

Ficha Técnica: Proyecto 2 Hipótesis

Título del Proyecto

Hipótesis

Objetivo

Obtener una estrategia efectiva para segmentar su base de clientes, utilizando el proceso de Análisis de Datos y desarrollando la metodología de RFM.

Identificar características técnicas, para encontrar qué hace que una canción sea más escuchada. Mediante técnica de validación (refutar o confirmar) las hipótesis mediante el análisis de los datos, y proporcionar recomendaciones estratégicas basadas en los hallazgos, para que la discográfica y el nuevo artista puedan tomar decisiones informadas que aumenten sus posibilidades de conseguir el “éxito”.

Equipo

Trabajo de equipo (B. Morales - M. Bobadilla).

Herramientas y Tecnologías

Listado de herramientas y tecnologías utilizadas en el proyecto:

- 1) Big Query (SQL).
- 2) Power BI (Python).
- 3) Workspace Google (Presentaciones, Gemini y Documentos).
- 4) Videos Looms.
- 5) Git Hub.

Procesamiento y análisis

1) Proceso de Análisis de Datos:

1.1 Procesar y preparar base de datos

- a) Conectar/importar datos a plataforma Big Query. Tablas importadas: track_in_competition, track_in_spotify y track_technical_info.
- b) Identificar y manejar valores nulos, encontrando 50 nulos in shazam charts en la tabla track_in_competition y 95 nulos en key en la tabla track_technical_info.
- c) Identificar y manejar valores duplicados, identificando duplicados en track_in_spotify.
- d) Identificar y manejar datos fuera del alcance del análisis, eliminamos las columnas key y mode considerando que no son relevantes en éste análisis.
- e) Identificar y manejar datos discrepantes en variables categóricas, realizando el manejo de símbolos en track_in_spotify.
- f) Identificar y manejar datos discrepantes en variables numéricas, identificación de discrepancias en valores numéricos utilizando las funciones MIN, MAX y AVG.
- g) Comprobar y cambiar tipo de dato de String a Integer.
- h) Crear nuevas variables como released_date.
- i) Unir tablas mediante Left Join
- j) Construir tablas auxiliares con la función WITH para conocer el total de canciones por artista y que conteo tienen en los playlist.

1.2 Análisis exploratorio

- a) Agrupar datos según variables categóricas.
- b) Visualizar las variables categóricas.
- c) Aplicar medidas de tendencia central.
- d) Visualizar distribución.
- e) Aplicar medidas de dispersión.
- f) Visualizar el comportamiento de los datos a lo largo del tiempo.
- g) Calcular cuartiles, deciles o percentiles.
- h) Calcular correlación entre variables.

1.3 Aplicar técnica de análisis

- a) Aplicar segmentación.
- b) Validar hipótesis.

Resultados y Conclusiones

Durante el desarrollo del proyecto se utilizó el proceso de Análisis de Datos y la metodología de segmentación (Cuartiles). El Periodo analizado fue el año 2023, la discográfica planteó una serie de hipótesis, sobre qué hace que una canción sea más escuchada. Estas hipótesis se detallan a continuación y de esto se puede concluir lo siguiente:

i) Las canciones con un mayor BPM (Beats Por Minuto) tienen más éxito en términos de cantidad de streams en Spotify.

Se refuta esta hipótesis ya que a mayor BPM no hay mayor cantidad de streams, por lo tanto las canciones de mayor bpm no son las que tienen mas cantidad de streams. Se deduce de gráfica de dispersión BPM VS streams_limpio.

Esto también se ve reflejado en el valor de la correlación de Pearson = -0,002 lo cual es un valor alejado del 1 para el ajuste perfecto.

ii) Las canciones más populares en el ranking de Spotify también tienen un comportamiento similar en otras plataformas como Deezer.

Se refuta esta hipótesis ya que, las canciones populares en ranking e Spotify no presentan el mismo comportamiento de éxito en las diferentes plataformas. Ver gráfica track_name limpio vs suma_plataforma_charts (spotify, deezer, apple y shazam).

iii) La presencia de una canción en un mayor número de playlists se relaciona con un mayor número de streams.

Se refuta la hipótesis ya que no se puede comprobar dicho comportamiento, dado que el contar con un mayor número de playlists por canción no es una relación directamente

proporcional al número de streams, esto se verifica en gráfico de track name limpio VS suma playlists , suma streams.

iv) Los artistas con un mayor número de canciones en Spotify tienen más streams.

Esta hipótesis se refuta con la gráfica de artist_name_limpio vs total tracks, artist_name_limpio vs suma de streams_limpio, quien tiene mas canciones no son quienes presentan mayor cantidad de streams ver ejemplo de Ed Sheeran vs sza.

v) Las características de la canción influyen en el éxito en términos de cantidad de streams en Spotify.

Esta hipótesis es refutada ya que las características técnicas como bpm, acousticness, danceability, energy, instrumentalness, liveness, speechiness y valence todas presentan un bajo valor en su correlación de Pearson respectivamente, ver gráficos característica técnica VS streams_limpio excepto la correlación entre total_playlists vs streams_limpio que es la más cercana a 1 con un valor de 0,7835, sin embargo no es una característica técnica.

Limitaciones/Próximos Pasos

Sin observaciones.

Enlaces de interés

[Bigquery](#)

[PowerBi archivo](#)

[Power Bi PDF](#)

[Presentación archivo](#)

[Presentación video](#)

