

# 21bit0466-assignment5

October 5, 2023

## 1 21BIT0466 Aarjav Jain

```
[1]: #import necessary libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

## 2 Understanding the data

```
[2]: #Load the Dataset
df=pd.read_csv("/content/Mall_Customers.csv")
df.head()
```

```
[2]:
```

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   CustomerID                            200 non-null    int64
1   Gender                                200 non-null    object
2   Age                                    200 non-null    int64
3   Annual Income (k$)                    200 non-null    int64
4   Spending Score (1-100)                 200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```
[ ]: df.shape
```

```
[ ]: (200, 5)
```

```
[ ]: df.describe()
```

```
[ ]:      CustomerID      Age  Annual Income (k$)  Spending Score (1-100)
count  200.000000  200.000000      200.000000      200.000000
mean    100.500000   38.850000      60.560000      50.200000
std     57.879185   13.969007      26.264721      25.823522
min       1.000000   18.000000      15.000000       1.000000
25%     50.750000   28.750000      41.500000      34.750000
50%    100.500000   36.000000      61.500000      50.000000
75%    150.250000   49.000000      78.000000      73.000000
max    200.000000   70.000000     137.000000      99.000000
```

```
[ ]: df.isnull().sum() #No Null values present
```

```
[ ]: Gender      0
Age      0
Annual Income (k$)  0
Spending Score (1-100)  0
dtype: int64
```

```
[ ]: df.Gender.value_counts()
```

```
[ ]: 0    112
     1     88
     Name: Gender, dtype: int64
```

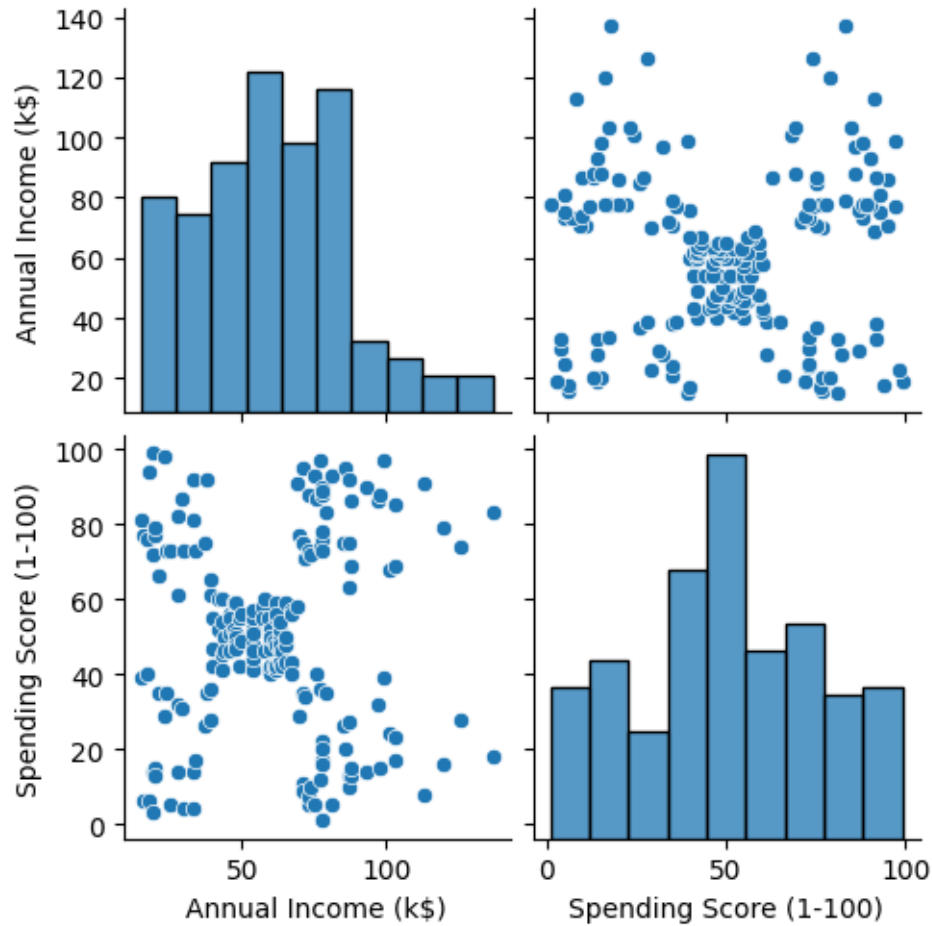
### 3 Data Preprocessing

```
[4]: #Removing unnecessary columns i.e. Customer ID, gender and age
new_df = df.iloc[:, -2:]
new_df.head()
```

```
[4]:      Annual Income (k$)  Spending Score (1-100)
0              15              39
1              15              81
2              16               6
3              16              77
4              17              40
```

```
[5]: sns.pairplot(new_df)
```

```
[5]: <seaborn.axisgrid.PairGrid at 0x7c07f2bef280>
```



## 4 K-means Clustering Algorithm

```
[ ]: from sklearn import cluster
```

```
[ ]: error=[]
      for i in range(1,11):
          kmeans = cluster.KMeans(n_clusters=i,init = 'k-means++',random_state=0)
          kmeans.fit(new_df)
          error.append(kmeans.inertia_)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
```

```

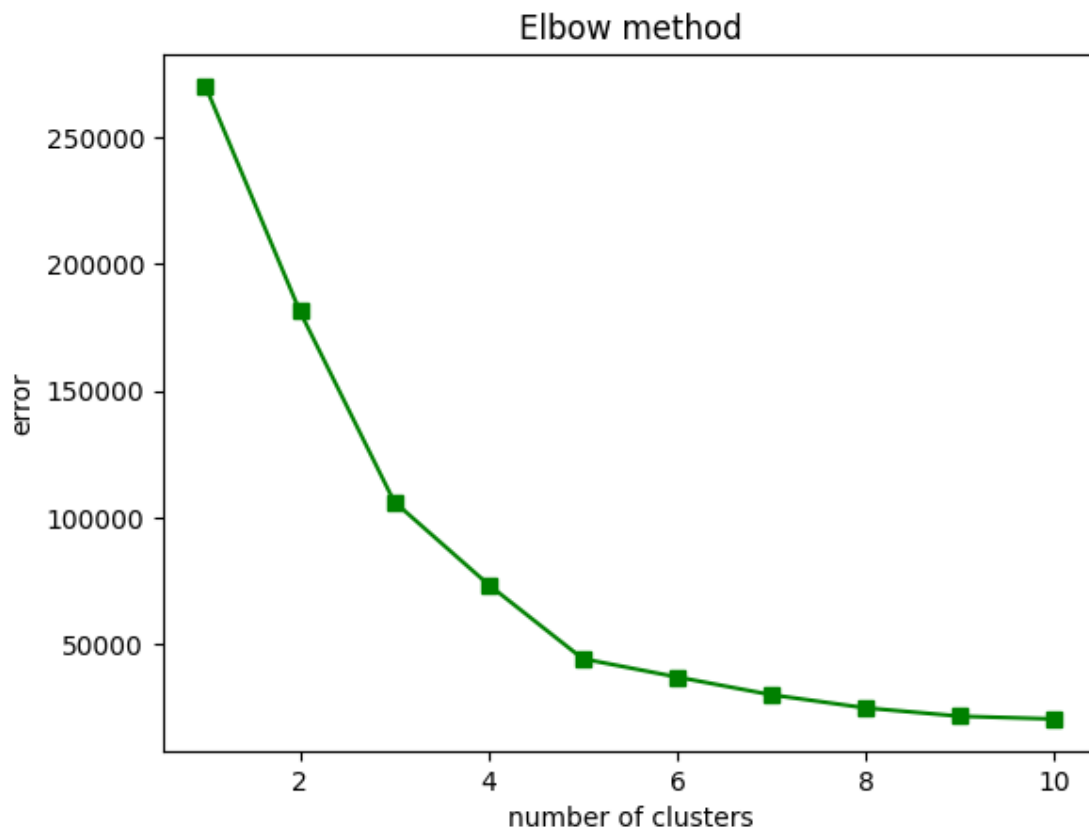
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(

```

```
[ ]: error
```

```
[ ]: [269981.28,
181363.59595959593,
106348.37306211122,
73679.78903948836,
44448.4554479337,
37265.86520484346,
30259.65720728547,
25095.70320999756,
21830.041978049434,
20736.679938924128]
```

```
[ ]: plt.plot(range(1,11),error, 'gs-')
plt.title('Elbow method')
plt.xlabel('number of clusters')
plt.ylabel('error')
plt.show()
```



```
[ ]: km_model = cluster.KMeans(n_clusters=5,init = 'k-means++',random_state=0)
```

```
[ ]: km_model.fit(new_df)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
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
warnings.warn(
```

```
[ ]: KMeans(n_clusters=5, random_state=0)
```

```
[ ]: pred = km_model.predict(new_df)
pred
```

```
[ ]: # Test the model with random observation
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but KMeans was fitted with feature names
  warnings.warn(
```

```
[ ]: km_model.predict([[20,40]])
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but KMeans was fitted with feature names
warnings.warn(
```

```
[ ]: km_model.predict([[60,50]])
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but KMeans was fitted with feature names
    warnings.warn(
```