POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Gestionale

Tesi di Laurea Magistrale

Progettazione e sviluppo di una metodologia di Data Analytics e modelli di Business Intelligence a supporto della gestione di una scuola sci

Relatore
prof.ssa Tania Cerquitelli

Candidato
Fabio Brignolo

Anno Accademico 2019-2020

Indice

Introduzione 1

Business Intelligence 2

1.1 Origine e significato 2
1.2 Il processo decisionale 4
1.3 Digital Transformation 7
1.4 Scienza dei dati 10

Il caso di studio 13

2.1 La società 13
2.2 Contesto geografico e competitivo 15
2.3 Panoramica del settore 18
2.4 Stato dell'arte 22
2.5 Il perché dell'analisi 25
2.5.1 Bisogni 26
2.5.2 Sfide 27
2.5.3 Opportunità 32

Il progetto 34

3.1 Data Management 34
3.2 Extract, Transform, Load 39
3.3 Database 41
3.4 Knowledge Discovery Process 44

Data Analysis 49

4.1 Premesse 49
4.2 Anomaly detection 51
4.3 Analisi temporali 53
4.4 Analisi sui soci 59
4.5 Analisi sulle famiglie 66
4.6 Analisi sui settori 70
4.7 Analisi proposte 77
4.7.1 Regole di associazione 78

4.7.2 Analisi predittive 79
4.7.3 Analisi incrociate 81
Conclusioni 82
Bibliografia 85

Tavola delle Figure

[Figura 1] Processo decisionale generico in ottica Business Intelligence.....5

[Figura 2] Digitization, Digitalization e Digital Transformation Fonte: coresystems.net 7

[Figura 3] Crescita dei dati memorizzati a livello globale.....11

[Figura 4] Struttura della scuola sci.....13

[Figura 5] Distribuzione delle partecipazioni, in termini relativi, nel corso della settimana. Fonte: dati scuola sci e snowboard.....16

[Figura 6] Impatto delle condizioni del manto nevoso sull'affluenza media di periodo 17

[Figura 7] Rappresentazione del fenomeno "Stuck in the middle" 19

[Figura 8] Trend discipline a livello nazionale.....20

[Figura 9] Esempio semplificato di un tabellone 23

[Figura 10] Software gestionali della scuola sci.....23

[Figura 11] Bisogni, sfide e opportunità per una scuola sci 25

[Figura 12] VRIO Analysis Fonte: Business2You.....31

[Figura 13] Problematiche relative ai dati.....35

[Figura 14] Aspetti chiave del Data Management Fonte: MKT Softwares.....36

[Figura 15] Struttura semplificata del database.....42

[Figura 16] Knowledge Discovery Process.....45

[Figura 17] Risultato del DBSCAN sugli iscritti ai corsi stagionali. Il cluster_0 contiene gli elementi etichettati come outlier. 52

[Figura 18] Trend iscritti per stagione 53

[Figura 19] Distribuzione delle lezioni nell'arco della stagione invernale 54

[Figura 20] Distribuzione delle lezioni durante la giornata 55

[Figura 21] Matrice di correlazione sui dati meteo 58

[Figura 22] Distribuzione dei soci per età e stagione 61

[Figura 23] Distribuzione per età e fedeltà nelle stagioni 2013-2014 e 2014-2015 62

[Figura 24] Fedeltà dei soci per range di età 64

[Figura 25] Esempio di distribuzione dei soci nel comune di Torino 65

[Figura 26] Trend numerosità media famiglie 67

[Figura 27] Distribuzione numerosità famiglie per stagione 68

[Figura 28] Percentuale rinnovi per numerosità famiglia 69

[Figura 29] Distribuzione dei settori per età 70

[Figura 30] Esempio di grafo orientato rappresentativo dei rinnovi dei soci 72

[Figura 31] Matrice dei rinnovi 74

[Figura 32] Classifica dei settori per gli indici PA, PPA e IF 76

[Figura 33] Esempio di input per l'algoritmo FP-growth 78

[Figura 34] Esempio di regola di associazione. Tra parentesi i valori dei parametri Support e Confidence 79

Introduzione

Con l'affermarsi della digitalizzazione e delle più svariate tecnologie, il dato rappresenta sempre più un valore aggiunto per tutte le attività commerciali che sfruttano un sistema informativo per l'organizzazione delle risorse interne o per la gestione dei servizi. Le imprese immagazzinano quotidianamente una consistente mole di informazioni che vengono utilizzate principalmente per il monitoraggio delle attività in tempo reale e il corretto svolgimento delle funzioni operative, ottimizzando il processo lavorativo in termini di tempo e minimizzando gli errori di comunicazione. Tuttavia, è solo negli ultimi anni che le imprese hanno iniziato a comprendere come sfruttare i dati in proprio possesso come supporto ai processi decisionali di medio e lungo periodo, con l'obiettivo di valutare e incrementare le performance aziendali attraverso un approccio data-driven. Sempre più spesso infatti, le scelte del management sono guidate dagli insights forniti da modelli di Business Intelligence che trasformano i dati grezzi in informazioni significative in grado di evidenziare gli aspetti chiave del business.

Oggetto del presente studio è l'applicazione delle metodologie di Business Intelligence all'attività di una scuola sci, con l'obiettivo di comprendere come un approccio data-driven possa fornire un valido supporto indipendentemente dalla grandezza dell'azienda e dal settore in cui opera.

La tesi affronta il tema della Business Intelligence soffermandosi sull'origine del termine e sul valore che essa rappresenta nel guidare il processo decisionale di

un'impresa; vengono inoltre introdotti i concetti di Digital Transformation e Data Science per comprendere il ruolo che ricopre la tecnologia in termini di gestione dei dati e approccio al business.

Si offre una panoramica della società in esame e del contesto geografico e competitivo in cui opera, con un focus sul settore di appartenenza per introdurre quelli che saranno poi confermati come i temi chiave dello studio.

Alla presentazione dello stato dell'arte si aggiungono considerazioni riguardanti le motivazioni del lavoro svolto, evidenziando i bisogni della scuola, a quali sfide deve saper rispondere e le possibili opportunità da cogliere per consolidare il proprio vantaggio competitivo sul mercato.

Il lavoro ha avuto come oggetto la progettazione di un database omogeneo, risultato dell'accorpamento dello storico dati comprensivo delle informazioni immagazzinate in 10 anni di attività e proveniente da tre diversi software gestionali. Sono state quindi eseguite operazioni di data cleaning, data mining e analisi, i cui principali risultati sono stati qui presentati per fornire un esempio di come alcuni aspetti chiave del business siano osservabili dal semplice studio dei dati, anche da un agente esterno al settore.

L'obiettivo della tesi è quello di fornire alla scuola uno strumento utile alla definizione del target d'impresa e alla valutazione delle performance aziendali, in grado di esplicitare relazioni e fenomeni intrinseci ai dati e che possa rappresentare una base per il processo di evoluzione da semplice scuola sci ad azienda.

Capitolo 1

Business Intelligence

1.1 Origine e significato

"Throughout Holland, Flanders, France, and Germany, he maintained a complete and perfect train of business intelligence. The news of the many battles fought was thus received first by him, and the fall of Namur added to his profits, owing to his early receipt of the news."

Così scriveva Richard Millar Devens nel 1865 [1] per descrivere il successo di Sir Henry Furnese, banchiere e politico britannico tra XVII e XVIII secolo. Furnese guadagnava infatti raccogliendo informazioni politiche e di mercato, agendo di conseguenza e in anticipo rispetto ai competitor. Si tratta del documento più antico di cui si abbia conoscenza in cui compare il termine *business intelligence* ed è sorprendente come una terminologia così attuale abbia in realtà radici tanto antiche. Per trovare un significato più vicino a quello dei giorni nostri bisogna però

aspettare il 1958 quando Hans Peter Luhn, ricercatore presso IBM, utilizzò il termine facendo riferimento a quanto riportato nel dizionario alla voce *intelligence*.

"The ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal." [2]

Luhn, nel suo articolo, espone un sistema in grado di automatizzare la trasmissione delle informazioni a diverse aree di organizzazioni scientifiche, industriali e governative, ovvero quegli organismi che maggiormente erano rimasti coinvolti nella terza rivoluzione industriale e necessitavano di affiancare all'innovazione tecnologica un sistema per organizzare i dati e renderli fruibili. In quegli anni IBM presentava al mondo i primi *hard disk*, il cui impiego avrebbe consentito la memorizzazione di grandi quantità di dati e, quindi, l'effettivo sviluppo del sistema proposto da Luhn e la creazione del primo Decision Support System (DSS).
Henk Gerard Sol, professore olandese di Business Engineering presso l'università di Groningen, scrive come la definizione di DSS sia mutata nel tempo [3].
Nei primi anni '70 ci si riferiva al DSS come *"a computer-based system to aid decision making"* mentre verso la fine del decennio l'attenzione è stata rivolta per la prima volta ai dati:

"interactive computer-based systems which help decision-makers utilize data bases and models to solve ill-structured problems".

Negli anni '80 i sistemi DSS avrebbero inoltre dovuto supportare le attività di decision-making ai fini di migliorare l'efficienza delle attività di management:

"using suitable and available technology to improve effectiveness of managerial and professional activities".

L'origine della terminologia *business intelligence* risiede tutta qui. Per comprenderla meglio facciamo riferimento alla definizione che ne dà Howard Dresner nel 1989, allora analista presso Gartner Group, riportata da Daniel J. Power [4], professore di Information Systems and Management presso la University of Northern Iowa:

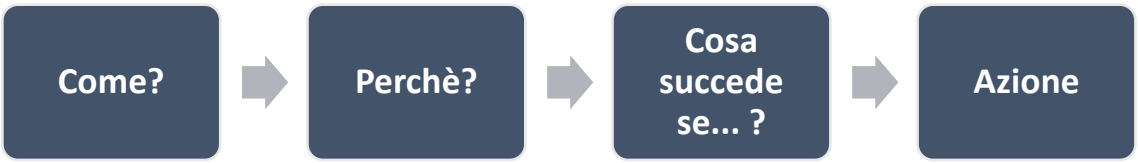
"It describes a set of concepts and methods to improve business decision making by using fact-based support systems."

La Business Intelligence è quindi una disciplina di supporto alle decisioni strategiche aziendali il cui obiettivo risiede nella trasformazione dei dati grezzi in informazioni utili: si tratta un approccio *data-driven* in quanto le dinamiche del decision-making sono fortemente influenzate dagli studi sui dati, storici o in tempo reale, dell'azienda.

L'attenzione è quindi rivolta ai dati come componente chiave della conoscenza su cui basare le decisioni di business, per capire quali azioni intraprendere e quando intraprenderle.

1.2 Il processo decisionale

Una volta compresi il significato e le origini del termine è possibile approfondire gli aspetti più concreti della disciplina. Il processo che sta alla base del decision-making è riassumibile schematicamente in 4 step [Figura 1].



[Figura 1] Processo decisionale generico in ottica Business Intelligence

- **Come?** Alla base di ogni attività vi sono gli obiettivi e il management deve essere in grado di stabilire dei parametri (*Key Performance Indicator*, KPI) per la valutazione delle performance in base agli obiettivi prefissati. Il processo decisionale parte dalla valutazione periodica, quantitativa e qualitativa dei fenomeni di interesse del business e dei KPI fondamentali per la comprensione della situazione.

- **Perché?** Individuate le criticità principali si passa all'analisi dettagliata dei fenomeni di interesse. L'obiettivo è indagare sul perché di un determinato evento scomponendo il dato complessivo e analizzandolo a partire da diverse dimensioni di analisi. Questi primi due step contribuiscono a comprendere la situazione e a focalizzare l'attenzione sugli aspetti critici del business in modo da evidenziare le cause di un determinato fenomeno per poterlo evitare in futuro (se negativo) o replicare (se positivo).

- **Cosa succede se...?** Nel terzo step si cerca di capire quali siano le misure correttive da effettuare per poter dare una risposta al problema evidenziato. Attraverso *what-if analysis*, simulazioni e analisi predittive è possibile visualizzare gli scenari futuri probabili conseguenti ad un'azione intrapresa, permettendo al management di intraprendere una strada piuttosto che un'altra.

- **Azione** Il quarto step è la presa di posizione una volta esaminati i risultati delle analisi. L'intero processo è studiato per ottimizzare la scelta dell'azione da intraprendere nel più breve tempo possibile.

Tutte le attività di Business Intelligence hanno come obiettivo ultimo il miglioramento delle performance aziendali, che sia in termini di tempo, efficienza degli interventi, ottimizzazione dei servizi offerti o riduzione dei costi.

Imprescindibile per il processo decisionale è il sistema informativo che deve permettere un rapido accesso ai dati e l'elaborazione degli stessi per consentire al management di reagire tempestivamente agli eventi esterni e intraprendere azioni efficaci. Il concetto che viene però spesso travisato è il ruolo della tecnologia nel sistema di Business Intelligence: al centro dell'attenzione deve essere posto il contesto aziendale ed il sistema informativo deve essere studiato e strutturato per seguire le logiche del business, non la tecnologia.

La struttura dei dati alla base del processo, le analisi, la reportistica, le componenti del sistema informativo devono tutte essere poste nel contesto del business di riferimento affinché i risultati possano avere una certa rilevanza ed essere utilizzati con efficacia dai decision-maker.

Tornando alla [Figura 1] soffermiamoci sugli step 1, 2 e 3. In primo luogo, il dato viene esaminato così com'è in base ai KPI di interesse, poi scomposto e studiato nel dettaglio per analizzare le cause degli eventi, infine utilizzato per la creazione di modelli predittivi. Il centro dell'attenzione è il dato: l'intero concetto di Business Intelligence non può prescindere dallo studio dei dati e dalla definizione di un processo metodologico di elaborazione ed analisi delle informazioni.

1.3 Digital Transformation

Per parlare di business intelligence è opportuno quindi capire come processi decisionali e tecnologia si integrino e quale sia il ruolo di quest'ultima nelle logiche del business di un'azienda. L'affermarsi della tecnologia digitale può essere riassunto in tre grandi periodi [Figura 2]:

[Figura 2] Digitization, Digitalization e Digital Transformation

Fonte: coresystems.net

- **Digitization** Con il termine *digitization* si intende il processo di conversione delle informazioni in formato digitale [5]. La tecnologia si sviluppa a partire dagli anni '70, quando la digitalizzazione dei dati venne adottata nel settore delle telecomunicazioni per migliorare le performance del network in termini di tempi, costi, sostenibilità e qualità del suono. L'era della *digitization* ha quindi come centro di interesse il dato; è di cruciale importanza per i meccanismi di trasmissione, storage e processing dei dati e ha posto le basi per quella che viene definita come Terza Rivoluzione Industriale.

- **Digitalization** Nonostante per la traduzione italiana non vi sia alcuna differenza con la digitization, la sfumatura anglosassone è importante e sottolinea come l'era della *digitalization* sia rivolta alle imprese invece che ai dati; più in generale, il focus della digitalizzazione è concentrato ora sullo scambio di informazioni e sull'estrapolazione di conoscenza dai dati. La digitalization interessa le aziende e i loro processi produttivi ed è tutt'ora in atto: Internet of Things, Blockchain, Big Data, Machine Learning e Industry 4.0 sono esempi di nuove tecnologie legate alla digitalization e rappresentano una sfida alla quale le imprese moderne dovranno saper rispondere per non uscire dal mercato. L'era della digitalization ha aperto le porte per nuovi modelli di business e per l'evoluzione dei processi esistenti, migliorandone l'efficienza e l'efficacia.
- **Digital Transformation** L'integrazione della digitization, intesa come evoluzione tecnologica, e digitalization (evoluzione dei processi) ha posto le basi per una nuova era, quella della *digital transformation*.

  "*Digital transformation is the integration of digital technology into all areas of a business, fundamentally changing how you operate and deliver value to customers. It's also a cultural change that requires organizations to continually challenge the status quo, experiment, and get comfortable with failure.*" [6]

  Con digital transformation si intende quindi un cambiamento dei processi produttivi, operativi e decisionali basato sulla digitalizzazione delle informazioni che influenza non solo il business dell'impresa, ma anche la cultura ed il modus operandi della stessa.

Nell'era della DT le aziende sono portate a reinventare il proprio business; questo vuol dire essere disposti a cambiare il modo in cui le attività vengono svolte e percepite dagli addetti ai lavori, sostituire macchinari, installare nuovi sistemi informativi, interfacciarsi in maniera diversa con l'utente finale e renderlo parte del business. Il processo richiede quindi investimenti importanti tanto nelle tecnologie quanto nel capitale umano: le competenze del personale sono un fattore cruciale nell'esito del cambiamento in quanto le tecnologie adottate necessitano spesso di persone dotate di determinate skills, e la prospettiva di una gestione delle attività affidata a tecnologie non familiari è un ulteriore freno per la digitalizzazione.

Nonostante l'argomento DT sia di grande attualità, le imprese italiane pronte a fare il grande passo sono in netta minoranza. L'Istat stima che le imprese italiane con un elevato livello di digitalizzazione siano soltanto il 4,7% a fronte di un 80% caratterizzato da un profilo tecnologico di basso livello [7].

In un articolo sul sito di Harvard Business Review [8] sono elencati i principali ostacoli al cambiamento, risultato di sondaggi e interviste a imprese di tutto il mondo:

1. Rallentamenti del processo decisionale dovuti a politiche interne, competitività tra dipendenti e ricerca del raggiungimento del consenso;
2. Incapacità di calcolare il beneficio del cambiamento dovuto alla digitalizzazione in termini quantitativi e scetticismo del management;
3. Eccessiva attenzione alla tecnologia e difficoltà nel reinventare il business dell'impresa e il modo in cui le persone lavorano;

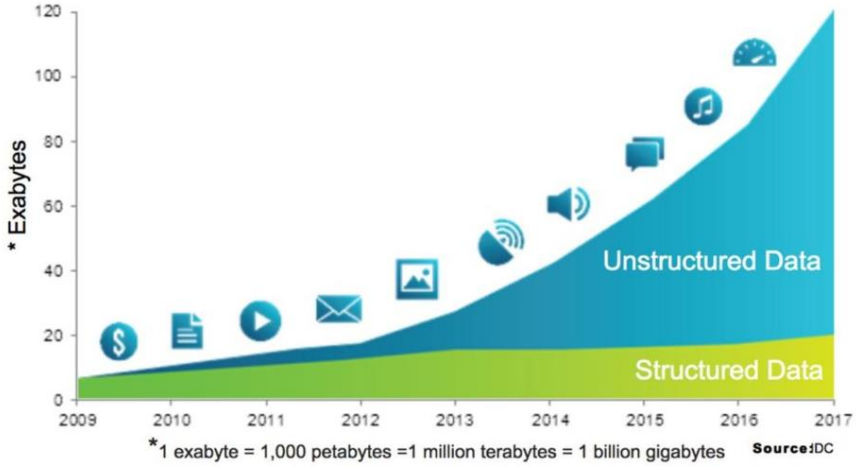
4. Mancanza di comprensione delle problematiche operative a livello decisionale e difficoltà nel passaggio dalla teoria alla pratica;
5. Paura di perdere il controllo delle operazioni e mancanza di fiducia negli addetti ai lavori.

Le imprese sono restie ad affrontare il cambiamento, e lo scetticismo è in parte giustificato: si stima infatti [9] che nel solo 2019 circa il 70% delle operazioni aventi come obiettivo la digital transformation sia risultato in un fallimento. In termini economici, su 1300 miliardi di investimento, 900 miliardi sono stati persi.

1.4 Scienza dei dati

Nell'era di digitalization e digital transformation, così come nella disciplina della business intelligence, lo studio dei dati ricopre un ruolo fondamentale.

A partire dagli ultimi anni del XX secolo, nel mondo si sono generati e memorizzati dati in quantità e velocità esponenziali [Figura 3]. Con la diffusione di internet e in particolar modo dei social media, sono stati generati e scambiati dati sotto forma di e-mail, immagini, video, messaggi e documenti; fonti eterogenee di informazioni definiti come *unstructured data*, ovvero i dati non strutturati. Questi comprendono circa l'80% dei dati attualmente memorizzati e risulta complesso analizzarli con i tradizionali metodi di analisi, utili invece ad esaminare quelli che prendono il nome di *structured data*, ovvero i dati che seguono uno schema standardizzato (pagine html, tabelle di database).



[Figura 3] Crescita dei dati memorizzati a livello globale.
Fonte: IDC

Con l'affermarsi delle tecnologie relative alla digitalization, quali per esempio Internet of Things, e le opportunità di business offerte dalla DT, la mole di dati generati è destinata ad aumentare vertiginosamente nei prossimi anni.

Al fine di gestire ed elaborare un tale volume di informazioni sono stati definiti processi metodologici e modelli di analisi complessi che, nell'insieme, prendono il nome di *data science*. Con data science ci si riferisce a un campo multidisciplinare che utilizza metodi scientifici, processi, algoritmi e sistemi per estrapolare conoscenza dai dati, siano essi strutturati o non strutturati. [10]

James Nicholas Gray, meglio noto come Jim Gray, computer scientist statunitense vincitore del Turing Award nel 1998, vede nella data science il quarto paradigma della scienza (empirica, teorica, computazionale e ora *data-driven*) e afferma:

> *"Everything about science is changing because of the impact of information technology" and the data deluge [11]*

Si parla quindi di una vera e propria scienza legata al dato, che guarda oltre le semplici analisi (*query driven*, approccio top-down) e ha come obiettivo il riconoscimento di pattern e schemi interni ai dati (approccio bottom-up) e la definizione di nuovi modelli replicabili.

I temi legati alla data science esulano da quello che è lo studio qui proposto, ma il concetto di fondo da cui si vuole prendere spunto è l'approccio scientifico tipico che sta alla base dell'attività svolta.

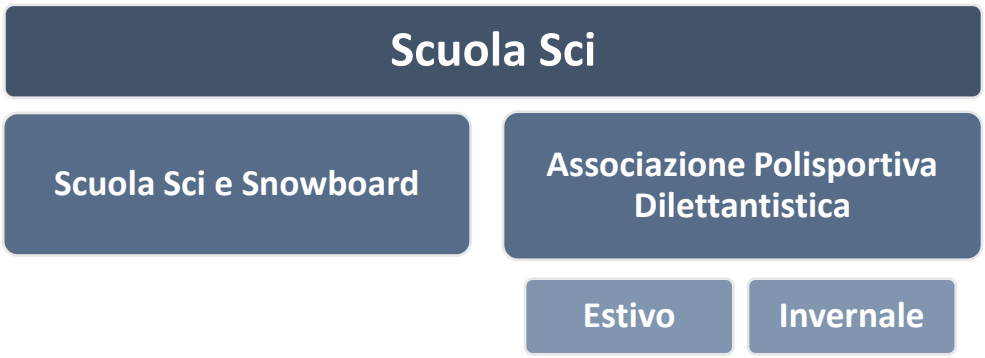
"The key word in 'Data Science' is not Data, it is Science" [12]

Capitolo 2

Il caso di studio

2.1 La società

La società di riferimento per lo studio è una scuola di sci piemontese, in provincia di Torino. La scuola nasce nel 2004 e con il passare degli anni si impone come una delle scuole di maggior rilievo del panorama piemontese. La società si occupa di gestire corsi e lezioni, sia in periodo invernale che estivo, e le attività sono rivolte a tutte le fasce di età a partire dai 4 anni.



[Figura 4] Struttura della scuola sci

La scuola è organizzata come mostrato in [Figura 4]:

- **Scuola Sci e Snowboard** Si occupa di gestire e organizzare le lezioni individuali e i corsi di breve durata (anche dell'ordine della singola giornata) del periodo invernale. Le attività proposte sono rivolte prevalentemente a chi vuole imparare a sciare e a chi scia occasionalmente. Si tratta quindi di lezioni di medio-basso livello la cui natura primaria è quella didattica, dedicate al cliente casual che non intende impegnarsi a lungo termine con la scuola.
- **L'Associazione Polisportiva Dilettantistica** Si occupa di gestire e organizzare i corsi stagionali della società. Le iscrizioni ai corsi invernali coprono il periodo tra ottobre e aprile mentre i corsi estivi si svolgono tra maggio e settembre. I corsi organizzati dall'A.P.D. sono indirizzati ad un pubblico già capace di sciare e hanno l'obiettivo di fidelizzare il cliente nelle diverse aree sportive. Le attività proposte per i mesi estivi sono pensate per lo svolgimento in assenza di neve (mountain bike, arrampicata...), rivolte a una clientela giovane e finalizzate ad ampliare l'offerta in termini di servizi e customer experience.

Le attività eseguite dalla scuola riguardano quindi l'organizzazione dei corsi, delle lezioni e degli eventi (gare, manifestazioni) e comprendono la gestione delle iscrizioni dei soci, la schedulazione delle lezioni e l'assegnazione dei maestri ai gruppi di allievi. La scuola offre inoltre la possibilità di affittare agli iscritti parte dell'attrezzatura sportiva necessaria per le lezioni e gestisce quindi un semplice magazzino di poche unità per prodotto.

2.2 Contesto geografico e competitivo

La località in cui è insediata la scuola sci e snowboard del nostro caso di studio rappresenta una delle mete più gettonate del Piemonte per quanto riguarda il turismo. La dotazione turistica comprende ristoranti, alberghi, discoteche, cinema, campi sportivi (calcio, tennis, pallavolo...) e palazzetti per lo sport, e nei periodi di alta stagione si contano oltre 30.000 presenze turistiche giornaliere che utilizzano, oltre alla ricettività alberghiera ed extra-alberghiera, i numerosi appartamenti in affitto e abitazioni secondarie.

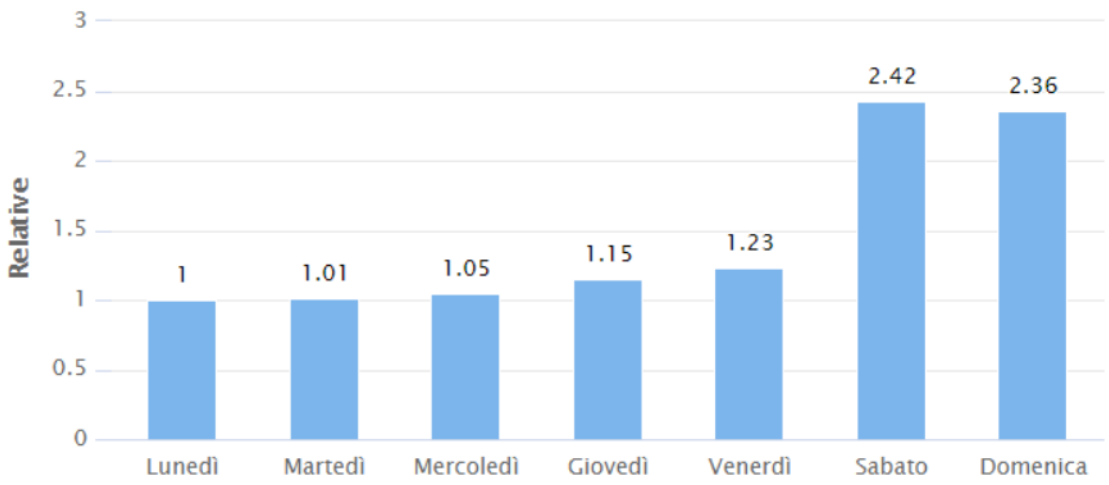
La scuola sci e snowboard è collocata nella zona in cui sorsero i primi impianti sciistici d'Italia agli inizi del '900, e si pone in un contesto molto competitivo. La località ospita infatti diverse scuole sci che si spartiscono la gran parte degli appassionati del comprensorio; la loro posizione ravvicinata offre al cliente la totale libertà di scelta e la strategia di business riveste quindi un ruolo molto importante per emergere e imporsi in uno scenario di questo tipo.

Le attività sportive invernali, in particolar modo quelle legate a sci e snowboard, sono generalmente rivolte ad una clientela con buona disponibilità economica; attrezzature, ski pass e corsi non sono alla portata di tutti e alcune scuole sci competono sul prezzo di abbonamenti e lezioni per attirare la fetta di clientela "low budget". Altre, come nel caso della nostra scuola, competono sulla diversificazione del servizio al fine di offrire al socio una customer experience di livello.

Il trend della domanda è di natura ciclica, non solo nel corso dell'anno dove per ovvie ragioni l'affluenza è più elevata nei mesi freddi, ma anche nel corso della

settimana. Durante il weekend infatti, il numero di partecipanti ai corsi e alle lezioni è superiore al doppio rispetto agli altri giorni della settimana [Figura 5].

La ciclicità della domanda porta le scuole sci a competere maggiormente durante il weekend per attrarre più clienti possibili, mentre nei giorni feriali di bassa affluenza è conveniente la collaborazione tra le scuole per attirare l'attenzione dei turisti. Nel corso dell'anno la competizione resta elevata in quanto le scuole si impegnano nell'offerta dei servizi estivi; attività come mountain-bike, arrampicata ed escursioni attirano clienti anche nelle stagioni calde e offrono alle scuole l'opportunità di catturare pubblico anche quando le piste sono chiuse.



[Figura 5] Distribuzione delle partecipazioni, in termini relativi, nel corso della settimana.
Fonte: dati scuola sci e snowboard

Il meteo rappresenta un fattore chiave per l'attività delle scuole sci. Le condizioni meteorologiche associate all'altezza del manto nevoso influenzano pesantemente la domanda, anche quando le previsioni del meteo non dovessero essere corrette.

Il cliente medio, proveniente da zone limitrofe al comprensorio sciistico, basa la programmazione del suo weekend esclusivamente sulle condizioni meteorologiche;

se queste non dovessero rispettare le previsioni, il cliente sarebbe in ogni caso perso dalla scuola. Questo rende le analisi predittive a lungo termine particolarmente complicate in quanto l'affluenza dipende in parte dalle condizioni meteo, non prevedibili con largo anticipo. Secondo i dati raccolti da JFC Tourism & Management in un report [13] che verrà approfondito nella prossima sezione, il fattore che influenza maggiormente la domanda è la presenza di neve al suolo. Nella tabella [Figura 6] qui riportata si nota come in alcuni periodi di minor affluenza (quali per esempio il periodo prenatalizio e dalla seconda metà di marzo fino a Pasqua) risentano maggiormente delle condizioni del manto nevoso (definito "snowfall impact").

Altri periodi, principalmente in corrispondenza delle festività, risentono meno delle condizioni meteo in quanto l'organizzazione delle vacanze, dovuta al periodo dell'anno, prescinde dai fattori casuali.

SNOWFALL IMPACT		
PERIODO	**Total white**	**Partial white**
Pre 8 dicembre	+14%	+6%
Ponte dell'Immacolata	+25%	-8%
Periodo pre-Natale	+77%	+23%
Natale, Capodanno ed Epifania	+5%	=
Dal 7 a fine gennaio	+38%	+11%
Febbraio	+24%	-3%
Carnevale (1/10 marzo)	+11%	-21%
Dall'11 a fine marzo	+41%	-8%
Aprile (sino a Pasqua)	+79%	-9%

[Figura 6] Impatto delle condizioni del manto nevoso sull'affluenza media di periodo

2.3 Panoramica del settore

Affrontiamo ora una prima analisi del settore sciistico presentando alcuni degli aspetti più rilevanti dello studio effettuato dal centro ricerca specializzato Skipass Panorama Turismo di JFC Tourism & Management [15]. L'attività di ricerca di JFC è finalizzata alla previsione del trend del settore a livello nazionale Italiano e il report di riferimento è relativo alla stagione invernale 2019-2020.

Per concludere le considerazioni sul contesto geografico, lo studio offre una classifica dei comprensori sciabili, aventi oltre 100km di piste, più gettonati d'Italia che tiene in considerazione una serie di fattori (quanto una meta è famosa, quanto è trendy, family friendly, meglio servita, migliori alberghi e migliori piste).

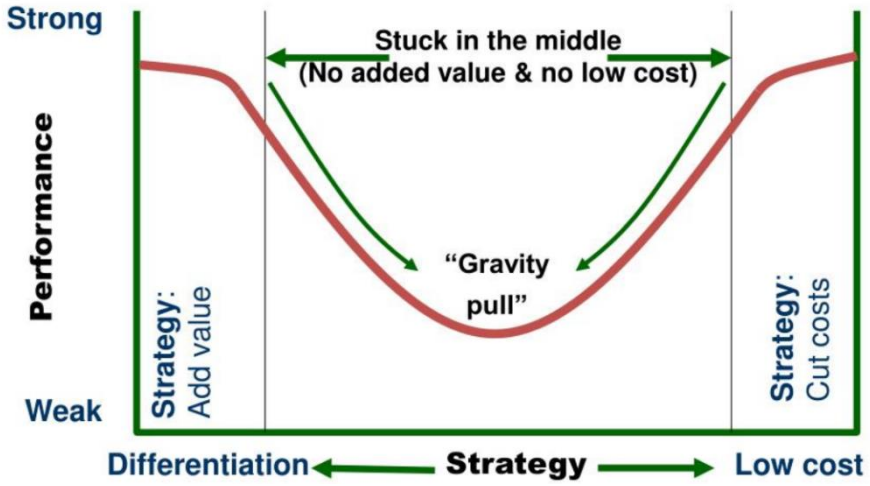
La nostra località si colloca al 10° posto della classifica generale, dietro alle maggiori località del Trentino, ma prima in Piemonte. Da segnalare il 7° posto nella classifica dei comprensori aventi le migliori piste da sci.

Se si guarda invece alle destinazioni bianche più gettonate, la nostra località scende al 15° posto in Italia, seconda in Piemonte nella classifica generale ma all'11° posto nella graduatoria "family friendly". Si conferma quindi una località gettonata in primo luogo dalle famiglie, offrendo quindi alle scuole sci l'opportunità per un certo target di mercato.

Un primo dato di interesse evidenziato dallo studio riguarda la distinzione tra clientela *low budget* (ovvero quella più attenta alla spesa o con meno disponibilità economica) e *big spender* (disposto a spendere pur di non rinunciare alla qualità del servizio). Il report suggerisce come questa distinzione sia sempre più netta per quanto riguarda la clientela italiana e fornisce da spunto per una riflessione di

carattere strategico. All'aumentare del divario tra le due categorie di clienti, cresce il rischio per le scuole sci (ed in generale per tutti gli organismi che si dedicano all'intrattenimento e al turismo) di rimanere "intrappolati nel mezzo".

Il fenomeno definito *stuck in the middle* [Figura 7] ha luogo quando un'azienda non compete sul prezzo ma non adotta nemmeno una strategia di differenziazione, offrendo quindi una discreta qualità ad un prezzo non sufficientemente basso.



[Figura 7] Rappresentazione del fenomeno "Stuck in the middle"

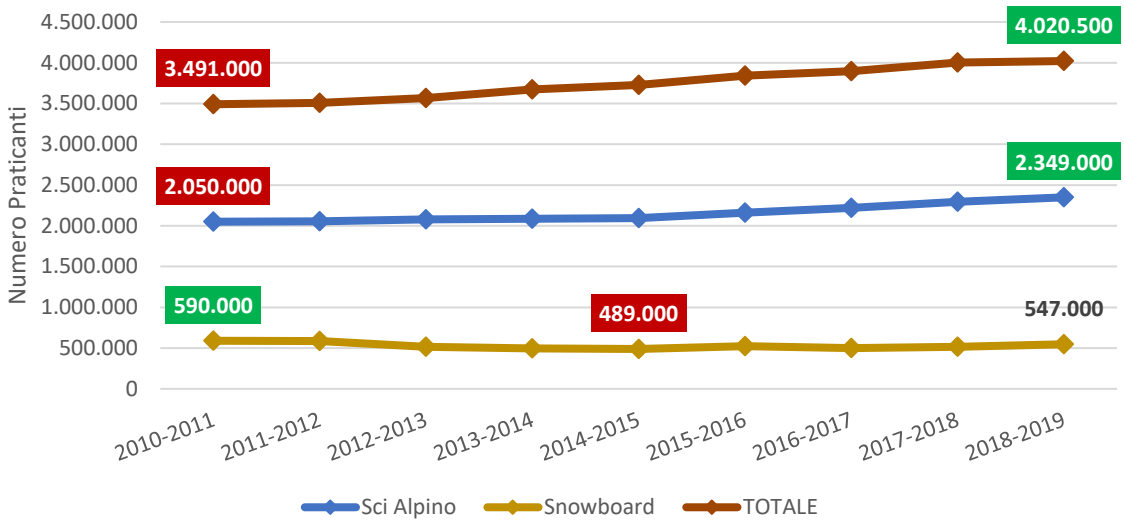
Il rischio è quindi quello di non accontentare né il cliente low budget, né quello big spender, con la possibilità concreta di uscire dal mercato.

Altro dato interessante è quello relativo al trend dei praticanti di sport invernali negli ultimi 10 anni. Il periodo coperto dallo studio va dalla stagione 2010-2011 alla stagione 2018-2019 che ha sostanzialmente confermato le previsioni del report dell'anno passato, leggermente sovrastimate.

Il grafico riportato [Figura 8] è stato generato dai dati dello studio JFC e mostra il trend delle due principali discipline sportive, sci alpino e snowboard, oltre al totale

delle discipline (comprensivo anche di tutte le discipline minori escluse dal grafico quali telemark, ciaspole, sci di fondo, freestyle, scialpinismo e altre).

Il trend generale, così come quello dello sci alpino, è in costante ascesa mentre lo snowboard, dopo il boom iniziale dei primi anni 2000, non riportati nel grafico, ha subito una leggera flessione (fino a un minimo di -17% rispetto al dato 2010-2011) per poi tornare in crescita.



[Figura 8] Trend discipline a livello nazionale

Per l'analisi sono state considerate soltanto le discipline "primarie". Lo studio è quindi particolarmente utile per rispondere alla domanda "qual è la disciplina più gettonata da chi si avvicina per la prima volta al mondo dello sport invernale?".

Tra le varie, lo snowboard è quella che incuriosisce maggiormente dopo lo sci alpino (scelta dal 24% di chi si avvicina al mondo dello sport invernale), ma si evince dallo studio che solo in minima parte viene poi effettivamente praticato da chi vorrebbe provarlo; questo fatto apre le porte alle scuole sci per l'offerta di corsi mirati a conquistare una potenziale fetta di pubblico finora non raggiunta.

Dal report dell'anno precedente [16] emerge una considerazione interessante per quanto riguarda i frequentatori della montagna italiana. JFC ha chiesto ai rispondenti quali fossero le motivazioni alla base della scelta delle mete per le vacanze invernali, prendendo in considerazione come parametri la natura motivazionale (soddisfazione della propria passione, sport), economica, sociologica (esigenza di sicurezza, comfort) e aggregativa (stare insieme a famigliari e amici, socializzazione). Il risultato evidenzia una netta propensione per il senso di aggregazione e voglia di compagnia (oltre il 31% dei voti) e un forte interesse per lo sport (natura motivazionale al 27%). Chi frequenta la montagna predilige quindi l'attività sportiva e tende a praticarla in compagnia: solamente il 6,3% dei rispondenti frequenta la montagna in solitudine.

Si conferma quindi la prospettiva per la quale la maggior parte degli interessati alle mete bianche trascorra le vacanze in compagnia, principalmente insieme a famigliari. Quello che emerge dallo studio è che questi gruppi contengano nella maggior parte dei casi almeno una persona non interessata all'attività sportiva bensì a quella che viene definita "enjoy nature", ovvero la ricerca di luoghi da visitare per goderne della vista, aria pulita e tranquillità dell'ambiente. Si tratta di persone che durante l'attività sportiva di amici e famigliari entrano nella fase del "che fare?", domanda la cui risposta può aprire le porte a nuovi modelli di business che porterebbero le scuole sci ad un nuovo livello di organizzazione, trasformandole in veri e propri centri servizi.

Un esempio perfetto è quello dei "kinder park", aree delle scuole sci dedicate ai bambini. Queste svolgono una doppia funzione: se da un lato rappresentano un

aiuto per i genitori che possono concentrarsi nello svolgimento dell'attività sportiva senza pensieri, dall'altro permettono al bambino di avvicinarsi al mondo dello sci fin dalla tenera età e di trovarsi in un ambiente preparato e famigliare.

Concludendo, dallo studio JFC emergono spunti interessanti in ottica scuola sci per quanto riguarda i target di mercato (gruppi e famiglie), le strategie da adottare (competizione sul costo o differenziazione del servizio) e nuovi modelli di business (centro servizi, kinder park). Possiamo inoltre prendere come base di partenza per le nostre analisi il trend generale del settore degli ultimi 10 anni e l'influenza stimata delle condizioni meteorologiche sull'affluenza turistica invernale, con l'obiettivo di confrontarli con i dati della scuola sci del nostro caso di studio.

2.4 Stato dell'arte

La scuola sci e snowboard oggetto di questa tesi si occupa esclusivamente delle attività operative del business. I dati dei nuovi soci vengono registrati manualmente dagli operatori della scuola e gestiti dal sistema informativo che ne memorizza le anagrafiche, le iscrizioni ai corsi, le prenotazioni delle lezioni e gestisce il cosiddetto *tabellone* [Figura 9], strumento indispensabile agli addetti per monitorare le attività della scuola.

Il tabellone offre agli operatori una visione d'insieme della schedulazione delle attività della giornata. Per ogni maestro sono riportate le lezioni previste in ogni slot temporale, con la specifica del corso ed eventualmente dell'allievo in caso di lezione individuale. La principale caratteristica del tabellone è quella di poter fornire, grazie alle celle e ai colori contrastanti, un forte impatto visivo all'operatore

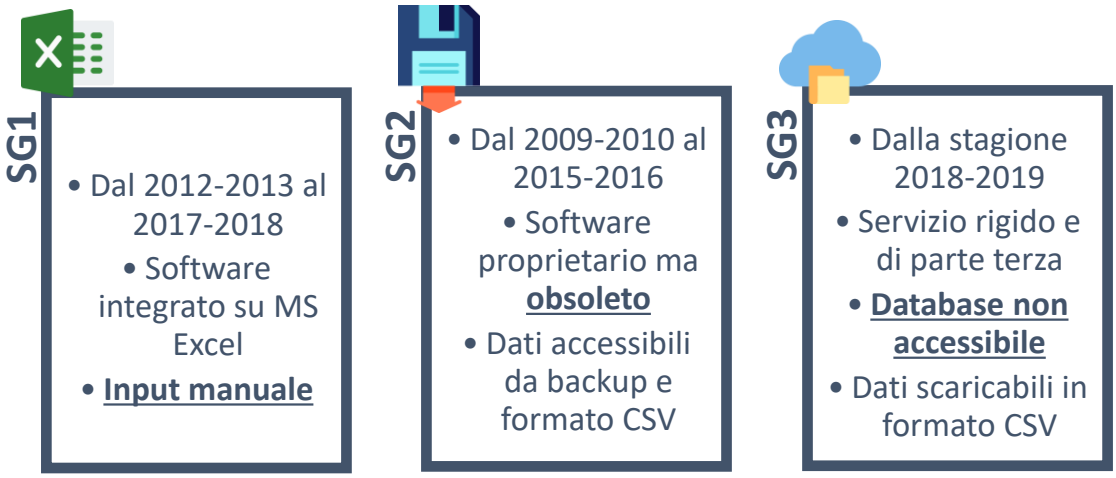
01/01/2019	09:00 - 09:30	09:30 - 10:00	10:00 - 10:30	10:30 - 11:00	11:00 - 11:30	11:30 - 12:00
Maestro 1		Corso 1	Corso 1		Corso 2	Corso 2
Maestro 2						
Maestro 3			Corso 3	Corso 3	Corso 3	
Maestro 4	Corso 4	Corso 4	Individuale	Individuale		
Maestro 5						
Maestro 6			Corso 5	Corso 5	Corso 5	Corso 5

[Figura 9] Esempio semplificato di un tabellone

che deve schedulare le nuove attività, permettendogli di individuare facilmente gli slot liberi e i maestri disponibili per la lezione.

L'insieme dei dati gestiti dal sistema informativo viene memorizzato in un database e presentato tramite interfaccia grafica che permette di recuperare i dati e aggiornarli in tempo reale andando a programmare nuovi corsi, schedulare lezioni, iscrivere nuovi soci, creare nuovi servizi e compilare fatture.

Allo stato attuale delle cose, la nostra scuola possiede uno storico dati di circa 10 anni e ha monitorato la propria attività attraverso l'utilizzo di tre diversi software gestionali [Figura 10].



[Figura 10] Software gestionali della scuola sci

- **SG1** Periodo di utilizzo: dalla stagione 2012-2013 alla stagione 2017-2018. Il software è proprietario e integrato su Microsoft Excel che gestisce le relazioni tra le tabelle attraverso l'utilizzo di macro: non è presente quindi un vero database. I dati sono accessibili in formato .xlsx e contengono le informazioni relative esclusivamente al periodo invernale dell'area A.P.D.
- **SG2** Periodo di utilizzo: dalla stagione 2009-2010 alla stagione 2015-2016. Il software è proprietario e gira esclusivamente su una vecchia release di Windows, non essendo stato più supportato dallo sviluppatore. Il software costituisce l'interfaccia per l'operatore della scuola e dialoga con un DBMS per la gestione del database relazionale attraverso query SQL.

  I dati sono quindi accessibili tramite backup del DBMS che gira in locale sui server della nostra scuola sci e fanno riferimento al periodo sia estivo che invernale dell'area Scuola Sci.
- **SG3** Periodo di utilizzo: dalla stagione 2018-2019 in avanti. Il software consiste in una applicazione web gestita da terze parti che offre il vantaggio di supportare la connessione da dispositivi diversi (operatori da postazione pc, maestri da dispositivo mobile) ma non consente l'accesso diretto ai dati. Il database è infatti gestito dagli sviluppatori del software e i dati sono disponibili esclusivamente tramite download di file .csv senza possibilità di personalizzare il download. Non è quindi possibile ottenere le informazioni nella loro totalità e i file ottenuti sono di difficile lettura e modifica. Il gestionale contiene i dati di tutte le aree della scuola sci.

Lo stato dell'arte è quindi rappresentato da una considerevole quantità di dati eterogenei, difficoltosi da acquisire, leggere e analizzare. La maggior parte di queste informazioni non è mai stata elaborata o posta come base per una presa di posizione nelle decisioni del management che si è quasi sempre affidato all'esperienza del personale e alla percezione degli eventi legati al business.

2.5 Il perché dell'analisi

L'analisi dei dati proposta vuole quindi fornire alla scuola uno strumento in grado di evolvere il processo decisionale attuale, basato su esperienza e sensazioni, in un approccio *data-driven* in cui i dati guidano le decisioni.

L'obiettivo è quello di porre le basi per trasformare la scuola sci del nostro caso di studio da semplice scuola a vera e propria azienda capace di offrire ai propri clienti servizi mirati e rispondere tempestivamente alle esigenze del mercato.

BISOGNI	SFIDE	OPPORTUNITÀ
Monitoraggio delle attività Individuazione delle **criticità** **Proposte efficaci**	**Attrattività** **Competitività** **Innovazione**	**Espansione** **Fidelizzazione** del cliente **Ottimizzazione** del servizio

[Figura 11] Bisogni, sfide e opportunità per una scuola sci

Una volta definiti lo stato dell'arte e il contesto in cui si colloca l'azienda, è importante focalizzare l'attenzione su quelli che possono essere i bisogni della

scuola, le sfide da affrontare per emergere nel settore e le opportunità che è possibile cogliere a seguito di un'analisi dati [Figura 11].

2.5.1 Bisogni

Le necessità della scuola sci a cui può rispondere un'analisi dati sono quelle anticipate in [Figura 1] nel primo capitolo, le fondamenta della disciplina che abbiamo chiamato Business Intelligence: essere in grado di rispondere alle domande "Come?", "Perché?" e "Cosa succede se...?" è la base di partenza per intraprendere azioni correttive e quindi fornire risposte efficaci.

- **Monitoraggio delle attività** Il primo bisogno della scuola sci è comprendere cosa stia succedendo all'interno del business in termini quantitativi. La percezione degli addetti è importante ma avere una fotografia di ciò che è stato e di ciò che è la situazione attuale permette di avere chiara la direzione che si sta intraprendendo. Come sta andando una determinata area della scuola? Quanto sta vendendo un servizio? Quanti allievi sono assegnati a un maestro? Sono tutte domande a cui può rispondere una semplice analisi dei dati, fornendo una prima panoramica dell'andamento della scuola.
- **Individuazione delle criticità** Conseguenza diretta del monitoraggio delle attività è l'individuazione degli aspetti critici del business. Se mi aspetto un trend stabile e scopro che invece è in decrescita, è importante saper rispondere alla domanda "Perché?". La risposta è nell'analisi di dettaglio. Il dato può essere scomposto e studiato a differenti livelli di

dettaglio per capire quali siano le caratteristiche e le correlazioni tra gli attributi che influiscono sull'aspetto che abbiamo ritenuto critico. Capite le cause è possibile quindi offrire una risposta mirata al problema.

- **Proposte efficaci** L'analisi effettuata è quindi in grado di suggerire proposte (o quanto meno fornire spunti) efficaci per correggere o migliorare le attività. Una volta analizzati i trend e individuate le criticità, il processo decisionale per intraprendere una strada piuttosto che un'altra è semplificato dai risultati delle analisi. È tuttavia importante sottolineare come il ruolo dell'analisi nel processo decisionale non sia sostitutivo dell'esperienza e dell'intuizione del management, bensì rappresenti uno strumento complemento ad esse. In base ai dati storici è inoltre possibile effettuare analisi predittive che rispondano alla domanda "Cosa succede se...?" grazie alla creazione di modelli che permettano ai software di auto apprendere in base alle correlazioni tra i dati.

2.5.2 Sfide

L'obiettivo finale dello strumento in esame è quello di facilitare la scuola sci in esame nell'affrontare le sfide del settore, considerando un contesto molto competitivo come quello presentato in 2.2, al fine di permettere alla scuola di emergere ed essere un punto di costante riferimento per gli appassionati dello sci. Rispondere con efficacia alle principali sfide è l'obiettivo di quell'insieme di tecniche e processi decisionali operativi, finanziari e commerciali che viene definito come *strategia d'impresa*.

- **Attrattività** L'esigenza primaria della scuola è quella di espandere la propria rete di clienti. Il business della nostra azienda consiste infatti nel "vendere divertimento" pertanto l'attrattività dell'azienda passa principalmente da marketing, servizi offerti e *customer experience*. Pubblicità, offerte mirate, corsi ed eventi ad-hoc, percorsi offerti ed esperienze positive sono fattori cruciali per l'espansione della clientela raggiunta. Catturare la curiosità del cliente casual o convincere un abituale scontento della concorrenza sono obiettivi chiave per rispondere alla sfida dell'attrattività.
- **Competitività** Aspetto fondamentale alla base della strategia d'impresa è la definizione di vantaggio competitivo: si tratta di un aspetto del business unico e difficilmente imitabile dai competitors in grado di garantire all'azienda un vantaggio in termini di performance.

  La sfida che la nostra scuola si trova ad affrontare è quella di costruire un proprio vantaggio competitivo che ne consenta, oltre alla permanenza sul mercato, la possibilità di eccellere e di affermarsi come leader di settore. Il vantaggio competitivo viene così definito da Enrico Valdani: "*capacità distintiva di un'impresa di presidiare, sviluppare e difendere nel tempo, con maggiore intensità dei rivali, una capacità market driving o una risorsa critica che possono divenire fattori critici di successo*" [15].
- **Innovazione** Tratto da sempre distintivo e fonte del successo dell'azienda è la capacità di rinnovarsi e offrire ai propri clienti servizi unici che garantissero loro un'ottima esperienza: è ragionevole pensare quindi che la customer experience rappresenti attualmente il vantaggio competitivo

della scuola sci in esame. In realtà, stando alla definizione di Valdani, risulta complicato pensare alla customer experience come un vero vantaggio competitivo, quanto più come caratteristica distintiva, ma non duratura, dell'azienda. Questo perché la percezione dell'esperienza varia da persona a persona, è strettamente correlata alle capacità sia tecniche che interpersonali degli istruttori e le richieste dei clienti possono variare di anno in anno. Inoltre, i servizi offerti possono essere in parte imitati dai competitor.

Il vantaggio competitivo può essere perseguito attraverso il cambiamento, sia esso interno o esterno all'azienda. Esempi di mutamenti esterni possono essere l'ingresso di nuovi competitors, l'evoluzione della tecnologia o l'affermarsi di un nuovo sport invernale. Il vantaggio competitivo può quindi essere raggiunto attraverso la capacità di prevedere questi cambiamenti e la velocità nel rispondere ad essi. Il modello di analisi proposto vuole essere invece un esempio di cambiamento interno all'azienda (leggasi innovazione): un mutamento del processo decisionale a partire dai dati che, se acquisito in anticipo rispetto ai rivali e reso parte integrante del business della nostra scuola sci e può rappresentare un vantaggio competitivo per l'azienda.

Per individuare nella caratteristica un vantaggio competitivo bisogna effettuare quella che viene chiamata VRIO Analysis [Figura 12]. La caratteristica, sia essa una risorsa o una competenza, deve infatti essere:

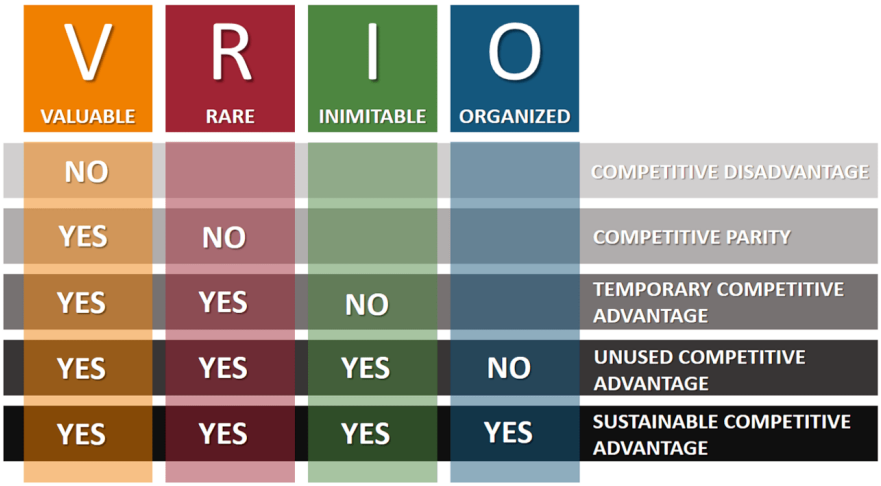
- **Valuable (V)** Di valore per l'azienda. Una risorsa è di valore quando permette all'impresa di implementare strategie o processi che ne migliorino

l'efficienza, consentano di cogliere opportunità o mitigare le minacce del business. Sono quindi di valore tutte quelle caratteristiche (brand, posizione geografica), risorse (attrezzature, persone, tecnologie) e competenze (know-how, figure professionali ed esperienza) che permettono all'azienda di svolgere la propria attività.

- **Rare (R)** Rara. La risorsa deve essere a disposizione di poche aziende del settore. Se la risorsa è di valore ma non rara si parla di *competitive parity* in quanto la concorrenza è in grado di sfruttare la stessa risorsa annullando di fatto il potenziale vantaggio competitivo.
- **Inimitable (I)** La caratteristica deve essere difficilmente imitabile dalle imprese competitor. Una risorsa è considerabile come non imitabile quando è troppo costosa per essere imitata; in particolare se è frutto di decisioni prese dall'azienda in passato, se il legame tra vantaggio competitivo e risorsa non è ben definibile o se il contesto da cui scaturisce il vantaggio non è imitabile per una combinazione di know-how aziendale, relazioni, asset.
- **Organization (O)** L'impresa deve essere organizzata per estrarre valore dalla risorsa di valore, rara e inimitabile. Se manca l'organizzazione attorno alla caratteristica, si parla di potenziale vantaggio competitivo.

  Nel caso della nostra scuola sci, l'organizzazione si riflette nel modo in cui i dati vengono acquisiti, nelle relazioni con i clienti, nella gestione delle attività, nell'organizzazione dei corsi, nel modo in cui gli addetti comunicano (staff, istruttori, management).

Questo è il motivo per il quale il modello proposto non può limitarsi ad essere un semplice tool in aiuto al processo decisionale bensì rappresentare una base di partenza per l'innovazione della scuola sci, sposando il concetto di digital transformation.



[Figura 12] VRIO Analysis
Fonte: Business2You

Ottimizzazione dei servizi, gestione più accurata delle attività, interventi mirati e tempestivi: organizzare la società attorno a questi principi e renderli parte integrante e persistente del business consentirebbe alla scuola sci di trasformarsi in azienda e consolidare ed evolvere le proprie *core competencies*. Una *core competency* è una competenza chiave attorno alla quale è basato il business dell'impresa. Nel nostro caso può essere definita una core competency la capacità di insegnare a sciare o, più in generale, vendere divertimento. Il vantaggio competitivo consisterebbe nel rendere unica sul mercato una propria core competency: vendere divertimento in maniera innovativa per quanto riguarda il settore di riferimento, offrire al cliente un'esperienza nuova che non potrebbe ripetere in altre scuole.

2.5.3 Opportunità

Rispondendo efficacemente ai propri bisogni e alle sfide del settore, la scuola sci e snowboard in questione, sarà in grado di cogliere le opportunità di un mercato, come quello dello sport invernale, tanto competitivo quanto poco dinamico.

Come anticipato nella descrizione dello scenario competitivo infatti, il settore delle scuole sci è caratterizzato da aziende, se così possono essere definite, molto classiche e poco volte all'innovazione, le cui offerte (intese come corsi, eventi, lezioni, attività) sono dettate dall'esperienza di staff e dirigenza. Si tratta di un settore piuttosto statico in cui le opportunità sono diverse per una scuola in grado di adottare un processo decisionale basato su business intelligence.

- **Espansione** Un approccio data-driven permetterebbe in primo luogo di focalizzare l'attenzione sulle opportunità di crescita della scuola sci. Capire come attrarre nuovi clienti, che tipo di pubblicità occorre e dove conviene effettuarla, quali offerte possono attrarre più appassionati e in quale momento dell'anno sono più proficue. Questi sono solo alcuni dei quesiti ai quali la scuola può attualmente rispondere in maniera approssimativa e che in seguito ad un'analisi di dettaglio possono notevolmente incrementare il bacino di clientela della scuola sci. Più soci conta la scuola, maggiori sono le opportunità nel proporre nuovi modelli di business e al tempo stesso le analisi aumentano di efficacia grazie al maggior numero di dati diversi a disposizione.

- **Fidelizzazione** Aspetto fondamentale del business della nostra scuola sci è la fidelizzazione del cliente. In un mercato come quello delle scuole sci, molto altalenante in termini di domanda nel corso delle stagioni e al tempo stesso molto competitivo, i nuovi clienti rappresentano una minoranza rispetto alla totalità dei soci iscritti annualmente. Questo perché se il cliente si trova bene in una determinata scuola, sarà poco incline a provarne un'altra. Fidelizzare il cliente vuol dire avere un costante bacino di utenza, anno dopo anno, in grado di ridurre la variabilità della domanda e conferire stabilità anche nelle stagioni di bassa affluenza. L'intento dell'analisi è quello di migliorare la customer experience attraverso corsi, eventi ed attività mirate in base ai trend annuali della domanda e all'età degli iscritti, considerando la loro storia all'interno della scuola per fornire la miglior esperienza possibile.
- **Ottimizzazione** L'individuazione degli aspetti critici e gli spunti forniti dai modelli di business intelligence si traducono principalmente nell'ottimizzazione di attività e servizi offerti. Guidare le proposte di nuove offerte in determinati giorni della settimana o ore della giornata permette una migliore distribuzione del carico di lavoro tra maestri e staff, così come l'organizzazione dei corsi in base al livello dei partecipanti.

  Inoltre, capire in termini quantitativi quali corsi riscuotono più successo e quali siano invece i percorsi meno gettonati, a seconda del tipo di cliente, permette di allocare al meglio le risorse (organizzazione, istruttori, eventi) e intervenire dove necessario di stagione in stagione.

Capitolo 3

Il progetto

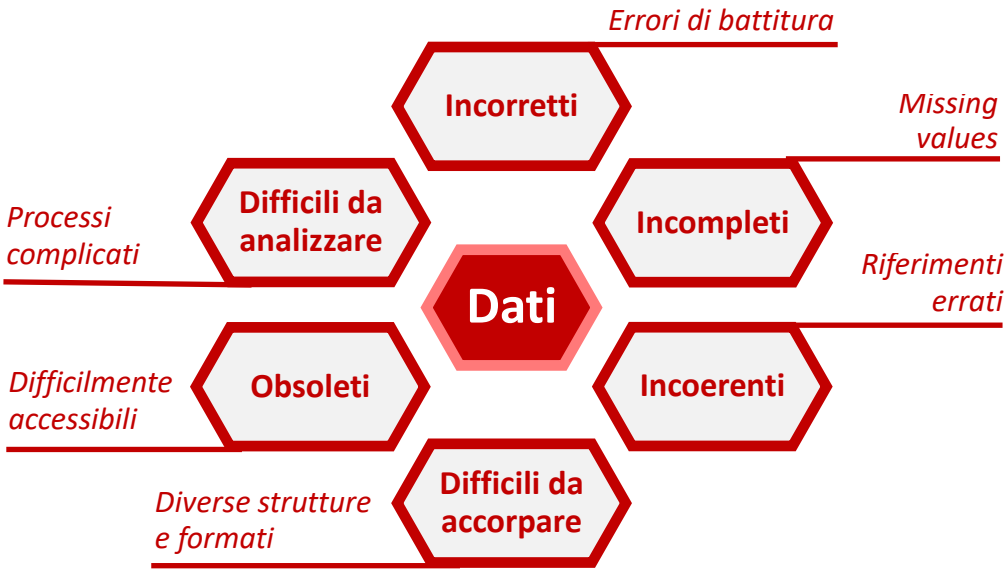
3.1 Data Management

Come presentato nel paragrafo 2.4, lo stato dell'arte consiste nello storico dati di 3 software che organizzano e gestiscono le informazioni in modi totalmente distinti. Ognuno di questi contiene i dati relativi a determinate aree della scuola (APD o Scuola Sci) e solo in determinate stagioni; inoltre ogni stagione è gestita in modo separato dalle altre, quindi per ogni stagione abbiamo una porzione di storico dati gestita da uno dei tre software.

La richiesta della scuola sci di avere un unico database deriva dalla necessità di rendere omogenea e facilmente analizzabile la mole di informazioni raccolte nel corso delle stagioni e in particolare risolvere alcune criticità relative all'interrogazione dei dati provenienti da fonti differenti [Figura 13].

In primo luogo, l'input manuale dei dati per opera degli addetti della scuola è fonte di errori di battitura, inesattezze e imprecisioni che si traducono nell'incorrettezza degli stessi. Inoltre, una gestione poco efficace dei dati da parte dei software e una

politica di data input approssimativa (quali sono i dati da richiedere al cliente? Quali sono quelli da conservare? Quali aggiornare?) rendono le informazioni incomplete e, inevitabilmente, anche le analisi.



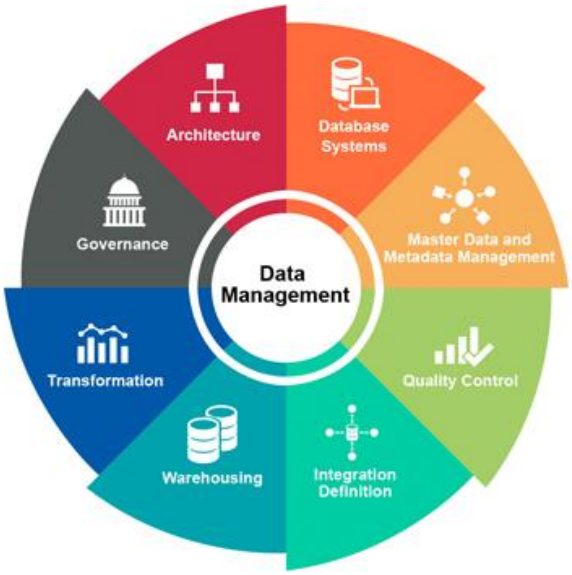
[Figura 13] Problematiche relative ai dati

La libertà di inserimento lasciata agli operatori ha il vantaggio di agevolare e velocizzare alcune operazioni che, seppur funzionali allo svolgersi delle attività operative, rappresentano un grosso ostacolo per il successivo lavoro di business intelligence. Un primo passo per risolvere il problema a lungo termine sarebbe quello di ripensare dalla base i meccanismi di interazione con il cliente, coniugando, seguendo i principi della digital transformation, una corretta gestione dei dati con la necessità dello staff di accelerare i tempi di iscrizione dei soci, in particolar modo nei giorni di alta affluenza nei quali i clienti affollano la sede della scuola.

Una delle principali problematiche della corrente gestione dei dati è relativa alla coerenza degli stessi. Poiché ogni stagione di ciascuna area della scuola sci è gestita indipendentemente dalle altre, i riferimenti tra una stagione e l'altra sono completamente inesistenti. Lo stesso socio viene etichettato con ID diversi a seconda della stagione e lo stesso corso o servizio offerto vengono identificati non solo con ID diversi, ma anche con nomi differenti in seguito all'azzeramento del database tra una stagione e l'altra.

All'interno della stessa stagione è possibile trovare incoerenze tra i dati o addirittura riferimenti mancanti; se a questo si aggiunge la provenienza da fonti differenti e quindi il diverso formato dei dati, risulta chiaro come tutto ciò renda impossibile ogni tipo di analisi, in particolar modo incrociata tra aree della scuola o pluriennale.

La necessità primaria della scuola è quindi quella di migliorare la gestione del dato e renderlo un asset di valore per il business, ottimizzando il processo di *data management* [Figura 14].



[Figura 14] Aspetti chiave del Data Management
Fonte: MKT Softwares

Senza entrare nel merito del significato del termine e delle sue caratteristiche, basti sapere che con data management si intende l'insieme di tutte quelle discipline volte al controllo dei dati dal momento della raccolta degli stessi alla preparazione per l'analisi.

La prima carenza della scuola da questo punto di vista risiede nel meccanismo di acquisizione delle informazioni. Allo stato attuale, la registrazione del nuovo cliente avviene per mano degli operatori che devono però gestire migliaia di nuovi soci per stagione e, dovendo inserire i dati comunicati a voce sul momento, sono spesso inclini ad errori. Inoltre, i dati richiesti al cliente sono essenziali alla scuola per le attività operative ma non sono pensati per il lavoro di analisi; per esempio, come vedremo nel prossimo capitolo, un aspetto chiave del business è quello relativo al comportamento delle famiglie che è però trascurato a livello di informazioni memorizzate: non vi è alcun riferimento al fatto che un socio sia parente di un altro, pur quando il software gestionale ne preveda la possibilità.

Una possibile soluzione potrebbe essere quella di permettere al cliente di inserire autonomamente i propri dati attraverso form strutturati che minimizzino il rischio di errori e informazioni mancanti, oltre ad alleggerire il carico di lavoro dello staff.

Il secondo aspetto da tenere in considerazione parlando di data management è quello relativo allo *storage*. La scuola attualmente gestisce i propri dati per stagione; al termine di ogni stagione le informazioni memorizzate vengono archiviate e la base dati azzerata. Se per quanto riguarda le attività operative questo metodo attribuisce al sistema ordine e una certa leggerezza, per le attività di business intelligence è deleterio.

Infatti, suddividere la gestione dei dati per stagione porta a:

- Dati duplicati: i dati di uno stesso socio che frequenta la scuola per più stagioni vengono inseriti numerose volte;
- Riferimenti errati: ID diversi per uno stesso oggetto (che sia socio, corso, servizio, lezione...) e ID uguali per oggetti diversi;
- Difficile gestione dei rinnovi: come posso sapere chi è un nuovo socio e chi ha rinnovato se lo storico dati è archiviato?

Oltre agli esempi presentati si potrebbero riportare decine di altri scenari emersi durante il lavoro di aggregazione dati, ma l'obiettivo è semplicemente quello di capire come una corretta gestione dello storage sia fondamentale per l'attività di analisi. Quella che viene definita *storage management* è la disciplina figlia del data management che si occupa della gestione del dato all'interno del sistema, in particolare della struttura della base dati e dell'architettura che la implementa.

Obiettivo dello storage management è standardizzare il processo di memorizzazione delle informazioni, oltre al formato dei dati, e permetterne il recupero; su quest'ultimo punto è importante precisare come al momento la scuola non abbia la possibilità di accedere ai propri dati. Il software gestionale attualmente utilizzato è infatti gestito da terzi e accessibile da web app, ma il servizio non consente l'accesso diretto alla base dati, impedendo alla scuola di ottenere una panoramica completa dello storico.

L'architettura che sta alla base del data management della scuola andrebbe quindi ripensata per consentire un migliore controllo dell'acquisizione delle informazioni e la corretta gestione dei dati all'interno del sistema, oltre al loro recupero.

3.2 Extract, Transform, Load

Scendiamo ora nel dettaglio di quanto svolto nell'attività di tirocinio e di quale sia l'obiettivo dello studio. Lo scopo del lavoro può essere idealmente suddiviso in due macro-obiettivi:

1. Aggregare lo storico dati della scuola sci in un unico database al fine di rendere i dati omogenei, aggiornabili e semplici da gestire.
2. Analizzare in maniera completa i dati per capire, dove possibile, le interazioni tra le diverse aree della scuola e fornire al management indicazioni e spunti utili per capire dove e come intervenire.

Il primo obiettivo è stato raggiunto attraverso quello che viene definito ETL (Extract, Transform, Load) ovvero un insieme di processi di estrazione, trasformazione e caricamento dei dati in un unico sistema di sintesi.
Il lavoro di ETL è stato eseguito con il fine di risolvere, almeno in parte, i problemi elencati nella sezione 3.1. In particolare, il processo è stato definito come segue:

1. **Estrazione dati** In primo luogo è stato necessario estrarre i dati dalle rispettive fonti. Per tutti i software gestionali sono stati ripristinati, dove possibile, i dati memorizzati stagione per stagione e convertiti in un formato leggibile in maniera semplice dagli strumenti di business intelligence in nostro possesso.
2. **Prima aggregazione** Per ogni fonte sono stati aggregati i dati delle diverse stagioni al fine di avere tre dataset comprensivi di tutte le informazioni memorizzate nei tre software gestionali.

3. **Trasformazione** Una volta studiata la struttura dei dati, ogni dataset è stato trasformato in base alle informazioni contenute per adattarsi a un modello generale che potesse rappresentare i dati in modo coerente ed efficace. La base del modello, su richiesta della scuola, è stata fornita dalla struttura del database più completo dei tre software, poi modificata ed adattata al fine di raccogliere gli altri dati e permettere determinate analisi indicate come particolarmente interessanti dal management.

   Sono state effettuate inoltre le prime operazioni di data cleaning e creazione di nuovi attributi che permettessero alle future analisi di esaminare anche informazioni implicite nei dati e approfondire alcuni aspetti del business.

4. **Aggregazione** Sono stati quindi riuniti tutti i dati in un'unica struttura, creando nuovi attributi che specificassero stagione e software di riferimento al fine di permettere la ricostruzione delle relazioni tra oggetti (tabelle) ed eliminando, dove possibile, i dati duplicati. Infine, è stato possibile esplicitare alcune caratteristiche dei record (per esempio se un'iscrizione fosse relativa a un nuovo socio o a un rinnovo) in seguito alla reindicizzazione di soci e corsi, offrendo una visione d'insieme più chiara dei dati a disposizione.

Aspetto fondamentale del processo di creazione del dataset unificato è stato il data cleaning; la quantità di dati "sporchi" e riferimenti errati avrebbe reso totalmente inutilizzabile il dataset. Sono stati quindi eliminati i record non significativi (blank record, test record), ripristinati gran parte dei riferimenti errati e corretti alcuni dati fondamentali, quali i nominativi di anagrafiche e iscrizioni ai corsi.

3.3 Database

Il risultato dei processi di ETL è quindi un dataset coerente e comprensivo dei dati relativi a 10 anni di attività della scuola sci, per un totale di 13 tabelle e una mole di circa 3 milioni di celle. La scelta di un database relazionale tradizionale è dovuta alla tipologia dei dati in possesso della scuola, tutti strutturati seppur diversi nella forma. Per lo scopo del lavoro, quindi per l'aggregazione e la trasformazione di questi dati, un database relazionale rende i processi di ETL più immediati oltre a permettere a management e staff di esaminare i dati in un formato a loro familiare, ma nulla vieta alla scuola di poter adottare in futuro un approccio basato su NoSQL. Con il termine NoSQL si intende un sistema di organizzazione dei dati caratterizzato dal fatto di non utilizzare il modello relazionale tipico dei tradizionali database RDBMS (Relational DataBase Management System).

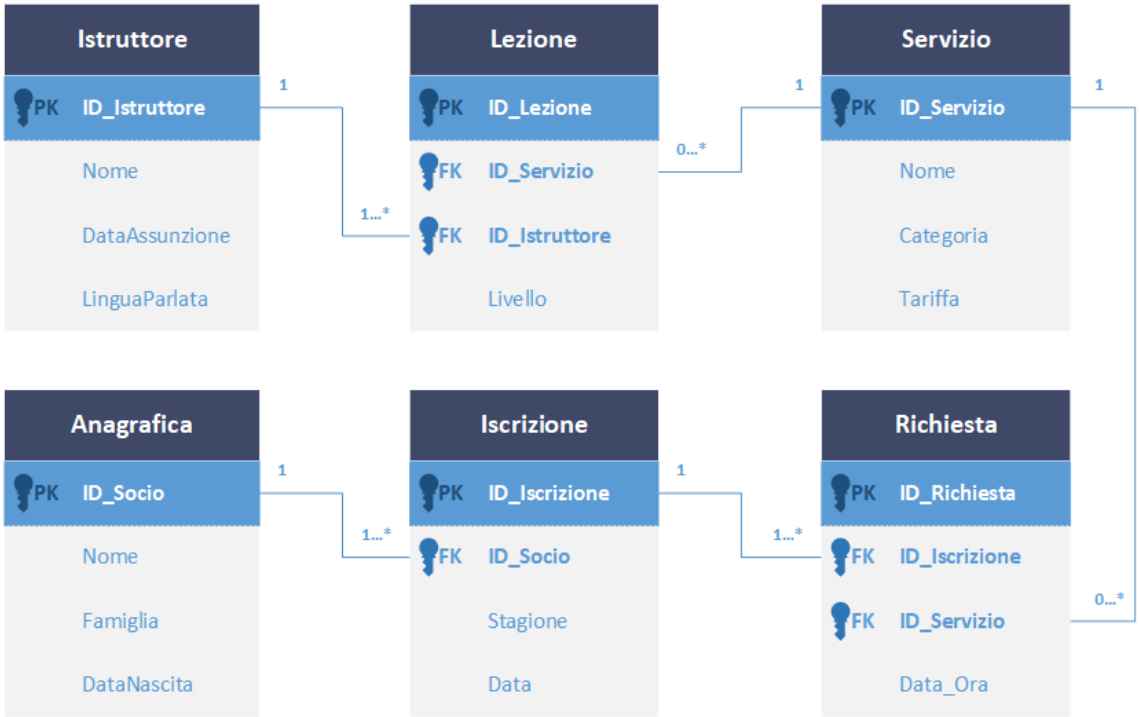
La principale caratteristica dei sistemi non relazionali risiede nell'assenza di uno schema fisso (tabelle) che permette al modello di gestire dati eterogenei ed aventi attributi diversi, rendendolo particolarmente funzionale in ambito big data. Considerando la realtà in cui si trova, la nostra scuola sci non trarrebbe al momento un reale beneficio dall'utilizzo di tale tecnologia e renderebbe forse le operazioni di data management meno immediate.

Il database è quindi strutturato in tabelle, ciascuna delle quali rappresenta un oggetto di interesse (socio, iscrizione, istruttore, lezione, corso...) caratterizzato da una lista di attributi (per un socio abbiamo il nome, la data di nascita, l'indirizzo...). Ogni istanza di un oggetto è caratterizzata da un attributo (o combinazione di attributi) che ne identifica la *chiave primaria*, univoca per le istanze di ogni tabella:

per esempio, un socio è identificato con un ID univoco, mentre un corso è identificato dal proprio ID e dalla stagione di riferimento perché, provenendo da fonti dati differenti, uno stesso ID è associato a più di un corso.

Le chiavi primarie permettono di mettere in relazione tra di loro gli oggetti: dal momento che ogni record della tabella "Iscrizioni" contiene l'ID univoco del socio a cui quell'iscrizione fa riferimento, è possibile mettere in relazione la tabella dei soci con quella delle iscrizioni. Queste relazioni permettono di incrociare i dati contenuti nelle tabelle attraverso l'operatore *join* del linguaggio SQL e quindi effettuare query articolate, ottenendo un dataset ad hoc per le analisi da effettuare.

Per fornire una panoramica dei dati a nostra disposizione su cui saranno basate le analisi del prossimo capitolo, mostriamo una visione semplificata dello schema del



[Figura 15] Struttura semplificata del database

database che nella sua versione completa risulterebbe di difficile comprensione a causa dell'elevata numerosità di tabelle, relazioni e attributi.

La [Figura 15] mostra i principali oggetti di interesse presenti nel database e le relazioni tra di essi. Ogni oggetto è identificato dalla chiave primaria (PK, Primary Key) e da una serie di attributi, alcuni dei quali fanno riferimento ad altre tabelle attraverso le chiavi esterne (FK, Foreign Key).

Gli elementi di maggior importanza per le attività di analisi sono:

- **Anagrafica** Contiene le informazioni relative a famiglia di appartenenza (o meglio, l'informazione è ricostruita attraverso il cognome del socio), età, sesso e indirizzo di residenza. Ogni record della tabella è relativo ad un socio che all'interno della tabella compare una sola volta, identificato univocamente dall'ID.
- **Iscrizione** Tiene traccia di ogni iscrizione ad un servizio effettuata da un socio, al quale possono essere assegnate diverse iscrizioni durante le stagioni, ma anche nell'arco della stessa stagione. Sono inoltre disponibili le informazioni riguardanti la data di iscrizione, la competenza (Scuola Sci o A.P.D.) e l'importo relativo all'iscrizione. Inoltre, è stato possibile ricavare se un'iscrizione appartenesse ad un nuovo socio (quindi mai iscritto prima) o se fosse corrispondente ad un rinnovo (socio già iscritto nella stagione precedente).
- **Servizio** Con servizio si intende qui, per comodità, un corso generico; in realtà i servizi offerti si distinguono in corsi di breve durata, eventi o iscrizioni di durata stagionale che nelle analisi verranno trattati in

maniera distinta. In generale sono specificati il prezzo e la categoria alla quale quel servizio appartiene.

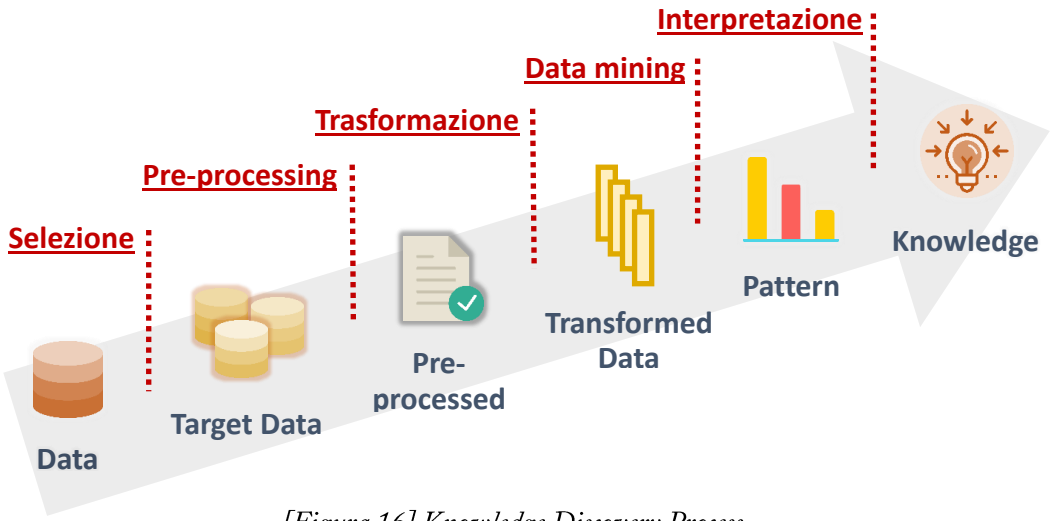
- **Lezione** Per ogni lezione sono indicati il corso di riferimento, l'istruttore che ha sostenuto la lezione, il livello, data e slot orario.
- **Richiesta** Ogni record della tabella si riferisce alla richiesta di un iscritto per una determinata lezione e contiene i dati relativi all'iscrizione a cui la richiesta fa riferimento, tipologia del corso, data e ora. Una terza tabella, che mette in relazione la richiesta con la lezione, permette di individuare i partecipanti ad ogni singola lezione attraverso il join (molto oneroso in termini complessità computazionale) tra le tabelle.

Tramite l'interrogazione del database con query SQL strutturate è possibile incrociare i dati di tutte queste tabelle per effettuare analisi complesse e comprendere quali fattori influenzano maggiormente il business della scuola.

3.4 Knowledge Discovery Process

Il lavoro di ETL svolto si pone come base di un processo di più ampio raggio avente come obiettivo ultimo l'acquisizione di conoscenza dai dati, definito Knowledge Discovery Process (KDP) [Figura 16].

Attraverso il KDP il dato grezzo viene trattato ed analizzato secondo modelli di analisi descrittiva e predittiva, a seconda delle esigenze, per poi essere interpretato e preso in considerazione a valle del processo decisionale. L'analisi descrittiva consiste nell'analisi dei dati storici alla ricerca di cause pregresse ed è particolarmente utile per capire l'andamento del business a più livelli di dettaglio e



[Figura 16] Knowledge Discovery Process

fornire indicazioni sugli aspetti critici delle attività. L'estrapolazione di modelli interpretabili in grado di descrivere i dati è un altro traguardo raggiungibile grazie all'analisi descrittiva: un esempio può essere la segmentazione dei clienti.

L'analisi predittiva ha invece lo scopo di predire eventi o comportamenti futuri: esplicitando alcune variabili conosciute è possibile creare modelli in grado di prevedere, partendo dai dati storici, il valore di altre variabili.

Il KDP può essere visto come un processo a step, che esaminiamo qui nel dettaglio, del quale il lavoro di ETL ha posto le basi per la preparazione dei dati. Infatti, nel nostro caso, il risultato delle operazioni di ETL elencate in 3.2 è stato semplicemente la creazione di un database a partire dai dati grezzi della scuola che rappresenta il punto di partenza del processo.

A differenza dell'ETL, che manipola e modifica i dati in forma permanente per la creazione di una struttura dati, il KDP agisce su una copia dei dati del database generato per effettuare operazioni volte all'analisi dei dati stessi.

Si può scomporre in:

- **Selezione** Una volta generato il database è necessario selezionare i dati utili per le analisi da effettuare. Nel lavoro svolto il processo di selezione è stato affrontato in due modi diversi. In un primo momento le operazioni di analisi sono state eseguite in totale libertà, senza alcuna richiesta da parte degli esperti di dominio, con l'obiettivo di "lasciare parlare i dati". L'attività di selezione è stata quindi eseguita da un agente totalmente esterno al business della scuola mentre, in un secondo momento, ottenuti i primi risultati, le successive analisi sono state effettuate su richiesta degli esperti di dominio e quindi la selezione dei dati è stata guidata dall'obiettivo finale delle analisi.
- **Pre-processing** Le operazioni di pre-processing hanno il fine di rendere il dato pronto per essere trasformato ed analizzato. Queste si possono raggruppare in due macro categorie dette *data cleaning* e *data integration*. Il data cleaning si occupa di ridurre il "rumore" dei dati, di identificare (e talvolta eliminare) gli *outliers* (ovvero quei record i cui attributi contengono valori molto al di fuori della norma, i blank record e i test record) e di risolvere le inconsistenze dei dati. Il processo di data integration è invece volto a uniformare i dati provenienti da fonti diverse e risolvere i conflitti tra essi. Parte del lavoro di pre-processing è stato oggetto dell'attività di ETL ma, tenendo presente che l'obiettivo primario era quello di unificare i dati preservando il più possibile le informazioni originali, le attività di data cleaning in termini di riduzione del rumore e individuazione

degli outlier non sono state affrontate precedentemente alla creazione del database.

- **Trasformazione** Questo step si differenzia da quello affrontato durante le attività di ETL in quanto i dati vengono trasformati per adattarsi allo scopo delle analisi da eseguire. Vengono qui effettuate operazioni di campionamento, filtraggio, aggregazione, eliminazione degli attributi poco significativi e la creazione di nuovi in base all'analisi da effettuare. Le operazioni di trasformazione dei dati sono il nucleo attorno al quale si sviluppano le analisi, in quanto ad ogni obiettivo prefissato corrisponde un differente lavoro di data transformation. Al termine dell'attività, il risultato è un dataset studiato e pronto per l'analisi.
- **Data mining** Le operazioni di data mining hanno lo scopo di estrarre informazioni utili da una grande quantità di dati. Queste informazioni sono spesso implicite e nascoste nei dati e non osservabili attraverso una semplice analisi descrittiva, ma necessitano di metodi matematici e statistici in grado di mettere in relazione tra loro le variabili e scoprire dei pattern significativi. Tra le principali tecniche di data mining troviamo il *clustering*, *alberi di decisione*, *classificazione*, *regressione*, *matrici di correlazione* e *regole di associazione*. Per la natura dei dati a disposizione è stato possibile eseguire solo parte di queste analisi, o meglio, solo alcune hanno permesso di estrapolare pattern significativi.
- **Interpretazione** Eseguite le operazioni di data mining occorre saper interpretare i risultati delle analisi e visualizzarli per evidenziarne gli

aspetti critici e più significativi. L'interpretazione deve essere guidata dall'esperto di dominio il quale ha una visione chiara e definita delle dinamiche interne al business ed è in grado di assegnare la giusta importanza ad ogni aspetto dell'analisi.

Il prossimo capitolo si concentrerà sugli ultimi due step del KDP, presentando i risultati delle analisi effettuate, la motivazione alla base delle analisi e alcune considerazioni in merito agli aspetti più significativi. Dove necessario saranno approfonditi i modelli di data mining utilizzati per comprenderne il funzionamento e la natura dello studio.

Capitolo 4

Data Analysis

4.1 Premesse

Per lo svolgersi delle analisi sono stati considerati quasi esclusivamente i dati relativi all'area A.P.D. invernale della scuola di riferimento in quanto più completi e meglio gestiti. Grazie al lavoro di data integration e data cleaning è stato possibile svolgere parte delle analisi anche sui dati di Scuola Sci e A.P.D. estivo, utili a comprendere il diverso comportamento delle aree. I dati dell'A.P.D. invernale comprendono le attività delle ultime 7 stagioni e sono frutto dell'integrazione dei dati raccolti da due diversi software gestionali. Si sottolinea inoltre che, prendendo in considerazione solo una porzione delle attività della scuola, i risultati delle analisi non siano rappresentativi dell'andamento del business della stessa, ma limitatamente ad un'area di essa.

La quasi totalità delle analisi qui presentate appartiene alla categoria dell'analisi descrittiva; questo è dovuto alla natura dei dati in possesso della scuola i quali contengono informazioni relative esclusivamente agli aspetti operativi delle

attività. I dati riguardanti le operazioni commerciali, le offerte, gli eventi e qualsiasi attività esterna all'effettivo svolgersi di lezioni e corsi non sono attualmente a nostra disposizione e quindi la maggior parte delle possibili analisi predittive perde di interesse. Tuttavia, è stato possibile effettuare analisi diversificate, a diversi livelli di dettaglio e sono emerse considerazioni particolarmente interessanti dal momento che questi dati non erano stati esaminati precedentemente.

Tenendo presente che i dati comprendono informazioni sensibili dei soci e di primaria importanza per la scuola, i risultati delle analisi sono qui riportati in termini relativi per assicurare la riservatezza degli stessi, permettendo comunque di effettuare interessanti considerazioni di carattere strategico.

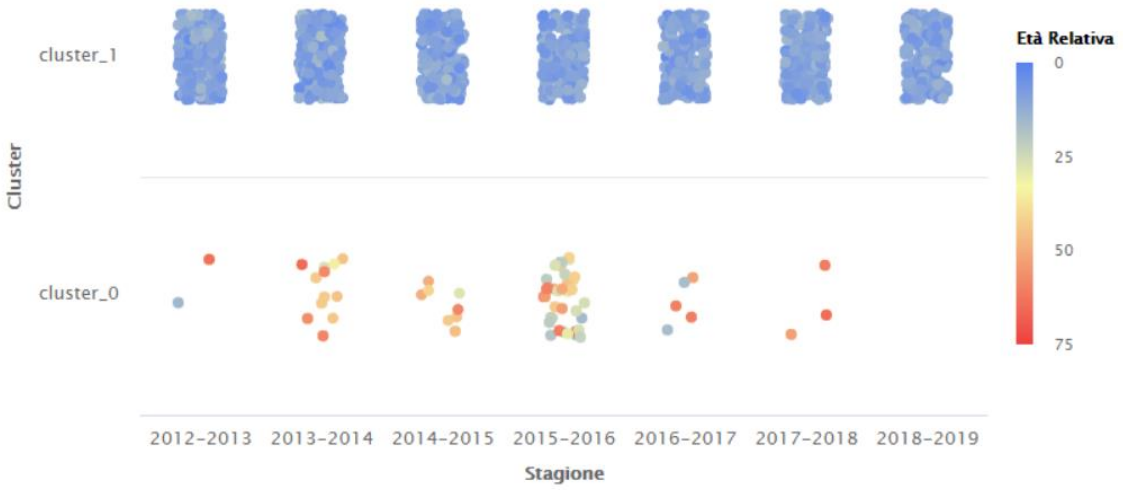
Si sottolinea come le osservazioni tratte dai risultati siano esclusivamente personali e frutto dell'attività accademica e di ricerca eseguita per questo studio; non emergeranno eventuali conclusioni dettate dagli esperti di dominio, i quali hanno comunque guidato il processo di analisi, ma ogni osservazione è da intendersi come spunto di carattere personale.

Per lo svolgimento delle analisi si è lavorato principalmente su Excel e RapidMiner (piattaforma dedicata alla data science già utilizzata in ambito accademico) con l'ausilio di Python per le analisi più complesse e specifiche del dominio. Le operazioni relative alla gestione dei dati sono state eseguite attraverso Excel e Power BI mentre per le interazioni con il database sono stati utilizzati SQL Server, MySQL e HeidiSQL.

4.2 Anomaly detection

Prima di iniziare con le analisi è importante effettuare un'operazione preliminare di individuazione delle anomalie dei dati. In statistica, un'anomalia consiste in un valore molto distante da quello della normalità della popolazione, in generale al di fuori dell'intervallo tra il primo e il terzo quartile, e viene indicato con il termine di *outlier*. L'individuazione degli outlier consente di capire quanti e quali elementi potrebbero inquinare le analisi, essendo non rappresentativi della popolazione in quanto significativamente differenti, permettendo quindi di escluderli dalle osservazioni o considerarli per solo una parte di esse. Nel nostro caso, gli elementi della popolazione sono i soci iscritti ai corsi dell'area A.P.D. invernale e tra i principali attributi di interesse troviamo l'età del socio relativa ad una specifica stagione, il settore sportivo di appartenenza e la categoria del settore.

Per l'individuazione degli outlier è stato utilizzato l'algoritmo di clustering DBSCAN (Density-Based Spatial Clustering of Application with Noise) la cui idea di base è quella di racchiudere gli elementi più simili tra loro in cluster distinti. L'algoritmo necessita di due parametri fondamentali: il numero minimo di elementi necessari a formare un nuovo cluster (*min_points*) e la dimensione del vicinato (ε), ovvero la distanza da un elemento entro la quale devono ricadere almeno *min_points* elementi per poter formare un cluster. Entrambi i parametri ottimali sono stati trovati attraverso l'operatore *Data to Similarity* di RapidMiner il quale calcola, per ogni elemento del dataset, la *similarity* con ogni altro elemento della popolazione. Con similarity ci si riferisce a una misura numerica di quanto due elementi siano simili tra loro.



[Figura 17] Risultato del DBSCAN sugli iscritti ai corsi stagionali.
Il cluster_0 contiene gli elementi etichettati come outlier.

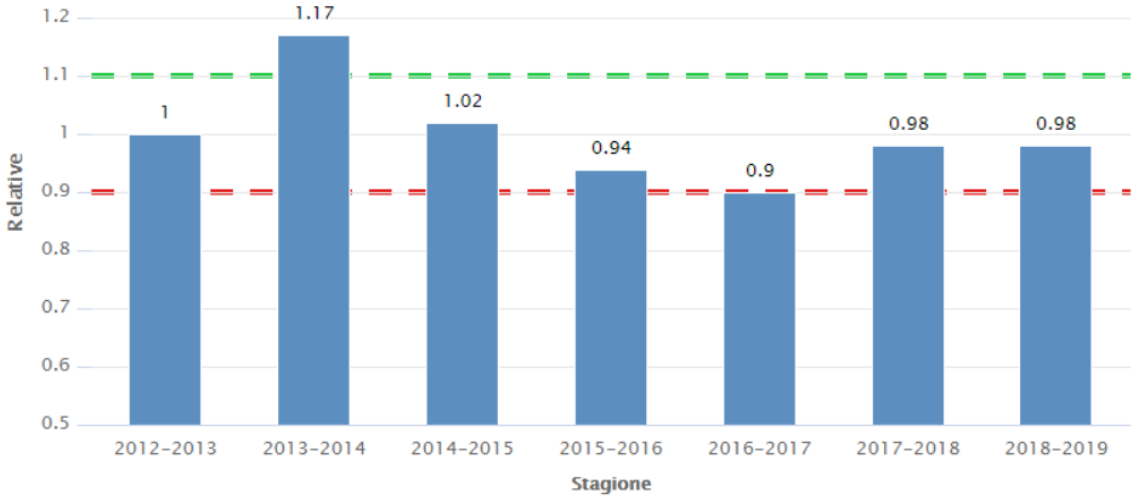
Il risultato dell'algoritmo DBSCAN è riportato nel grafico a dispersione in [Figura 17], nella quale ogni elemento del dataset è rappresentato da un punto. Il cluster_0 è quello relativo al rumore, ovvero a quei punti che non hanno trovato posto negli altri cluster e sono etichettati quindi come outlier. Al fine di comprendere la natura di questi elementi, il diagramma è stato suddiviso in stagioni e per ogni iscritto è stata evidenziata l'età relativa alla stagione di riferimento.

Notiamo quindi come nel cluster_0 ci siano quasi esclusivamente adulti (over 20 anni) e come questi rappresentino una minima parte della popolazione, inferiore al 5%, oltre ad essere totalmente assenti nel cluster_1; questo vuol dire che l'algoritmo ha rilevato essere determinante l'età del socio per l'individuazione delle anomalie. Essendo gli outlier concentrati principalmente nelle stagioni 2013-2014 e 2015-2016, mentre addirittura assenti nel 2018-2019, si può pensare a un cambio di organizzazione interna alla scuola nel corso delle stagioni per quanto riguarda gli adulti. Si è quindi scelto di escludere il cluster_0 da alcune delle analisi in quanto

potrebbe essere controproducente tenere in considerazione gli adulti, presenti solo in alcune stagioni, dal momento che la scuola stessa si è organizzata in maniera differente.

4.3 Analisi temporali

La prima necessità della scuola è quella di capire, in termini quantitativi, l'andamento del proprio business con il passare delle stagioni. La [Figura 18] mostra il numero dei singoli iscritti ai corsi stagionali dell'A.P.D. invernale negli ultimi 7 anni. Si sottolinea come il grafico faccia riferimento al socio e non all'iscrizione: esiste quindi la possibilità che uno stesso socio abbia acquistato più di un abbonamento nella stessa stagione, ma il diagramma non ne tiene conto.

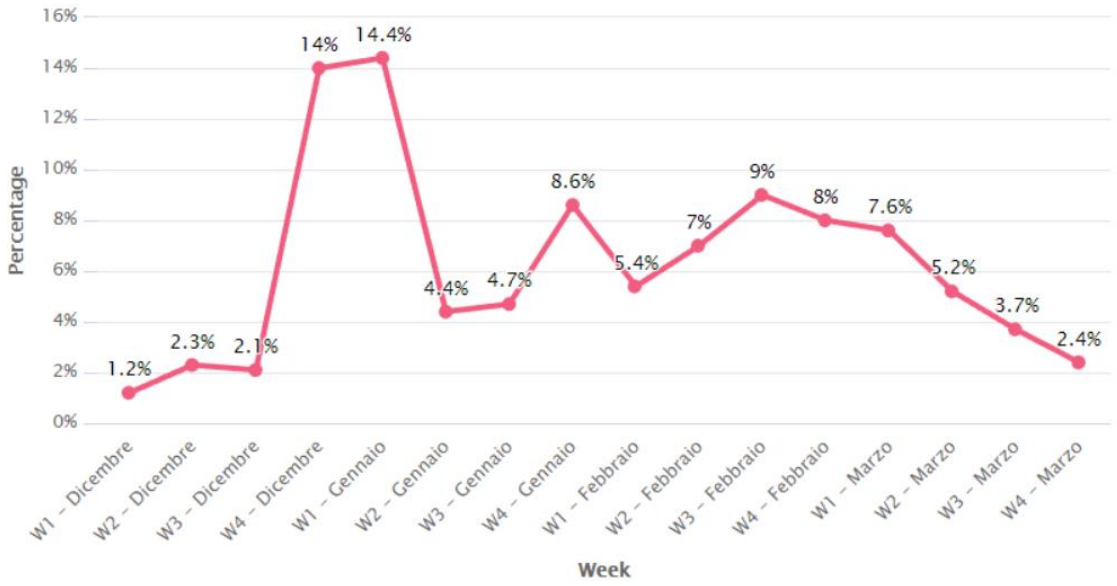


[Figura 18] Trend iscritti per stagione

I dati, come anticipato in 4.1, sono riportati in valori relativi rispetto alla media alla quale viene associato il valore 1. Il grafico mostra un trend altalenante, in leggera risalita nelle ultime stagioni ma comunque sotto la media, che viene alzata

dai valori delle prime 3 stagioni. I dati che attirano immediatamente l'attenzione sono però quelli relativi alla stagione 2013-2014 (+17% sulla media) e 2016-2017 (-10%). Per le attività della scuola sarebbe fondamentale comprendere quali eventi abbiano influito sui risultati delle due stagioni per capire cosa ripetere e cosa invece evitare. Al fine di approfondire quali potrebbero essere le dinamiche legate all'andamento del business nelle stagioni sopra citate, torneremo ad effettuare le analisi per stagione nei prossimi paragrafi, incrociando i dati con altre caratteristiche dei soci iscritti.

Passiamo ora ad analizzare l'attività della scuola a diversi livelli di dettaglio: oltre alle considerazioni per stagione è infatti possibile esaminare i dati temporali su base mensile, giornaliera e oraria. Analizzando i dati relativi alle lezioni tenute dagli istruttori dell'area Scuola Sci riusciamo a capire quali siano i mesi e giorni dell'anno con i maggiori carichi di lavoro **Error! Reference source not found.**].

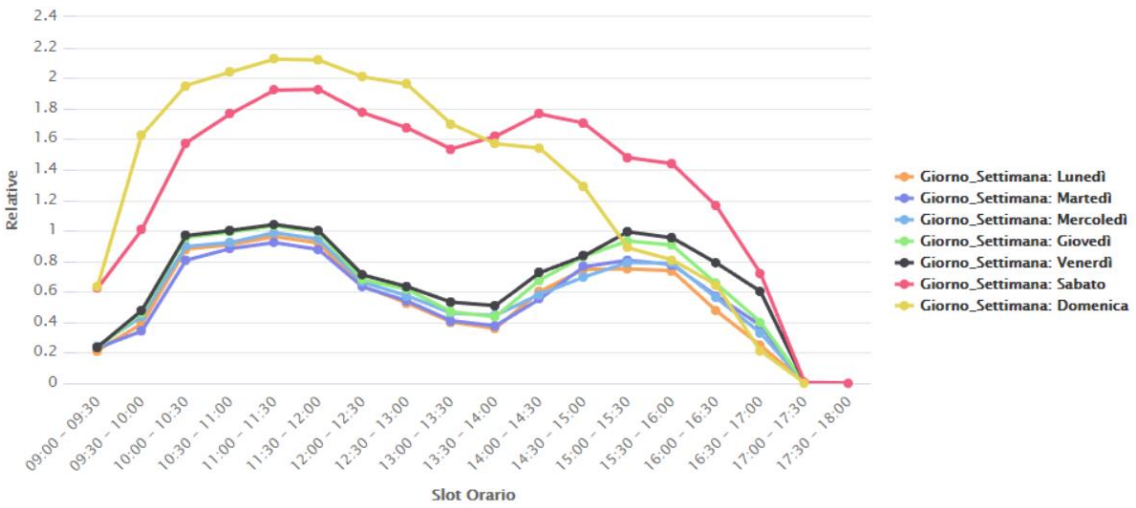


[Figura 19] Distribuzione delle lezioni nell'arco della stagione invernale

Il grafico mostra, in termini percentuali, la distribuzione del numero di lezioni nell'arco della stagione invernale, da dicembre a marzo. Per semplicità abbiamo aggregato i dati settimanali, al fine di comprendere l'andamento generale del carico di lavoro, senza entrare nel dettaglio dei singoli giorni; si ritiene che le analisi potrebbero essere altrimenti inquinate dalla ciclicità della domanda che, come visto nel paragrafo 2.2, varia a seconda che il giorno sia feriale o meno.

Si nota quindi come i periodi di punta siano in corrispondenza delle festività natalizie (ultima settimana di dicembre e prima di gennaio) e come il livello delle attività diminuisca gradualmente nel mese di marzo, mentre si trovano valori piuttosto elevati nella quarta settimana di gennaio e nella terza di febbraio.

Aumentando ulteriormente il livello di dettaglio è possibile esaminare la distribuzione delle lezioni nell'arco della settimana e della giornata [Figura 20].



[Figura 20] Distribuzione delle lezioni durante la giornata

Come anticipato in [Figura 5] del paragrafo 2.2, la distribuzione delle lezioni è strettamente legata al giorno della settimana e l'andamento della curva varia tra i giorni feriali e il weekend. Guardando il grafico emergono considerazioni

interessanti sulle abitudini degli iscritti durante la giornata, come il calo di richieste in corrispondenza dell'ora di pranzo, particolarmente visibile durante i giorni feriali, e il diverso comportamento nel pomeriggio di sabato e domenica. Dalle ore 15:00 della domenica infatti la frequenza delle lezioni subisce un crollo che non viene riscontrato nel pomeriggio del sabato, a causa degli impegni lavorativi dei genitori dei bambini iscritti alla scuola, e porta la curva (in giallo) ad essere addirittura superata da quelle di giovedì e venerdì, cosa che non avviene per la giornata di sabato.

Basandosi sulle osservazioni tratte da [Figura 5], **Error! Reference source not found.**] e [Figura 20] l'esperto di dominio è in grado, grazie alla propria esperienza, di valutare se una differente gestione di corsi e offerte possa giovare o meno a livello di affluenza. Chiaramente, le informazioni ottenute in seguito a questo tipo di analisi sono già note ai membri dello staff che vivono la realtà della scuola quotidianamente, ma basare le proprie valutazioni sui numeri invece di affidarsi alla percezione degli eventi rende più precisi ed efficaci gli interventi correttivi.

Come anticipato nel paragrafo 2.2, il meteo rappresenta un fattore determinante per il business della scuola che, vendendo divertimento all'aria aperta, risente particolarmente delle condizioni meteorologiche sfavorevoli, come confermato dagli stessi soci. Abbiamo quindi deciso di incrociare i dati provenienti dal tabellone invernale della Scuola Sci con i dati del bollettino meteo dell'Arpa Piemonte [16] con il fine di quantificare l'effetto del meteo sul numero di lezioni tenute dalla scuola.

L'obiettivo è stato raggiunto costruendo una matrice di correlazione. L'algoritmo prevede di calcolare, per ogni coppia di attributi del dataset, l'indice di correlazione tra le variabili. Con indice di correlazione si intende, in statistica, un coefficiente il cui valore è compreso nell'intervallo chiuso [-1,1] e si calcola come segue:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

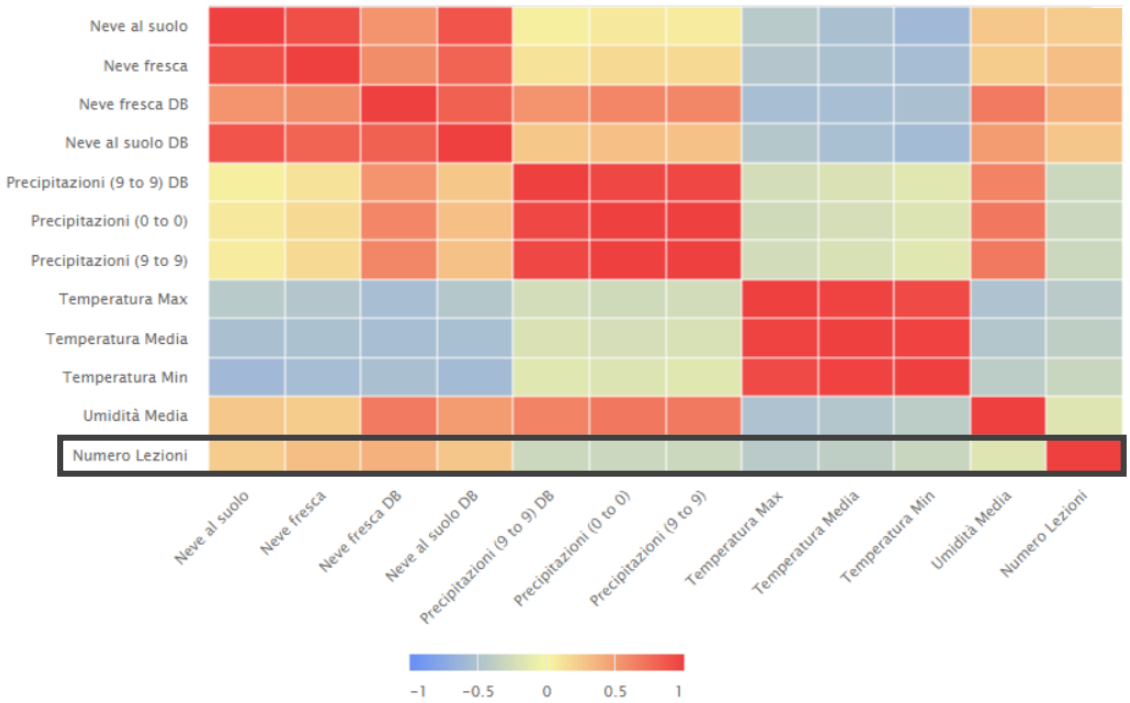
dove X e Y sono le variabili di interesse, σ_X e σ_Y le deviazioni standard di X e Y, e σ_{XY} la covarianza tra le variabili.

Un coefficiente positivo indica una diretta correlazione tra gli attributi, mentre un coefficiente negativo ne indica una correlazione inversa; un valore nullo dell'indice indica l'indipendenza tra le variabili. In generale, si dice che la correlazione è debole quando il valore assoluto del coefficiente è inferiore di 0,3 mentre un coefficiente superiore a 0,7 indica una forte correlazione.

Per la nostra analisi abbiamo aggregato i dati meteo per stagione, filtrando i mesi di maggior interesse, e calcolato la media dei parametri a disposizione. Per ogni stagione è stato quindi calcolato il numero delle lezioni tenute dalla scuola ed è stata generata la matrice di correlazione [Figura 21].

La matrice presenta quindi la correlazione tra ogni variabile, presente sia sull'asse delle ascisse che su quello delle ordinate, ed è speculare rispetto alla sua diagonale che è formata da valori unitari, in quanto la correlazione di ogni variabile verso sé stessa è sempre 1. Si notano forti correlazioni, positive o negative, in alcune zone della mappa in quanto vengono legate tra di loro variabili meteo strettamente correlate come le varie temperature o l'altezza del manto nevoso calcolati sotto

diversi punti di vista. La riga di nostro interesse è l'ultima, evidenziata in nero, facente riferimento alla variabile denominata "Numero Lezioni".



[Figura 21] Matrice di correlazione sui dati meteo

I parametri presi in considerazione sono principalmente quelli relativi alle precipitazioni, sia piovose che nevose, e alla temperatura; tutti i parametri sono calcolati come media giornaliera per la stagione di riferimento e il campione utilizzato comprende i dati di 10 anni di rilevamenti.

Si può notare come i primi quattro fattori, legati alle condizioni del manto nevoso, siano correlati positivamente con il numero delle lezioni svolte; l'indice assume il suo massimo valore in corrispondenza degli attributi relativi al livello della neve fresca (seppur rimanga moderato con un valore di 0,4) e in particolare in corrispondenza del parametro "Neve fresca DB" dove la sigla sta per *day before*.

Il dato ci dice quindi che un alto livello di neve fresca nel giorno precedente induce il cliente a sciare nella giornata successiva e che, quindi, si svolgano più lezioni (in particolare individuali).

Al contrario, i valori relativi alle precipitazioni e in particolare alla temperatura (tra -0,25 e -0,32) sono negativi e indicano una debole correlazione inversa: all'aumentare di precipitazioni e temperatura diminuiscono le lezioni.

L'analisi è stata ripetuta con un diverso livello di dettaglio aggregando i dati giornalieri e mensili, ma se da un lato ci si aspettava una migliore osservazione della correlazione tra le variabili meteorologiche con il numero di lezioni tenute quotidianamente, dall'altro l'influenza di fattori esterni (ciclicità della domanda, differente distribuzione tra weekend e giorni feriali) ha reso il fenomeno molto meno evidente; pur confermando il segno dei coefficienti, il valore assoluto dell'indice non ha superato lo 0,2 per nessun parametro.

Il risultato dell'analisi ha quindi confermato quella che era la percezione dello staff attribuendo una certa importanza al fattore meteorologico nelle attività della scuola. La correlazione tra le variabili è comunque moderatamente bassa, come ci si aspettava, in quanto gli aspetti più rilevanti e influenti sulla domanda sono altri e variano dalle offerte della scuola all'attività della concorrenza, ma si considerano in egual modo soddisfacenti i risultati ottenuti.

4.4 Analisi sui soci

Se le analisi temporali forniscono informazioni riguardanti il trend della scuola e sulle abitudini degli iscritti ai corsi, le analisi sui soci offrono utili indicazioni sulla

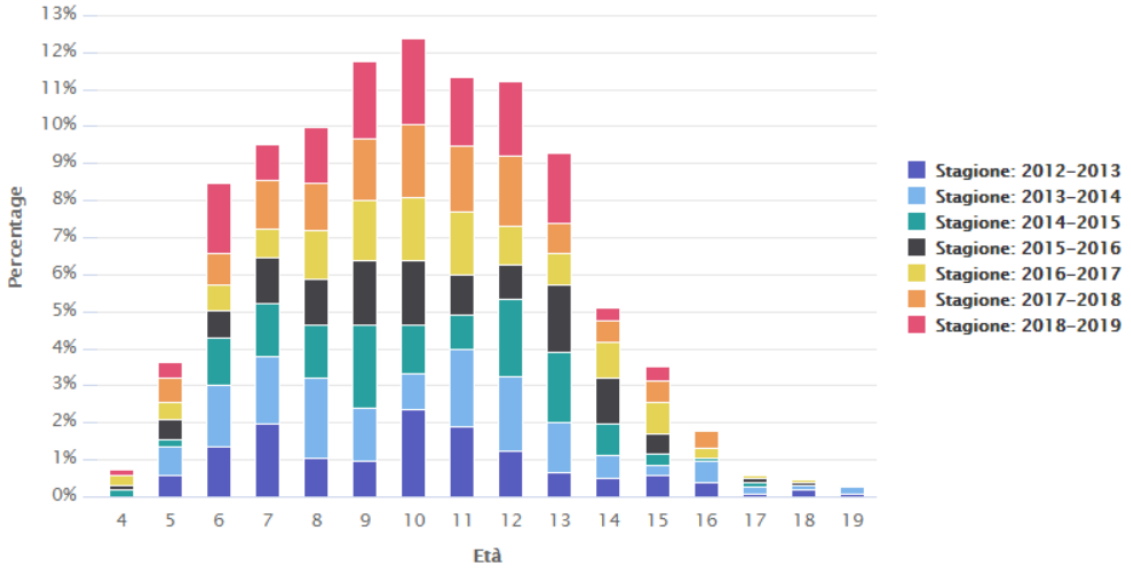
loro natura. Le informazioni riguardanti età e sesso dei frequentatori della scuola possono consentire una migliore gestione delle offerte da parte del management, mentre la provenienza dei soci è di fondamentale importanza in ottica data-driven marketing. Un altro parametro non trascurabile è quello relativo alla fedeltà dei clienti, intesa in termini di ritorni o rinnovi, che permetterà di effettuare analisi più approfondite legate alla storia dei soci in un prossimo paragrafo.

La prima analisi effettuata riguarda l'età degli iscritti. Come detto in 4.2, in ottica anomaly detection, per alcune analisi è stato scelto di non includere i soci di maggiore età dal momento che la scuola è diversamente organizzata in tal senso e, oltre a rappresentare meno del 5% della popolazione, i soci adulti presentano comportamenti differenti rispetto ai ragazzi in termini di corsi e sottoscrizioni ai servizi, andando quindi ad intaccare la bontà delle analisi.

La curva raffigurata in [Figura 22] è stata generata aggregando i dati anagrafici riferiti a 7 stagioni dell'area A.P.D. invernale e mostra la distribuzione degli iscritti, in valori percentuali sulla popolazione totale, in termini di età.

La figura non è perfettamente rappresentativa della popolazione effettiva in quanto i dati anagrafici in nostro possesso non sono completi per tutti i soci, ma rende comunque l'idea della distribuzione reale.

Come ci si potrebbe aspettare da una scuola sci, i primi clienti sono i ragazzi ed in particolare i bambini tra i 6 e i 13 anni che compongono oltre l'80% del totale dei soci; la distribuzione segue un andamento normale e l'età media della popolazione in esame è di circa 10 anni. Lo studio di questo grafico è importante per iniziare a comprendere non solo quale sia la natura dei propri soci (e di conseguenza il target



[Figura 22] Distribuzione dei soci per età e stagione

dell'azienda) ma anche quali siano stati i risultati della gestione nelle stagioni passate, per ogni fascia di età. Suddividere il grafico per stagioni consente di fare le prime osservazioni riguardo quali aree della scuola abbiano ottenuto i risultati migliori, in quali stagioni e in quale misura gli effetti degli eventi si siano propagati negli anni.

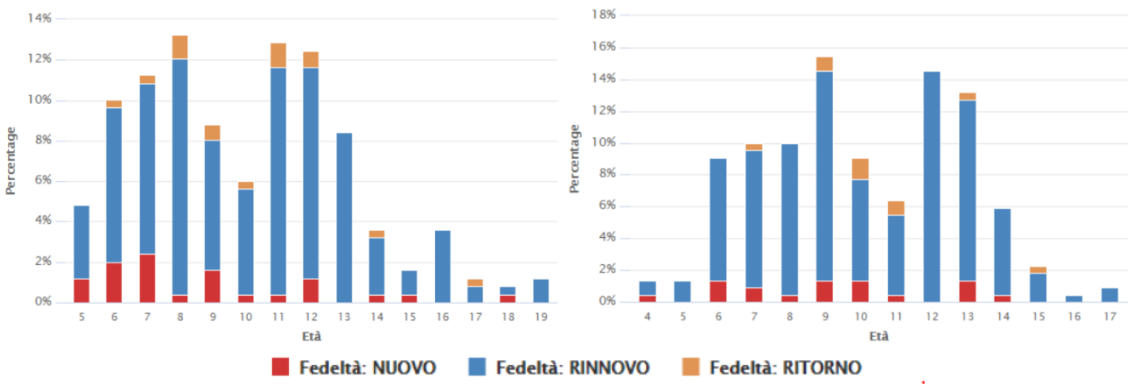
Osservando la curva corrispondente alla stagione 2012-2013, per esempio, si nota come i valori relativi alle fasce d'età degli 8 e 9 anni si discostino notevolmente dal trend della distribuzione. Non essendoci dati su cui basare delle ipotesi risulta complicato capire quale sia la causa del fenomeno che può riguardare diversi fattori. Ciò che si può fare è osservare come questo evento abbia avuto delle ripercussioni sulle stagioni seguenti; guardando la curva 2013-2014 si osserva come il gap si sia spostato sulla fascia dei 9 e 10 anni, e nella stagione successiva in particolar modo sugli quella degli 11. Questo è dovuto al fatto che una gran parte dei soci iscritti alla scuola, come vedremo successivamente, sia solita rinnovare la propria iscrizione

nella stagione successiva: se in una determinata stagione si è lavorato particolarmente bene o male su una specifica fascia d'età, è probabile che l'effetto di quanto prodotto si propaghi nella stagione successiva.

Esiste la possibilità quindi che osservando la stagione 2011-2012 il fenomeno sia evidente tra i 7 e gli 8 anni e che in generale la gestione di quella particolare fascia di età non sia stata ottimale in tali stagioni.

Per comprendere l'importanza dei rinnovi mostriamo a titolo di esempio le stagioni 2013-2014 e 2014-2015, suddividendo i soci per fascia d'età e fedeltà.

La fedeltà del cliente viene qui misurata in base al fatto che, nella stagione di riferimento, l'iscrizione del socio sia relativa a una nuova iscrizione, un rinnovo rispetto all'anno precedente oppure un ritorno dopo alcuni anni di assenza dalla scuola. La fedeltà è calcolata prendendo in considerazione anche le eventuali iscrizioni del socio alle altre aree della scuola, compreso l'estivo. Dal momento che la scuola non tiene conto dello stato delle iscrizioni, la fedeltà è stata calcolata con un semplice algoritmo prendendo in considerazione esclusivamente il nominativo del socio che è però soggetto ad errori di input da parte dello staff.



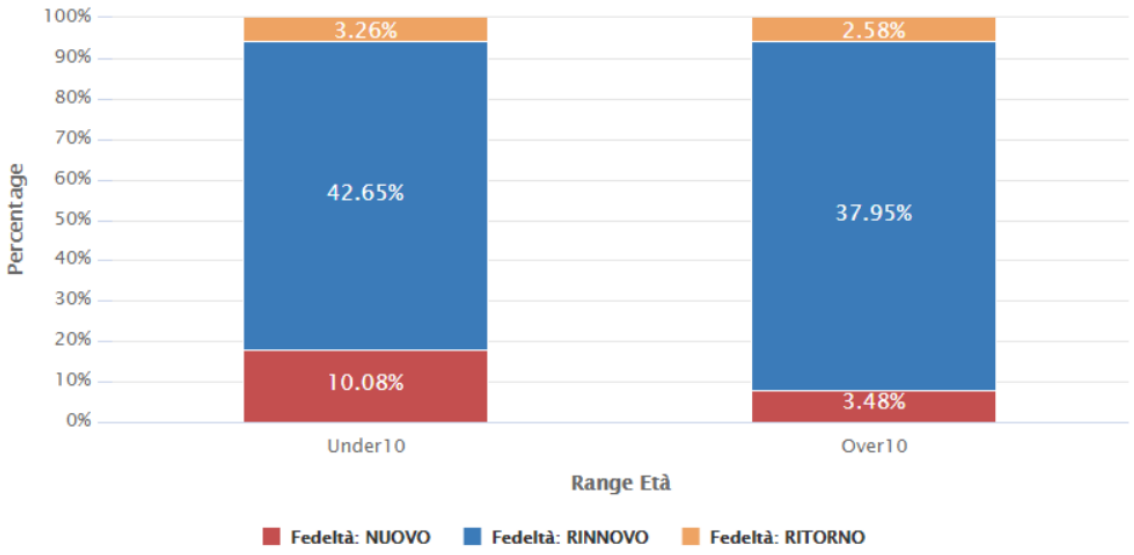
[Figura 23] Distribuzione per età e fedeltà nelle stagioni 2013-2014 e 2014-2015

Siamo comunque convinti che questi non incidano in larga misura su quella che è l'idea generale dell'analisi e sulle osservazioni tratte dai risultati.

Dai grafici in [Figura 23] si nota come a prescindere dalla stagione o dalla fascia di età, la percentuale di iscrizioni relative ad un rinnovo sia di gran lunga superiore a quella dei nuovi iscritti; quasi l'80% dei soci annuali ha infatti frequentato la scuola negli anni precedenti.

La figura mette in evidenza come un gap negativo in determinate fasce d'età si confermi nella stagione successiva, così come un picco positivo: si guardi ad esempio la fascia di età dei 9 anni nel 2013-2014 e quella dei 10 anni nella stagione dopo. Questa osservazione, unita al fatto che il numero dei soci tende a decrescere rapidamente da una certa età in poi, in particolar modo per quanto riguarda i nuovi iscritti, porta a pensare che il segmento target di mercato a cui si dovrebbe rivolgere una generica scuola sci sia quello degli Under-10.

Attirare e fidelizzare una clientela giovane, sin dai primi anni di età, sembrerebbe portare a buoni risultati non solo su questa fascia d'età ma anche in quelle successive grazie ai rinnovi degli iscritti. Discretizzando il range di età e suddividendo i soci in Under-10 (dai 10 anni compresi in giù) e Over-10 il fenomeno è infatti apparso più evidente: accorpando i dati delle stagioni a nostra disposizione si nota come tra gli Over-10 i record relativi a una nuova iscrizione rappresentino meno del 10% del totale di categoria, percentuale che aumenta a oltre il 18% per quanto riguarda i soci più piccoli. Il divario aumenta ulteriormente guardando la percentuale sul dato complessivo dove i nuovi iscritti Over-10 rappresentano solo un quarto del totale [Figura 24].



[Figura 24] Fedeltà dei soci per range di età

I dati ci dicono quindi che l'attenzione della scuola debba essere rivolta ai più piccoli per quanto riguarda l'attività di marketing, perchè è più semplice attirare nuovi clienti in quelle fasce d'età e sono propensi a rinnovare una volta cresciuti all'interno della scuola.

Non è quindi un caso che, come suggerito dal report di JFC Tourism & Management approfondito nel paragrafo 2.3, le prime scuole sci abbiano iniziato ad occuparsi dei bambini più piccoli creando i cosiddetti "kinder park", con il duplice scopo di attirare i genitori sciatori e fidelizzare i figli già in tenera età.

Per raggiungere l'obiettivo di fidelizzare una clientela molto giovane è però opportuno organizzare ogni aspetto della scuola in quest'ottica, in quanto le esigenze sono diverse e più specifiche. Lezioni, offerte ed eventi devono essere organizzati su misura, così come i servizi, la sicurezza, l'interazione con i clienti, il marketing. Attrarre questo genere di clientela vuol dire fare leva sulle esigenze dei genitori o sulla curiosità del bambino, o dove possibile entrambe.

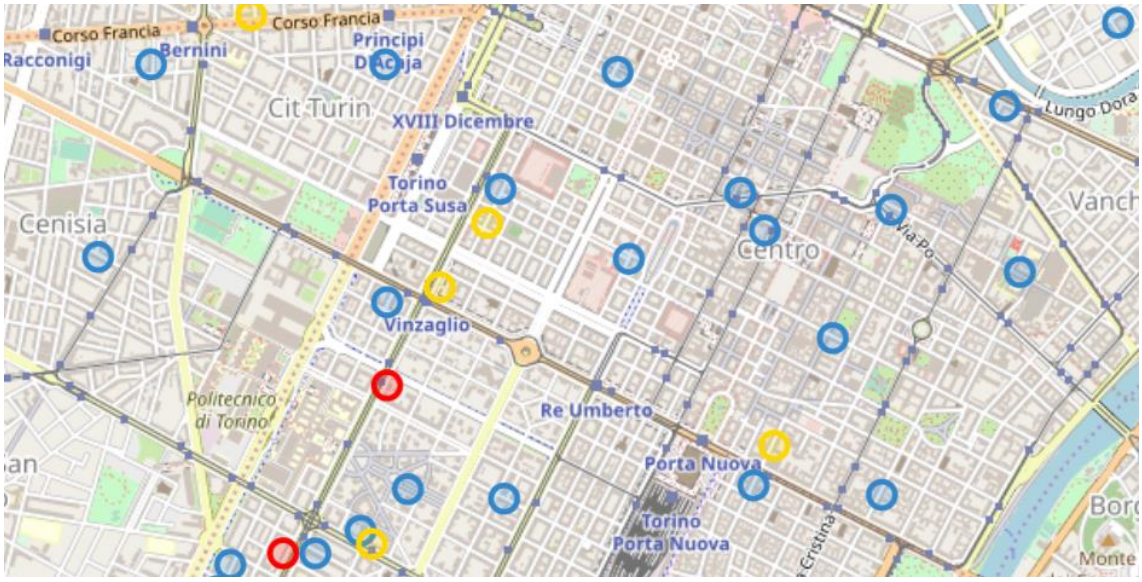
Se per la prima strada la soluzione consiste nell'offrire servizi che accontentino i genitori, per incuriosire un bambino bisogna prima sapere dove trovarlo.

A questo scopo è stata effettuata un'analisi geografica della distribuzione dei soci sul territorio per comprenderne la provenienza. In un primo momento la geolocalizzazione è stata eseguita, attraverso l'utilizzo di una heatmap, a livello comunale su tutto il territorio del nord ovest Italia, constatando un forte presidio sui comuni di Milano e Torino a fronte di una presenza più sporadica sul resto del territorio principalmente piemontese.

L'analisi è stata quindi ripetuta con un diverso livello di dettaglio focalizzandosi esclusivamente sulla distribuzione nel comune di Torino.

Per lo svolgersi dell'analisi sono state utilizzate le librerie Pandas, Geopy e Folium di Python, con il supporto delle API di Google Maps per la geolocalizzazione.

La [Figura 25] mostra una porzione del comune di Torino evidenziando la posizione dei soci. Il colore del segnalino anticipa il concetto di famiglia che



[Figura 25] Esempio di distribuzione dei soci nel comune di Torino

approfondiremo nel prossimo paragrafo, specificandone la numerosità: il colore rosso indica una famiglia di almeno 3 persone, il colore giallo una famiglia di 2 persone mentre il colore blu indica un singolo socio.

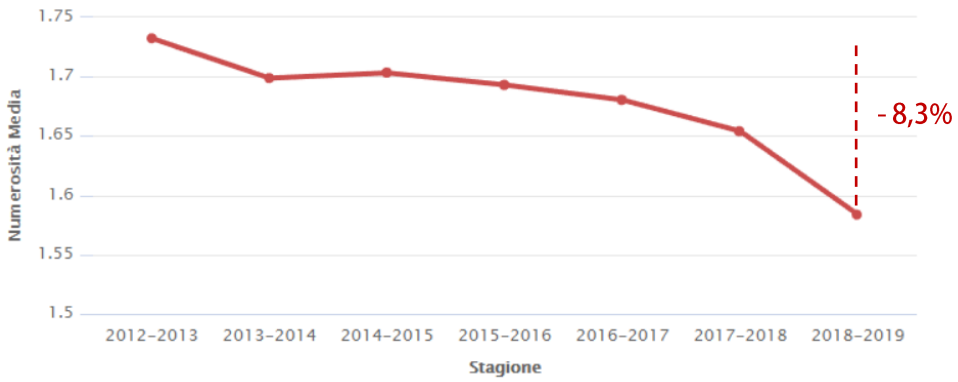
La porzione di mappa visibile è stata ristretta al fine di preservare i dati della scuola che dalla provenienza dei propri soci può ottenere utili informazioni riguardo la natura e le abitudini degli iscritti e quindi trarre conclusioni di carattere strategico. La geolocalizzazione dei clienti rappresenta un aspetto chiave del data-driven marketing poiché permette di capire quali siano le principali aree di presidio della scuola e in quali sia invece possibile catturare nuovi soci. Inoltre, incrociando i risultati dell'analisi con i dati relativi al reddito medio per quartiere di Torino è possibile ottimizzare la ricerca dei luoghi in cui una pubblicità studiata potrebbe risultare più efficace. Com'è noto infatti, praticare l'attività sciistica con continuità richiede un certo sforzo economico per le famiglie ed è quindi chiaro come concentrare i propri sforzi in ottica marketing su quelle che possono essere le aree più ricche della città possa portare maggiori frutti.

Tenendo in considerazione il target della scuola, ovvero i bambini Under-10, si potrebbe pensare ad una pubblicità mirata nei luoghi principalmente frequentati da questo tipo di clientela: scuole (in particolar modo private), parchi, palazzetti dello sport, oratori.

4.5 Analisi sulle famiglie

Per approfondire il discorso relativo all'individuazione del target del business, introduciamo ora il concetto di famiglia. Intendiamo con famiglia un gruppo di

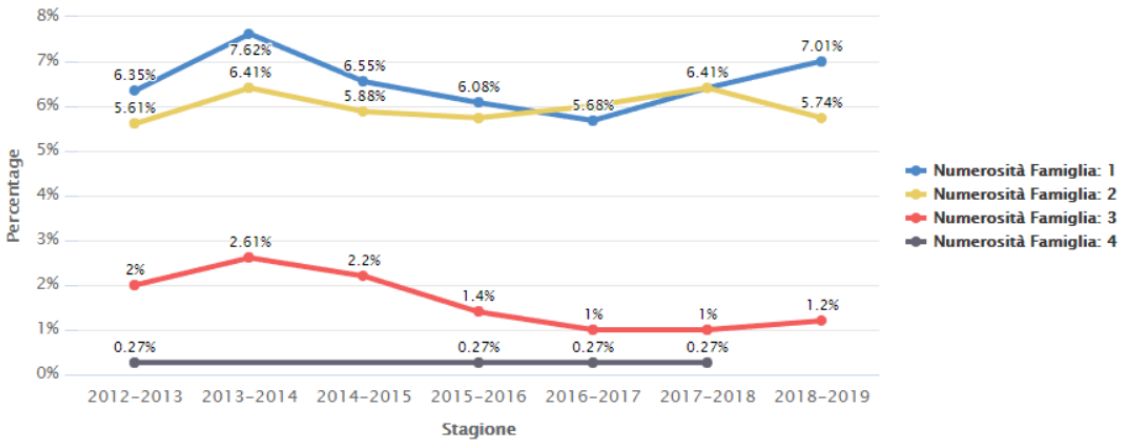
persone, idealmente parenti, iscritte insieme nella stessa stagione per godere dell'esperienza in compagnia. La scuola non ha tenuto traccia della famiglia di un socio se non assegnando alla famiglia il cognome della stessa che, però, in caso di omonimia risulterebbe allargata a più persone; viceversa un socio iscritto insieme ad un parente dal diverso cognome non viene considerato come una famiglia.
Non avendo altri dati su cui basare l'individuazione delle famiglie abbiamo quindi seguito lo stesso criterio.



[Figura 26] Trend numerosità media famiglie

In [Figura 26] è mostrato il trend della numerosità media delle famiglie durante le stagioni che risulta in costante decrescita. Ricordiamo che la fascia di età della popolazione è interamente compresa tra i 3 e i 20 anni, quindi tutte le famiglie di cui si parla sono composte da fratelli; una rapida analisi ha inoltre confermato come i soci esclusi in seguito all'anomaly detection non siano i genitori dei ragazzi oggetto dell'analisi (con rare eccezioni) ma singoli soci iscritti ai corsi per adulti.
Il grafico mostra un trend sicuramente non positivo soprattutto alla luce delle analisi che seguono. Esaminando il diagramma in [Figura 27], che raffigura in termini relativi il trend stagionale del numero di soci appartenente a una famiglia

di una certa numerosità, si percepisce subito quale importanza rivestano le famiglie nel business della scuola sci.



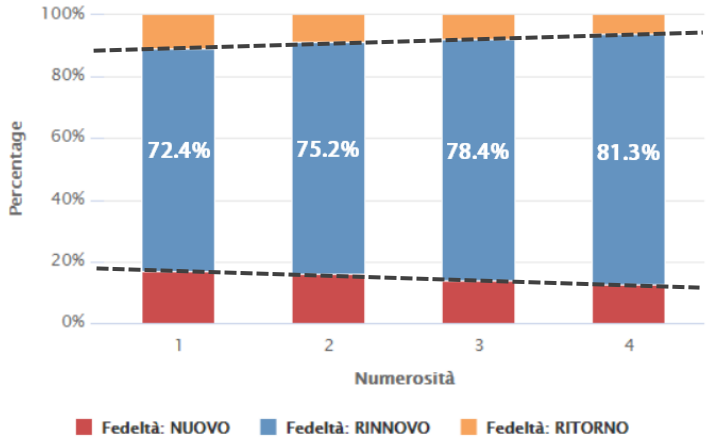
[Figura 27] Distribuzione numerosità famiglie per stagione

Il grafico mostra come la curva dei soci appartenenti ad una famiglia di due componenti sia molto vicina a quella dei soci singoli; se si considera anche la curva delle famiglie di numerosità 3 è chiaro come i soci che solitamente si iscrivono in compagnia di un parente siano in maggioranza rispetto a quelli che si iscrivono da soli. Questo risultato conferma nuovamente gli spunti tratti dal report di JFC che afferma come per la maggior parte dei turisti l'attività di montagna rappresenti un'occasione per coltivare la propria passione in compagnia di amici o famigliari.

Il calo della numerosità media di [Figura 26] è giustificato dal trend decrescente della curva rossa che corrisponde alle famiglie di 3 elementi e dalla scomparsa delle famiglie di numerosità pari a 4 nell'ultima stagione.

Le ragioni per le quali si vuole dare importanza al concetto di famiglia sono principalmente due. In primo luogo, una famiglia è una preziosa fonte di clienti.

Una volta catturato l'interesse di un componente della famiglia, offrire un accesso facilitato anche agli altri membri vuol dire aprire le porte a fratelli, genitori e potenziali amici. Inoltre, i soci facenti parte di una famiglia sono più propensi a rinnovare l'iscrizione negli anni a seguire [Figura 28].



[Figura 28] Percentuale rinnovi per numerosità famiglia

Il fenomeno evidenziato nel grafico, che mostra come i soci siano più propensi a rinnovare quanto più grande è la famiglia di cui fanno parte, conferma ancora una volta l'importanza dell'ambiente quando si parla di attività invernali. Infatti, oltre alla motivazione puramente commerciale, la seconda ragione per la quale è importante concentrare l'attenzione sulle famiglie risiede proprio nel fatto che un maggior numero di famiglie contribuisce a rendere l'atmosfera della scuola più accogliente, oltre ad attirare l'attenzione dei genitori di altri bambini. L'obiettivo è quello di elevare lo status della scuola sci agli occhi dei clienti, affinché possano considerarla non solo come passatempo invernale ma come una tappa fissa con cui divertirsi con i propri amici e famigliari. La fidelizzazione del cliente è un aspetto

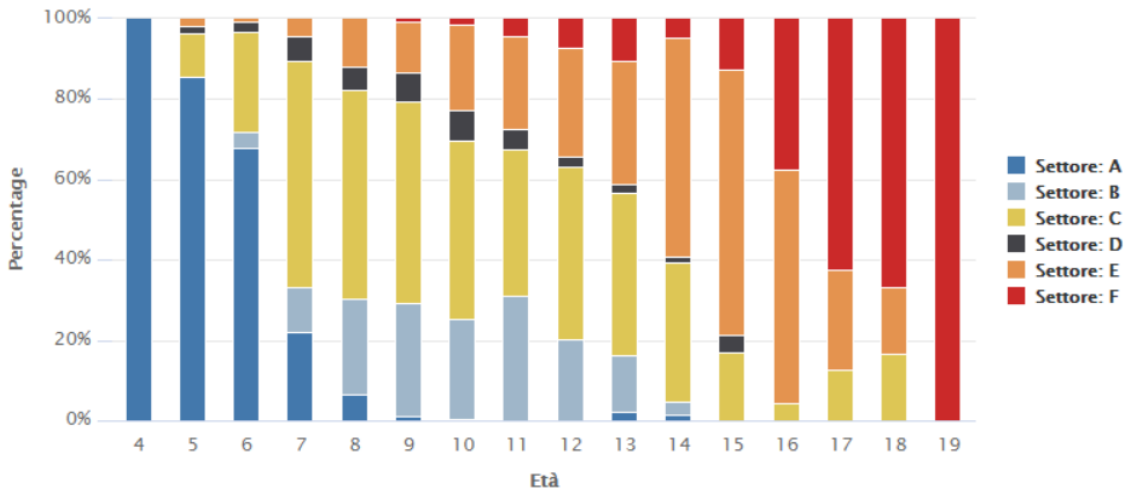
chiave del business della scuola sci e passa soprattutto dal concetto di famiglia e dall'attenzione per la customer experience.

4.6 Analisi sui settori

Completate le analisi sulla natura degli iscritti alla scuola, passiamo ora ad analizzare i servizi intesi sia come tipologia di corsi erogati (che suddividiamo in settori della scuola) sia come eventi particolari quali gare e manifestazioni.

Ricordiamo che i dati oggetto d'esame sono relativi all'area della scuola che si occupa degli abbonamenti stagionali; con "settore" ci riferiamo quindi alla famiglia di corsi e lezioni alla quale il socio sottoscrive l'abbonamento (esempi di settore possono essere "Snowboard" o "Sci di fondo"). Nel corso del paragrafo tralasceremo ogni analisi relativa alle performance dei settori nel corso delle stagioni, perchè poco interessanti ai fini dell'oggetto del lavoro, mentre ci soffermeremo su alcuni aspetti legati alle loro caratteristiche e relazioni.

In primo luogo cerchiamo di capire quali soci si abbonano a quali tipologie di corsi.



[Figura 29] Distribuzione dei settori per età

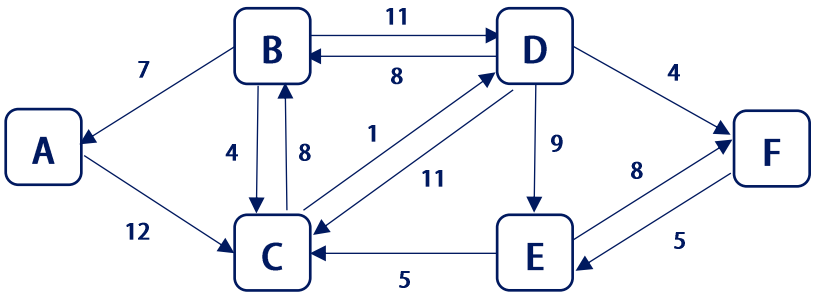
Come mostra la [Figura 29] ad ogni fascia d'età sono corrispondenti uno o più settori, che per semplicità abbiamo rinominato con le lettere dell'alfabeto.

Dal grafico si nota come ogni settore abbia un suo range di età, più o meno netto, e come questo non sia imposto dalle regole interne alla scuola. I soci scelgono infatti liberamente a quale famiglia di corsi iscriversi e lo fanno seguendo una certa logica che può dipendere dalle attività svolte, dalla difficoltà del corso o dal livello del socio stesso. Si delinea quindi un certo percorso naturale di attività che porta un socio, per esempio, a sottoscrivere un abbonamento al settore "E" dopo aver provato il settore "B" o "C". Staff e addetti della scuola sono ovviamente a conoscenza di questi aspetti, ma quantificare il fenomeno aiuta nel momento in cui si vuole studiare quali settori potrebbero influire maggiormente sul trend di ogni fascia di età e viceversa.

Organizzare la scuola rivolgendo particolare attenzione ai settori "E" ed "F" porterebbe a incrementare i propri soci oltre i 13 anni mentre, prendendo come esempio le osservazioni riguardanti i gap degli iscritti nel paragrafo 4.4, per attirare i bambini Under-10 bisognerebbe concentrare i propri sforzi sui settori "A", "B" e "C". Il grafico offre quindi un colpo d'occhio istantaneo che consente di confrontare il range di età previsto per ogni corso con i dati reali e conseguentemente intervenire con precisione dove ci si aspetta una diversa proporzione tra gli iscritti.

In seguito, è stata effettuata una rapida analisi riguardo la fedeltà dei soci abbonati ai vari settori, scoprendo come mediamente l'80% delle iscrizioni siano relative a rinnovi rispetto alla stagione precedente; un settore in particolare ha però

evidenziato una percentale superiore al 90% che ha destato il nostro interesse e ha dato spunto per un ulteriore approfondimento. Essendo in grado di ricavare quali iscrizioni appartengano alla categoria dei rinnovi e quali no, è possibile calcolare quella che definiamo come la "storia del socio", ovvero quali settori abbia frequentato ogni singolo cliente, dalla sua prima stagione nella scuola, in ordine cronologico. L'obiettivo ideale dell'analisi è quello di trovare i percorsi più gettonati dai soci, rappresentabili su un grafo orientato i cui vertici sono i settori della scuola, gli archi rappresentano i rinnovi da settore a settore mentre i pesi dipendono dal numero dei rinnovi [Figura 30].



[Figura 30] Esempio di grafo orientato rappresentativo dei rinnovi dei soci

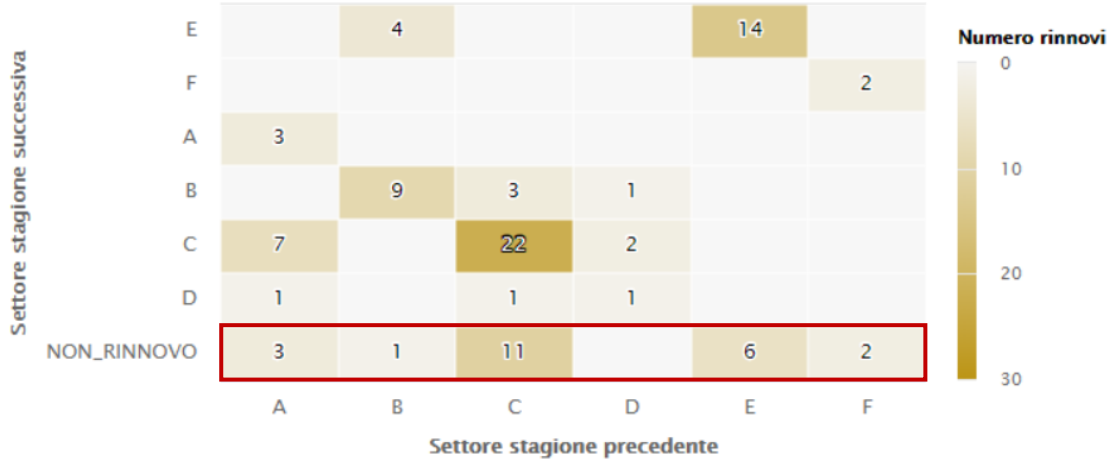
Quella che era l'idea originale dell'analisi è però venuta meno a causa della grande varietà di percorsi possibili e delle abitudini dei soci. E' infatti difficile tenere in considerazione eventuali interruzioni momentanee del rapporto tra cliente e scuola sci per determinate stagioni, o la semplice voglia di un socio di provare un settore diverso per poi tornare a quello precedente, senza intaccare la bontà dell'analisi. Gran parte dei soci, inoltre, rinnova l'abbonamento allo stesso settore per diversi anni di seguito e questo rende il grafico di difficile comprensione.

Se lo scopo era quindi quello di trovare percorsi tipici affrontati dai soci nel corso della loro storia, il modello non ha fornito risposte rilevanti ed è stato quindi modificato e adattato per un nuovo e diverso tipo analisi. Per capire meglio le motivazioni che stanno alla base di questa esigenza è opportuno fare una considerazione sulle analisi di performance di tipo descrittivo riguardanti i settori che, come detto all'inizio del paragrafo, sono state escluse da questo studio.

Dal momento in cui la scuola gestisce settori diversi, legati tra di loro da un percorso logico di attività che tiene conto del livello degli iscritti, tenere in considerazione l'andamento di un settore indipendentemente da quello degli altri non consente di valutarne correttamente le performance.

Il trend decrescente di una certa famiglia di corsi non si traduce necessariamente in un calo di quel settore ma, poichè i soci sono soliti rinnovare e passare da un'area della scuola ad un'altra, la causa è da cercarsi anche nel trend degli altri settori.

Se si considerasse, per esempio, un settore "A" i cui soci sono soliti rinnovare verso un settore "B" e in una determinata stagione il rendimento di "B" dovesse risultare in calo, la causa potrebbe essere una prestazione non positiva di "B" tanto quanto di "A". Se i soci frequentanti i corsi di "A" decidessero di interrompere il proprio rapporto con la scuola nella stagione successiva, a risentirne sarebbe infatti il settore "B" che conterebbe meno iscrizioni a causa dei soci persi. Per la prossima analisi l'attenzione è stata quindi spostata dalla storia del socio a quella del settore, con l'obiettivo di comprendere i movimenti degli iscritti tra un settore e l'altro al fine di quantificarne e valutarne le performance interne.



[Figura 31] Matrice dei rinnovi

In [Figura 31] viene mostrata la *matrice dei rinnovi* che mette in evidenza, per ogni settore, il numero di rinnovi verso un'altra area della scuola. I numeri in figura sono puramente simbolici e non rispecchiano il reale panorama dei settori della scuola, seppur rispettino i dati originali con le giuste proporzioni.

Dal grafico si possono notare immediatamente alcune relazioni tra le aree della scuola: i settori "C" ed "E" rinnovano principalmente verso loro stessi mentre i soci del settore "A" sono soliti passare a "C" nelle stagioni successive, così come buona parte di soci di "B" rinnova verso "E". Si osserva inoltre come nessun iscritto alla scuola rinnovi verso i settori "A" ed "F" se non chi già proviene da questi settori.

Rispetto al modello precedente è stata qui aggiunta la riga dei non rinnovi, evidenziata in rosso, ovvero tutti quei soci che, iscritti ad un determinato settore nel corso di una stagione, hanno deciso di interrompere il proprio rapporto con la scuola nella stagione successiva. La presenza di questa riga permette di capire quali siano i settori che perdono più clienti e in generale il modello consente di comprendere in modo quantitativo le reali performance delle diverse aree.

A tale scopo abbiamo elaborato dei KPI in grado di valutare le prestazioni dei settori in termini di soci acquisiti (la somma delle celle di una riga relativa a un settore), partenti (soci che nella stagione successiva hanno rinnovato verso un diverso settore, nella matrice ci si riferisce alla somma delle celle di una colonna, esclusa quella dei non rinnovi e quella relativa al corso stesso) e persi (soci che hanno interrotto il proprio rapporto con la scuola).

Sono stati definiti tre indici. Il primo, che chiameremo PA (Persi su Acquisiti), indica il numero di soci che nel corso della stagione successiva non ha frequentato la scuola per ogni socio acquisito nella stagione precedente:

$$PA = \frac{Soci\ persi}{Soci\ acquisiti}$$

Il valore dell'indice può variare in linea teorica da 0 a +∞. Chiaramente, un valore prossimo allo 0 è sinonimo di un'ottima performance da parte di un settore che ha attirato molti clienti e ne ha persi pochi, mentre un indice tendente all'unità suggerisce di apportare modifiche alla gestione delle attività riguardanti i corsi di quell'area. L'indicatore PA si limita però a valutare come la performance di un determinato settore abbia influito sul business della scuola, senza dare indicazioni sulla capacità del settore di fidelizzare i clienti, che abbiamo visto essere un aspetto chiave del nostro studio.

Il secondo indice è stato pensato al fine di valutare anche la capacità di un'area della scuola di fidelizzare il cliente ed è stato denominato PPA (Persi e Partenti su Acquisiti):

$$PPA = \frac{Soci\ persi + Soci\ partenti}{Soci\ acquisiti}$$

L'indicatore PPA è simile al PA ma, considerando anche i soci partenti, concentra l'attenzione su tutti quei soci che hanno rinnovato verso lo stesso settore. Un indice che considerasse esclusivamente i soci partenti e non quelli persi è stato reputato meno affidabile in quanto anche un valore pari a 0 (quindi ottimo) non avrebbe permesso di considerare come valide le prestazioni del settore senza avere l'indicazione di quanti soci abbia perso.

Il terzo indice valuta esclusivamente la fidelizzazione del settore in termini percentuali ed è appunto denominato IF (Indice di Fidelizzazione):

$$IF = \frac{Rinnovi}{Soci\ totali} \times 100$$

L'indicatore IF non tiene in considerazione i soci acquisiti ma soltanto il totale dei soci relativi alla stagione precedente, dei quali una certa porzione rinnova. Fornisce quindi un'indicazione più precisa riguardo il grado di fidelizzazione ma non dice nulla a livello di performance in termini di attrattiva verso i soci di altri corsi.

Calcolando gli indici PA, PPA e IF per i numeri della matrice in [Figura 31] abbiamo stilato una classifica di esempio:

Indice PA	
Settore	Valore
D	0
B	0,077
E	0,333
C	0,355
F	1,00
A	1,00

Indice PPA	
Settore	Valore
E	0,33
B	0,39
C	0,48
D	1,00
F	1,00
A	3,67

Indice IF	
Settore	Valore
E	70,0%
B	64,3%
C	59,5%
F	50%
A	27,3%
D	25%

[Figura 32] Classifica dei settori per gli indici PA, PPA e IF

Dalle classifiche si nota come i settori dalle prestazioni migliori siano "B" ed "E"; il primo perde pochi soci mentre il secondo è il settore che fidelizza di più. Il settore "D" non perde soci ma fidelizza poco mentre la prestazioni apparentemente pessime del settore "A" sono in parte giustificate dalla fascia d'età di riferimento: si tratta di un'area della scuola dedicata ai più piccoli e quindi di passaggio, non destinata alla fidelizzazione interna e sicuramente non rappresenta una destinazione per i rinnovi da altri settori.

Il modello proposto non considera invece i nuovi iscritti, che potrebbero rappresentare un'implementazione futura per analisi più approfondite, così come viene trascurato l'impatto dei singoli settori sulle attività della scuola in termini quantitativi. Gli indici prestazionali vanno infatti incrociati con il trend dei settori per capire le reali performance degli stessi e come queste incidano nel complesso sulla scuola.

4.7 Analisi proposte

In questo paragrafo tratteremo alcune tipologie di analisi che, a causa della natura dei dati a disposizione, non sono state svolte o non hanno presentato risultati particolarmente interessanti per la scuola. Dal momento in cui questo studio vuole offrire alla società uno strumento utile per guidare le decisioni del management, si vogliono mostrare anche le analisi potenzialmente interessanti per il futuro della scuola, affinchè la stessa si possa organizzare in ottica gestione dati per far si che queste rappresentino un valore aggiunto a medio-lungo termine.

4.7.1 Regole di associazione

Una delle analisi proposte, al fine di comprendere le relazioni tra i vari corsi e servizi offerti dalla società, prevede di esaminare le sottoscrizioni ai servizi da parte dei soci alla ricerca di regolarità negli acquisti. L'obiettivo è quello di capire quali siano i servizi più gettonati dai soci che si iscrivono a determinati corsi e come questi siano legati tra di loro; in generale, si cerca quindi di trovare tutti quei servizi che vengono acquistati spesso insieme.

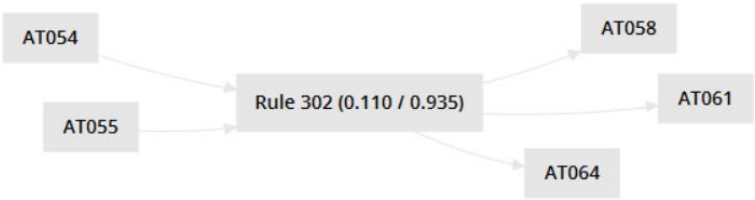
A questo scopo, la metodologia di data mining utilizzata consiste nelle regole di associazione. L'algoritmo selezionato, FP-growth [17], riceve in input una lista di

Cod_Socio	Servizi
s0127	AT005
s0128	AT005,AT042,AT026
s0129	AT005,AT026
s0130	AT012,AT014
s0131	AT007,AT029,AT033,AT040,AT054,AT055,AT058,AT061,AT064,AT067,AT075

[Figura 33] Esempio di input per l'algoritmo FP-growth

transazioni (*itemset*) ognuna delle quali contiene una serie di servizi sottoscritti (*item*), come mostrato in [Figura 33], dove ogni itemset è relativo all'insieme di sottoscrizioni del socio in una determinata stagione. In base a parametri settati appositamente, l'algoritmo genera un *frequent-pattern tree* che fornisce informazioni quantitative sulla frequenza degli itemset e permette di calcolare successivamente le regole di associazione.

L'output del processo è rappresentato da una serie di regole, la cui affidabilità è valutabile attraverso alcuni parametri (in primo luogo *lift* e *confidence*), che mettono in correlazione tra loro i servizi; un esempio è mostrato in [Figura 34].



[Figura 34] Esempio di regola di associazione.
Tra parentesi i valori dei parametri Support e Confidence

La "Rule 302" afferma che le sottoscrizioni ai servizi "AT054" e "AT055" sono spesso accompagnate da quelle di "AT058", "AT061" e "AT064".

Il parametro *support* indica che l'itemset di riferimento è presente nell'11% dei record e la regola è confermata nel 93,5% dei casi (parametro *confidence*). Per questa regola la correlazione calcolata (*lift*) assume un valore molto elevato: 8,201.

Questi parametri suggeriscono una certa bontà della regola ma non dell'analisi, che risulta essere invece di scarso interesse per il caso in esame. Il motivo risiede nel fatto che la scuola invita i propri iscritti ai vari settori a sottoscrivere servizi mirati, alcuni dei quali fanno già parte dell'offerta, e le conseguenti regole di associazione calcolate risultano scontate. Un secondo fattore da tenere in considerazione consiste nella granularità dei dati che non rende possibile applicare questo modello ad alcuni dei servizi, offerti separatamente dalla scuola ma non espliciti nei dati a disposizione dove risultano invece aggregati.

4.7.2 Analisi predittive

Alla famiglia delle possibili analisi descrittive, che abbiamo ampiamente trattato nel corso del capitolo, si aggiunge quella delle analisi predittive, utili invece a stimare il valore futuro di determinate variabili. Nel caso in esame, i modelli predittivi assumono una certa importanza per la possibilità di capire come le scelte

di gestione influiscano sull'andamento del business, in termini di iscrizioni ai corsi o sottoscrizioni ai servizi. Attraverso una tecnica di regressione siamo infatti in grado di stimare il numero e la natura dei soci iscritti a un settore nel corso di una stagione prendendo in considerazione prezzo dell'abbonamento, offerte specifiche, presenza di eventi correlati, iniziative di marketing.

Il modello di analisi predittiva è un tema di grande interesse per qualsiasi attività commerciale perchè è in grado di rappresentare un supporto affidabile per il processo decisionale, aiutando a comprendere la possibile reazione dei clienti ai cambiamenti in ogni aspetto del business.

Il metodo di regressione riceve come input il dataset che viene in primo luogo utilizzato come *training set*, ovvero come modello sul quale costruire e settare i parametri da utilizzare; in un secondo momento, il modello generato viene applicato su un secondo dataset (il *testing set,* che nel nostro caso risulta essere il medesimo) e ne vengono valutate le performance, ovvero l'accuratezza e il coefficiente di determinazione (R^2) che indica la bontà della curva di regressione generata. La valutazione del modello è possibile dal momento in cui esso viene utilizzato, in fase di testing, per stimare i valori di interesse per ogni record del nostro dataset, i quali sono quindi direttamente confrontabili con i valori reali dei dati. Come anticipato nella premessa in 4.1, la società non ha tenuto traccia dei dati relativi a promozioni e iniziative commerciali delle stagioni passate, rendendo di fatto impossibile la creazione di un modello previsionale affidabile; le informazioni in nostro possesso sono esclusivamente legate alla memoria dei singoli e pertanto di difficile utilizzo per analisi di questo tipo.

4.7.3 Analisi incrociate

Nel corso di questo studio abbiamo preso in esame principalmente le attività dell'area A.P.D. della scuola per una questione di semplicità e chiarezza dovute alla maggior completezza dei dati a disposizione.

In realtà, le stesse analisi qui presentate possono essere replicate, dove possibile, anche sui dati della Scuola Sci per offrire una panoramica più completa del business della società, andando a confrontare il target delle due aree, il trend dei settori e la numerosità degli iscritti, così come la loro fedeltà. Le due aree della scuola sono gestite e organizzate indipendentemente l'una dall'altra, così come i soci iscritti sono soliti frequentarne solo una delle due, nel corso della singola stagione.

Quello che può succedere, invece, è che un socio iscritto alla Scuola Sci frequenti poi l'A.P.D. nella stagione successiva. In quest'ottica, è molto interessante capire quale sia la tipologia di soci che affrontano questo passaggio, in termini di età dei singoli e famiglia. È inoltre possibile implementare la matrice dei rinnovi vista in 4.6 tenendo in considerazione tutti i corsi di Scuola Sci e A.P.D. per comprendere come le due aree siano legate tra di loro e ricostruire il percorso dei soci per ogni settore. Incrociare i dati delle due aree consente di esplicitare fenomeni intrinseci all'intero business della scuola, come per esempio la relazione che lega le performance di un'area con i trend dell'altra.

Dal risultato di queste analisi si potrebbe capire come la gestione indipendente delle aree possa in realtà essere migliorabile andando ad agire sui punti in comune, facendo leva sui giusti fattori in grado di incrementare le prestazioni di entrambe.

Conclusioni

Questo studio ha mostrato come l'applicazione delle metodologie di Business Intelligence consenta di estrapolare conoscenza dai dati grezzi, immagazzinati e mai analizzati in anni di attività, e offrire validi spunti per incrementare le performance d'impresa. L'individuazione delle criticità e degli aspetti chiave del business rappresenta la base per accrescere la propria competitività nel settore, qualunque esso sia, e per consolidare un vantaggio competitivo che permetta all'azienda di distinguersi nel mercato, a prescindere dalla propria dimensione.

I risultati dello studio hanno infatti evidenziato fenomeni intrinseci ai dati e difficilmente osservabili e quantificabili dagli addetti della scuola, permettendo la definizione di un target di mercato ben preciso e fornendo spunto per azioni correttive, volte al miglioramento delle prestazioni d'impresa e all'allineamento con le aspettative del management.

Al fine di conseguire un vantaggio competitivo duraturo nel tempo, la società ha il compito di organizzare la propria struttura per sfruttare con continuità i modelli di Business Intelligence adoperati. A tale scopo, il lavoro ha avuto come oggetto la progettazione di un database relazionale in grado di conservare i dati in maniera strutturata e standardizzata con l'obiettivo di renderli facilmente recuperabili e analizzabili.

Il presente studio ha fornito quindi alla scuola uno strumento utile come base nel processo di data management e, attraverso il lavoro di data analysis, una certa

varietà di tipologie di analisi volte al monitoraggio delle attività e alla valutazione delle performance. Tra gli aspetti presi in esame troviamo le anagrafiche dei soci, che hanno permesso di comprendere la natura dei clienti della scuola, definire il principale target di mercato e individuare alcune criticità nella gestione dei rinnovi e nella capacità di attrarre nuovi soci. Sono state analizzate le zone di provenienza dei clienti e la tendenza ad iscriversi come gruppo familiare piuttosto che come singolo individuo, andando quindi a focalizzare l'attenzione sugli aspetti relativi al marketing per ottimizzare offerte, eventi e servizi.

Sono state condotte analisi temporali sfruttando i dati relativi alle lezioni svolte per evidenziare il trend di affluenza annuale, settimanale e giornaliero, e gli stessi dati sono stati aggregati e incrociati con quelli meteorologici per capire come il fattore meteo influisca sull'attività della scuola.

È stata condotta un'approfondita analisi sulle performance dei vari settori della scuola dalla quale è emerso come i clienti seguano un percorso naturale che può, o meno, rispecchiare quello ideale pensato dal management della società. Il tool fornito è stato pensato per valutare oggettivamente e quantitativamente le prestazioni di un settore prendendo in considerazione i movimenti dei soci all'interno della scuola e può essere replicato filtrando o aggregando i dati a piacimento.

Infine, sono state proposte ulteriori analisi che, a causa della natura dei dati a disposizione, non hanno condotto a considerazioni in linea con le aspettative ma hanno esteso le possibilità di esaminare i dati da un punto di vista più scientifico

attraverso analisi predittive e regole di associazioni utili nel supporto al processo decisionale.

Il presente studio nasce dalla necessità di una scuola sci (o forse dalla semplice curiosità della dirigenza) di avviare un processo evolutivo che possa avvicinare la società a un pensiero "aziendale", sposando i concetti di innovazione e digital transformation fondamentali, in un futuro prossimo, per consolidare e rafforzare la propria presenza nel mercato. L'obiettivo originale consisteva nel prendere una consistente mole di dati, mai analizzati, e dar loro voce al fine di scoprire relazioni nascoste o fenomeni non direttamente osservabili dalla quotidiana esperienza lavorativa. Nel complesso si ritiene che lo scopo sia stato raggiunto e che ora la società in esame abbia le basi per compiere questo salto, partendo dal concetto di data management per sfruttare una risorsa interna già presente e fonte di un potenziale vantaggio competitivo.

Bibliografia

[1] R. M. Devens, Cyclopaedia of Commercial and Business Anecdotes, 1865.

[2] H. P. Luhn, «A Business Intelligence System,» *IBM Journal*, 1958.

[3] H. G. Sol, Expert Systems and Artificial Intelligence in Decision Support Systems, 1987.

[4] D. J. Power, «A Brief History of Decision Support Systems,» Marzo 2007. [Online]. Available: http://dssresources.com/history/dsshistory.html.

[5] M. Rouse, «WhatIs.com,» Aprile 2007. [Online]. Available: https://whatis.techtarget.com/definition/digitization.

[6] R. Hat, «The Enterprisers Project,» [Online]. Available: https://enterprisersproject.com/what-is-digital-transformation.

[7] D. Aliperto, «CorCom,» 20 Giugno 2019. [Online]. Available: https://www.corrierecomunicazioni.it/lavoro-carriere/competenze/istat-solo-il-47-delle-imprese-italiane-ha-un-elevato-livello-di-digitalizzazione/.

[8] J. McConnell, «Harvard Business Review,» 28 Agosto 2015. [Online]. Available: https://hbr.org/2015/08/the-company-cultures-that-help-or-hinder-digital-transformation.

[9] E. L. K. G. V. I. Behnam Tabrizi, «Harvard Business Review,» 13 Marzo 2019. [Online]. Available: https://hbr.org/2019/03/digital-transformation-is-not-about-technology?referral=03759&cm_vc=rr_item_page.bottom.

[10] V. Dhar, «Data Science and Prediction,» 2012.

[11] S. Tansley, The Fourth Paradigm: Data-intensive Scientific Discovery, 2009.

[12] J. Leek, «Simplystatistics,» 12 Dicembre 2013. [Online]. Available: https://simplystatistics.org/2013/12/12/the-key-word-in-data-science-is-not-data-it-is-science/.

[13] J. T. &. Management, «Situazione congiunturale Montagna Bianca Italiana, Inverno 2019-2020, Previsioni e Tendenze,» 2019.

[14] J. T. &. Management, «Situazione congiunturale Montagna Bianca Italiana, Inverno 2018-2019,» 2018.

[15] E. Valdani, Marketing strategico. Un'impresa proattiva per sviluppare capacità market driving e valore, 2003.

[16] A. Piemonte, «Arpa,» [Online]. Available: http://www.arpa.piemonte.it/.

[17] J. Han, «Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach,» Kluwer Academic Publishers, 2000.

[18] «Ziotis,» 2 Luglio 2019. [Online]. Available: https://www.ziotis.com/rpa-ai-solutions/.

[19] C. Hayashi, Data Science, Classification, and Related Methods, 1998.

[20] maticmind, «Big Data Analytics: l'evoluzione della business intelligence,» 27 Marzo 2019. [Online]. Available: https://www.maticmind.it/approfondimenti/big-data-analytics-levoluzione-della-business-intelligence.

[21] Thibaud, «Data Science in China: The Rise of Big Data,» 5 Luglio 2016. [Online]. Available: https://daxueconsulting.com/data-science-china-the-rise-of-big-data/.

[22] D. Laney, «3D Data Management: Controlling Data Volume, Velocity and Variety,» 6 Febbraio 2001. [Online]. Available: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[23] H. Rosling, Factfulness, 2018.

[24] B. Marr, «Why only one of the 5 Vs of big data really matters,» IBM, 19 Marzo 2015. [Online]. Available: https://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters.

Ringraziamenti

Giunto al termine del percorso di laurea, mi sento di dedicare qualche riga per ringraziare tutte le persone con le quali ho collaborato, trascorso gran parte del mio tempo o che mi sono state vicine per tutti questi anni.

Ringrazio in primo luogo i miei genitori, che mi hanno sempre supportato e sopportato, in particolare dall'ormai lontano Settembre 2012 quando ebbe inizio la mia avventura al Politecnico di Torino.

Ringrazio i miei tutor, Flavio e Alessia, che mi hanno seguito durante la stesura di questa tesi e dai quali ho imparato e spero di poter imparare ancora molto, collaborando nel prossimo futuro.

Ringrazio la professoressa Cerquitelli per aver reso possibile questo lavoro con la sua supervisione e disponibilità, nonostante i tantissimi impegni.

Durante il percorso accademico ho incontrato persone fantastiche con le quali ho condiviso gioie, ansie, momenti da ricordare e bocciature, anch'esse da ricordare. Persone con le quali sono cresciuto e che mi hanno insegnato tanto al di fuori del mondo fatto di libri, slide ed esami; alcune di queste rimarranno ottimi colleghi mentre altre condivideranno con me tante altre avventure.

Ringrazio in particolare Lorenzo, Alessia, Marco e Gian con i quali ho affrontato e superato esami e trascorso tante ore di studio fianco a fianco, tra mal di testa e risate.

Ringrazio Laura, conosciuta tra i banchi del Politecnico, che mi ha cambiato la vita e che ha reso un piacere affrontare le difficoltà dell'ultimo anno di università.

Ringrazio Cristian, Alessandro, Francesco, Domenico, Giulia, Cosmina, Mattia e tutti gli amici che mi hanno arricchito la vita e continueranno a farlo nei prossimi anni.

Ringrazio tutti gli amici del Fantacammello, una competizione unica e giocosa che ha regalato tante risate, soddisfazioni e rosicate, giorno dopo giorno, per anni.

Senza tutti voi non sarebbe la stessa vita. Grazie.