**Graduate Admission**

(Using Linear Regression model and -means clustering in Rapid Miner 9.2.001)



**Submitted by:**

**Akanksha Mohite**

**Prajakta Lahankar**

**Kshitija Jadhav**

**Manali Mane**

**Under the Guidance of**

**Prof. Raihan Siddique**

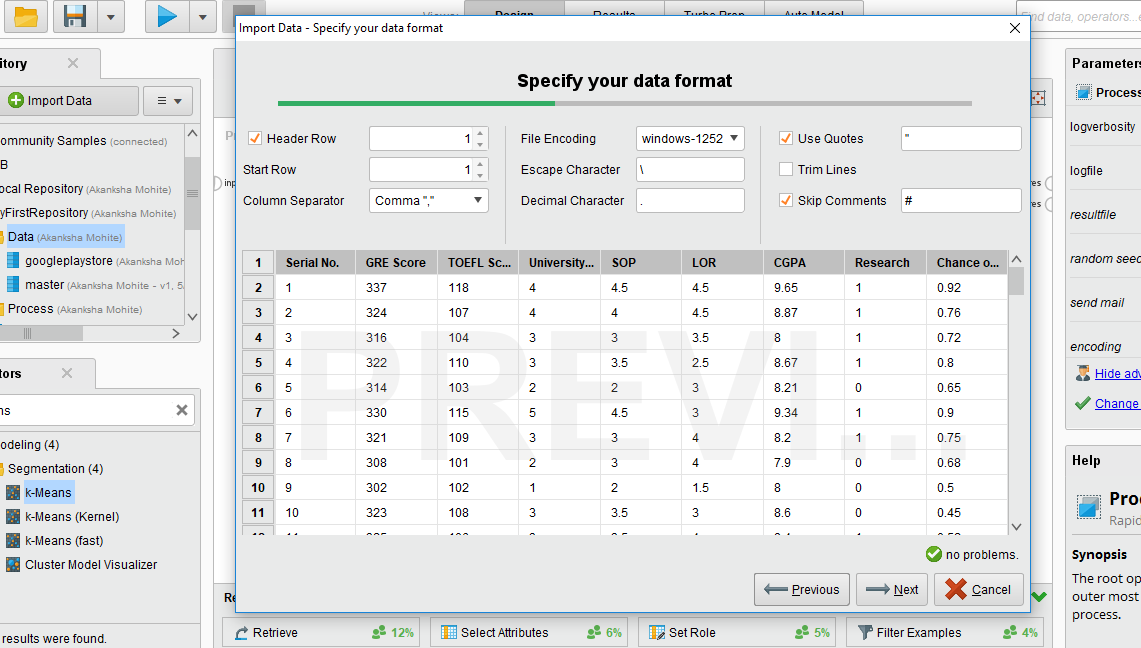
# **Purpose of the project:**

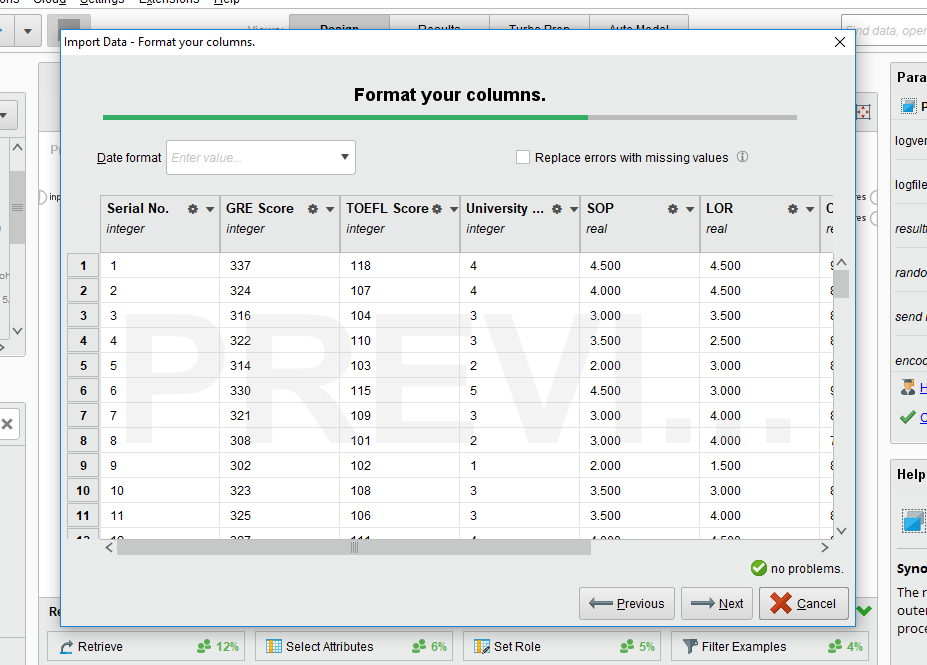
**Our project is based on Graduate Admission Students. We have gathered the data of all graduate students including international and Domestic as well. We have calculated the student’s chance of getting admitted based on various factors like scores, SOP, LOR and so many.**

**For our 400 students we know: GRE score, TOEFL Score, University ranking to which student have applied, SOP, LOR, CGPA and Research done by the student.**

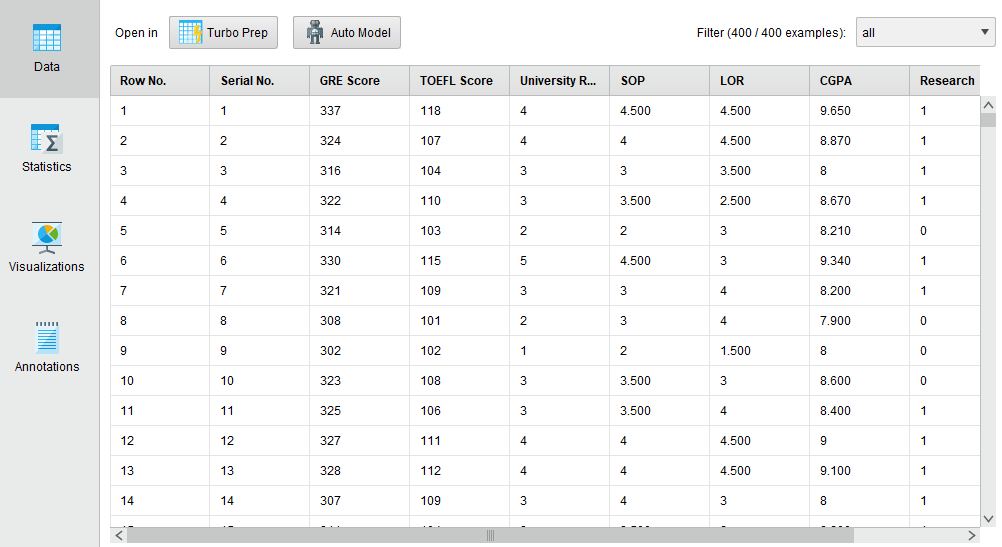
* Dataset we found online on https://www.kaggle.com/datasets as below:
* Dataset has total 400 examples/records

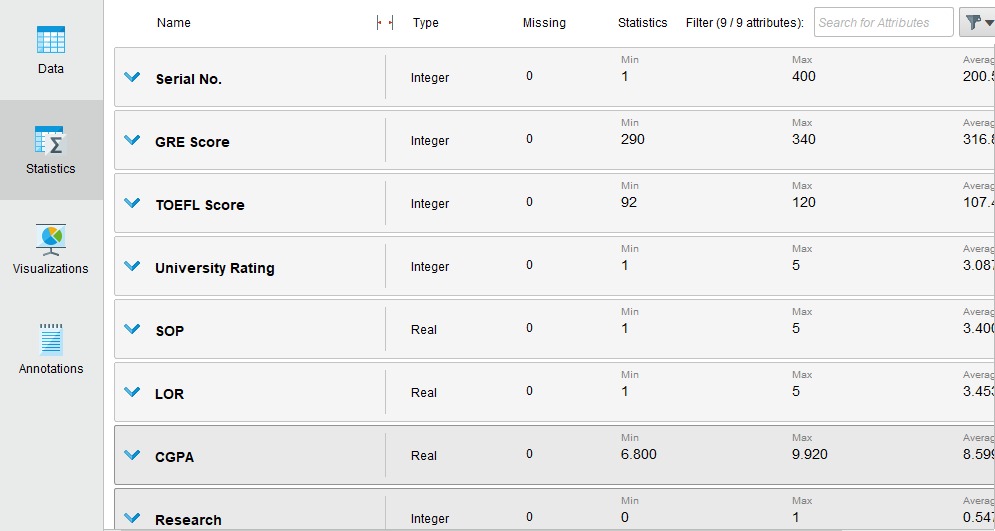
First, we imported the dataset.





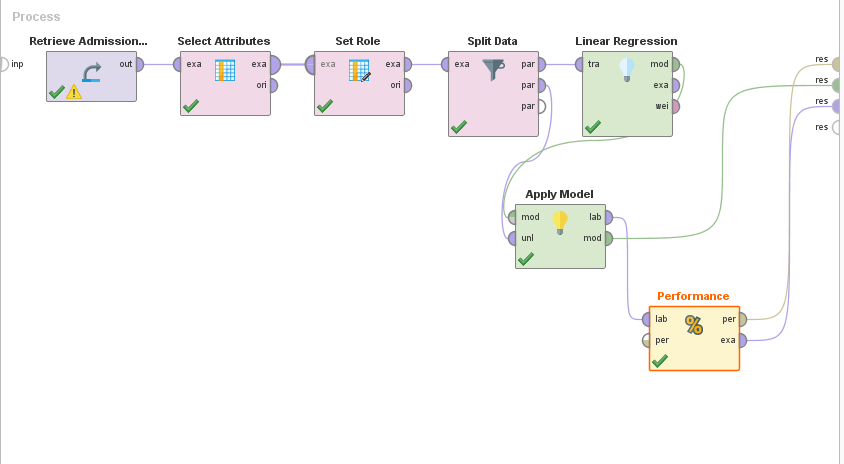
We did exploratory data analysis to familiarize with data and discover initial insights. We found out that select attribute operator from finalize dataset will be best fit for this analysis.



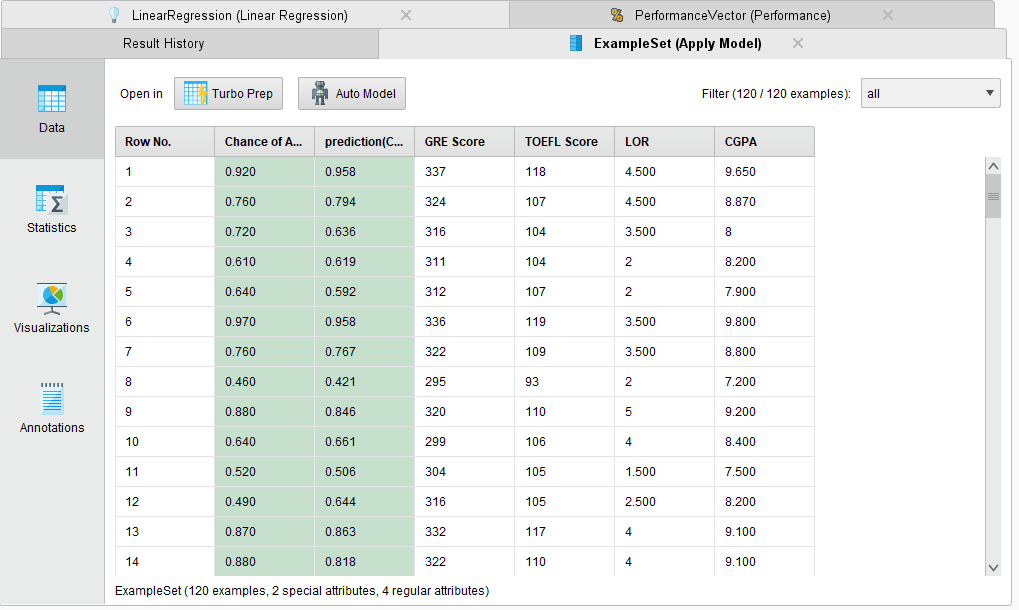


**Design:**

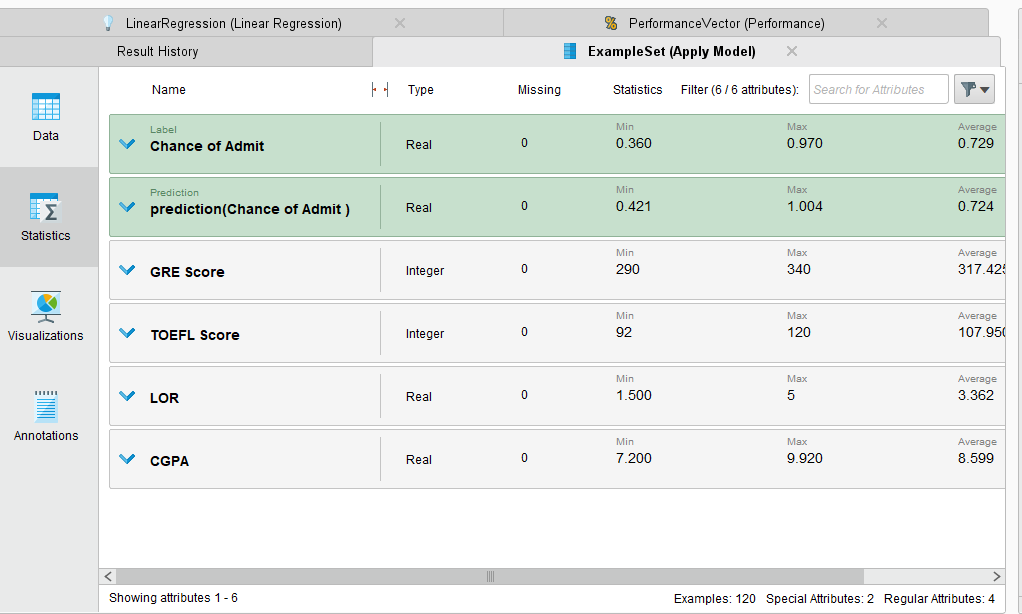
First, we retrieved the data of every student and after that we applied select attribute also set role operator. After this have split the data in 70% and 30% and later we applied Regression model, Performance vector and K-means clustering.



This is the result of Apply Model. In this we got prediction value of chance of getting admitted. How many chances of admission of every student.

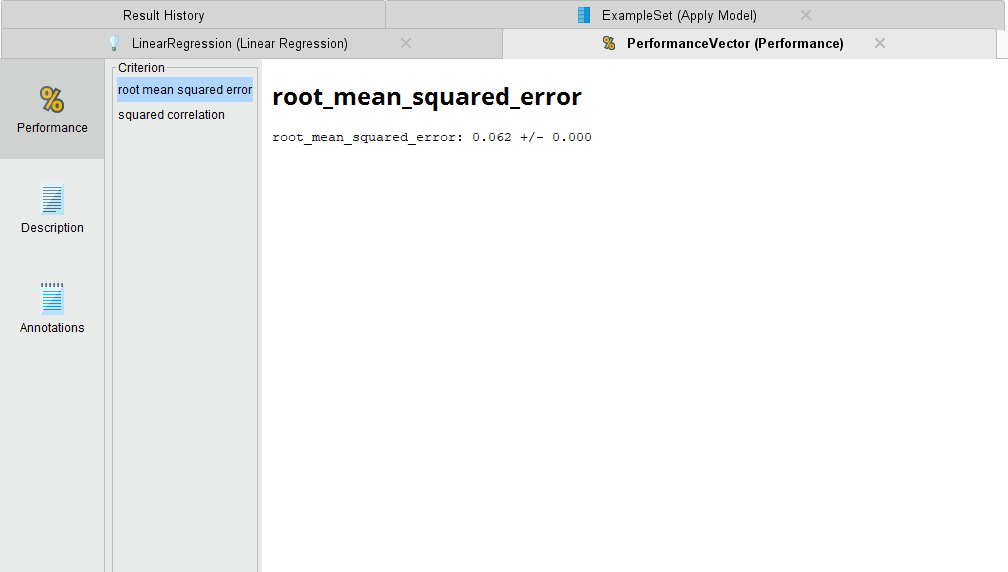


Statistics approach of Apply Model:

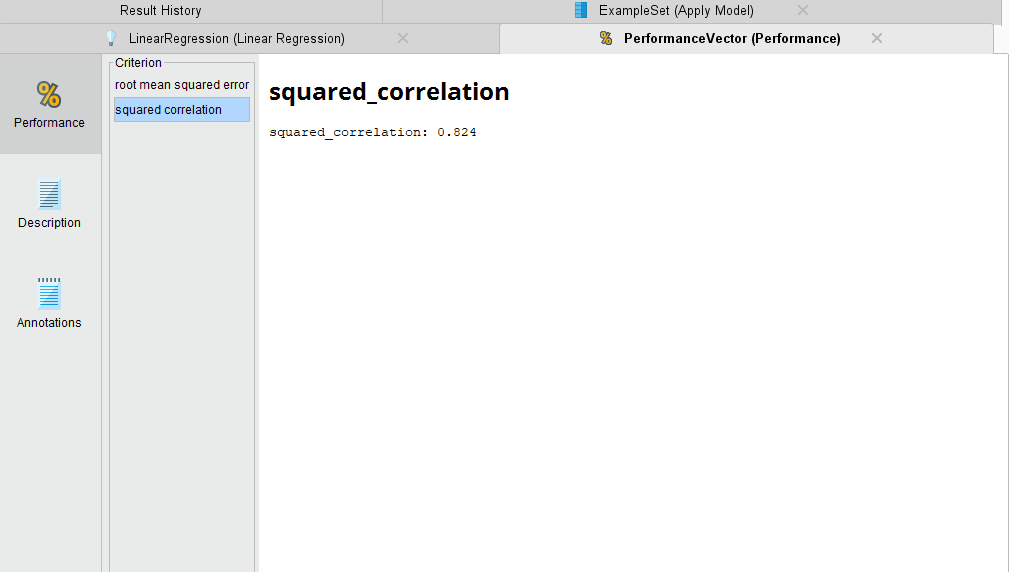


Performance approach of Performance Vector. We cannot determine the accuracy since the data is numerical in nature. There are 2 types of performance vector approach:

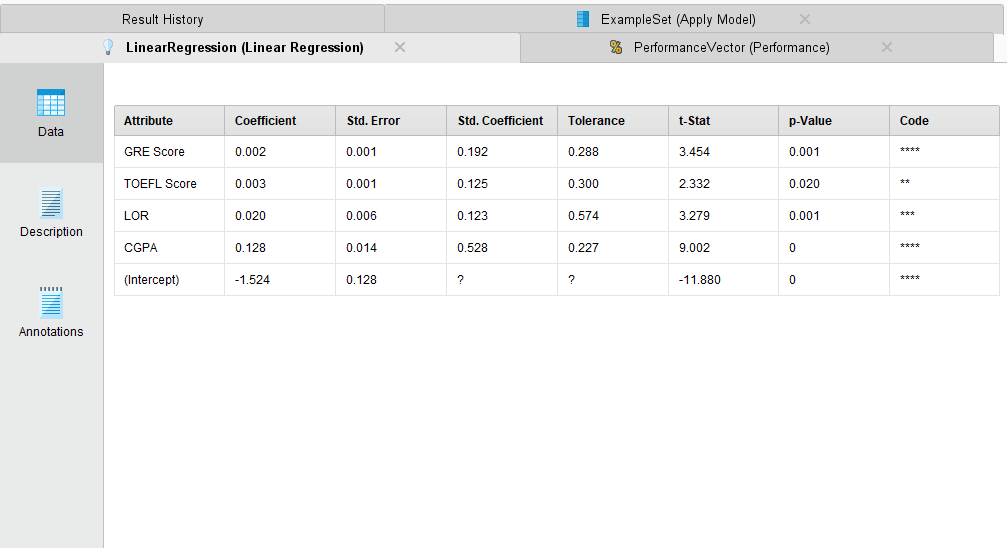
1. Root mean squared error



1. Squared correlation



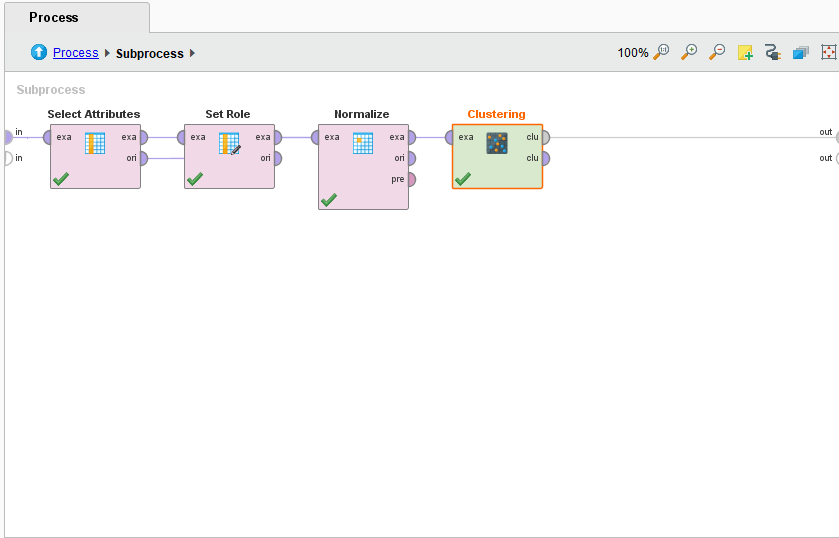
Then we have applied Linear Regression on dataset.



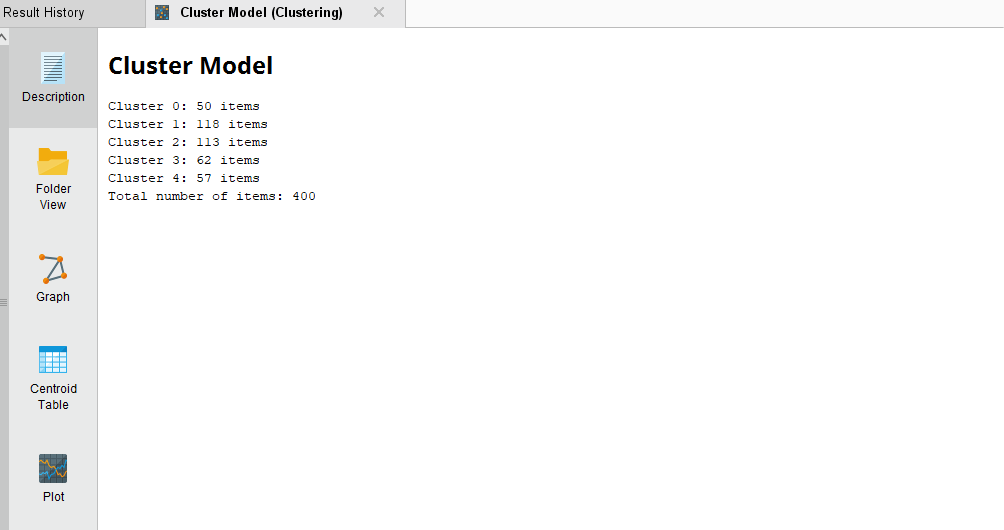
Equation of Linear Regression:



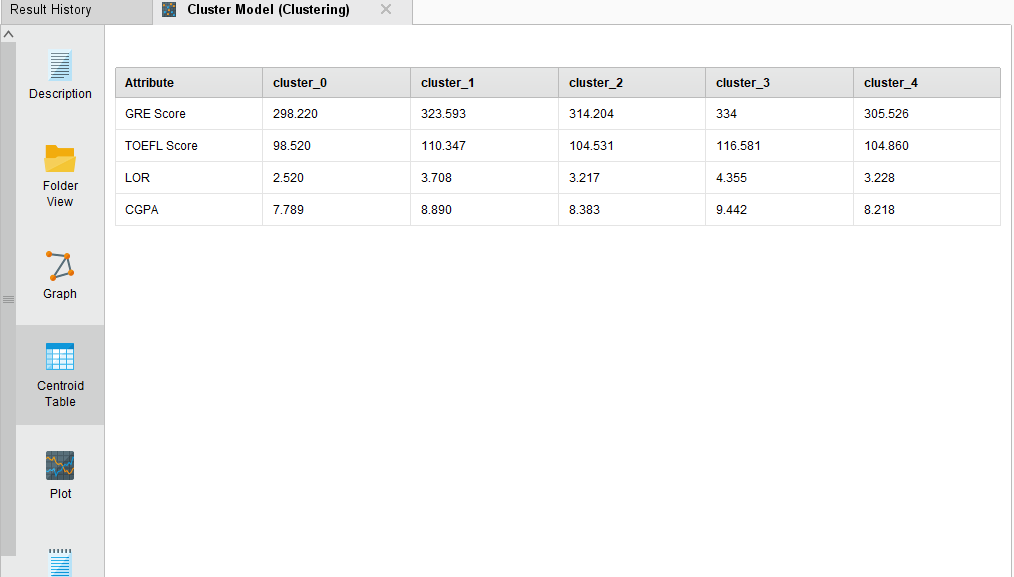
After above process we used another modelling technique which is clustering. Here we have used K-means clustering where k=5



Here, we can see that there are 5 clusters formed and which cluster consist of how many items.



This is the Centroid table of clustering. Clusters are formed based on the GRE score, TOEFL score, LOR and CGPA.



Plot of Clustering:

