A FLU-SHOT VACCINE MACHINE LEARNING PROJECT

BY GERALD MWANGI

Overview

Vaccines are essential for disease prevention, providing immunity against communicable diseases like COVID-19, swine flu, influenza, and Tuberculosis. The first vaccine was developed in 1796 by Edward Jenner. Today's vaccines have evolved due to advancements in technology and health industry research. Measures are put in place to ensure quality, efficiency, and distribution. The H1N1 vaccine, approved in 2010, was developed in 2009 after the swine flu pandemic, killing over two million people.
Vaccines are essential for disease prevention, providing immunity against communicable diseases like COVID-19, swine flu, influenza, and Tuberculosis. The first vaccine was developed in 1796 by Edward Jenner. Today's vaccines have evolved due to advancements in technology and health industry research. Measures are put in place to ensure quality, efficiency, and distribution. The H1N1 vaccine, approved in 2010, was developed in 2009 after the swine flu pandemic, killing over two million people.

Problem statement

Vaccines are critical for preventing communicable diseases such as COVID-19, swine flu, seasonal influenza, and tuberculosis. However, getting people to get vaccinated is still a huge challenge. The CDC seeks to understand how individual features influence uptake of the vaccines so as to provide guidance for future public health efforts

Project objectives

. To predicting H1N1 and Seasonal Vaccines uptake

. To determine distribution of vaccine uptake.

. To determine correlation between vaccine uptake.

. To identify characteristics influencing vaccine uptake.

Data source

Data was downloaded from https://www.drivendata.org/competitions/66/flu-shot-learning/data/ which was a phone survey done in 2009 courtesy of the United states National Center for Health Statistics

Data Description

Features

These data set contained features to be used for predicting our target variable

These features are sourced from respondents reply during the phone survey

Labels

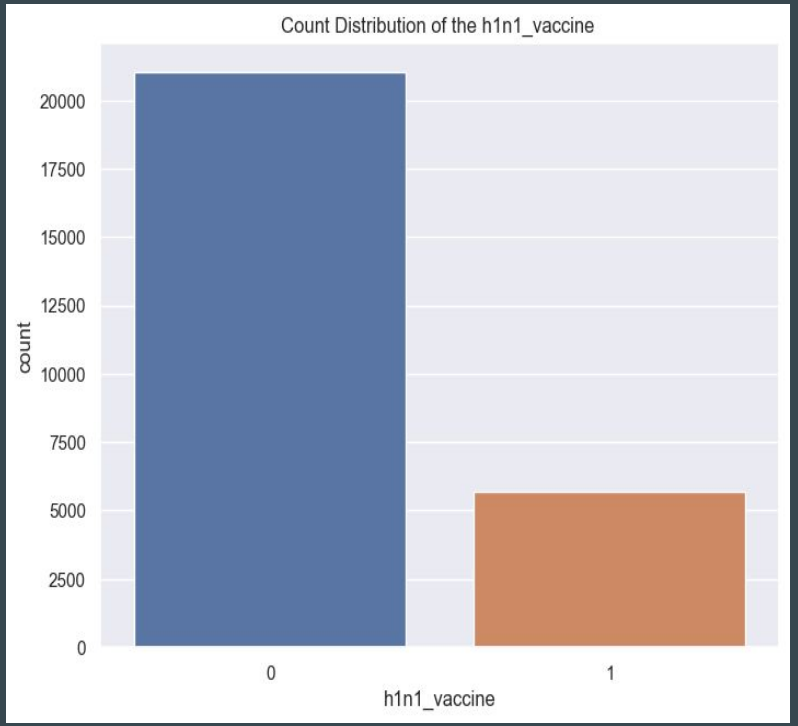
These data set contain our target variables H1N1 and seasonal vaccine take encoded in binary form [0,1]

Data preparation

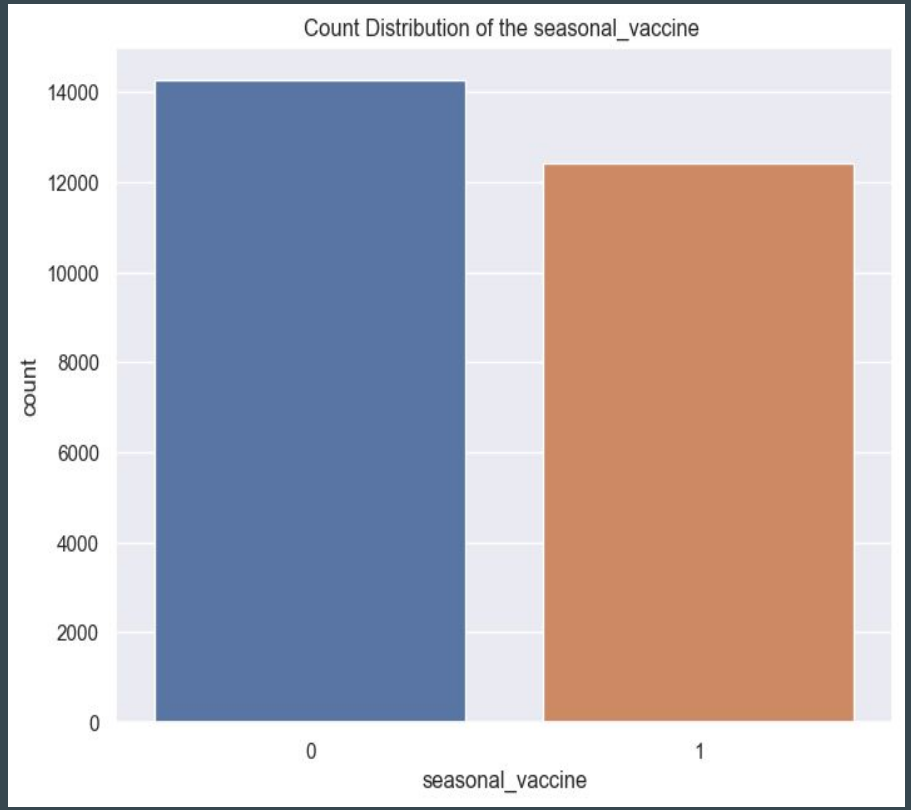
Handled missing values by filling them with median(for numerical values) ,mode(for categorical values) and dropping columns which had more than 40% missing values

Vaccine Distribution

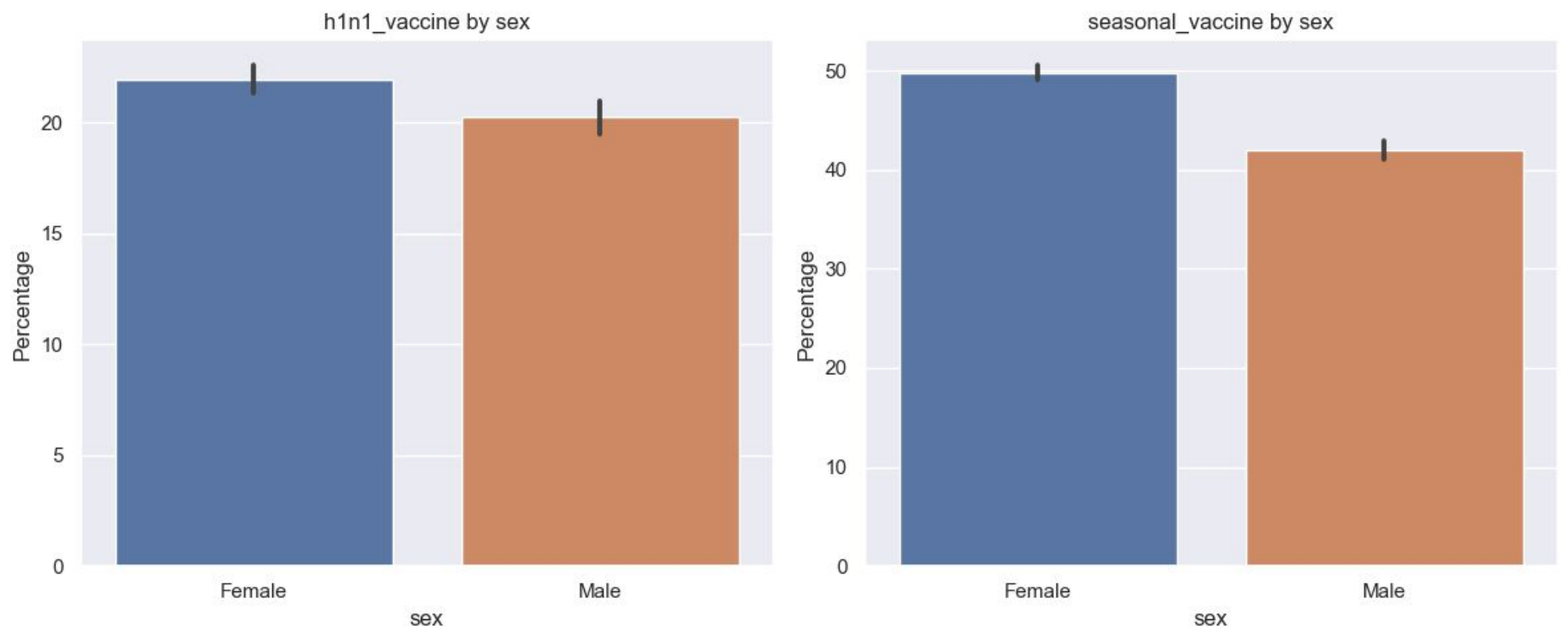
The H1N1 vaccine had the largest population of people who were not vaccinated with it



Seasonal vaccine had a larger population of people who were vaccinated with it

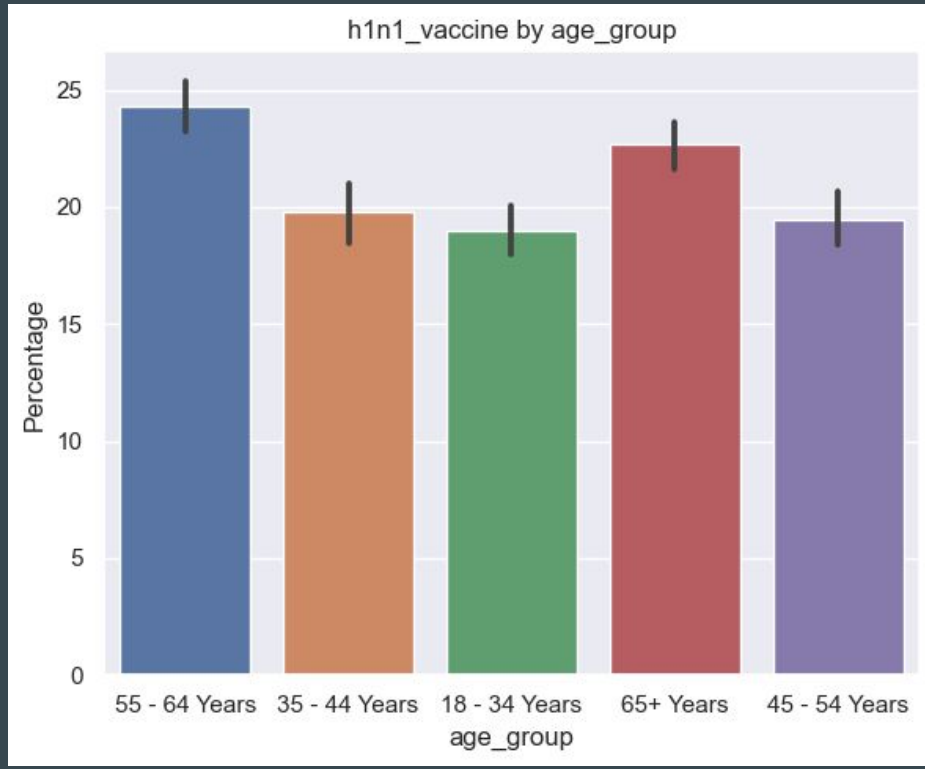


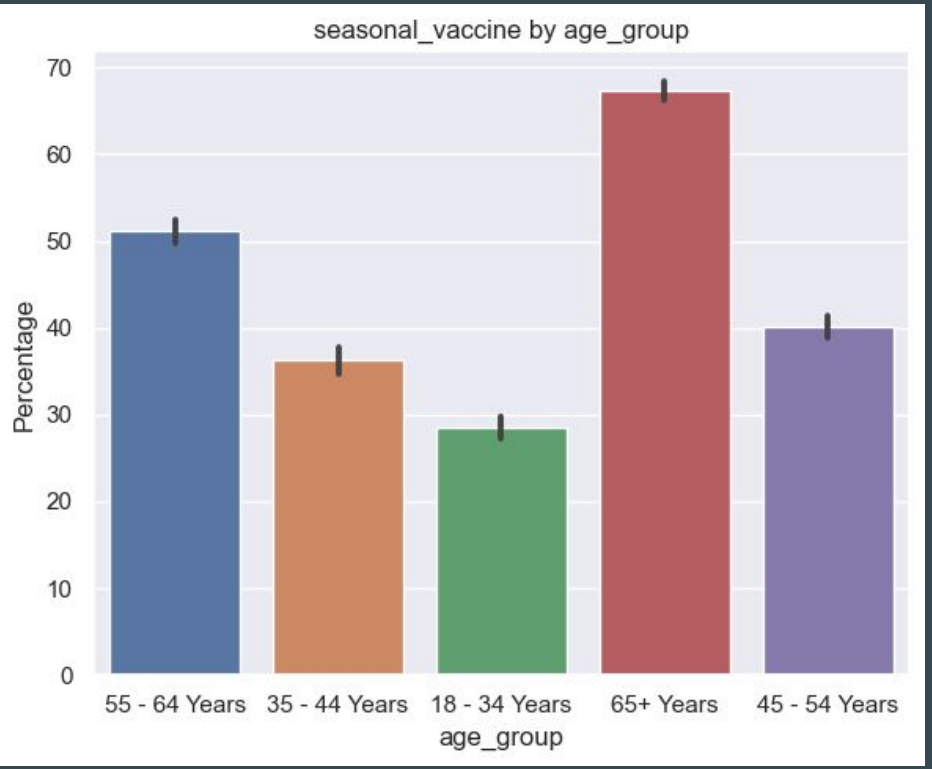
Gender vs Vaccine



Seems to be more vaccinated women than men

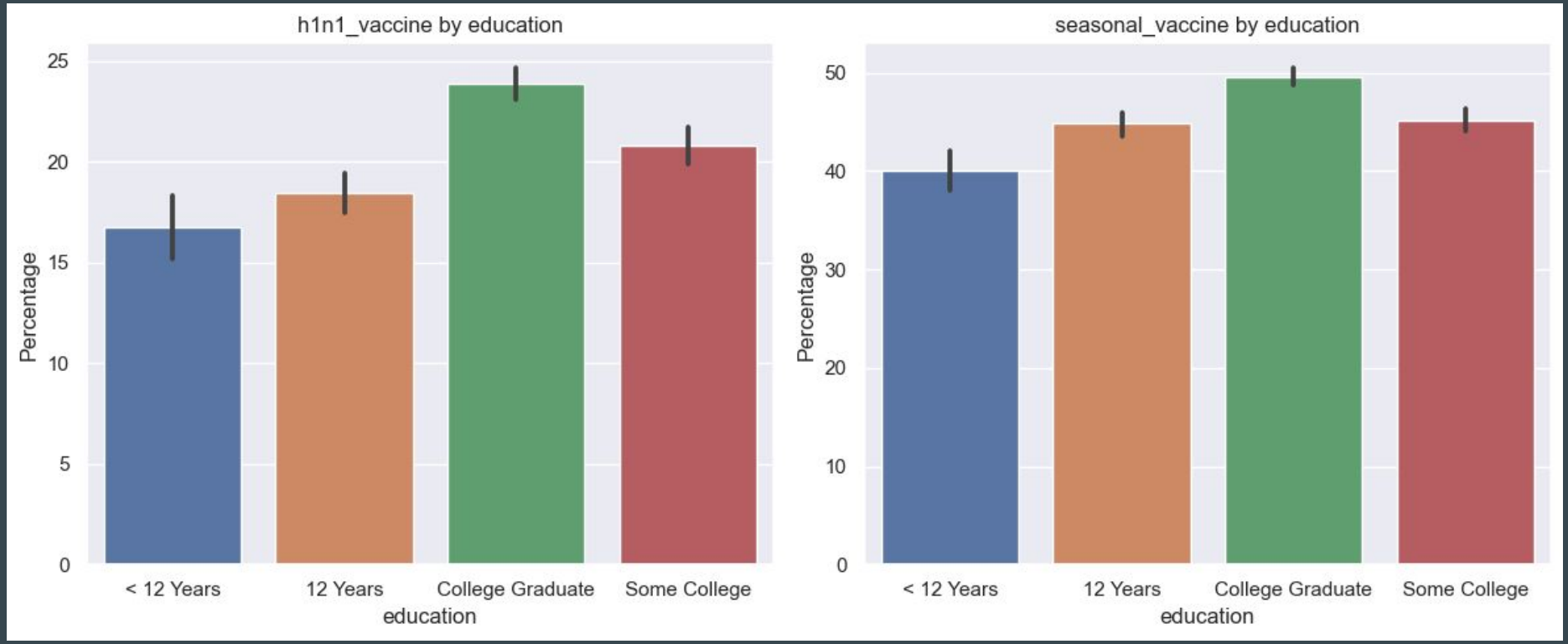
Age vs Vaccine





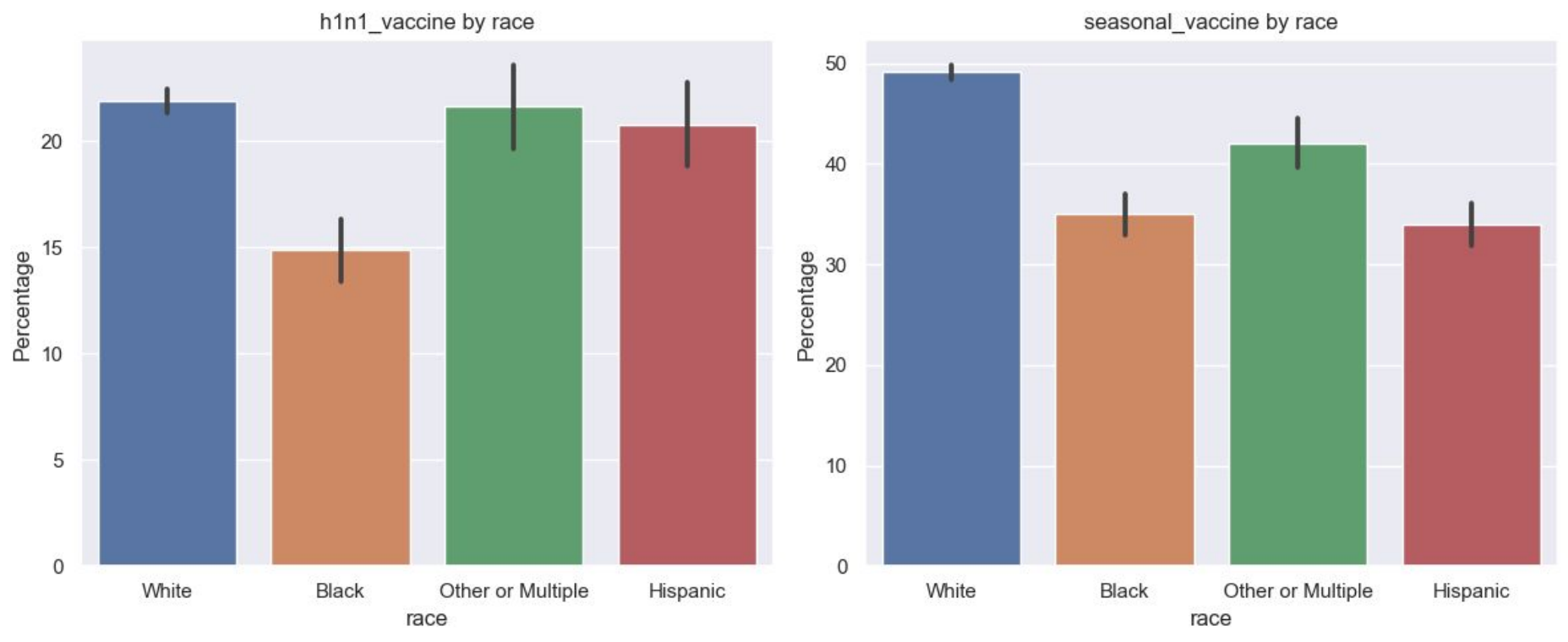
Most of the vaccinate age group was the older generation above age of 55 years

Education vs Vaccine



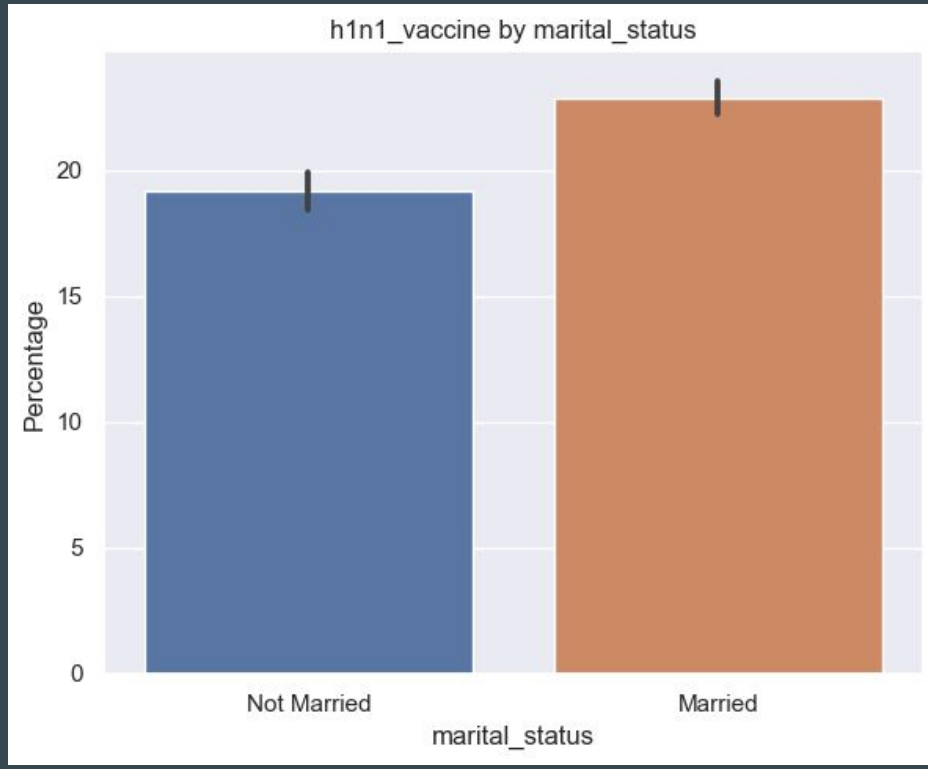
College graduates seem to be the majority taking the vaccines

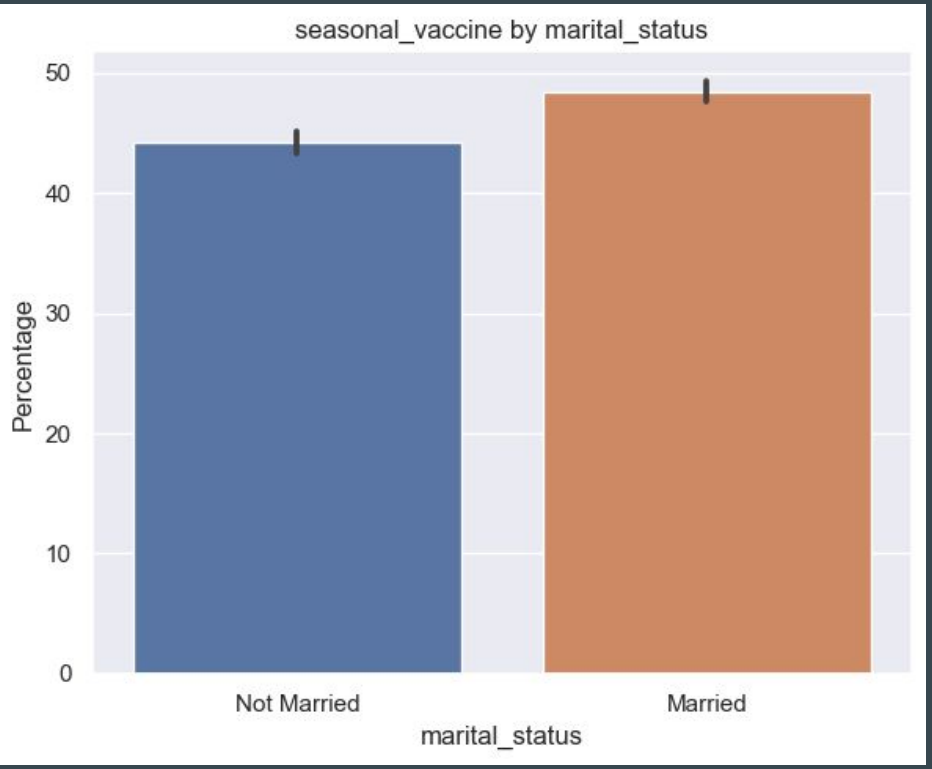
Race vs Vaccine



The white race seemed to be the majority taking the vaccines

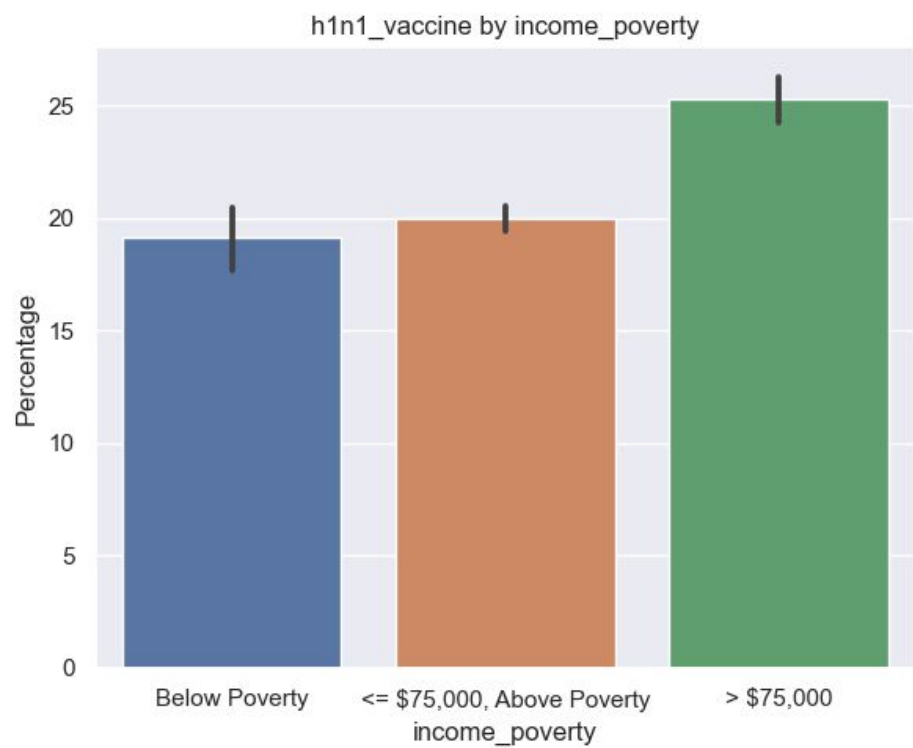
Marital status vs Vaccine

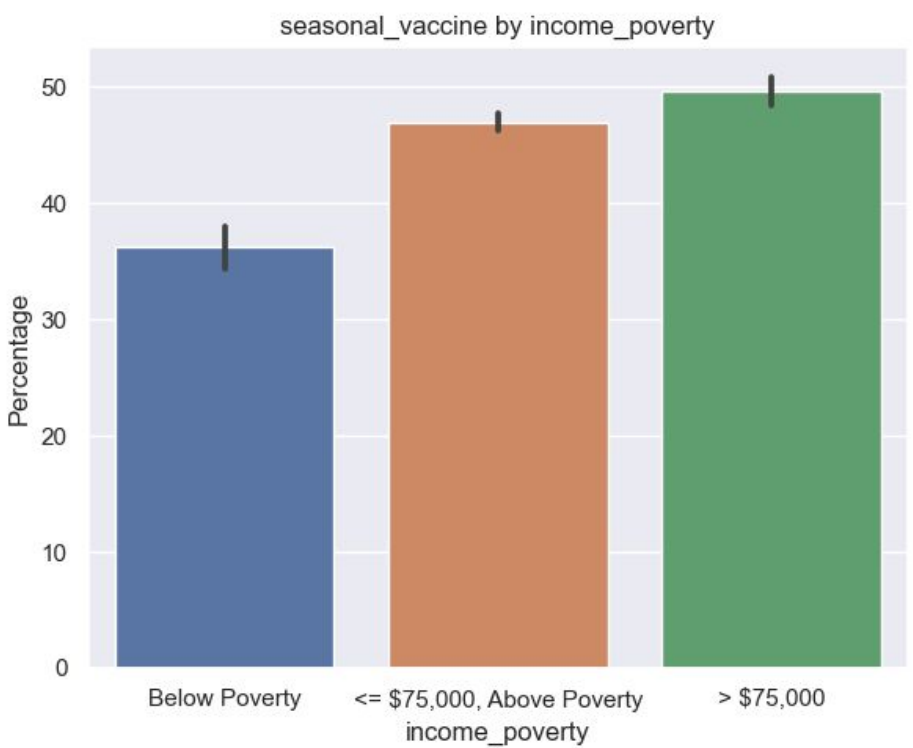




Majority of the population taking the vaccines are married

Income level vs Vaccine





Majority of the population who are vaccinated earn more than $75000 per year

Modelling

We created 5 models and evaluated them to find the best fitting models for our target variables

MODEL	ACCURACY H1N1	ACCURACY SEASONAL	AUC H1N1	AUC SEASONAL
Decision tree	0.59	0.67	0.6	0.68
Logistic regression	0.82	0.78	0.64	0.78
KNN	0.63	0.71	0.64	0.71
Naïve Bayes	0.65	0.74	0.69	0.74
Random Forest	0.63	0.77	0.65	0.78
Tuned KNN	0.83	0.78	0.65	0.77

Evaluation

With an AUC score of 0.65 and an accuracy of 0.83, the Tuned KNN model is the best model for predicting the "h1n1" class while the Random Forest model is the best option for predicting the "seasonal" class because it has an accuracy of 0.77 and AUC score of 0.78,

Conclusions

- Majority of respondents have received seasonal vaccines, more than H1N1.
- Majority are female, 55 years and above.
- Most have higher education levels, mostly college graduates.
- Majority are white, married, earn over $75000 a year.
- H1N1 and seasonal flu vaccinations are often combined due to weak correlation.
- Tuned KNN model is the best for predicting H1N1 vaccine and random forest for seasonal vaccine.

Recommendations

- Targeting females, people aged 55 and over, higher education students, and high income earners for vaccination programs.
- Monitoring racial discrepancies through public education and awareness campaigns.
- Emphasizing the importance of taking both vaccines through targeted messaging.
- Utilizing Tuned KNN and Random Forest models for accurate H1N1 vaccine prediction.
- Conducting campaigns to educate people with high disease risk to take the two vaccines.

Next Steps

. More updated data is required to train our models for optimum performance

. Collecting individual data for the two target variables so as to create better models for each variable

. Good quality data to avoid misguidance from the models

THANK YOU

ANY QUESTIONS ?