

Avocado Price Analysis

Jie Guo

October, 2020

CONTENT

1. Background
2. Data cleaning and Adaptation
3. Visualization
4. Cluster Analysis
5. Recommendations



Background

Stakeholder:

Hass Avocado

Missions:

1. How to price avocados in different cities?
2. How to price avocados in different seasons?
3. How to determine the supply of different avocados?



Data Cleaning:

According to our research, we drop "Total Bags", "Small Bags", "Large Bags" and "XLarge Bags":

```
avocado=avocado[["year", "Date", "AveragePrice", "Total Volume", "4046", "4225", "4770", "type", "region"]]  
avocado.head()
```

	year	Date	AveragePrice	Total Volume	4046	4225	4770	type	region
0	2015	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	conventional	Albany
1	2015	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	conventional	Albany
2	2015	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	conventional	Albany
3	2015	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	conventional	Albany
4	2015	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	conventional	Albany

Data Adaptation

Add a column called "No_label"

This column means the avocados that are not "4046", "4225" or "4770".

```
avocado["No_label"]=avocado.iloc[:,3]-avocado.iloc[:,4]-avocado.iloc[:,5]-avocado.iloc[:,6]  
avocado.head()
```

	year	Date	AveragePrice	Total Volume	4046	4225	4770	type	region	No_label
0	2015	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	conventional	Albany	8696.87
1	2015	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	conventional	Albany	9505.56
2	2015	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	conventional	Albany	8145.35
3	2015	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	conventional	Albany	5811.16
4	2015	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	conventional	Albany	6183.95

Because we need the numeric parameter to assign the color, we change "conventional" type into 0 and "organic" type equals to 1. We use for loop and if statement. The "18249" is the length of dataset. So I set for loop as range(18249).

```
j=0
for i in range(18249):
    if (avocado.iloc[i,7] == "conventional"):
        avocado.iloc[i,7]=0

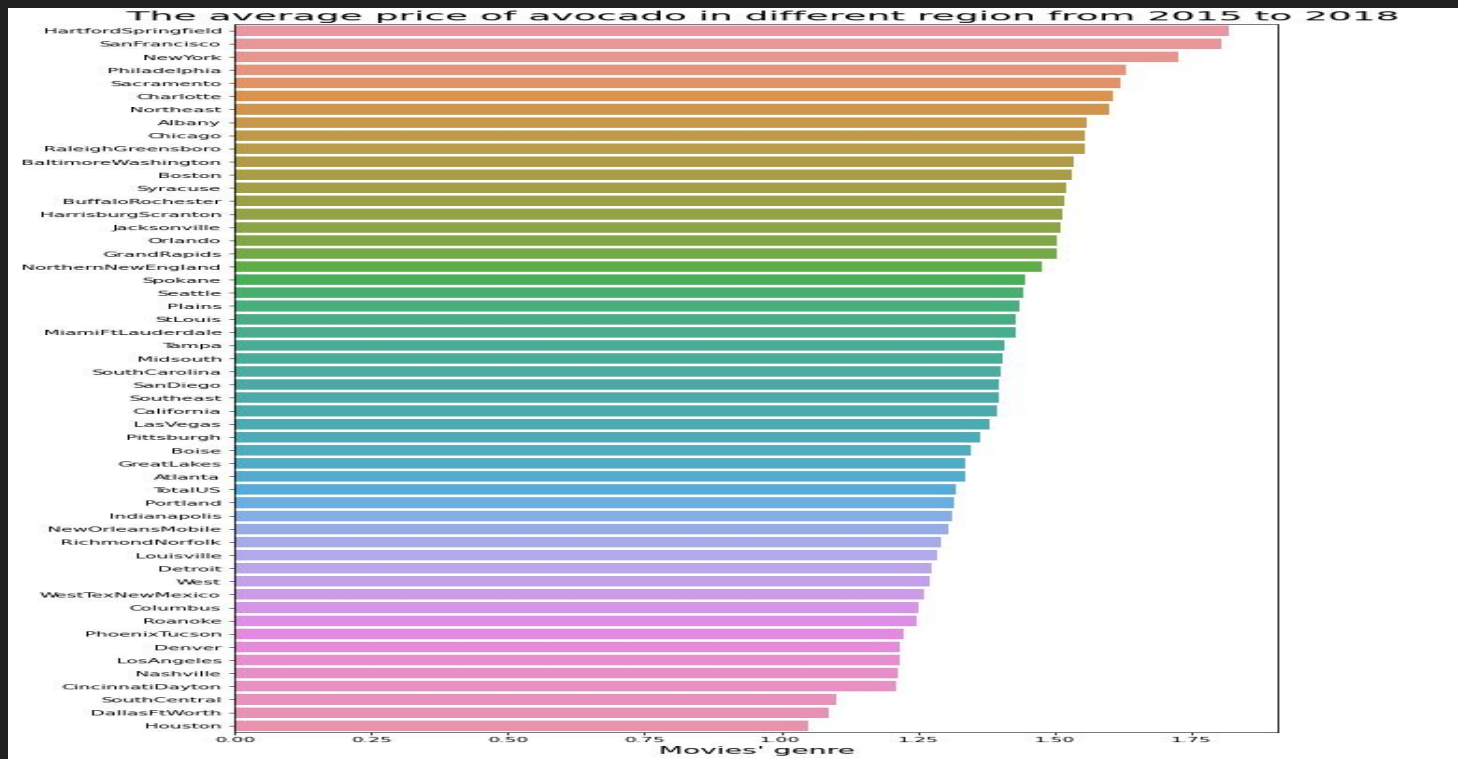
    if (avocado.iloc[i,7] == "organic"):
        avocado.iloc[i,7]=1

avocado.head()
```

	year	Date	AveragePrice	Total Volume	4046	4225	4770	type	region	No_label
0	2015	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	0	Albany	8696.87
1	2015	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	0	Albany	9505.56
2	2015	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	0	Albany	8145.35
3	2015	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	0	Albany	5811.16
4	2015	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	0	Albany	6183.95

Visualization 1:

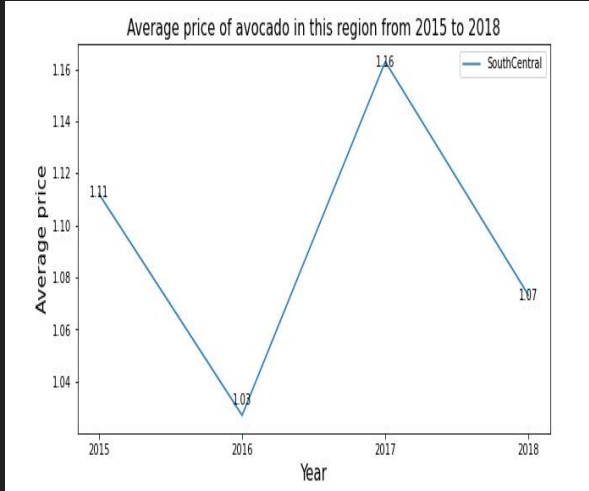
The average avocado price during 2015-2018 in different region.



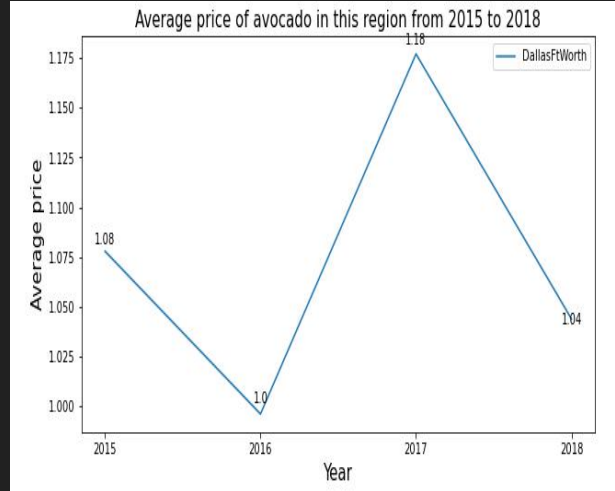
Visualization 2:

The trend of avocado price in different years in the lowest three region

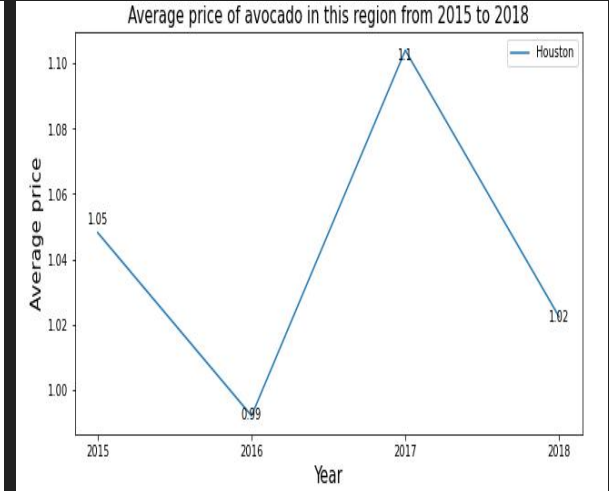
SouthCentral



DallasFtWorth



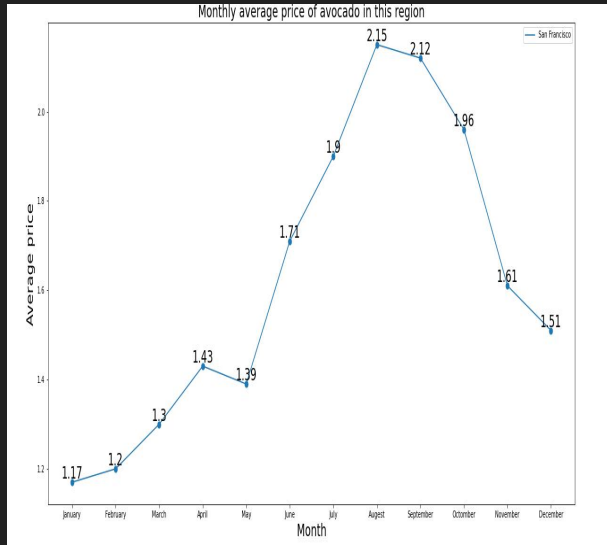
Houston



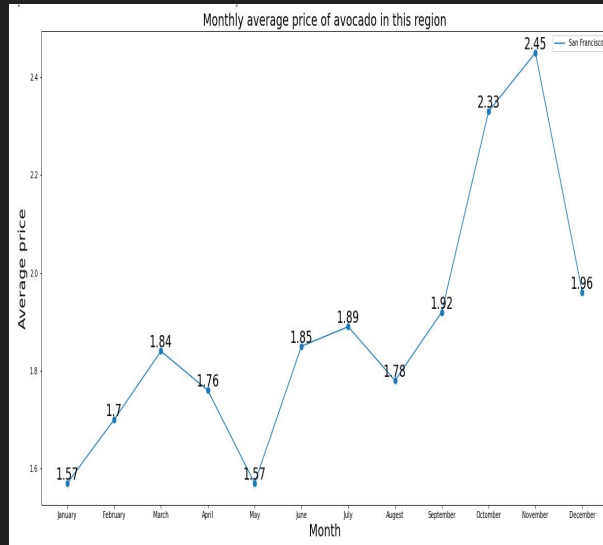
Visualization 3:

More detail in **San Francisco**: The trend of avocado price in different year

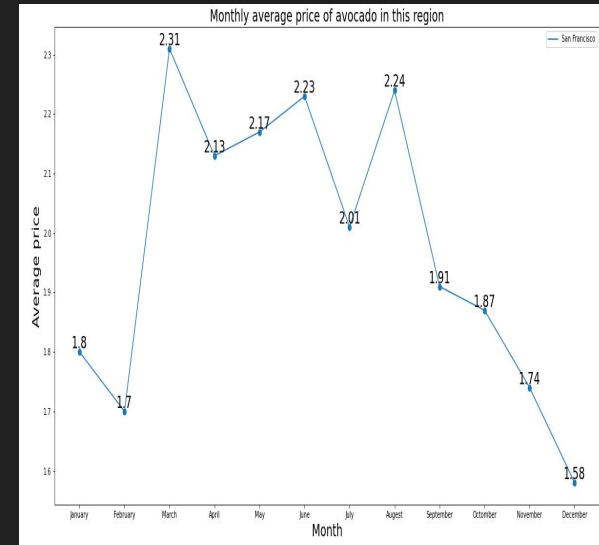
2015



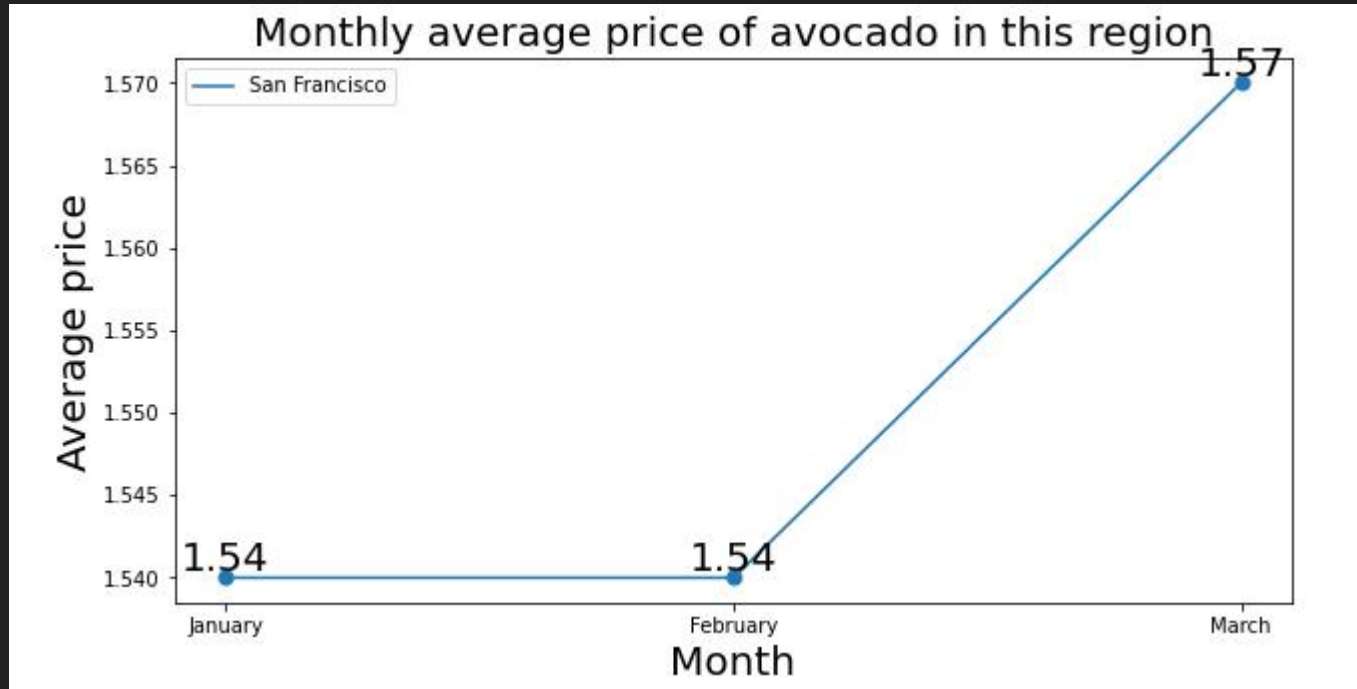
2016



2017

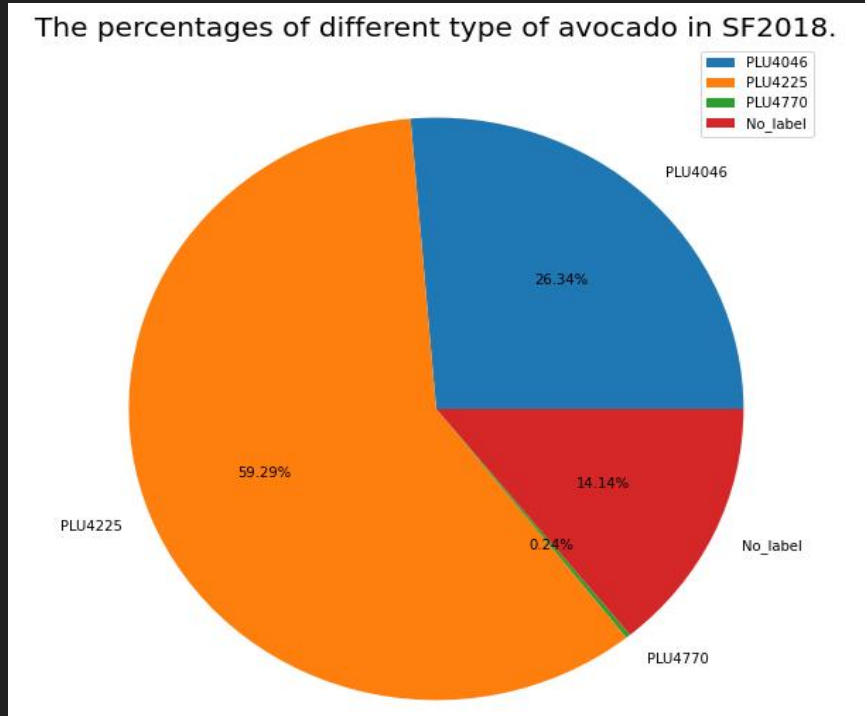


2018



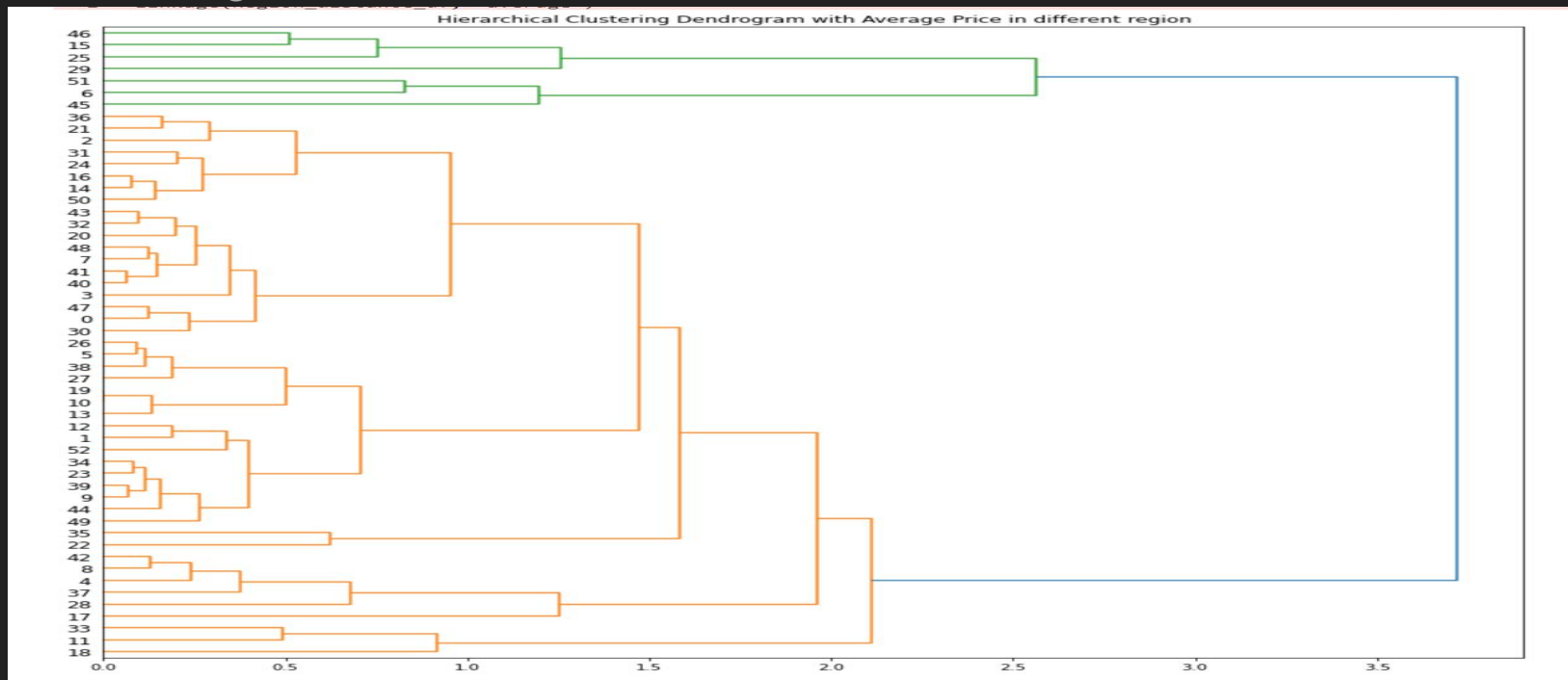
Visualization 4:

The percentages of **different type of avocado** in SF2018.



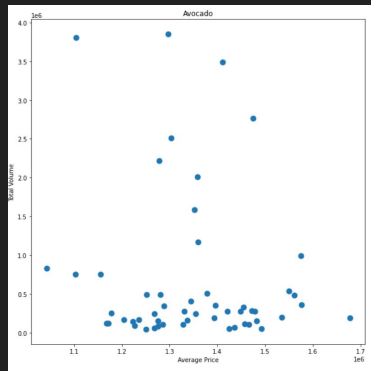
Cluster Analysis

1. Dendrogram

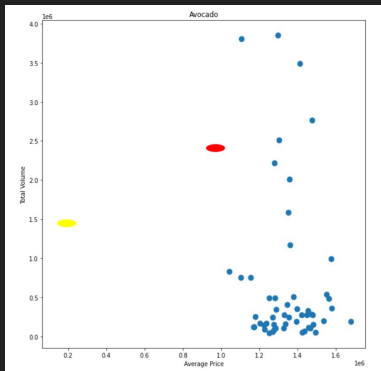


Step-by-step K-means:

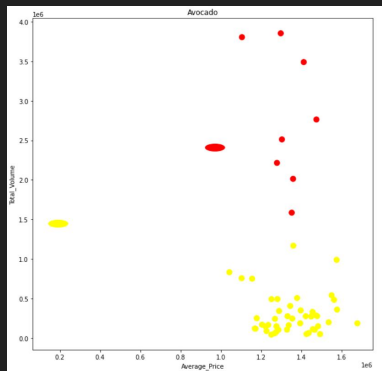
1



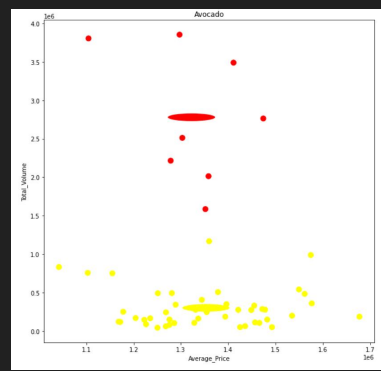
2



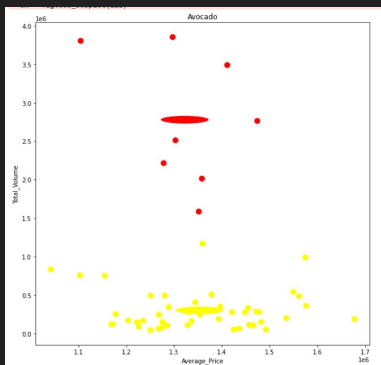
3



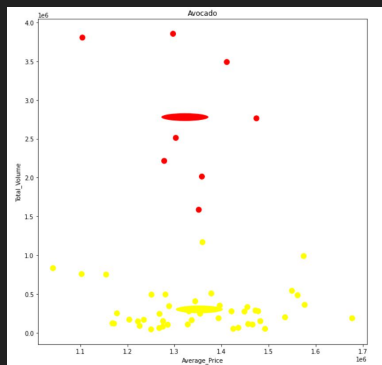
4



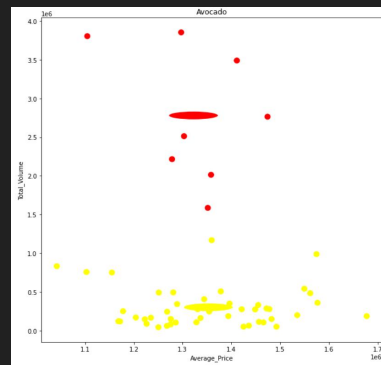
7



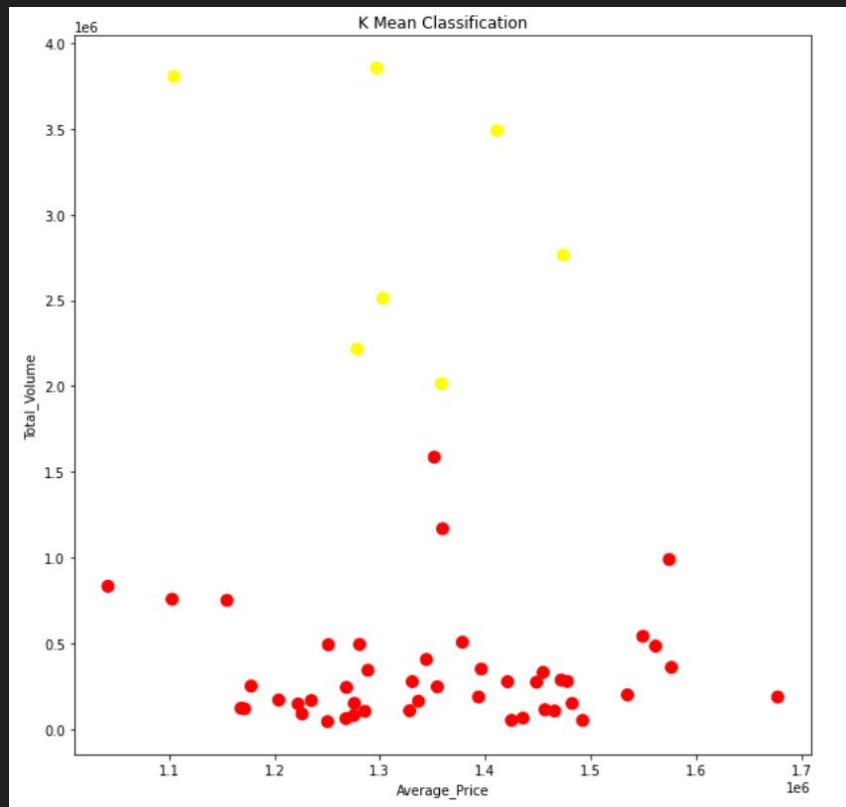
6



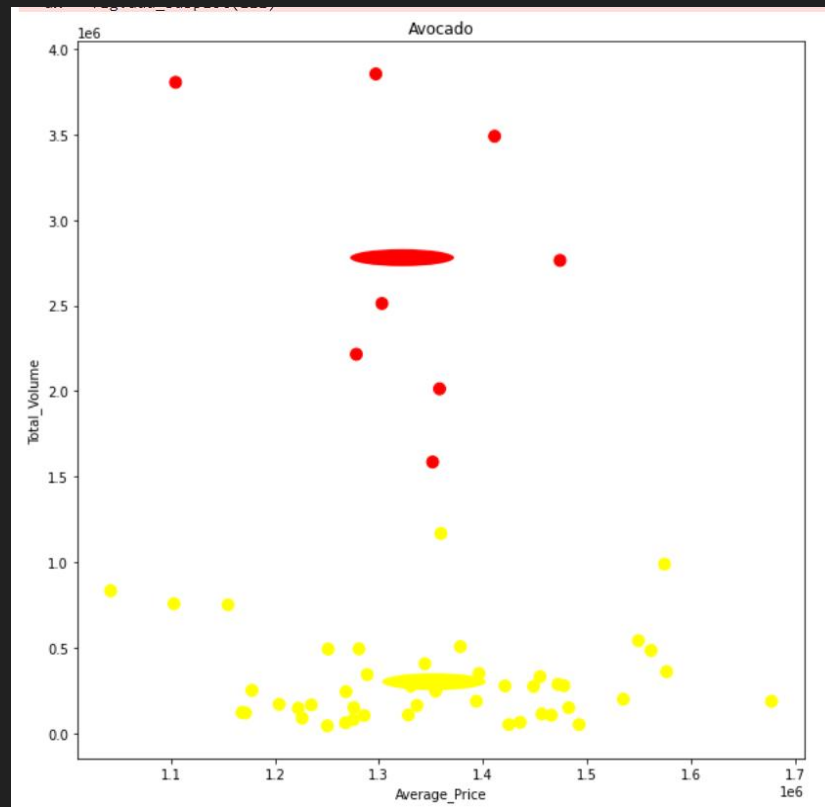
5



Sklearn.Cluster



Step by Step



PredictKNN

make prediction

A data-point called "test_region"

```
In [90]: knn_cluster_data.head()
```

```
Out[90]:
```

	Average_Price	Total_Volume	Association
0	1.435833	0.642494	1
1	1.288750	3.429759	1
2	1.378333	5.066210	1
3	1.492500	0.506150	1
4	1.576667	3.598752	1

```
In [91]: knndf=knn_cluster_data[["Average_Price", "Total_Volume","Association"]]
```

```
In [92]: knndf.head()
```

```
Out[92]:
```

	Average_Price	Total_Volume	Association
0	1.435833	0.642494	1
1	1.288750	3.429759	1
2	1.378333	5.066210	1
3	1.492500	0.506150	1
4	1.576667	3.598752	1

```
In [93]: from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n_neighbors = 2, p = 2)
knn.fit(knndf[['Average_Price', 'Total_Volume']],
        knndf.Association)
```

```
Out[93]: KNeighborsClassifier(n_neighbors=2)
```

Make prediction

```
In [94]: Average_Price = float(input('Average_Price'))
Total_Volume = float(input('Total_Volume'))

data_class = knn.predict(np.array([Average_Price, Total_Volume]).reshape(1, -1))[0]

class_name = ["N/A", "Yellow", "Red"]

print('Prediction: Loan is #', 3- data_class, class_name[data_class])
```

```
Average_Price1.06
Total_Volume25
Prediction: Loan is # 1 Red
```



SUGGESTIONS

1. Overall, In 2019, The price of avocado will **higher** than 2018.
In 2020, The price of avocado will **lower** than 2019.
2. Trends throughout a year, the price will keep **increasing** in **Spring and Summer**, and keep **decreasing** in **Fall and Winter**.
3. For the type of avocados, preparing more PLU4225, and PLU4046.
Additionally, reducing the number of PLU4770.



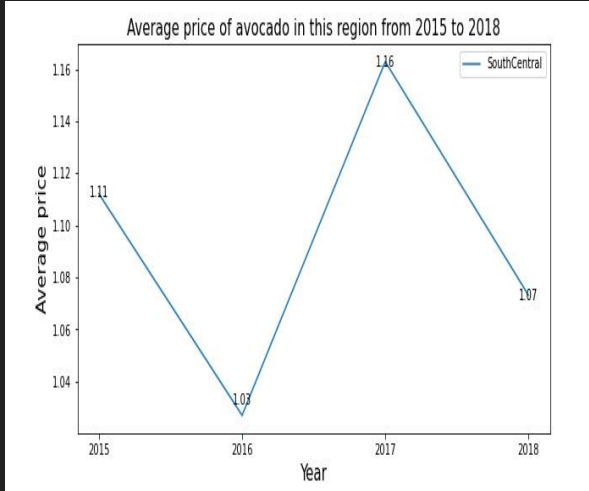
SUGGESTIONS

1. Overall, In 2019, The price of avocado will **higher** than 2018.
In 2020, The price of avocado will **lower** than 2019.
2. Trends throughout a year, the price will keep **increasing** in **Spring and Summer**, and keep **decreasing** in **Fall and Winter**.
3. For the type of avocados, preparing more PLU4225, and PLU4046.
Additionally, reducing the number of PLU4770.

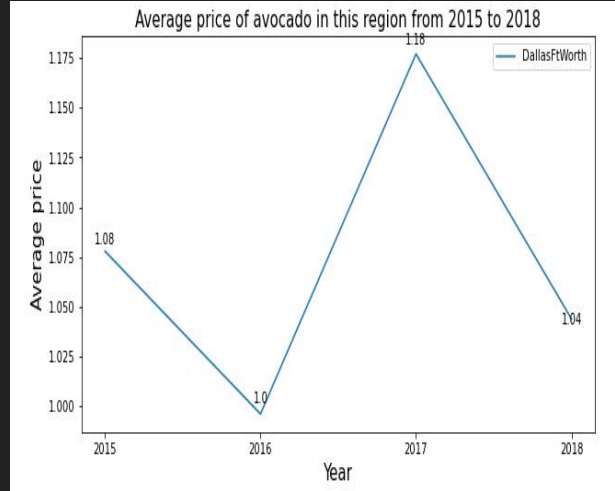
Visualization 2:

The trend of avocado price in different years in the lowest three region

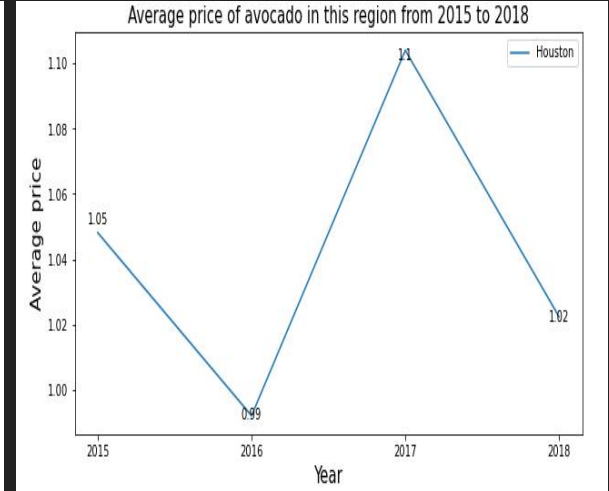
SouthCentral



DallasFtWorth



Houston





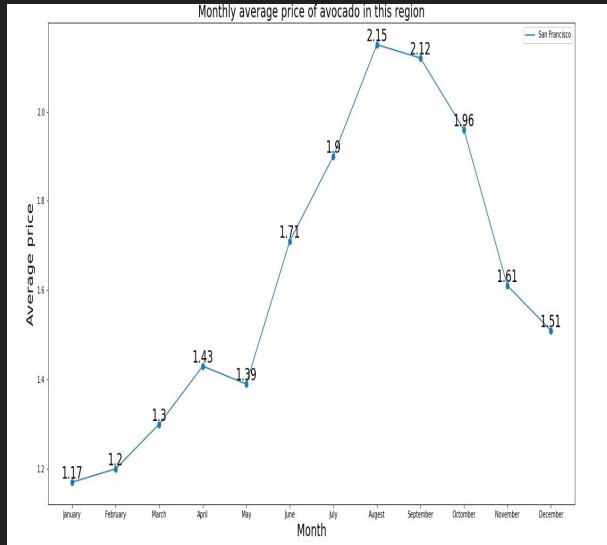
SUGGESTIONS

1. Overall, In 2019, The price of avocado will **higher** than 2018.
In 2020, The price of avocado will **lower** than 2019.
2. Trends throughout a year, the price will keep **increasing in Spring and Summer**, and keep **decreasing in Fall and Winter**.
3. For the type of avocados, preparing more PLU4225, and PLU4046.
Additionally, reducing the number of PLU4770.

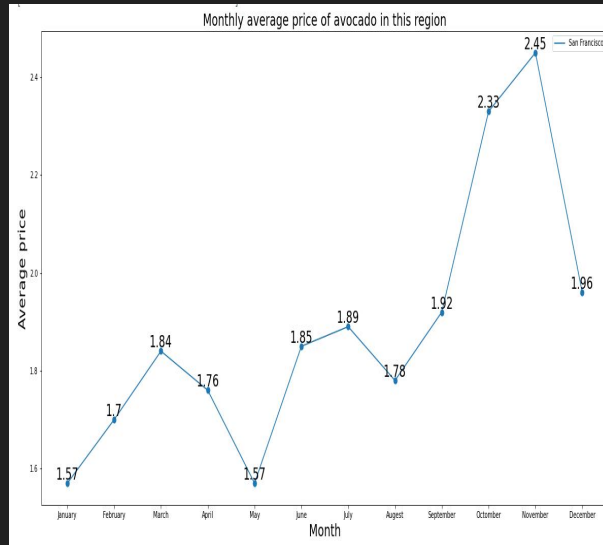
Visualization 3:

More detail in **San Francisco**: The trend of avocado price in different year

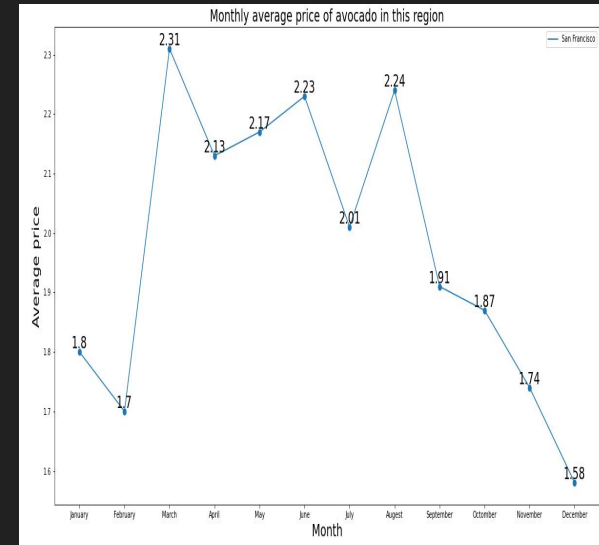
2015



2016



2017



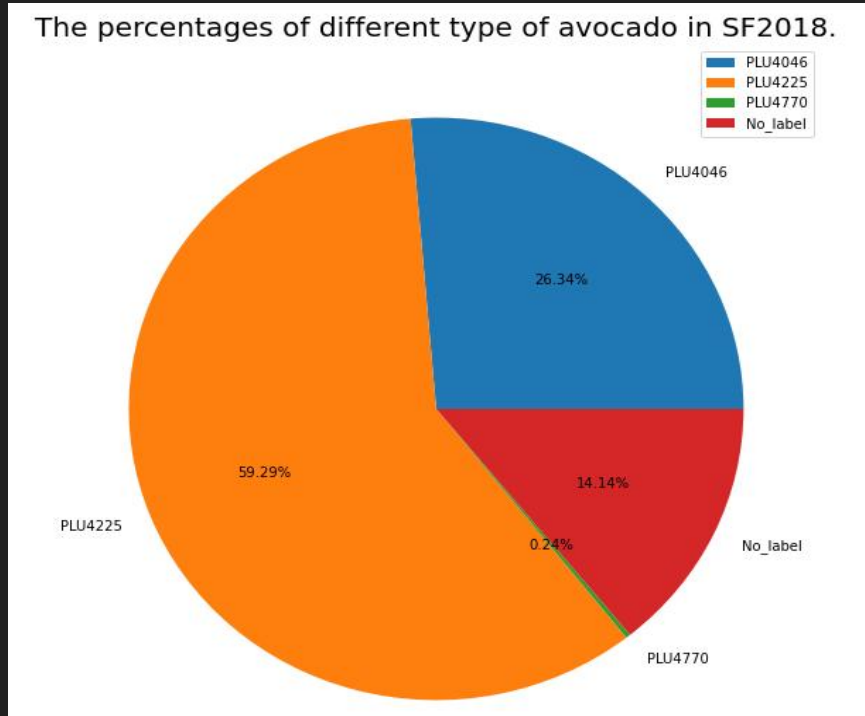


SUGGESTIONS

1. Overall, In 2019, The price of avocado will **higher** than 2018.
In 2020, The price of avocado will **lower** than 2019.
2. Trends throughout a year, the price will keep **increasing** in **Spring and Summer**, and keep **decreasing** in **Fall and Winter**.
3. For the type of avocados, preparing more PLU4225, and PLU4046.
Additionally, reducing the number of PLU4770.

Visualization 4:

The percentages of **different type of avocado** in SF2018.



Thank You!