**University of Texas at Dallas**

MIS 6380.501 - Data Visualization – Thu 7 PM Class

**GROUP- 10**

**Explore the relationship between comorbidities and other conditions with the number of COVID-19 deaths in the USA from 2020 to 2023.**

Report Prepared by:

Kruthagama Divakar-kxd220014

Jacob Samuel Mathew-jxm220007

Nihal Darak-nxd220015

Korrapati Naga Ruthvika-kxn210050

Harish Radhakrishnan Krishnamoorthy-hxr210030

**EXECUTIVE SUMMARY:**

The aim of this project is to explore the relationship between comorbidities and other conditions with the number of COVID-19 deaths in USA from 2020 to 2023.

The Primary dataset shows the number of COVID-19 deaths and the underlying medical conditions among Americans who eventually died due COVID. We are going to utilize this data along with a few other secondary datasets to better understand how factors such as immunity, pre-existing conditions, population, climate, and vaccinations affected the severity of the disease.

INDEX

Content Page number

Data Description 4

Data Cleaning 6

General Introduction 6

Insights and Findings 7

Hypothesis: 1 7

Hypothesis: 2 8

Hypothesis: 3 10

Hypothesis: 4 11

Hypothesis: 5 13

Conclusion 15

**Data Description:**

The Primary Data set that we chose for this project was from the Center for Disease Control and Prevention’s (CDC’s) official website. It shows the contributing conditions prevalent in those individuals who died due to COVID in USA between 2020 and 2023.

The Data Set contains 559K rows that contain information about COVID deaths and are segregated by age group and jurisdiction of occurrence in USA.

The Various columns in the dataset provide the following information:

| Column Name | Description | Type |
| --- | --- | --- |
| **Data As Of** | Date of analysis | Date & Time |
| **Start Date** | First week-ending date of data period | Date & Time |
| **End Date** | Last week-ending date of data period | Date & Time |
| **Group** | Time-period indicator for record: by Month, by Year, Total | Plain Text |
| **Year** | Year in which death occurred | Number |
| **Month** | Month in which death occurred. | Number |
| **State** | Jurisdiction of occurrence | Plain Text |
| **Condition Group** | Condition Group | Plain Text |
| **Condition** | Condition contributing to deaths involving COVID-19 | Plain Text |
| **ICD10\_codes** | ICD-10 code for condition | Plain Text |
| **Age Group** | Age group | Plain Text |
| **COVID-19 Deaths** | COVID 19 Deaths | Number |
| **Number of Mentions** | Number of positive COVID Cases | Number |
| **Flag** | Counts less than 10 supressed. | Plain Text |

For Analysis, we also used several secondary data sets from reliable sources, these were

1. https://www.visualcrossing.com/weather/weather-data-services

2. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/reporting-vaccinations.html>

3. https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html

**Data Cleaning:**

The primary data set 599K rows and each of the Secondary Dataset has a few hundred to several thousand rows which made it was quite challenging for us to clean the dataset. We cleaned the data by using MS-Excel. We segregated each record based on age groups.

**General Introduction:**

The COVID-19 Pandemic took numerous lives in USA between 2020 and 2023. During its entire existence there were numerous ambiguities on which underlying heath condition had a deadly combination with the COVID virus eventually leading to death. There were questions on how the population and social distancing in a particular area affected the spread of the virus. Some even speculated that the weather too affected the spread of the virus and its strength.

Through our analysis, we aim to explore the various trends that underlying health conditions, population, weather and vaccination had on the number of deaths due to COVID in USA

**INSIGHTS AND FINDINGS:**

**Hypothesis 1:**

**Individuals aged 75 and above in California had elevated mortality rate due to COVID compared to any other age group because of their relatively weaker immune systems.**

Chart, bubble chart

Description automatically generated

In the first graph we see the number of underlying health conditions that were present in people who died due to COVID in various age groups. The bars of different colors show different underlying health conditions. It is observed that in age groups 75-84 and 85+, the number of underlying health condition the individuals had were greater than any other age groups.

The bubble chart further supports our hypothesis by proving that individuals aged 75 and plus accounted to around 18% of deaths which is a significant percentage when compared to other age group. An interesting observation is that even though the number of cases was highest in age groups 18-34, deaths were highest in age group 80+. This further supports our hypothesis that people aged 75+ because of their multiple underlying conditions and weekend immunity saw the highest number of deaths

**Hypothesis 2:**

**The general trend shows most Americans died due to Heart Diseases, but between 2020 and 2023, majority of Americans died due to Respiratory diseases because COVID was at it peak and COVID virus caused respiratory problems**

**Chart, bar chart

Description automatically generated**

In the years 2020 to 2023, we see that the major contributor to death in USA was COVID 19 followed by Respiratory failure and Influenza. But when we look at historic data, we find that 33% of all deaths in USA was caused by heart diseases which is seen in the bubble chart below.

The bar chart below further shows that between 2020 and 2023 deaths due to Respiratory diseases and COVID itself occupy the first two spots, then followed by heart and circulatory diseases.

This clearly states that the COVID Virus which primarily affected an individuals respiratory system was the sole contributor to majority deaths in the United states. If there had been no pandemic, majority of US population who have continued to die due to circulatory diseases.

**Chart, bubble chart

Description automatically generated**

**Hypothesis 3:**

**California and Texas have more covid deaths when compared to other states because   
the population of California and Texas is high.**

**Chart, treemap chart

Description automatically generated**

It was a known fact that COVID virus spread rapidly when the affected individuals were in close contact with other un-affected individuals. Thus we can assume that highest populated regions must have seen more deaths due to covid than other regions.

Our analysis supported this assumption, the heath maps shows the population of each state through the size of each tile. Since California and Texas are the 2 highly populated states in USA, its tile occupies a greater area. The intensity of red color in each tile show the number of people who died due to COVID. This is also seen in the visualisation with the map of USA. California. And Texas have been represented with different colors as they had the highest population and the highest number of deaths due to COVID.

**Hypothesis 4:**

**Deaths due to COVID are higher in December and January when compared to other months because covid transmission rate is higher in colder weather.**

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Even though close proximity was considered the major factor contributing to the rapid spread of COVID virus, we also observed that weather had an affect on COVID Transmission. In the graph above we see the pink line representing average temperature in USA and the blue line representing the number of deaths due to COVID in US. It is clearly seen that as the average temperature reduced during the winter months, the number of deaths due to covid increased.

Additionally, in the area graph below, we also see that the number of people who tested positive for COVID were higher in winter months than in summer months. Thus we can conclude that temperature significantly affected COVID transmission and also deaths due to COVID

**Chart

Description automatically generated**

**Hypothesis 5:**

**Map

Description automatically generatedCOVID cases in 2022 got flattened when compared to 2021 because a greater proportion of Americans were fully vaccinated by 2022.**

**Map

Description automatically generated**

Chart, histogram

Description automatically generated

As the number of US Citizens who completed their vaccination series increased we see that the number of people who died due to COVID gradually reduced.

The visualisation with US map shows that as the total number of people who got fully vaccinated in the US increased from 2020 to 2023, the total number of deaths due to COVID reduced and in the graph right above our discussion we that the line, that is representing the number of people who died due to COVID reached its peak in December 2021 but as the number of people who got completely vaccinated i.e. represented by the yellow bars increased the line started stabilising in 2022 when compared to 2021.

**Conclusion:**

There are several other factors that affected the transmission of COVID in US from 2020 to 2023 and eventually led to death of Americans, from all our analysis through the 5 hypothesis we can conclude that:

* Since individuals respiratory system was affected by COVID. Respiratory disease took over as major contributor to death among Americans than Heart Diseases
* Americans in the age group of 75 and above had more deaths due to COVID because of a weaker immune system.
* Since COVID spread through close contact among individuals, the states with highest population also saw the highest number of covid deaths.
* As Transmission of COVID increased in months with lower average temperature, number of deaths due to COVID also increased.
* Number of deaths due to COVID reduced and also stabilized in 2022 than in 2021 as more and more Americans were fully vaccinated to the COVID Virus.