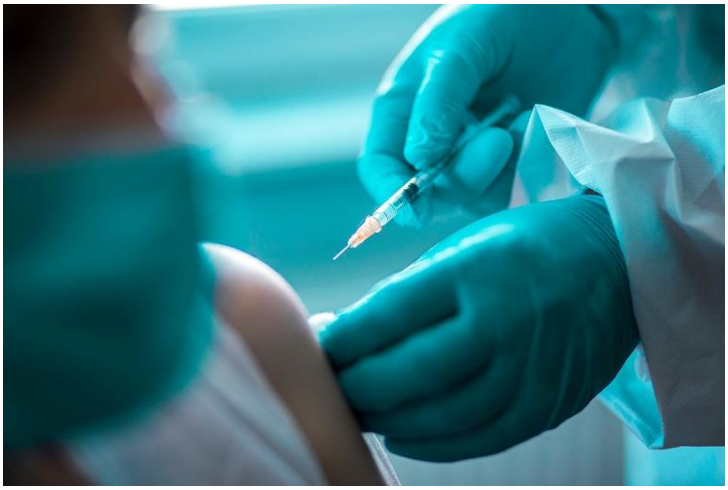
GROUP PROJECT PROPOSAL
GROUP – 4

MIS 6380.004 DATA VISUALIZATION



AMOL BHADANE

RADHIKA DUBEY

PRANAV MODY

HEMANGI PATIL

PRIYA VERMA

SUJAN NAIDU BHASKAR SURISETTY

Project Description:

As New York State grappled with the COVID-19 pandemic, its healthcare workforce stood at the forefront, valiantly battling the virus. This project delves into the intricate patterns of COVID-19 infections, vaccinations, and hospitalizations across the state's diverse regions. By scrutinizing vaccination rates among healthcare workers and exploring potential disparities between metropolitan and non-metropolitan areas, the analysis aims to unravel the nuances that shaped the pandemic's trajectory. Additionally, it investigates the relationship between hospital density and vaccination uptake, shedding light on the factors influencing the adoption of this crucial line of defence.

Furthermore, the project expands its scope beyond New York State, exploring the nationwide impact of the pandemic. It examines the surge in internet usage during the crisis, potentially driven by the widespread adoption of remote work arrangements. Moreover, it delves into the intriguing hypothesis of gender-specific COVID-19 symptom manifestations, specifically concerning fever prevalence in May 2020. Through this comprehensive investigation, the study endeavours to uncover valuable insights that could inform future preparedness strategies and shape our understanding of this unprecedented public health challenge.

Objective:

This project aims to conduct a comprehensive analysis of COVID-19 patterns across New York State and the U.S., unveiling disparities in case distribution between metropolitan and non-metropolitan counties, especially for the 20-44 age group. It explores potential correlations between hospital density and vaccination uptake rates. Additionally, it examines the nationwide surge in internet usage during the pandemic, potentially linked to remote work adoption, and investigates hypothesis surrounding gender-specific COVID-19 symptom manifestations. Through data cleaning, standardization, and visualization, the goal is to present a visual model capturing the pandemic's impact on healthcare infrastructure, hospitalization rates, and vaccination trends in New York over time. The insights generated will inform future preparedness strategies, resource allocation, and public health policies, enhancing resilience against similar crises.

Hypothesis:

Hypothesis 1: In New York State, in 2022, COVID vaccination rates among healthcare workers peaked at 65% fully vaccinated and 55% boosted in Q2 as compared to Q1 before declining by 5 percentage in Q3, due to decreasing COVID cases over time.

Hypothesis 2: In New York State, from March 2020 to February 2022, the 20-44 age group experienced approximately 30% higher COVID-19 cases in metropolitan counties compared to non-metropolitan counties, likely due to around 50% greater workplace exposure from office attendance among metropolitan residents.

Hypothesis 3: In New York metropolitan counties, from January 2021 – December 2023, metropolitan counties are 45% more vaccinated compared to non-metropolitan counties, due to metropolitan counties having a greater number of hospitals.

Hypothesis 4: Due to high population in New York, the COVID hospitalization rate of New York County was 0.5% of the total population i.e. 1,578,801 versus 0.3% of the populations of other counties that are not in metropolitan areas from October 2021 - July 2023.

Hypothesis 5: The United States has witnessed the use of a daily internet communication surge of over 25% during the COVID-19 pandemic in 2020 compared to the pre-COVID era, due to increased preference for remote work / WFH.

Hypothesis 6: In United States, during May 2020, more than 60% females are likely to experience fever as a symptom of COVID-19 than males, in the month of May in the year 2020 in the USA, due to a higher tendency of women to go outside for shopping and grocery.

Data Sources:

The following link contains all attributes included with our dataset.

Dataset1:https://health.data.ny.gov/Health/New-York-State-Statewide-Hospital-Staff-COVID-19-V/qfps-y8ta/about_data

Dataset2: https://www2.census.gov/programs-surveys/popest/tables/2020-2022/counties/totals/co-est2022-pop-36.xlsx - (https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html)

Dataset3:https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Hospitalizations/jw46-jpb7/about_data

Dataset 4: Psycological Effects of COVID | Kaggle

Dataset 5.a: https://www.kaggle.com/datasets/thedevastator/us-adult-covid-19-impact-survey-data

Dataset 5.b: https://data.world/associatedpress/covid-impact-survey-public-data

Number of records:

The dataset 1 contains 17,900 rows, 14 columns.

The dataset 2 contains 125 rows, 4 columns.

The dataset 3 contains 240956 rows, 37 columns.

The dataset 4 contains 1175 rows, 24 columns.

The dataset 5 contains 8975 rows, 177 columns.

Data Cleansing Tool:

We chose Microsoft Excel to conduct the data cleansing because of its powerful capabilities in data cleaning, transformation, and statistical analysis. This selection guarantees preparation of our data for data exploration and hypothesis testing, taking advantage of the powerful capacities of the Excel platform.

Visualization Tool:

In this project, we will utilize two data visualization tools: Microsoft Excel and Tableau. These are because they facilitate the development of informative charts, graphs, and visualizations. These tools will let us present the findings appropriately and tell captivating stories on the effect COVID-19 had on different aspects of adult US Hospital Staff residents' life.