



Semana Tec

El Arte de la analítica

Luis Ángel Alba Alfaro | A01640314

Obtención de estadísticas descriptivas

Actividad 1

Utilice el conjunto de datos de canciones de Spotify que obtuvimos de Kaggle creado con la intención de crear un algoritmo de machine learning para crear recomendaciones de música basado en los datos recopilados que representan conceptos abstractos como “bailabilidad” “energía”

Estos datos consisten de 200 canciones(renglones) y 14 parámetros(columnas) lo que nos da 2,800 datos, las variables son todas valores Float con la excepción de: key, mode, duration_ms, time_signature y liked, Estas últimas siendo Int.

Los rangos de las variables de tempo y duración que escogí son:

Para el tempo en BPM de 60.171 hasta 180.036

Y para la duración en milisegundos de 77203 hasta 655213

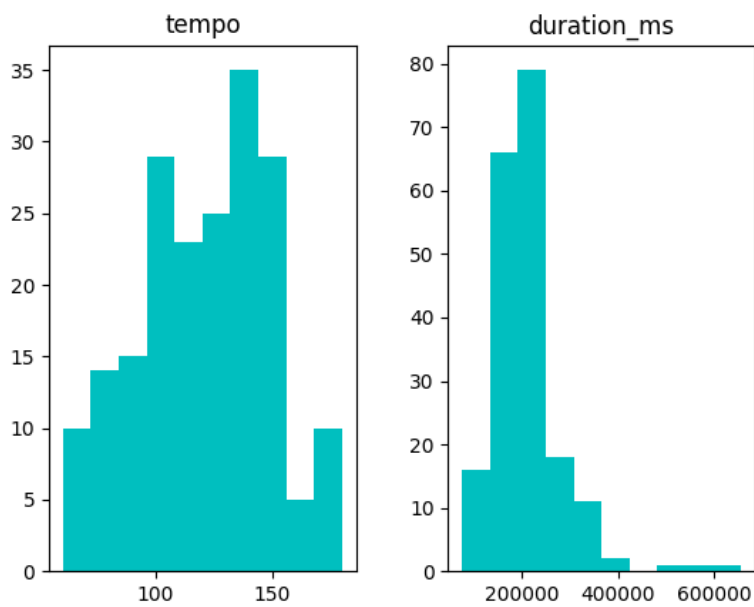
Mi hipótesis fue que entre menor el tempo de la canción más larga será esta, pero basándome en los datos que analice y en la media, mediana y desviación que no son muy dispersos, parece que mi asunción fue errónea ya que, a pesar de que en algunos casos se cumple mi hipótesis inicial en su mayoría parece haber una relación débil entre el tempo y la duración de la canción.

Actividad 2

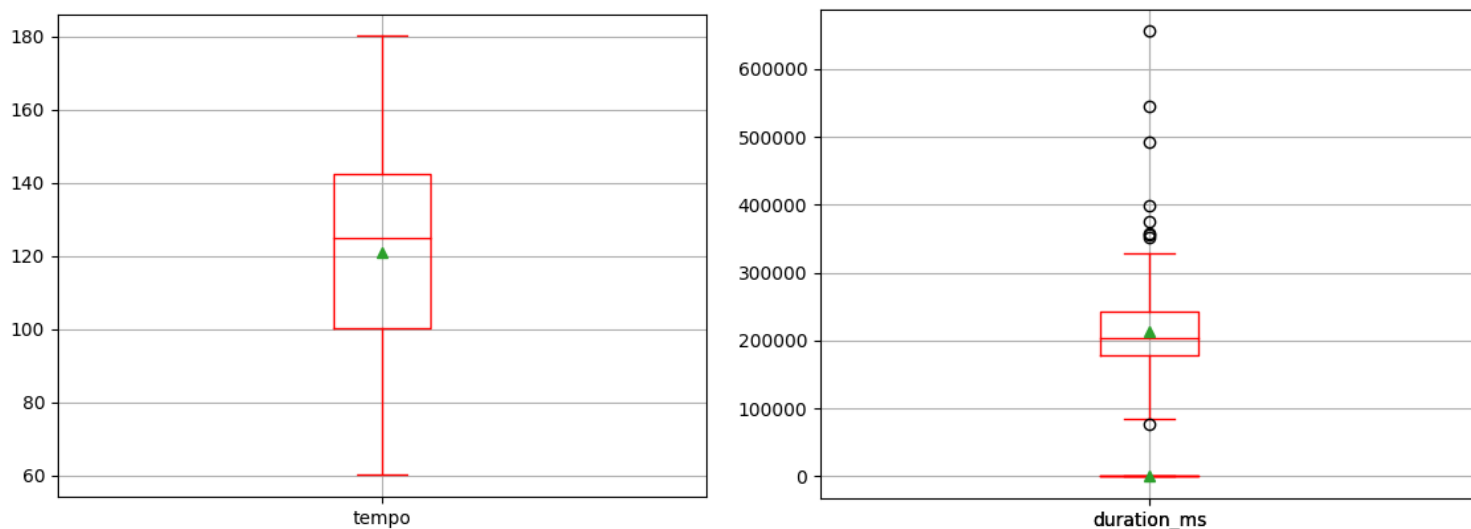
Las 2 variables seleccionadas para análisis fueron las siguientes:

- **Tempo.**
Está representado en Pulsaciones Por Minuto o BPM y se puede describir como la “velocidad” de una canción. Cabe mencionar un BPM más alto no siempre significa una canción más rápida.
- **Duración.**
Esta está representada en milésimas de segundo(ms) y es el factor temporal de las canciones. Entre mas alto sea esta variable más larga será la canción y por lo tanto su duración aumenta.

Histogramas:



Diagramas de cajas y bigotes:



Mapa de Calor:

Gracias al mapa de calor de la correlación entre el tempo y la duración podemos observar que entre mayor es el tempo menor es la duración lo cual concuerda con la hipótesis inicial.

