```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import pandas as pd
```

```
ava=pd.read_csv("/content/drive/MyDrive/STUDY2/DATA ANALYSIS LAB/LABCYCLE/DATASETS/avocado_10thQuestion.csv")
```

```
ava
```

| | Unnamed: 0 | Date | AveragePrice | Total Volume | 4046 | 4225 | 4770 | Total Bags | Small Bags | Large Bags | XLarge Bags | type | year | region |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 12/27/2015 | 1.33 | 64236.62 | 1036.74 | 54454.85 | 48.16 | 8696.87 | 8603.62 | 93.25 | 0.0 | conventional | 2015 | Albany |
| **1** | 1 | 12/20/2015 | 1.35 | 54876.98 | 674.28 | 44638.81 | 58.33 | 9505.56 | 9408.07 | 97.49 | 0.0 | conventional | 2015 | Albany |
| **2** | 2 | 12/13/2015 | 0.93 | 118220.22 | 794.70 | 109149.67 | 130.50 | 8145.35 | 8042.21 | 103.14 | 0.0 | conventional | 2015 | Albany |
| **3** | 3 | 12/6/2015 | 1.08 | 78992.15 | 1132.00 | 71976.41 | 72.58 | 5811.16 | 5677.40 | 133.76 | 0.0 | conventional | 2015 | Albany |
| **4** | 4 | 11/29/2015 | 1.28 | 51039.60 | 941.48 | 43838.39 | 75.78 | 6183.95 | 5986.26 | 197.69 | 0.0 | conventional | 2015 | Albany |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **18244** | 7 | 2/4/2018 | 1.63 | 17074.83 | 2046.96 | 1529.20 | 0.00 | 13498.67 | 13066.82 | 431.85 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18245** | 8 | 1/28/2018 | 1.71 | 13888.04 | 1191.70 | 3431.50 | 0.00 | 9264.84 | 8940.04 | 324.80 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18246** | 9 | 1/21/2018 | 1.87 | 13766.76 | 1191.92 | 2452.79 | 727.94 | 9394.11 | 9351.80 | 42.31 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18247** | 10 | 1/14/2018 | 1.93 | 16205.22 | 1527.63 | 2981.04 | 727.01 | 10969.54 | 10919.54 | 50.00 | 0.0 | organic | 2018 | WestTexNewMexico |
| **18248** | 11 | 1/7/2018 | 1.62 | 17489.58 | 2894.77 | 2356.13 | 224.53 | 12014.15 | 11988.14 | 26.01 | 0.0 | organic | 2018 | WestTexNewMexico |

18249 rows × 14 columns

```
#a) How to identify the unique values in the region column.
pd.Series(ava["region"].unique())
```

```
0                 Albany
1                Atlanta
2     BaltimoreWashington
3                  Boise
4                 Boston
5        BuffaloRochester
6              California
7               Charlotte
8                Chicago
9        CincinnatiDayton
10               Columbus
11          DallasFtWorth
12                Denver
13               Detroit
14           GrandRapids
15             GreatLakes
16     HarrisburgScranton
17    HartfordSpringfield
18                Houston
19           Indianapolis
20           Jacksonville
21                LasVegas
22              LosAngeles
23              Louisville
24      MiamiFtLauderdale
25                Midsouth
26              Nashville
27       NewOrleansMobile
28                NewYork
29              Northeast
30     NorthernNewEngland
31                Orlando
32           Philadelphia
33          PhoenixTucson
34             Pittsburgh
35                 Plains
36               Portland
37      RaleighGreensboro
38        RichmondNorfolk
39                Roanoke
40             Sacramento
41               SanDiego
42          SanFrancisco
43                Seattle
44          SouthCarolina
45           SouthCentral
46              Southeast
47                Spokane
48                StLouis
49               Syracuse
50                  Tampa
```

```
51                TotalUS
52                   West
53       WestTexNewMexico
dtype: object
```

]:
```
#b) What is the maximum price for an avocado in the dataset.
ava["AveragePrice"].max()
```

]:
```
3.25
```

]:
```
#c) Identify the type distribution  and take a single avocado in the dataset and find out the median price ,mean, and standard deviation.
print(ava["type"].value_counts())
single=ava[ava["type"]=='conventional']
print(single["AveragePrice"].mean())
print(single["AveragePrice"].median())
print(single["AveragePrice"].std())
```

]:
```
conventional    9126
organic         9123
Name: type, dtype: int64
1.1580396668858208
1.13
0.26304060411401714
```

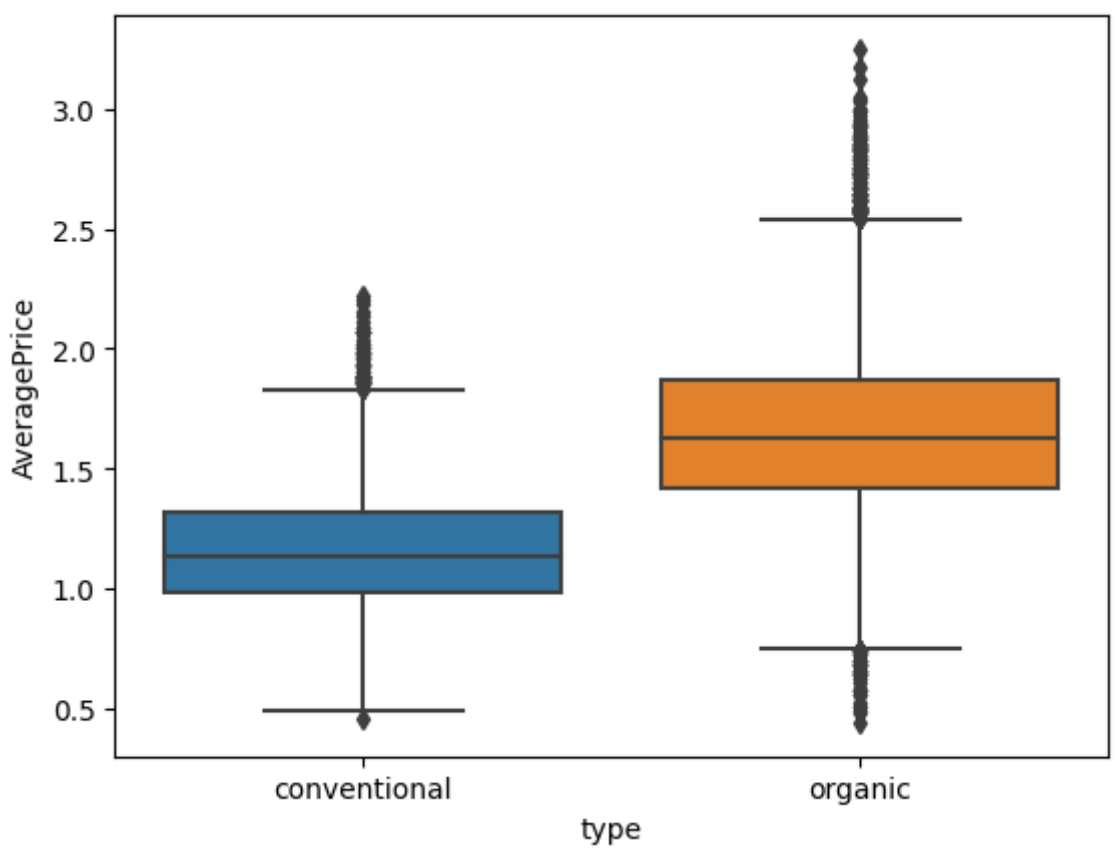
]:
```
#d) Find the highest, lowest price for conventional avocado’s in year with location.
conv=ava[ava["type"]=="conventional"]
p=conv.groupby(['year', 'region'])["AveragePrice"]
print(p.max())
print(p.min())
```

]:
```
year  region
2015  Albany                1.45
      Atlanta               1.15
      BaltimoreWashington   1.29
      Boise                 1.19
      Boston                1.28
                            ...
2018  Syracuse              1.45
      Tampa                 1.41
      TotalUS               1.20
      West                  1.08
      WestTexNewMexico      0.94
Name: AveragePrice, Length: 216, dtype: float64
year  region
2015  Albany                0.93
      Atlanta               0.93
      BaltimoreWashington   0.99
      Boise                 0.71
      Boston                0.94
                            ...
2018  Syracuse              1.10
      Tampa                 0.98
      TotalUS               0.87
      West                  0.83
      WestTexNewMexico      0.75
Name: AveragePrice, Length: 216, dtype: float64
```

]:
```
#e) Draw the plots of the distribution of average price for different types of Avocados
import seaborn as sns
sns.boxplot(x='type', y='AveragePrice', data=ava)
```

]:

]:




]:

```
#f) Find the correlation matrix to measure the strength of the correlation between variables.
ava.select_dtypes(include=['float64', 'int64']).corr()
```

]:

| | Unnamed: 0 | AveragePrice | Total Volume | 4046 | 4225 | 4770 | Total Bags | Small Bags | Large Bags | XLarge Bags | year |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unnamed: 0 | 1.000000 | -0.133008 | 0.014035 | 0.017628 | 0.019829 | 0.041752 | -0.002219 | 0.000347 | -0.009196 | -0.011546 | -0.171667 |
| AveragePrice | -0.133008 | 1.000000 | -0.192752 | -0.208317 | -0.172928 | -0.179446 | -0.177088 | -0.174730 | -0.172940 | -0.117592 | 0.093197 |
| Total Volume | 0.014035 | -0.192752 | 1.000000 | 0.977863 | 0.974181 | 0.872202 | 0.963047 | 0.967238 | 0.880640 | 0.747157 | 0.017193 |
| 4046 | 0.017628 | -0.208317 | 0.977863 | 1.000000 | 0.926110 | 0.833389 | 0.920057 | 0.925280 | 0.838645 | 0.699377 | 0.003353 |
| 4225 | 0.019829 | -0.172928 | 0.974181 | 0.926110 | 1.000000 | 0.887855 | 0.905787 | 0.916031 | 0.810015 | 0.688809 | -0.009559 |
| 4770 | 0.041752 | -0.179446 | 0.872202 | 0.833389 | 0.887855 | 1.000000 | 0.792314 | 0.802733 | 0.698471 | 0.679861 | -0.036531 |
| Total Bags | -0.002219 | -0.177088 | 0.963047 | 0.920057 | 0.905787 | 0.792314 | 1.000000 | 0.994335 | 0.943009 | 0.804233 | 0.071552 |
| Small Bags | 0.000347 | -0.174730 | 0.967238 | 0.925280 | 0.916031 | 0.802733 | 0.994335 | 1.000000 | 0.902589 | 0.806845 | 0.063915 |
| Large Bags | -0.009196 | -0.172940 | 0.880640 | 0.838645 | 0.810015 | 0.698471 | 0.943009 | 0.902589 | 1.000000 | 0.710858 | 0.087891 |
| XLarge Bags | -0.011546 | -0.117592 | 0.747157 | 0.699377 | 0.688809 | 0.679861 | 0.804233 | 0.806845 | 0.710858 | 1.000000 | 0.081033 |
| year | -0.171667 | 0.093197 | 0.017193 | 0.003353 | -0.009559 | -0.036531 | 0.071552 | 0.063915 | 0.087891 | 0.081033 | 1.000000 |

]:

```
#g) Find out the volume of avocado sales has increased in the last 5 years.
current_year = ava['year'].max()
five_years_ago = current_year - 5

last_5_years_volume = ava[ava['year'] >= five_years_ago]['Total Volume'].sum()
total_volume = ava['Total Volume'].sum()
last_5_years_volume - total_volume
```

]:

```
0.0
```