

SpotifyStats

Raimondo Rapacciuolo 0522501266

October 2022



1 Introduzione

L'idea principale dietro SpotifyStats è realizzare una web application che permetta l'interrogazione di un database NoSQL orientato ai documenti, contenente informazione riguardo i brani più ascoltati su Spotify presenti nella Top 200 Weekly globale.

2 Tecnologie

Le tecnologie usate per la realizzazione di SpotifyStats sono state:

- **Database:** come database è stato scelto MongoDB, con l'applicazione MongoDB Compass, la scelta è stata effettuata per l'utilità di avere un db orientato ai documenti e per la facilità di integrazione con i maggiori linguaggi di programmazione grazie alla sua ampia diffusione.
- **Backend:** per il lato server è stato scelto NodeJS con il framework ExpressJS, sono state scelte queste tecnologie perchè permettono una facile implementazioni di server e forniscono un framework per il parsing dei dati dal client.
- **Frontend:** per il lato client sono state impiegate le più comuni tecnologie: HTML, CSS e JavaScript e il framework per lo sviluppo frontend Bootstrap, la scelta è ricaduta su queste tecnologie poichè l'interfaccia non prevede elementi di elevata complessità e per la facilità di implementazione.

3 Dataset

Il dataset utilizzato è consultabile al seguente link: <https://www.kaggle.com/datasets/sashankpillai/spotify-top-200-charts-20202021>. Il dataset contiene tutte le canzoni che sono state presenti nella Top 200 Weekly globale nel 2020 e nel 2021 e contiene le seguenti informazioni:

- Posizione più alta raggiunta in classifica
- Numero di presenze nella classifica
- Settimana in cui è stata raggiunta la posizione più alta
- Nome della canzone
- ID della canzone (Fornito da Spotify)
- Numero di riproduzioni della canzone
- Artista
- Numero di follower dell'artista
- Genere
- Data di rilascio della canzone
- Settimane in cui la canzone è stata in classifica
- Popolarità della canzone
- Tonalità della canzone
- Altre informazioni utili per analisi statistiche: Danceability, Acousticness, Energy, Instrumentalness, Liveness, Loudness, Speechiness, Tempo e Valence

Il dataset è stato estratto tramite la libreria Python Spotipy da spotifycharts.com.

4 Preprocessing

La fase di preprocessing si è focalizzata su tre fasi:

1. **Data Cleaning:** sono stati identificati ed eliminati i documenti contenenti dati nulli, è stato scelto di eliminare i documenti affetti dal problema poichè in numero molto piccolo rispetto alle dimensioni del dataset
2. **Feature Selection:** sono state identificate le feature utili allo scopo del progetto, in particolare sono state eliminate tutte le informazioni riguardanti Danceability, Acousticness, Energy, Instrumentalness, Liveness, Loudness, Speechiness, Tempo e Valence

3. **Index Selection:** è stato scelto un indice secondario per MongoDB in modo da velocizzare le query più frequenti, in particolare è stato scelto l'attributo "Popularity" ordinato in modo ascendente.

5 Funzionalità

Le principali funzionalità implementate sono state:

- **Visualizzazione Home:** funzionalità base, permette di visualizzare la pagina principale del sito in cui viene mostrato un sample di 100 canzoni che abbiano una popolarità maggiore di 75, nella pagina è possibile ordinare gli elementi per Nome, Artista, Posizione più alta raggiunta, e Popolarità, la popolarità è la feature di ordinamento iniziale. Funzioni di mongo utilizzate: **aggregate** con **match** e **sample** e **sort**
- **Ricerca per Nome:** dalla barra di ricerca è possibile ricercare canzoni per nome che può essere completo, in questo caso viene mostrato l'unico brano cercato, o una sottostringa del nome, in questo caso vengono visualizzate tutte le canzoni che la contengono. Funzioni di mongo utilizzate: **find** con **regex** e **sort**
- **Ricerca per Artista:** la funzionalità è implementata esattamente come la Ricerca per Nome. Funzioni di mongo utilizzate: **find** con **regex** e **sort**
- **"Most Followed Artists":** permette di visualizzare una pagina in cui sono presenti i 50 artisti più seguiti con Nome, Numero di follower e Numero di canzoni all'interno della classifica, la lista è ordinabile per Numero di follower e per Numero di canzoni in classifica. Funzioni di mongo utilizzate: **aggregate** con **group**, **push sort** e **limit**
- **"Most Streamed Songs":** permette di visualizzare una pagina in cui sono presenti le 50 canzoni più riprodotte con Nome, Artista e Numero di riproduzioni. Funzioni di mongo utilizzate: **aggregate** con **sort** e **limit**
- **Query "Looking For":** la funzionalità che permette di effettuare delle query parametriche al database, è visualizzato come menu laterale e permette di scegliere il genere della canzone, un campo di ricerca tra Popolarità, Posizione più alta raggiunta e Numero di riproduzioni da confrontare con un parametro inserito dall'utente. Funzioni di mongo utilizzate: **find** con **regex** e **sort**