H1N1 VACCINES PREDICTION

FELIX MBURU NJOROGE

DSF-PT06

BACKGROUND

- **Vaccination is a key public health measure to fight infectious diseases.**
- **Vaccines provide immunization for individuals and enough vaccination in a community can reduce further spread of a diseases through herd immunity.**
- **This phone survey asked respondents whether they had received the H_1N_1 and seasonal flu vaccines, in conjunction with questions about themselves.**
- **These additional questions covered their social, economic, and demographic background, opinions on risks of illness and vaccine effectiveness, and behaviors towards mitigating transmission.**

Objectives

1. Predict Vaccination Uptake Based on Demographic and Socioeconomic Factors:

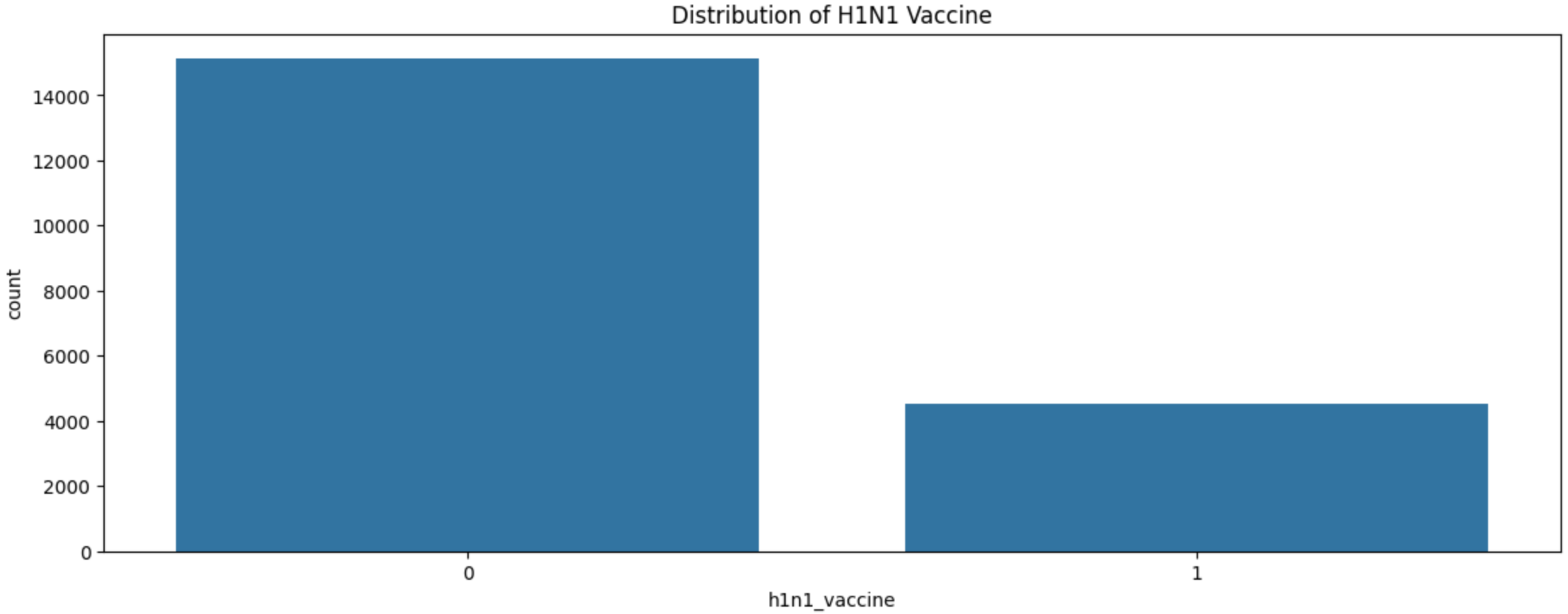
2. Forecast Future Vaccination Rates:

3. Analyze the Impact of Public Perceptions on Vaccination Behavior:

4. Evaluate the Effectiveness of Preventive Behaviors on Vaccination Rates:

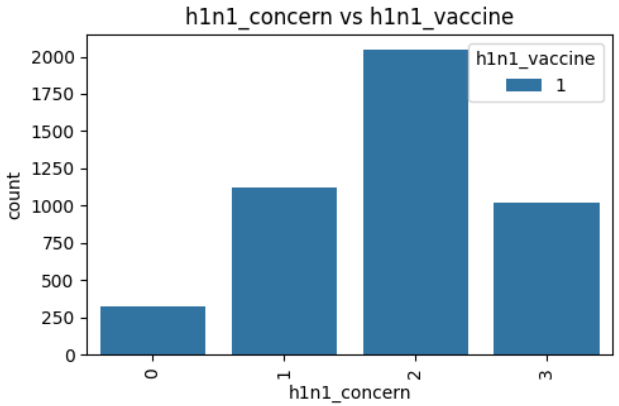
EDA

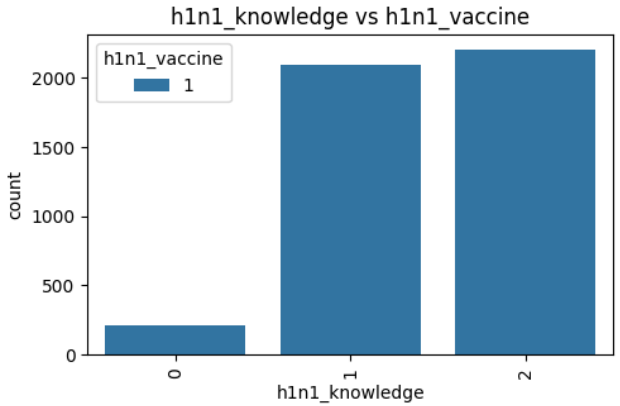
- Univariate analysis

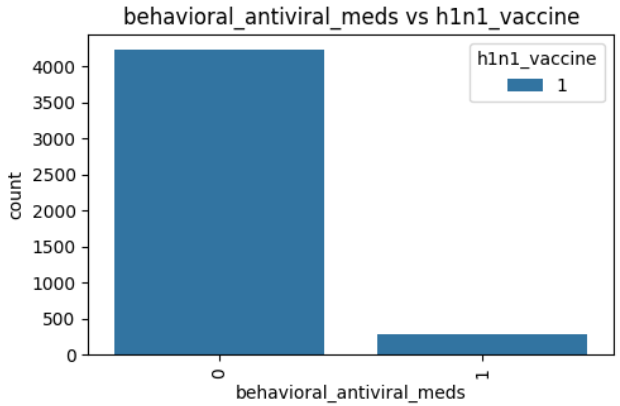


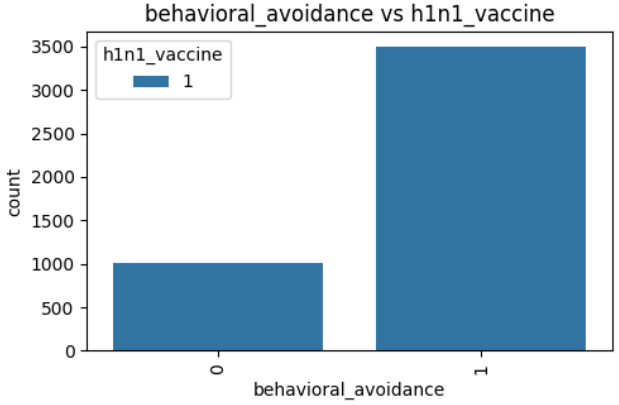
Class imbalance problem

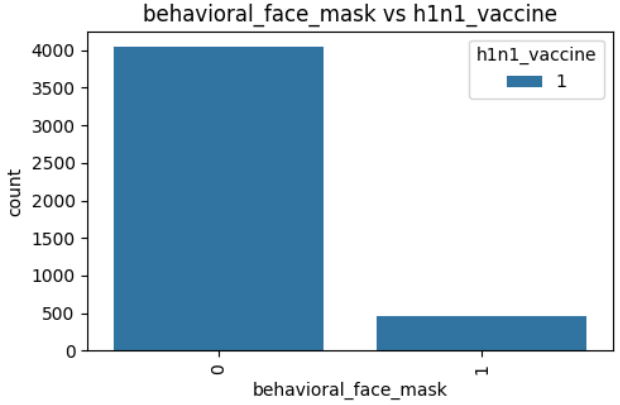
Bivariate analysis

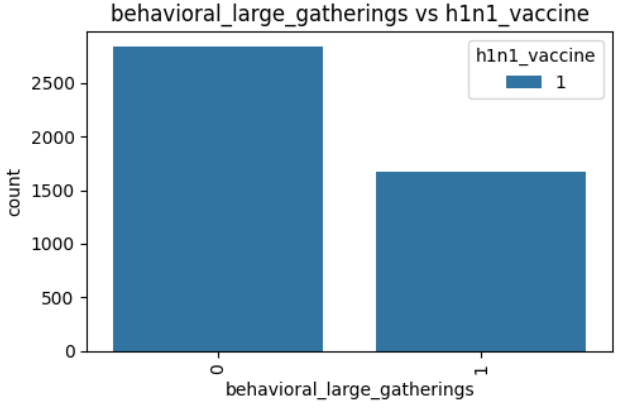


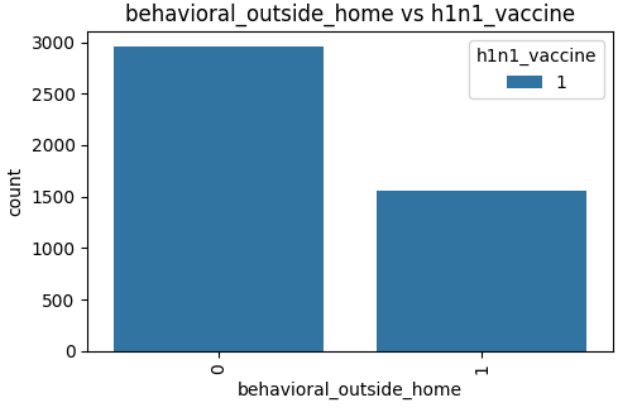


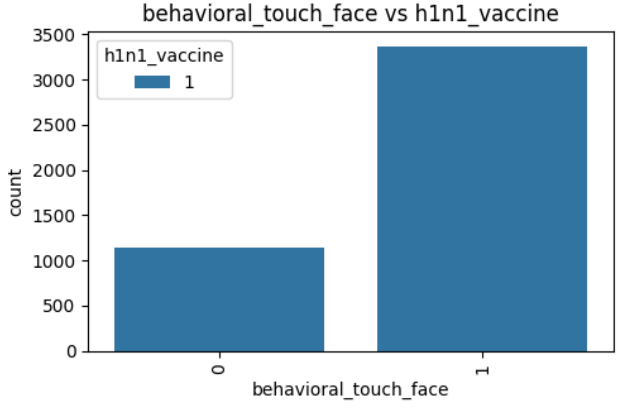


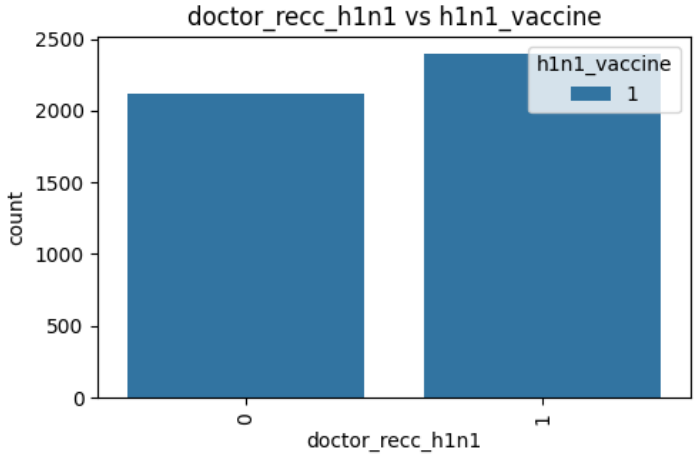


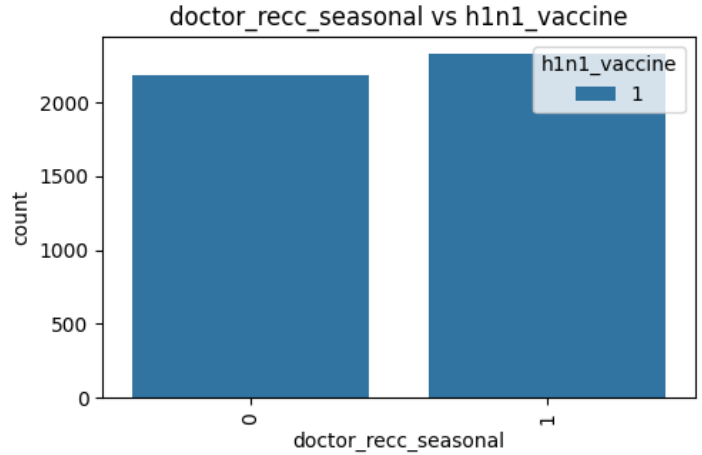


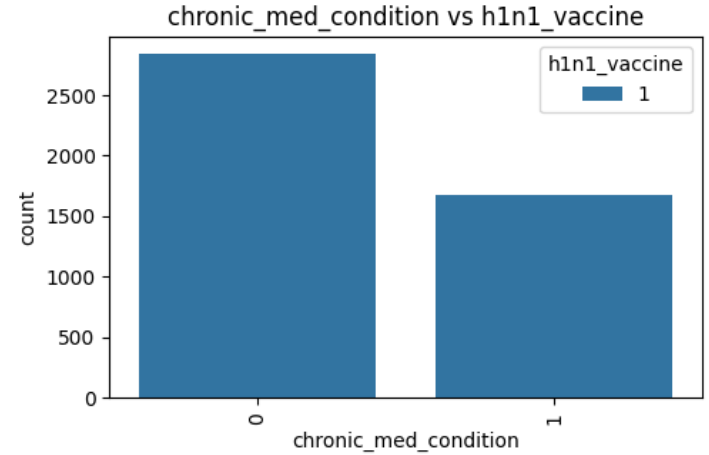


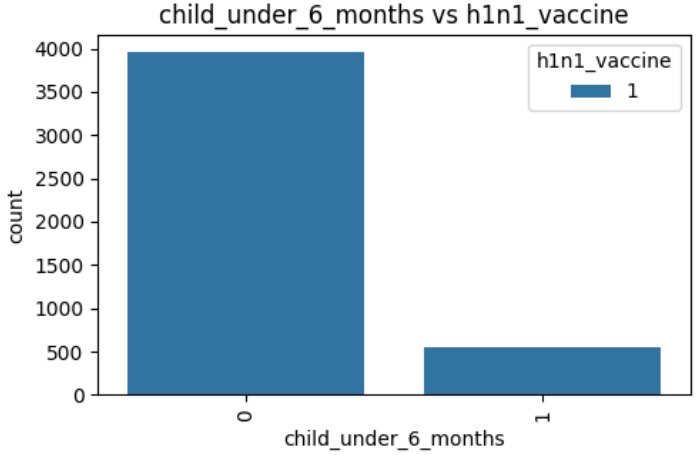


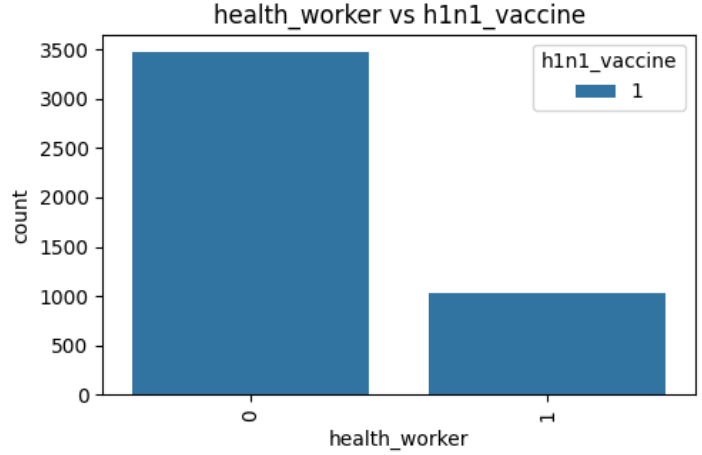


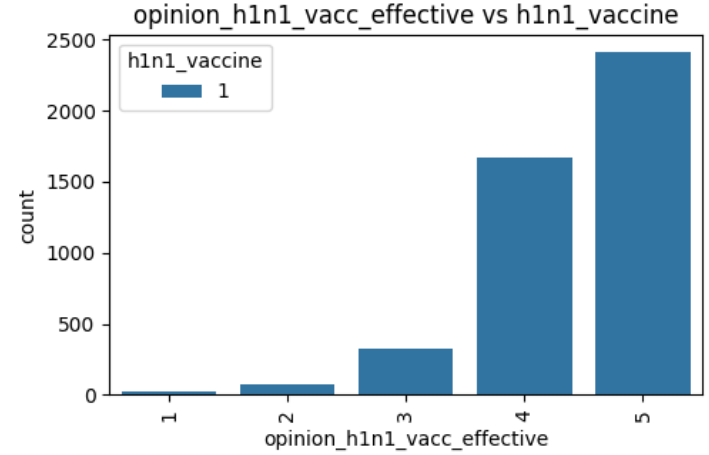

doctor_recc_h1n1 vs h1n1_vaccine
h1n1_vaccine
1
count
doctor_recc_h1n1
0
1

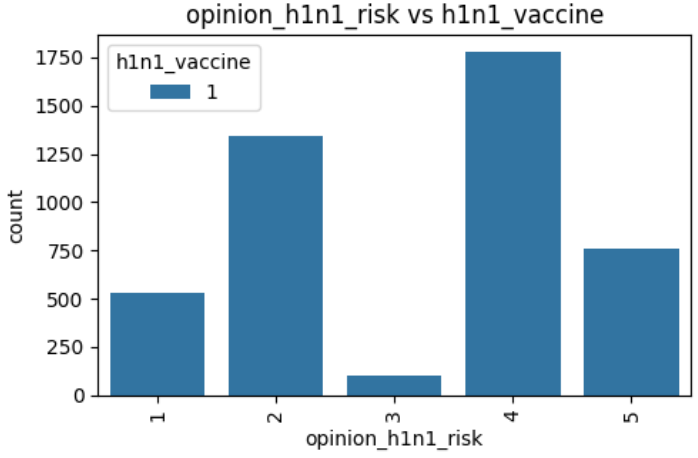

doctor_recc_seasonal vs h1n1_vaccine
h1n1_vaccine
1
count
doctor_recc_seasonal
0
1

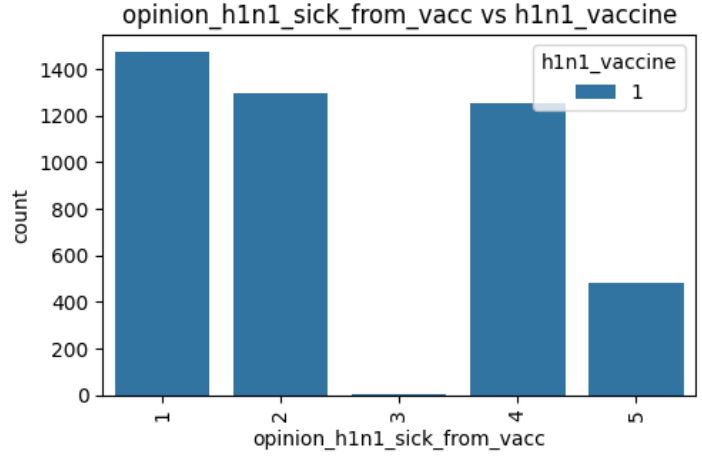

chronic_med_condition vs h1n1_vaccine
h1n1_vaccine
1
count
chronic_med_condition
0
1

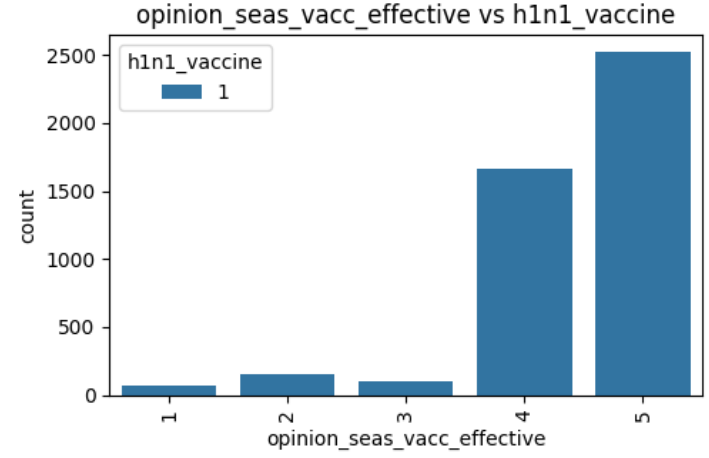

child_under_6_months vs h1n1_vaccine
h1n1_vaccine
1
count
child_under_6_months
0
1

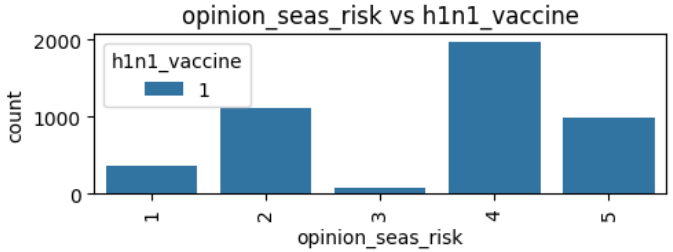

health_worker vs h1n1_vaccine
h1n1_vaccine
1
count
health_worker
0
1

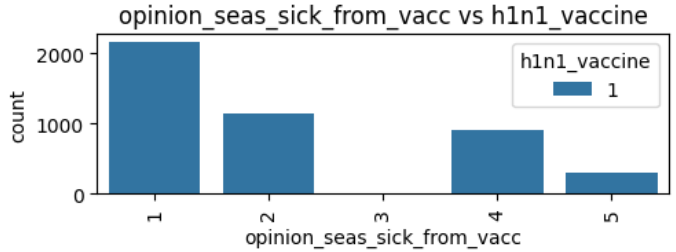

opinion_h1n1_vacc_effective vs h1n1_vaccine
h1n1_vaccine
1
count
opinion_h1n1_vacc_effective
1
2
3
4
5

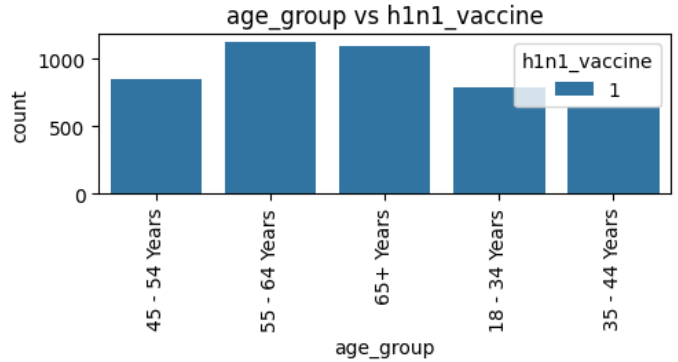

opinion_h1n1_risk vs h1n1_vaccine
h1n1_vaccine
1
count
opinion_h1n1_risk
1
2
3
4
5

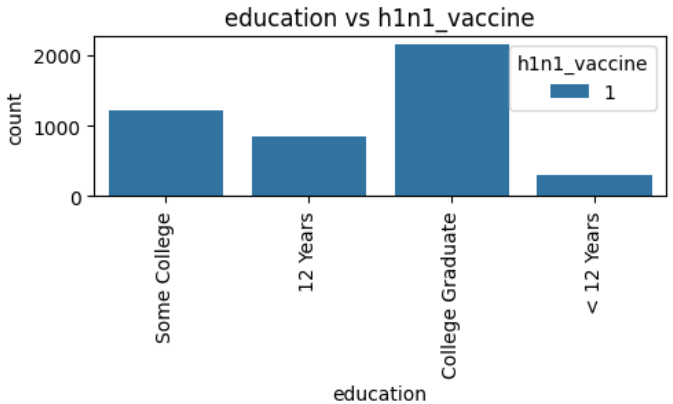

opinion_h1n1_sick_from_vacc vs h1n1_vaccine
h1n1_vaccine
1
count
opinion_h1n1_sick_from_vacc
1
2
3
4
5

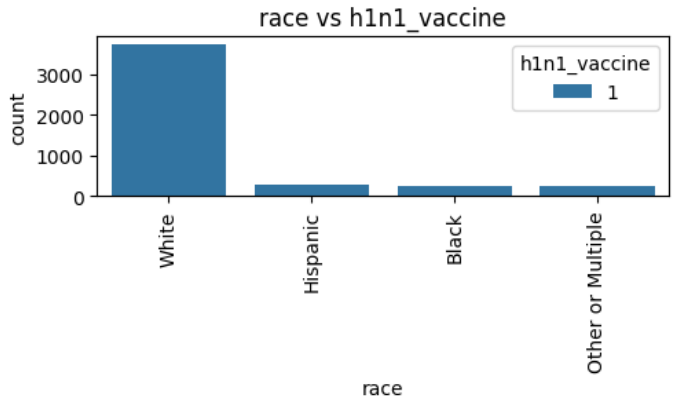

opinion_seas_vacc_effective vs h1n1_vaccine
h1n1_vaccine
1
count
opinion_seas_vacc_effective
1
2
3
4
5

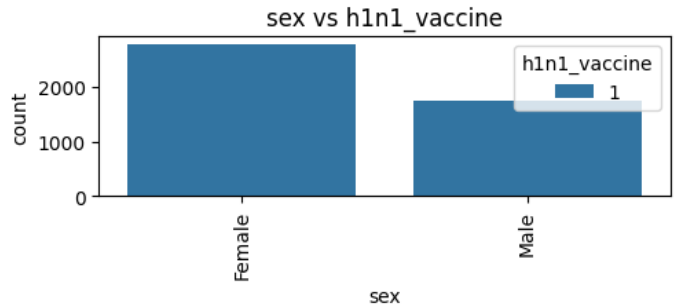

opinion_seas_risk vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
1
2
3
4
5
opinion_seas_risk

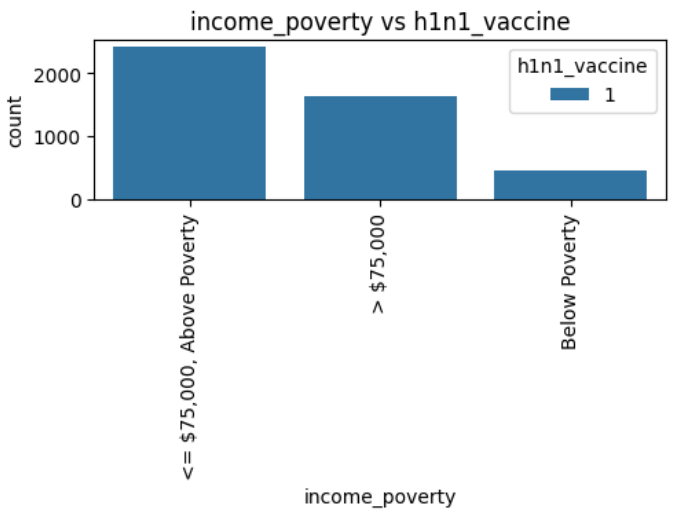

opinion_seas_sick_from_vacc vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
1
2
3
4
5
opinion_seas_sick_from_vacc

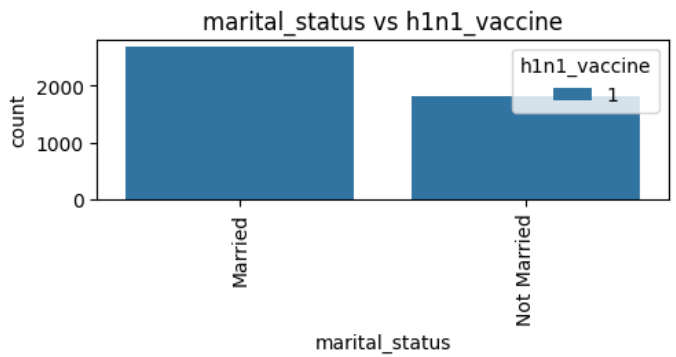

age_group vs h1n1_vaccine
h1n1_vaccine
1
count
1000
500
0
45 - 54 Years
55 - 64 Years
65+ Years
18 - 34 Years
35 - 44 Years
age_group

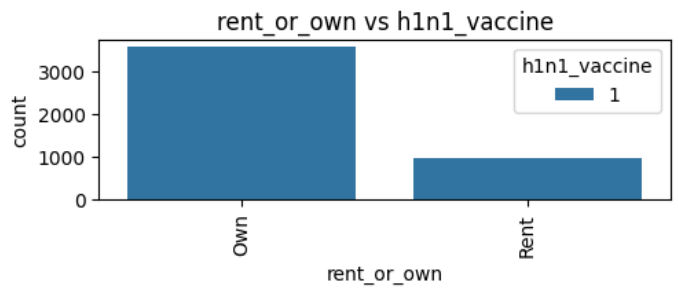

education vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
Some College
12 Years
College Graduate
< 12 Years
education

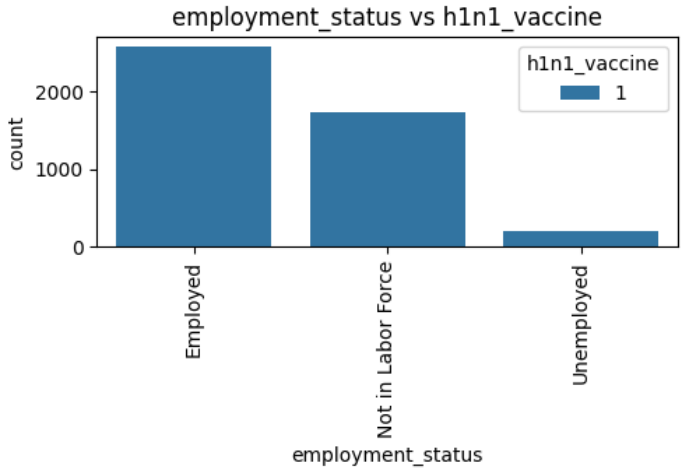

race vs h1n1_vaccine
h1n1_vaccine
1
count
3000
2000
1000
0
White
Hispanic
Black
Other or Multiple
race

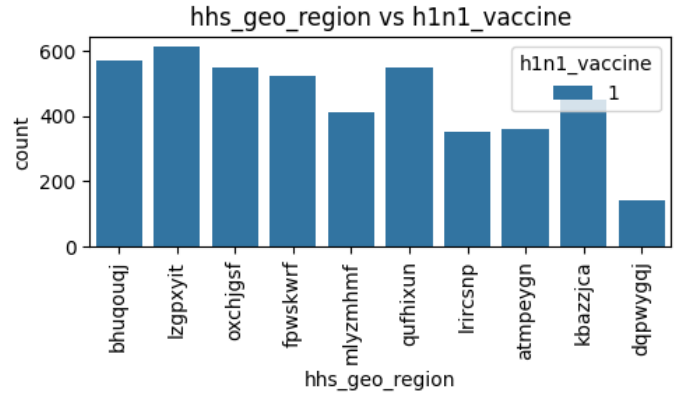

sex vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
Female
Male
sex

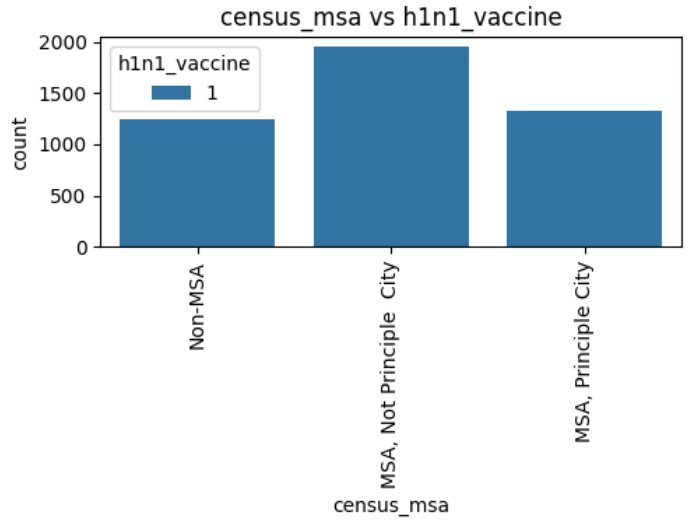

income_poverty vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
<= $75,000, Above Poverty
> $75,000
Below Poverty
income_poverty

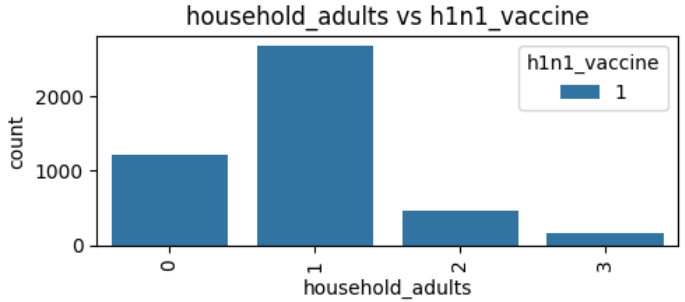

marital_status vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
Married
Not Married
marital_status

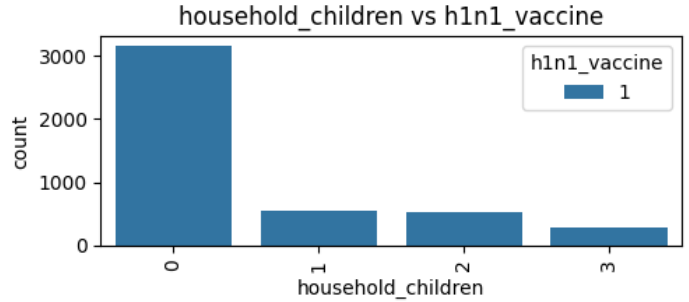

rent_or_own vs h1n1_vaccine
h1n1_vaccine
1
count
3000
2000
1000
0
Own
Rent
rent_or_own

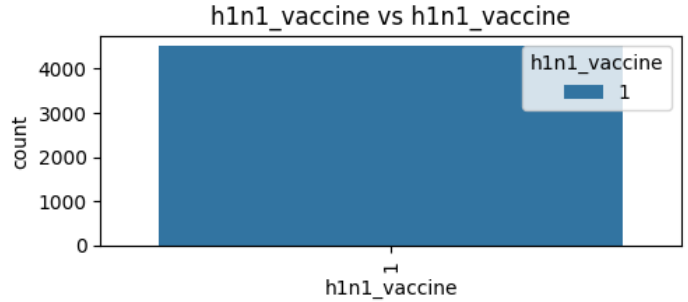

employment_status vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
Employed
Not in Labor Force
Unemployed
employment_status

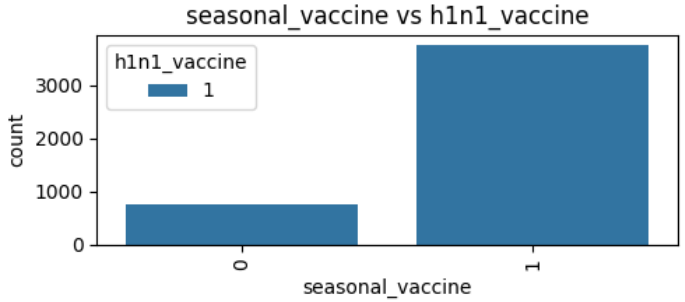

hhs_geo_region vs h1n1_vaccine
h1n1_vaccine
1
count
600
400
200
0
bhuqouqj
lzgpxyit
oxchjgsf
fpwskwrf
mlyzmhmf
qufhixun
lrircsnp
atmpeygn
kbazzjca
dqpwygqj
hhs_geo_region


census_msa vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1500
1000
500
0
Non-MSA
MSA, Not Principle City
MSA, Principle City
census_msa

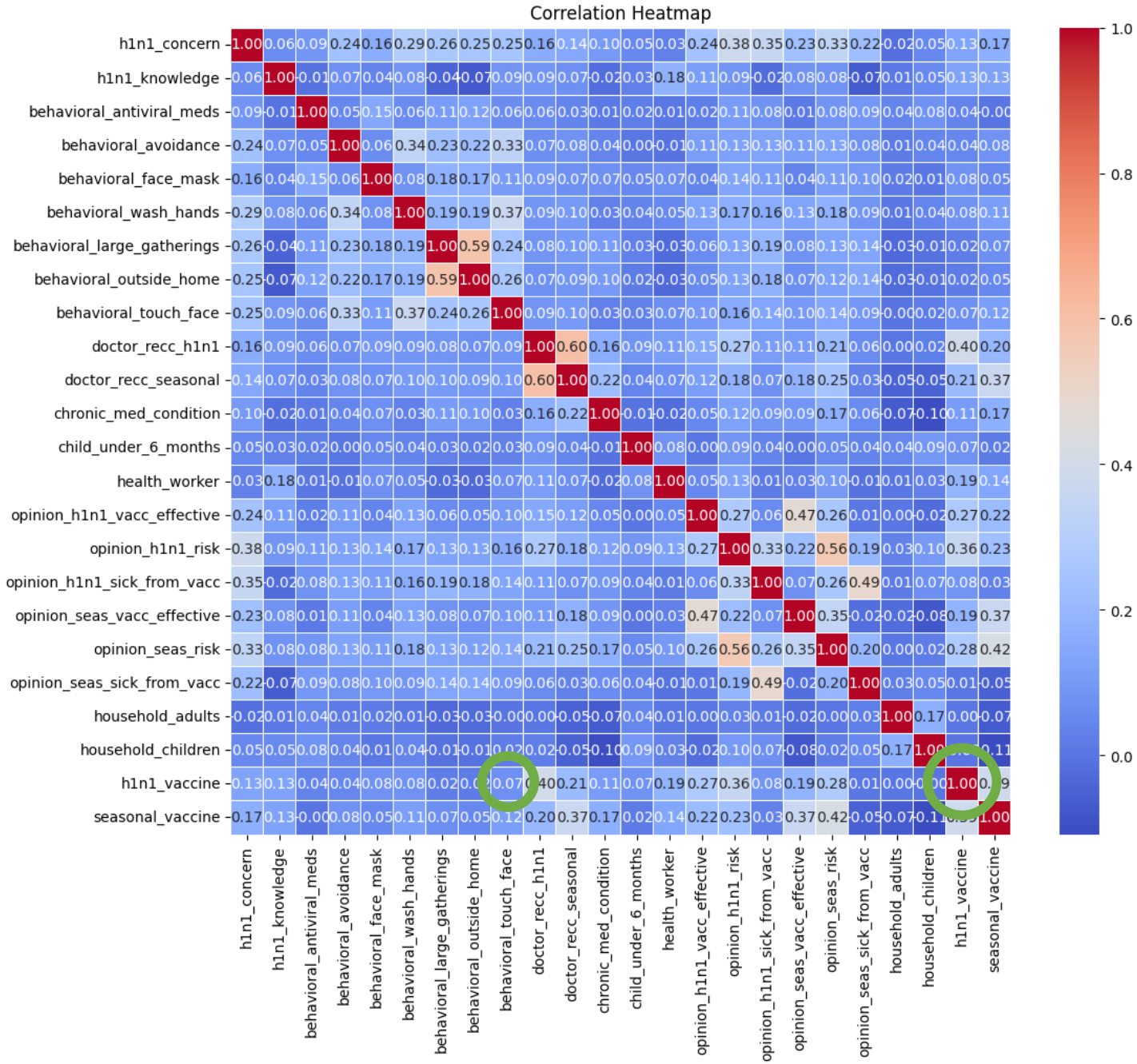

household_adults vs h1n1_vaccine
h1n1_vaccine
1
count
2000
1000
0
0
1
2
3
household_adults


household_children vs h1n1_vaccine
h1n1_vaccine
1
count
3000
2000
1000
0
0
1
2
3
household_children


h1n1_vaccine vs h1n1_vaccine
h1n1_vaccine
1
count
4000
3000
2000
1000
0
1
h1n1_vaccine

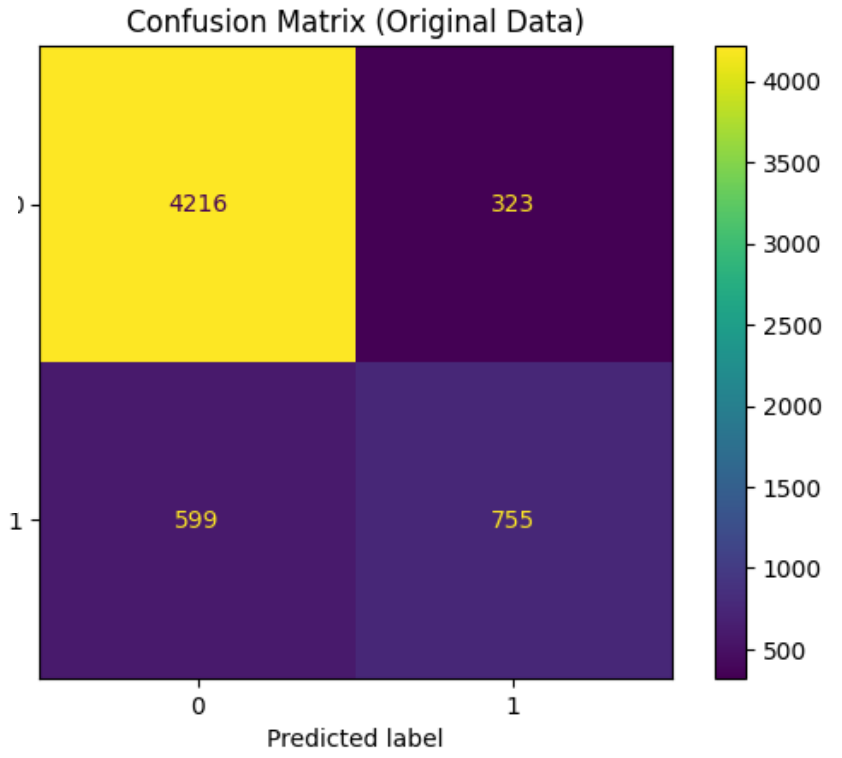

seasonal_vaccine vs h1n1_vaccine
h1n1_vaccine
1
count
3000
2000
1000
0
0
1
seasonal_vaccine

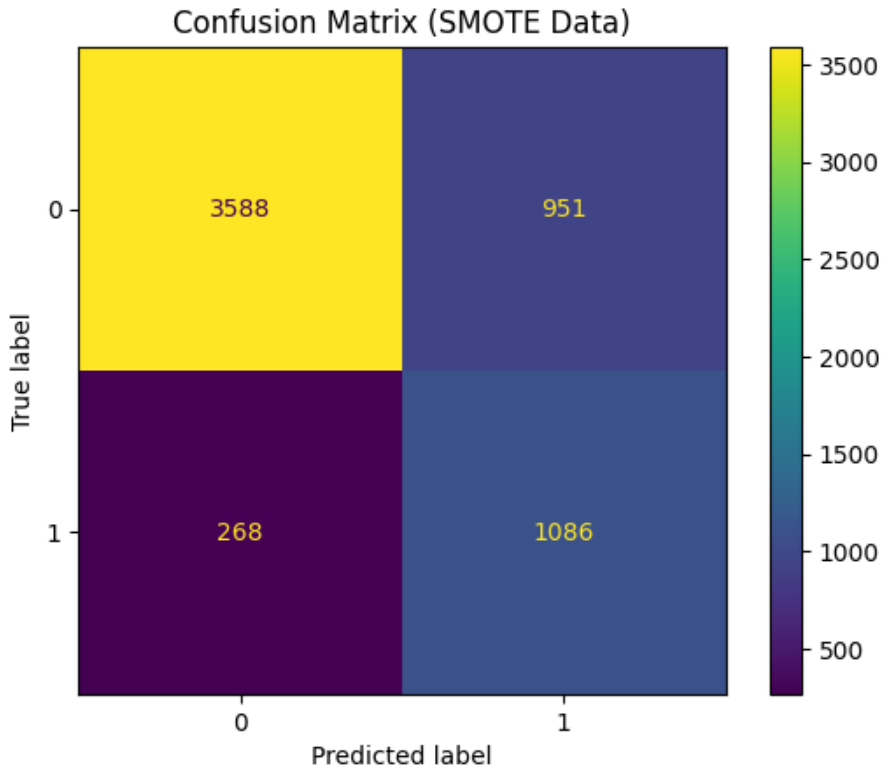
MULTIVARIATE ANALYSIS



CLASS IMBALANCE HANDLED USING SMOTE

```
Class distribution before SMOTE:
h1n1_vaccine
0    10589
1     3160
Name: count, dtype: int64
Class distribution after SMOTE:
h1n1_vaccine
1    10589
0    10589
Name: count, dtype: int64
```





LOGISTIC REGRESSION PERFORMANCE METRICS

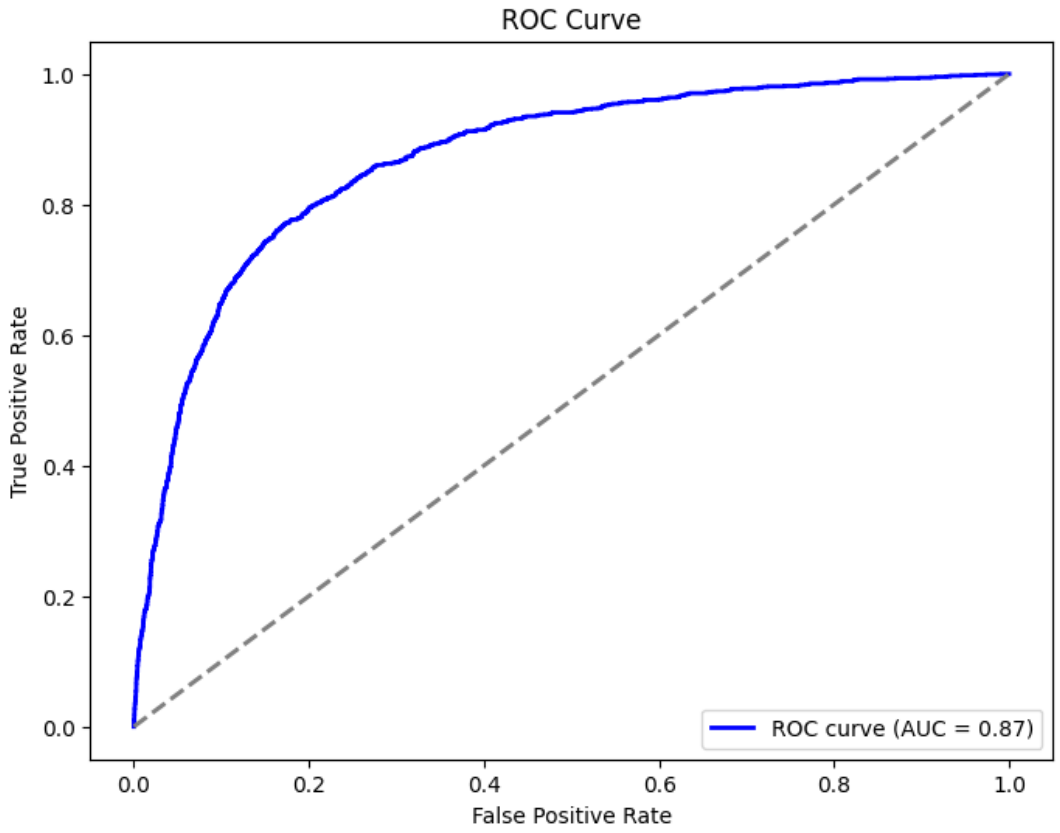
Train-test split

Classification Report (SMOTE Data):

	precision	recall	f1-score	support
0	0.93	0.79	0.85	4539
1	0.53	0.80	0.64	1354
accuracy			0.79	5893
macro avg	0.73	0.80	0.75	5893
weighted avg	0.84	0.79	0.81	5893

Cross validation

```
Cross-validation scores:
Accuracy: [0.80807365 0.81208687 0.82507082 0.81841795 0.826682
Mean Accuracy score: 0.8180663409043527
Precision: [0.81415503 0.80182648 0.82097902 0.80369203 0.81303
Mean Precision score: 0.810738028094392
Recall: [0.79839471 0.82908404 0.83144476 0.8427762  0.84837034
Mean Recall score: 0.8300140104188273
F1_score: [0.80619785 0.81522748 0.82617875 0.82277022 0.83032825]
Mean F1_score score: 0.820140511411752
```



AUC = 0.87

DECISION TREES PERFORMANCE

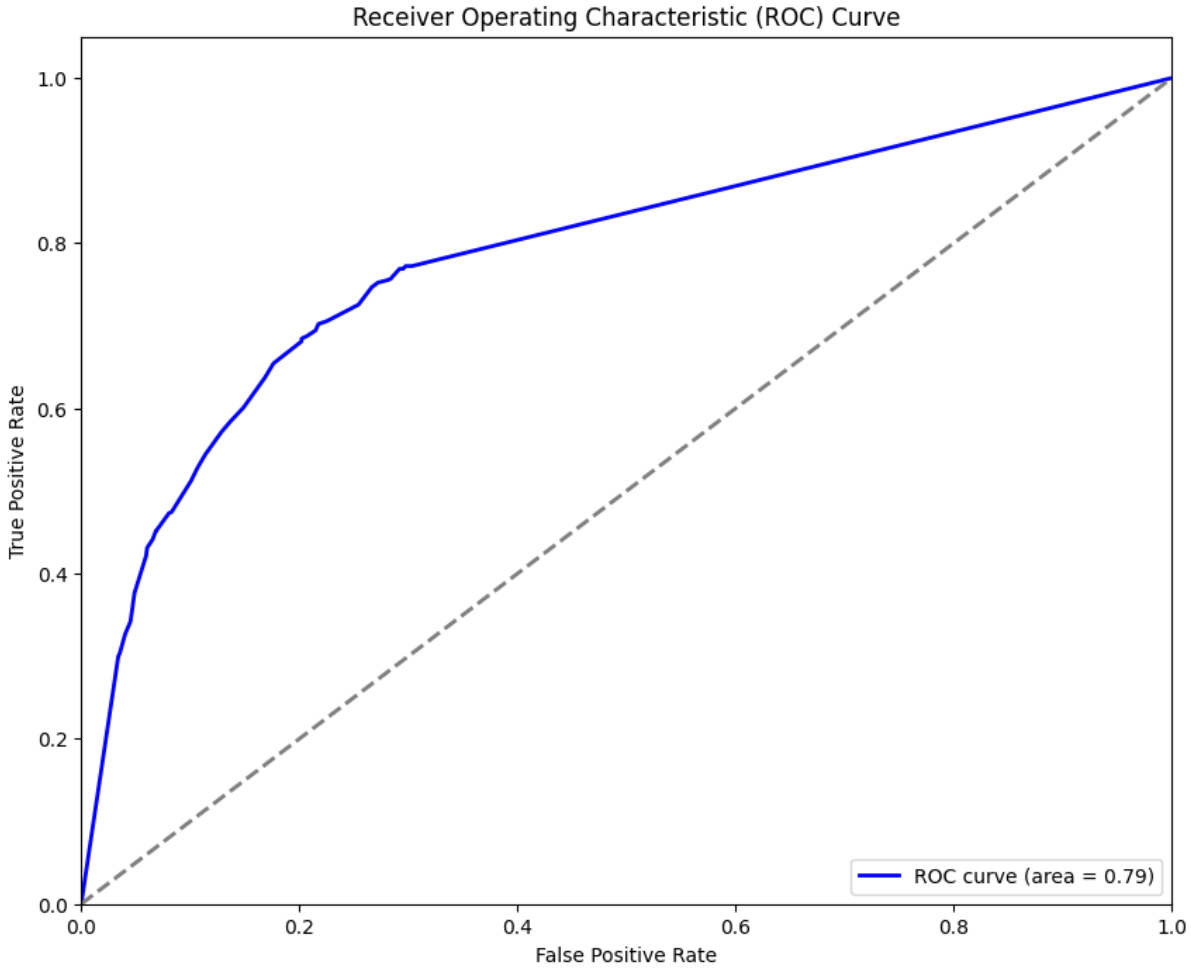
Accuracy

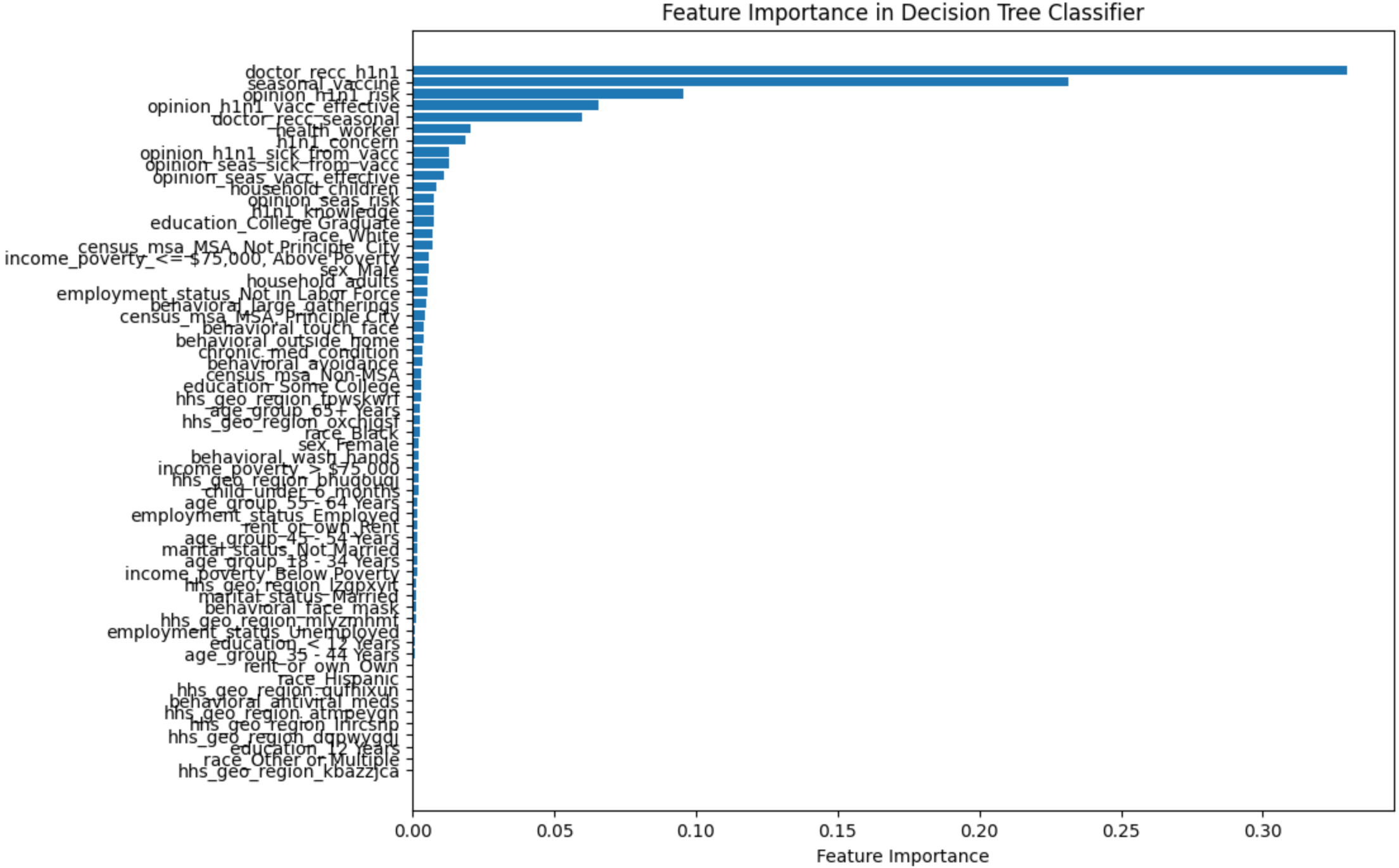
```
(0.7671163145838635,
    precision    recall  f1-score   support\r
     1       0.49      0.52      0.51        9
o avg       0.67      0.68      0.68       3
```

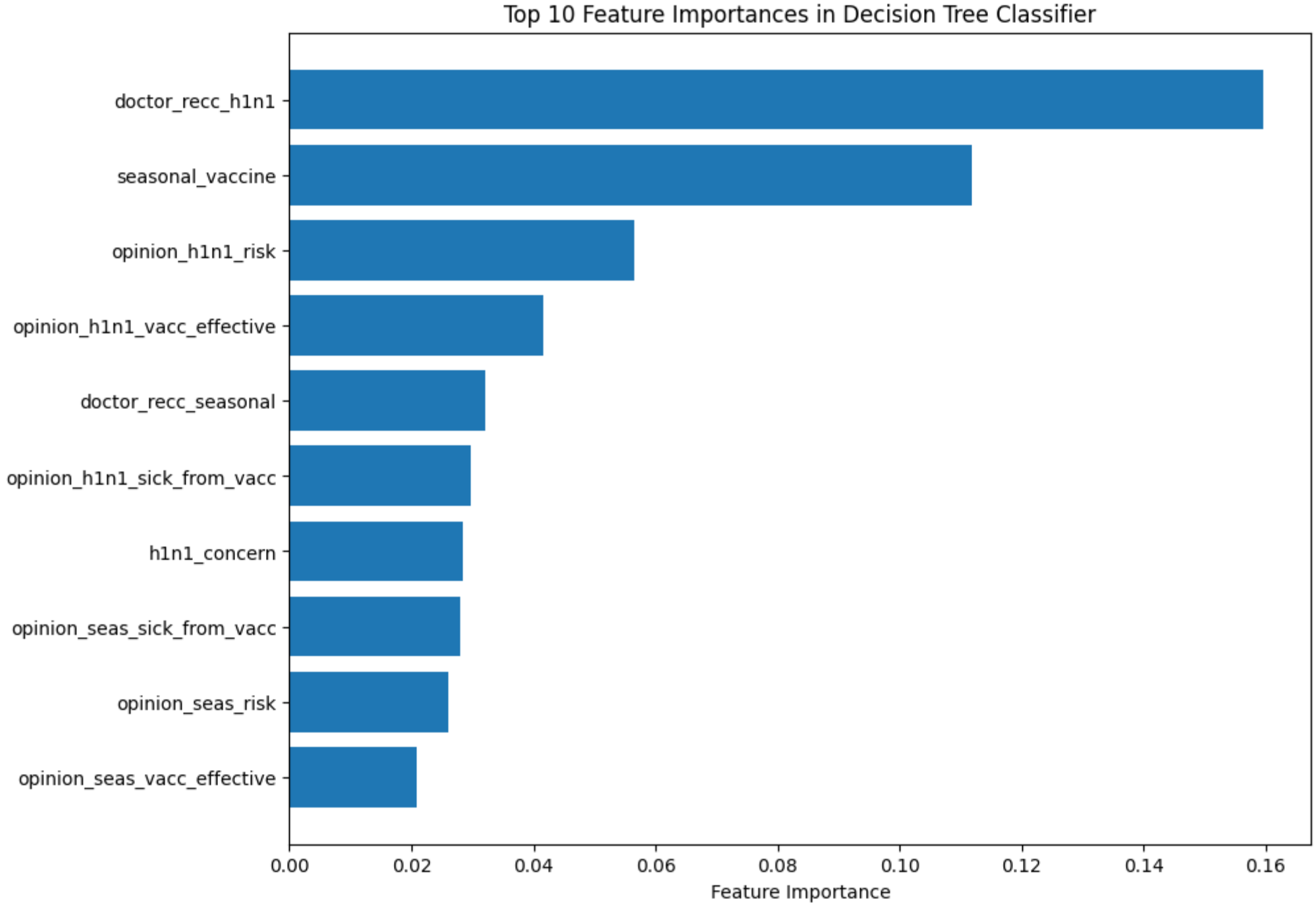
REPORT AFTER HYPERPARAMETER TUNING

```
Accuracy: 0.8078391448205651
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.89      0.88      3029
           1       0.59      0.54      0.56       900
```




Feature Importance in Decision Tree Classifier
doctor_recc_h1n1
seasonal_vaccine
opinion_h1n1_risk
opinion_h1n1_vacc_effective
doctor_recc_seasonal
health_worker
h1n1_concern
opinion_h1n1_sick_from_vacc
opinion_seas_sick_from_vacc
opinion_seas_vacc_effective
household_children
opinion_seas_risk
h1n1_knowledge
education_College Graduate
race_White
census_msa_MSA, Not Principle City
income_poverty_<=$75,000, Above Poverty
sex_Male
household_adults
employment_status_Not in Labor Force
behavioral_large_gatherings
census_msa_MSA, Principle City
behavioral_touch_face
behavioral_outside_home
chronic_med_condition
behavioral_avoidance
census_msa_Non-MSA
education_Some College
hhs_geo_region_fpwskwrf
age_group_65+ Years
hhs_geo_region_oxchjgsf
race_Black
sex_Female
behavioral_wash_hands
income_poverty_> $75,000
hhs_geo_region_bhuqbuqj
child_under_6_months
age_group_55 - 64 Years
employment_status_Employed
rent_or_own_Rent
age_group_45 - 54 Years
marital_status_Not Married
age_group_18 - 34 Years
income_poverty_Below Poverty
hhs_geo_region_lzgpxyit
marital_status_Married
behavioral_face_mask
hhs_geo_region_mlyzmhmf
employment_status_Unemployed
education_< 12 Years
age_group_35 - 44 Years
rent_or_own_Own
race_Hispanic
hhs_geo_region_qufhixun
behavioral_antiviral_meds
hhs_geo_region_atmpeygn
hhs_geo_region_lrircsnp
hhs_geo_region_dqpwygqj
education_12 Years
race_Other or Multiple
hhs_geo_region_kbazzjca
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Feature Importance

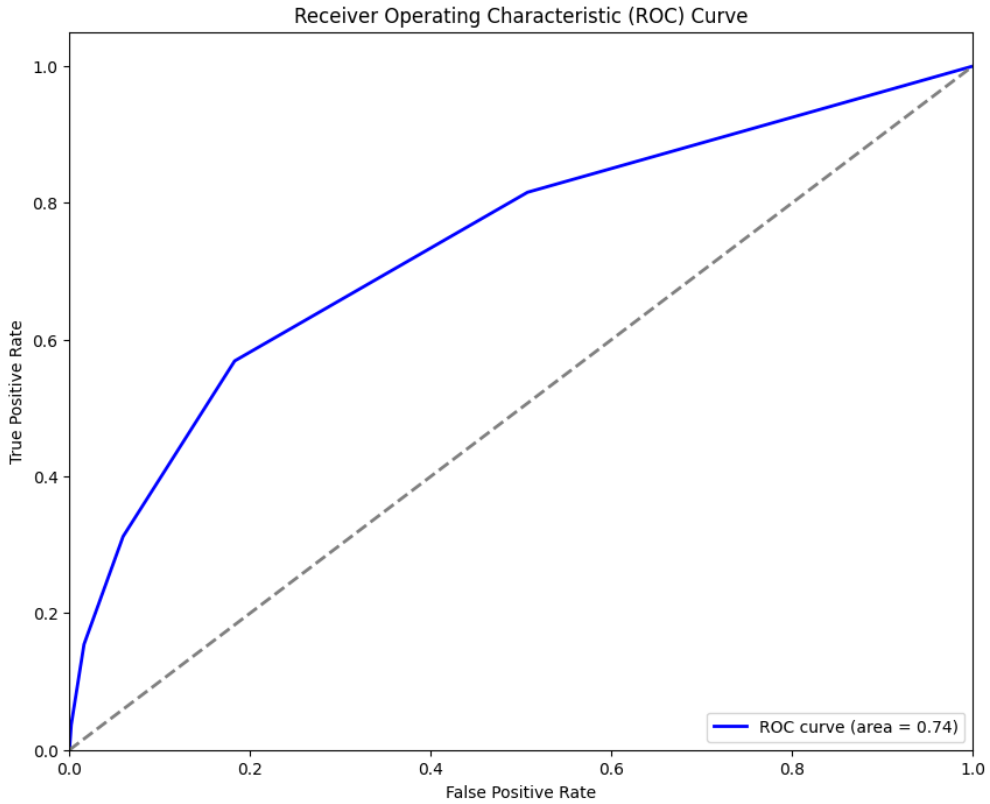

Top 10 Feature Importances in Decision Tree Classifier
doctor_recc_h1n1
seasonal_vaccine
opinion_h1n1_risk
opinion_h1n1_vacc_effective
doctor_recc_seasonal
opinion_h1n1_sick_from_vacc
h1n1_concern
opinion_seas_sick_from_vacc
opinion_seas_risk
opinion_seas_vacc_effective
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
Feature Importance

K NEAREST NEIGHBOUR

BEFORE HYPERPARAMETER TUNING

Accuracy: 0.7966403665054721
Classification Report:

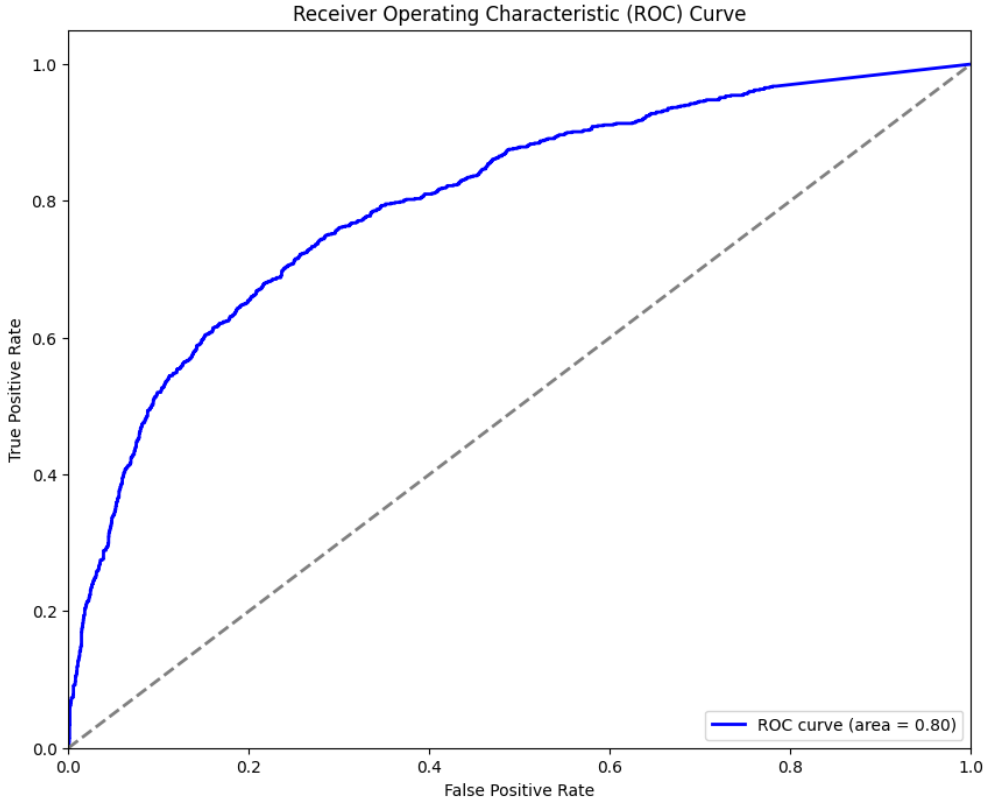
	precision	recall	f1-score	support
0	0.82	0.94	0.88	3029
1	0.61	0.31	0.41	900



AFTER TUNING

Accuracy: 0.8055484856197506
Classification Report:

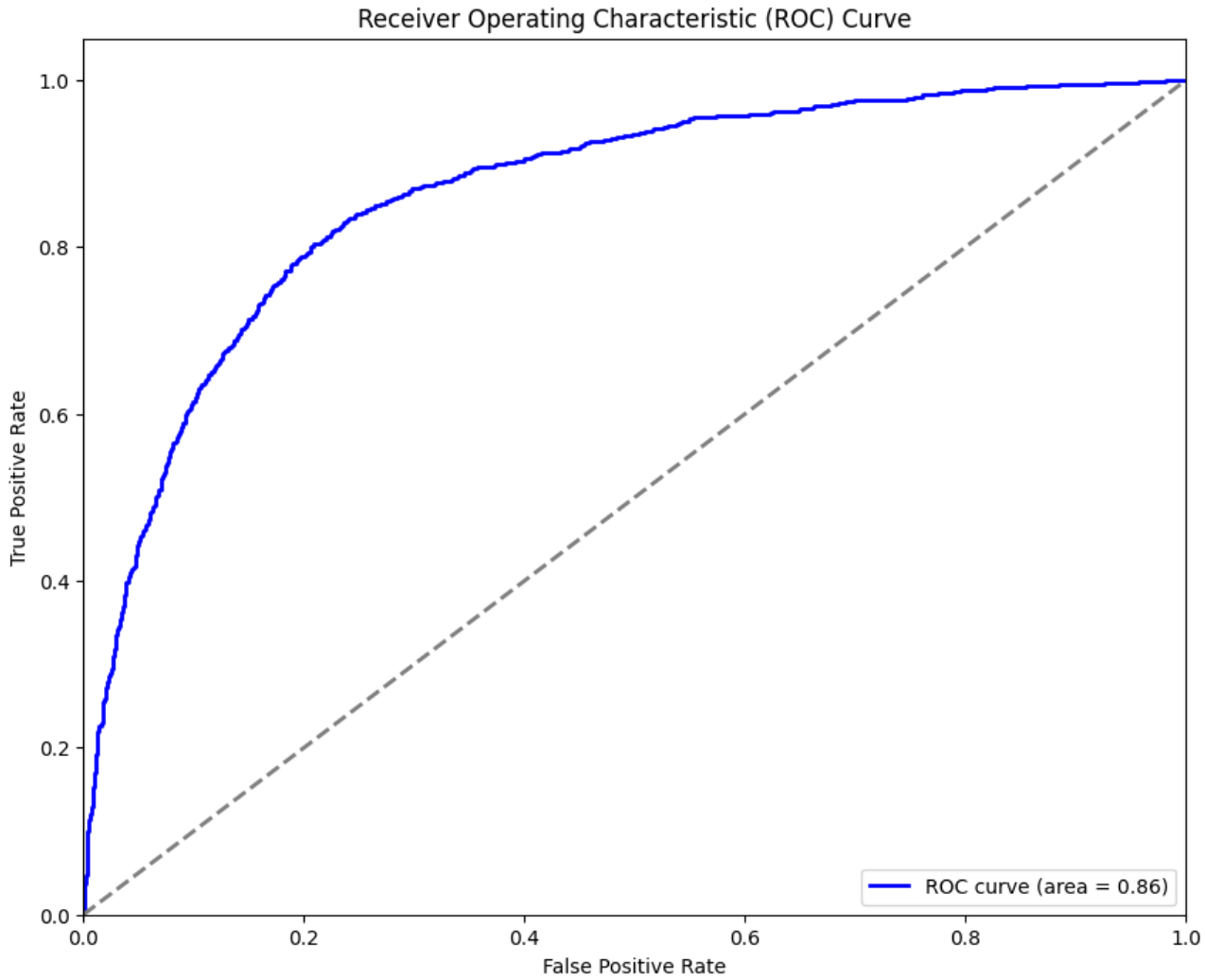
	precision	recall	f1-score	support
0	0.82	0.96	0.88	3029
1	0.67	0.29	0.41	900



XGBOOST

Accuracy: 0.8360906082972767
Classification Report:

	precision	recall	f1-score	support
0	0.88	0.91	0.90	3029
1	0.66	0.57	0.62	900



SUMMARY

- **doctor_recc_h1n1, seasonal_vaccine, opinion_h1n1 risk and doctor_recc_seasonal were the most important features for predicting h1n1 vaccine use.**

- **Best model performance was observed after XGboost**