

## **DATA FELLOWSHIP PROJECT**

Project Title: COVID-19 National Healthcare Safety Network (NHSN) Data Analysis

Project Brief:

Project Overview:

The goal of this project is to perform a comprehensive analysis of the COVID-19 National Healthcare Safety Network (NHSN) data. The data is available in four zip files representing the years 2020, 2021, 2022, and 2023.

Objectives:

- Extract and import the data from the zip files into a SQL database.
- Clean and transform the data to ensure consistency and accuracy.
- Conduct exploratory data analysis to identify trends, patterns, and insights.
- Perform statistical analysis to understand the impact of COVID-19 on healthcare facilities.
- Generate meaningful visualizations and reports to present the findings to stakeholders.
- Identify potential areas for improvement in healthcare practices based on the analysis.

Data Sources:

The data will be sourced from the following URLs:

2020 data: [https://download.cms.gov/covid\\_nhsn/faclevel\\_2020.zip](https://download.cms.gov/covid_nhsn/faclevel_2020.zip)

2021 data: [https://download.cms.gov/covid\\_nhsn/faclevel\\_2021.zip](https://download.cms.gov/covid_nhsn/faclevel_2021.zip)

2022 data: [https://download.cms.gov/covid\\_nhsn/faclevel\\_2022.zip](https://download.cms.gov/covid_nhsn/faclevel_2022.zip)

2023 data: [https://download.cms.gov/covid\\_nhsn/faclevel\\_2023.zip](https://download.cms.gov/covid_nhsn/faclevel_2023.zip)

Methodology:

- Extract the zip files and import the data into a SQL database.
- Perform data cleaning and preprocessing to address missing values, duplicates, and inconsistencies.
- Explore the data using SQL queries to identify key metrics, trends, and patterns.
- Conduct statistical analysis to understand the impact of COVID-19 on different healthcare facilities.
- Utilize SQL functions, aggregations, and joins to derive meaningful insights from the data.
- Create visualizations and reports using SQL or export the data to external tools for visualization.

### Questions:

1. Which healthcare facilities had the highest average number of daily COVID-19 cases in 2021? Display the top 10 facilities.
2. Calculate the 7-day moving average of new COVID-19 cases for each healthcare facility. Which facility had the highest peak in the moving average?
3. Determine the total number of COVID-19 cases, deaths, and recoveries for each state. Include the state's name and the corresponding counts in the result.
4. Find the top 5 states with the highest mortality rate (deaths per COVID-19 case) in 2022.
5. Identify the healthcare facilities that experienced a significant increase in COVID-19 cases from 2020 to 2021 (more than 50% increase). Display the facility names and the percentage increase.
6. Calculate the average length of hospital stay for COVID-19 patients.
7. Identify the top 3 states with the highest overall COVID-19 testing rates (tests per 1000 people) in 2023.
8. What is the total number of COVID-19 cases, deaths, and recoveries recorded in the dataset?
9. Find the healthcare facilities that had a consistent increase in COVID-19 cases for at least 5 consecutive months. Display the facility names and the corresponding months.
10. Calculate the mortality rate (deaths per COVID-19 case) for each healthcare facility.
11. Are there any significant differences in COVID-19 outcomes based on the type of healthcare facility (e.g., hospital, nursing home)?
12. How has the number of COVID-19 cases evolved over time (monthly or quarterly)?
13. What is the distribution of healthcare facilities by state?
14. Find the healthcare facilities with the highest occupancy rates for COVID-19 patients.
15. Find the healthcare facilities with the highest occupancy rates for COVID-19 patients.

### Deliverables:

- SQL database containing the cleaned and transformed data.
- Exploratory data analysis report highlighting key findings and insights.
- Statistical analysis report showcasing the impact of COVID-19 on healthcare facilities.
- Visualizations and dashboards presenting the analysis results.
- Documentation of the data cleaning, preprocessing, and analysis steps.
- Presentation slides for stakeholders summarizing the project and its outcomes.

### Suggested Timeline:

Day 1: Data extraction and import

Focus on quickly extracting and importing the data from the zip files into the SQL database.

Days 2-3: Data cleaning and preprocessing

Perform essential data cleaning tasks, such as addressing missing values, duplicates, and inconsistencies.

#### Days 4-5: Exploratory data analysis

Conduct high-level exploratory data analysis, focusing on identifying key trends, patterns, and initial insights.

#### Day 6: Statistical analysis

Perform basic statistical analysis, such as calculating averages, totals, or simple correlations, to gain initial insights into the data

.

#### Day 7: Visualization and reporting

Create basic visualizations (e.g., bar charts, line charts) to present the key findings from the exploratory and statistical analysis.

Prepare a brief summary report outlining the major observations and insights.