H1N1 AND SEASONAL FLU VACCINE PREDICTION

Presented by Kitts Kikumu

OVERVIEW

- Abstract
- Introduction
- Problem
- Objectives
- Methodology
- Result
- Conclusion

ABSTRACT

Influenza, commonly known as the flu, is a contagious respiratory illness caused by influenza viruses. It can cause mild to severe illness, and at times can lead to death.

Vaccination is the most effective way to prevent influenza and its severe outcomes.

Public health organizations, like the Centers for Disease Control and Prevention (CDC), invest heavily in annual vaccination campaigns. However, the success of these campaigns depends on public willingness to get vaccinated, often referred to as "vaccine uptake."

INTRODUCTION

The data provided is from the National 2009 H1N1 Flu Survey (NCFS), a telephone survey conducted in the United States. It captures a snapshot of public opinion, knowledge, and behavior during the H1N1 pandemic. This includes demographic information, personal health behaviors (like hand washing), opinions about vaccine effectiveness and risk, and ultimately, whether the respondent received the H1N1 and seasonal flu vaccines.

PROBLEM STATEMENT

- Public health officials face the significant challenge of effectively allocating limited resources (vaccines, personnel, marketing funds) to maximize vaccination rates and protect the population during seasonal flu seasons and pandemics.

- The core problem is to understand the demographic, behavioral, and opinion-based factors that influence an individual's decision to receive the H1N1 and seasonal flu vaccines.

- How can we identify and characterize the segments of the population that are least likely to be vaccinated against H1N1 and seasonal flu, in order to design targeted and effective public health interventions?

OBJECTIVE

To build and evaluate a robust machine learning model that accurately predicts the likelihood of an individual receiving the h1n1_vaccine and the seasonal_vaccine based on their survey responses.

To identify the most influential predictors of vaccination behavior. For example, is a doctor's recommendation more impactful than a person's perceived risk of getting sick?

To provide clear, data-driven recommendations that can help public health stakeholders refine their communication strategies and outreach efforts.

DATA OVERVIEW

Features Included

Many features are encoded as numbers (e.g., h1n1_concern from 0-3) which represent ordinal scales.
Features like age_group, race, and employment_status are represented as strings and will require encoding for use in most machine learning models.

Target Variables

The two columns we aim to predict are h1n1_vaccine and seasonal_vaccine. Both are binary, where 1 means the respondent received the vaccine and 0 means they did not.

Dataset size

The dataset contains 26,707 rows and 38 columns.

DATA PREPARATION

Handling Missing Values: Missing values in various columns were imputed. For categorical columns, the mode was used, while for numerical columns, the mean was used

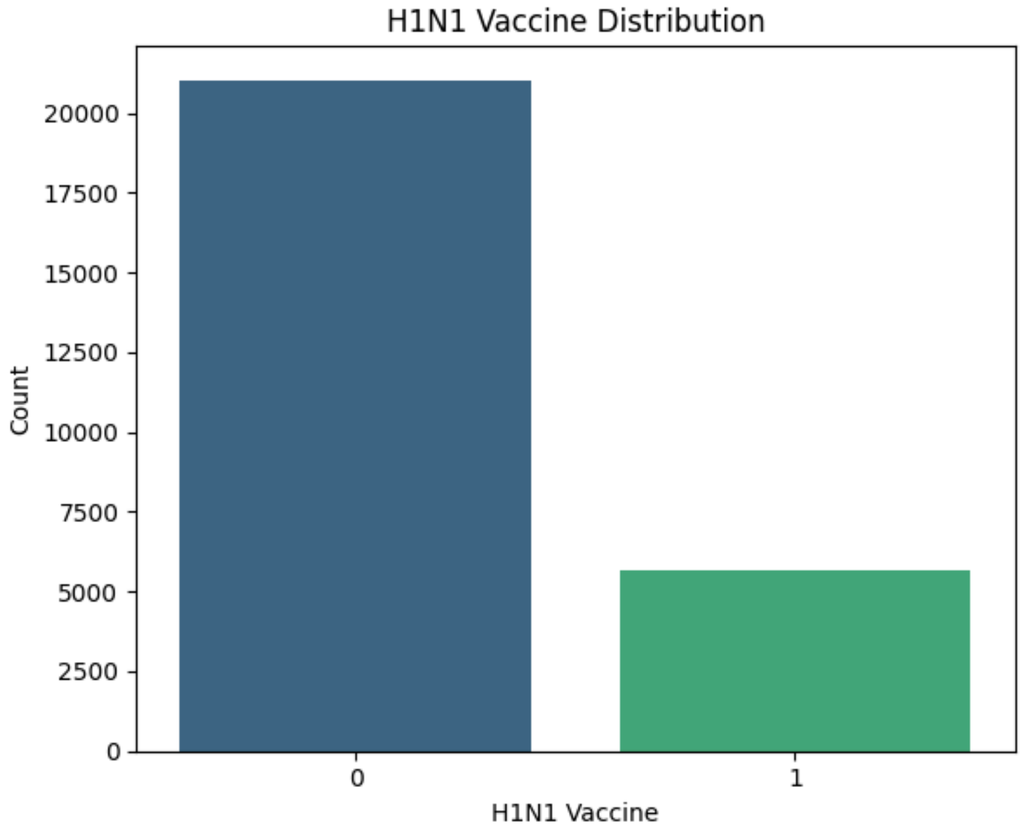
Encoding Categorical Variables: Categorical features were converted into a numerical format using one-hot encoding.

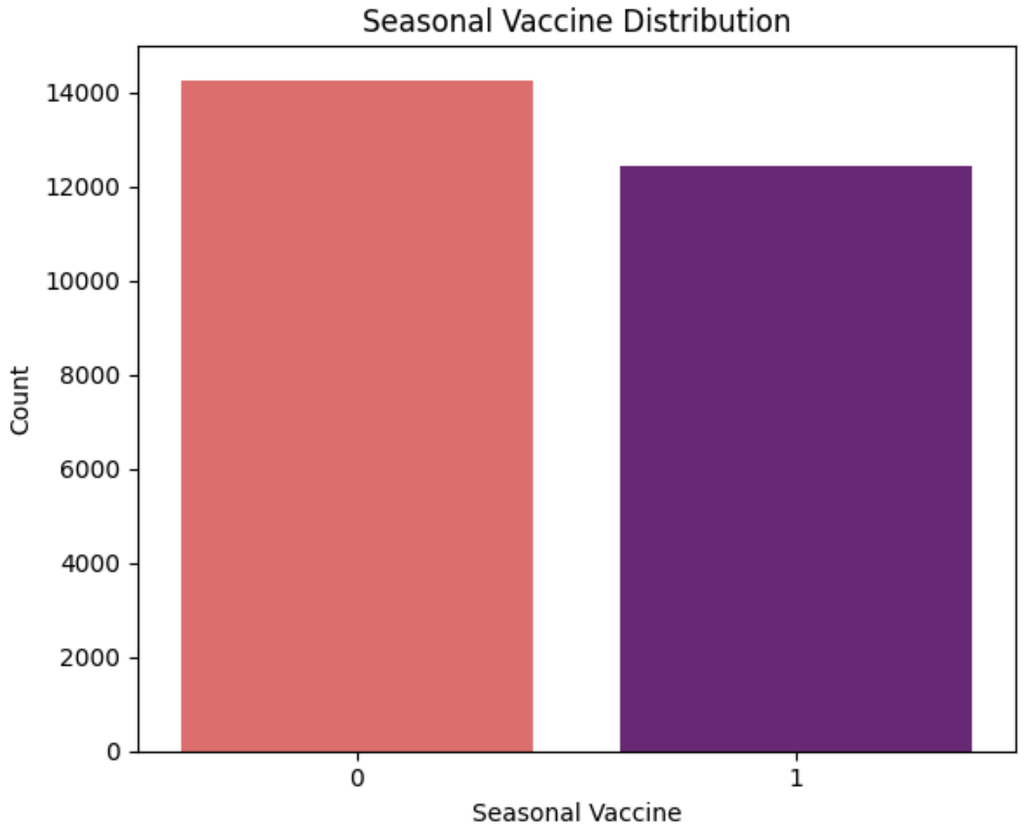
Feature Engineering: Some columns were transformed or combined to create new, more informative features.

EXPLORATORY DATA ANALYSIS

H1N1 VACCINE & SEASONAL VACCINE DISTRIBUTION

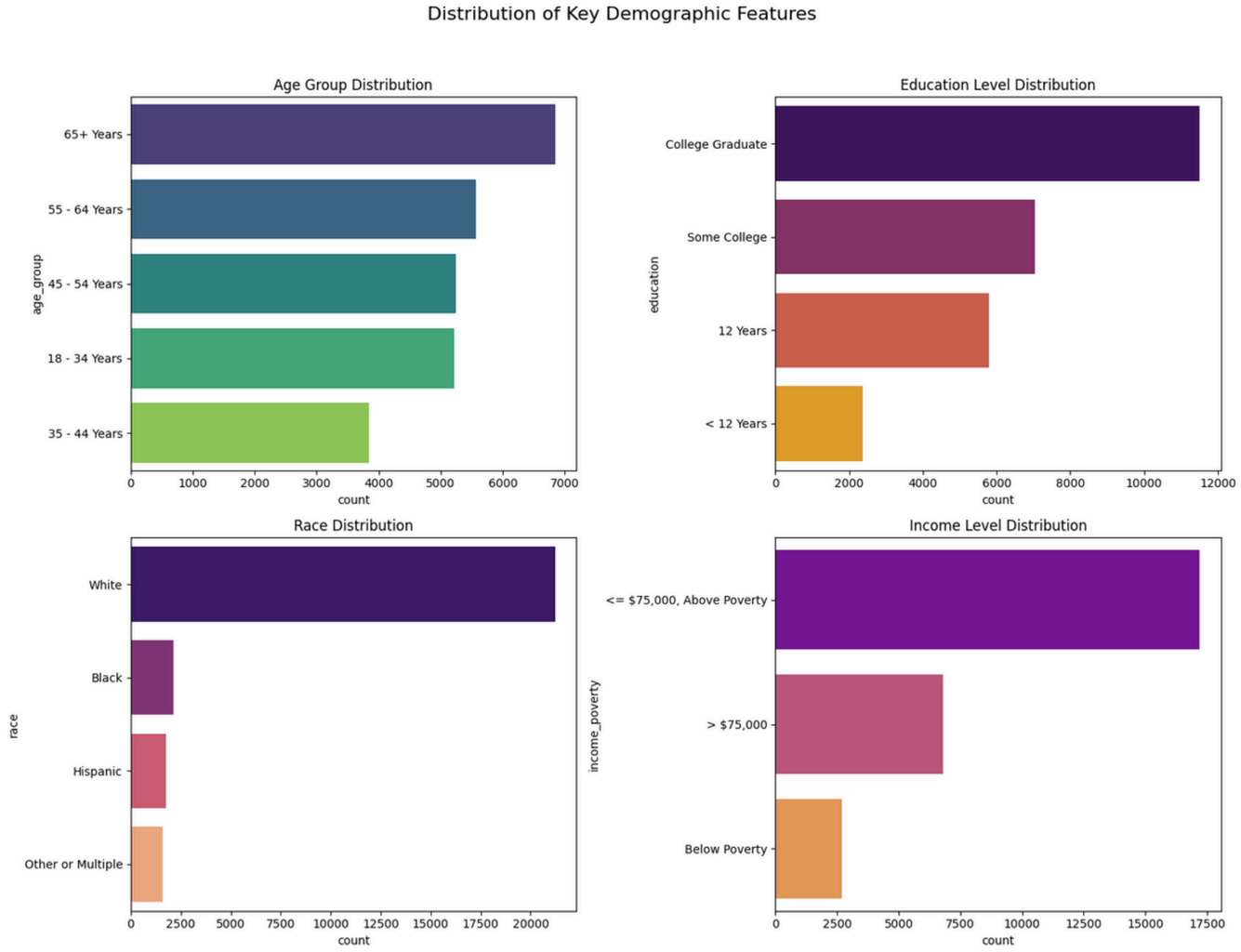
From the visualization we can see that a larger demographic have taken the seasonal vaccine as compared to the H1N1 vaccine and we can see that there is a significant class imbalance on H1N1 vaccine compared to the Seasonal Vaccine





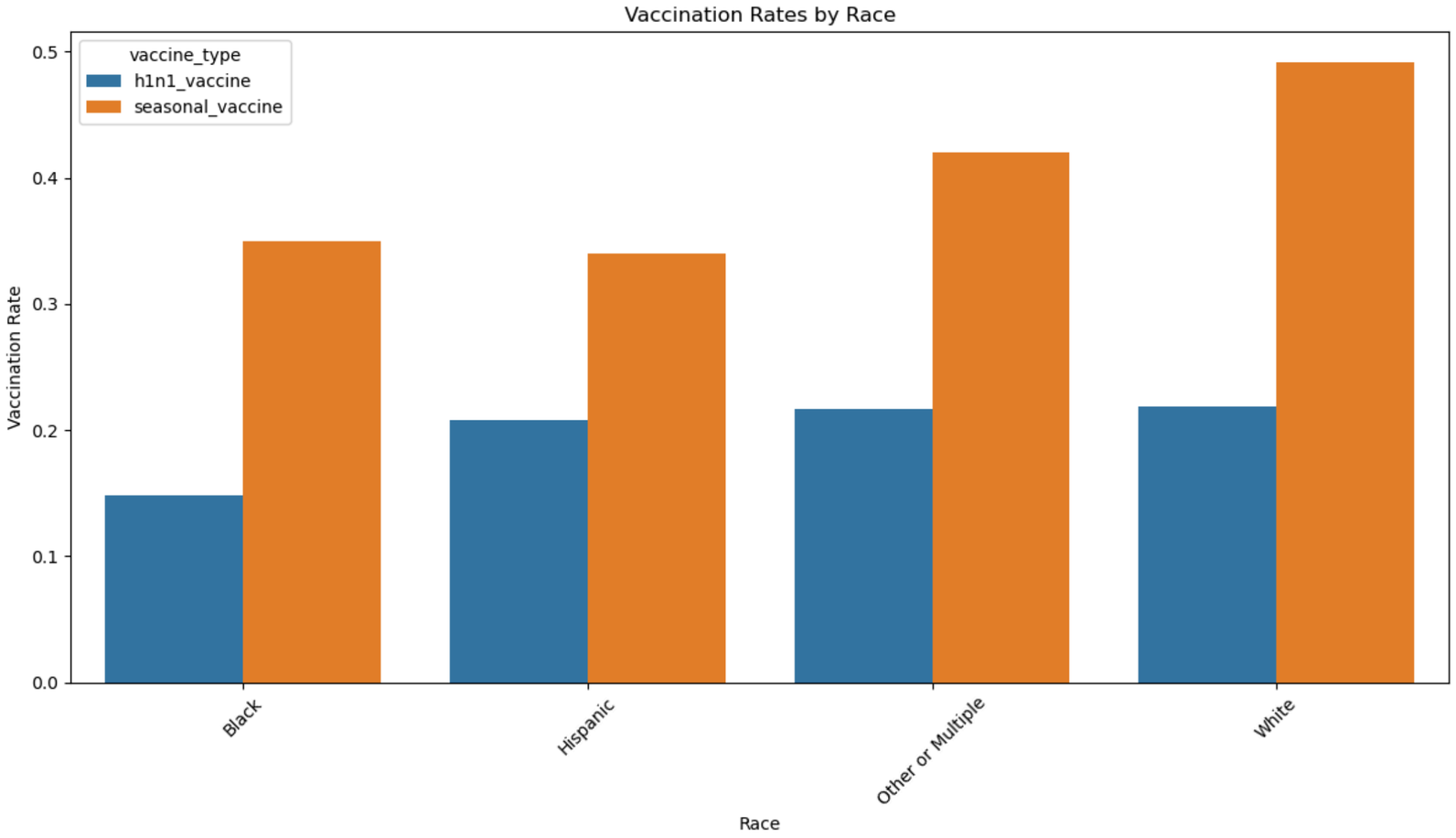
DISTRIBUTION OF KEY DEMOGRAPHIC FEATURES

There is a slightly even age group distribution with the most coming from ages 65+ years and above. We can get some conclusive observations such as the majority of the demographic are white people



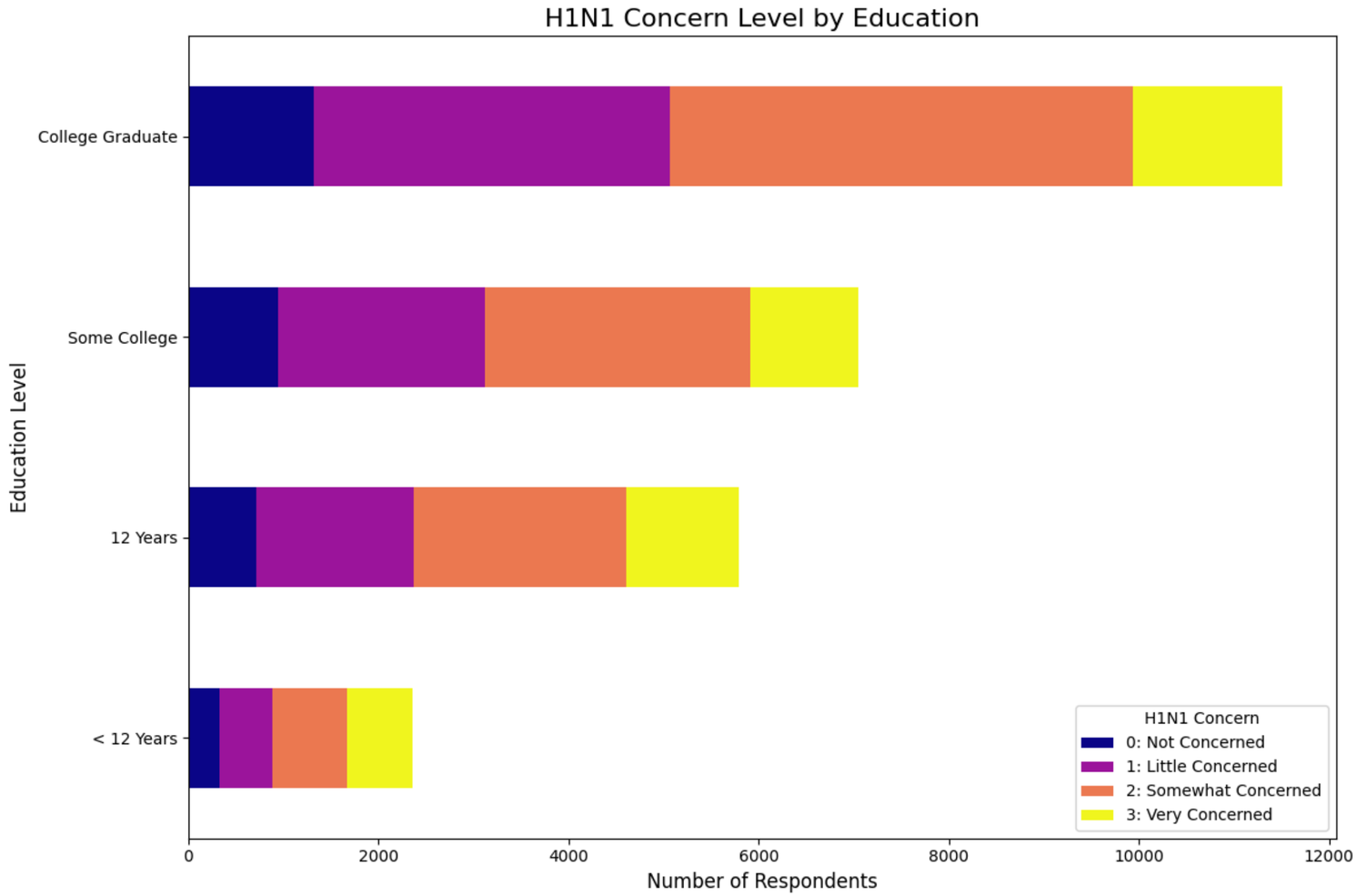
VACCINATION RATES BY RACE

The relative order of vaccination rates among the racial groups (White > Hispanic > Other > Black) is consistent for both H1N1 and seasonal vaccines.



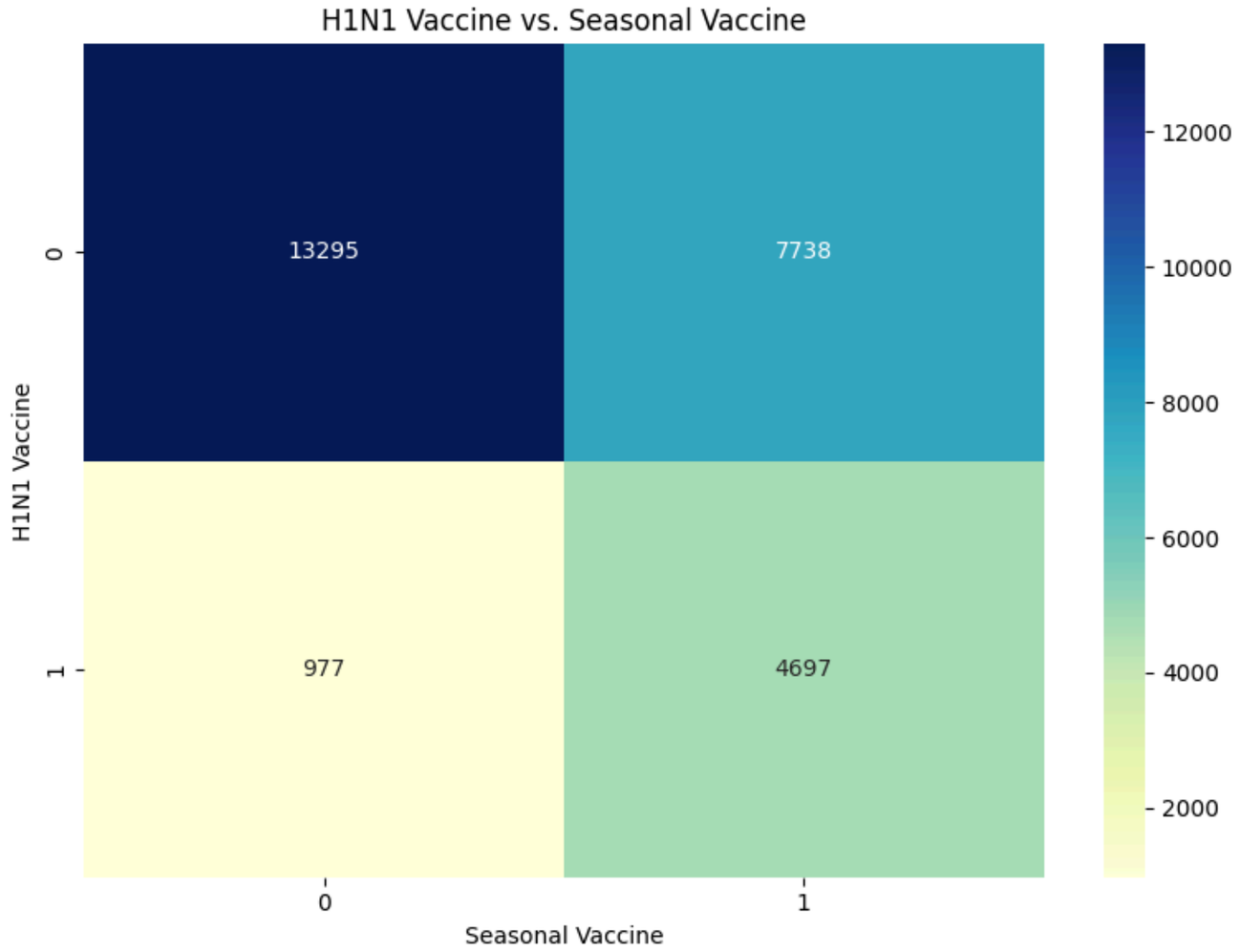
H1N1 CONCERN LEVEL BY EDUCATION

The level of concern appears to shift with higher education. This could imply that higher education is correlated with a greater awareness of health risks and, consequently, higher concern.



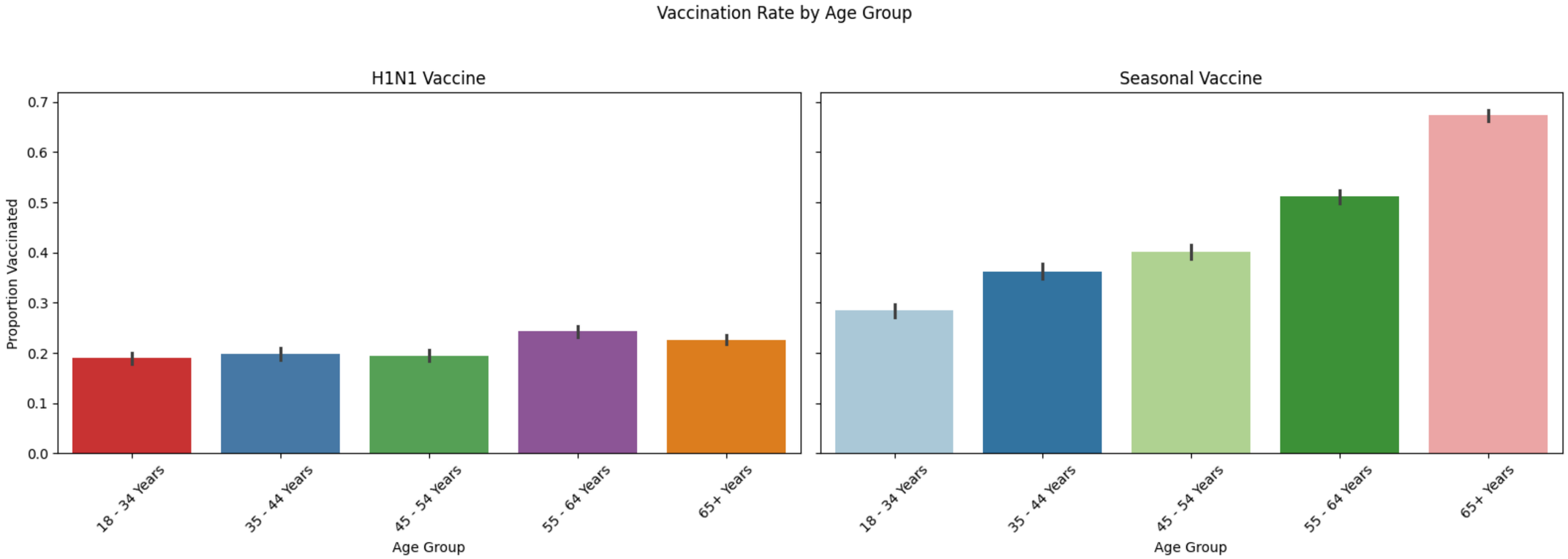
H1N1 VACCINE VS. SEASONAL VACCINE

There is a strong positive correlation between getting the H1N1 vaccine and getting the seasonal vaccine. This suggests that a person's decision to get vaccinated is not vaccine-specific but rather a general health behavior.



VACCINATION RATE BY AGE GROUP

There is a clear trend showing that older age groups are significantly more likely to receive the seasonal flu vaccine. The trend is also present but less pronounced for the H1N1 vaccine



PREDICTIVE MODELING

MODELING APPROACH

Logistic Regression: A baseline model that provides a good starting point for comparison

Random Forest: An ensemble method that combines multiple decision trees to improve predictive performance and control for overfitting.

XGBoost: A powerful and efficient gradient boosting algorithm that is known for its high performance in classification tasks.

For both the H1N1 and seasonal flu vaccine predictions, two separate models were developed. The class imbalance was addressed using SMOTE to ensure the models were not biased towards the majority class.

H1N1 VACCINE MODEL PERFORMANCE

Model Comparison (H1N1 Vaccine) : Logistic Regression vs Decision Tree (Pruned) vs Random Forest (Tuned) vs XGBoost (Tuned)

Metric	Logistic Regression	Decision Tree (Pruned)	Random Forest (Tuned)	XGBoost (Tuned)	Observation
Training Accuracy	79%	79%	84%	91%	XGBoost shows signs of overfitting with a high 91% training accuracy that doesn't transla to validation. Random Forest also shows higher performance on training data.
Validation Accuracy	78%	79%	83%	78%	Random Forest has the highest validation accuracy, but this is misleading. The other three models have very similar and stable validation scores.
Functional Class (0) (Recall)	0.81	0.83	0.96	0.79	The Random Forest model is extremely biased towards Class 0, achieving an almost perfe recall at the expense of the other class.
Functional Class (1) (Recall)	0.64	0.64	0.36	0.73	XGBoost is the clear winner here, being the most effective model at identifying the positi (and likely minority) class. The Random Forest model fails significantly on this metric.
ROC AUC	0.79	0.79	0.82	0.83	XGBoost has the highest AUC score, indicating it is the best model at distinguishing between the positive and negative classes across all thresholds.

SEASONAL VACCINE MODEL PERFORMANCE

Model Comparison (Seasonal Vaccine) : Logistic Regression vs Decision Tree (Pruned) vs Random Forest (Tuned) vs XGBoost (Tuned)

Metric	Logistic Regression	Decision Tree (Pruned)	Random Forest (Tuned)	XGBoost (Tuned)	Observation
Training Accuracy	77%	72%	80%	79%	All models show good generalization, with small gaps between training and validation accuracy. There are no major signs of overfitting
Validation Accuracy	78%	73%	78%	78%	The Decision Tree is the clear underperformer. The other three models are tied with a strong validation accuracy of 78%.
Functional Class (0) (Recall)	0.80	0.80	0.80	0.79	All models are very effective at identifying the negative class, with nearly identical recall scores.
Functional Class (1) (Recall)	0.77	0.66	0.76	0.77	The Decision Tree struggles significantly to identify the positive class. Logistic Regression and XGBoost are the top performers here
ROC AUC	0.85	0.79	0.85	0.86	XGBoost has the highest AUC, indicating it has the best ability to distinguish between the positive and negative classes

WHY XGBOOST WINS

- Superior Recall for the Positive Class
- Best Overall Discriminative Power (AUC)
- A Good Performance Balance

CONCLUSION

- Doctor's Recommendation is the Most Powerful Predictor: Across all models, the strongest and most consistent predictor of whether an individual gets vaccinated for both H1N1 and seasonal flu is a recommendation from their doctor or a healthcare professional. This factor outweighs demographic and behavioral variables.

- Different Motivations for H1N1 vs. Seasonal Flu Vaccination: While there are some overlapping drivers, the predictive models indicate that the motivations for getting the H1N1 vaccine are more strongly tied to risk perception and concern about a pandemic, while the decision to get the seasonal flu vaccine is also influenced by more routine factors like age and having health insurance.

RECOMMENDATION

- Leverage Healthcare Providers as Key Messengers: Public health campaigns should prioritize equipping doctors and other healthcare professionals with the information and tools they need to confidently and effectively recommend flu vaccines to their patients. This could include providing them with talking points, informational materials, and reminders to discuss vaccination during patient visits.

- Address Vaccine Efficacy and Safety Concerns Directly: Public health communications should proactively address common concerns and misinformation about vaccine effectiveness and safety. Providing clear, evidence-based information can help to build public trust and counter vaccine hesitancy.

Thank You

Presented By:
Kitts Kikumu