Suicide Prevention

DDS Projects Showcase

May 11, 2023

Trigger Warning
topics include suicide, mental health issues, drug use, and crime

Team Lead: Zairan Xiang
Team Members: Courtney Cheung, Vicky Li, Lana Murray, Josh Puray

Presentation Overview

- Background/Significance
- Dataset & Our Question
- Pre-Processing & EDA
- Models and Analysis
- Limitations/Challenges
- Conclusion

Background/Significance

Background/Significance

Stats (World Health Organization - 2021)

- More than 700,000 people die due to suicide every year.
- Suicide is the fourth leading cause of death among 15-29 year-olds.
- 77% of global suicides occur in low and middle income countries.

Significance

- By analyzing relevant features and attempting to predict suicide risk, we aim to spread awareness and emphasize that suicide is preventable.
- Suicide risk assessment is used by schools, healthcare services, and mental health professionals
- Important to understand what contributes to suicide risk
- Wanted to explore the relationship between drug use and mental health issues
- The better we understand risk, the faster we can provide quality mental health and harm reduction services

Dataset & Research Question

Our Dataset

- The National Survey on Drug Use and Health (NSDUH) Series
- Years 2010-2014
- Over 3000 features and 50,000 observations for each year
- Detailed documentation codebook with 900+ pages
- Needed to reduce number of features

Main Feature Categories (52 in totals)

- Mental Health
- Demographics
- Crime
- Drug Use
- Religiosity

Research Question

Which features are most predictive of suicide risk?

Pre-Processing & EDA

Roadmap for Pre-Processing

Before merging the 2010-2014 dataframes, we had to:

1. Only use features shared by each dataframe
2. EDA by topic category
3. Delete features with too much missingness and impute missingness
4. Find the most meaningful features
5. Rename columns for clarity and replace values in dataset
6. normalize all independent variables

Generating Suicide Risk Feature (Predictive Feature)

Suicide Risk is calculated using the following variables

- Think: Have you ever thought about suicide in the past year
- Plan: Have you ever planned suicide
- Attempt: Have you ever attempted suicide

Suicide Risk = Think + 2*Plan + 3*Attempt

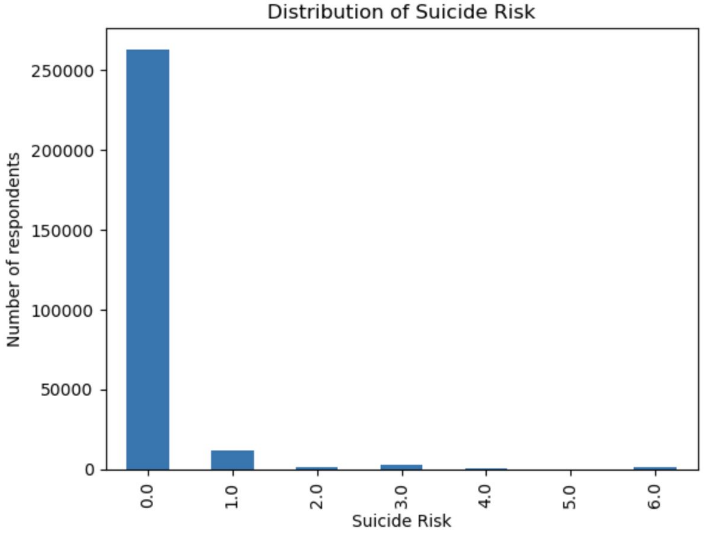


Figure 1: Distribution of Suicide Risk

EDA

Checking for Multicollinearity

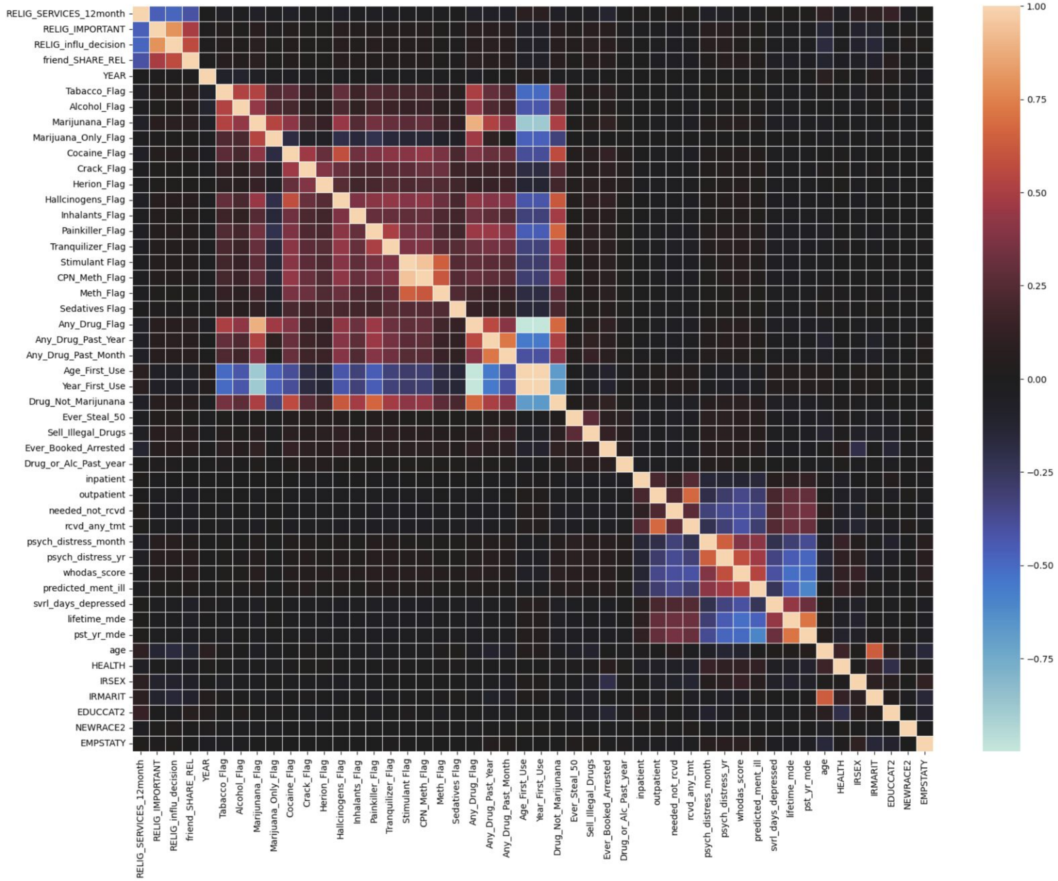
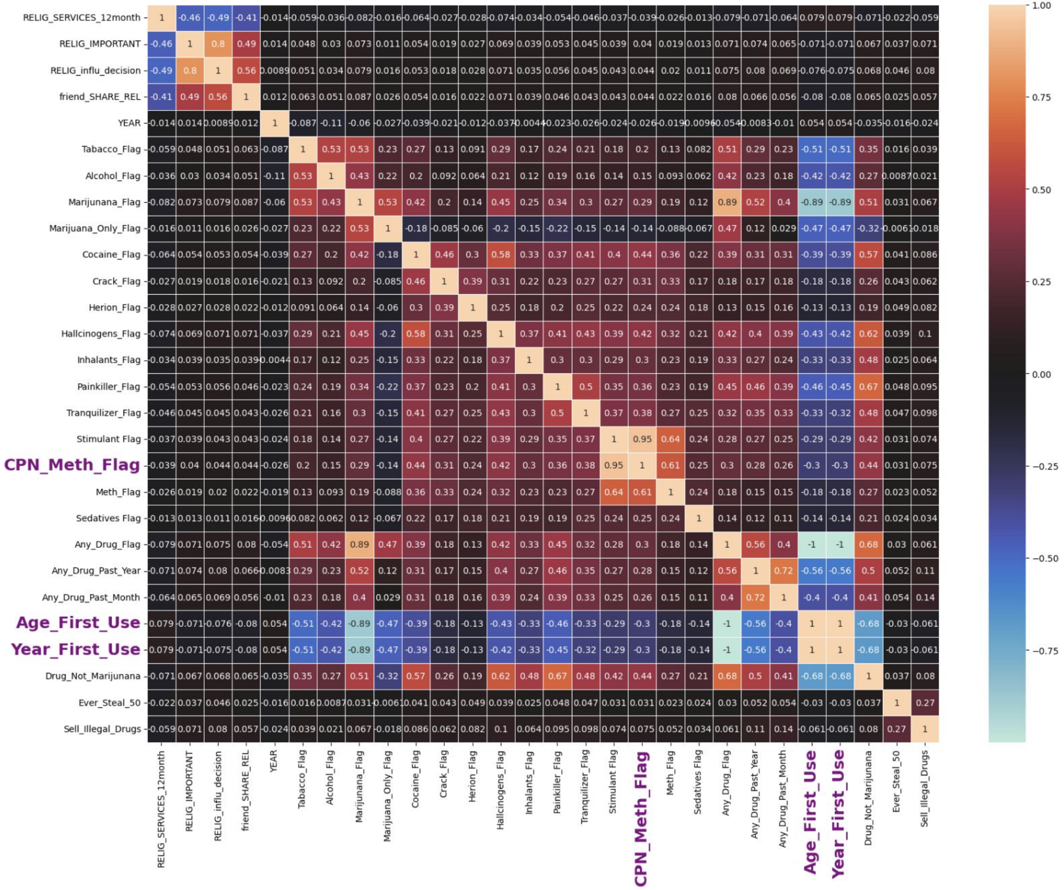


Figure 2A: Correlation Heatmap

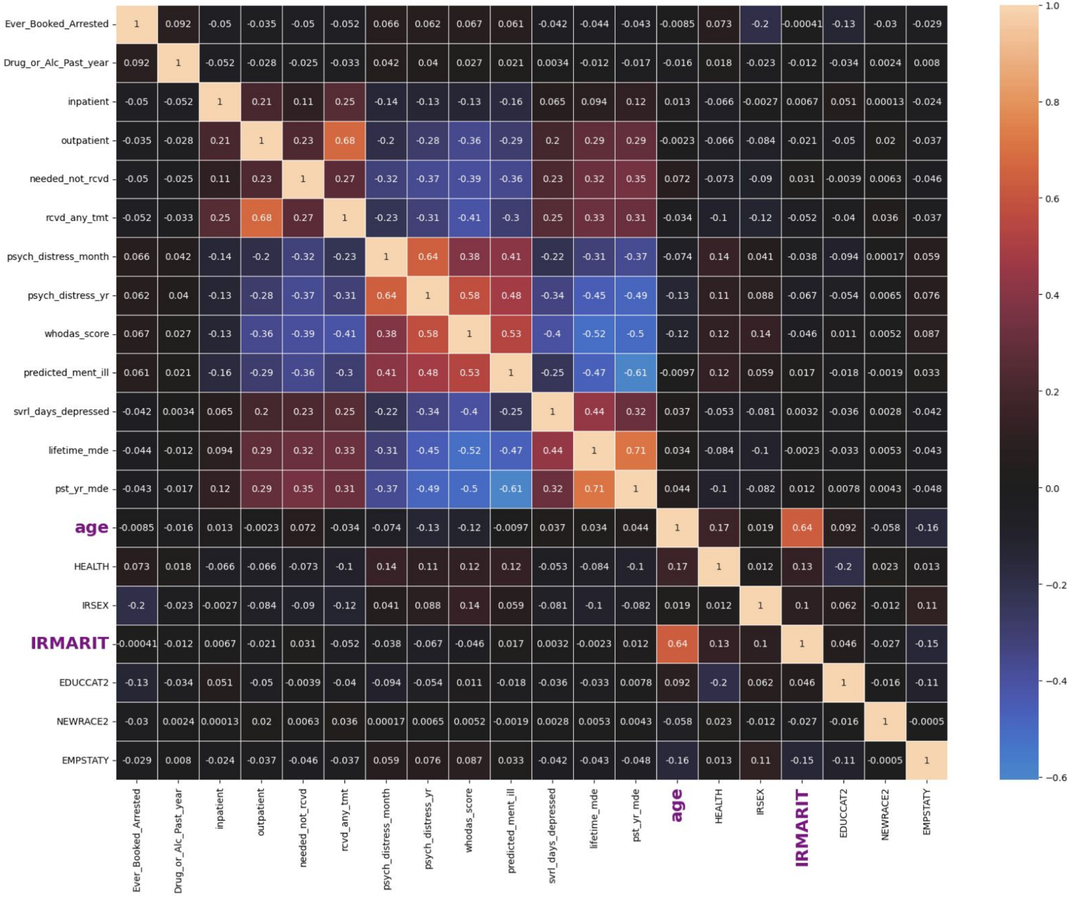
Correlation Heatmap of subset 1



	RELIG_SERVICES_12month	RELIG_IMPORTANT	RELIG_influ_decision	friend_SHARE_REL	YEAR	Tabacco_Flag	Alcohol_Flag	Marijunana_Flag	Marijuana_Only_Flag	Cocaine_Flag	Crack_Flag	Herion_Flag	Hallcinogens_Flag	Inhalants_Flag	Painkiller_Flag	Tranquilizer_Flag	Stimulant Flag	CPN_Meth_Flag	Meth_Flag	Sedatives Flag	Any_Drug_Flag	Any_Drug_Past_Year	Any_Drug_Past_Month	Age_First_Use	Year_First_Use	Drug_Not_Marijunana	Ever_Steal_50	Sell_Illegal_Drugs
RELIG_SERVICES_12month	1	-0.46	-0.49	-0.41	-0.014	-0.059	-0.036	-0.082	-0.016	-0.064	-0.027	-0.028	-0.074	-0.034	-0.054	-0.046	-0.037	-0.039	-0.026	-0.013	-0.079	-0.071	-0.064	0.079	0.079	-0.071	-0.022	-0.059
RELIG_IMPORTANT	-0.46	1	0.8	0.49	0.014	0.048	0.03	0.073	0.011	0.054	0.019	0.027	0.069	0.039	0.053	0.045	0.039	0.04	0.019	0.013	0.071	0.074	0.065	-0.071	-0.071	0.067	0.037	0.071
RELIG_influ_decision	-0.49	0.8	1	0.56	0.0089	0.051	0.034	0.079	0.016	0.053	0.018	0.028	0.071	0.035	0.056	0.045	0.043	0.044	0.02	0.011	0.075	0.08	0.069	-0.076	-0.075	0.068	0.046	0.08
friend_SHARE_REL	-0.41	0.49	0.56	1	0.012	0.063	0.051	0.087	0.026	0.054	0.016	0.022	0.071	0.039	0.046	0.043	0.043	0.044	0.022	0.016	0.08	0.066	0.056	-0.08	-0.08	0.065	0.025	0.057
YEAR	-0.014	0.014	0.0089	0.012	1	-0.087	-0.11	-0.06	-0.027	-0.039	-0.021	-0.012	-0.037	-0.0044	-0.023	-0.026	-0.024	-0.026	-0.019	-0.0096	-0.054	-0.0083	-0.01	0.054	0.054	-0.035	-0.016	-0.024
Tabacco_Flag	-0.059	0.048	0.051	0.063	-0.087	1	0.53	0.53	0.23	0.27	0.13	0.091	0.29	0.17	0.24	0.21	0.18	0.2	0.13	0.082	0.51	0.29	0.23	-0.51	-0.51	0.35	0.016	0.039
Alcohol_Flag	-0.036	0.03	0.034	0.051	-0.11	0.53	1	0.43	0.22	0.2	0.092	0.064	0.21	0.12	0.19	0.16	0.14	0.15	0.093	0.062	0.42	0.23	0.18	-0.42	-0.42	0.27	0.0087	0.021
Marijunana_Flag	-0.082	0.073	0.079	0.087	-0.06	0.53	0.43	1	0.53	0.42	0.2	0.14	0.45	0.25	0.34	0.3	0.27	0.29	0.19	0.12	0.89	0.52	0.4	-0.89	-0.89	0.51	0.031	0.067
Marijuana_Only_Flag	-0.016	0.011	0.016	0.026	-0.027	0.23	0.22	0.53	1	-0.18	-0.085	-0.06	-0.2	-0.15	-0.22	-0.15	-0.14	-0.14	-0.088	-0.067	0.47	0.12	0.029	-0.47	-0.47	-0.32	-0.0061	-0.018
Cocaine_Flag	-0.064	0.054	0.053	0.054	-0.039	0.27	0.2	0.42	-0.18	1	0.46	0.3	0.58	0.33	0.37	0.41	0.4	0.44	0.36	0.22	0.39	0.31	0.31	-0.39	-0.39	0.57	0.041	0.086
Crack_Flag	-0.027	0.019	0.018	0.016	-0.021	0.13	0.092	0.2	-0.085	0.46	1	0.39	0.31	0.22	0.23	0.27	0.27	0.31	0.33	0.17	0.18	0.17	0.18	-0.18	-0.18	0.26	0.043	0.062
Herion_Flag	-0.028	0.027	0.028	0.022	-0.012	0.091	0.064	0.14	-0.06	0.3	0.39	1	0.25	0.18	0.2	0.25	0.22	0.24	0.24	0.18	0.13	0.15	0.16	-0.13	-0.13	0.19	0.049	0.082
Hallcinogens_Flag	-0.074	0.069	0.071	0.071	-0.037	0.29	0.21	0.45	-0.2	0.58	0.31	0.25	1	0.37	0.41	0.43	0.39	0.42	0.32	0.21	0.42	0.4	0.39	-0.43	-0.42	0.62	0.039	0.1
Inhalants_Flag	-0.034	0.039	0.035	0.039	-0.0044	0.17	0.12	0.25	-0.15	0.33	0.22	0.18	0.37	1	0.3	0.3	0.29	0.3	0.23	0.19	0.33	0.27	0.24	-0.33	-0.33	0.48	0.025	0.064
Painkiller_Flag	-0.054	0.053	0.056	0.046	-0.023	0.24	0.19	0.34	-0.22	0.37	0.23	0.2	0.41	0.3	1	0.5	0.35	0.36	0.23	0.19	0.45	0.46	0.39	-0.46	-0.45	0.67	0.048	0.095
Tranquilizer_Flag	-0.046	0.045	0.045	0.043	-0.026	0.21	0.16	0.3	-0.15	0.41	0.27	0.25	0.43	0.3	0.5	1	0.37	0.38	0.27	0.25	0.32	0.35	0.33	-0.33	-0.32	0.48	0.047	0.098
Stimulant Flag	-0.037	0.039	0.043	0.043	-0.024	0.18	0.14	0.27	-0.14	0.4	0.27	0.22	0.39	0.29	0.35	0.37	1	0.95	0.64	0.24	0.28	0.27	0.25	-0.29	-0.29	0.42	0.031	0.074
CPN_Meth_Flag	-0.039	0.04	0.044	0.044	-0.026	0.2	0.15	0.29	-0.14	0.44	0.31	0.24	0.42	0.3	0.36	0.38	0.95	1	0.61	0.25	0.3	0.28	0.26	-0.3	-0.3	0.44	0.031	0.075
Meth_Flag	-0.026	0.019	0.02	0.022	-0.019	0.13	0.093	0.19	-0.088	0.36	0.33	0.24	0.32	0.23	0.23	0.27	0.64	0.61	1	0.24	0.18	0.15	0.15	-0.18	-0.18	0.27	0.023	0.052
Sedatives Flag	-0.013	0.013	0.011	0.016	-0.0096	0.082	0.062	0.12	-0.067	0.22	0.17	0.18	0.21	0.19	0.19	0.25	0.24	0.25	0.24	1	0.14	0.12	0.11	-0.14	-0.14	0.21	0.024	0.034
Any_Drug_Flag	-0.079	0.071	0.075	0.08	-0.054	0.51	0.42	0.89	0.47	0.39	0.18	0.13	0.42	0.33	0.45	0.32	0.28	0.3	0.18	0.14	1	0.56	0.4	-1	-1	0.68	0.03	0.061
Any_Drug_Past_Year	-0.071	0.074	0.08	0.066	-0.0083	0.29	0.23	0.52	0.12	0.31	0.17	0.15	0.4	0.27	0.46	0.35	0.27	0.28	0.15	0.12	0.56	1	0.72	-0.56	-0.56	0.5	0.052	0.11
Any_Drug_Past_Month	-0.064	0.065	0.069	0.056	-0.01	0.23	0.18	0.4	0.029	0.31	0.18	0.16	0.39	0.24	0.39	0.33	0.25	0.26	0.15	0.11	0.4	0.72	1	-0.4	-0.4	0.41	0.054	0.14
Age_First_Use	0.079	-0.071	-0.076	-0.08	0.054	-0.51	-0.42	-0.89	-0.47	-0.39	-0.18	-0.13	-0.43	-0.33	-0.46	-0.33	-0.29	-0.3	-0.18	-0.14	-1	-0.56	-0.4	1	1	-0.68	-0.03	-0.061
Year_First_Use	0.079	-0.071	-0.075	-0.08	0.054	-0.51	-0.42	-0.89	-0.47	-0.39	-0.18	-0.13	-0.42	-0.33	-0.45	-0.32	-0.29	-0.3	-0.18	-0.14	-1	-0.56	-0.4	1	1	-0.68	-0.03	-0.061
Drug_Not_Marijunana	-0.071	0.067	0.068	0.065	-0.035	0.35	0.27	0.51	-0.32	0.57	0.26	0.19	0.62	0.48	0.67	0.48	0.42	0.44	0.27	0.21	0.68	0.5	0.41	-0.68	-0.68	1	0.037	0.08
Ever_Steal_50	-0.022	0.037	0.046	0.025	-0.016	0.016	0.0087	0.031	-0.0061	0.041	0.043	0.049	0.039	0.025	0.048	0.047	0.031	0.031	0.023	0.024	0.03	0.052	0.054	-0.03	-0.03	0.037	1	0.27
Sell_Illegal_Drugs	-0.059	0.071	0.08	0.057	-0.024	0.039	0.021	0.067	-0.018	0.086	0.062	0.082	0.1	0.064	0.095	0.098	0.074	0.075	0.052	0.034	0.061	0.11	0.14	-0.061	-0.061	0.08	0.27	1

Figure 1B: Corr Heatmap of subset 1

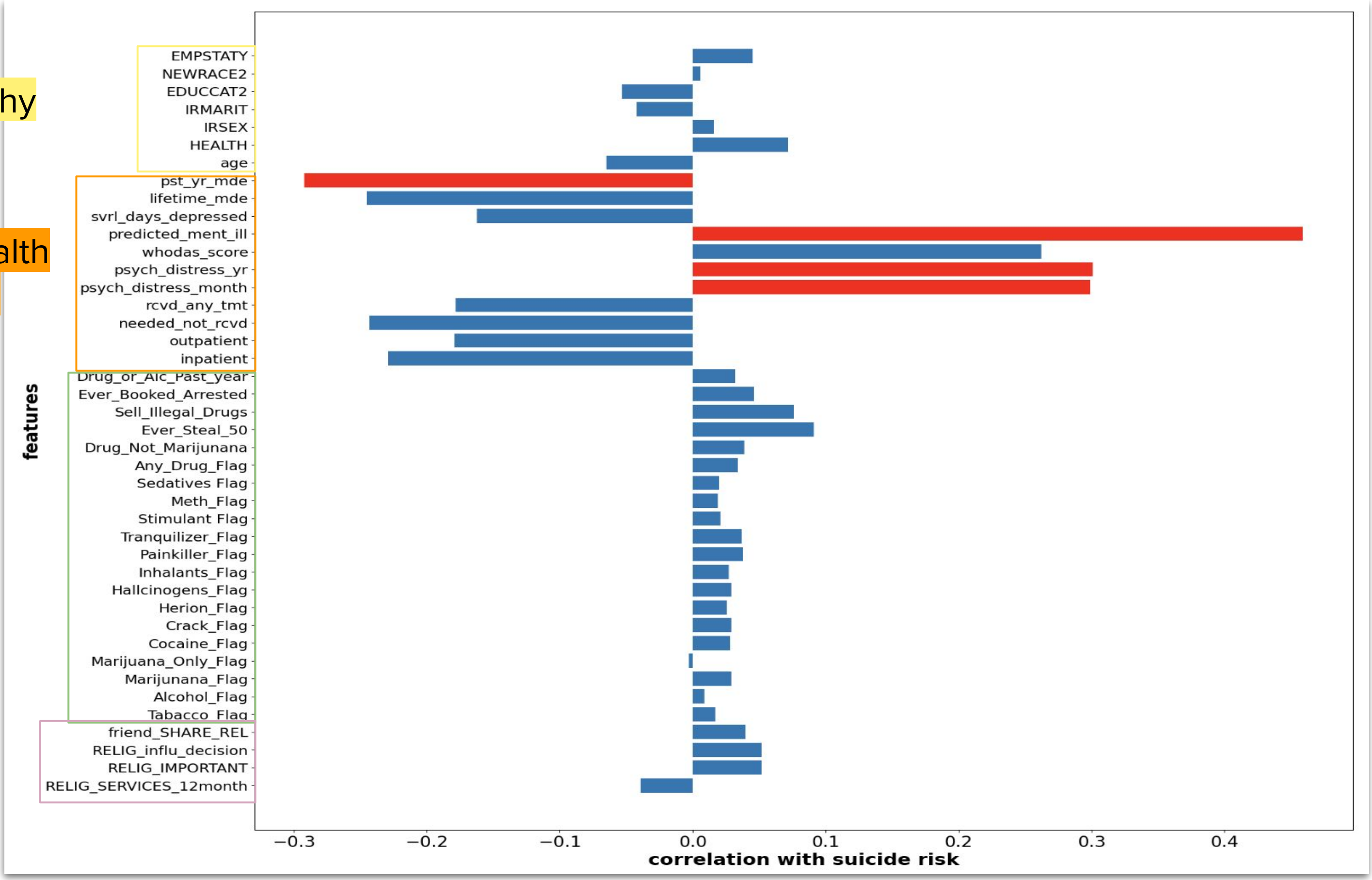
Correlation Heatmap of subset 2



	Ever_Booked_Arrested	Drug_or_Alc_Past_year	inpatient	outpatient	needed_not_rcvd	rcvd_any_tmt	psych_distress_month	psych_distress_yr	whodas_score	predicted_ment_ill	svrl_days_depressed	lifetime_mde	pst_yr_mde	**age**	HEALTH	IRSEX	**IRMARIT**	EDUCCAT2	NEWRACE2	EMPSTATY
Ever_Booked_Arrested	1	0.092	-0.05	-0.035	-0.05	-0.052	0.066	0.062	0.067	0.061	-0.042	-0.044	-0.043	-0.0085	0.073	-0.2	-0.00041	-0.13	-0.03	-0.029
Drug_or_Alc_Past_year	0.092	1	-0.052	-0.028	-0.025	-0.033	0.042	0.04	0.027	0.021	0.0034	-0.012	-0.017	-0.016	0.018	-0.023	-0.012	-0.034	0.0024	0.008
inpatient	-0.05	-0.052	1	0.21	0.11	0.25	-0.14	-0.13	-0.13	-0.16	0.065	0.094	0.12	0.013	-0.066	-0.0027	0.0067	0.051	0.00013	-0.024
outpatient	-0.035	-0.028	0.21	1	0.23	0.68	-0.2	-0.28	-0.36	-0.29	0.2	0.29	0.29	-0.0023	-0.066	-0.084	-0.021	-0.05	0.02	-0.037
needed_not_rcvd	-0.05	-0.025	0.11	0.23	1	0.27	-0.32	-0.37	-0.39	-0.36	0.23	0.32	0.35	0.072	-0.073	-0.09	0.031	-0.0039	0.0063	-0.046
rcvd_any_tmt	-0.052	-0.033	0.25	0.68	0.27	1	-0.23	-0.31	-0.41	-0.3	0.25	0.33	0.31	-0.034	-0.1	-0.12	-0.052	-0.04	0.036	-0.037
psych_distress_month	0.066	0.042	-0.14	-0.2	-0.32	-0.23	1	0.64	0.38	0.41	-0.22	-0.31	-0.37	-0.074	0.14	0.041	-0.038	-0.094	0.00017	0.059
psych_distress_yr	0.062	0.04	-0.13	-0.28	-0.37	-0.31	0.64	1	0.58	0.48	-0.34	-0.45	-0.49	-0.13	0.11	0.088	-0.067	-0.054	0.0065	0.076
whodas_score	0.067	0.027	-0.13	-0.36	-0.39	-0.41	0.38	0.58	1	0.53	-0.4	-0.52	-0.5	-0.12	0.12	0.14	-0.046	0.011	0.0052	0.087
predicted_ment_ill	0.061	0.021	-0.16	-0.29	-0.36	-0.3	0.41	0.48	0.53	1	-0.25	-0.47	-0.61	-0.0097	0.12	0.059	0.017	-0.018	-0.0019	0.033
svrl_days_depressed	-0.042	0.0034	0.065	0.2	0.23	0.25	-0.22	-0.34	-0.4	-0.25	1	0.44	0.32	0.037	-0.053	-0.081	0.0032	-0.036	0.0028	-0.042
lifetime_mde	-0.044	-0.012	0.094	0.29	0.32	0.33	-0.31	-0.45	-0.52	-0.47	0.44	1	0.71	0.034	-0.084	-0.1	-0.0023	-0.033	0.0053	-0.043
pst_yr_mde	-0.043	-0.017	0.12	0.29	0.35	0.31	-0.37	-0.49	-0.5	-0.61	0.32	0.71	1	0.044	-0.1	-0.082	0.012	0.0078	0.0043	-0.048
age	-0.0085	-0.016	0.013	-0.0023	0.072	-0.034	-0.074	-0.13	-0.12	-0.0097	0.037	0.034	0.044	1	0.17	0.019	0.64	0.092	-0.058	-0.16
HEALTH	0.073	0.018	-0.066	-0.066	-0.073	-0.1	0.14	0.11	0.12	0.12	-0.053	-0.084	-0.1	0.17	1	0.012	0.13	-0.2	0.023	0.013
IRSEX	-0.2	-0.023	-0.0027	-0.084	-0.09	-0.12	0.041	0.088	0.14	0.059	-0.081	-0.1	-0.082	0.019	0.012	1	0.1	0.062	-0.012	0.11
IRMARIT	-0.00041	-0.012	0.0067	-0.021	0.031	-0.052	-0.038	-0.067	-0.046	0.017	0.0032	-0.0023	0.012	0.64	0.13	0.1	1	0.046	-0.027	-0.15
EDUCCAT2	-0.13	-0.034	0.051	-0.05	-0.0039	-0.04	-0.094	-0.054	0.011	-0.018	-0.036	-0.033	0.0078	0.092	-0.2	0.062	0.046	1	-0.016	-0.11
NEWRACE2	-0.03	0.0024	0.00013	0.02	0.0063	0.036	0.00017	0.0065	0.0052	-0.0019	0.0028	0.0053	0.0043	-0.058	0.023	-0.012	-0.027	-0.016	1	-0.0005
EMPSTATY	-0.029	0.008	-0.024	-0.037	-0.046	-0.037	0.059	0.076	0.087	0.033	-0.042	-0.043	-0.048	-0.16	0.013	0.11	-0.15	-0.11	-0.0005	1

Figure 1B: Corr Heatmap of subset 2

Linear Correlation with Suicide Risk



demography

Mental health (our focus)

Drug use

Religion

Linear Relationship Between Features and Suicide Risk

The most related features to the risk of committing suicide are also related to mental health/emotions. The top 5 are:

Binary:

- Serious Mental Illness (SMI) indicator: 0.459,
- Experiencing Serious Psychological Distress in Last Year: 0.301,
- Experiencing Serious Psychological Distress in Last Month: 0.299,
- Experiencing Major Depressive Episode in Past Year (For Adult): -0.292

Continuous:

- Level of Difficulty in Performing Daily Activities due to Problems with Mental Health: 0.262

Discovery From the Most Correlated Features

- ~62% of individuals predicted to have **mental illness** (SMI) have a suicide risk > 0, with near 25% having a risk >= 3, 10% having a highest risk of 6.
- ~13% of individuals experiencing Major Depressive Episode in the last year have a risk >= 3
- ~10% of individuals experiencing Serious Psychological Distress in the last year have a risk >= 3
- A risk >= 3 means one has at least thought about and planed for committing suicide

Hypothesis

We realized the risk of suicide is most correlated to features associated with recent mental health, so we are interested in

- whether having a good mental health is predicted to be able to help decrease the risk of someone committing suicide, and
- to what extent does it help

Our Models

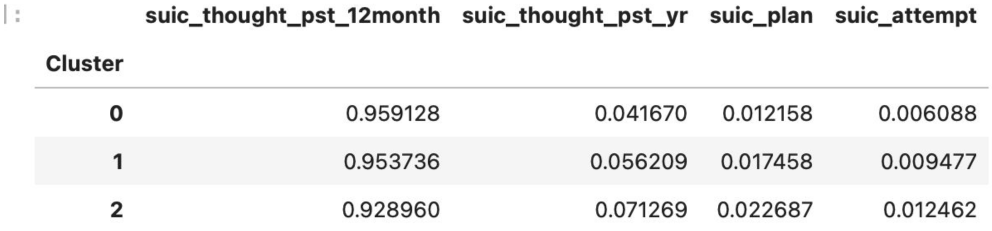
K-Modes Clustering

K-modes is used for clustering categorical variables

Choosing Optimal K = 3

EDA:

- Generalize why one cluster would have higher suicide risk
- Create distribution graphs comparing the clusters



Cluster	suic_thought_pst_12month	suic_thought_pst_yr	suic_plan	suic_attempt
0	0.959128	0.041670	0.012158	0.006088
1	0.953736	0.056209	0.017458	0.009477
2	0.928960	0.071269	0.022687	0.012462

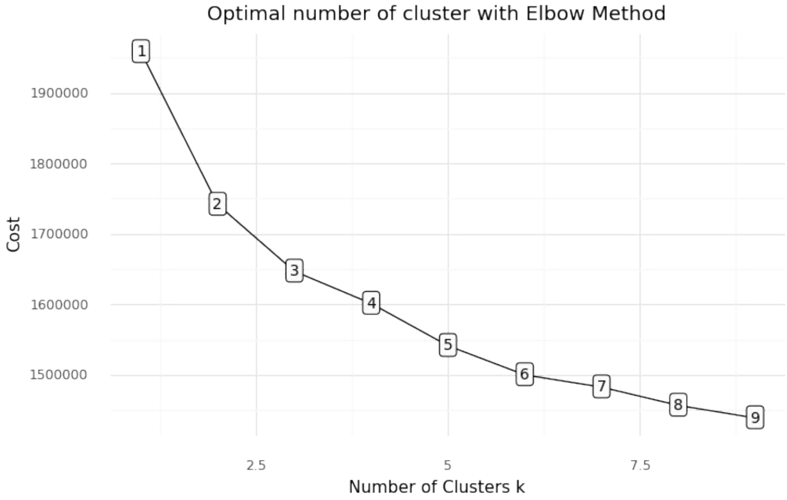


Figure 3: Elbow Method

K-modes clustering

Permutation Testing

- **Null Hypothesis:** Both samples are drawn from the same population thus the observed TVD **is due to chance alone**.

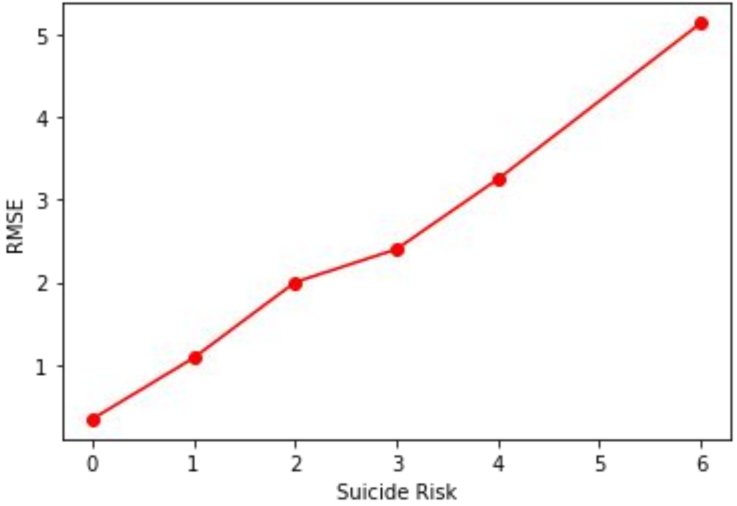
- **Alternative Hypothesis:** Samples are drawn from two different populations thus the TVD **is not due to chance alone**, and there are other factors at hand.

Results: Significant difference in distributions of suicide risk, but hard to tell which features are most important.

Regression Model Baseline

K Neighbors Regressor

Overall RMSE: 0.6286

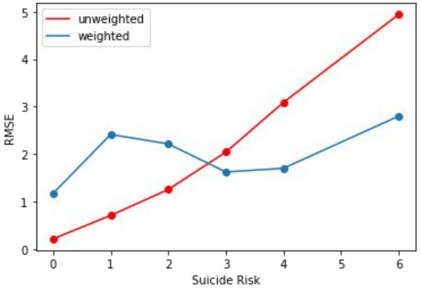
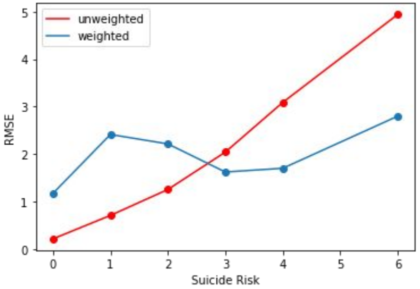
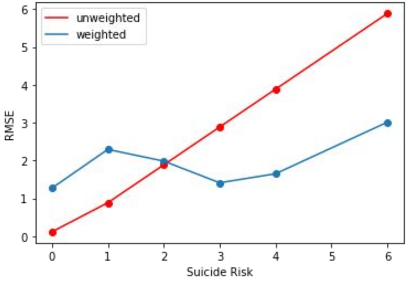
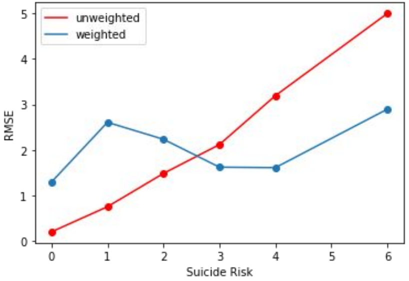


Regression Model Criteria

Better to have better prediction performance (low RMSE) for high risk respondents at the expense of performance for low risk.

Idea: add sample weights

Weight = 2^suicide_risk

	Least Squares	Ridge	LASSO	Random Forest
RMSE overall	0.5208	0.5209	0.6042	0.5253
RMSE weighted	1.8405	1.8405	1.9511	1.9381
RMSE by risk				

Important Features

7 important features common in the least squares and ridge regression models

1. Predicted_ment_ill: predicted to have a serious mental illness
2. Inpatient: received inpatient mental health treatment in the past year
3. Psych_distress_yr: past year serious psychological distress indicator
4. Ever_Steal_50: ever stolen or tried to steal anything > $50
5. Drug_or_Alc_Past_year: used drugs or alcohol in the past year
6. Psych_distress_month: past month serious psychological distress indicator
7. EDUCCAT2: education level

*note: some features are still highly correlated with each other, and can be removed in further analysis

Limitations/Challenges

Challenges

Limitations include large dataset with many features, took a lot of time to understand and look through the dataset, reformulating the hypothesis question, reducing down our features, finding features that weren’t highly correlated

Some features have many values representing missingness and we had to skim through each one. Some binary features use inconsistent values representing yes/no.

Conclusion

Conclusion

The features we found that impact suicide include not only mental health variables, but variables pertaining to education level, drug and alcohol use, and crime.

Although our modeling is not robust, it gives a good idea of the different signs that may indicate higher sucide risk, as well as the process others may use in risk modeling. It is important to understand the bias that will be present in risk modeling, as well as the socioeconomic and environmental factors that contribute to risk. Idealily, we would want to address the causes that lead to declining mental health, and learn how to promote and maintain positive mental health for all people.

Thank you for your attention!