

Laboratorio Big Data Analytics



Progetto finale di:
Belardinelli Marco
Galeazzi Lorenzo
Silvi Pietro

Docente:
Luca Virgili

Corso di Laurea Magistrale in Data Science per l'Economia e
le Imprese

ANNO ACCADEMICO 2022-2023

INDICE

1) QLIK

- 1.1 Descrizione del dataset
- 1.2 Collegamento tabelle
- 1.3 Analisi:
 - 1.3.1 Homepage & Navigazione
 - 1.3.2 Water Withdraw vs Water Consumption
 - 1.3.3 Water Withdraw by Economical Sector
 - 1.3.4 Agricultural Water Withdraw Focus

2) Tableau

- 2.1 Descrizione del dataset
- 2.2 Collegamento tabelle
- 2.3 Analisi:
 - 2.3.1 Home
 - 2.3.2 Team Info
 - 2.3.3 Player Stats
 - 2.3.4 Player Comparison

3) Power BI

- 3.1 Descrizione del dataset
- 3.2 Collegamento tabelle
- 3.3 Analisi:
 - 3.3.1 Home
 - 3.3.2 Analisi delle categorie di prodotti
 - 3.3.3 Analisi temporale per stato
 - 3.3.4 Analisi di un campione di venditori

QLIK

1.1 Descrizione del dataset:

Diversi dataset sono stati scelti e uniti tra loro per la creazione del progetto in Qlik, attingendo dal database "AQUASTAT" della Food and Agriculture Organization of the United Nations (FAO). Questo database colleziona dati sulla gestione dell'acqua e sulle principali risorse di questa presenti in ogni stato fin dal 1960. Per motivi di qualità e quantità dei dati sono stati utilizzati solamente dati dal 1980 in poi.

Per la creazione del progetto sono state prese in considerazione le tabelle sotto riportate:

Dipendenza (rappresenta la dipendenza idrica da altre nazioni):

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Livello di dipendenza espresso in percentuale

Quantità di acqua rinnovabile disponibile:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua rinnovabile disponibile espressa in 10^9 m^3

Quantità di acqua rinnovabile disponibile pro capite:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua rinnovabile disponibile pro capite espressa in m^3

Consumo di acqua totale:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua consumata dalla nazione espressa in 10^9 m^3

Consumo pro capite:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua consumata pro capite espressa in m^3

Consumo di acqua per settori:

Country	Nazioni
Variables	Variabile che identifica i tre settori (agricoltura, industria, servizi)
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua consumata espressa in 10^9 m^3

Settore industriale:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua consumata dalla nazione per il settore industriale espressa in 10^9 m^3

Apporto dell'industria al GDP della nazione:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di denaro espresso in dollari

Settore servizi:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua consumata dalla nazione per il settore dei servizi espressa in 10^9 m^3

Apporto dei servizi al GDP della nazione:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di denaro espresso in dollari

Settore agricolo del paese:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di acqua consumata dalla nazione per il settore agricolo espressa in 10^9 m^3

Apporto dell'agricoltura al GDP del paese:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di denaro espresso in dollari

GDP della nazione:

Country	Nazioni
M49	Codice identificativo delle nazioni
Year	Anno
Value	Quantità di denaro espresso in dollari

Coltivazione:

Country	Nazioni
M49	Codice identificativo delle nazioni
Variables	Variabile che identifica i vari tipi di coltivazione nel paese
Year	Anno
Value	Quantità di terreno occupato dalla coltivazione espresso in migliaia di ettari

Tipo di irrigazione utilizzata:

Country	Nazioni
M49	Codice identificativo delle nazioni
Variables	Variabile che identifica i vari tipi di irrigazione utilizzata nel paese
Year	Anno
Value	Quantità di terreno irrigata espresso in migliaia di ettari

Tipo di acqua per irrigazione utilizzata:

Country	Nazioni
M49	Codice identificativo delle nazioni
Variables	Variabile che identifica i vari tipi di acqua per irrigare utilizzata nel paese
Year	Anno
Value	Quantità di litri usati per irrigare nel paese espresso in 10^9 m^3

1.2 Collegamento tabelle

Le tabelle sono collegate tra loro nel seguente modo:

Fonte dati: https://tableau.apps.fao.org/views/ReviewDashboard-v1/country_dashboard?%3Aembed=y&%3AisGuestRedirectFromVizportal=y

1.3 Analisi

Sono stati creati tre fogli al fine di analizzare, sotto diversi punti di vista, il consumo di acqua. In particolare si è scelto di approfondire la quantità di risorse idriche disponibili e consumate nei diversi paesi nel tempo, anche a livello pro-capite. Ci si è poi focalizzati sull'utilizzo dell'acqua in diversi settori economici considerando anche l'apporto degli stessi in termini di GDP. In ultimo si è considerata l'agricoltura, un settore particolarmente legato all'acqua, per la quale sono stati approfonditi fonti e metodi di irrigazione.

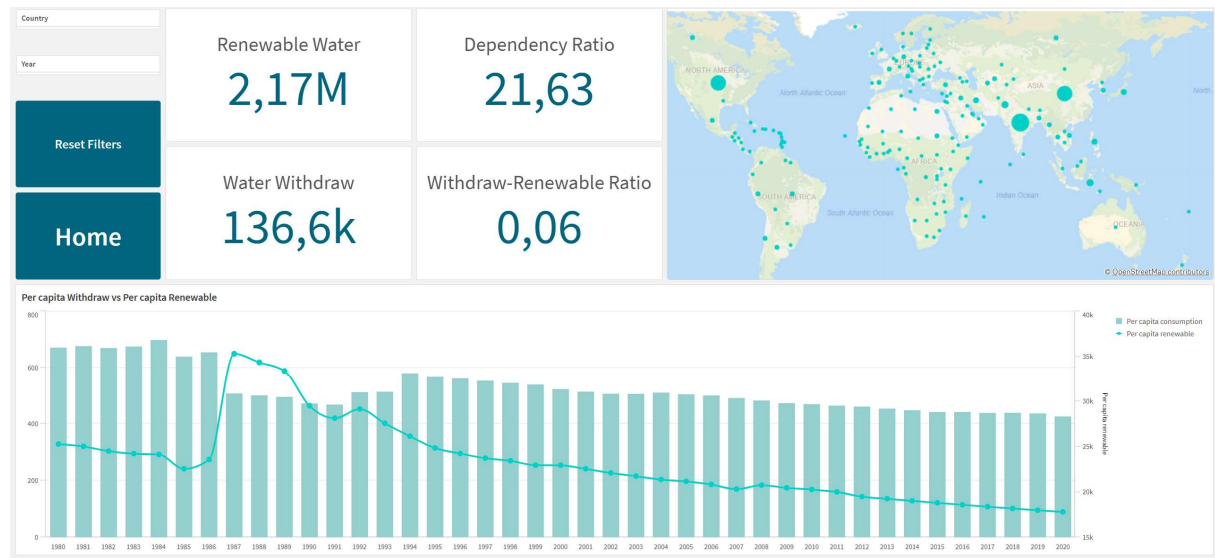
1.3.1 Homepage & navigazione

È stata creata una dashboard iniziale al fine di semplificare l'utente nella navigazione tra dashboard.



Cliccando quindi uno dei vari pulsanti è possibile raggiungere la dashboard interessata, in ognuna di esse è infine presente un tasto "HOME" per permettere all'utente di ritornare alla homepage.

1.3.2 Water withdraw vs Water Consumption

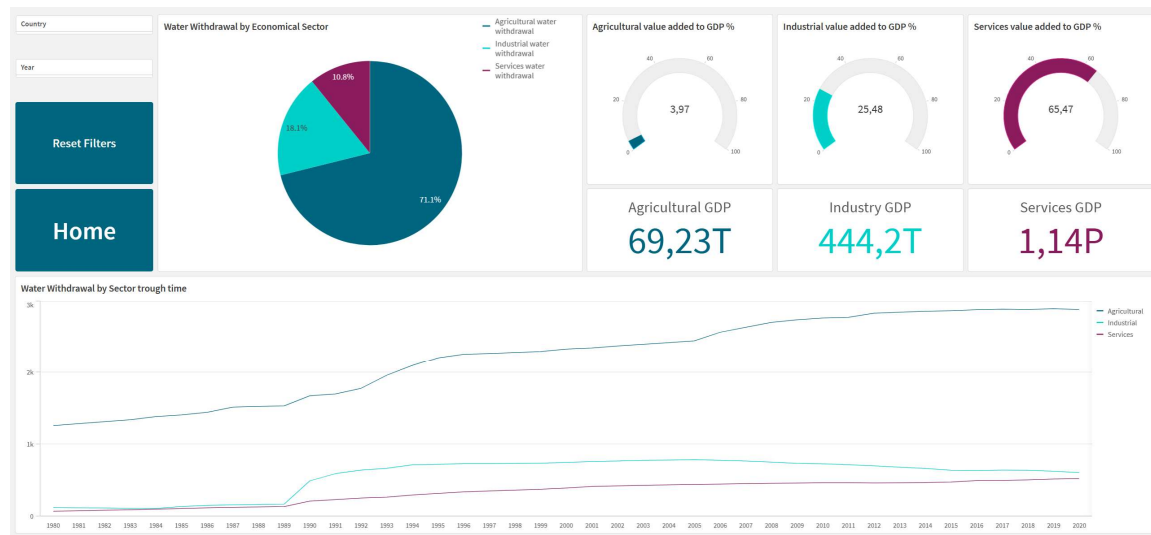


Il primo foglio fornisce una panoramica sulla quantità di acqua rinnovabile presente nel mondo e su quanta di questa viene utilizzata dall'uomo, nello specifico:

- I quattro KPI in alto a sinistra mostrano: quantità di acqua rinnovabile disponibile, consumo di acqua, indice di dipendenza (intesa come dipendenza idrica dalle nazioni limitrofe), rapporto tra acqua rinnovabile e consumata.
- La mappa in alto a destra mostra il consumo di acqua nelle nazioni presenti nel dataset.
- Il grafico in basso mostra l'andamento nel tempo di: consumo di acqua pro-capite (barre) e disponibilità di acqua rinnovabile pro-capite (linea).

L'utente ha la possibilità di eseguire operazioni di filtraggio sia attraverso i filtri posti nella parte in alto a sinistra del foglio (stato e anno), sia selezionando autonomamente uno stato dalla mappa interattiva o selezionando un anno dal grafico temporale.

1.3.3 Water Withdraw by Economical Sector



Il secondo foglio mette in risalto la differenza in termini di consumo di acqua che c'è tra tre differenti settori economici: agricoltura, industria e servizi.

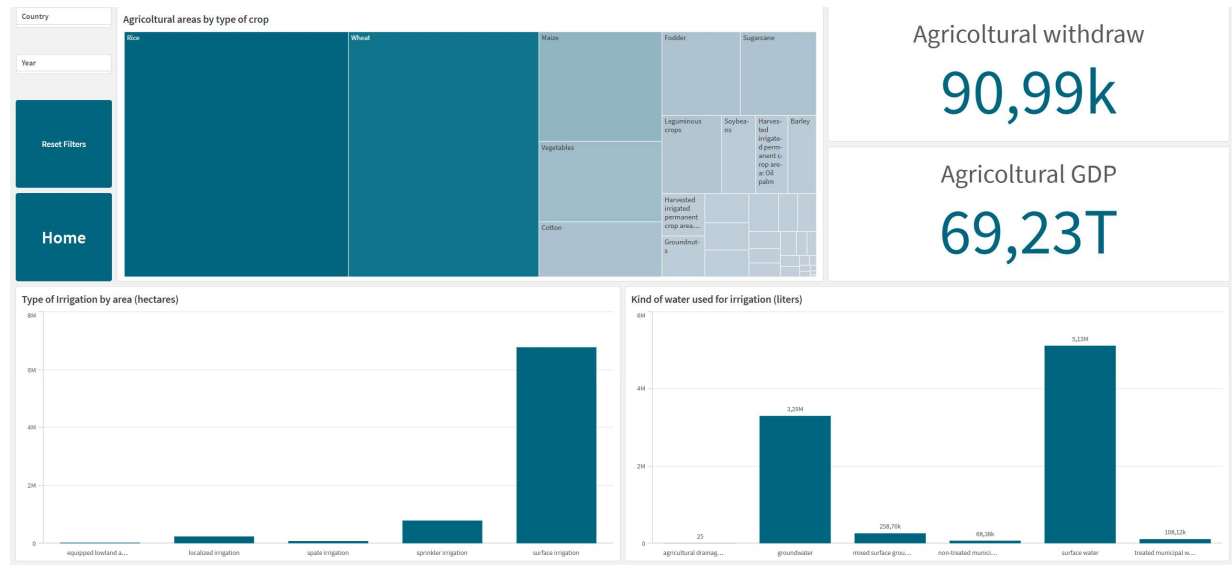
Nello specifico nel foglio possiamo trovare:

- Grafico a torta in alto a sinistra che mostra la percentuale di acqua utilizzata da ogni settore.
- Tre misuratori, in alto a destra, che mostrano quanto, in percentuale, un settore abbia contribuito al GDP del paese. Al di sotto di ognuno il rispettivo KPI mostra la quantità di ricavi generata dal settore in USD.
- Grafico a linee, in basso, che descrive l'andamento nel tempo del consumo di acqua dei settori presi in analisi.

È possibile, come nella prima dashboard, filtrare per anno e stato di interesse.

È interessante notare come l'agricoltura, nella maggior parte dei paesi, sia contemporaneamente il settore con più alti consumi idrici e quello che contribuisce meno in termini di GDP. In virtù di questo elevato consumo e dell'importanza del settore per la sussistenza umana abbiamo deciso di dedicare un foglio al suo approfondimento.

1.3.4 Agricultural Water Withdraw Focus



Il terzo e ultimo foglio mira ad effettuare un'analisi approfondita del settore agricolo, sono presenti:

- Due KPI, in alto a destra, riguardanti il consumo d'acqua per l'agricoltura e l'apporto di questo settore al GDP del paese.
- Una TreeMap in alto a sinistra che mostra il tipo di coltivazione piantata, ordinato per quantità di ettari occupati da ciascuna coltura.
- Il grafico a barre di sinistra mostra il tipo di irrigazione utilizzata per la coltivazione per area irrigata. Ricordiamo che diversi tipi di irrigazione possono essere usati sulla stessa coltura.
- Il grafico a barre di destra mostra, infine, il tipo di acqua (misurata in litri) utilizzata per irrigare la coltura.

Anche qui è possibile filtrare per stato e anno di interesse.

TABLEAU

2.1 Descrizione del dataset

Il dataset utilizzato deriva dall'unione di più tabelle relative a giocatori e team dell'NBA. Di seguito una breve descrizione di ciascuna tabella.

Salari

playerName	Nome e cognome dei giocatori
seasonStartYear	Anno di inizio della stagione. (Es. per la stagione 2015-2016 troveremo 2015 nel dataset)
salary	Stipendio annuale del giocatore
inflationAdjSalary	Valore attuale (2023) dello stipendio annuale dei giocatori.

Fonte: <https://www.kaggle.com/datasets/loganlauton/nba-players-and-team-data?select=NBA+Salaries%281990-2023%29.csv>

NBA Player Stats(1950-2022)

Per via delle numerose variazioni nel regolamento NBA si è scelto di considerare solo le stagioni dal 1990 al 2022 così che vi fosse maggiore comparabilità tra i giocatori.

Season	Anno di inizio della stagione. (Es. per la stagione 2015-2016 troveremo 2015 nel dataset)
Player	Nome e cognome dei giocatori
Pos	Ruolo del giocatore
Age	Età del giocatore l'anno di inizio stagione
Tm	Abbreviazione della squadra per cui il giocatore ha giocato in stagione. Le 30 abbreviazioni presenti nel dataset fanno riferimento alla NBA attuale. Per le franchigie che hanno cambiato città o hanno subito altre vicende societarie si è proceduto a ricostruire tali avvenimenti per ricondurle a quelle attuali
G	Numero partite giocate in stagione
GS	Numero partite in cui il giocatore era titolare in stagione
MP	Minuti giocati in stagione
FG	Tiri segnati dal campo(2P+3P) in stagione
FGA	Tiri tentati dal campo in stagione
FG%	Percentuale realizzativa dal campo in stagione o
3P	Tiri segnati da 3 punti in stagione
3PA	Tiri tentati da 3 punti in stagione
3P%	Percentuale realizzativa da 3 punti in stagione
2P	Tiri segnati da 2 punti in stagione
2PA	Tiri tentati da 2 punti in stagione
2P%	Percentuale realizzativa da 2 punti in stagione
eFG%	Effective field goal percentage in stagione
FT	Tiri liberi realizzati in stagione
FTA	Tiri liberi tentati in stagione

FT%	Percentuale realizzativa dal tiro libero in stagione
ORB	Rimbalzi offensivi conquistati in stagione
DRB	Rimbalzi difensivi conquistati in stagione
TRB	Rimbalzi conquistati in totale (ORB+DRB) in stagione
AST	Assist realizzati in stagione
STL	Palle rubate in stagione
BLK	Stoppate in stagione
TOV	Palle perse in stagione
PF	Falli commessi in stagione
PTS	Punti realizzati in stagione

Fonte: <https://www.kaggle.com/datasets/loganlauton/nba-players-and-team-data?select=NBA+Player+Stats%281950+--+2022%29.csv>

Team Details

Abbreviation	Abbreviazione della squadra
Nickname	Nome della franchigia
Yearfounded	Anno di creazione della franchigia
City	Città in cui gioca la squadra
Arena	Arena della squadra
arenacapacity	Capienza (in persone) dell'arena
Owner	Proprietario della franchigia
generalmanager	General manager
Headcoach	Allenatore
dleagueaffiliation	Squadra di D-league affiliata
Latitude	Latitudine dell'Arena (ricavata tramite Google Maps)
Longitude	Longitudine dell'Arena (ricavata tramite Google Maps)

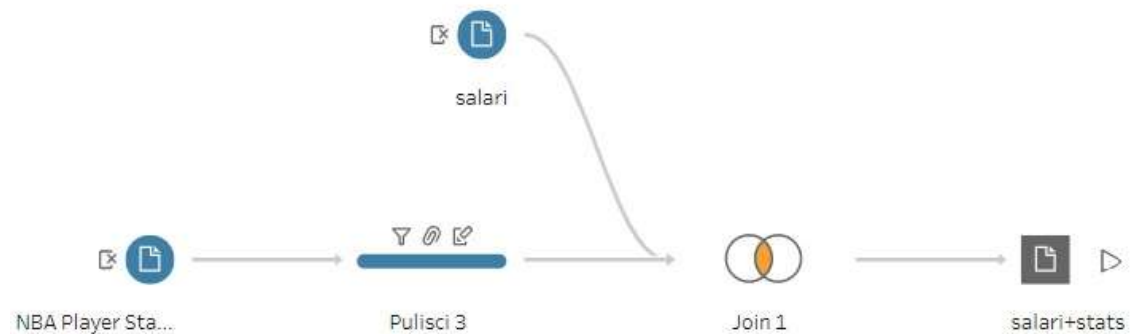
Fonte: <https://www.kaggle.com/datasets/nathanlauga/nba-games?select=teams.csv>

Record finals

Abbreviation	Abbreviazione della squadra
City	Città in cui gioca la squadra
Nickname	Nome della franchigia
Finali Vinte	Finali NBA (no conference) vinte
Finali Perse	Finali NBA (no conference) perse
Finali Totali	Finali NBA (no conference) disputate
Anni del titolo	Anni in cui la squadra ha vinto le finali

Fonte: webscraping + <https://www.dunkest.com/it/nba/notizie/22773/record-titoli-squadre-nba>

2.2 Collegamento Tabelle



La tabella NBA Player Stats (1950-2022), precedentemente ridotta per via delle numerose variazioni nel regolamento NBA che hanno portato a scegliere di considerare solo le stagioni dal 1990 al 2022 ha subito una fase di modifica tramite Tableau Prep Builder. Dalla colonna Tm sono state escluse le righe che presentavano il valore TOT al fine di evitare duplicati, sempre sulla stessa colonna sono stati raggruppati alcuni valori al fine di ricondurre le denominazioni degli anni '90 a quelle attuali.

Le tabelle salari e NBA Player Stats (1950-2022) sono state unite tramite inner-join basato sulle colonne playerName-Player e seasonStartYear-Season.

Si è poi proceduto a instaurare una relazione con la tabella team details tramite il campo abbreviation e Tm. Una relazione è stata instaurata anche con la tabella record finals tramite la sua colonna abbreviation.

2.3 L'Analisi

Con questa analisi si vuole offrire la possibilità di consultare in maniera rapida e confortevole le principali informazioni riguardanti il mondo NBA, sia per quanto riguarda le squadre che i giocatori. I dati presentati sono organizzati secondo diverse modalità (squadra, ruolo, stagione, ecc.) così da consentire diversi livelli di analisi e di interazione.

Sono state realizzate quattro dashboard:

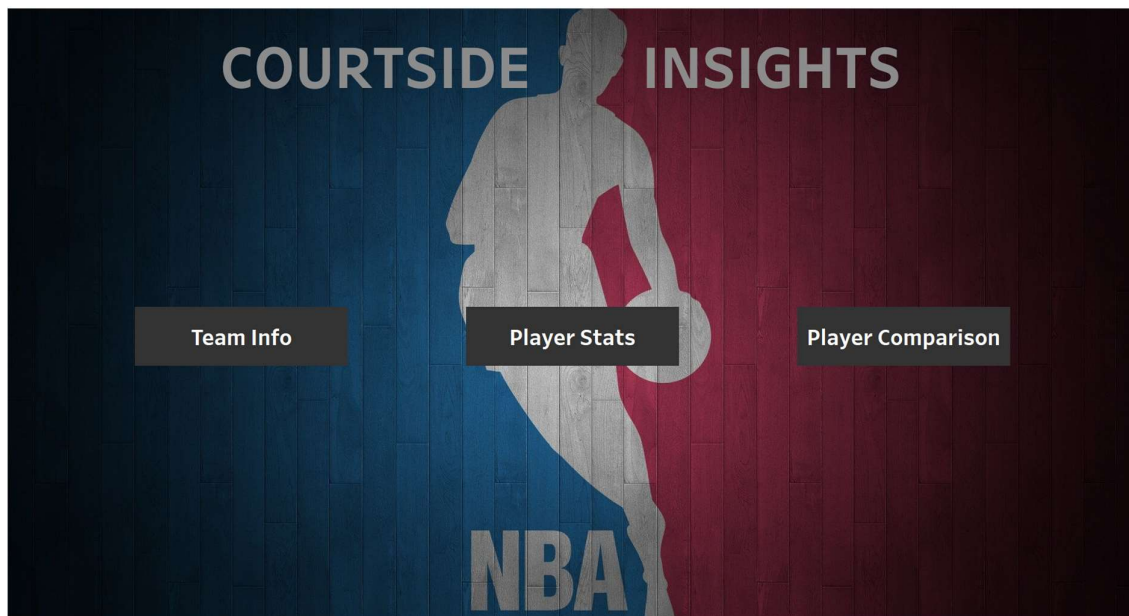
- “Home” la cui funzione è quella di consentire la navigazione tra le altre dashboard, è provvista di collegamenti a ciascuna di esse. Da ogni altra dashboard è possibile tornare alla home.

- “Team info” contiene informazioni riguardanti le diverse franchigie, le localizzazioni delle arene, le spese salariali e i giocatori titolari.

- “Player stats” contiene le principali statistiche relative ai singoli giocatori tra cui minuti giocati e stipendio annuale.

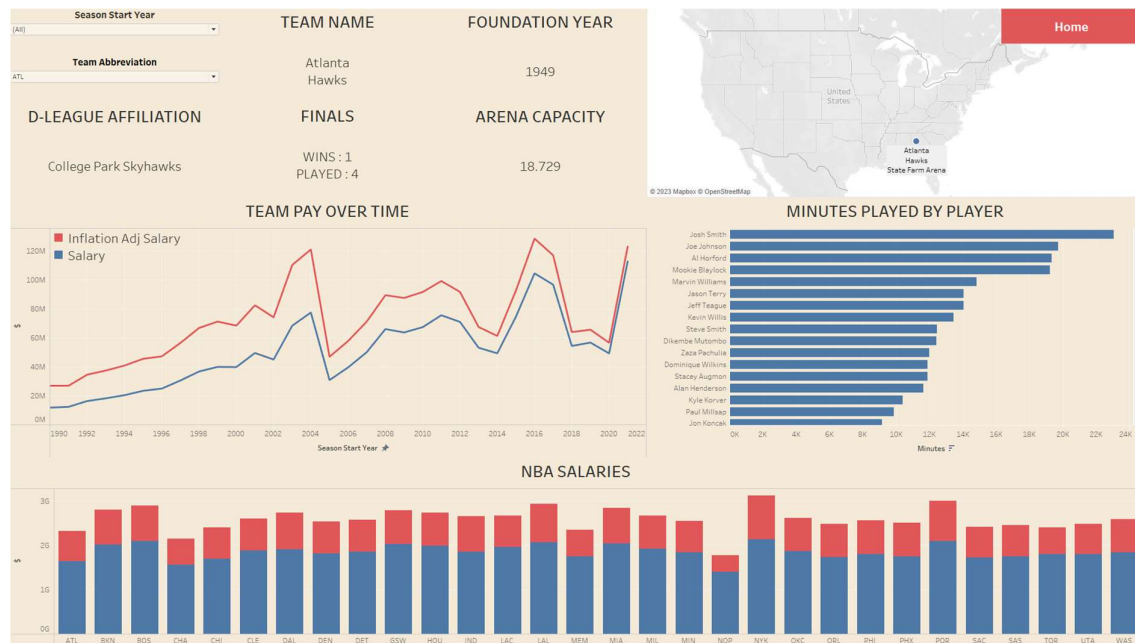
- “Player comparison” permette di comparare le prestazioni di un giocatore ad altri secondo diversi criteri e di confrontarlo con i migliori marcatori all-time per diverse tipologie di tiri.

2.3.1 Home



La Home si presenta come una semplice pagina che consente la navigazione verso le altre dashboard tramite i tre pulsanti. In ciascuna delle altre dashboard sarà presente un tasto “Home” così da poter agevolmente tornare alla pagina iniziale e da qui spostarci nuovamente in base alle nostre necessità.

2.3.2 Team info



Questa dashboard presenta due filtri: SeasonStartYear e TeamAbbreviation.

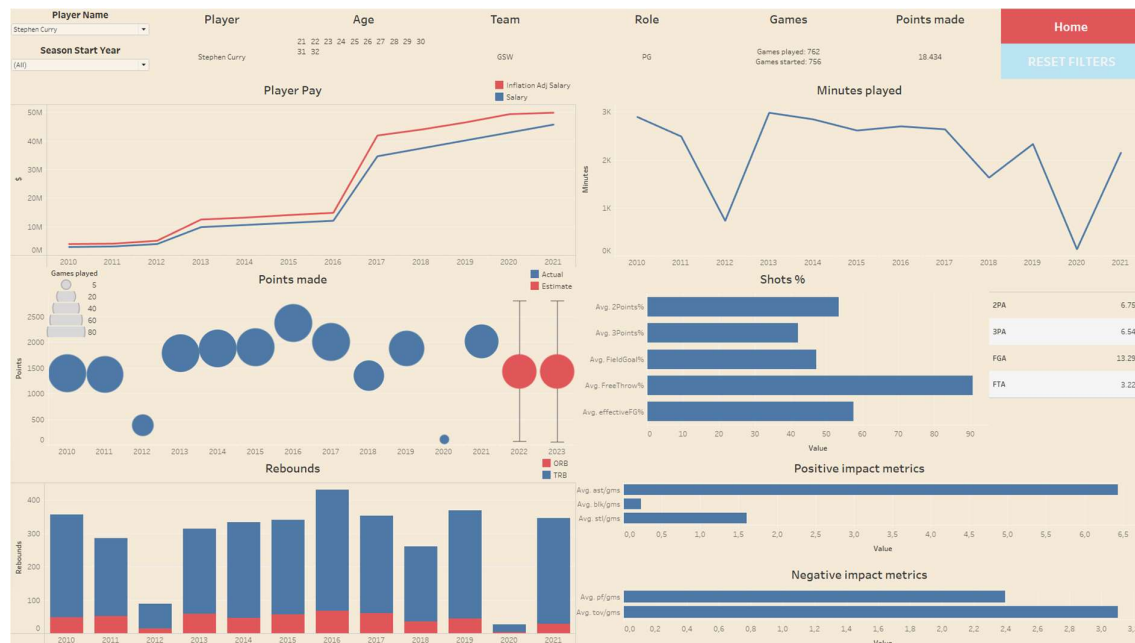
Le informazioni generali sulla squadra (Team name, foundation year, D-league affiliation, finals, arena capacity), la mappa e il grafico a linee che rappresenta l'andamento dei salari nel tempo sono soggetti al solo filtro TeamAbbreviation. Questo filtro consente di selezionare la sigla di ciascuno dei 30 Team della lega aggiornando di conseguenza queste rappresentazioni.

Anche il grafico a barre relativo ai minuti giocati dai diversi giocatori è soggetto a questo filtro. Il grafico è inoltre ordinato in base ai minuti mostrando quindi una classifica dei giocatori che sono stati in campo più a lungo per il team, questo quando il filtro SeasonStartYear è impostato su "All". Modificando il valore del filtro SeasonStartYear è infatti possibile scoprire quali giocatori hanno passato più minuti in campo in una specifica stagione ricavando quello che di fatto è il quintetto titolare.

Il grafico NBA salaries rappresenta la spesa in stipendi per tutti i 30 team, di conseguenza si aggiorna in base al solo filtro SeasonStartYear.

La dashboard nel suo complesso ci consente una rapida ricostruzione della storia della franchigia, con i suoi giocatori chiave e le sue infrastrutture. Ci consente inoltre di capire in che anni ha optato per una maggiore spesa salariale e di confrontare questa scelta con quella di tutte le altre franchigie.

2.3.3 Player stats



Questa dashboard presenta due filtri: PlayerName e SeasonStartYear.

Tramite il filtro PlayerName è possibile selezionare un giocatore sul quale concentrare la nostra analisi. A seguito della selezione avremo quindi a disposizione numerosi indicatori utili a valutarne la performance durante la sua carriera NBA.

In alto troviamo informazioni generali quali l'età del giocatore durante le stagioni giocate nella lega, i team di appartenenza, il ruolo ricoperto, le partite giocate ed i punti fatti.

Scendendo si trovano due grafici a linee che rappresentano l'andamento del giocatore negli anni da un punto di vista salariale (grafico a sx) e in relazione ai minuti giocati (grafico a dx).

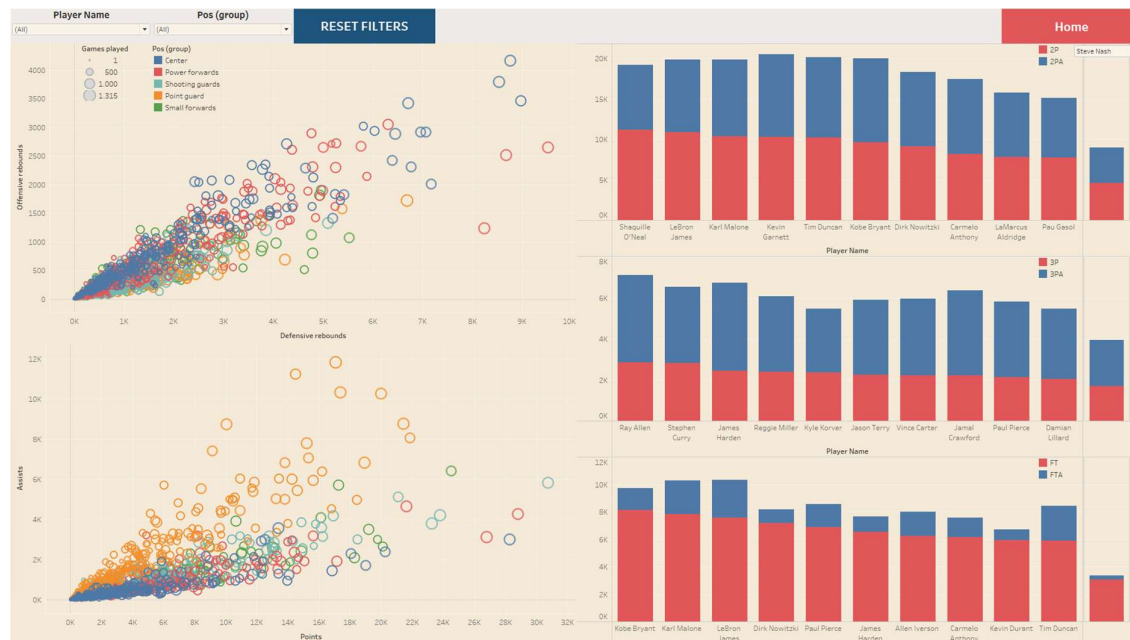
Troviamo poi un grafico a bolle che mostra l'andamento del giocatore per punti realizzati negli anni. La dimensione delle bolle è legata al numero di partite giocate in quella stagione. Quando per i giocatori in analisi si ha un numero sufficiente di stagioni giocate appariranno delle bolle rosse che rappresentano un forecast legato sia al numero di partite che avrebbero disputato che al numero di punti che avrebbero potuto segnare. A fianco si nota un grafico a barre che mostra le percentuali realizzative del giocatore per diverse tipologie di tiro e una tabella con il numero di tiri tentati da ciascuna posizione.

In basso a sinistra abbiamo realizzato un grafico a barre sovrapposte che rappresenta le prestazioni del giocatore a rimbalzo. In blu troviamo il numero totale di rimbalzi (TRB) presi dal giocatore, in rosso i rimbalzi offensivi (ORB).

In basso a destra troviamo due grafici a barre che rappresentano statistiche relative rispettivamente ad aspetti positivi (assist, stoppage e palle rubate) e ad aspetti negativi (falli commessi e palle perse) della performance di un giocatore.

In alto a dx troviamo il tasto "Home" e il tasto "RESET FILTERS" che riporta in automatico SeasonStartYear su "All".

2.3.4 Player comparison



Questa dashboard presenta tre filtri: due diversi filtri PlayerName e un filtro Pos(group).

Pos(group) è un gruppo creato a partire dalla variabile Pos. Consente di selezionare i giocatori in base a quello che è stato il loro ruolo principale in campo. Questo filtro agisce sull'intera dashboard aggiornando sia i grafici a dispersione che i grafici a barre.

Nei grafici a dispersione troviamo rappresentati i giocatori tramite cerchi la cui circonferenza dipende dal numero di partite giocate. Il colore dei cerchi corrisponde alla posizione del giocatore. Possiamo trovare più cerchi relativi a un singolo giocatore in virtù del fatto che può aver ricoperto molteplici ruoli durante la sua carriera. Nel grafico superiore analizziamo il contributo a rimbalzo. L'asse verticale rappresenta i rimbalzi offensivi mentre quello orizzontale quelli difensivi. Nel grafico inferiore analizziamo il contributo offensivo del giocatore. Sull'asse verticale troviamo gli assist realizzati mentre su quello orizzontale i punti. Utilizzando il filtro PlayerName in alto a sinistra i due grafici a dispersione vengono aggiornati consentendo di visualizzare la prestazione del singolo giocatore nei diversi ruoli ricoperti in carriera ma anche di confrontare più giocatori tra loro data la possibilità di effettuare una selezione multipla. Selezionando un ruolo tramite Pos(group) anche il filtro PlayerName si aggiornerà eliminando dalla lista i giocatori che non hanno mai ricoperto il ruolo selezionato. Tramite il tasto RESET FILTERS entrambi i filtri fin qui descritti vengono riportati sul valore "All".

La parte destra della dashboard è caratterizzata da grafici a barre sovrapposte ordinati in modo decrescente secondo il numero di tiri realizzati da diverse posizioni così da costituire delle top ten. In rosso troviamo il numero di tiri realizzati dalla posizione considerata mentre in blu il numero di tiri tentati da quella posizione, entrambi per l'intera carriera del giocatore. Sulla destra di queste top ten troviamo altri grafici a barre sovrapposte che rappresentano i dati relativi alle stesse tipologie di tiro e i cui assi sono tarati sulle stesse dimensioni. Questi tre grafici si aggiornano secondo il filtro in alto a destra (sotto il tasto Home) così da poter selezionare un giocatore in particolare e confrontare le sue performance con quelle dei giocatori nella top ten, tale filtro non è influenzato dal tasto RESET.

POWER BI

3.1 Descrizione del dataset

Il dataset utilizzato tratta dati commerciali reali che sono stati anonimizzati e resi pubblici da Olist, il più grande E-commerce del mercato brasiliano. In tutto il Brasile, Olist collega le aziende ai canali di vendita e stipulando un solo contratto permette ai commercianti di vendere i loro prodotti attraverso i propri partner logistici. Il dataset contiene informazioni su 100.000 ordini dal 2016 al 2018 effettuati in diversi mercati in Brasile. Tramite il dataset è possibile analizzare gli ordini da molteplici prospettive: stato dell'ordine, prezzo, prestazioni di pagamento, trasporto, posizione del cliente, attributi del prodotto e recensioni scritte dai clienti.

Questo primo dataset è stato integrato con un secondo dataset contenente informazioni di marketing sui venditori che hanno contattato Olist per vendere i loro prodotti sulla piattaforma. Le informazioni contenute riguardano la categoria d'impresa, il profilo comportamentale dell'azienda e altri aspetti di grande utilità per un'analisi di marketing, come il tempo trascorso tra il primo contatto con Olist e la sottoscrizione del contratto per la vendita dei prodotti.

Dei dati di partenza ne sono stati analizzati solo una parte. Di seguito la struttura delle tabelle:

Clienti

customer_id	Identificativo cliente
customer_unique_id	Identificativo cliente in relazione all'ordine
customer_zip_code_prefix	Codice Postale
customer_city	Città del cliente
Customer_state	Stato del cliente

InfoGeografiche

geolocation_city	Città
geolocation_lat	Latitudine della città
geolocation_lng	Longitudine della città
geolocation_state	Stato della città
geolocation_zip_code_prefix	Codice postale della città

Oggetti

Order_id	Codice identificativo ordine
Order_item_id	Identificativo dell'oggetto all'interno dell'ordine
Price	Prezzo del prodotto
Freight_value	Costo di spedizione relativo all'oggetto
Product_id	Identificativo prodotto
Seller_id	Identificativo venditore

Ordinamento giorno

NomeGiorno	Giorno della settimana
Ordinamento	Numero del giorno della settimana per creare l'ordinamento (es. lun-1, mar-2 ecc.)

Ordini_date_stato

Order_id	Identificativo ordine
Customer_id	Identificativo cliente
Order_status	Stato dell'ordine
Order_purchase_timestamp	Data e ora del pagamento
Order_delivered_carrier_date	Data e ora della conferma del partner logistico per la spedizione
Order_delivered_customer_date	Data e ora della consegna al cliente
Order_purchase_dayofweek_name	Giorno della settimana

Pagamenti

Order_id	Identificativo ordine
Payment_type	Tipo di pagamento
Payment_value	Valore del pagamento

Prodotti

Product_id	Identificativo prodotto
Product_category_name	Categoria prodotto in brasiliano
categoriaProdottoEND	Traduzione della categoria prodotto in inglese

Recensioni

Order_id	Identificativo ordine
Review_id	Identificativo recensione
Review_score	Recensione dell'ordine (Voto)

TraduzGeo

Sigla	Sigla dello stato
Intero	Nome completo dello stato

Traduzione_Categoria

CategoriaProdottoENG	Traduzione categoria prodotto
CategoriaProdottoESP	Categoria prodotto in spagnolo

Venditori

Seller_city	Città del venditore
Seller_id	Identificativo venditore
Seller_state	Stato del venditore
Seller_zip_code_prefix	Codice postale della città del venditore

Fonte: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>

Olist marketing qualified leads dataset

First_contact_date	Primo contatto tra venditore e Olist
Mql_id	Codice identificativo del venditore
origin	Tipo di ricerca che ha portato al contatto

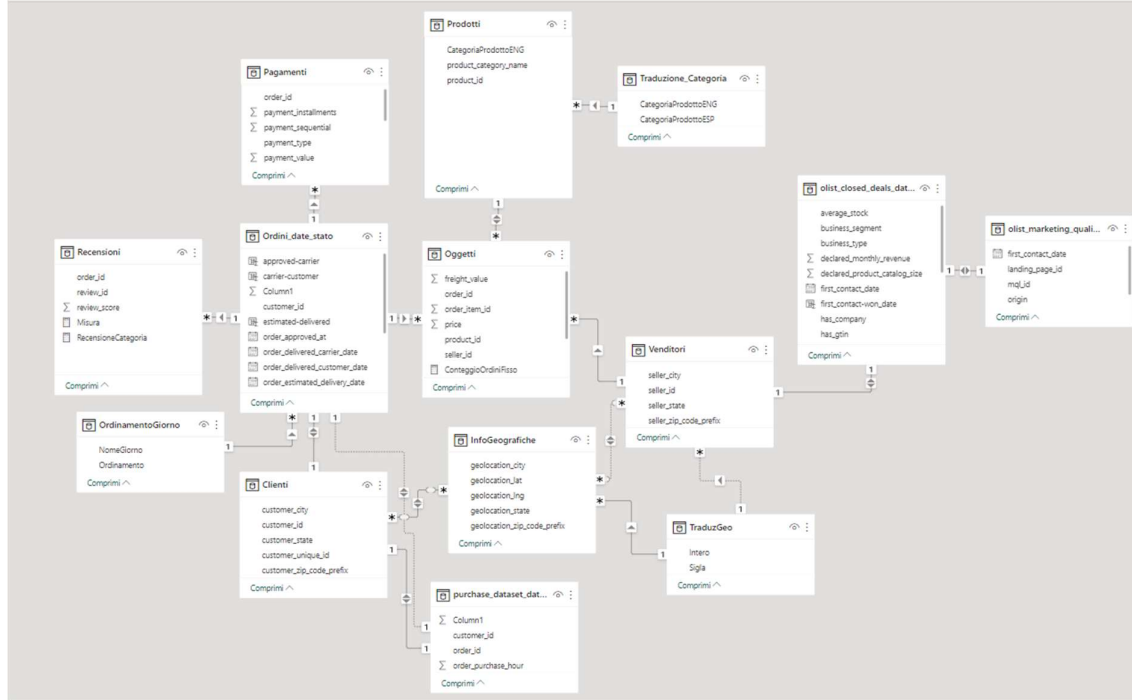
Olist closed deals dataset

Mql_id	Identificativo venditore
Seller_id	Identificativo venditore per join con l'altro dataset
Business_segment	Segmento dell'attività
Business_type	Tipo di attività (reseller-manufacturer)
Declared_monthly_revenue	Guadagni mensili dichiarati
Declared_product_catalog_size	Dimensioni catalogo prodotti
Lead_type	Tipologia d'impresa (online/offline, dimensione)
Lead_behaviour_profile	Profilo comportamentale dell'impresa

Fonte: <https://www.kaggle.com/datasets/olistbr/marketing-funnel-olist>

3.2 Collegamento Tabelle

Di seguito lo schema delle relazioni tra le tabelle del primo e del secondo dataset:



3.3 ANALISI

L'analisi avviene attraverso tre pagine:

- La prima analizza gli ordini del 2017 dal punto di vista del cliente e consente di ottenere una panoramica degli acquisti sulla base della specifica categoria di prodotto. È possibile apprezzare attraverso la dashboard quali sono le categorie di prodotto più popolari, come si compone il prezzo pagato dal cliente, in quale regione del Brasile sono stati acquistati più prodotti e la frequenza d'acquisto nei vari mesi, giorni e orari.
- Nella seconda è possibile analizzare il mercato nelle varie regioni e città brasiliane. Viene mostrato per stato e città l'andamento delle spese effettuate dai clienti ed una previsione dei successivi 12 mesi, quali sono i prodotti più acquistati nei singoli stati brasiliani e il tempo di consegna in relazione alle vendite. Sono informazioni utili per capire quali sono i mercati più proficui e l'attuale livello di soddisfazione dei clienti.
- La terza sposta l'attenzione dai clienti ai venditori e analizza un campione di 377 aziende che si dividono in manifatturiere e rivenditrici. Per ognuna di queste aziende viene registrata la data nella quale è avvenuto il primo contatto con Olist, la data nella quale la stessa azienda ha firmato il contratto per pubblicare il proprio catalogo di prodotti sulla piattaforma e una serie di informazioni sulle caratteristiche dell'azienda. Tramite queste è possibile analizzare questo sotto-campione che collabora con Olist.

3.3.1 Home



I tre pulsanti consentono di navigare all'interno della dashboard. Su tutte e tre le pagine è presente il tasto "Preview" che consente di tornare alla pagina iniziale.

3.3.2 Analisi delle categorie di prodotti

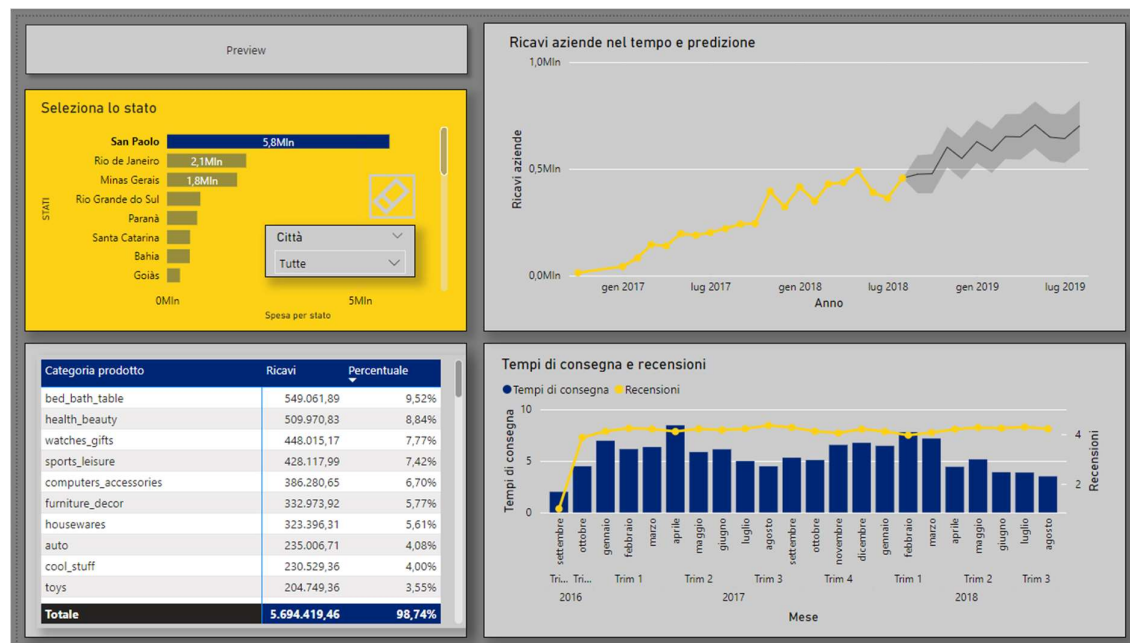


I filtri applicati alla pagina sono la categoria di prodotto e l'anno di riferimento. Per il filtro sulla categoria prodotto è stato utilizzato un grafico a barre in modo tale da avere informazioni anche sulla percentuale di ordini con riferimento alla singola categoria. Il filtro anno è stato invece inserito come menu a discesa con selezione singola.

A destra del filtro "Anno" è presente una mappa ad albero che mostra quali sono gli Stati del Brasile nei quali è stata effettuata la maggior spesa sulla piattaforma (più grande è il rettangolo e scura la tonalità di verde, maggiore è la spesa). Al di sopra della mappa ad albero si trova un grafico a ciambella che mostra come si compone il prezzo pagato dal cliente: in blu la percentuale del prezzo totale riferito al prodotto, in giallo la parte relativa alle spese di trasporto. Tale rappresentazione, che varia in base ai filtri selezionati, è affiancata da un benchmark. Il benchmark scelto è la percentuale media delle spese di trasporto considerando tutte le categorie (varia in base all'anno analizzato).

Sulla destra, si trovano tre grafici a barre, uno sotto l'altro, che rappresentano rispettivamente (dal basso verso l'alto) il numero di acquisti effettuati in base a orario, giorni della settimana e mesi. Questi grafici variano per tipologia di prodotto e anno, permettendo di analizzare in che momenti avvengono la maggior parte degli acquisti da parte dei clienti.

3.3.3 Analisi temporale per stato e per città



In questa pagina i filtri sono relativi alla dimensione geografica e si riferiscono in particolare a stato e città. Lo scopo della pagina è quello di analizzare le caratteristiche del mercato delle specifiche regioni brasiliane.

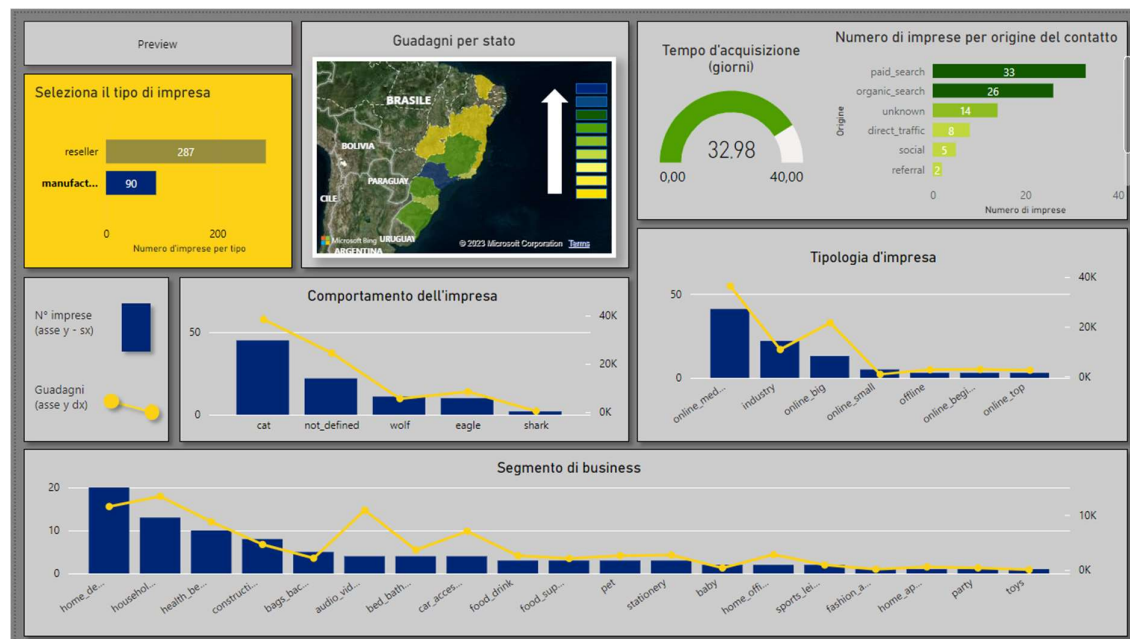
Il filtro principale si riferisce allo stato. Come nella pagina precedente è stato inserito sotto forma di grafico a barre. Esso rappresenta quanti guadagni fanno capo al singolo stato. Tramite menu a discesa si può affinare l'indagine a livello di città.

Sotto a tali filtri possiamo osservare una tabella che scompone la spesa dello stato tra le varie categorie di prodotto. Essa mostra i ricavi generati da ciascuna categoria, nonché quanto influisce in percentuale sulla spesa totale dell'area geografica selezionata.

A destra in alto troviamo un grafico a linee che rappresenta l'andamento dei ricavi nel tempo. Sono inoltre disponibili le previsioni per i 12 mesi successivi con i relativi intervalli di confidenza al 95%.

In basso a destra troviamo un grafico a barre e linee. Ogni barra rappresenta il tempo di consegna medio in giorni in uno specifico mese. La linea rappresenta invece la media mensile delle recensioni. Le due informazioni sono state rappresentate insieme così da permettere di vedere se le due misure sono in qualche modo collegate.

3.3.4 Analisi di un campione di venditori



Questa pagina si focalizza su 377 imprese, una piccola parte di quelle che hanno stipulato un contratto con Olist, sulle quali è stato possibile raccogliere maggiori informazioni. Tramite il filtro in alto a sinistra è possibile analizzare separatamente le imprese manifatturiere e i venditori. Anche in questo caso il filtro è rappresentato sotto forma di grafico a barre così da mostrare il numero di imprese appartenente alle 2 categorie.

Per ciascuna delle due categorie possiamo analizzare i guadagni per stato attraverso la mappa colorata. Il giallo rappresenta scarsi guadagni mentre il blu colora gli stati nei quali le imprese hanno guadagnato di più.

A destra della mappa abbiamo un misuratore che mostra il tempo trascorso tra il primo contatto con Olist e la stipula del contratto. A fianco troviamo un grafico a barre che mostra il numero di imprese che sono venute a contatto con Olist attraverso i diversi canali.

Al di sotto troviamo 3 grafici a barre e linee. Tutti e tre i grafici mostrano il numero di imprese (colonne) e i guadagni (linea). Ciò che li contraddistingue sono le caratteristiche delle imprese utilizzate per raggrupparle al fine di calcolare i valori mostrati: il comportamento dell'impresa (rilevato dal reparto di marketing di Olist nel momento della sottoscrizione del contratto), la tipologia d'impresa e il segmento di mercato nel quale l'impresa svolge la propria attività.