```
In [1]: import pandas as pd
        from matplotlib import pyplot as plt
        %matplotlib inline
```

```
In [4]: df=pd.read_csv(r"C:\Users\DELL\Downloads\Income.csv")
        df
```

Out[4]:

	Gender	Age	Income($)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17
...	...	...	...
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

200 rows × 3 columns

```
In [5]: df.head()
```

Out[5]:

	Gender	Age	Income($)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17

```
In [6]: df.tail()
```

Out[6]:

	Gender	Age	Income($)
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

```
In [7]: plt.scatter(df["Age"],df["Income($)"])
        plt.xlabel("Age")
        plt.ylabel("Income($)")
```

Out[7]: Text(0, 0.5, 'Income($)')
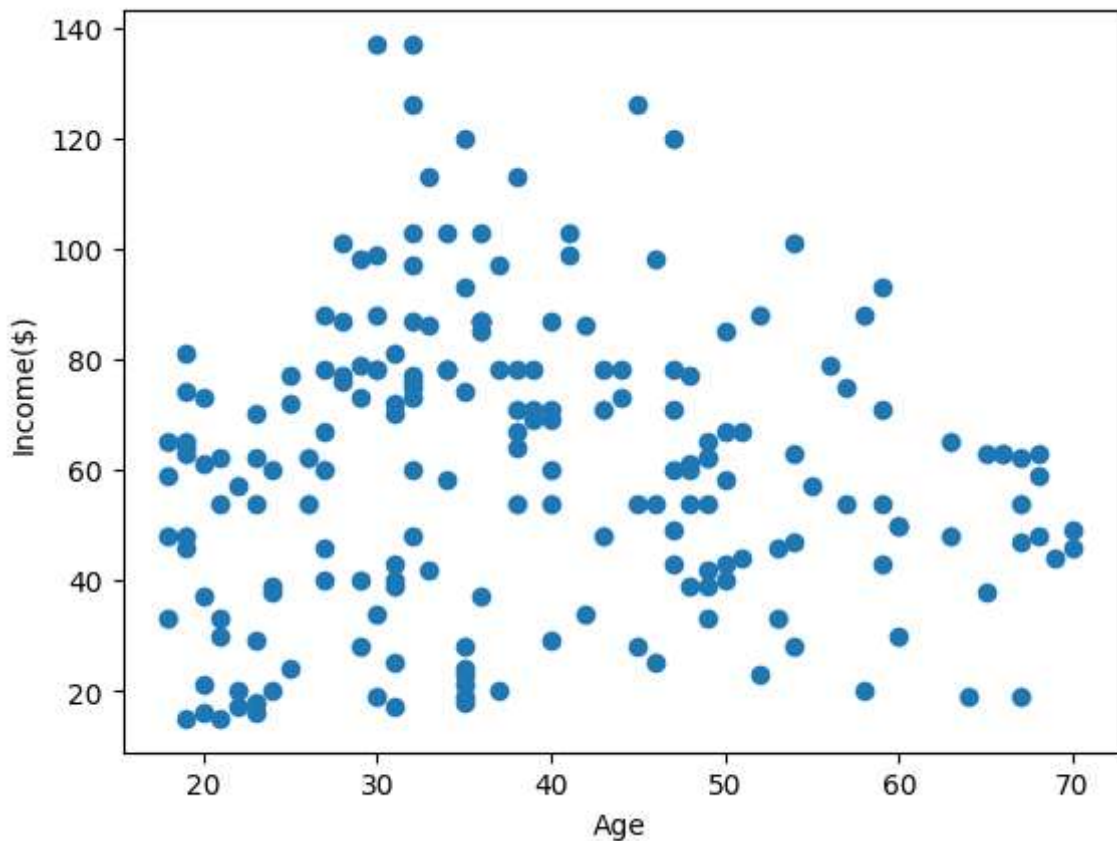

```
In [8]: from sklearn.cluster import KMeans
        km=KMeans()
        km
```

Out[8]: ▾ KMeans

         KMeans()

```
In [9]: y_predicted=km.fit_predict(df[["Age","Income($)"]])
        y_predicted
```

C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(

```
Out[9]: array([6, 6, 6, 6, 6, 6, 6, 6, 0, 6, 0, 6, 0, 6, 6, 6, 6, 6, 0, 6, 6, 6,
               0, 6, 0, 6, 0, 6, 0, 6, 0, 2, 0, 2, 0, 2, 2, 2, 0, 2, 0, 2,
               0, 2, 0, 2, 2, 2, 0, 2, 2, 0, 0, 0, 0, 3, 2, 0, 3, 2, 3, 0, 3, 2,
               0, 3, 2, 2, 3, 0, 3, 3, 3, 2, 5, 5, 2, 5, 3, 5, 3, 5, 2, 5, 3, 7,
               5, 5, 3, 7, 5, 5, 7, 7, 5, 7, 5, 7, 7, 5, 3, 7, 5, 7, 3, 5, 3, 3,
               3, 7, 5, 7, 7, 7, 3, 5, 5, 5, 7, 5, 5, 5, 7, 7, 5, 5, 5, 5, 5, 5,
               7, 7, 7, 7, 5, 7, 7, 7, 5, 7, 7, 7, 7, 7, 5, 7, 7, 7, 5, 5, 5, 7,
               5, 7, 7, 7, 7, 7, 5, 7, 7, 7, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
               1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 4, 4, 4, 4, 4,
               4, 4])
```

```
In [10]: df["cluster"]=y_predicted
         df.head()
```

Out[10]:

	Gender	Age	Income($)	cluster
0	Male	19	15	6
1	Male	21	15	6
2	Female	20	16	6
3	Female	23	16	6
4	Female	31	17	6

```
In [11]: df1=df[df.cluster==0]
         df2=df[df.cluster==1]
         df3=df[df.cluster==2]
         plt.scatter(df1["Age"],df1["Income($)"],color="red")
         plt.scatter(df2["Age"],df2["Income($)"],color="green")
         plt.scatter(df3["Age"],df3["Income($)"],color="blue")
         plt.xlabel("Age")
         plt.ylabel("Income($)")
```

Out[11]: Text(0, 0.5, 'Income($)')
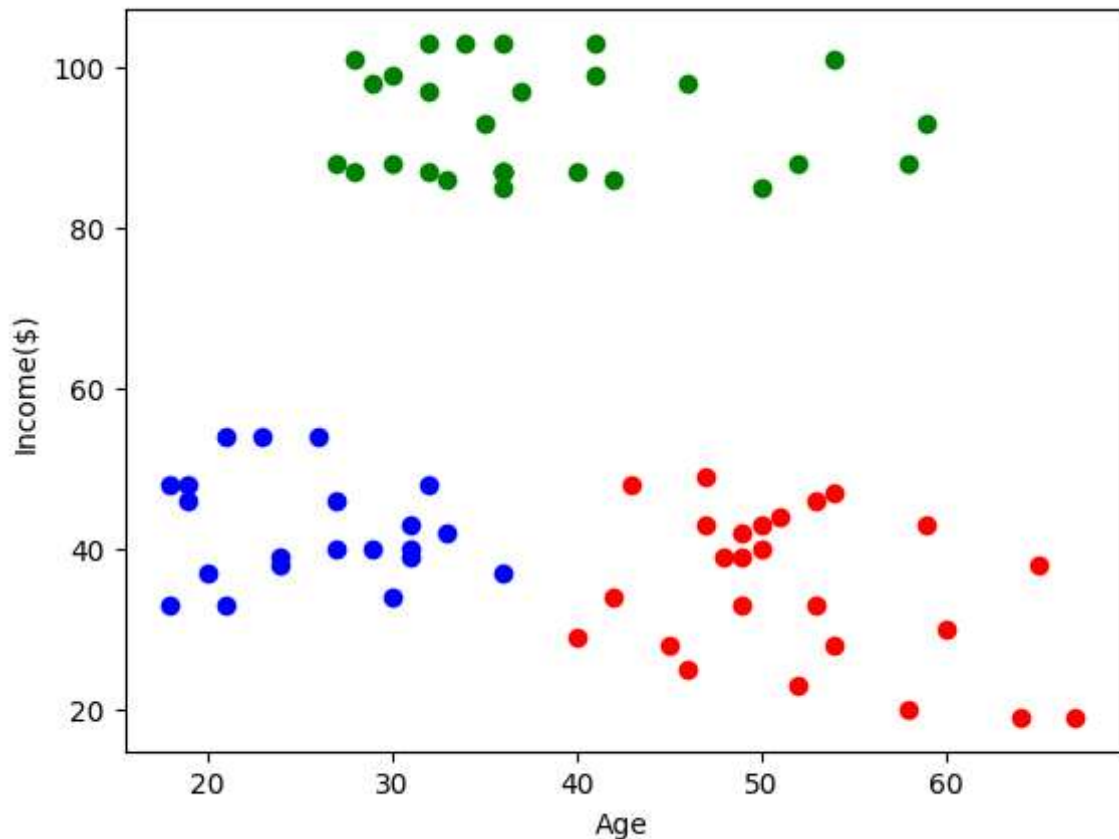

```
In [12]: from sklearn.preprocessing import MinMaxScaler
         scaler=MinMaxScaler()
         scaler.fit(df[["Income($)"]])
         df["Income($)"]=scaler.transform(df[["Income($)"]])
         df.head()
```

Out[12]:

	Gender	Age	Income($)	cluster
0	Male	19	0.000000	6
1	Male	21	0.000000	6
2	Female	20	0.008197	6
3	Female	23	0.008197	6
4	Female	31	0.016393	6

```python
In [13]: scaler.fit(df[["Age"]])
         df["Age"]=scaler.transform(df[["Age"]])
         df.head()
```

Out[13]:

	Gender	Age	Income($)	cluster
0	Male	0.019231	0.000000	6
1	Male	0.057692	0.000000	6
2	Female	0.038462	0.008197	6
3	Female	0.096154	0.008197	6
4	Female	0.250000	0.016393	6

```python
In [14]: km=KMeans()
```

```python
In [15]: y_predicted=km.fit_predict(df[["Age","Income($)"]])
         y_predicted
```

```
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
    warnings.warn(
```

```
Out[15]: array([1, 1, 1, 1, 6, 1, 6, 1, 5, 6, 5, 6, 3, 1, 6, 1, 6, 1, 3, 6, 6, 1,
                3, 6, 3, 6, 3, 6, 6, 1, 5, 1, 3, 1, 3, 1, 3, 6, 6, 1, 5, 1, 3, 6,
                3, 1, 3, 6, 6, 6, 3, 6, 6, 5, 3, 3, 3, 5, 7, 3, 5, 7, 5, 3, 5, 7,
                3, 5, 7, 6, 5, 3, 5, 5, 5, 7, 0, 4, 7, 0, 5, 4, 5, 0, 7, 0, 0, 7,
                4, 0, 5, 7, 0, 4, 4, 7, 0, 7, 0, 7, 7, 0, 5, 7, 0, 7, 5, 0, 5, 5,
                5, 7, 4, 7, 7, 7, 5, 0, 0, 0, 7, 4, 4, 4, 7, 4, 0, 4, 0, 4, 0, 4,
                7, 4, 7, 4, 0, 4, 7, 4, 0, 4, 4, 4, 7, 4, 0, 4, 4, 4, 0, 4, 0, 4,
                0, 4, 4, 4, 4, 4, 0, 4, 7, 4, 0, 4, 4, 4, 4, 4, 4, 4, 4, 4, 0, 4,
                0, 4, 0, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
                2, 2])
```

```python
In [16]: df["New Cluster"]=y_predicted
         df.head()
```

Out[16]:

	Gender	Age	Income($)	cluster	New Cluster
0	Male	0.019231	0.000000	6	1
1	Male	0.057692	0.000000	6	1
2	Female	0.038462	0.008197	6	1
3	Female	0.096154	0.008197	6	1
4	Female	0.250000	0.016393	6	6

```
In [17]: df1=df[df["New Cluster"]==0]
         df2=df[df["New Cluster"]==1]
         df3=df[df["New Cluster"]==2]
         plt.scatter(df1["Age"],df1["Income($)"],color="red")
         plt.scatter(df2["Age"],df2["Income($)"],color="green")
         plt.scatter(df3["Age"],df3["Income($)"],color="blue")
         plt.xlabel("Age")
         plt.ylabel("Income($)")
```

Out[17]: Text(0, 0.5, 'Income($)')
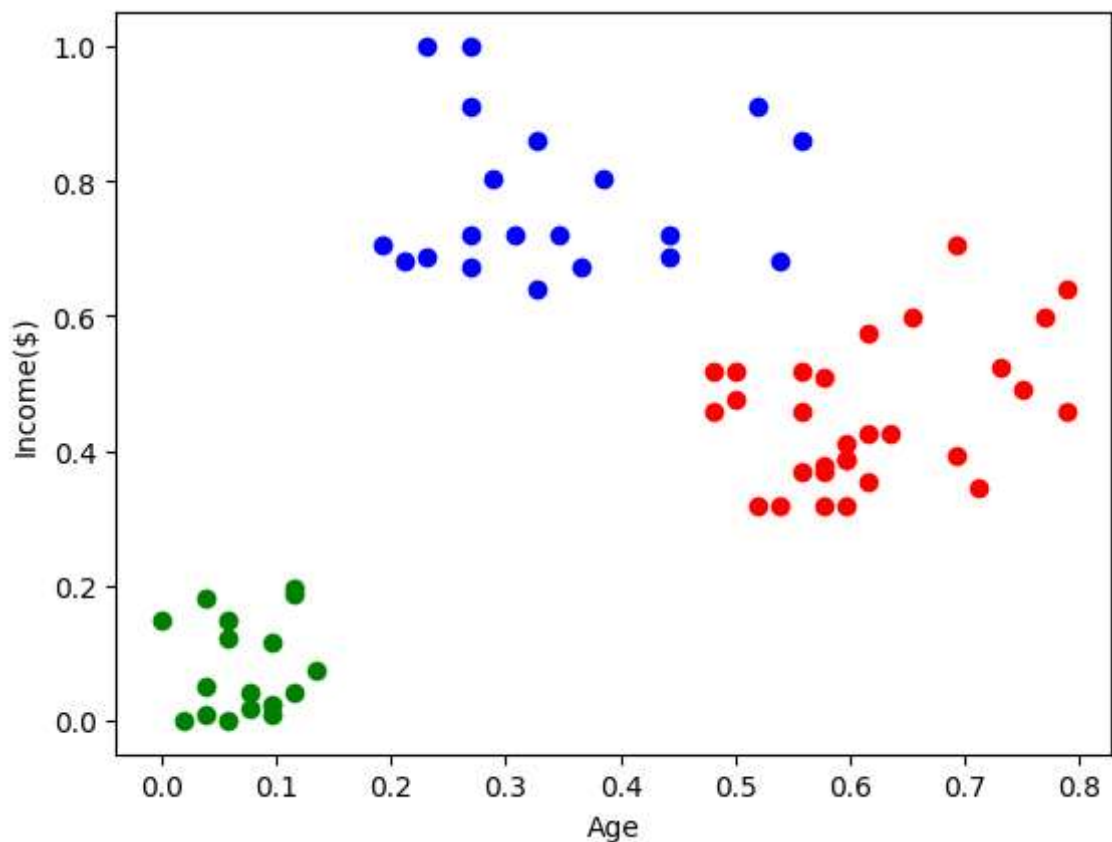

```
In [18]: km.cluster_centers_
```

Out[18]: array([[0.61474359, 0.45191257],
                [0.07239819, 0.08003857],
                [0.33942308, 0.77295082],
                [0.60526316, 0.18075928],
                [0.31162587, 0.49701937],
                [0.89799331, 0.28011404],
                [0.28388278, 0.1245121],
                [0.07322485, 0.38272383]])

```
In [19]: df1=df[df["New Cluster"]==0]
         df2=df[df["New Cluster"]==1]
         df3=df[df["New Cluster"]==2]
         plt.scatter(df1["Age"],df1["Income($)"],color="red")
         plt.scatter(df2["Age"],df2["Income($)"],color="green")
         plt.scatter(df3["Age"],df3["Income($)"],color="blue")
         plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="orange",ma
         plt.xlabel("Age")
         plt.ylabel("Income($)")
```

Out[19]: Text(0, 0.5, 'Income($)')
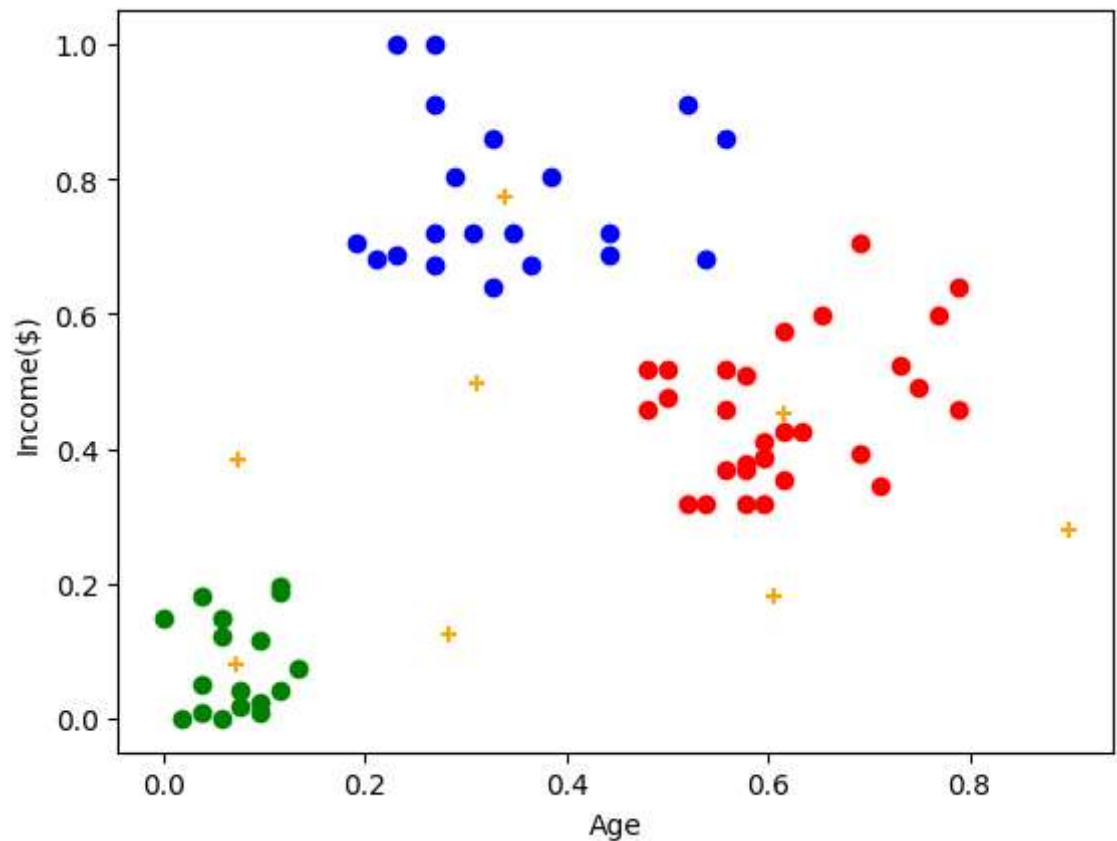

```
In [20]: k_rng=range(1,10)
         sse=[]
```

In [21]:
```
for k in k_rng:
  km=KMeans(n_clusters=k)
  km.fit(df[["Age","Income($)"]])
  sse.append(km.inertia_)
#km.inertia_ will give you the value of sum of square error
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

```
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\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 supp
ress the warning
  warnings.warn(
```
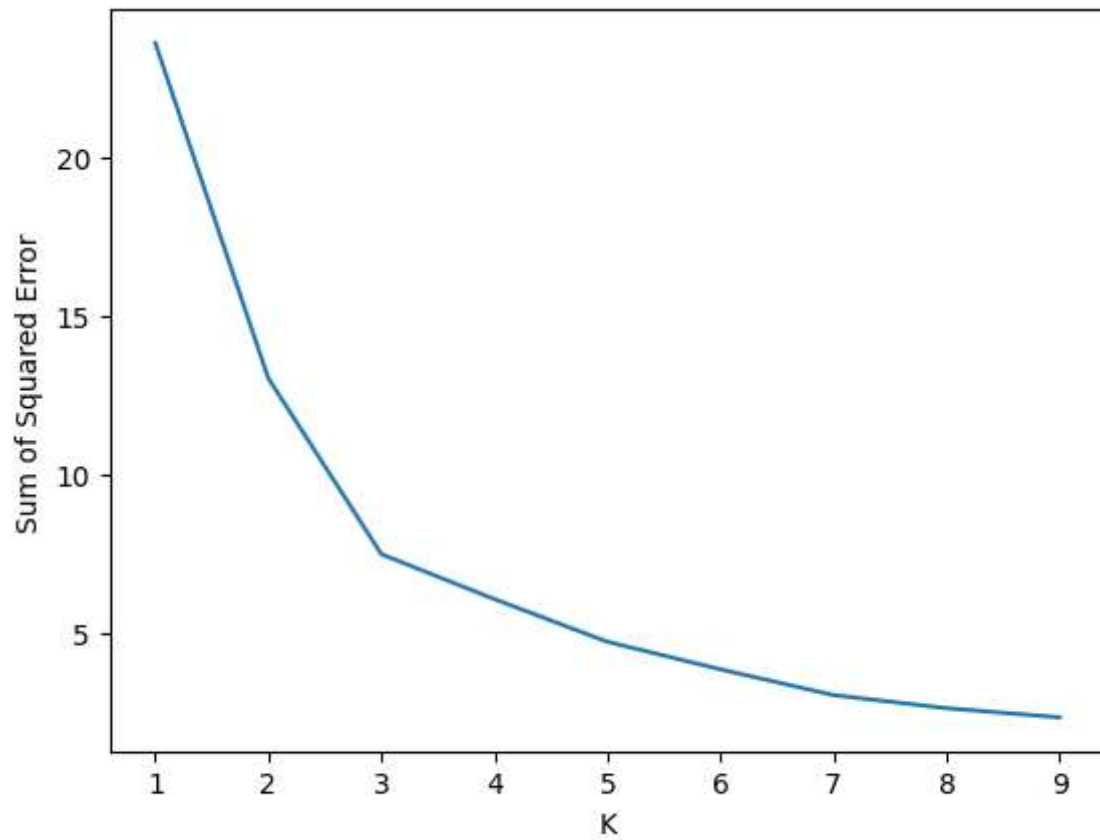
```
[23.58390615036361, 13.028938428018284, 7.492113413237458, 6.075958471031436,
4.733776701093291, 3.866293704700089, 3.0547174363693586, 2.6425203435360713,
2.354371853596292]
```

Out[21]: Text(0, 0.5, 'Sum of Squared Error')

In [ ]: