# ***ABSTRACT***

L’evoluzione della Business Intelligence è iniziata decenni fa con i primi report mainframe, chiamati output di sistema. Essi venivano principalmente stampati su carta, per poi essere periodicamente distribuiti ai manager.Le prime query hanno velocizzato il processo e hanno consentito ai manager tecnicamente esperti di creare report personalizzati ad hoc, ma pochi manager avevano il tempo e le competenze per farlo. L'emergere del data warehouse ha dato un grande impulso alla BI aggregando tutti i dati in un'unica posizione, dove potrebbe essere interrogato in modo interattivo senza impatto sulle applicazioni tramite l’uso di Query e rapporti online con interfacce grafiche sempre più facili da utilizzare.

L’avvento delle data warehouse, delle data mart e gli strumenti di analisi analitica hanno reso la BI accessibile a più gestori e hanno permesso ai manager di ottenere informazioni e risposte critiche in modo efficiente e rapido.

Il progetto propostosarà dedicato alla descrizione in dettaglio della creazione di una data mart dedicata alle vendite di una azienda del fashion attraverso un soluzione ottimale di best practice di un processo ETL ottenendo come risultato lo Snowflake Schema e lo Star Schema, ideale per la data. Inoltre, utilizzando il processo di classificazione comprendente sia i dati aziendali che i dati ISTAT ricavati dall’omonimo sito, si ha avuto la possibilità di localizzare le area più efficaci per aprire un nuovo negozio e di offrire una spiegazione sul perché alcuni negozi sono stati chiusi in un recente passato.

Le conclusioni, sono visualizzate in Power BI, software Microsoft per la data Visualization..

# ***INTRODUZIONE***

Man mano che diventiamo una società digitale, la quantità di dati creati e raccolti cresce e accelera in modo significativo. L'analisi di questi dati diventa una sfida per gli strumenti analitici tradizionali che fanno sempre più fatica a stare al passo. È necessaria quindi una costante innovazione per colmare il divario tra i dati generati e i dati che possono essere analizzati in modo efficace.

I grandi strumenti e le tecnologie di dati offrono opportunità e sfide nel poterli studiare in modo proficuo per comprendere meglio le preferenze dei clienti, ottenere un vantaggio competitivo sul mercato e far crescere il loro business.

Le architetture di gestione dei dati si sono evolute dal tradizionale modello di data warehousing ad architetture più complesse che soddisfano requisiti differenti, come l'elaborazione in tempo reale e in batch, dati strutturati e non strutturati, transazioni ad alta velocità, etc.

Mediamente Consulting s.r.l., nata nel 2007, è una società di consulenza e progettazione specializzata in sistemi a supporto delle decisioni. In particolare, si occupa di progetti di Business Intelligence, Data Warehouse e Advanced Analytics in ambito Big Data.

Grazie a tali sistemi, il cliente ha la possibilità di visualizzare le informazioni e conseguentemente di prendere una decisione più consapevole basata su fatti. Pertanto, Mediamente Consulting s.r.l. supporta il cliente nella conoscenza delle proprie performance e lo aiuta ad incrementarle attraverso decisioni migliori.

Per lo svolgimento del mio progetto di tesi, sono stato inserito in un team che segue un’importante azienda di Moda, che risponde alle esigenze analitiche in ambito fashion.

Il cliente si occupa di gestire il settore occhialeria in campo “Fashion Retail”. Si tratta di una holding multinazionale responsabile delle vendite di prodotti in tutto il mondo.

Al fine di massimizzare lo sviluppo del proprio marchio, l’azienda ha deciso di puntare su una nuova finestra di mercato sui prodotti Sneakers, lanciati nel febbraio 2019 alla Fashion week tenutasi a Milano su cui elaborerò una analisi e gestione delle vendite.

Anche attraverso questo progetto, l’azienda sta mettendo a punto una forma innovativa di gestione delle operazioni di analisi strategiche e operative che consente di cogliere appieno il potenziale di crescita dei suoi brand, all’interno di un mercato globale considerevole e molto competitivo, in cui il segmento di mercato sta godendo di una crescita sostanziale.

In questo modo è stato possibile garantire le attività di analisi operative e strategiche attraverso report Real-Time e parallelamente consolidare le regole per la costruzione dei report finali. Quest’ultima fase precede la costruzione del Data Mart nell’ambito della Produzione e sarà successivamente integrato nell'Enterprise Data Warehouse già esistente.

Di seguito andrò a presentare brevemente e schematicamente i capitoli in cui si compone il mio elaborato. L’illustrazione dell’intero lavoro svolto, è stato possibile grazie all’ausilio di concetti teorici fondamentali e all’aiuto fornito dai miei colleghi in ambito lavorativo.

Il primo capitolo può essere considerato un’introduzione ai concetti fondamentali Dei Big Data e in generale della Business Intelligence, che rappresentano lo Stato Dell’arte. Inoltre, si spiegheranno i meccanismi usati e i concetti chiave del Data Mining e del Machine Learning.

La parte centrale del mio elaborato, che corrisponde ai capitoli 2-3, sarà dedicato alla descrizione delle analisi condotte e della metodologia usata da cui sono partito per lo sviluppo del mio progetto di tesi. Sarà spiegato in dettaglio il processo ETL aziendale ottenendo come risultato lo Star Schema, ideale per la Data Visualization, e lo Snowflake Schema, ideale per la struttura di un processo Extraction, Trasformation & Loading. Utilizzando il processo di classificazione comprendente sia i dati aziendali che i dati ISTAT ricavati dall’omonimo sito, si ha avuto la possibilità di localizzare le area più efficaci per aprire un nuovo negozio e di offrire una spiegazione sul perché alcuni negozi sono stati chiusi.

In questo modo si potrà avere una panoramica generale dei concetti applicati per svolgere il lavoro.

Infine, nell’ultimo capitolo, descriverò l’intero processo del progetto per poi concludere andando a delineare nel dettaglio i dati ottenuti tramite l’utilizzo di Power BI sul contesto applicativo in cui si è svolto il progetto, osservando e analizzando le caratteristiche e scelte implementative ottenendo così una visione strategica e decisionale che l’azienda potrà decidere di implementare o meno.

Questo capitolo risulta il più importante, in quanto verrà descritto il mio contributo al progetto, ovvero lo sviluppo di un metodo per fare reportistica su dati live, attraverso una analisi top-down, partendo dalle esigenze dell’utente e una datawarehouse già completa, per costruire le Query e Fact Table per la realizzazione di Data Mart formando un Database contenuto, che sarà essenziale per la progettazione e visualizzazione dei report finali.

## ***METODOLOGIA***

In questo paragrafo saranno illustrate le fasi principali del progetto aziendale.

1. Top-Down Analysis per comprendere e analizzare in modo completo e efficace tutte le caratteristiche rilevanti del progetto, ponendosi domande e provando a capire come ottenere risposte da esse, per poi arrivare ad ottenere una base i dati che serviranno per raggiungere l’obiettivo finale.
2. Costruzione della Data Mart tramite l’utilizzo di Talend Open Source, software usato per il Traditional ETL, evidenziando tutti i processi aziendali dalla Staging area della datawarehouse, dove si importeranno i file di origine (csv, excel…)
3. Creazione dell‘ETL Best Practice Model e dello Snowflake Schema grazie alla creazione di surrogate key per le relazioni tra le tabelle, ottimizzandone le prestazioni.
4. Creazione della Data Visualization Best Practice Model e dello Star Schema, entrambi utili per essere efficienti nella rappresentazione di vari report utili per le finali decisioni aziendali del cliente.
5. Creazione di un modulo per velocizzare il processo di aggiornamento delle tabelle, raggruppando i job appartenenti alla stessa famiglia (Anagrafiche e Movimenti) in modo che si eseguano in contemporanea e grazie ad un sistema di avvisi di Warning/Error con conseguente invio di email, di avere subito un Audit del problema.
6. Utilizzo di software di Data Mining e Machine Learning che per mezzo dell’intelligenza artificiale, implementano autonomamente diversi algoritmi in grado di eseguire un processo sempre più ottimizzato unendo i dati del datawarehouse e i dati ISTAT italiani, precedentemente estratti.
7. Creazione di una Dashboard in Power BI per avere un controllo e una serie di dati effettivi su tutto quello che riguarda le vendite. Grazie a questo processo è possibile estrarre risultati da fornire al clienti in modo chiaro ed esaustivo, aiutandoli a comprendere i problemi dei loro sistemi di produzione e commercio, facendo sì che migliorino in futuro tramite la scelta di giuste e nuove azioni manageriali.

## ***SCHEDULING***

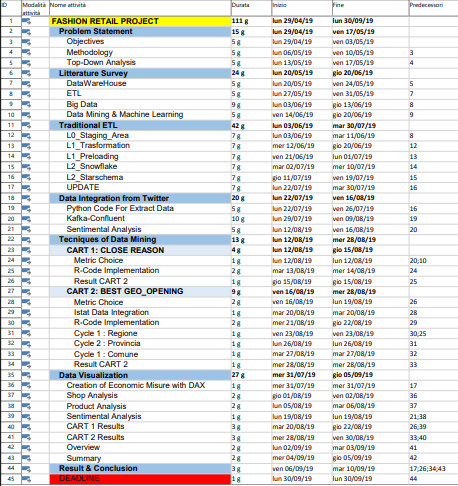


Figura 1: Project Scheduling

Immagine che contiene testo, mappa

Descrizione generata con affidabilità molto elevata

Figura 2: Gaant Chart

## ***TOP-DOWN ANALISI DEL PROBLEMA***

**Attributi aziendali di successo nel campo della moda:**

* posizione strategica
* qualità/prezzo
* diversificazione dei prodotti
* innovazione
* comodità

**Obiettivo aziendale:**

* Guadagnare il più possibile avendo il minimo inventario in magazzino = costi
* Aumentare le vendite

**Competitors:**

Gucci, Michael Kors, Coach

**Data Mining:**

Regole di associazione tramite algoritmo Apriori

* (Borsa, Portafoglio)
* (Zaino, Portafoglio)
* (Borsa, Zaino, Portafoglio)

Clustering tramite algoritmo K-Means & Classificazione tramite algoritmo CART:

* Causa di Chiusura
* Country
* Tipo di negozio
* Rimarrà aperto?
* Best Geo-Positioning per aprire un uovo negozio.

### ***QUESTION & ANSWER***

1. Meglio una vetrina elegante o stravagante per attirare i clienti?

Osservo i top 5 negozi e tramite un backward analisi guardo se son negozi più eleganti e molto statici oppure son negozi più “colorati e vivaci” nell’esporre i propri prodotti.

1. Dopo quanto devo spostare la merce da un negozio retail ad uno outlet?

Valuto il tempo che un oggetto rimane nello store, calcolato dal tempo di consegna al tempo attuale/ tempo di vendita.

1. Quante volte un mio cliente ritorna a comprare? Quante volte invece torna a comprare ma attraverso l’e-commerce?

Valutazione negozi. Utilizzo il count della gente che entra ed esce e associa ad essi le vendite.

1. Quanto magazzino può permettersi questo negozio? Quanto posso rischiare di lasciare invenduto?

Magazzino= costi. Calcolare la percentuale merce venduta/merce arrivata e sfilare una classifica dei WORST 5. In questi negozi bisogna spedire meno merce.

1. Quale particolare mi differenzia e mi fa subito riconoscere da competitors?

Bisogna far si che l’azienda si focalizzi su pochi prodotti che innalzano la Brand Loyalty, valutazione tramite Pareto.

1. Meglio aprire negozi di nicchia nelle grandi città o aprire negozi negli outlet?

* Calcolare il rapporto vendite profitto per tipo di shop.
* Calcolare #visitatori nei negozi outlet e nei corner per vedere se l’affluenza e le vendite mi posso far cambiare strategia.
* Preferisco visibilità o fatturato?
* Classificazione Data Mining.
* Quali Outlet hanno un numero maggiore di visibilità?
* Analisi territori tramite dati ISTAT.
* In quali zone l’azienda è debole in termini di mercato?

1. Quando devo creare una promozione e per chi? Meglio a breve o lungo termine? Mi conviene fare promozioni su un singolo prodotto o basandomi sulle regole di associazione? Quando il mio numero delle vendite è sotto la media?

Come faccio a riequilibrare questo dato? Con delle promozioni singolari per determinati oggetti o promozioni tramite regole di associazione in questo modo determino anche il periodo dell’offerta.

1. Quanto cresce/diminuisce il mio marchio rispetto al settore?

Dove il mio marchio fa parlare e coinvolge la gente?

Tramite real time tweets con l’uso di Python e Confluent inserendo i dati in un db grazie ai sink connettors. Osservazione dei dati e dei tweet inerenti al mio marchio e a quelli dei competitors. Quale è il feedback dei miei clienti?

Il numero di citazioni sta salendo o decrescendo rispetto ad un periodo passato?

In quale luogo il mio marchio è piu conosciuto? In quali paesi invece devo investire per aumentare la mia visibilità? Dati real time by Twitter.

1. Come si stanno muovendo i miei competitors?

Continua osservazione dei puri dati delle vendite di ogni competitors con relativo simbolo di trading positivo o negativo per evidenziare l’andamento odierno della azienda rispetto ai vari avversari del settore, il tutto riassunto in una tabella.

1. Il mio cliente è premium?

Offerte speciali per utenti premium individuati tramite creazione di un boolean.

1. Mi conviene puntare piu su una clientela puramente femminile o anche maschile?

Calcolare la percentuale di clientela femminile sul totale

Calcolare la percentuale delle vendite di un prodotto femminile.

1. Quanto un periodo di saldi incide sul numero di vendite? Quanto incide, invece, sul Margine complessivo aziendale?

Offerte speciali per i clienti mensilmente e confrontarli con il margine.

### ***DATI PER L’ANALISI***

1. Prodotto (id, descrizione, categoria, data di uscita, stagionalità, caratteristiche, sesso)
2. Negozi (id, posizione geografica, descrizione, tipo/canale)
3. Scontrini (id, id\_data, prodotto, id\_cliente, valore, sconto, quantità)
4. Data (id, giorno, mese, quadrimestre, semestre, anno)
5. Real time delle Vendite e dei tweet in base alla localizzazione (Italia principalmente) e rispetto ai competiors.
6. Dati pubblici commerciali: bilancio, Dati ISTAT su popolazione,turismo,indicatori economici.

## ***OBIETTIVO***

Il panorama della moda è stato ampiamente studiato per migliorare capire come i leaders influenzano i loro seguaci e, in cambio, fornire informazioni preziose ai decisori sul lancio e sul marketing del prodotto strategie. L'obiettivo del nostro studio è ottenere tramite il processo ETL, una Best Practice per la creazione di una Datawarehouse o di una Data Mart per un Cliente, sfruttando i dati ottenuti per ottenere delle analisi di business sulle vendite, con risultati chiari ed efficaci.Faremo un confronto con i più grandi competitors sulla tematica Sentiment & Social Analitycs, sfruttando e si terrà particolare attenzione alle vendite del 2019 con l’arrivo delle Sneakers, lanciate dall’azienda alla Fashion Week di Milano di Febbraio, per capirne le potenziali capacità di inserimento nel mercato e il valore aggiunto apportato.

I capi di moda hanno sempre un’influenza potente e irresistibile sui seguaci della moda. Pertanto, è importante scoprire le relazioni nascoste tra le caratteristiche dei consumatori e le proprietà dei loro prodotti preferiti, al fine di aiutare i fashion designer nello sviluppo del prodotto e nel marketing in lanciando le loro strategie o campagne di marketing.

In particolare, svolgerò una analisi approfondita sul territorio Italiano e sui negozi evidenziando le cause principali della chiusura di essi e tramite l’utilizzo di algoritmi di machine learning e data mining per evidenziare la tipologia e la zona geografica ideale per aprire un nuovo store.