K-Means Clustering

* It uses the Logic of “Centroid Linkage”
* Here we have to initially supply the value of “K”
* Data will be provided based on your k-value here the data will be clustered into groups
* It’s is an Iterative process at Each and Every time when the data point is added or Removed from the cluster. The centroid or Center value of the cluster will be changed
* Clusters with Homogeneous (similar observations) should fall on same cluster
* Cluster with Non-Homogeneous (Dis-Similar observations) should fall on another cluster
* No Hierarchy
* No Linkages
* No Dendrograms are needed in K-Means Clustering

***How can we know exactly how many clusters are Required?***

***Here we Use a Method called Elbow method which shows how many clusters to form based on the data set that your provided***

***Note: Here we use K-Means Clustering for Large Data sets when we have more than 1000 records***

Example:

Data Description:

The file East-West-Airlines contains information on passengers who belong to an airline’s frequent flier program. For each passenger the data include information on their mileage history and on different ways they accrued or spent miles in the last year. The goal is to try to identify clusters of passengers that have similar characteristics for the purpose of targeting different segments for different types of mileage offers

1. ID --Unique ID
2. Balance--Number of miles eligible for award travel
3. Qual\_mile--Number of miles counted as qualifying for Topflight status
4. cc1\_miles -- Number of miles earned with freq. flyer credit card in the past 12 months:
5. cc2\_miles -- Number of miles earned with Rewards credit card in the past 12 months:
6. cc3\_miles -- Number of miles earned with Small Business credit card in the past 12 months:

1 = under 5,000

2 = 5,000 - 10,000

3 = 10,001 - 25,000

4 = 25,001 - 50,000

1. = over 50,000

7.Bonus miles--Number of miles earned from non-flight bonus transactions in the past 12 months

8.Bonus\_trans--Number of non-flight bonus transactions in the past 12 months

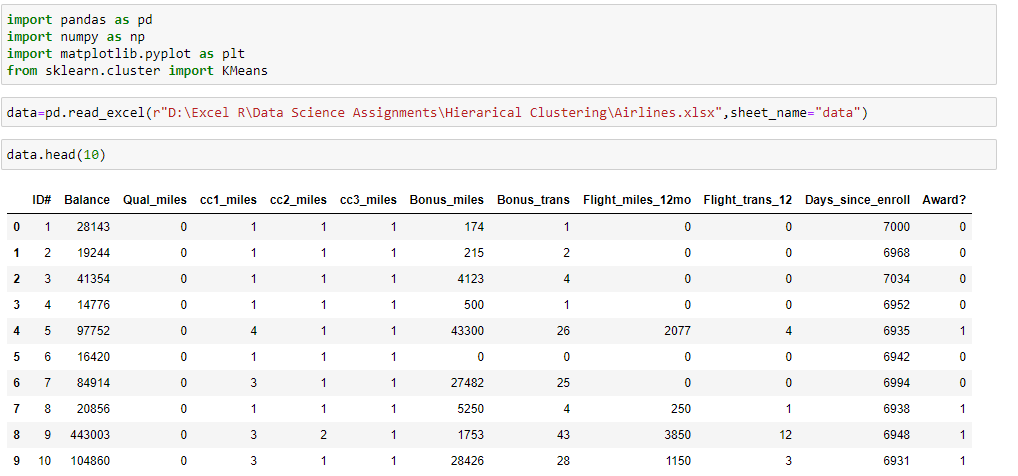
9.Flight\_miles\_12mo--Number of flight miles in the past 12 months

10.Flight\_trans\_12--Number of flight transactions in the past 12 months

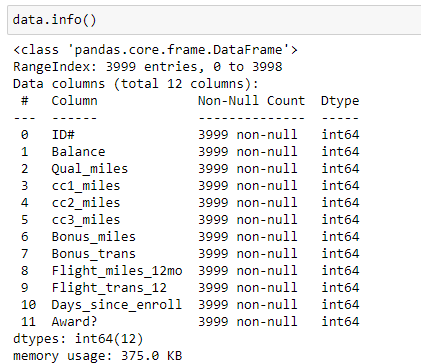
11.Days\_since\_enrolled--Number of days since enrolled in flier program

12.Award--whether that person had award flight (free flight) or not

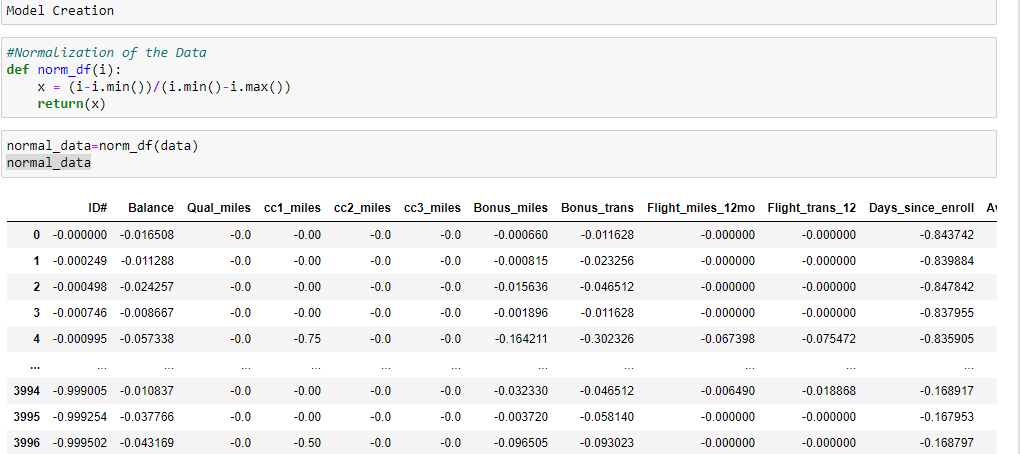
Import The data:



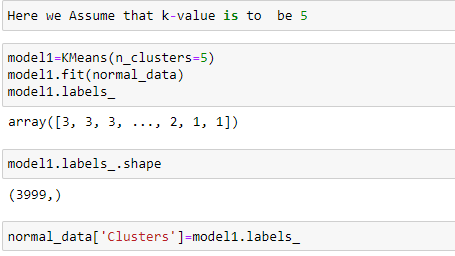
Pre-Processing



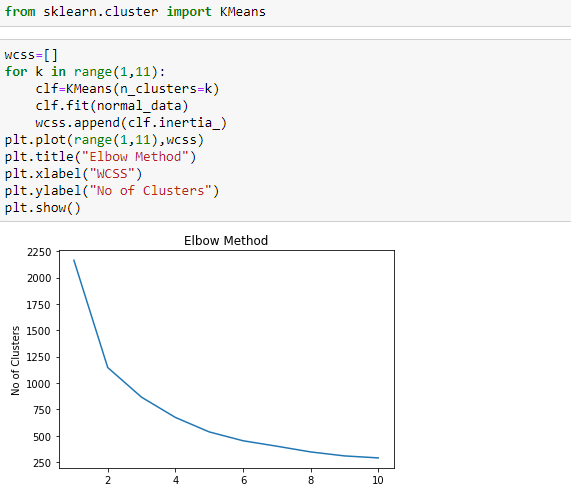
Normalization



Model Creation:



Elbow Method:



Clusters:

