

**Actividad 1: Análisis de correlación**

Alberto Rodríguez Porras A01721497

26 de septiembre del 2022

Plataformas de analítica de negocios para organizaciones

Profesor: Alfredo García Suárez

En este trabajo, se analizó una base de datos acerca de diamantes. En este formato Excel, se observaban diversas columnas, que se intentan relacionar con el precio del diamante. Las columnas en el Excel son carat, cut, color, clarity, depth, table, Price, x, y, z. De estas variables, analizamos las cuantitativas para buscar una correlación con la variable “Price” que nos indica el precio del diamante. De esta manera, podremos observar que factores influyen mayormente en el precio de uno.

El objetivo es encontrar las variables que puedan generar un modelo que explique de mejor manera el comportamiento del precio del diamante, basándonos únicamente en la correlación. Por lo tanto, se realizó un análisis done encontramos 5 excelentes modelos, todos con una correlación muy fuerte y positiva.

|  |  |
| --- | --- |
| **Variables a buscar correlación con precio:** | **Correlación (R)** |
| Carat | 0.92 |
| Depth | 0.01 |
| Table | 0.13 |
| Carat y Depth | 0.92 |
| Carat y Table | 0.92 |
| Depth y Table | 0.13 |
| X,Y,Z | 0.88 |
| X,Y,Z, Carat y Depth | 0.93 |
| X,Y,Z, Carat, Table y Depth | 0.93 |

El primer modelo uno de regresión lineal simple, utilizando únicamente la variable “carat”, obteniendo una correlación de 0.92. Después encontramos que las variables “carat” y “depth”, al igual que las variables “carat” y “table”, generan un modelo que igualmente tiene una correlación de 0.92. Por último, se obtuvo igualmente otros dos modelos con las variables “x”, “y”, “z”, “carat” y “depth”, y otro con “x”, “y”, “z”, “carat”, “depth” y “table” que obtienen un resultado de correlación de 0.93.

Los 5 modelos obtienen una correlación bastante fuerte, por lo que nos indica que el comportamiento en las variables indicadas por modelo tiene una gran influencia en la manera en que se comporta la variable del precio. Por lo tanto, analizando los 5 modelos, el que considero como mejor modelo es el primero, donde únicamente se toma en cuenta la variable de carat para observar el comportamiento de precio. Este modelo obtuvo una correlación bastante alta, mostrando claramente que es la variable que mayor influencia tiene sobre el precio. Además, es el modelo más sencillo, por lo que es preferible para explicárselo e interpretárselo a algún ejecutivo o empresa. Considero yo que la variable “carat” es responsable por la alta correlación en todos los modelos, y pienso que esa debería de ser la única variable por observar.