Safer Together: Washington DC

July 3, 2018

Table of Contents

1. Introduction
2. Research and Data Sources
3. Models and Key Findings
4. Recommendations
5. Next Steps

The Team

Brian Collins

- Data Scientist
- The Code

Harsha Goonewardana

- Data Scientist
- The Vision

Priya Kakkar

- Data Scientist
- The Facts

Kihoon Sohn

- Data Scientist
- The Executioner

Project objectives:

1. Inference of gunshots predictors

2. Prediction of gunshots frequency in the next 12 months

Research & Data Sources

Impacts of gun violence on the community

Across five cities, gun violence surges slowed neighborhood home value appreciation by ~4 percent. Also decreasing Average credit score and homeownership rates.

Learn more: "Gun Violence Affects the Economic Health of Communities"

In Washington, DC, every 10 fewer incidents of gunfire in a census tract are significantly related to

- one new business opening
- creation of 20 more jobs in new businesses
- $1.3 million more in sales at new businesses
- one less business closure

Learn more: "Is Gun Violence Stunting Business Growth?"

Source - https://repository.upenn.edu/cgi/viewcontent.cgi?article=1352&context=fnce_papers

Gun crimes and housing prices

Research suggests that violent crime, for which gunshots can be a proxy, is selectively underreported and ShotSpotter data provides an important alternative outcome measure.

There is a causal link between amount of violent crime and property values. Across United States of America, reduction in gun crimes have shown to increase property values by close to two thousand dollars.

Source - http://journals.sagepub.com/doi/abs/10.1177/0011128717714974

Importance of 311 service request resolution

- Increase in “Hot spots” or “micro places” translates into increased crime.
- Examples: bus stops, a busy street, or the lack of street lighting etc.
- A correlation between calls for service and crime have been shown by the literature. Researchers have suggested that 311 calls for service are a leading indicator of municipal neglect.
- A delay in providing services such as graffiti removal, street light repair etc. gives gang members a sense of lack of police presence encouraging their motives.

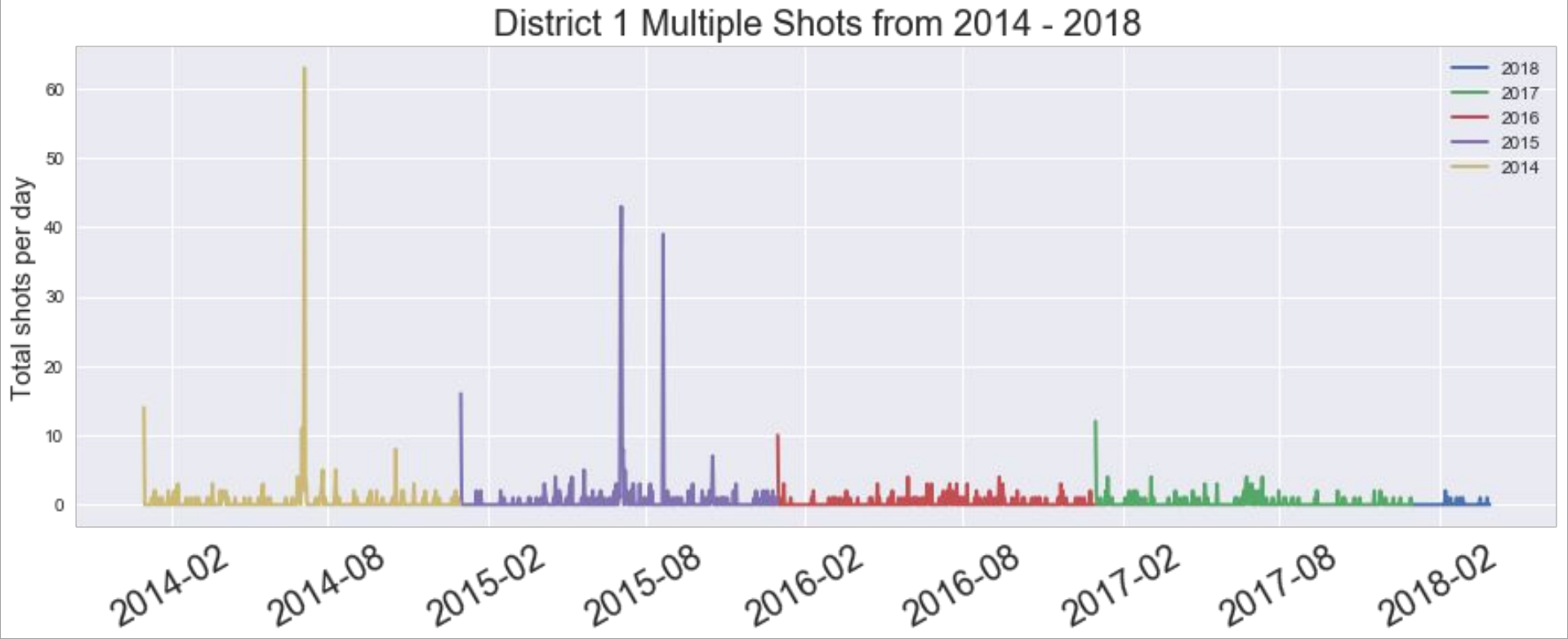
Source - https://www.huduser.gov/portal/periodicals/em/summer16/highlight2.html

Data Sources

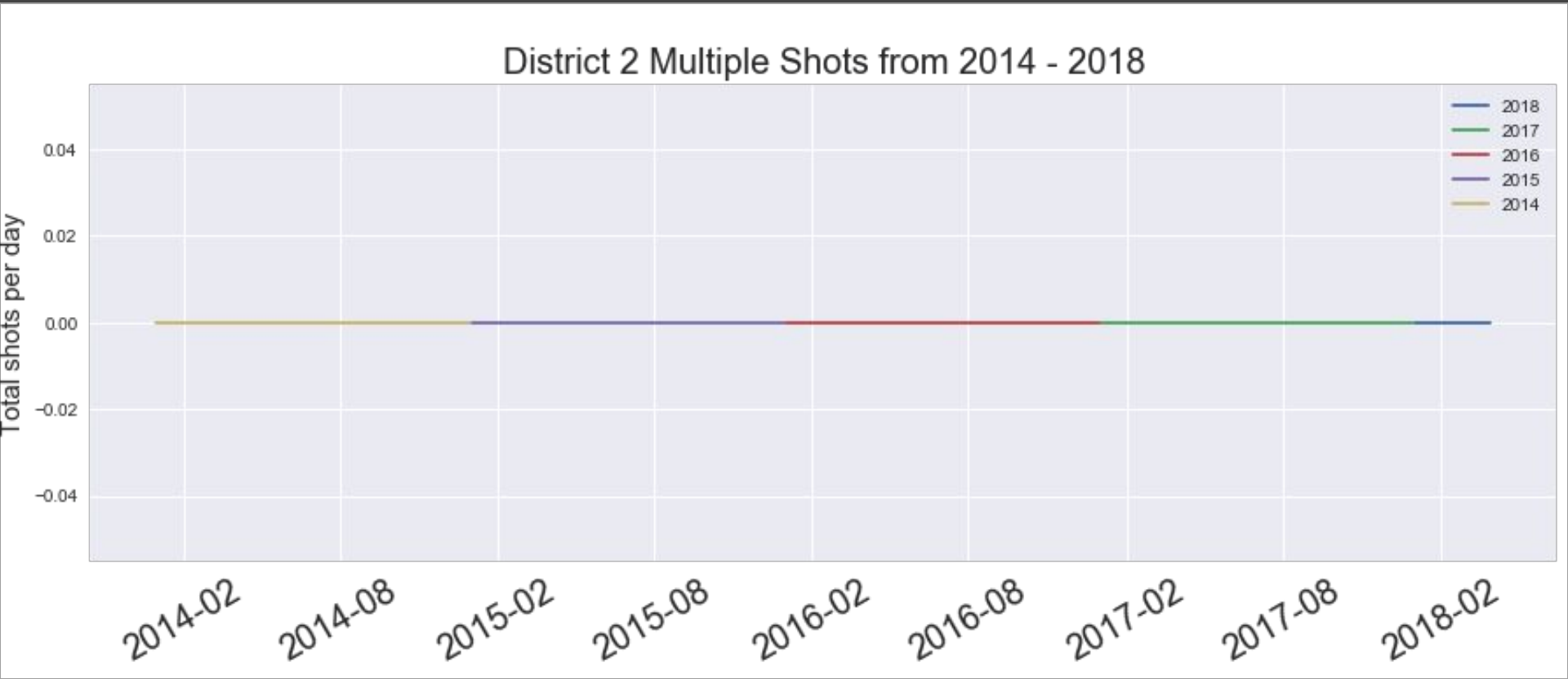
Target periods: 2014 ~ 2018 (retrieved June 18~20, 2018)

1. **ShotSpotter Data**, Metropolitan Police Department [1]
2. **City Service Requests Data**, Open Data DC [2]
3. **Housing Price Data**, Redfin [3]
    1. Source: https://mpdc.dc.gov/publication/shotspotter-data-disclaimer-and-dictionary
    2. Source: https://opendata.dc.gov/ (2014 / 2015 / 2016 / 2017)
    3. Source: https://www.redfin.com/blog/data-center

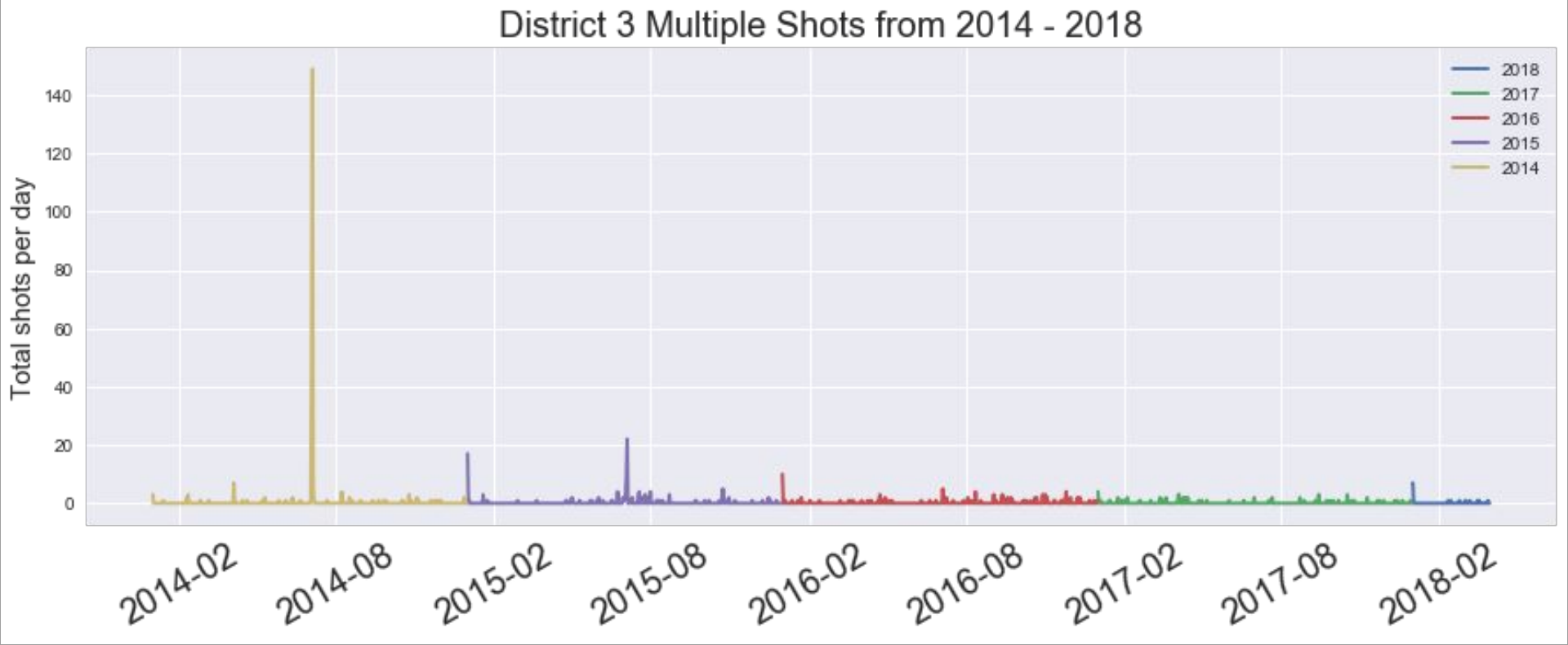
Basic Breakdown of Data - District 1



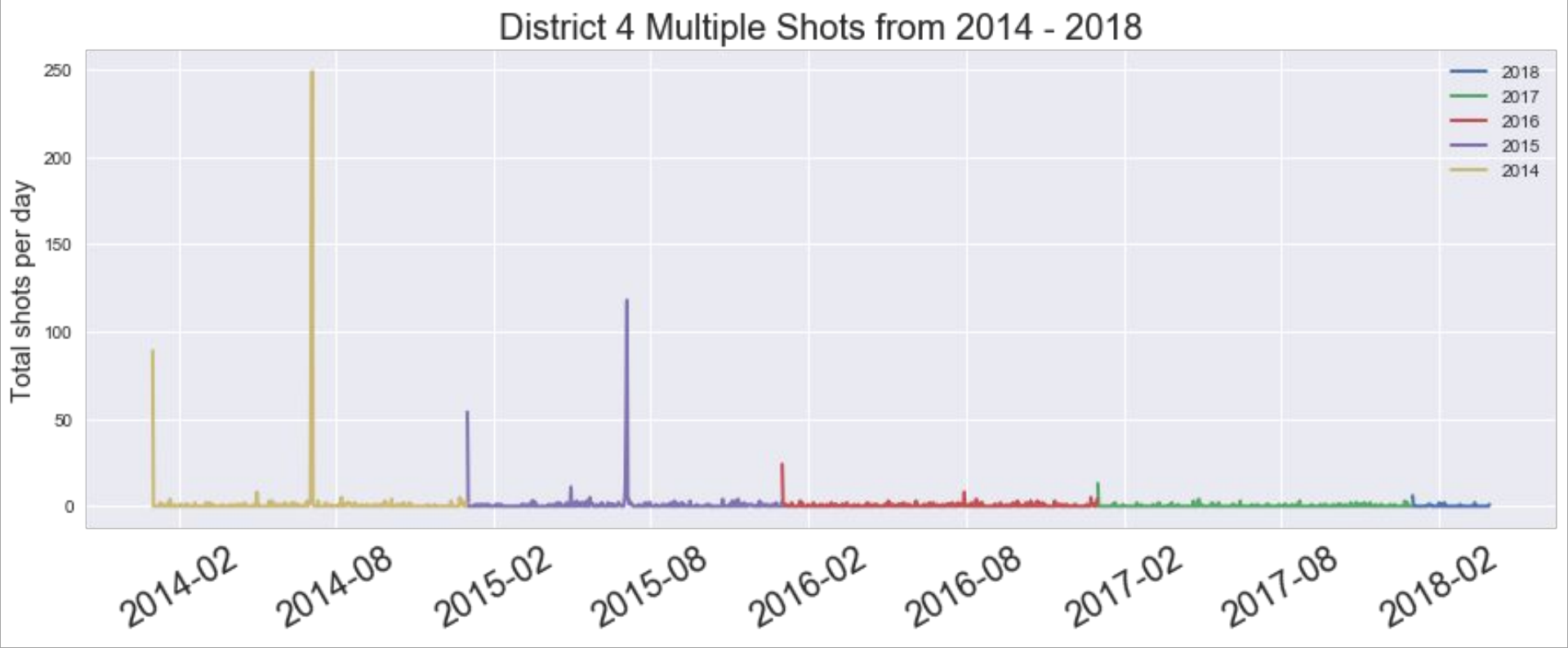
Basic Breakdown of Data - District 2



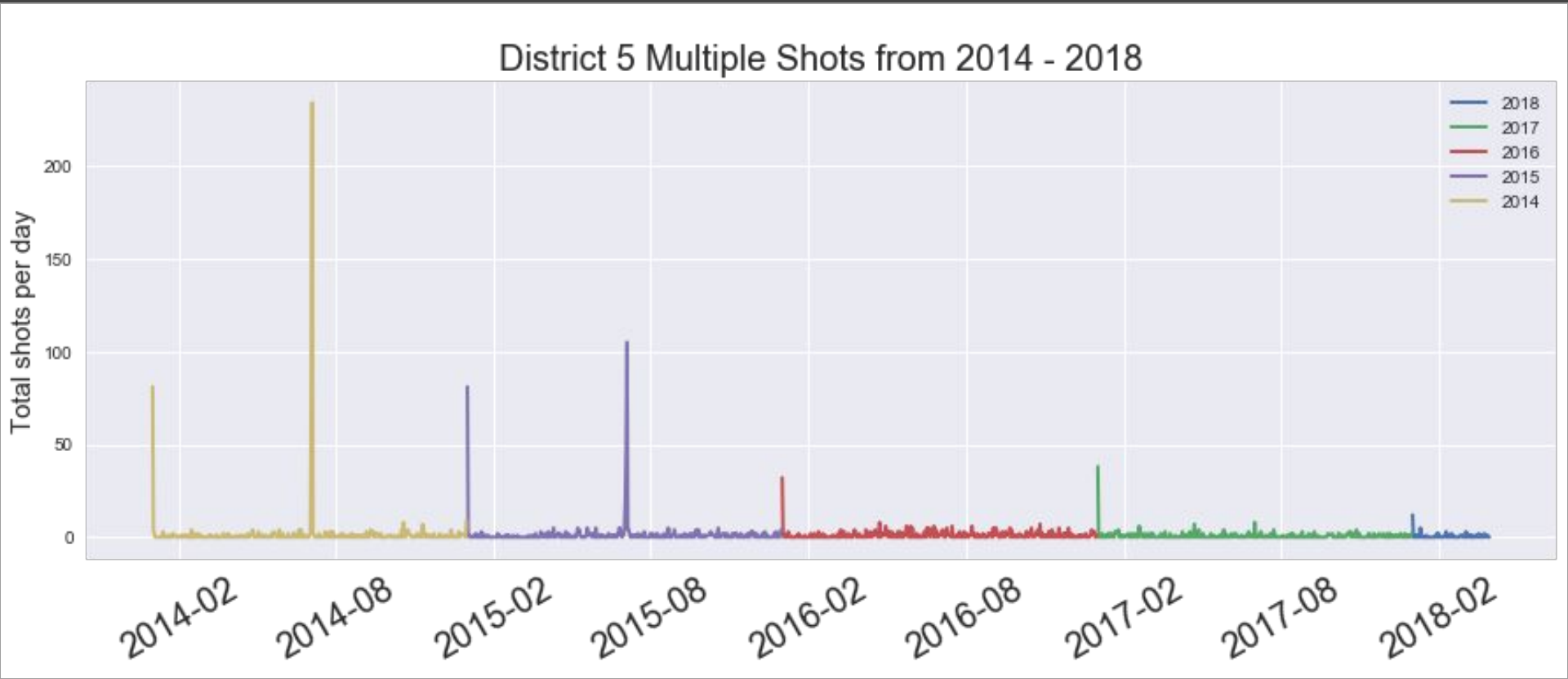
Basic Breakdown of Data - District 3



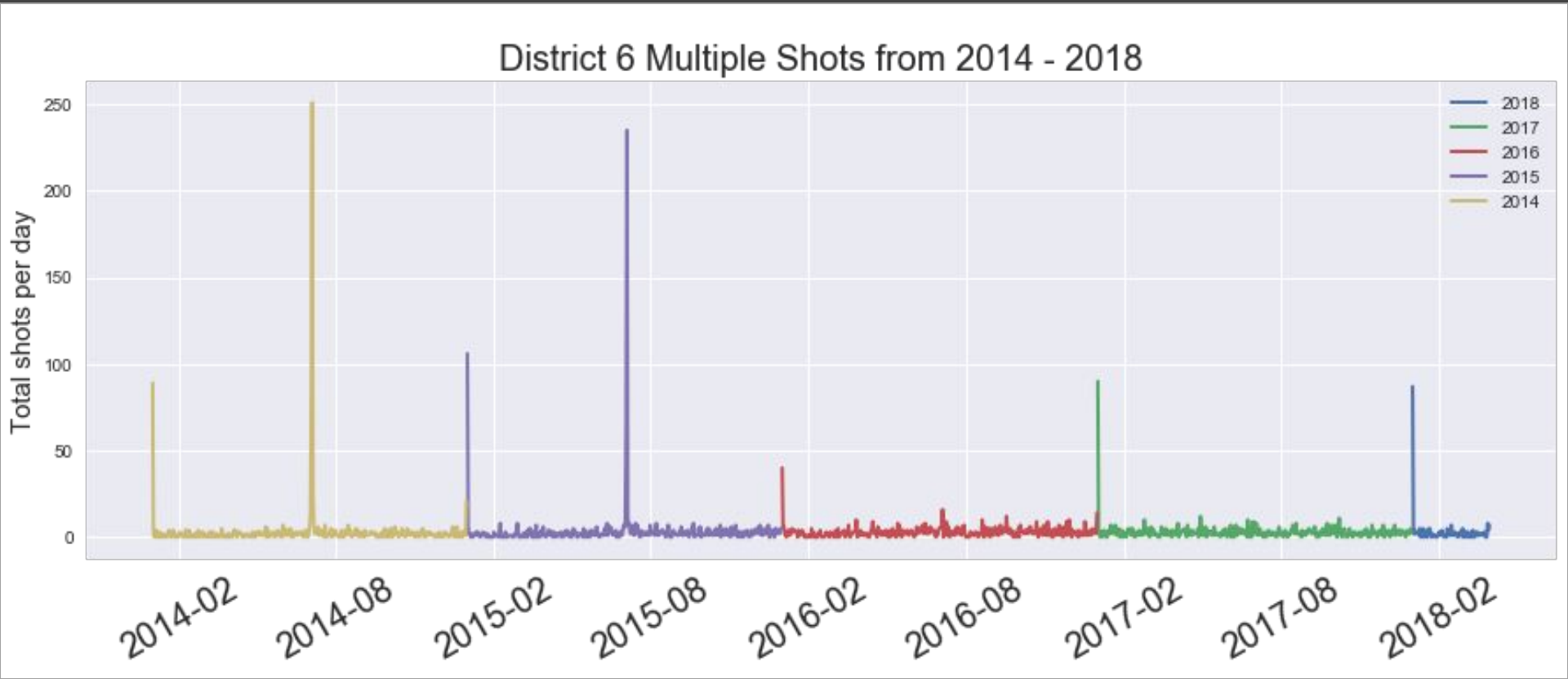
Basic Breakdown of Data - District 4



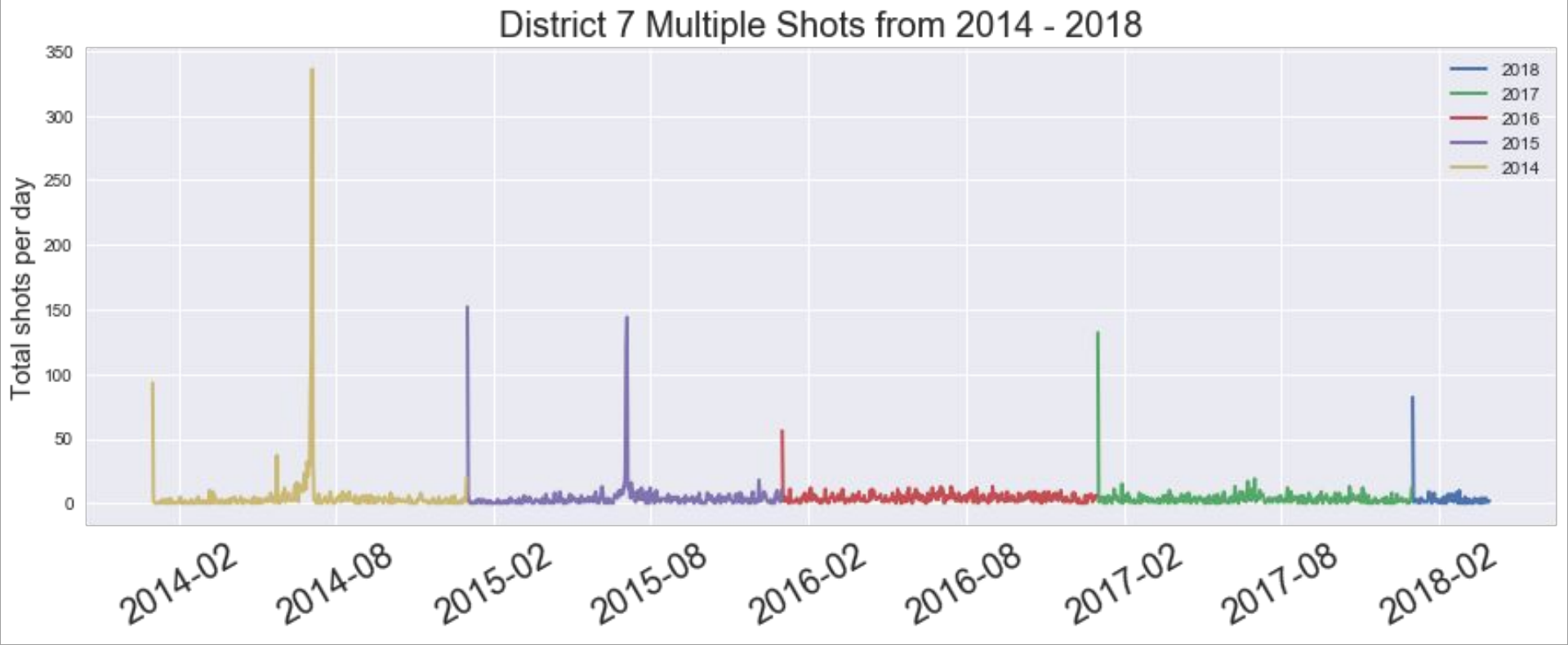
Basic Breakdown of Data - District 5



Basic Breakdown of Data - District 6



Basic Breakdown of Data - District 7

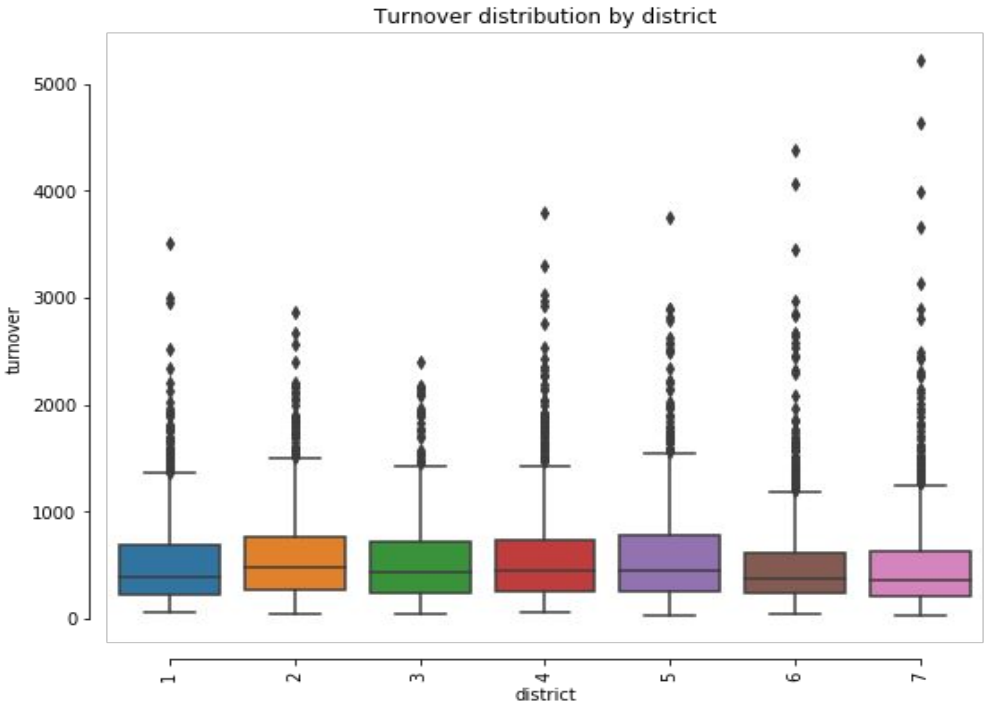


Summary of data

- No gunshot data from District 2 [1]
- Trend each year towards less shots as a result of better calibration in sensors
- The ShotSpotter data showed high false positive rates in 2014-2015 with fireworks often being mistaken for gunshots around the holiday periods
- Districts 1, 3, and 4 consistently show low levels of shots in 2018

1. District 2 data is not included as there was no collection of data

City Service Request (311) data



Turnover(hours) = {Resolution Date} - {Add Date}

Consolidated 311 service

collection	301067
parking meter repair	223577
parking enforcement	206177
maintenance	127182
tree related	54692
street cleaning	51731
light repair	47930
signs	33080
sanitation	29834
dmv related	23625
snow related	21979
dumping	20510
graffiti removal	15873
pesticide	12537
report	11253
safety	8644
transportation	5621
bicycle related	5052
dc gov	3402

Selected 6 categories for the model

311 Call Reviews - frequency of 6 categories by district

district	Collection	Graffiti Removal	Light Repair	Maintenance	Parking Meter Repair	Street Cleaning
1	40887.0	1267.0	7277.0	17453.0	81637.0	7939.0
2	49237.0	3306.0	12720.0	29893.0	110277.0	2863.0
3	28430.0	6120.0	4718.0	11092.0	23315.0	7719.0
4	72719.0	3337.0	12172.0	27120.0	5254.0	10601.0
5	43806.0	1493.0	6457.0	17392.0	2349.0	9475.0
6	47797.0	242.0	5824.0	17057.0	655.0	8723.0
7	26030.0	159.0	2879.0	10810.0	87.0	4411.0

- Largest observation for each categories
  - Collection(D4), Graffiti Removal(D3), Light Repair(D2)
  - Maintenance(D2), Parking Meter Repair(D2), Street Cleaning(D4)

Models and Key Findings

Time Series Regression

Two parallel models deployed at tandem to cross validate the predictions

ARIMA (Auto Regression Integrated Moving Average)

- Dependency between the variable in different time periods
- Difference between current values and the previous values of each observation
- Relationship between the observations and the error term

FB Prophet*

- Piecewise linear or logistic growth curve trend.
- Yearly seasonal component modeled using Fourier series
- Weekly seasonal component using dummy variables.

* Source - https://research.fb.com/prophet-forecasting-at-scale/

Classification

Model: Random Forest Classifier

Characteristics of this model: Random Forest model measures the relationship between the categorical dependent variable and one or more independent variables by using a greedy algorithm to calculate a gini coefficient to pick the best route.

Why choose it: Random Forest is a strong classifying model used to find categorical values that quickly reach the optimal value. It is also very easy to interpret.

Selected Features

Time series data*

ShotSpotter data

311 Service calls

Tested for stationarity

10,220 daily data points

(4 years * 365 days * 7 districts)

Geo-Spatial

PSA location

Districts

Longitude/Latitude

Other

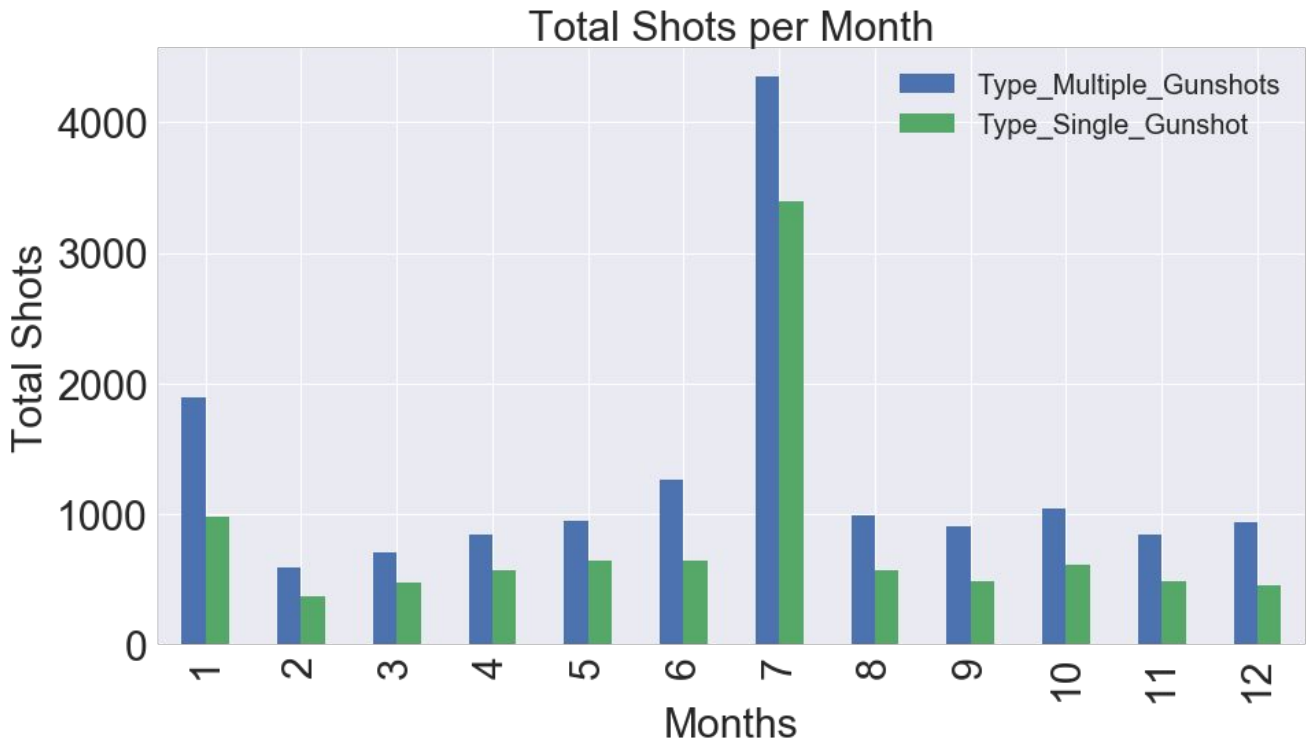
Turnover Time

Specific Type of Requests

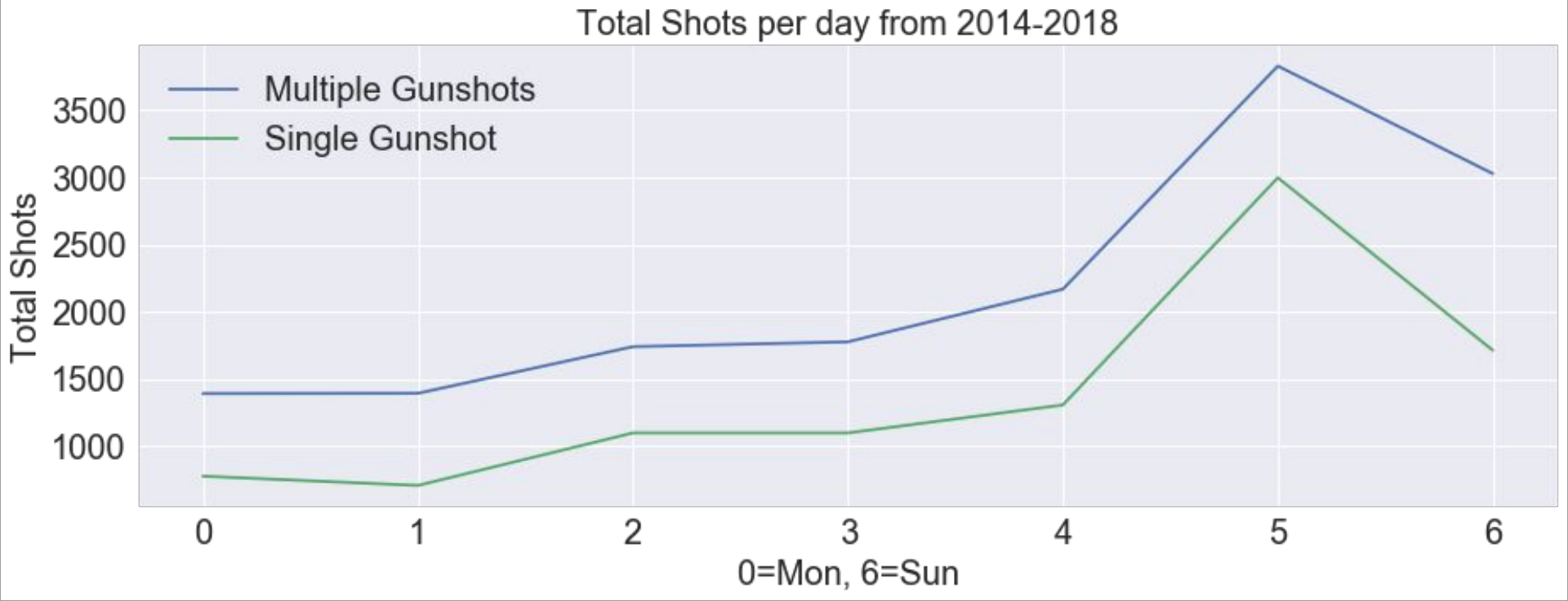
Key findings

- Presence of false positive clustered around holiday periods.
- Likelihood of shots substantially higher on weekends
- Correlation between gunshot over 311 cases/turnover time
- Strong evidence of autocorrelation on gunshot data

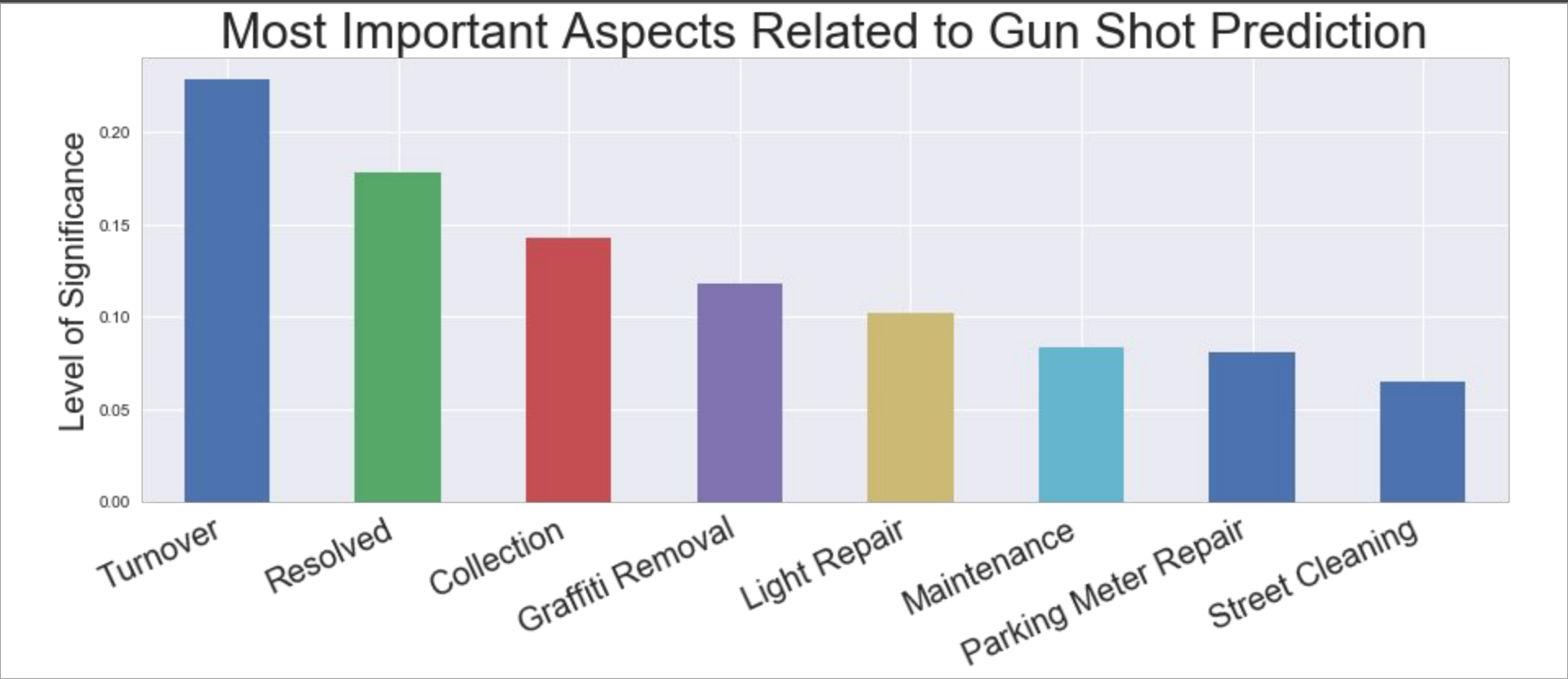
Presence of false positive clustered around the holiday periods.



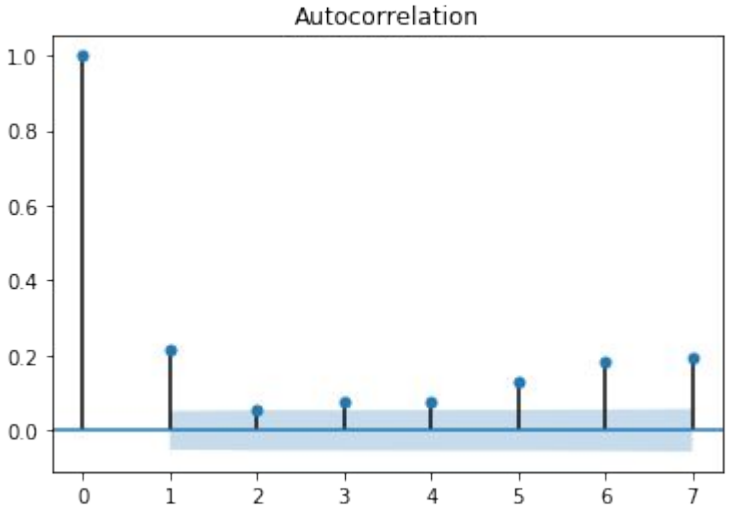
Likelihood of shots substantially higher on weekends



Correlation between gunshots over 311 cases/turnover time



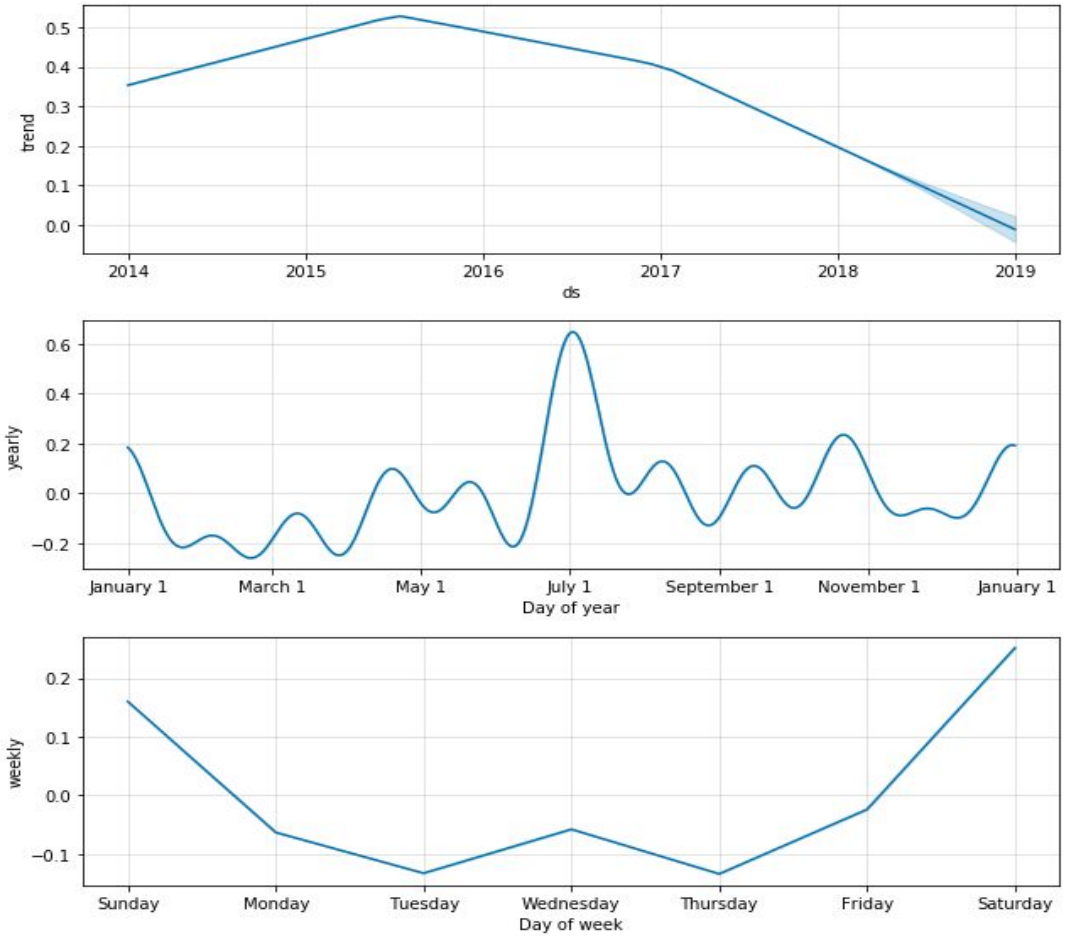
Strong evidence of autocorrelation on gunshot data



- The graph represents a first order autocorrelation.
- This could translate into gunshots leading to other firearm
- incidents within 1 to 2 days of the original incident.

Seasonal and trend analysis: DC-wide

- Gunshot incidence are declining from an apogee in 2015 and expected to fall.
- Gunshots are most likely in July, lowest in February.
- Weekends are also most likely time of the week to hear gunshots



12 months Time series forecast with prediction in DC

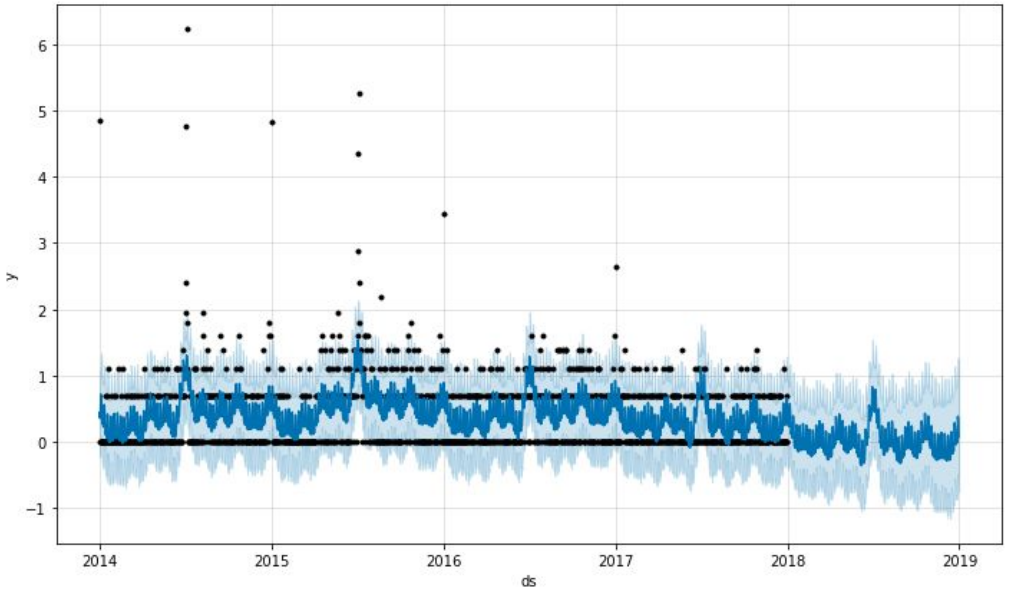


Figure 1) FB Prophet

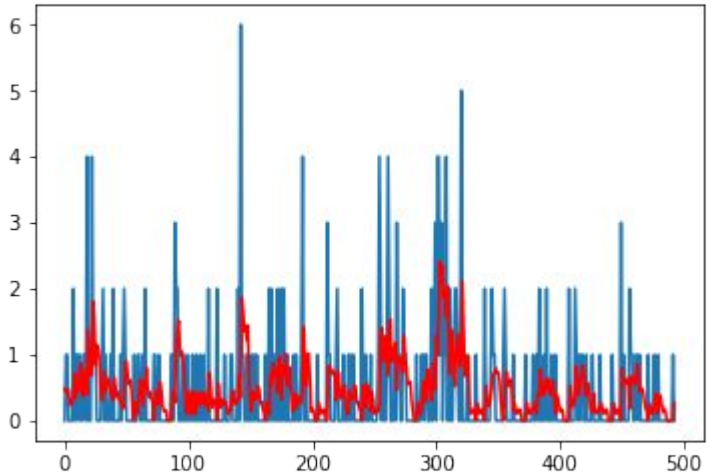
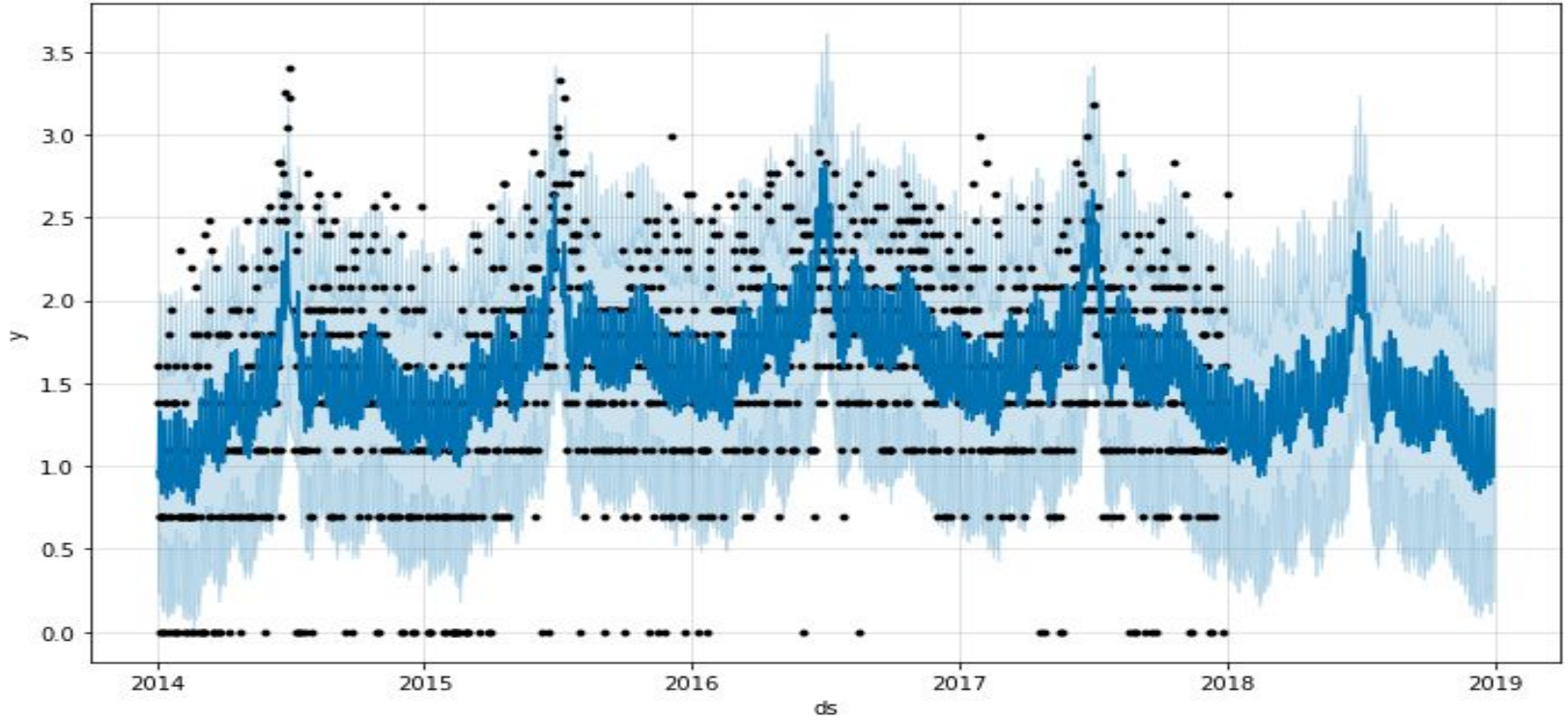
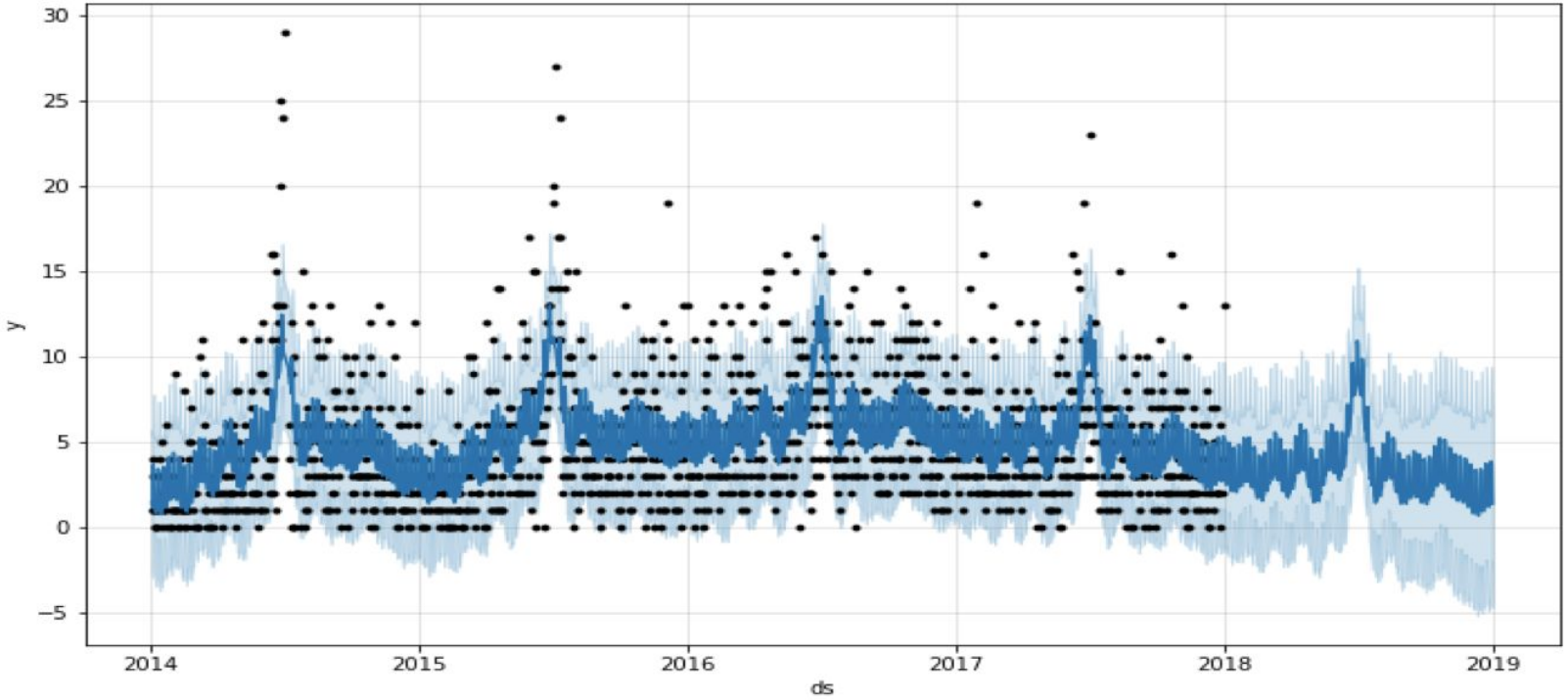


Figure 2) ARIMA (units, days)

District 6: no decrease in gunshots in the next 12 months



District 7: no decrease in gunshots in the next 12 months



Recommendations

Recommendations

Police Deployment	• Optimize schedules to increase police presence during weekends
Speedy Results	• Resolve 311 calls quickly in all areas
Target District Gunshot	• Focus more attention on District 6 and 7
High Gunshot Auto correlation	• Increase police presence in the days after the firearm incident in that location

Challenges

Housing Data & PSA

- Neighborhoods are clustered unevenly across all PSA
- It is captured in monthly interval as the other datasets in daily interval

311 Calls Data

- Data is mislabeled
- Description and units are inconsistent across the time period
- Data labels changed in 2015

Next steps

- Adjust ShotSpotter calibration ShotSpotter to further reduce False Positives
- Obtain housing data with more granular temporospatial variables
- Use housing data to better calculate economic loss from shots
- Investigate correlations of traffic tickets associated with crime

Brian Collins

Harsha Goonewardana

Priya Kakkar

Kihoon Sohn

Any questions?

Thank you!!

Project GitHub link
http://bit.ly/TheLabDC_GA