



Semana Tec

El Arte de la analítica

Luis Ángel Alba Alfaro | A01640314

Reporte Final

Introducción

En este reporte hago un análisis de un set de datos, haciendo uso de Python y las herramientas que este lenguaje tiene para sintetizar y analizar datos.

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 abstraídos como “bailabilidad” “energía”

Estos datos consisten en 200 canciones(renglones) y 14 variables (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.

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 más alto sea esta variable más larga será la canción y por lo tanto su duración aumenta.

	Tempo	Duration
count	195	195
mean	121.086174	213408.933333
std	28.084829	72152.392864
min	60.171	77203
25%	100.242	178300.5
50%	124.896	204000
75%	142.4605	242373.5
max	180.036	655213

Algunas de las cosas a resaltar de esta tabla son:

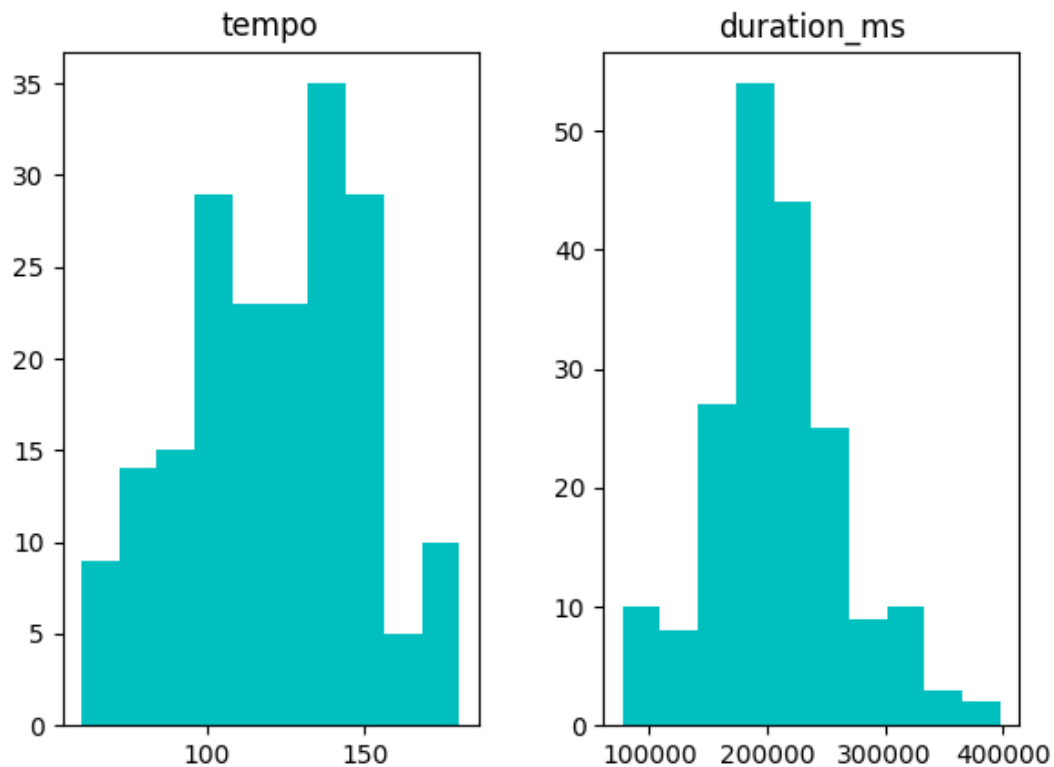
- El rango del tempo en BPM es de **60.171** hasta **180.036**.
- El rango de la duración en milisegundos es de **77203** hasta **655213**.
- La duración media es de **213408** Lo que es igual a **3.55** minutos.

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.

Visualizaciones

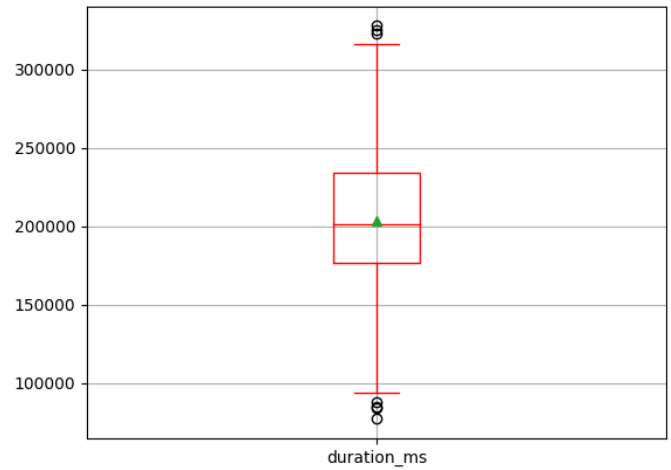
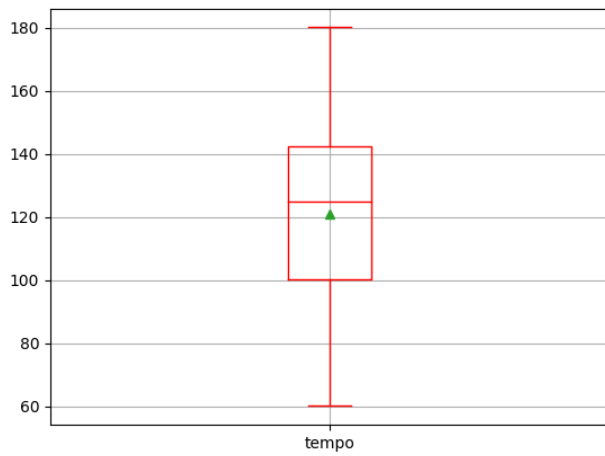
Histogramas:

Con el histograma podemos visualizar como están distribuidos los datos de nuestras variables y podemos extrapolar cuales son los rangos de BPM y Duración donde se concentran los datos (poner rangos). También pude observar que en el caso del tiempo la mayoría de sus datos se encuentran entre **100** y **160** BPM aproximadamente, y en la duración desde **150000** y **280000** milisegundos que equivale a entre **2.5** y **4.6** minutos.



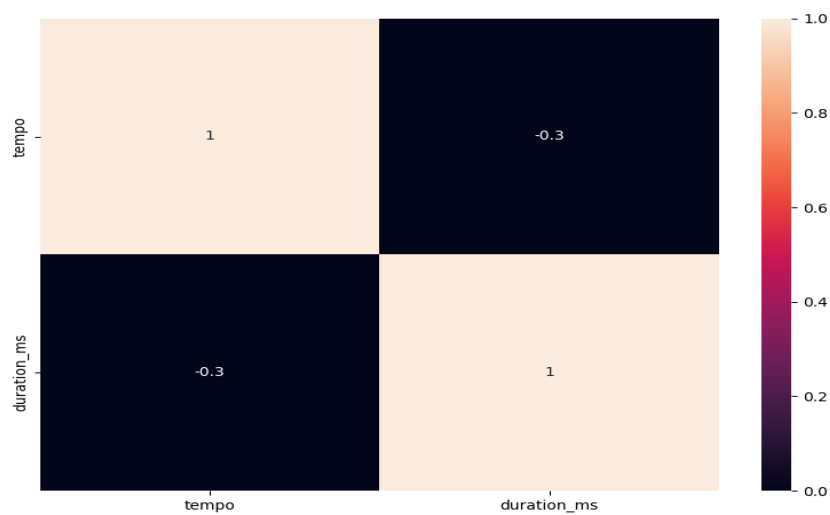
Diagramas de cajas y bigotes:

Con los Boxplot's podemos ver que los datos no son muy simétricos, también en el caso del tempo la agrupación no parece tan ajustada como la de la duración, aunque en esta ultima podemos ver que hay 2 grupos de *outliers* en los extremos estos se deben a que la distribución de los datos

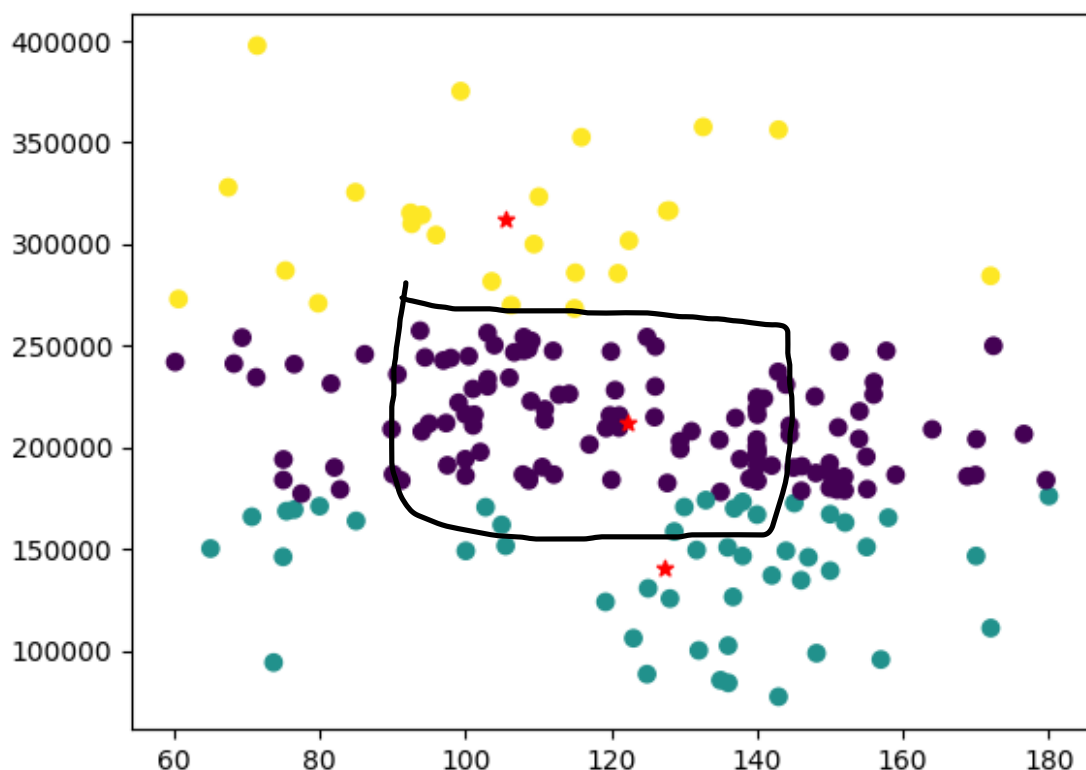


Mapa de Calor:

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.



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 noté 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.

También observo que la mayoría de los datos se encuentran dentro de los rangos que delineados en los histogramas creando un cuadro en alrededor del primer centroide

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.

Conclusión

Ahora puedo ver que efectivamente existe una relación entre el tempo y la duración, aunque esta parece pequeña, me pareció interesante el echo de que la duración media de los datos que utilice sea mayor a la duración promedio de las canciones actuales, yo creo que esto se debe a que la persona que recolecto los datos tiene gustos particulares en este caso la mayoría de las canciones recolectadas son de: rap francés, rap americano, rock y música electrónica los cuales son géneros que tienden a tener un BPM en promedio más bajo, lo cual si nos basamos en lo la relación que encontré entre el BPM y la duración de una pieza resulta en canciones más largas.

Referencias:

Wolf, T. (2021, December 13). *Genre Classification of Electronic Dance Music Using*

Spotify's Audio Analysis. Medium. [https://towardsdatascience.com/genre-](https://towardsdatascience.com/genre-classification-of-electronic-dance-music-using-spotifys-audio-analysis-7350cac7daf0)

[classification-of-electronic-dance-music-using-spotifys-audio-analysis-](https://towardsdatascience.com/genre-classification-of-electronic-dance-music-using-spotifys-audio-analysis-7350cac7daf0)

[7350cac7daf0](https://towardsdatascience.com/genre-classification-of-electronic-dance-music-using-spotifys-audio-analysis-7350cac7daf0)

Vernoug, B. (2021, July 27). *Spotify Recommendation*. Kaggle. Retrieved January 14,

2022, from <https://www.kaggle.com/bricevergnou/spotify-recommendation>

Average Song Length / scatter chart made by Rhettallain / plotly. (2014). Plotly. Retrieved

January 14, 2022, from [https://chart-](https://chart-studio.plotly.com/%7ERhettAllain/131/average-song-length/#/)

[studio.plotly.com/%7ERhettAllain/131/average-song-length/#/](https://chart-studio.plotly.com/%7ERhettAllain/131/average-song-length/#/)

Morris, C. (2021, June 8). *Songs Are Getting Shorter. Blame the Economics of Streaming*

Music. Fortune. Retrieved January 14, 2022, from

<https://fortune.com/2019/01/17/shorter-songs-spotify/>