

## TALLER 1

### Objetivo

Evaluar las capacidades del estudiante para realizar análisis de datos utilizando técnicas estadísticas y de visualización que permitan entender un conjunto de datos, descubrir insights y plantear hipótesis de negocio.

### Descripción

Para este taller se utilizará una base de datos de Spotify en donde se encuentran las canciones más escuchadas incluyendo características tanto intrínsecas como extrínsecas de cada canción. A continuación, se relaciona el diccionario de datos:

| Field            | Description   |
|------------------|---|
| artist           | Song's artist   |
| song             | Name of the Track   |
| duration_ms      | Duration of the track in milliseconds   |
| explicit         | Explicit content  |
| year             | Release Year of the track   |
| popularity       | The higher the value the more popular the song is                                 |
| danceability     | A value of 0.0 is least danceable and 1.0 is most danceable                       |
| energy           | Represents a perceptual measure of intensity and activity                         |
| key              | The key the track is in   |
| loudness         | The overall loudness of a track in decibels (dB)                                  |
| mode             | The modality (major or minor) of a track  |
| speechiness      | Speechiness detects the presence of spoken words in a track                       |
| acousticness     | A confidence measure from 0.0 to 1.0 of whether the track is acoustic             |
| instrumentalness | Predicts whether a track contains no vocals                                       |
| liveness         | Detects the presence of an audience in the recording                              |
| valence          | A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track |
| tempo            | The overall estimated tempo of a track in beats per minute (BPM)                  |
| genre            | Genre of the track  |

## Actividades

Responda cada una de las siguientes preguntas de negocio utilizando las técnicas y herramientas vistas en clase. En la medida de lo posible, responda cada pregunta aplicando una técnica estadística acompañada de una visualización.

### Punto 1 (5 pts)

Describa el conjunto de datos, tipos de variables y destaque cualquier problema de calidad de datos y procesos de limpieza que deba implementar.

### Punto 1 (5 pts)

¿Cuál es el top 10 de artistas más activos de los últimos 10 años?

### Punto 2 (10 pts)

¿Cómo se diferencian las canciones de los géneros de Latin y Folk/Acoustic en relación con su duración? Halle la diferencia del tiempo promedio de ambos géneros.

### Punto 3 (15 pts)

Halle el top 5 de los géneros del 2019 según la cantidad de canciones. ¿Cómo ha variado la cantidad de canciones del Top de géneros en los años 2000, 2005, 2010, 2015, 2019?

### Punto 4 (15 pts)

¿Cómo ha sido la progresión de nuevos artistas? Asuma que un artista nuevo es aquel del cual no se tiene registros pasados y solo es nuevo durante el primer año de aparición.

### Punto 5 (15 pts)

Grafique la progresión de la popularidad promedio por género y por año. Concluya sobre la gráfica, ej: ¿existen tendencias?

### Punto 6 (20 pts)

Compare los géneros Pop y Rock según sus características de: energy, valence y tempo. Concluya sobre su análisis.

### Punto 7 (15 pts)

Plantee una pregunta de negocio de su interés, ya sea por tipo de música, artistas u otra dimensión, mediante la cual se analicen al menos 3 variables del dataset y concluya.

## Criterios de aceptación

- El taller debe ser desarrollado individualmente.
- Debe ser entregado en los tiempos estipulados y solo a través de BloqueNeón. No se admiten entregas por otros medios como correo electrónico.
- El entregable debe consistir de un notebook subido a un repositorio público de GitHub, el cual debe incluir los outputs de la ejecución de cada celda pero también deberá poder ser ejecutado en su totalidad. En BloqueNeón se debe subir solo la URL del repositorio, no se admitirán commits posteriores a la fecha máxima de entrega.
- Dentro del notebook, haga uso de celdas de texto tipo markdown para exponer sus resultados y/o conclusiones de cada punto. También puede utilizar el archivo Readme del repositorio para concluir lo que considere necesario.
- Debe utilizar únicamente el dataset provisto en este taller.
- Algunas de las librerías que pueden utilizar para generar los gráficos son: Matplotlib, Seaborn, Ggplot, Plotly, Altair.