Analysis of 2019 Homicides in Chicago

Antonio Santarsiero, MBA

Capstone Project for IBM Data Science Certificate

1. **Introduction**

Chicago’s reputation as one of America’s most violent cities sparked my curiosity to investigate a dataset of the homicides happening in the “Windy City” in 2019. Between 2010 and 2015, Chicago had 16.4 murders per 100,000 residents; numbers that place the city far from the top – taken by New Orleans, Detroit, and St. Louis – yet still in the top 20.

**1.2 Problem**

Given the reputation Chicago has as a dangerous city, residents and tourists alike may wonder where and when most of the murders take place. I analyzed the trend of murders by location – distribution of the homicides by neighborhood – and by time - month, day, and part of the day. I also threw in the analysis a question not many may wonder: what are the most common venues around homicides?

**1.3 Interest**

It goes without saying that the residents of the city of Chicago as well as casual tourists – or the most frequent ones from the suburbs – would be interested from my findings. Moreover, enforcement authorities would have an idea of how to better distribute their workforce throughout the neighborhoods and throughout the year.

1. **Analysis of 2019 Homicides in Chicago**

***2.1 Data Sources***

All the data regarding crimes can be found on the Chicago Data Portal, where the city officials periodically publish official data about the Windy City. Unfortunately, it is only possible to download data regarding the 2019 crimes.

The other source used in my research is the Foursquare database. Through API calls, I was able to obtain the category type of the venues in a 200m radius from the homicide location.

***2.2 Data Cleaning and Feature Selection***

The dataset extracted from the Chicago Data Portal contains specific data about the crimes such as case number, date, location data (block, latitude, longitude, community area), and data regarding the crime (category, divided in primary and secondary type, and then whether the crime ended up requiring in an arrest or not, whether it was domestic, and so on.)

For my research, I sliced the data frame to obtain only data regarding crimes with primary type “Homicide”, disregarding the more selective secondary type column. I also dropped columns like “Arrest” or “Domestic” to focus only on the location data (coordinates and neighborhood) and date information.

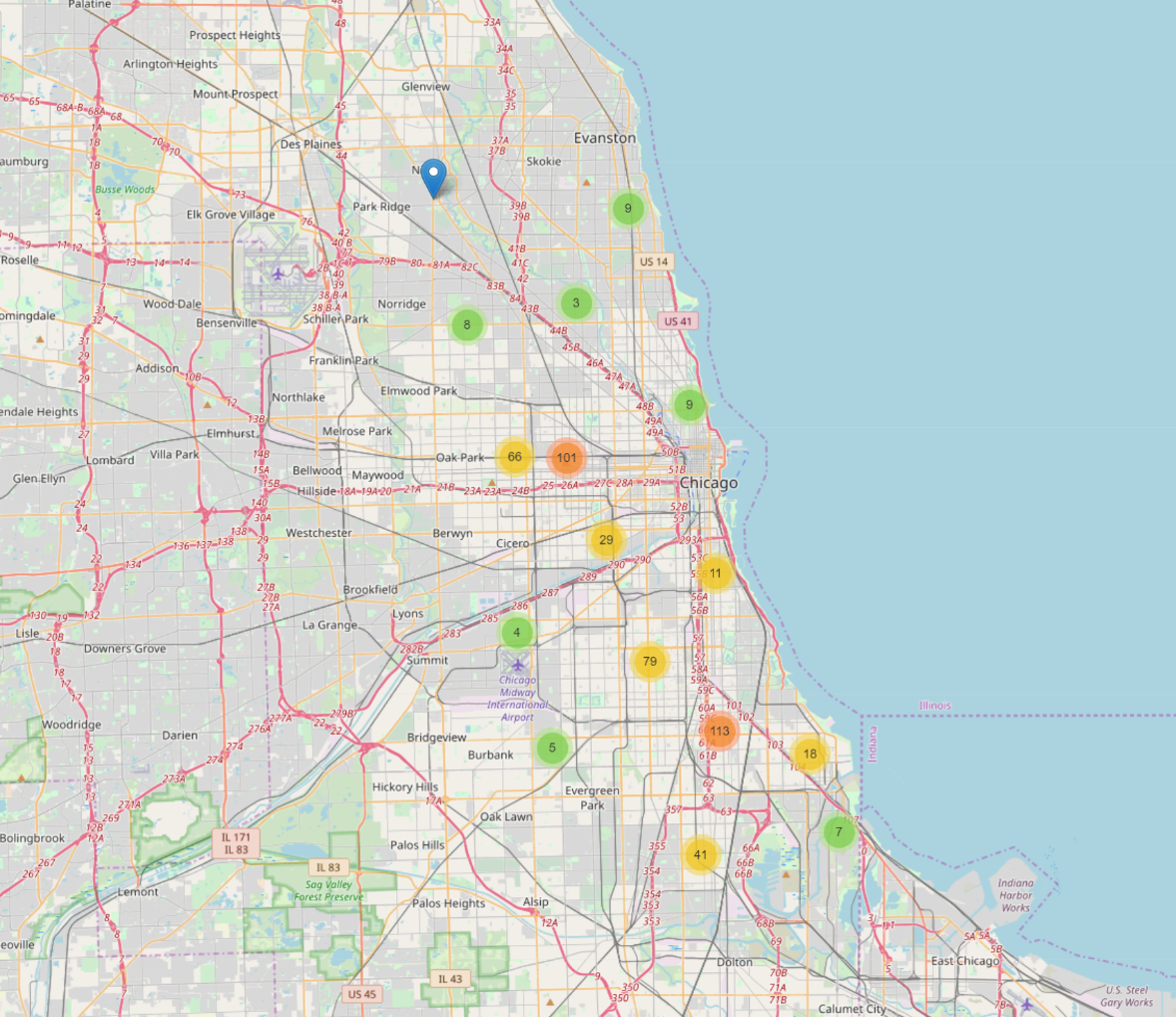
1. **Methodology**

After slicing the data frame, I mapped the homicides using the Folium library in Python and clustered them to have a better representation (Figure 1). I then created a function to extract the closest venues in a 200 meters (220 yards) radius from all of the homicides location, grouped the results by venue category, and transferred the findings into a data frame. From there, I plotted the 5 most frequent venue categories that appeared the most near the murder locations (Figure 3) and that accounted for almost 20% of the total venue categories – on a total of 212 different ones.

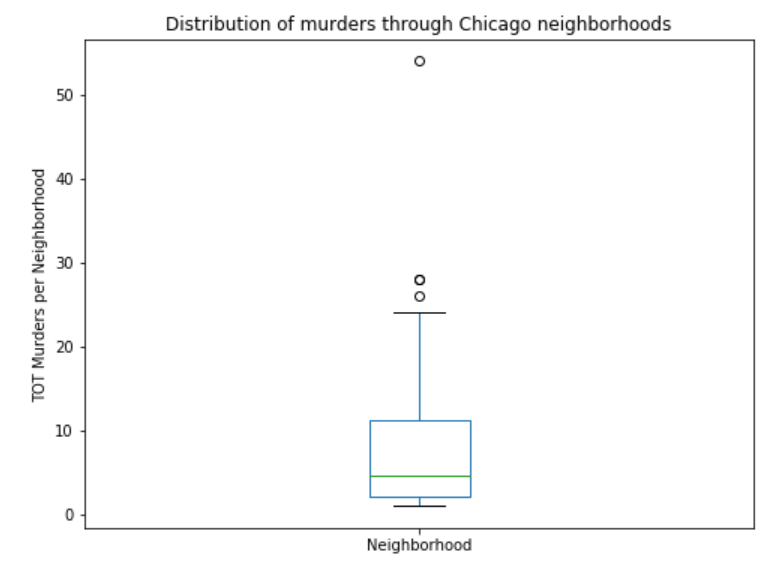
The second part of my analysis involved slicing the original data frame to obtain only time data – the columns ‘Time’, ‘Month’, and ‘Day’. I explored the distribution of murders throughout the day (Figure 4) - considering ‘Night’ from 10pm to 6am, ‘Morning’ from 6am-12pm, ‘Afternoon’ from 12pm-6pm, and ‘Evening’ from 6pm-10pm - then throughout the year (Figure 5), and throughout the different days in a week (Figure 6.)

1. **Results and Observations**

Using the Folium library on Python, I was able to obtain a map depicting the distribution of homicides throughout Chicago. The highest concentration of homicides happened in the southern and western neighborhoods, with a peak of murders in Austin (54 total murders compared to a median of 6 per neighborhood, as shown in the box plot in Figure 2).



*Figure 1. Map of 2019 murders in Chicago.*



*Austin*

*Figure 2. Box plot of murder distribution by neighborhood.*



*Table 1. Count of 2019 murders per neighborhood.*

Using Foursquare, I then discovered the most frequent venues near the murder location (in a 200m radius). In the chart below, the top 5 venue categories followed by the list of the top 50. As one can notice, the most “suspected” places like train stations or parks fall, respectively, 14th and 15th in the ranking.

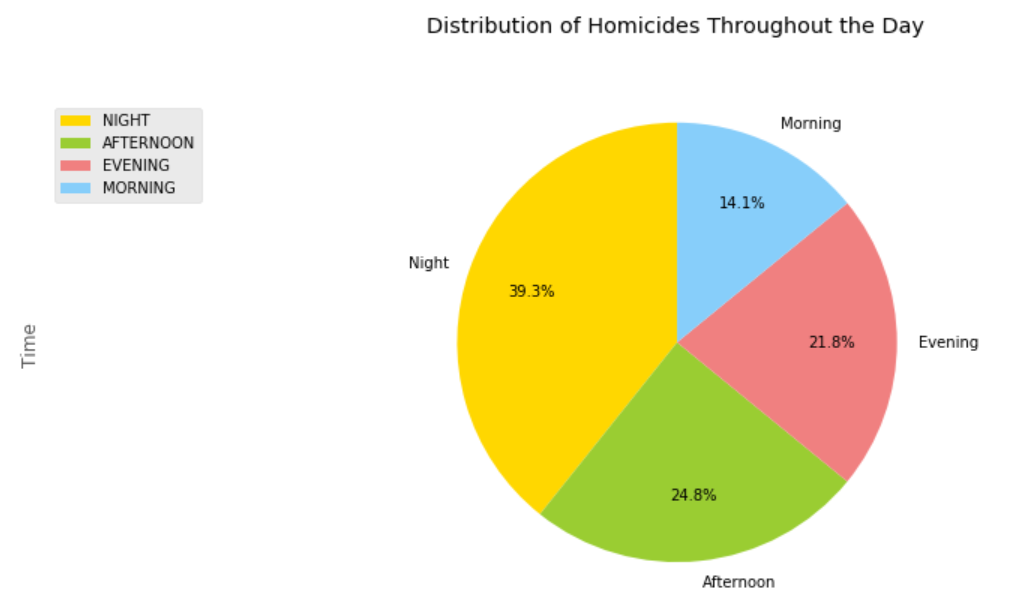


*Figure 3. Venue categories that appeared the most in a 200m radius from murders*

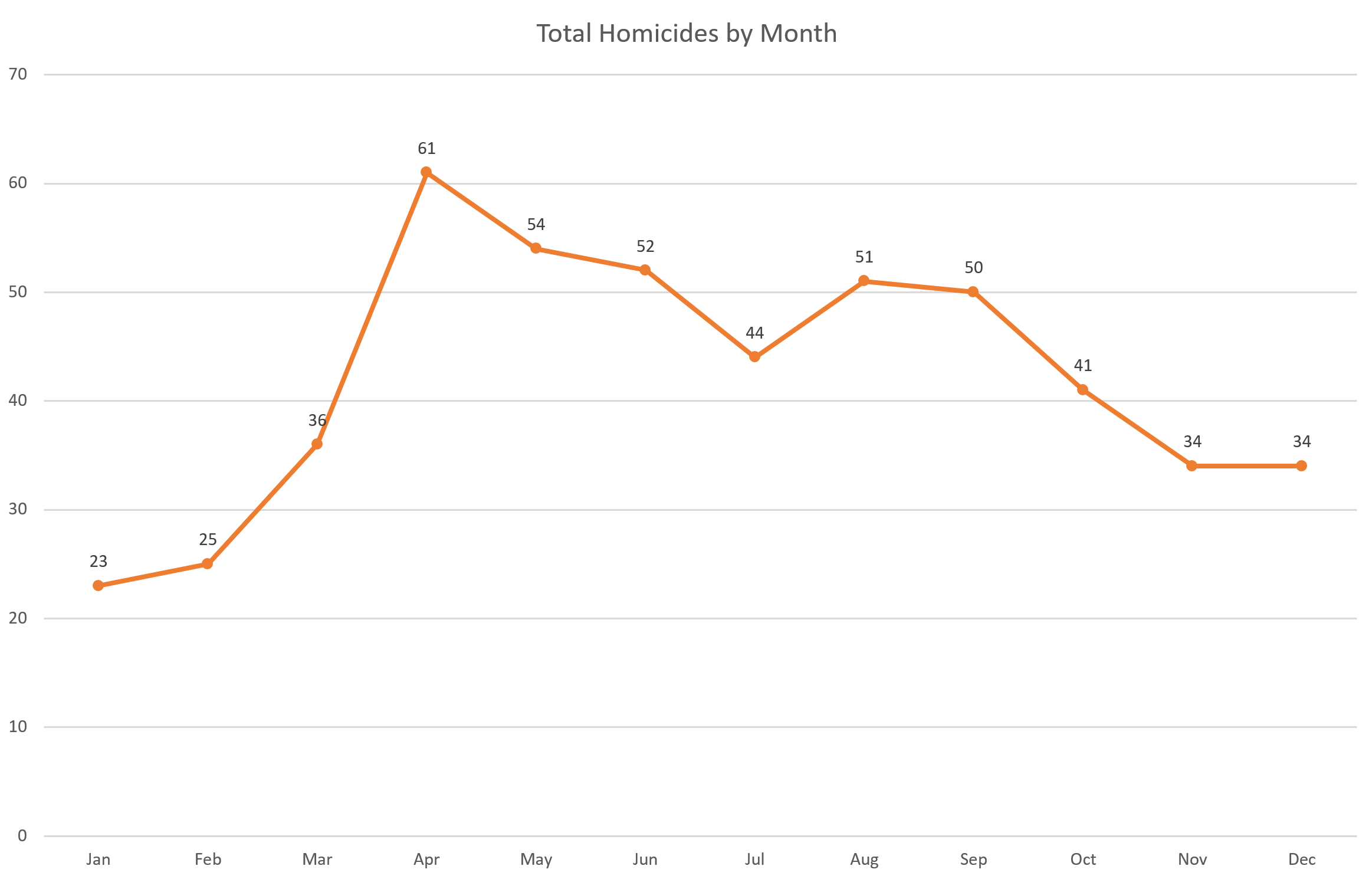
**

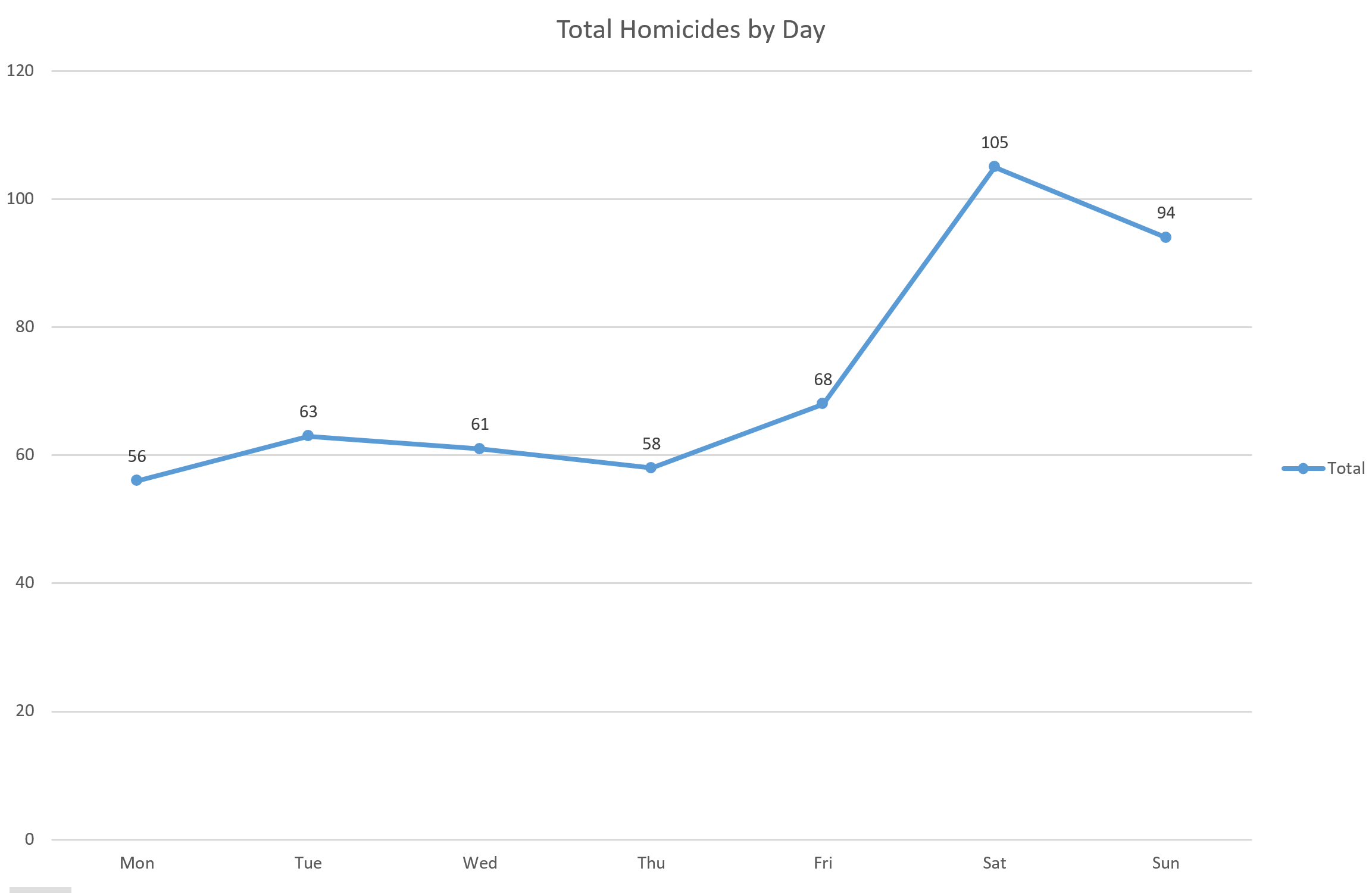
*Table 2. Top 40 venue categories in a 200m radius from murders.*

The following charts, instead, show the distribution of murders based on time data. As one can notice, most of the murders happened in the night (almost 40%), during the spring (34%) and the summer (28%), and generally in the weekend (Saturday and Sunday account for almost 40% of the total homicides).



*Figure 4. Most frequent time of 2019 murders in Chicago.*

*Figure 5. Distribution of 2019 Murders by Month*



*Figure 6. Distribution of homicides by day.*

1. **Conclusion**

In this study, I analyzed the 505 murders that happened in Chicago in 2019. My analysis focused on location and time data. It surprisingly found out that the neighborhood with the highest number of homicides was not in the southern area; indeed, Austin – a western neighborhood – topped the list with an astonishing 54 total crimes, marking itself as an outlier. Also, by extracting data from Foursquare, I located the most common venue categories in a 200m radius from each crime. As shown in figure 3, murders happened the most near fast foods, Mexican restaurants, and pizza places, leaving out train stations, parks, and bus stations away from the top 10. With regards to time data, my analysis found out that most of the homicides happened during spring and summer (leaving the winter season last), during weekends (almost 40% happened on a Saturday or Sunday), and mainly in the night (40%.) I am sure residents of the Chicagoland will find this data interesting; my analysis wanted to shed some light on a topic that is surely widely discussed in Chicago, yet without the proper data.