In [1]:

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

In [2]:

```
data=pd.read_excel('EastWestAirlines.xlsx',sheet_name='data')
data.head()
```

Out[2]:

| | ID# | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | F |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 | |
| **1** | 2 | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 | |
| **2** | 3 | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 | |
| **3** | 4 | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 | |
| **4** | 5 | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |

In [3]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3999 entries, 0 to 3998
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   ID#                3999 non-null   int64
 1   Balance            3999 non-null   int64
 2   Qual_miles         3999 non-null   int64
 3   cc1_miles          3999 non-null   int64
 4   cc2_miles          3999 non-null   int64
 5   cc3_miles          3999 non-null   int64
 6   Bonus_miles        3999 non-null   int64
 7   Bonus_trans        3999 non-null   int64
 8   Flight_miles_12mo  3999 non-null   int64
 9   Flight_trans_12    3999 non-null   int64
 10  Days_since_enroll  3999 non-null   int64
 11  Award?             3999 non-null   int64
dtypes: int64(12)
memory usage: 375.0 KB
```

In [4]:

```
data.shape
```

Out[4]:

```
(3999, 12)
```

In [5]:

```
air=data.drop(['ID#','Award?'], axis=1)
air
```

Loading [MathJax]/extensions/Safe.js

Out[5]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 | |
| **1** | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 | |
| **2** | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 | |
| **3** | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 | |
| **4** | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 200 | |
| **3995** | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | 0 | |
| **3996** | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | 0 | |
| **3997** | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 500 | |
| **3998** | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | |

3999 rows × 10 columns

In [6]:

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

In [7]:

```
df_norm = norm_func(air.iloc[:,:])
df_norm
```

Out[7]:

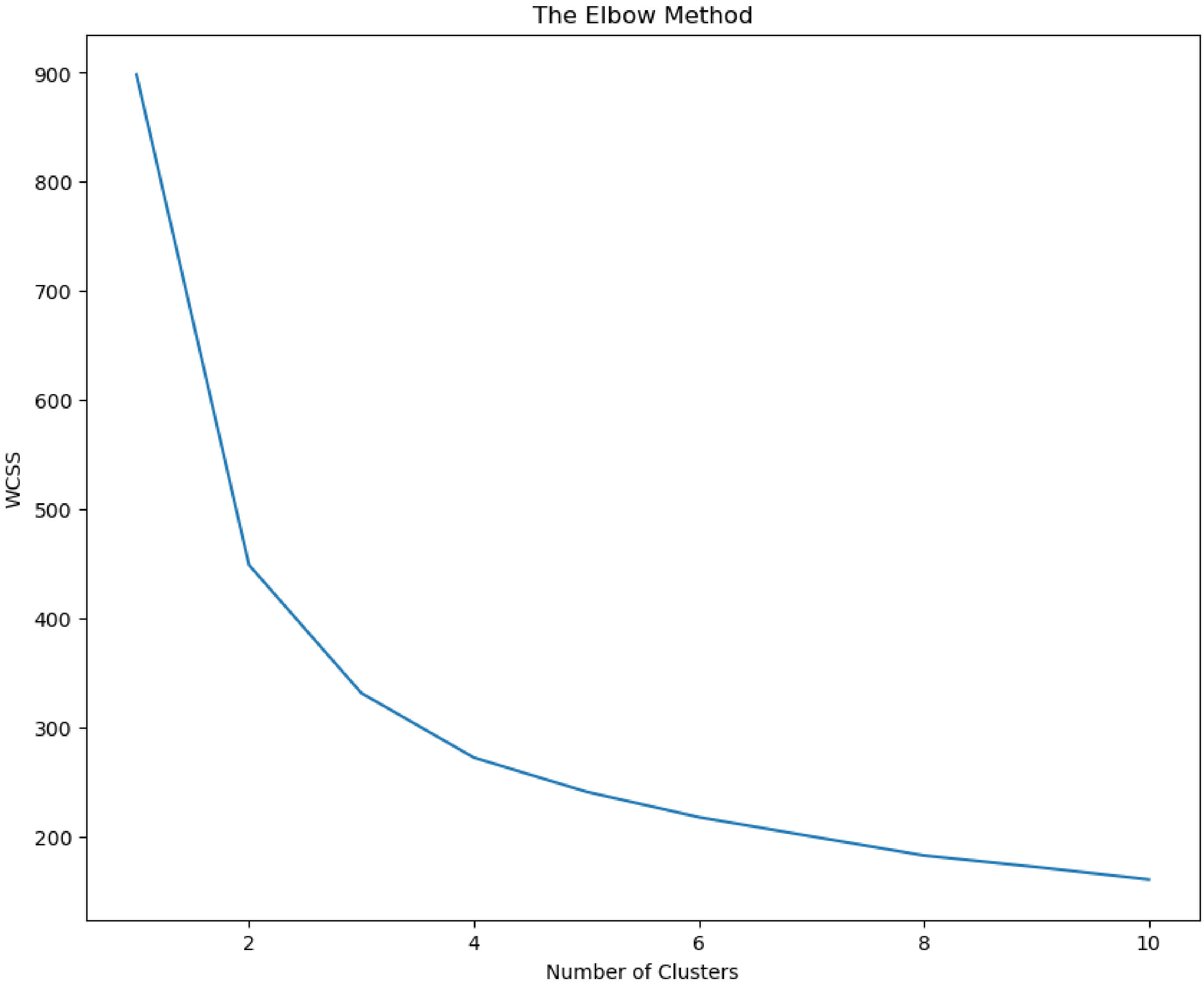
| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fl |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 0.016508 | 0.0 | 0.00 | 0.0 | 0.0 | 0.000660 | 0.011628 | 0.000000 | |
| **1** | 0.011288 | 0.0 | 0.00 | 0.0 | 0.0 | 0.000815 | 0.023256 | 0.000000 | |
| **2** | 0.024257 | 0.0 | 0.00 | 0.0 | 0.0 | 0.015636 | 0.046512 | 0.000000 | |
| **3** | 0.008667 | 0.0 | 0.00 | 0.0 | 0.0 | 0.001896 | 0.011628 | 0.000000 | |
| **4** | 0.057338 | 0.0 | 0.75 | 0.0 | 0.0 | 0.164211 | 0.302326 | 0.067398 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 0.010837 | 0.0 | 0.00 | 0.0 | 0.0 | 0.032330 | 0.046512 | 0.006490 | |
| **3995** | 0.037766 | 0.0 | 0.00 | 0.0 | 0.0 | 0.003720 | 0.058140 | 0.000000 | |
| **3996** | 0.043169 | 0.0 | 0.50 | 0.0 | 0.0 | 0.096505 | 0.093023 | 0.000000 | |
| **3997** | 0.032202 | 0.0 | 0.00 | 0.0 | 0.0 | 0.001896 | 0.011628 | 0.016225 | |
| **3998** | 0.001769 | 0.0 | 0.00 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.000000 | |

3999 rows × 10 columns

In [8]:

```
fig = plt.figure(figsize=(10, 8))
WCSS = []
for i in range(1, 11):
    clf = KMeans(n_clusters=i)
    clf.fit(df_norm)
    WCSS.append(clf.inertia_)
plt.plot(range(1, 11), WCSS)
plt.title('The Elbow Method')
plt.ylabel('WCSS')
plt.xlabel('Number of Clusters')
```

Loading [MathJax]/extensions/Safe.js



In [9]:

```
WCSS
```

Out[9]:

```
[898.2705822007467,
 448.9989194445697,
 331.4315185424921,
 272.39254909046446,
 241.07572112696283,
 217.7640341949783,
 200.21286104789934,
 182.7151065322523,
 172.10646043690986,
 160.7291592574553]
```

In [10]:

```
clf = KMeans(n_clusters=4)
y_kmeans = clf.fit_predict(df_norm)
```

In [11]:

```
y_kmeans
```

Out[11]:

```
array([2, 2, 2, ..., 1, 0, 0])
```

In [12]:

```
clf.cluster_centers_
```

Out[12]:

```
array([[2.36724363e-02, 1.09788344e-02, 2.37926136e-02, 1.10085227e-02,
        1.24289773e-03, 1.69214392e-02, 8.35177722e-02, 1.28437586e-02,
        2.19500429e-02, 2.65314415e-01],
       [5.01893530e-02, 1.09101606e-02, 6.26102941e-01, 7.35294118e-04,
        3.30882353e-03, 1.18152084e-01, 2.09917921e-01, 1.28678021e-02,
        2.18091010e-02, 3.63864204e-01],
       [3.87082781e-02, 1.52707004e-02, 3.21888412e-02, 1.07296137e-02,
        1.50214592e-03, 1.96348371e-02, 9.29234455e-02, 1.58009047e-02,
        2.81156369e-02, 6.93991623e-01],
       [8.05499702e-02, 1.47843877e-02, 7.54021448e-01, 6.70241287e-04,
        8.71313673e-03, 1.78247067e-01, 2.29082237e-01, 1.93799495e-02,
        3.37144013e-02, 7.44413255e-01]])
```

In [13]:

```
clf.inertia_
```

Out[13]:

```
272.3927662116429
```

In [14]:

```
md=pd.Series(y_kmeans)
air['clust']=md
```

Out[14]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 | |
| **1** | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 | |
| **2** | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 | |
| **3** | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 | |
| **4** | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 200 | |
| **3995** | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | 0 | |
| **3996** | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | 0 | |
| **3997** | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 500 | |
| **3998** | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | |

3999 rows × 11 columns

In [15]:

```
air.groupby(air.clust).mean()
```

Out[15]:

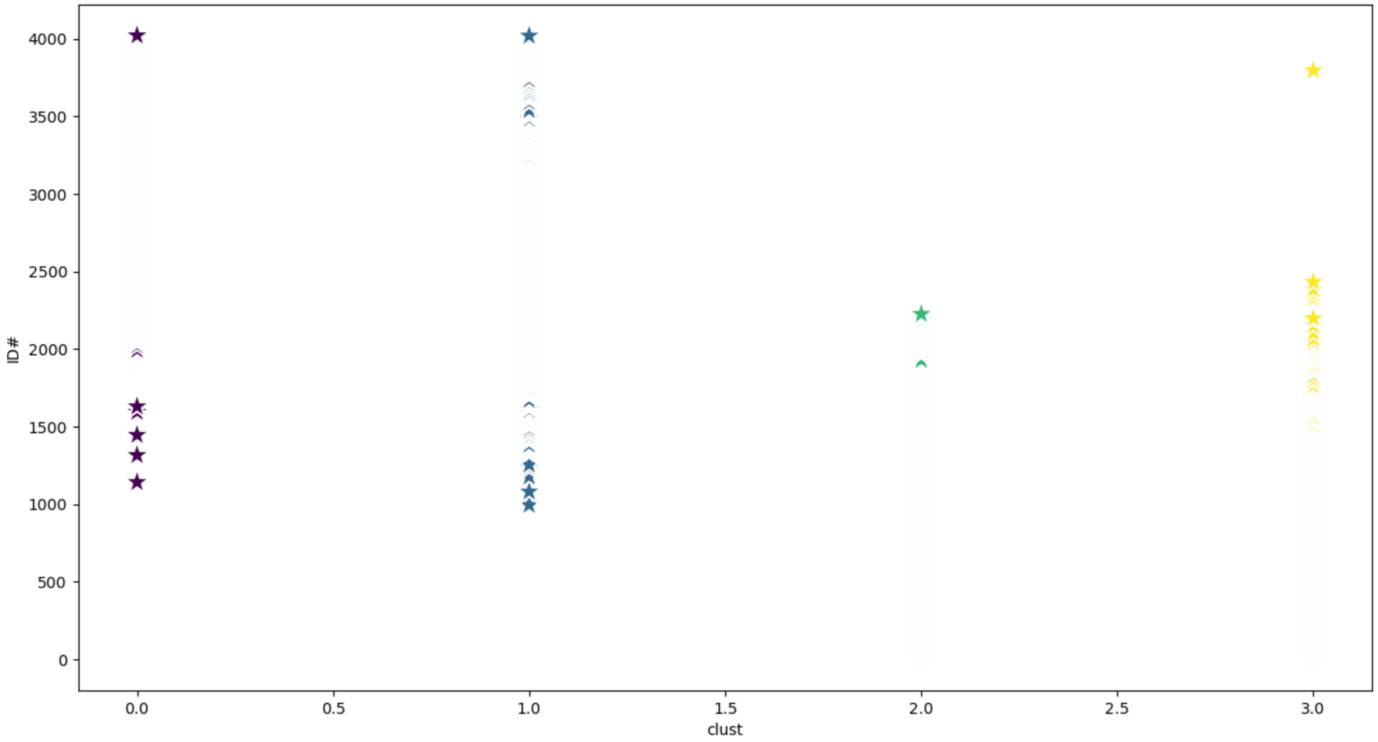
| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12r |
|---|---|---|---|---|---|---|---|---|
| **clust** | | | | | | | | |
| **0** | 40357.669034 | 122.392045 | 1.095170 | 1.022017 | 1.004972 | 4461.929688 | 7.182528 | 395.8061 |
| **1** | 85554.287187 | 121.805596 | 3.503682 | 1.001473 | 1.013255 | 31160.790869 | 18.055965 | 397.1310 |
| **2** | 65991.343348 | 170.237768 | 1.128755 | 1.021459 | 1.006009 | 5177.412017 | 7.991416 | 486.9364 |
| **3** | 137264.839357 | 164.595716 | 4.016064 | 1.001339 | 1.034806 | 46974.539491 | 19.696118 | 596.4323 |

In [16]:

```
plt.figure(figsize=(15,8))
sn.scatterplot(air['clust'],data['ID#'],c=clf.labels_,s=300,marker='*')
plt.show();
```

Loading [MathJax]/extensions/Safe.js

```
C:\Users\ROHIT\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pas
s the following variables as keyword args: x, y. From version 0.12, the only valid posit
ional argument will be `data`, and passing other arguments without an explicit keyword w
ill result in an error or misinterpretation.
  warnings.warn(
```



In [17]:
```
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
```

In [18]:
```
air
```

Out[18]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 | |
| **1** | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 | |
| **2** | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 | |
| **3** | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 | |
| **4** | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 200 | |
| **3995** | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | 0 | |
| **3996** | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | 0 | |
| **3997** | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 500 | |
| **3998** | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | |

3999 rows × 11 columns

In [19]:
```
array=air.values
array
```

```
Out[19]: array([[28143,     0,     1, ...,     0,  7000,     2],
                [19244,     0,     1, ...,     0,  6968,     2],
                [41354,     0,     1, ...,     0,  7034,     2],
                ...,
                [73597,     0,     3, ...,     0,  1402,     1],
                [54899,     0,     1, ...,     1,  1401,     0],
                [ 3016,     0,     1, ...,     0,  1398,     0]], dtype=int64)
```

```
In [20]: stscaler = StandardScaler().fit(array)
         X = stscaler.transform(array)
         X
```

```
Out[20]: array([[-4.51140783e-01, -1.86298687e-01, -7.69578406e-01, ...,
                 -3.62167870e-01,  1.39545434e+00,  6.04380405e-01],
                [-5.39456874e-01, -1.86298687e-01, -7.69578406e-01, ...,
                 -3.62167870e-01,  1.37995704e+00,  6.04380405e-01],
                [-3.20031232e-01, -1.86298687e-01, -7.69578406e-01, ...,
                 -3.62167870e-01,  1.41192021e+00,  6.04380405e-01],
                ...,
                [-4.29480975e-05, -1.86298687e-01,  6.83121167e-01, ...,
                 -3.62167870e-01, -1.31560393e+00, -2.75138241e-01],
                [-1.85606976e-01, -1.86298687e-01, -7.69578406e-01, ...,
                 -9.85033311e-02, -1.31608822e+00, -1.15465689e+00],
                [-7.00507951e-01, -1.86298687e-01, -7.69578406e-01, ...,
                 -3.62167870e-01, -1.31754109e+00, -1.15465689e+00]])
```

```
In [21]: dbscan = DBSCAN(eps=0.70, min_samples=10)
         dbscan.fit(X)
```

```
Out[21]: DBSCAN(eps=0.7, min_samples=10)
```

```
In [22]: dbscan.labels_
```

```
Out[22]: array([0, 0, 0, ..., 5, 6, 6], dtype=int64)
```

```
In [23]: c=pd.DataFrame(dbscan.labels_,columns=['cluster'])
```

```
In [24]: c
```

Out[24]:

|  | cluster |
|---|---|
| **0** | 0 |
| **1** | 0 |
| **2** | 0 |
| **3** | 0 |
| **4** | -1 |
| **...** | ... |
| **3994** | 6 |
| **3995** | 6 |
| **3996** | 5 |
| **3997** | 6 |
| **3998** | 6 |

3999 rows × 1 columns

In [25]:

```
df = pd.concat([data,c],axis=1)
df
```

Out[25]:

| | ID# | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12m |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | |
| **1** | 2 | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | |
| **2** | 3 | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | |
| **3** | 4 | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | |
| **4** | 5 | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 207 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 4017 | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 20 |
| **3995** | 4018 | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | |
| **3996** | 4019 | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | |
| **3997** | 4020 | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 50 |
| **3998** | 4021 | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | |

3999 rows × 13 columns

In [26]:

```
d1=dbscan.labels_
d1
```

Out[26]:

```
array([0, 0, 0, ..., 5, 6, 6], dtype=int64)
```

In [27]:

```
import sklearn
sklearn.metrics.silhouette_score(X, d1)
```

Out[27]:

```
0.0983201597948486
```

In [28]:

```
from sklearn.cluster import KMeans
clf = KMeans(n_clusters=5)
y_kmeans = clf.fit_predict(X)
```

In [29]:

```
y_kmeans
```

Out[29]:

```
array([1, 1, 1, ..., 4, 4, 4])
```

In [30]:

```
cl1=pd.DataFrame(y_kmeans,columns=['Kcluster'])
cl1
```

Loading [MathJax]/extensions/Safe.js

Out[30]:

|  | Kcluster |
|---|---|
| **0** | 1 |
| **1** | 1 |
| **2** | 1 |
| **3** | 1 |
| **4** | 3 |
| **...** | ... |
| **3994** | 4 |
| **3995** | 4 |
| **3996** | 4 |
| **3997** | 4 |
| **3998** | 4 |

3999 rows × 1 columns

In [31]:
```
df1 = pd.concat([df,cl1],axis=1)
df1
```

Out[31]:

|  | ID# | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12m |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 28143 | 0 | 1 | 1 | 1 | 174 | 1 |  |
| **1** | 2 | 19244 | 0 | 1 | 1 | 1 | 215 | 2 |  |
| **2** | 3 | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 |  |
| **3** | 4 | 14776 | 0 | 1 | 1 | 1 | 500 | 1 |  |
| **4** | 5 | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |  |
| **3994** | 4017 | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 2 |
| **3995** | 4018 | 64385 | 0 | 1 | 1 | 1 | 981 | 5 |  |
| **3996** | 4019 | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 |  |
| **3997** | 4020 | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 5 |
| **3998** | 4021 | 3016 | 0 | 1 | 1 | 1 | 0 | 0 |  |

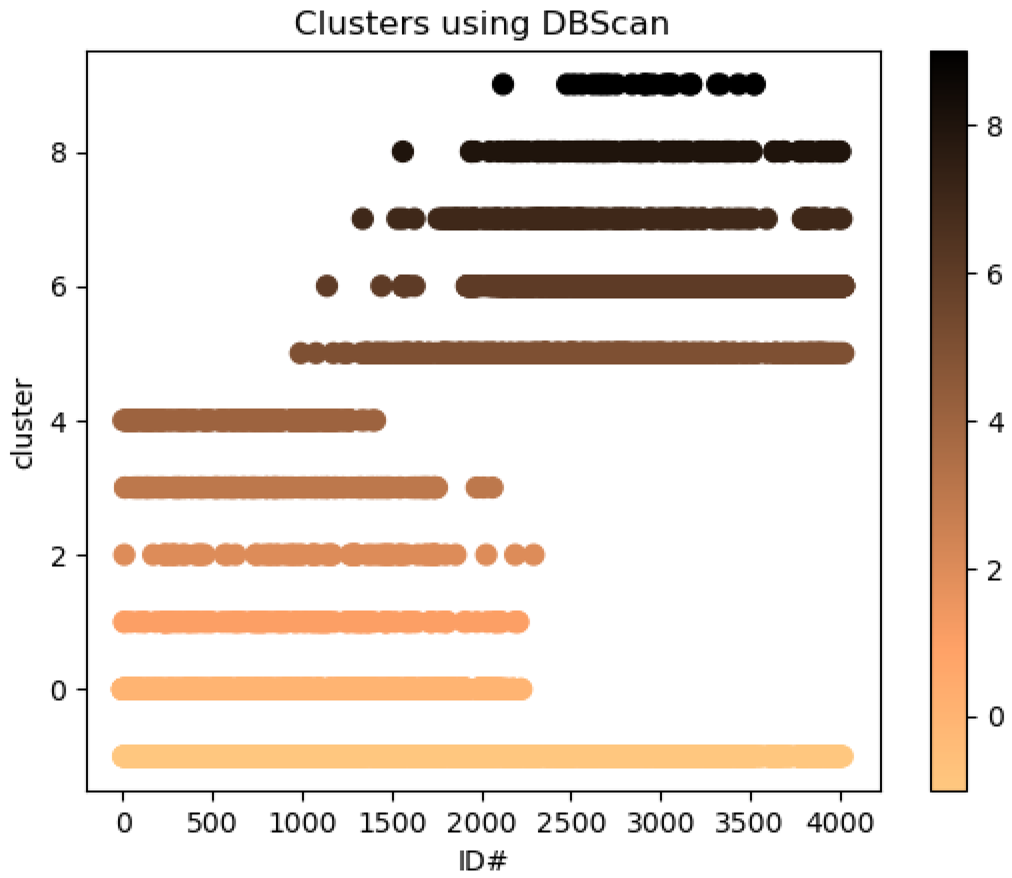
3999 rows × 14 columns

In [32]:
```
sklearn.metrics.silhouette_score(X, y_kmeans)
```

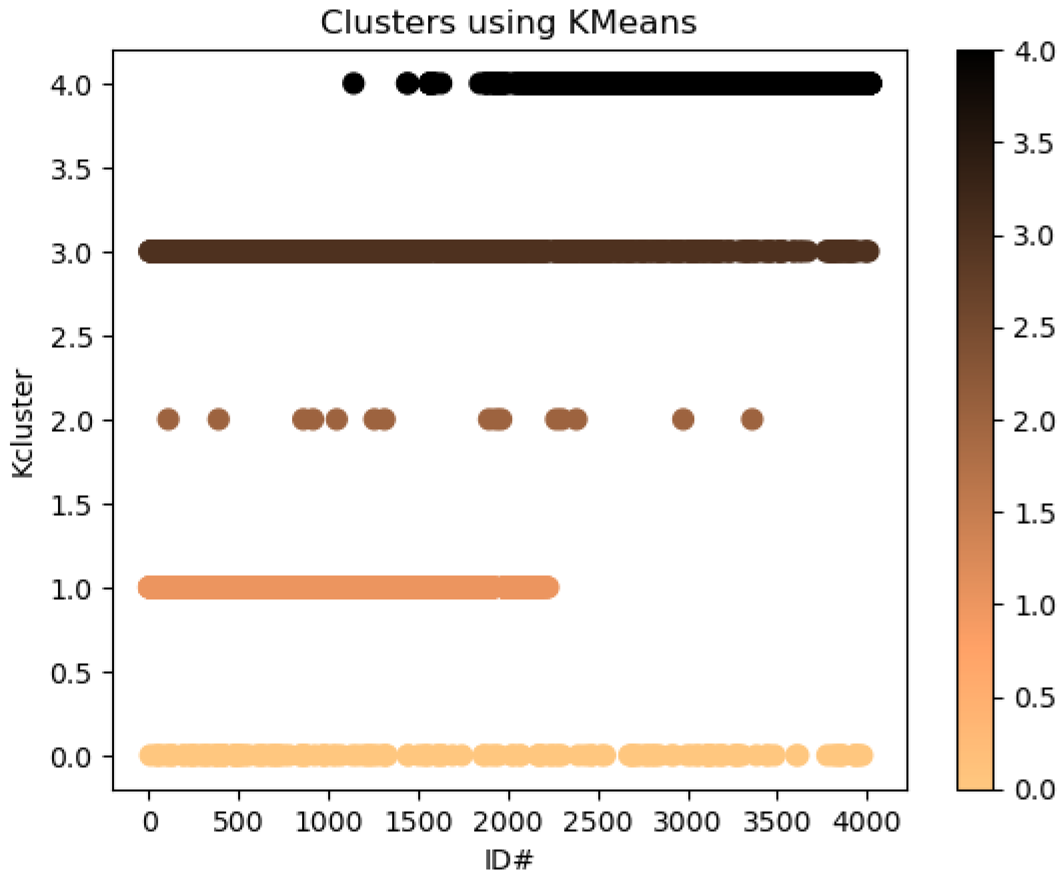
Out[32]:
```
0.30326300970492936
```

In [33]:
```
df.plot(x="ID#",y ="cluster",c=dbscan.labels_ ,kind="scatter",s=50 ,cmap=plt.cm.copper_r
plt.title('Clusters using DBScan')
plt.xlabel("ID#")
plt.ylabel("cluster")
```

Out[33]:
```
Text(0, 0.5, 'cluster')
```



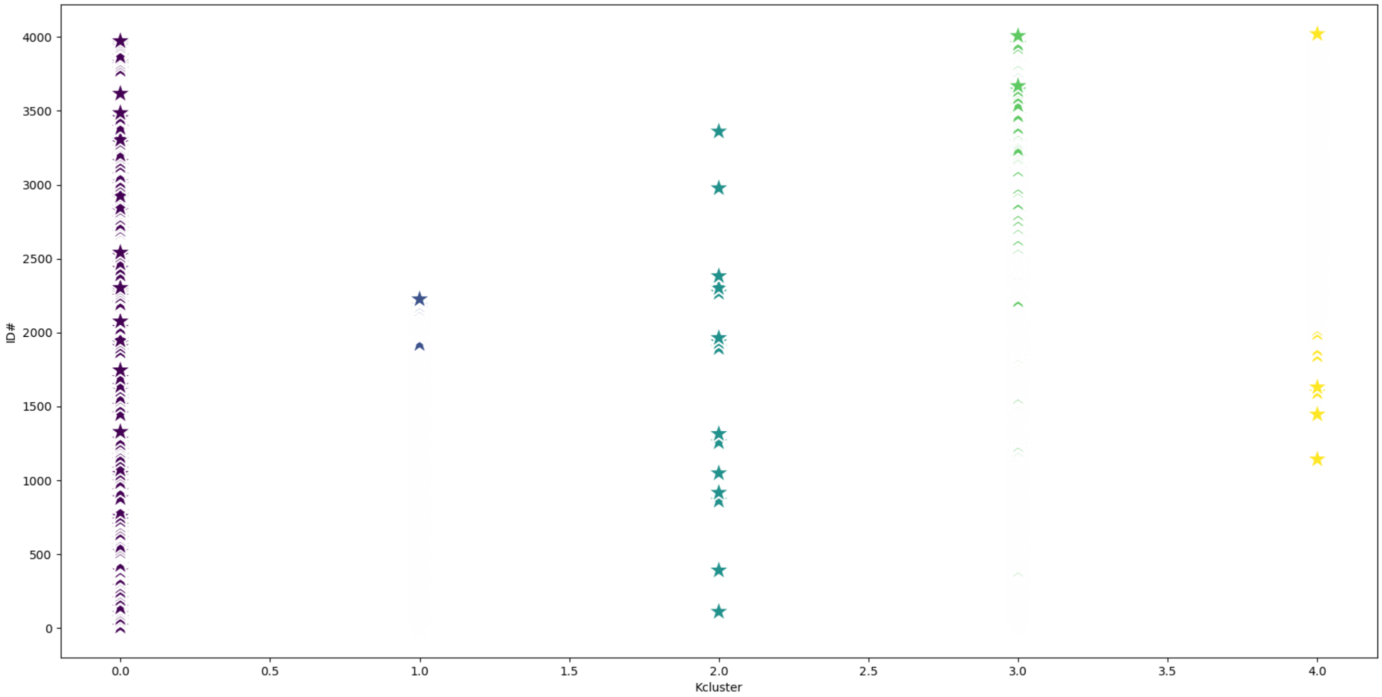
```
In [34]: df1.plot(x="ID#",y ="Kcluster",c=y_kmeans ,kind="scatter",s=50 ,cmap=plt.cm.copper_r)
         plt.title('Clusters using KMeans')

Out[34]: Text(0.5, 1.0, 'Clusters using KMeans')
```



```
In [35]: plt.figure(figsize=(20,10))
Loading [MathJax]/extensions/Safe.js t(df1['Kcluster'],df1['ID#'],c=clf.labels_,s=400,marker='*')
```

```
plt.show();
```

```
C:\Users\ROHIT\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pas
s the following variables as keyword args: x, y. From version 0.12, the only valid posit
ional argument will be `data`, and passing other arguments without an explicit keyword w
ill result in an error or misinterpretation.
  warnings.warn(
```



In [36]:
```
data
```

Out[36]:

|  | ID# | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12m |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 28143 | 0 | 1 | 1 | 1 | 174 | 1 |  |
| **1** | 2 | 19244 | 0 | 1 | 1 | 1 | 215 | 2 |  |
| **2** | 3 | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 |  |
| **3** | 4 | 14776 | 0 | 1 | 1 | 1 | 500 | 1 |  |
| **4** | 5 | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 20 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |  |
| **3994** | 4017 | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 2 |
| **3995** | 4018 | 64385 | 0 | 1 | 1 | 1 | 981 | 5 |  |
| **3996** | 4019 | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 |  |
| **3997** | 4020 | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 5 |
| **3998** | 4021 | 3016 | 0 | 1 | 1 | 1 | 0 | 0 |  |

3999 rows × 12 columns

In [37]:
```
air=data.drop(['ID#','Award?'],axis=1)
```

In [38]:
```
air
```

Loading [MathJax]/extensions/Safe.js

Out[38]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 | |
| **1** | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 | |
| **2** | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 | |
| **3** | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 | |
| **4** | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 200 | |
| **3995** | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | 0 | |
| **3996** | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | 0 | |
| **3997** | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 500 | |
| **3998** | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | |

3999 rows × 10 columns

In [39]:

```python
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
air_subset = pd.DataFrame(scaler.fit_transform(air.iloc[:,1:7]))
air_subset
```

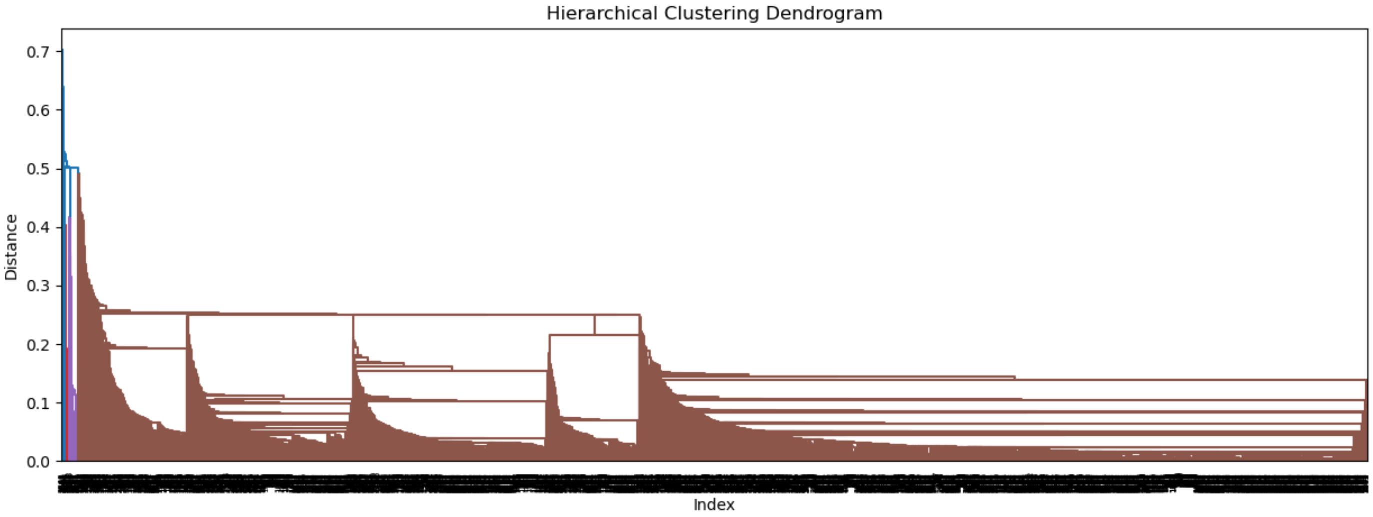
Out[39]:

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| **0** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.702786 | -1.104065 |
| **1** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.701088 | -0.999926 |
| **2** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.539253 | -0.791649 |
| **3** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.689286 | -1.104065 |
| **4** | -0.186299 | 1.409471 | -0.098242 | -0.062767 | 1.083121 | 1.499394 |
| **...** | ... | ... | ... | ... | ... | ... |
| **3994** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.356960 | -0.791649 |
| **3995** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.669367 | -0.687511 |
| **3996** | -0.186299 | 0.683121 | -0.098242 | -0.062767 | 0.343804 | -0.375096 |
| **3997** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.689286 | -1.104065 |
| **3998** | -0.186299 | -0.769578 | -0.098242 | -0.062767 | -0.709992 | -1.208203 |

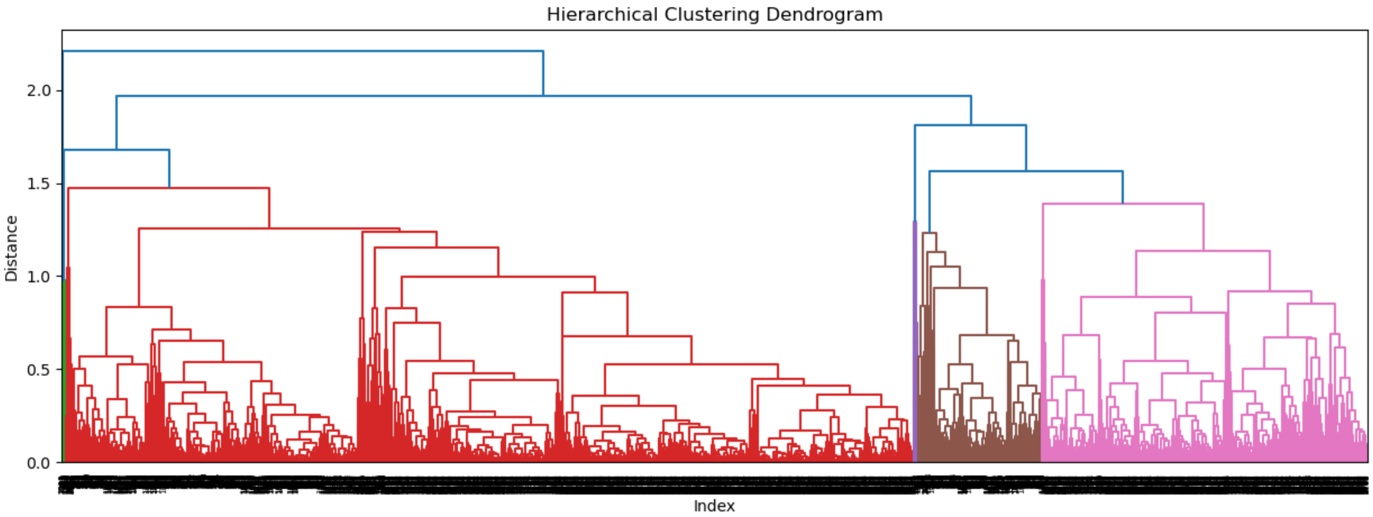
3999 rows × 6 columns

In [50]:

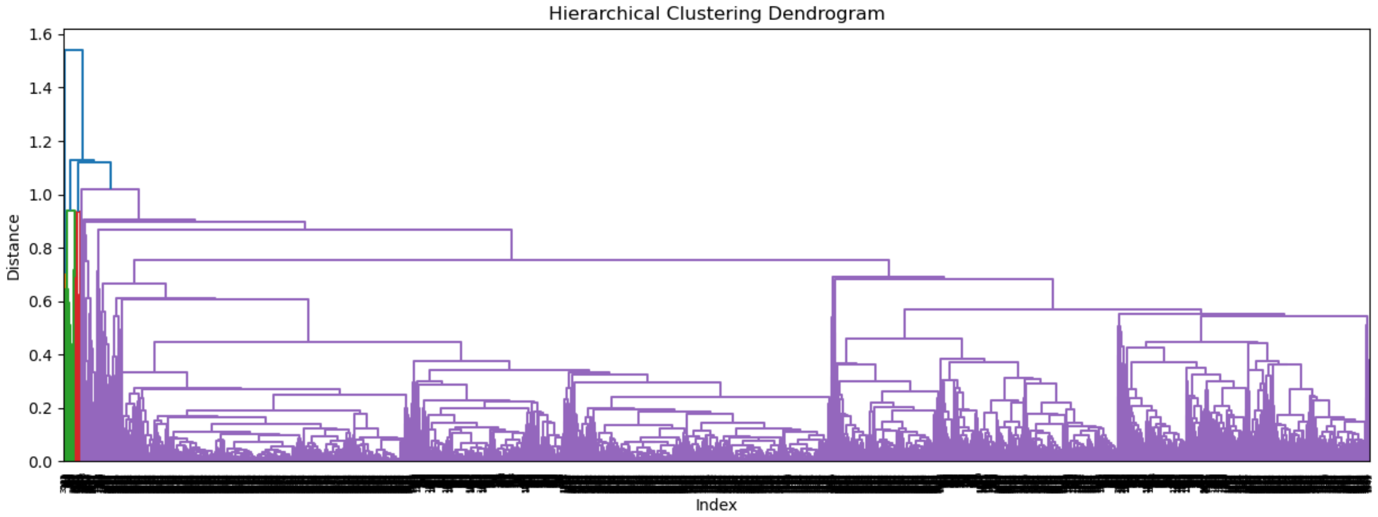
```python
from scipy.cluster.hierarchy import linkage
import scipy.cluster.hierarchy as sch # for creating dendrogram
p = np.array(df_norm) # converting into numpy array format
z = linkage(df_norm, method="single",metric="euclidean")
plt.figure(figsize=(15, 5))
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Index')
plt.ylabel('Distance')
sch.dendrogram(z,)
plt.show()
```



In [51]:
```
p = np.array(df_norm)
z = linkage(df_norm, method="complete",metric="euclidean")
plt.figure(figsize=(15, 5))
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Index')
plt.ylabel('Distance')
sch.dendrogram(z,)
plt.show()
```



In [52]:
```
p = np.array(df_norm)
z = linkage(df_norm, method="average",metric="euclidean")
plt.figure(figsize=(15, 5))
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Index')
plt.ylabel('Distance')
sch.dendrogram(z,)
plt.show()
```



In [53]:
```
from sklearn.cluster import AgglomerativeClustering
h_complete = AgglomerativeClustering(n_clusters=5, linkage='complete',affinity = "euclid

cluster_labels=pd.Series(h_complete.labels_)
cluster_labels
air['clust']=cluster_labels
air
```

Out[53]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 | |
| **1** | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 | |
| **2** | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 | |
| **3** | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 | |
| **4** | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3994** | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 200 | |
| **3995** | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | 0 | |
| **3996** | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | 0 | |
| **3997** | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 500 | |
| **3998** | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | |

3999 rows × 11 columns

In [54]:
```
air.iloc[:,1:].groupby(air.clust).mean()
```

Out[54]:

| | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Flight_trans |
|---|---|---|---|---|---|---|---|---|
| **clust** | | | | | | | | |
| **0** | 149.668605 | 3.667151 | 1.001453 | 1.001453 | 38389.199128 | 18.653343 | 477.122093 | 1.428 |
| **1** | 0.000000 | 3.642857 | 1.000000 | 4.142857 | 97132.785714 | 28.214286 | 378.571429 | 1.071 |
| **2** | 137.437211 | 1.201079 | 1.021572 | 1.001156 | 5388.251541 | 7.672188 | 414.703005 | 1.245 |
| **3** | 347.000000 | 2.500000 | 1.000000 | 1.000000 | 65634.250000 | 69.250000 | 19960.000000 | 49.250 |
| **4** | 1355.000000 | 1.222222 | 1.000000 | 1.000000 | 14267.222222 | 15.555556 | 2392.666667 | 9.111 |

Loading [MathJax]/extensions/Safe.js

In [55]:

```
data = air[(air.clust==0)]
data
```

Out[55]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **4** | 97752 | 0 | 4 | 1 | 1 | 43300 | 26 | 2077 | |
| **8** | 443003 | 0 | 3 | 2 | 1 | 1753 | 43 | 3850 | |
| **11** | 96522 | 0 | 5 | 1 | 1 | 61105 | 19 | 0 | |
| **15** | 28495 | 0 | 4 | 1 | 1 | 49442 | 15 | 0 | |
| **16** | 51890 | 0 | 4 | 1 | 1 | 48963 | 16 | 0 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **3979** | 57793 | 0 | 3 | 1 | 1 | 20959 | 15 | 1198 | |
| **3980** | 28867 | 0 | 3 | 1 | 1 | 19169 | 28 | 0 | |
| **3985** | 59017 | 0 | 4 | 1 | 1 | 34746 | 25 | 0 | |
| **3991** | 39142 | 0 | 3 | 1 | 1 | 14981 | 28 | 0 | |
| **3996** | 73597 | 0 | 3 | 1 | 1 | 25447 | 8 | 0 | |

1376 rows × 11 columns

In [56]:

```
data = air[(air.clust==1)]
data
```

Out[56]:

| | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **384** | 28193 | 0 | 5 | 1 | 4 | 103456 | 32 | 0 | |
| **850** | 25629 | 0 | 4 | 1 | 5 | 107308 | 36 | 600 | |
| **904** | 15835 | 0 | 2 | 1 | 5 | 112171 | 16 | 0 | |
| **1036** | 154664 | 0 | 4 | 1 | 3 | 73881 | 33 | 0 | |
| **1244** | 166465 | 0 | 4 | 1 | 4 | 98717 | 30 | 0 | |
| **1300** | 737514 | 0 | 3 | 1 | 4 | 43993 | 40 | 2150 | |
| **1878** | 128302 | 0 | 5 | 1 | 4 | 132263 | 32 | 400 | |
| **1917** | 82981 | 0 | 5 | 1 | 3 | 105325 | 44 | 1000 | |
| **1946** | 126873 | 0 | 4 | 1 | 5 | 95598 | 32 | 500 | |
| **2251** | 125679 | 0 | 3 | 1 | 4 | 68151 | 27 | 650 | |
| **2283** | 128456 | 0 | 5 | 1 | 3 | 70125 | 28 | 0 | |
| **2364** | 108081 | 0 | 5 | 1 | 5 | 217006 | 24 | 0 | |
| **2955** | 61474 | 0 | 1 | 1 | 4 | 47717 | 14 | 0 | |
| **3338** | 84148 | 0 | 1 | 1 | 5 | 84148 | 7 | 0 | |

In [57]:

```
data = air[(air.clust==2)]
data
```

Out[57]:

|  | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 28143 | 0 | 1 | 1 | 1 | 174 | 1 | 0 |  |
| **1** | 19244 | 0 | 1 | 1 | 1 | 215 | 2 | 0 |  |
| **2** | 41354 | 0 | 1 | 1 | 1 | 4123 | 4 | 0 |  |
| **3** | 14776 | 0 | 1 | 1 | 1 | 500 | 1 | 0 |  |
| **5** | 16420 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |  |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |  |
| **3993** | 3974 | 0 | 1 | 1 | 1 | 365 | 3 | 0 |  |
| **3994** | 18476 | 0 | 1 | 1 | 1 | 8525 | 4 | 200 |  |
| **3995** | 64385 | 0 | 1 | 1 | 1 | 981 | 5 | 0 |  |
| **3997** | 54899 | 0 | 1 | 1 | 1 | 500 | 1 | 500 |  |
| **3998** | 3016 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |  |

2596 rows × 11 columns

In [58]:

```
data=air[(air.clust==3)]
data
```

Out[58]:

|  | Balance | Qual_miles | cc1_miles | cc2_miles | cc3_miles | Bonus_miles | Bonus_trans | Flight_miles_12mo | Fli |
|---|---|---|---|---|---|---|---|---|---|
| **2015** | 53232 | 888 | 4 | 1 | 1 | 80696 | 65 | 22100 |  |
| **3235** | 287033 | 0 | 1 | 1 | 1 | 26161 | 58 | 12873 |  |
| **3583** | 160114 | 500 | 1 | 1 | 1 | 71954 | 86 | 30817 |  |
| **3594** | 27619 | 0 | 4 | 1 | 1 | 83726 | 68 | 14050 |  |

In [60]:

```
data = air[(air.clust==4)]
```

In [ ]:

```
```