```
from google.colab import files
upload=files.upload()
```

Choose Files EastWestAirlines1.csv

- **EastWestAirlines1.csv**(text/csv) - 157167 bytes, last modified: 2/25/2023 - 100% done

```
Saving EastWestAirlines1.csv to EastWestAirlines1.csv
```

```
import pandas as pd
df=pd.read_csv("EastWestAirlines1.csv")
df.head()
```

	ID#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_t
0	1	28143	0	1	1	1	174	
1	2	19244	0	1	1	1	215	
2	3	41354	0	1	1	1	4123	
3	4	14776	0	1	1	1	500	
4	5	97752	0	4	1	1	43300	

```
df.dtypes
```

```
ID#                 int64
Balance             int64
Qual_miles          int64
cc1_miles           int64
cc2_miles           int64
cc3_miles           int64
Bonus_miles         int64
Bonus_trans         int64
Flight_miles_12mo   int64
Flight_trans_12     int64
Days_since_enroll   int64
Award?              int64
dtype: object
```

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

```
(3999, 12)
```

```
df1=df.corr()
X=df.iloc[:,0:12]
X.head()
```

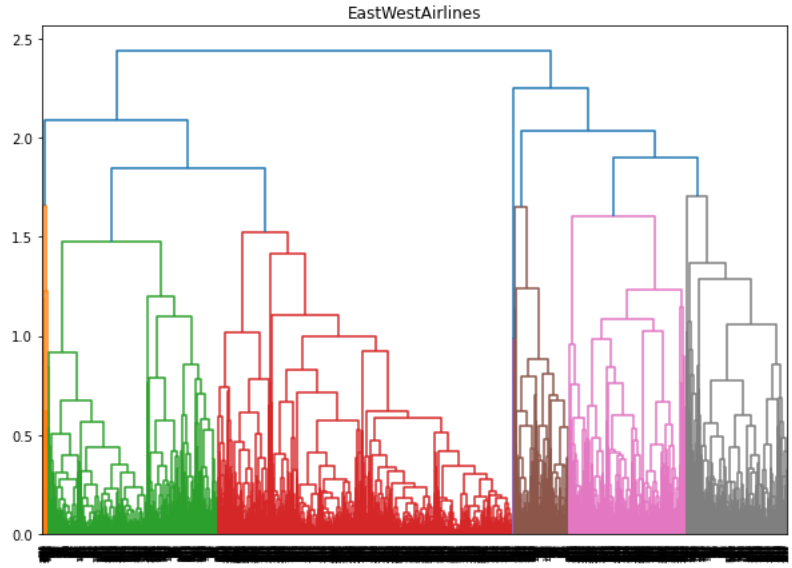
	ID#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_t
0	1	28143	0	1	1	1	174	
1	2	19244	0	1	1	1	215	
2	3	41354	0	1	1	1	4123	
3	4	14776	0	1	1	1	500	
4	5	97752	0	4	1	1	43300	

```
from sklearn.preprocessing import MinMaxScaler
MM=MinMaxScaler()
x=MM.fit_transform(X)
x
```

```
array([[0.00000000e+00, 1.65077268e-02, 0.00000000e+00, ...,
        0.00000000e+00, 8.43742464e-01, 0.00000000e+00],
       [2.48756219e-04, 1.12878760e-02, 0.00000000e+00, ...,
        0.00000000e+00, 8.39884254e-01, 0.00000000e+00],
       [4.97512438e-04, 2.42568502e-02, 0.00000000e+00, ...,
        0.00000000e+00, 8.47841813e-01, 0.00000000e+00],
       ...,
       [9.99502488e-01, 4.31694976e-02, 0.00000000e+00, ...,
        0.00000000e+00, 1.68796721e-01, 1.00000000e+00],
       [9.99751244e-01, 3.22018866e-02, 0.00000000e+00, ...,
        1.88679245e-02, 1.68676151e-01, 0.00000000e+00],
```

```
       [1.00000000e+00, 1.76908304e-03, 0.00000000e+00, ...,
        0.00000000e+00, 1.68314444e-01, 0.00000000e+00]])
```

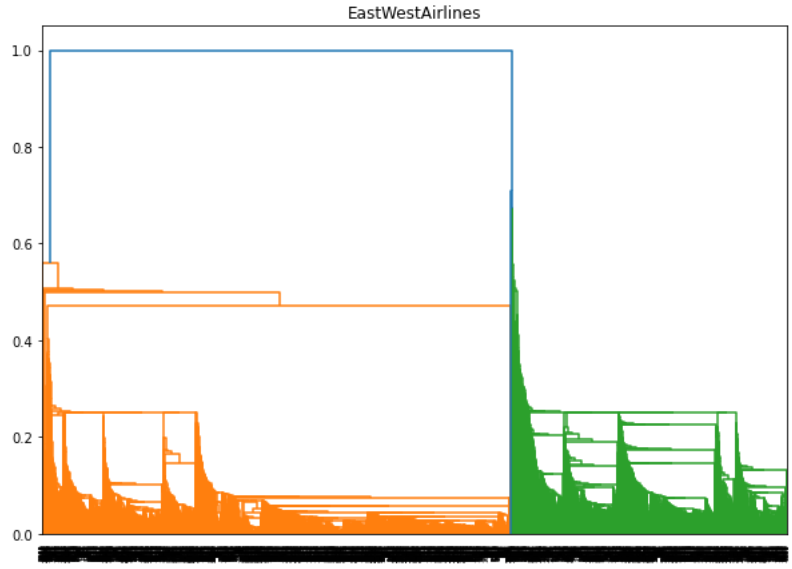
```
import scipy.cluster.hierarchy as shc
import matplotlib.pyplot as plt
plt.figure(figsize=(10,7))
plt.title("EastWestAirlines")
dend=shc.dendrogram(shc.linkage(x,method="complete"))
```




```
from sklearn.cluster import AgglomerativeClustering
AMC=AgglomerativeClustering(n_clusters=6,affinity="euclidean",linkage="complete")
y=AMC.fit_predict(x)
Y=pd.DataFrame(y)
Y.value_counts()
```

```
0    2493
2     627
1     542
4     298
3      35
5       4
dtype: int64
```

```
import scipy.cluster.hierarchy as shc
import matplotlib.pyplot as plt
plt.figure(figsize=(10,7))
plt.title("EastWestAirlines")
dend=shc.dendrogram(shc.linkage(x,method="single"))
```



```
from sklearn.cluster import AgglomerativeClustering
AMC=AgglomerativeClustering(n_clusters=3,affinity="euclidean",linkage="single")
y=AMC.fit_predict(x)
Y=pd.DataFrame(y)
Y.value_counts()
```

```
1    2518
0    1480
2       1
dtype: int64
```

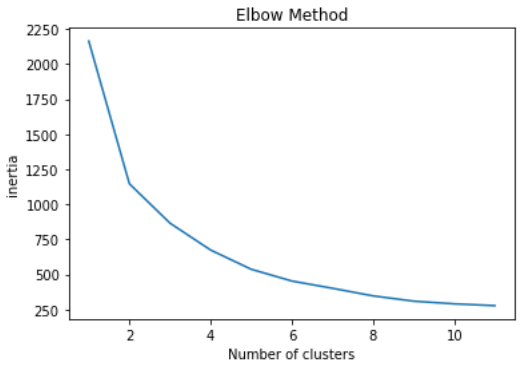
########## KMEANS ############*italicized text*

```
from sklearn.cluster import KMeans
Km=KMeans(n_clusters=5,n_init=25)
y=Km.fit_predict(X)
Y=pd.DataFrame(y)
Y.value_counts()
```

```
0    2493
4    1053
2     345
3      89
1      19
dtype: int64
```

```
inertia = []
for i in range(1,12):
    km = KMeans(n_clusters=i,random_state=30)
    km.fit(x)
    inertia.append(km.inertia_)

plt.plot(range(1, 12), inertia)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('inertia')
plt.show()
```



############### DBSCAN #################

```
from sklearn.cluster import DBSCAN
DBSCAN()
dbscan = DBSCAN(eps=3, min_samples=2)
dbscan.fit(X)
dbscan.labels_
d_new=pd.DataFrame(dbscan.labels_,columns=["cluster"])
d_new
d_new["cluster"].value_counts()
```

```
-1    3999
Name: cluster, dtype: int64
```

✓ 0s completed at 10:14 PM