Harnessing Data to Combat Vaccine Hesitancy and Distribution Challenges: Acme Health Care Company's Analytical Blueprint for Pandemic Preparedness

MIS-690 Capstone Project Thesis

Submitted to Grand Canyon University

Graduate Faculty of the Colangelo College of Business

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Business Analytics

by

Project Group A2

December 20, 2023

Approved by:

____ ____________________

Professor Date

Table of Contents

Harnessing Data to Combat Vaccine Hesitancy and Distribution Challenges: Acme Health Care Company's Analytical Blueprint for Pandemic Preparedness

As the world grapples with the unprecedented challenges brought about by the COVID-19 pandemic, the efficient distribution and uptake of vaccines have emerged as critical factors in controlling the virus's spread. Amidst this backdrop, the Acme Health Care Company stands as a beacon of hope in New York, having played a significant role in testing and vaccine administration during the pandemic's peak. However, with the onset of the North American winter season, a convergence of flu, RSV, and COVID-19 cases poses new challenges. The project titled "Leveraging Analytics to Optimize COVID-19 Vaccine Distribution and Uptake: A Case Study on Acme Health Care Company" delves into the complexities of these challenges, exploring how data-driven strategies can provide actionable insights to streamline vaccine distribution, enhance public awareness, and ensure an adequately trained healthcare workforce. This case study not only serves as an exemplar of the fusion of healthcare and analytics but also emphasizes the role of informed decision-making in navigating the multifaceted challenges of a global health crisis.

Business Problem Identification

The Acme Health Care Company, situated in New York, played a pivotal role during the COVID-19 pandemic by conducting essential testing and facilitating vaccine administration. However, as the winter season approaches, this organization faces numerous challenges, including the efficient distribution of vaccines, effective administration of these vaccines (particularly to individuals hesitant about vaccination), and the pressing requirement for skilled professionals to educate the community and raise awareness (Vaccine Distribution, Supply Chain, Testing Still Present Challenges in Federal Pandemic Response, n.d.-b).

Background

Describe the organizational history.

The Acme Health Care Company has established itself as an urgent care facility located in New York. Its significant contribution during the COVID-19 pandemic included conducting vital COVID-19 testing and managing vaccine administration.

Has this problem been previously encountered?

Yes, the problem of effective distribution of COVID-19 vaccines has been previously encountered. Since the onset of the pandemic, challenges linked to vaccine distribution, affordability, accessibility, and acceptability have persisted. The federal government's initiatives to boost the development and manufacturing of vaccines have unfortunately fallen short of the year-end expectations regarding distribution and administration (Vaccine Distribution, Supply Chain, Testing Still Present Challenges in Federal Pandemic Response, n.d.-c).

How long has this problem existed?

The problem has been ongoing since the inception of the COVID-19 pandemic in March 2020.

What are the potential factors that are influenced by this problem or factors that the problems influence?

The problem influences several critical factors:

- The efficient availability and distribution of vaccines.
- The willingness of the public to undergo vaccination.
- The capability of healthcare professionals to administer vaccines and educate the public.

What stakeholders can you interview to ascertain root causes of the problem?

Key stakeholders that can provide insights into the problem include:

- Healthcare Professionals: Doctors, nurses, and pharmacists.
- Public Health Experts: Public health officials, epidemiologists, and researchers.
- Community Leaders: Local community leaders such as mayors, council members, and community organizers (With Virus Season Looming and Awareness Campaign Funding Slashed, Experts See a Scattered Covid-19 Vaccine Rollout Ahead, n.d.).

Is the business problem stated in such a way that analytical tools and techniques can be applied to address it?

Yes, the business problem is articulated in a manner that facilitates the application of analytical tools and techniques, and the necessary data related to the identified business problem can be sourced (Cordes, 2023).

Can the necessary data for your identified business problem be obtained?

Yes, obtaining the required data for the identified business problem is feasible. It involves defining the problem, data collection, data analysis, and result communication (Science, 2023).

What are specific benefits to the organization if the business problem is solved using analytics?

By leveraging analytics, the company can not only address the immediate issue of vaccine distribution but also realize broader positive outcomes in public health, enhance its reputation, and solidify its competitive positioning within the healthcare sector.

Business Problem Statement

Acme Health Care Company is confronting challenges tied to the effective distribution of COVID-19 vaccines, which encompass transportation, storage, and maintaining the cold chain integrity of the vaccine. Ensuring adequately trained healthcare professionals for vaccine administration and public education is paramount for augmenting vaccine uptake, especially with the North American winter season approaching, which coincides with the flu and RSV season.

Analytics Assumptions

Identify any analytics-based assumptions involved in the scope of the project.

We operate under the assumptions that we have relatively clean data, sufficient for testing. We anticipate that there is data from previous cases, and that the forecasting models Acme Health Care Company will most likely not have missing points. The distribution of the COVID-19 vaccine is a challenge due to transportation, storage, and maintaining the vaccine's cold chain integrity. Ensuring that healthcare professionals are adequately trained in vaccine administration and public education is essential for improving vaccine uptake.

Identify the specific output measures that would be improved (e.g., profit, retention) as a result of implementing the analytics-based solution.

Acme Health Care Company aspires to enhance education in regions or demographics with suboptimal vaccine uptake. This would lead to more patients getting vaccinated, thus offering protection against COVID-19 and its variants.

Identify the cost of not implementing the analytics-based solution.

The repercussions of not adopting the analytics-based solution could be severe, with a potential increase in patients contracting COVID-19 and its variants. As immunity from the vaccines may wane over time and with the emergence of new variants, breakthrough infections might increase, complicating infection prevention (Katella, 2023). This could also result in a rise in hospitalizations and patients becoming susceptible to other illnesses. A lack of proper training for hospital staff could lead to short-staffing and increased burnout, "with as many as 60 percent to 75 percent of front-line workers reporting symptoms that can compromise patient care and safety" (Glicksman, 2022).

Identify specific inputs and how they relate to the specific output measures.

The data inputs, such as clinical data, experimental data, and prior knowledge, will help in formulating a robust model. The output would allow for potential localized outbreak forecasts based on historical data and emerging trends. These inputs are crucial in identifying regions or demographics with lower vaccine uptake, enabling targeted public education campaigns.

Describe the analytics focus area, including justification for selecting the focus area to address the specific analytics problem.

The primary focus is on data related to population demographics, geographical distribution, health infrastructure, and COVID-19 case rates. After gathering the data, it will be cleaned, errors addressed, and missing values managed. Subsequent analysis will spotlight trends, and predictive models will be developed. "Variables such as environmental factors and market conditions can be factored into the calculations to provide more comprehensive views" (Ali, 2020). These models require consistent monitoring and adjustments based on feedback.

Include information related to the ability to obtain the necessary data for your identified business problem. Be as specific as possible.

Health organizations worldwide conduct tests for the virus, and this data is relayed to local and national health departments. Hospitals provide data on COVID-19 patients, their conditions, available resources, and bed availability. Contact tracing offers insights into the spread of the virus. Demographic data provides information on those testing positive. Vaccination data provides details on vaccine recipients, the type of vaccine they received, and post-vaccination responses. The CDC, among other organizations, maintains extensive databases on cases, deaths, recoveries, and testing.

Specify data mining problem type (e.g., classification, prescriptive, prediction, or clustering).

The goal is to ensure more patients are vaccinated and informed about COVID-19 and its variants. Prescriptive analytics will be the primary approach, determining the most effective measures to curb the spread of the virus.

Analytics Problem Statement

Considering the COVID-19 pandemic, Acme Health Care Company is compelled to implement a robust analytics solution. By leveraging both historical and current data trends, the company aims to anticipate potential outbreaks, streamline vaccine distribution logistics, and identify under-vaccinated demographics for targeted education. Additionally, the company aims to gauge the preparedness of healthcare professionals for vaccine administration, emphasizing prompt training and recruitment. Adopting this analytics-driven approach, Acme Health Care Company is committed to informed decision-making, optimal resource allocation, and effectively addressing the pandemic's challenges in the upcoming Fall and Winter seasons.

Data Understanding, Acquisition, and Preprocessing for Acme Health Care Company's Vaccine Distribution Strategy

Collection of Initial Data

Acme Health Care Company is embarking on a mission to effectively distribute COVID-19 vaccines while addressing vaccine hesitancy in New York State. To navigate this complex task, we have collated datasets that provide a detailed account of the vaccination campaign's progress. This includes geographical and temporal data crucial for tracking vaccination rates and metrics such as percentages of different age groups that have completed the vaccine series.

Description of Data

In our mission to understand and improve vaccine uptake, we identified a need for comprehensive data that covers both vaccination progress and the factors influencing vaccine hesitancy. We sourced two primary datasets for this purpose:

1. The first dataset is provided by the Centers for Disease Control and Prevention (CDC), detailing vaccination data from New York between January 2021 and March 2023 (*COVID-19 Vaccinations in the United States, County | Data | Centers for Disease Control and Prevention*, n.d.) . It includes data points such as rates of initial doses and

completed series by age group, booster dose uptake, Social Vulnerability Index (SVI) categories, and population metrics based on the 2019 census.

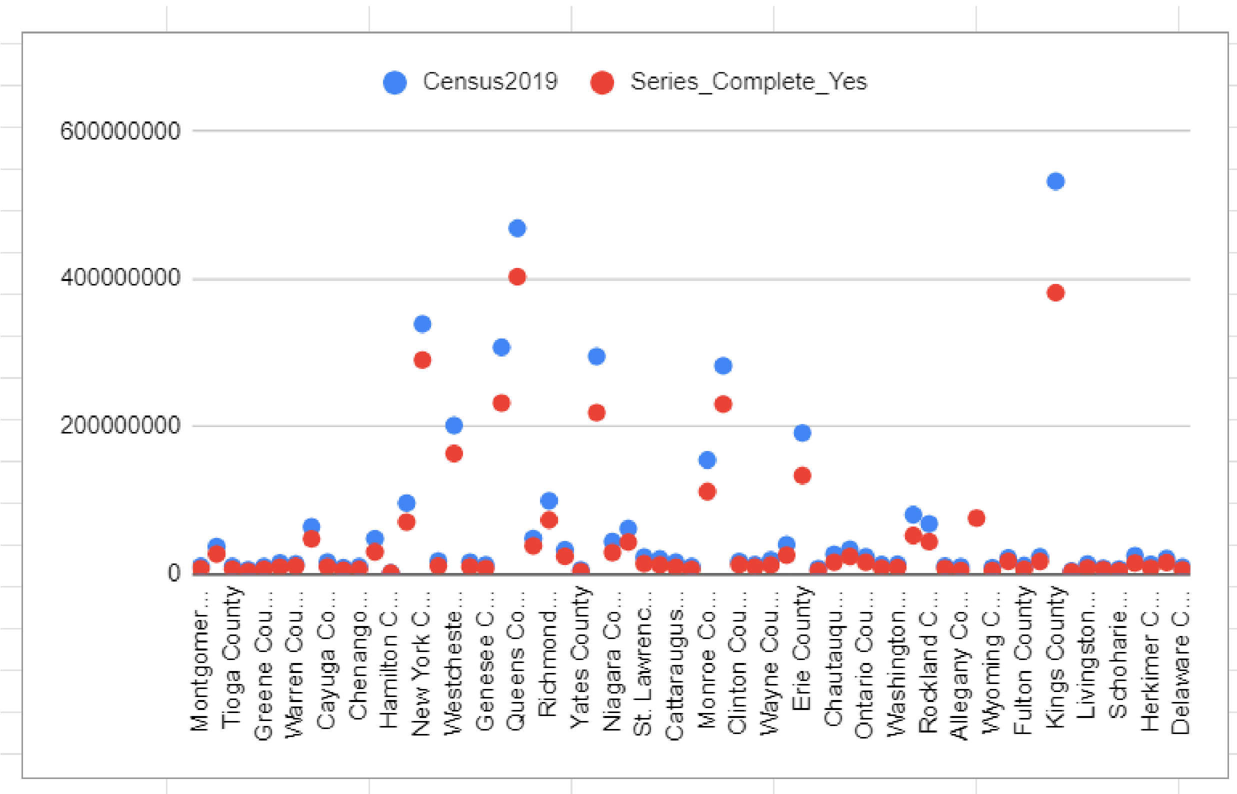
2. The second dataset, titled "COVID-19 County Hesitancy," also sourced from the CDC, focuses on the multifaceted issue of vaccine hesitancy (*Vaccine Hesitancy for COVID-19: County and Local Estimates | Data | Centers for Disease Control and Prevention*, 2021) . It includes variables such as county density and demographics, trust indices for science, media, and government, personal health and risk perception, political and educational backgrounds, and household income.

Our data comprises both continuous variables, such as index scores and income levels, and discrete variables, including the number of graduates and vaccination numbers. This blend allows us to capture a nuanced picture of the variables at play.

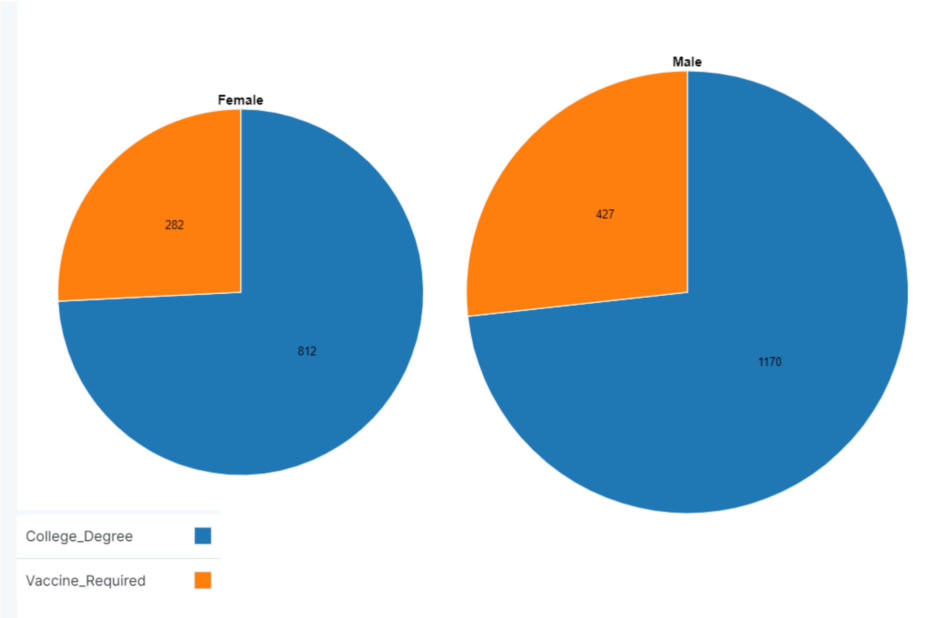
Exploration of Data

To explore these variables, we created visualizations, revealing trends like the correlation between education level and vaccine uptake. A preliminary analysis of the New York dataset showed that among 1170 male graduates, 427 had taken the vaccine, while 812 female graduates saw 282 vaccinations. These trends provide initial evidence of the relationship between education and vaccine acceptance.

The scatter box plot shows the comparison of the 2019 census to the number of completed vaccines in each New York County.



This pie chart is a comparison of a sample of college graduates who are vaccinated by gender:



Verify Data Quality

Visualization techniques have been employed to ensure data consistency and quality. In doing so, we've taken steps to avoid Simpson's paradox, ensuring that aggregations do not mask true trends and relationships within the data.

Hypothesis Statements

Based on our preliminary analysis, we propose the following hypotheses:

- Null hypothesis: Education level does not significantly affect vaccine uptake in New York State.

- Alternative hypothesis: There is a significant difference in vaccine uptake among different education levels in New York State.

The synthesized insights from these datasets are integral to tackling vaccine hesitancy and are expected to contribute significantly to a proactive public health response by Acme Health Care Company in New York State. Our strategic, data-driven approach is designed to position Acme Health Care Company as a leader in public health response.

The raw data and code utilized in our analysis will be made available in the appendix of our final report. This will allow for transparency and reproducibility of our findings.

Here is a sample of the first CDC dataset:

Date	FIPS	MMWR_week	Recip_County	Recip_Stat	Completeness_pct	Administered_Dose1_Recip	Administered_Dose1_Pop_Pct	Administered_Dose1_Recip_5Plus	Administered_Dose1_Recip_5PlusPop_Pct
3/29/2023	36057	13	Montgomery County	NY	97.5	37233	75.6	37141	80.6
3/29/2023	36111	13	Ulster County	NY	97.5	153849	86.6	152873	90
3/29/2023	36107	13	Tioga County	NY	97.5	34981	72.6	34855	76
3/29/2023	36049	13	Lewis County	NY	97.5	14763	56.1	14745	59.8
3/29/2023	36039	13	Greene County	NY	97.5	34244	72.6	34109	75.4
3/29/2023	36053	13	Madison County	NY	97.5	47462	66.9	47204	69.9
3/29/2023	36113	13	Warren County	NY	97.5	56933	89	56600	92.6
3/29/2023	36001	13	Albany County	NY	97.5	269336	88.2	266652	91.9
3/29/2023	36011	13	Cayuga County	NY	97.5	49490	64.6	49277	67.7
3/29/2023	36073	13	Orleans County	NY	97.5	24356	60.4	24277	63.3
3/29/2023	36017	13	Chenango County	NY	97.5	31719	67.2	31616	70.8
3/29/2023	36065	13	Oneida County	NY	97.5	163306	71.4	162737	75.4
3/29/2023	36041	13	Hamilton County	NY	97.5	4162	94.2	4153	95
3/29/2023	36067	13	Onondaga County	NY	97.5	377785	82	374340	86.1
3/29/2023	36061	13	New York County	NY	97.5	1913341	95	1892474	95

Here is a sample of the second CDC dataset:

A	B	C	D	E	F	G	H	I	J	K	L	M
31161	Sheridan County, Nebraska	NEBRASKA	0.0842	0.1287	0.0516	0.6	Moderate Vulnerability	0.43	Moderate Concern	0.29	0.0552	0.0673
36073	Orleans County, New York	NEW YORK	0.0851	0.1558	0.0603	0.64	High Vulnerability	0.38	Low Concern	0.453	0.0492	0.0033
36057	Montgomery County, New York	NEW YORK	0.0796	0.1489	0.0553	0.83	Very High Vulnerability	0.39	Low Concern	0.586	0.141	0.0015
36083	Rensselaer County, New York	NEW YORK	0.0678	0.1273	0.0476	0.34	Low Vulnerability	0.15	Very Low Concern	0.619	0.0493	0.0007
36075	Oswego County, New York	NEW YORK	0.0798	0.1477	0.056	0.65	High Vulnerability	0.23	Low Concern	0.516	0.0259	0.0013
36065	Oneida County, New York	NEW YORK	0.074	0.1385	0.052	0.69	High Vulnerability	0.21	Low Concern	0.548	0.0577	0.0021
36081	Queens County, New York	NEW YORK	0.051	0.1014	0.036	0.62	High Vulnerability	0.32	Low Concern	0.611	0.2804	0.0023
36063	Niagara County, New York	NEW YORK	0.0721	0.1343	0.0501	0.33	Low Vulnerability	0.2	Low Concern	0.574	0.0302	0.0098
36077	Otsego County, New York	NEW YORK	0.0722	0.1359	0.0509	0.42	Moderate Vulnerability	0.2	Very Low Concern	0.524	0.0369	0.0008
36067	Onondaga County, New York	NEW YORK	0.0655	0.1241	0.0453	0.52	Moderate Vulnerability	0.17	Very Low Concern	0.659	0.0491	0.0049
36061	New York County, New York	NEW YORK	0.0457	0.0894	0.0313	0.55	Moderate Vulnerability	0.18	Very Low Concern	0.649	0.2583	0.0012
36069	Ontario County, New York	NEW YORK	0.0663	0.1213	0.0465	0.23	Low Vulnerability	0.1	Very Low Concern	0.621	0.0474	0.0024
36079	Putnam County, New York	NEW YORK	0.0578	0.1043	0.0408	0.08	Very Low Vulnerability	0.04	Very Low Concern	0.619	0.1497	0.0009
36071	Orange County, New York	NEW YORK	0.0667	0.1251	0.0471	0.55	Moderate Vulnerability	0.2	Low Concern	0.564	0.2049	0.0023
36059	Nassau County, New York	NEW YORK	0.0477	0.0868	0.0343	0.24	Low Vulnerability	0.05	Very Low Concern	0.661	0.169	0.0012

In summary, these datasets, when combined, offer a robust framework for analyzing the factors influencing vaccine uptake. The vaccine hesitancy dataset delves into attitudes and beliefs, while the vaccination progress dataset tracks the outcomes of those attitudes in terms of behavior. This comprehensive analysis will empower Acme Health Care Company to make informed decisions to enhance vaccine distribution, tailor educational campaigns, and ensure healthcare professional training, leading to a successful pandemic response in the upcoming seasons.

Data Diagnostics and Descriptive Summary

Sampling Model

For ACME Health Care Company, our focus has been on two datasets for the project; the COVID-19 County Hesitancy dataset from the CDC (Centers for Disease Control) (Vaccine Hesitancy for COVID-19: County and Local Estimates | Data | Centers for Disease Control and Prevention, 2021) and the COVID-19 Vaccinations in the United States dataset, specifically the vaccination data from New York between January 2021 and March 2023 (COVID-19 Vaccinations in the United States, County | Data | Centers for Disease Control and Prevention, n.d.). The County Hesitancy dataset is narrowed down to the State of New York, centering on the number of estimated hesitant individuals and the corresponding percentage of vulnerability. This data was chosen because it represents the diverse landscape of vaccine hesitancy and social vulnerability across various counties in New York State. This strategic selection allows for a comprehensive understanding of these factors in vaccine distribution and public health.

Data Construction and Integration

Integrating two key variables, the estimated vaccine hesitancy percentage, and the Social Vulnerability Index (SVI) for each county in New York, the County Hesitancy dataset provides a multi-dimensional perspective. The SVI categories, ranging from "Very Low Vulnerability" to "Very High Vulnerability," offer insights into each county's socio-economic status, which is a crucial factor influencing vaccine hesitancy.

The dataset on vaccine hesitancy comprises 21 columns and 3142 rows, focusing on specific columns including State, County, Estimated Hesitant, Social Vulnerability Index (SVI), and SVI Category. The Modified New York dataset was introduced, rearranging the Social Vulnerability Index Category alphabetically and incorporating filters to facilitate searches based on specific categories. These modifications did not change the dataset's essence.

Descriptive Summary

Overview of Data: The data encompasses a total of 62 counties in New York State, detailing estimated vaccine hesitancy percentages (ranging from 5% to 9%) and SVI percentages. The Pivot Summary categorizes these counties based on their SVI ratings, providing aggregated insights into average hesitancy and SVI in different vulnerability categories, including high, low, moderate, very high, and very low vulnerability areas. The overall state average for estimated hesitancy and SVI is also highlighted. The statistical analysis reveals a slight negative skewness in hesitancy, suggesting a concentration of lower hesitancy rates, and the variance in hesitancy and SVI illustrates the diversity within the dataset.

A	B	C	D	E
NEW YORK	Orleans County, New York	0.0851	0.64	High Vulnerability
NEW YORK	Montgomery County, New York	0.0796	0.83	Very High Vulnerability
NEW YORK	Rensselaer County, New York	0.0678	0.34	Low Vulnerability
NEW YORK	Oswego County, New York	0.0798	0.65	High Vulnerability
NEW YORK	Oneida County, New York	0.074	0.69	High Vulnerability
NEW YORK	Queens County, New York	0.051	0.62	High Vulnerability
NEW YORK	Niagara County, New York	0.0721	0.33	Low Vulnerability
NEW YORK	Otsego County, New York	0.0722	0.42	Moderate Vulnerability
NEW YORK	Onondaga County, New York	0.0655	0.52	Moderate Vulnerability
NEW YORK	New York County, New York	0.0457	0.55	Moderate Vulnerability
NEW YORK	Ontario County, New York	0.0663	0.23	Low Vulnerability
NEW YORK	Putnam County, New York	0.0578	0.08	Very Low Vulnerability
NEW YORK	Orange County, New York	0.0667	0.55	Moderate Vulnerability
NEW YORK	Nassau County, New York	0.0477	0.24	Low Vulnerability

Pivot Summary Insights: The Pivot Summary categorizes counties based on their SVI ratings providing aggregated insights:

High Vulnerability: Average estimated hesitancy is 8% with an average SVI of 68%.

Low Vulnerability: Shows an average hesitancy of 7% with a lower average SVI of 33%.

Moderate Vulnerability: Here the average hesitancy is 7% with an SVI average of 49%.

Very High Vulnerability: These areas average 7% in hesitancy with a significantly higher SVI at 88%.

Very Low Vulnerability: Both estimated hesitancy and SVI are lower averaging 7% and 14% respectively.

Overall State Average: Across New York, the average estimated hesitancy is 7% with an average SVI of 47%.

Row Labels	Average of Estimated hesitant	Average of Social Vulnerability Index (SVI)
NEW YORK	**7%**	**47%**
High Vulnerability	8%	68%
Low Vulnerability	7%	33%
Moderate Vulnerability	7%	49%
Very High Vulnerability	7%	88%
Very Low Vulnerability	7%	14%
Grand Total	**7%**	**47%**

Statistical Analysis:

Kurtosis and Skewness: The data exhibits a slight negative skewness in hesitancy (-0.447) suggesting a concentration of lower hesitancy rates.

Sample Variance: The variance in hesitancy (0.00011) and SVI (0.03698) illustrates the diversity within the dataset.

Estimated hesitant		*Social Vulnerability Index (SVI)*	
Mean	0.071785484	Mean	0.468065
Standard Error	0.001353994	Standard Error	0.024423
Median	0.0722	Median	0.46
Mode	0.0722	Mode	0.39
Standard Deviation	0.010661357	Standard Deviation	0.19231
Sample Variance	0.000113665	Sample Variance	0.036983
Kurtosis	-0.199668202	Kurtosis	0.003354
Skewness	-0.446668357	Skewness	0.202433
Range	0.044	Range	0.94
Minimum	0.0457	Minimum	0.05
Maximum	0.0897	Maximum	0.99
Sum	4.4507	Sum	29.02
Count	62	Count	62

Exploratory Data Analysis

Handling Missing Data and Data Construction:

Missing Data: While the current data set displays no missing data, it is essential to be prepared for such instances. Should missing data be encountered in future datasets, robust imputation methods would be utilized. These methods could include mean or median substitution for continuous variables, and mode substitution or predictive modeling techniques for categorical variables. The choice of imputation method will depend on the nature of the data and the missingness mechanism—whether it is missing completely at random, missing at random, or missing not at random.

Derived Attributes: To deepen the analysis, new attributes could be derived. For instance, counties could be categorized into 'low', 'medium', and 'high' hesitancy levels based on their estimated hesitancy percentages. Another potential attribute could be the creation of an 'Access to Healthcare' index, derived from existing data on healthcare facilities, transportation infrastructure, and demographic factors influencing healthcare access.

Data Integration: Integrating data from diverse sources is crucial for a multi-faceted understanding of vaccine hesitancy. This could involve merging demographic data, such as age distribution, income levels, education status, and ethnic composition, with the existing dataset. Additionally, incorporating data on healthcare infrastructure, like the number of hospitals or clinics per county, could provide insights into accessibility issues that may influence vaccine hesitancy.

Syntactic Modifications:

Purpose: To enhance the compatibility of the dataset with various analytical tools, syntactic modifications are necessary. This includes standardizing formats for dates, numerical values, and categorical variables. For instance, ensuring consistent date formats (YYYY-MM-DD) across all data points, or converting categorical data into a format suitable for machine learning algorithms.

Preserving Meaning: While making these syntactic modifications, it is paramount to preserve the original meaning and integrity of the data. This involves careful documentation of all transformations and maintaining a raw version of the dataset for reference. This practice ensures that any changes made can be traced back to their original state, thus maintaining transparency and reliability in the data analysis process.

Additional Considerations: For datasets with varying scales, data normalization techniques would be applied to bring different variables onto a common scale without distorting differences in the ranges of values. Where applicable, a temporal analysis could be conducted to understand the trends and patterns in vaccine hesitancy over time. This might involve analyzing changes in hesitancy levels before and after public health campaigns or significant events related to the COVID-19 pandemic.

Trends Analysis

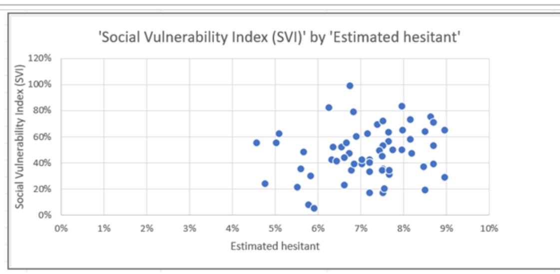
Data Segmentation and Anomaly Identification: The data was segmented based on SVI ratings, with careful attention to avoid violating Simpson's rule, ensuring that the segmentation did not obscure any underlying trends or relationships.

Anomalies: Any identified anomalies, such as unexpectedly high or low vaccine hesitancy in a particular segment, were investigated to determine their cause. They could represent data entry errors, unique local factors, or genuine deviations from the norm.

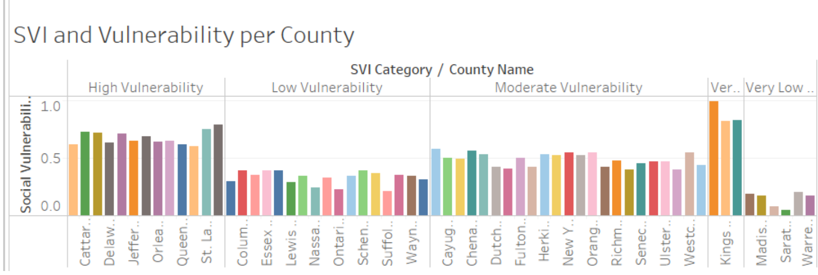
Avoiding Anomalies: Anomalies were avoided by thorough data cleaning and verification processes. If anomalies represented true outliers, they were documented and analyzed separately to understand their implications.

Descriptive Analytics

Visualization: Graphs and charts are used to visually represent the distribution of vaccine hesitancy and SVI across counties, providing an intuitive understanding of the data. Please refer to the Tableau that visualizes our work.



SVI Category	Average of Estimated hesitant	Average of Social Vulnerability Index (SVI)
High Vulnerability	8%	68%
Low Vulnerability	7%	33%
Very Low Vulnerability	7%	14%
Very High Vulnerability	7%	88%
Moderate Vulnerability	7%	49%
Grand Total	**7%**	**47%**



Verify Data Quality

Data Reliability and Problem-Solving:

Reliability Check: Cross-referencing the data with reputable sources (e.g., CDC data) ensures reliability. Consistency checks within the dataset were also performed.

Problem Resolution: Any identified problems like outliers or data entry errors were addressed by revising the dataset or applying statistical methods to mitigate their impact.

Investigating Data Quality

Verification Steps: To ensure data integrity, steps such as rigorous data profiling, data quality analysis, and data cleaning were undertaken. A scatter chart was used for data quality checks, showing no outliers that could impact the analysis significantly. The reliability of the dataset was assessed using a statistical measure (alpha), with a threshold set at 0.5. The obtained value of 0.59 indicates a robust level of reliability, suggesting the dataset's accuracy and cleanliness.

Estimated hesitant	Social Vulnerability Index (SVI)	CVAC level of concern for vaccination rollout	Total
0.09	0.64	0.38	1.11
0.08	0.83	0.39	1.30
0.07	0.34	0.15	0.56
0.08	0.65	0.23	0.96
0.07	0.69	0.21	0.97
0.05	0.62	0.32	0.99
0.07	0.33	0.2	0.60
0.07	0.42	0.2	0.69
0.07	0.52	0.17	0.76
0.05	0.55	0.18	0.78
0.07	0.23	0.1	0.40
0.06	0.08	0.04	0.18
0.07	0.55	0.2	0.82
0.05	0.24	0.05	0.34

Transformations and Discards: If variables were found to be irrelevant or redundant, they would be transformed or discarded. For instance, if a variable does not significantly impact vaccine hesitancy, it might be omitted from further analysis.

Conclusion

The comprehensive analysis of the relationship between social vulnerability and vaccine hesitancy across New York counties offers valuable insights for Acme Health Care Company's strategic planning in vaccine distribution. By integrating and examining data on estimated vaccine hesitancy and the Social Vulnerability Index (SVI) across various counties, we have identified key patterns and trends that are crucial for developing targeted public health interventions. Our findings highlight a slight negative skewness in hesitancy rates and significant variance both in hesitancy and SVI, suggesting a diverse landscape in vaccine acceptance and social vulnerability across New York State. The Pivot Summary further emphasizes these variations, showing how vaccine hesitancy and SVI interact differently across counties with varying levels of vulnerability.

These insights are not only pivotal in guiding Acme Health Care Company's efforts in combating vaccine hesitancy but also in enhancing the efficiency and effectiveness of vaccine distribution strategies. By understanding the nuanced relationship between social vulnerability and vaccine hesitancy, Acme Health Care can tailor its approach to address the specific needs and challenges of different communities, thereby improving overall public health outcomes. The thoroughness of this analysis, underpinned by rigorous data integration and descriptive analytics, ensures a robust and accurate understanding of these complex dynamics. This study serves as a foundation for our modeling, aimed at mitigating vaccine hesitancy and enhancing public health preparedness in the face of ongoing and future health crises.

Model Evaluation and Analysis

Model 1 Overview

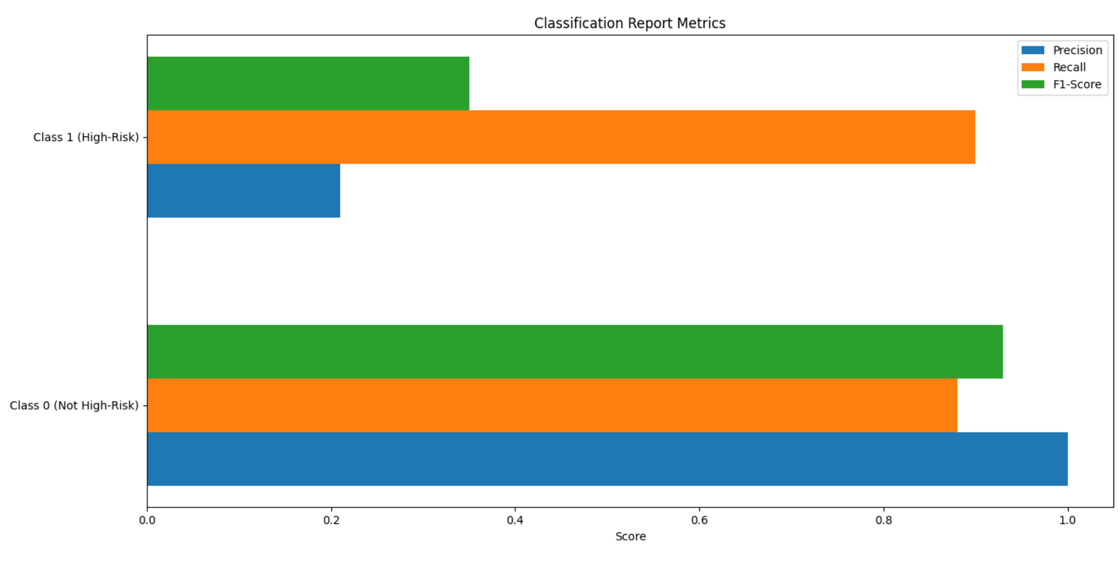
We have Predictive Analytics for Anticipating Outbreaks and Assessing Preparedness. The objective is to predict potential COVID-19 hotspots and assess healthcare readiness for vaccine administration. The dataset being used is the Testing Dataset. The metrics used are critical for identifying trending hotspots of COVID-19 cases, enabling early intervention and preparation (Huang et al., 2023). The secondary dataset is the Vaccinations Dataset. The key features are first dose recipients, the population percentage of Vaccines, and complete series data. This data helps evaluate vaccine coverage and immunity levels, which is essential for outbreak risk assessment.

Model Purpose

The logistic regression model developed for this thesis is designed to identify high-risk counties for public health planning purposes, particularly in the context of a pandemic response. The model operates by analyzing a dataset that includes key variables such as the number of new positive cases, total cases per 100k, first-dose vaccinations, and complete vaccination percentages in various counties. The primary objective is to classify each county as either high-risk or not, based on a defined threshold for new cases and vaccination rates. This classification is crucial as it informs public health officials about areas that might need increased attention or resources.

Model Performance and Evaluation

The model's performance was rigorously evaluated through a series of metrics, including cross-validation recall scores, precision, accuracy, and a confusion matrix. Notably, the model achieved an average recall score of approximately 0.81 during cross-validation, indicating its robustness in correctly identifying high-risk counties across different data samples. Adjusting the classification threshold to 0.45, the model exhibited a high recall of 0.90 for high-risk counties, demonstrating its effectiveness in capturing many potential hot spots. This high recall, however, came with a trade-off in precision (0.21), leading to a considerable number of false positives.

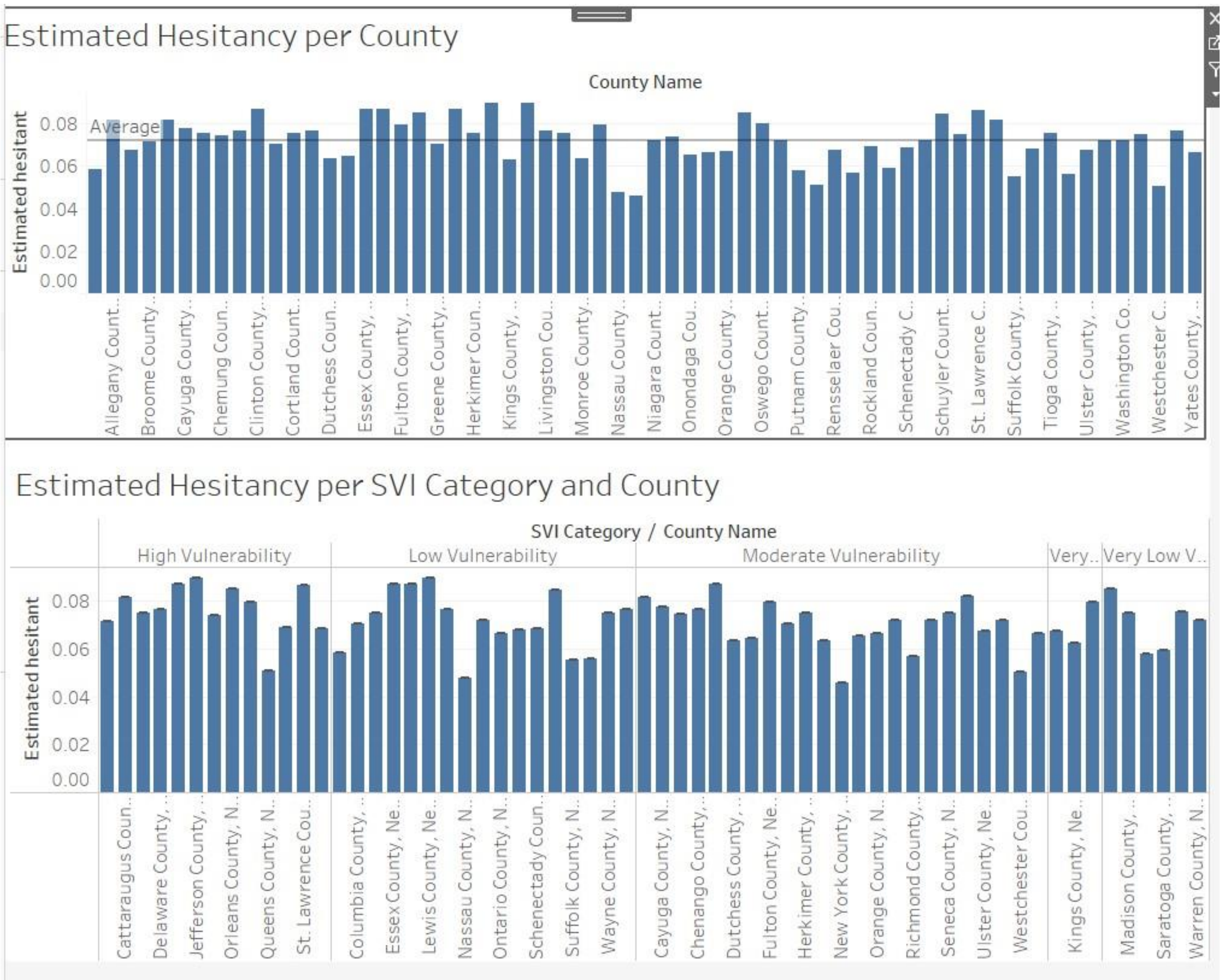


Incorporating Risk Evaluation in the Model Process

The development of our predictive model integrates risk evaluation, a pivotal aspect in public health and pandemic management. By forecasting the risk levels of different counties using current data on new positive cases and vaccination rates, the model serves as a tool for predicting potential high-risk areas. This prediction, grounded in logistic regression, is a statistical approach to estimating probabilities of outcomes, aligning with forecasting methodologies. Furthermore, the model's design incorporates optimization strategies, notably in adjusting the classification threshold (e.g., from 0.5 to 0.45) to balance precision and recall, enhancing its ability to accurately identify high-risk zones. The implementation of cross-validation and oversampling techniques further optimizes the model's robustness, ensuring its applicability to diverse and unseen data scenarios. Although not explicitly a simulation, the process of threshold adjustment and methodological fine-tuning effectively simulates various scenarios, providing insights into how the model responds to different data configurations and real-world situations. This approach underscores our commitment to a comprehensive risk evaluation framework, essential for effective public health decision-making.

Interpreting the Trade-offs and Practical Implications

In the context of public health, where the cost of overlooking a potential outbreak is significantly high, this trade-off is justifiable. A high recall ensures that the model errs on the side of caution, flagging more regions as high-risk than they might be. This approach is preferable in scenarios were failing to identify a high-risk area could lead to severe consequences, such as uncontrolled outbreaks or delayed responses. The lower precision, indicative of false positives, suggests that while some counties may be classified as high-risk unnecessarily, this approach enables a more comprehensive and preventive strategy in public health planning.



Conclusion and Future Directions

Overall, the model serves as a valuable tool for early identification of potential hot spots, enabling proactive measures in pandemic management. The current analysis opens avenues for further refinement, such as experimenting with different models or incorporating more features to improve precision without significantly compromising recall. Additionally, the model's framework could be extended to forecasting future risk levels, offering a more dynamic and predictive element to public health strategies.

In the realm of data analytics, the necessity to adapt and evolve models to suit specific objectives is paramount. While Model 1 was adept at addressing a particular set of challenges, its architecture and design were primarily tailored for classification tasks. In contrast, Model 2 diverges significantly in its purpose and construction, focusing on regression analysis. This shift was necessitated by the different nature of the problem at hand – predicting continuous outcomes rather than categorizing discrete ones. Thus, Model 2 was developed with the goal of predicting vaccination rates, a continuous variable, requiring a fundamentally different approach.

Model 2 Overview

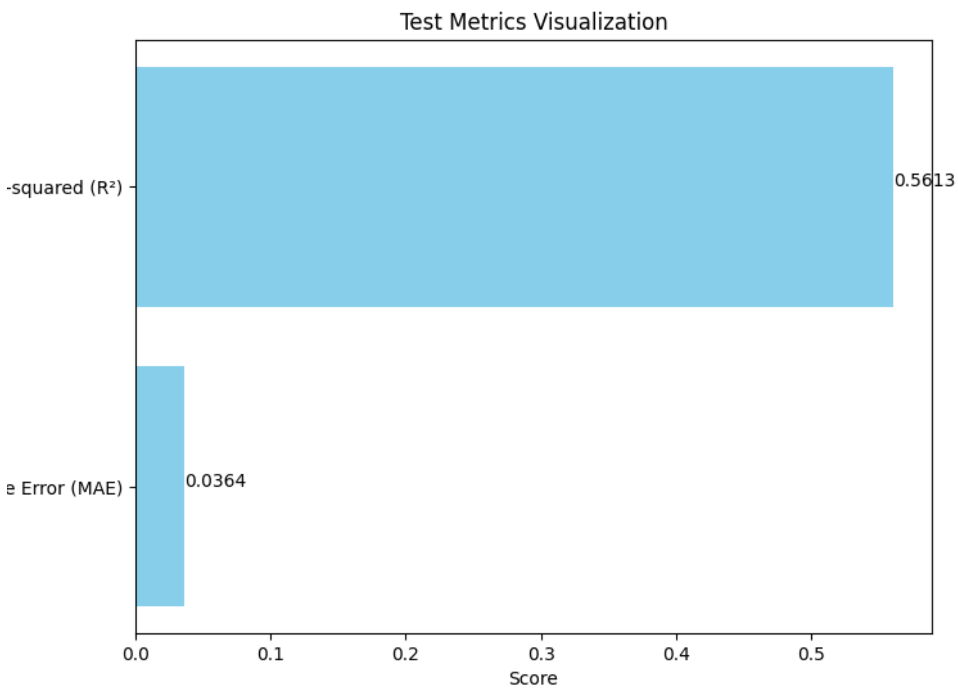
Model 2 utilizes Ridge Regression, a type of linear regression that introduces regularization to prevent overfitting. This is particularly important when dealing with datasets that have numerous features, as it helps to maintain the model's generalizability. The model is built upon a robust set of features derived from vaccination and hesitancy data, including but not limited to vaccination hesitancy rates, demographic information, and social vulnerability indices. These features were preprocessed and transformed, with a focus on standardizing numerical data and encoding categorical variables for optimal model performance.

Feature Engineering and Selection

A critical aspect of Model 2 was the careful engineering and selection of features. Unlike Model 1, which primarily dealt with discrete features for classification, Model 2 required the manipulation of continuous variables to better capture the nuances in predicting vaccination rates. Polynomial features were introduced to model non-linear relationships, albeit with a controlled complexity to avoid overfitting (Chen et al., 2020). Furthermore, feature selection was rigorously performed using Recursive Feature Elimination (RFE) and other techniques, ensuring that only the most relevant features were included in the final model.

Model Performance and Evaluation

The performance of Model 2 was thoroughly evaluated using standard regression metrics, including the Mean Absolute Error (MAE) and the R-squared (R^2) score. The model achieved an R^2 score of approximately 0.57, indicating that it could explain about 57% of the variance in the vaccination rates. This performance, while not perfect, represents a significant capability in predicting continuous outcomes based on a range of influencing factors. Additional validation techniques, such as cross-validation, were also employed to assess the model's robustness and generalizability.

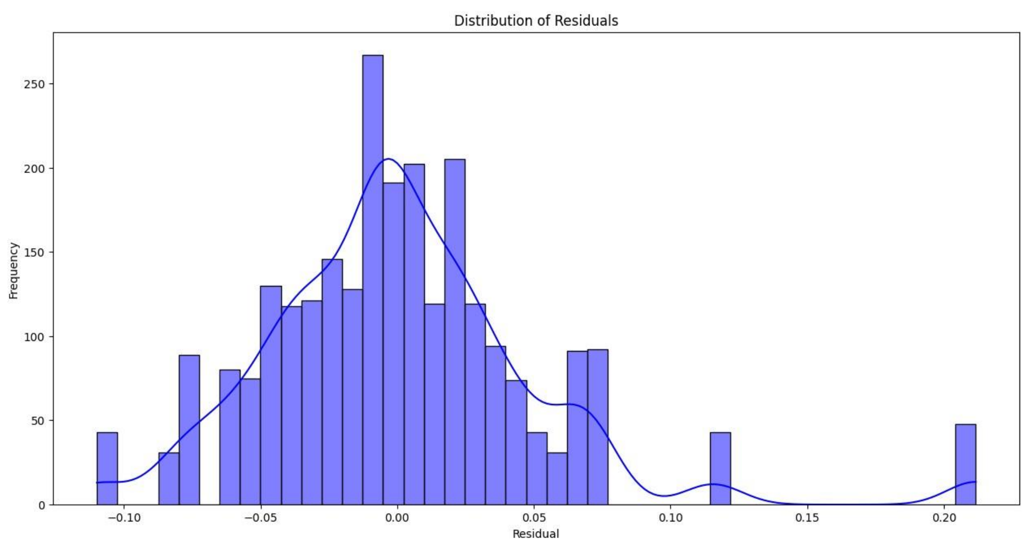


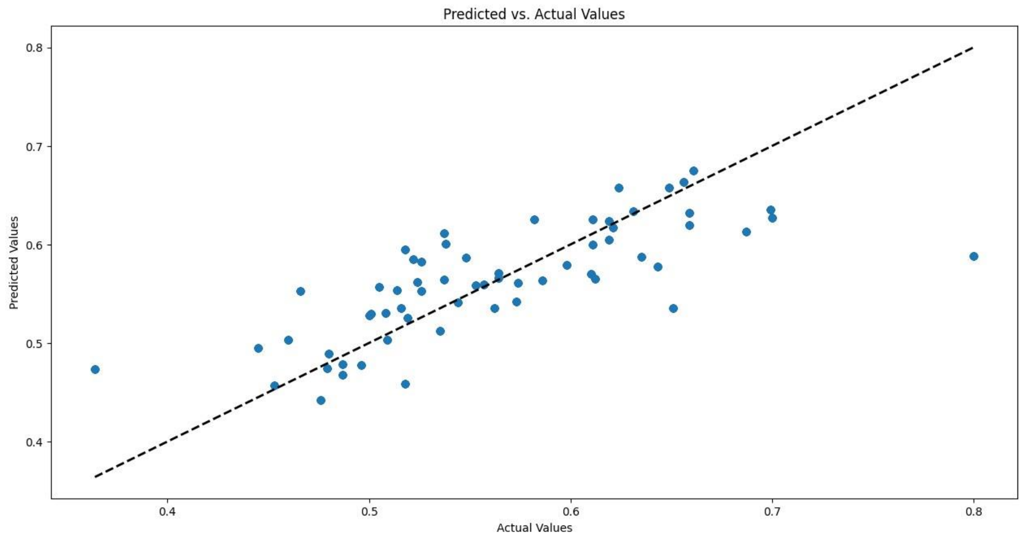
Incorporating Risk Evaluation in the Model Process:

The development of Model 2 inherently incorporates elements of risk evaluation, which is critical in managing public health initiatives. The model's primary function is to forecast vaccination rates across different regions, utilizing current and historical data related to vaccination uptake and hesitancy. This predictive capability aligns with forecasting methodologies in risk evaluation, as it allows health authorities to identify areas potentially lagging in vaccination efforts.

Model 2's approach to risk evaluation extends beyond mere forecasting. By integrating features such as social vulnerability indices and demographic data, the model offers a nuanced understanding of the factors that might influence vaccination rates. This understanding aids in simulating various scenarios, wherein health authorities can gauge the effectiveness of different intervention strategies. Such simulations can inform policy decisions, like where to focus educational campaigns or allocate additional vaccination resources.

Additionally, Model 2 incorporates optimization strategies in its feature selection and regularization processes. The careful selection of features and the application of Ridge Regression ensure that the model captures essential patterns in the data without overfitting. This balance is crucial for making reliable predictions in diverse scenarios. The use of polynomial features, while controlled to avoid excessive complexity, allows the model to capture more intricate relationships in the data, further enhancing its predictive accuracy.





Conclusion and Future Directions

Model 2 stands as a testament to the adaptability required in data analytics. Its development from a different foundational approach than Model 1 underscores the diverse methodologies needed to tackle varied analytical problems. There are opportunities to further refine this model. Potential improvements could include more sophisticated feature engineering, exploring other forms of regression analysis, or even integrating machine learning algorithms to capture more complex patterns within the data. As the landscape of data analytics continues to evolve, so will the models and techniques we employ to extract meaningful insights.

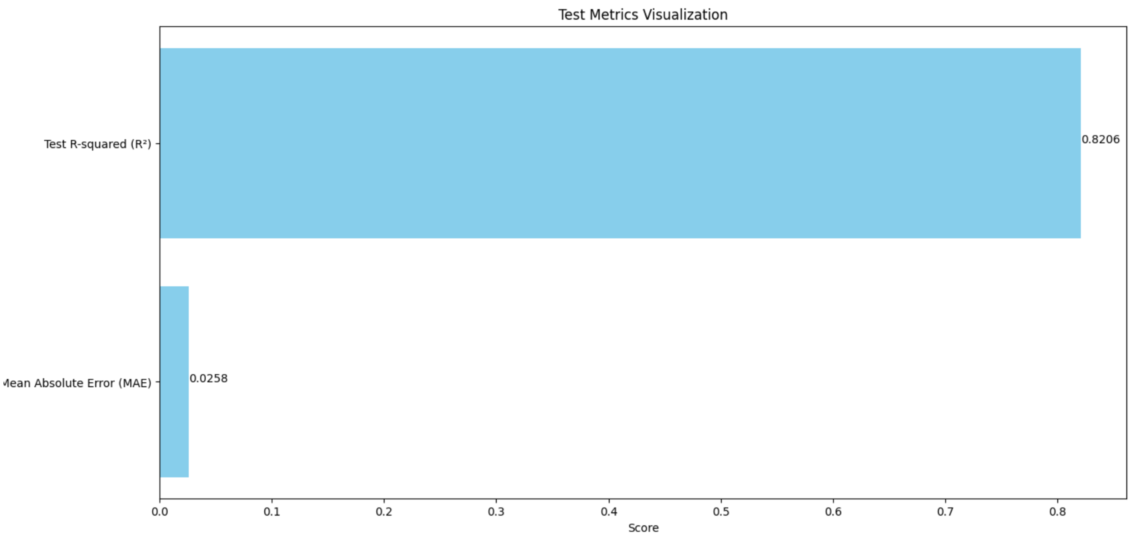
External Model Verification and Calibration

In our case study, we initially validated two models. Model 1, using logistic regression, underwent cross-validation, achieving an average recall score of about 0.808. While this model showed high recall, its precision was lower at 0.21, indicating some false positives in identifying vaccine hesitancy. For Model 2, we initially employed Ridge regression, but it only achieved moderate predictability, with an R^2 score of 0.56 and a Mean Absolute Error (MAE) of 0.036.

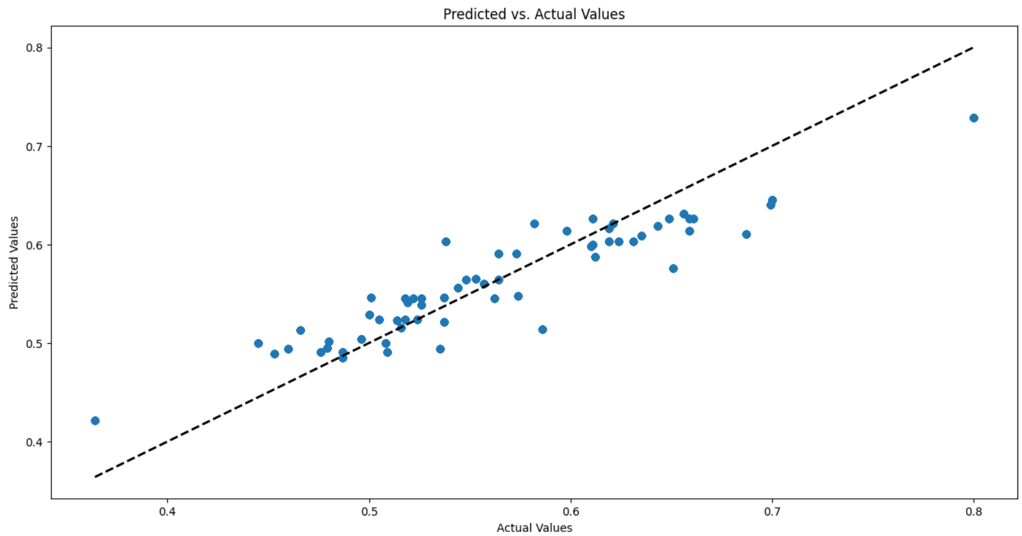
Calibration

Unsatisfied with Model 2's performance, we explored alternative models and settled on using Gradient Boosting Machines (GBM), specifically employing XGBoost. GBM is an advanced ensemble technique that builds multiple decision trees sequentially, with each tree trying to correct the errors of its predecessor. This approach is particularly effective in handling complex datasets with intricate patterns, as it combines the predictive power of multiple trees to improve accuracy. XGBoost, or Extreme Gradient Boosting, is a more efficient version of GBM that offers faster computation and better performance. It can handle a variety of data types, distributions, and relationships, making it particularly suited for our complex healthcare analytics dataset. The switch to XGBoost likely resulted in better results due to its ability to manage complex interactions in the data more effectively and its robustness against overfitting.

However, initial iterations of this model showed signs of overfitting, indicated by extremely high R^2 values. To address this, we systematically adjusted the model's complexity and regularization parameters. By reducing the number of estimators and the depth of trees, and increasing the regularization, we aimed to balance the model's fit to the data against its ability to generalize. These adjustments brought the model's performance closer to Model 1, achieving more realistic R^2 scores in the low 80s and a MAE of about 0.0258. This iterative process demonstrates the nuanced approach required in predictive modeling, especially when dealing with complex datasets in healthcare analytics.



For the validation of our new GBM model using XGBoost, we employed a combination of hold-out testing and early stopping techniques. The hold-out method involved splitting our dataset into training and testing sets, allowing us to evaluate the model's performance on unseen data. Early stopping was implemented to prevent overfitting; the training process would halt once the model's performance on a validation set ceased to improve, ensuring a balanced approach between learning from the data and generalizing well to new data. This strategy, pivotal in modern machine learning practices, particularly for complex models like GBM, significantly contributed to refining our model's predictive accuracy and reliability.



Literature Review

To complement our research, we reviewed three studies: "Global Trends and Correlates of COVID-19 Vaccination Hesitancy: Findings from the iCARE Study" by Stojanovic et al. (2021), "Revisiting COVID-19 Vaccine Hesitancy around the World using Data from 23 Countries in 2021" by Lazarus et al. (2022), and "Vaccine Allocation and Distribution: A Review with a Focus on Quantitative Methodologies and Application to Equity, Hesitancy, and COVID-19 Pandemic" by Blasioli et al. (2023). These studies provided valuable insights into global vaccination hesitancy and the need for quantitative methodologies in addressing this challenge.

1. "Global Trends and Correlates of COVID-19 Vaccination Hesitancy: Findings from the iCARE Study" by Stojanovic et al. (2021): This study analyzed vaccine hesitancy in eight countries, using data to understand public attitudes towards COVID-19 vaccination. Its

insights are crucial for developing targeted communication strategies to address hesitancy (Stojanovic et al., 2021).

2. "Revisiting COVID-19 Vaccine Hesitancy around the World using Data from 23 Countries in 2021" by Lazarus et al. (2022): Investigating vaccine hesitancy across 23 countries, this research found no direct correlation between hesitancy and a country's COVID-19 case burden, highlighting the complexity of vaccine acceptance (Lazarus et al., 2022).

3. "Vaccine Allocation and Distribution: A Review with a Focus on Quantitative Methodologies and Application to Equity, Hesitancy, and COVID-19 Pandemic" by Blasioli et al. (2023): This review emphasizes the importance of quantitative methodologies in equitable vaccine distribution, particularly in the context of the COVID-19 pandemic. It addresses how these methods can help overcome vaccine hesitancy and ensure fair access to vaccines (Blasioli et al., 2023).

Future Recommendations

Moving forward, our model could be refined further by incorporating more diverse datasets and considering additional variables that might affect vaccine hesitancy. One potential area for improvement is the integration of socio-demographic factors, which could enhance the model's applicability across different populations. If given another opportunity, we would focus on expanding our data sources and exploring more sophisticated machine learning techniques, like deep learning, to capture the nuanced patterns in vaccine hesitancy better.

While our model has demonstrated promising results, continuous refinement and validation of new datasets and case studies are essential. The insights gained from our study and the comparison with other research in the field reinforce the importance of adaptable and robust modeling techniques in understanding and addressing public health challenges like vaccine hesitancy.

Model Deployment and Model Lifecycle

In addressing the deployment strategy and life cycle of the data model for Acme Health Care Company, we delve into several key aspects: an improvement plan, model verification and calibration, a deployment strategy statement, cost benchmarking, milestones and timelines, training requirements, model quality tracking, and organizational benefits.

Improvement Plan: The retrospective analysis from Topic 6 highlighted crucial areas for enhancement. Over three months, we'll conduct a comprehensive cost analysis and expand the scope to incorporate new features, addressing risks in data management and model complexity through rigorous testing and validation.

Model Verification and Calibration: The model, while robust, shows room for improvement in handling edge cases. We propose further iterations to refine these aspects and adopt a phased approach for future projects. The potential for repetitive use is promising, contingent on continuous data updates and periodic recalibration.

Deployment Strategy Statement: Our deployment strategy aims for seamless integration into Acme Health Care Company’s workflow. This includes comprehensive staff training, regular evaluations, and a feedback mechanism, beginning with a pilot program and culminating in a full-scale rollout. We anticipate challenges such as system compatibility and staff adaptability, planning to address them through detailed preparation and responsive support mechanisms.

Cost Benchmarking: In our quest to establish a robust framework for cost benchmarking, we thoroughly examined various studies, ultimately selecting two distinct yet complementary models: the WHO report on COVID-19 vaccine delivery in 92 AMC countries (Griffiths et al., 2021), and a comprehensive U.S. study on the cost-benefit of vaccinating older adults against four diseases (Carrico et al., 2021).

The WHO report offers a detailed analysis of COVID-19 vaccine delivery costs in 92 AMC countries, estimating the total financial cost at approximately US$ 2.018 billion. This translates to US$ 1.66 per dose and US$ 3.70 per individual vaccinated with two doses. Major cost components include in-country delivery and upfront costs, as well as technical assistance at various levels.

Complementing this, the study on the cost-benefit of vaccination against four diseases in older adults in the U.S. highlights the economic viability of vaccination programs, showing significant long-term healthcare savings and a positive benefit-cost ratio. These studies collectively inform the cost benchmarking for Acme Health Care Company’s data model deployment.

Budget Justification: While the WHO report establishes a baseline, the inclusion of the U.S. study underscores the complexity of vaccine-related data analysis, suggesting a potentially higher budget allocation for Acme's model due to its advanced technological needs and extensive data requirements.

Cost Savings and Efficiency: Acme's model could potentially achieve greater savings and efficiency compared to the scenarios in both studies. By leveraging advanced analytics, the model can optimize vaccine distribution processes, echoing the U.S. study's findings on economic benefits of vaccination in an aging population.

Long-term Financial Implications: The long-term financial benefits of Acme's model are significant. Efficient vaccine distribution, underpinned by precise data analytics, aligns with the U.S. study's findings on societal cost savings, including reduced healthcare expenses and improved health outcomes, leading to substantial cost savings over time.

Milestones and Timeline: We have structured the deployment into key milestones over six weeks. This includes detailed staff training, system integration, a pilot phase, and final evaluations, each stage designed to build upon the previous, ensuring smooth progression to full-scale implementation.

Our structured six-week plan includes:

- **Weeks 1-2**: Staff training with a focus on model functionality and data interpretation.
- **Weeks 3-4**: System integration and pilot phase, testing the model in a controlled environment.
- **Week 5**: Evaluation and feedback collection from pilot testing.
- **Week 6**: Full-scale implementation, based on the insights gained from the pilot phase.

Training Requirements: Training will encompass both technical and practical aspects of the model. This includes understanding the predictive analytics process, from identifying objectives to testing and refining the model. Tailored sessions for different user groups will ensure effective knowledge transfer, with continuous support and resources available post-training.

Model Quality Tracking: We will utilize a combination of performance metrics like accuracy, recall, and user satisfaction. Regular monthly checks, supplemented by quarterly in-depth reviews, will ensure the model remains effective and relevant. User feedback will be a crucial component of this process, allowing us to adapt and refine the model continually.

Organizational Benefits: The model's deployment is expected to revolutionize vaccine distribution, leading to improved public health outcomes and operational efficiencies. Quantifiable benefits include reduced vaccine wastage, optimized resource allocation, and enhanced data-driven decision-making capabilities. Communicating these benefits organization-wide will be key to fostering a culture of innovation and data literacy.

In conclusion, the deployment strategy and life cycle of our model are designed to be dynamic and responsive, not only addressing current challenges in vaccine distribution and hesitancy but also setting a precedent for future data-driven healthcare initiatives. Through meticulous planning, rigorous testing, and continuous improvement, the model is set to offer substantial long-term benefits, marking a significant step in pandemic preparedness and response for Acme Health Care Company.

Recommendations for Practice, Future Research, and Conclusions

1. Recommendations for Practice

Based on our findings from the study, several key recommendations for practice emerge:

- **Data-Driven Decision-Making**: Healthcare organizations should integrate data analytics into their decision-making processes, particularly for vaccine distribution and managing hesitancy. This involves using predictive models to identify potential hotspots and under-vaccinated demographics for targeted interventions.
- **Tailored Communication Strategies**: Understanding the nuances of vaccine hesitancy across different demographics is crucial. Healthcare providers should develop customized communication and educational campaigns based on insights gained from data analysis, addressing specific concerns and misinformation prevalent in various communities.
- **Training and Skill Development**: Continuous training for healthcare professionals in data analytics and interpretation is vital. This will ensure that they are well-equipped to make informed decisions and can effectively contribute to public health strategies.
- **Partnership and Collaboration**: Collaborating with governmental and non-governmental organizations can enhance the reach and effectiveness of vaccination campaigns. Sharing data and insights can lead to more coordinated and comprehensive public health responses.

2. Recommendations for Future Research

- **Exploring Socio-Demographic Factors**: Future research should delve deeper into the impact of socio-demographic factors on vaccine hesitancy and uptake. This could involve analyzing more diverse datasets and integrating advanced machine learning techniques.
- **Longitudinal Studies**: Conducting longitudinal studies to monitor the long-term effectiveness of different strategies in combatting vaccine hesitancy can provide valuable insights for sustained public health efforts.
- **Comparative Studies**: Comparative analysis of different regions or countries could offer insights into the effectiveness of various vaccine distribution strategies and public health policies.
- **Ethical Considerations in Data Usage**: There are potential ethical implications when using large-scale health data for public health purposes. Safeguards protecting the public should include considerations around privacy, consent, and data security.

3. Conclusions

Our thesis successfully demonstrates the power of data analytics in addressing public health challenges, particularly in the context of a pandemic. The models developed provide a blueprint for healthcare organizations to optimize vaccine distribution, address vaccine hesitancy effectively, and make informed decisions.

This research underscores the necessity of a data-driven approach in healthcare, highlighting the potential for predictive analytics to play a pivotal role in pandemic preparedness and response. However, it also brings to light the need for continuous adaptation and evolution of models and strategies to suit changing circumstances and emerging challenges.

The findings and recommendations of this study serve not only as a guide for current pandemic management but also as a foundation for future public health initiatives. They emphasize the importance of integrating data analytics into healthcare strategies to enhance public health outcomes and operational efficiencies.

References

Ali, R. (2020, September 22). Predicting your business's future. Oracle NetSuite. https://www.netsuite.com/portal/resource/articles/financial-management/predictive-modeling.shtml

Blasioli, E., Mansouri, B., Tamvada, S. S., & Hassini, E. (2023, March). Vaccine Allocation and Distribution: A Review with a Focus on Quantitative Methodologies and Application to Equity, Hesitancy, and COVID-19 Pandemic. In Operations Research Forum (Vol. 4, No. 2, p. 27). Cham: Springer International Publishing. https://doi.org/10.1007/s43069-023-00194-8

Carrico, J., Talbird, S. E., La, E. M., Poston, S., Poirrier, J. E., DeMartino, J. K., & Hogea, C. (2021). Cost-benefit analysis of vaccination against four preventable diseases in older adults: Impact of an aging population. *Vaccine*, *39*(36), 5187-5197.

Chen, R., Dewi, C., Huang, S., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00327-4

Cordes, S. (2023, March 9). Business analysis problem solving techniques. The Business Analyst's Toolkit. https://www.businessanalyststoolkit.com/business-analysis-problem-solving-techniques/

COVID-19 vaccinations in the United States, County | Data | Centers for Disease Control and Prevention. (n.d.). Data.CDC.gov. https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh/data

Glicksman, E. (2022, January 12). Workplace shortages, Burnout Strain Hospital HR staff. SHRM. https://www.shrm.org/resourcesandtools/hr-topics/organizational-and-employee-development/pages/workplace-shortages-burnout-strain-hospital-hr-staff.aspx

Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. (n.d.-b). https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(22)00320-6/fulltext

Griffiths, U., Adjagba, A., Attaran, M., Hutubessy, R., Van de Maele, N., Yeung, K., ... & Alkenbrack, S. (2021, February 8). Costs of delivering COVID-19 vaccine in 92 AMC countries. *World Health Organization*, *8*.

Huang, S., Chen, Y., Lin, X., Ly, Y., Lien, S., Chen, P., Wang, C., Wu, S., Chen, C., & Lin, C. (2023). In silico prediction of immune-escaping hot spots for future COVID-19 vaccine design. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-40741-1

Katella, K. (2023, January 6). 9 things everyone should know about the coronavirus outbreak. Yale Medicine. https://www.yalemedicine.org/news/2019-novel-coronavirus

Lazarus, J. V., Wyka, K., White, T. M., Picchio, C. A., Rabin, K., Ratzan, S. C., Leigh, J. P., Hu, J., & El-Mohandes, A. (2022). Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-31441-x

Overview of Barriers and Facilitators in COVID-19 Vaccine Outreach. (n.d.). https://www.aspe.hhs.gov/sites/default/files/2021-08/Vaccine%20Outreach%20Research%20Report%208-27-2021%20FINAL.pdf

Risk prediction. (2023, March 13). Columbia University Mailman School of Public Health. https://www.publichealth.columbia.edu/research/population-health-methods/risk-prediction

Science, D. (2023, October 2). What are the most effective ways to use data to solve business problems? www.linkedin.com. https://www.linkedin.com/advice/1/what-most-effective-ways-use-data-solve-business-problems

Stojanovic, J., Boucher, V. G., Gagne, M., Gupta, S., Joyal-Desmarais, K., Paduano, S., Aburub, A. S., Sheinfeld Gorin, S. N., Kassianos, A. P., Ribeiro, P. A. B., Bacon, S. L., & Lavoie, K. L. (2021, June 17). Global trends and correlates of COVID-19 vaccination hesitancy: Findings from the iCARE Study. MDPI. https://www.mdpi.com/2076-393X/9/6/661

Vaccine distribution, supply chain, testing still present challenges in federal pandemic response. (n.d.-b). U.S. GAO. https://www.gao.gov/blog/vaccine-distribution%2C-supply-chain%2C-testing-still-present-challenges-federal-pandemic-response

Vaccine Hesitancy for COVID-19: County and local estimates | Data | Centers for Disease Control and Prevention. (2021, June 17). https://data.cdc.gov/Vaccinations/Vaccine-Hesitancy-for-COVID-19-County-and-local-es/q9mh-h2tw

With virus season looming and awareness campaign funding slashed, experts see a scattered Covid-19 vaccine rollout ahead. (n.d.). https://www.cnn.com/2023/10/18/health/covid-vaccine-awareness-campaigns/index.html