



Prediction of Seasonal vaccine uptake

BY Lorna Gatimu



Table of Contents

- 1 Business and Data Understanding
- 2 Modelling
- 3 Evaluations
- 4 Conclusions and Recommendations

Business and Data Understanding

Business Understanding

Goal: Predict whether individuals will take the seasonal flu vaccine based on demographics, socioeconomic factors, and behaviors.

Business Objective: Improve vaccination rates by identifying individuals likely to get vaccinated.

Key Benefit:

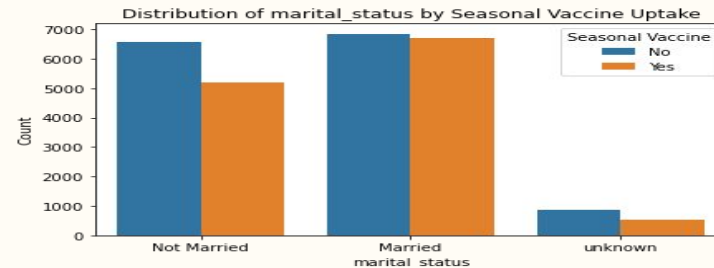
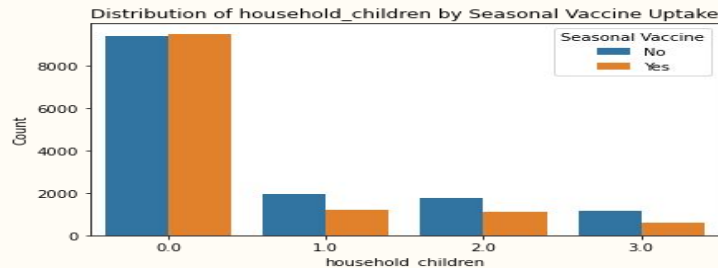
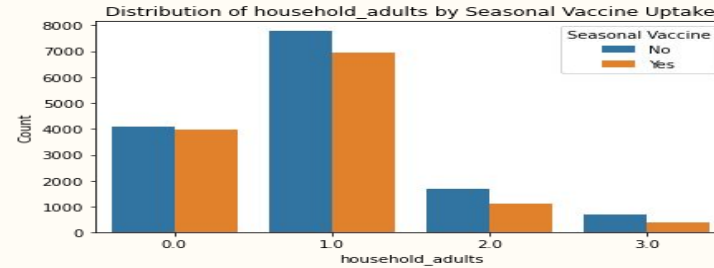
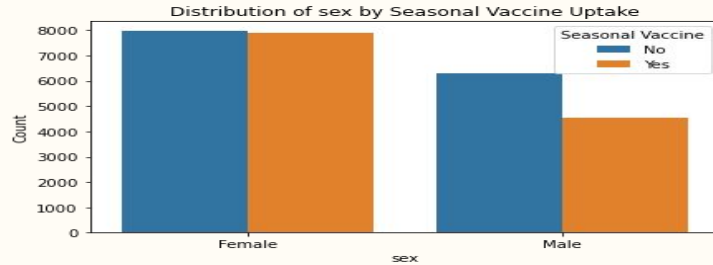
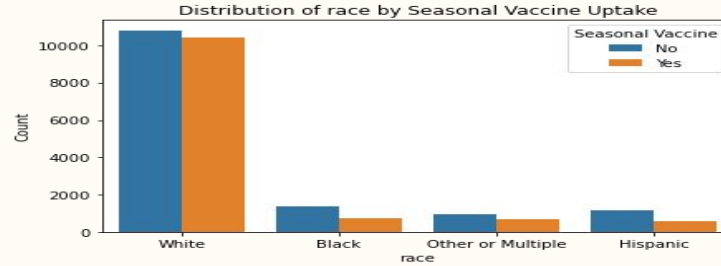
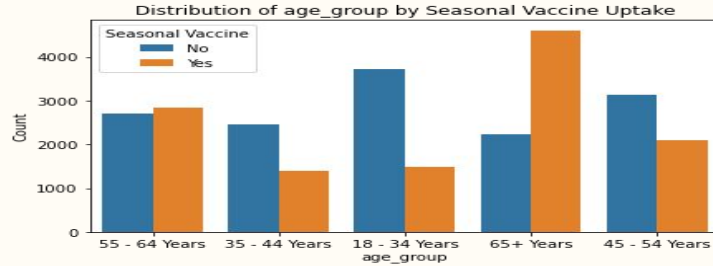
- Enable targeted interventions to encourage vaccination.
- Improve public health outcomes.
- Reduce healthcare costs associated with flu-related illnesses.



Business and Data Understanding

4

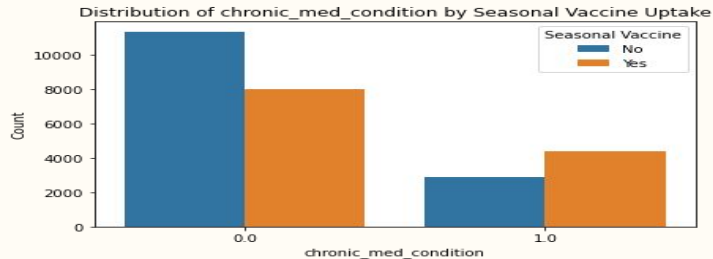
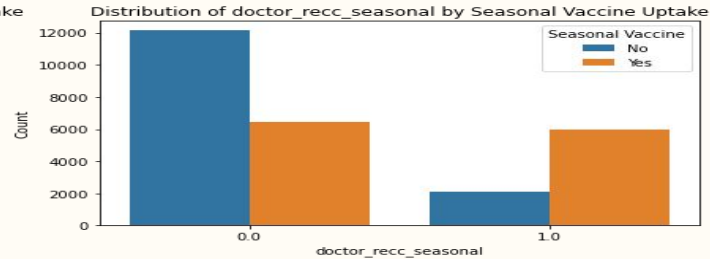
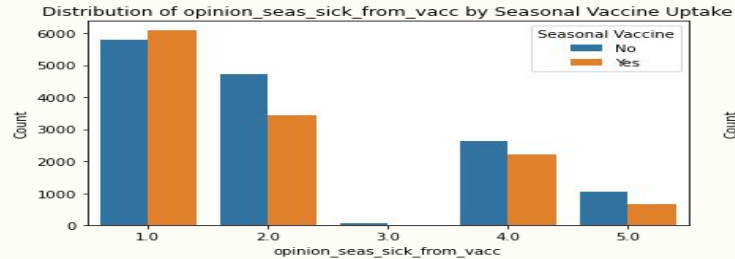
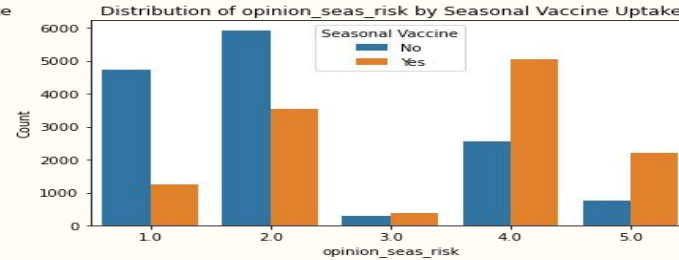
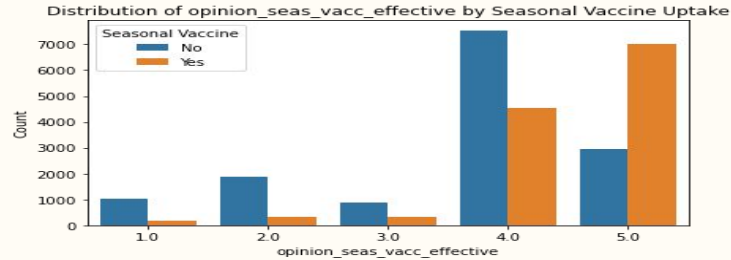
Data Understanding



Business and Data Understanding

5

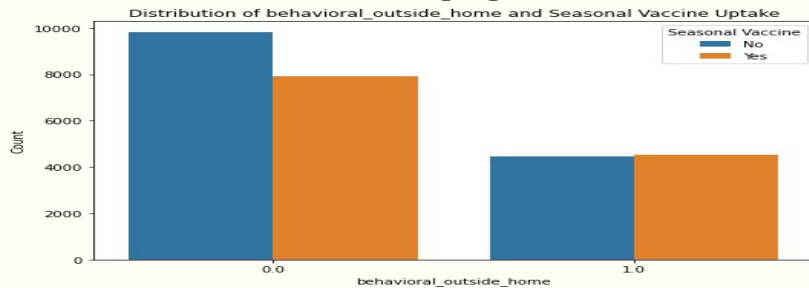
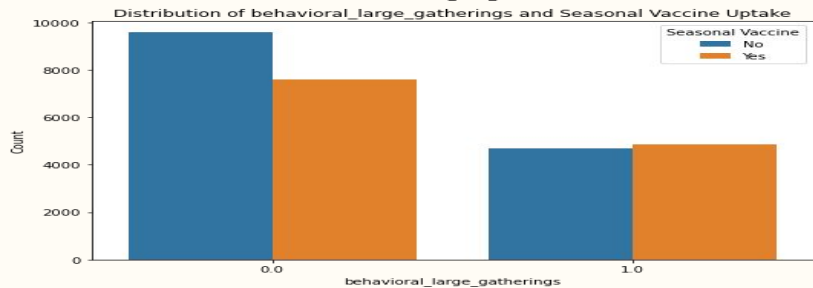
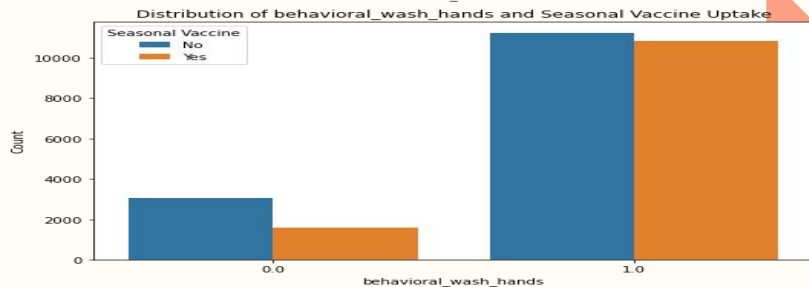
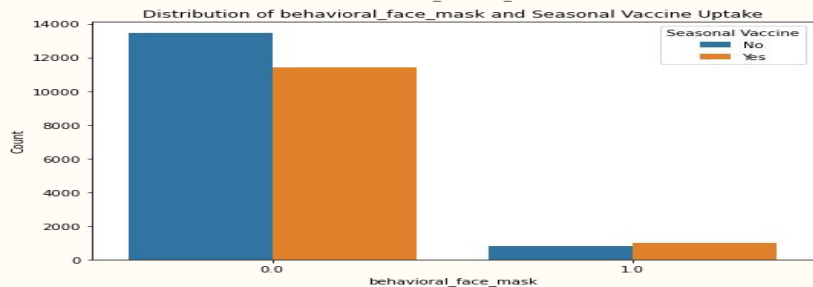
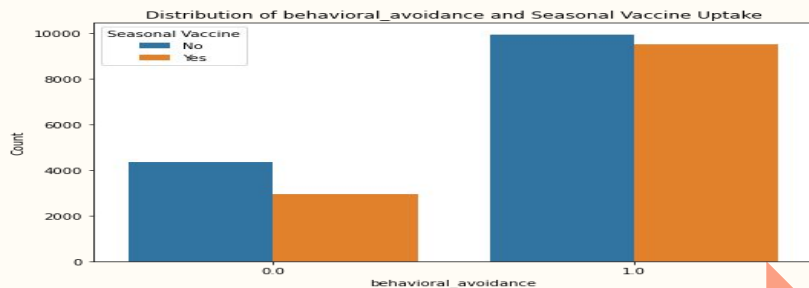
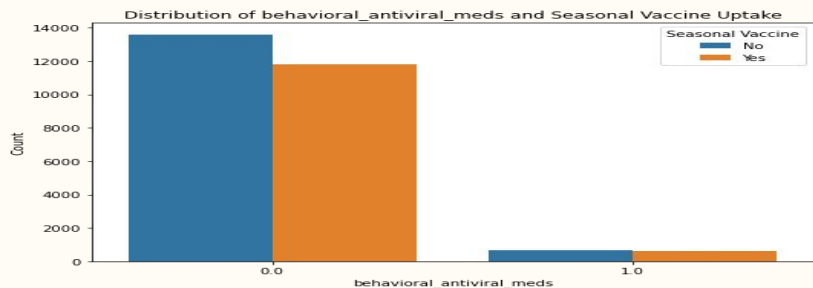
Data Understanding



Business and Data Understanding

6

Data Understanding



Modelling

Baseline Model: The Dummy Classifier (baseline) has limited predictive value, with a high recall (100%) but low precision (45.81%), resulting in an F1-Score of 62.84%. It predicts all instances as vaccinated, which is not practical.

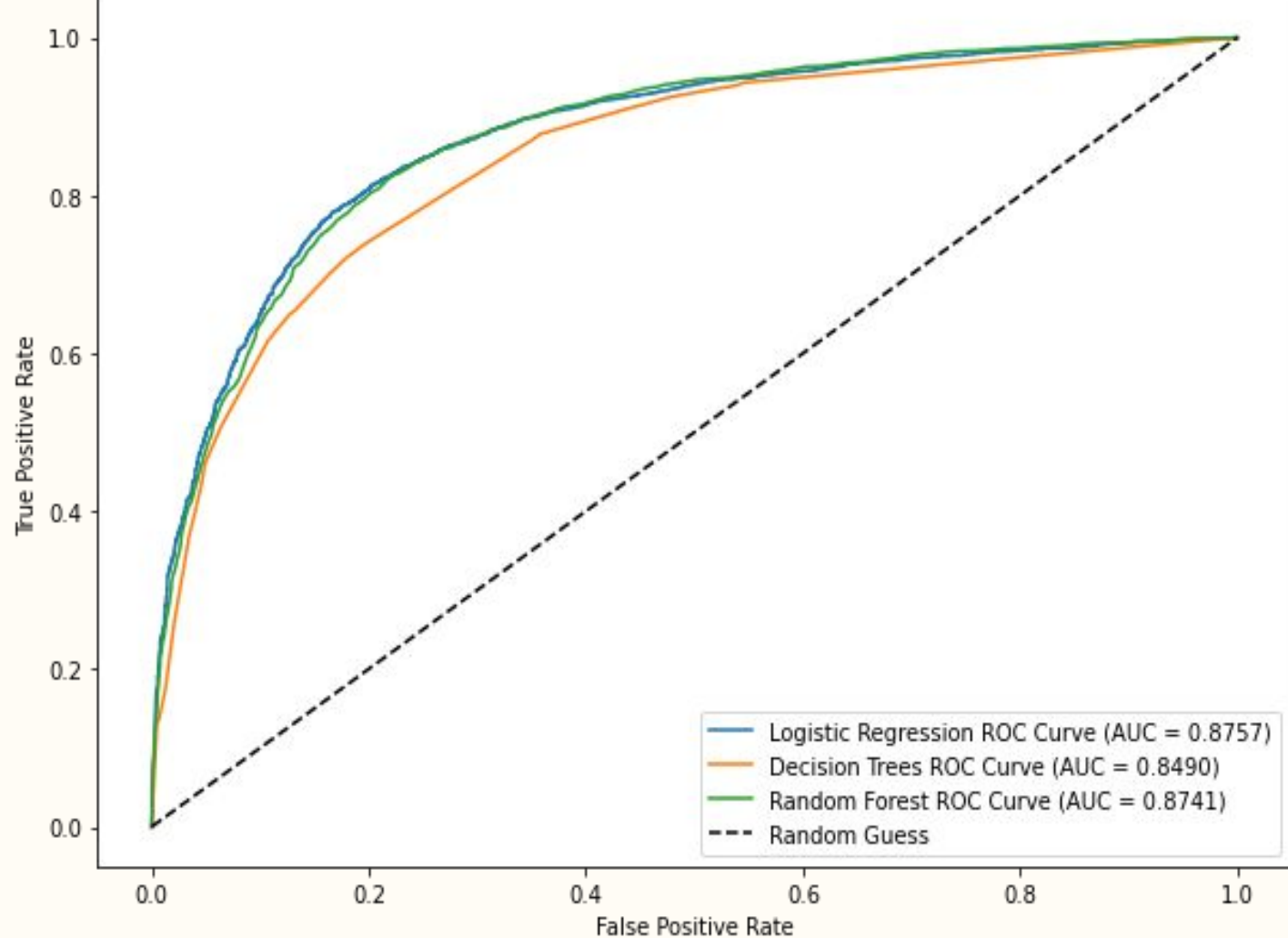
Logistic Regression: This model is the strongest, with an accuracy of 80.7%, precision of 80.6%, and recall of 76.0%. Its F1-Score of 78.3% and ROC AUC of 0.876 indicate strong performance in distinguishing between vaccinated and non-vaccinated individuals. It's preferred for its simplicity and interpretability.

Random Forest: With an accuracy of 80.1%, precision of 79.2%, and recall of 76.8%, Random Forest performs similarly to Logistic Regression. Its F1-Score is 78.0% and ROC AUC is 0.874. It is useful for more complex, non-linear relationships.

Decision Trees: Decision Trees performed slightly worse than Logistic Regression and Random Forest, with an accuracy of 77.5% and recall of 72.0%. The ROC AUC score of 0.849 suggests lower class discrimination but still a functional model.

ROC Curve and Class Discrimination: All models (except baseline) show strong performance in distinguishing between classes, with ROC AUC scores above 0.8.

Real-World Application: Logistic Regression is ideal for simple, interpretable models, while Random Forest can handle more complexity. Decision Trees may be less favorable unless simplicity is prioritized.



CONCLUSIONS AND RECOMMENDATIONS

Conclusion:

- **Baseline Model:** Simple reference model, predicting everyone will get vaccinated.
- **Logistic Regression:** Best overall performance, strong ability to predict vaccine uptake.
- **Decision Tree:** Less accurate but still viable for predicting vaccine uptake.
- **Random Forest:** Promising results, good prediction accuracy

Recommendations:

- **Improving Vaccination Rates:** Target at-risk individuals with personalized communication and incentives.
- **Tailored Campaigns:** Use Logistic Regression and Random Forest for targeted educational campaigns and reminders.
- **Resource Allocation:** Efficiently allocate resources by focusing on segments predicted to need intervention.
- **Continuous Improvement:** Regularly monitor and retrain models to maintain prediction accuracy as trends change.



THANK YOU!!!