Group ALPHA

Northeastern University Assignment 03

ALY6040 Data Mining Applications

https://northeastern.blackboard.com

Handed out: 07/26/2019 Due by 11:59 PM, 07/30/2019

Group ALPHA **Dataset:** Crimes in Boston

Maryam Heidari, Thi Thai Ngan Le, Findlee Odili-Obi

Final Project Proposal.

About the dataset.

Crime rate has long been a concerning issue. Our team will dive into the crime situation in Boston using "Crimes in Boston" dataset which retrieved from https://www.kaggle.com/AnalyzeBoston/crimes-in-boston/. The data is provided by Boston Police Department including records of crime in the Boston between June 14, 2015, and September 3, 2018. There are 17 variables with 319,073 rows in the dataset which provides information about location, date, time, and type of crimes.

Data Cleaning

For this project, we will be performing **EDA** which involves cleaning the data, choosing the important variables, showing the correlation between the variables, and detecting outliers. Data cleaning is essential before doing the analysis part to make sure the data is clean and well explored to avoid bias in drawing conclusions.

Methods and Progress.

We will be making use of the **Clustering** and **Time Series** methods;

- The Time Series would be applied to show the tendency of crimes in different period of time, then based on that observation, we can predict the number of crimes to expect in the future. This would answer the question of **when was the time happened the most crimes? Does the frequency change over time? What is the trend for the next coming time?**

- We will use the Clustering method to partition the whole data set into clusters which have the same patterns. With this, we can analyze each cluster to figure out their features - the crime categories would be divided into 3 or 4 groups, then we would find out further details from each of the groups based on the type of crime, location, time or season, etc.