

## Reporte Comparativo

Para realizar la actividad, primero realicé un análisis de 6 variables con respecto a 1 de salida, todas agrupadas en 4 distintas categorías. Estas 6 variables eran: "host\_acceptance\_rate", "price", "availability\_365", "review\_scores\_rating", "review\_scores\_cleanliness", y "review\_scores\_communication". La variable objetivo fue: "number\_of\_reviews". Y todo el dataset fue agrupado por el tipo de cuarto, que, en este caso, fueron 4. Para este primer caso, que fue la Ciudad de México, realicé un análisis de regresión simple, comparando la correlación de las 6 variables con respecto a la variable objetivo, esto para cada tipo de cuarto. Al obtener las correlaciones, obtuve el modelo matemático que mejor describiera el comportamiento de las variables, utilizando la que tuviera el mejor índice de correlación. Al tener los 4 distintos modelos, obtuve la correlación y determinación para cada diferente tipo de cuarto. Estos fueron los resultados obtenidos:

Room Type	Coef Deter	Coef Corr
Entire home/apt	0.009648	0.098224
Private room	0.015311	0.123736
Hotel room	0.050175	0.223997
Shared room	0.029711	0.172370

Después realicé el mismo análisis con la ciudad de Ámsterdam, pero en esta ocasión, no fue con regresión lineal simple, sino múltiple. Después de calcular las correlaciones individuales, elegí las 3 que tuvieran mayor correlación con la variable objetivo, y las utilicé para crear el modelo, igualmente, para cada tipo de cuarto. Al final también calculé el coeficiente de determinación y de correlación para cada tipo de cuarto, y estos fueron los resultados:

Room Type	Coef Deter	Coef Corr
Private room	0.038362	0.195863
Entire home/apt	0.009214	0.095988
Hotel room	0.121449	0.348495
Shared room	0.164196	0.405211

También calculé lo mismo para la ciudad de Berlín, dando los siguientes resultados:

Room Type	Coef Deter	Coef Corr
Entire home/apt	0.045552	0.213429
Private room	0.030998	0.176063
Shared room	0.125187	0.353817
Hotel room	0.162568	0.403198

Como podemos observar, el modelo que mejor representa el comportamiento del dataset es el de Berlín. Sin embargo, este sigue siendo demasiado bajo. Al observar el de Ámsterdam, vemos como tiene valores acercados, pero poco más bajos. Finalmente, cuando observamos los valores de correlación y determinación de la Ciudad de México, vemos cómo son significativamente más bajos que los otros 2. Con esto podemos concluir que, aunque para estos 3 casos, no existe realmente una correlación entre los datos y el comportamiento de estos para la variable objetivo, al menos una regresión lineal múltiple da un mejor acercamiento a éste que una regresión lineal simple. Esto es porque, al ser múltiple, varias variables son factores, por lo que se puede manipular de mejor manera para acercarse al valor objetivo. Sin embargo, cuando es simple, sólo hay una variable que se utiliza, por lo que, siendo tan baja, no es posible aumentar esta correlación.