

Natalia María Ovando Flores A01368118  
Christian Jesús Soto-Vieyra Gil A01707759  
Renata Emilia Chávez Martínez A01351716  
Miguel Saúl Fernández Ávalos A01707491

## Evaluación: Actividad 6.2 (Regresión Lineal Múltiple)

Empezamos la actividad cargando todos los archivos de csv. Para que los modelos de regresión funcionaran correctamente se limpiaron y rellenaron datos nulos. Para los datos string, se utilizó un before fill y forward fill y para los datos numéricos se usó una mediana. Posteriormente, se aplicó una reducción de valores extremos por medio del rango intercuartil, para reducir el ruido y valores demasiado dispersos. Este paso se repitió para los 3 datasets por medio de un FOR.

A continuación, para cada uno de los data frames ya preprocesados, en un FOR, se pasan ciertos datos string a datos que deberían ser enteros para poder trabajar con ellos, para después seleccionar solo las filas que tuvieran como característica 'Entire home/apt' y 'Private room'.

También, aislamos las variables de cada tipo de cuarto como variables dependientes o independientes según sea el caso. Ahora, en vez de solo usar una variable independiente, usamos 3 variables para explicar una sola dependiente. Además dentro de un FOR, se saca la ecuación del modelo matemático para cada caso, y cada país.

Ya teniendo nuestros datos listos, se llevaron a cabo las regresiones en las cuales tuvimos los siguientes resultados:

- California
  - host\_acceptance\_rate vs host\_response\_rate, Price, number\_of\_reviews , review\_scores\_rating: En esta estamos midiendo el efecto que tiene que un host responda rápido, el precio del alojamiento y la calificación en las reseñas, en la calificación final que recibe. Podemos notar que en ambos, tienen una baja correlación, podríamos decir que el host que responde rápido/lento, el precio del alojamiento, el número de reseñas y la calificación de esta no necesariamente significa que el host tendrá buena tasa de aceptación. También, nuestro modelo nos da un R2 muy bajo en ambos cuartos, dándonos a entender que la fuerza de la relación lineal es baja.
    - Entire home/apt:
      - Determinación:0.06
      - Correlación:0.24
    - Private room:
      - Determinación:0.04
      - Correlación: 0.20
  - Price vs availability\_365, host\_response\_rate, number\_of\_reviews, accommodates: En este caso, medimos el efecto de la ocupación, la tasa de respuesta del host, el número de reseñas y cuántos inquilinos caben, en el precio del alojamiento. De igual manera, vemos correlaciones muy bajas y poca fuerza en la relación entre variables. Podríamos decir que estas variables no son explicativas para el precio del alojamiento.
    - Entire home/apt:

Natalia María Ovando Flores A01368118  
Christian Jesús Soto-Vieyra Gil A01707759  
Renata Emilia Chávez Martínez A01351716  
Miguel Saúl Fernández Ávalos A01707491

- Determinación:0.08
  - Correlación:0.28
  - Private room:
    - Determinación:0.017
    - Correlación: 0.13
  - host\_acceptance\_rate vs number\_of\_reviews, review\_scores\_cleanliness, review\_scores\_communication: En este caso, medimos el efecto del número de reseñas, la calificación de limpieza y la calificación en comunicación en la tasa de aceptación del host. De igual manera, estas variables no son explicativas para la tasa de aceptación del host.
    - Entire home/apt:
      - Determinación:0.01
      - Correlación:0.10
    - Private room:
      - Determinación:0.02
      - Correlación: 0.15
- México: En este país podemos dar el mismo veredicto general que en EUA, ninguna podría considerarse variable explicativa para las variables dependientes. Aunque vemos un aumento en general en las métricas, ninguna tiene la significancia suficiente para poder considerar el modelo matemático explicativo.
  - host\_acceptance\_rate vs host\_response\_rate, Price, number\_of\_reviews , review\_scores\_rating:
    - Entire home/apt:
      - Determinación:0.13
      - Correlación:0.36
    - Private room:
      - Determinación:0.12
      - Correlación: 0.36
  - Price vs availability\_365, host\_response\_rate, number\_of\_reviews, accommodates:
    - Entire home/apt:
      - Determinación:0.0011
      - Correlación:0.03
    - Private room:
      - Determinación:0.00038
      - Correlación: 0.019
  - host\_acceptance\_rate vs number\_of\_reviews, review\_scores\_cleanliness, review\_scores\_communication:
    - Entire home/apt:
      - Determinación:0.02
      - Correlación:0.14
    - Private room:
      - Determinación:0.02
      - Correlación: 0.15

Natalia María Ovando Flores A01368118  
Christian Jesús Soto-Vieyra Gil A01707759  
Renata Emilia Chávez Martínez A01351716  
Miguel Saúl Fernández Ávalos A01707491

- Girona: Aquí vemos métricas muy parecidas a los dos países anteriores. En general, podemos decir que las tasas de aceptación del host y el precio del alojamiento, no depende del desempeño de éstos.
  - host\_acceptance\_rate vs host\_response\_rate, Price, number\_of\_reviews , review\_scores\_rating:
    - Entire home/apt:
      - Determinación:0.08
      - Correlación:0.28
    - Private room:
      - Determinación:0.08
      - Correlación: 0.28
  - Price vs availability\_365, host\_response\_rate, number\_of\_reviews, accommodates:
    - Entire home/apt:
      - Determinación:0.02
      - Correlación:0.14
    - Private room:
      - Determinación:0.02
      - Correlación: 0.14
  - host\_acceptance\_rate vs number\_of\_reviews, review\_scores\_cleanliness, review\_scores\_communication:
    - Entire home/apt:
      - Determinación:0.00016
      - Correlación:0.01
    - Private room:
      - Determinación:0.00016
      - Correlación: 0.01