# ANALISI DEI DATI DEL FESTIVAL DI SANREMO

## ACCESSO AL REPORT

Di seguito, è riportato il link che consente l’accesso al report realizzato su Google Looker:

[**https://lookerstudio.google.com/reporting/759e211b-dee0-4e69-8ca9-f20870cc35e4**](https://lookerstudio.google.com/reporting/759e211b-dee0-4e69-8ca9-f20870cc35e4)

## OBIETTIVI DELL’ANALISI

Con questo progetto ci si propone di effettuare due indagini:

* Analisi degli ascolti del festival di Sanremo
* Analisi dei cantanti del festival di Sanremo, con focus sulla disparità di genere dei vincitori

## PREPARAZIONE DEI DATI

Per queste analisi sono state utilizzate come sorgenti di dati primarie le seguenti tabelle:

1. Vincitori
2. AscoltiTotaliDisponibili
3. AscoltiPerSerataDisponibili

### SORGENTE DATI “Vincitori”

Per ottenere tale sorgente dati si è partiti dal file “dati-festival-sanremo-1951-2023.xlsx”, dotato dei seguenti campi:

* Anno
* Periodo
* Sede
* Presentatore
* Partecipanti
* Vincitore

Da un’analisi esplorativa dei dati riportati in tale file è emerso, alla luce di quanto osservato sul web, sul sito <https://www.sorrisi.com/musica/sanremo/festival-di-sanremo-vincitori-degli-ultimi-10-anni/>, che il campo vincitore non conteneva valori corretti. Infatti, laddove era presente una coppia di artisti come vincitore, risultava riportato un solo nome. Per tale ragione, ho effettuato uno scraping tramite Python del sito sopra riportato, per ottenere i vincitori corretti.

Di seguito, è riportato il codice Python utilizzato:

Immagine che contiene testo, schermata, software, Pagina Web

Descrizione generata automaticamente

Immagine che contiene testo, schermata, software, Pagina Web

Descrizione generata automaticamente

Con tale codice Python, in particolare, è stato creato un file Excel con al suo interno riportata una colonna Anno-Vincitori, contenente le coppie Anno-Vincitore associate alle varie edizioni.

A questo punto, sono state eseguite delle operazioni in Power Query che hanno consentito di estrarre una colonna riportante i soli vincitori.

Dopodiché, tale colonna è stata inserita nel file “dati-festival-sanremo-1951-2023.xlsx”, al posto della colonna Vincitore iniziale.

A questo punto, il file è stato importato in un foglio di Google, all’interno del quale, laddove in un dato record erano presenti n vincitori, è stata effettuata un’esplosione del record in n record, ciascuno relativo ad un vincitore.

Inoltre, è stato aggiunto un campo “Categoria”, che indica se il vincitore è un uomo, una donna, oppure un gruppo musicale e tre campi booleani calcolati “Uomini”, “Donne”, “Gruppi”, che si basano sul campo “Categoria”.

Tali campi sono risultati funzionali alle rappresentazioni grafiche effettuate in Google Looker.

Di seguito, è riportata, una rappresentazione visiva del foglio di Google sopra citato, denominato Vincitori.

Immagine che contiene testo, schermata, numero, linea

Descrizione generata automaticamente

### SORGENTE DATI “AscoltiTotaliDisponibili”

Per poter effettuare un’analisi sugli ascolti del festival di Sanremo negli anni, è stata acquisita, mediante Power Query, la tabella riportata in Wikipedia, visualizzabile all’interno del link <https://it.wikipedia.org/wiki/Festival_di_Sanremo>.

Sono stati utilizzati gli ascolti rilevati dall’ Auditel, entrato in funzione il 7 dicembre 1986.

Di seguito, è riportata un’immagine di una porzione di tale tabella:

Immagine che contiene testo, numero, software, Icona del computer

Descrizione generata automaticamente

In questa tabella i campi utilizzati sono stati:

* Edizione
* Anno
* Telespettatori I
* Telespettatori II
* Telespettatori III
* Telespettatori IV
* Telespettatori V
* Media

Di particolare interesse sono i campi relativi ai telespettatori, di cui abbiamo i campi “Telespettatori I”, “Telespettatori II”, Telespettatori III”, Telespettatori IV”, e Telespettatori V”, che descrivono, per il festival considerato, il numero di telespettatori nella serata considerata (I, II, III, IV, V) ed il campo “Media”, che descrive il numero medio di telespettatori nelle serate in cui si è suddiviso il festival considerato.

A questo punto, è stato creato un campo “NumeroSerate”, che calcola il numero di serate in cui è risultato suddiviso ciascun festival e, successivamente, è stato creato un campo “Totale Ascolti”, calcolato effettuando il prodotto tra “NumeroSerate” e “Media”.

Il file Excel contenente tale tabella, ottenuta mediante Power Query, è stato importato, successivamente, in un foglio di Google, utilizzato come sorgente dati per il report finale in Google Looker.

Un campo “Presentatore” è stato inoltre aggiunto, per integrare, all’interno di tale tabella, l’informazione relativa ai presentatori che hanno diretto i festival.

Di seguito, è riportata una rappresentazione visiva di tale foglio di Google, denominato “AscoltiTotaliDisponibili”:

Immagine che contiene testo, schermata, numero, Parallelo

Descrizione generata automaticamente

### SORGENTI DATI: “AscoltiPerSerataDisponibili”

A partire dal foglio di Google “AscoltiTotaliDisponibili”, è stato creato un foglio di Google “AscoltiPerSerataDisponibili”, avente i seguenti campi:

* Edizione
* Anno
* Serata
* Presentatore
* Ascolti

Rispetto alla sorgente di dati “AscoltiTotaliDisponibili”, in cui ogni record è relativo ad un festival, in tale sorgente dati ogni record è relativo ad una certa serata (I,II,III,IV, o V) di un festival.

Tale sorgente dati è risultata funzionale alle rappresentazioni grafiche effettuate tramite Google Looker.

Di seguito, è stata riportata una rappresentazione visiva di una porzione di tale foglio di Google:

Immagine che contiene testo, elettronica, schermata, software

Descrizione generata automaticamente