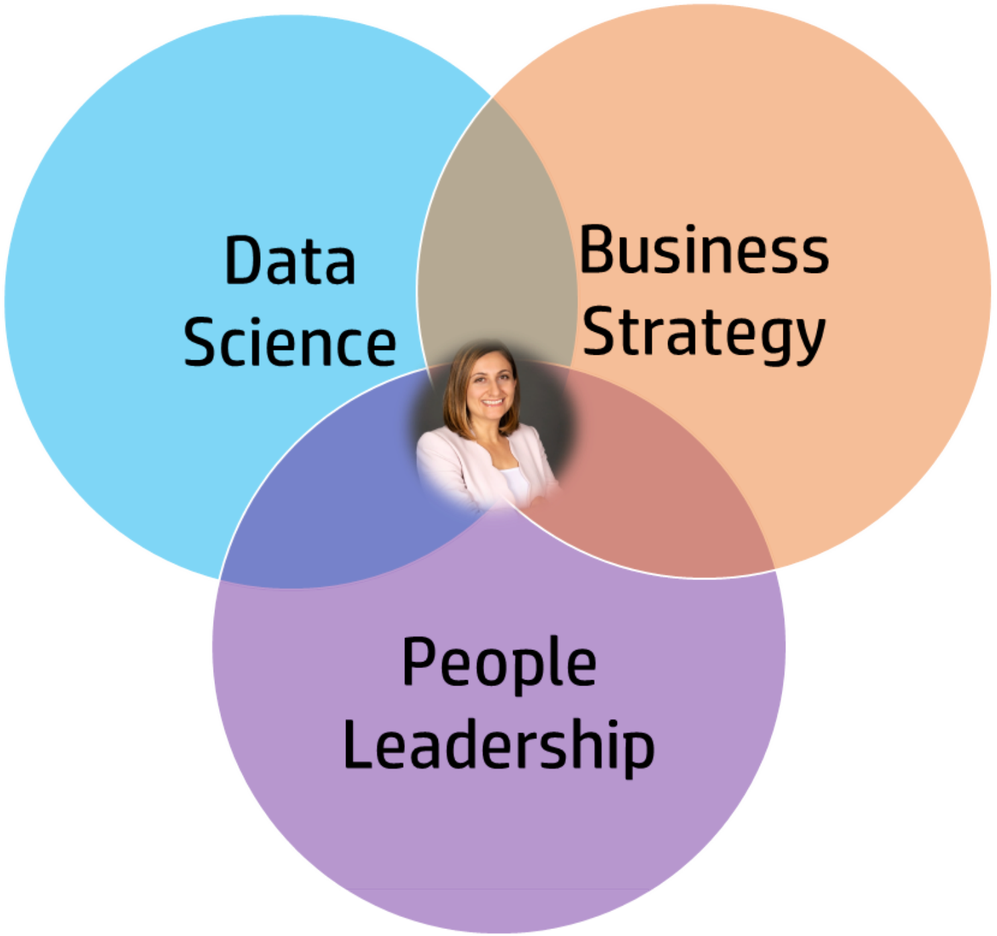
SAFE NEIGHBORHOODS

Predicting crime outcomes in England

by: Marina Mnoyan

July 7, 2022

I DELIVER VALUE AT THE INTERSECTION OF DATA, BUSINESS AND PEOPLE



38%

of investigations in England are completed without having identified a suspect.

In other words, **close to 4 out of 10 crimes do not get solved!**

WHY IS THIS IMPORTANT?

What if it were your car that was the target of a crime?

Your property?

What if it were your loved ones... and they never found who did it?

RESEARCH QUESTION

What factors influence whether crimes are solved in the UK?

Null hypothesis:
Crime types are the sole predictors of crime outcome

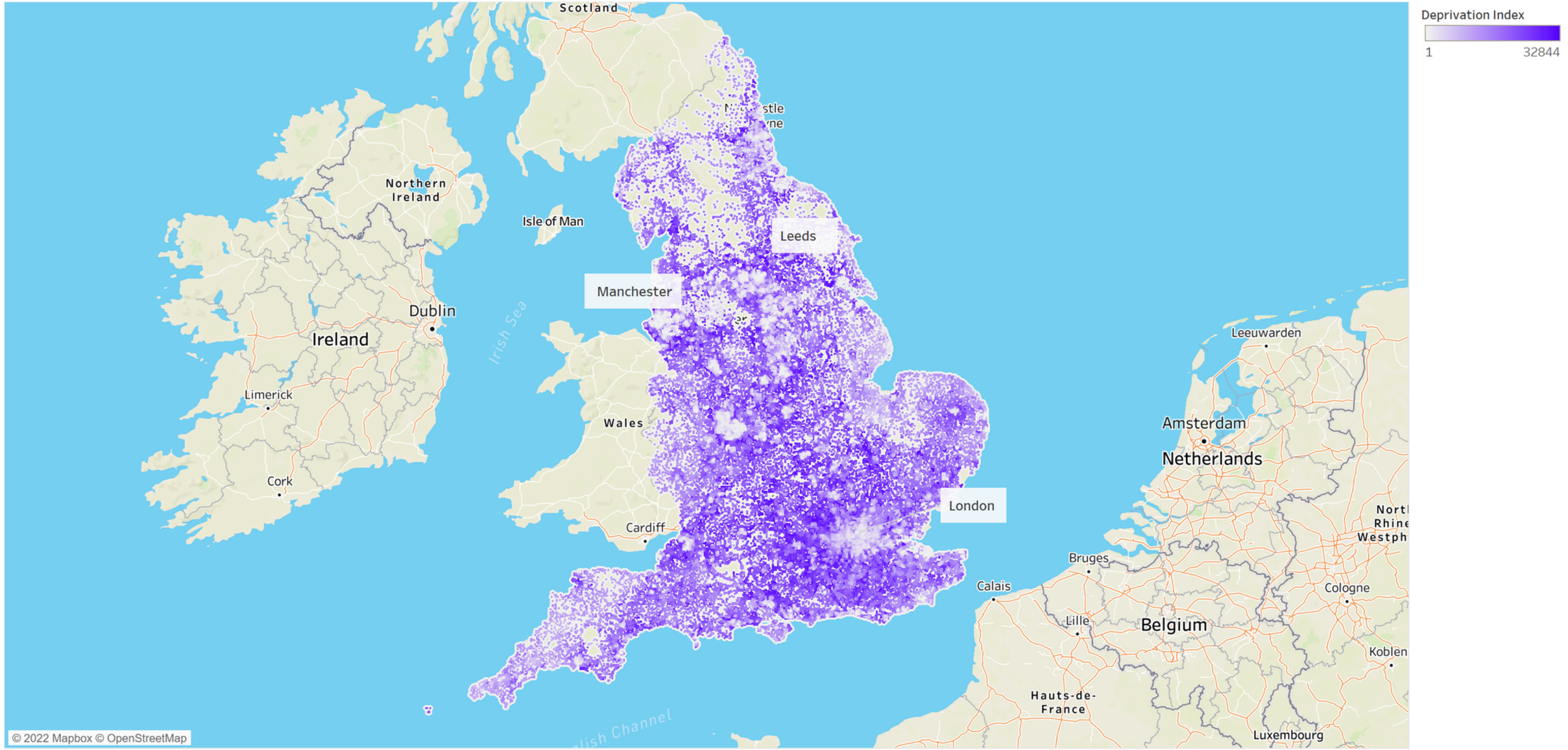
SOCIETAL VALUE

Insights from this project may help the UK Government identify (and possibly address) the socieconomic factors and systemic barriers which are correlated with poor outcomes of crime investigations.

DATA SNAPSHOP: POVERTY ACROSS ENGLAND

Poverty map across England and Wales (as measured by Deprivation Index where 1=Most Deprivated)

Poor areas are spread across the country, mostly concentrated around large cities

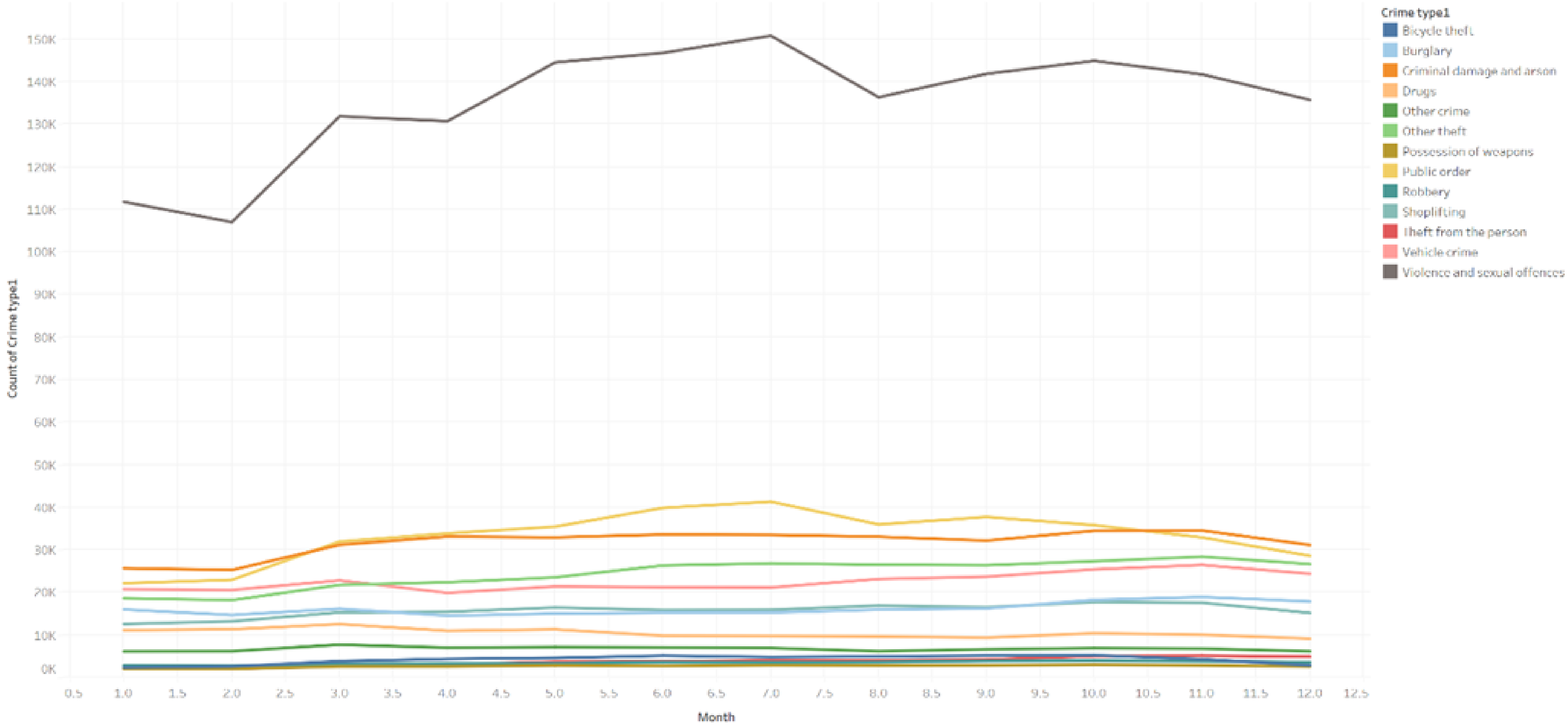


Map based on Longitude1 and Latitude1. Color shows details about Deprivation Index. The view is filtered on Exclusions (Latitude,Longitude), which keeps 472,313 members.

DATA SNAPSHOT: SEASONALITY OF VIOLENCE

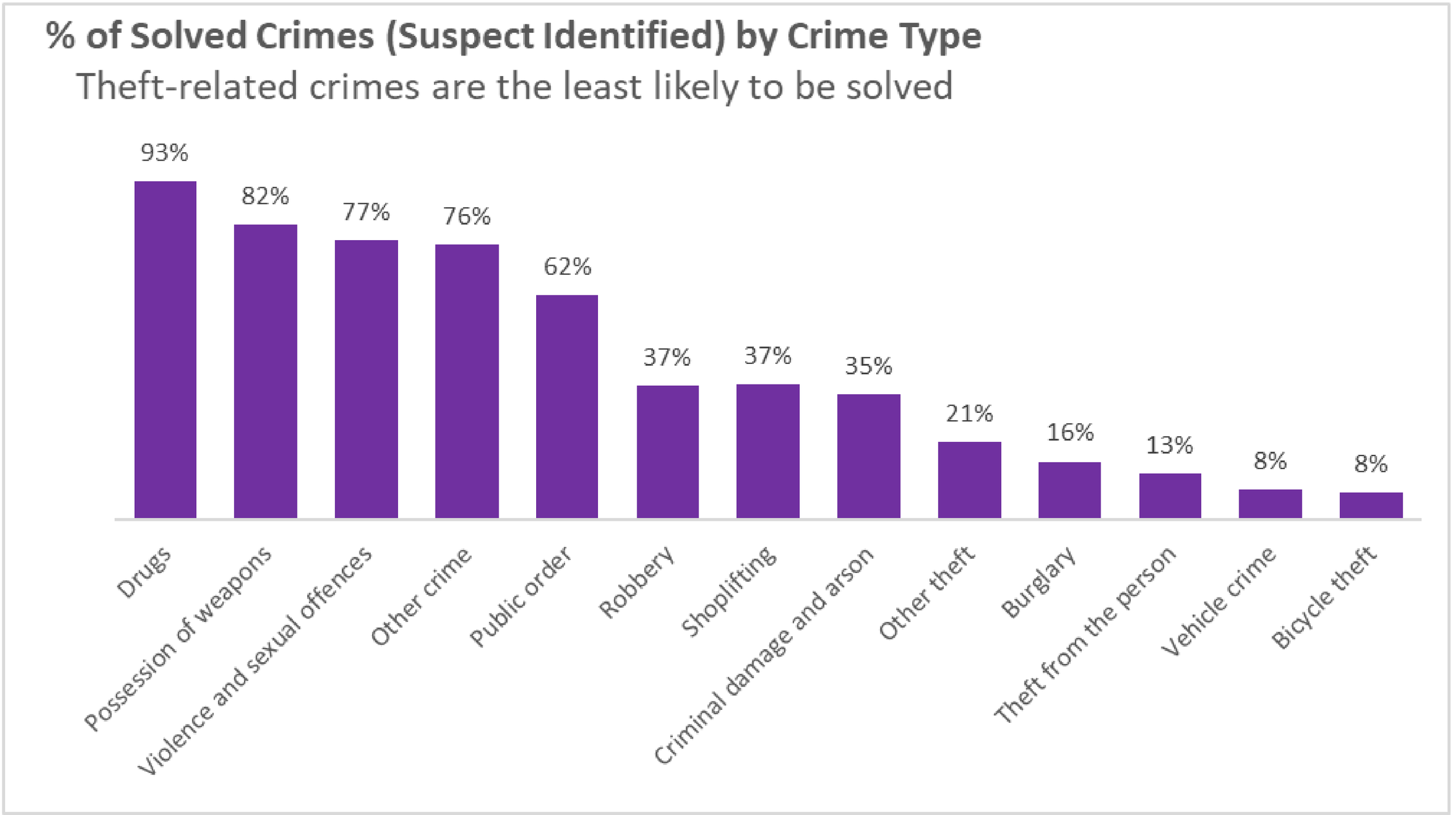
Seasonality of Crime

Violence & Sexual Offences follow a clearly seasonal pattern, peaking during summer months

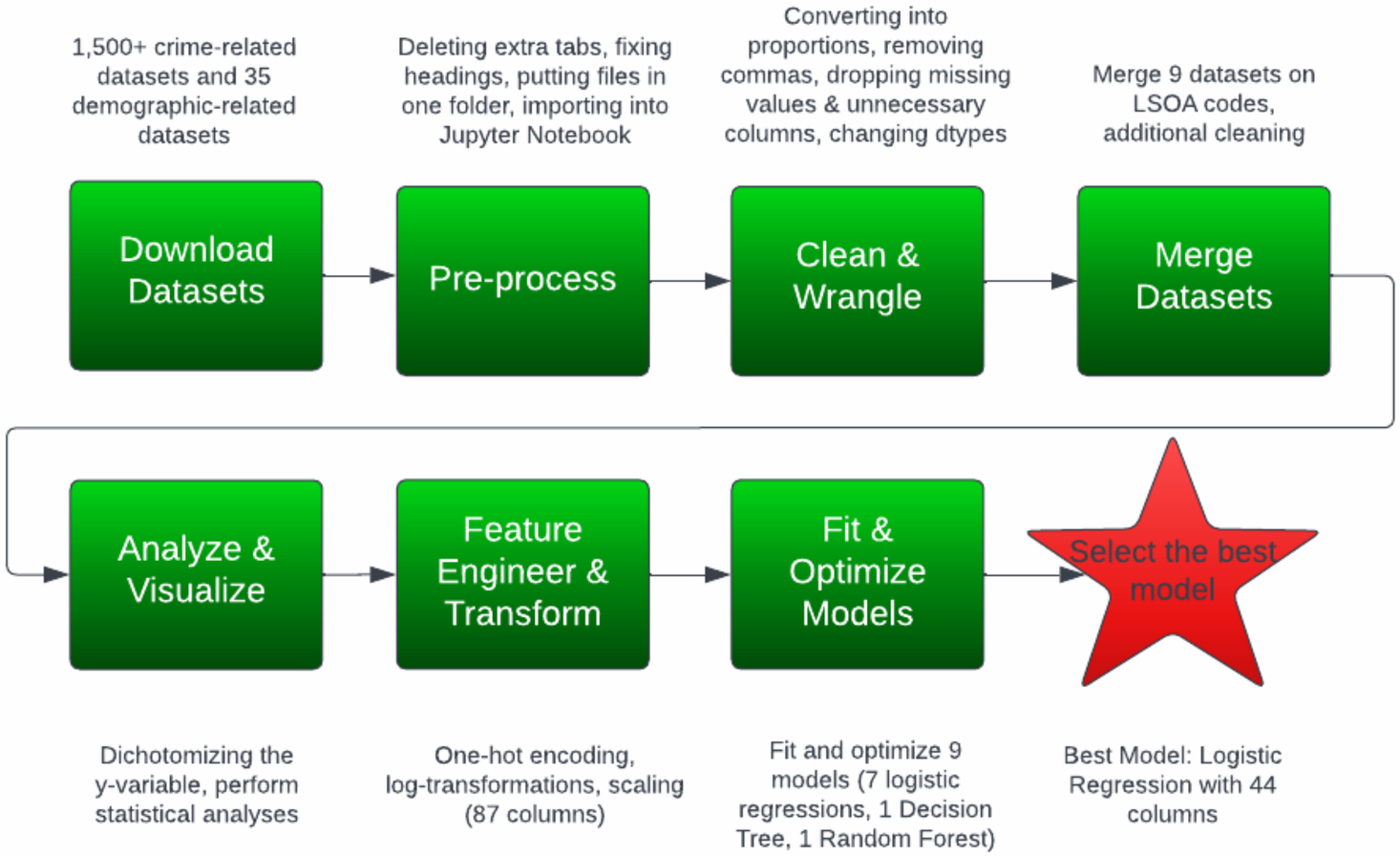


The trend of count of Crime type1 for Month. Color shows details about Crime type1.

DATA SNAPSHOT: CRIME OUTCOMES PER TYPE



IT WAS A LONG PROCESS...



which included lots, and lots, and lots of crashing

IT WAS A CLOSE CALL... BUT THE BEST MODEL WON

8

logistic regressions

1

decision tree

1

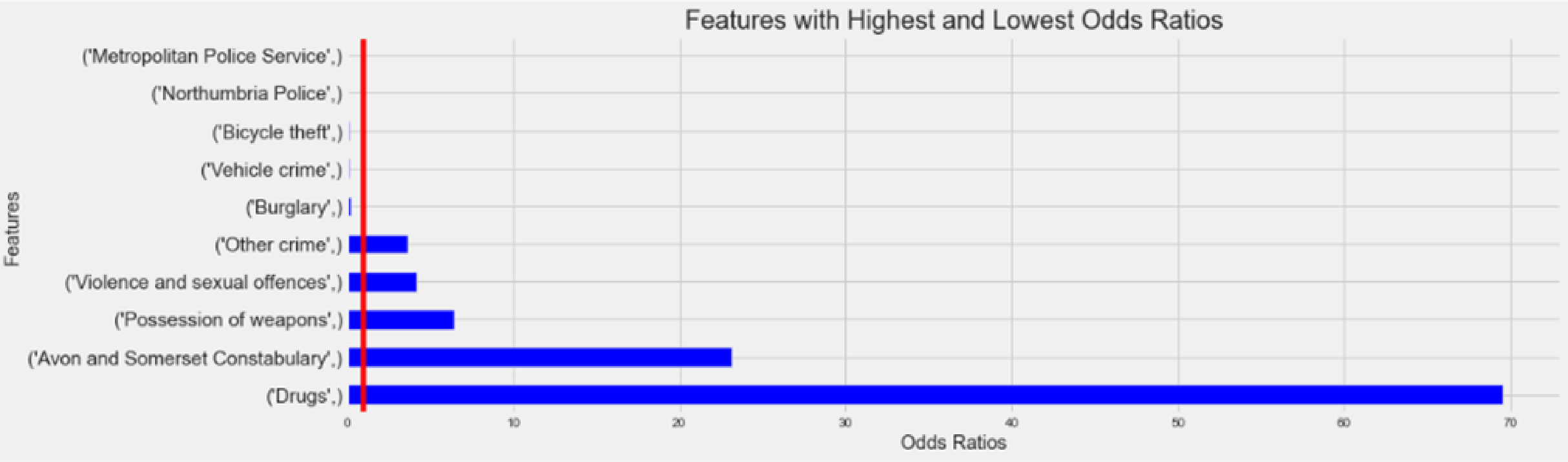
random forest

	Name	Transformation	Optimization	Features	Accuracy on Validation	F1-score on Validation
0	Base Logistic Model	None	None	87	61.45	69.17
1	Logistic Model 1	Scaled data	None	87	81.32	83.06
2	Logistic Model 2	Scaled data	C=0.1	87	81.32	83.05
3	Logistic Model 3	Scaled & log-transformed data	C=0.1	87	81.32	83.06
4	Logistic Model 4	Scaled data	C=0.1	81 (VT)	81.33	83.06
5	Logistic Model 5	Scaled data	C=0.1	44 (VT & Feature Selection)	81.38	83.10
6	Logistic Model 6	Scaled data	C=0.1	28 (KBest)	63.58	71.56
7	Decision Tree	Scaled data (optional)	max_depth = 8	87	81.17	82.55
8	Random Forest	None	n_estimators=31, max_depth=5	87	78.47	81.29

KEY FINDINGS: TOP FACTORS

- Crime Type and Police Jurisdiction are the most significant outcome predictors

- It is recommended that the UK Government organize a knowledge sharing program between **best performing (Avon & Somerset Constabulary)** and **worst performing (Metropolitan Police Service, Northumbria Police)** Jurisdictions



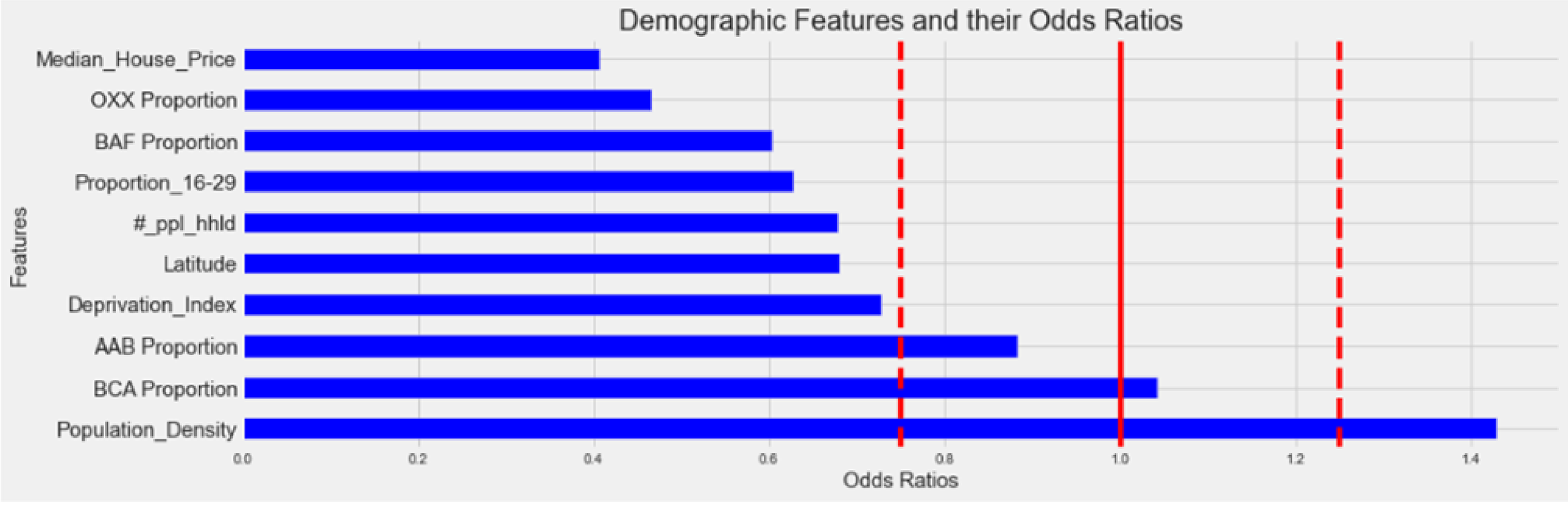
KEY FINDINGS: DEMOGRAPHIC FACTORS

Lower likelihood of crime being solved in neighborhoods that...

- Are affluent
- Have a high proportion of Black African of Other Ethnicities
- Have a younger population (16-29)
- Higher number of people in household
- Are in the North of the country

Higher likelihood of crime being solved in neighborhoods that...

- Have high population density



SUMMARY & NEXT STEPS

- Apart from crime type, demographic factors and police jurisdiction have the highest impact
  - This may indicate potential overpolicing and underpolicing in certain areas and disparities in the police force training

- In the future, Search & Order (i.e., carding) data could be added to confirm the hypothesis around overpolicing in racialized neighborhoods and neighborhoods with low socie-economic status

Thank you!

Marina Mnoyan

Marketing Leader >> Data Scientist | Python/R/Bash/SQL | ...

HOW WAS THE DATA OBTAINED?

1,500+ CSV files downloaded from the *UK Police, Office of National Statistics, Open Data Portal and UK Data Service* websites

- **Individual crime data** from 04/2019 to 03/2022, e.g., crime type, outcome. jurisdiction & location

- **Neighborhood-level demographic** data based on 2011 Census, e.g. age, ethnicity and socioeconomic status