**The Evolution of COVID-19 in New York State: A Data-Driven Analysis**

**Group Members:**

Bhoomi Parikh

L. Sameera Velama

Matiullah Mati

Ramya P. Jandhyala

Vikramaditya Devarakonda

**Group - 7**

**Executive Summary**

The COVID-19 pandemic significantly affected New York State, a densely populated region that became one of the earliest epicenters in the United States. This project aims to understand the trajectory of the pandemic by analyzing hospitalization rates, fatalities, vaccination efforts, and healthcare infrastructure across key counties, including Kings, Queens, and New York County. The data used in this analysis were sourced from the New York State Health Department and cover critical periods, from the initial outbreak in 2020 to subsequent vaccination rollouts.

This project explores five hypotheses through detailed statistical analysis and data visualization to identify patterns and correlations among these variables. The findings reveal critical insights into how age, vaccination rates, and healthcare capacity shaped outcomes during the pandemic. Specifically, individuals aged 65 and older were disproportionately affected, experiencing the highest hospitalization and fatality rates compared to other age groups. This underscores the importance of targeted public health interventions for vulnerable populations during health crises.

Vaccination campaigns emerged as a pivotal factor in reducing the burden on healthcare systems, particularly ICU bed occupancy. Counties with higher vaccination rates reported fewer newly occupied ICU beds, highlighting the efficacy of vaccines in preventing severe cases. However, the study also found that first-dose vaccination rates did not strongly predict hospitalization trends in subsequent quarters. This suggests that other factors, such as virus variants and individual health behaviors, played significant roles in driving hospital admissions.

Healthcare infrastructure, particularly the availability of staffed acute care beds, was another critical factor influencing outcomes. Facilities with more staffed beds demonstrated higher patient discharge rates, reflecting their capacity to provide timely and effective care during surges in demand. This finding emphasizes the need for robust healthcare resource planning to mitigate the impact of future pandemics.

In summary, this project comprehensively analyzes COVID-19 dynamics in New York State, offering valuable lessons for public health policymakers, healthcare administrators, and researchers. The findings emphasize prioritizing vulnerable populations, strengthening healthcare systems, and sustaining vaccination efforts to better prepare for future health emergencies. These insights contribute to understanding the pandemic’s impact and serve as a guide for enhancing resilience in the face of similar challenges.

**Index**

1. Executive Summary
2. Introduction
3. Data Overview
4. Data Processing
5. Hypotheses and Findings

* Hypothesis-1: Age and Hospitalization
* Hypothesis-2: Vaccinations and ICU Beds
* Hypothesis-3: Case Fatality Correlation
* Hypothesis-4: Vaccination Trends and Hospitalization
* Hypothesis-5: Bed Capacity and Discharges

1. Key Insights
2. Conclusion
3. References

**General Introduction**

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, represented one of the most significant public health crises of the 21st century. Its global impact has been profound, disrupting economies, overwhelming healthcare systems, and causing widespread social challenges. In the United States, New York State emerged as a critical hotspot during the early stages of the pandemic due to its dense population, interconnected transportation systems, and global prominence as a travel hub. The pandemic exposed systemic vulnerabilities in healthcare infrastructure while also demonstrating the resilience and adaptability of public health responses.

This project focuses on analyzing COVID-19 data in New York State, with a specific emphasis on metrics such as hospitalization rates, fatality counts, and vaccination efforts. By examining these factors, the study aims to provide a comprehensive understanding of the pandemic's trajectory and its implications for public health policy and healthcare planning.

Key objectives of this project include:

1. **Identifying correlations and trends**: The analysis seeks to uncover meaningful relationships between key metrics, such as the link between vaccination rates and ICU admissions or the connection between infection rates and fatalities.
2. **Highlighting the pandemic’s impact on vulnerable populations**: By examining age-specific data, particularly for individuals aged 65 and older, the project explores how different demographics were disproportionately affected.
3. **Assessing healthcare system responses**: This includes understanding the role of resource allocation, such as staffed acute care and ICU beds, in managing patient outcomes during surges.
4. **Evaluating the efficacy of vaccination campaigns**: Statistical insights are used to measure the impact of vaccination efforts in reducing severe outcomes and alleviating the strain on healthcare infrastructure.

The analysis leverages data visualization techniques to present findings in a clear and accessible manner, offering insights that are not only statistically robust but also actionable for stakeholders. By examining how New York State managed its response to COVID-19, this project sheds light on critical factors that shaped the course of the pandemic.

This study is timely and relevant, as the lessons learned from COVID-19 can inform preparedness for future public health emergencies. It underscores the importance of data-driven decision-making, the need to prioritize vulnerable populations, and the role of healthcare capacity in determining outcomes. The findings aim to contribute to a growing body of knowledge that supports more resilient healthcare systems and equitable public health strategies in the face of future crises.

**Data Description**

The analysis is based on a comprehensive dataset that provides insights into the evolution of COVID-19 in New York State. The data spans critical aspects of the pandemic, offering detailed metrics to analyze trends, correlations, and the impact of various public health measures. The dataset includes:

* **Hospitalizations**: Daily and weekly counts of hospital admissions, segmented by age groups (e.g., 0–18, 19–64, 65+ years) and categorized by periods to track trends during different waves of the pandemic.
* **Fatalities**: County-level mortality data, detailing the number of deaths associated with COVID-19, including demographic breakdowns where available.
* **Vaccinations**: Data on vaccine administration, including first dose and fully vaccinated counts, as well as cumulative totals. This includes county-wise distribution to highlight geographic disparities.
* **Healthcare Infrastructure**: Information on the availability of acute care and ICU beds, staffing levels, and the number of patient discharges, which are critical for understanding healthcare capacity during surges.

**Key Counties Analyzed**

The analysis focuses on three major counties in New York State that were heavily impacted by the pandemic:

* **Kings County (Brooklyn)**: One of the most populous counties, with diverse demographics and significant early outbreaks.
* **Queens County**: Known for its high population density and international connectivity, which contributed to the early virus spread.
* **New York County (Manhattan)**: A commercial and cultural hub that faced unique challenges due to its population density and transient population.

**Temporal Coverage**

The dataset covers the initial phase of the pandemic in 2020, capturing the emergence of cases and hospitalizations, and extends through subsequent infection waves. It also includes data from the vaccine rollout in late 2020 and 2021, highlighting the impact of vaccinations on hospitalization and fatality trends.

**Relevance of the Dataset**

This data set provides a comprehensive basis for analyzing the pandemic's effects. Including hospitalization, fatality, and vaccination data enables insights into localized trends and the effectiveness of interventions across counties and demographics.

The scope and depth of this dataset enable a detailed exploration of how the pandemic unfolded in New York State, offering valuable lessons for public health planning and crisis management.

**Data Cleaning**

The dataset underwent extensive preprocessing to ensure its accuracy, reliability, and suitability for statistical analysis and visualization. Given the complexity of the data and its diverse sources, the following steps were implemented to address potential issues and enhance the quality of the dataset:

1. **Handling Missing Values**:  
   Missing data points, particularly in hospitalization and vaccination records, posed challenges for trend analysis. To address this, interpolation techniques were applied to estimate missing values while preserving the natural flow of the data. In cases where interpolation was not feasible, county-level averages or data from similar timeframes were used to fill gaps.
2. **Standardization of Formats**:  
   The dataset contained metrics reported in varying formats, such as inconsistent date ranges and units of measurement (e.g., daily vs. weekly counts). Standardization efforts ensured uniformity, aligning all data to a consistent temporal and spatial framework. For example, vaccination rates and ICU bed availability metrics were scaled to enable direct comparisons across counties.
3. **Outlier Detection and Removal**:  
   Outliers, such as anomalously high hospitalization or fatality numbers, were identified using statistical methods (e.g., z-scores) and domain expertise. These outliers were flagged and investigated, with corrections made where appropriate, such as adjusting for known reporting errors. Records deemed invalid were excluded to avoid skewing the analysis.
4. **Normalization for Population Differences**:  
   To facilitate equitable comparisons across counties of varying population sizes, normalization was applied. Metrics such as hospitalization and fatality rates were expressed per 100,000 residents. This adjustment allowed for more meaningful analysis, highlighting trends that were independent of raw population numbers.
5. **Ensuring Temporal Consistency**:  
   Data was aligned across the same timeframes to ensure consistency in trends. For instance, hospitalization data was synchronized with vaccination and ICU bed metrics to analyze temporal relationships effectively.
6. **Data Aggregation and Segmentation**:  
   To focus on specific insights, data was aggregated into meaningful categories, such as age groups and vaccination phases. Segmentation allowed for a deeper understanding of subgroup dynamics, such as hospitalization trends among individuals aged 65 and older.

**Hypothesis 1**

COVID-19 patients aged 65+ are more likely to experience hospitalization and fatalities than patients in other age groups.

**Conclusion**: The visualizations indicate that COVID-19 patients aged 65 years and above experienced the highest rates of hospitalization and fatalities, providing empirical evidence to support the hypothesis.

A screenshot of a computer

Description automatically generated

**Hypothesis 2**

There is a correlation between the total number of COVID-19 vaccination doses administered and the number of newly occupied staffed ICU beds

**Conclusion**: The visualization indicates a correlation between the cumulative number of COVID-19 vaccination doses administered and the number of newly occupied staffed ICU beds in Queens County. Specifically, areas with higher vaccination rates tend to have lower numbers of newly occupied ICU beds due to COVID-19

A screenshot of a computer

Description automatically generated

**Hypothesis 3**

There is a positive correlation between the number of COVID-19 positive cases and the number of fatalities in a county in New York State.

**Conclusion**: There is indeed a positive correlation between the number of COVID-19 positive cases and the number of fatalities in a county in New York State. This implies that areas with a higher number of positive cases are more likely to experience a higher number of fatalities due to COVID-19

A screenshot of a graph

Description automatically generated

**Hypothesis 4**

There exists a strong and significant correlation between the number of individuals who received their first dose of COVID-19 vaccination during a given quarter in a specific location and the number of newly admitted patients with COVID-19 in the subsequent quarter in the same location.

**Conclusion**: There is no strong and significant correlation between the number of individuals who received their first dose of COVID-19 vaccination in a specific location during a given quarter and the number of newly admitted COVID-19 patients in the same location during the subsequent quarter.

A screenshot of a computer screen

Description automatically generated

**Hypothesis 5**

There is a significant relationship between the number of staffed acute care beds in a healthcare facility and the number of patient discharges.

**Conclusion**: There is a significant relationship between the number of staffed acute care beds in a healthcare facility and the number of patient discharges. The analysis suggests that facilities with more staffed beds can discharge more patients compared to those with fewer staffed beds, potentially due to their higher capacity to provide care.

A screenshot of a graph

Description automatically generated

**Insights and Findings**

The analysis of COVID-19 metrics in New York State provided several critical insights that highlight the multifaceted impact of the pandemic. These findings offer valuable lessons for improving public health responses and healthcare infrastructure during future health crises.

**1. Age-Based Vulnerability**

* Individuals aged 65 and older consistently demonstrated the highest rates of hospitalization and fatalities compared to other age groups.
* This demographic's heightened vulnerability underscores the need for targeted intervention strategies, such as prioritizing vaccinations, early treatment access, and protective measures.
* Policies aimed at safeguarding older adults, including tailored healthcare resources and outreach programs, are crucial during pandemics.

**2. Vaccination Impact**

* Vaccination efforts played a pivotal role in reducing severe outcomes, particularly ICU stress. Counties with higher vaccination rates reported fewer newly occupied ICU beds, suggesting the vaccines' effectiveness in preventing critical cases.
* However, the data indicated that vaccination rates were less predictive of overall hospitalization trends. This suggests that while vaccines reduce severe cases, hospitalization drivers may include factors like pre-existing conditions and viral variants.
* The findings highlight the importance of sustaining vaccination campaigns, especially in regions with lower uptake, to protect healthcare resources.

**3. Healthcare Infrastructure**

* A strong correlation was observed between the number of staffed acute care beds and patient discharge rates. Facilities with more staffed beds managed higher volumes of patients efficiently, reflecting their ability to provide timely care during surges.
* This emphasizes the importance of resource planning, including staffing and bed allocation, to ensure healthcare systems remain resilient during crises.
* Investment in flexible healthcare infrastructure, such as surge capacity for beds and staff, is vital for managing future outbreaks.

**4. Case-Fatality Trends**

* Higher infection rates directly correlated with increased fatalities, particularly during periods of uncontrolled spread. This reinforces the need for proactive containment measures, such as testing, contact tracing, and public awareness campaigns, to reduce transmission in high-risk zones.
* Counties experiencing surges in infections often faced simultaneous strain on healthcare resources, leading to compounded challenges in managing outcomes.
* These trends highlight the importance of early interventions to prevent outbreaks from escalating into widespread crises.

**Visual Summaries**

Visualizations played a critical role in this analysis by offering clear, accessible insights into the data trends and relationships. The graphical representations highlighted key findings and helped communicate the pandemic's impact in New York State. The following visualizations were used:

**1. Line Graphs Showing Hospitalization Trends by Age Group**

* These graphs displayed hospitalization rates for different age demographics over time.
* The steep increases during pandemic peaks revealed the disproportionate burden on individuals aged 65 and older, reinforcing their higher vulnerability to severe COVID-19 outcomes.
* Age-specific hospitalization trends also provided insights into the timing and effectiveness of interventions like lockdowns and vaccination campaigns.

**2. Scatter Plots of Vaccination Rates vs. ICU Bed Occupancy in Queens County**

* Scatter plots illustrated the inverse relationship between vaccination rates and ICU bed occupancy, particularly in Queens County.
* Counties with higher vaccination rates showed significantly lower ICU admissions, emphasizing the role of vaccines in preventing severe cases.
* These plots offered a clear visual representation of the impact of vaccination campaigns on healthcare system stress.

**3. Heatmaps Comparing Case and Fatality Rates Across Counties**

* Heatmaps provided a geographic representation of infection and mortality rates across counties.
* High-case areas often corresponded to higher fatality rates, particularly in counties with limited healthcare resources.
* This visualization made it easy to identify hotspots, aiding in understanding regional disparities in COVID-19 outcomes.

**4. Bar Charts Depicting Discharge Rates Relative to Staffed Bed Capacity**

* Bar charts illustrated the relationship between the number of staffed acute care beds and patient discharge rates.
* Facilities with more beds and adequate staffing showed higher discharge rates, highlighting the importance of healthcare capacity during surges.
* These charts underscored the need for resource allocation strategies to ensure efficient patient care.

**Conclusion**

This project provides a comprehensive analysis of COVID-19's impact on New York State, focusing on hospitalization, fatalities, and vaccination data. The findings highlight several critical insights that can guide future public health strategies and healthcare planning.

**Key Takeaways**

1. **Age-Based Risk**:
   * Older adults, particularly those aged 65 and above, faced the highest risks of severe outcomes, including hospitalization and fatalities.
   * These findings emphasize the importance of prioritizing vulnerable populations during health crises, ensuring timely vaccination and access to healthcare resources.
2. **Vaccination Efficacy**:
   * Vaccination campaigns played a crucial role in reducing ICU occupancy, demonstrating their effectiveness in preventing severe cases.
   * However, their influence on overall hospitalization trends was less pronounced, suggesting that factors such as pre-existing conditions and virus variants also play significant roles.
3. **Healthcare Infrastructure**:
   * The availability of staffed acute care and ICU beds emerged as a decisive factor in patient outcomes.
   * Facilities with better resources were able to manage patient volumes effectively, achieving higher discharge rates and better overall care delivery.

**Broader Implications**

The analysis underscores the need for **proactive and balanced healthcare policies** that address multiple aspects of pandemic management:

* **Preventative Measures**: Early interventions, such as vaccination campaigns and public health education, are critical for mitigating severe outcomes.
* **Resource Readiness**: Investments in healthcare infrastructure, including surge capacity planning and staffing, are essential to handle sudden increases in demand.
* **Equitable Access**: Ensuring that all populations, particularly those in underserved areas, have access to vaccinations, testing, and treatment is key to reducing disparities in outcomes.

**References**

**Data Sources:**

* New York State Statewide COVID-19 Hospitalizations and Beds | State of New York

<https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Hospitalizations/jw46-jpb7/data_preview>

* New York State Statewide COVID-19 Testing (Archived) | State of New York

<https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Testing-Archived/xdss-u53e/data_preview>

* New York State Statewide COVID-19 Vaccination Data by County (Archived, Initial) | State of New York

<https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Vaccination-Data/duk7-xrni/data_preview>