WORLD HEALTH ORGANIZATION

FLU SHOT LEARNING: PREDICT H1N1 AND SEASONAL FLU VACCINES

Phase 3 project

WORLD HEALTH ORGANIZATION | VACCINES

Outline

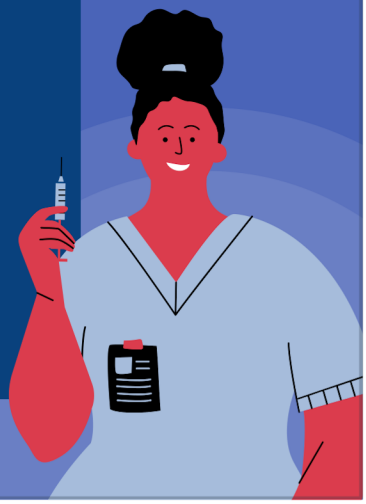
01 Business Problem

02 Research Questions:

03 DATA ANALYSIS

04 MODELLING

05 RECOMMENDATIONS



WORLD HEALTH ORGANIZATION | VACCINES

Business problem

WORLD HEALTH ORGANIZATION | VACCINES

STAKEHOLDER: World Health Organization (WHO)

- Given the importance of vaccinations, particularly in light of global pandemics like COVID-19 and the H1N1 flu, understanding the factors that influence an individual's decision to get vaccinated can be crucial for public health planning
- By analyzing data from previous pandemics, we can identify patterns and trends that have emerged over time
- By demonstrating an understanding of the concerns and behaviors of different groups, public health officials can build trust and engage more effectively with communities.
- Insights from data classification can guide where resources (like awareness campaigns, vaccination centers, or community health workers) might be most effectively deployed.

Research Questions:

As the world struggles to vaccinate the global population against COVID-19, an understanding of how people's backgrounds, opinions, and health behaviors are related to their personal vaccination patterns can provide guidance for future public health efforts.

can you predict whether people got H1N1 and seasonal flu vaccines using data collected in the National 2009 H1N1 Flu Survey?

DATA SOURCE: DrivenData. (2020). Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines. Retrieved [10/17/2023] from https://www.drivendata.org/competitions/66/flu-shot-learning.

WORLD HEALTH ORGANIZATION | VACCINES

DATA EXPLORATION

We have two data sets that we have merged training_set_features

contains information about the respondents, such as their level of concern about the H1N1 virus, knowledge about H1N1, behavioral habits, and demographic details.

training_set_labels

provides the target variables for each respondent, indicating whether they received the H1N1 vaccine (h1n1_vaccine) and the seasonal vaccine (seasonal_vaccine).

WORLD HEALTH ORGANIZATION | VACCINES

MISSING VALUES AND DUPLICATES

From our merged data set we have no duplicates

Several columns have missing values. The columns employment_occupation, employment_industry, and health_insurance have notably high percentages of missing data, with 50.44%, 49.91%, and 45.96% missing respectively.

- **One-hot encoding used for categorical data. Concatenated the original dataframe with the encoded dataframe**
- **Imputed missing values For numerical columns, we'll use median;For categorical (encoded) columns, we'll use mode**
- **Scaling the numerical features using standardd scaler**

WORLD HEALTH ORGANIZATION | VACCINES

CORRELATION

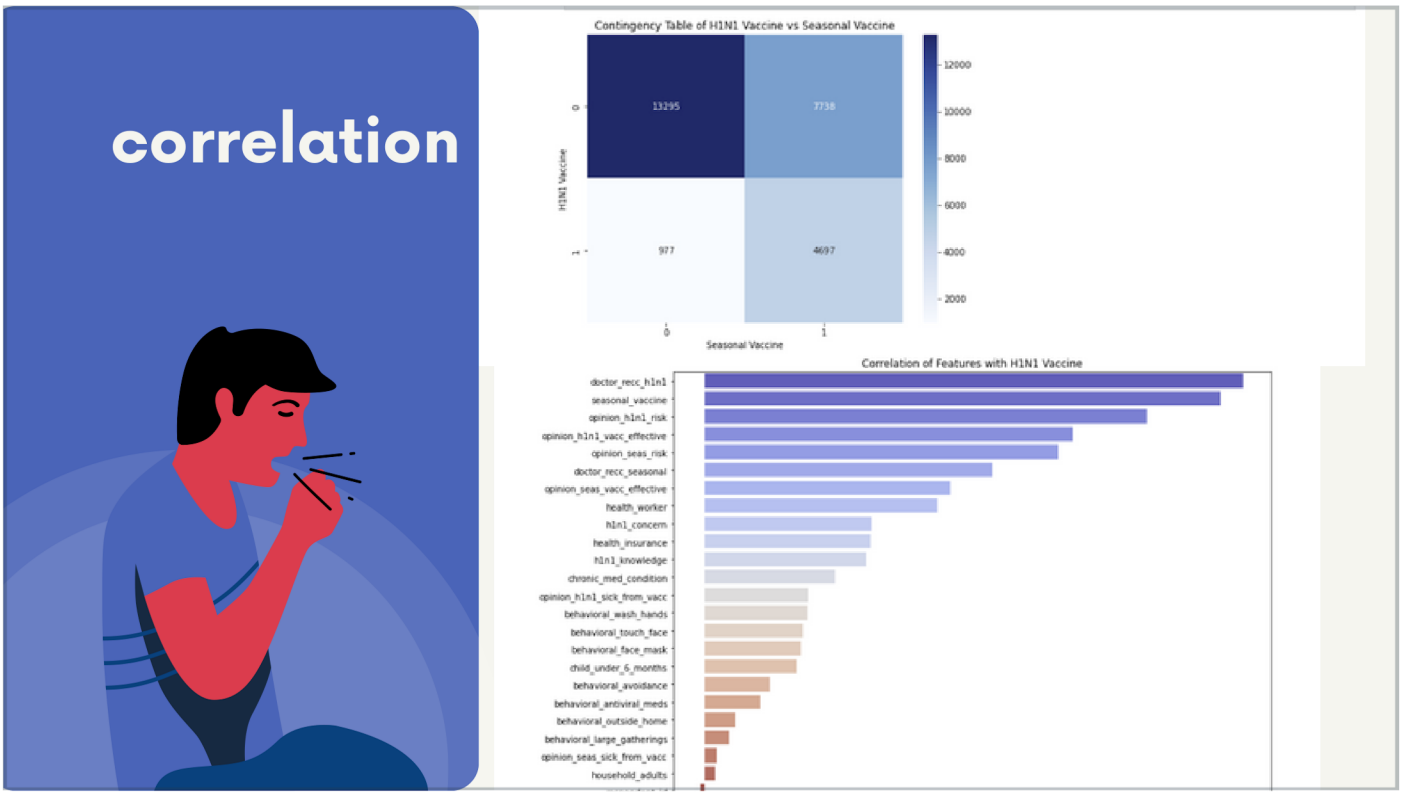
H1N1 Vaccine Correlations:

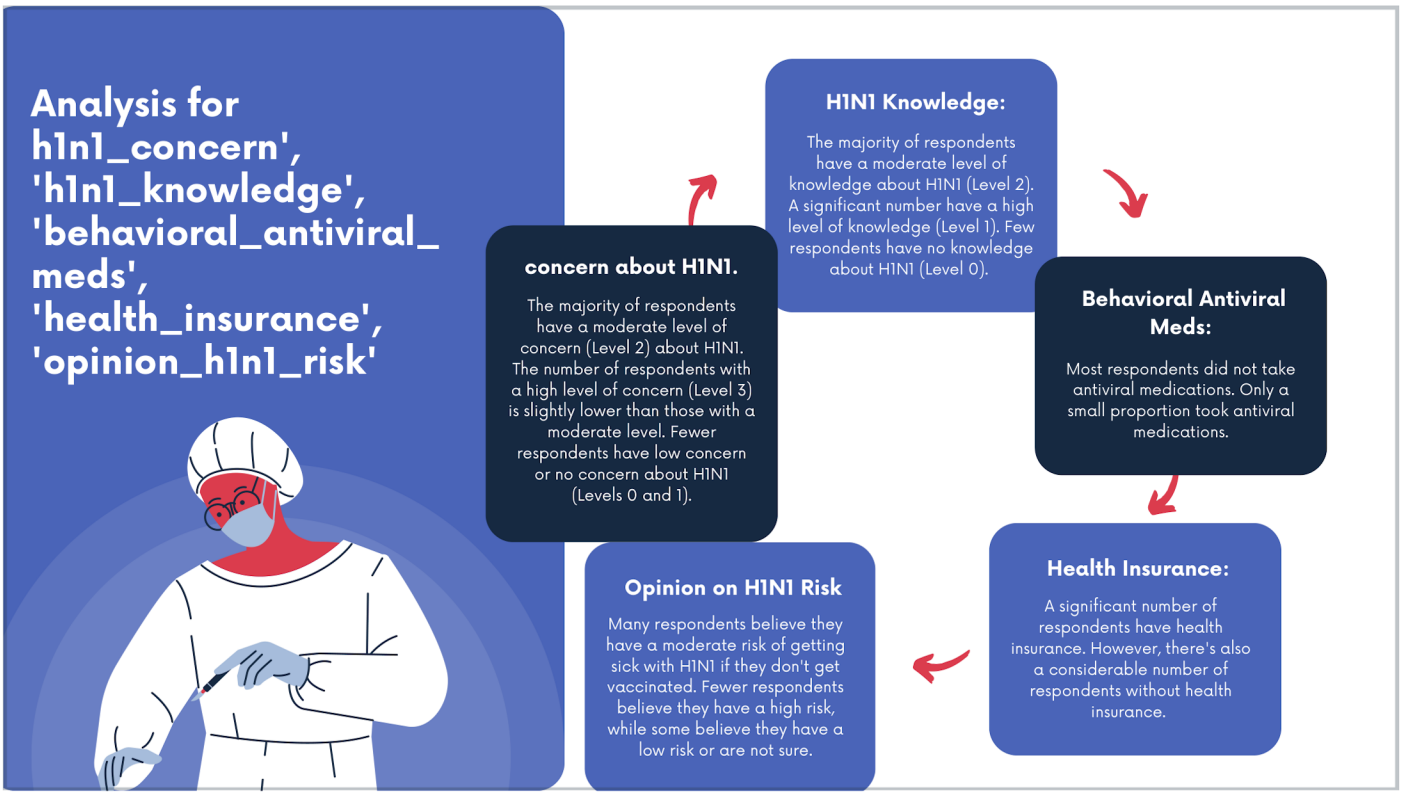
Doctor recommendations (doctor_recc_h1n1) have the highest positive correlation with getting the H1N1 vaccine. This suggests that individuals are more likely to get vaccinated if recommended by a healthcare professional. Respondents' opinions on the risks and effectiveness of the H1N1 vaccine (opinion_h1n1_risk, opinion_h1n1_vacc_effective, and opinion_h1n1_sick_from_vacc) also show significant correlations.

Seasonal Vaccine Correlations:

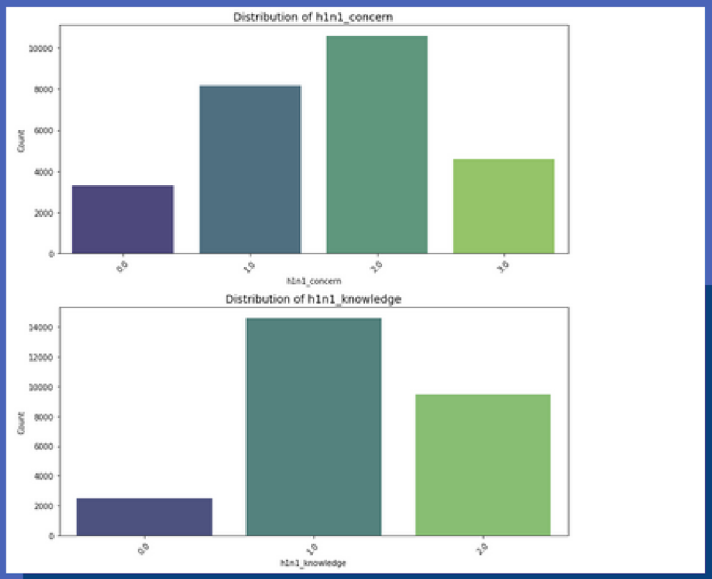
The age group of the respondent (age_group) has a strong positive correlation with receiving the seasonal vaccine. Doctor recommendations for the seasonal flu vaccine (doctor_recc_seasonal) and opinions about its risk and effectiveness are also significantly correlated. Interestingly, the correlation of h1n1_vaccine with the seasonal vaccine is also evident, reinforcing our earlier observation that the two are not independent.

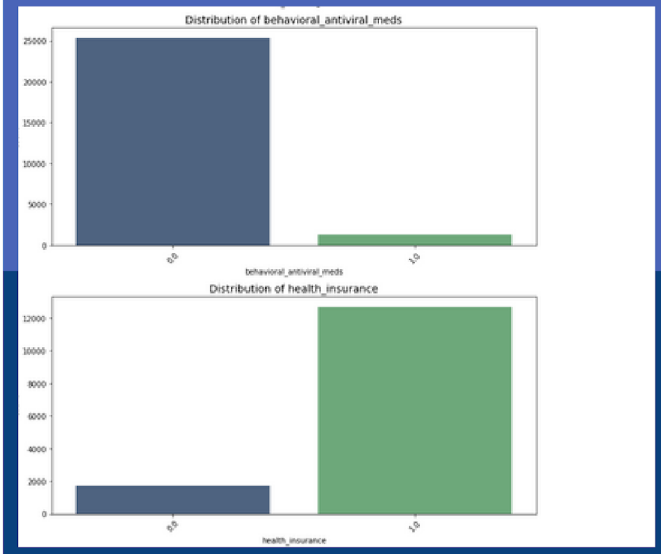
WORLD HEALTH ORGANIZATION | VACCINES


correlation
Contingency Table of H1N1 Vaccine vs Seasonal Vaccine
13295
7738
977
4697
H1N1 Vaccine
Seasonal Vaccine
Correlation of Features with H1N1 Vaccine
doctor_recc_h1n1
seasonal_vaccine
opinion_h1n1_risk
opinion_h1n1_vacc_effective
opinion_seas_risk
doctor_recc_seasonal
opinion_seas_vacc_effective
health_worker
h1n1_concern
health_insurance
h1n1_knowledge
chronic_med_condition
opinion_h1n1_sick_from_vacc
behavioral_wash_hands
behavioral_touch_face
behavioral_face_mask
child_under_6_months
behavioral_avoidance
behavioral_antiviral_meds
behavioral_outside_home
behavioral_large_gatherings
opinion_seas_sick_from_vacc
household_adults

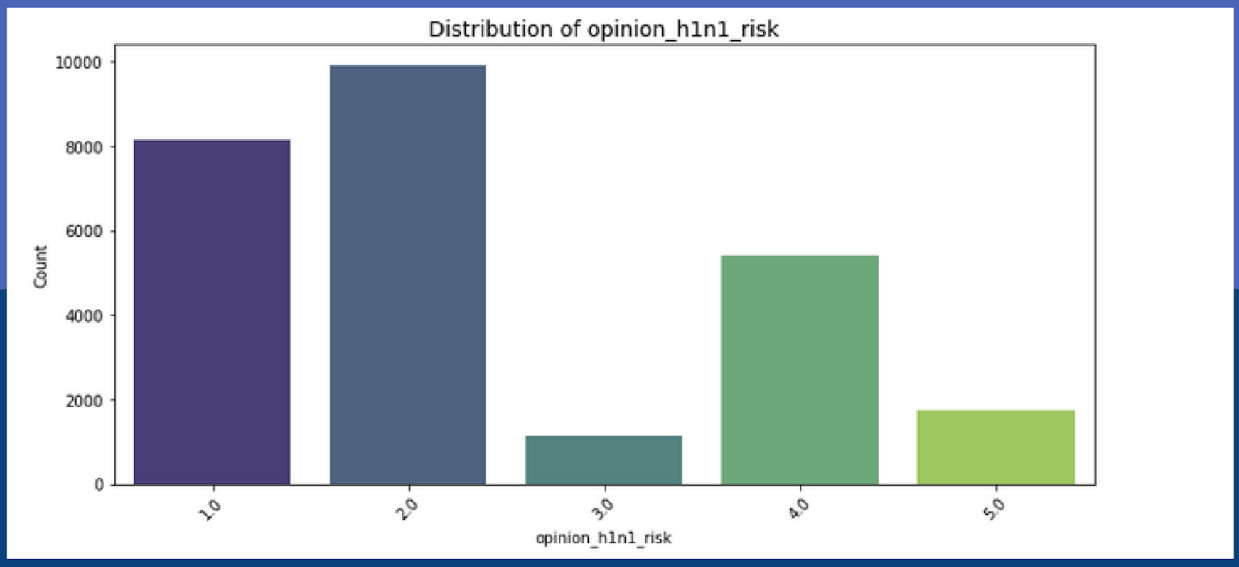

Analysis for h1n1_concern', 'h1n1_knowledge', 'behavioral_antiviral_meds', 'health_insurance', 'opinion_h1n1_risk'
concern about H1N1.
The majority of respondents have a moderate level of concern (Level 2) about H1N1. The number of respondents with a high level of concern (Level 3) is slightly lower than those with a moderate level. Fewer respondents have low concern or no concern about H1N1 (Levels 0 and 1).
H1N1 Knowledge:
The majority of respondents have a moderate level of knowledge about H1N1 (Level 2). A significant number have a high level of knowledge (Level 1). Few respondents have no knowledge about H1N1 (Level 0).
Behavioral Antiviral Meds:
Most respondents did not take antiviral medications. Only a small proportion took antiviral medications.
Health Insurance:
A significant number of respondents have health insurance. However, there's also a considerable number of respondents without health insurance.
Opinion on H1N1 Risk
Many respondents believe they have a moderate risk of getting sick with H1N1 if they don't get vaccinated. Fewer respondents believe they have a high risk, while some believe they have a low risk or are not sure.

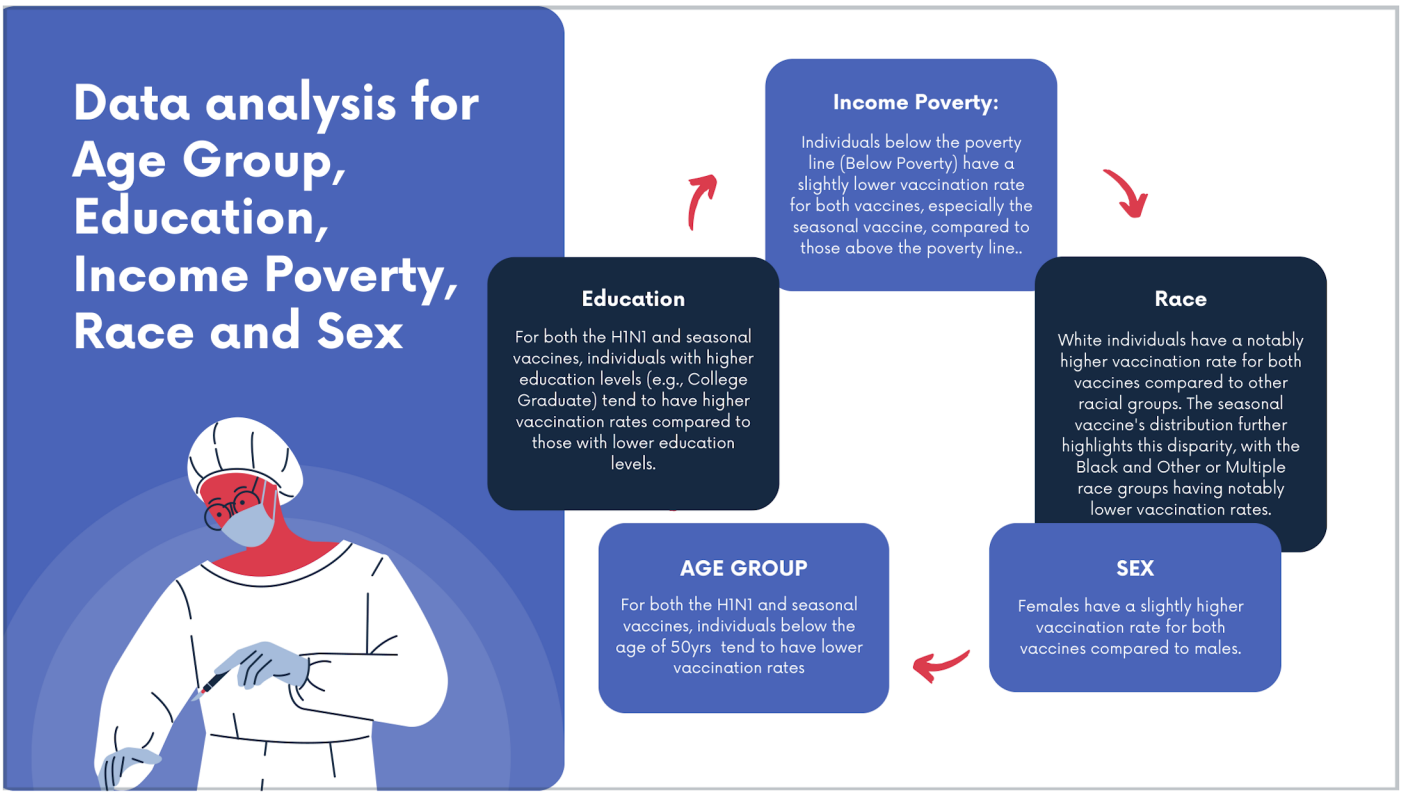
Analysis for h1n1_concern', 'h1n1_knowledge', 'behavioral_antiviral_meds', 'health_insurance', 'opinion_h1n1_risk'



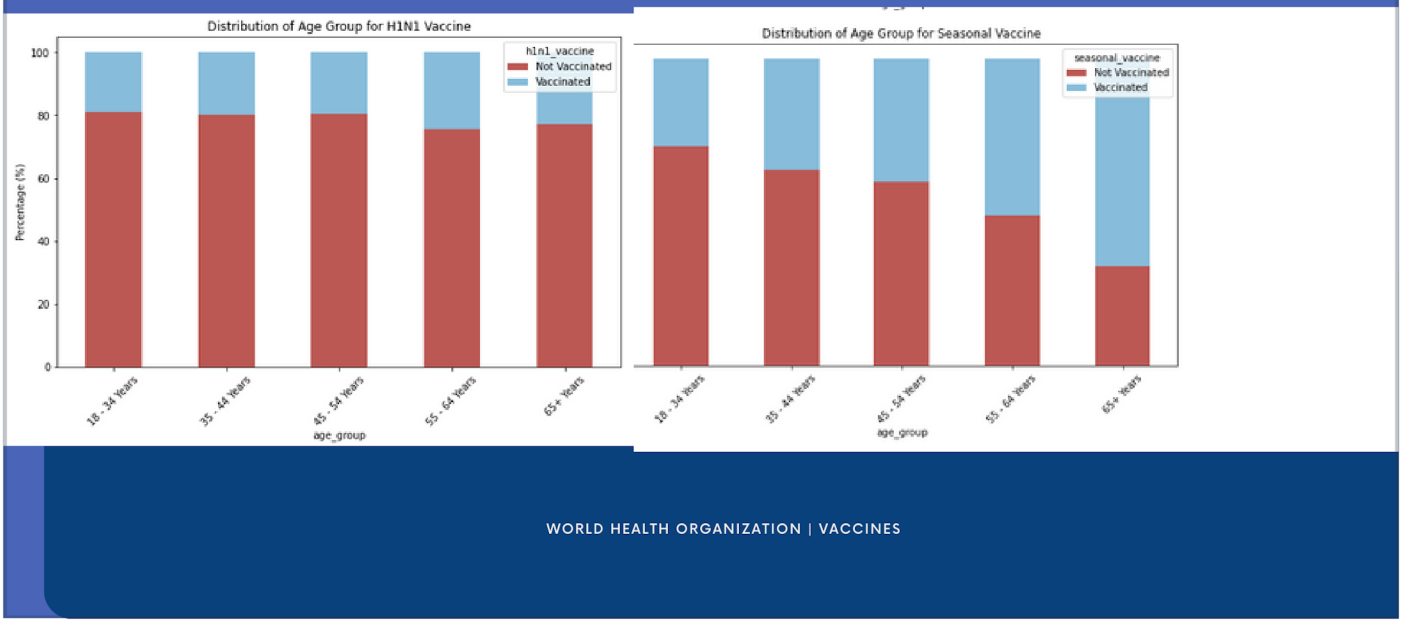


Analysis for h1n1_concern', 'h1n1_knowledge', 'behavioral_antiviral_meds', 'health_insurance', 'opinion_h1n1_risk'



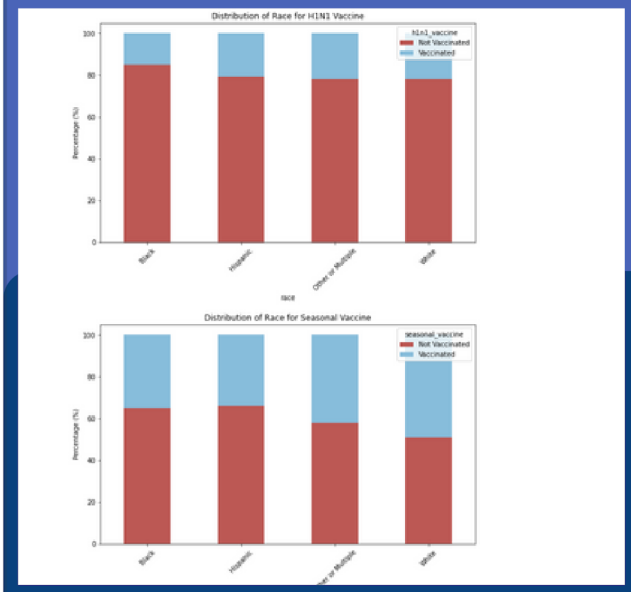

Data analysis for Age Group, Education, Income Poverty, Race and Sex
Income Poverty:
Individuals below the poverty line (Below Poverty) have a slightly lower vaccination rate for both vaccines, especially the seasonal vaccine, compared to those above the poverty line..
Race
White individuals have a notably higher vaccination rate for both vaccines compared to other racial groups. The seasonal vaccine's distribution further highlights this disparity, with the Black and Other or Multiple race groups having notably lower vaccination rates.
SEX
Females have a slightly higher vaccination rate for both vaccines compared to males.
AGE GROUP
For both the H1N1 and seasonal vaccines, individuals below the age of 50yrs tend to have lower vaccination rates
Education
For both the H1N1 and seasonal vaccines, individuals with higher education levels (e.g., College Graduate) tend to have higher vaccination rates compared to those with lower education levels.

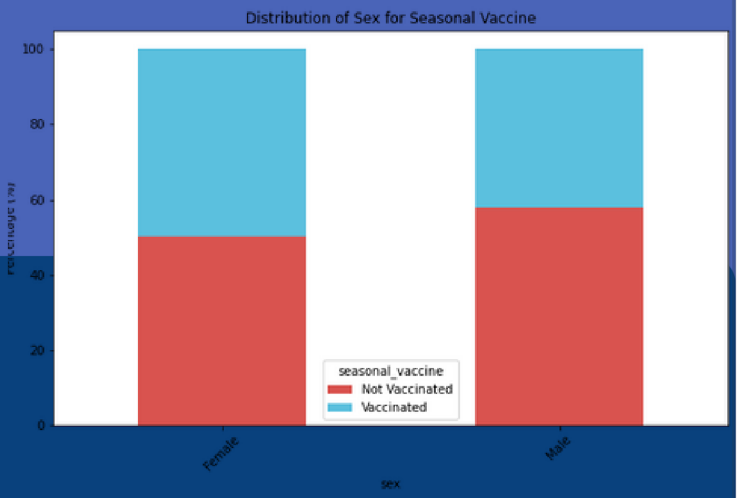
Data analysis for Age Group, Education, Income Poverty, Race and Sex



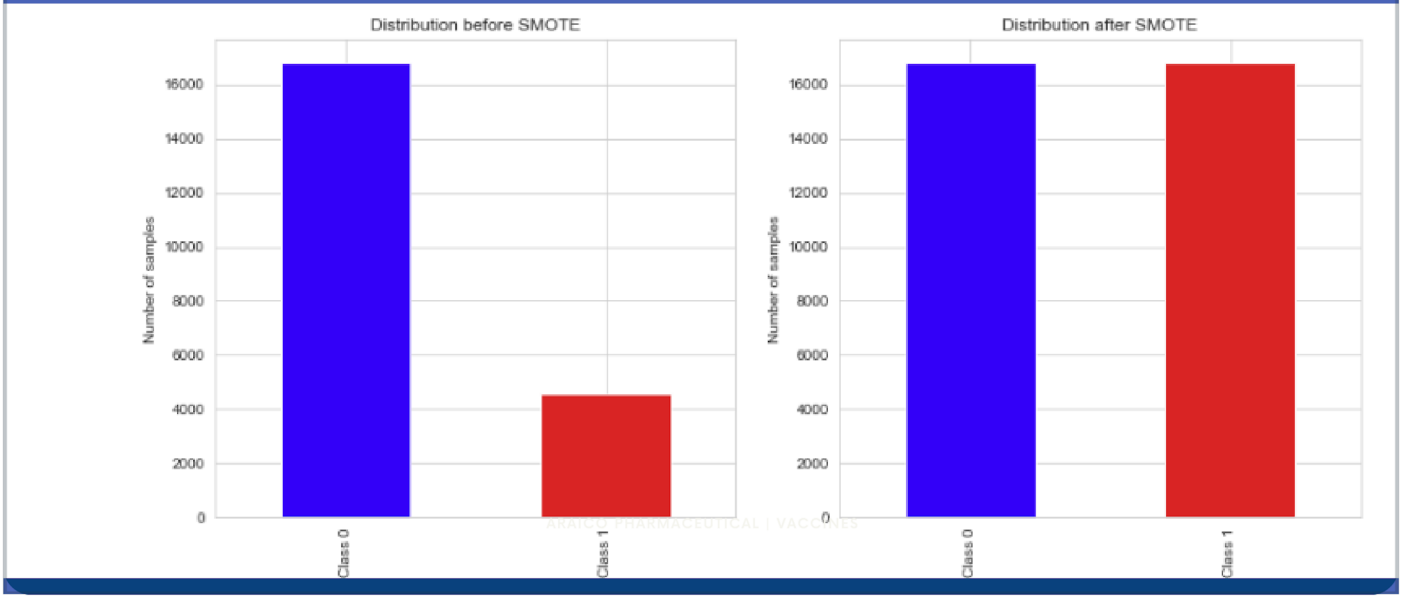
WORLD HEALTH ORGANIZATION | VACCINES

Data analysis for Age Group, Education, Income Poverty, Race and Sex





CLASS DISTRIBUTION

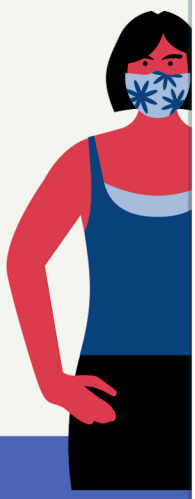


FIRST MODEL: LOGISTIC REGRESSION

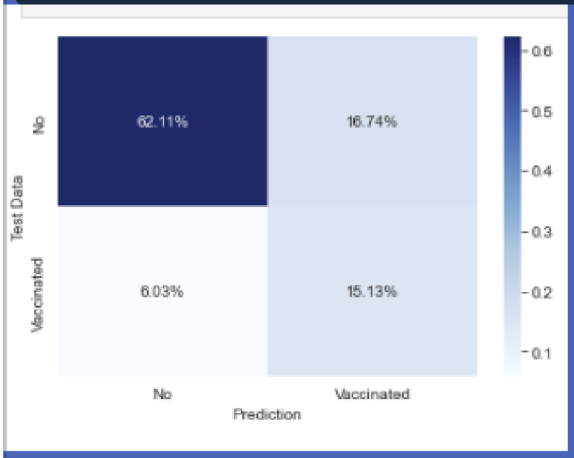
The model performs relatively well for both vaccines, but there's room for improvement, especially in recall for h1n1_vaccine. The lower recall indicates that the model might be missing a significant portion of individuals who actually received the H1N1 vaccine. The results for seasonal_vaccine are more balanced, with both precision and recall being in the mid-70s. These metrics provide a comprehensive view of the model's performance.

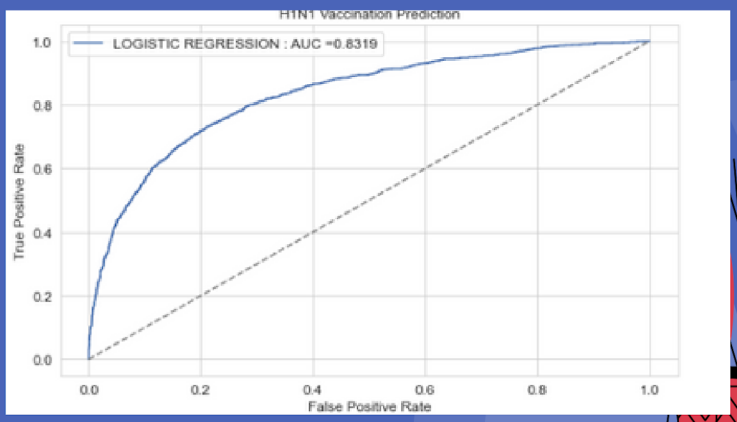
SECOND MODEL: LOGISTIC REGRESSION AFTER HANDLING CLASS IMBALANCE h1n1_vaccine

The model's accuracy after applying SMOTE is slightly lower than the model trained on the original imbalanced dataset. However, the recall for Class 1 (Vaccinated) has seen a significant improvement (from 43% in the orignal model to 72% in the SMOTE model). This improved recall indicates that the model is better at identifying individuals who actually received the H1N1 vaccine.
There is a trade-off in precision for Class 1, which has decreased to 49%.



MODEL 2 :Confusion Matrix for H1N1 ; ROC-AUC





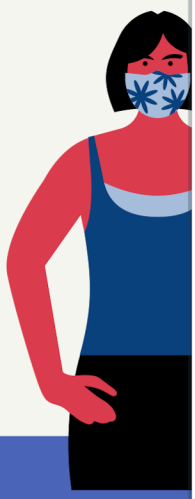
WORLD HEALTH ORGANIZATION | VACCINES

THIRD MODEL: RANDOM FOREST CLASSIFIER

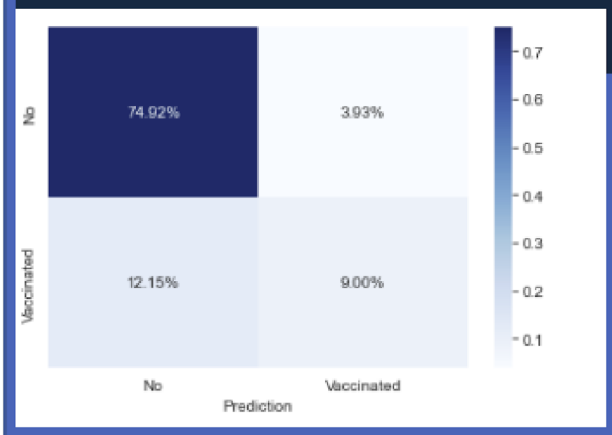
The model's accuracy for the h1n1_vaccine target is quite high at 83.86%. For predicting individuals who did not receive the H1N1 vaccine (Class 0), the model performs exceptionally well with a high precision, recall, and F1-score. For predicting individuals who received the H1N1 vaccine (Class 1), while the precision is decent, the recall is relatively low, indicating that there are a significant number of false negatives (individuals who received the vaccine but were predicted as not receiving it). This is evident from the F1-score of 53% for Class 1, indicating that there's a balance to be achieved between precision and recall for this class.

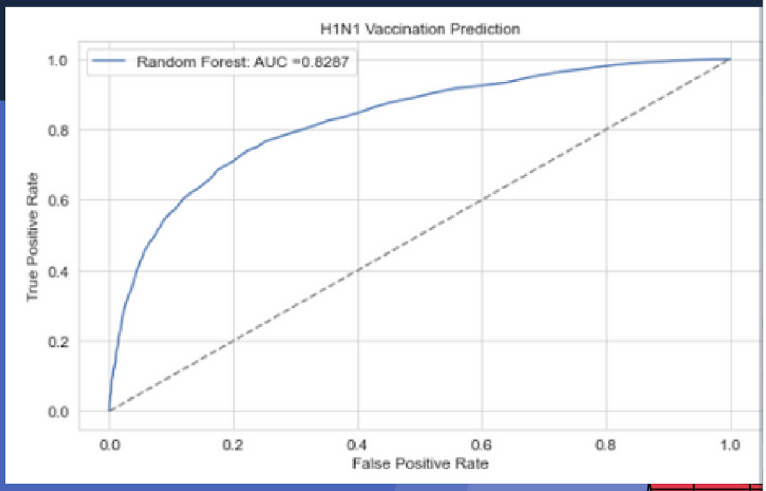
Final Model (Gradient Booster Classifier)

For H1N1 Flu predictions, the model performs relatively well with an accuracy of 84%. It performs better at identifying true negatives than true positives.
For Seasonal Flu, the model has a lower overall accuracy of 63%. The model is very good at identifying negatives (high recall for class 0) but struggles significantly with identifying true positives (low recall for class 1).
In both cases, the model seems to be more biased towards predicting the negative class (either for H1N1 or Seasonal Flu), as indicated by the higher recall for class 0 in both models



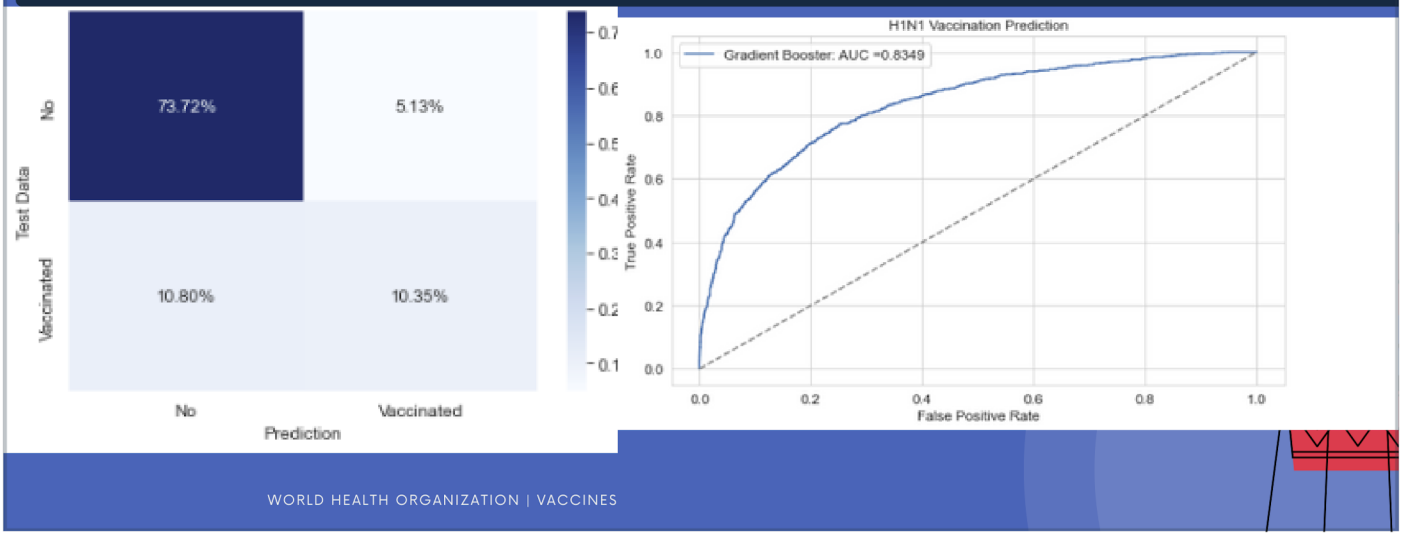
MODEL 3:Confusion Matrix for H1N1 ; ROC-AUC





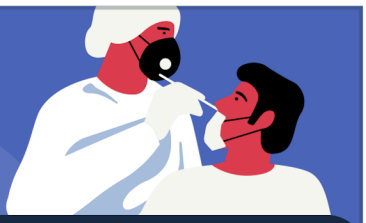
WORLD HEALTH ORGANIZATION | VACCINES

MODEL 4:Confusion Matrix for H1N1 ; ROC-AUC



WORLD HEALTH ORGANIZATION | VACCINES

RESULTS AND CONCLUSIONS



Seasonal Flu:

- Precision: Gradient Boosting has the highest precision, followed closely by Logistic Regression and then Random Forest.
- Recall: Logistic Regression has the highest recall, followed closely by Random Forest, while Gradient Boosting lags significantly.
- F1-score: Logistic Regression and Random Forest have much higher F1-scores than Gradient Boosting for seasonal_vaccine, given their balanced precision and recall.
- Accuracy: Logistic Regression has the highest accuracy, followed closely by Random Forest, with Gradient Boosting having notably lower accuracy.

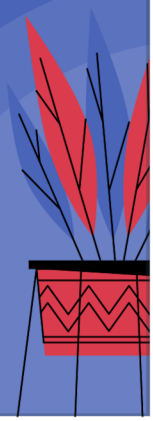
H1N1 Flu:

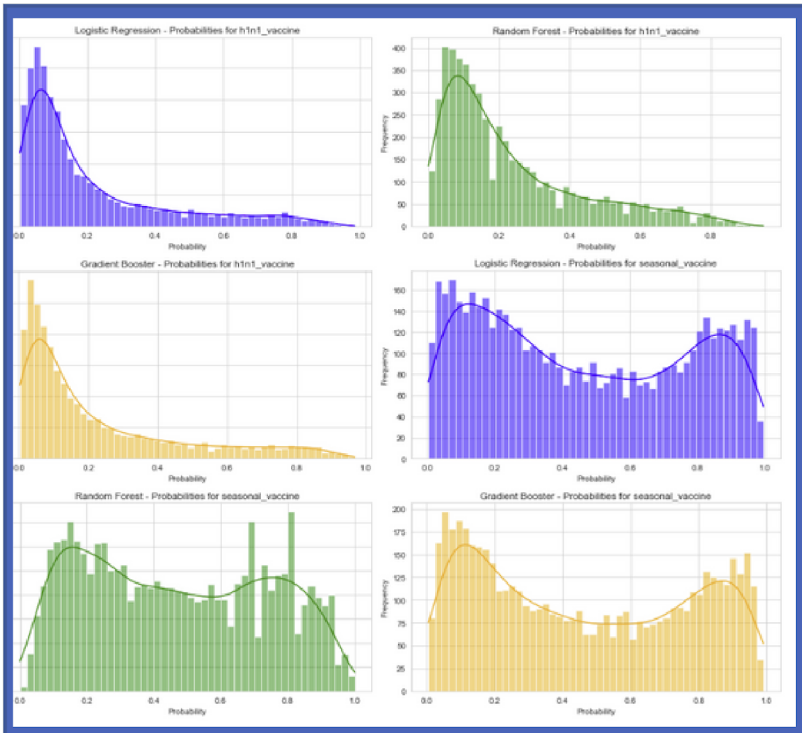
- Precision: Random Forest and Gradient Boosting have similar precision, both higher than Logistic Regression.
- Recall: Logistic Regression has the highest recall, meaning it correctly identifies a larger percentage of actual positives than the other two models.
- F1-score: Given the balance of precision and recall, the F1-scores are quite close for all three models.
- Accuracy: Both Gradient Boosting and Random Forest have the same accuracy, which is higher than that of Logistic Regression.

MODEL COMPARISON

Metric/Model	Gradient Boosting	Logistic Regression	Random Forest
Precision (Class 1)	66%	48%	69%
Recall (Class 1)	48%	72%	42%
F1-score (Class 1)	56%	58%	52%
Accuracy	84%	78%	84%
Macro Avg F1-score	73%	71%	71%

Metric/Model	Gradient Boosting	Logistic Regression	Random Forest
Precision (Class 1)	79%	78%	77%
Recall (Class 1)	26%	75%	74%
F1-score (Class 1)	40%	76%	76%
Accuracy	63%	79%	78%
Macro Avg F1-score	56%	79%	78%





WORLD HEALTH ORGANIZATION | VACCINES

H1N1 Flu:
Overall: For predicting h1n1_vaccine, all three models offer competitive performance, with slight variations in precision, recall, and F1-score.
For predicting seasonal_vaccine, while Gradient Boosting offers the highest precision, its recall is significantly lower than that of Logistic Regression and Random Forest, leading to a much lower F1-score and overall accuracy.

RECOMMENDATIONS

Public Awareness and Education:

- H1N1 Concern & Knowledge: Since a significant number of respondents have moderate to high concern and knowledge about H1N1, it suggests that public awareness campaigns have been somewhat effective. However, there's still room for improvement. .
- Education Level: Vaccination rates are higher among those with higher education. Efforts should be made to target awareness campaigns to

Health Infrastructure and Support:

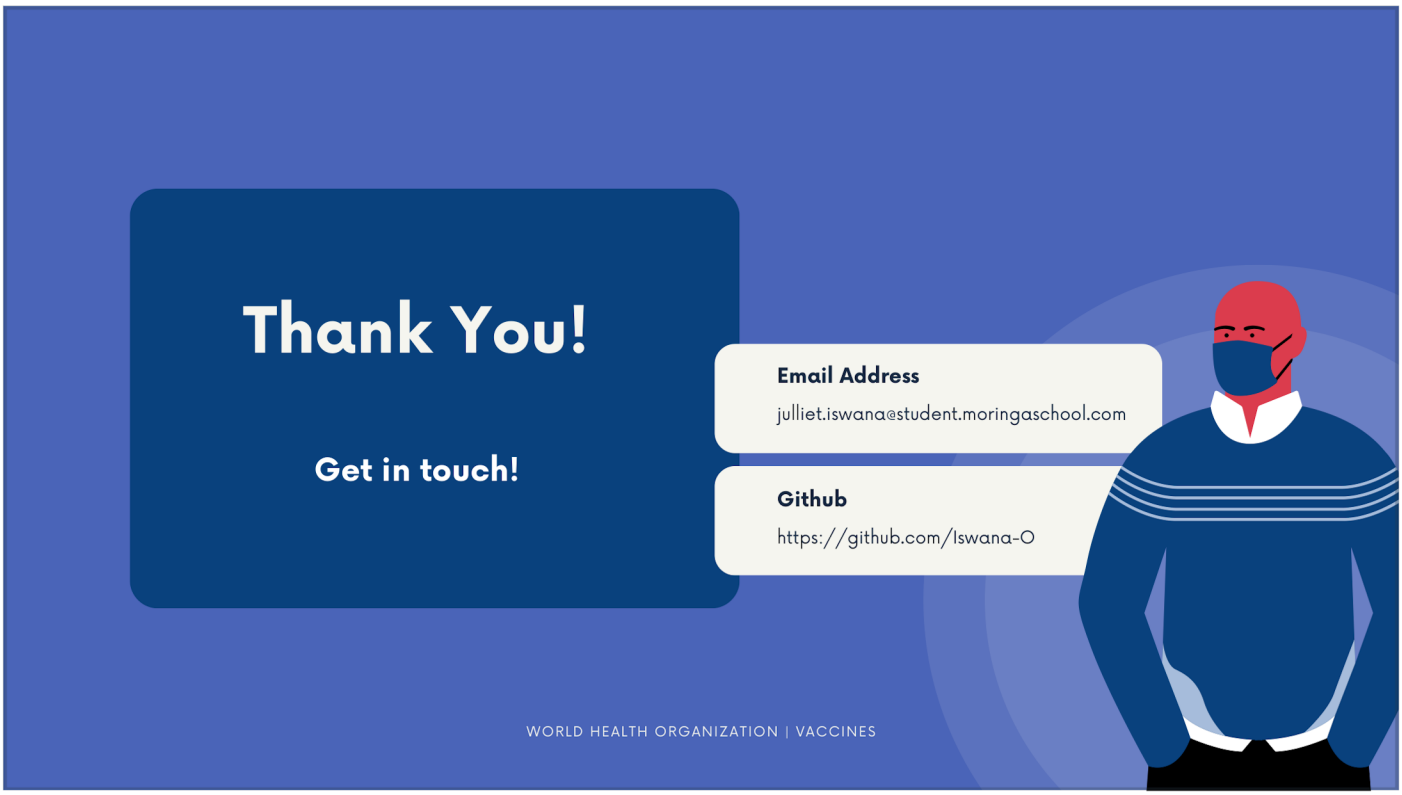
Health Insurance: A significant number of respondents do not have health insurance. Policymakers should consider expanding access to affordable health insurance, which could indirectly improve vaccination rates and general health outcomes.

Targeted Interventions:

- Race: There's a notable disparity in vaccination rates among racial groups. Targeted interventions and campaigns should be developed to address the specific concerns and barriers faced by the racial groups with lower vaccination rates.
- Sex: While the difference is slight, efforts can be made to ensure that both males and females have equal access to information and vaccination opportunities

Model Recommendations for Predictive Analytics:

- **H1N1 Flu Predictions: Given that all three models (Random Forest, Gradient Boosting, and Logistic Regression) have competitive performance, it might be useful to consider an ensemble approach**
- **Seasonal Flu Predictions: Given the significantly lower recall of Gradient Boosting, it might not be the best choice for predicting seasonal_vaccine, especially if identifying actual positive cases is crucial. Logistic Regression seems to be a balanced choice for this task.**


Thank You!
Get in touch!
Email Address
julliet.iswana@student.moringaschool.com
Github
https://github.com/Iswana-O
WORLD HEALTH ORGANIZATION | VACCINES