7PAM2000-0901-2023: APPLIED DATA SCIENCE 1

Assignment 1: Visualization

By Chiamaka Expensive Joy Ezenezi, student id no: 22058942

GitHub: https://github.com/CEJ-Ezenezi/ADS1-Projects/blob/main/ADS1%20assignment1.py

Dataset link: https://www.data.gov.uk/dataset/43b2334a-6970-4e71-a27a-630dc4901e04/coronavirus-covid-19-deaths#licence-info

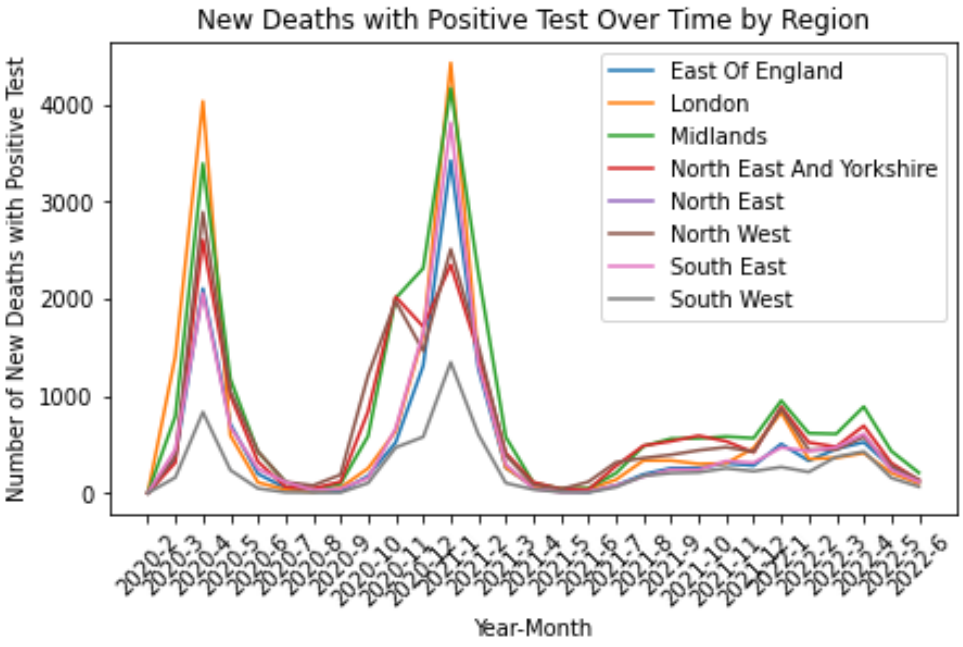
DATASET SUMMARY

The dataset portrays the entries of deaths by regions of England during the period of the Corona virus (covid-19) pandemic, that is, from the year 2020 to 2022. The total deaths are relative to deaths tested positive to covid-19 and deaths that are not tested positive. In a bid to draw insight from the data, I decided to use the following three types of visualization:

1. **Line plot:**

A line plot is a graph that shows datapoints joined by a line. It is generally used to show change in values over time. From the data used, I tried to show the evolution of deaths that were incidental (or tested positive) to covid-19, categorized by the regions of England. A multiple line graph with each line representing a region in England was used to give a picture of the data at a glance.

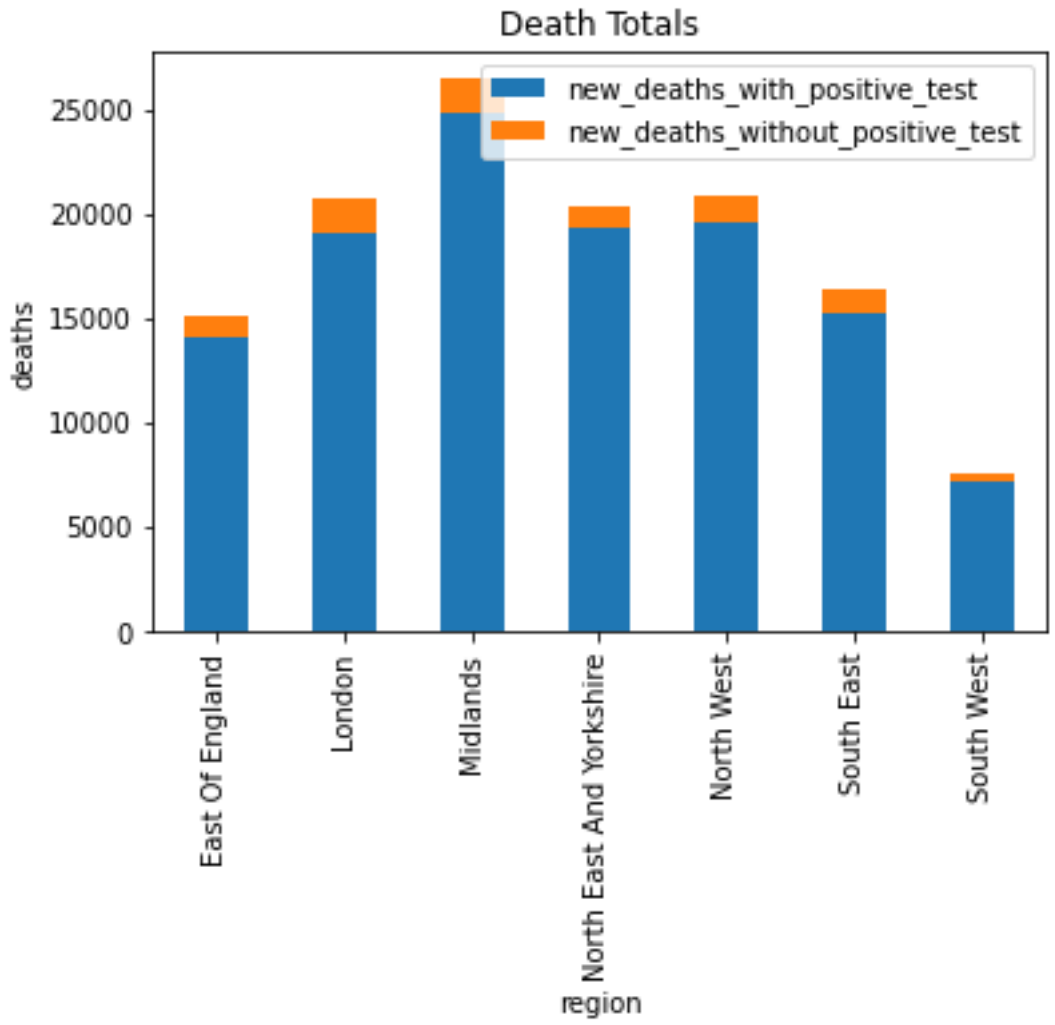
The plot below shows that the evolution of deaths that tested positive for the regions are closely moving in the same pattern with peaks at the beginning of year 2020 and 2021. There was recorded decline in deaths from the middle of year 2021 to the point when the data ended. Also, notice that the Southwest region records the least deaths.



2.**Stacked Bar Chart:**

A traditional bar chart or graph is used to present categorical data as bars with their heights or lengths proportional to values they represent.

Stacked bar chart is used in representing a grouping variable (categorical data), where group counts or relative proportions are being plotted in a stacked manner, with the height and length of the bars, a combined result of the variables. The plot below shows a stacked bar plots for total deaths in each region of England. From the visualization, we can deduce that during the time under consideration the number of deaths that were not tested positive to covid-19 was relatively small compared to the deaths tested positive. We also see that the Midlands region records the highest number of deaths in the period.



3.**Pie Chart:**

A pie chart is a circle chart used to represent a nominal data with each divided segment representing a category relative to each other. This visualisation also suits this data as the proportion of deaths by region in the England is represented by each segment/colour of the circle. The percentage of the death in each region with relation to the total covid tested deaths can also be calculated. From the visualisation is we can still deduce that the Midlands region records the most deaths with 20.8%, followed by the North West with 16.4%, closely followed by the North East and Yorkshire with 16.2% and we still see the South West region as the least recorded deaths with only 6.0%.