

Clustering

Note: Use jupyter notebook and explain each and every step you do. If you fail to explain you would not get any marks.

- You can use sklearn, numpy pandas, matplotlib
- Apply Kmean Clustering using sklearn
- Steps include
 - Read the data using pandas dataframe
 - Do feature engineering and plot the data using PCA
 - Remove those features which are not contributing
 - Through Elbow plot determine the value of K using kmean
 - Apply kmean clustering
 - plot cluster using t-SNE plot

Bonus part:

Do detail Exploratory data analysis to get good features.

Apply Hierarchical Clustering

Help:

python, pandas numpy:

<http://cs231n.github.io/python-numpy-tutorial/>

<https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/>
Sklearn

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html#sphx-glr-auto-examples-cluster-plot-cluster-iris-py