

Link al report:

<https://lookerstudio.google.com/s/jBCgGWeHHaU>

Passaggi del progetto:

1)EDA

Sono stati esaminati i file forniti per capire la tipologia di dato disponibile. Dopo una prima analisi, l'idea è stata di unire i 3 file seguendo una struttura ben precisa:

Per ogni canzone, si memorizza l'edizione del festival (id, anno), il vincitore del festival, la sede, il presentatore, il cantante (id e nome), il piazzamento in classifica della canzone, il tipo (se è stata cantata da solista o con un altro cantante), durata della canzone, popolarità e url di Spotify.

2)Data Cleaning su Excel e Python

71	NF	Carla Bonie Gino Latilla e Duo Fasano e Quartetto Cetra
72	NF	Carla Boni e Duo Fasano e Giorgio Consolini

Come si vede dall'estratto soprastante, la colonna Interprete del file con i dati sulla classifica è quella più sporca e conseguentemente soggetta a più pulizia. Ci sono stati molti casi in cui la congiunzione “e”, che divide i cantanti della canzone, non fosse separata correttamente; a volte il delimitatore era diversa dalla “e” (“&” o “-“). Quindi è stato deciso di normalizzare il delimitatore usando il carattere “-“.

Su python, sono stati caricati i 3 file e si è cercato di costruire il file finale provando a simulare una join seguendo una logica che prevedeva l'utilizzo di 2 for annidati utilizzando come “chiave” l'anno del festival. Oltre le colonne ricavate dalla “join”, sono state aggiunte 5 colonne calcolate. Una che dice il numero di partecipazioni, una con il numero di vittorie del singolo cantante (sono colonne con valori ripetuti quindi), un'altra colonna che mostra il numero di autori della canzone, una che dice se il cantante ha partecipato alla scrittura della canzone (per tracciare i cantautori) ed infine una che indica il genere del cantante (Uomo, Donna, Gruppo), quest'ultima è stata ricavata grazie ad un altro file ausiliare.

Tutto questo è stato salvato in un csv.

Oltre ai dati ricavati dai 3 file principali, verranno usati anche quelli di un file che raccoglie il dati sullo share ed il numero dei telespettatori.

3) Visualization su Google Looker - Data Storytelling

La scelta è stata di dare un overview generale dei vincitori del festival cercando di mostrare le diverse sfaccettature annesse, come il genere, il numero di vittorie, le volte in cui si è arrivati a podio, la percentuale di vittoria, i dati sui cantautori, ecc...

Dopo un quadro sui vincitori, sono stati inseriti anche dei report per i presentatori e per lo share.