```
#To implement k-Means using any inbuild and external data set.
```

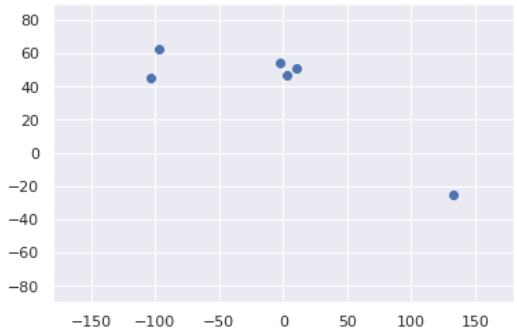
```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
from sklearn.cluster import KMeans
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use t
  import pandas.util.testing as tm
```

```
data = pd.read_csv('Countryclusters.csv')
data
```

	Country	Latitude	Longitude	Language
0	USA	44.97	-103.77	English
1	Canada	62.40	-96.80	English
2	France	46.75	2.40	French
3	UK	54.01	-2.53	English
4	Germany	51.15	10.40	German
5	Australia	-25.45	133.11	English

```
plt.scatter(data['Longitude'],data['Latitude'])
plt.xlim(-180,180)
plt.ylim(-90,90)
plt.show()
```




```
x = data.iloc[:,1:3] # 1t for rows and second for columns
x
```

	Latitude	Longitude
0	44.97	-103.77
1	62.40	-96.80
2	46.75	2.40
3	54.01	-2.53
4	51.15	10.40
5	-25.45	133.11

```
kmeans = KMeans(2)
kmeans.fit(x)
```

```
KMeans(n_clusters=2)
```

```
identified_clusters = kmeans.fit_predict(x)
identified_clusters
```

```
array([0, 0, 0, 0, 0, 1], dtype=int32)
```

```
data_with_clusters = data.copy()
data_with_clusters['Clusters'] = identified_clusters
plt.scatter(data_with_clusters['Longitude'],data_with_clusters['Latitude'],c=data_with_clusters['Clusters'],cmap='rainbow')
```

```
<matplotlib.collections.PathCollection at 0x7ff9064c9f90>
```