Predicting Seasonal Flu Vaccine Uptake Using Demographic, Health, and Opinion Data

Presented by: Yvonne Mwangi

19th May 2024

Objective and Goals

Objective

- Develop a machine learning model to predict the likelihood of individuals receiving their seasonal flu vaccines based on their demographic information, health status, and opinions on health risks and vaccine effectiveness. The model aims to identify key predictors of vaccine uptake to inform targeted public health interventions and improve vaccination rates.

- Goals of Data Analysis and Prediction:
  - **Identify Key Predictors**: Determine which socioeconomic factors are most strongly associated with the likelihood of receiving a seasonal flu vaccine.
  - **Model Development**: Build a predictive model that accurately estimates the probability of an individual receiving the vaccine based on their socioeconomic characteristics.
  - **Inform Public Health Strategies**: Provide actionable insights through reports and dashboards that can be used by public health officials to design targeted interventions aimed at increasing flu vaccination rates, particularly in underserved or high-risk populations.
  - R**esource Allocation**: Help allocate resources more effectively by identifying demographic groups that are less likely to receive the vaccine and may benefit from additional outreach and support.

Data Understanding

- The original data set contained:
  - Training Features Dataset
    - Demographic, health, and opinion data of respondents.
    - 26707 rows and 36 columns
  - Training Labels Dataset
    - Labels indicating seasonal vaccine uptake.
    - 26707 rows and 3 columns (respondent_id, h1n1_vaccine and seasonal_vaccine)
    - I focus on Seasonal vaccine for this project
  - Testing Features Dataset
    - Demographic, health, and opinion data for prediction.
    - 26708 rows and 36 columns

Data Preparation – Features used

- **Vaccine feature**
  - Seasonal vaccine
- **Demographic Features**
  - Age Group
  - Education
  - Race
  - Sex
  - Income Relative to Poverty
  - Marital Status
  - Rent or Own
  - Employment Status
- **Health and Opinion Features**
  - Doctor Recommendation for Seasonal Vaccine
  - Health Insurance
  - Opinion on Seasonal Vaccine Effectiveness
  - Opinion on Risk of Sickness
  - Opinion on Sickness from Vaccine

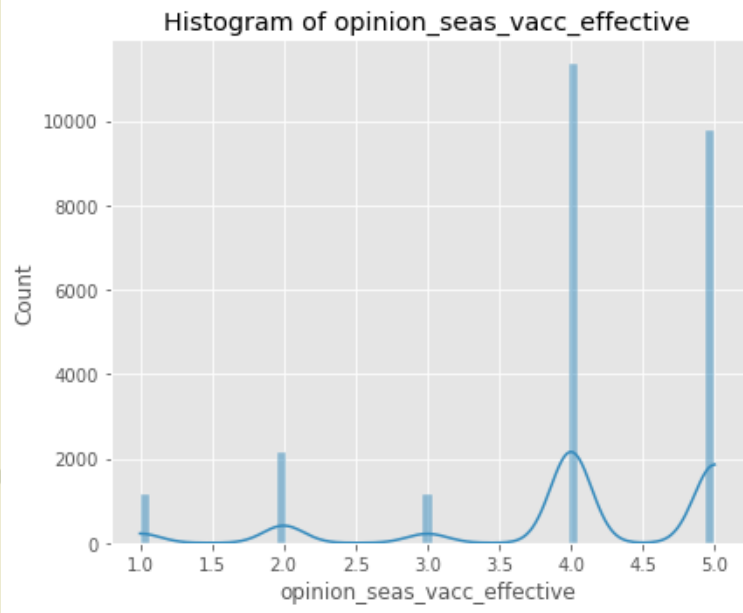
Steps taken:

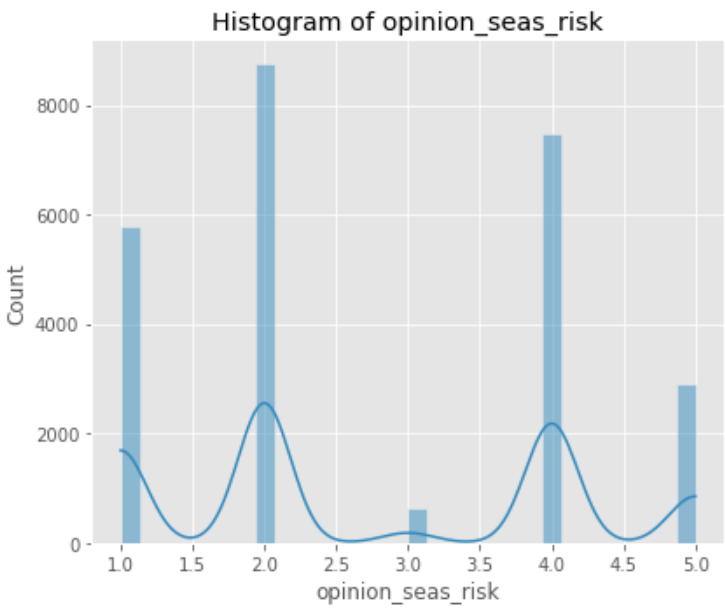
- Merged the labels and training feature data sets
- Cleaned Data
  - Handled Missing data (mode , category creation)
  - Checked for Duplicates
- Formatted the data
- Result - 25,548 rows and 17 columns

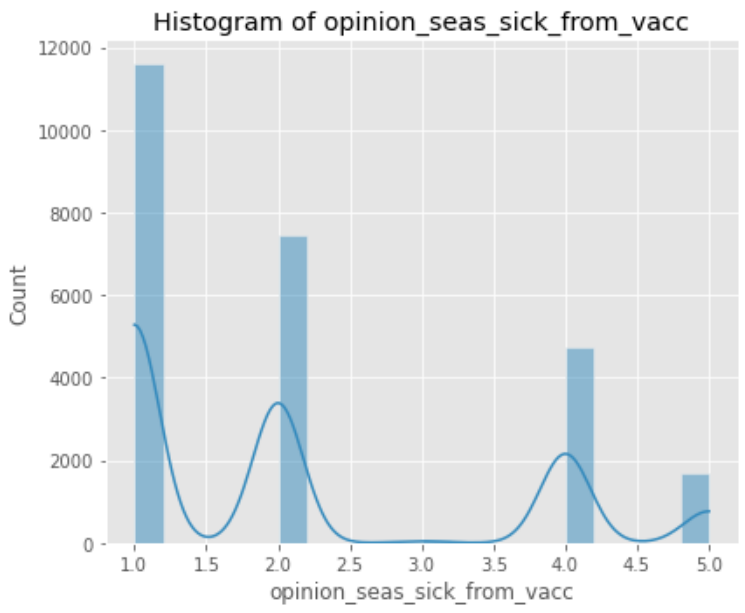
Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Univariate Analysis



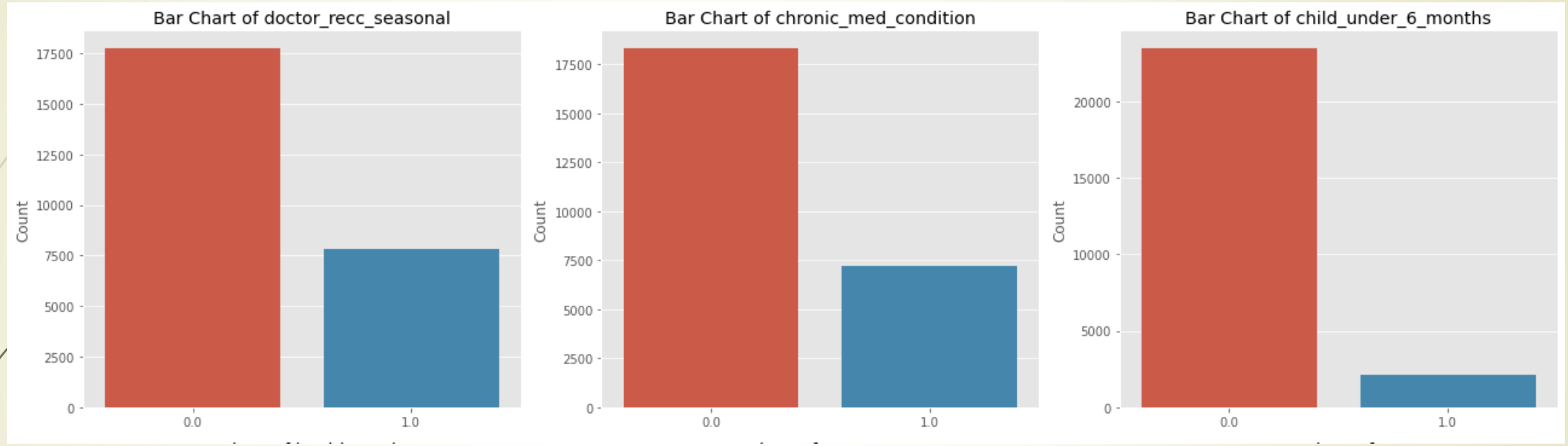




- Vaccine Effectiveness: Most respondents trust the vaccine's effectiveness.
-
- Risk: Opinions on the risk of falling sick by not getting vaccinated are varied, with a significant portion recognizing some level of risk.

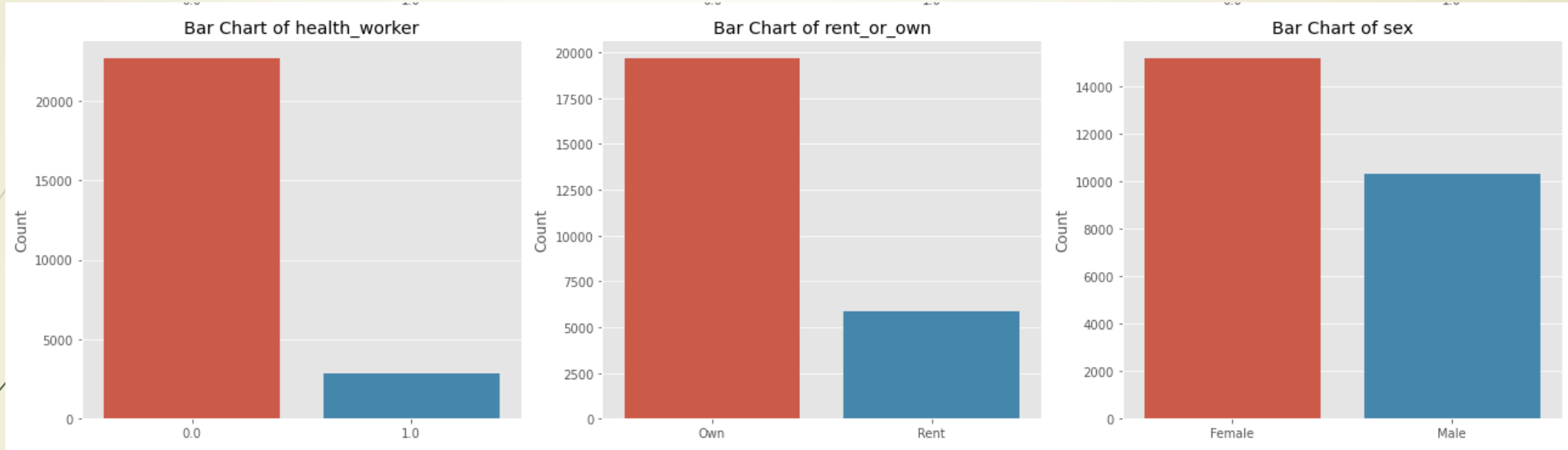
- Sickness: The majority believe that the vaccine does not cause sickness.

Univariate Analysis



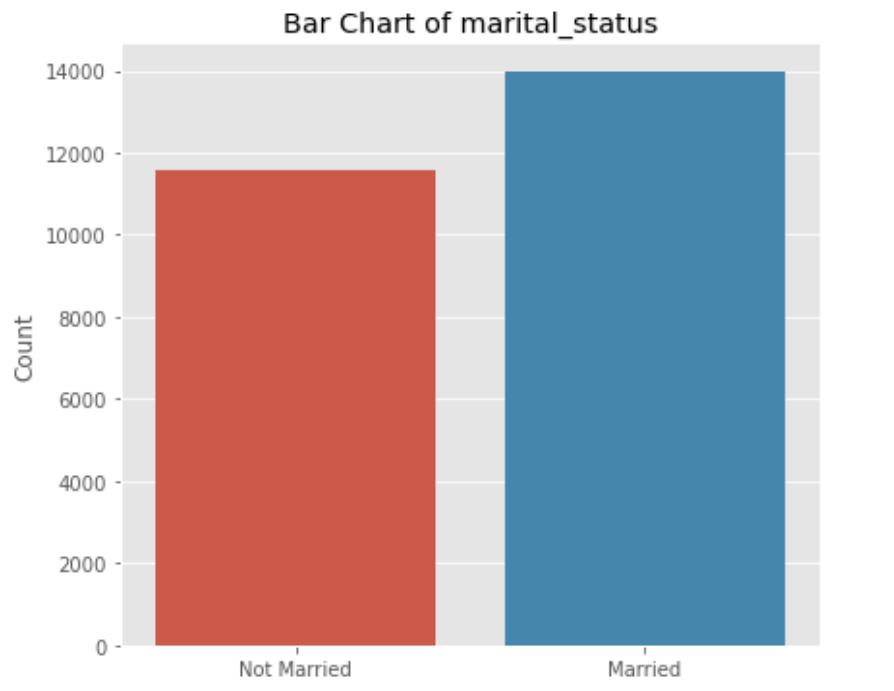
- Doctor Recommendations: More respondents did not receive a recommendation for the seasonal vaccine than those who did.
-
- Chronic Medical Conditions: The majority of respondents do not have chronic medical conditions.
-
- Children Under 6 Months: Most respondents do not have children under 6 months.

Univariate Analysis



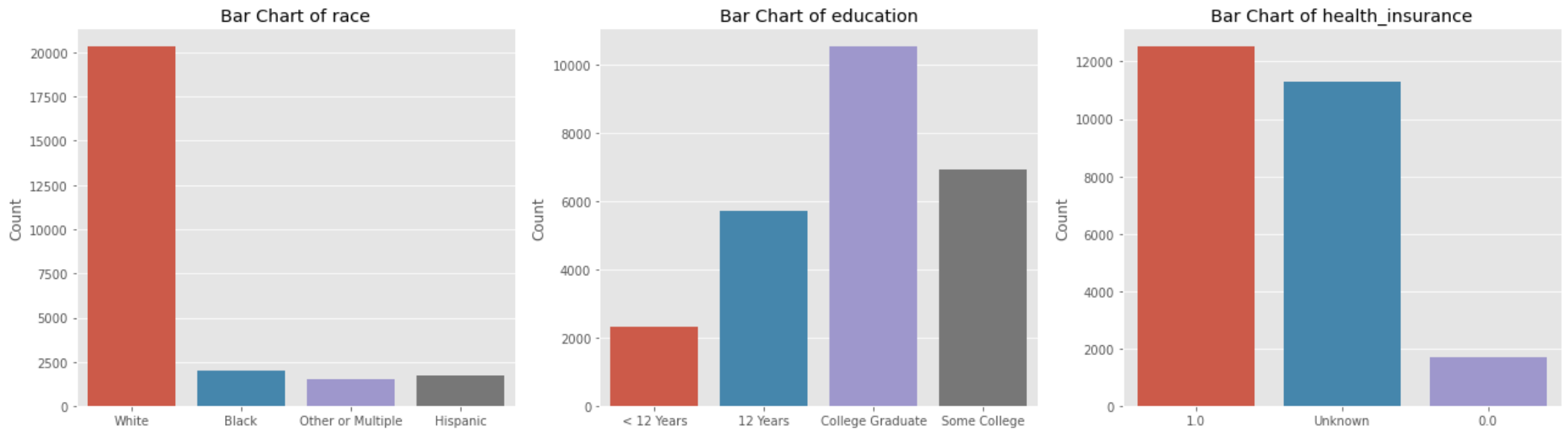
- Health Workers: A small portion of respondents are health workers.
-
- Home Ownership: The majority of respondents own their homes.
-
- Sex: There are more female respondents than male respondents.

Univariate Analysis



Marital Status: More respondents are married compared to those who are not married.

Univariate Analysis

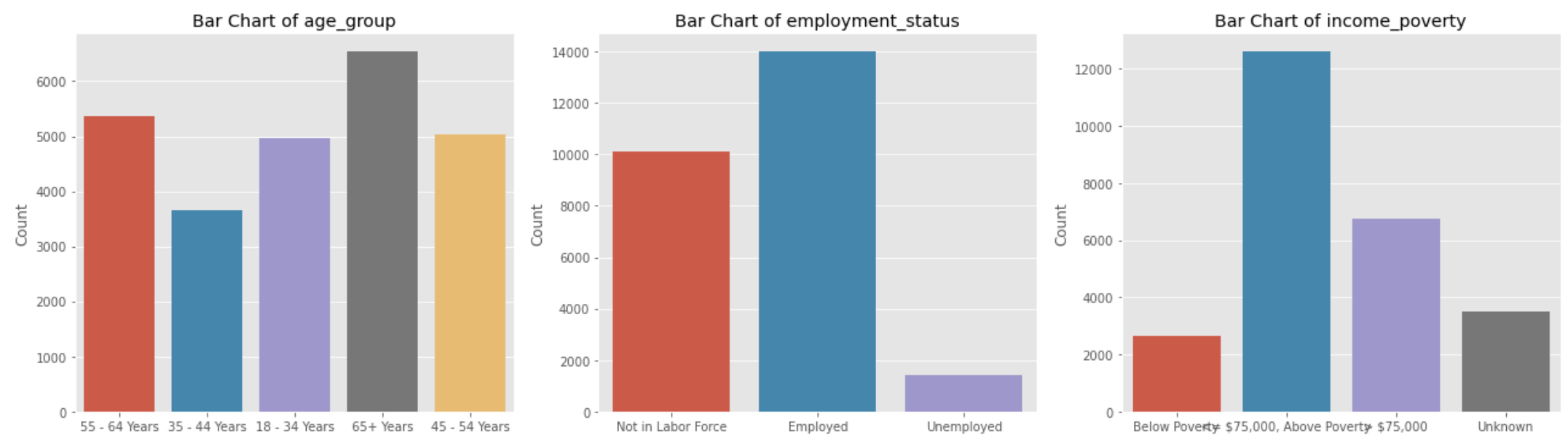


Race: The majority of respondents are White, with smaller representations from Black, Hispanic, and other races.

Education: Most respondents are college graduates, with significant groups having some college or 12 years of education.

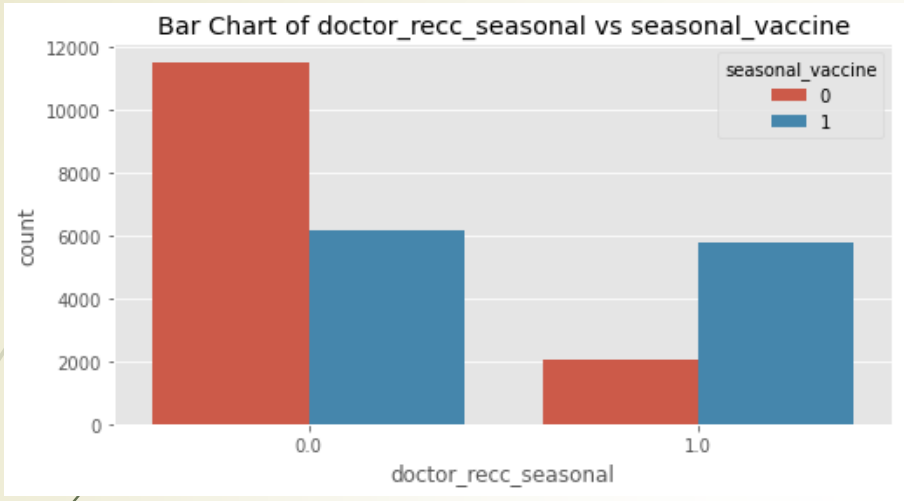
Health Insurance: Many respondents have health insurance, though a substantial number have unknown insurance status.

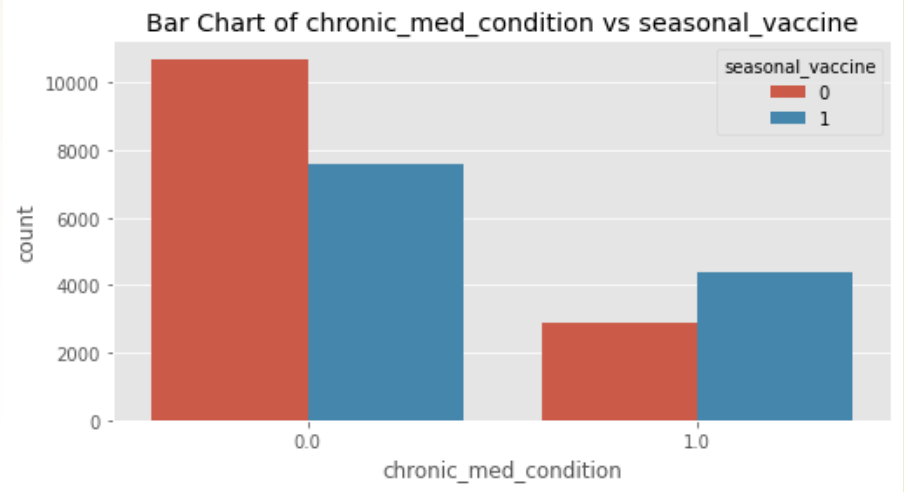
Univariate Analysis

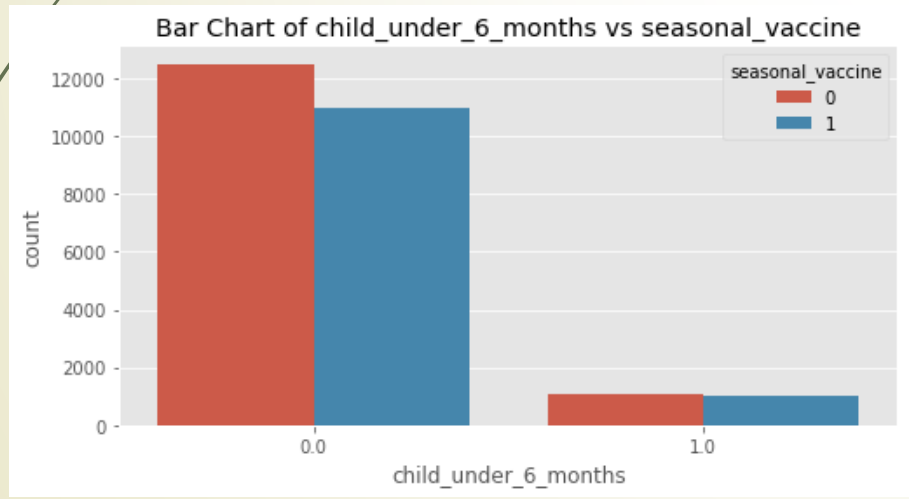


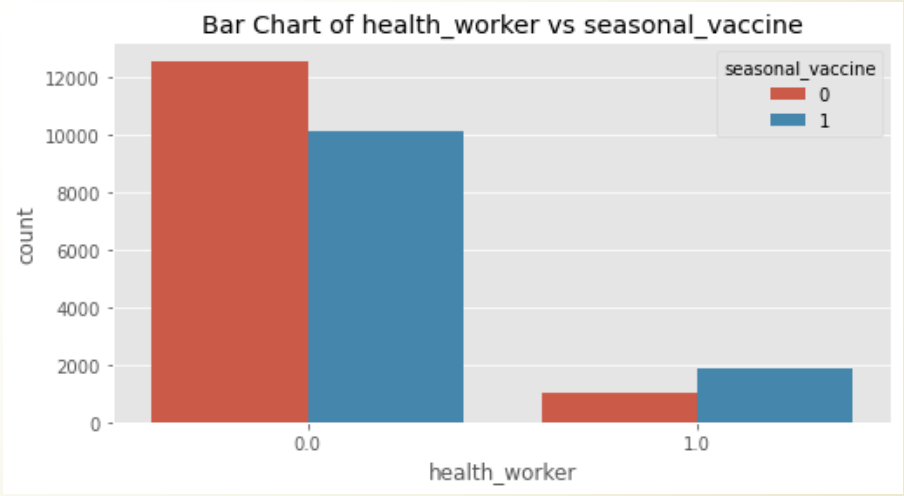
- Age Group: Respondents are distributed across various age groups, with the largest group being 65 years and older.
- Employment Status: The majority of respondents are employed, with significant groups not in the labor force.
- Income and Poverty: Most respondents have an income above the poverty line, with a significant number earning less than $75,000.

Bivariate Analysis

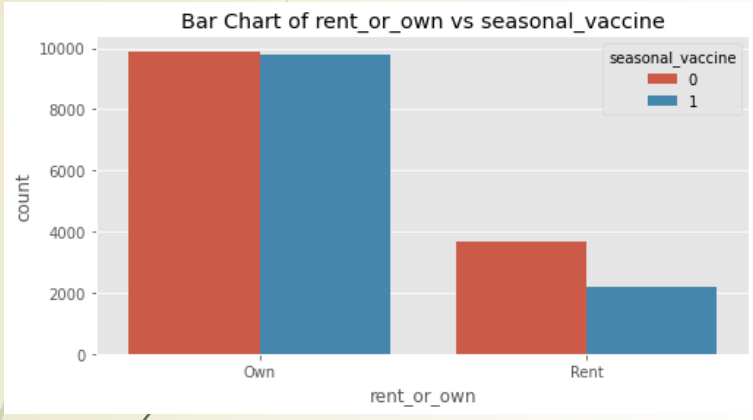


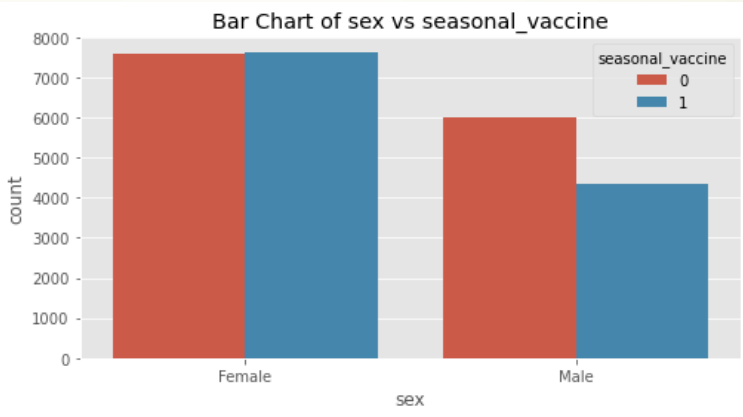


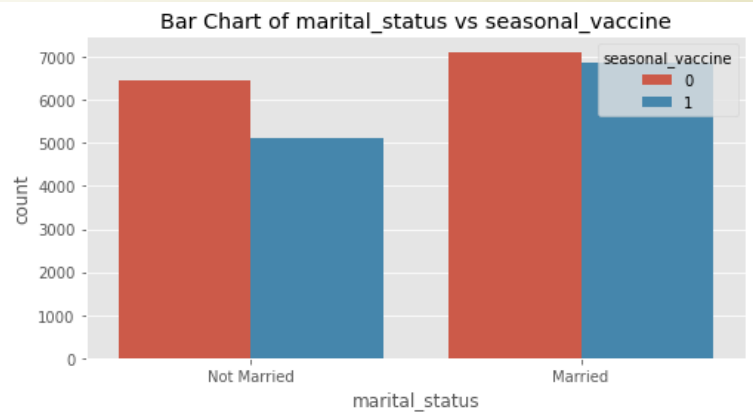


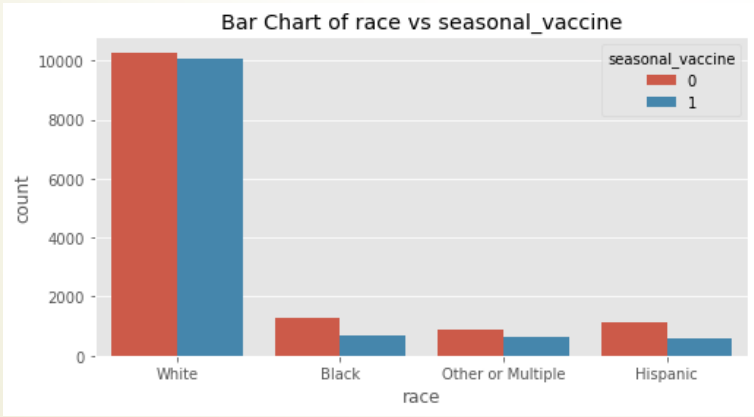


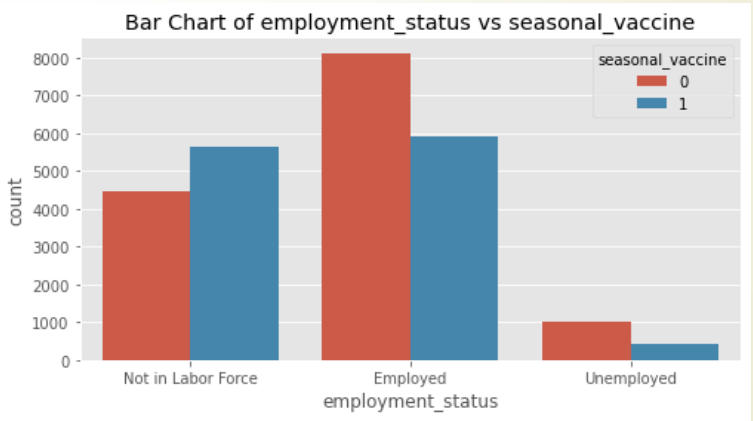
Bivariate Analysis



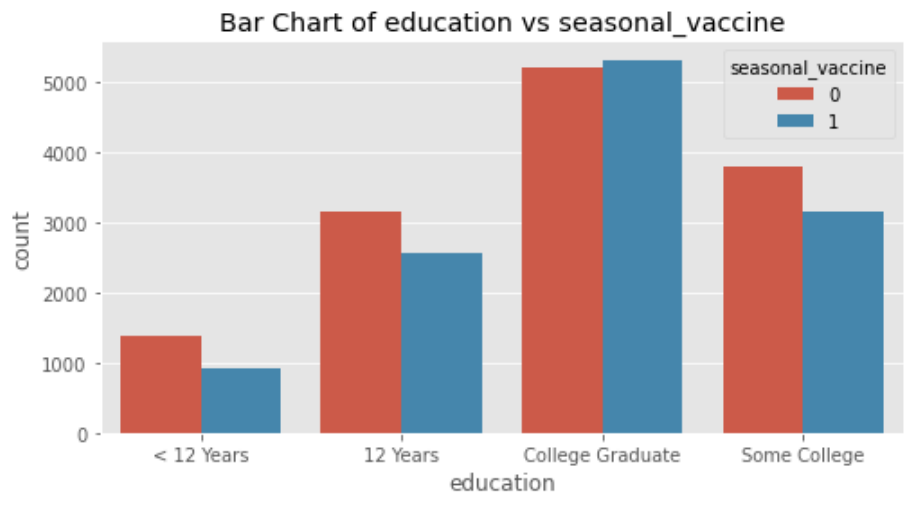


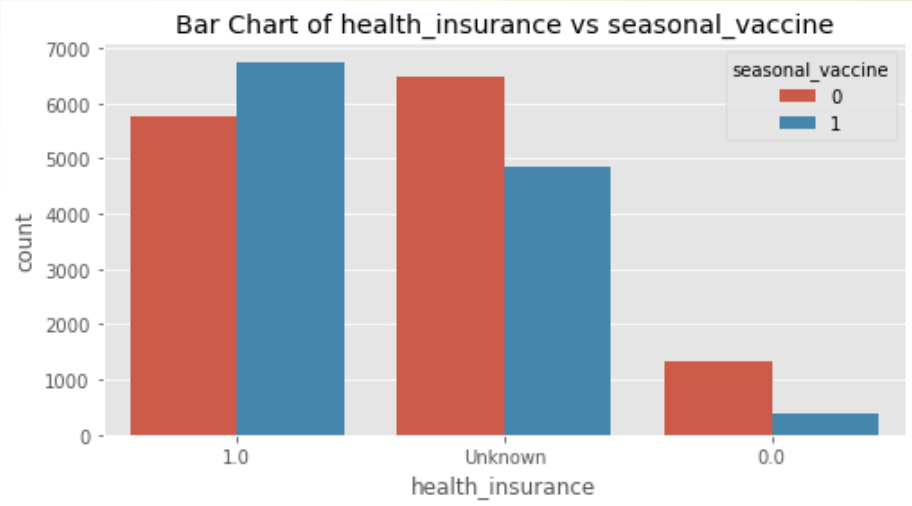


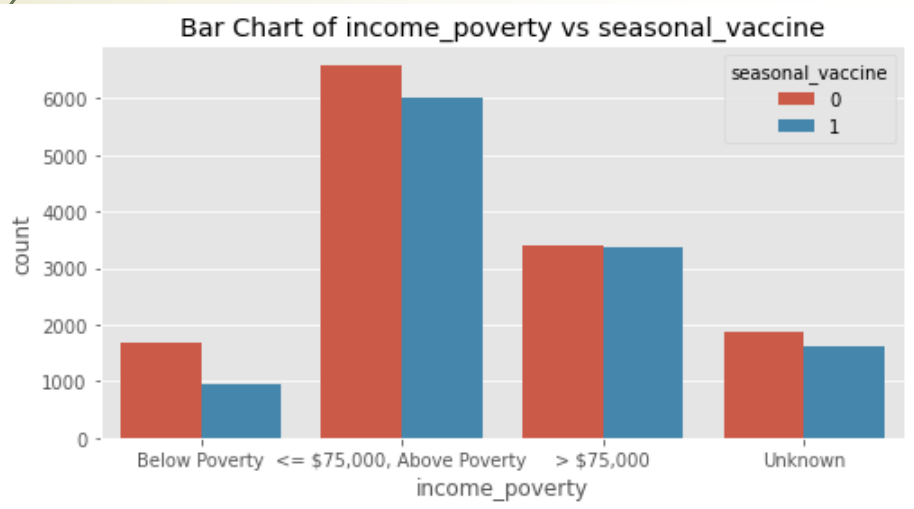


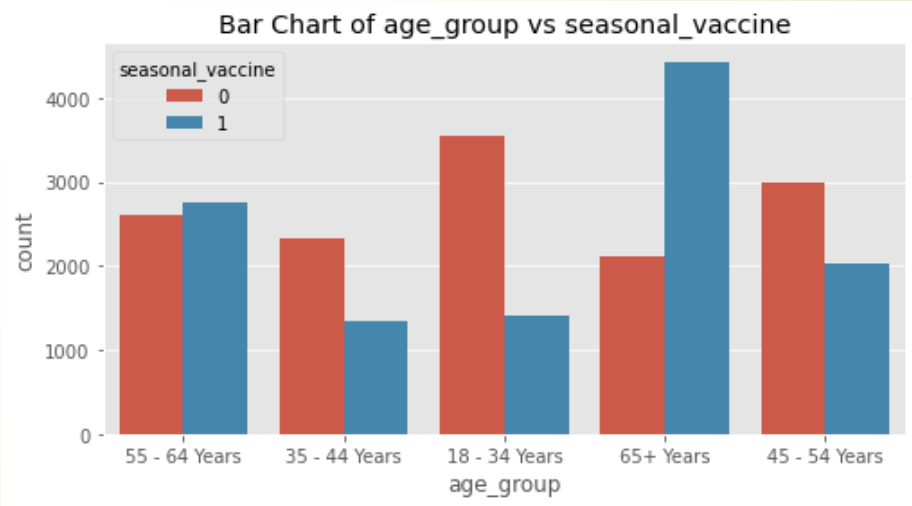


Bivariate Analysis









Bivariate Analysis – Categorical Variables

Variable	Chi-square Statistic	p-value
Strongest Predictors		
Doctor_recc_seasonal	3311.98	0.0
Chronic_med_condition	743.98	8.17e-164
Employment_status	613.33	6.58e-134
Health_insurance	721.77	1.86e-157
Age_group	2095.26	0.0
Health_worker	415.40	2.45e-92
Rent_or_own	290.66	3.58e-65
Race	292.51	4.16e-63
Significant Predictors		
Sex	163.84	1.64e-37
Marital_status	58.67	1.86e-14
Education	115.81	6.17e-25
Income_poverty	145.72	2.21e-31
Weak Predictor		
Child_under_6_months	3.85	0.0499

Strongest Predictors with very strong and highly significant relationships with receiving the seasonal vaccine:

- Doctor recommendations
- Chronic medical condition
- Health insurance
- Age group
- Race
- employment status

Significant Predictors (show significant relationships with vaccine uptake.):

- Gender
- marital status
- education level
- income

Weak Predictor (shows a weaker, but still statistically significant, relationship):

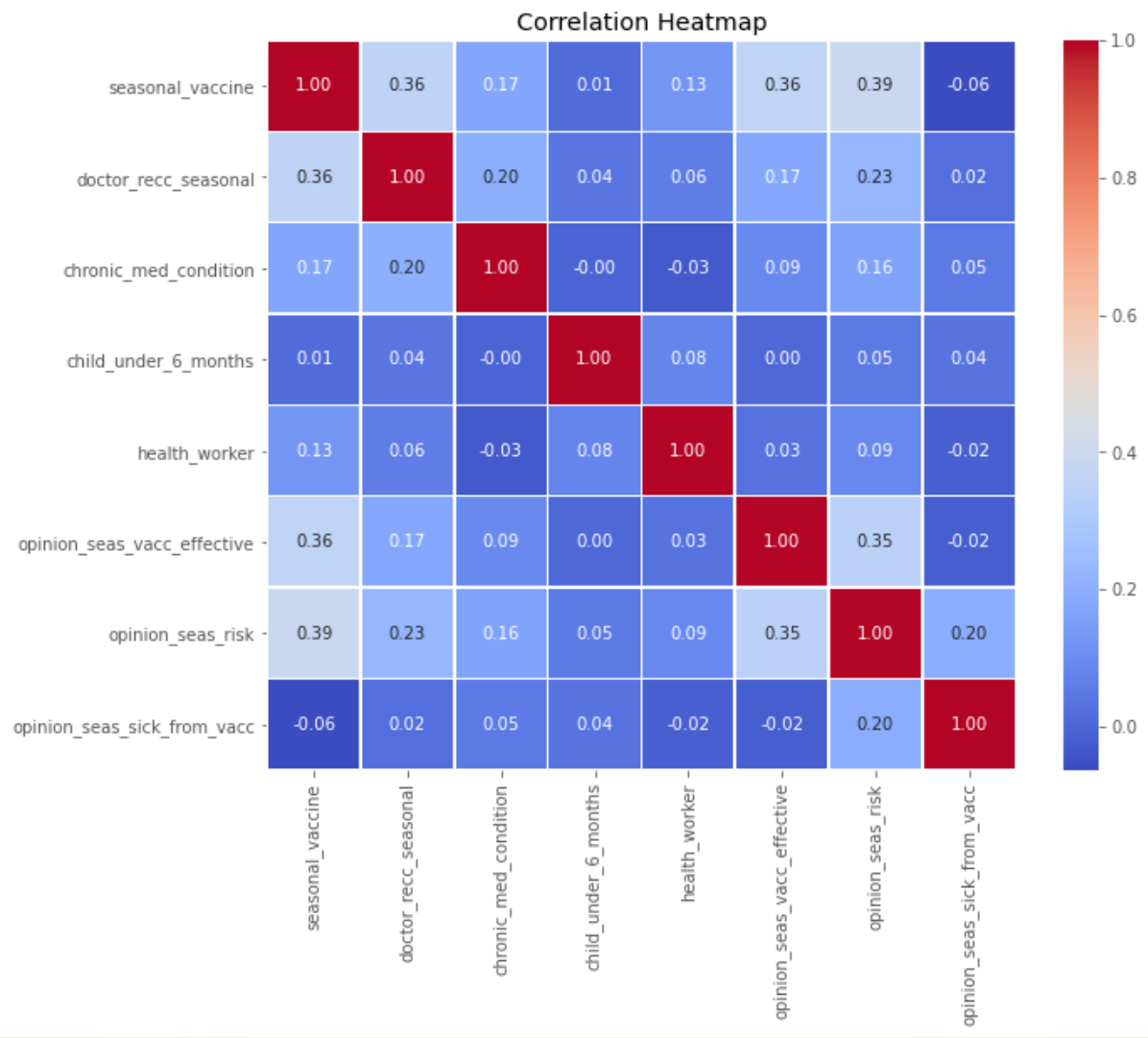
- Having a child under 6 months

Bivariate Analysis – Numerical Variables

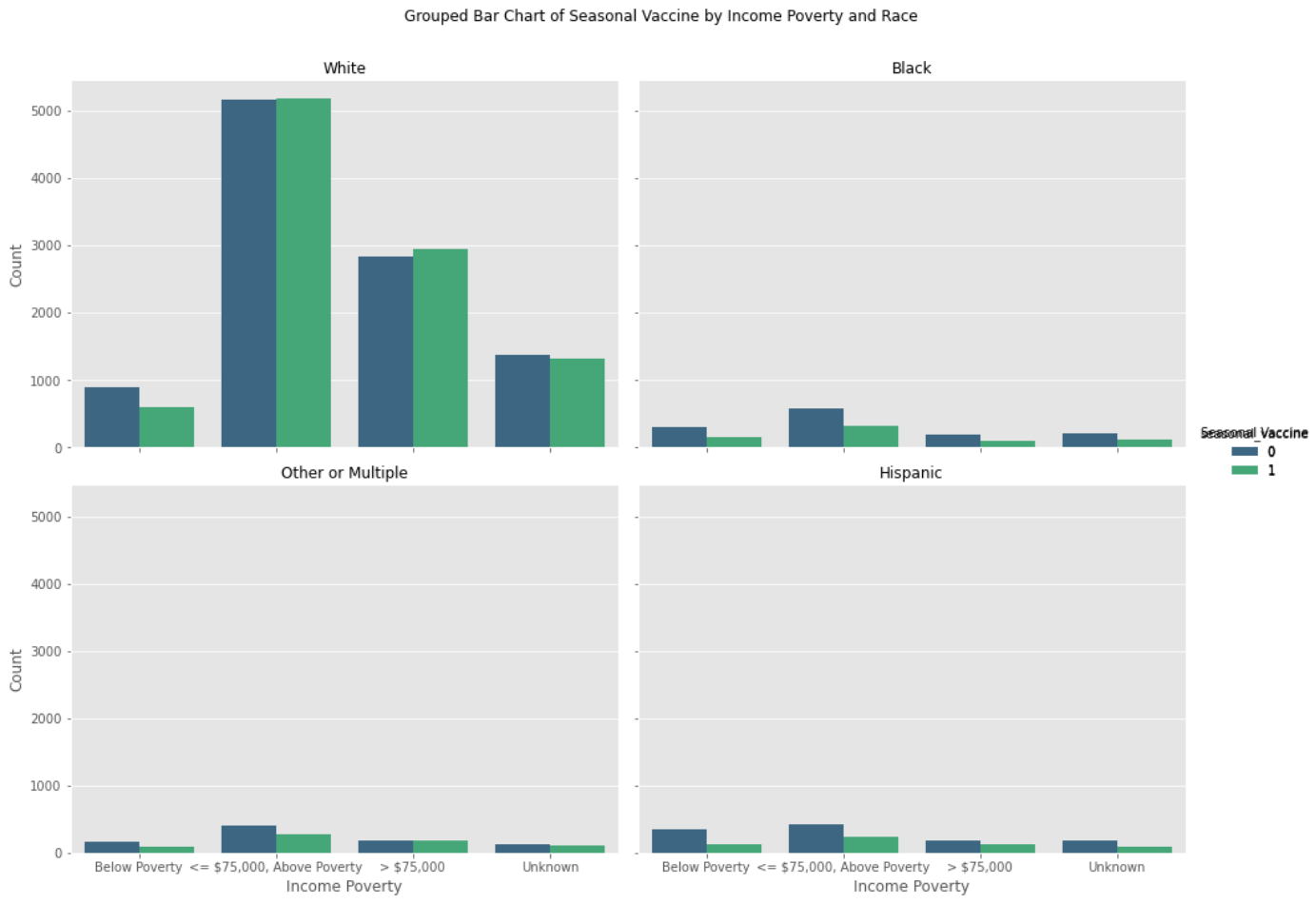
Variable	t-statistic	p-value
Opinion_seas_vacc_effective	62.38	0.0
Opinion_seas_risk	68.35	0.0
Opinion_seas_sick_from_vacc	-10.13	4.60e-24

- **Vaccine Effectiveness:** Those who received the vaccine believe it is significantly more effective than those who did not receive it.
- **Risk Perception:** Those who received the vaccine perceive a significantly higher risk from not getting vaccinated compared to those who did not receive it.
- **Sickness Concern**: Those who received the vaccine are significantly less concerned about getting sick from the vaccine compared to those who did not receive it.

Feature relationships - Correlation

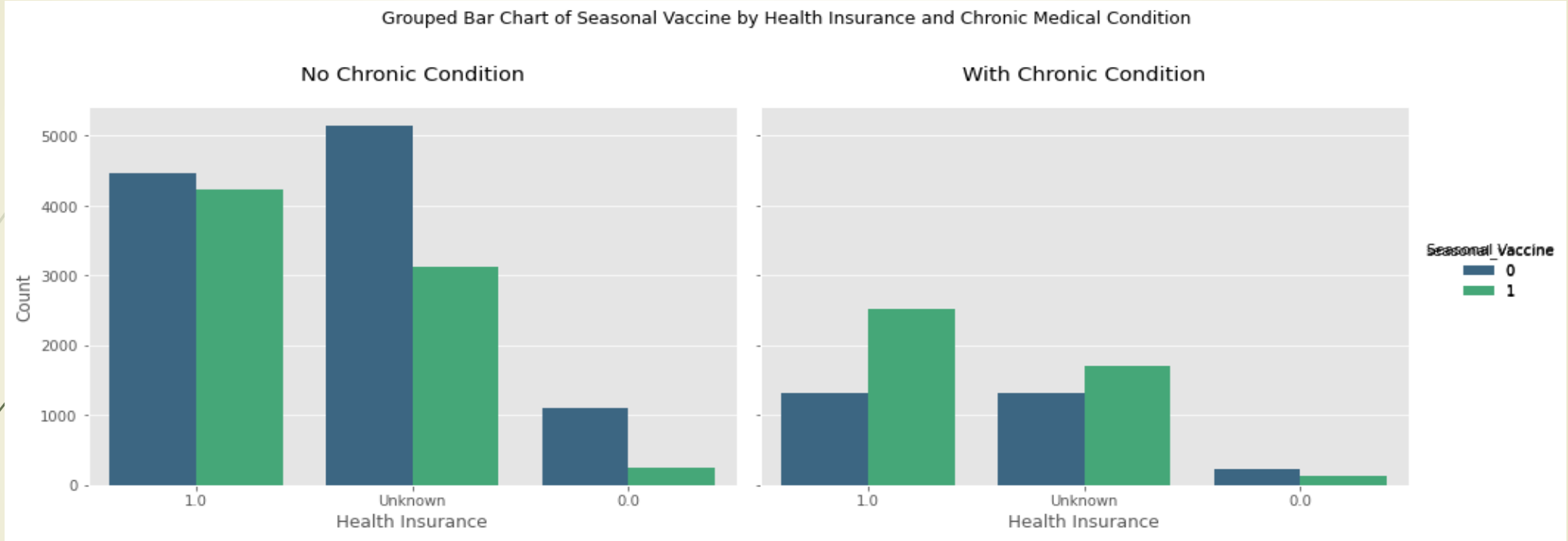


Multivariate Analysis



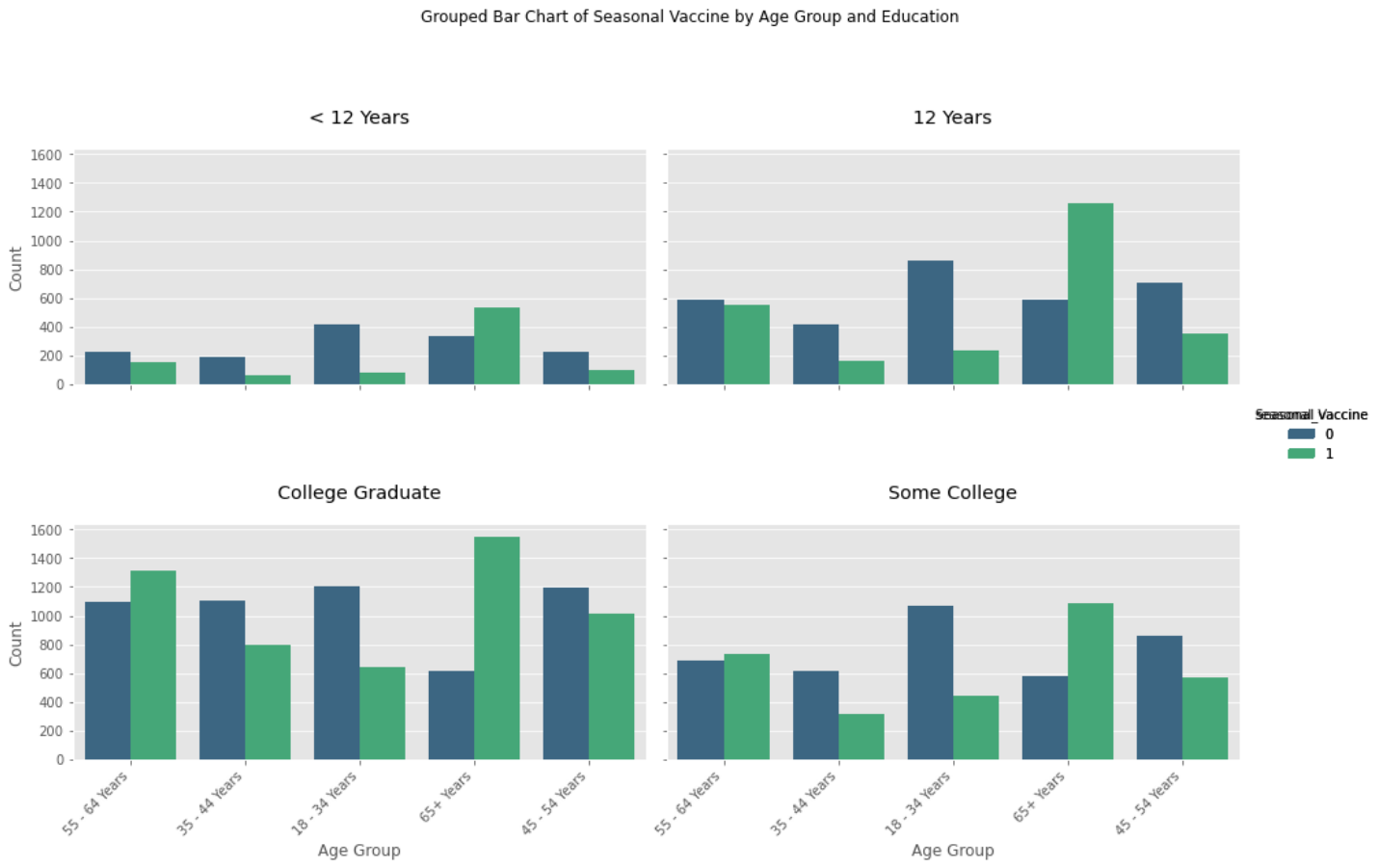
- Income levels and race both influence the likelihood of receiving the seasonal vaccine.
- For most races, individuals below the poverty line and those with unknown income status are less likely to be vaccinated.
- However, for White individuals in higher income categories, vaccination rates are relatively higher.

Multivariate Analysis



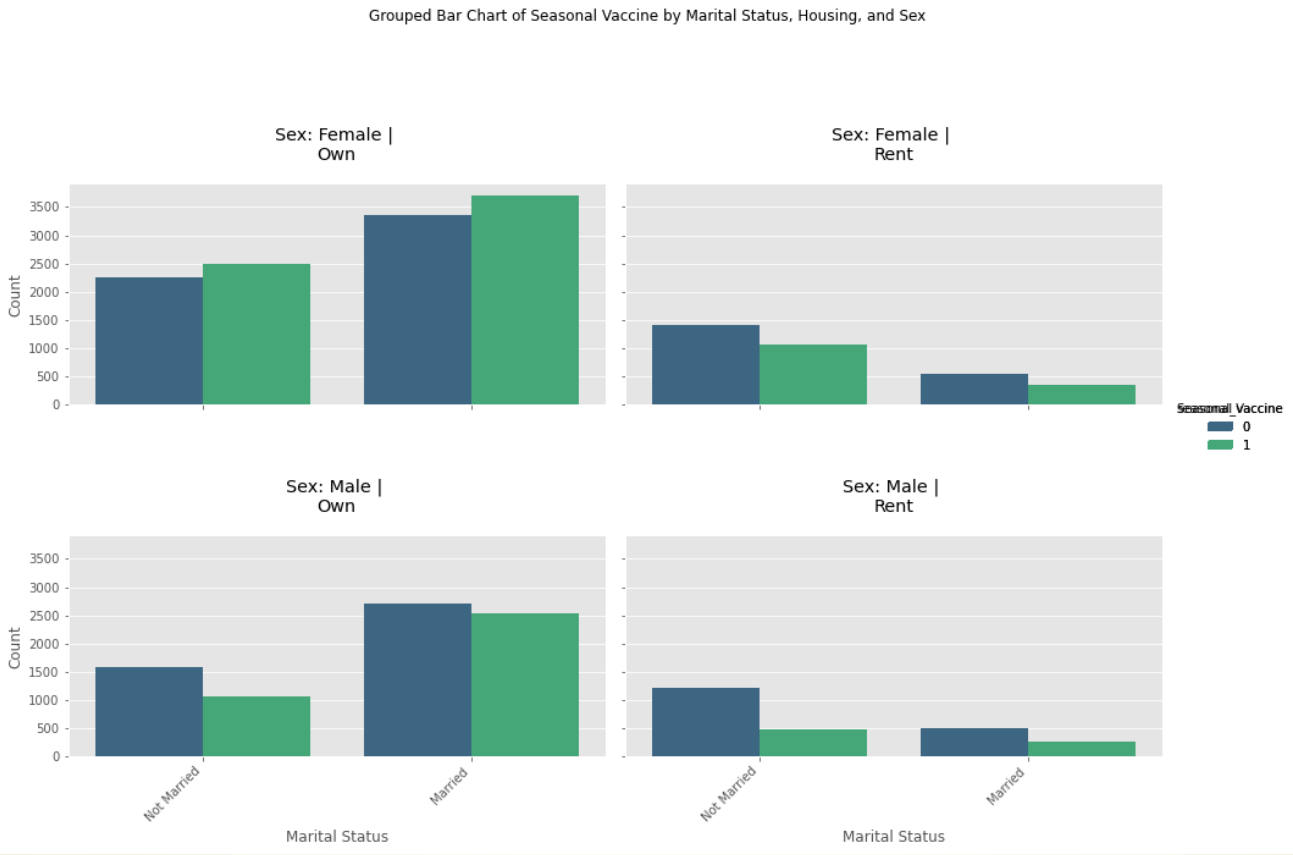
- Health insurance status and the presence of chronic medical conditions influence the likelihood of receiving the seasonal vaccine.
- Having health insurance seems to be a significant factor, particularly for individuals with chronic medical conditions, in increasing the likelihood of vaccination.

Multivariate Analysis



Higher education levels, especially college graduates, tend to show higher vaccination rates across most age groups, indicating the influence of education on vaccine uptake.

Multivariate Analysis

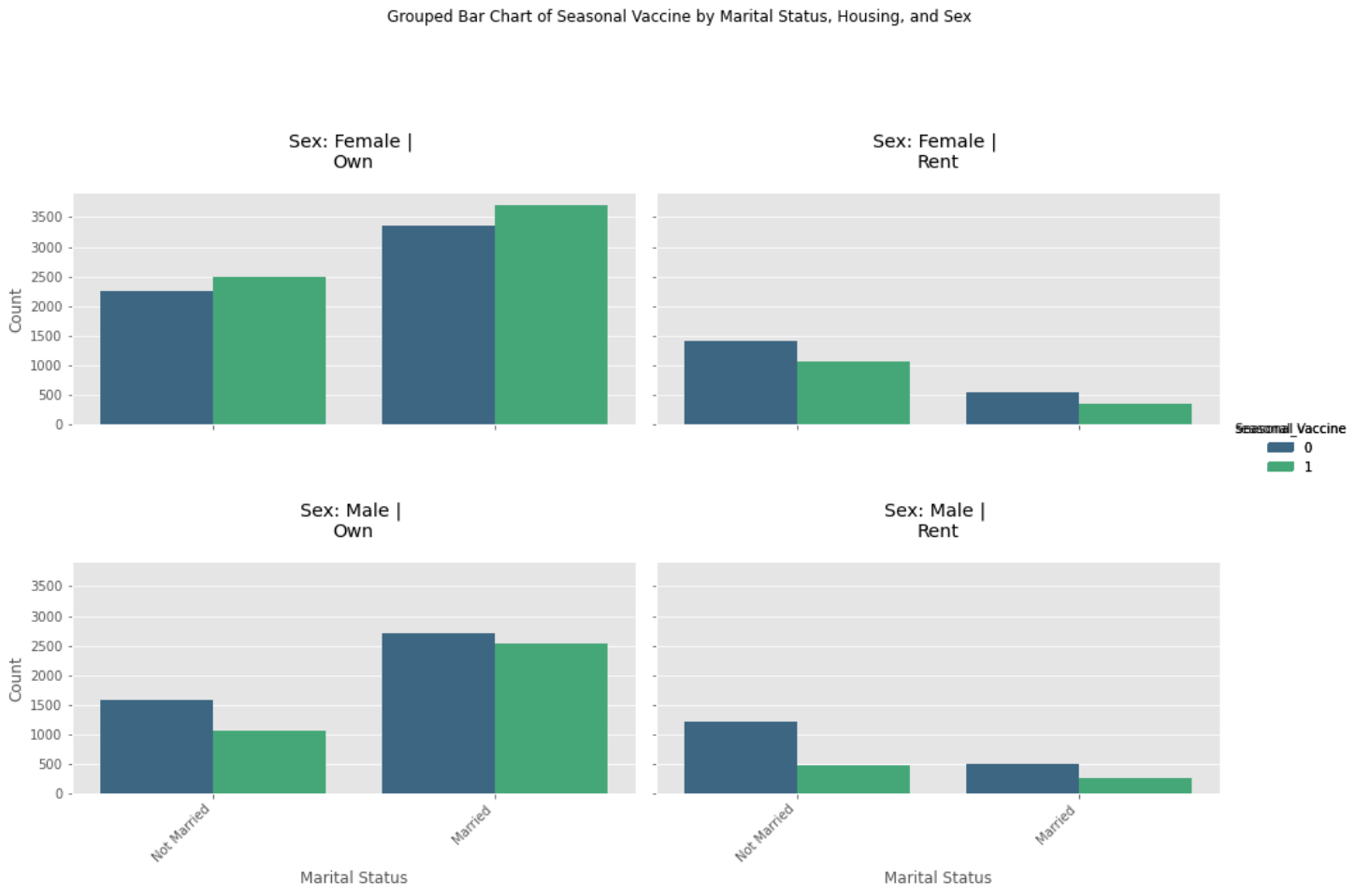


- Homeownership and being married are associated with higher vaccination rates, with females showing a generally higher inclination towards getting vaccinated compared to males.

Multivariate Analysis

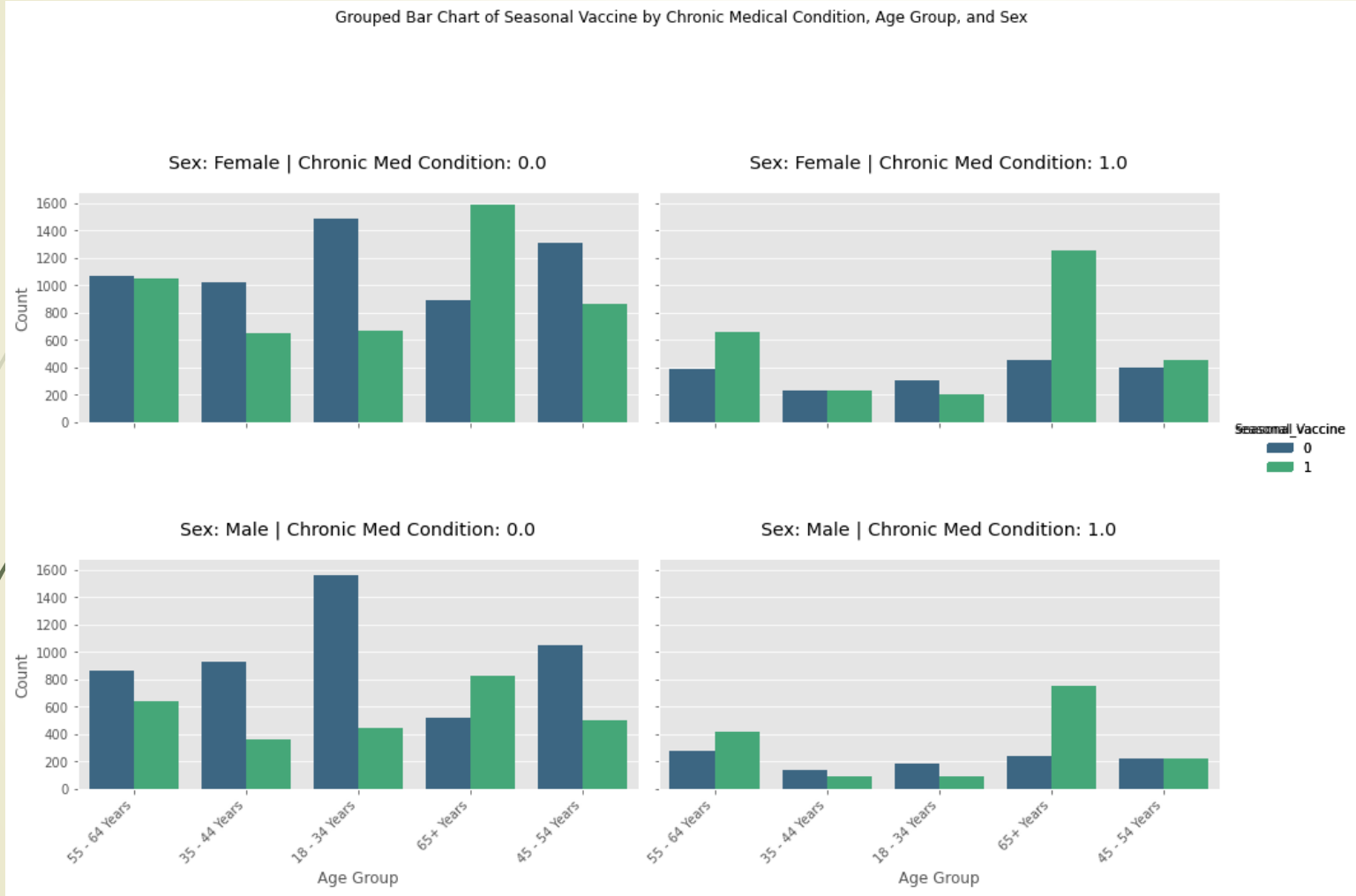
- This shows the importance of healthcare professional recommendations and personal beliefs in vaccine effectiveness in influencing vaccination uptake.
- Health workers, generally being more informed, tend to follow recommendations more consistently, and their opinions strongly align with their vaccination decisions.

Multivariate Analysis

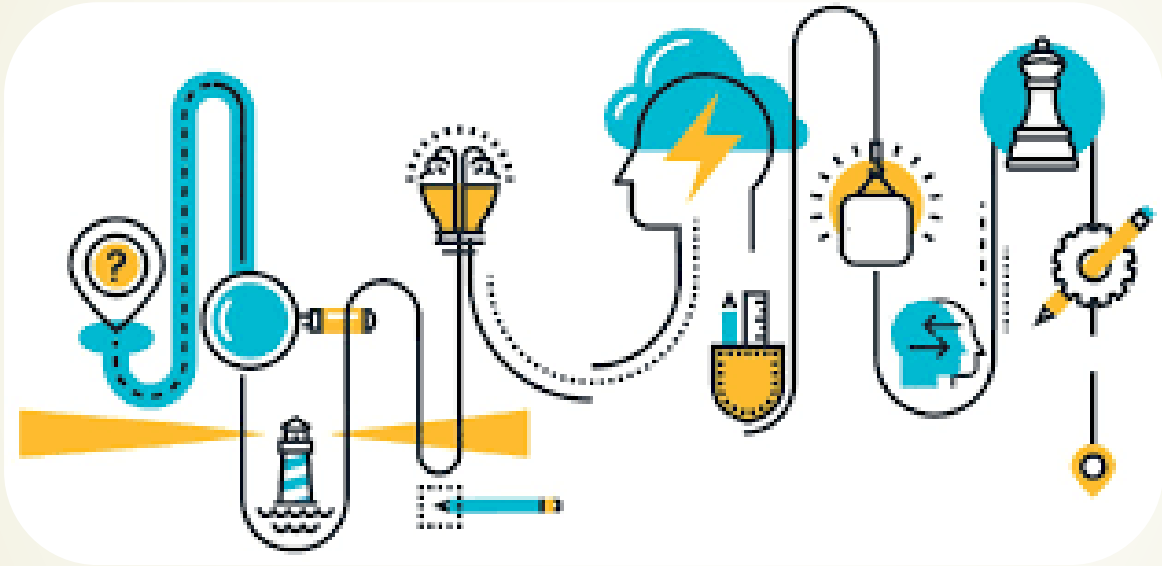


Homeownership and being married are associated with higher vaccination rates, with females showing a generally higher inclination towards getting vaccinated compared to males.

Multivariate Analysis

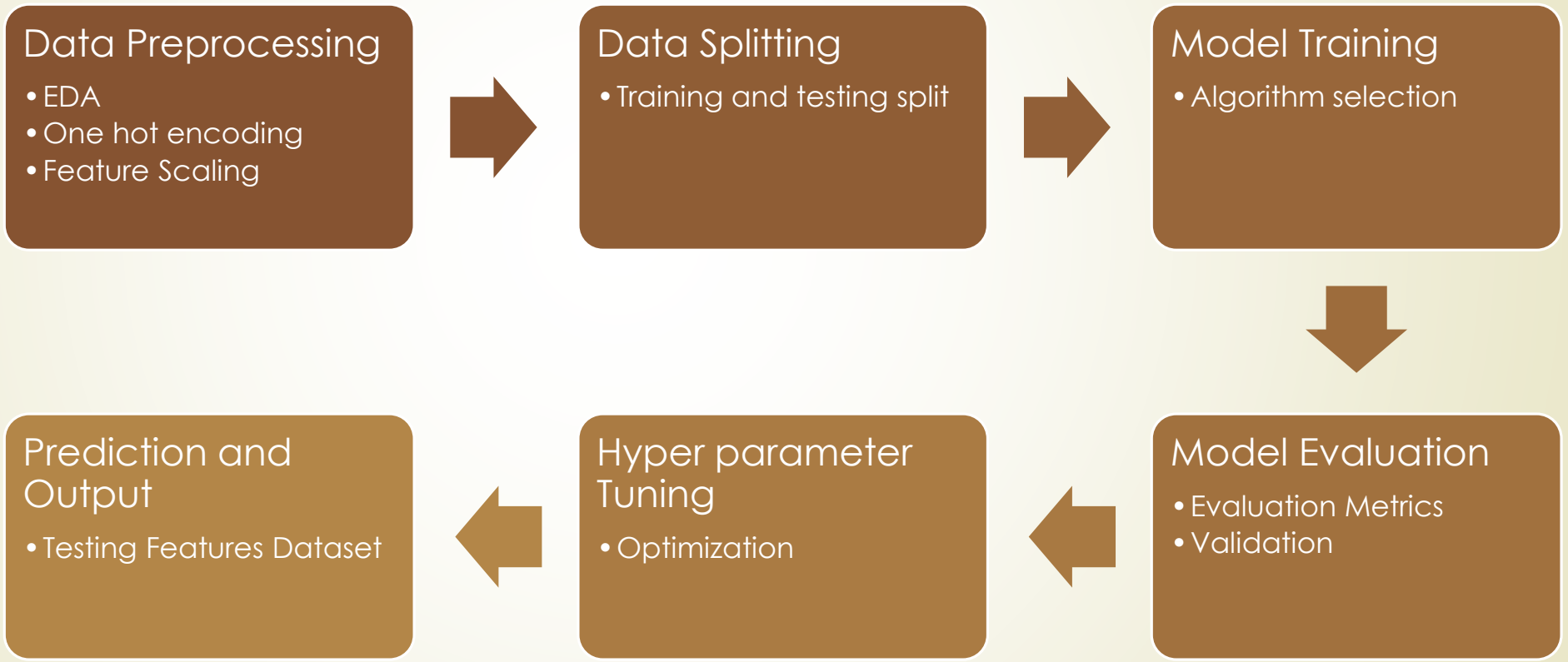


- Vaccination rates tend to be higher among older individuals and those without chronic medical conditions, with females showing a higher inclination towards getting vaccinated compared to males.



Modelling

Process taken



Model Training

- Algorithm Selection
  - Logistic Regression
  - Support Vector Machine
  - Gradient Boosting
  - Naive Bayes
  - Random Forest
  - Decision Tree
  - KNN
- Training Process
  - The model was trained using the merged and cleaned training features and labels dataset.

Model Evaluation

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC Score
Logistic Regression	0.7802	0.7800	0.7802	0.7800	0.8510
Support Vector Machine	0.7796	0.7793	0.7796	0.7793	0.8502
Gradient Boosting	0.7838	0.7838	0.7838	0.7838	0.8539
Naive Bayes	0.7139	0.7349	0.7139	0.7129	0.8164
Random Forest	0.7542	0.7542	0.7542	0.7542	0.8181
Decision Tree	0.6814	0.6814	0.6814	0.6814	0.6879
KNN	0.7393	0.7392	0.7393	0.7392	0.7948

- Overall, Gradient Boosting is the best performing model based on these metrics, followed closely by Logistic Regression, Support Vector Machine, and Random Forest.
- Naive Bayes and KNN have moderate performance, while the Decision Tree model has the lowest performance among the models evaluated.

Model Validation

- Optimization
  - Fine-tuned the Logistic Regression and the Gradient Boosting models hyper parameters to improve performance.
  - Techniques used
    - Grid Search and
    - Random Search
  - Final model - The **tuned Gradient Boosting** model showed slightly improved performance
    - Consistency: The model exhibits consistent performance across training and test datasets
    - Balanced Performance: The F1 scores are close to the precision and recall values, suggesting a balanced performance
    - Good Discriminatory Ability: The high ROC-AUC scores for both training and test sets.

Hyper parameter tuning	**Logistic Regression using grid search**		**Gradient Boosting using random search**	
Metric	**Default Logistic Regression**	**Tuned Logistic Regression (Best Parameters: {'C': 0.1, 'penalty': 'l2', 'solver': 'liblinear'})**	**Default Gradient Boosting**	**Tuned Gradient Boosting (Best Parameters: {'subsample': 0.8, 'n_estimators': 100, 'max_depth': 3, 'learning_rate': 0.1})**
Accuracy	0.7802	0.7800	0.7838	0.7832
Precision (weighted)	0.7800	0.7797	0.7838	0.7833
Recall (weighted)	0.7802	0.7800	0.7838	0.7832
F1 Score (weighted)	0.7800	0.7798	0.7838	0.7832
ROC-AUC Score	0.8510	0.8507	0.8539	0.8545
True Negatives (TN)	2263	2264	2248	2244
False Positives (FP)	540	539	555	559
False Negatives (FN)	583	585	550	549
True Positives (TP)	1724	1722	1757	1758

Prediction and Output

- Testing Features Dataset
  - Used the trained model to predict vaccine uptake in the testing data provided
  - The result output in a CSV file.

Expected Outcome

- Expectation - A machine learning model that accurately predicts seasonal vaccine uptake.
- Outcome - The final model has an accuracy score of 78.93%, precision score of 78.92% F1 score of 78.91% and an ROC –AUC score of 86.56%
- Facilitates targeted public health interventions and resource allocation.

Summary Findings

Summary Findings

Income Levels:

- People living above the poverty line and earning up to USD 75,000 are more likely to get vaccinated.
- Those below the poverty line are the least likely to receive the vaccine.

Vaccine Perception:

- Recipients of the vaccine believe it is significantly more effective and perceive a higher risk from not getting vaccinated.
- Those vaccinated are less concerned about getting sick from the vaccine itself.

Demographic Influences:

- Income and race influence vaccination likelihood. Lower-income and unknown income status correlate with lower vaccination rates across most races.
- White individuals with higher incomes have higher vaccination rates.

Health Insurance and Chronic Conditions:

- Having health insurance, especially for individuals with chronic conditions, increases the likelihood of getting vaccinated.

Summary Findings

Education:

- Higher education levels, particularly among college graduates, are associated with higher vaccination rates.

Homeownership and Marital Status:

- Homeowners and married individuals show higher vaccination rates.
- Females generally have a higher inclination towards vaccination compared to males.

Health Workers:

- Health workers, being more informed, tend to follow vaccination recommendations more consistently, and their opinions strongly align with their vaccination decisions.

Age and Health:

- Older individuals and those without chronic medical conditions have higher vaccination rates.

Recommendations

- Public Health Officials
- Healthcare providers
- Policymakers
- Community Organizations
- Researchers
- Cross cutting strategies

Recommendations- Public Health Officials

- Targeted Campaigns:
  - Develop campaigns specifically aimed at populations below the poverty line, emphasizing the importance and safety of the flu vaccine.
  - Use data to identify regions with low vaccination rates and deploy targeted interventions.
- Education and Awareness:
  - Increase public awareness about the effectiveness and benefits of the flu vaccine.
  - Highlight the risks of not getting vaccinated and address common concerns about vaccine safety.
- Partnerships:
  - Partner with local community organizations and faith-based groups to reach underserved populations.
  - Collaborate with schools and universities to promote vaccination among students and staff.

Recommendations- Healthcare Providers

- Patient Education:
  - Educate patients, especially those with chronic conditions, about the benefits of flu vaccination during regular check-ups.
  - Provide clear and accurate information to address concerns about vaccine safety and effectiveness.
- Accessibility:
  - Make the vaccine easily accessible by offering it in clinics, pharmacies, and through mobile vaccination units.
  - Extend clinic hours to accommodate working individuals and families.
- Recommendation and Reminders:
  - Actively recommend the flu vaccine to all patients, especially those at higher risk.
  - Use reminder systems (calls, texts, emails) to remind patients about flu vaccination appointments.

Recommendations Policymakers

- Funding and Resources:
  - Increase funding for vaccination programs targeting low-income and underserved communities.
  - Provide financial incentives or subsidies for vaccines to make them affordable for all income groups.
- Legislation:
  - Enact policies that require health insurance plans to cover flu vaccination costs without copayments.
  - Support policies that facilitate workplace vaccination programs, especially in high-risk industries.
- Data Utilization:
  - Use data analytics to monitor vaccination rates and identify gaps in coverage.
  - Implement policies based on data-driven insights to address disparities in vaccine uptake.

Recommendations - Community Organizations

- Outreach Programs:
  - Organize community vaccination drives in collaboration with health departments and local healthcare providers.
  - Use community leaders to spread awareness and encourage vaccination among residents.
- Education Initiatives:
  - Host workshops and informational sessions about the flu vaccine's benefits and safety.
  - Distribute educational materials in multiple languages to reach diverse populations.
- Support Services:
  - Provide assistance with transportation to vaccination sites for those in need.
  - Offer support to navigate health insurance and access vaccination services.

Recommendations - Researchers

- Ongoing Studies:
  - Conduct research to understand the barriers to vaccination in different demographics.
  - Investigate the effectiveness of various interventions aimed at increasing vaccination rates.
- Publications and Dissemination:
  - Publish findings in accessible formats and disseminate them to policymakers, healthcare providers, and public health officials.
  - Collaborate with media to share research insights and promote vaccine awareness.
- Community Engagement:
  - Engage with communities to understand their concerns and perceptions about vaccines.
  - Develop community-based participatory research projects to co-create solutions for increasing vaccine uptake.

Recommendations- Cross-cutting Strategies

- Inclusive Messaging:
  - Ensure that all communication and educational materials are culturally sensitive and inclusive.
  - Address specific concerns and misconceptions prevalent in various communities.
- Leveraging Technology:
  - Use social media and digital platforms to reach a wider audience with vaccination messages.
  - Implement mobile apps and online portals for easy access to vaccination information and appointment scheduling.
- Role Models and Influencers:
  - Use trusted community leaders and influencers to advocate for flu vaccination.
  - Highlight stories of individuals who have benefited from getting vaccinated.

Conclusion

- By adopting the recommendations , stakeholders can work together to improve vaccination rates, thereby reducing the incidence and impact of seasonal flu. This model serves as a crucial tool in informing and guiding these efforts, ultimately contributing to better public health outcomes.

Thank you

Q&A