Analyzing the Effect of Demographic and Economic Indicators on Wave Three COVID 19 out Comes in New York State Counties

Gabriel Rutherford[1], **Gerald Farai Chigwada**[1] **and Heather Padziri**[1]

[1] *Clarkson University, Data Science Department, Potsdam, New York, United States of America*

Correspondence*:

ABSTRACT

The COVID-19 pandemic exposed and worsened existing inequalities in health outcomes across different communities. The purpose of this study is to investigate the impact of demographic and economic factors on the health outcomes during the third wave of the pandemic across New York counties. Using integrated data sets from the US Census Bureau and the New York State Department of Health hospital reporting systems, we applied k-means clustering, classifying the counties into three distinct profiles based on hospitalization patterns. The results revealed that counties with higher education and income levels experienced the highest hospitalization and ICU utilization rates, while counties characterized by blue-collar occupations showed lower hospitalization metrics. Statistical analysis using Pearson's correlation coefficients and regression modeling (linear regression, decision tree regression, and random forest regression) was conducted to examine relationships between socioeconomic variables such as income, education, occupation, unemployment, and poverty and the hospitalization metric of the Third Wave of the Pandemic. This study provides strong evidence that certain economic, occupational, and educational demographic variables are significantly associated with the hospitalized outcomes of COVID-19 in New York state counties during the 3rd wave. This research provides a foundation for data-driven health equity strategies, emphasizing the complex interplay between community characteristics and public health outcomes.

1 INTRODUCTION

The COVID-19 pandemic not only posed a critical threat to public health but it also exposed deep-rooted social and economic inequalities in the United States. In New York State, one of the hardest hit regions, these differences became especially apparent during the third wave (October 2021 – March 2022). This was when hospitalizations surged and resources such as ICU and general hospital beds were severely strained. Notably, the impact was not uniform across the state, some counties were disproportionately affected thus revealing stark differences in community vulnerability and healthcare preparedness.

Numerous studies have highlighted how demographic and economic factors such as income, education, race, and occupation have significantly influenced COVID-19 outcomes. Karmakar et al. (2021)Karmakar et al. (2021) revealed that U.S counties with higher proportions of minority and low-income residents experienced elevated COVID-19 mortality rates, emphasizing the structural vulnerabilities present across U.S. communities. Similaly, Athavale and Rahim (2021)Athavale and Rahim (2021) demonstrated differential impacts of risk factors across ethnic groups, thus underscoring the need for localized and demographically informed health interventions. Chaipitakporn et al. (2022)Chaipitakporn et al. (2022) showed how sociodemographic and environmental factors shaped pandemic dynamics during the pre-vaccination period particularly highlighting the role of socioeconomic disadvantage and air pollution. More recent studies have shifted focus from biological risk factors to social determinants of health, including economic stability, educational attainment, and occupational exposure Machado et al. (2021)(Machado et al., 2021; Chen Krieger, 2021).

Despite this growing research, fewer studies have specifically examined how these socioeconomic factors influenced hospitalization patterns during later pandemic waves, particularly in geographically diverse regions like New York State. This research gap is significant, as later waves occurred within different contexts of healthcare adaptation, vaccination availability, and evolving viral variants. New York offers a compelling case study due to its geographic and socioeconomic diversity, ranging from densely populated urban centers to rural counties. These differences shaped counties' vulnerability, healthcare access, and outcomes during the pandemic. Understanding how demographic and economic factors influenced hospitalization patterns during the third wave can help inform equitable public health strategies, guide resource allocation, and strengthen pandemic preparedness efforts in future health emergencies.

The primary objective of this study is to analyze the associations between demographic and economic indicators and COVID-19 hospitalization outcomes across New York counties during the third wave. Specifically, this research aims to:

1. Identify distinct clusters of counties based on COVID-19 hospitalization metrics using k-means clustering analysis.
2. Examine the distribution of demographic and economic indicators within each cluster.
3. Determine statistically significant correlations between socioeconomic factors and COVID-19 outcomes using multivariate regression models.
4. Assess the predictive capacity of regression models using demographic and economic variables to forecast hospitalization metrics.

This study is grounded in the social determinants of health framework, which emphasizes that health outcomes are shaped by the conditions in which people are born, grow, live, work, and age Marmot and Allen (2014)(Marmot Allen, 2014). Economic stability, education, healthcare access, neighborhood characteristics, and social context all contribute to health disparities. We hypothesized that counties with different socioeconomic profiles will demonstrate distinct patterns in COVID-19 hospitalization metrics. We anticipated that factors such as educational attainment, occupation type, income levels, and poverty rates will correlate significantly with hospitalization outcomes, though the direction and strength of these relationships may vary across county clusters.

2 METHODOLOGY

2.1 Utilized Data set Overview

The main data set used was the New York State Department of Health "New York State Statewide COVID-19 hospitalizations and beds" data setof Health (2025). The data set was originally created on October 15th 2021 and has been updated to this day. The version of the data used in this research was uploaded on March 21st 2025. The data set contains nearly 300,000 rows and 37 columns which includes features such as the date of reporting, location information, and a significant amount of numerical counts of daily hospitable intake information regarding COVID 19. The second data set used was the United States Census Bureau's American Community Survey "INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS)" U.S. Census Bureau (2021c). The data set was originally created in 2021 using data from 2017-2021, the data collected was used to create 5-year estimates which are used in this research. The version of the data used for the report was extracted on April 1, 2025, and it is the original dataset from 2021. The dataset contains 806 rows with 18 columns included features such as "Median Income" "Mean Income" and a more detailed break down of income ranges by county in New York State. The third data set used was the United States Census Bureau's American Community Survey "EDUCATIONAL ATTAINMENT."U.S. Census Bureau (2021b). The data set was originally created on 2021 using data from 2017-2021 the data collected was used to create five year estimates which are used in this research. The data set contains 1178 rows and 69 columns including features related to an individuals educational attainment sorted by county as a percentage of population as well as broken down by age groups. The forth data set used was the United States Census Bureau's American Community Survey "OCCUPATION BY CLASS OF WORKER FOR THE CIVILIAN EMPLOYED POPULATION 16 YEARS AND OVER"U.S. Census Bureau (2021d). The data set was originally created on 2021 using data from 2017-2021 the data collected was used to create five year estimates which are used in this research. The data set contains 1178 rows and 7 columns which includes features breaking down the percentage of population in a given occupational by county "Service Occupations" is an example of an area present in the dataset. The fifth data set used was the United States Census Bureau's American Community Survey "COMPARATIVE ECONOMIC CHARACTERISTICS"U.S. Census Bureau (2021a). The data set was originally created on 2021 using data from 2017-2021 the data collected was used to create five year estimates which are used in this research. The data set contains 248 rows and 146 columns including features related to a given counties economic characteristics examples of these characteristics include "Employment Percentage" broken down by age group and "Unemployment Percentage" broken down by age group.

2.2 Data Grouping and Wave Creation

The original New York State data set was broken down by county with each counties data being taken out and put into a new data set. After all the data was add to the new data sets it was then broken down into three more data sets based on the waves of COVID 19 infections in New York State. These waves were determined by analyzing the high and low points of infection present in the John Hopkins COVID 19 Dash bored for New York State Center (2023). The waves were determined as follows wave one was between March 14th 2020 till June 28th 2020, wave two was October 10th 2020 till June 1st 2021, and wave three was October 26th 2021 till March 8th 2022. The dataset were then grouped by date of reporting as to ensure that the data contained summations of any given feature on a given day for a given county. After six dataset were created one for the raw data for each wave and one for county averages present in each wave. The categorical data such as Facility Identifiers, Names, DOH Regions, Networks, and "NY Forward Region" were removed as they exist in the hierarchy that the county would also be in.

2.3 Additional Feature Selection

The fourth data set containing additional data which would be used for the analysis portion of the report also had to be prepared and features had to be selected and/or constructed. The Income data set provided the feature "Median Income" this was the only viable feature from the dataset as the rest was income bracket break down for each county which the authors felt didn't provide enough of a variance in result compared to just utilizing the Median Income feature. The Education data set provide the features "Less than high school", "High school graduate", "Some college or associate's degree", and "Bachelor's degree or higher". All of these features were developed by combining the separated age group data into one feature. The remaining features then had no additional use as they provided the same thing in a more granular state and so they were removed. The Occupations data set provided the features "Management, business, science, and arts occupations", "Service occupations", "Sales and office occupations", "Natural resources, construction, and maintenance occupations", and "Production, transportation, and material moving occupations". These features like the last data set were created by combining the separated age group data into one feature. The remaining features then had no additional use as they provided the same thing in a more granular state and so they were removed. The Economic Characteristics data set provided the features "Unemployment" and "Poverty". Though there was many other feature to choose from in this data set they existed on a different plain of granularity then the rest of the features chosen as so for the purpose of easier comparison those features were removed. The utilized features like in the case of the other data sets were created by combining the separated age group data into one feature. The remaining features then had no additional use as they provided the same thing in a more granular state and so they were removed.

2.4 Clustering

The main New York State data set also had to under go feature reduction as evident when clustering was originally attempted. The number of clusters was determined for this data set through a scree plot and by using the elbow method it was determined that three was optimal. After the initial attempt with all features in the State data set many features were removed as they didn't contribute meaningfully to the clusters and the final seven features were selected as they all contributed meaningfully to the clustering. The seven features selected are "Patients Admitted Due to COVID", "Patients Admitted Not Due to COVID", "Patients Newly Admitted", "Patients Discharged", "Patients Currently in ICU", "Patients Currently ICU Intubated", and "Patients Expired". The clustering method was K-Means and the data was scaled the result was 13 counties present in cluster one, 20 in two, and 24 in three. The broad results of the clustering indicated that cluster two had the highest level of every feature, cluster 1 had the lowest of every feature, and cluster 3 was present in between the other two.

2.5 Significance Testing

After the dataset were all combined into one with all key features present from the original State data set the other comparison features were added and the three dataset were created through division by cluster. The key features in each cluster were tested for both significances and to ensure they were normally distributed. The significances testing was processed using Pearson Correlation testing and the normality testing was done using Anderson Darling as the data sets were large and so other methods were not viable.

2.6 Liner Regression Models

Liner Regression models were then trained and tested for each key feature while utilizing both select comparison features determined through the results of significance testing and all comparison features. The random state utilized was 42 and the test size was 20 percent of the dataset. The scoring metric used in the report was decided to be R-Squared and Mean Squared Error as both imply the quality of a given linear regression model. The additional component of feature model significance was included and both F-Scores and P-Values were recorded for each model and feature.

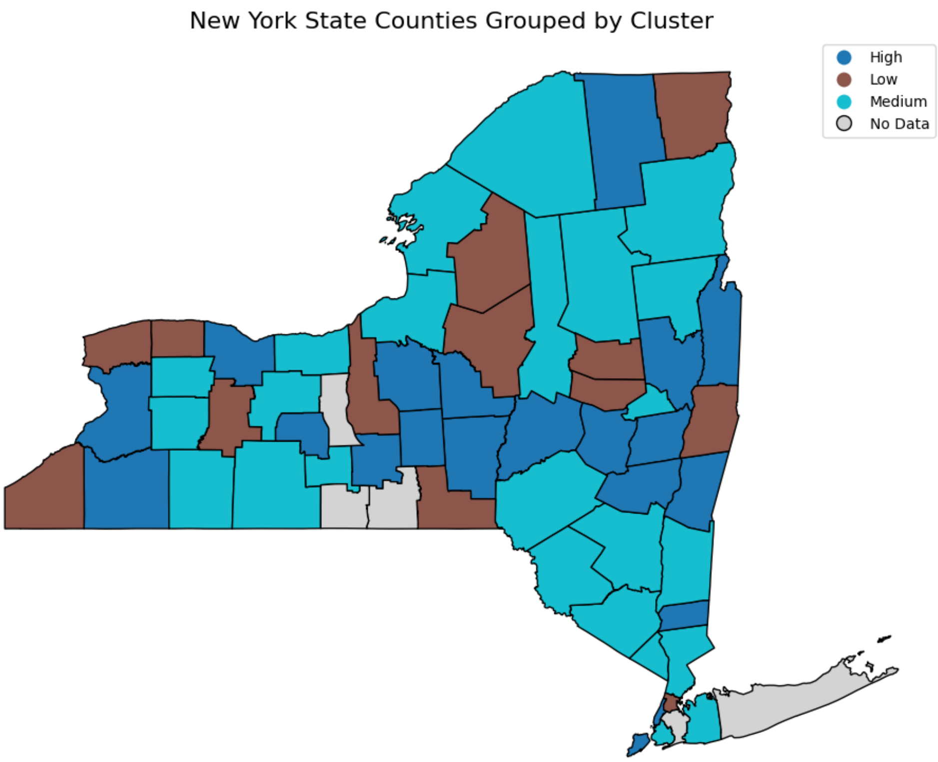

Figure 1. Map of New York State by County With Clusters Filled in

2.7 Decision Tree Regression Models

Decision Tree Regression models were then trained and tested for each key feature while utilizing both select comparison features determined through the results of significance testing and all comparison features. The random state utilized was 42 and the test size was 20 percent of the dataset the max depth of the tree was determined to be best at 6. The scoring metric used in the report was decided to be R-Squared and Mean Squared Error as both imply the quality of a given linear regression model. The additional component of feature model significance was included and both F-Scores and P-Values were recorded for each model and feature.

2.8 Random Forest Regression Models

Random Forest Regression models were then trained and tested for each key feature while utilizing both select comparison features determined through the results of significance testing and all comparison features. The random state utilized was 42 and the test size was 20 percent of the dataset the max depth of the forest was determined to be best at 7. The scoring metric used in the report was decided to be R-Squared and Mean Squared Error as both imply the quality of a given linear regression model. The additional component of feature model significance was included and both F-Scores and P-Values were recorded for each model and feature.

3 RESULTS

The analysis of hospital admission data during the third wave of COVID-19 revealed several insights into the distribution and clustering of patient statistics across counties, as well as their correlation with demographic

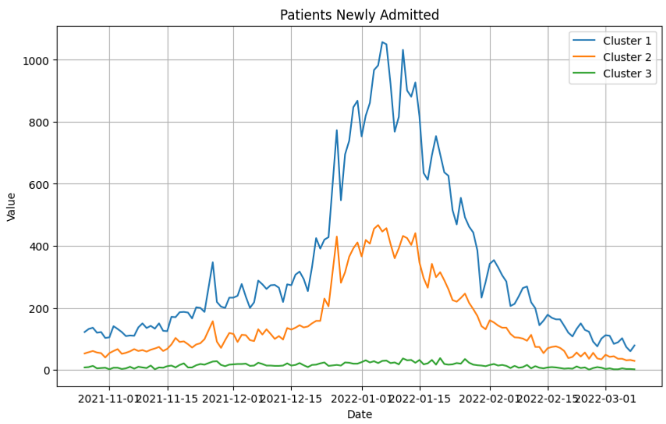

Figure 2. Newly Admitted COVID-19 Patients

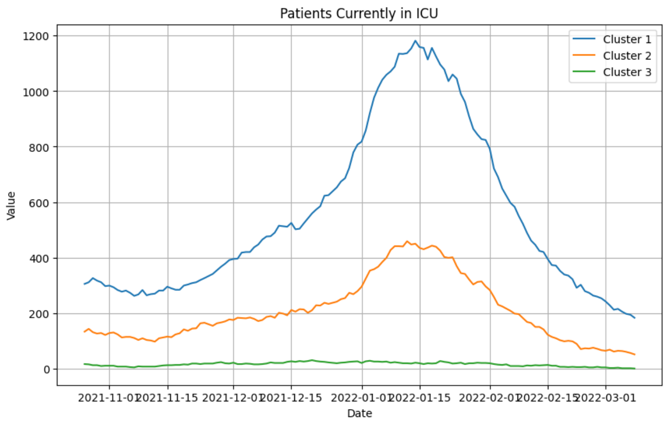

Figure 3. Patients Currently In ICU

and economic indicators. All features examined were confirmed to follow a normal distribution using the Anderson-Darling test at a 0.05 significance level.

3.1 Distribution of COVID 19 in hospitals within NYS counties

The daily average number of patients admitted to hospitals due to COVID-19 across counties had a mean of 7.15 patients per day, with a median of 6.88 and a range of 15.07. When broken down by cluster, Cluster 1 showed the lowest average at 3.5 patients per day, Cluster 2 had the highest at 10.3, and Cluster 3 had an intermediate average of 6.5 (Figure 1). For patients admitted to hospitals for reasons other than COVID-19, the mean was 4.2 patients per day. Cluster 1 again showed the lowest number (1.17), while Cluster 2 saw the highest (7.64). New COVID-19 admissions had a daily average of 2.55 patients per county, with Cluster 1 showing 1.26, Cluster 2 with 3.45, and Cluster 3 with 2.51. Discharges due to COVID-19 followed a similar trend, with an overall mean of 2.86 patients per day and Cluster 2 having the highest average at 3.87. For patients in intensive care units (ICU) due to COVID-19, the average was 3.75 across all counties. Cluster 2 again had the highest number, averaging 5.5 patients per day (Figure 2). A more severe subset, patients both in the ICU and intubated, had an average of 1.8 patients per day. Cluster 2 led this category with 2.98, while Cluster 1 had just 0.35 (Figure 3). The average number of daily COVID-related deaths across counties was low, at 0.38 patients per day. Cluster 2 again had the highest count, with 0.49 patients per day (Figure 4).

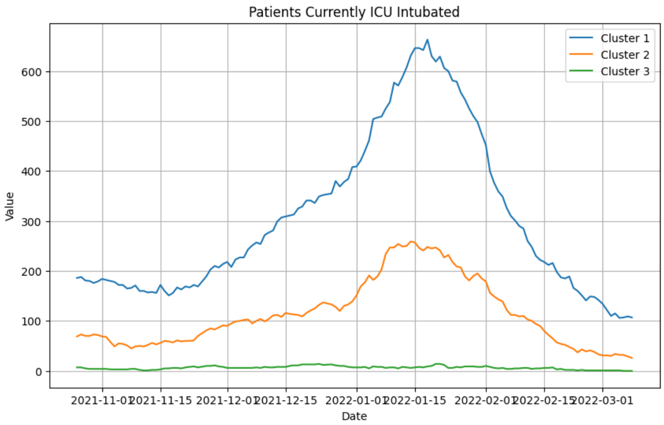

Figure 4. Patients Currently In ICU and Intubated

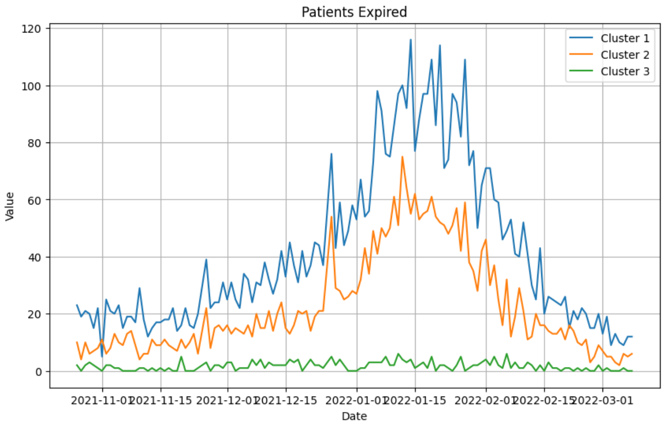

Figure 5. Number of Patients Dying

3.2 Distribution of economic and demographic education level indicators on COVID 19

Demographic and economic data, also normally distributed across counties, were analyzed to investigate their statistical significance in relation to the health data. The percentage of the population over 16 without a high school diploma remained consistent across clusters at 10 percent, with little variation in the median. High school graduates made up an average of 31 percent, with Cluster 1 slightly higher and Cluster 2 slightly lower. About 31 percent had some college education or an associate's degree, and 28 percent had a bachelor's degree or higher, with Cluster 2 again having the highest proportion of degree holders. Median household income was highest in Cluster 2, averaging 73,991 dollars, compared to Cluster 1's 66,060 dollars . Workforce composition varied slightly between clusters, with Cluster 2 having a greater percentage of individuals in management, science, and arts roles, while Cluster 1 had more representation in construction and maintenance occupations. Unemployment rates were generally consistent, averaging 3.34 percent. Poverty levels averaged 5.61 percent, again with small variations between clusters but large variation within the clusters as demonstrated by a range of 3.89 at the lowest and 7.19 at its highest..

3.3 Significance of features

Pearson correlation analysis (at a = 0.10) revealed numerous statistically significant relationships between healthcare outcomes and demographic variables. In Cluster 1, COVID-related hospital admissions negatively correlated with lower educational attainment and manual labor occupations. Non-COVID

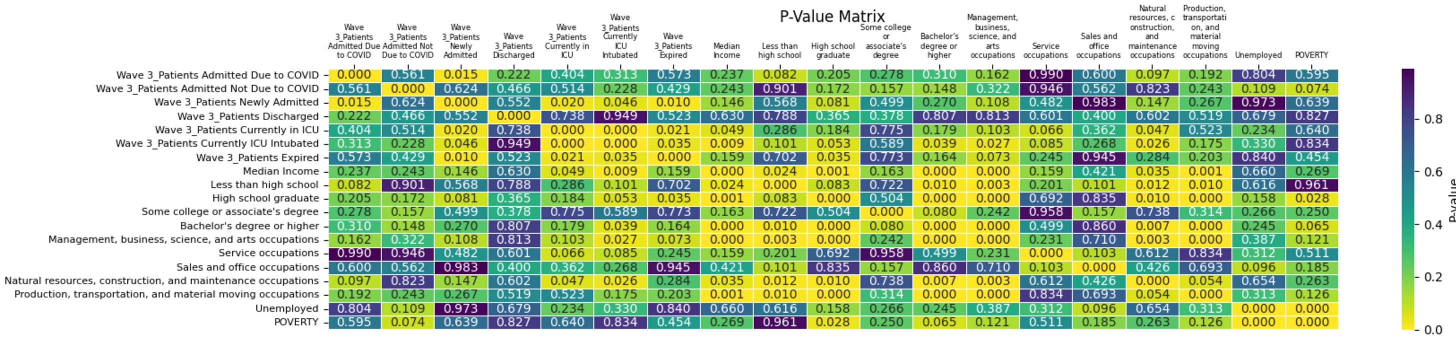

Figure 6. P-Value Matrix

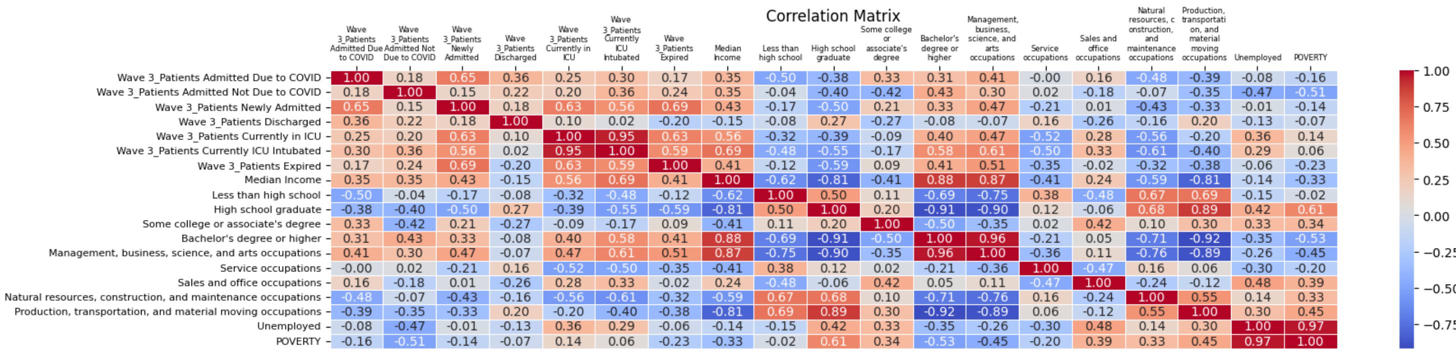

Figure 7. Correlation Matrix

admissions were significantly linked to poverty levels. New admissions were negatively correlated with high school graduation rates. ICU occupancy in Cluster 1 showed significant correlations with median income, service occupations, and manual labor occupations. Intubated ICU patients had strong positive correlations with higher education levels, income, and professional occupations, while being negatively associated with service and labor jobs. COVID-related deaths were negatively associated with high school education and positively with professional occupations. Notably, patient discharges in this cluster showed no significant demographic correlation. In Cluster 2, COVID-19 admissions were positively associated with professional occupations, and negatively with both service jobs and poverty. Non-COVID admissions were significantly related to both lower education and higher professional employment. Admissions also showed strong negative correlations with construction and maintenance roles. COVID-19 deaths in Cluster 2 were found to be significantly related to lower income, lower professional employment, and higher levels of blue-collar work and unemployment. Interestingly, no significant correlations were found for ICU and intubated patients in Cluster 2, nor for discharges. Cluster 3 revealed a dense web of significant correlations, particularly between non-COVID hospital admissions and nearly all education, occupation, and income variables. These included positive relationships with education levels from less than high school through bachelor's degrees, as well as with all occupation types, income, unemployment, and poverty. ICU admissions in Cluster 3 also showed consistent negative correlations with lower education, income, and employment metrics. However, most other patient features in this cluster, such as COVID-related admissions, ICU intubation, and deaths, did not show significant correlations with demographic indicators.

3.4 Linear Regression Results

Regression analysis largely indicated that most models trained to predict COVID-related healthcare outcomes based on demographic and economic indicators had poor predictive value. Across all three clusters, models attempting to predict general COVID-19 admissions, non-COVID admissions, discharges, and deaths resulted in low R-squared values and high mean squared error (MSE), indicating poor model fit. However, a few models from Cluster 1 provided moderate predictability. The model predicting ICU occupancy based on all COVID-related features had an R-squared of 0.40 and an MSE of 1.79. Predictions for new COVID admissions and ICU intubation were also modestly predictive, with R-squared values of 0.30 and 0.49, respectively. Models trained in Clusters 2 and 3 had overall poor predictive capacity. One exception was a model in Cluster 3 predicting ICU occupancy using a wide set of demographic features, which achieved a relatively higher R-squared of 0.11, though this was still considered weak. Service occupations were found to be a non-significant feature in this model.

Cluster	Predicted Value	Number of Features	R-Square	Mean Squared Error
1	Patients Currently in ICU	3	0.40	1.79
1	Patients Currently ICU Intubated	6	0.30	0.49
2	Patients Admitted Not Due to Covid	5	0.02	386.08
2	Patients Expired	4	0.006	0.59
3	Patients Currently in ICU	9	0.11	14.67
3	Patients Admitted Not Due to COVID	11	0.10	247.84

Table 1. Summary of Predicted Values by Cluster

3.5 Decision Tree Regression Results

Decision Tree Regression analysis largely indicated that most models trained to predict COVID-related healthcare outcomes based on demographic and economic indicators had moderate predictive value. Across all three clusters, models attempting to predict general ICU related features resulted in moderate R-squared values and low mean squared error (MSE), indicating moderate model fit. The model predicting ICU occupancy based on all COVID-related features had an R-squared of 0.54 and an MSE of 1.41. Predictions for new COVID admissions and ICU intubation were also modestly predictive, with R-squared values of 0.39 and 0.43, respectively. Models trained in Clusters 2 and 3 had overall poor predictive capacity. But, had improved scoring relative to the models created for linear regression one example of this is present in cluster 2. The model for Patients Currently ICU Intubated had an R-squared value of 0.11 and a MSE of 44.78 which is a dramatic improvement from the linear Regression Result.

Cluster	Predicted Value	Number of Features	R-Square	Mean Squared Error
1	Patients Currently in ICU	3	0.54	1.41
1	Patients Currently ICU Intubated	6	0.39	0.43
2	Patients Currently in ICU	4	0.11	44.78
2	Patients Currently ICU Intubated	3	0.13	16.26
3	Patients Currently in ICU	9	0.14	13.56
3	Patients Currently ICU Intubated	5	0.18	5.45

Table 2. Summary of Predicted Values by Cluster

3.6 Random Forest Regression Results

Regression analysis largely indicated that most models trained to predict COVID-related healthcare outcomes based on demographic and economic indicators had poor predictive value. Across all three clusters, models attempting to predict general COVID-19 admissions, non-COVID admissions, discharges, and deaths resulted in low R-squared values and moderately high mean squared error (MSE), indicating poor model fit. However, a few models from Cluster 1 provided moderate predictability. The model predicting ICU occupancy based on all COVID-related features had an R-squared of 0.19 and an MSE of 2.49. Predictions for new COVID admissions and ICU intubation with R-squared values of 0.22 and 0.56, respectively was also not to the standards of Linear Regression results or the Decision Tree results. Models trained in Clusters 2 and 3 had overall poor predictive capacity across the board in a similar manor to the linear regression results.

Cluster	Predicted Value	Number of Features	R-Square	Mean Squared Error
1	Patients Currently in ICU	3	0.19	2.49
1	Patients Currently ICU Intubated	6	0.22	0.56
2	Patients Currently in ICU	4	0.02	70.24
2	Patients Currently ICU Intubated	3	0.05	25.38
3	Patients Currently in ICU	9	0.03	22.50
3	Patients Currently ICU Intubated	5	0.07	9.01

Table 3. Summary of Predicted Values by Cluster

Overall, while many features showed statistically significant correlations with COVID-19 health outcomes, translating those findings into predictive regression models proved challenging. Only a small number of models, mostly in Cluster 1, offered modest predictive power, suggesting that while demographic and economic indicators may influence outcomes, they do not alone provide sufficient predictive capability for modeling hospital metrics during the third wave of the pandemic.

4 DISCUSSION

The results of the statistical analyzes in this study provide strong evidence that certain economic, occupational and educational demographic variables are significantly associated with the hospitalization outcomes of COVID-19 in New York state counties during wave 3 of the pandemic. The discussion in the following synthesizes the statistical patterns found in Chapters 2 and 3 to highlight key implications and interpretations.

4.1 Interpretation of Cluster Profiles

The clustering analysis revealed three distinct profiles among counties: - Cluster 1 (Low Impact): Characterized by the lowest averages in all hospital-related COVID-19 metrics. Counties in this group had relatively lower median incomes and higher prevalence of blue-collar occupations, such as natural resources and service occupations. - Cluster 2 (High Impact): Counties in this group consistently recorded the highest numbers in admissions, ICU occupancy, and patient expiration. In particular, these counties also had higher median income, higher levels of education (including bachelor's degrees), and a higher percentage of professional occupations. - Cluster 3 (Moderate Impact): Positioned between the other two clusters, with moderate values for most indicators and mixed demographic characteristics.

Socioeconomic diversity between clusters provides a valuable lens for understanding how resource allocation, local occupational risks, and community education levels influence the burden of the health system during pandemics.

4.2 Regression Insights

Regression models trained using selected features were generally weak predictors of hospitalization outcomes, with R-squared values rarely exceeding 0.30. However, some models, particularly for ICU admissions in Cluster 1, reached moderate predictive power. For instance: - The ICU model in Cluster 1 had an R-squared value of 0.40 and included median income, service occupations, and natural resources employment as statistically significant predictors. - The ICU Intubation model in Cluster 1 had an R-squared value of 0.30 and identified strong contributions from education level and management-related occupations.

This confirms that while demographic and economic factors have meaningful statistical associations, they do not fully explain the variation in health outcomes, pointing to the importance of other latent factors such as comorbidities, access to healthcare, and test rates.

5 CONCLUSION

This study explored the relationship between the health outcomes of COVID-19 and socioeconomic indicators in New York State counties during the third wave. The analysis utilized k-means clustering, Pearson correlation, and linear regression to uncover patterns and test significance.

5.1 Key findings

- Statistically distinct clusters of counties emerged based on COVID-19 hospital metrics.
- Higher-income, more educated counties showed both higher hospitalization rates and higher ICU usage, possibly due to better testing infrastructure or greater elderly populations.
- Statistically significant correlations were found between hospitalization outcomes and variables such as income, education, occupation type, and poverty rate.
- Despite the significance of many features, predictive modeling using regression techniques yielded limited accuracy, with most models performing weakly.

5.2 Recommendations

- Public health policy should integrate demographic indicators into resource distribution planning.
- Further studies should include variables like age distribution, health insurance coverage, and pre-existing conditions to improve predictive modeling.
- Investment in data infrastructure and standardized reporting practices will improve real-time pandemic response across diverse socioeconomic settings.

Overall, this study provides a foundation for data-driven health equity strategies, emphasizing the complex interplay between community characteristics and public health outcomes during a pandemic.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGMENTS

The authors of the paper would like to thank Dr. Sumona Mondal and Prof. Naveen Reddy for their guidence through our research process.

DATA AVAILABILITY STATEMENT

U.S. Census Bureau. (2021). INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS). can be found here

U.S. Census Bureau. ”EDUCATIONAL ATTAINMENT.” can be found here

U.S. Census Bureau. (2021). OCCUPATION BY CLASS OF WORKER FOR THE CIVILIAN EMPLOYED POPULATION 16 YEARS AND OVER. can be found here

U.S. Census Bureau. (2021). COMPARATIVE ECONOMIC CHARACTERISTICS. can be found here

REFERENCES

Athavale, P. and Rahim, Y. (2021). Impact of covid-19 on ethnic minorities and marginalised communities: A systematic review. *Journal of Health Disparities Research and Practice*

[Dataset] Center, J. H. C. R. (2023). New york - covid-19 overview

Chaipitakporn, P., Singh, R., and Goldstein, A. (2022). Sociodemographic and environmental determinants of covid-19 spread during the pre-vaccination period. *Environmental Research and Public Health*

Karmakar, M., Lantz, P. M., and Tipirneni, R. (2021). Association of social and demographic factors with covid-19 incidence and death rates in the us. *JAMA Network Open*

Machado, M. R., Jayaweera, D. T., and Eyal, N. (2021). Covid-19 and a retrospective look at the social determinants of health disparities. *International Journal of Health Services*

Marmot, M. and Allen, J. J. (2014). Social determinants of health equity. *American Journal of Public Health*

[Dataset] of Health, N. Y. S. D. (2025). New york state statewide covid-19 hospitalizations and beds

[Dataset] U.S. Census Bureau, A. C. S. (2021a). Comparative economic characteristics

[Dataset] U.S. Census Bureau, A. C. S. (2021b). Educational attainment.

[Dataset] U.S. Census Bureau, A. C. S. (2021c). Income in the past 12 months (in 2021 inflation-adjusted dollars)

[Dataset] U.S. Census Bureau, A. C. S. (2021d). Occupation by class of worker for the civilian employed population 16 years and over