

# Where to commit a crime and get away with it... using PySpark.

An analysis of street crime data in England & Wales.

Disclaimer: **DO NOT COMMIT CRIME**

Alistair Rogers  
November 2019

# Who am I?

- Data Scientist/Engineer/Jack of all Trades - **OnCorps** (Jun 18 - Present)
- Data Scientist - **KPMG UK** (London) (Oct 16 - Jun 18)
- West Swindon born and raised





**We are hiring at OnCorps!**



- Senior/Mid Level Engineers - Any of
  - JavaScript/Angular/TypeScript
  - PHP
  - Python

**Email me:** [alistair@oncorps.io](mailto:alistair@oncorps.io)

---




# An Angry Alistair

DO NOT COMMIT CRIME PLEASE


# How did I get involved in this project?

**BristolLive**

NEWS ▾IN YOUR AREAWHAT'S ON ▾CITYROVERS RUGBYSPECIAL FEATURES



[JOBS](#) [FOOTIES](#) [BOOK AN AD](#) [BUY A PHOTO](#) [SUBSCRIBE](#) [ADVERTISE WITH US](#) [PROPERTY](#) [VOUCHER CODES](#) [MARKETPLACE](#) [DIRECTORY](#) [FUNERAL](#)









News ▸ Bristol News ▸ Crime

## The cars most likely to be stolen in and around Bristol revealed


Car theft is on the rise in Avon and Somerset

Taken from Bristol Post

SHARE 

 COMMENTS

By **Debora Aru** & [Zosia Eyres](#) Content editor  
07:00, 1 DEC 2018

NEWS 



# Calling about my car being broken in to..

POLICE.UK

Find your neighbourhood

Police.uk uses cookies to make the site simpler. [Find out more about cookies](#)

[Home](#) > [Information & advice](#) > [Contact the police](#) >

## 101 - The police non-emergency number

101 is the number to call when you want to contact your local police in England, Wales, Scotland or Northern Ireland - when it's less urgent than a 999 call.

## When should I use 101?

You should call 101 to report crime and other concerns that do not require an emergency response. For example, if:

- Your car has been stolen
- Your property has been damaged
- You suspect drug use or dealing in your neighbourhood

Or to:

- Give the police information about crime in your area
- Speak to the police about a general enquiry



# The record of my call should be here

DATA.POLICE.UK

Cymraeg

Home

Data

API

Changelog

Contact

About


## Welcome to data.police.uk

This is the site for open data about crime and policing in England, Wales and Northern Ireland.

You can download [street-level crime, outcome, and stop and search data](#) in clear and simple CSV format and explore the API containing detailed crime data and information about individual police forces and neighbourhood teams.


You can also download data on police activity, and a range of data collected under the [police annual data requirement \(ADR\)](#) including arrests and 101 call handling.

All the data on this site is made available under the [Open Government Licence v3.0](#).




DOWNLOADS

[Download Police.uk data in](#)



API DOCS

[Access Police.uk data via](#)



CHANGELOG

[See what's new, see what's](#)



# Are there any patterns to street crime?

- Where it happened?
- When it happened?
- What was the outcome?
- What was the crime?

Are there any... **associations?**



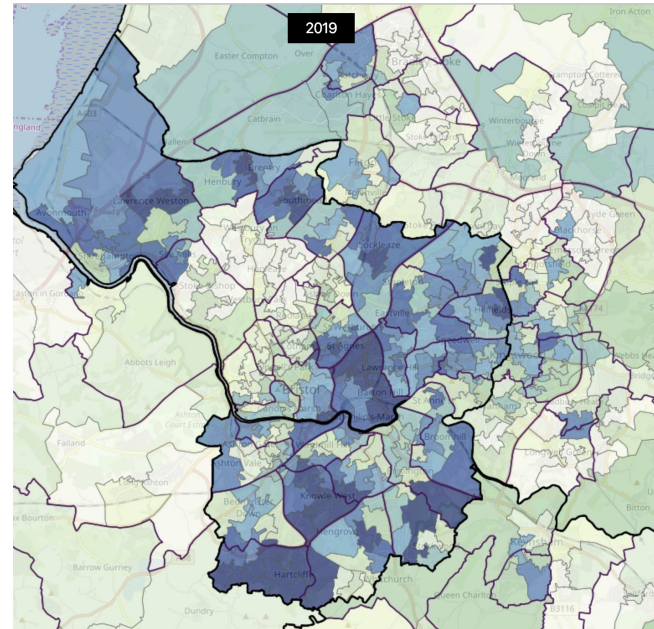


## Example of a record

Crime ID	04b467fb6189761e9ca766964af1a0dc9285390febcec0...
Month	2017-12
Reported by	Avon and Somerset Constabulary
Falls within	Avon and Somerset Constabulary
Longitude	-2.48985
Latitude	51.4369
Location	On or near Palmdale Close
LSOA code	E01014932
LSOA name	South Gloucestershire 032C
Crime type	Burglary
Last outcome category	Awaiting court outcome
Context	NaN

# If you are wondering what LSOA is..

- Lower Layer Super Output Area
- Geographic area - mean of 1500 people each
- **34,753** different areas
- Used in the Indices of Multiple Deprivation, check it out!
- **Darker => More Deprived**



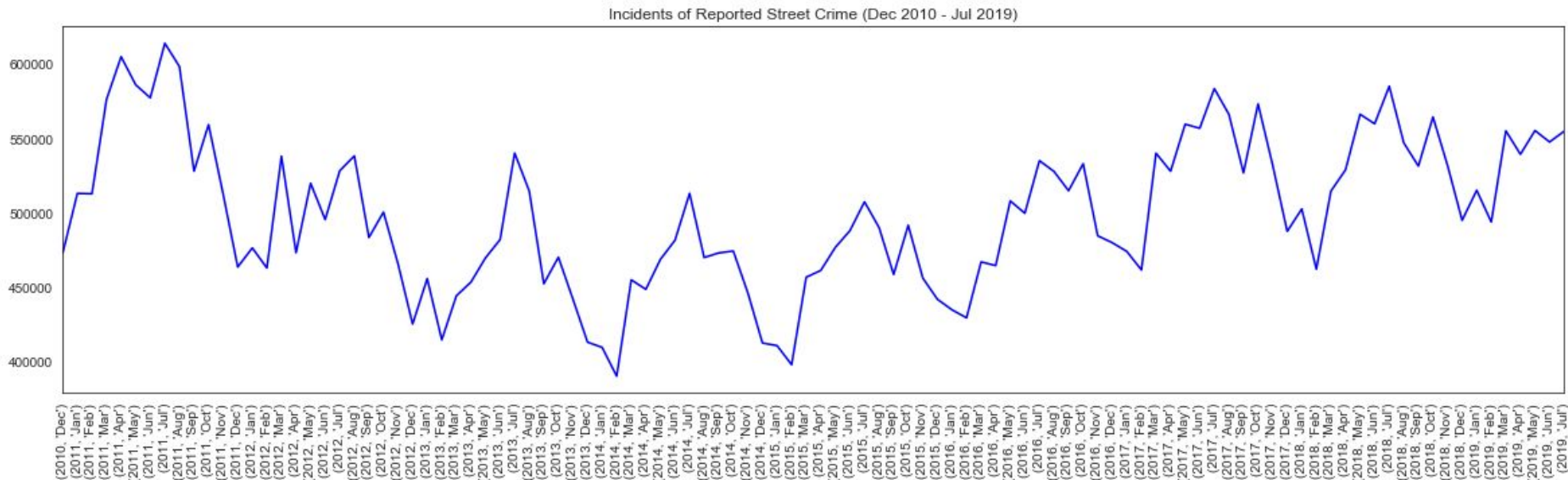
<- Bristol

---

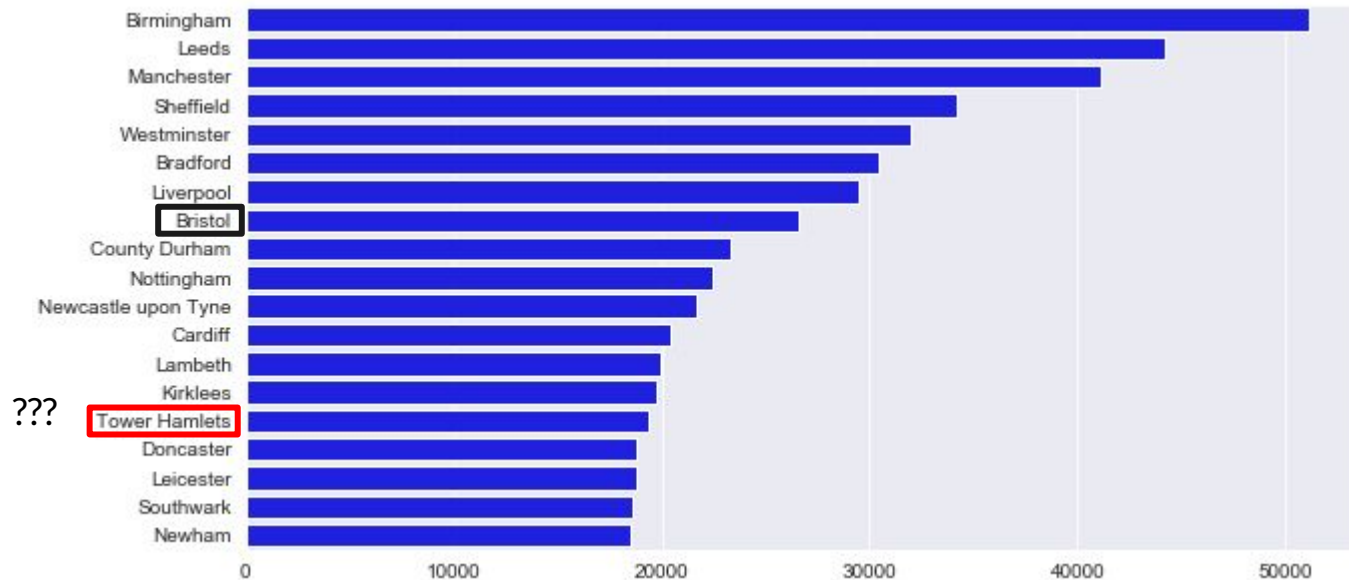
# Some Findings...

DO NOT COMMIT CRIME PLEASE

# A stationary time series?



## Total Incidents of Reported Street Crime by Location (Dec 10 - Jul 19)



---

# Crime Associations

DO NOT COMMIT CRIME PLEASE



# Association Rule Mining

transaction ID	milk	bread	butter	beer	diapers
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1
4	1	1	1	0	0
5	0	1	0	0	0

→ {butter, bread} ⇒ {milk}

# A List of Crime Transactions

items
[Metropolitan Police Service, Aug, Arun, Violent crime, Under investigation]
[Other crime, Metropolitan Police Service, Aug, Aylesbury Vale, Under investigation]
[Metropolitan Police Service, Aug, Violent crime, Under investigation, Babergh]
[Burglary, Barking and Dagenham, Metropolitan Police Service, Aug, Under investigation]
[Barking and Dagenham, Metropolitan Police Service, Investigation complete; no suspect identified, Aug, Criminal damage and arson]
[Offender given a drugs possession warning, Drugs, Barking and Dagenham, Metropolitan Police Service, Aug]
[Barking and Dagenham, Metropolitan Police Service, Aug, Vehicle crime, Under investigation]
[Barking and Dagenham, Metropolitan Police Service, Aug, Vehicle crime, Under investigation]
[Barking and Dagenham, Metropolitan Police Service, Aug, Vehicle crime, Under investigation]
[Barking and Dagenham, Metropolitan Police Service, Investigation complete; no suspect identified, Aug, Violent crime]
[Offender sent to prison, Barking and Dagenham, Metropolitan Police Service, Aug, Violent crime]
[Barking and Dagenham, Metropolitan Police Service, Aug, Violent crime, Under investigation]
[Offender given a caution, Barking and Dagenham, Metropolitan Police Service, Aug, Violent crime]
[Barking and Dagenham, Metropolitan Police Service, Aug, Violent crime, Under investigation]
[Barking and Dagenham, Metropolitan Police Service, Aug, Violent crime, Under investigation]



{Bristol, Aug}  $\Rightarrow$  {Drugs} ?





# Association Rule Mining Terms

## Rule Confidence ( $X \rightarrow Y$ )

- Indication of how often the rule has been found to be true
- Conditional probability of Y given X
- E.g. Out of all of the bike thefts, % resulted in a prosecution?

## Rule Lift ( $X \rightarrow Y$ )

- Measure of positive/negative association
- $> 1$  implies presence of X increases the chances of the presence of Y
- $< 1$  implies presence of X decreases the chances of the presence of Y

## Support (X)

- Frequency Density of the occurrence of X in the data
- E.g. Antisocial Behaviour occurs in 20% of crime transactions



## Conditional Probability Statements

$\mathbb{P}(\text{a drugs offence} | \text{offence committed in Bristol, offence committed in August}) =$   
Confidence(offence committed in Bristol during August  $\rightarrow$   
drugs offense)

An association rule confidence maps directly to a conditional probability statement.

The probability of Y, given the information you know X



# How to find association rules?

## Apriori Algorithm

1. Set a minimum threshold for support and confidence.

Only interested in finding rules for the items that have certain default existence (**support**) and have a minimum value for co-occurrence with other items (**confidence**).

2. Extract all the subsets having higher value of support than minimum threshold.
3. Select all the rules from the subsets with confidence value higher than minimum threshold.
4. Order the rules by descending order of Lift.

# Inefficient on large data... use FP-Growth

How many Rows do we have?

```
[14]: num_rows = police_data_df.count()
```

```
[14]: num_rows
```

```
[14]: 52835178
```

[Overview](#)[Programming Guides](#)[API Docs](#)[Deploying](#)

## MLlib: Main Guide

- Basic statistics
- Data sources
- Pipelines
- Extracting, transforming and selecting features
- Classification and Regression
- Clustering
- Collaborative filtering

## Frequent Pattern Mining

Mining frequent items, itemsets, subsequences, or other substructure dataset, which has been an active research topic in data mining for [learning](#) for more information.

### Table of Contents

- FP-Growth
- PrefixSpan

## FP-Growth

# Rules found...

Probability of consequent  
given antecedent

	antecedent	consequent	confidence	lift
12	Bicycle theft	Investigation complete; no suspect identified	0.719270	1.525113
11	Burglary	Investigation complete; no suspect identified	0.711254	1.508117
10	Vehicle crime	Investigation complete; no suspect identified	0.692018	1.467329
9	Other theft	Investigation complete; no suspect identified	0.666711	1.413669
8	Manchester, Greater Manchester Police	Investigation complete; no suspect identified	0.651027	1.380413
7	Manchester	Investigation complete; no suspect identified	0.650860	1.380060
6	Criminal damage and arson	Investigation complete; no suspect identified	0.641662	1.360556
5				
4	Birmingham, West Midlands Police	Investigation complete; no suspect identified	0.608291	1.289796
3	Birmingham	Investigation complete; no suspect identified	0.607874	1.288914
2	West Midlands Police	Investigation complete; no suspect identified	0.606495	1.285988
1	Greater Manchester Police	Investigation complete; no suspect identified	0.603717	1.280098
0	Theft from the person	Investigation complete; no suspect identified	0.601021	1.274382



## Conclusion

- Steal a bike - 28% chance of getting caught!
- Burgle a house - 29% chance of getting caught!
- Commit any crime in Manchester - 35% chance of getting caught!

**Disclaimer: Don't do any of these**

# Police services are stretched..

## Greater Manchester Police: 43% of crimes not fully investigated

By Andy Bell and Dan Box  
BBC News

23 April 2019

[f](#) [v](#) [t](#) [e](#) [Share](#)



GETTY IMAGES

More than 40% of the crimes reported to Greater Manchester Police are 'screened out'

## Bike thefts rise 'not a police focus'

An A&E doctor has spoken of his frustration after being told police "could do nothing" to find his stolen bikes, even though they were secured with "gold standard" locks in an area covered by CCTV.

Mike Brooks stopped commuting by bike after two of them were stolen from Cambridge North railway station in the space of a month. In one case the entire bike rack was removed from the ground.

He said: "We're the fifth-richest nation on the planet yet I don't understand why we're having to pick the crimes we follow up."

British Transport Police had 6,000 reports of stolen bikes this year and the number of thefts from railway stations has increased by 40% over four years.

# STAY SAFE!

---

DO NOT COMMIT CRIME PLEASE