

**UNIVERSITA' DEGLI STUDI DI PADOVA**



**Facoltà di SCIENZE STATISTICHE**

**CORSO DI LAUREA IN STATISTICA E  
TECNOLOGIE INFORMATICHE**

**Relazione Finale di Laurea Triennale**

**ANALISI DELLA CLIENTELA  
di Lotto Sport Italia S.p.A.**

Relatore: Ch. mo Prof. Bruno Scarpa

Laureando: Michel Comazzetto

Matricola: 554356 – STI

Anno Accademico: 2008/2009



# Indice:

---

|                                                           | <b>Pagina</b> |
|-----------------------------------------------------------|---------------|
| <b>INDICE</b>                                             | <b>3</b>      |
| <b>Introduzione</b>                                       | <b>5</b>      |
| <b>CAPITOLO I</b>                                         |               |
| <b>PRESENTAZIONE DELL’AZIENDA E OBIETTIVO DELLO STAGE</b> | <b>8</b>      |
| 1.1 L’azienda e il distretto montebellunese               | 8             |
| 1.1.1 Il distretto dello “sportsystem” montebellunese     | 8             |
| 1.1.2 Lotto e Lotto Sport Italia S.p.A.                   | 13            |
| 1.2 Obiettivo stage                                       | 17            |
| <b>CAPITOLO II</b>                                        |               |
| <b>PRODUCT MARKETING E DATA MINING</b>                    | <b>19</b>     |
| 2.1 Concetti di marketing utili nel contesto aziendale    | 19            |
| 2.1.1 Il Prodotto                                         | 19            |
| 2.2 Cenni sul Product Marketing                           | 22            |
| 2.2.1 Le innovazioni di prodotto                          | 22            |
| 2.2.2 Organizzazione del product management               | 23            |
| 2.2.3 Il product manager                                  | 24            |
| 2.3 Concetti utili di data mining                         | 26            |
| 2.3.1 Ambiti applicativi del data mining                  | 27            |
| 2.3.2 Raggruppamento in classi                            | 28            |
| 2.3.3 Associazioni tra variabili                          | 29            |
| 2.3.4 Analisi della correlazione canonica                 | 31            |

## **CAPITOLO III**

|                                                                            |               |
|----------------------------------------------------------------------------|---------------|
| <b>ANALISI</b>                                                             | <b>35</b>     |
| 3.1 Organizzazione e illustrazione del dataset                             | 35            |
| 3.2 <i>Prima Analisi: Raggruppamento dei clienti di Lotto Sport Italia</i> | 38            |
| 3.2.1 Applicazioni dei metodi di raggruppamento                            | 38            |
| 3.2.2 Conclusioni                                                          | 44            |
| 3.3 <i>Seconda Analisi: Analisi della correlazione canonica</i>            | 51            |
| 3.3.1 Analisi della correlazione canonica                                  | 51            |
| 3.3.2 Analisi uomo – donna                                                 | 52            |
| 3.3.3 Analisi Calcio-Calciotto – Tennis-Running                            | 56            |
| 3.3.4 Conclusioni                                                          | 59            |
| 3.4 <i>Terza Analisi: Algoritmo ‘A priori’</i>                             | 60            |
| 3.4.1 Applicazione dell’algoritmo “apriori”                                | 60            |
| <br><b>CONCLUSIONI</b>                                                     | <br><b>62</b> |
| <br><b>BIBLIOGRAFIA E SITOGRAFIA</b>                                       | <br><b>64</b> |

## **Introduzione**

Il presente elaborato è frutto della mia esperienza di stage che si è svolto all'interno della ditta Lotto Sport Italia S.p.A. nel periodo che va dal 10 marzo 2009 al 5 giugno 2009 nella sede principale dell'azienda posta a Trevignano in provincia di Treviso.

Lo stage prevedeva l'inserimento all'interno dell'ufficio Product Marketing, più specificatamente nel settore delle calzature cosiddette "performance", che all'interno di Lotto comprendono i settori Calcio, Calcetto, Tennis, Running e una parte di accessori legati prevalentemente al mondo del calcio e del calcetto quali palloni, parastinchi e guanti da portiere.

Nella prima parte della tesina seguirò a presentare l'azienda nel suo complesso e la realtà territoriale in cui è nata e in cui continua da anni la sua attività, l'area a cui facciamo riferimento è quella che viene definita come "Area dello sportsystem montebellunese", una zona che raggruppa numerose aziende del settore calzaturiero attorno alla città di Montebelluna (in provincia di Treviso) per un raggio di una trentina di chilometri.

In seguito vi sarà una parte dedicata all'illustrazione di alcuni elementi teorici per comprendere la realtà aziendale in cui ho effettuato lo stage e le tecniche utilizzate per le analisi. Saranno quindi presentate le caratteristiche principali del product marketing e come viene utilizzata la statistica in questo contesto, vi sarà poi spazio per gli strumenti statistici specifici utilizzati nelle analisi.

Nella terza parte sono applicati alcuni modelli per l'analisi dei dati, attraverso i quali cercheremo di arrivare a delle conclusioni che potrebbero risultare utili nell'ambito delle decisioni aziendali da prendere in futuro.

Dal punto di vista aziendale le analisi condotte in questo elaborato avevano due fini principali. Il primo era quello di avere una riclassificazione dei clienti differente da quella già utilizzata dall'azienda stessa per raggruppare le diverse tipologie di clienti, la seconda era quella di ricercare delle relazioni tra i prodotti per capire se era in qualche modo possibile collegare la vendita di un prodotto ad un altro. Partendo da queste necessità dell'azienda abbiamo quindi condotto tre tipi di analisi per rispondere alle questioni poste dall'azienda.

La prima analisi consiste nel raggruppare i clienti dell'area EMEA (Europe, Middle East and Africa) in base ad una serie di caratteristiche dei singoli clienti e alla quantità ordinata dei diversi modelli di calzature "performance".

Abbiamo proceduto in questa fase alla raccolta di dati che riguardano i clienti di Lotto Sport Italia S.p.A., cercando di riassumere in un dataset il più completo possibile informazioni riguardanti il posizionamento geografico, la tipologia distributiva, l'agente che si occupa dei contatti con quel negozio e molte altre informazioni oltre alle quantità di calzature ordinate nella stagione invernale del 2008.

Attraverso l'applicazione di modelli differenti abbiamo cercato di tracciare dei profili che raggruppassero clienti con caratteristiche simili di acquisto per poter eventualmente attuare delle iniziative promozionali uniche a clienti che presentano profili comuni.

Nella seconda parte abbiamo applicato l'analisi della correlazione canonica per evidenziare (nel caso esistessero) interdipendenze tra gruppi di prodotti.

Abbiamo costruito un dataset che risultasse indipendente dai vari clienti ma che permettesse di stabilire l'esistenza di correlazioni tra modelli concettualmente differenti. Un modo per utilizzare tale informazione potrebbe essere quella di proporre dei pack che comprendano questi modelli per incentivare le vendite, un esempio potrebbe essere la vendita di un determinato tipo di scarpa da calcio assieme ad una scarpa da running da utilizzare durante la preparazione atletica ad inizio stagione.

Infine, l'ultima analisi, si propone di andare a verificare attraverso l'algoritmo "a priori", un metodo per l'apprendimento di regole di associazione, l'esistenza di sottoinsiemi di item che risultano nell'insieme delle transazioni almeno S volte. Queste associazioni saranno rappresentate attraverso un grafo dove la presenza di un ramo che unisca due modelli sarà da considerarsi come "presenza di associazione" mentre l'assenza di questo ramo indica un'assenza di associazione tra i modelli non collegati nel grafo.

Le tipologie di analisi sviluppate in questa tesina sono solo alcune delle moltissime metodologie che si possono utilizzare nell'ambito del data mining.

La scelta di applicare al dataset questo tipo di analisi è dovuta principalmente al fatto che l'azienda non ha dato un indirizzamento specifico su quali analisi condurre, quindi sono state applicate tecniche per ottenere risultati più generici possibili, senza andare a ricercare una determinata caratteristica.

## **CAPITOLO I**

### **PRESENTAZIONE DELL'AZIENDA E OBIETTIVO DELLO STAGE**

#### **1.1 L'azienda e il distretto montebellunese**

##### ***1.1.1 Il distretto dello “sportsystem” montebellunese***

La zona montebellunese rappresenta uno dei più importanti distretti industriali della calzatura sportiva italiana, situato nei pressi di Treviso e comprendente un'area di una trentina di chilometri nei dintorni di Montebelluna. A partire dall'inizio del '900 nascono le prime aziende frutto di un'evoluzione dai precedenti laboratori artigianali delle calzature già presenti nel territorio. È possibile suddividere lo sviluppo del distretto in alcune fasi, caratterizzate dall'evoluzione delle aziende, passate da una lavorazione pressoché artigianale ad una lavorazione industriale con l'introduzione di macchinari e con l'inserimento in alcuni casi di multinazionali nella gestione delle aziende stesse.

Un semplice quadro dell'evoluzione distrettuale potrebbe essere sintetizzato in questo modo:

##### **DISTRETTO ARTIGIANO-INDUSTRIALE (1911-1954)**

Nel 1951 sono dieci i comuni che formano il nucleo storico: Montebelluna, Pederobba, Arcade, Cornuda, Nervesa della Battaglia, Caerano San Marco, Trevignano, Volpago del Montello, Maser, Giavera del Montello.

Alcuni laboratori calzaturieri si trasformano in aziende vere e proprie dove viene attuata la divisione del lavoro e dove vi è l'introduzione delle prime macchine, nonostante la parte manuale costituisca ancora la maggioranza del lavoro.

Inizialmente la lavorazione si concentrava principalmente nella realizzazione di scarpe da lavoro, ma la svolta decisiva si ebbe nel dopoguerra con la progettazione di calzature adatte alle escursioni in montagna e, successivamente, della produzione di scarponi da sci, utilizzando come base di partenza la calzatura da montagna precedentemente elaborata.

### **DISTRETTO CONCORRENZIALE (1954-1974)**

Il miracolo economico segna il tempo della plastica, delle diversificazioni produttive, del benessere e del consumismo.

La crescita, dovuta al boom economico italiano, fa aumentare anche la domanda di scarponi da sci, che diventano il prodotto leader del Distretto.

Le innovazioni tecnologiche culminano con la rivoluzione della plastica, passaggio fondamentale nell'evoluzione del distretto in quanto, rappresenterà un notevole vantaggio per le aziende in grado di adattare la propria produzione alle nuove tecnologie, mentre altre decideranno di esplorare altri settori abbandonando quello tradizionale della montagna.

La gerarchia del Distretto è ormai netta: alcuni grandi marchi, qualche decina di piccoli e medi marchi, ed una miriade di subfornitori.

### **DISTRETTO INTERNAZIONALE (1974-1989)**

1985 – COMUNI (15): Nucleo storico (10) + Nuovi comuni (5): Asolo, Altivole, Castelcucco, Veduggio, Istrana.

Altro passaggio fondamentale nell'evoluzione del distretto è rappresentato dall'internazionalizzazione di alcune aziende. La vendita della Caber alla Spalding (azienda statunitense di abbigliamento e attrezzatura sportiva) apre le porte del Distretto alle multinazionali, assumendo la fisionomia che lo rende peculiare nel panorama italiano e mondiale.

Le diversificazioni produttive inoltre favoriscono la nascita di numerose altre piccole aziende.

Lotto e Diadora introducono prodotti estivi (scarpe da tennis e da jogging) in competizione con le multinazionali tedesche ed americane; iniziano inoltre nello stesso periodo a delocalizzare nel Far East parte della produzione.

### **DISTRETTO GLOBALIZZATO (1989-2008)**

2005 – COMUNI (26): Nucleo storico (10) + Nuovi comuni 1985 (5) + Valdobbiadene, Cavaso del Tomba, Monfumo, Castelcucco, Fonte, Vidor, Castello di Godego, Castelfranco Veneto, Ponzano Veneto, Villorba, San Biagio di Callalta, Arcade.

Con gli anni '90 la globalizzazione avviluppa nella sua rete tutte le aziende del Distretto che vive una profonda metamorfosi. Si allenta, infatti, il rapporto con il territorio di origine. Dopo la caduta del muro di Berlino e la fine dei regimi comunisti nell'Europa Orientale, parti rilevanti della produzione si spostano nell'Est europeo ed in Asia. Molte grandi aziende assumono le caratteristiche di commerciali.

Il fenomeno delle concentrazioni, avviato con l'acquisto della Nordica da parte della Benetton, sconvolge gli antichi equilibri del Distretto che accentua la sua verticalizzazione. Si assottiglia il numero dei laboratori e cresce il peso della delocalizzazione.

Con il nuovo Millennio il distretto vede ulteriormente intensificarsi il fenomeno delle concentrazioni. Nel distretto si forma il più importante gruppo mondiale per prodotti da neve, costituito dai marchi Nordica, Dolomite, Tecnica, Lowa (a cui si aggiunge Rollerblade).

Grande successo delle scarpe da città con i marchi Geox, Stonefly. La definizione tradizionale di distretto (un territorio, un prodotto specifico, ecc...) negli ultimi anni si è molto modificata. Il territorio è diventato il mondo e i prodotti sono una gamma molto varia e in continua crescita. (Durante A., Distretto cosmopolita)

A supporto di quanto illustrato in queste pagine possiamo vedere con questa serie di grafici l'andamento generale del mercato italiano della calzatura e in particolare dell'apporto che il distretto montebellunese dà a questo importante mercato.

I grafici in questione sono stati presi dal rapporto OSEM sul distretto montebellunese, questo rapporto, redatto quasi annualmente grazie al museo dello scarpone di Montebelluna e all'aiuto di Veneto Banca, offre una panoramica generale sull'andamento del distretto dello sport system nei settori del fatturato, dell'occupazione, dell'esportazione e di tutto ciò che riguarda le aziende del territorio.

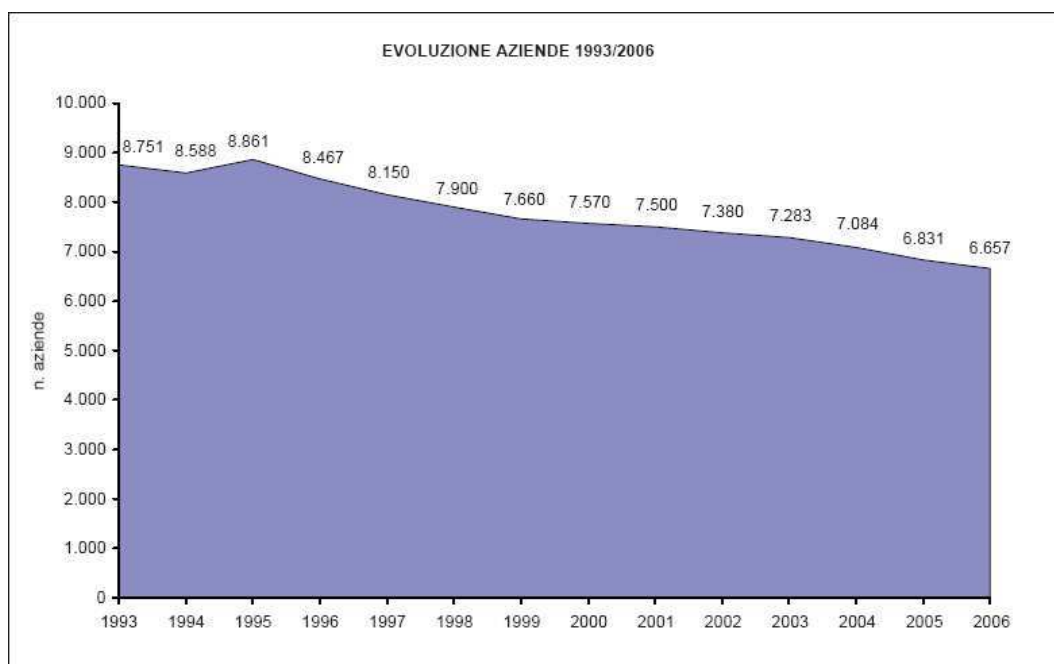


Figura 1. Evoluzione del settore calzaturiero italiano

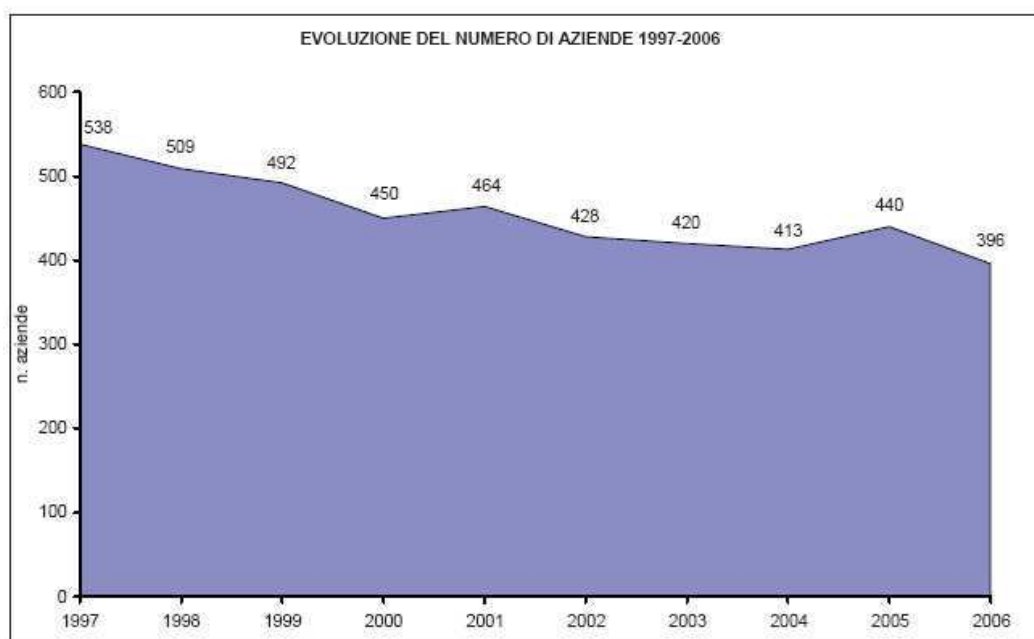


Figura 2. Evoluzione del distretto montebellunese

Le figure 1 e 2 ci offrono una panoramica generale del numero di aziende che operano nel settore della calzatura. Possiamo notare che, nel corso del periodo considerato, vi sia un trend negativo costante che sta portando alla diminuzione delle aziende operanti nel settore. Anche il distretto montebellunese segue il trend

nazionale passando dalle 538 aziende del 1997 alle 396 aziende del 2006, anno in cui è stato redatto il rapporto da cui sono stati presi i grafici.

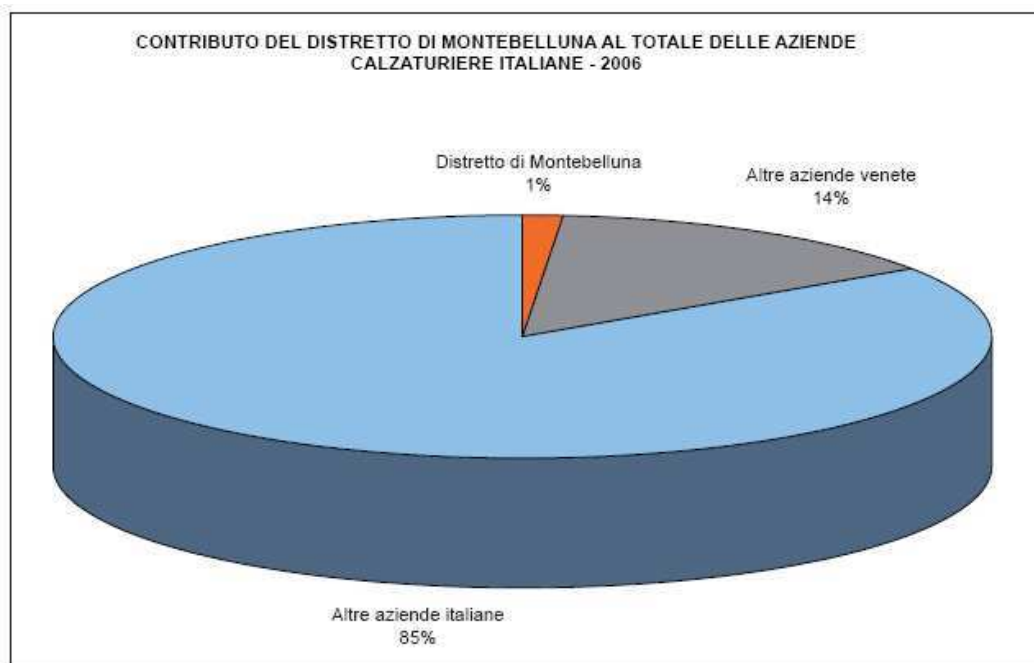


Figura 3. Contributo del distretto di Montebelluna al settore calzaturiero italiano

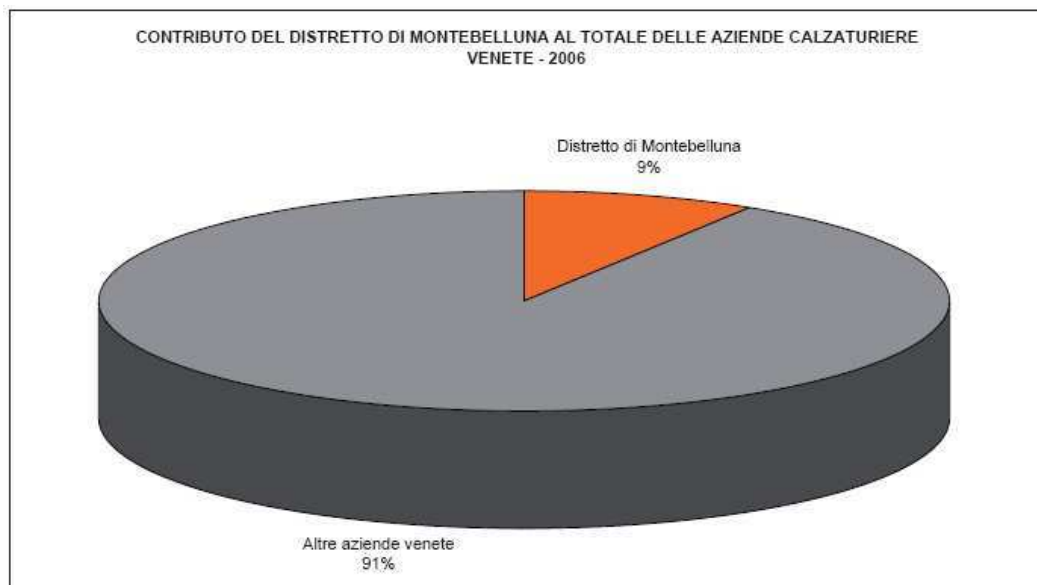


Figura 4. Contributo del distretto di Montebelluna al settore calzaturiero veneto

Le figure 3 e 4 ci illustrano il peso che il distretto montebellunese ha rispetto al totale italiano nel primo caso e al totale delle aziende venete nel secondo. Dal primo grafico possiamo notare come il peso del veneto nel settore sia del 15% del

totale nazionale mentre dal secondo vediamo la forte influenza del distretto nell'ambito regionale arrivando a sfiorare il 10% sul totale delle aziende venete operanti nel settore.

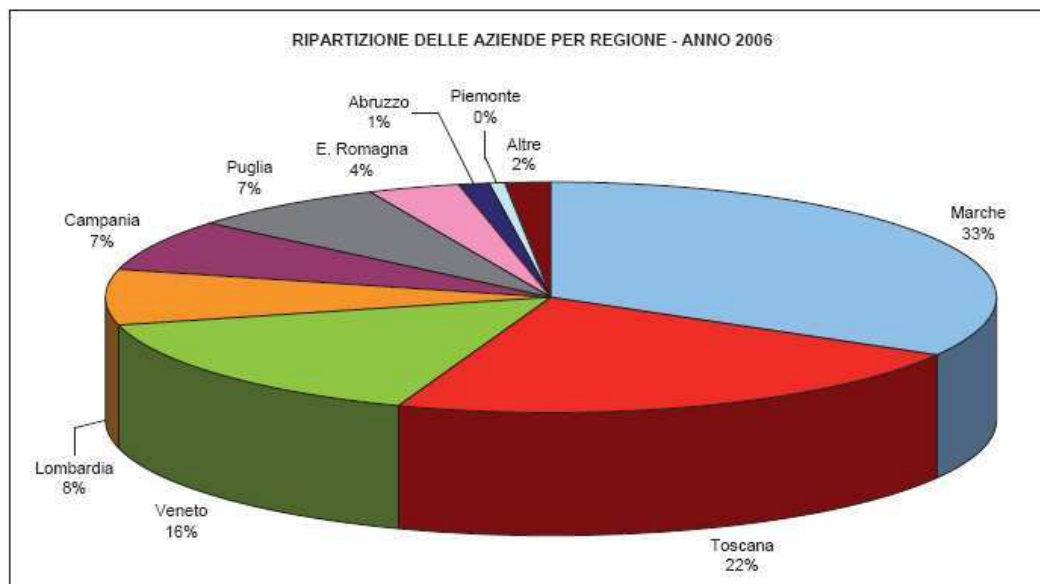


Figura 5. Ripartizione delle aziende per regione – Anno 2006

L'ultimo grafico ci dà una visione globale del settore calzaturiero suddiviso per regione, da cui possiamo osservare la quota notevole del Veneto dietro solo a Marche e Toscana.

Grafici tratti da Rapporto OSEM 2006 a cura di Aldo Durante.

### 1.1.2 Lotto e Lotto Sport Italia S.p.A.

Ecco ciò che possiamo trovare nel sito dell'azienda per quanto riguarda la storia e la sua evoluzione dalla fondazione fino ad oggi.

*“Lotto fa il suo esordio tra le aziende produttrici di calzature sportive nel 1973. Le scarpe da tennis segnano l'inizio della produzione, a cui si aggiunge quella di modelli da basket, pallavolo, atletica, calcio e in seguito l'abbigliamento sportivo. Durante i primi 10 anni, Lotto si concentra sul mercato italiano, diventando uno dei marchi di riferimento nel settore dell'articolo sportivo e una delle aziende leader nel tennis.*

*Negli anni '80 inizia la realizzazione dei primi modelli di scarpe da calcio e*

*iniziano le grandi collaborazioni con atleti e squadre di importanza internazionale: John Newcombe, Andrei Gomes e José Luis Clerc nel tennis, Ruud Gullit e l'Olanda nel calcio. Il coinvolgimento degli atleti, sia nella progettazione che nella messa a punto dei prodotti abbinati alla loro immagine, porta l'azienda a diventare leader nel tennis, nel calcetto e nel calcio. Negli stessi anni, comincia la distribuzione nei mercati esteri. La crescita a livello internazionale procede velocemente e, nel giro di un decennio, il marchio viene distribuito in oltre 60 Paesi in tutto il mondo.*

*Nel Giugno del 1999 l'Azienda viene rilevata da un gruppo di imprenditori locali, già attivissimi nel settore dello sport, con a capo Andrea Tomat che assume il ruolo di Presidente e Direttore Generale della nuova Società, ribattezzata **Lotto Sport Italia S.p.A.** Obiettivo della nuova proprietà, quello di valorizzare i punti di forza del marchio - dinamismo, innovazione, qualità, design made in Italy e passione per lo sport - uniti ad un servizio sempre più attento ed efficace nei confronti del Cliente.*

*Oggi, in linea con la nuova missione aziendale, è stato potenziato il settore performance dando particolare enfasi alle calzature e all'abbigliamento tecnico calcio e tennis, a sostegno anche della leadership mondiale del marchio attraverso prodotti d'avanguardia per innovazione e design. In parallelo, sulla base di un know how produttivo, tecnico e stilistico, è stata sviluppata un'idea di abbigliamento e calzature leisure per uomo e donna con un'immagine e un gusto sport inspired nella scelta dei tessuti, dei colori e delle forme, calzature e abbigliamento dedicate al tempo libero. Lo stesso know how è stato utilizzato per dare nuovo impulso alle collezioni di abbigliamento e calzature junior, con la nascita della linea Air Zone, studiata per massimizzare il comfort e il benessere dei bambini durante tutte le attività. Sempre con marchio Lotto vengono, inoltre, prodotte e distribuite su licenza, da aziende terze, linee di intimo sportivo, cartotecnica e cosmesi“. (Fonte: Sito internet Lotto Sport Italia S.p.A.)*

Strutturalmente l'azienda può essere descritta come una struttura divisionale poiché vi è una separazione delle varie Business Unit per area geografica e per

marchio (Lotto, Lotto Works e Lotto Leggenda) ed all'interno delle varie Business Unit vi è un'ulteriore suddivisione in base alla tipologia di distribuzione. Le tipologie distributive adottate da Lotto Sport Italia si articolano in tre varianti: Agenzie, Distributori e Licenziatari.

- Agenzie: sono presiedute direttamente da Lotto Sport Italia, infatti, è l'azienda che si occupa direttamente della vendita ai clienti attraverso degli agenti che operano nella nazione in cui è presente questa tipologia di distribuzione. In questi casi è l'azienda che si fa carico di tutti i rischi legati alla vendita del prodotto.

In ogni paese è presente un'Agenzia che attraverso una rete di agenti distribuiti sul territorio seguono i vari clienti.

I paesi in cui Lotto vende attraverso le agenzie sono: Austria, Belgio, Croazia, Francia, Germania, Grecia, Lussemburgo, Olanda, Slovenia, Spagna, Regno Unito e Ungheria.

È quindi presente una sola agenzia per ogni nazione in cui viene praticata questa tipologia di vendita e ogni agenzia al suo interno opera con più agenti che si occupano di diverse aree geografiche.

- Distributori: costituiscono un passaggio aggiuntivo rispetto alle nazioni in cui l'azienda opera con le agenzie. In questi casi i distributori si occupano di comprare da Lotto i prodotti che successivamente venderanno direttamente ai negozi finali nel proprio paese.

Il distributore quindi assume propriamente il compito di rivenditore, assumendosi i rischi che ora non gravano più direttamente sull'azienda.

Il distributore quindi diventa per il paese nel quale opera l'unico "rivenditore" ufficiale del marchio.

I paesi in cui Lotto distribuisce i suoi prodotti utilizzando questa tipologia sono: Arabia Saudita, Bolivia, Bosnia, Bulgaria, Cipro, Cuba, Danimarca, Emirati Arabi, Ecuador, Estonia, Ghana, Irlanda, Islanda, Kuwait, Lettonia, Lituania, Malta, Norvegia, Paraguay, Perù, Polonia, Portogallo, Qatar, Repubblica Ceca, Romania, Russia europea, Serbia, Siberia, Slovacchia, Suriname, Svezia, Svizzera, Ucraina, Uruguay.

- Licenziatari: si tratta di aziende a cui è permesso produrre e vendere prodotti con marchio Lotto in cambio di un pagamento di *royalties*, ovvero un compenso che varia in base al peso del venduto. Questa tipologia permette di adeguarsi velocemente alle esigenze di mercati che magari sono differenti in alcuni aspetti rispetto a mercati più tradizionali quale può essere quello italiano ed europeo.

I licenziatari possono quindi rifornirsi di prodotti Lotto oppure creare delle vere e proprie linee parallele a quelle originali della casa, specifiche per le caratteristiche del mercato dove andranno a rivendere, sempre sottostando ad una serie di regolamenti che vincolano le aziende per quanto riguarda la tipologia e la qualità dei prodotti che deve essere certificata dalla sede centrale.

I paesi in cui sono presenti licenziatari Lotto sono: Argentina, Australia, Brasile, Brunei, Canada, Cile, Cina, Colombia, Egitto, Giappone, Corea, Hong Kong, India, Indonesia, Israele, Malaysia, Messico, Nuova Zelanda, Sabah, Sarawak, Singapore, Taiwan, Thailandia, Tunisia, Turchia, USA, Venezuela, Vietnam.

Oltre a queste tre forme di distribuzione Lotto possiede anche una serie di negozi monomarca in Italia e in Europa che vendono esclusivamente prodotti a marchio Lotto.

Il mercato italiano viene gestito direttamente attraverso una serie di agenti che operano in regioni o gruppi di regioni diverse e sono in contatto direttamente con i negozi che vendono prodotti marchiati Lotto.

L'azienda, presente in più di 80 paesi nel mondo, è in continua evoluzione alla ricerca di ritagliarsi una fetta sempre più consistente nel mercato mondiale della calzatura e dell'abbigliamento. Il settore performance, essendo il "core business" dell'azienda, è quello di maggior interesse, ma anche nei settori del tempo libero e del lavoro vi è un notevole apporto di risorse per arrivare a guadagnare quote di mercato significative.

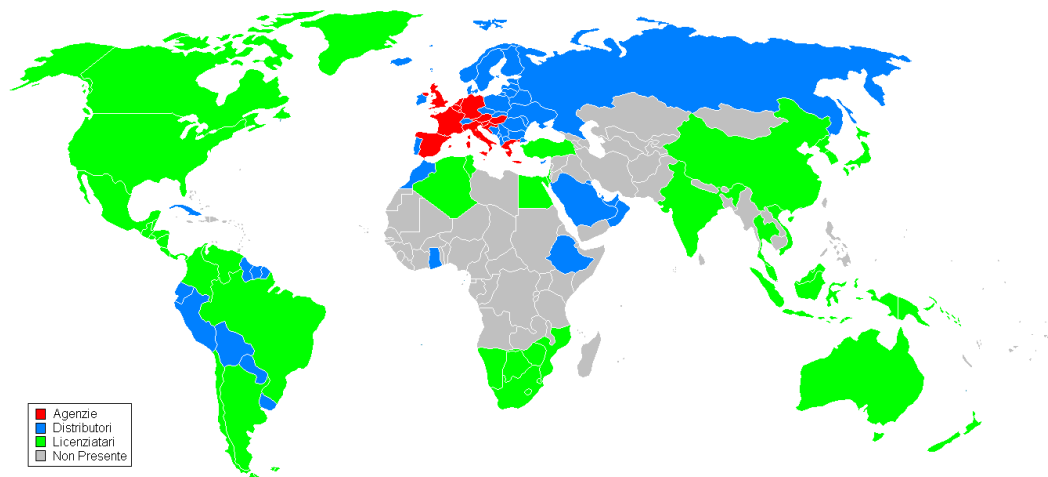


Figura 6. Nazioni di vendita di Lotto Sport Italia S.p.A. (Fonte: interna Lotto)

## 1.2 Obiettivi dello stage

Qui sotto riporto le caratteristiche dello stage così come mi è stato proposto:

*“Progetto nell’area della programmazione commerciale: analisi dei dati di vendita, sviluppo di modelli previsionali, analisi delle caratteristiche del prodotto calzatura sportiva, disegno dei prezzi, valutazione e implementazione delle strategie previsionali, analisi dei canali distributivi e della rete di vendita”.*

Poiché lo stage non prevedeva un progetto vero e proprio da sviluppare nell’arco temporale in cui sarei rimasto in azienda e spinto dall’interesse per l’analisi dei dati, sia nella forma prettamente descrittiva sia in quella previsionale, ho deciso di utilizzare parte del mio tempo in azienda nella creazione di una serie di analisi su alcuni fattori che potevano risultare utili ma che non erano ancora state eseguite.

Dopo una fase iniziale di progettazione delle analisi e una preparazione degli strumenti per lo svolgimento di queste ultime, si è passati alla realizzazione di alcuni studi di caso che riguardano principalmente le vendite.

Alcune possibili analisi da sviluppare nel contesto delle vendite di Lotto Sport Italia sono:

1. *Segmentazione dei clienti*: raggruppamento dei clienti, atto a evidenziare aspetti comuni anche tra clienti che, nel database aziendale, sono catalogati in modo diverso. All’interno dell’azienda vi è già una suddivisione dei clienti in base ad

alcune caratteristiche come, ad esempio, la quantità di prodotti ordinata e il tipo di distribuzione con cui operano nel mercato.

2. *Analisi della correlazione canonica*: ha come scopo quello di cercare delle interdipendenze tra gruppi di variabili quantitative, nella nostra analisi queste variabili sono rappresentate dalle quantità vendute dei modelli di calzature di Lotto Sport Italia.

I gruppi creati per effettuare queste analisi sono stati pensati in modo da avere già elementi degli stessi gruppi che sono concettualmente legati e che quindi i due gruppi formati appaiano il più possibile distante l'uno dall'altro.

3. *Analisi "a priori"*: consiste nella ricerca di eventuali relazioni che esistono tra i prodotti nel momento della vendita. Vengono evidenziati quindi i casi in cui l'acquisto di un determinato prodotto risulti collegato all'acquisto di un altro prodotto in listino. La conoscenza di queste relazioni risulta utile ai fini del marketing, in quanto sarà possibile creare delle apposite campagne per la promozione delle vendite di alcuni prodotti insieme ad altri.

Gli esempi qui riportati sono solo un gruppo ristretto di analisi possibili applicando il data mining al database aziendale, questo database presentando una quantità numerosissima di dati provenienti dalle diverse attività aziendali potrebbe essere utilizzato per una grande quantità di analisi nei campi più disparati come accade già in azienda attualmente.

## CAPITOLO II

### PRODUCT MARKETING E DATA MINING

#### 2.1 Concetti generali di marketing utili nel contesto aziendale

##### 2.1.1 *Il Prodotto*

La variabile fondamentale del marketing-mix è il prodotto.

Il concetto di marketing mix sintetizza le principali aree d'intervento che compongono il programma di marketing nei mercati e nei segmenti in cui l'impresa ha deciso di operare. Le leve su cui opera questo concetto sono: prodotto, prezzo, distribuzione e comunicazione.

Nell'ambito aziendale in cui lo stage si è svolto, ovvero l'ufficio di marketing prodotto, le principali leve su cui si opera sono il prezzo e il prodotto.

Il prodotto è una variabile complessa, rappresenta, infatti, la complessità che l'azienda deve affrontare nell'interfaccia del mercato.

Il prodotto quindi è visto come un insieme di attributi, tra questi possiamo evidenziare come primo grande sottoinsieme gli attributi tangibili, che sono le caratteristiche oggettive del bene-oggetto, mentre un secondo sottoinsieme è rappresentato dagli attributi intangibili, quali posso essere per esempio i servizi integrativi, che rappresentano un valore aggiunto al prodotto commercializzato. Infine vi sono gli attributi di prezzo e marca che rappresentano una categoria a parte rispetto agli attributi elencati in precedenza in quanto, questi due aspetti, si collocano a metà strada tra i due sottoinsiemi.

Proprio il prezzo è una leva fondamentale da analizzare ed interpretare. La corretta collocazione nella fascia di prezzo del nuovo prodotto risulterà fondamentale per il successo o il fallimento del prodotto nel mercato.

Il corretto posizionamento nelle varie fasce di prezzo in ambito aziendale viene fatta basandosi sullo studio della concorrenza. Vedere a che "punto prezzo" si trovano i principali concorrenti dell'azienda aiuta a capire quali sono i punti e la tipologia di prodotti su cui "spingere" e quali magari abbandonare in quanto poco redditizi.

Ho avuto modo di incontrare un esempio di quanto appena descritto nel momento in cui ho affrontato lo studio dei concorrenti del settore calzaturiero nella disciplina “Calcio”. Utilizzando una rappresentazione grafica con i vari modelli di Lotto e dei principali concorrenti sul mercato è emerso che la fascia di prezzo intermedia 60-80 euro presenta una quantità di prodotti molto bassa e la maggior parte delle linee sono concentrate su fasce alte (sopra i 100 euro) e sulla fascia cosiddetta “entry” dove sono presenti modelli base a un prezzo contenuto. Confrontando quest’analisi grafica con un’analisi compiuta qualche stagione prima si nota che vi è stata, nell’ultimo periodo, un’inversione di tendenza per quanto riguarda il posizionamento dei prodotti di quel mercato. Infatti, se fino a qualche stagione fa la fascia prezzo intermedia era popolata da una grande quantità di modelli differenti, adesso questa fascia è composta da circa un modello per ogni marchio. Questo evidenzia un cambiamento di abitudini da parte dei consumatori, in quanto, attraverso l’analisi delle vendite, si è visto che il consumatore se non era interessato ad un prodotto “top” orientava la sua scelta sulla fascia prezzo più bassa, cosa di cui le aziende hanno tenuto conto nel momento in cui sono andati a realizzare i modelli per le stagioni successive, riducendo il numero di prodotti nel segmento che stava perdendo quota di mercato a favore di quelle maggiormente prese in considerazione da parte dei consumatori.

| età    |  |  |  |  |  |
|--------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| > 101  |  |  |  |   |  |
| 81-100 |  |  |  |   |  |
| 71-80  |  |                                                                                   |                                                                                   |   |  |
| 61-70  |  |  |  |                                                                                     |                                                                                     |
| < 60   |  |  |  |   |  |

Figura 7. - Tabellone benchmark di analisi della concorrenza del settore Calcio FW09 (Fonte interna Lotto)

La figura è piuttosto rappresentativa e permette in modo semplice e chiaro di comprendere il fenomeno descritto a parole.

Questo tipo di analisi aiuta a far capire anche a chi non ha familiarità con il metodo di lavoro dell'ufficio marketing prodotto qual è la strategia che sta dietro alla progettazione e realizzazione delle nuove collezioni e strumenti di questo genere vengono utilizzati spesso nelle riunioni con le diverse componenti aziendali per presentare e motivare le decisioni prese sulle varie collezioni.

La figura rappresenta solo una sintesi finale di tutta la fase di studio e le considerazioni fatte sui dati di vendita delle stagioni passate per massimizzare le vendite nella stagione che ci si appresta ad affrontare nel mercato.

## **2.2 Cenni sul Product Marketing**

### **2.2.1 Le innovazioni di prodotto**

Un efficiente ed efficace processo di sviluppo di prodotti innovativi richiede il contributo del marketing, nello specifico possiamo dire che l'approccio manageriale deve contemplare l'azione di diverse funzioni aziendali e di una pluralità di apporti esterni.

I nuovi prodotti sono collocati dagli esperti di marketing in una sorta di piramide, dove alla base vi sono le più frequenti innovazioni di tipo incrementale, ovvero quelle innovazioni che aggiungono valore ad un prodotto già esistente, mentre al vertice di questa piramide troviamo le innovazioni radicali, poco frequenti e che comportano una vera e propria novità nel settore a cui appartengono o sono in ogni caso un prodotto nuovo che si discosta da quelli già esistenti.

Ad un livello intermedio troviamo infine quelle innovazioni che si collocano appunto tra questi due tipi di innovazioni, le quali non rappresentano delle innovazioni rivoluzionarie ma neanche totalmente marginali.

Le variabili che influenzano il successo di un determinato prodotto nel mercato sono sostanzialmente tre: l'appropriabilità dell'innovazione, la fluidità dell'innovazione e le risorse complementari.

L'appropriabilità si riferisce alle condizioni che consentono di proteggere l'impresa innovatrice dalle imitazioni, la fluidità indica quanto la prima versione del prodotto si differisce dalla versione finale che vede l'inserimento del prodotto nel mercato, infine le risorse complementari rappresentano un insieme di risorse che non riguardano direttamente lo sviluppo dell'innovazione ma che comunque sono necessarie per determinare il successo del prodotto nel mercato.

In quest'ultimo contesto si colloca il marketing, questo fattore risulta molto importante soprattutto per mantenere o incrementare la propria posizione e importanza nel settore di mercato in cui ci si è inseriti, per fare ciò è necessario studiare tutta una serie di tecniche di marketing al fine di raggiungere gli obiettivi prefissati dall'azienda. (Grandinetti, 2008)

### 2.2.2 Organizzazione del Product Management

Per quanto riguarda il modo di organizzare il processo innovativo all'interno dell'impresa esistono diversi metodi, tra le varie configurazioni possibili quella che viene maggiormente utilizzata nelle imprese è un modello che valorizza la "coralità" interfunzionale del processo. Si articola su due livelli: il comitato nuovi prodotti e il project team.

Il comitato nuovi prodotti è costituito da rappresentanti del marketing, della produzione, della ricerca & sviluppo e di altre aree funzionali dell'impresa, questo gruppo ha il compito di valutare i progetti relativi ai nuovi prodotti nei momenti critici del loro iter.

I project team o gruppi di lavoro hanno invece il compito di sviluppo vero e proprio dei nuovi prodotti. I compiti di queste due figure quindi non si sovrappongono, ma anzi, sono complementari tra loro permettendo il perseguimento degli obiettivi dell'impresa.

Non sempre però l'utilizzo di modelli organizzativi "puri" riescono a configurare risposte adeguate alla complessità aziendale, ecco quindi che si inseriscono in questo contesto figure nuove come i product manager e i responsabili di specifici mercati. Attraverso l'inserimento di queste nuove figure viene introdotto un meccanismo flessibile e di coordinamento organizzativo nella struttura aziendale. I Product manager appartengono nello specifico alla sezione marketing di un'azienda e rispondono a un direttore marketing il quale ha il compito di supervisionare il lavoro di più product manager.

Il product management è una struttura organizzativa a matrice in cui il product manager svolge la maggior parte del suo lavoro presso le diverse unità operative e a stretto contatto con i team interfunzionali, quasi come se egli dirigesse "un'azienda all'interno dell'azienda".

Nella figura vediamo come può essere composto il product management e quali sono le funzioni aziendali che sono comprese in questo settore molto importante per qualsiasi azienda in un mercato competitivo come quello attuale (Tieto, 2006).

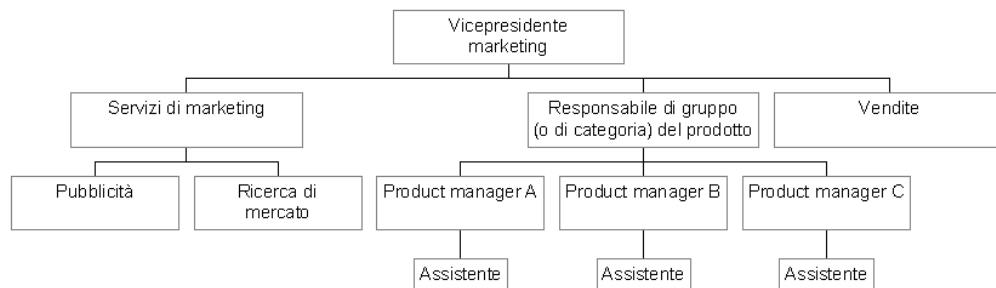


Figura 8. Struttura product management (Fonte: sito internet Eccellere)

### 2.2.3 *Il Product Manager*

Il ruolo del product manager è quello di coordinare risorse e persone per la progettazione, la realizzazione e il controllo di tutte le attività inerenti al marketing per quanto riguarda un prodotto o una linea di prodotti.

Esistono inoltre due diverse declinazioni all'interno del termine product manager, raccolti sotto questo nome possiamo, infatti, trovare sia i product manager veri e propri sia i cosiddetti merchandiser.

Vediamo ora quali sono le differenze e le azioni che queste due figure devono compiere all'interno delle realtà in cui operano.

Il *Product manager* che progetta una collezione che deve essere venduta su scala internazionale deve: ideare – spiegare – gestire dei prodotti che devono incontrare il gusto del più ampio numero di consumatori possibili, nei mercati in cui opera l'azienda, sempre rimanendo in linea con gli obiettivi aziendali (per esempio un'azienda che opera in un settore di nicchia studierà dei prodotti che vadano incontro ai bisogni e al gusto della ristretta clientela a cui il prodotto è rivolta); indirizzare i product manager “locali” nella scelta della selezione secondo quelle che sono le strategie e l'immagine che l'azienda vuole comunicare con la nuova collezione.

Il *Product Manager/Merchandiser* che seleziona i prodotti da una collezione internazionale deve: scegliere tutti i prodotti ritenuti idonei per il proprio mercato di competenza, seguendo le strategie che vengono stabilite dall'azienda di appartenenza ed eventualmente integrando la collezione con prodotti specifici creati appositamente per il proprio mercato qualora non siano disponibili nella collezione internazionale (Special Make Up).

Il product manager inoltre per svolgere il suo lavoro ha bisogno di una notevole quantità d'informazioni e deve quindi fare affidamento su numerose entità funzionali interne ed esterne all'azienda per raggiungere gli obiettivi prefissati. Il product manager si trova quindi a fare da collegamento tra numerose attività aziendali quali la produzione, il reparto vendite, la sezione di ricerca e sviluppo e molte altre (Tieto, 2006).

Nella figura vediamo i principali organi con cui il PM si trova più frequentemente a collaborare.



Figura 9. Product Manager (Fonte: sito internet Eccellere)

### 2.3 Cenni utili di Data Mining

Con il termine data mining vengono indicate una serie di attività di elaborazione in forma grafica o numerica di grandi raccolte o flussi continui di dati con lo scopo di estrarre informazione utile a chi detiene i dati stessi (Azzalini-Scarpa, 2004).

Questa disciplina, nata recentemente con lo sviluppo di nuove tecnologie in grado di memorizzare una quantità di dati sempre maggiore, si colloca come punto d'intersezione tra diverse aree scientifiche, in particolare: la statistica, l'intelligenza artificiale e la gestione dei data base.

Lo sviluppo di queste tecniche è stato necessario in quanto vengono applicati in situazioni in cui altri strumenti di analisi di dati più tradizionali non porterebbero a risultati soddisfacenti. I motivi principali per cui l'utilizzo di tecniche classiche sono sconsigliate sono sostanzialmente due: la grande numerosità campionaria e gli obiettivi dell'analisi.

La prima motivazione, ovvero la numerosità "campionaria", rappresenta una problematica da affrontare in quanto la crescita del numero di unità statistiche e la crescita del numero di caratteristiche rilevate sulle unità statistiche, detta dimensionalità dei dati, fanno sì che vi sia una crescita esponenziale del livello di complessità sia dal punto di vista computazionale (numerosità) che per complessità concettuale (dimensionalità). Si dovranno quindi applicare dei metodi che abbiano un basso carico computazionale, altrimenti i tempi di elaborazione risulterebbero eccessivamente lunghi, ovviamente la scelta di un metodo rispetto ad un altro dovrà essere fatta in base alle esigenze riscontrate nella trattazione di quel determinato problema.

La seconda problematica che ci si trova ad affrontare è quella della definizione degli obiettivi delle analisi, a differenza, infatti, delle tecniche statistiche classiche qui non andiamo a definire a priori un modello, anzi, molto spesso l'oggetto d'interesse viene individuato "scavando" tra i dati (Azzalini - Scarpa, 2004)

### 2.3.1 Ambiti applicativi del data mining

All'interno del "mondo" che definiamo come data mining esistono numerose tecniche per estrarre dai dati indicazioni a supporto dell'attività di business; la scelta di quale tecnica utilizzare dipende, essenzialmente, dal tipo di problema oggetto di studio e dal tipo di dati disponibili per l'analisi.

Tra le applicazioni di interesse nel nostro specifico caso vi sono:

- Customer Profiling (Market Bubble): applicazione di tecniche di clustering al fine di individuare gruppi omogenei in termini di comportamenti e di caratteristiche socio-demografiche; l'individuazione delle diverse tipologie di clienti permette di effettuare campagne di marketing diretto e di valutarne gli effetti, nonché di ottenere indicazioni su come modificare la propria offerta, e rende possibile monitorare nel tempo l'evoluzione della propria clientela e l'emergere di nuove tipologie.
- Scoring System (Predictive Modelling): sistema integrato di procedure volte alla costruzione di un modello che mette in relazione le caratteristiche di un sistema di individui (la lista dei potenziali clienti) con una variabile obiettivo. Tale variabile rappresenta l'adesione ad una proposta commerciale, mentre le variabili esplicative sono legate al comportamento passato degli individui nei confronti di analoghe iniziative e alle loro caratteristiche anagrafiche.  
L'obiettivo di tale tecnica consiste nell'assegnazione di un punteggio numerico (*score*) ad ogni individuo; tale *score* sarà collegato alla probabilità che ogni individuo risponda all'azione commerciale. In questo modo è possibile ordinare i clienti in base allo score e selezionare quelli a cui è associata una probabilità di adesione più elevata (Saccardi, 2003).

### 2.3.2 Raggruppamento in classi

Uno dei problemi applicativi più frequenti è quello di dover allocare un'unità a una categoria (o classe) tra  $K$  possibili alternative utilizzando le osservazioni relative a delle variabili rilevate su quella unità statistica.

La prima classe di metodi che affronteremo saranno i metodi di analisi interna, la caratteristica principale di queste tecniche è data dall'assenza di una variabile (quantitativa o qualitativa che sia) da interpretare come variabile risposta e da mettere in relazione con le altre. Tutte le variabili, infatti, sono poste sullo stesso piano, e solo lo studio complessivo di tutte le variabili contemporaneamente porterà a risultati utili e sensati.

L'obiettivo, che ci prefiggiamo quando utilizziamo metodi di raggruppamento, è quello di voler raggruppare le  $n$  unità disponibili in  $K$  gruppi ma, a differenza dei metodi di classificazione non abbiamo una variabile pre assegnata, e quindi non abbiamo una variabile di riferimento con cui mettere in relazione tutte le altre.

Generalmente, non abbiamo alcuna prefigurazione sul numero e sulla natura dei gruppi e cerchiamo un metodo per formarli a partire dalle variabili disponibili, e spesso si cerca una lettura interpretativa dei risultati così ottenuti.

Sarà necessario quindi avere dei profili in cui vengono individuati alcuni aspetti salienti dei vari gruppi, dove vorremmo allocare ad uno stesso gruppo quegli individui che sono tra loro più simili e, viceversa, a gruppi diversi individui dissimili.

Le possibili scelte per come quantificare questa 'vicinanza' sono tantissime e per riferirci a queste misure di lontananza tra unità statistiche si utilizza il termine "dissimilarità".

In generale la matrice  $X$  contiene misurazioni effettuate su tutte le scale prese in considerazione. Un giudizio complessivo circa la similarità tra gli oggetti di riga si ottiene dall'indice di similarità di Gower:  $c_{ij} = \sum_{k=1}^p c_{ij,k} / \sum_{k=1}^p \delta_{ij,k}$ .

dove  $c_{ij,k}$  è una misura di similarità fra le unità  $i$  e  $j$  calcolata con riferimento al  $k$ -esimo attributo, mentre  $\delta_{ij,k}$  è una variabile nominale che assume valore unitario se le unità possono essere confrontate con riferimento all'attributo  $k$  e zero altrimenti. In altre parole, essa denota l'ammissibilità del confronto.

I metodi di raggruppamento presentano al loro interno delle suddivisioni che permettono di scegliere la tecnica più adatta in base ai dati che abbiamo a disposizione e ai risultati che stiamo cercando. Le principali tecniche di nostro interesse sono i metodi non gerarchici e i metodi gerarchici, con un'ulteriore suddivisione di questi ultimi in metodi agglomerativi e metodi divisivi.

I metodi non gerarchici, tra i quali il più famoso è quello delle K-medie destinato a variabili continue, hanno come idea di base quella di individuare dei punti di aggregazione attorno ai quali costruire i gruppi attribuendo le osservazioni a quello dei centri individuati che è più prossimo. Questi centri peraltro non sono fissati, ma sono soggetti ad aggiornamento man mano che l'algoritmo procede nel calcolo. In questi metodi si cercherà quindi di minimizzare la dissimilarità complessiva all'interno dei gruppi, quest'obiettivo corrisponde anche a massimizzare la distanza tra i gruppi.

Esiste, però una forte limitazione di questo metodo, e consiste nel fatto che la sua applicazione richiede la scelta iniziale di diversi elementi. Un esempio di parametro da fissare a priori è il numero K di classi in cui suddividere le unità statistiche in quanto, spesso, non vi sono informazioni sulla struttura dei dati. Per ovviare a questo problema sono spesso utilizzati metodi che producono una struttura gerarchica dei dati e della loro organizzazione in gruppi, ciò viene fatto associando all'insieme dei punti una struttura ad albero binario.

La suddivisione tra metodi agglomerativi e metodi divisivi sta fondamentalmente nella situazione di partenza. Un metodo agglomerativo procede aggregando successivamente gruppi precedentemente formati e con bassa dissimilarità tra di loro, a partire da uno stato iniziale in cui  $K=n$ , ovvero ogni unità statistica rappresenta un gruppo a se stante e il metodo consiste nel raggruppare insieme quelle unità che sono più simili.

Un metodo divisivo segue una logica analoga alla precedente, ma partendo dall'estremità opposta, cioè formando prima un solo gruppo che include tutte le unità e poi procedendo per suddivisioni successive. Questi metodi sono stati peraltro meno esplorati, e meno utilizzati, rispetto a quelli agglomerativi.

### ***2.3.3 Associazioni tra variabili***

Altro possibile problema da affrontare nell'utilizzo del data mining nel contesto aziendale è quello dell'associazione tra variabili. Finora, infatti, le associazioni che cercavamo erano legate alle unità statistiche mentre qui sono le variabili che occupano il ruolo principale nell'analisi.

Con i dati a disposizione si cercheranno delle associazioni per vedere se vi sono delle variabili in qualche modo collegate tra loro, e in caso di risposta affermativa, si potranno prendere delle decisioni avendo ben presente queste associazioni.

Il più noto concetto nell'ambito della dipendenza tra due variabili è quello di correlazione. La matrice di correlazione rappresenta, infatti, uno strumento fondamentale per l'investigazione di strutture di dipendenza.

L'utilizzo di questo strumento presenta però alcuni svantaggi; la prima osservazione che si può fare è che la correlazione riflette unicamente le dipendenze di tipo lineare tra le variabili, altra fonte di difficoltà è legata alla corretta interpretazione dei valori della matrice risultante.

Uno dei possibili problemi d'interpretazione può essere data dalla presenza di altre variabili che interagiscono con le variabili prese in esame rendendo i risultati poco sensati e questo problema potrebbe essere risolto utilizzando la correlazione parziale tra le due variabili prese in esame eliminando, di fatto, l'effetto della variazione di altre variabili che interferiscono con i risultati.

La matrice di correlazione parziale risulterà quindi di più facile interpretazione rispetto a quella con le correlazioni marginali.

Nel caso in cui si operi con variabili categoriali, entrano in gioco le probabilità condizionate tra due eventi che permettono di definire il livello di associazione che lega le variabili prese in esame. Con queste variabili vengono utilizzate delle funzioni legame logaritmiche per la costruzione dei modelli, attraverso le quali i modelli sono in tutto e per tutto simili ai modelli log-lineari.

La rappresentazione dei risultati avviene attraverso un grafo, dove ad ogni variabile è associato un nodo del grafico e dove gli archi rappresentano un'associazione fra le variabili e l'assenza di un arco indica che le corrispondenti variabili nel modello sono indipendenti condizionatamente dal valore delle altre variabili.

Infine, attraverso l'analisi grafica, è possibile dare una prima lettura interpretativa dei dati che abbiamo elaborato e trarre delle semplici conclusioni sui livelli di associazione tra le diverse variabili.

#### 2.3.4 Analisi della correlazione canonica e Analisi delle corrispondenze

L'Analisi della correlazione canonica (o Analisi Canonica) ha lo scopo di studiare le relazioni di interdipendenza tra due o più gruppi di variabili manifeste *quantitative*. E' stata proposta da Hotelling (1938) in particolare per analizzare i legami tra le caratteristiche fisiologiche (variabili antropometriche) e quelle mentali (risposte a test psicologici) di un insieme di individui.

Nel dettaglio, si considera di avere osservato su un collettivo di unità statistiche, due gruppi di  $l$  e  $m$  variabili quantitative. Volendo studiare la correlazione tra le variabili di un gruppo e quelle dell'altro si tratterebbe di calcolare un numero pari a  $l \times m$  coefficienti di correlazione semplice.

Al fine di sintetizzare tali legami, la tecnica si propone di utilizzare un doppio sistema di variabili latenti che riproducano la correlazione tra i due gruppi di variabili manifeste, al netto di quella presente al loro interno.

In pratica si costruiscono due nuovi gruppi di variabili che sono incorrelati al loro interno e massimamente correlati tra di loro.

In questo caso, quindi, si parte da una tabella unità per variabili (le quali, anche in questo caso, sono *quantitative*), alla quale si possono, in seguito, aggiungere le variabili latenti così determinate, che vengono chiamate *Variabili canoniche*.

| Unità | $V_1$    | $V_2$    | ... | $V_l$    | $W_1$    | $W_2$    | ... | $W_m$    | ← | $C_1$    | $C_2$    | ... | $C_q$    | $G_1$    | $G_2$    | ... | $G_q$    |
|-------|----------|----------|-----|----------|----------|----------|-----|----------|---|----------|----------|-----|----------|----------|----------|-----|----------|
| 1     | $v_{11}$ | $v_{12}$ | ... | $v_{1l}$ | $w_{11}$ | $w_{12}$ | ... | $w_{1m}$ | ← | $c_{11}$ | $c_{12}$ | ... | $c_{1q}$ | $g_{11}$ | $g_{12}$ | ... | $g_{1q}$ |
| 2     | $v_{21}$ | $v_{22}$ | ... | $v_{2l}$ | $w_{21}$ | $w_{22}$ | ... | $w_{2m}$ | ← | $c_{21}$ | $c_{22}$ | ... | $c_{2q}$ | $g_{21}$ | $g_{22}$ | ... | $g_{2q}$ |
| ...   | ...      | ...      | ... | ...      | ...      | ...      | ... | ...      | ← | ...      | ...      | ... | ...      | ...      | ...      | ... | ...      |
| n     | $v_{n1}$ | $v_{n2}$ | ... | $v_{nl}$ | $w_{n1}$ | $w_{n2}$ | ... | $w_{nm}$ | ← | $c_{n1}$ | $c_{n2}$ | ... | $c_{nq}$ | $g_{n1}$ | $g_{n2}$ | ... | $g_{nq}$ |

Figura 10. – Tabelle generiche per analisi della correlazione canonica (Progetto AdaMSoft, 2007)

In ciascuno dei due gruppi le variabili canoniche vengono costruite come combinazione lineare delle variabili manifeste, con pesi espressi dal legame di

ciascuna variabile manifesta col fenomeno latente che la variabile canonica rappresenta. Queste si ottengono simultaneamente per l'uno e per l'altro gruppo, in modo tale da massimizzare congiuntamente la variabilità spiegata all'interno di ciascun gruppo e la correlazione tra i due gruppi.

La prima coppia di variabili canoniche è quella che presenta la maggiore correlazione; il legame espresso diminuisce man mano che si determinano le coppie successive. Pertanto anche in questo caso, come negli altri metodi di riduzione dati, si può trascurare una piccola parte dell'informazione e scegliere un numero di dimensioni inferiore a quello di partenza. Si può ottenere, in tale maniera, una consistente semplificazione dei dati di partenza. Essendo ogni variabile canonica una combinazione lineare delle variabili manifeste, la sua interpretazione si può avere sulla base del contributo alla sua costruzione da parte di ciascuna variabile manifesta.

L'analisi delle corrispondenze fu introdotta negli anni '70 da Benzecri per lo studio di tabelle di contingenza generate dall'incrocio di due variabili qualitative (Analisi delle Corrispondenze Semplice). In seguito fu estesa ai casi con un numero qualsiasi di variabili (Analisi delle Corrispondenze Multiple).

E' un'analisi di tipo fattoriale che ha come scopo quello di individuare dimensioni soggiacenti alla struttura dei dati, dimensioni intese a riassumere l'intreccio di relazioni di "interdipendenza" tra le variabili originarie.

Trasforma una tabella di contingenza in una rappresentazione grafica al fine di facilitare l'interpretazione dell'informazione contenuta nella tabella stessa.

Andremo ora ad illustrare il tipo di analisi di nostro interesse, ovvero le corrispondenze multiple.

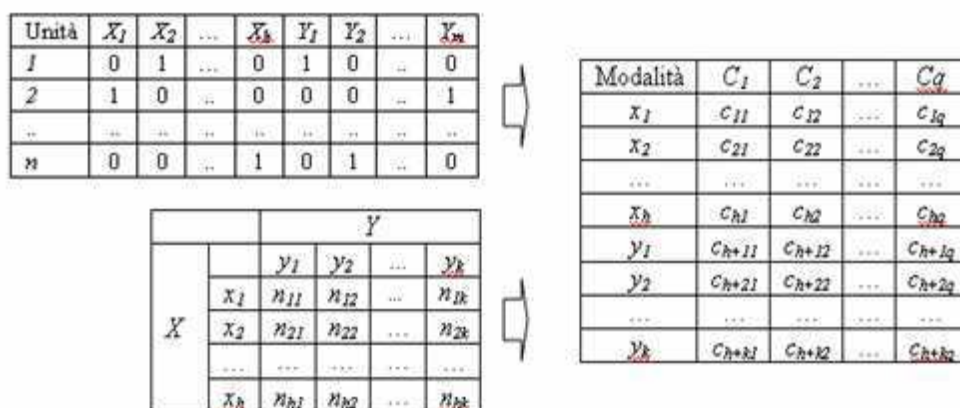


Figura 11. – tabelle generiche sull’analisi delle corrispondenze (Progetto AdaMSoft, 2007)

### Analisi delle corrispondenze multiple

Il punto di partenza è una tabella di contingenza “particolare”, nota come tavola di Burt, in cui le modalità riga coincidono con le modalità colonna.

### Tavola di Burt

|       | $M_1$    | $M_2$    | ... | $M_j$    | ... | $M_q$    |
|-------|----------|----------|-----|----------|-----|----------|
| $M_1$ | $n_{11}$ | $n_{12}$ | ... | $n_{1j}$ | ... | $n_{1q}$ |
| $M_2$ | $n_{21}$ | $n_{22}$ | ... | $n_{2j}$ | ... | $n_{2q}$ |
| ...   | ...      | ...      | ... | ...      | ... | ...      |
| $M_j$ | $n_{j1}$ | $n_{j2}$ | ... | $n_{jj}$ | ... | $n_{jq}$ |
| ...   | ...      | ...      | ... | ...      | ... | ...      |
| $M_q$ | $n_{q1}$ | $n_{q2}$ | ... | $n_{qi}$ | ... | $n_{qq}$ |

Figura 12. – Tavola di Burt caso Analisi Corrispondenze Multidimensionali (Ronconi)

Nell’ACM possiamo distinguere due tipologie di variabili:

- *attive*, cioè variabili che entrano direttamente nell’analisi concorrendo alla formazione degli assi fattoriali.
- *supplementari* o *illustrative*, cioè variabili di tipo “passivo” che sono escluse dalla fase di estrazione dei fattori ma si utilizzano successivamente considerando la loro posizione sugli assi fattoriali come ausilio per la loro interpretazione.

Come nell'Analisi delle corrispondenze semplici, viene utilizzata la metrica del  $\chi^2$  per valutare la dispersione dei profili, riga e colonna, rispetto al loro "centro di gravità".

L'inerzia totale è proporzionale a  $\chi^2$  ma non è esattamente uguale al rapporto tra il  $\chi^2$  e il numero di osservazioni. Essa è funzione del numero di modalità (q) e del numero di variabili attive (p):  $\text{Inerzia} = (q-p)/p$ .

Vengono estratti degli assi fattoriali, ortogonali tra loro, che spiegano ciascuno, in ordine decrescente, il massimo della variabilità della matrice dei dati (inerzia).

Il numero massimo di fattori estraibili è pari al numero di modalità (q) meno il numero di variabili attive (p).

I fattori da considerare si determinano in base alla loro rilevanza, ovvero alla quota di inerzia totale che spiegano.

Benzecri indica un limite, pari a: 1 diviso il numero di variabili attive ( $=1/p$ ), al di sotto del quale l'autovalore, e dunque il fattore ad esso associato, diviene insignificante.

Per interpretare il significato degli assi fattoriali, assieme alle coordinate fattoriali, si utilizzano i seguenti indicatori:

- la *massa*, ovvero la frequenza relativa della modalità rapportata al numero di variabili attive
- il *contributo assoluto* e il *contributo relativo (coseno quadrato)*

Talora può risultare interessante ripetere l'ACM dopo aver ripulito la matrice dei dati compattando le categorie che nella prima analisi non presentano contributi assoluti abbastanza elevati sui fattori considerati; in questo modo, infatti, si riduce l'inerzia globale ( $q$  diminuisce mentre  $p$  rimane fisso).

## CAPITOLO III

### ANALISI

#### 3.1 Organizzazione e illustrazione del dataset

Le analisi che condurremo nel proseguo di questa relazione sono state fatte basandosi su un dataset da me creato con le informazioni prese dal data warehouse aziendale.

Se volessimo descrivere in modo più generale possibile la composizione del dataset potremmo dire che esistono due grandi gruppi di variabili: il primo comprende tutte le caratteristiche di varia natura del cliente, si va dal nome del negozio, alla provincia in cui si trova, all'agente con cui tratta e così via; il secondo gruppo invece comprende delle variabili che rappresentano i modelli di scarpe su cui vogliamo andare ad effettuare gli studi con le relative quantità vendute.

Poiché l'universo dei dati disponibili nel database aziendale era molto ampio ho deciso di circoscrivere l'area geografica di analisi alla cosiddetta area EMEA (Europe, Middle East And Africa), quest'area rappresenta il mercato più importante per Lotto Sport Italia in quanto la presenza dell'azienda in questo mercato è molto forte rispetto a mercati più lontani quale, per esempio, potrebbe essere quello americano.

Altra limitazione sui dati raccolti riguarda i modelli di calzature su cui vengono effettuate le analisi. Nel dataset sono presenti infatti solo le calzature del settore performance (calcio, calcetto, tennis e running) in quanto l'obiettivo è quello di studiare il comportamento dei negozi che trattano prodotti sportivi.

Ultima considerazione da non trascurare sul dataset è l'assenza delle calzature da bambino. L'esclusione di tali variabili è stata resa necessaria principalmente da due fattori: il primo, di natura puramente tecnica, consiste nel problema che le scarpe da bambino presentano taglie diverse per indicare lo stesso modello, nel database aziendale queste taglie vengono gestite separatamente cosa che nello studio sul comportamento di acquisto dei negozi avrebbe portato presumibilmente a conclusioni diverse dalla realtà esistente; la seconda, suggeritami dal mio tutor

aziendale, riguarda il comportamento di acquisto dei negozi che è stato appreso dall'azienda con studi precedenti. Il comportamento di acquisto dei negozi che acquistano i prodotti da bambino risultano molto diversi in base alla tipologia di negozio, quindi l'analisi dei prodotti da bambino assieme a quelli da adulto avrebbe potuto creare delle differenziazioni già note all'azienda e quindi non utili ai fini di queste nuove analisi.

Vediamo ora brevemente qual è la struttura del dataset utilizzato per i nostri studi. Ogni riga del dataset rappresenta un cliente con le proprie caratteristiche e con la quantità acquistata di ogni singolo modello.

Nel dataset sono presenti 1913 clienti e per ogni cliente abbiamo raccolto 85 variabili.

Per concludere riporto la lista delle variabili considerate con una breve descrizione.

**Business Unit:** Area di vendita;

**Ragione sociale:** Nome Negozio;

**Titolare:** Nome del titolare del negozio (unica informazione disponibile sul titolare del negozio);

**Tipologia Cliente:** Suddivisione utilizzata da Lotto per suddividere i clienti;

**Tipologia distributiva:** suddivisione in base alla tipologia distributiva adottata (dipende dal paese);

**Gruppo Acquisto:** nome del gruppo di acquisto se il cliente ne fa parte;

**Team Business Cliente:** tipo di negozio del cliente (specialista, generico, ecc);

**Caratteristica Negozio:** tipologia di negozio;

**Città:** città in cui si trova il negozio;

**Provincia:** provincia in cui si trova il negozio;

**Nazione Vendita:** nazione dove opera il cliente;

**Priorità ordine:** priorità di invio merce al momento dell'ordine da parte del cliente;

**Tipo tassativa:** Tassativa a cui è sottoposto l'ordine del cliente;

**Agente:** Agente che si occupa di quel cliente;

**Provincia Agente:** provincia dell'agente (se in Italia);

**Località Agente:** città dell'agente;

**Q.tà ordinata:** quantità ordinata dal cliente di tutti i modelli;

**Importo lordo:** Ammontare in euro dell'ordinato;

**Prezzo medio:** prezzo medio tra tutti i prodotti acquistati dal cliente;

**% sconto:** percentuale di sconto sulla merce ordinata;

**% annulli:** percentuale di prodotti annullati sul totale ordinato;

**Count Modelli:** numero di modelli totali che il cliente ordina;

Oltre a queste variabili identificative del cliente sono presenti nel dataset tutti i modelli del settore performance di Calcio, Calcetto, Running e Tennis delle categorie Uomo e Donna. In totale il numero di modelli presenti è 63, e per ogni modello è stata presa la quantità ordinata nella stagione che stiamo esaminando.

## 3.2 Raggruppamento dei clienti di Lotto Sport Italia S.p.A.

### 3.2.1 Applicazione dei metodi di raggruppamento

Per effettuare questo tipo di analisi abbiamo utilizzato il programma statistico “R”. All’interno di questo programma esistono varie librerie per eseguire dei raggruppamenti partendo dal dataset di interesse. Abbiamo utilizzato per questi studi la libreria “cluster” la quale, al suo interno, presenta una serie di algoritmi gerarchici e non gerarchici (cap. 2, par. 2.3.2) che utilizzeremo per tracciare dei profili dei clienti presi in esame.

Dallo studio preliminare del dataset si è visto che alcune variabili risultavano di scarso interesse in questa tipologia di analisi.

Le variabili che sono state eliminate prima di iniziare lo studio sono: *Business unit*, in quanto avendo considerato solo l’area EMEA tutte le righe presentavano lo stesso valore di quella variabile, successivamente sono state eliminate *Ragione Sociale* e *Titolare*, in quanto, poiché il raggruppamento si prefigge l’obiettivo di tracciare dei gruppi con un profilo ben definito, queste due variabili che assumono numerose modalità non portavano dell’informazione utile ai fini dell’aggregazione dei dati.

Dopo aver eliminato queste variabili abbiamo proceduto abbiamo quindi proceduto con la creazione della matrice di dissimilarità, questo passaggio ci è servito in quanto, costituirà la base per l’applicazione dei vari modelli di raggruppamento, i quali devono avere come dati in ingresso una matrice di dissimilarità e non semplicemente un dataset. La matrice di dissimilarità calcola la distanza tra le variabili utilizzando una distanza che è diversa in base alle tipologie di dati che si hanno, nel nostro caso siamo in presenza sia di dati quantitativi sia di dati qualitativi quindi, come distanza, abbiamo utilizzato la distanza di Gower che ci permette di calcolare distanze avendo entrambe le tipologie i dati.

Nella prima parte abbiamo applicato due modelli gerarchici già implementati nella libreria cluster, che sono il metodo Agglomerative Nesting e il metodo Divisive Analysis Clustering.

A questo punto abbiamo applicato i metodi per il raggruppamento gerarchico ottenendo, dopo una serie di prove per stabilire quale fosse la funzione legame migliore, questi due alberi binari come risultato.

Il legame migliore è stato scelto dopo una serie di prove utilizzando i diversi legami disponibili per questi metodi.

Per l'analisi agglomerativa il legame utilizzato è il legame ward, basato sulla minimizzazione della devianza entro i gruppi che è uguale a quando tutti i casi sono separati, ed è massima quando essi appartengono tutti a un gruppo unico.

Per l'analisi divisiva invece, è stato utilizzato il metodo manhattan, calcolato attraverso questa formula  $d(i,k) = \sum_{j=1}^d |x_{ij} - x_{kj}|$ .

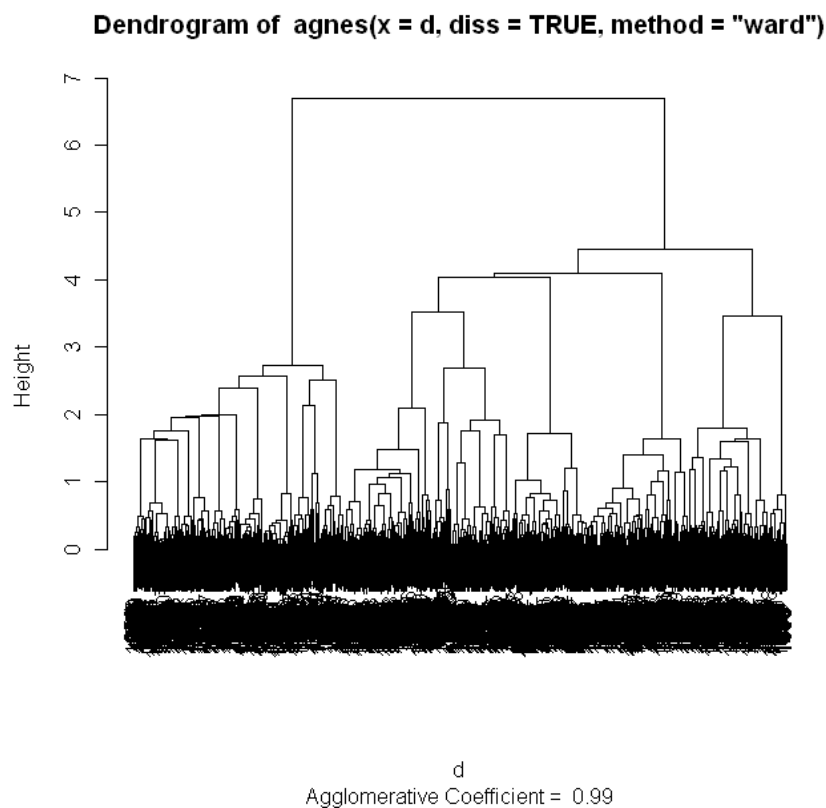


Figura 13. Albero metodo “agnes” con legame “ward”

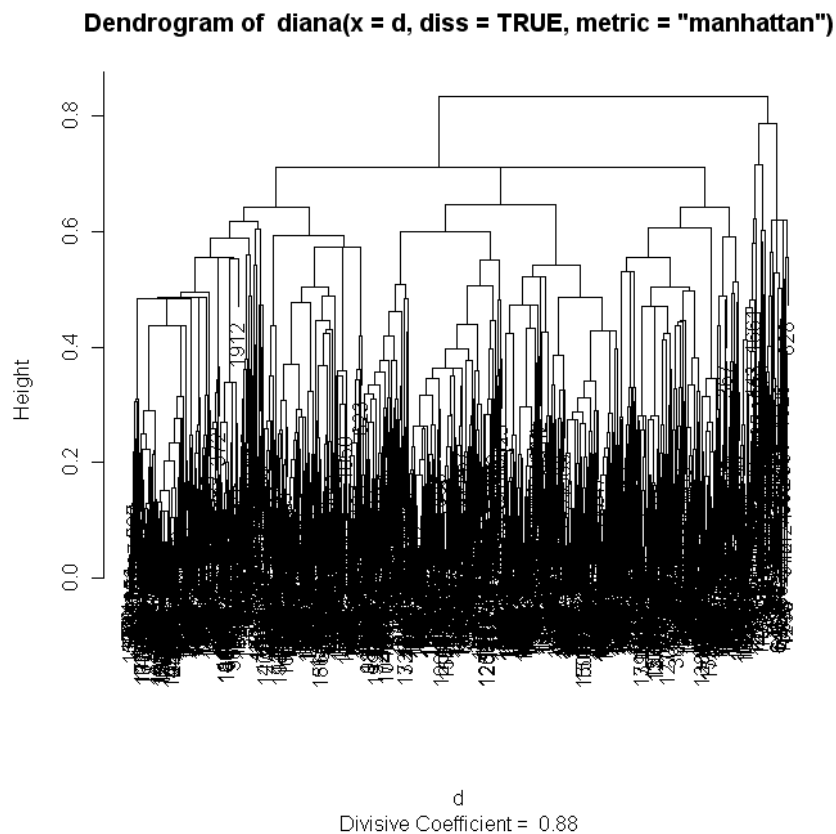
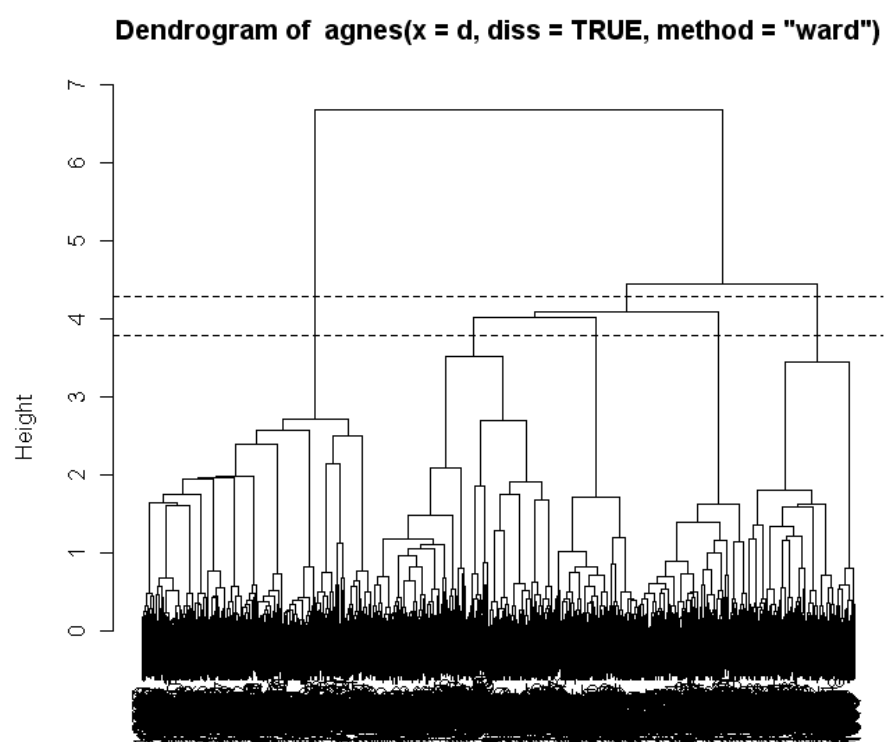


Figura 14. Albero metodo “diana” legame “Manhattan”

Il confronto tra l’albero ottenuto dall’algoritmo agglomerativo e l’algoritmo divisivo nonostante siano presenti delle differenze piuttosto evidenti per quanto riguarda il peso dei rami ci porta a concludere che il numero di gruppi più ragionevole sia 3. Poiché la lunghezza dei rami è proporzionale alla caduta di dissimilarità passando da  $K$  a  $K+1$  gruppi vediamo che i rami con lunghezza maggiore, dove in genere è sensato andare a potare, si trovano in entrambi i grafici al livello con  $K=3$ , anche se una potatura al livello  $K=5$  o  $K=6$  potrebbero essere comunque una buona suddivisione.

I grafici qui sotto rappresentano i possibili tagli che si potrebbero quindi effettuare per la formazione dei gruppi di clienti.



Agglomerative Coefficient = 0.99

Figura 15. Albero potato ai livelli K=3 e K=5

**Dendrogram of diana(x = d, diss = TRUE, metric = "manhattan")**

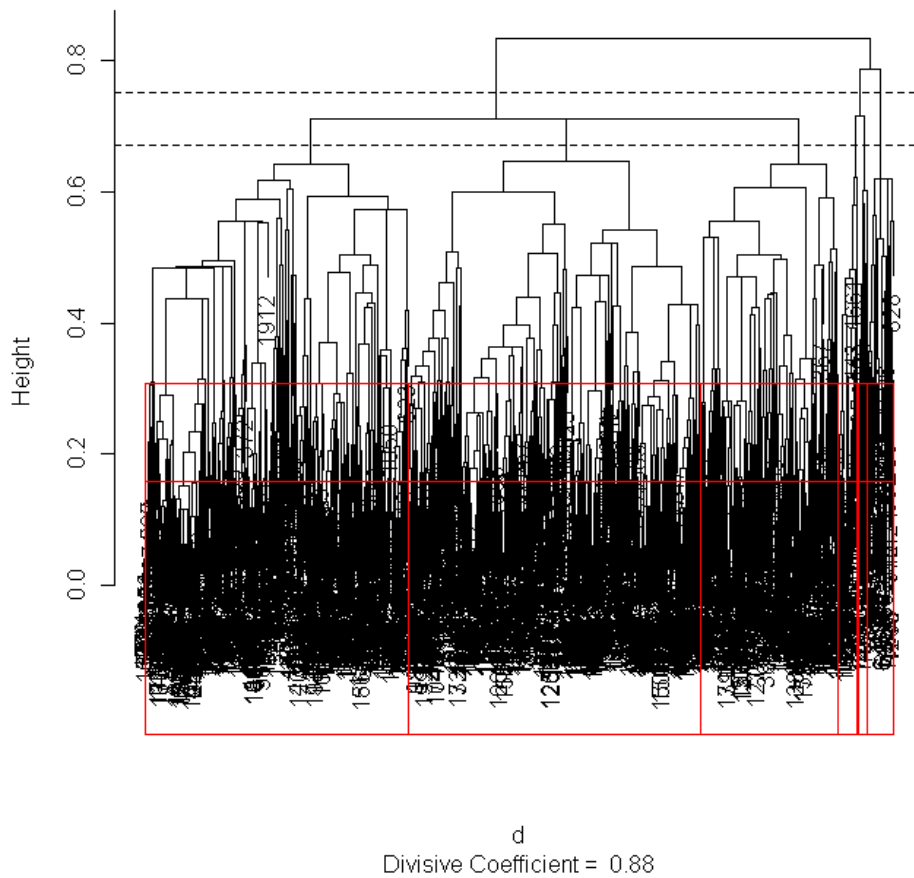


Figura 16. Albero potato ai livelli K=3 e K=6

Andando a potare il primo albero ai livelli 3 e 5 otteniamo queste numerosità dei gruppi.

| Gruppo | 1   | 2    | 3   |
|--------|-----|------|-----|
| num.   | 615 | 1009 | 289 |

| Gruppo | 1   | 2   | 3   | 4   | 5   |
|--------|-----|-----|-----|-----|-----|
| num.   | 615 | 491 | 292 | 289 | 226 |

Mentre andando a potare il secondo albero ai livelli 3 e sei otteniamo:

| Gruppo | 1    | 2  | 3  |
|--------|------|----|----|
| num.   | 1771 | 77 | 65 |

| Gruppo | 1   | 2   | 3   | 4  | 5  | 6  |
|--------|-----|-----|-----|----|----|----|
| num.   | 673 | 748 | 350 | 51 | 65 | 26 |

Come possiamo vedere dalle numerosità ottenute con i due metodi con il metodo divisivo i gruppi formati risultano pesantemente sbilanciati presentando una numerosità molto elevata nei primi gruppi e veramente piccola in quelli rimanenti. Con il proseguo delle analisi andremo ad utilizzare la suddivisione ottenuta con il metodo agglomerativo, ovvero il metodo “agnes”.

Come già detto in precedenza il nostro dataset presenta al suo interno variabili sia quantitative che qualitative e questo non ci permette di utilizzare delle funzioni incluse nella libreria “cluster” per determinare i profili dei gruppi.

Per ovviare a questo problema e giungere ugualmente a delle conclusioni su come sono stati formati i gruppi andremo a svolgere delle analisi descrittive su alcune variabili più significative per verificare se ci sono differenze su quelle variabili tra i clienti di un gruppo e quelli di un altro gruppo.

### ***Analisi descrittiva dei risultati***

Giunti a questo punto risulta utile e sensato andare a fare delle considerazioni su quanto emerge dall’applicazione delle tecniche fin qui adottate per il raggruppamento dei clienti.

Dal primo raggruppamento, con livello di gruppi  $K=3$ , emergono una serie di risultati più o meno significativi. Per prima cosa osserviamo che le numerosità dei gruppi non sono omogenee, abbiamo infatti nel primo gruppo 615, 1009 nel secondo e solamente 289 nel terzo.

Riportiamo ora una serie di tabelle di confronto per i dati utilizzando il raggruppamento per il momento in *tre classi*.

### **Tipologia Distributiva**

|          | Agenzia | Distributore | Licenziatario |
|----------|---------|--------------|---------------|
| Gruppo 1 | 615     | 0            | 0             |
| Gruppo 2 | 972     | 35           | 2             |
| Gruppo 3 | 289     | 0            | 0             |

### Priorità Ordine

|          | Alta | Media | Bassa |
|----------|------|-------|-------|
| Gruppo 1 | 95   | 395   | 125   |
| Gruppo 2 | 123  | 728   | 158   |
| Gruppo 3 | 23   | 149   | 114   |

### Nazioni di vendita

|          | Italia | Francia | Germania | Spagna | Grecia | Olanda | Belgio | Austria | Other | TOTALE      |
|----------|--------|---------|----------|--------|--------|--------|--------|---------|-------|-------------|
| Gruppo 1 | 615    | 0       | 0        | 0      | 0      | 0      | 0      | 0       | 0     | <b>615</b>  |
| Gruppo 2 | 257    | 300     | 226      |        | 58     | 34     | 32     |         | 102   | <b>1009</b> |
| Gruppo 3 | 0      | 0       | 0        | 236    | 0      | 0      | 0      | 53      | 0     | <b>289</b>  |

### Valori Medi

|          | Qtà.<br>Ordinata | Importo | Prezzo<br>Medio |
|----------|------------------|---------|-----------------|
| Gruppo 1 | 169,2            | 4217,2  | 25,87           |
| Gruppo 2 | 268,3            | 7213    | 28,96           |
| Gruppo 3 | 57,74            | 1600    | 29,26           |

### 3.2.2 Conclusioni

Le analisi svolte ci portano a dire che è possibile e sensato raggruppare i clienti in tre o cinque gruppi. Vediamo ora, analizzando più approfonditamente i risultati dei raggruppamenti, quale dei due risulta migliore per i nostri obiettivi.

Dalla sintesi delle informazioni riportate qui sopra iniziamo col tracciare dei profili dei tre gruppi.

Nel primo gruppo quello che risalta maggiormente sono la tipologia distributiva e la nazionalità. Vediamo che in questo gruppo sono raccolti tutti negozi italiani dove vi è una tipologia distributiva gestita attraverso agenzie.

Questa seconda informazione però risulta inutile ai fini di tracciare un profilo del gruppo in quanto la nazione Italia è gestita attraverso una rete di agenzie come unica tipologia distributiva quindi è normale questa concomitanza. Più interessante è notare che la quantità media ordinata si attesta sulle 170 paia con un prezzo medio attorno ai 26 euro.

Il secondo gruppo presenta la maggioranza di negozi gestiti da agenzie, cosa che possiamo dedurre dal fatto che le principali nazioni di vendita di questo gruppo

sono rappresentate da Italia, Francia e Germania, tutte e tre gestite tramite agenzie.

Interessante anche qui vedere prezzo medio e quantità ordinate, ci troviamo di fronte ad un gruppo che acquista in media di più del primo gruppo, 268 paia a fronte di un prezzo medio dei modelli che si aggira attorno ai 29€.

Infine l'ultimo gruppo composto in totale da 289 clienti raggruppa dei clienti di nazionalità spagnola e austriaca. Il prezzo medio qui supera di poco i 29€ e le quantità si riducono notevolmente rispetto al gruppo 2 stabilizzandosi attorno alle 60 paia.

Vi è quindi una tendenza ad acquistare prodotti di fascia superiore ma in quantità minore, presupponiamo quindi di trovarci di fronte a dei negozi specialisti o che comunque trattano con una porzione piccola di mercato.

Queste conclusioni è possibile trarle poiché il prezzo con cui Lotto vende ai propri clienti è lo stesso in tutti i paesi, mentre il prezzo al consumatore varierà in base al paese di destinazione.

Poiché il secondo gruppo presenta al suo interno, clienti di diverse nazionalità e tipologie distributive, proviamo a confrontare i risultati fin qui ottenuti con quelli di una suddivisione in 5 gruppi.

Confrontando i risultati delle due suddivisioni vediamo come le principali differenze interessino il gruppo 2, il quale, in questo nuovo raggruppamento risulterà diviso in tre, formando gruppi suddivisi principalmente per nazionalità. Il gruppo misto formato dall'Italia e da una presenza importante di clienti di altre nazionalità, il gruppo che comprende i clienti francesi e infine un gruppo con solamente clienti di nazionalità tedesca.

Con questa suddivisione è possibile notare come, nonostante il comportamento di acquisto differenti, essi possano essere raggruppati in un unico gruppo che li contenga tutte e tre.

|                             | Prezzo Medio | Quantità Ordinata |
|-----------------------------|--------------|-------------------|
| Gruppo Italia e altri paesi | 29,36        | 474,5             |
| Gruppo Francia              | 24,31        | 62,86             |
| Gruppo Germania             | 34,12        | 85,59             |

Vediamo come vi siano differenze sostanziali tra i prezzi medi e le quantità ordinate, tuttavia l'algoritmo sulla base di tutte le variabili prese in considerazione passando da un raggruppamento da 5 ad uno con tre 3 gruppi considera tutti questi clienti come simili.

Particolare importante da notare sul gruppo francese è che il 52% dei negozi fanno parte del gruppo di acquisto Intersport France.

In definitiva nonostante il raggruppamento con tre gruppi sia sufficientemente esplicativo e contraddistingua bene i gruppi la soluzione con cinque gruppi permette di cogliere delle particolarità che altrimenti non ci sarebbe stato permesso di individuare.

Manterremo quindi una suddivisione in cinque gruppi e riassumiamo ora le caratteristiche principali dei gruppi così formati:

Gruppo 1: clienti italiani che tendono all'acquisto di prodotti in tutte le fasce di prezzo, ma in quantità non molto elevate. Abbiamo quindi di fronte negozi con un mercato limitato ma che comunque necessita di un buon assortimento per poterlo soddisfare.

Gruppo 2: i clienti che fanno parte di questo gruppo sono di diverse nazionalità, sono accomunati dalle quantità ordinate piuttosto elevate e dal prezzo medio piuttosto alto, si tratta dunque di negozi che ricoprono un ruolo importante nelle vendite di Lotto.

Gruppo 3: il terzo gruppo è composto da clienti in prevalenza francesi, legati per più della metà di essi, all'acquisto attraverso Intersport France. Prezzo medio e quantità ordinate sono piuttosto bassi facendoci intuire che i prodotti principalmente acquistati siano quelli di fascia medio - bassa e in piccole quantità. In questo caso azioni di promozione dovrebbero avere come soggetto il gruppo di acquisto più che il singolo cliente in modo da incentivare le vendite attraverso promozioni o stock di modelli a prezzi convenienti magari di fascia leggermente più alta rispetto al target attuale.

Gruppo 4: siamo di fronte a clienti tedeschi che trattano quantità di prodotti piccole ma con un prezzo elevato. I clienti quindi cercano prodotti di livello alto in quanto lo proporranno presumibilmente ad una clientela disposta a spendere per

avere prodotti di alto livello. Di fatto questo gruppo raggruppa una maggioranza di negozi specialisti o con un team business ben definito a differenza degli altri paesi dove troviamo negozi più generici e che quindi cercano di soddisfare un range più grande di clientela.

Gruppo 5: questo gruppo risulta formato principalmente da clienti di nazionalità spagnola, il prezzo medio pur non essendo ai livelli del gruppo tedesco resta piuttosto alto e la quantità media ordinata ci può far pensare di trovarci di fronte a clienti di medio - piccole dimensioni che propongono ai loro clienti prodotti di fascia medio - alta.

La figura che segue qui sotto propone un grafico dei prezzi medi suddivisi per gruppi. Vediamo come il prezzo medio del quinto gruppo risulti sensibilmente più alto rispetto agli altri.

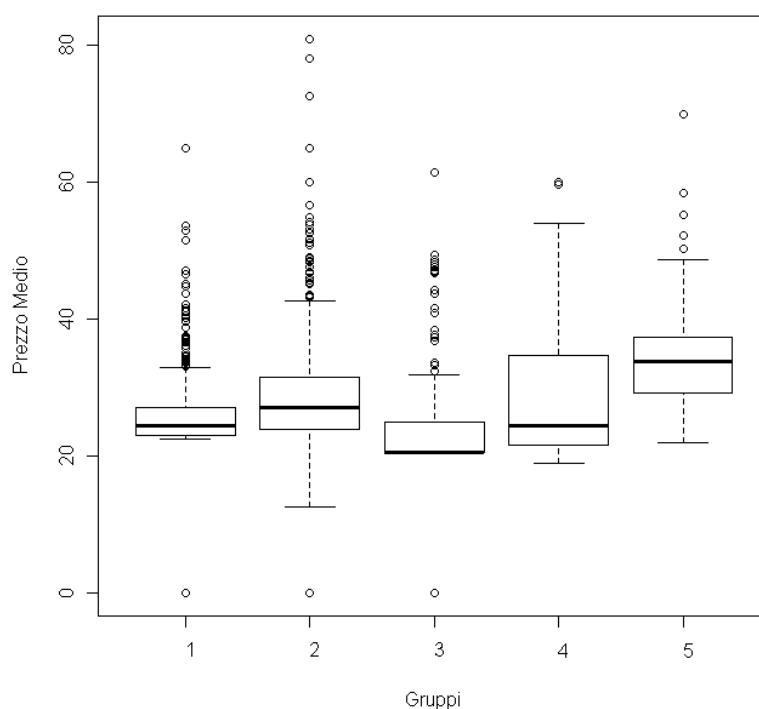


Figura 17. – Boxplot dei prezzi medi dei gruppi

Un ulteriore analisi, che potrebbe risultare utile all'azienda, è vedere se vi sono differenze di preferenza per quanto riguarda i modelli.

Abbiamo quindi analizzato i vari modelli presi singolarmente osservando come si distribuivano le quantità vendute nei vari gruppi.

La prima conclusione che possiamo trarre dall'analisi dei grafici è che il gruppo 3 acquista tendenzialmente un numero basso di modelli.

Questo fatto, con buona probabilità, è legata al fatto che molti clienti sono vincolati al gruppo di acquisto Intersport France. Questo fa sì che l'assortimento dei prodotti non sia molto vasto in quanto, la scelta dei prodotti, è fatta a monte dal gruppo di acquisto.

Un secondo aspetto evidenziato dall'analisi dei grafici sui modelli riguarda il gruppo 2, notiamo infatti che rispetto agli altri gruppi ha tendenzialmente una variabilità maggiore sulle quantità acquistate dai vari clienti. Le quantità acquistate dai singoli clienti variano molto, a differenza degli altri gruppi dove le quantità risultano più concentrate attorno al valore medio. Inoltre, questo gruppo presenta una quantità media di prodotti acquistati che è sempre maggiore rispetto a tutti gli altri gruppi.

Tra tutti i grafici dei modelli che abbiamo analizzato riportiamo ora quelli più caratteristici.

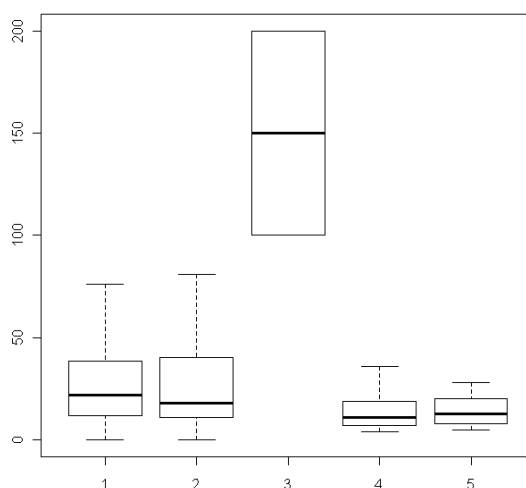


Figura 18. – Grafico delle quantità vendute del modello “Venice II”

La quantità media risulta sensibilmente diversa, vediamo quindi come il prodotto Venice II venga tenuto in considerazione in modo diverso in questo gruppo rispetto agli altri. La quantità media prossima alle 150 paia è sensibilmente differente rispetto alle 25 paia degli altri gruppi. Vediamo quindi come questo prodotto risulti particolarmente attraente nel mercato francese rispetto agli altri.

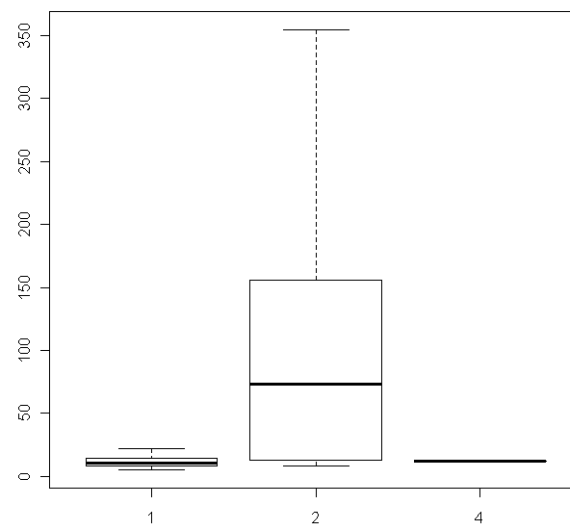


Figura 19. - Grafico delle quantità vendute del modello “Pabellon Lth TF”

L’assenza dei gruppi 3 e 5 dal grafico indicano una assenza di prodotti acquistati da parte di quel gruppo. Inoltre, anche i gruppi 1 e 4 presentano quantità molto basse. L’unico gruppo in cui questo modello ottiene dei buoni risultati di vendita è il gruppo 2.

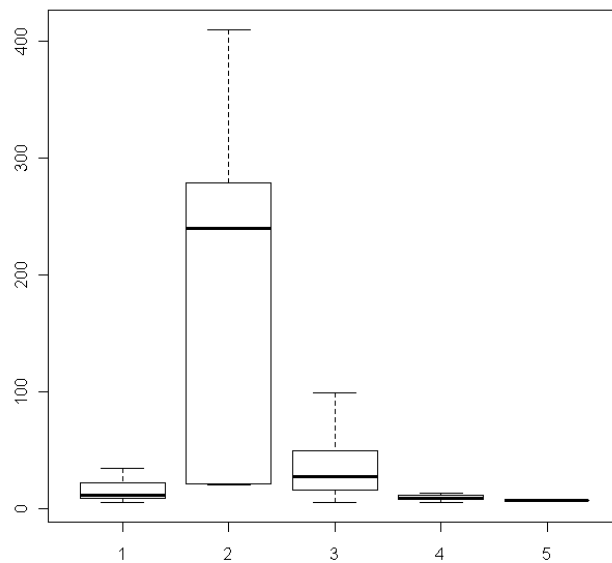


Figura 20. - Grafico delle quantità vendute del modello “Zhero Vertigo Free TF”

In questo caso, nonostante il prodotto sia acquistato da clienti di tutti i gruppi, vediamo come le quantità del secondo gruppo siano nettamente maggiori rispetto a tutti gli altri. Questo è l'esempio più lampante rispetto a quanto detto prima riguardo la variabilità presente nel gruppo 2 e alla media di prodotti acquistati che è più alta rispetto a tutti gli altri.

### 3.3 Analisi della correlazione canonica

#### 3.3.1 Applicazione dell'analisi della correlazione canonica

L'analisi della correlazione canonica ha come scopo quello di cercare delle interdipendenze tra gruppi di variabili quantitative. Nell'analisi che ci apprestiamo a fare queste variabili sono rappresentate dai modelli di calzature performance di Lotto Sport Italia.

Nel dataset utilizzato per effettuare questa analisi sono presenti i modelli uomo e donna mentre non sono stati inclusi i modelli da bambino in quanto, questa categoria, merita una trattazione a se poiché presenta delle caratteristiche di vendita tali da non poter essere considerate nello stesso modo delle calzature da adulto.

Abbiamo utilizzato dei metodi per ricercare la correlazione tra i gruppi di variabili utilizzando una libreria già implementata per l'ambiente R, ovvero la libreria "CCA".

L'utilizzo di tali metodi prevede una suddivisione a priori in due gruppi delle variabili che andremo a considerare. È stato deciso di attuare due suddivisioni distinte e successivamente di studiare questi due casi separatamente per cercare di trarre maggior informazione possibile dai risultati dell'applicazione di tali metodi. Le due suddivisioni che abbiamo adottato sono state: la prima, più classica, è stata la suddivisione tra calzature maschili e femminili, si cercheranno quindi eventuali interdipendenze tra questi due gruppi per vedere se calzature destinate a sessi diversi risultano in qualche modo collegate; la seconda è stata fatta sulla base dell'organizzazione aziendale dei prodotti, sono stati raggruppati insieme i settori calcio e calcetto e messi in relazione con i settori tennis e running.

Dall'andamento stagionale delle vendite è noto a priori che le calzature del primo gruppo sono vendute in quantità maggiore nella stagione invernale mentre l'altro raggruppamento ha un notevole incremento di venduto nel periodo estivo.

La suddivisione è stata scelta appunto per evitare di trovare correlazioni già note all'azienda su questi modelli.

### 3.3.2 Analisi uomo-donna

Dopo aver suddiviso le variabili del dataset nei due gruppi di interesse abbiamo proceduto ad applicare i metodi per l'analisi della correlazione canonica tra i gruppi.

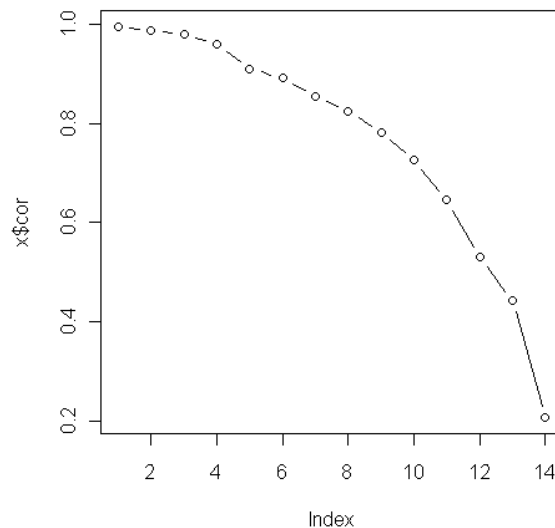


Figura 21. Grafico della correlazione spiegata

Come è possibile vedere dal grafico della correlazione spiegata sembra non ci sia una correlazione particolarmente significativa tra i due gruppi.

Cosa che ci viene confermata da questi altri grafici che rappresentano in maniera differente i due gruppi di variabili applicati ad un grafico radiale e la distribuzione degli individui, ovvero le righe che rappresentano in sostanza un cliente.

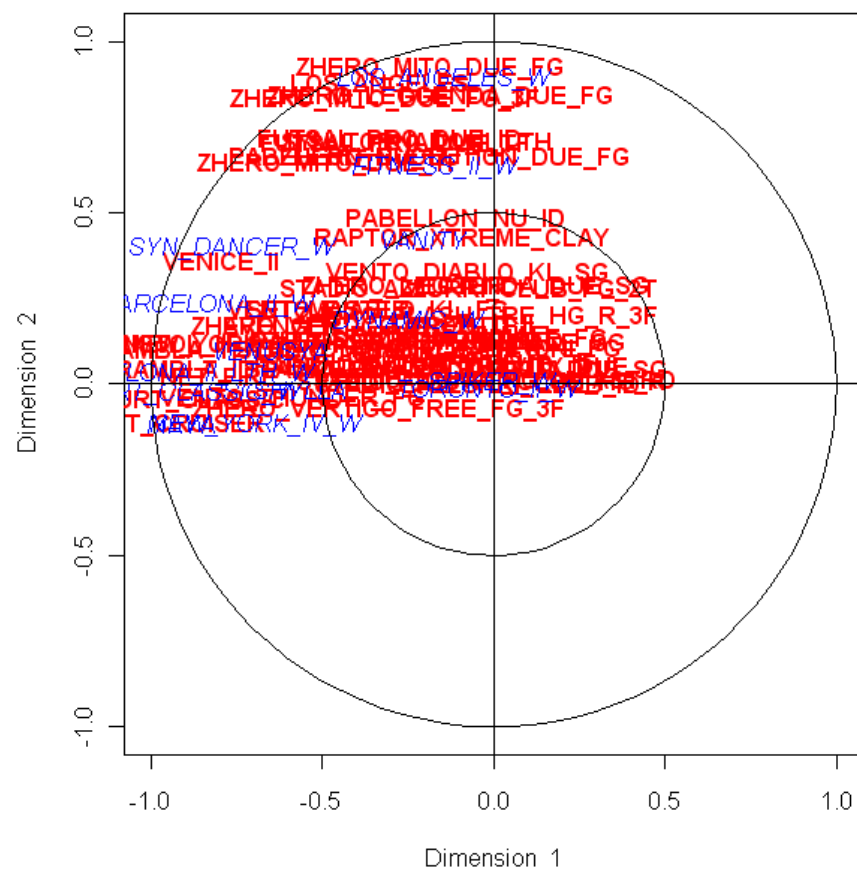


Figura 22. Grafico delle variabili con la suddivisione uomo-donna delle calzature

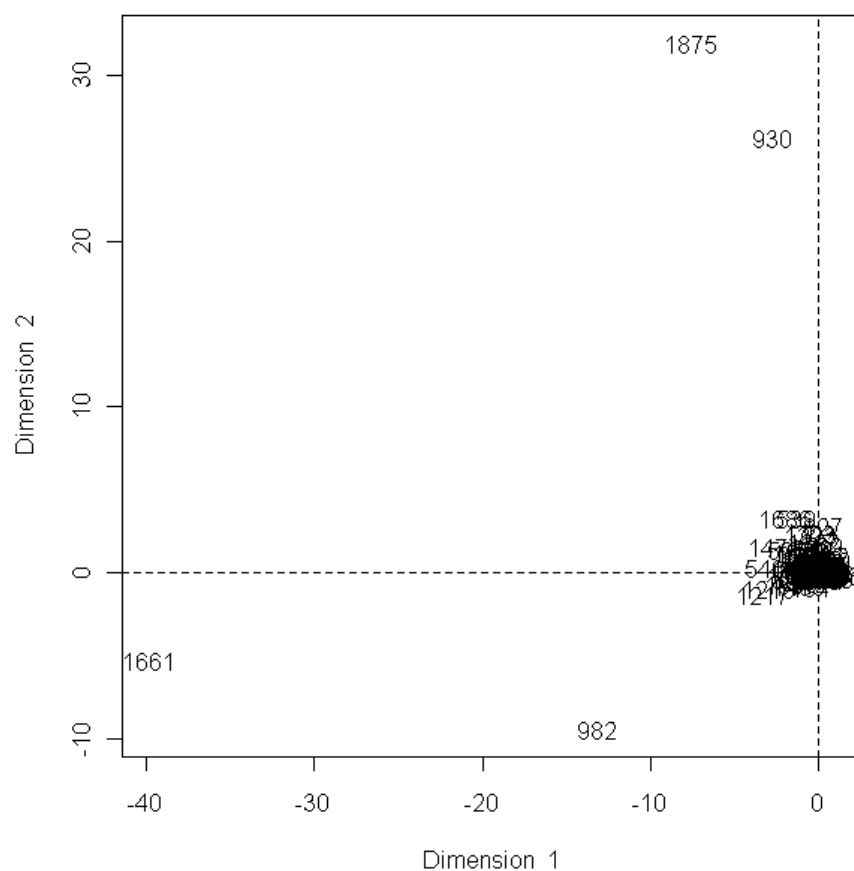


Figura 23. Grafico degli individui con la suddivisione uomo-donna delle calzature

Dalla figura 22 è possibile notare come le variabili siano distribuite su due assi, il primo gruppo lo vediamo molto vicino all'asse della Dimensione 1 nella sua parte positiva mentre il secondo gruppo lo troviamo nella parte negativa dell'asse della Dimensione 2.

I nomi scritti in blu rappresentano le calzature femminili e, come possiamo osservare dal grafico, risultano essere un numero molto inferiore rispetto alle calzature maschili.

Sostanzialmente vi sono tre modelli che si comportano in maniera differente rispetto agli altri articoli femminili e questi sono: Los Angeles W, Fitness II W e Vanity.

Poiché la suddivisione tra calzature uomo e donna viene fatta solo nei settori running e tennis andiamo ad analizzare un grafico che metta in relazione solamente queste due discipline, escludendo il calcio e il calcetto.

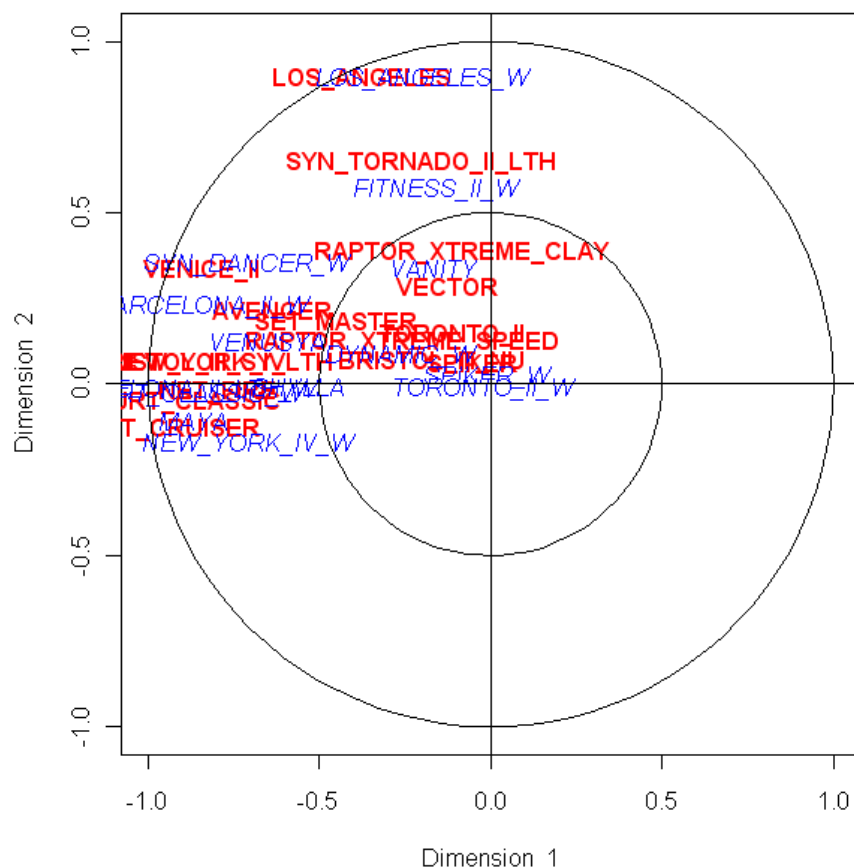


Figura 24. Grafico delle variabili con la suddivisione uomo e donna considerando solo le discipline tennis e running.

Analizzando il posizionamento dei modelli che si dispongono sull'asse verticale osserviamo che questi risultano essere i modelli top di gamma. Come era facile prevedere quindi il comportamento dei prodotti più tecnici e con un prezzo maggiore hanno un comportamento che si differenzia in modo sostanziale dagli altri prodotti.

### 3.3.3 Analisi Calcio, Calcetto - Tennis, Running

Abbiamo qui suddiviso le calzature secondo la disciplina ed il risultato del grafico della correlazione è questo.

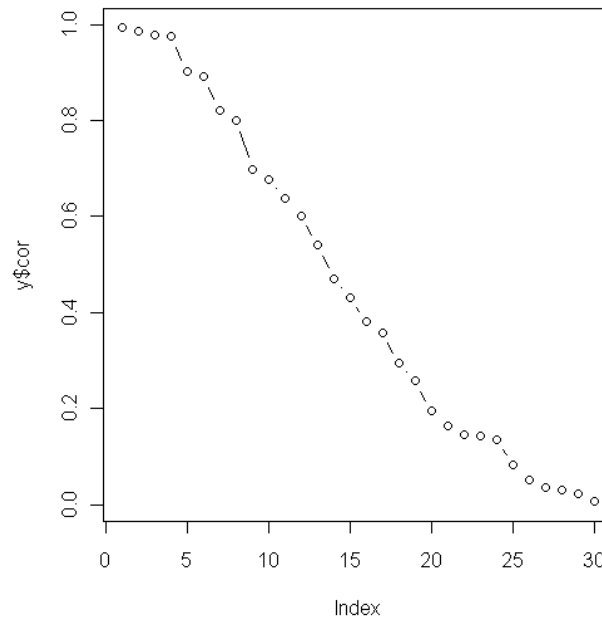


Figura 25. Grafico della correlazione spiegata

Anche qui come possiamo osservare la correlazione spiegata è piuttosto bassa. Sembra esserci anche qui una scarsa correlazione tra i due gruppi formati. Vediamo i grafici delle variabili e degli individui realizzato utilizzando gli stessi metodi applicati nella precedente analisi.

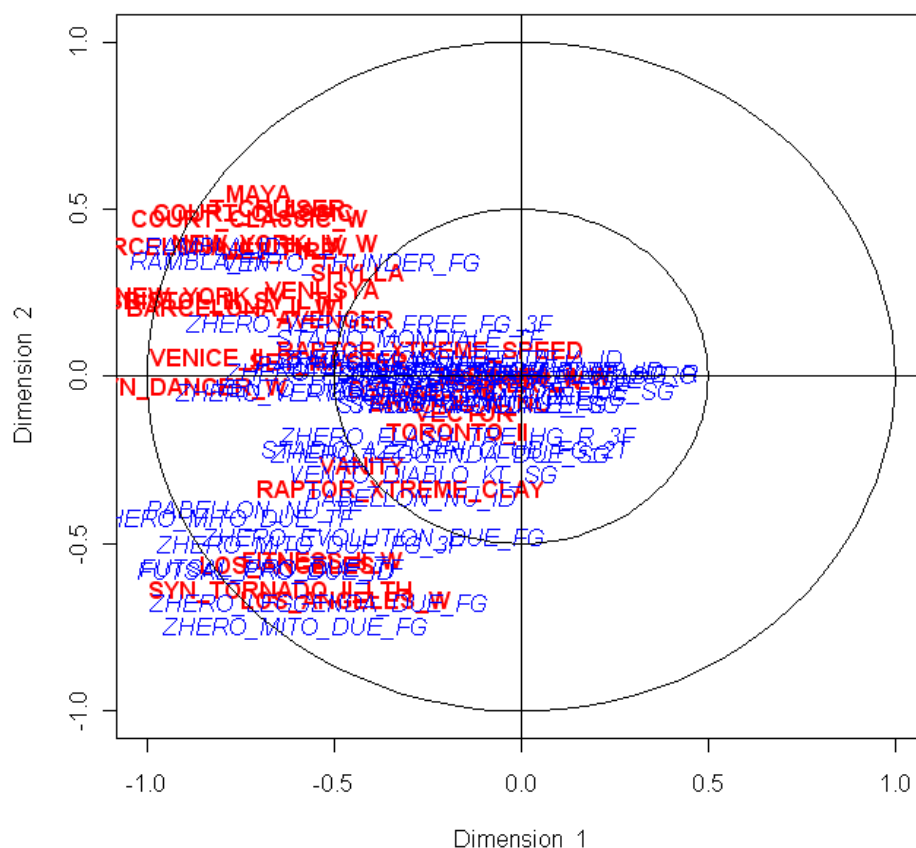


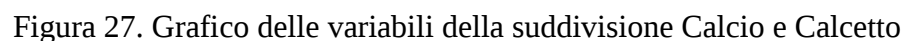
Figura 26. Grafico delle variabili della suddivisione Calcio-Calciotto e Tennis-Running

Da questo grafico osserviamo due raggruppamenti che, a differenza del grafico in figura 22, non si dispongono attorno alle due dimensioni degli assi ma che risultano spostati di circa 45°.

Il primo gruppo è possibile individuarlo sulla bisettrice del terzo quadrante attorno alla quale risultano piuttosto concentrati, mentre il secondo si dispone in un'area più estesa compresa tra l'asse della dimensione 2 e la bisettrice del secondo quadrante.

Da questo grafico non riusciamo a trarre molte informazioni, l'unica informazione che risulta evidente è che vi siano tendenzialmente comportamenti differenti tra le due categorie. I prodotti di Tennis e running che vengono inseriti nel gruppo posto

Vista lo scarso legame che lega queste due categorie di prodotti proviamo a studiare brevemente una correlazione solamente tra i prodotti calcio e calcetto. Come nei casi precedenti la correlazione spiegata è molto bassa e dal grafico delle variabili non è possibile trarre delle informazioni su diversità di comportamento dei prodotti. Eccezion fatta per due modelli che risultano separati dagli altri probabilmente per le quantità molto basse di venduto.



### **3.3.4 Conclusioni**

L'analisi svolta ci permette di trarre alcune conclusioni che tuttavia potrebbero già essere note all'azienda.

La suddivisione uomo – donna dei prodotti ci ha permesso di verificare che prodotti top di gamma hanno comportamenti di vendita comuni indipendentemente dal sesso.

Abbiamo, infatti, visto nel grafico delle variabili che i prodotti top di gamma del mondo femminile risultavano raggruppati assieme ai prodotti top maschili.

La conclusione che l'azienda può trarre da ciò è che potrebbe essere possibile accomunare le tecniche di promozione e vendita per i prodotti top indipendentemente dalla categoria, sia essa tennis, running oppure calcio.

La seconda suddivisione, ovvero quella che raggruppava prodotti tennis e running messi in relazione con i prodotti calcio e calcetto, non ha portato alla luce caratteristiche particolarmente significative. L'unica nota è che è possibile accomunare la vendita del settore calcio e calcetto con i prodotti tecnici del settore running e tennis. La cosa, però, non è una nuova informazione in quanto i prodotti legati al mondo del calcio sono venduti nella maggioranza dei casi da clienti specializzati o comunque che tengono nel loro negozio prodotti tecnici. Mentre i prodotti di fascia medio - bassa è possibile incontrarli anche in negozi di scarpe generici o in negozi della grande distribuzione.

L'ultimo grafico di questa analisi dove vengono messi a confronto prodotti calcio con prodotti legati al calcetto non ha portato a nessuna differenziazione significativa nelle vendite tra i due gruppi.

### 3.4 Algoritmo “a priori”

#### 3.4.1 Applicazione dell’algoritmo “a priori”

L’ultima analisi svolta sul dataset di riferimento riguarda l’associazione di variabili. Attraverso l’utilizzo dell’algoritmo “a priori”, implementato nella libreria R “arules”, siamo andati alla ricerca di associazioni tra i singoli modelli di calzature del settore performance di Lotto.

Vengono calcolate attraverso il metodo “apriori” le correlazioni esistenti tra i modelli presi in considerazione nell’analisi. Queste correlazioni verranno successivamente rappresentate attraverso un grafo dove ad ogni nodo del grafo viene associata una variabile, un arco indica un’associazione fra le variabili mentre l’assenza di un arco indica che le corrispondenti variabili nel modello sono indipendenti condizionatamente dal valore delle altre variabili.

Dopo aver estratto dal dataset le variabili corrispondenti alle calzature costruiamo un grafico che ci permette di individuare quali sono i modelli più richiesti dai clienti.

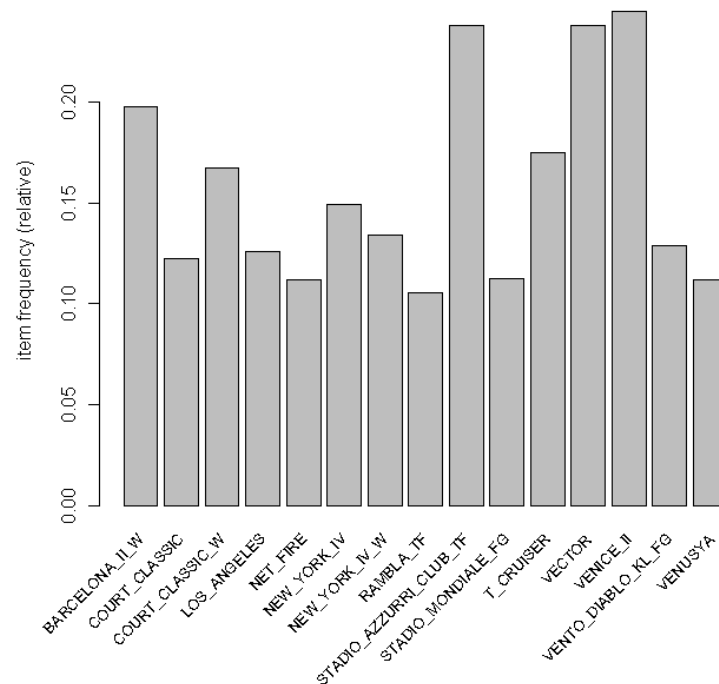


Figura 28. Grafico delle frequenze relative dei modelli più venduti

I quattro modelli che spiccano rispetto a tutti gli altri sono: Barcelona II W (running donna), Stadio Azzurri Club TF (calcetto), Vector (tennis) e infine Venice II (running uomo).

Tutti questi modelli sono modelli che hanno un posizionamento intermedio tra i modelli top di gamma e i modelli cosiddetti “entry”.

Applicando ora il modello “apriori” scopriamo che non esistono regole associative che leghino queste variabili.

Una possibile motivazione per spiegare l’assenza di regole associative è il dataset da cui siamo partiti. L’analisi, infatti, si basa solo sulle quantità di venduto dei vari modelli, e non vi sono quindi altre caratteristiche per trovare associazioni tra i prodotti.

Per applicare correttamente questo tipo di analisi si dovrebbe procedere con la creazione di un nuovo dataset, il quale contenga al suo interno variabili che identifichino più dettagliatamente i modelli e a quel punto vedere se vi sono relazioni tra i vari prodotti.

Dal dataset a disposizione avremmo quindi potuto trarre informazioni per quanto riguarda le associazioni tra i vari clienti, la cosa però avrebbe comportato delle analisi che risultavano di scarso interesse in quanto essendo già i clienti raggruppati in più modi risultava inutile trovare correlazioni tra singoli clienti.

## CONCLUSIONI

Nell'elaborato appena proposto si è cercato di analizzare parte della clientela di Lotto Sport Italia S.p.A. per trarre delle informazioni che potrebbero essere applicate in azioni di vendita o che comunque risultassero di un certo interesse per l'azienda.

La parte introduttiva all'azienda e alle problematiche del marketing di prodotto servivano per visualizzare a grandi linee il contesto di inserimento dello stage e di conseguenza quali fossero le possibili parti da analizzare partendo dai dati sulla clientela che avevamo a disposizione.

Riassumiamo ora le conclusioni che abbiamo potuto trarre da queste analisi.

### ***- Analisi di Raggruppamento***

L'applicazione della cluster analysis ai nostri dati ci porta a tracciare dei raggruppamenti, per la clientela EMEA di Lotto, di clienti con caratteristiche di acquisto simili.

Abbiamo individuato dalla nostra analisi cinque categorie di clienti.

La caratteristica principale dei raggruppamenti ottenuti la possiamo osservare nelle loro nazionalità.

Ad eccezione del secondo gruppo, che comprende clienti di diversi paesi, gli altri presentano un raggruppamento omogeneo nella nazionalità.

Abbiamo, infatti, un gruppo costituito da negozi solo italiani, uno di francesi, uno di spagnoli e infine uno di tedeschi.

L'indicazione che possiamo trarre da questa analisi è che nonostante le differenze di tipologie di negozio, di distribuzione o altro, clienti della stessa nazionalità seguono un comportamento di acquisto comune.

Inoltre, dall'analisi condotta sui singoli prodotti, sono emerse delle particolarità interessanti. La prima riguarda il secondo gruppo, ovvero quello che conteneva al suo interno clienti di diverse nazionalità, tendenzialmente acquista in media quantità più elevate di prodotti con una grande varietà di prodotti. Al contrario il terzo gruppo (clienti francesi) presenta una varietà di prodotti molto bassa, dalle analisi grafiche condotte si è visto infatti che numerosi prodotti non erano acquistati da nessun cliente compreso in quel gruppo.

#### **- *Analisi della correlazione canonica***

L'analisi della correlazione canonica ha avvalorato dei risultati che, nonostante potessero essere noti all'azienda, possono risultare comunque importanti da utilizzare come informazione aggiuntiva per prendere decisioni sulla promozione e distribuzione dei prodotti.

Dall'analisi è emerso che prodotti uomo e donna si comportino sostanzialmente nello stesso modo. Ciò che abbiamo evidenziato con l'analisi è stata la correlazione che vi è tra prodotti top di gamma e prodotti di fascia medio – bassa. I prodotti top di gamma si comportano tutti allo stesso modo, indipendentemente dal sesso a cui è destinato il prodotto.

Non abbiamo trovato inoltre una correlazione significativa in linea generale tra i prodotti calcio - calcetto e i prodotti tennis – running.

La particolarità emersa in questa analisi è stata la possibilità di accomunare il comportamento dei prodotti calcio e calcetto con i prodotti top di tennis e running. Infatti, se i prodotti tennis e running di fascia medio - bassa risultavano raggruppati separatamente a tutti gli altri i prodotti top di queste due categorie presentavano numerosi tratti in comune con l'altro gruppo, ovvero quello legato al mondo del calcio.

#### **- *Analisi “a priori”***

Come già spiegato ampiamente nel paragrafo dell'analisi non è stato possibile condurre questa analisi per una carenza di informazioni nel dataset iniziale.

Per arrivare a delle conclusioni sulle associazioni tra prodotti non è stato sufficiente avere a disposizione le quantità del venduto, ma erano necessarie variabili che specificavano più nel dettaglio le caratteristiche dei vari modelli e a quel punto risultava efficace un'analisi così come era stata pensata inizialmente.

## **Bibliografia e sitografia:**

---

### **BIBLIOGRAFIA**

Aldo Durante - Rapporto OSEM 2006

Roberto Grandinetti – Marketing, mercati prodotti e relazioni, Carocci Editore

Alberto Tieto – “Il Product manager, lo sviluppo di una collezione” (2006)

Alberto Saccardi – Data Mining e Marketing Intelligence (2003)

Introduzione all’uso di SAS System - Modulo Analisi delle Corrispondenze -

Lucia Ronconi

Il progetto ADaMSoft - L’analisi della struttura latente – 05/03/2007

Aldo Durante – Montebelluna Sportsystem distretto cosmopolita

### **SITOGRAFIA**

<http://it.wikipedia.org/wiki/>

<http://www.lottosport.com>

<http://www.eccellere.com/Rubriche/Marketing/PM.htm>

### **Librerie R:**

CCA

arules

cluster