

Santiago Gabian Perez A01658280

Grupo 222

Mayo 2022

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Ciudad de México

Herramientas computacionales: el arte de la analítica

TC1002.S

Carga de datos:

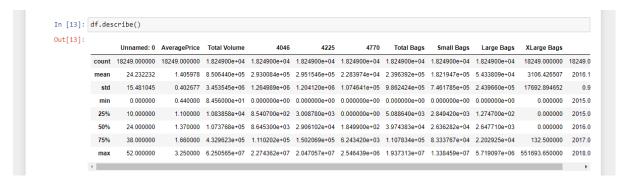


Cantidad de Datos:

```
In [24]: df.count()
Out[24]: Unnamed: 0
                                 18249
                                 18249
18249
            Date
            AveragePrice
            Total Volume
                                 18249
            4946
                                 18249
            4225
4770
                                 18249
            Total Bags
Small Bags
Large Bags
                                 18249
                                 18249
18249
            XLarge Bags
                                 18249
            type
year
region
                                 18249
            dtype: int64
```

En el archivo de datos de avocado se encuentran 18249 datos de la cual todos excepto región y tipo son enteros y región y tipo son strings.

Análisis de variables:



Media, Mediana y Desviación estándar:

```
In [22]: df.std(axis = 0)
             C:\Users\santi\anaconda3\envs\Semana tec\lib\site-packages\ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only v alid columns before calling the reduction.

"""Entry point for launching an IPython kernel.
Out[22]: Unnamed: 0
                                     1.548104e+01
              AveragePrice
                                     4.026766e-01
              Total Volume
                                     3.453545e+06
              4046
4225
                                     1.204120e+06
              4770
                                     1.074641e+05
                                     9.862424e+05
7.461785e+05
              Total Bags
              Small Bags
              Large Bags
XLarge Bags
                                     2.439660e+05
              year
                                     9.399385e-01
              dtype: float64
In [20]: df.mean(axis = 0)
              in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only v alid columns before calling the reduction.

"""Entry point for launching an IPython kernel.
              AveragePrice
                                            1.405978
              Total Volume
4046
                                    850644.013009
293008.424531
              4225
                                     295154.568356
                                     22839.735993
239639.202060
              4770
              Total Bags
              Small Bags
Large Bags
                                     182194,686696
                                      54338.088145
3106.426507
              XLarge Bags
             year
dtype: float64
                                        2016.147899
In [21]: df.median(axis = 0)
             C:\Users\santi\anaconda3\envs\Semana tec\lib\site-packages\ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only v alid columns before calling the reduction.

""Entry point for launching an IPython kernel.
Out[21]: Unnamed: 0
                                          24.00
              AveragePrice
                                            1.37
                                     107376.76
              Total Volume
              4946
                                        8645.30
              4225
                                      29061.02
              4779
                                         184.99
              Total Bags
Small Bags
                                      39743.83
26362.82
              Large Bags
XLarge Bags
                                       2647.71
                                        0.00
2016.00
              dtype: float64
```

Esto nos dice que el precio promedio es alrededor de 1.4 y una desviación estándar de 0.402 lo cual nos dice que el precio no varía tanto. Con el precio promedio mínimo siendo 0.44 y el máximo 3.25. La poca varianza tiene que ver que ya es un precio promedio y los promedios varían menos que si fueran todos los datos sin promediar.

Luego en el Volumen total se puede ver que hay una gran desviación estándar dando a entender la gran varianza que hay de volúmenes totales.

También se puede ver que varia mucho él numero total de bolsas incluyendo grandes y chicas. Según los datos podemos concluir que normalmente se usan mucho más bolsas pequeñas que grandes. De que en promedio hay 239,639 bolsas totales y de esas 182,194 son chicas, 54,338 son grandes y 3,106 son extra grandes.

Así que en conclusión el precio promedio total no suele variar mucho estando en promedio en 1.4 mientras que el volumen total si. También podemos concluir de que de todas las bolsas compradas se compran mucho más que las grandes y extragrandes.