Project: Explore the recorded crime rate in London and the number of police officers all within the last 24 months.

I set out to explore the well known decreasing number of metropolitan police force officers and the increasing number of crimes in London.

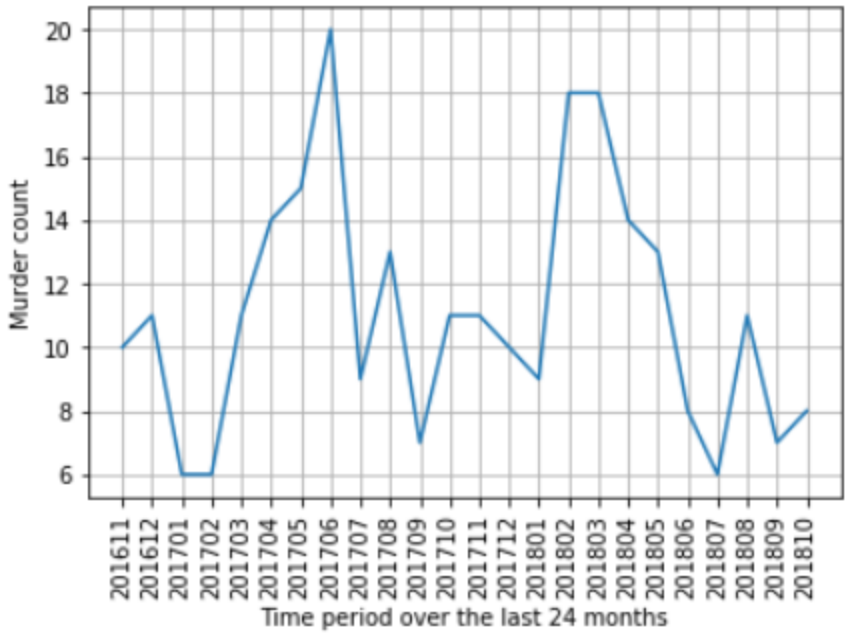
https://www.statista.com/topics/4627/crime-in-london/ - summary of the increasing crime rate. Any local London news outlet will have made clear the increasing levels of crimes in London.

Recorded crime rate data is available from https://data.london.gov.uk/dataset/recorded_crime_summary it is a regularly updated data set, which made exploration more relevant. The reason why I focused on the last 24 months is due to the fact that the police recommend that last 24 months should always be used (on the website). All data was in a CSV format from the London Datastore and was imported using the python io and requests library.

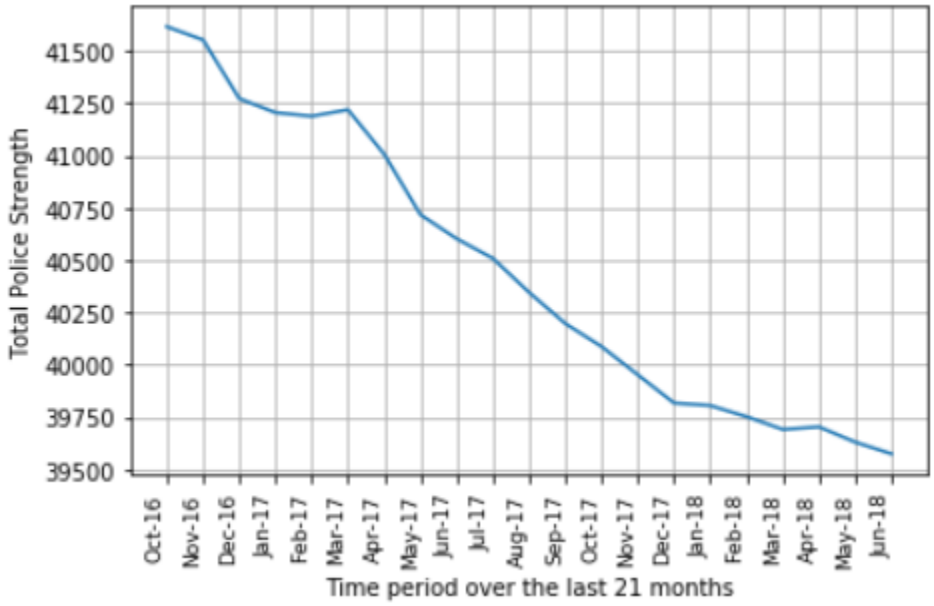
Preprocessing was not needed for the data, as the datasets were updated regularly. I was able to select any column and row quite simply using column labels, however I struggled to process the current date format into a cleaner format.

I had to narrow down the search space, data for recorded crime was grouped by borough in London. Focusing on each individual borough would be too complex, so I added all of the specific crime counts of each borough to make a specific total crime count of the whole of London. Narrowing down the search space and considering all the factors that can contribute to a phenomenon such as increasing crime rate was one of the most important lesson I learnt.

All data was recorded monthly. The first variables I explored was the Murder count. This is only variable in the crime data where every occurrence must be recorded, it is almost impossible for a murder to be underreported or overreported. All other crime types have drawbacks whereby crimes can happen, but the victim does not report them for various reasons such as a belief the crime would not be solved, distrust in the police etc...



I compared this variable with the Metropolitan police strength data. The MPS strength data represents a current head count but is split between regular Police officers, Police staff and PCSO (volunteer part-time officers), all of these different parts of the police force were combined to create a general indicator of the police count.

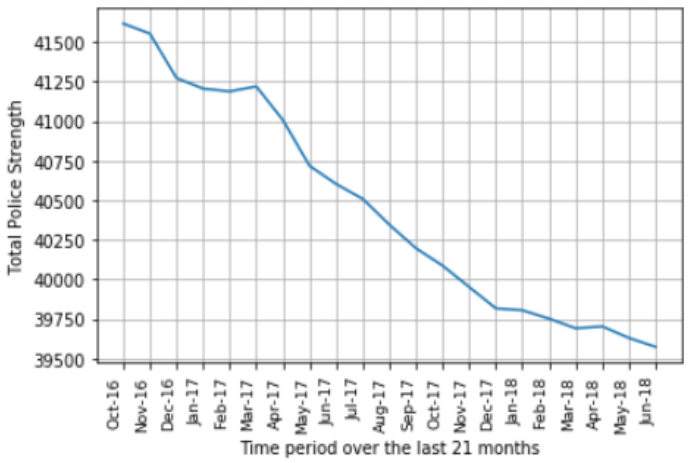


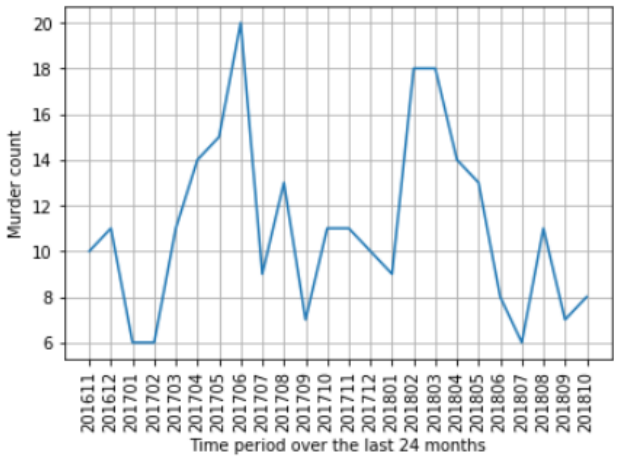
https://data.london.gov.uk/dataset/police-force-strength

Unlike the recorded crime data, the last update for the police force strength dataset was June 2018, a 3-month gap. But it provides a clear indicator that there are less police officers and staff. Line graphs were used because they are the most suitable graph for measuring a variable over time and with a grid you can see a rough estimate of variable values. With line graphs you can also see the genral direction of a

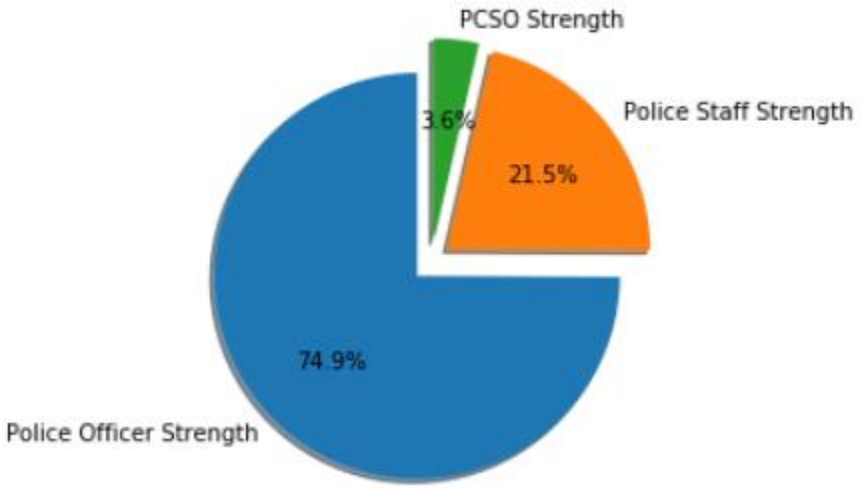
variable and any pattern, for example in the above police strength graph it is clear that there is a drastic decrease in total police strength over the last 21 months.

Line graphs also allow comparison between visualisations, which can help in finding any correlation between two datasets, for example the murder count can be compared with the total police strength;

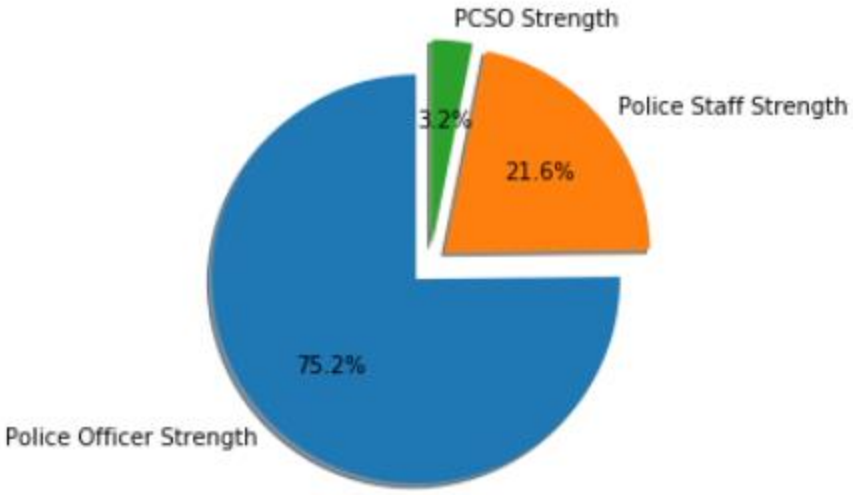




I also created pie charts to show the composition of the police force, as you can see there is a reduction in police numbers, not a restructuring of the police force, the human structure of the Metropolitan police force has remained relatively the same.

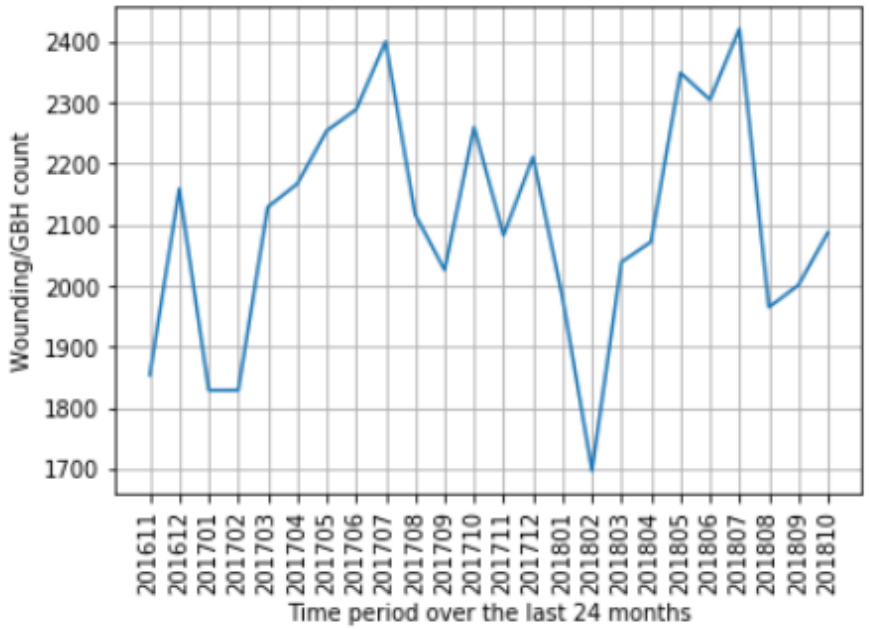


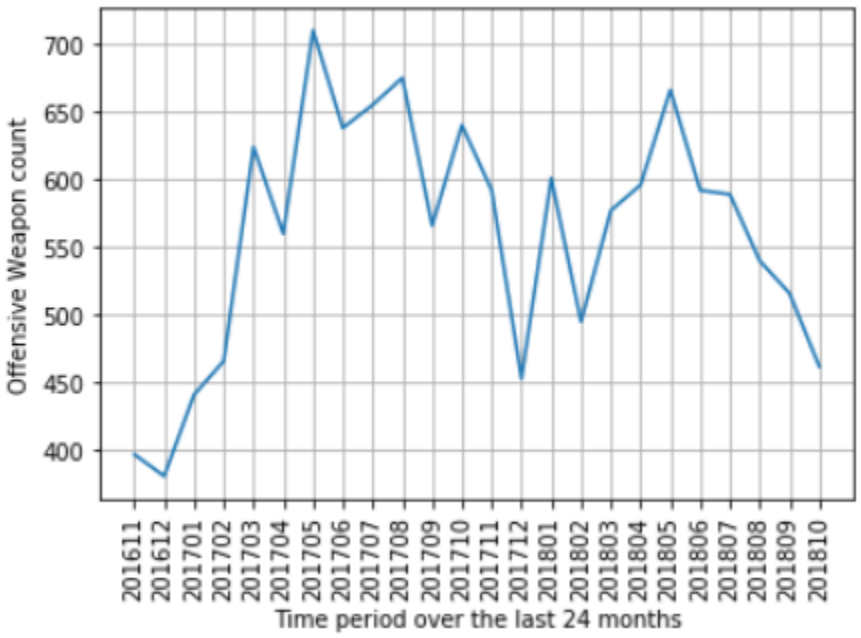
October 2016 Metropolitan Police Force composition

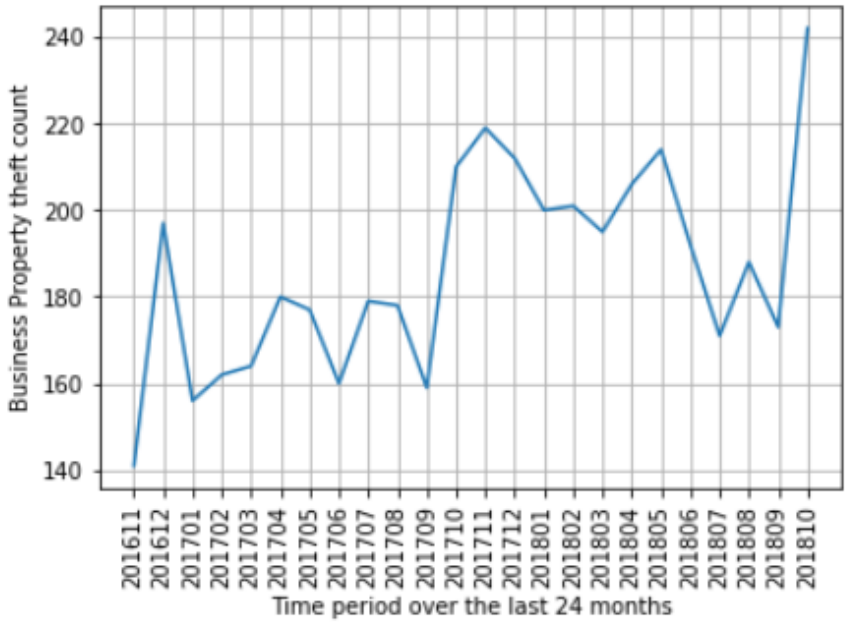


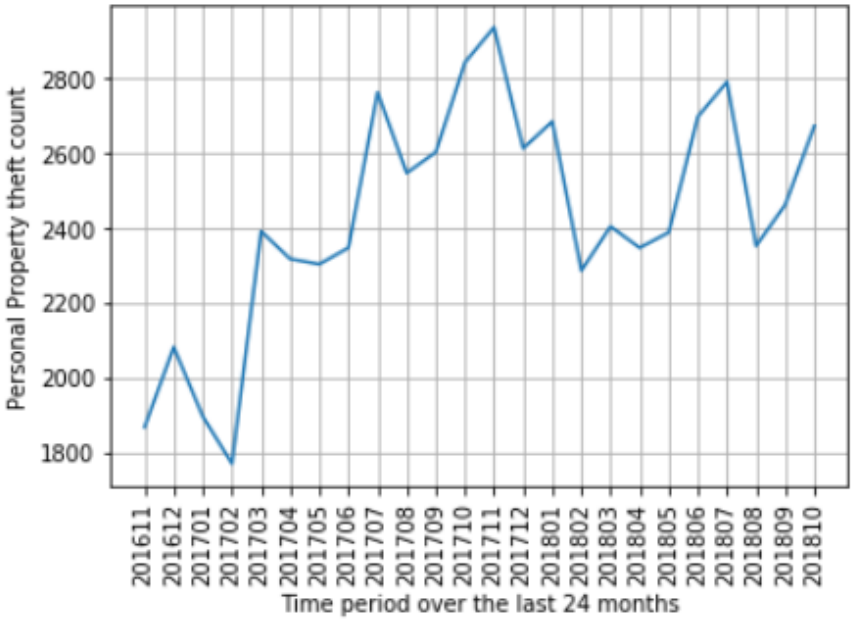
June 2018 Metropolitan Police Force composition

The line graph visualisation worked well compared to using a scatter or bar graph, it clearly shows any patterns, trends and general direction over time while occupying less visual space. I created other visualisations and compared these to the police strength, I was also interested in finding differences in similarly recorded crimes such as business property and personal property theft;

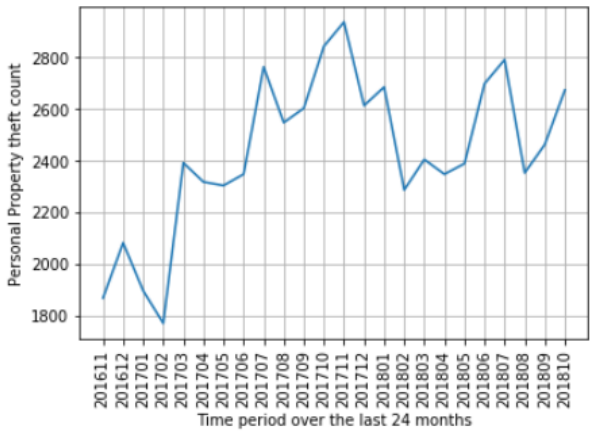








With all the crime count graphs, they are clear enough to compare them side by side to the police strength graph.



The problem with recorded crime data is that if the crime rate increases it could be that more crimes are being recorded and likewise if the crime rate decreases it may be that less people are reporting crimes or even less police are recording crimes. I believe that the crime count should not be used to form any correlations, however it can provide a rough indicator. The reduction of total police force strength does not necessarily correlate into an increase in criminal activity, other factors such as population density, poverty, location and many other factors which can complicate the matter further all could contribute to an increase in criminal activity.

The line graphs could be further improved by combining two-line graphs into one, for example I could have had a property theft graph with two lines representing personal and business property. However, the problem is that the range for personal property is roughly 10 times higher than the business property, the business property line count would be skewed, you wouldn't be able to see any pattern.

To conclude, I had problems with the axis labels, when importing the data, I could not change the date format from "201611" to 11 2016, I think it would have been best to split the graphs into different years either as a bucket or a sperate graph for each year, possibly to reduce the cognitive load. The axis labels are currently too overcrowded.