Street Crime Analysis In England's Major Cities Before and During COVID-19

Team members: Diana Mutheu, Cherry O'Connell, Lina Ziaka, Iwona Izdebska, Pallavi Atkarne

Introduction

In this project we are examining street crime levels of 6 major cities in England before and during COVID-19 namely: London, Leeds, Bradford, Sheffield, Liverpool and Bristol.
More specifically the aims and objectives of the project were: see how different crimes are affected by lockdown severity and pandemic period, see if crime can be predicted using machine learning methods given the covid and lockdown severity information and see what the top street crimes are before and after the pandemic in the different cities.
In the report we first give a background of the issue and our specific analysis as well as how it can be useful. Next we go through the analysis stages: the data gathering, preprocessing, plotting and Machine Learning analysis. We describe our methodology and resources we use as well as the results and our interpretation of them.

Background

Street crime is a very prevalent issue in England with many English cities having one of the highest criminality rates. Previous research has shown that the pandemic has influenced crime rates with many crimes decreasing while others increasing. For example, domestic assault has increased
(https://commonslibrary.parliament.uk/domestic-abuse-and-covid-19-a-year-into-the-pandemic/).

Questions:

1. Does the Covid-19 Pandemic have an effect on street crime levels?
2. What are the top street crimes, crime locations and crime outcomes before and during the Pandemic?
3. Can Machine Learning predict street crimes and give us insights about the data?

In this project we are specifically looking at the effects of the pandemic period and lockdown severity (measured with 'stringency') on street crime. Upon starting the project we were expecting that street crime would decrease when lockdown severity increased (negative correlation). That is because while people stay more indoors street crime would naturally show a decrease with less people in public areas.
The England police departments spend £400 million per year on combating crime (https://www.gov.uk/government/news/police-to-receive-more-than-15-billion-to-fight-crime-and-recruit-more-officers). Thus, examining the effect that the pandemic has had in crime is

important since it can help the police devote more resources in street crime in time periods where it is most needed. Stringency can also provide added valuable information on what the police can expect in future lockdowns. So, the target audience of our project is the police departments as well as civilians of those cities since they are majorly affected by street crime.

Steps Specifications

Framing questions

Initial brainstorming discussions were conducted to decide on the topic and the specific direction we wanted to take. After we had a rough idea about the topic we started deciding on the specific questions and analyses we should use to answer those questions efficiently.

Data gathering

The crime data for the cities was obtained from the police database (https://data.police.uk/). The data was from January 2018 to September 2021 covering the pre and during Covid-19 period. This involved merging of the datasets since they were stored in a monthly order for each year. Information about the lockdown severity was added to the dataset using the Oxford COVID-19 Government Response Tracker API (https://covidtracker.bsg.ox.ac.uk/). This API provides the lockdown severity in a measure called stringency ranging from 0 (no lockdown) to 100 (strictest lockdown measures). The resources were found by browsing the internet and testing various different ones until appropriate ones were found.

Preprocessing

The police dataset for London initially consisted of more than 4 million rows with 12 columns (csv file). The other 5 cities were combined in a separate csv file that consisted of 2 million rows and 12 columns. For the 2 police datasets, missing values and duplicates were first removed. Next unnecessary and blank columns were also removed (unrelated column 'crime_id' and column 'context' which was empty). The 'Falls within' column was also removed as it was a duplicate of the 'Reported by' column which both showed the police force in which the crime was reported to. The columns that we kept were: 'month' (YYYY-MM), 'location', 'longitude', 'latitude', 'lsoa_code' (neighbourhood area numerical information), 'crime type', and 'last outcome category'. We also added the 'borough' column for the London dataset. We realised that some crimes reported were not within London by use of borough extraction and had to drop those rows as well.

After that, the numerical columns were also one-hot encoded in order to be used in machine learning (ML) and other analyses where categorical data cannot be provided as input. The month column was also turned from date-time type to numerical using the astype.int() method in order to be inputted in the ML algorithms.

Additional columns that were added:

1. Stringency: the stringency value for each month was extracted from the API and was averaged per month since the month column in the police dataset did not have day information.
2. Covid: a covid binary column was added with 0 representing before covid and 1 during covid.
3. Crimes_count: in order to use ML on the total crimes a column with the total crimes per month was also added.

Lastly, feature analysis and selection was also conducted using 2 ML models Random Forests and xgBoost. That was done as a preprocessing step before ML in order to keep only the most important features.
An issue that occurred from having the same values repeating across many rows (month, crime count, stringency) was that this created many new duplicate values after keeping only the most important columns for ML. So before ML, the data was again tested for duplicates and any found were removed.

In depth analysis

After preprocessing the data as both datasets were millions of rows long, and analysis would be challenging to conduct (especially for ML), an SQL database was created. In order for all of us (and the instructors) to be able to access it and derive data from it, it was hosted online in AWS' Relational Database Management (RDS) system.
Next, we performed exploratory analysis, to start answering our research questions. In this step, we plotted before and after covid to explore how covid has affected crimes. Bubble plots of the most important crimes for both datasets were also plotted along with other various plots. This helped us understand the nature and underlying patterns of the data better. We also plotted the crimes in city maps using the Mapbox tool and API (https://www.mapbox.com/).
After the explanatory analysis, we focused on exploring the relationships deeper, by examining the effects of stringency on the crime types. For that we created a correlation matrix and did multiple regression. However, the dataset was nonlinear so we also did ML regression and classification. We also performed clustering using K-means to see if there are separate clusters for the data before and after COVID-19.

Implementation and execution

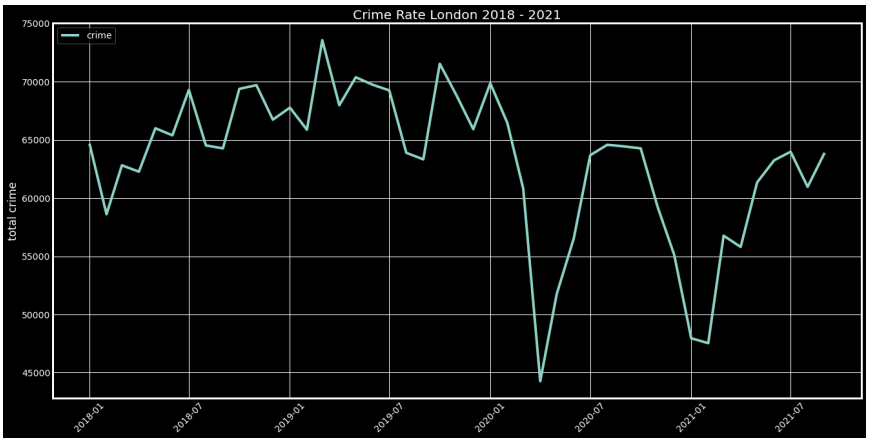
Approach: The approach we took for the project used agile methodology as we did short scrum meetings and also longer meetings (sprint reviews) at the end of a sprint. We also used Jira to keep track of the tasks that needed to be done and all the deadlines. For version control we used GitHub, where we uploaded all the code we finished for the tasks we were assigned and this also made code review easy. We also used Google Colab as it allowed all of us to edit code at the same time and also to run ML since our laptops were taking too long due to the data size.

Work allocation: Team members were responsible for different tasks and for implementing different methods, picked from a product backlog. For example, some members conducted the preprocessing, some the visualisations and some the ML and some the correlation analysis etc. For some tasks all of us contributed for example for data gathering, report writing etc. The product backlog was updated during a weekly sprint planning meeting, which served also as a retrospective meeting for the previous sprint.
Tools: As for the tools and resources we used various different Python libraries like matplotlib, scikit, seaborn, pandas, scikit learn, datetime, xgboost, plotly, numpy. Other tools were also SQL, AWS and Jira and Github issues.
Challenges: Some of the challenges we faced had to do with the large amount of data which was also one of our accomplishments since we managed to tackle the problem by creating an AWS database. Working with such large data gave us a lot of insight on Big Data and we gained new skills. Another challenge was that the data averaged across each month, which in the end gave us a lot of values that were repeating which hindered in part the ML analysis.

Result reporting

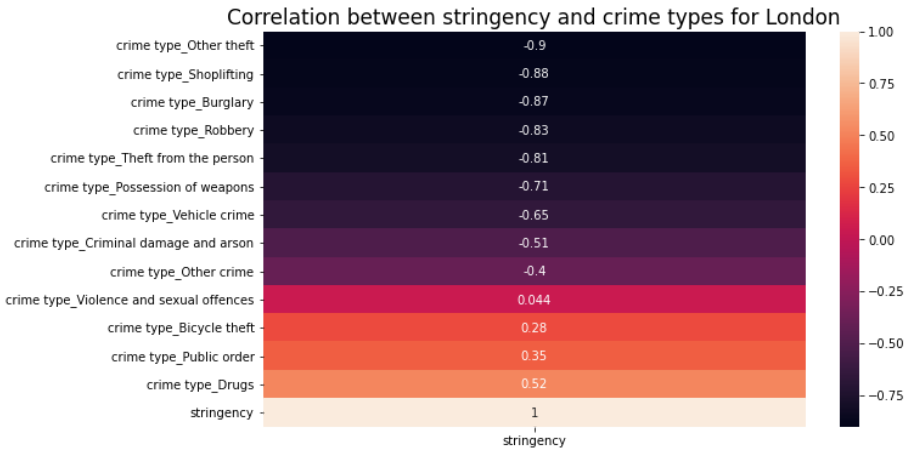
1. Does the Covid-19 Pandemic have an effect on street crime levels?
Overall, street crime levels have significantly dropped during the pandemic.

Plot 1: Crime rate London 2018-2021



The results of the correlation matrix showed a negative correlation between most of the street crimes. The negatively correlated street crimes can be best described as relating to theft apart from weapon possession and vehicle crime. The only theft-related crime that was positively correlated with stringency was bicycle theft. The most interesting finding is that drug misuse was moderately positively correlated with stringency. Public order showed a weak positive correlation.

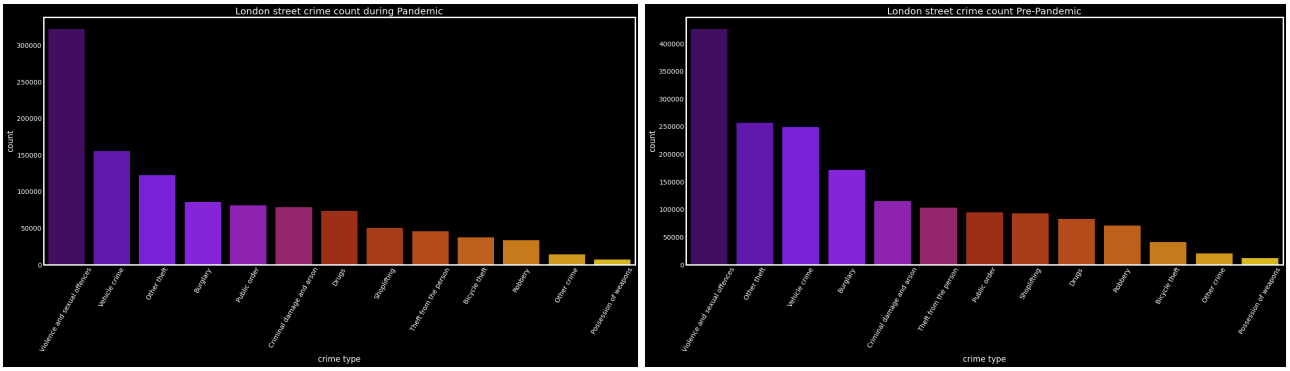
Plot 2: Correlation between stringency and crime types



2. What are the top street crimes, crime locations and crime outcomes before and during the Pandemic?

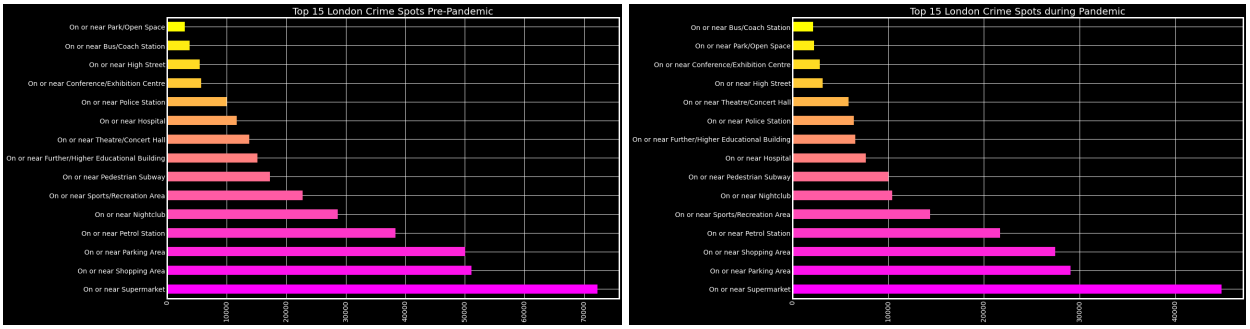
The relative importance of crime types was affected by the pandemic. While violence and sexual offences was by far the most prevalent crime in London both before and during the pandemic, the immediately following crime types have switched places: other thefts and vehicle crime. Drugs and public order have increased in prominence during the pandemic and moved before theft from person. Same pattern applied to bicycle theft and robbery.

Plot 3: London street crime count during the pandemic



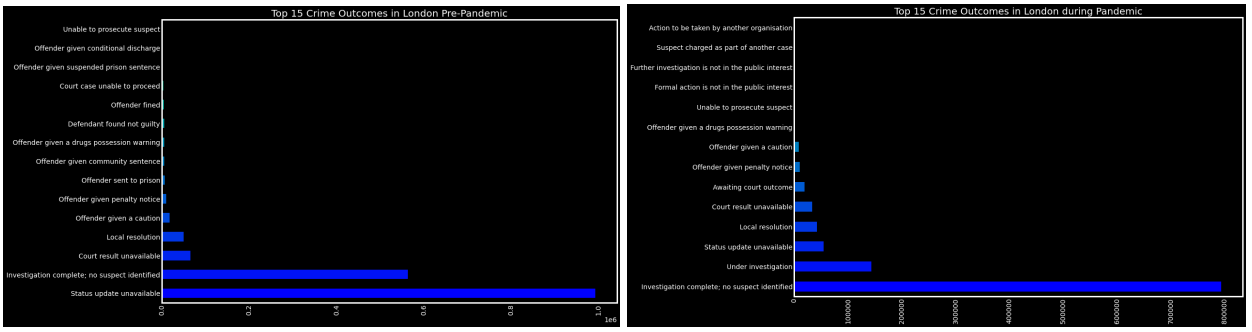
With relation to crime spots on or near supermarkets being the most frequent crime spot, there was a change in relative importance of other crime spots.

Plot 4: Top 15 London crime spots pre and during pandemic



During the pandemic the number of unresolved crimes has risen:

Plot 5: Top 15 crime outcomes in London pre and during the pandemic



With relation to the remaining cites, the patterns mirrored those in London, therefore we refer the reader to the main analysis.

3. Can Machine Learning predict street crimes and give us insights about the data?

The results of the regression using Random Forests (RF) and K-Nearest Neighbors (KNN) are shown in the table below:

Table 1 and 2: ML results for London

Regression				
	R2 score	MSE	MAE	N
Multiple regression	0.87	11.3 (RMSE)	9.36	45
RF/KNN	1	0	0	160.686

Classification					
	Accuracy	Precision	Recall	F1 score	N
RF/KNN	1	1	1	1	65.606

N is the number of samples. Results were the same for both RF and KNN. The features were the columns deemed important during the feature analysis and the target (y) variable was the crime count.

Table 3: Clustering Results for London

Kmeans cluster	Covid mean value	Crime count mean
0	1	60.052

1	0	66.225
2	0.2	64.877

There were 3 clusters with two occurring completely before and after covid. The last clusters belonged 20% before covid and 80% after as can be seen by the mean.

Conclusion

Among the crimes in England, public order, bicycle theft and drug misuse are seeing the most dramatic increase in recent years, and violence and sexual assault has a slight increase in some cities.

One strategy that local authorities may consider to fight those crime types is situational crime prevention. Intuitively, one might think this could be through tackling the increasing violence and drug-related crimes by limiting the number of nightlife venues and social gatherings as potential anti-social behaviour hotspots (What Works to Reduce Crime?: A Summary of the Evidence - gov.scot (www.gov.scot)). Our analysis shows, however, that while theft, robbery and burglary significantly decreased during pandemic restrictions, those restrictions had less influence on criminal damage and arson, or even some degree of aggravating influence on violence and sexual offences or social order disruption. Importantly, drug misuse increased during Covid, showing that people when deprived of opportunities to meet others, turn to recreational drug use. This means that social milieu might not be the root cause of these types of crime and socialising can even be ameliorative for people with such antisocial tendencies. We recommend, local authorities use their funding to set up youth clubs or other places where people could engage in meaningful activities (Drug misuse in England and Wales - Office for National Statistics (ons.gov.uk)).

Moreover from the correlation analysis we also see that there was an influence of stringency in crime types. The correlation matrix showed a negative correlation between most, but not all crimes and stringency.

The ML results for both regression and classification show 100% accuracy. After troubleshooting, we believed that maybe since the crime count was repeating for thousands of rows it was too easy for the model to predict as it was not a continuous value. If we removed all the repeats we were only left with 45 rows which is a very low number for ML. Many different versions of the ML models were tested and by tuning the hyperparameters to make the model ‘worse’ the accuracy did drop to 58%. This might indicate that instead of a problem in the model fitting the data were just easy for ML to predict. Unsupervised clustering however showed better results, showing different clusters forming before and after Covid. This indicates that there is a difference between before and during Covid.

To conclude, our analysis can contribute to the better understanding that Covid-19 and lockdown severity had in street crime and can help the police force in tackling crime during the pandemic.