

Prac5Shre

October 26, 2023

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: df = pd.read_csv('./sales_data_sample.csv', encoding='unicode_escape')
```

```
[3]: to_drop = ['ADDRESSLINE1', 'ADDRESSLINE2', 'STATE', 'POSTALCODE', 'PHONE']
df = df.drop(to_drop, axis=1)
```

```
[4]: df.isnull().sum()
```

```
[4]: ORDERNUMBER          0
QUANTITYORDERED         0
PRICEEACH               0
ORDERLINENUMBER         0
SALES                   0
ORDERDATE               0
STATUS                 0
QTR_ID                 0
MONTH_ID               0
YEAR_ID                0
PRODUCTLINE            0
MSRP                   0
PRODUCTCODE            0
CUSTOMERNAME           0
CITY                   0
COUNTRY                0
TERRITORY              1074
CONTACTLASTNAME         0
CONTACTFIRSTNAME        0
DEALSIZE                0
dtype: int64
```

```
[5]: df['ORDERDATE'] = pd.to_datetime(df['ORDERDATE'])
```

```
[6]: import datetime as dt
snapshot_date = df['ORDERDATE'].max() + dt.timedelta(days = 1)
df_RFM = df.groupby(['CUSTOMERNAME']).agg({
    'ORDERDATE' : lambda x : (snapshot_date - x.max()).days,
```

```

    'ORDERNUMBER' : 'count',
    'SALES' : 'sum'
})

#Rename the columns
df_RFM.rename(columns = {
    'ORDERDATE' : 'Recency',
    'ORDERNUMBER' : 'Frequency',
    'SALES' : 'MonetaryValue'
}, inplace=True)

```

```
[7]: df_RFM.head()
```

```
[7]:
```

	Recency	Frequency	MonetaryValue
CUSTOMERNAME			
AV Stores, Co.	196	51	157807.81
Alpha Cognac	65	20	70488.44
Amica Models & Co.	265	26	94117.26
Anna's Decorations, Ltd	84	46	153996.13
Atelier graphique	188	7	24179.96

```
[8]: df_RFM['M'] = pd.qcut(df_RFM['MonetaryValue'], q = 4, labels = range(1,5))
df_RFM['R'] = pd.qcut(df_RFM['Recency'], q = 4, labels = list(range(4,0,-1)))
df_RFM['F'] = pd.qcut(df_RFM['Frequency'], q = 4, labels = range(1,5))

df_RFM.head()
```

```
[8]:
```

	Recency	Frequency	MonetaryValue	M	R	F
CUSTOMERNAME						
AV Stores, Co.	196	51	157807.81	4	2	4
Alpha Cognac	65	20	70488.44	2	4	2
Amica Models & Co.	265	26	94117.26	3	1	2
Anna's Decorations, Ltd	84	46	153996.13	4	3	4
Atelier graphique	188	7	24179.96	1	2	1

```
[9]: df_RFM['RFM_Score'] = df_RFM[['R', 'M', 'F']].sum(axis=1)
df_RFM.head()
```

```
[9]:
```

	Recency	Frequency	MonetaryValue	M	R	F	RFM_Score
CUSTOMERNAME							
AV Stores, Co.	196	51	157807.81	4	2	4	10
Alpha Cognac	65	20	70488.44	2	4	2	8
Amica Models & Co.	265	26	94117.26	3	1	2	6
Anna's Decorations, Ltd	84	46	153996.13	4	3	4	11
Atelier graphique	188	7	24179.96	1	2	1	4

```
[10]: def rfm_level(df):
        if bool(df['RFM_Score'] >= 10):
            return 'High Value Customer'

        elif bool(df['RFM_Score'] < 10) and bool(df['RFM_Score'] >= 6):
            return 'Mid Value Customer'
        else:
            return 'Low Value Customer'
df_RFM['RFM_Level'] = df_RFM.apply(rfm_level, axis = 1)
df_RFM.head()
```

```
[10]:
```

	Recency	Frequency	MonetaryValue	M	R	F	\
CUSTOMERNAME							
AV Stores, Co.	196	51	157807.81	4	2	4	
Alpha Cognac	65	20	70488.44	2	4	2	
Amica Models & Co.	265	26	94117.26	3	1	2	
Anna's Decorations, Ltd	84	46	153996.13	4	3	4	
Atelier graphique	188	7	24179.96	1	2	1	

	RFM_Score	RFM_Level
CUSTOMERNAME		
AV Stores, Co.	10	High Value Customer
Alpha Cognac	8	Mid Value Customer
Amica Models & Co.	6	Mid Value Customer
Anna's Decorations, Ltd	11	High Value Customer
Atelier graphique	4	Low Value Customer

```
[11]: data = df_RFM[['Recency', 'Frequency', 'MonetaryValue']]
data.head()
```

```
[11]:
```

	Recency	Frequency	MonetaryValue
CUSTOMERNAME			
AV Stores, Co.	196	51	157807.81
Alpha Cognac	65	20	70488.44
Amica Models & Co.	265	26	94117.26
Anna's Decorations, Ltd	84	46	153996.13
Atelier graphique	188	7	24179.96

```
[12]: data_log = np.log(data)
data_log.head()
```

```
[12]:
```

	Recency	Frequency	MonetaryValue
CUSTOMERNAME			
AV Stores, Co.	5.278115	3.931826	11.969133
Alpha Cognac	4.174387	2.995732	11.163204
Amica Models & Co.	5.579730	3.258097	11.452297
Anna's Decorations, Ltd	4.430817	3.828641	11.944683

Atelier graphique 5.236442 1.945910 10.093279

```
[13]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(data_log)
data_normalized = scaler.transform(data_log)
data_normalized = pd.DataFrame(data_normalized, index = data_log.index,
    ↪ columns=data_log.columns)
data_normalized.describe().round(2)
```

```
[13]:
```

	Recency	Frequency	MonetaryValue
count	92.00	92.00	92.00
mean	0.00	-0.00	0.00
std	1.01	1.01	1.01
min	-3.51	-3.67	-3.82
25%	-0.24	-0.41	-0.39
50%	0.37	0.06	-0.04
75%	0.53	0.45	0.52
max	1.12	4.03	3.92

```
[14]: import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans

sse = {}

for k in range(1, 21):
    kmeans = KMeans(n_clusters = k, random_state = 1)
    kmeans.fit(data_normalized)
    sse[k] = kmeans.inertia_
```

```
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
```



```

/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)

```

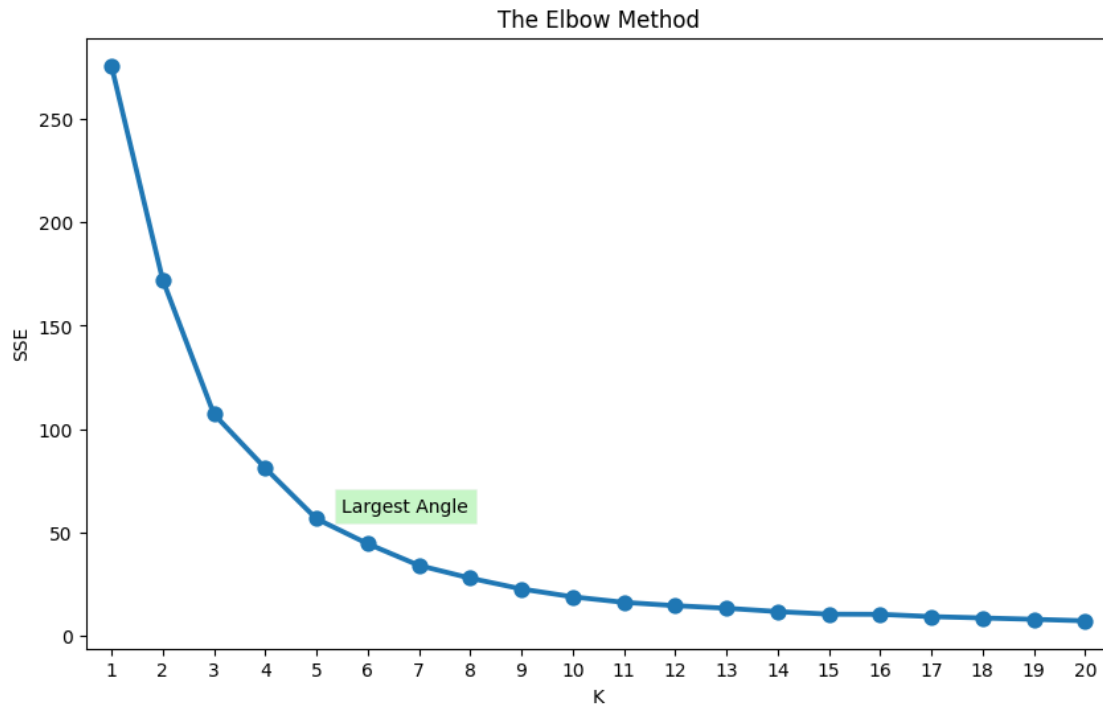
```

[15]: plt.figure(figsize=(10,6))
      plt.title('The Elbow Method')

      plt.xlabel('K')
      plt.ylabel('SSE')
      plt.style.use('ggplot')

      sns.pointplot(x=list(sse.keys()), y = list(sse.values()))
      plt.text(4.5, 60, "Largest Angle", bbox = dict(facecolor = 'lightgreen', alpha=
↳ 0.5))
      plt.show()

```



```
[16]: kmeans = KMeans(n_clusters=5, random_state=1)
kmeans.fit(data_normalized)
cluster_labels = kmeans.labels_

data_rfm = data.assign(Cluster = cluster_labels)
data_rfm.head()
```

```
/home/anand007/.local/pipx/venvs/jupyterlab/lib/python3.11/site-
packages/sklearn/cluster/_kmeans.py:1416: FutureWarning: The default value of
`n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```

```
[16]:
```

CUSTOMERNAME	Recency	Frequency	MonetaryValue	Cluster
AV Stores, Co.	196	51	157807.81	1
Alpha Cognac	65	20	70488.44	2
Amica Models & Co.	265	26	94117.26	2
Anna's Decorations, Ltd	84	46	153996.13	1
Atelier graphique	188	7	24179.96	0

```
[ ]:
```