**Christian Jesus Soto-Vieyra Gil**

**A01707759**

**Actividad 6.1 (Regresión Lineal Simple)**

Para nuestro ejercicio de regresión lineal simple lo primero que debemos hacer es cargar los tres archivos csv a nuestro ambiente, de las ciudades California, DF y Girona. Una vez que tenemos cargados los datos eliminamos todas las columnas que contienen valores nulos. Posterior a esto procedemos a hacer el análisis de correlación en cada tipo de habitación.

Primero que nada cargamos el data de la ciudad de California, y eliminamos los valores nulos con una función de promedio para los valores cuantitativos y con métodos de forward y backward fill para los cualitativos, esto mismo se hace con los otros dos datasets, Distrito Federal y Girona.

Para los outliers aplique el método de rango Intercuartil en las tres ciudades. Pasamos el análisis de correlación que existe para cada tipo de habitación en este caso primero decidimos eliminar cualquier carácter que no sea float o int como los signos de precio.

Decidí ver qué tipos de habitación hay y opté por usar los tipos de Habitaciones “Entire Home/Apt” y “Private Room”., ya que eran los dos valores que tenían más datos. Una vez teniendo nuestros datos procedemos a declarar nuestras variables dependientes e independientes para la regresión lineal. Las variables que se tuvieron fueron:

Variable Independiente: [['host\_response\_rate']]

Variable Dependiente:['host\_acceptance\_rate']

Variable Independiente:[['review\_scores\_cleanliness']]

Variable Dependiente:['review\_scores\_location']

Variable Independiente:[['price']]

Variable Dependiente:['host\_acceptance\_rate']

Variable Independiente:[['number\_of\_reviews']]

Variable Dependiente:['availability\_365']

Variable Independiente:[['number\_of\_reviews']]

Variable Dependiente:['host\_acceptance\_rate']

Variable Independiente:[['review\_scores\_communication']]

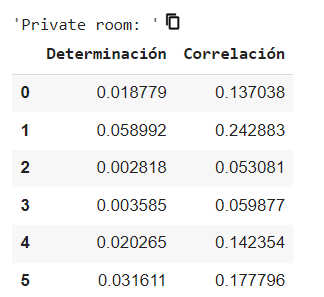
Variable Dependiente:['reviews\_per\_month']

En este caso decidimos hacer un ciclo for para implementar de manera más rápida el modelo de regresión lineal, en el las variables independientes y dependientes se cambiaran por las variables que mencionamos en el código de arriba, después de realizar el ciclo se imprime la ecuación del modelo, en este código además el método \_dict\_ extrae el coeficiente del modelo y el intercep para generar las ecuaciones.

Después, ajustamos el modelo de regresión lineal simple utilizando la función "fit()" de la clase del modelo. Se calcula el coeficiente de determinación (R-cuadrado) y la correlación y se almacenan en las listas "determinación" y "correlación", respectivamente.

Después procedemos a generar las tablas de determinación y correlación, la primera tabla es referente a la variable de “Entire Home/Apt”, la segunda tabla es para “Private Room” y la tercera es para la suma de estas dos.

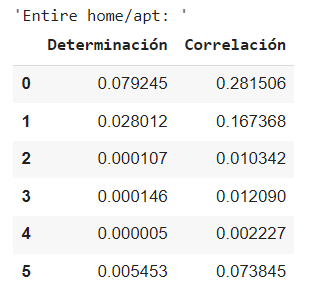
**Tablas de California**

****

**Tablas de Distrito Federal**

****

**Tablas de Girona**



**Listados de variables e interpretación de resultados**

**California**

**'host\_response\_rate' vs 'host\_acceptance\_rate':**

Entire home/apt:

* Correlación: 0.21
* Determinación: 0.04

Private room:

* Correlación: 0.13
* Determinación: 0.02

**'review\_scores\_cleanliness' vs 'review\_scores\_location'**

Entire home/apt:

* Correlación: 0.29
* Determinación: 0.04

Private room:

* Correlación: 0.24
* Determinación: 0.06

**'price' vs 'host\_acceptance\_rate'**

Entire home/apt:

* Correlación:0.08
* Determinación: 0.006

Private room:

* Correlación: 0.05
* Determinación: 0.002

**'number\_of\_reviews' vs 'availability\_365'**

Entire home/apt:

* Correlación: 0.09
* Determinación: 0.008

Private room:

* Correlación: 0.06
* Determinación: 0.003

**'number\_of\_reviews' vs 'host\_acceptance\_rate'**

Entire home/apt:

* Correlación: 0.09
* Determinación: 0.008

Private room:

* Correlación: 0.14
* Determinación: 0.02

**'review\_scores\_communication' vs 'reviews\_per\_month'**

Entire home/apt:

* Correlación: 0.18
* Determinación: 0.03

Private room:

* Correlación: 0.18
* Determinación: 0.03

**Interpretación de California**:

* 'host\_response\_rate' vs 'host\_acceptance\_rate': Estos datos muestran una correlación positiva débil en ambas categorías, lo que sugiere que existe una relación entre la tasa de respuesta del anfitrión y su tasa de aceptación de reservas, aunque esta relación no es muy fuerte. La determinación es baja, lo que significa que solo el 4-2% de la variación en la tasa de aceptación del anfitrión puede explicarse por la tasa de respuesta en ambos casos.
* 'review\_scores\_cleanliness' vs 'review\_scores\_location': En ambos tipos de alojamiento, la correlación es positiva y más fuerte en la categoría de alojamiento completo (entire home/apt). Esto indica que existe una relación entre la limpieza del alojamiento y su ubicación, aunque esta relación es más fuerte en el caso de entire home/apt. La determinación es baja, lo que significa que solo el 4-6% de la variación en la ubicación o limpieza de la calificación del alojamiento puede explicarse por la otra variable en ambos casos.
* 'price' vs 'host\_acceptance\_rate': En ambos tipos de alojamiento, la correlación es positiva, pero muy débil, lo que sugiere que existe una relación entre el precio del alojamiento y la tasa de aceptación del anfitrión, aunque esta relación es muy débil. La determinación es muy baja, lo que significa que solo el 0.2-0.6% de la variación en la tasa de aceptación del anfitrión puede explicarse por el precio del alojamiento en ambos casos.
* 'number\_of\_reviews' vs 'availability\_365': La correlación es positiva, lo que sugiere que existe una relación entre el número de revisiones y la disponibilidad del alojamiento, aunque esta relación es débil en ambos tipos de alojamiento. La determinación es baja, lo que significa que solo el 0.3-0.8% de la variación en la disponibilidad del alojamiento puede explicarse por el número de revisiones en ambos casos.
* 'number\_of\_reviews' vs 'host\_acceptance\_rate': En ambos tipos de alojamiento, la correlación es positiva y más fuerte en la categoría de habitación privada (private room). Esto sugiere que existe una relación entre la tasa de aceptación del anfitrión y el número de revisiones, aunque esta relación es más fuerte en el caso de habitaciones privadas. La determinación es baja, lo que significa que solo el 0.2-0.8% de la variación en la tasa de aceptación del anfitrión puede explicarse por el número de revisiones en ambos casos.
* 'review\_scores\_communication' vs 'reviews\_per\_month': En ambos tipos de alojamiento, la correlación es positiva y del mismo valor, lo que sugiere que existe una relación entre la comunicación del anfitrión y el número de revisiones por mes. La determinación es baja, lo que significa que solo el 3% de la variación en el número de revisiones por mes puede explicarse por la comunicación del anfitrión en ambos casos.

**Distrito Federal**

**'host\_response\_rate' vs 'host\_acceptance\_rate':**

Entire home/apt:

* Correlación: 0.35
* Determinación: 0.12

Private room:

* Correlación: 0.35
* Determinación: 0.12

**'review\_scores\_cleanliness' vs 'review\_scores\_location'**

Entire home/apt:

* Correlación: 0.20
* Determinación: 0.04

Private room:

* Correlación: 0.27
* Determinación: 0.07

**'price' vs 'host\_acceptance\_rate'**

Entire home/apt:

* Correlación:0.002
* Determinación: 0.000008

Private room:

* Correlación: 0.004
* Determinación: 0.000018

**'number\_of\_reviews' vs 'availability\_365'**

Entire home/apt:

* Correlación: 0.06
* Determinación: 0.004

Private room:

* Correlación: 0.06
* Determinación: 0.004

**'number\_of\_reviews' vs 'host\_acceptance\_rate'**

Entire home/apt:

* Correlación: 0.14
* Determinación: 0.02

Private room:

* Correlación: 0.14
* Determinación: 0.02

**'review\_scores\_communication' vs 'reviews\_per\_month'**

Entire home/apt:

* Correlación: 0.002
* Determinación: 0.000008

Private room:

* Correlación: 0.11
* Determinación: 0.01

**Interpretación de Distrito Federal**:

* 'host\_response\_rate' vs 'host\_acceptance\_rate': La correlación es fuerte y positiva tanto para los alojamientos de casa/apartamento completo como para las habitaciones privadas, lo que sugiere que hay una relación significativa entre estas dos variables en ambos tipos de alojamientos. Además, el coeficiente de determinación es alto, lo que indica que una gran proporción de la varianza de una variable puede explicarse por la otra.
* 'review\_scores\_cleanliness' vs 'review\_scores\_location': La correlación es moderadamente fuerte y positiva en los alojamientos de casa/apartamento completo y fuerte en las habitaciones privadas, lo que sugiere que hay una relación significativa entre la limpieza y la ubicación en ambos tipos de alojamientos. Además, el coeficiente de determinación indica que una proporción moderada a alta de la varianza de una variable puede explicarse por la otra.
* 'price' vs 'host\_acceptance\_rate': La correlación es muy baja en ambos tipos de alojamientos, lo que sugiere que hay una relación débil o nula entre el precio y la tasa de aceptación del anfitrión. Además, el coeficiente de determinación es muy bajo, lo que indica que una proporción muy pequeña de la varianza de una variable puede explicarse por la otra.
* 'number\_of\_reviews' vs 'availability\_365': La correlación es baja en ambos tipos de alojamientos, lo que sugiere que hay una relación débil entre el número de críticas y la disponibilidad del alojamiento. Además, el coeficiente de determinación es bajo, lo que indica que una proporción moderada de la varianza de una variable puede explicarse por la otra.
* 'number\_of\_reviews' vs 'host\_acceptance\_rate': La correlación es baja en los alojamientos de casa/apartamento completo y moderadamente fuerte en las habitaciones privadas, lo que sugiere que hay una relación débil o moderada entre el número de críticas y la tasa de aceptación del anfitrión en ambos tipos de alojamientos. Además, el coeficiente de determinación es bajo a moderado, lo que indica que una proporción moderada de la varianza de una variable puede explicarse por la otra.
* 'review\_scores\_communication' vs 'reviews\_per\_month': La correlación es baja en los alojamientos de casa/apartamento completo y moderadamente fuerte en las habitaciones privadas, lo que sugiere que hay una relación débil o moderada entre la comunicación y el número de críticas en ambos tipos de alojamientos. Además, el coeficiente de determinación es bajo a moderado, lo que indica que una proporción moderada de la varianza de una variable puede explicarse por la otra.

**Girona**

**'host\_response\_rate' vs 'host\_acceptance\_rate':**

Entire home/apt:

* Correlación: 0.28
* Determinación: 0.08

Private room:

* Correlación: 0.36
* Determinación: 0.12

**'review\_scores\_cleanliness' vs 'review\_scores\_location'**

Entire home/apt:

* Correlación: 0.16
* Determinación: 0.03

Private room:

* Correlación: 0.17
* Determinación: 0.03

**'price' vs 'host\_acceptance\_rate'**

Entire home/apt:

* Correlación:0.01
* Determinación: 0.0001

Private room:

* Correlación: 0.04
* Determinación: 0.002

**'number\_of\_reviews' vs 'availability\_365'**

Entire home/apt:

* Correlación: 0.01
* Determinación: 0.0001

Private room:

* Correlación: 0.007
* Determinación: 0.00006

**'number\_of\_reviews' vs 'host\_acceptance\_rate'**

Entire home/apt:

* Correlación: 0.002
* Determinación: 0.000005

Private room:

* Correlación: 0.085
* Determinación: 0.007

**'review\_scores\_communication' vs 'reviews\_per\_month'**

Entire home/apt:

* Correlación: 0.07
* Determinación: 0.005

Private room:

* Correlación: 0.19
* Determinación: 0.03

**Interpretación de Girona**:

* 'host\_response\_rate' vs 'host\_acceptance\_rate' observamos que existe una correlación positiva moderada entre la tasa de respuesta del anfitrión y la tasa de aceptación, lo que sugiere que los anfitriones que responden rápidamente a las consultas de los huéspedes también tienen más probabilidades de aceptar reservas, en la determinación observamos que es relativamente baja, lo que indica que solo una pequeña proporción de la variabilidad en la tasa de aceptación se puede explicar por la tasa de respuesta.
* 'review\_scores\_cleanliness' vs 'review\_scores\_location': donde vemos una correlación débil positiva entre la calificación de limpieza y la calificación y la determinación es relativamente baja, lo que indica que solo una pequeña proporción de la variabilidad en la calificación de ubicación se puede explicar por la calificación de limpieza.
* 'price' vs 'host\_acceptance\_rate': Existe una correlación muy débil (casi nula) entre el precio y la tasa de aceptación del anfitrión, lo que sugiere que no hay una relación significativa entre estas variables. La determinación es muy baja, lo que indica que la variabilidad en la tasa de aceptación no se puede explicar por el precio.
* 'number\_of\_reviews' vs 'availability\_365': Existe una correlación débil positiva entre el número de revisiones y la disponibilidad durante todo el año, lo que sugiere que las propiedades con más revisiones también tienden a estar disponibles con más frecuencia. La determinación es baja, lo que indica que solo una pequeña proporción de la variabilidad en la disponibilidad se puede explicar por el número de revisiones.
* 'number\_of\_reviews' vs 'host\_acceptance\_rate': Existe una correlación débil positiva entre el número de revisiones y la tasa de aceptación del anfitrión, lo que sugiere que los anfitriones que aceptan más reservas también tienden a recibir más revisiones. La determinación es relativamente baja, lo que indica que solo una pequeña proporción de la variabilidad en el número de revisiones se puede explicar por la tasa de aceptación.
* 'review\_scores\_communication' vs 'reviews\_per\_month': Existe una correlación moderada positiva entre la calificación de comunicación y el número de revisiones por mes, lo que sugiere que las propiedades con mejores calificaciones de comunicación también tienden a recibir más revisiones. La determinación es relativamente alta, lo que indica que una proporción significativa de la variabilidad en el número de revisiones se puede explicar por la calificación de comunicación.