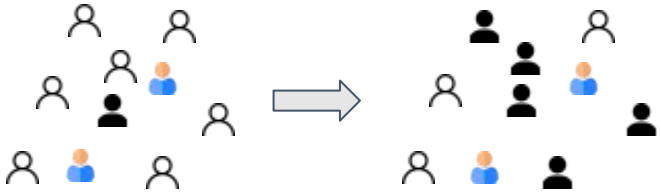
Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines

Anubhav Pareek(Ridge)

Project-3 Presentation
10th Feb

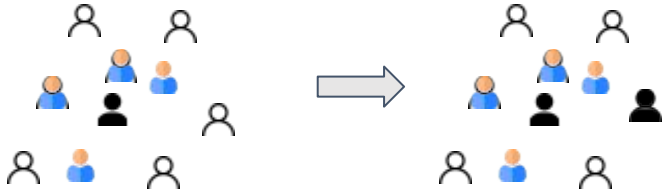
Business Case

Few people are Vaccinated



Then **disease** spreads very fast

Lots of people are Vaccinated



We can reach Herd Immunity

Herd Immunity = Logistics + Personal Choice

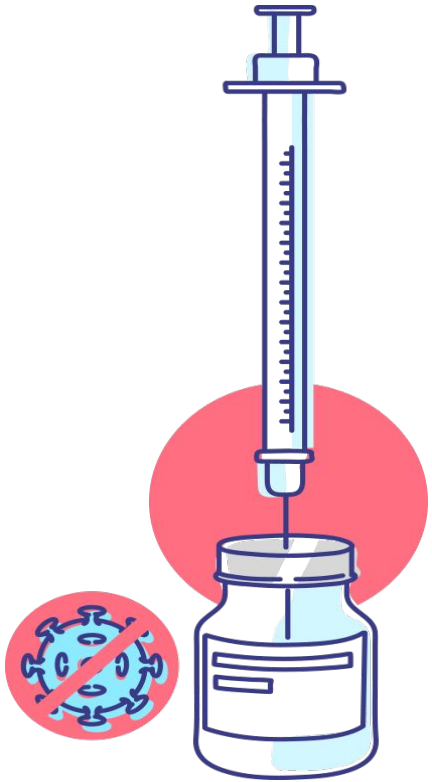
Business Case ⇨ Problem Statement

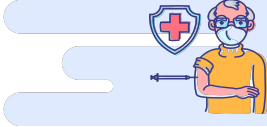
"You'll have a choice of whether you take the vaccine or not"

\- Bill Gates

So what determines if a person will take the vaccine or not ?

Problem Statement ⇨ Data





1 Source

The National 2009 H1N1 Flu Survey

2 Provider

Center for Disease Control and Prevention

3 Data Definition

Total Records - 26K
Features - 36

4 Target

H1N1 & Seasonal Flu - Yes/No

Data | Feature Overview

36 Features in Data Set

Background

- Age group
- Race
- Education

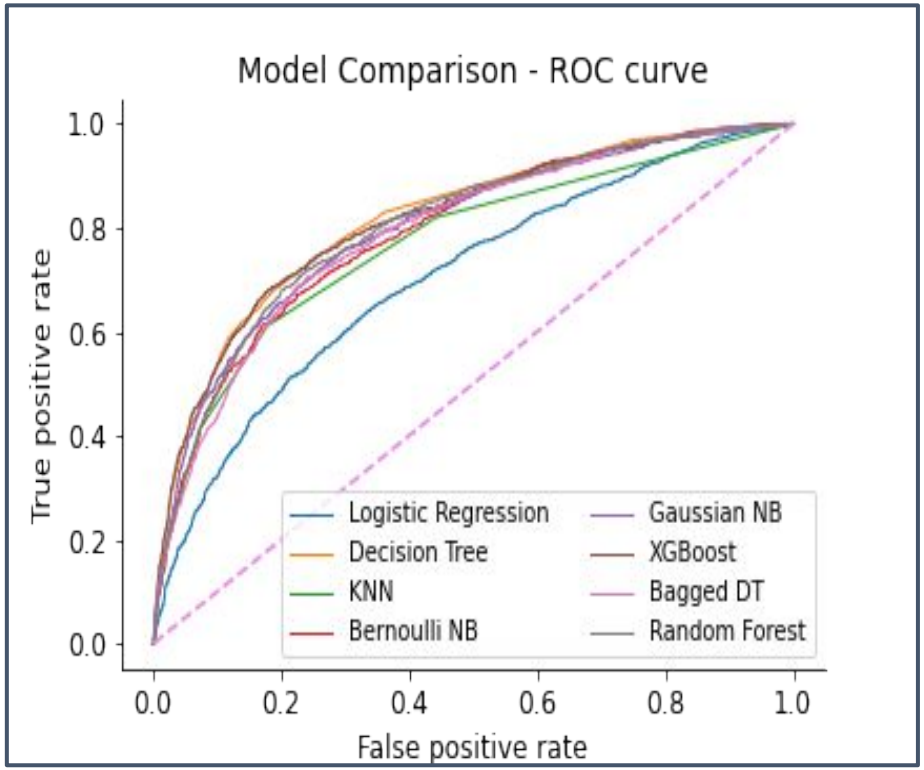
Opinion

- Effectiveness
- Risk
- Confidence

Health Behaviour

- Wash Hands
- Gathering
- Face Mask

Data ⇨ Model


Model Comparison - ROC curve
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression
Decision Tree
KNN
Bernoulli NB
Gaussian NB
XGBoost
Bagged DT
Random Forest

Model | Overview

Model : Logistic Regression

Features Category :

- Background
- Opinion
- Behavioural

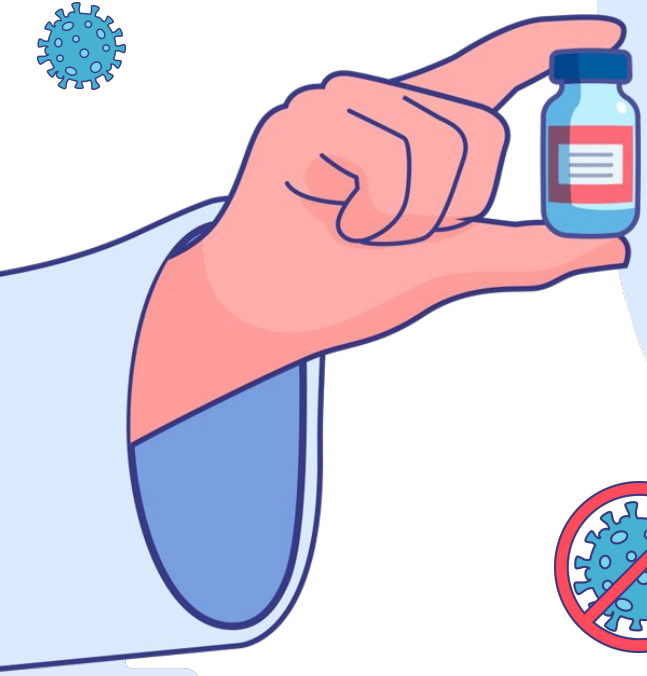
Metric : f_β score

Results Similar for H1N1 and Seasonal Flu

Results

Baseline Likelihood for **H1N1** Vaccination

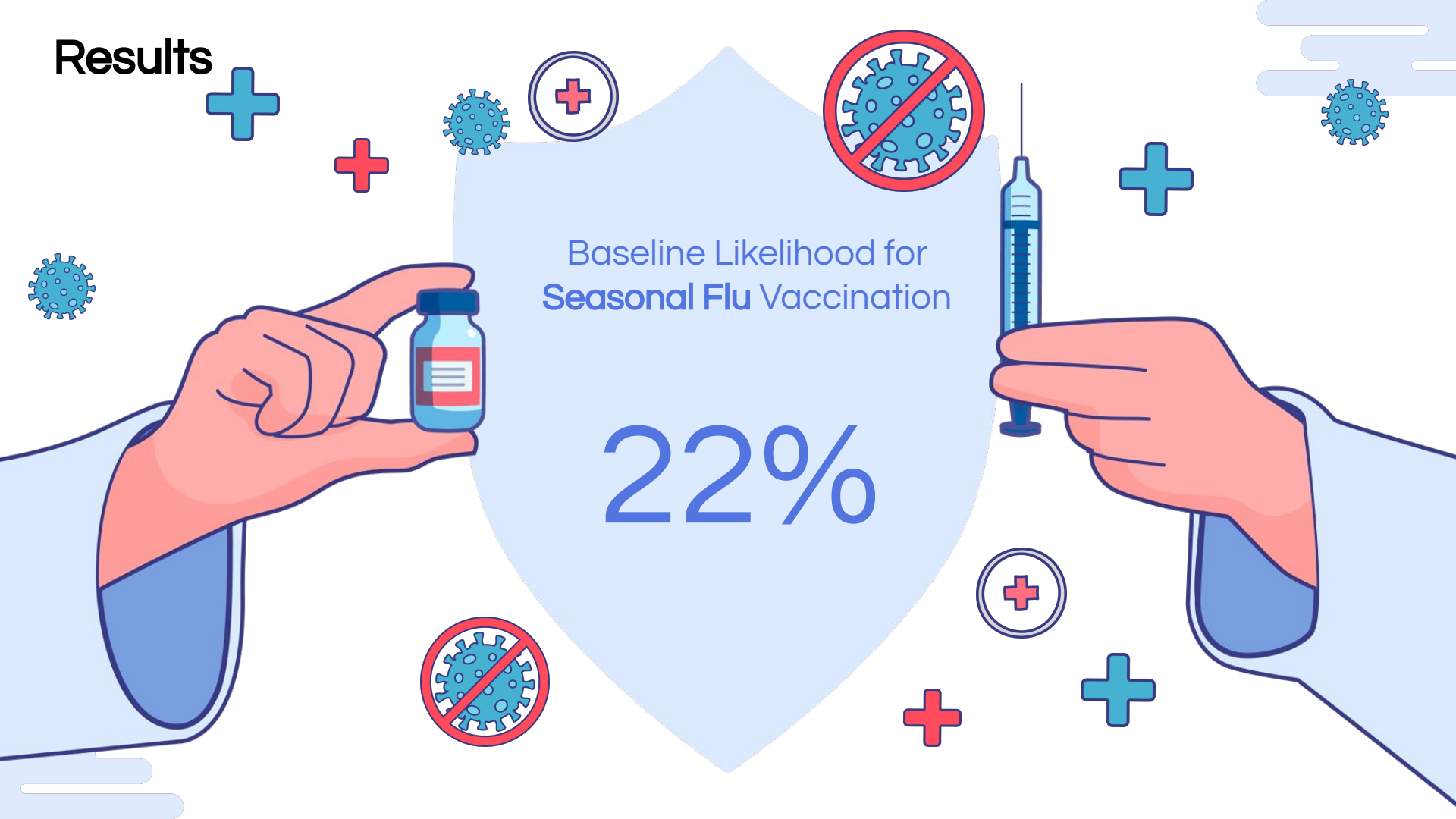
10%



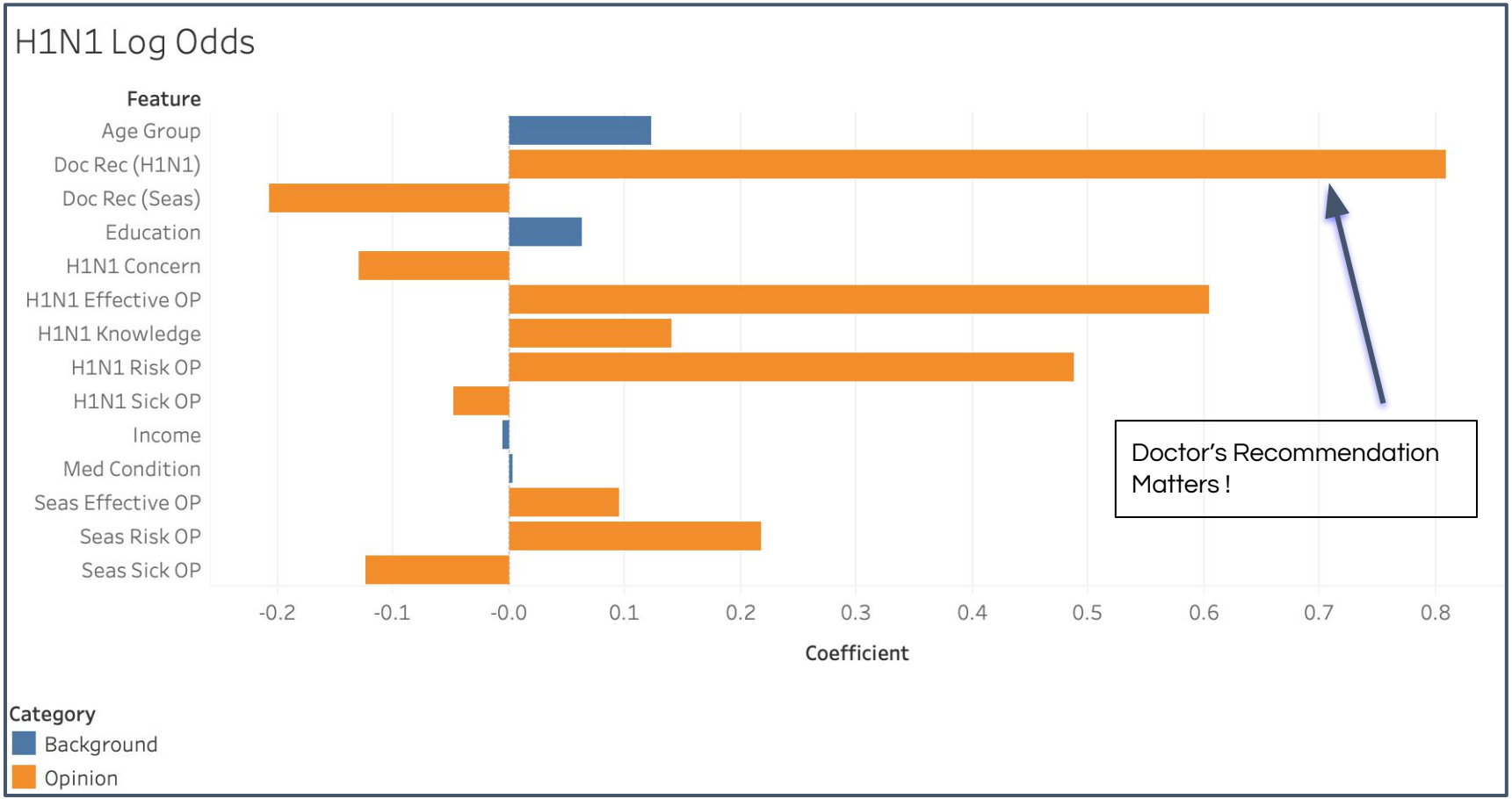
Results

Baseline Likelihood for
Seasonal Flu Vaccination

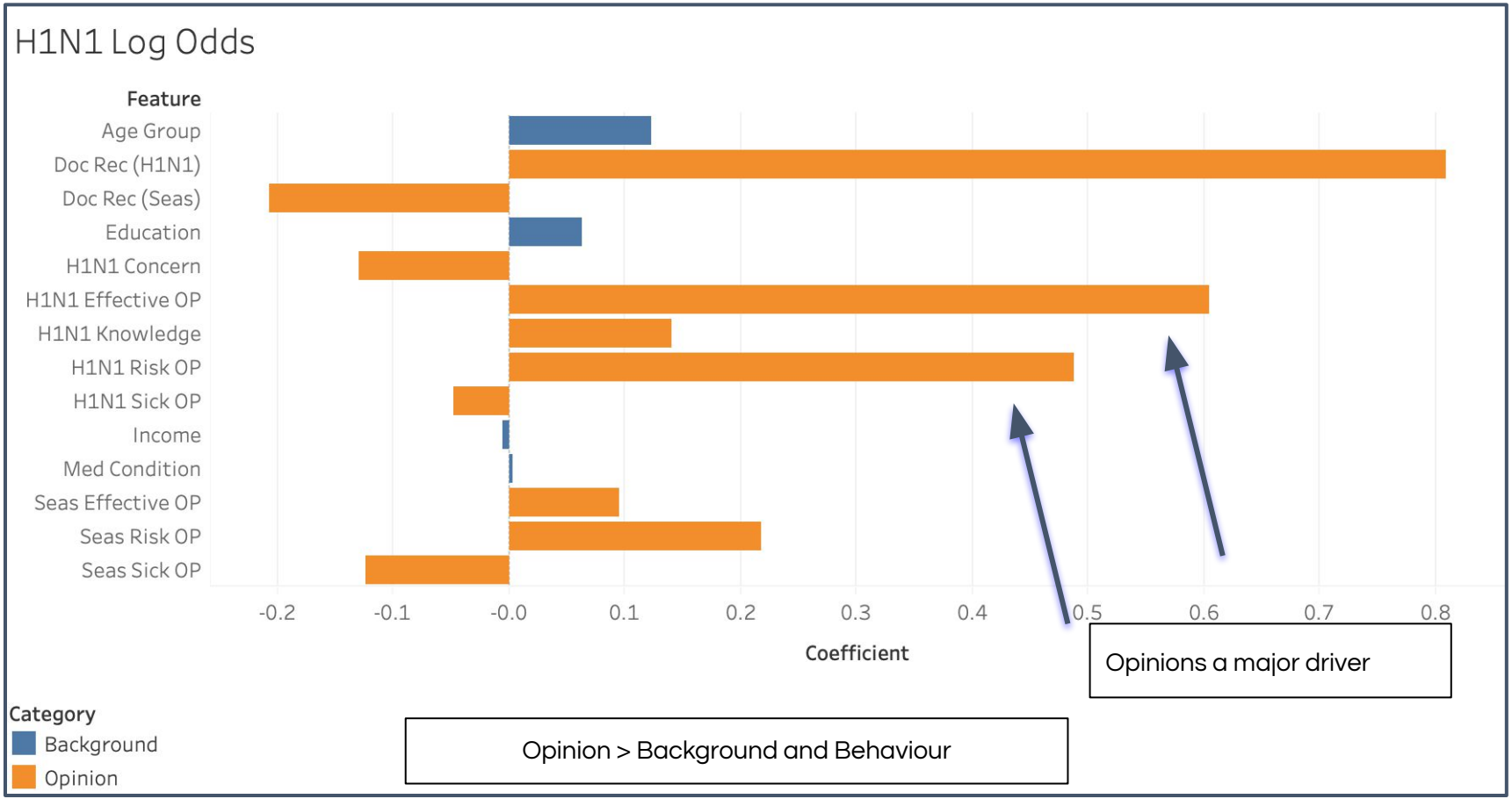
22%



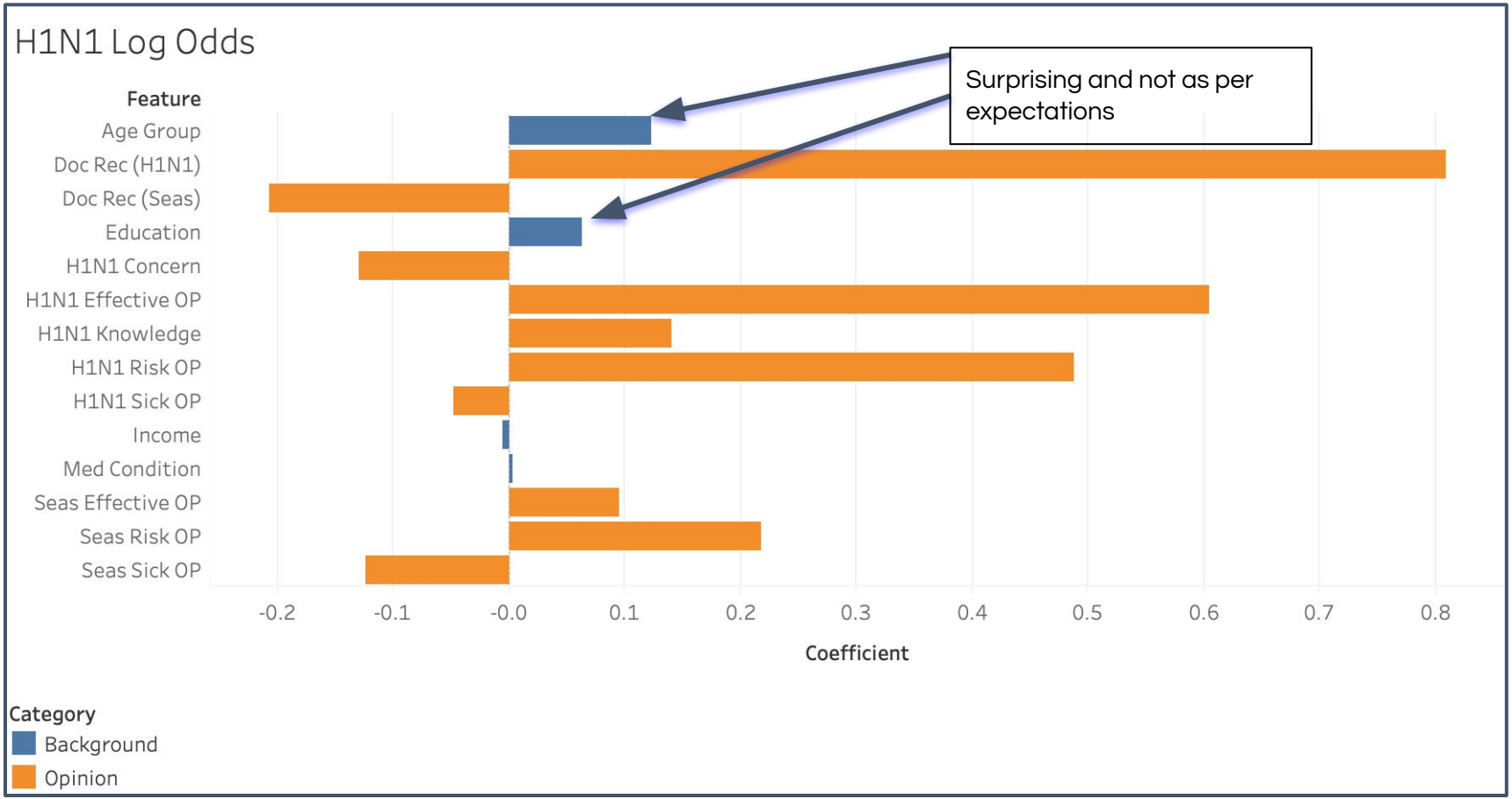
Results


H1N1 Log Odds
Feature
Age Group
Doc Rec (H1N1)
Doc Rec (Seas)
Education
H1N1 Concern
H1N1 Effective OP
H1N1 Knowledge
H1N1 Risk OP
H1N1 Sick OP
Income
Med Condition
Seas Effective OP
Seas Risk OP
Seas Sick OP
-0.2
-0.1
-0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Coefficient
Doctor's Recommendation Matters !
Category
Background
Opinion

Results


H1N1 Log Odds
Feature
Age Group
Doc Rec (H1N1)
Doc Rec (Seas)
Education
H1N1 Concern
H1N1 Effective OP
H1N1 Knowledge
H1N1 Risk OP
H1N1 Sick OP
Income
Med Condition
Seas Effective OP
Seas Risk OP
Seas Sick OP
-0.2
-0.1
-0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Coefficient
Category
Background
Opinion
Opinions a major driver
Opinion > Background and Behaviour

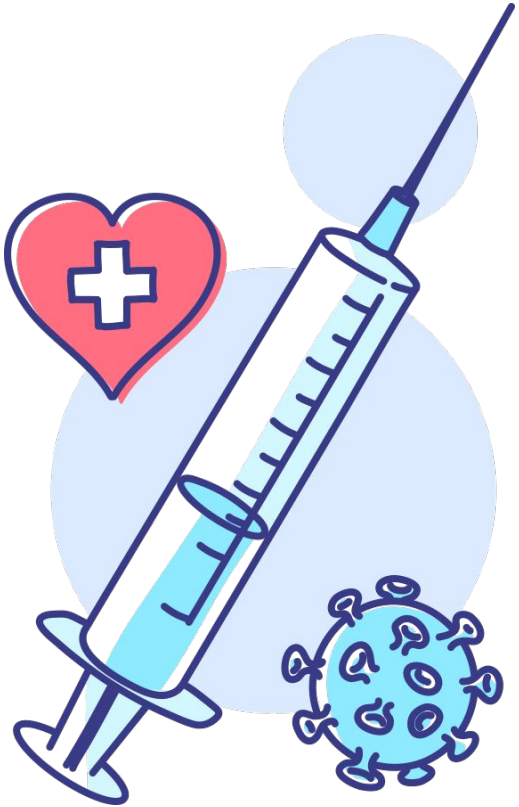
Results


H1N1 Log Odds
Feature
Age Group
Doc Rec (H1N1)
Doc Rec (Seas)
Education
H1N1 Concern
H1N1 Effective OP
H1N1 Knowledge
H1N1 Risk OP
H1N1 Sick OP
Income
Med Condition
Seas Effective OP
Seas Risk OP
Seas Sick OP
-0.2
-0.1
-0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Coefficient
Surprising and not as per expectations
Category
Background
Opinion

Results

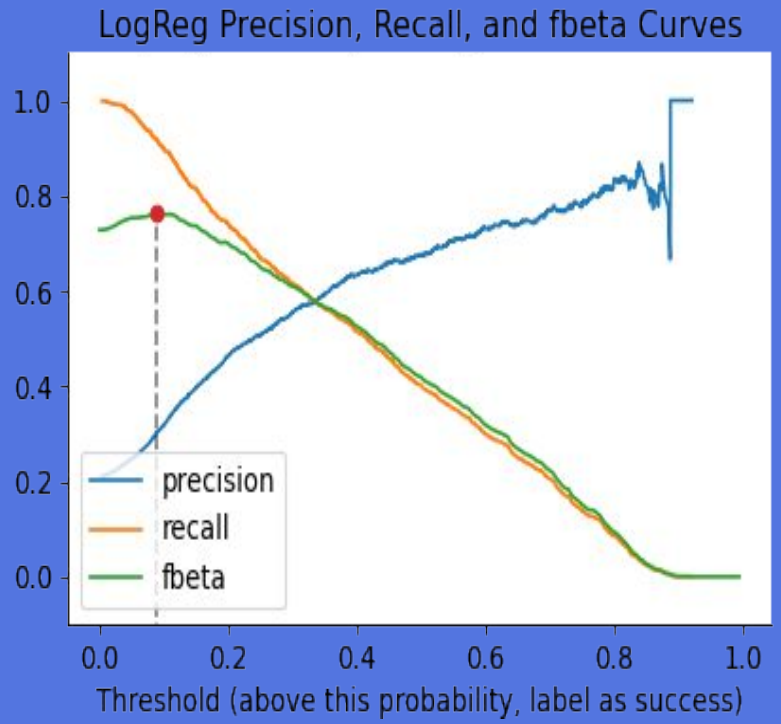
$$f_{\beta^{(H1N1)}} = 0.76$$

$$f_{\beta^{(Flu)}} = 0.84$$

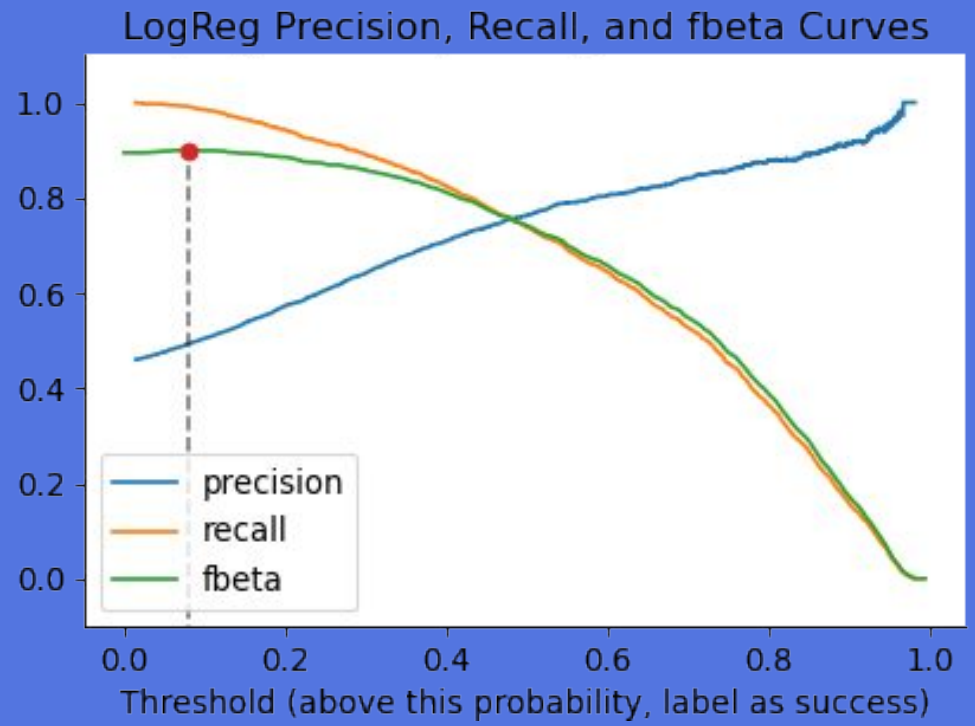


THANK YOU

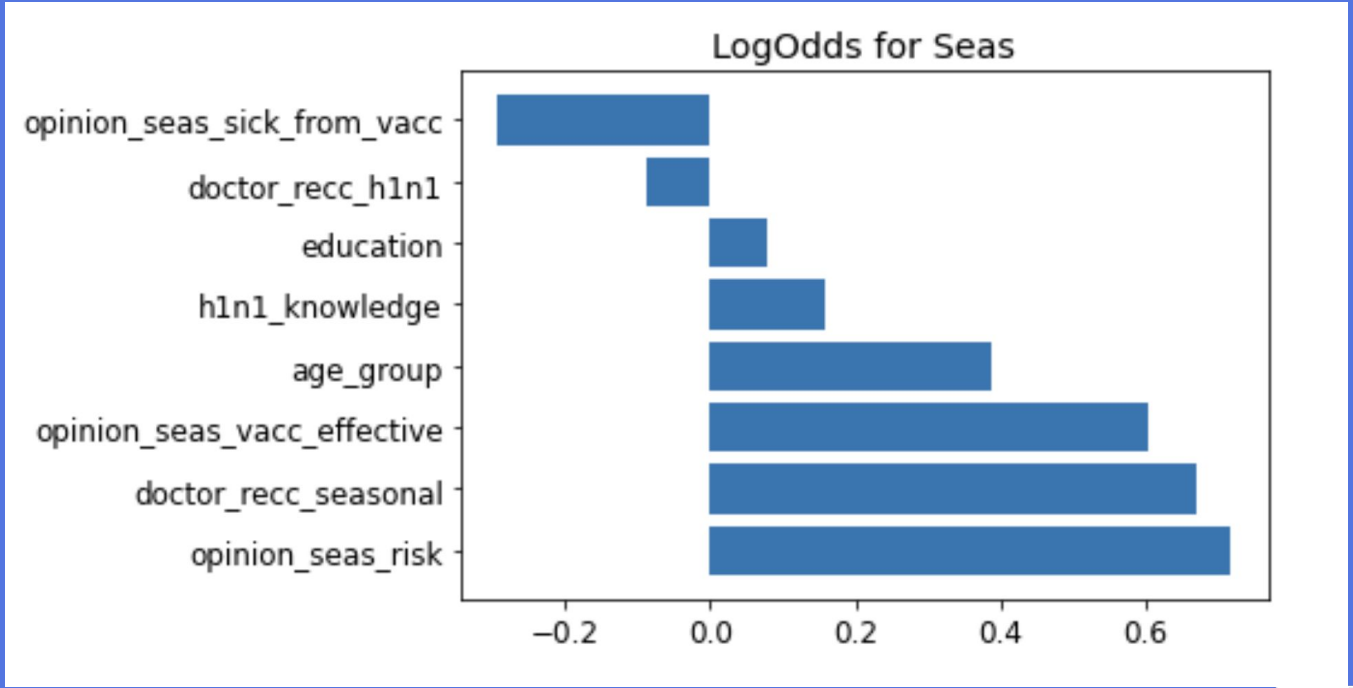
Precision Recall Curve - H1N1



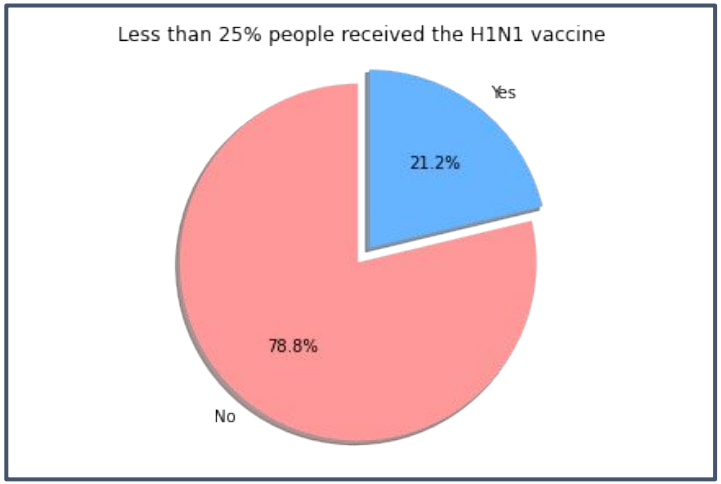
Precision Recall Curve - Seasonal Flu

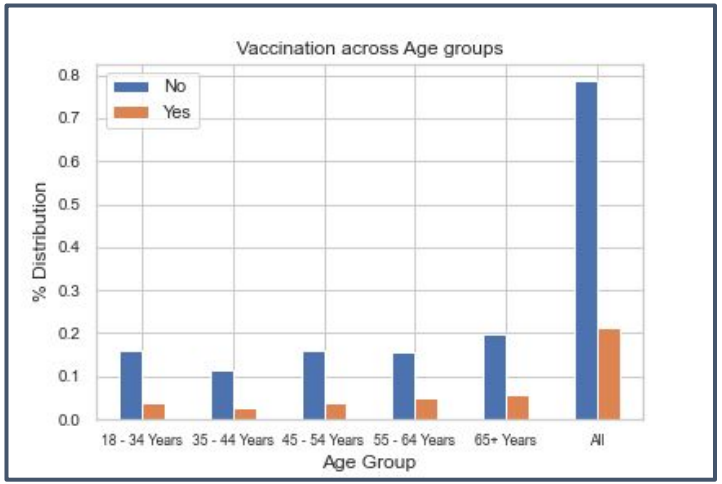


Log Odds - Seasonal Flu



Exploratory Data Analysis - H1N1 Data

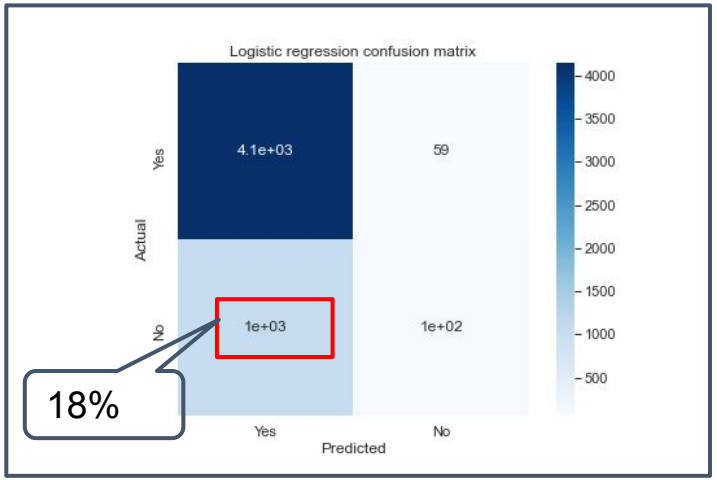


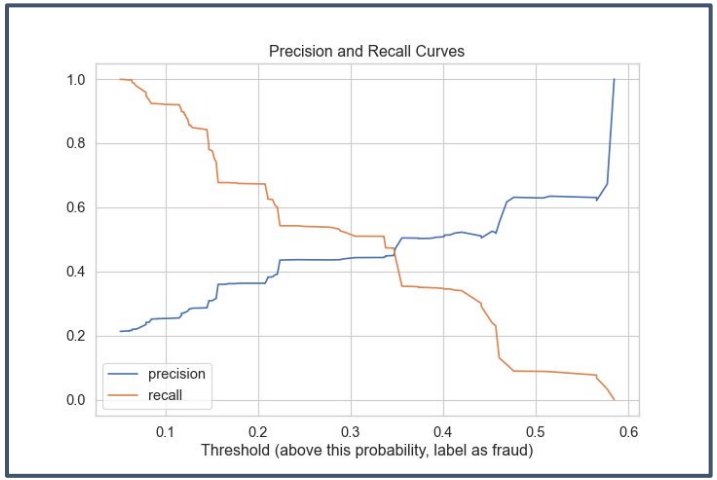


Key Observations:

- **50% Population -** Not vaccinated for H1N1 and Seasonal Flu
- Vaccination trends **consistent** across age groups
- **High Class Imbalance**

Model





Key Observations:

- **Logistic Regression** Model
- **Test Set** ~ 5000 records
- **4** Features used for training
- **F1_score :** 0.15586 (0.5 Threshold)
- **AUC :** 0.7312