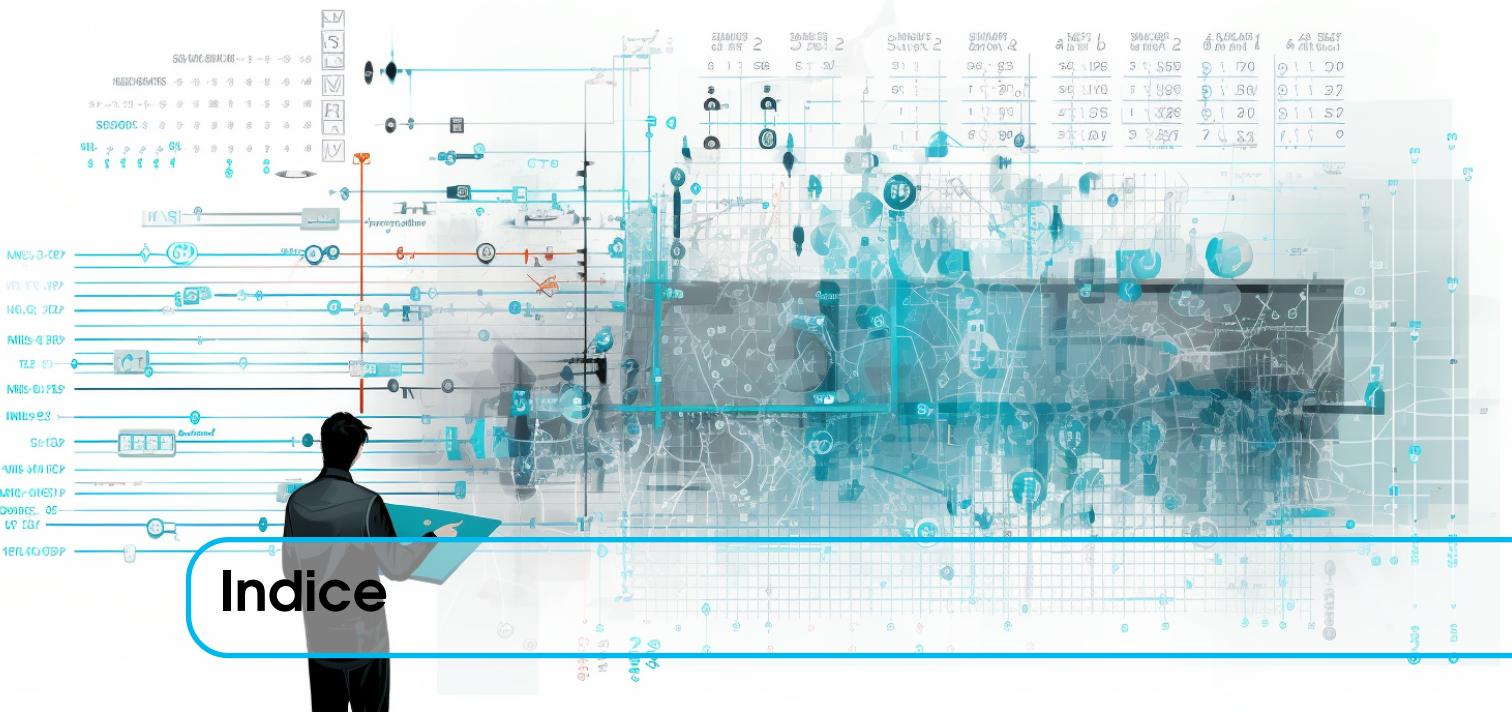


Progetto di Business Intelligence

Simone Recinelli

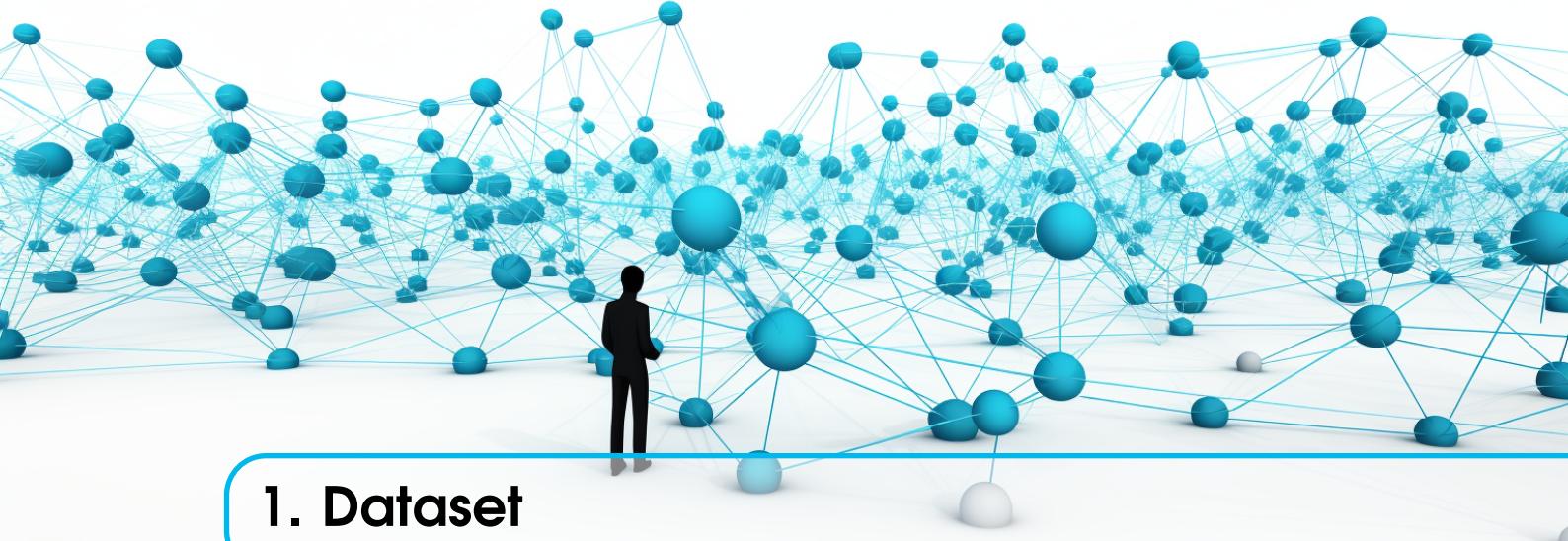
Diego Santarelli

Andrea Marini



1	Dataset	4
2	Qlik	7
2.1	Introduzione a Qlik	7
2.2	Caricamento dei dati	7
2.3	Preprocessing	8
2.4	Analisi dei dati e rappresentazione grafica	8
2.5	Foglio 1: Canzoni in top 50 in un Paese in una data selezionata	9
2.6	Foglio 2: Popolarità di una canzone in base ai suoi valori	10
2.7	Foglio 3: Artisti più popolari con durata media delle loro canzoni e i loro valori	11
3	Tableau	14
3.1	Introduzione a Tableau	14
3.2	Dashboard realizzate con Tableau	15
3.2.1	Preprocessing	15
3.3	Dashboard 1: Analisi dei brani con una notevole componente acustica	17
3.3.1	Considerazioni finali	18
3.4	Dashboard 2: Analisi della popolarità e danazabilità di un brano	19
3.4.1	Considerazioni finali	20
3.5	Dashboard 3: Analisi del daily movement dei top 6 brani in Spagna	20
3.5.1	Considerazioni finali	22
3.6	Dashboard 4: Analisi e previsione dell'andamento di un brano	22
3.6.1	Stime e previsioni sull'andamento	25

3.6.2	Qualità delle previsioni	28
4	Power BI	29
4.1	Introduzione a Power BI	29
4.2	Caricamento e pulizia dei dati	30
4.3	Data analysis e rappresentazione grafica	31
4.3.1	Foglio 1 - Confronto tra "top songs"	31
4.3.2	Foglio 2 - Analisi degli artisti con cui effettuare un featuring	32
4.3.3	Foglio 3 - Andamento giornaliero in classifica di "All I Want For Christmas Is You" con previsione temporale	36
4.3.4	Foglio 4 - Confronto tra paesi in Nord America	38
4.3.5	Foglio 5 - Continuo del confronto tra Paesi in Nord America	41



1. Dataset

Il dataset considerato comprende dati statistici completi sulle caratteristiche dei brani di Spotify in tendenza in oltre 70 Paesi; esse sono aggiornate ogni giorno al fine di fornire informazioni il più aggiornate possibile sulla popolarità delle canzoni nel mondo. Tale dataset è reperibile al seguente link: <https://www.kaggle.com/datasets/asaniczka/top-spotify-songs-in-73-countries-daily-updated>.

Naturalmente, l'obiettivo del progetto è quello di ottenere una conoscenza approfondita dai dati considerati.

Si è scelto di utilizzare e studiare un dataset di questo tipo per diversi fattori, il primo tra tutti è dettato dalla *"usability"*, cioè una valutazione di quanto il dataset sia accessibile, comprensibile e utilizzabile da parte degli utenti che ne fanno uso per analisi, modellazione o altre attività; infatti, la *"usability"* di questo dataset è massima, ovvero pari a 10.0.

Inoltre, il dataset è molto ben curato in quanto viene aggiornato giornalmente in base alle classifiche giornaliere di Spotify.

L'ultimo motivo, ma non per importanza, per il quale abbiamo deciso di utilizzare un dataset relativo alla musica è perché ormai, ai giorni d'oggi, la musica gioca un ruolo fondamentale nella vita di tutti i giorni, soprattutto tra i giovani. In aggiunta, è molto stimolante ed utile definire delle analisi e metodi per capire quando un brano, un album o un'artista possa essere definito popolare.

Il nome del file all'interno del quale è contenuto il Dataset è `universal_top_spotify_songs.csv` e le colonne che lo caratterizzano sono le seguenti:

Tabella 1.1: Tabella `universal_top_spotify_songs.csv`

Attributo	Significato
<code>spotify_id</code>	L'identificatore univoco del brano nel database di Spotify. (tipo: str)

Continua nella pagina successiva

Tabella 1.1 – continua dalla pagina precedente

Attributo	Significato
name	Titolo della canzone (tipo: str)
artists	Nome/nomi dell'artista/artisti associati alla canzone (tipo:str)
daily_rank	Posizione giornaliera di una canzone (tipo: int)
daily_movement	Cambiamento giornaliero di una canzone nella classifica (tipo: int)
weekly_movement	Cambiamento settimanale di una canzone nella classifica (tipo: int)
country	Codice ISO del Paese nella classifica (tipo: str)
snapshot_date	Data nella quale sono stati reperiti i dati dall'API di Spotify (tipo: date)
popularity	Misura della popolarità corrente della canzone su Spotify (tipo: int)
is_explicit	Indica se la canzone contiene del testo esplicito o meno (tipo: bool)
duration_ms	Durata della canzone in millisecondi (tipo: int)
album_name	Titolo dell'album al quale appartiene la canzone (tipo: str)
album_release_date	Data di uscita dell'album al quale appartiene la canzone (tipo: date)
danceability	Una misura di quanto la canzone sia adatta alla danza basata su vari elementi musicali (tipo: float)
energy	Una misura dell'intensità e del livello di attività della canzone (tipo: float)
key	Chiave della canzone (tipo: int)
loudness	Il volume complessivo della canzone in decibel (tipo: float)
mode	Indica se la canzone è in chiave maggiore o minore (tipo: int)
Continua nella pagina successiva	

Tabella 1.1 – continua dalla pagina precedente

Attributo	Significato
speechiness	Misura della presenza di parole dette nella canzone (tipo: float)
acousticness	Misura della qualità acustica della canzone (tipo: float)
instrumentalness	Misura della probabilità che la canzone non contenga parti vocali (tipo: float)
liveness	Misura della presenza di un pubblico dal vivo nella registrazione (tipo: float)
valence	Misura della positività musicale trasmessa dalla canzone (tipo: float)
tempo	Tempo della canzone in battiti al minuto (tipo: float)
time_signature	Durata complessiva stimata del brano (tipo: int)

Per quanto riguarda la fase di **preprocessing**, cioè la fase volta alla preparazione dei dati per l'analisi e l'addestramento dei modelli, essa viene descritta in tutti e 3 gli ambienti di analisi dei dati: *Qlik*, *Tableau*, *PowerBI*.



2. Qlik

2.1 Introduzione a Qlik

La crescente complessità del panorama aziendale e la proliferazione di dati provenienti da varie fonti hanno reso cruciale l'adozione di strumenti avanzati di Business Intelligence (BI) per supportare le decisioni organizzative. In questo contesto, Qlik emerge come una piattaforma innovativa e potente, fondata nel 1993 da Björn Berg e Staffan Gestrelius a Lund in Svezia e si distingue per il suo approccio unico noto come "associative model". Questo modello consente agli utenti di connettere e analizzare dinamicamente dati provenienti da diverse fonti, facilitando l'esplorazione intuitiva e la scoperta di relazioni significative. La piattaforma Qlik, con il suo strumento principale **Qlik Sense** per le «Self-service Analytics», offre la possibilità di creare dashboard interattive e report personalizzati, fornendo agli utenti la libertà di analizzare dati in modo autonomo. Oltre a Qlik Sense, l'azienda svedese ha messo a disposizione un altro software, **Qlik View**, che, al contrario, si basa su «Guided Analytics», in cui l'utente finale non ha la capacità di aggiungere elementi all'applicazione fornita e risulta, quindi, uno strumento per le aziende per aggiornare grafici secondo la navigazione dell'utente e le fonti di dati sottostanti.

2.2 Caricamento dei dati

L'accesso e l'analisi dei dati sono diventati pilastri essenziali per il successo delle organizzazioni in un panorama aziendale sempre più complesso. In questo contesto, la corretta gestione del caricamento dei dati riveste un ruolo centrale nell'ecosistema di Qlik.

Il processo di caricamento dei dati in Qlik è cruciale per garantire che le informazioni siano non solo accessibili, ma anche comprensibili e utilizzabili per prendere decisioni informate. Qlik offre strumenti avanzati di Estrazione, Trasformazione e Caricamento dei Dati (**ETL**¹) che consentono di raccogliere, organizzare e preparare i dati provenienti da svariate fonti e di centralizzarli in un unico repository. Un elemento distintivo di Qlik è la sua architettura in memoria, che consente di caricare grandi quantità di dati in una cache di memoria per

¹Extract, Transform, e Load

garantire un accesso rapido e interattivo. Ciò significa che gli utenti possono analizzare i dati in modo immediato, senza dover attendere lunghi tempi di caricamento. Durante il processo di ETL, è possibile applicare trasformazioni e manipolazioni ai dati per renderli più adatti all'analisi. Ad esempio, è possibile unire tabelle, filtrare dati non necessari, o convertire formati per uniformare le informazioni provenienti da diverse fonti. Questa fase è fondamentale per ottenere una visione coerente e completa dei dati. La piattaforma Qlik consente inoltre la connessione a una vasta gamma di fonti dati (come server o file CSV, Excel...), sia interne che esterne, facilitando l'integrazione con sistemi aziendali esistenti. Questa flessibilità è essenziale per comprendere appieno l'ambiente operativo e poter prendere decisioni basate su informazioni complete e aggiornate.

2.3 Preprocessing

Nel caso di Qlik, la fase di preprocessing è stata molto rapida, in quanto il software ha effettuato automaticamente le conversioni di cui aveva bisogno per la corretta interpretazione dei dati. L'unico cambiamento manuale è avvenuto per l'attributo *duration_ms*, per il quale è stata effettuato il passaggio a minuti. In questo caso, abbiamo eseguito la query visibile in 2.1 per fare la media dei minuti di ogni canzone.

```
1 Avg(duration_ms/(1000*60))
```

Listato 2.1: Query eseguita su Qlik.

2.4 Analisi dei dati e rappresentazione grafica

Una volta completata con successo la fase di caricamento dei dati all'interno della piattaforma Qlik e consolidati tutti i collegamenti, se si stanno utilizzando più fonti di dati, siamo pronti a procedere alla creazione dei fogli. I **fogli** sono il nome che identifica le varie dashboard all'interno della piattaforma di Qlik e permettono di strutturare le idee e realizzare lo scopo di un'app, rappresentando ognuno un'idea o un obiettivo; una **dashboard** è un'interfaccia visiva interattiva che raccoglie e presenta dati in modo chiaro e comprensibile.

Le dashboard che abbiamo utilizzato nella nostra analisi sono:

- Grafico a barre: visualizza una barra per ogni valore di dimensione². La lunghezza della barra corrisponde al relativo valore numerico della misura³;
- Grafico lineare: visualizza le linee dei dati tra i valori. Vengono spesso utilizzati per visualizzare una tendenza nei dati su un intervallo di tempo;
- Casella di filtro: consente di filtrare contemporaneamente i dati di più dimensioni;
- Tabella: visualizza i valori sotto forma di record, in modo che ogni riga della tabella contenga i campi calcolati utilizzando le misure;
- KPI: viene utilizzato per presentare dati quantitativi fondamentali relativi alle prestazioni;
- Mappa ad albero: mostra dati generici. Può mostrare un gran numero di valori simultaneamente all'interno di uno spazio limitato;
- Grafico a torta: visualizza la relazione tra una singola dimensione ed una singola misura.

In particolare, la nostra analisi in Qlik conterrà i seguenti fogli:

- Foglio 1: Canzoni in top 50 in un Paese in una data selezionata;
- Foglio 2: Popolarità di una canzone in base ai suoi valori;

²Campi dati con valori qualitativi

³Campi dati con valori quantitativi da misurare e/o aggregare

- Foglio 3: Artisti più popolari con durata media delle loro canzoni e i loro valori.

2.5 Foglio 1: Canzoni in top 50 in un Paese in una data selezionata

Come prima analisi svolta nell'ambiente del software Qlik si è scelto di incentrare l'attenzione sulla classifica dei top 50 brani. Nella dashboard illustrata in Figura 2.1, si possono osservare dati dettagliati e applicare filtri variabili per condurre un'analisi sul ranking di diverse canzoni in un Paese e data selezionati. Inoltre, si può scegliere una determinata canzone di cui si vuole sapere l'andamento temporale della posizione in classifica e la sua popolarità.

La dashboard è composta da 8 elementi, ovvero una tabella, due grafici lineari e tre caselle di filtraggio. In particolare, come descritto precedentemente, i filtri avranno come dimensioni la data, il Paese e il nome della canzone. Per quanto riguarda la tabella abbiamo messo in relazione il nome della canzone (dimensione) e la sua posizione giornaliera (misura); invece, nei grafici lineari abbiamo nell'asse delle ascisse le varie date, in modo da ricostruire l'andamento temporale, e negli assi delle ordinate la posizione giornaliera e la popolarità della singola canzone. Infine abbiamo inserito 2 KPI per riportare valori significativi per la nostra analisi.

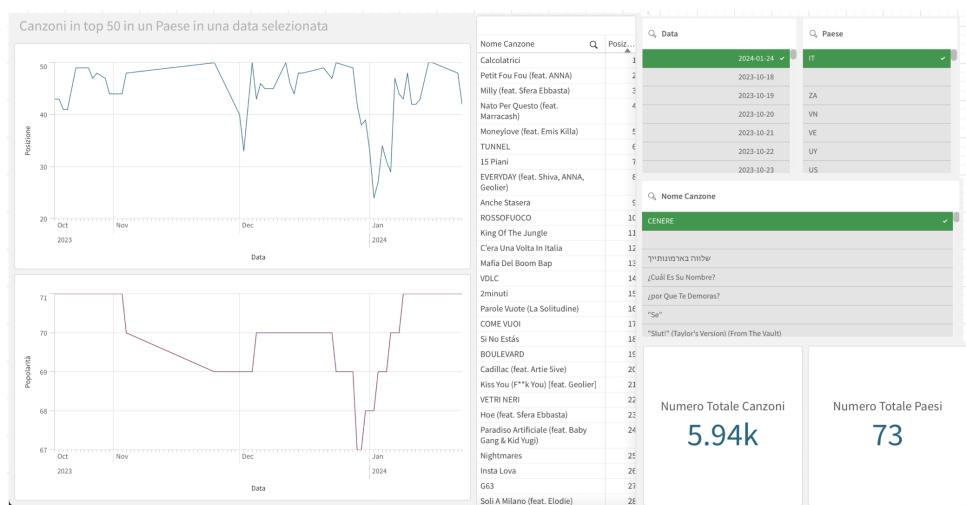


Figura 2.1: Canzoni in top 50 in un Paese in una data selezionata

In particolare si è scelto di analizzare la classifica (top 50) dei brani in Italia in data 24-01-2024, in cui possiamo notare che le prime tre posizioni sono occupate dalle canzoni "Calcolatrici", "Petit Fou Fou" e "Milly" per poi proseguire con i brani in posizioni successive. Ad ogni modo, l'applicazione dei filtri relativi alla classifica è visibile nella parte in alto a destra, in cui è appunto selezionata la data "24-01-2024" e come Paese "Italia".

Nella parte in basso a destra, nelle due KPI, è possibile notare il numero totale di canzoni su cui viene condotta l'analisi, che risultano essere **5.940**, e il numero totale dei Paesi, ovvero 73. È importante sottolineare il fatto che i filtri precedentemente citati non sono applicati a questi due grafici e quindi indicano la totalità dei dati presenti nel dataset.

Infine, nella parte sinistra della dashboard, sono presenti due grafici in cui viene visualizzato rispettivamente l'andamento temporale della posizione di un brano specifico e della sua popolarità. Il brano in questione è "Cenere", che è stato selezionato dall'ultima casella di filtro non precedentemente menzionata, presente nella parte destra della dashboard. Dal primo grafico risulta quindi che, durante i mesi di ottobre e novembre, "Cenere" ha mantenuto

posizione in classifica, nonostante bassa, per poi scalare la classifica ad inizio dicembre e raggiungere la sua posizione più alta ad inizio gennaio.

Nota: Nel grafico lineare relativo all'andamento temporale della posizione in classifica di una canzone, picchi più bassi corrispondono a posizioni più alte in classifica; viceversa, picchi più alti indicano una posizione più bassa.

Nel secondo grafico, invece, è possibile notare che il brano ha mantenuto una buona popolarità nel mese di ottobre per poi avere un decremento durante il mese di novembre. Successivamente, ad inizio dicembre e ad inizio gennaio, si registra un aumento della popolarità del brano, per poi continuare a mantenere una buona popolarità durante il mese di gennaio. L'andamento della popolarità può essere, in un certo senso, riscontrato anche nel primo grafico: ad un aumento della popolarità nel secondo grafico, corrisponde un aumento della posizione in classifica della canzone.

Nota: Nonostante nel grafico la canzone "Cenere" sembra avere una popolarità alta, in verità il valore massimo delle ordinate, che corrisponde al tasso di popolarità, è 71, quando il valore massimo raggiungibile da una canzone è 100.

2.6 Foglio 2: Popolarità di una canzone in base ai suoi valori

Nella dashboard illustrata in Figura 2.2 si possono osservare dati dettagliati sulla popolarità di alcune canzoni e su alcuni loro valori. In particolare, lo scopo della dashboard è quello di individuare le cinque canzoni più popolari e riportare i loro valori di acustica, danzabilità, loquacità ed energia, stabilendone una soglia minima. In questo modo si cerca di individuare i valori che dovrebbe avere una canzone per avere maggiore popolarità. Gli elementi grafici presenti nel foglio, combinati, forniscono un'analisi complessiva dell'appeal audio delle canzoni e possono essere utilizzati per guidare decisioni di marketing, tramite l'analisi di tendenze musicali.



Figura 2.2: Popolarità di una canzone in base ai suoi valori

La dashboard è formata da 6 elementi, ovvero un grafico a barre raggruppato verticale, un grafico a barre raggruppato orizzontale ed infine quattro KPI. In particolare, nel primo abbiamo come dimensione nell'asse delle ascisse il nome delle canzoni e la popolarità nell'asse delle ordinate; nel secondo, nell'asse verticale i nomi delle canzoni selezionate e in quella

2.7 Foglio 3: Artisti più popolari con durata media delle loro canzoni e i loro valori

orizzontale i valori di acustica, danzabilità, loquacità ed energia; infine vengono riportati i valori esatti delle grandezze sopra citate (valori che vanno da 0 a 1).

In particolare, nel primo grafico in alto a sinistra, vengono individuate le cinque canzoni più popolari, in base al loro valore medio di popolarità nei 73 Paesi in cui viene condotta l'analisi, e risultano essere "*greedy*", "*Cruel Summer*", "*Si no Estas*", "*My Love Mine All Mine*" e "*Strangers*".

Successivamente, nel grafico a destra, vengono individuati graficamente i valori di *acustica*, *danzabilità*, *loquacità* ed *energia* per le cinque canzoni selezionate.

Infine, per esplicitarli al meglio, nella parte in basso a sinistra della dashboard, vengono riportati i valori esatti di *acustica*, *danzabilità*, *loquacità* ed *energia*. La soglia minima individuata risulta essere 0.12 per l'*acustica*, 0.5 per la *danzabilità*, 0.03 per la *loquacità* e 0.31 per l'*energia*.

Per verificare ciò che è stato appena detto, abbiamo realizzato un'altra dashboard (Figura 2.3) in cui vengono riportate a sinistra due KPI, contenenti la popolarità media di due canzoni, una con valore più elevato ("*As it was*") e l'altra con uno più basso ("*I Wish it could be Christmas everyday*"). Nella parte destra, invece, vengono riportati, sempre tramite KPI, i valori di *acustica*, *danzabilità*, *loquacità* ed *energia* delle due canzoni.

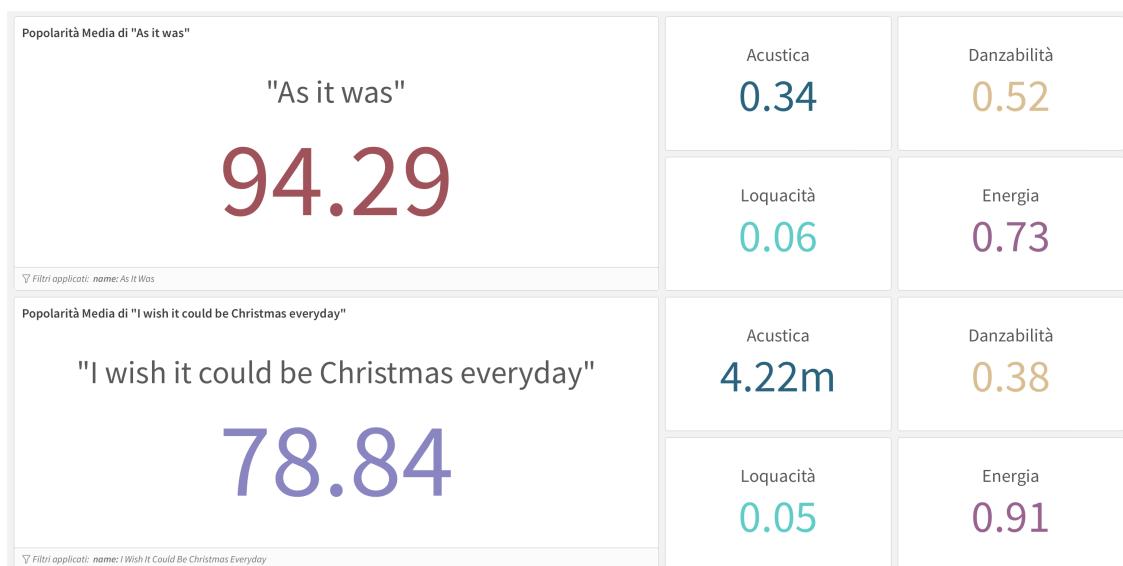


Figura 2.3: Confronto tra due canzoni

Come possiamo notare, la canzone più popolare rispetta la soglia minima di valori trovati nella dashboard precedente, mentre il secondo brano presenta valori di acustica e danzabilità inferiori. Infatti dalla dashboard in figura 2.2 risulta che il valore di acustica deve essere almeno pari a 0.12, mentre quello di danzabilità pari a 0.5; il brano "*I Wish it could be Christmas everyday*" ha, invece, valori 0.00422 per l'*acustica* e 0.38 per la *danzabilità*. I valori di *loquacità* ed *energia* vengono, al contrario, rispettati.

2.7 Foglio 3: Artisti più popolari con durata media delle loro canzoni e i loro valori

Nella dashboard illustrata in Figura 2.4 abbiamo unito due fogli, uno in cui vengono selezionati gli otto artisti più popolari e viene studiata la durata media delle loro canzoni e i loro valori medi di acustica, danzabilità, loquacità ed energia, mentre nell'altro l'andamento temporale della popolarità di un artista selezionato.

La dashboard è composta 8 elementi, ovvero una mappa ad albero in cui vengono indicati i nomi degli artisti e mediante il gradiente del colore blu la loro popolarità media (un colore blu più scuro indica una popolarità maggiore, mentre una tonalità più chiara indica una popolarità minore), un grafico a torta in cui vengono indicati i valori medi delle durate delle canzoni degli otto artisti selezionati, quattro KPI in cui vengono riportati i valori medi di acustica, danzabilità, loquacità ed energia delle loro canzoni ed infine un grafico lineare con l'andamento temporale della popolarità (anche in questo caso nelle ascisse abbiamo le date e nelle ordinate la popolarità) di un artista selezionato mediante la casella di filtro adiacente.

In particolare, nel primo grafico in alto a sinistra risultano come artisti più popolari (in ordine decrescente di popolarità): Inigo Quintero, Mitsky, Kenya Grace, Tate McRae, Bad Bunny, Mariah Carey, Billie Eilish e Brenda Lee. Nel grafico in alto a destra, invece, viene indicata, per ogni artista, la durata media delle loro canzoni. Da questo risulta: la durata media delle canzoni di Inigo Quintero è 3.07 minuti, quella di Mitsky è 2.3 minuti, Kenya Grace 2.88 minuti, Tate McRae 2.26 minuti, Bad Bunny 2.71 minuti, Mariah Carey 3.97 minuti, Billie Eilish 3.71 minuti e Brenda Lee 2.1 minuti.

Nota: La durata effettiva in minuti e secondi è ottenibile semplicemente separando la parte intera dei valori del grafico (che rappresenta i minuti) e moltiplicando la parte decimale per 60; il valore ottenuto rappresenta i secondi da aggiungere ai minuti.

Successivamente, nella parte centrale della dashboard vengono riportati i valori medi di acustica, danzabilità, loquacità ed energia delle canzoni degli artisti sopra citati. Questi risultano essere 0.53 per l'acustica, 0.62 per la danzabilità, 0.07 per la loquacità e 0.53 per l'energia. Questi valori risultano, inoltre, in linea con quelli trovati nella dashboard in figura 2.2.

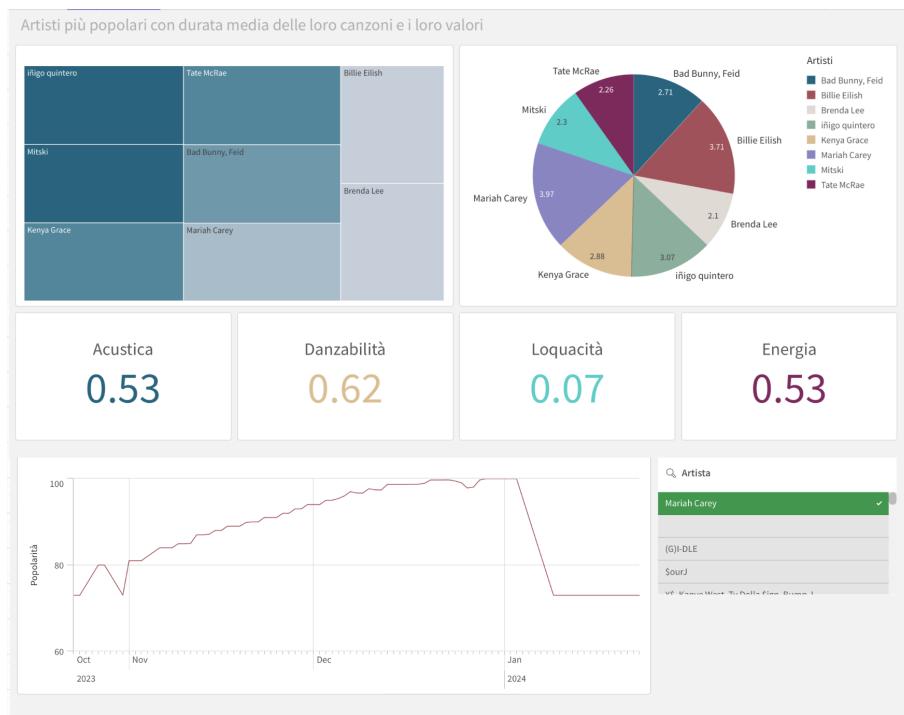


Figura 2.4: Artisti più popolari con durata media delle loro canzoni e i loro valori

Infine, per completare l'analisi, abbiamo riportato l'andamento temporale della popolarità di un artista selezionato tramite una casella di filtro. L'artista in questione è Mariah Carey, celebre per la sua canzone natalizia "All I want for Christmas is you". Come ci potevamo aspettare, l'artista presenta una popolarità "scarsa" durante il mese di ottobre, per poi

2.7 Foglio 3: Artisti più popolari con durata media delle loro canzoni e i loro valori

cominciare a salire nel mese di novembre e raggiungere il suo picco nel mese di dicembre, in particolare alla fine, ovvero nel periodo natalizio. Finite le festività, da poco dopo l'inizio di gennaio, l'artista ritorna ad avere una popolarità più bassa.



3. Tableau

3.1 Introduzione a Tableau

Tableau è un'azienda americana fondata nel 2003 e, rispetto a Qlik e Power BI, costituisce lo strumento più avanzato per quanto concerne la parte di visualizzazione dei dati offrendo un range di personalizzazione molto vario al fine di produrre dashboard adatte al contesto che si vuole analizzare. Un altro vantaggio di Tableau è il suo sistema interno che offre la possibilità di eseguire delle query direttamente sui dati. Il suddetto software, nel 2019, è stato venduto a cifre esorbitanti a Salesforce.

All'interno del contesto di questo progetto di Business Intelligence, il focus di questo capitolo è dedicato alla creazione di dashboard (composte da più fogli) mediante l'utilizzo di Tableau Desktop, riconosciuto come uno dei leader indiscutibile nel panorama degli strumenti per la visualizzazione e l'analisi dati. Tableau si distingue per aver reso l'analisi dati potente e, al contempo, accessibile. Grazie alla sua interfaccia utente intuitiva, gli analisti possono creare facilmente grafici complessi e dashboard interattive attraverso un semplice sistema di trascinamento. Questo approccio intuitivo ha aperto le porte dell'analisi dati a un vasto pubblico di utenti aziendali, consentendo loro di ottenere approfondimenti significativi senza la necessità di dipendere esclusivamente da analisti di dati o specialisti nell'ambito dell'Information Technology (IT).

Ovviamente su Tableau vi è la possibilità di importare e collegare tra loro dataset eterogenei in quanto consente di importare dati da file *Excel*, file di testo, file *JSON*, file *PDF*, database e altre fonti. I connettori di Tableau seguono due modalità:

- **Live:** è preferibile quando la sorgente dati è veloce o è ottimizzata per sostenere analisi interattive, quando è necessario vedere i dati aggiornati continuamente (magari minuto per minuto) oppure quando la mole di dati è veramente enorme. Questo viene realizzato attraverso l'invio di query SQL dinamiche o MDX statement (MultiDimensional eXpressions è un linguaggio informatico sintatticamente simile al linguaggio SQL ma ha una semantica diversa) direttamente alla sorgente dei dati invece di importarli. Questa modalità rappresenta la prima differenza col software Qlik in quanto esso non la possiede.
- **In-memory:** è ideale quando la sorgente dati è troppo lenta per sostenere analisi

interattive, quando i dati vanno presi da una sorgente transazionale oppure quando si deve lavorare offline. Il motore di Tableau è ottimizzato per sfruttare il sistema sottostante (sia in termini di disco e RAM) per poter eseguire query veloci su milioni di righe di dati.

3.2 Dashboard realizzate con Tableau

Analogamente a quanto fatto con Qlik, il dataset è stato importato in Tableau con la modalità In-Memory (Extract). Una volta importato il dataset, si possono esplorare le prime righe dei dati per ottenere una visione iniziale.

Tableau riconosce e classifica autonomamente quali attributi costituiscono delle misure e quali delle dimensioni: nel nostro caso, in Figura 3.1, è possibile visionare come, ad esempio, l'attributo "Acusticità", che corrisponde al tasso di acustica di un brano musicale, sia classificato come una dimensione in quanto è un valore numerico non aggregabile, a differenza dell'attributo "Danzabilità", il quale indica il tasso di danzabilità di un brano ed è classificato come misura in quanto è un valore numerico aggregabile.

3.2.1 Preprocessing

Per quanto riguarda il preprocessing, si è proceduto alla traduzione dei nomi degli attributi del dataset, inizialmente in lingua inglese, in lingua italiana.

Successivamente, sono state eseguite operazioni di pulizia e trasformazione dei dati attraverso gli strumenti integrati di Tableau. La colonna del dataset denominata "Durata in Ms" contenente la durata di ogni traccia in millisecondi, è stata utilizzata come punto di partenza per il calcolo di un nuovo campo denominato "Durata". Questo calcolo è stato effettuato per convertire la durata dall'unità di misura iniziale (millisecondi) al formato di visualizzazione "ore:minuti:secondi". Tali manipolazioni sono state eseguite mediante apposite query direttamente all'interno dell'ambiente di Tableau, attraverso lo specifico linguaggio di calcolo e manipolazione dei dati noto come "Tableau Calculation Language" (TCL).

Le query eseguite sono le seguenti:

```
1 INT([Duration ms] / 60000)
```

Listato 3.1: Prima query eseguita su Tableau.

```
1 ROUND(([Duration ms] % 60000) / 1000, 3)
```

Listato 3.2: Seconda query eseguita su Tableau.

```
1 MAKETIME(0, [Minuti], INT([Secondi]))
```

Listato 3.3: Terza query eseguita su Tableau.

La prima query è stata utilizzata per creare un campo denominato "Minuti" contenente i valori dei minuti di ogni traccia mentre la seconda query per produrre un campo "Secondi"

Tabelle

#	Acustica
#	Chiave
⌚	Data uscita album
⌚⌚	Durata
Abc	Energia
T F	Is Explicit
Abc	Loquacità
Abc	Loudness
Abc	Nome album
Abc	Nome artista
Abc	Nome traccia
🌐	Paese
⌚	Snapshot Date
Abc	Spotify Id
Abc	Strumentalità
Abc	Valenza
Abc	Vitalità
Abc	<i>Nomi misure</i>
⌘	Danzabilità
=⌘	Durata in minuti
⌘	Durata in Ms
⌘	Indicazione del tempo
=⌘	Minuti
⌘	Modalità
⌘	Movimento giornaliero
⌘	Movimento settimanale
⌘	Popolarità
⌘	Posizione
=⌘	Secondi
⌘	Tempo
🌐	<i>Latitudine (generata)</i>
🌐	<i>Longitudine (generata)</i>
⌘	<i>universal_top_spotify_son...</i>
⌘	<i>Valori misure</i>

Figura 3.1: Campi presenti nel dataset scelto.

contenente i secondi di ogni traccia. Eseguendo la terza query, invece, è stato possibile creare il campo di nome "Durata" in modo tale da contenere dati del formato "*ore:minuti:secondi*" relativi ad ogni traccia musicale.

Questo passaggio è cruciale per garantire che i dati siano pronti per l'analisi, affrontando eventuali dati mancanti o modificando tipi di dati secondo necessità.

Approfondendo la discussione sulle dashboard create attraverso la piattaforma Tableau, si delineano come un passo significativo nell'analisi e nella presentazione dei dati riguardanti le tracce musicali più popolari in 73 paesi diversi durante un periodo specifico. Grazie a un'ampia varietà di grafici, mappe e filtri interattivi, queste dashboard contribuiscono a rendere i dati più accessibili e comprensibili, offrendo agli utenti la possibilità di esplorare le tendenze in

modo intuitivo.

Le dashboard realizzate con il suddetto software sono le seguenti:

- Dashboard 1: Analisi degli artisti che hanno prodotto brani con una notevole componente acustica, esplorando la correlazione tra questa caratteristica musicale e i Paesi in cui tali brani sono ascoltati;
- Dashboard 2: Analisi della relazione tra popolarità e danzabilità dei brani più popolari a livello globale nell'ultimo mese;
- Dashboard 3: Analisi del daily movement dei top 6 brani in Spagna dall'1 al 4 Gennaio 2024;
- Dashboard 4: Analisi e previsione dell'andamento di "All I Want For Christmas";

3.3 Dashboard 1: Analisi dei brani con una notevole componente acustica

La prima analisi svolta riguarda l'approfondimento degli artisti che hanno composto brani con un elevato tasso di acustica, esaminando la relazione tra questa caratteristica musicale e i paesi in cui tali opere sono apprezzate. La dashboard presentata è una rappresentazione visiva composta da due sezioni principali che forniscono un'analisi comparativa del tasso di acustica a livello globale: una mappa geografica e un grafico a bolle.

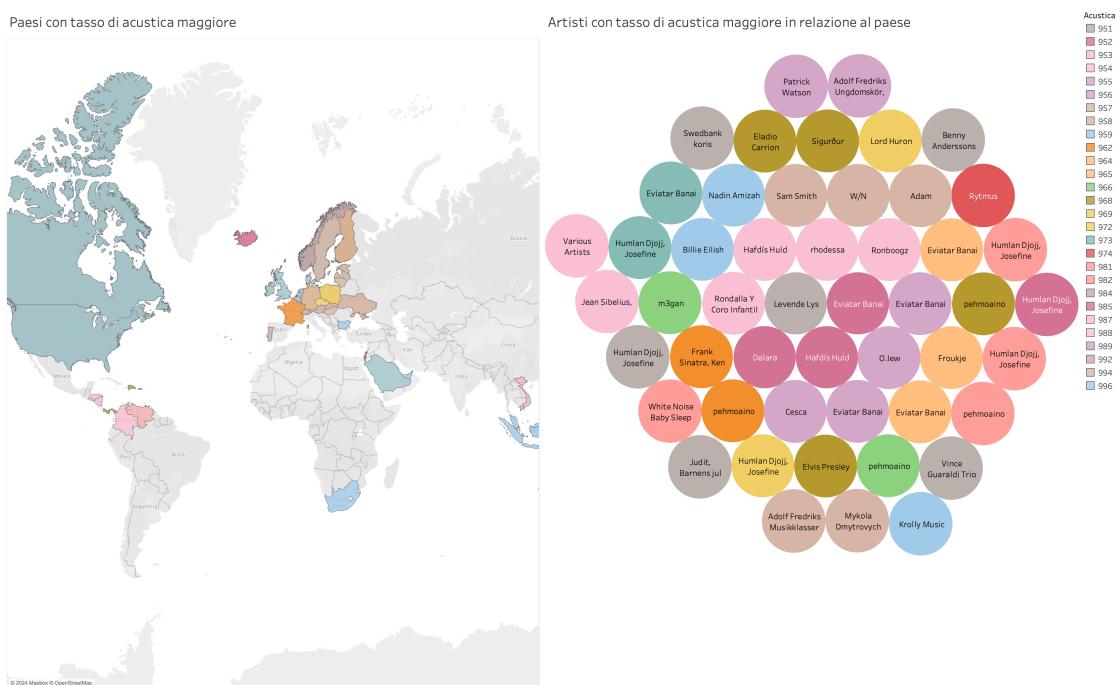


Figura 3.2: Dashboard 1

Partendo da sinistra, la prima sezione è una mappa geografica, intitolata "Paesi con tasso di acustica maggiore", che utilizza una scala cromatica per rappresentare l'intensità del tasso di acustica tra i diversi paesi. I colori variano dal grigio chiaro (tasso più basso) all'azzurro chiaro (tasso più alto), permettendo un immediato riconoscimento visivo delle regioni con maggiore prevalenza di musica acustica.

A destra si trova, invece, la seconda sezione che è un grafico a bolle denominato "Artisti con tasso di acustica maggiore in relazione al Paese", il quale mette in relazione gli artisti musicali con il loro rispettivo tasso di acustica e ogni artista presenta lo stesso colore del Paese in cui i suoi brani sono ascoltati. Inoltre ogni artista è rappresentato da una bolla

il cui diametro è proporzionale alla grandezza del dato. Il grafico utilizza una varietà di colori vivaci per differenziare gli artisti, rendendo quest'ultimo non solo informativo ma anche esteticamente piacevole.

Come si può notare nelle Figure 3.3 e 3.4, i tre artisti con tasso di acustica maggiore sono etichettati direttamente sul grafico in azzurro chiaro e in marrone. Essi sono *Krolly Music*, ascoltato in Bulgaria e, a pari merito, *Adolf Fredriks Musikklasser* e *Mykola Dmytryovych*, ascoltati in Canada, Irlanda, Germania, Svizzera, Austria, Svezia, Polonia e Ucraina.



Figura 3.3: Krolly Music

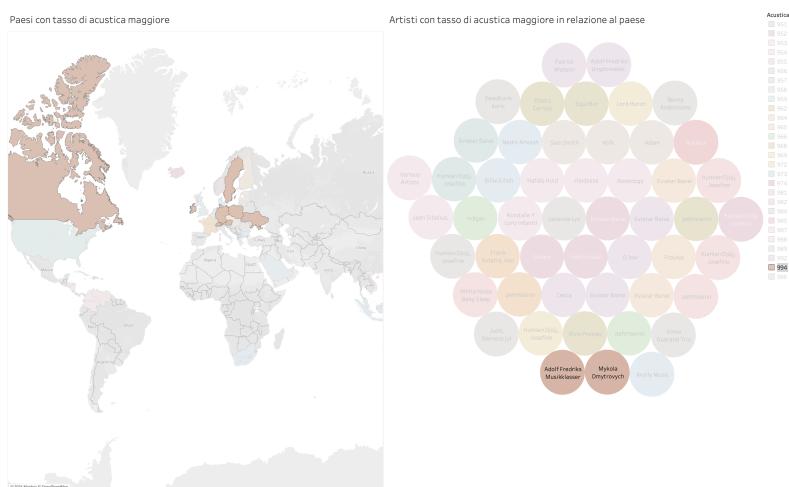


Figura 3.4: Adolf Fredriks Musikklasser e Mykola Dmytryovych

3.3.1 Considerazioni finali

L'integrazione delle due visualizzazioni offre un rapido sguardo sulla distribuzione geografica della musica acustica e sugli artisti che la caratterizzano. Quest'analisi rivela informazioni cruciali per progetti di business intelligence mirati a identificare tendenze di mercato e opportunità di targeting geografico basate sulle preferenze musicali. Entrambi i grafici, complementari tra loro, forniscono dati preziosi per strategie di marketing mirate. Ad esempio, le aziende potrebbero utilizzare queste informazioni per pianificare campagne promozionali in specifici paesi in base agli artisti localmente popolari o per sviluppare prodotti legati alla

musica acustica. Inoltre, l'analisi evidenzia regioni in cui la presenza della musica acustica è limitata, suggerendo opportunità di crescita o mercati ancora inesplorati.

3.4 Dashboard 2: Analisi della popolarità e danzabilità di un brano

La seconda analisi effettuata su Tableau risulta molto interessante in quanto si presuppone di studiare se vi è una relazione tra la popolarità di una canzone e il suo tasso di danzabilità, prendendo in considerazione il periodo che va dal 24/12/2023 al 24/01/2024.

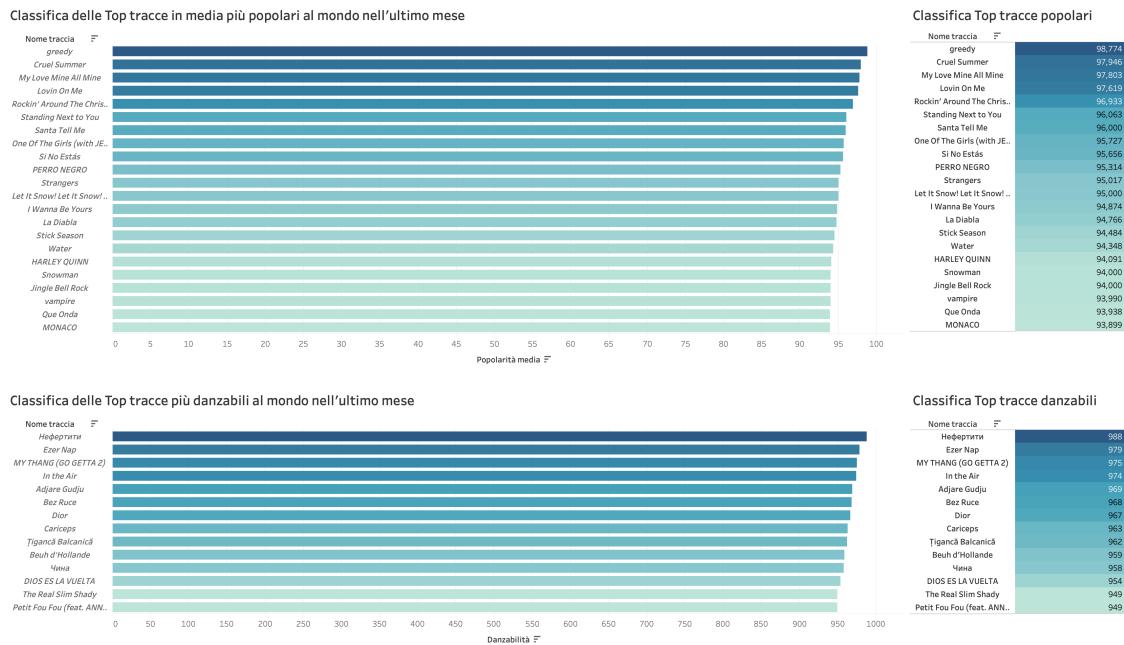


Figura 3.5: Dashboard 2

La Figura 3.5 illustra due grafici a barre orizzontali generati da Tableau, strumento di visualizzazione dati, che classificano le tracce musicali in base a diversi attributi.

Il grafico in alto, intitolato "Classifica delle Top tracce in media più popolari al mondo nell'ultimo mese", mostra le canzoni più popolari nell'ultimo mese. Ogni barra rappresenta una canzone, con il nome della traccia sull'asse delle ordinate e la media di popolarità sull'asse delle ascisse, che varia da 0 a 100. La popolarità è indicata con barre colorate, dove una tonalità più intensa di blu corrisponde a un valore più elevato di popolarità, una tonalità più chiara, corrispondente al verde acqua, si riferisce ad un valore più basso di popolarità. Ad esempio, la canzone "greedy" è raffigurata in blu ed ha la popolarità più alta con un valore approssimativo di 98,774, mentre il brano "Monaco" è colorato in verde acqua ed è il brano meno popolare tra quelli presi in considerazione, con un tasso di popolarità medio pari a 93,899.

Il grafico in basso, invece, denominato "Classifica delle Top tracce più danzabili al mondo nell'ultimo mese", ordina le canzoni in base alla loro danzabilità. Allo stesso modo, l'asse delle ordinate elenca i nomi delle tracce, mentre l'asse delle ascisse mostra il valore di danzabilità, che in questo caso varia da 0 a 1000. Le barre, di nuovo, utilizzano varie tonalità, dal blu al verde acqua, per indicare i livelli di danzabilità, con tonalità più scure che rappresentano valori superiori. Tra i brani presi in considerazione, la traccia più danzabile spicca con un punteggio prossimo a 988, mentre quella meno danzabile ha un punteggio pari a 949.

Al fianco di questi due grafici, vengono mostrati rispettivamente la classifica delle tracce più popolari con i rispettivi indici di popolarità e la classifica delle tracce più danzabili con i rispettivi indici di danzabilità. Entrambi i grafici seguono una metodologia simile ai due grafici precedenti: con un'asse delle ordinate per i nomi delle tracce e lo stesso schema di colorazione e disposizione.

3.4.1 Considerazioni finali

In conclusione, si nota come non vi sia una correlazione evidente tra canzoni più popolari e canzoni più ballabili, evidenziando la diversità di preferenze nel pubblico musicale. Mentre alcune tracce guadagnano popolarità grazie a caratteristiche musicali diverse, come testi coinvolgenti o melodie accattivanti, altre possono primeggiare per la loro capacità di far ballare le persone. Questa varietà sottolinea la complessità del panorama musicale e la necessità di considerare diversi fattori per comprendere a pieno le preferenze dell'ascoltatore.

Inoltre, questi grafici forniscono un'analisi dettagliata delle tendenze musicali basate su popolarità e danzabilità, offrendo insights preziosi per le strategie di marketing nel settore dell'intrattenimento. I dati potrebbero essere utilizzati per identificare potenziali successi nelle playlist, per strategie di promozione nelle piattaforme di streaming o per organizzare eventi e festival che rispondano alle preferenze del pubblico globale. Questa approfondita analisi potrebbe contribuire a ottimizzare le decisioni nell'industria musicale, garantendo un allineamento più preciso con le aspettative e le tendenze del pubblico.

3.5 Dashboard 3: Analisi del daily movement dei top 6 brani in Spagna

Il documento presenta una serie di grafici interattivi concepiti per esaminare in dettaglio il comportamento dei brani musicali nel mercato spagnolo nel periodo compreso tra l'1 e il 4 gennaio 2024. Il fulcro di questa visualizzazione è rappresentato dal grafico principale, un diagramma lineare denominato "Andamento dei Top 6 brani in Spagna dal 01/01/2024 al 04/01/2024".

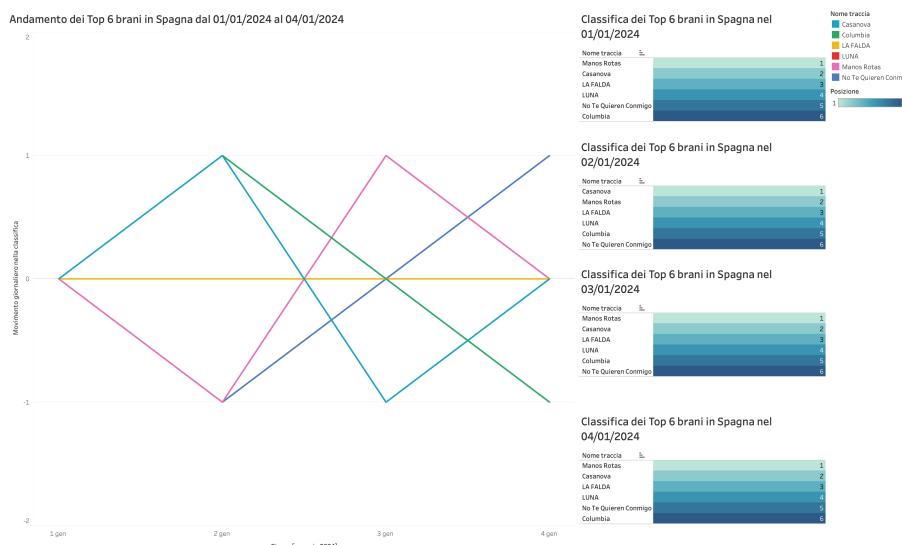


Figura 3.6: Dashboard 3

Questo grafico delinea le fluttuazioni quotidiane nella classifica di sei brani distinti, con linee colorate che tracciano la traiettoria di ciascun brano su un asse delle ordinate. Tale asse

indica il movimento giornaliero, positivo o negativo, rispetto alla posizione occupata il giorno precedente.

Accanto a questo grafico principale, sono collocati quattro grafici a barre orizzontali che dettagliano la classifica giornaliera dei Top 6 brani per ciascuno dei giorni considerati. Le barre, variamente colorate dal verde acqua al blu scuro, riflettono la posizione di ciascun brano in classifica, contribuendo a una visualizzazione immediata delle dinamiche di classifica nel periodo specifico. Infatti, i quattro grafici a barre orizzontali, sono stati inseriti in modo tale da dimostrare le variazioni dell'andamento dei brani presi in considerazione nella classifica giornaliera dall'1 al 4 gennaio 2024 in Spagna.

Nella Figura 3.7 è evidenziato come, ad esempio, il brano "Manos Rotas" appare, nei giorni considerati, rispettivamente primo, secondo, primo e primo in classifica con un daily movement pari a "-1" nel 2 Gennaio 2024, pari a "+1" nel 3 Gennaio 2024 e pari a "0" nel 4 Gennaio 2024.



Figura 3.7: Andamento del brano denominato "Manos Rotas"

Riguardo all'andamento del brano denominato "Casanova" (Figura 3.9), il quale mostra un daily movement pari a "+1" nel 2 Gennaio 2024, a "-1" nel 3 Gennaio 2024 e a "0" nel 4 Gennaio 2024, come si può ben notare, il brano occupa la seconda, prima, seconda e di nuovo la seconda posizione.

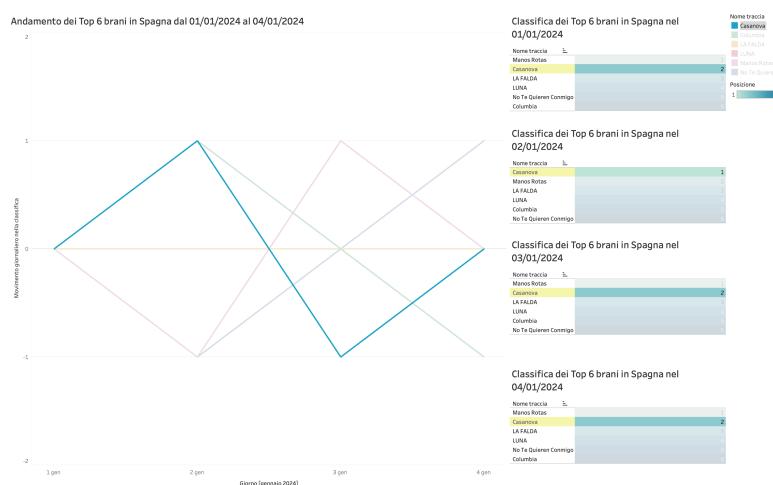


Figura 3.8: Andamento del brano denominato "Casanova"

L'ultimo esempio di andamento di un brano inserito è quello di "LA FALDA" che mostra un andamento lineare del daily movement in quanto, nel periodo di riferimento, rimane sempre in terza posizione nella classifica delle canzoni più popolari in Spagna.

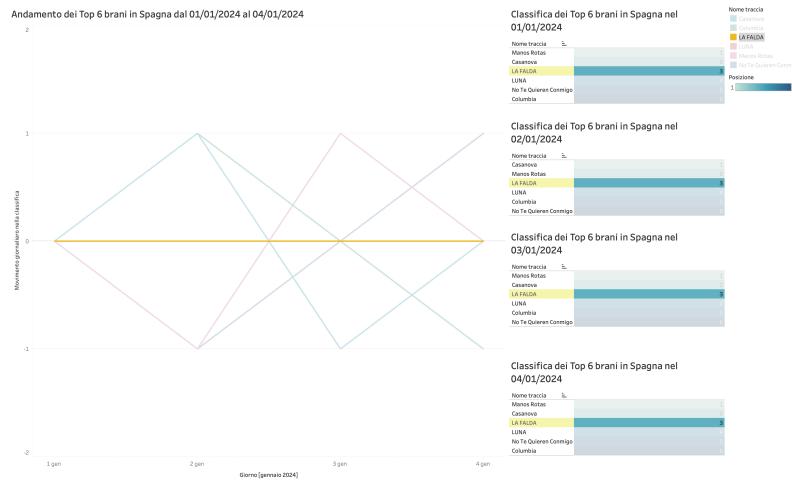


Figura 3.9: Andamento del brano denominato "LA FALDA"

3.5.1 Considerazioni finali

Nell'analisi specifica condotta sui brani "Manos Rotas", "Casanova" e "LA FALDA", emergono variazioni di rilievo nelle loro posizioni giornaliere all'interno della classifica spagnola. Questi cambiamenti evidenziano la fluidità e la competitività dell'ambiente musicale durante il periodo preso in esame, sottolineando la mutevolezza delle preferenze dell'ascolto e l'impatto dinamico degli utenti sulla disposizione delle tracce nella classifica.

"Manos Rotas", ad esempio, attraversa una sequenza di posizioni, oscillando tra il primo, il secondo e il primo posto nei giorni considerati. Questo pattern di daily movement, con valori come "-1", "+1", e "0", suggerisce una risposta variabile da parte del pubblico, che può essere influenzata da diversi fattori, come nuovi rilasci, eventi o tendenze culturali.

Il brano "Casanova" mostra un andamento simile, evidenziando un dinamismo nella sua posizione in classifica. Con daily movement quali "+1", "-1", e "0", il brano mantiene una presenza significativa, ma le fluttuazioni giornaliere indicano una competizione attiva con altri brani per ottenere la massima visibilità e preferenza da parte degli ascoltatori.

Per quanto riguarda "LA FALDA", la sua costante permanenza in terza posizione nel periodo analizzato sottolinea una stabilità relativa rispetto agli altri brani presi in considerazione. La linearità del daily movement indica una posizione consolidata nella classifica, con possibili implicazioni sulla sua popolarità o sulla fedeltà dell'audience.

In conclusione, la Dashboard 3 si rivela un potente strumento di analisi per comprendere non solo le variazioni quotidiane nelle posizioni dei brani, ma anche le dinamiche complesse del mercato musicale spagnolo. La sua interattività fornisce agli utenti una visione dettagliata delle tendenze e delle opportunità emergenti, contribuendo così a informare decisioni strategiche nel settore della gestione e promozione della musica in Spagna.

3.6 Dashboard 4: Analisi e previsione dell'andamento di un brano

La quarta analisi condotta attraverso Tableau si concentra sull'andamento medio giornaliero e sulla posizione media della canzone "All I Want for Christmas Is You". La dashboard presenta analisi di dati, con particolare enfasi sul grafico a sinistra, nel periodo compreso tra il 28

ottobre 2023, quando la traccia fa la sua entrata nelle Top 50, e il 16 gennaio 2024. Nel contempo, il grafico a destra si estende ulteriormente, coprendo un periodo più prolungato, dal 28 ottobre 2023 al 1 febbraio 2024. Al fine di anticipare e comprendere le dinamiche in evoluzione, sono state effettuate previsioni sull'andamento futuro della canzone, sfruttando l'efficace meccanismo di previsione di Tableau. Di seguito, è allegata un'immagine che include la suddetta dashboard per una visione più chiara e dettagliata.

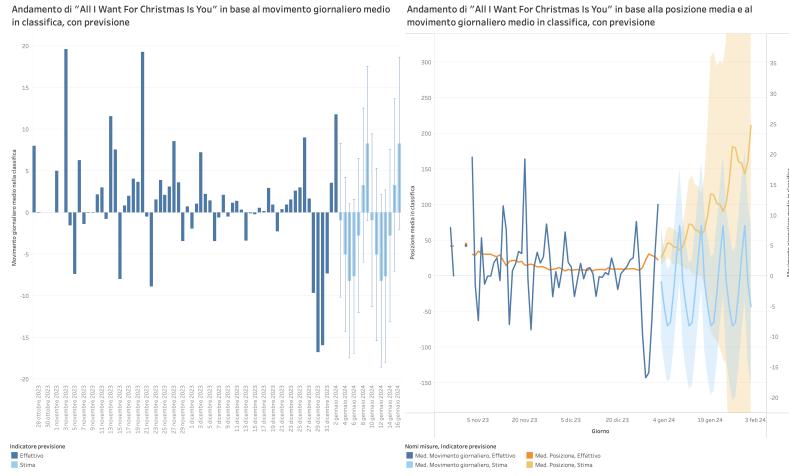


Figura 3.10: Andamento di "All I Want For Christmas Is You" con previsione dello stesso

Essa è composta da due grafici:

- Il grafico a barre sulla sinistra offre una panoramica del movimento giornaliero medio della canzone dal 28 ottobre 2023 al 16 gennaio 2024, evidenziando le fluttuazioni nella sua posizione giornaliera in classifica. Il colore delle barre distingue tra i valori effettivi dei dati (in blu scuro) e le stime (in azzurro chiaro), fornendo inoltre degli intervalli di confidenza, i quali indicano il livello di incertezza associato alle previsioni. L'asse delle ordinate indica il numero di posizioni guadagnate o perse, mentre l'asse delle ascisse mostra le date in cui sono state effettuate queste analisi. Le barre ascendenti indicano un miglioramento nella posizione, mentre quelle discendenti segnalano un declino. Emerge un andamento distintivo e ciclico (Figura 3.11): ad esempio, nel corso del mese di novembre 2023, si osserva un aumento di posizione ogni lunedì, specificamente il 6, 13, 20 e 27.

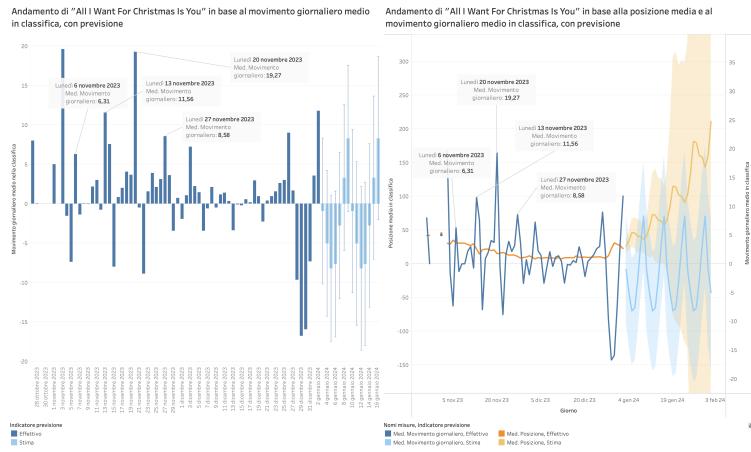


Figura 3.11: Andamento di "All I Want For Christmas" nel mese di Novembre 2023

Nel grafico a doppia linea sulla destra della stessa figura, viene documentato come, all'inizio di ogni settimana nel mese di novembre 2023, si verifichi un picco di aumento del movimento medio giornaliero, associato a un calo della posizione media nella classifica giornaliera. Questo fenomeno evidenzia una correlazione diretta: all'aumentare del movimento giornaliero, la canzone tende a posizionarsi più in basso nella classifica giornaliera, acquisendo così posizioni di maggior rilievo.

- Il grafico a doppie linee a destra, presenta appunto due linee che tracciano sia il rank medio giornaliero (linea in arancione scuro) che il movimento medio giornaliero della canzone (linea in blu scuro), con delle conseguenti previsioni rispettivamente in arancione chiaro e in azzurro chiaro.

Il grafico ha un aspetto più fluido e mostra anche intervalli di confidenza (le aree colorate), indicando il livello di incertezza associato alle previsioni. L'asse delle ordinate di sinistra rappresenta la posizione giornaliera media della canzone, mentre l'asse delle ordinate di destra indica il movimento medio giornaliero in termini di posizioni nella classifica. L'asse delle ascisse mostra le date, che coprono i mesi invernali e includono il periodo festivo, quando ci si aspetterebbe che una canzone natalizia come "All I Want for Christmas Is You" abbia un picco di popolarità.

Nella Figura 3.12, si osserva l'andamento del menzionato brano nei giorni conclusivi di dicembre 2023, in prossimità delle festività natalizie. Dopo un notevole picco del movimento medio giornaliero nella classifica delle canzoni più popolari, registrato il 26 dicembre 2023, si evidenzia come, due giorni successivi, cioè il 28 dicembre 2023, il brano subisca una brusca perdita di posizioni, quantificata in una media di "-9,66". Tale declino raggiunge il punto minimo il 29 dicembre 2023, con una media di "-16,74" posizioni perse giornalmente.

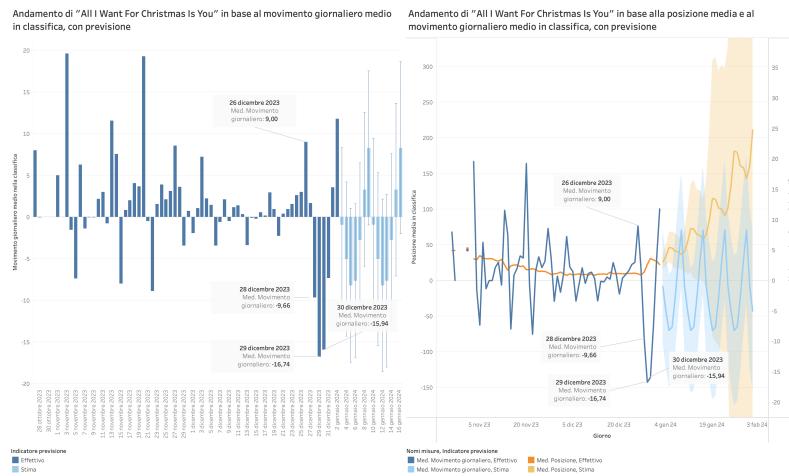


Figura 3.12: Andamento di "All I Want For Christmas" nel mese di Dicembre 2023

Come già ampiamente evidenziato, esiste una chiara correlazione tra il movimento giornaliero medio e la posizione media giornaliera di ciascun brano in classifica. Questo legame diventa particolarmente evidente nei giorni successivi alle festività natalizie, dove un picco negativo nel movimento medio giornaliero corrisponde a un picco positivo nel daily rank. Tale relazione conferma che, nel periodo post-natalizio, la canzone "All I Want For Christmas Is You" subisce una perdita di posizioni significativa nella classifica giornaliera.

3.6.1 Stime e previsioni sull'andamento

Tableau offre anche la possibilità di fare stime e previsioni: la capacità di generare stime con Tableau deriva dalla sua integrazione di strumenti di analisi avanzati e dalla facilità con cui gli utenti possono manipolare e interpretare i dati, consentendo a quest'ultimi di ottenere una comprensione più profonda dei trend e delle dinamiche presenti nei dati. Dalla visualizzazione delle tendenze storiche alla previsione di risultati futuri, Tableau offre un ambiente dinamico per esplorare e analizzare dati in modo innovativo e predittivo.

La quarta Dashboard contiene previsioni sull'andamento del movimento medio giornaliero, in particolare nel primo grafico si ha una previsione delle successive 2 settimane, ossia dal 3 al 16 Gennaio 2024 e sull'andamento della posizione media in classifica, previsto nel secondo grafico dal 3 Gennaio 2024 all'1 Febbraio 2024.

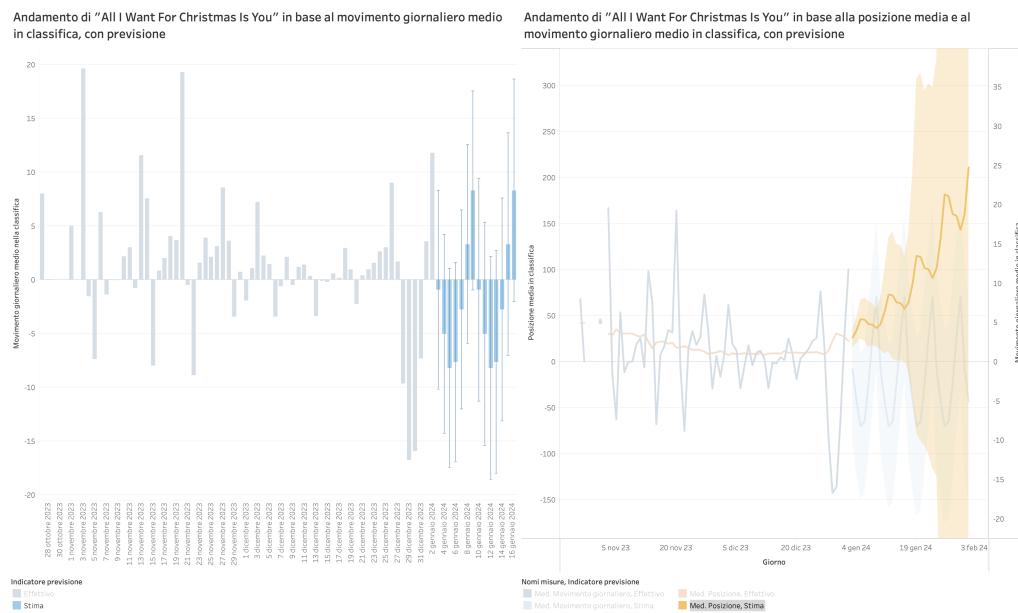


Figura 3.13: Previsioni effettuate con Tableau

Le previsioni sul daily movement fino al 16 Gennaio 2024 sono riassunte nella tabella 3.1:

Data	Valore effettivo	Stima
3 Gennaio 2024	-12,50	-0,95
4 Gennaio 2024	/	-5,03
5 Gennaio 2024	/	-8,22
6 Gennaio 2024	/	-7,63
7 Gennaio 2024	/	-2,75
8 Gennaio 2024	/	3,31
9 Gennaio 2024	/	8,32
10 Gennaio 2024	/	-0,95
11 Gennaio 2024	/	-5,03
12 Gennaio 2024	/	-8,22
13 Gennaio 2024	/	-7,63
14 Gennaio 2024	/	-2,75
15 Gennaio 2024	/	3,31
16 Gennaio 2024	/	8,32

Tabella 3.1: Previsioni sul daily movement del brano dal 3 al 16 Gennaio 2024

Le previsioni sull'andamento del movimento medio giornaliero, riportate nella Tabella 3.1, presentano una panoramica dettagliata delle stime previste per il periodo dal 3 al 16 gennaio 2024. I principali punti di interesse sono i seguenti:

- Le stime effettuate forniscono una visione della possibile variazione del movimento medio giornaliero della canzone "All I Want for Christmas Is You". Il confronto con il valore effettivo del 3 gennaio 2024 ("−12,50") indica una previsione di andamento negativo iniziale. Tuttavia, è importante notare che i valori effettivi mancanti per i giorni successivi sono attribuibili a dati non ancora disponibili.
- Le stime si estendono su un ampio spettro, con previsioni che variano da "−8,22" a "8,32". Questo suggerisce una certa incertezza nelle previsioni, come indicato dagli intervalli di confidenza. L'andamento previsto mostra una alternanza tra giorni con previsioni di aumento e giorni con previsioni di diminuzione del movimento medio giornaliero.
- Le limitazioni inferiori e superiori, nel contesto delle previsioni per il movimento medio giornaliero, costituiscono margini di errore che delimitano la gamma di possibili valori previsti. La limitazione inferiore, che varia da "−20,48" a "−0,93", rappresenta il limite inferiore entro il quale ci si aspetta che il movimento medio giornaliero possa cadere. In altre parole, questo intervallo indica il valore più basso plausibile previsto per il movimento medio giornaliero.

Dall'altro lato, la limitazione superiore, con un intervallo da "1,03" a "20,58", indica il limite superiore al di sopra del quale ci si aspetta che il movimento medio giornaliero non superi. Questo rappresenta il massimo valore previsto per il movimento medio giornaliero. Questi intervalli di confidenza forniscono una valutazione della variabilità associata alle stime, offrendo un quadro chiaro dei possibili scenari in cui il movimento medio giornaliero potrebbe evolversi. La presenza di questi margini di errore è cruciale per comprendere la robustezza delle previsioni.

- Va sottolineato che l'incertezza nelle previsioni potrebbe essere attribuibile alle dimensioni limitate del dataset disponibile. In presenza di un dataset più ampio e rappresentativo, le previsioni potrebbero essere più accurate e affidabili. Complessivamente, le previsioni consentono di anticipare e comprendere le dinamiche dell'andamento della canzone nel periodo considerato, integrando efficacemente gli strumenti di analisi avanzati offerti da Tableau.

Le previsioni sull'andamento della posizione media giornaliera in classifica riportate nella Tabella 3.2, presentano una panoramica dettagliata delle stime previste per il periodo dal 3 Gennaio 2024 all'1 Febbraio 2024. Le limitazioni inferiori e superiori per la previsione della media della posizione in classifica forniscono informazioni cruciali sulla variabilità dei valori previsti. L'intervallo di previsione più basso, che varia da "−759,5" a "25,0", indica la gamma in cui ci si aspetta che la media della posizione possa variare. Il valore inferiore di "−759,5" rappresenta il limite più basso plausibile previsto per la posizione media.

Contemporaneamente, l'intervallo di stima superiore, compreso tra "36" e "1.080", indica il limite superiore al di sopra del quale ci si aspetta che la posizione media non superi. Il valore massimo di "1.080" rappresenta il limite massimo previsto per la posizione media giornaliera.

Questi intervalli di confidenza sono fondamentali per valutare la precisione delle previsioni, fornendo una visione completa della variabilità associata alla posizione media prevista. Consentono di comprendere i possibili scenari di fluttuazione della posizione media, contribuendo così a prendere decisioni informate e a interpretare correttamente i risultati delle analisi previsionali.

Data	Valore effettivo	Stima
3 Gennaio 2024	27,0	25,8
4 Gennaio 2024	/	34,0
5 Gennaio 2024	/	46,2
6 Gennaio 2024	/	45,6
7 Gennaio 2024	/	40,8
8 Gennaio 2024	/	40,3
9 Gennaio 2024	/	36,4
10 Gennaio 2024	/	40,8
11 Gennaio 2024	/	53,7
12 Gennaio 2024	/	72,8
13 Gennaio 2024	/	72,1
14 Gennaio 2024	/	64,3
15 Gennaio 2024	/	63,5
16 Gennaio 2024	/	57,4
17 Gennaio 2024	/	64,4
18 Gennaio 2024	/	84,7
19 Gennaio 2024	/	115,0
20 Gennaio 2024	/	113,7
21 Gennaio 2024	/	101,6
22 Gennaio 2024	/	100,3
23 Gennaio 2024	/	90,6
24 Gennaio 2024	/	101,6
25 Gennaio 2024	/	133,8
26 Gennaio 2024	/	181,5
27 Gennaio 2024	/	179,5
28 Gennaio 2024	/	160,3
29 Gennaio 2024	/	158,3
30 Gennaio 2024	/	143,0
31 Gennaio 2024	/	160,3
1 febbraio 2024	/	211,1

Tabella 3.2: Previsioni sul daily rank del brano dal 3 Gennaio 2024 all'1 Febbraio 2024

Durante l'intervallo temporale preso in considerazione, emergono alcune tendenze degne di nota: i dati rivelano una variazione nei valori previsti rispetto ai valori effettivi. Ad esempio, il 3 Gennaio 2024 presenta un valore effettivo di "27,0", mentre la stima è di "25,8". Nel corso delle previsioni, si notano oscillazioni significative nella posizione media giornaliera, evidenziando la dinamicità del posizionamento del brano in classifica. Nel 3 Gennaio quindi, la stima supera il valore effettivo, indicando una previsione più positiva rispetto alla realtà osservata.

È interessante notare che, verso la fine del periodo di previsione, le stime suggeriscono un aumento significativo della posizione media giornaliera, raggiungendo valori considerevolmente più alti rispetto all'inizio del periodo preso in esame. Addirittura, dall'11 Gennaio 2024 in poi, si prevede che, in media, il brano esca dalla classifica delle 50 canzoni più popolari a livello globale.

Complessivamente, queste previsioni forniscono una panoramica delle possibili dinamiche del posizionamento in classifica del brano nel periodo indicato, sottolineando l'importanza di monitorare attentamente l'effettivo andamento per adattare strategie e decisioni di conseguenza.

3.6.2 Qualità delle previsioni

Le suddette previsioni sono state eseguite attraverso la scelta di un intervallo di confidenza pari al 90% e ignorando l'ultimo giorno in cui i dati sono presenti. Inoltre sono stati adottati dei modelli di previsione automatici per i campi su cui è stata effettuata la stima, ossia sono stati scelti dei modelli che Tableau reputa i migliori di tutti.

Tutte le previsioni sono state calcolate utilizzando un livellamento esponenziale.

Med. Movimento giornaliero

Livello	Modello			Metriche della qualità					Coeffienti di livellamento		
	Trend	Stagione		RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Aggiuntivo	Nessuno	Aggiuntivo		5,62	3,99	0,96	166,4%	241	0,008	0,000	0,500

Med. Posizione

Livello	Modello			Metriche della qualità					Coeffienti di livellamento		
	Trend	Stagione		RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Moltiplicativo	Moltiplicativo	Moltiplicativo		3,4	2,1	1,03	12,8%	180	0,500	0,073	0,340

Figura 3.14: Modello di previsione utilizzato da Tableau

Tutte le previsioni sono state calcolate utilizzando un livellamento esponenziale. Come mostrato in Figura 3.14, i dati forniti da Tableau per la previsione del movimento medio giornaliero indicano l'utilizzo di un modello che incorpora una componente di livello, una componente stagionale aggiuntiva e nessuna componente di trend. Le metriche di qualità del modello includono il Root Mean Squared Error (RMSE) pari a 5,62, il Mean Absolute Error (MAE) di 3,99 e il Mean Absolute Scaled Error (MASE) di 0,96. Queste metriche forniscono una valutazione della precisione del modello, con un RMSE più basso e un MASE più vicino a 1 indicativi di una previsione più accurata. Il Mean Absolute Percentage Error (MAPE) è del 166,4%, indicando l'errore percentuale medio tra i valori previsti e quelli effettivi. I coefficienti di livellamento associati al modello includono Alpha (α) pari a 0,008, Beta (β) pari a 0,000 e Gamma (γ) pari a 0,500, che influenzano il modo in cui il modello tiene conto del livello, del trend e della stagionalità. Questi dati offrono una panoramica della metodologia di previsione e della sua precisione: complessivamente, il modello potrebbe beneficiare di miglioramenti, specialmente considerando l'alta percentuale di errore.

Per quanto riguarda la previsione sulla posizione media giornaliera, i dati forniti da Tableau indicano l'uso di un modello moltiplicativo con componenti moltiplicative per livello, trend e stagionalità. Le metriche di qualità del modello comprendono un RMSE di 3,4 e un MAE di 2,1, indicando una certa precisione nelle previsioni. Il MASE è di 1,03, suggerendo una buona performance rispetto a un modello di base. Il MAPE è del 12,8%, indicando un errore percentuale medio relativamente basso. I coefficienti di livellamento associati al modello includono Alpha (α) pari a 0,500, Beta (β) pari a 0,073 e Gamma (γ) pari a 0,340, influenzando il modo in cui il modello tiene conto del livello, del trend e della stagionalità. Complessivamente, questi dati indicano che il modello moltiplicativo è in grado di fornire previsioni con un livello accettabile di precisione per la posizione media.

Poiché non tutte le previsioni effettuate su Tableau mostrano un livello di precisione uniforme, sarà necessario un ulteriore studio e analisi dettagliata nel capitolo successivo, utilizzando la piattaforma Power BI. Questo approccio consentirà una valutazione più approfondita delle previsioni, con l'obiettivo di migliorare la precisione e fornire una comprensione più completa dell'andamento del brano "All I Want For Christmas Is You" nella classifica musicale nel periodo considerato.

Power BI



4. Power BI

4.1 Introduzione a Power BI

Power BI è una suite di strumenti di business analytics sviluppata da Microsoft che consente agli utenti di connettersi a una vasta gamma di origini di dati, di visualizzare i dati in modi interattivi e di condividere le visualizzazioni con gli altri. Inoltre, Power BI offre diverse funzionalità per l'analisi e la presentazione dei dati, rendendo più facile per le organizzazioni trarre informazioni significative dalle loro fonti di dati.

L'utilizzo di una suite di strumenti di business analytics come Power BI offre molti vantaggi, i quali, come si evince da uno studio effettuato da Gartner in Figura 4.1 (società di ricerca e consulenza, con sede in America, che ogni anno pubblica il report Gartner Magic Quadrant) in data 2019, lo rendono il primo tra i "Leaders" delle piattaforme utilizzate per l'analisi e la presentazione dei dati, confermando la sua supremazia su piattaforme come "Qlik" e "Tableau" descritte ed utilizzate precedentemente.



Figura 4.1: Magic quadrant di Gartner relativo alle piattaforme di Business Intelligence

4.2 Caricamento e pulizia dei dati

Una volta avviato Power BI, il primo passo è effettuare la connessione con il dataset in questione. Dunque, occorre caricare su Power BI il file CSV contenente tutti i dati da studiare, ovvero il file `universal_top_spotify_songs.csv`.

Power BI è uno strumento cruciale nelle fasi di *ETL (Extract, Transform, Load)*, consentendo di preparare e organizzare i dati per analisi più approfondite. Durante il caricamento iniziale dei dati su Power BI, sono state eseguiti diversi aggiustamenti utilizzando il tool apposito per questi compiti, *Power Query*. Inizialmente, poiché Power BI ha erroneamente sostituito le virgole che caratterizzano i numeri decimali con dei punti, è stato caricato il file CSV senza alcun tipo di distinzione delle colonne in base ad un delimitatore. Successivamente, le colonne sono state divise in base al delimitatore "virgola". Così facendo, la prima riga della tabella era caratterizzata dal nome delle colonne originali del dataset, mentre l'intestazione di ogni colonna è stata ridenominata automaticamente da Power BI.

Proseguendo, al fine di evitare il problema di punteggiatura descritto in precedenza, tramite il tool di sostituzione di Power BI, sono stati sostituiti tutti i punti con le virgole, nelle sole colonne interessate. Infine, gli ultimi cambiamenti riguardano l'adeguamento dei formati delle colonne, poiché Power BI ha talvolta categorizzato erroneamente alcune colonne; ad esempio, come "stringhe" anziché "numeri decimali" e l'eliminazione di eventuali righe contenenti valori nel dataset originale.

Una volta terminata questa fase, si può chiudere Power Query e caricare il dataset con gli opportuni cambiamenti effettuati, mediante il tasto "*Chiudi e Applica*" nella sezione "*Home*" di Power Query.

In Figura 4.2 è fornito un riassunto di tutti i passaggi effettuati.

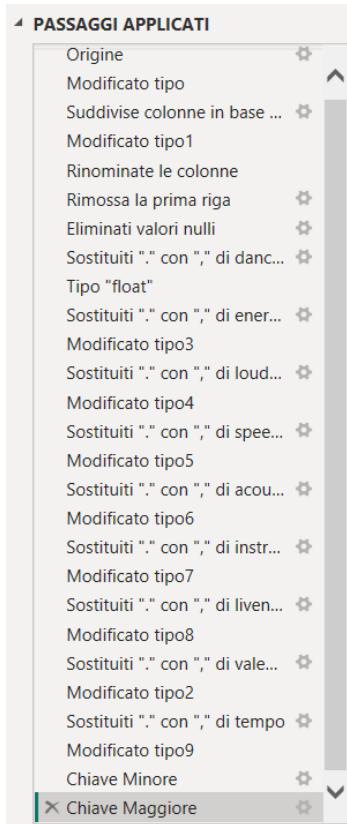


Figura 4.2: Albero dei passaggi applicati in Power Query

4.3 Data analysis e rappresentazione grafica

Per creare dashboard interattive e informative su Power BI, si utilizzano i fogli che ospitano i report. Ogni report è progettato per esaminare in dettaglio un particolare argomento e comprende un insieme di grafici che agevolano l'analisi. In termini più semplici, i fogli in Power BI sono come pagine virtuali su cui si costruiscono report contenenti grafici interattivi per esplorare e comunicare informazioni in modo efficace. I fogli che sono stati realizzati in PowerBI sono i seguenti:

- *Foglio 1*: Confronto tra "top songs";
- *Foglio 2*: Analisi degli artisti con cui effettuare un featuring;
- *Foglio 3*: Andamento giornaliero in classifica di "*All I Want For Christmas Is You*" con previsione temporale;
- *Fogli 4 e 5*: Confronto tra Paesi in Nord America.

4.3.1 Foglio 1 - Confronto tra "top songs"

La dashboard (Figura 4.3) fornisce una comparazione tra le "top songs" esplicite e non esplicite più popolari al mondo, insieme all'andamento temporale della loro posizione nelle classifiche nei mesi di Novembre e Dicembre. La dashboard è strutturata in quattro grafici principali, due bar chart e due line chart, ciascuno fornendo insights distinti sul dataset analizzato.

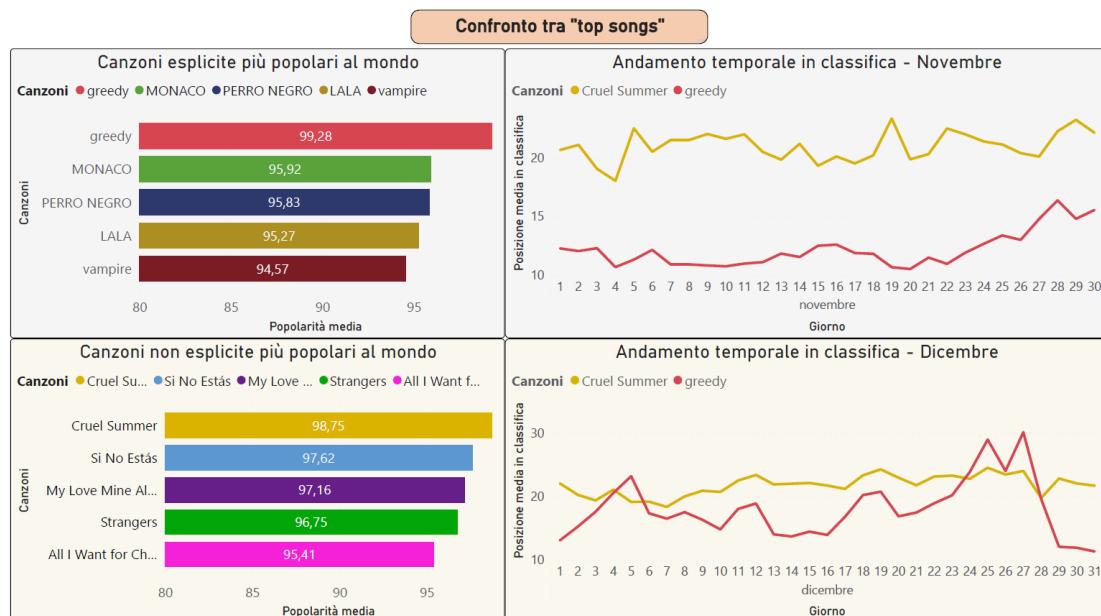


Figura 4.3: Dashboard 1

Analizzando la dashboard mostrata:

- Il primo grafico, un bar chart, illustra le "canzoni esplicite più popolari al mondo", con i valori di popolarità media rappresentati sull'asse delle ordinate. Le canzoni sono elencate sull'asse delle ascisse e colorate diversamente per una facile distinzione. Dall'analisi effettuata, risulta che la canzone "greedy" appare come la più popolare con una popolarità media vicina a **100**, precisamente **99,28**, seguita da "MONACO", "PERRO NEGRO", "LALA" e "vampire". Al fine di eseguire questa analisi e selezionare solamente le canzoni con il tasso di popolarità più alto è stato inserito un filtro che permette di visualizzare nel grafico le canzoni con una popolarità media maggiore o

uguale a **94,00**. Inoltre, l'asse X è stato ridimensionato ponendo come valore iniziale **80,00** così da mostrare con maggiore chiarezza le differenze tra le barre, dunque le canzoni, che fanno parte del grafico.

- Il secondo grafico, concettualmente uguale al primo ma per canzoni non esplicite, mostra "*Cruel Summer*" come la canzone non esplicita più popolare, con una popolarità media pari a **98,75**, seguita da "*Si No Estás*", "*My Love Mine All Mine*", "*Strangers*" e "*All I Want for Christmas Is You*". Analogamente al grafico precedente, sono stati inseriti dei filtri grazie ai quali possono essere selezionate le canzoni non esplicite più popolari al mondo. Anche in questo caso, l'asse X è stata ridimensionata per mostrare le differenze tra le popolarità delle varie canzoni.

Nella parte destra della dashboard, sono presenti due line chart che tracciano l'andamento temporale in classifica delle due canzoni più popolari, trovate come risultato delle due analisi precedenti, nei mesi di Novembre e Dicembre. Questi grafici mostrano fluttuazioni giornaliere nelle posizioni in classifica, filtrando le canzoni "*Cruel Summer*" e "*greedy*" ed evidenziandole in colori diversi per confrontare il loro rendimento nel tempo:

- Nel mese di Novembre, è interessante notare la tendenza di "*greedy*" ad aumentare la propria posizione in classifica, mentre "*Cruel Summer*" mostra un andamento più instabile. Inoltre, poiché le curve non tendono ad intersecarsi durante tutto il mese, è facile arrivare alla conclusione che, in questo periodo, le persone hanno preferito di gran lunga ascoltare "*Cruel Summer*", rappresentante delle canzoni non esplicite più popolari, rispetto a "*greedy*", canzone esplicita più popolare.
- Situazione diversa si ha nel mese di Dicembre, dove si osserva una caduta significativa in classifica di "*greedy*", che può indicare cambiamenti nelle preferenze di ascolto o in altri fattori esterni influenzanti la popolarità. Infatti, nei periodi dal 4 al 6 dicembre e dal 24 al 28 dicembre, risulta che le persone hanno preferito l'ascolto di "*Cruel Summer*", simboleggiando un aumento di interesse nell'ascolto di canzoni non esplicite rispetto a quelle esplicite.

Riassumendo, questa dashboard fornisce una visione chiara e diretta delle tendenze della musica popolare, permettendo agli utenti di identificare quali canzoni e quale genere di esse catturano l'attenzione del pubblico e come questa cambia nel tempo. Gli insights tratti da questi dati possono essere utilizzati per prendere decisioni informate nel campo del marketing musicale e per adattare le strategie di promozione a seconda della popolarità e dell'andamento delle canzoni.

4.3.2 Foglio 2 - Analisi degli artisti con cui effettuare un featuring

La dashboard (Figura 4.4) fornisce un'analisi approfondita del successo degli artisti musicali in termini di giorni trascorsi in Top 10, longevità degli album in classifica e la prevalenza di brani in Top 10, il che li rendono dei potenziali candidati per collaborazioni (featuring) in modo da creare una canzone, o album, che abbia una buona probabilità di diventare popolare e andare in tendenza.

Il layout è diviso in tre sezioni principali, ognuna con un tipo diverso di visualizzazione dei dati che, insieme, forniscono una panoramica complessiva del dataset analizzato:

- Nella prima sezione, un treemap mostra gli artisti che hanno trascorso più giorni in Top 10. Ogni artista è rappresentato da un rettangolo la cui area è proporzionale al numero di giorni trascorsi in Top 10. *Bad Bunny* domina questa classifica con **3270** giorni, seguito da Feid e altri artisti.

Per realizzare questa visualizzazione, è stata eseguita una query in linguaggio "**M**", linguaggio mashup (da qui la lettera M) progettato per creare query che combina-

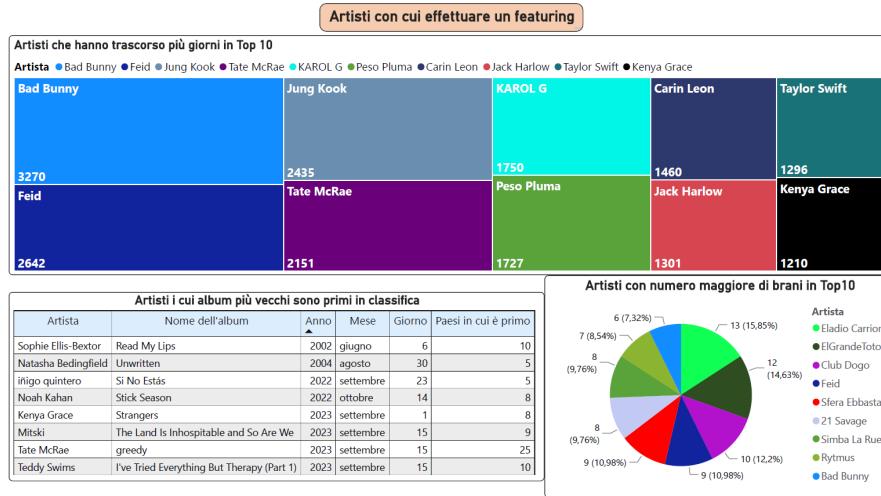


Figura 4.4: Dashboard 2

no dati, che ha permesso la creazione di un dataset contenente solamente i dati di interesse per la visualizzazione trattata, a partire dai dati del dataset originale: *"universal_top_spotify_songs"*.

Grazie ad M, è stata generata la tabella *"Giorni_Top10"* tramite la seguente query:

```

1 let
2   // Caricamento del dataset originale
3   Source = universal_top_spotify_songs,
4
5   // Espansione della colonna "artists" separando gli artisti
6   ExpandedTable = Table.ExpandListColumn(Table.TransformColumns(Source, {"artists",
7     Splitter.SplitTextByDelimiter(", "), let itemType = (type nullable text) meta
8     [Serialized.Text = true] in type {itemType}}), "artists"),
9
10  // Filtraggio delle righe con daily_rank tra 1 e 10
11  FilteredTable = Table.SelectRows(ExpandedTable, each [daily_rank] >= 1 and
12    [daily_rank] <= 10),
13
14  // Raggruppamento per artista e conta dei giorni in cui sono stati in Top 10
15  GroupedTable = Table.Group(FilteredTable, {"artists"}, {"DaysInTop10", each
16    Table.RowCount(_, type number)}),
17
17  // Ordinamento in base ai giorni in Top 10 in ordine decrescente
18  SortedTable = Table.Sort(GroupedTable, {"DaysInTop10", Order.Descending})
in
SortedTable

```

Listato 4.1: Creazione della tabella *"Giorni_Top10"*

Grazie a tale query, la tabella risultante, sulla quale sono state effettuate le analisi della visualizzazione, è quella presente in Figura 4.5.

- La seconda sezione della dashboard presenta un elenco tabellare degli artisti i cui album più vecchi sono primi in classifica, fornendo dettagli come il nome dell'artista, il titolo dell'album, l'anno di uscita, e il numero di paesi in cui l'album è primo in classifica. Questo fornisce una visione della longevità e del successo globale di un album,

	A_B_C_artists	1.2_DaysinTop10
1	Bad Bunny	3270
2	Feid	2642
3	Jung Kook	2435
4	Tate McRae	2151
5	KAROL G	1750
6	Peso Pluma	1727
7	Carin Leon	1460
8	Jack Harlow	1301
9	Taylor Swift	1296
10	Kenya Grace	1210

Figura 4.5: Tabella "Giorni_Top10"

esprimendo come gli artisti di tali album con successo in più paesi siano ancora dei "top" con cui potrebbe essere interessante effettuare delle collaborazioni.

Analogamente a quanto fatto per la prima visualizzazione, anche per questa tabella è stato realizzata una query in linguaggio M per creare la tabella "Album" che permettesse di filtrare, dal dataset originale, solamente quegli artisti i quali album più "vecchi" siano ancora in prima posizione in classifica. La query realizzata è la seguente:

```

1 let
2     // Caricamento del dataset
3     Source = universal_top_spotify_songs,
4     // Divisione degli artisti
5     SplitArtists = Table.ExpandListColumn(Table.TransformColumns(Source, {"artists",
6         => Splitter.SplitTextByDelimiter(", ", ))}, "artists"),
7     // Rimozione degli spazi extra dagli artisti
8     TrimArtists = Table.TransformColumns(SplitArtists,{{"artists", Text.Trim}}),
9     // Calcolo dell'album meno recente per ogni artista
10    MinAlbumDates = Table.Group(TrimArtists, "artists", {[{"MinAlbumDate", each
11        => List.Min([album_release_date]), type date}}),
12    // Unione con il dataset originale per ottenere i dettagli completi
13    Result = Table.Join(MinAlbumDates, {"artists", "MinAlbumDate"}, Source, {"artists",
14        => "album_release_date"}),
15    // Filtraggio delle righe in cui daily_rank è tra 1 e 10 e snapshot_date è il 24/01/2024
16    FilteredRows = Table.SelectRows(Result, each [daily_rank] >= 1 and [daily_rank] <= 10 and
17        => [snapshot_date] = #date(2024, 1, 24)),
18    // Conta delle occorrenze per ogni artista
19    CountByArtist = Table.Group(FilteredRows, "artists", {[{"Count", each Table.RowCount(_),
20        => Int64.Type}}),
21    // Ordinamento del risultato in base al conteggio decrescente
22    SortedResult = Table.Sort(CountByArtist, {[{"Count", Order.Descending}}),
23    // Join con la tabella originale
24    JoinedTable = Table.Join(SortedResult, {"artists"}, Result, {"artists"}),
25    // Selezione delle colonne desiderate
26    SelectedColumns = Table.SelectColumns(JoinedTable, {"artists", "album_name", "MinAlbumDate",
27        => "snapshot_date", "daily_rank", "Count"}),
28    // Raggruppamento per "artists"
29    GroupedTable = Table.Group(SelectedColumns, "artists", {
30        {"album_name", each List.First([album_name]), type text},
31        {"MinAlbumDate", each List.First([MinAlbumDate]), type date},
32        {"snapshot_date", each List.First([snapshot_date]), type date},
33        {"Count", each List.First([Count]), type number}
34    })
35 in
36 GroupedTable

```

Listato 4.2: Creazione della tabella "Album"

Tramite questa query, la tabella "Album" generata è quella presente in Figura 4.6.

	A _C artists	A _C album_name	MinAlbumDate	snapshot_date	1.2 Count
1	Tate McRae	greedy	15/09/2023	24/01/2024	25
2	Tyla	Water	28/07/2023	24/01/2024	1
3	Flo Milli	Never Lose Me	30/11/2023	24/01/2024	2
4	Noah Kahan	Stick Season	14/10/2022	24/01/2024	8
5	KAROL G	MAÑANA SERÁ BONITO	24/02/2023	24/01/2024	4
6	Mitski	The Land Is In hospitable and So Are We	15/09/2023	24/01/2024	9
7	Teddy Swims	I've Tried Everything But Therapy (Part 1)	15/09/2023	24/01/2024	10

Figura 4.6: Tabella "Album"

- Infine, l'ultima visualizzazione include un grafico a torta che visualizza la percentuale di brani che ogni artista ha in Top 10. I dati mostrano una distribuzione relativamente omogenea tra gli artisti, con *Eladio Carrion* e *Bad Bunny* che si distinguono per avere una percentuale leggermente superiore agli altri.

Anche per la realizzazione di questa visualizzazione è stato prodotto un codice in M che permette la creazione della tabella sulla quale verranno effettuate le analisi: "%_Top10". La query è la seguente:

```

1 let
2   // Caricamento del dataset
3   Source = universal_top_spotify_songs, // Inserisci qui la tua sorgente dati,
4
5   // Divisione degli artisti
6   SplitArtists = Table.ExpandListColumn(Table.TransformColumns(Source, {"artists",
7     Splitter.SplitTextByDelimiter(", ",)}), "artists"),
8
9   // Filtraggio delle righe in cui daily_rank è tra 1 e 10 e snapshot_date è nel
10  // periodo specificato
11  FilteredRows = Table.SelectRows(SplitArtists, each [daily_rank] >= 1 and
12    [daily_rank] <= 10 and [snapshot_date] >= #date(2023, 12, 1) and
13    [snapshot_date] <= #date(2024, 1, 24)),
14
15  // Rimozione delle colonne non necessarie
16  RemoveColumns = Table.RemoveColumns(FilteredRows, {"daily_rank", "snapshot_date"}),
17
18  // Eliminazione delle righe duplicate basate sul nome della canzone
19  DistinctRows = Table.Distinct(RemoveColumns, {"name"}),
20
21  // Conta delle occorrenze per ogni artista
22  CountByArtist = Table.Group(DistinctRows, "artists", {"Count", each
23    Table.RowCount(_), Int64.Type}),
24
25  // Ordinamento del risultato in base al conteggio decrescente
26  SortedResult = Table.Sort(CountByArtist, {"Count", Order.Descending})
27
28 in
29   SortedResult

```

Listato 4.3: Creazione della tabella "%_Top10"

La tabella, risultato della query, avrà il seguente aspetto (Figura 4.7):

	ABC 123 artists	i23 Count
1	Eladio Carrion	13
2	ElGrandeToto	12
3	Club Dogo	10
4	Sfera Ebbasta	9
5	Feid	9
6	21 Savage	8
7	Simba La Rue	8
8	Robi	7
9	Rytmus	7
10	Hafdis Huld	7
11	Keung To	7
12	ODUMODUBLVCK	7
13	Bad Bunny	6

Figura 4.7: Tabella "%_Top10"

In sintesi, questa dashboard è un potente strumento per analisti e decision-maker nell'industria musicale, offrendo dati chiave per strategie di marketing, selezione di artisti per collaborazioni e valutazioni di successo a lungo termine degli album.

4.3.3 Foglio 3 - Andamento giornaliero in classifica di "All I Want For Christmas Is You" con previsione temporale

La presente dashboard di PowerBI fornisce analisi predittive e descrittive riguardanti il confronto tra la popolarità di "All I Want For Christmas Is You" e le canzoni esplicite e non più popolari, trovate nella dashboard 1, nel tempo e a livello europeo. Inoltre, è presente uno studio più approfondito della popolarità media e della posizione media in classifica della canzone "All I Want For Christmas Is You" concentrandosi nei mesi di Novembre e Dicembre, e più precisamente nei giorni di Natale, per osservare dei cambiamenti interessanti.

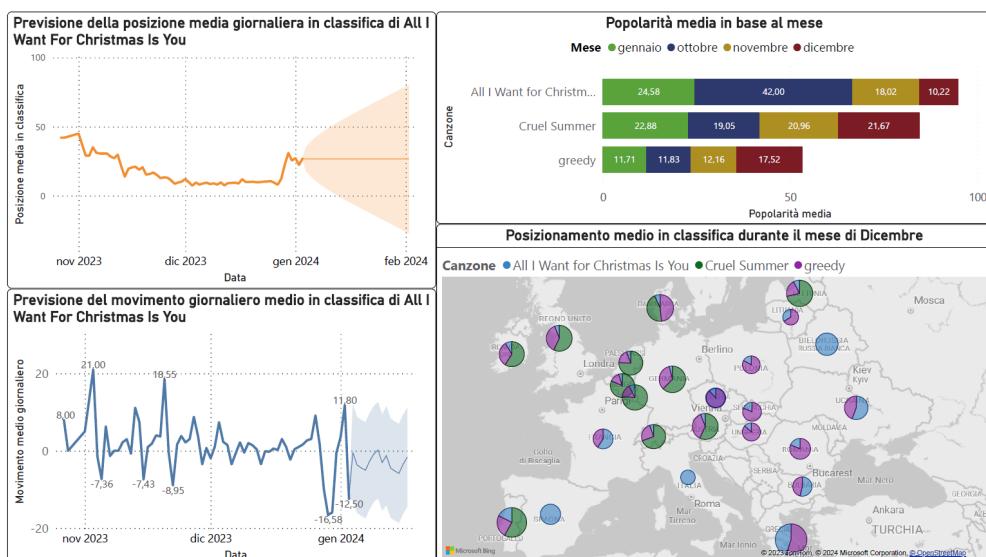


Figura 4.8: Dashboard 3

La dashboard è suddivisa in quattro parti principali, ognuna rappresentando un aspetto diverso dell'analisi dei dati:

- La prima parte presenta un grafico di previsione intitolato "Previsione della posizione media giornaliera in classifica di All I Want For Christmas Is You". Il grafico mostra un trend ascendente a partire da novembre 2023, culminante in dicembre 2023, con una proiezione che si estende fino a febbraio 2024. La regione ombreggiata indica l'intervallo di confidenza della previsione. Il valore più basso, ovvero la posizione più alta in classifica, viene toccato in data 27 dicembre 2023, confermando l'ascesa in classifica della canzone nel mese di Dicembre ed il suo picco di popolarità nei giorni di Natale (Figura 4.9).



Figura 4.9: Posizione media in classifica a Natale

Per quanto riguarda la previsione, data la poca mole di dati che PowerBI può utilizzare per effettuarla, essa non è precisa, indicando una stabilità perenne della canzone in classifica al 27esimo posto. Chiaramente, una situazione di questo tipo non è reale, arrivando alla conclusione che la precisione della previsione risulta essere troppo bassa.

- Nel grafico sottostante, è illustrato il movimento giornaliero medio del brano nel periodo analizzato. Il grafico a linee rivela fluttuazioni significative, con valori che oscillano tra picchi positivi e negativi. Questa visualizzazione permette di verificare e confermare quanto detto nella prima dashboard. Infatti, in corrispondenza dei giorni di Natale, il grafico, ovvero il movimento medio giornaliero in classifica, subisce una brusca diminuzione, definendo un aumento in classifica vertiginoso da parte della canzone "All I Want For Christmas Is You" (Figura 4.10).

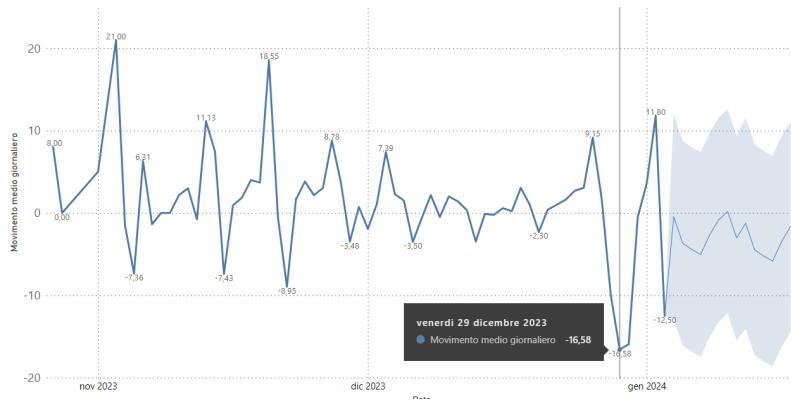


Figura 4.10: Movimento giornaliero medio in classifica a Natale

In questo grafico, la previsione sembra essere molto più accurata e verosimile rispetto a quella precedente; infatti, la previsione definisce che la canzone continua ad avere un aumento del movimento medio giornaliero, constatando che, a mano a mano che ci si allontana dai giorni di Natale, la posizione della canzone in classifica, sarà sempre più alta, allontanandosi dai primi posti destinati alle canzoni di tendenza e più popolari.

- Il terzo elemento è un bar chart denominato "Popolarità media in base al mese", che confronta la popolarità media mensile di tre canzoni: *"All I Want For Christmas Is You"*, *"greedy"* (la canzone esplicita mediamente più popolare) e *"Cruel Summer"* (la canzone non esplicita mediamente più popolare). I dati sono distinti per i mesi di Gennaio, Ottobre, Novembre e Dicembre. La situazione del mese di Dicembre conferma quanto detto in precedenza, *"All I Want For Christmas Is You"* è nettamente più in alto in classifca rispetto alle altre, con una media pari a **10,22**. La situazione è ribaltata nel mese di Ottobre, dove *"greedy"* risulta la canzone posizionata meglio in classifica con un valore medio pari **11,83** e *"Cruel Summer"* non troppo distante con un valore medio pari a **19,05**. Invece, *"All I Want For Christmas Is You"*, con un valore di posizione media in classifica pari a **42,00**, non si avvicina neanche alla Top 30 delle canzoni più popolari di quel periodo.
- Infine, la quarta parte della dashboard è una mappa geografica indicante il posizionamento medio in classifica durante il mese di Dicembre in Europa. Questa visualizzazione utilizza bolle sovrapposte per rappresentare la posizione media delle canzoni in diversi paesi del mondo, con colori distinti per ogni canzone. La situazione è piuttosto chiara, in ogni Paese europeo, escluse poche eccezioni come Francia, Ucraina e Bulgaria, la parte di bolla più piccola, e quindi il posizionamento medio migliore in classifica, è dominato da *"All I Want For Christmas Is You"*. Questa è l'ennesima prova e verifica del dominio incontrastato in classifica da parte della canzone suddetta nel mese di Dicembre.

In sintesi, questa dashboard fornisce insights critici riguardo alla performance temporale e geografica del brano *"All I Want For Christmas Is You"* ed un suo confronto con i brani *"greedy"* e *"Cruel Summer"*, essenziali per informare decisioni strategiche nell'industria musicale e per capire i gusti musicali e l'andamento di determinate canzoni e generi in classifica durante l'anno.

4.3.4 Foglio 4 - Confronto tra paesi in Nord America

La dashboard realizzata nel Foglio 4 (Figura 4.11) offre un'analisi comparativa sulla popolarità e la distribuzione delle canzoni in chiave maggiore e minore tra il Canada (CA) e gli Stati Uniti (US), due colossi del Nord America.

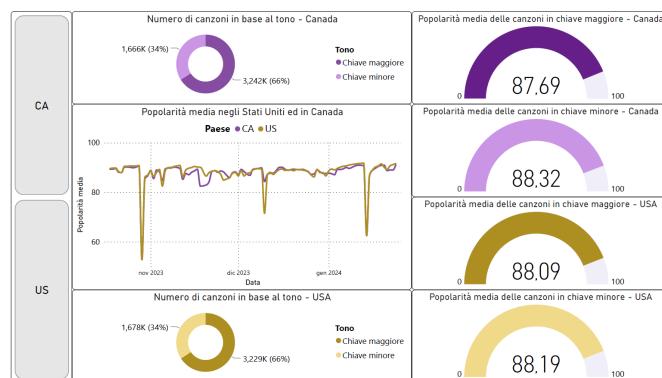


Figura 4.11: Dashboard 4

La Dashboard è divisa nelle seguenti visualizzazioni:

- Il primo elemento, al centro, è un grafico a linee intitolato "Popolarità media negli Stati Uniti ed in Canada" che traccia la popolarità media delle canzoni da Novembre 2023 a Gennaio 2024, con linee distinte per i due Paesi. Il grafico mostra una tendenza generale stabile, con due principali differenze che risaltano all'occhio, datate 18 Novembre 2023 e 10 Dicembre 2023. Inoltre, prima di analizzare le differenze appena citate, è possibile anche determinare una regolarità nella discesa della popolarità; infatti, tutti i picchi verso il basso presenti nel grafico, più o meno accentuati che siano, cadono in corrispondenza di ogni domenica presente da Ottobre a fine Gennaio. Questa analisi dimostra quindi la voglia delle persone, in entrambi i Paesi, di ascoltare musica meno popolare e celebre nei fine settimana. In data 18 Novembre (Figura 4.12) si nota la vera e propria discrepanza tra le due curve; infatti, il valore della popolarità media in Canada (CA) ammonta a **82,48** e quella negli Stati Uniti (USA) è pari a **90,12**. Questo simboleggia come, in quel periodo temporale, negli Stati Uniti le persone preferiscono ascoltare le canzoni più popolari in circolo, mentre in Canada i gusti musicali ricadono su canzoni meno popolari e conosciute.

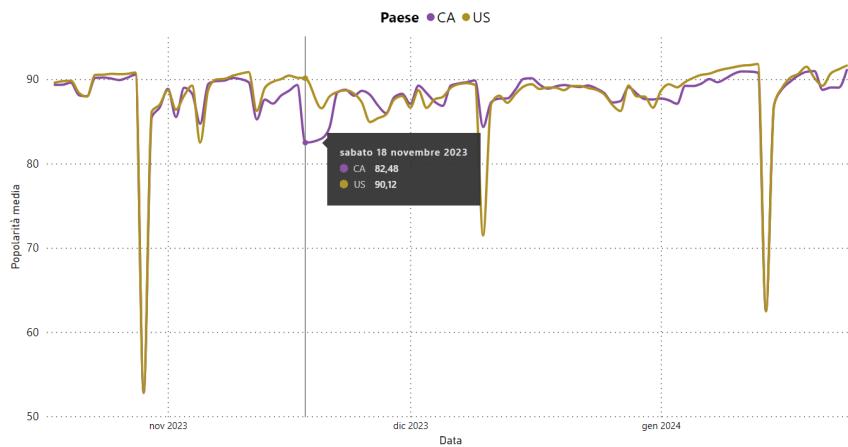


Figura 4.12: Situazione in data 7 novembre 2023

Una situazione analogica, ma opposta e ancora più marcata, si presenta in data 10 Dicembre (Figura 4.13), dove la popolarità media in Canada è pari a **84,34** mentre negli Stati Uniti è **71,46**.

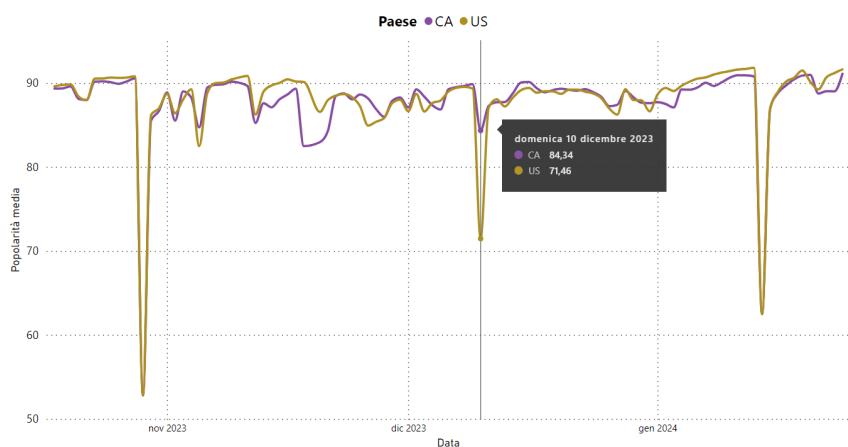


Figura 4.13: Situazione in data 10 dicembre 2023

- Le due visualizzazioni presenti nella parte superiore e inferiore del grafico appena descritto sono due donut chart, che rappresentano il numero di canzoni in base al tono in entrambi i Paesi, evidenziando una maggioranza del **66%** dei brani in chiave maggiore sia in Canada che negli Stati Uniti. Questa equivalenza mostrata dai due donut chart definisce una similitudine nei modi di concepire musica nei due Paesi Nord Americani, preferendo la produzione di più della metà dei brani in chiave maggiore piuttosto che in chiave minore.
- Infine, sulla destra, quattro gauge charts separati quantificano la popolarità media delle canzoni in chiave maggiore e minore per entrambi i paesi. In Canada, le canzoni in chiave minore risultano leggermente più popolari rispetto a quelle in chiave maggiore, mentre negli USA, la popolarità è pressoché identica tra le due chiavi musicali, con una minima prevalenza per le canzoni in chiave minore. Anche in questa visualizzazione, analizzando il primo grafico della dashboard, è presente una data dove l'equilibrio tra le popolartà descritte nei gauge charts viene stravolto e la differenza tra le due è massima. Infatti, analizzando la data 29 Ottobre nel primo Paese (Figura 4.14), il Canada, la popolarità media delle canzoni in chiave maggiore è **38,80**, quella delle canzoni in chiave minore, invece, è pari a **85,40**.

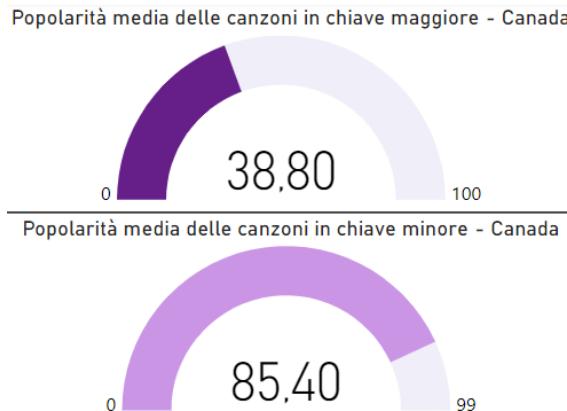


Figura 4.14: Situazione in data 29 ottobre 2023 in Canada

Analogamente per il secondo Paese (Figura 4.15), gli Stati Uniti. Infatti, sempre nella stessa data, la popolarità media delle canzoni in chiave maggiore ammonta a **37,35**, quella delle canzoni in chiave minore, invece, è pari a **85,69**.

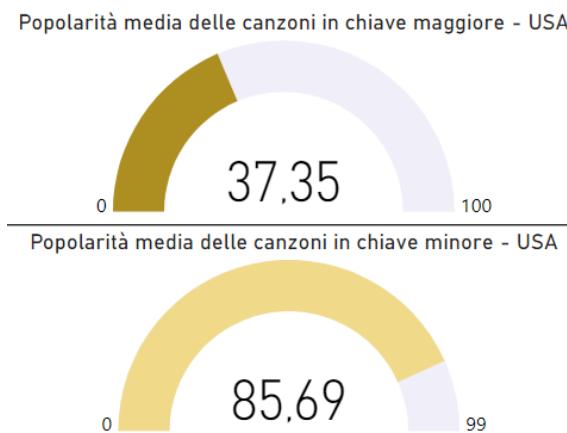


Figura 4.15: Situazione in data 29 ottobre 2023 negli Stati Uniti

Dunque, questo dimostra come, in quel periodo, ci fosse una netta preferenza per le canzoni in chiave minore rispetto a quelle in chiave maggiore, e quest'aspetto si ritrova in entrambi i Paesi poiché, anche nel caso di divergenza massima tra le due medie, continua ad essere presenta una certa equivalenza tra i due.

Quindi, in generale, questa dashboard fornisce:

1. insights sulle preferenze musicali in termini di tonalità negli importanti mercati dei due colossi Nord Americani: il Canada e gli Stati Uniti;
2. informazioni utili per editori musicali e artisti nella scelta delle chiavi per nuove composizioni o nella selezione di brani per promozioni mirate.

4.3.5 Foglio 5 - Continuo del confronto tra Paesi in Nord America

La dashboard presente nel Foglio 5 (Figura 4.16) è un continuo dell'analisi effettuata nel Foglio precedente. Infatti, essa visualizza la correlazione presente tra la valenza media delle canzoni in base al tono (chiave maggiore o minore) per il Canada e gli Stati Uniti.

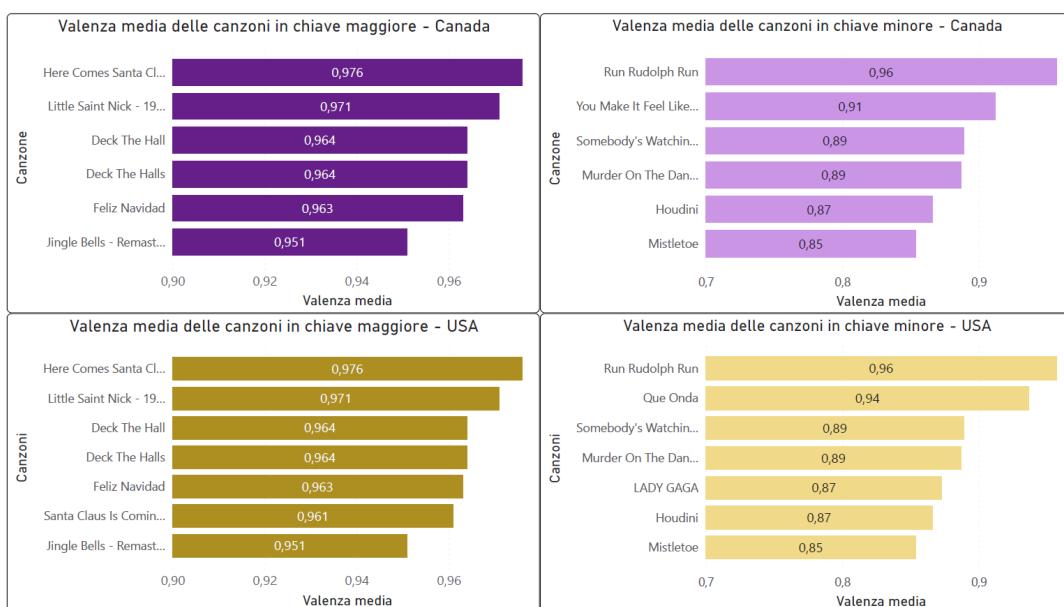


Figura 4.16: Dashboard 5

L'analisi è rappresentata tramite quattro bar chart:

- Il primo grafico, in alto a sinistra, mostra la valenza media delle canzoni in chiave maggiore per il Canada, con *"Here Comes Santa Claus"* che registra il valore più alto. Similmente, il grafico corrispondente in basso a sinistra presenta i dati per gli Stati Uniti, con risultati simili. Questo definisce come, in entrambi i Paesi, le canzoni in chiave maggiore forniscano un tasso di felicità molto simile, dettato, per la maggior parte, dagli stessi brani a patto di qualche differenza.
- Sulla destra, i grafici rappresentano il caso contrario, ovvero la valenza media indotta dalle canzoni in chiave minore, con *"Run Rudolph Run"* che mostra la valenza più alta sia per il Canada sia per gli USA. Inoltre, è interessante osservare che, sebbene le canzoni in chiave minore abbiano generalmente una valenza più bassa rispetto a quelle in chiave maggiore, i valori massimi registrati sono simili tra le due tonalità.

Questi dati possono fornire intuizioni importanti agli editori musicali e agli artisti riguardo alle tendenze emotive associate alle canzoni di differenti tonalità.