



# Análisis de canciones y géneros musicales

# Justificación

Se decidió tomar la variable de género como nuestra target para el problema de clasificación , y las reproducciones promedio al año de cada canción como nuestra variable objetivo de regresión. Las razones para elegir el género como objetivo son:

- **Recomendación Personalizada:** Al etiquetar cada canción con su género musical correspondiente, podemos ofrecer recomendaciones personalizadas a los usuarios, lo que mejora significativamente su experiencia de usuario y fomenta la retención.
- **Organización Eficiente:** La clasificación por género simplifica la organización de la biblioteca de canciones y ayuda a los usuarios a encontrar la música que más les gusta de manera más eficiente.
- **Optimización de Contenido:** Al comprender el género de cada canción, estamos en una posición sólida para promocionar y destacar canciones populares

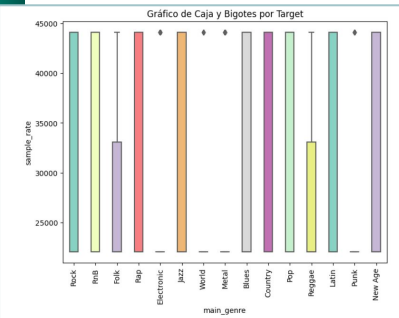
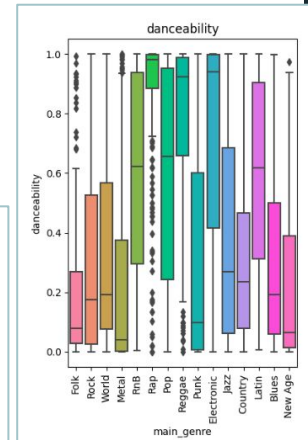
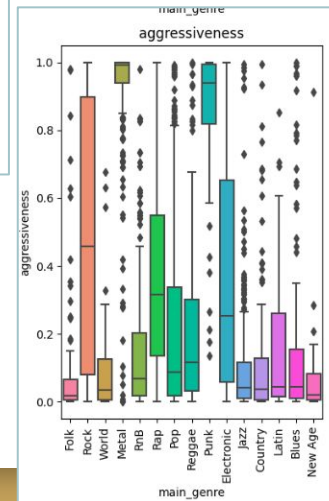
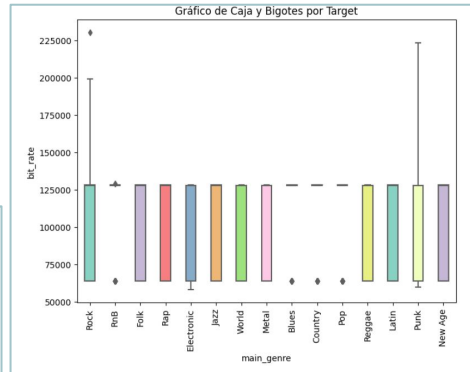
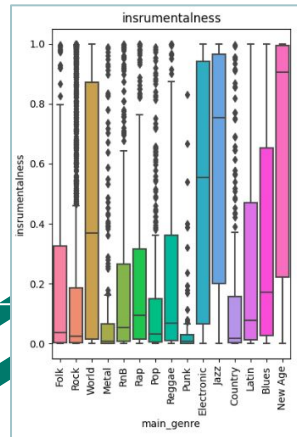
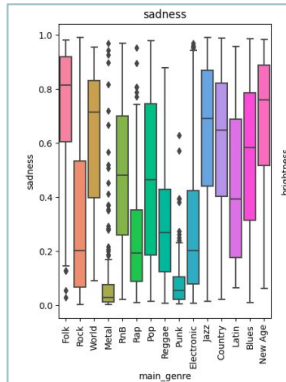
# Justificación

La elección de la variable objetivo de regresión, que en este caso son las reproducciones promedio al año de cada canción, es crucial para nuestro negocio por las siguientes razones:

- **Medición de Popularidad:** Las reproducciones promedio al año son un indicador claro de la popularidad de una canción. Cuantas más reproducciones tenga una canción, mayor será su relevancia en la plataforma. Esta métrica nos permite evaluar el impacto de una canción a lo largo del tiempo y es fundamental para comprender qué contenido atrae a nuestros usuarios.
- **Monetización y Negociación de Licencias:** Las canciones con altas reproducciones promedio al año suelen ser más valiosas en términos de monetización a través de publicidad o suscripciones. Además, es útil en las negociaciones de licencias con artistas y sellos discográficos, ya que les permite evaluar el valor potencial de su música en nuestra plataforma.

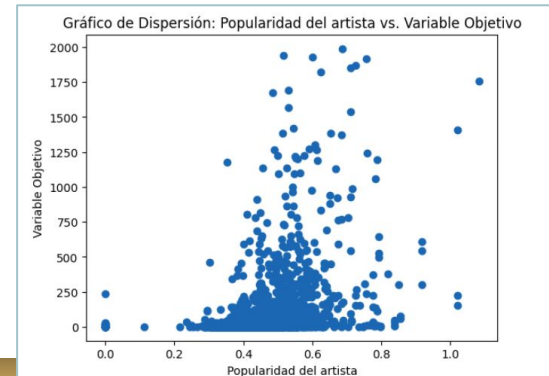
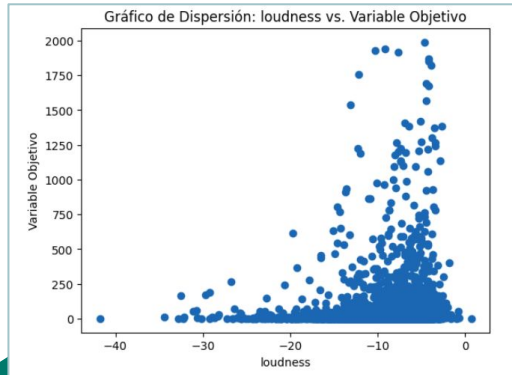
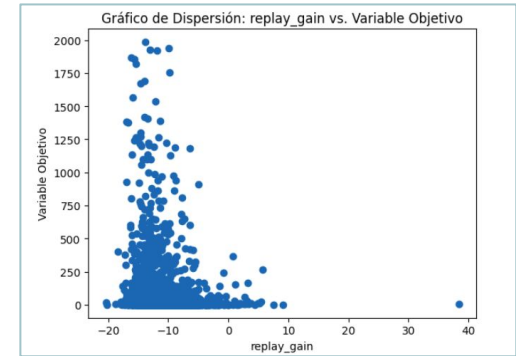
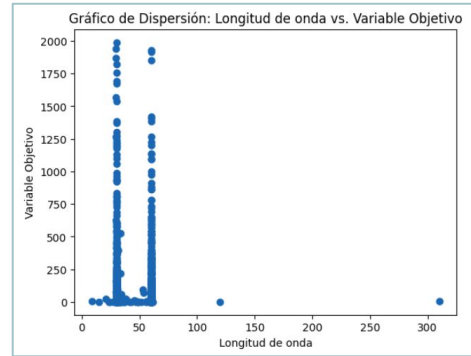
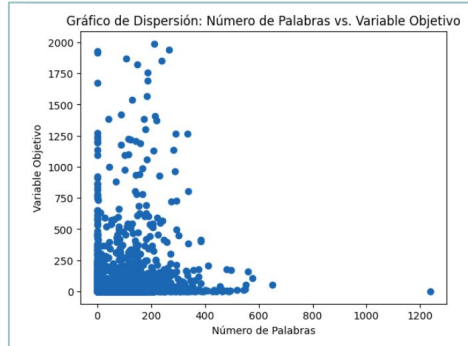
# Análisis

Durante el análisis exploratorio de datos, se encontró que las variables que más discriminan entre los géneros eran: danceability, sadness, aggressiveness, electronicity, instrumentality, sample rate y bit rate (entre otras), en las gráficas se puede ver que la distribución entre categorías no es la misma.



# Análisis

Para el caso de el número medio de reproducciones anual, se detectó que las variables que más tienen influencia son: `artist_hottnesss`, `loudness`, `length`, `replay_gain`, `n_palabras` (entre otras)



# Modelo y resultados

Para el modelo de clasificación de género se usó un modelo Gradient Boosting y el Accuracy fue de 0.5187, se detectó que la causa de no obtener un mejor Accuracy fue debido a que las clases estaban muy desbalanceadas, para poner en contexto en el conjunto la clase Rock tenía 2066 registros, le seguía Electronic con 596 registros, cerca de una cuarta parte de los registros de Rock, además, la clase con menos registros era 'New Age' con 29 registros, se aplicaron técnicas de submuestreo pero no se encontró una mejora significativa.

Para la regresión se usó un Random Forest, se obtuvo un RMSE de 183.77, la desviación estándar de la variable objetivo fue de 201.9, lo que indica que el RMSE es menor a 1 desviación estándar de la variable, este resultado es bueno y cumple con las expectativas.