—---------Import Important Libraries—--------------

```
import pandas as pd
import scipy.cluster.hierarchy as sch
from sklearn.cluster import AgglomerativeClustering
import numpy as np
import matplotlib.pyplot as plt
```

—-----------Read the Datesets—-----------

```
xls = pd.ExcelFile('Downloads/EastWestAirlines.xlsx')
data = pd.read_excel(xls,'data')
data
```

	ID#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?
0	1	28143	0	1	1	1	174	1	0	0	7000	0
1	2	19244	0	1	1	1	215	2	0	0	6968	0
2	3	41354	0	1	1	1	4123	4	0	0	7034	0
3	4	14776	0	1	1	1	500	1	0	0	6952	0
4	5	97752	0	4	1	1	43300	26	2077	4	6935	1
...	...	...	...	...	...	...	...	...	...	...	...	...
3994	4017	18476	0	1	1	1	8525	4	200	1	1403	1
3995	4018	64385	0	1	1	1	981	5	0	0	1395	1
3996	4019	73597	0	3	1	1	25447	8	0	0	1402	1
3997	4020	54899	0	1	1	1	500	1	500	1	1401	0
3998	4021	3016	0	1	1	1	0	0	0	0	1398	0

3999 rows × 12 columns

—-----------------Read first 6 rows—---------------------

```
data.head(6)
```

	ID#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?
0	1	28143	0	1	1	1	174	1	0	0	7000	0
1	2	19244	0	1	1	1	215	2	0	0	6968	0
2	3	41354	0	1	1	1	4123	4	0	0	7034	0
3	4	14776	0	1	1	1	500	1	0	0	6952	0
4	5	97752	0	4	1	1	43300	26	2077	4	6935	1
5	6	16420	0	1	1	1	0	0	0	0	6942	0

—-------------Check shape—----------------

```
data.shape
(3999,12)
```

—------------Define Normalization function—--------------

```
def norm_func(i):
    x = (i-i.min())/(i.max()-i.min())
    return(x)
```

—-----------Drop first column—----------------

```
divide = norm_func(data.iloc[:,1:])
divide
```

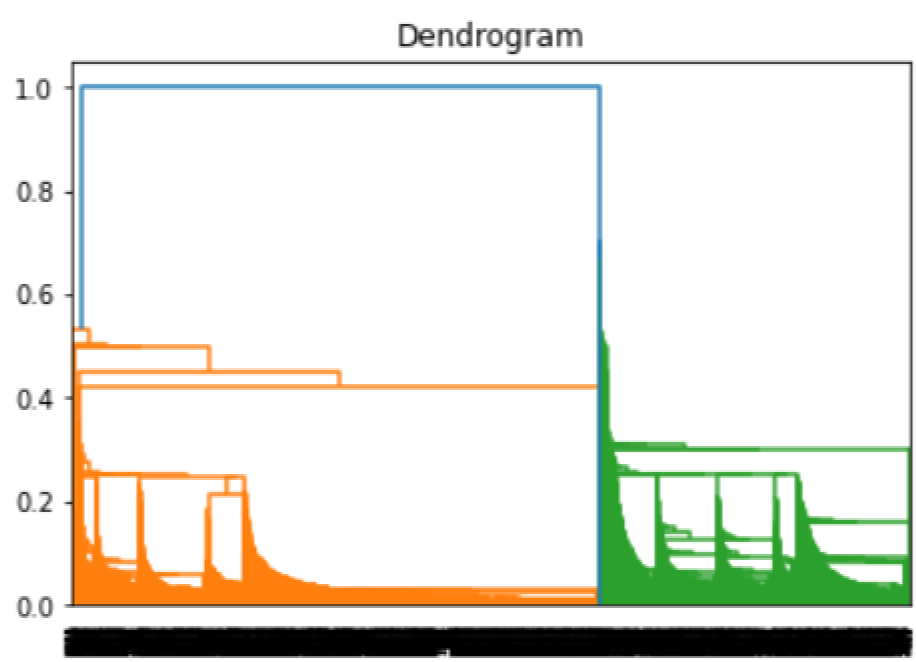
	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?
0	0.016508	0.0	0.00	0.0	0.0	0.000660	0.011628	0.000000	0.000000	0.843742	0.0
1	0.011288	0.0	0.00	0.0	0.0	0.000815	0.023256	0.000000	0.000000	0.839884	0.0
2	0.024257	0.0	0.00	0.0	0.0	0.015636	0.046512	0.000000	0.000000	0.847842	0.0
3	0.008667	0.0	0.00	0.0	0.0	0.001896	0.011628	0.000000	0.000000	0.837955	0.0
4	0.057338	0.0	0.75	0.0	0.0	0.164211	0.302326	0.067398	0.075472	0.835905	1.0
...	...	...	...	...	...	...	...	...	...	...	...
3994	0.010837	0.0	0.00	0.0	0.0	0.032330	0.046512	0.006490	0.018868	0.168917	1.0
3995	0.037766	0.0	0.00	0.0	0.0	0.003720	0.058140	0.000000	0.000000	0.167953	1.0
3996	0.043169	0.0	0.50	0.0	0.0	0.096505	0.093023	0.000000	0.000000	0.168797	1.0
3997	0.032202	0.0	0.00	0.0	0.0	0.001896	0.011628	0.016225	0.018868	0.168676	0.0
3998	0.001769	0.0	0.00	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.168314	0.0

3999 rows × 11 columns

—--------------Plot Dendrogram—--------------------

dendrogram = sch.dendrogram(sch.linkage(divide, method = 'single'))
plt.title('Dendrogram')

```
Text(0.5, 1.0, 'Dendrogram')
```



—------------Agglomerative Clustering—--------------------

a = AgglomerativeClustering (n_clusters=4, affinity='euclidean', linkage='single')
a

```
AgglomerativeClustering(linkage='single', n_clusters=4)
```

—------------Prediction—----------------

predict = a.fit_predict(divide)
cluster = pd.DataFrame(predict, columns = ['cluster'])

cluster

	cluster
0	1
1	1
2	1
3	1
4	0
...	...
3994	0
3995	0
3996	0
3997	1
3998	1

3999 rows × 1 columns

—-------------Read Data—------------------

	ID#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?
0	1	28143	0	1	1	1	174	1	0	0	7000	0
1	2	19244	0	1	1	1	215	2	0	0	6968	0
2	3	41354	0	1	1	1	4123	4	0	0	7034	0
3	4	14776	0	1	1	1	500	1	0	0	6952	0
4	5	97752	0	4	1	1	43300	26	2077	4	6935	1
...	...	...	...	...	...	...	...	...	...	...	...	...
3994	4017	18476	0	1	1	1	8525	4	200	1	1403	1
3995	4018	64385	0	1	1	1	981	5	0	0	1395	1

—----------Import Important Libraries (KMeans)—----------------

import pandas as pd
import numpy as np
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

—-----------Read Dataset—------------------------

xls = pd.ExcelFile('Downloads/EastWestAirlines.xlsx')
data = pd.read_excel(xls,'data')
data

	ID#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?
0	1	28143	0	1	1	1	174	1	0	0	7000	0
1	2	19244	0	1	1	1	215	2	0	0	6968	0
2	3	41354	0	1	1	1	4123	4	0	0	7034	0
3	4	14776	0	1	1	1	500	1	0	0	6952	0
4	5	97752	0	4	1	1	43300	26	2077	4	6935	1
...	...	...	...	...	...	...	...	...	...	...	...	...
3994	4017	18476	0	1	1	1	8525	4	200	1	1403	1
3995	4018	64385	0	1	1	1	981	5	0	0	1395	1
3996	4019	73597	0	3	1	1	25447	8	0	0	1402	1
3997	4020	54899	0	1	1	1	500	1	500	1	1401	0
3998	4021	3016	0	1	1	1	0	0	0	0	1398	0

3999 rows × 12 columns

—-------------Scale Dataset—-----------------

```
scaler = StandardScaler()
scaled = scaler.fit_transform(data.iloc[:,1:])
scaled
```

```
array([[-4.51140783e-01, -1.86298687e-01, -7.69578406e-01, ...,
        -3.62167870e-01,  1.39545434e+00, -7.66919299e-01],
       [-5.39456874e-01, -1.86298687e-01, -7.69578406e-01, ...,
        -3.62167870e-01,  1.37995704e+00, -7.66919299e-01],
       [-3.20031232e-01, -1.86298687e-01, -7.69578406e-01, ...,
        -3.62167870e-01,  1.41192021e+00, -7.66919299e-01],
       ...,
       [-4.29480975e-05, -1.86298687e-01,  6.83121167e-01, ...,
        -3.62167870e-01, -1.31560393e+00,  1.30391816e+00],
       [-1.85606976e-01, -1.86298687e-01, -7.69578406e-01, ...,
        -9.85033311e-02, -1.31608822e+00, -7.66919299e-01],
       [-7.00507951e-01, -1.86298687e-01, -7.69578406e-01, ...,
        -3.62167870e-01, -1.31754109e+00, -7.66919299e-01]])
```

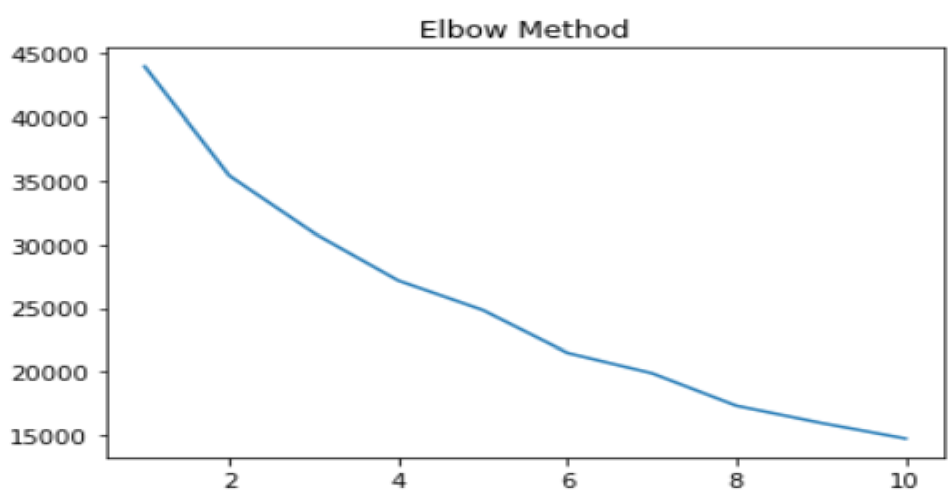
—-----------Define function—------------------

```
wss = []
for i in range (1,11):
    kmeans = KMeans (n_clusters = i)
    kmeans.fit(scaled)
    wss.append(kmeans.inertia_)
```

—---------Plot function—--------------------

```
plt.plot(range(1,11),wss)
plt.title('Elbow Method')
```

```
Text(0.5, 1.0, 'Elbow Method')
```



—------------KMeans Clustering—-----------------------

```
clusters = KMeans (n_clusters = 4)
clusters.fit(scaled)
```

```
KMeans(n_clusters=4)
```

```
clusters.labels_
```

```
array([0, 0, 0, ..., 2, 0, 0])
```

```
data['New Clusters'] = clusters.labels_
```

```
clusters.cluster_centers_
```

```
array([[-2.98337183e-01, -5.97435910e-02, -6.14895460e-01,
         3.28947498e-02, -6.07426678e-02, -5.18339919e-01,
        -4.91701345e-01, -1.84890271e-01, -1.97468685e-01,
        -2.07730169e-01, -3.48824498e-01],
       [ 1.20311027e+00,  8.16669419e-01,  9.94472314e-02,
         1.43672710e-01, -6.27665798e-02,  6.37799889e-01,
         1.59051553e+00,  3.49563610e+00,  3.76072941e+00,
         2.79571174e-01,  9.21799343e-01],
       [ 4.22858440e-01,  1.18887629e-02,  1.18619612e+00,
        -8.24280146e-02, -5.47937838e-02,  9.00445591e-01,
         7.40521626e-01, -9.32365605e-02, -1.03427805e-01,
         3.69804678e-01,  5.62606696e-01],
       [ 6.39719256e-01, -8.44329231e-02,  1.02208440e+00,
        -9.82418871e-02,  1.56462993e+01,  3.17969131e+00,
         1.71461374e+00,  3.32926913e-02,  5.96953922e-02,
         2.39872612e-01,  3.37527346e-01]])
```

—----------------Grouping data—----------------

data.groupby('New Clusters').agg(['mean']).reset_index()

	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?
	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
396922	43617.468824	97.782952	1.215075	1.019337	1.000395	4646.683110	6.890292	200.964878	0.623915	3688.902920	0.201657
340476	194830.404762	775.863095	2.196429	1.035714	1.000000	32546.404762	26.875000	5354.065476	15.636905	4695.839286	0.815476
238690	116226.687988	153.670047	3.694228	1.002340	1.001560	38932.005460	18.720749	330.292512	0.983619	4886.372855	0.643526
366667	138061.400000	78.800000	3.466667	1.000000	4.066667	93927.866667	28.066667	506.666667	1.600000	4613.866667	0.533333

—-----------Read Data—--------------------------

data

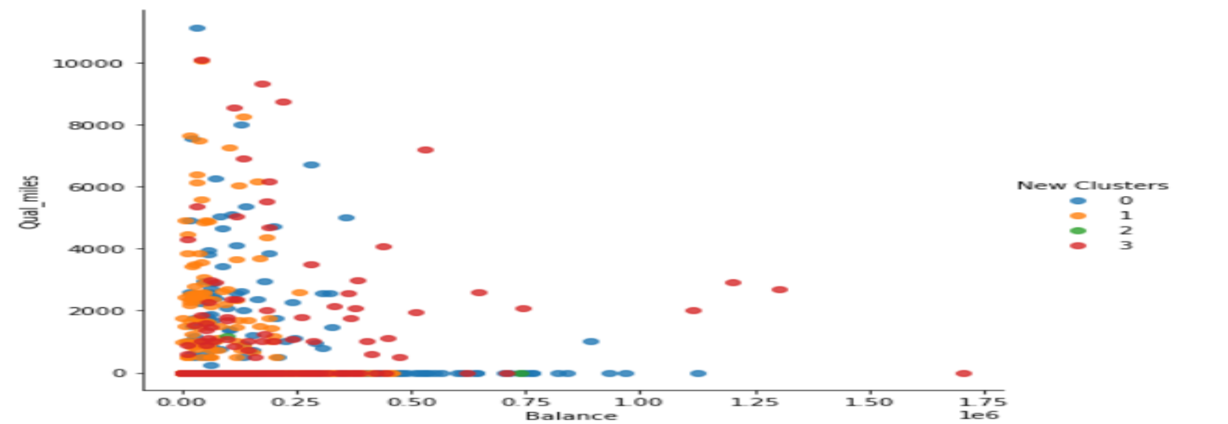
D#	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award?	New Clusters
1	28143	0	1	1	1	174	1	0	0	7000	0	1
2	19244	0	1	1	1	215	2	0	0	6968	0	1
3	41354	0	1	1	1	4123	4	0	0	7034	0	1
4	14776	0	1	1	1	500	1	0	0	6952	0	1
5	97752	0	4	1	1	43300	26	2077	4	6935	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...
017	18476	0	1	1	1	8525	4	200	1	1403	1	1
018	64385	0	1	1	1	981	5	0	0	1395	1	1
019	73597	0	3	1	1	25447	8	0	0	1402	1	0
020	54899	0	1	1	1	500	1	500	1	1401	0	1
021	3016	0	1	1	1	0	0	0	0	1398	0	1

—-----------Plot 1—----------------------------------

import seaborn as sns

sns.lmplot('Balance','Qual_miles', data=data, hue='New Clusters', fit_reg=False,size=6)

<seaborn.axisgrid.FacetGrid at 0x12ea07e1970>



—---------------Plot 2—---------------------------

sns.lmplot('Balance','Bonus_miles', data=data, hue='New Clusters', fit_reg=False,size=6)

<seaborn.axisgrid.FacetGrid at 0x12ea18625e0>