Alma Mater Studiorum - Università di Bologna

SCUOLA DI INGEGNERIA E ARCHITETTURA

Corso di Laurea Magistrale in Ingegneria Informatica

Tesi di Laurea
in Processi e Tecniche di Data Mining M

LOCATION INTELLIGENCE PER VENDITA AL DETTAGLIO

Tesi di laurea di:

Riccardo Mencucci

Relatore:

Chiar. mo Prof. Ing. **Claudio Sartori**

Tutor:

Dott. **Matteo Zanirati**

Anno Accademico 2017-2018 Sessione I

Indice

Introduzione

Tra i settori in grado di offrire opportunità interessanti, il *retail* ha avuto in questi anni una notevole evoluzione. Per capire come funziona, bisogna innanzitutto conoscere cos'è. Il termine retail, che tradotto dall'inglese significa commercio al dettaglio, è una distribuzione specializzata che avviene attraverso una rete organizzata di punti vendita. Grazie al retail, le aziende possono mettere a disposizione i prodotti con modalità, tempi e luoghi scelti dai consumatori. Il retail è un nuovo modo di concepire produzione, logistica e servizio commerciale a vantaggio dell'utente. Per far funzionare questo meccanismo, oltre alla logistica è fondamentale la garanzia di numerosi servizi accessori che determinino le modalità con cui il consumatore può accedere agli articoli. L'utente non vuole acquistare semplicemente i prodotti offerti: per questo motivo le aziende devono differenziare l'offerta per soddisfare al meglio la diversa tipologia di domanda.[LET43]. Inoltre, grazie allo sviluppo tecnologico e alla messa in produzione di nuovi dispositivi cellulari è possibile al giorno d'oggi usufruire dei *social network* che possono essere utilizzati, sia per uso personale per esprimere pensieri e opinioni su un particolare fatto da condividere con tutti, sia per effettuare acquisti/ vendite cambiando la sfera del business.

In base a queste richieste sempre più esigenti dei clienti le aziende di consulenza moderne devono farsi trovare preparate, cercando di capire come si evolve il mercato del *business* che le circonda confrontandosi ogni giorno con una mole di dati sempre crescente. Al fronte di queste esigenze nascono i Data Warehouse in grado di avere un supporto di memorizzazione dei "Big Data" che possano in qualche modo garantire l'estrapolazione delle informazioni principali, detti anche KPI, che possano guidare le aziende nel business di mercato.

Nell'ambito "fashion retail", in particolare nel mondo delle vendite ("*Sell-out*"), si vuole cercare di capire qual'è il trend corrente, come si evolve nel tempo e nello spazio (determinando le zone che riscontrano più vendite) , osservando se si verificano pattern interessanti da segnalare.

Il lavoro realizzato rappresenta un progetto di ricerca e sviluppo, rappresentato sotto forma di

P.o.C. (prova di concetto), che permetta di vedere dove si può arrivare partendo da dati inerenti a vendite di occhiali derivanti da scontrini. Grazie al supporto della **Location Intelligence** è stato possibile concatenare dati di vendite con informazioni inerenti allo **spazio geografico** come *territorio* (popolazione, occupazione e salario medio), *dati social e dati turistici*.
Mettendo insieme queste informazioni si è determinato un nuovo KPI che rappresenta la **potenzialità dell'area**. Tutto questo permette di avere un analisi completa su un caso di studio che va a contestualizzare i dati di vendita sul cliente *Kering Eyewear* in negozi che si trovano negli Stati Uniti nell'arco di tempo che va da Luglio a Ottobre 2017. La visualizzazione dell'informazione è realizzata attraverso delle dashboard interattive che possano raccontare una *storia* che possa essere facilmente interpretata dai manager e da coloro che hanno comptetenze sul campo. Il lavoro è stato strutturato in questo modo:

1) *Concetti principali:* introduzione ai concetti fondamentali della Business Intelligence, Location Intelligence e Social Media Analysis dando una panoramica generali dei concetti applicati per svolgere il lavoro;

2) *Contesto applicativo e tecnologie:* presentazione del contesto applicativo dove si è svolto il progetto e il cliente interessato. Descrizione delle tecnologie usate per svolgere il progetto;

3) *Caso di studio:* Presentazione del caso di studio esplicitando le fasi salienti che hanno portato alla visualizzazione dei risultati permettendo di dare una valutazione generale sugli output prodotti.

4) *Conclusione:* discussione sul lavoro effettuato permettendo di focalizzarsi sugli sviluppi futuri.

1

Concetti principali

1.1 Business Intelligence

Nell'ambito lavorativo moderno, soprattutto nel campo della consulenza, le aziende si trovano sempre più spesso a confrontarsi col bisogno di conoscere il modo in cui si stanno comportando e come si evolve il mercato in cui si trovano ad operare con il principale obiettivo di fare *business* e diventare *leader di mercato.*
Data l'elevata mole di dati , con cui è si è costretti a confrontarsi quotidianamente, è necessario riuscire a trovare un modo da:

- permettere di raccogliere e processare dati ad alta velocità (spesso si parla di processi *real-time);*

- fornire un servizio di pulizia del dato stesso eliminando dati sporchi, duplicati e/o errati (***ETL Processing);***

- definire un sistema solido di memorizzazione per i dati certificati (***Data Warehouse);***

- trasformare *l'informazione* in fonte di *conoscenza* attraverso analisi di business sui dati stessi, determinando nuovi KPI che apriranno la strada a nuove analisi.
Questo processo è noto già da molti anni ma mai come in questo periodo la sua applicazione sta portando le aziende ad avere dei risultati operativi concreti, anche grazie allo sviluppo scientifico e tecnologico, concentrandosi principalmente nell'impresa di realizzare *Sistemi di Supporto alle Decisioni(DSS).*

Il riassunto di questa introduzione alla Business Intelligence può essere rappresentato attraverso la seguente figura.

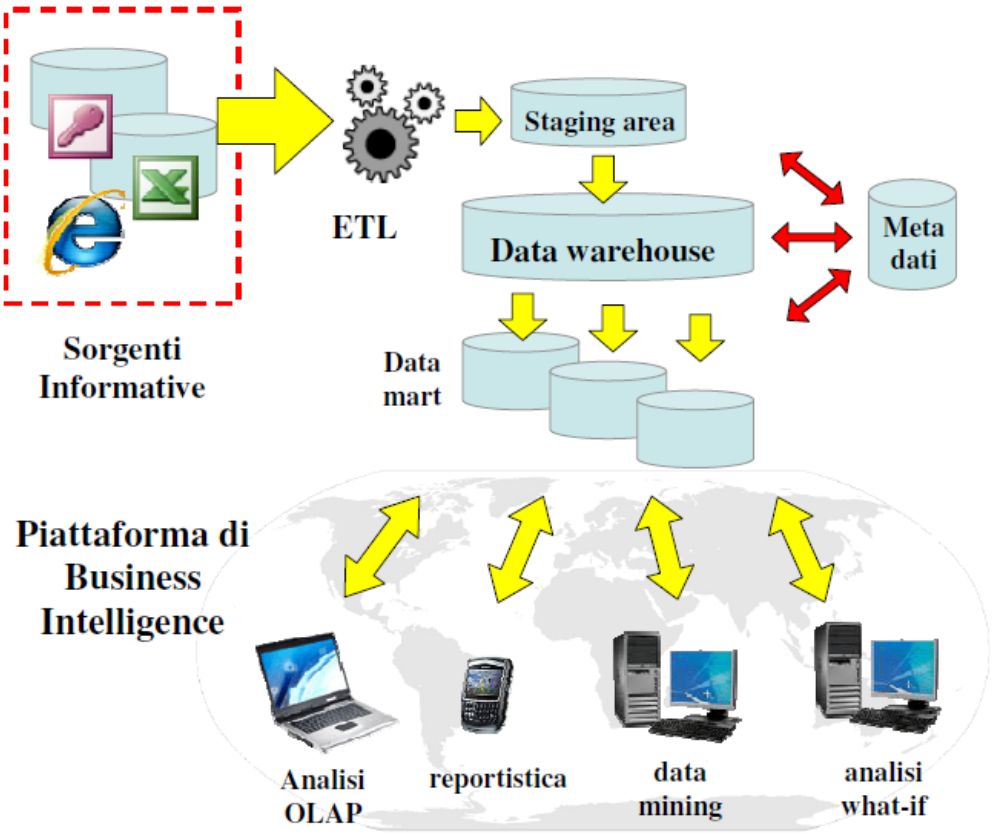


Figura 1.1 -Piattaforma di Business Intellgence

1.1.1 Definizione e contesto

Con Business Intelligence (BI) si intende il processo di trasformare dati grezzi in utili informazioni per supportare strategie di mercato efficaci e consapevoli. Si vogliono catturare i dati di business e fornire le giuste informazioni, alle giuste persone, nel perfetto momento[RF03].

Queste poche righe riportate sopra vanno a dare la definizione più semplice e completa da assegnare alla Business Intelligence. Come già accennato, lo scopo della **Business Intelligence** è riuscire a trasformare il mare di dati disomogenei, disorganizzati e spesso anche ridondanti in vere e proprie fonti di conoscenza, attraverso una serie di tecniche, strumenti e processi definiti come *decision making*. Ecco che a questo punto nasce nelle aziende un nuovo tipo di figura di fondamentale importanza: il *Businnes Analyst*

1.1.2 Business analysis

All'interno di ogni organizzazione aziendale ogni dipendente/consulente si trova a svolgere attività e carichi di lavoro diversi a seconda del ruolo che svolge al suo interno. Spesso viene fatta una suddivisione in *team* in modo che l'azienda può concentrarsi su più azioni (in ambito della consulenza spesso si parla di *progetti*) in maniera parallela. I progetti possono essere associati a una figura di grande rappresentanza e prestigiosità: " **Il Project Manager"**.
Il project Manager è il responsabile unico dell'avvio, pianicazione, svolgimento, controllo e chiusura di un progetto. Questo arduo compito non viene fatto da una persona sola, infatti è compito del *team leader (*una figura carismatica) di guidare lo stesso team nelle azioni e nel modo su come agire e quindi definire quale sono le strategie di business con opportune analisi[LUISS15]. A prescindere dai ruoli che una persona ricopre la vera forza motrice che sta alla base delle *Decision Support Systems* è la *Businnes Analisyst.*
La definizione viene riportata sottostante:

La Business Analyst è l'esplorazione di dati storici di una attività, provenienti da fonti molto diverse, attraverso l'uso di analisi statistica, qualitativa, data mining, modellazione percettiva e altre tecnologie, con lo scopo di identificare pattern ed estrarre informazioni che possano guidare i Business Manager nel

prendere le proprie decisioni[RF03].

Figura 1.2: Business Analysis

1.1.3 Big Data

Abbiamo parlato spesso di Big Data senza essere mai entrati nel dettaglio in cosa essi realmente rappresentano e come possono essere classificati. Le caratteristiche, che deve avere un grande dataset per essere definito come "Big Data", possono essere racchiuse in 6 V caratteristiche[RF02]:

1)*Volume:* I Big Data sono sempre in grandi dimensioni (ordine di centinaia di TeraByte), sono sia da memorizzare che da trasmettere, quindi sia infrastrutture che reti devono essere in grado di sostenerli. Inoltre sono quasi sempre distribuiti;

2) *Velocity*: I Big Data non solo sono tanti, ma vengono acquisiti e devono essere processati e restituiti ad una velocità altissima, molto spesso si parla di real-time/stream process. In più solitamente devono subire elaborazioni molto complesse;

3) *Value*: Il valore sia economico che statistico dei Big Data è sempre molto elevato;

4) *Variability*: I Big Data sono totalmente dinamici, soprattutto quando associati a processi real-time;

5)Veracity: Dal momento che vengono utilizzati per il decision-making automatico, i Big Data devono essere estremamente accurati, puliti da elementi di rumore, outliers e replicati. Il processo di pulizia viene però fatto dopo la raccolta, inizialmente infatti si prendono tutti i dati che si trovano, accumulandone in gran quantità e senza preoccuparsi della qualità;

6) *Variety*: I Big Data non solo numeri e stringhe, essi comprendono dati strutturati, non strutturati, probabilistici, suoni, immagini, video, locazioni geografiche, opinioni di utenti e molto molto altro ancora.

Inoltre si è ultimamente aggiunta una settima V:

7)*Visualization*: parte integrante dei Big Data è come vengono visualizzati dopo la procedura di elaborazione, la conoscenza estratta deve essere presentata nel modo più efficiente possibile.

Queste caratteristiche completano la seguente definizione:

Con il termine Big Data si intende una collezione di dataset così grandi e complessi che sono difficilmente processabili con normali database relazionali[RF02].

Serve quindi un supporto alla Business Intelligence che permetta di memorizzare, gestire e caricare i dati in una struttura ad hoc, nascono quindi i "*Data Warehouse*".

Figura 1.3 – Big Data Features

1.1.4 Data Warehouse

I Data Warehouse (DWH) sono il principale strumento a supporto della Business Intelligence, possiamo vederli come dei repository ottimizzati che memorizzano le informazione per il processo di decision-making.[RF04]

L'obiettivo di un DW è di supportare il "knowledge worker"(dirigenti, amministratori, manager, analisti) per condurre analisi sui dati finalizzate ad attuare processi decisionali e guidare strategie di business. Dovranno quindi essere attentamente progettati per gestire in maniera efficiente ed efficacie le caratteristiche dei Big Data.

I Data Warehouse sono realizzati come principale base per i Decision Support System (DSS) quindi risaltano le seguenti caratteristiche:

- Sono *Subject-Oriented,* in quanto si focalizzano su concetti specifici di analisi dell'impresa aziendale (ad esempio cliente, vendite,ordini ecc...);

- Sono *integrati e consistenti*: aggregano dati da diverse fonti (interne e\o esterne dall'impresa);

- *Si evolvono* costantemente nel tempo, si parla spesso di dati dinamici;

- Sono dati *non volatili* e quindi mai sovrascritti o cancellati ("write once read many").

Tra I Big Data possiamo fare una distinzione in base alla dimensione crescente e valore calante:

1. *Azioni:* sono le decisioni prese post-analisys, sono poche e di grande valore;
2. *Presentazioni:* permettono di prendere decisioni;
3. *Informazioni:* sono estratte dai dati e permettono di fare prfesentazioni;

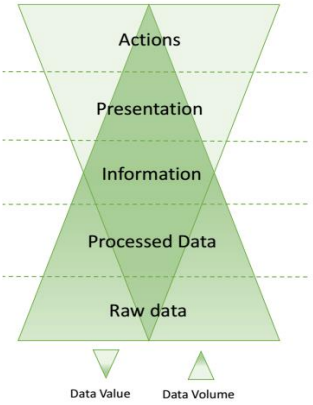


Figura 1.4- Scala valore-volume

4. *Dati processati:* I dati vengono selezionati e puliti quindi processati;
5. *Dati grezzi:* sono di enorme quantità e di pochissimo valore;

Una volta processati I dati vengono caricati nel DWH. Per fare interrogazioni sui dati è necessario capire quali di essi rappresentano una coordinate di analisi comune. Grazie a questo concetto nascono i *Data Mart.*

1.1.4.1 Data Mart

Con Data Mart si intende un sottoinsieme o un aggregazione dei dati salvati in un DWH, essi includono i dati rilevanti per una determinata applicazione o business area[RF04].

Un Data Mart viene estratto da un **Data Warehouse**, dal quale possono essere ottenuti tanti Data Mart quante sono le finalità che si vogliono perseguire con le successive analisi, tuttavia può essere costruito anche in assenza di un sistema di dati integrato. Detto in termini più tecnici, un Data Mart è un sottoinsieme logico o fisico di un Data Warehouse di maggiori dimensioni.
Sulla base di questa definizione si apre un grande dibattito in base al punto di vista top-down o

bottom-up tra Data Warehouse e Data Mart presentando differenti tipi di architettura.

1.1.4.2 Architettura di un Data Warehouse

Come anticipato precedentemente, il DWH costituisce il principale sistema di supporto della BI e, in fase di progettazione, risulta fondamentale stabilire quale tipologia di architettura adottare. Una buon architettura deve soddifare I seguenti requisiti[KEL97]:

- *Separazione*: I processi analitici e transazionali devono essere tenuti il più possibile separate.
- *Scalabilità*: le architetture hardware e software dovrebbero poter essere facilmente ridimensionate a fronte della crescita nel tempo dei volumi di dati e del numero di utenti.
- *Estendibilità*: l'architettura dovrebbe essere in grado di accogliere nuove applicazioni e tecnologie senza dover riprogettare l'intero Sistema.
- *Sicurezza*: il controllo sugli accessi è essenziale a causa della natura strategica dei dati memorizzati in un DWH.
- *Amministrabilità*: la complessità di gestione del DWH non dovrebbe essere eccessiva.

Modello di Inmon - Corporate Information Factory: I DWH si costruiscono nella loro totalità fin da subito come un unico blocco monolitico, non è possibile vederli come la composizione dei DM. Viene adottata una vision Top-Down

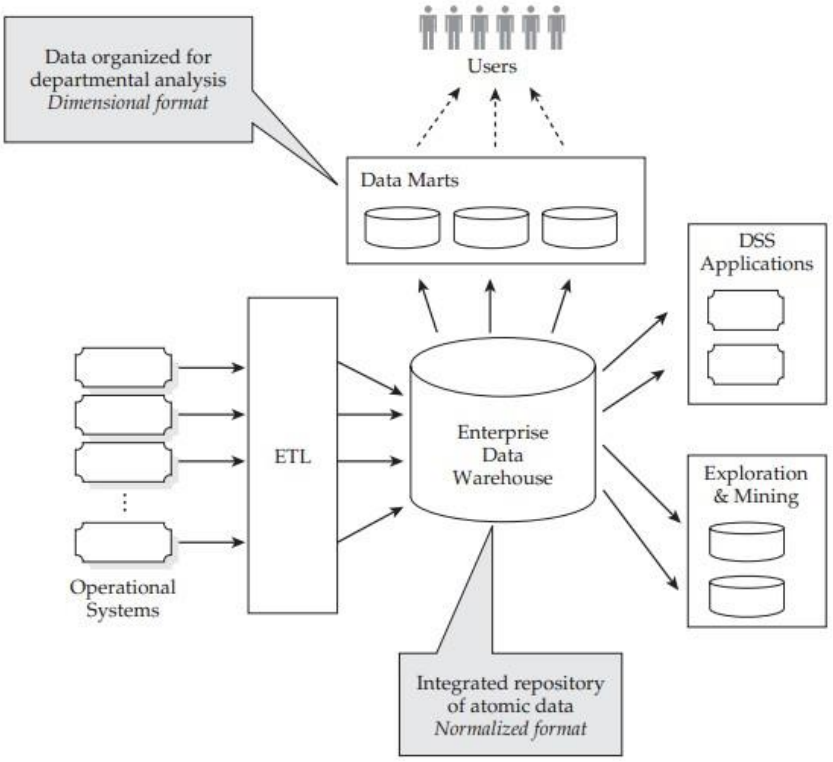


Figura 1.6: Modello di Inmon

Modello di Kimball - Dimensional Model: adotta un approccio Bottom-up in cui il DWH nasce dall'unione dei vari *Data Mart*, che riferiscono ognuno ad una specifica area di business[INM08].

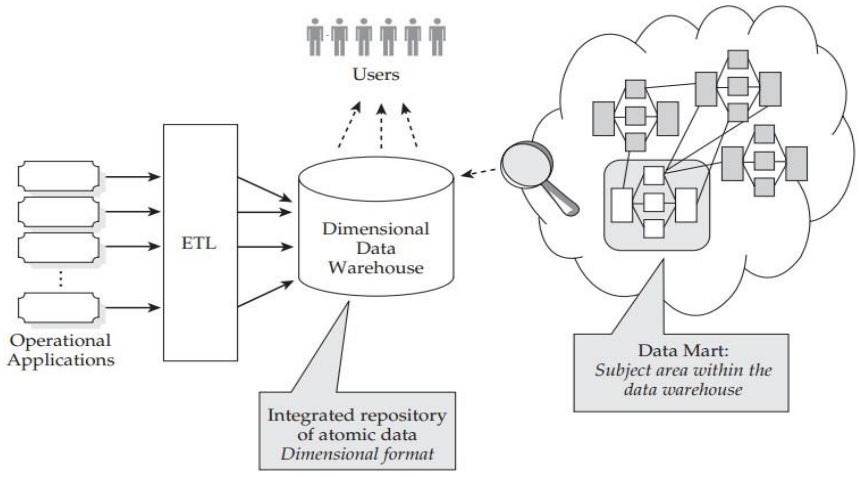


Figura 1.7: Modello di Kimball

1.1.4.3 ETL (Extraction Transformation Loading)

Come si può osservare nelle figure delle architetture precedenti, entrambi I sistemi presentano un flusso evolutivo comune: i dati vengono raccolti attraverso fonti interne e/o esterne dalle proprie aree di business per poi essere integrati nel DWH secondo un processo di pulizia **ETL (Extraction-Cleaning-Transformation-Loading)**. Andiamo ad analizzare nello specifico ognuna di queste fasi[RF04]:

- **Extraction:** I dati, strutturati o meno, vengono estratti dalle loro sorgenti informative, questa estrazione può essere di due tipi:

  -*Statica*: Il DWH viene popolato per la prima volta, quindi non ci sono problemi di integrazione con dati preesistenti;

  - *Incrementale*: Si incrementano i dati presenti nel DWH, aggiornando i dati presenti con le nuove versioni (senza cancellare o sovrascrivere nulla), serve quindi un timestamp o un trigger associato all'azione per conoscere l'epoca di appartenenza di un dato.

- **Cleaning:** Il dato viene pulito al fine di migliorare la qualità ed eliminare tutte le informazioni inconsistenti e inutili per la fase di analisi, in particolare:

  -Dati duplicati

  -Dati mancanti

  -Dati con attibuti errati

  -Valori impossibili

-Inconsistenza per errori lessicografici

- **Transformation:** Una volta pulito il dato viene trasformato in una struttura comprensibile da un DWH.

- **Loading:** Il dato viene finalmente caricato sul DWH, il processo può essere di:

  - *Refresh*: si riscrive completamente il DWH, usato solitamente in combinazione con estrazione statica;
  - *Update*: si modificano i dati aggiungendo quelli nuovi e mantenendo la consistenza con quelli presenti, i dati attuali non vengono cancellati o modificati. Solitamente usato in combo con l'estrazione Incrementale

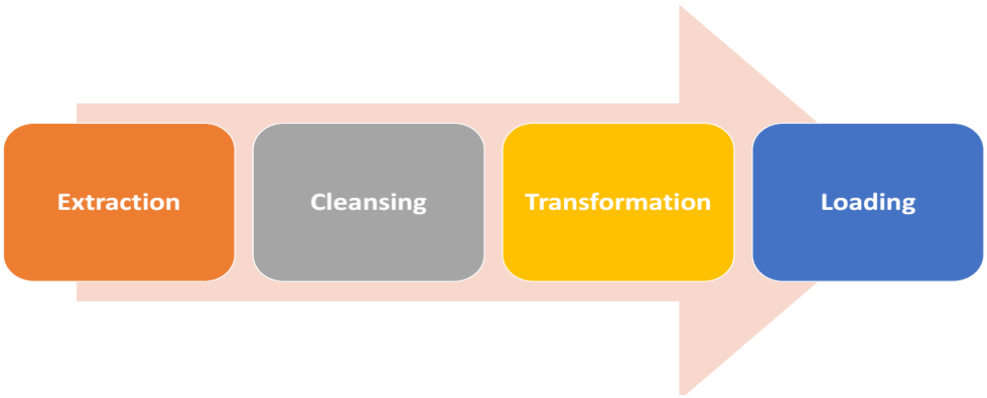


Figura 1.7: ETL Processing

1.1.4.4 Strumenti di analisi

Quando le informazioni sono state raccolte, trasformate, integrate e salvate all'interno del DWH si può passare a sfruttarle per tutti i processi di *decision making*. Attraverso gli *strumenti di analisi* si effettua l'accesso ai dati contenuti nel DWH e vengono permesse la creazione di report e l'indagine approfondita delle informazioni. Tra gli strumenti più importanti troviamo:

- **OLAP** *(On-Line Analytical Processing)*, ovvero tutte quelle tecniche di analisi attraverso cui è possibile navigare in modo semplice grandi quantità di dati, in

maniera rapida e interattiva [WW07].

- **Report e Cruscotti**, che hanno lo scopo di presentare graficamente le informazioni sotto forma di grafici e tabelle, in maniera facilmente comprensibile, per permettere di prendere velocemente decisioni.

- **Analisi What-If**, ossia quei metodi che consentono di determinare, attraverso delle simulazioni, quali possano essere i possibili scenari futuri, modificando alcune variabili.

- **Data Mining**, che applica tecniche statistiche e di intelligenza ar- tificiale per la ricerca di correlazioni significative, tendenze o pattern prima sconosciuti perché celati dietro la grande massa dei dati.

1.1.5 Modello Multidimensionale

Il modello E-R, diffuso per progettare sistemi informativi relazionali, non è adatto per esprimere e analizzare in modo dettagliato grosse moli di dati del business; da qui la necessità di adottare un nuovo modello concettuale: il **modello multidimensionale** chiamato **DFM (Dimensional Fact Model) [RF04].**

Questo modello è stato creato per supportare analisi sulle relative coordinate di analisi.

Attraverso questa modellazione è possibile rappresentare i dati all'interno di ipercubi che forniscono immediatamente il raggio di azione del DWH: vengono evidenziate istantaneamente le **dimensioni** di analisi e i **fatti** di interesse legati al business.

Gli scopi principali del sistema sono:

- Fornire supporto al design concettuale;
- Creare un ambiente dove gli utenti possano fare query in maniera intuitiva e formale;
- Favorire la comunicazione tra designer a utenti al fine di formalizzare i requisiti di progetto;

- Costruire un stabile piattaforma di design logico;
- Fornire una documentazione chiara e efficace.

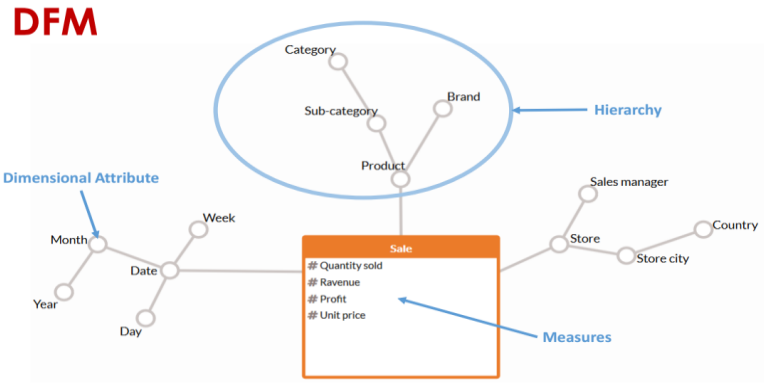


Figura 1.8: Struttura DFM

1.1.5.1 Concetti chiave di un DFM

- *Fatto:* concetto rilevante per il processo di decision-making, tipicamente modella un specifica area di business (Vendite,Ordini ecc.).

- *Misura:* rappresenta l'aspetto quantitativo del fatto che risulta di elevata importanza per l'analisi, proprio dalle *Misure* vengono estratti i *KPI (Key Performance Indicator)* che guideranno le imprese nelle proprie strategie di business. Esempio (Quantità venduta, Sconti, Profitto…)

- *Dimensione:* rappresenta la coordinate di analisi del *Fatto*. Esempio (Data, Prodotto,Negozio)

- *Attributo Dimensionale: Dimensioni* riletive alle *Dimensioni* del *Fatto*, hanno un dominio discreto. Esempio (Categoria di prodotto, Città del Negozio,Mese ecc)

1.1.5.2 Concetti avanzati di un DFM

- *Attributi Descrittivi:* sono unsati per aggiungere informazioni a un attributo dimensionale ma non sono usati come criteri di aggrtegazione.

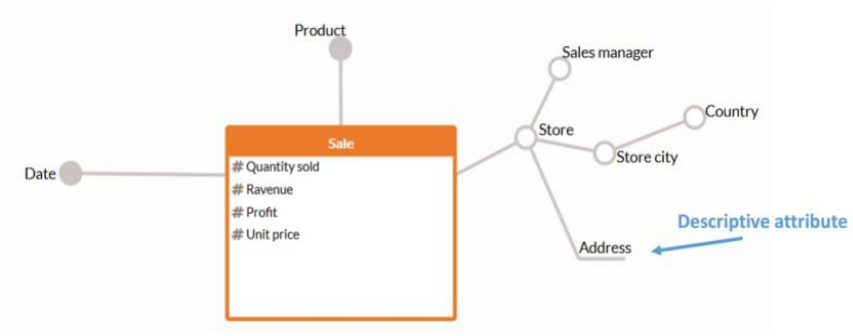


Figura 1.9: Attributo descrittivo

- *Attributi Cross-Dimensionali:* sono attributi dimensionali o descrittivi il cui valore è definito dalla combinazione di due o più attribute dimensionali, anche appartenenti a diverse gerarchie

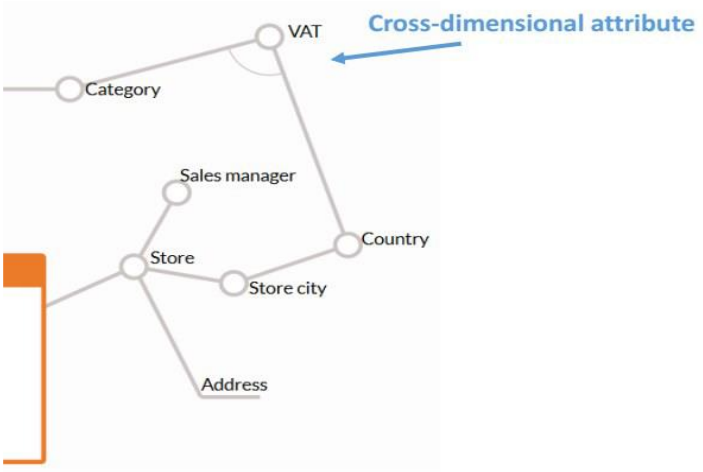


Figura 1.10: Attributo cross-dimensionale

- *Convergenza:* Si ha quando due o più archi, appartenenti alla stessa gerarchia, portano allo stesso attributo dimensionale.

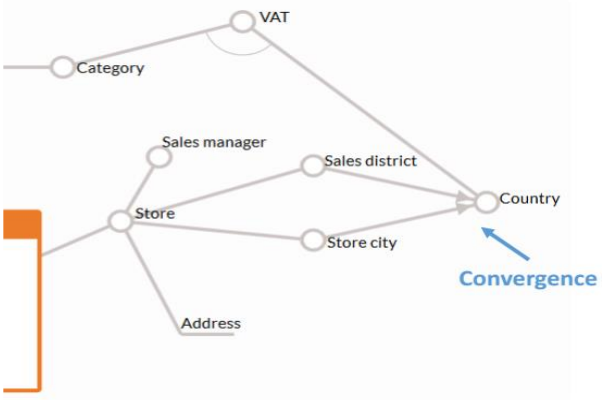


Figura 1.11: Convergenza

- *Gerarchia condivisa:* Un doppio cerchio rappresenta e enfatizza il primo attributo condiviso dalle gerarchie. Tutti gli attributi discendenti da quello sarranno condivisi. Per ogni arco entrante deve essere definito un "ruolo".

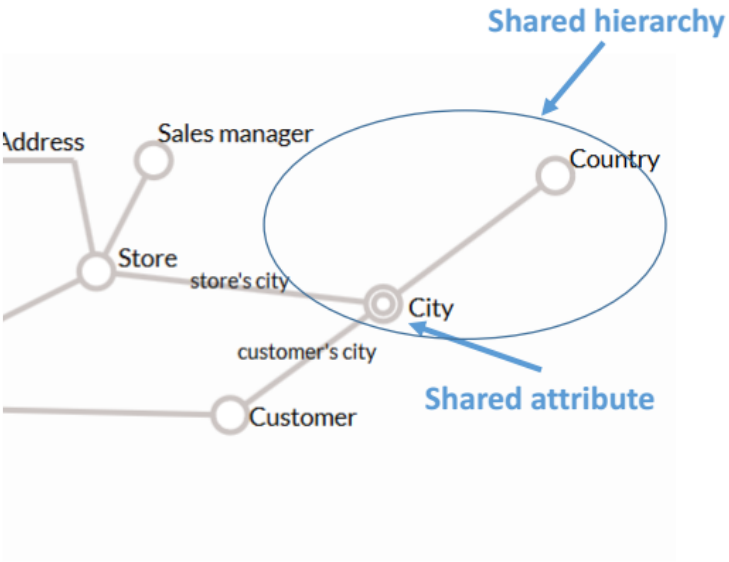


Figura 1.12: Gerarchia condivisa

- *Gerarchia ricorsiva:* sono rappresentate da relazioni padre-figlio. Sono rappresentati con un doppio cerchio proprio per evitare il problema del "*double counting*". L'esempio descrive la gerarchia di un ambiente di lavoro per ruoli di importanza. Uno dei più classici KPI che si può estrarre è la somma degli stipendi di tutti I dipendenti di un'azienda. Il totale è dato dalla somma di tutti i

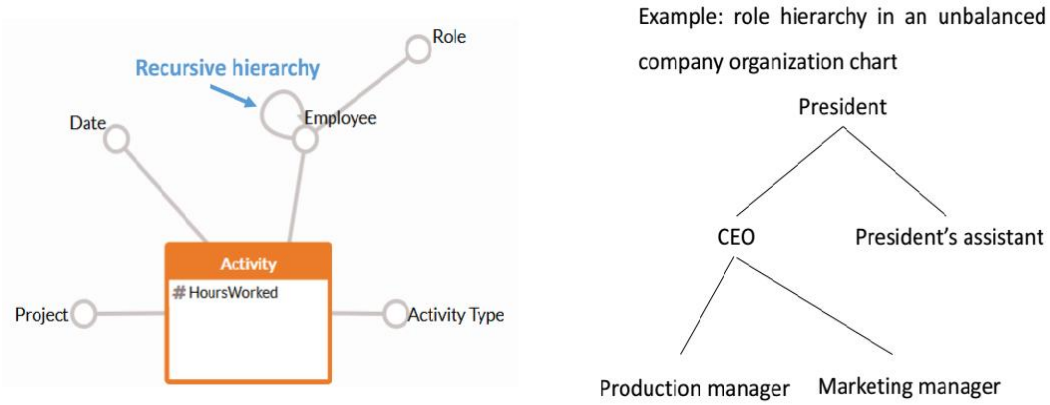


Figura 1.12: Gerarchia ricorsiva

Manager, Team Leader e consulenti. Il Manager, essendo esso stesso un dipendente deve risultare nel conteggio una volta sola, quindi per risolvere il problema del "*double counting*" si usa la *Gerarchia condivisa.*

1.1.5.3 Star Schema e Snowflake schema

A questo punto dato il *DFM* si definisce la struttura dei dati all'interno dei Data Mart in accordo con il modello logico scelto, al fine di ottimizzare le performance.
Possiamo suddividere la fase in tre step logici:

- Si convertono i fatti in schemi logici;
- Si definiscono viste secondarie che aggregano i fatti primari al fine di migliorare le performances della query;
- Si frammentano le tabelle verticalmente e orizzontalmente.

Esistono tre approcci per implementare i *Data Warehouse*:

- *Relational OLAP (ROLAP):* implementazione basata su DBMSs, cioè su un modello relazionale che include i concetti dei DWH.

- *Multidimensional OLAP (MOLAP):* implementazione basata su DBMSs multidimensionali, fortemente dipendente dai concetti dei DWH, ha performances di query molto elevate.

- *Hybrid OLAP(HOLAP):* modello ibrido rispetto ai due precedenti.

Una volta costruito il *Dimensional Fact Model* si passa allo schema logico sotto rappresentazione di **Star Schema** dove i fatti e le dimensioni diventano delle tabelle e ogni tabella include tutti gli attributi della gerarchia.

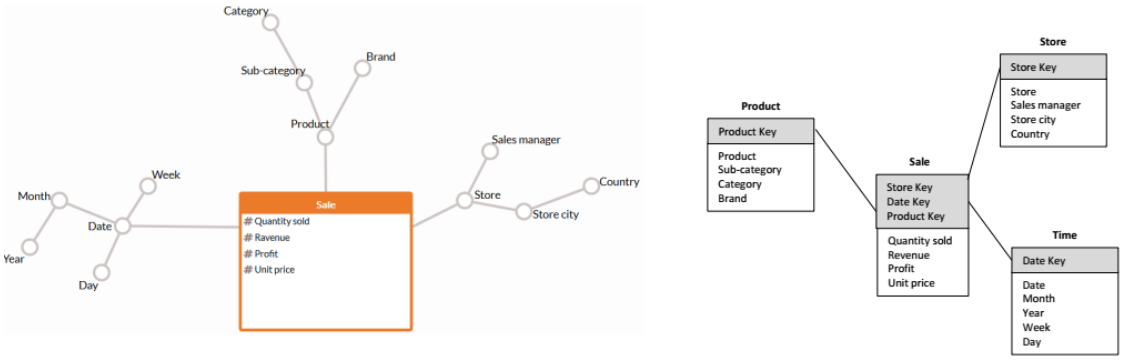


Figura 1.13: Star Schema

Se invece le tabelle dimensionali sono normalizzate allora viene definito **Snowflake Schema**

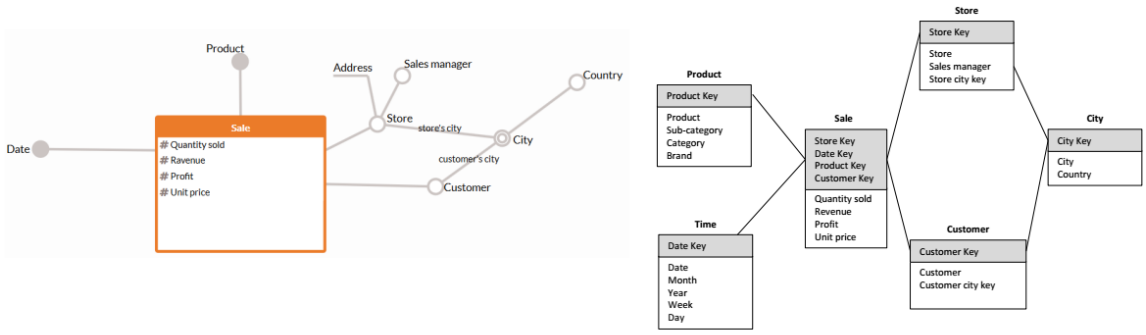


Figura 1.14: Snowflake Schema

1.1.6 OLAP

Come anticipato in precedenza una delle caratteristiche principali in un architettura di un DWH consiste nel cercare di separare il più possibile l'analisi transazionale da quella analitica.

L'Analisi **OLAP**(**OnLine Analytical Processing**) permette all'utente di navigare in maniera interattiva le informazioni contenute nel DWH. Tipicamente i dati sono analizzati a differenti livelli di aggregazione applicando operatori OLAP in sequenza, ognuno composto da una o più differenti query. OLAP si contrappone al metodo tradizionale OLTP (OnLine Transactional

Processing) basato su database relazionali.
In una Sessione OLAP l'utente può scandagliare il modello multidimensionale scegliendo il successivo operatore da utilizzare, basandosi sul risultato dell'operatore precedente. In questo modo gli utenti creano un percorso di
navigazione che corrisponde a un processo di analisi dei fatti appartenenti a diversi punti e diversi livelli di dettaglio.
A tal proposito vengono offerti alcuni Operatori OLAP[RF04]:

- *Roll-Up:* Provoca un aumento di aggregazione dei dati rimuovendo un livello di dettaglio dalla gerarchia

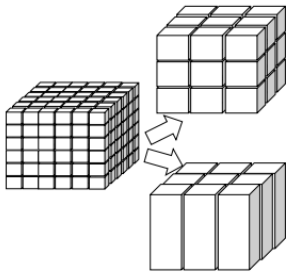


Figura 1.15: Roll-up

- *Drill-Down:* Procedura inversa del Roll-Up, diminiusce il grado di aggregazione, aggiungengo un nuovo livello di dettaglio alla gerarchia

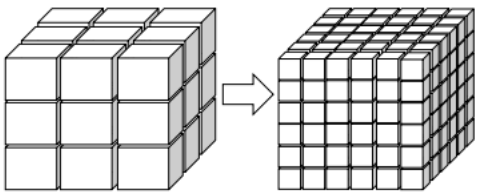


Figura 1.16: Drill-down

- *Slice and Dice :* Riduce il numero delle dimensioni del cubo multidimensionale dopo aver settato una dimensione a uno specifico valore (per esempio imposto "Categoria Film = Horror"); la procedura di Dicing riduce la dimensione dei dati sulla base di un filtro multidimensionale.

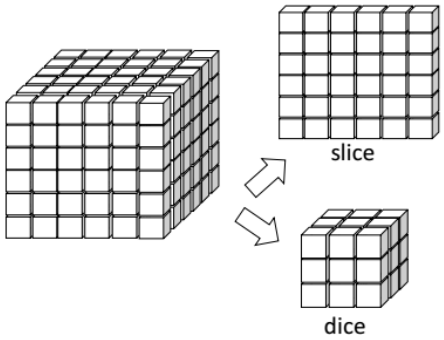


Figura 1.17: Slice and Dice

- *Pivoting:* Cambia il modo di vedere I dati per effettuare differenti analisi in base ai punti di vista

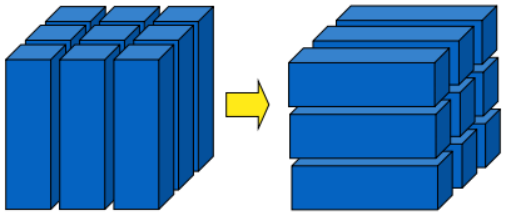


Figura 1.18: Pivoting

- *Drill-Across:* Permette di fare connessione tra I fatti di cubi correlate e fare confronti

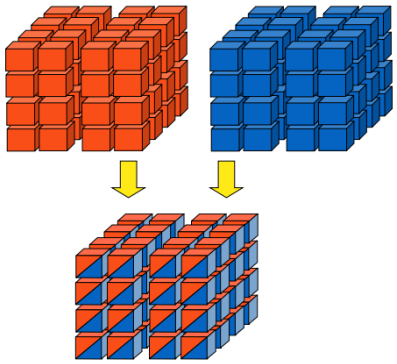


Figura 1.19: Drill-Across

1.1.7 Data Visualization

Abbiamo studiato ed analizzato come estrarre le informazioni dai dati, tuttavia se non siamo in grado di presentarle in maniera efficace ed efficente l'utilità di queste informazioni potrebbe essere nettamente minore del previsto. Va quindi studiato qual'è il modo più opportune di visualizzare i dati in modo da *trasmettere in maniera immediata ed intuitiva il messaggio da trasmettere.*
Il dato inoltre deve essere presentato nel modo adatto senza sporcare il contenuto informative che trasmette.
Una buona visualizzazione aiuta ad identificare velocemente i punti chiave della ricerca e focalizzarsi sul d*ecision-making.*
Tutto sta nel saper scegliere il mezzo di comunicazione più opportuno in base al tipo di dato, ricordando di essere certi di rappresentare [RF10]:

- *Qualità:* presentare solo le informazioni necessarie di grande importanza;
- *Quantità:* mostrare sempre sia i punti di forza che le incertezze in modo da effettuare nuove analisi;
- *Rilevanza:* Mostrare solo le informazioni pertinenti al contesto di analisi, evitare informazioni ridondanti e inutili.

1.1.7.1 Rappresentazione delle informazioni

Esistono vari modi per rappresentare i *KPI,* come già accennato precedentemente è importante capire quale sia il modo più adatto a rappresentare l'informazione in base al tipo di dato.
In particolare:

-*Tabelle:* rappresentano dati per righe e colonne. Sono adatte per:

- Dare attenzione al valore numerico;
- Confrontare numeri;
- Mostrare valori precisi;
- Includere unità di misura;
- Mostrare dettagli e somme.

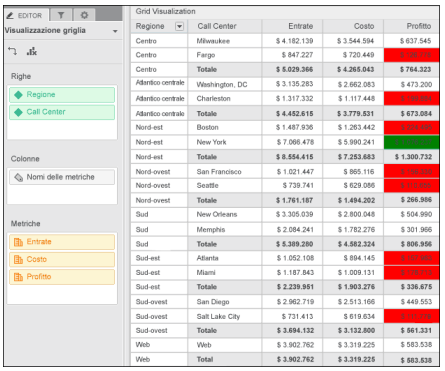


Figura 1.20: Tabella

-*Grafici:* Rappresentano i dati su assi cartesiani, o su diagrammi a torta.
Sono adatti per:

- Mostrare pattern;
- Mostrare Trend;
- Mostrare Eccezioni

Figura 1.21: Grafici

-*Report:* un generico documento contenente dati consultabili, presentati attraverso tabelle o grafici. I KPI vengono percepiti in maniera immediata.
La standardizzazione dei documenti consente inoltre - secondo l'approccio all'informazione come bene

Figura 1.22: Report

aziendale – una miglior distribuzione delle conoscenze ed una visione dell'attività più conforme e concorde fra le varie funzioni dell'organizzazione, oltreché aggiornata secondo la disponibilità della fonte - o delle fonti - dei dati.La rappresentazione dei KPI avviene in maniera *statica.* I report infatti spesso vengono stampati, letti, documentati e analizzati.

-*Dashboard:* Si definisce *dashboard*, o *cruscotto*, un contenitore di report e altri elementi di analisi, contenenti grafici quali *istogrammi*, *mappe* e *gauge* in modo da essere intuitivi, pertinenti e di facile comprensione. Le dashboard dovrebbero offrire, sulla base del ruolo dell'utente, una panoramica più o meno generale di un ambito di business, presentandogli i dati nel modo più efficace possibile.

La caratteristica principale delle dashboard è data da una schermata che ti permette di monitorare in tempo reale l'andamento dei report e delle metriche aziendali più importanti permettono ai reparti marketing e vendite di essere sempre allineati in merito ai dati più importanti. La rappresentazione

Figura 1.23: Dashboard

dell'informazione avviene quindi in maniera *dinamica* a differenza della staticità dei semplici report.

1.2 Location Intelligence

E' un dato di fatto che le informazioni estratte dai dati di business raccontano una storia che avviene in uno *spazio* e in un determinato *tempo*. Rimane quindi riduttivo rappresentare le informazioni su semplici tabelle o grafici. Esempi evidenti di attività per cui questo genere di dati costituisce un ruolo di fondamentale importanza sono aziende di telecomunicazioni, agenzie immobiliari ed assicurative o società di trasporti, per cui riuscire a sfruttare appieno queste informazioni può significare un notevole impatto a livello di efficienza e, di conseguenza, in una potenziale riduzione dei costi. Per questo motivo definire una *dimensione spaziale* di rappresentazione del dato fornisce un alto livello di astrazione dell'informazione che permette di:

- facilitare le analisi;
- estrrarre anomalie o punti di interesse in maniera veloce e intuitiva;
- velocizzare il processo di decision-making;

Ecco quindi che nasce la **Location Intelligence,** il principale supporto alla Business Intelligence geo-spaziale.

1.2.1 Definizione e contesto

La Location Intelligence è l'estensione della Business Intelligence tradizionale con l'aggiunta della dimensione spaziale. Si guadagna la capacità di unire le tipiche visualizzazioni a mappa con le informazioni ottenute tramite i tradizionali metodi di BI.Con il termine Location Intelligence, si fa riferimento alla capacità di organizzare e comprendere processi, tecnologie e applicazioni che permettono di mettere in relazione dati spaziali elaborati da sistemi GIS, con dati di business trattati da applicazioni di BI, al fine di sviluppare funzionalita di analisi più complete, supportare decisioni efficaci ed ottimizzare le attività di business. [RF06]

La *Location Intelligence* nasce dall'integrazione in un unico Sistema di *Supporto alle Decisioni (DSS)* degli aspetti caratteristici di un sistema di *Business Intelligence* e di un sistema *Geographical Information System (GIS).*

Questo binomio in un unico sistema è dato dall'esigenza di riuscire a dare risposta a quelle analisi che cercano di mettere in luce una relazione tra gli eventi di business e le caratteristiche del contesto geografico in cui si verificano gli eventi stessi.

La Location Intelligence è veramente utile nei casi in cui si riescano ad aggregare grandi quantità di dati all'interno di indicatori numerici di sintesi, da poter collocare su mappe interattive, per la comprensione di pattern spaziali che non si sarebbero mai potuti evincere, con la stessa pragmaticità, attraverso l'utilizzo di normali tabelle.

1.2.2 Cenni storici

I principi della LI, se pur molto recenti e innovativi, furono già anticipati e utilizzati nel 1854 dal Dr. John Snow. In quel periodo in tutta Europa erano frequenti molte epidemie di colera dovute alle scarse condizioni igieniche, anche Londra non fu risparmiata e si contarono più di 15.000 deceduti a causa della malattia. Durante questo episodio di epidemia il Dr. John Snow si mise a ricercare e a studiare le cause della diffusione così repentina del colera e notò che il maggior numero di casi erano situati in quartieri londinesi che erano serviti da due società di approvvigionamento idrico, da qui il sospetto di un veleno contenuto dell'acqua. Per valorizzare la sua tesi Snow sovrappose in un unica mappa di Londra i casi registrati di colera e le pompe di acqua utilizzate dai londinesi.
Questa rappresentazione visiva dei dati mise subito in luce il fatto che il maggior numero di casi erano in prossimità della pompa dell'acqua posizionata a Broad Street e che altri casi piu lontani erano comunque collegati dal fatto che usufruivano di tale pompa. Tramite i risultati portati da questa analisi, venne chiusa la pompa dell'acqua di Broad Street limitando drasticamente il diffondersi della malattia. La figura 1.24 mostra la Mappa della zona di Broad Street (Londra) che mette in relazione i casi di colera con le pompe di acqua. Gli studi e i risultati ottenuti dall'analisi di Snow evidenziano come la componente spaziale, se messa in relazione con eventi, possa portare all'estrapolazione di informazioni implicite in grado di arricchire l'analisi e di supportare il decisore con un nuovo punto di vista.Il sistema informativo territoriale che è alla base del sistema ideato dal Dr. Snow fu utilizzo in larga scala, soprattutto per contribuire a spiegare fenomeni complessi. Le tecniche furono via via raffinate passando dall'utilizzare mappe cartacee a mappe separate in più livelli fino ad arrivare al mapping digitale con i primi computer degli anni 60.

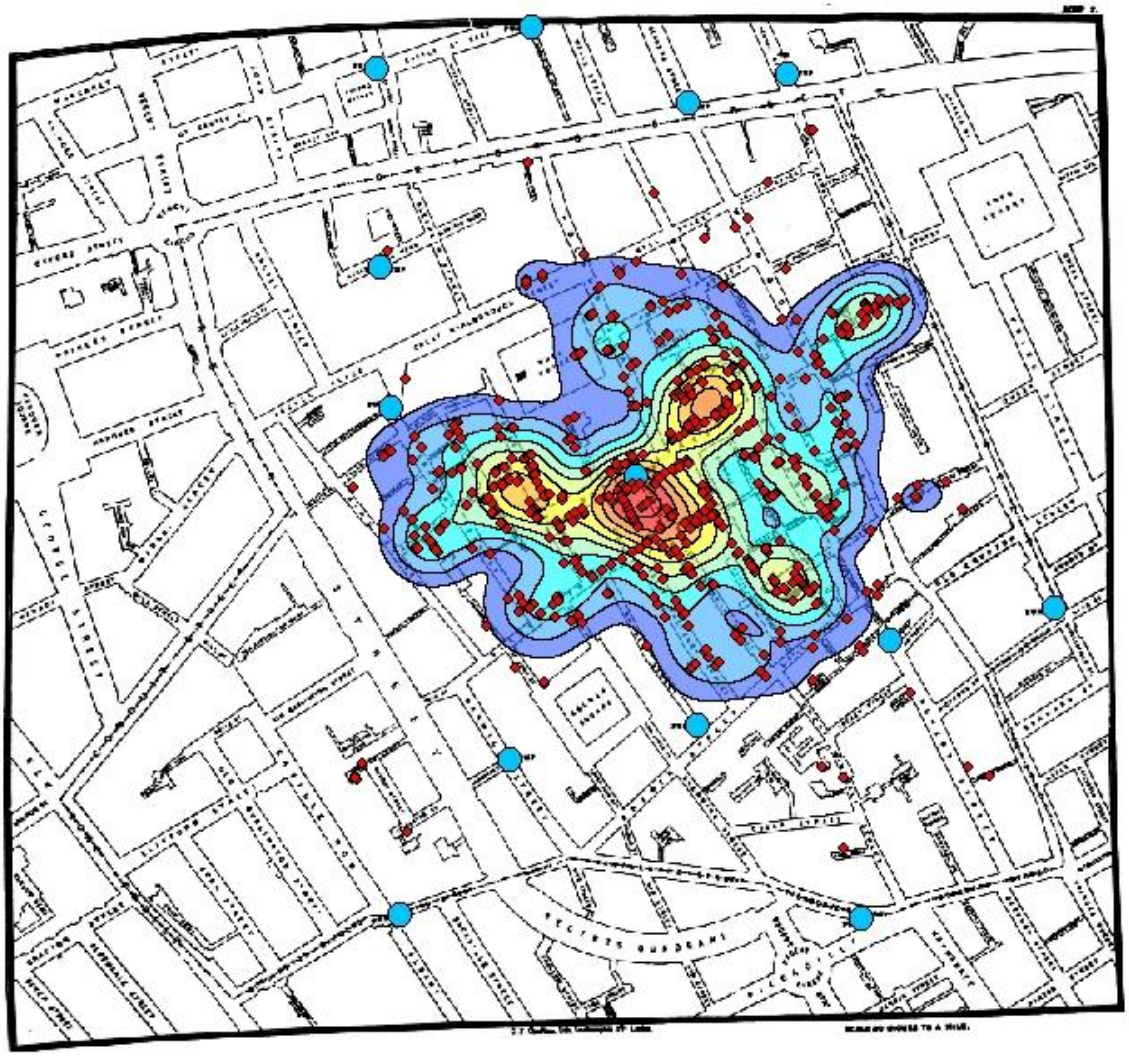


Figura 1.24: Casi di colera nella zona di Broad Street (Londra)

Successivamente negli anni 80 ci furono i primi sviluppatori dei sistemi GIS come ESRI che misero sul mercato applicazioni per utenti esperti e solo negli ultimi anni con l'arrivo di applicazioni GIS fruibili attraverso internet come Google Maps, aperto ad un pubblico più ampio, che si è iniziato a capire la potenzialità offerta da piattaforme di LI. [AAS04].

1.2.3 Geographical Information System (GIS)

Con il termine *Geographical Information System* (GIS) si intende un sistema informativo progettato per la ricezione, l'immagazzinamento, l'elabo- razione, la gestione, l'analisi e la presentazione di tutti i dati caratterizzati da dimensioni geografiche[CLA97].

Esso rappresenta quindi un importante strumento utile all'individuazione di pattern nascosti

all'interno dei dati spaziali, i quali possono essere fondamentalmente di due tipi:

- **Informazioni geometriche**: relative a caratteristiche metriche, quali *posizione, forma, dimensione e distribuzione*
- **Informazioni topologiche**: relative alle relazioni tra i vari oggetti geografici, quali *connessioni, adiacenze, inclusioni*, etc...

L'architettura GIS si compone di una rappresentazione gearchica, ovvero composto da tre diversi strati, ognuno dei quali dedicato rispettivamente a *interfaccia utente, logica funzionale e gestione dei dati[PLGR05] :*

1. **Interfaccia utente**: livello di presentazione contenente tutti i compo- nenti atti alla visualizzazione dei dati e ai controlli per l'interazione.
2. **Logica Funzionale**: livello che permette il collegamento tra l'inter- faccia grafica e il database sottostante, fornendo alla prima tutte le funzionalità di analisi necessarie all'utente finale.
3. **Gestione dei dati**: livello che si occupa della modellazione e dell'ar- chiviazione dei dati, ottimizzandone l'accesso per servire le richieste che provengono dal livello superiore.

1.2.4 Rappresentazione dell'informazione geografica

Uno dei problemi con cui si deve scontrare un sistema GIS è quello della rappresentazione del contesto geografico e delle caratteristiche che lo descrivono, attraverso la costruzione di un modello digitale in grado di catturare tutte le peculiarità. La scelta di come rappresentare tali informazioni diventa fondamentale, in quanto definisce il livello di dettaglio con il quale andrà a lavorare il sistema. Tipicamente vengono utilizzati due metodi per la rappresentazione e la memorizzazione dei dati all'interno di un sistema GIS:

- *Raster* : adatto per rappresentare temi che variano in modo continuo

sul dominio spaziale come le proprietà fisiche del territorio ad esempio elevazione, temperatura, tipo di suolo, ecc.. I luoghi sono referenziati per mezzo di una griglia di celle un array bidimensionale e gli attributi sono rappresentati come valori associati alle celle. I dati di solito derivano da immagini da Remote Sensing o da mappe scannerizzate.

- *Vettoriale*: adatto per rappresentare temi con confini ben definiti sul dominio spaziale come confini politici, confini amministrativi, strade, ecc.. I luoghi sono referenziati da una coppia di coordinate (X,Y) e possono essere collegate per formare linee e poligoni, gli attributi sono referenziati per mezzo di identificatori univoci in tabelle.

Le informazioni geografiche modellate dal GIS sono rappresentate in data layer o layer tematici, utilizzati per descrive una singola caratteristica dell'area geografica che si vuole rappresentare. Al fine di fornire una modellazione dettagliata e precisa i layer vengono aggregati in collezioni e messi in relazione tra loro tramite collegamenti e sovrapposizione geografica. Come mostra la Figura 1.25, sono stati definiti più layer, con metodologie differenti e rappresentativi di singoli aspetti, per essere combinati tra loro al fine di avere in ununico modello molteplici aspetti che con un singolo layer non sarebbe statopossibile descrivere.

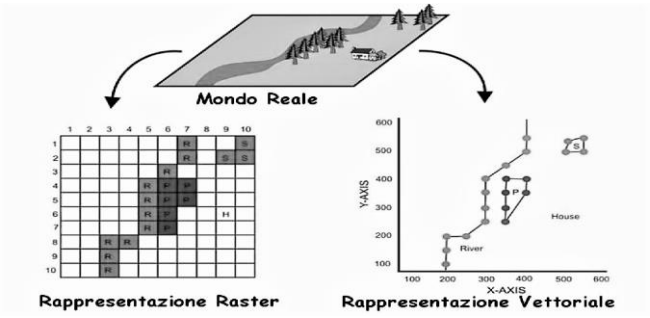


Figura 1.25: Rappresentazione del mondo reale attraverso le tecniche Raster e Vettoriale

1.2.5 Architettura di un sistema di LI

In base alle considerazioni appena fatte per costruire un architettura adatta a trattare questi tipi di dati è necessario rispettare le caratteristiche relative a un DWH aggregando la

dimensione spaziale. I principali sistemi archietturali adottati da un Sistema di Location Intelligence sono [RF06]:

- *Bridge:* mantiene separato il Sistema di un DWH da un Sistema GIS e quindi i due mondi comunicano attraverso dei server;

- *Geo-Data Warehouse:* viene proposto un Sistema architetturale unico dove la dimensione spaziale viene aggregata al DWH permettendo di avere un livello di astrazione altissimo e di grande rappresentanza.

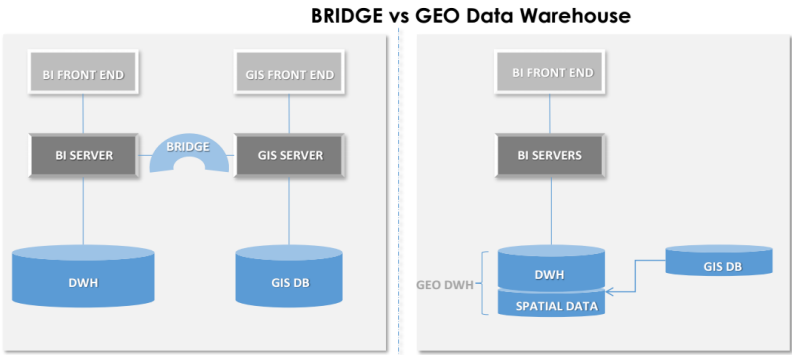


Figura 1.26: Architettura Bridge Vs Geo Data Warehouse

Includere la dimensione geografica in un DWH da la possibilità di aggiungere innumerevoli nuove possibilità di aggregare elementi e di fare query miste.
Il punto focale della Location Intelligence sta nel riconoscere dei KPI complessi, ricavati attraverso query miste. Infatti molto spesso è possibile fare un merge tra le informazioni sui dati di business con informazioni legate al terriotorio come ad esempio popolazione, stipendio medio, flussi turistici, dati social ecc.Per questi motivi è stato necessario adattare i classici *DBMS relazionali (RDBMS)*, estendendoli con funzionalità che permettessero la memorizzazione, l'elaborazione e l'accesso efficiente a enormi moli di informazioni spaziali [SRV01]; si è arrivati così a parlare di *Spatial DBMS (SDBMS)*, attraverso i quali, con le stesse analogie relative la BI, otterremo un'infra struttura di *Geo-Data Warehouse* e si rivedrà ad esempio il concetto di ETL in *Spatial ETL.*

1.3 Social Media Intelligence

Negli ultimi vent'anni il tipo di dato che ha riscontrato un impatto evolutivo dinamico e veloce è senza dubbio il *dato social*. Grazie all'utilizzo dei social network infatti l'informazione viene trasmessa a velocità lampo in ogni parte del mondo.
L'informazione generata dall'uomo è ormai interamente digitalizzata e salvata in locale e in rete. Questo tipo di dati debolmente strutturati e spesso indefiniti viene classificata come *User Generated Content.* Informazioni di questo tipo vengono da:

- *Social Network*: Facebook, Instagram,Twitter ecc…
- *Video*: Youtube ecc
- *Motori di ricerca*: Google, Bing, Firefox
- *Immagini:* foto
- *Messaggistica:* E-mail,chat e messaggi

Figura 1.27: Social Media Intelligence

1.3.1 Definizione e contesto

Con il termine Social Media Analysis si intende l'insieme di tool mirati al collezionare, monitorare, analizzare e visualizzare dati social permettendo di effettuare un analisi per supportare il decision-making dall'User Generated Content (UCG)[RF09].

Il tipo di dato generato dall'utente, essendo non strutturato, è un tipo di dato per cui occorre un analisi molto accurata, esso è:

- *Real-time*
- *Non anonimo*
- *Geo-localizzato*

In base al tipo di dato possiamo identificare due diversi tipi di Social BI:

- *Brand:* si vanno ad analizzare i dati in ordine cronologico
- *Industriale:* range applicativo molto ampio, dai trasporti al commercio fino ai trend di moda, dove le quantità di UGC variano molto in base all'impresa, solitamente si hanno grandi volume di dati per marchi commerciali, bassi volume per marchi finanziari.

1.3.2 Fasi salienti

Le fasi della Social Media Analysis sono riassunte in sei fasi salienti [RF09]:

1. *Mapping dei KPI:* Ogni azienda deduce i propri Key-Performances Indicator dalla strategia che ha adottato, i KPI determinano la categoria da analizzare e, di conseguenza le parole e i termini che possono essere utilizzate in quell'ambito. Inoltre non tutti i KPI dedotti possono essere valide misure dell'opinione della popolazione riguardo l'azienda, bisogna valutare i KPI validi disponibili. Per esempio, un'azienda di elettronica non andrà a valutare gli UGC che contengono

parole come “cotoletta” o “cravatta”. La regola base è quella di selezionare i KPI connessi a relazioni con il cliente, prodotti e immagine dell’azienda;

2. *Collecting dei KPI:* È necessario identificare le piattaforme dalle quali estrarre UGC, il tipo di comunità e di discussione dipenderà dalla piattaforma, per esempio in un Social Network solitamente le discussioni sono breve ma spesso contenenti molte informazioni, in un blog potrebbero invece esserci discussioni molto lunghe e poco interessanti

3. *Pre-Processing:* I dati estratti, solitamente molto eterogenei, vengono puliti, resi anonimi e aggregati secondo differenti attributi di analisi e processi ETL;

4. *Processing:* Vengono effettivamente analizzati I dati attraverso tecniche mirate a estrarre pattern al loro interno come:

   - *Natural Language Processing*: il processo di trattamento automatico mediante un calcolatore elettronico delle informazioni scritte o parlate in una lingua naturale.

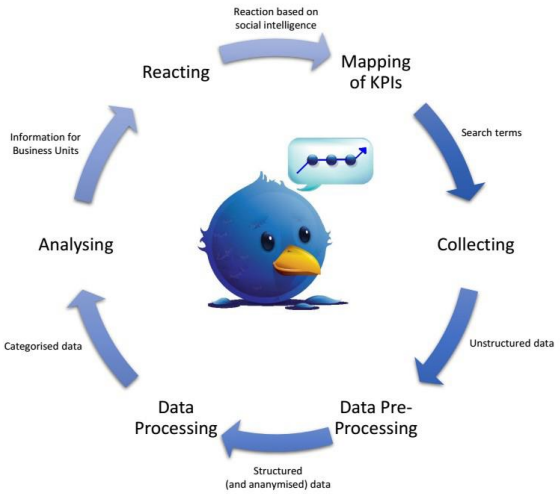
   - *World Sense Disambiguation:* basate su ontologie e modelli statistici

   -*Sentiment Analysis:* è la maniera a cui ci si riferisce all'uso dell'elaborazione del linguaggio naturale, analisi testuale e linguistica computazionale per identificare ed estrarre informazioni soggettive da diverse fonti.

5. *Analysis:* Una volta analizzati i dati vengono estratte le informazioni chiave utili per la Business Intelligenece, queste vengono memorizzate in DWH e visualizzate con le tecniche di Data Visualizzation;

6. *Reaction:* Si utilizzano infine le informazioni per prendere decisioni, magari

rivisitando un prodotto o intervenendo in discussioni sui social. Potrebbero inoltre cambiare i KPI di interesse o determinane altri nuovi e complessi per l'analisi successiva



1.28: Fasi salient Social Media Analysis

1.3.3 Social Media Architecture

Come abbiamo già anticipato questo tipo di dati, essendo non strutturato, è il più difficile da analizzare, per questo motivo abbiamo bisogno di un architettura robusta che possa supportare un processo di estrazione delle informazioni per I *Decision Making Support*.
L'architettura proposta si compone di tre macro-fasi presentate nella figura sottostante.

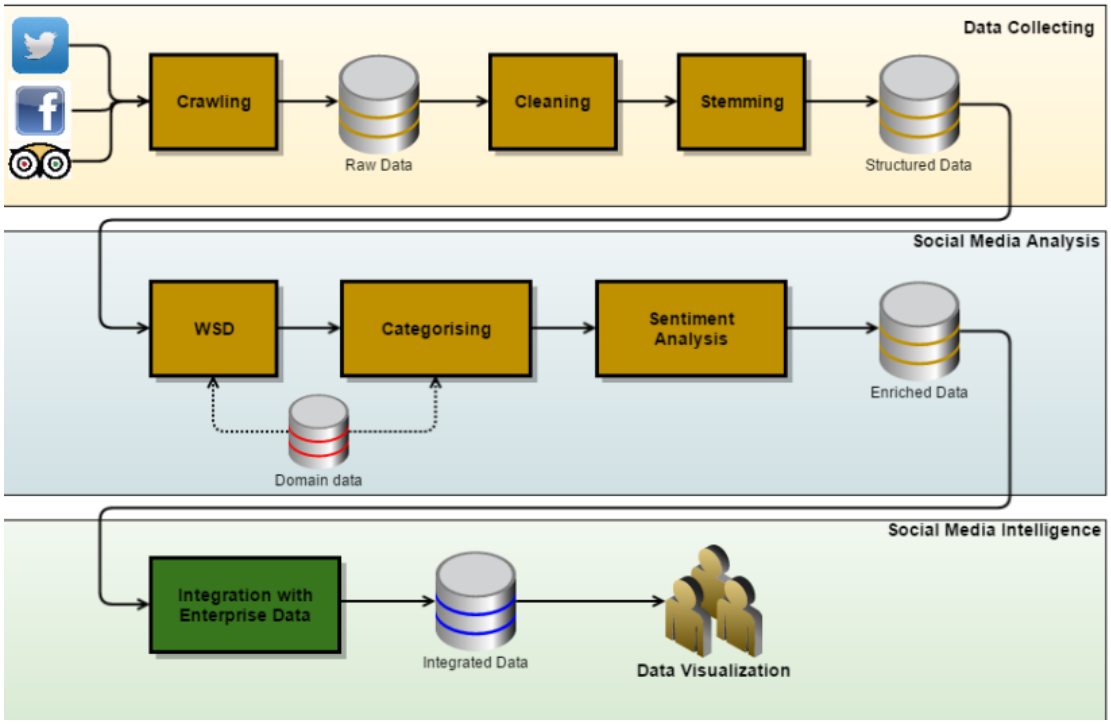


Figura 1.29: Social Media Intelligence processing

Andiamo ad analizzare ognuna delle fasi nello specifico:

1. *Data Collecting:* Rappresenta il processo di raccolta dei dati social,vengono effettuate le seguenti operazioni:

    - *Crawling:* rappresenta la fase di raccolta dei dati, solitamente per KeyWord.

    - *Cleaning:* Consiste ovviamente nella pulizia dei dati, essendo generati dall'utente sono quasi sicuramente molto sporchi e senza dubbio non strutturati.

    - *Stemming:* Processo che riduce le parole, modificate o derivate, alla loro forma base base o forma radice, per esempio "I play Game**s**"
      diventa "I play Game", un tool molto usato è Freeling.

2. *Social Media Analysis:* Dopo una prima fase di pulizia il dato ora è strutturato e può essere sottoposto ad analisi complesse:

- *Word Sense Disambiguation(WSD):* Tecnica che permette di comprendere il significato di un parola in un determinato contesto. A supporto di questo meccanismo interviene *il Social Media Mining*.
  Il *Social Media Mining (SMM)* rappresenta il processo di estrazione della conoscenza dai dati social utilizzando tecniche e algoritmi di *Data Mining* che come sappiamo possono essere:

  - *Supervisionati*: restituiscono risultati molto accurati ma hanno un costo elevato, soprattutto a causa del mantenimento di molteplici database in base al contesto
  - *Non-supervisionati:* sono totalmente knowledge-free
    ma permettono di trovare in maniera relativamente semplice tutti i termini che non sono dipendenti dal contesto, risultano quindi meno performanti degli altri. Le tecniche non supervisionate fanno largo uso delle Ontologie, delle risorse lessicali organizzate in ordine alfabetico (*Dizionario*), o raggruppate per significato (*Thesaurus*), a volte mischiate tra di loro.
- *Categorising:* una volta analizzate le parole all'interno di un contesto inzia la fase di categorizzazione in base al contestopermettendo di raggruppare tutte le parole che appartengono un contesto specifico comune.

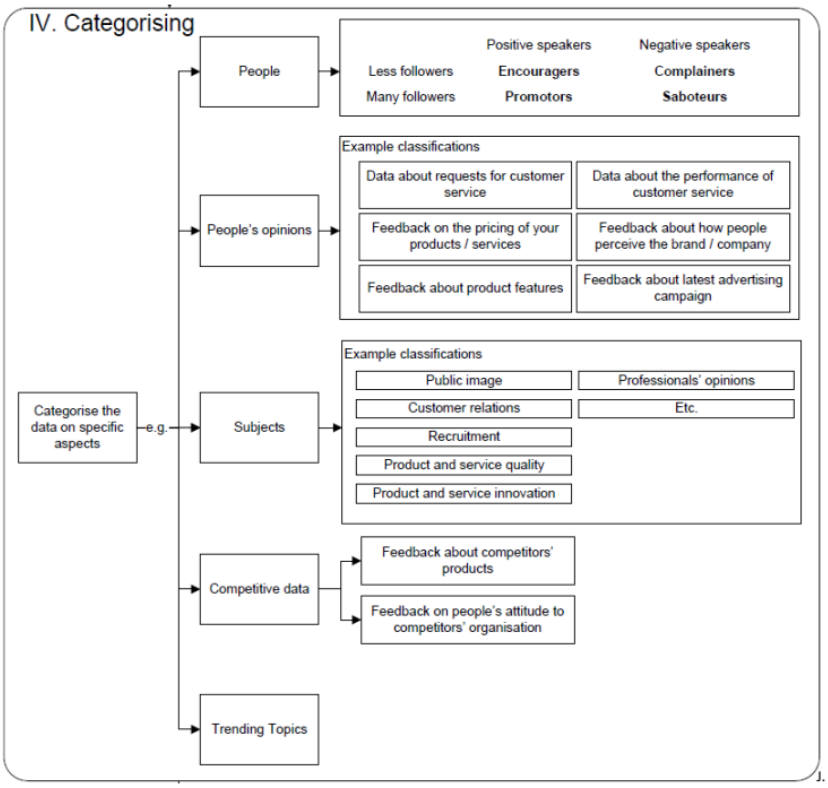


Figura 1.30: Categorasing

- *Sentiment Analysis:* Insieme di tecniche di Natual Language Process (NLP) atte ad automatizzare l'estrazione della polarità dell'opinione pubblica da dati. I problemi principali stanno nel capire se, una determinata parola, in un determinato contesto, ha un significato positivo o negativo. Per esempio:

  - "Il tempo di risposta è molto LUNGO", ha significato negativo;
  - "La durata delal batteria è molto LUNGA", ha significato positivo.

3. *Social Media Intelligence:* dopo le relative operazioni di pulizia e analisi i dati vengono integrati con dati già esistenti nel DWH. L'informazione viene finalemnte visualizzata e in base alle informazioni passate e alle informazioni nuove nasce un nuovo tipo di analisi.

2

Contesto applicativo

2.1 Il cliente

Prima di andare a parlare del caso di studio è necessario introdurre il contesto applicativo esplictando il luogo di svolgimento del progetto, il cliente interessato e lo stack tencologico grazie al quale è stato possibile creare un output nei tempi di lavoro stabiliti.

Il lavoro svolto è stato realizzato durante un periodo di sei mesi di tirocinio per Tesi presso *Iconsulting S.p.*A, un azienda fortemente interessata a imporsi sul mercato proponendo sistemi a supporto delle decisioni, specializzandosi in progetti sul campo della Business Intelligence. Essendo un'azienda di consulenza, Iconsulting collabora con diversi clienti al fine di raggiungere obiettivi comuni sulla modalità domanda-offerta. Come già detto in precedenza più clienti vengono assegnati a un team condividendo assieme idee, opinioni e soprattutto progetti.

Il mio team (gli "*All-In*"), attualmente composto da una totalità di 15 persone distribuite tra manager, team-leader e consulenti, lavora in maniera paralella su più progetti al fine di ottimizzare le performaces dei tempi di richiesta. Il progetto realizzato è stato presentato tramite *P.o.C. (Proof of Concept)* per il cliente *"Kering Eyewear"*.Nel 2014 *Kering,* una **holding** multinazionale francese responsabile delle vendite di prodotti comprendente un gruppo mondiale di marchi distribuiti in 120 paesi, lancia una start-up per sviluppare *occhiali* fatti in casa divulgandosi in campo *"Fashion Retail"* [KER14].Al fine di massimizzare lo sviluppo dei propri marchi, Kering ha deciso di interiorizzare la catena del valore per le sue attività sugli occhiali, su creazione e sviluppo sulla gestione delle vendite del marketing. Attraverso questo progetto, Kering sta mettendo a punto un modo innovativo di gestire le sue operazioni *Eyewear*, che porterà a significative opportunità di creazione di valore e consentirà a Kering di cogliere a pieno il potenziale di crescita delle sue case in questa categoria, in un mercato globale che è considerevole e in cui il segmento di mercato sta godendo di una crescita sostanziale.Oggi Kering Eyewear sviluppa e distribuisce occhiali per *Gucci, Bottega Veneta, Saint Laurent, Alexander McQueen, Boucheron, Tomas Maier, Puma e Christopher Kane.*

KERING
EYEWEAR

Figura 2.3: Logo Kering Eyewear

Grazie all'attuale progetto, il cliente è in grado di analizzare a fondo come si sta evolevendo il trend di mercato incrociando dati di "*Sell-out*"(parliamo quindi di vendite) con dati riguardanti il territorio (popolazione, salario medio, tweets, flussi turistici e lavoro) astraendo KPI complessi che permettono di valutare la *potenzialità* di un area. Il lavoro è realizzato tramite dashboard interattive che includono mappe, tabelle e grafici. Il caso di studio verrà approfondito nel capitolo 3.

2.2 Stack tecnologico

Per facilitare l'interazione con il cliente Iconsulting ha realizzato per il cliente Kering Eyewear un Data Warehouse e un sistema di Business Intelligence dedicato nel quale ad oggi confluiscono diversi dati (ordini e fatture di sell-in, vendite di sell-out, campagne di raccolta ordini ecc..). Il tutto è stato realizzato con la tecnologia full-SAP (SAP HANA, SAP Data Services, SAP Business Object) in cloud. Per la realizzazione del P.o.C. è stato utilizzato tutto ciò che è offerto da questa tecnologia con l'aggiunta di Tableau come front-end.

Per sua natura, la BI necessita di strumenti che permettano di interrogare e manipolare le informazioni. Tra i piu` importanti troviamo i sistemi per la gestione di basi di dati (DBMS), ovvero strumenti tecnologici in grado di gestire efficientemente grandi collezioni di dati, persistenti e condivisi, che offrono dei meccanismi per la garanzia dell'affidabilita` dei dati, per il controllo degli accessi e per la concorrenza. In questa sezione, verranno introdotti le tecnologie utilizzate all'interno di questo progetto di tesi, esplicitando le motivazioni che ne hanno portato all'adozione. In particolare approfondiremo:

- *SAP HANA*: caratteristiche, architettura, gestione della memoria.
- *SAP HANA Spatial:* caratteristiche, tipo di dato e sistemi di riferimento.
- *SAP Data Services:* caratteristiche, prorietà e utilizzo.
- *SAP Business Oject:* caratteristiche, architettura, tool.
- *Tableau:* caratteristiche, architettura,utilizzo.

2.2.1 SAP HANA

Il mercato attuale è fortemente influenzato dalle esigenze della società. Il fatto che essa si evolva in maniera costante e con ritmi veloci, obbliga le aziende a rivedere e ripensare costamntemente il proprio business. In questo scenario gioca un ruolo fondamentale la scelta tecnologica.
La strategia di SAP è proprio quella di riuscire ad abbattere le latenze e i colli di bottiglia legati all'accesso del patrimonio informativo aziendale, creando una soluzione che sia in grado di analizzare in maniera tempestiva il proprio business per prendere decisioni consapevoli sul futuro prossimo. [SAPAG16]. La soluzione sviluppata da SAP è la prima appliance analitica basata *sull'in-memory computing* e da qui il nome di *High Performance Analytic Appliance*, sul mercato conosciuta con il suo acronimo HANA:

- *Appliance*: la piattaforma costituita da un insieme di hardware e software progettati per riuscire a ad eseguire particolari e complesse funzioni applicative. Le Appliance permettono di ottenere ottime diverse prestazioni in quanto sono supportate da ottimizzazioni hardware e software.

- *Analytic*: è una piattaforma orientata all'analisi di enormi quantità di dati che descrivono il business supportando in maniera eccellente tutte le operazioni OLAP, mettendo a disposizione anche tecniche di data mining per fare analisi predittiva come algoritmi di clustering, algoritmi di classificazione, algoritmi di regressione, reti neurali, funzionalità statistiche e algoritmi per l'analisi per i contenuti dei social network [SAPHAPA16]

- *High Performance*: raggiunge ottime prestazioni grazie alla politica dell'in memory computing, abbattendo il collo di bottiglia legato all'elevata latenza su operazioni di I/O dei dischi sui quali risiedono informazioni aziendali, infatti la totalità dei dati risiede direttamente sulla memoria centrale permettendo di avere tempi bassissimi di elaborazione dei dati. Le alte prestazioni sono dovute anche alla

parallellizzazione dei processi possibile grazie alla memorizzazione dei dati in colonna e all'architettura multi-core.

Figura 2.4: Tecnologia SAP

Il sistema SAP HANA è stato completamente sviluppato in C++ ed è progettato per eseguire un sistema operativo Linux Enterprise Server. La figura sottostante mostra I componenti principali che costituiscono l'architettura di SAP HANA[SAPHATU12]:

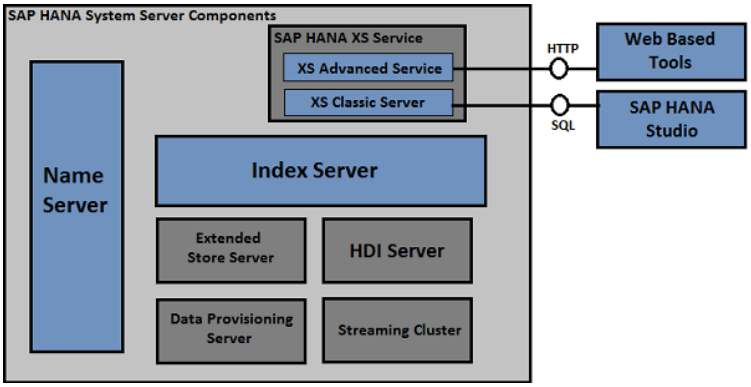


Figura 2.5: Architettura SAP

Andiamo ad analizzare I singoli elementi e le loro funzionalità:

- *Index Server:* è il componente principale, anche detto *SAP HANA Database*. Questo componente permette di archiviare i dati ed elaborarli. Il server richiede di gestire richieste tramite query SQL e memorizza le sessioni precedenti.

- *Name Server:* è il componente che contiene informazioni riguatrdanti la tecnologia del sistema SAP HANA. In un sistema distribuito, con più istanze di HANA su server differenti, il Name Server tiene traccia per ogni server quali siano le componenti in esecuzione e su quale porzione di dati sta avvenendo l'elaborazione.

- *XS Engine:* XS è l'acronimo di eXtended Service ed è un estensione del database che consente alle applicazioni esterne di comunicare con la piattaforma mediante richieste HTTP. *XS Classic Server* è un vero e proprio server HTTP utilizzato per l'esecuzione di operazioni web che svincola gli sviluppatori dall'appoggiarsi ad un server esterno.

- *Extended Store Server:* è il componente predisposto per fornire supporto e garantire ottime prestazioni anche nel momento in cui si lavora con grandi quantità di dati.

- *Data Provisioning Server:* è il componente predisposto alla preparazione dei dati da fornire all'utente o a qualche altra risorsa in remote, in real-time o in modalità batch.

- *Streaming Cluster:* conosciuto anche come SAP Event Stream Processor è il componente predisposto alla computazioni di dati stream e alla gestione di eventi complessi.

In base a questa caratterizzazione va data importanza alla gestione della tabelle in SAP HANA.
Tipicamente, una tabella all'interno di un database è una struttura bidimensionale organizzata in righe e colonne, la memoria del computer va in constrasto con questa gestione dato che organizza i dati come una struttura lineare. Per salvare una tabella in una memoria lineare sono possibili due soluzioni:

- *Memorizzazione orientata per righe (row oriented):* le tabelle vengono gestite come una sequenza di record, ognuno dei quali contiene tutti gli attributi che descrivono la tabella. Questa tecnica è adottata dalla maggior parte dei database
- *Memorizzazione orientata per colonne (column oriented):* le tabelle venono gestite come sequenze di colonne, ognuna delle quali immagazzina valori di un attributo per tutti quanti i record.

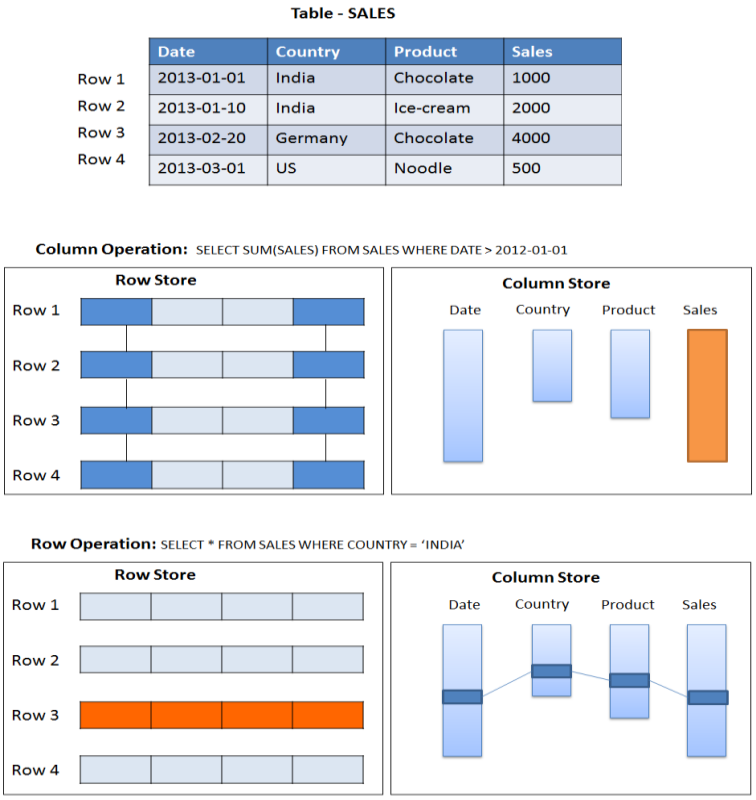


Figura 2.6: Tipi di memorizzazione

La figura mostra, attraverso un esempio sulla tabella delle vendite, le due tecniche di memorizzazione. I vantaggi apportati dalla tecnica colonnare sono notevoli, soprattutto in termini di performance, questa tecnica garantisce:

- Accesso ad dati veloce in quanto è sufficiente leggere lsolo la colonna interessata alla query.

- Qualsiasi colonna può essere utilizzata come indice. Salvare i dati in colonne è, dal

punto di vista funzionale, molto simile a costruire un indice per ogni colonna.

- Maggior compressione dei dati perché la maggior parte dei dati contenuti all'interno di una colonna è composta da pochi valori distinti.

- Procedure parallelizzabili, se più di una colonna è coinvolta in operazioni di ricerca, aggregazione o altro, ognuna di queste operazioni può essere eseguita da un differente processore.

SAP HANA Database supporta entrambe le tecniche, ma è particolarmente ottimizzato per la memorizzazione colonnare dei dati, riuscendo a gestire in maniera efficiente la memoria.
Un'applicazione analitica si compone principalmente di operazioni di lettura e di aggregazione, le quali richiedono velocità nel recuperare il dato e velocità di elaborazione. La memorizzazione in colonna ottiene prestazioni notevoli per le operazioni di questo tipo ma non non per operazioni di scrittura. Per riuscire ad essere performante su qualsiasi tipo di operazione, la gestione dei dati in memoria centrale viene gestita attraverso due strutture di memoria differenti, ottimizzate per singole funzioni, che cooperano assieme: *Main Storage (MS)* e *Delta Storage (DS)*. Come si può vedere dalla Foto le operazioni di lettura vengono eseguite da entrambe le strutture ma sono ottimizzate solo per il MS, mentre quelle di scrittura sono eseguite e ottimizzate solo per il DS. Il MS è quella parte di memoria centrale ottimizzata per operazioni come la compressione dei dati che viene effettuata prima di salvare il dato in memoria, la lettura, la ricerca e le operazioni di aggregazione sui dati salvati in maniera colonnare. Per poter eseguire queste operazioni il MS deve poter avere al suo interno la totalità dei dati, i quali sono salvati in maniera compressa. L'operazione di scrittura sui dati compressi e memorizzati all'interno del MS è un'operazione molto onerosa e dispendiosa in termini di tempo e risorse. Proprio per questo motivo, tali operazioni non vanno a modificare direttamente i dati compressi, ma tutti i cambiamenti sui dati vanno a scrivere su una struttura separata chiamata Delta Storage.

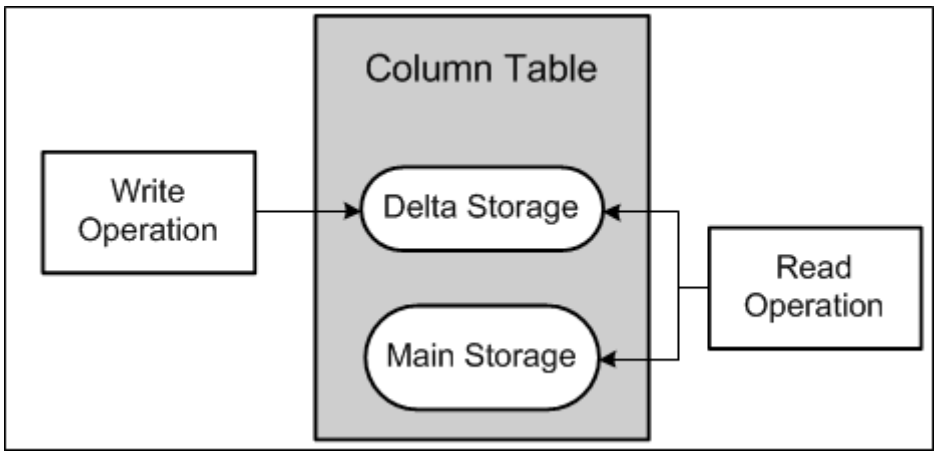


Figura 2.7: Operazioni di read e write

Questa struttura ausiliaria è incaricata e ottimizzata per la scrittura dei nuovi dati in memoria centrale e successivamente in memoria di massa, al fine di allineare i dati per operazioni di analisi con quelli di backup. Quindi il Delta Storage rappresenta la chiave di ottimizzazione dei tempi per la gestione della memoria. La procedura viene definita *Delta Merge* ed è rappresentata in figura:

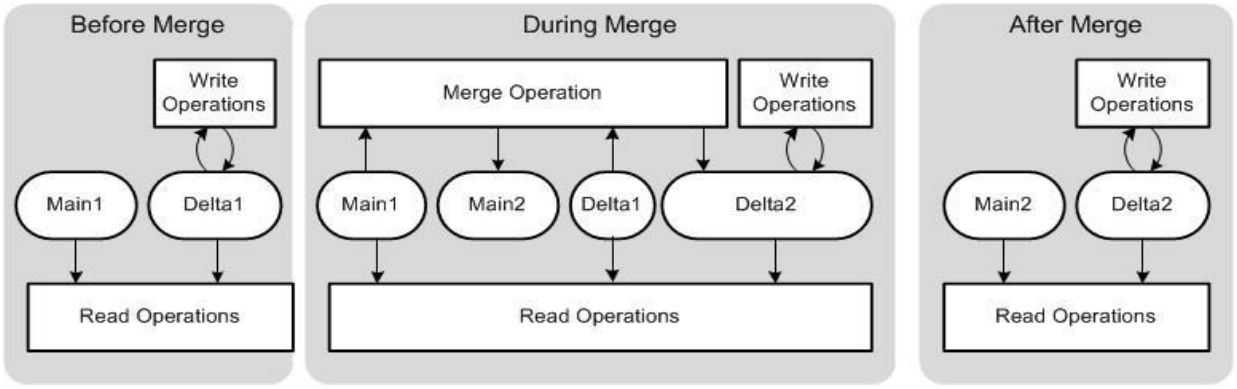


Figura 2.8: Before-During-After Merge

La procedura è rappresentata in 3 step:

- *Before Merge:* tutte le operazioni di scrittura sono gestate dal Delta Storage mentre le operazioni di lettura sono gestate sia dal Delta Storage che dal Main Storage.
- *During Merge:* si accavallano le seuguenti attività:
    1. Tutte le operazioni di scrittura sono gestitte dal Delta Storage 2
    2. La lettura viene fatta da entrambi i Main Storage e dal Delta Storage 1
    3. I nuovi dati presenti solo nel DS1 (dati "uncommitted) vengono trascritti da

DS1 a DS2.

4. I dati presenti all'interno di MS1 e i dati "committed" all'interno del DS1, vengono uniti e scritti all'interno del MS2.

- *After Merge*: si verificano le seguenti attività:
    1. MS1 E DS1 vengono eliminate.
    2. Viene ottimizzata e rivalutata la memorizzazione dei dati all'interno dello spazio di memoria MS2. Queste operazioni consistono nel riordinamento delle righe e la compressione dei dati. Nel caso in cui le operazioni generano un cambiamento dei dati, tali dati verranno ricaricati nuovamente all'interno del MS2.
    3. Il contenuto del MS2, viene scritto in maniera permanente su disco.

2.2.2 SAP HANA Spatial

Le caratteristiche chiave scelte da SAP nello sviluppo di HANA: IMDB (In Memory Database) e strutture dati colonnari, hanno dato luce ad uno strumento prestante ed efficiente per le applicazioni aziendali di oggi che sempre più sono volte all'analisi del business. Tali analisi però potevano essere arricchite e assumere maggiore significatività con l'aggiunta della componente spaziale, per questo motivo nasce un nuovo componente *SAP HANA Spatial.* Venne integrato nel sistema dalla versione SP6 permettendo di organizzare e gestire le informazioni geografiche, potendole mettere in relazione con tutti gli altri tipi di dato presenti sul database e fornendo la possibilità di interrogarli attraverso specifiche funzioni geospaziali come il calcolo della distanza, l'unione o l'intersezione di più oggetti e altro ancora ,al fine di riuscire a mantenere alte prestazioni anche per gestire ed elaborare e l'analisi dei dati spaziali.

SAP HANA include un "motore" multilayer dedicato interamente ai dati spaziali il quale si basa sullo standard SQL Multimedia (SQL/MM), supporta la struttura colonnare e incorpora metodi di accesso ottimizzati per tali dati. SAP HANA Spatial modella i dati spaziali attraverso forme geometriche in 2 e 3 dimensioni, ogni tipologia di forma può essere compresa in una collezione di figure omogenee (solo oggetti dello stesso tipo) o eterogenea. Nella Figura sootostante viene mostrata la gerarchia del tipo di dato spaziale *ST_Geometry* e tutti i sottotipi che derivano da esso.

Questo tipo di dato, all'interno di HANA, è gestito in un ottica orientata agli oggetti, infatti rispetta le seguenti proprietà: un sottotipo è più specifico di un supertipo, un sottotipo eredita tutti i metodi del suo supertipo, il valore di un sottotipo può essere convertito automaticamente al valore di un supertipo, una colonna definita come un sottotipo può assumere tutti i valori di ogni suo supertipo proprio come accade nel concetto di ereditatrietà tra classi Java [SAPHASR16].

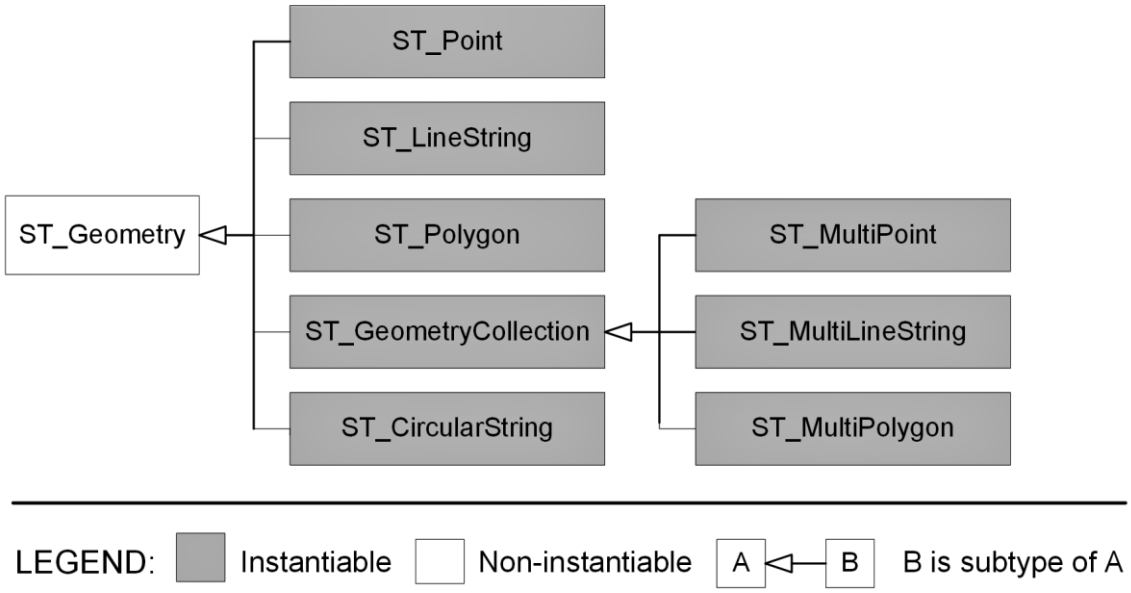


Figura 2.9: Tipo ST_Geometry

I tipi di dati derivabili da ST_Geometry e utilizzabili da HANA sono:

- *Geometry:* è l'emeneto alla base della gerarchia ed è il "padre" di tutti i tipi di dato spaziale.

- *Point:* è la forma geotrica più semplice, rappresenta un insolo punto nello spazio. Un punto è rappresentato tramite una coppia di coordinate (X,Y). Tipicamente rappresentati tramite *latitudine* e *longitudine*.

- *MultiPoint:* è una collezione di uno o più punti. Tipicamente vengono utilizzati per rappresentare un insieme di luoghi clusterizzati da caratteristiche comuni.

- *LineString:* connette una serie di punti (o vertici), consiste in una serie di coordinate,

ordinate, connesse fra loro da linee rette. Una linea viene definita semplice quando le parti che la compongono non si sovrappongono incrociandosi mentre viene definita chiusa nel momento in cui il punto di inizio e di fine coincidono. Una linea possiede lunghezza ma non possiede l'area. Tipicamente vengono utilizzate per rappresentare
strade, fiumi o percorsi prestabiliti.

- *MultiLineString:* è una collezione di una o più linee, tipicamente utilizzati per descrivere grafi stradali o percorsi d'acqua.

- *Polygon:* definisce una regione geometrica racchiusa da una linea chiusa. Un poligono possiede un area ma non una lunghezza. La rappresentazione di un poligono è composta da una struttura esterna che definisce il confine della regione e da zero o più figure interne che definiscono delle aree escluse dalla regione in considerazione. Tipicamente è utilizzato per descrivere un territorio politico come una città, una regione, uno stato o un continente.

- *MultiPolygon:* collezione di uno o più poligoni.

- *CircularString:* rappresenta una sequenza di punti connessi tra loro da segmenti circolari detti anche cerchi.

- *GeometryCollection:* rappresenta un insieme di un numero di qualsiasi di forme delle tipologie precedenti.

Per interagire con questo tipo di dato è necessario specificare lo spazio geometrico (o sistema di riferimento) *Spatial Reference System(SRS)*. Esso si compone di :

- Un identificatore univoco chiamato *Spatial Reference Identifier (SRID)*
- Unità di misura del sottostante sistema di riferimento, HANA riconosce metri, radianti e misure derivate.
- Coordinate massime e minime accettate, tipicamente per definire i limiti spaziali

- Se i dati sono planari o sferoidali

- Informazioni aggiuntive in cui viene specificata la formula di proiezione per la conversione da un SRS a un altro.

Le informazioni riguardanti i sistemi di riferimento sono memorizzate all'interno delle informazioni di sistema all'interno della tabella ST_SPATIAL_REFERENCE_SYSTEM, la quale inizialmente viene popolata con gli SRID di default ma è possibile inserirne di nuovi attraverso il tool *Geospatial Metadata Installer* che permette di inserire nuovi SRID.
Gli SRID di default presenti nella tabella ST_SPATIAL_REFERENCE_SYSTEM sono:

SRS	**SRID**	**Descrizione**
Default	*0*	Rappresenta il sisstema di riferimento utilizzato di default nle momento in cui non viene specificato nessun SRID. Ogni punto è descritto da una coppia di coordinate (X,Y) ,il cui valore è compreso tra [-.1000.000, 1.000.000]
WGS84	*4326*	E' il sistema di riferimento standard per descrivere la superficie terrestre sferoidale, il sistema WGS84 rappresenta lo standard per tutti i sistemi GPS.L'origine delle coordinate è il centro della Terra. Le coordinate sono espresse in gradi, la prima rappresenta la longitudine e può assumere un valore compreso tra [-180,180] mentre la seconda rappresenta la latitudine e può assumere un valore tra [-90,90]. L'unità di misura è espressa in metri.
WGS84 (planar)	*1000004326*	E un sistema di riferimento simile al WGS84 ma differisce per il semplice fatto che usa proiezioni basate su rettangoli che distorcono il calcolo delle distanze , dell'area e altre misure. Tipicamente sono utilizzati per valutare la relazione tra due figure su predicati come: una figura contiene l'altra, le due figure si toccano ecc.

sa_planar_unbounded (unbounded planar)	*2147483646*	Accetta qualsiasi valore di coordinata venga utilizzato perché descrive uno spazio senza limiti. Viene utilizzato solo per usi interni.

2.2.3 SAP Data Services

SAP HANA rappresenta il principale sistema di supporto di memorizzazione di grandi flussi di dati, permettendo di presentare attraverso strategie di elaborazione dei dati delle tabelle finali dalle quali verranno estratte informazioni fondamentali per fare analisi per guidare nuove strategie di business. Risulta quindi determinante la fase di elaborazione del dato. Spesso quando si ha a che fare con flussi di dati e script SQL decisamente complessi SAP HANA ha capacità limitate. Viene offerto quindi un sistema di supporto per l'integrazione e trasformazione dei dati a se stante: *SAP Data Services.*

SAP Data Services è un efficiente strumento *ETL* (Extraction, Transformation and Loading) certificato da SAP per eseguire il caricamento in "batch" su SAP HANA.

Consente agli utenti di sviluppare ed eseguire flussi di lavoro che acquisiscono dati da più fonti permettendo all'utente di combinare, trasformaree perfezionare i dati.

L'integrazione e le trasformazioni dei dati possono essere eseguite utilizzando linguaggi di programmazione di database come SQL e PLSQL, tuttavia sarà costoso gestire/ mantenere il panorama. E' qui che gli strumenti ETL giocano un ruolo fondamentale. Un nuon strumento ETL infatti è progettato per avere una piattaforma unica in cui gli sviluppatori possono costruire la logica per le trasformazioni e gli amministratori possono anche manterenere facilmente il sistema [SAPDASE16].

Figura 2.10: SAP Data Services

Il caricamento da più fonti su SAP HANA riscontra differenti problemi infatti i dati potrebbero essere[SAPBO16]:

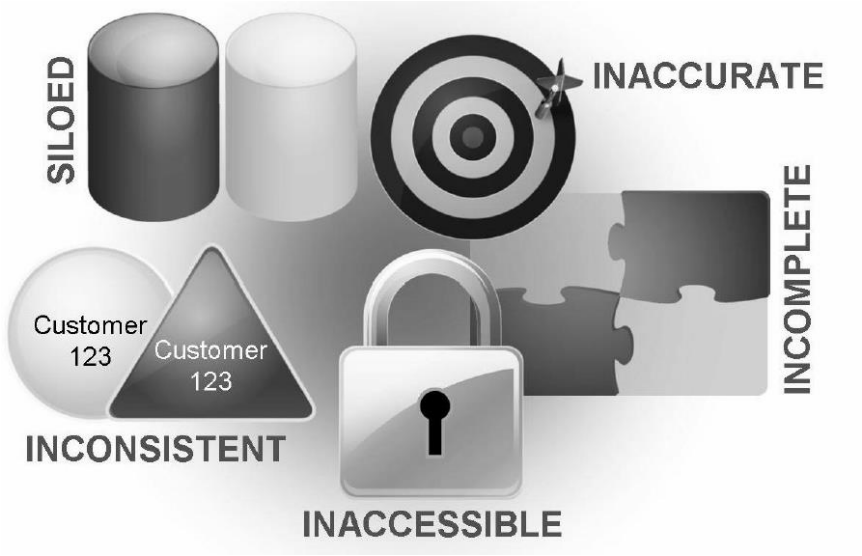


Figura 2.11: Caratteristiche SAP HANA

- *sporchi*: Con dati sparsi in diversi sistemi ERP è più probabile trovare diverse versioni della verità limitando la capacità di ottenere una completa visione del business.

- *Inaccurati:* in molti sistemi del cliente, i dati sono intrinsecamente incoerenti poiché le cose cambiano e i requisiti aziendali continuano ad evolversi per raggiiungere nuovi obiettivi. Problemi comuni vengono riscontrati con nomi di clienti, indirizzi e nomi di prodotto errati.

- *Inconsistenti:* Le definizioni di entità aziendali comuni come clienti, prodotti, fornitori, nomi materiali e codici hanno "*naming convention*" diverse da sistema a sistema, creando

incoerenze tra i dati stessi. Nasce la necessità di avere un modo efficiente di riconciliazione dei dati in maniera veloce e accurata.

- *Incompleti:* Spesso alcuni dati risultano incompleti come codice postale, codice paese ecc.. A volte questo tipo di dato è inutilizzato e quindi risulta inunfluente in fase di analisi mentre se viene utilizzato potrebbe creare problemi.

- *Inaccessibili:* A volte i dati sono in un formato non strutturato come in un testo in formato libero. La sfida sta nel come sbloccare approfondimenti e il potenziale da tutte le fonti di dati.

SAP Data Services è la prima e unica soluzione all-in-one per l'integrazione dei dati (ETL), la gestione della qualità dei dati, gestione delle informazioni e dei meta-dati che permette di avere un servizio efficiente e allo stesso tempo veloce e sicuro.

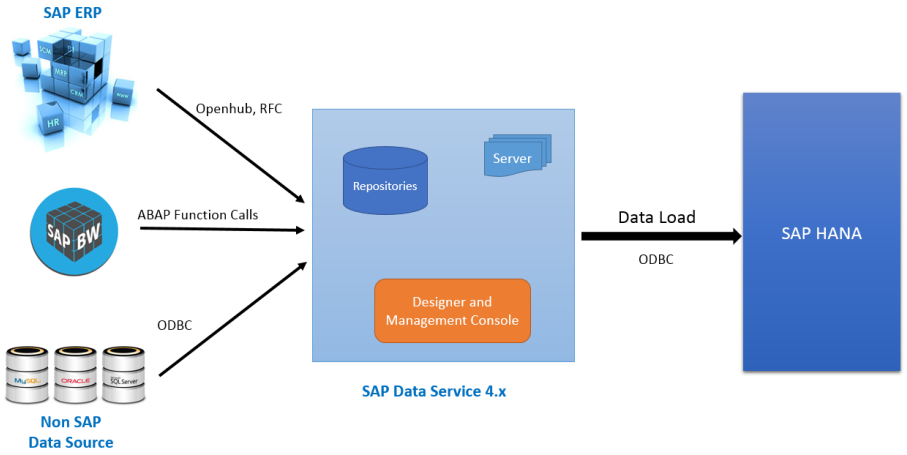


Figura 2.12: Processo di caricamento dei dati su SAP HANA

Inoltre, come possiamo vedere in figura, esso permette di caricare le informazioni su SAP HANA tramite ODBC, permettendo di risparmare notevole tempo in scrittura di script SQL complessi con una grande quantità di dati [SAPDASE]. Data Services è costituita dai seguenti componenti:

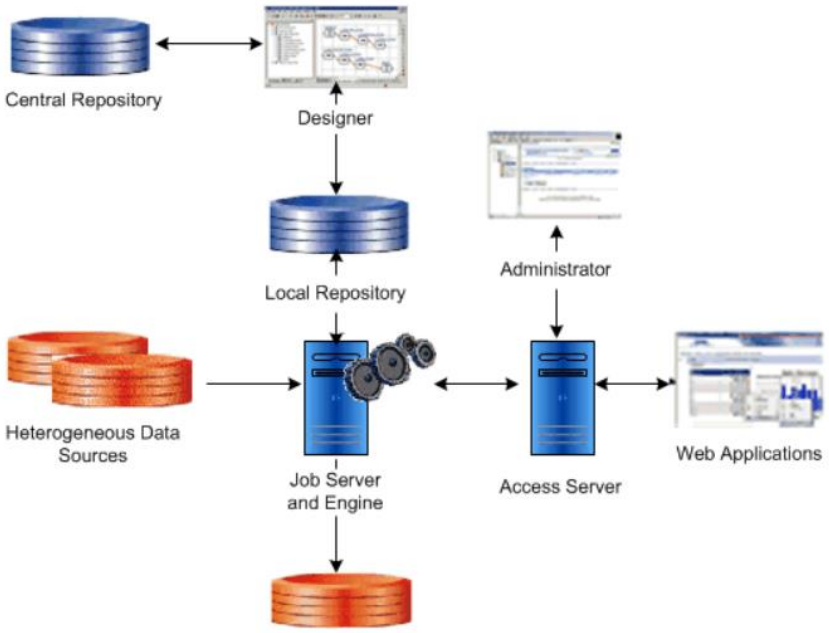


Figura 2.13: Componenti di SAP Data Services

- *Designer:* rappresenta lo struymento GUI front-end in cui gli sviluppatori possono accedere e creare Job in SAP Data Services.

- *Repository:* rappresenta il database che memorizza il Job fatto da *designer*. Gli oggetti possono venir salvati in due repository:
    - Repository Locale
    - Repositoty Centrale

- *Access Server:* Rappresenta il server che esegue real-time i Job creati dagli sviluppatori nei repository.

- *Job Server:* questo è uno dei principali componenti server nei servizi dati e viene utilizzato per eseguire i batch creati dagli sviluppatori nel sistema. I repository devono essere collegati ad almeno un Job server per eseguire i Job all'interno del repository.

- *Managment Console:* Rappresenta una console web per la gestione dllo scheduling dei Job su Data Services, mettendo in luce sistemi statistici nella memoria, esecuzione run-time

dei Job e utilizzo della CPU.

Abbiamo parlato di Job senza andare a specificare cosa fosse, quale sia la sua utilità e da cosa è composto.
Un Job in SAP Data Services rappresenta l'unico "oggetto" da eseguire nel designer. I Job possono essere eseguiti in modalità:

- *sviluppo:* si usa in genere per fare dei test sul Job stesso
- *produzione:* è possibile pianificare "Job batch" in tempo reale.

Ogni modalità legge e scrive su un DWH apposito.
All'interno dei Job è possibile specificare dei diagrammi di lavoro che definiscono una sequenza delle operazioni da effettuare.

All'interno di Job troviamo:

- *Work Flow:* rappresentano un "flusso di dati" che contengono anch'essi una sequenza di operazioni da eseguire. Al suo interno possiamo trovare una sequenza di uno o più Data Flow oppure sorgenti e traformazioni.

- *Data Flow:* rappresenta il contenitore di tutte le operazioni effettuate sui dati quindi (tabelle, scripts, condizioni, trasformazioni, Try/Catch)

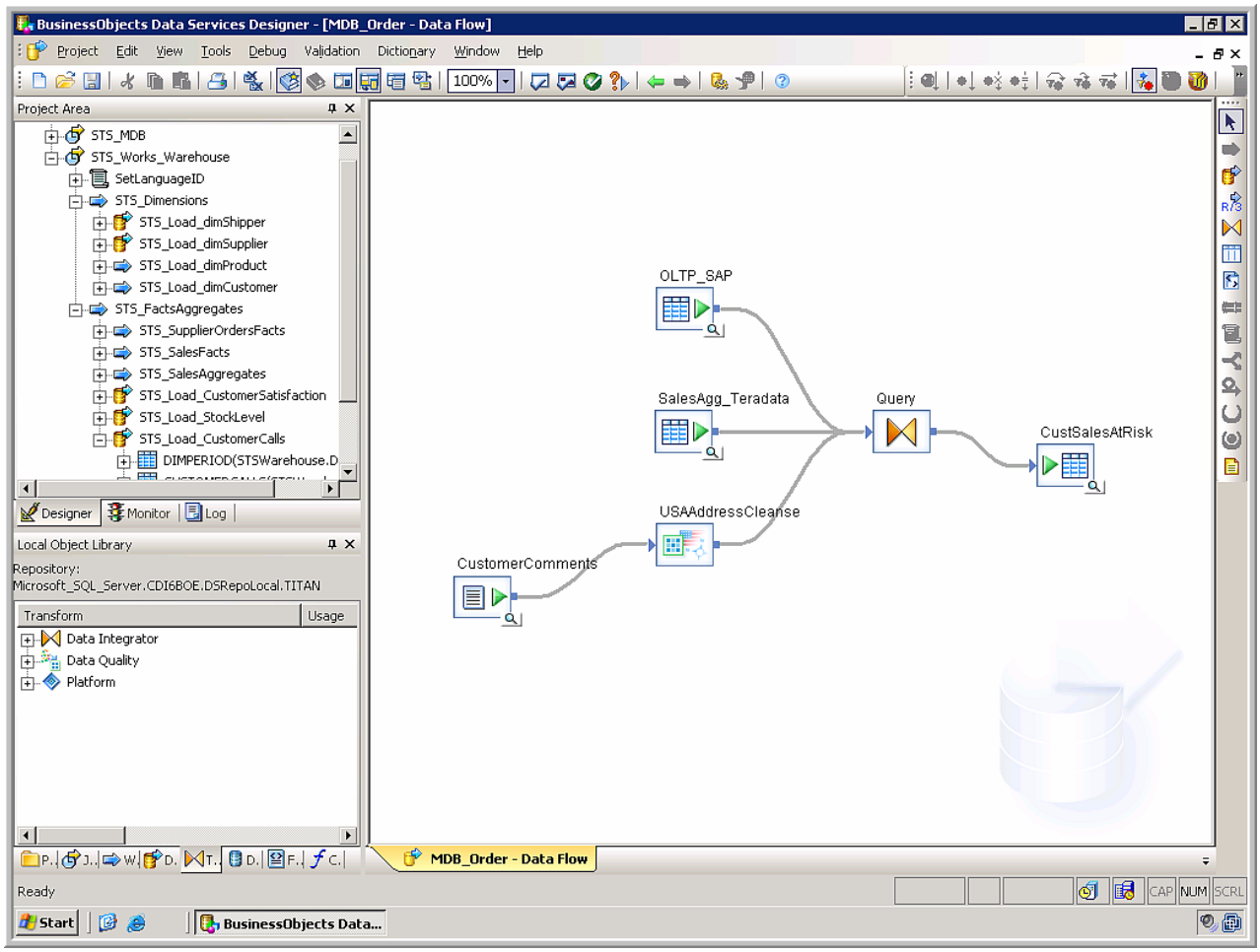


Figura 2.14: Ambiente di sviluppo SAP Data Services

2.3.4. SAP Business Object

Una volta che i dati sono stati puliti tramite il servizio ETL SAP DataServices e quindi integrati e caricati all'interno del DWH SAP HANA, inzia la fase di studio e analisi dei dati garantita attraverso *SAP Business Object*.

SAP Business Objects è la soluzione di business intelligence scelta dalle imprese che vogliono migliorare i processi aziendali, individuare nuove opportunità e ottenere un vantaggio sulla concorrenza. Offre soluzioni in grado di soddisfare tutte le esigenze di BI, da strumenti flessibili di reporting, query e analisi ad hoc a cruscotti e visualizzazioni avanzate. SAP Business Objects è un portafoglio di strumenti e applicazioni perfettamente integrate con i software gestionali SAP, ampiamente scalabili per ogni esigenza di business che si presenti. Non solo semplice reportistica,

ma analisi strategica dei dati, qualità delle informazioni, pianificazione, budget e consolidato. La BI è la naturale evoluzione della piattaforma SAP dove BMS, attraverso la sua business unit dedicata, è in grado di dare supporto a tutti i livelli.

SAP Business Object presenta una piattaforma per la BI *flessibile, scalabile e potente*.

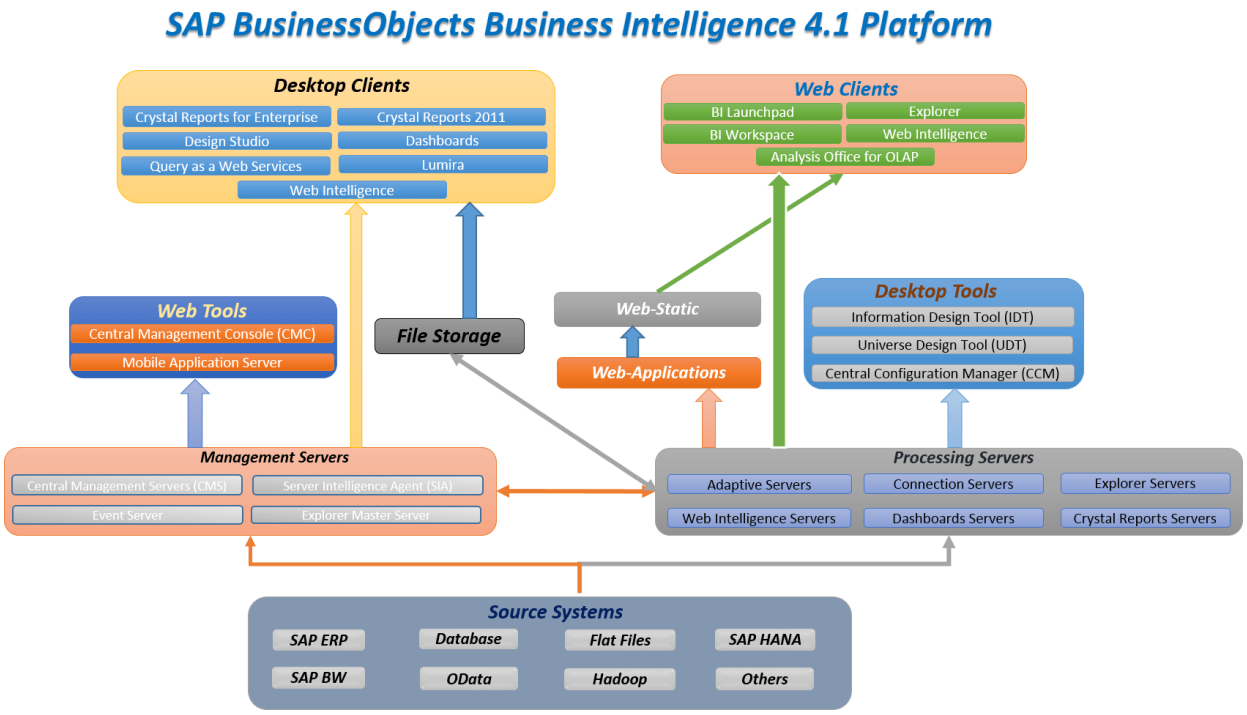


Figura 2.15: Piattaforma SAP Business Object 4.1

Dalla figura notiamo i principali componenti della piattaforma SAP BOBI:

- *Central Managment Console*: è l'interfaccia web principale per eseguire attività amministrative nella piattaforma SAP BOBI, compresa la sicurezza dell'utente, il contenuto e la gestione del server. Ci consente inoltre di pubblicare e organizzare i contenuti e configurare le impostazioni.

- *Server Intelligence Agent:* gestisce e controlla tutti i server su un nodo assicurandone il corretto funzionamento. Un nodo è un gruppo di server SAP BOBI Platform che vengono eseguiti con lo stesso account utente. Una macchina può contenere più nodi quindi è

possibile eseguire processi paralleli.

- *Central Configuration Managment:* rappresenta uno strumento dei problemi del server e di gestione dei nodi fornito in due forme. In un ambiente Microsoft Windows, CCM consente di gestire server locali e remoti tramite l'interfaccia grafica utente (GUI) o da una riga di comando. In un ambiente UNIX, lo script della shell CCM (ccm.sh) consente di gestire i server dalla riga di comando.

- *Semantic Layer Tool:* Alcuni degli strumenti di reportistica che abbiamo in SAP BOBI possono essere collegati direttamente ai sistemi di origine per creare report, mentre alcuni di essi necessitano di un livello semantico / aziendale intermedio. Il livello semantico / aziendale, noto anche come *Universo* in ambiente SAP BOBI, può essere creato utilizzando Information Design Tool (IDT) o Universe Design Tool (UDT).
  Information Design Tool (IDT) è un ambiente di progettazione di metadati SAP BusinessObjects che consente a un progettista di estrarre, definire e manipolare i metadati dalle origini relazionali e OLAP per creare e distribuire universi UNX. Gli *Universi* sono uno strato semantico che nasconde la complessità del database dagli utenti finali. Estrae la complessità dei dati utilizzando il business piuttosto che il linguaggio tecnico per accedere, manipolare e organizzare I dati. Lo strumento viene utilizzato per effettuare report.

- *Client Tools:* SAP BusinessObjects Mobile consente agli utenti di accedere in remoto agli stessi report di business intelligence (BI), metriche e dati in tempo reale disponibili per i client desktop e Web, il tutto da un dispositivo mobile.
  SAP Business Object Mobile aggiunge dei tool client di supporto:

  -*Lumira*: SAP Lumira è una soluzione self-service che consente agli Analisti e ai responsabili delle decisioni di accedere, analizzare trasformare e visualizzare i dati in maniera efficacie,veloce ed efficiente per via grafica. Esistono due versioni disponibili per SAP Lumira:

1. *Lumira Desktop:* SAP Lumira Desktop viene utilizzata per preparare i dati da più fonti, visualizzarli e quindi comporre storie da quelle visualizzazioni che possono essere condivise con altri responsabili decisionali.

Figura 2.16: SAP Lumira Desktop

Interangendo con SAP Lumira Server è possibile fornire l'utilizzo del browser e dispositivi mobili per analizzare ulteriormente i dati.

2. *Lumira Sever:* SAP Lumira, piattaforma server per Business Intelligence (BI), porta il contenuto di SAP Lumira nella distribuzione della piattaforma SAP BusinessObjects BI, consentendo così di utilizzare la sicurezza, la scalabilità e la facilità d'uso fornite dalla piattaforma BI. Il visualizzatore di SAP Lumira è integrato nella piattaforma SAP BusinessObjects BI come applicazione Web, consentendo agli utenti di avere la stessa esperienza di visualizzazione ed esplorazione di storie in BI Launch Pad come avviene nel desktop SAP Lumira. La piattaforma SAP BusinessObjects BI rafforza la sicurezza dei documenti Lumira e consente l'accesso e la categorizzazione allo stesso modo degli altri contenuti della piattaforma BI, consentendo di adottare perfettamente SAP Lumira all'interno.

-

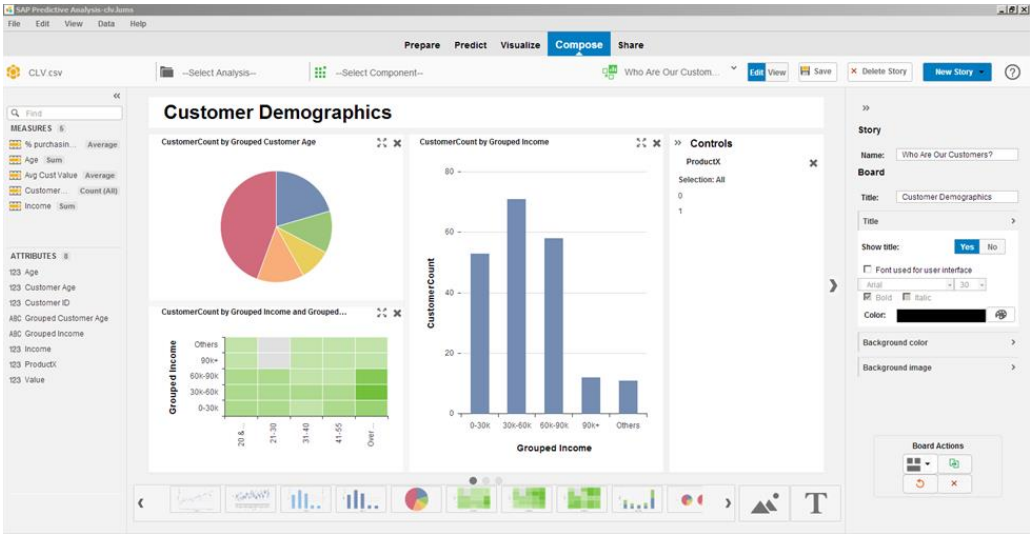


Figura 2.17: Dashboard Lumira

-*Design Studio:* SAP BusinessObjects Design Studio consente ai progettisti di creare applicazioni di analisi e dashboard, basati su SAP NetWeaver BW, SAP HANA per browser e dispositivi mobili (iPad, ad esempio). È il prodotto di scelta quando è richiesto il pieno supporto per i modelli di dati SAP NetWeaver BW e SAP HANA e le funzionalità del motore. Il prodotto offre uno strumento di progettazione che consente di creare applicazioni in modo semplice e intuitivo senza la necessità di competenze native di programmazione dell'interfaccia utente HTML. la scelta di creare dashboard completamente compatibili con HTML è data dal fatto che possono essere eseguite su qualsiasi dispositivo, da un semplice pc fisso per arrivare fino ai dispositivi portatili e mobili come smartphone e tablet. Questa scelta non è casuale, ma bensì in linea con gli elementi cardine della nuova strategia di SAP, volta a supportare le aziende nel loro processi di analisi fornendo un supporto che sia quanto più in real-time e usufruibile anche al di fuori delle mura aziendali, attraverso l'utilizzo di dispositivi mobili ovunque essi si trovino. Attraverso Design Studio lo sviluppatore come gia accennato è svincolato dal conoscere la tecnologia HTML5. Infatti lo sviluppo di una dashboard, complessa o basilare, si ottiene dal semplice *drag and drop* dei componenti messi a disposizione da Design Studio e un breve passo di setup di quelli utilizzati all'interno dell'applicazione. Questo tipo di tool che mostrano in anteprima il risultato finale, in questo caso ottenuto dal drag and drop, vengono chiamati WYSIWYG che rappresenta l'acronimo di "What You See Is What You Get".

Design Studio offre una vasta gamma di componenti utilizzabili per creare le proprie dashboard, ci sono componenti analitici che mostrano i dati provenienti dai vari datasource in formato tabellare o attraverso grafici, componenti basilari come testo, immagini, pulsanti e una serie di filtri che permettono di ideare un percorso interattivo che permetta all'utente di interagire con la dashboard e consultarla a piacimento navigando anche altre porzioni di dati. Infine ci sono anche componenti strutturali che permettono di creare componenti che contengono altri componenti per poi poterli riutilizzare anche in altre dashboard. Questi ultimi permettono anche la creazione di dashboard con pagine multiple. Oltre a questi componenti standard creati da SAP, gli sviluppatori di terze parti possono creare i loro componenti attraverso l'SDK Design Studio Software Development Kit. Molteplici componenti di terze parti sono già disponibili attraverso la vasta community online che mette a disposizione in maniera open-source o a pagamento i priori componenti.

Figura 2.18: Dashboard con Design Studio

2.2.5 Tableau

L'ultima sezione dedicata alle tecnologie presenta un altro strumento di Data Visualization e Data Discovery: *Tableau*.

Tableau prende il nome dall'azienda americana *Tableau Software* specializzata nella generazione

di prodotti di visualizzazione di dati interattivi focalizzati sulla BI. La società è stata fondata presso il Dipartimento di Informatica dell'Università di Stanford tra il 1997 e il 2002.
Su Tableau sono disponibili due diverse versioni [RB05]:

Figura 2.19: Tableau

- *Tableau Desktop:* è un'applicazione di visualizzazione dei dati che consente di esaminare praticamente qualsiasi tipo di dati strutturati e generare *grafici, dashboard e report* altamente interattivi in pochi minuti. Dopo un'installazione rapida, è possibile legare virtualmente qualsiasi origine dati dai fogli di calcolo ai Data Warehouse e visualizzare le informazioni in diverse prospettive grafiche. E' stato progettato per rendere facile l'apprendimento e l'utilizzo.

- *Tableau Server:* rappresenta un 'applicazione di Business Intelligence che offre analisi basate su browser. E' una soluzione online pensata per condividere, distribuire e collaborare su contenuti creati in Tableau.Tableau fornisce di un'architettura client-

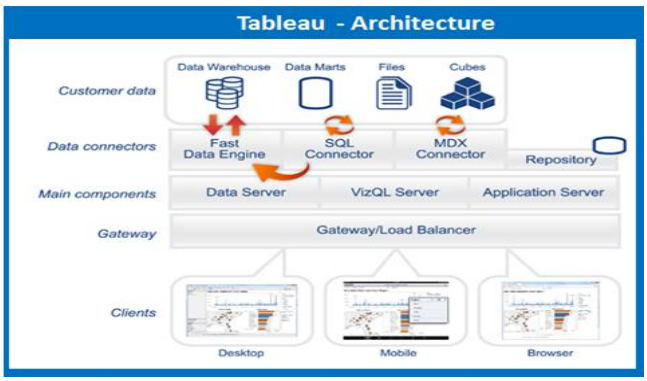


Figura 2.20: Tableau Architecture

server altamente scalabile che serve client mobili, client Web e software installato sul

Desktop. Tablueau Desktop è lo strumento di creazione e pubblicazione utilizzato per creare viste condivise su Tableau Server.

I layer vengono suddivisi in questo modo:

- *Data Layer:* Una delle caratteristiche fondamentali di Tableau è supportare la scelta dell'architettura dei dati. Non ha bisogno che i dati siano immagazzinati in un singolo sistema, proprietario o altro. Quasi tutte le organizzazioni hanno un ambiente di dati eterogeneo. Non è necessario recuperare tutti i dati in memoria finché non si sceglie di farlo. Se la piattaforma di dati esistenti è veloce e scalabile, allora ti permette di controllare direttamente il tuo investimento sfruttando la potenza del database per i problemi di risposta. In caso contrario, fornisce semplici opzioni per migliorare i dati in modo rapido e reattivo con il Data Engine in memoria.

- *Data Connectors:* Consiste in una quantità di connettori dati ottimizzati per i database. Esistono anche connettori ODBC comuni progettati per qualsiasi sistema senza un connettore nativo. Offre due modalità a supporto dell'interazione con i dati: connessione Live o In-memory. I client possono passare tra una connessione attiva e in memoria

- *Live connection:* rappresentano connettori dati di Tableau che controllano l'infrastruttura dati disponibile trasferendo direttamente istruzioni dinamiche SQL o MDX nel database di origine, ad eccezione dell'importazione di tutti i dati. Inoltre, questo significa che Tableau può utilizzare efficacemente quantità illimitate di dati, infatti Tableau è il client di analisi front-end per molti dei più grandi database del mondo. Ha ottimizzato ogni connettore per sfruttare le caratteristiche uniche di ogni fonte di dati.

- *In memory:* Presenta un Data Engine veloce e in memoria per l'ottimizzazione per l'analisi.E' possibile connettersi direttamente al DWH in maniera semplice. Il Data Engine di Tableau consuma completamente l'intero sistema per ottenere risposte rapide alle query su milioni di righe di dati su hardware di base. Poiché il Data Engine può utilizzare

l'archiviazione su disco e la memoria RAM e cache, non è limitato dalla quantità di memoria su un sistema.

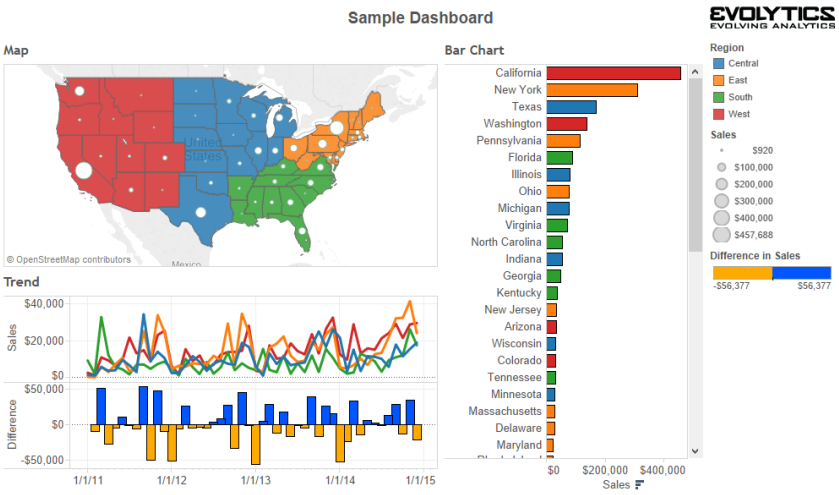


Figura 2.21: Dashboard Tableau

3

Caso di studio: P.o.C. di Location Intelligence in ambito Fashion Retail

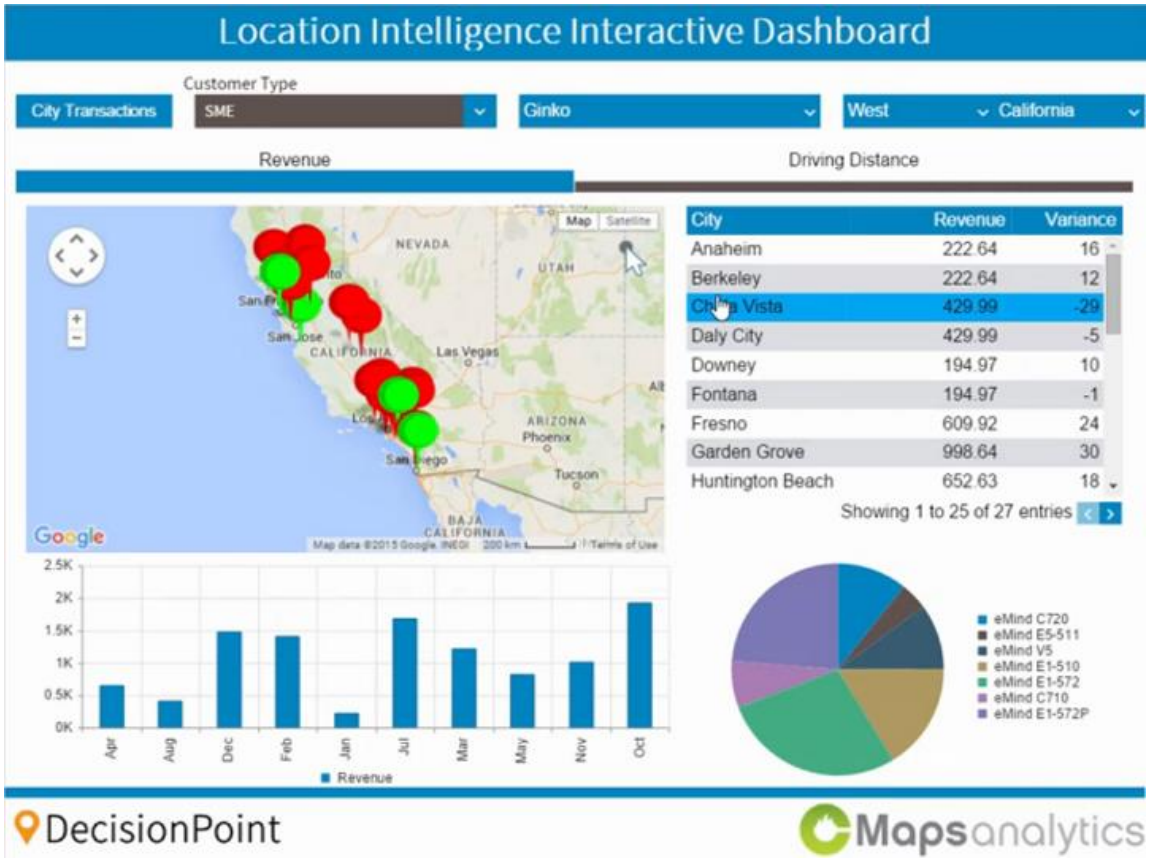


Figura 3.1: Dashboard Location Intelligence

Questo terzo e ultimo capitolo è esclusivamente dedicato alla presentazione del caso di studio, presentando tutti i passaggi effettuati per la realizzazione del progetto e i risultati ottenuti che saranno le basi per gli sviluppi futuri.

Il riassunto del lavoro di 6 mesi è stato presentato sotto forma di "*P.o.C.*". In locuzione inglese *P.o.C. (Proof of Concept)*, che si può tradurre in italiano come *prova di concetto,* si intende un'incompleta realizzazione o abbozzo (sinapsi) di un certo progetto, con lo scopo di dimostrare la fattibilità o la fondatezza di alcuni principi o concetti costituenti. Un esempio tipico è quello di un *prototipo*. In Italia il termine Proof of Concept è utilizzato prevalentemente in ambito informatico e consiste nella dimostrazione pratica dei funzionamenti di base di un applicativo o intero sistema integrandolo all'interno di un ambiente già esistente.

Il P.o.C. è stato realizzato in ambito "*Fashion Retail*", partendo da dati di scontrini relativi a *vendite* sui brand di occhiali di Kering Eyewear citati nel capitolo 2 negli Stati Uniti, sfruttando le potenzialità che offre la *Location Intelligence*. L'obiettivo non è solo di rappresentare il *Sell-out* (le vendite) su mappa ma cercare di integrare questi dati con informazioni relative allo **spazio geografico** come *territorio*(e quindi *popolazione, salario medio e occupazione)* , *dati social* e *flussi turistici*, grazie all'utilizzo di KPI complessi che messi insieme determinano un' **"area potenziale**". Il tutto viene visualizzato tramite dashboard interattive contenenti mappe, grafici e KPI che vanno a raccontare una *storia*, che deve essere perrcettibile e allo stesso tempo accattivante agli occhi di chi la vede.

Figura 3.2: Location Intelligence

3.1 Fasi salienti

Il lavoro è stato realizzato attraverso le seguenti fasi salienti presentate in figura che saranno analizzate nelle seguenti sotto-sezioni:

1. *Raccolta dati*: i dati vengono raccolti da diverse fonti rappresentati da file CSV.

2. *Processo ETL* : i dati vengono estratti, puliti elaborati e caricati all'interno del Data Warehouse SAP HANA tramite SAP Data Services. Si introduce il dato geo-refenziato. Al termine di questa fase i dati sono pronti per essere analizzati.

3. *Analisi sui dati:* Determinazione di KPI semplici e complessi. La visualizzazione delle dashboard è stata realizzata tramite Tableau.

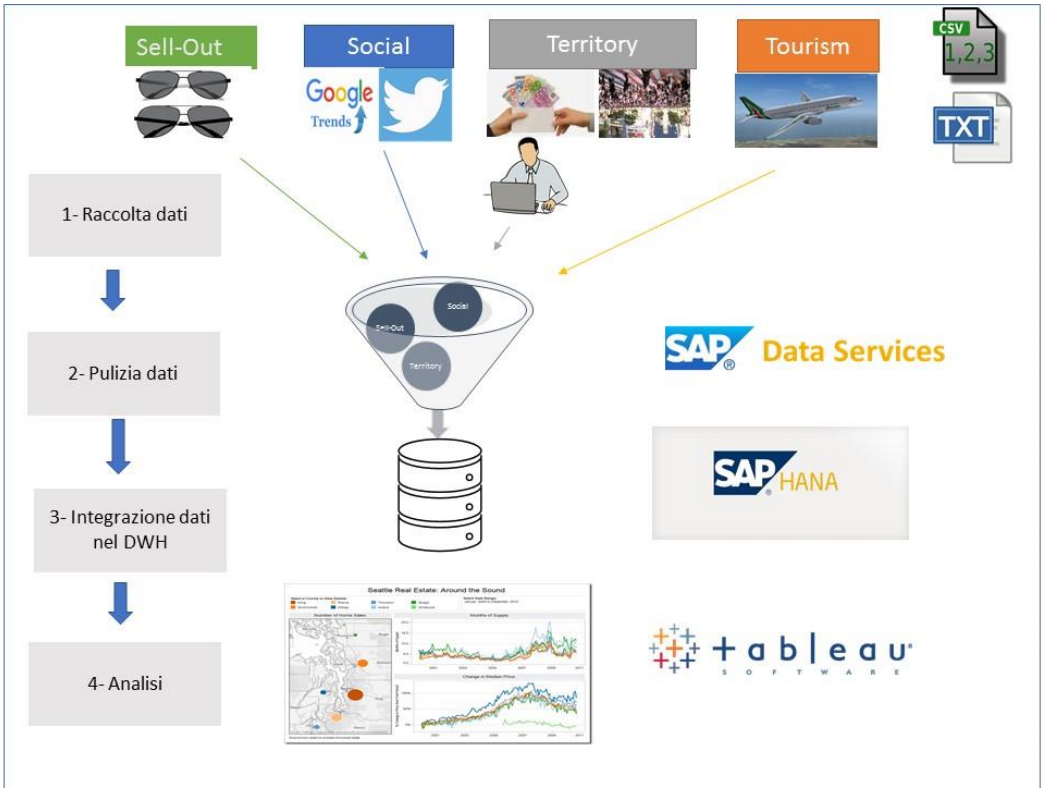


Figura 3.3: Fasi salienti

3.1.1 Raccolta dati

Il P.o.C in ambito "Fashion Retail" è stato realizzato con il principale obiettivo di individuare pattern e analogie interessanti integrando insieme dati di diverso ambito.
La prima fase saliente rappresenta la "*raccolta dati*". Andiamo ad analizzare nello specifico, per ogni ambito, come è stata effettuata questa fondamentale fase:

- *Sell-Out:* rappresentano dati su vendite di occhiali dei brand di Kering Eyewear. I dati vengono forniti da SPS, un system integrator che ha raggiunto un accordo commerciale con Kering Eyewear in quanto possa garantire la trasformazione di scontrini in file csv rappresentando le informazioni in una struttura ben precisa. La procedura è semplificata nella figura sottostante.
  I file contengono informazioni su *negozio di vendita, materiale venduto, prezzo del materiale ,quantità venduta e istante di tempo dell'acquisto.*

Figura 3.4: Raccolta dati di "Sell-out"

- *Social:* un'altra componente fondamentale per la fase di analisi rappresenta il dato "social", un tipo di dato non-strutturato quindi molto ostico da gestire. I social network al giorno d'oggi sono usati costantemente sia per uso personale e quindi per esprimere stati d'animo, pensieri, opinioni riguardo a un evento sia per fare pubblicità/ business. In questo caso l'obiettivo è cercare di capire *quanto si parla del business* riguandante l'analisi e quindi estrarre informazioni *temporali* e *spaziali*. I dati sono stati raccolti dai seguenti social

netwotk seguendo strategie diverse:

- *Twitter:* comunemente diffuso e utilizzato da persone di ogni fascia d'età. L'interazione con Twitter è risultata abbastanza semplice ed accessibile. Le informazioni sono state estratte secondo i seguenti passaggi:
  1. *Registrarsi a Twitter:* per poter richiedere di fare un applicazione è necessario essere registrati a Twitter.
  2. *Richiedere un applicazione:* per fare delle query sul Db di Twitter è necessario richiedere l'autorizzazione e richiede chiavi pubbliche e private.
  3. *Connessione a Twitter:* Attraverso lo scambio di chiavi si è connessi e quindi si è autorizzati ad interagire con Twitter.
  4. *Estrazione dei dati:* grazie alla tecnologia *Python* e in particolare alla libreria *tweepy* è stato realizzato uno script che permettesse di cercare, sfruttando la ricerca per hashtag, informazioni sui brand di Kering e parole correlate come ad esempio "Gucci", "Saint Laurent", "Sunglasses", "Fashion"ecc.. La ricerca permette di reperire anche informazioni su tempo, spazio e numero di Tweet da ricercare. Inoltre per poter effettuare analisi più

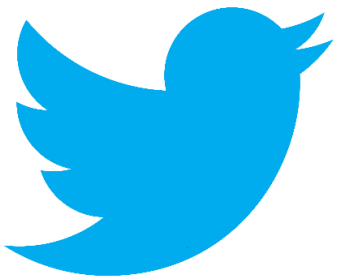


Figura 3.5: Logo Twitter

approfondite sono state ricavate anche informazioni sui brand della concorrenza. I dati vengono salvati su file csv riportando informazioni

sul *tweet* stesso, informazioni *sull'utente* e *location*.

- *Google Trends:* rappresenta un motore di ricerca che offre Google che permette di conoscere la frequenza di ricerca sul web di una determinata parola o frase. I risultati (cioè i *trends*, ovvero le "tendenze" correnti) sono mostrati accompagnando l'occorrenza con un grafico che sintetizza, nel tempo, l'andamento della sua popolarità (ricerca o visualizzazione). Inoltre è possibile effettuare la ricerca contemporanea su più brand riportando informazioni nello spazio e nel tempo. Nel nostro caso sono state scaricati dataset che permettessero di confrontare il brand di maggior rappresentanza di Kering Eyewear, nonchè *Gucci,* con i brand più potenti della concorrenza come *Oakley, Cartier e Vogue* espresso per *un valore di interesse*. Questo confronto viene rappresentato sia negli Stati Uniti che nel mondo intero in due diversi dataset.

Figura 3.6: Logo Google Trends

- *Territory:* rappresentano informazioni che descrivono il territorio degli Stati Uniti, ricavati su Data.Gov un sito web che mette a disposizione Open Data che vanno a descrivere la situazione territoriale per contea. Per ogni contea vengono forniti i seguenti dati: *reddito personale pro-capite, il numero annuo di posti di lavoro a tempo pieno e part-time pro-capite, il salario medio per posto di lavoro in dollari, popolazione e numero pro capite di posti di lavoro*. Il reddito personale pro capite è calcolato come reddito personale totale dei residenti di una contea diviso per la popolazione residente della contea. Le stime annuali della popolazione

a metà anno del Census Bureau sono state utilizzate nel calcolo. Il numero medio annuo di lavori a tempo pieno e part-time comprende tutti i lavori per i quali sono pagati i salari e le retribuzioni, eccetto la giuria e il servizio di testimonianza e l'impiego retribuito di prigionieri. I posti di lavoro vengono contati allo stesso modo e tutti i dipendenti, i proprietari individuali e i partner attivi sono inclusi. I familiari e i volontari non retribuiti non sono inclusi. La retribuzione media per lavoro è rappresentata dalle erogazioni salariali e salariali divise per il numero di posti di lavoro salariali nella contea. Le erogazioni salariali e retributive consistono nella retribuzione monetaria dei dipendenti, compreso il compenso degli esponenti aziendali; commissioni, consigli e bonus; e ricevute in natura, come i pasti forniti ai dipendenti dei ristoranti. Riflette l'importo dei pagamenti erogati, ma non necessariamente guadagnati durante l'anno. Il numero pro capite di posti di lavoro è calcolato come numero medio annuo di posti di lavoro a tempo pieno e part-time in una contea divisi per la popolazione residente della contea. Le stime annuali della popolazione a metà anno del Census Bureau sono state utilizzate nel calcolo. Tutte le stime del dollaro sono in dollari correnti, non adeguate all'inflazione.

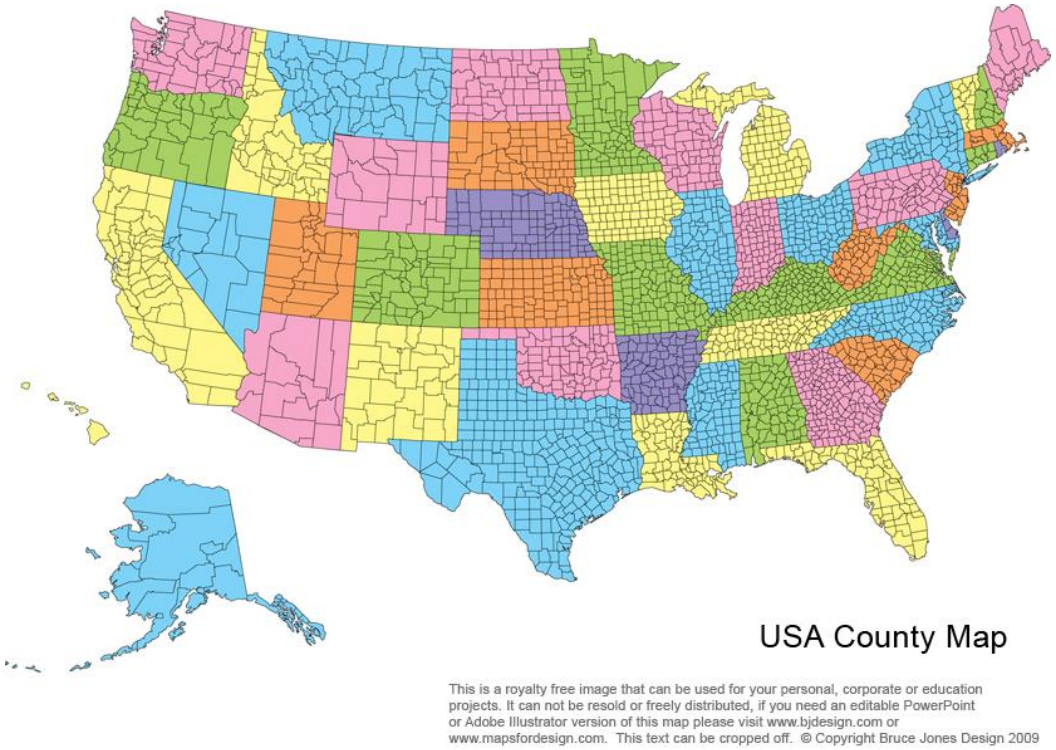


Figura 3.7: Counties Stati Uniti

- *Tourism:* Turismo e Retail sono sempre più spesso coesi ed insieme collaborano per contribuire alla crescita economica del Paese, soprattutto per quanto riguarda il "*travel retail*".

  Il travel retail, tuttavia, non riguarda però soltanto la vendita in aeroporto, ma include quell'attività commerciale al dettaglio svolta in luoghi e rivolta a consumatori in viaggio: aeroporti, autogrill, hotel, outlet. Il punto chiave del "travel retail" è costituito quindi dal flusso turistico in continuo sviluppo.

  Per riuscire a capire quali sono le zone turistiche più calde all'interno degli Stati Uniti si sono scaricati dataset sui *voli esterni* ed *interni*, sempre sotto forma di Open Data, assieme ad informazioni sugli *aereoporti* in modo da poter collegare voli e areoporti in fase di ETL.

Figura 3.8: Turismo Stati Uniti

3.1.2 Processo ETL

La sezione precedente permette di dare uno scenario principale su quali informazioni sono state reperite esplicitando il motivo e la sorgente di esse, in questa sotto-sezione invece andiamo ad analizzare, per ogni ambito, il processo di ETL e quindi estrazione,

pulizia e caricamento dei dati nel DataWarehouse SAP HANA.
Come abbiamo detto in precedenza, per motivi di efficienza e comodità per l'elaborazione dei dati viene utilizzato SAP Data Services.
L'intero processo di ETL si può riassumere nelle seguenti fasi:

1. *Estrazione dei dati:* i dati, che al termine della fase di raccolta sono sotto forma di file csv o txt con un separatore che va a distinguere le colonne, diventano tabelle vere e proprie. In questa prima fase non viene fatto alcun tipo di elaborazione sui dati rappresentando l'informazione esattamente così come è.

2. *Certificazione sui dati:* questa fase è completamente dedicata all'elaborazione dei dati. Ecco le principali operazioni che si fanno su di essi:

    - *Rinominare le colonne:* dare un nome consono alle colonne per rappresentare l'informazione in maniera da rispettare la naming convention anche per facilitare operazioni chiave come il "join".

    - *Trattare il dato nullo:* una delle questioni più importanti da affrontare è sicuramente la gestione del dato nullo. Se manca un' informazione su una colonna potrebbe portare a dei fallimenti del sistema, soprattutto quando si vanno ad effettuare operazioni di join con altre tabelle. Onde evitare questi problemi si va a sostituire il dato mancante con un "dato fittizio", stabilito in base anche alla naming convention, che permette di giustificare la mancanza del dato stesso.

    - *Eliminare dati duplicati:* un'altra situazione da evitare rappresenta il dato duplicato che va ad aggiungere informazioni inutili e superflue. Inoltre come sappiamo, alcune tabelle possono venir identificate da chiavi primarie che per definizione sono univoche, quindi è necessario effettuari controlli su chiavi primarie e/o importate per evitare il dato duplicato.

Al termine di queste operazioni effettuate i dati vengono puliti e quindi "*certificati*".

3. *Integrazione dati nel DWH:* dati vengono trasformati in un formato multidimensionale per facilitare le operazioni di analisi calcolando KPI semplici e complessi.

4. *Presentazione tabella finale:* al termine di tutte le operazioni le tabelle sono esplicitate al massimo livello di dettaglio e sono pronte per essere utilizzate in fase di analisi.

Per ognuna di queste fasi andiamo a vedere i passaggi effettuati in ogni ambito. La fase di estrazione dei dati è comune per tutti. A livello di elaborazione e integrazione dei dati verranno fatte diversi tipi di operazioni in base all'ambito. Andiamo a vedere nello specifico le operazioni effettuate:

- *Sell-out:* una volta caricati i dati su SAP Data Services, vengono rinominate le colonne e trattati i dati mancanti seguendo la naming convention. In seguito viene fatto il controllo di univocività sulle chiavi primarie (*codice cliente del negozio e numero del negozio*) . AL termine di queste operazioni la fase 2 termina e i dati sono certificati. Nella fase 3 i dati vengono aggregati all'interno DWH assieme alle informazioni sulle vendite memorizzate precedentemente.
  All'interno del Data Warehouse è presente una tabella dedicata a tutti i negozi (attivi e non) di Kering Eyewear, contetendo informazioni anche sull*'indirizzo* corrente. Quindi collegando queste due informazioni si ottiengono le vendite sullo spazio. A questo punto, per poter sfruttare tutti i vantaggi che offre la Location Intelligence, è necessario introdurre il tipo di dato spaziale come tipo di dato a se stante, modellando l'informazione a 2 dimensioni (*latitudine e longitudine*) in una tabella che estende quella dei negozi già esistente. Questo processo è sintetizzato nella figura sotto.

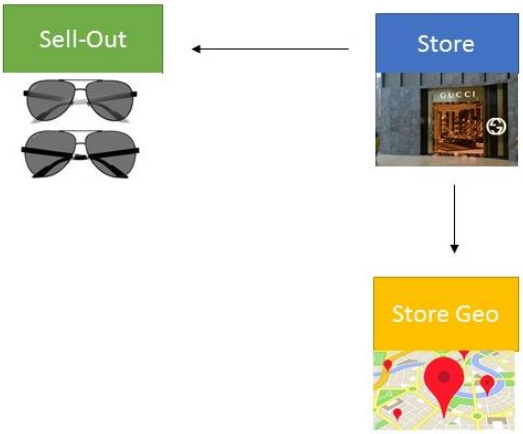


Figura 3.9: ETL Sell-out

- *Social:* come già accennato precedentemente i dati social derivano dalle piattaforme di Twitter e Google Trends. Per entrambe, vengono effettuati diversi tipi di operazioni. Quindi i risultati verranno divisi in tabelle diverse:

  -*Twitter:* dopo essere stati caricati, i dati di Twitter vengono puliti secondo tutte le operazioni citati nella fase 2. Così come nel Sell-Out anche per i Tweet è stato necessario geo-referenziare le location in una tabella a parte.

Figura 3.10: ETL Twitter

-*Google Trends:* anche per Google Trends è stato necessario fare pulizia sugli hashtag. La ricerca viene effettuata per regione nel dataset interno agli Stati

Uniti e per stato nel dataset su ricerca mondiale.

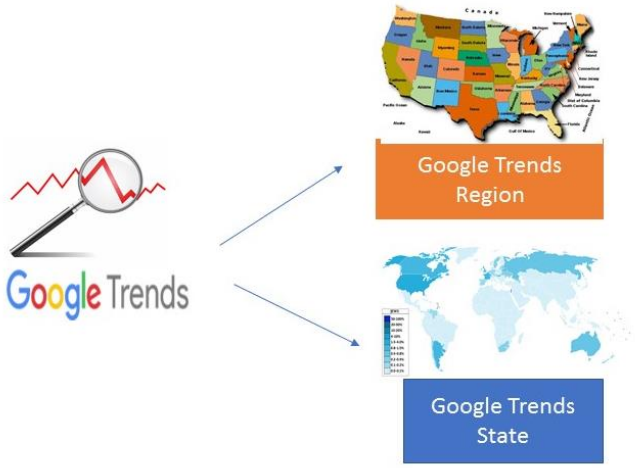


Figura 3.11: ETL Google Trends

- *Tourism*: l'ultimo filone da trattare rappresenta l'elaborazione su dati riguardanti i flussi turistici, che come già accennato si ricavano incrociando i dati sui voli nei rispettivi aereoporti. Dopo le usuali operazioni di pulizia descritte nella fase 2, i dati dei voli vengono incrociati con i dati degli aereoporti che hanno già a disposizione il dato spaziale. Quindi grazie a un semplice join è stato possibile geo-localizzare le tratte dei voli.

Figura 3.12: ETL Turismo

3.1.3 Analisi sui dati

Finito il processo ETL i dati sono rappresentati in tabelle all'interno di SAP HANA al massimo livello di dettaglio permettendo di effettuare analisi in maniera accurata e immediata.

L'obiettivo principale di questa ultima fase è di mettere insieme il tutte le informazioni in maniera incrementale determinando pattern e/o anomalie interessanti che possano attirare l'attenzione dei manager e in particolare del cliente Kering Eyewear. Per visualizzare le informazioni sono state realizzate delle dashboard interattive facilmente interpretabili dagli occhi di chi ha competenze sul campo.

Al termine del capitolo 2 sulle tecnologie sono state argomentate due tipi diversi di tecnologie di data visualization: SAP Design Studio (presente nel pacchetto SAP BO) e Tableau. Essendo un "P.o.C" lo scopo principale è la realizzazione di un prototipo che permetta di racchiudere un analisi completa. La scelta è stata riversata su Tableau, realizzando dashboard in maniera agile con un design accattivante includendo tutte le informazioni necessarie. Design Studio possiede degli elementi di analisi più potenti di Tableau (come ad esempio mappe di calore) ma ha dei tempi di sviluppo notevoli, tuttavia è stata realizzata una parte del lavoro anche con Design Studio ma non ancora terminata per motivi di tempo.

Le dashboard permettono di raccontare le informazioni astratte da ogni ambito tramite KPI, mappe e grafici. Esse possono essere:

- *Esplorative*: viene presentato il singolo ambito focalizzandosi in profondità su esso come *Sell-out,Tweets, Google Trends, US Flights.*
- *Incrociate:* analizzando il Sell-Out con ogni altra coordinata di analisi (*Sell-out and Territory, Sell-out and Tweets, Sell-out and Google Trends, Sell-out and Flights*).

L'ultima dashboard della storia è dedicata all'output che permette di mettere insieme le informazioni determinando **l'area potenziale**.

Andiamo ora ad analizzare dashboard per dahsboard tutte le caratteristiche presentate:

1) *Sell-out:* la prima dashboard non può che essere quella sul Sell-Out, introducendo il principale contesto di analisi. Il KPI estratto da rappresentare è la *quantità venduta* (in alto a sinistra). Il KPI viene argomentato per *Top SKU*(il materiale venduto), *Brand, Release* (comprende anno, versione e codice del materiale venduto) , *Città e Negozio cliente.* Altro aspetto molto importante è la rappresentazione del KPI nel *tempo* permettendo di capire in maniera istantanea, quando è avvenuto il "picco" di vendite, ma la vera e propria novità rappresenta la visualizzazione dei dati su mappa. La grandezza delle bolle in figura è

direttamente proporzionale alla quantità di vendita nello spazio. La dashboard è presentata nella figura sottostante 3.13.

Si può notare che sono state totalizzati *12.144* materiali venduti tra cui il Brand che ha riscosso maggior successo è senza dubbio "*Gucci*" , riscontrando un picco a *Luglio*. Le città dove sono state riscontrate più vendite sono a *San Diego, Dallas e Los Angeles* presso i negozi dei clienti "*Nordstrom*" e "*Neimain Marcus*".

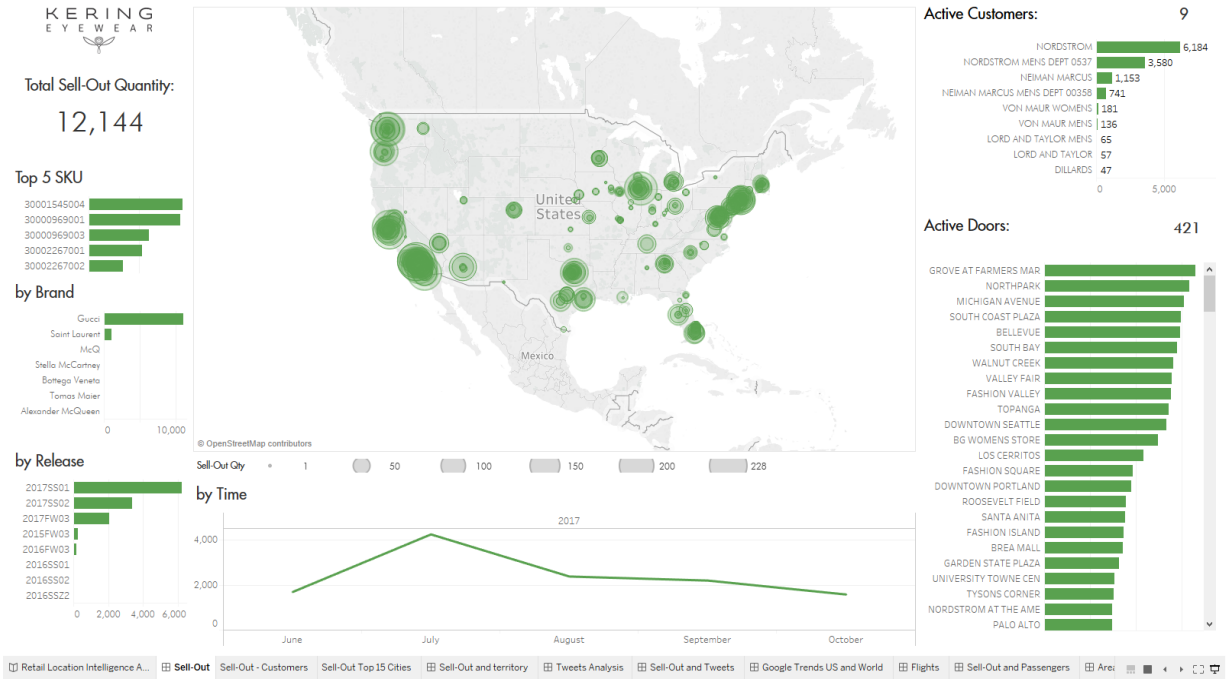


Figura 3.13: Dashboard 1 –" Sell-out(Doors)"

2) *Sell-out and Territory:* la prima dashboard incrociata mette insieme informazioni sulle vendite con dati legati al territorio (quindi considerando le variabili *di popolazione, stipendio medio e occupazione*). Il KPI da evidenziare rimane sempre la quantità venduta (ancora spezzata per Brand, Release e Time) mostrando gli stessi risultati della dashboard precedente, ma grazie ad essa e ai dati sul territorio si sono ricavati KPI complessi pertinenti al tipo di analisi. In alto a destra troviamo il grafico di correlazione tra il Sell-out e i nuovi KPI, contenuti in una una checklist che permette di selezionare il KPI desiderato. La retta che taglia il piano è definita *mediana* e permette di mappare la distribuzione dei dati in base alle coordinate di analisi, i dati che si allontanano dal "centroide" determinano

potenziale (elementi che si trovano sopra la retta mediana) e *anomalie* (elementi che si trovano sotto la mediana) da segnalare. Il filtro applicato cambia ogni barra/grafico o mappa della dashboard in base al KPI selezionato. Andiamo ad analizzare i KPI complessi e vediamo come si sono determinati e per ognuno di essi se ci sono casi interessanti:

- *Sell-Out Qty every 1.000 resident:* KPI che permette di descrivere le vendite sul territorio ogni 1000 abitanti. Come già accennato l'analisi sul terriorio viene fatta per contea. Banalmente il KPI relativo alla popolazione è determinato dal *numero di residenti*. Di conseguenza il calcolo viene effettuato in questo modo:

  Sell-Out Qty every 1.000 resident = SUM([*SellOutQuantity*]) / MAX([*Resident Population*]) * 1000

- *Sell-Out every 1B$ income*: permette di rappresentare il numero di vendite sul prodotto interno lordo della popolazione. In questo fattore incide naturalmente il KPI inerente allo stipendio (*income*) Il calcolo è dato da:

  Sell-Out every 1B$ income = SUM([*SellOutQuantity*]) / MAX(*[Income*]) * 1000000000

- *Sell-Out over Jobs:* rappresenta le vendite sull'occupazione. Il KPI è ricavato dal numero di lavori (*Jobs*). Il calcolo è dato da:

  Sell-Out over Jobs = SUM([*SellOutQuantity*]) / MAX(*[Jobs])* * 1000

- *Doors covarage over population (x 100000):* numero dei negozi (*DoorKeringCod)* di cliente Kering Eyewear ogni 100.000 abitanti :

  Doors covarage over population (x100000) = COUNTD(*[DoorKeringCod])* / MAX([*Resident Population])* * 100000

- *Avarage Sell-Out for door:* vendita media per ogni negozio :

  Avarage Sell-Out for door = SUM(*[SellOutQuantity]*) / COUNTD*([DoorKeringCod])*

- *Income for capita:* rappresenta lo stipendio medio sulla popolazione

  Income for capita = MAX(*[Income*]) / MAX([*Resident Population])*

- *Jobs for capita:* anaologo come il precedente, occupazione su poplazione.

  Jobs for capita = MAX(*[Jobs]) /* MAX([*Resident Population])*

In base ai risultati riscontrati per il calcolo di ognuno di questi KPI complessi andiamo ad analizzare le zone con potenziale e anomalie nella tabella sotto-stante:

	Potenziale	*Anomalie*
Sell-Out Qty every 1.000 resident	Tanta popolazione e tante vendite: *Los Angeles(California)*	Tanta popolazione e poche vendite: *-Wayne (Michigan)* *-Broward (Florida)*
Sell-Out every 1B$ income	Stipendio medio alto e tante vendite: *- Marin(California)* *-Arlinton(Virginia)*	Stipendio medio alto ma poche vendite: *-Farfield (Connecticut)* *-San Mateo (California)* Stipendio medio basso ma tante vendite: *-Multnomah (Oregon)* *-Scott(Iowa)* *-Miami (Florida)*

Sell-Out over Jobs	Tanta occupazione e tante vendite: *-Marin(California)* *-Arlinton(Virginia)* *-King (Washington)*	Tanta occupazione e poche vendite: *-Oklahoma (Oklahoma)* *-St.Luis (MIssouri)* Poca occupazione e tante vendite: *-Contra Costa (California)* *-Placer (California)*
Doors covarage over population (x100000)	Molti negozi che coprono tanta poplazione: *-Marin(California)* *-Arlinton(Virginia)* *-King (Washington)* *-San Francisco(California)*	Molti negozi che coprono poca popolazione: *-New York (new York)* *-San Diego(California)* *Washington(Oregon)* *Los Angelese (California)* Pochi negozi che coprono tanta popolazione: *-Ontario(New York)* *-Black Hawk (Iowa)*

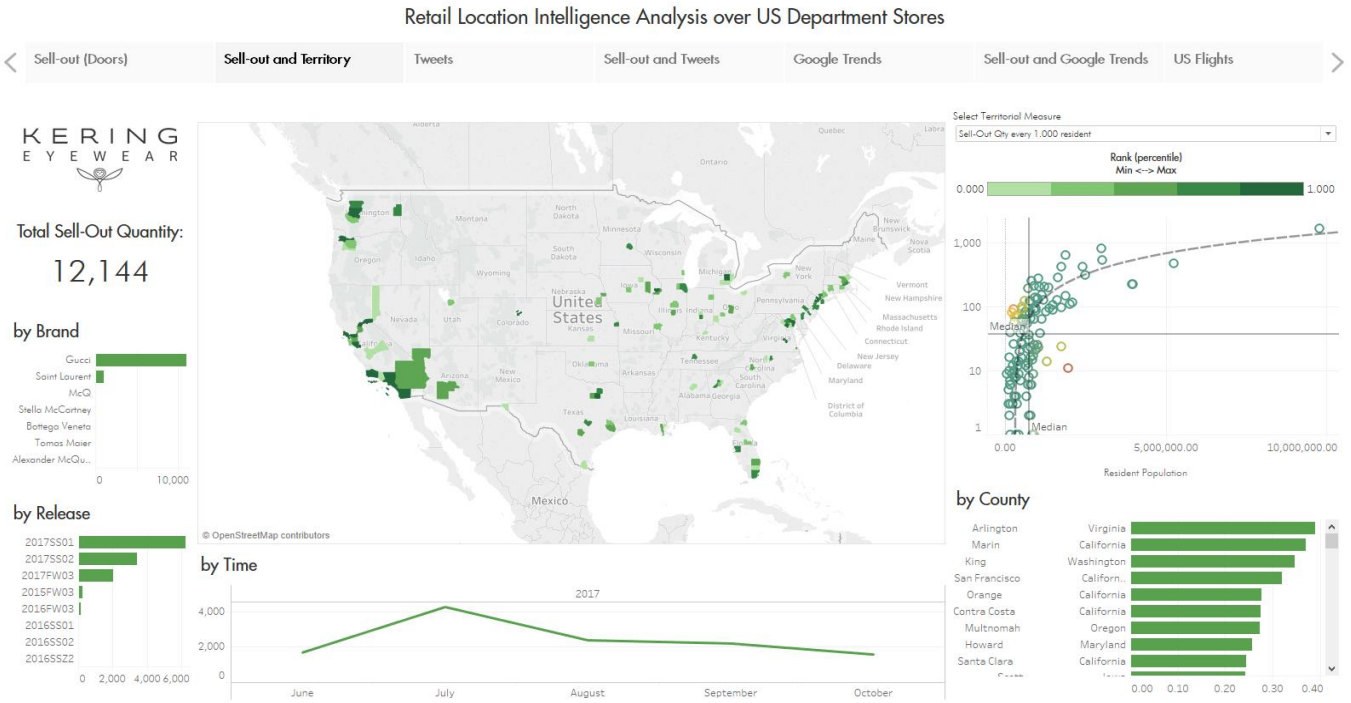


Figura 3.14: Dashboard 2 – "Sell-out and Territory"

3) *Tweets:* la dashboard "social" che viene presentata per prima è quella sui Tweets. Il KPI che viene estratto è il *numero totale di Tweets di interesse*. L'informazione viene spacchettata per *Brand* (di Kering e concorrenza), parole di business correlate interessanti *(" Fashion" , "Style"," Sunglasses"*), visualizzando il tutto su mappa colorando le zone più "chiacchierate". Anche questa dashboard esprime il trend nel tempo nel periodo tra l'11 Ottobre e il 23 di Ottobre (è stato possibile riportare solo 2 settimane di tempo a causa del limite dell'applicazione gratutita Twitter). In base ai risultati il numero dei Tweet totali calcolati in questo asso di tempo è pari a *26.892,* Il Brand più discusso rimane ancora *Gucci,* in vantaggio rispetto ai competitor, spesso correlato a parole come *Fashion, Style* e *Sunglasses*. I luoghi più twittati sono *Missuori, Colorado e Texas* per quanto riguarda invece l'analisi per contea *Hennepin (Minnesota), St.Luis (Missouri) e Denver(Colorado).*

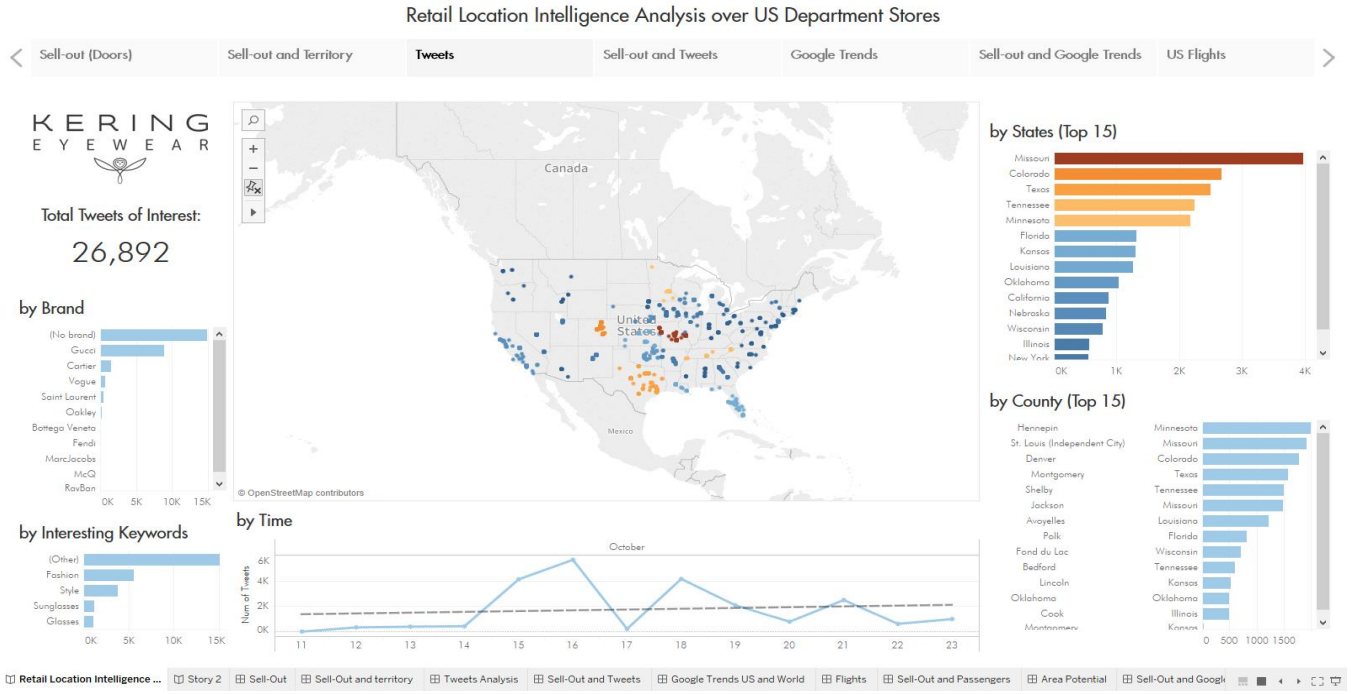


Figura 3.15: Dashboard 3 – "Tweets"

4) *Sell-out and Tweets:* dopo aver introdotto la dashboard esplorativa sui Tweets presentiamo la prossima dashboard che va a incrociare dati di vendita e Tweets, considerando esclusivamente i Brand di Kering Eyewear poiché non abbiamo a disposizione le vendite della concorrenza.

I KPI quindi da mettere in luce quindi sono *la quantità totale venduta(12.444)* e il *numero totale di Tweets (3.171*) inerenti ai Brand di Kering. Il Brand più venduto e chiacchierato è di nuovo "*Gucci*". Come nella dashboard sull'analisi del territorio è stato ricavato il grafico di correlazione tra quantità venduta e numero di Tweets. I casi più interessanti si sono riscontrati a *Los Angeles* in California dove si vende tanto e si parla tanto mentre le anomalie sono state riscontrate a *Orange* sempre in California dove si vende tanto ma non sono stati riscontrati Tweets, ancor più interessanti sono la situazioni notate a *Douglas e Lebraska in Nevraska e Oklahoma* dove si vende poco ma si parla tanto. Il grafico a barre in basso a destra mostra il rapporto Sell-out su Tweets, i leader sono *San Diego (California), King (Washington) e San Franscisco(California).*

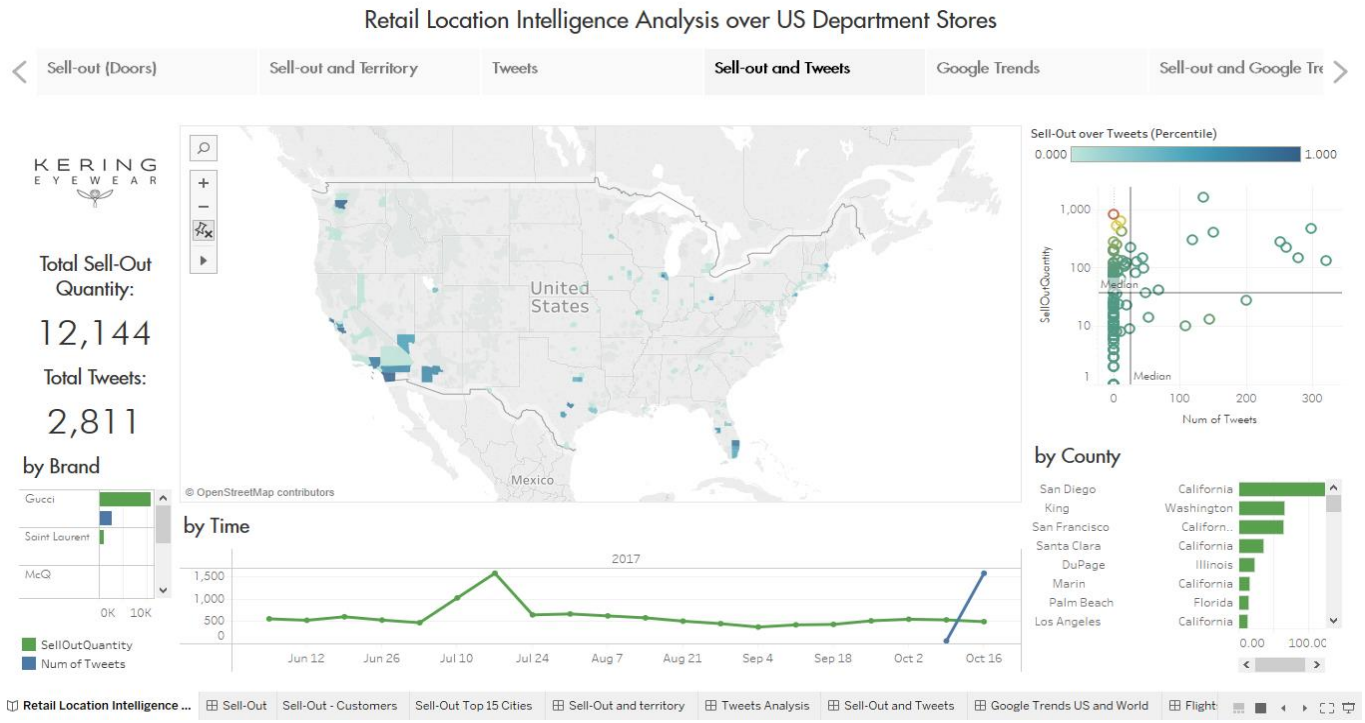


Figura 3.16: Dashboard 4 – " Sell-out and Tweets"

5) *Google Trends:* la seconda dashboard esplorativa "social" rappresenta la informazioni scaricate da Google Trends, come già anticipato viene analizzato il confronto, sia negli Stati Uniti che in tutto il mondo, tra il Brand più popolare di Kering nonché Gucci con i Brand più potenti della concorrenza (Oakley, Cartier e Vogue). Cosi come per i Tweets la valutazione viene effettuata per *valore di interesse medio* (KPI da rappresentare).
Il KPI viene presentato *su mappa, Region* e nel *tempo* colorando questi due ultimi grafici per Brand caratteristico.

- *Stati Uniti:* i luoghi dove si riscontra un valore di interesse più alto sono *New York, California e Distretto di Colombia*, dove in tutte e tre le regioni il Brand più interessante è ancora Gucci che sovrasta la concorrenza l'analisi è effettuata nell'arco di tempo che va da Ottobre 2016 fino a Ottobre 2017 segnalando però dei picchi di Oakley a Gennaio e Marzo.

- *Mondiale:* per avere un analisi più completa vediamo come si riperquote il valore

di interesse non limitandoci solo agli Stati Uniti ma considerando il mondo intero. In questo caso vediamo che le Nazioni dove si parla di più sono *Canada, Italia e Regno Unito* con Gucci che mantiene ancora il primato.

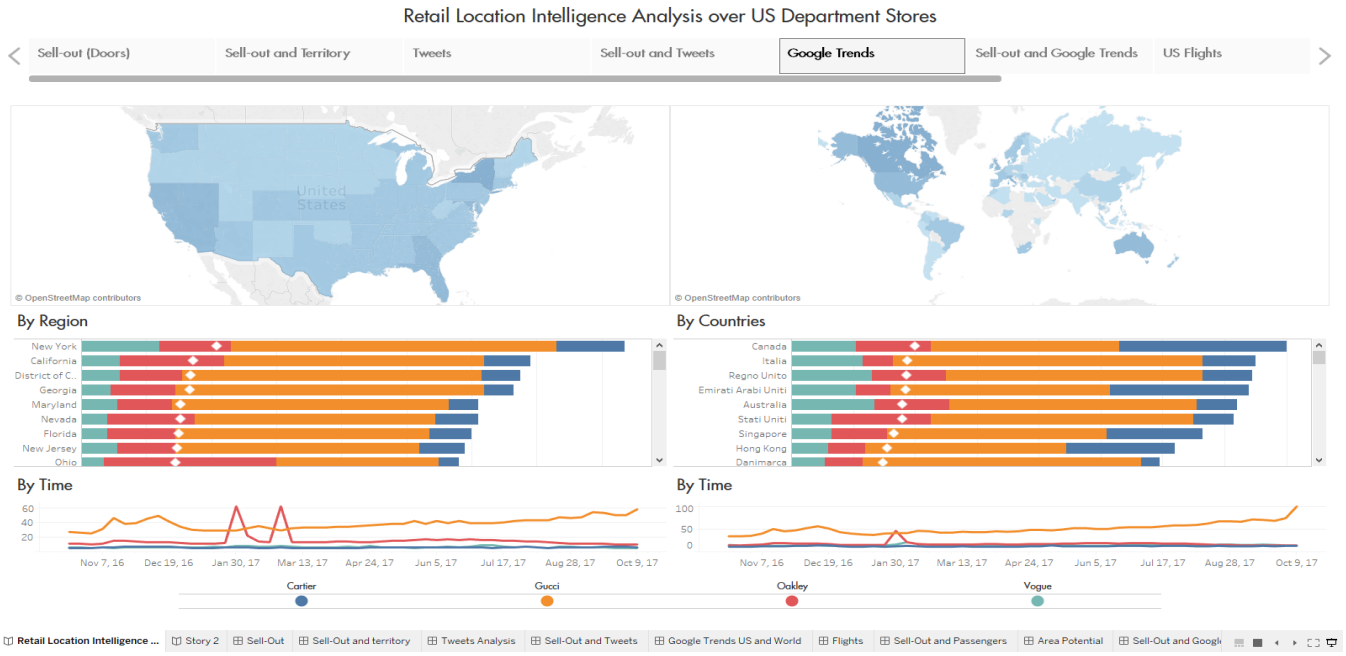


Figura 3.17: Dashboard 5 – "Google Trends"

6) *Sell-out e Google Trends:* dopo aver analizzato il valore di interesse ricavato da Google Trends andiamo a concaterare le informazioni con le vendite nelle contee degli Stati Uniti. La dashboard presenta la mappa, il grafico di correlazione tra i due KPI e il grafico che restituisce il rapporto per contee. La zona che presenta grandi vendite e grandi valori di interesse è ancora la *California,* i casi critici li troviamo a *Washington* dove si vende tanto ma se ne parla poco mentre il caso contrario è segnalato in *Alabama*. Il rapporto Sell-out su interesse è maggiore a *Los Angeles e Orange (California) e King(Washington).*

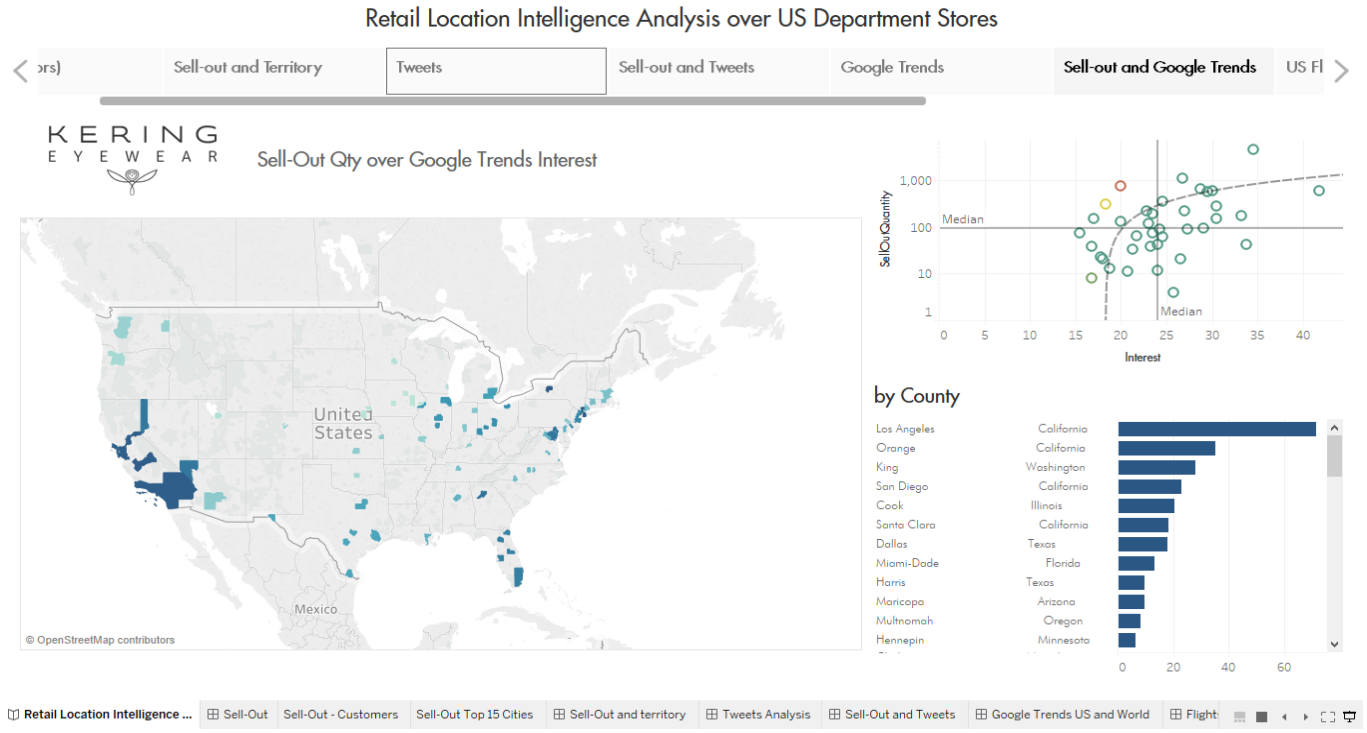


Figura 3.18: Dashboard 6- "sell-out and Google Trends"

7) *Voli:* l'ultima dashboard esplorativa è dedicata al flusso turistico, che abbiamo detto che si riscontra osservando i voli effettuati nel tempo. Il KPI da considerare quindi è banalmente il *numero di passeggeri* presenti nei voli.

   I voli sono catalogati come:

   - *Voli interni:* segnalando quali sono i movimenti che avvengono interni al paese considerando la città di partenza.
   - *Voli esterni:* segnalando quali sono i movimenti che provengono dall'esterno considerando la Nazionalità.

In base a questa distinzione è possibile ricavare quali sono le mete turistiche più calde negli Stati Uniti (tramite voli esterni ed interni).

In base ai risultati ottenuti si analizza che i voli interni sono in maggioranza ad *Athlanta, Chicago* e *Los Angeles*. Le persone che viaggiano di più verso gli Stati Uniti provengono da *Canada, Messico, Inghilterra e Germania* mentre le mete turistiche che riscontrano più voli (da americani e esterni) sono ad *Athlanta, Chicago, New York e Los Angeles.*

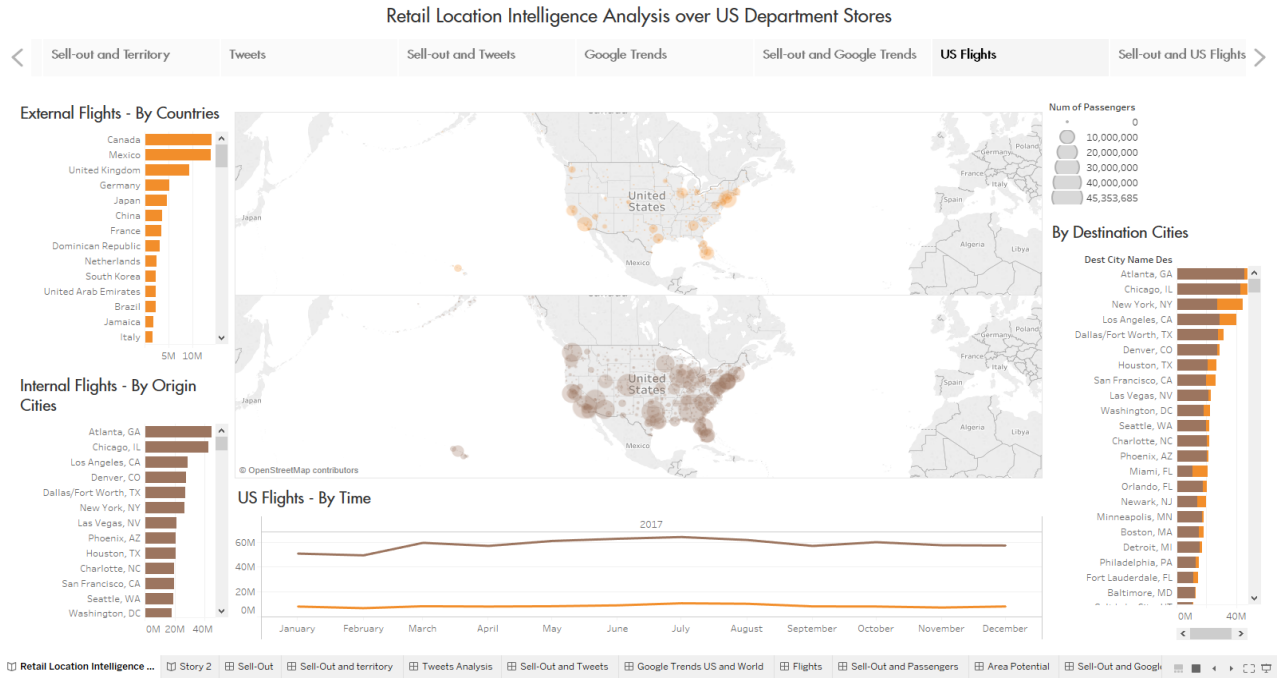


Figura 3.19: Dashboard 7: " US Flight"

8) *Sell-out and Flights:* rappresenta la dashboard che riscontra l'influenza del flusso turistico sulle vendite. Per effettuare un'analisi pertinente si vanno a considerare le vendite ogni 100.000 passeggeri nella relativa contea. Il grafico temporale segnala un picco a Luglio con *4.266* materiali venduti e *11.054.419* passeggeri. Il grafico che esprime la correlazione tra Sell-out e numero di passeggeri riscontra *Los Angeles* ancora in testa proponendo tante vendite in una zona pienamente turistica, mentre le anomalie sono registrate sempre in *California* a *San Mateo* zona turistica ma con poche vendite, questa situazione si riperquote a *Wayne (Michigan)* e *Broward* in *Florida*. Le zone invece che riscontrano vendite con meno turismo sono *Riverside* e *San Bernardino* in *California*.

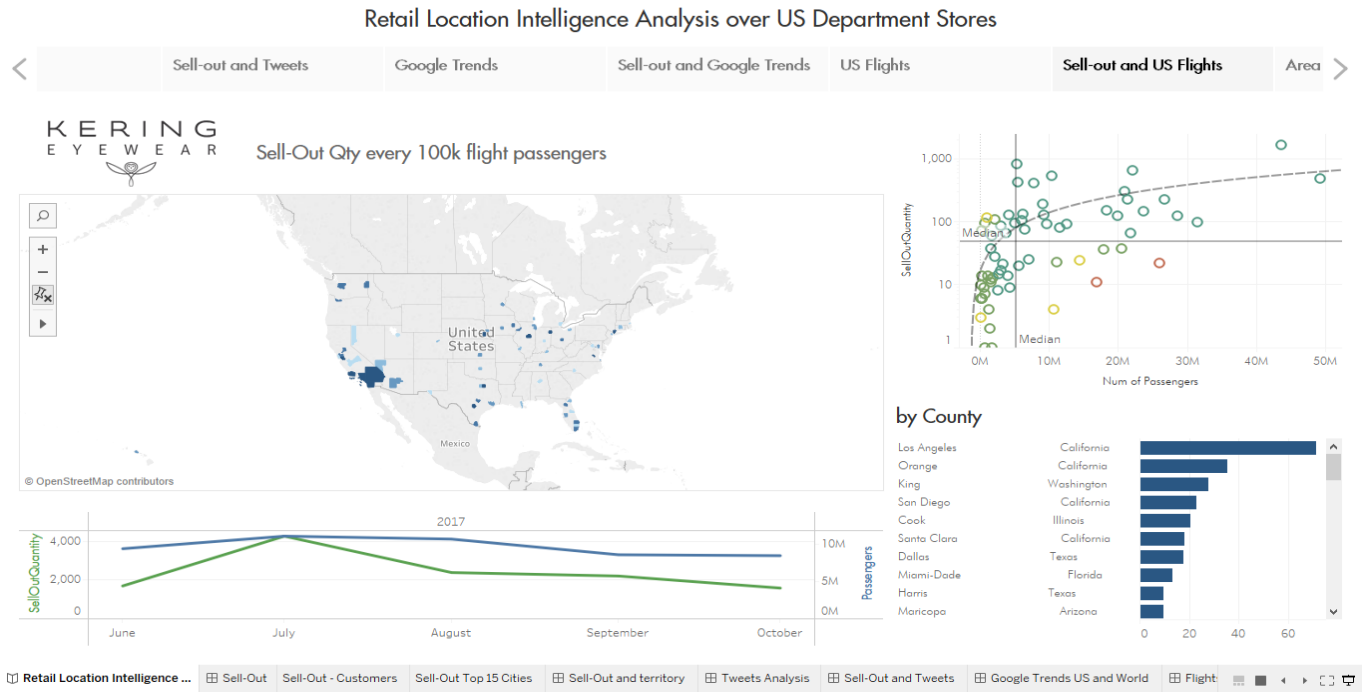


Figura 3.20: Dashboard 8 – "Sell-out and US Flights"

9) *Sell-out and Potential:* In base a tutte le analisi effettuate fino ad ora abbiamo considerato tutte le dashboard incrociate in maniera separata. Questa ultima e importante dashboard permette di connettere le dimensioni permettendo di determinare le **aree potenziali**, avendo quindi a dispozione un output finale unico che riassume tutta l'analisi fatta fino ad ora. Il KPI calcolato è definito come **Potential** ed è stato calcolato mettendo assieme i KPI normalizzati ognuno moltiplicati per un parametro che permette di dare dei pesi (da 0 a 1 con scalo di 0.1) in modo da poter cambiare dinamicamente il suo impatto:

Area Potential = *[ResidentPopulation Weight] * [Resident Population (Normalized)] +*
[External Flights Passengers Weight]] [Num of Passengers (External Flights - Normalized)] +*
*[Num of Tweets Weight] * [Num of Tweets (Normalized)] +*
*[Income Weight] * [Income(Normalized)] +*
*[Jobs Weight] * [Jobs(Normalized)] +*
*[Google Trends Weight] * Google Trends (Normalized)]*

Considerando tutti i pesi massimi le contee che offrono un grande potenziale con tante vendite sono *Los Angeles (California), Cook (Illinois), New York e Denver (Colorado) .*

Il grafico di correlazione Sell-out su Potenziale segnala dei casi critici a *San Mateo (California), Distretto di Colombia , Montgomery (Texas)* dove ci sono poche vendite ma alto potenziale mentre *Washngthon* e *Monomah* ad *Oregon* riscontrano tante vendite ma poco potenziale.

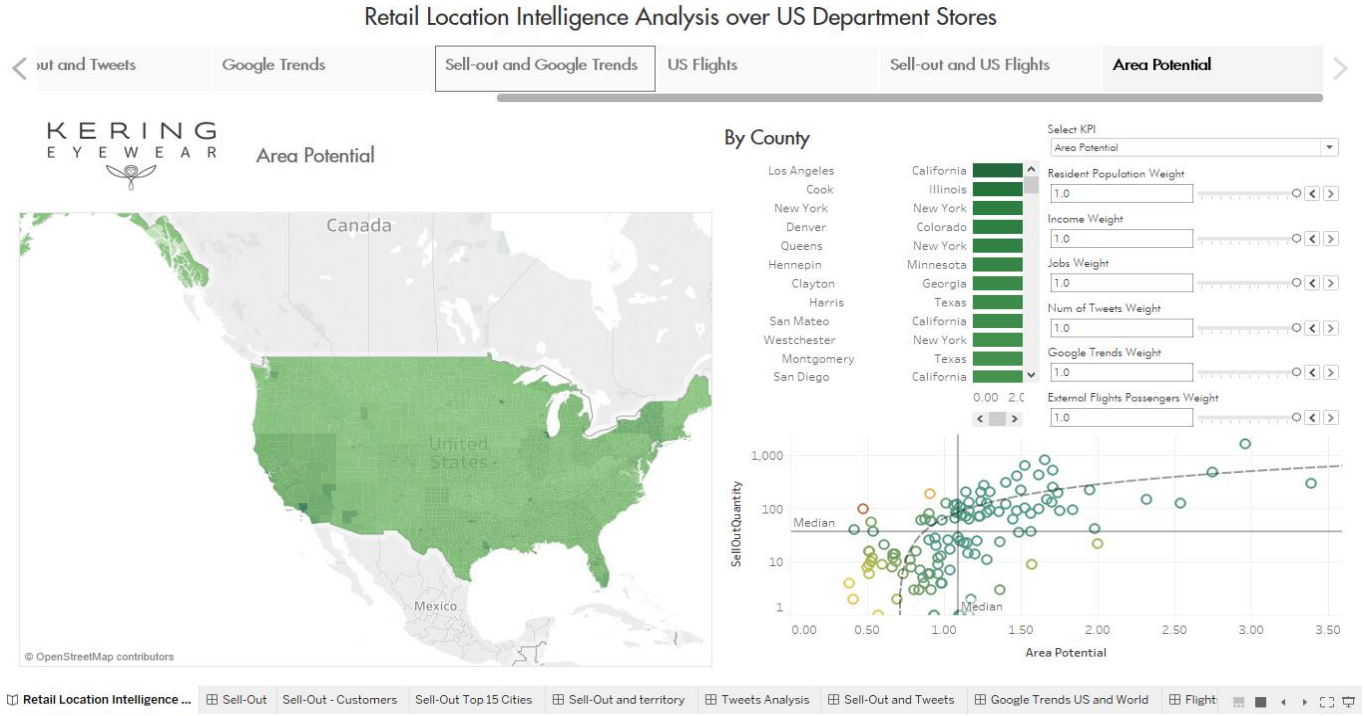


Figura 3.21: Dashboard 9 – "Area Potential"

3.2 Valutazione dei risultati

Finita la fase di analisi, analizzando dashboard per dashboard i risultati ottenuti, arriva il momento di effettuare una valutazione finale sugli output in modo da osservare se si verificano eventuali *pattern.*

Dalla formula del potenziale, calcolato alla fine del capitolo precedente si può notare che :

- tre addendi (population, income e jobs) costituiscono la *dimensione terriroriale.*
- due addendi (Twitter e Google Trends) costituiscono la *dimensione social*.
- un addendo (voli) rappresenta la *dimensione turistica.*

Per vedere meglio quindi l'impatto del territorio , dei social e del turismo sulle vendite vengono calcolati tre nuovi KPI:

Territory = *[ResidentPopulation Weight] * [Resident Population (Normalized)]* + *[Income Weight] * [Income(Normalized)]*
*+ [Jobs Weight] * [Jobs(Normalized)]*

Social = *[Num of Tweets Weight] * [Num of Tweets (Normalized)]* +
*[Jobs Weight] * [Jobs(Normalized)]*

Tourism = *[External Flights Passengers Weight] * [Num of Passengers (External Flights - Normalized)]*

A questo punto riassumiamo i risultati generalizzando per ambito quali sono le zone potenziali e le anomalie.
Le anomalie possono essere classificate come:

- *positive:* se la vendità è alta nonostante il potenziale sia basso;

- *negative:* vendita bassa ma alto potenziale;

Entrambe le anomalie risultano interessanti. Il tutto è rappresentato nella tabella qui sotto che riassume il tutto riportando i 3 top campioni ordinati per valore del KPI relativo decrescente:

	Potential	Anomalie
Sell-out and Territory	*1)Los Angeles(California)* *2)Cook (Illinois)* *3)Orange(California)*	Positive: *1) Washngthon (Oregon)* *2)Cadmen(New Jersey)* *3)Sponake (Washingthon)* Negative: *1)Distict of Columbia,* *2)Fairfield (Connecticut),* *3)San Mateo(California)*

Sell-out and Social	*1)Los Angeles(California)* *2)Cook (Illinois)* *3)San Diego(California)*	Positive: *1)Multnomah(Oregon)* *2)Washington(Oregon)* *3)Snohomigh(Washingthon)* Negative: *1)Montgomery (Texas)* *2)Ontario (New York)* *3)Fresno(California)*
Sell-out and Tourism	*1)Los Angeles(California)* *2)King(Washingthon)* *3)Miami(Florida)*	Positive: *1)DuPage(Illinois)* *2)Montgomery(Pensylvania)* *3)Marin(California)* Negative: *1)Multonah(Oregon)* *2)Jhonson (Indiana)* *3) Washngthon (Oregon)*
Sell-out and Potential	*1)Los Angeles(California)* *2)Cook (Illinois)* *3)San Diego(California)*	Positive: *1)Washingthon(Oregon)* *2)Snohomish(Washingthon)* *3)Spokane(Oregon)* Negative: *1)San Mateo(California)* *2)District of Columbia* *3)Orange(Florida)*

La tabella finale permette di focalizzarsi su pattern ricorrenti e interessanti che permette di

aprire nuovi contesti su cui effettuare nuove analisi. Come si può notare nell'ultima colonna *Los Angeles* riscontra tantisse vendite, facendo parlare tanto sui social dei Brand venduti (Gucci in particolare), in una zona pienamente turistica dove il territorio offre tantissima popolazione e occupazione con stipendio medio alto questo significa che le *strategie di marketing stanno riportando le aspettative attese.*

Le analisi riportate sulle anomalie sono le più interessanti e rappresentano la chiave che apre allo studio di nuove analisi. Per esempio sarebbe interessante capire il perché a *Washigthon* sono state riscontrate tantissime vendite rispetto al potenziale basso, o ancor meglio al contrario perché a *San Mateo* il numero di vendite è strettamente basso per il potenziale che offre la *California.*

4

Conclusione e sviluppi futuri

Il "P.o.C" in ambito Fashion Retail messo a disposizione del cliente è stato presentato sotto forma di progetto di ricerca e sviluppo. Il lavoro incaricato è partito da una richiesta del cliente di elaborazione di dati di vendite inerenti a scontrini in negozi con location statunitense e grazie all'opportunità di poterci effettuare una Tesi sopra (quindi offerto come un servizio in più) si è sviluppato un progetto innovativo che permetterà di aprire un nuovo mondo nell'ambito domanda/ offerta consulenziale. Come da definizione l'obiettivo raggiunto tramite il "P.o.C." è stato quello di realizzare un prototipo dimostrando la fattibilità in maniera parziale di un concetto da rappresentare in tempi non esageratamente lunghi in modo da avere degli output interessanti in maniera "agile". Questo principio permette di avere notevoli sviluppi futuri che potranno occupare il tempo in nuove richieste:

- *Sviluppare il lavoro su Design Studio:* il lavoro è stato realizzato con Tableau per motivi di tempo e praticità, tuttavia il tutto può essere replicato su Design Studio, uno strumento di Data Visualization che permette di realizzare dashboard più complete permettendo anche di calcolare "mappe di calore" in base alla distribuzione dei dati.

- *allargare l'analisi nello spazio e nel tempo:* il progetto è stato realizzato contestualizzando dati sulle vendite di negozi che si trovano negli Stati Uniti nell'arco di tempo da Luglio a Ottobre 2017, presentando quindi un analisi in tempi brevi e in spazi ristretti, si potrebbe pensare quindi di allargare lo studio in

negozi di Kering che si trovano al di fuori degli Stati Uniti con tempi più lunghi. Questo principio permette di vedere come si sviluppa il trend nello spazio e nel tempo vedendo se i risultati si riperquoteranno o se andranno a delinerare nuovi pattern.

- *Allargare il concetto di "Potenziale":* come già detto la formula del potenziale è calcolata sommando tre diverse dimensioni di analisi: *terriorio, social e turismo.* Queste tre dimensioni presentano variabili che possono essere allargate:

    - *Social:* la dimensione social è costituita dalla ricerca sugli hashtag sui Brand di Kering e/o concorrenza effettuata tramite i motori di ricerca di Twitter e Google Trends. Questa dimensione può essere allargata aggiungendo le informazioni estrapolate da *Facebook* e *Intsagram* che assieme a Twitter rappresentano i social maggiormente utilizzati al giorno d'oggi permettendo di vedere se la parte social conferma i risultati trovati o trova altre caratteristiche di output interessanti.
      Un altro filone che si potrebbe aprire riguardo la parte social è quello di effettuare *sentiment analysis* sui risultati trovati cercando di estrarre informazioni sulla valutazione positiva/negativa sui materiali acquistati dai clienti.
    - *Territorio:* la dimensione territoriale è abbastanza ampia e presenta un' analisi completa. Tuttavia un altro concetto che può legare la vendita degli occhiali e il territorio può essere rappresentata dalla *miopia.*
      Capire la percentuale di persone che sono soggette a miopia nel territorio può essere fondamentale per trovare nuovi pattern e incrementare l'interesse di vendite di occhiali (chiaramente da vista).

- *Tarare la formula del potenziale:* Come già spiegato nel capitolo 3 la formula del potenziale è stata calcolata grazie al calcolo di KPI normalizzati ognuno moltiplicati per un parametro che permette di dare dei pesi (da 0 a 1 con scalo di

0.1) permettendo all'utente di "pesare" le coordinate a proprio piacimento. Si potrebbe includere in questa parte un'algoritmo di *Data Mining* che in base alla distribuzione dei dati potrebbe calcolare i pesi in maniera automatica in modo da tarare la formula del potenziale in maniera consona.

- *Analisi sugli output:* in base ai pattern riscontrati come output parte lo studio di una nuova analisi sulle aree potenziali. Gli studi più interessanti su cui focalizzare le attanzioni sono su:
    - *Aree potenziali*: le aree dove le vendite sul potenziale è maggiore deliando quindi che il trend di mercato sta rispettando nella maniera giusta le strategie adottate. Quello che si può fare è cercare di capire da dove derivano i punti di forza e se ci possono essere dei principi che possano incrementare questo potenziale.
    - *Aree con anomalie:* Lo studio sulle anomalie rappresenta sicuramente la parte più interessante su cui focalizzarsi. Come abbiamo già detto, le anomalie sono state classificate come positive e negative in base al rapporto vendita su potenziale. Sulle anomalie "positive" ci interessa sapere perché in luoghi dove il potenziale è basso si sono riscontrate tantissime vendite. Si potrebbero essere verificate eventuali "campagne di marketing" nell'esposizione di nuovi prodotti. Mentre nelle "anomalie negative" abbiamo riscontrato pochissime vendite rispetto all'elevato potenziale che offre la zona. Questo principio può essere causato dalla mancanza di negozi nella zona e quindi in base ai risultati ottenuti pensare di aprire un nuovo punto vendita in una zona "fertile". Tutte queste risposte possono essere svelate grazie allo studio di una nuova analisi.

Bibliografia

[AAS04]]Anandarajan, M., Anandarajan, A., and Srinivasan, C. A., *Business Intelligence Techniques, A Perspective from Accounting and Finance*,
Springer-Verlag Berlin Heidelberg, 2004

[CLA97]Clarke, K., *Getting started with geographic information systems*,
Prentice Hall, 1997

[INM08] Inmon, W. H, *DW 2.0: The Architecture for the Next Generation of Data Warehousing*, Morgan Kaufman Publishers, New York, 2008

[KEL97] Kelly, S., *Data Warehousing in Action*, John Wiley & Sons, 1997

[KER14] Kering, *Kering Eyewear,* 20/03/2015

[LET43] Lettura 43, *Cos'è e come funziona retail,* 30/01/2016

[LUISS15] Luiss Business School, *La professione del Project Manager. Chi è , cosa fa e perche le azienda cercano questa figura? ,* 08/10/ 15

[RF02] Ravaldi, F., *Lesson 02 – Big Data Tecnhiques- Big Data fundamentals ,* 04/11/2016

[RF03] Ravaldi, F., *Lesson 03 - Big Data Tecnhiques -Introduction to Business Intelligence,* 08/11/2016

[RF04] Ravaldi, F., *Lesson 04 – Big Data Tecnhiques – Data Warehouse, Big Data and databases,* 16/11/2016

[RF06] Ravaldi, F., *Lesson 06 – Big Data Tecnhiques – Location Intelligence,* 30/11/2016

[RF09] Ravaldi, F., *Lesson 09 – Big Data Tecnhiques – Social Media Analysis & Social BI,* 14/12/2016

[RF10] Ravaldi, F., *Lesson 10 – Big Data Techinques – Data Visualization,* 14/12/2016

[SAPAG16] SAP HANA Administration Guide, 2016. [Online]

[SAPDASE16] SAP Data Services, Guide references, 2016.

[SAPHAPA16] SAP HANA Predictive Analysis Library (PAL), 2016. [Online]

[SAPHASR16] SAP HANA Spatial Reference, 2016. [Online]

[SAPHATU12] SAP HANA Tutorial, *SAP HANA Architecture Overview*, 2012

[SAPBO16] Administrator Guide: SAP BusinessObjects Design Studio based on SAP BusinessObjects BI Platform, 2016

[RB05]Rich Baker, Tableau Tutorial, *Introduction of Tableau,* 10/11/2005

[PLGR05]Maguire, D., Longley, P., Goodchild, M., and Rhind, D., *Geographical Information Systems and Science*, Wiley and Sons, 2005

[WW07]Watson, H. J., and Wixom, B. H., *The Current State of Business Intelligence*, IEEE Computer (40:9), pp. 96-99, 2007

Ringraziamenti

Dopo sei lunghi anni finalmente arriva il momento di concludere gli studi e di iniziare un nuovo percorso di carriera professionale nel mondo del lavoro, un grandissimo traguardo che è stato conseguito seguendo un percorso specifico ponendosi di volta in volta degli obiettivi e portati al termine con determinazione e voglia di imparare e di migliorarsi. Tuttavia tutto questo non sarebbe stato possibile senza il sostegno di persone a me molto care e ci terrei vivamente di ringraziare in questa sezione dedicata a loro. Ringrazio in primis, i miei genitori Leonardo e Cristina che oltre al sostegno economico hanno saputo aiutarmi a livello morale in momenti di difficoltà dandomi la forza di non mollare e spronandomi a fare sempre di meglio. Sono stati determinanti per avermi fatto vivere questa esperienza con serenità.

Ringrazio mio fratello Giacomo che si è dimostrato un bravo fratello sempre disponibile nei miei confronti permettendo di condividere insieme idee e opinioni su come affrontare al meglio il percorso universitario.

Fondamentali e non di secondo piano sono stati i miei nonni (Virginio , Luciana e Giovanna) che mi hanno cresciuto e mi hanno visto crescere e io compenserò nel regalare a loro questa gioia da condividere e celebrare insieme il prossimo 8 febbraio.

Durante questo percorso oltre i miei familiari sono stati determinanti i miei amici, in particolare ci tengo a ringraziare Michele, Giovanni e Andrea con il quale sto avendo l'onore di trascorrere un bellissimo rapporto da coinquilini in Andrea Costa 8 a Bologna, loro per me rappresentano una seconda famiglia e siamo sempre pronti a sostenerci nei momenti difficile e vivere con gioia i momenti più sereni. Facciamo parte di una compagnia di amici paragonabile a una grande flotta dove ogni elemento è determinante. Insieme portiamo avanti uno spirito di amicizia particolare coltivato giorno dopo giorno sotto una guida di riferimento, il nostro capitano Thomas che ogni giorno ci identifica la rotta da seguire dandoci la giusta carica per affrontare la giornata. I ringraziamenti vanno ad ognuno di loro: Thomas (il capitano),Tommaso,Davide, Rodolfo, Nicolò e Diego.

Ringrazio inoltre i miei amici Guglielmo, Alan, Nicolò, Marco, Alberto con il quale ho passato e

sto passando bei momenti di amicizia.

Gli ultimi ringraziamenti ma non meno importanti vanno a tutte le persone con cui ho avuto l'onore di condividere la mia prima vera e propria esperienza lavorativa in Iconsulting permettendomi di produrre questa tesi. Durante questa esperienza sono cresciuto tanto sia a livello professionale che a livello umano coltivando nuove conoscenze e amicizie. Ringrazio in primis il mio tutor Matteo che mi ha insegnato tantissime cose non solo a livello applicativo ma anche a livello umano, dimostrandosi una gran persona e un grande amico. Ringrazio inoltre Luca, Fabio e Paolo con il quale ho il piacere di condividere momenti extra-lavorativi importanti con ognuno di loro. Ringrazio Tommaso che mi ha spronato più di una volta a fare il massimo dimostrando un grande ineteresse nella mia crescita professionale così come ogni altro elemento del team: Andrea, Leonardo, Luca, Sara, Federico e Alberto. Infine ringrazio Vasil e Federico che mi hanno accolto in questo fantastico team facendomi sentire a casa.