DATA SCIENCE ASSIGNMENT

Abhilasha Lokannavar
North Carolina State University
alokann@ncsu.edu

April 3, 2019

1   Problem Statement

The goal of this assignment is to demonstrate the ability to manage a real-world data set, to communicate and contextualize summary statistics, and, most importantly, creativity and analytical rigor in generating and testing hypotheses. The assignment is intentionally nonspecific and exploratory in nature.

2   Introduction

A city like New York is ever progressing. Therefore, construction scenes are quite common. With growing businesses, residential areas and attractions one can imagine the number of construction activities taking place. A great way to figure that out is by analysing the number of building permits the New York City Department of Buildings issues each year, since permits are required before beginning any construction in the city.

3   Dataset

The data set at hand captures permits issued by the New York City Department Of Buildings (DOB), and can be accessed here . I have merged the NYC permit data with the NYC Zip Code Boundaries data set to contextualize my investigation and visualize the permits on a heat map.

4   Research Question

The project answers the following questions:

- How many building permits are issued in NYC each year?
- Where are the most building permits issued?
- What are the most common types of permits issued?

And, conclusions are made on the basis of the answers to the above questions.

5   Methodology and Experiment

5.1   Importing the libraries

Initially, all the necessary libraries are imported:

- pandas- reading the data
- seaborn, geopandas, matplotlib- visualizations

5.2 Reading the data

To get started with the analysis, data set is loaded from the source. This is done in order to manipulate the data using the pandas library.

5.3 Experimenting with the data

5.3.1 How many building permits are issued in NYC each year?

To answer this question, data is subset for 'Issued' permit status since we are not interested in any other permit status, and then the number of permits by year(excluding 2019) are counted.

5.3.2 Which types of permits are often issued?

Here, the subset data with permit status as 'Issued' is further grouped on the basis of Permit Type which can be EQ: Equipment work,PL: Plumbing, etc. To make the output more readable the abbreviations are mapped with the actual meanings to make it more readable.

5.3.3 Where the most building permits issued between 1989-2018?

Here, the subset data with permit status as 'Issued' is grouped on the basis of BoroughsManhattan, Brooklyn, Queens, Bronx, Staten Island.

5.3.4 What percentage of borough permits are for residential projects?

To answer this question, for each borough residential projects are calculated and the percentage is calculated by dividing it by total number of projects in the specific borough.

5.3.5 How many permits have been issued by zip code?

The subset data is grouped on the basis of zip code to identify the top five areas with maximum construction activity. The erroneous zip codes with values less than 10000 are dropped before plotting the data. In order to visualize the permits on a heatmap, I merged the permit count data with a geometric shape file for NYC zip code boundaries. Then the zip code boundary data is downloaded from New York's open data portal. While merging the data, it was ensured that there must be a shared column in the two dataframes and the shared column must have the same data type.

The heatmap is darker in areas with more permits issued, and lighter in areas with fewer permits issued.

6 Result

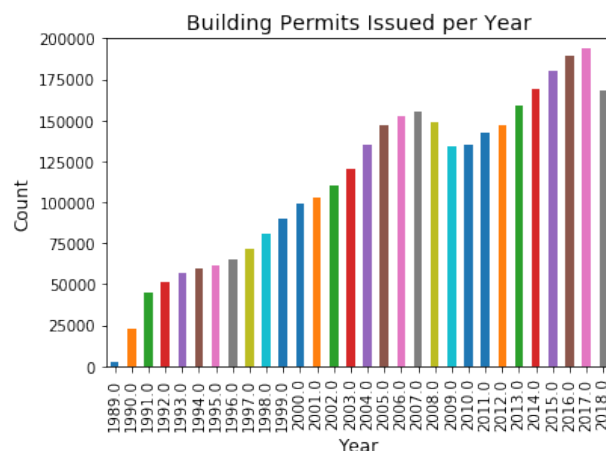The code can be accessed here:Github link
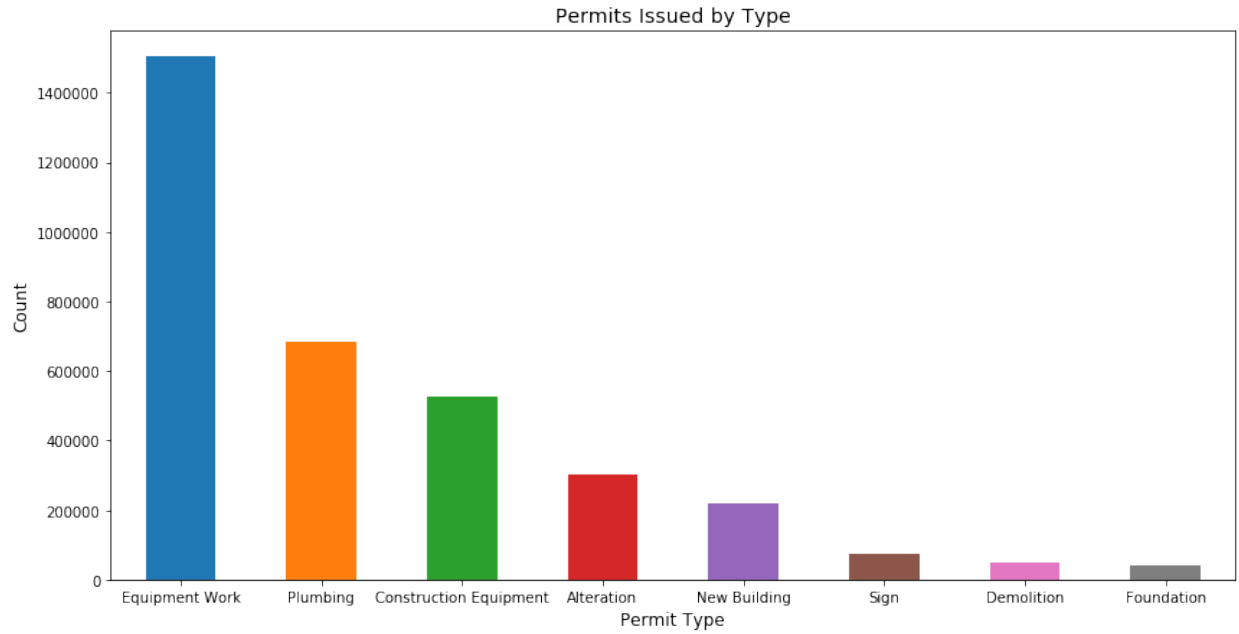

Figure 1: Building permits issued per year

2

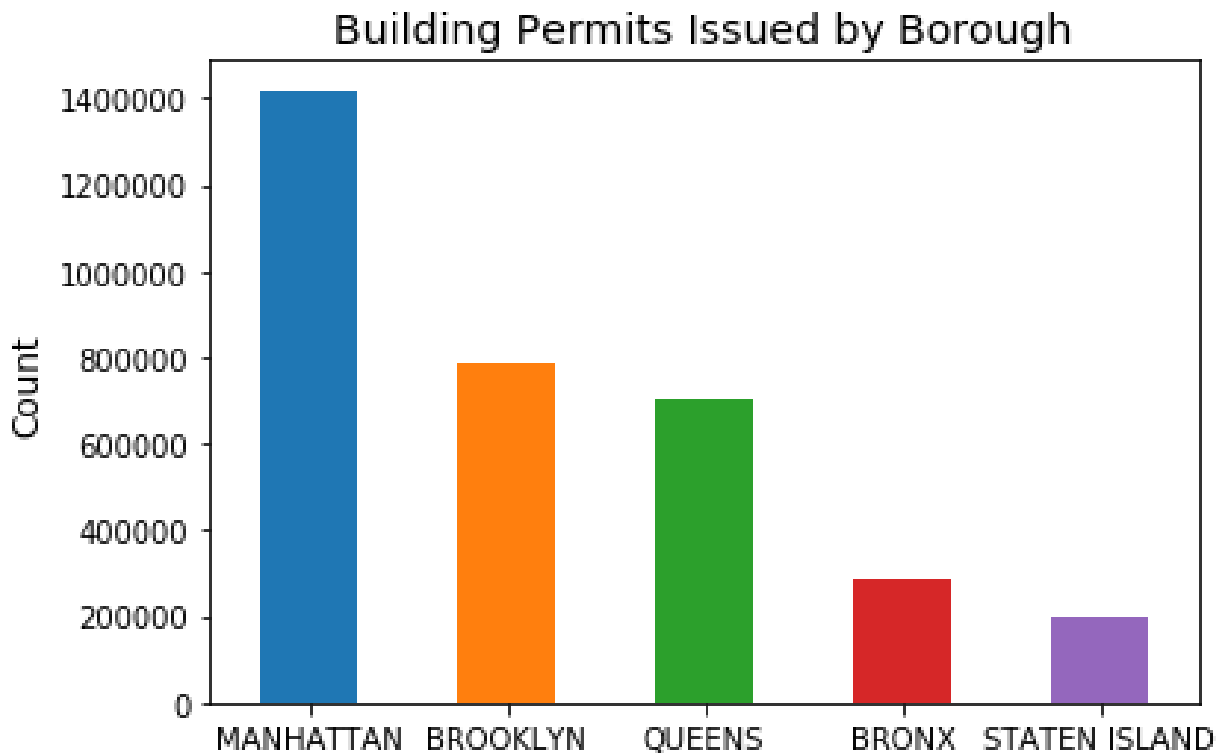Figure 2: Permits issued by type

Figure 3: Building permits issued by borough
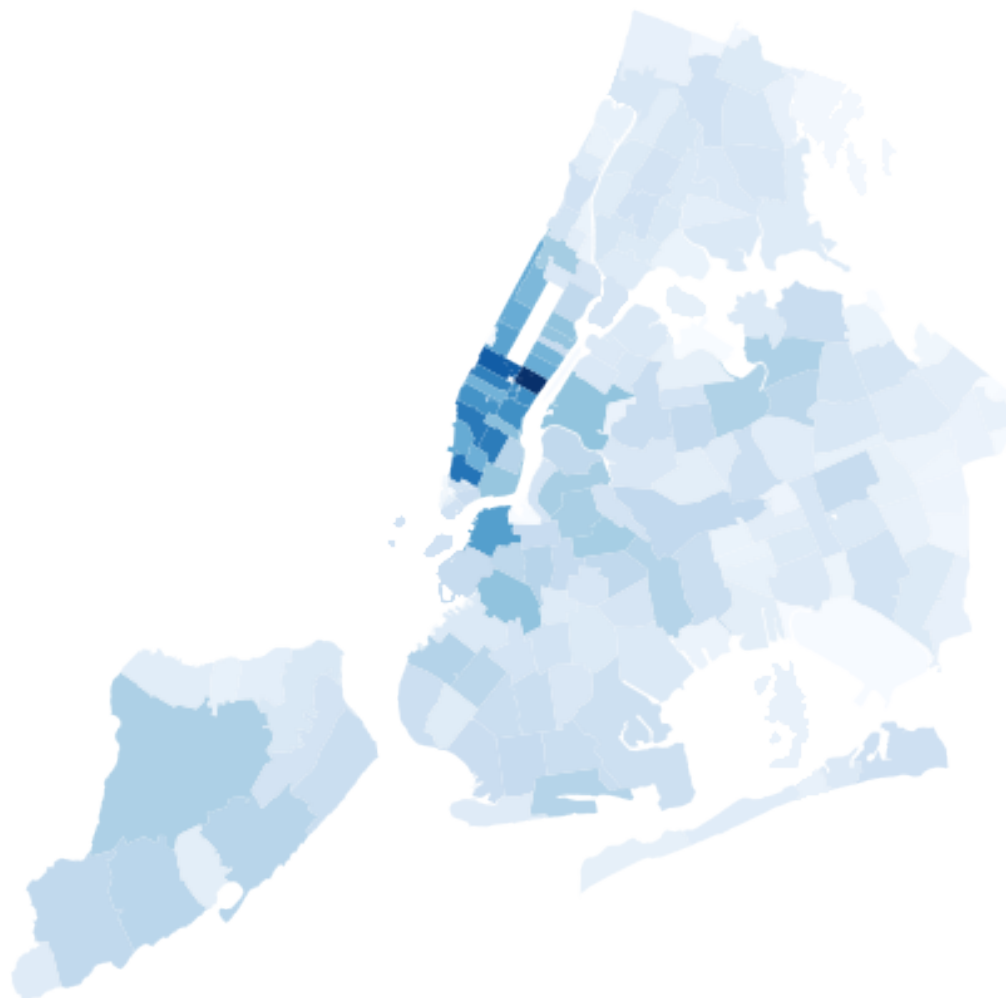
Building Permits by Zipcode

Figure 4: Building permits by Zipcode

7   Conclusion

Between 1989-2018, the New York City Department of Buildings issued over 3200,000 building permits. The New York City neighborhood Midtown East (zip code 10022) had the greatest number of permits issued between 1989- 2018. Midtown East is the home for various attractions. Considering the area generates a ton of traffic from tourists, shoppers, and workers alike, it is unsurprising that a high volume construction work is done to maintain the area and continue to drive the NYC economy.

Visualizing the building permits issued by zipcode provides better context for where permits have been issued. Merging the building data with a shape file allows you to display the density of building permits issued and locate the hotspots. For someone unfamiliar with zip codes in New York, the visual representation provides clues to spatial positioning. For example, one can see that the darkest spot is right beneath Central Park, which is more informative than simply identifying the darkest spot as zip code 10022.