

Exploring the Preferences of Robbery Crimes

1. Introduction

1.1 Background

Robbery is one of the most common violent crimes in New York City. Statistics show, from 2000 to 2019, there are an average of about 20,000 robberies reported in NYC each year, ranging from 12,913 in 2018 to 32,562 in 2000. Robbery makes up for at least 13% of all violent crimes reported in NYC. However, the frequency robbery occurs in different areas of the city varies. It would be valuable for both business owners as well as the local police department to know in which areas there is a higher risk of robbery.

1.2 Problem

The risk of robbery in a specific area of a city is dependent on various factors. One of them is the distribution of different categories of business and stores in that area. Based on this assumption, this project aims to find a correlation between the robbery risk and distribution of business categories and to apply it to predict the robbery risks in the future, even for areas where the data of crime history is not available.

2. Data Acquisition and Cleaning

2.1 Data Sources

All reported crimes in 2018 are available through the Open Data NYC program by NYC government agencies. The Raw data can be downloaded from the following url:

<https://data.cityofnewyork.us/api/views/8h9b-rp9u/rows.csv?accessType=DOWNLOAD>

The numbers of different categories of venues in a specific area is obtained by using the explore endpoint of Foursquare API.

2.2 Data Cleaning and Usage

For the crime data, the downloaded raw data contains all sorts of crimes that have been reported and recorded. Given that this project is focused on robbery, the first step is to exclude all the records that are not a robbery crime from the dataset. Secondly, the dataset contains crime records starting from January 1, 2006 to December 31, 2018. Since there could have been a huge change to a neighborhood during the 12 years of timespan, only the records in 2018 were used for the project for a better accuracy. Finally, since the crime data is utilized to locate the areas with high density of robbery crimes, only the latitude and longitude of each record are kept while other features are dropped from the dataset.

Venue information for different neighborhoods can be obtained through Foursquare API. The explore endpoint is used for the purpose of this project. Results from each API call were summarized together into a table showing the number of different categories of venues in different neighborhoods.