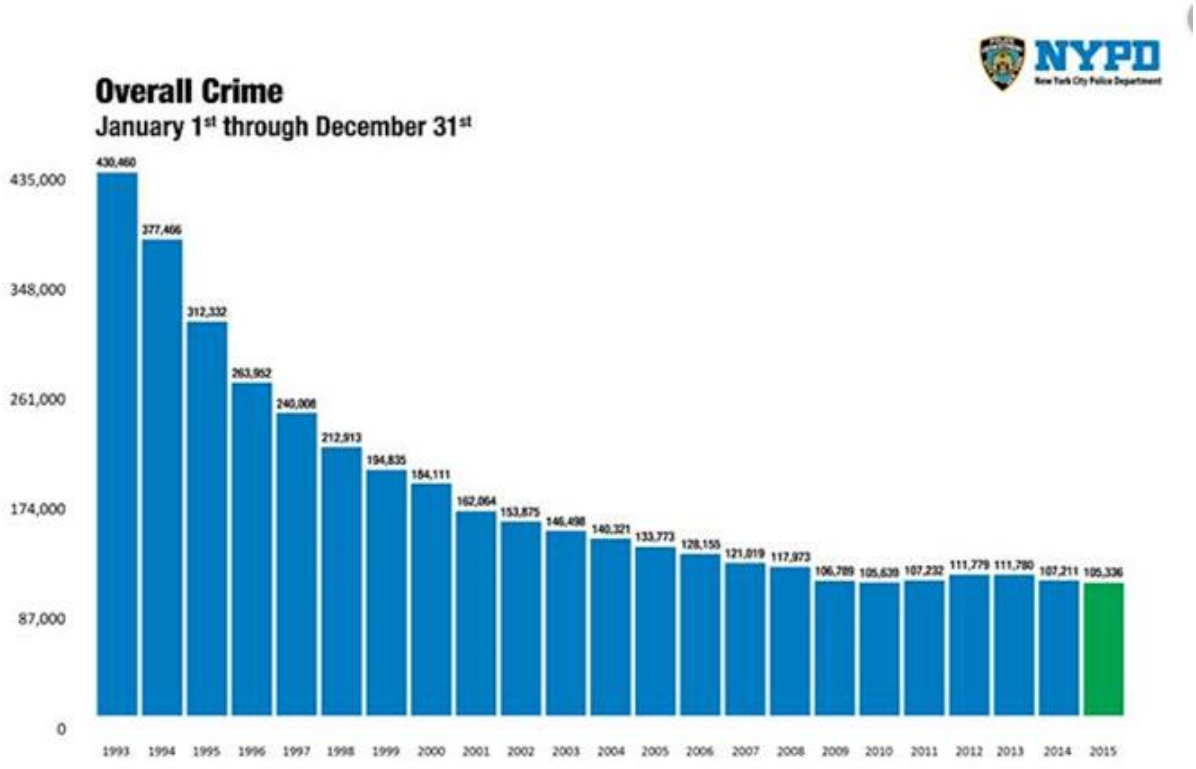
1. Introduction

1.1. Background

The crime rate in NYC has consistently decreased in the last several decades as shown in the figure below provided by NYPD.



But there are still, as shown above over 100,000 crimes committed every year and the decrease seems to have leveled off in the last few years. Reducing the crime rate stays a priority for the city.

1.2. Problem Description

It would be interesting to understand what are the areas with higher crime rate and study whether the type of venues that are established in those areas are in any way correlated to the type or number of arrests performed in those areas. Having that knowledge could help improve the allocation of security resources and focus safety messages to patrons.

It is understood the study could yield inconclusive results or no result at all. It is still interesting to try to answer the question and demonstrate the tools and technics learned throughout the data science course.

2. Data

2.1. Arrest Data

The site https://opendata.cityofnewyork.us/ provides useful public data about NYC. In particular it contains data listing all the arrests in NYC since 2013: NYPD Arrests Data (Historic) – Note: you need to create an account (it's free) to access the data.

This is a considerable amount of data that goes back several years, it contains millions of records. The venue data that we have access to through Foursquare (described in the next section) is current data, not historic data, so we will limit our study to the arrests recorded in the first 6 months of 2019. The site provides these data as well in a separate data set: NYPD Arrests Data (Year to Date). Reducing the data set could impact the study but it will help us manage the limited computing resource that are available to us for free (IBM Watson)

The Arrests data set contains a log of all the arrests performed in NYC from Jan-1-2019 to June-30-2019. There are a dozen field included for each record. The field of interests for this exercise are OFNS_DESC which the type of offense committed and the location of the arrest (Latitude and Longitude) that we will use to query Foursquare for venues.

2.2. Venue Data

We will use the Foursquare API (https://developer.foursquare.com/) to explore the venues around the arrest locations. We will limit the number of locations we will explore to 500 and we will set a radius of 200m around the locations.