How to run the program:

The entire code for the program has been written using python 3.9.6. The codes have been broken into modules namely:

- GUI Main File: This module interacts with other modules and serve as a GUI/interactive console
- data_scraping_via_api_module: This module extracts the dataset from the U.K police database
- data_cleaning_module: This module cleans the entire dataset extracted
- visualization_module: This module contains functions that generate visual charts from the dataset

To run the entire program, open the GUI Main File in your python environment and run the code by pressing F5 (for python shells), this automatically interacts with the other modules and generate the visual charts in matplotlib windows. Close each figure to generate the next figure.

Dependency Information:

The dependency library used are: pandas, matplotlib, seaborn, pywaffle and squarify. These libraries can be installed on your local system by using the pip command in the command prompt terminal:

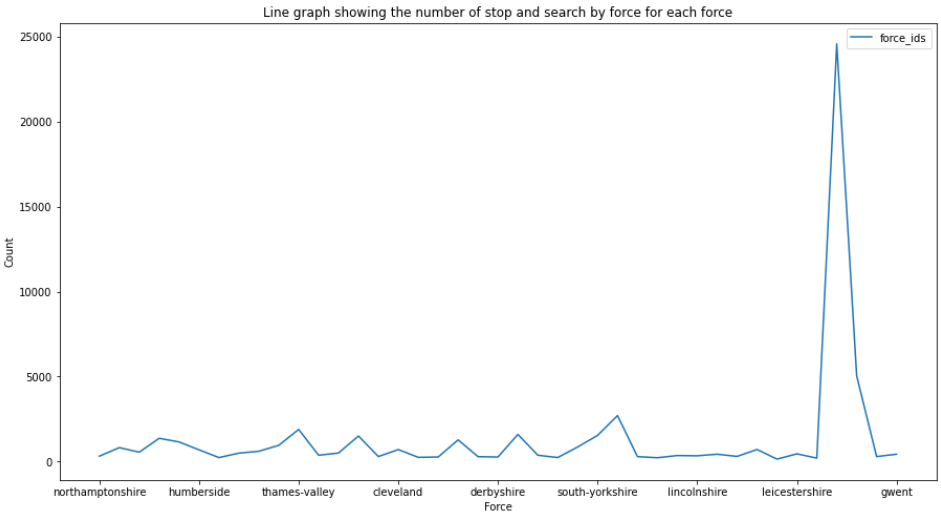
- pip install pandas
- pip install matplotlib
- pip install seaborn
- pip install pywaffle
- pip install squarify

How to run the unit tests:

Each of the visualization chart code have been written as a function, beneath each plot function lies another cell that contains the code for its unit testing, you can run the unit test by running the cell; although all codes including the unit tests automatically runs when you run the entire program as described in the section: “How to run the program” above.

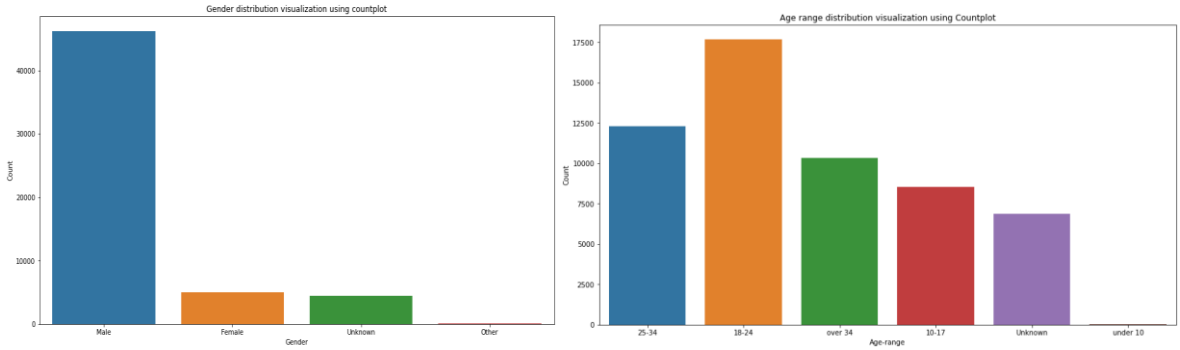
Details of charts produced:

1. Line graph showing the number of stop and search conducted by each force unit:

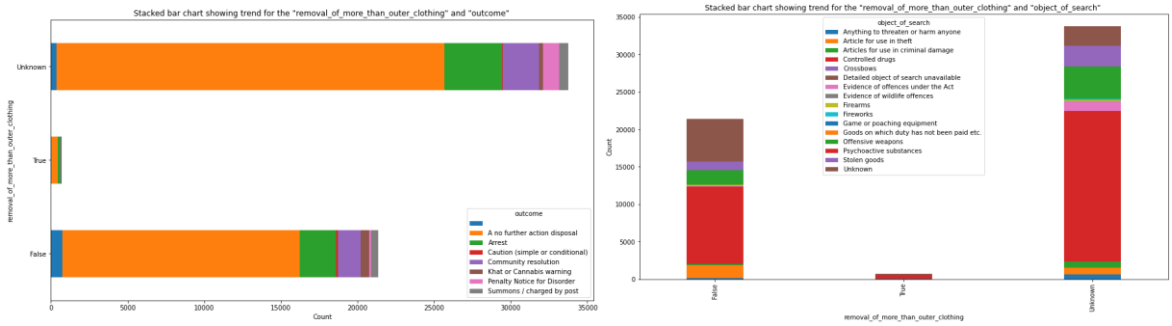


2. Count plots showing Gender distribution and age distribution across the dataset.

- X-axis: Gender/Age-distribution
- Y-axis: Number of searches conducted on them
- Description: The gender count plot shows that males were most searched while the age count plot shows that people of ages 18-24 were most searched.

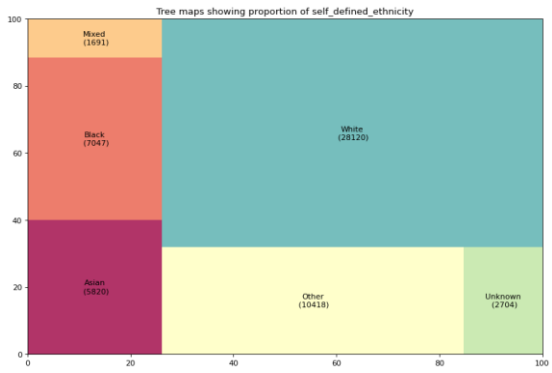


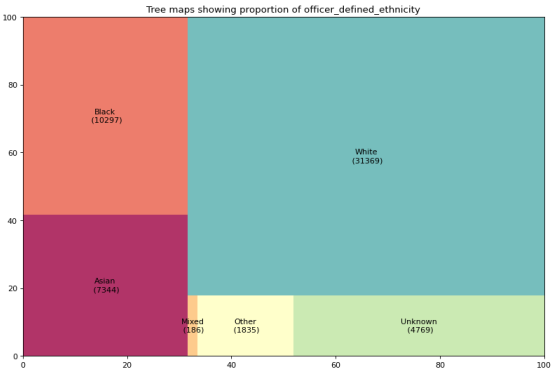
3. Stacked bar graphs showing the relationship that exists between searched people who had “more than an outer clothing removed” and “outcome/object of search” respectively:



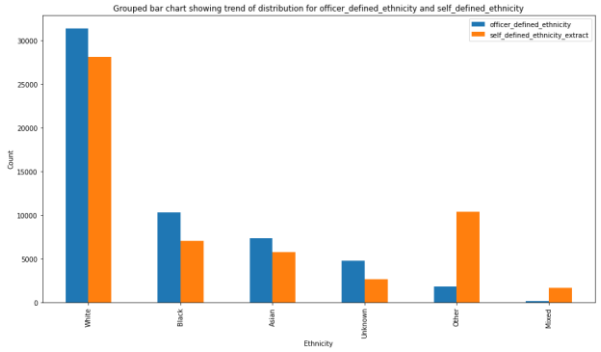
4. Tree maps showing the distribution of “officer defined ethnicity” and “self-defined ethnicity” of people searched:

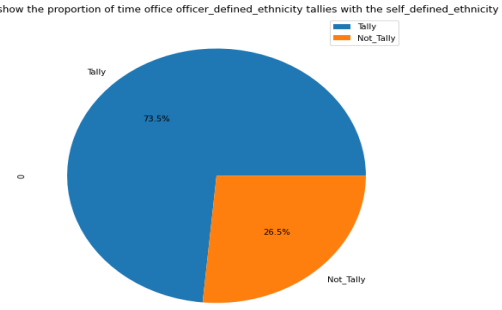
- Self-defined-ethnicity: The self-defined ethnicity of the person stopped
- Officer-defined-ethnicity: The officer-defined ethnicity of the person stopped



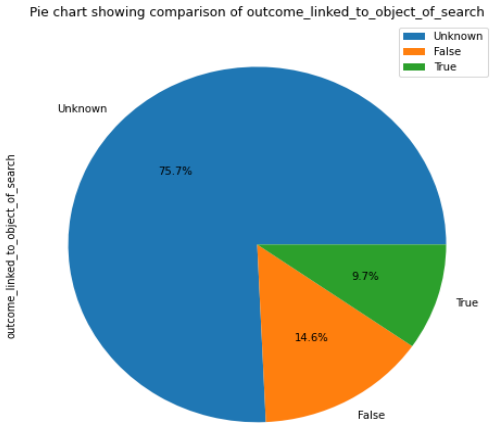


5. The grouped bar chart below shows the proportion of ethnicity of searched people has stated by the “person searched [self-defined ethnicity]” and “officer who searched [officer-defined ethnicity]”, the pie chart shows the number of times the “self-defined ethnicity” tallies with that of the “officer-defined ethnicity”.



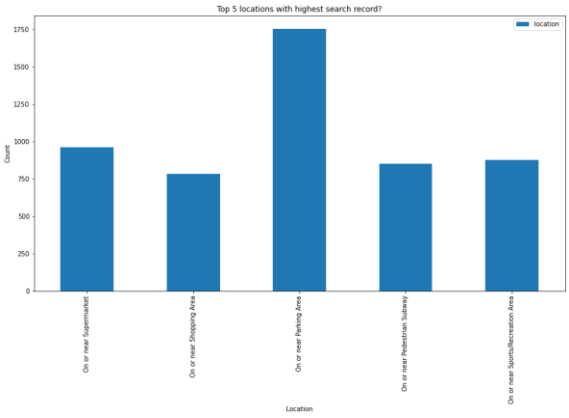


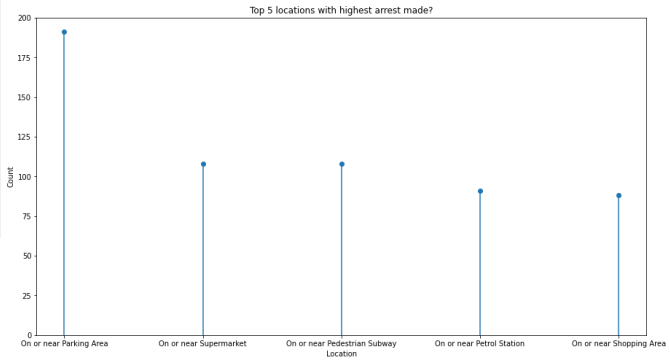
6. Pie chart which shows the proportion of time in which the outcome of the stop and search was related to the reason the stop and search was carried out in the first place.



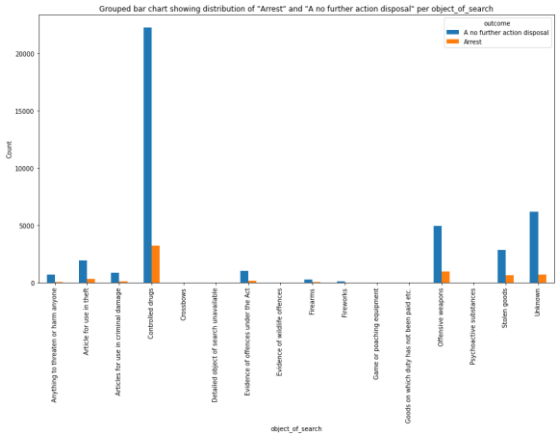
7. A grouped bar chart and stem plot showing respectively the top 5 locations stop and search is most carried out and the top 5 locations arrests were made respectively:

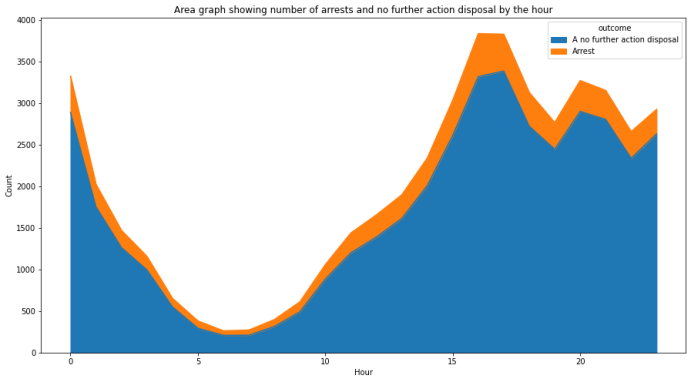
- X-axis: Location [location of the incident]
- Y-axis: Count [Number of searches/arrests made in that location]
- Description: The plots show that searches and arrests most occur around parking areas.



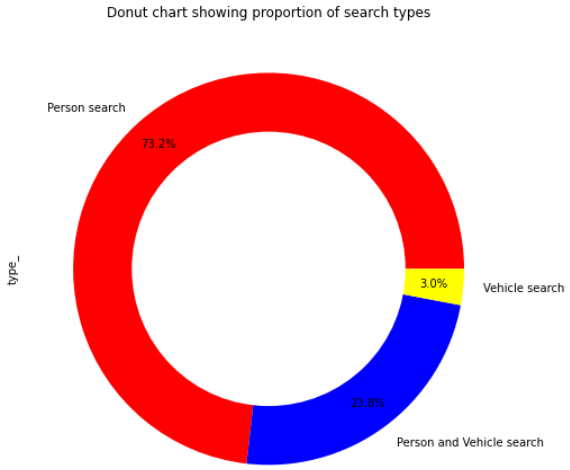


8. A grouped bar chart and an area chart showing the proportion of “arrests made” and “no further action disposal” made per “object of search [The reason the stop and search was carried out]” and hour in which search was conducted. The graphs reveal that for most searches, more no further action disposal was carried out as compared to arrests and that the hour of the day with most arrests made is the 16th-17th hour and the object of search people were most arrested for was controlled drugs

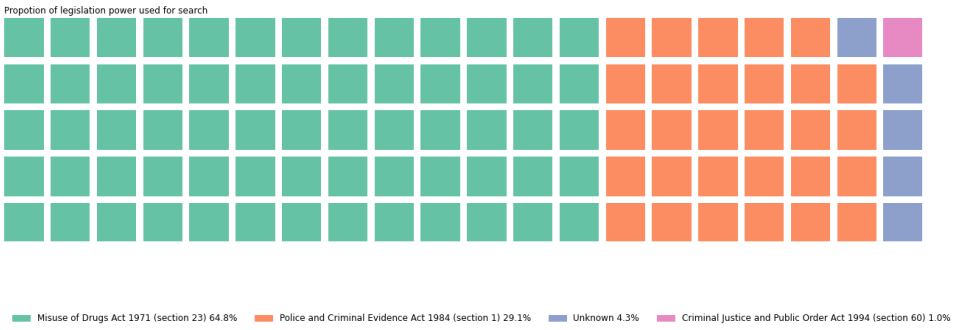




9. A donut chart showing proportion of search types [Whether this was a 'Person search', a 'Vehicle search', or a 'Person and Vehicle search'], the chart shows that most search was a person search.



10. The waffle plot shows the top-most used legislation power used in carrying out stop and search.



11. The bar graph below gives information about the age-group distribution of the people searched by the Cleveland police force in the month of march, 2021.