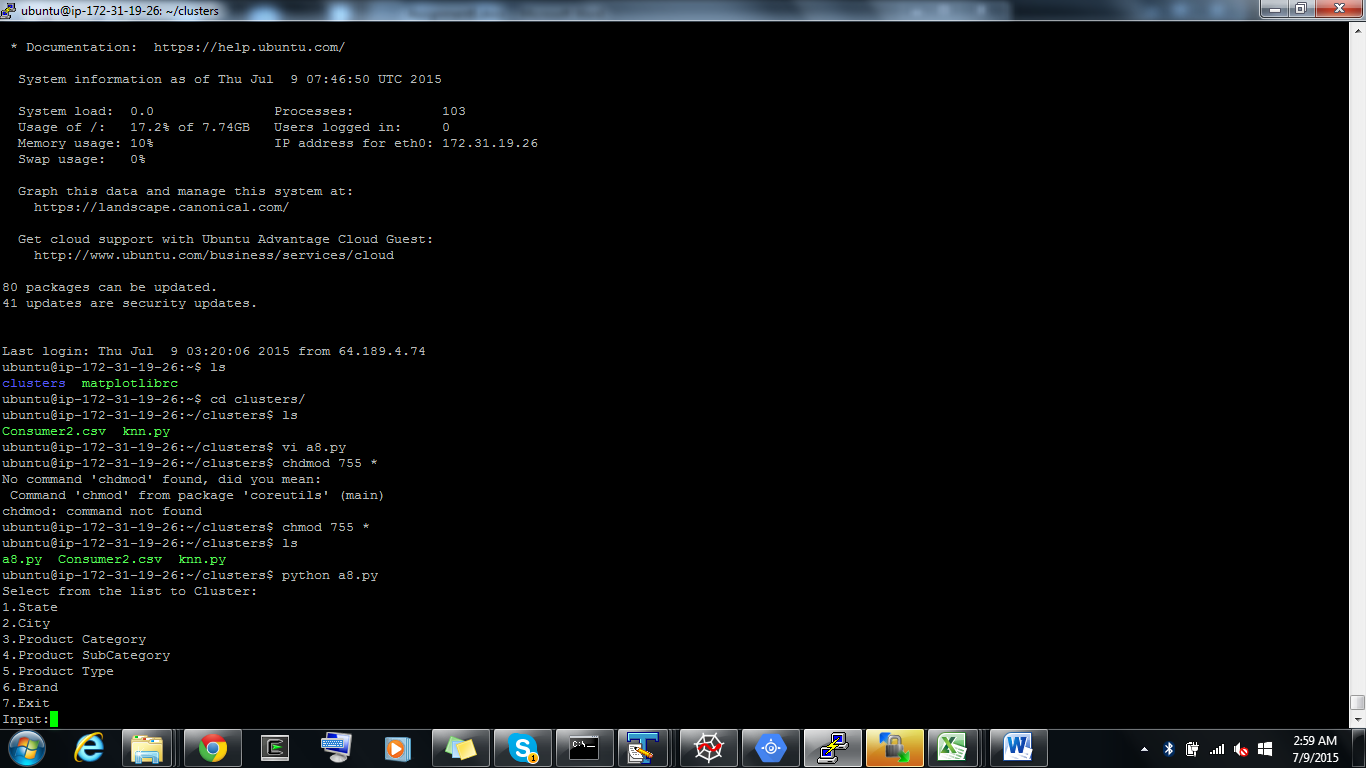
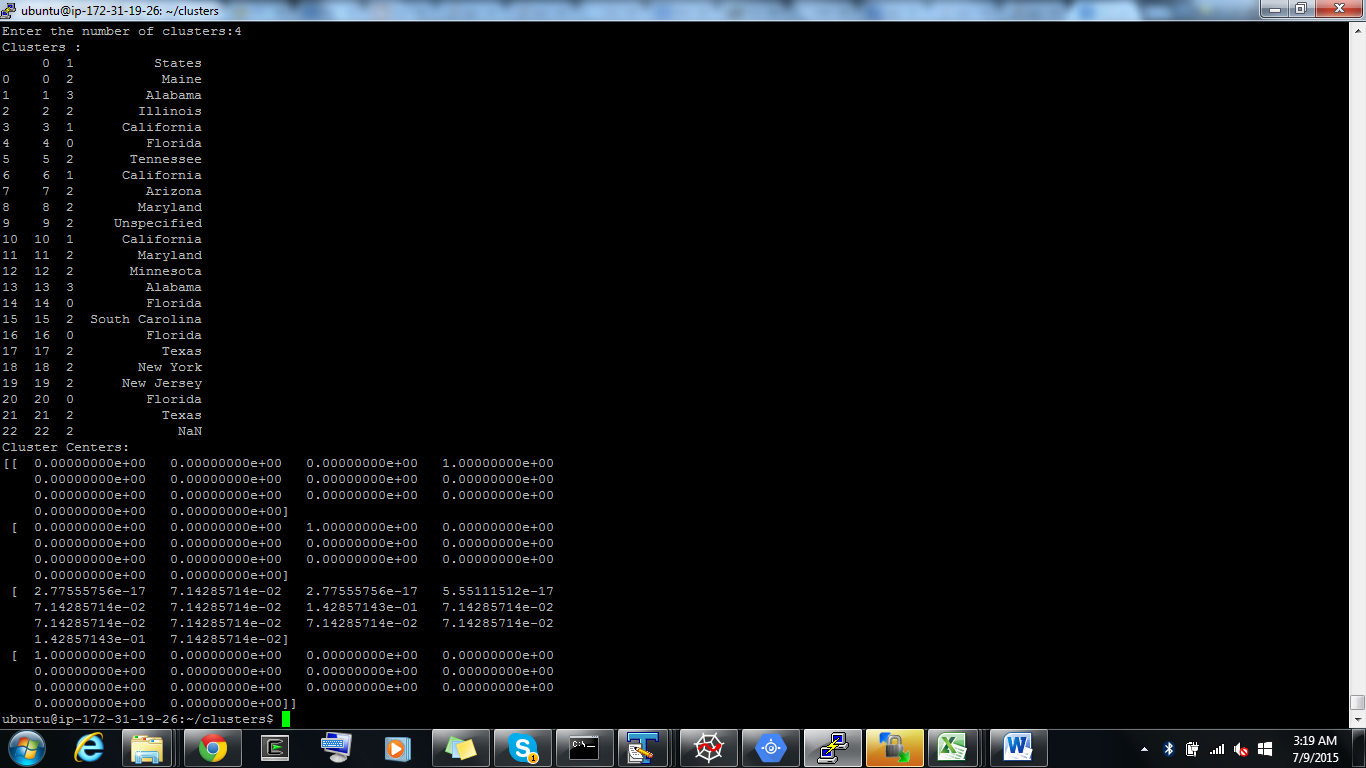
# Assignment 8

Name:Harish Kamuju

UTA ID: 1001120930

1. Implemented using amazon ec2 engine and python 2.7
2. Used pandas for dataframes and sklearn for kmeans clustering
3. Uploaded Consumer\_complaints.csv file from the gov data to the local storage in ec2.
4. Provides an interface to allow the user to input the number of clusters and selected attributes from the file.
5. As per the options the clustering is performed on those attributes and the number of complaints received based on the attribute are clustered .
6. Centroids and the distance between them and also the clusters are displayed on the screen.





References:

<http://scikit-learn.org/stable/modules/clustering.html>

<https://github.com/scikit-learn/scikit-learn>

<http://stackoverflow.com/>