

Análisis-Extracción de características

Leonardo Isaías Guevara Pioquinto

Predicciones y análisis de correlación en cada tipo de habitación (Hotel room, entire home/apt) respecto a las siguientes variables.

"host_acceptance_rate vs host_response_rate":

1. En México, el modelo matemático obtenido es $y = 0.67835143x + 0.26812711148753066$, con un coeficiente de correlación de 0.7627476762774626. Esto indica que hay una correlación positiva moderada entre la tasa de aceptación del anfitrión y la tasa de respuesta en México. Por cada aumento en la tasa de respuesta, se espera un aumento en la tasa de aceptación.

En España, el modelo matemático obtenido es $y = 0.99791475x + 0.005231914297750051$, con un coeficiente de correlación de 0.9949740354637561. Aquí también se observa una correlación positiva, pero más fuerte. Por lo tanto, en España, la tasa de respuesta del anfitrión tiene una influencia significativa en la tasa de aceptación.

"review_scores_location vs review_scores_cleanliness":

2. En México, el modelo matemático obtenido es $y = 0.62852689x + 0.3277244893054201$, con un coeficiente de correlación de 0.6926441955416891. Esto muestra una correlación positiva moderada entre la ubicación y la limpieza de las reseñas en México. Si un alojamiento tiene una ubicación mejor calificada, es probable que también tenga una calificación más alta en cuanto a la limpieza.

En España, el modelo matemático obtenido es $y = 0.62852689x + 0.3277244893054201$, con un coeficiente de correlación de 0.5445350497122934. Aquí también se encuentra una correlación positiva, pero más débil que en México. Aunque la ubicación puede influir en la limpieza, no es tan determinante como en México.

"host_acceptance_rate vs price":

3. En México, el modelo matemático obtenido es $y = -1.47290562e-05x + 0.9112905846283037$, con un coeficiente de correlación de 0.351128069152858. Aquí se observa una correlación débil entre la tasa de aceptación del anfitrión y el precio en México. Parece que la tasa de aceptación del anfitrión no tiene un impacto significativo en los precios de los alojamientos.

En España, el modelo matemático obtenido es $y = 0.00095754x + 0.6647844385514933$, con un coeficiente de correlación de 0.16495424012440707. En este caso, la correlación es aún más débil. Parece que en España, la tasa de aceptación del anfitrión no tiene una influencia clara en los precios de los alojamientos.

"availability_365 vs number_of_reviews":

4. En México, el modelo matemático obtenido es $y = -4.66947697e-07x + 0.8743659535692$, con un coeficiente de correlación de 0.02356056503512942. Aquí se encuentra una correlación prácticamente nula entre la disponibilidad anual y el número de reseñas en México. La disponibilidad del alojamiento no parece afectar

En base a los modelos y análisis de correlación realizados para cada tipo de habitación en México y España, podemos extraer las siguientes conclusiones:

"host_acceptance_rate vs host_response_rate":

1. Tanto en México como en España, se observa una correlación positiva entre la tasa de aceptación del anfitrión y la tasa de respuesta. Sin embargo, en España esta correlación es más fuerte, lo que sugiere que la tasa de respuesta del anfitrión tiene una influencia más significativa en la tasa de aceptación en comparación con México.

"review_scores_location vs review_scores_cleanliness":

2. En ambos países, se encuentra una correlación positiva moderada entre la ubicación y la limpieza de las reseñas. Esto indica que los alojamientos mejor calificados en cuanto a ubicación tienden a tener mejores calificaciones en términos de limpieza. Sin embargo, esta correlación es más fuerte en México que en España.

"host_acceptance_rate vs price":

3. Tanto en México como en España, se encontró una correlación débil entre la tasa de aceptación del anfitrión y el precio de los alojamientos. Esto sugiere que la tasa de aceptación del anfitrión no tiene un impacto significativo en la determinación de los precios.

"availability_365 vs number_of_reviews":

4. En México, no se encontró una correlación clara entre la disponibilidad anual y el número de reseñas de los alojamientos. Esto indica que la disponibilidad no afecta directamente la cantidad de reseñas recibidas. No se dispone de datos específicos para España en esta comparación.

La regresión lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente y una o más variables independientes. En este caso, se aplicó la regresión lineal para analizar la relación entre las variables mencionadas. El modelo de regresión lineal utiliza los datos disponibles para encontrar la mejor línea recta que se ajuste a los puntos, minimizando la suma de los errores cuadráticos.

La ecuación de la línea recta obtenida se utiliza para predecir los valores de la variable dependiente en función de los valores de la variable independiente. El coeficiente de correlación indica la fuerza y dirección de la relación entre las variables, con valores cercanos a 1 o -1 que indican una correlación fuerte, mientras que valores cercanos a 0 indican una correlación débil.

Es importante tener en cuenta que los modelos de regresión lineal son herramientas estadísticas que proporcionan una estimación de la relación entre las variables basada en los datos disponibles. Sin embargo, no implican una relación causal y otros factores pueden influir en los resultados. Por lo tanto, es necesario interpretar los resultados con cautela y considerar otros factores relevantes en el análisis.