Pandas demo

Import of the libraries

```
import pandas as pd
```

Import of the dataset

```
df = pd.read_csv('/content/avocado.csv')
```

```
df
```

	Unnamed: 0	Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags
0	0	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	8696.87
1	1	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	9505.56
2	2	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	8145.35
3	3	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	5811.16
4	4	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	6183.95
...	...	...	...	...	...	...	...	...
18244	7	2018-02-04	1.63	17074.83	2046.96	1529.20	0.00	13498.67
18245	8	2018-01-28	1.71	13888.04	1191.70	3431.50	0.00	9264.84
18246	9	2018-01-21	1.87	13766.76	1191.92	2452.79	727.94	9394.11
18247	10	2018-01-14	1.93	16205.22	1527.63	2981.04	727.01	10969.54
18248	11	2018-01-07	1.62	17489.58	2894.77	2356.13	224.53	12014.15

18249 rows × 14 columns

```
df.tail(10)
```

	Unnamed: 0	Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags	S

```
df.shape
```

```
(18249, 14)
```

03-04

```
df.head(10)
```

	Unnamed: 0	Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags	Sm B
0	0	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	8696.87	8603
1	1	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	9505.56	9408
2	2	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	8145.35	8042
3	3	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	5811.16	5677
4	4	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	6183.95	5986
5	5	2015-11-22	1.26	55979.78	1184.27	48067.99	43.61	6683.91	6556
6	6	2015-11-15	0.99	83453.76	1368.92	73672.72	93.26	8318.86	8196
7	7	2015-11-08	0.98	109428.33	703.75	101815.36	80.00	6829.22	6266
8	8	2015-11-01	1.02	99811.42	1022.15	87315.57	85.34	11388.36	11104
9	9	2015-10-25	1.07	74338.76	842.40	64757.44	113.00	8625.92	8061

```
df.columns
```

```
Index(['Unnamed: 0', 'Date', 'AveragePrice', 'Total Volume', '4046', '4225',
       '4770', 'Total Bags', 'Small Bags', 'Large Bags', 'XLarge Bags', 'type',
       'year', 'region'],
      dtype='object')
```

```
df['AveragePrice'].nlargest(10)
```

```
14125    3.25
17428    3.17
14124    3.12
16055    3.05
16720    3.04
13037    3.03
16715    3.00
16985    3.00
14123    2.99
15814    2.99
Name: AveragePrice, dtype: float64
```

```
df.nsmallest(10, 'AveragePrice')
```

amed: 0	Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags	Sm E
43	2017-03-05	0.44	64057.04	223.84	4748.88	0.00	59084.32	63
47	2017-02-05	0.46	2200550.27	1200632.86	531226.65	18324.93	450365.83	11375
43	2017-03-05	0.48	50890.73	717.57	4138.84	0.00	46034.32	138

```
df['AveragePrice'].median()
```

```
1.37
```

02-26

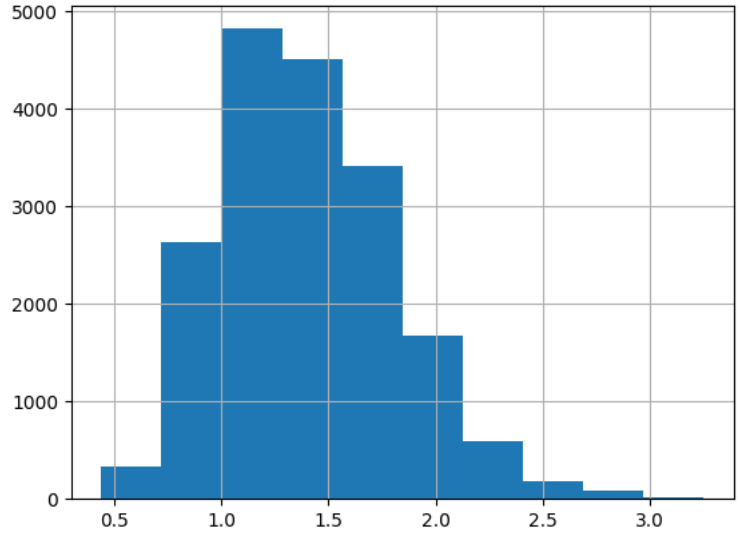
```
df.groupby(by='type').mean()
```

```
<ipython-input-12-dc68b57112ab>:1: FutureWarning: The default value of numeric_only :
  df.groupby(by='type').mean()
```

type	Unnamed: 0	AveragePrice	Total Volume	4046	4225	
conventional	24.236686	1.158040	1.653213e+06	578611.649925	574805.318859	454(
organic	24.227776	1.653999	4.781121e+04	7311.281600	15411.857724	26

```
df['AveragePrice'].hist()
```

```
<Axes: >
```




```
df['type']
```

```
0        conventional
1        conventional
2        conventional
3        conventional
4        conventional
            ...
18244         organic
18245         organic
18246         organic
18247         organic
18248         organic
Name: type, Length: 18249, dtype: object
```

```
df['type'].map({'conventional': True, 'organic': False})
```

```
0         True
1         True
2         True
3         True
4         True
         ...
18244    False
18245    False
18246    False
18247    False
```

```
18248    False
Name: type, Length: 18249, dtype: bool
```

```
df.median()
```

```
<ipython-input-22-6d467abf240d>:1: FutureWarning: The default value of numeric_only in DataFrame.median is deprecated. In a future
  df.median()
Unnamed: 0          24.00
AveragePrice         1.37
Total Volume    107376.76
4046              8645.30
4225             29061.02
4770               184.99
Total Bags       39743.83
Small Bags       26362.82
Large Bags        2647.71
XLarge Bags          0.00
year              2016.00
dtype: float64
```

```
df.describe()
```

	Unnamed: 0	AveragePrice	Total Volume	4046	4225	4770	Total Bags	Small Bags	Large Bags	
count	18249.000000	18249.000000	1.824900e+04	1.824900e+04	1.824900e+04	1.824900e+04	1.824900e+04	1.824900e+04	1.824900e+04	1
mean	24.232232	1.405978	8.506440e+05	2.930084e+05	2.951546e+05	2.283974e+04	2.396392e+05	1.821947e+05	5.433809e+04	
std	15.481045	0.402677	3.453545e+06	1.264989e+06	1.204120e+06	1.074641e+05	9.862424e+05	7.461785e+05	2.439660e+05	1
min	0.000000	0.440000	8.456000e+01	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	
25%	10.000000	1.100000	1.083858e+04	8.540700e+02	3.008780e+03	0.000000e+00	5.088640e+03	2.849420e+03	1.274700e+02	
50%	24.000000	1.370000	1.073768e+05	8.645300e+03	2.906102e+04	1.849900e+02	3.974383e+04	2.636282e+04	2.647710e+03	
75%	38.000000	1.660000	4.329623e+05	1.110202e+05	1.502069e+05	6.243420e+03	1.107834e+05	8.333767e+04	2.202925e+04	
max	52.000000	3.250000	6.250565e+07	2.274362e+07	2.047057e+07	2.546439e+06	1.937313e+07	1.338459e+07	5.719097e+06	5

```
df['Total Volume'].head(10)
```

```
0     64236.62
1     54876.98
2    118220.22
3     78992.15
4     51039.60
5     55979.78
6     83453.76
7    109428.33
8     99811.42
9     74338.76
Name: Total Volume, dtype: float64
```

```
df['Date'].min()
```

```
'2015-01-04'
```

```
df['Date'].max()
```

```
'2018-03-25'
```