****

Mini Project Report

*for*

Data Analysis on Homicides in Toronto

using Python

*Submitted by*

CHAMARTY KUMAR RAJA

under the guidance of Tareq J

*In partial fulfillment for the award of the degree of*

***Diploma in Advanced DataScience***

**1. Introduction:**

We have used the Homicide data set from the year 2004 to 2018 to analyze the different variables and answer different questions that have been mentioned later section. The data has been downloaded from <http://data.torontopolice.on.ca/search?collection=Dataset&q=Homicide> which is a public domain. The data consists of basic information such as different types of homicides committed, when they were committed and where they were committed. The data has

**2. Objectives:**

To demonstrate basic data analytical skills in Python using Pandas, Numpy and other necessary packages. It involves, but not limited to the below steps:

* Read the raw data from source, clean it and convert to a structured format (in this scenario, a ‘Data Frame’.
* Analyze the data and gain insight into various attributes that are available.
* Determine or Identify
  + The safest Neighbourhood to live in Toronto.
  + Which Police Division may need more resources

**3. Software and packages used to complete this project :**

* OS: Windows 10
* Interpreter: Python – 3.7
* IDE: Spyder
* Packages:
  + Pandas
  + Numpy
  + GeoPandas
  + Matplotlib
* Tools: Anaconda Navigator

**4. Project Goal:**

The goal of the project is to gain hands on experience with various Data science and Analytics related packages that are available in Python and in the process, demonstrate our skills.

**5. Problem Description:**

We have identified the below Analysis Questions and tried to provide the best possible solutions to them:

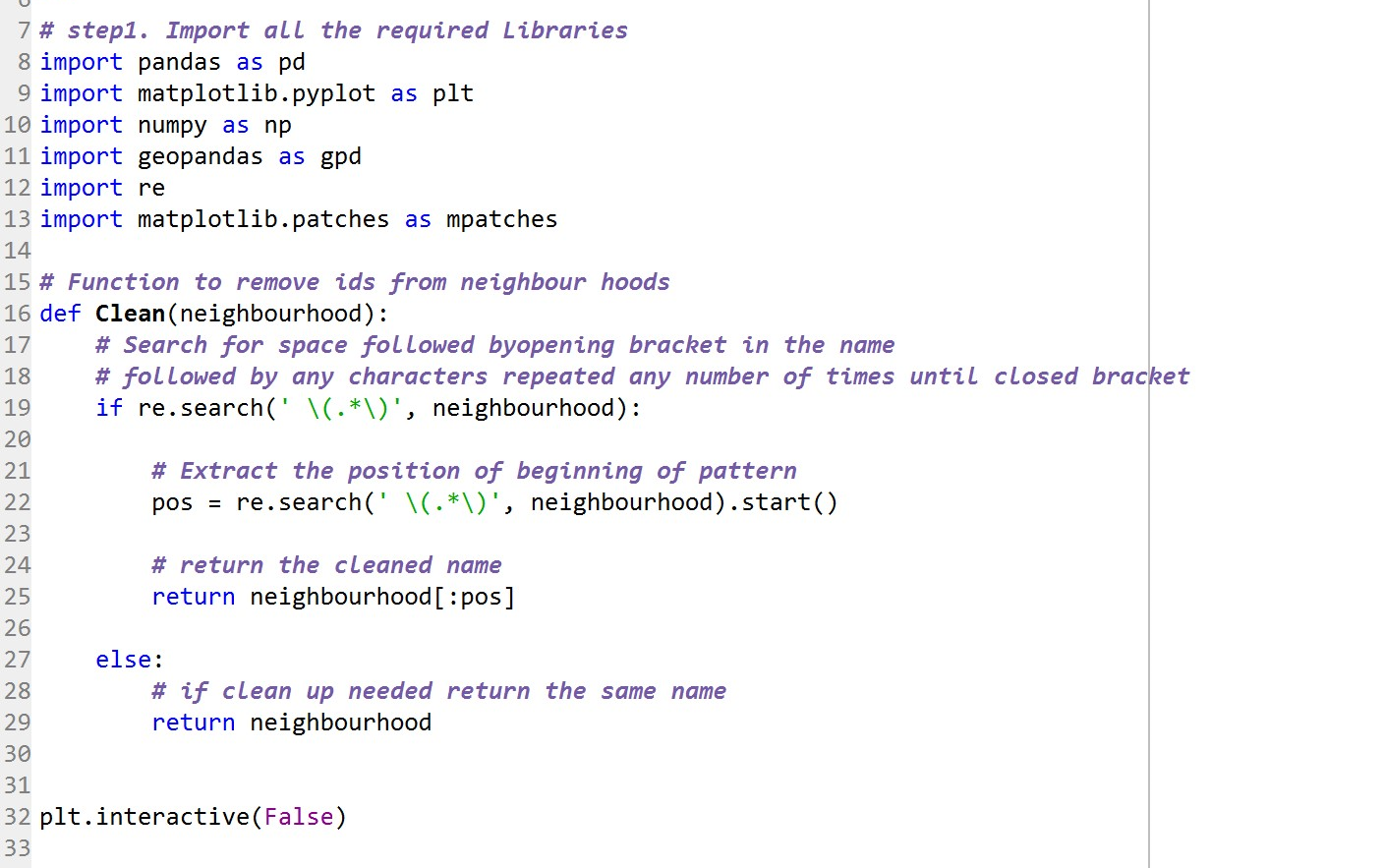
* Q1: What are the total number of Homicides per year from 2004 to 2018? And provide the Visual as a Bar Graph
* Q2: What are the total number of Homicides grouped by Divisions from 2004 to 2018? And provide the Visual as a pie chart and indicate the highest division by separating from the rest.
* Q3: What are the total number of Homicides by Neighborhoods for top two divisions?
* Q4: Is Gun Violence the primary cause of Homicides in Toronto?
* Q5: What are the total number of homicides grouped by year and type?
* Q6: Get the total Homicide counts in months when grouped by Homicide type.
* Q7: Has Gun Violence increased or decreased in 2018 from 2004?
* Q8: Identify the Top 5 safe neighborhoods to live in Toronto using ranking concept.
* Q9: Geo Plotting of Homicides in the year 2004 Per Division
* Q10: Geo Plot of cumulative Homicides Per Division and analyze the trend in each division.

**6. Problem Solution:**

We will provide sample solutions to two of the questions (Q6 and Q9) here and the rest will be provided in the form of python code and a ppt.

**Q6: Get the total Homicide counts in months when grouped by Homicide type.**

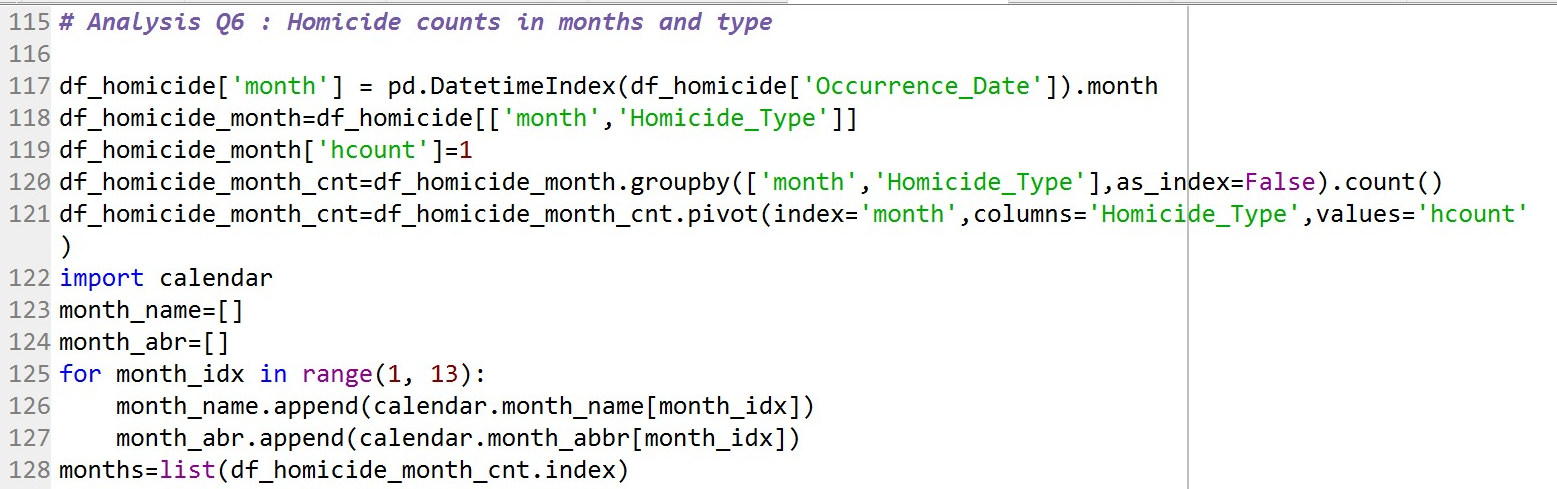
* **Step1:** Import all the necessary libraries and the define the necessary functions:



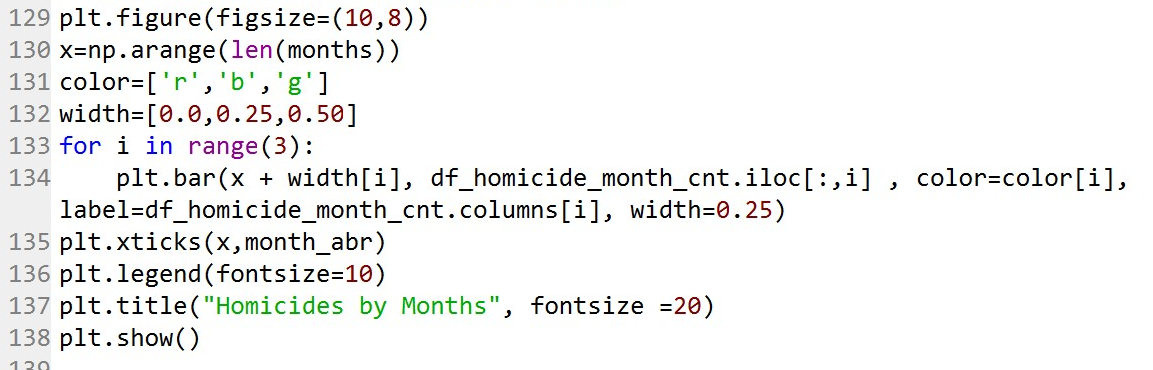
* **Step2:** Read the data from raw files and Prepare the data for Analysis.



* **Step3:** Perform the required Analysis and prepare the data for visual presentation.



* **Step4:** The final step is to present the data in a visual form.



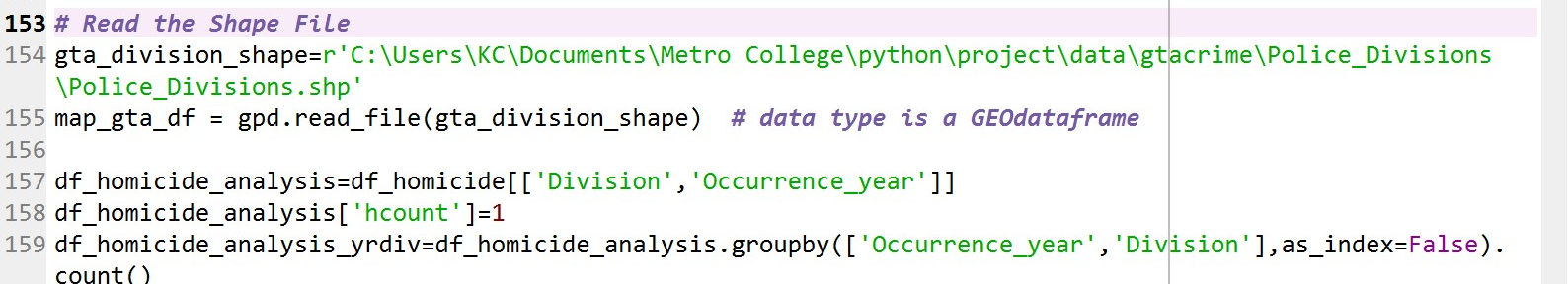
* **Result:**

A screenshot of a cell phone

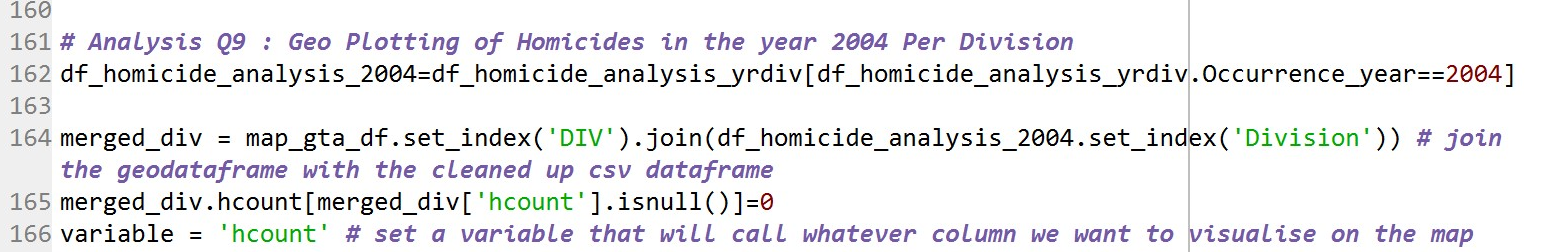
Description automatically generated

**Q10: Geo Plot of cumulative Homicides Per Division and analyze the trend in each division**

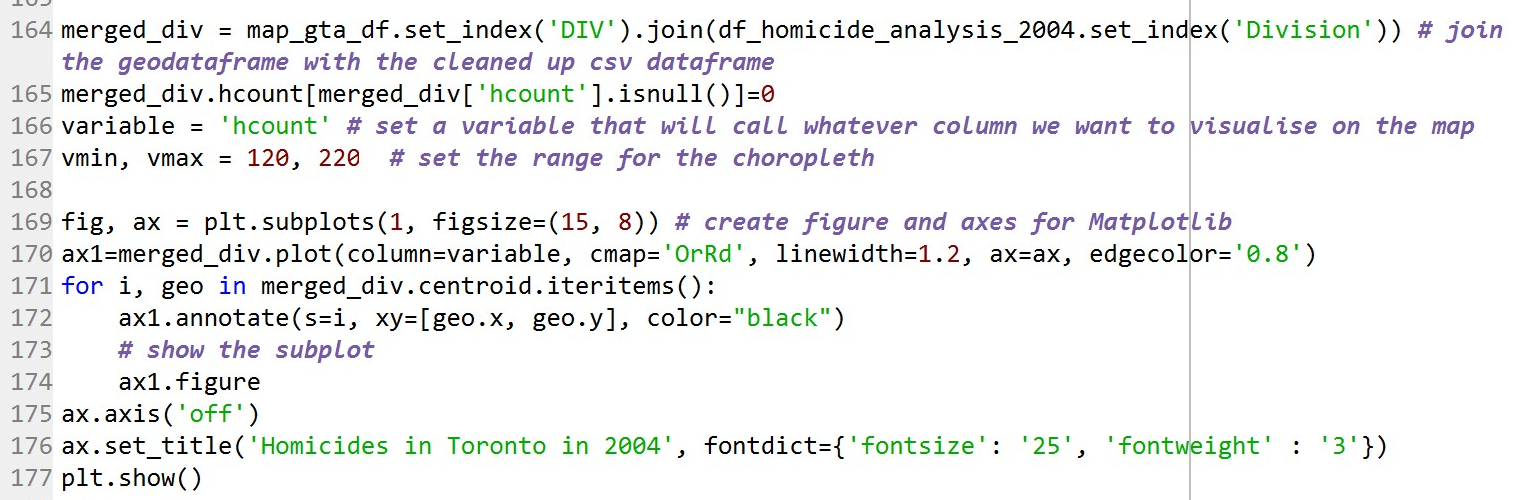
* Repeat the **steps 1 and 2** from above.
* **Step3:** Read the Shape file and store in a Geo Data Frame.



* **Step4:** Perform the required Analysis and merge the data with our Geo Data Frame

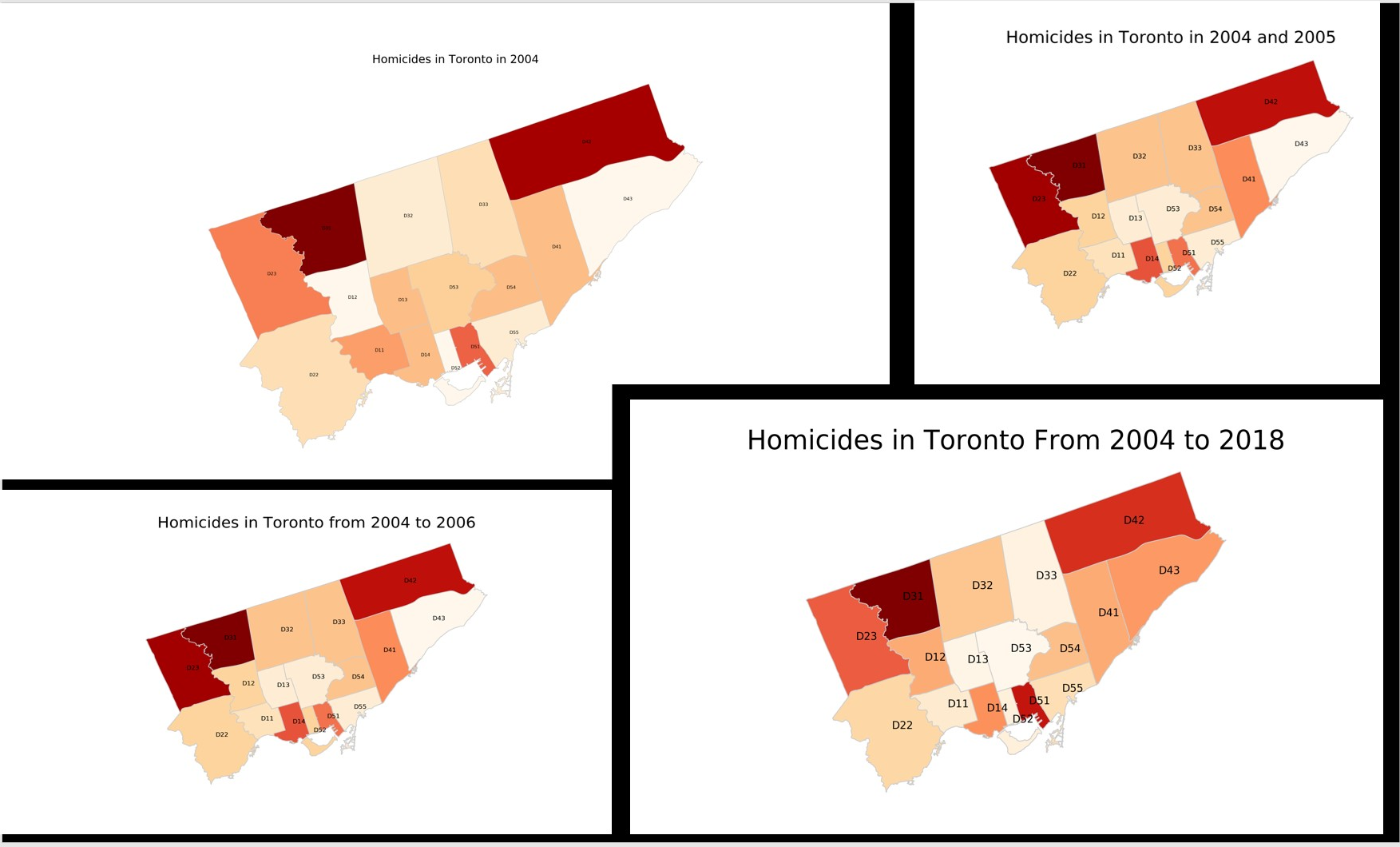


* **Step5:** Prepare the data for Visual representation



* **Result:**

In order to get get all the four maps, we need modify and repeat the above analysis step for all the years required.



**7. Conclusion:**

After performing a detailed analysis from above, we can conclude the following:

1. Division ‘D31’ is the most vulnerable division and the trend has not seen much improvement since 2004.
2. We have noticed significant improvements in certain divisions like D13 and D11, while certain divisions like D12 and D42 have been observed to witness a downward trend.
3. Gun Violence has been the primary reason for Homicides across all districts.
4. Most of the Homicides tend to occur in the Month of July.
5. The top 5 safe neighborhoods to live in Toronto w.r.t Homicides are Forest Hill North, Bedford Park-Nortown, Forest Hill North, Playter Estates-Danforth and Yonge-Eglinton.