

import modules and dataset

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df=pd.read_csv("C:\\Users\\hp\\Downloads\\prodigy infotech\\task 2\\Mall_Customers.csv")
df.head()
```

```
↗
```

| | 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['Gender']=pd.get_dummies(df['Gender'],drop_first=True)
df.head() # male = 1 , female = 0
```

```
↗
```

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

```
df['Gender']=df['Gender'].astype(int)
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   int32
2   Age                    200 non-null   int64
3   Annual Income (k$)     200 non-null   int64
4   Spending Score (1-100) 200 non-null   int64
dtypes: int32(1), int64(4)
memory usage: 7.2 KB
```

```
df.drop('CustomerID',axis=1,inplace=True)
```

```
df.head()
```

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

```
df.isnull().sum()
```

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

```
#finding value of k
```

```
from sklearn.cluster import KMeans
```

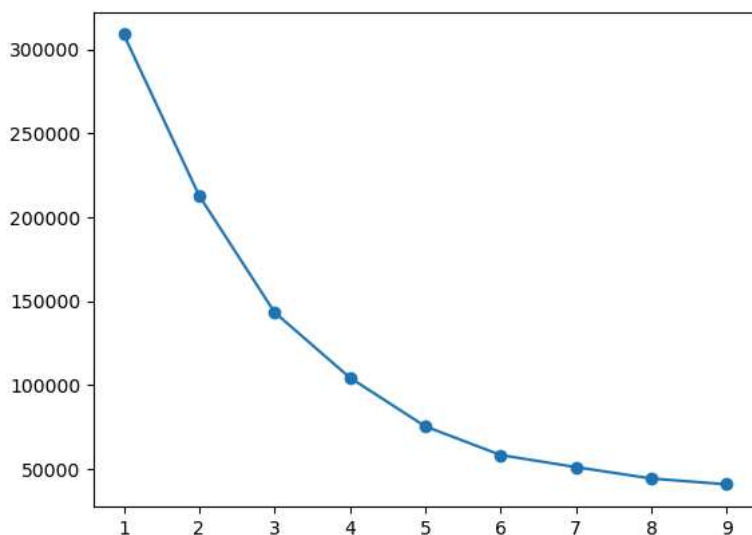
```
sse=[]
k_rng=range(1,10)
```

```
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[['Gender', 'Age', 'Annual Income (k$)', 'Spending Score (1-100)']])
    sse.append(km.inertia_)
```

```
C:\Users\hp\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:1036: UserWarning: KMeans is known to have a memory leak on Windows w
warnings.warn(
```

```
plt.scatter(k_rng,sse)
plt.plot(k_rng,sse)
```

```
[<matplotlib.lines.Line2D at 0x215ebb746d0>]
```



```
plt.plot(df)
```

[illegible]

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

```
df.head()
```

```

Gender  Age  Annual Income (k$)  Spending Score (1-100)  cluster
0      1   19                  15                  39          0

```

```

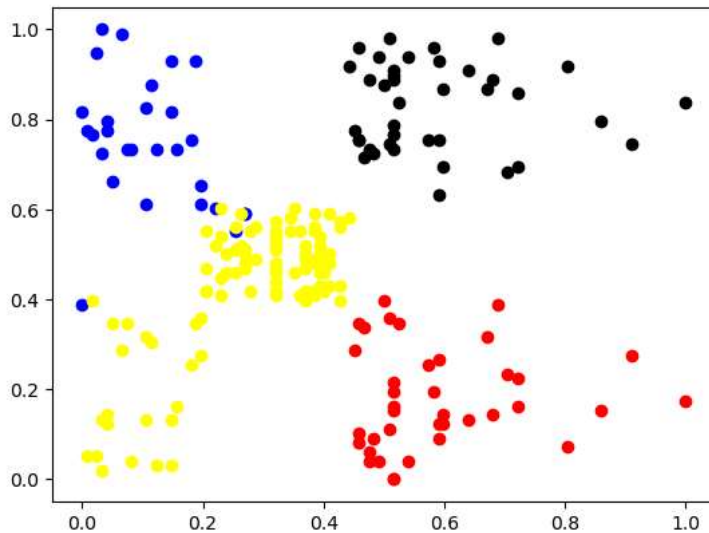
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
df4=df[df.cluster==3]

4      0   31                  17                  40          3

df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
df4=df[df.cluster==3]
plt.scatter(df1['Annual Income (k$)'],df1['Spending Score (1-100)'],color='blue')
plt.scatter(df2['Annual Income (k$)'],df2['Spending Score (1-100)'],color='red')
plt.scatter(df3['Annual Income (k$)'],df3['Spending Score (1-100)'],color='black')
plt.scatter(df4['Annual Income (k$)'],df4['Spending Score (1-100)'],color='yellow')

```

```
<matplotlib.collections.PathCollection at 0x215ed6aaf10>
```



#this algorithm used to to group customers of a retail store based on their purchase history.

Start coding or [generate](#) with AI.