STAT/CS 287 FINAL PROJECT REPORT

The Relation Among Parking Violations, Time, And Weather

Chia-Chun Chao

12-14-2017

A. Abstract

This project analyzed the relation between parking violations and date or time. Also, the weather on the same day or in the same hour is also examined to check if there is strong relation between parking violations and weather. The aim of this project is to answer following questions. First, does parking violations happen more on specific days? Second, if there is a parking violation, does it tend to happen in the morning, afternoon, or evening? Third, does outside temperature influence the number of parking violations? After analyzing parking violations in NYC in 2016, we can conclude that more parking violations occur on Sundays and between 8 AM to 2 PM. Also, the weather and temperature seem to be not strong enough to influence the number of parking violations.

B. Data

There are three categories of CSV datasets applied in this project and all were found from Kaggle.

1. NYC Parking Tickets

- Parking_Violations_Issued_-_Fiscal_Year_2014.csv
- Parking_Violations_Issued_-_Fiscal_Year_2015.csv
- Parking_Violations_Issued_-_Fiscal_Year_2016.csv
- Parking_Violations_Issued_-_Fiscal_Year_2017.csv

These datasets covering Aug 2013 to June 2017 store the information of parking ticket issued in NYC. There are more than 40 columns in the data, and only Summons Number, Violation Time, Issue Date, and Registration State were processed.

2. Weather data in New York City - 2016

- weather_data_nyc_centralpark_2016(1).csv

This dataset was collected from National Weather Service. It contains minimum temperature, maximum temperature, average temperature, precipitation, new snow fall, and current snow depth for every day in 2016 in central park. Only average temperature was used in this project.

3. New York City Taxi Trip - Hourly Weather Data

- Weather.csv

This dataset was collected for NYC taxi trip duration challenge. The purpose is different from this project, but it was chosen because it includes hourly weather data in 2016. Columns including date and time of day, temperature in Fahrenheit, and the boolean value of raining or not were applied in this project.

There are more than 10 million data in parking violations dataset, so list-wise deletion was applied to clean missing values. All data about date or time were processed to fit the datetime object. For example, 1037A was formatted to 10:37 AM, and 0025A was changed to 12:25 AM. With the datetime object, three dataset were able to be successfully merged.

C. Results

This project was organized in 7 steps.

1. Process the original parking violations data

Simplified datasets, where only attributes that may be analyzed are maintained, can reduce the processing time for future analysis. The range of the dataset in 2016 is from June, 2015 to June, 2016, so it is safer to process three datasets and only store data in 2016 to an intermediary file. In this step, the original datasets are narrowed down to only contain Summons Number, Violation Time, Issue Date, and Registration State. After narrowing down the number of columns, all intermediary datasets are also organized to only contain the parking violations occurring in the specific year.

2. Answer if parking violations happen more on specific days

Before answering this question, missing values should be checked and processed. List-wise deletion is performed to drop any observations where at least one value is missing.

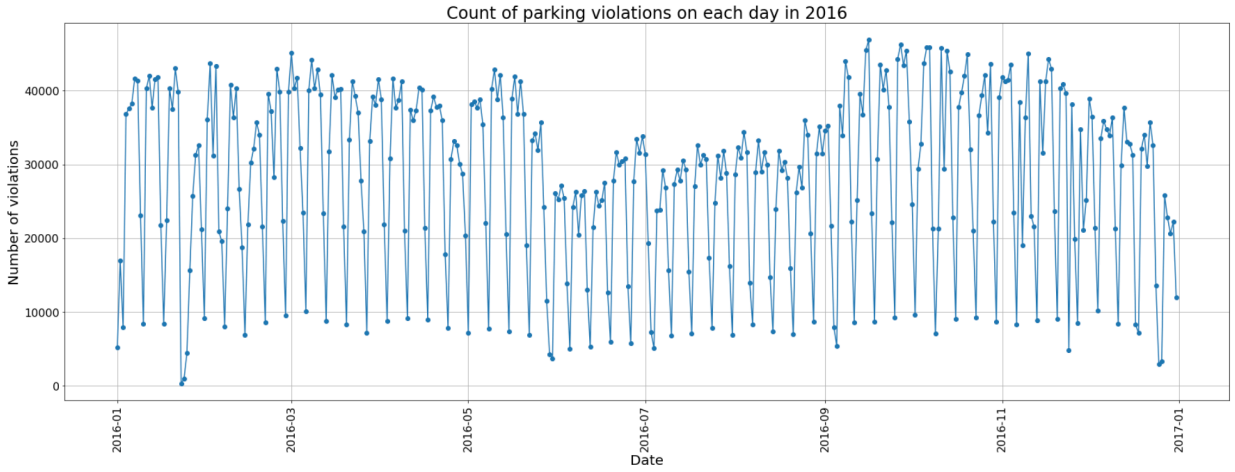

Figure 1: Count of parking violations on each day in 2016

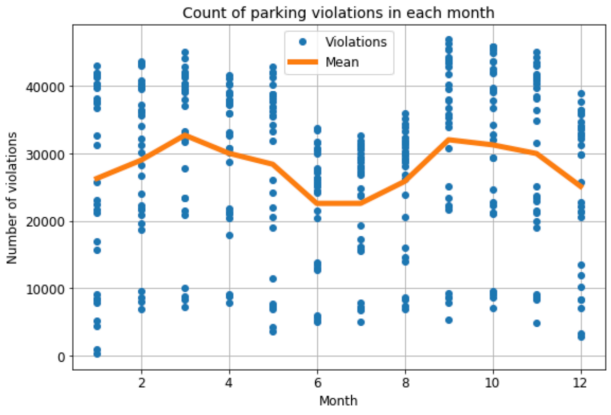

Figure 2: Count of parking violations in each month

Then, we need to count the violations occurring on the same date. Dates of violations are stored in a list and sorted by counter in collections. Figure 1 shows that the numbers of parking violations in between June to August are obviously fewer than other months. Therefore, the largest 50 numbers of parking violations are printed to show if there is any date in June, July, or August and the result shows none of date in that range. Figure 2 is also presented to show the mean of violation number in each month.

Vertices with lower values seem to be in some pattern, where distance between two adjacent lower vertices are similar, so the smallest 30 numbers of parking violations are printed. Most of the weekdays with fewest violations are Sunday. There are exactly 52 Sundays in the year 2016. The result shows that there are 22 Sundays over 30 weekdays with fewest number, which is a very large proportion. Therefore, this proves that parking violations occurred less on Sunday or parking violations were penalized less on Sunday. Besides Sundays, there are also some holidays included in the top 30 days. For example, Memorial Day, Independence Day, Labor Day, Thanksgiving, and Christmas.

3. Answer if it tends to happen in the morning, afternoon, or evening when there is a parking violation.

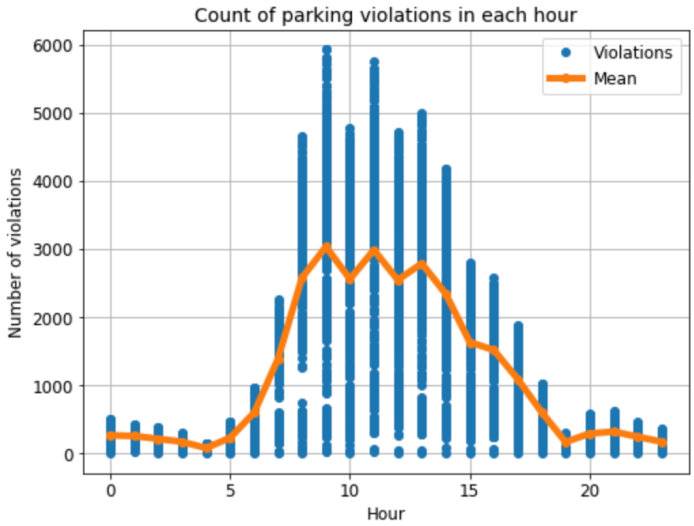

Figure 3: Count of parking violations in each hour

In this step, issue date and violation time are combined to form a complete datetime object. Strings of violation time are processed to fit the format of datetime object. For example, violation time, 1037A, is changed to be 10:37 AM, and 0025A is changed to 12:25 AM because 12-hour clock starts from 01 to 12. Some abnormal minute formats are also processed when parsed to datetime object.

A plot is made to show the number of violations in each hour and the mean of violations in each hour. Figure 3 demonstrates that during 8 AM to 2 PM, there were relatively higher numbers of parking violation. Therefore, we can not conclude that parking violations occurred more in the morning or in the afternoon. Instead, the higher result ranges between 8 to 14 in 24-hour clock.

4. Answer if daily average temperature influences the number of parking violations.

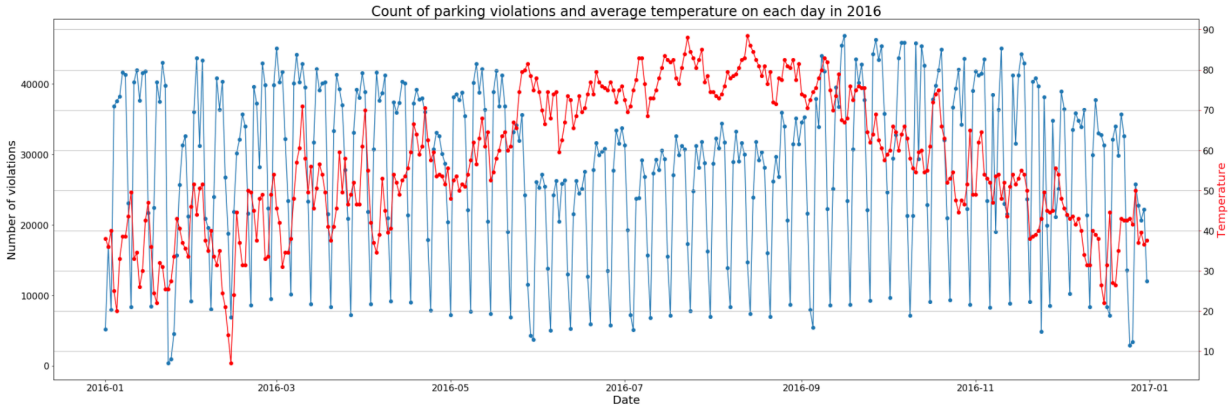

Figure 4: Count of parking violations and average temperature on each day in 2016

To answer this question, NYC weather dataset is processed first. This dataset contains the highest and lowest temperatures on each day in 2016. Figure 4 shows the relation between violation count and temperature, but it is too ambiguous because when temperature is too high or too low, people feel worse than a medium temperature. Therefore, the base temperature is set as 70 and the difference between the temperature on that day and the base is divided by 5 to determine the degree of uncomfortable feeling, as shown in figure 5.

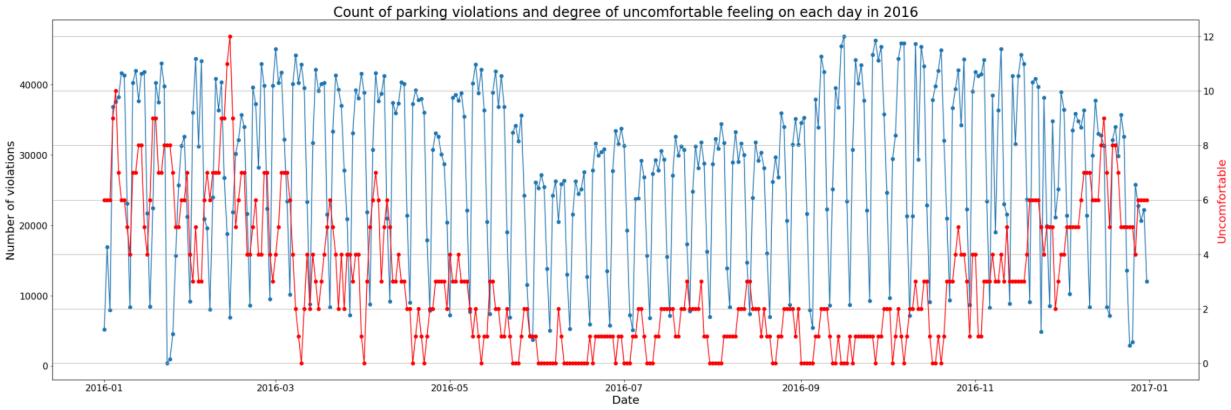

Figure 5: Count of parking violations and degree of uncomfortable feeling on each day in 2016

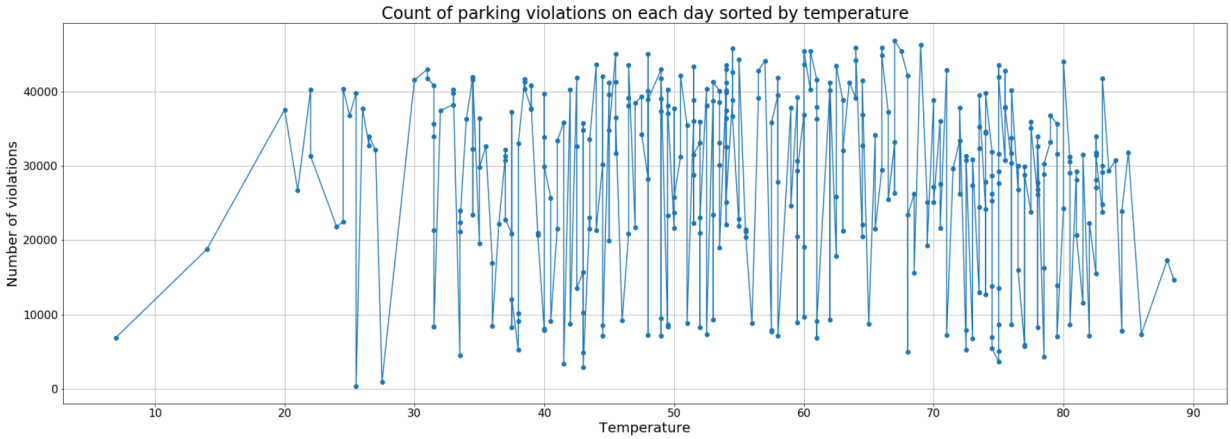

Figure 6: Count of parking violations on each day sorted by temperature

Figure 5 still does not show the relation, so issue days are sorted by the average temperature on that day and Figure 6 is created to show the direct relation between temperature and violation count. The numbers of violations at every temperature are scattered widely, so we can conclude that the relation of parking violations and temperature might be too weak to be shown in figures.

5. Answer if hourly outside temperature influences the number of parking violations.

To answer this question, hourly weather dataset is processed first and data is stored in a dictionary. There are 24 hours in a day, so the amount of data might be too large and hard to process. Therefore, the dictionaries of violations and hourly temperature were classified into different months. Because previous result showed that during 8 AM to 2

PM, there were relatively higher numbers of parking violation, only the violations and temperatures during this period are maintained. This process makes the figure more simple.

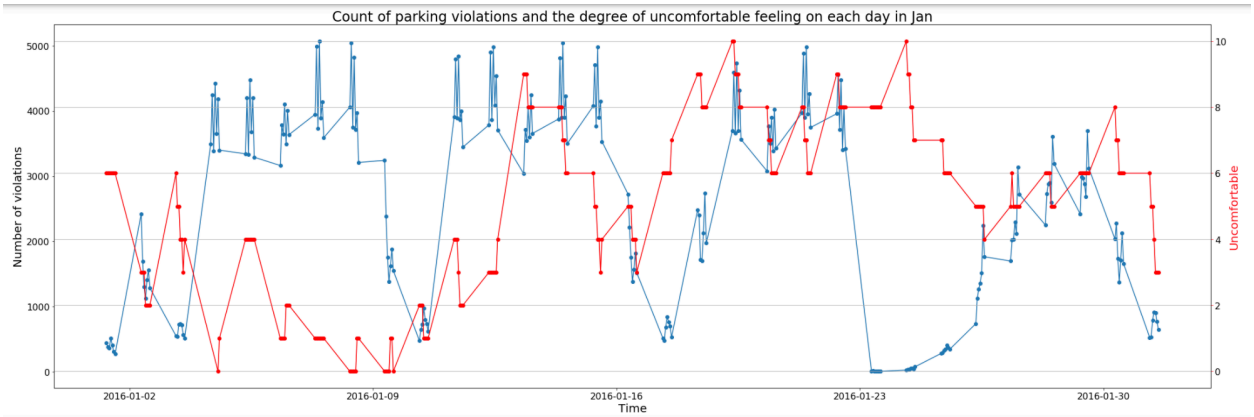

Figure 7: Count of parking violations and the degree of uncomfortable feeling on each day in Jan

Because of the ambiguous issue mentioned previously, only the degree of uncomfortable feeling is considered here. There are 12 figures showing data in 12 months and there seems not to be any trend in the figures. Therefore, even if temperature is examined hourly, the relation between parking violations and temperature is still not strong enough to be observed.

6. Answer if the number of parking violations is higher or lower when it is raining.

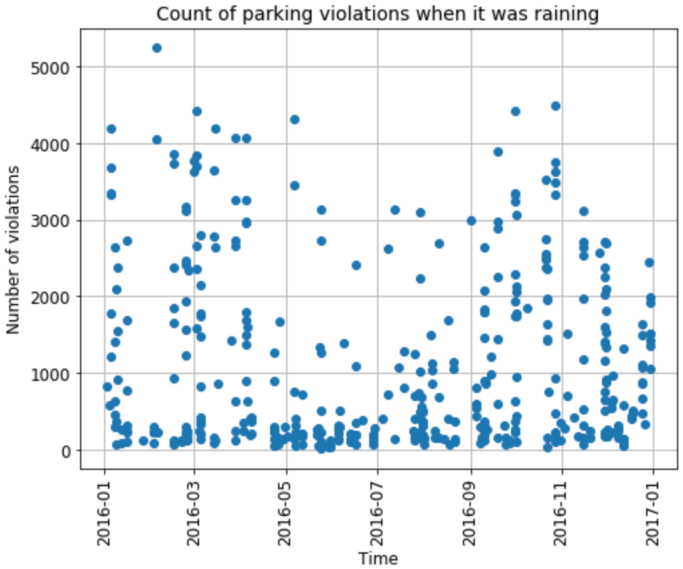

Figure 8: Count of parking violations when it was raining at the violation time

To do this, a boolean attribute of rain is used to find the hourly data where the boolean value is true. This figure shows that the relation of raining and violation is also weak to answer the question.

7. Find the registration states with the largest numbers of parking violations.

The result shows that interestingly. When the numbers of parking violations were calculated by the registration state of vehicles, Florida is the fifth. However, the distance between FL the farthest from NYC in these top ten states, including NY, NJ, PA, CT, FL, MA, IN, VA, MD, and NC.

D. Discussion and Conclusion

After analyzing these datasets, we can conclude that date influences the number parking violations the most. Parking violations occurred less on Sunday and some specific holidays or parking violations were penalized less on Sunday and some holidays. When the temperature is too high or too low, or it is raining outside, the number of parking violations may not be affected. There are more parking violations issued during 8AM to 2PM, which may result from the facts that there are more vehicles on the streets or more parking enforcement officers are working during the period.

In my proposal, I assume that when it is raining, parking spaces are harder to find because there are more vehicles on streets, which will result in more parking violations. However, all my results about temperature and weather show that there is not strong relation shown in the figures. I may need more statistic and analysis methods to develop a more precise conclusion. In the future, if driver's gender, income levels, age, identity, or reason of violating parking laws is contained in some datasets, that might be much more interesting to analyze the relation between parking violations and the information.

E. Dataset Source

https://www.kaggle.com/new-york-city/nyc-parking-tickets

https://www.kaggle.com/mathijs/weather-data-in-new-york-city-2016

https://www.kaggle.com/meinertsen/new-york-city-taxi-trip-hourly-weather-data