

Universidad Jorge Tadeo Lozano

FACULTAD DE INGENIERIA SISTEMAS

PROYECTO INTELIGENCIA ARTIFICIAL

trip advisor-hotel-reviews

Autores: Nicolas Rojas Molina
Andres Felipe Barbosa Hernandez
Brandon Rambauth

Marzo 2021

Introduccion

¿Cuales son los mejores hoteles para los clientes segun su satisfacción o experiencia basada en como se sienten?

En el presente laboratorio se aplicarán las bases de python aprendidas, haciendo uso de ecuaciones estadísticas de los datos de calificaciones de hoteles usando librerías tales como: Numpy, pandas y Matplotlib con las cuales se analizarán los datos de calificación de 20490 hoteles, teniendo en cuenta que este análisis es basado en las opiniones de los usuarios, según lo que sienten a la hora de usar el servicio o después de usarlo.

Metodo y visualizacion

Tras haber realizado las pertinentes visualizaciones de los datos hemos notado que las opiniones de los huéspedes frente a la atención de los hoteles tiene calificación en un rango de 1 a 5 estrellas, según la experiencia de uso del servicio.

para lo cual mediante el uso de herramientas y métodos de estadística, se tomó la decisión de realizar un análisis de los datos, los cuales constan de hallar la media o media aritmética, la mediana, la moda, y la varianza, con las respectivas ecuaciones.

1. Ecuaciones

La **Media** o **Media aritmética** es el valor promedio de un conjunto el cual es calculado con la suma total de los valores del conjunto dividido entre el número total de valores.

$$Media(X) = \bar{x} = \frac{\sum_{i=1}^N X_i}{N} = \frac{x_1 + x_2 + x_3 \dots + x_n}{N} \quad (1)$$

la **Mediana**, para calcularla lo primero que debemos saber es si N es par o impar.

- si N es impar, la mediana es el valor que está al medio.

$$Mediana(X) = X_{\frac{N+1}{2}} \quad (2)$$

- si N es par, la mediana es la media de los valores del centro $N/2$ y $N/2+1$:

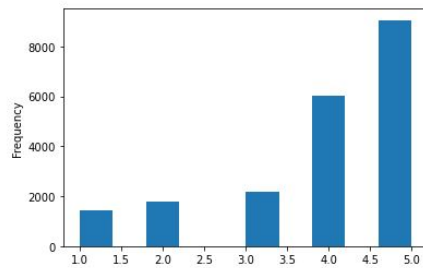
$$Mediana(X) = Media(X_{\frac{N}{2}}, X_{\frac{N}{2}+1}) = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2} \quad (3)$$

la **Moda** es el dato con mayor frecuencia en una distribucion de datos y se calcula de la siguiente forma:

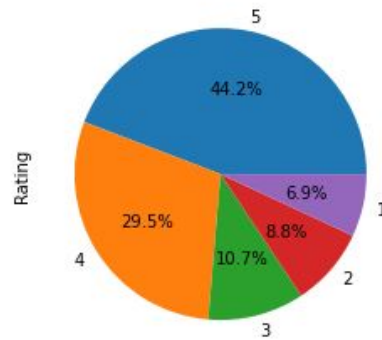
$$M_o = Li + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) * C \quad (4)$$

la **Varianza** es la unidad de medida correspondiente a los datos pero elevada al cuadrado, y siempre es mayor o igual a cero.

$$Var(X) = \frac{\sum_1^n (x_1 - \bar{X})^2}{n} \quad (5)$$



(a) Hist: hoteles segun la calificacion de usuario



(a) Porcentaje de calificacion segun experiencia del usuario

2. Analisis

- al sumar la calificacion de estrellas dadas dividido las cantidades de la misma (20490), podemos sacar la media lo que equivale a el promedio de calificacion que en este caso seria de 3.95.

- al dividir la calificación de estrellas podemos ver que la mediana es decir el 50 % de los datos es 4.0
- encontramos que el número de la calificación del hotel que más se repite es 5 en la calificación de los huéspedes.
- podemos evidenciar en el histograma la separación de los datos al rededor de la media

3. Conclusiones

Como se puede observar en la investigación realizada, los datos obtenidos al final nos da como conclusión:

1. Al analizar los datos nos encontramos con el promedio de las calificaciones de los hoteles visitados por los clientes, los cuales podemos evidenciar que la mitad de los datos son 4.0 estrellas y se puede demostrar que la calificación de los huéspedes que más se repite para los hoteles es de 5.0 estrellas

4. Links video

- <https://www.youtube.com/watch?v=KBqXbW3O1eQ>
- <https://github.com/IAUtaadeo/Enterega1>

5. Glosario

- López, J. F. (2021, 13 febrero). Media. Economipedia.
<https://economipedia.com/definiciones/media.html>
- Serra, B. R. (2020, 22 noviembre). Mediana. Universo Formulas.
<https://www.universoformulas.com/estadistica/descriptiva/mediana/>
- Descriptiva, E. (2021, 10 marzo). moda, para datos agrupados y no agrupados.
moda <http://descriptiva2010.blogspot.com/2010/03/moda-para-datos-agrupados-y-no.html>
- López, J. F. (2021, 27 enero). Varianza. Economipedia.
<https://economipedia.com/definiciones/varianza.html>