

Capstone Project-Battle of Neighbourhoods with Crime data in NewYork

1. Problem statement:

Crimes now days are increasing day by day and with different level of intensity and versatility. The result is great loss to society in terms of monetary loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. To overcome this problem the computing era can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. The crime rate prediction strategies can be applied on historical data available in the police records by examining the data at various angles like reason of crime, frequency of similar kind of crimes at specific location with other parameters to prepare model the crime prediction. It is the major challenge to understand the versatile data available with us then model it to predict the future incidence with acceptable accuracy and further to reduce the crime rate. In this project we are going to find in which particular locations or neighbourhoods are do we need to focus and take certain actions to reduce crimes. Trying to locate patterns among neighbourhoods related to crimes. Present we consider New York crime data set for analysis along with foursquare locations data.

2. Analytical Approach:

Unsupervised learning to find patterns of crimes in New York city and its neighbourhood places : K-means clustering will be used to compare crimes

3. Data Collection

1. New York crime data from <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243> (<https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243>) This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) for all complete quarters from 2016 to 2019. This data set consists data about neighbourhood name, crime, date ,place and location details. From this data we can able to extract the no. of crimes in each neighbourhood and stats about crimes etc..
2. Foursquare api for extracting neighbourhoods and venues details. The collected data consists of NYC venues with compact metadata like id, name, location, checkins count (total checkins ever done in that venue), users count (total users who have ever checked in), tip count (total number of tips written by users), associated categories, menu, etc. The Foursquare categories span a broad ontology, with the following 10 categories on the first level: (1) Arts and Entertainment, (2) College and University, (3) Event, (4) Food, (5) Nightlife Spot, (6) Outdoors and Recreation, (7) Professional and Other Places, (8) Residence, (9) Shop and Service, (10) Travel and Transport.