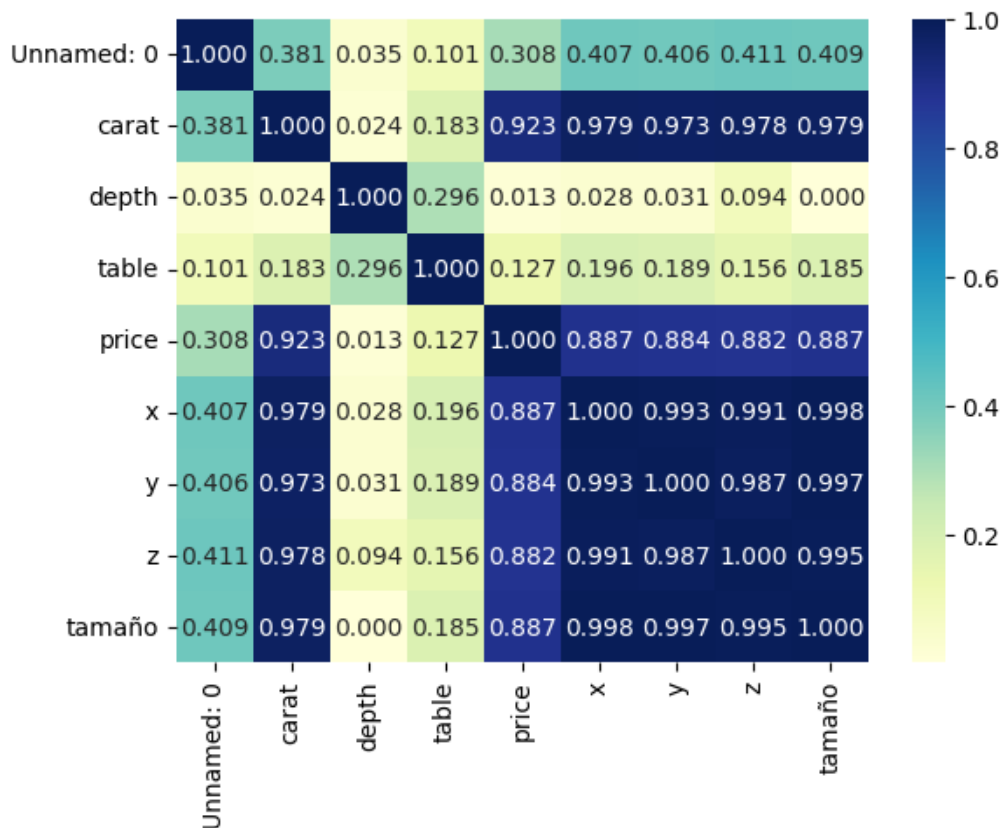


### Actividad 1 (Análisis de correlación utilizando plataformas de Analítica: Python)

Para esta actividad se hizo uso de la base de datos diamonds.csv, el cual contenía información sobre características de los diamantes, tales como depth, price, tamaño, etc. En primera instancia se realizó un análisis para identificar y eliminar los valores nulos. Este procedimiento se hizo de igual manera para localizar y lidiar con los outliers. Finalmente se obtuvieron los coeficientes de determinación, así como los de correlación de cada variable y se realizaron diferentes modelos de regresión lineal simple y múltiple. A continuación, se muestran los resultados obtenidos:

#### Correlaciones por variables



### Regresiones Múltiples

<b>Variables</b>	<b>Coef. De Determinación</b>	<b>Coef. Correlación</b>	<b>Tipo de Correlación</b>
<b>Price~Carat</b>	0.853	0.9235799911215054	<b>Correlación Fuerte Positiva</b>
<b>Price~ Depth</b>	0.000	0.0	<b>No existe correlación</b>
<b>Price~Table</b>	0.016	0.12649110640673517	<b>Correlación Débil Positiva</b>
<b>Price~ Carat y depth</b>	0.854	0.9241212041718337	<b>Correlación Fuerte Positiva</b>
<b>Price~ Carat y table</b>	0.855	0.9246621004453465	<b>Correlación Fuerte Positiva</b>
<b>Price~depth y table</b>	0.017	0.130384048104053	<b>Correlación Débil Positiva</b>
<b>Price~ tamaño (x, y, z)</b>	0.787	0.887130204648675	<b>Correlación Fuerte Positiva</b>
<b>Price~ tamaño, carat y depth</b>	0.862	0.9284395510748128	<b>Correlación Fuerte Positiva</b>
<b>Price~ tamaño, carat, depth y table</b>	0.865	0.9300537618869137	<b>Correlación Fuerte Positiva</b>

Después de analizar los modelos anteriores de regresión lineal simple y multiple, se logro observar que las variables que tienen mayor relación con “Price” son: Carat, Table y tamaño. Estas variables obtuvieron las correlaciones más altas en relación con el precio; Por otro lado, la variable depth es la que presenta una menor correlación con Price.

Podemos observar que al juntar estas variables en un solo modelo de regresión múltiple **(Price~ tamaño, carat y depth)** y **(Price~ tamaño, carat, depth y table)** la correlación se mantiene positiva y con un valor alto.

Se puede concluir que existen diferentes modelos que nos permiten comprender el comportamiento de Price con respecto a otras variables; Uno de los mejores es **(Price~ tamaño, carat, depth y table)**, el cual contiene el coeficiente de correlación más alto a comparación de los demás. Sin embargo, es importante considerar que entre más variables agreguemos al modelo, corremos el riesgo de que crezca nuestra varianza y exista un mayor error en nuestras predicciones futuras. Por lo que es importante considerar otros indicadores en el modelo como el BIC, AIC, etc.