



City's Safety Predictor

SIDDHANT MAHALLE

Introduction

- ▶ Imagine visiting a new city for the first time. You want to visit all the top sites and the best restaurants there. But you have no one to show you around or to suggest these places.
- ▶ You want to visit these places, but you want to be cautious of crime-riddled neighborhoods and avoid such routes.
- ▶ The idea is to combine the venue and location data from Foursquare along with open source crime data to provide the traveler with a list of attractions and restaurants along with a graphical representation of crime statistics in those areas

Approach

- ▶ The traveler decides on a city location [in this case, Chicago].
- ▶ The Foursquare website is scrapped for the top venues in the city.
- ▶ From this, a list of top venues is augmented with additional geographical data
- ▶ Using this data, the top nearby restaurants are selected.
- ▶ The historical crime statistics within a predetermined distance of all venues is obtained.
- ▶ A map is presented to the traveler showing the selected venues and crime statistics of the area.
- ▶ The future probability of a crime happening near or around the selected top sites is also presented to the user

Data acquisition

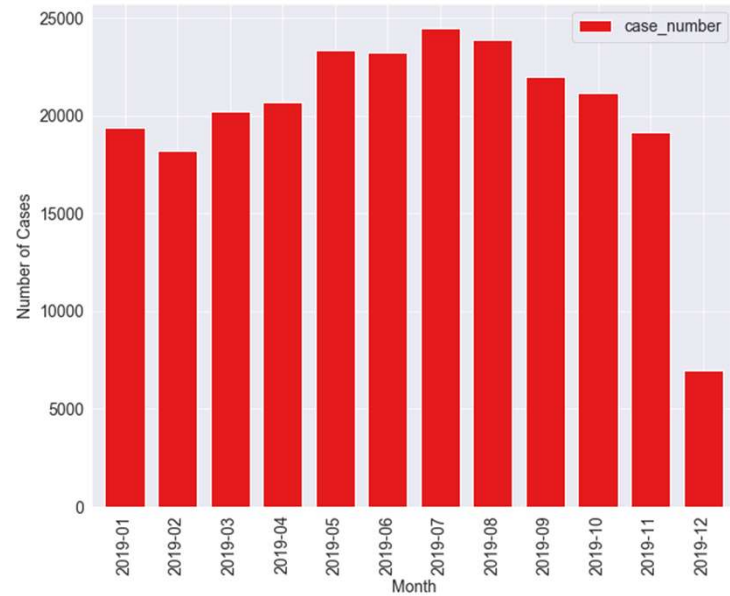
The data acquisition section is divided in two parts:

- ▶ Venues and Location data obtained from FourSquare
 - Query the Foursquare website for the top sites in Chicago
 - Use the Foursquare API to get supplemental geographical data about the top sites
 - Use the Foursquare API to get top restaurant recommendations closest to each of the top site
- ▶ Chicago Crime Data
 - This dataset can be obtained from the [Chicago Data Portal](#) and reflects reported incidents of crime (except for murders where data exists for each victim) that occurred in the City of Chicago.

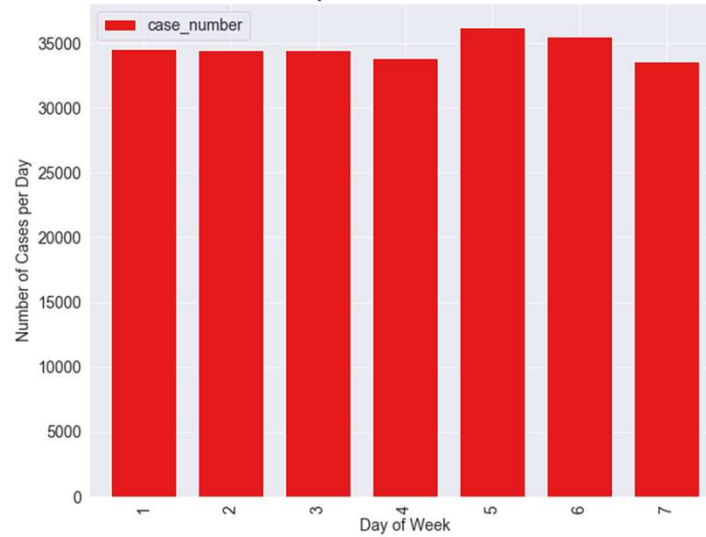
Data Visualization

Visualizing the Chicago crime data

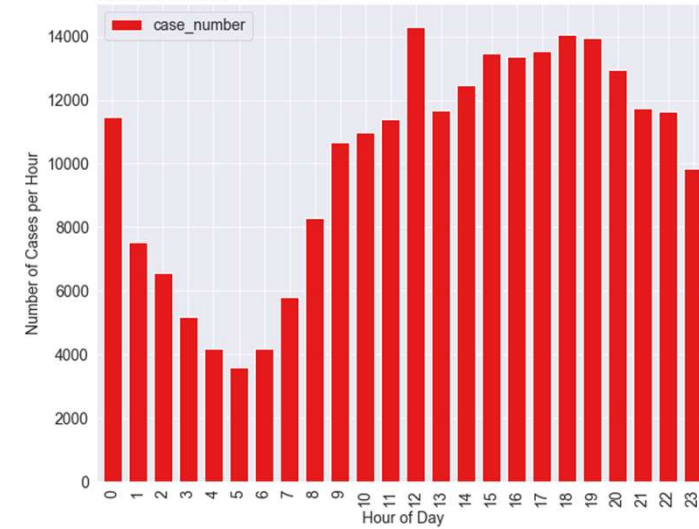
Number of Cases Per Month



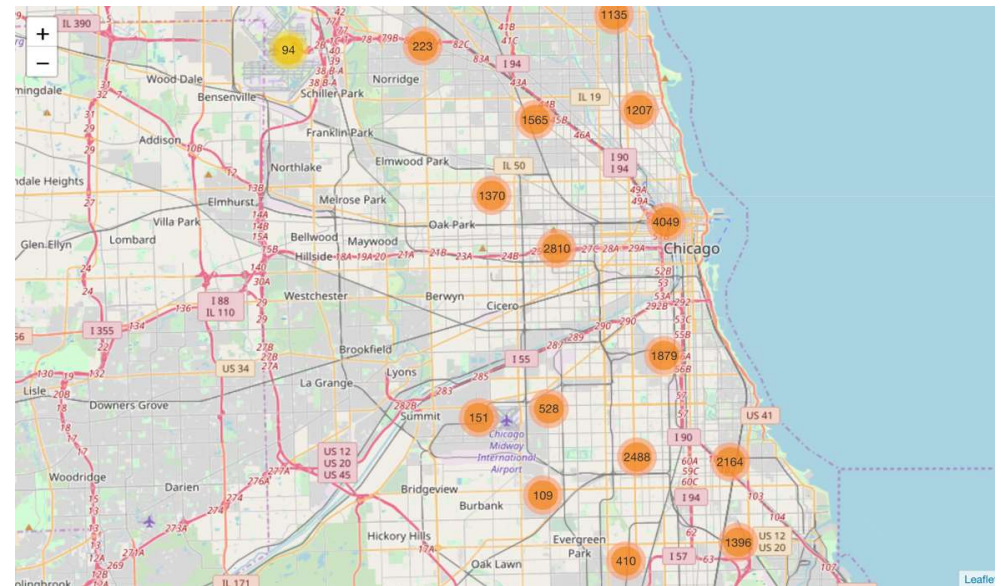
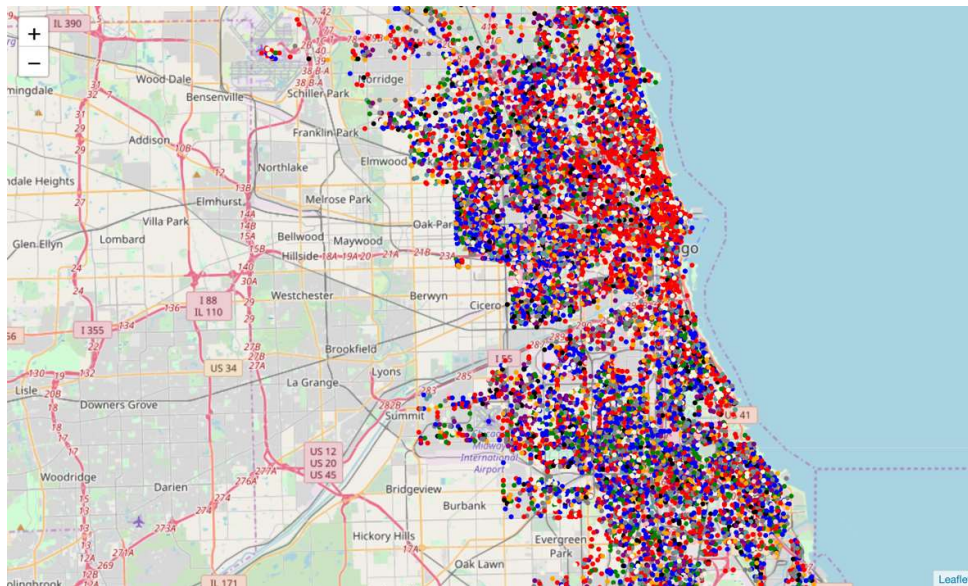
Number of Cases Per Day of Week



Number of Cases Per Hour



Data Visualization (Cont.)



Data Analysis

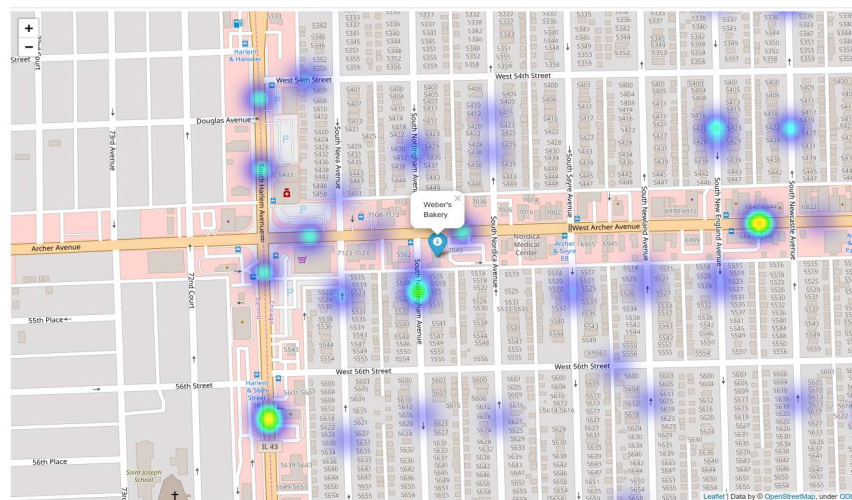
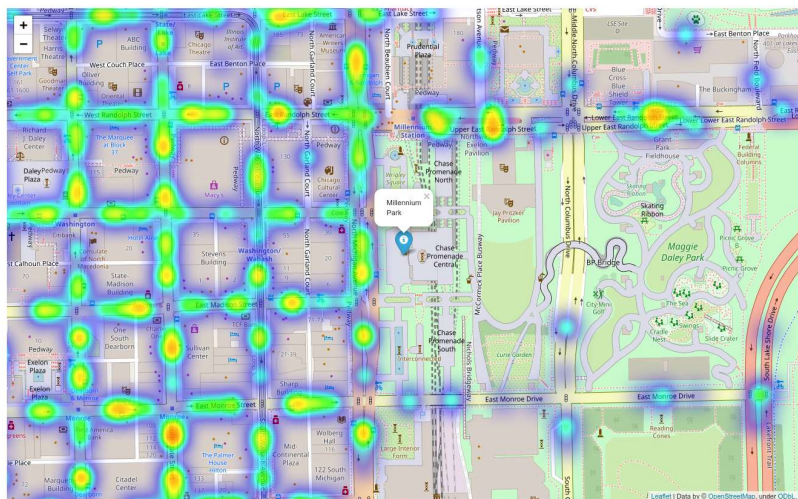
Five model type were then chosen to be evaluated:

1. K Nearest Neighbors
2. Decision Trees
3. Logistic Regression
4. Naive Bayes
5. Decision Forest using a Random Forest

	Jaccard	F1-Score	LogLoss
Algorithm			
KNN	0.687951	0.717168	10.777931
Decision Tree	0.694382	0.703628	10.555772
Bernoulli Naive Bayes	0.611236	0.658090	13.427587
Logistic Regression	0.621219	0.682511	13.082811
Random Forest	0.994625	0.995126	0.185660

Results and Prediction

- ▶ Of the top 20 venues 17 were identified as potentially dangerous to visit and 3 was deemed safe. As there is no data to compare the predictions against the best way, we will visualize the data again.
- ▶ We will look at the following 5 venues:
 1. Millennium Park
 2. Chicago Riverwalk
 3. The Chicago Theatre
 4. Weber's Bakery
 5. Chicago Lakefront



. Conclusions and Discussions

- ▶ Although all the goals of this project were met there is room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be developed into a fully-fledged application that could support the cautious traveler in an unknown location.
- ▶ Of the contributing data the Chicago Crime data is the one where more data would be good to have. Also, not every city in the world makes this data freely available so that is a drawback.
- ▶ FourSquare proved to be a good source of data but frustrating at times. Despite having a Developer account, I regularly exceeded my hourly limit locking me out for the day. Therefore, Pickle was used to store the captured data.

Further Development

The following are suggestions how this project could be further developed:

- ▶ Best time to visit each venue.
- ▶ Suggestions for morning, afternoon, evening and nighttime.
- ▶ Daily itineraries.
- ▶ Route planning and transportation.
- ▶ Time lapse of the crime in the area of the venue.
- ▶ Favorite dining preferences could be used to choose the restaurants.

GitHub Link: <https://github.com/siddhantmahalle/Applied-Data-ScienceCapstone->