Introduction

Accordingly, to an article from 2019, the average American moves once every 5 years. New York City (NYC) is the most populous city in united states, however is not the safest one. Nonetheless, NYC is a very popular city among migrants and imigrants due to the job opportunities that exist here. In turn, if the job opportunity is important at an early stage, security becomes important when families start having children. Thus, this characteristic become one of the most important when choosing a place to live. As NYC is a big city with so many opportunities, it is useful to know which neighbourhoods are the safest to live in when looking for a home.

This way, this project aims to find the safest borough in New York City based on the crimes committed between 2014 and 2015. Also, it is set as goal to select the 5 most common venues (of the safest borough) in each neighbourhood and cluster them using k-mean clustering.

Data

For this project it was used three different data sources. By using a dataset from Kaggle, it was possible to access to the crimes committed in NYC between 2014 and 2015. In this dataset it was possible to know the date of the crimes, the type and the borough where it happens and the latitude and longitude.

Second, it was used a Wikipedia page that had the list of NYC boroughs. From this table it was possible the population, gross domestic product, land area and density.

As soon as the more safety borough was selected, it was used a new Wikipedia page referred to the list of neighborhoods of the borough. In this case, the borough was Staten Island.