tipi di erogazione presenteranno quantità di acqua diverse.

Sono state quindi decise delle soglie sulla quantità di acqua erogata osservando la distribuzione delle frequenze, e si sono eliminate le erogazioni avente valore fuori soglia. Questo metodo si è ricavato particolarmente efficace in quanto spesso un gruppo viene utilizzato per effettuare sempre lo stesso tipo di erogazione: gli esperti di dominio hanno confermato che solitamente a ciascun gruppo è associato un portafiltro singolo oppure doppio, in particolare in macchine a tre gruppi. Questa informazione è confermata dai dati del mercato: alcuni clienti presentano infatti macchine con gruppi che erogano quantità di acqua differenti, in particolare uno dei due gruppi ha un erogato quasi dimezzato rispetto all'altro. In questo modo è possibile discriminare gruppi usati per caffè doppi da quelli usati per caffè singoli. Le conclusioni a cui si è giunti sono validate anche attraverso il confronto con i dati di vendita, che mostrano una corrispondenza tra numero di caffè singoli e doppi venduti e rilevazioni considerate singole e doppie.

Inoltre, il metodo di pulizia descritto è automatizzabile in quanto, individuato il picco nelle frequenze, si possono eliminare tutte quelle erogazioni al di fuori di un range dal valore più o meno stretto, a seconda dei casi. Una buona proposta potrebbe essere quella di mantenere le erogazioni aventi erogato nella fascia  $quantità_{picco} \pm 2ml$ .

### 5.3 Caratteristiche di un insieme di dati

Quando si hanno a disposizione numerosi dati è impossibile analizzare singolarmente ogni rilevazione e si ricorre allora a degli indicatori statistici utili per evidenziare le caratteristiche principali dei dati e identificare rilevazioni anomale. Inoltre permettono anche di comunicare in maniera efficace le informazioni contenute in quei dati.

Una volta ottenuto un campione di dati, si può calcolare la sua *distribuzione di frequenza*, in cui viene calcolato quanto spesso un certo valore, o un certo range di valori, viene rilevato all'interno del campione. La rappresentazione delle varie frequenze attraverso un istogramma, permette di capire quale sia l'andamento delle rilevazioni, quali siano i range di valori più frequenti e quali invece siano più rari.

Partendo dalla distribuzione dei valori, possiamo individuare alcuni elementi caratterizzanti, come il valore massimo e quello minimo. Inoltre possiamo individuare i *quartili*, ovvero i tre valori che dividono l'intero insieme in 4 gruppi, ognuno contenente il 25% delle

osservazioni. Il quartile centrale prende il nome di *mediana*, ed è quindi il valore che divide un campione di dati ordinati in due parti ugualmente numerose. Essa ci dà un'informazione della tendenza centrale e risente in modo limitato di valori estremi. Con *distanza interquartile* si indica invece la differenza tra i due quartili esterni, e identifica l'intervallo centrale della distribuzione di frequenza all'interno del quale cade il 50% delle osservazioni.

Prendendo in analisi variabili di tipo quantitativo[9], possiamo calcolare la *media* del campione in analisi:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

con  $N$  numero totale di rilevazioni e  $x_i$  osservazione  $i$ -esima.

Partendo dalla media, possiamo individuare due misure di dispersione basate sulle differenze tra le singole osservazioni e la media stessa:

- Varianza: media della deviazioni dal valore medio, al quadrato.

$$Var(X) = \frac{1}{N} \sum_{i=0}^n (x_i - \bar{x})^2$$

con  $N$  numero totale di rilevazioni e  $x_i$  osservazione  $i$ -esima.

- Deviazione standard: radice quadrata della Varianza, permette il confronto diretto con la media

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{N} \sum_{i=0}^n (x_i - \bar{x})^2}$$

con  $N$  numero totale di rilevazioni e  $x_i$  osservazione  $i$ -esima.

Grazie a questi indici statistici si riesce così a ottenere una panoramica generale del campione.

### 5.3.1 Relazione tra due insiemi di dati

Come abbiamo visto nella sezione precedente, si possono ricavare molte informazioni a partire da un insieme di dati. Siamo adesso interessati a degli indicatori che descrivano come due insiemi di osservazioni siano legati tra loro[7]. Considerando quindi i nostri due insiemi come variabili aleatorie  $X$  e  $Y$ , la statistica ci viene incontro con due indicatori:

- Covarianza:

$$Cov(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)$$

con  $\mu_X, \mu_Y$  medie delle due variabili;

- Correlazione:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

con  $\sigma_X, \sigma_Y$  deviazioni standard delle variabili  $X$  e  $Y$ ;

La *covarianza* ci indica se le due variabili hanno un comportamento concorde o discorde. In particolare, un valore positivo significa che le due variabili hanno lo stesso andamento crescente o decrescente, anche se non necessariamente della medesima intensità. Con un valore negativo, invece, si ha un comportamento opposto. Infine, un valore prossimo allo zero indica che i due insiemi non sono in relazione tra di loro. Il principale problema della correlazione è il fatto che dipende in maniera significativa dalla scala di misura utilizzata: con scale molto piccole la covarianza risulta prossima allo zero anche se le due variabili sono legate tra loro. Inoltre presenta un'unità di misura diversa da quella di partenza, ed è quindi difficilmente confrontabile.

La *correlazione* è la soluzione a questo problema, in quanto è un numero puro, adimensionale: inoltre può assumere valori solo nell'intervallo  $[-1,1]$ . In questo modo si riesce a vedere in maniera univoca e immediata se due variabili sono correlate positivamente o negativamente.

### 5.3.2 Grafici in Matplotlib

Matplotlib è una libreria grafica che mette a disposizione alcune semplici funzioni per visualizzare informazioni a partire dai dati. Di particolare interesse sono le funzioni:

- *hist*: permette la rappresentazione della distribuzione delle frequenze di un insieme di dati, fornendo una rappresentazione visiva dell'andamento dei valori assunti dalla variabile.

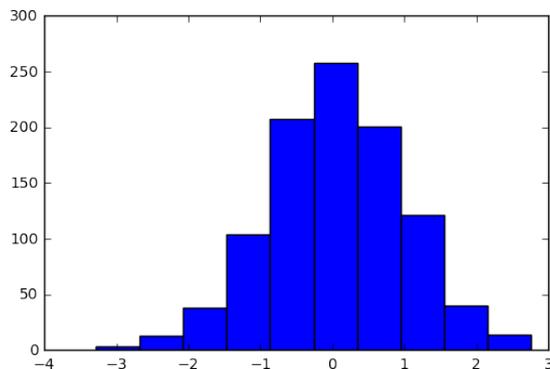


Figura 5.1: Esempio di istogramma

- *boxplot*: mette in risalto la distribuzione dei valori basandosi sulle informazioni dei quartili. In particolare rappresenta con un rettangolo la fascia di valori tra il primo e il terzo quartile, evidenziando quindi la fascia dei valori centrali e mettendo in risalto la mediana. Inoltre presenta dei "baffi" che si prolungano dal rettangolo per il valore della *distanza interquartile* moltiplicato per 1,5. I valori esterni a questo range sono rappresentati come dei punti, e rappresentano gli outliers. Questo tipo di grafico è particolarmente utile per confrontare l'andamento generale di più insiemi di dati.
- *plot*: permette di visualizzare i punti di cui si passano le coordinate x e y. Molto utile per mettere in relazione attributi diversi di uno stesso dato.

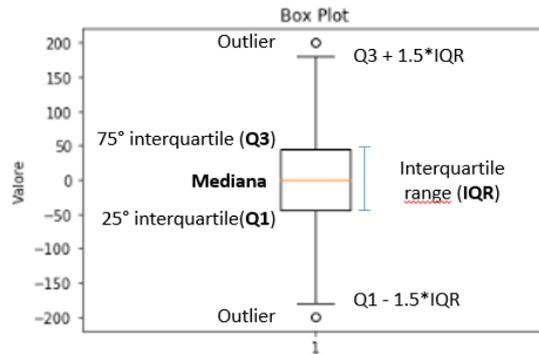
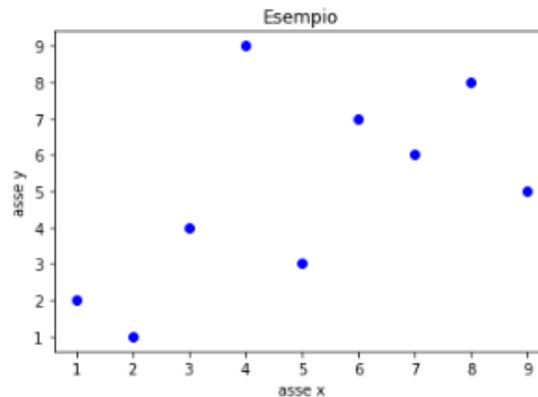


Figura 5.2: Esempio di boxplot e le sue caratteristiche

Figura 5.3: Esempio di grafico con *plot*

## 5.4 Regressione lineare

Come descritto nella sezione precedente, due variabili possono essere tra loro correlate, ovvero possono avere un trend in qualche modo legato. In particolare, due variabili sono fortemente correlate se l'indice di correlazione ha un valore prossimo a 1 (in valore assoluto). Ciò vuol dire che le variabili hanno una relazione lineare tra loro. Un metodo molto semplice per visualizzare questa relazione è quello di visualizzare i dati con uno scatter plot, dove x e y assumono il valore delle due caratteristiche per ogni punto.

In questo modo i dati sono plottati come punti, e assumono una certa distribuzione: in caso di relazione lineare, la distribuzione assume la forma di una retta, come in figura 5.4.

Poichè si ha quindi la sicurezza che i dati siano correlati e presentino un andamento lineare, è possibile applicare la regressione lineare.

La regressione lineare è un metodo che cerca di calcolare la retta di regressione tra le due variabili in esame, fornendo un'equazione nel formato

$$y = \alpha + \beta x$$

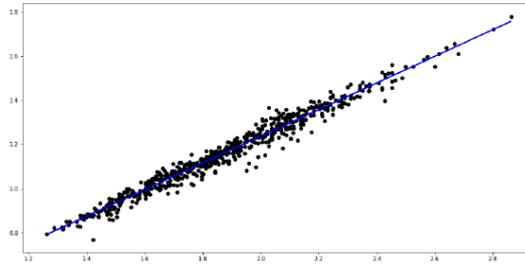


Figura 5.4: Esempio di regressione lineare

con  $x$  e  $y$  variabili indipendente e dipendente rispettivamente,  $\alpha$  il punto di intersezione con l'asse  $y$  e  $\beta$  la pendenza della retta.

Nel nostro caso, il metodo con cui la migliore retta di regressione è calcolata è il *OLS* (*Ordinary Least Square*), ovvero il metodo dei quadrati minimi, che minimizza la somma dei quadrati delle distanze tra i dati presi in considerazione e la curva che si sta cercando di ottenere. Quindi i valori di  $\alpha$  e  $\beta$  sono quelli che minimizzano queste distanze. È importante considerare le distanze al quadrato in modo tale che gli errori positivi e negativi non si compensino a vicenda.

Per valutare la bontà della retta di regressione trovata, si fa riferimento al *coefficiente di determinazione*[10], anche detto *r-squared* ( $R^2$ ). Esso misura la frazione della variabilità delle osservazioni che si riescono a predire attraverso il modello lineare appena creato. È un numero compreso tra 0 e 1, e più è alto più è possibile tenere in considerazione la variabilità dei dati appena analizzati.

Grazie alla retta di regressione è quindi possibile trovare una funzione di trasferimento che, data una variabile, calcola il valore dell'altra.

## 5.5 Regole di associazione

Le regole di associazione[6] permettono di estrarre in modo esaustivo patterns frequenti da un database transazionale. Ogni transazione è composta da un insieme di oggetti non ordinati. Un classico esempio di transazione è uno scontrino, che contiene l'insieme di prodotti comprati. L'insieme di questi oggetti è detto *itemset*.

Una regola di associazione si presenta nel formato:

$$A \Rightarrow B$$

con  $A$  e  $B$  itemsets, ovvero un insieme composto da uno o più oggetti.  $A$  è chiamato *corpo della regola*, mentre  $B$  è nominata *Testa della regola*. La regola così composta indica che, in un dataset transazionale, quando si trova  $A$ , si trova anche  $B$ . Si presti particolare attenzione al fatto che il simbolo  $\Rightarrow$  non indica causalità, bensì il fatto che corpo e testa della regola occorrono assieme. Ad ogni regola possono essere associati alcuni indici di qualità:

- *Supporto*: percentuale di transazioni contenenti sia A che B.

$$\text{supporto} = \frac{\#(A, B)}{\#T}$$

con  $\#(A, B)$  numero di transazioni contenenti sia A che B, e  $\#T$  numero totale di transazioni. Indica quanto la regola è frequente, ovvero quanto è supportata dai dati.

- *Confidence*: frequenza di B in transazioni contenenti A.

$$\text{confidence} = \frac{\text{supporto}(A, B)}{\text{supporto}(A)}$$

Indica la "forza" della regola, ovvero qual è la probabilità di avere la testa della regola avendo il corpo della regola.

- *Lift*: indica il grado di correlazione del corpo e della testa della regola.

$$\text{Lift} = \frac{\text{supporto}(A, B)}{\text{supporto}(A)\text{supporto}(B)}$$

In particolare,  $\text{lift}=1$  indica indipendenza statistica, ovvero la regola non è significativa; se il  $\text{lift}$  è maggiore di 1 si ha una correlazione positiva, mentre se è minore di 1 si ha correlazione negativa, ovvero l'occorrenza di A implica una non-occorrenza di B.

Per famigliarizzare con gli indici di qualità della regola, si propone un esempio:

ID	Items
1	Pane, Cola, Latte
2	Birra, Pane
3	Birra, Cola, Pannolini, Latte
4	Birra, Pane, Pannolini, Latte
5	Cola, Pannolini, Latte

Tabella 5.1: Esempio di dataset transazionale

Estraiamo alcune regole e ne calcoliamo gli indici:

- La regola  $\text{Latte} \Rightarrow \text{Pannolini}$  ha

$$\text{supporto} = \frac{\#\{\text{Latte}, \text{Pannolini}\}}{\#\text{trans}} = 3/5 = 60\%$$

$$\text{confidenza} = \frac{\#\{\text{Latte}, \text{Pannolini}\}}{\#\{\text{Latte}\}} = 3/4 = 75\%$$

$$\text{Lift} = \frac{\text{supporto}\{\text{Latte}, \text{Pannolini}\}}{\text{supporto}\{\text{Latte}\}\text{supporto}\{\text{Pannolini}\}} = \frac{3/5}{4/5 * 3/5} = 15/12 = 1,25$$

- La regola  $Pannolini \Rightarrow Latte$  ha

$$supporto = \frac{\#\{Pannolini, Latte\}}{\#trans} = 3/5 = 60\%$$

$$confidenza = \frac{\#\{Pannolini, Latte\}}{\#\{Pannolini\}} = 3/3 = 100\%$$

$$Lift = \frac{supporto\{Pannolini, Latte\}}{supporto\{Pannolini\}supporto\{Latte\}} = \frac{3/5}{3/5 * 4/5} = 15/12 = 1,25$$

Per utilizzare le regole di associazione bisogna quindi convertire un normale dataset in un dataset transazionale: ogni record del dataset diventa una transazione dove ogni oggetto è rappresentato nel formato  $nome\_attributo = valore\_attributo$ .

A questo punto viene usato un algoritmo di estrazione delle regole di associazione, impostando delle soglie per supporto e confidenza, in modo da estrarre solo regole significative e ridurre il tempo di computazione delle regole. Inoltre, le regole ottenute vengono in genere filtrate anche per valori di lift maggiori di 1, o comunque ordinate per valori decrescenti di esse, così da individuare le regole più importanti.

Sono presenti più algoritmi di estrazione di regole di associazione che operano in maniera diversa producendo risultati differenti. Tra questi, i migliori e più utilizzati sono l'algoritmo Apriori, che si basa sul concetto degli itemset frequenti, e l'algoritmo FP-Growth, che utilizza una struttura ad albero in modo da ridurre il numero di letture del database. Nel nostro caso, viene utilizzato l'algoritmo Apriori, che viene approfondito nella prossima sezione.

### 5.5.1 Algoritmo Apriori

Per poter estrarre le regole di associazione, bisogna per prima cosa individuare gli itemset frequenti, ovvero quegli insiemi di elementi che superano la soglia di supporto impostata. Il metodo più intuitivo per calcolare quali siano gli itemset più frequenti è quello di calcolare tutte le possibili combinazioni di elementi e calcolare la loro frequenza all'interno del dataset. Ciò però è molto dispendioso, in quanto con  $d$  item disponibile, si hanno  $2^d$  itemset disponibili.

L'algoritmo a priori si propone di ridurre il numero di itemset da generare basandosi sul principio Apriori, che dice che "se un itemset è frequente, allora anche tutti i suoi sottoinsiemi devono essere frequenti". Infatti il supporto di un itemset non può mai essere maggiore del supporto dei suoi subset.

Si ha un approccio a livelli: a ogni iterazione si estraggono itemset di lunghezza  $k$ . Ogni livello  $p$  composto da due passi principali:

- Generazione dei candidati: si generano i candidati di lunghezza  $k+1$  unendo i vari itemset frequenti di lunghezza  $k$ , e poi si eliminano i candidati di lunghezza  $k+1$  che contengono almeno un  $k$ -itemset che non è frequente.
- Generazione degli itemset frequenti: si scandisce il dataset per contare il supporto dei candidati di lunghezza  $k+1$  rimasti, eliminando quelli che non superano la soglia.

Il numero degli itemset frequenti può però essere molto elevato, e quindi oltre ad essere molto lenta la generazione dei candidati, anche il numero di accessi al dataset per il conteggio del supporto è elevato. In particolare il numero è elevato se si usa una soglia di supporto molto bassa e se il database è molto grande.

Alcune tecniche per velocizzare l'operazione è quella di alzare la soglia di supporto, rischiando di perdere però delle regole interessanti, usare funzioni statistiche per predire le frequenze invece di calcolarle e l'uso di tabelle di hash per velocizzare la ricerca dei candidati.

## 5.5.2 Implementazione in Python dell'algoritmo Apriori

In Python è presente una libreria che mette a disposizione delle funzioni per generare le regole di associazione a partire da un Dataframe di Pandas. La libreria si chiama *mlxtend*, in particolare si usano i moduli *mlxtend.preprocessing* per la conversione della dataframe in database transazionale e *mlxtend.frequent\_patterns* per estrarre le regole.

---

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

dataset = [['Milk', 'Onion', 'Nutmeg', 'Kidney Beans', 'Eggs', 'Yogurt'],
           ['Dill', 'Onion', 'Nutmeg', 'Kidney Beans', 'Eggs', 'Yogurt'],
           ['Milk', 'Apple', 'Kidney Beans', 'Eggs'],
           ['Milk', 'Unicorn', 'Corn', 'Kidney Beans', 'Yogurt'],
           ['Corn', 'Onion', 'Onion', 'Kidney Beans', 'Ice cream', 'Eggs']]

te = TransactionEncoder()
te_ary = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_ary, columns=te.columns_)
frequent_itemsets = apriori(df, min_support=0.6, use_colnames=True)

association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7)
```

---

Il codice riportato[3] permette l'estrazione delle regole di associazione utilizzando Python.

In particolare, dopo aver importato Pandas e i pacchetti necessari di Mlxtend, viene creato un Dataframe, che rappresenta il nostro dataset. Ogni riga è una rilevazione, e ogni colonna un oggetto. Con *TransactionEncoder* si trasforma il dataset in forma transazionale, in cui ogni riga è un insieme di oggetti del tipo *nome\_colonna=valore*. Dopo aver convertito il nuovo dataset in un dataframe (poiché la funzione successiva richiede in input un oggetto del genere), si chiama la funzione *apriori*, che calcola il supporto di tutti gli itemset restituendo solo quelli con supporto maggiore di *min\_threshold*.

A questo punto, con la funzione *association\_rules* sono estratte tutte le regole, a partire dagli itemset frequenti prima calcolati, che superano la soglia *min\_threshold* riferita alla metrica indicata in *metric*, che può essere una tra supporto, confidence e lift.



## Capitolo 6

# Risultati sperimentali

In questo capitolo verranno infatti descritti i vari studi effettuati e i risultati ottenuti. Il presente lavoro è stato diviso in più parti, ognuna delle quali affronta un diverso aspetto del problema, ed è in genere conseguenza dei risultati ottenuti nelle analisi precedenti. Nella sezione 6.1 si sono analizzati i dati a disposizione per capire quali aspetti fossero più importanti e quali legami ci fossero tra essi. Nella sezione 6.2 si analizzano invece i dati della telemetria con l'obiettivo di garantirne la consistenza. Nel capitolo successivo sono state messe in discussione le soglie di qualità fornite dagli esperti di dominio, proponendo dei nuovi valori, in particolare per il flusso, partendo dai dati ottenuti in laboratorio e da quelli in produzione. La sezione 6.4 parla dell'utilizzo delle regole di associazione per studiare l'influenza delle variabili esterne, quali dose, macinatura e pressione, sulla qualità del caffè, e per capire come gruppi diversi possano comportarsi diversamente a causa di differenza intrinseche oppure dovute a una mancanza di lavaggi macchina. Nella sezione 6.5 si sfruttano tutte le conoscenze ottenute nei punti precedenti per definire una sessione di test in grado di modellare un modello di macchina nel migliore dei modi, tenendo in considerazione i vari fattori su cui si vuole porre l'attenzione. Si passa, successivamente, allo studio dei dati provenienti dal mercato: in particolare si cerca di il comportamento di alcuni clienti di riferimento, cercando nuovamente delle soglie di qualità da applicare ai dati della telemetria. Inoltre, alcuni interventi tecnici vengono analizzati, al fine di verificare se l'intervento è stato effettuato e quali modifiche ha causato nei dati.

Come si può vedere dall'impostazione di questo capitolo, prima si analizzano dei dati provenienti da un ambiente controllato come il laboratorio, per poi analizzare informazioni forniteci dalle macchine in produzione. È necessario specificare come sono stati effettuati gli esperimenti in laboratorio: in particolare sono stati svolte due sessioni sperimentali atte a capire l'impatto della variazione di alcuni parametri esterni, quali macinatura dei grani, dose di macinato e pressione della pompa, sulle caratteristiche principali del caffè, ovvero tempo di erogazione, flusso e quantità erogata in tazza. Inoltre si cerca di evidenziare una differenza tra i due gruppi a causa della mancanza dei lavaggi in uno dei due. A tutte le erogazioni sono inoltre assegnate delle etichette di qualità per le rilevazioni di tempo di erogazione, flusso e erogato in tazza, forniteci dagli esperti e messe in discussione, come affrontato nella sezione 6.3. Inoltre, per ridurre il numero di variabili nell'esperimento, si è scelto di svolgere solo erogazioni doppie (le più comuni sul mercato) con l'uso della miscela Top Class, molto usata e che non è troppo influenzabile da fattori esterni.

Sono quindi state svolte due sessioni sperimentali diverse, entrambe svolte sulla *Cimbali M100*, in modo da poter testare diverse variazioni:

- Prima sessione sperimentale: svolta nel mese di marzo, mira a modellare l'effetto delle variazioni estreme dei parametri esterni:
  - *Macinatura*: partendo dalla macinatura impostata durante la taratura della macchina, considerata come ottimale, essa viene variata spostando di due tacche in senso orario la ghiera del macinadosatore così da avere una macinatura più fine(+2 tacche), oppure di due tacche nel senso opposto così da ottenere una macinatura più grossolana (+2 tacche)
  - *Dose*: la dose ottimale di macinato da inserire per una dose doppia è 14 grammi: essa viene variata aggiungendo o sottraendo due grammi di macinato.
  - *Pressione*: dal valore ottimale di pressione (8.9 bar) la pressione è diminuita di 3 bar e aumentata di 2 bar.

Vengono testate tutte le possibile combinazioni tra i valori appena elencate delle tre variabili esterne, per un totale di 27 test, ognuno composto di 20 erogazioni. Inoltre essi sono ripetuti su entrambi i gruppi così da confrontare i comportamenti ottenuti in presenza o mancanza di lavaggi. I dati di questa sessione sperimentale sono raccolti tramite rilevazioni in tazza e tramite scheda PC della macchina. Per maggiori dettagli sulla struttura del dataset fare riferimento alla sezione 4.2.

- Seconda sessione sperimentale: svolta a cavallo dei mesi di giugno e luglio, ha come obiettivo quello di modellare una variazione meno estrema dei parametri esterni. La struttura dell'esperimento è la stessa del precedente, con la differenza dei parametri testati:
  - *Macinatura*: Ottimale, -1 tacca (più fine), +1 tacca(più grossolana);
  - *Dose*: Ottimale(14 g), -1 grammo, +1 grammo;
  - *Pressione*: Ottimale(8.9 bar), -3 bar, +2 bar;

La seconda sessione sperimentale ci fornisce però un dato in più: infatti, oltre alle rilevazioni esterne fatte in tazza e a quelle tramite la scheda PC della macchina, le erogazioni sono misurate anche tramite il device di telemetria che si trova anche sulle macchine in produzione. Come analizzato in seguito, nella sezione 6.4.2, la maggior parte delle erogazioni rilevate in questa sessione sperimentale, risulta avere un erogato in tazza e dei tempi di erogazione troppo elevati, forse a causa di un errore nella macinatura. I dati raccolti sono però utili per capire la relazione tra i dati rilevati in tazza e quelli in telemetria(sezione 6.2), che permette il passaggio all'analisi dei dati provenienti dal mercato.

## 6.1 Analisi esplorativa della prima sessione sperimentale

L'analisi iniziale è stata svolta grazie all'analisi dei dati della prima sessione sperimentale raccolti nel mese di marzo e mirati a evidenziare come una variazione di alcuni parametri

esterni tra cui macinatura, dose e pressione, influenzi le caratteristiche del caffè e se una differenza nella manutenzione del gruppo macchina potesse incidere sulle prestazioni.

### 6.1.1 Selezione delle variabili principali

L'obiettivo iniziale è definire una struttura per il rilevamento dei dati in laboratorio, tale da includere le variabili più utili e significative per l'analisi, al fine di velocizzare il processo di raccoglimento dei dati. Si procede con la selezione degli attributi che forniscono maggiori informazioni.

Come descritto nella sezione 4.2.1, la struttura del dataset si può dividere in diverse macroaree:

- *Informazioni generali*: indica quando l'erogazione è stata effettuata, e a quale test appartiene, più precisamente a quali variazioni dei parametri esterni corrisponde;
- *Dati caratteristici dell'erogazione*: descrive l'erogazione grazie ai dati provenienti dalla macchina (modello, gruppo usato, tipologia, dose programmata), al tipo di miscela usata e alla sua pressatura. Generalmente i valori degli attributi di questa sezione restano fissi per l'intera durata dell'esperimento;
- *Parametri primari dell'erogazione rilevati esternamente*: rilevazioni effettuate manualmente, che offrono informazioni sul tempo di erogazione, e sui pesi di macinato e quantità in tazza dell'erogazione. Rappresentano le caratteristiche in tazza del caffè;
- *Dati provenienti dalla scheda PC*: dati rilevati dalla scheda elettronica della macchina. Sono la base dei dati rilevati dal device di telemetria;
- *Confronto con le soglie*: sono le informazioni più importanti dell'erogazione (tempo di erogazione, flusso e erogato in tazza), a cui vengono date delle etichette basandoci sulle soglie di qualità fornite dagli esperti di dominio;
- *Diagnostica della macchina*: forniscono informazioni sulla presenza dei lavaggi del gruppo di erogazione, oltre a fornire eventuali messaggi di errore della macchina.
- *Indici calcolati*: sono indici ricavati partendo dalle rilevazioni in tazza e dalla scheda PC, così da fornire nuove variabili in analisi.

Durante lo studio delle variabili è emerso come gli *Indici Calcolati* non forniscono molte informazioni aggiuntive, ed essendo calcolati a partire dai dati rilevati possono essere calcolate successivamente in caso di necessità. Quindi, si è optato per eliminarli dalla struttura dei dataset futuri e di non rilevare alcuni campi della diagnostica della macchina, in quanto non significativi o con valore nullo costante.

Per quanto concerne i *Dati caratteristici dell'erogazione*, assumono sempre lo stesso valore all'interno della prima sessione sperimentale, in quanto tutte le erogazioni sono doppie e svolte sulla stessa macchina, con la sola differenza del gruppo di erogazione. Esse risultano molto importanti in un contesto in cui dati di rilevazioni di diverso tipo, svolte su macchine diverse, siano analizzate assieme. Per l'analisi limitata alla prima sessione sperimentale risultano però di poca importanza poiché vengono mantenute le stesse condizioni per ogni erogazione.

È stato deciso allora di eliminare queste colonne solo per quanto riguarda l'analisi della prima sessione sperimentale, ottenendo il database ridotto le cui caratteristiche sono specificate nella sezione 4.2.2.

### 6.1.2 Analisi di correlazione

Partendo dal nuovo dataset, si è cercato di capire quali parametri delle erogazioni sono correlati. Dopo aver eliminato quegli attributi il cui valore è calcolato direttamente da altre variabili, è stata calcolata la matrice di correlazione tra le diverse variabili.

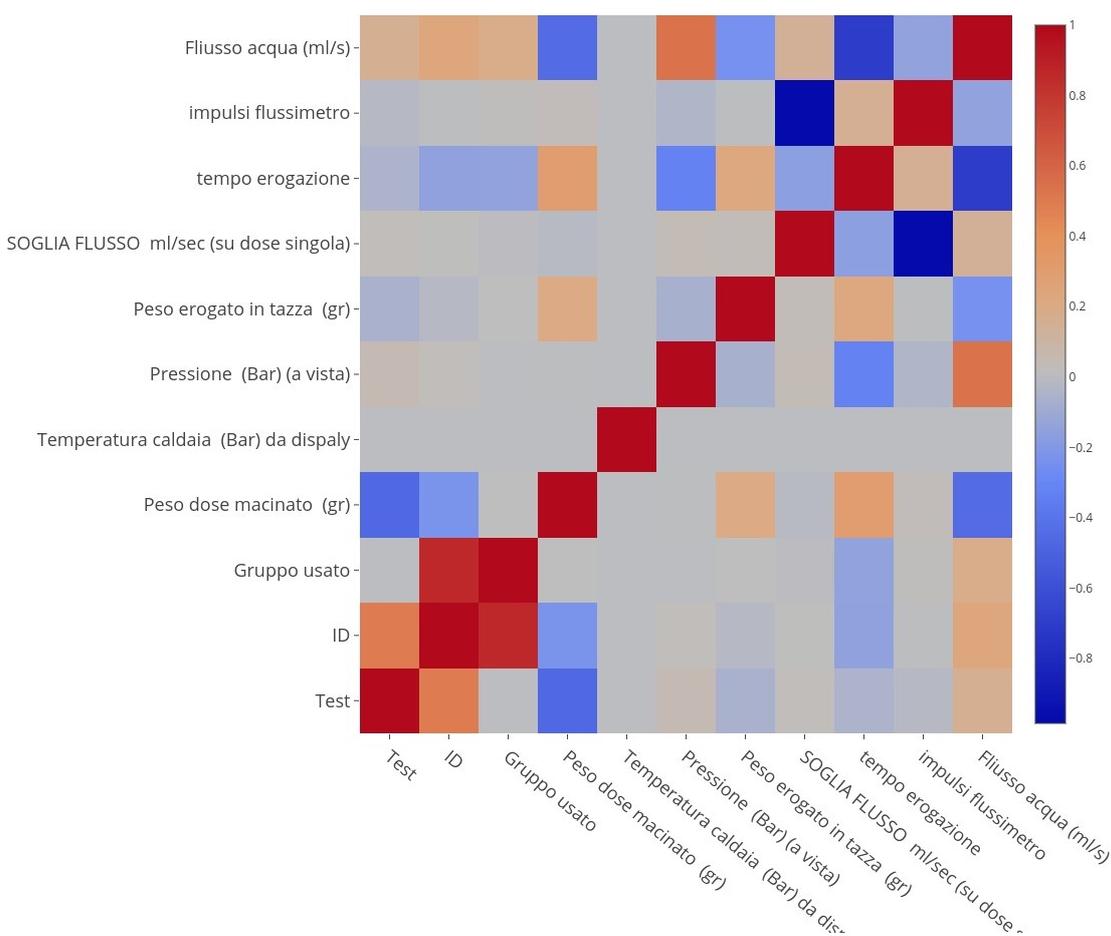


Figura 6.1: Matrice di correlazione

Il grafico riporta la correlazione tra le diverse variabili: le gradazioni di rosso indicano una correlazione positiva, mentre quelle di colore azzurro presentano una correlazione negativa. Variabili poco o non correlate sono invece caratterizzate da un colore grigio. Possiamo quindi notare come alcune variabili siano correlate tra loro:

- Tempo di erogazione e flusso della scheda PC: inversamente correlati, poichè più aumenta il tempo, più il flusso è lento;

- Pressione con tempo di erogazione e flusso della scheda PC: aumentando la pressione il tempo di erogazione diminuisce, e di conseguenza il flusso aumenta;
- Peso del macinato con tempo di erogazione e flusso della scheda PC: una dose minore di macinato causa un tempo di erogazione minore, e quindi un flusso maggiore;
- Tempo di erogazione e erogato in tazza: correlati positivamente, ovvero a un tempo maggiore corrisponde un erogato maggiore;
- Gruppo usato con tempo di erogazione e flusso della scheda PC: le due caratteristiche principali sembrano in qualche modo correlati col gruppo usato. Ciò è di particolare interesse in quanto i due gruppi presentano condizioni diverse, in quanto in uno dei due gruppi non sono mai stati effettuati i lavaggi.

Possiamo quindi affermare che, come già anticipato dagli esperti di dominio, le nostre variabili di interesse sono:

- Tempo di erogazione;
- Flusso;
- Erogato in tazza.

In particolare il tempo e il flusso variano a seconda di alcuni parametri esterni, come, ad esempio, la pressione e la dose di macinato, e sembra intercorrere una differenza di comportamento nei due gruppi: di questi due aspetti parleremo più approfonditamente nella sezione 6.4, dedicata allo studio delle regole di associazione.

Svolgiamo però una prima analisi per vedere l'andamento delle tre caratteristiche principali all'interno della sessione sperimentale, confrontando i due gruppi aventi processo di manutenzione diverso a causa della mancanza dei lavaggi in uno dei due gruppi.

### 6.1.3 Andamento delle variabili e confronto tra gruppi

Sfruttiamo i boxplot della libreria Matplotlib per visualizzare l'andamento delle tre variabili nei due gruppi.

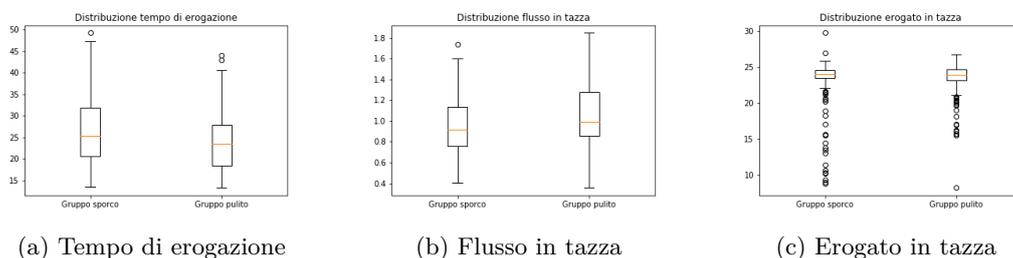


Figura 6.2: Distribuzione delle tre variabili principali nella prima sessione sperimentale.

A causa della struttura della sessione sperimentale i tempi di erogazione e i flussi sono molto variabili, poichè è prevista la variazione di macinatura, dose e pressione, che influenzano in modo significativo sui parametri generali dell'erogazione. In aggiunta, l'erogato,

esclusi alcuni outliers, ha una distribuzione più stretta, con una quantità in tazza che rimane praticamente costante. Questo è attribuibile alla taratura della macchina, che pone come obiettivo principale l'erogazione di una quantità costante di caffè.

Confrontando le distribuzioni dei due gruppi, risulta evidente come il gruppo sporco abbia dei tempi più elevati rispetto al gruppo pulito, e questo aspetto è dovuto, probabilmente, dovuti a una sporcizia accumulatasi nei condotti della macchina a causa della mancanza dei lavaggi. Analogamente, il flusso del gruppo sporco è più lento rispetto al gruppo pulito.

## 6.2 Relazione tra rilevazioni da fonti differenti

Durante le prime due sessioni sperimentali, i dati delle erogazioni sono raccolti da diverse fonti.

Nella prima sessione sperimentale si hanno a disposizione i dati rilevati in tazza, attraverso l'utilizzo di una bilancia per il calcolo della quantità di erogato in tazza, e attraverso la scheda Pc, ovvero la scheda elettronica della macchina che fornisce il numero di colpi di ventolino e il tempo di erogazione.

Nella seconda sessione sperimentale, invece, oltre alle due fonti appena citate, sono disponibili anche i dati rilevati attraverso il device di telemetria. Esso è collegato alla scheda PC e riporta i dati come tempo di erogazione e velocità del flusso.

Sulle macchine sul mercato, solo quest'ultima tipologia di fonte è disponibile: ci si concentra allora nel trovare la correlazione tra le rilevazioni provenienti dalle diverse fonti, così da poter trasformare le informazioni ottenute con la telemetria in informazioni comparabili con le conoscenze sul caffè in tazza.

### 6.2.1 Correlazione tra scheda PC e rilevazione in tazza

La prima sessione sperimentale, effettuata su una macchina Cimbali M100, presenta le rilevazioni esterne e dalla scheda PC. Come già anticipato, le misurazioni esterne sono effettuate con un bilancino per calcolare il peso della quantità erogata in tazza. Quest'ultima è poi moltiplicata per 1.02, peso specifico del caffè, così da ottenere il dato in millilitri.

La scheda PC invece restituisce solamente la durata dell'erogazione e i colpi di ventolino erogati. Ogni colpo di ventoli corrisponde, per la macchina in questione, a 0.5 ml di acqua.

Essendo la sessione sperimentale basata su erogazioni doppie, la quantità rilevata viene poi divisa per due, così da avere il dato riferito al singolo caffè.

L'obiettivo è quindi di trovare una correlazione tra due caratteristiche di qualità: flusso e quantità.

Il flusso è una grandezza derivata. Essa viene calcolata come segue nei due casi.

$$F_{tazza} = \frac{Q_{\{tazza\}}}{t} \quad (6.1)$$

dove  $Q_{tazza}$  rappresenta la quantità di caffè misurata in tazza riferita al singolo caffè e trasformata in millilitri, e  $t$  è il tempo di erogazione.

$$F_{PC} = \frac{Q_{PC}}{t} = \frac{Cdv * 0,5 * 1/2}{t} \quad (6.2)$$

dove  $Cdv$  è il numero di colpi di ventolino contati dal flussimetro, 0,5 sono i ml di acqua erogata per ogni colpo di ventolino,  $1/2$  è il fattore per passare dall'erogazione doppia a quella singola e  $t$  è il tempo di erogazione.

Osservando l'andamento dei due flussi (figura 6.3), si nota subito come essi abbiano un andamento simile.

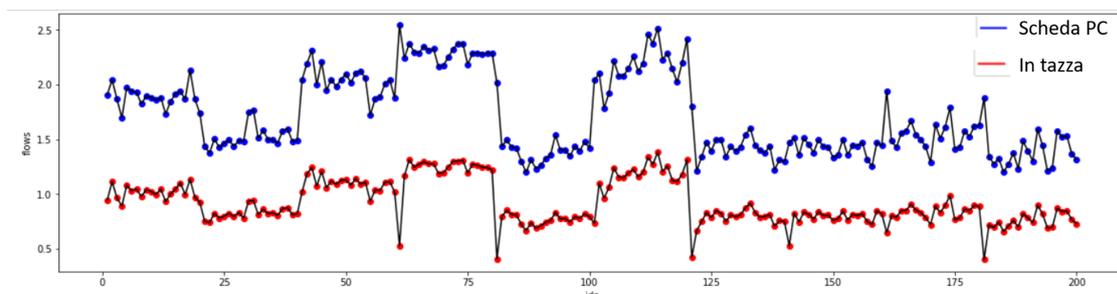


Figura 6.3: Andamento dei flussi rilevati in tazza e da scheda PC

Attraverso una retta di regressione, viene calcolata la funzione di trasferimento dalla flusso rilevato dalla scheda PC a quello in tazza.

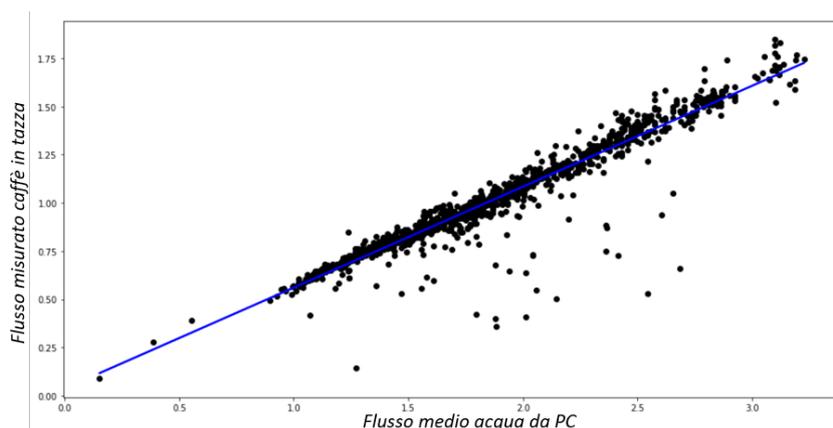


Figura 6.4: Retta di regressione da flusso scheda PC a flusso in tazza

Osservando la tabella 6.1 e la figura 6.4 è possibile notare come le due variabili siano correlate tra loro con un indice di Pearson del 98%, e che con la funzione descritta si ottiene una trasformazione del dato della scheda PC in dato tazza con una buona approssimazione, come visibile dalla retta in figura.

Quindi, data una certa taratura della macchina, e quindi una certa impostazione dei colpi di ventolino, è possibile calcolare il flusso in tazza corrispondente a quello misurato dalla scheda PC. Inoltre la funzione calcolata è valida per la macchina utilizzata, ovvero la Cimbali M100, con una programmazione di 177 colpi di ventolino.

Si è usato un approccio analogo per calcolare la relazione tra quantità in tazza e quantità di acqua erogata secondo la scheda PC. Come indicato già nella formula (6.2), la quantità

Corr. Pearson	R-squared	Funzione
0.98	0.90	$F_{tazza} = 0.52F_{PC} + 0.04$

Tabella 6.1: Funzione di trasferimento da flusso scheda PC a flusso in tazza

di acqua erogata secondo la scheda PC è calcolata come:

$$Q_{PC} = Cdv * 0,5 * \frac{1}{2} \quad (6.3)$$

dove  $Cdv$  è il numero di colpi di ventolino contati dal flussimetro, 0,5 sono i ml di acqua erogata per ogni colpo di ventolino,  $1/2$  è il fattore per passare dall'erogazione doppia a quella singola.

Dalla comparazione degli andamenti delle quantità di scheda PC e in tazza, si nota come essa rimane costante, anche se in tazza si verificano alcuni picchi inferiori, con una quantità non comparabile alle altre. In particolare sono 25 gli erogati anomali su 1080, e di questi 25,

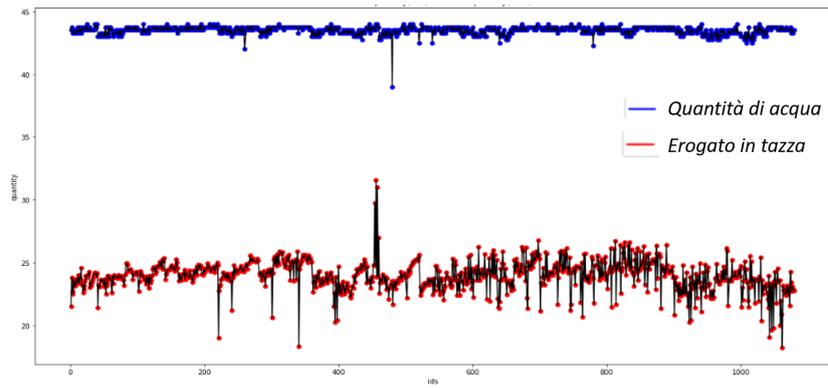


Figura 6.5: Andamento delle quantità della scheda PC e in tazza

ben 17 sono stati rilevati alla prima erogazione di ogni test, in cui veniva cambiato uno dei parametri esterni (dose, macinatura e pressione). Si può assumere quindi che questi valori rappresentino degli outliers, forse a causa di alcuni errori durante la rilevazione del peso con la bilancia, o durante la trascrizione manuale del dato. Si decide quindi di eliminare, per ogni test, l'erogazione con erogato minore e maggiore, come descritto nella sezione 5.2.1.

Essendo l'acqua erogata calcolata come numero di colpi di ventolino per una quantità fissa, essa risulta molto stabile. L'erogato in tazza risulta invece più variabile, presentando infatti una deviazione standard più elevata, mostrata in tabella 6.2.

In realtà è la misura della quantità della scheda PC ad essere troppo costante: infatti essa calcolata considerando la quantità emessa da ogni colpo di ventolino costante. In realtà, essendo un elemento meccanico, la quantità emessa può variare leggermente ogni volta, ma non abbiamo possibilità di rilevarlo. Questa potrebbe essere quindi la causa della maggiore variabilità dell'erogato in tazza, a cui vanno a sommarsi altri due effetti che influiscono nella misurazione diretta dell'erogato. Infatti, la lettura del peso della

Quantità	Media (ml)	Deviazione standard (ml)
In tazza	24.0	1.2
Scheda PC	43.5	0.3

Tabella 6.2: Statistiche delle quantità in tazza e della scheda PC

quantità avviene a mano, e possono esserci alcuni errori di trascrizione. Inoltre, un caffè appena erogato presenta una schiuma superficiale che evapora dopo alcuni istanti. In base al momento della rilevazione potrebbe essere conteggiata o meno.

Calcolando la regressione lineare per la quantità, si ottiene però che la funzione di trasferimento non ha una buona approssimazione (tabella 6.3), in quanto le due quantità non sono direttamente correlate, poichè a una stessa quantità della scheda PC vengono associati valori, seppur di poco, diversi in tazza (figura 6.6).

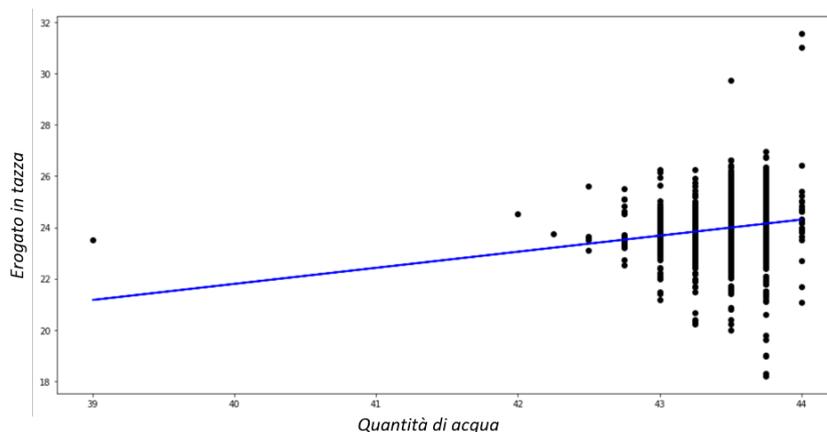


Figura 6.6: Retta di regressione da quantità della scheda PC a in tazza

Essendo però la deviazione dell'erogato in tazza molto limitato (1,2 ml), si può approssimare che una quantità di acqua erogata di 43.5 ml corrisponda a un caffè di 24 ml. Ciò è valido su una macchina Cimbali M100, con una impostazione di 177 colpi di ventolino.

Per confermare e generalizzare gli studi effettuati, la stessa analisi è stata effettuata nella seconda sessione sperimentale. È stata utilizzata la stessa macchina (Cimbali M100), che presenta però una impostazione di 207 colpi di ventolino iniziali, mantenendo però lo stesso tipo di miscela per effettuare gli esperimenti.

Attraverso un processo analogo, è stata calcolata nuovamente la funzione di trasferimento da scheda PC a tazza. Essa è nuovamente calcolabile e precisa, ma diversa dalla precedente (tabella 6.4).

Essa è però differente dalla precedente: essendo stata utilizzata stessa macchina e stessa miscela per effettuare le erogazioni, la differenza della funzione è riconducibile alla differenza nei colpi di ventolino, e quindi alla diversa taratura.

Corr. Pearson	R-squared	Funzione
0.17	0.03	$y = 0.63x + 3.28$

Tabella 6.3: Funzione di trasferimento da quantità scheda PC a in tazza

Corr. Pearson	R-squared	Funzione
0.96	0.97	$y = 0.60x + 0.03$

Tabella 6.4: Funzione di trasferimento da flusso scheda PC a flusso in tazza

Un discorso analogo si può fare per la quantità di erogato in tazza, che dopo la pulizia (sezione 5.2.1) non permette il calcolo accurato della dose in tazza a partire dai colpi di ventolino. È possibile vedere come però le due quantità siano pressochè costanti e legate tra loro.

In conclusione, grazie al confronto tra le funzioni calcolate nelle due sessioni sperimentali, è possibile affermare che non è possibile calcolare accuratamente la quantità in tazza partendo dal dato della scheda PC, anche se si rileva una perdita di acqua abbastanza costante (55-60% di caffè erogato rispetto alla quantità di acqua erogata). Si può invece calcolare accuratamente il flusso in tazza partendo dai colpi di ventolino, anche se essa varia a seconda della taratura, e quindi dell'impostazione dei colpi di ventolino.

Quantity	Media (ml)	Deviation Standard (ml)
In tazza	32.0	0.9
Scheda PC	51.7	0.2

Tabella 6.5: Statistiche della quantità rilevata in tazza e dalla scheda PC

## 6.2.2 Correlazione tra dato di telemetria e rilevazione in tazza

Nella seconda sessione sperimentale sono state aggiunte le variabili ricavate dal sensore di telemetria: il flusso medio e il tempo di telemetria. Sono sorti i primi problemi di mancanza e di allineamento dei dati di telemetria, mostrati nel capitolo 5.2: per questo motivo, il numero di erogazioni che è possibile analizzare, è di 708 erogazioni sulle 1160 effettuate (il 39% delle erogazioni non è presente). I dati della scheda PC rappresentano l'acquisizione del sensore, mentre quelli della telemetria sono un'elaborazione di questi ultimi. L'analisi di correlazione svolta sui dati della scheda PC aveva lo scopo di validare la correttezza dei dati acquisiti dal sensore e di verificare la possibilità di trovare le funzioni di trasferimento. L'obiettivo finale è però quello di trovare la correlazione tra i misurati in tazza e quelli di telemetria: per le macchine sul mercato infatti questi ultimi saranno gli unici disponibili.

Come per l'analisi della correlazione tra i dati della scheda PC e quelli misurati in tazza, si è partiti dallo studio del flusso: osservando il trend dei due flussi è evidente come anche in questo caso è possibile che i due siano correlati (grafico 6.7). Infatti, procedendo con l'analisi degli indici statistici e con la regressione lineare, la correlazione tra flusso medio di telemetria e flusso misurato in tazza risulta molto forte (tabella 6.6 e figura 6.8).

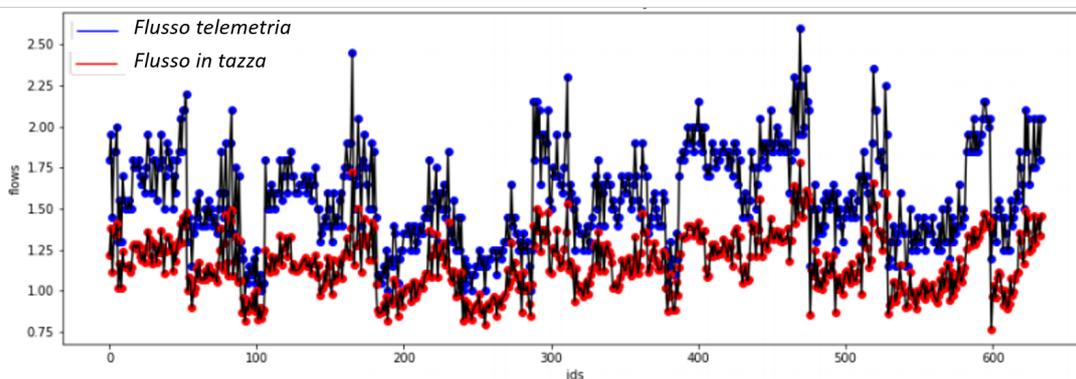


Figura 6.7: Andamento dei flussi in tazza e rilevati in telemetria

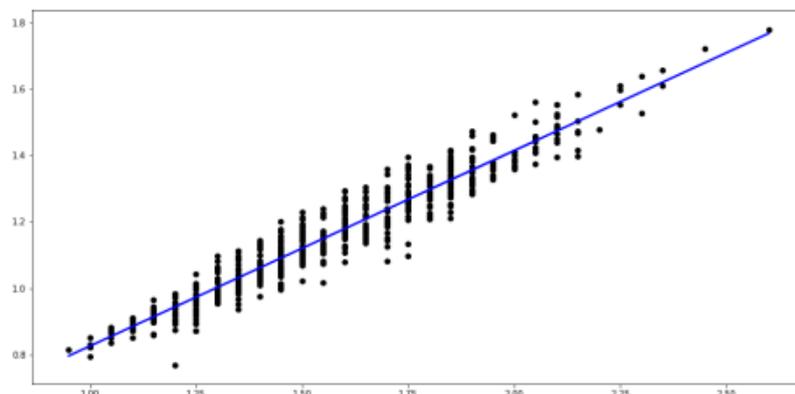


Figura 6.8: Retta di regressione da flusso di telemetria a flusso in tazza

Corr. Pearson	R-squared	Funzione
0.96	0.93	$y = 0.59x + 0.24$

Tabella 6.6: Funzione di trasferimento da flusso di telemetria a flusso in tazza

Considerando, invece, la quantità di acqua erogata, questa non è disponibile direttamente come dato di telemetria: può essere però derivata da flusso medio e tempo di erogazione attraverso la formula:

$$Q_{Tel} = F_{medio} * t \quad (6.4)$$

dove  $F_{medio}$  è il flusso medio di telemetria e  $t$  è il tempo di erogazione di telemetria

Calcolando quindi la indice di correlazione e la regressione lineare tra quantità pesata in tazza e quantità ricavata dalla telemetria, si ottengono i risultati in tabella 6.7. Anche in questo caso, non è possibile trovare una funzione di regressione precisa per il passaggio dalla quantità d'acqua di telemetria alla quantità di caffè erogato in tazza. Mentre in precedenza la quantità d'acqua misurata dalla scheda PC era più costante rispetto a quella misurata in tazza, ora è la quantità d'acqua di telemetria a presentare una varianza maggiore. Lo dimostrano, infatti, la media e la deviazione standard mostrate in tabella 6.8: il caso è molto diverso dal precedente, dal momento che la deviazione standard della quantità d'acqua di telemetria non è più trascurabile. Si possono vedere anche dal grafico 6.9 le grandi oscillazioni del flusso di telemetria e la poca correlazione tra gli andamenti delle quantità.

Corr. Pearson	R-squared	Funzione
0.07	0.04	$y = 0.07x + 28.97$

Tabella 6.7: Funzione di trasferimento da quantità della telemetria a quantità in tazza.

Quantità	Media (ml)	Deviazione standard (ml)
In tazza	32.0	0.8
Da telemetria	43.9	2.4

Tabella 6.8: Statistiche sulle quantità in tazza e rilevate dalla telemetria

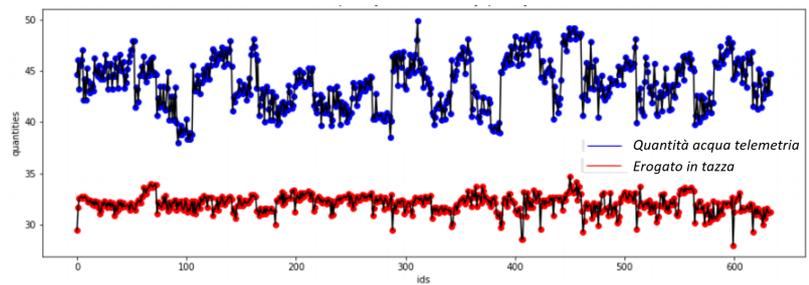


Figura 6.9: Andamento della quantità in tazza e in telemetria

La quantità di caffè erogato in tazza non può essere invece predetta con precisione dal valore della quantità di acqua di telemetria (vedi formula (6.4)). Inoltre il valore non coincide ed è molto meno costante rispetto alla quantità di acqua calcolata dai dati della scheda PC (vedi grafico 6.10): ciò suggerisce che, essendo il tempo di telemetria e quello calcolato dalla scheda PC coincidenti, il flusso medio di telemetria non venga calcolato direttamente a partire dai colpi di ventolino, ma che subisca un'elaborazione. Questo fatto è stato confermato dagli esperti di dominio di Lavazza, ma purtroppo solo il costruttore della macchina per ora è a conoscenza dell'algoritmo di elaborazione del dato.

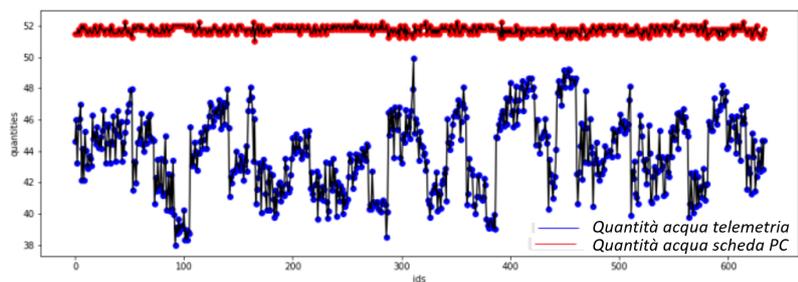


Figura 6.10: Andamento quantità d'acqua scheda PC e telemetria

Per approfondire e verificare le conclusioni precedenti, si è proposta una sessione sperimentale con dei test che mirano a calcolare le funzioni di trasferimento dalle variabili in telemetria a quelle in tazza per diverse tarature e diverse macchine. In particolare viene modificato il numero di colpi di ventolino mantenendo la stessa impostazione ottimale di macinatura e dose, così da poter ricavare quale sia la relazione tra acqua erogata dalla macchina e quella in tazza, e calcolare le funzioni di trasferimento del flusso con diverse tarature e diverse macchine.

### 6.3 Calcolo delle nuove soglie

Uno dei temi principali sviluppati nello studio riguarda la qualità in tazza del caffè: con qualità non si intende un parametro soggettivo come il gusto o l'odore, bensì alcuni parametri oggettivi. In particolare, per ottenere un buon caffè in tazza, il processo di erogazione deve rispettare alcune regole generali per il tempo di erogazione, la velocità del flusso e la quantità di erogato.

Come descritto nel capitolo 2, comunemente un caffè è definito come ottimo se presenta un tempo di erogazione nell'intervallo 20-30 secondi e se ha una quantità in tazza tra 20-30 millilitri.

Gli esperti della Lavazza hanno fornito alcune soglie compatibili con quelle di carattere generale, leggermente più restrittive per quanto riguarda il tempo di erogazione mentre un po' più lasche per l'erogato in tazza. Inoltre si presta particolare attenzione anche al flusso in tazza dell'erogazione, calcolato come rapporto tra quantità erogata e tempo: il flusso è infatti una variabile molto importante poiché l'acqua deve attraversare con la giusta velocità il panetto così da poter impregnarsi appieno dell'aroma del caffè senza però degradarne la qualità. Un caffè ottimo deve avere un flusso di 1 ml/s, come conseguenza

della corretta taratura, che prevede di avere 30 cc di erogato in 30 secondi. La tabella 6.9 contiene le soglie di qualità forniteci dagli esperti Lavazza.

	soglie
Tempo di erogazione (s)	20 - 27
Erogato in tazza (ml)	18 - 30
Flusso (ml/s)	0,49 - 1,33

Tabella 6.9: Soglie in tazza degli esperti di dominio

Si prova ora a mettere in discussione le soglie forniteci dagli esperti, analizzando le distribuzioni dei dati forniti.

### 6.3.1 Proposta di nuove soglie a partire dai dati di laboratorio

Durante l'analisi della prima sessione sperimentale, mirata a verificare l'effetto delle variazioni dei parametri esterni sulla qualità del caffè, è stato effettuato un test con tutti i parametri esterni (macinatura, dose e pressione) ottimali: abbiamo allora cercato di approfondire lo studio delle caratteristiche dei caffè ottimali. In particolare, sono state calcolate media e deviazione standard delle tre caratteristiche principali del caffè, e sono state definite delle soglie di qualità come  $media \pm dev$ , come indicate in tabella 6.10.

	soglie
Tempo di erogazione (s)	21 - 25
Erogato in tazza (ml)	21 - 25
Flusso (ml/s)	0,9 - 1,1

Tabella 6.10: Statistiche e soglie in tazza ricavate da erogazioni ottimali

Analizzando le soglie trovate a partire dai dati e confrontandole con quelle fornite dagli esperti possiamo trarre alcune osservazioni:

- Le soglie del tempo di erogazione sono comparabili con quelle degli esperti.
- Le soglie sull'erogato sono decisamente più strette rispetto a quelle fornite dalla Lavazza: ciò è dovuto al fatto che una volta impostato un certo numero di colpi di ventolino durante la taratura l'erogato in tazza rimane pressochè costante;
- Le soglie di flusso sono davvero molto strette e centrate su 1 ml/s, rispettando quindi la regola dei 30cc di erogato in 30 secondi.

Prendendo in considerazione le soglie di flusso degli esperti notiamo subito un particolare: esse non sono centrate su 1ml/s come potevamo aspettarci, ma accettano come buoni anche dei flussi molto lenti. Le soglie calcolate dai dati sembrano invece troppo restrittive, sebbene centrate su 1 ml/s. Proponiamo allora delle soglie che non siano troppo restrittive

ma che siano centrate su 1 ml/s. Essendo il flusso una misura derivata da tempo di erogazione e quantità erogata, si analizzano le soglie di queste due grandezze e si effettuano alcune osservazioni:

- le soglie del tempo comparabili nei due casi, si preferisce mantenere le soglie proposte dagli esterni in quanto meno dipendenti dal caso specifico;
- L'erogato in tazza, invece, si presenta abbastanza costante, poichè la taratura setta un certo numero di colpi di ventolino, ovvero di acqua erogata. Essa non varia modificando le condizioni esterne, ma varia effettuando una taratura diversa. Il valore delle soglie allora dovrebbe variare per ogni taratura. Essendo impossibile modellare tutti i possibili casi, si mantengono le soglie molto larghe fornite dagli esperti, che permettono al cliente di effettuare tarature diverse e quindi dosi in tazza diverse.

Possiamo però usare le soglie dell'erogato calcolate a partire dai dati per cercare delle soglie centrate del flusso in tazza: date quindi le soglie di tempo  $T_{inf}$  e  $T_{sup}$ , e di quantità  $Q_{inf}$  e  $Q_{sup}$ , possiamo calcolare le soglie del flusso  $F_{inf}$  e  $F_{sup}$  come:

$$F_{inf} = \frac{Q_{inf}}{T_{sup}} \quad F_{sup} = \frac{Q_{sup}}{T_{inf}} \quad (6.5)$$

Le soglie quindi proposte e utilizzate nelle analisi sono riportate nella tabella 6.11. La scelta di queste soglie permette così di discernere i vari comportamenti senza però restringere troppo la fascia ottimale, riuscendo così a generalizzare rispetto al caso specifico preso in analisi.

	soglie
Tempo di erogazione (s)	20 - 27
Erogato in tazza (ml)	18 - 30
Flusso (ml/s)	0,77 - 1,23

Tabella 6.11: Soglie in tazza proposte a partire dai dati.

### 6.3.2 Proposta di soglie di flusso a tre livelli

Per cercare gli effetti delle variazioni che non portano il flusso fuori soglia, ma che hanno un impatto su di esso, sono state concordate delle soglie a tre livelli, che permettono di dividere le erogazioni ottimali in erogazioni con un flusso leggermente veloce o leggermente lento. In particolare, si vuole porre attenzione alle erogazioni aventi flusso più veloce, perchè collegate a una dose minore o a una macinatura più larga.

La tabella 6.12 contiene le due proposte di soglie a tre livelli per il flusso.

	minore	ok_lento	ok_veloce	maggiore
Prima proposta	<0,73	[0,73-1]	[1-1,23]	>1,23
Seconda proposta	<0,73	[0,73-1,1]	[1-1,23]	>1,23

Tabella 6.12: Soglie di flusso a tre livelli.

Come sarà approfondito nella prossima sezione, quattro livelli di etichette dividono eccessivamente il dataset, non riuscendo quindi a dare un valore statistico alle conclusioni caratterizzanti ogni gruppo di erogazioni. In particolare, per lo studio attraverso le regole di associazione effettuato in seguito, si ha un supporto troppo ridotto per le regole estratte, ovvero si hanno regole poco frequenti. Inoltre si corre il rischio che erogazioni effettivamente ottimali, poichè effettuate senza variazioni delle caratteristiche esterne, siano divise nelle due categorie, dando quindi una interpretazione ambigua del comportamento corretto.

## 6.4 Regole di associazione

In questa sezione verrà approfondito lo studio della qualità in tazza attraverso le regole di associazione: esse permettono di identificare pattern e correlazioni frequenti in un insieme di variabili. In particolare si tratta di un metodo esaustivo, che restituisce tutte le possibili correlazioni: tra queste ne vengono selezionate alcune considerate come rilevanti, poichè presentano indici di qualità più importanti. Una descrizione più approfondita del funzionamento delle regole di associazione è fornita nella sezione 5.5.

I dati rilevati in laboratorio nelle prime due sessioni sperimentali hanno due principali obiettivi:

- Osservare l'influenza della variazione delle variabili esterne (dose, pressione e macinatura) sui tre parametri caratteristici dell'erogazione (tempo di erogazione, flusso e quantità in tazza);
- Rilevare eventuali differenze nel comportamento dei gruppi, dovuto alla mancanza di lavaggi in uno di essi.

In particolare, come affrontato nella sezione 6.3, vengono fornite delle soglie di qualità per le tre misurazioni che permettono l'assegnazione di un'etichetta ad ognuna delle tre grandezze di ogni erogazione. Inoltre ogni erogazione corrisponde a certi valori assunti dalle variabili esterne: infatti, ogni rilevazione fa parte di uno dei 27 test che presentano le varie combinazioni dei valori assunti da dose, pressione e macinatura. In questo modo, ad ogni erogazione può essere assegnato un insieme di etichette che descrivono le sue caratteristiche. In particolare è caratterizzata da sei etichette:

- Macinatura: indica le caratteristiche della macinatura utilizzata per l'erogazione. Può assumere valore "Ottimale", oppure il numero di tacche (che può essere anche negativo) di cui è stata spostata la ghiera del macinadosatore: un valore positivo indica una macinatura più grossolana, uno negativo, invece, una più fine;

- Dose: indica i grammi di macinato inserito nell'erogatore. Con ottimale sono indicati 14g di polvere, mentre con un numero con segno si indicano i grammi aggiunti o rimossi a partire dal valore standard;
- Pressione: indica il valore di pressione della pompa. Con "Ottimale" è indicata la pressione di 9 bar, mentre con un numero con segno si indicano i bar aggiunti o sottratti al valore di riferimento;
- Tempo di erogazione: indica la qualità dell'erogazione rispetto alle soglie prese in considerazione. Con "opt" si indica un tempo in soglia, con "lento" un tempo inferiore alla soglia inferiore e, infine, con "veloce" un tempo più elevato della soglia superiore.
- Flusso: indica la qualità dell'erogazione rispetto alle soglie di flusso. I valori delle etichette sono analoghi a quelli espressi al punto precedente;
- Erogato: indica la qualità dell'erogazione rispetto alle soglie della quantità di erogato in tazza. I valori possibili sono analoghi al punto precedente.

È proprio grazie alla possibilità di assegnare delle etichette alle erogazioni che è possibile l'utilizzo delle regole di associazioni. Esse infatti si basano sul concetto di *co-occorrenza*: le regole che estrarremo indicano che il corpo della regola e la testa della regola si presentano assieme con una certa frequenza, il che non implica causalità tra le due parti.

Il significato di causalità è una interpretazione che si può dare in questo specifico contesto in quanto le variazioni dei parametri esterni, rappresentate da etichette, portano delle variazioni nelle caratteristiche del caffè a cui sono assegnate un altro tipo di etichetta.

### 6.4.1 Tipi di regole

Nel corso di questo studio, le regole estratte sono classificate in tre tipi, che si basano sulla loro composizione:

- Regole di impatto: descrivono l'influenza delle variabili esterne sulla qualità del caffè. Presentano nella testa della regola un'etichetta di qualità e nel corpo i valori assunti dalle variabili esterne a cui è collegata la testa: tra esse possiamo trovare delle regole di compensazione quando la testa presenta un'etichetta "opt";
- Regole di origine: analizzano quale variazione esterna ha portato un certo indice di qualità. Presentano un'etichetta di qualità nel corpo della regola e una variabile esterna come testa;
- Regole di compensazione: studiano gli effetti compensazione, ovvero attraverso quali configurazioni delle variazioni dei parametri esterni si possa ottenere comunque una buona qualità del caffè. Presentano nel corpo della regola una etichetta di qualità ottimale e almeno un parametro esterno non ottimale, e nella testa della regola un altro parametro esterno non ottimale. La regola va quindi interpretata nel seguente modo: se ho un caffè buono nonostante un parametro esterno non ottimale, allora quest'ultimo è compensato dalla variazione di un altro fattore esterno.

Questo tipo di regole permettono di osservare l'influenza delle variabili esterne sulla qualità del caffè.

Esse possono anche essere usate per effettuare un confronto tra caffè erogati su gruppi con condizioni diverse a causa dell'assenza dei lavaggi: in questo modo è possibile vedere come i due gruppi reagiscono alle stesse variazioni.

### 6.4.2 Dati a disposizione

Come descritto nell'introduzione di questo capitolo, sono state effettuate due sessioni sperimentali volte a capire l'influenza di dose, macinatura e pressione sulla qualità del caffè. La prima sessione mira in particolare a modellare una variazione estrema dei fattori esterni, mentre la seconda una variazione più lieve di essi. Mentre il primo esperimento contiene dati in linea con quelli attesi, ovvero con dei caffè ottimali che rispettano le norme di qualità della Lavazza (46 grammi di erogato in 25 secondi circa, per un caffè doppio), la seconda sessione in laboratorio presenta dei dati che vanno oltre queste indicazioni: infatti, solo 11 erogazioni delle 1080 erogazioni complessive presenta una quantità in tazza all'interno delle soglie di qualità forniteci dagli esperti (18-30 ml). A causa di questa alta quantità in tazza, anche il tempo di erogazione risulta molto elevato e si presenta maggiore di 27 secondi (soglia fornita dalla Lavazza) in più della metà dei caffè considerati ottimali. Di conseguenza, molti caffè che presentano variazioni dei parametri esterni sono considerati ottimali in quanto portano nel range delle soglie un caffè non considerato ottimale. Purtroppo quindi non è possibile considerare i dati attendibili per valutare la qualità in tazza.

Quindi l'analisi viene effettuata sui dati della prima sessione sperimentale, che prevedono le seguenti etichette per quanto riguarda i parametri esterni:

- Macinatura: opt(corrispondente a ottimale), grossa(+2 tacche), fine(-2 tacche));
- Dose: ottimale, alta(+2 grammi), bassa(-2 grammi);
- Pressione: ottimale, alta(+2 bar), bassa(-3 bar).

### 6.4.3 Pulizia dei dati

I dati rilevati in laboratorio sono raccolti manualmente, quindi è probabile la presenza di alcuni errori di rilevazione, oppure la presenza di erogazioni non svolte nel migliore dei modi e quindi con caratteristiche diverse dal normale. Per avere dei risultati più attendibili, è stata quindi effettuata una pulizia dei dati. Essa si articola in due parti: per prima cosa si eliminano le erogazioni aventi quantità in tazza minore e maggiore all'interno del test a cui appartengono, come descritto nella sezione 5.2.1, in quanto in genere corrispondono alla prima erogazione di ogni test che ha un erogato basso a causa della variazione del parametro esterno. In seguito si impongono delle soglie sull'erogato e sul tempo di erogazione così da eliminare le rilevazioni non valide. In particolare sono scartate le erogazioni aventi tempo di erogazione al di fuori del range 10-45 secondi e quantità di erogato oltre 10-45 ml. Dopo la pulizia, le osservazioni disponibili sono 955, ovvero all'incirca 18 per ognuno dei 27 test effettuati sui due gruppi.

#### 6.4.4 Soglie di qualità

Le etichette di qualità per le tre caratteristiche del caffè sono assegnate in base a delle soglie di qualità. Nella sezione 6.3 abbiamo approfondito l'argomento: mentre per quanto riguarda le soglie di tempo e erogato in tazza non si hanno dubbi su quali soglie usare, poichè coincidono, per il flusso le soglie degli esperti e quelle proposte a partire dai dati sono decisamente differenti. Per scegliere quali utilizzare, in tabella 6.13, si confrontano i supporti delle etichette di qualità del flusso assunte nei due casi, ovvero quale è la loro frequenza all'interno del dataset.

	Valori	basso (%)	opt (%)	alto (%)
Soglie esperti	0,49 - 1,33 ml/s	1	82	17
Soglie proposte	0,77 - 1,23 ml/s	21	56	23

Tabella 6.13: Supporto delle etichette di qualità del flusso con le diverse soglie

Si può subito vedere come con le soglie degli esperti, nonostante le variazioni estreme dei parametri, più dell'80% delle erogazioni sia in soglia: in questo modo è difficile estrarre delle regole valide, poichè quelle con etichetta diversa da "opt" hanno un supporto troppo basso, mentre quelle con etichetta "opt" hanno un lift prossimo a 1, ovvero rappresentano correlazioni casuali. Con le soglie proposte invece si può vedere come, nonostante il 50% delle erogazioni abbiano flusso in soglia, sia possibile modellare anche il comportamento di un flusso lento, che nel caso precedente non era rilevato.

Si opta quindi per le soglie di flusso proposte a partire dai dati.

Per quanto riguarda la quantità di caffè erogato, dopo aver effettuato la pulizia dei dati, il 100% delle erogazioni risulta in soglia. Ciò perchè l'erogato rimane pressochè costante dopo aver effettuato la taratura. La quantità in tazza però varia a seconda della taratura effettuata e non sembra quindi ragionevole effettuare una analisi sulla quantità di erogato utilizzando delle soglie valide solo per questo esperimento, e che quindi porterebbero a regole non generalizzabili.

Altri parametri da impostare per lo studio delle regole di associazione sono le soglie degli indici di qualità che permettono di capire quali regole hanno valore e quali no. Gli indici di qualità delle regole sono tre, e una loro spiegazione dettagliata è fornita alla sezione 5.5. Brevemente si riprende il concetto e si elencano le soglie stabilite con le motivazioni della scelta:

- **Supporto:** indica quanto è frequente la regola, ovvero quanto è supportata dai dati. Ogni test, come detto, è composto da 18 erogazioni circa per gruppo. Esse rappresentano quindi il 3-4% dell'intero dataset. Per estrarre regole più generali, che non si limitino a modellare un solo test, si sceglie un supporto minimo dell'8%, così da estrarre regole valide per più casi.
- **Confidenza:** indica la "forza" della regola, ovvero quale è la probabilità che dato il corpo della regola, si abbia anche la testa della regola. Un ottimo valore per filtrare regole interessanti è il 50%, così da considerare regole che presentano la testa in almeno la metà dei casi in cui compare il corpo;

- Lift: indica quanto il valore della testa della regola sia dovuto al corpo, ovvero quanto essi siano correlati. Un lift=1 indica che le due parti della regola sono indipendenti, e quindi la regola non è utile. Per scremare le regole si impone un lift>1,2;

In alcuni casi di analisi, si accetteranno comunque regole con valori sotto soglia se ritenute interessanti.

Si procede quindi all'analisi delle regole di associazione per tempo di erogazione e flusso: in particolare, per ognuna delle due saranno affrontate due tipi di analisi:

- Studio dell'effetto delle variabili esterne sulla qualità del caffè: viene considerato l'intero dataset, considerando le condizioni in cui sono svolte le erogazioni come uguali.
- Studio dell'effetto della mancanza dei lavaggi sul comportamento dell'erogazione: il dataset viene diviso in due in base al gruppo di erogazione, poiché mentre un gruppo è stato sempre lavato, l'altro non ha effettuato lavaggi. Attraverso il confronto delle regole estratte si possono ricavare alcune differenze dovute proprio alla mancanza di manutenzione in uno dei due.

### 6.4.5 Regole di associazione per il tempo di erogazione

Come descritto precedentemente, lo studio delle regole di associazione si divide, per ogni caratteristica, in due parti: la prima considera tutte le erogazioni a disposizione per valutare l'effetto delle variazioni dei parametri esterni sulla qualità, mentre la seconda considera due dataset differenti provenienti da due gruppi diversi e attraverso il confronto cerca di dare un'interpretazione alle loro differenze. In questa analisi sul tempo di erogazione, le etichette di qualità sono state assegnate secondo le soglie forniteci dagli esperti, ovvero 20-27 secondi.

#### Influenza delle variabili esterne

La tabella 6.14 contiene le regole riguardanti il tempo di erogazione: esse sono state raggruppate per tipo e per valore dell'etichetta di qualità. Si procede all'analisi delle regole in questione.

Le regole 1-9 indicano la correlazione tra le variabili esterne e un tempo di erogazione alto: una erogazione lenta porta un caffè slavato e poco gustoso, poiché il panetto di macinato viene degradato dall'esposizione prolungata al flusso d'acqua.

La regola 1 indica che una pressione bassa, una macinatura fine e una dose di macinato elevata portano a un tempo maggiore del desiderato. La regola è supportata da una confidenza del 100%, indice del fatto che essa è valida per tutte le erogazioni di questo tipo effettuate nel test. Ciò è comprensivo in quanto il flusso d'acqua ha una spinta minore in quanto la pressione è bassa, e la dose elevata e la macinatura fine presentano un ostacolo che rallenta ulteriormente il passaggio dell'acqua.

Le regole dalla 2 alla 6 considerano solo due delle variabili esterne: si può vedere come la combinazione di macinatura alta e dose elevata, senza tener conto del valore della pressione, produce nella quasi totalità dei casi un caffè lento: ciò è molto interessante per il barista in quanto, mentre la pressione della macchina è impostata periodicamente ed è in genere fissa, dose e macinatura possono variare da erogazione a erogazione. Quindi è interesse del barista aggiustare questi due parametri per evitare un caffè slavato.

Id	Corpo	Testa	Conf (%)	Lift	Supp (%)
1	pressione='bassa', macinatura='fine',dose='alta'	tempo='alto'	100	2.68	4
2	macinatura='fine' , dose='alta'	tempo='alto'	97	2.61	11
3	pressione='bassa' , macinatura='fine'	tempo='alto'	92	2.45	10
4	pressione='bassa', dose='alta'	tempo='alto'	87	2.34	9
5	pressione='bassa',dose ='opt'	tempo='alto'	83	2.23	9
6	macinatura='fine', dose=opt	tempo='alto'	79	2.11	9
7	macinatura='fine'	tempo='alto'	69	1.86	23
8	pressione='bassa'	tempo='alto'	66	1.77	22
9	dose='alta'	tempo='alto'	57	1.54	19
10	pressione='alta', macinatura ='grossa',dose='bassa'	tempo='basso'	100	3.62	4
11	pressione='alta', macinatura ='grossa'	tempo='basso'	81	2.93	9
12	pressione = 'opt', dose='bassa'	tempo='basso'	69	2.49	8
13	macinatura ='grossa'	tempo='basso'	49.0	1.78	16
14	pressione='alta'	tempo='basso'	48.0	1.75	16
15	dose='bassa'	tempo='basso'	46.0	1.67	15
16	pressione='opt',macinatura='opt',dose='opt'	tempo='opt'	97	2.77	4
17	pressione='bassa', macinatura='opt', dose='bassa'	tempo='opt'	97	2.77	4
18	pressione='bassa', macinatura ='grossa',dose='bassa'	tempo='opt'	97	2.77	4
19	pressione='alta', macinatura='fine',dose='bassa'	tempo='opt'	80	2.28	3
20	pressione='bassa',dose='bassa'	tempo='opt'	72	2.05	8
21	pressione='opt',macinatura='fine',dose='bassa'	tempo='opt'	69	1.98	3
22	tempo='opt', dose='bassa'	pressione='bassa'	57.0	1.71	8
23	tempo='opt', dose='bassa'	macinatura='fine'	45.0	1.34	6
24	tempo='alto'	macinatura='fine'	62.0	1.86	23
25	tempo='alto'	pressione='bassa'	58.0	1.77	22
26	tempo='alto'	dose='alta'	51.0	1.54	19
27	tempo='basso'	macinatura ='grossa'	59.0	1.78	16
28	tempo='basso'	pressione='alta'	58.0	1.75	16
29	tempo='basso'	dose='bassa'	56.0	1.67	15

Tabella 6.14: Regole di associazione per il tempo. Studio dell'influenza delle variabili esterne

Le regole 3,4 e 5 portano all'attenzione come una pressione bassa porti spesso fuori soglia il tempo di erogazione, sia che essa sia combinata con una macinatura fine (regola 3),oppure con una dose elevata(regola 4), ma anche in caso di quantità di macinato ottimale. É quindi utile monitorare spesso il valore della pressione della pompa in quanto essa porta un caffè molto lungo anche usando una dose ottimale di macinato.

Per concludere l'analisi riguardante i caffè con tempi elevati, con le regole 7-9 si prende in considerazione l'influenza della singola variabile esterna. Si può vedere come un caffè lento è normalmente più dovuto a una macinatura fine dei grani oppure a una pressione ridotta, rispetto a un incremento del macinato usato. Ciò è comprensibile in quanto una macinatura più fine rende più difficile il passaggio dell'acqua attraverso il panetto, mentre la pressione ridotta riduca la forza penetrante dell'acqua. Un aumento della dose incrementa sì l'ostacolo che l'acqua deve superare, ma al contempo rende anche il caffè più corposo e gustoso.

Le regole 10-15 analizzano invece le cause di un caffè molto veloce, che quindi risulta

poco gustoso in quanto l'acqua non riesce ad assorbire appieno l'aroma del caffè.

La regola 10 cattura il comportamento opposto a quello descritto dalla regola 1: una pressione elevata, combinata con una macinatura grossolana e una dose ridotta di macinato, permette all'acqua di fluire più velocemente in tazza, in quanto la spinta maggiore trova una superficie di contatto ridotta a causa della macinatura grossolana e uno spazio da penetrare ridotto a causa della scarsa dose di macinato. Ciò è valido per tutti gli esperimenti svolti con queste caratteristiche, come indicato dalla confidenza del 100%. La regola successiva spiega come la combinazione di una macinatura grossolana e una pressione elevata produca nell'80% dei casi un caffè molto veloce, a prescindere dalla dose utilizzata. Se invece si opera con pressione ottimale, una dose ridotta porta con buona probabilità una erogazione rapida (regola 12). Le ultime tre regole di questa sezione (13-15) indicano la relazione tra le singole variabili esterne con il tempo di erogazione basso: esso sembra dovuto maggiormente a una macinatura grossolana e a una pressione elevata, e leggermente meno a una dose ridotta. Si vede come l'ordine di influenza sia lo stesso di quello trovato nell'analisi del tempo di erogazione elevato. La differenza è però nel valore della confidenza: si vede come essa sia leggermente sotto il 50%, indice che l'influenza della singola variabile dipende dalla combinazione con le altre: infatti, la regola 12 dimostra come una dose ridotta combinata con una pressione ottimale, porti molto spesso un caffè veloce. Si può concludere quindi che la Lavazza è in particolar modo interessata all'analisi di questo tipo di caffè in quanto dovuti all'uso ridotto di macinato e a una macinatura grossolana dei grani: ciò arreca danno all'azienda sia dal punto di vista economico, poichè l'uso di una dose ridotta di macinato porta ad un acquisto minore di grani, e quindi a vendite ridotte per la Lavazza, sia dal punto di vista dell'immagine, poichè il caffè risulta poco gustoso. Le regole della sezione successiva descrivono invece le combinazioni dei parametri esterni che producono un caffè con tempo di erogazione ottimale. Esse rappresentano delle regole di compensazione e permettono di modellare il comportamento assunto dal test a cui si riferiscono: infatti presentano un supporto molto basso, del 3-4%, che rappresenta appunto il numero di erogazioni svolto con una determinata combinazione. La regola 16 rappresenta le erogazioni ottimali: si può vedere dalla confidenza del 97% che esse risultano avere sempre tempo in soglia, a meno di una erogazione che, analizzando il valore rilevato, esce dalle soglie di 0,3 secondi, e può quindi essere considerata come buona. Le regole 17 e 18 dimostrano come pressione bassa e dose ridotta si compensino sempre tra loro, a prescindere dal fatto che si utilizzi una macinatura ottimale o grossolana. Questa conclusione è anche supportata dall'alta confidenza della regola 20, che indica che nel 72% dei casi le due variabili si compensano: è plausibile affermare che il 28% mancante sia dovuto a una mancata compensazione in presenza di macinatura fine, che rallenta ulteriormente l'acqua. Infine, le regole 19 e 20 dimostrano come macinatura fine e dose ridotta si compensino meglio in presenza di una pressione elevata (80%) rispetto a una condizione della pompa ottimale.

Le regole 22 e 23 sono anch'esse regole di compensazione che però affrontano il problema da un'altra prospettiva: ci si chiede quale sia il fattore di compensazione più frequente nei caffè aventi per certo un tempo ottimale e un altro parametro non ottimale. In particolare si pone l'attenzione a come si possa ottenere un caffè ottimale utilizzando una dose di macinato ridotta: è più frequente ottenere un caffè ottimale abbassando la pressione, rispetto all'uso di una macinatura più fine. Ciò è di grande interesse per l'azienda, in quanto essendo difficile compensare una bassa quantità con una macinatura più fine, l'unico metodo rimanente è l'abbassamento della pressione. Esso è però monitorabile tramite telemetria,

ed è quindi rilevabile una sua variazione. In conclusione l'utilizzo di una dose ridotta di macinato è individuabile tramite il riconoscimento di un tempo di erogazione ridotto.

In ultimo, sono analizzate le cosiddette regole di origine, il cui obiettivo è quello di estrarre informazioni su quale sia la causa più probabile di tempi elevati e ridotti per l'erogazione di un caffè. Le regole 24 e 27 spiegano che la causa di tempi di erogazione fuori soglia è causata più frequentemente da una variazione della macinatura, mentre è meno dovuta ad una dose di macinata diversa da quella standard. In particolare per quanto riguarda i tempi di erogazione ridotti, è però difficile discriminare efficacemente la causa esterna, in quanto le confidenze delle regole sono molto simili.

Si può quindi concludere che un tempo maggiore è causato da una macinatura fine, da una pressione bassa e infine da una dose maggiore di macinato. Un tempo minore invece è causato dalle rispettive azioni complementari, ovvero da macinatura grossolana, pressione alta e dose ridotta.

### Impatto dei lavaggi sul tempo di erogazione

Viene ora svolta una analisi analoga a quella appena effettuata, svolta però su due dataset differenti. Le erogazioni analizzate sono infatti state svolte su due gruppi differenti. In un uno, dopo ogni erogazione, viene svolta una purge, ovvero un lavaggio di 8-10 secondi che elimina residui di polvere e olii dai condotti del gruppo. Nell'altro gruppo ciò non avviene. Attraverso il confronto di alcuni comportamenti ottenuti grazie all'analisi con le regole di associazione sui due gruppi si cercano differenze dovute al diverso grado di pulizia dei gruppi.

La tabella 6.15 contiene alcune regole interessanti, con i valori di confidenza, lift e supporto assunti nei due differenti gruppi.

Id	Corpo	Testa	Non pulito			Pulito		
			Conf (%)	Lift	Supp (%)	Conf (%)	Lift	Supp (%)
1	pressione='bassa', dose='opt'	tempo='alto'	96	2.20	11	70	2.27	8.0
2	pressione='opt', dose='alta'	tempo='alto'	67.0	1.53	8.0	44.0	1.43	5.0
3	pressione='alta', macinatura='grossa'	tempo='basso'	68.0	2.91	8.0	94.0	2.96	10.0
4	pressione='alta', macinatura='opt'	tempo='basso'	40.0	1.70	4.0	81.0	2.56	9.0
5	pressione='alta'	tempo='basso'	36.0	1.56	12.0	60.0	1.88	20.0
6	pressione='bassa', macinatura='grossa'	tempo='opt'	31.0	0.95	3.0	80.0	2.15	9.0
7	pressione='opt', macinatura='opt'	tempo='opt'	37.0	1.12	4.0	60.0	1.63	7.0
8	tempo='opt', macinatura='grossa'	pressione='bassa'	31.0	0.94	3.0	69.0	2.07	9.0
9	tempo='opt', pressione='bassa'	dose='bassa'	94.0	2.79	7.0	60.0	1.79	9.0
10	tempo='opt', dose='opt'	pressione='alta'	58.0	1.74	6.0	27.0	0.83	3.0
12	tempo='alto'	macinatura='fine'	59.0	1.77	26	67.0	1.99	21
13	tempo='alto'	pressione='bassa'	57.0	1.76	25	60.0	1.79	19
14	tempo='alto'	dose='alta'	49.0	1.50	21	53.0	1.60	17
15	tempo='basso'	pressione='alta'	52.0	1.56	12	62.0	1.88	20
16	tempo='basso'	macinatura='grossa'	64.0	1.91	15	56.0	1.69	18
17	tempo='basso'	dose='bassa'	65.0	1.93	15	49.0	1.48	16

Tabella 6.15: Regole di associazione per il tempo. Studio dell'influenza dei lavaggi

La regola 1 porta alla luce come, combinando una pressione bassa con una dose di macinato ottimale, in entrambi i gruppi si ottenga un tempo di erogazione alto. In particolare la regola si presenta più volte nel gruppo sporco (supporto del 11% contro 8% del gruppo pulito) e che essa sia valida nella quasi totalità dei casi in mancanza di pulizia. Le purge permettono la rimozione di oli e residui di macinato dai condotti della macchina, rimuovendo quindi gli ostacoli che possono rallentare lo scorrere dell'acqua. Ciò che la regola mette in risalto è come nel gruppo sporco, che presenta quindi alcuni rallentamenti, l'azione rallentante di una pressione ridotta è amplificata proprio a causa di questi residui. Nel gruppo pulito invece il flusso è più veloce e quindi, grazie anche alla compensazione con altri parametri esterni, l'impatto della pressione bassa è minore. La regola successiva è quella complementare alla precedente: con una pressione ottimale, una dose di macinato alta provoca un tempo di erogazione elevato. In particolare si nota come la confidenza della regola nel gruppo pulito sia inferiore al 50%, ovvero che la sua validità sia limitata. Invece, come spiegato per la regola precedente, nel gruppo sporco il degrado accumulato rallenta ulteriormente il fluire dell'acqua, aumentando il tempo di erogazione. Le regole 3, 4 e 5 prendono in analisi la pressione alta e la sua relazione con un tempo di erogazione ridotto: si può notare come supporto e confidenza delle tre regole siano molto più elevate nel caso del gruppo pulito. Infatti, il gruppo pulito permette il passaggio più fluido dell'acqua, che presenta una velocità maggiore a causa dell'elevata pressione. In particolare, le regole 3 e 4 mettono in risalto come la macinatura influisca in modo diverso nei due gruppi: se infatti nel gruppo pulito, utilizzando una macinatura grossolana o una ottimale, il tempo di erogazione resta basso in entrambi i casi (infatti supporto e confidenza sono molto simili), nel gruppo sporco il comportamento varia molto, e raramente si ottiene un caffè molto veloce con una macinatura ottimale e pressione alta. Ciò è spiegabile con il fatto che l'effetto della pressione è nascosto dal rallentamento dovuto al degrado, e quindi è solo la presenza di una superficie di impatto ridotta a rallentare l'erogazione.

Le ultime due regole di questa sezione indagano invece il diverso comportamento di compensazione con cui si può ottenere un caffè ottimale. La regola 6 infatti dimostra come nell'80% dei casi, nel gruppo pulito, una macinatura grossolana si compensi con una pressione bassa. Ciò invece non avviene nel gruppo sporco, dove si ha una confidenza del 30% e soprattutto un lift minore di 1, che indica una correlazione negativa, ovvero che la combinazione di macinatura e pressione non presenti un tempo di erogazione ottimale. In conclusione, la regola 7 dimostra come, considerati pressione e macinatura ottimali, il tempo di erogazione sia ottimale nel 60% dei casi in presenza di lavaggi, e solo nel 37% in loro assenza. Ciò è dovuto al fatto che il degrado accumulato rallenta il fluire dell'acqua e porta a caffè lenti non solo quando la dose è maggiore dello standard, ma anche quando coincide con lo standard.

La sezione successiva analizza invece le regole di compensazione, ovvero quale, dato un tempo ottimale e una variabile esterna, quale sia la variazione che permette la compensazione. La regola 8 mostra come sia più frequente la compensazione di una macinatura grossolana con una pressione bassa nel gruppo pulito. Anzi, nel gruppo sporco la regola ha lift inferiore a 1, che indica che la compensazione della macinatura non avviene con la pressione: ciò poichè l'effetto della pressione è ostacolato dai residui presenti nei condotti e non è in grado di compensare il rapido flusso causato dalla presenza di una superficie di impatto col panetto inferiore. La regola successiva ci informa di un comportamento in qualche modo inverso a quello appena descritto: nel gruppo sporco, una pressione bassa

può essere compensata da una dose ridotta di macinato nella quasi totalità dei casi. Nel gruppo pulito ciò avviene invece nel 70% dei casi. Questo indica che nel gruppo sporco, qualsiasi sia il valore della macinatura, pressione bassa e dose ridotta si compensano. Invece nel gruppo pulito l'effetto della macinatura è più rilevante. Con la regola 10 si approfondisce invece il caso in cui, in un erogazione che conosciamo avere tempo ottimale e dose ottimale, si presenti anche una pressione elevata. Mentre nel gruppo pulito, come ci si aspetterebbe, ciò si verifica raramente ed è caratterizzato da un lift negativo, nel gruppo sporco questo evento ha una certa rilevanza. Si può quindi concludere che, in un gruppo non lavato, è necessaria una pressione più elevata per ottenere un tempo di erogazione in soglia nonostante l'utilizzo della giusta quantità di macinato.

Infine, nell'ultima sezione si confrontano le regole di origine. Dalle regole 11-13 si può concludere che la causa principale di un tempo elevato è la macinatura fine, seguita da una pressione bassa e una dose di macinato elevata. Ciò non varia a seconda del grado di pulizia, ma in un gruppo lavato la macinatura sembra influire maggiormente rispetto a quello sporco. Per quanto riguarda le cause di un tempo di erogazione ridotto, invece, si nota come il comportamento dei due gruppi sia opposto. Infatti, mentre nel gruppo pulito è la pressione la principale causa di una erogazione veloce (regola 14), seguita da macinatura grossolana e dose ridotta (regole 15 e 16 rispettivamente), nel gruppo non pulito è proprio la dose ridotta la prima causa di una erogazione corta. Anche la macinatura ha una confidenza comparabile con quella della dose ridotta, mentre più staccata è presente la pressione. Quindi si può concludere che l'effetto dei residui nei condotti della macchina riduce molto l'effetto di un aumento della pressione, producendo quindi caffè comunque in un tempo ottimale. In questo modo la dose ridotta di macinato e la macinatura grossolana dei grani assume una rilevanza maggiore come causa di un tempo ridotto.

In conclusione, grazie alle regole di associazione è possibile distinguere il comportamento dei due gruppi causata dalla differenza nella pulizia dei condotti:

- La sporcizia accumulata attenua in particolar modo la riduzione del tempo di erogazione a causa di un aumento di pressione, mentre amplifica il rallentamento causato da un abbassamento di essa;
- Una adeguata pulizia permette una maggiore compensazione tra le variabili esterne. Quindi un cliente che effettua i lavaggi è in grado di ottenere un maggiore effetto di compensazione. Ciò è utile poiché, sapendo se un cliente effettua o meno i lavaggi, possiamo essere abbastanza certi che non sia possibile per lui mascherare una riduzione della dose di macinato utilizzato con una modifica della pressione e della macinatura.
- Se si è a conoscenza che un cliente non effettua i lavaggi, e presenta dei tempi di erogazione lenti, la causa potrebbe essere associata a un uso ridotto di macinato, e si può quindi effettuare un'analisi delle vendite così da approfondire se il cliente effettua un numero di caffè compatibili con la quantità di caffè comprato.

#### 6.4.6 Regole di associazione per il flusso

Dopo aver affrontato l'analisi delle regole di associazione per il tempo di erogazione, si segue lo stesso impostazione con l'analisi del flusso. Si ricorda che per l'assegnazione delle etichette sono state utilizzate le soglie proposte a partire dai dati (0.77-1.23), in quanto,

come analizzato nella sezione 6.4.4, permettono una distinzione migliore dei valori di flusso registrati, permettendo così una migliore modellazione del comportamento. Questa sezione, come la precedente, è divisa in due parti: la prima studia l'influenza delle variabili esterne sul flusso, mentre la seconda confronta le erogazioni svolte in un gruppo pulito e in un gruppo sporco, così da modellare come la mancanza dei lavaggi influenzi l'erogazione.

### Influenza delle variabili esterne

La tabella 6.16 contiene le regole riguardanti il flusso: esse sono state raggruppate per tipo e per valore dell'etichetta di qualità. Si procede all'analisi delle regole in questione.

Id	Corpo	Testa	Conf (%)	Lift	Supp (%)
1	pressione='opt', macinatura='grossa', dose='bassa'	flusso='alto'	100	4.19	4
2	pressione='alta', macinatura='grossa', dose='bassa'	flusso='alto'	100	4.19	4
3	pressione='alta', macinatura='opt', dose='bassa'	flusso='alto'	89	3.72	3
4	pressione='opt', macinatura='grossa', dose='opt'	flusso='alto'	86	3.59	3
5	pressione='alta', macinatura='grossa'	flusso='alto'	77	3.23	8
6	pressione='opt', macinatura='grossa'	flusso='alto'	64	2.66	7
7	pressione='alta', dose='bassa'	flusso='alto'	63	2.65	7
8	pressione='opt', dose='bassa'	flusso='alto'	56	2.33	6
9	macinatura='grossa'	flusso='alto'	47	1.97	16
10	pressione='alta'	flusso='alto'	43	1.80	14
11	pressione='bassa', macinatura='fine', dose='alta'	flusso='basso'	82	4.47	3
12	pressione='bassa', macinatura='opt', dose='alta'	flusso='basso'	82	4.44	3
13	pressione='opt', macinatura='fine', dose='alta'	flusso='basso'	75	4.07	3
14	macinatura='fine', dose='alta'	flusso='basso'	66	3.60	7
15	pressione='bassa', dose='alta'	flusso='basso'	65	3.51	7
16	pressione='bassa', macinatura='fine'	flusso='basso'	61	3.33	7
17	dose='alta'	flusso='basso'	38	2.06	12
18	pressione='bassa'	flusso='basso'	37	2.01	12
19	macinatura='fine'	flusso='basso'	35	1.91	12
20	pressione='bassa', macinatura='grossa', dose='bassa'	flusso='opt'	100	1.73	4
21	pressione='alta', macinatura='fine', dose='bassa'	flusso='opt'	100	1.73	4
22	pressione='alta', macinatura='fine'	flusso='opt'	86	1.49	9
23	dose='bassa', macinatura='fine'	flusso='opt'	82	1.43	9
24	flusso='opt', dose='bassa'	pressione='bassa'	51	1.54	9
25	flusso='opt', dose='bassa'	macinatura='fine'	50	1.49	9
26	flusso='basso'	dose='alta'	68	2.06	12
27	flusso='basso'	pressione='bassa'	66	2.01	12
28	flusso='basso'	macinatura='fine'	64	1.91	12
29	flusso='alto'	macinatura='grossa'	65	1.97	16
30	flusso='alto'	pressione='alta'	60	1.80	14
31	flusso='alto'	dose='bassa'	56	1.66	13

Tabella 6.16: Regole di associazione per il flusso. Studio dell'influenza delle variabili esterne

Le prime dieci regole indicano la correlazione tra variabili esterne e flusso alto, ovvero quando l'acqua attraversa il panetto troppo velocemente senza riuscire ad assorbire appieno l'aroma del caffè. Le regole 1 e 2 indicano che in presenza di macinatura grossolana, dose ridotta e pressione, sia ottimale che bassa, il flusso risulta sempre elevato. Infatti le due

regole hanno confidenza del 100%. Ciò corrisponde al comportamento evidenziato con le regole di associazione per il tempo. Infatti, con una macinatura grossolana la superficie di contatto dell'acqua risulta inferiore e quindi essa è più veloce nell'attraversare il panetto. Inoltre la bassa dose di macinato utilizzato rende l'ostacolo da superare inferiore, e quindi si riduce il tempo necessario per la penetrazione. L'aumento della pressione incrementa quindi ulteriormente la velocità dell'acqua, e di conseguenza il flusso. Anche senza un aumento della pressione il caffè viene erogato comunque con un flusso molto elevato.

La regola 3 analizza il caso in cui si abbiano macinatura ottimale, pressione alta e dose ridotta. Nel 89% dei casi questa combinazione fornisce un flusso elevato, poichè l'acqua è spinta maggiormente dalla pompa e incontra un ostacolo ridotto a causa della bassa dose di macinato utilizzata. La regola successiva invece evidenzia come nell'86% dei casi una macinatura grossolana influisca negativamente sul flusso: è interesse del barista quindi regolare correttamente la macinatura dei grani di caffè in fase di taratura così da ottenere un caffè corposo e molto gustoso. A conferma della regola precedente le regole 5 e 6 danno l'idea di come una macinatura grossolana inducano un flusso elevato sia con una pressione elevata, che amplifica la velocità dell'acqua, sia con una pressione ottimale: in quest'ultimo caso si ha una confidenza del 64% che rappresenta, sulla base delle scoperte precedenti, che la macinatura a grani grossi può portare a caffè dal flusso ottimale solo in un caso su tre, ovvero quando si usa una dose ridotta. Le due regole successive riportano lo stesso comportamento ottenuto con una macinatura grossa, considerando però una dose ridotta di macinato. Essa, in presenza di una pressione ottimale, produce un flusso elevato nel 56% dei casi (regola 8). Con l'effetto combinato dell'aumento della pressione aumenta la probabilità di avere uno scorrere dell'acqua molto veloce. Infine le regole 9 e 10 considerano l'influenza di una sola variabile esterna alla volta: si può notare come macinatura grossolana e pressione alta producano un flusso alto, però solo nel 45% dei casi circa: un indice così basso è dovuto probabilmente a effetti di compensazione tra le varie variabili.

Con le regole 11-19 si approfondisce invece il caso in cui un caffè presenti un flusso molto lento: una erogazione lenta permette all'acqua di impregnarsi appieno dell'aroma del caffè, ma la lunga esposizione del panetto alla pressione esercitata dall'acqua ne deteriora le caratteristiche. Le prime tre regole rispecchiano le caratteristiche delle prime tre della sezione precedente: si nota infatti la contrapposizione tra pressioni alte e basse, macinature fini e grossolane e dosi ridotte e elevate. Si nota però che la confidenza delle regole per il flusso basso non hanno una confidenza uguale o simile al 100%, ma leggermente più bassa. Questo comportamento stupisce poichè le dinamiche dell'erogazione dovrebbero essere duali, ma non è così: ciò potrebbe essere causato da una differenza tra il comportamento dei due gruppi su cui sono state effettuate le erogazioni, e saranno indagate nella sezione successiva con l'analisi dell'influenza dei lavaggi sulla qualità del flusso in tazza. Le successive regole analizzano l'influenza della combinazione di due parametri esterni su un flusso lento: le tre regole contengono le combinazioni di macinatura fine, dose alta e pressione bassa. Tra queste, combinazione dose elevate e macinatura fine è quella che produce più spesso un flusso lento, poichè il panetto di caffè risulta molto spesso e compatto, rendendo difficile il passaggio dell'acqua. Quasi con la stessa frequenza, una dose elevata e una pressione bassa producono una erogazione lenta, come la restante combinazione tra pressione bassa e macinatura fine. Le ultime tre regole indicano invece quale è l'influenza di una sola variabile su un flusso lento: si può subito notare che la confidenza è molto bassa, poichè l'effetto delle variabili presenta diversi effetti di compensazione.

Nella terza sezione, con le regole 20-21, si analizzano gli effetti di compensazione tra variabili esterne. Le regole 20 e 21 dimostrano come, con l'uso di una dose ridotta di macinato, pressione e macinatura non ottimali permettano di avere comunque un flusso ottimale. In particolare, la pressione bassa compensa una dose ridotta e una macinatura grossolana. Invece con una pressione elevata e una macinatura molto fine si può mascherare l'effetto velocizzante dell'uso di una dose ridotta di macinato. Il cliente potrebbe quindi usare sempre dosi ridotte ottenendo un buon caffè regolando in modo corretto macinatura e pressione. In questo modo ricaverebbe un guadagno economico in quanto in ogni erogazione vengono risparmiati alcuni grammi di caffè. La Lavazza può però individuare un comportamento del genere monitorando la pressione della macchina, rilevata tramite telemetria, e correggere il comportamento scorretto. Le regole 22 e 23 mostrano invece gli effetti di compensazione che si ottengono tenendo in considerazione solo due variabili, e ammettendo per la terza qualsiasi valore. Come visto precedentemente, pressione alta e macinatura si compensano producendo un caffè con flusso ottimale nell'86% dei casi. Inoltre anche modificando solamente macinatura e dose si possono ottenere caffè in soggia: infatti una dose ridotta, se macinata finemente, oppone maggiore resistenza al flusso d'acqua e produce nell'82% dei casi un caffè ottimale. Considerando poco probabile una variazione della pressione, poichè richiede una manovra meccanica non banale ed è facilmente rilevabile attraverso la telemetria, un cliente che vuole usare una dose di macinato ridotto così da risparmiare sulle spese per i grani da macinare può macinare finemente i grani stessi così da rallentare il flusso e ottenere un caffè che, dai dati rilevati, sembra effettuato correttamente.

A supporto di questo ultimo risultato, le regole 24 e 25 descrivono come in un caffè con flusso ottimale, una dose di macinato sia compensata in maniera equivalente sia da una pressione bassa che da una macinatura fine.

L'analisi si conclude con l'investigazione sulla principale causa di un flusso veloce e lento in un caffè: per quanto riguarda i caffè con erogazione lenta, nessuna delle tre variabili ha un'influenza più elevata delle altre nel rallentare lo scorrere dell'acqua. Seppur di poco, una dose elevata di macinato porta in più casi a caffè lenti(68%) poichè il panetto di macinato da superare è più spesso e rallenta sensibilmente l'erogazione. La pressione è il secondo fattore esterno più caratterizzante un'erogazione lenta e infine si ha l'effetto rallentante della macinatura fine che aumenta la superficie di contatto dell'acqua con il macinato. Per quanto riguarda le erogazioni veloci, la principale causa è rappresentata dalla macinatura grossolana dei grani(64%), seguita da una pressione elevate(60%) e dall'uso di una dose ridotta di macinato(56%).

### **Impatto dei lavaggi sul tempo di erogazione**

Si passa quindi all'analisi dell'impatto dei lavaggi sul flusso in tazza. Calcoliamo le regole di associazione separatamente sui dati dei due gruppi aventi caratteristiche di manutenzione diverse, poichè solo su uno sono state svolte sempre le purge. La tabella 6.17 contiene alcune regole interessanti, con i valori di confidenza, lift e supporto assunti nei due differenti gruppi.

Risultati sperimentali

Id	Corpo	Testa	Non pulito			Pulito		
			Conf (%)	Lift	Supp (%)	Conf (%)	Lift	Supp (%)
1	pressione='alta', macinatura='grossa'	flusso='alto'	58	2.97	7	96	3.44	10
2	pressione='alta', macinatura='opt'	flusso='alto'	28	1.44	3	74	2.65	8
3	pressione='alta'	flusso='alto'	29	1.49	10	56	2.01	19
4	macinatura='grossa'	flusso='alto'	39	1.97	13	55	1.97	18
5	dose='bassa'	flusso='alto'	42	2.13	14	37	1.33	12
6	pressione='bassa', macinatura='fine', dose='alta'	flusso='basso'	100	4.29	3	67	4.88	2
7	pressione='bassa', macinatura='opt', dose='alta'	flusso='basso'	100	4.29	3	67	4.88	2
8	macinatura='fine', dose='alta'	flusso='basso'	84	3.60	9	50	3.66	6
9	pressione='bassa', dose='alta'	flusso='basso'	85	3.67	9	46	3.39	5
10	dose='alta'	flusso='basso'	71	2.19	17	25	1.86	8
11	pressione='opt', dose='alta'	flusso='basso'	54	2.30	6	19	1.36	2
12	pressione='opt', macinatura='opt'	flusso='opt'	50	0.88	6	89	1.52	10
13	pressione='opt', dose='alta'	flusso='opt'	46	0.81	5	78	1.33	9
14	pressione='alta', dose='opt'	flusso='opt'	87	1.52	10	44	0.76	5
15	flusso='opt', pressione='alta'	macinatura='fine'	41	1.24	9	75	2.24	10
16	flusso='opt', macinatura='grossa'	pressione='bassa'	42	1.30	7	74	2.21	11
17	flusso='alto'	pressione='alta'	49	1.49	10	67	2.01	19
18	flusso='alto'	macinatura='grossa'	66	1.97	13	65	1.97	18
19	flusso='alto'	dose='bassa'	72	2.13	14	44	1.33	12
20	flusso='basso'	pressione='bassa'	61	1.87	14	76	2.26	10
21	flusso='basso'	macinatura='fine'	58	1.75	14	73	2.17	10
22	flusso='basso'	dose='alta'	51	2.19	17	62	1.86	8

Tabella 6.17: Regole di associazione per il flusso. Studio dell'influenza dei lavaggi

Le prime regole analizzano il comportamento che provoca un flusso alto nei due gruppi. La regola 1 evidenzia che una pressione alta e una macinatura grossolana provocano un flusso elevato. Ciò si verifica in entrambi i gruppi, ma mentre nel gruppo pulito questo comportamento comprende praticamente la totalità delle erogazioni (96%), nel gruppo sporco solo il 58% di esse rispetta la regola. Ciò è causato dai residui di caffè e oli presenti nella macchina e non puliti, che rallentano il fluire dell'acqua, che nonostante la pressione elevata e la facilità nel passare attraverso il macinato in quanto molto largo, risulta in soglia molto spesso. La regola successiva evidenzia maggiormente questo comportamento, in quanto solo nel 28% delle erogazioni aventi una macinatura ottimale ma una pressione elevata, il flusso risulta elevato. Nel gruppo pulito invece addirittura il 74% di queste erogazioni presenta un flusso alto. Si può in pratica dedurre che il degrado accumulato compensa quasi completamente la pressione elevata della pompa. Con le tre regole successive si può infatti vedere come nel gruppo pulito alte pressioni e macinature larghe portano più della metà dei caffè ad avere un flusso elevato, mentre nel gruppo sporco soprattutto la pressione alta non influenza il flusso in maniera sostanziale.

Successivamente si confronta il comportamento dei gruppi nel caso di variabili esterne che producono caffè lenti. Durante l'analisi dell'influenza delle variabili esterne sul flusso, si è notato come con pressione elevata, macinatura grossolana e dose ridotta si ottenesse sempre un flusso elevato, mentre con i valori duali degli stessi parametri esterni si ottenesse un flusso basso solo nell'82% dei casi. Osservando le prime due regole della sezione è chiaro che questa differenza sia attribuibile alla presenza dei lavaggi nel gruppo pulito, che permettono una maggiore velocità dell'acqua e quindi una maggiore forza penetrante nel panetto: in questo modo la quantità d'acqua impostata in taratura riesce ad attraversare il panetto in un tempo minore nonostante le caratteristiche dei parametri esterni rallentino

il flusso d'acqua. Si parla di piccole quantità di tempo(1-2 secondi) che permettono però al flusso, che è una misura derivata, di rimanere in soglia. Infatti si nota come nel gruppo sporco la totalità delle erogazioni abbia flusso maggiore della soglia impostata, mentre nel gruppo pulito solo il 62% presenti un flusso elevato.

Le regole 8, 9 mostrano come l'effetto combinato di una dose elevata con una pressione bassa oppure una macinatura porti l'84% delle erogazioni ad avere un flusso lento nel gruppo sporco, mentre nel gruppo pulito solo metà delle erogazioni siano lente. Si può quindi nuovamente vedere l'effetto dei residui presenti nei condotti che rallentano sensibilmente le erogazioni. Infine le regole 10 e 11 ribadiscono il fatto che una dose elevata di macinato, produce un flusso lento solo nel gruppo sporco e non nel gruppo pulito. Quindi si può concludere che è la dose di macinato la variabile esterna che più risente della mancata pulizia del gruppo, mentre in presenza di lavaggi non modifica in maniera significativa il flusso.

La regola 12 risulta interessante poichè fornisce la prova che una variazione di dose di macinato non è molto influente nel gruppo pulito, ma diventa determinante nel gruppo sporco. Infatti la regola descrive che nel gruppo pulito macinatura e pressioni ottimali portano un flusso ottimale nell'89% dei casi: considerando che solo il 33% delle erogazioni con pressione e macinatura ottimale presenta anche dose ottimale, essa non influisce nel modificare il flusso nel gruppo pulito. Anche con la regola 13 si vede come con una pressione ottimale e una dose elevata il flusso nel gruppo pulito risulti comunque ottimale, mentre nel gruppo sporco si ha un lift minore di 1, che indica una correlazione negativa tra le due variabili esterne e il flusso ottimale. La regola 14 invece evidenzia un comportamento analogo ma invertito. Considerata una dose ottimale e una pressione alta, il gruppo sporco non risente dell'accelerazione dell'acqua e produce caffè ottimali nell'87% dei casi mentre nel gruppo pulito solo il 44% delle erogazioni presenta flusso ottimale. Inoltre nel gruppo pulito la regola ha un supporto più basso e in particolare un lift negativo, che indica che la regola con la testa negata è più efficace, ovvero che una pressione alta e una dose elevata portino a flussi non ottimali con una probabilità più alta.

Le regole 15 e 16 esprimono come la pressione alta sia compensata più volte nel gruppo pulito rispetto a quello sporco, poichè l'effetto della pressione nel gruppo sporco è limitato dalla presenza di residui e oli in eccesso. Allo stesso modo, una pressione bassa compensa meglio una macinatura grossolana nel gruppo pulito rispetto a quello in cui non sono svolti i lavaggi.

Per concludere, con le regole 17-21 si conferma il fatto che la dose abbia un impatto rilevante nel gruppo sporco mentre l'effetto della variazione della pressione è oscurato dalla pulizia. Nel gruppo pulito è invece proprio la pressione il primo fattore che porta il flusso fuori soglia.

Quindi se si è a conoscenza che un cliente non effettua i lavaggi della macchina e presenta molti caffè con flusso fuori soglia, molto probabilmente sta usando dosi di macinato diverse da quelle ottimali.

### **6.4.7 Regole di associazione per il flusso con tre livelli di soglie**

Dopo aver effettuato l'analisi dell'influenza delle variabili esterne con le regole di associazione ci si è resi conto che l'interesse dell'azienda è quello di modellare in particolare il comportamento di un caffè che presenta flusso veloce. La motivazione di questo interesse

è dovuta al fatto che un caffè veloce è portato da una macinatura grossolana e soprattutto all'utilizzo di una dose ridotta di macinato. Per la Lavazza questo rappresenta un danno economico, oltre che di immagine, in quanto due grammi di macinato risparmiati per ogni caffè rappresentano un risparmio di 1 kg di caffè in un bar che effettua 500 caffè al giorno. Si è deciso di imporre un'ulteriore soglia per il flusso così da poter evidenziare quei caffè che presentano un flusso non eccessivamente veloce, ma che è maggiore dell'1 ml/s ideale. Sono quindi state effettuate due proposte di soglie a 3 livelli per il flusso, come descritto nella sezione 6.3.2. Alle erogazioni sono state quindi assegnate 4 possibili etichette sulla base del valore del flusso: 'basso', 'ok\_basso', 'ok\_alto', 'alto'. La tabella 6.18 contiene i supporti delle varie categorie individuate con le due proposte di soglie effettuate.

	Soglie	basso(%)	ok_basso(%)	ok_alto(%)', alto(%)
Prima proposta	0,73-1-1,23	21	35	21 23
Seconda proposta	0,73-1,1-1,23	21	46	10 23

Tabella 6.18: Supporti con soglie di flusso a tre livelli.

Si può subito vedere come la seconda proposta presenti, per l'etichetta 'ok\_alto' un supporto del 10%, decisamente troppo basso per estrarre regole di validità generale. Infatti, disegnando su un grafico(figura 6.11) le regole riguardanti l'etichetta 'ok\_alto' in base al loro supporto(asse x) e la loro confidenza(asse y) si può vedere come le uniche regole con un'alta confidenza siano quelle con supporto molto basso, che si riferiscono a un solo test effettuato, e che quindi non forniscono informazioni generali.

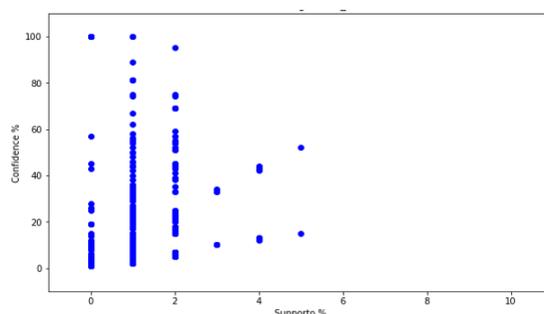


Figura 6.11: Regole estratte per 'ok\_alto' con soglie a tre livelli per il flusso 0,73-1,1-1,23

Si sceglie quindi la prima proposta di soglie (0,77-1-1,23) e si estraggono le regole di associazione per il flusso in questo caso. Si noti come non verranno affrontate le regole per il caso di flusso basso e alto, in quanto coincidenti con il caso precedente. In particolare l'interesse è rivolto alle regole con 'ok\_alto', poichè si vuole porre l'attenzione nei casi in cui il cliente aumenta la velocità dell'erogazione per effettuare più caffè e per risparmiare dose di macinato.

Si analizzano sia l'influenza delle variabili esterne che le differenze tra i due gruppi dovute alla mancanza di pulizia.

### Impatto delle variabili esterne con soglie a tre livelli

La tabella 6.19 contiene le regole di associazione che indagano l'influenza delle variabili esterne sul flusso. In particolare sono analizzate le regole per flusso 'ok\_alto' e 'ok\_basso', in quanto le regole per 'alto' e 'basso' sono già state analizzate precedentemente.

Id	Corpo	Testa	Conf (%)	Lift	Supp (%)
1	pressione='alta', macinatura='opt', dose='opt'	flusso='ok_alto'	90	4.39	3
2	pressione='opt', macinatura='opt', dose='opt'	flusso='ok_alto'	89	3.72	2
3	pressione='opt', macinatura='fine', dose='bassa'	flusso='ok_alto'	70	3.26	3
4	pressione='opt', macinatura='grossa', dose='alta'	flusso='ok_alto'	82	4.84	3
5	pressione='alta', dose='opt'	flusso='ok_alto'	44	2.06	5
6	pressione='bassa', macinatura='opt', dose='bassa'	flusso='ok_basso'	95	2.90	4
7	pressione='bassa', dose='bassa'	flusso='ok_basso'	70	2.14	8
8	macinatura='fine', dose='opt'	flusso='ok_basso'	68	1.95	8
9	pressione='bassa'	flusso='ok_basso'	52	1.51	17
10	flusso='ok_alto'	pressione='opt'	47	1.41	10
11	flusso='ok_alto'	macinatura='opt'	44	1.32	9
12	flusso='ok_alto'	pressione='alta'	41	1.32	9
13	flusso='ok_alto'	dose='bassa'	37	1.10	8
14	flusso='ok_basso'	pressione='bassa'	50	1.51	17
15	flusso='ok_basso'	macinatura='fine'	47	1.41	16
16	flusso='ok_basso'	dose='opt'	44	1.31	15

Tabella 6.19: Regole di associazione per il flusso con soglie a 3 livelli. Studio dell'influenza delle variabili esterne.

Si nota come le regole per il flusso 'ok\_alto' abbiano dei supporti molto ridotti. In questo però le confidenze sono molto buone. Invece, con l'aumento del supporto la confidenza cala drasticamente. La regola 1 descrive come, in una situazione di macinatura e dose ottimale, una pressione elevata produce nel 90% dei casi un flusso leggermente elevato. Ciò dimostra come una pressione elevata aumenti il flusso ma non abbastanza da superare le soglie imposte, ovvero non in maniera eccessiva. La regola 2 è molto importante: con tutte le variabili esterne ottimali, il flusso rientra nella fascia tra 1 e 1,27 ml/s. Quindi l'analisi delle erogazioni appartenenti a questa fascia non discerne completamente le erogazioni ottimali da quelle leggermente veloci, in quanto il range di valori comprende buona parte delle erogazioni ottimali. Le regole 3 e 4 sono quindi da considerare come regole di compensazione, in quanto nell'82% dei casi, dose elevata e macinatura grossolana portano il flusso nella fascia dei valori 'ok\_alto', che come scoperto nella regola precedente è la fascia dei valori ottimali. La regola 5 infatti descrive come una pressione alta combinata con una dose ottimale, producono erogazioni ottimali solo nel 44% dei casi: ciò è comprensibile perchè non si tiene in considerazione la macinatura, che in caso sia grossolana porta sicuramente fuori soglia il flusso.

Per quanto riguarda il flusso nella fascia 0,77-1, che corrisponde quindi alla fascia di flusso leggermente lento, si nota come le confidenze siano decisamente più elevati del caso precedente, anche in presenza di un supporto più elevato. Con la regola 6 si precisa come, con una macinatura ottimale, pressione bassa e dose ridotta il flusso sia leggermente al

di sotto dell'ottimale nel 95% dei casi. Quindi una dose ridotta non è compensata appieno dalla riduzione della pressione. Questa è una buona notizia per la Lavazza, poiché così può individuare possibili sgrammature in presenza di una pressione ridotta e valori del flusso in questa fascia di valori, come confermato dalla regola 7 che considera il caso in cui non si conosca il valore di macinatura. La regola 8 indica come la macinatura fine assieme a una dose ottimale non porti sempre sotto soglia il flusso, ma esso viene comunque leggermente rallentato. Infine nel 50% dei casi in cui si presenta una pressione bassa, l'erogazione non è eccessivamente lenta, ma si presenta comunque più lenta dell'ottimale.

Con le regole di origine si cerca di capire quali siano le cause dell'assegnazione delle tue etichette in analisi. Si nota subito, analizzando le regole 10-13 che il flusso nel range ok-alto è dovuto maggiormente alle condizioni ottimali di pressione e macinatura, mentre è meno dovuto a pressioni elevate e dosi basse, in quanto esse portano un flusso molto veloce, superiore alle soglie impostate. Al contrario un flusso nel range 0,77-1 ml/s è dovuto maggiormente a una pressione bassa (confidenza 50%) e a una macinatura fine.

In conclusione, la suddivisione in ulteriori due fasce delle erogazioni erogazioni con flusso ottimale non si rivela così utile poiché la fascia di interesse 'ok\_alto' corrisponde in realtà alla fascia dei valori ottimali. L'analisi delle erogazioni leggermente più lente non è così interessante se non per il fatto che una dose ridotta di macinato, in presenza di una pressione bassa, cada all'interno di questa fascia di valori.

### Impatto dei lavaggi sul tempo di erogazione

Si analizza comunque il comportamento dei due gruppi con queste soglie, in modo da evidenziare eventuali differenze nel comportamento all'interno della fascia dei valori considerati ottimali. La tabella 6.20 contiene alcune regole interessanti, con i valori di confidenza, lift e supporto assunti nei due differenti gruppi.

Id	Corpo	Testa	Non pulito			Pulito		
			Conf (%)	Lift	Supp (%)	Conf (%)	Lift	Supp (%)
1	pressione='alta', macinatura='opt'	flusso='ok_alto'	67	3.16	7	22	0.99	2
2	pressione='alta', dose='opt'	flusso='ok_alto'	50	2.37	6	16	0.84	2
3	pressione='opt', macinatura='fine'	flusso='ok_basso'	32	0.97	4	52	1.42	6

Tabella 6.20: Regole di associazione per il flusso con soglie a 3 livelli. Studio dell'influenza dei lavaggi.

Si può notare grazie alle due regole che, nel gruppo sporco, una pressione alta in condizione di macinatura o dose ottimale produce una erogazione con flusso nella fascia 1-1.23 ml/s nel 667% dei casi, mentre nel gruppo pulito ciò non si verifica praticamente mai in quanto la regola presenta lift circa uguale a 1, ovvero indica che non c'è relazione tra causa e effetto. Questo è un comportamento atteso in quanto il gruppo sporco attenua l'effetto della pressione alta, permettendo così che l'erogazione presenti comunque un flusso ottimale. Nel gruppo pulito invece la pressione alta aumenta considerevolmente il flusso, che quindi esce dai valori di soglia. In caso di pressione ottimale e macinatura fine (regola 3), il gruppo pulito presenta un flusso leggermente lento, mentre il gruppo sporco non presenta

un comportamento del genere in quanto i residui nei condotti rallentano ulteriormente il flusso, facendo ricadere l'erogazione nella categoria flusso='basso'.

L'analisi conferma quindi quanto ottenuto nello studio del flusso con le soglie a due livelli, ovvero che il gruppo sporco rallenti in generale l'erogazione.

### 6.4.8 Conclusioni

In conclusione si afferma che le variabili esterne hanno una certa influenza sulla qualità del caffè. In particolare si scopre come una pressione alta, una macinatura grossolana e una dose di macinato ridotto diminuisca il tempo di erogazione, e di conseguenza aumenti la velocità del flusso d'acqua. Nel modo inverso invece, tempo di erogazione e flusso lenti sono caratterizzati dall'uso di una pressione ridotta, di una macinatura fine che aumenta la superficie di esposizione del panetto all'acqua e l'uso di una dose eccessiva di macinato che aumenta l'ostacolo che il getto d'acqua deve superare.

L'analisi svolta considera delle variazioni estreme dei parametri esterni: le soglie di tempo e flusso utilizzate producono risultati coerenti e sono quindi ritenute adatte all'analisi.

Non è stato possibile al momento svolgere un'analisi con delle variazioni lievi dei parametri poiché gli esperimenti effettuati presentano un erogato in tazza e un tempo troppo elevati rispetto agli standard. In una futura analisi di queste variazioni ridotte si suppone che anche le variazioni di flusso e tempo saranno meno evidenti. Con l'utilizzo di queste soglie potrebbe verificarsi il caso che tutte le erogazioni siano considerate ottimali. Gli approcci da seguire possono essere due:

- restringere il range delle soglie così da poter valutare le variazioni più efficacemente;
- utilizzare tre livelli di soglie che nell'analisi effettuata non forniscono regole molto interessanti poiché variazioni estreme dei parametri esterni portano a variazioni estreme delle caratteristiche del caffè, che non sono individuabili nella fascia centrale dei valori assunti dal flusso, bensì in quella più esterna.

Dall'analisi effettuata è emerso come vari parametri si compensino tra loro per produrre comunque un caffè ottimale. In particolare sono interessanti alcune regole che dimostrano come dose e macinatura, due parametri non rilevabili dalla Lavazza nei suoi bar, se programmate correttamente possano compensarsi a vicenda. Anche la pressione presenta un effetto compensativo, però risulta meno interessante poiché monitorata dalla telemetria della Lavazza, e quindi si può individuare immediatamente una sua variazione.

È inoltre individuabile una differenza tra gruppi che effettuano lavaggi e quelli su cui non viene effettuata pulizia. In particolare si nota come l'effetto della pressione viene limitato dal degrado accumulato, e come l'influenza della dose aumenti parecchio. Ciò potrebbe essere applicabile sul mercato: individuati tramite telemetria clienti che non effettuano lavaggi, la presenza di erogazioni veloci potrebbe essere un indice di riduzione della dose. In questo modo si può effettuare una analisi approfondita e intervenire.

Dato l'interesse a individuare caffè veloci poiché recano danno maggiore all'azienda in quanto permettono l'erogazione di un numero maggiore di caffè in un tempo ridotto, compromettendo però il gusto del prodotto in tazza e che potrebbero essere causati dall'utilizzo di una dose ridotta di macinato, che riduce il guadagno della Lavazza, si è proposta una sessione di test futuri che si concentra appunto sulla riduzione della dose e sull'uso di una

macinatura grossolana, analizzando sia una variazione lieve che estrema di esse così da poter effettuare un confronto diretto. Inoltre si è scelto di non analizzare variazioni di pressione poiché poco frequenti sul mercato e individuabili trami telemetria. I dettagli sulla nuova proposta di test sono presenti nella sezione 6.5.5.

## 6.5 Proposta sessione sperimentale

Dopo aver analizzato i dati della prima e della seconda sessione sperimentale e concordato alcuni obiettivi con la Lavazza, è stato progettato un master di test per modellare i comportamenti di interesse su diversi tipi di macchina e con diversi tipi di miscela.

La sessione sperimentale ha come fine modellare in particolare 5 tipi di fenomeni:

- Differenza intrinseca tra i gruppi;
- Relazione tra colpi di ventolino e erogato in tazza;
- Modellazione di lavaggi anomali;
- Rilevazione del degrado dovuto alla mancanza dei lavaggi;
- Influenza dei parametri esterni sulla qualità del caffè;

È importante che i test siano svolti nell'ordine proposto, così da mantenere le stesse condizioni il più a lungo possibile. Questo fattore risulta molto importante in quanto, nelle prime due sessioni sperimentali, è stata notata una grande variazione causata da una ritaratura della macchina e dalla variazione della macinatura del caffè. Con l'ordine indicato si cercano di ridurre al minimo queste variazioni, così da avere le stesse condizioni per l'intera durata dei test. Si procede dunque con la descrizione dei vari test; si faccia riferimento alla sezione **!!! descrizione dataset sessione sperimentale!!!** per l'impostazione del dataset apposito per la raccolta dei dati per la sessione sperimentale proposta.

### 6.5.1 Differenza intrinseca dei gruppi

Uno degli obiettivi principali della Lavazza è capire l'impatto dell'assenza dei lavaggi: secondo le aspettative, dovrebbe presentarsi con un aumento del tempo di erogazione e un rallentamento del flusso rispetto alla condizione iniziale, dove la macchina è pulita. Inoltre il degrado dovuto all'assenza di lavaggi è rilevabile differenziando il comportamento dei due gruppi: nel primo vengono effettuati sia lavaggi giornalieri che purge, mentre nel secondo non viene effettuato alcun tipo di pulizia. È quindi molto importante conoscere eventuali differenze intrinseche tra i due gruppi, così da attribuire un'eventuale differenza nei tempi di erogazione e nel flusso a cause di natura meccanica, e non alla mancanza dei lavaggi.

Il test si struttura in molteplici punti:

- Per prima cosa si effettuano lavaggi ripetuti sui gruppi, in modo da portare la macchina alla condizione ideale.
- Si effettua poi la taratura della macchina come indicato dagli esperti del Training Center, con l'obiettivo di avere, per la dose doppia, 46 grammi di erogato in 25 secondi.

- Si svolgono quindi 60 erogazioni ottimali per gruppo, mantenendo la giusta dose di macinato e la giusta macinatura dei grani. Inoltre, dopo ogni erogazione, vengono svolte le pulizie al fine di mantenere il gruppo pulito.

Per ogni erogazione si raccolgono le rilevazioni del device di telemetria e la misura dell'erogato in tazza, per fornire la relazione tra la taratura della macchina e il caffè erogato.

È necessario che questa operazione si svolga all'inizio di ogni test, e in particolare quando sono modificate macinatura e taratura della macchina, in modo da avere sempre un punto di partenza noto e ideale.

### 6.5.2 Relazione tra colpi di ventolino e erogato in tazza

Durante l'analisi della seconda sessione sperimentale, è stato notato che l'impostazione dei colpi di ventolino durante la taratura della macchina si differenziava notevolmente rispetto a quella della prima sessione sperimentale. In particolare, il numero di colpi di ventolino era molto più elevato, e di conseguenza, anche l'erogato rilevato in tazza risultava molto abbondante, superando le soglie di qualità fornite dagli esperti. Il numero di colpi di ventolino impostato può variare da macchina a macchina, in quanto la quantità di acqua erogata da ventolini diversi potrebbe essere diversa. Inoltre la loro impostazione è effettuata al momento della taratura della macchina, procedura che è però influenzata da diversi fattori, come ad esempio il tipo di miscela, la quantità di macinato utilizzato e le condizioni del macinato.

Si effettuano alcune erogazioni variando i colpi di ventolino, con l'obiettivo di valutare la variazione dell'erogato in tazza e di calcolare la funzione di trasferimento da dato di telemetria a valore in tazza nelle diverse programmazioni. In particolare:

- Si svolgono 10 erogazioni ottimali, con taratura della macchina secondo istruzioni degli esperti (46gr di erogato in 25 secondi, per dose doppia);
- Si svolgono altre 9 sessioni, ognuna composta da 10 erogazioni, mantenendo la macinatura ottimale ma aggiungendo oppure diminuendo un numero di colpi di ventolino costante (5-10, a seconda della quantità erogata da ogni colpo) rispetto a quello ottimale, calcolato durante la taratura.

I 10 test sono svolti su entrambi i gruppi, per un totale di 200 erogazioni. In alcuni casi, potrebbe essere accettabile ridurre il numero di test basandosi sul range di colpi di ventolino accettabili.

### 6.5.3 Modellazione dei lavaggi anomali

Esistono due tipi principali di pulizia per una macchina del caffè: i lavaggi lunghi e i lavaggi brevi. I primi sono effettuati giornalmente mentre i secondi, detti anche pulizie, sono svolti dopo ogni erogazione. Vi sono, essenzialmente, tre modalità per questi lavaggi:

- Senza il gruppo inserito: l'acqua scorre nelle tubature senza ostacoli;
- Con coppetta cieca: l'acqua erogata non trova uscita nell'erogatore, e pulisce più efficacemente i condotti;

- Con coppetta ceca e pastiglia: oltre alla coppetta ceca, si utilizza una pastiglia che garantisce una pulizia migliore del gruppo.

Sebbene sia previsto un programma apposito per i lavaggi, un numero considerevole di baristi usano i tasti per la dose doppia e per la dose libera per la pulizia della macchina. In aggiunta, spesso, si sceglie di utilizzare delle pastiglie non rilasciate dall'azienda, spesso di costo inferiore, che potrebbero compromettere l'utilizzo della macchina.

Sono previste 10 erogazioni per ognuna delle combinazioni di questi lavaggi anomali, per un totale di 6 test (brevi e lunghi, nei tre casi sopra specificati). Inoltre, prima di essi si svolgono 60 erogazioni ottimali senza effettuare i lavaggi. Ripetendo le erogazioni per entrambi i gruppi, ne risulta un totale di 240 erogazioni.

L'obiettivo è, in particolare, scoprire le modalità secondo cui vengono rilevati i lavaggi anomali dalla telemetria, al fine di riconoscerli analizzando i dati provenienti dal mercato e correggere il comportamento errato del cliente.

#### 6.5.4 Degrado dovuto alla mancanza di lavaggi

La Lavazza si pone come obiettivo identificare i clienti che non effettuano la giusta manutenzione della macchina, accorciando o compromettendo la vita della stessa, con il conseguente aumento delle richieste di intervento tecnici a spese dell'azienda.

Si imposta un test che cerca di replicare un cliente che non effettua i lavaggi, cercando di rilevare un degrado nelle prestazioni con l'aumentare del numero di erogazioni. Inoltre, si è posta l'attenzione sull'effetto che un lavaggio presenta nell'azzerare il degrado accumulato, e quali risultati si ottengano con lavaggi successivi.

Inoltre si è anche interessati all'effetto che un lavaggio ha nell'azzerare il degrado accumulato, e quali risultati si ottengano con lavaggi successivi.

Una delle problematiche che emerge dal test è l'elevato numero di erogazioni necessarie: un bar con un buon ritmo di lavoro produce in una giornata circa 500 caffè. Il numero minimo di erogazioni richieste per rilevare un degrado è quindi stimato in 500, anche se il numero più appropriato si aggira sulle 2000 erogazioni per gruppo, così da notarsi l'accumulo del degrado di più giorni. Per velocizzare le operazioni di rilevazione dei dati, l'erogato in tazza è misurato solo all'inizio e dopo un certo numero di erogazioni, così da fornire informazioni su eventuali cambi di taratura indesiderati. Si effettuano quindi, per ogni gruppo, almeno 500 erogazioni successive senza lavaggi e purge, dopo aver effettuato 60 erogazioni ottimali con la macchina allo stato ideale, che fungono da riferimento per l'analisi. Se dopo 500 erogazioni non si è registrato un degrado, si prosegue con le erogazione senza lavaggi fin quando esso è rilevato. A questo punto il comportamento dei due gruppi varia, così da poter modellare due comportamenti diversi e effettuare il confronto:

- in un gruppo si effettua un lavaggio macchina e si continua con le erogazioni senza lavaggi, così da vedere l'effetto del lavaggio e se quest'ultimo riporta lo stato della macchina a quello iniziale. Si può procedere con lavaggi ulteriori per vedere il loro effetto e dopo che numero la condizione ritorna a quella iniziale;
- Nell'altro gruppo invece non viene effettuato un lavaggio, ma si inizia a effettuare le erogazioni con le purge, così da vedere se i lavaggi brevi dopo ogni erogazione portano un miglioramento.

### 6.5.5 Influenza delle variabili esterne

Ulteriore obiettivo della Lavazza è garantire una buona qualità del caffè: alcuni clienti, però, riducono la quantità di macinato utilizzato così da ridurre la quantità di caffè da comprare. È una pratica comune utilizzare una macinatura più “grossolana” in modo tale che il tempo di erogazione diminuisca e che si effettuano quindi più caffè in meno tempo). Le prime due sessioni sperimentali miravano a individuare l’effetto di una variazione, lieve o estrema, di dose di macinato, macinatura dei grani e pressione, sull’erogazione del caffè. Attraverso il confronto con gli esperti è stato possibile variare la pressione della pompa non è banale, e un numero limitato di clienti la modificano. Inoltre, la telemetria rileva la pressione della pompa, ed è quindi subito individuabile una sua variazione.

Si è allora deciso di non prendere in considerazione la pressione, ma di analizzare unicamente la variazione di dose e macinatura. Vengono presi in esame i seguenti casi:

- Dose: -2grammi, -1 grammo, ottimale e + 1 grammo, dove ottimale è il valore di 14 grammi previsto dagli esperti.
- Macinatura: +2 tacche, +1 tacca, ottimale e -1 tacca, dove ottimale è la macinatura impostata in taratura. Un aumento delle tacche corrisponde a una macinatura più grossolana, mentre una riduzione indica una macinatura più fine.

In particolare maggiore importanza è attribuita alla riduzione del macinato e alla macinatura grossolana in quanto sono atteggiamenti più comuni sul mercato, che arrecano un danno sia economico che di immagine alla Lavazza.

Inoltre, per questo studio, si è differenziato il comportamento dei due gruppi: infatti mentre su uno si effettuano sempre le purge e i lavaggi, l’altro opera in mancanza di manutenzione.

Il test si articola in più parti:

- 60 erogazioni ottimali in presenza di lavaggi in entrambi i gruppi, effettuati dopo aver garantito il loro stato ottimale. Essi rappresentano i nostri caffè di ottima qualità e servono come riferimento.
- 15 test composti di 20 erogazioni ciascuno, dove vengono testate tutte le combinazioni di dose e macinature descritte precedentemente, mantenendo il più a lungo possibile costante la macinatura;
- 20 erogazioni ottimali ogni volta che viene cambiata la macinatura dei grani di caffè, così da avere un confronto con la situazione iniziale e confrontare un eventuale degrado tra i due gruppi.

Si ottiene un totale di 20 test, con 440 erogazioni per gruppo. Per chiarezza, nell’immagine successiva è riportata la sequenza dei vari test.

## 6.6 Analisi dei dati provenienti dal mercato

Una volta che il device di telemetria è stato installato su un certo numero di macchine, si è deciso di analizzare alcuni dati reali, in modo da capire se fosse possibile estendere le analisi

TEST	NUMERO DI EROGAZIONI	ID COINVOLTI	Macinatura	Dose
T_01	20	1-20	OTTIMALE	OTTIMALE
T_01	20	21-40	OTTIMALE	OTTIMALE
T_01	20	41-60	OTTIMALE	OTTIMALE
T_02	20	61-80	OTTIMALE	+1
T_03	20	81-100	OTTIMALE	-1
T_04	20	101-120	OTTIMALE	-2
T_05	20	121-140	ottimale	ottimale
T_06	20	141-160	+2	OTTIMALE
T_07	20	141-160	+2	+1
T_08	20	161-180	+2	-1
T_09	20	181 - 200	+2	-2
T_10	20	201 - 220	ottimale	ottimale
T_11	20	221 - 240	+1	OTTIMALE
T_12	20	221 - 240	+1	+1
T_13	20	241 - 260	+1	-1
T_14	20	261-280	+1	-2
T_15	20	281-300	ottimale	ottimale
T_16	20	301-320	-1	OTTIMALE
T_17	20	321-340	-1	+1
T_18	20	341-360	-1	-1
T_19	20	341-360	-1	-2
T_20	20	361-380	ottimale	ottimale

Figura 6.12: Tracciato del test sull'influenza delle variabili esterne

effettuate in laboratorio al caso reale. In questo modo abbiamo potuto familiarizzare con i dati provenienti dal mercato, così da capire come analizzarli.

Il dispositivo di telemetria rileva numerose informazioni sullo stato della macchina e sulle erogazioni effettuate. Tra questi, ci focalizziamo in particolare su tempo di erogazione e flusso. A partire da essi viene calcolata anche la quantità di acqua erogata, ottenuta come prodotto tra le due grandezze rilevate. I dati sono reperiti tramite il portale della telemetria, dove, selezionando un certo cliente, si può scegliere un periodo di massimo 41 giorni da cui estrarre i dati rilevati. Come descritto nella sezione 5.2.2, le rilevazioni, salvate in un file excel, presentano inizialmente alcuni problemi di allineamento, risolti in parte con i successivi aggiornamenti del firmware del device: per velocizzare l'operazione di pulizia si utilizza uno script che, dato l'excel ottenuto dal portale, restituisce un file avente tempo e flusso allineati sulla base dell'orario di rilevazione e del gruppo su cui è stata svolta l'erogazione.

I dati provenienti dal mercato presentano una quantità di informazione decisamente ridotta rispetto a quelli ottenuti dagli esperimenti in laboratorio. In particolare non possiamo conoscere:

- Informazioni esterne alla macchina: non si conosce la miscela utilizzata, la macinatura dei grani e la dose di macinato utilizzato per l'erogazione;

- Tasto selezionato: ogni erogazione non porta l'informazione sul tipo (singola, doppia o continua), e sulla dose programmata (corto, normale e lungo).

In particolare, la mancanza dell'informazione del tipo di erogazione collegata alle misurazioni rappresenta un grosso limite, in quanto le analisi sono compromesse da dati eterogenei, e quindi non confrontabili. Per ovviare a questo problema, si è cercato inizialmente di prendere in analisi solo periodi in cui le vendite riportano per la maggior parte caffè doppi, scelti poichè la dose doppia è il tipo di erogazione sui cui si basano i test delle prime due sessioni sperimentali in laboratorio. In questo modo si cerca di garantire l'omogeneità dei dati e quindi risultati dell'analisi più attendibili. Per limitare ulteriormente l'approssimazione dovuta ai diversi tipi di erogazioni presenti, si effettua un'ulteriore pulizia basandosi sul valore dell'acqua erogata, come descritto nella sezione 5.2.3 (**Pulizia basata sull'erogato**).

È importante osservare come tutte le analisi dei dati in produzione prenderanno quindi come riferimento le erogazioni doppie, che producono due caffè in tazza contemporaneamente. Lo studio è però interessato al singolo caffè in tazza, a cui si riferiscono le soglie di qualità: mentre il tempo di erogazione rimane costante sia per l'erogazione doppia che per quella singola, il flusso rilevato per la dose doppia risulterà il doppio di quello rilevato per la dose singola. Si è deciso allora di modificare il flusso proveniente dalla telemetria così da riportarlo alla singola erogazione, dimezzando sempre il flusso, e di conseguenza, anche l'acqua erogata calcolata a partire da esso.

Vengono affrontate in particolare due tematiche: in primis si cerca di modellare alcuni clienti ritenuti fidati dall'azienda, con l'obiettivo di avere un riferimento per i dati in telemetria e per stabilire delle soglie di qualità su di essi.

Successivamente si affronta il discorso degli interventi tecnici: vengono analizzati alcuni periodi in cui sono effettuati degli interventi tecnici, con l'obiettivo di confrontare le caratteristiche delle erogazioni prima e dopo di esso.

### 6.6.1 Modellazione di clienti fidati

Passando per la prima volta all'analisi dei dati provenienti dal mercato, abbiamo analizzato dei periodi di clienti suggeriti dall'azienda poichè ritenuti fidati, ovvero che seguono tutte le buone pratiche necessarie per ottenere un caffè eccellente. A partire da questi clienti, abbiamo confrontato alcuni periodi di rilevazioni sulla stessa macchina, su macchine dello stesso modello e su macchine di modello diverso. In particolare sono due i clienti migliori: la caffetteria della Nuvola, con sede a Torino, e il Top Shop, con sede a Milano.

Entrambe le caffetterie hanno in dotazione il modello Faema E71: purtroppo non abbiamo avuto modo di testare in laboratorio il funzionamento di questo modello di macchina, e quindi non abbiamo potuto ricavare la relazione tra caratteristiche dell'erogazione in tazza e in telemetria.

Per ovviare a questo problema, il barista della caffetteria della Nuvola si offerto disponibile per svolgere 20 erogazioni di cui ha rilevato il peso dell'erogato in tazza, rendendo possibile il calcolo della funzione di trasferimento che lega tazza a telemetria attraverso la regressione, come spiegato nella sezione 6.2. In questo modo è stato trovato che la relazione tra tazza e telemetria sulla Faema F72 è indicata dalle seguenti formule:

- Dato  $F_{tazza}$  misurazione del flusso in tazza, e indicando con  $F_{tel}$  il flusso rilevato dalla telemetria

$$F_{tel} = 1,65F_{tazza} - 0,15 \quad (6.6)$$

- Dato  $F_{tel}$  rilevazione del flusso dalla telemetria, e indicando con  $F_{tazza}$  il flusso in tazza

$$F_{tazza} = 0,57F_{tel} + 0,12 \quad (6.7)$$

Per valutare la qualità delle erogazioni rilevate dalla telemetria, si ricavano le soglie a partire da quelle in tazza utilizzando le funzioni di trasferimento appena calcolate: in particolare si utilizzano le soglie di tempo e flusso forniteci dagli esperti, e le soglie proposte per il flusso partendo dai dati della prima sessione sperimentale, come descritto nelle sezione 6.3, che chiameremo rispettivamente *Soglie flusso originali* e *Soglie flusso proposte*.

La tabella 6.21 riassume le soglie in tazza e in telemetria utilizzate.

	In tazza	Telemetria
Soglie di tempo	20 - 27 s	20 - 27 s
Soglie di flusso originali	0,49 - 1,33 ml/s	0,65 - 2,04 ml/s
Soglie di flusso proposte	0,77 - 1,23 ml/s	1,17 - 1,86 ml/s

Tabella 6.21: Soglie in tazza e in telemetria

In aggiunta, è stata calcolata anche la funzione di trasferimento da quantità di acqua erogata secondo la telemetria a erogato in tazza: la seguente relazione non è così attendibile, in quanto la misurazione dell'erogato in tazza è molto più variabile di quello rilevato dalla telemetria, poichè il peso in tazza varia a seconda dell'istante in cui avviene la misurazione, a causa della presenza iniziale della schiuma sopra il caffè, che col passare del tempo diventa liquida. Essa però ci permette di avere un'idea approssimativa di quanti millilitri di acqua sono riassorbiti dalla macchina durante il processo di erogazione, e quindi poter avere una stima di quale sia la quantità di erogato in tazza rispetto a quella erogata complessivamente dalla macchina.

- Dato  $Q_{tel}$  quantità di acqua erogata, calcolata come tempo di erogazione per flusso rilevato dalla telemetria, e indicando con  $Q_{tazza}$  la quantità in tazza

$$Q_{tazza} = 0,78Q_{tel} - 1,83 \quad (6.8)$$

L'analisi di ogni periodo avviene attraverso una serie di passi:

- Descrizione del periodo scelto, in cui avviene il confronto tra numero di rilevazioni della telemetria e conteggio delle vendite, in modo da verificare la loro corretta corrispondenza e il tipo di erogazioni effettuate in quel periodo;
- Pulizia dei dati, con l'obiettivo di eliminare erogazioni accorpate e erogazioni corte e lunghe, come descritto nella sezione 5.2.3. Durante l'analisi dell'acqua erogata è anche possibile riconoscere il tipo di erogazione effettuata su ogni gruppo, se singola o doppia, come illustrato nell'introduzione di questa sessione;

- Assegnazione delle etichette di qualità da assegnare alle erogazioni sulla base delle soglie indicate nella tabella 6.21, e analisi generale di esse;
- Calcolo delle statistiche (media e deviazione standard) delle caratteristiche principali (tempo di erogazione, flusso e quantità erogata) di ogni gruppo macchina nel periodo selezionato;
- A partire dalle statistiche calcolate, vengono ricavate delle soglie di qualità riferite al dato di telemetria: esse vengono calcolate in particolare per i clienti fidati (Nuvola e Top Shop), per essere usate come termine di paragone con le erogazioni degli altri clienti;
- Trasformazione delle soglie appena ricavate in telemetria in soglie riferite al dato in tazza, così da effettuare un paragone con le soglie forniteci dagli esperti;
- Osservazioni e conclusioni sulle analisi sopra citati.

Sono stati scelti quattro periodi: in particolare, due periodi sulla Faema E71 installata presso la Nuvola Lavazza di Torino, uno sulla Faema E71 presso il Top Shop di Milano e, infine, un periodo sulla Wega My Concept presso il bar del Gruppo Spes SCS Etiko. La scelta non è casuale: infatti, è possibile effettuare un confronto tra periodi diversi della stessa macchina, tra dati di macchine diverse ma dello stesso modello e, per finire, tra macchine di costruttori diversi. Inoltre è giusto ricordare che mentre i primi due clienti (Nuvola e Top Shop) sono clienti fidati, dei quali si è consapevoli della buona qualità del caffè erogato, il terzo è un rappresentante qualsiasi dell'insieme dei clienti Lavazza, di cui si vorrebbe monitorare il comportamento.

Nella tabella 6.22, sono elencati i dettagli di ogni periodo: si noti che la scelta del periodo è stata effettuata sulla base della percentuale di erogazioni normali (doppie e singole), così da avere dati il più possibile omogenei. In particolare si è data preferenza a periodi in cui le vendite riportano una maggiore percentuale di erogazioni normali doppie.

Cliente	Periodo selezionato	Modello di macchina	Percentuale erogazioni normali
Nuvola Lavazza	27/06/2019 – 11/07/2019	Faema E71	91%
Nuvola Lavazza	17/05/2019 – 30/05/2019	Faema E71	85%
Top Shop	08/07/2019 – 27/07/2019	Faema E71	88%
Gruppo Spes SCS Etiko	06/06/2019 – 16/06/2019	Wega My Concept	92%

Tabella 6.22: Dettagli dei periodi analizzati

Si procede allora con l'analisi dei vari periodi: si tenga presente che l'analisi effettuata non tiene in considerazione le diverse miscele utilizzate e possibili differenze nella macinatura del caffè.

### Nuvola Lavazza - periodo 27/06/2019–11/07/2019

Il primo periodo analizzato è sul miglior cliente possibile, la Nuvola della Lavazza. Infatti, la caffetteria della Nuvola effettua lavaggi giornalieri multipli e utilizza dosi addirittura abbondanti di macinato. Esso sarà il nostro vero punto di riferimento durante tutta l'analisi.

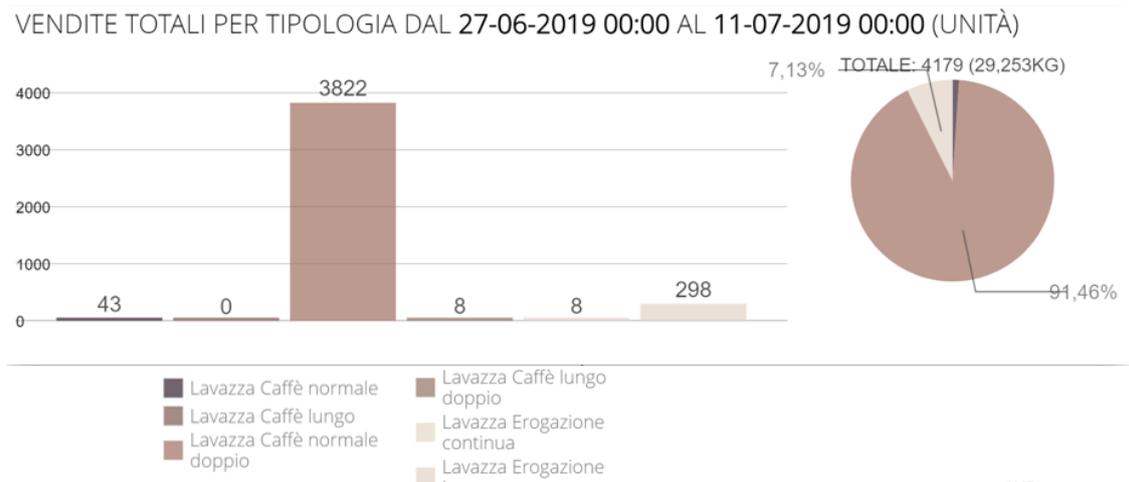


Figura 6.13: Dati di vendita della Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.

Grazie ai dati delle vendite, riportati in figura 6.13, scopriamo che, nel periodo analizzato, sono stati effettuati:

- 3830 caffè doppi, corrispondenti a 1915 erogazioni, poichè ogni erogazione doppia corrisponde a due caffè venduti;
- 349 caffè di altro tipo(singoli, erogazioni continue e leva);
- **Numero totale di erogazioni: 2264.**

La telemetria invece rileva un totale di 1471 erogazioni: si nota subito che i dati non corrispondono, e vengono rilevate all'incirca solo i 2/3 delle erogazioni attese. La causa di questa discrepanza è dovuta probabilmente al fatto che la Faema E71 è una macchina prodotta dal Gruppo Cimbali, e presenta quindi un device di telemetria del costruttore che elabora i dati rilevati, scartando le erogazioni che non rispettano certe soglie impostate durante la prima taratura della macchina. Procediamo comunque con l'analisi, consci dell'assenza di un buon numero di erogazioni.

La tabella 6.23 riassume i risultati del processo di pulizia, indicando sia le soglie scelte manualmente per filtrare le erogazioni, sia il numero di rilevazioni prima e dopo di essa. Inoltre si prova a dare anche un'interpretazione dell'utilizzo del gruppo.

	Soglie quantità acqua (ml)	Erogazioni totali	Erogazioni rimanenti	Utilizzo del gruppo
Gruppo 1	17 - 30	297	290	caffè doppio
Gruppo 2	8 - 27	519	498	caffè singolo
Gruppo 3	18 - 30	659	621	caffè doppio

Tabella 6.23: Dettagli della pulizia dei dati per gruppo nella Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.

Il gruppo due è stato identificato come gruppo singolo in quanto è caratterizzato da un'acqua erogata minore rispetto agli altri gruppi: analizzando le distribuzioni degli erogati del gruppo 2 e 3, notiamo infatti che il gruppo 2 presenta la maggior parte delle erogazioni con un acqua erogata all'incirca dimezzata rispetto a quella del gruppo 1. Inoltre si nota sempre che il gruppo singolo presenta un erogato più variabile, e di conseguenza delle soglie di pulizia più larghe, poichè probabilmente più usato per svolgere caffè con dosi corte e lunghe, in quanto la probabilità che sia uno solo il cliente a chiedere una dose diversa da quella standard è più alta rispetto a richieste multiple per la stessa dose. Occorre notare come i dati di vendita non riportano erogazioni singole, nonostante si abbia avuto la conferma diretta da parte del barista che il gruppo due effettua solo erogazioni di questo tipo.

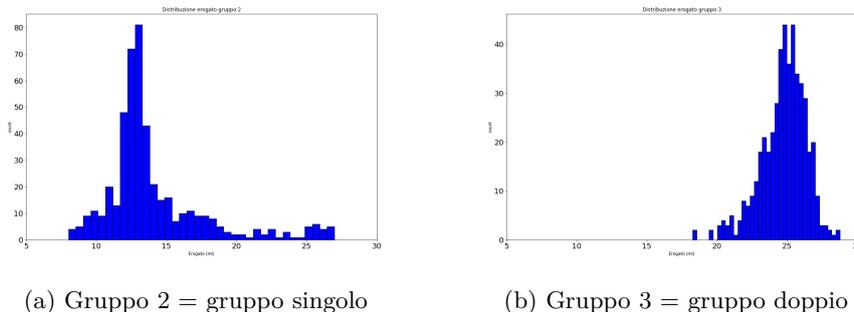


Figura 6.14: Differenza tra le distribuzioni dell'acqua erogata dei gruppi 2 e 3

Dopo aver moltiplicato per due il flusso di tutte le rilevazioni del gruppo 2, così da renderle confrontabili con le erogazioni doppie, si procede ad assegnare le etichette di qualità ai caffè secondo le soglie di telemetria indicate in tabella 6.21: la tabella 6.24 contiene il conteggio delle varie etichette considerando tre casi: andamento generale della macchina, andamento dei gruppi doppi e andamento del gruppo singolo.

Soglie	Complessivo			Gruppi doppi			Gruppo singolo		
	<	OK	>	<	OK	>	<	OK	>
Soglie di tempo	113	882	405	60	643	204	53	239	201
Soglie di flusso originali	44	1348	8	42	865	0	2	483	8
Soglie di flusso proposte	851	524	25	744	163	0	107	361	25

Tabella 6.24: Conteggio delle etichette di qualità per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.

Si può notare subito che le soglie di flusso originali non differenziano il comportamento dell'erogazioni: il 95% delle erogazioni risulta in soglia, che potrebbe essere in linea con il fatto che la Nuvola produce dei caffè ottimi, anche se è da tenere in conto che le soglie proposte dagli esperti sono molto larghe e potrebbero accettare come buoni in particolare caffè con flusso molto lento. Le soglie di flusso proposte, invece, evidenziano una grande presenza di flussi sotto soglia, in particolare nei gruppi doppi, dove solo il 18% dei caffè risulta in soglia. Nel gruppo singolo invece la maggioranza delle erogazioni rientra nei limiti impostati a partire dalle analisi di laboratorio. Prendendo in considerazione i dati del tempo di erogazione, vediamo che la maggior parte dei caffè sta nelle soglie, anche se il 30% circa di essi presenta un tempo maggiore della soglia. Questa grande presenza di caffè con tempi di erogazione alti e flussi molto ridotti potrebbe essere riconducibile al fatto che la caffetteria della Lavazza usa una dose di macinato superiore alla media, in modo da rendere l'aroma del caffè più gustoso, e portando quindi a una riduzione della velocità del flusso. Inoltre si ricorda che le soglie proposte si riferiscono all'utilizzo della miscela Top Class, diversa da quelle utilizzate nella caffetteria, che necessitano di base una dose di macinato maggiore.

Si analizza ora la distribuzione dei dati grazie ai boxplot di figura 6.15, e calcoliamo media e deviazione standard in modo da comprendere meglio l'andamento delle tre caratteristiche principali, mostrati nella tabella 6.25.

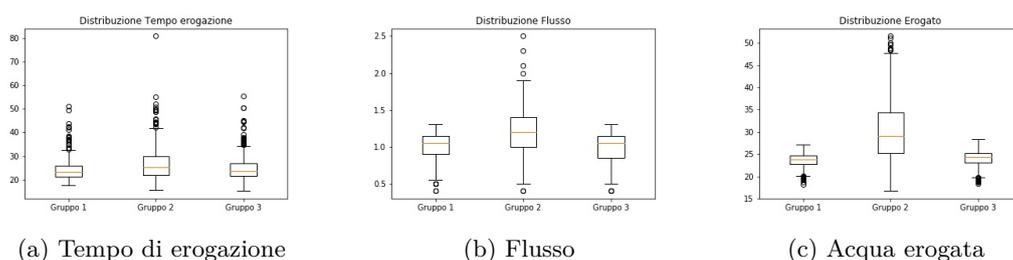


Figura 6.15: Distribuzione delle tre variabili principali per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.

	Flusso (ml/s)		Tempo di erogazione (s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Gruppo 1	1,00	0,20	24,58	4,98	23,56	1,63
Gruppo 2	1,16	0,30	26,83	7,01	30,04	7,20
Gruppo 3	1,00	0,19	24,95	4,81	24,07	1,77

Tabella 6.25: Calcolo statistiche per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.

È subito chiaro come il gruppo 2, che eroga singole dosi, abbia un comportamento diverso dai due gruppi doppi: infatti, l'acqua erogata ha una variazione molto più ampia, rappresentativa probabilmente del fatto che il gruppo singolo è usato per svolgere erogazioni libere. Di conseguenza, anche i tempi di erogazione sono leggermente maggiori e più variabili. Per quanto riguarda il flusso, anch'esso è leggermente maggiore di quello dei gruppi doppi, probabilmente poichè la dose di macinato utilizzato per la dose singola è la metà di quella della dose doppia, e quindi la resistenza del panetto all'acqua risulta minore.

Analizzando in particolare la tabella 6.25, possiamo notare come la media del flusso e del tempo di erogazione dei gruppi doppi concordi perfettamente con le regole forniteci dagli esperti, che prevedono un tempo di erogazione di 24-25 secondi con un flusso di 1 ml/s, in modo da avere un erogato in tazza di circa 24-25 ml. Si viene però a creare una discrepanza: infatti, le informazioni forniteci dagli esperti sono riferite alle misurazioni in tazza, mentre le statistiche sono calcolate su dati di telemetria. Appurato che i gruppi 1 e 3 sono gruppi doppi caratterizzati dallo stesso andamento, sono calcolate le soglie di qualità in telemetria come  $media \pm dev.std.$ , utilizzando il gruppo 1 come riferimento. Inoltre, utilizzando le funzioni di trasferimento per la Faema E71, calcolate ai punti (6.7) e (6.8), si ricavano le soglie in tazza.

	Flusso (ml/s)		Tempo di erogazione (s)		Acqua erogata (ml)	
	inf	sup	inf	sup	inf	sup
Telemetria	0,80	1,2	20	30	22	25
In tazza	0,57	0,80	20	30	15	17

Tabella 6.26: Calcolo soglie per la Nuvola Lavazza nel periodo 27/06/2019–11/07/2019.

Effettuando un confronto con le soglie in tazza ricavate dalla prima sessione sperimentale, emerge che, nonostante, le soglie del tempo di erogazione siano paragonabili, con una leggera tendenza a tempi maggiori in produzione, l'erogato in tazza e il flusso sono decisamente inferiori.

L'erogato minore è sicuramente dovuto alla procedura di taratura: probabilmente il barista imposta il tempo di erogazione correttamente, usando una macinatura più fine del caffè e preferendo una quantità minore di caffè in tazza, così da avere un gusto più concentrato. Infatti, la Nuvola è da considerare come un cliente unico nel suo genere, dove la qualità del caffè ha la priorità su tutto, e quindi viene utilizzata una dose maggiore del necessario e probabilmente una macinatura più fine. In questo modo, l'acqua che riesce ad

attraversare il panetto nel tempo ottimale(24-25 secondi) è minore in quanto l'ostacolo da superare è maggiore, e di conseguenza anche il flusso è rallentato.

Arrivati alla fine dell'analisi del primo periodo possiamo quindi concludere che:

- È possibile riconoscere se un gruppo è utilizzato con erogatore singolo o doppio osservando la quantità di acqua erogata nelle varie erogazioni.
- In genere, un gruppo singolo presenta un erogato più variabile, perchè più usato per svolgere erogazioni continue;
- La Nuvola della Lavazza è un cliente molto particolare, poichè per ottenere un caffè di qualità eccellente utilizza dosi di macinato superiori allo standard, cosa che avviene raramente tra i comuni clienti della Lavazza;
- A causa della dose elevata di macinato, si ottengono delle soglie di flusso molto basse.

L'analisi continua con il confronto di due periodi sulla stessa macchina, così da validare le conclusioni appena elencate e verificare se, in periodi diversi, il comportamento del cliente varia.

### **Confronto tra periodi diversi sulla stessa macchina**

È stato preso in analisi un nuovo periodo sulla stessa macchina analizzata in precedenza, ovvero la Faema E71 della Nuvola Lavazza. In particolare è stato selezionato il periodo che va dal 17/05/2019 al 30/05/2019, dove sappiamo essere state effettuati, grazie ai dati di vendita:

- 3534 caffè doppi, corrispondenti a 1767 erogazioni rilevate in telemetria;
- 702 caffè di altro tipo(singoli, erogazioni continue e leva).
- **Numero totale di erogazioni:** 2470.

Anche in questo caso è stata trovata una discrepanza tra numero di caffè venduti e erogazioni rilevate: infatti sono presenti solo 1384 rilevazioni, ovvero poco più della metà di quelle attese. Ciò conferma la mancanza di rilevazioni da parte della telemetria di alcune erogazioni a causa della telemetria del costruttore Cimbali che filtra i dati prima del device della Lavazza.

La tabella 6.27 contiene i dati relativi all'operazione di pulizia dei dati basata sulla quantità di acqua erogata. Si verifica che anche in questo caso il gruppo 2 presenta un erogato in media dimezzato rispetto a quello dei restanti gruppi, confermandosi gruppo con erogatore singolo.

È importante notare come le soglie utilizzate per la pulizia dei dati siano le stesse del primo periodo analizzato, fatto che mette in risalto una certa continuità tra le caratteristiche delle erogazioni di questa stessa macchina.

A questo punto si procede con l'assegnazione delle etichette di qualità secondo le soglie riportate nella tabella 6.21, e si effettua il conteggio di esse, così da valutare la qualità delle erogazioni.

	Soglie quantità acqua (ml)	Erogazioni totali	Erogazioni rimanenti	Utilizzo del gruppo
Gruppo 1	17 - 30	379	354	caffè doppio
Gruppo 2	8 - 27	483	467	caffè singolo
Gruppo 3	18 - 30	522	483	caffè doppio

Tabella 6.27: Dettagli della pulizia dei dati per gruppo per Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.

Soglie	Complessivo			Gruppi doppi			Gruppo singolo		
	<	OK	>	<	OK	>	<	OK	>
Soglie di tempo	104	812	391	55	564	338	69	248	153
Soglie di flusso originali	7	1292	8	7	831	0	1	461	8
Soglie di flusso proposte	839	447	21	724	113	0	115	334	21

Tabella 6.28: Conteggio delle etichette di qualità per la Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.

Confrontando i numeri del periodo in analisi con i conteggi contenuti nella tabella 6.24, si può notare un andamento praticamente identico tra i due periodi, con molte erogazioni con flussi lenti e tempi alti, a causa dell'elevata dose di macinato utilizzata, mentre il gruppo singolo produce la maggior parte dei caffè in soglia. Inoltre, dal paragone delle statistiche calcolate per i due periodi, e riportate nelle tabelle 6.25 e 6.29 si può notare che i valori medi delle tre caratteristiche praticamente coincidono, con la sola differenza che nel secondo periodo analizzato (maggio), le deviazioni standard sono leggermente più elevate.

Partendo dalle statistiche calcolate si ricavano nuovamente delle soglie di qualità sui dati provenienti dal mercato, utilizzando nuovamente come riferimento il gruppo 1.

	Flusso (ml/s)		Tempo di erogazione (s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Gruppo 1	0,98	0,16	24,90	3,5	23,90	1,20
Gruppo 2	1,14	0,26	25,50	5,6	28,20	3,80
Gruppo 3	0,99	0,16	25,50	4,20	24,70	1,60

Tabella 6.29: Calcolo statistiche per la Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.

	Flusso (ml/s)		Tempo di erogazione (s)		Acqua erogata (ml)	
	inf	sup	inf	sup	inf	sup
Telemetria	0,82	1,14	21	28	23	25
In tazza	0,59	0,77	21	28	16	17

Tabella 6.30: Calcolo statistiche per la Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.

Dal confronto delle soglie ricavate dai due periodi, contenute nelle tabelle 6.26(giugno) e 6.30(maggio), si nota che queste ultime sono più stringenti: in particolare, le soglie nel tempo sono coerenti con quelle degli esperti.

In conclusione, si può dire che i dati ricavati sulla stessa macchina sono coerenti.

Grazie all'analisi di questi due periodi della Nuvola Lavazza, si sono definite le soglie di flusso e tempo di erogazione usate nel portale di telemetria, che permette un'analisi ad alto livello dei dati. In particolare si sono scelte le seguenti soglie:

- Per il flusso si è optato per le soglie meno stringenti calcolate con i dati del mese di giugno: 0,8 - 1,2 ml/s;
- Per il tempo di erogazione invece sono state scelte le soglie più stringenti ricavate dai dati del mese di maggio, in quanto più coerenti con quelle degli esperti: 21 - 28 secondi.

Le soglie proposte sono solo una soluzione iniziale per avere un riferimento sulla piattaforma online: infatti, come citato precedentemente, la Nuvola è un cliente che effettua caffè "troppo perfetti", e che possono essere difficilmente paragonabili a quelli dei normali bar.

Per avere delle soglie più accurate, quando in futuro si avrà la possibilità di selezionare il tipo di erogazione desiderata, si potranno monitorare periodi più lunghi di un insieme di clienti ritenuti fidati e calcolare la media delle caratteristiche in analisi: in questo modo le soglie non saranno affette da approssimazioni dovute alla presenza di erogazioni diverse da quelle doppie e potranno comprendere caratteristiche di miscele diverse, in quanto clienti diversi possono utilizzare miscele diverse. Inoltre potrebbe essere possibile creare delle soglie che si differenziano a seconda del tipo di erogazione in analisi(singola, doppia, lunga, corta).

### Confronto tra macchine diverse ma dello stesso modello

Si procede ora con il confronto tra periodi di macchine diverse, ma dello stesso modello. Si prende in analisi allora il periodo 08/07/2019 – 27/07/2019 del cliente Top Shop di Milano, che ha in dotazione una Faema E71. Il modello è quindi lo stesso della Nuvola Lavazza, con l'unica differenza che presenta solo due gruppi invece di tre. Inoltre, anche il Top Shop, come la Nuvola, è un cliente che effettua il caffè nel migliore dei modi.

Dai dati di vendita si ricava che, nel periodo selezionato, sono stati venduti:

- 1432 caffè doppi, corrispondenti a 716 erogazioni registrate dalla telemetria;
- 48 caffè di altro tipo (singoli, erogazioni continue o leva);
- **Numero totale di erogazioni: 764.**

La telemetria ha rilevato 597 erogazioni, ovvero  $3/4$  circa di quelli attesi: si nota che la differenza tra le erogazioni attese e rilevate in questo caso è minore, dovuto al fatto che le erogazioni libere sono in numero minore rispetto ai casi precedenti, e quindi sono meno le rilevazioni filtrate dalla telemetria Cimbali.

Dopo la pulizia dei dati, otteniamo un numero di erogazioni come descritto nella tabella 6.31.

	Soglie quantità acqua (ml)	Erogazioni totali	Erogazioni rimanenti	Utilizzo del gruppo
Gruppo 1	17 - 30	291	258	caffè doppio
Gruppo 2	21 - 27	306	272	caffè doppio

Tabella 6.31: Dettagli della pulizia dei dati per gruppo per Top shop nel periodo 08/07/2019 – 27/07/2019

Si nota come in questo caso non ci sia un gruppo dedicato alle erogazioni singole: infatti, normalmente, è presente un gruppo dedicato ad esse in caso di macchine con più gruppi, così da poter comunque permettere l'erogazione di più caffè contemporaneamente.

Vengono allora assegnate le etichette di qualità secondo le soglie riportate nella tabella 6.21.

Soglie	Complessivo			Gruppo 1			Gruppo 2		
	<	OK	>	<	OK	>	<	OK	>
Soglie di tempo	63	417	63	45	186	31	18	231	32
Soglie di flusso originali	5	538	0	4	258	0	1	280	0
Soglie di flusso proposte	472	71	0	236	26	0	236	45	0

Tabella 6.32: Conteggio delle etichette di qualità per Top shop nel periodo 08/07/2019 – 27/07/2019.

Dal conteggio delle etichette (tabella 6.32) si nota che, nonostante la maggior parte delle erogazioni sia in soglia per quanto riguarda il tempo, quasi tutte presentano un flusso sotto

soglia. Ciò potrebbe essere causato da una dose maggiore di macinato, come avvenuto per la Nuvola della Lavazza, in quanto cliente di buona qualità.

Si procede con il calcolo delle statistiche per le varie caratteristiche nel periodo in analisi.

	Flusso (ml/s)		Tempo di erogazione (s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Gruppo 1	1,00	0,15	23,00	3,40	22,00	1,15
Gruppo 2	1,00	0,13	23,60	2,60	23,80	1,20

Tabella 6.33: Calcolo statistiche per la Nuvola Lavazza nel periodo 17/05/2019-30/05/2019.

Dalla tabella 6.33 si può vedere come il comportamento dei due gruppi sia coerente, con una piccola differenza nella acqua erogata, che potrebbe essere causata da delle differenze intrinseche dei gruppi che potrebbe portare a un rallentamento del processo di erogazione del caffè nel gruppo 2. Dal confronto con la tabella 6.25 emerge che i due periodi presentano statistiche molto simili, con l'unica differenza che il Top Shop eroga i caffè in un tempo leggermente inferiore, portando quindi anche un acqua erogata (ricavata come flusso per tempo) leggermente inferiore.

Si può dunque concludere che il funzionamento delle due macchine dello stesso modello è analogo: in particolare è possibile affermare ciò poichè i due clienti analizzati hanno caratteristiche simili e si è sicuri della buona qualità dei loro caffè.

### Confronto tra periodi su macchine di costruttori diversi

Per concludere l'analisi, si effettua lo studio di un periodo su una macchina diversa, la Wega My Concept. Si è preso in considerazione il periodo che va dal 06/06/2019 al 16/06/2019 sulla macchina del cliente Spes SCS Etiko. A differenza dei casi precedenti, i dati di vendita riportano che solo il 65% dei caffè venduti è doppio, fattore che rende difficile l'analisi in quanto i dati sono disomogenei tra loro. Oltre alle erogazioni doppie però, si ha la presenza di erogazioni singole con una percentuale del 27%: essendo erogazioni singole e doppie normali comparabili tra loro, si procede con lo studio, in particolare perchè ci si aspetta che uno dei tre gruppi di erogazione effettuati sono erogazioni singole.

Nel particolare, le vendite riportano:

- 2754 caffè doppi, corrispondenti a 1377 erogazioni registrate di caffè doppi;
- 1104 caffè singoli corrispondenti altrettante erogazioni in telemetria;
- 190 caffè di altro tipo(lungo, corto, erogazioni continue);
- **Numero totale di erogazioni: 2671.**

La telemetria ha rilevato 2665 erogazioni, ovvero il numero di osservazioni attese: ciò dimostra, come previsto, che la discrepanza tra dati di vendita e rilevazioni della telemetria è dovuto non a un problema del device di telemetria della Lavazza, ma all'operazione di pulizia effettuata dalla telemetria del costruttore Cimbali. Infatti My Concept appartiene al costruttore Wega, e non elabora i dati.

	Soglie quantità acqua (ml)	Erogazioni totali	Erogazioni rimanenti	Utilizzo del gruppo
Gruppo 1	6 - 21	1246	1136	caffè singolo
Gruppo 2	21 - 31	985	831	caffè doppio
Gruppo 3	17 - 33	434	352	caffè doppio

Tabella 6.34: Dettagli della pulizia dei dati per gruppo per Gruppo Spes SCS Etiko nel periodo 06/06/2019–16/06/2019.

La tabella 6.34 presenta le operazioni di pulizia effettuate e i dati rimanenti: inoltre, il gruppo 1, che presenta un erogato ridotto, viene considerato come gruppo singolo.

Si precisa che a differenza dei due precedenti clienti, Spes SCS Etiko è un cliente consigliato dalla Lavazza per l’alto numero di erogazioni a disposizione, non per la qualità del caffè. Lo studio delle etichette di qualità e delle statistiche quindi non fornirà una differenziazione esatta tra macchine, ma terrà conto anche delle abitudini del cliente. Sicuramente il cliente in questione non usa una dose di macinato superiore al normale. Inoltre, purtroppo, non si è a conoscenza della funzione di trasferimento da telemetria a tazza (e viceversa) sulla macchina Wega My Concept: non sarà quindi possibile risalire alle statistiche in tazza del flusso. Per quanto riguarda l’assegnazione delle etichette di qualità vengono usate, per coerenza, comunque le soglie in telemetrie calcolate con la funzione di trasferimento per la Faema E71.

Sono quindi assegnate le etichette di qualità alle erogazioni:

Soglie	Complessivo			Gruppi doppi			Gruppo singolo		
	<	OK	>	<	OK	>	<	OK	>
Soglie di tempo	450	708	1161	93	397	693	357	311	468
Soglie di flusso originali	146	1997	176	104	1051	28	42	946	148
Soglie di flusso proposte	1434	669	216	904	249	30	530	420	186

Tabella 6.35: Conteggio delle etichette di qualità per Gruppo Spes SCS Etiko nel periodo 06/06/2019–16/06/2019.

In questo caso si nota subito come la distribuzione delle etichette sia differente dai casi precedenti: in primis abbiamo che la maggior parte delle erogazioni ha un tempo maggiore rispetto alle soglie: ciò è molto particolare in quanto normalmente il cliente cerca di velocizzare il processo di erogazione, non di allungarlo. Inoltre abbiamo per la prima volta dei caffè aventi un flusso sopra soglia, sebbene esso sia in numero ridotto rispetto a quelli sotto soglia.

Osservando le statistiche calcolate sui dati si vede come il gruppo singolo abbia un flusso leggermente maggiore, e in generale una deviazione standard maggiore, che rappresenta la presenza di erogazioni di tipo diverso (continue, corte, lunghe). Si nota anche come il tempo di erogazione sia più elevato e variabile rispetto al modello di riferimento rappresentato dalla Nuvola (tabella 6.25). Ciò potrebbe essere dovuto all’uso di una miscela diversa che

necessita una macinatura più fine, oppure alla presenza di sporczia o guasti interni alla macchina che rallentano il fluire dell'acqua: sicuramente però l'erogato in tazza del cliente in analisi sarà maggiore di quello della Nuvola Lavazza, a causa di una diversa taratura.

	Flusso (ml/s)		Tempo di erogazione (s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Gruppo 1	1,17	0,68	26,50	11,50	25,80	6,40
Gruppo 2	1,02	0,40	28,50	8,30	26,10	3,10
Gruppo 3	0,98	0,31	28,80	7,20	26,30	1,60

Tabella 6.36: Calcolo statistiche per Gruppo Spes SCS Etiko nel periodo 06/06/2019–16/06/2019.

In quest'ultimo caso possiamo vedere come il cambio di modello, e soprattutto il cambio di cliente, possano influire nelle caratteristiche dell'erogazione.

## Conclusioni

Per riassumere l'analisi dei clienti sul mercato, possiamo individuare alcuni elementi chiave:

- Sono stati scelti solo periodi con erogazioni doppie dai dati di vendita, in quanto la telemetria non ci permette di discernere il tipo di erogazione;
- È stata rilevata una discrepanza tra numero di rilevazioni e dati di vendita sulle macchine del costruttore Cimbali, poiché il dato viene prima filtrato dalla macchina;
- Grazie all'analisi dell'erogato si può individuare la presenza di gruppi con erogatore singolo: essi normalmente hanno erogazioni più variabili perché la singola erogazione è normalmente preferita per effettuare dosi corte, lunghe e libere;
- È stato riscontrato che la dose in tazza nei clienti di riferimento è molto inferiore rispetto a quella indicata dagli esperti: si parla di 17-18 ml di erogato rispetto ai 24 indicati.
- Le funzioni di trasferimento calcolate nelle sessioni sperimentali si basano su un certo valore dell'erogato in tazza: tarature diverse possono portare a funzioni di trasferimento diverse, poiché cambia la quantità di caffè erogato e di conseguenza anche il tempo di erogazione. Nella sessione sperimentale proposta è stata definita una sequenza di test per approfondire il tema della taratura e di come reagisce la macchina a una variazione dei colpi di ventolino impostati.
- Sono state prese come riferimento per il portale della telemetria le soglie calcolate a partire dai dati estratti dalla Nuvola della Lavazza: queste soglie potrebbero avere dei valori troppo bassi poiché i caffè da cui sono calcolate sono effettuati con una dose di macinato maggiore al normale: per gli studi futuri, avendo più macchine controllate dalla telemetria, e conoscenza del tipo di erogazione di cui è stata fatta la rilevazione, si consiglia di prendere periodi più lunghi e su macchine diverse, e fare una media delle statistiche calcolate così da trovare delle soglie più resistenti e generali.

## 6.6.2 Intervento tecnico

Uno degli obiettivi finali della Lavazza è quello di poter effettuare una manutenzione predittiva della macchina grazie ai dati di telemetria, così da ridurre il numero di interventi tecnici necessari e segnalare quando la macchina presenta un degrado delle prestazioni o delle variazioni inaspettate. Infatti, gli interventi tecnici rappresentano un grande costo economico per la Lavazza, che mette a disposizione al cliente la macchina del caffè in comodato d'uso garantendone la manutenzione necessaria. Generalmente le richieste di intervento sono effettuate dal cliente stesso, e hanno tempi di attuazione che vanno dalla settimana al giorno stesso. Potendo predire la necessità dell'intervento e, successivamente, la sua causa, la Lavazza mira a ottimizzare i tempi di intervento e la gestione dei suoi tecnici.

Il tema in questione non è affatto banale da affrontare: il principale limite al momento sta nel fatto che non si possiede l'informazione del tipo di erogazione collegata ai dati rilevati. Risulta quindi complicato il paragone di flusso e tempo di erogazione, poichè appartenenti a erogazioni con caratteristiche diverse. Inoltre, è difficile reperire al momento lunghi periodi di rilevazione in quanto il device di telemetria è stato installato sulle macchine in produzione solo recentemente, e l'unico accesso ai dati disponibile al momento è attraverso il portale della telemetria, che permette solo la selezione di periodi di massimo 41 giorni.

Questa sezione si propone di svolgere un'analisi esplorativa sulla fattibilità della manutenzione predittiva della macchina: sono stati presi in considerazione alcuni interventi tecnici svolti su macchine in telemetria, e studiare il loro impatto sulle tre variabili caratteristiche dell'erogazione (tempo, flusso e acqua erogata, ottenuta come prodotto delle prime due). In particolare si cerca di verificare la presenza effettiva dell'intervento e se esso porta le variazioni che ci aspettiamo.

Il nostro punto di partenza è il file excel contenente gli interventi tecnici (vedi sezione 4.6). Esso contiene tutti gli interventi tecnici effettuati dalla Lavazza, riportando molte informazioni, come il cliente che ha richiesto l'intervento, la data di richiesta e quella di avvenimento, il tipo di attività svolte e i pezzi sostituiti, e infine il costo. Il file presenta però tutti i tipi di intervento, anche quelli svolti su altri tipi di attrezzatura, come i macinadosatori, e anche su macchine non ancora controllate dalla telemetria. È stato necessario filtrare allora i dati in modo da ottenere solo gli interventi che possono essere monitorare in telemetria: inoltre, essi sono stati compattati e ordinati secondo il costo dell'intervento, così da porre l'attenzione ai casi in cui la manutenzione ha avuto peso più rilevante. Gli interventi analizzati prevedono in genere la sostituzione di alcuni componenti. Tra quelli analizzati, i pezzi sostituiti sono:

- Brita Filter: un filtro dell'acqua che riduce il calcare presente in nell'acqua che entra nella macchina. È un pezzo molto costoso, e la sostituzione è esterna alla macchina e non dovrebbe quindi alterare le caratteristiche dell'erogazione.
- Brew Unit: è l'unità di infusione, dove viene inserito il macinato nell'erogazione.
- Shower Head: le doccette che si occupano di distribuire uniformemente sul panetto l'acqua che viene spinta dalla pompa. Normalmente ci si aspetta che l'erogazione del caffè diventi più rapida dopo la sostituzione;

- Internal Gasket: guarnizione interna, che permette una corretta chiusura dell'erogatore e un'erogazione ottimale. In genere sostituita assieme alle doccette.

Normalmente, durante un intervento, vengono sostituiti almeno due dei componenti sopra citati: l'intervento base e più numeroso è la sostituzione di doccette e guarnizione.

Come descritto in precedenza, in mancanza del tipo di erogazione collegato ai dati rilevati, si deve effettuare una pulizia in modo da mantenere solo erogazioni dello stesso tipo, così che possano essere puntualmente confrontate tra loro. Il processo di pulizia è effettuato attraverso l'imposizione di soglie sulla quantità di acqua erogata, ottenuta come tempo di erogazione per flusso. Le soglie cercano di essere molto strette così da ridurre il numero di erogazioni diverse da quella più frequente: si noti che il tipo di erogazione in analisi potrebbe non essere un caffè normale, ma anche un caffè lungo a seconda delle caratteristiche del cliente. Inoltre, nonostante la pulizia, potrebbero rimanere comunque erogazioni libere che inficiano in parte il calcolo delle statistiche: si è però deciso di non imporre delle soglie su tempo di erogazione e flusso in quanto alcuni loro valori estremi potrebbero indicare un segnale della presenza dell'intervento tecnico.

Analizzando i vari clienti si è notato che il comportamento dei gruppi è coerente tra loro, anche a cavallo di un intervento, e si riportano allora solo le statistiche calcolate per un gruppo scelto come rappresentate poiché presenta un erogato più uniforme e erogazioni doppie.

Una volta effettuata la pulizia, viene inizialmente preso in considerazione il periodo a cavallo dell'intervento, in particolare i 7 giorni precedenti e i 7 successivi ad esso. Vengono quindi calcolate media e deviazione standard di tempo di erogazione, flusso e quantità erogata nei due periodi e confrontati. La tabella 6.16 contiene le statistiche degli interventi analizzati, ordinati per costo decrescente.

Cliente	Macchina	Data	Intervento	Periodo	Tempo erogazione		Flusso		Erogato	
					Media	Dev	Media	Dev	Media	Dev
LEA SRL	FAEMA e71	05/06	BF+SH	Prima	22	9	1,60	0,41	33,1	3,2
				Dopo	27	10	1,22	0,18	32,7	1,8
FARMACIA DEL CAMBIO	LACIMBALI M100	17/07	BF+BU	Prima	22	3	1,16	0,16	25,0	1,7
				Dopo	24	3	1,01	0,15	24,1	1,5
SPES ETIKO	WEGA MY CONCEPT	17/06	SH+IG	Prima	28	8	1,08	0,43	26,6	3,1
				Dopo	26	8	1,15	0,45	26,8	3,7
LA ROSA SAS	WEGA MY CONCEPT	26/07	SH+IG	Prima	21	6	1,41	0,37	28,3	2,9
				Dopo	29	5	0,99	0,40	25,9	3,0
IL PEPE	RANCILIO CL11	20/09	SH+IG	Prima	20	10	1,90	0,66	33,7	1,9
				Dopo	17	11	2,19	0,61	33,9	1,6
RAL SRL	RANCILIO CL11	24/06	SH+IG	Prima	27	5	1,57	0,55	38,0	2,9
				Dopo	27	7	1,62	0,70	38,4	3,0

Figura 6.16: Statistiche degli interventi analizzati

Gli interventi analizzati si possono dividere in due categorie:

- Sostituzione del filtro Brita, assieme ad altri componenti come doccette e unità di infusione: si ha un incremento, anche significativo, del tempo di erogazione e conseguente diminuzione del flusso. La sostituzione del filtro è però esterna alla macchina e non dovrebbe influenzare le erogazioni, e la sostituzione dei restanti componenti dovrebbe variare le erogazioni in modo contrario rispetto a quanto rilevato.
- Sostituzione di doccette e guarnizioni: sono gli interventi più frequenti e meno costosi, che portano a una leggera diminuzione del tempo di erogazione, mantenendo l'erogato costante. Ci sono però due casi anomali, in cui la variazione o non è rilevata o è molto elevata e contrario rispetto a quella attesa.

Si nota quindi che il comportamento delle erogazioni a cavallo dell'intervento non è sempre uguale. Vanno tenute in considerazione infatti due caratteristiche:

- Il periodo precedente all'intervento è un periodo in cui la macchina non funziona correttamente, ed può quindi avere un comportamento diverso da quello standard;
- Dopo l'intervento può essere riefettuata la taratura della macchina che può cambiare completamente le caratteristiche dell'erogazione poichè cambia la quantità di acqua erogata.

Che sia dovuto a una ritaratura oppure all'effetto dell'intervento, dovrebbe essere possibile quindi verificare la presenza effettiva dell'intervento.

Sono analizzati ora alcuni casi di interventi che hanno suscitato il nostro intervento, poichè i dati hanno riferito una variazione inaspettata, troppo evidente oppure troppo poco.

### Intervento con variazione inattesa delle statistiche

Si prende in considerazione l'intervento più costoso, effettuato dal cliente Lea Srl San Tommaso il 05/06/2019. Esso presenta la sostituzione del filtro Brita e delle doccette. È di nostro interesse in quanto, come riportato in tabella 6.37 dopo l'intervento si ha un incremento significativo dei tempi di erogazione e conseguente diminuzione del flusso, mantenendo un erogato costante. Inoltre si nota una deviazione standard molto elevata, soprattutto prima dell'intervento, simbolo della presenza di erogazioni diverse non eliminate attraverso la pulizia dei dati oppure di oscillazioni.

periodo	Tempo di erogazione (s)		Flusso (ml/s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Prima	22	9	1,60	0,41	33,1	3,2
Dopo	27	10	1,22	0,18	32,7	1,8

Tabella 6.37: Statistiche dei periodi a cavallo dell'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso.

Il cliente analizzato effettua, in media, 100 erogazioni al giorno. Analizzando l'andamento dell'erogato nelle due settimane a cavallo dell'intervento (figura 6.17) notiamo che prima dell'intervento (indicato col rettangolo giallo) si hanno delle oscillazioni, mentre dopo è più stabile.

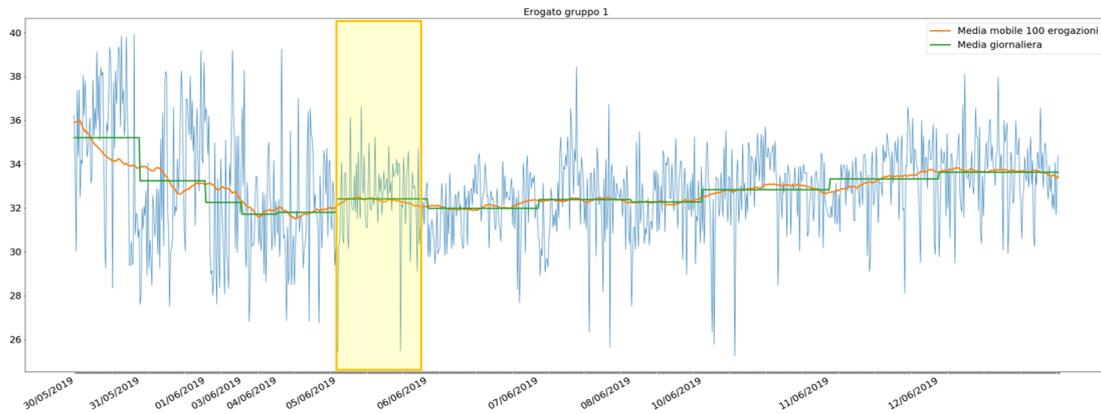


Figura 6.17: Andamento dell'acqua erogata nel periodo a cavallo dell'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso.

Osservando perciò il periodo precedente all'intervento (figura 6.18), si può notare che a un certo punto hanno iniziato a verificarsi delle oscillazioni prima non presenti. In particolare si può notare come la data di richiesta di intervento, evidenziata con una linea gialla, sia proprio nelle giornate in cui queste oscillazioni si verificano: questa ondulazione probabilmente è dovuta al barista che cerca di regolare la macinatura del caffè e la quantità in tazza per evitare l'effetto del problema della macchina.

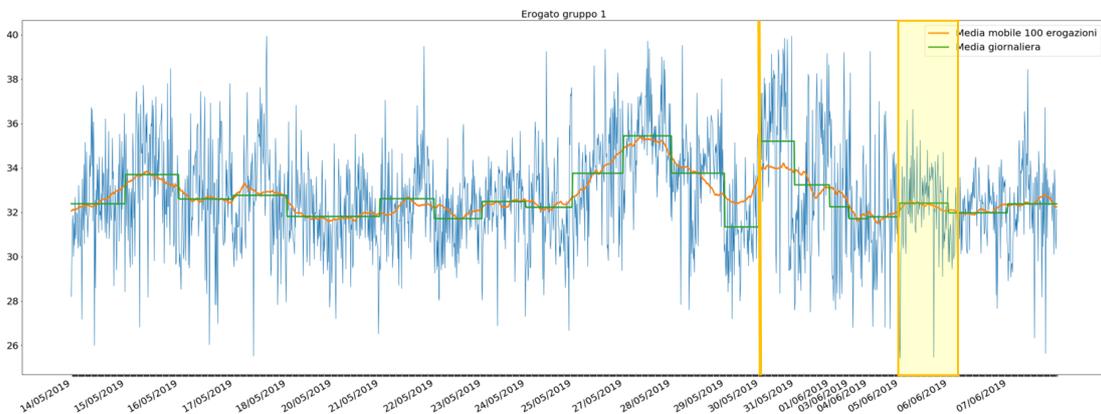
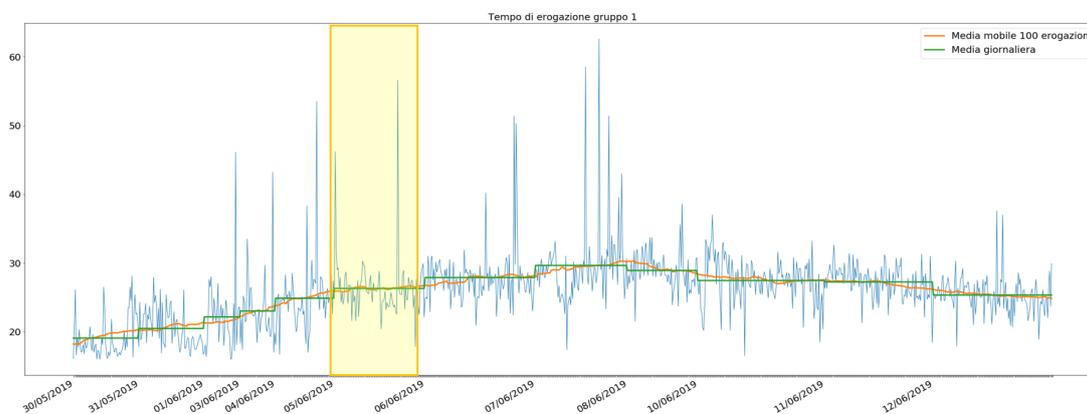


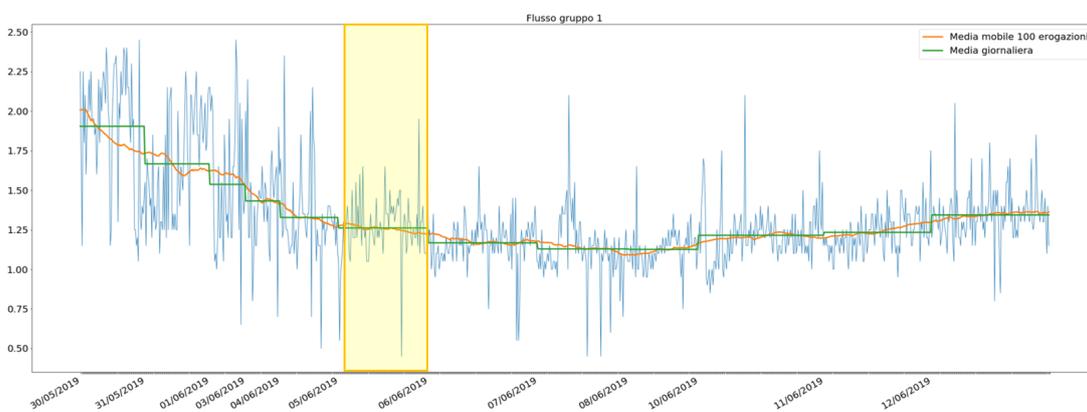
Figura 6.18: Andamento dell'acqua erogata nel periodo precedente all'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso .

Si riportano in figura 6.19, per completezza, anche gli andamenti del tempo di erogazione e del flusso nel periodo a cavallo dell'intervento, così da poter verificare la stabilizzazione degli andamenti e la corrispondenza con le statistiche calcolate precedentemente (aumento del tempo e diminuzione del flusso): esse non hanno salti improvvisi e non sembrano quindi dovuti a una ritaratura della macchina, bensì all'effetto dell'intervento che, nonostante prevedesse una riduzione del tempo di erogazione, in presenza di una taratura effettuata

precedentemente ad esso, porta gli effetti osservati.



(a) Tempo di erogazione



(b) Flusso

Figura 6.19: Andamento di flusso e tempo nel periodo a cavallo dell'intervento del 05/06/2019 dal cliente Lea Srl San Tommaso .

### Intervento con elevata variazione delle statistiche

Viene preso in considerazione l'intervento del 26/07/2019 al cliente La Rosa Sas poichè è rilevato un inaspettato aumento del tempo di erogazione e una diminuzione del flusso, oltre a una variazione evidente anche della quantità di acqua erogata, come riportato nella tabella 6.38. L'intervento prevede la sostituzione di doccette e guarnizione.

periodo	Tempo di erogazione (s)		Flusso (ml/s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Prima	21	6	1,41	0,37	28,3	2,9
Dopo	29	5	0,99	0,40	25,9	3,0

Tabella 6.38: Statistiche dei periodi a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas.

Analizzando l'andamento della quantità di acqua erogata nel periodo a cavallo dell'intervento (figura 6.20), si nota come dal giorno dell'intervento si veda una brusca variazione. Essa è confermata anche dalle variazioni di flusso e tempo (figura 6.21): particolare il fatto che dopo l'intervento però il tempo risulti troppo elevato.

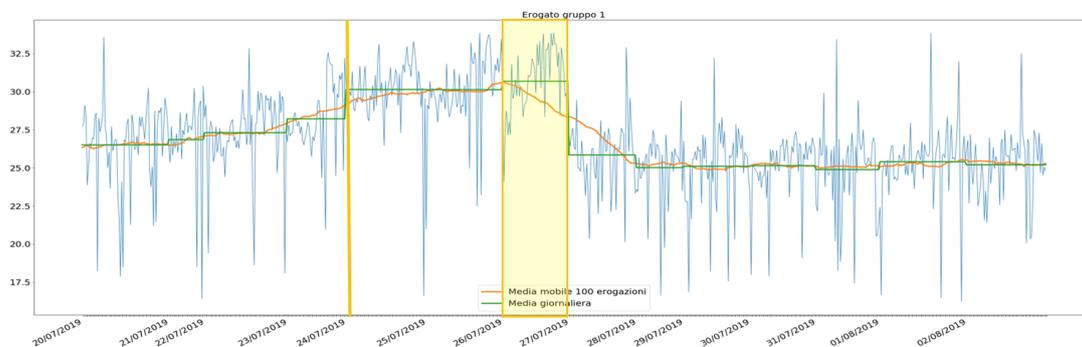


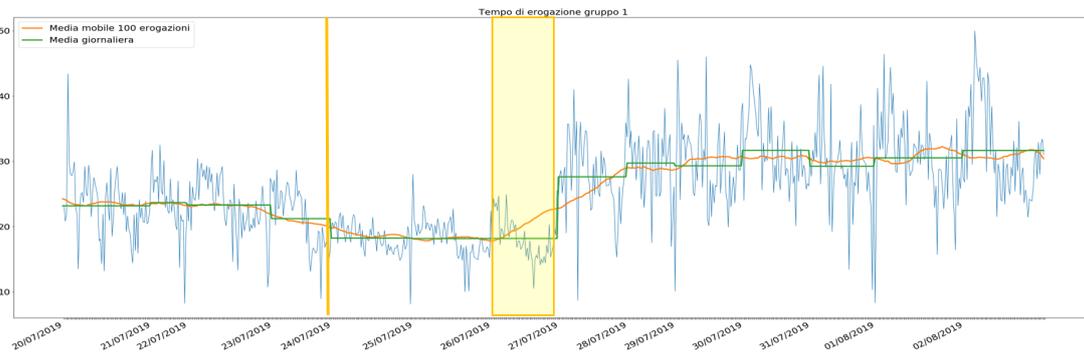
Figura 6.20: Andamento dell'acqua erogata nel periodo a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas.

Si tratta molto probabilmente di una nuova taratura della macchina, che però non sembra ottimale.

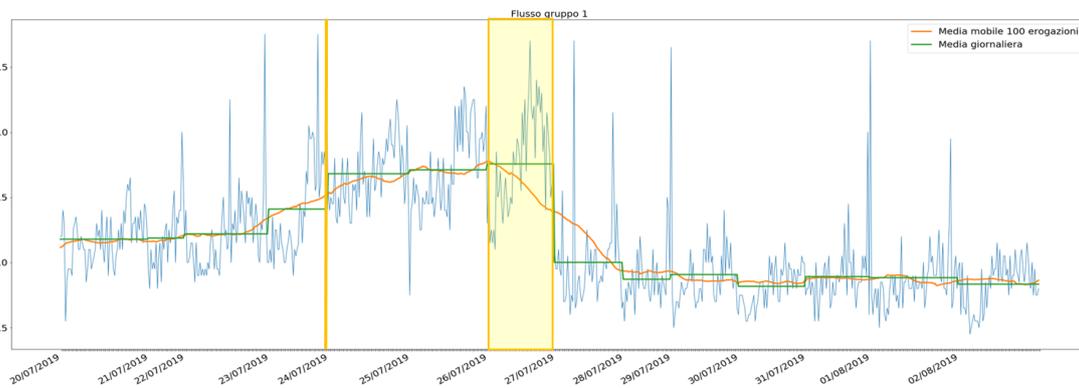
Possiamo quindi individuare l'intervento grazie alla taratura della macchina, che però non permette un confronto significativo dei due periodi.

Da notare però come nei giorni tra la richiesta dell'intervento e l'arrivo del tecnico si abbia un tempo di erogazione ancora minore, probabilmente a causa delle doccette che non distribuendo l'acqua uniformemente sul panetto di caffè aumentano la forza penetrante dell'acqua riducendo così il tempo di erogazione.

Dalla figura 6.22 si può vedere l'andamento del flusso in un periodo più ampio a cavallo dell'intervento: si nota come prima dell'intervento esso fosse più oscillante, mentre successivamente esso è più stabile: quindi l'intervento ha avuto successo.



(a) Tempo di erogazione



(b) Flusso

Figura 6.21: Andamento di flusso e tempo nel periodo a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas.

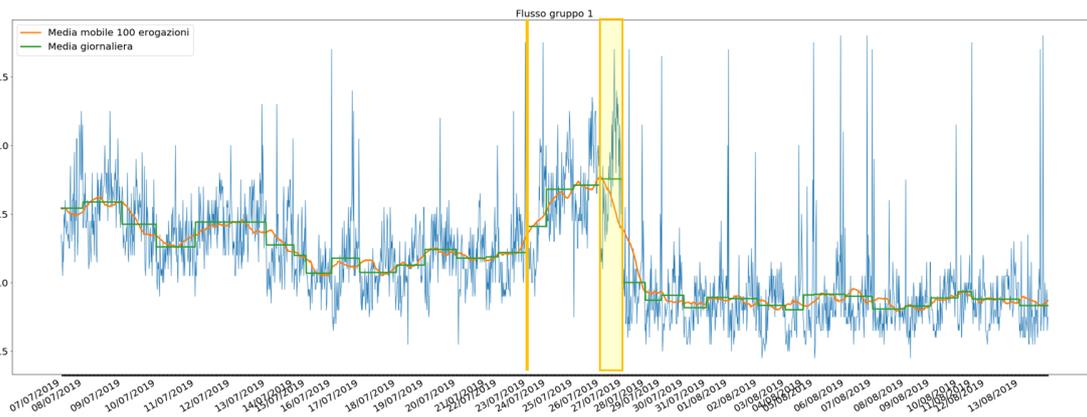


Figura 6.22: Andamento del flusso nel periodo a cavallo dell'intervento del 26/07/2019 al cliente La Rosa Sas.

### Intervento senza variazioni delle statistiche

Si analizza come ultimo intervento quello del 17/06/2019 al cliente Spes SCS Etiko. Dalle statistiche riportate in tabella 6.39, si nota come l'intervento abbia avuto un impatto minimo sulle statistiche.

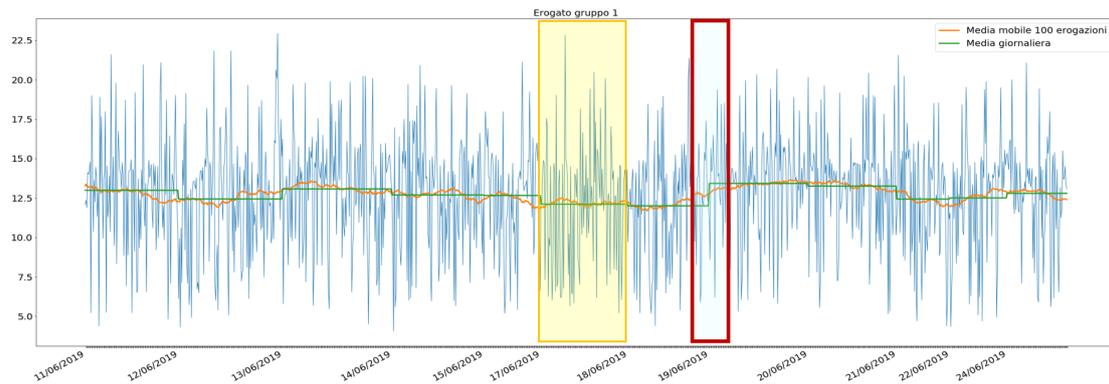
periodo	Tempo di erogazione (s)		Flusso (ml/s)		Acqua erogata (ml)	
	media	dev. std.	media	dev. std.	media	dev. std.
Prima	28	8	1,08	0,43	26,6	3,1
Dopo	26	8	1,15	0,45	26,8	3,7

Tabella 6.39: Statistiche dei periodi a cavallo del 17/06/2019 al cliente Spes SCS Etiko.

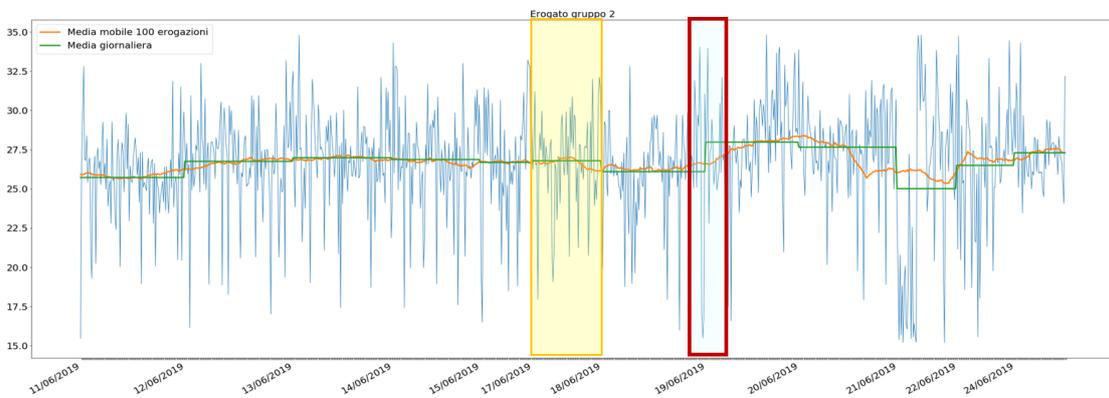
Analizzando le distribuzioni della quantità di acqua erogata dei gruppi presenti sulla macchina (figura 6.23), ci si è però resi conto che, nella data dell'intervento, non si verifica alcuna variazione: inoltre si osserva che in data 19/06/2019 (evidenziata in verde) si registra invece una leggera variazione. Questo andamento è supportato anche dall'analisi del tempo di erogazione e del flusso. Ciò potrebbe indicare che l'intervento non è stato effettuato nella giornata indicata, bensì in quella successiva: attraverso uno studio più approfondito si potrebbe quindi monitorare l'operato del tecnico che effettua l'intervento.

L'elemento che però salta maggiormente all'occhio è il fatto che prima dell'intervento non ci siano delle oscillazioni, a differenza degli interventi analizzati precedentemente.

Ci si chiede allora se l'intervento fosse davvero necessario.



(a) Gruppo 1



(b) Gruppo 2

Figura 6.23: Andamento della quantità di acqua erogata nel periodo a cavallo dell'intervento del 17/06/2019 al cliente Spes SCS Etiko.

## Conclusioni

Riassumendo quindi questa analisi esplorativa riguardante gli interventi tecnici, possiamo concludere che è possibile individuare un intervento tecnico poichè si registrano delle variazioni nelle rilevazioni, oppure perchè associata a una nuova taratura. In alcuni casi, una variazione minima può essere indicatore di una discrepanza tra data reale dell'intervento e quella riportata negli archivi della Lavazza. Inoltre si nota che in genere l'andamento delle caratteristiche delle erogazioni si presenta oscillatorio prima dell'intervento e più piatto dopo di esso.

Si può quindi continuare con uno studio mirato alla manutenzione predittiva della macchina: ciò richiede un periodo di rilevazione prima dell'intervento maggiore, così da studiare meglio il comportamento della macchina in situazioni ottimali e poi individuare delle variazioni che portano alla richiesta dell'intervento.

Attualmente non è possibile analizzare lunghi periodi in quanto sulla maggioranza delle macchine il device di telemetria è appena stato installato: inoltre il reperimento di lunghi periodi tramite portale della telemetria risulta particolarmente oneroso.

Una delle proposte per i prossimi studi potrebbe essere quella di monitorare il comportamento delle macchine dalla loro installazione così da modellare il suo comportamento. Per migliorare ulteriormente l'analisi si rivela fondamentale l'informazione sul tipo di erogazione a cui corrispondono i dati rilevati: in questo modo si velocizzerebbe il processo di pulizia e potrebbero confrontare meglio i periodi a cavallo dell'intervento che conterrebbero così solo uno stesso tipo di erogazione.

## Capitolo 7

# Conclusioni

Questa tesi presenta un'analisi esplorativa di dati riguardanti le caratteristiche dei caffè. In particolare, avviene il confronto tra i dati rilevati in tazza e quelli ottenuti grazie a un device di telemetria installato sulle macchine da bar. In primis sono stati svolti dei test in laboratorio così da poter raccogliere i dati da diverse fonti attraverso misurazioni dirette, scheda elettronica e telemetria. Inoltre è stato possibile effettuare degli esperimenti in condizioni ottimali in cui si è studiata l'influenza della variazione di parametri esterni quali dose del macinato, macinatura dei grani e pressione della pompa, sulla qualità dell'influenza. In aggiunta, si è valutato anche le differenti caratteristiche delle erogazioni svolte su gruppi in condizioni di pulizia diverse.

Dopo aver scoperto che la taratura di una macchina stabilisce la quantità di acqua erogata dalla macchinetta, l'attenzione è stata posta sugli andamenti di flusso e tempo di erogazione dei vari test: sono state calcolate delle nuove soglie di qualità per il flusso a partire dai dati, così che fossero più coerenti ai dati rilevati e centrate sul flusso ottimale fornitoci dagli esperti, ovvero 1 ml/s. È stata però necessaria una pulizia dei dati, in quanto in 17 casi la prima erogazione del test presenta un'erogazione con quantità in tazza ridotta, probabilmente perché la macchina necessita di un'erogazione per adattarsi alla variazione dei parametri. Grazie alla presenza di queste soglie, è stato applicato delle etichette di qualità alle varie erogazioni, che hanno permesso l'uso delle regole di associazione per trovare alcuni pattern frequenti. In particolare sono stati effettuati due studi: il primo mirato a capire l'influenza delle variabili esterne sulla qualità del caffè, mentre il secondo per capire quale sia l'impatto dell'assenza dei lavaggi della macchina sulle caratteristiche dell'erogazione. Attraverso l'analisi di regole con tre semantiche diverse, è stato possibile capire che macinatura grossolana, dose ridotta e pressione elevata producono caffè più veloci, e quindi più slavati, mentre macinatura fine, dose elevata e pressione bassa riducono la velocità con cui l'erogazione avviene, compromettendo la qualità del panetto di caffè. Sono stati rilevati anche alcuni elementi di compensazione tra essi.

In aggiunta, è evidente come il gruppo sporco presenti flussi più lenti rispetto a quello pulito, e come un aumento della pressione sia molto attenuato a causa della mancata pulizia. Molto interessante rilevare anche come gli effetti di compensazione siano più probabili in cliente che effettua i lavaggi, e come, la causa di un caffè lento in un cliente che non lava sia maggiormente dovuta all'utilizzo di una dose ridotta di macinato.

Confrontando le rilevazioni in tazza e di telemetria, è stato possibile calcolare la funzione che lega le due fonti, così da poter ricondurre le osservazioni in telemetria allo studio in tazza effettuato e viceversa. Si è scoperto come la funzione di trasferimento per il flusso dipenda dalla tipologia di macchina e dalla taratura effettuata.

Grazie alle conoscenze acquisite da questi studi, e agli interessi principali dell'azienda, è stata proposta una sessione sperimentale da applicare a diversi tipi di macchina così da poter modellarne appieno il comportamento. In particolare sono stati proposti alcuni test per effettuare delle analisi sulle tarature, e quindi sulla relazione dei dati in tazza e in telemetria, sull'individuazione dei lavaggi anomali, sulla rilevazione di un degrado dovuto a una mancanza di pulizia e infine alle variazioni dei parametri esterni.

A questo punto, l'analisi si è concentrata sull'esplorazione dei dati provenienti dal mercato. Sono emerse alcune criticità nel sistema di rilevazione, in particolare su alcune macchine. Inoltre è stato creato un metodo di pulizia e allineamento dei dati rilevati.

Si è cercato allora di analizzare il comportamento di alcuni clienti fidati, con l'obiettivo di trovare delle soglie di riferimento sui dati di telemetria. Non avendo a disposizione l'informazione sul tipo di erogazione analizzata, è stato proposto e applicato un metodo di pulizia basato sulla quantità di acqua erogata dalla macchina, calcolata come tempo per flusso medio. Ciò permette inoltre di individuare e discernere gruppi atti a erogazioni singole e doppie.

Infine si svolge un'analisi esplorativa di alcuni interventi tecnici, che cercano di verificare gli effetti dovuti agli interventi, e che rilevano come l'intervento sia spesso associato a una ritaratura della macchina. Gli studi futuri risultano promettenti anche grazie all'individuazione di un'oscillazione degli andamenti delle erogazioni, chiaro segnale di problemi all'interno della macchina.

## 7.1 Limiti attuali dell'analisi

L'analisi svolta ha presentato però alcune difficoltà di carattere tecnico.

In primis, il lungo tempo necessario per svolgere gli esperimenti in laboratorio a causa del grande numero di erogazioni da effettuare.

Inoltre, è stato necessario parecchio tempo per l'allineamento dei dati provenienti da diverse fonti, che non presentano elementi in comune per essere allineati. Ciò è stato in parte velocizzato grazie alla scrittura di alcuni script.

Infine, alcune analisi risultano limitate a causa della possibilità di selezione di periodi limitati dal portale web della telemetria: inoltre, manca un'informazione essenziale sul tipo di erogazione a cui corrisponde i dati. Infatti non è possibile discernere tra erogazioni corte, lunghe, singole, doppie e continue. Si è provato a proporre un metodo di pulizia basato sulla quantità di acqua erogata che rende le analisi più attendibili in attesa dell'aggiornamento del device che renderà disponibile il dato.

## 7.2 Possibili studi futuri

Gli studi compiuti si mostrano comunque molto promettenti. In particolare, con l'informazione del tipo di erogazione collegata alla rilevazione, aumenterà notevolmente la precisione delle analisi e i possibili studi.

Inoltre, la possibilità di accedere a periodi più lunghi grazie a un accesso diretto al database, permetteranno il monitoraggio di clienti nel lungo periodo, oltre al calcolo di soglie basate su una media calcolata tra più clienti rappresentativi in periodi più lunghi. In aggiunta, col monitoraggio di lunghi periodi, può essere approfondito il tema della predictive maintenance, che permette di predire quando un intervento è necessario osservando delle variazioni anomale nei dati.

Per quanto riguarda i test in laboratorio, lo svolgimento delle sessioni sperimentale proposte fornirà molte informazioni utili sul comportamento della macchina e su come differenti macchine reagiscano in maniera diversa.

In particolare sarà possibile analizzare l'influenza di variazioni meno estreme dei parametri esterni sulla qualità del caffè, con la possibilità di trovare delle soglie a partire dai dati di telemetria e applicare lo studio con le regole di associazione, svolto all'interno di questa tesi con dati riferiti in tazza, ai dati di telemetria, così da avvicinarsi al caso delle macchine sul mercato.

Infine sarà molto interessante analizzare i comportamenti studiati in questa tesi con l'uso di diverse miscele.



# Bibliografia

- [1] S. Andueza et al. «Influence of coffee/water ratio on the final quality of espresso coffee». In: *Journal of the Science of Food and Agriculture* (2007).
- [2] Rajat Vaidya Ashish Kumar Singh Aditya Sinha e Hrishikesh Vijay Kulkarni. «A Review Paper on IoT Based Coffee Vending Machine». In: *International Journal of Advanced Research in Computer and Communication Engineering* (2007).
- [3] *Association Rules Generation from Frequent Itemsets*. [http://rasbt.github.io/mlxtend/user\\_guide/frequent\\_patterns/association\\_rules/](http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/). [Online].
- [4] M. Marone C. Severini I. Ricci, A. Derossi e T. De Pilli. «Changes in the Aromatic Profile of Espresso Coffee as a Function of the Grinding Grade and Extraction Time: A Study by the Electronic Nose System». In: *Journal of Agricultural and food chemistry* (2015).
- [5] *Come regolare la macinatura del caffè per un espresso perfetto*. <https://www.ilcaffeespressoitaliano.com/2018/come-regolare-la-macinatura-del-caffe-per-un-espresso-perfetto/>. [Online].
- [6] T. Cerquitelli E. Baralis. *Association Rules Foudamentals*. Database e data mining group, Politecnico di Torino.
- [7] R. Fontana. *Joint Probability Distributions and Random Samples. Expected Values, Covariance, and Correlation*. Politecnico di Torino.
- [8] *Funzionamento della macchina espresso*. <https://www.carioka.it/funzionamento-della-macchina-espresso/>. [Online].
- [9] G. Bertorelle. *Statistica Descrittiva*. [http://m.docente.unife.it/giorgio.bertorelle/didattica\\_insegnamenti/biostatistica-1/Slide2.pdf](http://m.docente.unife.it/giorgio.bertorelle/didattica_insegnamenti/biostatistica-1/Slide2.pdf). [Online].
- [10] L.A. Kirkpatrick e B.C. Feeney. *A Simple Guide to IBM SPSS: For Version 20.0*. Cengage Learning, 2012. URL: <https://books.google.it/books?id=HR\1CAAAQBAJ>.
- [11] Benjamin Koke et al. «Sensor retrofit for a coffee machine as condition monitoring and predictive maintenance use case». In: feb. 2019.
- [12] *matplotlib*. <https://matplotlib.org/>. [Online].
- [13] L. Navarini e D. Rivetti. «Water quality for Espresso coffee». In: *Food Chemistry* 122.2 (2010). 5th Conference on Water in Food, pp. 424–428. ISSN: 0308-8146. DOI: <https://doi.org/10.1016/j.foodchem.2009.04.019>. URL: <http://www.sciencedirect.com/science/article/pii/S0308814609004725>.

- [14] *Python Data Analysis Library*. <https://pandas.pydata.org/>. [Online].
- [15] Viani Rinantonio e Illy Andrea. *Espresso Coffee: The Science of Quality*. Elsevier Academic Press, 2005. URL: <https://books.google.it/books?hl=en&lr=&id=AJdlfSFCmVIC&oi=fnd&pg=PP2&dq=espresso+coffee+quality&ots=ml7p4fzSkN&sig=f39Yh0JODvMDpnURLb1iXC9FK5E#v=onepage&q=espresso%20coffee%20quality&f=false>.
- [16] Anrew Ross. «Costa Coffee to deploy IoT-enabled vending machines». In: (set. 2019). URL: <https://www.information-age.com/costa-coffee-iot-vending-machine-123485369/>.
- [17] Wikipedia contributors. *Project Jupyter* — *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Project\\_Jupyter](https://en.wikipedia.org/wiki/Project_Jupyter). [Online].
- [18] Wikipedia contributors. *scikit-learn* — *Wikipedia, The Free Encyclopedia*. <https://en.wikipedia.org/wiki/Scikit-learn>. [Online].