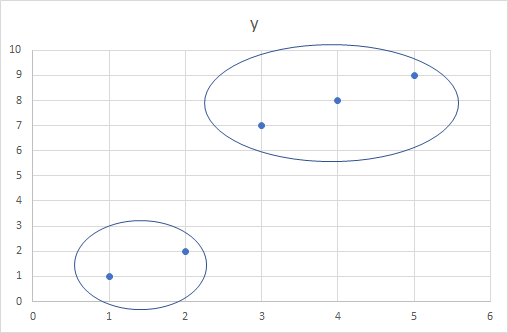
Clustering



Cluster = Group of points that are “close” to each other.

Close? → Standard distance (sqrt((x2-x1)^2 + (y2-y1)^2)

How many groups? →

Decree the number of groups = 2

import numpy as np

import pandas as pd

from matplotlib import pyplot as plt

from sklearn.cluster import KMeans

x = [1, 2, 3, 4, 5]

y = [1, 2, 7, 8, 9]

X = [[1, 1], [2, 2], [3, 7], [4, 8], [5, 9]]

#kmeans is the algorithm we described

kmeans = KMeans(n\_clusters = 2)

groups = kmeans.fit\_predict(X)

plt.scatter(x, y, c=groups)

plt.show()

Problem

Create 3 clusters from this points.

|  |  |
| --- | --- |
| x | y |
| 1 | 8 |
| 2 | 9 |
| 3 | 8.5 |
| 4 | 1 |
| 5 | 2 |
| 6 | 1.5 |
| 7 | 12 |
| 8 | 13.5 |
| 9 | 12.5 |

Get a colored scatter plot with 3 clusters

|  |  |
| --- | --- |
| Group | Cluster Picture |
| 1 |  |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |
| 6 |  |
| 7 |  |

Problem

Take the marketing.csv file and create a cluster that uses all the questions.

List the people in each of the three clusters

Cluster 1 - [1, 3, 4, 5…]

Cluster 2 - [2, 6, 9, …]

Cluster 3 - [7, 8, …]

|  |  |
| --- | --- |
| Group | Cluster Members |
| 1 | cluster 1 - [2, 6, 7, 8, 13, 14, 18]  cluster 2 - [0, 4, 5, 10, 12]  cluster 3 - [1, 3, 9, 11, 15, 16, 17] |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |
| 6 |  |
| 7 |  |

data = pd.read\_csv("I:\LWTech\Winter2021\DataAnal\marketing.csv")

#print(data)

kmeans = KMeans(n\_clusters = 3)

data = data[['V1','V2','V3','V4','V5','V6']]

#print(data)

groups = kmeans.fit\_predict(data)

data['Group'] = groups

#print(data)

#print(data[data['Group'] == 0])

group0 = data[data['Group'] == 0]

group1 = data[data['Group'] == 1]

group2 = data[data['Group'] == 2]

#print(group0)

#print(group1)

#print(group2)

print("Group 0")

print(group0['V1'].mean())

print(group0['V1'].std())

print("Group 1")

print(group1['V1'].mean())

print(group1['V1'].std())

print("Group 2")

print(group2['V1'].mean())

print(group2['V1'].std())

**Break - Back at 7:50**

**\* What kind of customer this is (i.e. frequent shopper, happy shopper, likely online shopper, etc.)**

**L Low - M Medium - H High**

**Group 1**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **M** | **H** | **L** | **H** | **M** | **H** | **Bargainers** |
| **1** | **H** | **M** | **H** | **H** | **L** | **H** | **Eager Shoppers** |
| **2** | **L** | **H** | **M** | **H** | **H** | **H** | **Reluctant Shoppers** |

**Group 2**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **H** | **M** | **H** | **H** | **L** | **H** |  |
| **1** | **L** | **H** | **M** | **H** | **H** | **H** |  |
| **2** | **M** | **H** | **L** | **H** | **M** | **H** |  |

**Group 3**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **H** | **M** | **H** | **H** | **L** | **H** | **Love Shopping** |
| **1** | **L** | **H** | **M** | **M** | **H** | **H** | **Reluctant Shoppers** |
| **2** | **M** | **H** | **L** | **H** | **M** | **H** | **Frugal Shoppers** |

**Group 4**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **L** | **M** | **M** | **L** | **H** | **H** | **Research Shopper** |
| **1** | **M** | **H** | **L** | **M** | **M** | **M** | **Necessity Shopper** |
| **2** | **H** | **L** | **H** | **H** | **L** | **L** | **Impulsive Shopper** |

**Group 5**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **M** | **H** | **L** | **H** | **M** | **H** | **Deal shopper** |
| **1** | **L** | **H** | **M** | **H** | **H** | **H** | **Frugal** |
| **2** | **H** | **M** | **H** | **H** | **M** | **H** | **Happy shopper** |

**Group 6**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **M** | **M** | **H** | **H** | **H** | **H** | **Online Shopper** |
| **1** | **H** | **L** | **M** | **H** | **M** | **H** | **Shopper Holic** |
| **2** | **L** | **H** | **L** | **H** | **L** | **H** | **Cautious** |

**Group 7**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **v1** | **v2** | **v3** | **v4** | **v5** | **v6** | **Name?\*** |
| **0** | **M** | **H** | **L** | **H** | **M** | **H** | **Shopper** |
| **1** | **L** | **H** | **M** | **H** | **H** | **H** | **Cautious Spender** |
| **2** | **H** | **M** | **H** | **H** | **L** | **H** | **Happy shopper** |

**Cluster Analysis**

**Group points together, understand the characteristics of the points, and then you try find commonality between those characteristics**

**Marketing Research - Customer and Customer groups (cohorts)**

**import yfinance as yf**

**from matplotlib import pyplot as plt**

**import statsmodels.api as sm**

**import numpy as np**

**apple = list(yf.download('AAPL', '2015-01-01', '2021-03-01')['Adj Close'])**

**msft = list(yf.download('MSFT', '2015-01-01', '2021-03-01')['Adj Close'])**

**spy = list(yf.download('SPY', '2015-01-01', '2021-03-01')['Adj Close'])**

**tsla = list(yf.download('TSLA', '2015-01-01', '2021-03-01')['Adj Close'])**

**xom = list(yf.download('XOM', '2015-01-01', '2021-03-01')['Adj Close'])**

**spy = sm.add\_constant(spy)**

**msft = sm.add\_constant(msft)**

**tsla = sm.add\_constant(tsla)**

**model = sm.OLS(xom, tsla)**

**result = model.fit()**

**print(result.summary())**

**#Market Sectors**

**#SPY, MSFT, AAPL**

**#TSLA**

**#XOM**

**Problem**

**For these stocks. Group them with one of the previous ones or start a new group.**

**FB, CRM, CVX, AIG**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group** | **FB** | **CRM** | **CVX** | **AIG** |
| **1** | **SPY** | **SPY** | **unique** | **unique** |
| **2** | **SPY** | **SPY** | **CVX** | **AIG** |
| **3** | **SPY** | **SPY** | **underperforming** | **underperforming** |
| **4** | **SPY** | **SPY** | **CVX** | **AIG** |
| **5** | **SPY** | **SPY** |  |  |
| **6** | **SPY** | **SPY** | **unique** |  |
| **7** | **SPY** | **SPY** | **unique** | **unique** |