**Capstone Project/Final Report**

**1. Introduction:**

The purpose of this Capstone Project was to find the location for a new restaurant in Austin, Texas, USA. The location has to be near the downtown and supposed to be with minimum criminal activity.

Aims to create an analysis of crime reports from the Austin Police Department from 2018 and analysis of Austin restaurants. To determine the location for a future venue.

**2. Data Section**

Data Link: https://data.austintexas.gov/Public-Safety/Crime-Reports-2018/vmn9-3bvu

I used the APD dataset. Yet it required a lot of cleaning and analyzation since it has data that was not important. In addition, I wanted to show details like charts and top crime report to help to determine the final decision. Dataset consisting of:

Crime Reports

|  |  |
| --- | --- |
| Field | Description |
| Incident Number | Incident report number |
| Highest Offense Description | Description |
| Highest Offense Code | Code |
| Family Violence | Incident involves family violence? Y = yes, N = no |
| Occurred Date Time | Date and time (combined) incident occurred |
| Occurred Date | Date the incident occurred |
| Occurred Time | Time the incident occurred |
| Report Date Time | Date and time (combined) incident was reported |

|  |  |
| --- | --- |
| Report Date | Date the incident was reported |
| Report Time | Time the incident was reported |
| Location Type | General description of the premise where the incident occurred |
| Address | Incident location |
| Zip code | Zip code where incident occurred |
| Council District | Austin city council district where the incident occurred |
| APD Sector | APD sector where incident occurred |
| APD District | APD district where incident occurred |
| PRA | APD police reporting area where incident occurred |
| Census Tract | Census tract where incident occurred |
| Clearance Status | How/whether crime was solved (see Clearance lookup) |
| Clearance Date | Date crime was solved |
| UCR Category | Code for the most serious crimes identified by the FBI as part of its Uniform Crime Reporting program |
| Category Description | Description for the most serious crimes identified by the FBI as part of its Uniform Crime Reporting program |
| X-coordinate | X-coordinate where the incident occurred |
| Y-coordinate | Y-coordinate where incident occurred |
| Latitude | Latitude where incident occurred |
| Longitude | Longitude where the incident occurred |
| Location | 3rd party generated spatial column |

**Foursquare API Data:**

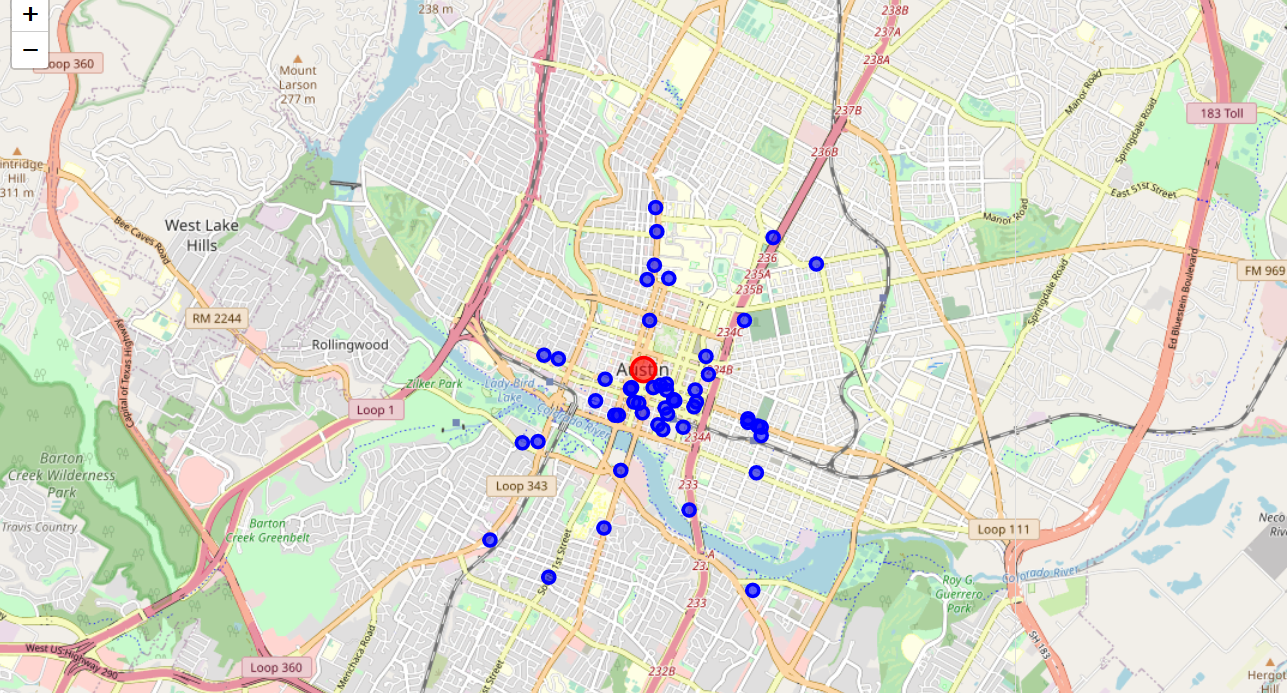
Helped me to id restaurants in the Austin area.

To gain that information, we will use “Foursquare” locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus, and even photos. However, I did not use all of this information. Since it was not a goal. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After, I cleaned the Crime Report Dataset and used a folium map to visualize crime activity. I then connect to the Foursquare API to gather information about venues inside the Austin area.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Name
2. categories address
3. crossStreet
4. lat
5. lng
6. labeledLatLngs
7. distance
8. postalCode
9. country
10. city
11. state
12. country
13. formattedAddress
14. neighborhood
15. id

Map of Austin with a restaurants.  


**Austin Crime Data**

This dataset can be download from the [Austin Data Portal](https://data.austintexas.gov/Public-Safety/Crime-Reports-2018/vmn9-3bvu) and reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Austin in 2018. I choose 2018 dataset since 2019 and 2020 data was not ready. A full description of the data is available on the site.

Not all of the attributes are required so on the following data was removed:

The following columns are removed:

Family Violence

Occurred Date

Occurred Time

Report Date Time

Report Date

Report Time

Census Tract

Clearance Status

Category Description

UCR Category

X COORDINATE

Y COORDINATE

LOCATION

The rest columns are kept and name was changed for easier use. Also, I split the Data column into several columns: year, month, day, hour etc. It helped me for visualization the data. Also, this data was processed further:

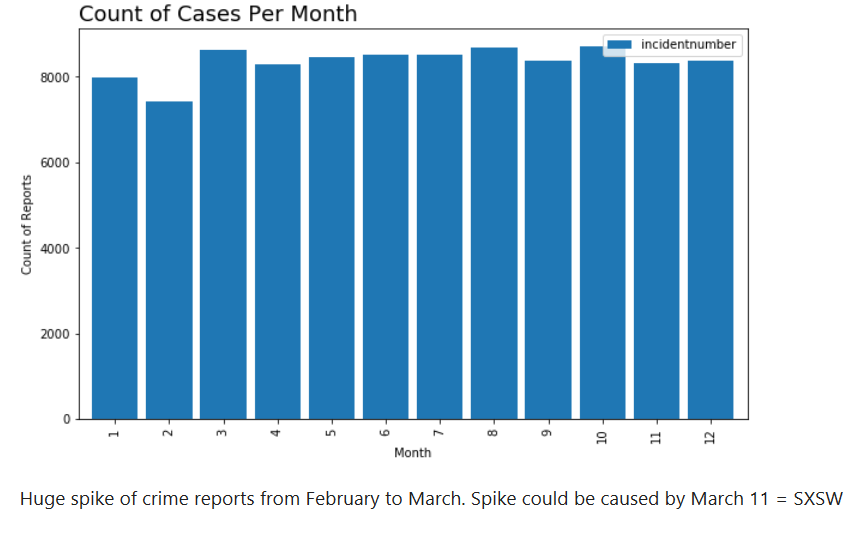
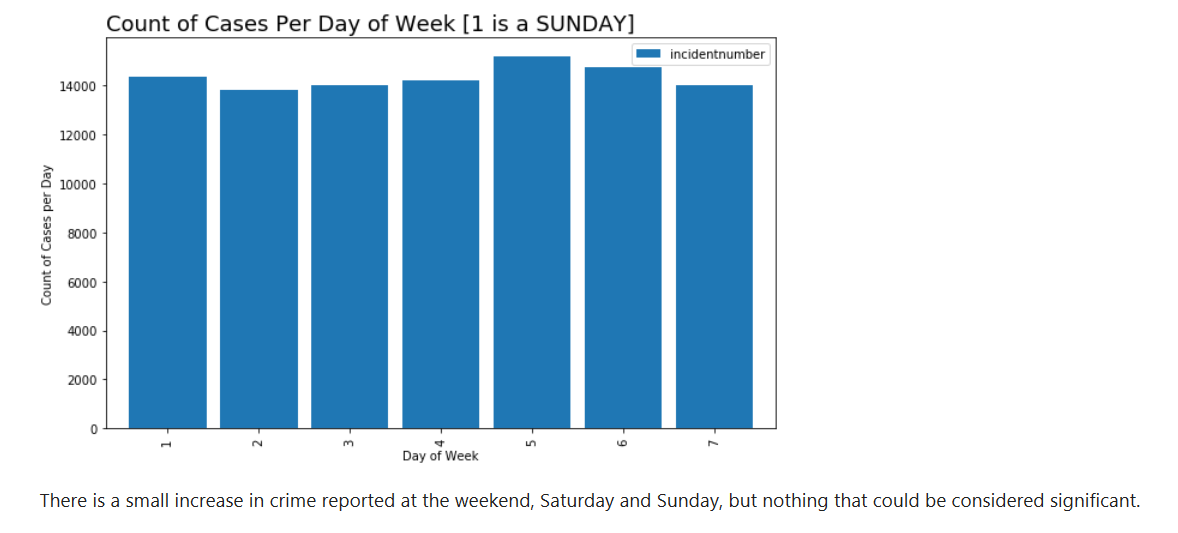
1. Clean up the column names:
   1. Strip leading & trailing whitespace
   2. Replace multiple spaces with a single space
   3. Remove # characters
   4. Replace spaces with \_
   5. Convert to lowercase
2. Change the date of occurance field to a date / time object
3. Add new columns for:
   1. Hour
   2. Day
   3. Month
   4. Year
   5. etc.
4. Split Block into zip\_code and street
5. Verify that all rows have valid data

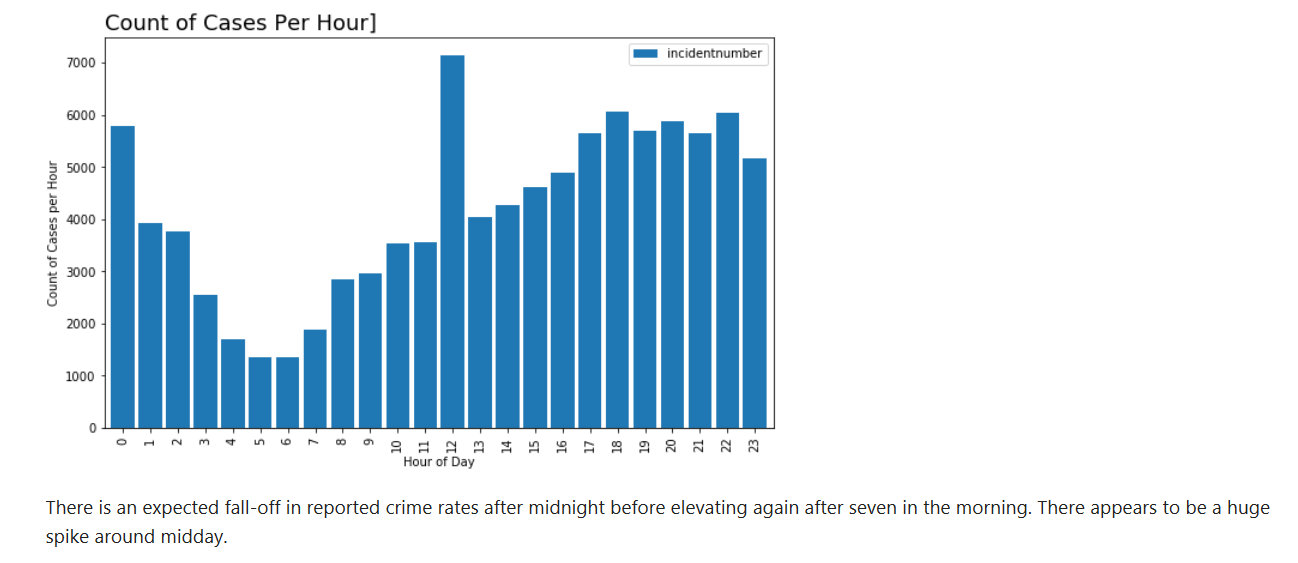
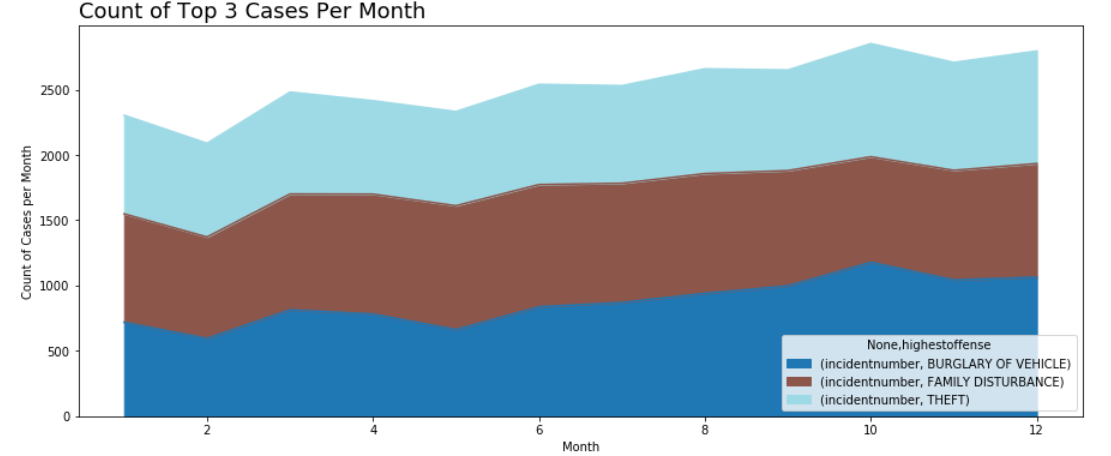


**Data Analysis and Visualisation**

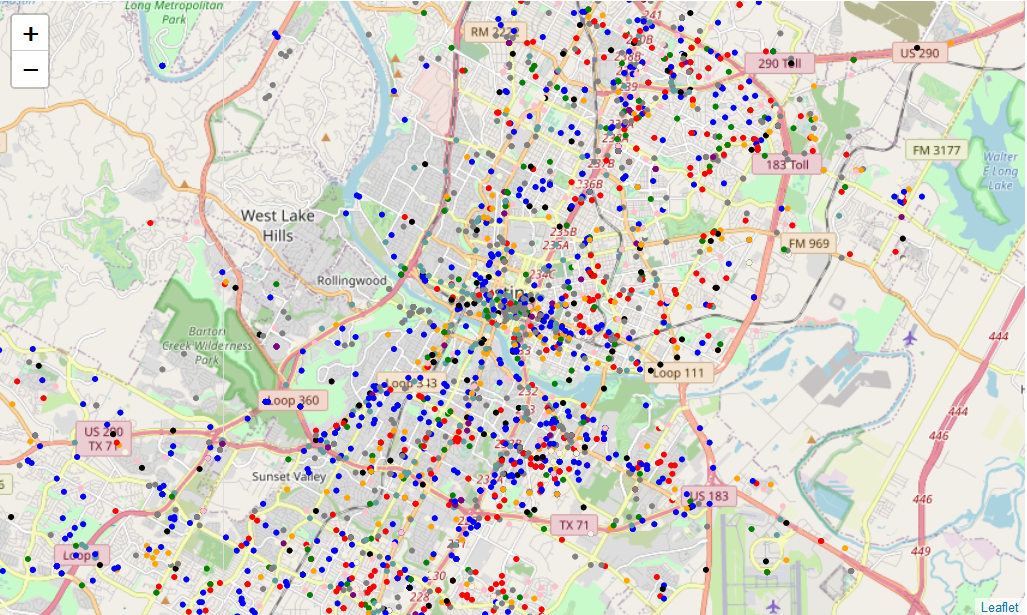
Now let's look at some of the attributes and statistics of the crime dataset. We will start by looking at the top three crimes and a total count for each crime type:

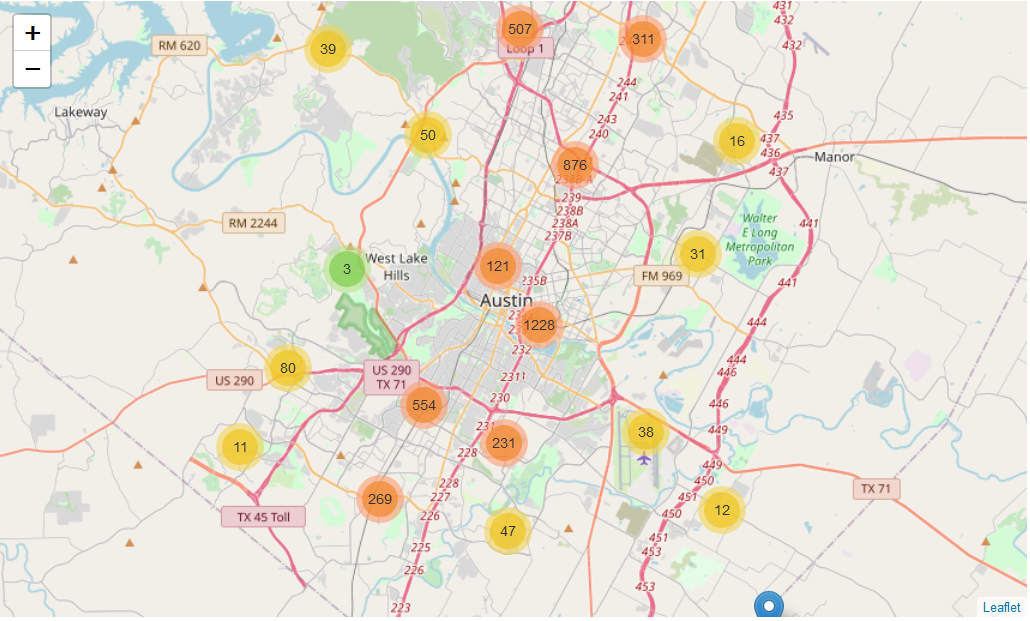
To get a better understanding of the data we will now visualize it. The number of crimes per month, day and hour were calculated:

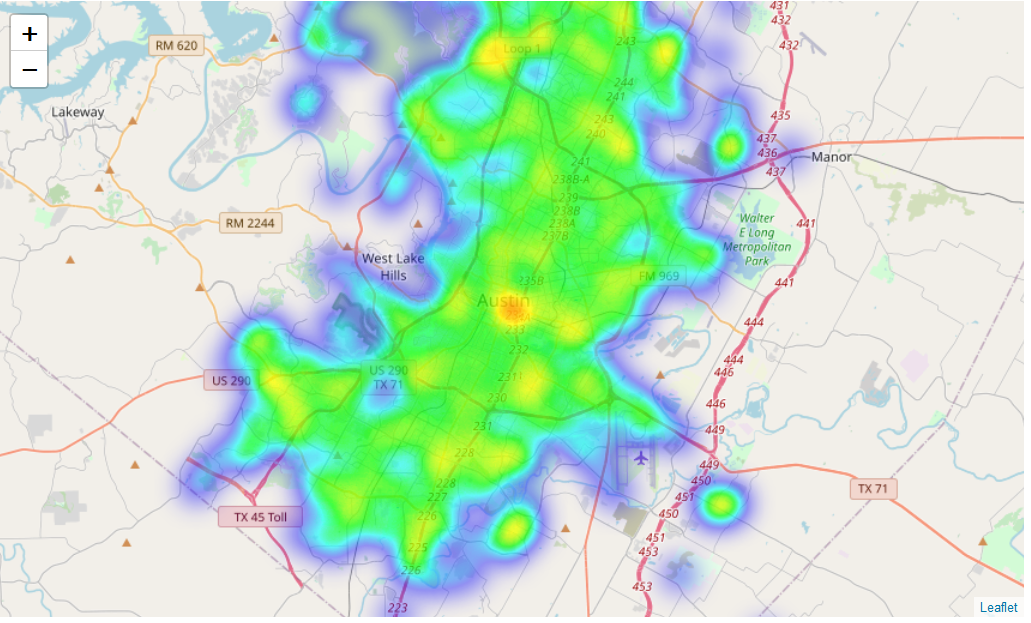


Unsurprisingly there little obvious variation in the number of crimes committed per month other than an apparent drop-off in February. There is a small increase in crime reported at the weekend, Saturday and Sunday, but nothing that could be considered significant. There is an expected fall-off in reported crime rates after midnight before elevating again after seven in the morning. There appears to be a huge spike around midday.

Finally the crimes data for a single month, October, was super-imposed over a map of Austin to visualize the distribution of that data:



The higher frequency of the top two crimes can be easily seen. Red for Family Disturbance and Blue for Burglary of Vehicle. Next, the crimes were clustered:

Several obvious clusters of crime locations were visible, particularly in the center of Austin. Finally, a heat map of the August crimes was created:

This reinforces the cluster chart where it can clearly be seen that the center of Austin and the area around downtown have a high crime rate occurrence. It will be interesting to see later if there is a high probability of crime in these areas if one of the top listed venues are located in these areas.

#### Foursquare API Let’s check the Austin map with restaurants using Foursquare API.

#### There are not a lot of restaurants in Austin. Let’s integrate crime reports from earlier into the venues map:

#### As we can see here:

#### 

#### Area marked with red circle is the best location for the restaurant to build.

Although all of the goals of this project were met, there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some, more work.