



Shahid Beheshti University

Anita Soroush

Crime reports

A bitter truth, crimes do exist. Collecting data about crimes and analyzing them may lead us to their underlying cause. *FBI NIBRS* is a rich database that captures details on each single crime incident.

Download:

- From the Louisville government's [website](#) download five consecutive yearly datasets.

After concatenation of all past five years datasets, the result final dataset consists of 15 columns and around 500,000 rows. The columns are as follow:

- INCIDENT_NUMBER
- DATE_REPORTED
- DATE_OCCURED
- UOR_DESC
- CRIME_TYPE
- NIBRS_CODE
- UCR_HIERARCHY
- ATT_COMP
- LMPD_DIVISION
- LMPD_BEAT
- PREMISE_TYPE
- BLOCK_ADDRESS
- CITY
- ZIP_CODE
- ID

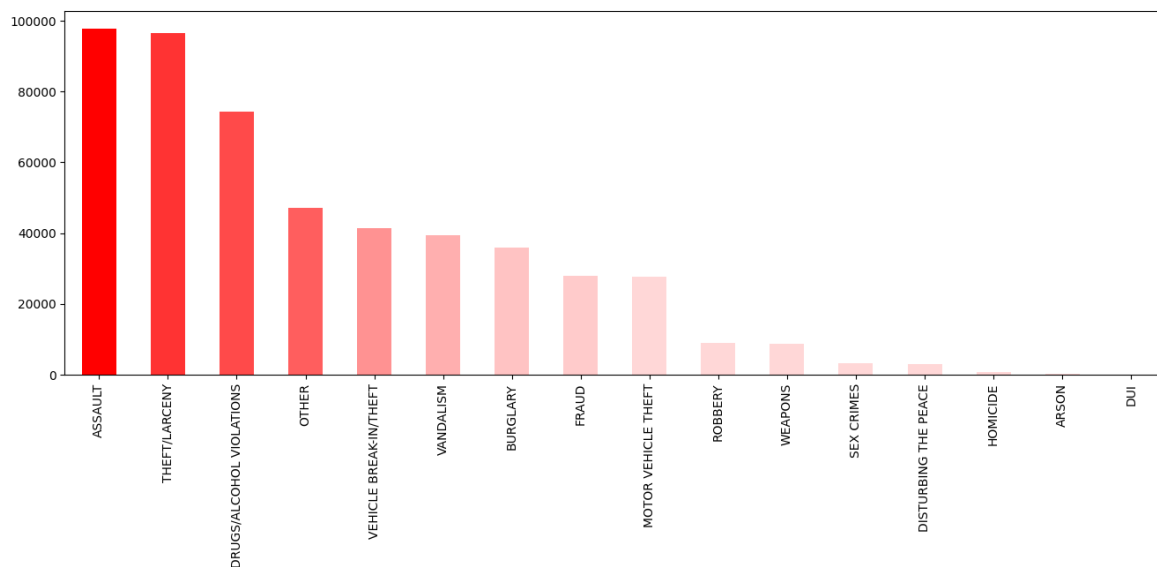
After cleaning the dataset, let's analyze our data:

- Which crimes are most common?

In the first place, we even do not know all types of crimes. So with the use of `value_counts()` function, we can both:

- 1) Make a list of all crime types in our dataset
- 2) Count the repetition of each crime type

The result is shown in the bar chart below:



As we can see in the top plot assault and theft/larceny are most common crimes with nearly 100,000 records for each of them in the last 5 years.

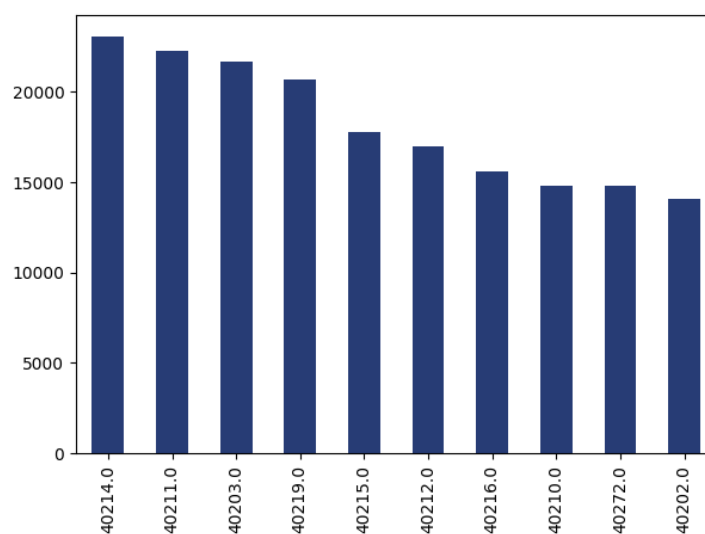
And as in mentioned, now we can make a list of all crime types, based on the x-axis of plot:

- 0) "ASSAULT"
- 1) "THEFT/LARCENY"
- 2) "DRUGS/ALCOHOL VIOLATIONS"
- 3) "OTHER"
- 4) "VEHICLE BREAK-IN/THEFT"
- 5) "VANDALISM"
- 6) "BURGLARY"
- 7) "FRAUD"
- 8) "MOTOR VEHICLE THEFT"
- 9) "ROBBERY"
- 10) "WEAPONS"
- 11) "SEX CRIMES"
- 12) "DISTURBING THE PEAS"
- 13) "HOMICIDE"
- 14) "ARSON"
- 15) "DUI"

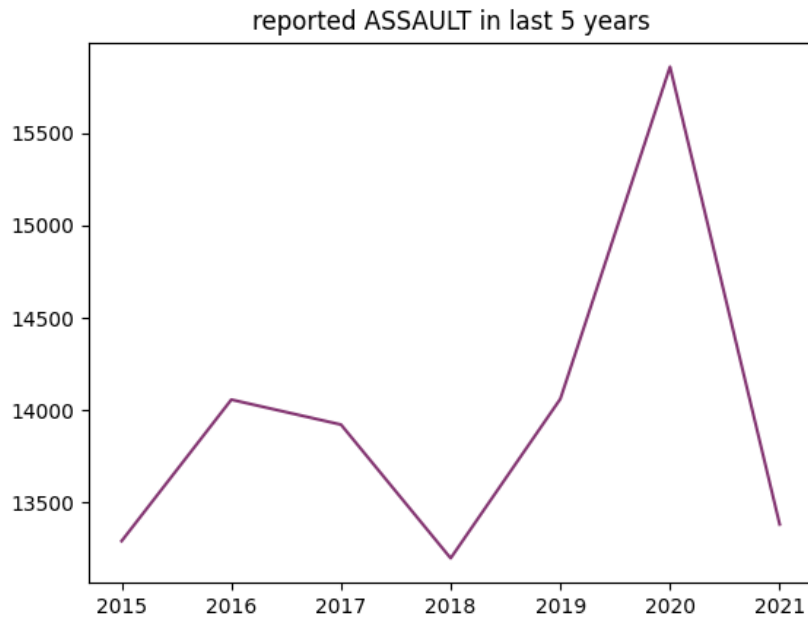
- **In which zip-codes are crimes more likely to occur?**

We just need to count the repetition of each zip code number and sort the zip codes based on it in a descending trend and make a list of them.

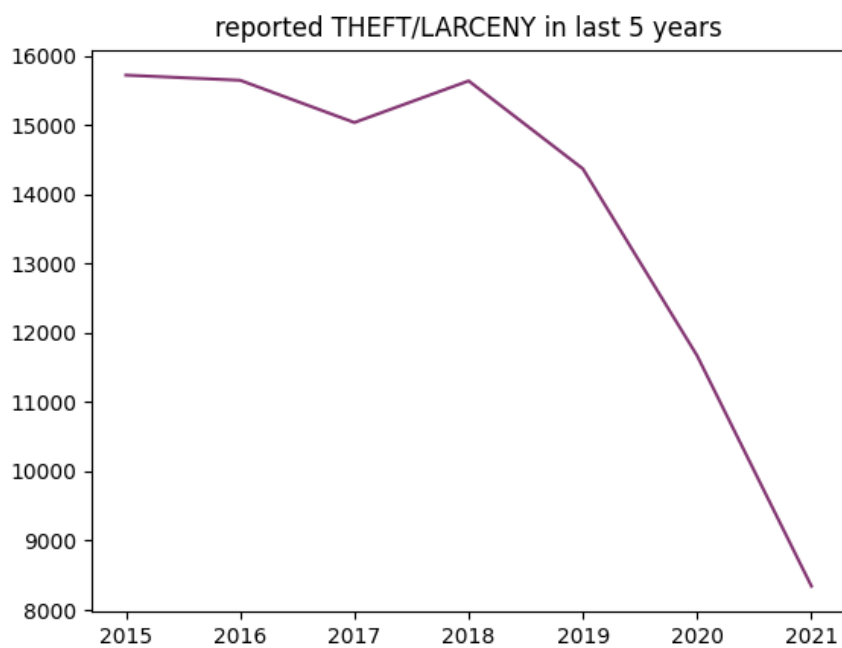
Showing all zip codes list is not necessary so let's just take a look at top 10 of them which are repeated the most:



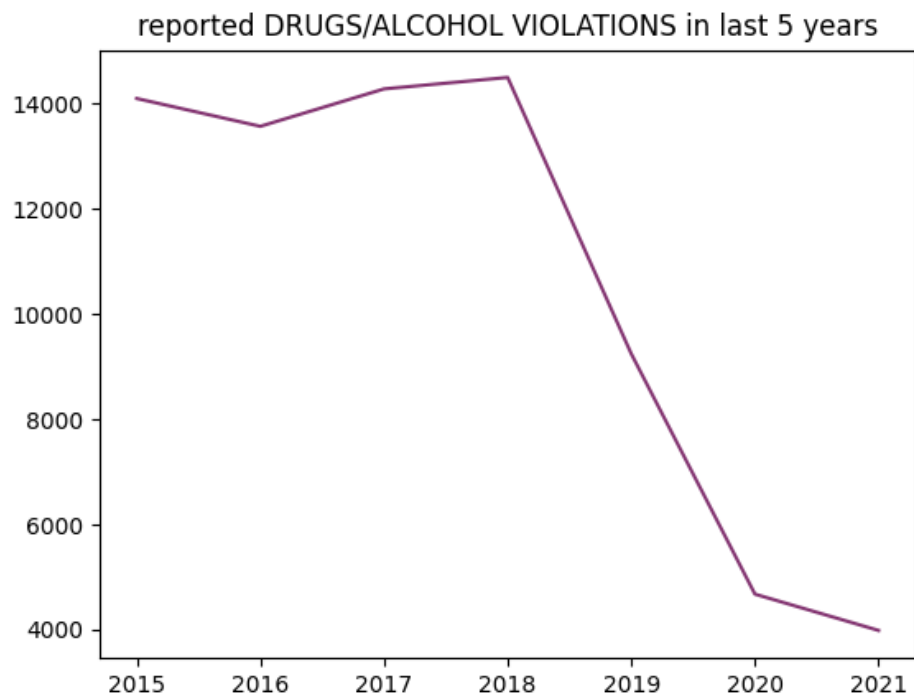
- Is there a trend of some crimes increasing and others decreasing in number over these five years?



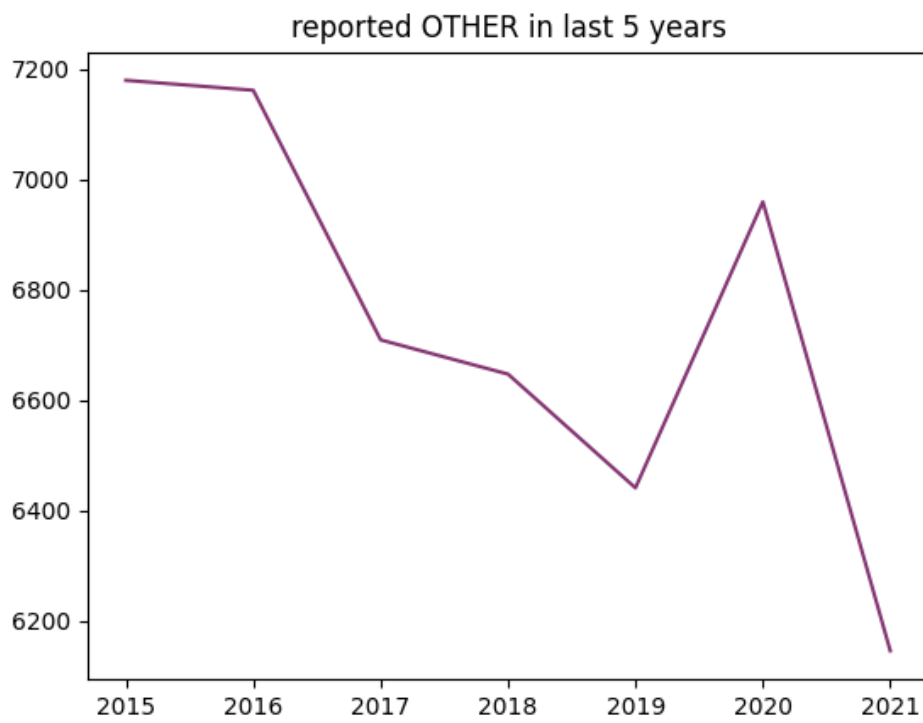
not a meaningful trend



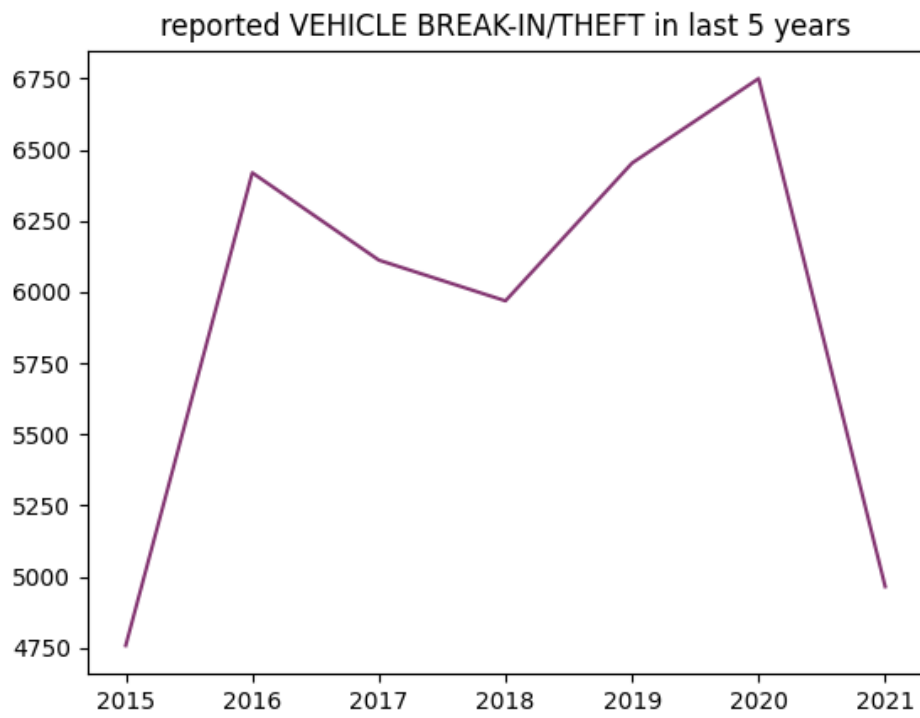
Decreasing trend



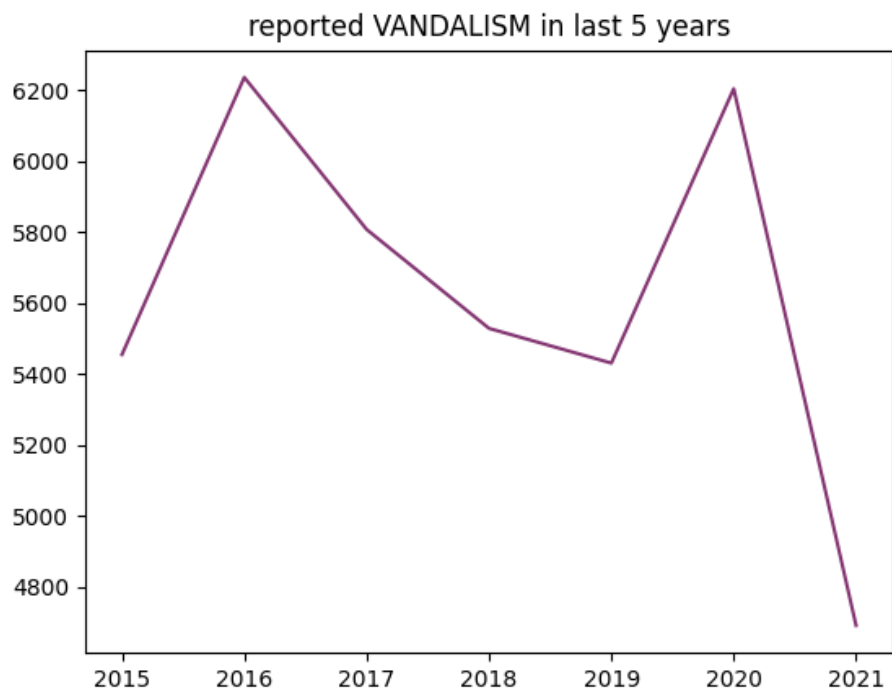
Decreasing trend



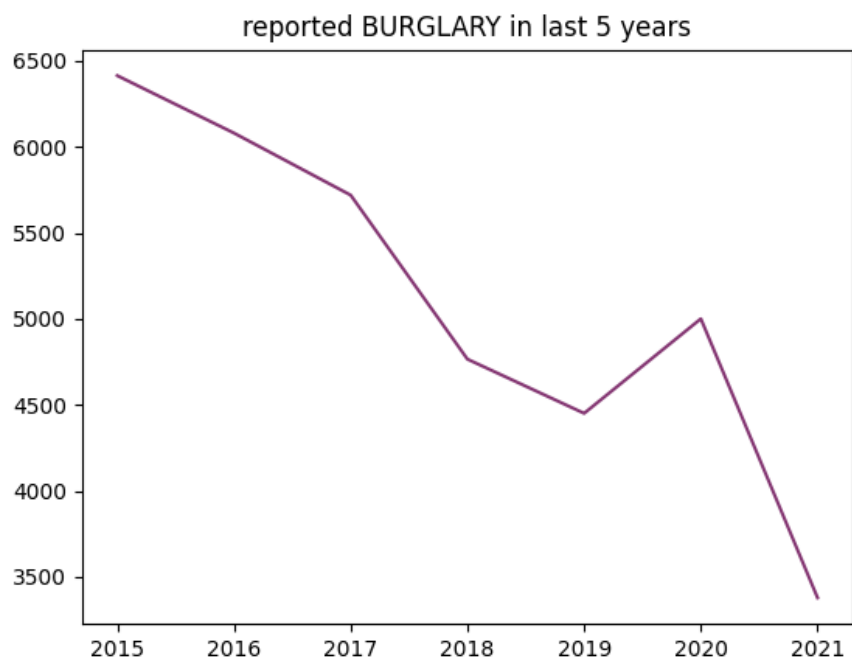
not a meaningful trend



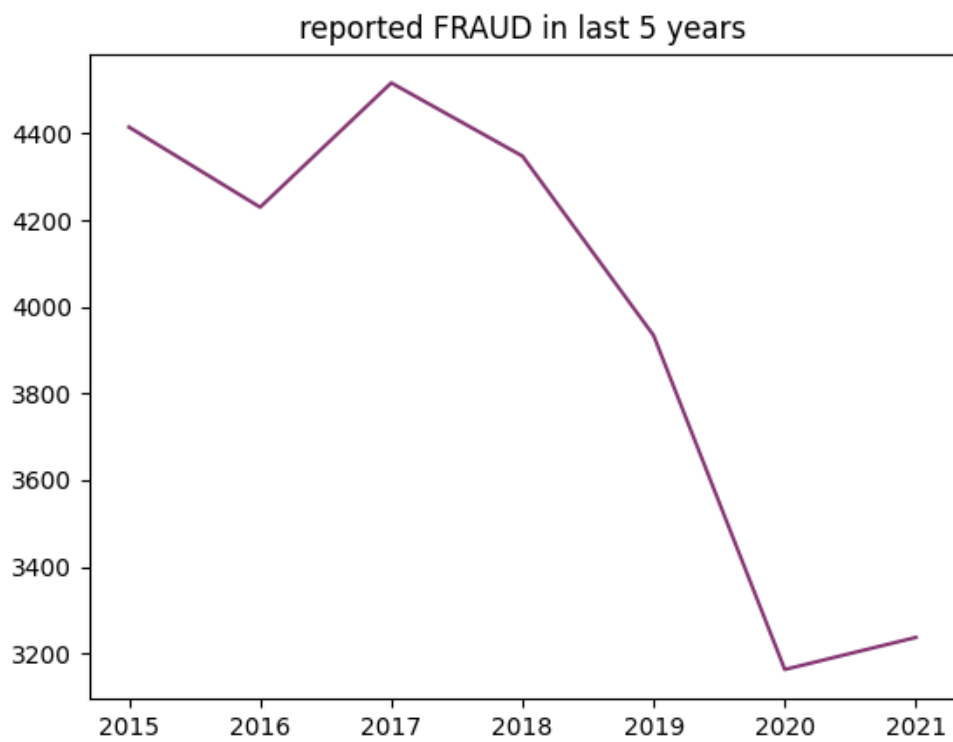
not a meaningful trend



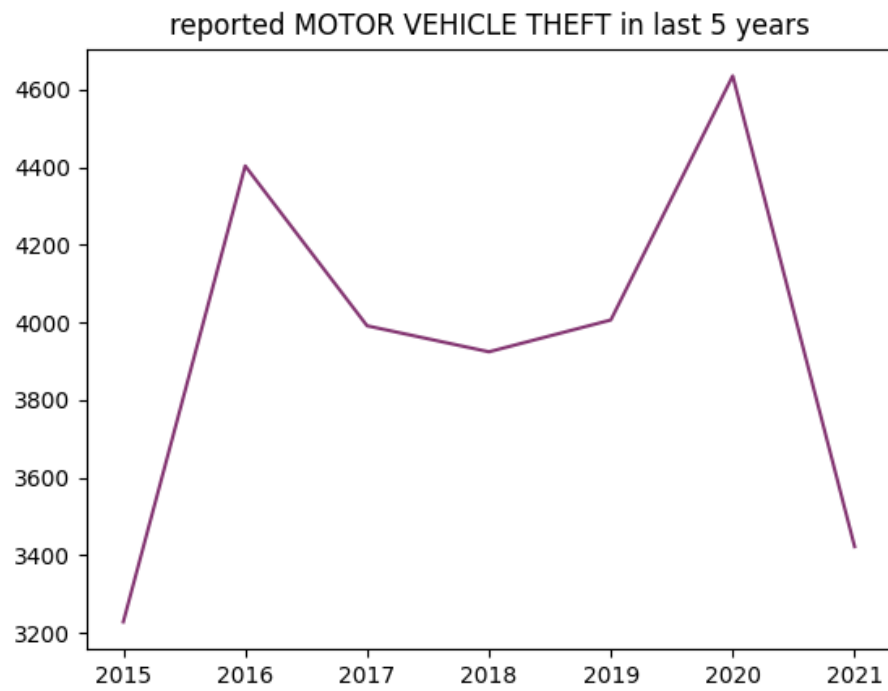
not a meaningful trend



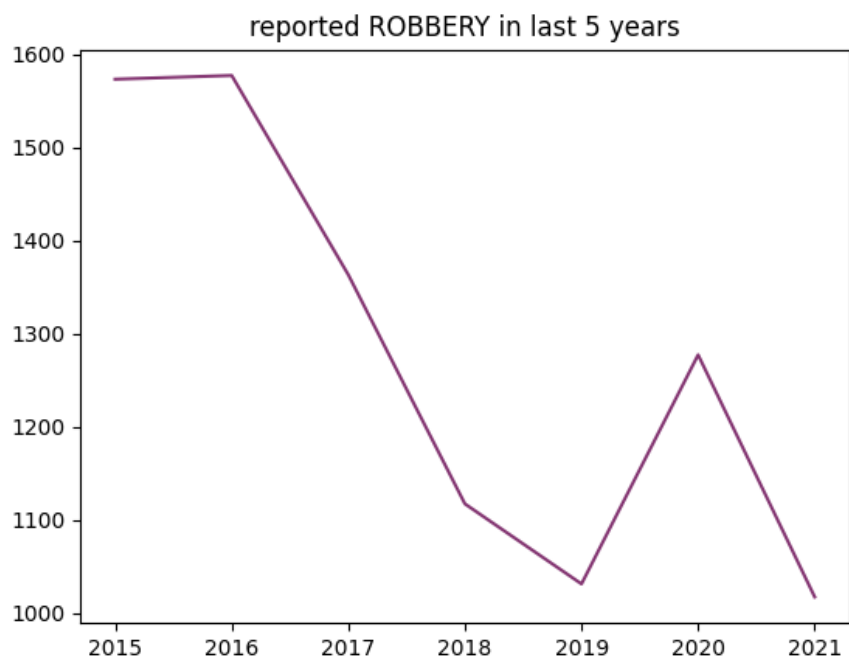
not a meaningful trend



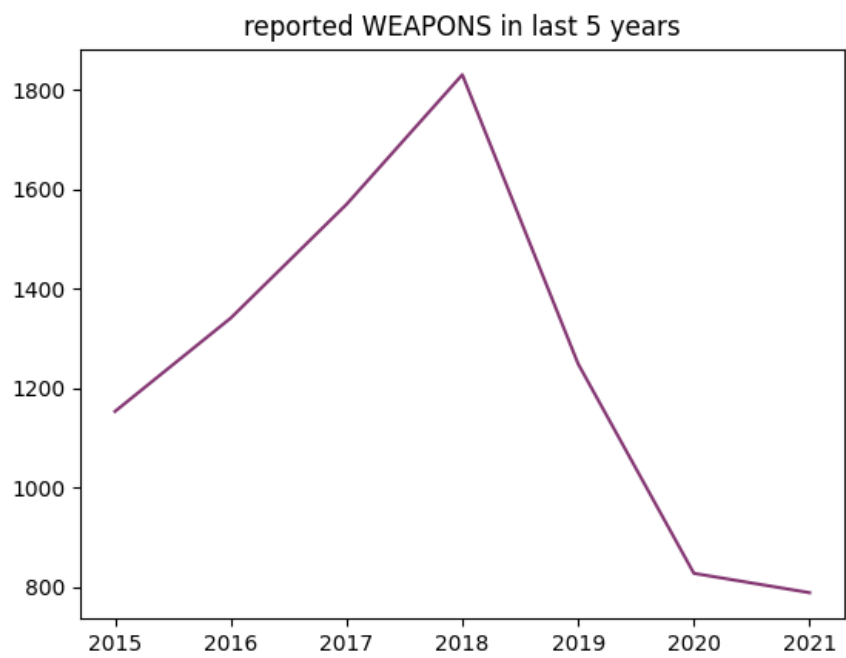
not a meaningful trend



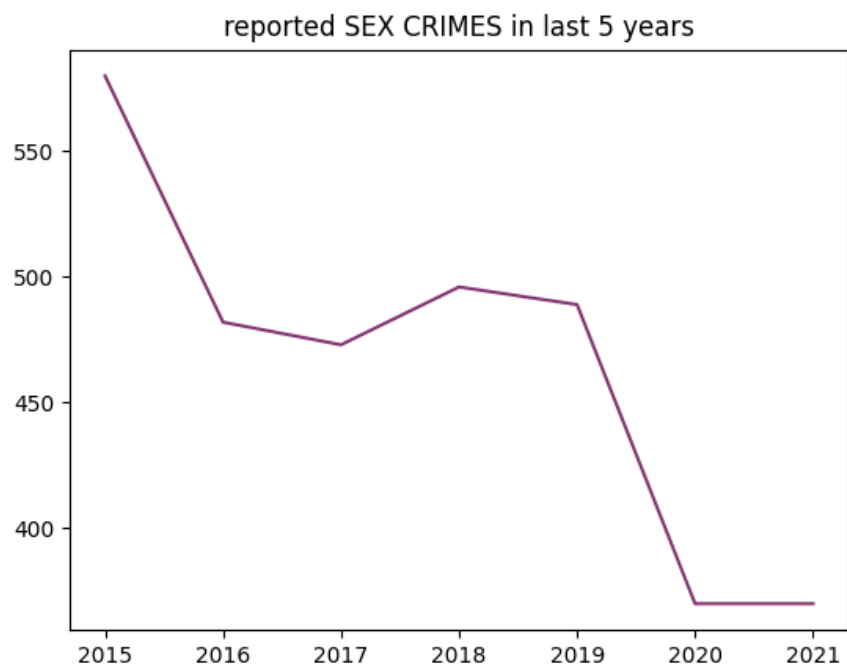
not a meaningful trend



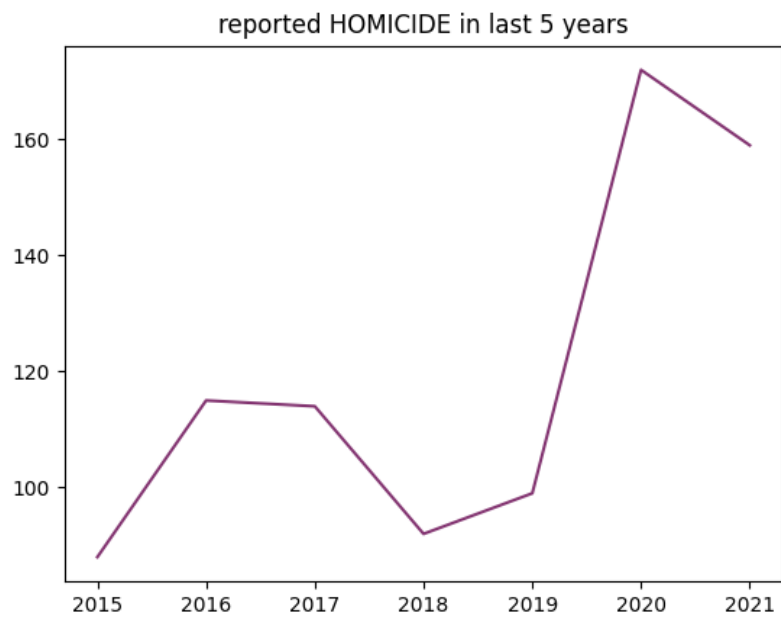
not a meaningful trend



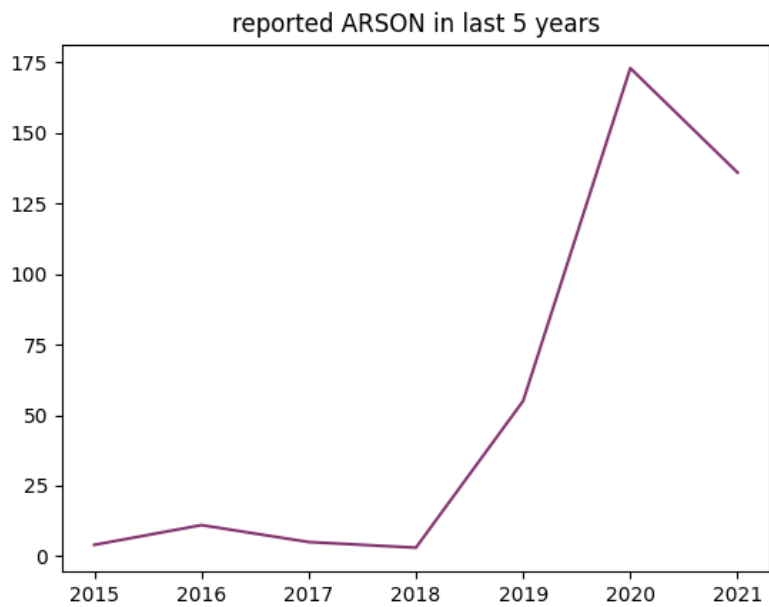
Decreasing trend (after 2018)



Decreasing trend

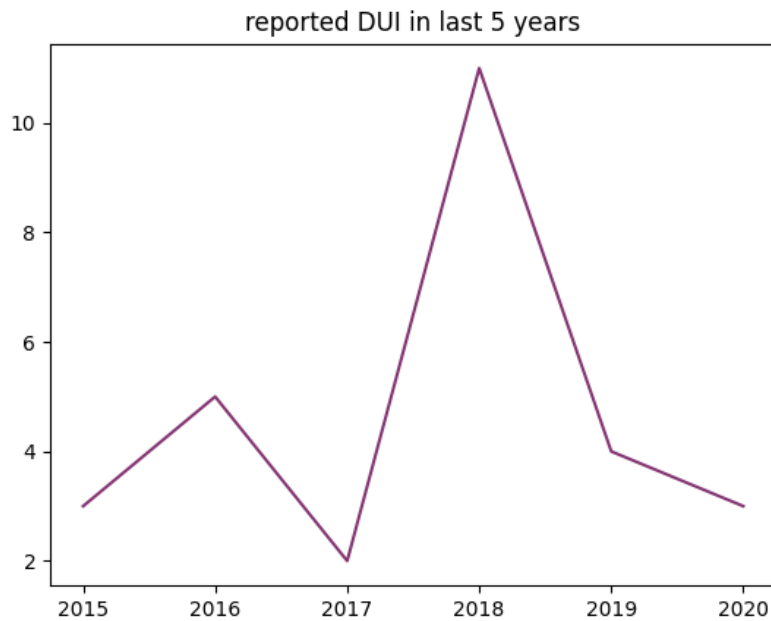


not a meaningful trend



Increasing trend

(as 2021 is not finished yet, I ignored it)



not a meaningful trend

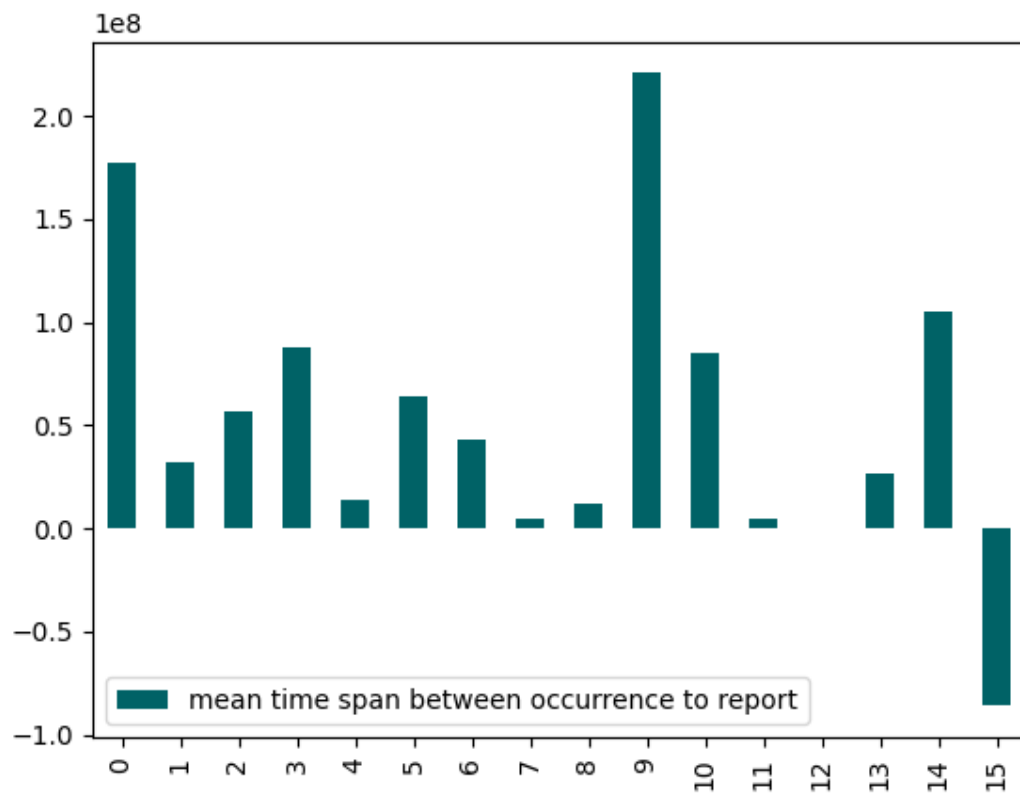
let's move on to the next question:

- Which crimes take the longest to report?

There are two columns of Date in our data set, one of them is the date of occurrence of a crime and the other one is the date of that crime being reported.

Now we can extract a new column out of these two and call it OCCURRENCE_TO_REPORT which contain the time interval between happening and report date of each crime.

The bar chart below shows that, on average, which type of crime took the longest to be reported, in comparison to other types:



If we match the numbers on x-axis with the list of crimes which was mentioned before, we can see that first of all, number 9 (ROBBERY) and after that, number 0 (ASSAULT) took by far the longest time to be reported.

Resources:

https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html

<https://www.geeksforgeeks.org/python-pandas-dataframe-sum/>

https://www.rapidtables.com/web/color/RGB_Color.html

<https://www.geeksforgeeks.org/how-to-merge-multiple-csv-files-into-a-single-pandas-dataframe/>

Assignment Source:

<https://sk7w4tch3r.github.io/CS-SBU-DataScience/chapters/chapter2-3/02-q/>