



Tecnológico de Monterrey

Instituto Tecnológico de Estudios Superiores de Monterrey

Campus Puebla

Plataformas de analítica de negocios para organizaciones

Alberto Valiente Lobato A01732596

Actividad: reto final

30 de ago. de 23

Análisis descriptivo

Este Proyecto habla acerca de una base de datos de airbnb que cuenta con 74 columnas y 24,226 filas, esta incluye variables de tipo float, int y object, las cuales tienen deficiencias entorno a valores atípicos y valores nulos. Mediante el estudio de las variables, observamos que se tienen algunas variables que son más importantes y sobresalen en la exploración de los datos, estas se definen de la siguiente manera:

a) host_response_time:

Host response time es una variable de categoría objeto, en esta variable podemos encontrar respuestas como “within in a day”, “within in a few hours”, “within an hour”, “a few days or more” y “Desconocido”. Al no ser una variable numérica no tiene rangos, en realidad no se puede evaluar con valores, solo con categorizaciones; y en este caso, se crean categorías con el tiempo que tarda en contestar el host del lugar. En esta misma, el valor que predomina es el de en unas horas, ya que el 42% de los usuarios tenían este método, mientras que los demás utilizaban en una hora o pocos días.

b) host_acceptance_rate:

Esta es una variable porcentual, denominada como objeto por python, esta variable se utiliza para evaluar al host del lugar en airbnb como el porcentaje de aceptación, es decir, cada vez que el host acepta o declina una reservación este valor crece o decrece. El 97% de las personas que tienen un lugar suelen aceptar sin ningún problema secundario.

c) host_identity_verified:

En esta variable tenemos valores que representa una variable categórica, en la cual están incluidas las variables f y t, de esta forma, podemos explicar esta variable como una variable de tipo objeto que brinda la categoría hacia la identidad verificada del host. En este caso el 8% de la base de datos tiene verificada su identidad con Airbnb.

d) property_type:

En esta variable tenemos valores que son de tipo objeto, nos hablan acerca del tipo de propiedad que se está rentando, en este caso, encontramos variables como entire loft o private room, etc. En esta clase, podemos observar que solo el 10% de los valores son cuartos de hotel, lo demás incluyen datos de departamentos, cuartos privados etc.

e) number_of_reviews

Esta variable es una variable de tipo float, es decir, con puntos decimales, y con esta variable podemos observar el número de retroalimentaciones recibidas por cada inmueble. Al ser una variable numérica, lleva un proceso de limpieza de datos y outliers diferente que al de las variables categóricas o numéricas.

f) availability_365

Esta variable es de tipo float, por lo tanto tiene valores numéricos con los que se pueden generar la clase de limpieza en la que tomamos en cuenta outliers y valores nulos, esta variable nos dice los días en los que se encuentra disponible el lugar en días. EN este caso, suele coincidir el número de estrellas en los reviews con la ocupación en días.

g) review_scores_cleanliness:

Esta variable es de tipo numérico, al mismo tiempo podemos observar que habla sobre la calificación entorno a la limpieza del lugar, es decir, una evaluación del 1-5 donde 5 es máximo y 1 es el mínimo. En la limpieza de los datos es posibles observar los nulos y los outliers y reemplazarlos. En esta calificación todas las reviews se encontraban en un promedio entre 3.5 y 3.8 estrellas de calificación.

h) review_scores_communication

De la misma forma de la variable anterior, hablamos de una variable de tipo numerico que genera una evaluación en la comunicación del host con los huéspedes, al evaluar se generan números del 1-5 donde 5 es el mayor y 1 es el menor. En esta calificación todas las reviews se encontraban en un promedio entre 3.5 y 3.8 estrellas de calificación.

i) review_scores_location

Esta es una variable de tipo numérico, dicha variable es una calificación que nos ayuda a evaluar el lugar en el que se encuentra el inmueble. En esta variable se puede generar una limpieza de outliers y valores nulos, en el que 1-5 es el rango de evaluación. En esta calificación todas las reviews se encontraban en un promedio entre 3.5 y 3.8 estrellas de calificación.

j) review_scores_value

Esta variable es de tipo numérico, esta mism nos ayuda a evaluar el valor del lugar, es decir, del 1-5 que valor que el usuario cubrió por hospedarse en el inmueble, dado que 5 es un valor positivo y 1 un valor negativo. En esta calificación todas las reviews se encontraban en un promedio entre 3.5 y 3.8 estrellas de calificación.

Con estas variables definidas, es posible iniciar con el estudio de airbnb en la ciudad de México, ya que, estas nos ayudan a generar filtros que son de beneficio para el estudio.