



**Politecnico
di Torino**

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea Magistrale

A.a. 2025/2026

Digital Voice of Customer e applicazione di tecniche di
Topic Modeling e carte di controllo X-S

Relatore:

Professore: Federico Barravecchia

Candidato:

Riccardo Manna

Introduzione	4
1. La Voice of Customer	6
1.1 Introduzione alla Voice of Customer.....	6
1.2 La Voice of Customer nel contesto della Quality 4.0.....	7
1.3 Tecniche di analisi della Voice of Customer digitale	8
1.4 Indicatori della VoC: MTP, IMTP e MRP	10
1.5 Le anomalie nella VoC digitale	14
1.5.1 Spike Anomalies.....	15
1.5.2 Level Anomalies.....	16
1.5.3 Trend Anomalies	17
1.5.4 Seasonal Anomalies.....	18
1.6 Il rilevamento delle anomalie attraverso le carte di controllo	19
1.7 Introduzione alle carte di controllo	19
1.8 Struttura e interpretazione delle carte di controllo	20
1.9 Tipologie di carte di controllo	20
1.9.1 Carta $\bar{X}$ -R (Media e Range)	21
1.9.2 Carta $\bar{X}$ -S (Media e Deviazione Standard).....	23
1.9.3 Carte p e np	24
1.9.4 Carte c e u.....	26
1.10 Analisi dei residui con carta X-S e utilizzo della regressione polinomiale di terzo grado	29
2. Introduzione al caso studio: Descrizione del servizio.....	32
2.1 Implementazione dell'algoritmo di Topic Modeling.....	33
2.2 Analisi degli indicatori derivati dal Topic Modeling	36
2.3 Analisi MTP (Mean Topic Prevalence)	36
2.4 Analisi MRP (Mean Rating Proportion)	38
2.5 Analisi IMTP (Interval Mean Topical Prevalence)	39
3. Costruzione del Database e Pre-elaborazione dei dati	41
3.1 Struttura della Dashboard.....	43
3.2 Funzionamento interattivo.....	44

3.3	Costruzione Carte di controllo X-S	45
4.	Analisi dei risultati Topic rappresentativi dei determinanti di qualità	50
4.1	Struttura dell'analisi	50
4.1.1	Disneyland California – Show and Parade.....	52
4.1.2	Disneyland California – Special Event's celebration.....	56
4.1.3	Disneyland Hong-Kong – Judgment on Staff	60
4.1.4	Disneyland Paris – Disneyland Trip for all ages	66
4.1.5	Disneyland Paris – Ticket Purchase	70
5.	Conclusione Analisi: Analisi dei risultati.....	75
5.1	Punti di forza dell'analisi	76
5.2	Debolezze e limiti.....	76
5.3	Migliorie e sviluppi futuri dell'approccio	77
	Bibliografia.....	78
	Sitografia:	81
	Appendice A: Codice in R per automazione carte di controllo X-S	82

Introduzione

La trasformazione digitale ha profondamente modificato il modo in cui i consumatori interagiscono con le organizzazioni e condividono le proprie esperienze di acquisto e di utilizzo dei servizi. La diffusione delle piattaforme online, dei social network e dei siti di recensioni ha infatti reso possibile la produzione e la condivisione continua di grandi quantità di contenuti generati dagli utenti, comunemente definiti User-Generated Contents (UGC). Tali contenuti rappresentano oggi una delle principali fonti di informazione a disposizione delle imprese, poiché consentono di raccogliere in modo spontaneo e non sollecitato le opinioni, le aspettative e le valutazioni espresse dai clienti.

In questo contesto assume particolare rilevanza il concetto di Digital Voice of Customer (DVoC), intesa come l'insieme delle informazioni derivanti dalle interazioni digitali dei consumatori. La DVoC costituisce un'evoluzione delle tradizionali metodologie di raccolta del feedback, permettendo alle organizzazioni di monitorare in tempo reale la percezione dei clienti e di individuare tempestivamente criticità, opportunità di miglioramento e cambiamenti nelle preferenze degli utenti. Grazie alla crescente disponibilità di dati digitali, la Voice of Customer è diventata un elemento centrale nei moderni sistemi di gestione della qualità e si inserisce pienamente nel paradigma della Quality 4.0, che promuove l'integrazione di tecnologie digitali, strumenti di analisi avanzata e tecniche di intelligenza artificiale per supportare il miglioramento continuo dei processi.

L'elevata quantità di dati non strutturati generati dagli utenti pone tuttavia importanti sfide metodologiche. Le recensioni online sono caratterizzate da contenuti testuali eterogenei, spesso complessi da interpretare attraverso strumenti tradizionali. Per questo motivo, negli ultimi anni si è assistito a una crescente diffusione di tecniche di text mining e machine learning finalizzate all'estrazione automatica di conoscenza dai dati testuali. Tra queste, il Topic Modeling rappresenta uno degli approcci più utilizzati per identificare i principali argomenti discussi dagli utenti e trasformare grandi volumi di testo in informazioni sintetiche e interpretabili.

La letteratura ha ampiamente dimostrato l'utilità del Topic Modeling nell'individuazione dei principali determinanti della qualità percepita e nell'analisi della soddisfazione dei clienti. Tuttavia, gran parte degli studi esistenti si concentra prevalentemente sull'identificazione dei topic e sulla loro interpretazione statica, mentre risulta ancora limitata l'attenzione dedicata all'analisi della loro evoluzione temporale e all'individuazione di eventuali anomalie nei comportamenti osservati. In particolare, emerge l'esigenza di sviluppare strumenti in grado non solo di identificare i temi maggiormente discussi dai clienti, ma anche di monitorarne l'andamento nel tempo, distinguendo le normali variazioni del fenomeno da cambiamenti potenzialmente significativi.

Alla luce di tali considerazioni, il presente lavoro si propone di analizzare l'evoluzione temporale della qualità percepita attraverso lo studio della Digital Voice of Customer derivante da un ampio database di recensioni online. L'obiettivo della ricerca è comprendere come i principali determinanti della qualità percepita evolvano nel tempo e verificare la possibilità di individuare anomalie e variazioni significative attraverso l'utilizzo integrato di tecniche di Topic Modeling e strumenti di Statistical Process Control.

In particolare, il lavoro intende rispondere alla seguente domanda di ricerca:

Come evolve nel tempo la rilevanza dei principali determinanti della qualità percepita individuati attraverso la Digital Voice of Customer e in che modo è possibile identificare anomalie significative mediante l'applicazione di strumenti di Statistical Process Control?

Per rispondere a tale quesito è stato adottato un approccio metodologico articolato in diverse fasi. In primo luogo, è stato utilizzato un modello di Topic Modeling basato sullo Structural Topic Model (STM), finalizzato all'individuazione automatica dei principali temi presenti nelle recensioni analizzate. Successivamente, sulla base dei topic estratti, sono stati calcolati specifici indicatori quantitativi, tra cui il Mean Topic Prevalence (MTP), il Mean Rating Proportion (MRP) e l'Interval Mean Topical Prevalence (IMTP), utilizzati rispettivamente per misurare la rilevanza dei diversi topic, il livello di soddisfazione associato e la loro evoluzione temporale.

Le serie storiche ottenute mediante l'indicatore IMTP sono state successivamente analizzate attraverso modelli di regressione polinomiale di terzo grado, al fine di descrivere l'andamento atteso dei fenomeni osservati e isolare eventuali deviazioni significative. Infine, i residui derivanti dai modelli di regressione sono stati monitorati mediante carte di controllo X-S secondo i principi dello Statistical Process Control, con l'obiettivo di distinguere le normali oscillazioni del processo da possibili segnali di instabilità o eventi anomali.

Il contributo originale della presente ricerca consiste nell'integrazione tra tecniche di Topic Modeling e strumenti di Statistical Process Control per il monitoraggio dinamico della Digital Voice of Customer. In particolare, l'applicazione delle carte di controllo X-S ai residui delle serie storiche ottenute dagli indicatori di prevalenza dei topic consente di individuare anomalie e cambiamenti significativi nella percezione dei clienti, offrendo un approccio innovativo per il monitoraggio della qualità percepita nel contesto dei servizi.

La tesi è organizzata come segue. Il Capitolo 1 presenta il quadro teorico di riferimento relativo alla Voice of Customer, alla Digital Voice of Customer e al paradigma della Quality 4.0. Vengono inoltre illustrate le principali tecniche di analisi della VoC digitale, con particolare attenzione al Topic Modeling, agli indicatori derivati dall'analisi dei topic e agli strumenti di Statistical Process Control utilizzati per il monitoraggio delle anomalie.

Il Capitolo 2 introduce il caso studio oggetto dell'analisi, descrivendo il servizio considerato, il dataset di recensioni online utilizzato e l'implementazione del modello Structural Topic Model, illustrando inoltre il processo di identificazione e validazione dei topic.

Il Capitolo 3 descrive le attività di costruzione del database, preelaborazione dei dati e sviluppo degli strumenti di analisi, presentando gli indicatori utilizzati per la misurazione della rilevanza e della soddisfazione associate ai diversi topic.

Il Capitolo 4 riporta i risultati dell'analisi empirica e approfondisce l'evoluzione temporale dei principali determinanti della qualità percepita attraverso l'utilizzo delle tecniche descritte.

Infine, il Capitolo 5 presenta una discussione complessiva dei risultati ottenuti, evidenziando i principali contributi della ricerca, i limiti dell'approccio adottato e le possibili direzioni di sviluppo futuro.

1. La Voice of Customer

1.1 Introduzione alla Voice of Customer

Nell'attuale era digitale, i clienti condividono in maniera crescente le proprie esperienze, opinioni e giudizi sui prodotti e sui servizi attraverso piattaforme online come siti di recensioni, forum e social media. Questo insieme di contenuti generati dagli utenti viene definito **Voice of Customer (VoC)**, o in contesto digitale **digital VoC**, ed è riconducibile al più ampio concetto di **electronic Word of Mouth (eWOM)** ([Mastrogiacomo et al., 2021](#); [Verma and Yadav, 2021](#)).

La VoC digitale rappresenta quindi una fonte di dati preziosa per le organizzazioni, poiché riflette in tempo reale le **percezioni, le aspettative e i livelli di soddisfazione** dei consumatori ([Colicev et al., 2019](#); [Sykora et al., 2022](#)). Rispetto agli strumenti tradizionali di raccolta del feedback, come questionari, interviste e focus group, la Digital VoC consente di accedere a una quantità più ampia di informazioni, spesso disponibili a costi inferiori e generate direttamente dagli utenti durante o dopo l'esperienza di consumo ([Barravecchia et al., 2022](#); [Barravecchia et al., 2023b](#)). Un aspetto rilevante della VoC digitale è il suo carattere non strutturato. Le recensioni online, infatti, non seguono uno schema predefinito, ma contengono testi liberi nei quali i clienti descrivono aspetti differenti dell'esperienza vissuta. Questo permette alle imprese di intercettare bisogni, criticità e aspettative che potrebbero non emergere attraverso strumenti di indagine tradizionali. Allo stesso tempo, la natura eterogenea di tali dati rende necessaria l'applicazione di tecniche analitiche avanzate, capaci di trasformare grandi volumi di testo in informazioni interpretabili e utilizzabili nei processi decisionali.

In questa prospettiva, la VoC digitale non rappresenta soltanto un insieme di opinioni espresse dai clienti, ma una vera e propria fonte informativa per comprendere i determinanti della qualità percepita. L'analisi sistematica dei contenuti generati dagli utenti consente infatti di individuare quali attributi di un prodotto o servizio influenzano maggiormente la soddisfazione o l'insoddisfazione dei clienti ([Barravecchia et al., 2022](#)). Un utilizzo efficace di tali informazioni permette quindi alle imprese di allineare meglio la propria offerta ai bisogni del mercato, reagire tempestivamente a eventuali criticità emergenti e supportare il miglioramento continuo della qualità.

1.2 La Voice of Customer nel contesto della Quality 4.0

Con l'avvento della **Quality 4.0**, il ruolo della VoC digitale è diventato sempre più importante per i sistemi di gestione della qualità. La Quality 4.0, infatti, si basa sull'integrazione di **big data analytics, intelligenza artificiale e strumenti digitali** nei processi di controllo e miglioramento continuo ([Escobar et al., 2021](#); [Watson, 2019](#); [Zonnenshain and Kenett, 2020](#)).

In questo nuovo paradigma, la qualità non viene più monitorata con approcci tradizionali e statici, ma attraverso sistemi dinamici capaci di auto-monitorarsi e di prevedere i problemi prima che si manifestino ([Agrawal et al., 2024](#); [Antony et al., 2022](#)). La diffusione delle piattaforme digitali ha reso disponibili grandi quantità di contenuti generati dagli utenti, offrendo nuove opportunità per comprendere come la qualità di prodotti e servizi venga percepita nel tempo ([Barravecchia et al., 2023b](#)). In questo senso, la Digital VoC si inserisce pienamente nella logica della Quality 4.0, poiché permette di integrare la prospettiva del cliente nei processi di monitoraggio della qualità. La letteratura evidenzia come le tradizionali tecniche di quality tracking, basate su indagini periodiche, questionari o interviste, presentino alcuni limiti: possono richiedere tempo e risorse, coinvolgere un numero ridotto di clienti e risultare intrusive per gli utenti ([Barravecchia et al., 2023b](#)). Al contrario, la VoC digitale consente di analizzare feedback spontanei, continuamente aggiornati e provenienti da campioni più ampi di consumatori. Ciò permette alle organizzazioni di osservare non solo il livello complessivo di soddisfazione, ma anche l'evoluzione dei determinanti della qualità percepita.

In tale contesto, l'impiego di tecniche di text mining e topic modeling consente di estrarre dai testi non strutturati i principali temi discussi dai clienti, interpretandoli come attributi o determinanti della qualità. Questi strumenti permettono quindi di trasformare le recensioni online in indicatori utili per supportare decisioni manageriali e attività di miglioramento continuo ([Mastrogiacomo et al., 2021](#); [Barravecchia et al., 2022](#)).

La VoC digitale diventa quindi uno strumento strategico che permette alle aziende di rilevare in tempo reali **trend emergenti, anomalie e cambiamenti nei bisogni dei clienti**, favorendo un passaggio da una logica **reattiva** a una logica **proattiva** di gestione della qualità ([Broday, 2022](#); [Sony et al., 2020](#)).

L'integrazione tra Digital VoC e Quality 4.0 risulta particolarmente rilevante nei contesti di servizio, nei quali la qualità percepita dipende fortemente dall'esperienza vissuta dal cliente. Monitorare nel tempo le opinioni espresse online consente infatti di individuare segnali precoci di criticità, comprendere quali aspetti del servizio generano soddisfazione o insoddisfazione e orientare gli interventi di miglioramento verso gli elementi maggiormente rilevanti per l'utente finale.

1.3 Tecniche di analisi della Voice of Customer digitale

La grande quantità di dati non strutturati che caratterizza la VoC digitale richiede l'impiego di strumenti avanzati di analisi testuale e machine learning. In letteratura vengono identificate principalmente quattro categorie di tecniche:

- **Sentiment Analysis:**

La Sentiment Analysis è una tecnica di text mining finalizzata all'identificazione dell'orientamento emotivo espresso all'interno di un testo. L'obiettivo principale consiste nel classificare i contenuti generati dagli utenti in categorie quali sentiment positivo, negativo o neutro, consentendo alle organizzazioni di monitorare il livello di soddisfazione dei clienti e di individuare tempestivamente eventuali criticità ([Kim, 2021](#)). Le principali metodologie utilizzate comprendono approcci basati su dizionari lessicali e modelli di Machine Learning addestrati su dataset etichettati ([Barravecchia et al., 2025](#)). Sebbene tale tecnica permetta di ottenere rapidamente una visione sintetica delle percezioni degli utenti, presenta alcune limitazioni legate alla difficoltà di interpretare correttamente sarcasmo, ironia e sfumature contestuali del linguaggio naturale ([Liang and Wang, 2019](#); [Barravecchia et al., 2025](#)).

- **Topic Modeling:**

Un'ulteriore Tecnica di analisi per la DVoC è il Topic Modeling ([Mastrogiacomo et al., 2021](#), [Ozdagoglu et al., 2018](#)). Il Topic Modeling occupa un ruolo centrale poiché consente di estrarre automaticamente i principali argomenti discussi all'interno di grandi collezioni di documenti testuali ([Blei et al., 2003](#); [Roberts et al., 2014](#)). L'obiettivo principale di tali algoritmi è identificare strutture latenti presenti nei testi, senza richiedere una classificazione preventiva da parte dell'analista.

Il Topic Modeling appartiene all'ambito del text mining e può essere definito come una famiglia di metodi statistici finalizzati all'identificazione dei topic, ovvero insiemi di parole che tendono a comparire frequentemente insieme e che rappresentano uno specifico tema trattato nei documenti ([Blei, 2012](#)). Quando applicato alla Digital Voice of Customer, il Topic Modeling permette di individuare automaticamente i principali aspetti di un prodotto o servizio discussi dagli utenti, trasformando grandi quantità di recensioni non strutturate in informazioni sintetiche e interpretabili.

Il funzionamento generale di un algoritmo di Topic Modeling può essere descritto attraverso tre passaggi fondamentali. In primo luogo, viene analizzato un corpus di documenti costituito dalle recensioni raccolte online. Successivamente, l'algoritmo identifica un insieme di topic latenti caratterizzati da specifiche parole chiave che tendono a comparire frequentemente nello stesso contesto. Infine, per ogni documento viene stimata una distribuzione di probabilità che indica in quale misura ciascun topic è presente all'interno del testo ([Blei et al., 2003](#); [Mastrogiacomo et al., 2021](#)).

L'output prodotto da tali algoritmi è generalmente costituito da due elementi principali ([Roberts et al., 2016](#)). Il primo è il Topical Content, che descrive ciascun topic attraverso le parole maggiormente rappresentative. Il secondo è la Topical Prevalence, che indica il peso relativo di ogni topic all'interno di ciascuna recensione. In altre parole, una singola recensione può affrontare contemporaneamente più temi, ciascuno con una diversa intensità ([Barravecchia et al., 2023b](#)).

Secondo Barravecchia et al. (2023b), l'applicazione di un algoritmo di Topic Modeling alla Digital Voice of Customer permette di affrontare tre problemi fondamentali: identificare i topic presenti nel corpus documentale, associare a ciascun topic le parole chiave che lo caratterizzano e determinare la distribuzione dei topic all'interno di ogni recensione.

Tra gli algoritmi maggiormente utilizzati in letteratura si distinguono la Latent Dirichlet Allocation (LDA) e lo Structural Topic Model (STM). La LDA rappresenta il modello probabilistico più diffuso e assume che ogni documento sia costituito da una combinazione di topic latenti ([Blei et al., 2003](#)). Lo STM estende tale approccio consentendo di incorporare variabili aggiuntive, come il rating assegnato dall'utente, la data della recensione o altre informazioni contestuali, migliorando l'interpretabilità dei risultati e rendendolo particolarmente adatto all'analisi della Digital Voice of Customer ([Roberts et al., 2014](#)).

Nel presente lavoro viene adottato lo Structural Topic Model, poiché consente di integrare informazioni aggiuntive associate alle recensioni e di ottenere indicatori quantitativi utili per il monitoraggio temporale dei determinanti della qualità percepita.

- **Clustering:**

Il clustering rappresenta una tecnica di apprendimento non supervisionato finalizzata al raggruppamento automatico di documenti o recensioni caratterizzati da elevata similarità. Attraverso algoritmi quali K-Means e Hierarchical Clustering, le recensioni vengono suddivise in gruppi omogenei che consentono di individuare segmenti di clienti, problematiche ricorrenti o pattern comportamentali comuni ([Tabianan et al., 2022](#); [Nainggolan and Purba, 2020](#); [Nguyen et al., 2020](#)). Questa metodologia risulta particolarmente utile per supportare la segmentazione delle opinioni espresse dagli utenti e l'identificazione di categorie emergenti di feedback. Tuttavia, l'efficacia dei risultati dipende fortemente dalla dimensione del dataset disponibile e dalla scelta dell'algoritmo di clustering adottato ([Barravecchia et al., 2025](#)).

- **Deep Learning:**

Negli ultimi anni l'evoluzione dell'Intelligenza Artificiale ha favorito l'impiego di modelli di Deep Learning per l'analisi della Voice of Customer. Tra questi, particolare

rilevanza assume il modello BERT, in grado di comprendere il significato delle parole considerando il contesto in cui esse compaiono ([Durairaj and Chinnalagu, 2021](#); [Bilal and Almazroi, 2023](#)). A differenza degli approcci tradizionali, i modelli basati su reti neurali profonde consentono di cogliere relazioni semantiche complesse e di analizzare efficacemente grandi volumi di dati testuali. Tali tecniche offrono generalmente livelli di accuratezza elevati nell'identificazione di sentiment, argomenti e intenzioni degli utenti, ma richiedono notevoli risorse computazionali e grandi quantità di dati per l'addestramento dei modelli ([Mutinda et al., 2023](#); [Barravecchia et al., 2025](#)).

1.4 Indicatori della VoC: MTP, IMTP e MRP

Uno dei principali contributi della letteratura recente è lo sviluppo di metriche derivate dal Topic Modeling, utili per monitorare l'evoluzione della Voice of Customer digitale ([Barravecchia et al., 2023a, b](#)). In particolare, i risultati del Topic Modeling possono essere sintetizzati attraverso la frequenza con cui determinati topic vengono discussi dai clienti all'interno delle recensioni online. Tale frequenza viene comunemente misurata attraverso il Mean Topic Prevalence (MTP), ossia la prevalenza media di ciascun topic sull'intero dataset di recensioni analizzate ([Mastrogiacomo et al., 2021](#)).

L' MTP misura quindi la prominenza media di uno specifico topic all'interno dell'intero corpus di recensioni. Formalmente, esso è calcolato come:

$$MTP_t = \frac{\sum_{i=1}^N TP_{i,t}}{N}$$

dove N rappresenta il numero totale di recensioni analizzate, mentre $TP_{i,t}$ indica la prevalenza del topic (t) all'interno della recensione (i). La somma dei valori di MTP relativi a tutti i topic considerati è pari a 1, poiché ogni recensione distribuisce la propria prevalenza tra i diversi topic individuati.

Accanto all'indicatore MTP, la letteratura propone anche l'Interval Mean Topical Prevalence (IMTP), che estende il calcolo della prevalenza media dei topic a specifici intervalli temporali. A differenza dell'indicatore MTP, che restituisce una fotografia complessiva dei topic prevalenti nell'intero dataset, l'IMTP consente di osservare l'evoluzione dei topic nel tempo, trasformando la prevalenza tematica in una serie storica ([Barravecchia et al., 2023a, b](#)).

L'IMTP misura quindi la prevalenza media di un determinato topic (d) all'interno di un preciso periodo di osservazione (t), ad esempio una settimana, un mese o un altro intervallo temporale definito dal ricercatore. Formalmente, può essere espresso come:

$$IMTP_{t,n} = \frac{\sum_j^{R_t} TP_{t,t}}{|R_n|}$$

dove R_t rappresenta l'insieme delle recensioni appartenenti al periodo di campionamento (n), mentre $|R_n|$ indica la numerosità di tale insieme. Anche in questo caso, per ogni intervallo temporale, la somma degli IMTP relativi a tutti i topic è pari a 1.

L'analisi dell'IMTP risulta particolarmente utile per rilevare cambiamenti improvvisi o graduali nelle percezioni dei clienti, poiché permette di osservare come l'importanza relativa dei diversi topic vari nel tempo. In questo senso, l'IMTP può supportare l'identificazione di segnali precoci di criticità, anomalie o cambiamenti nelle preferenze dei clienti, fornendo una base quantitativa per il monitoraggio dinamico della qualità percepita ([Barravecchia et al., 2023a, b](#)).

Per chiarire il funzionamento degli indicatori MTP e IMTP, ([Barravecchia et al. 2023a](#)) propongono un esempio semplificato basato su dieci recensioni online distribuite in tre periodi temporali e riferite a tre determinanti fittizie della qualità (A, B e C). Le Figure 1 e 2 riportano rispettivamente il calcolo del MTP e dell'IMTP.

Digital VoC record (j)	Date	Rating (k)	Sampling period (t)	Topical prevalence		
				Quality determinant A (TP_{j1})	Quality determinant B (TP_{j2})	Quality determinant C (TP_{j3})
1	3 January 2022	1	1	0.5	0.3	0.2
2	15 January 2022	4		0.1	0.7	0.2
3	17 January 2022	2		0.8	0.15	0.05
4	11 February 2022	5	2	0.7	0.25	0.05
5	16 February 2022	3		0.1	0.7	0.2
6	18 February 2022	1		0.55	0.1	0.35
7	9 March 2022	4	3	0.65	0.15	0.2
8	13 March 2022	2		0.2	0.1	0.7
9	15 March 2022	5		0.8	0.15	0.05
10	22 March 2022	3		0.1	0.7	0.2
MTP_d				0.45	0.33	0.22

Figura 1 Esempio: risultato dell'algoritmo di Topic Modeling e calcolo dell'MTP per tre determinanti di qualità. Fonte: Barravecchia et al., 2025.

Calcolando il MTP sull'intero dataset come mostrato in Figura 1, si ottiene una misura complessiva della rilevanza media dei tre topic. Nell'esempio riportato dalla letteratura, i valori medi risultano pari a:

$$MTP_A = 0,45$$

$$MTPB = 0,33$$

$$MTPC = 0,22$$

Questi risultati indicano che, considerando tutte le recensioni analizzate, il topic A è mediamente il più rilevante, seguito dal topic B e infine dal topic C.

Digital VoC record (j)	Date	Rating (k)	Sampling period (t)	Topical prevalence		
				Quality determinant A ($TP_{j,1}$)	Quality determinant B ($TP_{j,2}$)	Quality determinant C ($TP_{j,3}$)
1	3 January 2022	1	1	0.5	0.3	0.2
2	15 January 2022	4		0.1	0.7	0.2
3	17 January 2022	2		0.8	0.15	0.05
			$IMTP_{d,1}$	0.47	0.38	0.15
4	11 February 2022	5	2	0.7	0.25	0.05
5	16 February 2022	3		0.1	0.7	0.2
6	18 February 2022	1		0.55	0.1	0.35
			$IMTP_{d,2}$	0.45	0.35	0.20
7	9 March 2022	4	3	0.65	0.15	0.2
8	13 March 2022	2		0.2	0.1	0.7
9	15 March 2022	5		0.8	0.15	0.05
10	22 March 2022	3		0.1	0.7	0.2
			$IMTP_{d,3}$	0.44	0.28	0.29

Figura 2 Esempio: risultato dell' algoritmo di Topic Modeling e calcolo dell'IMTP per tre determinanti di qualità. Fonte: Barravecchia et al., 2025.

Tuttavia, il solo MTP non consente di comprendere se la rilevanza dei topic cambi nel tempo. Per questo motivo viene calcolato l'IMTP, come viene mostrato in Figura 2, che permette di osservare la prevalenza media dei topic in ciascun periodo di campionamento. Nell'esempio, i valori risultano:

Periodo 1:

$$IMTPA,1 = 0,47$$

$$IMTPB,1 = 0,38$$

$$IMTPC,1 = 0,15$$

Periodo 2:

$$IMTPA,2 = 0,45$$

$$IMTPB,2 = 0,35$$

$$IMTPC,2 = 0,20$$

Periodo 3:

$$\text{IMTPA},3 = 0,44$$

$$\text{IMTPB},3 = 0,28$$

$$\text{IMTPC},3 = 0,29$$

L'esempio mostra che, pur rimanendo il topic A il più rilevante in tutti i periodi, la sua prevalenza tende leggermente a diminuire nel tempo. Il topic B registra anch'esso una riduzione, mentre il topic C aumenta progressivamente, passando da 0,15 nel primo periodo a 0,29 nel terzo. Questo tipo di analisi consente quindi di cogliere dinamiche temporali che il solo MTP complessivo non sarebbe in grado di evidenziare.

Per integrare l'analisi dei topic con una misura esplicita della valutazione espressa dai clienti, può essere utilizzato l'indicatore Mean Rating Proportion (MRP) ([Barravecchia et al., 2022](#)). L'MRP utilizza il valore di valutazione oggettivo generalmente assegnato dai clienti alle proprie recensioni, spesso su una scala da 1 a 5 ([Barravecchia et al., 2022](#)). L'MRP quantifica la prevalenza di un topic all'interno delle recensioni associate a uno specifico punteggio, calcolata mediante la seguente formula:

$$\text{MRP}_{t,k} = \frac{\sum_{i \in R_k} \text{TP}_{i,t}}{|R_k|}$$

dove (t) rappresenta il topic; (k) indica il livello di valutazione assegnato dall'utente, ad es. su una scala da 1 a 5; R_k il sottoinsieme delle recensioni associate a un livello di valutazione pari a k; $\text{TP}_{i,t}$ la prevalenza del topic d nella recensione i-esima; $|R_k|$ indica il numero di recensioni con rating pari a (k). L'MRP consente quindi di collegare i topic individuati dal modello alle valutazioni espresse dagli utenti, permettendo di comprendere quali aspetti del servizio siano maggiormente associati a giudizi positivi o negativi.

1.5 Le anomalie nella VoC digitale

L'analisi della **digital Voice of Customer** non si limita a una rappresentazione statica dei temi discussi dai clienti, ma deve considerare la dimensione temporale dei dati e le tecniche di analisi delle serie storiche, nell'ambito delle quali assume particolare rilevanza l'identificazione delle anomalie. L'indicatore **IMTP**, infatti, assume la forma di una **serie storica** che descrive l'evoluzione della rilevanza dei diversi topic nel tempo. Proprio in questa prospettiva diventa essenziale l'individuazione delle **anomalie**, ossia deviazioni significative dai comportamenti attesi ([Hamilton, 2020](#); [Esling and Agon, 2012](#)).

Le anomalie non rappresentano semplici rumori casuali, ma spesso costituiscono **segnali di cambiamenti importanti** nel sistema: un improvviso aumento delle lamentele su un aspetto del servizio, oppure un calo inatteso di interesse per una caratteristica di prodotto, possono anticipare problematiche di qualità o mutamenti nelle preferenze dei consumatori ([Hawkins, 1984](#); [Aggarwal, 2017](#)). Secondo ([Barravecchia et al., 2025](#)), sulla base della letteratura sullo studio delle serie temporali, le anomalie solitamente ricadono in due principali caratteristiche ([Box-Steffensmeier et al., 2014](#)):

- Noise or erroneous data: punti derivanti da errori di raccolta o di elaborazione, che non riflettono trend significativi
- Indicator of unusual phenomena: punti che segnalano deviazioni significative dovute a cambiamenti nel sistema sottostante, come insoddisfazione dei clienti o guasti/problemi del prodotto.

L'identificazione di tali anomalie richiede tecniche di analisi avanzate in grado di distinguere tra variazioni casuali e deviazioni significative. Dalla letteratura sull'analisi delle serie temporali ([Aggarwal, 2017](#)), le anomalie possono essere racchiuse in quattro categorie chiave:

- Point anomaly: punto isolato o una piccola sequenza di punti che si distacca dal range atteso
- Contextual anomaly: punto anomalo rispetto uno specifico contesto
- Collective anomaly: un gruppo di punti che rappresenta un pattern anomalo rispetto alla norma
- Other anomaly types: Ulteriori categorie specializzate che catturano variazioni più granulari, basate su specifici pattern presenti nel dataset.

Sulla base dei principi della **time series analysis** e adattandoli alle peculiarità della VoC digitale, ([Barravecchia et al. 2025](#)) propongono una tassonomia che distingue quattro categorie principali di anomalie:

1. **Spike anomalies** – deviazioni improvvise (picchi verso l'alto o verso il basso).
2. **Level anomalies** – cambiamenti duraturi e stabili del livello medio di attenzione su un tema.
3. **Trend anomalies** – variazioni gradualmente e persistenti che modificano la direzione della serie.
4. **Seasonal anomalies** – rottura di pattern stagionali attesi, con intensità o periodicità anomala.

L'identificazione di tali anomalie consente alle aziende di distinguere tra eventi temporanei e cambiamenti strutturali nelle percezioni dei clienti, guidando le decisioni di miglioramento continuo, nei paragrafi successivi viene fornito un approfondimento sulle principali tipologie di anomalie.

1.5.1 Spike Anomalies

Le **Spike Anomalies** rappresentano deviazioni improvvise e localizzate rispetto all'andamento atteso della serie IMTP. Si manifestano come picchi isolati verso l'alto (*upward spike*) oppure verso il basso (*downward spike*) e sono generalmente associati a eventi specifici o inattesi che producono un cambiamento brusco nella discussione relativa a un determinato topic ([Barravecchia et al., 2025](#)).

A differenza delle variazioni graduali, le spike anomalies hanno carattere temporaneo e tendono a distinguersi chiaramente dalla baseline della serie storica. Un improvviso aumento delle recensioni negative su un servizio, ad esempio, può generare un upward spike, mentre un rapido calo di attenzione verso un tema può produrre un downward spike. **Come mostrato in Figura 3**, tali anomalie possono manifestarsi sia in presenza di trend stabili sia all'interno di trend già crescenti o decrescenti.

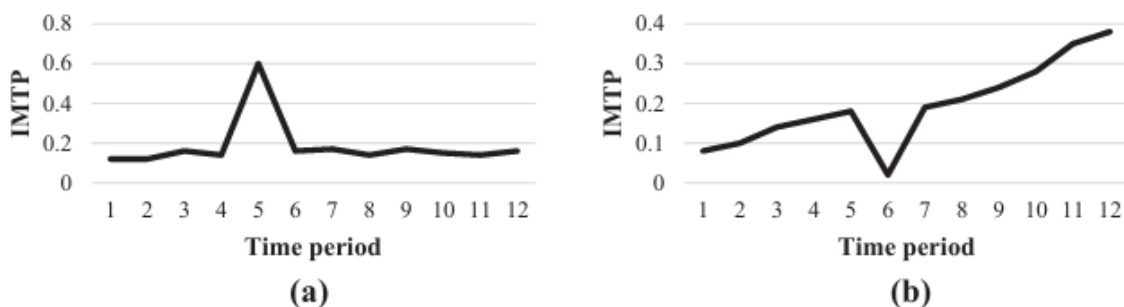


Figura 3 Esempio: Spike Anomalies nel Grafico dell'indicatore IMTP. Picchi verso l'alto(a) picchi verso il basso (b). Fonte: Barravecchia et al., 2025.

1.5.2 Level Anomalies

Le Level Anomalies si verificano quando la serie IMTP subisce un cambiamento improvviso e persistente del proprio livello medio. In questo caso, dopo una discontinuità iniziale, la serie tende a stabilizzarsi attorno a un nuovo valore, segnalando un possibile cambiamento strutturale nella percezione dei clienti ([Barravecchia et al., 2025](#)).

La letteratura distingue due principali configurazioni:

- upward level anomaly, caratterizzata da un incremento stabile del livello della serie;
- downward level anomaly, associata invece a una riduzione persistente della rilevanza attribuita a un topic.

Queste anomalie sono particolarmente rilevanti perché suggeriscono trasformazioni durature nella Voice of Customer, più che semplici eventi episodici. Come evidenziato in Figura 4, dopo il cambiamento iniziale la serie si stabilizza su un nuovo livello di equilibrio.

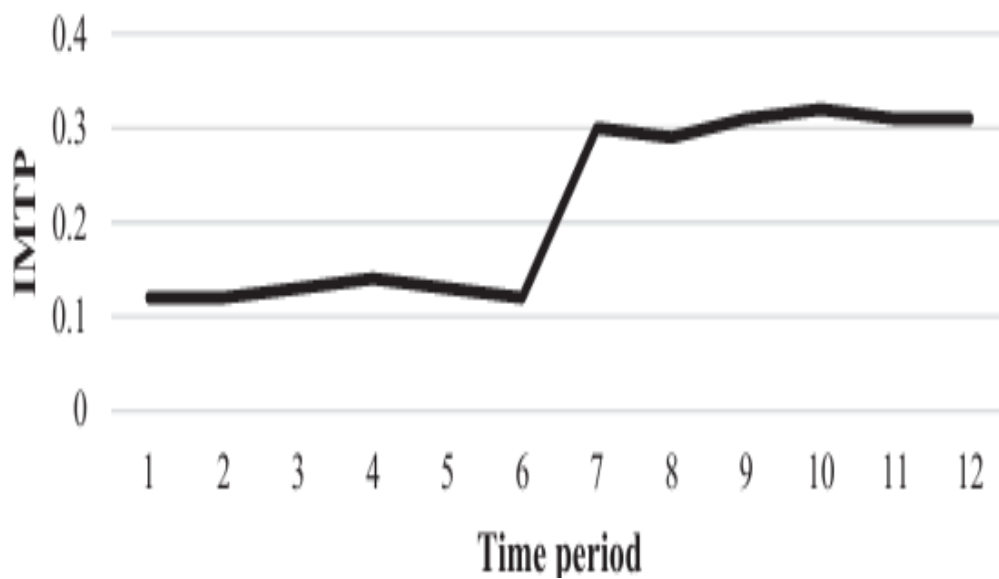


Figura 4 Esempio: Level Anomalies nel grafico IMTP. Fonte: Barravecchia et al., 2025.

1.5.3 Trend Anomalies

Le **Trend Anomalies** descrivono variazioni graduali ma persistenti nella direzione della serie IMTP. Diversamente dalle spike anomalies, non si manifestano attraverso un singolo picco improvviso, bensì tramite una progressiva modifica del comportamento della serie nel tempo (Barravecchia et al., 2025).

Queste anomalie indicano spesso cambiamenti più profondi nelle preferenze o nelle percezioni dei clienti. Secondo la tassonomia proposta da (Barravecchia et al., 2025) le trend anomalies possono assumere diverse configurazioni, tra cui:

- **Plateau**, ossia passaggio da trend crescente a stabilizzazione;
- **Escalation**, con accelerazione crescente della serie;
- **Decline**, caratterizzato da una progressiva riduzione;
- **Inversion**, che rappresenta un cambio di direzione del trend;
- **Stabilization**, in cui un trend decrescente evolve verso una fase stabile.

Come mostrato in **Figura 5**, le trend anomalies possono assumere configurazioni differenti a seconda della direzione iniziale e finale del trend osservato.

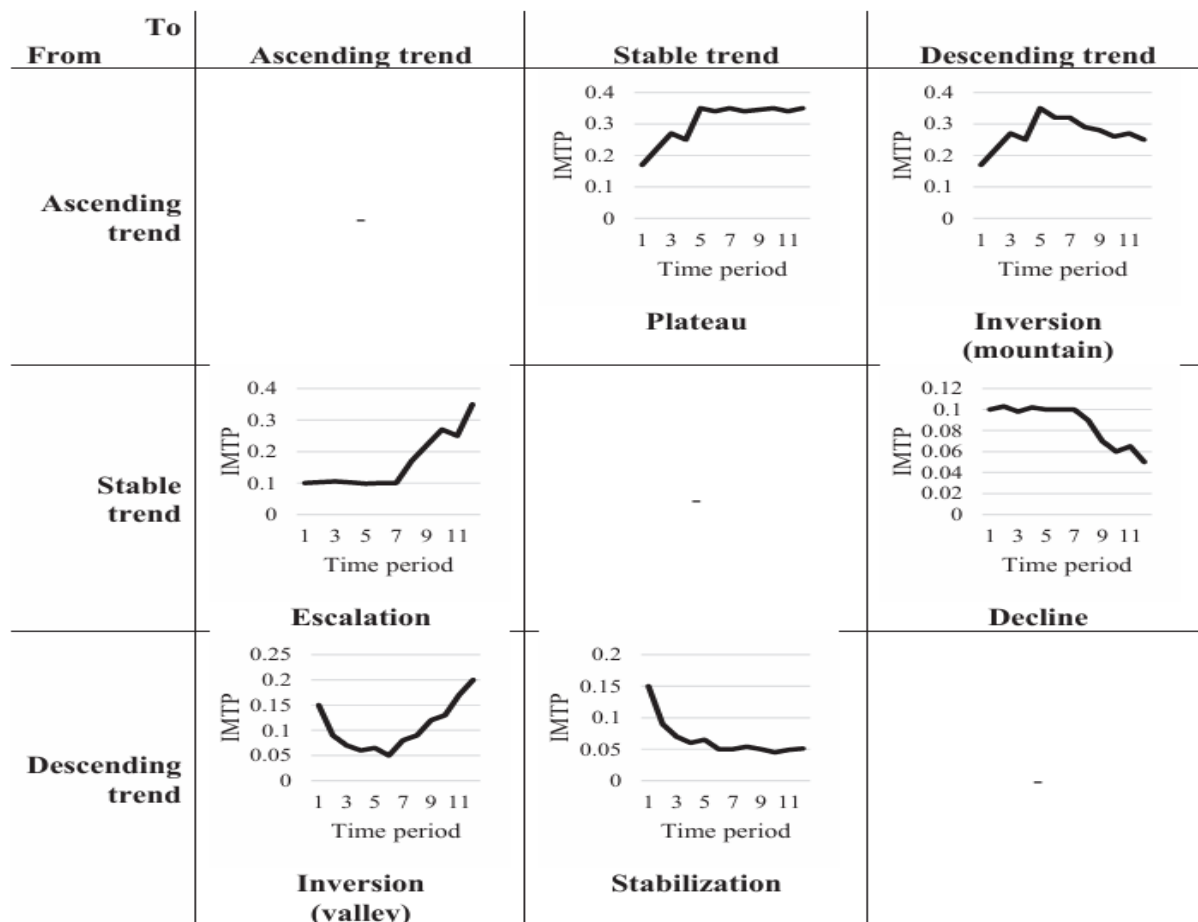


Figura 5 Esempi: Trend Anomalies dell'Indicatore IMTP. Fonte: Barravecchia et al., 2025.

1.5.4 Seasonal Anomalies

Le Seasonal Anomalies riguardano alterazioni dei pattern stagionali presenti nella serie IMTP. In condizioni normali, alcuni topic mostrano oscillazioni periodiche prevedibili, legate ad esempio a stagionalità di mercato, festività o abitudini ricorrenti dei consumatori. L'anomalia emerge quando tali pattern vengono modificati in termini di intensità, periodicità o continuità ([Barravecchia et al., 2025](#)).

Tra le principali configurazioni individuate in letteratura vi sono:

- variazioni nell'intensità della stagionalità;
- modifiche nella periodicità dei pattern;
- comparsa o scomparsa di trend stagionali consolidati.

L'analisi delle Seasonal Anomalies permette quindi di intercettare cambiamenti nel comportamento dei clienti che potrebbero riflettere trasformazioni del contesto competitivo o delle preferenze di consumo. Come illustrato in Figura 6, tali anomalie possono manifestarsi attraverso una riduzione dell'intensità stagionale, una variazione della periodicità oppure un progressivo affievolimento del pattern ricorrente.

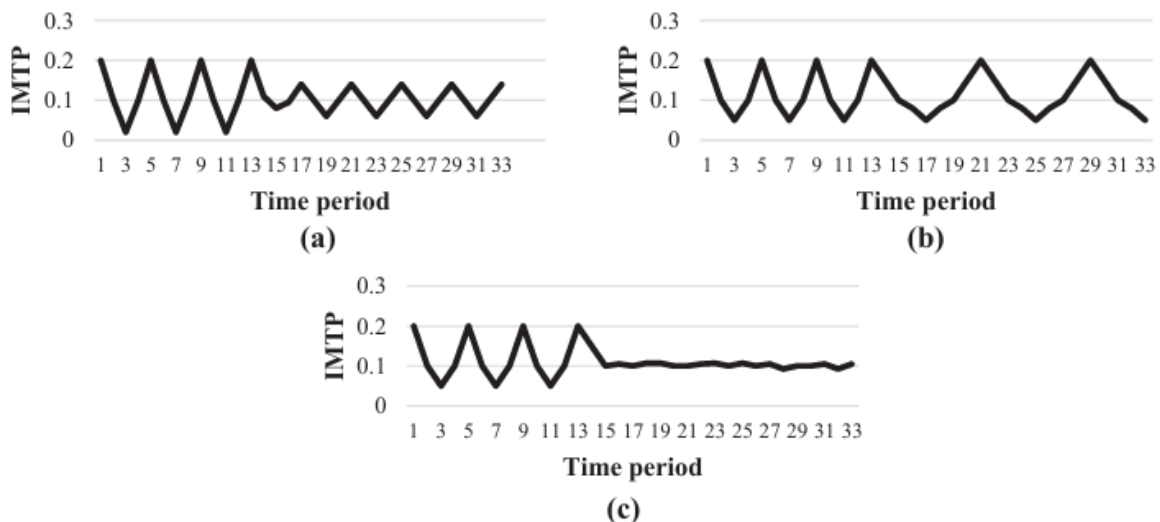


Figura 6 Esempi di Seasonal Anomalies su Indicatore IMTP. Fonte: Barravecchia et al., 2025.

1.6 Il rilevamento delle anomalie attraverso le carte di controllo

Per quanto concerne l'analisi di questa tesi, ovvero lo studio dei dati VoC relativi ai parchi Disney, in particolare in California, Hong-Kong e Parigi, l'individuazione di tali anomalie assume un ruolo strategico per comprendere l'evoluzione della soddisfazione dei visitatori nei diversi contesti geografici e temporali. Ogni parco infatti può essere soggetto a eventi locali – come festività – che possono influenzare la rilevanza dei vari Topic estratti dal modello STM. Tali deviazioni rappresentano indicatori di discontinuità e possono essere legate a eventi operativi, problematiche di servizio o variazioni nella percezione del marchio e delle attrazioni. Per il monitoraggio della stabilità del processo rappresentato dal campione UGC, per questa tesi, introduciamo un approccio mutuato dal Statistical Process Control (SPC). Le carte di controllo permettono di analizzare la variabilità dei dati nel tempo e di distinguere tra fluttuazioni casuali (*common cause variation*) e scostamenti significativi attribuibili a cause specifiche (*special cause variation*) ([Montgomery, 2009](#)).

1.7 Introduzione alle carte di controllo

L'introduzione delle carte di controllo nel contesto della Digital Voice of Customer rappresenta un'estensione del tradizionale approccio di Statistical Process Control (SPC) al dominio dei dati testuali e digitali, in linea con i principi della Quality 4.0, che promuove l'integrazione di strumenti analitici avanzati per il monitoraggio e il miglioramento continuo dei processi ([Sader et al., 2021](#)). Originariamente sviluppate da Walter A. Shewhart presso i Bell Laboratories negli anni Venti del secolo scorso, le carte di controllo nascono con l'obiettivo di monitorare la stabilità dei processi produttivi, distinguendo la variabilità dovuta a cause comuni (*common causes*) da quella derivante da cause speciali (*special causes*) ([Montgomery, 2009](#)).

1.8 Struttura e interpretazione delle carte di controllo

Una carta di controllo si compone di un asse temporale (o sequenziale) e di un asse della misurazione che rappresenta il valore dell'indicatore monitorato. A partire dai dati raccolti, e a seconda del tipo di carta, si procede con il calcolo dei tre limiti fondamentali:

- La linea centrale (CL), che corrisponde al valor medio del processo;
- Il limite di controllo superiore (UCL)
- Il limite di controllo inferiore (LCL)

I limiti che sono stati calcolati vanno a definire un intervallo entro il quale la variabilità è considerata accettabile. Nel momento in cui un punto supera questi limiti o mostra pattern non casuali (ad esempio più punti tutti sopra la media), il processo viene considerato fuori controllo e si richiede un'analisi delle cause che hanno generato tale deviazione ([Montgomery, 2009](#)).

Utilizzando grandi quantità di dati, come nel caso dei flussi di recensioni digitali, le carte di controllo permettono di osservare le anomalie in maniera tempestiva, offrendo un vantaggio competitivo nella gestione della qualità percepita ([Sader et al., 2021](#)). L'utilizzo delle carte di controllo è stato scelto anche per la loro capacità di fornire una rappresentazione visiva immediata della variabilità del processo, facilitando l'identificazione di comportamenti anomali e di possibili condizioni di fuori controllo."

1.9 Tipologie di carte di controllo

Esistono vari tipi di carte di controllo, nel dettaglio:

- Carta X-R: utilizzata per monitorare media e intervallo di sottogruppi di piccole dimensioni;
- Carta X-S: simile alla precedente ma basata sulla deviazione standard anziché sull'intervallo; scelta per questo caso studio perché fornisce una stima più accurata della variabilità del processo attraverso la deviazione standard ed è particolarmente appropriata per sottogruppi di dimensione medio-grande e variabili;
- Carta p e np: applicate a proporzioni o conteggi di non conformità;
- Carta c e u: dedicate al controllo di conteggi di difetti per unità di misura o di tempo;

Nel contesto della Digital Voice of Customer, dove gli indicatori derivano da misure continue (come valori medi di prevalenza o sentiment), le carte X-S risultano particolarmente adatte al monitoraggio temporale di indicatori continui derivati dalla Voice of Customer, come la prevalenza media dei topic o altri indicatori sintetici di percezione dei clienti ([Montgomery, 2009](#)). Di seguito un approfondimento per ciascuna di esse.

1.9.1 Carta $\bar{X}$ -R (Media e Range)

La carta $\bar{X}$ -R è una delle carte di controllo più utilizzate nello Statistical Process Control (SPC) per monitorare processi caratterizzati da sottogruppi di piccole dimensioni ([Montgomery, 2009](#)). Essa si compone di due grafici distinti ma complementari: la carta delle medie ($\bar{X}$), che controlla la variazione della media del processo, e la carta dei range (R), che monitora la variabilità interna ai sottogruppi. Nel contesto della Digital Voice of Customer, questa tipologia di carta può essere utilizzata per osservare l'andamento medio di un indicatore – ad esempio il sentiment o la prevalenza di un topic – su intervalli temporali limitati, verificando contemporaneamente se la variabilità interna rimane stabile. Come mostrato in Figura 7, eventuali punti al di fuori dei limiti di controllo rappresentano segnali di possibili anomalie e richiedono un'analisi approfondita delle cause ([Montgomery, 2009](#)).

Formule della carta $\bar{X}$ -R

Per ogni sottogruppo i:

Media campionaria:

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{i,j}$$

Range del sottogruppo:

$$R_i = X_{max} - X_{min}$$

Il Range medio e linea centrale della carta R:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m}$$

Linea centrale della carta delle medie:

$$CL_{\bar{X}} = \bar{\bar{X}}$$

Limiti di controllo:

$$\begin{aligned} UCL_{\bar{X}} &= \bar{\bar{X}} + A_2 \bar{R} \\ LCL_{\bar{X}} &= \bar{\bar{X}} - A_2 \bar{R} \\ UCL_R &= D_4 \bar{R} \\ LCL_R &= D_3 \bar{R} \end{aligned}$$

Dove A_2, D_3, D_4 sono coefficienti tabellati dipendenti dalla numerosità del sottogruppo ([Montgomery, 2009](#)).

Montgomery (2009) illustra l'utilizzo delle carte $\bar{X}$ -R per il monitoraggio della stabilità dei processi attraverso l'analisi congiunta della media e della variabilità dei sottogruppi. Come mostrato in Figura 7, la carta delle medie consente di identificare variazioni anomale della performance del processo, mentre la carta R permette di verificare la stabilità della variabilità interna ai campioni. Nel grafico è possibile osservare la presenza della linea centrale (CL) e dei limiti di controllo superiore e inferiore (UCL e LCL), utilizzati per individuare eventuali condizioni fuori controllo.

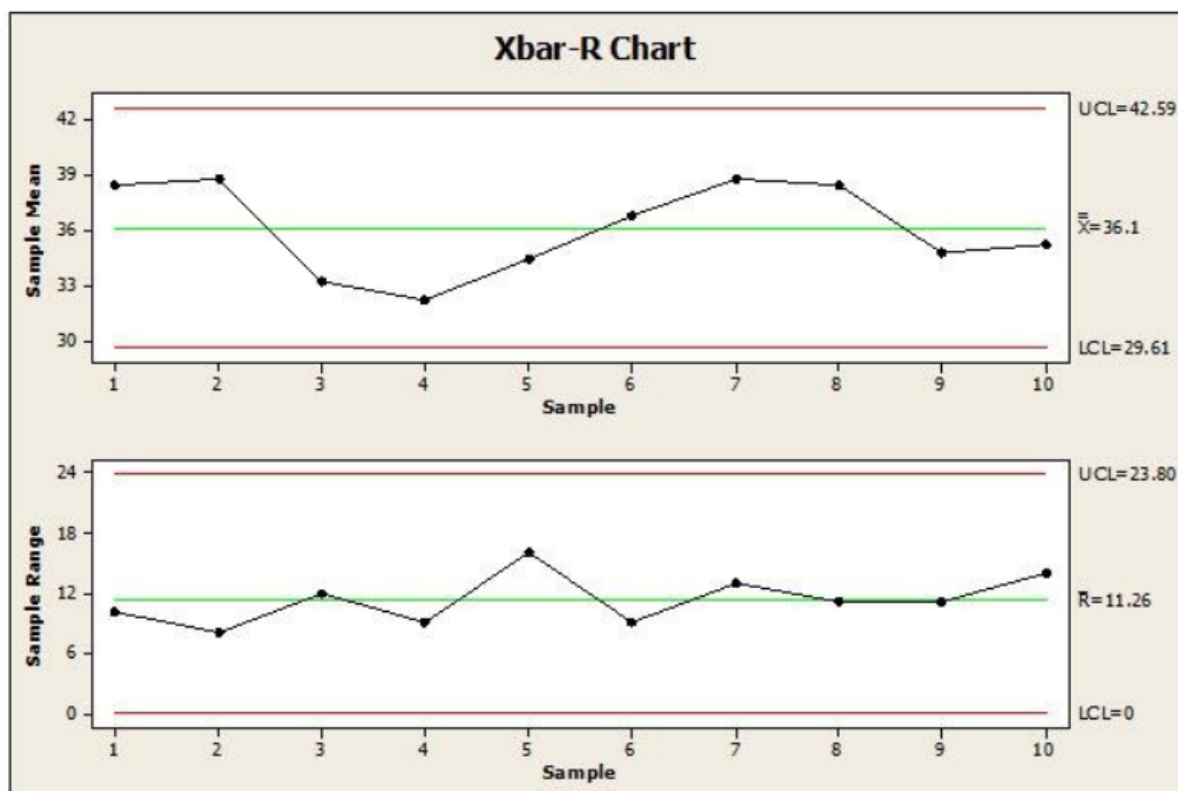


Figura 7 Esempio di carta $\bar{X}$ -R per il monitoraggio della stabilità di un processo produttivo. Fonte: adattato da Montgomery, 2009.

1.9.2 Carta $\bar{X}$ -S (Media e Deviazione Standard)

La carta $\bar{X}$ -S rappresenta un'evoluzione della carta $\bar{X}$ -R ed è particolarmente indicata quando i sottogruppi presentano dimensione medio-grande, poiché la deviazione standard fornisce una stima della variabilità più accurata rispetto al range. ([Montgomery, 2009](#)). Nel contesto della Digital Voice of Customer, questa carta risulta particolarmente utile per monitorare indicatori continui derivanti dalle recensioni online, come la prevalenza dei topic estratti tramite STM o gli score di sentiment. L'utilizzo della deviazione standard permette di descrivere meglio la variabilità degli indicatori nel tempo, aspetto particolarmente rilevante nell'analisi della VoC.

Formule della carta $\bar{X}$ -S:

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{i,j}$$

Deviazione standard:

$$S_i = \sqrt{\frac{\sum_{j=1}^n (X_{i,j} - \bar{X}_i)^2}{n - 1}}$$

Limiti delle carte:

$$UCL_{\bar{X}} = \bar{\bar{X}} + A_3 \bar{S}$$

$$LCL_{\bar{X}} = \bar{\bar{X}} - A_3 \bar{S}$$

$$UCL_S = B_4 \bar{S}$$

$$LCL_S = B_3 \bar{S}$$

Dove A_3, B_3, B_4 sono coefficienti dipendenti dalla numerosità campionaria.

[Montgomery \(2009\)](#) evidenzia come le carte $\bar{X}$ -S siano particolarmente adatte in presenza di sottogruppi di dimensioni variabili o numericamente più elevati, poiché la deviazione standard fornisce una misura della dispersione più robusta rispetto al range. Come illustrato in Figura 8, il superamento dei limiti di controllo nella carta X-S segnala un incremento anomalo della variabilità del processo.

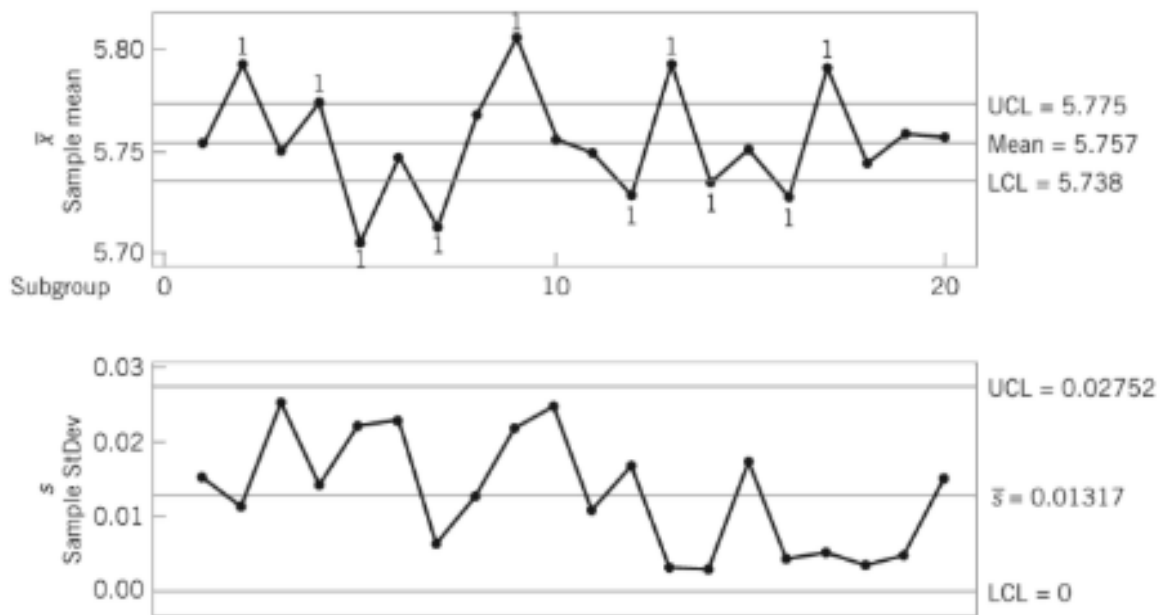


Figura 8 Esempio di carta $\bar{X}$ -S utilizzata per il monitoraggio della variabilità di processo. Fonte: adattato da Montgomery, 2009.

1.9.3 Carte p e np

Le carte p e np vengono utilizzate per monitorare processi basati su dati attributivi, ossia espressi tramite proporzioni o conteggi di unità non conformi (Montgomery, 2009). La carta p controlla la proporzione di elementi difettosi rispetto al totale osservato, mentre la carta np monitora direttamente il numero assoluto di non conformità. Nel contesto della Digital Voice of Customer, tali carte possono essere applicate per monitorare la percentuale di recensioni negative o il numero di commenti contenenti determinate problematiche, come lamentele sul servizio o sulla pulizia.

Formule della carta p: (dove p rappresenta la percentuale media dei difettosi nel campione di riferimento)

$$p_i = \frac{di}{ni}$$

Linea centrale

$$CL = \bar{p}$$

Limiti di controllo

$$UCL = \bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

Formule della carta np:

$$np_i = n\bar{p}$$

Limiti di controllo:

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1 - \bar{p})}$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1 - \bar{p})}$$

Le carte p e np vengono frequentemente utilizzate nel controllo della qualità per monitorare proporzioni di prodotti difettosi o numeri di non conformità. ([Montgomery, 2009](#)) riporta esempi applicativi in ambito manifatturiero nei quali il superamento dei limiti di controllo indica la presenza di cause speciali responsabili dell'aumento delle unità difettose. Come mostrato in Figura 9, la carta np permette di osservare il numero assoluto di elementi difettosi presenti nei campioni.

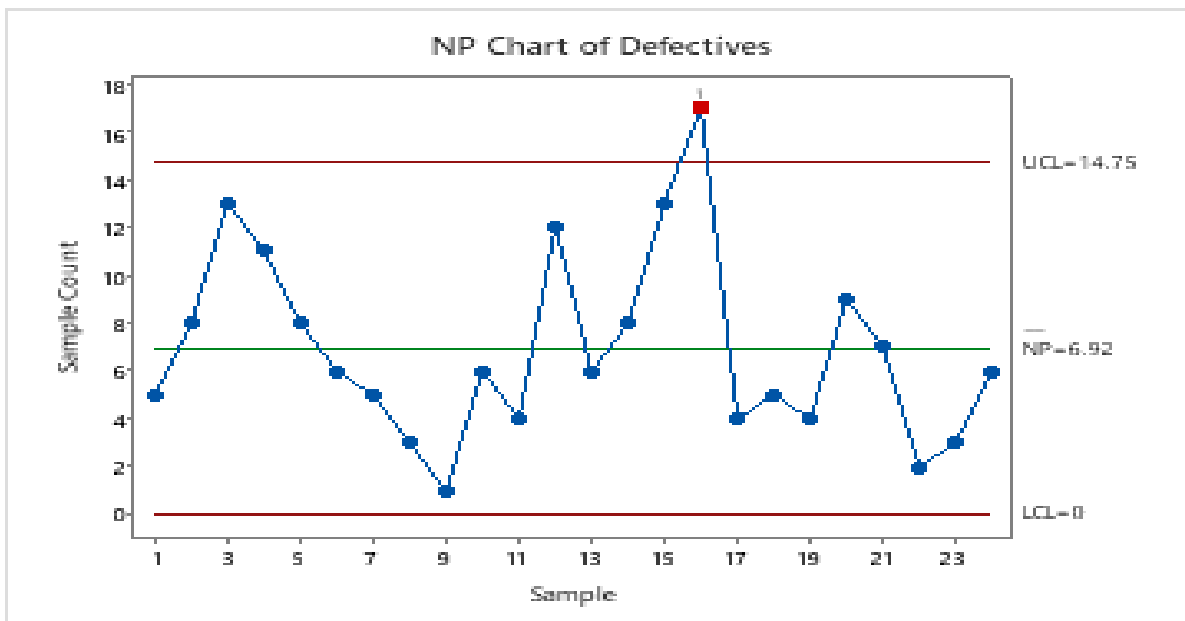


Figura 9 Esempio di carta np. Fonte: adattato da Minitab Support, "Example of NP Chart"

Mentre nella Figura 10, la carta p rappresenta l'andamento temporale della proporzione di unità non conformi rispetto ai limiti di controllo.

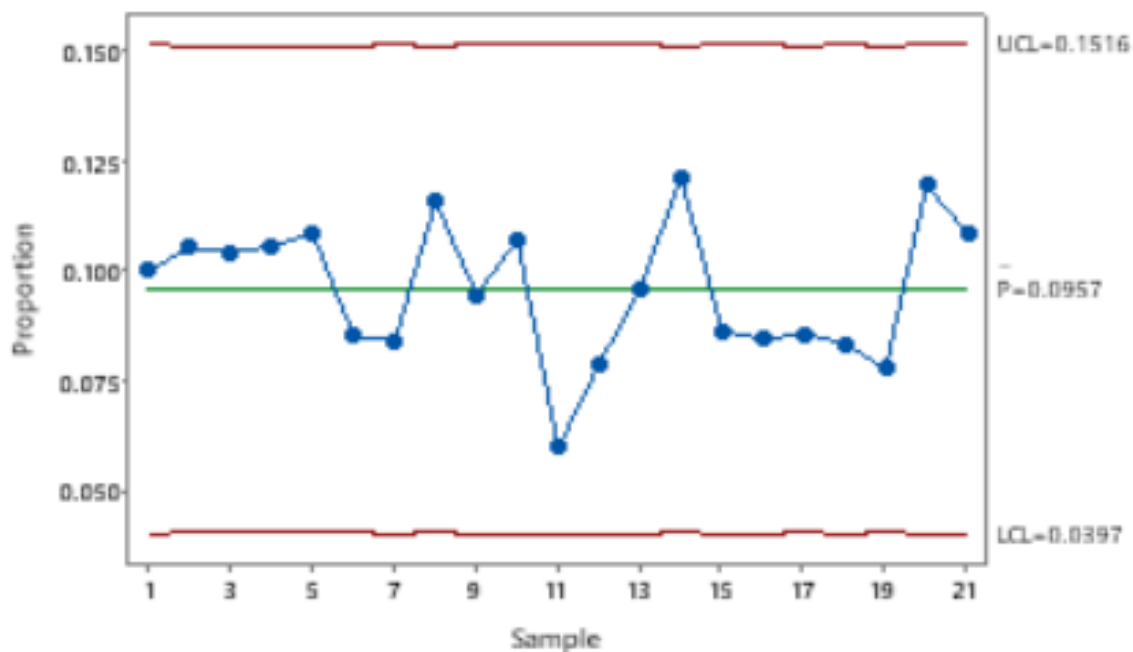


Figura 10 Esempio di carta p. Fonte: adattato da Minitab Support “Example of P Chart”

Sebbene le carte p e np possano essere adattate all’analisi di specifiche caratteristiche della Voice of Customer espresse in forma attributiva, nel presente lavoro si è scelto di utilizzare carte per variabili, in quanto gli indicatori analizzati assumono natura continua.

1.9.4 Carte c e u

Le carte c e u sono dedicate al monitoraggio del numero di difetti osservati in un processo. La carta c viene utilizzata quando l’unità di osservazione è costante, mentre la carta u risulta più adatta quando il numero di unità osservate varia nel tempo ([Montgomery, 2009](#)). Nel caso della Digital Voice of Customer, queste carte possono essere applicate per monitorare il numero di problematiche specifiche segnalate nelle recensioni, ad esempio reclami relativi ai tempi di attesa o ai servizi di ristorazione.

Formule della carta c

Numero medio di difetti (c):

$$CL = \bar{c}$$

Limiti di controllo:

$$UCL = \bar{c} + 3\sqrt{\bar{c}}$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}}$$

Formule della carta (u) numero medio osservato di non conformi per unità in un insieme preliminare di dati:

$$u_i = \frac{ci}{ni}$$

Linea centrale:

$$CL = \bar{u}$$

Limiti di controllo:

$$UCL = \bar{u} + 3 \sqrt{\frac{\bar{u}}{ni}}$$

$$LCL = \bar{u} - 3 \sqrt{\frac{\bar{u}}{ni}}$$

Le carte c e u sono utilizzate per monitorare il numero di difetti osservati all'interno di un processo produttivo o di servizio. Secondo Montgomery (2009), tali strumenti risultano particolarmente efficaci nell'identificazione di anomalie legata all'aumento improvviso dei difetti.

Come illustrato in Figura 11, la carta c monitora il numero totale di difetti rilevati in unità di osservazione costanti. La carta u, rappresentata in Figura 12, consente invece di controllare il numero medio di difetti per unità anche in presenza di campioni di dimensione variabile.

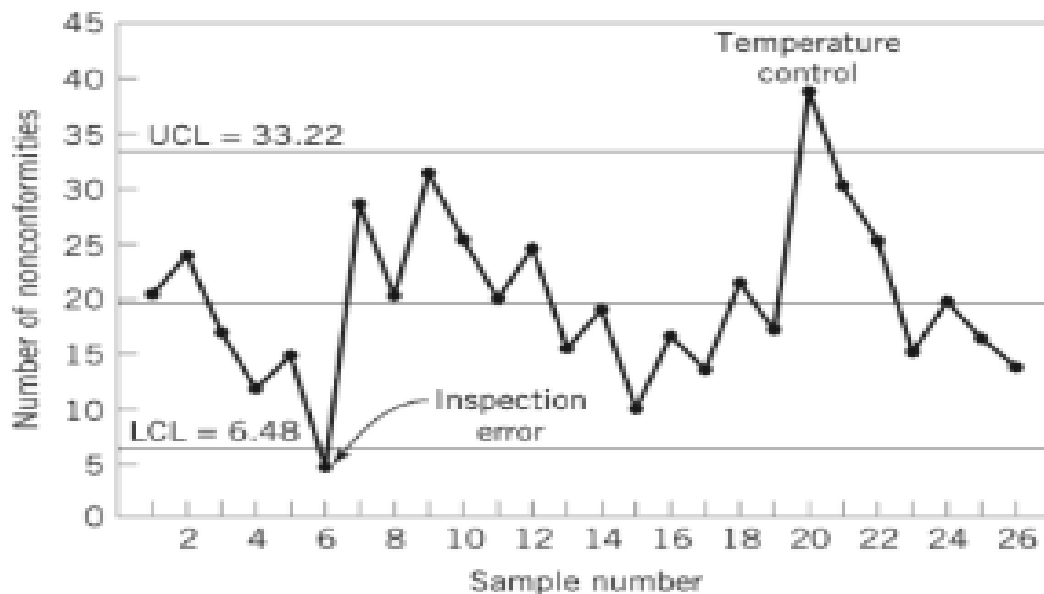


Figura 11 Esempio: Carta C esempio non conformi in un processo di circuiti stampati. Fonte: adattata da Montgomery, 2009.

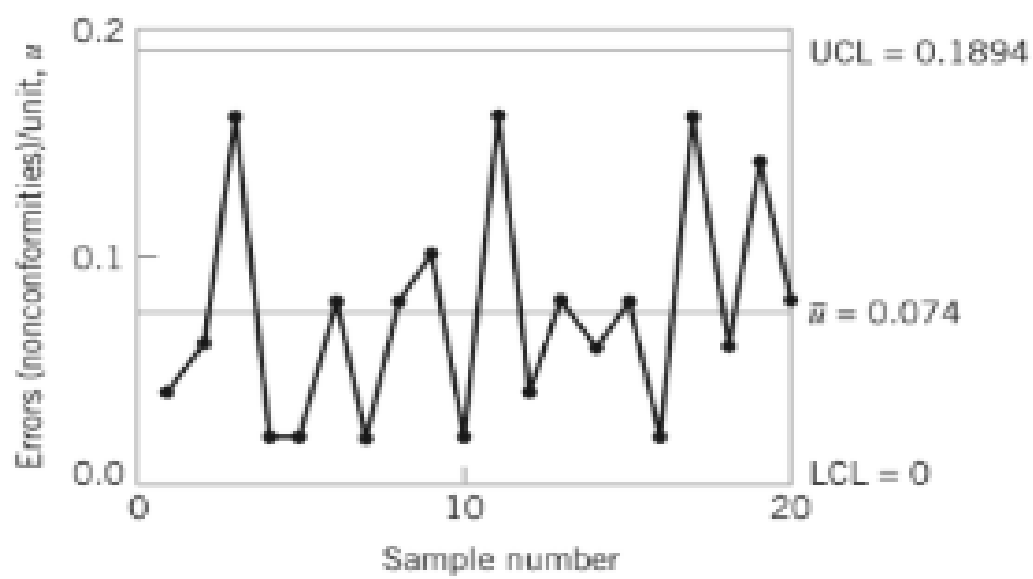


Figura 12 Esempio: Carta U non conformi per unità in un dataset preliminare. Fonte: adattata da Montgomery, 2009.

1.10 Analisi dei residui con carta X-S e utilizzo della regressione polinomiale di terzo grado

Tra le diverse tipologie di carte di controllo presentate, per questa tesi si è scelto di utilizzare la carta $\bar{X}$ -S. Tale scelta deriva principalmente dalla natura dei dati analizzati e dalle caratteristiche del dataset relativo alla Digital Voice of Customer dei parchi Disney.

Nel contesto di questa ricerca, l'applicazione delle carte di controllo non viene effettuata direttamente sulla serie storica originale degli indicatori dVoC, bensì sui residui ottenuti da un modello di regressione polinomiale. Tale scelta metodologica deriva dalla necessità di distinguere le variazioni strutturali e sistematiche del fenomeno dalle anomalie effettive del processo.

Le serie temporali generate dagli indicatori derivanti dal modello STM, come la prevalenza dei topic o gli score di sentiment, presentano infatti andamenti non stazionari e caratterizzati da trend evolutivi nel tempo. In presenza di trend marcati, l'applicazione diretta delle carte di controllo potrebbe produrre numerosi falsi segnali di anomalia, poiché le carte interpreterebbero variazioni fisiologiche del trend come deviazioni fuori controllo ([Montgomery, 2009](#)).

Per questo motivo, si procede preliminarmente alla stima di una curva di regressione in grado di rappresentare l'andamento medio atteso della serie. Successivamente vengono calcolati i residui, ossia le differenze tra i valori osservati e quelli stimati dal modello:

$$e_i = y_i - \hat{y}_i$$

Dove y_i rappresenta il valore osservato e $\hat{y}_i$ il valore stimato dalla regressione. I residui costituiscono quindi la componente “depurata” dal trend strutturale e risultano più adatti all'applicazione delle carte di controllo, poiché evidenziano esclusivamente le deviazioni inattese del processo. Il monitoraggio dei residui consente di individuare anomalie reali senza confondere i cambiamenti sistematici del processo con normali evoluzioni temporali. In questo modo, le carte di controllo risultano maggiormente sensibili agli eventi eccezionali e meno influenzate dalle dinamiche di lungo periodo.

La scelta di utilizzare una regressione polinomiale di terzo grado deriva dalle caratteristiche delle serie temporali analizzate e dall'obiettivo di ottenere residui privi delle principali componenti sistematiche del fenomeno. Gli indicatori derivati dal modello STM, come l'IMTP e l'MRP, non presentano infatti un andamento costante nel tempo, ma mostrano variazioni progressive dovute all'evoluzione delle percezioni dei visitatori e alle modifiche apportate al servizio.

Osservando le serie storiche dei quality determinants, è possibile notare come alcuni topic presentino inizialmente una crescita della rilevanza nelle recensioni, seguita da una fase di stabilizzazione e successivamente da una diminuzione o da un'inversione di tendenza. Un modello lineare, rappresentato da una semplice retta, sarebbe in grado di descrivere esclusivamente una crescita o una diminuzione costante nel tempo. In presenza di una serie che aumenta nei primi periodi e successivamente rallenta o decresce, una regressione lineare produrrebbe residui sistematicamente positivi in alcune aree della serie e negativi in altre,

segnalando erroneamente come anomalie variazioni che in realtà appartengono alla normale evoluzione del fenomeno.

Una regressione quadratica consentirebbe di modellare una sola curvatura, descrivendo ad esempio una fase di crescita seguita da una diminuzione, oppure viceversa. Tuttavia, le serie analizzate mostrano spesso dinamiche più articolate, caratterizzate da variazioni del tasso di crescita e da possibili cambiamenti nella direzione del trend. In tali situazioni un modello quadratico potrebbe risultare ancora insufficiente a rappresentare adeguatamente l'andamento osservato.

La regressione polinomiale di terzo grado introduce invece un termine cubico che consente la presenza di un punto di flesso, permettendo di descrivere serie storiche caratterizzate da una crescita iniziale, una successiva fase di rallentamento e un eventuale cambio di tendenza. Tale comportamento risulta coerente con l'evoluzione temporale osservata negli indicatori derivati dalle recensioni dei visitatori Disney, nei quali la rilevanza di un determinato argomento può aumentare, stabilizzarsi e successivamente ridursi in seguito a cambiamenti nell'esperienza percepita dai clienti.

L'utilizzo di polinomi di grado superiore avrebbe certamente garantito una maggiore capacità di adattamento ai dati, ma avrebbe incrementato il rischio di overfitting, portando il modello a seguire anche oscillazioni casuali e rumore statistico. Poiché l'obiettivo dell'analisi è identificare anomalie reali attraverso lo studio dei residui mediante carte di controllo, è stato necessario adottare un modello sufficientemente flessibile da rappresentare il trend generale della serie, ma al contempo sufficientemente semplice da non assorbire le deviazioni anomale che si intendono individuare. Per tali ragioni il polinomio di terzo grado è stato ritenuto la soluzione più appropriata per la successiva analisi SPC.

La funzione che esprime il modello è la seguente:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

Dove rispettivamente:

- β_0 evidenzia l'intercetta, ovvero il valore iniziale atteso della serie quando la variabile temporale assume valori pari a zero;
- $\beta_1, \beta_2, \beta_3$ sono i coefficienti del modello, I coefficienti β sono ottenuti tramite il metodo dei minimi quadrati (OLS), che minimizza gli errori di stima della regressione.
- ε rappresenta il termine di errore, non è visualizzato nei calcoli successivi poiché rappresenta la componente residuale del modello, successivamente analizzata tramite le carte di controllo.

Come mostrato in Figura 13, la regressione polinomiale di terzo grado consente di rappresentare il trend generale della serie temporale senza adattarsi eccessivamente alle oscillazioni di breve periodo. Nel caso illustrato, relativo al quality determinant "Ticket

purchase", la curva di regressione descrive l'evoluzione media dell'indicatore IMTP nel tempo, fornendo una stima del comportamento atteso del fenomeno. Tuttavia, l'obiettivo principale dell'analisi non è il trend in sé, bensì lo scostamento tra i valori osservati e quelli stimati dal modello.

La differenza tra il valore effettivamente osservato e il valore previsto dalla regressione costituisce il residuo. Tale quantità rappresenta la porzione di variabilità che non può essere spiegata dall'andamento generale della serie e che potrebbe quindi essere associata a eventi specifici, cambiamenti improvvisi nelle percezioni dei clienti o anomalie del processo. In altri termini, mentre la regressione descrive ciò che ci si aspetta di osservare sulla base del trend storico, il residuo misura quanto il comportamento reale si discosti da tale aspettativa.

La scelta di una regressione polinomiale di terzo grado assume quindi particolare rilevanza poiché consente di modellare in modo sufficientemente accurato la componente strutturale della serie senza assorbire le variazioni locali che costituiscono il vero oggetto di interesse. Se il modello fosse troppo semplice, parte del trend rimarrebbe nei residui; viceversa, un modello eccessivamente complesso tenderebbe a incorporare anche le oscillazioni anomale, riducendo la capacità di individuarle. Il polinomio di terzo grado rappresenta pertanto un compromesso adeguato tra capacità descrittiva e preservazione dell'informazione contenuta nei residui.

Nella Figura 13, ad esempio, i picchi e le diminuzioni che si discostano dalla curva di regressione generano residui di maggiore entità, evidenziando periodi nei quali la rilevanza del topic risulta significativamente diversa da quella attesa. Sono proprio tali scostamenti a costituire l'input delle successive carte di controllo, le quali permettono di verificare se le deviazioni osservate siano compatibili con la normale variabilità del processo oppure rappresentino segnali anomali meritevoli di ulteriori approfondimenti.

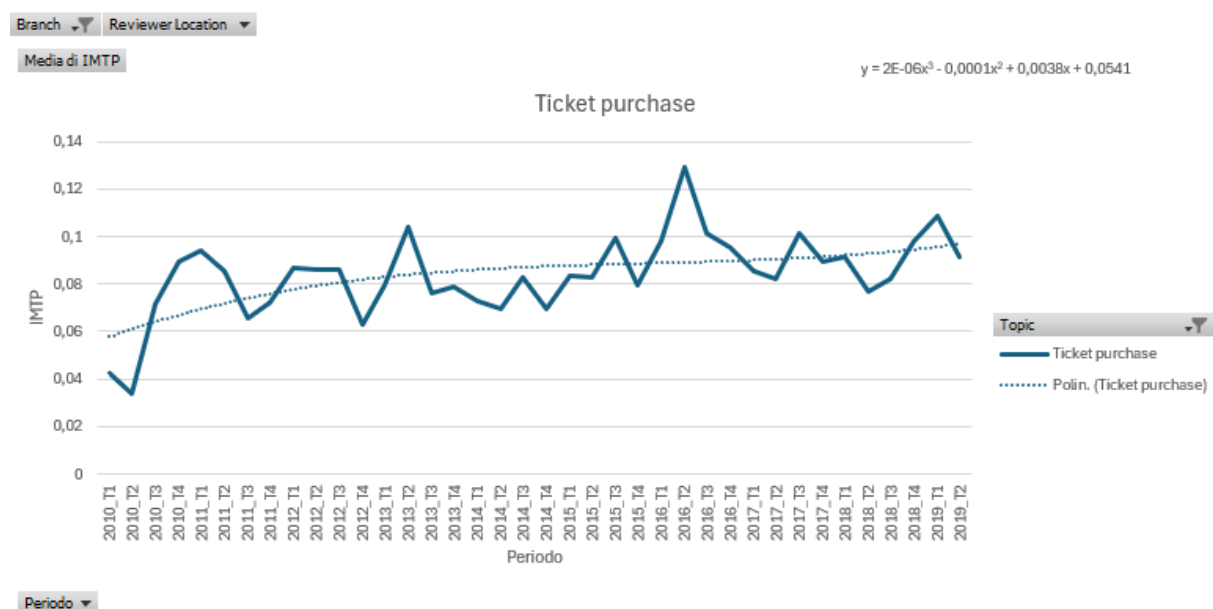


Figura 13 Esempio di regressione polinomiale di terzo grado applicata a una serie temporale con successiva analisi dei residui.

2. Introduzione al caso studio: Descrizione del servizio

Disneyland rappresenta un esempio emblematico di servizio esperienziale complesso, caratterizzato da un'elevata integrazione tra intrattenimento, organizzazione operativa e gestione della qualità percepita. Il primo parco Disneyland venne inaugurato nel 1955 ad Anaheim, in California, su iniziativa di Walt Disney, con l'obiettivo di creare un ambiente in cui visitatori di diverse fasce d'età potessero vivere un'esperienza immersiva all'interno di un mondo tematico coerente e riconoscibile ([The Walt Disney Company, 2023](#)).

Nel corso del tempo, il modello Disneyland è stato esteso a livello internazionale, dando origine a una rete di parchi tematici accomunati da una struttura concettuale comune, ma adattati ai diversi contesti geografici e culturali. Tutti i parchi Disneyland condividono una suddivisione in aree tematiche e un'offerta articolata di servizi che comprende attrazioni, spettacoli, ristorazione, strutture ricettive, negozi e servizi di supporto alla visita ([The Walt Disney Company, 2023](#)). L'esperienza del cliente non si limita quindi alla fruizione delle singole attrazioni, ma coinvolge l'intero percorso di visita, dalla pianificazione fino al termine del soggiorno.

Nel presente lavoro di tesi, l'analisi della Digital Voice of Customer si concentra su tre parchi Disneyland, selezionati in base alla disponibilità dei dati all'interno del database di recensioni utilizzato. I parchi analizzati sono Disneyland Park in California, Disneyland Paris e Hong Kong Disneyland, che rappresentano tre contesti geografici e culturali differenti ma accomunati da una struttura del servizio comparabile ([Disneyland Resort, 2023](#); [Disneyland Paris, 2023](#); [Hong Kong Disneyland Resort, 2023](#)).

Disneyland Park, situato ad Anaheim, è il primo parco Disneyland realizzato ed è l'unico inaugurato in vita da Walt Disney. Il complesso, oggi denominato Disneyland Resort, comprende due parchi a tema, strutture alberghiere e un'area dedicata a ristorazione e shopping ([Disneyland Resort, 2023](#)). Disneyland Paris, localizzato a Marne-la-Vallée, costituisce uno dei principali poli turistici europei e include due parchi a tema, un'area commerciale e numerose strutture ricettive, configurandosi come un sistema di servizi integrato rivolto a una clientela internazionale ([Disneyland Paris, 2023](#)). Hong Kong Disneyland, parte dell'Hong Kong Disneyland Resort, rappresenta il primo parco Disney realizzato in Cina e combina il parco a tema con strutture alberghiere e aree verdi, adattando il modello Disneyland alle specificità del contesto asiatico ([Hong Kong Disneyland Resort, 2023](#)).

Nonostante le differenze dimensionali e organizzative, i tre parchi presentano una struttura del servizio comparabile, rendendo possibile un'analisi congiunta della percezione dei clienti. La complessità del servizio offerto, unita alla sua forte componente esperienziale, genera una Digital Voice of Customer particolarmente ricca e articolata, che riflette valutazioni ad aspetti eterogenei quali la qualità delle attrazioni, l'interazione con il personale, la gestione delle code, l'organizzazione degli spazi e il rapporto tra prezzo e valore percepito ([Nair & Raju, 2017](#)).

2.1 Implementazione dell'algoritmo di Topic Modeling

L'implementazione dell'algoritmo di topic modeling utilizzata nel presente lavoro si basa su un'analisi precedentemente condotta nell'ambito del corso di riferimento "Qualità nei servizi", dalla quale sono stati ripresi sia il dataset pre-processato sia i principali risultati relativi all'individuazione e alla validazione dei topic. Tale analisi preliminare ha consentito di identificare le principali criticità e i punti di forza del campione di User-Generated Content (UGC), costituendo una base metodologica solida per le successive fasi di studio.

L'algoritmo di topic modeling è stato implementato sulla piattaforma R, utilizzando il modello Structural Topic Model (STM) ([Roberts et al., 2014](#)). Al termine della fase di pre-processing, il dataset risultava composto da 42.649 recensioni, a fronte delle 42 655 iniziali, in quanto la lunghezza mediamente elevata dei testi non ha consentito l'eliminazione di un numero significativo di osservazioni.

Una fase cruciale dell'analisi ha riguardato la determinazione del numero ottimale di topic. A tal fine, è stato utilizzato il parametro di held-out likelihood, che consente di valutare il compromesso tra capacità descrittiva del modello ed eccessiva ridondanza dei topic individuati. Sono stati effettuati diversi tentativi, considerando valori alternativi del numero di topic. In particolare, una configurazione con 10 topic si è rivelata insufficiente a rappresentare in modo adeguato l'informazione contenuta nel dataset, mentre una soluzione con 18 topic ha mostrato un'elevata ridondanza semantica. Sulla base di tali considerazioni, è stato selezionato un numero pari a 15 topic, valore oltre il quale la funzione di held-out likelihood assume un andamento pressoché stazionario, come riportato in Figura 14.

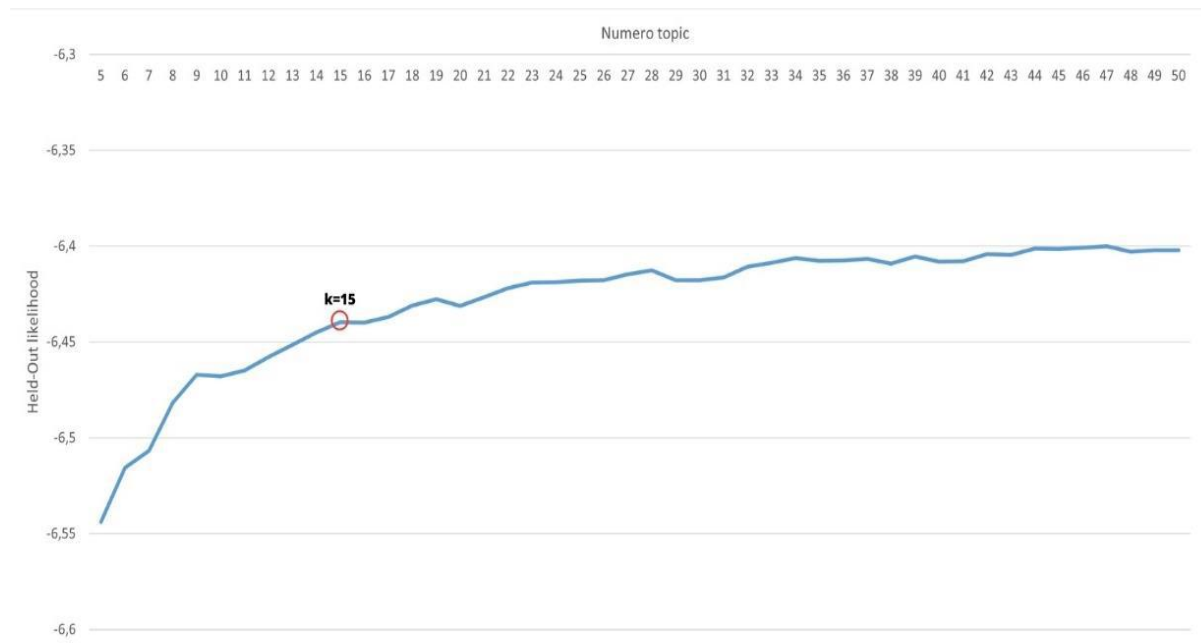


Figura 14 Individuazione del numero ottimale di topic pari a 15. Fonte: Dalponte et al., 2023.

I quindici topic individuati rappresentano le principali dimensioni tematiche emerse dalle recensioni e coprono aspetti eterogenei dell'esperienza del visitatore, quali l'organizzazione del parco, le attrazioni, i servizi di ristorazione, l'interazione con il personale, la gestione delle code, gli eventi speciali e il rapporto qualità-prezzo.

Al fine di verificare la correttezza dell'assegnazione dei topic, è stata condotta una procedura di validazione su un campione casuale e rappresentativo di 100 recensioni, confrontando l'assegnazione automatica con una classificazione manuale. I risultati della validazione mostrano un numero complessivo di 110 veri positivi, 1.274 veri negativi, 39 falsi positivi e 77 falsi negativi. Gli indicatori di performance ottenuti risultano nel complesso soddisfacenti; le principali criticità, in particolare il False Discovery Rate e il Miss Rate, sono attribuibili alla lunghezza elevata delle recensioni, che può influenzare la corretta stima delle distribuzioni di probabilità dei topic.

Considerata la bontà complessiva degli indicatori di validazione, le etichette dei topic individuati sono state confermate e utilizzate come base per le successive analisi sviluppate nel presente lavoro, focalizzate sull'evoluzione temporale della Digital Voice of Customer e sull'individuazione dei periodi anomali tramite modelli di regressione e carte di controllo, come dimostrato nella Figura 15.

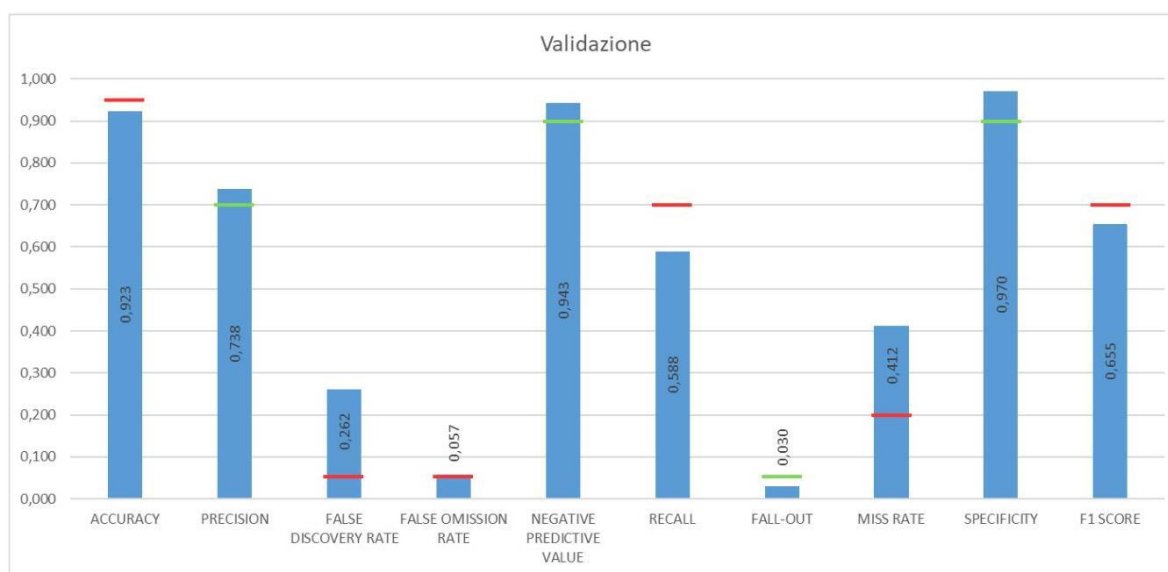


Figura 15 Report Riassuntivo di validazione dei dati. Fonte: Dalponte et al., 2023.

Di seguito i 15 Topic individuati:

- T1: “Lack in the organization”, disagi all'interno del parco a livello organizzativo;
- T2: “Disneyland trip for all ages”, la capacità del parco di essere attrattivo per tutte le fasce d'età;
- T3: “Visiting disneyland with children”, suggerimenti da parte degli utenti su come trascorrere al meglio l'esperienza al parco con i propri figli;
- T4: “Hotel experience”, giudizi sui servizi di hospitality offerti a corredo dell'esperienza;

- T5: “Judgment on staff”, giudizi sullo staff presente all’interno del parco;
- T6: “Waiting time”, considerazioni sui tempi di attesa presso le attrazioni e i vari servizi;
- T7: “Disneyland dining recommendations”, suggerimenti sui servizi di ristorazione;
- T8: “Disneyland rides”, commenti sulle attrazioni del parco;
- T9: “Meet and Greet with Disney’s characters”, considerazioni sulle interazioni con i vari personaggi;
- 10: “Special Event’s Celebrations”, visite al parco di divertimento in occasioni di celebrazioni speciali o festività;
- T11: “FastPass”, commenti sul servizio extra offerto per evitare le code alle attrazioni;
- T12: “Show and Parade”, considerazioni sugli spettacoli disponibili presso i parchi;
- T13: “Ticket purchase”, considerazioni sull’esperienza di acquisto dei biglietti;
- T14: “Disney World vs Disneyland”, elementi distintivi tra il parco Disney World in Florida e gli altri parchi Disneyland in giro per il mondo;
- T15: “Overpricing in Disney parks”, considerazioni sui prezzi proposti in relazione al servizio offerto.

2.2 Analisi degli indicatori derivati dal Topic Modeling

Validata la base di dati precedente, si sono effettuate diverse analisi finalizzate al controllo statistico del processo. L'analisi riguarda vari indicatori quali:

- MTP: Mean Topic Prevalence per avere una visione di quanto i topic fossero discussi, messi a confronto tra i tre parchi
- MRP: Mean Rating Proportion con la possibilità di avere una percezione polarizzata della discussione del topic in determinati periodi e contesti
- IMTP: Interval Mean Topical Proportion con la possibilità di avere una visione nel tempo del trend del topic. Nella presente ricerca, l'andamento temporale IMTP è stato modellato mediante regressione polinomiale al fine di descrivere la componente di trend della serie e ottenere i residui da sottoporre successivamente ad analisi tramite carte di
- Carte di controllo X-S: con la possibilità di studiare l'andamento dei residui nel tempo e la loro variabilità

2.3 Analisi MTP (Mean Topic Prevalence)

Come primo passo dell'analisi dei risultati, è stato calcolato il **Mean Topical Prevalence (MTP)**, al fine di valutare il livello medio di discussione dei diversi topic all'interno delle recensioni analizzate. L'indicatore calcolato mediante la seguente formula matematica ([Mastrogiacomo et al., 2021](#)):

$$MTP_t = \frac{\sum_{i=1}^N TP_{i,t}}{N}$$

Dove N è il numero di recensioni considerate, mentre $TP_{i,t}$ la prevalenza tematica del topic (t) (key-attribute) nella recensione i-esima.

L'indicatore consente di quantificare la rilevanza media dei diversi topic all'interno della *Digital Voice of Customer*, fornendo una prima caratterizzazione degli aspetti maggiormente discussi dagli utenti. Il grafico in Figura 16 riporta il confronto dei valori di MTP per i tre parchi analizzati – **Disneyland California, Disneyland Paris e Hong Kong Disneyland** – consentendo di evidenziare analogie e differenze nella distribuzione dei topic tra contesti geografici differenti. In generale, i topic risultano distribuiti in maniera relativamente omogenea tra i tre parchi, a conferma del fatto che la struttura del servizio e gli aspetti maggiormente discusse dai visitatori presentano elementi comuni.

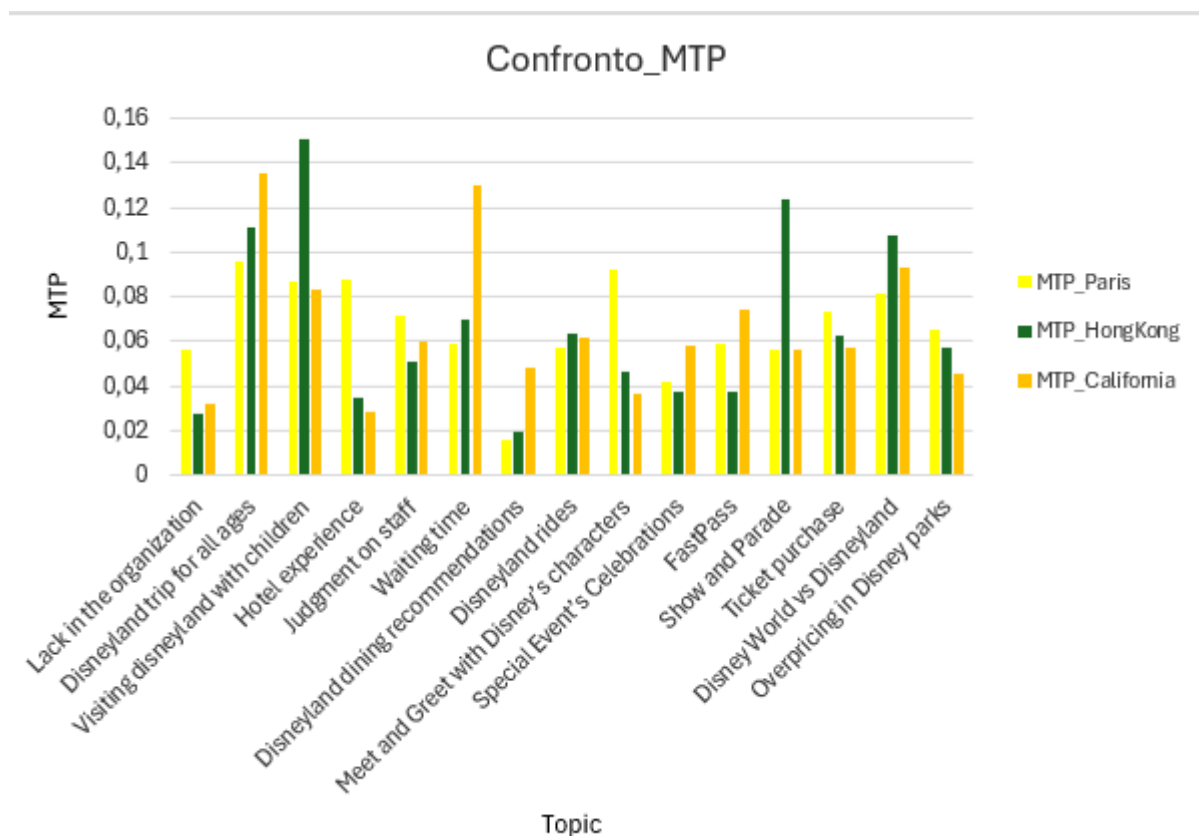


Figura 16 Confronto MTP tra Branch

Alcuni topic emergono tuttavia come più rilevanti rispetto ad altri. In particolare, i topic “Disneyland trip for all ages” e “Visiting Disneyland with children” mostrano valori di MTP elevati in tutti e tre i parchi, suggerendo che l’esperienza familiare rappresenti un elemento centrale del servizio offerto. Tale risultato è coerente con il posizionamento di Disneyland come destinazione rivolta a un pubblico trasversale, con una forte focalizzazione sulle famiglie con bambini, che costituiscono uno dei principali segmenti di clientela.

Altri topic mostrano invece differenze più marcate tra i parchi. Ad esempio, il topic “Waiting time” risulta particolarmente discusso nel parco di Disneyland California, indicando che la gestione delle code rappresenta un aspetto rilevante per i visitatori di tale contesto. Analogamente, il topic “Show and Parade” assume una maggiore rilevanza nel caso di Hong Kong Disneyland, suggerendo una maggiore attenzione o sensibilità dei visitatori verso gli spettacoli e gli eventi programmati.

Nel complesso, l’analisi del MTP consente di individuare i temi che caratterizzano maggiormente l’esperienza del visitatore nei diversi parchi e fornisce una prima indicazione sugli aspetti del servizio che risultano più frequentemente oggetto di discussione. Tali risultati costituiscono la base per le successive analisi temporali e qualitative, nelle quali l’attenzione verrà posta sull’evoluzione dei topic nel tempo e sull’individuazione di periodi anomali attraverso indicatori più avanzati.

2.4 Analisi MRP (Mean Rating Proportion)

Accanto all'analisi della prevalenza dei topic, il presente lavoro introduce l'indicatore Mean Rating Proportion (MRP) con l'obiettivo di approfondire la dimensione percettiva associata ai contenuti espressi nelle recensioni ([Barravecchia et al., 2022](#)). Calcolato attraverso la seguente formula matematica:

$$MRP_{t,k} = \frac{\sum_{j \in R_k} TP_{i,t}}{|R_k|}$$

Dove (t) rappresenta il topic considerato; k il livello della scala di valutazione; R_k il sottoinsieme delle recensioni associate al livello rating k; $TP_{i,t}$ la prevalenza tematica del topic (t) nella recensione i-esima; e $|R_k|$ la cardinalità dell'insieme R_k , ossia il numero di recensioni appartenenti a tale sottoinsieme.

Mentre l'MTP consente di quantificare quanto un determinato tema venga discusso all'interno della Digital Voice of Customer, l'MRP permette di analizzare come tale tema venga valutato dagli utenti, fornendo quindi un'informazione complementare.

L'MRP si basa sulla distribuzione delle recensioni all'interno di cinque classi ordinali di giudizio, che rappresentano un gradiente di valutazione dalla più negativa (1st) alla più positiva (5th). In questo modo, l'indicatore consente di associare a ciascun topic non solo un livello medio di rilevanza, ma anche una misura sintetica della percezione espressa dagli utenti. Il grafico in Figura 17 mostra un esempio dell'andamento dell'MRP nel tempo, evidenziando come la composizione delle classi di giudizio vari nei diversi periodi analizzati.

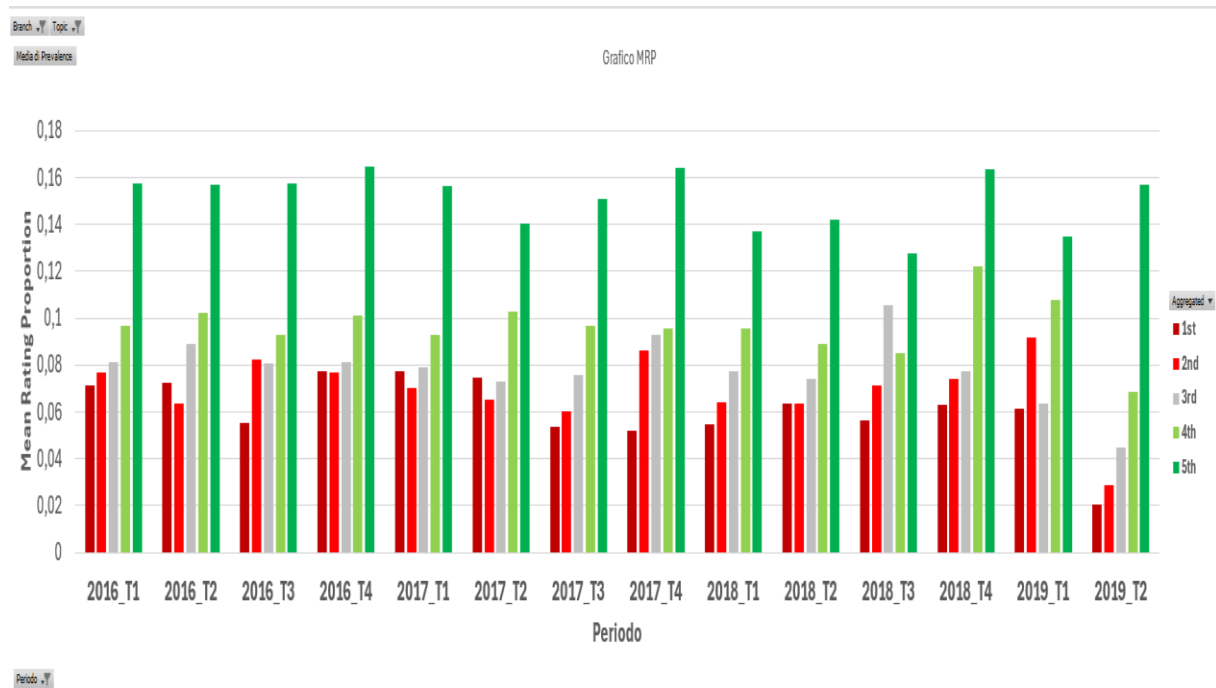


Figura 17 Es. Mean Rating Proportion Branch Disneyland California su Topic "Disneyland Trip for all ages" Periodo di riferimento dal 2016_T1 al 2019_T2

In questa analisi fatta su un campione di dati, appartenenti al Topic “Disneyland Trip for all ages” con periodo compreso tra 2016_T1 al 2019_T2, si può evincere come il gradiente di valutazione si sia modificato sempre più nel corso degli anni passando da valutazione complessivamente positiva del 2016_T1, con una percentuale prossima allo 0,08 delle classi rating più basse (1st e 2nd) ad una valutazione nettamente positiva del 2019_T2 dove le classi rating più basse si assestano intorno ad un valore pari allo 0,02, sei punti percentuali in meno rispetto all’anno di partenza.

Rispetto agli indicatori aggregati tradizionali, l’MRP permette di raggiungere un maggiore grado di profondità analitica, in quanto consente di distinguere tra topic discussi prevalentemente in termini positivi, negativi o caratterizzati da una percezione più eterogenea. Tale aspetto è particolarmente rilevante nel contesto di servizi complessi e fortemente esperienziali come i parchi Disneyland, nei quali uno stesso elemento del servizio può generare valutazioni molto diverse tra i visitatori.

In continuità con il lavoro di analisi svolto nelle fasi precedenti, l’utilizzo dell’MRP consente quindi di arricchire l’interpretazione della Digital Voice of Customer, passando da una lettura puramente quantitativa della rilevanza dei topic a una lettura qualitativa della percezione associata. Questo livello di dettaglio risulta fondamentale per comprendere in modo più completo le dinamiche dell’esperienza del cliente e costituisce la base per le successive analisi temporali più approfondite sviluppate nel capitolo successivo.

2.5 Analisi IMTP (Interval Mean Topical Prevalence)

L’interval Mean Topical Prevalence (IMTP) rappresenta una misura sintetica della rilevanza dei topic emersi dal processo di Topic Modeling nel corso del tempo. Tale indice permette di mantenere sotto controllo l’evoluzione della tematica nelle recensioni, fornendo una visione d’insieme dell’intensità e della frequenza con cui diversi argomenti vengono discussi dai visitatori ([Barravecchia et al., 2023a, b](#)). Si ottiene attraverso la seguente formulazione matematica:

$$IMTP_{t,n} = \frac{\sum_j^{R_t} TP_{t,n}}{|R_n|}$$

Dove R_t identifica l’insieme delle recensioni della Digital Voice of Customer raccolte nel t-esimo intervallo temporale, mentre $|R_n|$ rappresenta il numero totale di recensioni appartenenti a tale insieme. Per ciascun periodo temporale considerato, la somma dei valori di $IMTP_{t,n}$ relativi a tutte le determinanti di qualità individuati è pari a 1

L’indicatore IMTP si basa sul concetto di Topic Prevalence, ossia sulla proporzione con cui un determinato topic è presente all’interno di un documento. Tale concetto deriva dai modelli probabilistici di Topic Modeling, introdotti originariamente dal Latent Dirichlet Allocation

(LDA), nel quale ciascun documento è rappresentato come una distribuzione di probabilità sui topic e ciascun topic come una distribuzione di probabilità sulle parole ([Blei et al., 2003](#)).

Nel contesto di questa ricerca, l'IMTP viene utilizzato come indicatore sintetico per descrivere l'andamento temporale della discussione dei topic estratti dalle recensioni pubblicate su piattaforme digitali relative ai parchi Disneyland nel periodo che va dal 2010-2019.

Per la fase di sviluppo dell'indicatore IMTP e della Carta di controllo X-S è stato previsto un capitolo descrittivo di Costruzione e Pre-elaborazione dei dati, che segue.

3 Costruzione del Database e Pre-elaborazione dei dati

Per il calcolo dell'IMTP si è resa necessaria una fase di integrazione e pulizia del dataset originari, realizzata con l'ausilio di Microsoft Excel tramite l'editor di Power Query.

- Database di partenza: Il dataset iniziale conteneva tutte le recensioni dei visitatori, con informazioni su Review ID, Rating, Anno e Mese, Testo della recensione, Branch (parco di appartenenza) e Reviewer Location (paese di provenienza). Inoltre, ciascun record includeva i valori di prevalenza associati ai vari topic identificati dal modello di Topic Modeling.
- Ristrutturazione con Power Query:
Tramite Power Query si è proceduto a:
 1. Normalizzare i nomi delle colonne e convertire i tipi di dati;
 2. Trasformare le colonne tramite UnPivot, in modo da ottenere una struttura lunga con tre variabili principali: ReviewID, Topic, Prevalence
 3. Si sono aggiunte colonne derivate per la classificazione temporale (Periodo, Quarter, Year) e per la separazione geografica (Branch, Reviewer Location);
 4. Infine, sono stati rimossi i duplicati e salvato il risultato in un nuovo foglio denominato Database_PowerQuery_Lungo.
- Creazione dell'Indice IMTP:
Successivamente, i dati sono stati sintetizzati in una Tabella Pivot, in cui:
 1. Le etichette di riga rappresentano i periodi temporali, scelti i Trimestri per avere un'analisi più rigorosa;
 2. Le etichette di colonna i diversi topic emersi nel modello;
 3. E come valori è stata utilizzata la media aritmetica della prevalenza per ciascun periodo (Media di Prevalence).

La media così ottenuta per ogni periodo costituisce l'Indice Medio di Topic Prevalence (IMTP). Esempio mostrato nella Figura 18.

Periodo	Disney World vs Disneyland	Disneyland dining recommendations	Disneyland rides	Disneyland trip for all ages	FastPass	Hotel experience	Judgment on staff	Lack in the organization	Meet and Greet with Disney's char	Overpricing in Disney parks	Show and Parade	Special Event's Celebrations	Ticket purchase	Visiting disneyland with children	Waiting time	TOTALE
2010_T2	0,06899	0,04642	0,02921	0,16018	0,04092	0,0847	0,044	0,02882	0,04621	0,02911	0,04754	0,05361	0,05206	0,16298	0,10523952	1
2010_T3	0,09086	0,05713	0,02209	0,14595	0,04036	0,04035	0,02873	0,01664	0,02144	0,03024	0,07424	0,05107	0,04628	0,19243	0,14218013	1
2010_T4	0,17829	0,03209	0,04244	0,14329	0,03445	0,03912	0,05558	0,03846	0,04234	0,04137	0,08152	0,054	0,04894	0,08548	0,08261383	1
2011_T1	0,06866	0,10767	0,0824	0,10804	0,05433	0,04407	0,0515	0,02468	0,04484	0,03746	0,0872	0,06621	0,08157	0,0622	0,0791722	1
2011_T2	0,0948	0,04879	0,04333	0,15122	0,07754	0,04742	0,06205	0,0261	0,0357	0,04979	0,04658	0,0369	0,07163	0,10008	0,10809179	1
2011_T3	0,09169	0,03985	0,0606	0,12214	0,09511	0,0415	0,03758	0,05153	0,03034	0,047	0,05821	0,04438	0,06399	0,09572	0,12035389	1
2011_T4	0,14092	0,02624	0,05173	0,1233	0,05466	0,02657	0,06358	0,02677	0,06572	0,0517	0,05356	0,04191	0,07553	0,10534	0,09247036	1
2012_T1	0,09557	0,02422	0,06958	0,16425	0,04751	0,02402	0,04716	0,02212	0,04446	0,04101	0,0885	0,07005	0,06708	0,08771	0,10675481	1
2012_T2	0,1006	0,0394	0,07244	0,14339	0,07067	0,02687	0,04796	0,02282	0,03448	0,06881	0,05051	0,04761	0,05152	0,09561	0,12730416	1
2012_T3	0,08444	0,02804	0,07962	0,12907	0,06154	0,02721	0,05101	0,02732	0,04578	0,05801	0,0592	0,05024	0,06306	0,1045	0,1309502	1
2012_T4	0,09879	0,04086	0,06592	0,14579	0,049	0,03079	0,04322	0,03144	0,03653	0,05375	0,06731	0,04308	0,06978	0,11679	0,10694157	1
2013_T1	0,10439	0,04043	0,06867	0,15397	0,06801	0,0249	0,0447	0,02262	0,03993	0,05017	0,05415	0,04372	0,05197	0,11093	0,12143741	1
2013_T2	0,09012	0,03435	0,05766	0,13033	0,0734	0,03878	0,04548	0,0283	0,0396	0,05886	0,07155	0,04145	0,08202	0,10558	0,10253601	1
2013_T3	0,086	0,03277	0,08228	0,12563	0,06262	0,02826	0,05531	0,02706	0,0488	0,05792	0,07814	0,06164	0,06029	0,10168	0,09161208	1
2013_T4	0,09332	0,02752	0,05914	0,15812	0,04919	0,02672	0,05873	0,0331	0,03818	0,06387	0,07425	0,04708	0,05209	0,11279	0,1058892	1
2014_T1	0,12238	0,03366	0,07531	0,12156	0,05413	0,02263	0,05093	0,02895	0,0351	0,04807	0,07288	0,04016	0,05808	0,12874	0,10740672	1
2014_T2	0,08087	0,04044	0,0856	0,13956	0,07532	0,02496	0,05318	0,02248	0,03624	0,04528	0,07476	0,05057	0,04681	0,10524	0,11868923	1
2014_T3	0,09959	0,03284	0,06168	0,15165	0,05999	0,03279	0,05886	0,03568	0,04274	0,04364	0,0652	0,06496	0,05437	0,11239	0,08361607	1
2014_T4	0,10238	0,03579	0,04092	0,15795	0,04879	0,03704	0,04869	0,0385	0,03869	0,05017	0,05571	0,04406	0,09212	0,12422	0,08496259	1
2015_T1	0,1209	0,02477	0,10688	0,11809	0,06779	0,02212	0,05389	0,02465	0,04066	0,05081	0,04873	0,0631	0,05829	0,11385	0,08547445	1
2015_T2	0,08351	0,02596	0,04807	0,13897	0,08611	0,02482	0,06141	0,02639	0,04231	0,04185	0,06469	0,05017	0,07436	0,11684	0,11453207	1
2015_T3	0,10037	0,03109	0,06844	0,14563	0,0742	0,02801	0,04544	0,03468	0,03454	0,04525	0,06715	0,04711	0,06151	0,10533	0,11126976	1
2015_T4	0,08739	0,04579	0,05631	0,12677	0,05452	0,02296	0,06738	0,03127	0,04209	0,0523	0,06082	0,05263	0,06883	0,11008	0,12088217	1
2016_T1	0,08761	0,0276	0,06544	0,13526	0,06799	0,02073	0,04995	0,02459	0,05756	0,04502	0,07421	0,04849	0,06158	0,12735	0,1066081	1
2016_T2	0,09431	0,01924	0,07482	0,13874	0,06986	0,02417	0,0424	0,02277	0,04625	0,04673	0,08128	0,04756	0,05725	0,13092	0,10367901	1
2016_T3	0,10164	0,02242	0,05334	0,13247	0,07958	0,02417	0,05771	0,02359	0,03551	0,05539	0,07265	0,03876	0,06524	0,12436	0,11317866	1
2016_T4	0,10736	0,02392	0,04834	0,14603	0,06863	0,03165	0,0484	0,0245	0,04027	0,06151	0,04876	0,05823	0,05518	0,11649	0,12072599	1
2017_T1	0,11101	0,05538	0,08352	0,12325	0,07247	0,02835	0,05185	0,02224	0,04593	0,03579	0,0651	0,03522	0,06293	0,09754	0,10943423	1
2017_T2	0,12132	0,03168	0,06962	0,09905	0,0858	0,0248	0,0449	0,03549	0,028	0,04943	0,05089	0,03601	0,07785	0,12369	0,12145906	1
2017_T3	0,08666	0,02792	0,0574	0,14371	0,08041	0,0299	0,06442	0,03375	0,03563	0,04929	0,06918	0,04293	0,0777	0,1015	0,09953296	1
2017_T4	0,09774	0,02497	0,06921	0,13612	0,08225	0,02345	0,0677	0,02696	0,03914	0,0567	0,05401	0,04541	0,0607	0,11823	0,09740209	1
2018_T1	0,07098	0,06062	0,052	0,08787	0,11158	0,03101	0,05048	0,03754	0,032	0,04985	0,05595	0,03943	0,07499	0,09473	0,15096098	1
2018_T2	0,08192	0,02795	0,06373	0,12102	0,08127	0,03142	0,05276	0,02813	0,04357	0,06816	0,06755	0,04301	0,0863	0,09848	0,10473561	1
2018_T3	0,10144	0,03374	0,04339	0,09981	0,10917	0,03082	0,06899	0,03252	0,04132	0,05614	0,04752	0,04218	0,0882	0,08819	0,11658594	1
2018_T4	0,08012	0,03004	0,0419	0,11182	0,1181	0,02539	0,07668	0,03883	0,06066	0,03946	0,0523	0,04954	0,05185	0,08823	0,13506071	1
2019_T1	0,10025	0,01918	0,04405	0,13204	0,08264	0,0422	0,05972	0,02582	0,039	0,06784	0,05033	0,05272	0,07893	0,11477	0,09052561	1
2019_T2	0,07229	0,02386	0,03476	0,18718	0,10355	0,04733	0,05152	0,03988	0,03858	0,0533	0,05347	0,04439	0,07535	0,08995	0,08458867	1
IMTP Medio	0,09728	0,0358	0,06032	0,13511	0,07063	0,03195	0,05308	0,02938	0,04081	0,05003	0,06339	0,04837	0,06557	0,11003	0,10824102	1

Figura 18 Esempio Tabella IMTP per Branch

Per consentire una lettura visiva e dinamica dei risultati, l'indice IMTP è stato integrato all'interno di una Dashboard interattiva sviluppata attraverso l'ausilio di Excel.

3.1 Struttura della Dashboard

La Dashboard, mostrata in Figura 19, si compone di tre sezioni principali:



Figura 19 Dashboard IMTP

- Area filtri (slicer): consente la selezione dei parchi (Branch), della provenienza dei recensori (Reviewer_Location) e dei topic di interesse. I filtri sono interconnessi e permettono di aggiornare dinamicamente i grafici e le metriche sottostanti.
- Indicatori sintetici (KPI): sono stati definiti quattro indicatori sintetici, per la selezione corrente:

- L'IMTP medio,
 - L'IMTP massimo,
 - La deviazione standard
 - E il numero totale di recensioni analizzate
- Tali valori si aggiornano automaticamente mediante le formule GETPIVOTDAT e intervalli dinamici definiti tramite funzioni di SCARTO e CONTA.VALORI.
- Sezione grafica:
 - la dashboard include due grafici principali:
 - Grafico linea che rappresenta l'andamento temporale dell'IMTP (con linea di tendenza polinomiale);
 - Grafico a barre che mostra, per lo stesso periodo, il volume delle recensioni raccolte.

3.2 Funzionamento interattivo

Grazie all'integrazione tra Power Query, Tabella Pivot e Slicer, la dashboard consente una visualizzazione dinamica e multilivello:

- Filtrando per parco, si osserva l'evoluzione dell'indice limitatamente a quella sede;
- Selezionando una Reviewer Location, si evidenziano le differenze geografiche nella percezione dei topic;
- Infine, selezionando un topic specifico, il grafico dà la possibilità di osservare l'andamento della rilevanza di quell'argomento nel tempo.

L'analisi temporale con l'aiuto della Dashboard consentirà di evidenziare **picchi di attenzione**, spesso legati a eventi o cambiamenti significativi nei parchi, **periodi di stabilità** con sentiment costante e le **fasi di flessione**, che possono suggerire momenti di criticità o minore coinvolgimento.

3.3 Costruzione Carte di controllo X-S

Al fine di analizzare in modo strutturato l'evoluzione temporale dei topic e individuare eventuali comportamenti anomali, nel presente lavoro sono state utilizzate le carte di controllo statistiche. L'applicazione di tali strumenti consente di distinguere le normali fluttuazioni della Digital Voice of Customer da variazioni non attribuibili al comportamento ordinario del processo.

Le carte di controllo sono state costruite a partire dai residui dell'IMTP, ottenuti come differenza tra il valore osservato, come mostrato in Figura 20,

Periodo	Media di IMTP															TOTALE
	Disney World vs Disneyland	Disneyland dining recommendations	Disneyland rides	Disneyland trip for all ages	FastPass	Hotel experience	Judgment on staff	Lack in the organization	Meet and Greet with Disney's characters	Overpricing in Disney parks	Show and Parade	Special Event's Celebrations	Ticket purchase	Visiting disneyland with children	Waiting time	
2010_T2	0,069	0,046	0,029	0,16	0,041	0,085	0,044	0,029	0,046	0,029	0,048	0,054	0,052	0,163	0,10524	1
2010_T3	0,091	0,057	0,022	0,146	0,04	0,04	0,029	0,017	0,021	0,03	0,074	0,051	0,046	0,192	0,14218	1
2010_T4	0,178	0,032	0,042	0,143	0,034	0,039	0,056	0,038	0,042	0,041	0,082	0,054	0,049	0,085	0,08261	1
2011_T1	0,069	0,108	0,082	0,108	0,054	0,044	0,051	0,025	0,045	0,037	0,087	0,066	0,082	0,062	0,07917	1
2011_T2	0,095	0,049	0,043	0,151	0,078	0,047	0,062	0,026	0,036	0,05	0,047	0,037	0,072	0,1	0,10809	1
2011_T3	0,092	0,04	0,061	0,122	0,095	0,041	0,038	0,052	0,03	0,047	0,058	0,044	0,064	0,096	0,12035	1
2011_T4	0,141	0,026	0,052	0,123	0,055	0,027	0,064	0,027	0,066	0,052	0,054	0,042	0,076	0,105	0,09247	1
2012_T1	0,096	0,024	0,07	0,164	0,048	0,024	0,047	0,022	0,044	0,041	0,088	0,07	0,067	0,088	0,10675	1
2012_T2	0,101	0,039	0,072	0,143	0,071	0,027	0,048	0,023	0,034	0,069	0,051	0,048	0,052	0,096	0,1273	1
2012_T3	0,084	0,028	0,08	0,129	0,062	0,027	0,051	0,027	0,046	0,058	0,059	0,05	0,063	0,105	0,13095	1
2012_T4	0,099	0,041	0,066	0,146	0,049	0,031	0,043	0,031	0,037	0,054	0,067	0,043	0,07	0,117	0,10694	1
2013_T1	0,104	0,04	0,069	0,154	0,068	0,025	0,045	0,023	0,04	0,05	0,054	0,044	0,052	0,111	0,12144	1
2013_T2	0,09	0,034	0,058	0,13	0,073	0,039	0,045	0,028	0,04	0,059	0,072	0,041	0,082	0,106	0,10254	1
2013_T3	0,086	0,033	0,082	0,126	0,063	0,028	0,055	0,027	0,049	0,058	0,078	0,062	0,06	0,102	0,09161	1
2013_T4	0,093	0,028	0,059	0,158	0,049	0,027	0,059	0,033	0,038	0,064	0,074	0,047	0,052	0,113	0,10589	1
2014_T1	0,122	0,034	0,075	0,122	0,054	0,023	0,051	0,029	0,035	0,048	0,073	0,04	0,058	0,129	0,10741	1
2014_T2	0,081	0,04	0,086	0,14	0,075	0,025	0,053	0,022	0,036	0,045	0,075	0,051	0,047	0,105	0,11869	1
2014_T3	0,1	0,033	0,062	0,152	0,06	0,033	0,059	0,036	0,043	0,044	0,065	0,065	0,054	0,112	0,08362	1
2014_T4	0,102	0,036	0,041	0,158	0,049	0,037	0,049	0,038	0,039	0,05	0,056	0,044	0,092	0,124	0,08496	1
2015_T1	0,121	0,025	0,107	0,118	0,068	0,022	0,054	0,025	0,041	0,051	0,049	0,063	0,058	0,114	0,08547	1
2015_T2	0,084	0,026	0,048	0,139	0,086	0,025	0,061	0,026	0,042	0,042	0,065	0,05	0,074	0,117	0,11453	1
2015_T3	0,1	0,031	0,068	0,146	0,074	0,028	0,045	0,035	0,035	0,045	0,067	0,047	0,062	0,105	0,11127	1
2015_T4	0,087	0,046	0,056	0,127	0,055	0,023	0,067	0,031	0,042	0,052	0,061	0,053	0,069	0,11	0,12088	1
2016_T1	0,088	0,028	0,065	0,135	0,068	0,021	0,05	0,025	0,058	0,045	0,074	0,048	0,062	0,127	0,10661	1
2016_T2	0,094	0,019	0,075	0,139	0,07	0,024	0,042	0,023	0,046	0,047	0,081	0,048	0,057	0,131	0,10368	1
2016_T3	0,102	0,022	0,053	0,132	0,08	0,024	0,058	0,024	0,036	0,055	0,073	0,039	0,065	0,124	0,11318	1
2016_T4	0,107	0,024	0,048	0,146	0,069	0,032	0,048	0,025	0,04	0,062	0,049	0,058	0,055	0,116	0,12073	1
2017_T1	0,111	0,055	0,084	0,123	0,072	0,028	0,052	0,022	0,046	0,036	0,065	0,035	0,063	0,098	0,10943	1
2017_T2	0,121	0,032	0,07	0,099	0,086	0,025	0,045	0,035	0,028	0,049	0,051	0,036	0,078	0,124	0,12146	1
2017_T3	0,087	0,028	0,057	0,144	0,08	0,03	0,064	0,034	0,036	0,049	0,069	0,043	0,078	0,101	0,09359	1
2017_T4	0,098	0,025	0,069	0,136	0,082	0,023	0,068	0,027	0,039	0,057	0,054	0,045	0,061	0,118	0,0974	1
2018_T1	0,071	0,061	0,052	0,088	0,112	0,031	0,05	0,038	0,032	0,05	0,056	0,039	0,075	0,095	0,15096	1
2018_T2	0,082	0,028	0,064	0,121	0,081	0,031	0,053	0,028	0,044	0,068	0,068	0,043	0,086	0,098	0,10474	1
2018_T3	0,101	0,034	0,043	0,1	0,109	0,031	0,069	0,033	0,041	0,056	0,048	0,042	0,088	0,088	0,11659	1
2018_T4	0,08	0,03	0,042	0,112	0,118	0,025	0,077	0,039	0,061	0,039	0,052	0,05	0,052	0,088	0,13506	1
2019_T1	0,1	0,019	0,044	0,132	0,083	0,042	0,06	0,026	0,039	0,068	0,05	0,053	0,079	0,115	0,09053	1
2019_T2	0,072	0,024	0,035	0,187	0,104	0,047	0,052	0,04	0,039	0,053	0,053	0,044	0,075	0,09	0,08459	1
IMTP Medio	0,097	0,036	0,06	0,135	0,071	0,032	0,053	0,029	0,041	0,05	0,063	0,048	0,066	0,11	0,10824	1

Figura 20 Matrice IMTP di partenza per le carte di controllo X-S

e il valore predetto da una funzione di regressione polinomiale di terzo grado, fornisco es. di equazione ottenuta su un grafico IMTP di una determinante di qualità, Figura 21:

$$y = 2E-06x^3 - 0,0001x^2 + 0,0038x + 0,0541$$

Figura 21 Esempio di regressione polinomiale di 3 grado su grafico IMTP di una determinante di qualità

Ottenuta la funzione di regressione Y (valore predetto) si è poi calcolato il valore del residuo come differenza tra valore osservato e valore predetto, come dimostrato attraverso foglio di calcolo Excel in Figura 21.

Indice Periodo	x^3	x^2	x	y Disney World vs Disneyland	RESIDUO
1	1	1	1	0,1021927	-0,03320
2	8	4	2	0,10153048	-0,01067
3	27	9	3	0,10097827	0,07731
4	64	16	4	0,10052704	-0,03187
5	125	25	5	0,1001678	-0,00537
6	216	36	6	0,09989152	-0,00820
7	343	49	7	0,09968919	0,04123
8	512	64	8	0,09955179	-0,00399
9	729	81	9	0,09947032	0,00113
10	1000	100	10	0,09943576	-0,01500
11	1331	121	11	0,09943908	-0,00064
12	1728	144	12	0,09947129	0,00492
13	2197	169	13	0,09952337	-0,00941
14	2744	196	14	0,0995863	-0,01359
15	3375	225	15	0,09965106	-0,00634
16	4096	256	16	0,09970865	0,02267
17	4913	289	17	0,09975005	-0,01888
18	5832	324	18	0,09976624	-0,00017
19	6859	361	19	0,09974822	0,00263
20	8000	400	20	0,09968696	0,02122
21	9261	441	21	0,09957346	-0,01606
22	10648	484	22	0,0993987	0,00097
23	12167	529	23	0,09915367	-0,01177
24	13824	576	24	0,09882934	-0,01122
25	15625	625	25	0,09841672	-0,00410
26	17576	676	26	0,09790678	0,00373
27	19683	729	27	0,09729051	0,01007
28	21952	784	28	0,09655889	0,01445
29	24389	841	29	0,09570292	0,02562
30	27000	900	30	0,09471358	-0,00806
31	29791	961	31	0,09358185	0,00416
32	32768	1024	32	0,09229872	-0,02132
33	35937	1089	33	0,09085518	-0,00893
34	39304	1156	34	0,08924221	0,01220
35	42875	1225	35	0,08745079	-0,00733
36	46656	1296	36	0,08547192	0,01478
37	50653	1369	37	0,08329659	-0,01101

Figura 22 Esempio di calcolo del Residuo, foglio di lavoro su Excel

Questo approccio consente di isolare le componenti non spiegate dal trend di lungo periodo, rendendo più agevole l'individuazione di variazioni anomale.

Sui residui così ottenuti sono state costruite le carte di controllo di tipo X-S, in cui la Carta X monitora l'andamento dei residui medi per periodo, ottenuta con le formule presentate nel

capitolo 1.9.2 Carta X-S (Media e Dev.Standard), applicate poi su foglio di calcolo Excel come mostrato in Figura 23:

MEDIA	DEV.STANDARD	REVIEWCOUNT	UCLx	LCLx	STANDARD ERROR
3,70E-12	0,057431627	8	6,10E-02	-6,10E-02	0,02
3,70E-12	0,137262194	12	4,98E-02	-4,98E-02	0,017
3,70E-12	0,106338799	52	2,39E-02	-2,39E-02	0,008
3,70E-12	0,054856084	64	2,16E-02	-2,16E-02	0,007
3,70E-12	0,079984922	144	1,44E-02	-1,44E-02	0,005
3,70E-12	0,097619483	250	1,09E-02	-1,09E-02	0,004
3,70E-12	0,085028752	493	7,77E-03	-7,77E-03	0,003
3,70E-12	0,085009542	492	7,78E-03	-7,78E-03	0,003
3,70E-12	0,079152006	574	7,20E-03	-7,20E-03	0,002
3,70E-12	0,083408983	621	6,92E-03	-6,92E-03	0,002
3,70E-12	0,078920295	620	6,93E-03	-6,93E-03	0,002
3,70E-12	0,083160577	435	8,27E-03	-8,27E-03	0,003
3,70E-12	0,090877428	625	6,90E-03	-6,90E-03	0,002
3,70E-12	0,09938639	713	6,46E-03	-6,46E-03	0,002
3,70E-12	0,07492517	554	7,33E-03	-7,33E-03	0,002
3,70E-12	0,085040868	430	8,32E-03	-8,32E-03	0,003
3,70E-12	0,088861732	638	6,83E-03	-6,83E-03	0,002
3,70E-12	0,112851269	848	5,92E-03	-5,92E-03	0,002
3,70E-12	0,084739024	727	6,40E-03	-6,40E-03	0,002
3,70E-12	0,088112836	554	7,33E-03	-7,33E-03	0,002
3,70E-12	0,097559775	885	5,80E-03	-5,80E-03	0,002
3,70E-12	0,093612987	1010	5,43E-03	-5,43E-03	0,002
3,70E-12	0,078085619	790	6,14E-03	-6,14E-03	0,002
3,70E-12	0,078746771	646	6,79E-03	-6,79E-03	0,002
3,70E-12	0,099089749	766	6,23E-03	-6,23E-03	0,002
3,70E-12	0,094053482	743	6,33E-03	-6,33E-03	0,002
3,70E-12	0,088622403	645	6,79E-03	-6,79E-03	0,002
3,70E-12	0,089731125	508	7,65E-03	-7,65E-03	0,003
3,70E-12	0,099211405	570	7,22E-03	-7,22E-03	0,002
3,70E-12	0,112716215	515	7,60E-03	-7,60E-03	0,003
3,70E-12	0,090688506	471	7,95E-03	-7,95E-03	0,003
3,70E-12	0,083660543	379	8,86E-03	-8,86E-03	0,003
3,70E-12	0,090934832	432	8,30E-03	-8,30E-03	0,003
3,70E-12	0,117644895	416	8,46E-03	-8,46E-03	0,003
3,70E-12	0,079258312	274	1,04E-02	-1,04E-02	0,003
3,70E-12	0,092019181	238	1,12E-02	-1,12E-02	0,004
3,70E-12	0,130855271	54	2,35E-02	-2,35E-02	0,008

Figura 23 Esempio di calcolo della Carta X

mentre la Carta S consente di analizzare la variabilità associata a ciascun periodo, ottenuta anch'essa dalle formule sopra riportate e applicate su foglio di calcolo Excel come mostrato in Figura 24.

s medio	UCLs	LCLs	C4(n)	B4	B3
0,0911	0,1649	0,0172	0,9653	1,8112	0,1888
0,0911	0,1498	0,0323	0,9776	1,6452	0,3548
0,0911	0,1182	0,064	0,9951	1,2977	0,7023
0,0911	0,1155	0,0667	0,996	1,2678	0,7322
0,0911	0,1072	0,0749	0,9983	1,1775	0,8225
0,0911	0,1033	0,0788	0,999	1,1345	0,8655
0,0911	0,0998	0,0824	0,9995	1,0957	0,9043
0,0911	0,0998	0,0823	0,9995	1,0958	0,9042
0,0911	0,0991	0,083	0,9996	1,0886	0,9114
0,0911	0,0988	0,0833	0,9996	1,0852	0,9148
0,0911	0,0988	0,0833	0,9996	1,0853	0,9147
0,0911	0,1003	0,0818	0,9994	1,1019	0,8981
0,0911	0,0988	0,0833	0,9996	1,0849	0,9151
0,0911	0,0983	0,0838	0,9996	1,0795	0,9205
0,0911	0,0993	0,0829	0,9995	1,0902	0,9098
0,0911	0,1004	0,0817	0,9994	1,1024	0,8976
0,0911	0,0987	0,0834	0,9996	1,0841	0,9159
0,0911	0,0977	0,0844	0,9997	1,0729	0,9271
0,0911	0,0982	0,0839	0,9997	1,0787	0,9213
0,0911	0,0993	0,0829	0,9995	1,0902	0,9098
0,0911	0,0976	0,0846	0,9997	1,0714	0,9286
0,0911	0,0972	0,085	0,9998	1,0668	0,9332
0,0911	0,0979	0,0842	0,9997	1,0755	0,9245
0,0911	0,0987	0,0835	0,9996	1,0835	0,9165
0,0911	0,0981	0,0841	0,9997	1,0767	0,9233
0,0911	0,0982	0,084	0,9997	1,0779	0,9221
0,0911	0,0987	0,0835	0,9996	1,0836	0,9164
0,0911	0,0996	0,0825	0,9995	1,0942	0,9058
0,0911	0,0992	0,083	0,9996	1,0889	0,9111
0,0911	0,0996	0,0825	0,9995	1,0936	0,9064
0,0911	0,1	0,0822	0,9995	1,0979	0,9021
0,0911	0,101	0,0811	0,9993	1,1091	0,8909
0,0911	0,1004	0,0818	0,9994	1,1022	0,8978
0,0911	0,1006	0,0816	0,9994	1,1042	0,8958
0,0911	0,1028	0,0794	0,9991	1,1284	0,8716
0,0911	0,1036	0,0785	0,9989	1,1379	0,8621
0,0911	0,1177	0,0645	0,9953	1,292	0,708

Figura 24 Esempio di calcolo della Carta S

I limiti di controllo non sono stati assunti come costanti, ma sono stati calcolati in modo dinamico, tenendo conto del numero di recensioni disponibili in ciascun periodo come rappresentato in Figura 23 nella colonna “Review Count”. Tale scelta metodologica permette di adattare i limiti di controllo alla diversa numerosità campionaria, rendendo le carte più robuste e coerenti con la struttura dei dati analizzati. I calcoli permettono così di ottenere le due carte di controllo, $\bar{X}$ riportata in Figura 25 e la Carta S riportata in Figura 26, che sono state il punto di partenza per l’evoluzione successiva.

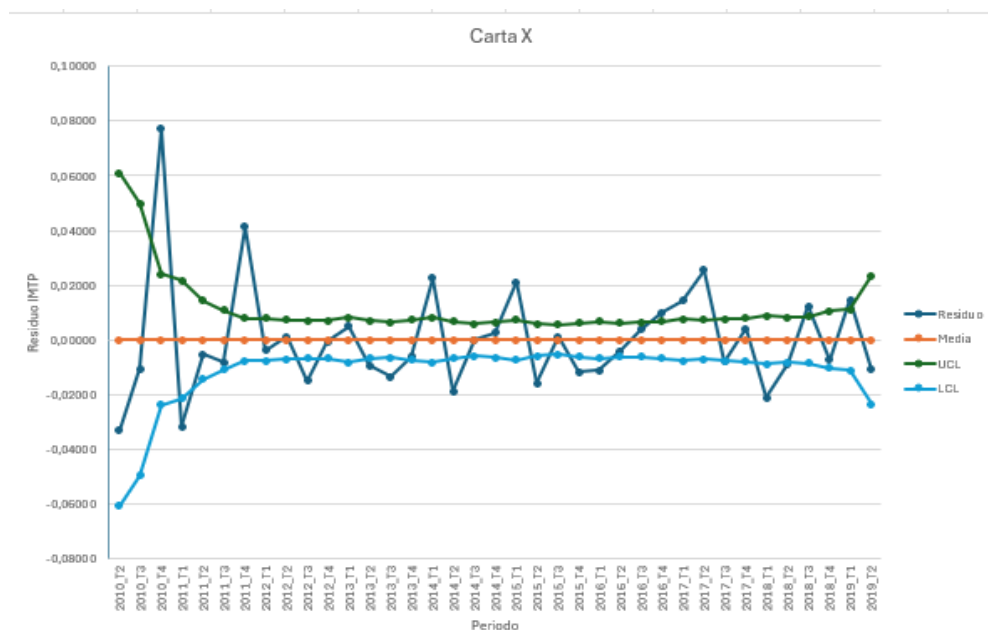


Figura 25 Esempio di Carta X per una determinante di qualità: Disney Word vs Disneyland

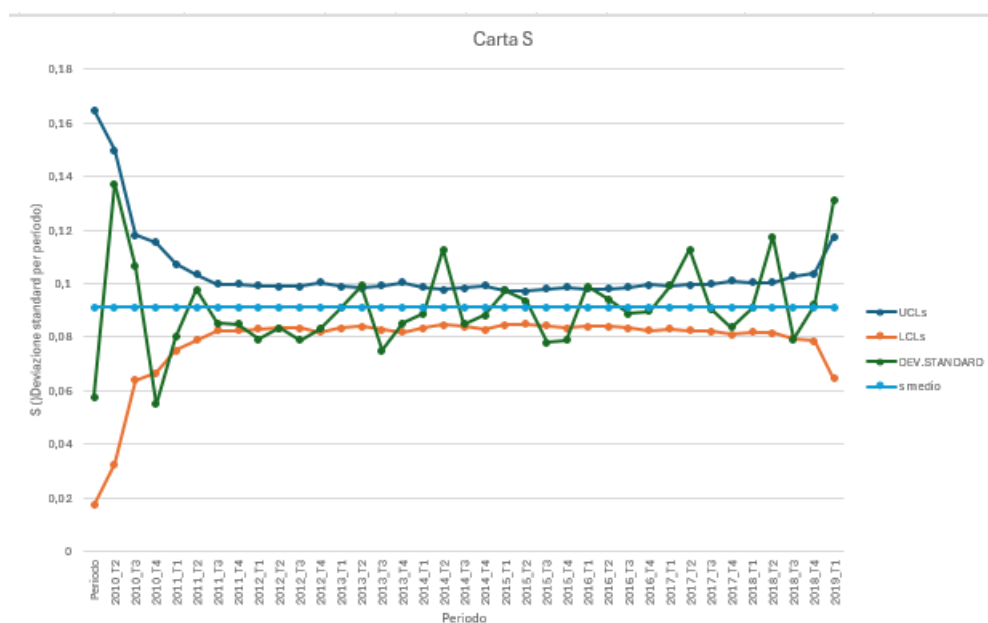


Figura 26 Esempio di Carta S per una determinante di qualità: Disney Word VS Disneyland

L'intero processo di calcolo dei residui, costruzione delle carte di controllo e determinazione dei limiti è stato automatizzato mediante uno script sviluppato in ambiente R. L'utilizzo di un approccio automatizzato ha consentito di applicare in modo sistematico la stessa procedura a tutti i topic analizzati e ai diversi parchi considerati, garantendo coerenza metodologica e riproducibilità dei risultati. (Si veda Appendice A.)

4. Analisi dei risultati Topic rappresentativi dei determinanti di qualità

L'analisi dei risultati si concentra sull'evoluzione temporale dei principali **determinanti di qualità dell'esperienza percepita** nei parchi Disney, individuati attraverso la matrice di *Topical Prevalence*. Data l'elevata dimensionalità dei topic estratti e la necessità di fornire un'interpretazione approfondita e coerente, si è scelto di focalizzare l'attenzione su un sottoinsieme di **topic rappresentativi**, selezionati per ciascun parco.

L'obiettivo non è quello di analizzare in modo descrittivo tutti i topic, ma di approfondire quelli che risultano:

- maggiormente rilevanti in termini di peso medio (IMTP),
- strutturalmente presenti nel tempo,
- presenza di punti al di fuori dei limiti di controllo

4.1 Struttura dell'analisi

Per ciascun topic selezionato l'analisi segue una procedura articolata in **tre livelli complementari**, che permettono di osservare il fenomeno sia dal punto di vista descrittivo sia dal punto di vista statistico-controllistico.

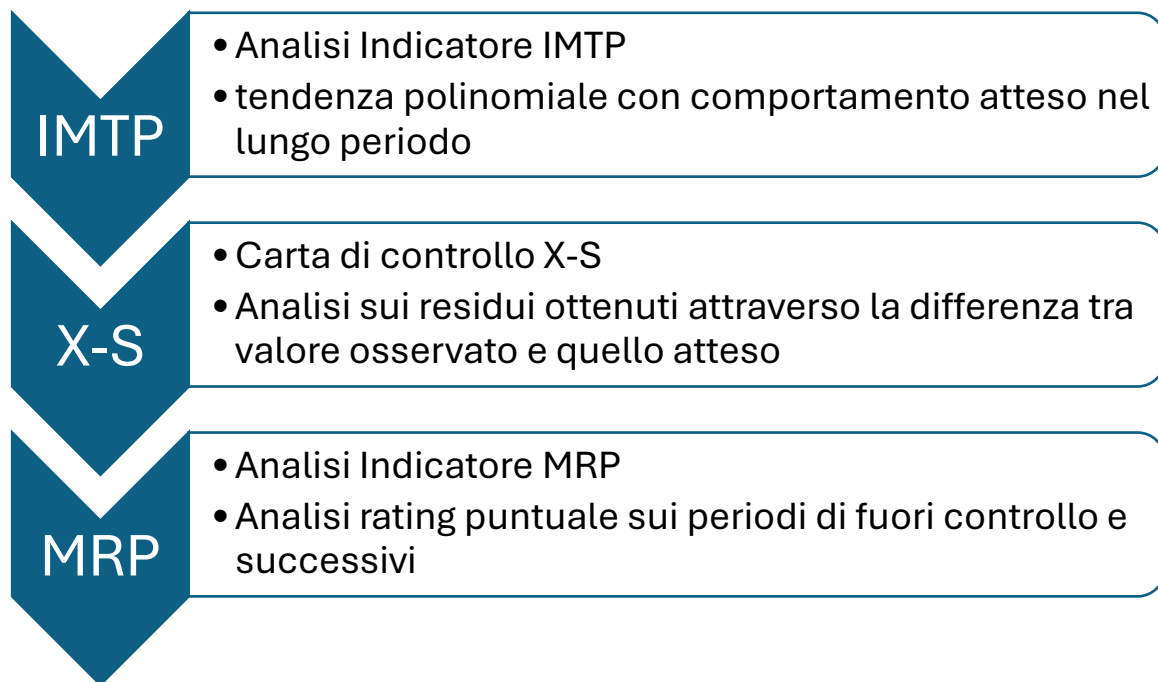


Figura 27 Struttura Analisi

1. Analisi dell'IMTP (Interval Mean Topical Prevalence)

Il primo livello di analisi riguarda l'andamento temporale dell'IMTP medio per periodo. Questo grafico consente di:

- osservare come la rilevanza del topic evolve nel tempo,
- identificare trend di crescita o declino,
- individuare eventuali discontinuità o inversioni di tendenza.

La tendenza polinomiale associata all'IMTP fornisce una rappresentazione del **comportamento atteso** del topic nel lungo periodo.

2. Carte di controllo X-S sui residui IMTP

Il secondo livello di analisi introduce un approccio di **Statistical Process Control (SPC)** applicato al contesto testuale.

- **Carta X (residui IMTP)** Analizza la deviazione del valore osservato rispetto al valore atteso (trend), consentendo di individuare periodi **fuori controllo**, ossia momenti in cui il topic viene discusso in modo anomalo rispetto al comportamento storico.
- **Carta S (deviazione standard)** Analizza la variabilità della discussione del topic nel tempo.
Un aumento della variabilità indica una percezione più eterogenea tra i visitatori, mentre una riduzione segnala maggiore uniformità.

L'utilizzo di limiti di controllo **variabili**, calcolati in funzione del numero di recensioni per periodo, garantisce una corretta interpretazione anche in presenza di campioni di dimensione diversa.

3. Analisi dell'MRP (Mean Rating Proportion)

Il terzo livello di analisi si concentra sull'**MRP**, calcolato esclusivamente per i topic selezionati, andando a descrivere esattamente il periodo del fuori controllo per capire la percezione (rating). Questo indicatore permette di:

- qualificare la natura **positiva o negativa** della discussione,
- interpretare semanticamente i periodi individuati come fuori controllo,
- comprendere se l'aumento o la diminuzione dell'IMTP sia associato a una percezione favorevole o critica.

L'MRP rappresenta quindi il collegamento finale tra **dinamica quantitativa** e **significato qualitativo**.

4.1.1 Disneyland California – Show and Parade

- Analisi IMTP: con andamento polinomiale di terzo grado

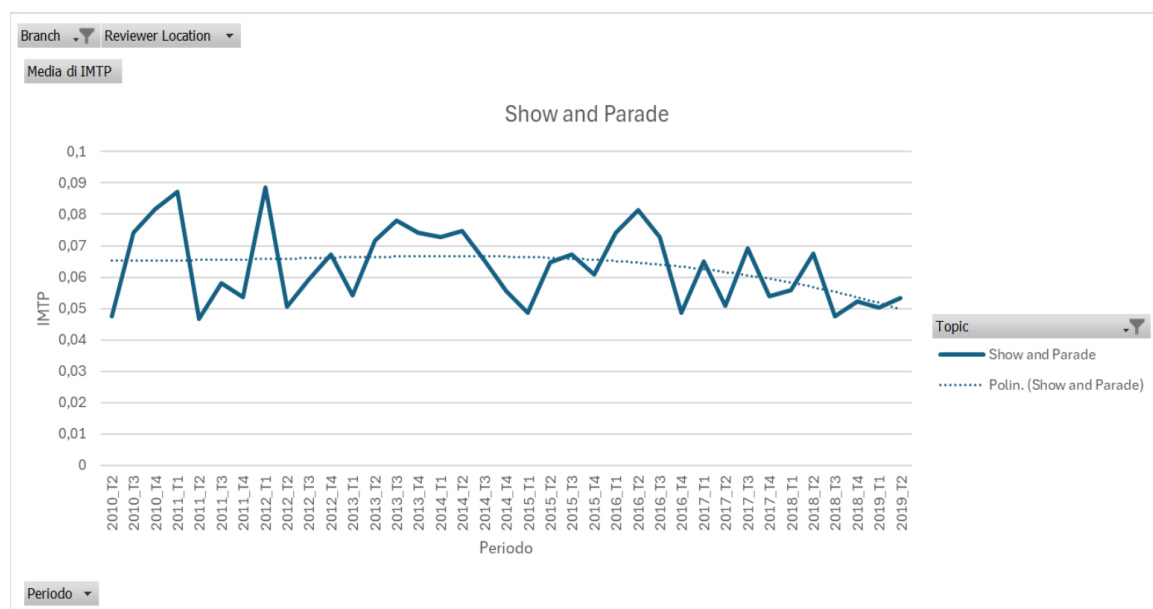


Figura 28 IMTP Show and Parade Branch Disneyland California

L'andamento dell'Interval Mean Topical Prevalence (IMTP), mostrato in Figura 28, relativo al topic "Show and Parade" mostra una dinamica caratterizzata da oscillazioni attorno a un valore medio compreso tra 0,05 e 0,07. La regressione polinomiale applicata alla serie temporale consente di descrivere il trend di fondo del fenomeno, evidenziando una lieve tendenza decrescente nel lungo periodo, che rappresenta il comportamento medio atteso del livello di discussione del topic.

Con riferimento all'arco temporale oggetto di approfondimento, compreso tra 2013_T4 e 2015_T4, l'IMTP mostra inizialmente valori prossimi all'andamento descritto dalla polinomiale. Nei periodi successivi si osservano tuttavia alcune deviazioni positive e negative rispetto al trend, che indicano variazioni temporanee della rilevanza del topic Show and Parade. In particolare, tra 2014_T3 e 2016_T2 si rileva un incremento dell'IMTP, con valori che si collocano stabilmente al di sopra dell'andamento medio, suggerendo una maggiore attenzione degli utenti verso spettacoli e parate in tale fase.

Questa analisi preliminare dell'IMTP consente quindi di individuare i periodi in cui il topic assume una rilevanza anomala rispetto al trend di lungo periodo, fornendo la base per l'analisi dei residui della regressione polinomiale.

- Analisi carta di controllo X-S:

Carta X:

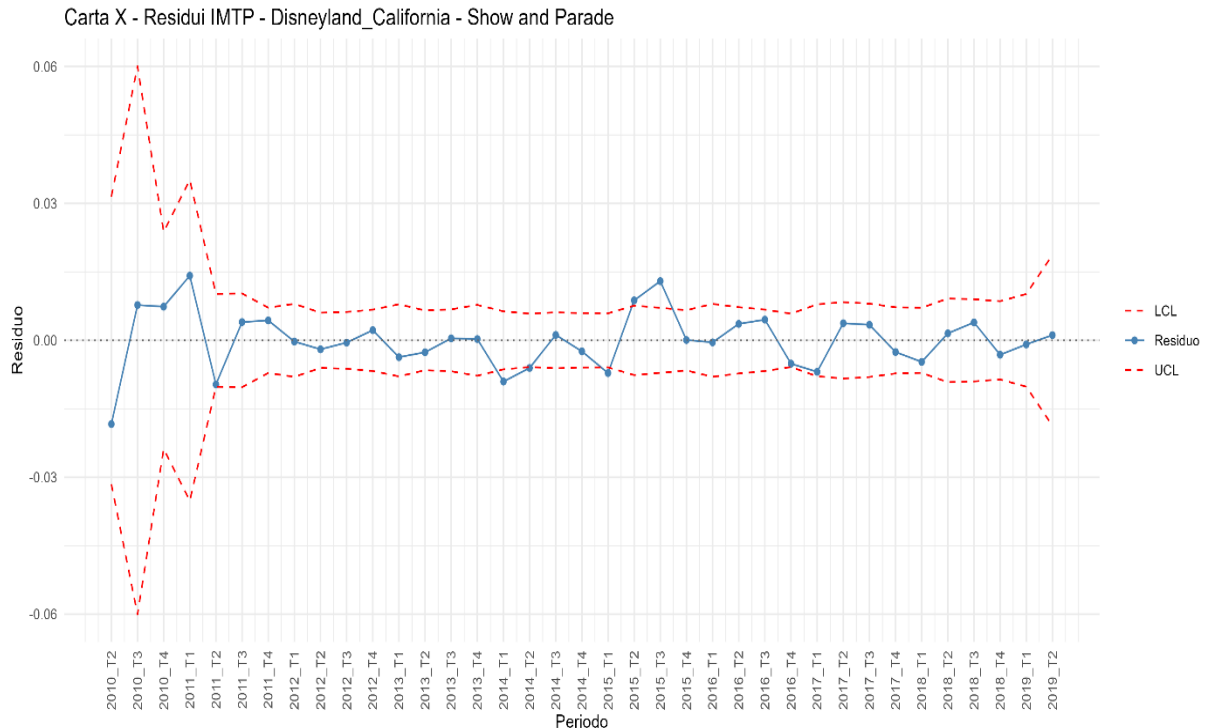


Figura 29 Carta X Show and Parade Branch Disneyland California

A partire dalla regressione polinomiale applicata all'IMTP, sono stati analizzati i residui mediante una carta X, al fine di valutare la presenza di segnali di fuori controllo nella dinamica temporale del topic Show and Parade. I residui rappresentano lo scostamento puntuale dell'IMTP rispetto al valore stimato dalla polinomiale e consentono di isolare le componenti anomale non spiegate dal trend. Nel periodo 2013_T4 – 2015_T4 si osserva, Figura 29, inizialmente un punto al di sotto del limite inferiore di controllo, che rappresenta il primo segnale di anomalia rispetto al comportamento atteso. Successivamente, nel corso del 2014, i residui rientrano all'interno dei limiti di controllo, indicando una fase di apparente stabilizzazione del processo. A partire da 2015_T2, si osservano due punti consecutivi al di sopra del limite superiore di controllo, che segnalano un incremento anomalo e persistente del livello di discussione del topic rispetto al trend polinomiale. Tali segnali indicano una condizione di fuori controllo del processo, suggerendo la presenza di fattori esterni o eventi specifici che hanno influenzato in modo significativo la Digital Voice of Customer. Nei periodi successivi, i residui rientrano progressivamente all'interno dei limiti di controllo, segnando un ritorno a condizioni di stabilità.

Carta S:

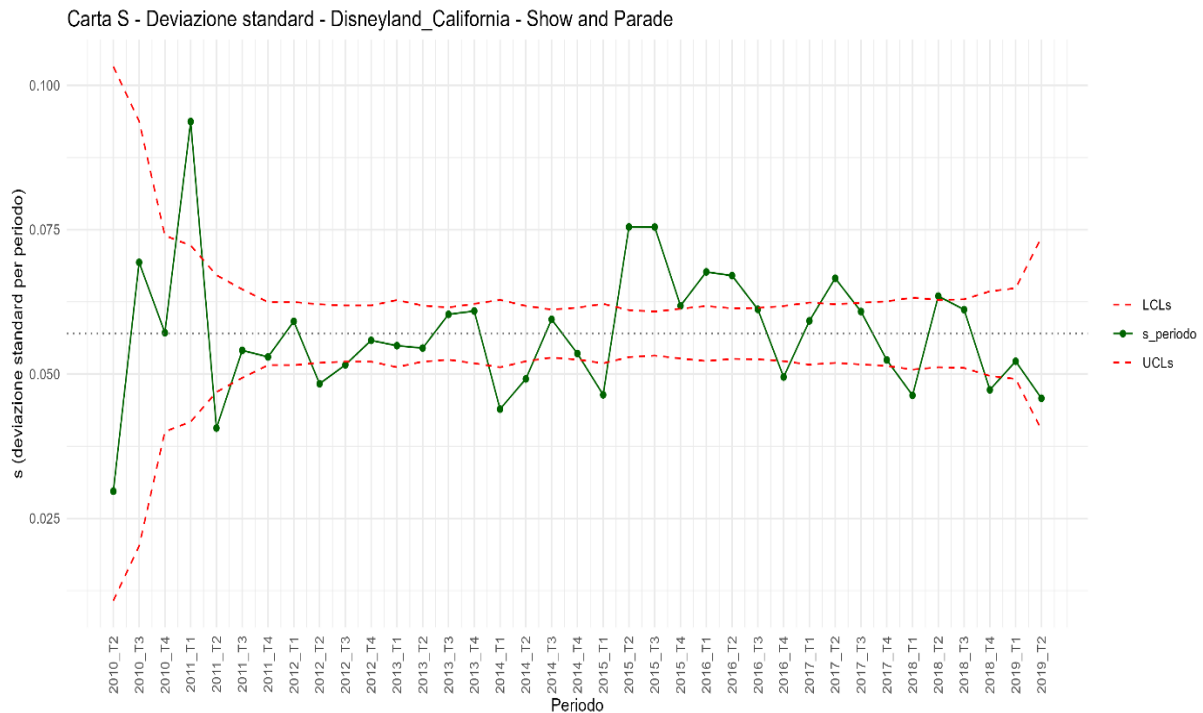


Figura 30 Carta S Topic “Show and Parade” Branch Disneyland California

L’analisi della carta S, relativa alla deviazione standard dell’IMTP per periodo, consente di valutare l’evoluzione della variabilità associata al topic Show and Parade. Nel periodo considerato, la deviazione standard mostra oscillazioni attorno a un valore medio sostanzialmente stabile, con alcuni picchi localizzati.

In particolare, nel periodo 2015_T2 – 2015_T3, Figura 30, in corrispondenza dei punti fuori controllo individuati nella carta X, si osserva un incremento della variabilità, che indica una maggiore eterogeneità delle opinioni espresse dagli utenti. Questo comportamento suggerisce che l’aumento della rilevanza del topic non è stato uniforme, ma accompagnato da percezioni differenziate da parte dei visitatori.

Il successivo rientro della deviazione standard entro i limiti di controllo conferma una progressiva normalizzazione del processo, coerente con quanto osservato nell’analisi dei residui.

- Analisi MRP (specifica per periodo di interesse)

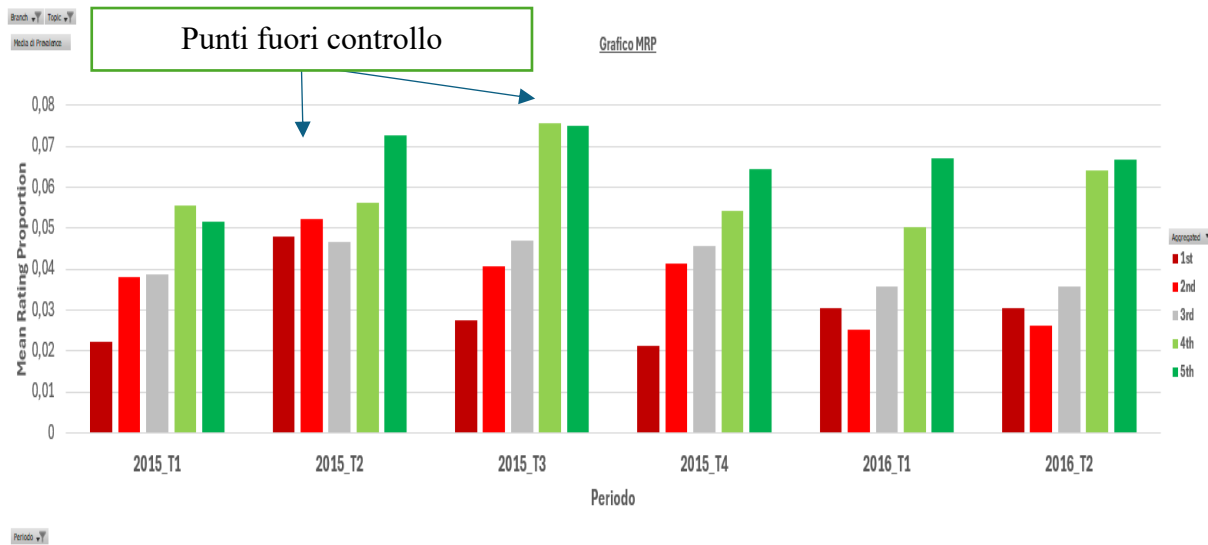


Figura 31 MRP Topic "Show and Parade" Branch Disneyland California

Per interpretare la percezione associata ai periodi di fuori controllo individuati nell'analisi dei residui dell'IMTP, è stata analizzata la *Mean Rating Proportion* (MRP) nel periodo compreso tra **2015_T1** e **2016_T2**, **Figura 31**. L'MRP consente di valutare la distribuzione delle recensioni sulle cinque classi di rating (1st, 2nd, 3rd, 4th, 5th), che variano da valutazioni **molto negative** (1st) a valutazioni **molto positive** (5 th), fornendo una misura sintetica della soddisfazione percepita dai clienti.

La dinamica che viene osservata nei periodi **2015_T2** e **2015_T3**, coincide con i due punti fuori controllo che si collocano al di sopra del limite di controllo superiore (UCL) individuati nella carta X dei residui. In questi periodi, la prevalenza delle classi di rating più elevate (4th e 5 th) indica che l'aumento della rilevanza del topic *Show and Parade* è stato accompagnato da una percezione complessivamente positiva da parte dei visitatori, soprattutto sul secondo periodo di fuori controllo dove le classi positive sono nettamente superiori raggiungendo un livello di MRP pari a 0,07.

Il successivo periodo **2015_T4**, in cui il processo rientra in controllo, mostra una stabilizzazione della distribuzione dell'MRP, con un mantenimento di livelli di soddisfazione elevati. Questo suggerisce che le condizioni che hanno determinato i fuori controllo non hanno generato un peggioramento della percezione del servizio, ma al contrario sono state assimilate positivamente dagli utenti.

Nel complesso, l'analisi dell'MRP evidenzia come i periodi di fuori controllo osservati nell'IMTP siano associati a un **cambio di percezione verso le classi di rating più elevate**, suggerendo che l'aumento dell'attenzione verso il topic *Show and Parade* sia riconducibile a elementi percepiti favorevolmente dai clienti, piuttosto che a criticità del servizio.

4.1.2 Disneyland California – Special Event’s celebration

- **Analisi dell’IMTP: con andamento polinomiale di terzo grado**

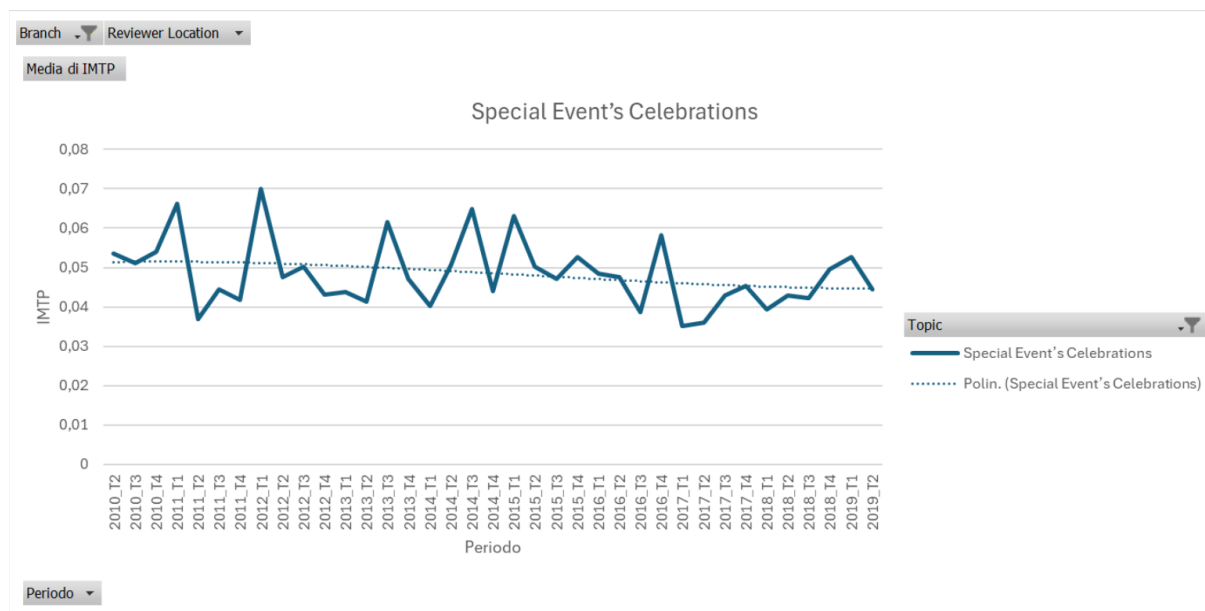


Figura 32 IMTP Topic “Special Event's celebration” Branch Disneyland California

L’andamento dell’Interval Mean Topical Prevalence (IMTP) relativo al topic Special Event’s Celebrations evidenzia una dinamica caratterizzata da una variabilità marcata rispetto ad altri topic analizzati, Figura 32. La regressione polinomiale applicata alla serie temporale descrive un trend di fondo lievemente decrescente, che rappresenta l’andamento medio della rilevanza del topic nel periodo osservato.

Con riferimento all’intervallo 2011_T2 – 2017_T4, l’IMTP presenta numerosi picchi positivi rispetto alla polinomiale, distribuiti lungo l’intero orizzonte temporale. In particolare, si osservano incrementi rilevanti nei periodi 2011_T4, 2012_T2, 2013_T4, 2014_T1, 2014_T4, 2016_T1 e 2016_T4, nei quali il livello di discussione del topic risulta sensibilmente superiore al trend di lungo periodo.

Questa configurazione suggerisce che la rilevanza del topic Special Event’s Celebrations non segua un’evoluzione lineare o strutturale, ma sia fortemente influenzata da eventi puntuali e ricorrenti, coerenti con la natura episodica delle celebrazioni e degli eventi speciali.

- **Analisi carta di controllo X-S:**

Carta X:

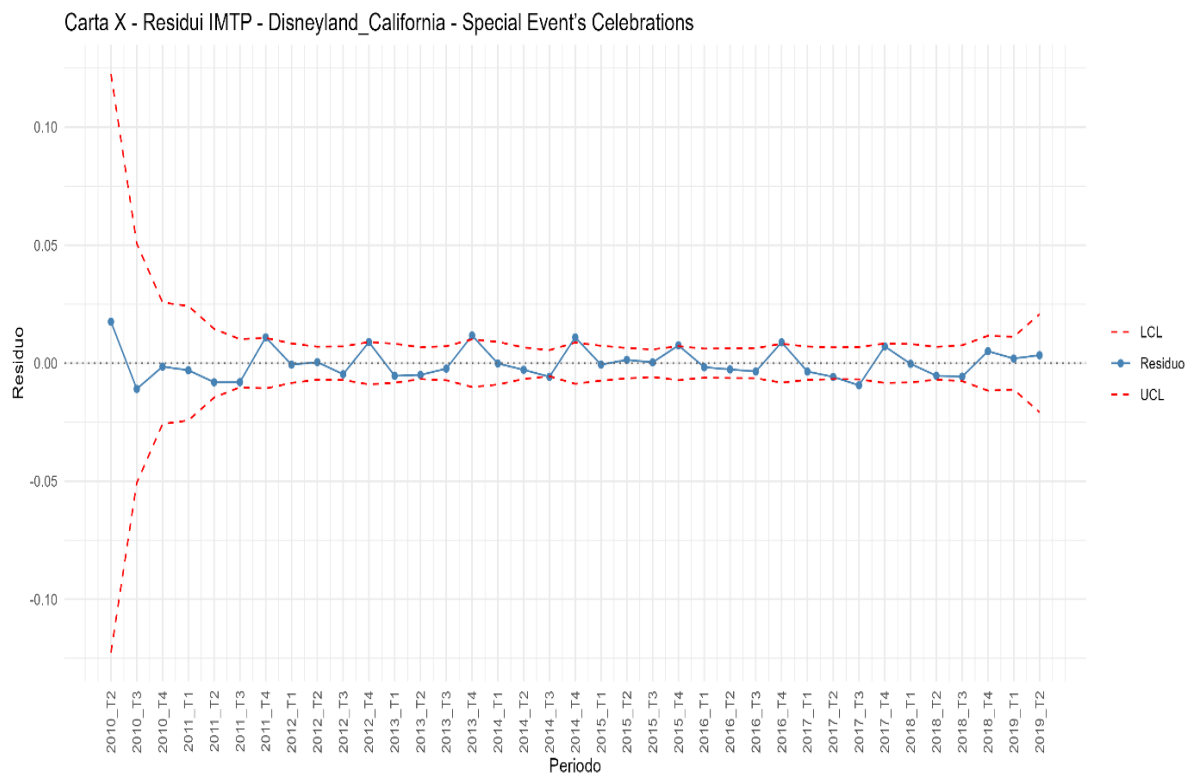


Figura 33 Carta X Topic “Special Event's Celebration” Branch Disneyland California

L’analisi dei residui della regressione polinomiale, condotta mediante carta X, consente di individuare con maggiore precisione i periodi caratterizzati da comportamenti anomali rispetto al trend atteso. Nel periodo 2011_T2 – 2017_T4 si osservano diversi punti fuori controllo, sia superiori sia inferiori, distribuiti in più fasi temporali, Figura 33.

In particolare, si rilevano residui positivi in corrispondenza dei periodi 2011_T4, 2012_T4, 2013_T4, 2014_T4 e 2016_T4, che indicano un livello di discussione del topic superiore rispetto a quanto stimato dalla polinomiale, analisi che evidenzia un comportamento particolare sul quarto quadrimestre degli anni sopra citati. Tali segnali suggeriscono incrementi improvvisi dell’attenzione degli utenti verso il topic analizzato. Particolare attenzione anche per un punto di fuori controllo al disotto del limite inferiore, in particolare nel periodo 2017_T3, suggerendo un lieve abbassamento dell’attenzione verso il topic.

Carta S:

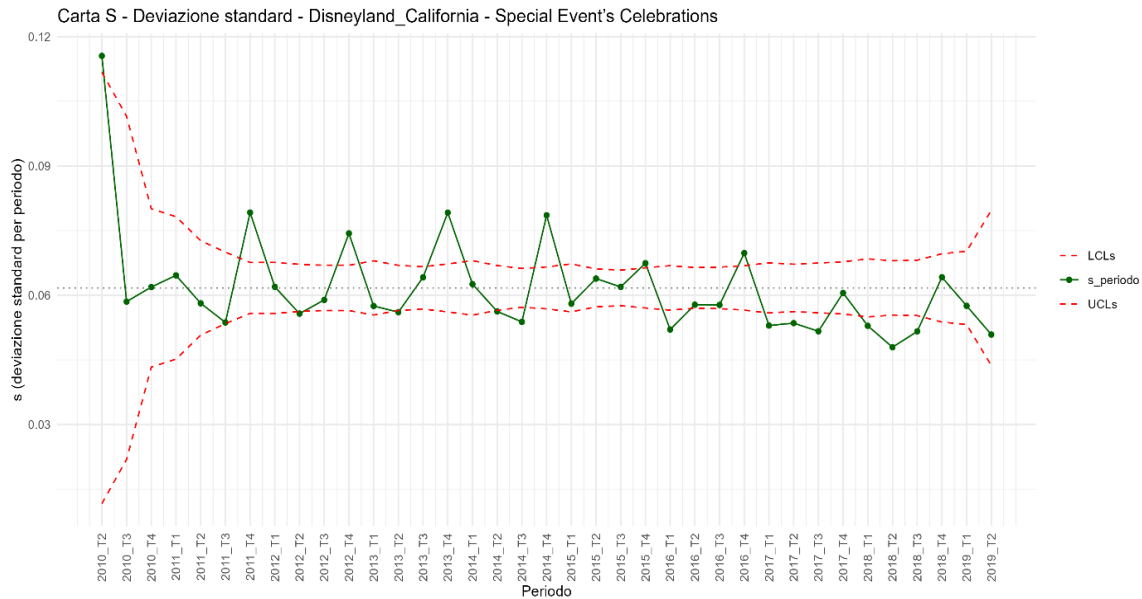


Figura 34 Carta S Topic “Special Event's Celebration” Branch Disneyland California

L'analisi della carta S, relativa alla deviazione standard dell'IMTP per periodo, mostra una variabilità generalmente stabile ma con incrementi localizzati, Figura 34, in corrispondenza di alcuni dei periodi fuori controllo individuati già nella carta X. In particolare, aumenti della deviazione standard si osservano nei periodi 2011_T4, 2012_T4, 2013_T4, 2014_T4 e 2016_T4, indicando una maggiore dispersione delle opinioni degli utenti. Effetto contrario si può osservare nei periodi finali, in particolare nei periodi che vanno dal 2017_T1 al 2018_T4, sintomo di riallineamento aziendale sulla tematica con una dispersione sotto la media delle opinioni tra gli utenti.

Questo comportamento suggerisce che, durante alcuni eventi speciali, la percezione dei visitatori risulti più eterogenea, come denota l'andamento della carta. In particolare, notiamo una particolare ricorrenza sul quarto trimestre di ogni anno studiato.

- **Analisi MRP (specifica per periodo di interesse)**

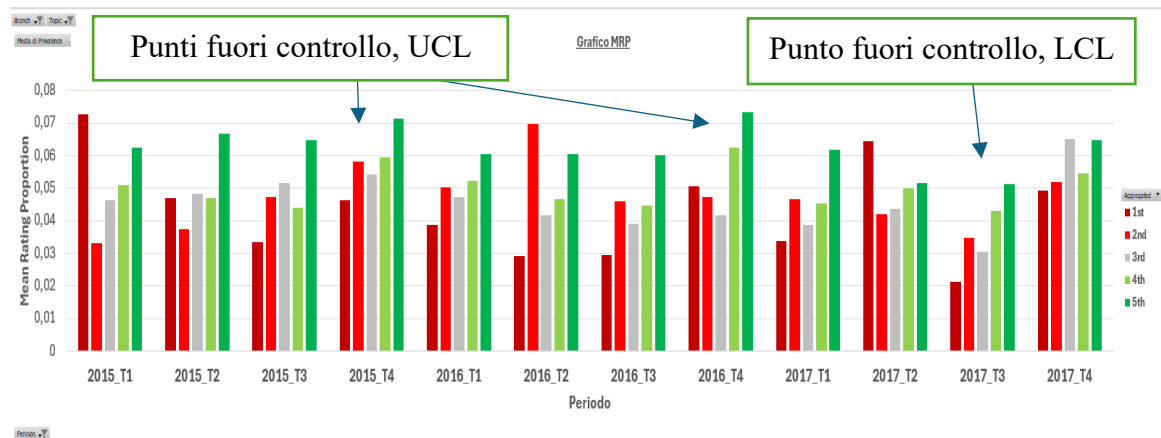


Figura 35 MRP Topic “Special Event’s Celebration” Branch Disneyland California

Attraverso lo studio della Carta X si può notare che l’andamento del topic presenta dei fuori controllo a partire dal 2011_T4 fino al 2017_T3, con ricorrenza puntuale sul quarto trimestre di ogni anno, come mostrato in Figura 33. In particolare, nell’analisi MRP condotta sul topic “Special Event’s Celebration” viene analizzato il periodo che va dal 2015_T1 al 2017_T4. Come mostrato in Figura 35, il 2015_T4 e 2016_T4 corrispondono a due punti di fuori controllo che superano il limite di controllo superiore (UCL), evidenziando in particolare nel primo periodo una presenza netta della prima classe rating (1st), indicando una percezione molto negativa del topic, con un livello medio MRP pari 0,05, nonostante una buona presenza delle classi positive (4th e 5th) e neutra (3rd). Nel secondo periodo di fuori controllo, si può notare un lieve miglioramento, la classe rating negativa (2nd) subisce una lieve riduzione di circa un punto percentuale, passando da un valore prossimo allo 0,06 a 0,05 a favore della quarta classe rating (4th) e quinta (5th) che tocca il picco della serie temporale osservata, con un valore > 0.07 . Nel 2017_T3 periodo di fuori controllo al di sotto del limite di controllo inferiore (LCL) si nota una riduzione generale del valore MRP, congruo rispetto ai valori osservati nella carta X. Nonostante il livello MRP più basso rispetto ai trimestri precedenti, prevalgono le classi rating 4th e 5th, con una valutazione complessivamente positiva del topic. Dal 2017_T3 i valori rientrano all’interno del range dei limiti (Figura 33) con una valutazione più omogenea tra i visitatori. Figura 34 e l’indicatore MRP che nel 2017_T4 ci mostra una situazione più equilibrata tra le cinque classi rating.

4.1.3 Disneyland Hong-Kong – Judgment on Staff

- Analisi dell'IMTP: con andamento polinomiale di terzo grado

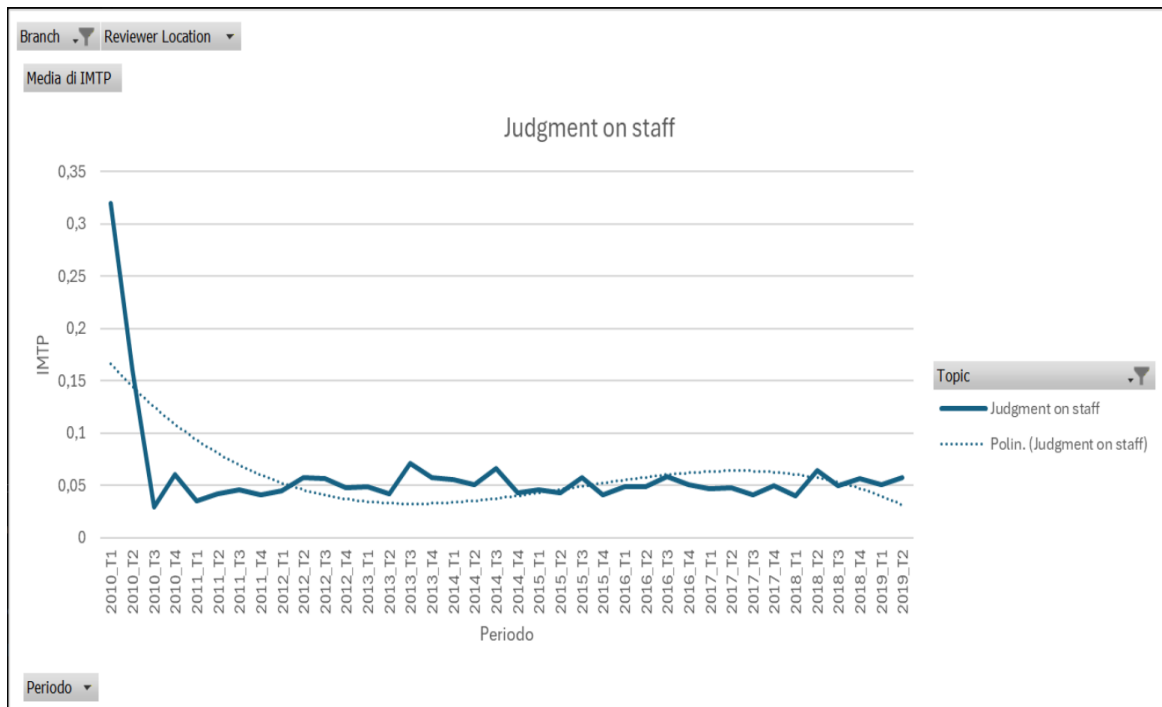


Figura 36 IMTP Topic “Judgment on staff” Branch Disneyland Hong-Kong

L’andamento dell’Interval Mean Topical Prevalence (IMTP), mostrato in Figura 36, relativo al topic Judgment on staff mostra una dinamica fortemente influenzata dai primi periodi della serie temporale. In particolare, nel periodo 2010_T1 si osserva un valore di IMTP eccezionalmente elevato, che determina una forte discontinuità rispetto ai periodi successivi.

La regressione polinomiale applicata alla serie evidenzia una rapida decrescita iniziale, seguita da una fase di stabilizzazione a partire dal 2010_T3, con valori di IMTP che si assestano intorno a livelli significativamente più contenuti. Nel lungo periodo, il trend descritto dalla polinomiale risulta sostanzialmente stabile, suggerendo che, dopo la fase iniziale, la rilevanza del topic Judgment on staff rimanga costante nel tempo.

Questa configurazione indica che l’elevata attenzione iniziale verso il personale non rappresenta un comportamento strutturale, ma un fenomeno transitorio che si esaurisce rapidamente nei periodi successivi.

• Analisi Carta di controllo X-S:

Carta X:

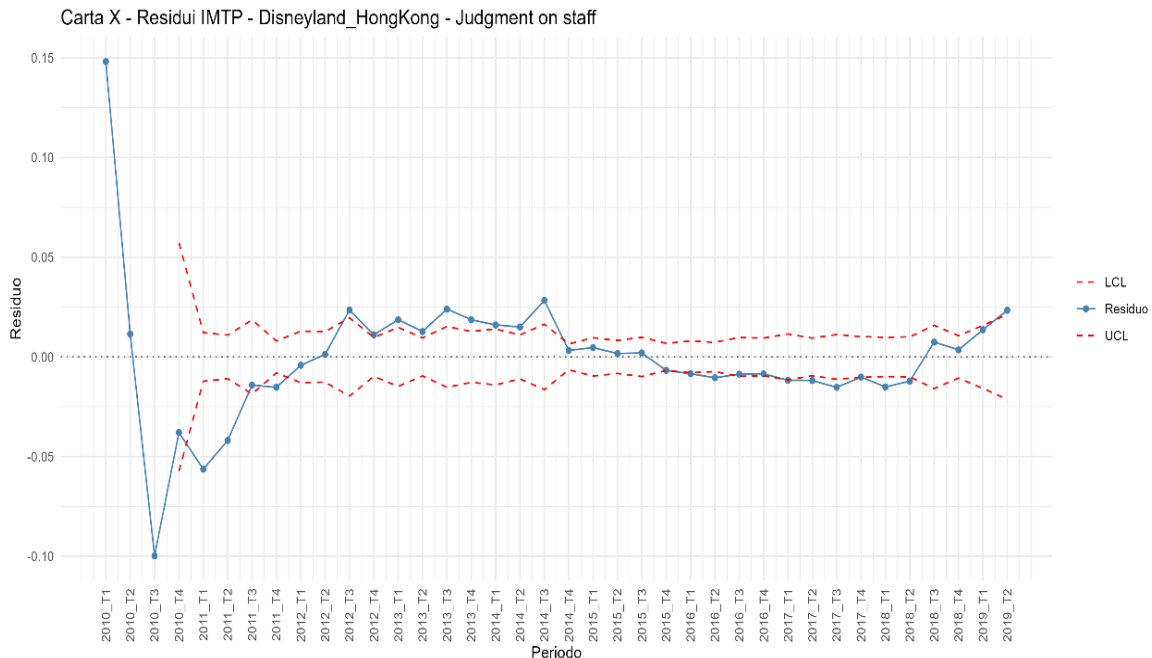


Figura 37 Carta X Topic “Judgment on Staff” Branch Disneyland Hong-Kong

La Carta X dei residui dell’IMTP evidenzia una marcata instabilità del processo, con numerosi punti al di fuori dei limiti di controllo lungo la serie temporale. In particolare, osservando la Figura 37, nella fase iniziale si osserva un residuo fortemente positivo nel periodo 2010_T1, che supera nettamente il limite superiore di controllo (UCL), indicando un livello di discussione, sul tema Judgment on staff, significativamente più elevato rispetto a quanto stimato dalla regressione polinomiale. Questo segnale è seguito, nei periodi immediatamente successivi, da una rapida inversione con residui che scendono al di sotto del limite inferiore di controllo (LCL), in particolare nel periodo 2010_T4 e nei primi trimestri del 2011. Tali scostamenti indicano una fase di forte discontinuità, in cui il topic passa da una condizione di attenzione eccezionalmente elevata a una di sottorappresentazione rispetto al trend atteso. Successivamente, pur in presenza di una maggiore stabilità complessiva, la Carta X continua a mostrare episodi puntuali di fuori controllo, prevalentemente in direzione positiva. In particolare, residui oltre il limite superiore di controllo (UCL) si osservano nei periodi compresi tra il 2012_T3 e 2014_T3, indicando incrementi improvvisi e statisticamente significativi della prevalenza del topic non spiegati dall’andamento di fondo, in particolare un comportamento riconducibile al Plateau con passaggio da trend crescente a stabilizzazione (Figura 5). Dal 2015_T3 al 2018_T2 notiamo una netta inversione, in quanto vi è la presenza di più punti al di sotto del limite inferiore di controllo (LCL) suggerendo un netto calo della prevalenza del topic rispetto al trend stimato verso una fase stabile, Stabilization Figura 5. Nella parte finale della serie temporale, tra 2018_T4 e 2019_T2, si rileva un nuovo avvicinamento e superamento del limite superiore, suggerendo un ritorno di attenzione verso il tema del giudizio sul personale. Nel complesso, la presenza sia di fuori controllo positivi sia negativi conferma che il comportamento del topic è caratterizzato da fasi in cui intervengono cause speciali, rendendo

l'analisi dei periodi fuori controllo centrale per l'interpretazione della dinamica della Digital Voice of Customer.

Carta S:

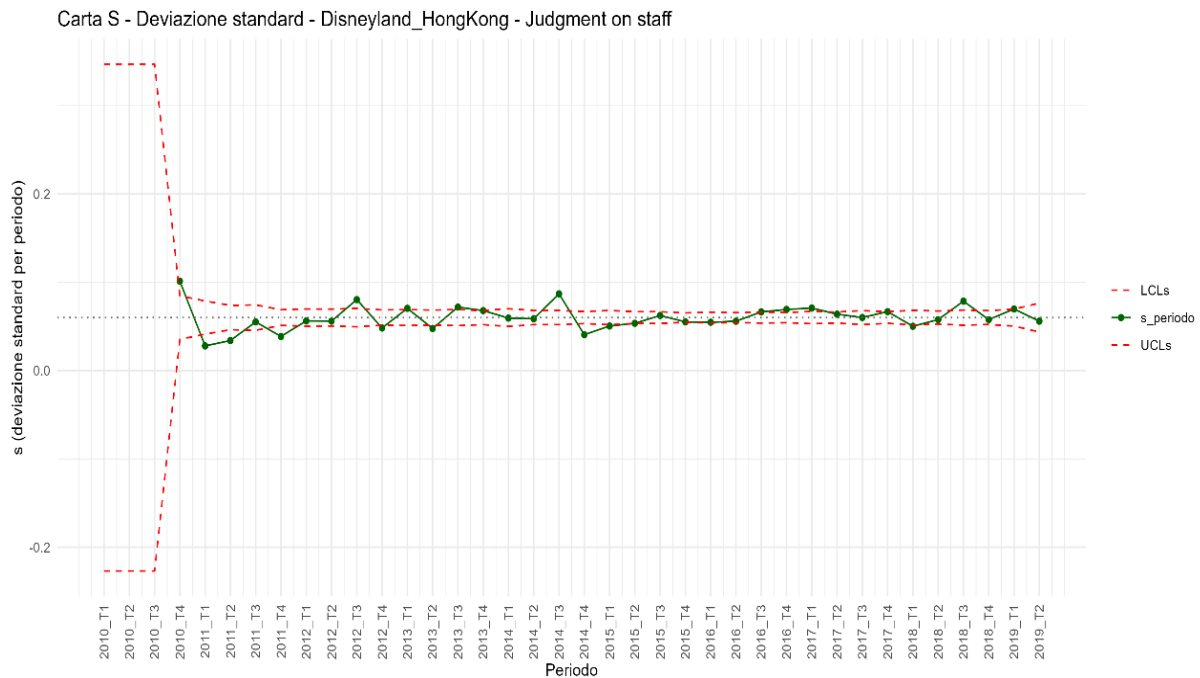


Figura 38 Carta S Topic “Judgment on Staff” Branch Disneyland Hong-Kong

L'analisi della Carta S, relativa alla deviazione standard dell'IMTP per periodo, fornisce un supporto fondamentale all'interpretazione dei segnali emersi nella Carta X. Nei primi periodi della serie, come mostrato nella Figura 38, in particolare tra 2010_T1 e 2010_T3, la deviazione standard assume valori anomali e si colloca al di fuori dei limiti di controllo, indicando una variabilità delle opinioni espresse dagli utenti sul personale. Questa elevata dispersione è coerente con la presenza di residui fuori controllo in entrambe le direzioni e suggerisce una forte eterogeneità delle esperienze vissute dai visitatori in quella fase. A partire dal 2011, la deviazione standard si riduce sensibilmente e rientra stabilmente all'interno dei limiti di controllo, mantenendosi su livelli contenuti per gran parte dell'orizzonte temporale. Tuttavia, in corrispondenza di alcuni dei periodi caratterizzati da residui oltre il limite superiore (ad esempio 2012_T3 e 2014_T3), si osservano lievi incrementi della variabilità, pur senza superamenti persistenti dei limiti. Questo comportamento indica che, durante i picchi di attenzione verso il topic, l'aumento di prevalenza è talvolta accompagnato da una maggiore dispersione delle valutazioni, ma non sempre da una perdita completa di controllo del processo. Nel complesso, la Carta S conferma l'esistenza di una fase iniziale fortemente instabile e di una successiva fase di controllo statistico, all'interno della quale permangono tuttavia segnali puntuali associati ai periodi di fuori controllo individuati nella Carta X.

- **Analisi MRP: (specifica per periodo di interesse)**

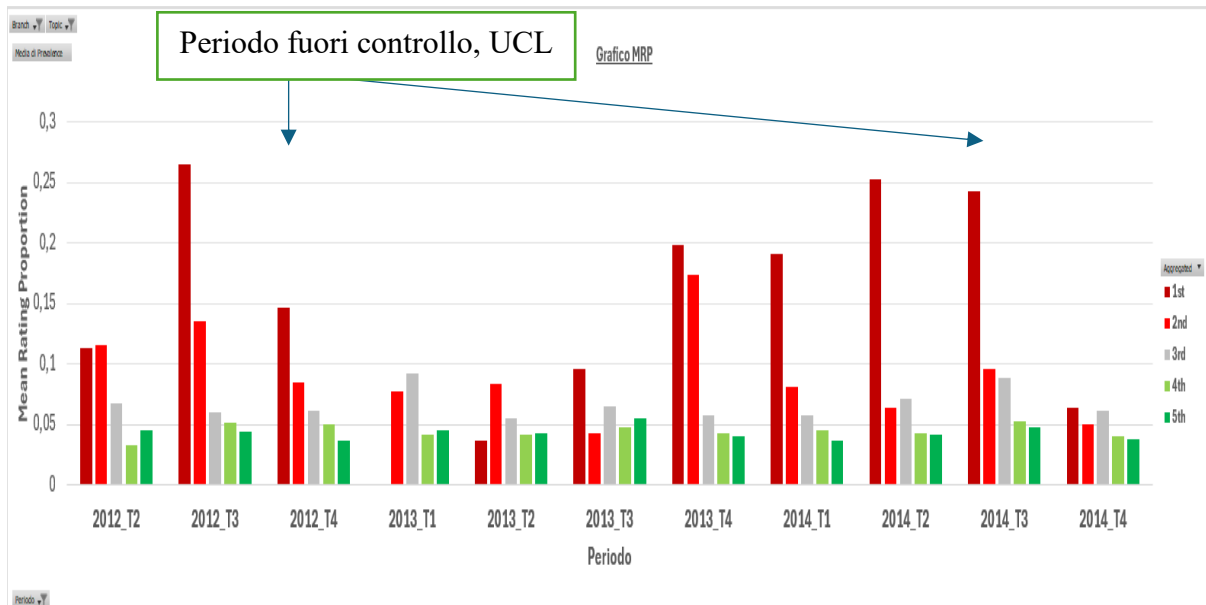


Figura 39 MRP Topic: “Judgment on Staff” Prima fase fuori controllo Branch Disneyland Hong-Kong

L’analisi della Mean Rating Proportion (MRP), articolata sulle cinque classi ordinali di giudizio (1st, 2nd, 3rd, 4th e 5th), mostrata in Figura 39, ci permette di interpretare i primi fuori controllo, dopo una prima fase di stabilizzazione. Sono dei periodi caratterizzati da fuori controllo oltre il limite di controllo superiore (UCL) nella Carta X, in particolare dal periodo 2012_T3 al 2014_T3. L’MRP evidenzia una distribuzione polarizzata tra le classi ma comunque fortemente negativa. In questi trimestri, in particolare nel 2012_T3 la serie temporale tocca il valore più alto con la prima classe (1st) raggiungendo un valore pari a circa lo 0,25, consolidando una discussione fortemente negativa nei confronti del topic “Judgment on staff”. Nell’ultimo periodo di fuori controllo 2014_T3, si nota un leggero miglioramento con una riduzione di circa un punto percentuale delle classi 1st e 2nd a favore della classe neutra 3rd e 5th, congruo con l’andamento della carta X-S nel primo periodo in cui la serie rientra all’interno del range, 2014_T4, notiamo in Figura 39 una situazione più equilibra tra le classi. Questa polarizzazione è coerente con gli incrementi di variabilità osservati nella Carta S e suggerisce una percezione non uniforme della qualità del servizio.

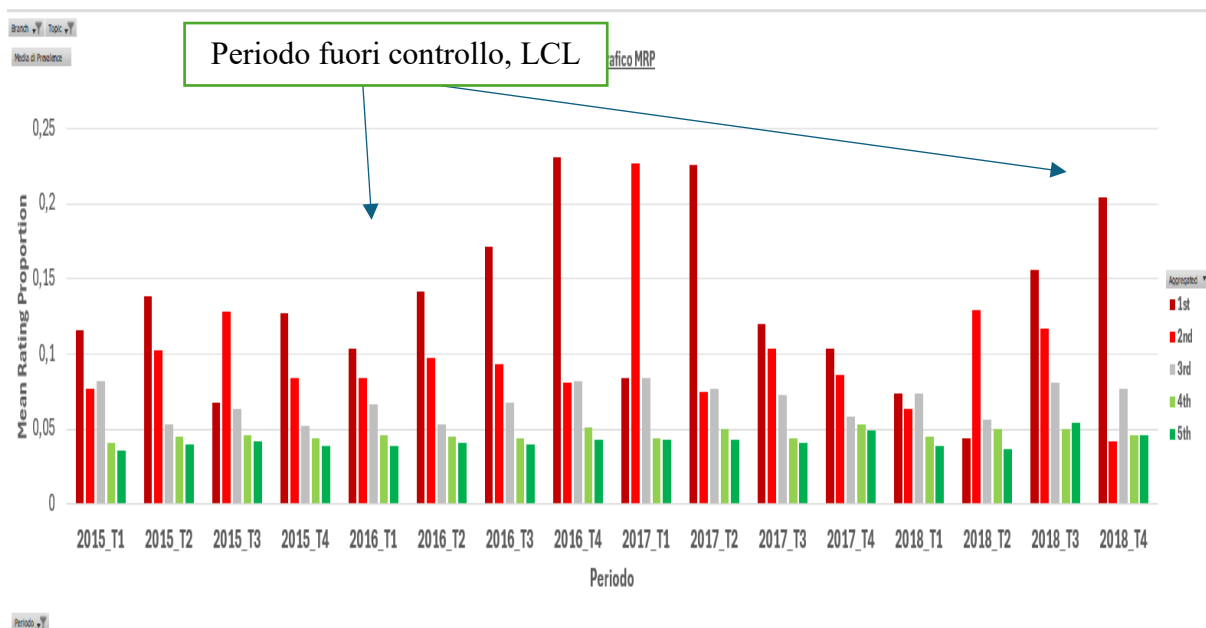


Figura 40 MRP Topic "Judgment on staff" fase centrale presenza di punti al di sotto del LCL Branch Disneyland Hong-Kong

Nella fase centrale, invece, come mostrato in Figura 40, in corrispondenza dei residui al di sotto del limite di controllo inferiore (LCL) a partire dal 2015_T4 fino al 2018_T4 notiamo ancora una netta polarizzazione, orientata verso classi più basse (1st e 2nd) in particolare nel periodo compreso tra il 2016_T4 e 2017_T2 la serie raggiunge il suo massimo con le classi 1st e 2nd, prossime al valore di 0,20. Quindi, nonostante il valore di discussione scende al di sotto del limite inferiore, non notiamo nessun miglioramento, anzi viene consolidato un trend di discussione molto negativo nei confronti del topic.

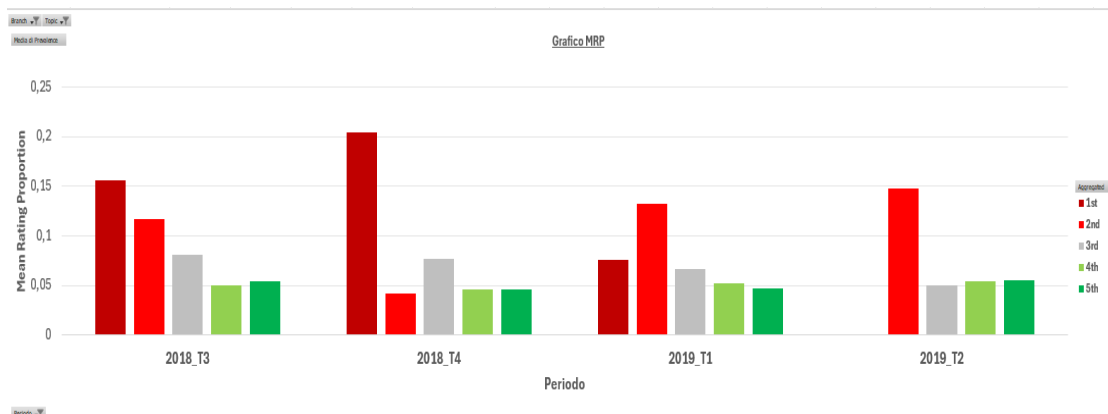


Figura 41 MRP Topic "Judgment on staff" fase finale punti all'interno dei limiti nella carta X Branch Disneyland Hong-Kong

Verso la fine del periodo analizzato invece notiamo una situazione che migliora leggermente tra 2018_T4 e 2019_T2, la distribuzione dell'MRP risulta meno estrema, come mostrato in Figura 41, rispetto ai primi anni, ma comunque orientata verso le classi di giudizio negative. In particolare, nel 2019_T1 notiamo una riduzione di circa un punto percentuale della prima classe

rating (1st) che passa da un valore di 0,20 a 0,10, fino ad essere totalmente assente nel 2019_T2, periodo in cui notiamo Figura 37 una rinnovata attenzione verso il topic.

Nel complesso, l'analisi dell'MRP conferma che i punti fuori controllo individuati nelle carte di controllo dei residui corrispondono a cambiamenti significativi nella distribuzione delle classi di giudizio. I fuori controllo positivi risultano associati a fasi in cui la percezione del Topic era molto negativa, segno di un eventuale problematica presente all'interno del comparto del personale. Successivamente negli anni più recenti le classi di rating negative sembrano parzialmente più basse e ci si orienta verso una discussione positiva della determinante di qualità.

4.1.4 Disneyland Paris – Disneyland Trip for all ages

- **Analisi dell'IMTP: con andamento polinomiale di terzo grado**

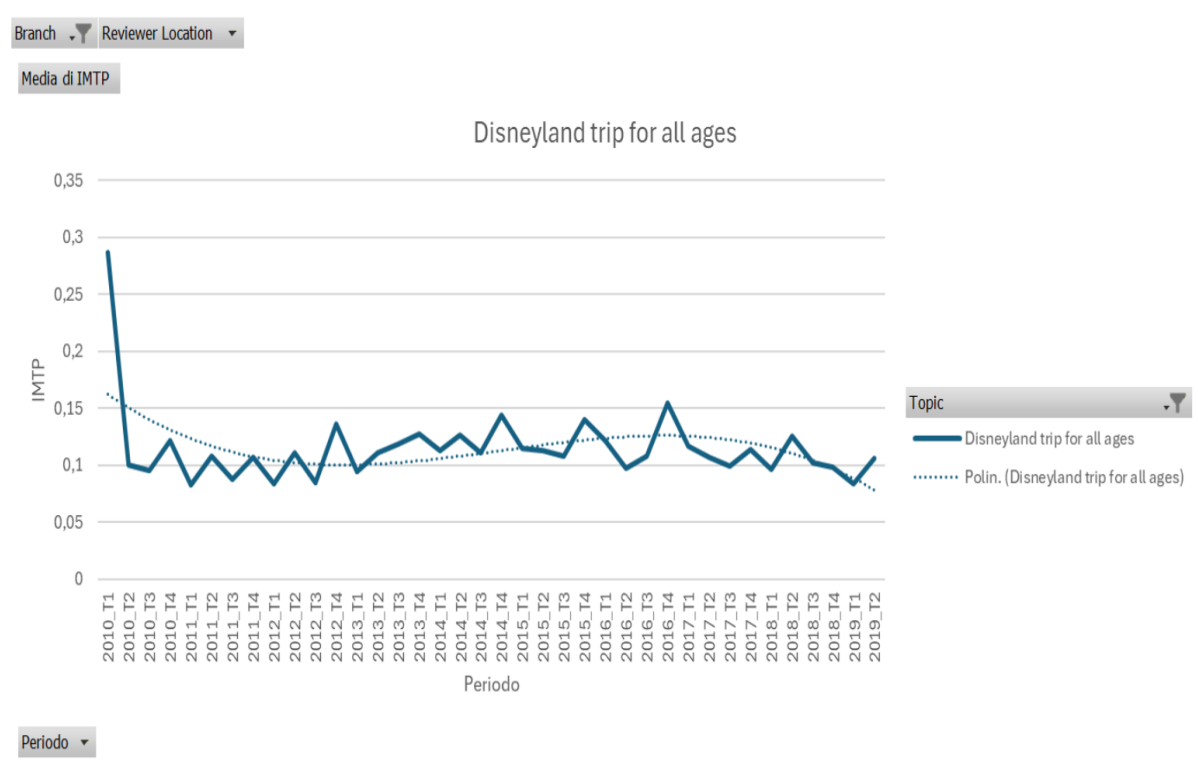


Figura 42 IMTP Topic: “Disneyland Trip for all ages” Branch Disneyland Paris

L'andamento dell'IMTP relativo al topic “Disneyland trip for all ages” mostra, come si può notare nella Figura 42, una dinamica caratterizzata da un valore iniziale estremamente elevato nel periodo 2010_T1, seguito da un brusco ridimensionamento già nei trimestri successivi e da una successiva stabilizzazione su livelli inferiori ma relativamente costanti nel tempo. La regressione polinomiale applicata alla serie intercetta questa dinamica evidenziando una forte decrescita iniziale, che può essere interpretata come una fase di assestamento del tema all'interno delle recensioni, seguita da un andamento più stabile nel medio-lungo periodo per finire con una fase leggermente decrescente a livello di discussione del Topic.

- **Analisi Carta di controllo X-S:**

Carta X:

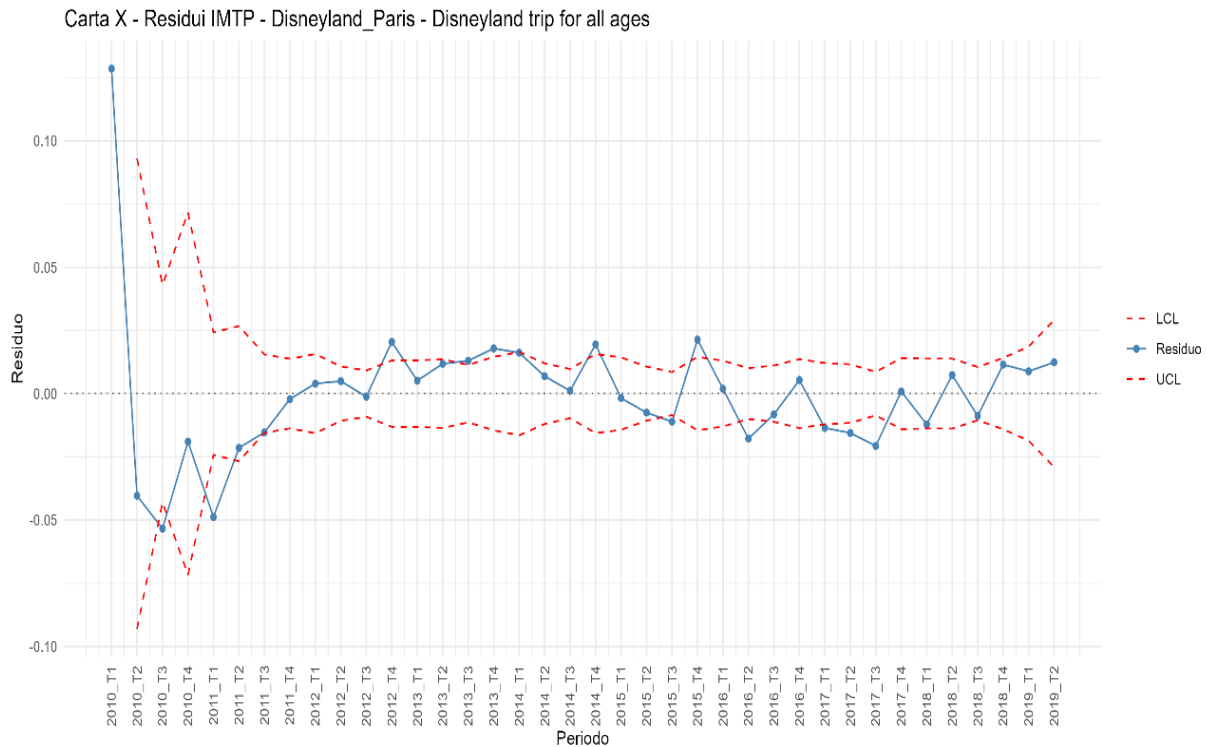


Figura 43 Carta X Topic: “Disneyland Trip for all ages” Branch Disneyland Paris

La Carta X dei residui evidenzia una chiara fase di instabilità, Figura 43. Nella parte iniziale della serie temporale, con numerosi punti al di fuori dei limiti di controllo. Il periodo 2010_T1 presenta un residuo fortemente positivo, nettamente oltre il limite superiore di controllo, coerente con il picco osservato nell’IMTP. Questo segnale indica una prevalenza del topic significativamente superiore a quanto previsto dalla regressione polinomiale e suggerisce una fase in cui il tema “Disneyland come esperienza adatta a tutte le età” risulta eccezionalmente centrale nella narrazione degli utenti, risultato ottenuto anche per la poca presenza di recensioni nel periodo.

Nei periodi immediatamente successivi (2010_T2 – 2011_T1) si osserva una sequenza di residui negativi, al di sotto del limite di controllo inferiore (LCL). Questa inversione indica una fase in cui il topic risulta sottorappresentato rispetto al trend stimato, probabilmente come effetto di normalizzazione dopo l’enfasi iniziale. Tali fuori controllo negativi segnalano una discontinuità significativa nel processo, che non può essere interpretata come semplice variabilità casuale.

A partire dal 2011_T2, la maggior parte dei residui rientra all'interno dei limiti di controllo, indicando una maggiore stabilità del processo. Tuttavia, si osservano comunque episodi puntuali di residui prossimi o superiori all'UCL (ad esempio intorno al 2012–2013 e in alcuni periodi successivi), che suggeriscono riattivazioni temporanee dell'attenzione verso il topic. Questi segnali, pur meno estremi rispetto al 2010, indicano che il tema mantiene la capacità di emergere in modo anomalo in specifici periodi.

Nel periodo che intercorre dal 2016_T2 al 2017_T3 si osserva un leggero calo rispetto al trend stimato, in particolare con tre punti sotto al limite di controllo inferiore (LCL), per poi rientrare nella stabilità ordinaria del processo.

Carta S:

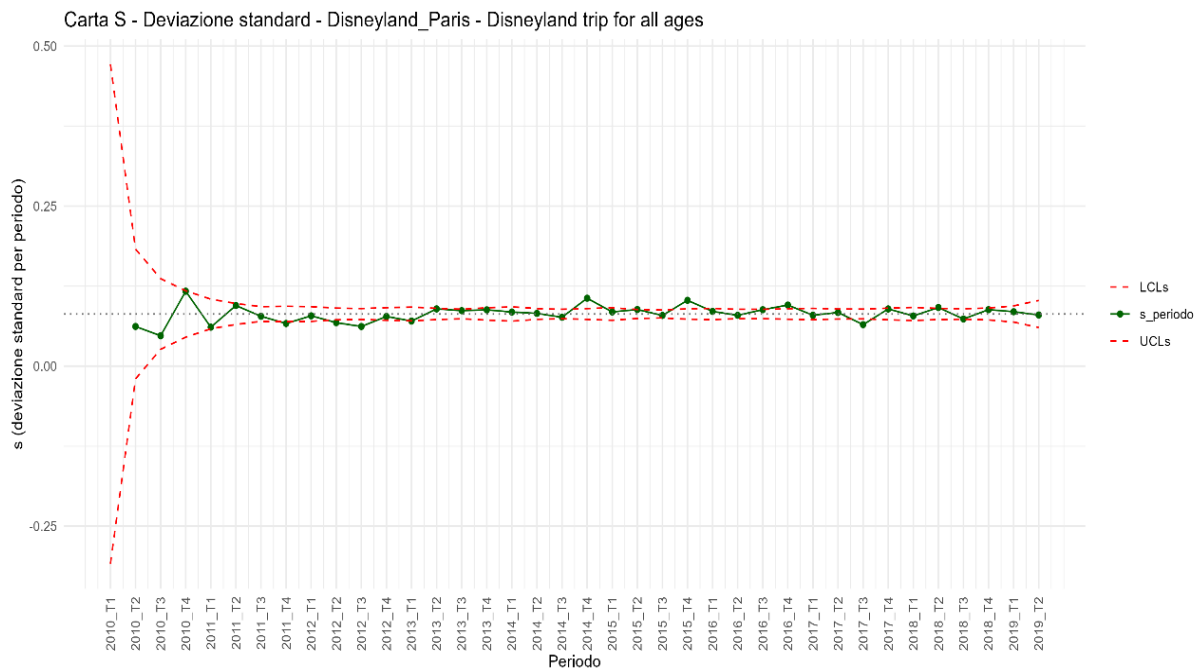


Figura 44 Carta S Disneyland Trip for all ages

Osservando la Carta S, in Figura 44, si colloca stabilmente all'interno dei limiti mantenendosi su livelli relativamente costanti per il resto della serie temporale. Nei periodi in cui la Carta X evidenzia residui positivi, oltre il limite UCL, si osservano lievi aumenti della variabilità (2014_T4), che tuttavia non compromettono il controllo complessivo del processo. Questo suggerisce che, nelle fasi di rinnovata attenzione verso il topic, l'aumento di prevalenza non è accompagnato da una dispersione eccessiva delle opinioni, ma piuttosto da un rafforzamento relativamente omogeneo del messaggio.

- **Analisi MRP (specifica per periodo di interesse)**

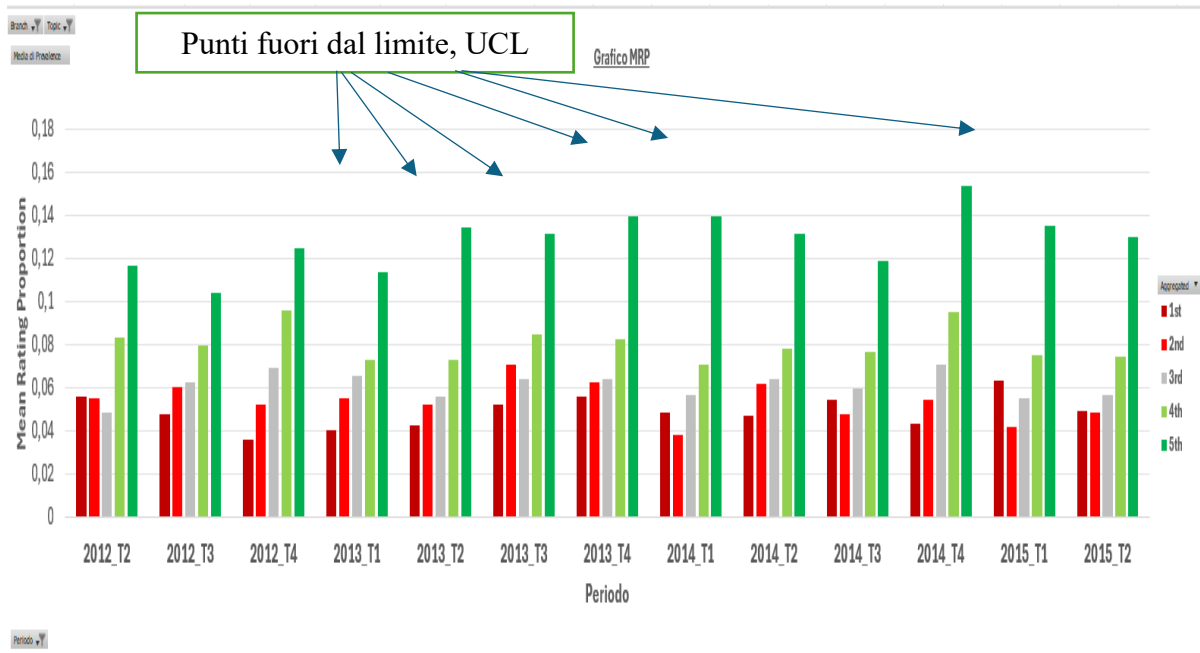


Figura 45MRP Topic: "Disneyland trip for all ages" Branch Disneyland Paris

L'analisi del Mean Rating Proportion (MRP), legata alla serie temporale che va dal 2012_T2 al 2015_T2, Figura 45, consente di interpretare come i periodi fuori controllo individuati nella Carta X, oltre UCL siano stati percepiti dagli utenti in maniera positiva. Le prime fasi, appartenenti ai periodi 2012_T2 e 2012_T3, sono antecedenti al primo punto fuori controllo, come mostrato in Figura 44. Notiamo che in queste fasi le classi negative sono presenti in una buona percentuale, in particolare 2nd con un valore pari a 0,06, nonostante un'ottima percentuale delle classi positive, in particolare, la quinta classe (5th) con un valore di 0,12. Successivamente si nota, nella Figura 42, un punto fuori dal limite superiore, corrispondente al 2012_T4, accompagnata da una crescita delle classi positive 4th e 5th con un aumento di mezzo punto percentuale rispetto alla prima e alla seconda (1st e 2nd) classe rating. La serie temporale osservata raggiunge il valore massimo (prossimo allo 0,16) in corrispondenza del 2014_T4, punto al di sopra del limite di controllo superiore, andando a rafforzare la valutazione positiva della determinante di qualità.

4.1.5 Disneyland Paris – Ticket Purchase

- **Analisi IMTP: con andamento polinomiale di terzo grado**

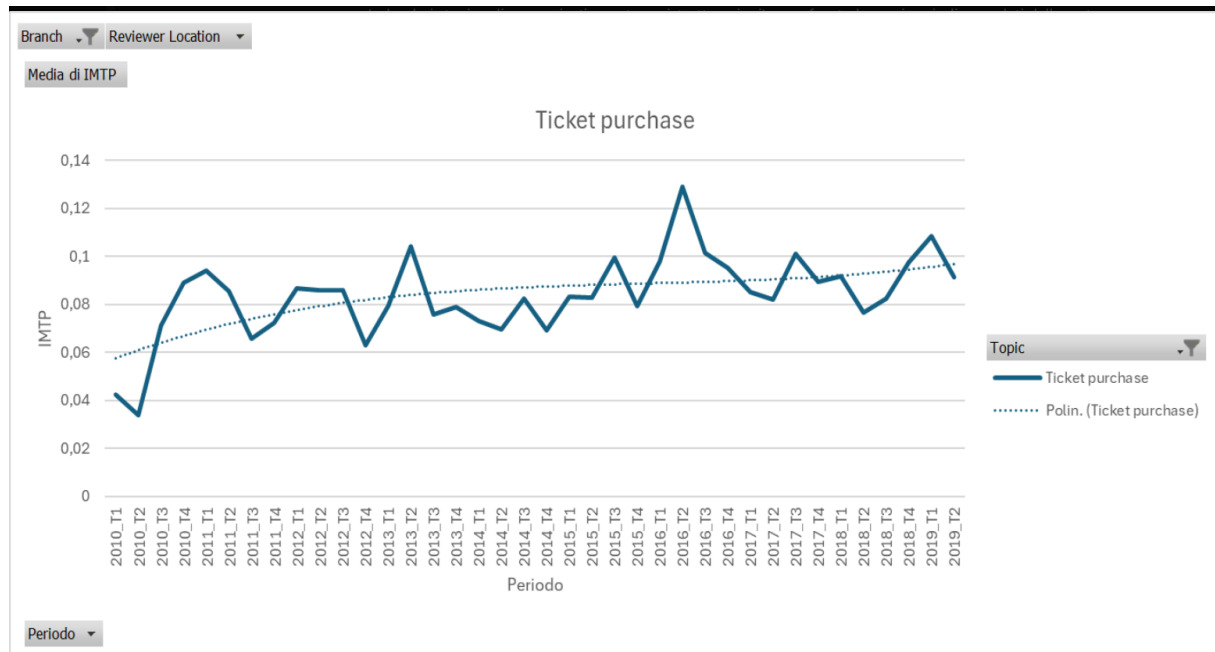


Figura 46 IMTP Topic: “Ticket Purchase” Branch Disneyland Paris

L’andamento dell’IMTP relativo al topic Ticket purchase mostra una dinamica complessivamente crescente lungo l’orizzonte temporale considerato. Dopo una fase iniziale caratterizzata da valori relativamente bassi (2010_T1–2010_T2), come mostrato in Figura 46, la prevalenza del topic aumenta progressivamente, con oscillazioni locali ma con un chiaro trend di fondo positivo. La regressione polinomiale applicata alla serie cattura questo comportamento, descrivendo un incremento graduale dell’attenzione verso il tema dell’acquisto dei biglietti, coerente con una crescente centralità di questo aspetto nell’esperienza dei visitatori. Notiamo un forte picco dell’attenzione localizzato nel periodo 2016_T2 con un valore maggiore di 0,12, periodo in cui la prevalenza del topic si discosta in maniera significativa rispetto al valore IMTP osservato.

• Analisi carta di controllo X-S

Carta X:

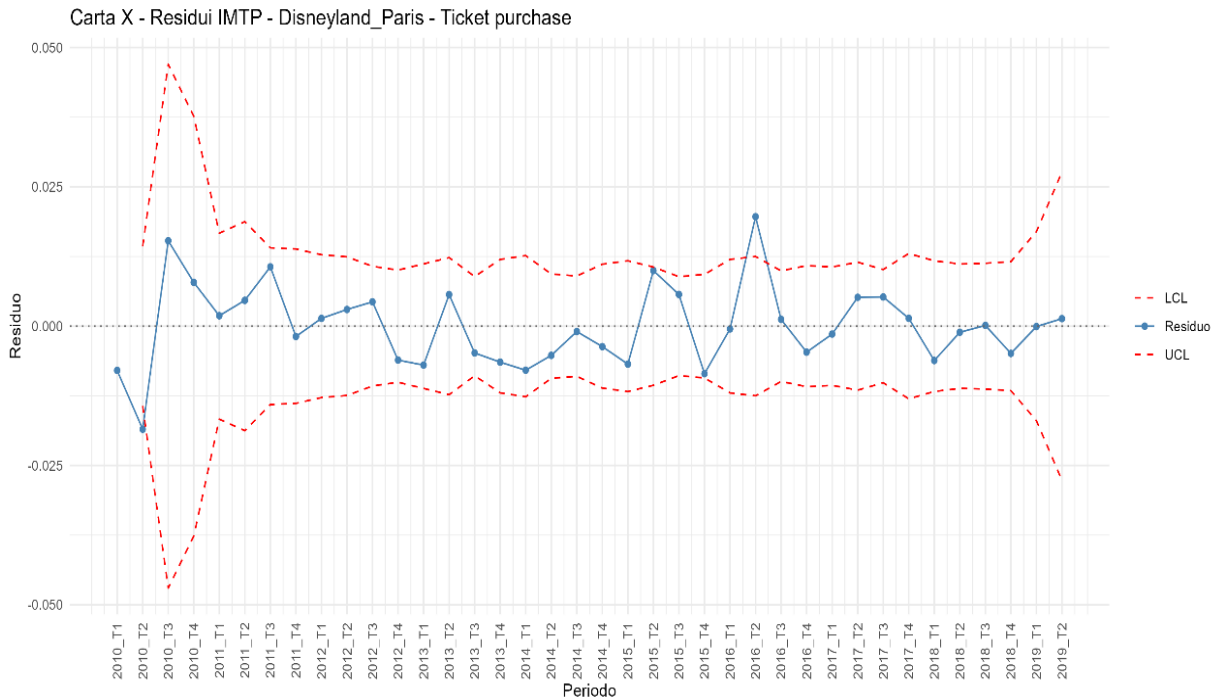


Figura 47 Carta X Topic: “Ticket Purchase” Branch Disneyland Paris

La Carta X, mostrata in Figura 47, dei residui mostra una fase iniziale di instabilità nei primi periodi della serie, pur rimanendo all'interno dei limiti. Successivamente, il processo entra in una fase più stabile, con residui che oscillano attorno al valore medio con punti localizzati all'interno dei limiti di controllo. Tuttavia, nel periodo 2016_T2 si osserva un residuo che supera il limite di controllo superiore (UCL), configurando un chiaro segnale di fuori controllo, periodo di instabilità già osservati nella figura 46 e confermato nella Carta X. Nei periodi immediatamente successivi (2016_T3–2016_T4) i residui rientrano progressivamente entro i limiti di controllo, suggerendo una fase di riassorbimento dell'anomalia e un ritorno a un comportamento più coerente con il trend stimato.

Il fuori controllo positivo del 2016_T2 rappresenta quindi il punto centrale dell'analisi, in quanto segnala un cambiamento temporaneo ma significativo nell'attenzione degli utenti verso il processo di acquisto dei biglietti.

Carta S:

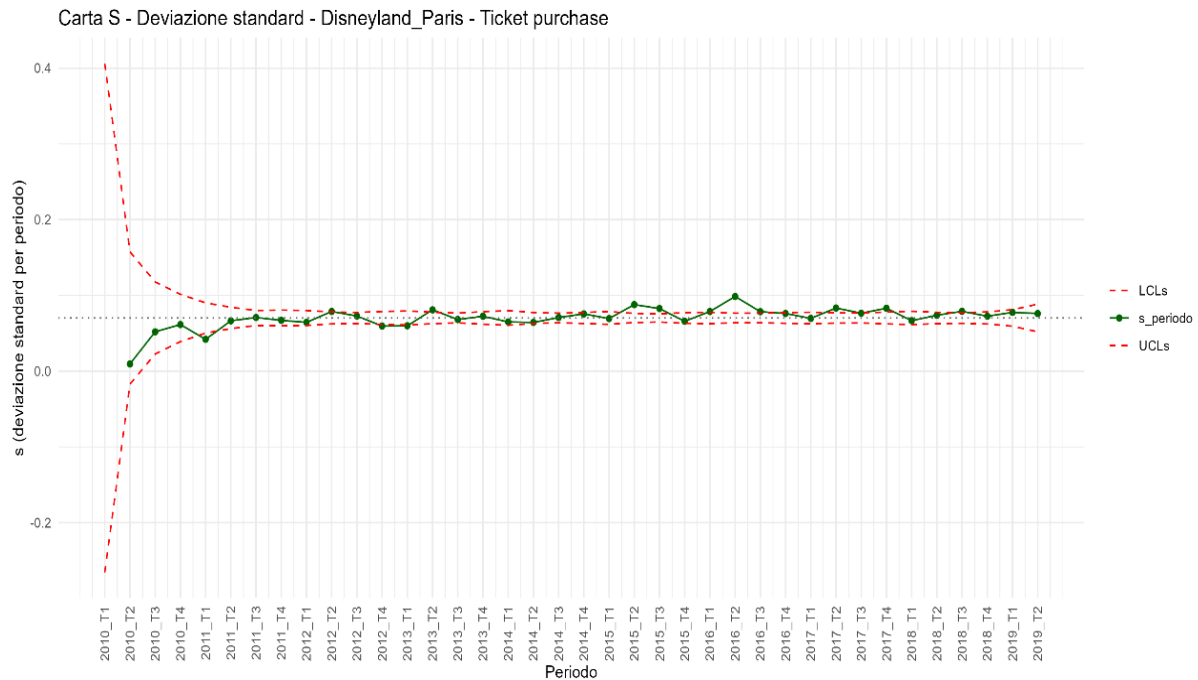


Figura 48 Carta S Topic: "Ticket Purchase" Branch Disneyland Paris

A partire dal 2010, la deviazione standard si assesta stabilmente entro i limiti di controllo, mantenendosi su livelli relativamente costanti per il resto dell'orizzonte temporale. In corrispondenza del periodo di fuori controllo individuato nella Carta X (2016_T2), come mostrato in Figura 48, la Carta S mostra un superamento significativo del limite di controllo superiore (UCL), con un incremento della variabilità. Questo suggerisce che l'aumento anomalo della prevalenza del topic è accompagnato da una dispersione delle opinioni, relativamente al tema dell'acquisto dei biglietti.

- **Analisi MRP (specifica per periodo di interesse)**

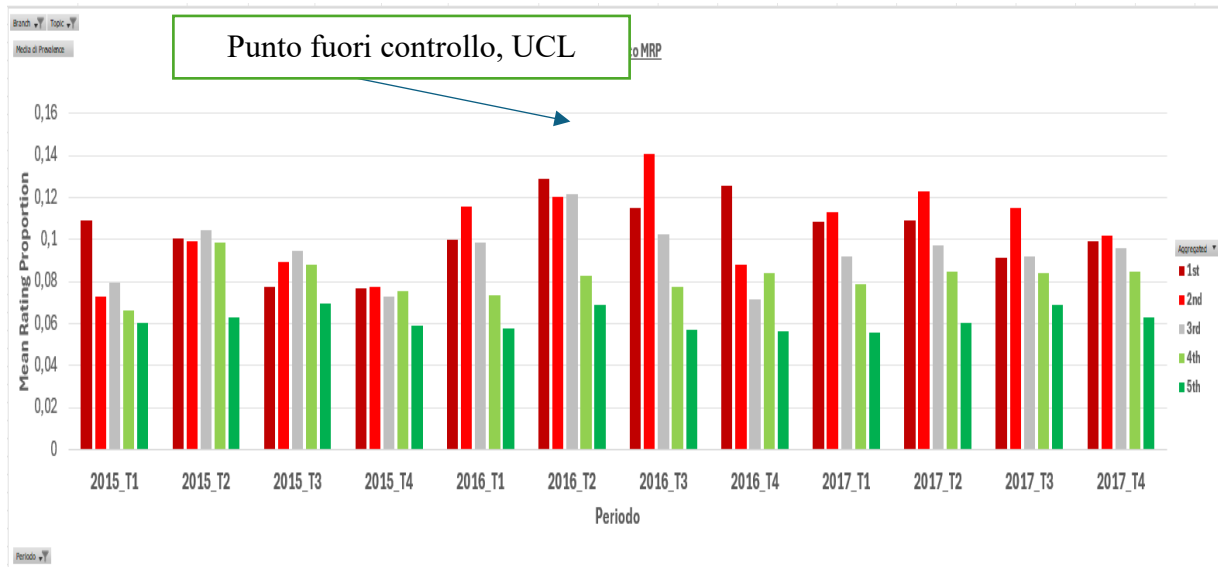


Figura 49 MRP Topic: "Ticket Purchase" Branch Disneyland Paris

L'analisi della Mean Rating Proportion (MRP) è stata condotta su un sottoinsieme temporale che include alcuni periodi precedenti al fuori controllo, il periodo di fuori controllo e i periodi immediatamente successivi, con l'obiettivo di comprendere come l'anomalia individuata nella Carta X sia stata percepita dagli utenti.

Nel periodo precedente al fuori controllo (2016_T1), Figura 49, la distribuzione dell'MRP risulta relativamente equilibrata, con una presenza significativa delle classi intermedie (2nd e 3rd), con valori prossimi allo 0,12 e una quota leggermente più contenuta della classe più estrema (1st), le classi positive (4th e 5th) presenti in una percentuale inferiore con valori prossimi allo 0,08. Questo indica una percezione tendenzialmente negativa prossima ad una valutazione neutra del processo di acquisto dei biglietti.

Nel periodo di fuori controllo positivo (2016_T2), si osserva un chiaro cambiamento nella distribuzione delle classi di giudizio: aumentano in modo significativo le classi (1st e 2nd) di circa un punto percentuale, mentre la quota delle classi più elevate (4th e 5th) rimane più contenuta. Questo suggerisce che l'incremento anomalo della prevalenza del topic Ticket purchase sia stato guidato prevalentemente da esperienze negative o critiche legate al processo di acquisto, che hanno spinto gli utenti a focalizzare maggiormente le recensioni su questo aspetto.

Nei periodi successivi al fuori controllo (2016_T3–2016_T4), la distribuzione dell'MRP mostra un progressivo riequilibrio: le classi più negative diminuiscono e cresce nuovamente il peso delle classi intermedie e medio-alte, notiamo infatti nel 2017_T4 un aumento della quarta classe rating (4th) di quasi due punti percentuali rispetto al 2016_T2 passando da un valore prossimo

allo 0,08 a 0,1. Questo andamento è coerente con il rientro dei residui entro i limiti di controllo e suggerisce che le criticità percepite nel periodo di fuori controllo siano state in parte assorbite o risolte, riportando il processo verso una percezione più stabile.

L'analisi condotta rafforza l'interpretazione del segnale statistico come indicatore di un evento o di una criticità concreta nel processo di acquisto dei biglietti, rendendo i periodi fuori controllo elementi centrali per l'analisi della Digital Voice of Customer.

5. Conclusione Analisi: Analisi dei risultati

L'analisi condotta ha dimostrato come la Digital Voice of Customer possa essere utilizzata non soltanto come fonte di feedback qualitativo, ma anche come strumento quantitativo per il monitoraggio continuo della qualità percepita. Attraverso l'integrazione tra Topic Modeling, indicatori temporali derivati dalle recensioni online e strumenti di Statistical Process Control, è stato possibile trasformare grandi volumi di dati testuali non strutturati in informazioni misurabili, confrontabili nel tempo e utilizzabili a supporto dei processi decisionali. L'approccio adottato ha consentito di osservare l'evoluzione dei principali quality determinants individuati nelle recensioni dei visitatori Disneyland e di identificare tempestivamente periodi caratterizzati da comportamenti significativamente diversi rispetto all'andamento atteso del processo.

Per ciascun topic analizzato, è stato applicato un trend polinomiale al fine di descrivere l'evoluzione temporale del livello di discussione e individuare eventuali dinamiche non lineari o cambiamenti strutturali nella percezione del servizio. L'approccio adottato ha permesso di evidenziare come alcuni topic presentino andamenti relativamente stabili nel tempo, mentre altri mostrino variazioni significative associate a particolari periodi o eventi specifici.

L'analisi dei residui ottenuti dalla regressione polinomiale ha rappresentato un elemento centrale dello studio, in quanto ha consentito di valutare gli scostamenti rispetto al comportamento atteso del processo. I residui sono stati successivamente monitorati mediante carte di controllo X-S, utilizzate per verificare la stabilità statistica del processo e identificare eventuali segnali di fuori controllo. In questo contesto, il residuo assume il ruolo di indicatore sintetico della distanza tra il comportamento osservato e quello atteso del processo. Mentre la regressione polinomiale descrive l'evoluzione strutturale della rilevanza di un topic nel tempo, il residuo evidenzia le variazioni non spiegate dal trend generale, permettendo di concentrare l'attenzione sugli eventi realmente anomali e potenzialmente rilevanti dal punto di vista gestionale. Attraverso questo approccio è stato possibile distinguere le normali oscillazioni fisiologiche da anomalie riconducibili a criticità operative, variazioni organizzative o modifiche nell'esperienza percepita dai visitatori.

Parallelamente, l'analisi dell'MRP specifico ha consentito di approfondire il livello di soddisfazione associato ai singoli topic nel tempo, permettendo di confrontare l'evoluzione della rilevanza dei temi discussi con la relativa percezione qualitativa espressa dagli utenti. Questo confronto ha evidenziato come alcuni topic altamente discussi siano associati a livelli di soddisfazione negativi, configurandosi come elementi critici del servizio, mentre altri presentano valutazioni positive e contribuiscono maggiormente alla qualità percepita complessiva dell'esperienza Disneyland.

Nel complesso, i risultati ottenuti mostrano come l'integrazione tra Topic Modeling, analisi dei trend temporali e carte di controllo rappresenti uno strumento efficace per il monitoraggio dinamico della qualità percepita, consentendo di individuare tempestivamente variazioni significative del processo e supportare attività di miglioramento continuo del servizio.

5.1 Punti di forza dell'analisi

L'approccio proposto consente di trasformare dati testuali non strutturati (recensioni) in indicatori quantitativi interpretabili nel tempo, rendendo possibile un monitoraggio continuo della Digital Voice of Customer. L'utilizzo del Topic Modeling permette di scomporre l'esperienza del visitatore in dimensioni tematiche (topic) e di misurarne la rilevanza tramite l'IMTP, ottenendo serie temporali confrontabili tra parchi e nel tempo.

Un ulteriore punto di forza riguarda l'integrazione tra regressione polinomiale e controllo statistico di processo. La polinomiale fornisce una baseline del trend atteso e consente di separare l'andamento di fondo dalle oscillazioni di breve periodo; l'analisi dei residui tramite carte di controllo permette invece di evidenziare in modo oggettivo i periodi anomali (fuori controllo), che rappresentano il focus principale dell'interpretazione. Questo consente di passare da una lettura puramente descrittiva a una lettura diagnostica del fenomeno, individuando quando il comportamento della dVoC non è spiegato dalla dinamica ordinaria e richiede ulteriori approfondimenti.

Infine, l'uso dell'MRP consente di associare i segnali di anomalia alla dimensione percettiva della qualità. In corrispondenza dei periodi fuori controllo è infatti possibile comprendere se l'attenzione verso un topic aumenta per motivi positivi, negativi o polarizzati, grazie alla distribuzione delle recensioni sulle diverse classi di giudizio. Tale integrazione permette non solo di individuare quando si verifica un'anomalia, ma anche di interpretarne il possibile impatto sull'esperienza percepita dai visitatori.

5.2 Debolezze e limiti

Il primo limite riguarda la natura stessa dei dati: le recensioni online sono soggette a bias di selezione (chi recensisce non è necessariamente rappresentativo di tutti i visitatori) e possono presentare variazioni di volume nel tempo. Inoltre, la composizione dei recensori può cambiare per provenienza geografica, lingua e motivazioni, influenzando la dVoC in modo non controllabile.

Dal punto di vista metodologico, il Topic Modeling introduce elementi di sensibilità legati a scelta del numero di topic, parametri del modello e stabilità semantica dei topic nel tempo. In particolare, un topic può "spostarsi" leggermente di significato o sovrapporsi ad altri, rendendo necessaria una validazione qualitativa continua.

Un ulteriore limite è legato alla regressione polinomiale: se da un lato è utile per catturare andamenti non lineari, dall'altro può essere influenzata da valori estremi (outlier) o da fasi iniziali molto anomale, con il rischio di condizionare la baseline e quindi i residui. Anche la costruzione delle carte di controllo sui residui richiede attenzione: i residui dovrebbero essere il più possibile indipendenti e con varianza relativamente stabile; in presenza di stagionalità, autocorrelazione o cambiamenti strutturali, i limiti di controllo potrebbero non essere pienamente robusti.

Infine, l'interpretazione dei fuori controllo resta in parte "inferenziale": il metodo individua quando avvengono anomalie, ma non fornisce automaticamente perché avvengono, se non attraverso una lettura integrata con altri dati e con analisi qualitative.

5.3 Migliorie e sviluppi futuri dell'approccio

Un primo miglioramento potrebbe riguardare una gestione ancora più precisa della componente temporale. Nei parchi a tema, infatti, l'esperienza del visitatore è fortemente influenzata da stagionalità, festività, vacanze scolastiche, aperture straordinarie ed eventi speciali. Integrare questi elementi in modo più esplicito all'interno dei modelli statistici consentirebbe di distinguere meglio ciò che è "normale" per il periodo da ciò che rappresenta realmente un'anomalia. Un secondo sviluppo riguarda il perfezionamento delle carte di controllo. L'utilizzo di strumenti più evoluti potrebbe aumentare la capacità di rilevare cambiamenti gradualmente nel tempo, evitando al contempo falsi segnali dovuti a oscillazioni fisiologiche del fenomeno osservato. Dal lato testuale, sarebbe utile rafforzare ulteriormente l'analisi dei topic, ad esempio migliorando la gestione delle diverse lingue presenti nelle recensioni oppure utilizzando modelli linguistici più avanzati. Nel caso di brand internazionali come Disneyland, infatti, le recensioni provengono da utenti con background culturali differenti, e ciò può incidere sul modo in cui vengono espressi giudizi e percezioni. Un altro miglioramento concreto potrebbe essere l'utilizzo di una frequenza temporale più dettagliata, ad esempio mensile invece che trimestrale, laddove il numero di recensioni lo consenta. Questo permetterebbe di collegare con maggiore precisione eventuali picchi o cali a eventi realmente accaduti in quello specifico periodo.

Infine, l'approccio risulterebbe ancora più efficace integrando l'analisi quantitativa con una lettura qualitativa delle recensioni nei periodi anomali. Dopo aver individuato un punto fuori controllo, analizzare direttamente i commenti pubblicati in quel momento consentirebbe di comprendere con maggiore chiarezza le cause reali del fenomeno, trasformando il segnale statistico in un'informazione manageriale concreta e utilizzabile.

Bibliografia

- Aggarwal, C.C. (2017), *Outlier Analysis*, 2nd ed., Springer, Cham.
- Agrawal, R., Nurmamatovna, I.S., Xudayarovich, T.S., Karimovich, S.J. and Khaldarova, G. (2024), "Sustainable development with Industry 4.0: a study with design, features and challenges", *Journal of Integrated Science and Technology*, Vol. 12 No. 2, p. 737.
- Antony, J., McDermott, O. and Sony, M. (2022), "Quality 4.0 conceptualisation and theoretical understanding: a global exploratory qualitative study", *The TQM Journal*, Vol. 34 No. 5, pp. 1169-1188, doi: 10.1108/tqm-07-2021-0215.
- Barravecchia, F., Mastrogiamco, L. and Franceschini, F. (2022), "KA-VoC map: classifying product key-attributes from digital voice-of-customer", *Quality Engineering*, Vol. 34 No. 3, pp. 344-358, doi: 10.1080/08982112.2022.2057805.
- Barravecchia, F., Mastrogiamco, L. and Franceschini, F. (2023b), "Product quality tracking based on digital Voice-of-Customers", *Total Quality Management and Business Excellence*, Vol. 34 Nos 11-12, pp. 1386-1409, doi: 10.1080/14783363.2023.2177147.
- Barravecchia, F., Mastrogiamco, L. and Franceschini, F. (2025), "Detecting digital voice-of-customer anomalies to improve product quality tracking", *International Journal of Quality & Reliability Management*, Vol. ahead-of-print No. ahead-of-print, pp. 1-29, DOI 10.1108/IJQRM-07-2024-0229.
- Barravecchia, F., Mastrogiamco, L., Casadesùs Fa, M. and Franceschini, F. (2023a), "MOBI-Qual: a common framework to manage the product-service system quality of shared mobility", *Flexible Services and Manufacturing Journal*, Vol. 36 No. 4, pp. 1359-1398, doi: 10.1007/s10696-023-09520-y.
- Bilal, M. and Almazroi, A.A. (2023), "Effectiveness of fine-tuned BERT model in classification of helpful and unhelpful online customer reviews", *Electronic Commerce Research*, Vol. 23 No. 4, pp. 2737-2757, doi: 10.1007/s10660-022-09560-w.
- Blei, D.M. (2012), "Probabilistic topic models", *Communications of the ACM*, Vol. 55 No. 4, pp. 77-84, doi: 10.1145/2133806.2133826.
- Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.
- Box-Steffensmeier, J.M., Freeman, J.R., Hitt, M.P. and Pevehouse, J.C. (2014), *Time Series Analysis for the Social Sciences*, Cambridge University Press.
- Broday, E.E. (2022), "The evolution of quality: from inspection to quality 4.0", *International Journal of Quality and Service Sciences*, Vol. 14 No. 3, pp. 368-382, doi: 10.1108/ijqss-09-2021-0121.
- Colicev, A., Kumar, A. and O'Connor, P. (2019), "Modeling the relationship between firm and user generated content and the stages of the marketing funnel", *International Journal of Research in Marketing*, Vol. 36 No. 1, pp. 100-116, doi: 10.1016/j.ijresmar.2018.09.005.
- Durairaj, A.K. and Chinnalagu, A. (2021), "Transformer based contextual model for sentiment analysis of customer reviews: a fine-tuned bert", *International Journal of*

Advanced Computer Science and Applications, Vol. 12 No. 11, doi: 10.14569/ijacsa.2021.0121153.

- Escobar, C.A., McGovern, M.E. and Morales-Menendez, R. (2021), “Quality 4.0: a review of big data challenges in manufacturing”, *Journal of Intelligent Manufacturing*, Vol. 32 No. 8, pp. 2319-2334, doi: 10.1007/s10845-021-01765-4.
- Esling, P. and Agon, C. (2012), “Time-series data mining”, *ACM Computing Surveys (CSUR)*, Vol. 45 No. 1, pp. 1-34.
- Hamilton, J.D. (2020), *Time Series Analysis*, Princeton University Press.
- Hawkins, D.L. (1984), “Sequential procedures for detecting deviations in the parameters of the autoregressive model from specified targets”, *Sequential Analysis*, Vol. 3 No. 2, pp. 121-154.
- Kim, R.Y. (2021), “Using online reviews for customer sentiment analysis”, *IEEE Engineering Management Review*, Vol. 49 No. 4, pp. 162-168, doi: 10.1109/emr.2021.3103835.
- Liang, Q. and Wang, K. (2019), “Monitoring of user-generated reviews via a sequential reverse joint sentiment-topic model”, *Quality and Reliability Engineering International*, Vol. 35 No. 4, pp. 1180-1199, doi: 10.1002/qre.2452.
- Mastrogiacomio, L., Barravecchia, F., Franceschini, F. and Marimon, F. (2021), “Mining quality determinants of product-service systems from user-generated contents”, *Quality Engineering*, Vol. 33 No. 3, pp. 425-442, doi: 10.1080/08982112.2021.1877305.
- Montgomery, D.C. (2009), *Introduction to Statistical Quality Control*, 6th ed., Wiley, Hoboken, NJ.
- Mutinda, J., Mwangi, W. and Okeyo, G. (2023), “Sentiment analysis of text reviews using lexiconenhanced bert embedding (LeBERT) model with convolutional neural network”, *Applied Sciences*, Vol. 13 No. 3, p. 1445, doi: 10.3390/app13031445.
- Nainggolan, R. and Purba, E. (2020), “The cluster analysis of online shop product reviews using K-means clustering”, *Data Science: Journal of Computing And Applied Informatics*, Vol. 4 No. 2, pp. 111-121, doi: 10.32734/jocai.v4.i2-2855.
- Nair, V. K. and Raju, M. (2017). Formation of memorable tourism experience and its influence on destination image-evidences from literature. *International Journal of Qualitative Research in Services*, 2(4), 325–334.
- Nguyen, J., Armisen, A., Sánchez-Hernández, G., Casabayò, M. and Agell, N. (2020), “An OWAbased hierarchical clustering approach to understanding users’ lifestyles”, *Knowledge-Based Systems*, Vol. 190, 105308, doi: 10.1016/j.knosys.2019.105308.
- Ozdagoglu, G., Kapucugil_ Ikiz, A. and Çelik, A.F. (2018), “Topic modelling-based decision framework for analysing digital voice of the customer”, *Total Quality Management and Business Excellence*, Vol. 29 Nos 13-14, pp. 1545-1562, doi: 10.1080/14783363.2016.1273106.
- Park, E., Chae, B., Kwon, J. and Kim, W.-H. (2020), “The effects of green restaurant attributes on customer satisfaction using the structural topic model on online customer reviews”, *Sustainability*, Vol. 12 No. 7, p. 2843, doi: 10.3390/su12072843.
- Roberts, M.E., Stewart, B.M. and Airoidi, E.M. (2016), “A model of text for experimentation in the social sciences”, *Journal of the American Statistical Association*, Vol. 111 No. 515, pp. 988-1003, doi: 10.1080/01621459.2016.1141684.

- Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G. (2014), “Structural topic models for open-ended survey responses”, *American Journal of Political Science*, Vol. 58 No. 4, pp. 1064-1082, doi: 10.1111/ajps.12103.
- Sami Sader, Istvan Husti & Miklos Daroczi (2021): A review of quality 4.0: Sader, S., Husti, I. and Daroczi, M. (2021), “A review of quality 4.0: definitions, features, technologies, applications, and challenges”, *Total Quality Management and Business Excellence*, Vol. 33 Nos 9-10, pp. 1-19, doi: 10.1080/14783363.2021.1944082.
- Sony, M., Antony, J. and Douglas, J.A. (2020), “Essential ingredients for the implementation of Quality 4.0”, *TQM Journal*, Vol. 32 No. 4, pp. 779-793, doi: 10.1108/tqm-12-2019-0275.
- Sykora, M., Elayan, S., Hodgkinson, I.R., Jackson, T.W. and West, A. (2022), “The power of emotions: leveraging user generated content for customer experience management”, *Journal of Business Research*, Vol. 144, pp. 997-1006, doi: 10.1016/j.jbusres.2022.02.048.
- Tabianan, K., Velu, S. and Ravi, V. (2022), “K-means clustering approach for intelligent customer segmentation using customer purchase behavior data”, *Sustainability*, Vol. 14 No. 12, p. 7243, doi: 10.3390/su14127243.
- Verma, S. and Yadav, N. (2021), “Past, present, and future of electronic word of mouth (EWOM)”, *Journal of Interactive Marketing*, Vol. 53 No. 1, pp. 111-128, doi: 10.1016/j.intmar.2020.07.001.
- Watson, G.H. (2019), “The ascent of quality 4.0: how the new age of quality came to be and what it might look like in 20 years”, *Quality Progress*, Vol. 52 No. 3, pp. 24-30.
- Zonnenshain, A. and Kenett, R.S. (2020), “Quality 4.0—the challenging future of quality engineering”, *Quality Engineering*, Vol. 32 No. 4, pp. 614-626, doi: 10.1080/08982112.2019.1706744.

Sitografia:

- The Walt Disney Company (2023). Official corporate website. Available at: <https://thewaltdisneycompany.com>
- Disneyland Resort (2023). Official website. Available at: <https://disneyland.disney.go.com>
- Disneyland Paris (2023). Official website. Available at: <https://www.disneylandparis.com>
- Hong Kong Disneyland Resort (2023). Official website. Available at: <https://www.hongkongdisneyland.com>

Appendice A: Codice in R per automazione carte di controllo X-S

```
#####
# APPENDICE A - Codice R per le carte di controllo X-S
#####

# -----
# Librerie
# -----
library(readxl)
library(dplyr)
library(ggplot2)
library(forcats)

# -----
# Parametri iniziali
# -----

file_path <- "Analisi dati Excel.xlsx"
sheet_sent <- "Database Sentiment"

branch_sel <- "Disneyland_California"

# Cartella output grafici
output_dir <- "Grafici_Disneyland_California"

if (!dir.exists(output_dir)) {
  dir.create(output_dir)
}

# -----
# Lettura database
# -----

db <- read_excel(file_path, sheet = sheet_sent)

# Individuazione colonne topic
idx_branch <- which(names(db) == "Branch")
idx_agg <- which(names(db) == "Aggregated")

topic_cols <- names(db)[(idx_branch + 1):(idx_agg - 1)]

# -----
# Funzione principale per analisi topic
# -----

analizza_topic <- function(topic_sel) {

  # Verifica esistenza topic
  if (!topic_sel %in% names(db)) {
    warning(paste("Topic", topic_sel, "non trovato"))
    return(invisible(NULL))
  }

  # Copia locale database
  db_loc <- db
}
```

```

db_loc[[topic_sel]] <- as.numeric(db_loc[[topic_sel]])

# -----
# Calcolo IMTP medio, deviazione standard e numerosità
# -----

df_topic <- db_loc %>%
  filter(Branch == branch_sel) %>%
  select(Periodo, Branch, all_of(topic_sel)) %>%
  filter(!is.na(Periodo)) %>%
  group_by(Periodo) %>%
  summarise(
    IMTP_mean = mean(.data[[topic_sel]], na.rm = TRUE),
    s_periodo = sd(.data[[topic_sel]], na.rm = TRUE),
    n_periodo = sum(!is.na(.data[[topic_sel]])),
    .groups = "drop"
  ) %>%
  arrange(Periodo) %>%
  mutate(
    indice_periodo = row_number(),
    Periodo = factor(Periodo,
                     levels = unique(Periodo))
  )

# Controllo numero minimo di periodi
if (nrow(df_topic) < 4) {
  warning(paste("Topic", topic_sel,
                ": meno di 4 periodi"))
  return(invisible(NULL))
}

# -----
# Regressione polinomiale di terzo grado
# -----

mod_poly <- lm(
  IMTP_mean ~ poly(indice_periodo, 3, raw = TRUE),
  data = df_topic
)

df_topic <- df_topic %>%
  mutate(
    IMTP_pred = predict(mod_poly, newdata = df_topic),
    residuo = IMTP_mean - IMTP_pred
  )

# -----
# Carta X dei residui
# -----

media_residuo <- mean(df_topic$residuo, na.rm = TRUE)

df_topic <- df_topic %>%
  mutate(
    SE = s_periodo / sqrt(n_periodo),
    UCLx = media_residuo + 3 * SE,
    LCLx = media_residuo - 3 * SE
  )

```

```

# Grafico carta X
plot_X <- ggplot(df_topic,
                 aes(x = indice_periodo)) +

  geom_line(aes(y = residuo,
                colour = "Residuo")) +

  geom_point(aes(y = residuo,
                 colour = "Residuo")) +

  geom_line(aes(y = UCLx,
                colour = "UCL"),
            linetype = "dashed") +

  geom_line(aes(y = LCLx,
                colour = "LCL"),
            linetype = "dashed") +

  geom_hline(yintercept = media_residuo,
             colour = "grey40",
             linetype = "dotted") +

  scale_colour_manual(
    values = c(
      "Residuo" = "steelblue",
      "UCL"     = "red",
      "LCL"     = "red"
    )
  ) +

  scale_x_continuous(
    breaks = df_topic$indice_periodo,
    labels = df_topic$Periodo
  ) +

  labs(
    title = paste(
      "Carta X - Residui IMTP -",
      branch_sel,
      "-",
      topic_sel
    ),
    x = "Periodo",
    y = "Residuo"
  ) +

  theme_minimal() +

  theme(
    axis.text.x = element_text(
      angle = 90,
      vjust = 0.5,
      hjust = 1
    ),
    legend.title = element_blank()
  )

print(plot_X)

```

```

# -----
# Salvataggio grafico carta X
# -----

safe_topic_name <- gsub(
  "[^A-Za-z0-9]+",
  "_",
  topic_sel
)

ggsave(
  filename = file.path(
    output_dir,
    paste0("CartaX_", safe_topic_name, ".png")
  ),
  plot = plot_X,
  width = 12,
  height = 6,
  dpi = 300
)

# -----
# Carta S
# -----

S_medio <- mean(df_topic$s_periodo, na.rm = TRUE)

df_topic <- df_topic %>%
  mutate(
    C4 = 1 - 1 / (4 * n_periodo)
      - 7 / (32 * n_periodo^2),

    B4 = 1 + 3 * sqrt(1 - C4^2) / C4,

    B3 = 1 - 3 * sqrt(1 - C4^2) / C4,

    UCLs = S_medio * B4,
    LCLs = S_medio * B3
  )

# Grafico carta S
plot_S <- ggplot(df_topic,
  aes(x = indice_periodo)) +

  geom_line(aes(y = s_periodo,
    colour = "s_periodo")) +

  geom_point(aes(y = s_periodo,
    colour = "s_periodo")) +

  geom_line(aes(y = UCLs,
    colour = "UCLs"),
    linetype = "dashed") +

  geom_line(aes(y = LCLs,
    colour = "LCLs"),
    linetype = "dashed") +

```

```

geom_hline(yintercept = S_medio,
           colour = "grey40",
           linetype = "dotted") +

scale_colour_manual(
  values = c(
    "s_periodo" = "darkgreen",
    "UCLs"      = "red",
    "LCLs"      = "red"
  )
) +

scale_x_continuous(
  breaks = df_topic$indice_periodo,
  labels = df_topic$Periodo
) +

labs(
  title = paste(
    "Carta S - Deviazione standard -",
    branch_sel,
    "-",
    topic_sel
  ),
  x = "Periodo",
  y = "s (deviazione standard)"
) +

theme_minimal() +

theme(
  axis.text.x = element_text(
    angle = 90,
    vjust = 0.5,
    hjust = 1
  ),
  legend.title = element_blank()
)

print(plot_S)

# -----
# Salvataggio grafico carta S
# -----

ggsave(
  filename = file.path(
    output_dir,
    paste0("CartaS_", safe_topic_name, ".png")
  ),
  plot      = plot_S,
  width     = 12,
  height    = 6,
  dpi       = 300
)

invisible(NULL)
}

```

```

# -----
# Applicazione funzione a tutti i topic
# -----

for (topic_sel in topic_cols) {
  analizza_topic(topic_sel)
}

cat(
  "\nAnalisi completata per tutti i topic di",
  branch_sel,
  "\n"
)

# -----
# Validazione automatica dei topic
# -----

validazione_topic <- function(topic_sel) {

  # Verifica esistenza topic
  if (!topic_sel %in% names(db)) {

    return(data.frame(
      Topic = topic_sel,
      Status = "COLONNA NON TROVATA"
    ))
  }

  # Copia locale database
  db_loc <- db

  db_loc[[topic_sel]] <- as.numeric(
    db_loc[[topic_sel]]
  )

  # -----
  # Calcolo statistiche
  # -----

  df <- db_loc %>%
    filter(Branch == branch_sel) %>%
    select(Periodo, all_of(topic_sel)) %>%
    filter(!is.na(Periodo)) %>%
    group_by(Periodo) %>%
    summarise(
      IMTP_mean = mean(
        .data[[topic_sel]],
        na.rm = TRUE
      ),

      s_periodo = sd(
        .data[[topic_sel]],
        na.rm = TRUE
      ),

      n_periodo = sum(
        !is.na(.data[[topic_sel]])
      ),

```

```

        .groups = "drop"
    )

# -----
# Statistiche di validazione
# -----

n_periodi_tot <- nrow(df)

n_validi <- sum(!is.na(df$IMTP_mean))

n_NA <- sum(is.na(df$IMTP_mean))

s_medio <- mean(df$s_periodo,
               na.rm = TRUE)

n_min <- min(df$n_periodo,
            na.rm = TRUE)

n_max <- max(df$n_periodo,
            na.rm = TRUE)

# -----
# Condizioni di validità
# -----

ok <- TRUE

if (n_validi < 4) ok <- FALSE

if (any(df$n_periodo == 0)) ok <- FALSE

if (is.nan(s_medio) | is.na(s_medio)) ok <- FALSE

# -----
# Output finale
# -----

return(data.frame(
  Topic          = topic_sel,
  Periodi_totali = n_periodi_tot,
  Periodi_validi = n_validi,
  Periodi_NA     = n_NA,
  S_medio        = round(s_medio, 5),
  N_min          = n_min,
  N_max          = n_max,
  OK             = ok
))
}

# -----
# Applicazione validazione a tutti i topic
# -----

risultati_validazione <- do.call(
  rbind,
  lapply(topic_cols, validazione_topic)
)

```

```
print(risultati_validazione)
```