Department of Computer Science and Engineering (Data Science)

Experiment 2

(Data Visualization)

NAME: Dev Patel   SAP ID: 60009200016

BATCH: K/K1   DATE: 24/11/2021

Aim: Perform time series aggregation, apply filters on a given dataset, create line and area charts.

Theory:

Time aggregation:

Time aggregation is the aggregation of all data points for a single resource over a specified period (the granularity). Data aggregations in Resource Time Series reports are of the time aggregation type.
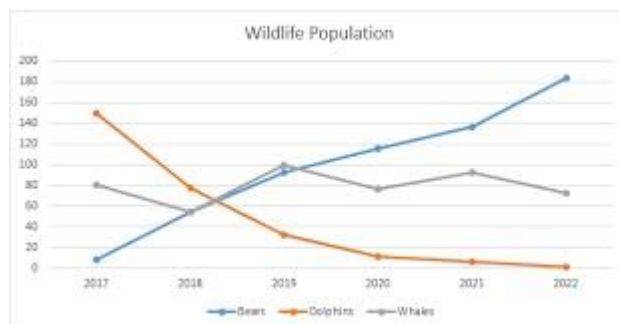
The result of the aggregation is one data point that reflects a statistical view of the collected and aggregated data points. For example, average, minimum, maximum, sum, or count. Typically, multiple aggregated data points are presented in a report for a given reporting period.

Data Filtering:

Data filtering is the process of choosing a smaller part of your data set and using that subset for viewing or analysis. Filtering is generally (but not always) temporary – the complete data set is kept, but only part of it is used for the calculation.

Line Charts:

A line chart displays information as a series of data points called 'markers' connected by straight line segments. It is a basic type of chart common in many fields.
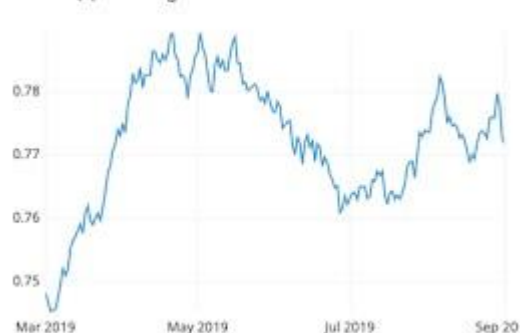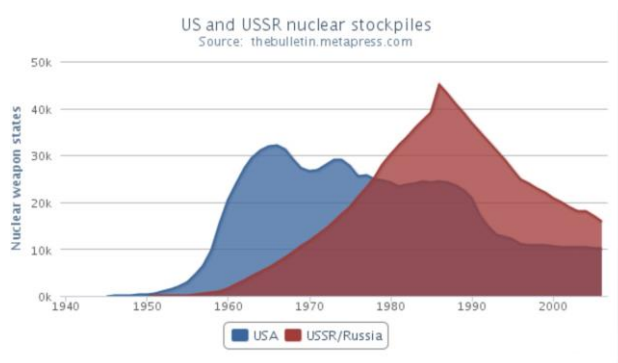
Area Charts:

An area chart is like a line chart in terms of how data values are plotted on the chart and connected using line segments. In an area chart, however, the area between the line segments and the x-axis is filled with colour. It helps in comparison of multiple series, relationships between two series and analysis of various trends.

Department of Computer Science and Engineering (Data Science)

Tool Used:

Tableau Public 2021.3

Dataset:

https://raw.githubusercontent.com/nytimes/covid-19-data/master/us.csv Compatible Web Data Connector for easy importing:

https://basic-csv-wdc.herokuapp.com/

Visualizations:

a) **Vaccination in the USA started in full swing in Early Jan 2021. Analyse the trend of Cases per day from Jan 2021 (Vaccines begun) to the current day and report if this policy was useful.**
On X axis: Dates in the year 2021
On Y axis: Cases Per Day (Difference in Cases)

The vaccination drive started in Jan 2021. We can clearly see the number of cases reducing by the day for the next few months. Thus, we can report that the policy was indeed useful.


USA COVID Cases in 2021
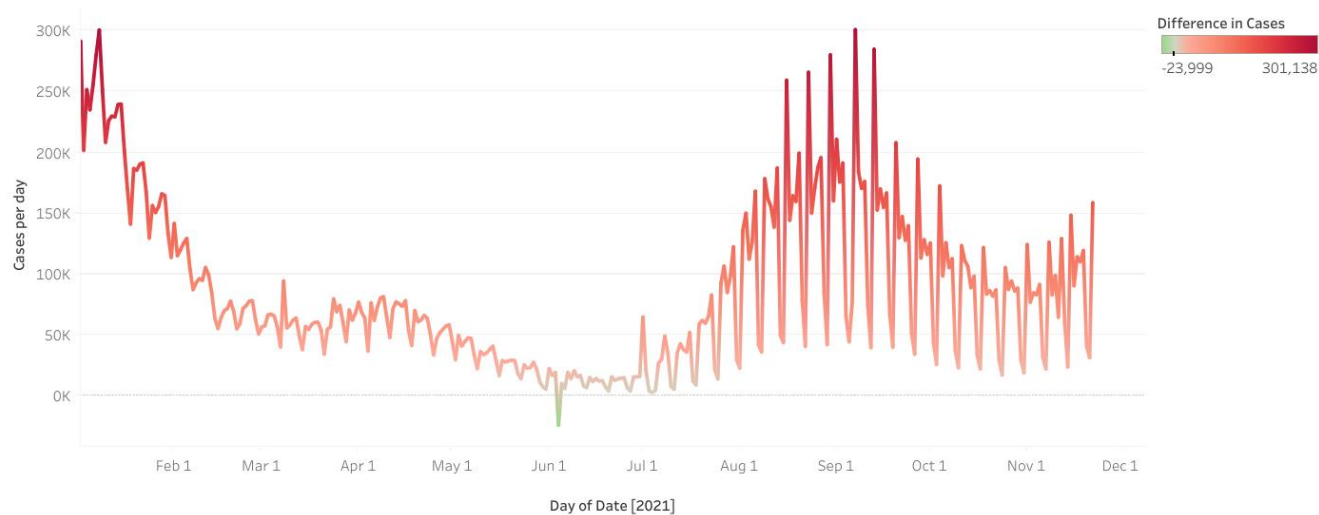
b) **The USA implemented a policy where Vaccinated Individuals could avoid wearing masks in public from June 2021. The UN wants to view the impact of making such a move in the long term as well as the short term.**
On X axis: Dates from the 1st of June 2021
On Y axis: Cases Per Day (Difference in Cases), Deaths per day (Difference in Deaths)

While the cases were in the decline, thanks to the hugely successful vaccination drive, the policy of allowing vaccinated individuals to not wear masks turned out to be a failure. While it did seem alright at first, in the long term, the cases started rising and the deaths per day due to the pandemic also went up.

USA COVID Impact from June 2021
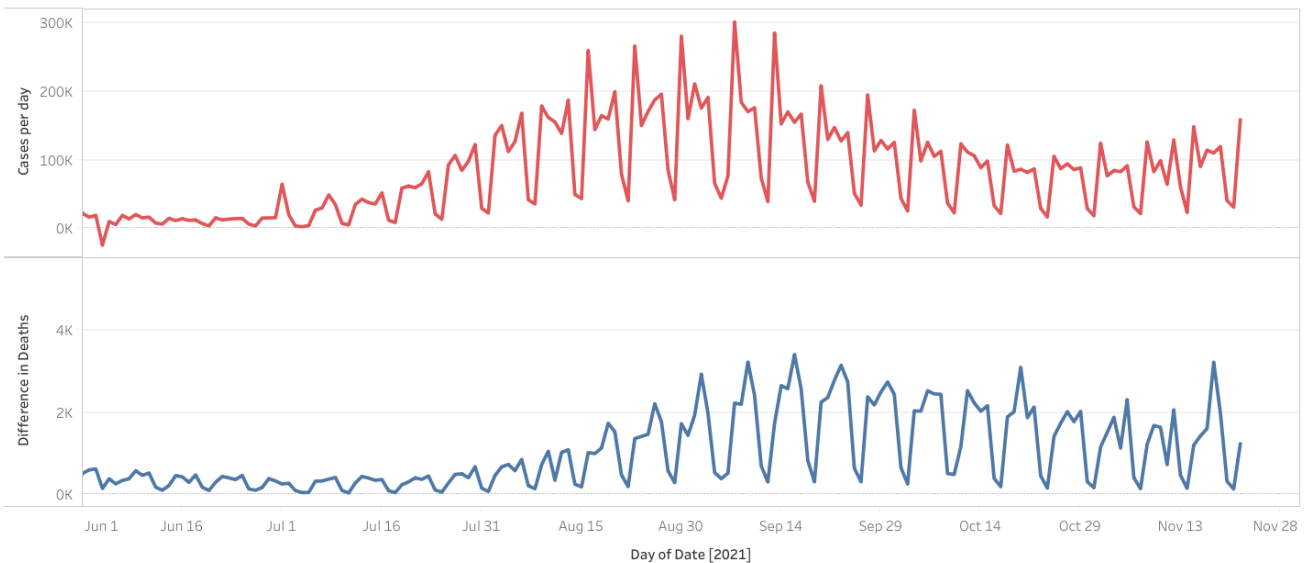
c) **Analysis requires the creation of a calculated field in terms of Cases Per day and a Weekly moving average of cases per day.**
On X-axis: Dates
On Y-axis: Weekly Moving Average (Yellow Area), Cases per Day (Red spots)

The visualization using Cases per Day and Weekly Moving average has been made. The red spots indicate the cases per day and the yellow area chart shows to weekly moving average of cases per day so that invaluable inferences can be made.

USA Daily COVID Cases and Weekly Moving Average
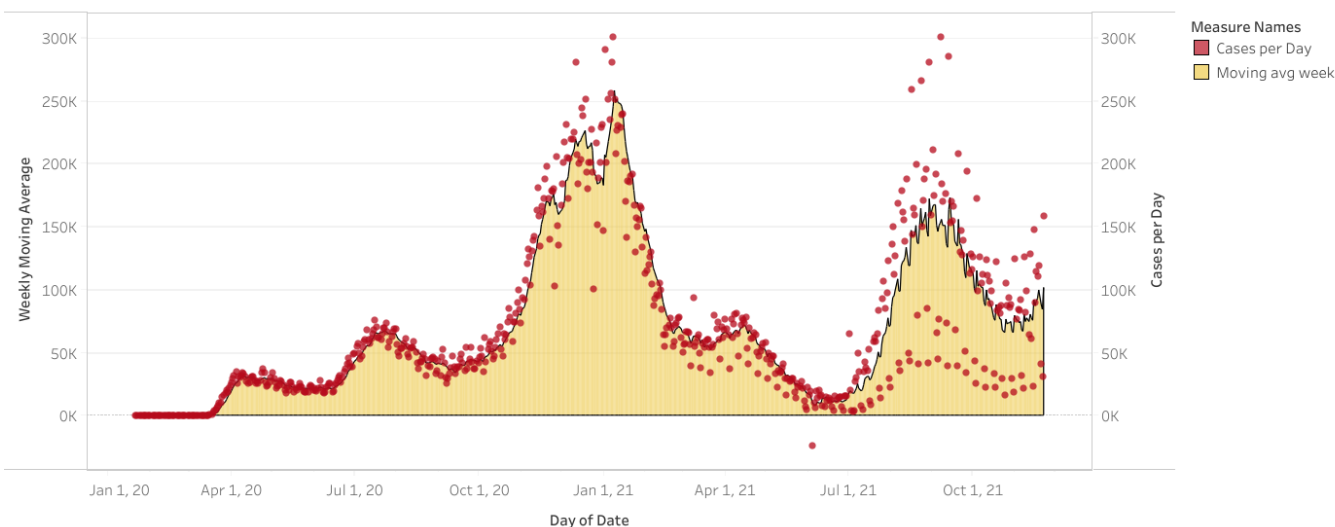
d) **Justify why you have chosen a particular visualization for this dataset. Also discuss the details of your visualization.**

The line chart has been used to perform time series aggregation. The dataset had cumulative number of cases and deaths until the particular day. So, a quick table calculation of difference was used, so that we can get the number of cases or deaths of each individual day. As data had continuous data, we used to a line chart to give us a better idea of how the curve is moving.

In the first and second visualization, we used a line chart to observe and analyse the trends of the number of daily cases and deaths. We also used colour grading in the first question. In the third visualization, we have an area chart which denotes the weekly moving average of the number of cases in the USA since the start of the pandemic. This weekly moving average of daily cases was a calculated field. We also have red spots that show the number of deaths on each particular day, so that comparison can be made between them. Dots were used because line chart along with the area chart looked indistinct. Hence, we learnt how to perform time series aggregation, apply filters on a given dataset, create line and area charts.