

Gabriela Cantú Sánchez
19/09/2022
A01735256

Gestión de proyectos de plataformas tecnológicas

Es importante mencionar que para el análisis realizado no se reemplazaron los valores atípicos con el propósito de obtener diferentes resultados con métodos distintos. A través de la modelación de diferentes variables independientes con la variable dependiente de precio fue posible encontrar los coeficientes de correlación de Pearson del modelo. El coeficiente de correlación es una prueba estadística que permite medir la relación entre variables. Como variables independientes se tomaron en cuenta; carat, depth, table, x, y, y z. Cabe recalcar que se llevaron a cabo tanto modelos simples como múltiples para poder hacer combinaciones de variables independientes y encontrar el modelo ideal. Como se puede observar, el modelo con el coeficiente de correlación más alto fue el modelo múltiple en el que se incluyen todas las variables independientes incluidas en la base de datos.

Modelo	Coeficiente de correlación		Coeficiente de determinación	
Precio, tamaño, carat, depth, table	0.926822529	92.7%	0.8590000	86%
Precio, tamaño, carat, depth	0.92520268	92.5%	0.8560000	86%
Precio, tamaño, table y carat	0.92520268	92.5%	0.8560000	86%
Precio, tamaño y carat	0.924121204	92.4%	0.8540000	85%
Precio, carat, depth	0.922496612	92.2%	0.8510000	85%
Precio, carat, table	0.922496612	92.2%	0.8510000	85%
Precio, tamaño (x,y,z)	0.884872872	88.5%	0.7830000	78%
Precio y carat	0.849312472	84.9%	0.7213317	72%
Precio y x	0.782217539	78.2%	0.6118643	61%
Precio y y	0.748953557	74.9%	0.5609314	56%
Precio y z	0.741722381	74.2%	0.5501521	55%
Precio, depth, table	0.130384048	13.0%	0.0170000	2%
Precio y table	0.016164063	1.6%	0.0002613	0%
Precio y depth	0.000113312	0.0%	0.0000000	0%

Asimismo, gracias a los modelos simples se pudo obtener el coeficiente de correlación de las variables individualmente. Con lo anterior fue posible encontrar que carat es la variable independiente que más influye sobre el precio de venta ya que tiene un coeficiente de correlación de .84 y un coeficiente de determinación de .72. Es importante mencionar que tamaño toma en cuenta 3 variables diferentes, por lo que sería considerado un modelo múltiple. De cualquier manera, después de la variable carat las variables con más influencia sobre el precio de venta de los diamantes son x, y, y z. Es decir, las variables que se toman en cuenta para calcular el tamaño. Asimismo, es importante mencionar que se obtuvieron coeficientes muy altos en la mayoría de los casos.

Estos coeficientes demuestran que hay una relación alta y positiva entre las variables y el precio a excepción de las variables de *depth* y *table*. Al conocer los coeficientes de correlación entre variables o de influencia de la independiente en la dependiente podemos generar modelos múltiples que tal vez permitan explicar de mayor manera el modelo.

Se puede observar que el modelo que más permite explicar la influencia de variables en el precio de los diamantes es el múltiple en el que se toman en cuenta las variables de *tamaño*, *carat*, *depth* y *table*. Esto se concluye debido a que tiene un coeficiente de correlación de 0.92 y uno de determinación de 0.85, siendo el más alto de los modelos generados. De cualquier manera, considero pertinente mencionar que otros modelos con menos variables son casi igual de eficaces. Por ejemplo, el modelo que incluye el *tamaño* y el *carat* como variables independientes solamente es .002 menor en cuanto a correlación y 0.005 en cuanto a determinación a comparación con el modelo anteriormente mencionado. Dado que tiene menos variables es más sencillo manejar el modelo y tener las variables para poder predecir el precio del diamante a futuro. Por ende, sería recomendable trabajar con ese modelo obteniendo algo como una ganancia-pérdida ya que se pierde muy poco en los coeficientes pero se gana sencillez para trabajar con el modelo.