H1N1 VACCINE UPTAKE PREDICTION

BUSINESS PROBLEM

During flu outbreaks, low vaccination rates contribute to increased hospitalizations, economic burdens, and public health crises.

Understanding the factors that influence an individual's willingness to get vaccinated is critical for government agencies, healthcare providers, and pharmaceutical companies.

Traditional vaccination campaigns rely on broad, one-size-fits-all messaging, which may not effectively target hesitant populations.

Come up with a model that predicts how likely individuals are to receive their H1N1 vaccine.

OBJECTIVES

- Identify the relationship between risk perception and vaccine uptake.

- Determine if higher knowledge about H1N1 increases the likelihood of getting vaccinated.

- Analyze if concerns about vaccine side effects reduce uptake.

- Segment population groups based on trust in vaccine effectiveness.

- Come up with a model that predicts how likely individuals are to receive their H1N1 vaccine.

DATA OVERVIEW

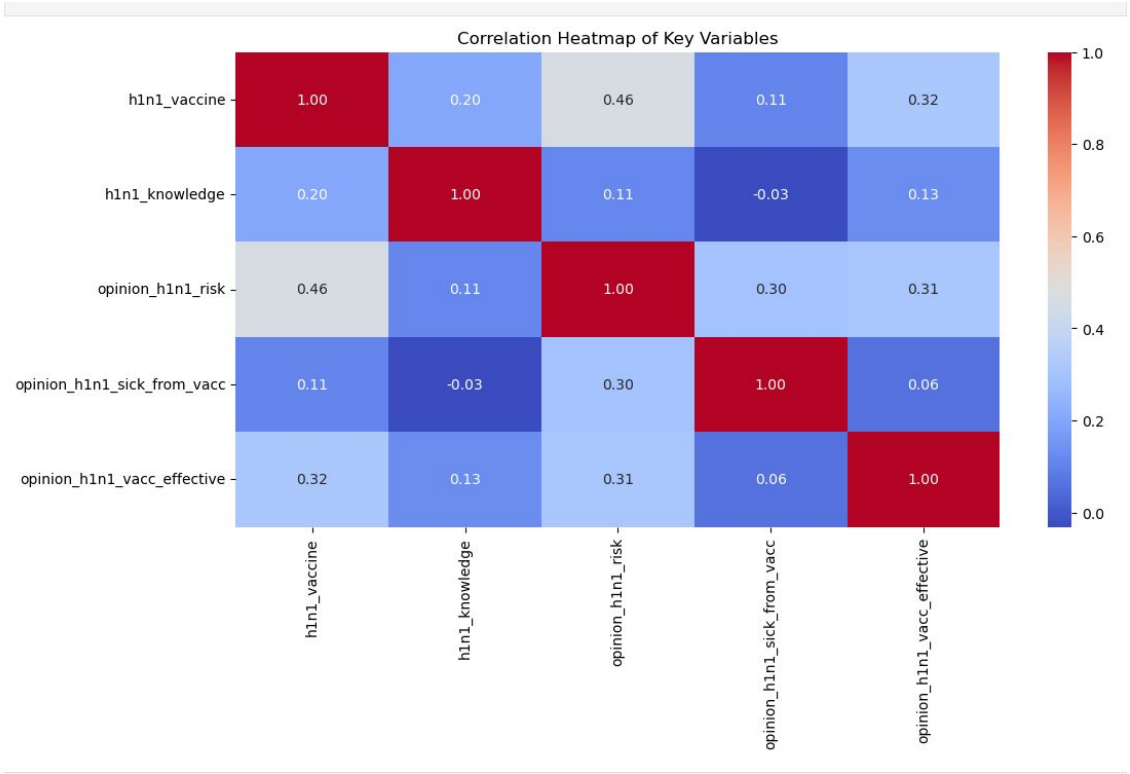
There are 36 columns in this dataset.

respondent_id,h1n1_concern, h1n1_knowledge, behavioral_antiviral_meds, behavioral_avoidance,behavioral_face_mask, behavioral_wash_hands, behavioral_large_gatherings, behavioral_outside_home, behavioral_touch_face, doctor_recc_h1n1, doctor_recc_seasonal, chronic_med_condition, child_under_6_months, health_worker, health_insurance, opinion_h1n1_vacc_effective, opinion_h1n1_risk, opinion_h1n1_sick_from_vacc, opinion_seas_vacc_effective, opinion_seas_risk, opinion_seas_sick_from_vacc, age_group, education, race, sex, income_poverty, marital_status,rent_or_own, employment_status, hhs_geo_region, census_msa, household_adults, household_children, employment_industry, employment_occupation.

DATA PREPARATION.

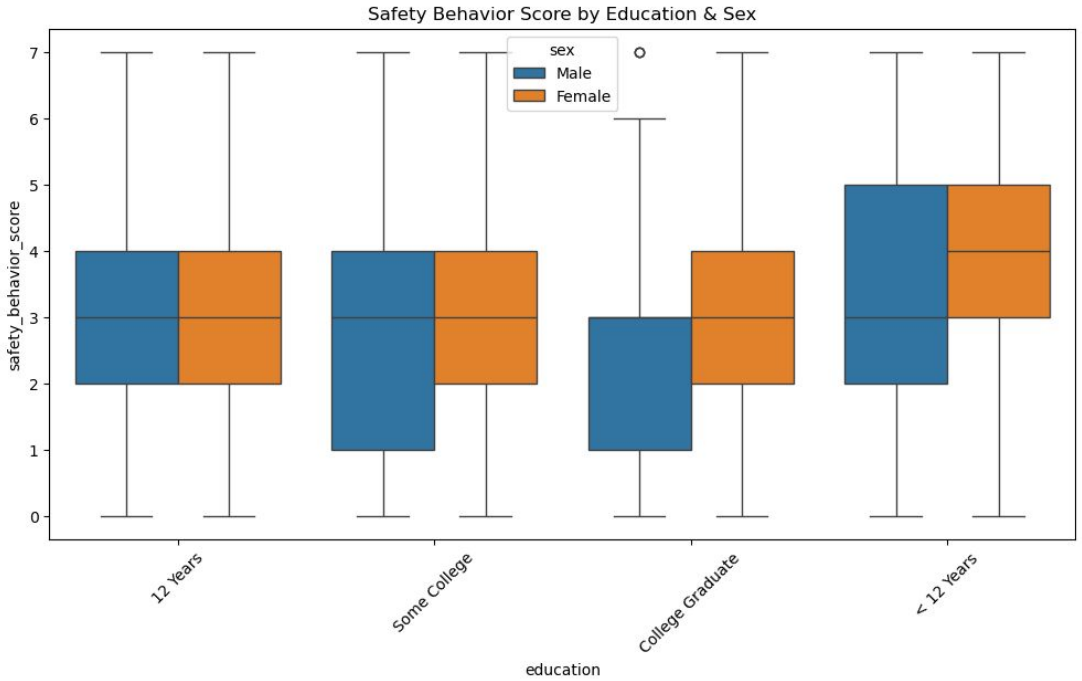
1. Dropping null values
2. Feature Engineering of columns like : safety_behavior_score, doctor_recc_total, health_risk_score,household_vulnerability.

DATA ANALYSIS



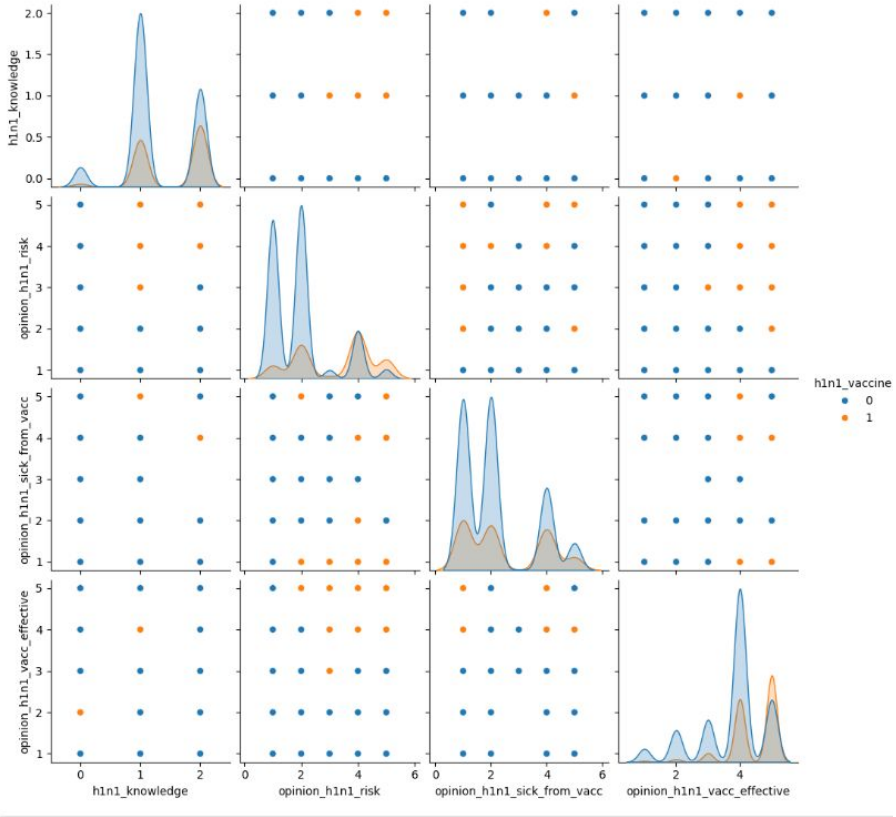
The heatmap provides a valuable overview of how different beliefs and knowledge about H1N1 relate to vaccination decisions. It highlights the importance of perceived risk and belief in vaccine effectiveness as factors influencing vaccine uptake.

SAFETY BEHAVIOR VS EDUCATION VS SEX



The grouped boxplot reveals a strong positive association between education level and safety behavior scores, with higher education generally linked to better safety practices. While sex might play a minor role, its influence is less pronounced compared to education.

VACCINE UPTAKE ANALYSIS



The plot provides visual evidence of factors associated with vaccine uptake. The strongest relationships appear to be between vaccination status and perceived risk and belief in vaccine effectiveness

MODELING

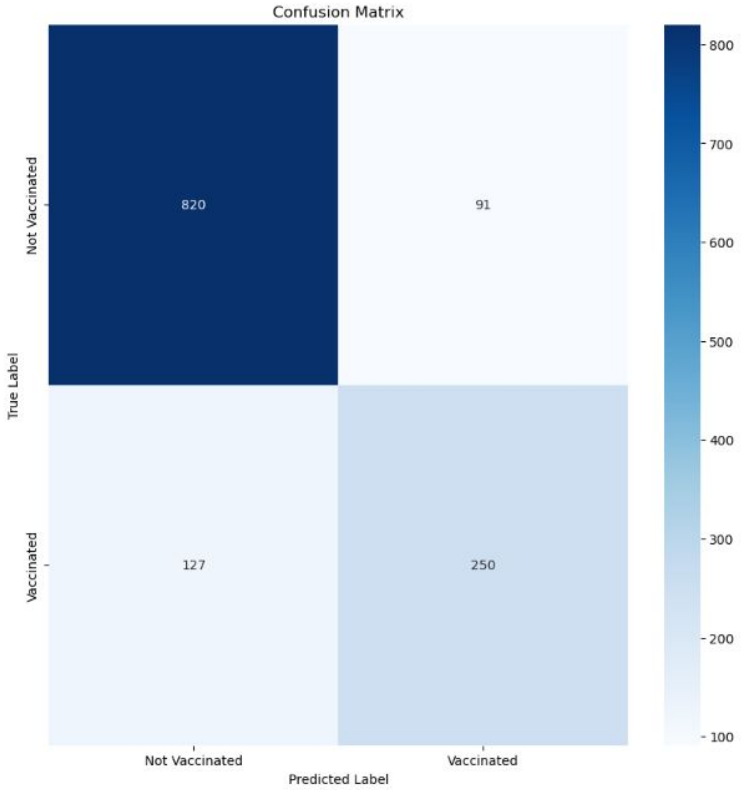
The modeling was done using two classification models : Logistic regression and Gradient boosting model.

The key metrics for analysis was receiver operator characteristics curve(roc) where the models performed 0.876953 and 0.873427 roc respectively.

Later did a feature importance analysis on gradient boosting model which suggested that socio-demographic factors and individual attitudes are key predictors.

Logistic Regression Preferred (Slightly): While both models performed very similarly (ROC AUC around 0.87), logistic regression has a slight edge in cross-validated performance (0.8734 vs 0.8762 for gradient boosting). More importantly, logistic regression offers greater interpretability. In public health contexts, understanding *why* a model makes a prediction is often as important as the prediction itself. Therefore, unless there's a strong reason to prioritize the slightly higher (but likely not statistically significant) performance of gradient boosting, logistic regression is the recommended choice.

CONFUSION MATRIX



Key Findings;
820 (True Negative): This is the number of individuals who actually did not get vaccinated, and the model correctly predicted they wouldn't.
91 (False Positive): This is the number of individuals who actually did not get vaccinated, but the model incorrectly predicted they would be vaccinated. (Also known as a Type I error)
127 (False Negative): This is the number of individuals who actually did get vaccinated, but the model incorrectly predicted they would not be vaccinated. (Also known as a Type II error)
250 (True Positive): This is the number of individuals who actually did get vaccinated, and the model correctly predicted they did.

RECOMMENDATIONS

Highlight Success Stories: Share real-life stories of how vaccines have prevented serious illnesses and protected communities.

Develop safety behavior campaigns that are specifically tailored to different education levels. Use language, channels, and examples that resonate with each group. For lower education levels, focus on clear, simple messages and practical demonstrations.

Emphasize the Benefits of Vaccination: Clearly communicate how vaccination reduces the risk of contracting the disease, experiencing severe symptoms, and spreading it to others.

Address Concerns about Safety: Provide information on the rigorous safety testing that vaccines undergo.

Provide Balanced Information: Present information about both the risks and benefits of vaccination, allowing individuals to make informed decisions.

Encourage Dialogue: Create opportunities for open dialogue and discussion about vaccines and safety behaviors.