

COVID-19, Crime and Effects: Harlow

Data Validity: Can Official Nacional Statistics be trusted?

Author: Experience Omozee,
University of East Anglia, CMP

Introduction

I wanted to find out the validity of the ONS report on crime levels nationally dropping due to COVID-19 this posed a very interesting question that I wanted to explore which was if crime had been perceived to have lowered during covid lockdowns why is it that crime had stayed the same? Applying EDA (Exploratory data Analysis) I was set out on finding out whether crime had fallen locally, I chose Essex Harlow since it is the town I'm from which has a relatively low population and compared it to Great Yarmouth since they both have relatively the same population size .

Methods

When completing my research I had to use the data science pipeline to get my data, clean it and then explore it and come up with a visualisation presenting the levels of crime in Harlow and finding out if the crime levels stayed the same between the years.

Data Collection

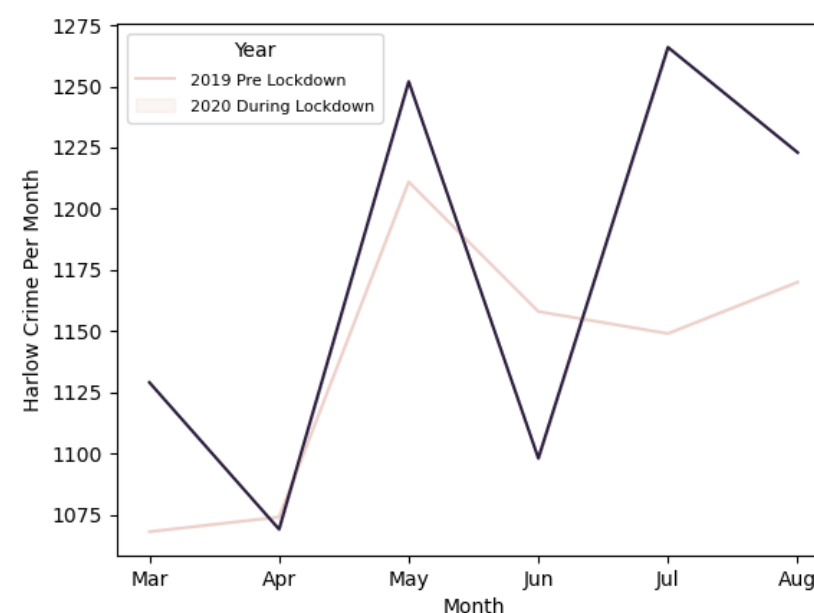
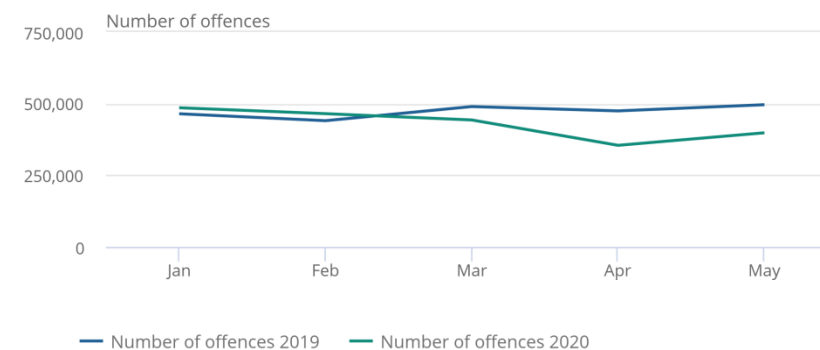
To collect my data I went to the open data webpage. To extract the data that I needed I used a library called remote zip because I was searching through the Open data archives and downloading them would have been space inefficient it allowed me to access the URL of the zip that I needed. I could have used a csv library but I didn't because it was far more in space inefficient that's why I used the remote zip library. I took the data for

Data Wrangling

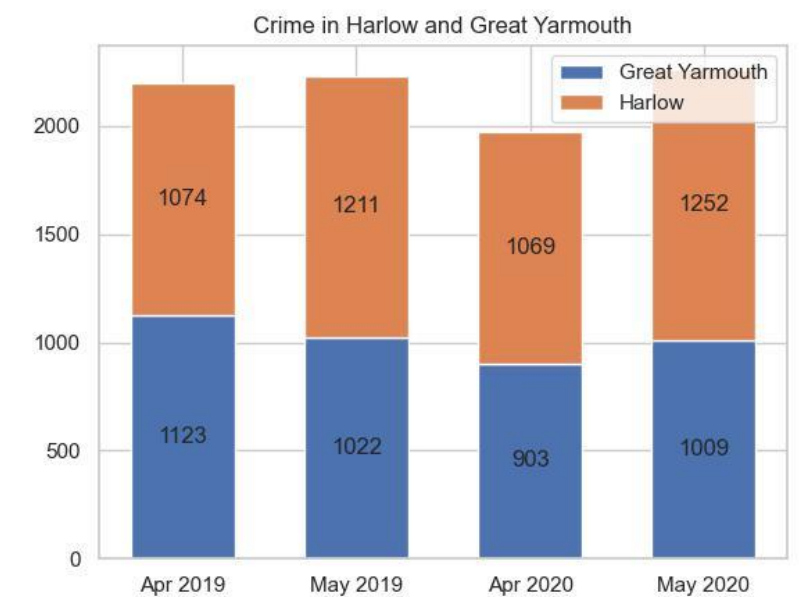
After I collected my data I had to clean it because it had invalid and missing values that would have corrupted the data. I used the 'fillna' method and replaced the missing values within the 'LSAO name' section, I did this because I would then need to search within that column to find the result 'Harlow'. Once I found 'Harlow' results I used the 'dropna' method in pandas to get rid off the unnecessary columns. I could have used the 'fillna' method and replaced the null values with the modal type but that would have altered my search results, that's why I turned the null values to 0 instead.

Exploratory Data Analysis

Before I plotted the data I had to add an additional column which would represent the year to my data set using 'insert method in pandas. I used this method because it updated the data set with my new column, I could have used the 'assign' method but this would have returned a new data frame after the addition of the column'. I then used the 'groupby method' I used it because it would allow me to group the amount of crime committed per month between the years and would then allow me to use the 'transform' method to place the results of the grouped data into a new column the per month, this was inefficient because I needed to create a new column for the 'crimes When plotting the data of Harlow's crime numbers per month between 2019 pre lockdown and 2020 during lockdown I used a the seaborn library and utilize a line plot. I used seaborn library since it would collect all of the values in my pandas data frame and since the two have a good compatibility, I did not plot it using matplotlib lib . I utilized the line plot since it is has the best type of visualisation when displaying change over time which was important. I could have used a bar graph but since the numbers are not so high the comparison wouldn't be as impactful when showing whether crime stayed the same or rose, that's why I chose the line plot.



Before I plotted the data I had to perform the same data collection methods and wrangling to get the data for the Great Yarmouth data set since I was going to compare the two towns to see their crime levels during the lock down and prior to it. I used a matplotlib lib stacked bar chart because it would have shown the crime levels per month within the April and may months where lockdown was at its most secure.



I had to collect the data for the amount of crime that occurred in May and April I used the pivot method on my data set to get the exact values for the different amounts of crime for each crime type and the year it happened it. I utilised the matplotlib lib library to make the grouped bar chart because would be good when showing the comparison between the two years and month ranges. I chose a grouped bar chart to show the comparisons of the crimes that were committed in the lock down period and in the previous year.

