

Predicting London Crime

Laila Sprejer, Oct 31th, 2018

“Police forces in the UK have access to a vast amount of digital data but lack the technological capabilities to make effective use of it.”

Big data and policing. Sept. 2017.



RUSI
www.rusi.org

Royal United Services Institute for Defense and Security Studies

The data

kaggle™

2008 - 2016

33 boroughs
4835 LSOAs

9 major categories
32 minor categories

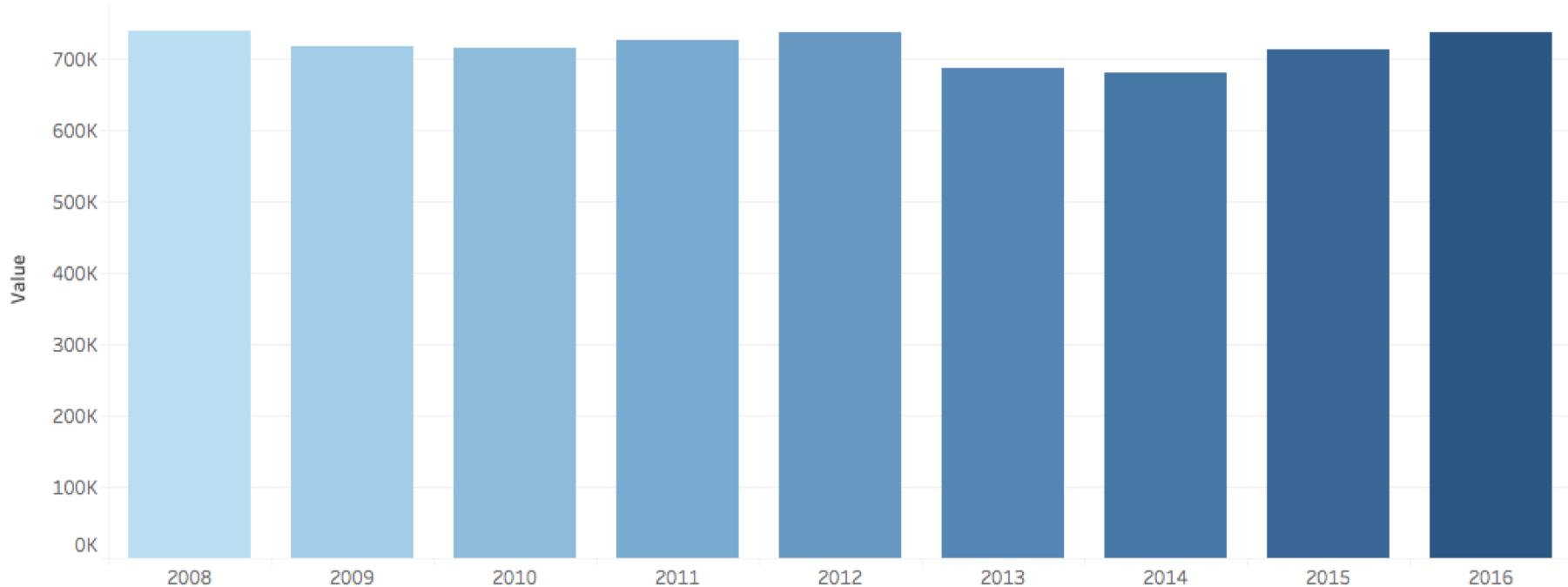
Number of committed crimes per month

13M observations

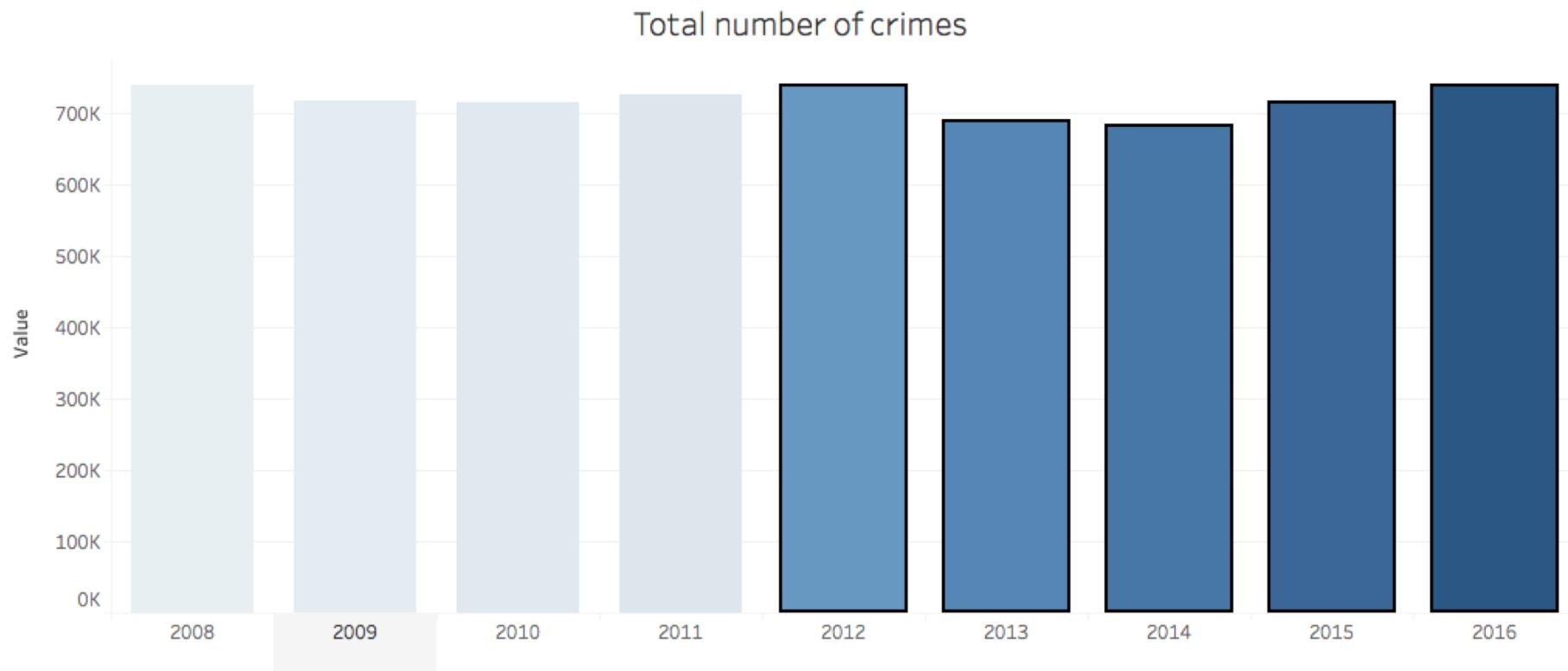
LSOA = Lower Level Super Output Areas

Taking a look at the data

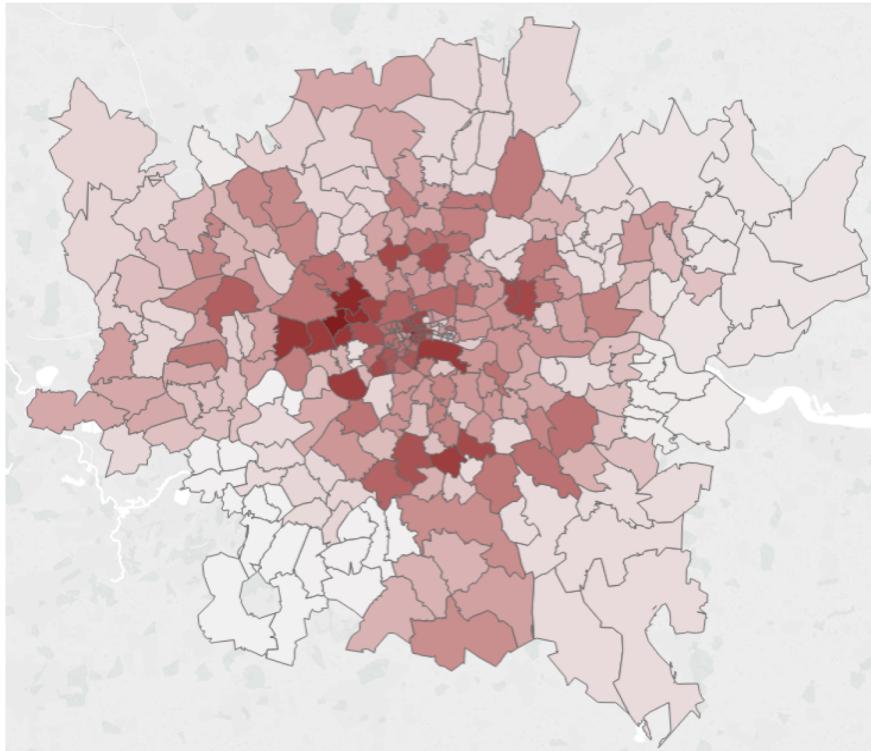
Total number of crimes



Focused on 2012 - 2016 data



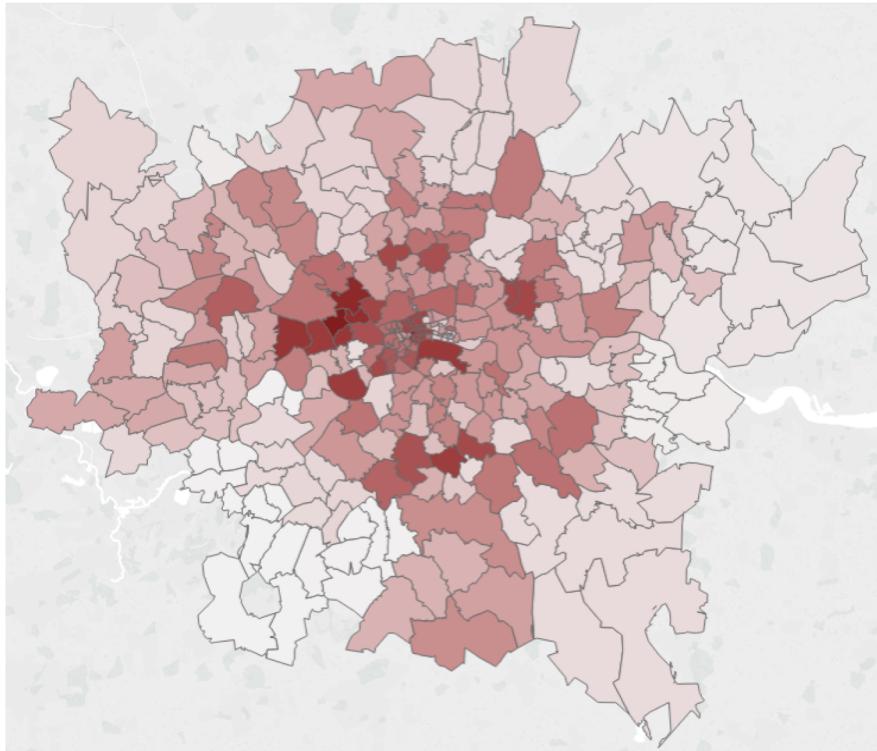
Taking a look at the borough crime distribution



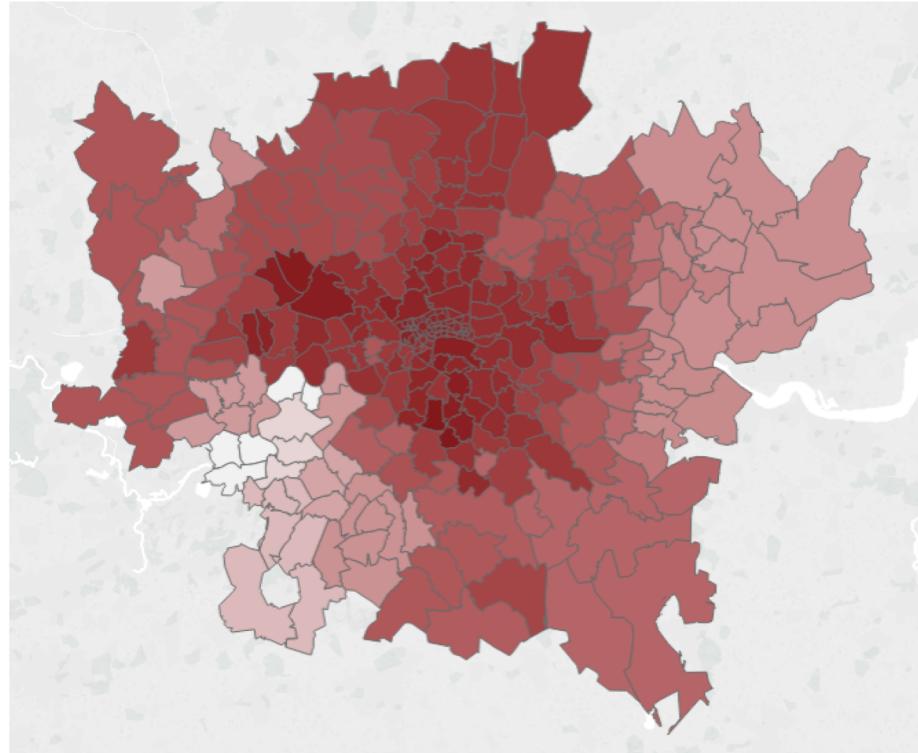
One or more crime committed for a category, month, Isoa

No crimes committed for a category, month, Isoa

Taking a look at the borough crime distribution



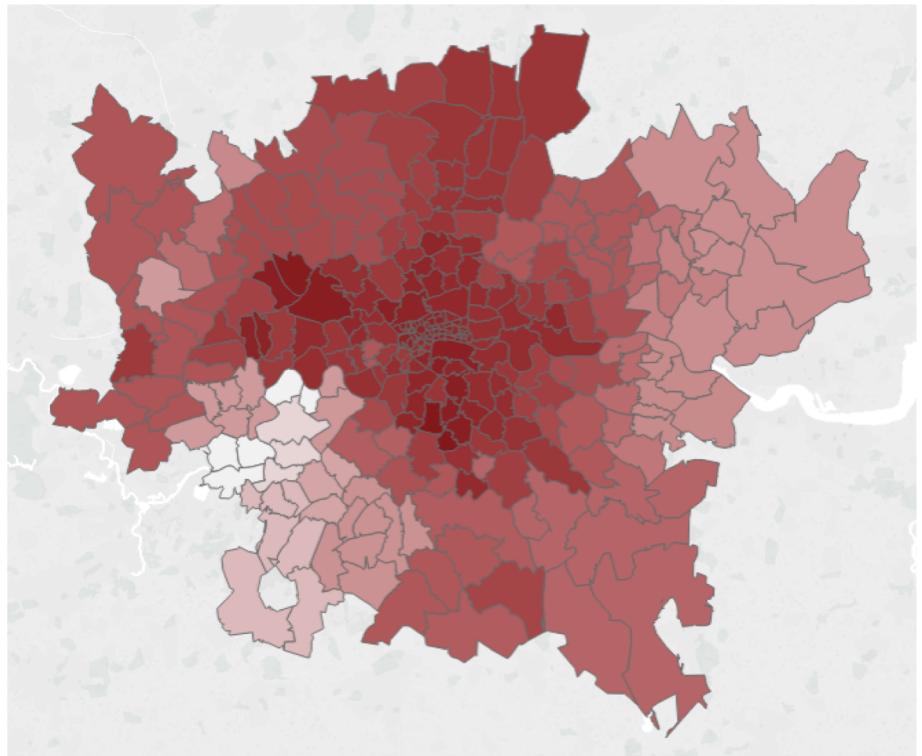
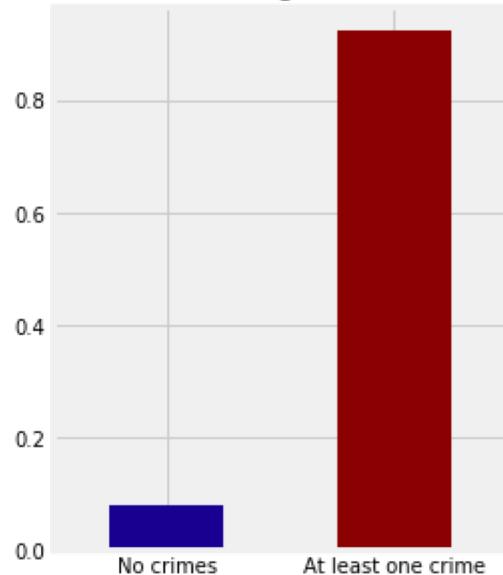
Number of crimes



% of months with at least one crime

Taking a look at the borough crime distribution

Borough distribution of target variable for 2012-2016



% of months with at least one crime

General model specifications

1 Optimizing F1

- Want to prevent as many crimes as possible
- Limited resources

2 Time series modeling

- Time based feature engineering
- Time based train test split:
 - Train: 2012 - 2015
 - Test: 2016

3 Resampling

- To correct imbalanced data

Features

Crime Data

- Previous crime data

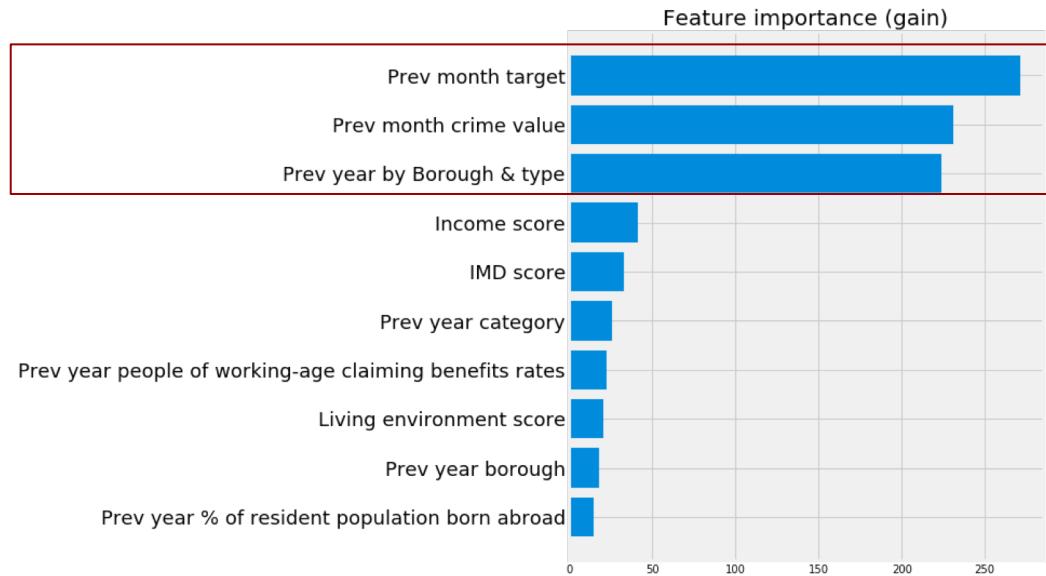
London Data Store

- Previous year demographic and socioeconomic data

- Census data (2011)

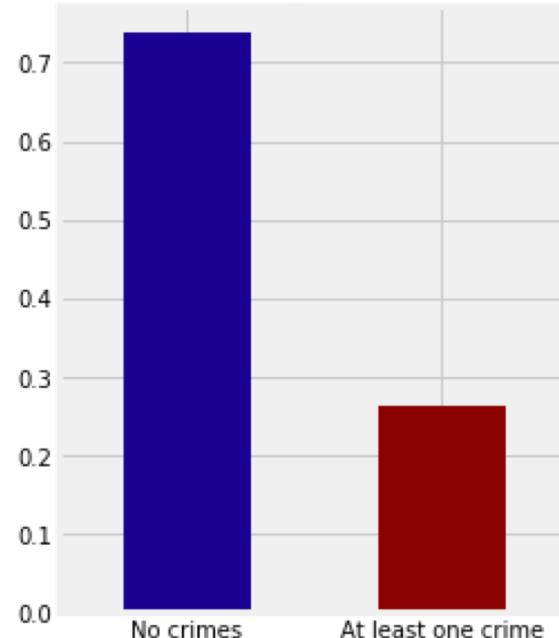
Borough modeling:

XGBoost results	
F1	0.98
Precision	0.98
Recall	0.98
Accuracy	0.96
ROC	0.99

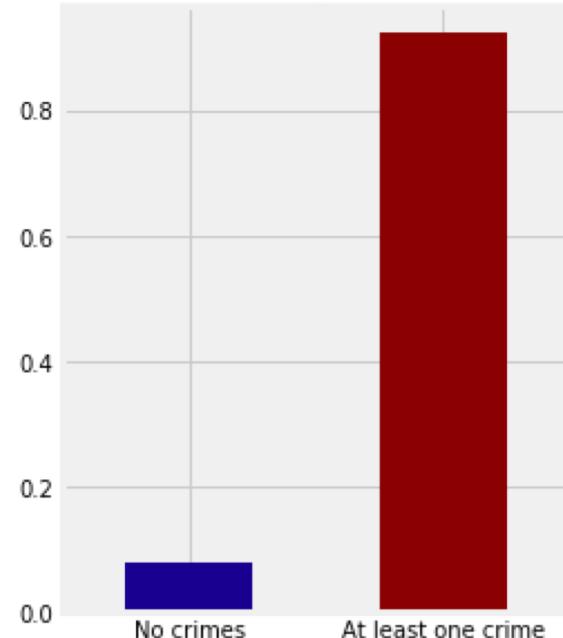


LSOA modeling

LSOA Distribution of target variable for 2012-2016

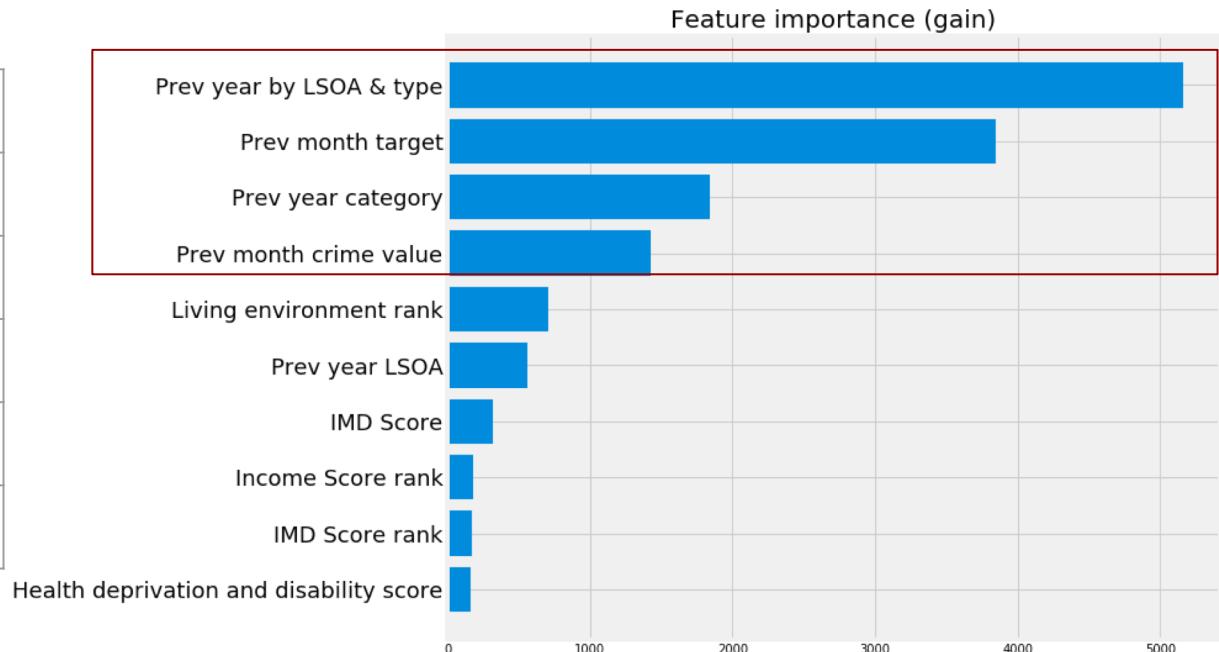


Borough distribution of target variable for 2012-2016



LSOA modeling

XGBoost results	
F1	0.58
Precision	0.56
Recall	0.60
Accuracy	0.77
ROC	0.81



Conclusions

-  Past crime is the main predictor of future crime
-  Socioeconomic and demographic factors are also relevant
-  Further investigation to improve prediction



Questions?

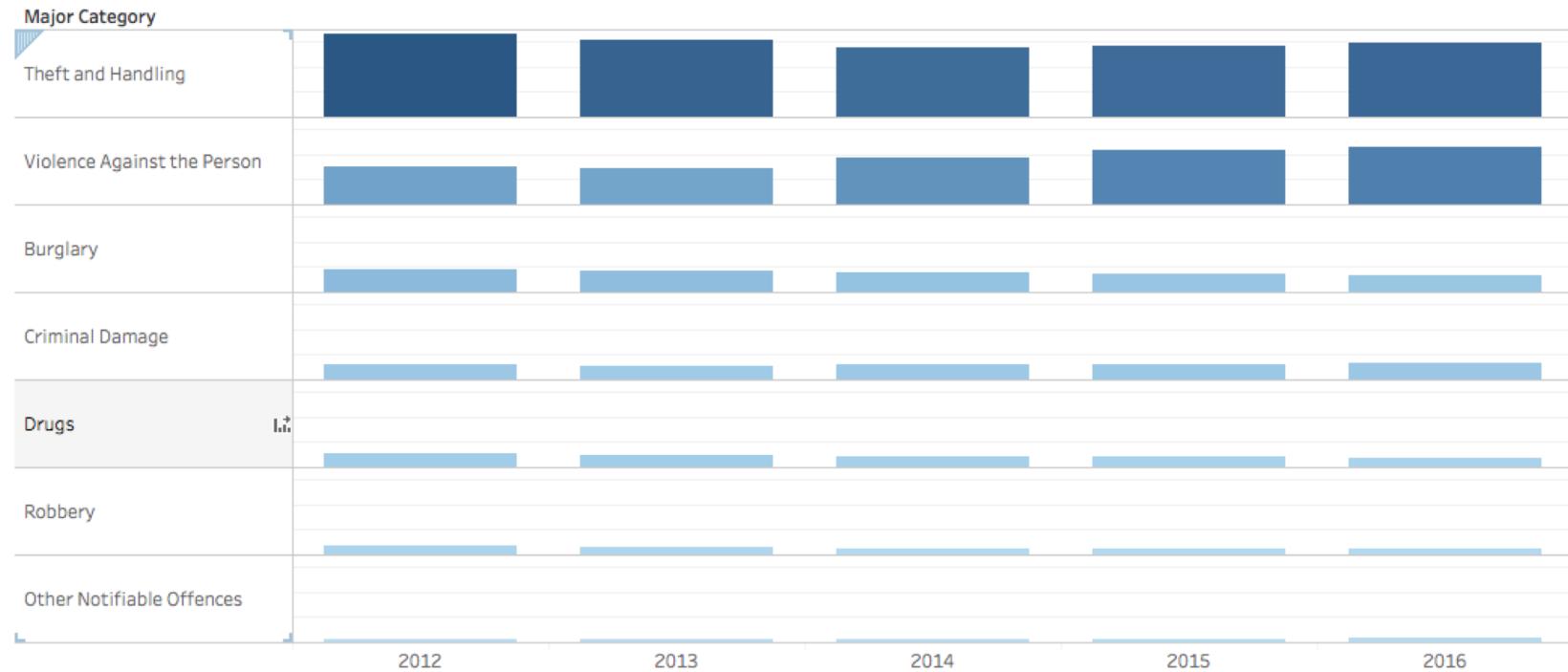
Thanks!

Appendix:

Is there any particular feature associated with
the increment on violence against the person?

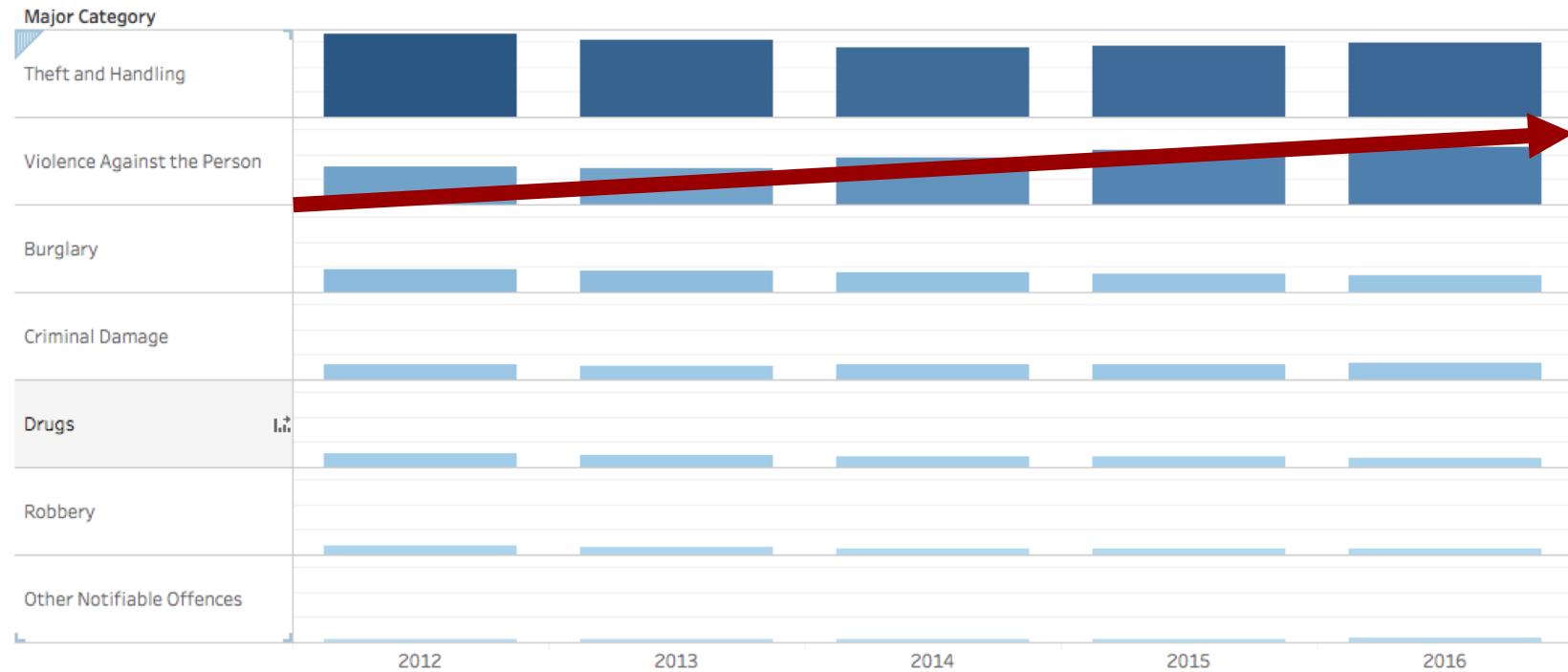
Increment in violence against the person

Number of crimes by major category

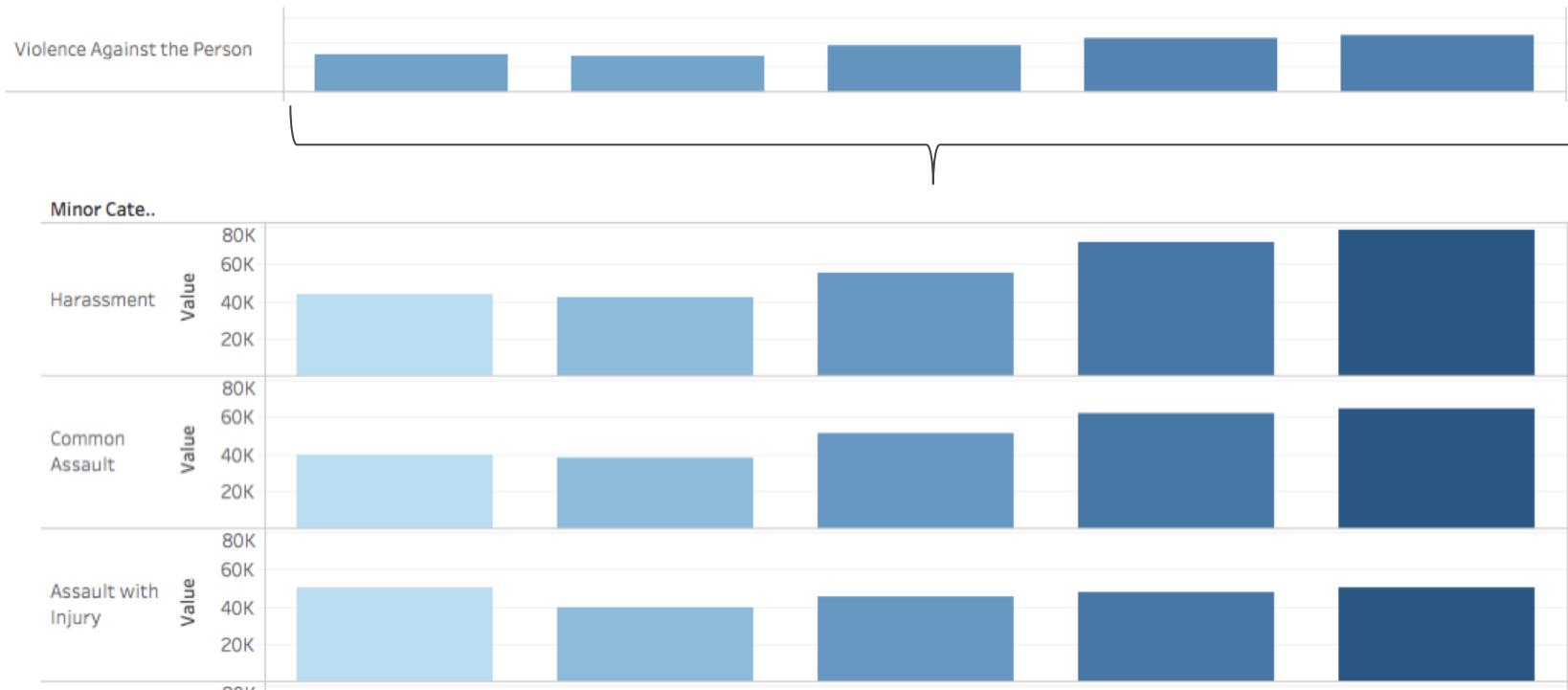


Increment in violence against the person

Number of crimes by major category

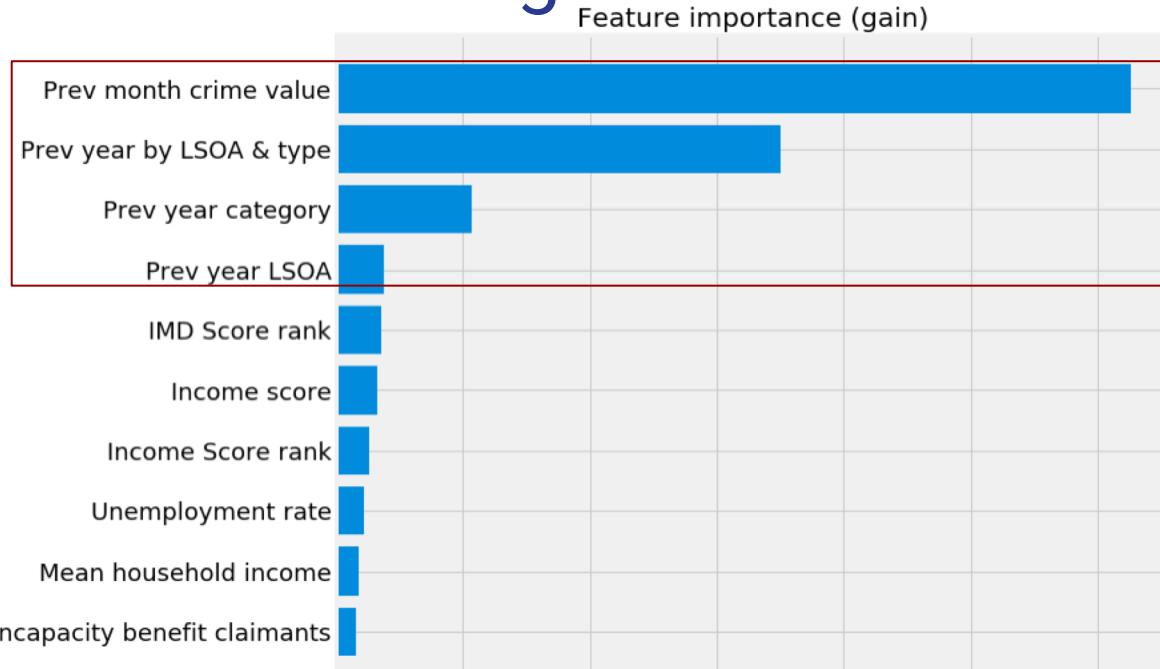


Increment in violence against the person



Violence against person modeling

Model results	
ROC	0.82
Precision	0.68
Recall	0.60
F1	0.64
Accuracy	0.77



Takeouts

Crime history is the best predictor of future crime
Other factors: IMD, Income, Unemployment