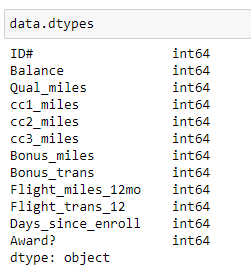
***Hierarchal Clustering***

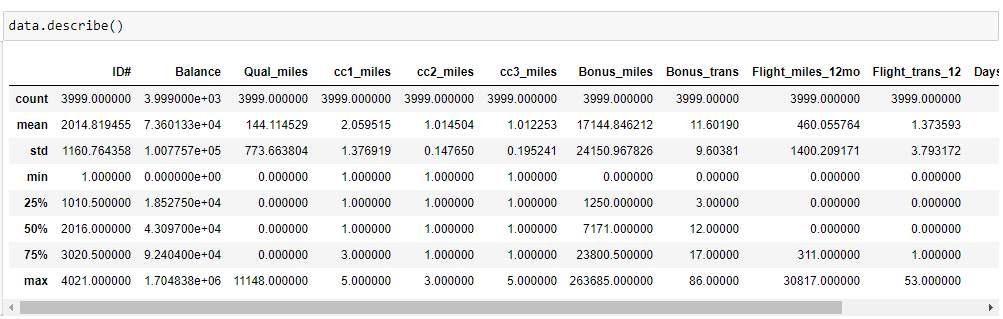
Example Problem:

* East West Airlines Data The file 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
* Data Description:
* ID --Unique ID
* Balance--Number of miles eligible for award travel
* Qual\_mile--Number of miles counted as qualifying for Topflight status
* cc1\_miles -- Number of miles earned with freq. flyer credit card in the past 12 months:
* cc2\_miles -- Number of miles earned with Rewards credit card in the past 12 months:
* 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
* 5 = over 50,000
* Bonus\_miles--Number of miles earned from non-flight bonus transactions in the past 12 months
* Bonus\_trans--Number of non-flight bonus transactions in the past 12 months
* Flight\_miles\_12mo--Number of flight miles in the past 12 months
* Flight\_trans\_12--Number of flight transactions in the past 12 months
* Days\_since\_enrolled--Number of days since enrolled in flier program
* Award--whether that person had award flight (free flight) or not

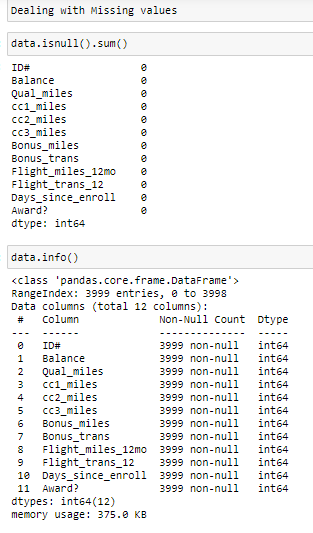
Data Col names and Description:



Data Describe (): description of data to get fill view please check the python eastwestairlines file

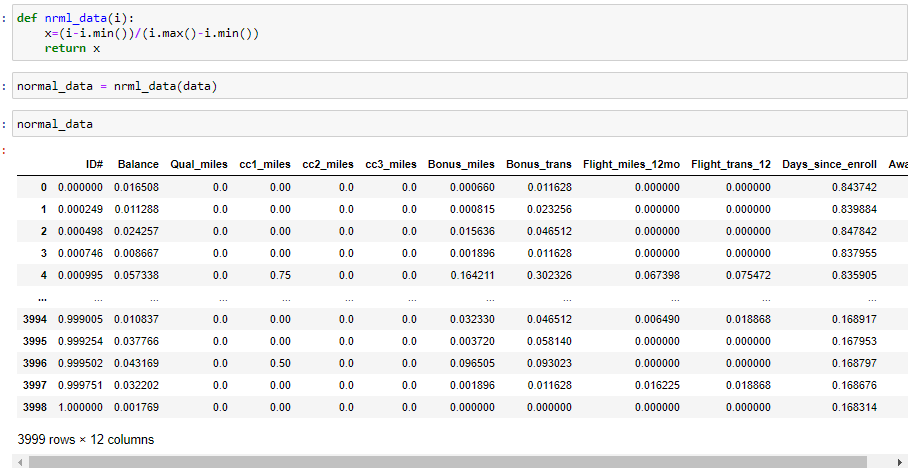


Missing Values in Data set:

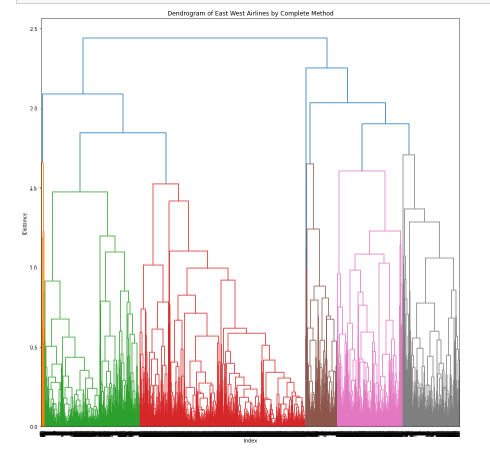


They are No Missing values and Nan Values in the data set

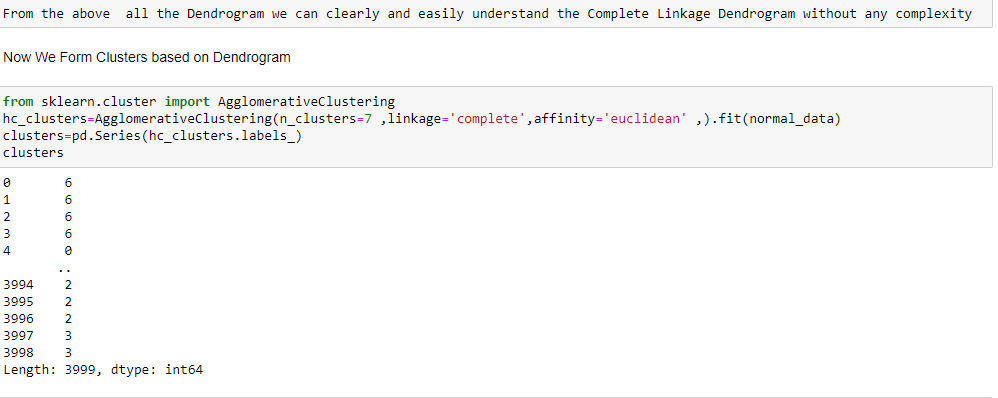
Normalization OF Data : while dealing with the distances we always need to Normalize the data or Standardize the data which results in between (0 to 1).



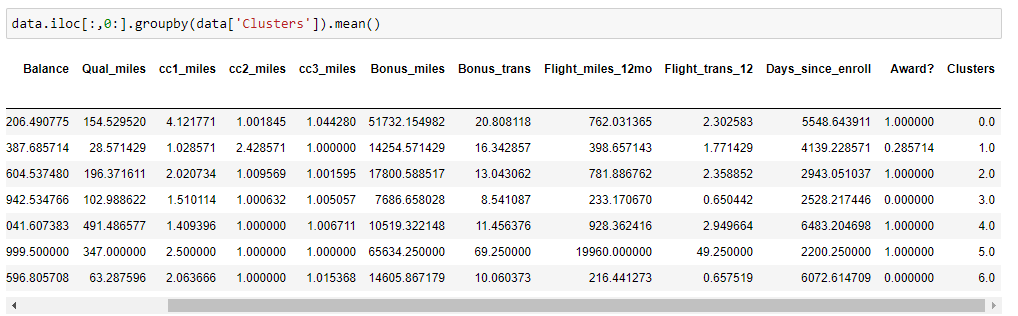
Dendrogram Creation: Creating Dendrogram with normalized data and with the complete linkage and metrics is Euclidean distance the below figure is the only Dendrogram which is easily understandable and with no complexity of data I try with the other Linkages like Single, Centroid, average etc. but those are very complicated when compared to Complete Dendrogram



Forming of Clusters: based on dendrogram we created the Clusters using Agglomerative clustering



Clusters with the mean of Data: which data is gone into which Cluster we will be seen into the below figure



To get Detail please go through the python file which I attached near this file