Predicting Depression

Ting Ting | Data Science Part-Time Capstone Project

Problem Statement

According to the World Health Organization, "**Depression is one of the leading causes of disability. Suicide is the second leading cause of death among 15-29-year-olds.** People with severe mental health conditions die prematurely – as much as two decades early – due to preventable physical conditions.

Project Goal

This project seeks to use information from an **online assessment** which includes test scores on the **Depression, Anxiety and Stress Scale, Ten Personality Item Scale** and demographic information to **predict depression** in an individual and **explore possible interactions between personality, demographic and depression.**

Approach and Process

- Computing scores on the Depression, Anxiety and Stress Scale and Ten-Personality Item Scale
- Cleaning Data
- Exploratory Data Analysis
- Modelling and model evaluation
- Interpreting the results

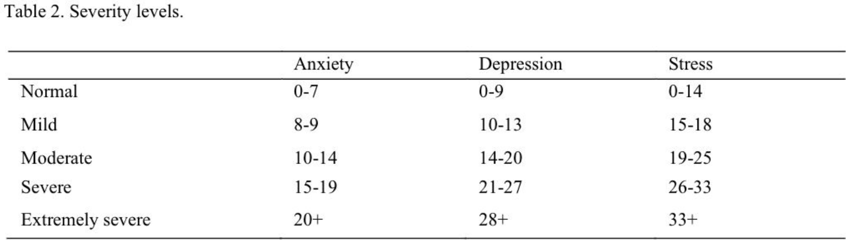


Table 2. Severity levels.

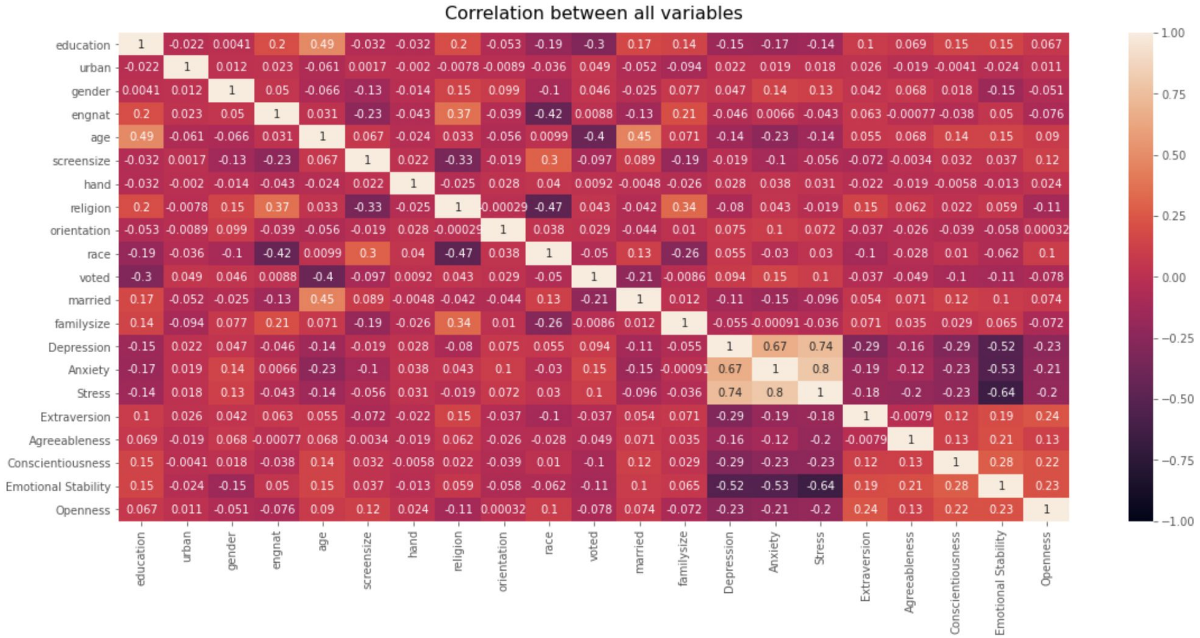
	Anxiety	Depression	Stress
Normal	0-7	0-9	0-14
Mild	8-9	10-13	15-18
Moderate	10-14	14-20	19-25
Severe	15-19	21-27	26-33
Extremely severe	20+	28+	33+

Cleaning Data

- Data source was from an **online assessment** with 3 different components
- Only included data where **test takers marked that the results can be used for research**
- Data set was relatively clean and required **minimal data cleaning** (removing outliers and null values were in features that are inconsequential to our target e.g. major)

Exploratory Data Analysis

- None of the independent variables saw particularly high correlation with another independent variable



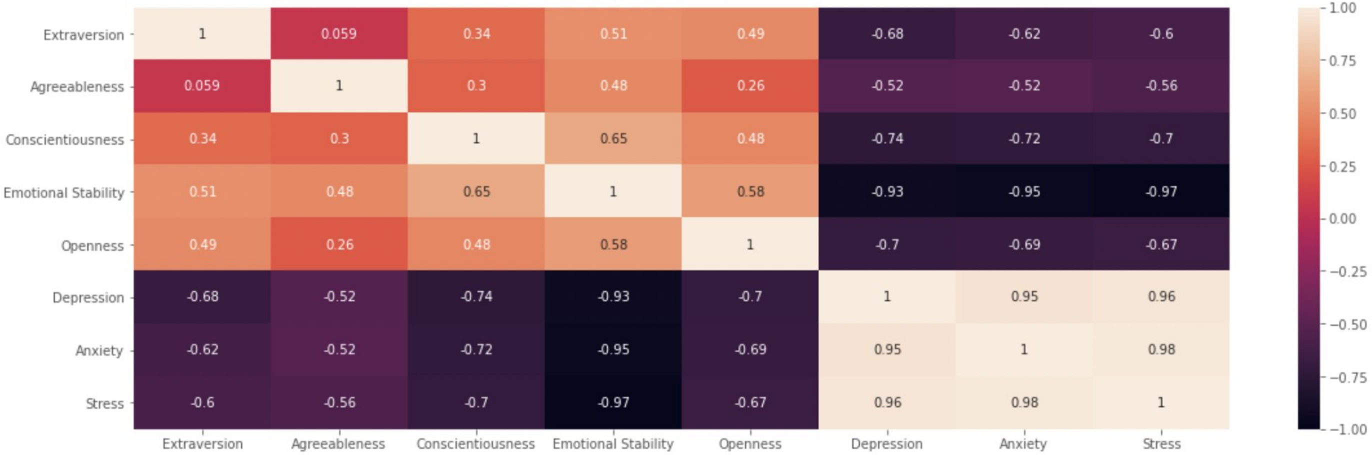
Correlation between all variables

	education	urban	gender	engnat	age	screensize	hand	religion	orientation	race	voted	married	familysize	Depression	Anxiety	Stress	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
education	1	-0.022	0.0041	0.2	0.49	-0.032	-0.032	0.2	-0.053	-0.19	-0.3	0.17	0.14	-0.15	-0.17	-0.14	0.1	0.069	0.15	0.15	0.067
urban	-0.022	1	0.012	0.023	-0.061	0.0017	-0.002	-0.0078	-0.0089	-0.036	0.049	-0.052	-0.094	0.022	0.019	0.018	0.026	-0.019	-0.0041	-0.024	0.011
gender	0.0041	0.012	1	0.05	-0.066	-0.13	-0.014	0.15	0.099	-0.1	0.046	-0.025	0.077	0.047	0.14	0.13	0.042	0.068	0.018	-0.15	-0.051
engnat	0.2	0.023	0.05	1	0.031	-0.23	-0.043	0.37	-0.039	-0.42	0.0088	-0.13	0.21	-0.046	0.0066	-0.043	0.063	-0.00077	-0.038	0.05	-0.076
age	0.49	-0.061	-0.066	0.031	1	0.067	-0.024	0.033	-0.056	0.0099	-0.4	0.45	0.071	-0.14	-0.23	-0.14	0.055	0.068	0.14	0.15	0.09
screensize	-0.032	0.0017	-0.13	-0.23	0.067	1	0.022	-0.33	-0.019	0.3	-0.097	0.089	-0.19	-0.019	-0.1	-0.056	-0.072	-0.0034	0.032	0.037	0.12
hand	-0.032	-0.002	-0.014	-0.043	-0.024	0.022	1	-0.025	0.028	0.04	0.0092	-0.0048	-0.026	0.028	0.038	0.031	-0.022	-0.019	-0.0058	-0.013	0.024
religion	0.2	-0.0078	0.15	0.37	0.033	-0.33	-0.025	1	-0.00029	-0.47	0.043	-0.042	0.34	-0.08	0.043	-0.019	0.15	0.062	0.022	0.059	-0.11
orientation	-0.053	-0.0089	0.099	-0.039	-0.056	-0.019	0.028	-0.00029	1	0.038	0.029	-0.044	0.01	0.075	0.1	0.072	-0.037	-0.026	-0.039	-0.058	0.00032
race	-0.19	-0.036	-0.1	-0.42	0.0099	0.3	0.04	-0.47	0.038	1	-0.05	0.13	-0.26	0.055	-0.03	0.03	-0.1	-0.028	0.01	-0.062	0.1
voted	-0.3	0.049	0.046	0.0088	-0.4	-0.097	0.0092	0.043	0.029	-0.05	1	-0.21	-0.0086	0.094	0.15	0.1	-0.037	-0.049	-0.1	-0.11	-0.078
married	0.17	-0.052	-0.025	-0.13	0.45	0.089	-0.0048	-0.042	-0.044	0.13	-0.21	1	0.012	-0.11	-0.15	-0.096	0.054	0.071	0.12	0.1	0.074
familysize	0.14	-0.094	0.077	0.21	0.071	-0.19	-0.026	0.34	0.01	-0.26	-0.0086	0.012	1	-0.055	-0.00091	-0.036	0.071	0.035	0.029	0.065	-0.072
Depression	-0.15	0.022	0.047	-0.046	-0.14	-0.019	0.028	-0.08	0.075	0.055	0.094	-0.11	-0.055	1	0.67	0.74	-0.29	-0.16	-0.29	-0.52	-0.23
Anxiety	-0.17	0.019	0.14	0.0066	-0.23	-0.1	0.038	0.043	0.1	-0.03	0.15	-0.15	-0.00091	0.67	1	0.8	-0.19	-0.12	-0.23	-0.53	-0.21
Stress	-0.14	0.018	0.13	-0.043	-0.14	-0.056	0.031	-0.019	0.072	0.03	0.1	-0.096	-0.036	0.74	0.8	1	-0.18	-0.2	-0.23	-0.64	-0.2
Extraversion	0.1	0.026	0.042	0.063	0.055	-0.072	-0.022	0.15	-0.037	-0.1	-0.037	0.054	0.071	-0.29	-0.19	-0.18	1	-0.0079	0.12	0.19	0.24
Agreeableness	0.069	-0.019	0.068	-0.00077	0.068	-0.0034	-0.019	0.062	-0.026	-0.028	-0.049	0.071	0.035	-0.16	-0.12	-0.2	-0.0079	1	0.13	0.21	0.13
Conscientiousness	0.15	-0.0041	0.018	-0.038	0.14	0.032	-0.0058	0.022	-0.039	0.01	-0.1	0.12	0.029	-0.29	-0.23	-0.23	0.12	0.13	1	0.28	0.22
Emotional Stability	0.15	-0.024	-0.15	0.05	0.15	0.037	-0.013	0.059	-0.058	-0.062	-0.11	0.1	0.065	-0.52	-0.53	-0.64	0.19	0.21	0.28	1	0.23
Openness	0.067	0.011	-0.051	-0.076	0.09	0.12	0.024	-0.11	0.00032	0.1	-0.078	0.074	-0.072	-0.23	-0.21	-0.2	0.24	0.13	0.22	0.23	1

Exploratory Data Analysis

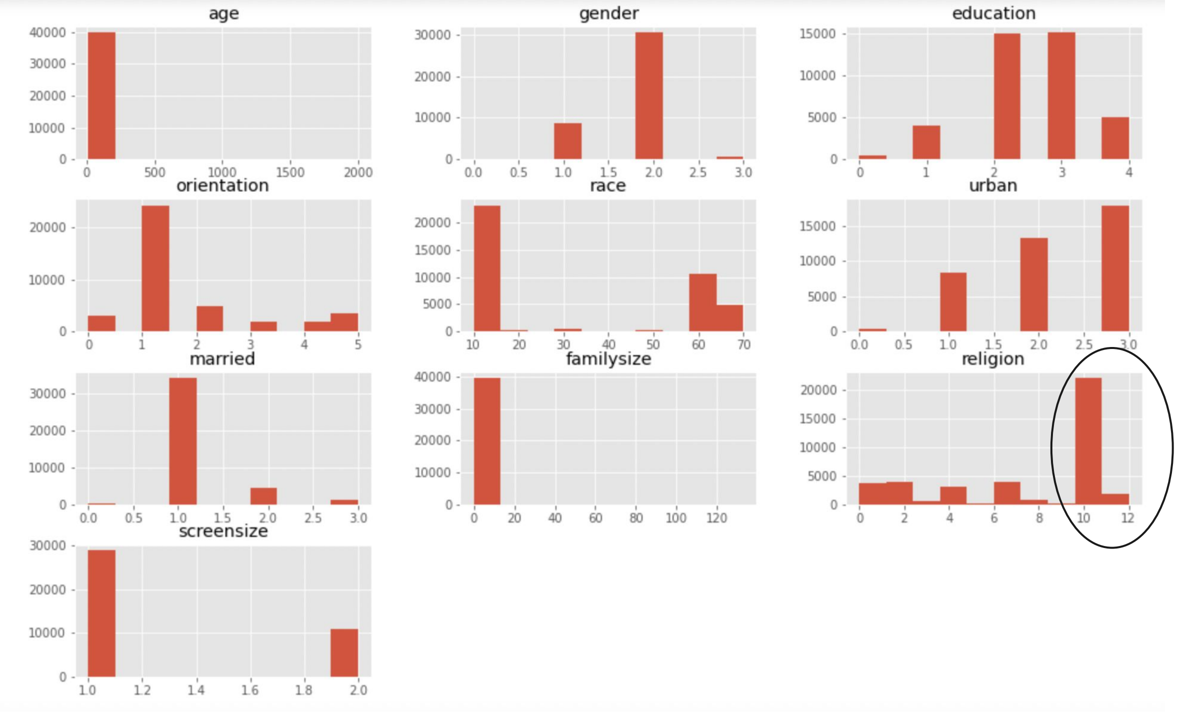
- Isolating personality traits and depression, anxiety and stress levels, anxiety and stress correlates strongly with depression and emotional stability negatively correlatives with all 3

Correlation between personality traits and DASS levels

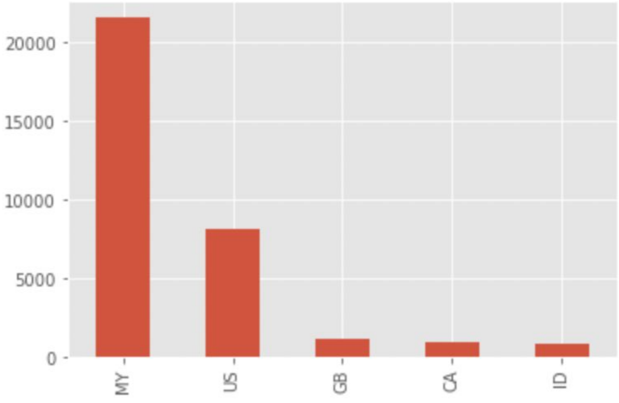


	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness	Depression	Anxiety	Stress
Extraversion	1	0.059	0.34	0.51	0.49	-0.68	-0.62	-0.6
Agreeableness	0.059	1	0.3	0.48	0.26	-0.52	-0.52	-0.56
Conscientiousness	0.34	0.3	1	0.65	0.48	-0.74	-0.72	-0.7
Emotional Stability	0.51	0.48	0.65	1	0.58	-0.93	-0.95	-0.97
Openness	0.49	0.26	0.48	0.58	1	-0.7	-0.69	-0.67
Depression	-0.68	-0.52	-0.74	-0.93	-0.7	1	0.95	0.96
Anxiety	-0.62	-0.52	-0.72	-0.95	-0.69	0.95	1	0.98
Stress	-0.6	-0.56	-0.7	-0.97	-0.67	0.96	0.98	1

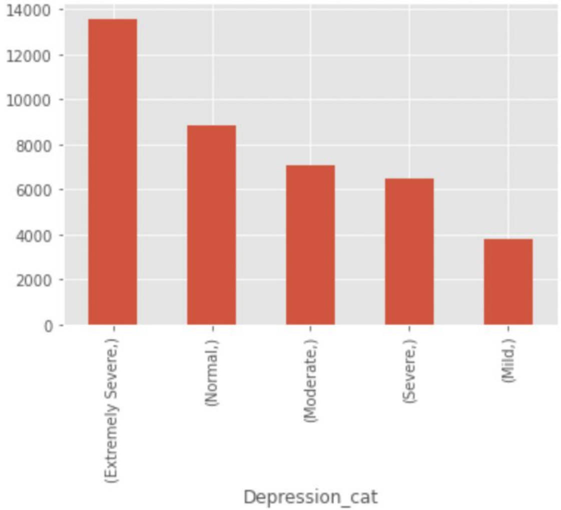
Who were the test takers?

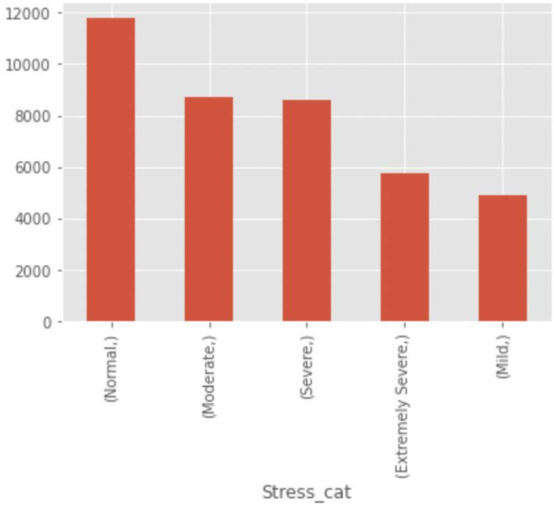


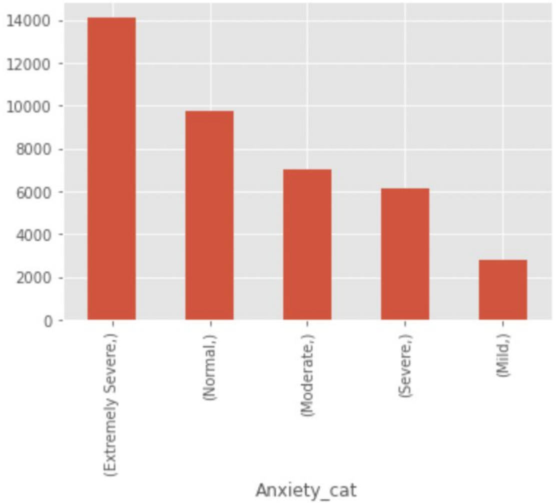
country



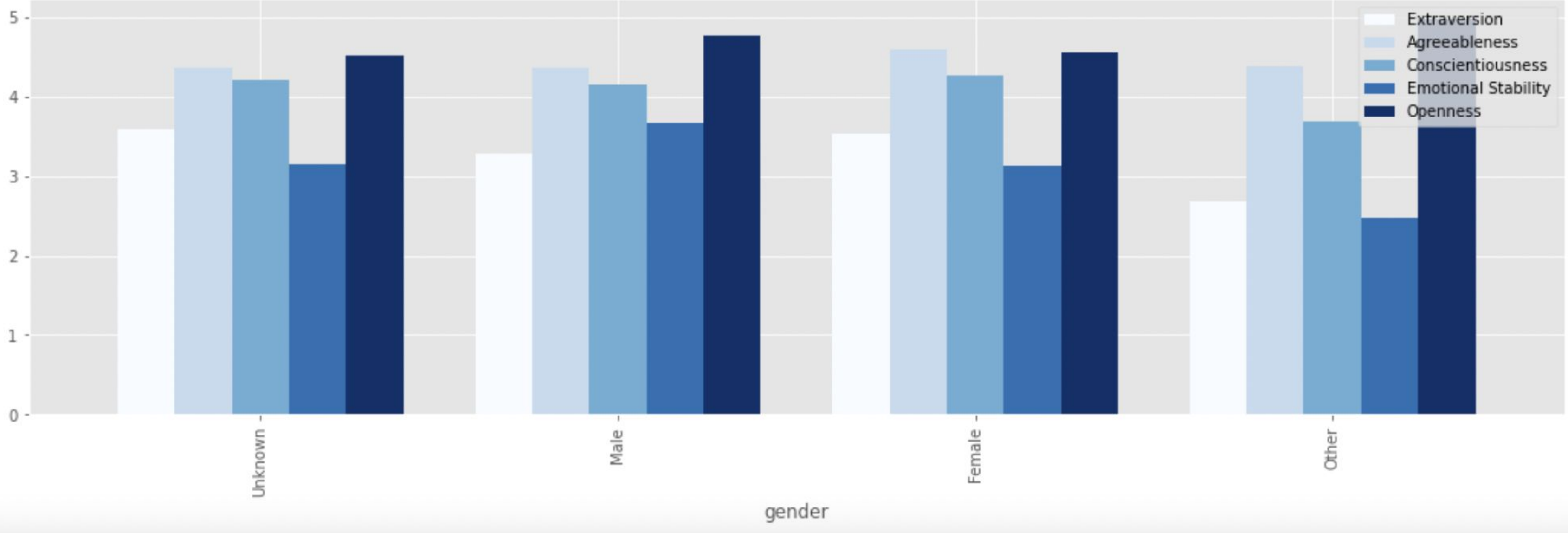
How participants scored on the DASS



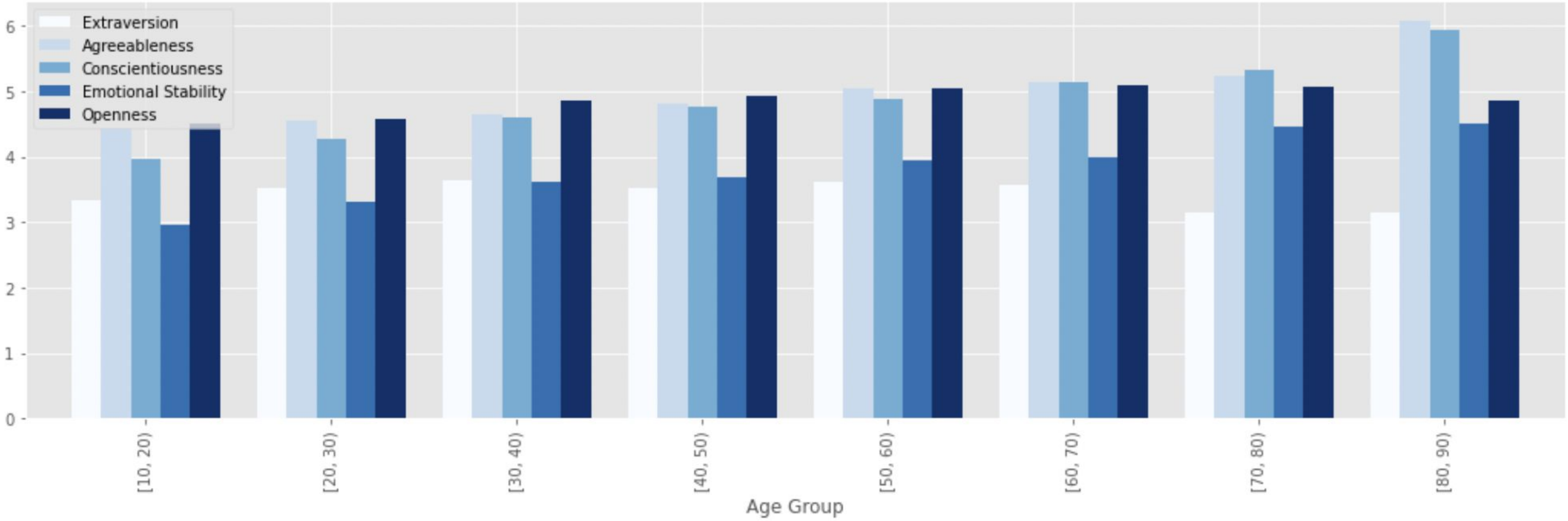




How participants scored on the Personality Scale by gender



How participants scored on the Personality Scale by age



Model and solution

List of models I applied before parameter hypertuning and their corresponding accuracy and f1 scores

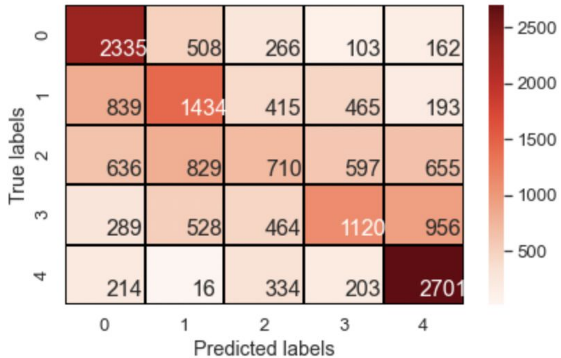
Model	Before SMOTE	After SMOTE
Random Forest Classifier	0.51, 0.35	0.67, 0.67
Support Vector Classification	0.52, 0.33	-
Logistic Regression	0.52, 0.33	0.49, 0.47
K Nearest Neighbors Classifier	0.45, 0.36	0.65, 0.64
XGBoost	0.51, 0.37	0.55, 0.54
Dumb Classifier	0.34 Zero Rate, 0.23 Random Weight Guessing	

Performance evaluation

- SVC scored similarly to other models but needed 35X more run time. Dropped after the first round of model evaluation
- Through randomized search and hyperparameter tuning, we found the best model (**Random Forest Classifier**) and **improved f1 score by 5% (from 0.67 to 0.7)**

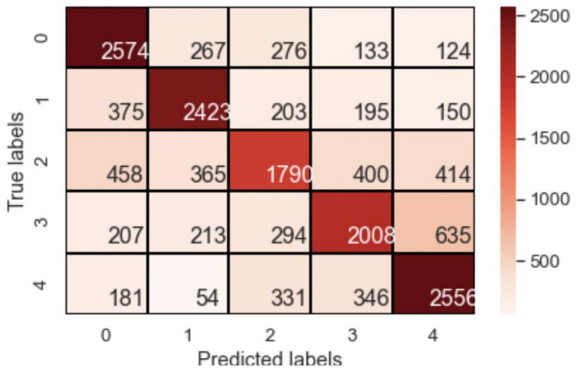
Confusion Matrix

LR



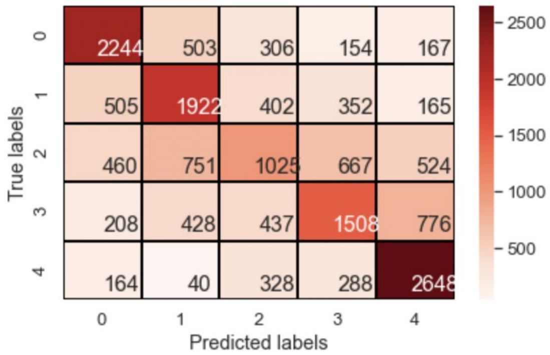
	0	1	2	3	4
0	2335	508	266	103	162
1	839	1434	415	465	193
2	636	829	710	597	655
3	289	528	464	1120	956
4	214	16	334	203	2701

RF



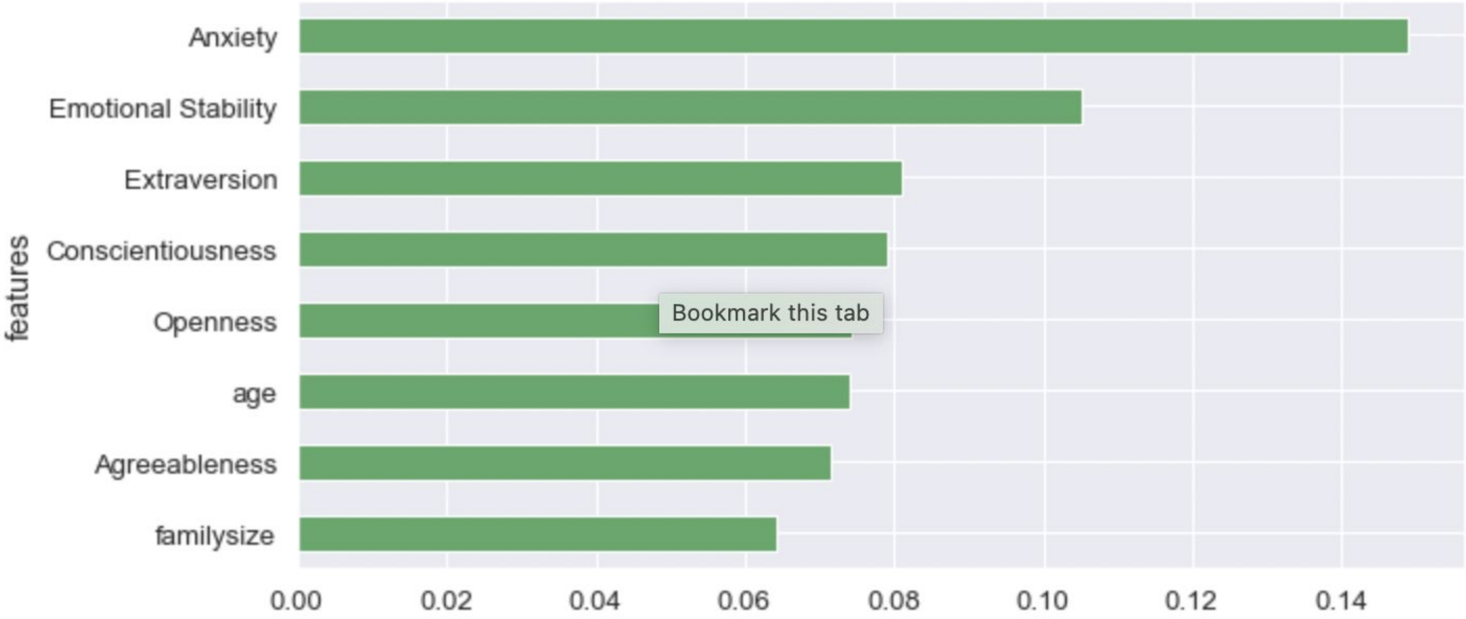
	0	1	2	3	4
0	2574	267	276	133	124
1	375	2423	203	195	150
2	458	365	1790	400	414
3	207	213	294	2008	635
4	181	54	331	346	2556

XGB



	0	1	2	3	4
0	2244	503	306	154	167
1	505	1922	402	352	165
2	460	751	1025	667	524
3	208	428	437	1508	776
4	164	40	328	288	2648

Feature Importance



Findings

- **Personality traits, particularly Emotional Stability**, are strong predictors of depression. They take precedence over test takers' background (e.g. gender, education received etc.)
- **Younger people on average scored lower on emotional stability**, males in general scored higher on emotional stability
- **Stress and anxiety levels had the largest effects on predicting depression**. They were also strongly positive correlated.

Limitations

- Data collected was skewed to test takers in one country, gender, age, marriage status and religion
- Classes were imbalanced: online assessments can be biased due to self-selection

Recommendations

- Collecting more data and model improvement: **diversify dataset** to include test takers in other countries, genders and age so that the model will be more representative of the general population
- Address class imbalance by administering test to a target group of people through random sampling or representative sampling instead