**ACTIVIDAD 1- Análisis de regresión lineal**

**Objetivo**

Analizar el número de vistas en distintos tipos de habitación y ciudad con base a la correlación de distintas variables independientes, con el fin de crear un modelo el cual describa el número de vistas incluyendo el uso de la regresión lineal y la regresión lineal múltiple.

**Desarrollo**

Se utilizaron bases de datos proporcionadas por Airbnb de las siguientes ciudades:

* Ciudad de México
* Barcelona
* Amsterdam

Se llevo a cabo un proceso de limpieza y transformación de los datos con el fin de tener modelos más precisos.

Las variables independientes que se decidieron analizar fueron las siguientes:

* host\_acceptance\_rate
* Price
* availability\_365
* review\_scores\_rating
* review\_scores\_cleanliness

Se realizaron distintos modelos de regresión lineal con el fin de modelar el numero de vistas por tipo de habitación.

Se usarán conceptos de estadística como el coeficiente de correlación y determinación para evaluar la precisión de los modelos.

**Resultados**

Una vez realizados los distintos modelos de regresión lineal para cada una de las variables independientes, por cada tipo de habitación, los resultados fueron los siguientes:

**Tabla

Descripción generada automáticamente**

Podemos observar coeficientes de correlación y determinación bajos que nos indican que un modelo de regresión lineal no es un buen modelo para ajustarse a los datos del número de vistas.

Sin embargo, utilizaremos aquellas variables con un mayor coeficiente de correlación, para realizar un modelo de regresión múltiple. Este modelo de regresión múltiple lo aplicaremos en otras ciudades para evaluar su comportamiento.

En la siguiente gráfica se muestran los distintos coeficientes de correlación:

Las 3 variables que tienen una mayor correlación con el numero de vistas en la ciudad de México son:

* Host acceptance rate
* Review scores cleanliness
* Review scores communication

**Análisis de regresión lineal múltiple**

Usando las 3 variables con mayor correlación en la CDMX se realizo un modelo de regresión lineal múltiple para las ciudades de Barcelona y Ámsterdam.

Tabla

Descripción generada automáticamente

Los coeficientes siguen siendo demasiado bajos, sin embargo, aumentaron en comparación con el modelo de regresión lineal. No se considera un buen modelo, habría que evaluar el uso de mayores variables independientes.

**Conclusión**

Analizando los distintos resultados en las distintas ciudades podemos observar que a medida que cambia la ciudad y cambia el tipo de habitación la correlación entre variables cambia, en donde en una ciudad pudo haber sido el precio una variable determinante, en otra ciudad no lo es.

Esto nos indica que no podemos generalizar en cuanto a que variables tienen una mayor correlación para el numero de vistas, cada ciudad y cada tipo de cuarto tiene un comportamiento diferente.

Por otro lado, el modelo de regresión lineal y el modelo de regresión lineal múltiple, no se ajustan de manera adecuada a los datos, habrá que evaluar el uso de otros modelos, con el fin de obtener un mayor coeficiente de correlación.