How to run the program:

The entire code for the program has been written using python 3.7.10 in Jupyter notebook; Jupyter notebook is an IDE (Interactive Development Environment). The environment already creates a console menu system as you can create readable analyses, keep codes, images, comments, formulae and plots together.

To run the entire code and view the visualization charts simultaneously; Press **Ctrl + Shift + Alt + Enter** OR **Ctrl + F9** (if you are running from google colab), these automatically runs all the programs and gives the charts without need for any command line arguments.

Dependency Information:

The dependency library used are: pandas, matplotlib, seaborn, pywaffle and squarify. Libraries such as pandas, matplotlib/seaborn are data analysis and visualization libraries respectively and they come pre-installed on jupyter notebook. They can however be installed manually using the pip command:

- ➢ ! pip install pandas
- ➢ ! pip install matplotlib
- ➢ ! pip install seaborn

While pywaffle and squarify are dependency libraries that do not come pre-installed on jupyter notebook and must be installed manually before use, they can also be installed using the pip command:
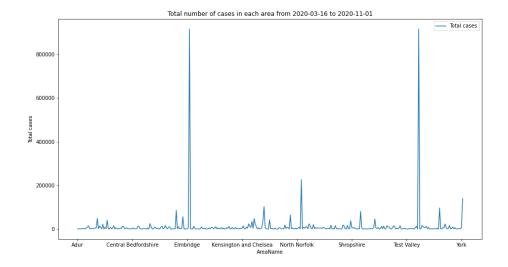
- ➢ ! pip install pywaffle
- ➢ ! pip install squarify
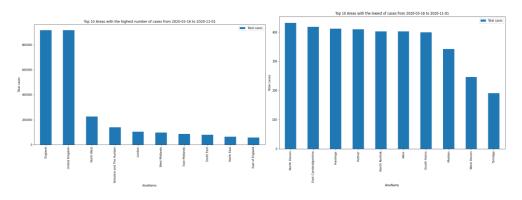
How to run the unit tests:

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.
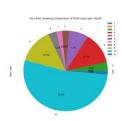
Details of charts produced:

1. Line Graph showing total Number of cases in each area from 2020-03 to 2020-11:
✓ X-axis: AreaName [Names of the areas in the dataset]
✓ Y-axis: Total Cases [Total cases observed in the respective area through the dataset]
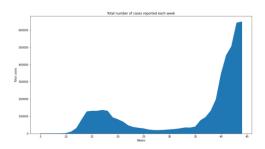✓ Description: The graph shows the certain areas have spiked number of total cases compared to others.

Total number of cases in each area from 2020-03-16 to 2020-11-01

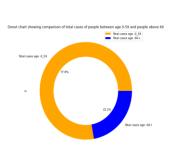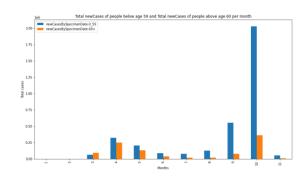2. Bar graphs showing the Top 10 areas with the highest number and lowest number of cases:

✓ X-axis: AreaName [Names of the areas in the datasets]
✓ Y-axis: Total Cases [Total cases observed in the respective area through the dataset]
✓ Description: The graphs shows that England/Uk recorded the highest cases while Torridge the lowest.

3. A Pie chart showing the comparison of total cases by month and an area chart showing the comparison of total cases through the weeks:
✓ X-axis: Weeks/Months [The Months and Weeks cases were recorded through the dataset]
✓ Y-axis: Total Cases [Total cases observed in the respective Months/weeks through the dataset]
✓ Description: The graphs shows that the month with the highest number of cases is 10(October); with more than 50% of the total cases recorded in that month and the week with the highest case is week 40-45 with extreme number of cases recorded as compared with other weeks.
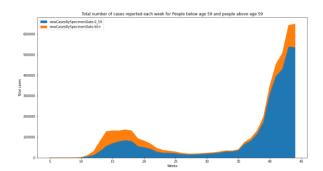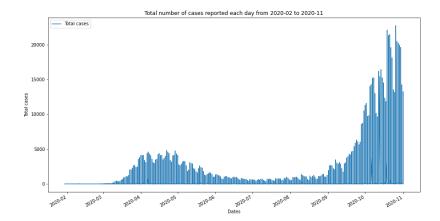
4. A Donut chart, a Grouped bar chart and a Stacked area chart all showing the distribution of the cases by unique age range (those below 59 and those above 60):

✓ X-axis: Months/Weeks [The Months/Weeks cases were recorded through the dataset]
✓ Y-axis: Total Cases [Total cases observed in the respective Months/Weeks through the dataset]
✓ Description: The grouped bar graph and area chart shows that through the months and weeks, more cases were recorded for people below age 59 and as seen in the pie-chart 77% of the total cases recorded was from people in the age-group 59 and below.
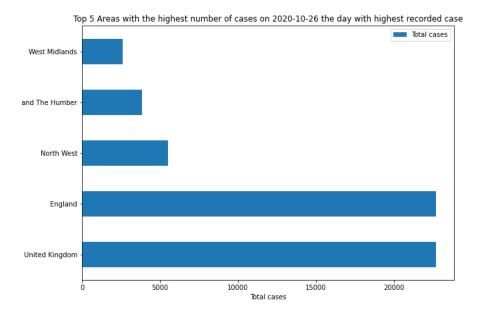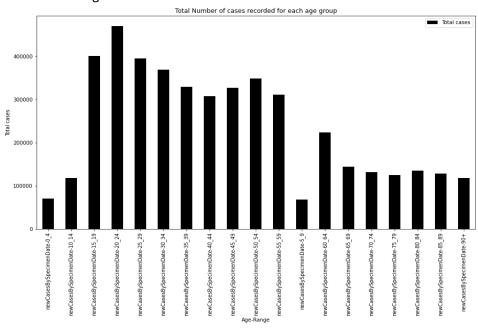

5. Graph showing distribution of cases per Day:
✓ X-axis: Dates [The Days cases were recorded through the datasets]
✓ Y-axis: Total Cases [Total cases observed in the respective days through the dataset]
✓ Description: The graph shows that a first spike occurred between 22-04 and 2020-06 and a second spike much greater than the first occurred from 2020-10 through 2020-11.
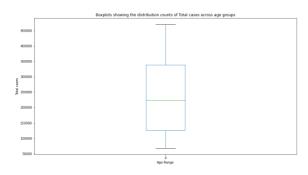
Total number of cases reported each day from 2020-02 to 2020-11

6. Graph showing the areas that made the top contributions to cases on the day with highest recorded cases:
   ✓ Y-axis: AreaName [Names of the areas that made the most contribution to recorded cases on 2020-10-26]
   ✓ X-axis: Total Cases [Total cases observed in each area that particular day]
   ✓ Description: The graph shows that England/Uk contributed the most to the spike experienced that day, followed by North west.
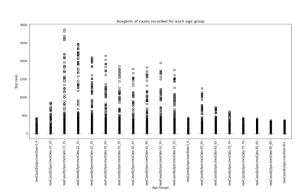

Top 5 Areas with the highest number of cases on 2020-10-26 the day with highest recorded case

7. Bar graph showing total number of cases recorded for each age groups:
   ✓ X-axis: Age-Groups [The age-range of people as grouped in the dataset]
   ✓ Y-axis: Total Cases [Total cases observed for each age-range through the dataset]
   ✓ Description: The graph shows that age-group 20-24 contributed the most to total recorded cases while ages 5-9 contributed the least.
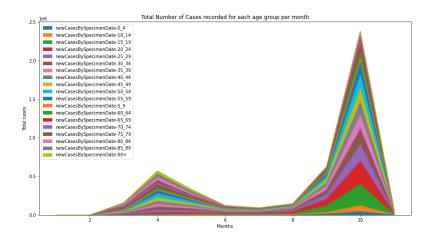

8. Box-plots showing the distribution of Total cases as a whole and across respective age groups:

9. Area graph showing the contribution of each age groups across the months:
   ✓ X-axis: Months [The Months with recorded cases in the dataset]
   ✓ Y-axis: Total Cases [Total cases observed for each month through the dataset]
   ✓ Description: The area graph shows the proportion of each age-group contribution to total cases recorded as the disease progresses through the months.

Total Number of Cases recorded for each age group per month

10. Waffle chart showing proportion of each age-group contribution on the 10th Month; the month with highest recorded cases. The chart reveals that age-group 20-24 and 15-19 contributed most significantly to the recorded cases that month.

11. Graph showing total number of cumulative cases through the days: