IMAT5322 BIG DATA ANALYTICS

ASSESSMENT 1

POLICE OPEN-SOURCE DATA

HOW THE CRIME HAS CHANGED WITH COVID-19

Done By

Kaushik Chandrasekhar

P2656623

**Contents**

1.Abstract…………………………………………………………………………………………………………………………….3

2.Introduction of Datasets……………………………………………………………………………………………………4

3.Exploring Datasets……………………………………………………………………………………………………………..5

4.Cleaning the datasets………………………………………………………………………………………………………..7

5.Preparing the datasets for analysis…….………………………………………………………………………………8

6.Plotting Graph using matplotlib……………………………………………………………………………………….10

7.Plotting Graph with Pixiedust…………………………………………………………………………………………..12

8.Map View of crime…………………………………………………………………………………………………………..18

9.Conclusion……………………………………………………………………………………………………………………….20

10.Future Works…………………………………………………………………………………………………………………21

11.Reference………………………………………………………………………………………………………………………22

1. ABSTRACT

This is a report based on police data set from 2019 and 2020 which are compared to analyse the occurrence of crime in the COVID19 season. Initially the datasets are introduced, cleansed and then the analysis is done using various libraries such as matplotlib and pixiedust. Using matplotlib 3 graphs are plotted with total crime and months. First on depicting the graph of 2019, the second representing 2020 and the third depicting both the years side by side. Pixiedust is used to depict Crime type vs Count. Pixiedust is also used to plot geographical graphs. Graphs are used to analyse the difference in crime from 2019 to 2020. At the end both 2019 and 2020 datasets are combined to compare the data side by side and get proper analysis out of the data.

2.INTRODUCTION OF DATASET

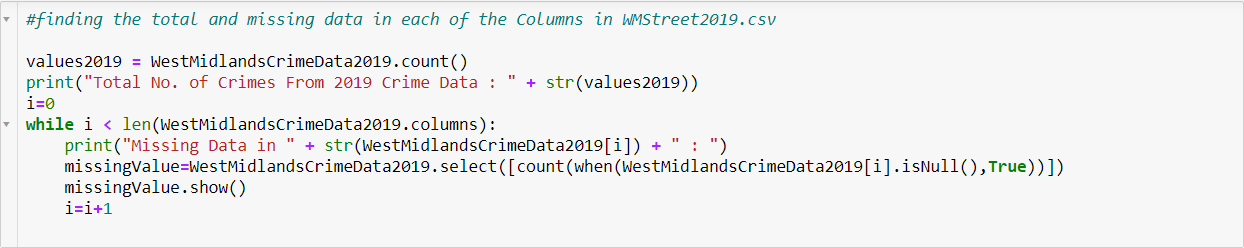
Three Datasets are used for this assessment which are crime data from West Midlands Police for the year 2019 and 2020 and a dataset that is made from combining 2019 and 2020 datasets. All of them are CSV files have the following variables:

1. Crime ID – Unique ID for each crime
2. Month – Month in which the crime is reported
3. Reported By – Person responsible for reporting the crime
4. Falls Within – Jurisdiction of the crime
5. Longitude – Longitude of the crime
6. Latitude – Latitude of the crime
7. Location – Location of the crime
8. LSOA code – LSOA Code
9. LSOA Name – Name of the LSOA
10. Crime Type – Type of crime
11. Last Outcome Category – Outcome of the crime
12. Context

3.EXPLORING THE DATASETS

The 2019 Dataset has a total of 29618 observations. These crimes are recorded from the

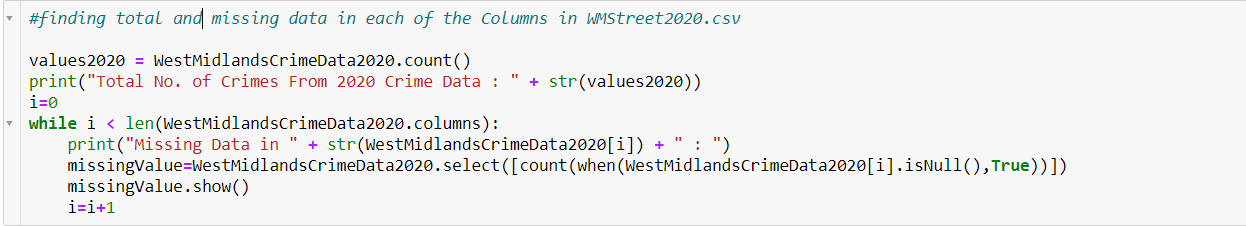
month of January to December in 2019



The missing data from this file are as follows:

|  |  |
| --- | --- |
| Crime ID | 36489 |
| Month | 1 |
| Reported By | 1 |
| Falls Within | 1 |
| Longitude | 12 |
| Latitude | 12 |
| Location | 1 |
| LSOA code | 4 |
| LSOA name | 4 |
| Crime Type | 1 |
| Last outcome category | 36490 |
| Context | 296173 |

While the 2020 Dataset has a total of 322809 observations. These crimes are recorded from the month of January to December in 2020



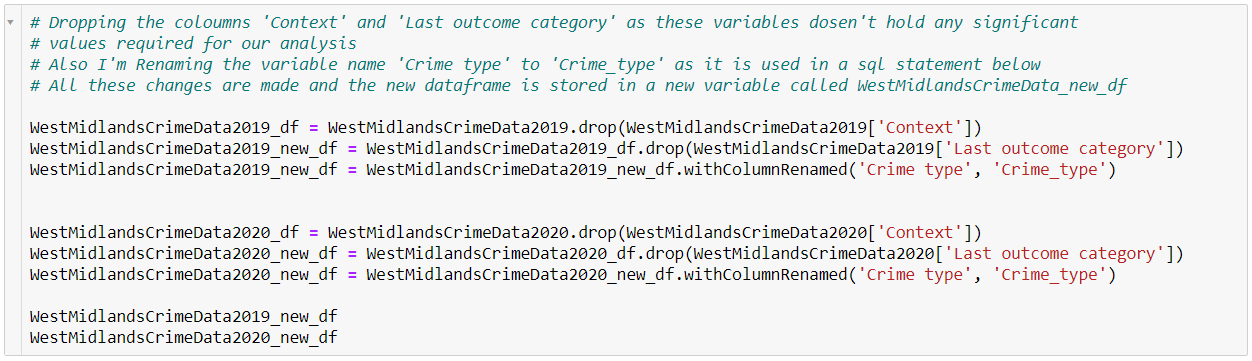
The Missing data from Dataset 2020 is as follows:

|  |  |
| --- | --- |
| Crime ID | 54704 |
| Month | 1 |
| Reported By | 1 |
| Falls Within | 1 |
| Longitude | 12 |
| Latitude | 12 |
| Location | 1 |
| LSOA code | 11 |
| LSOA name | 11 |
| Crime Type | 1 |
| Last outcome category | 54705 |
| Context | 322798 |

4.CLEANING THE DATA

Most of the missing values is in the variable context followed by Last outcome category and crime ID. Because Crime ID is a unique identifier for each crime, I’m keeping the variable. Since there are Lots of missing data and these two variables does not help us in performing any analysis I am dropping ‘Context’ and ‘Last outcome category’.

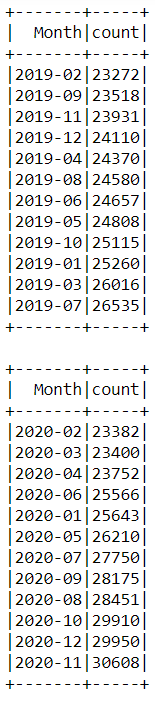
I am also renaming the “Crime type” to “Crime\_type” as this helps me later when performing the actions with SQL queries.

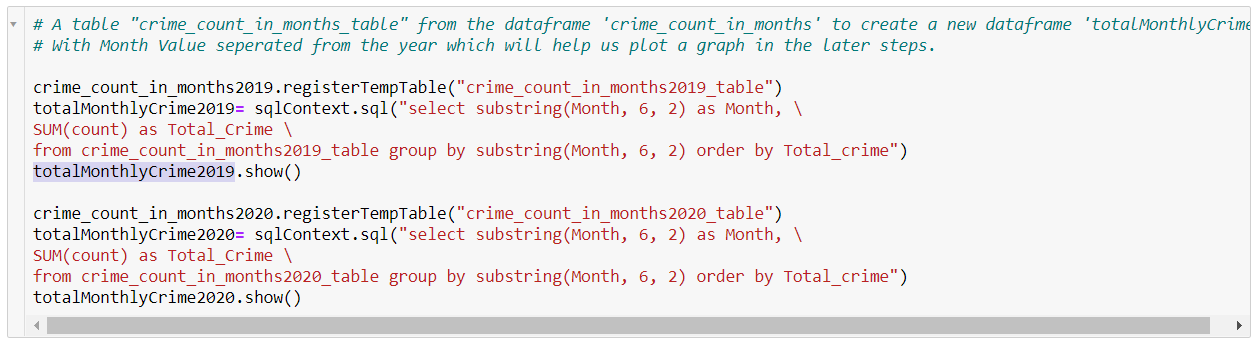
The above-mentioned changes are made to the data frame and stored in the variables “WestMidlandsCrimeData2019\_new\_df” and “WestMidlandsCrimeData2020\_new\_df” respectively.

5.PREPARING THE DATA FOR ANALYSIS

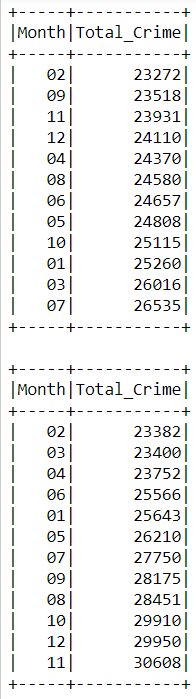
Now I am creating 2 new data frames “crime\_count\_in\_months2019” and “crime\_count\_in\_months2020” from “WestMidlandsCrimeData2019\_new\_df” and “WestMidlandsCrimeData2020\_new\_df” respectively that will later help me create a 2 new data frames “totalMonthlyCrime2019” and “totalMonthlyCrime2020” which will help me plot a graph to compare the 2 Data frames

OUTPUT:

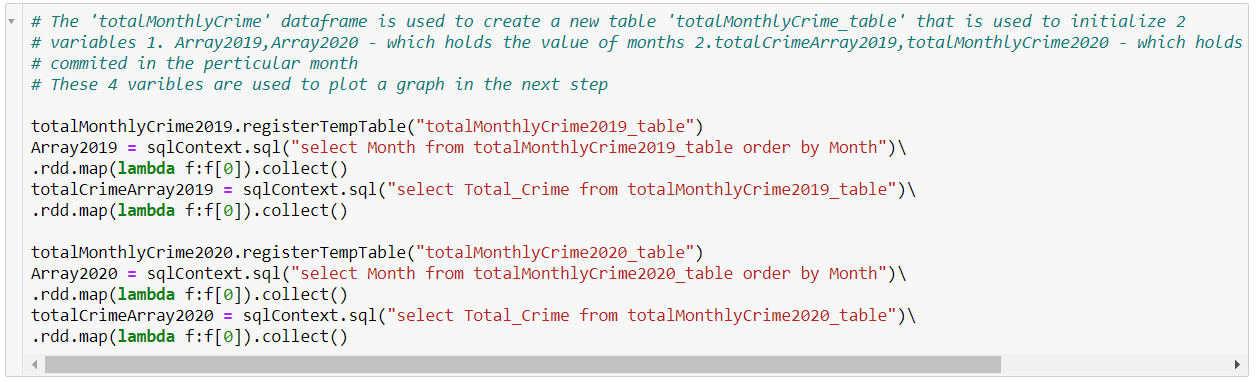




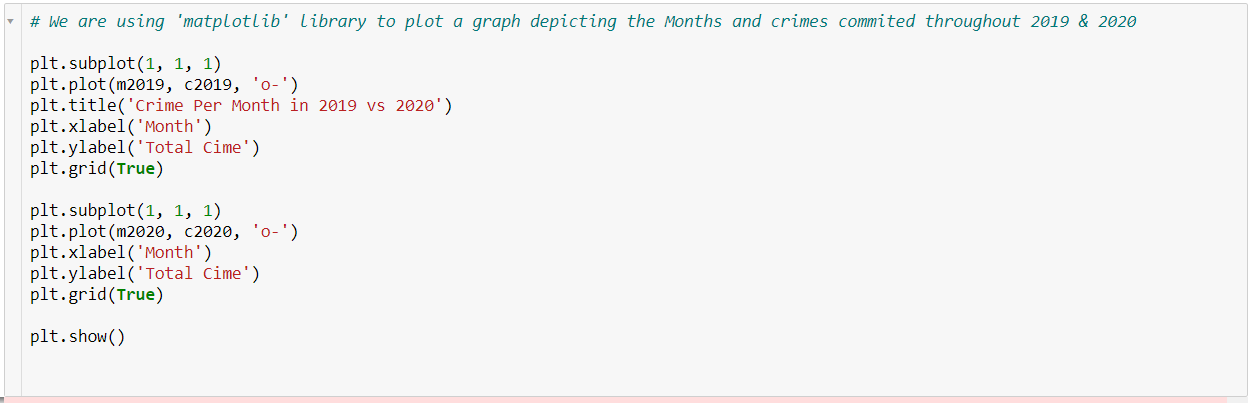
OUTPUT:



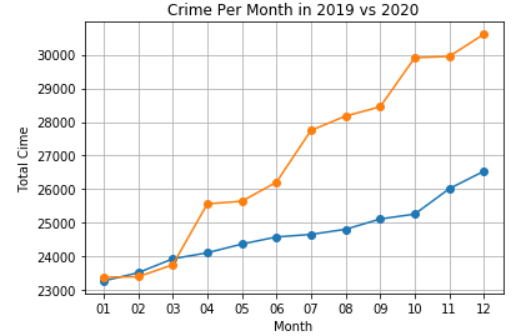
6.PLOTTING A GRAPH USING MATPLOTLIB

After creating “totalMonthlyCrime2019” and “totalMonthlyCrime2020” I am registering a temp table to “totalMonthlyCrime2019\_table” and “totalMonthlyCrime2020\_table” to create rdd map that will help me plot a graph using matplotlib 

I am using matplotlib library to plot the graph of both the years in a single chart so that it will be easy for us to compare them



OUTPUT:



In the above graph the yellow line represents the crime in 2020 while the blue line represents crime in 2019.

From the graph we can clearly see the dip in crime in 2020 in the months of February and March (02 and 03). This is the time the lockdown was getting implemented. But as the lockdown continued over the months the crime has increased significantly compared to 2019. With the increase in lockdown restrictions and people initially being scared of COVID19 have probably restricted themselves indoors, this could be the reason why we can see a dip in crimes in the 2 months when the lockdown has started.

Once people started getting out and was more used to the COVID19 restrictions and limitations and with local authorities in half their strength, it was a mixture of fire and oil. People started committing more crimes as there was little the authorities could do as well.

Thus, we can conclude that with the local authorities working with half their force and with the COVID 19 restrictions being implemented the crime rate in the West Midlands has skyrocketed compared to 2019. In our further analysis we can look at what type of crime contributed to such an increase in crime in 2020

7.PLOTTING GRAPH WITH PIXIEDUST

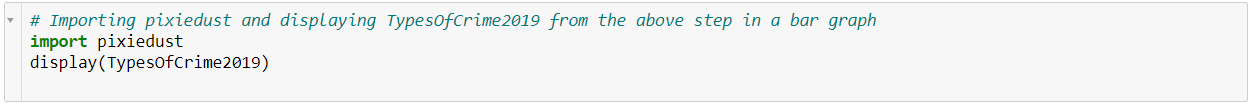
PIXIEDUST is a opensource Python library that helps in improving the user experience when working with data.

Initially I’m using the registertemptable command to register 2 temporary tables "WM\_CrimeTypes2019\_table" and "WM\_CrimeTypes2020\_table" from “WestMidlandsCrimeData2019\_new\_df” and “WestMidlandsCrimeData2020\_new\_df” respectively.

From the tables created above I’m creating 2 new data frames “TypesOfCrime2019” and “TypesOfCrime2020” with the variables Crime\_type, Month, Year, and Count. These 2 data frames will help us display a graph in pixiedust

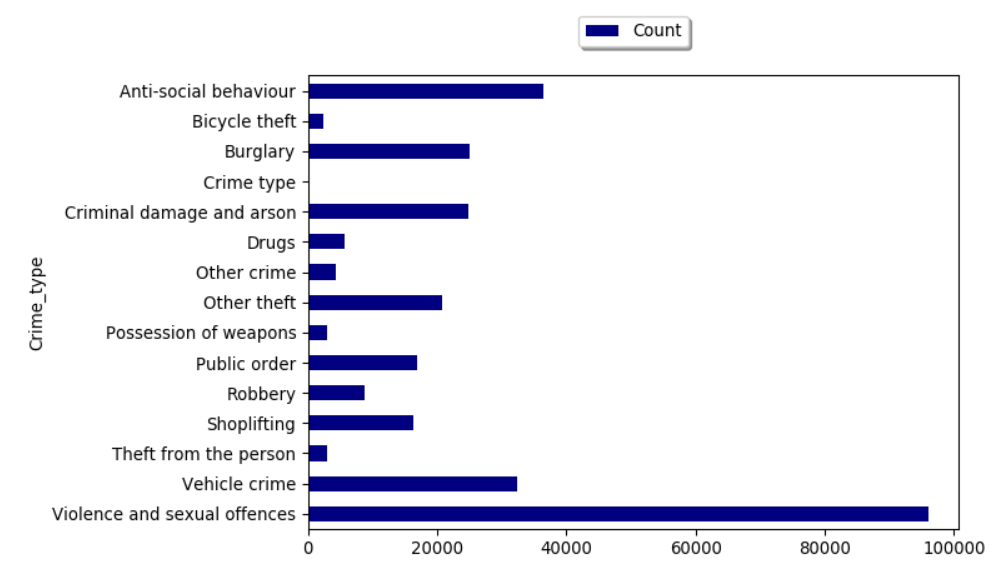


A value “Month\_S” is introduced as the month variable is stored in the following format: “2020-01”. To separate the month from the year I’m creating a separate variable “Month\_S” and dropping the month variable and them Renaming “Month\_S” as “Month”.



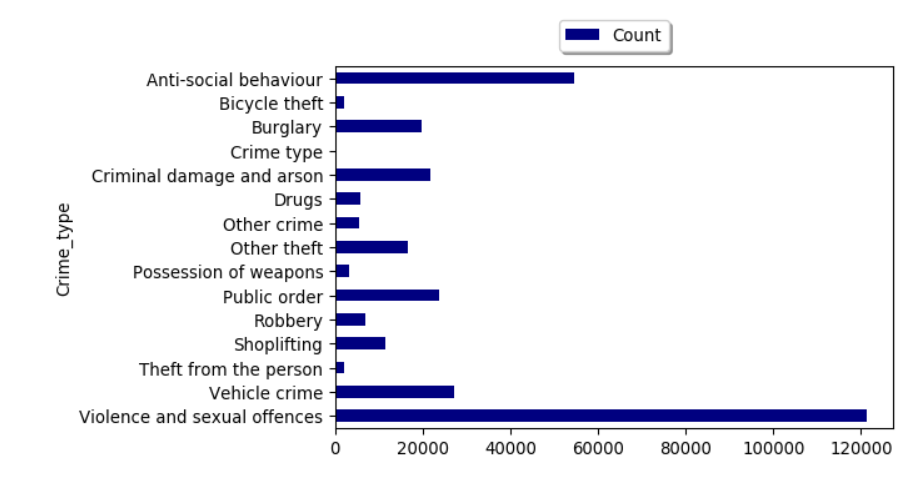
In the above code I’m importing the pixiedust library and using it’s display function to display a Bar Graph that displays the data Types of crimes committed in 2019 vs the Count which can be matched in the pixiedust user interface.

OUTPUT:



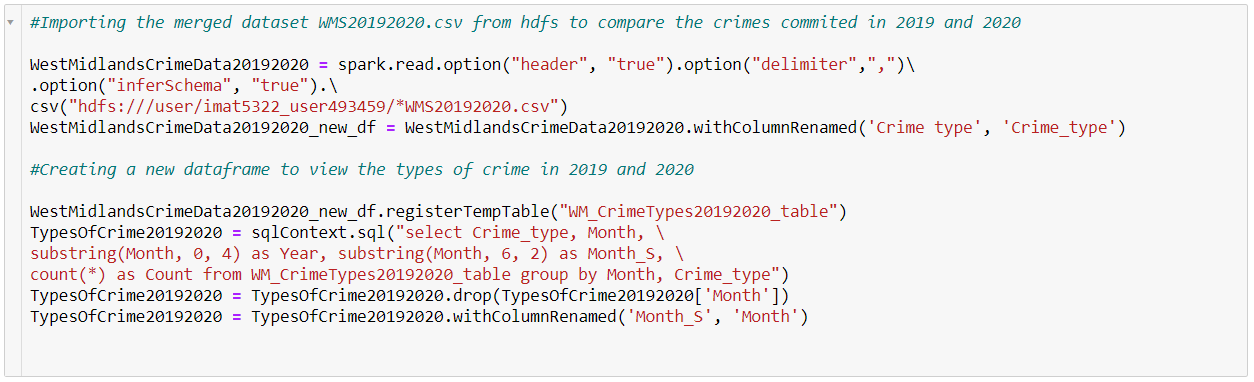
From the above graph we can understand that most crime comes under “Violence and sexual offences” followed by “Anti-social behaviour” then “Vehicle crime” and so on.

Now we display the same graph for 2020

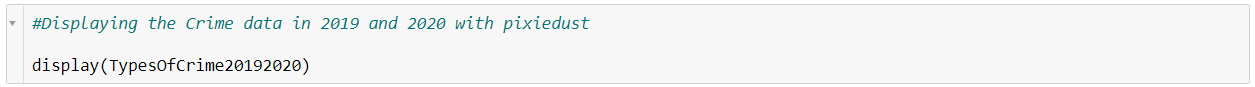
OUTPUT:

From the above graph we get similar observation from 2019 where “Violence and sexual offences” followed by “Anti-social behaviour” then “Vehicle crime” are the most committed crimes in 2020. But compared to 2019 the numbers of “Violence and sexual offences” have grown significantly.

Now we are importing a new dataset “WMS20192020.csv” which is obtained from merging “WMStreet2019.csv” & “WMStreet2020.csv”. This dataset is then used to create a temporary table “WM\_CrimeTypes20192020\_table” from which a new data frame “TypesOfCrime20192020” is derived.

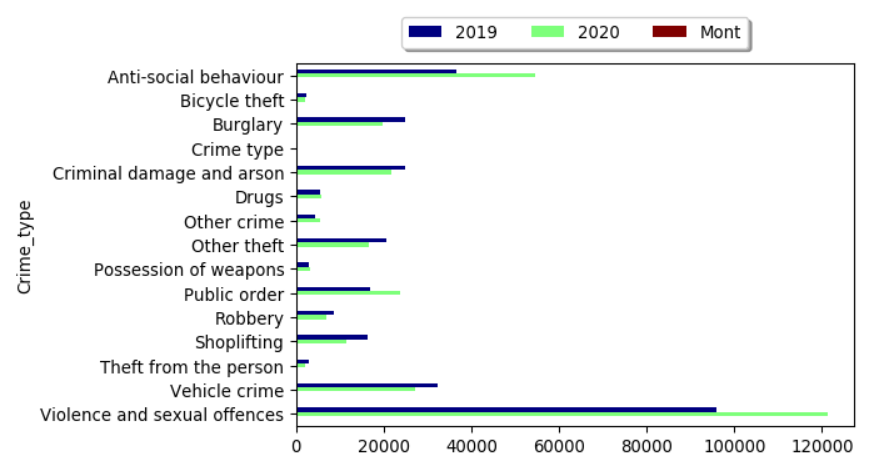


This new data frame is used to create a graph to depict the Crime type and count of 2019 and 2020 side by side.



OUTPUT:

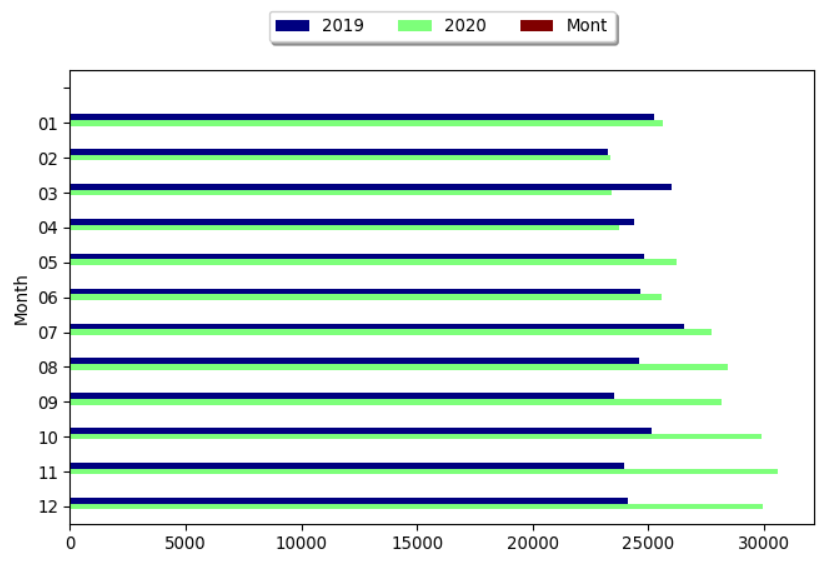
The graph below depicts the crime that occurred in 2019 (in Blue) and 2020 (in Green) respectively.



From the graph we can see that “Violence and sexual offences”, “Anti-social behaviour” & “Public order” crimes have increased significantly in 2020 while “Robbery”, “Burglary”, “Criminal Damage and arson”, “Vehicle Crime” & “Shoplifting” have gone down in 2020.

From this data we can conclude that during the covid period people are more prone towards committing “Violence and sexual offences” & “Anti-social behaviour” crimes. This could be due to the fact that everyone is in their houses and are frustrated with the lockdown. With the lockdown being implemented all other type of crimes that were significant such as “Robbery”, “Burglary”, “Criminal Damage and arson”, “Vehicle Crime” & “Shoplifting” in 2019 have gone down in 2020.

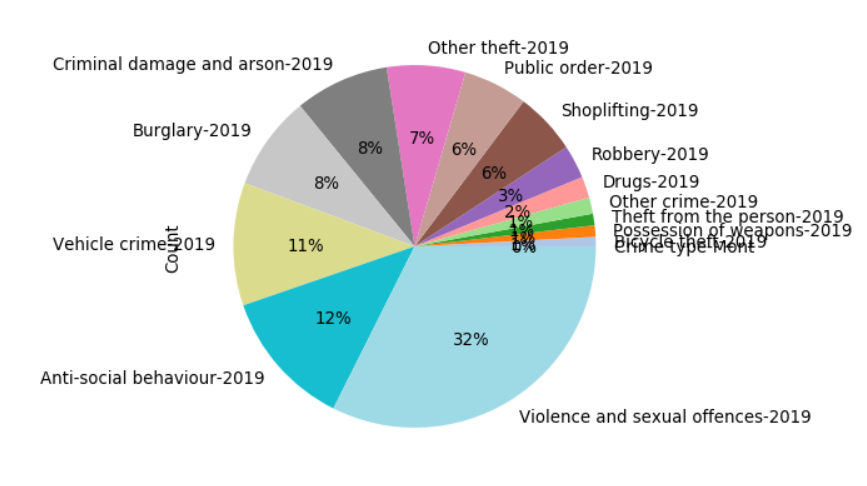
The graph below represents the crimes committed in each moth of 2019(in Blue) and 2020(in Green).



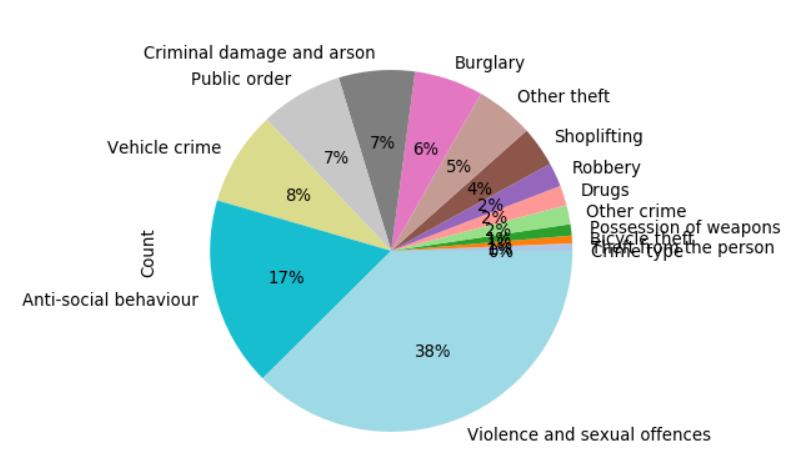
With the starting of covid19 in December 2019 we can see that the crime rate has also been reduced. Through the months from January 2020 to March 2020 the crime has reduced due to the lockdown. As the restrictions were eased on June the crime committed started skyrocketing and on the month of November 2020 it reached its peak.

By comparing the 2 tables above we can conclude that in the star of the COVID19 lockdown the crime has significantly gone down, but the intensity of “Violence and sexual offences” & “Anti-social behaviour” has skyrocketed over the period of lockdown. With the lockdown keeping people isolated has a negative affect on the people’s mental health causing them to behave in an anti-social manner.

The above PIE chart represents the percentage of crimes committed in 2019



The above PIE chart represents the percentage of crimes committed in 2020

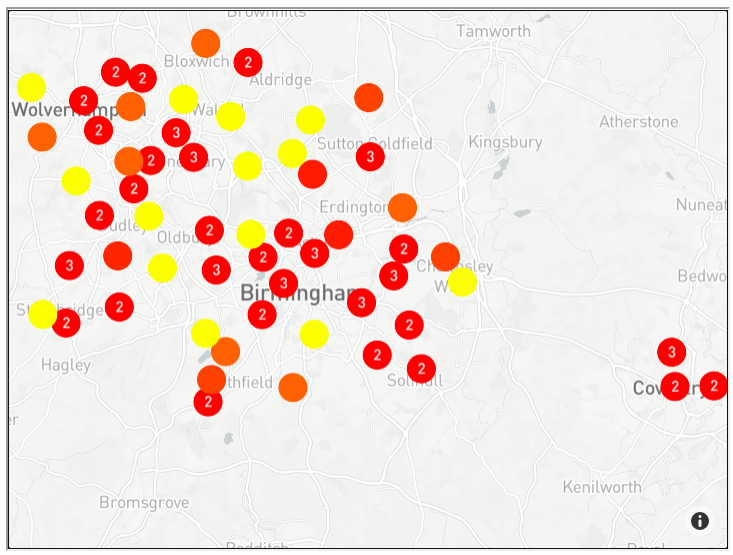


With reference to the above PIE charts from 2019 and 2020 we can clearly see how the types of crime committed in both years have changed. In 2019 the percentage of “Violence and sexual offences” & “Anti-social behaviour” is 32% and 12% respectively compared to 2020 where they have gone up to 38% and 17% respectively.

From the 2 PIE charts and the 2 graphs above we can clearly say that the lockdown has affected the mental health of the people, which in turn has affected the crime rate to skyrocket in the months that followed the lockdown.

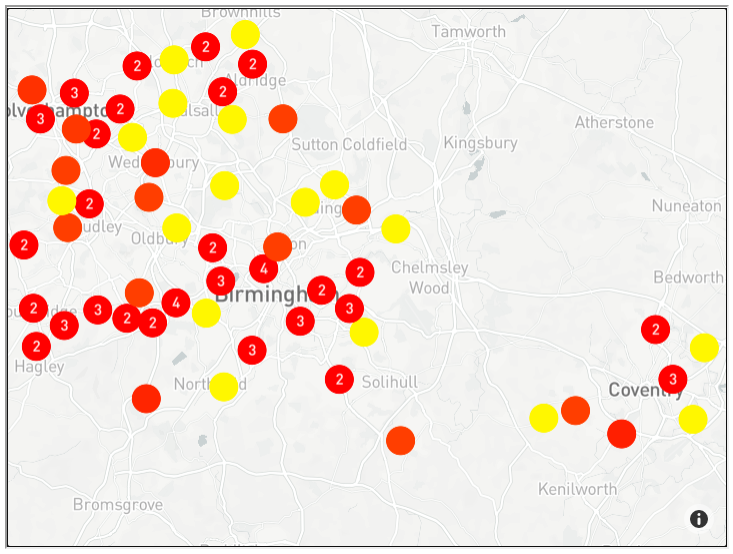
8.MAP VIEW OF THE CRIMES

With the Longitude and Latitude data provided in the police data set we can plot a graph of the data we have using pixiedust, from this data we can get an overview of the locations where the crime is being committed.



The Map represents the crime committed in 2019 under the West Midlands Police Jurisdiction. We can see that lot of the crime is committed in and around Birmingham and Wolverhampton. Some amount of crime has also been recorded in Coventry as well. This could be because Birmingham is the in the middle of England and acts as a connecting point to all other cities.

Also as Birmingham is the 2nd largest city in England the police force may also be more vigilant compared to other cities like Coventry. From this map we can conclude that in places in and around the city is vigilantly watched over by the police.



This map depicts the crime committed in 2020. From the above map we can clearly see the increase in amount of crime in and around Coventry. While there were only few crimes registered in Coventry in 2019 the number has significantly increased in the COVID19 and Post-COVID19 era. This is probably because, with the smaller number of police officials on the job, people are more tempted towards doing an offence.

9.CONCLUSION

After doing all the above analysis we can conclude that during the initial months of the COVID19 pandemic this crime rate as plummeted due to the initial fear of people towards the virus. There has been significant drop in all kinds of crimes that were being committed.

But when people started getting tired of being in their houses for a long period of time during the lockdown it started affecting their mental health were by making them do anti-social activities. We can clearly see in the geographical map that crime in places that reported lower crimes in 2019 has considerably gone up in 2020 where the crime peaking out to 30000 by November 2020.

With the police force working in half their strength the crime rate has gone up significantly during the COVID19 pandemic. With the vaccines on the way let’s hope the crimes are kept under check by 2021.

10.FUTURE WORKS

With the help of all the resources I was only able to pull of the above report. The things I would like to improve in the future would be:

1. Merging the dataset in python (Now I have combined the 2019 and 2020 dataset outside jupyter as CSV and brought them in)
2. Create more cleaner and efficient graphs for analysing data
3. Create a more accurately plotted geographical data

11.REFERENCES

* Data Sets from <https://data.police.uk/data/>
* CSV Merging - <https://www.rondebruin.nl/win/s3/win021.htm>
* Week 6 Module in IMAT5322
* <https://stackoverflow.com/questions/tagged/hadoop>
* <https://discourse.jupyter.org/t/del-drop-column-axis-in-data-frames-python-jupyter-notebook/1791/2>
* <https://pixiedust.github.io/pixiedust/#:~:text=PixieDust%20is%20an%20open%20source,is%20hosted%20on%20the%20cloud>.
* <https://www.w3schools.com/python/matplotlib_pyplot.asp>
* https://datacarpentry.org/python-ecology-lesson/05-merging-data/