Sophia Mir

DATS 6103

Data Mining

06/22/2022

Project Proposal

For the data mining project, I will be using a crime dataset consisting of primary data collected by DC’s Metropolitan Police Department (source: <https://crimecards.dc.gov>). It consists of over 55,000 observations and the incident data is dated from January 2020 to May 2022. Some of the questions I will address during the EDA are:

1. Is there a relationship between time of day and incidences of crime?
2. Is there a relationship between time of year and incidences of crime?
3. Is there a relationship between time of day and type of crime?
4. Is there a relationship between geographical area and type of offense?
5. Which crimes happen most frequently in each ward?
6. Is there a relationship between property crime and violent crime?
7. What types of crimes happen during the day/evening/midnight?
8. Is there a relationship between the offense type and the method?

For the modeling aspect, I will be creating a classification model to choose between the two categories of crime – property crime and violent crime.