



Università Politecnica delle Marche

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

Laboratorio Big Data Analytics

Analisi dati mediante l'utilizzo dei tools Microsoft Power BI, Qlik e Tableau

Docente

Prof. Virgili Luca

Studenti

Galassi Leonardo

Parisi Sheril

Sartini Alberto

Anno Accademico 2023/2024

Indice

Microsoft Power BI

1.1 Dataset	3
1.2 Trasformazione dei dati	5
1.3 Obiettivo dell'analisi	6
1.4 Visualizzazioni	6
1.4.1 Storia della UEFA Champions League	7
1.4.2 UCL 1955-2023	8
1.4.3 UCL 2021-2022	10
1.4.4 Classificazione delle Squadre 2021-22	12
1.5 Visualizzazione mobile	14

Qlik

2.1 Dataset	15
2.2 Trasformazione dei dati	16
2.3 Obiettivo dell'analisi	16
2.4 Visualizzazioni	16
2.4.1 The Richest People in the World	17
2.4.2 Distribution of Wealth	18
2.4.3 Evolution of Wealth	20
2.4.4 Number of Billionaires and Economic Indicators	21

Tableau

3.1 Dataset	23
3.2 Trasformazione dei dati	25
3.3 Obiettivo dell'analisi	25
3.4 Visualizzazioni	25
3.4.1 Customer Overview	26
3.4.2 Campaign Overview (pt. 1)	28
3.4.3 Campaign Overview (pt. 2)	30

CAPITOLO 1

Microsoft Power BI**1.1 Dataset**

Per il progetto di Power BI sono stati utilizzati tre dataset riguardanti la UEFA Champions League, competizione calcistica europea tra squadre di alto livello.

Il primo¹ fornisce informazioni riguardanti l'albo d'oro della competizione dal 1955 al 2023. In particolare, la tabella importata dal web presenta le seguenti variabili:

- Stagione
- Data della finale
- Nazionalità della squadra vincitrice
- Vincitore
- Finalista perdente
- Risultato
- Sede della finale
- Spettatori

Il secondo² è costituito da diverse tabelle che riguardano delle statistiche complessive riferite a tutta la storia della Champions (1955-2023):

<i>Tabella</i>	<i>Variabili</i>
All time ranking by club	Squadra, paese della squadra, numero di volte che la squadra ha partecipato alla competizione, titoli vinti, partite giocate, vittorie, pareggi, sconfitte, goal segnati, goal subiti, punteggio
All time ranking by country	Paese, numero di volte che il paese ha partecipato alla competizione, titoli vinti, partite giocate, vittorie, pareggi, sconfitte, goal segnati, goal subiti, punteggio

¹ https://it.wikipedia.org/wiki/Albo_d%27oro_della_UEFA_Champions_League

² <https://www.kaggle.com/datasets/basharalkuwaiti/champions-league-era-stats>

Coaches appear details	Allenatore, squadre che ha allenato, numero di presenze
Coaches appear totals	Allenatore, numero di presenze totali
Player appear details	Giocatore, squadre in cui ha giocato, numero di presenze
Player appear totals	Giocatore, numero di presenze totali
Player goal details	Giocatore, squadra, goal
Player goal totals	Giocatore, goal totali segnati

Il terzo³ si concentra sulla stagione 2021/22 della Champions League e contiene le seguenti tabelle:

<i>Tabella</i>	<i>Variabili</i>
Attacking	Giocatore, squadra, ruolo, assists, calci d'angolo battuti, fuorigioco, dribblings, partite giocate
Attempts	Giocatore, squadra, ruolo, numero di tiri tentati in porta, tiri riusciti, tiri non riusciti, partite giocate
Defending	Giocatore, squadra, ruolo, palle recuperate, contrasti, contrasti vinti, contrasti persi, tentativi di disimpegno, partite giocate
Disciplinary	Giocatore, squadra, ruolo, falli commessi, falli subiti, cartellini gialli, cartellini rossi, minuti giocati, partite giocate
Distribution	Giocatore, squadra, ruolo, passaggi tentati, passaggi completati, cross tentati, cross completati, calci di punizione, partite giocate. <u>Misure calcolate: % accuratezza passaggi (completati/tentati), % accuratezza cross (completati/tentati)</u>
Goalkeeping	Portiere, squadra, parate, goal subiti, cleansheets, partite giocate

³ <https://www.kaggle.com/datasets/azminetoushikwasi/ucl-202122-uefa-champions-league>

Goals	Giocatore, squadra, ruolo, numero di goal segnati, goal con piede destro, goal con piede sinistro, goal di testa, goal da dentro l'area, goal fuori area, partite giocate
Key Stats	Giocatore, squadra, ruolo, minuti giocati, partite giocate, numero di goal segnati, assists, distanza corsa (km)

Oltre alle suddette tabelle sono stati utilizzati dati provenienti da altre fonti in merito a:

- Foto giocatori
- Logo squadre

1.2 Trasformazione dei dati

Primo dataset

Essendo una tabella importata dal web si è proceduto alla pulizia di quest'ultima con varie operazioni:

- Rimozione valori nulli
- Rimozione variabili non utili ai fini dell'analisi
- Divisione della colonna stagione (es. 1955-56) in due colonne che presentassero la singola data e trasformazione di quest'ultima in formato data
- Divisione della colonna sede della finale (es. Parigi-Parco dei Principi) in due variabili che fornissero separatamente l'informazione sulla sede e sullo stadio
- Sostituzione di diversi valori per risolvere eventuali incongruenze (es. sostituito Real Madrid (2) con Real Madrid)

Secondo dataset

- Join tra tabelle
- Rimozione valori nulli e duplicati
- Rimozione variabili non utili ai fini dell'analisi
- Risoluzione incongruenze (es. sostituito Manchester Utd con Manchester United)

Terzo dataset

- Join tra tabelle
- Rimozione valori nulli, duplicati ed errori
- Rimozione variabili non utili ai fini dell'analisi
- Sostituzione di diversi valori da inglese a italiano (es. sostituito goalkeeper con portiere)
- Risoluzione incongruenze

1.3 Obiettivo dell'analisi

L'obiettivo della seguente analisi è quello di fornire a un utente appassionato di calcio, o semplicemente curioso, una panoramica generale della storia della UEFA Champions League dal 1955 al 2023. Inoltre, l'utente può esplorare, in modo interattivo, statistiche maggiormente dettagliate sui giocatori e sulle squadre che hanno partecipato alla stagione del 2021-2022.

1.4 Visualizzazioni

A tal proposito, sono state realizzate quattro dashboard:

- Storia della UEFA Champions League
- UCL 1955-2023
- UCL 2021-2022
- Classificazione delle Squadre 2021-22

1.4.1 Storia della UEFA Champions League



La prima è una dashboard introduttiva che permette, attraverso tre pulsanti, di spostarsi alle dashboard successive. Consente di osservare diverse informazioni riguardanti le finali di Champions League disputate dal 1955 al 2023.

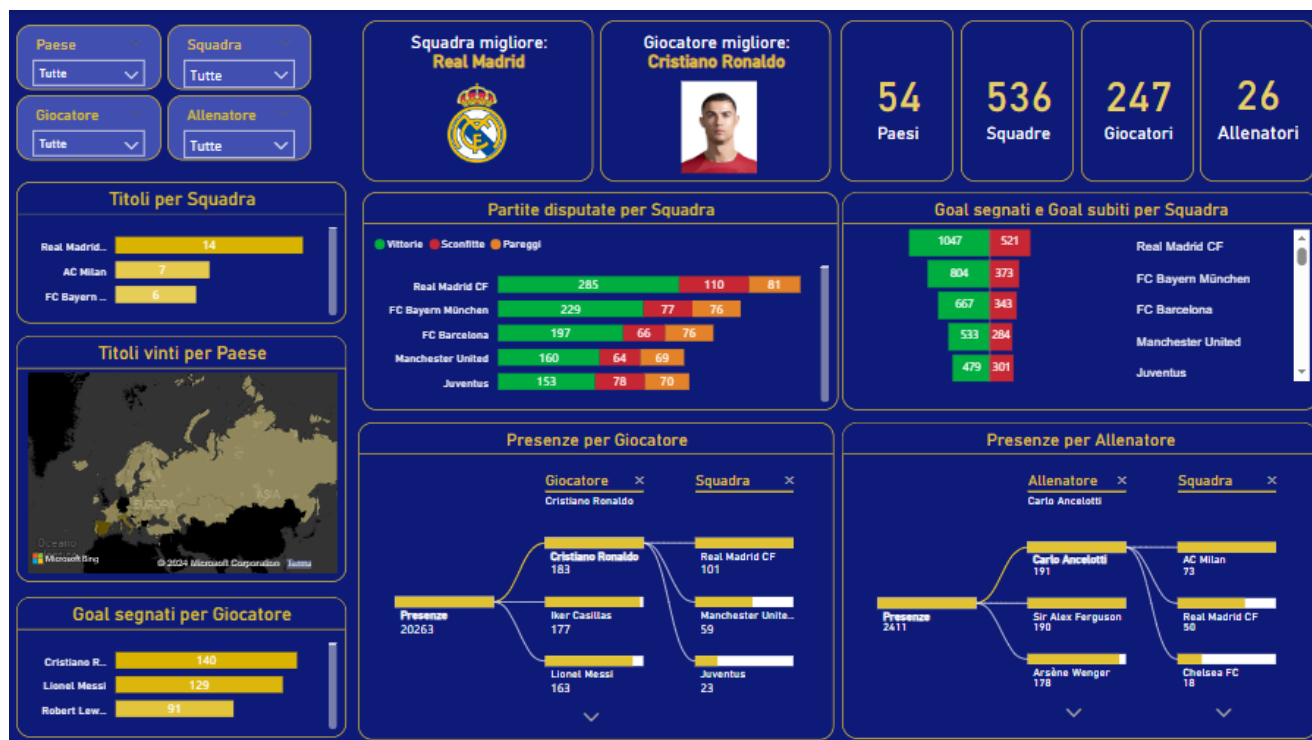
Attraverso la selezione dell'anno, tramite l'apposito filtro situato in alto a sinistra, è possibile ottenere una visualizzazione dinamica che permette di vedere, per ogni stagione, quali sono state le due squadre che si sono scontrate durante la finale e qual è stato il risultato della partita.

Filtrando la dashboard, anche la sede della finale, lo stadio, il numero di spettatori e la mappa che rappresenta il luogo in cui è stata giocata la partita, cambiano di conseguenza. Una curiosità è quella in riferimento all'anno 2020 in cui, a causa del Covid, la partita è stata giocata a porte chiuse e dunque non c'è stato nessuno spettatore.

Infine, in basso a sinistra è stato inserito un "Animated Bar Chart" che permette di visualizzare, senza dover filtrare, la squadra vincitrice di ogni stagione grazie all'anno che cambia automaticamente.

Scorrendo i vari anni, si nota già da questa prima visualizzazione che la squadra più vincente della storia della Champions League è il Real Madrid.

1.4.2 UCL 1955-2023



La seconda dashboard offre un'analisi approfondita della storia della competizione, risultando particolarmente utile per gli utenti interessati a maggiori dettagli su squadre e giocatori più performanti, ma anche sugli allenatori.

Grazie ai quattro filtri (paese, squadra, giocatore, allenatore), è possibile interagire dinamicamente con la dashboard.

A colpo d'occhio si può vedere che la squadra e il giocatore migliori statisticamente della storia della Champions League sono rispettivamente il Real Madrid e Cristiano Ronaldo.

In alto a destra sono presenti quattro KPI che, in modo riassuntivo e secondo i dati utilizzati, mostrano il numero di paesi, squadre, giocatori e allenatori presenti nella storia della competizione.

Squadre e Paesi

Il grafico "Titoli per Squadra" mostra le squadre che hanno vinto più titoli. Si nota come il Real Madrid abbia vinto 14 volte la competizione.

Passando con il mouse su ciascuna barra, è possibile anche vedere il numero di volte che la squadra ha partecipato alla competizione.



La mappa "Titoli vinti per Paese" utilizza diverse sfumature di colore per indicare i paesi le cui squadre hanno vinto più titoli. La Spagna risulta il paese con il maggior numero di titoli (19), seguita dall'Italia con 12 titoli.

Il grafico a barre in pila "Partite Disputate per Squadra" mostra il totale delle partite giocate in Champions da ciascuna squadra, differenziando per colore l'esito (vittorie, sconfitte e pareggi). Anche in questo caso, il Real Madrid è in prima posizione.

Per completare l'analisi sulle squadre si può guardare al grafico "Tornado" che mostra i goal segnati e subiti per squadra.

Giocatori

Il grafico in basso a sinistra mostra i capocannonieri: Ronaldo e Messi si confermano i migliori dal punto di vista dei goal effettuati.

A destra del suddetto grafico si trova un albero di scomposizione che permette di analizzare il numero di presenze dei giocatori, ordinato in modo crescente, e la distribuzione di queste presenze nelle diverse squadre. Ad esempio, Cristiano Ronaldo si conferma il primo anche sotto questo aspetto, con un totale di 183 presenze, così suddivise: 101 al Real Madrid, 59 al Manchester United e 23 alla Juventus.

Allenatori

Infine, in basso a destra, è presente un altro albero di scomposizione che permette di focalizzarsi sugli allenatori. Carlo Ancelotti è l'allenatore con il maggior numero di presenze, con la maggior parte di esse accumulate al Milan, seguito da altre squadre elencate in ordine crescente.

1.4.3 UCL 2021-2022



La terza dashboard si concentra sulla competizione calcistica nella stagione 2021/22, presentando statistiche più dettagliate con particolare riferimento ai giocatori.

I due filtri (giocatore e squadra) permettono di osservare il tutto in modo dinamico.

In alto a destra ci sono due KPI che mostrano il numero di giocatori e il numero di partite giocate in questa stagione. Sotto sono mostrate anche due misure (accuratezza dei passaggi e dei cross), calcolate a partire dai dati disponibili (passaggi completati/tentati, cross completati/tentati).

Inoltre, sono state studiate delle **metriche disciplinari**; nel grafico ad anello compare il numero di cartellini gialli e rossi e, nel grafico a barre, i falli commessi da ciascun giocatore.

Anche in questo caso, come nella dashboard precedente, è possibile capire chi sono i giocatori migliori per goal segnati. Il colore delle barre è differente in base al ruolo e, passando con il mouse sulla barra, si può vedere anche la squadra di appartenenza del giocatore. Come atteso, sono soprattutto gli attaccanti, ma anche i centrocampisti, a segnare il maggior numero di goal. Il migliore è Benzema, mentre **Cristiano Ronaldo, che in media era primo, in questa stagione risulta essere il decimo per numero di goal.**

Si può approfondire l'analisi consultando le schede che indicano con quale parte del corpo è stato segnato il goal e se è stato battuto da dentro o da fuori dall'area. Il grafico "synoptic panel" consente, passando il mouse sulle varie parti del corpo, di visualizzare il numero totale di goal realizzati di testa, con il piede destro e con il piede sinistro.

Al centro sono presenti due grafici a barre impilate che mostrano i giocatori più offensivi e più difensivi in base a diverse metriche. I primi sono valutati in base al numero di dribblings, movimento

tecnico che permette di superare un avversario, e al numero di assists, passaggio chiave che porta a un goal; queste ultime sono azioni tipiche degli attaccanti il cui compito è quello di creare e finalizzare azioni di attacco. Per i secondi, invece, si fa riferimento a due metriche maggiormente associate al ruolo di difesa ovvero quello di proteggere la propria porta e prevenire i goal avversari. In particolare, il numero di contrasti vinti è indicatore della capacità di un giocatore di vincere il confronto fisico con l'avversario per il possesso della palla, mentre il numero di disimpegni tentati **è un'azione che permette di allontanare la palla dalla propria area di rigore**. In formato tooltip, oltre alla squadra di appartenenza, è stato inserito il ruolo del giocatore. Come atteso, le due top 5 sono costituite rispettivamente da attaccanti e da difensori.

Il grafico a dispersione in basso a sinistra ha lo scopo di capire la resistenza di ciascun giocatore, in base ai minuti giocati in una partita e alla distanza che riesce a percorrere. Vinícius Júnior, attaccante del Real Madrid, emerge come il migliore della stagione 2021/22, avendo giocato 1.199 minuti e percorso 133 km. Inoltre, è da notare che i portieri, identificati dai pallini arancioni, si collocano più in basso nel grafico, il che è conforme alle aspettative, poiché di norma coprono distanze minori rispetto ai giocatori di campo.

Infine, in basso a destra, si può esaminare la performance dei portieri. **La tabella è ordinata in modo crescente in base al numero di cleansheets, cioè il numero di partite in cui un portiere riesce a mantenere la propria porta inviolata**. Si possono vedere anche il numero di partite giocate e, attraverso delle barre, le parate e i goal subiti (il numero è mostrato in formato tooltip).

1.4.4 Classificazione delle Squadre 2021-22



La quarta e ultima dashboard rappresenta un'analisi maggiormente approfondita delle squadre partecipanti alla Champions League 2021/2022. Le quattro rappresentazioni mirano a riprodurre il medesimo studio svolto per i giocatori nella dashboard 1.4.3 ma riferendosi alle squadre, cercando di comprendere quali sono più forti nell'attacco e quali nella difesa. Si è deciso dunque di raggrupparle con l'algoritmo di clustering K-Means, attraverso l'integrazione Python di Power BI. Ogni grafico mette in relazione due differenti metriche; i due in alto rappresentano caratteristiche più offensive di una squadra mentre i due in basso quelle più difensive.

Assist vs Corner Taken

Il grafico in alto a sinistra mostra la relazione tra il numero di assist e i calci d'angolo battuti dalle squadre. È evidente che alcune squadre eccellono sia negli assist che nei corner, suggerendo una forte capacità offensiva sia nei passaggi che nelle situazioni da palla ferma.

Assist vs Dribbles

Nel grafico in alto a destra viene esaminata la relazione tra assists e dribblings. Qui si nota che alcune squadre si distinguono per un alto numero di assists, mentre altre puntano maggiormente sui dribblings. Questo indica che ci sono diverse strategie offensive: alcune squadre preferiscono il gioco di squadra e i passaggi decisivi, mentre altre si affidano alle abilità individuali nei dribblings per creare occasioni da goal.

Balls Recovered vs Tackles Won

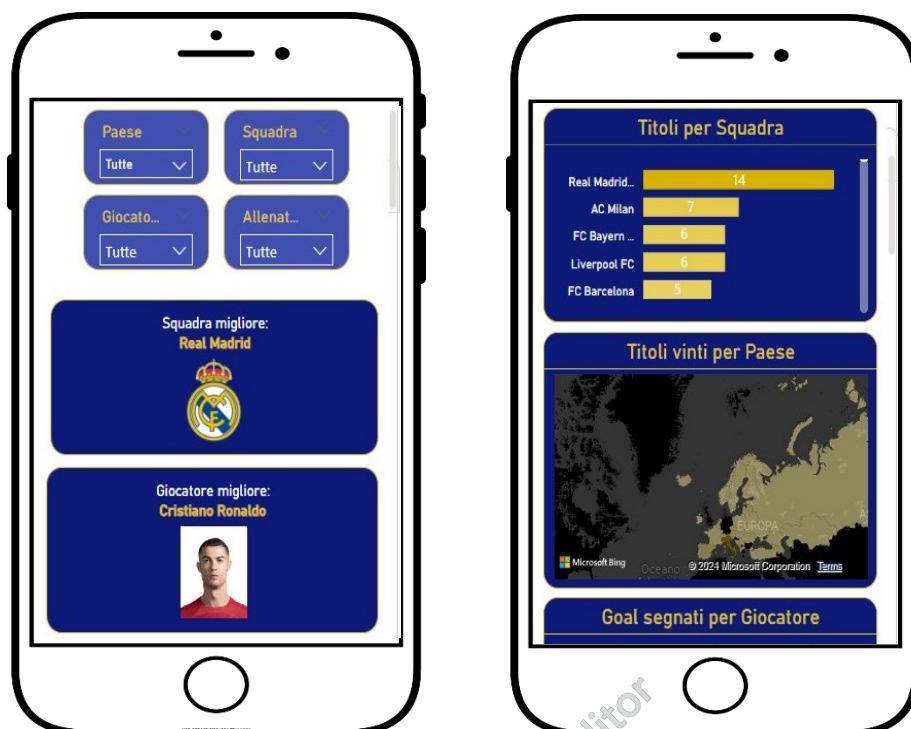
Il grafico in basso a sinistra illustra il numero di palloni recuperati rispetto ai contrasti vinti. Squadre con un alto numero di entrambe le metriche mostrano una forte capacità difensiva, suggerendo che sono abili nel recuperare il possesso e interrompere le azioni avversarie. Altre squadre mostrano valori più bassi, indicando aree di potenziale miglioramento nella fase difensiva.

Balls Recovered vs Clearances Attempted

Infine, l'ultimo grafico mette in relazione i palloni recuperati con i tentativi di spazzata. Squadre che eccellono in entrambe le metriche dimostrano non solo una buona capacità di recupero palla, ma anche una propensione a liberare la propria area di rigore in modo efficace. Ciò suggerisce una strategia difensiva solida, in grado di gestire la pressione avversaria.

In definitiva, il K-means ha prodotto tre cluster: i punti in rosso indicano le squadre che eccellono in una specifica metrica (es. Manchester United), quelli in giallo identificano i team che mostrano buone prestazioni in più metriche (es. Real Madrid), suggerendo un equilibrio tra abilità offensive e difensive e quelli in verde (es. Sporting CP) rappresentano squadre con prestazioni inferiori e che dunque potrebbero necessitare di miglioramenti in diversi aspetti del gioco.

1.5 Visualizzazione mobile



Le visualizzazioni desktop sono state riadattate per favorire la visualizzazione delle dashboard tramite cellulare, consentendo così agli utenti di consultare, seppur in modo limitante, le diverse informazioni più velocemente e in qualsiasi luogo essi si trovino.

CAPITOLO 2

Qlik

2.1 Dataset

Per l'analisi su Qlik sono stati utilizzati due dataset riguardanti i miliardari nel mondo secondo la rivista economica americana Forbes.

Il primo⁴ è caratterizzato da diversi file csv, ognuno dei quali rappresenta la lista redatta annualmente da Forbes degli uomini più ricchi del mondo dal 1997 al 2023. Per lo studio sono state prese in considerazione nove tabelle, unite a formare un unico dataset (dal 2015 al 2023) che presenta le seguenti variabili:

- **Year:** anno in cui è stata pubblicata la lista dei miliardari
- **Rank:** posizione in classifica del miliardario in termini di ricchezza
- **Net worth:** patrimonio netto finale del miliardario (espresso in dollari USA)
- **Full name:** nome completo del miliardario
- **Last name:** cognome
- **First name:** nome
- **Age:** età
- **Birth date:** data di nascita
- **Gender:** indica se il miliardario è uomo, donna o una famiglia ("family")
- **Country of citizenship:** paese di cittadinanza
- **Country of residence:** paese di residenza
- **City of residence:** città di residenza
- **Business category:** categoria o settore in cui opera l'azienda del miliardario
- **Organization name:** nome dell'organizzazione o della società associata al miliardario
- **Self made:** indica se il miliardario si è fatto da sé ("self-made") o ha ereditato la sua ricchezza ("not self-made")
- **Position in organization:** posizione ricoperta nell'organizzazione (es. CEO, cofondatore)

Il secondo⁵ fornisce un'analisi specifica per l'anno 2023 e include, oltre alle variabili appena menzionate, anche indicatori specifici per ciascun paese di residenza:

- **CPI country:** indice dei prezzi al consumo (CPI)

⁴ <https://www.kaggle.com/datasets/guillemservera/forbes-billionaires-1997-2023>

⁵ <https://www.kaggle.com/datasets/nelgiryewithana/billionaires-statistics-dataset>

- **CPI change country:** variazione del CPI
- **GDP country:** prodotto interno lordo (PIL)
- **Gross tertiary education enrollment:** iscrizione all'istruzione terziaria
- **Gross primary education enrollment country:** iscrizione all'istruzione primaria
- **Life expectancy country:** aspettativa di vita
- **Tax revenue country:** entrate fiscali
- **Total tax rate country:** aliquota fiscale totale
- **Population country:** popolazione

2.2 Trasformazione dei dati

- Rimozione valori nulli e duplicati
- Rimozione variabili non utili ai fini dell'analisi
- Modifica del tipo di dato
- Risoluzione incongruenze varie (es. sostituito Amazon.com con Amazon)

2.3 Obiettivo dell'analisi

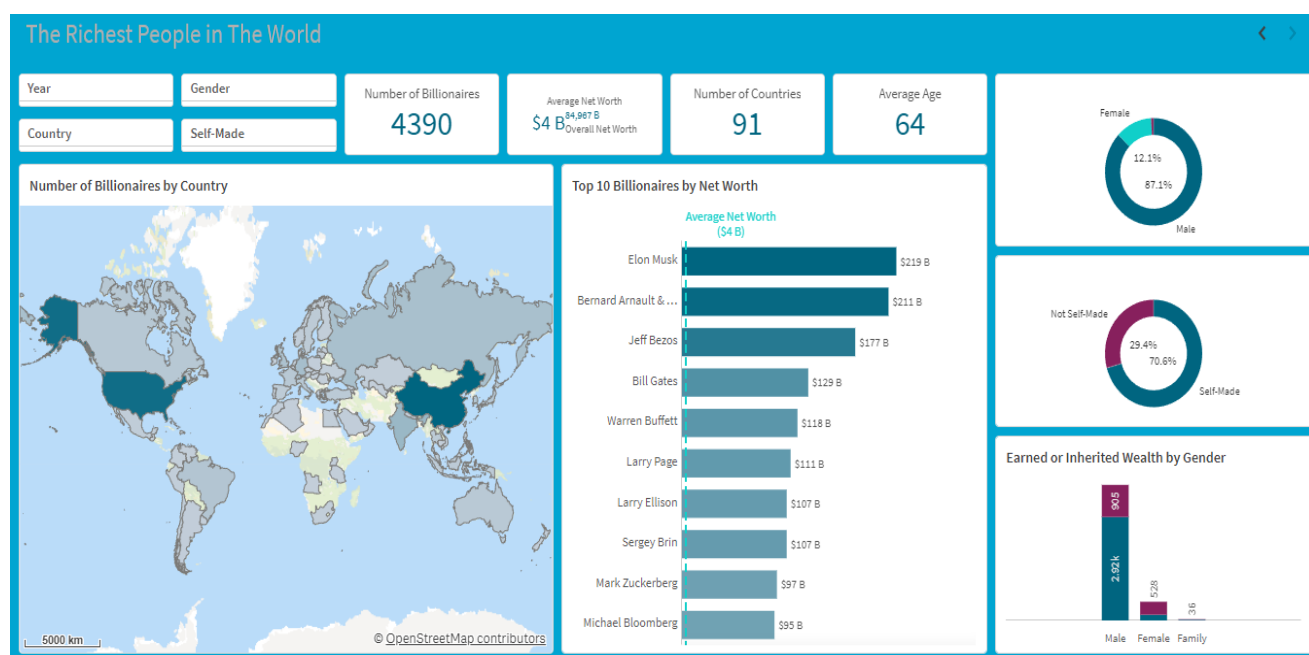
L'obiettivo della seguente analisi è quello di fornire una panoramica sui miliardari nel mondo e sull'evoluzione della loro ricchezza nel tempo. Lo studio mira a comprendere come si distribuisce la ricchezza in relazione a fattori come età, sesso e nazionalità, oltre a identificare le categorie di business e i paesi maggiormente profittevoli. Infine, si cerca di stabilire una relazione tra il numero di miliardari in un paese e diversi indicatori economici.

2.4 Visualizzazioni

A tal proposito, sono state realizzate quattro dashboard:

- The Richest People in The World
- Distribution of Wealth
- Evolution of Wealth
- Number of Billionaires and Economic Indicators

2.4.1 The Richest People in The World



La prima dashboard offre una panoramica interattiva sui miliardari e sulle loro caratteristiche, consentendo all'utente di selezionare quattro filtri:

- Year: permette di analizzare i miliardari nel corso degli anni
- Country: modifica i grafici a seconda del paese selezionato
- Gender: consente di concentrarsi su un genere specifico
- Self-Made: mostra la differenza tra coloro che hanno guadagnato autonomamente la propria ricchezza e quelli che l'hanno ereditata

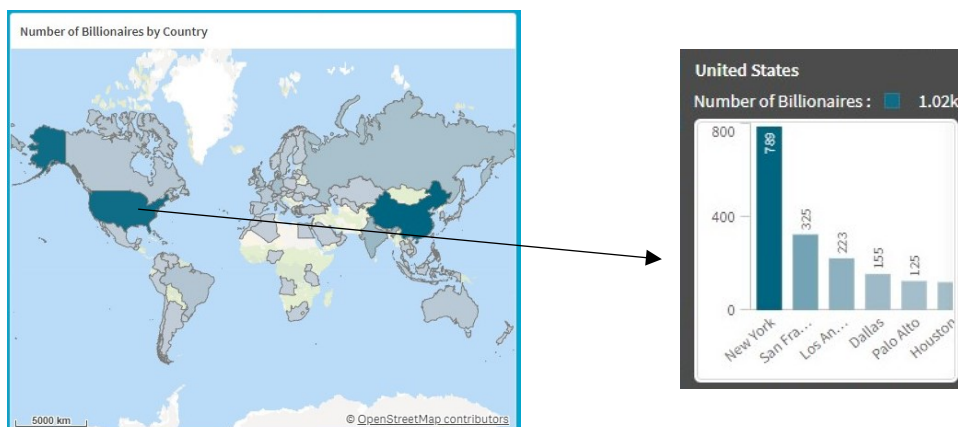
Sono presenti quattro KPI che forniscono una visione d'insieme: il numero totale di miliardari dal 2015 al 2023 è 4.390, la media del patrimonio netto è 4 miliardi di dollari e la ricchezza complessiva ammonta a 85 mila miliardi. Inoltre, i miliardari si collocano in 91 paesi diversi e hanno un'età media di 64 anni.

Grazie al primo grafico ad anello si vede come la maggioranza dei soggetti siano uomini (87.1%) mentre il 12.1% donne; lo 0.8% rientra nella categoria "family" (es. Bernard Arnault & Family).

Il secondo grafico ad anello distingue i miliardari in self-made (70.6%) e in coloro che hanno ereditato la propria ricchezza (29.4%).

Il grafico a barre impilate mostra come la maggioranza degli uomini si è fatta da sé (in blu), mentre la maggior parte delle donne ha ereditato la propria ricchezza (in rosso).

La mappa geografica utilizza diverse sfumature di colore e le tonalità più scure rappresentano i paesi in cui il numero di miliardari è più elevato. Si distinguono, in ordine, la Cina e gli Stati Uniti. Inoltre, passando con il mouse sulla mappa, è possibile vedere, grazie ad un grafico a barre colorato per numero di miliardari, la distribuzione di questi ultimi per città.



Infine, il grafico centrale mostra la classifica dei 10 miliardari più ricchi del mondo. Senza applicare il filtro per anno, si nota che Elon Musk è il primo in classifica con un patrimonio netto di 219 miliardi di dollari, nettamente superiore alla media della ricchezza complessiva pari a 4 miliardi (rappresentata da una linea di riferimento).

2.4.2 Distribution of Wealth



Questa dashboard mostra come si distribuisce la ricchezza in base a diverse variabili.

In alto a destra è possibile selezionare quattro filtri mentre di fianco è presente una tabella che contiene varie informazioni relative ai miliardari, mostrati in ordine crescente per patrimonio netto.

Senza applicare il filtro per anno vengono mostrati tutti i primi classificati nei diversi anni ma, selezionando, ad esempio, l'anno 2021 si può vedere che Jeff Bezos, fondatore del gruppo Amazon, occupa la prima posizione.

Country	Rank	Name	Net Worth	Age	Country	Organization Name	Business Category	Position
Year	1	Jeff Bezos	\$177 B	57	United States	Amazon	Technology	Entrepreneur
Gender	2	Elon Musk	\$151 B	49	United States	Tesla	Automotive	CEO and Chairman
Organization	3	Bernard Arnault & family	\$150 B	72	France	LVMH Moët Hennessy Louis Vuitton	Fashion & Retail	Chairman and CEO
	4	Bill Gates	\$124 B	65	United States	Bill & Melinda Gates Foundation	Technology	Cofounder
	5	Mark Zuckerberg	\$97 B	36	United States	Facebook	Technology	Entrepreneur

I grafici in basso hanno lo scopo di far capire come complessivamente si distribuisce la ricchezza:

- *Wealth by Business Category*: mappa ad albero che consente di vedere quale categoria di business è maggiormente profittevole. Nel 2021, ad esempio, prevale il settore della tecnologia con un valore complessivo del patrimonio netto pari a 2.455 miliardi di dollari.
- *Wealth by Group Age*: la ricchezza in questo caso è stata suddivisa in fasce d'età. Gli individui più ricchi hanno un'età compresa tra i 51 e i 70 anni.
- *Wealth by Gender*: gli uomini, complessivamente, possiedono un patrimonio netto più alto rispetto a quello delle donne.
- *Wealth by Self-Made or Not*: il patrimonio di coloro che non hanno ereditato la propria ricchezza è maggiore.
- *Wealth by Country*: nonostante il numero di miliardari sia più elevato in Cina, come detto precedentemente, la ricchezza complessiva dei soggetti è maggiore negli Stati Uniti. Al terzo posto si colloca l'India con, nel 2021, 577 miliardi di dollari.
- *Wealth by Organization*: l'azienda che negli anni ha contribuito ad un maggior patrimonio dei soggetti interessati è Amazon seguita dalla Bill & Melinda Gates Foundation.

2.4.3 Evolution of Wealth



La tabella mostra la classifica, ordinata per patrimonio netto, dei più ricchi del 2023. La terza colonna presenta un indicatore utile a comprendere se il patrimonio del miliardario ha subito un aumento (freccia verde verso l'alto) o un decremento (freccia rossa verso il basso) rispetto all'anno precedente (2022). Ad esempio, Bernard Arnault nel 2023 registra un patrimonio netto di 211 miliardi di dollari, superiore rispetto ai 158 miliardi del 2022. Infine, la colonna "Wealth History" permette di non tralasciare le informazioni relative agli altri anni, riepilogando graficamente l'evoluzione della ricchezza per ciascun individuo; ogni pallino rappresenta un anno specifico, dal 2015 al 2023.

A destra sono presenti tre grafici a linee. Il primo mostra il cambiamento della media del patrimonio complessivo dei miliardari negli anni, che sembra aver subito un leggero ma progressivo aumento fino ad arrivare al 2023 con 5 miliardi di dollari. Nel secondo si distinguono due linee, quella blu rappresenta gli uomini e quella rossa le donne; in media, il patrimonio di queste ultime risulta essere inferiore. L'ultimo rappresenta la differenza tra il patrimonio di coloro che hanno guadagnato la propria ricchezza (in blu) e quelli che l'hanno ereditata (in rosso), i quali sembra che mediamente possiedano una ricchezza maggiore.

2.4.4 Number of Billionaires and Economic Indicators



L'ultima dashboard si focalizza sull'anno 2023 e mira a comprendere, mediante dei grafici a dispersione, se è possibile individuare una relazione tra il numero di miliardari in un paese e vari indicatori economici per quel paese.

- *GDP*: il PIL è un indicatore utile per valutare le prestazioni economiche di un paese ed è comunemente utilizzato come misura generale del tenore di vita e del benessere della popolazione. Dal grafico si nota che, in linea con quanto atteso, gli Stati Uniti e la Cina, che si collocano rispettivamente al primo e secondo posto per numero di miliardari, hanno un PIL più alto (21 e 19 bilioni di dollari). Infatti, un PIL elevato indica un'economia più sviluppata e dunque una gamma più ampia di opportunità di business e investimento per i miliardari. Inoltre, un alto PIL corrisponde spesso a un elevato livello di reddito pro capite e dunque a un mercato di consumatori più redditizio da cui i miliardari possono trarre vantaggio. Tuttavia, questa relazione non è sempre rispettata poiché l'India, che si colloca al terzo posto per numero di miliardari, presenta un PIL relativamente basso, ad indicare un notevole divario economico tra i ricchi e il resto della popolazione.
- *Gross Tertiary Education Enrollment*: è una misura utilizzata per valutare la partecipazione all'istruzione post-secondaria. Un alto tasso di iscrizione indica che un'ampia percentuale della popolazione in età universitaria è iscritta a programmi di istruzione terziaria. Si è ritenuto potesse essere interessante valutare la sua relazione con il numero di miliardari in un paese poiché è probabile che una popolazione maggiormente istruita tenda ad assumere competenze avanzate, incoraggiando le persone a intraprendere, per esempio, carriere imprenditoriali. Questo sembra essere confermato dagli Stati Uniti, che presentano un tasso di iscrizione dell'88%, mentre paesi come la Grecia e l'Australia, in cui l'indicatore è più alto rispetto agli

Stati Uniti, hanno un numero di miliardari inferiore. Ciò può essere dovuto a vari fattori come un ambiente imprenditoriale ostile.

- *CPI*: il consumer price index, è un indicatore che misura la variazione nel tempo del livello medio dei prezzi pagati dai consumatori per un insieme specifico di beni e servizi di consumo. Potrebbe essere collegato al numero di miliardari poiché, essendo utilizzato per valutare l'inflazione, quest'ultima può influenzare la crescita economica personale. Paesi come la Nigeria o l'Ucraina, con un CPI molto elevato, presentano un numero inferiore di miliardari poiché un'elevata inflazione può creare un ambiente economico instabile e poco favorevole agli investimenti.
- *Life Expectancy*: l'aspettativa di vita è un indicatore utilizzato come misura del livello di salute e qualità di vita in un paese. Un valore più elevato è spesso associato a un'economia stabile e prospera. Anche se la relazione con il numero di miliardari non è così immediata, si nota come la Nigeria, con un'aspettativa di vita di 54 anni, conti solamente 3 miliardari.
- *Total Tax Rate*: il TTR è una misura che rappresenta la somma delle imposte pagate dalle imprese, espressa come percentuale dei loro profitti commerciali. Un TTR più basso, può incentivare l'imprenditorialità e gli investimenti come nel caso degli Stati Uniti che sono primi per numero di miliardari e presentano un TTR del 36%.
- *Population*: un paese con una popolazione più ampia potrebbe offrire un mercato più vasto e dunque una maggiore domanda. I paesi con il maggior numero di miliardari (Cina, India e Stati Uniti) sono anche quelli che presentano un numero maggiore di abitanti.

CAPITOLO 3

Tableau

3.1 Dataset

Il dataset⁶ utilizzato per l'analisi su Tableau riguarda una campagna di marketing diretto condotta da un istituto bancario portoghese, il cui obiettivo è quello di promuovere la sottoscrizione di depositi a termine tramite chiamate telefoniche ai clienti.

Il set di dati è caratterizzato dalle seguenti variabili:

<i>Variabile</i>	<i>Tipo</i>	<i>Descrizione</i>
Age	Numerica	Età del cliente
Job	Categorica	Tipo di lavoro (admin, blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown)
Marital	Categorica	Stato civile (married, single, divorced, unknown)
Education	Categorica	Livello di istruzione (primary, secondary, tertiary, unknown)
Default	Categorica	Indica se il cliente ha un credito in default (yes, no)
Balance	Numerica	Saldo medio annuo su un conto
Housing	Categorica	Indica se il cliente ha richiesto un prestito per l'acquisto di una casa (yes, no)
Loan	Categorica	Indica se il cliente ha richiesto un prestito per esigenze personali (yes, no)
Contact	Categorica	Indica il dispositivo tramite il quale il cliente ha risposto alla chiamata (cellular, telephone, unknown)

⁶ <https://archive.ics.uci.edu/dataset/222/bank+marketing>

Day	Data	Giorno del mese in cui il cliente è stato contattato l'ultima volta
Month	Data	Mese dell'anno in cui il cliente è stato contattato l'ultima volta
Duration	Numerica	Durata dell'ultima chiamata (in secondi)
Campaign	Numerica	Numero di volte che il cliente è stato contattato durante l'attuale campagna
Previous	Numerica	Numero di volte che il cliente è stato contattato prima dell'attuale campagna
Pdays	Numerica	Numero di giorni trascorsi dall'ultima volta che il cliente è stato contattato nella campagna precedente
Poutcome	Categorica	Esito della campagna precedente (success, failure, nonexistent)
Y	Binaria	Indica se il cliente ha sottoscritto (yes) o meno (no) il deposito a termine

Misure calcolate:

Yes	SUM(IF [Y] = 'yes' THEN 1 ELSE 0 END)
No	SUM(IF [Y] = 'no' THEN 1 ELSE 0 END)
Conversion Rate	([Yes]/COUNT([Y]))*100
Rejection Rate	([No]/COUNT([Y]))*100
Total Conversion Rate	([Yes]/SUM([Campaign]))*100
Total Rejection Rate	([No]/SUM([Campaign]))*100

3.2 Trasformazione dei dati



Si è proceduto alla pulizia dei dati mediante alcune operazioni:

- Rimozione valori nulli e duplicati
- Trasformazione del tipo di dato
- Rimozione variabili non utili per l'analisi
- Analisi ed eventuale rimozione dei valori "unknown"
- Rimozione errori (es. sostituito "unknown" con "nonexistent" nel caso di "Poutcome")

3.3 Obiettivo dell'analisi

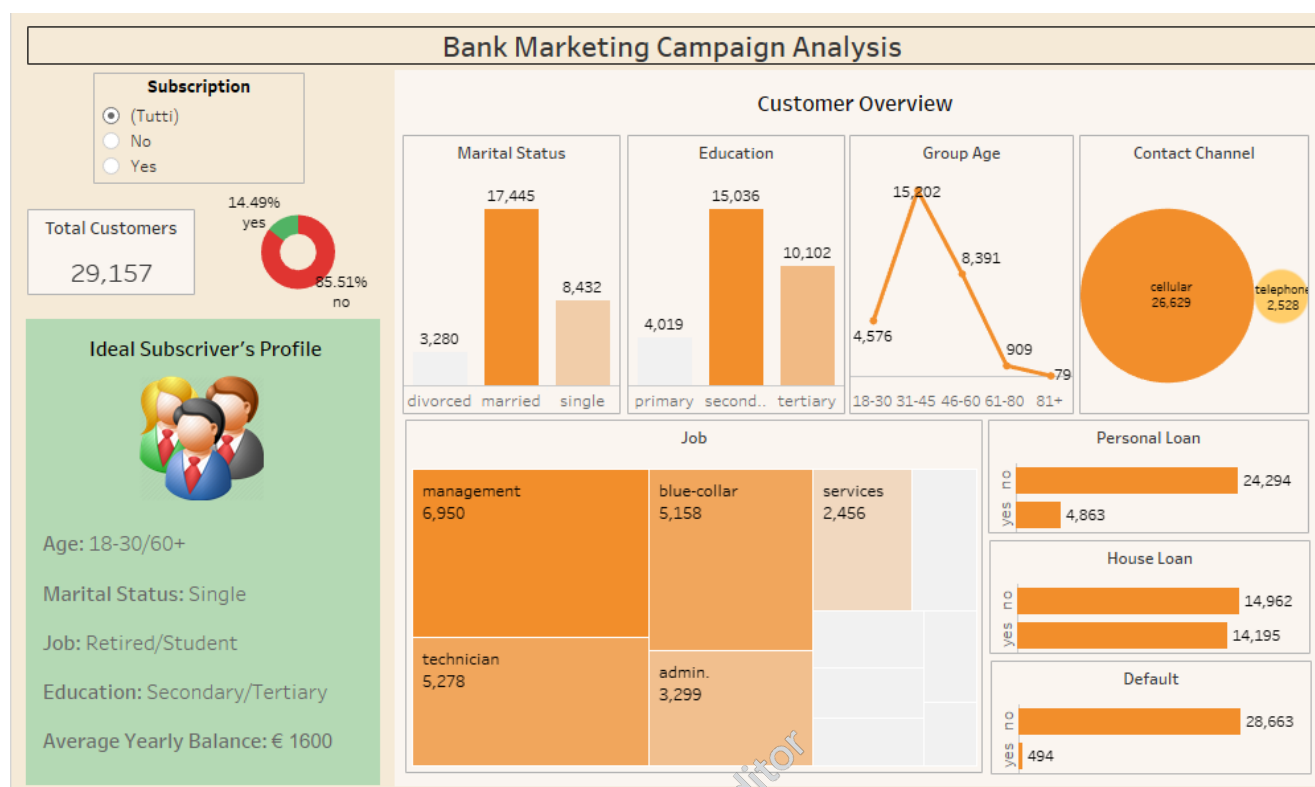
L'obiettivo di questa analisi è quello di rispondere all'esigenza dell'istituto bancario di comprendere chi sono gli individui più inclini ad accettare la proposta di sottoscrizione di un deposito a termine. Infatti, le visualizzazioni forniscono una panoramica dei modelli comportamentali dei clienti, consentendo l'identificazione dei segmenti di clientela più promettenti. Inoltre, si cerca anche di valutare l'efficacia della campagna di marketing diretto, in modo da ottimizzare le future iniziative commerciali e migliorare i risultati complessivi dell'istituto bancario.

3.4 Visualizzazioni

Per analizzare la "Bank Marketing Campaign Analysis" sono state realizzate tre dashboard:

- Customer Overview
- Campaign Overview (pt. 1)
- Campaign Overview (pt. 2)

3.4.1 Customer Overview



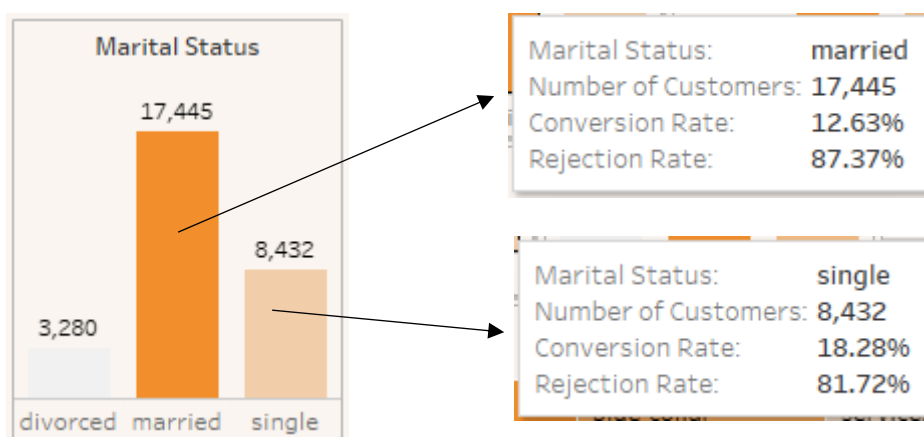
La prima è una dashboard introduttiva che fornisce una panoramica generale sugli individui che sono stati contattati dall'istituto bancario, al fine di comprendere coloro che sono più inclini a sottoscrivere un deposito a termine.

In alto a sinistra è presente un filtro denominato "Subscription" il quale permette di filtrare in base all'esito della campagna di marketing: selezionando "Yes" si ottengono gli individui sottoscrittori, viceversa selezionando "No" si ottengono gli individui che non hanno accettato la proposta commerciale.

Il KPI mostra il totale dei clienti contattati che risulta essere 29.157 mentre il grafico ad anello rappresenta la proporzione in percentuale di "Yes" e "No". Si nota che i soggetti non sottoscrittori sono in maggioranza.

Sulla destra si trova un pannello dedicato alla "Customer Overview":

I primi due grafici a barre riportano il numero di clienti raggruppati rispettivamente per stato civile e livello di istruzione. Passando il cursore sui grafici, è possibile visualizzare per ciascuna barra il tasso di conversione e quello di rifiuto.



Nonostante gli individui sposati e con un'istruzione secondaria siano quelli che accettano più frequentemente la proposta commerciale, queste categorie sono anche le più numerose nel dataset. Pertanto, per individuare un target ideale al quale l'istituto bancario può rivolgersi, si è deciso di considerare il tasso di conversione, che rapporta il numero di persone che accettano con il totale degli individui. Così facendo emerge che i single e coloro che possiedono una laurea hanno una probabilità maggiore di sottoscrivere un deposito a termine.

Il grafico a linee analizza i gruppi di età. In questo caso, le occorrenze di individui di età compresa tra 31 e 45 anni sono in maggioranza. Nonostante ciò, considerando il conversion rate, emerge che i soggetti più inclini ad accettare sono i più giovani (18-30 anni) e quelli più anziani (60+).

Il grafico a bolle mostra che gli individui contattati rispondono maggiormente da cellulare rispetto che da telefono fisso.

Il grafico ad albero rappresenta il numero di clienti raggruppati in base alla categoria lavorativa. Anche in questo caso, il settore del management è quello che figura maggiormente nei dati ma il tasso di conversione è più elevato per pensionati e studenti.

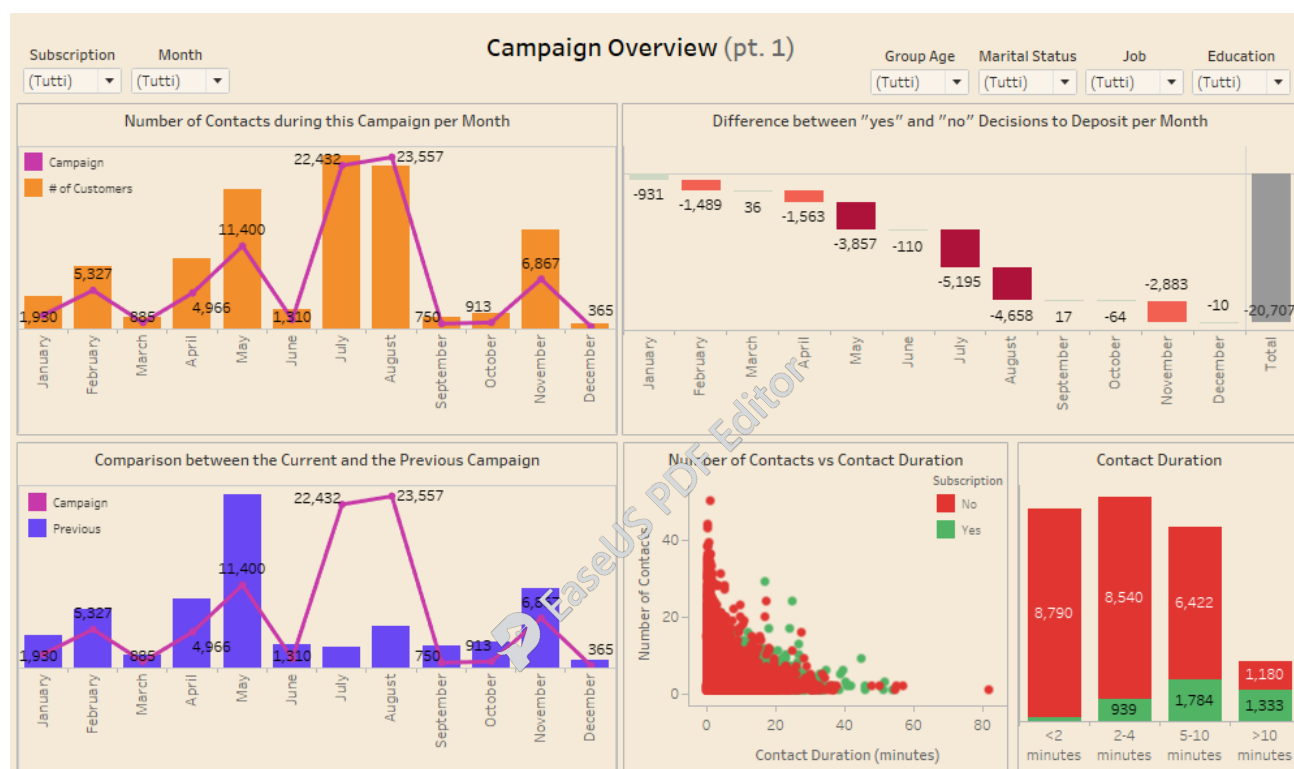
In basso a destra vengono studiate le variabili Personal Loan, House Loan e Default. Ciascuno dei grafici confronta la probabilità di accettare il prestito da parte dei clienti con le informazioni riguardanti le loro posizioni debitorie (prestito personale e prestito per l'acquisto di un'abitazione) e i loro inadempimenti (credito in default). Si nota che vi è un conversion rate maggiore per persone che non sono indebitate e che non si sono mostrate inadempienti.

Alla luce dei risultati emersi, per individuare un cliente ideale che l'istituto bancario può contattare, si è definita una Buyer Persona.

L'individuo target è una persona single, con istruzione secondaria o terziaria e appartenente alle fasce d'età 18-30 o 60+. Inoltre, risulta essere studente o pensionato, con un saldo medio annuo di €1.600. Ciò potrebbe essere spiegato dal fatto che i pensionati tendono a risparmiare una quantità considerevole di liquidità, manifestando una propensione maggiore a cedere risorse finanziarie con

lo scopo di trarne profitto. Per quanto riguarda gli studenti, questi potrebbero essere più propensi ad aprire depositi per pianificare il proprio futuro finanziario. Un altro aspetto rilevante è il livello di istruzione poiché, all'aumentare di quest'ultimo, aumenta la probabilità di accettare la proposta commerciale.

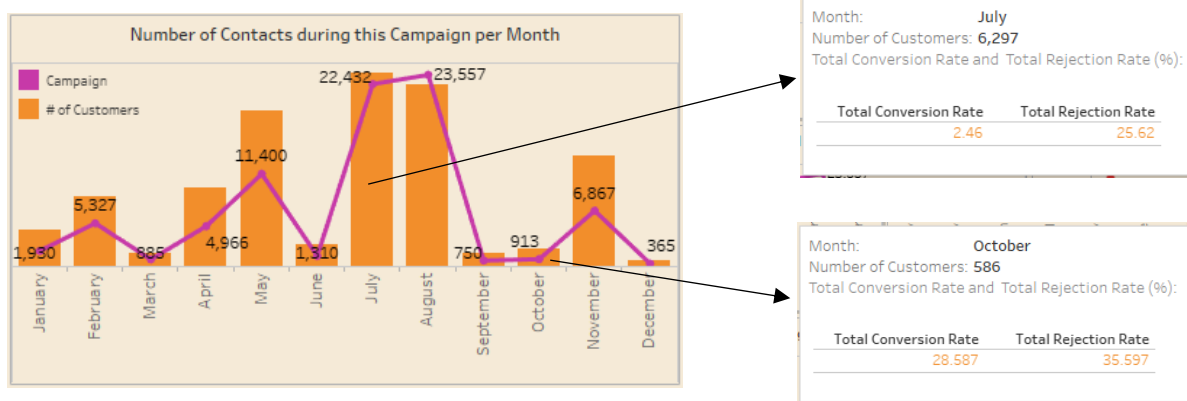
3.4.2 Campaign Overview (pt.1)



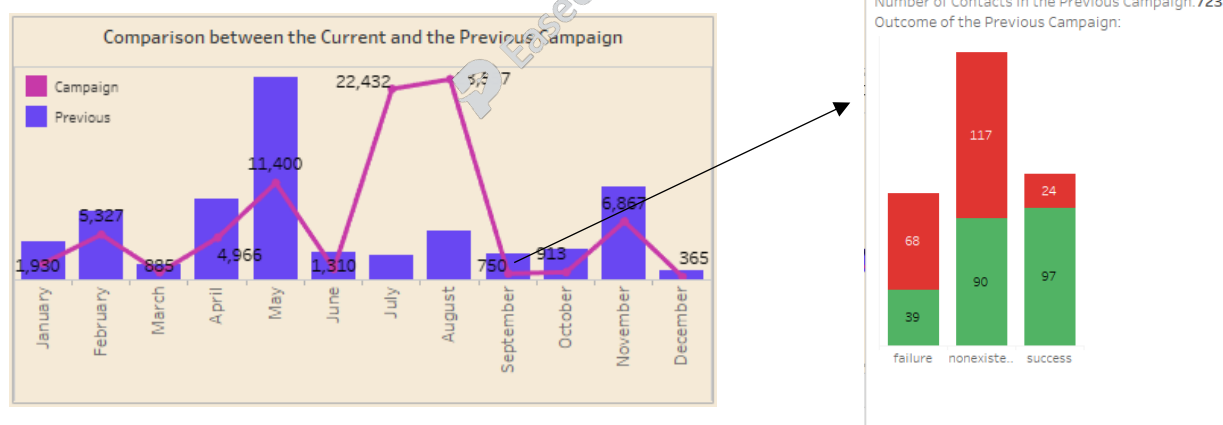
La seconda dashboard offre un'analisi dettagliata delle performance della campagna di marketing della banca nel corso dell'anno 2011. L'obiettivo è di monitorare diverse metriche per fornire all'istituto finanziario informazioni utili volte al miglioramento delle campagne future.

La dashboard è interattiva, permettendo all'utente di filtrare i dati per stato di sottoscrizione, mese e vari attributi demografici come età, stato civile, lavoro e livello di istruzione.

Il grafico in alto a sinistra mostra, oltre al numero di persone contattate mensilmente, anche il numero di volte che queste ultime sono state contattate. Confrontando le barre con la linea, si nota che le chiamate effettuate ad ogni cliente siano spesso più di una. I mesi di picco sono luglio, con 22.432 contatti, e agosto, con 23.557 contatti, indicando uno sforzo di marketing intensificato durante i mesi estivi. Osservando il tooltip inserito all'interno delle barre, si nota che non è detto che un maggior numero di chiamate porti ad un tasso di conversione più alto rispetto ai mesi dove l'attività di marketing è meno massiva.



Il grafico in basso a sinistra presenta un confronto mensile tra il numero di volte che i clienti sono stati contattati nella campagna precedente e in quella attuale. Le barre indaco mostrano i contatti mensili nella campagna antecedente. I picchi di attività attuali si evidenziano a luglio e agosto mentre nella precedente a maggio e novembre. Si nota dunque come l'istituto bancario abbia adottato una strategia diversa rispetto alla precedente. Inoltre, è stato inserito un tooltip che indica l'esito della campagna pregressa. Nonostante la maggior parte delle persone non siano state contattate durante la campagna precedente, confrontando il successo e il fallimento di quest'ultima, si nota che i sottoscrittori che hanno usufruito in modo soddisfacente del servizio, presentano una propensione maggiore ad una nuova sottoscrizione del deposito.

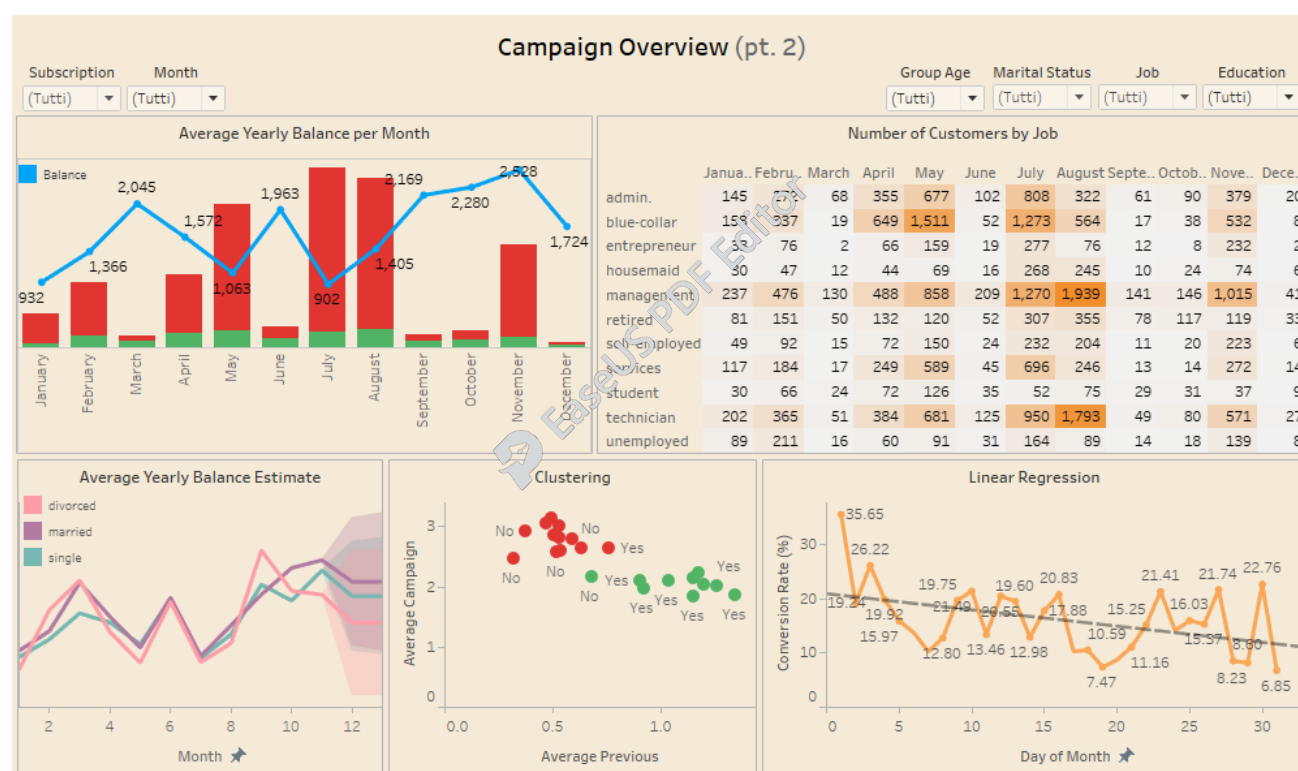


Il grafico in alto a destra presenta la differenza tra le decisioni positive e negative di effettuare un deposito, suddivisa per mese. Luglio è il mese con la maggiore differenza negativa, con un saldo di 5.525 decisioni negative rispetto a quelle positive, indicando un basso tasso di conversione. Marzo e settembre presentano un saldo positivo di accettazione.

Il grafico a dispersione esamina la relazione tra il numero di volte che i clienti sono stati contattati e la durata di ciascuna chiamata, segmentando i risultati in base all'esito della sottoscrizione. Gli esiti negativi registrano durate di contatto più brevi mentre per gli individui sottoscrittori sia il numero di chiamate che la durata di queste ultime sono più elevate. Questo suggerisce che interazioni più

prolungate con i clienti tendono a portare a conversioni di successo, probabilmente perché consentono di fornire informazioni più dettagliate e rispondere a eventuali domande o preoccupazioni dei clienti. Infine, il grafico a barre in pila conferma quanto detto poiché, quando la durata della chiamata è tra i 5 e i 10 minuti, si registra il maggior numero di sottoscrizioni; questa fascia potrebbe essere costituita da persone potenzialmente interessate ma ancora indecise. Invece, quando la durata supera i 10 minuti coloro che sottoscrivono sono di più rispetto a chi non lo fa, indicando una probabilità sempre più elevata di ricevere riscontro positivo.

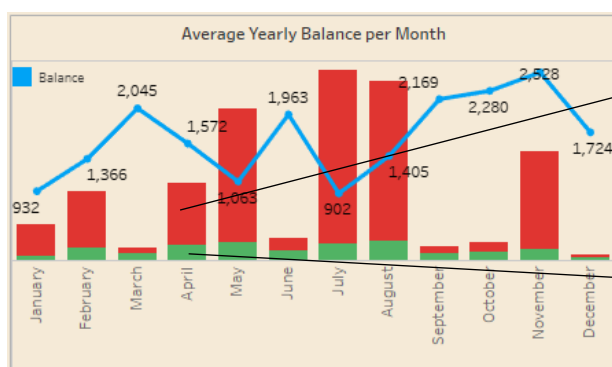
3.4.3 Campaign Overview (pt.2)



La terza dashboard continua l'analisi della campagna di marketing.

Il grafico in alto a sinistra, oltre a mostrare il numero di persone contattate ogni mese divise tra soggetti sottoscrittori e non sottoscrittori, evidenzia anche l'andamento del saldo medio annuo sui conti bancari degli individui. Quest'ultimo presenta dei picchi in particolari mesi dell'anno, dunque, l'istituto bancario dovrebbe concentrare i propri sforzi promozionali nei mesi in cui il saldo medio è più elevato. Infatti, un saldo più alto indica una maggiore disponibilità finanziaria, aumentando la probabilità che i clienti possano permettersi di vincolare una parte dei loro fondi in un deposito a termine. Inoltre, osservando i KPI inseriti in formato tooltip, si nota, come prevedibile, che per la

maggior parte dei mesi, il saldo medio annuo sul conto dei soggetti sottoscrittori è maggiore rispetto a quello dei non sottoscrittori.



Month: April
Subscription: No
Number of Customers: 2,067
Average Yearly Balance by Subscription (€):
No
1,490

Month: April
Subscription: Yes
Number of Customers: 504
Average Yearly Balance by Subscription (€):
Yes
1,910

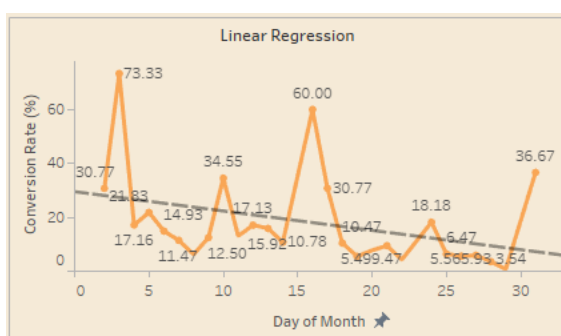
Il grafico in basso a sinistra mostra la previsione per il primo mese dell'anno successivo (2012) del saldo medio annuo sul conto dei soggetti, distinti sulla base dello stato civile (sposati, divorziati, single). Si nota che i soggetti sposati presentano mediamente un saldo più elevato pari a €2.257, seguiti dai single e dai divorziati rispettivamente con un saldo medio di €2.000 e di €1.500.

La tabella in alto a destra mostra il numero di contatti mensili, suddivisi per categoria lavorativa. Si osserva che gli individui con un'occupazione nell'ambito del management, evidenziati con una colorazione più intensa, sono stati contattati maggiormente nel mese di agosto. In generale, il mese di dicembre è quello in cui si è registrato il minor numero di contatti. Attivando il filtro "Subscription" con l'opzione "Yes", è possibile concentrarsi esclusivamente sui soggetti più inclini ad accettare la proposta di sottoscrizione.

La variabile "Job" è stata utilizzata anche per fare clustering al fine di classificare gli individui in gruppi, in base al numero di volte che sono stati contattati durante la campagna attuale e in quella precedente. Nonostante la categoria lavorativa non sia determinante nella discriminazione dei gruppi, poiché sono presenti lavoratori dello stesso settore in entrambi i cluster, è risultata utile per una visualizzazione più comprensibile dei risultati, rispetto a considerare l'intero campione. Sono stati identificati due gruppi piuttosto distinti: uno composto principalmente da soggetti che hanno accettato la proposta (gruppo verde) e l'altro da coloro che non lo hanno fatto (gruppo rosso). Si può notare che gli individui che sono stati contattati almeno una volta nella campagna precedente e hanno ricevuto mediamente due chiamate nell'attuale campagna sono più propensi ad accettare l'offerta. Al contrario, coloro che non sono mai stati contattati o sono stati contattati al massimo una volta nella campagna pregressa, ma hanno ricevuto mediamente tre chiamate nella campagna attuale, sono più inclini a rifiutare la proposta di sottoscrizione. Queste informazioni sono utili poiché suggeriscono all'istituto bancario di adottare una politica che stabilisca un limite massimo di due contatti per ciascun potenziale cliente così da limitare costi superflui ed evitare di risultare insistenti, tenendo conto che

coloro che sono già stati contattati nella campagna precedente, e in particolare se quest'ultima ha avuto riscontro positivo, è più probabile che risultino nuovamente interessati.

Il grafico in basso a destra offre un livello di dettaglio maggiore concentrandosi sui singoli giorni del mese. L'obiettivo è individuare il periodo del mese in cui la probabilità di sottoscrivere un deposito a termine è più elevata. Analizzando i dati relativi a tutti i mesi dell'anno, emerge che all'inizio del mese il tasso di conversione è più alto rispetto ai giorni successivi, suggerendo all'istituto bancario di concentrare le chiamate soprattutto nei primi giorni del mese. Ciò è confermato dalla linea di tendenza, poiché, con il passare dei giorni, si nota una diminuzione della probabilità di sottoscrizione. Tuttavia, è importante sottolineare che la tendenza generale potrebbe cambiare considerando i singoli mesi. Pertanto, se si è interessati ad un mese specifico, è consigliabile applicare un filtro.



Ad esempio, nel mese di agosto si ha un conversion rate più alto non solo all'inizio ma anche a metà mese. Inoltre, si nota anche una sorta di ricorsività poiché i picchi del tasso di conversione si verificano nei primi giorni di ogni settimana, suggerendo di privilegiare le chiamate durante questo periodo.