Flu Shot Prediction: Predicting Whether Individuals got their H1N1 and seasonal Vaccines

KIPCHUMBA BRIAN

INTRODUCTION

- The spread of infectious diseases is a major public health concern.
- The recent COVID-19 pandemic highlighted the importance of vaccines in controlling the spread of diseases.
- Vaccines play a crucial role in protecting public health and preventing loss of life and economic disruption
- In this project, we are using a computer program called machine learning to study how likely people are to get vaccinated against the H1N1 and seasonal flu.
- We want to learn more about why people make the decisions they do when it comes to getting vaccinated. This will help us understand how to encourage more people to get vaccinated in the future.

OUTLINE

Business Understanding

Data Understanding

Exploratory Data Analysis

Modelling and Evaluation

Results

Conclusions

recommendation

BUSINESS UNDERSTANDING

Business Problem

- Purpose: To understand the public's behavior towards receiving vaccines by revisiting the response to the H1N1 flu pandemic in 2009 and developing a classification model that predicts the likelihood of individuals receiving the H1N1 and seasonal flu vaccines.

- Data: The National 2009 H1N1 Flu Survey data that includes demographic, social, economic, and health-related information.

- Method: A machine learning algorithm will be used to build the classification model.

- Evaluation: The success of the project will be measured by the accuracy, recall, and precision scores of the model.

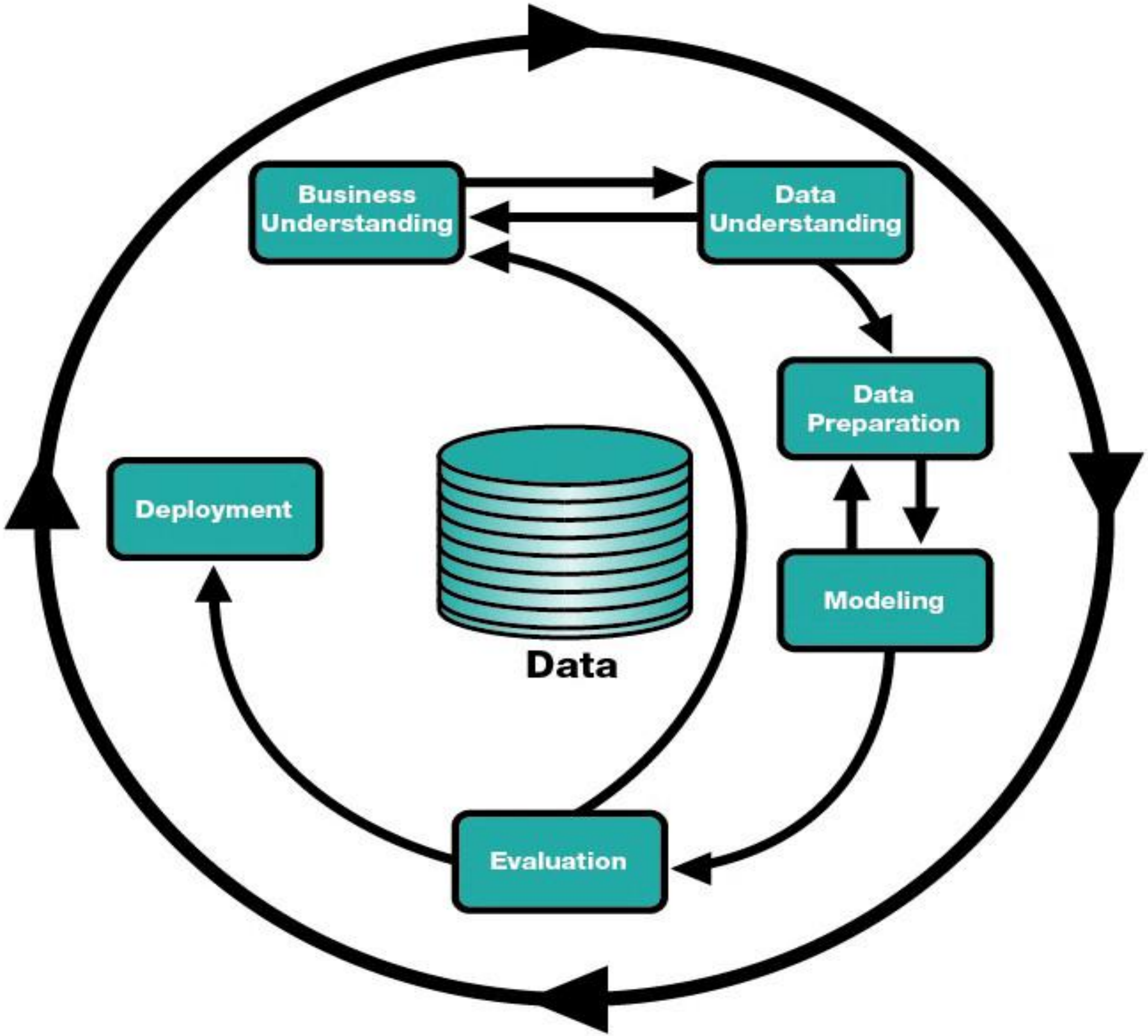
OBJECTIVES

- The main objective of this project is to develop a classification model to predict the response of individuals to H1N1 and seasonal flu vaccines.

Other Objectives

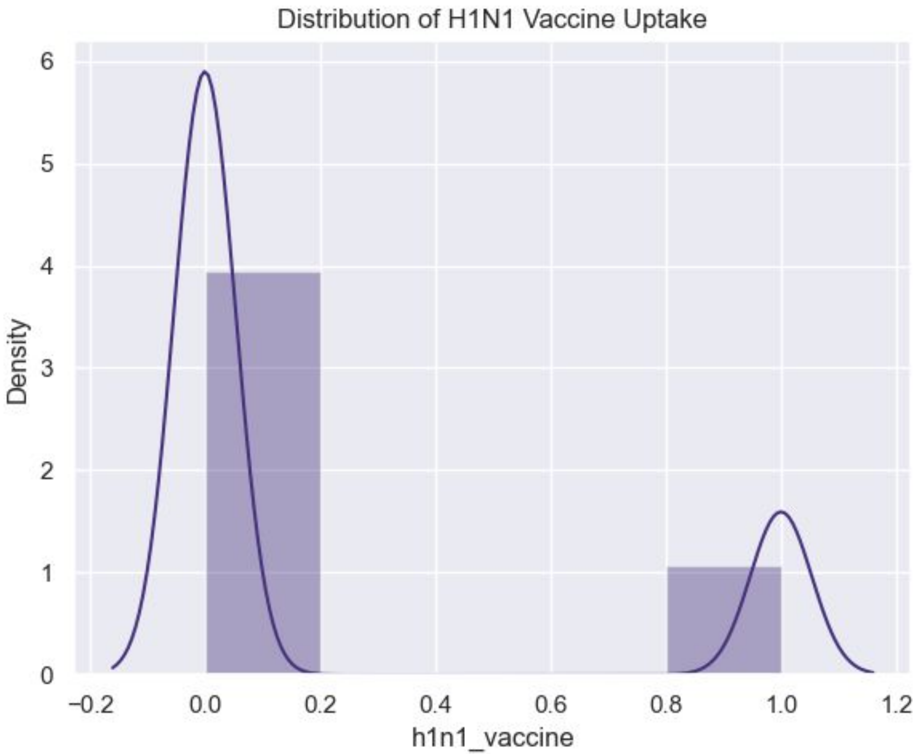
- Explore and visualize the data to gain a better understanding of the relationship between the variables and the target variable.
- Select and train different classification models, including logistic regression, decision trees, and random forest and evaluate their performance.
- To provide actionable insights for public health officials and policymakers to reduce the spread of contagious infections and promote herd immunity.

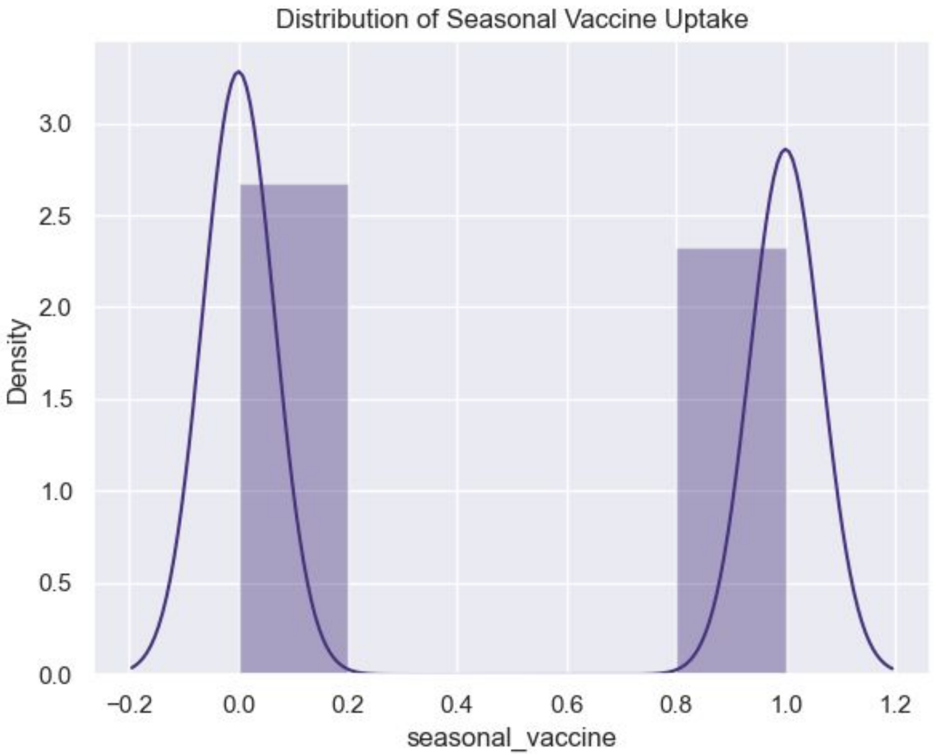
Methodology



EXPLORATORY DATA ANALYSIS

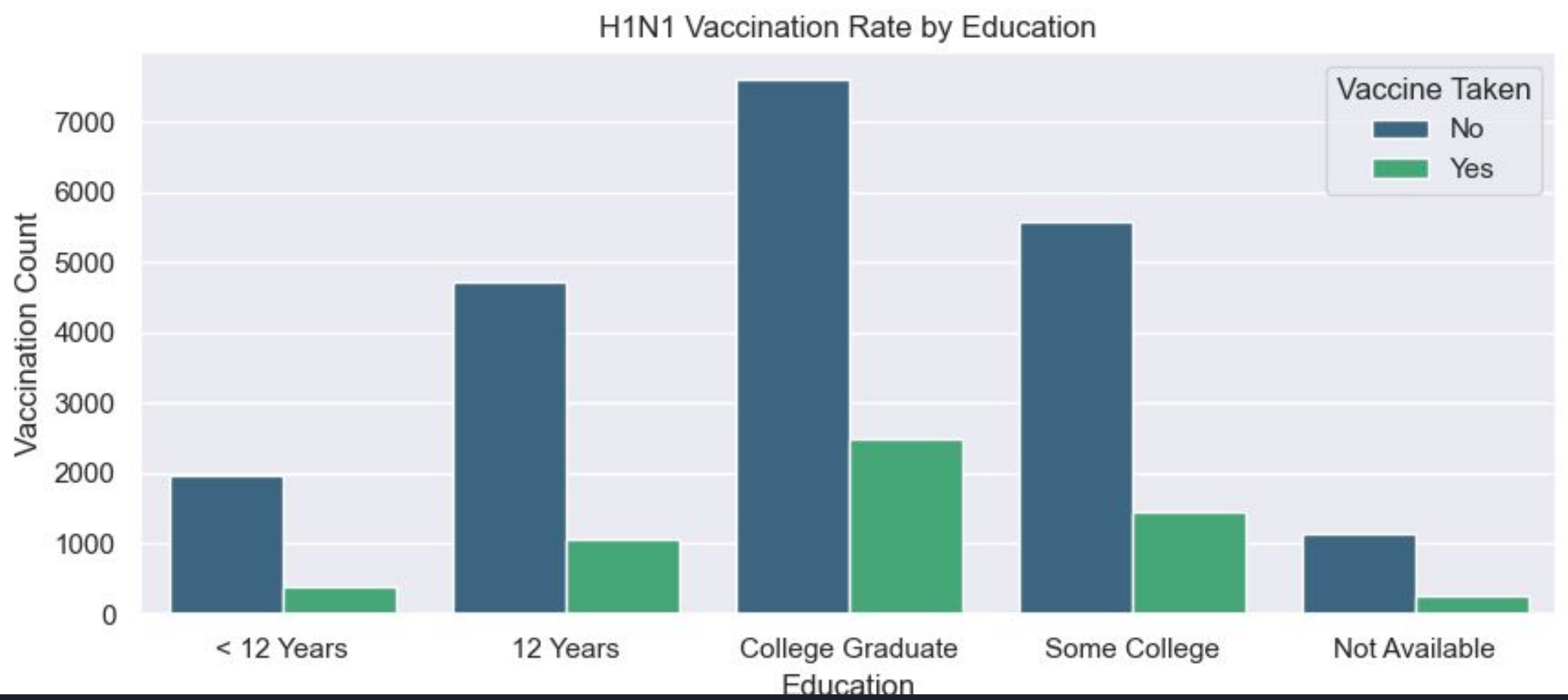
Distribution h1n1_vaccine and seasonal_vaccine



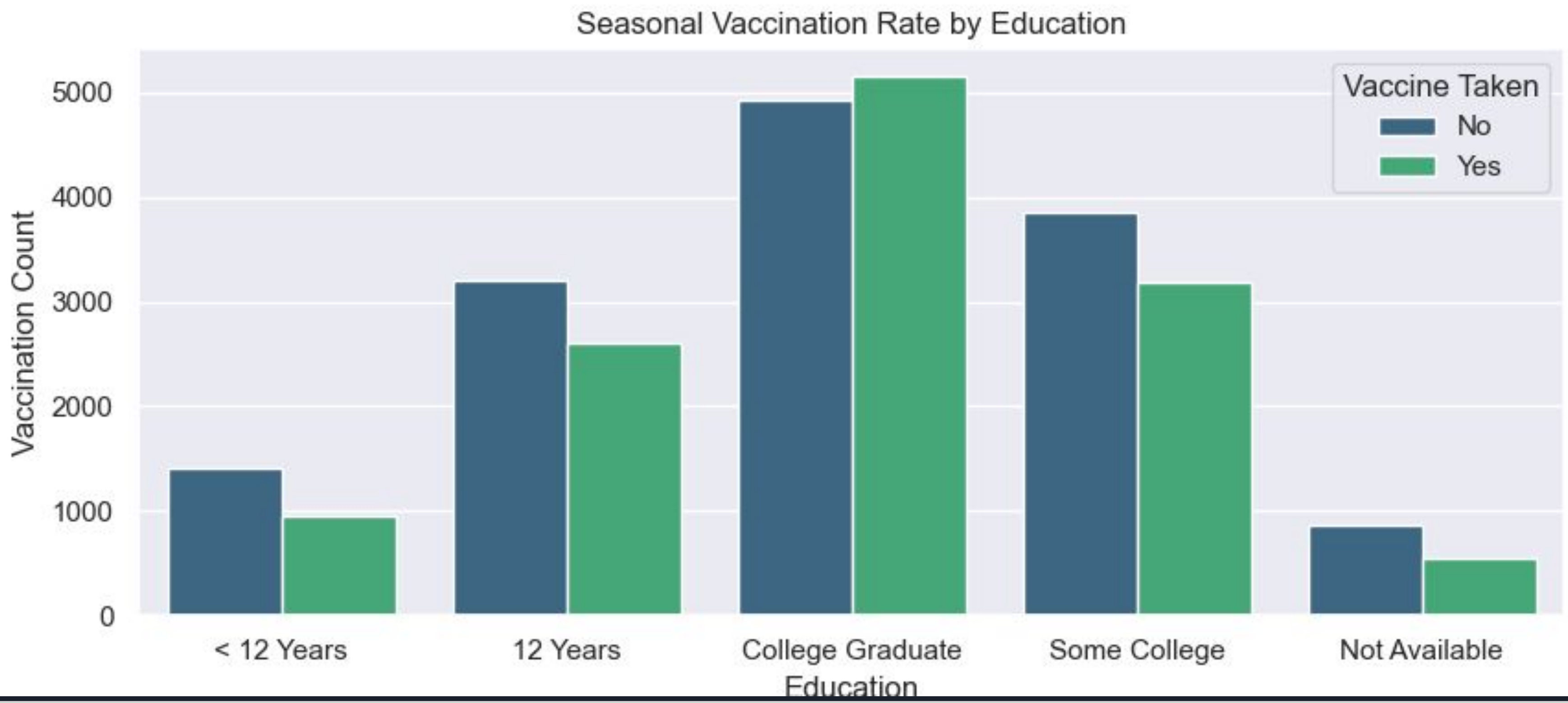


The vaccine uptake variables have slightly right-skewed distributions with light tails, showing that both H1N1 and seasonal vaccine uptake have fewer outliers and more spread out values

Does education affect the in_take of H1N1 vaccine ?

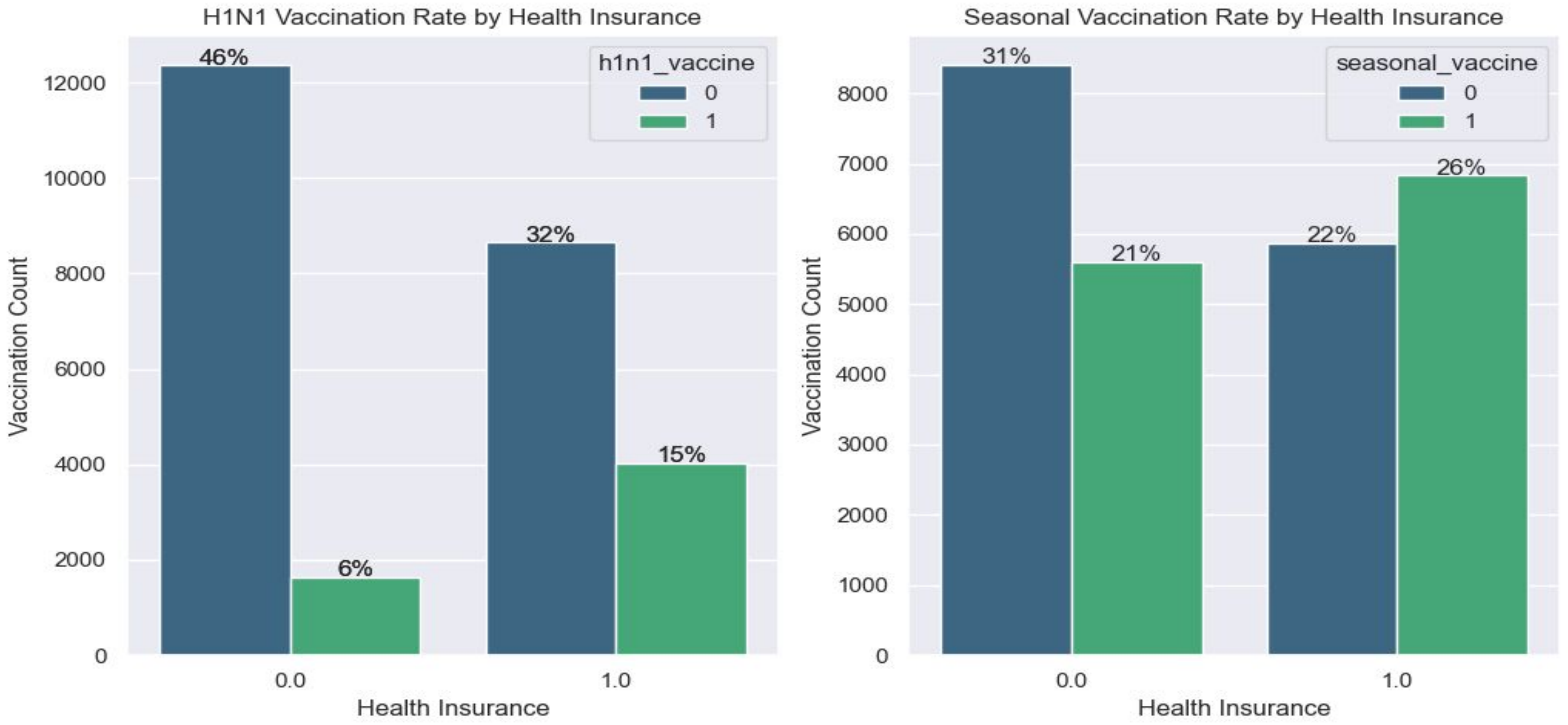


Does education affect the in_take of seasonal vaccine ?

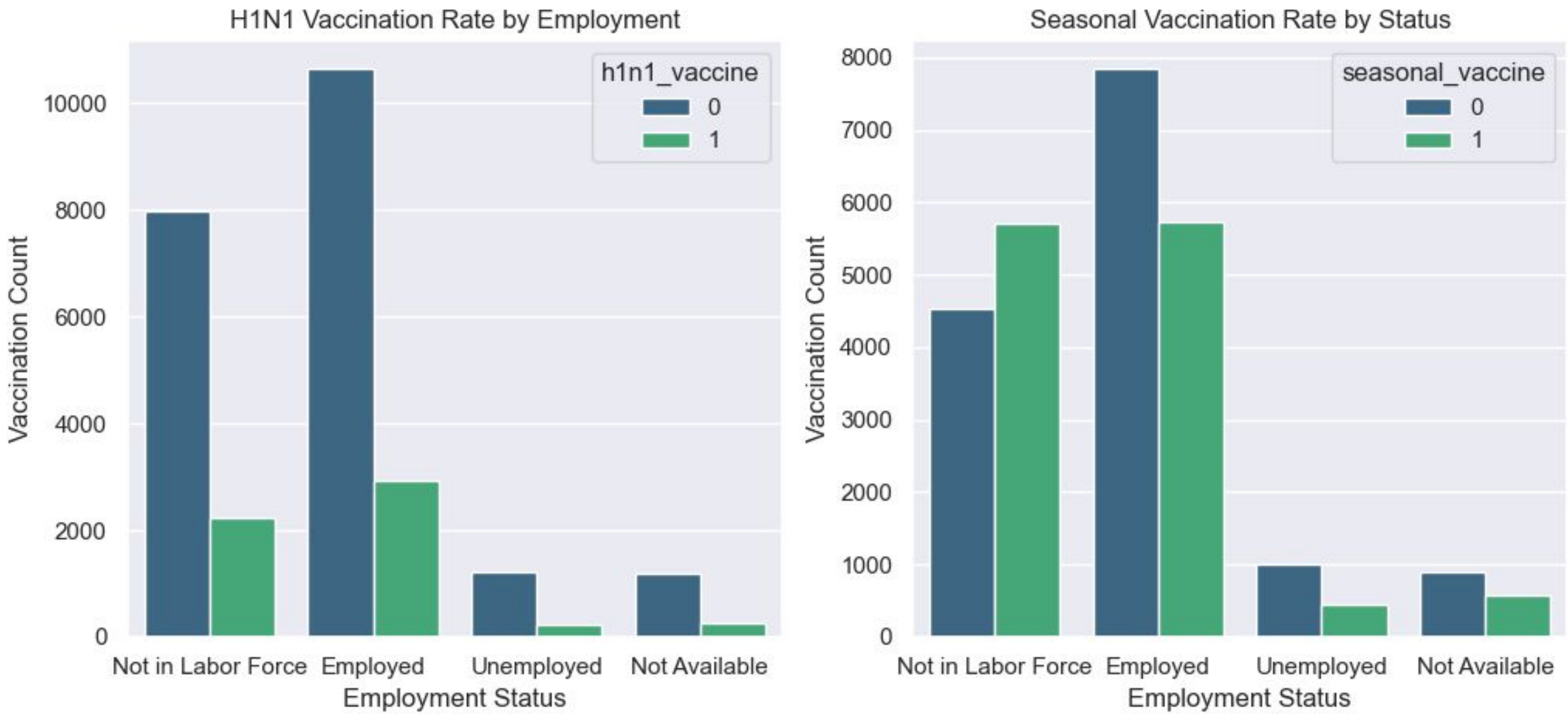


Individuals with higher levels of education tend to have higher vaccination rates for both H1N1 and seasonal vaccines, as seen from the bar plots of vaccine uptake by education.

Does Insurance affect the intake of H1N1 vaccine and seasonal vaccine ?



Does employment_status affect the intake of H1N1 vaccine and seasonal_vaccine ?



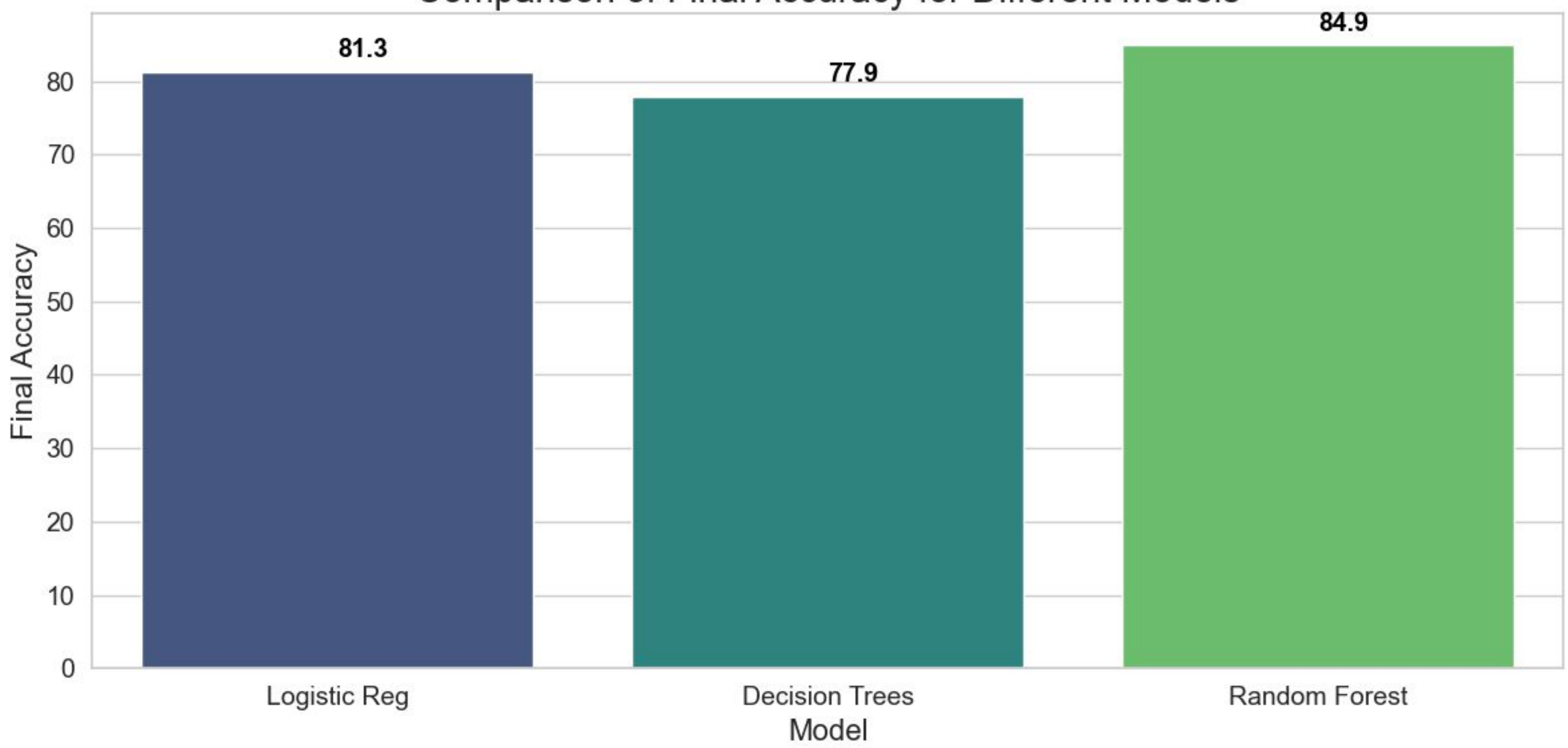
The bar plots indicate that employment status affects vaccine uptake, with employed and not in labor force individuals having higher rates of h1n1 and seasonal vaccine uptake compared to the unemployed. This suggests that employment status plays a role in vaccine uptake

MODELLING AND EVALUATION

Methodology for Modelling

- Methodology: Data preparation for machine learning models by splitting into training and testing sets, scaling numerical features using StandardScaler, and encoding categorical features using OrdinalEncoder
- Testing of three machine learning algorithms: Logistic Regression, Decision Trees, and Random Forest
- Evaluation of accuracy, precision, recall, and final accuracy of each model
- Final choice of Random Forest as the best performer with an accuracy of 84.9% for H1N1 and 78.5% for the seasonal flu vaccine. Overall entropy accuracy for both datasets was 0.78.

RESULTS



Looking at the final accuracy of the models, it seems that the Random Forest model performed the best with an accuracy of 81.67%, followed by Logistic Regression with an accuracy of 81.32%, and Decision Tree with an accuracy of 77.90%.

CONCLUSION

The **final models** for predicting H1N1 and seasonal flu vaccine include a Decision Tree, Random Forest, and Logistic Regression. The **Random Forest model** achieved the **highest accuracy with 84.9%** for H1N1 and 78.5% for the seasonal flu vaccine, while the Logistic Regression model had an accuracy of 81.3% for both datasets. The Decision Tree model had an accuracy of 77.9%. Precision and recall scores varied between the models and datasets. The entropy accuracy for both datasets combined was 0.78

RECOMMENDATION

1. Targeted Education and Communication: Focus vaccine education and communication efforts towards groups identified by the models as being hesitant or reluctant to vaccinate.

2. Tailored Messaging and Strategies: Develop tailored messaging and strategies that address specific concerns and barriers identified by the models.

3. Effective Interventions: Implement interventions such as reminder systems, incentives, and reducing access barriers, that have been shown to be effective in improving vaccination rates.

4. Conduct further research to identify and address factors that may not have been captured by the current models.

5. Use the models to forecast vaccine demand and allocate resources effectively to ensure that vaccines are available to those who need them.

6. Continuously monitor vaccination rates and adjust interventions as needed to ensure that vaccination goals are being met.