

Boston Crime Data Analysis

Andrea Marcoccia (775451), Aurora
Menegatto (783301), Giulia Scalas
(772601), Irene Fellin (785421)

Python and R Group Project



Overview of "Crimes in Boston" Dataset

1. Purpose of the Dataset

- Documents initial details of incidents responded to by Boston Police Department officers.
- Emphasizes capturing the nature of crime incidents.

2. Structural Emphasis

- Highlights key details: type of offence, timing, and location.
- Critical for understanding crime patterns and trends in Boston.

3. Time Frame

- Spans from June 14, 2015, to September 3, 2018.
- Encompasses records crucial for historical analysis.

4. Additional Dataset

- Complementary dataset mapping OFFENSE_CODE to respective names.
- Enhances the interpretability of the primary dataset.

Key Takeaways

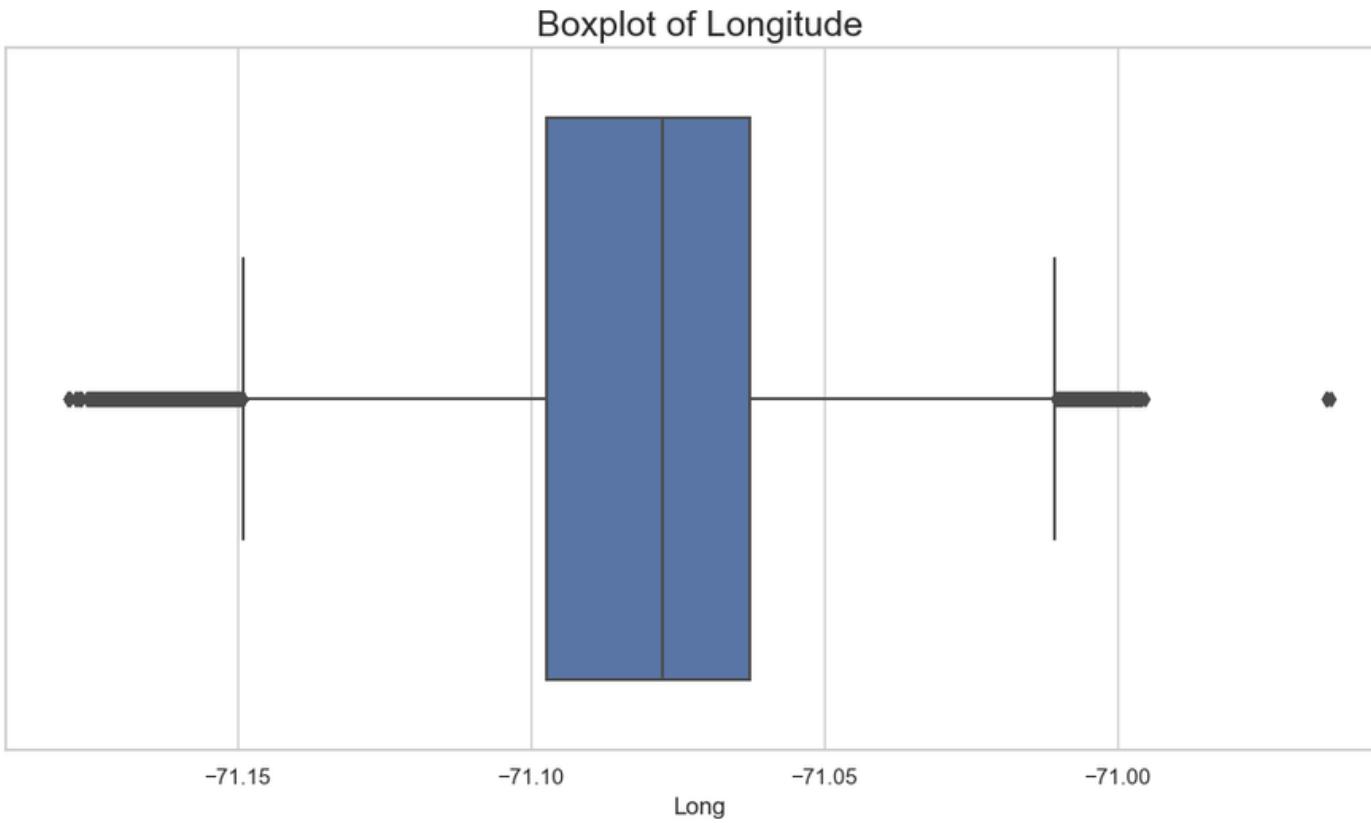
- Temporal and spatial attributes enable in-depth analysis of crime patterns.
- Integration with an additional dataset enhances the clarity of offense codes.

Data Preparation



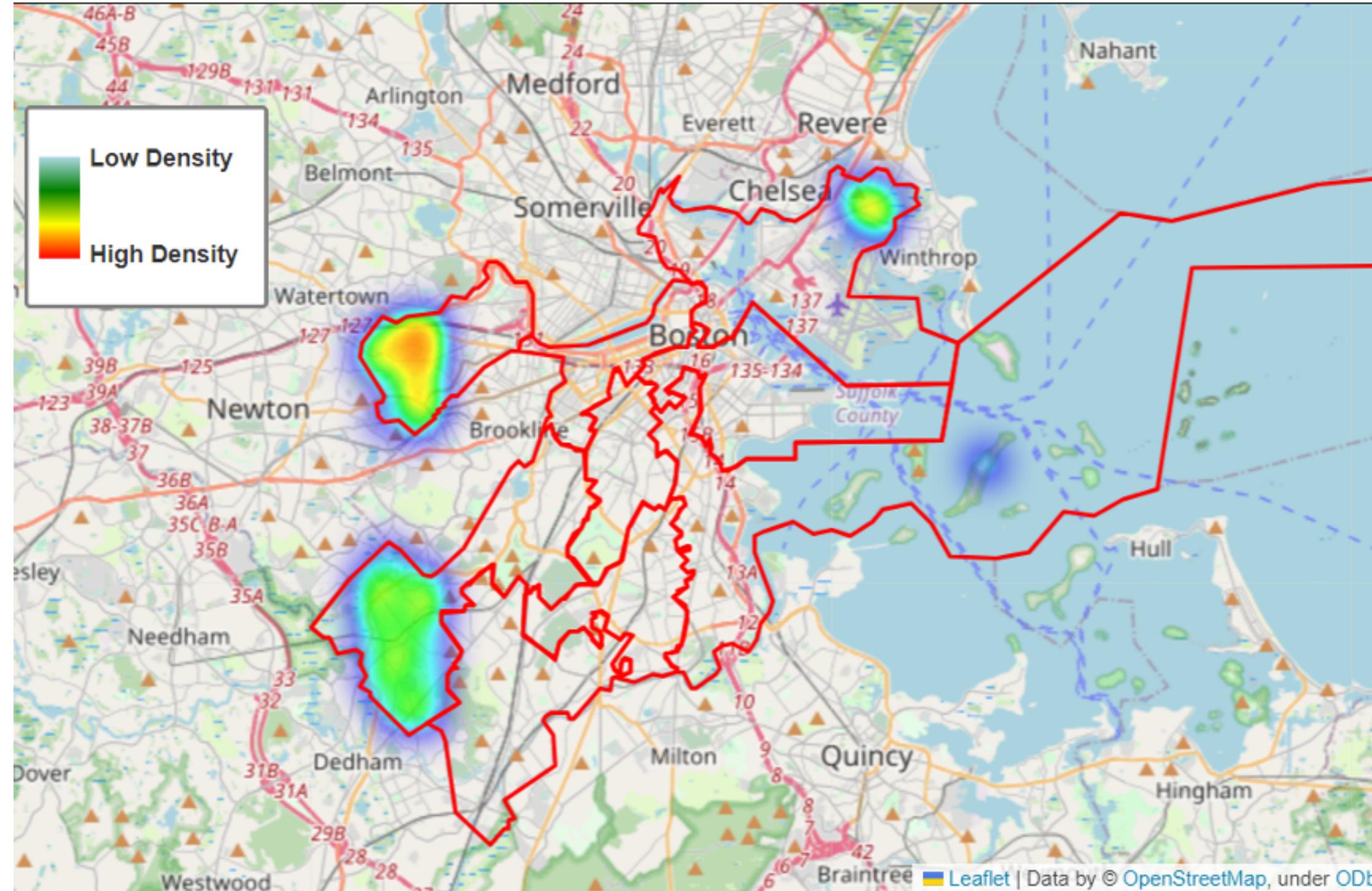
INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH
I182070945	619	Lancerny	LARCENY ALL OTHERS	D14	808	NaN	2/9/18 13.00	2018	9
I182070943	1402	Vandalism	VANDALISM	C11	347	NaN	21/8/18 0.00	2018	8
I182070941	3410	Towed	TOWED MOTOR VEHICLE	D4	151	NaN	3/9/18 19.27	2018	9
I182070940	3114	Property	INVESTIGATE PROPERTY	D4	272	NaN	3/9/18 21.16	2018	9
I182070938	3114	Property	INVESTIGATE PROPERTY	B3	421	NaN	3/9/18 21.05	2018	9

DAY_OF_WEEK	HOUR	UCR_PART	STREET	LAT	LONG	LOCATION	OFFENSE_NAME
Sunday	13	Part One	LINCOLN ST	42.357791	-71.139371	(42.35779134, -71.13937053)	LARCENY OTHER \$200 & OVER
Tuesday	0	Part Two	HECLAST	42.306821	-71.060300	(42.30682138, -71.06030035)	VANDALISM
Monday	19	Part Three	CAZENOUE ST	42.346589	-71.072429	(42.34658879, -71.07242943)	TOWED MOTOR VEHICLE
Monday	21	Part Three	NEWCOMB ST	42.334182	-71.078664	(42.33418175, -71.07866441)	INVESTIGATE PROPERTY
Monday	21	Part Three	DELHI ST	42.275365	-71.090361	(42.27536542, -71.09036101)	INVESTIGATE PROPERTY



- **Data cleaning and preparation**
- We have **added** the “**OFFENSE_NAME**” from the secondary dataset to the main dataset
- Handling **Missing Values**
- **Outlier** detection and correction
- Descriptive statistics revealed **anomalies** in Lat and Long, removed for accurate spatial analysis.
- **Boxplots** were used to identify outliers; and confirmed longitude data deviations.

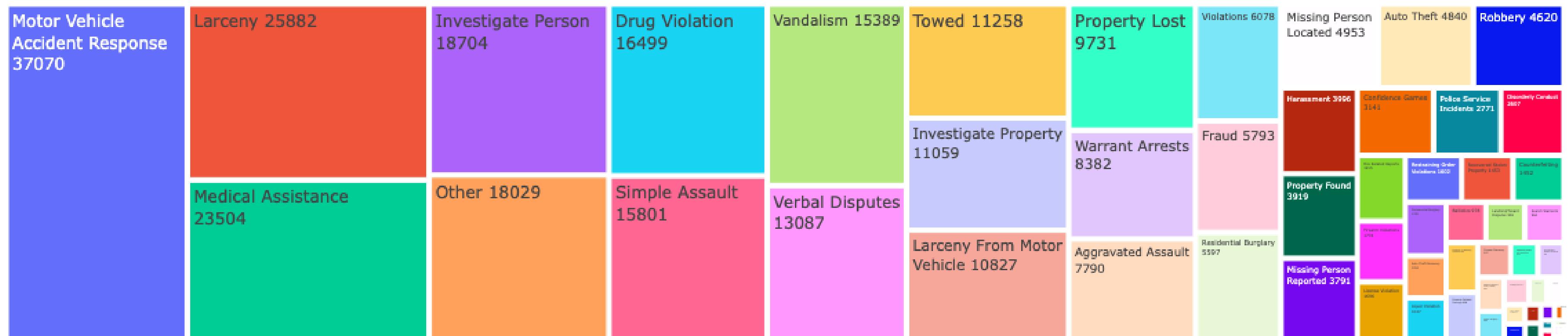
Data Preparation



- **Heatmap** visualization with Boston City Council Districts for outlier validation.
- Ensuring **Data Integrity**
- **EDA** has been concentrated on subsets of features from the dataset.

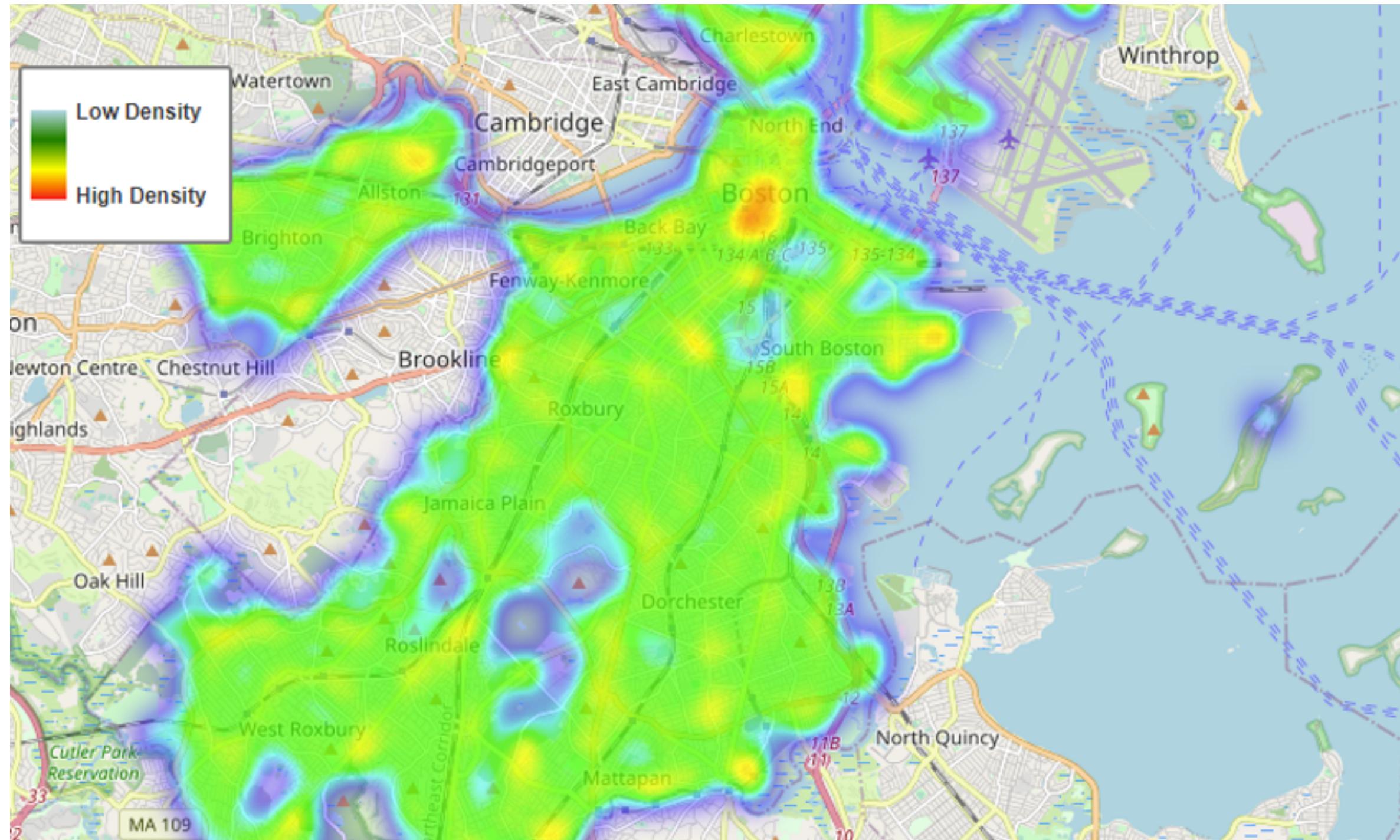
Crime distribution

Major Crimes in Boston from mid 2015 until 2018



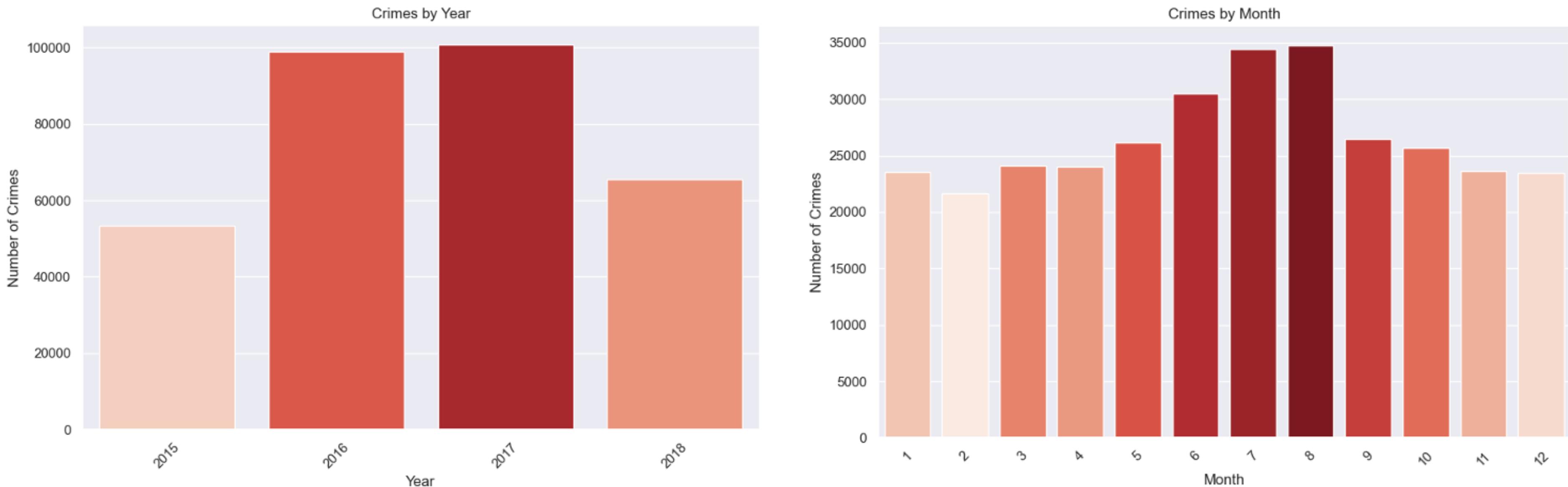
- Wide range of crime types
- The most common are the Motor Vehicle Accident and Larceny
- Least Common are Biological Threat, Burglary and Human Trafficking

Crime distribution



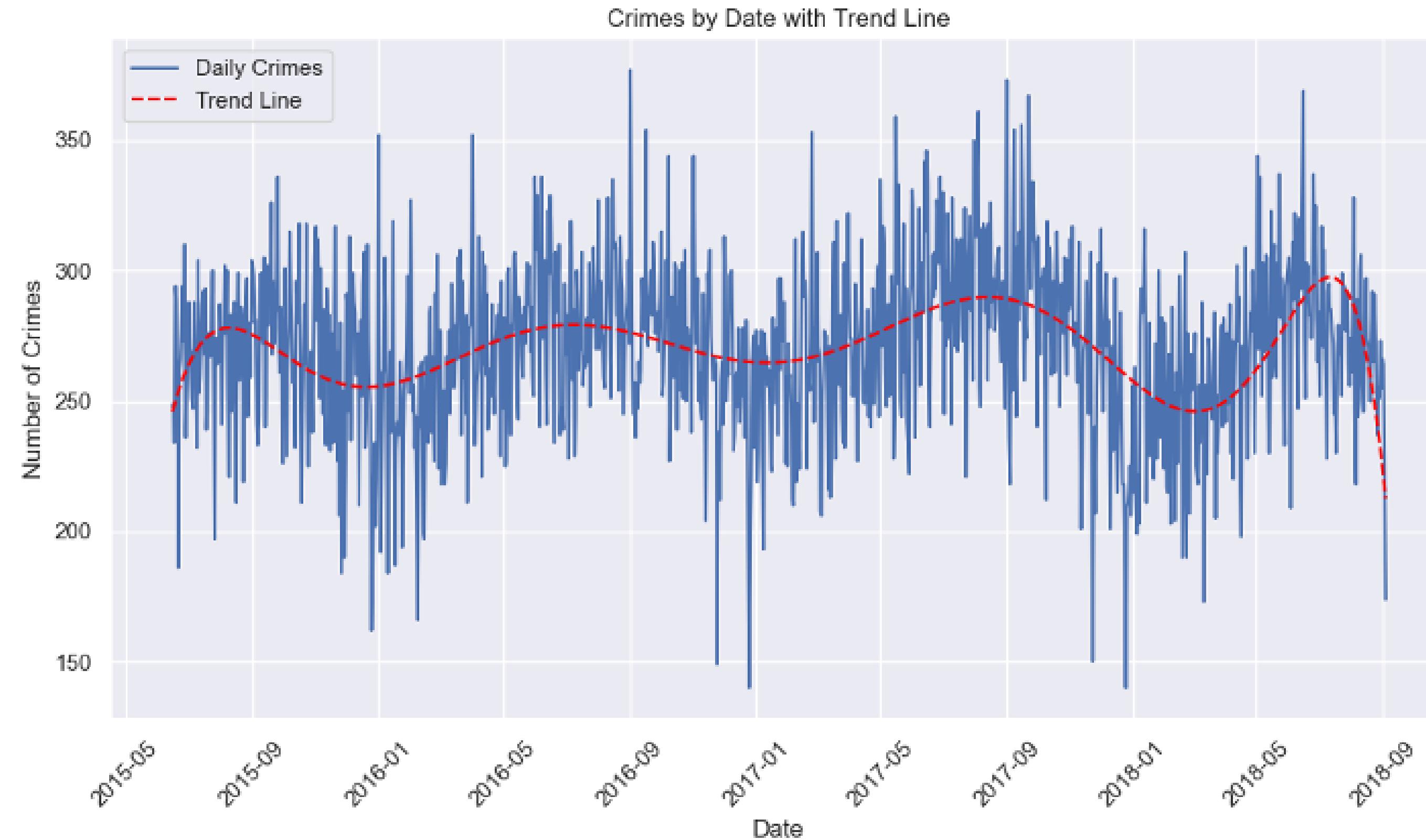
The heatmap represents the **density of the crimes** across Boston districts. The **city centre** has the highest density of crimes, followed by the **South Boston district**

Crime distribution over time



- **Uneven distribution** across the years attributable to the dataset's timeframe (starting June 14, 2015, ending on September 3, 2018)
- **2016 and 2017** have almost **double** the number of crimes compared to **2015 and 2018**
- It seems that crimes tend to **increase during the summer months** however it is not reliable due to partial presence of data in 2015 and 2018

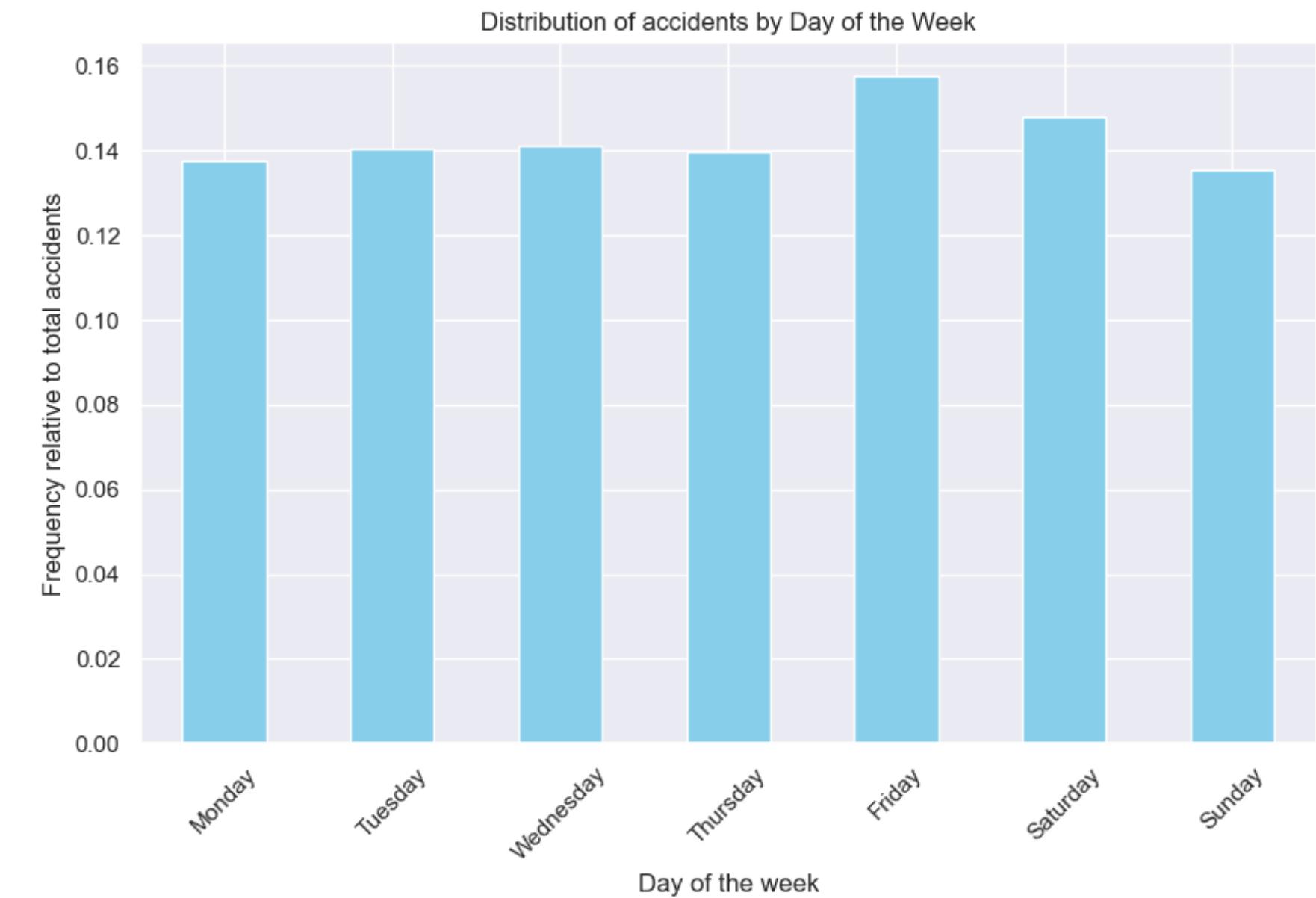
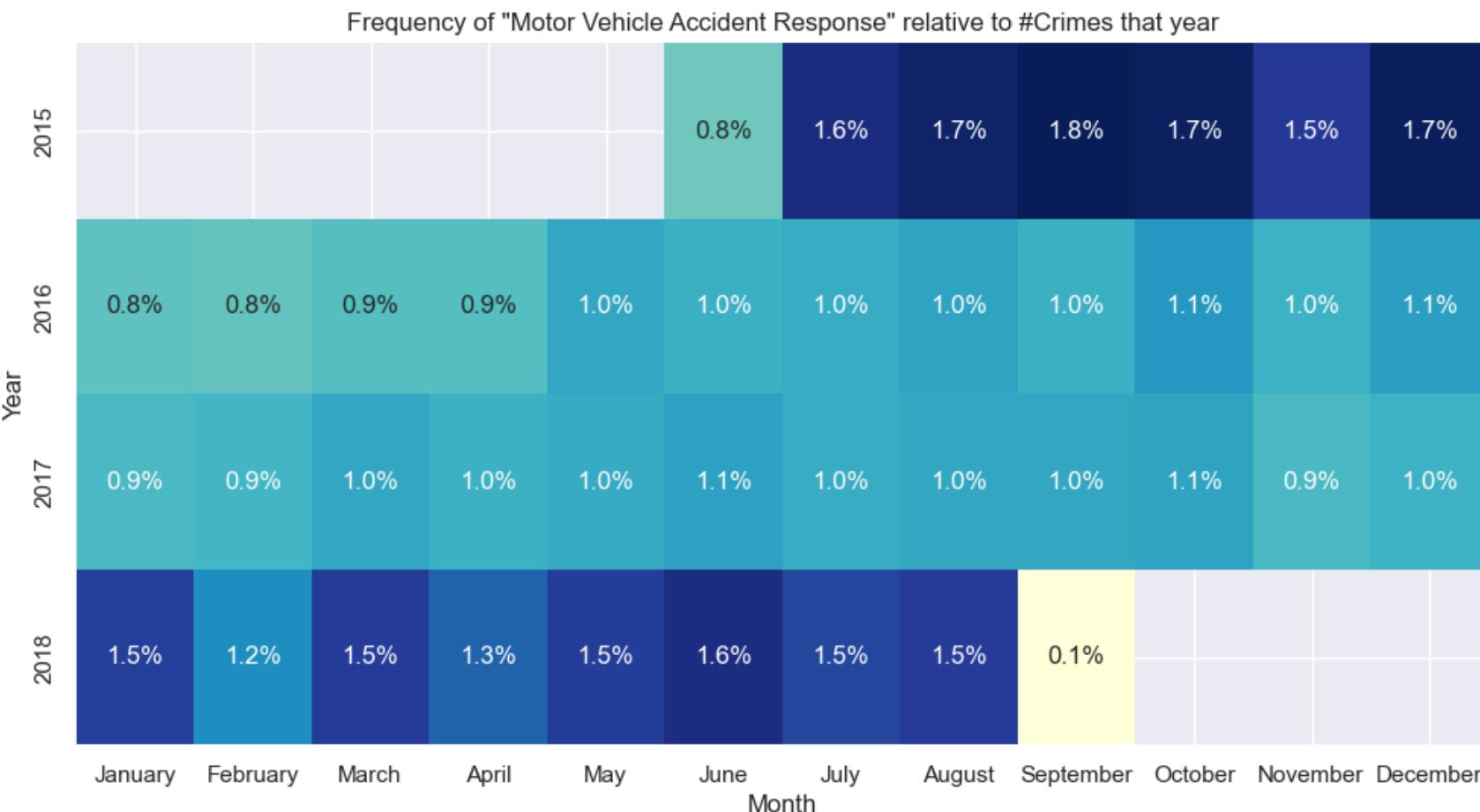
Seasonal trend in crimes



- The graph highlights a pattern that suggests a **strong seasonal influence**
- The crime rates **decrease towards January**, rising to a **peak around July**

Specific Crime Analysis

Motor vehicle accidents

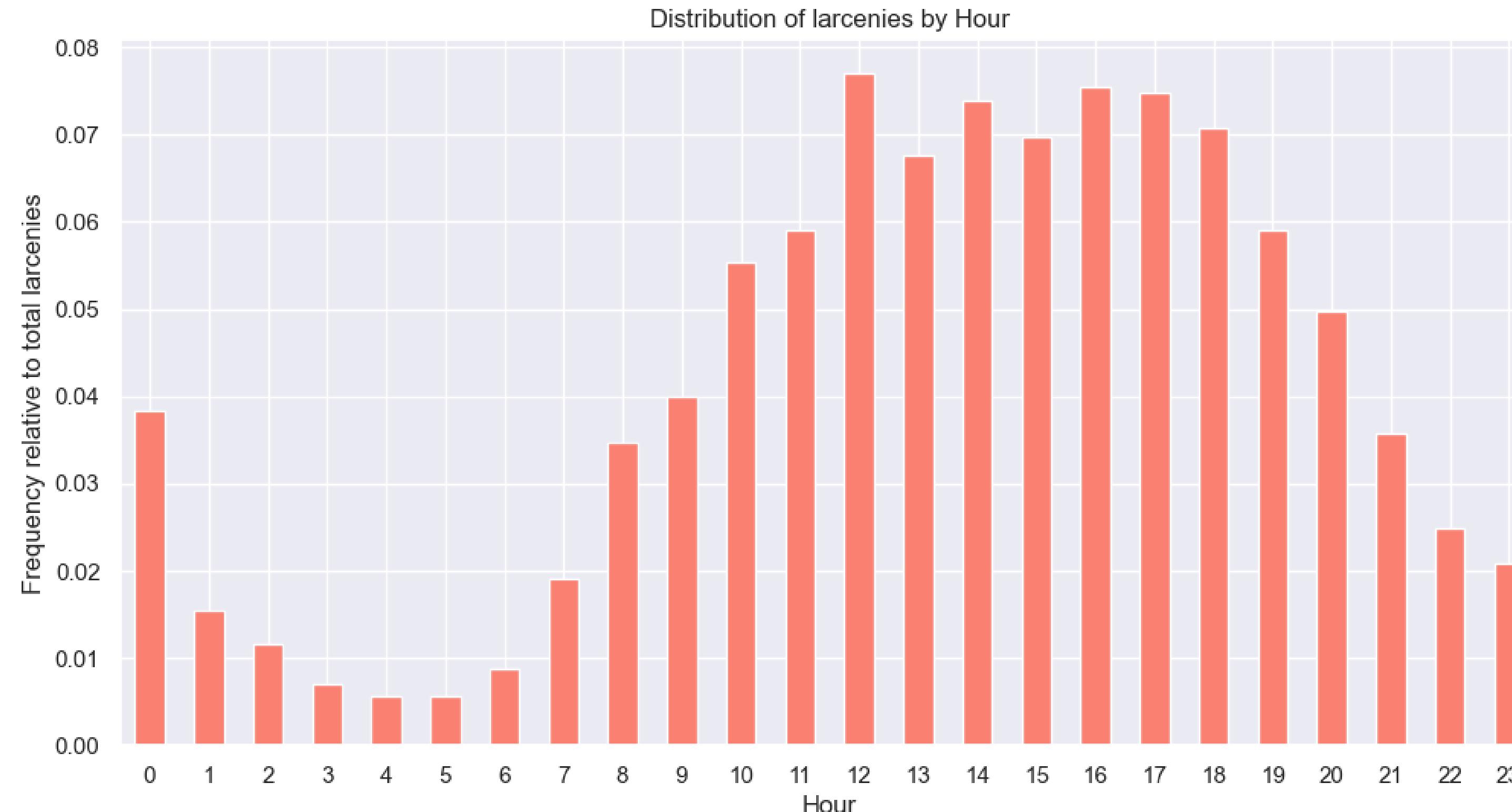


Frequency w.r.t. #crimes on specific years

Accidents distribution by day

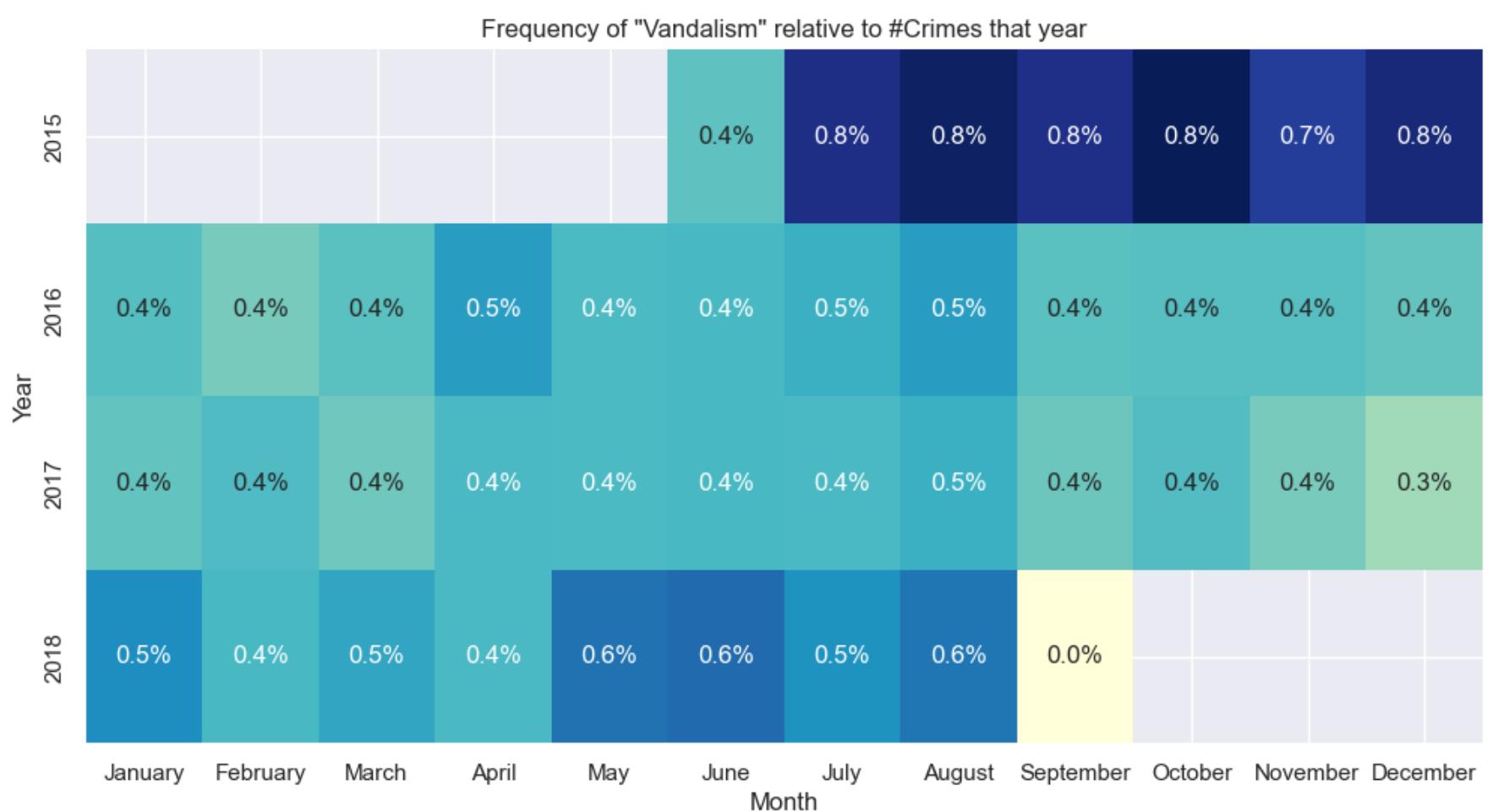
Specific Crime Analysis

Larceny

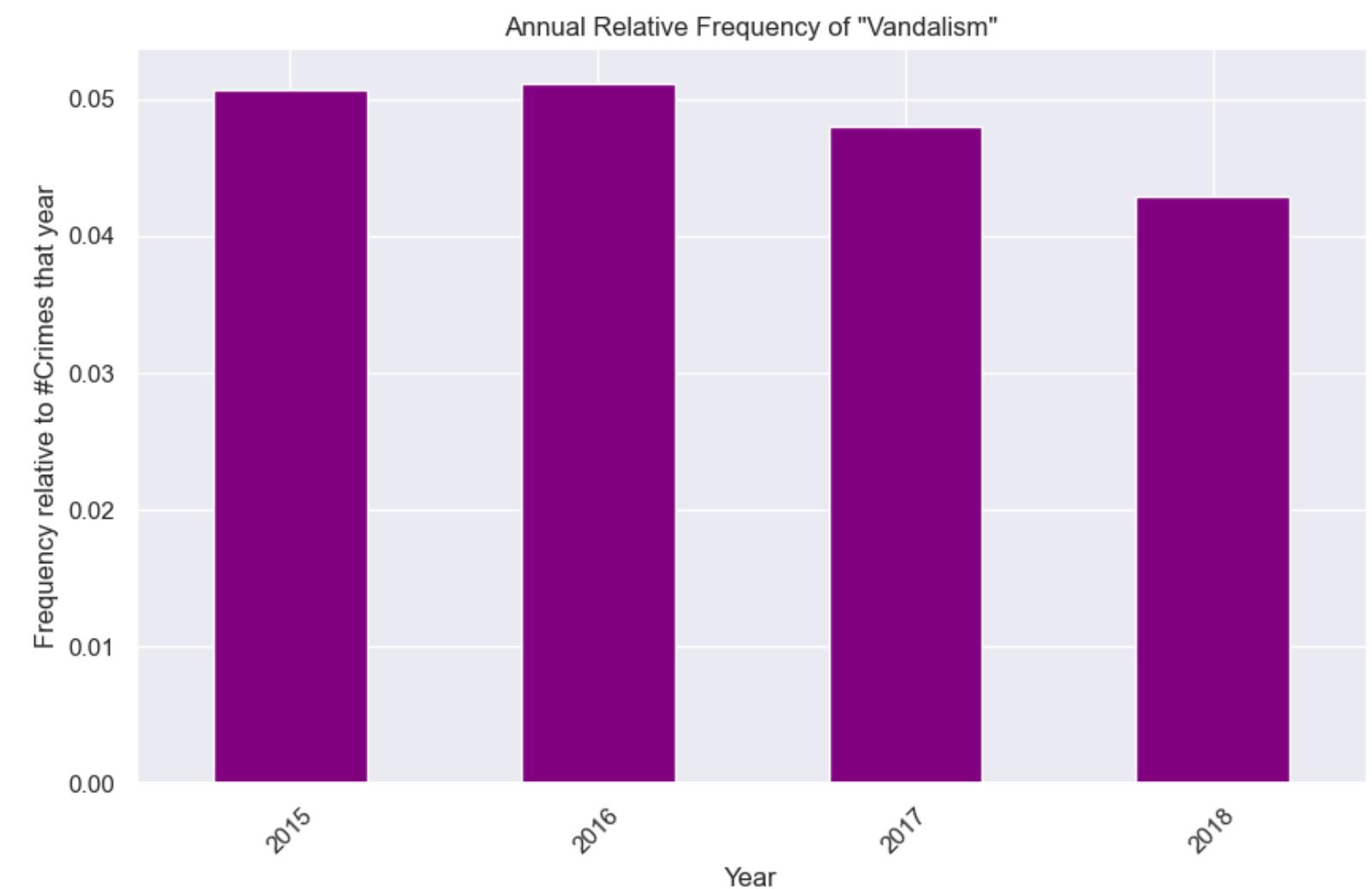


Specific Crime Analysis

Vandalism



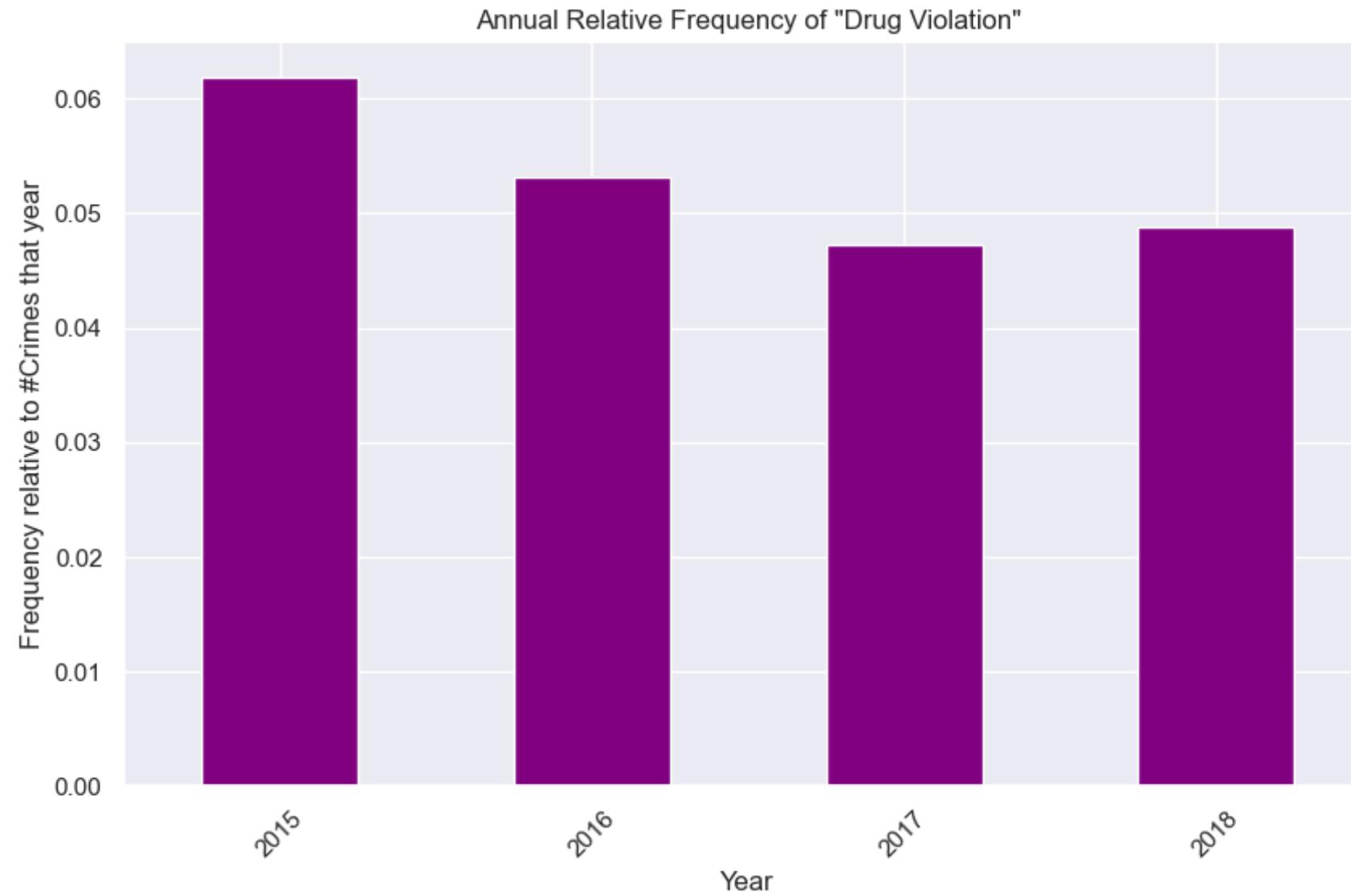
Frequency w.r.t. #crimes on specific years



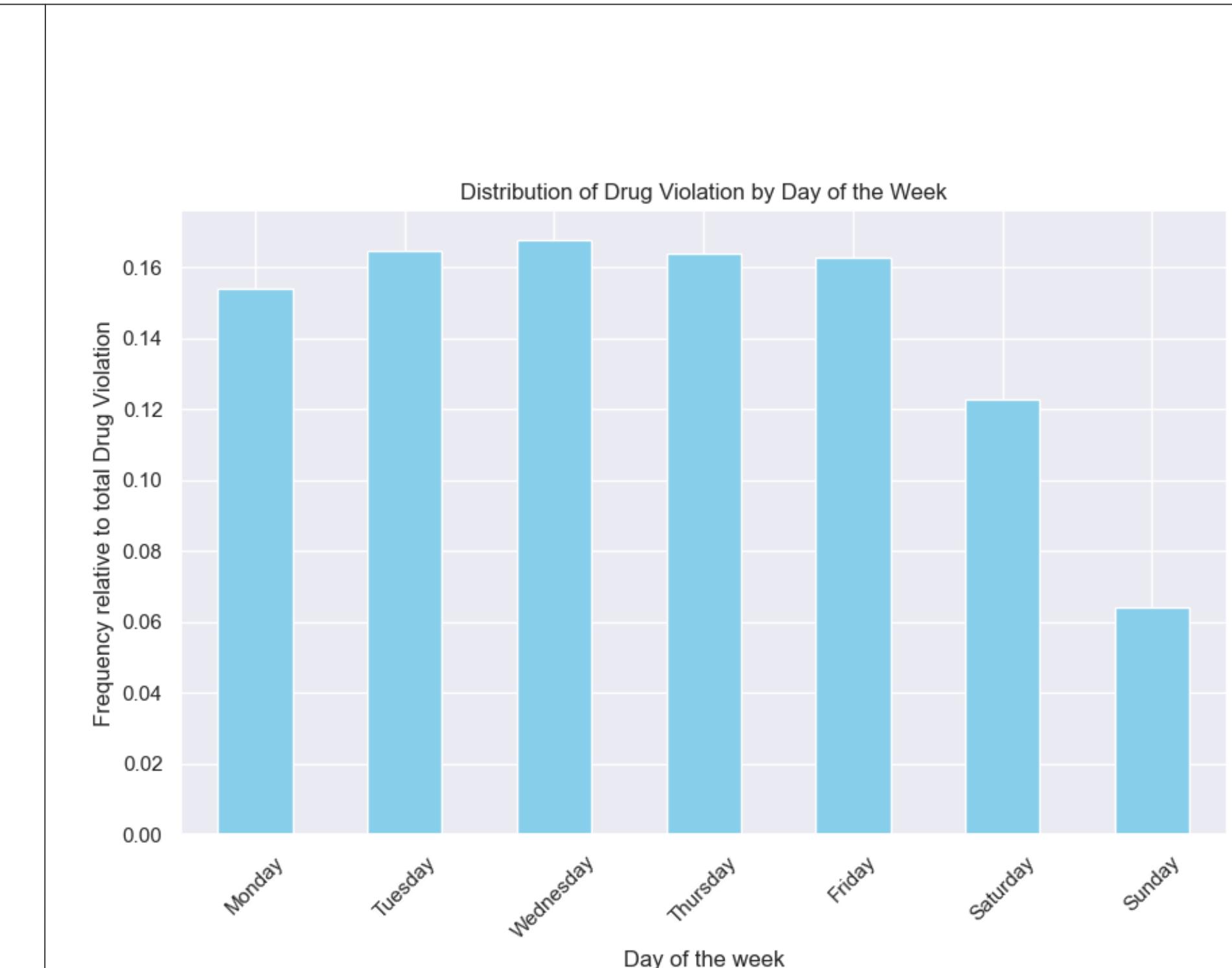
Frequency w.r.t. #crimes on specific years

Specific Crime Analysis

Drug Crime



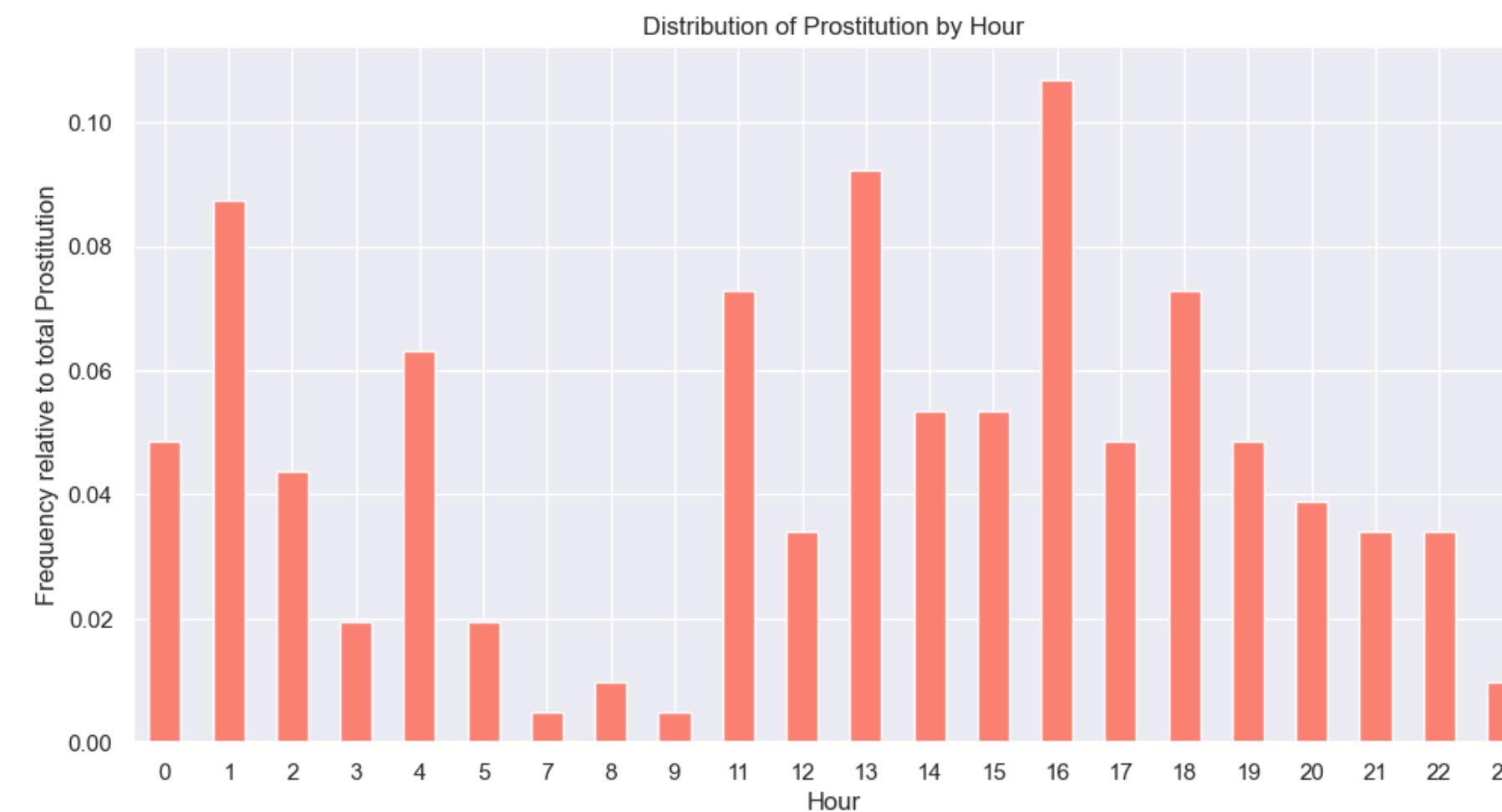
Frequency w.r.t. #crimes on specific years



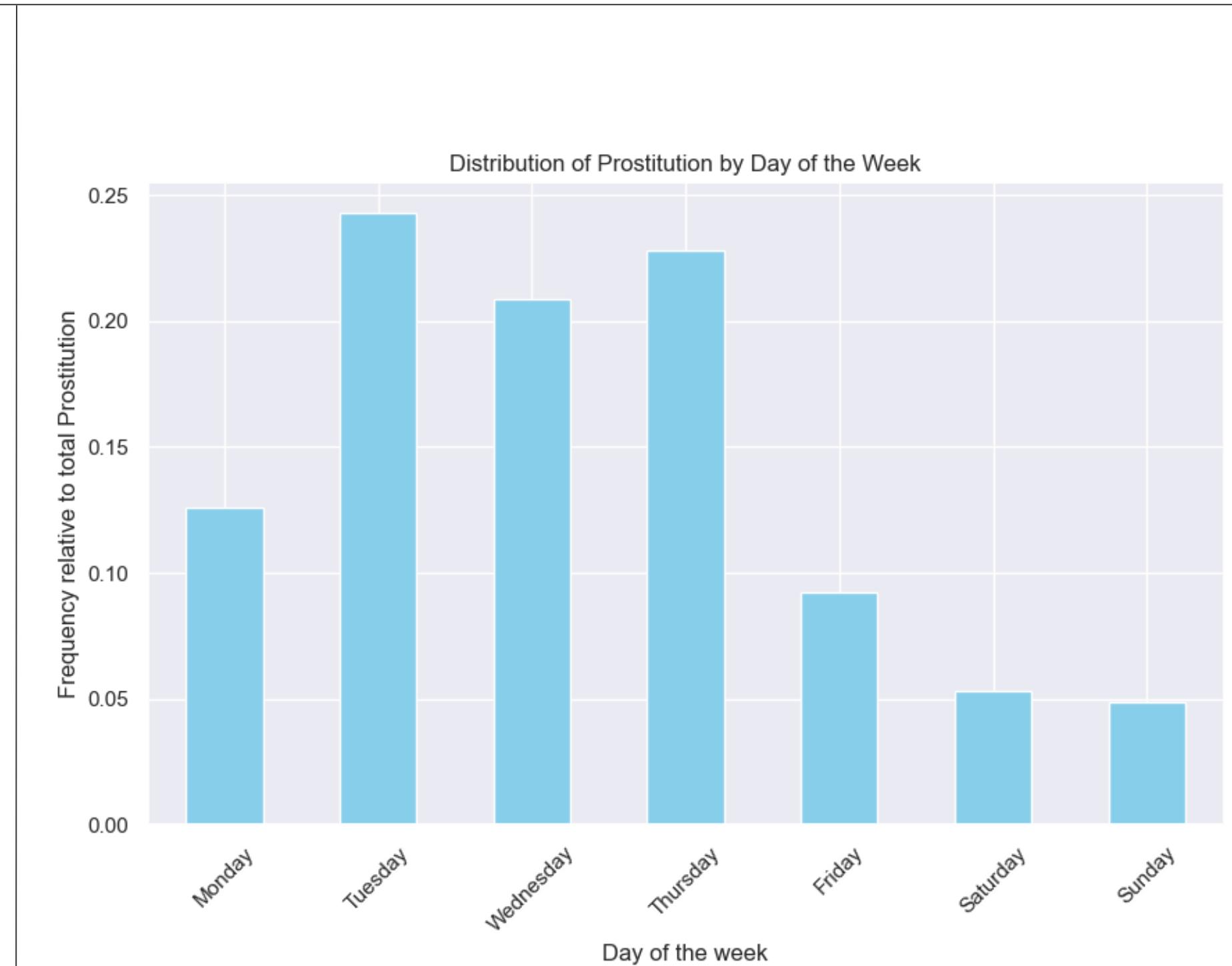
Drug crimes distribution by day

Specific Crime Analysis

Prostitution



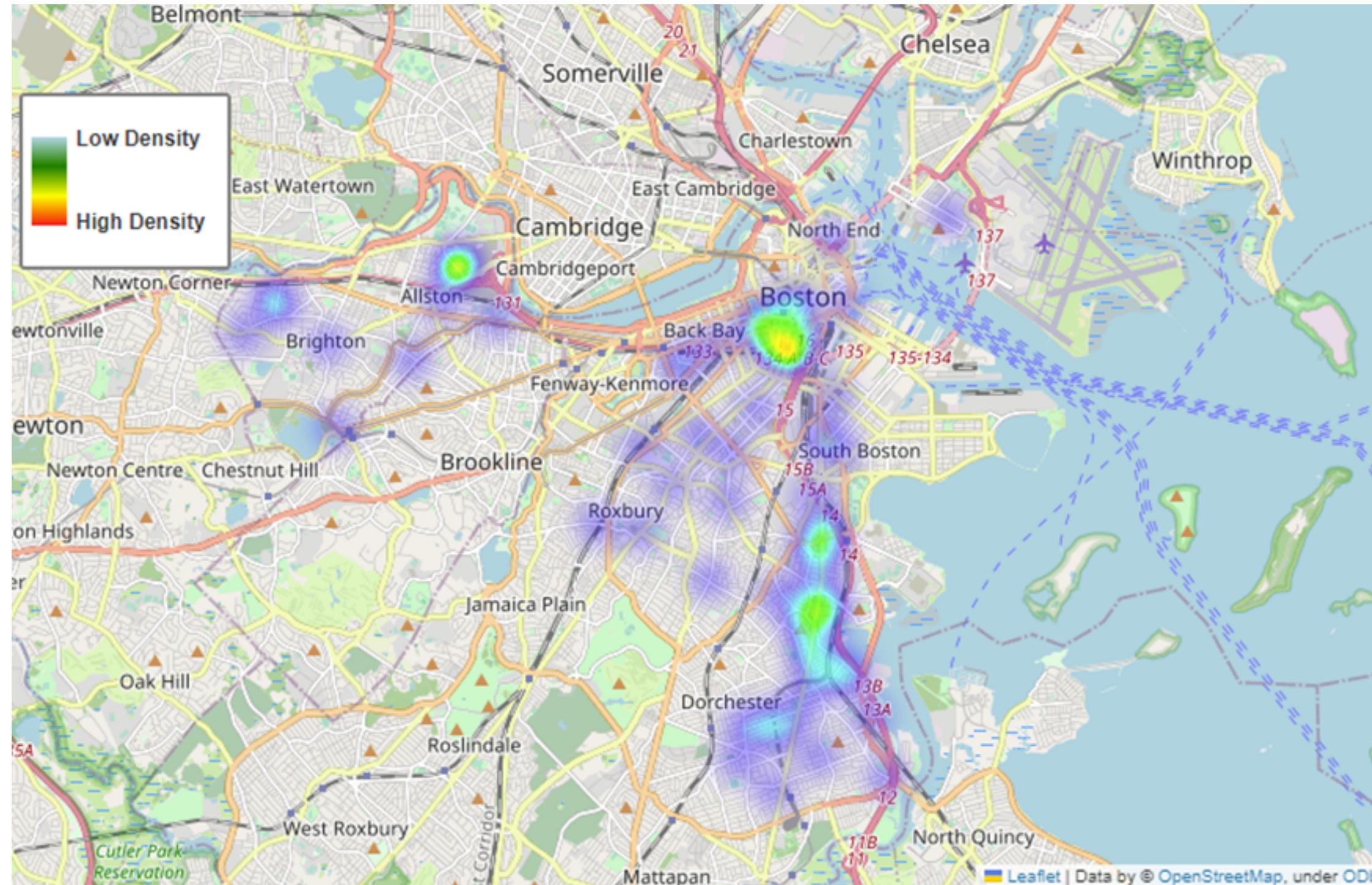
Prostitution distribution by hour



Prostitution distribution by day

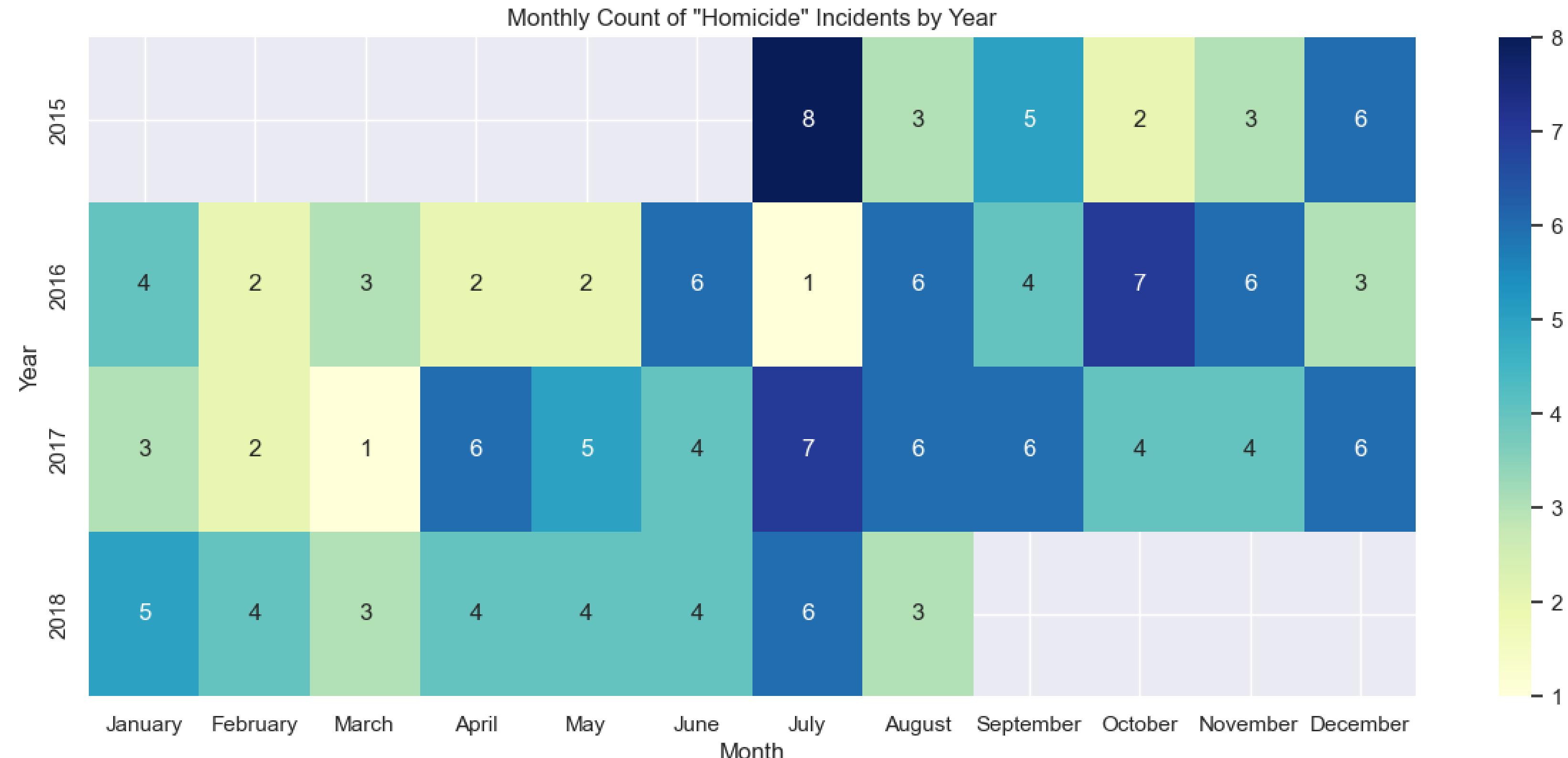
Specific Crime Analysis

Prostitution



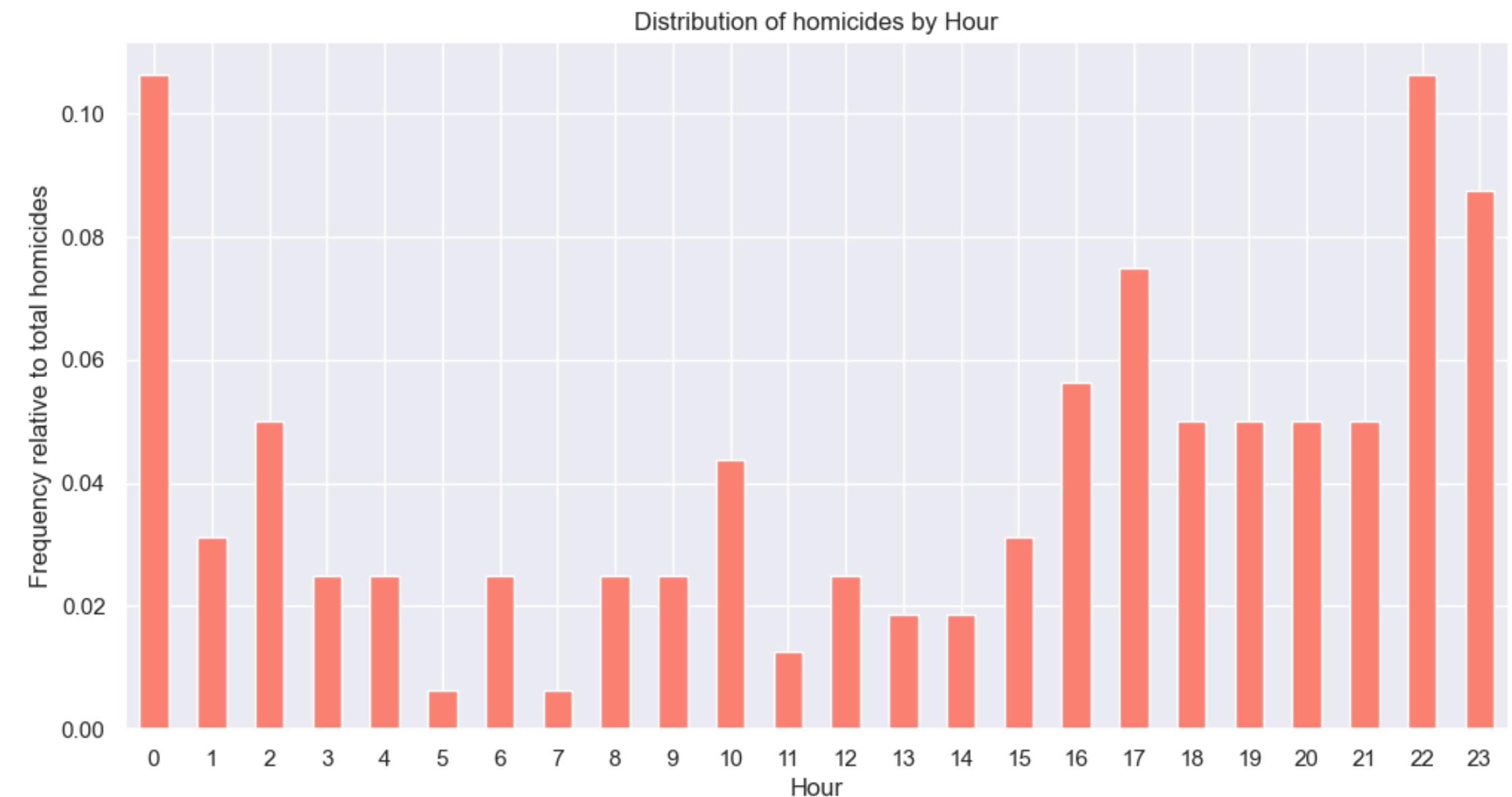
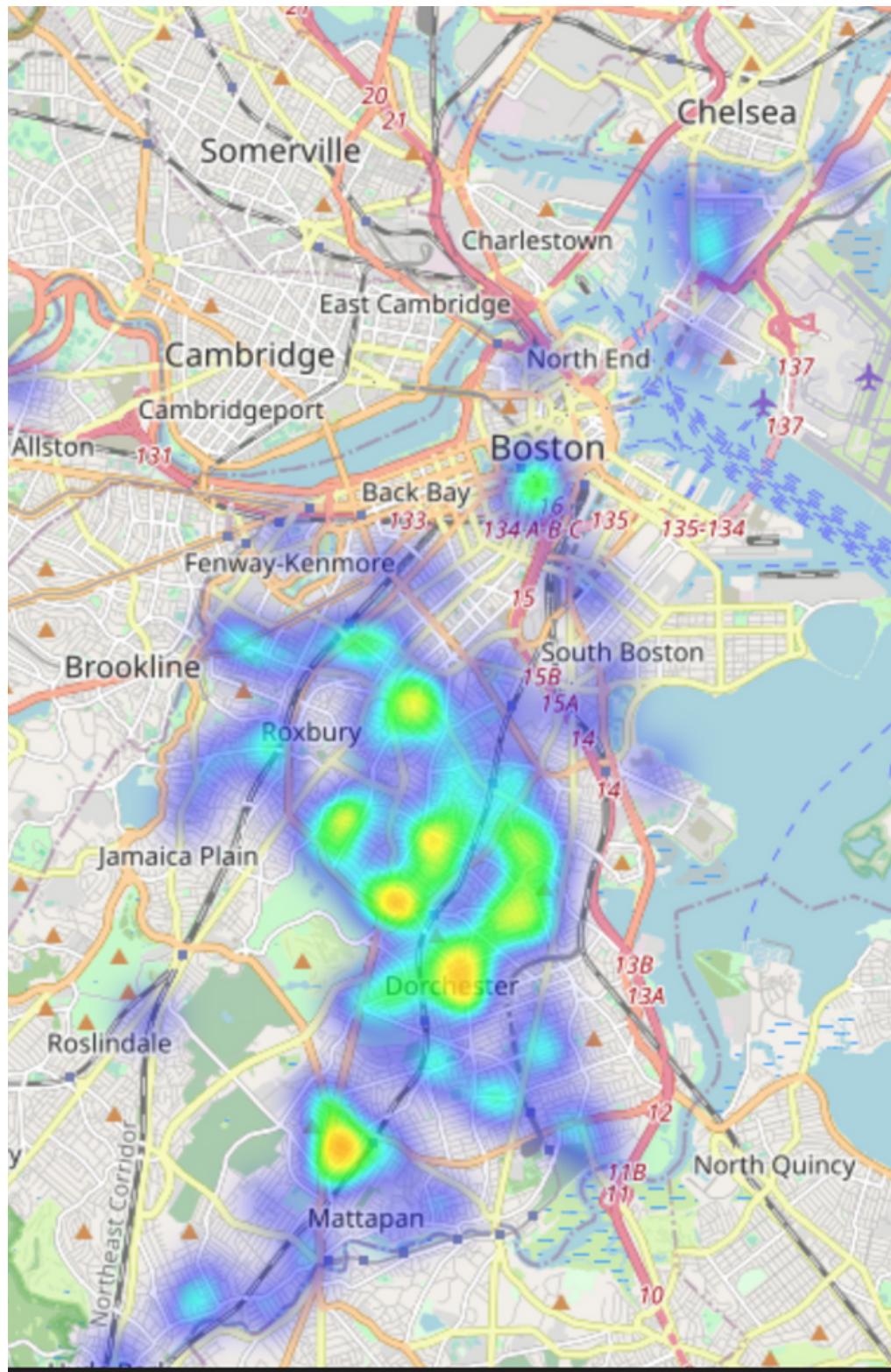
Specific Crime Analysis

Homicide



Specific Crime Analysis

Homicide



Homicide distribution by hour

Findings and Conclusion

1. Critical Emphasis:

- Crime analysis requires understanding socioeconomic, demographic, and cultural context.
- Lack of additional data on population, income, education, unemployment hampers accurate conclusions.

2. Importance of Contextual Data:

- Detailed demographic data (age, gender) is crucial for a comprehensive understanding of crime landscape.

3. Analysis as a Starting Point:

- Current analysis provides a foundation.
- Emphasizes need for further research with broader contextual data.

4. Call for Additional Research:

- Highlight the necessity for sociological, economic, and demographic analyses.
- Identify more significant correlations and trends.

5. Conclusion:

- Data collected is a starting point.
- Solid foundation for future criminology and crime prevention research in Boston.

6. Potential Impact:

- Work could lead to more detailed analyses.
- Facilitate the development of targeted intervention strategies.

7. Final Message:

- Stress the role of ongoing research in improving safety and quality of life in the city.

Thank you for your attention!

Do you have any question?

