**Esercizio**: Analisi e raccolta dei requisiti per un progetto di analisi dei dati

1. Descrizione del caso di studio scelto

Si intende realizzare un sistema per l’analisi di dati per i supermercati del gruppo Esselunga, azienda che opera nel settore della grande distribuzione organizzata e che conta oltre 170 punti vendita sul territorio nazionale italiano. Lo scopo è analizzare le tendenze di acquisto dei clienti fidelizzati. In particolare si intende esplorare le relazioni tra i possessori di tessera fedeltà e le tipologie di prodotti acquistati, provenienza geografica, metodo di pagamento scelto, tipologia di cassa utilizzata.

2. Analisi e raccolta dei requisiti (generici e poi più specifici)

Abbiamo a disposizione i dati anagrafici dei clienti fidelizzati, le caratteristiche dei prodotti venduti da Esselunga ed i dati relativi ai supermercati della catena.

3. Strutturazione dei requisiti in gruppi di frasi omogenee

Per i **clienti fidelizzati** (5,7 milioni), identificati da numero di tessera fedeltà, si hanno a disposizione cognome, nome, data di nascita, luogo di nascita, codice fiscale, sesso, indirizzo di residenza, indirizzo e-mail, numero di telefono.

Per i **prodotti** (circa 10.000), identificati da codice a barre, si hanno a disposizione marca, categoria merceologica, paese di produzione, azienda produttrice, prezzo, eventuale sconto sul prezzo, disponibilità in magazzino, numero corsia. Tessera fedeltà del cliente, numero pezzi acquistati, data d’acquisto e tipologia di pagamento sono dati memorizzati nella tabella ‘Spesa’.

Per i **supermercati** (171)**,** dislocati in 7 regioni d’Italia come da tabella sottostante, identificati da codice univoco, si hanno a disposizione indirizzo, città, provincia, regione, numero di casse tradizionali, numero di casse automatiche.

**Tabella A**

| **Regione** | **Numero di negozi** | **Province coperte** |
| --- | --- | --- |
| Emilia Romagna | 13 | BO, MO, PC, PR, RE |
| Lazio | 2 | LT, RM |
| Liguria | 2 | GE, SP |
| Lombardia | 102 | BG, BS, CO, CR, LC, MI, MB, MN, PV, VA |
| Piemonte | 17 | AL, AT, BI, NO, TO, VB |
| Toscana | 31 | AR, FI, LI, LU, MS, PI , PO, PT |
| Veneto | 4 | VR, VI |

4. Glossario dei termini

| Termine | Descrizione | Sinonimi | Collegamenti |
| --- | --- | --- | --- |
| cliente fidelizzato | cliente che fa acquisti presso il supermercato, in possesso di una tessera fedeltà | possessore di tessera fedeltà | supermercato, |
| prodotto | bene che viene venduto dal supermercato e acquistato dal cliente | merce | supermercato |
| supermercato | ente che vende il prodotto al cliente | punto vendita | prodotto, cliente fidelizzato |

5. Descrivere nel dettaglio: tipologia di dati necessari (strutturati/non strutturati), outcome atteso dal sistema (analisi di business / analisi predittive), come potrebbe essere di supporto nelle decisioni ed eventualmente quali ‘storie’ è possibile ricavare

Si intendono proporre le seguenti **analisi di business:**  [Data Engineer / Analyst]

* media, mediana, moda, max/min della quantità dei prodotti acquistati per **provenienza geografica** **del cliente** fidelizzato (regione/provincia secondo tabella A), incrociando con categoria merceologica, paese di produzione, azienda produttrice, prezzo, eventuale disponibilità di sconto sul prezzo. Questo fornisce un dato qualitativo sulle preferenze d’acquisto del cliente [dati strutturati]
* media, mediana, moda, max/min della quantità dei prodotti acquistati **per fascia di età del cliente** fidelizzato(le fasce d’età sono 18-24, 25-34, 35-44, 45-60, over 60), incrociando con categoria merceologica, paese di produzione, azienda produttrice, prezzo, eventuale disponibilità di sconto sul prezzo. Questo fornisce un dato qualitativo sulle preferenze d’acquisto del cliente [dati strutturati]
* media, mediana, moda, max/min della quantità dei prodotti acquistati per **sesso del cliente** fidelizzato (maschi, femmine, non specificato), incrociando con categoria merceologica, paese di produzione, azienda produttrice, prezzo, eventuale disponibilità di sconto sul prezzo. Questo fornisce un dato qualitativo sulle preferenze d’acquisto del cliente [dati strutturati]
* media, mediana, moda, max/min della quantità dei prodotti acquistati per **collocazione geografica del supermercato.** Questo fornisce un dato quantitativo sulle preferenze d’acquisto del cliente [dati strutturati].
* media, mediana, moda, max/min del**l’afflusso di clienti giornaliero/mensile/annuale** per **collocazione geografica del supermercato** considerando giorni feriali e festivi. Questo fornisce un dato quantitativo sulle affluenze di clienti in ogni punto vendita. [dati strutturati].
* media, mediana, moda, max/min del numero **di clienti** che scelgono giornalmente, mensilmente, annualmente di utilizzare le **casse** tradizionali e le casse automatiche
* media, mediana, moda, max/min del numero **di clienti** che scelgono di giornalmente, mensilmente, annualmente di utilizzare i **diversi metodi di pagamento**

Si intendono inoltre proporre le seguenti **tecniche predittive**: [Data Scientist]

* **Market Basket Analysis** sui prodotti acquistati per sondare le associazioni tra articoli venduti → utile per decidere quali prodotti mettere in vendita, quali prodotti mettere in promozione e come disporre la merce all’interno delle corsie del supermercato. [dati strutturati]
* **Analisi predittiva sul comportamento dei clienti**: ottenendo le informazioni sulle preferenze di acquisto è possibile descrivere le abitudini del cliente fidelizzato, sulla cui base prevedere il suo comportamento futuro.
* **Posizione della merce all’interno del supermercato**: confrontando la posizione in corsia degli articoli acquistati è possibile sapere quali aree del supermercato attirano maggiormente l’attenzione e quali sono i percorsi più frequentati all’interno del punto vendita [dati strutturati]
* **Analisi predittiva dei prezzi**: confrontando la domanda, lo storico dei prezzi, l’attività della concorrenza ed i livelli di inventario è possibile ottimizzare i prezzi, prevedendo ed impostando in automatico i prezzi migliori per rispondere ai cambiamenti del mercato [dati strutturati]
* **Creazione di campagne pubblicitarie mirate**: analizzando dati provenienti da fonti come l’andamento del mercato, domanda e offerta, tassi di risposta, analisi del cliente fidelizzato, è possibile determinare quali campagne avranno più successo. Scegliendo messaggio, prodotto ed il contenuto dell’offerta più adatta al singolo cliente sarà possibile raggiungere tassi di conversione più elevati rispetto alle campagne non personalizzate [dati strutturati e non strutturati]
* **Posizione geografica dei punti vendita**: l’analisi predittiva può essere utilizzata per prevedere e analizzare il successo di un supermercato. I dati coinvolti riguardano i comportamenti di acquisto, concorrenza, andamento del mercato immobiliare [dati strutturati]

Ruoli:

* **Data Enginee**r: Raccolta e memorizzazione i dati
* **Data Analyst**: Ispezione, trasformazione e analisi dei dati. Risponde alle domande di business legate all’affluenza al supermercato, alle preferenze negli acquisti.
* **Data Scientist**: utilizza modelli predittivi per aiutare gli stakeholder a prendere decisioni relative a inventario, prezzi, posizione degli articoli all’interno del supermercato, posizione geografica dei nuovi punti vendita.

F. Micheletti

F.Testa