**INFORME PROYECTO #2 Y #3**

**JUAN CAMILO GUTIERREZ VIVEROS - 2159874**

**ANDRES FELIPE ROJAS – 2160**

**PUNTO 1 - Analisis descriptivo del dataset**

**el dataset tiene 18.249 registros y 14 columnas, lo que es una muestra amplia y representativa.**

**Ninguna columna tiene valores nulos**

**Heatmap**

**Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.**

**La matriz de correlación muestra cómo se relacionan las variables numéricas del dataset entre sí.**

**El color representa la fuerza y dirección de la relación:**

**Cercano a 1 → fuerte correlación positiva (aumentan juntas).**

**Cercano a -1 → fuerte correlación negativa (una sube, la otra baja).**

**Cercano a 0 → sin relación lineal relevante**

**PUNTO 2 - Limpieza y Normalización de Datos**

**Para la normalización se hicieron los siguientes procesos detallados en el archivo limpieza.ipynb**

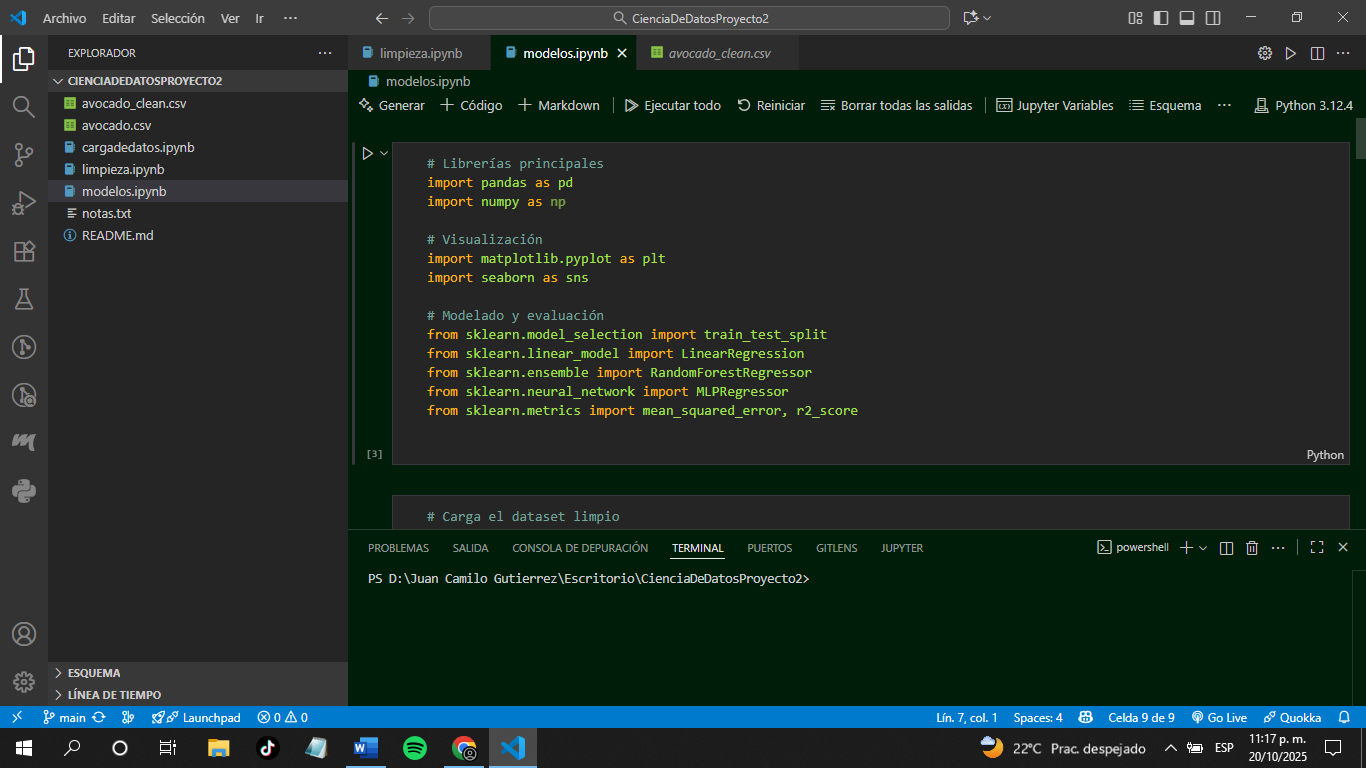
**Texto

El contenido generado por IA puede ser incorrecto.**

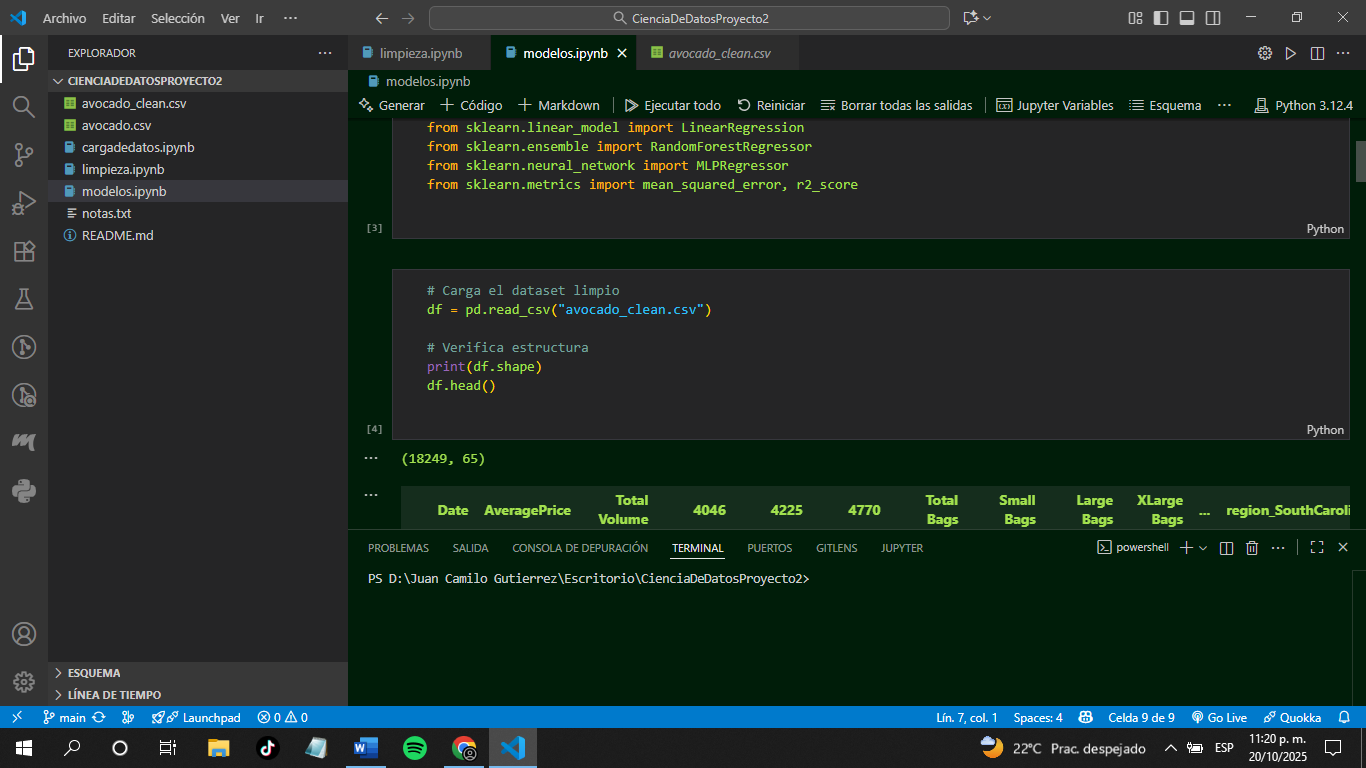
**PUNTO 3**

Se implementan los 3 modelos de machine learning, asimismo se analiza y compara sus métricas de desempeño.

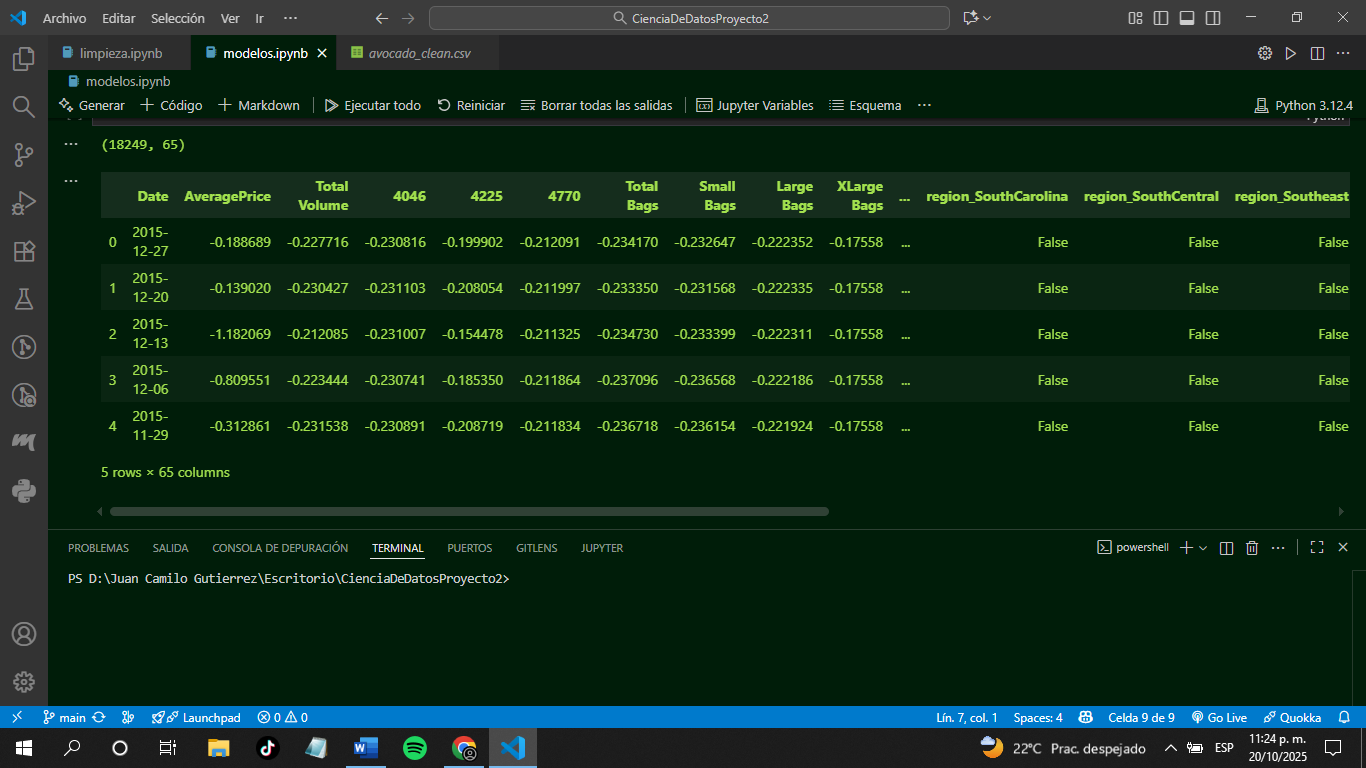
Se inicializa creando en la carpeta del proyecto el archivo *modelo.ipynb*, luego se importan las librerías necesarias (pandas, sklearn, matplotlib, etc.), se define estado aleatorio y muestra %matplotlib inline para Jupyter. Esta es la base del entorno:



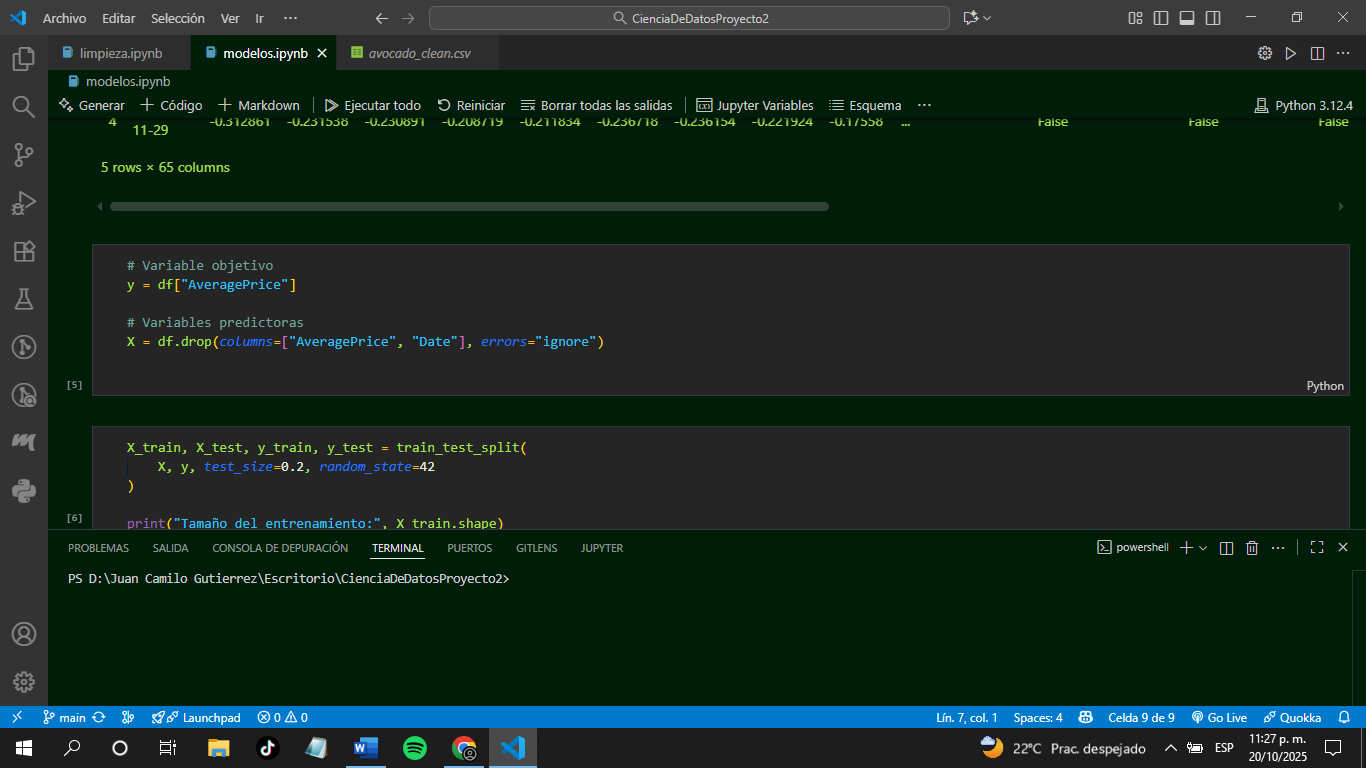
Se carga el dataset limpio (*avocado\_clean.csv*) para el respectivo uso en los modelos:



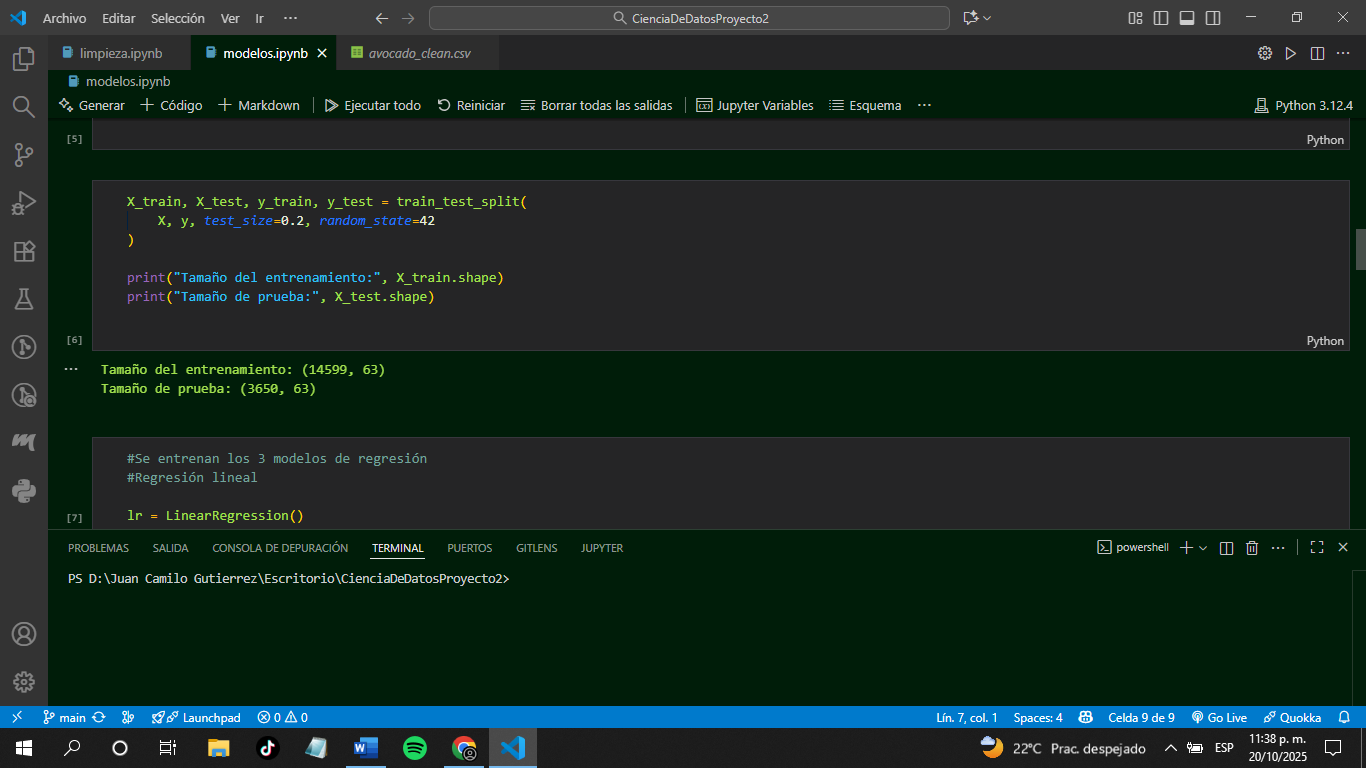
La celda anteriormente mostrada contiene *pd.read\_csv("avocado\_clean.csv"), df.head()* y *df.describe()* para comprobar forma, columnas y estadísticas básicas. Aquí se verifica que el dataset cargó correctamente:



Pequeña celda en a cual se separa la variable objetivo y = AveragePrice y las predictores X, y que se elimina Date si existe:



El siguiente fragmento de código crea X y Y, y realiza train\_test\_split (80/20). Muestra tamaños de train/test, también indica que los datos están listos para entrenar:



Están listas las métricas que se usarán: R², MSE, MAE.

**R² (Coeficiente de determinación):** mide la proporción de la varianza de la variable objetivo que el modelo logra explicar. Rango: 0 a 1. Cuanto más cercano a 1, mejor es el ajuste del modelo.

**MSE (Mean Squared Error / Error cuadrático medio):** mide el promedio de los errores al cuadrado entre los valores reales y predichos. Cuanto más bajo el MSE, menor es el error promedio.

**MAE (Mean Absolute Error / Error absoluto medio):** mide el error promedio en las predicciones en las mismas unidades del valor real.

