**Lab 10**

**Amar Fazlic**

**1.3.1. Based on the previous result, which two features have (on average) the smallest values and the largest values?**

- max value: Balance

- min value: QualMiles and FlightTrans

**1.3.2. Why is it important to normalize the data before clustering?**

- to reduce data duplication

- to improve efficiency

- to improve accuraccy

**1.4.1. What are the mean and standard deviation of the features in the standardized dataset?**

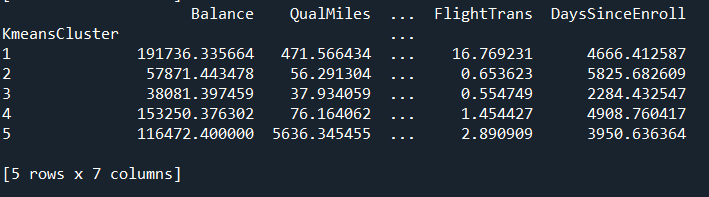
-the mean is the average value

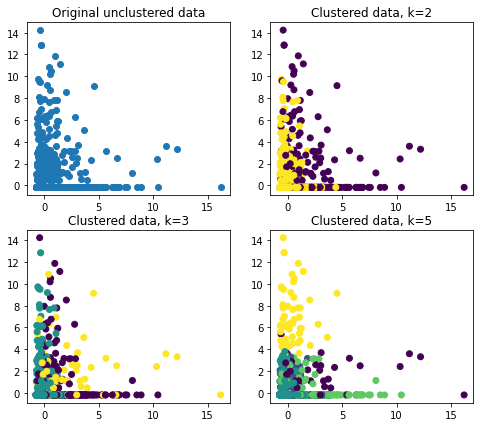
-the standard deviation describes the differences in the dataset, the higher the deviation the more differences there are in the dataset

**1.4.2. Based on the normalized dataset descriptive stats, which two features have (on average) the smallest values and the largest values?**

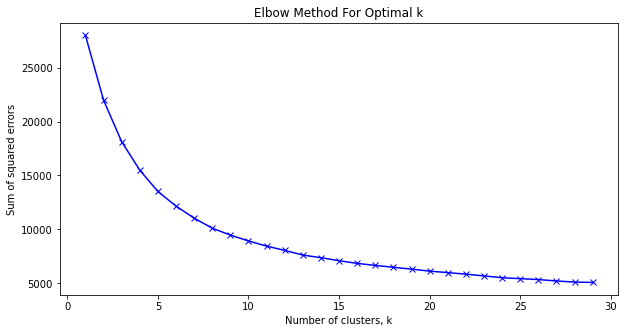
-max: 0

-min: 0





**2.1.1.1 Compared to the other clusters, Cluster 1 has the largest average values in which features (if any)? Based on this, how would you describe the Airline’s customers in Cluster 1?**

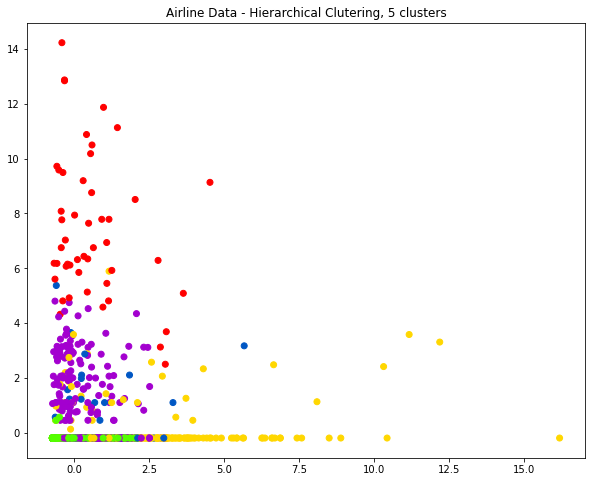
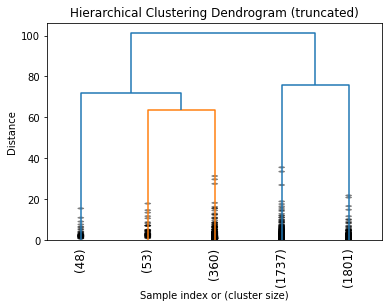
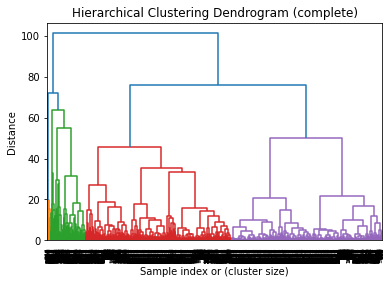


**2.1.2.1 Describe the relationship between the parameter k and the obtained sum of squared errors.**

- If we have more clusters the squared error is smaller, which makes sense since there is a lesser error rate, because the algorithm is more likely to put object in correct cluster

**2.1.2.2 Based on the results, indicate the optimal parameter k, that you consider appropriate to clustering the Airlines Cluster dataset.**

- From 5 to 12



2.2.1.1 Do you expect Cluster 1 of the Hierarchical clustering output to necessarily be similar to K-Means clustering? Compare the results obtained using K-Means and Hierarchical clustering.

2.2.1.2 Compared to the other hierarchical clusters, Cluster 1 has the largest average values in which features (if any)? Based on this, how would you describe the Airline’s customers in Cluster 1? Applied the previous analysis to the remaining four clusters and describe them.

2.2.1.3 Performed a new hierarchical clustering of the AirlineCluster Dataset using a different linkage method and compare with the clustering obtained using the ‘ward’ method. Explain what happens.