

Reporte

Obtención de estadísticas descriptivas

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.