

Semana Tec TC1002S

Obtención de estadísticas descriptivas

11 de Enero 2022



Herramientas computacionales: el arte de la analítica

Alumno y matrícula: Alison Magie Yáñez Dávila A01423011}

Profesor: Gilberto Huesca Juárez

Campus: Campus Cuernavaca

Instrucciones

En esta actividad, trabajarás con el conjunto de datos que seleccionaste. Tienes que replicar los pasos vistos en clase.


Realiza un código que haga lo siguiente:

1. Cargar los datos.
2. Mostrar el número de variables y el número de registros.
3. Mostrar el nombre de las columnas.
4. Mostrar los tipos de datos.
5. Seleccionar dos columnas que al momento parezcan interesantes. Para ellas:
6. Mostrar los valores únicos
7. Mostrar los valores máximos y mínimos
8. Mostrar la media, la mediana y la desviación estándar.

Realiza un pequeño reporte que muestre lo siguiente:

9. Una introducción al conjunto de datos ¿Qué es? ¿De dónde se obtuvo? ¿Qué representa?
10. Cantidad de datos que tienen, las variables que contiene cada vector de datos y el tipo de variables.

Para las dos variables que escogiste:

11. Los rangos de las variables que escogiste
 12. Basándote en la media, mediana y desviación estándar de cada variable, ¿Qué conclusiones o asunciones puedes obtener de los datos? ¿Parecen muy dispersos? ¿Alcanzas a vislumbrar algún patrón? ¿Parecen relacionados?
- 

Introducción

La música desde siempre ha representado un gran impacto entre nosotros, por lo que podemos definir la música como aquel arte de organización lógica y sensible donde la combinación de sonidos y silencios hacen uso de principios fundamentales de la armonía.

Es por lo anterior, que en la actualidad existen diversas plataformas de música que nos permiten escuchar y descargarla de forma legal, además de compartirla y puntuarla. Entre las plataformas de música streaming más conocidas por los usuarios se encuentra Spotify, el cual está dirigida principalmente hacia artistas musicales.

A continuación, dentro del siguiente trabajo mostraré un análisis de 200 canciones junto a sus estadísticas encontradas en la plataforma Spotify a través de un dataset desarrollado por Brice Vergnou dentro del website Kaggle.

Variables

En la dataset encontrada podemos encontrar una cantidad de datos de 14 columnas y 195 renglones. Dentro del csv las variables que podemos encontrar son:

- **Bailabilidad**

Esta variable describe qué tan adecuada es una pista para bailar en función de una combinación de elementos musicales como tiempo, ritmo, fuerza etc. Donde un valor de 0,0 representa una canción noailable y 1 másailable.

El tipo de esta variable es flotante.

- **Energía**

La energía representa una medida perceptiva de intensidad y actividad.

Podemos definir una pista energética como aquellas que se sienten rápidas, ruidosas y fuertes, mientras que por el contrario, podemos definir una pista no energética como aquellas que se sienten tranquilas.

Tal como la variable anterior, esta tiene un valor mínimo de 0,0 y un valor máximo de 1,0.

El tipo de esta variable es flotante.



- **Llave**

La llave es aquella tonalidad que es encontrada en una pista. Esta variable es definida en función a la notación estándar de clase de tono.

El tipo de esta variable es entero.

- **Volumen**

El volumen es definido como la amplitud y es medido en dB. Los valores típicos de esta oscilan entre -60 y 0dB.

El tipo de esta variable es flotante.

- **Modo**

Indica la modalidad (mayor o menor) de una pista, es decir, el tipo de escala en la que se deriva su contenido melódico, donde mayor es representado por 1 y menor por 0.

El tipo de esta variable es entero.

- **Habla**

El habla es aquella variable donde se detecta la presencia de palabras en una pista. El rango de esta se encuentra entre 0 (no palabras) y 1.

El tipo de esta variable es flotante.

- **Acústica**

La acústica es una medida de confianza que se encuentra entre los rangos de 0,0 y 1,0 (si la pista es acústica).

El tipo de esta variable es flotante.

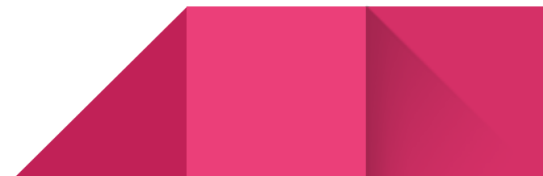
- **Instrumentalidad**

En esta variable se predice si una pista contiene voces, donde los rangos se encuentran entre 0,0 (no contiene voces) y 1,0 (contiene voces).

El tipo de esta variable es flotante.

- **Vivacidad**

La vivacidad es la variable que detecta si existe o no la presencia de una audiencia en la grabación, donde los rangos se encuentran entre 0,0 (no se detecta alguna audiencia) y



1,0 (se detecta una audiencia).

El tipo de esta variable es flotante.

- **Valencia**

La valencia describe la positividad musical transmitida por una pista. Esta variable es medida entre 0,0 (menor positividad) y 1,0 (mayor positividad).

El tipo de esta variable es flotante.

- **Tiempo**

La variable tiempo estima la cantidad de pulsaciones por minuto en una pista (BPM). Esta variable es encontrada entre los rangos 0,0 y 1,0.

El tipo de esta variable es flotante.

- **Duración en ms**

Duración Ms es la duración de una pista en milisegundos.

Esta variable tiene un valor mínimo de 77.2k y un valor máximo de 655k.

El tipo de esta variable es entero.

- **Signatura de compás**

La signatura de compás es una firma de tiempo general estimada de una pista.

Esta variable tiene un valor mínimo de 1 y un valor máximo de 5.

El tipo de esta variable es entero.

- **Gusto**

La variable gusto representa las canciones gustadas y no gustadas del desarrollador de la database.

Esta variable tiene un valor mínimo de 0 y un valor máximo de 1.

El tipo de esta variable es entero.



Análisis

Para el desarrollo del análisis, decidí hacer uso de las variables de balabilidad (rango entre 0 y 1) y energía (rango entre 0 y 1).

Dentro de la variable de bailabilidad podemos encontrar los siguientes resultados:

- **Media:** 0.6366564102564103
- **Mediana:** 0.705
- **Desviación estándar:** 0.2166139610959702
- **Valor máximo:** 0.946
- **Valor mínimo:** 0.13

Por otra parte, en la variable de energía podemos encontrar los siguientes resultados:


- **Media:** 0.6384314871794872
- **Mediana:** 0.659
- **Desviación estándar:** 0.2600958155534138
- **Valor máximo:** 0.996
- **Valor mínimo:** 0.0024

Conclusiones

En la variable bailabilidad, la gran mayoría de canciones registradas dentro de la dataset son consideradas pistas rítmicas y fuertes, ya que la media se encuentra cercano al valor máximo 1, por lo que podemos decir que es congruente el valor de la mediana. Así mismo podemos decir que la desviación estándar al mostrar una variabilidad menor a 1, los datos encontrados tienen un grado de dispersión menor y como consecuencia podemos determinar que son datos reales.

En cuanto a la variable de energía, observamos que la media es cercano al valor máximo 1, y que al igual que la variable de bailabilidad, el valor de la mediana arma sentido. De igual modo, en la desviación estándar la variabilidad encontrada es menor a 1, no obstante es ligeramente mayor en comparación al valor de desviación estándar de bailabilidad.

En conclusión, la gran mayoría de los 195 elementos son considerados bailables y energéticos, esto en función a la combinación de elementos musicales que podemos encontrar en las pistas. Además de que los valores entre las mismas es congruente, ya que a pesar de que la bailabilidad y la energía son medidas de forma distinta, los resultados del análisis apuntaron que se encuentra una relación entre ellas.



Referencias

- EcuRed. (s. f.). *Música - EcuRed*. Recuperado 11 de enero de 2022, de <https://www.ecured.cu/M%C3%BAsica>
- Vergnou, B. (2021, 27 julio). *Spotify Recommendation*. Kaggle. <https://www.kaggle.com/bricevergnou/spotify-recommendation?select=data.csv>