|  |
| --- |
| **Data Visualization Group 1 Report**  MIS6380.501  Data Visualization  Fall 2024  **Project Name**  COVID-19 Pandemic in the City of Chicago (Jan 2020 - Dec 2022): Data Visualization, Analysis, and Insights.  **Group Participants**  Deepmala Srivastava(dxs210033), Suman Anand(sxa230187), Rishabh Balaiwar (rxb220092), Mounika Kolle(mxk240034), Naveena Paleti(nxp230055), Gowtham Siddegowda(gxs240016) |

**TITLE:**

Confirmed COVID-19 cases among The city of Chicago residents who live and work in congregate living facilities within the city of Chicago for the reporting period.

**SUMMARY:**

This report provides a comprehensive analysis of the COVID-19 pandemic in the city of Chicago, covering the period from January 2020 to December 2022. Using publicly available datasets, the study investigates patterns in positive cases, hospitalizations, deaths, testing trends, and vaccination uptake. Key findings reveal strong correlations between cases and healthcare outcomes, disparities in testing and vaccination rates across demographics, and temporal trends in pandemic progression. Data visualizations created in Tableau underscore these insights, offering actionable conclusions for understanding and addressing public health challenges.

**INDEX:**

[1. Data Description](#_Data_Description:):……………………………………………………………………………. 3

[1.1. Dataset Overview:](#_Dataset_Overview)..……………………………………………………………… 3

[1.2. Key Features:](#_Key_Features) ……………………………………………………………………… 4

[2. Data Cleaning](#_Data_Cleaning):……………………………………………………………………………….. 5

[3. General Introduction:](#_General_Introduction:)………………………………………………………………………. 6

[4. Insights and Findings:](#_INSIGHTS_AND_FINDINGS:)……………………………………………………………………… 7

[5. Conclusion:](#_Conclusion:)…………………………………………………………………………………. 12

[6. References:](#_References:)…………………………………………………………………………………. 13

# **DATA DESCRIPTION:**

## Dataset Overview

The analysis relies on five datasets sourced from [data.gov](https://catalog.data.gov/dataset/), focusing on COVID-19 trends in the city of Chicago from January 2020 to December 2022. These datasets contain detailed information on positive cases, hospitalizations, deaths, testing rates, and vaccinations, broken down by demographics and geography.

**Datasets Used**

1. **COVID-19 Vaccinations by Age and Race-Ethnicity**
   * **Rows:** 5,150
   * **Columns:** 12
   * **Focus:** Demographic breakdown of vaccination rates, including age, race, and ethnicity.
2. **COVID-19 Daily Testing by Person**
   * **Rows:** 127
   * **Columns:** 59
   * **Focus:** Daily testing data by individuals, including demographic and geographic details.
3. **COVID-19 Vaccinations by ZIP Code**
   * **Rows:** 43,070
   * **Columns:** 36
   * **Focus:** Vaccination trends by ZIP code, highlighting regional disparities.
4. **COVID-19 Daily Rolling Averages (Cases, Deaths, and Hospitalizations)**
   * **Rows:** 730
   * **Columns:** 97
   * **Focus:** Daily averages for positive cases, hospitalizations, and deaths, reflecting the pandemic timeline.
5. **COVID-19 Vaccinations by Region, Age, and Race-Ethnicity**
   * **Rows:** 5,040
   * **Columns:** 12
   * **Focus:** Vaccination trends by region, age, and race, with insights into disparities across demographic groups.

## Key Features

1. **Temporal Data**
   * Tracks trends over three years (2020–2022) to analyze pandemic progression and public health responses.
2. **Demographic Breakdown**
   * Data includes age, race, and ethnicity, enabling the identification of disparities in testing and vaccination uptake.
3. **Geographic Insights**
   * ZIP code and regional data highlight geographic disparities in healthcare access and outcomes.
4. **Vaccination Metrics**
   * Tracks first-dose, second-dose, and booster uptake, revealing engagement levels and drop-offs.
5. **Healthcare Correlations**
   * Positive cases, hospitalizations, and deaths are analyzed for correlations and lagged effects.
6. **Testing Trends**
   * Provides weekly and demographic testing patterns, highlighting high-engagement groups and time periods.

These datasets, cleaned and normalized for visualization, form the basis of the report’s insights and findings, offering a comprehensive view of COVID-19’s impact in the city of Chicago.

# **DATA CLEANING:**

To ensure the accuracy and reliability of the analysis, the datasets underwent a rigorous data cleaning process. The steps are outlined below:

1. **Removal of Duplicates**
   * Duplicate records were identified and removed to maintain the uniqueness of entries. This step ensured that no data point was redundantly counted, which could skew analysis results.
2. **Handling Missing Values**
   * Missing data were addressed using a predefined threshold: entries with less than 2% missing data were filled with appropriate values (e.g., averages or medians).
   * Records with excessive missing data were excluded to maintain data integrity without compromising the overall dataset size.
3. **Error Correction**
   * Common errors such as inconsistent formats, typographical errors, and invalid data entries were identified and corrected.
   * Examples include standardizing date formats and ensuring numeric fields contained valid numbers.
4. **Deleting Empty Rows**
   * Rows with no meaningful data (completely empty rows) were removed from the dataset. This helped in optimizing storage and ensuring only relevant data was analyzed.

This cleaning process enhanced the quality of the data, ensuring that subsequent analyses and visualizations were based on consistent, accurate, and complete information.

# **GENERAL INTRODUCTION:**

The COVID-19 pandemic, which began in late 2019, has had an unprecedented impact on global health, economies, and everyday life. As governments, healthcare systems, and researchers worked tirelessly to mitigate the virus's spread and impact, data-driven approaches emerged as critical tools for guiding public health decisions, understanding the dynamics of the disease, and measuring the effectiveness of various interventions. The city of Chicago, like many other urban centers, faced significant challenges during this time, with the virus affecting diverse populations across different neighborhoods, age groups, and socioeconomic backgrounds.

This project aims to analyze and interpret key datasets related to COVID-19 in the city of Chicago, spanning from January 2020 to December 2022. The analysis focuses on several important factors, including the rates of hospitalization, mortality, COVID-19 testing, vaccination uptake, and booster doses across different demographic groups and regions. By leveraging these datasets, the project seeks to identify trends and patterns in how the virus impacted the population, the effectiveness of health interventions, and the disparities that arose based on race, age, and geographic location.

The purpose of this report is to provide a comprehensive understanding of COVID-19's impact on the city of Chicago, drawing insights from data to guide future public health strategies. These insights will contribute to the ongoing efforts to improve healthcare responses to future pandemics, as well as help policymakers and health officials design more effective, equitable public health programs that address both immediate and long-term needs.

Through this project, we will also highlight the importance of data-driven decision-making in public health, especially in urban environments, and suggest strategies for strengthening healthcare systems, reducing health disparities, and ensuring equitable access to medical resources and treatments.

# **INSIGHTS AND FINDINGS:**

1. **Hospitalization and Death Rate Trends in Chicago (Jan 2020 - Dec 2022)**

* **Hypothesis:** A relatively low hospitalization rate and death rate in the city of Chicago from January 2020 to December 2022 are attributed to the decreasing number of COVID-19 cases by the end of 2022.
* **Insights:**
  + Total Counts December 2021 shows significant counts across all categories, with positive cases, hospitalizations, and deaths each reaching notable totals.
  + Patterned Correlation: The data reveals a strong correlation among positive cases, hospitalizations, and deaths, even though they follow a staggered pattern.
  + Lagged Sequence: The peak in positive cases is followed by a delayed rise in hospitalizations and, subsequently, deaths, illustrating the typical progression and impact timeline of the outbreak on healthcare and mortality.
* **Visualization:**

A screenshot of a graph

Description automatically generated

* **Dataset:** In the city of Chicago COVID-19 Dataset (Jan 2020 - Dec 2022)
* **Number of Records:** 5150

1. **Testing Rates Among Different Age Groups in the city of Chicago (Jan 2021 - Dec 2021)**

* **Hypothesis:** Individuals aged 18-29 years in the city of Chicago exhibit the highest testing rates due to their large population base and increased testing frequency, particularly on Mondays.
* **Insight:**
  + Weekly Testing Trend: Testing rates are highest on Monday (17.71%) and remain relatively high through Wednesday, with a notable dip on Sunday (7.62%), indicating increased testing at the start of the week.
  + Testing by Age: Highest testing counts are among those aged 18-29, followed by 30-39, with significantly lower testing in older age groups
  + Testing by Race: Variation in testing rates among racial groups suggests different levels of engagement or access to testing across racial demographics.
* **Visualization:**

A screenshot of a computer

Description automatically generated

* **Dataset:** In the city of Chicago COVID-19 Testing Dataset (Jan 2021 - Dec 2021)
* **Number of Records:** 127

1. **Vaccination Doses in the city of Chicago (Jan 2020 - Dec 2022)**

* **Hypothesis:** The highest number of COVID-19 vaccination doses administered in the city of Chicago occurred in December 2022 due to the cumulative effect of widespread testing and vaccination campaigns during that period.
* **Insight:**
  + Dose Funnel: Drop-off in uptake from the first dose to the second and booster doses highlights a decrease in vaccine engagement over time. Regional Variations: The vaccination map shows certain regions with higher vaccination rates than others, indicating possible disparities in access.
  + Vaccination Trend: The cumulative trend shows an initial surge in first doses, followed by slower increases in second and booster doses, suggesting challenges in sustaining vaccination rates.
* **Visualization:**

A screenshot of a computer

Description automatically generated

* **Dataset:** In the city of Chicago COVID-19 Vaccination Dataset (Jan 2020 - Dec 2022)
* **Number of Records:** 43,070

1. **Booster Dose Rates by Race/Ethnicity in the city of Chicago (Jan 2020 - Dec 2022)**

* **Hypothesis:** White non-Latinx individuals in the city of Chicago have the highest average rate of receiving booster doses due to their higher population proportion and vaccination uptake.
* **Insight:**
  + The chart indicates that the White non-Latinx group has the highest booster dose rate relative to their population size, while Latinx and Black non-Latinx populations show lower booster uptake.
  + Asian non-Latinx and Other non-Latinx groups have relatively small population sizes and booster rates, indicating disparities in vaccine follow-through across racial groups.
* **Visualization:**

A screenshot of a computer

Description automatically generated

* **Dataset:** In the city of Chicago COVID-19 Data (Jan 2020 - Dec 2022)
* **Number of Records:** 730

1. **Regional Variations in First-Dose Vaccinations in the city of Chicago (2021-2022)**

* **Hypothesis:** The North/Central region in the city of Chicago experienced the highest average first-dose vaccination rates in late 2021 and early 2022, particularly in October, November, and December.
* **Insight:**
  + We observed that Regional Variation in First Dose Uptake: Areas like the Near South and North/Central zones have higher first-dose averages, while regions like the Far South and Southwest show lower averages, indicating uneven distribution of initial vaccine doses across the city.
* **Visualization:**

A screenshot of a graph

Description automatically generated

* **Dataset:** In the city of Chicago COVID-19 Data (Jan 2021 - Dec 2022)
* **Number of Records:** 5040

# **CONCLUSION:**

The analysis of COVID-19 data in the city of Chicago from January 2020 to December 2022 provides key insights into the pandemic's impact on hospitalization, mortality, testing, and vaccination. The decline in COVID-19 cases by December 2022 led to lower hospitalization (54.38%) and death rates (6.95%), while individuals aged 18-29 exhibited the highest testing rates. December 2022 saw the peak in total vaccine doses administered, and White non-Latinx individuals had the highest booster dose rates. Regional disparities in vaccination uptake, particularly in the North/Central region, highlight the need for targeted strategies.

To better prepare for future health crises, the following recommendations are crucial:

1. **Strengthen Public Health Education**: Implement school awareness programs and broad public awareness campaigns to promote health literacy.
2. **Invest in Vaccine and Therapeutic Research**: Continue research on next-generation vaccines and therapeutics to stay ahead of emerging variants.
3. **Build Resilient Healthcare Systems**: Ensure hospitals are equipped to handle surges, with adequate ICU beds, ventilators, and trained personnel.
4. **Address Socioeconomic Determinants of Health**: Focus on reducing health inequalities, ensuring access to healthcare for all.
5. **Enhance Quarantine and Isolation Protocols**: Improve protocols to manage future pandemics effectively.
6. **Promote Global Health Equity**: Ensure equitable access to vaccines and healthcare globally, particularly in low-income countries.

These strategies will help build a more resilient and equitable public health system, ensuring better preparedness for future pandemics.

# **REFERENCES:**

1. U.S. Government Website: <https://catalog.data.gov/dataset/>
2. COVID-19 Vaccinations by Age and Race-Ethnicity (Historical CSV).

Retrieved from: <https://catalog.data.gov/dataset/covid-19-vaccinations-by-age-and-race-ethnicity>

1. COVID-19 Daily Testing by Person (Historical CSV).

Retrieved from: <https://catalog.data.gov/dataset/covid-19-daily-testing-by-person>

1. COVID-19 Vaccinations by ZIP Code (Historical CSV).

Retrieved from: <https://catalog.data.gov/dataset/covid-19-vaccinations-by-zip-code>

1. COVID-19 Daily Rolling Average Case, Death, and Hospitalization Rates (Historical CSV).

Retrieved from: <https://catalog.data.gov/dataset/covid-19-daily-rolling-average-case-and-death-rates>

1. COVID-19 Vaccinations by Region, Age, and Race-Ethnicity (Historical CSV).

Retrieved from: <https://catalog.data.gov/dataset/covid-19-vaccinations-by-region-age-and-race-ethnicity>