***Primer Pre Entrega***

***Alumno: Oviedo Nelso Alejandro***

***Data Science CoderHouse [61750]***

***Introducción***

El presente análisis tiene como objetivo explorar un conjunto de datos musicales con el fin de identificar patrones y relaciones entre varias variables. A través de este análisis, buscamos responder preguntas clave, tales como: ¿existe alguna relación significativa entre las características de las canciones (como la energía, popularidad, y el tempo)? ¿Cómo influye el género musical en estas variables? ¿Podemos predecir características musicales basadas en el análisis de los datos? A través de esta exploración, también intentaremos hacer predicciones relacionadas con la popularidad y la energía en función de las características de las canciones.

***Fuente***

Los datos utilizados en este trabajo fueron obtenidos de Kaggle y se encuentran disponibles en el siguiente enlace:

https://www.kaggle.com/datasets/purumalgi/music-genre-classification/data?select=train.csv

Los siguientes datos parte de un conjunto de entrenamiento, muestran un data set con información musical, en la cual buscaremos realizar el análisis e implementación de las preguntas

***Determinación de condiciones iniciales***

* Al comenzar con la carga de datos, verificamos la presencia de valores nulos y aseguramos la integridad de los registros. Las columnas clave que se analizaron incluyen **'Popularity'**, **'energy'**, **'loudness'**, **'danceability'**, y **'instrumentalness'**.
* **Análisis exploratorio**
* Al inspeccionar los primeros registros del DataFrame, observamos que tenemos un conjunto de 17 columnas y 17,996 filas, con algunas variables faltantes que requieren tratamiento. Posteriormente, realizamos un análisis descriptivo para conocer las distribuciones de las variables numéricas y categóricas.
* **Detección y resolución de nulos**
* A través del análisis de nulos, identificamos varias columnas con valores ausentes, que fueron manejados adecuadamente mediante técnicas como la imputación o la eliminación de registros, dependiendo de la gravedad y el contexto de los datos faltantes.
* **Variables clave para el análisis**
* Tras un análisis inicial, identificamos las variables más relevantes para entender las características musicales de las canciones, tales como **'Popularity'**, **'energy'**, **'loudness'**, **'tempo'**, y **'danceability'**.

***Hipótesis***

* **¿Existe relación entre la popularidad y la energía de las canciones?**
* **¿Qué variables influyen más en la popularidad de una canción?**
* **¿Cómo se relacionan las características acústicas de las canciones (como 'danceability' y 'energy') con el género musical?**

#### **Relación entre la Popularidad y la Energía**

Al analizar la correlación entre las variables **'Popularity'** y **'energy'**, encontramos que existe una relación débil .La popularidad de una canción tiene una pequeña influencia sobre su energía, la popularidad está más influenciada por otros factores como el gusto del público, las tendencias culturales, y la exposición en medios de comunicación.

#### **Relación entre el Género Musical y las Variables**

Utilizando un **boxplot**, visualizamos la distribución de la variable **'Popularity'** por género musical (**'Class'**). Como era de esperarse, algunos géneros, como el pop, muestran una mayor concentración de popularidad, mientras que géneros como el jazz o la música clásica presentan una menor concentración. Esto sugiere que los géneros más comerciales tienen una mayor posibilidad de alcanzar una alta popularidad, lo cual es un patrón común en la industria musical.

**Distribución de Popularidad por Género:**

* En los géneros más populares, como el pop y el rock, la popularidad de las canciones tiende a concentrarse en valores altos.
* En géneros menos comerciales, como el jazz, la popularidad presenta una distribución más dispersa, lo que indica que estos géneros tienen menos canciones con alta popularidad.

#### **Energía por Género**

Al analizar la variable **'energy'** a través de un **boxplot**, observamos que los géneros con mayor energía, como el rock o la música electrónica, tienen un rango de valores mucho más alto que otros géneros como el jazz o la música clásica. Esto nos indica que ciertos géneros tienden a producir canciones con características energéticas más pronunciadas.

***Distribución de Energía por Género:***

* Géneros como la música electrónica y el rock tienen valores de energía más altos en comparación con géneros más suaves como la música clásica y el jazz.
* Esto refleja las expectativas del público en términos de la dinámica de cada género musical.

### ***Conclusiones Iniciales***

A través de este análisis, confirmamos algunas de nuestras hipótesis iniciales. Es evidente que **la popularidad de las canciones está estrechamente vinculada a factores externos**, como las tendencias culturales y la exposición mediática, más que a las características acústicas de las canciones. Sin embargo, **la energía de una canción sí está asociada con su género musical**, lo que nos ayuda a entender cómo ciertas características, como el ritmo y la intensidad, definen el estilo de una canción.