New Orleans Incidents and Venues

An Exploration through 911 Calls and Venue Data

Eric Goldberg

January 13, 2020

1. Introduction

New Orleans is a beautiful and complicated city in southern Louisiana with a 300 year history. It's known for jazz, Creole/Cajun cuisine, and Mardi Gras. The city has generally been the most populous of the state, but suffered a large outflux after the 2005 disaster, Hurricane Katrina. For 2019, it's estimated to be home to **423,656 people** in **13 districts** with **72 distinct neighborhoods**, coming out with a population density around 858 persons per square kilometer. However, along with its rich culture, the city has been riddled with crime, poverty, and flooding. The New Orleans Police Department is estimated to be understaffed by 30- 50% depending on the unit (see this article) and as such must stretch their resources by whatever means they have. In stretching those resources, they must be informed on high problem times, areas, and what to expect. This project focuses upon the relation between New Orleans locales and 911 (emergency/crime) incident calls, seeking to find patterns to better distribute local police patrols, response units, and city funds. It leverages data on venues, neighborhood zones, and 911 calls in exploring these patterns. Lastly, traffic and miscellaneous incidents will be *excluded* from the project.

Specifically, we want to answer the **question**:

- Can we use machine learning to cluster New Orleans areas based on 911 incidents and venue types present?

We will also *explore*:

- Incident totals by neighborhood, incident type, and month of occurrence
- Venue types by neighborhood

Which may help the following **stakeholders**:

- Local police and emergency responders
- Government officials allocating city resources, business licenses (if certain venue types trend with crime), or police funding/staffing

Additionally, if the information from this project is combined with housing data, it could be useful for:

- Realestate investors and developers
- Non-locals looking to move to the area

Considering the problem and stakeholders, we can create a map of New Orleans where each neighborhood is clustered according to frequency of 911 incidents and venue types. This may inform where and what type of emergency resource is most needed and around which venue type(s) the incidents occur.

2. Data

2.1 Data Sources

Data for this project comes from four main sources:

1. **Calls for Service 2018**
   This is a log of all 911 calls in New Orleans in 2018. It contains 175 different types of incidents ranging from traffic violations to homicide. Each incident's time, location, priority, and a range of other factors are listed. For this project, incidents related to neighborhood safety and crime will be used. Those such as traffic, medical, and miscellaneous will be *excluded*.
2. **Foursquare API**
   Venues and their types can be retrieved based on geolocations from a given neighborhood's centroid (listed on Wikipedia).
3. **Wikipedia list of Neighborhoods**
   A table of New Orleans neighborhoods and their centroid coordinates can be scraped from wikipedia using pandas.
4. **Neighborhood boundaries**
   This is a list of New Orleans neighborhoods and their geographic boundaries as defined by the US census. Incidents from the calls of service will be categorized into neighborhoods based on geolocation using this data.

2.2 Data Cleaning and Selection

Data from Calls for Service 2018 is a huge table consisting of over 460,000 entries. Many fields contain missing values or irrelevant content for this study. In cleaning the table, several things must be done. First, essential columns must be identified and erroneous ones dropped. Second, rows with an incident type not pertaining to neighborhood safety or crime are to be removed. Third, missing values need to be dropped or filled. Finally, naming schemes of rows are revised.

Scanning the columns in the Calls for Service data set, we can see an incident's police log number, type, the police code for the type, its initial type and code, 3 different time stamps for time created, arrived, and closed; the disposition (what happened when the responders arrived), and several other fields. For this study, the type and location fields are essential. We will also keep the block address, zip, beat, and police district for additional location information. One time stamp (TimeCreate) and the type code (Type_) are retained for reference and later analysis.

Summary of Columns	Column Names	Total
Columns kept	Type_, TypeText, TimeCreate, Beat, BLOCK_ADDRESS, Zip, PoliceDistrict, Location	8
Columns discarded	NOPD_Item, Priority, InitialType, InitialTypeText, InitialPriority, MapX, MapY, TimeDispatch, TimeArrive, TimeClosed, Disposition, DispositionText, SelfInitiated	13

The next big item for this data set is retaining rows with a relevant type field. As stated before, we want rows that are related to neighborhood safety and crime. Some examples of types to be discarded are traffic related incidents, MEDICAL, and AREA CHECK. For a full list, check the jupyter book on github. Sample below:

Type Removed	Note
'ABANDONED BOAT'	--
'ABANDONED VEHICLE'	--
'DIRECTED TRAFFIC ENFORCEMENT'	--
'DISTURBANCE (OTHER)'	May be safety related, but not enough info is available
'ELECTRONIC MONITORING'	--
'WARR STOP WITH RELEASE'	--
TOTAL REMOVED	**40**
TOTAL RETAINED	**135**

It could be argued that other types not removed should be or some removed should be retained. However, without consulting experts the above items are what we will discard.

After filtering the necessary columns and rows, we drop all rows with nan values in the Location field. Rows marked with (0, 0) are also dropped. Lastly, columns are renamed as follows:

Initial Name	Final Name
Type_	Code
TypeText	Type
TimeCreate	Time
Beat	Beat
BLOCK_ADDRESS	Address
Zip	Zip
PoliceDistrict	District
Location	Coordinates

Data from both Foursquare and Wikipedia is cleaned and formatted as it is retrieved. Foursquare venues are drawn from within a 750 m radius from each neighborhoods centroid. Their type and location will also be retrieved. As for neighborhood boundaries, the boundary information is stored as a MULTIPOLYGON type from a different kernel. We will change this into a POLYGON type from the shapely library in python. Later, we can pass the coordinates from Calls for Service as "points" to see if they are contained in the POLYGON, thus assigning their neighborhood. The necessary columns from this data set are "the_geom" and "Neighborhood". "the_geom" is renamed as "Poly". Last, the neighborhood listings in the Wikipedia list and Neighborhood boundaries are reconciled--the only difference being an additional space in "St. Anthony" in the Neighborhood Boundaries data set.

3. Methodology

As previously mentioned, the focus of this project is to explore relationships between venues and 911 incident occurrence in New Orleans neighborhoods, with the ultimate goal of providing useful information for government resource allocation. Towards that end, an exploratory analysis will be conducted using *descriptive statistics* to show:

1. Top Incident Types (what incidents occur most often and the neighborhoods where they most occur)
2. Frequency Distribution of Incidents (in what quantity do incident types usually occur)
3. Top Neighborhoods of Incident Occurrence (where incidents occur most often and what the types of incidents occur in these neighborhoods)
4. Frequency Distribution of Total Incident Occurrence by Neighborhood (how many incidents usually occur in neighborhoods)
5. Choropleth map of incident occurrence in New Orleans by neighborhood
6. Time Series of Incidents by Month
7. Top Venue Types and Frequencies

K-means machine learning will be used 3 separate times to cluster neighborhoods based on incident type frequency, venue type frequency, and then incident type frequency and venue type frequency together. K-means was chosen based on its efficiency and speed for working with a large data set. The elbow method is used to choose the best k for each clustering, checking both distortion and inertia. Neighborhoods will be displayed upon a folium map with their K-means cluster labels.

Finally, based on k-means cluster labels for incident type and venue type frequencies, connections will be examined and discussed.

3.1 Exploratory Analysis

In order to complete our exploratory analysis as mentioned in the methodology, we'll need to add neighborhood labels to the incidents data in civildf. We'll drop any rows that do not receive an appropriate neighborhood label. Afterward, we can explore incidents by neighborhood and top incidents overall. We will also convert our Time column information to a timestamp for exploring monthly incident occurrence by neighborhood and type. A completed csv is available in the github project folder.

Snippet of prepared data in civildf:

	Code	Type	Time	Beat	Address	Zip	District	Coordinates	Neighborhood
0	107	SUSPICIOUS PERSON	2018-02-08 16:55:15	2K03	018XX Cambronne St	70118.0	2	[29.9543012, -90.12741579]	LEONIDAS
1	107	SUSPICIOUS PERSON	2018-02-08 17:20:45	5I04	Andry St & N Claiborne Ave	70117.0	5	[29.9668437, -90.0168289]	LOWER NINTH WARD
2	62B	BUSINESS BURGLARY	2018-02-08 05:19:41	7L01	046XX Michoud Blvd	70129.0	7	[30.03203358, -89.92870826]	VILLAGE DE LEST
3	29	DEATH	2018-02-08 10:16:26	6P01	014XX General Taylor St	70115.0	6	[29.92592877, -90.09646117]	TOURO
4	67	THEFT	2018-02-08 10:54:10	5L01	023XX N Tonti St	70117.0	5	[29.98070722, -90.05459674]	ST. ROCH

3.1.1 Top Incident Types

To get our top incident types, we can group the incidents from civildf by type and get the type count. Following that, we can represent this data in a bar chart showing the top 20 incident types. Afterwards, we can explore where each of these incident types occur most.

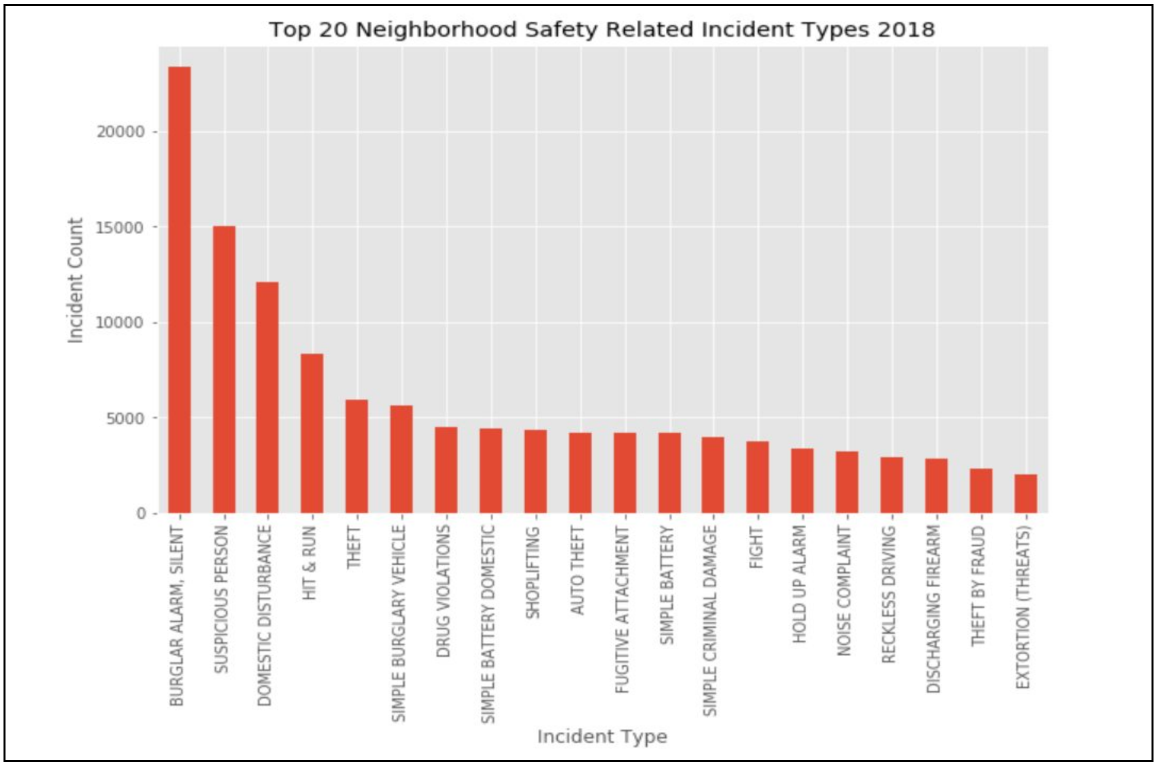


Figure 3.1.1a Top 20 Neighborhood Safety Related Incident Types 2018

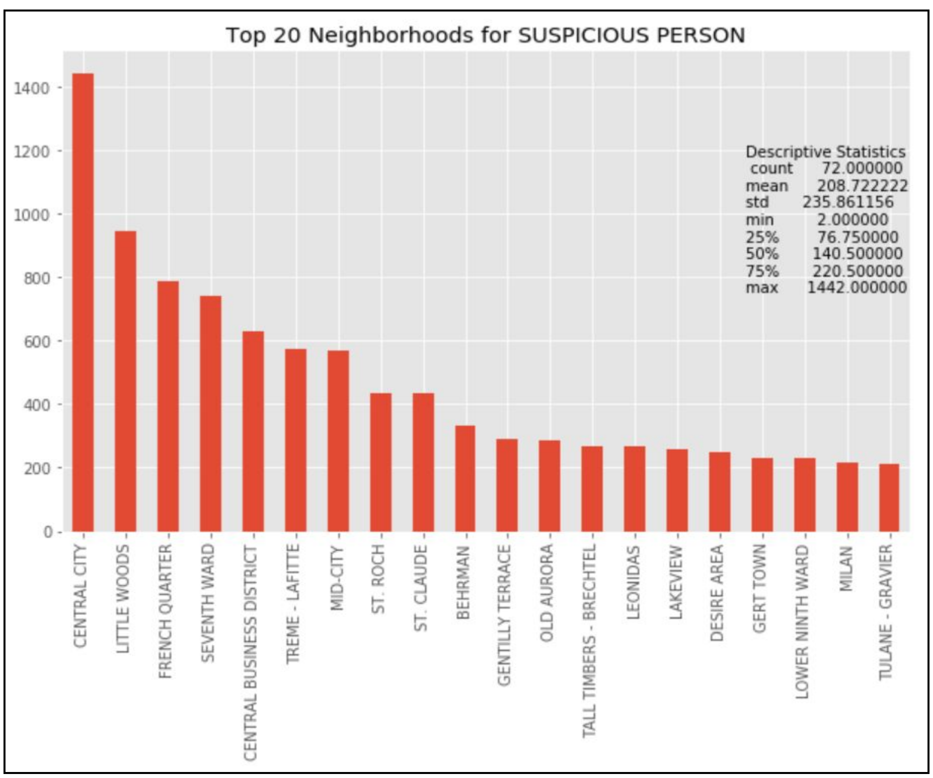


Figure 3.1.1b Sample of Top 20 Neighborhoods per Incident Type

3.1.2 Frequency Distribution of Incident Totals by Types

To see the distribution of incident type occurrence, we'll plot a histogram from the typetotals data frame.

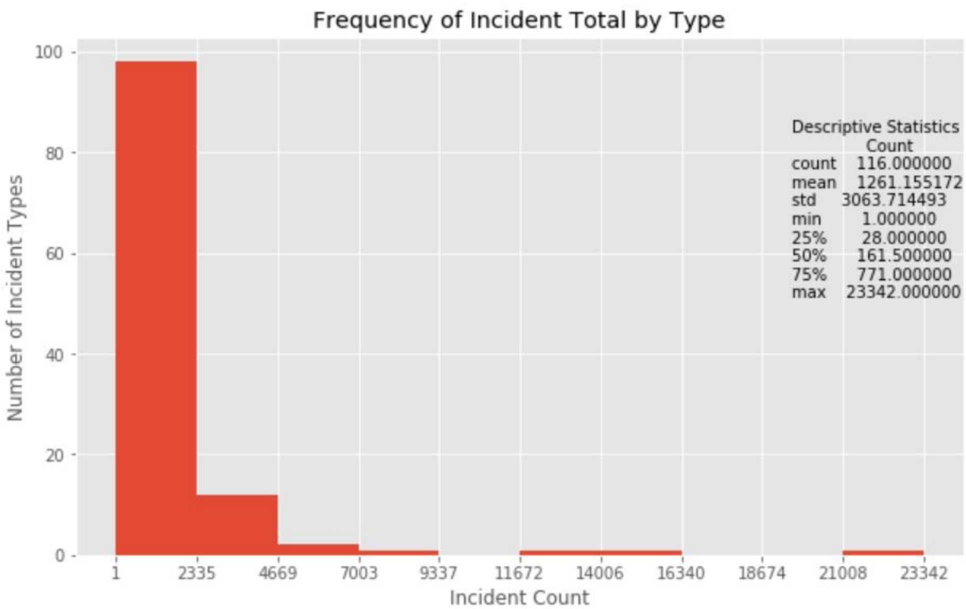


Figure 3.1.2 Frequency of Incident Totals by Type

3.1.3 Top Neighborhoods of Incident Occurrence

To get our top neighborhoods of incident occurrence, we can group the incidents from civildf by neighborhood and get the total for each. Following that, we can represent this data in a bar chart showing the top 20 neighborhoods of occurrence. We can then explore what incident types occur most in each of these neighborhoods.

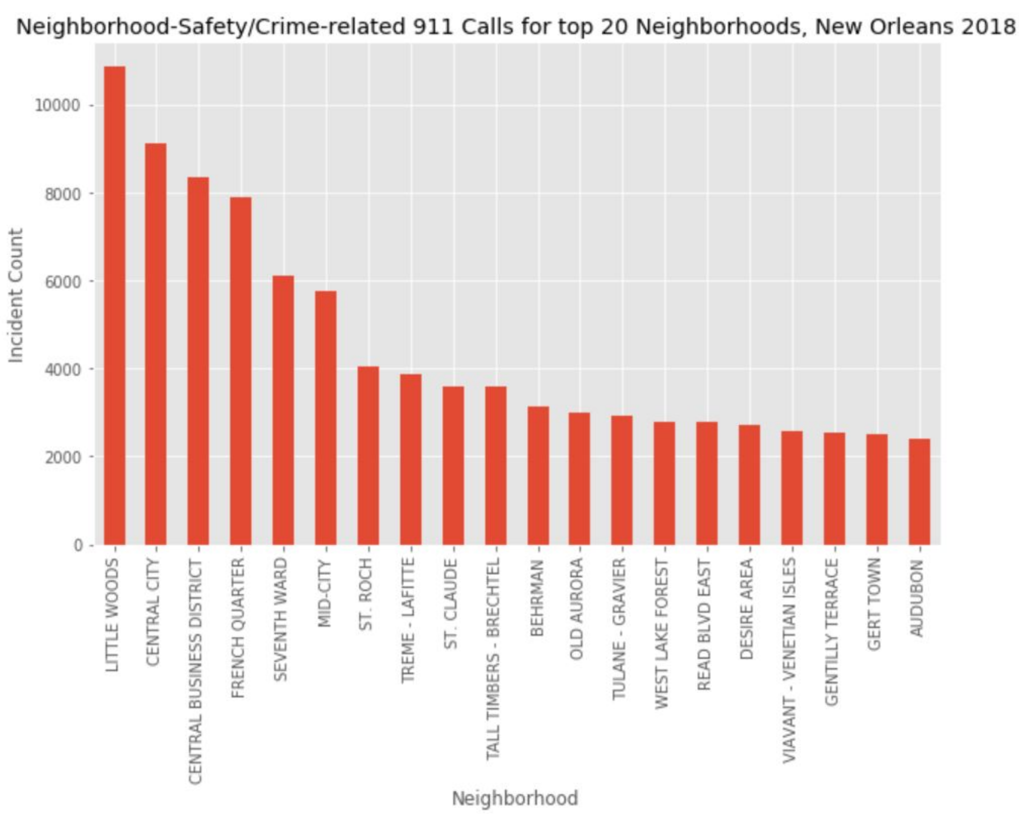


Figure 3.1.3a Top Neighborhoods of Incident occurrence

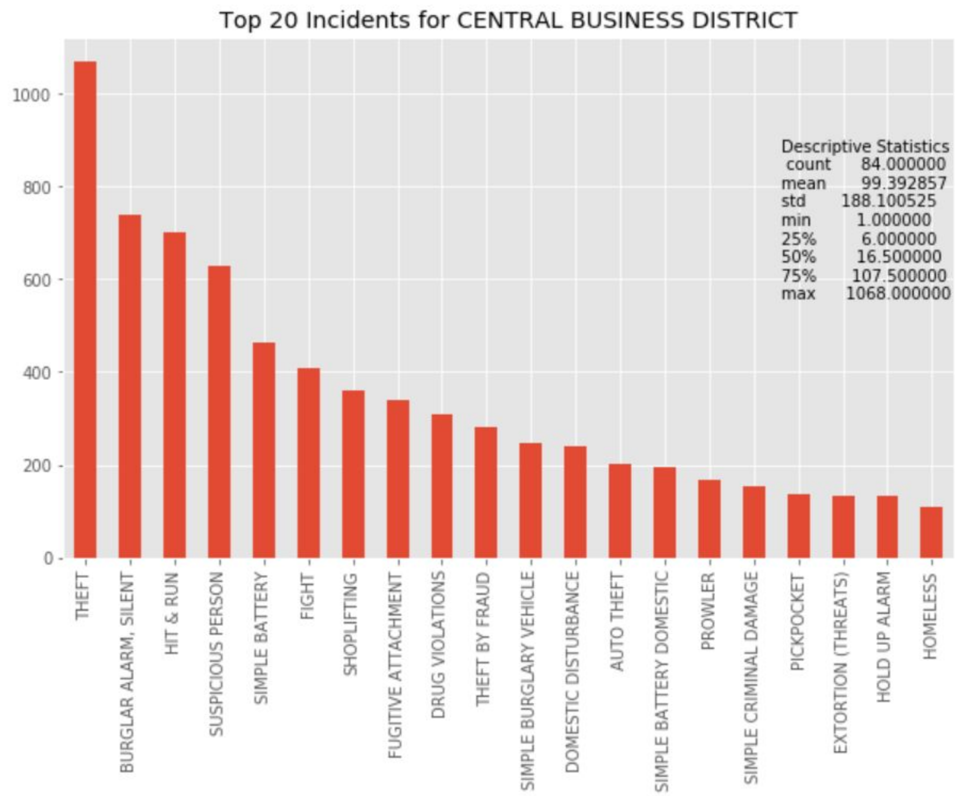


Figure 3.1.3b Sample of Top 20 Incidents by Neighborhood (CBD)

3.1.4 Frequency Distribution of Incident Count by Neighborhoods

---

How many incidents usually occur in neighborhoods. Like above, we'll plot a histogram from the hoodtotals data frame.

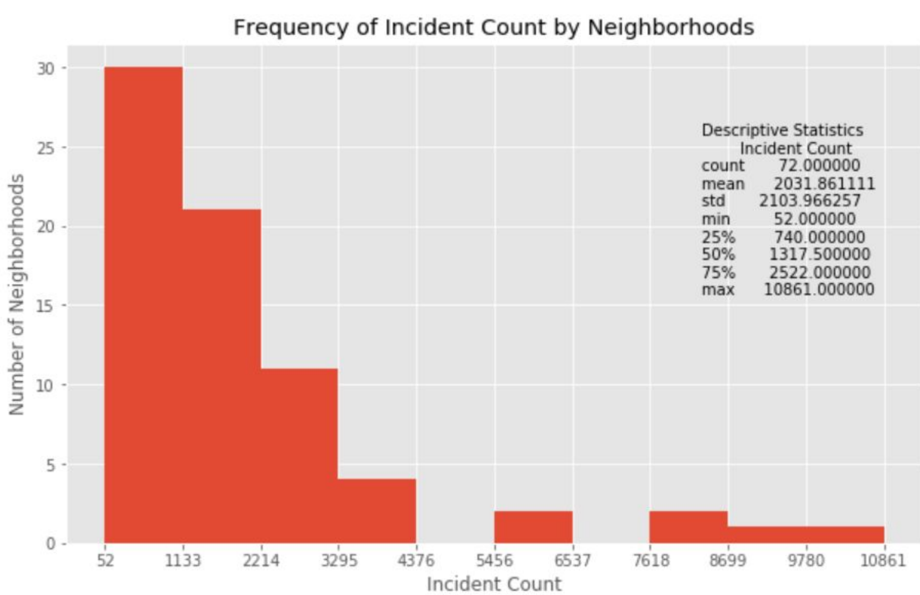


Figure 3.1.4 Frequency of Total Incidents by Neighborhood Count

3.1.5 Choropleth Map of Incidents in New Orleans by Neighborhood

We'll need to use a geojson file of our neighborhood boundary information to generate a choropleth map. This file can be found here. A corrected version is also available in the github project folder. Next, we will join our neighborhoods data frame with the hoodtotals data frame to get the neighborhood centroid coordinates.

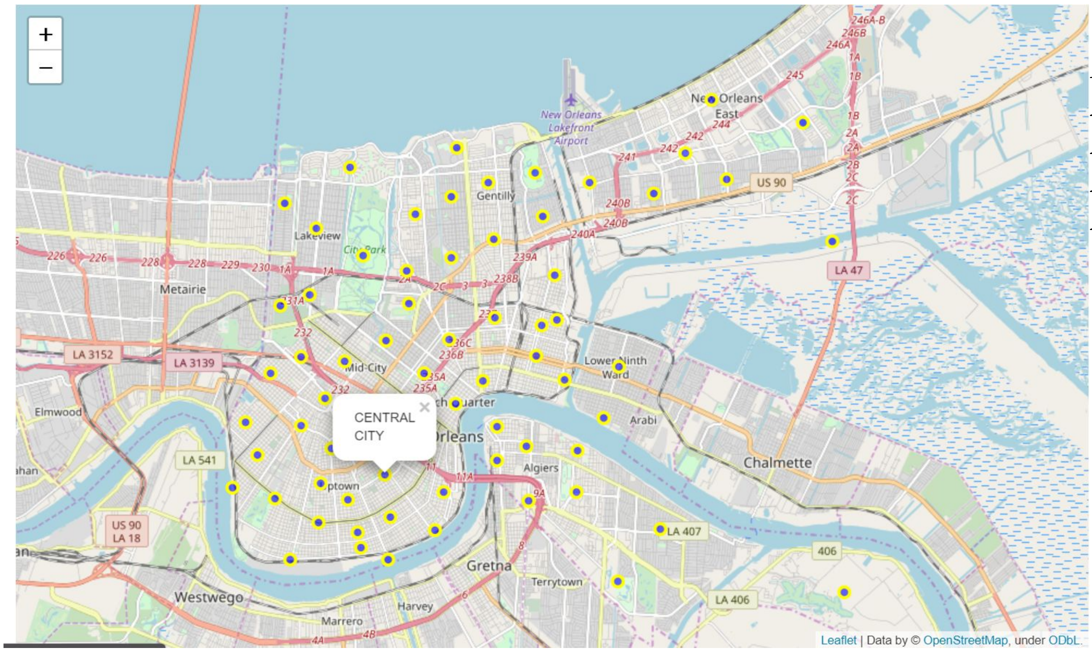


Figure 3.1.5a New Orleans Map with Neighborhood Markers

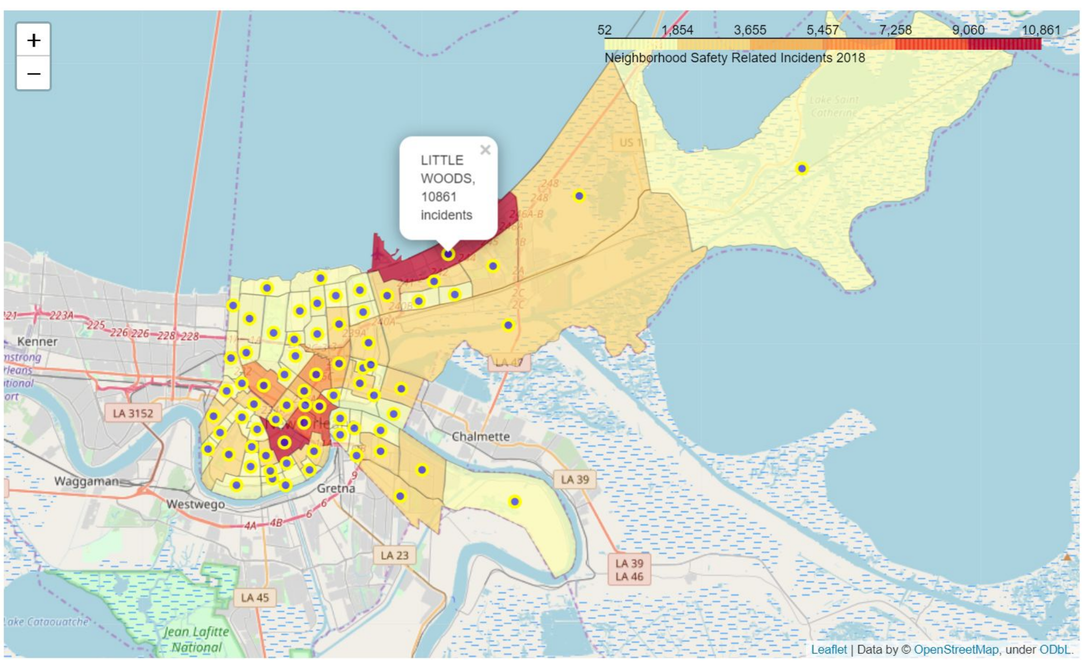


Figure 3.1.5b New Orleans Choropleth Map with Markers

We can also explore this data by quartile or by excluding certain quartiles. By excluding the top quartile, we may produce a map showing more desirable neighborhoods to live in for those moving to or around the city. The top quartile neighborhoods are blacked out in this map.

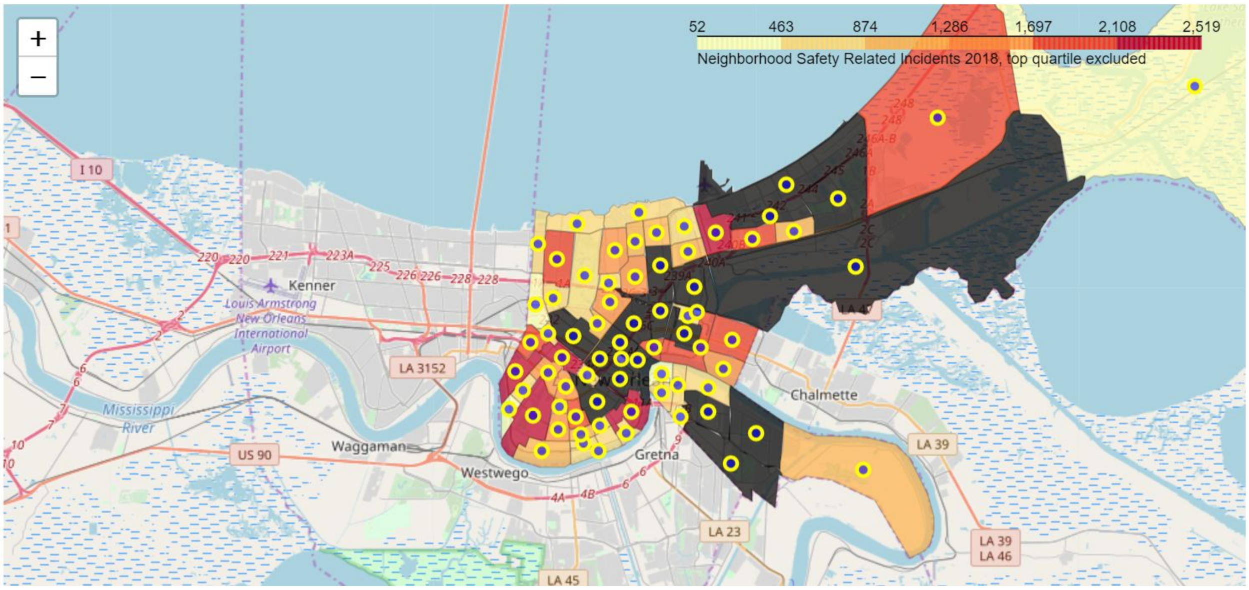


Figure 3.1.5c Choropleth Map of New Orleans Incidents, Top Quartile Removed

3.1.6 Time Series of Incidents by Month

For checking incidents by month, our final data frame needs to have the months as the index and the sum totals by neighborhood or incident type as columns. We will make a dataframe for each of these types using the pandas groupby on our timestamp data.

Sample of processed data frames:

Neighborhood	LITTLE WOODS	CENTRAL CITY	CENTRAL BUSINESS DISTRICT	FRENCH QUARTER	SEVENTH WARD	MID-CITY	ST. ROCH	TREME - LAFITTE	ST. CLAUDE	TALL TIMBERS - BRECHTEL	...	NAVARRE
Month												
1	798.0	736.0	614.0	646.0	465.0	495.0	319.0	330.0	268.0	302.0	...	34.0
2	838.0	773.0	776.0	776.0	463.0	455.0	347.0	378.0	304.0	274.0	...	45.0
3	887.0	717.0	634.0	634.0	562.0	504.0	329.0	348.0	274.0	304.0	...	36.0
4	912.0	702.0	636.0	654.0	531.0	474.0	375.0	323.0	329.0	301.0	...	48.0
5	995.0	861.0	713.0	617.0	569.0	497.0	374.0	329.0	336.0	325.0	...	32.0

5 rows × 73 columns

Type	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	DOMESTIC DISTURBANCE	HIT & RUN	THEFT	SIMPLE BURGLARY VEHICLE	DRUG VIOLATIONS	SIMPLE BATTERY DOMESTIC	SHOPLIFTING	AUTO THEFT
Month										
1	2297.0	1111.0	976.0	601.0	392.0	321.0	280.0	368.0	295.0	292.0
2	1785.0	1222.0	985.0	697.0	433.0	414.0	328.0	406.0	278.0	350.0
3	2090.0	1250.0	1027.0	771.0	502.0	377.0	416.0	381.0	340.0	314.0
4	1952.0	1231.0	1144.0	719.0	515.0	390.0	368.0	419.0	330.0	338.0
5	2204.0	1312.0	1144.0	765.0	516.0	418.0	438.0	416.0	346.0	387.0

5 rows × 117 columns

First, let's examine our incident totals by month.

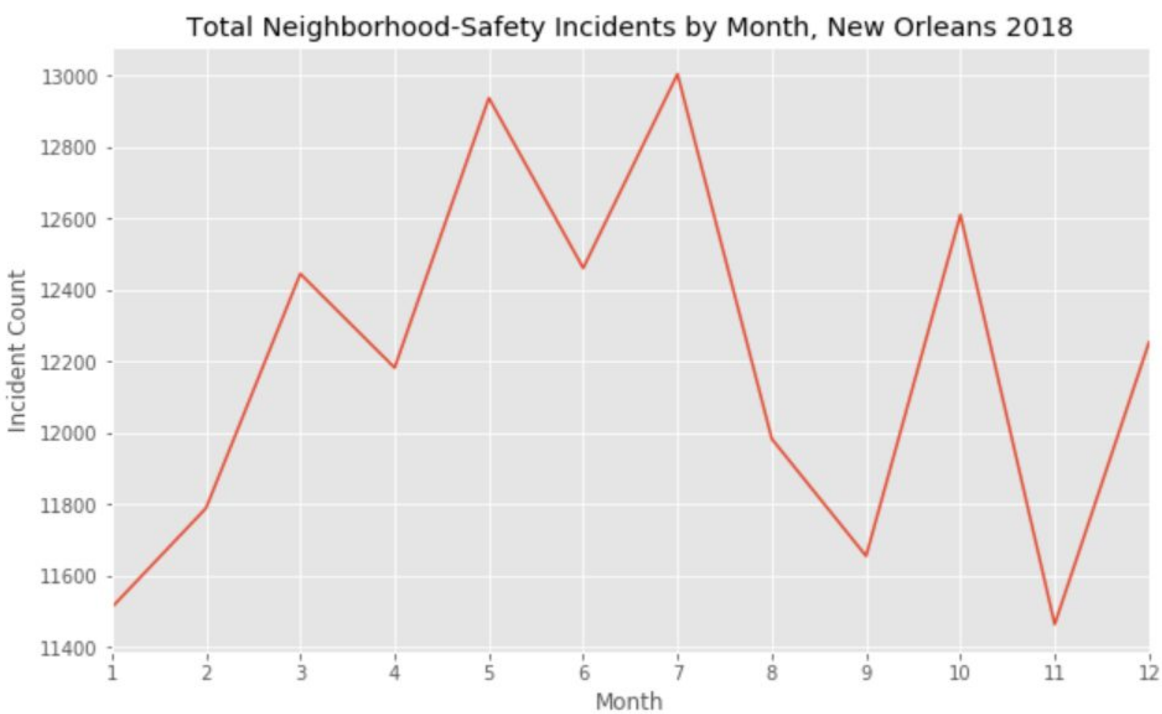


Figure 3.1.6a Total Incident Count by Month

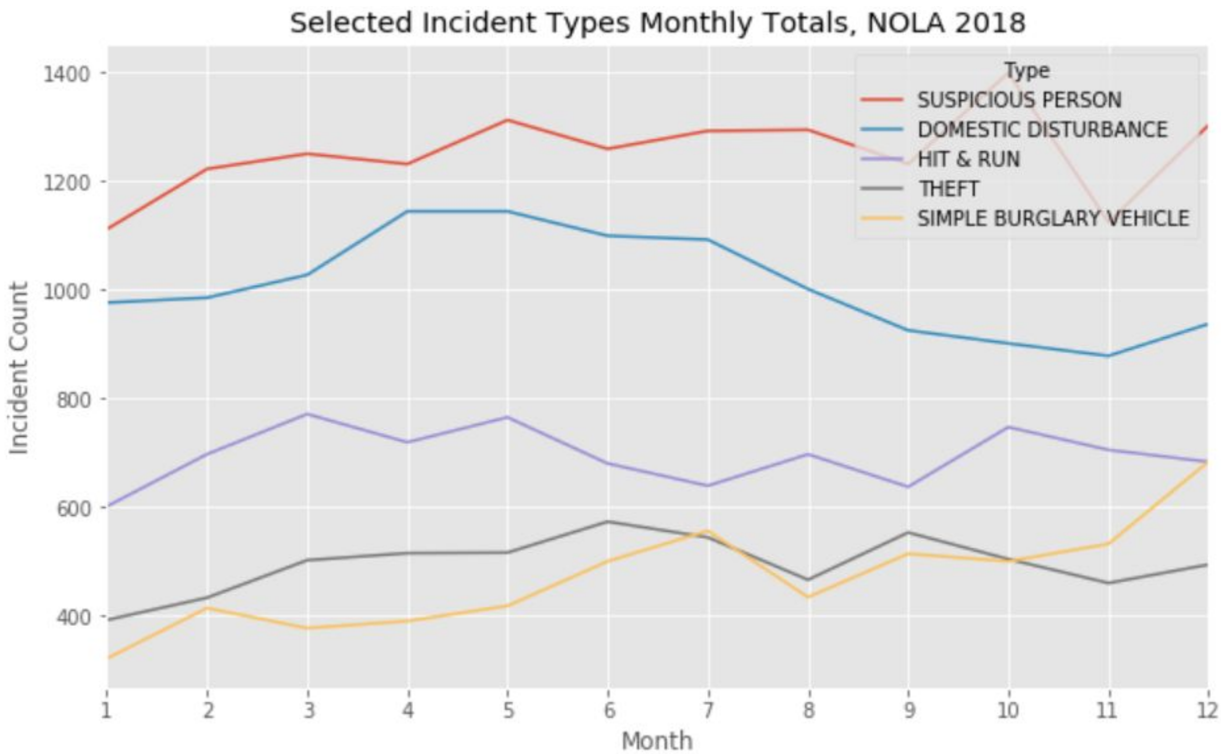


Figure 3.1.6b Selected Incident Types Time Series

Next, we'll plot a segment of our neighborhood incident totals over the year. Since 72 neighborhoods is too busy for one graph, let's look at the median neighborhoods of each quartile of neighborhoods by total incident count.

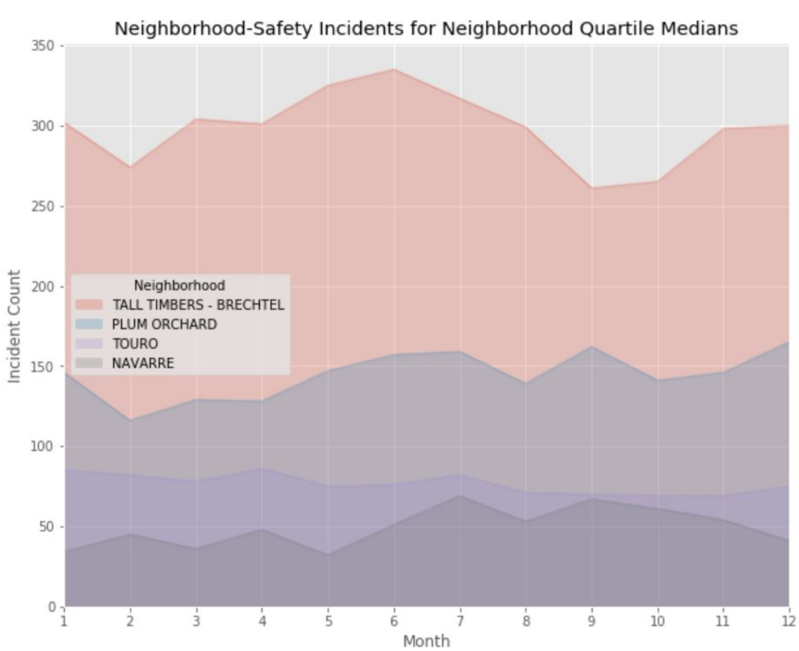


Figure 3.1.6c Timeseries of Incidents for Quartile Median Neighborhoods

3.1.7 Top Venue Types and Frequencies

We'll now use Foursquare API to retrieve the top 100 venues within a 750 m radius of each neighborhood centroid from the neighborhoods dataframe. Please note that this will not retrieve all venues and it will not be exhaustive of a neighborhood's boundaries due to the methods available with foursquare and their data that is freely available. Afterward, we'll see what the top venue types retrieved for New Orleans are.

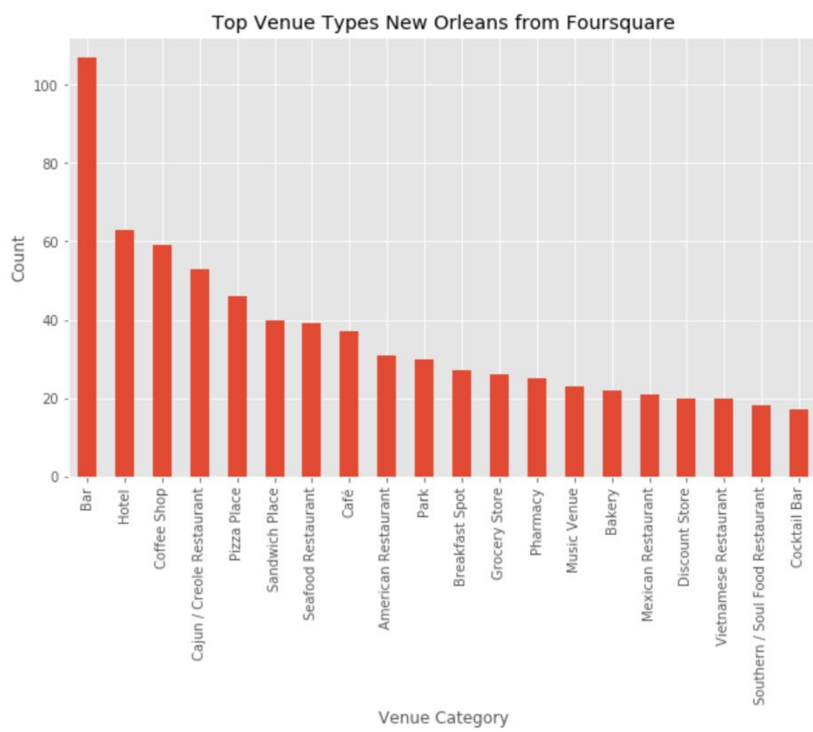
Please note: Foursquare was unable to return any venues for 2 neighborhoods--Village de Lest and Viavant - Venetian Isles. These will be excluded for k-means venue analysis.

Foursquare reached the limit set for 4 neighborhoods: Marigny, Iberville, Central Business District, and French Quarter. For others, it only retrieved a handful of venues. Lake Catherine only has 1.

Sample of Venue types retrieved:

	Count
Venue Category	
Bar	107
Hotel	63
Coffee Shop	59
Cajun / Creole Restaurant	53
Pizza Place	46

Figure 3.1.7a Top Venue Types in New Orleans, Foursquare



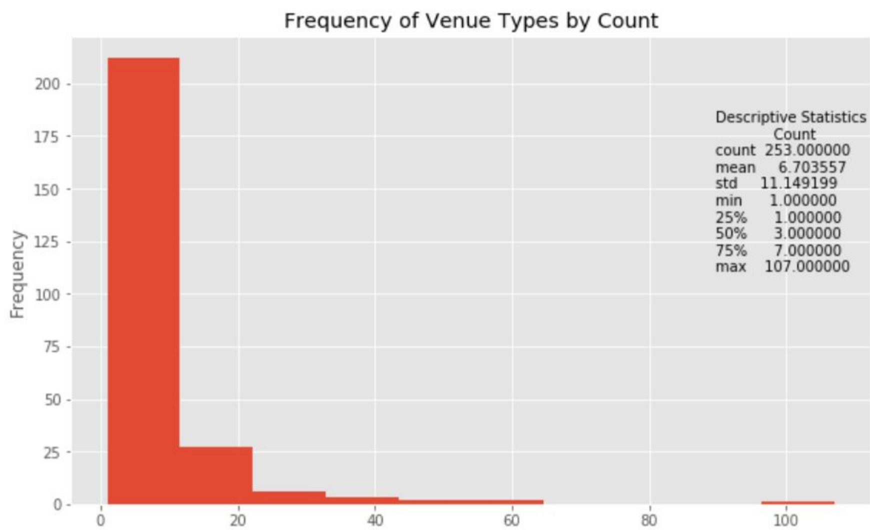


Figure 3.1.7b Frequency of Venue Types by Count

3.2 K-Means

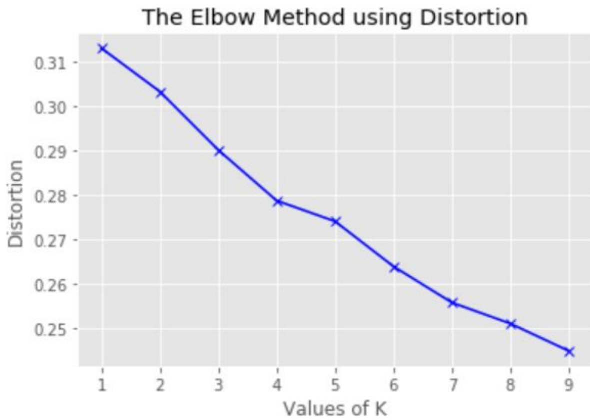
Data must first be prepared with neighborhoods in the index and venue type/incident type in the column fields with values being normalized incident frequency/venue frequency before we can complete our k-means clustering. We'll use one-hot encoding on the nola_venues and civildf data frames then group by neighborhood while using the mean function to achieve our target data structure.

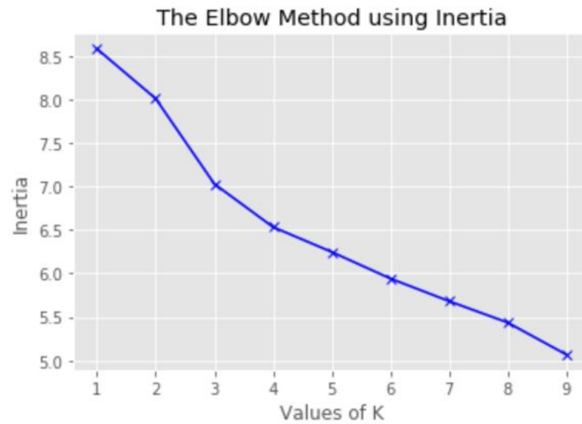
Neighborhoods with very little data will cause trouble for K-means, so we'll drop those with 3 or fewer venues before starting.

Example of prepared venue dataframe:

	Neighborhoods	Accessories Store	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auto Garage	Automotive Shop	...	Video Store	Vietnamese Restaurant
0	ALGIERS POINT	0.0	0.000000	0.0	0.00	0.000	0.04	0.000000	0.0	0.0	...	0.0	0.0
1	AUDUBON	0.0	0.000000	0.0	0.00	0.000	0.00	0.000000	0.0	0.0	...	0.0	0.0
2	B. W. COOPER	0.0	0.000000	0.0	0.05	0.000	0.00	0.000000	0.0	0.0	...	0.0	0.0
3	BAYOU ST. JOHN	0.0	0.023256	0.0	0.00	0.000	0.00	0.023256	0.0	0.0	...	0.0	0.0
4	BEHRMAN	0.0	0.000000	0.0	0.00	0.125	0.00	0.000000	0.0	0.0	...	0.0	0.0

Next, K-means must be run for several values of k and chosen using the elbow method. Here are the graphs for distortion and inertia on venue data:





K=4 was chosen for the venue data.

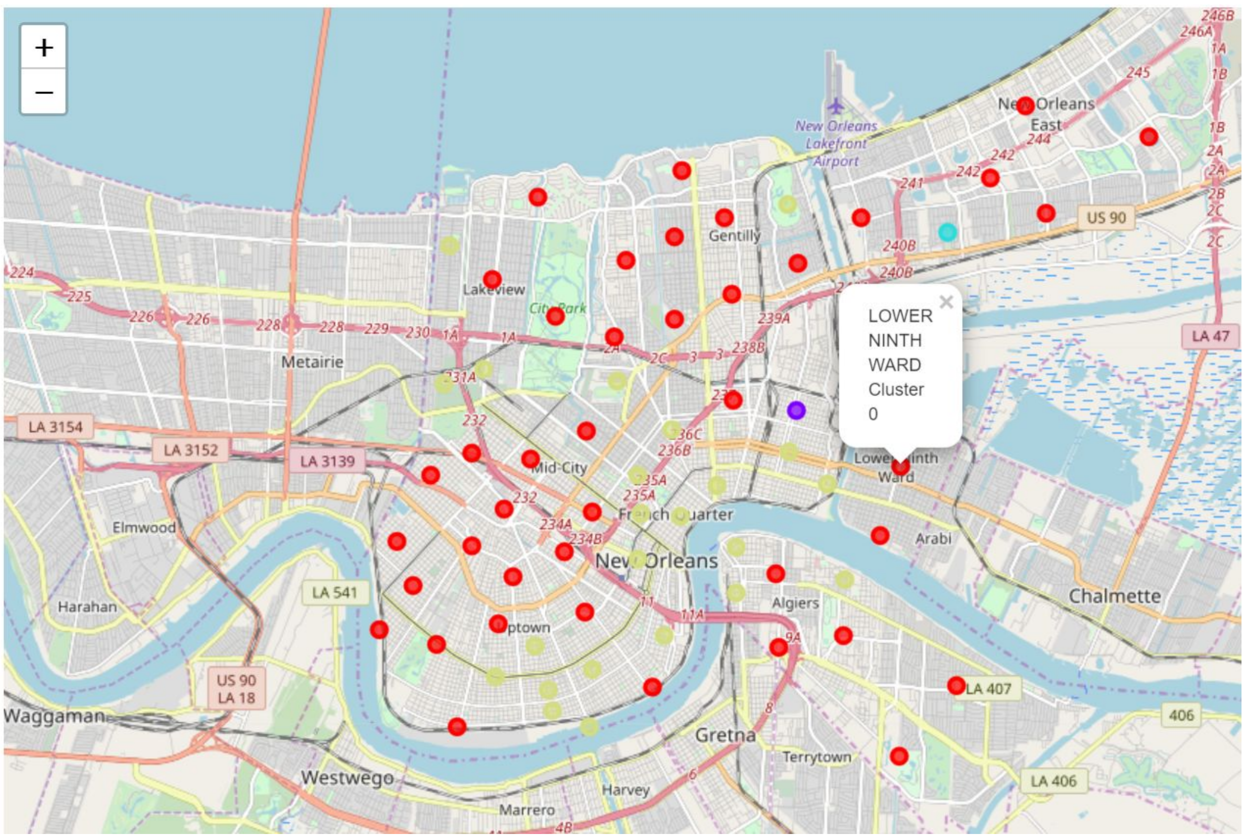


Figure 3.2.1 Map of New Orleans Neighborhoods with Colored Venue Cluster labels

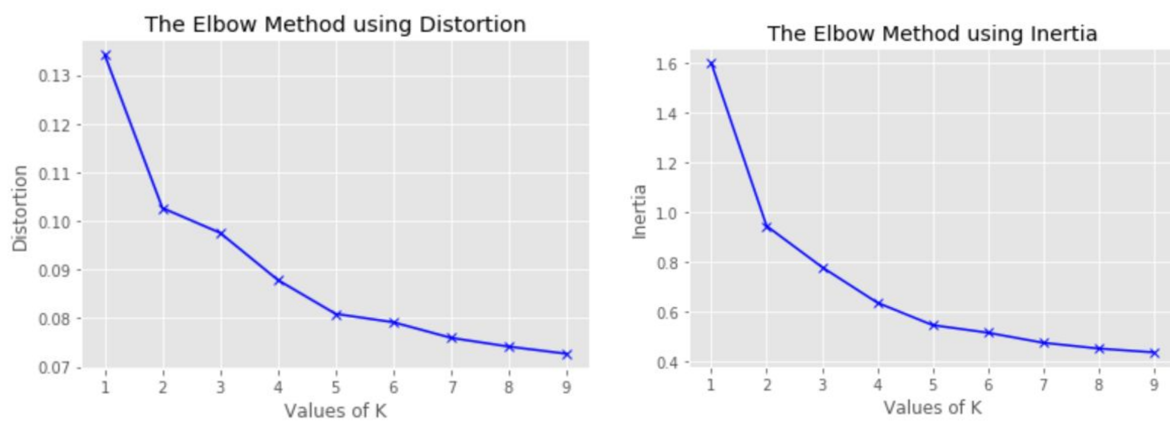
We’ll Repeat the process for incidents.

Sample of prepared dataframe for incident type frequency:

	Neighborhoods	AGGRAVATED ARSON	AGGRAVATED ASSAULT	AGGRAVATED ASSAULT DOMESTIC	AGGRAVATED BATTERY	AGGRAVATED BATTERY BY CUTTING	AGGRAVATED BATTERY BY KNIFE	AGGRAVATED BATTERY BY SHOOTING
0	LEONIDAS	0	0	0	0	0	0	0
1	LOWER NINTH WARD	0	0	0	0	0	0	0
2	VILLAGE DE LEST	0	0	0	0	0	0	0
3	TOURO	0	0	0	0	0	0	0
4	ST. ROCH	0	0	0	0	0	0	0

5 rows × 117 columns

Elbow method for Incident type frequency k-means:



K=5 was chosen for incident type frequency in neighborhoods for k-means.

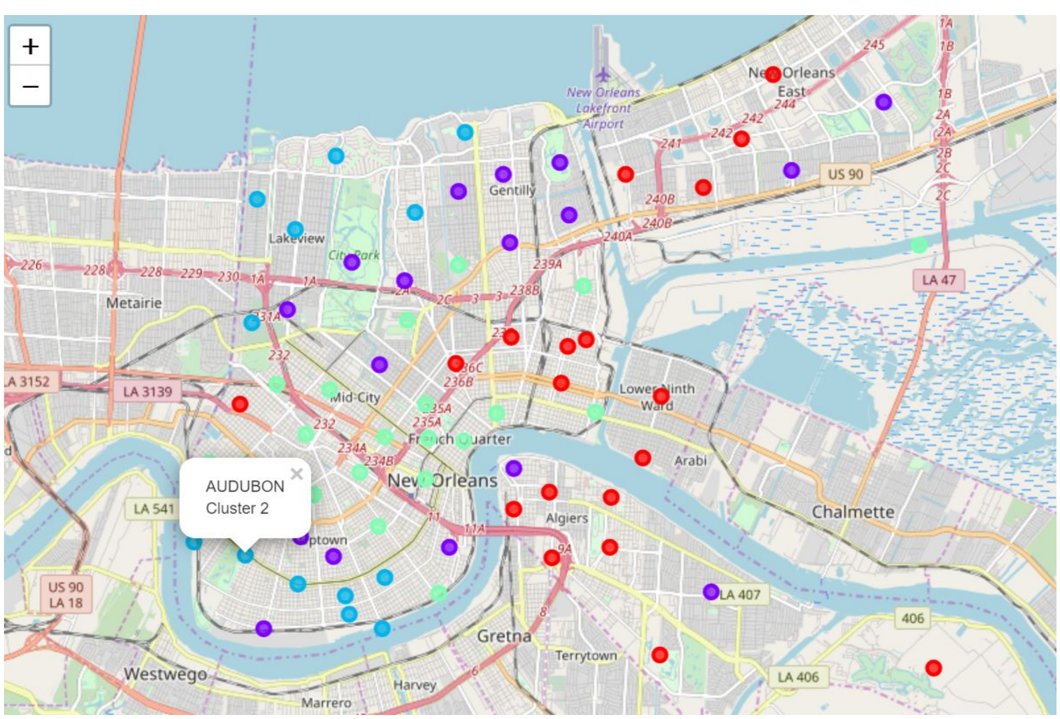


Figure 3.2.2 Map of New Orleans Neighborhoods with Colored Incident Frequency Clusters

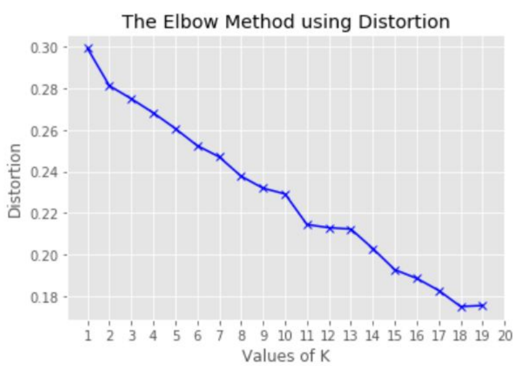
Finally, venues and incident type frequencies were combined into one table for a combined k-means clustering.

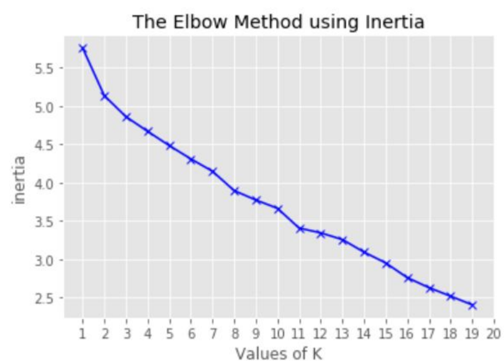
Sample of combined table:

	Neighborhoods	AGGRAVATED ARSON	AGGRAVATED ASSAULT	AGGRAVATED ASSAULT DOMESTIC	AGGRAVATED BATTERY	AGGRAVATED BATTERY BY CUTTING	AGGRAVATED BATTERY BY KNIFE	AGGRAVATED BATTERY BY SHOOTING	AGGRAVATED BATTERY DOMESTIC
0	ALGIERS POINT	0.0	0.002326	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	AUDUBON	0.0	0.000833	0.000000	0.000416	0.000000	0.000000	0.000000	0.000416
2	B. W. COOPER	0.0	0.013393	0.000000	0.002976	0.002976	0.001488	0.002976	0.000000
3	BAYOU ST. JOHN	0.0	0.000000	0.002146	0.002146	0.001073	0.002146	0.001073	0.002146
4	BEHRMAN	0.0	0.010804	0.004449	0.000636	0.000636	0.002860	0.004131	0.000636

5 rows × 370 columns

Elbow Method:





K=8 was chosen after examining distortion and inertia.

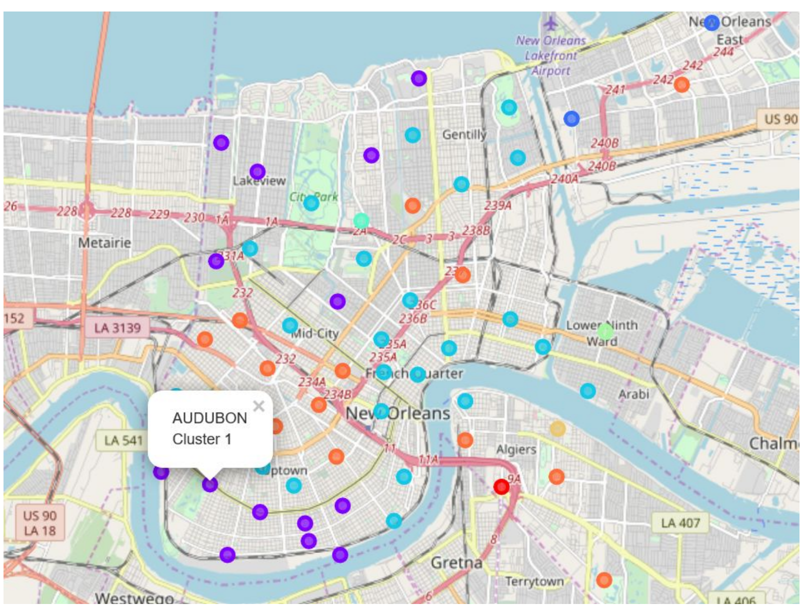


Figure 3.2.3 Map of New Orleans Neighborhoods with Colored Combined Frequency Clusters

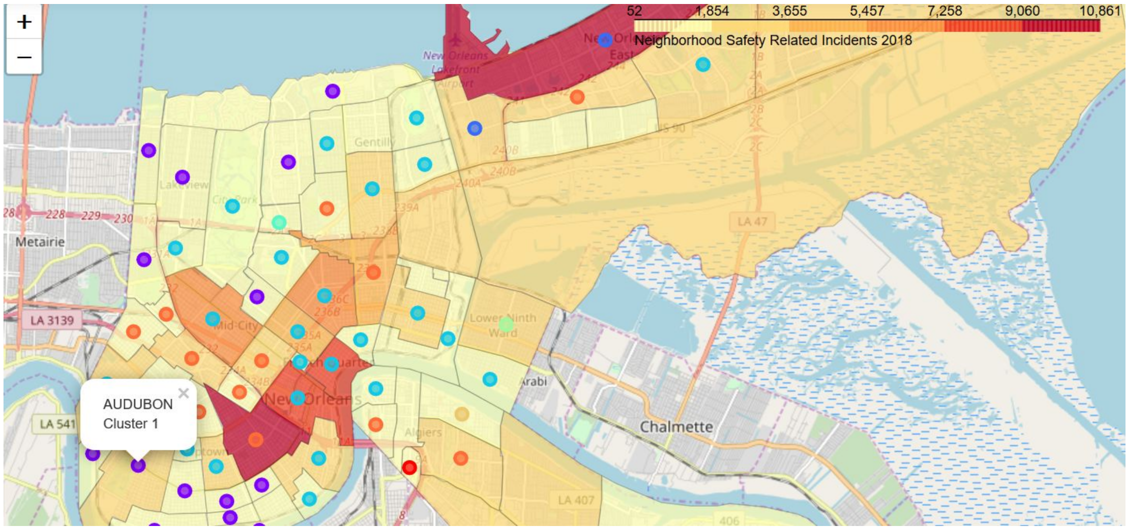


Figure 3.2.4 Choropleth Map of Incidents with Combined Clusters

4. Results

We've now processed a great deal of data. Our main incidents frame, civildf, consisted of over 146000 incidents from 2018. Foursquare returned 1675 different venues for New Orleans. First looking at the results from the analysis on incidents:

3.1.1a shows the top 3 incidents (Burglar alarm (silent), suspicious person, and domestic disturbance) greatly outweighing all others. It is not surprising that an automated call would be the top type, however the occurrence rate of suspicious person calls indicates a general distrust of others within the city. Looking at Top 20 Neighborhoods Suspicious Person in 3.1.1b, we can see that suspicious person calls are most problematic in Central City and Little Woods. Theft occurs most in the Central Business District and the French Quarter, perhaps because of businesses or tourists in the area.

In 3.1.2 we notice that nearly 100 out of 116 incident types occurred less than 2335 times. Our interquartile range is even smaller at 743, the second quartile starting at 28 incidents and the 3rd ending at 771 incidents total.This indicates that most incidents do not occur or are not reported nearly as often as a select few. Furthermore, city resources may be being spent at a disproportionate rate.

3.1.3 and 3.1.4 show that we have four neighborhoods with much higher incident occurrence than average (over 2 standard deviations above the mean). Referring to our table hoodtotals, these are Little Woods, Central City, Central Business District, and the French Quarter. These may require special attention from city officials.

The choropleth map produced in 3.1.5 clearly shows a group of problem areas near Central City and downtown New Orleans, along with the extreme red zone, Little Woods, in New Orleans East near Lake Pontchartrain. Looking at this map could be useful for those sightseeing around the city and trying to avoid problem areas.

With 3.1.6a, *Total Incident Count by Month*, we see large variations of over 1.25 std from the mean in January, May, July, and November. Our peaks are at July and May while the lowest periods for incidents are January and November. July is the all-time high month for incidents at 13004 total (over 1.5 std above the mean). November is the minimum with only 11464 incidents. The next plot, *Timeseries of Incidents for Quartile Median Neighborhoods*, shows neighborhood incident trends over the year. These remain fairly constant without significant variation.

Foursquare retrieved **1683 venues** with **252 different types** for 70 neighborhoods in New Orleans. Most venue types had a low count (mean of 6.7) while a few had very high counts. These top venue types are bars, hotels, and coffee shops with cajun/ creole restaurants following in fourth. For more, see 3.1.7.

Moving on to k-means--clustering did not perform very well when working with foursquare venue data. Many neighborhoods were underrepresented and some venue types were over represented. The inertia and distortion values were subsequently much larger for clustering with venues than with incidents or combined. K-means produced 4 clusters for Neighborhoods based on venue type frequency, 5 clusters based on incident type frequency, and 8 clusters on combined incident and venue type frequencies. Labelling of these clusters will follow in the discussion section.

5. Discussion

Overall, much was uncovered through data analysis of incidents, leading to possible insights for stakeholders in policy, policing, and non-profit. The ability to harness data of scale to precision will surely become more essential in decision making through all sectors. The following discussion provides points from the exploratory analysis and k-means clustering that can help towards the goal of mindfully allocating city resources.

Two major problems noted in the exploratory analysis are suspicious persons and theft. The high occurrence of suspicious person reports seen in Central City and Little Woods (refer to 3.1.1) may indicate low levels of social trust in the area. Organizing community events in order to familiarize members with one another may help mitigate the occurrence of suspicious person reports, freeing up police time and resources. Establishment or encouragement of local neighborhood watches may also be beneficial for these areas.

Theft is by far the biggest problem in the Central Business District and the French Quarter. There are already frequent police patrols in these areas. Though controversial, increasing surveillance cameras may act as a deterrent, with results able to be measured in years to come. This approach has problems however. One of the most obvious is that it does little to address the diverse causes of criminality. Social, economic, and other root forces are hardly alleviated by such simplistic solutions. An additional study in this direction would be helpful even so. Other solutions could be explored by seeing what cities in different regions have done to address theft occurrence.

The time series analysis conducted in 3.1.6 provides a snap-shot of how crime changes over the year. This data becomes more powerful if we combine it with multiple years worth of data. We could see trends of total incidents per month and suggest allocation of resources based on month helping those making staffing decisions for police personnel. Furthermore, we could see how different incident types and neighborhoods change based on month. Regression could even be used to predict the amount of incidents by type or neighborhood based on time of day or month. This may be a topic of interest to city officials for further research.

Unfortunately the Foursquare Venue data for New Orleans is rather lacking. It does not paint a clear picture of venues in some neighborhoods, and over represents others. This leads to a bias when clustering neighborhoods and also throws off insights when linking venues with crime type. With these caveats and warnings, we'll examine the combined clusterings of venue types and incident types with the choropleth map here and their data below.

Cluster 1

	Neighborhoods	1st Most Common Incident	2nd Most Common Incident	3rd Most Common Incident	4th Most Common Incident	5th Most Common Incident	6th Most Common Incident	7th Most Common Incident	8th Most Common Incident	9th Most Common Incident	...	1st Most Common Venue	2nd Most Common Venue	
18	FISCHER DEV	DOMESTIC DISTURBANCE	HIT & RUN	SUSPICIOUS PERSON	SIMPLE BATTERY DOMESTIC	BURGLAR ALARM, SILENT	FUGITIVE ATTACHMENT	DRIVING WHILE UNDER INFLUENCE	FIGHT	SIMPLE CRIMINAL DAMAGE	...	Sandwich Place	Women's Store	E

1 rows × 24 columns

Cluster 1 consists of a single neighborhood, Fischer Dev. It has a relatively low amount of incidents, but they consist of a slightly more serious kind, such as domestic battery and fight. These moderate-type incidents are paired with sub-urban venues. We'll call this cluster **moderate-suburban.**

Cluster 2

	Neighborhoods	1st Most Common Incident	2nd Most Common Incident	3rd Most Common Incident	4th Most Common Incident	5th Most Common Incident	6th Most Common Incident	7th Most Common Incident	8th Most Common Incident	9th Most Common Incident	...	1s Co
1	AUDUBON	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	HIT & RUN	NOISE COMPLAINT	SIMPLE BURGLARY VEHICLE	HOLD UP ALARM	THEFT	DRUG VIOLATIONS	AUTO THEFT	...	C B
3	BAYOU ST. JOHN	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	DOMESTIC DISTURBANCE	SIMPLE BURGLARY VEHICLE	HIT & RUN	HOLD UP ALARM	THEFT	AUTO THEFT	SIMPLE CRIMINAL DAMAGE	...	Sa
5	BLACK PEARL	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	SIMPLE BURGLARY VEHICLE	DOMESTIC DISTURBANCE	HIT & RUN	NOISE COMPLAINT	HOLD UP ALARM	AUTO THEFT	FIGHT	...	Lig
14	EAST CARROLLTON	BURGLAR ALARM, SILENT	NOISE COMPLAINT	SUSPICIOUS PERSON	HIT & RUN	SIMPLE BURGLARY VEHICLE	DOMESTIC DISTURBANCE	AUTO THEFT	HOLD UP ALARM	SIMPLE CRIMINAL DAMAGE	...	
15	EAST RIVERSIDE	BURGLAR ALARM, SILENT	SIMPLE BURGLARY VEHICLE	DOMESTIC DISTURBANCE	SUSPICIOUS PERSON	HIT & RUN	AUTO THEFT	THEFT	NOISE COMPLAINT	HOLD UP ALARM	...	
17	FILMORE	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	DOMESTIC DISTURBANCE	SIMPLE BURGLARY VEHICLE	HOLD UP ALARM	SHOPLIFTING	AUTO THEFT	FUGITIVE ATTACHMENT	EXTORTION (THREATS)	...	

Cluster 2 has 15 neighborhoods. We see low incident rate neighborhoods in cluster 2 in more expensive areas. There are universities, bars, coffee shops, and restaurants with less serious incidents and auto-related incidents. With lots of walkable, long-stay venues, we might connect auto-related incidents and theft with these urban-stay locales. We'll label cluster 2 as **car-beware, urban-stay.**

Cluster 3

	Neighborhoods	1st Most Common Incident	2nd Most Common Incident	3rd Most Common Incident	4th Most Common Incident	5th Most Common Incident	6th Most Common Incident	7th Most Common Incident	8th Most Common Incident	9th Most Common Incident	...	1st Most Common Venue	2nd Most Common Venue
37	LITTLE WOODS	BURGLAR ALARM, SILENT	DOMESTIC DISTURBANCE	SUSPICIOUS PERSON	SIMPLE BURGLARY VEHICLE	SIMPLE BATTERY DOMESTIC	AUTO THEFT	SIMPLE CRIMINAL DAMAGE	HIT & RUN	THEFT	...	Seafood Restaurant	Cosmetics Shop
49	PINES VILLAGE	BURGLAR ALARM, SILENT	DOMESTIC DISTURBANCE	SUSPICIOUS PERSON	HIT & RUN	SIMPLE BATTERY DOMESTIC	SHOPLIFTING	FIGHT	AUTO THEFT	SIMPLE BURGLARY VEHICLE	...	Pizza Place	Event Service

2 rows × 24 columns

Cluster 3 has 2 neighborhoods in high-incident zones. These are in New Orleans East, with mostly car-accessed venues. This cluster will be labelled **car-accessessed risk**.

Cluster 4

	Neighborhoods	1st Most Common Incident	2nd Most Common Incident	3rd Most Common Incident	4th Most Common Incident	5th Most Common Incident	6th Most Common Incident	7th Most Common Incident	8th Most Common Incident	9th Mo Comm Incide
0	ALGIERS POINT	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	SIMPLE BURGLARY VEHICLE	DOMESTIC DISTURBANCE	THEFT	QUALITY OF LIFE ISSUE	NOISE COMPLAINT	DISCHARGING FIREARM	HIT & RU
7	BYWATER	BURGLAR ALARM, SILENT	SHOPLIFTING	SUSPICIOUS PERSON	HIT & RUN	DOMESTIC DISTURBANCE	AUTO THEFT	THEFT	SIMPLE BURGLARY VEHICLE	DISCHARGIN FIREAR
8	CENTRAL BUSINESS DISTRICT	THEFT	BURGLAR ALARM, SILENT	HIT & RUN	SUSPICIOUS PERSON	SIMPLE BATTERY	FIGHT	SHOPLIFTING	FUGITIVE ATTACHMENT	DRU VIOLATION
10	CITY PARK	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	HIT & RUN	SIMPLE BURGLARY VEHICLE	RECKLESS DRIVING	THEFT	NOISE COMPLAINT	AUTO THEFT	DOMEST DISTURBANC

Cluster 4 is one of the largest clusters with 25 neighborhoods. Venues mostly consist of restaurants and incidents mostly consist of burglar alarms and suspicious persons. We can tenuously infer that an increased presence of restaurants gets more break-in attempts in an area, coinciding with suspicious persons. This cluster will be labelled **restaurant-break-in**.

Clusters 5, 6, and 7

Clusters 5-7 all consisted of single neighborhoods.

Cluster 8

	Neighborhoods	1st Most Common Incident	2nd Most Common Incident	3rd Most Common Incident	4th Most Common Incident	5th Most Common Incident	6th Most Common Incident	7th Most Common Incident	8th Most Common Incident	9th Mo Commo Incide
2	B. W. COOPER	HIT & RUN	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	DOMESTIC DISTURBANCE	DRUG VIOLATIONS	RECKLESS DRIVING	SIMPLE BURGLARY VEHICLE	FUGITIVE ATTACHMENT	AUTO THEF
4	BEHRMAN	DOMESTIC DISTURBANCE	SUSPICIOUS PERSON	BURGLAR ALARM, SILENT	SHOPLIFTING	DISCHARGING FIREARM	SIMPLE BATTERY DOMESTIC	FUGITIVE ATTACHMENT	HIT & RUN	SIMPL BATTER
6	BROADMOOR	BURGLAR ALARM, SILENT	SUSPICIOUS PERSON	DOMESTIC DISTURBANCE	DRUG VIOLATIONS	SIMPLE BURGLARY VEHICLE	HIT & RUN	SHOPLIFTING	FUGITIVE ATTACHMENT	AUTO THEF
9	CENTRAL CITY	SUSPICIOUS PERSON	BURGLAR ALARM, SILENT	DOMESTIC DISTURBANCE	HIT & RUN	DRUG VIOLATIONS	SHOPLIFTING	FUGITIVE ATTACHMENT	SIMPLE BATTERY	THEF

Our last cluster of 13 neighborhoods are in moderate-high incident zones. Incident types are more serious with theft, battery, and drug violations common. Frequent venue types are gas stations, discount stores, light-industrial venues, and sports-related venues. We'll call this cluster **Moderate-light industrial**.

We can draw insights from clusters 2, 4, and 8. With Clustering from 2 and we can reasonably say that the more long-period-stay venues in an area, the more car-break ins you can expect. Group 4 suggests that restaurants are targets for burglary, with high alarm and suspicious person incidents. Lastly, group 8 shows a tenuous relation between moderate-crime and light industrial areas.

6. Conclusion

The purpose of this project was to elucidate data-based insights and relations between incidents and venues in New Orleans. The data analysis on incidents and venues has been able to produce take-aways toward the efficient use of police and public resources showing where, what, and when incidents occur along with venues they may be related with. K-means loosely linked car-related incidents with long-stay venues, burglar alarms and suspicious persons with restaurants, and moderate-crime with light industrial areas. Lastly, the choropleth map is an immediate visual aid in identifying problem areas in the city. Combining this type of data along with that already available in resources such as the opportunity atlas can help us better relate diverse factors such as income, education, and housing price to ongoing crime.