

Depression Tendency Prediction

Presented by Duy Khanh

Abstract

The goal of this project is to predict characteristics that influence an individual's mental health utilizing classifier models such as logit regression and random forests. Data is gathered on a group of people with a wide range of living situations, health, and socioeconomic status.

Presented by Duy Khanh

Agenda Overview

01 Problem Statement

02 Data Understanding
and Preparation

03 EDA and Visualization

04 Predicting with Classification
Model and Evaluation

05 Actionable
Recommendation

06 Conslusion

Presented by Juliana Silva



Problem

01

What are the significant life, economic, and social factors that greatly influence the development of psychological disorders?

02

Exploring the key factors that influence and can predict the development of depression in an individual

03

Predicting, preventing, and assisting individuals who may have mental health issues and those who already have mental health problems.

Presented by Duy Khanh

Data Understanding and Preparation

This dataset contains information on individuals with various attributes related to their personal and lifestyle factors. It is designed to facilitate analysis in areas such as health, lifestyle, and socio-economic status. It includes 16 columns divided into 4 feature groups as follows:

Presented by Duy Khanh

Data Understanding and Preparation

Demographic Information

- Name
- Age
- Marital Status
- Number of Children
- Education Level

Health and Lifestyle

- Smoking Status
- Physical Activity Level
- Alcohol Consumption
- Dietary Habits
- Sleep Patterns
- Chronic Medical Conditions

Socioeconomic Status

- Employment Status
- Income

Mental Health History

- History of Mental Illness
- History of Substance Abuse
- Family History of Depression

Data Understanding and Preparation

Data Checking

Checking missing values, inconsistencies, potential outlier, duplicates,...

Data Preparing

- Remove the column containing personally identifiable information.
- Convert categorical features (e.g., Marital Status, Smoking Status) into numerical values.
- Scale numerical features (e.g., Age, Income) to a standard range to ensure the models perform optimally.

Presented by Duy Khanh

Data Understanding and Preparation

RangeIndex: 413768 entries, 0 to 413767

Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	Age	413768 non-null	int64
1	Marital Status	413768 non-null	int64
2	Education Level	413768 non-null	int64
3	Number of Children	413768 non-null	int64
4	Smoking Status	413768 non-null	int64
5	Physical Activity Level	413768 non-null	int64
6	Employment Status	413768 non-null	int64
7	Income	413768 non-null	float64
8	Alcohol Consumption	413768 non-null	int64
9	Dietary Habits	413768 non-null	int64
10	Sleep Patterns	413768 non-null	int64
11	History of Mental Illness	413768 non-null	int64
12	History of Substance Abuse	413768 non-null	int64
13	Family History of Depression	413768 non-null	int64
14	Chronic Medical Conditions	413768 non-null	int64

	Age	Number of Children	Income
count	413768.000000	413768.000000	413768.000000
mean	49.000713	1.298972	50661.707971
std	18.158759	1.237054	40624.100565
min	18.000000	0.000000	0.410000
25%	33.000000	0.000000	21001.030000
50%	49.000000	1.000000	37520.135000
75%	65.000000	2.000000	76616.300000
max	80.000000	4.000000	209995.220000

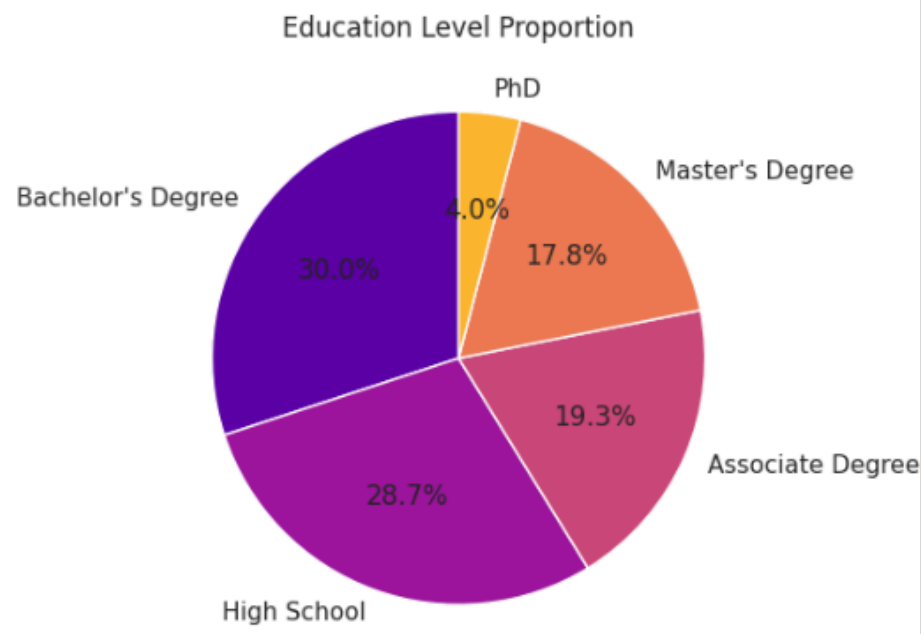
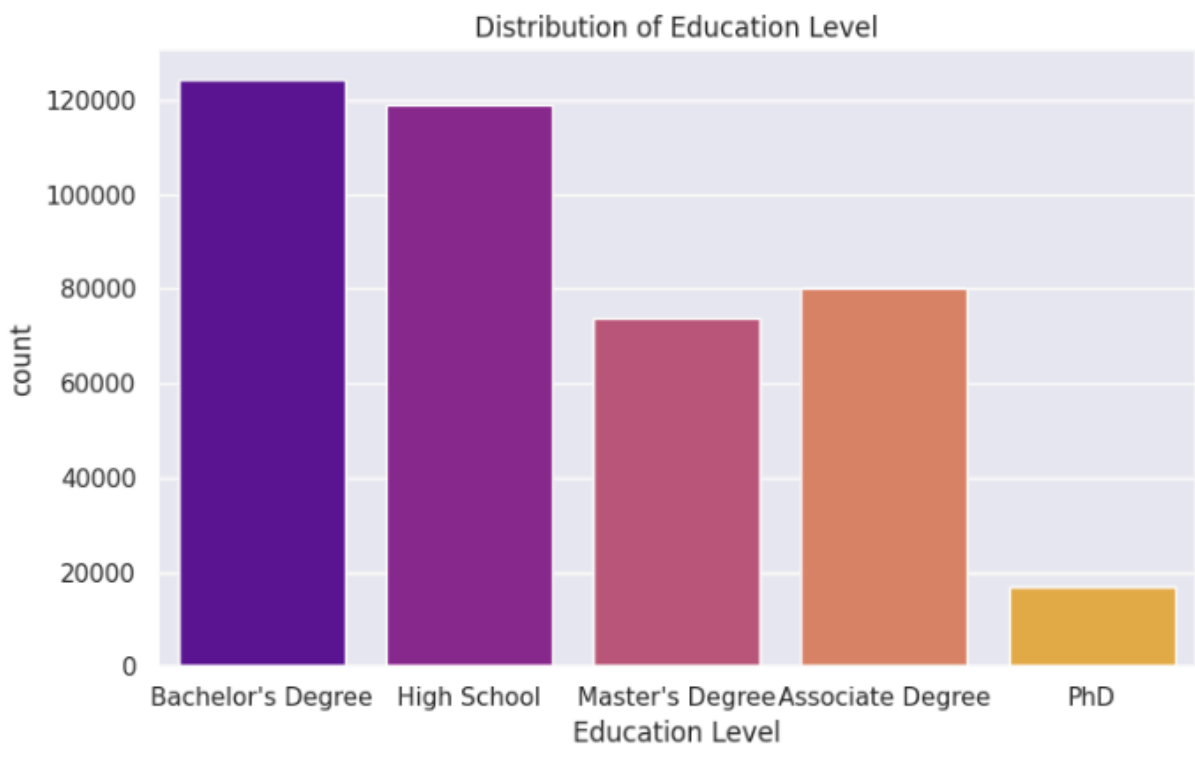
Presented by Duy Khanh

EDA and Visualization

Overview

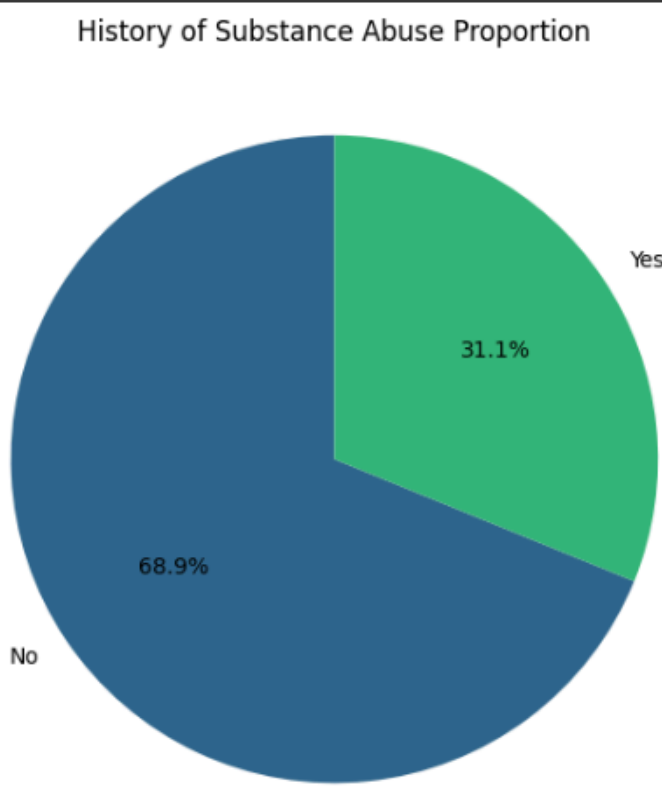
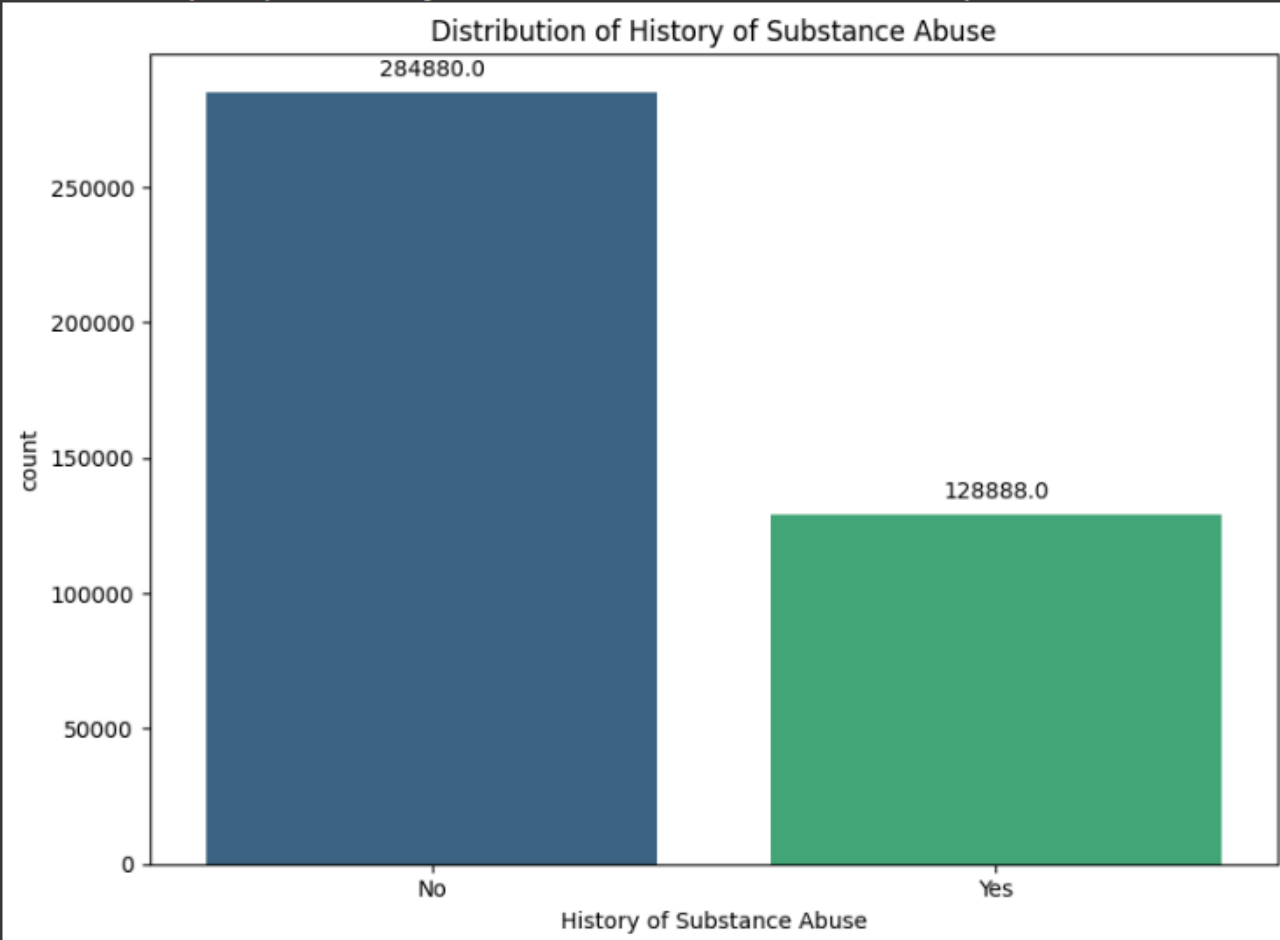
Grouping the factors, calculating, and visualizing them using charts show that the dataset is collected from individuals aged 18 and above, categorized quite thoroughly and evenly. The dataset is gathered from individuals with diverse aspects in all areas. Therefore, this dataset is highly comprehensive when applying mental health prediction models.

Presented by Duy Khanh



Education Level

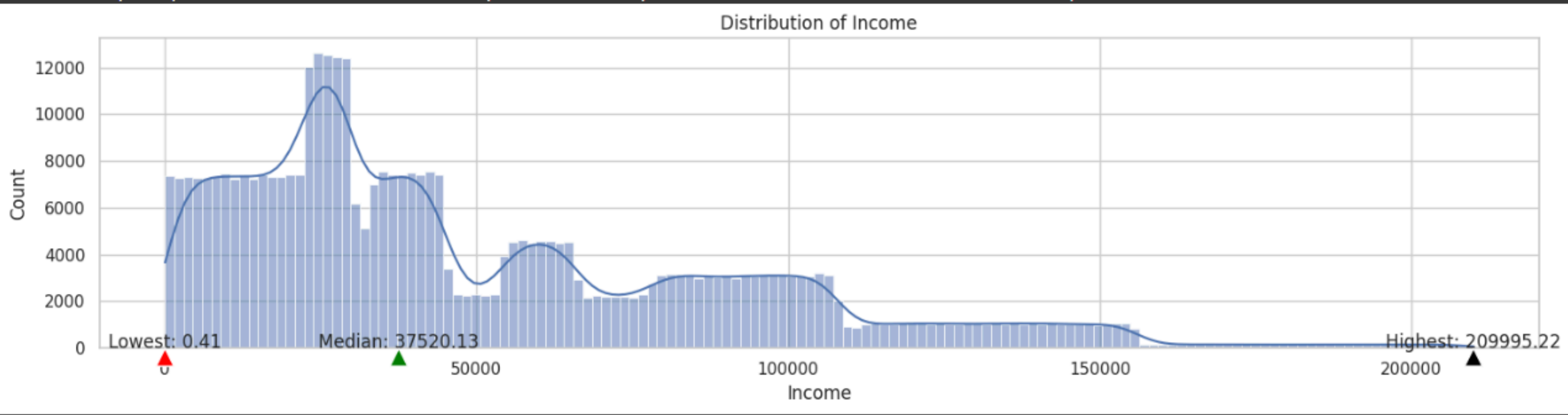
History of Substance Abuse





Employment Status

Income



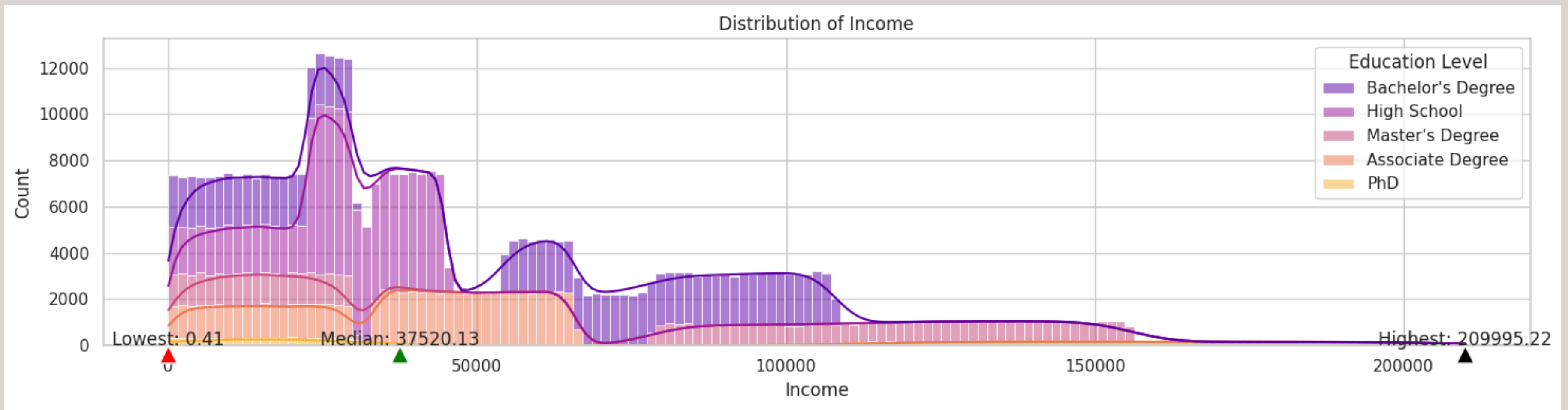
EDA and Visualization

Additional

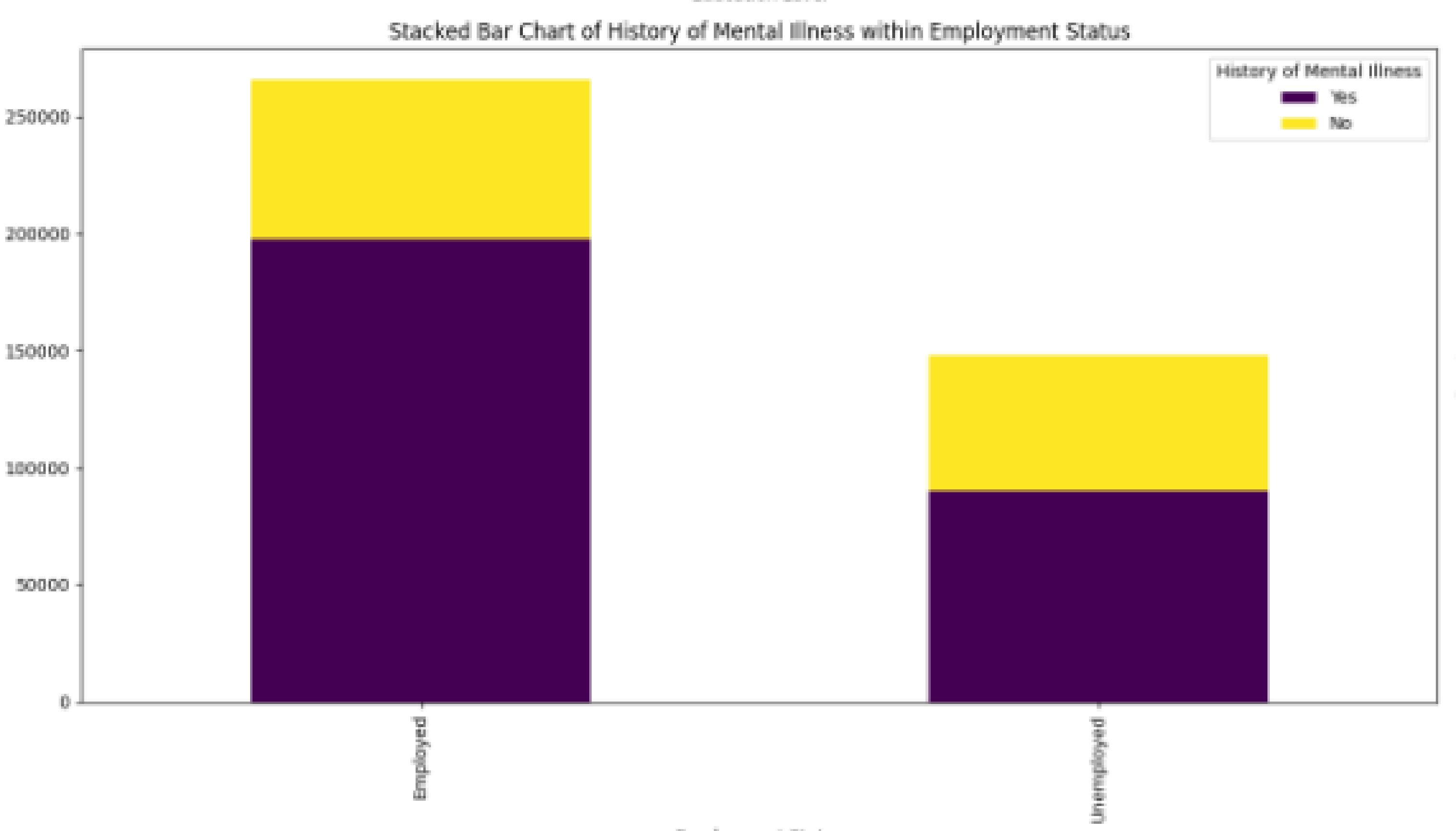
Grouping the factors into different categories helps to gain clearer insights into the relationships and influences between the features, especially the relationship between History of Mental Illness and other features.

Presented by Duy Khanh

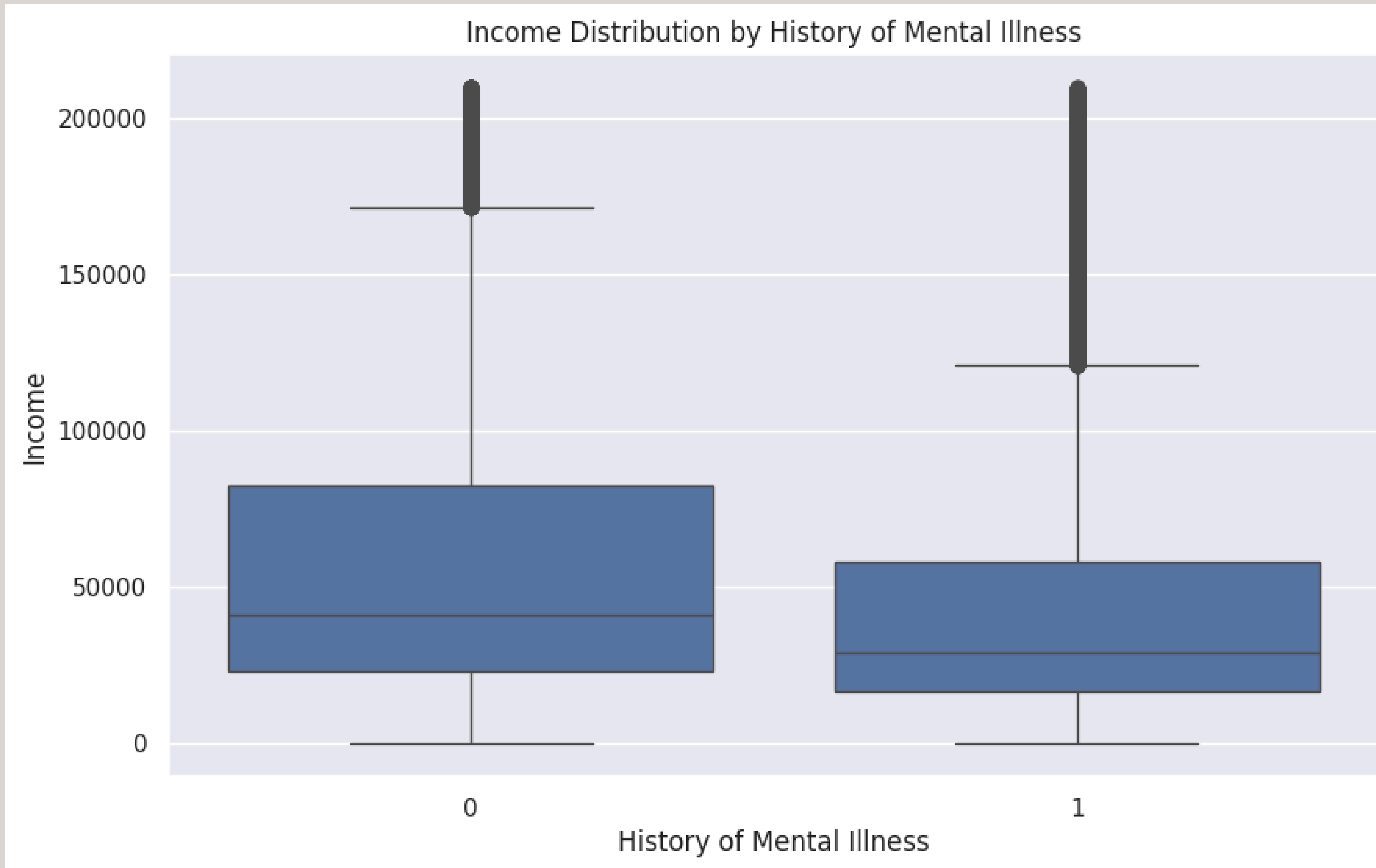
Relationship between Income and Education Level



Relationship between History of Mental Illness and Employment Status



Relationship between History of Mental Illness and Income



EDA and Visualization

Key Features

- 1. Marital Status:** Single, divorced, and widowed individuals show a higher count of history of mental illness compared to married individuals.
- 2. Education Level:** Individuals with lower education levels (e.g., no formal education or primary education) show a higher count of history of mental illness compared to those with higher education levels (e.g., college or university degree).
- 3. Employment Status:** Unemployed individuals have a higher count of history of mental illness compared to employed individuals.
- 4. History of Substance Abuse:** Individuals with a history of substance abuse show a higher count of history of mental illness compared to those without.
- 5. Family History of Depression:** A family history of depression is associated with a higher count of history of mental illness.
- 6. Income:** Individuals with no history of mental illness tend to have higher median incomes compared to those with a history of mental illness.

Presented by Duy Khanh

Depression Predicting with Classification Model and Evaluation

Presented by Duy Khanh

Correlation Matrix

1. Employment Status: 0.14

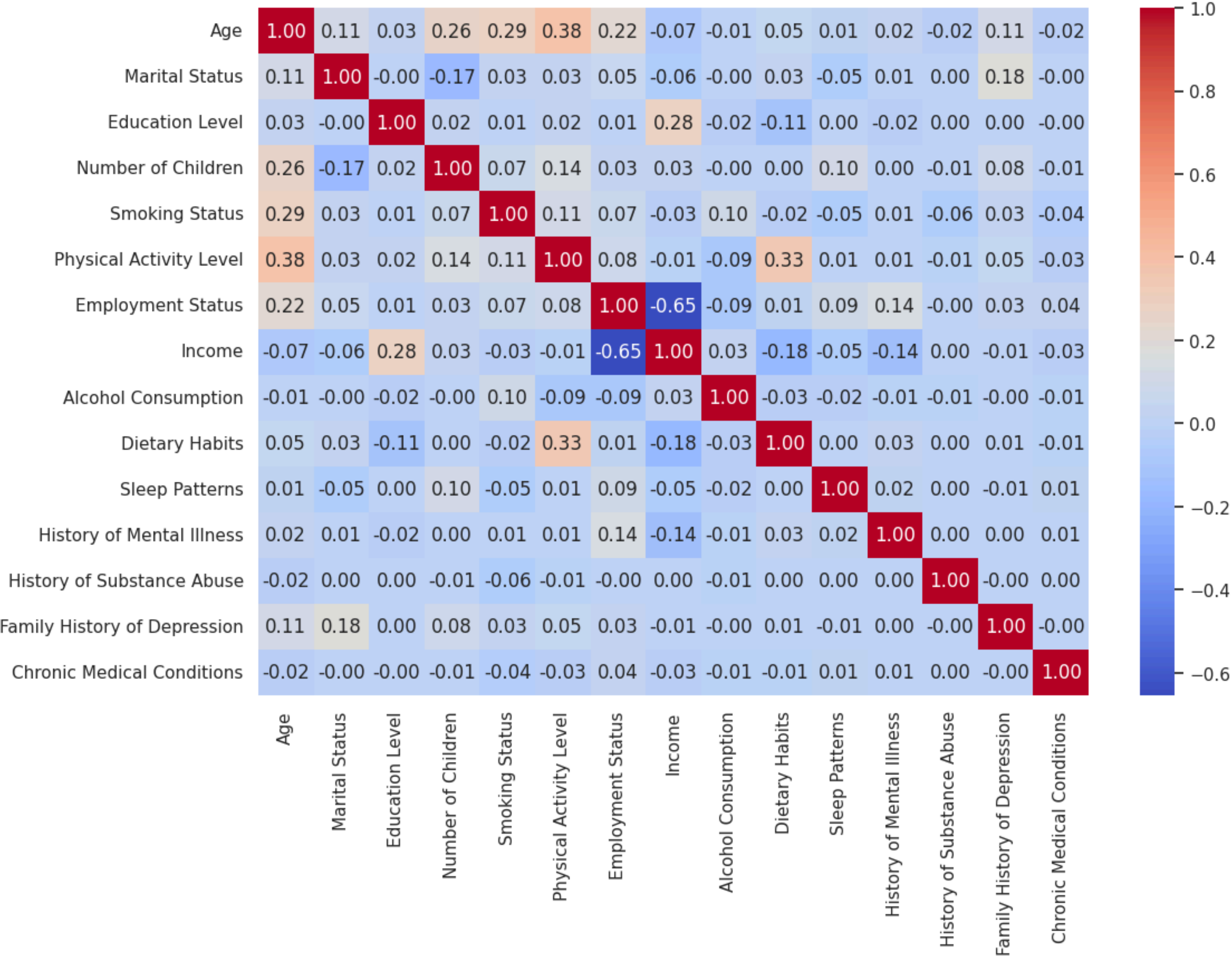
- Indicating that individuals with a history of mental illness are more likely to be unemployed.

2. Income: -0.14

- Suggesting that individuals with a history of mental illness tend to have lower incomes.

3. History of Substance Abuse: 0.14

- Indicating that individuals with a history of mental illness are more likely to have a history of substance abuse.



Optimization terminated successfully.
Current function value: 0.602386
Iterations 5

Logit Regression Results

=====						
Dep. Variable:	History of Mental Illness		No. Observations:	331014		
Model:	Logit		Df Residuals:	330999		
Method:	MLE		Df Model:	14		
Date:	Mon, 30 Dec 2024		Pseudo R-squ.:	0.01933		
Time:	13:41:25		Log-Likelihood:	-1.9940e+05		
converged:	True		LL-Null:	-2.0333e+05		
Covariance Type:	nonrobust		LLR p-value:	0.000		
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	-0.7887	0.020	-39.393	0.000	-0.828	-0.749
Age	0.0003	0.000	1.284	0.199	-0.000	0.001
Marital Status	-0.0043	0.005	-0.916	0.359	-0.013	0.005
Education Level	0.0032	0.004	0.862	0.388	-0.004	0.010
Number of Children	0.0011	0.003	0.331	0.740	-0.005	0.008
Smoking Status	-0.0118	0.006	-2.050	0.040	-0.023	-0.001
Physical Activity Level	-0.0176	0.006	-2.975	0.003	-0.029	-0.006
Employment Status	0.3807	0.011	33.570	0.000	0.358	0.403
Income	-4.558e-06	1.47e-07	-30.995	0.000	-4.85e-06	-4.27e-06
Alcohol Consumption	-0.0057	0.005	-1.174	0.240	-0.015	0.004
Dietary Habits	0.0405	0.006	7.011	0.000	0.029	0.052
Sleep Patterns	0.0071	0.004	1.613	0.107	-0.002	0.016
History of Substance Abuse	0.0101	0.008	1.218	0.223	-0.006	0.026
Family History of Depression	0.0059	0.009	0.673	0.501	-0.011	0.023
Chronic Medical Conditions	0.0059	0.008	0.728	0.467	-0.010	0.022
=====						

Accuracy: 0.69573676196921

Logictis Regression Model

- Model Accuracy: 69,5%
- Significant Predictors:
 - Income
 - Employment Status
 - History of Substance Abuse
 - Family History of Depression

Accuracy: 0.6641612490030694

Feature Importances:

	Feature	Importance
7	Income	0.388671
0	Age	0.249901
3	Number of Children	0.057259
8	Alcohol Consumption	0.040443
2	Education Level	0.036566
5	Physical Activity Level	0.033874
4	Smoking Status	0.029311
9	Dietary Habits	0.029125
1	Marital Status	0.027599
10	Sleep Patterns	0.026182
13	Chronic Medical Conditions	0.023782
11	History of Substance Abuse	0.023175
12	Family History of Depression	0.022739
6	Employment Status	0.011373

Random Forrest Model

- Model Accuracy: 66,4%

Actional Recommendation

Family

- Income support.
- Physical Activity, Dietary Habits and Sleep Pattern.

Social workers

Assessing economic and financial situations and providing support services for unemployment and job placement are essential. Additionally, workshops and outdoor activities that enhance understanding and awareness of mental health are also very beneficial.

Psychologists

A comprehensive assessment of patients, including factors such as family history, income, employment status, and substance use, is essential in the treatment process. Providing advice on improving mental health by eliminating addictive substances, or incorporating job placement and vocational training during treatment, is crucial.

Presented by Duy Khanh

Conclusion

These systems are based on the significant factors identified in the logistic regression analysis and are aimed at improving the mental health and well-being of individuals through targeted interventions and support. By addressing these factors, families, psychologists, and medical social workers can work together to reduce the tendency of depression and enhance overall mental health outcomes.

Presented by Duy Khanh



Thank You

Presented by Duy Khanh