



UNIVERSITY OF TORONTO  
SCHOOL OF CONTINUING STUDIES

# City of Baltimore

## 911 Calls

Team A

Adil Alkhateeb  
Mohammad Ali  
Asim Ali

Dec 15<sup>th</sup> 2017



# **Agenda**

---

**The City and the Data**

**Data Preparation and Geo-coding**

**Data Analysis**

**Maps Visualization**

**Conclusion**

**Q&A**

# The city of Baltimore

Largest US city in Maryland State

**30<sup>th</sup>** most populous city in the United States

Divided into nine major districts

Crime in Baltimore is above the national average

Baltimore's homicides in **2015** are the highest homicide rate in the city's recorded history

Second-highest for U.S!

**Baltimore's 911 operations are critical!**



# Dataset Overview

**2.8 Million actual Calls**

**Published on daily basis by the City of Baltimore**

**Kaggle's snapshot covering period Jan 2015 to Aug 2017**

	callDateTime	priority	district	description	callNumber	incidentLocation	location
1	2015-07-13 10:47:00	Medium	CD	911/NO VOICE	P151941003	600 E FAYETTE ST	(39.2906737,-76.6071600)
2	2015-07-13 10:42:00	NaN	CD	911/NO VOICE	P151941004	200 E BALTIMORE ST	(,)
3	2015-07-13 10:45:00	Low	CD	PRKG COMPLAINT	P151941005	800 PARK AV	(39.2985163,-76.6184754)

# Data Preparation

---

## Missing Data cleaning

Priorities

Descriptions

Coordinates Geo-coding

# Priorities Missing Data

**Forward Fill and Back Fill ?**

**Use the priorities of same or similar descriptions**

**Multiple priorities for same descriptions!**

**Identify the highest priority count per description**

**Create a *key* 'lookup' table to fill missing priorities**

	Before	After
Priority	Count	Count
Medium	1399431	1403898
Low	636380	637275
High	434022	435239
Non-Emergency	321619	321695
Emergency	1004	1004
Out of Service	803	803
<b>NaN</b>	<b>6655</b>	<b>0</b>

```
In [7]: grouped = data.groupby(['description','priority']).count()['callNumber'].loc[['BURGLARY']]
```

```
In [8]: grouped
```

```
Out[8]:
```

callNumber		
description	priority	
BURGLARY	High	80
	Low	3
	Medium	58416
	Out of Service	15

```
In [9]: key=grouped.loc[grouped.groupby(level=0).idxmax()['callNumber'].values]
```

```
In [10]: key
```

```
Out[10]:
```

callNumber		
description	priority	
BURGLARY	Medium	58416



# Coordinates Geo-coding

## Attempt 1

82,532 missing coordinates

Google Maps geocoding API

*Pygeocoder* Package a python wrapper based JSON

Validate all **2.8M** addresses and coordinates

callDateTime	priority	district	description	callNumber	incidentLocation	location	address	zipCode
2015-07-13 10:47:00	Medium	CD	911/NO VOICE	P151941003	600 E FAYETTE ST	(39.2906737,-76.6071600)	560-618 E Fayette St, Baltimore, MD 21202, USA	21202
2015-07-13 10:42:00	Medium	CD	911/NO VOICE	P151941004	200 E BALTIMORE ST	(39.2898910,-76.6120720)	200 E Baltimore St, Baltimore, MD 21202, USA	21202

Google Limitations:

**2,500** free requests only per day per IP with a limitation on the number of requests per second

\$0.50 USD per 1,000 additional requests up to **100,000** max daily

FAILED

# Coordinates Geo-coding contd.

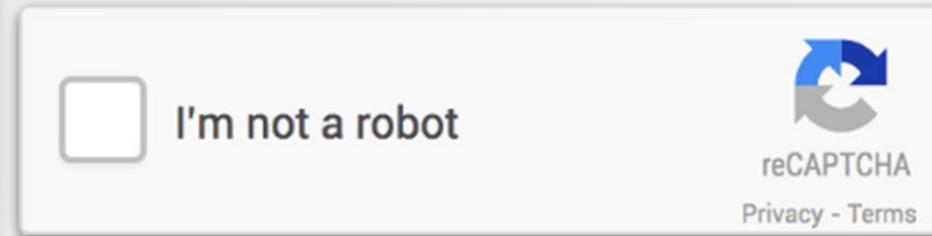
## Attempt 2

Google Maps Places Web Calls

***urllib*** Package used to initiate web requests

Slice and analyze response strings to capture coordinates

```
https://www.google.ca/maps/place/Baltimore,+MD/@39.2846854,-76.6905256,12z/data=
```



```
Out[3]: 'I am a Python robot ... :('
```

FAILED

# Coordinates Geo-coding contd.

## Attempt 3

**Validating all addresses is not possible**

**Workaround to resolve for the missing **82k** coordinates only**

*Back to **Pygeocoder** Package*

**Complex coordinates script developed:**

**Run missing dataset in **82** patches of length **1k****

**Robust algorithm to overcome throughput limitations**

**Disconnection sensors**

**Retrials infused with sleeps**

**Run time: 5-15 mins/patch**

**One week to complete!**

Renew IPs from:  
Home, Office  
Public Networks  
Mobile Data  
Class!

Patch: 10000-10999  
Starting missing: 1000  
Ending missing: 0  
Resolved coordinates: 1000  
Success rate: % 92.10

**SUCCESS**  
**74k**  
**Resolved!**

# Data Analysis

---

## High-level Analysis

Trend Analysis and Correlation

## Analysis Framework

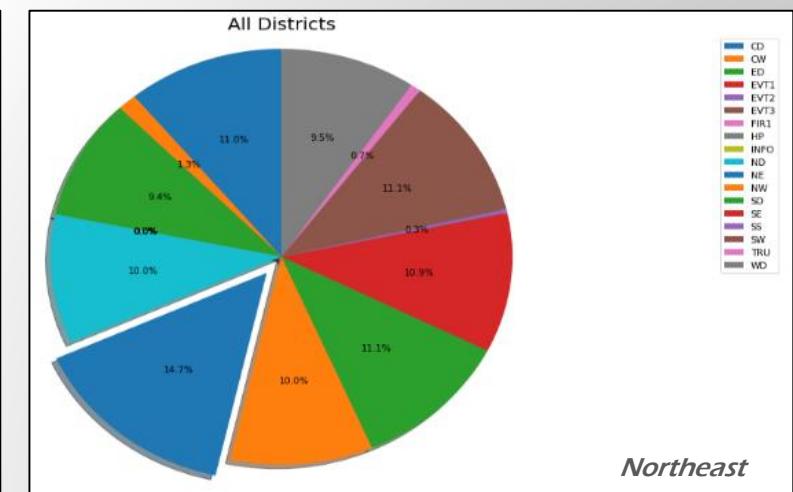
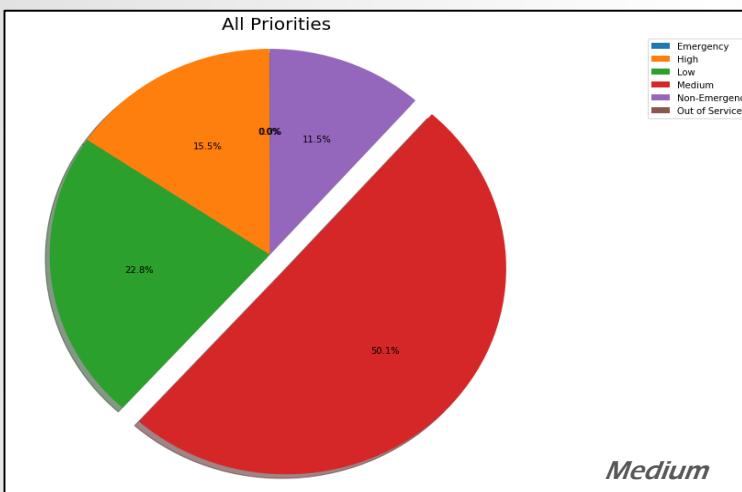
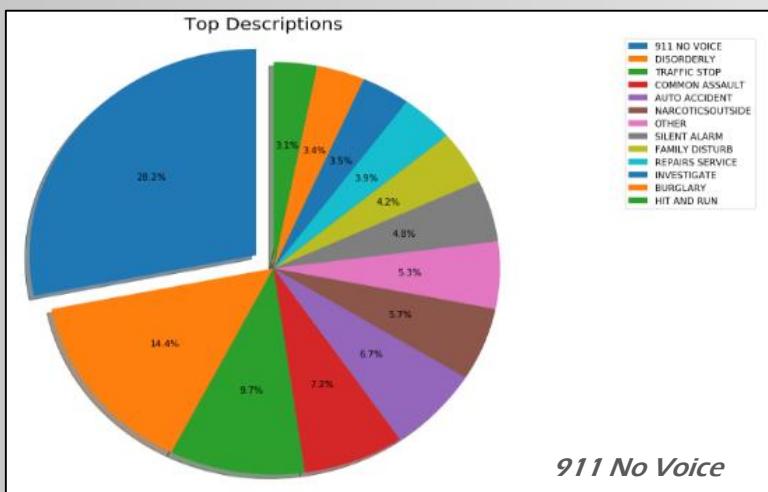
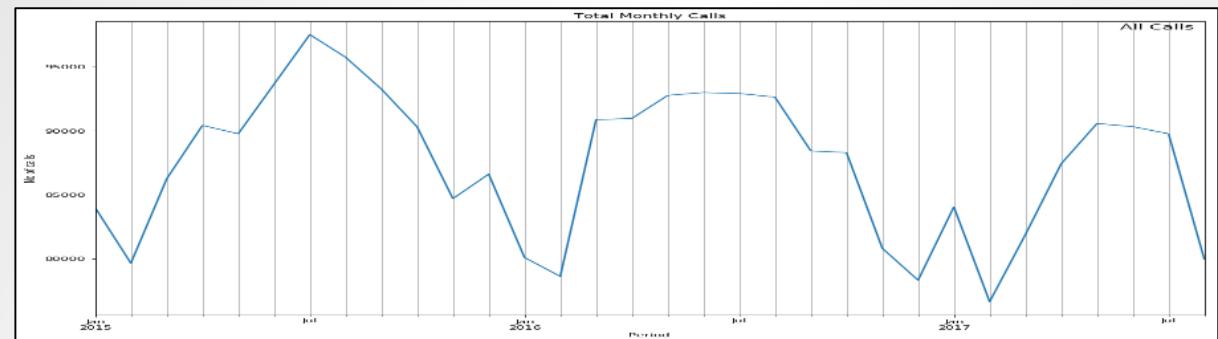
**Story 1:** Assaults Analysis

**Story 2:** Narcotics Analysis

# High Level Analysis

callDateTime	priority	district	description	callNumber	incidentLocation	location	lat	lon
2015-07-13 10:47:00	Medium	CD	911 NO VOICE	P151941003	600 E FAYETTE ST	(39.2906737,-76.6071600)	39.290674	-76.607160
2015-07-13 10:42:00	Medium	CD	911 NO VOICE	P151941004	200 E BALTIMORE ST	(39.2898910,-76.6120720)	39.289891	-76.612072
2015-07-13 10:45:00	Low	CD	PRKG COMPLAINT	P151941005	800 PARK AV	(39.2985163,-76.6184754)	39.298516	-76.618475

**2,799,860 records**  
**87,496 calls per month**  
**2,881 calls per day**  
**peaking on Fridays**  
**Slowing down on Sundays**



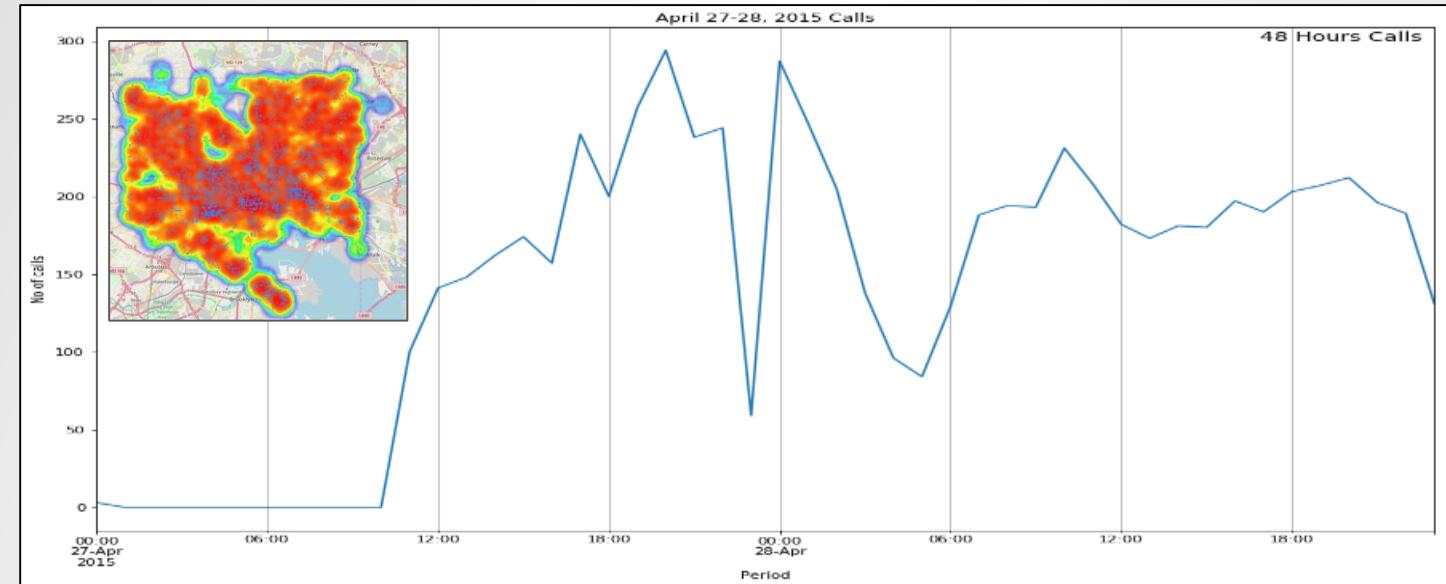
# High Level Analysis – Peak Overview

Peak month July 2015 at a record high of **97,511** calls  
Peak day April 28<sup>th</sup> 2015 at **4,441** calls

Baltimore was on violent riots  
after the Apr 27th funeral of  
Freddie Gray



A young Black American man arrested by the Baltimore Police and died from his injuries allegedly caused by the arresting officers



CNN U.S. » Crime + Justice | Energy + Environment | Extreme Weather | Space + Science Live TV • U.S. Edition + 🔍

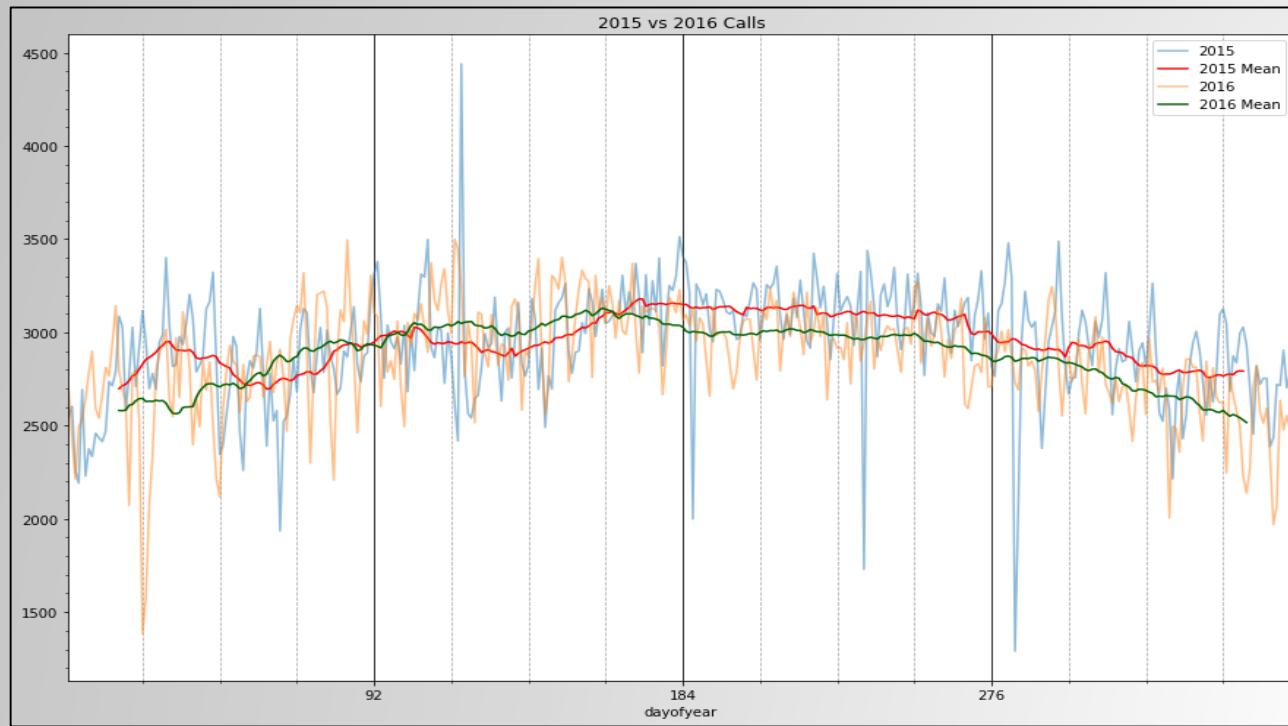
Baltimore riots: Looting, fires engulf city after Freddie Gray's funeral



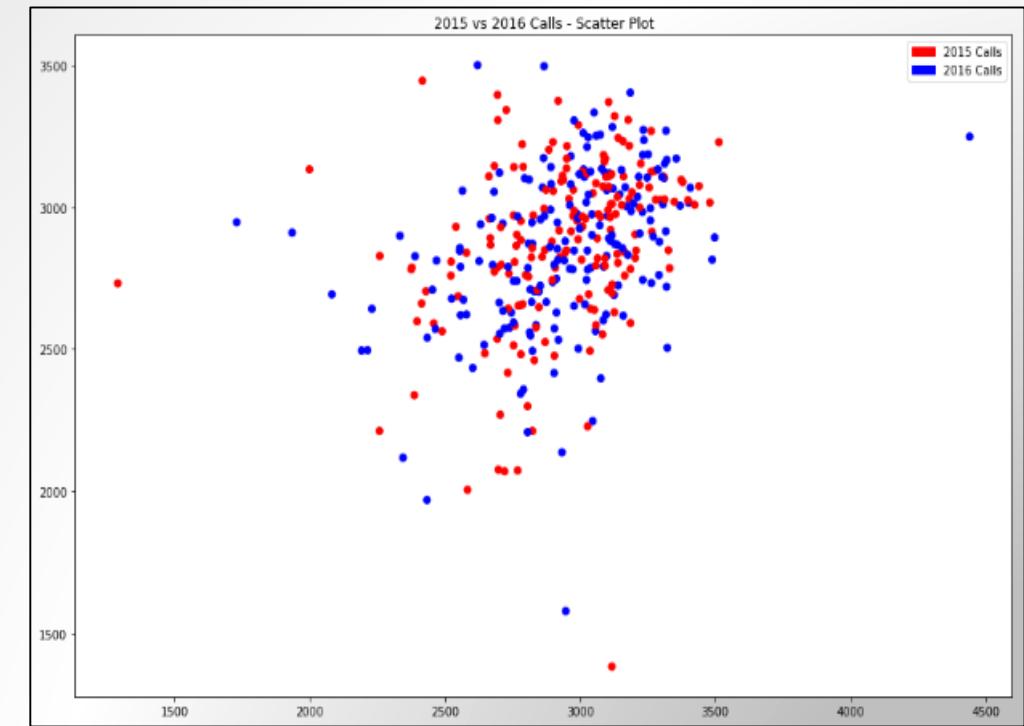
By Holly Yan and Dana Ford, CNN

© Updated 10:30 AM ET, Tue April 28, 2015

# Trend Analysis and Correlation



**Seasonal Trend:** normalized to smooth out high calls oscillation by taking the rolling mean of calls count per day at a 30 days window



No clearly noticeable correlation found

# Analysis Framework

**Accepts a text as search keyword criteria**

**Equipped with scripts to load, search and produce analysis results**

A search with 'assault' keyword produces full analysis on all assault records such as 'Common Assault', 'Other Assault', 'Any Assaults' etc

**Results grouped and analyzed by priorities and districts**

**Resampled:**

on daily basis to analyze weekdays trends

on monthly basis to produce visual analysis charts

**Pre-set charts configuration functions**

`DrawPieChart()` - Helps in drawing pie charts and calculating maximum percentage to 'explode' the highest value on the pie chart

`SetChartProperties()` - Pre-set chart properties for smoother and neater output

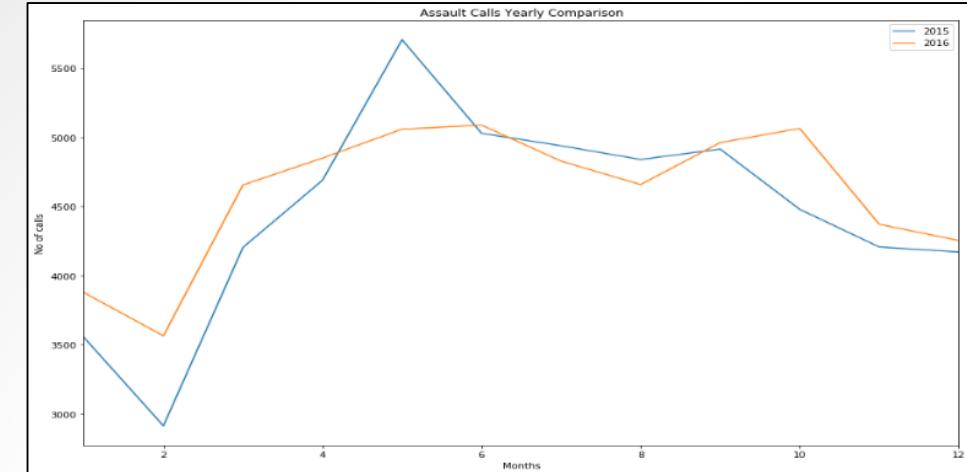
`DrawAutoCorrelation()` - Draws line charts along with autocorrelation analysis

# Story 1: Assaults Analysis

**146,280 assault calls, 5.22% of total calls**

**High frequency assaults:**

**COMMON ASSAULT  
AGGRAV ASSAULT  
DOMESTIC ASSAULT  
SEXUAL ASSAULT**



**Highest Assaults reported from:**

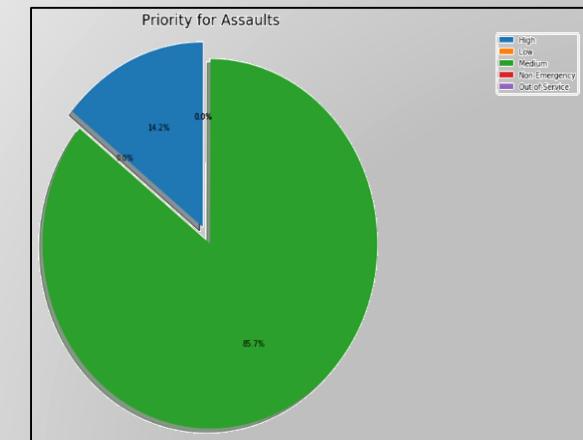
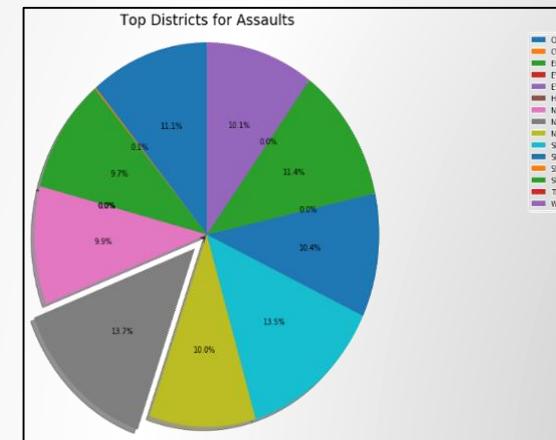
**'North East' district (14%)  
'South District' (13%)  
'South West' (11%)  
'Central District' (11%)**

THE BALTIMORE SUN



May 2015: Baltimore's deadliest month since 1970s

JUNE 11, 2015, 1:20 PM



## Story 2: Narcotics Analysis

**128,332 narcotics calls, 4.58% of total calls**

Analyzed calls includes Narcotics observed indoors and outdoors

# Highest Assaults reported from:

# 'West District' (19%)

## 'North West' (14%)

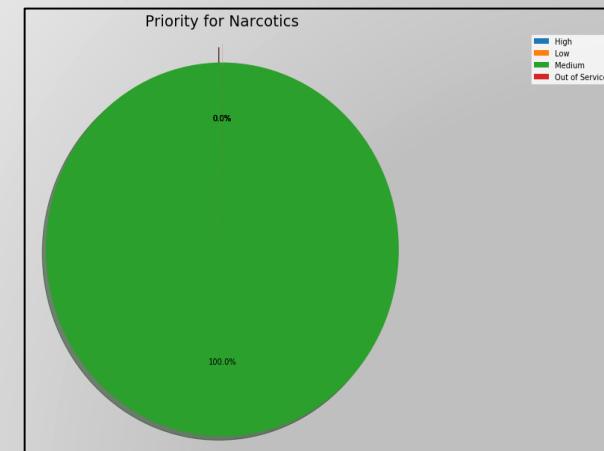
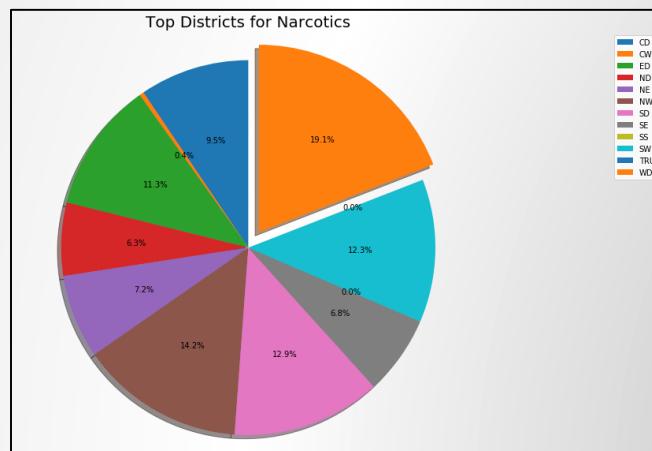
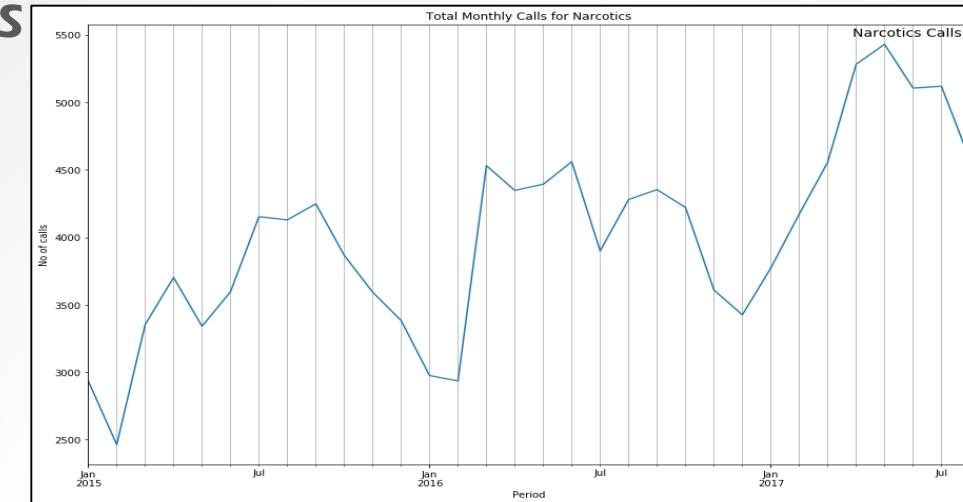
## 'South District' (13%)

# 'South West' (12%)

# Increasing trend Jan'17 Peaking on May'17



## 3 Men Face Charges After \$9 Million Worth Of Heroin Seized



# Maps Visualization

---

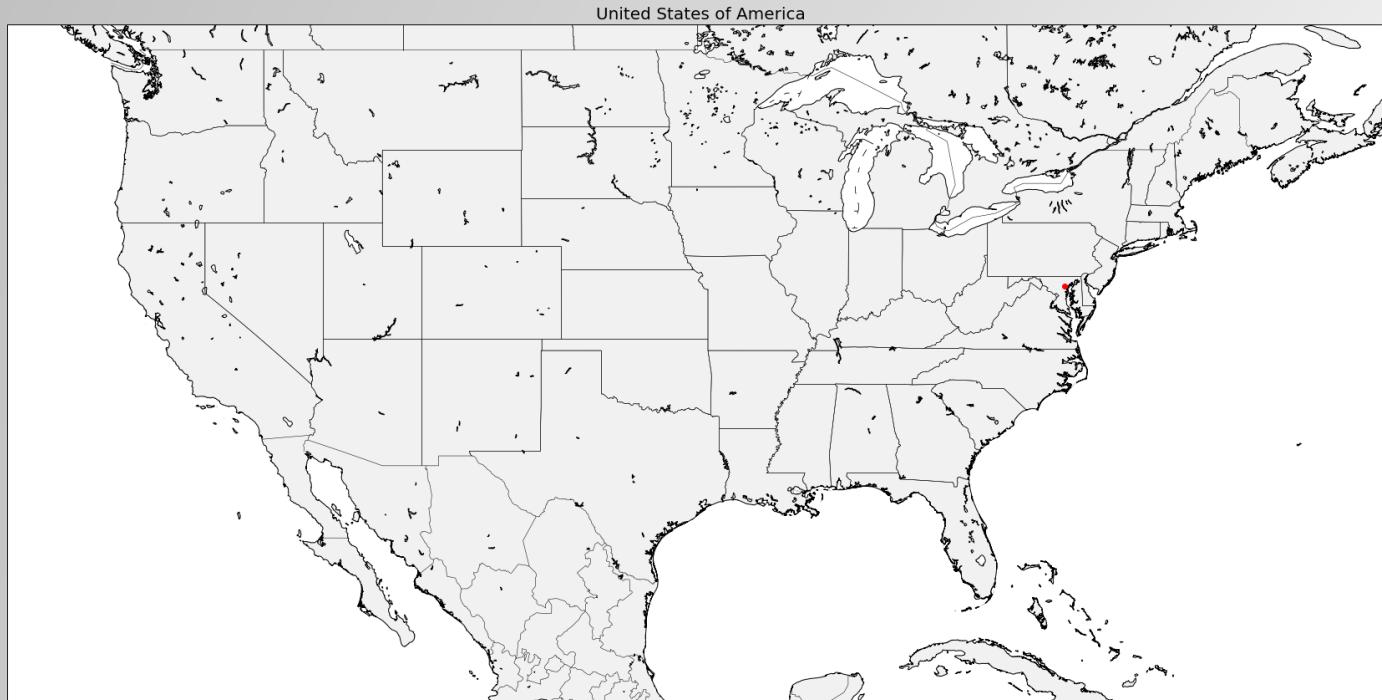
## Visualization Packages

**Basemap**

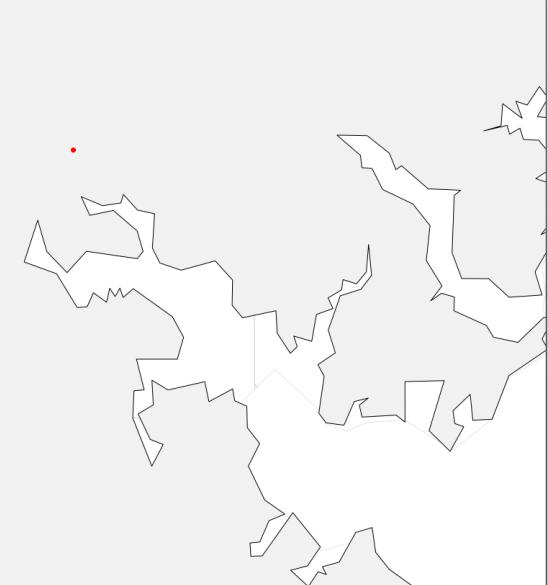
**Leaflet.js**

**Folium**

# Basemap Package



The City of Baltimore



## Unofficial Windows Binaries for Python Extension Packages

by Christoph Gohlke, Laboratory for Fluorescence Dynamics, University of California, Irvine.

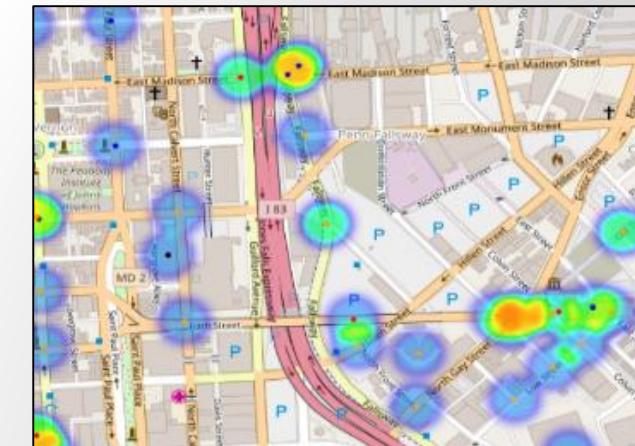
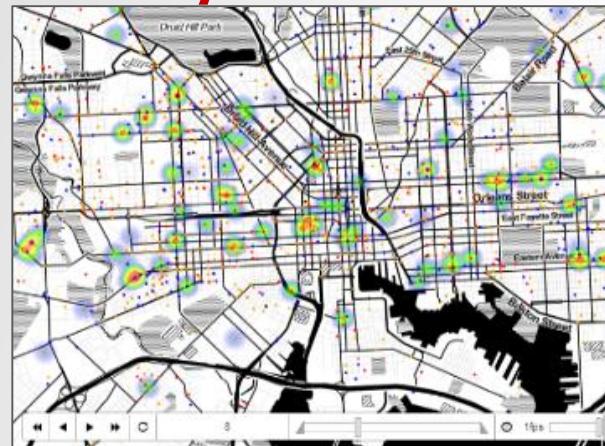
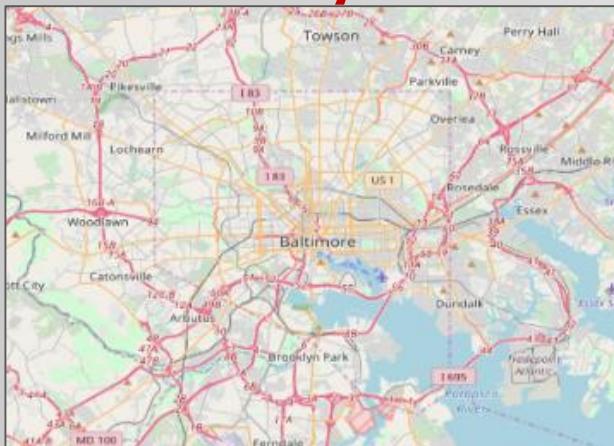
<https://www.lfd.uci.edu/~gohlke/pythonlibs/#basemap>  
basemap-1.1.0-cp36-cp36m-win\_amd64.whl

# Folium and Leaflet.js

***folium*** package brings python strength and ***Leaflet.js*** mapping library together

Allows to visualize data over interactive Leaflet maps

***OpenStreetMap*** and ***StamenToner*** tiles used along with ***HeatMaps*** and ***TimeHeatMaps*** to visualize calls



Jupyter was unable to render maps with large data points due to low iopub\_data\_rate\_limit value in Jupyter config file. It was necessary to set that parameter to a higher limit by starting Jupyter using the following: `jupyter notebook --NotebookApp.iopub_data_rate_limit=10000000`

# Conclusion

**911 operations at the City of Baltimore's is clearly very critical given the City's Crime rates and the supporting analysis results presented**

**Data capturing efficiency can be enhanced:**

Preset descriptions

Automated address capturing

**The large volume of calls can be used to predict a forecast of 2018 calls on various levels to support:**

Resources planning at the 911 operations center

Field resources planning utilizing the geo-tagging functionalities implemented.

# **Q&A**