



Tecnológico de Monterrey

**Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Querétaro**

Actividad 1

Analítica de Datos y Herramientas de Inteligencia Artificial II
Grupo 502

Rodrigo Mendoza Orea

A01658480

Profesor:

Alfredo García Suárez

Introducción

En la practica se realizo un analisis comparativo de distintas ciudades, en donde se eligieron datos de 3 diferentes ciudades, los datos brindados fueron tomados de la base de datos de airbnb, en donde las tres ciudades tomadas fueron Ciudad de México(México), Victoria(Canadá) y Bologna(Italia).

Se realizó la limpieza de valores nulos y valores atípicos, para posteriormente realizar el modelo matemático, en donde se dividió la base de datos en los 4 diferentes tipos de cuarto, los cuales fueron: Entire Home, Hotel room, Private room y Shared room.

Se tomaron como variables independientes:

- host_acceptance_rate
- review_scores_rating
- review_scores_cleanliness
- review_scores_communication
- availability_365

Se tomó como variable dependiente:

- number_of_reviews

Procedimiento

Entire Home (Mexico)

Hicimos una regresión lineal simple con cada una de las variables independientes donde nos dieron las siguientes correlaciones:

host_acceptance_rate: 0.000598329486470317

price: 0.007016473294397696

availability_365: 0.007493789640944604

review_scores_rating: 0.009353539997384974

review_scores_cleanliness: 0.018639942293028566

review_scores_communication: 0.031566428222442044

En donde observamos que la variable review scores communication es la que tiene más relación con los siguientes coeficientes de determinación y correlación

coeficiente de determinación: 0.031566428222442044

Coeficiente de correlación: 0.1776694352510922

Entire Home (Italia)

Hicimos una regresión lineal múltiple con las variables que más porcentaje de correlación con ayuda con un mapa de calor, las variables con más correlación fueron review_scores_rating, review_scores_cleanliness y review_scores_communication.

Con las anteriores variables se dio una correlación de: 0.004704907076065745

Entire Home (Canada)

Hicimos una regresión lineal múltiple con las variables que más porcentaje de correlación con ayuda con un mapa de calor, las variables con más correlación fueron review_scores_rating, price y review_scores_communication.

Con las anteriores variables se dio una correlación de: 0.010128732815515407

Hotel room(México)

Hicimos una regresión lineal simple con cada una de las variables independientes donde nos dieron las siguientes correlaciones:

host_acceptance_rate: 0.004710369616617305

price: 0.046699540453199906

availability_365: 0.03642711630552553

review_scores_rating: 0.026447522840118753

review_scores_cleanliness: 0.08655069790802594

review_scores_communication: 0.08653986216299459

En donde observamos que la variable review scores cleanliness es la que tiene más relación con los siguientes coeficientes de determinación y correlación

Coeficiente de determinación: 0.08655069790802594

Coeficiente de correlación: 0.2941949998011964

Hotel room(Italia)

Hicimos una regresión lineal múltiple con las variables que más porcentaje de correlación con ayuda con un mapa de calor, las variables con más correlación fueron availability_365, price y host_acceptance_rate.

Con las anteriores variables se dio una correlación de: 0.11967571310898617

Hotel room(Canadá)

Hicimos una regresión lineal múltiple con las variables que más porcentaje de correlación con ayuda con un mapa de calor, las variables con más correlación fueron review scores rating, price y availability 365

.

Con las anteriores variables se dio una correlación de: 0.37178019701248344

Private room(Mexico)

Hicimos una regresión lineal simple con cada una de las variables independientes donde nos dieron las siguientes correlaciones:

host_acceptance_rate: 0.000598329486470317

price: 0.007016473294397696

availability_365: 0.007493789640944604

review_scores_rating: 0.009353539997384974

review_scores_cleanliness: 0.018639942293028566

review_scores_communication: 0.031566428222442044

En donde observamos que la variable review scores communication es la que tiene más relación con los siguientes coeficientes de determinación y correlación

Coefficiente de determinación: 0.031566428222442044

Coefficiente de correlación: 0.1776694352510922

Private room(Italia)

Hicimos una regresión lineal múltiple con las variables que más porcentaje de correlación con ayuda con un mapa de calor, las variables con más correlación fueron review_scores_communication, price y review_scores_rating

.

Con las anteriores variables se dio una correlación de: 0.06857695293442989

Private room(Canada)

Hicimos una regresión lineal múltiple con las variables que más porcentaje de correlación con ayuda con un mapa de calor, las variables con más correlación fueron review_scores_communication, price y review_scores_rating

Con las anteriores variables se dio una correlación de: 0.046131544901256705

Conclusión

En los 4 tipos de cuarto vimos que no las variables independientes no se ajustaban a un modelo de regresión lineal a la variable dependiente, teníamos una relaciones muy bajas tanto en los modelos de regresión simple, tanto con el múltiple, esto nos indica que no se ajustan los datos al modelo.

Lo pudimos notar desde las gráficas de dispersión de los datos ya que no tenían la forma necesaria para que se ajustará el modelo, pero hicimos el proceso, esto para comprobar que las gráficas no se equivocan.

Entire Home

Como podemos observar en los coeficientes de correlación en las podemos observar que las variables independientes no tienen relación lineal con la variable dependiente, tanto cuando hicimos un modelo simple al igual con el modelo múltiple, lo máximo que obtuvimos fue una relación de 0.1, la cual es una relación muy baja casi inexistente

Hotel room

Como podemos observar en los coeficientes de correlación en las podemos observar que las variables independientes no tienen relación lineal con la variable dependiente, tanto cuando hicimos un modelo simple al igual con el modelo múltiple, lo máximo que obtuvimos fue una relación de 0.3, la cual es una relación muy baja.

Private room

Como podemos observar en los coeficientes de correlación en las podemos observar que las variables independientes no tienen relación lineal con la variable dependiente, tanto cuando hicimos un modelo simple al igual con el modelo múltiple, lo máximo que obtuvimos fue una relación de 0.1, la cual es una relación muy baja casi inexistente.