

# Actividad 2

Victoria Eugenia Téllez Castillo A01732258

Alberto Rodriguez Porras A01721497

7 de octubre del 2022

Nombre de la materia: Gestión de proyectos de plataformas tecnológicas.

**ANÁLISIS EXPLORATORIO**

Dentro de este reporte se analizarán los resultados obtenidos de una base de datos de Airbnb México. Las variables que se analizarán son las siguientes:

* Precio
* Minimum nights. (mínimo de noches)
* Number of reviews. (Número de reseñas)
* Reviews per month. (Reseñas al mes)
* Calculated host listings count. (Recuento de listados de hosts calculado)
* Availability 365. (Disponibilidad 365)
* Number of reviews ltm. (Número de reseñas).

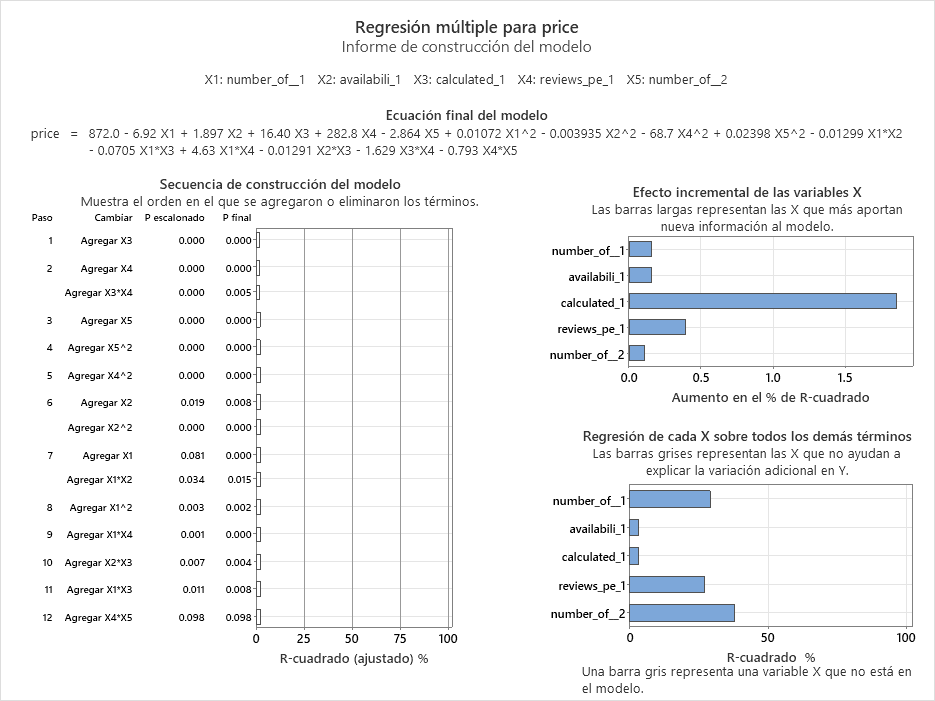
El objetivo final de todos estos análisis es hacer comparaciones de las mismas variables de la empresa Airbnb en México con otros países para saber cuál país es el que mejor coeficientes de correlación y determinación tiene al relacionarse con la variable de precio. Los países que vamos a comparar junto con México son Italia (Milán) y Canadá (Montreal). Se escogieron estas ciudades pensando en tener contextos diferentes, buscando encontrar patrones en ciudades con mucha densidad de población y que son consideradas como puntos turísticos alrededor del mundo, o lugares frecuentemente visitados. Primeramente se realizó el análisis con Ciudad de México, México:

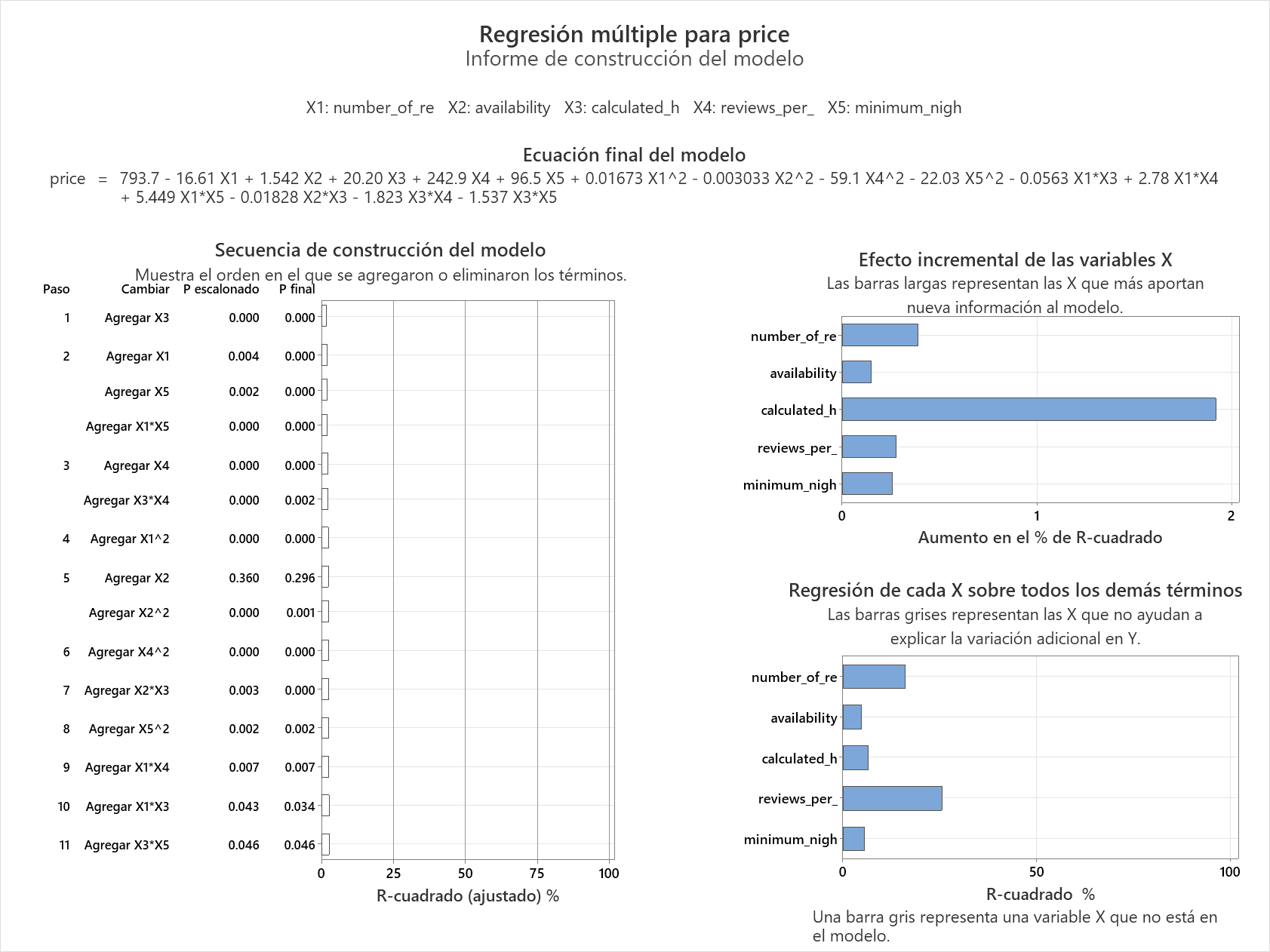
**CIUDAD DE MÉXICO**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coeficiente de Correlación** | **Coeficiente de Determinación** |
| Minimum nights | -0.015 | 0.000225 |
| Number of reviews. | -0.014 | 0.000196 |
| Reviews per month. | 0.046 | 0.002116 |
| Calculated host listings count. | 0.132 | 0.017424 |
| Availability 365. | 0.023 | 0.000529 |
| Number of reviews ltm. | -0.007 | 0.000049 |

En esta tabla se analizan las variables de Airbnb México, y se observa que los coeficientes de determinación son casi nulos. Sus resultados son exageradamente bajos, esto quiere decir que ninguna de estas variables tiene una fuerte relación con nuestra variable objetivo que es el precio. El coeficiente de determinación más alto es de tan solo 0.017424, lo cual tan solo explica el 1.74% de la variación en el precio con la variable de calculated host listings count.

Después se generó un modelo de regresión lineal múltiple. Este con el fin de tener un modelo más certero en cuanto a la variable de precio. En este modelo se utilizaron 5 variables, excluyendo de las variables listadas en la lista únicamente la variable de number of reviews, siendo la que menor correlación tuvo con la variable respuesta. El modelo que se generó es el siguiente:





Este modelo, obtuvo un coeficiente de determinación de 0.0267, logrando explicar el 2.67% de la variación en el precio en Airbnb México. Como se observa en la gráfica en el lado izquierdo, la variable que ayor influencia tiene en el modelo es el calculated host listing count. Este modelo de regresión lineal múltiple logra ser un modelo más certero, sin embargo sigue siendo bastante incierto. Por lo tanto, no se considera un modelo adecuado para observar la variación en el precio.

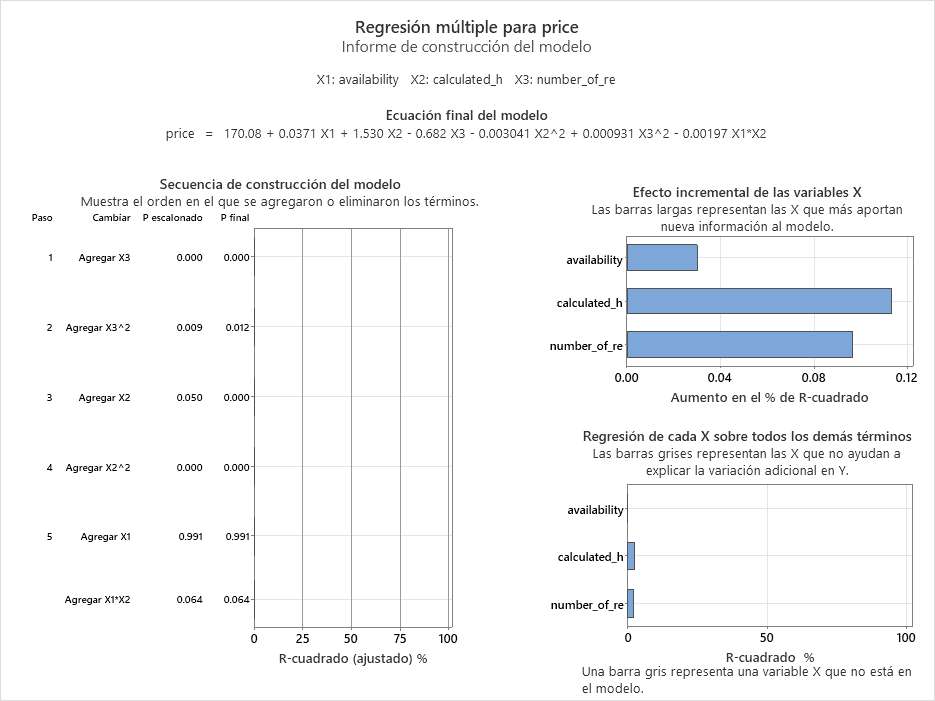
Ahora se hizo un análisis similar al pasado, sin embargo con otras 2 ubicaciones, comparando qué variables son más significantes en un lugar y otro, buscando encontrar patrones o similitudes.

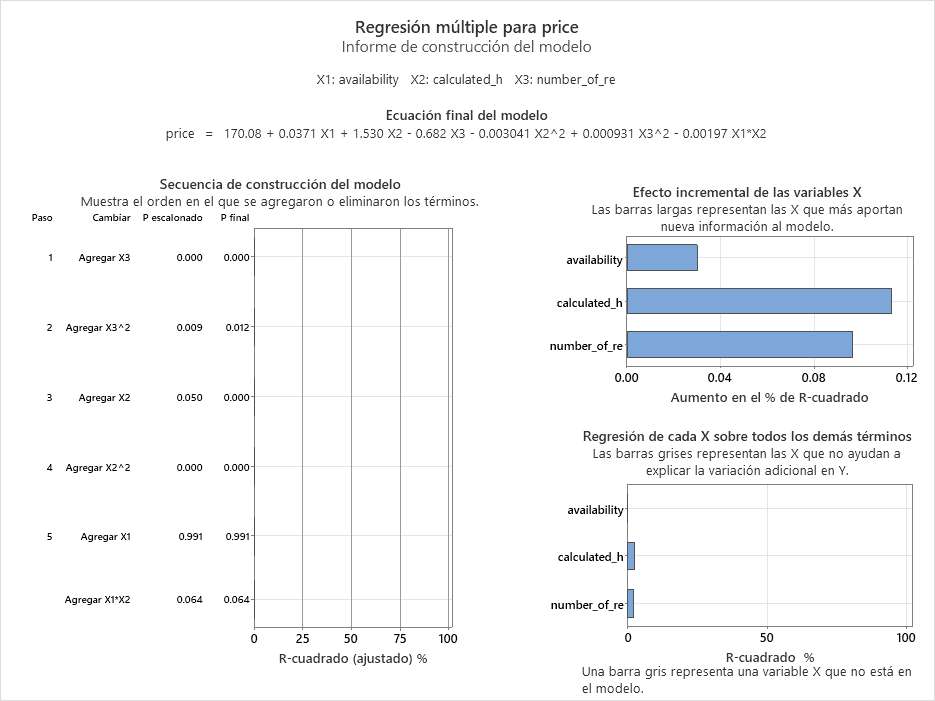
**MILÁN**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coeficiente de Correlación** | **Coeficiente de Determinación** |
| Minimum nights | -0.011 | 0.000121 |
| Number of reviews. | -0.029 | 0.000841 |
| Reviews per month. | -0.011 | 0.000121 |
| Calculated host listings count. | 0.017 | 0.000289 |
| Availability 365. | 0.003 | 0.000009 |
| Number of reviews ltm. | -0.021 | 0.000441 |

En esta tabla se analizan las mismas variables, pero en Airbnb de Milán, Italia. Aquí se observa que la mayoría de las correlaciones son negativas. Por lo tanto, entre más crece el valor de la columna de precio, menores son los valores en las columnas de minimum nights, number of reviews, reviews per month y number of reviews ltm. Sin embargo, las correlaciones son muy débiles. Por ello, los coeficientes de determinación son extremadamente bajos, por lo que tan solo logra explicar a lo mucho un 0.08% de la varianza en el precio con la variable seleccionada. Por lo tanto no podemos obtener un buen modelo de regresión lineal simple con estas variables, teniendo la variable de precio como la variable respuesta.

El mejor modelo de regresión lineal con estas variables es el siguiente:

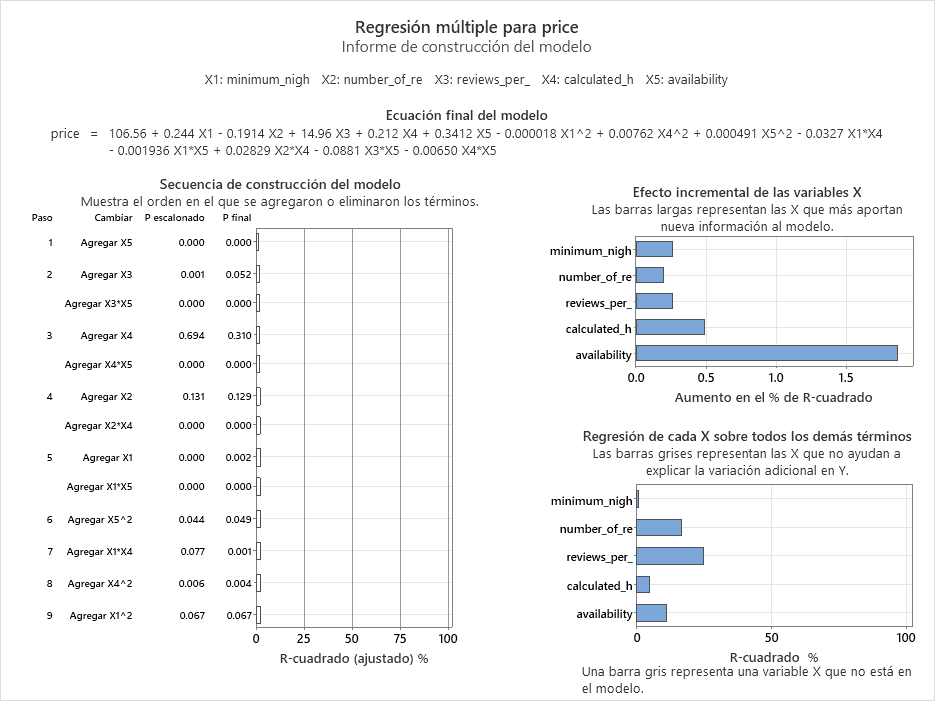


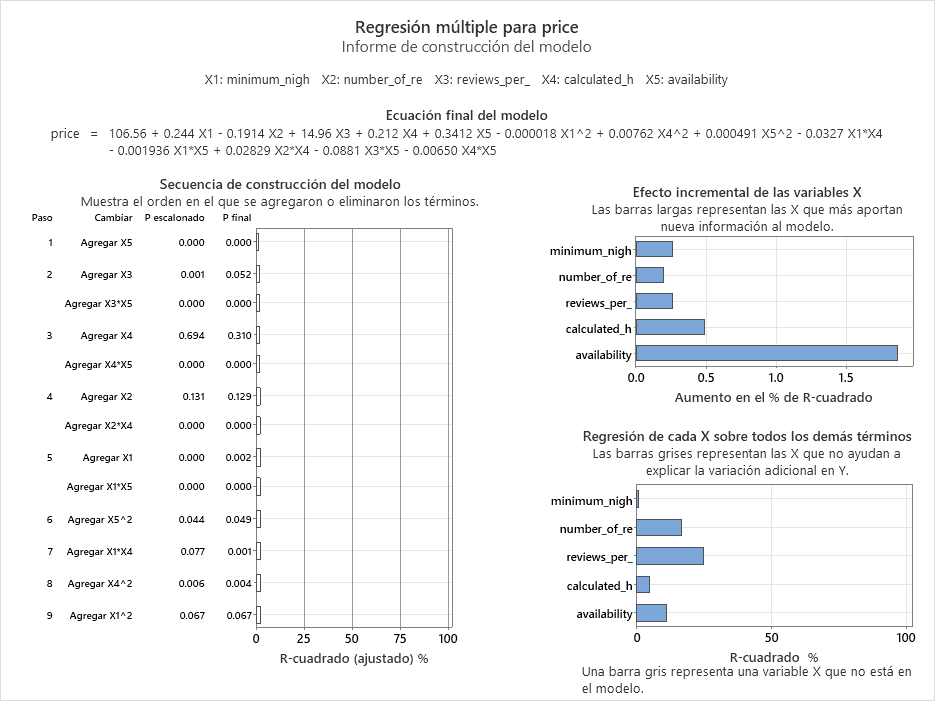
Con la base de datos de Airbnb de Milán, el mejor modelo consigue un coeficiente de determinación de 0.0024, Lo cual significa que el modelo tan solo logra explicar el 0.24% de la variación en el precio. Las variables que se toman en cuenta son “availability 365”, “calculated host listings” y “number of reviews” Como se observa en la tabla a continuación, la variable con mayor influencia en el modelo es la de “calculated host listings”, sin embargo el modelo sigue teniendo un coeficiente de determinación muy bajo por lo cual el modelo no se considera certero.

**MONTREAL**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coeficiente de Correlación** | **Coeficiente de Determinación** |
| Minimum nights | -0.003 | 0.000009 |
| Number of reviews. | 0.014 | 0.000196 |
| Reviews per month. | 0.043 | 0.001849 |
| Calculated host listings count. | 0.019 | 0.000361 |
| Availability 365. | 0.123 | 0.015129 |
| Number of reviews ltm. | 0.033 | 0.001089 |

En esta tercera tabla, analizamos las mismas variables de Airbnb en Montreal, Canadá. En esta tabla, observamos que a diferencia de la tabla pasada, la gran mayoría de las variables tienen una correlación positiva con la variable precio. Lo cual significa que entre mayor sea el valor de una de estas variables, el precio tiende a ser mayor también. Sin embargo, los coeficientes de correlación son muy bajos. La correlación más alta se presenta en la variabilidad de Availability 365 con el precio. Sin embargo, la correlación no es muy fuerte. Por lo tanto, ese modelo de regresión lineal simple, tomando esta variable como la independiente y la variable respuesta siendo precio, el modelo tan solo logra explicar un 1.51% de la variación en el precio. Por lo tanto no se considera un modelo certero. De acuerdo a las variables obtenidas, se busca una optimización mediante un modelo de regresión lineal múltiple. El mejor resultado obtenido es el siguiente:



De esta ecuación, se obtiene una gráfica donde se visualiza la influencia de las variables sobre el modelo como se muestra en la gráfica a la izquierda. En ella se observa que la variable de disponibilidad en el año, es la que mayor influencia tiene sobre el precio del inmueble. Sin embargo,observamos el grado de la influencia que es cerca de 2%, lo cuál es muy poco. El modelo completo consigue tener un coeficiente de determinación de 0.242, lo cuál logra explicar tan solo el 2.42% del comportamiento de la variación en el precio. 

En este análisis lamentablemente no se logró obtener un modelo con un coeficiente de determinación alto, por lo que no son certeros al explicar la variación del precio en los diferentes inmuebles alrededor del mundo. Sin embargo, se encontraron algunas similitudes. Se encontró que en los 3 análisis la variable más significativa en cuanto a la influencia en los modelos generados, fue el calculated host listing count.

Por otro lado, se observa que incluso esa variable tiene muy poca correlación. Por lo tanto ninguno de los modelos generados son adecuados para observar el comportamiento del precio. Con esto se concluye que la información encontrada en estas bases de datos, no es suficiente para generar un modelo de regresión lineal adecuado para la variable de precio.