1. Introduction

1.1 Background

      Young professionals moving to a big city need to figure out which neighborhoods are the safest while still within budget. However, it is difficult to research every neighborhood's crime levels efficiently. I am basing my research on the criteria of a single, first-time city inhabitant who is looking for a 1 bedroom apartment in Chicago.

1.2 Problem

      Chicago has 77 community areas grouped into 9 districts. Certain areas are safe overall, but have pockets of crime which are difficult to know about without advice from locals who have lived there for a long time. This led me to group crime frequency by zipcode so that it would be easier to see exactly where incidents are occurring.

1.3 Interest

      As a young professional who moved to the city recently, it would have been very helpful to have data on this issue to help me decide where to live. It would also be a useful tool for any colleagues who are interested in moving to Chicago.

2. Data acquisition and cleaning

2.1 Data sources

      To identify low-frequency of crime neighbourhoods I used two sources of data. The first was Zillow.com rentals in Chicago, and second was crime data from City of Chicago Data Portal from January 2020 until present.

2.2 Data cleaning

      Downloading rental data from Zillow.com was a challenge. I focused on retrieving the url, address, coordinates, and details. The coordinates were the most difficult to retrieve as an issue with version control caused BeautifulSoup to not work when dealing with the Json file. The next step was to separate values in the 'details' into 'price' and 'rooms'. The 'coordinates' were separated into two columns of 'latitude' and 'longitude' as well. Using these two values, I used Uszipcode to identify the zipcode from coordinates.

      For the crime data, it was more straightforward to isolate necessary values. Coordinates were also separated into 'lat' and 'long' values and then translated into zipcode. The descriptions of crimes as related to zipcode were then counted to see the concentration of crime occurrences in each location.
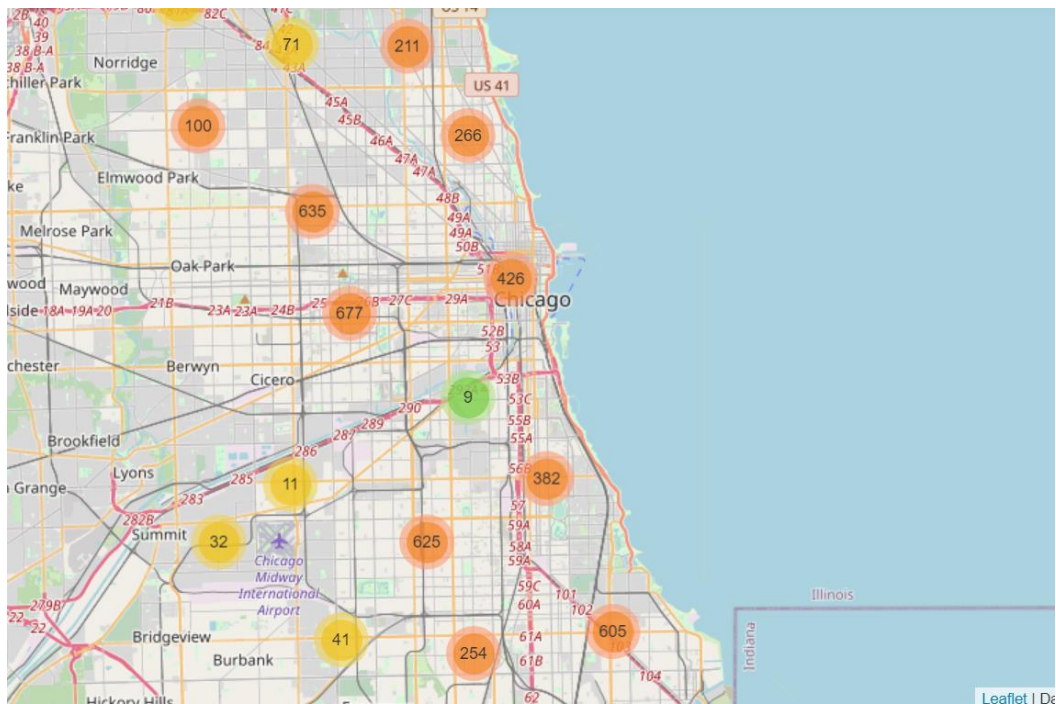
3. Exploratory Data Analysis

3.1 Examining the occurrence of crimes
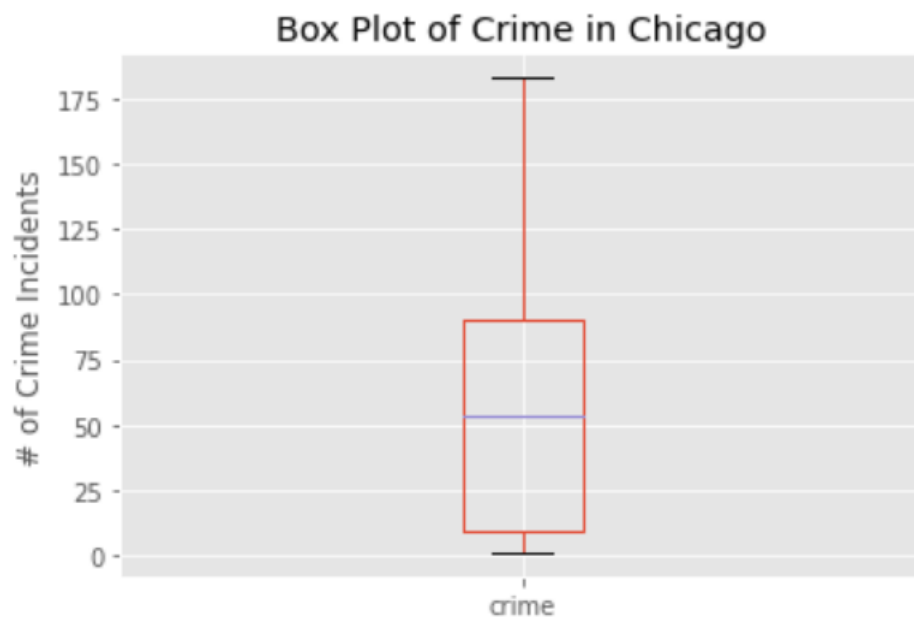

3.2 Plot of apartment locations
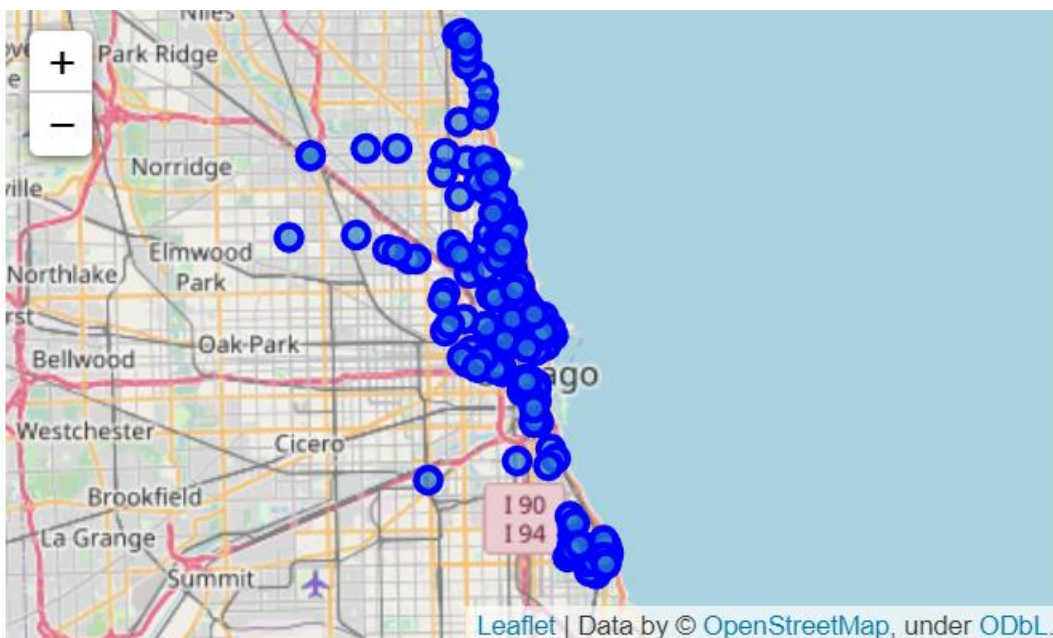
3.3 Crime location data mapped

3.4 After mapping the crime data by zipcode onto the rental data, the crime incident rate was sorted by 'High', 'Medium', and 'Low'. The attribution of 'High' was given to zipcodes where crimes occurred above 80. 'Low' was given to zipcodes where crimes occurred under 50 times. All others were assigned to 'Medium'.
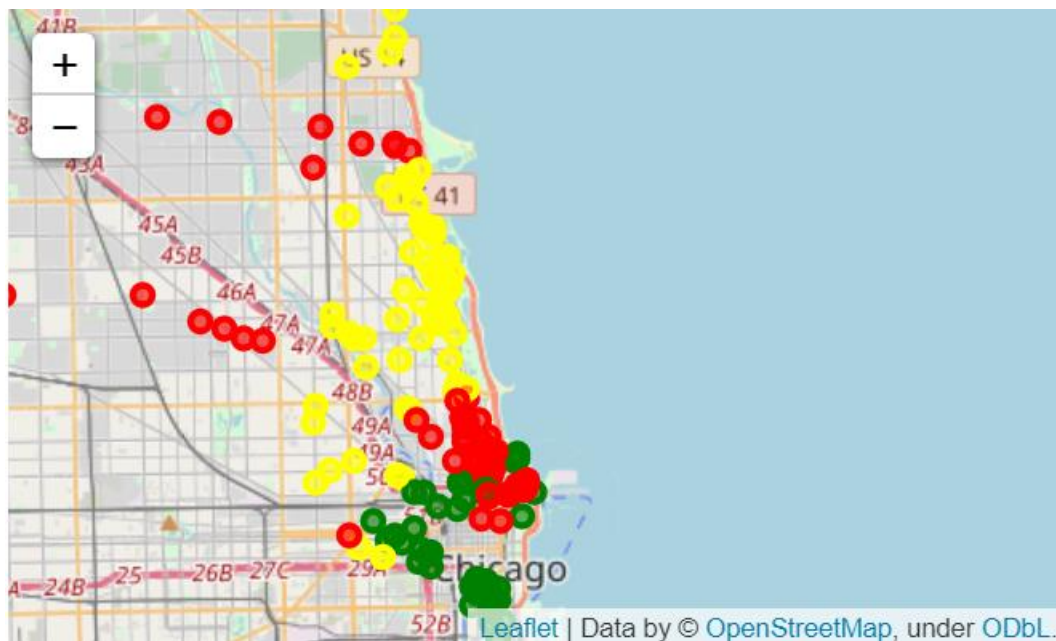
3.5 Using a box plot, I examined the correlation between price and crime levels.
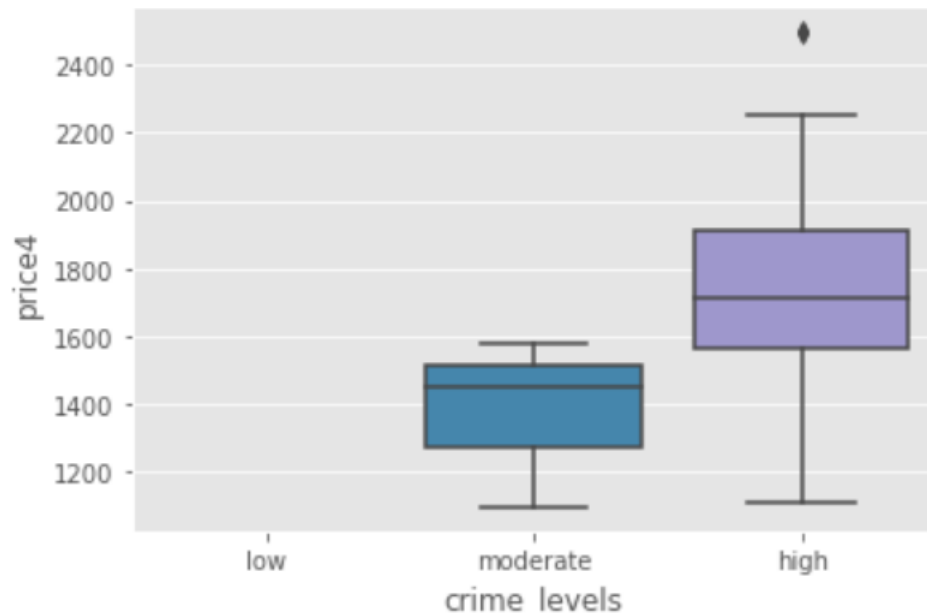
4. Conclusions

After evaluating the 77 neighborhoods in Chicago by zipcode and crime levels, rental data was mapped on to the locations. Visualizing by 'high, medium, and low' made it visually clear to see which neighborhoods and areas to avoid. Surprisingly, there was no clear correlation of price to crime incidence rates. This displays one of the difficulties of choosing 'safe' neighborhoods as each one is spread out. The map of crime locations is helpful in seeing the safety of certain streets.

5. Future directions

Only examining crime data from January 2020 does not present a full picture of the neighborhood as each one can change relatively quickly over the years. There are certainly more factors that can be considered, such as type of crime. In the future, it would interesting to update this project by including more categorizing of the crimes and also proximity to local transit.