# CHAPTER 4

# EXPERIMENTAL RESULT AND ANALYSIS

## 4.1 Result Obtained after applying k-prototype to a numeric and categorical data set

We have taken a small database of some companies located at different location. The attributes contain the company name, there revenue, Type of company and Location. There are basically 3 categorical attribute and 1 numeric attribute of this dataset. We have taken 100 objects which need to be clustered in 4 of these cluster.

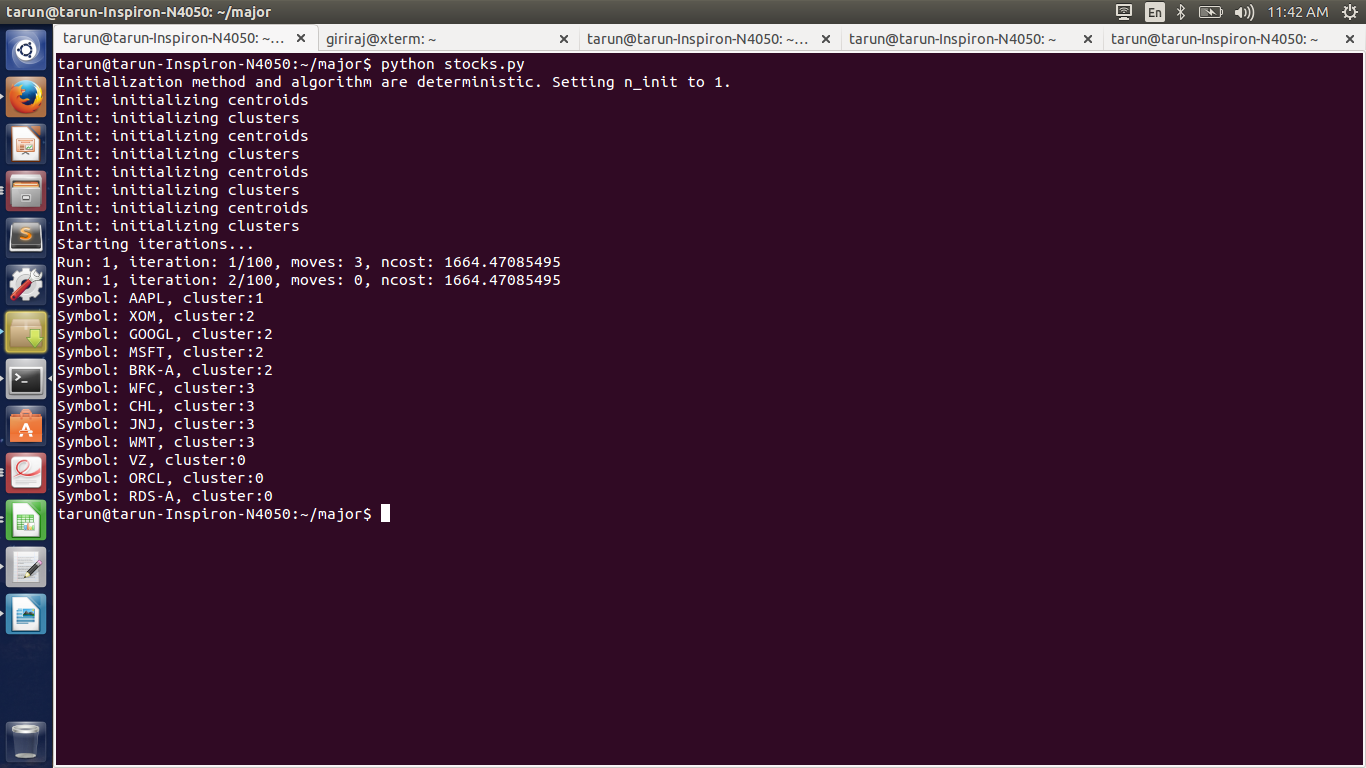
Here is the small portion of the database we have taken:

**Table 4.1: Dataset taken for analyzing k-prototype algorithm**

|  |  |  |  |
| --- | --- | --- | --- |
| Company | Market Price | Type of Company | Location |

|  |  |  |  |
| --- | --- | --- | --- |
| AAPL | 738.5 | Tech | USA |
| XOM | 369.5 | Nrg | USA |
| GOOGL | 368.2 | tech | USA |
| MSFT | 346.7 | Tech | USA |
| BRK-A | 343.5 | Fin | USA |
| WFC | 282.4 | Fin | USA |
| CHL | 282.2 | Tel | CN |
| JNJ | 279.7 | Cons | USA |
| WMT | 257.2 | Cons | USA |
| VZ | 205.2 | Tel | USA |
| ORCL | 192.1 | Tech | USA |
| RDS-A | 195.7 | Nrg | NL |

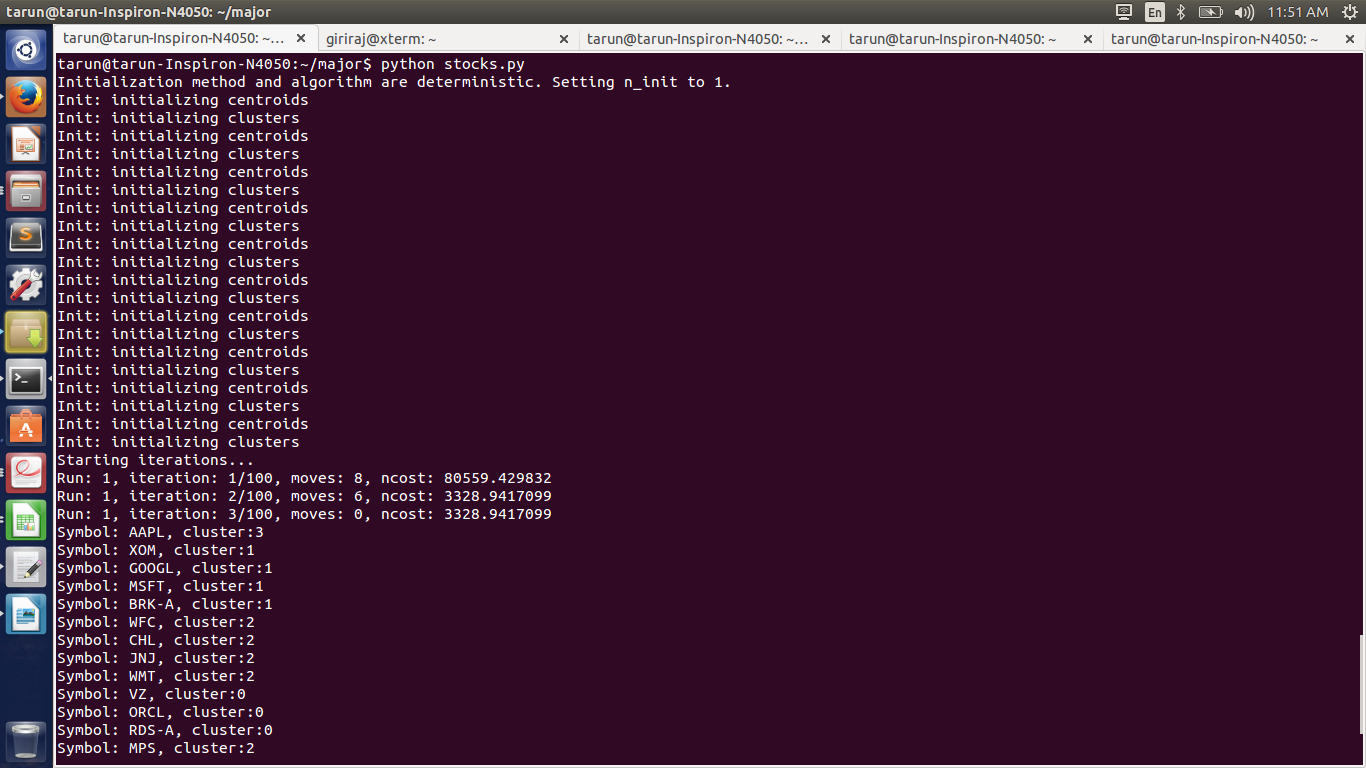
**4.2 Output on running k-prototype on the given dataset**



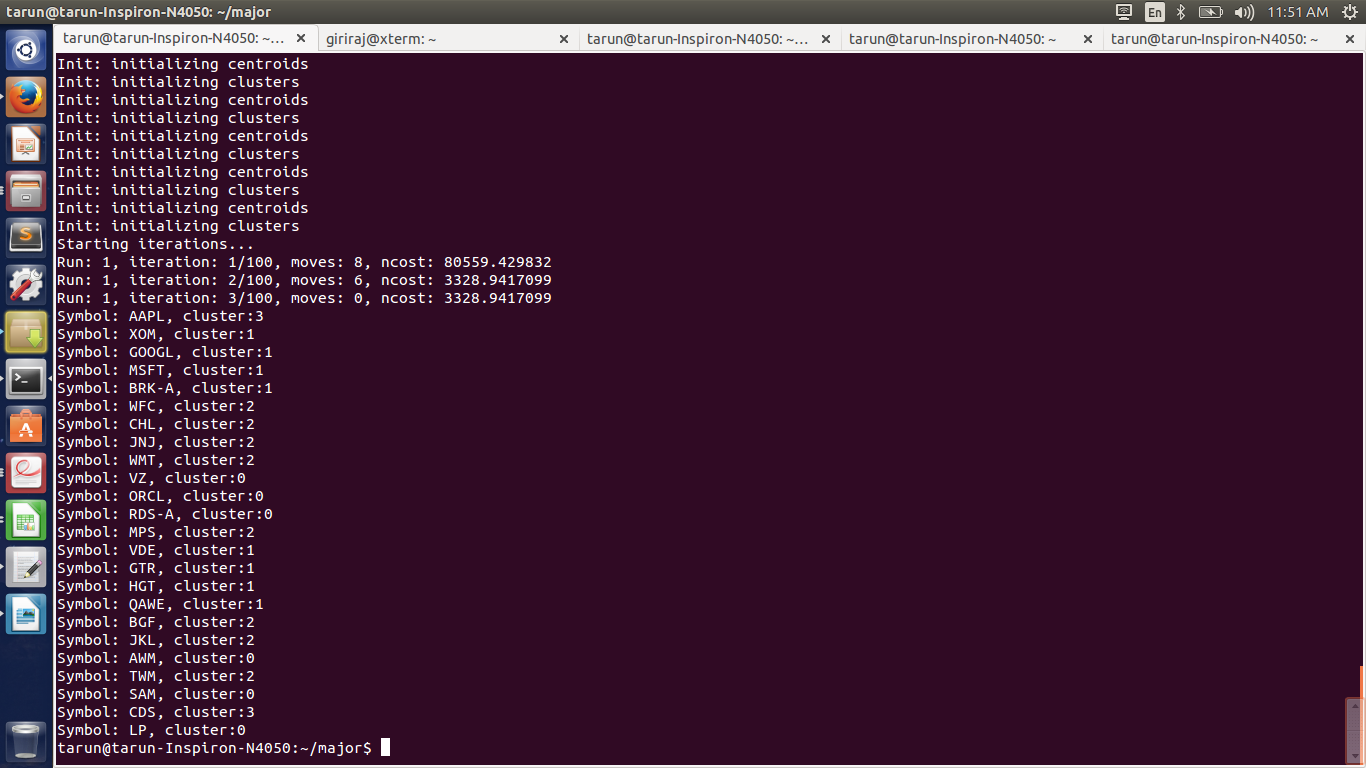
**Figure 4.2: Snapshot of the output of k-prototype for a dataset of 11 objects having 3 categorical and 1 numerical attribute**

Here we have clustered the objects into 4 cluster. The result is shown in these three snapshot.

In the figure 4.1 11 objects are clustered into 3 cluster whereas in figure 4.2 and figure 4.3 22 objects are clustered into 4 cluster.



**Figure 4.2: Snapshot of the output of k-prototype for a dataset of 22 objects having 3 categorical and 1 numerical attribute**



**Figure 4.2: Snapshot of the output of k-prototype for a dataset of 22 objects having 3 categorical and 1 numerical attribute**

**4.3 Analysis**

From the above result we have analyzed that clustering of points is dependent on both numeric and categorical values. The points, which are spatially close and have the same categorical values, are clustered together. If a point has a categorical value A but is close the majority of points having categorical value B and far away from the majority points having categorical value A, the point is clustered together with the points, which have categorical value B in majority. In the example we have taken, GOOGL and AAPL have two categorical values same so they should belong to same group but the distance between these two points is large so here XOM and GOOGL which has only one categorical attribute matching is taken into one cluster.

Other case in which categorical attribute is taken into higher consideration is object HGT and GTR. The distance between QAE and HGT is smaller as compared to HGT and GTR but none of their categorical attribute matches so HGT and GTR are placed in one cluster as per our example.