Team 7: Project Proposal

Team Members: Neil Balazon, Mei Chang, Andrew Ivanov, Sam Lee, Skye Morgan
Dataset: http://data.torontopolice.on.ca/datasets/98f7dde610b54b9081dfca80be453ac9_0
Source of Dataset: Toronto Police Service Public Safety Data Portal
Research Question: Where are the safest Neighbourhoods in Toronto?
Methods of Analysis: Classification, Clustering, Regression

Project Proposal

Our project's main objective is to study Toronto's crimes with the goal of finding the safest neighborhoods. Throughout our analysis, we will take multiple types of crimes into account, including robbery, assault, battery and theft. Furthermore, we will study each type of crime based on the specific hour, day, week or month of occurrence to see if there are any surprising correlations.

We obtained our dataset from Toronto's public safety data portal website. This particular dataset consists of 167,525 observations from the year 2014 to 2018 and includes a plethora of different attributes such as: Division(plices department), Neighborhood(different places who live in Toronto), occurancedate(which date the crimes occur), occurancetimee(when date the crimes occur), occurancemonth(which month the crimes occur), offence(what type of crime people didi), premisetype(what is the major type of crime people have been done), reporteddate(which date police report the crime). A large part of our study will be focused on time-based and date-based variables such os occurancetime and occurancedate, and both of these variables would help us in finding the correlation between the time/date and the occurrence of each type of crime. Additionally, the time that spans between the occurancedate and the reporteddate could prove enlightening as well. We believe that there are more than enough attributes and variables in this dataset to use in our research and analysis in order to draw several conclusive implications.

Throughout our project, we will explore several research questions:Where are the safest neighborhoods? Where in Toronto are crimes most likely to occur?What types of crimes are most frequent at each hour of the day? When identifying the safest neighborhood in Toronto is there any correlation between the type of crime and the occurrence of the crime?

Our team will use regression, clustering and classification methods to group neighborhoods with a high number of assaults together. Using K-Means clustering we will partition groups of data points in small clusters. As we measure the number of MCIs, the neighborhoods with high numbers of indicators will be grouped together.

The research questions pose several practical implications. Crime statistics can be useful in helping criminal justice professionals anticipate increased risk of crime in specific neighborhoods. By taking a closer look at the statistics for certain neighbourhoods as well as the hours of the day crimes are committed, we can gain a deeper understanding of crime, why it occurs, and when it is most likely to occur. From the data we can find out what type of crime is most common in which neighborhoods, which can then allow criminal justice professionals to look into why certain types of crime are occurring in some neighborhoods.