Parking Roulette

Introduction

Cars are the go-to means of transportation in everyday life in United States. With the largely based on highway system and lower investment of Public transport, the numbers of cars that are on the streets is relatively large. With the large number of private vehicles in street, the first problem that arises is of the parking. Here, we take on example of the City of Los Angeles, one of the busiest cities in the world and analyze the parking citations. Our goal is to identify the most common areas for the citations and see if there are enough public parking available and I not how can new parking lots be established so that the number of violations can be decreased.

The dataset to be analyzed is maintained by Socrata, and is publicly available and updated on a daily basis. Our goals would be to format the data to suit our analysis and visualize it on the map to see what part of the city are massively guilty of parking violations and leverage the Foursquare APIs to locate nearby parking lots which will help to understand better about the problem to be a negligence or a rather lack of proper infrastructure and management by the city. This analysis helps both the public office and private business pointing out where the infrastructure is lacking so that, a newer infrastructure with a nominal service fee would be able to generate a large revenue.

Data

As mentioned previously, the dataset is maintained by Socrata, an organization which helps to maintain and manage data for the public. The dataset is considerably large with 8.8 million rows and 19 different columns. Among the 19 columns, our primary focus would be on following features.

- Issued date
- Issue Time
- Location
- Violation
- Fine Amount
- Latitude
- Longitude

This would then help us determine the circumstance and the location and with the Foursquare API in hand, we can find out more on nearby parking spots, public sites to have better idea about the traffic problems.

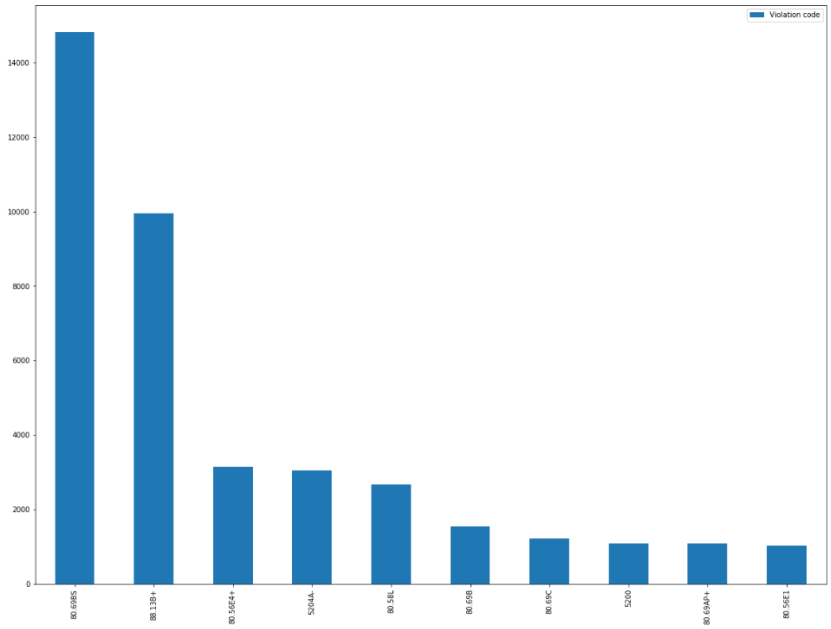
The dataset was organized to some extent, but we had a lot to do to make it ready for our analysis. We had several missing values in different columns while there were multiple columns which were irrelevant for the scope of our analysis. We dropped columns like “Meter Id”, “VIN”, “Make”, “Color”, “RP State Plate” etc. Because our analysis is more focused on when and where the incident occurs rather than on what.

One approach of handling the missing data was to fill it with the average of the column. This approach made more sense when dealing with the missing values in “Fine” column where we took an average of the remaining data and used it to fill the missing data.

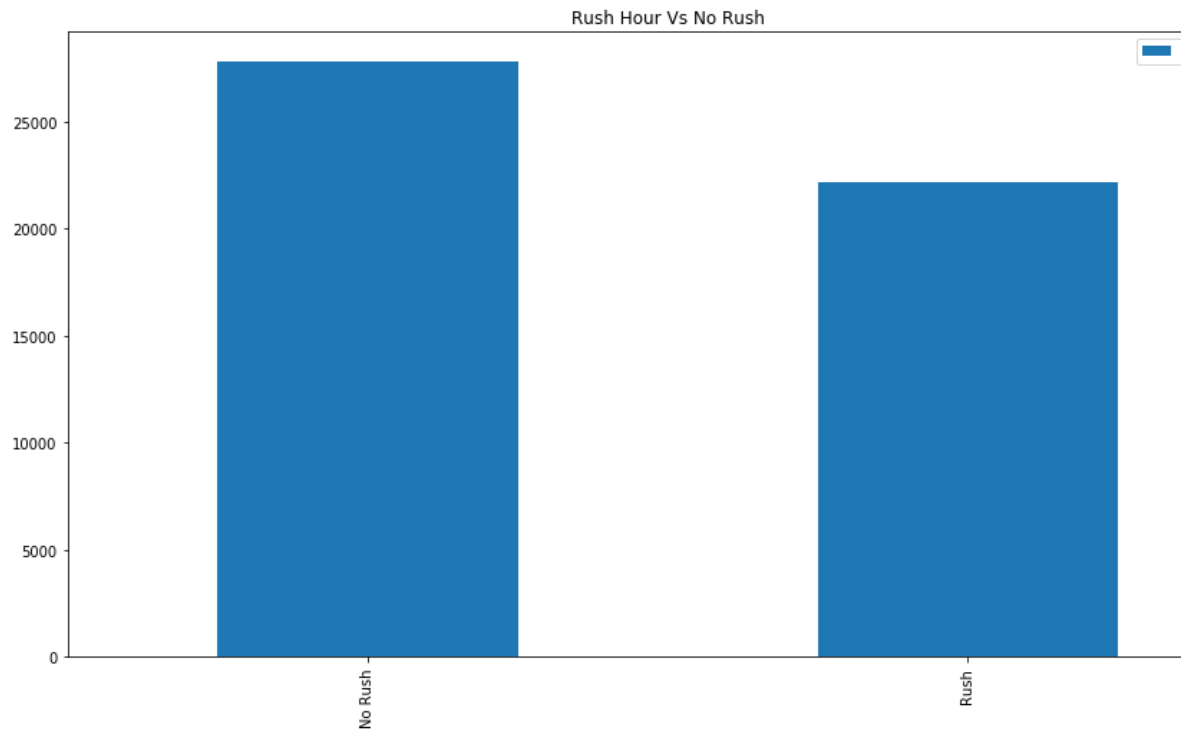
Since we are largely focused on using the location of the places, the location data needed to be processed to suit our needs. The dataset provides the latitude and longitude coordinates in US Feet coordinates according to the *NAD_1983_StatePlane_California_V_FIPS_0405_Feet projection* which isn’t suitable for the map plots. We filter out the bad data, from the latitude and longitude which was 99999.0 in our case and then use the Pyproj library to convert it into the latitude and longitude coordinates.

Methodology

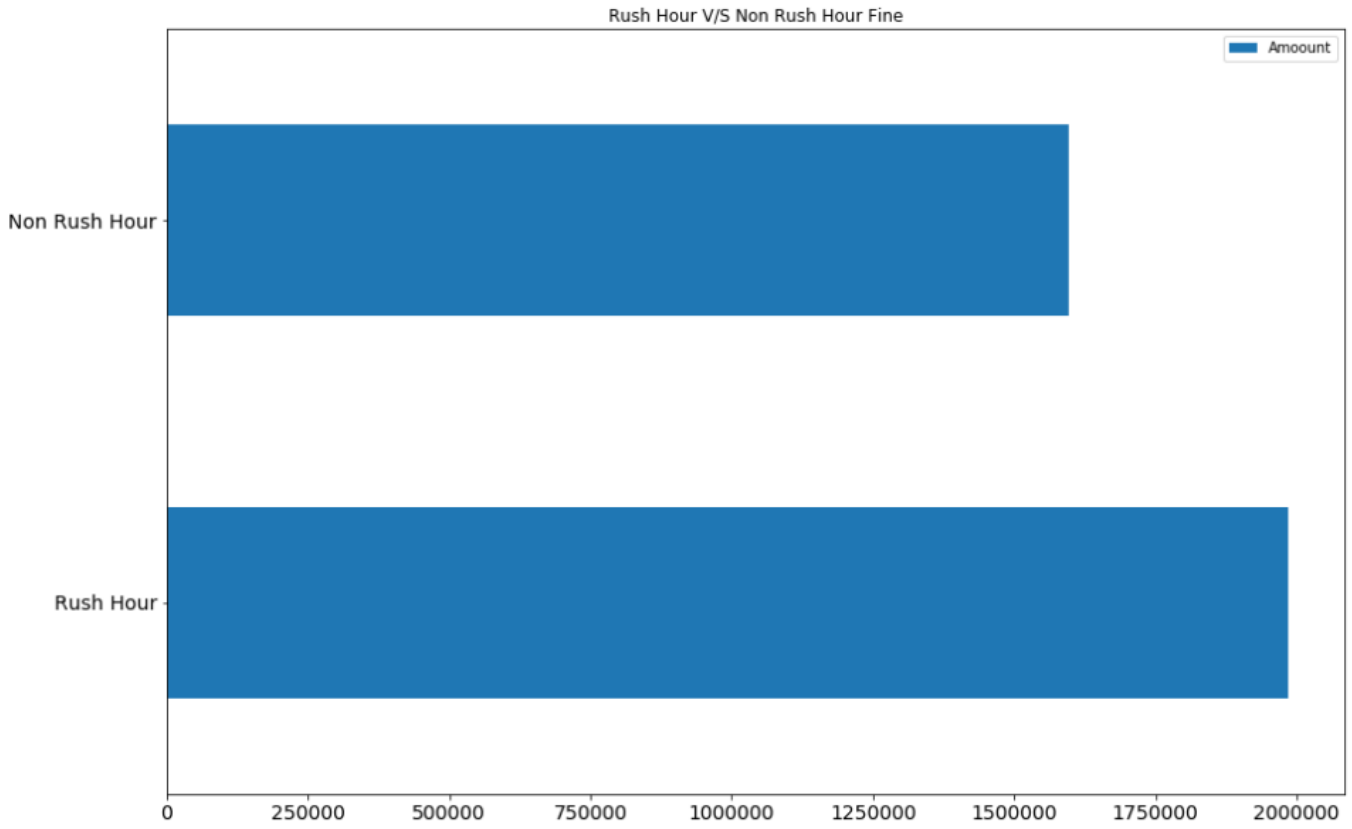
The dataset has loads of information which we can use to perform basic analysis. We proceed further with attempts to visualize various aspects of the datasets. For starters we try to visualize the 10 most common traffic violation to get a better sense of what is happening in the streets.



Furthermore, we decided to create a new feature called “Rush hour” to divide our violations based on the time of the day. We picked two timespans where the traffic is in peak hour in LA and categorized our dataset on its basis. We then try to visualize it and get a sense of what time the traffic violations are more frequent.



From this figure we can see, the parking violation in LA is more during the off-peak hours than it is during the peak hours. However, on further exploration into the amount of fines paid during the each category are vastly different.

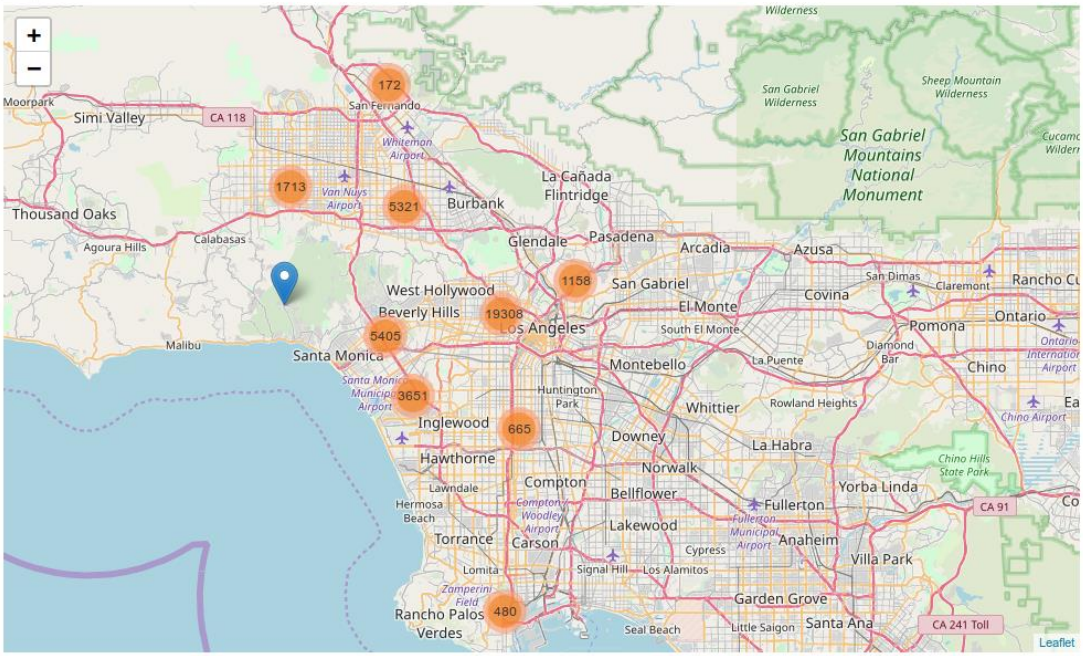


The city of LA paid almost 3.5 million in fine in less than two years but the amount of money spent on rush hour fines is a lot higher compared to the number of the parking violations.

One of the major parts of the project was to visualize the parking violation based on its geolocation. After the cleaning of the latitude and longitude data we use the Folium library to generate an interactive map that gives us a better sense of the traffic violation on the geospatial plane. Since we have large number of points, we cluster them together which is interactive and can be explored in detail when clicked on it.

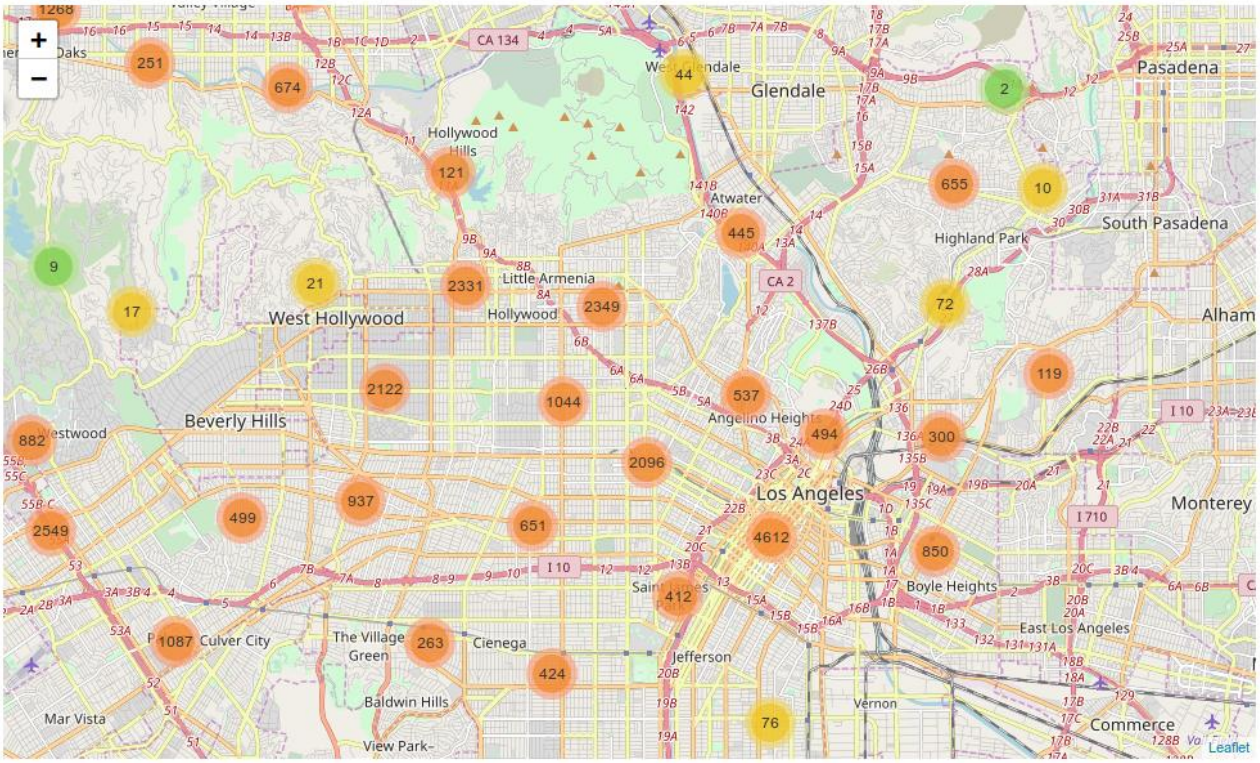
```
In [43]:
1 LAmap.add_child(mc)
2 LAmap
```

Out[43]:

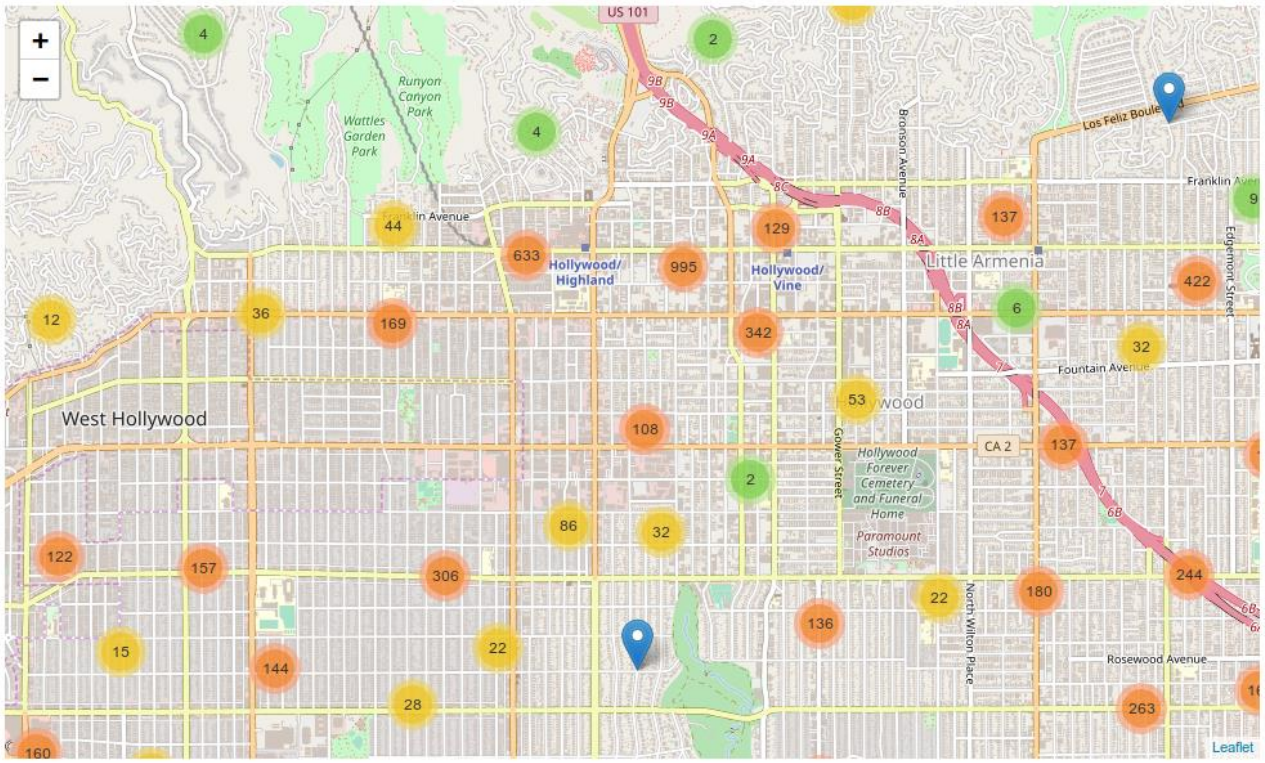


On further clicking into the clusters on the map we can see the more detailed plotting of the location of traffic violation.

3]:


CA 134
Glendale
Pasadena
Hollywood Hills
Atwater
South Pasadena
Highland Park
Little Armenia
Hollywood
West Hollywood
CA 2
Beverly Hills
Angeleno Heights
Los Angeles
Monterey
I 10
I 710
Boyle Heights
East Los Angeles
Culver City
The Village Green
Cienega
Jefferson
Baldwin Hills
Vernon
Mar Vista
View Park-
Commerce
Leaflet

Out[43]:


US 101
Runyon Canyon Park
Wattles Garden Park
Los Feliz Boulevard
Bronson Avenue
Franklin Avenue
Hollywood/ Highland
Hollywood/ Vine
Little Armenia
Edgemont Street
Fountain Avenue
West Hollywood
Gower Street
CA 2
Hollywood Forever Cemetery and Funeral Home
Paramount Studios
North Wilton Place
Rosewood Avenue
Leaflet

So now that we have our breakdown of our locations, we then compute the 10 most common location with traffic violation and leverage the power of the Foursquare API to find the parking lot around the most common places.

Results

After our analysis, we come to our final result where we find the number of parking lots around our most frequent area of traffic violation and calculated the ratio of parking violation to the number of parking lots in the area.

Out[57]:

	Location	counts_parking	Latitude	Longitude	counts	ratio
0	100 LARCHMONT BL N	20	34.073023	-118.323530	43	2.150000
1	101 LARCHMONT BL N	20	34.073023	-118.323530	35	1.750000
2	1301 ELECTRIC AVE	20	33.991691	-118.467910	52	2.600000
3	2377 MIDVALE AVE	20	34.040144	-118.429894	35	1.750000
4	4867 SUNSET BLVD W	20	34.098141	-118.295389	72	3.600000
5	5901 98TH ST W	20	33.947344	-118.385970	45	2.250000
6	2800 E OBSERVATORY	6	34.123408	-118.302409	58	9.666667
7	2800 W OBSERVATORY	6	34.123408	-118.302409	35	5.833333
8	1235 FIGUEROA PL	4	33.782415	-118.281029	52	13.000000
9	11601 SAN VICENTE BL	2	34.050751	-118.480653	42	21.000000

Conclusion

From the table above we can clearly see the San Vincent Boulevard had the worst ratio of traffic violations to the number of parking lot, so we can conclude that constructing new parking lots around the neighborhood can solve the issue and also generate revenue as parking costs.