



**Matriculas:** 2022-0457

**Estudiante:** Stiven De La Rosa Brito

**Carrera:** Desarrollo de software.

**Materia:** Minería de Datos e Inteligencia

**Sección:** Virtual

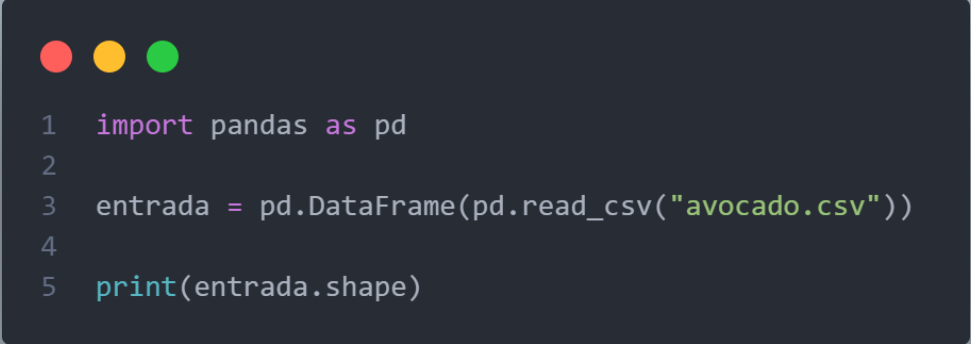
**Profesor/a:** Juan Luis Restituyo

**Trabajo :** Practica 3 - Exploración de Datos con  
Python

**Fecha:** 14/6/2023

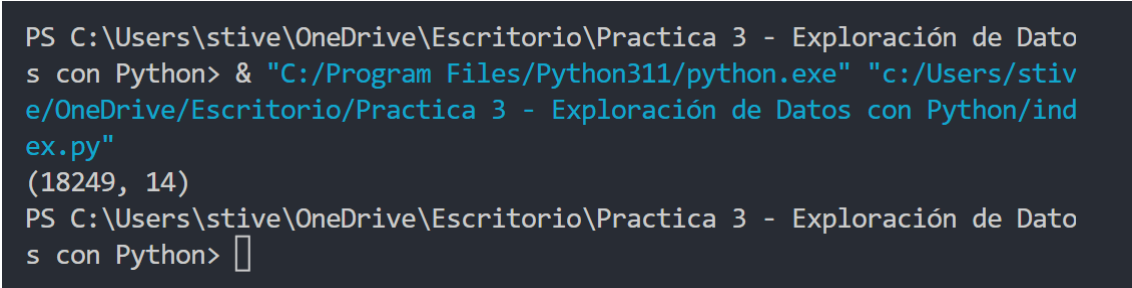
## Realizar en Python lo siguiente

### 1. Obtener Cuantas filas y cuantas columnas tiene el conjunto de datos

A dark-themed terminal window with three colored window control buttons (red, yellow, green) in the top-left corner. It contains five lines of Python code:

```
1 import pandas as pd
2
3 entrada = pd.DataFrame(pd.read_csv("avocado.csv"))
4
5 print(entrada.shape)
```

### Salida:

A dark-themed terminal window showing the execution of a Python script. The prompt is 'PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python>'. The command executed is '& "C:/Program Files/Python311/python.exe" "c:/Users/stive/OneDrive/Escritorio/Practica 3 - Exploración de Datos con Python/index.py"'. The output is '(18249, 14)'. The prompt is repeated at the bottom.

```
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> & "C:/Program Files/Python311/python.exe" "c:/Users/stive/OneDrive/Escritorio/Practica 3 - Exploración de Datos con Python/index.py"
(18249, 14)
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> 
```

## 2. Mostrar los primeros 100 registros

```
1 import pandas as pd
2
3 entrada = pd.DataFrame(pd.read_csv("avocado.csv"))
4
5 print(entrada.head(100))
```

Salida:

```
PS C:\Users\> & "C:/Program Files/Python311/python.exe" "c:/Users/stivs con Python> &
e/OneDrive/Escritorio/Practica 3 - Exploración de Datos con Python/ind
ex.py"rive/Escritorio/Practica 3 - Exploración de Datos con Python/in
   Unnamed: 0      Date  AveragePrice  ...      type  year  re
gion
0           0  2015-12-27           1.33  ...  conventional  2015  Al
bany
1           1  2015-12-20           1.35  ...  conventional  2015  Al
bany
2           2  2015-12-13           0.93  ...  conventional  2015  Al
bany
3           3  2015-12-06           1.08  ...  conventional  2015  Al
bany
4           4  2015-11-29           1.28  ...  conventional  2015  Albany
..         ...      ...           ...  ...      ...      ...      ...
95          43  2015-03-01           0.99  ...  conventional  2015  Atlanta
96          44  2015-02-22           1.10  ...  conventional  2015  Atlanta
97          45  2015-02-15           1.06  ...  conventional  2015  Atlanta
98          46  2015-02-08           1.03  ...  conventional  2015  Atlanta
99          47  2015-02-01           0.96  ...  conventional  2015  Atlanta

[100 rows x 14 columns]
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> 
```

### 3. Mostrar los últimos 20 registros

```
1 import pandas as pd
2
3 entrada = pd.DataFrame(pd.read_csv("avocado.csv"))
4
5 print(entrada.tail(20))
```

Salida:

```
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> & "C:/Program Files/Python311/python.exe" "c:/Users/stive/OneDrive/Escritorio/Practica 3 - Exploración de Datos con Python/index.py"
   Unnamed: 0  Date  AveragePrice  Total Volume  4046  ...  Large Bags  XLarge Bags  type  year  region
18229      4  2018-02-25      1.49    301985.61  34200.18  ...  118314.77    256.12  organic  2018    West
18230      5  2018-02-18      1.64    224798.60  30149.00  ...   35257.73     39.97  organic  2018    West
18231      6  2018-02-11      1.47    275248.53  24732.55  ...   99810.80    251.60  organic  2018    West
18232      7  2018-02-04      1.41    283378.47  22474.66  ...  134666.91    510.41  organic  2018    West
18233      8  2018-01-28      1.80    185974.53  22918.40  ...   51986.86    102.38  organic  2018    West
18234      9  2018-01-21      1.83    189317.99  27049.44  ...   51947.50    228.27  organic  2018    West
18235     10  2018-01-14      1.82    207999.67  33869.12  ...   37133.99     12.12  organic  2018    West
18236     11  2018-01-07      1.48    297190.60  34734.97  ...   95544.39     24.18  organic  2018    West
18237      0  2018-03-25      1.62    15303.40   2325.30  ...     236.64      0.00  organic  2018  WestTexNewMexico
18238      1  2018-03-18      1.56    15896.38   2055.35  ...     226.67      0.00  organic  2018  WestTexNewMexico
18239      2  2018-03-11      1.56    22128.42   2162.67  ...     252.25      0.00  organic  2018  WestTexNewMexico
18240      3  2018-03-04      1.54    17393.30   1832.24  ...     253.56      0.00  organic  2018  WestTexNewMexico
18241      4  2018-02-25      1.57    18421.24   1974.26  ...     266.06      0.00  organic  2018  WestTexNewMexico
18242      5  2018-02-18      1.56    17597.12   1892.05  ...     223.18      0.00  organic  2018  WestTexNewMexico
18243      6  2018-02-11      1.57    15986.17   1924.28  ...     256.22      0.00  organic  2018  WestTexNewMexico
18244      7  2018-02-04      1.63    17074.83   2046.96  ...     431.85      0.00  organic  2018  WestTexNewMexico
18245      8  2018-01-28      1.71    13888.04   1191.70  ...     324.80      0.00  organic  2018  WestTexNewMexico
18246      9  2018-01-21      1.87    13766.76   1191.92  ...      42.31      0.00  organic  2018  WestTexNewMexico
18247     10  2018-01-14      1.93    16205.22   1527.63  ...      50.00      0.00  organic  2018  WestTexNewMexico
18248     11  2018-01-07      1.62    17489.58   2894.77  ...      26.01      0.00  organic  2018  WestTexNewMexico

[20 rows x 14 columns]
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> █
```

#### 4. Cual es el precio mínimo, máximo y promedio del aguacate en ese conjunto de datos

```
1 import pandas as pd
2
3 entrada = pd.DataFrame(pd.read_csv("avocado.csv"))
4 AveragePrice = entrada["AveragePrice"]
5
6 valor_minimo = AveragePrice.min()
7 valor_maximo = AveragePrice.max()
8 precio_promedio = AveragePrice.mean()
9
10
11 print("el precio mínimo del aguacate es de : " , valor_minimo )
12 print("el Precio máximo del aguacate es de: " , valor_maximo)
13 print("Precio promedio del aguacate:" , precio_promedio)
```

#### Resultado

```
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> & "C:/Pr
ve/OneDrive/Escritorio/Practica 3 - Exploración de Datos con Python/index.py"
el precio mínimo del aguacate es de : 0.44
el Precio máximo del aguacate es de: 3.25
Precio promedio del aguacate: 1.405978409775878
PS C:\Users\stive\OneDrive\Escritorio\Practica 3 - Exploración de Datos con Python> □
```

5. Generar un gráfico de tipo scatter usando para la x la variable 'year' y para y la variable 'AveragePrice' para 3 regiones (las que usted elija),

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 salida = plt.subplot()
5 entrada = pd.DataFrame(pd.read_csv("avocado.csv"))
6
7 # print(entrada["region"].to_string())
8
9 Sacramento = entrada[entrada["region"] == "Sacramento"]
10 Southeast = entrada[entrada["region"] == "Southeast"]
11 Mexico = entrada[entrada["region"] == "WestTexNewMexico"]
12
13 g_Sacramento = Sacramento.plot(x = "year" , y = "AveragePrice" , kind = "scatter" , color = "red" , ax = salida)
14 g_Tampa = Southeast.plot(x = "year" , y = "AveragePrice" , kind = "scatter" , color = "green" , ax = salida)
15 g_Miami = Mexico.plot(x = "year" , y = "AveragePrice" , kind = "scatter" , color = "blue" , ax = salida)
16
17 plt.show()
```

Salida:

