LOS ANGELES SAVE BIKE LANES

Martin Ledl 108012012
Pawel Urbanowicz 108015016

TEAM MEMBERS

Martin Ledl	108012012
Pawel Urbanowicz	108015016

OUTLINE

- Introduction
- Datasets
- Data Preprocessing
- Data Analysis
- Conclusion

INTRODUCTION

Goal:

Provide an analysis of safety of bike lanes in Los Angeles

- Analysis of most frequent committed crimes
- Analyse which crimes might affect bikers
- Timely analysis of frequent committed crimes
- Analysis of shared bike trips
- Analysis of provided bikelanes
- Combine results to extract pattern

DATASETS

- Arrest Dataset
- Metro Bike Share Trip Dataset
- Bike Lanes Dataset

ARREST DATASET

Source

- https://data.lacity.org/A-Safe-City/Arrest-Data-from-2010-to-Present/yru6-6re4

Stats

- Owner: LAPD
- Refresh rate: Weekly
- Columns: 17
- Rows: 1.32M
- Each row represents the booking of an arrestee

METRO BIKE SHARE TRIP DATASET

Source

- https://data.lacity.org/A-Livable-and-Sustainable-City/Metro-Bike-Share-Trip-Data/sii9-rjps

Stats

- Owner: Mayor's Office
- Refresh rate: Quarterly
- Columns: 1
- Rows: 132K
- Each row represents a trip with a shared bike

BIKE LANES DATASET

Source

- https://data.lacity.org/A-Livable-and-Sustainable-City/Bikelanes/uzvv-a9xz

Stats

- Owner: Department of Transportation
- Refresh rate: Not specified
- Feature Collection of different named features that describe the different bikelanes
- Current Number of Features: 806

UNCLEANED DATASETS

Arrest:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1313405 entries, 0 to 1313404
Data columns (total 17 columns):
Report ID                    1313405 non-null int64
Arrest Date                  1313405 non-null object
Time                         1313209 non-null float64
Area ID                      1313405 non-null int64
Area Name                    1313405 non-null object
Reporting District           1313405 non-null int64
Age                          1313405 non-null int64
Sex Code                     1313405 non-null object
Descent Code                 1313405 non-null object
Charge Group Code            1225010 non-null float64
Charge Group Description     1224486 non-null object
Arrest Type Code             1313405 non-null object
Charge                       1313405 non-null object
Charge Description           1225060 non-null object
Address                      1313405 non-null object
Cross Street                 749250 non-null object
Location                     1313405 non-null object
dtypes: float64(2), int64(4), object(11)
memory usage: 170.3+ MB
```

Metro Bike Share Trip:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 132427 entries, 0 to 132426
Data columns (total 16 columns):
Trip ID                      132427 non-null object
Duration                     132427 non-null object
Start Time                   132427 non-null object
End Time                     132427 non-null object
Starting Station ID          132408 non-null object
Starting Station Latitude    132379 non-null float64
Starting Station Longitude   132379 non-null float64
Ending Station ID            132331 non-null object
Ending Station Latitude      131376 non-null float64
Ending Station Longitude     131376 non-null float64
Bike ID                      132417 non-null object
Plan Duration                131661 non-null float64
Trip Route Category          132427 non-null object
Passholder Type              132427 non-null object
Starting Lat-Long            98622 non-null object
Ending Lat-Long              131376 non-null object
dtypes: float64(5), object(11)
memory usage: 16.2+ MB
```

DATA PREPROCESSING

DATA PREPROCESSING

Extract the important features

- Delete unnecessary dimensions
- Can fix many of the missing values

Essential Features

- Drop data points (rows) with missing values
- If there are too many missing → replace with default value

Non-Essential Features

- Replace with default value
- E.g. ‘Cross Street’ missing values → ‘Unknown’

Combine/Split Features

- Combine ‘Arrest Date’ and ‘Time’ features into ‘Arrest Date’ feature
- Split ‘Location’ feature into ‘Latitude’ and ‘Longitude’ features

DATA PREPROCESSING

Arrest:

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1224346 entries, 2015-02-24 23:20:00 to 2019
Data columns (total 11 columns):
Report ID                   1224346 non-null int64
Arrest Date                 1224346 non-null datetime64[ns]
Area Name                   1224346 non-null object
Age                         1224346 non-null int64
Sex Code                    1224346 non-null object
Descent Code                1224346 non-null object
Charge Group Description    1224346 non-null object
Charge Description          1224346 non-null object
Address                     1224346 non-null object
Cross Street                1224346 non-null object
Location                    1224346 non-null object
dtypes: datetime64[ns](1), int64(2), object(8)
memory usage: 112.1+ MB
```

Metro Bike Share Trip:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 131336 entries, 25 to 132426
Data columns (total 9 columns):
Start Time                  131336 non-null object
End Time                    131336 non-null object
Starting Station ID         131336 non-null object
Starting Station Latitude   131336 non-null float64
Starting Station Longitude  131336 non-null float64
Ending Station ID           131336 non-null object
Ending Station Latitude     131336 non-null float64
Ending Station Longitude    131336 non-null float64
Trip Route Category         131336 non-null object
dtypes: float64(4), object(5)
memory usage: 10.0+ MB
```

DATA ANALYSIS

Arrest Dataset

- Analysis of most frequent committed crimes
- Analyse which crimes might affect bikers
- Timely analysis of frequent committed crimes

Metro Bike Share Trip Dataset

- Analysis of shared bike trips
- Analysis of provided bikelanes

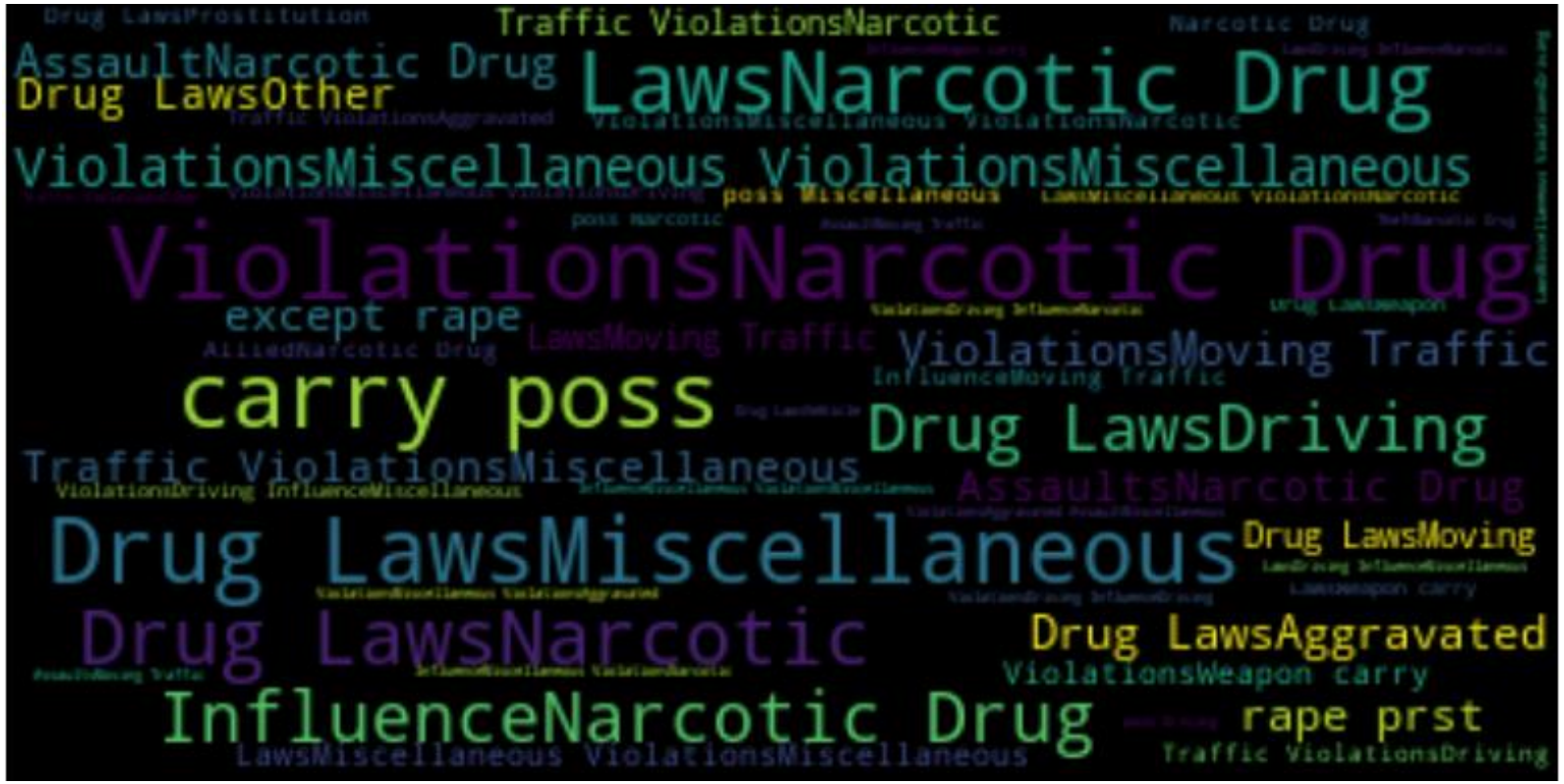
Bike Lane Dataset

- Extract bike lanes to observe

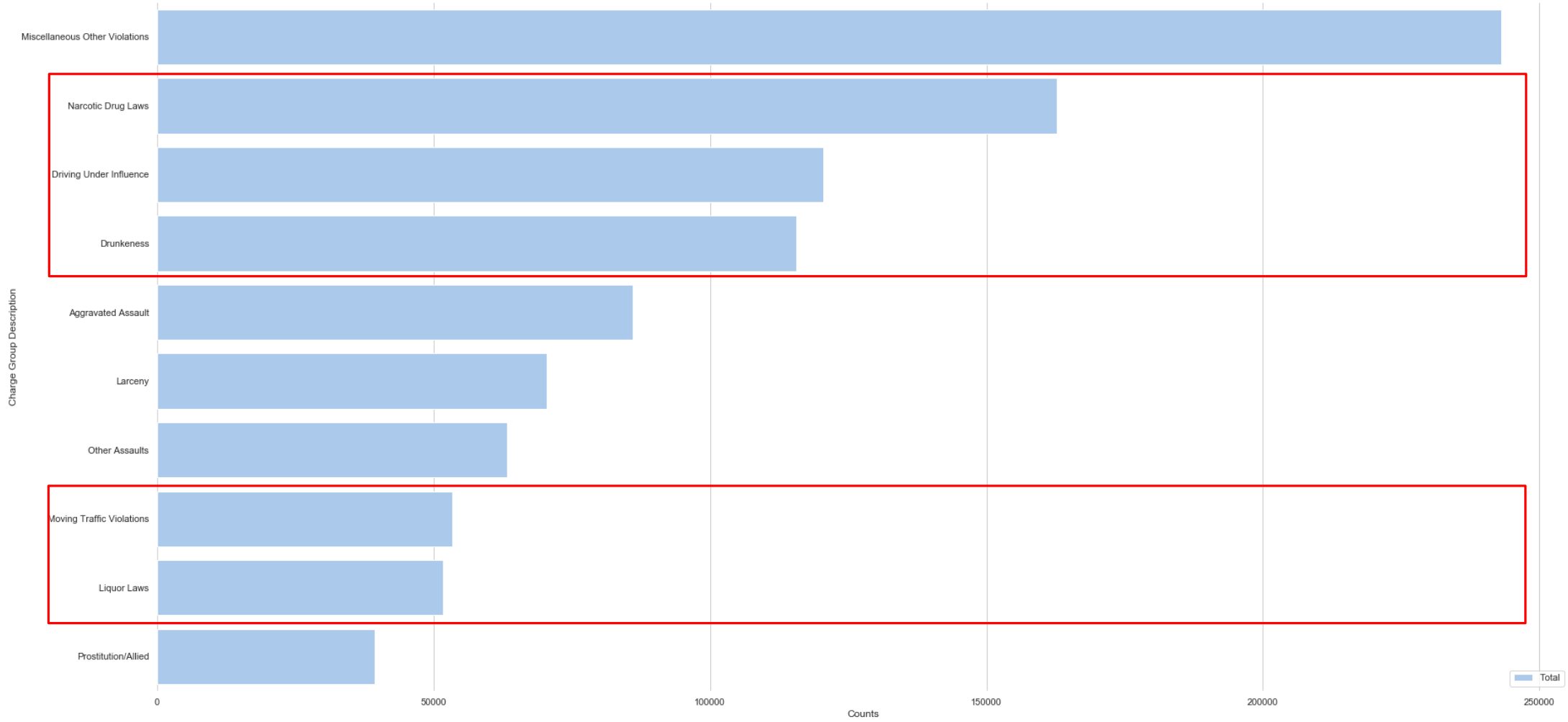
Integration of Analyzes

- Combine results to extract pattern

ARREST: CHARGE GROUP DESCRIPTION

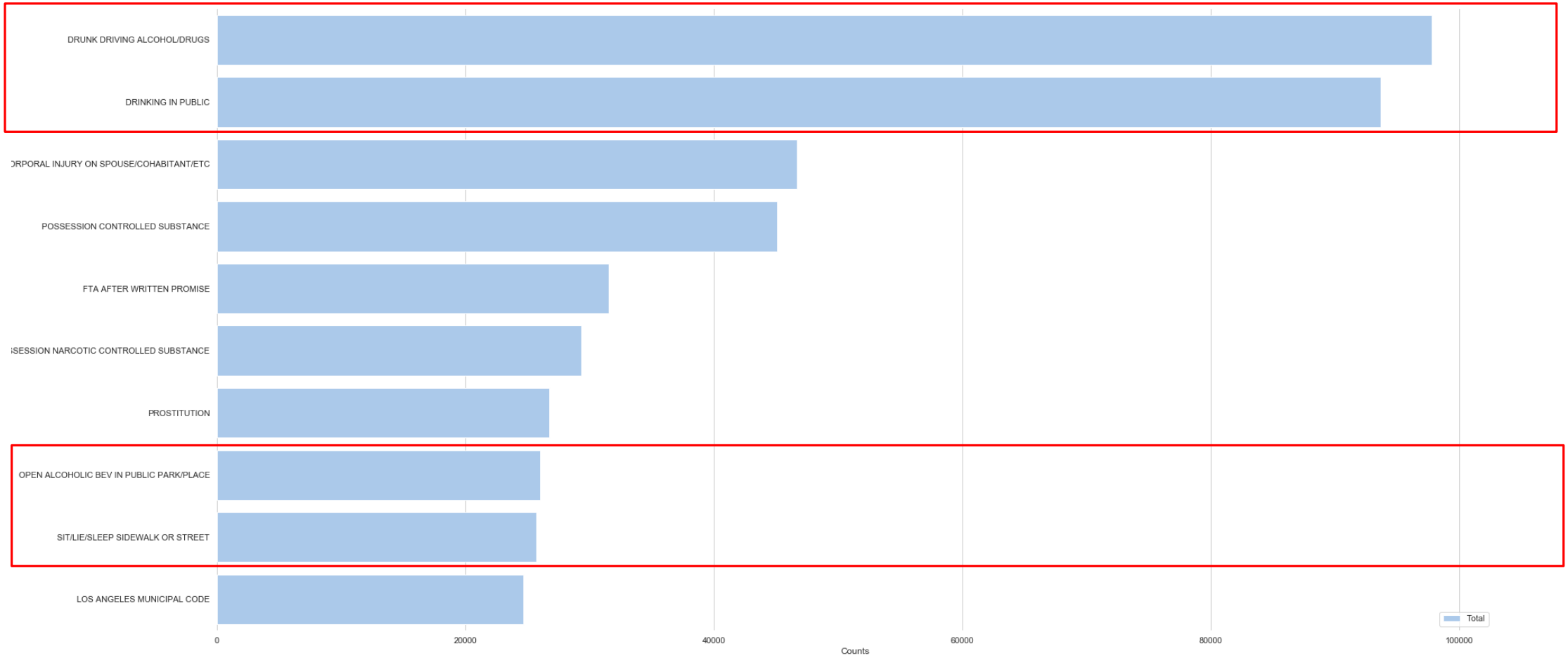


ARREST: CHARGE GROUP DESCRIPTION



ARREST: CHARGE DESCRIPTION

ARREST: CHARGE DESCRIPTION



ARREST: EXTRACT INTERESTING CHARGES

Extract the Charges/Crimes that possibly harm bikers

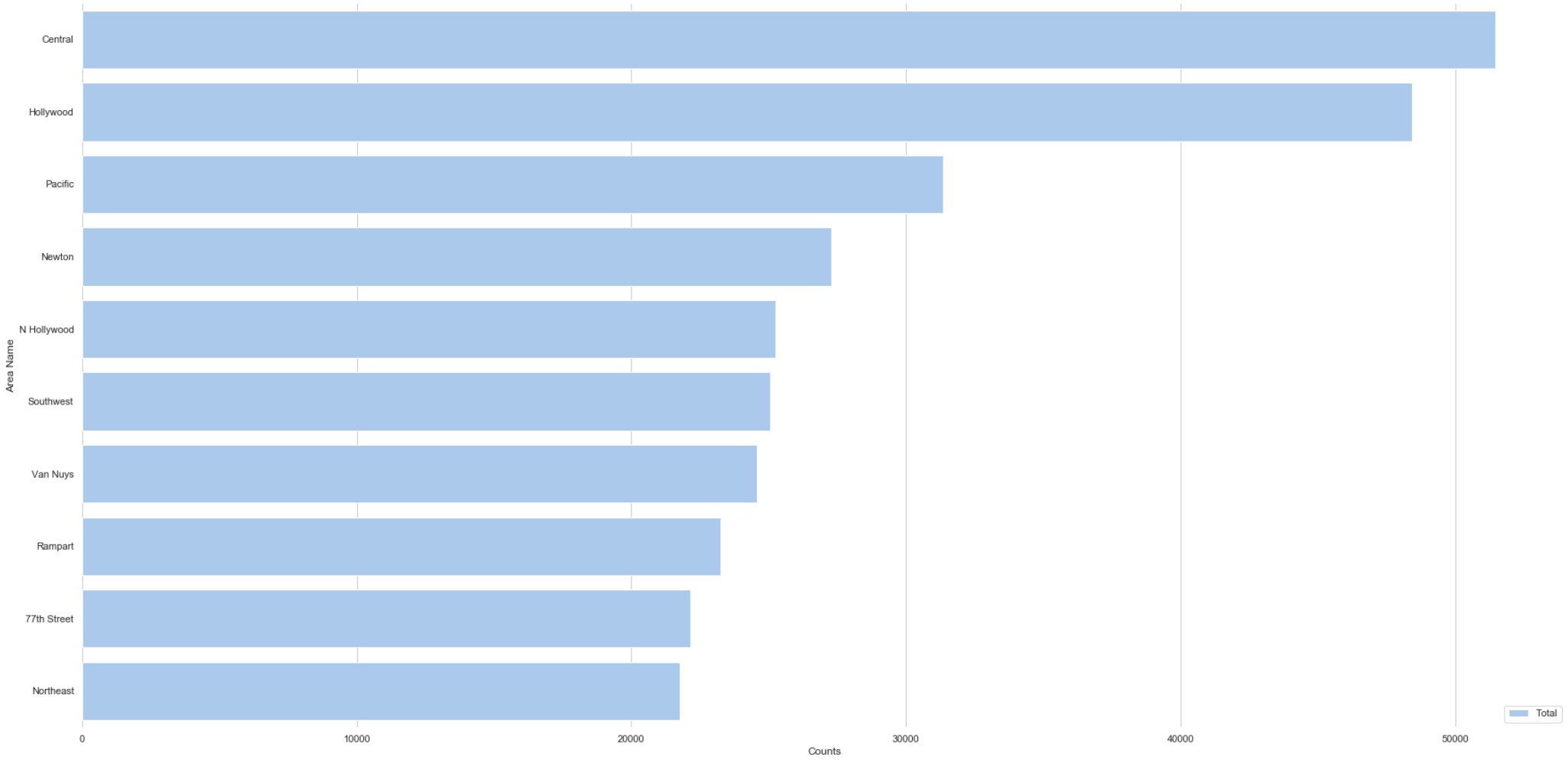
- Substrings within the most frequently committed crimes
- Generate subsets containing different substring
- Assembly those sets to one set of interest

Result

- Reduction from 2319 different Charges to 289 of interest
- 38.62% of entries are of interest!

ARREST: FREQUENT AREAS

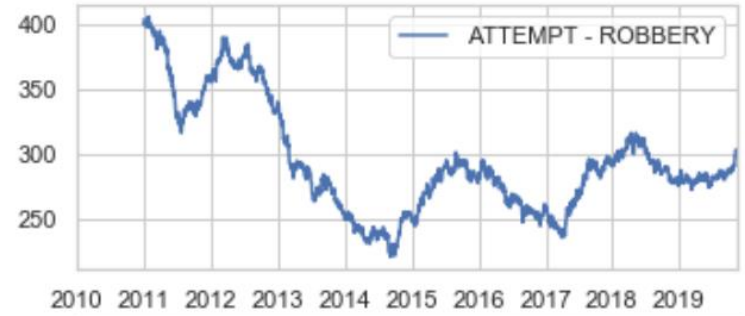
ARREST: FREQUENT AREAS

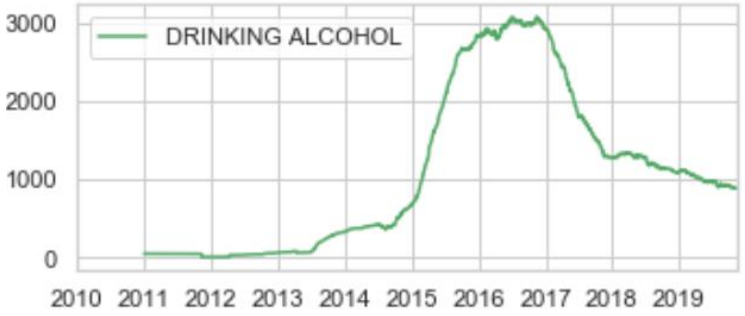


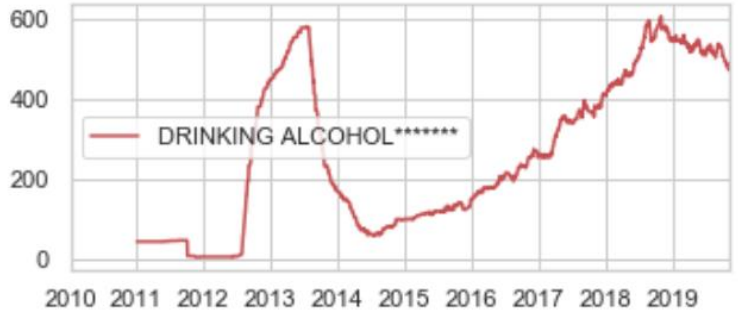
ARREST: FREQUENT ADDRESSES

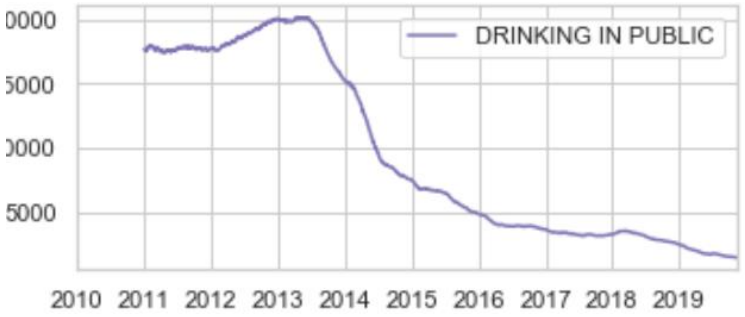
ARREST: EVOLUTION NUMBER OF CHARGES

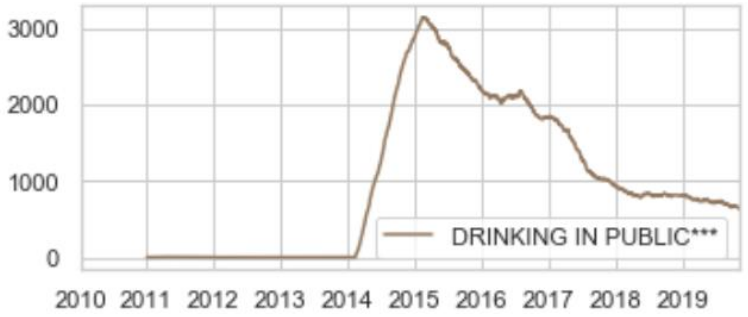
ARREST: TOP 30 CHARGES TRENDS





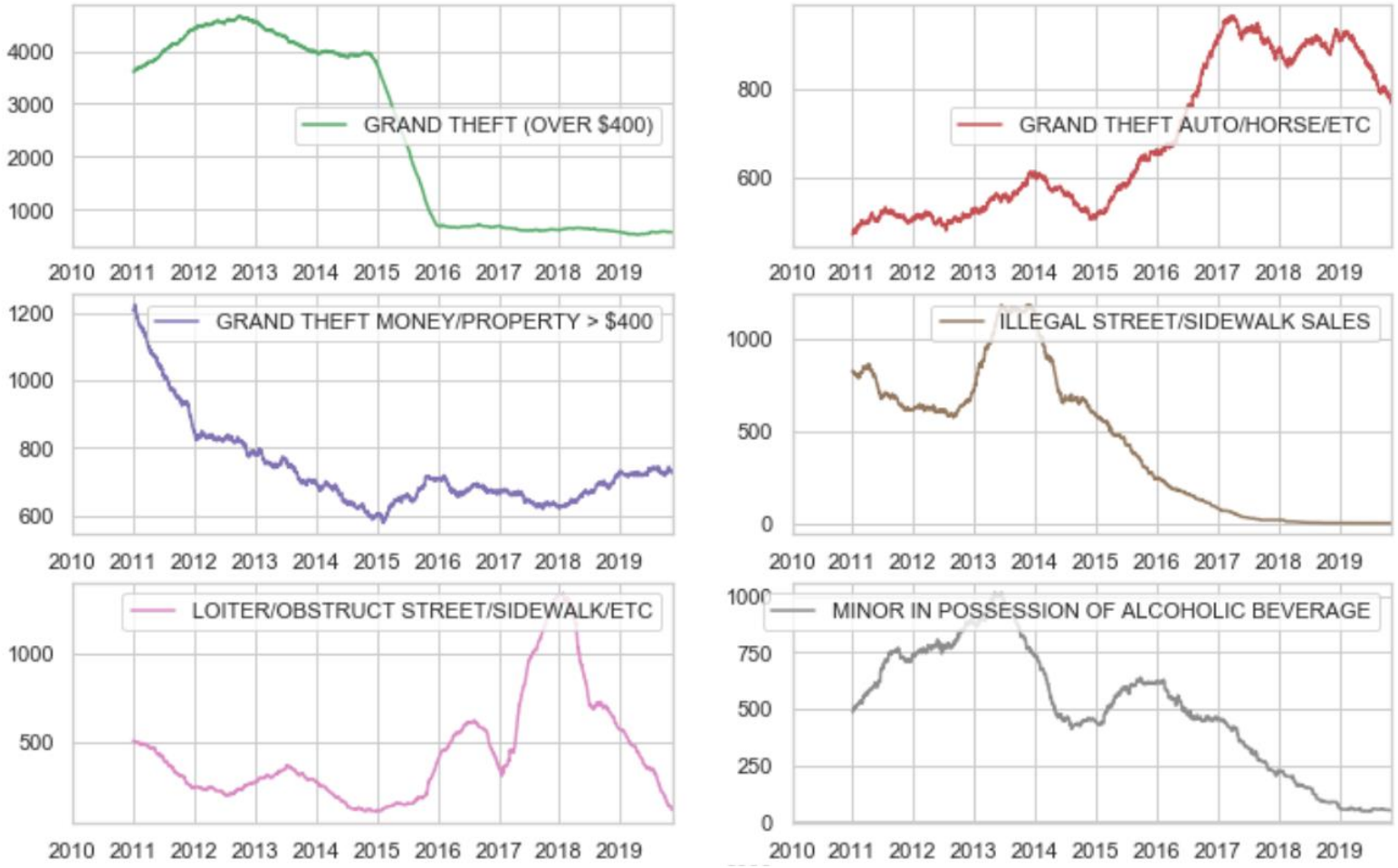




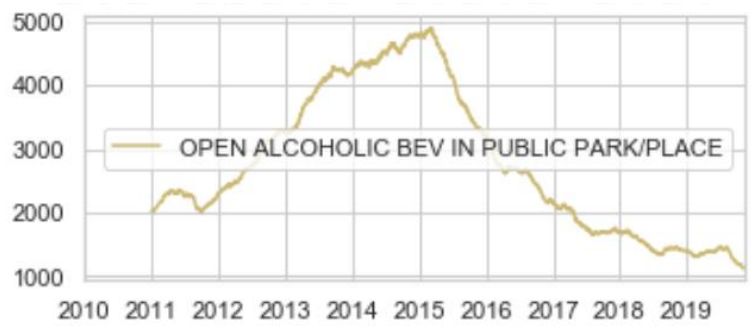


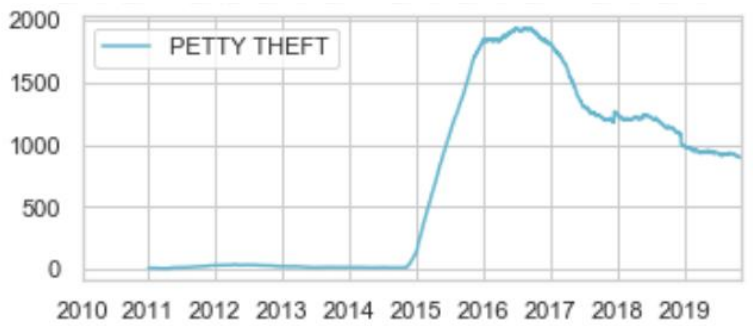
ARREST: TOP 30 CHARGES TRENDS

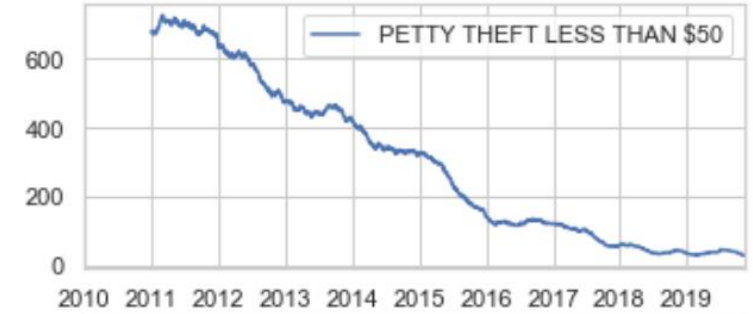
ARREST: TOP 30 CHARGES TRENDS

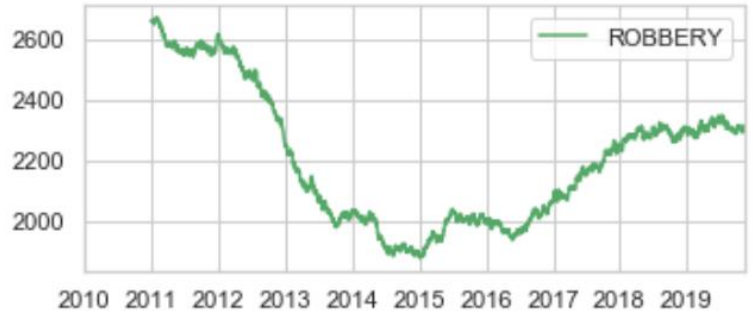


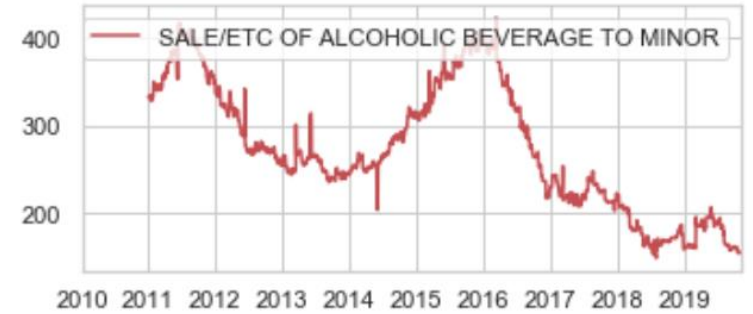
ARREST: TOP 30 FREQUENT CHARGES TRENDS



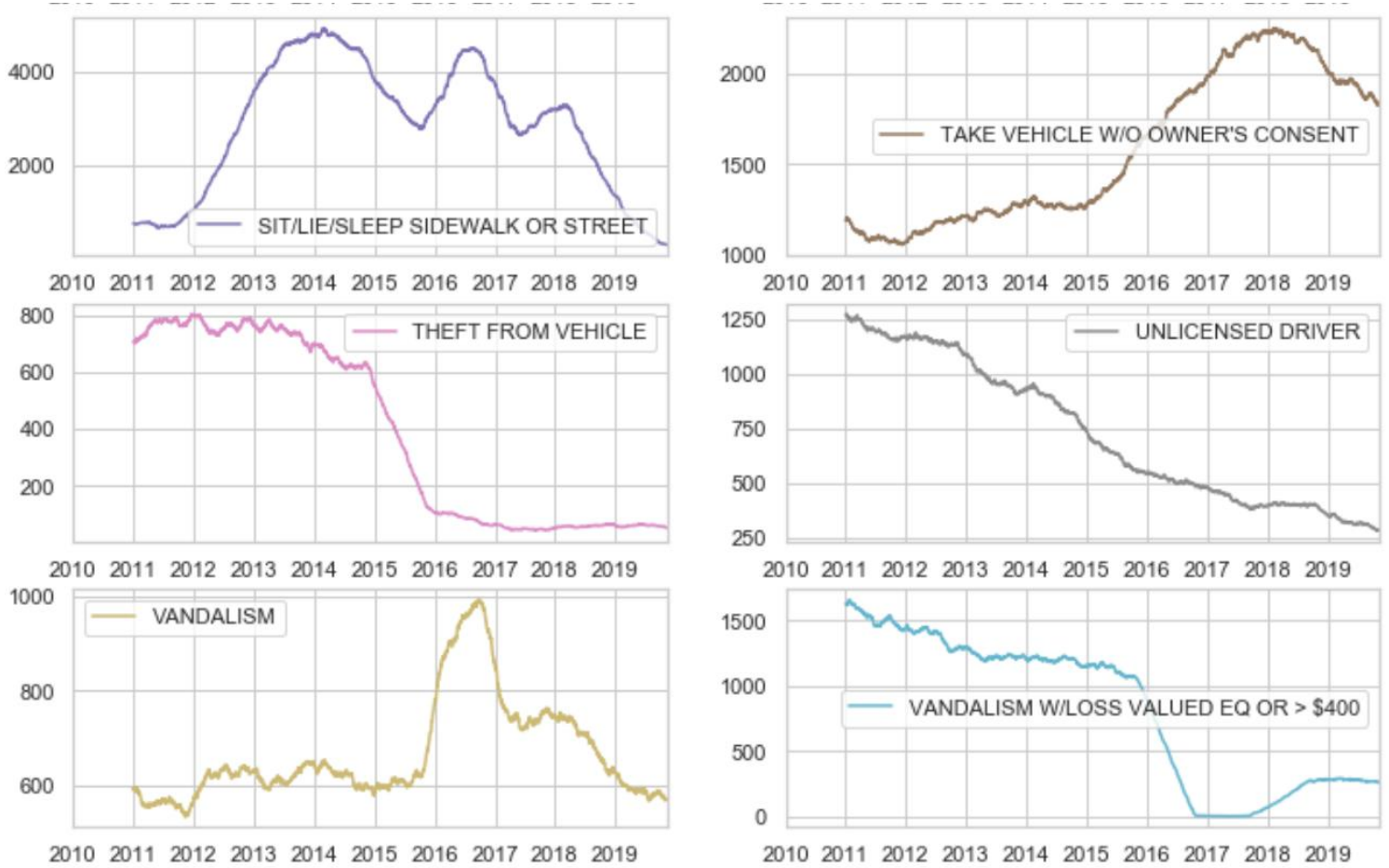




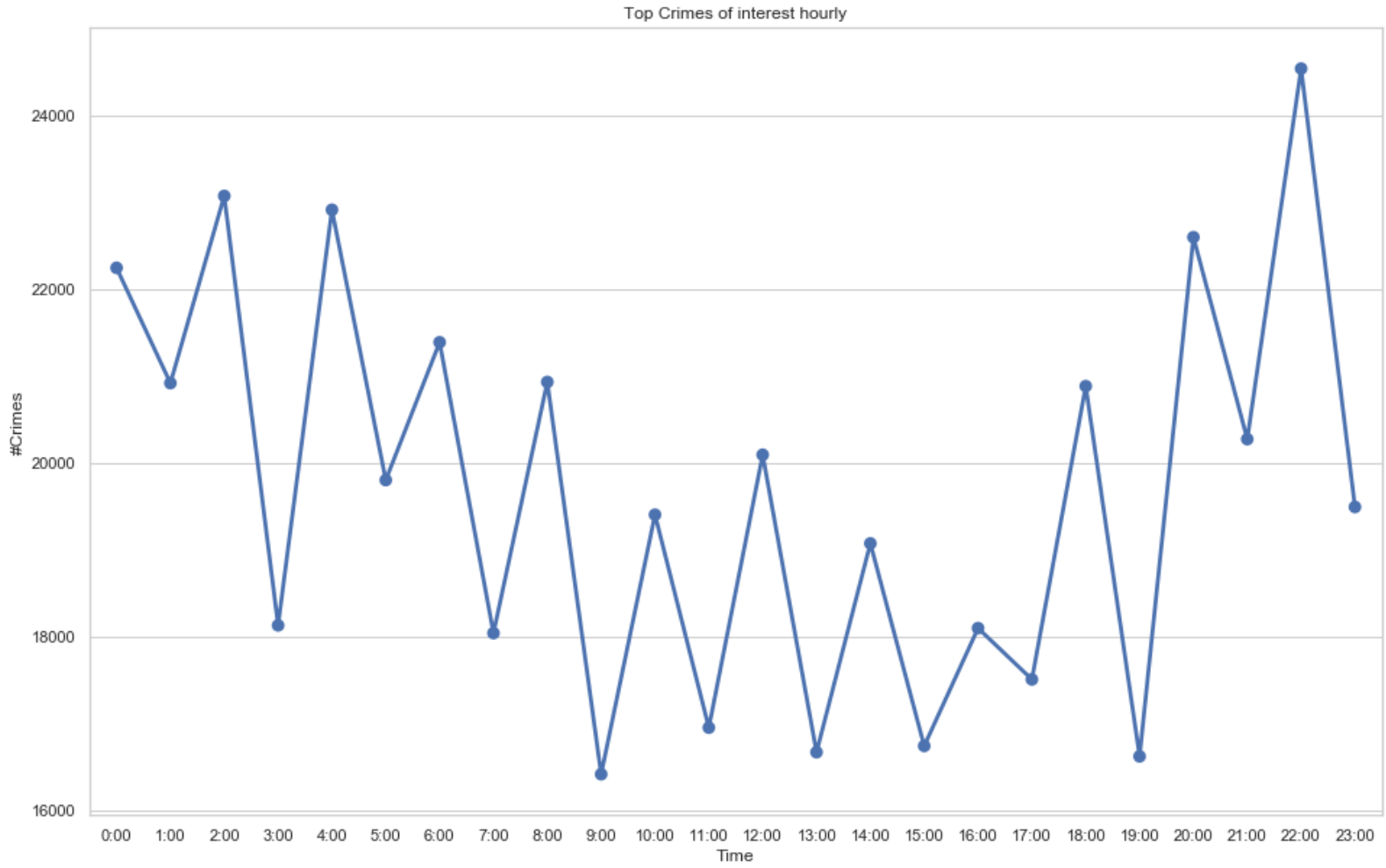




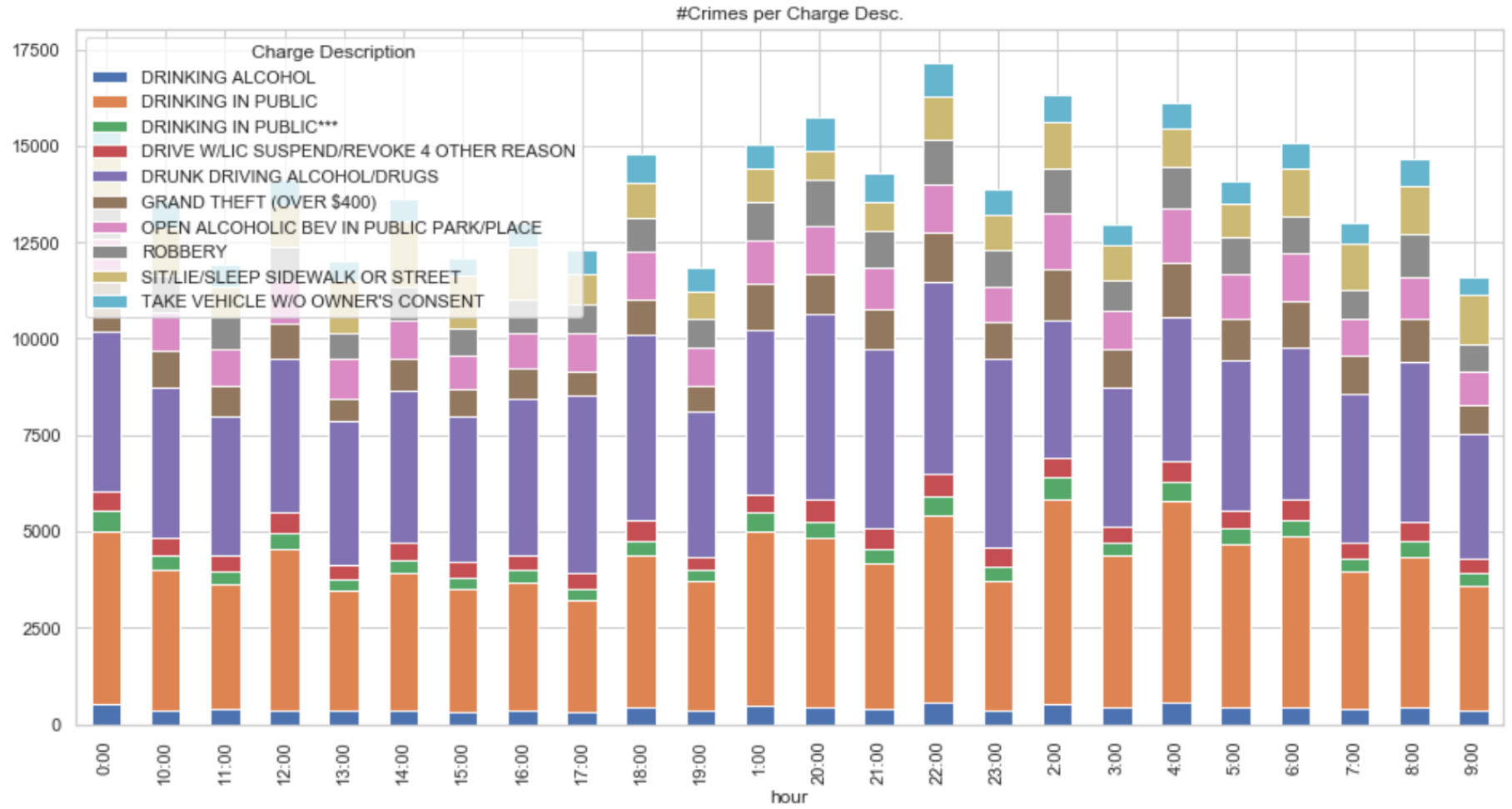
ARREST: TOP 30 CHARGES TRENDS



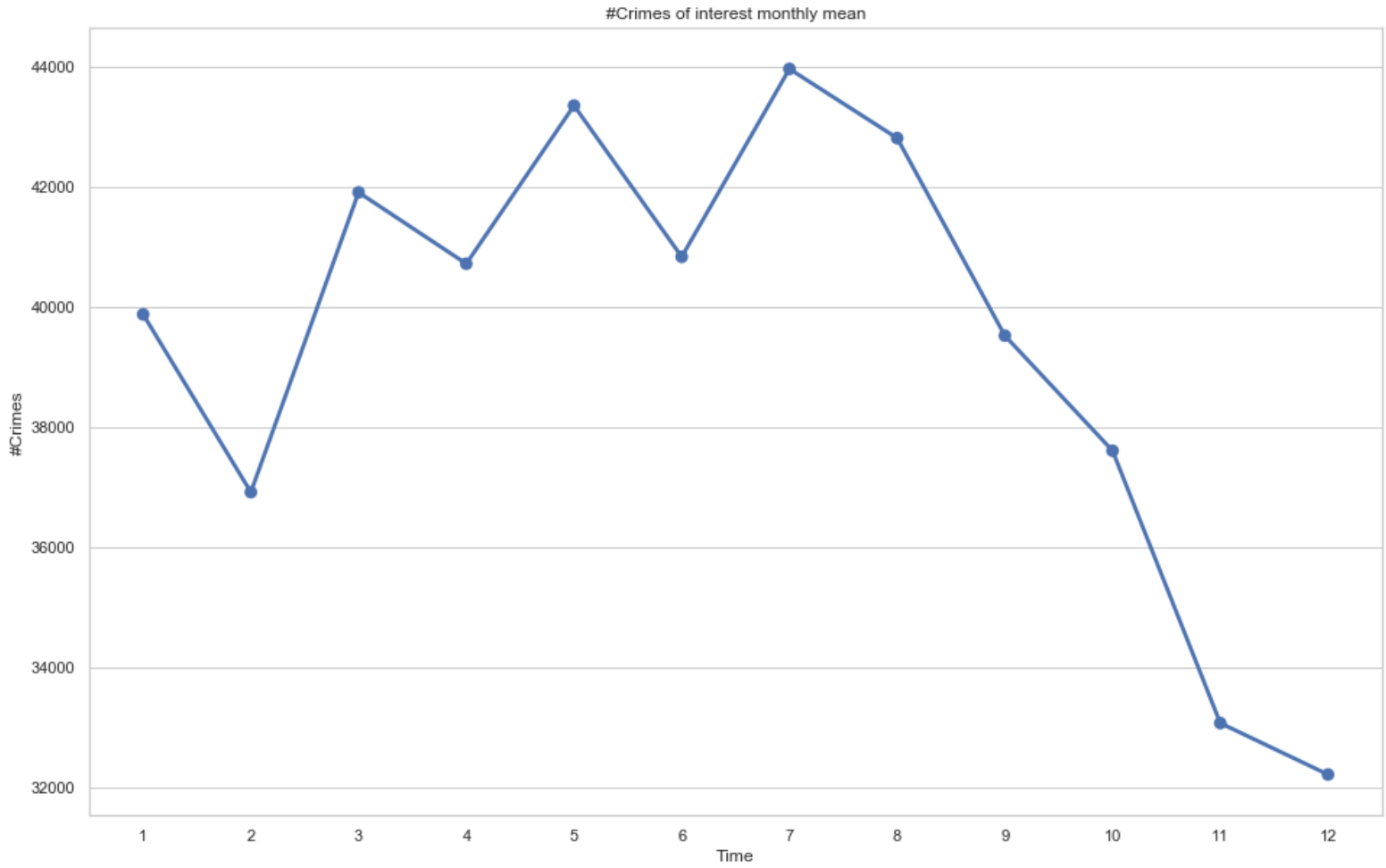
ARREST: FREQUENT CHARGES HOURLY



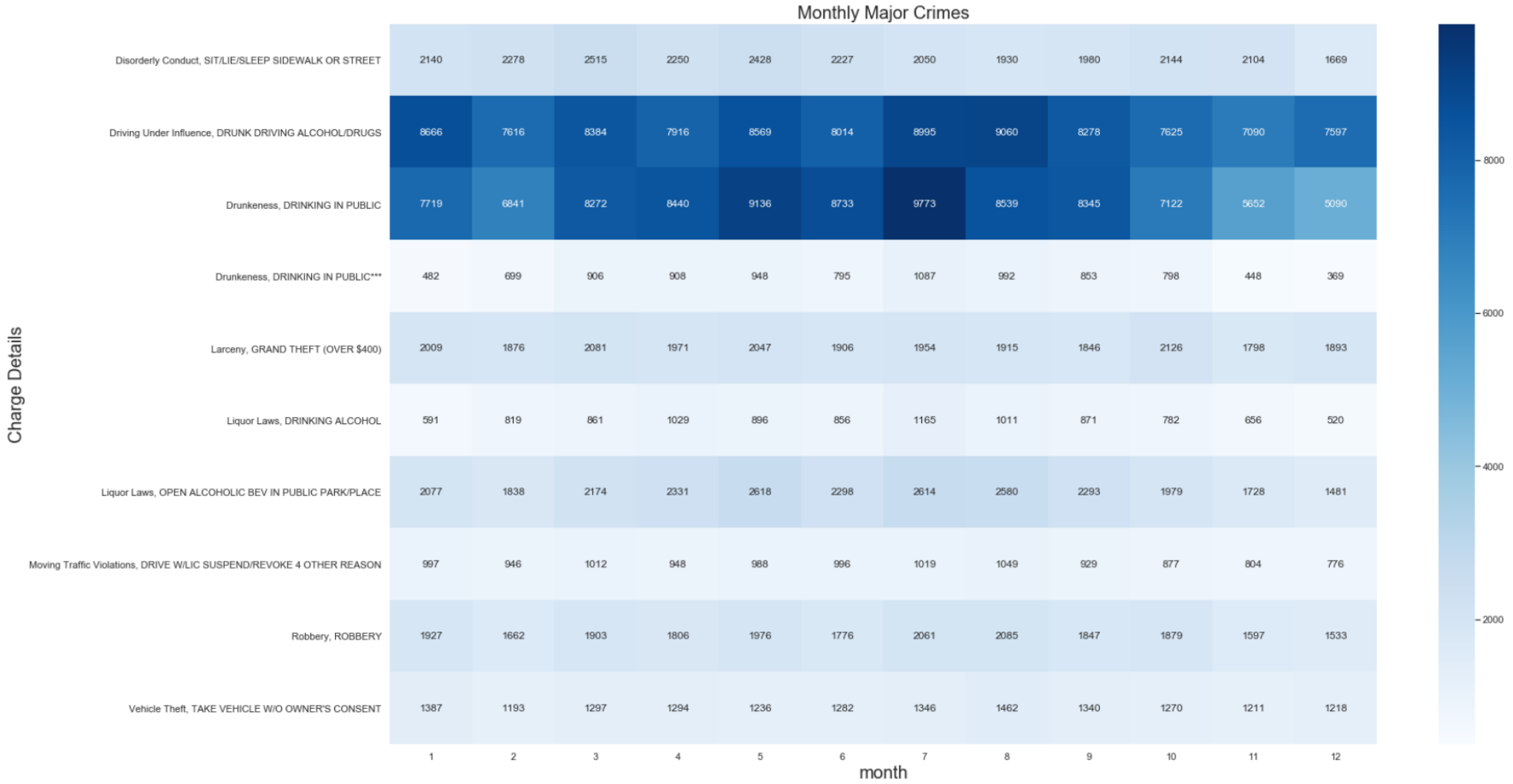
ARREST: TOP 10 FREQUENT CHARGES HOURLY



ARREST: FREQUENT CHARGES MONTHLY



ARREST: TOP 10 FREQUENT CHARGES MONTHLY

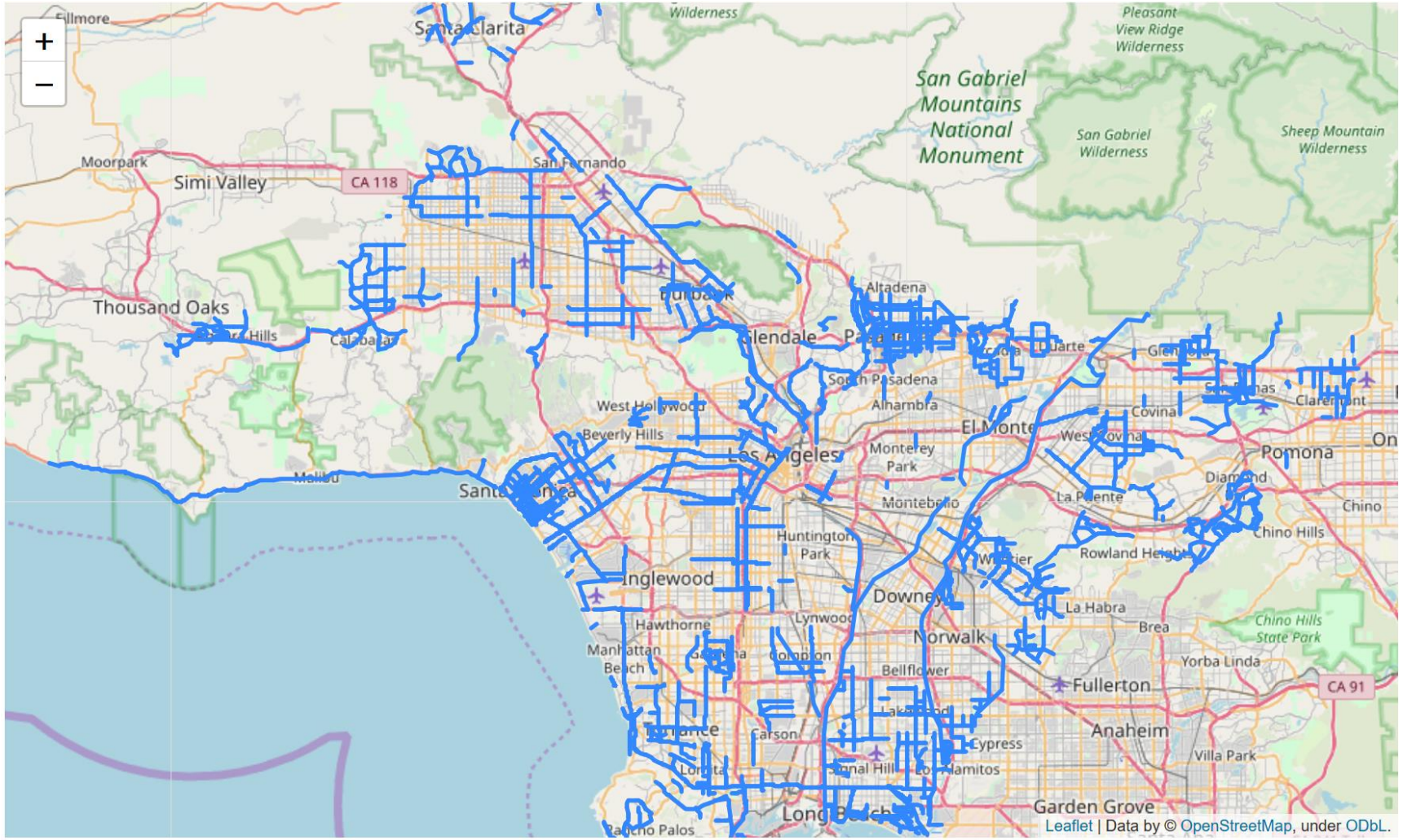


METRO BIKE SHARE TRIP

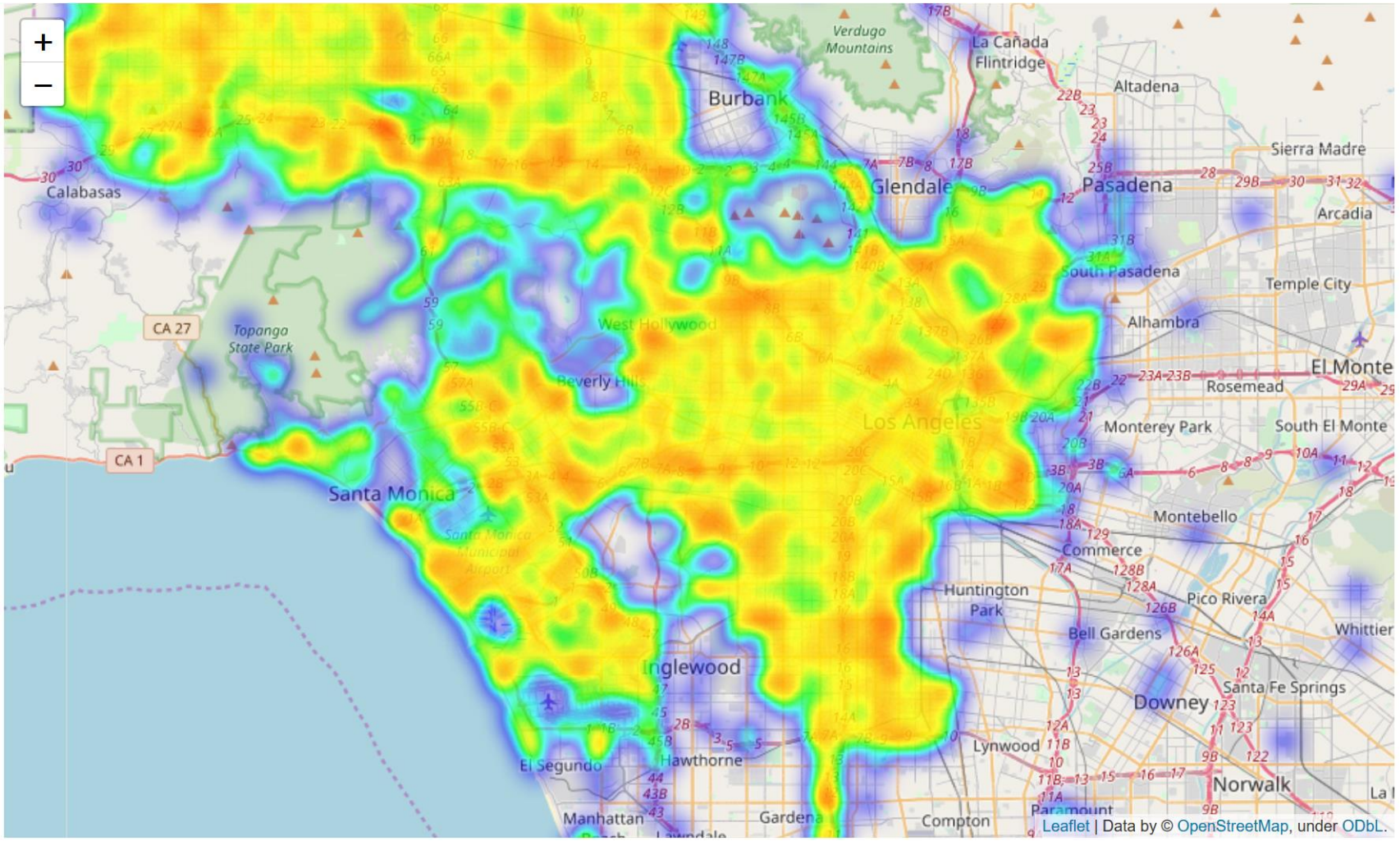
- Provides 131336 rides from start to end location
- Unique rides: 3580
- Top 10 popular rides:

```
3,030, 3,014    933
3,014, 3,030    676
3,031, 3,005    609
3,048, 3,048    569
3,005, 3,031    513
3,030, 3,042    506
3,022, 3,022    450
3,069, 3,069    440
3,082, 3,082    391
3,005, 3,034    389
Name: start_end, dtype: int64
```

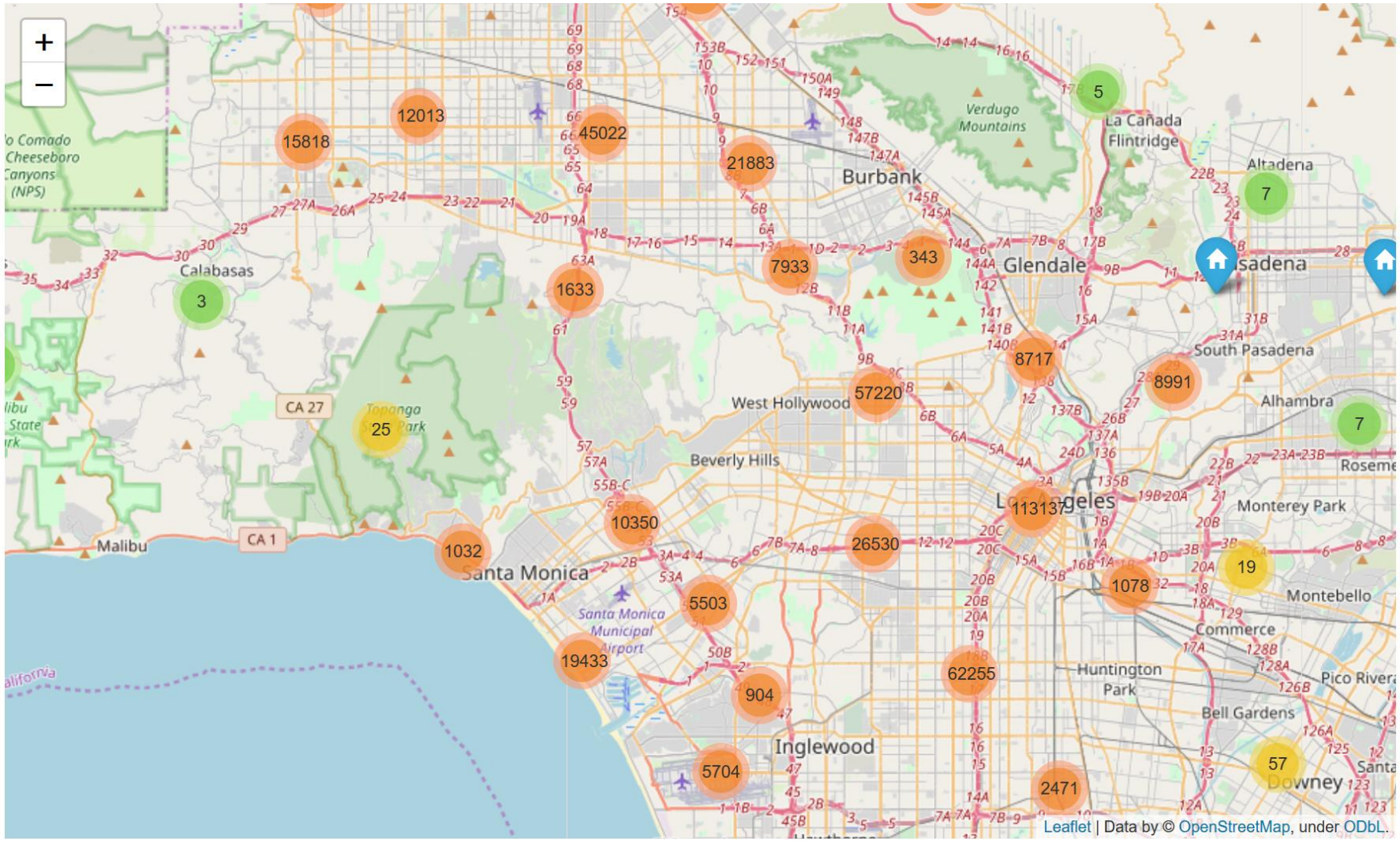
BIKE LANES: LANES MAP



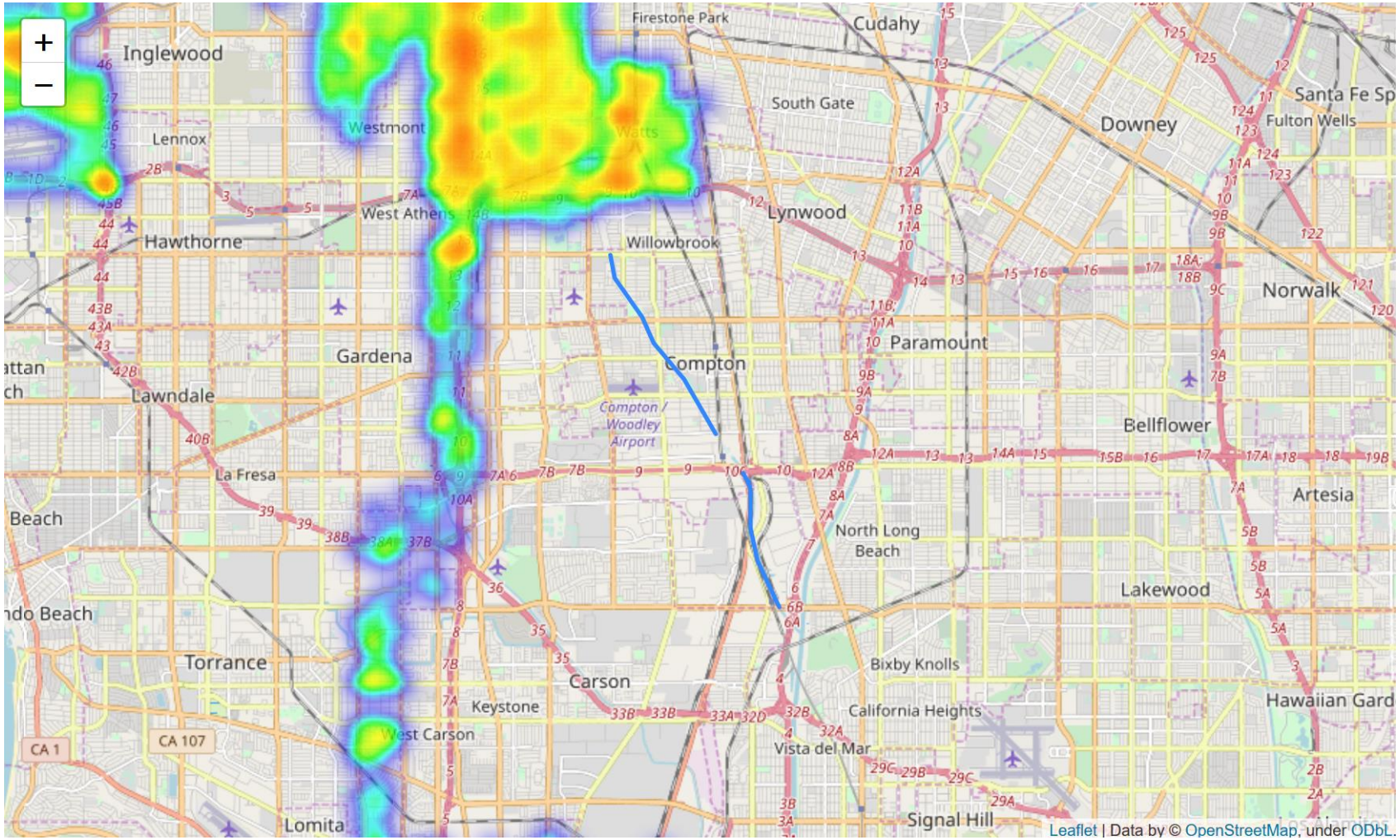
ARREST: INTERESTING CHARGES HEATMAP



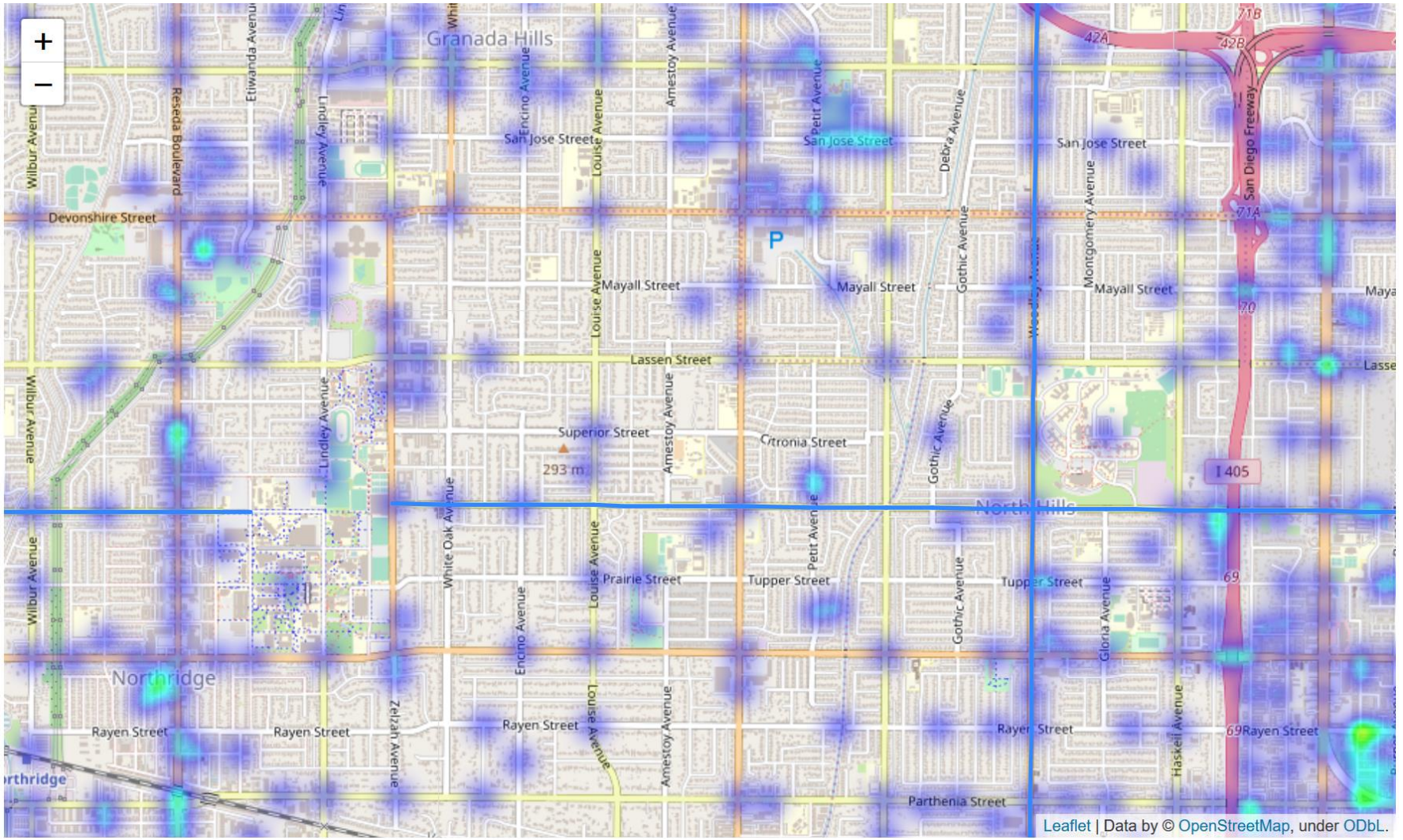
ARREST: INTERESTING CHARGES CLUSTERS



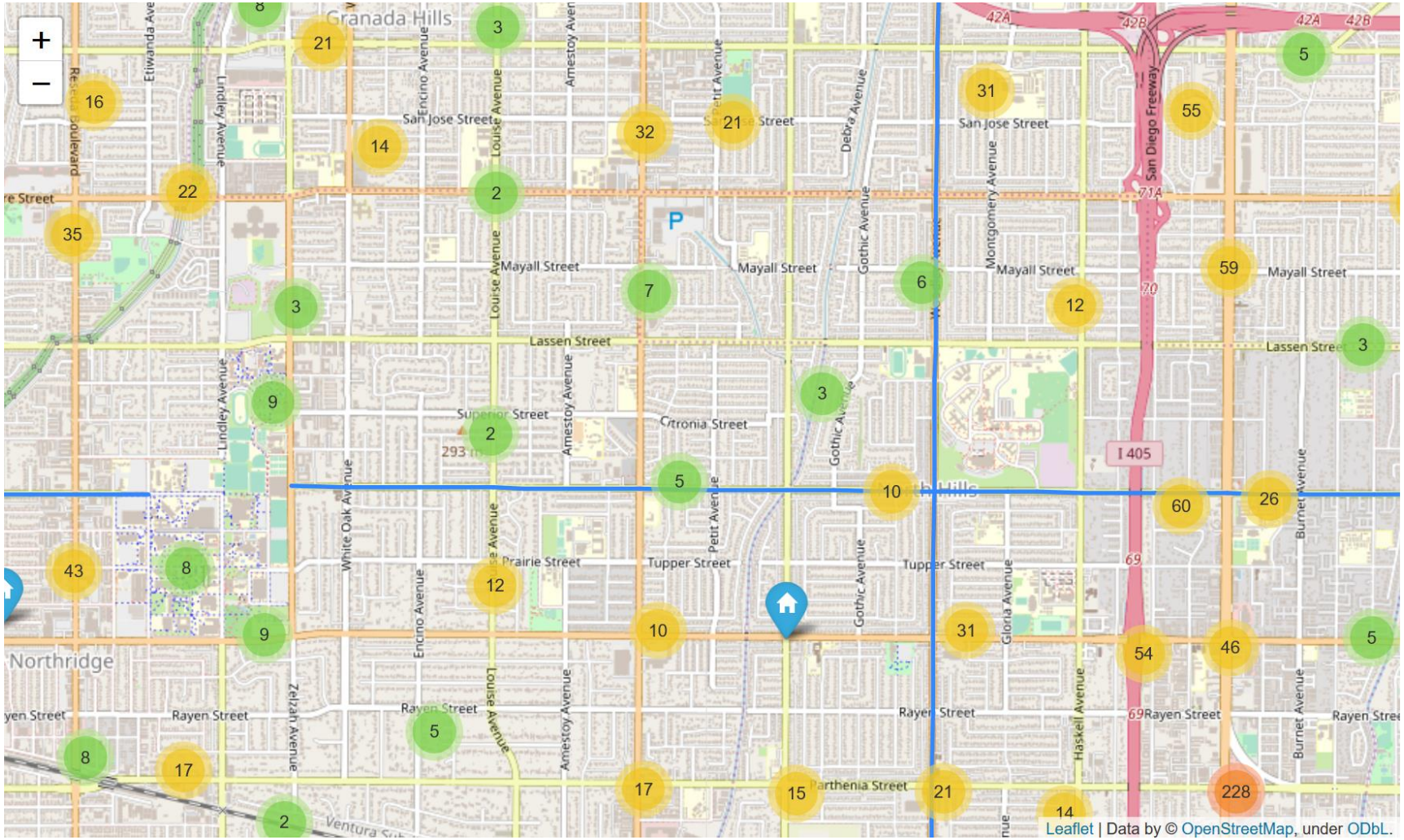
INTEGRATION: ANALYSE BIKE LANE



INTEGRATION: ANALYSE BIKE LANE



INTEGRATION: ANALYSE BIKE LANE



DEMO TIME

CONCLUSION

- Able to load specific bike lanes
- Overlay processed crime dataset
- Interactive map enables exploring
- Can be useful in combination with time analysis