



## **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

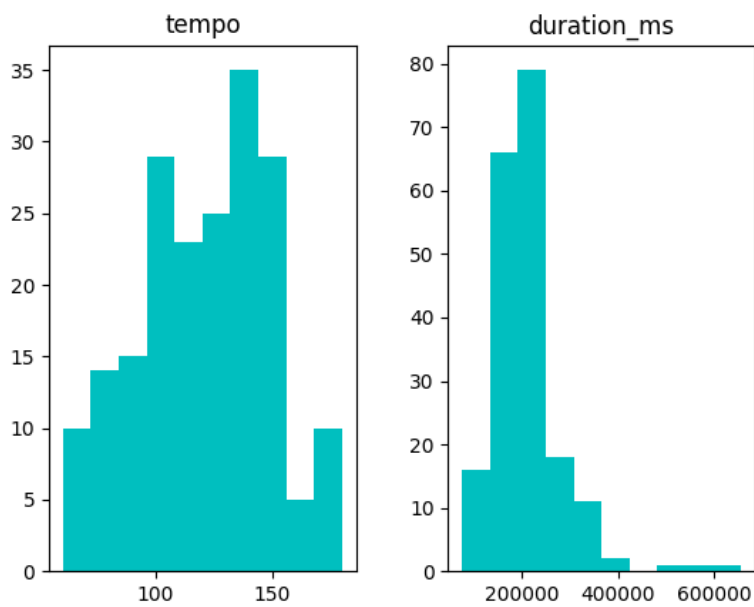
Mi hipótesis fue que entre menor el tempo de la canción más larga será esta, pero basándome en los datos que analice y en la media, mediana y desviación que no son muy dispersos, parece que mi asunción fue errónea ya que, a pesar de que en algunos casos se cumple mi hipótesis inicial en su mayoría parece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

## Actividad 2

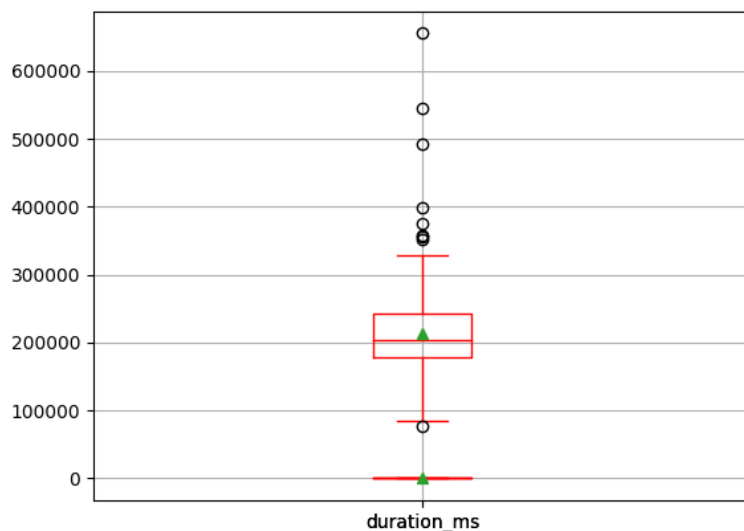
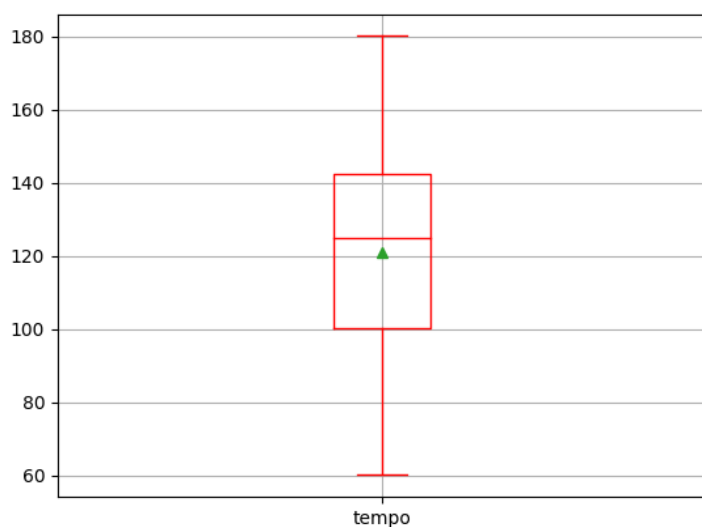
Las 2 variables seleccionadas para análisis fueron las siguientes:

- **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:

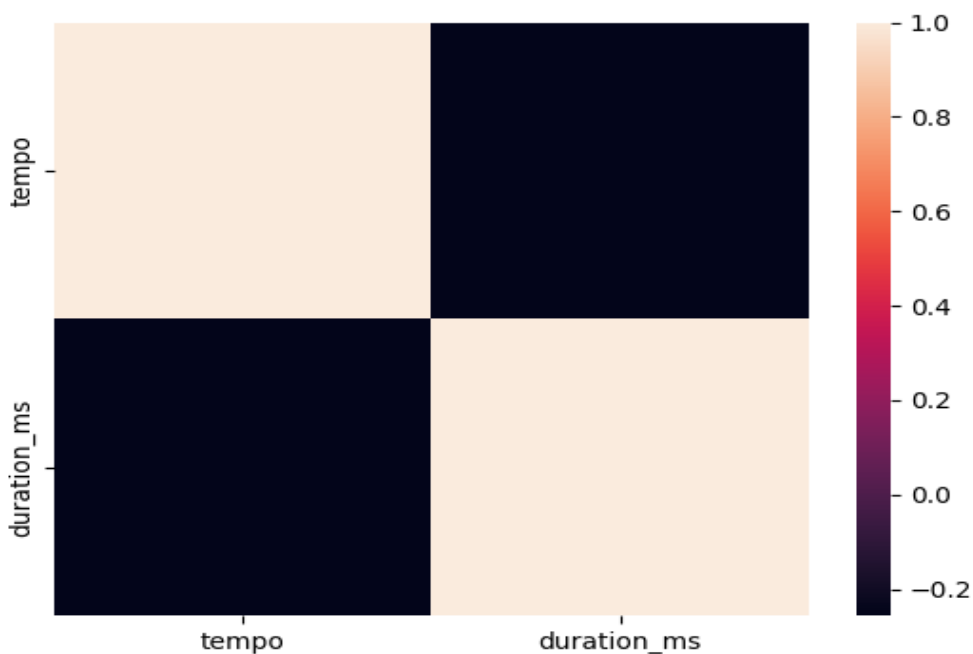


## Diagramas de cajas y bigotes:



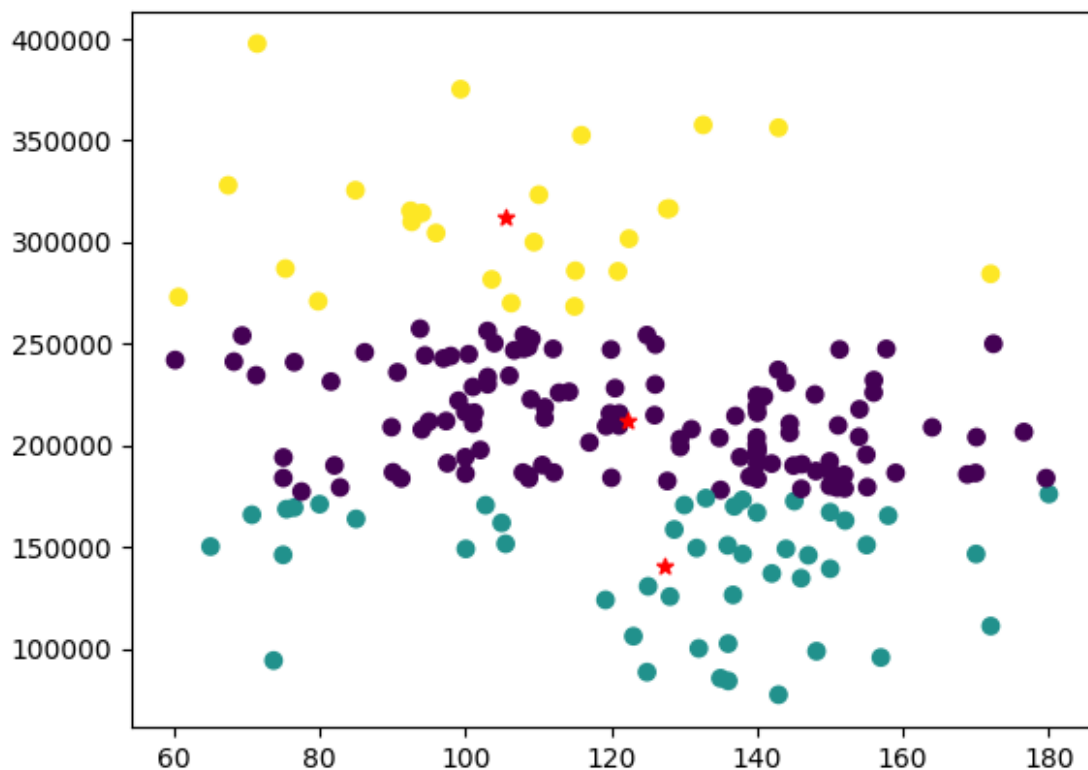
## Mapa de Calor:

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.



### Actividad 3

#### Grafica con centroides K-means:



La grafica no parece representar nada concreto a primera vista ya que la dispersión de los datos y la posición final de los centroides es muy amplia y muestra una orientación horizontal, después de un poco más de deliberación note que la gráfica está dividida en la duración de las canciones sin tomar mucho en cuenta el BPM lo cual me parece un comportamiento extraño.

Utilice el valor K de 3 ya que al experimentar con valores más altos o bajos el comportamiento no cambia, sigue dividiéndolos por tiempo, solo que, en más secciones, esto podría ser usado para agrupar canciones de cierta duración.

La distancia entre los centroides parece ser similar con el centroide de la parte superior estando un poco mas alejado, Si se presentaran muchos *outliers* en nuestros datos la posición de los centroides se vería afectada siendo esta más cercana a la sección donde se encuentren los centroides.

P.D.:

Maestro discúlpeme la falta de calidad en esta entrega, mañana voy a usar el tiempo para atender toda la retroalimentación que me proporciono para hacer una mejor entrega.

### **Referencias:**

Wolf, T. (2021, December 13). *Genre Classification of Electronic Dance Music Using Spotify's Audio Analysis*. Medium. <https://towardsdatascience.com/genre-classification-of-electronic-dance-music-using-spotifys-audio-analysis-7350cac7daf0>