CORONAVIRUS AND MENTAL HEALTH

How are you, really?

CONTENT

INTRODUCTION

In addition to the economical and social effects of the pandemic on the livelihoods of people, it has also brought to light the implications of such an unprecedented pandemic on the mental health of people.

Large scale disasters (SARS,9/11,hurricanes)) were almost always accompanied with increase in mental health and behavioural disorders

Together with past knowledge and existing data, what can we do to prevent such occurrences?

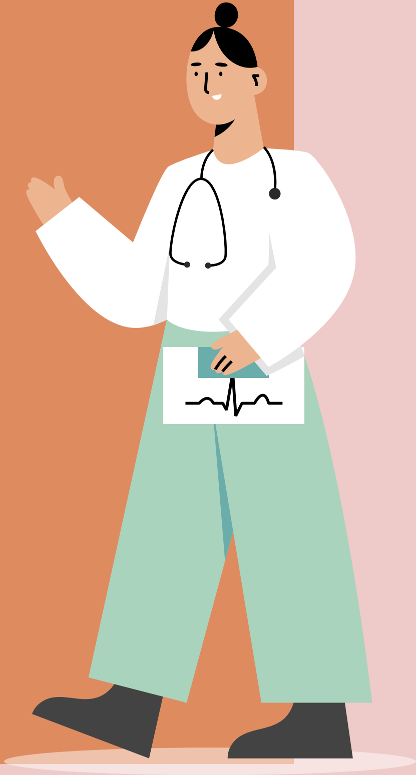
PROBLEM STATEMENT

Goal:

Aim to identify characteristics of respondents prone to developing avoidance behaviours in receiving mental health aid and build a binary classification model to predict likelihood of this tendency based on data from **Household Pulse Survey**. Models are evaluated using **ROC-AUC** and **recall** scores

Stakeholders:

Mental Health America, Local healthcare departments



METHODOLOGY

1. 2. 3. 4

Business Goal

- Develop a binary model to **identify vulnerable respondents** that exemplifies resistance to receiving mental health aid

Data Cleaning

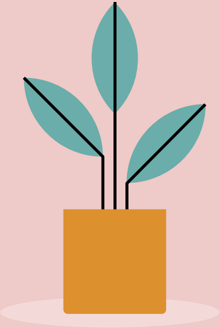
- Presence of imbalanced classes in target variable
- Presence of null values
- Presence of correlated predictor variables

Model Pre-processing

- Feature selection
- Feature engineering (creating new feature, PCA)
- Remove correlated categorical features (Pearson’s chi squared test)

Data Modeling

- Experimenting with different models
- Logreg, Xgboost, k-NN, RandomForest, Neural Nets)



HOUSEHOLD PULSE SURVEY

Time period: 19 Aug - 19 Sep)
Rows: 219,070
Columns: 188

DATA ANALYSIS

	T_BIRTHYEAR	INSURED	WKRLOSS	FOODSUF	MH_NOTGET
1	1989	Yes	Yes	Somewhat confident	No
2	1988	Yes	No	Very confident	No
3	1969	No	Yes	Not at all confident	Yes
4	1947	Yes	No	-	-

DATA CLEANING

IMPUTATION

- Impute missing values with averages
- Remove missing values

FEATURE ENGINEERING

- Creation of new feature ('HOUSEPAY') to reduce dependence between predictor variables
- PCA for dimensionality reduction

DROPPING REDUNDANT FEATURES

- Remove secondary features (travel plans, accessibility to internet etc)

FEATURE SELECTION

- Pearson's Chi Squared test to analyze correlation between categorical variables and target variable
- Reduce unrelated/redundant features(p-value > 0.05)

MODEL PREPARATION

TRAIN -TEST SPLIT

Splitting final data into train and test sets

STANDARD SCALING

Scaling X_train and X_test for algorithms that use distance metrics(k-NN) and gradient descent(Neural Nets)

UNDER - SAMPLING

Undersampling of majority classes in train set using RandomUnderSampler to balance dataset

Bimodel strategy

MODEL OPTIMIZATION

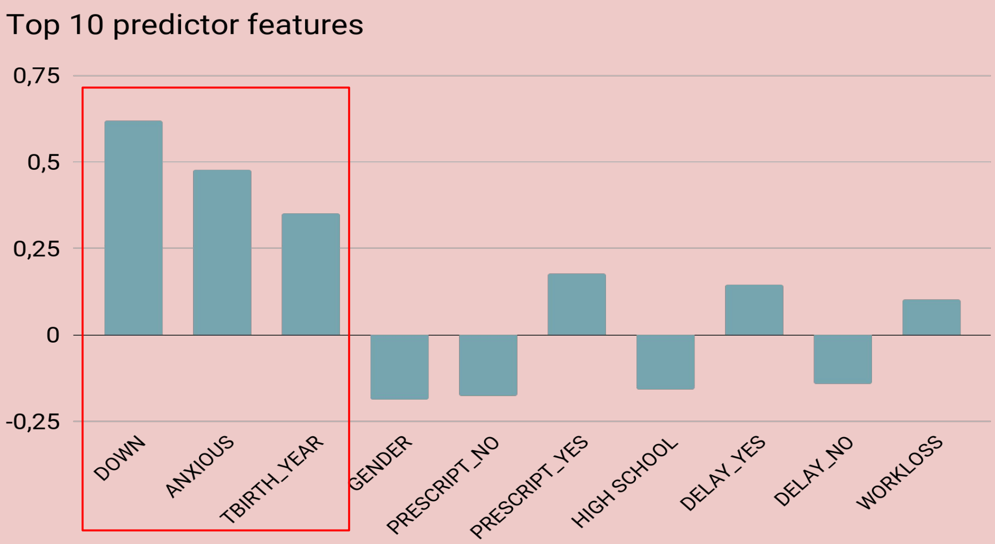
Logistic Regression

- Inference: Identify and quantify feature importance
- Hyperparameters: C:0.001, penalty: L2
- Recall: 0.802
  ROC-AUC: 0.790
- Pros: Rank feature importance, fast
- Cons: Assume linear relationship, lower recall score

ExtraTrees Classifier

- Prediction: Generate accurate predictions
- Hyperparameters: class_weight: balanced,max_depth: 40, max_features: auto, min_samples_leaf: 40
- Recall: 0.856
  ROC-AUC: 0.770
- Pros: Good balance of bias-variance trade off
- Cons: Slow implementation

Logistic Regression



FEATURE IMPORTANCE

Logistic Regression

1

DOWN :
How often do you feel sad/hopeless?
1 - Not at all, 4 - Everyday

2

ANXIOUS :
How often do you feel anxious?
1 - Not at all, 4 - Everyday

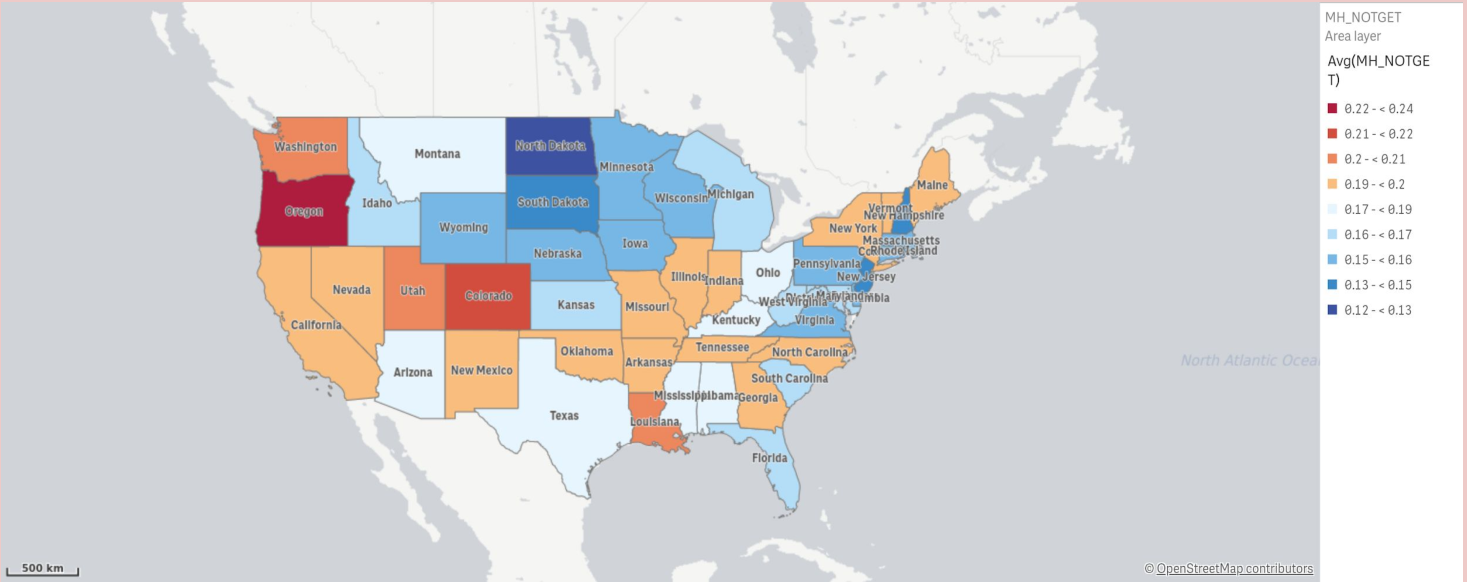
3

AGE :
How old are you?

FEATURE IMPORTANCE

Distribution of avoidance behaviour

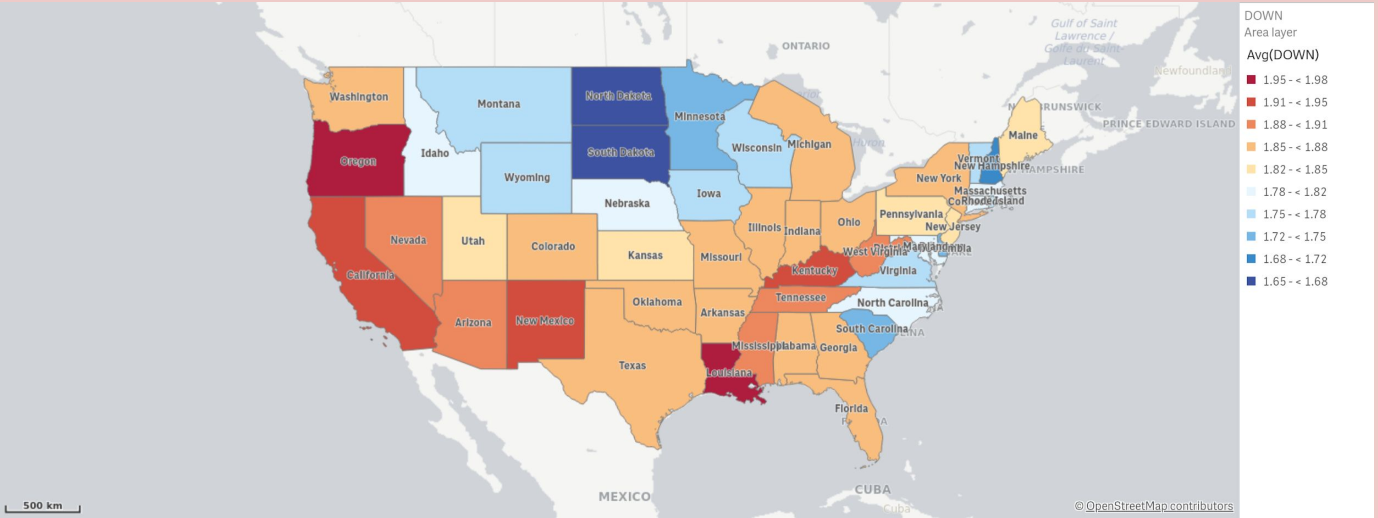
- High percentages: Oregon, Washington
- Low percentages: North and South Dakota
- Oregon and Colorado ranked **48th** and **47th** respectively in mental health



FINDINGS

1) Constantly feeling down

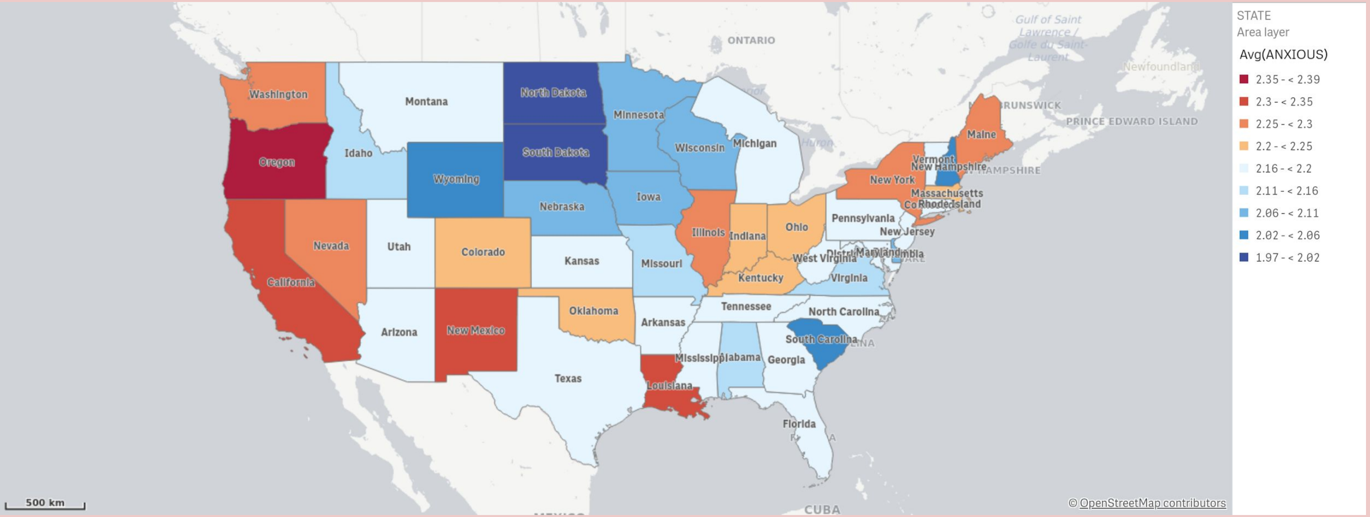
- Strong correlation between **high** frequency of feeling sad/hopeless and development of avoidance behaviours
- A **common symptom** of depression
- High anxiety levels: Oregon, Louisiana
- Low anxiety levels: North and South Dakota



FINDINGS

2) Constantly feeling anxious

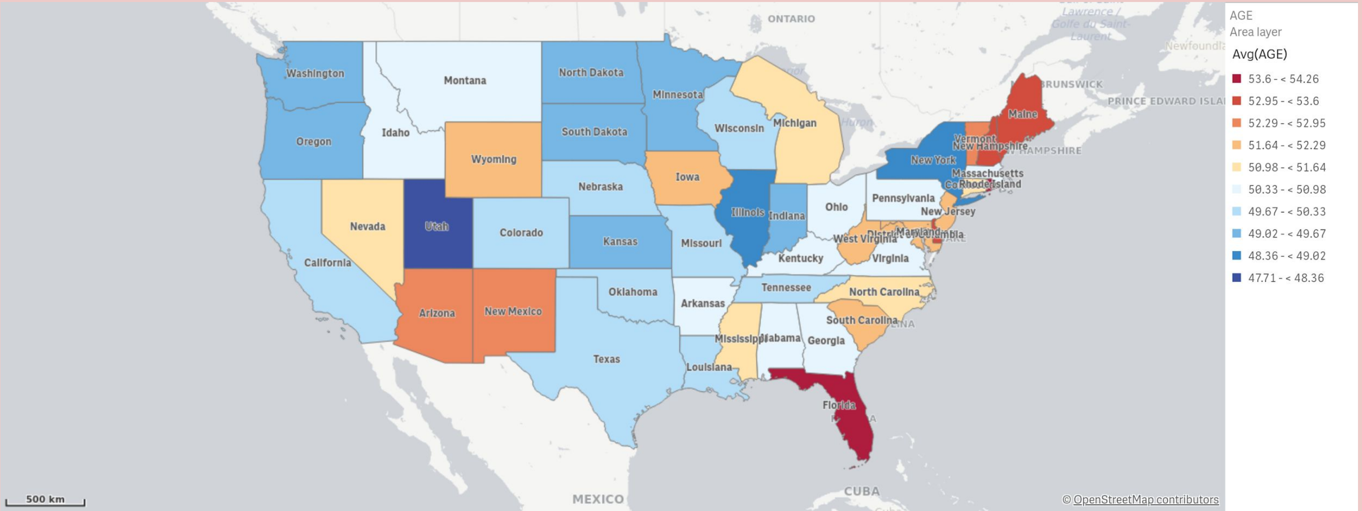
- Strong correlation between **high** frequency of feeling anxious and development of avoidance behaviour
- A **common symptom** of anxiety disorders
- High anxiety levels: Oregon, Washington
- Low anxiety levels: North and South Dakota



FINDINGS

3) Younger population

- Correlation between a younger population and development of avoidance behaviour
- Average age of respondents that do not exhibit such behaviours: **52.0**
- Average age of respondents that do exhibit such behaviours: **44.2**



FINDINGS

MODEL EVALUATION

	Logistic Regression	XgBoost	Random Forest	k-NN	Neural Nets	Extra Trees
CV Recall	0.776	0.921	0.824	0.733	0.702	0.840
Train Recall	0.777	0.999	0.861	0.744	0.843	0.870
Test Recall	0.799	0.940	0.839	0.750	0.836	0.856
Train ROC-AUC	0.775	0.881	0.807	0.769	0.799	0.797
Test ROC-AUC	0.787	0.743	0.771	0.771	0.790	0.770

Extra Trees Classifier

CONFUSION MATRIX

	PREDICT NEGATIVE	PREDICT POSITIVE
ACTUAL NEGATIVE	13816	6402
ACTUAL POSITIVE	647	3711

MODEL PERFORMANCE

Re-evaluating false positives

Criteria for False positives?
DOWN >= 3 , ANXIOUS >= 3

	False positives	False positives?
Total	6402	2117

	Age of actual positives	Age of false positives?
Average	44.2	46.2

FALSE POSITIVES

LIMITATIONS

1

Bias in responses:
Response bias and self-reported assessment of mental health status is highly subjective.

2

Imbalanced classes:
Undersampling was done randomly to reduce the number of majority class which would disregard potentially important features

3

Unrepresentative data:
Insufficient data on minority class and thus model could be overfitted with this particular class. Groups of people such as those without internet access/people who are institutionalized are excluded

CONCLUSION

Implement model on a smaller scale and as reference

Critical features identified (DOWN,AXNIOUS,age) would be the deciding factors on where to implement models. Models viable as references, not indicative of actual mental health disorders

Implement model in dire state - Oregon

Propose models to Oregon Health Authority are Oregon marked the checkboxes for high ANXIOUS, high DOWN and a younger population

Implement model in states that show similar trend - Utah, Illinois

Medium to high levels of DOWN and ANXIOUS with a growing population

- Utah has a resident: behavioural professionals ratio below national average
- Illinois has per capita expenditure on health services below that of national average

THANK YOU

ANY QUESTIONS?