**Project Semester January–April 2025**

**DATA SCIENCE MINOR PROJECT REPORT**

**ON**

**A DASHBOARD TO ANALYSE SALES AND PROFIT**

**INRODUCTION TO DATA MANAGEMENT**

**COURSE CODE: INT217**

1. **TECH COMPUTER SCIENCE AND ENGINEERING**

****

**LOVELY PROFESSIONAL UNIVERSITY**

**PHAGWARA, PUNJAB**

**PROJECT SUBMITTED BY:**

**Ragini Yadav (12311227)**

**Section: K23GR**

**Roll No.: 30**

**PROJECT SUBMITTED TO:**

**Ms. Ashu (23631)**

**DECLARATION**

I, **Ragini**, student of B.Tech – Computer Science and Engineering (Section K23GR) at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report titled:

**“A DASHBOARD TO ANALYSE COVID-19 CASES, HOSPITALIZATION AND DEATH”**

is based on my own intensive work and is genuine. The content of this report has not been submitted to any other university or institution for the award of any degree or diploma.

**Date:**17-04-2025  
**Regno:**12311227  
**Name:** Ragini Yadav

**CERTIFICATE**

This is to certify that **Ms. Ragini Yadav,** bearing Registration No. **12311227**, has successfully completed the **INT217** – Introduction to Data Management project titled:

**“A DASHBOARD TO ANALYSE COVID-19 CASES, HOSPITALIZATION AND DEATH”**

under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort, and study. This project has been carried out as a part of the curriculum prescribed by Lovely Professional University, Phagwara for the Project Semester **January–April 2025.**

**Name:   
Ms. Ashu**

**ACKNOWLEDGEMENT**

I sincerely thank Ms.Ashu, Assistant Professor, for her guidance and support throughout this project. I also thank the faculty of the CSE Department at Lovely Professional University for providing the necessary resources and assistance.

**Ragini   
Reg. No.: 12311227**

**INTRODUCTION:**

This project provides a comprehensive analysis of COVID-19 trends using historical data from New York City. The main objective of this project is to track the progression and impact of the COVID-19 pandemic by monitoring daily case counts, hospitalizations, and deaths. Using Excel's data processing and visualization tools, we created an interactive dashboard that presents time-series data and region-wise comparisons, helping stakeholders understand the effectiveness of public health responses and the evolving situation. The dashboard serves as a powerful analytical tool to assess outbreak patterns and guide decisions on public health measures.

**SOURCE OF DATASET:**

<https://catalog.data.gov/dataset/covid-19-daily-counts-of-cases-hospitalizations-and-deaths>

**LINKEDIN LINK :**

<https://www.linkedin.com/posts/ragini15_excel-dataanalysis-covid19-activity-7318482709280890880-oD56?utm_source=share&utm_medium=member_android&rcm=ACoAAEfooJYB_HDt33u-nUZpbieDy2IVMpGPR80>

**DATASET PRE-PROCESSING:**

The data pre-processing stage included:

The data pre-processing stage was critical to ensure the accuracy and reliability of the analysis. The key steps involved were:

**Data Cleaning:** The dataset was first examined for missing, null, or irrelevant values. Inconsistent labels or entries, such as misspelled borough names or undefined values in the hospitalization or death count columns, were corrected or removed. This step ensured that only valid and meaningful records were retained.

**Data Formatting:** To support numerical and temporal analysis, appropriate data types were assigned. Dates were converted into consistent formats (e.g., DD-MM-YYYY), and case, death, and hospitalization columns were formatted as numeric data. This allowed for calculations, sorting, and filtering based on these fields.

**Date Standardization:** Date entries were standardized to allow chronological sorting and time-series analysis. This made it possible to observe daily trends and calculate metrics over time intervals.

**Categorization and Grouping:** The dataset contained borough-wise counts (e.g., Bronx, Brooklyn, Manhattan). These columns were grouped together logically to perform regional comparisons. By categorizing data this way, borough-specific trends in cases, hospitalizations, and deaths were effectively analyzed.

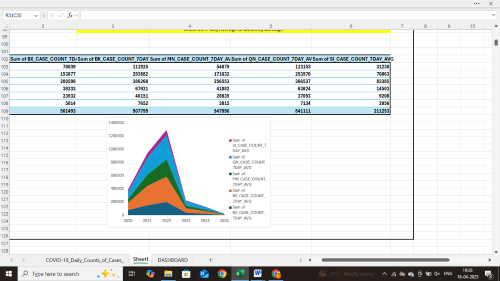
**Deriving Averages and Smoothing Trends:** 7-day rolling averages were calculated for various metrics (cases, hospitalizations, deaths). This technique smoothens daily fluctuations and highlights sustained trends, making it easier to detect outbreak waves and monitor their progression.

These preprocessing steps transformed the raw dataset into a structured, clean, and analysis-ready format, setting a solid foundation for accurate visualizations and insights.

**ANALYSIS ON DATASET**

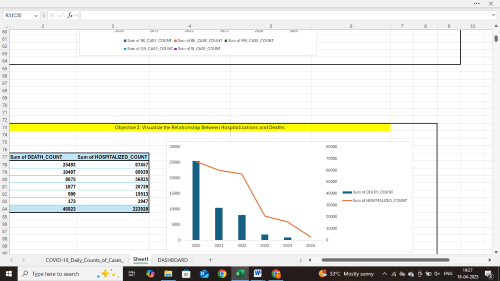
**1. Daily COVID-19 Case Trends**

* **Description:** Tracks daily confirmed and probable case counts and their 7-day averages.
* **Requirement:** Columns used: CASE\_COUNT, PROBABLE\_CASE\_COUNT, CASE\_COUNT\_7DAY\_AVG.
* **Analysis Result:** Spikes in cases align with known waves; moving averages highlight sustained trends.
* **Visualization:** Line chart showing daily and 7-day average case counts.



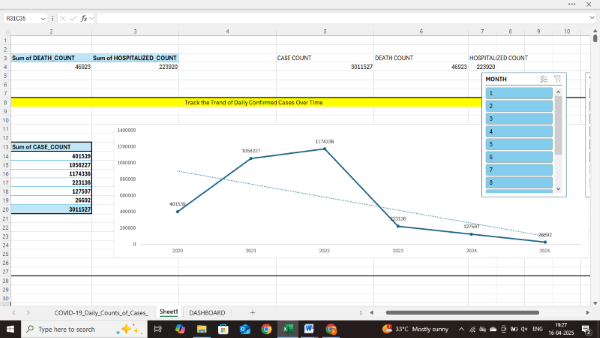
**2. Hospitalization Analysis**

* **Description:** Measures strain on healthcare by tracking COVID-related hospitalizations.
* **Requirement:** Columns used: HOSPITALIZED\_COUNT, HOSP\_COUNT\_7DAY\_AVG.
* **Analysis Result:** Hospitalizations peak slightly after case surges, reflecting lag in severe case development.
* **Visualization:** Area chart or line chart to visualize hospitalization trends.



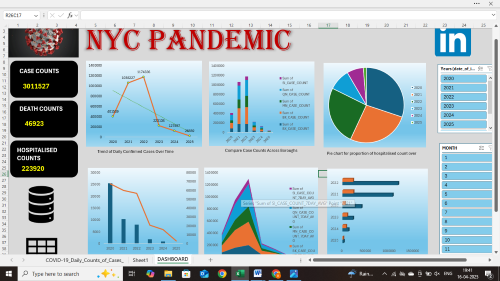
**3. Death Count and Mortality Trends**

* **Description:** Monitors fatality trends through daily and 7-day average death counts.
* **Requirement:** Columns used: DEATH\_COUNT, DEATH\_COUNT\_7DAY\_AVG.
* **Analysis Result:** Deaths peak after hospitalizations; trend helps evaluate virus severity over time.
* **Visualization:** Line chart with dual series for daily and average deaths.



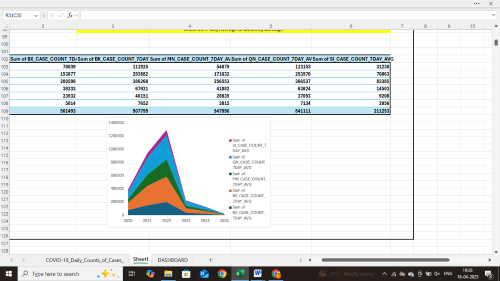
**4. Borough-Wise Comparison**

* **Description:** Compares case, hospitalization, and death counts across boroughs (Bronx, Manhattan, etc.).
* **Requirement:** Columns like BX\_CASE\_COUNT, BK\_CASE\_COUNT, etc.
* **Analysis Result:** Boroughs like Queens and Brooklyn showed higher case volumes.
* **Visualization:** Bar chart for region-wise metrics and stacked area charts for borough trend comparison.



**5. 7-Day Rolling Averages**

* **Description:** Smoothing time-series data using rolling averages to observe long-term patterns.
* **Requirement:** Use of 7-day average columns like CASE\_COUNT\_7DAY\_AVG.
* **Analysis Result:** Helps identify pandemic waves clearly by minimizing day-to-day variation.
* **Visualization:** Line charts combining daily counts and their 7-day averages.



**CONCLUSION:**

This project successfully demonstrates how Excel dashboards can be leveraged to monitor and analyze a public health crisis like COVID-19. By tracking daily cases, hospitalizations, and deaths, and incorporating 7-day averages and borough-level analysis, the dashboard offers a comprehensive view of pandemic trends. These insights enable authorities and the public to better understand the disease's spread and impact, facilitating timely and data-driven responses.

**FUTURE SCOPE:**

In future work, the dashboard can be enhanced by integrating real-time data sources such as public health APIs or cloud databases. Transitioning to tools like Power BI or Python-based dashboards can offer more advanced interactivity, predictive modeling (e.g., forecasting case spikes), and geospatial mapping. This project can also be extended to analyze vaccine distribution, testing rates, and policy impact assessments.

**REFERENCES:**

* NYC Health COVID-19 Data
* Microsoft Excel Documentation
* CDC COVID-19 Guidelines
* Towards Data Science (Time Series Analysis)