Analyzing Susceptibility to Mental Health Issues

Project Team #2 | Data Divas

Date: October 22, 2021

Team Members:
Whitney Brooks (Executive)
Catherine Greenman (Practitioner)
Margot Herman (Practitioner)
Michell Li (Practitioner)
Chiu-Feng (Steph) Yap (Practitioner)

TABLE OF CONTENTS

PROJECT DESCRIPTION

Problem Overview
While the stigma around mental health has decreased over the years, many providers have seen a spike in cases related to "diseases of despair." These include anxiety and depression, which often go untreated or lead sufferers to "self-medicate" with substances like drugs and alcohol. According to the Tufts Medical Center and One Mind at Work, depression alone accounts for about $44 billion in losses to workplace productivity. In 2019, national spending on mental health services totaled $225.1 billion and accounted for 5.5% of all health spending[1]. Furthermore, approximately 40% of Americans live in a designated mental health provider shortage area, which exacerbates the problem. Across the US, each state has discretionary funding allocated specifically for mental health. Sufficient funds and effective resource allocation are necessary for the diagnosis and treatment of mental health issues. Mental health issues are pervasive and, now more than ever, need to be better understood to address their causes and impacts in a meaningful way.

Objective
The goal of this project is to identify factors that make individuals more susceptible to mental health issues, based on self-administered substance use, demographics, and geographic information from the National Survey on Drug Use and Health (NSDUH).

Methodology
1. Exploratory Data Analysis
   We created a Tableau dashboard using the National Survey on Drug Use and Health (NSDUH) data to illustrate key findings of our dataset.
2. Statistical Analysis
   Chi-squared tests were used in our bivariate analysis, with $p<0.01$ considered significant. Multivariate logistic regression was performed to assess the relationship between our outcome variable and exposure variables. The outcome variable is captured in the data as a binary indicator of 1 (Yes) for 'Past Month Serious Psychological Distress Indicator' which is derived from a series of six questions, asking adults respondents how frequently they experienced the following symptoms in the past 30 days:
    - How often did you feel nervous?
    - How often did you feel hopeless?
    - How often did you feel restless or fidgety?
    - How often did you feel so sad/depressed that nothing could cheer you up?
    - How often did you feel that everything was an effort?
    - How often did you feel down on yourself, no good or worthless?

Questions are asked on a likert scale of 1-5, with a sum greater than 13 being the threshold for the outcome variable.

Problem Importance

Adequate funding is a key factor in most leading implementation science frameworks. According to the American Psychiatric Association (APA)[2], one in two Americans will have mental health conditions in their lifetime. When untreated, mental health challenges can have a negative effect on a person's economic solvency, leading to increased rates of homelessness and poverty, social isolation, deteriorating physical health and shorter life expectancy, and decreased profitability for employers and their shareholders due to lower employee efficiency. Not only that, suicide rates are increasing in all age groups, yet, federal funding for mental health comprises less than 7% of the total U.S. health care spending.[2] If we could identify the type of population who is more likely to struggle from mental health issues and identify features/variables features that we could potentially use to provide insights in terms of where the funding for mental health services should be distributed.

DATASETS & DETAILS

Dataset	Source	Description	Topics
National Survey on Drug Use and Health (NSDUH)	National Survey on Drug Use and Health	Survey level data on drug use and health (2015-2019)	Survey level information on the trends in specific substance use and mental illness measures
Health Professional Shortage Areas (HPSA)	Health Professional Shortage Areas	Clinic level data, including Date of Designation/Withdrawal	- Level of resource shortage - Clinic metadata - Clinic geospatial data
Census Core-Based Statistics (CBSA)	NBER National Bureau of Economic Research	Census provides delineation files listing (CBSAs) and combined statistical areas (CSAs) and their components by FIPS state and county	Relate to the CBSA column in the grant dataset so we would tie back to the CBSA column in the NSDUH dataset to extrapolate information by county-level

Health Resources and Services Administration (HRSA) Grants	HRSA Awarded Grants	Includes award amount, award year, project description, and grantee state and county level geographic data (2015-2019)	Determine total funding and mental-health related funding on a state and county level basis

DATA CLEANING AND FEATURE ENGINEERING

Data Cleaning

NSDUH

The NSDUH dataset originally has 210, 959 records and more than 2000 columns. As the NSDUH consolidates location data into three overarching categories and does not preserve state or county-level data, we decided to consolidate five years' worth of reports in order to use time as a metric for tracking trends in mental health indicators. The reports are yearly, from 2015 to 2019.

- Preserve only the columns relevant to our analysis (including answers to mental health screening questions, insurance status, and demographic data)
- Consolidate non-committal answers (don't know, refused, legitimate skip) into a single value

HPSA

The HPSA dataset has 27,813 rows and 65 columns. Rows are on the HPSA entity level. The following was done to clean the data:

- Remove variables with more than 30% data missing
- Remove variables without any variation (Break in Designation, Discipline Class Number, Data Warehouse Record Create Date Text)

The resulting dataset has 27,813 rows and 44 columns.

HRSA Awarded Grants and CBSA

HRSA Awarded Grants data contains awards from 2013 through 2021. For the Exploratory Data Analysis, we just decided to look at grants from 2015 to 2019.

- Only keep columns related to Financial Assistance, Award Year, Grant Program Description, and geographic information like State and County names
- Group by County and State and calculate the sum of the Financial Assistance as well as Financial Assistance related to mental health for all awards in a specific county from the HRSA dataset.

- Inner join HRSA with the CBSA dataset on State and County in order to relate Financial Assistance for awarded grants per county with the county population

Feature Engineering

HPSA

The first goal of feature engineering for the HPSA dataset is to create a column that allows the dataset to be joined with the NSDUH dataset. The second goal is to generate features using existing features that could enhance the dataset.

- Created a column for population density based on CBSA population density definitions for 2010. (PDEN10)
- Created Time spent as designated entity variable (DaysBeforeWithdrawn)

EXPLORATORY DATA ANALYSIS

NSDUH

Figure 1 - Distribution of susceptibility to mental health issues by age group

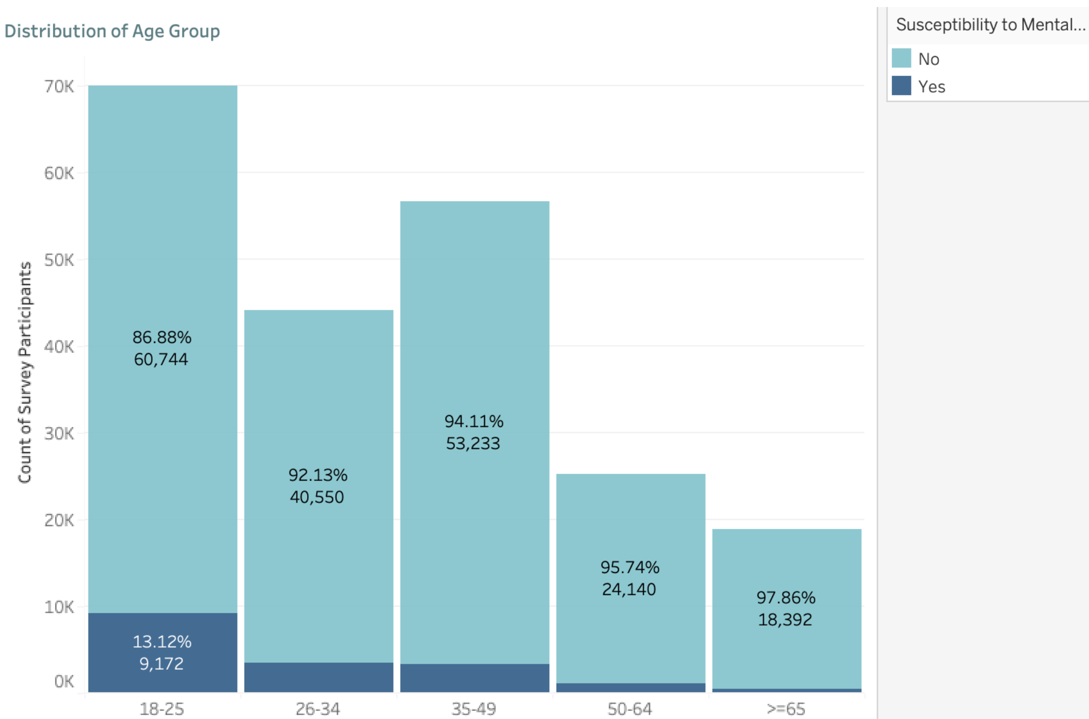


Figure 2 - Distribution of susceptibility to mental health issues across education level

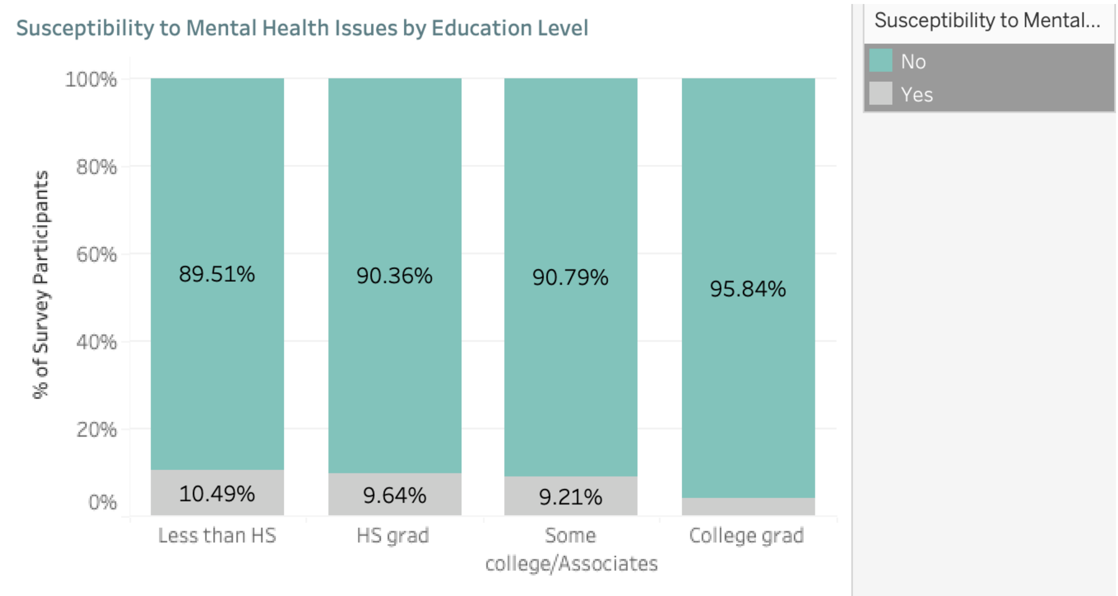


Figure 3 - Distribution of health insurance of the survey participants

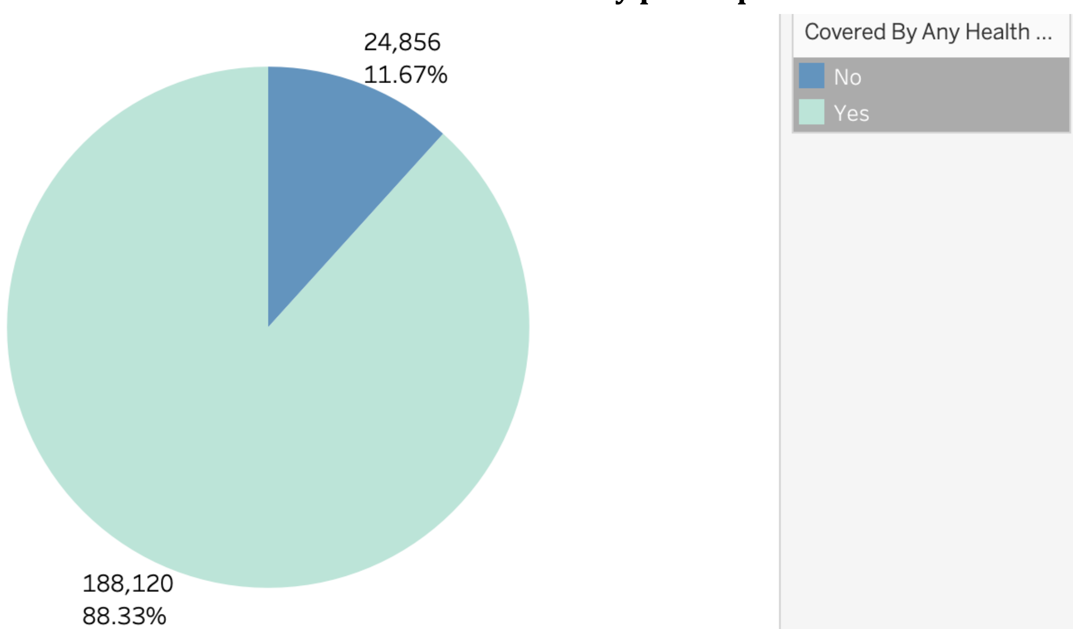


Figure 4 - Distribution of the susceptibility to mental health issues across all health insurance types (including people who reported do not have insurance)

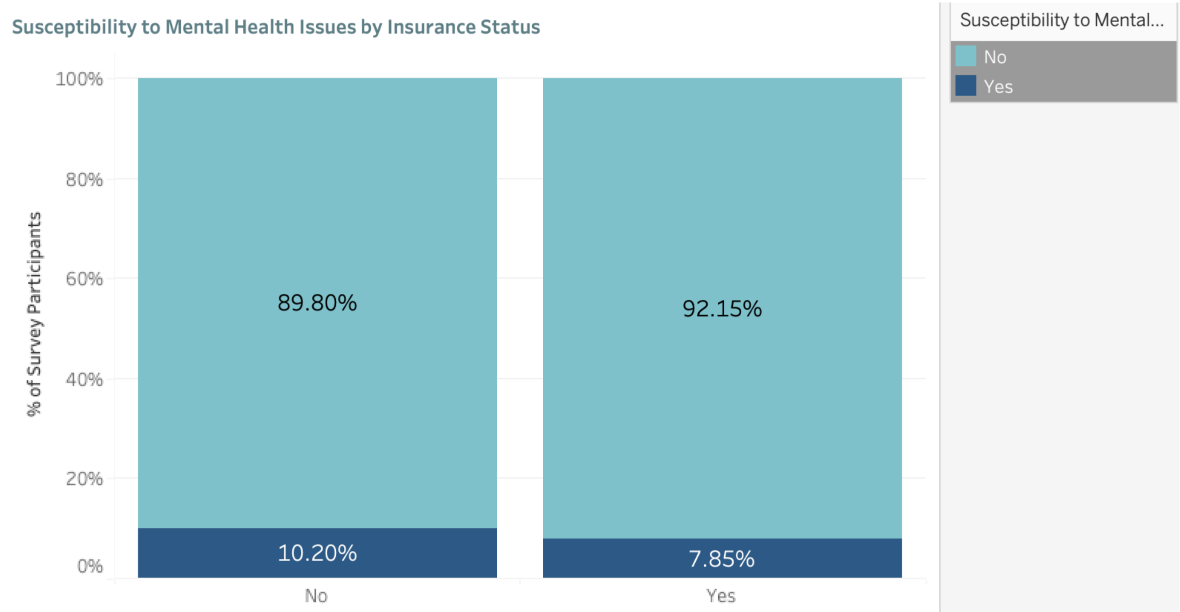
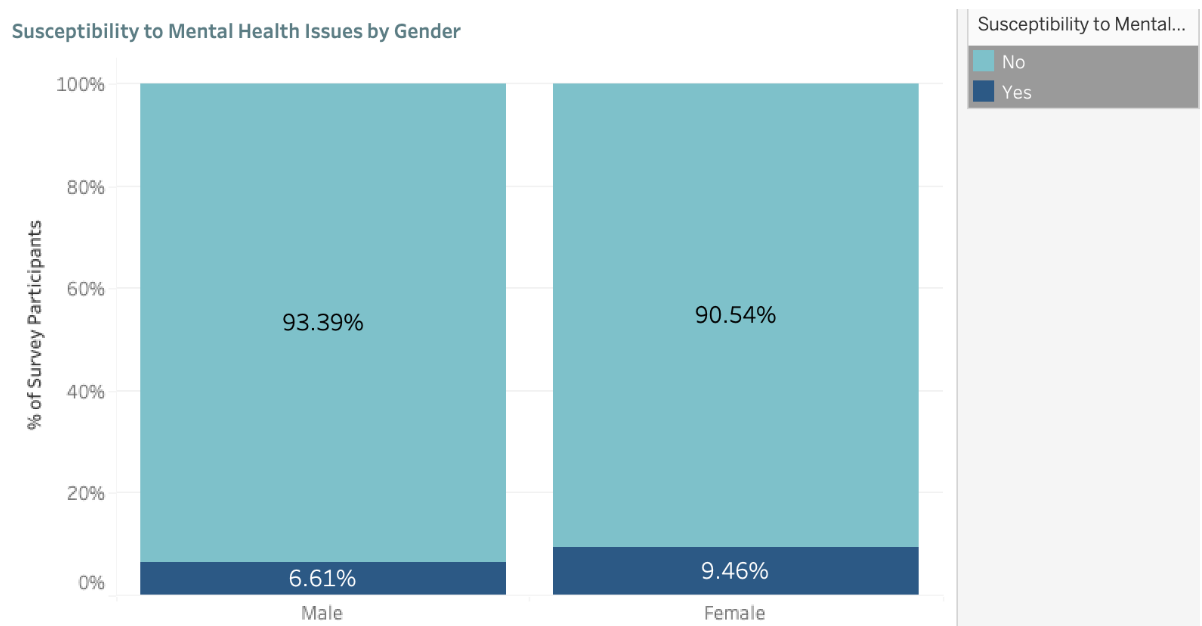


Figure 5 - Distribution of susceptibility to mental health issues across gender



HPSA

Figure 6 - Population Density Type by County

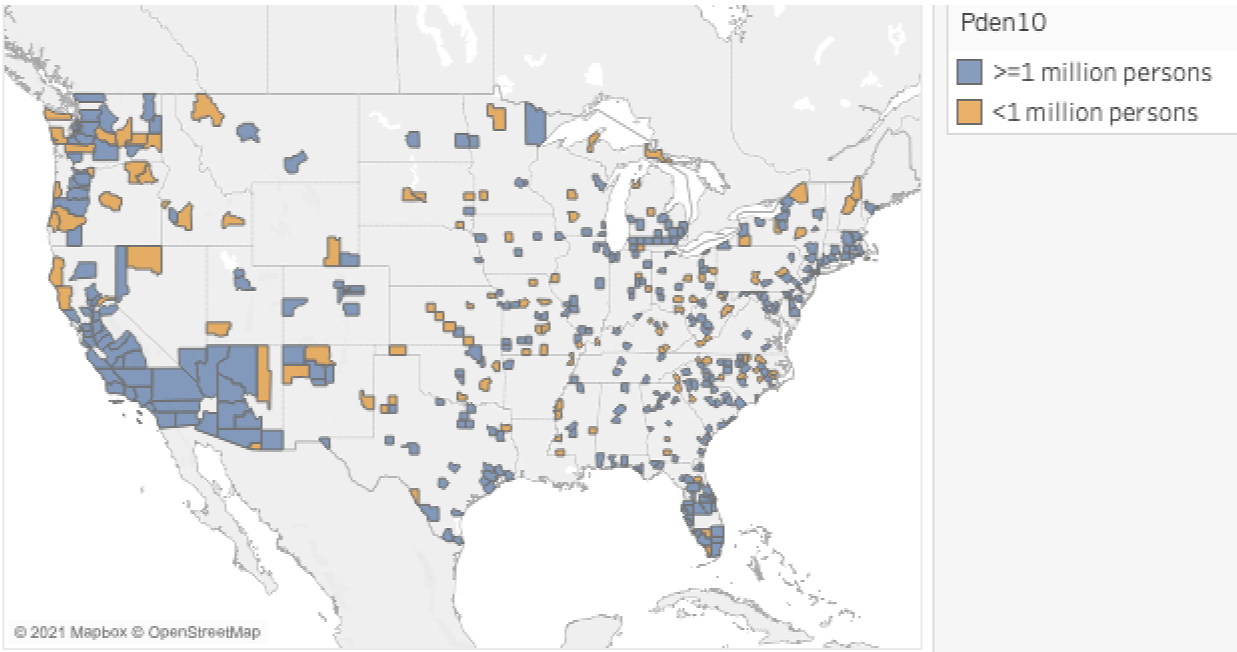


Figure 7 - The average number of withdrawn days by states

Average Number of Days before Withdraw by Population Density Type

>=1 million persons	963.8
<1 million persons	798.9

Figure 8 - Number of Counties with HPSA high score (>13) by Population Density Type

Number of County with High HPSA Score (>13) by Population Density Type

>=1 million persons	6,403
<1 million persons	383

VISUALIZATION

Our team used Tableau (Desktop Professional Edition 2021.1.5)[4], to create our visualizations (screenshots attached below). It also allowed us to publish our dashboard online so anyone can view and access the visualizations.

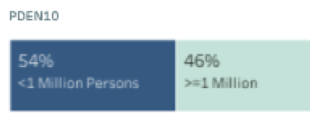
ANALYZING SUSCEPTIBILITY TO MENTAL HEALTH ISSUES AND COUNTY'S POPULATION DENSITY TYPE

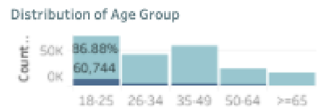
Mental health issues are pervasive and, now more than ever, need to be better understood to address their causes and impacts in a meaningful way. Approximately 40% of Americans live in a designated mental health provider shortage area, which exacerbates the problem. Across the US, each state has discretionary funding allocated specifically for mental health. Sufficient funds and effective resource allocation are necessary for the diagnosis and treatment of mental health issues.

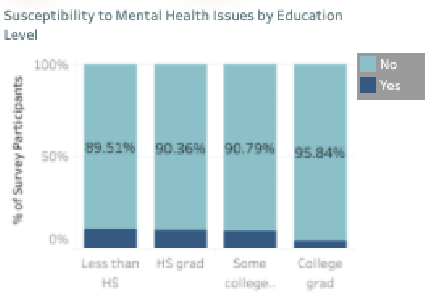
When untreated, mental health challenges can have a negative effect on a person's economic solvency, leading to increased rates of homelessness and poverty, social isolation, deteriorating physical health and shorter life expectancy, and decreased profitability for employers and their shareholders due to lower employee efficiency. Not only that, suicide rates are increasing in all age groups, yet, federal funding for mental health comprises less than 7% of the total U.S. health care spending.

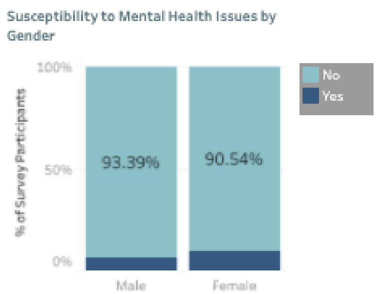
EXPLORATORY DATA ANALYSIS

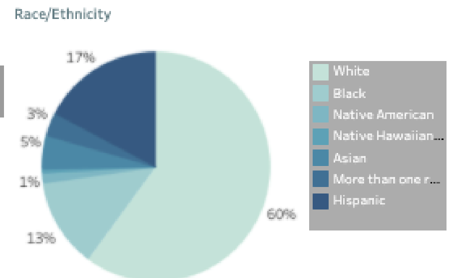
# Survey Participants	Male	Female
210,959	47%	53%

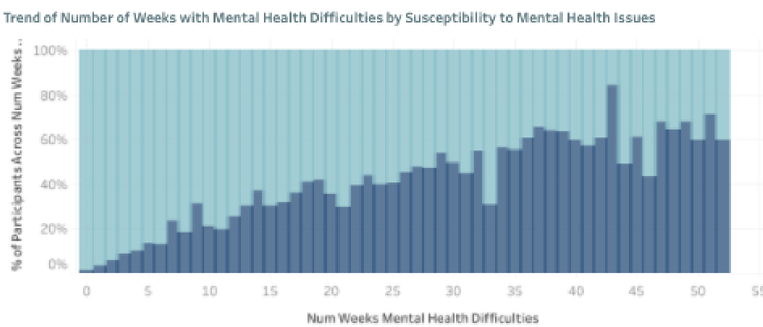


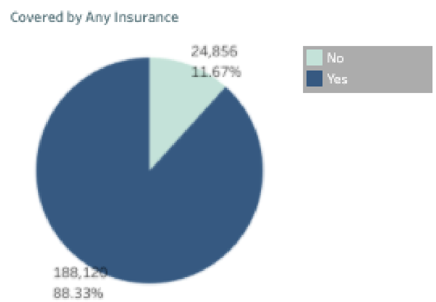






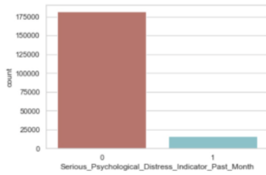






STATISTICAL ANALYSIS

Over-sampling using **SMOTE** (Synthetic Minority Oversampling Technique) by creating synthetic samples from the minor class (no-subscription) instead of creating copies.



A chi-squared statistical test was performed to find the relationship between county's population density type and the target variable on interested (susceptibility to mental health issues). These two variables are statistical significantly dependent with a p-value <0.01.

Table 2. Bivariate Associations Between Susceptibility to Mental Health Issues with County's Population Density Type in NSDUH Survey (2015-2019)

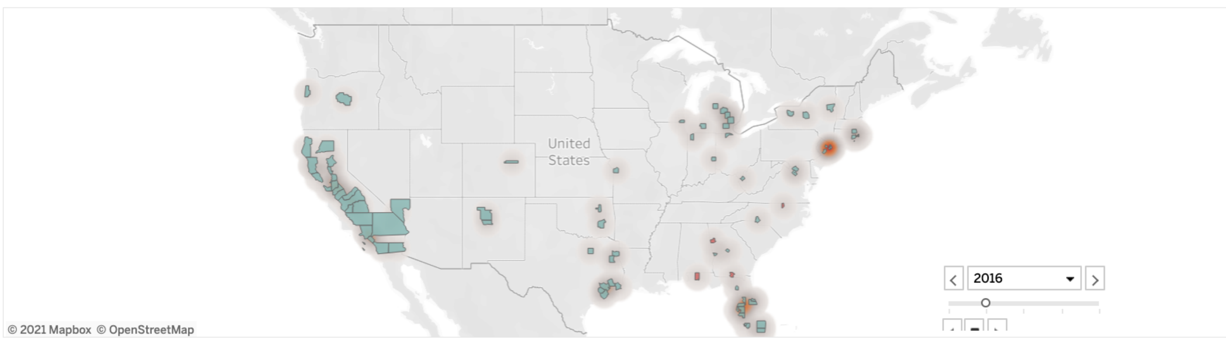
	High susceptibility to mental health issues (n = 16,078); n(%)	Low susceptibility to mental health issues (n= 181,505), n(%)	χ^2	p-value
Population Density Type			46.83	< 0.01
>=1 Million Persons	7,017 (44)	84,317 (46)	-	-
<1 Million Persons	9,061 (56)	97,188 (54)	-	-

Multivariate Logistic Regression was performed to understand interactions between variables selected from recursive feature selection and forward feature selection.

```
                           Logit Regression Results
==============================================================================================
Dep. Variable:     Serious_Psychological_Distress_Indicator_Past_Month   No. Observations:   197583
Model:                                                          Logit   Df Residuals:        197552
Method:                                                           MLE   Df Model:                30
Date:                                                Fri, 22 Oct 2021   Pseudo R-squ.:       0.5785
Time:                                                        19:50:56   Log-Likelihood:     -23494.
converged:                                                       True   LL-Null:            -55740.
Covariance Type:                                            nonrobust   LLR p-value:          0.000
==================================================================================================
                                                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
Year                                          -0.0037   3.52e-05   -106.381      0.000      -0.004      -0.004
Perceived_Unmet_Need                           0.5947      0.028     21.559      0.000       0.541       0.649
Worst_Psychological_Distress_Level             0.3715      0.003    139.619      0.000       0.366       0.377
Num_Days_Skipped_Work_Past_30_Days             0.0671      0.007     10.079      0.000       0.054       0.080
Treatment_Type_Past_Year_Inpatient_Only        0.4608      0.155      2.972      0.003       0.157       0.765
Race_Ethnicity_Native_American                -0.0726   9.69e+06  -7.49e-09      1.000    -1.9e+07     1.9e+07
Overall_Health_Unknown                        -0.3465        nan        nan        nan         nan         nan
Covered_By_Any_Health_Insurance_Imputation_Revised -0.0708  0.045    -1.569      0.117      -0.159       0.018
Gender_Male                                    0.1895      0.026      7.276      0.000       0.138       0.241
Age_Category_Six_Levels_26-34                 -0.2019      0.033     -6.128      0.000      -0.267      -0.137
Age_Category_Six_Levels_35-49                 -0.1999      0.034     -5.863      0.000      -0.267      -0.133
Age_Category_Six_Levels_50-64                 -0.1980      0.051     -3.874      0.000      -0.298      -0.098
Age_Category_Six_Levels_65_And_Above          -0.3432      0.075     -4.593      0.000      -0.490      -0.197
Race_Ethnicity_Black                           0.1270      0.041      3.110      0.002       0.047       0.207
Race_Ethnicity_Hispanic                       -0.0953        nan        nan        nan         nan         nan
Race_Ethnicity_Multiple_Races                 -0.0918      0.057     -1.622      0.105      -0.203       0.019
Race_Ethnicity_Native_American                -0.0726   9.66e+06  -7.52e-09      1.000   -1.89e+07    1.89e+07
Race_Ethnicity_Native_Hawaiian                 0.0810      0.162      0.501      0.616      -0.236       0.398
Race_Ethnicity_White                          -0.0985        nan        nan        nan         nan         nan
Education_Category_HS_Grad                     0.5248      0.040     13.002      0.000       0.446       0.604
Education_Category_Less_than_HS                0.6227      0.048     13.077      0.000       0.529       0.716
Education_Category_Some_College_Assoc          0.2459      0.036      6.812      0.000       0.175       0.317
Overall_Health_Fair_Poor                       0.9342      0.045     20.745      0.000       0.846       1.023
Overall_Health_Good                            0.4750      0.040     12.011      0.000       0.398       0.553
Overall_Health_Unknown                        -0.3465        nan        nan        nan         nan         nan
Overall_Health_Very_Good                       0.1361      0.039      3.455      0.001       0.059       0.213
Adult_Employment_Status_Employed_PartTime      0.0366      0.034      1.079      0.281      -0.030       0.103
Adult_Employment_Status_Other                  0.3047      0.033      9.278      0.000       0.240       0.369
Adult_Employment_Status_Unemployed             0.5636      0.045     12.641      0.000       0.476       0.651
Total_Income_Family_Recode_20000orLesser       0.0542      0.032      1.710      0.087      -0.008       0.116
Total_Income_Family_Recode_50000-75000        -0.0363      0.039     -0.935      0.350      -0.112       0.040
Total_Income_Family_Recode_75000orMore        -0.1497      0.035     -4.272      0.000      -0.218      -0.081
PDEN10_Less_than_1_Mil                        -0.0677      0.025     -2.708      0.007      -0.117      -0.019
==================================================================================================
```

Health Professional Shortage Areas & Awarded Mental Health Assistance

High HPSA Scored Areas (Bright Density Spot) vs. Received Mental Health Assistance (Map color)



Abbreviations: **CBSA**, Core Based Statistical Areas; **NSDUH**, National Survey of Drug Use and Health ; **HPSA**, Health Professional Shortage Areas; **PDEN**, Population Density; **AOR,** Adjusted Odds Ratio; **OR**,Unadjusted Odds Ratio

DATA ANALYSIS AND MODELING

Objective

To analyze the relationship between the susceptibility to mental health issues and exposure variables for those age 18 and above from 2015 to 2019 using the NSDUH survey. Our hypothesis is that the susceptibility to mental health issues will significantly differ across U.S. adults based on the Population Density type of their living areas between 2015 and 2019.

Data Collection

Data was obtained from the National Survey Drug Use and Health (NSDUH) Survey from 2015 to 2019, a national, cross-sectional population survey of adults 12 years of age and older who live in the United States. NSDUH contains data on the ongoing patterns in health needs, substance use, mental health, and other information including health behaviors. The survey was self-administered. For this project, we excluded individuals aged younger than 18. The final sample size contained 210, 959 unique records.

Measures

The dependent variable in this project was the susceptibility to mental health issues. The independent variable in this project was the population density type (PDEN10) of the living areas of the individuals. This was a binary variable (1 = >=1 million persons/2 = <1 million persons).

Sociodemographic variables include: age, gender, education level, race/ethnicity, employment status were controlled for in the multivariate analysis. All the variables above were treated as categorical variables: Gender (male/female); Total Income for Respondent/Family (<$10,000/$10,000 - $19,999/ $20,000-$29,999/ $30,000-$39,999/ $40,000-$49,999/$50,000-$74,999/>=$75,000); Education level (Less than HS/HS grad/Some college/Associate's degree/ College graduate/other); Race/Ethnicity (White/Black/ Native American/ Native Hawaiian/Pacific Islander/ Asian/ More than One Race/ Hispanic); Employment Status (Employed full-time/Employed part-time/ Unemployed/ Other/ Underage).

Data analysis was performed using Python (version 3.8.8) [5], using descriptive statistics, bivariate analysis, multivariate logistic regression.

Statistical Analysis

Descriptive Analysis

Table 1. Descriptive Statistics	
	n = 197,583 n (%)
Age Group	
18-25	65,025 (33)
26-34	40,828 (21)
35-49	52,137 (26)
50-64	22,814 (12)
>=65	16,779 (9)
Gender	
Male	91,869 (47)

Female	105,714 (54)
Race/Ethnicity	
White	115,554 (59)
Black	25,998 (13)
Hispanics	36,157 (18)
Other	19,874 (10)
Education Level	
Less than HS	25,264 (13)
HS graduate	50,978 (26)
Some college	66,371 (34)
College graduate	54,970 (28)
Overall Health	
Excellent	46,046 (23)
Very Good	74,475 (38)
Good	55,002 (28)
Fair/Poor	22,025 (11)
Total Household Income	
<$20,000	38,899 (20)
$20,000 - $49,999	60, 877 (31)
$50,000-$75,000	30,724 (16)
>=$75,000	67,083 (34)
Employment Status	
Employed full-time	103,066 (52)
Employed part-time	31,334 (16)
Unemployed	11,907 (6)
Other (including not in labor force)	51,276 (26)

Population Density Type	
1 (>=1 million persons)	91,334 (46)
2 (<1 million persons)	106,249 (54)

Bivariate Analysis

Chi-square statistical test was performed to find the relationship between the independent variable of interest (PDEN10) and the target variable of interest.

Table 2. Bivariate Associations Between Susceptibility to Mental Health Issues with County's Population Density Type in NSDUH Survey (2015-2019)				
	High susceptibility to mental health issues (n = 16,078); n(%)	**Low susceptibility to mental health issues (n= 181,505), n(%)**	χ^2	**p-value**
Population Density Type			46.83	< 0.01
>=1 Million Persons	7,017 (44)	84,317 (46)	-	-
<1 Million Persons	9,061 (56)	97,188 (54)	-	-

Multivariate Analysis

Multivariate Logistic Regression is utilized as an explanatory model to find the key drivers of our binary outcome variable: Experienced Serious Psychological Distress in the Past Month. Adjusted odds ratios (AORs) with 95% confidence intervals (CIs) were computed.

Outcome Variable

The outcome variable is a binary indicator of 1 for having experienced serious psychological distress in the past month and 0 for not. Our dataset contains 181,505 responses for 0 and 16,078 responses for 1 (excluding rows with null values). Therefore, all test accuracy metrics should be benchmarked against our baseline 181,505/197,583 = 91.862%.

Modeling Building

First, we exclude variables from areas, such as Substance Use Disorder, to only include mental health related, sociodemographics, and use of healthcare facility variables. We computed the VIFs of all features and drop variables with a high p-value and a high VIF (threshold = 10) . After re-fitting the model with the new set of features, we checked for statistically significant variables. If desirable, we proceeded with model evaluation.

Then, we used sklearn.linear_model's LogisticRegression() in Python to create logistic regression models using features selected from Recursive Elimination Method (k=10 and 20) and Forward Feature Selection Method. To control sociodemographic variables, they are later added to the selected features for our final modeling.

Table 3. Logistic Regression with Different Features and Their Respective R^2 Scores and Test Accuracy Score				
Dataset without HPSA and HRSA Awarded Grant Features				
Model	**Number of Features (k)**	**Features Selected**	**R^2**	**Test accuracy**
Recursive Elimination	12	('Inpatient_Past_Year', 'Serious_Psychological_Distress_Indicator_Past_Year', 'Mild_Mental_Illness_Indicator_Past_Year', 'Treatment_Type_Past_Year_Inp_Outp_PrescMed', 'Treatment_Type_Past_Year_Outpatient_Only', 'Treatment_Type_Past_Year_Outpatient_PrescMed', 'Treatment_Type_Past_Year_PrescMed_Only', 'Treatment_Type_Past_Year_Skipped', 'Treatment_Type_Past_Year_Unknown', 'Race_Ethnicity_White', 'Adult_Employment_Status_Employed_PartTime', 'Poverty_Level_In_Poverty', 'Poverty_Level_More_than_2x_Poverty', 'Poverty_Level_Up_to_2x_Poverty', 'PDEN10_Less_than_1_Mil')	0.6218	0.9351 *1.648% above baseline*
Forward ***Final Model***	**20**	**('Overall_Health_Fair_Poor', 'Education_Category_Less_than_HS', 'Perceived_Unmet_Need', 'Adult_Employment_Status_Unemployed', 'Education_Category_HS_Grad', 'Overall_Health_Good', 'Worst_Psychological_Distress_Level', 'Adult_Employment_Status_Other', 'Education_Category_Some_College_Assoc', 'Gender_Male', 'Num_Days_Skipped_Work_Past_30_Days', 'Year',**	**0.5797**	**0.9426** ***2.398% above baseline***

		'Total_Income_Family_Recode_75000orMore', 'Age_Category_Six_Levels_50-64', 'Age_Category_Six_Levels_35-49', 'Age_Category_Six_Levels_26-34', 'Age_Category_Six_Levels_65_And_Above', 'Overall_Health_Very_Good', 'Race_Ethnicity_Black', 'Treatment_Type_Past_Year_Inpatient_Only', 'PDEN10_Less_than_1_Mil')		
Dataset with HPSA ang HRSA Awarded Grant Features				
Model	**Number of Features (k)**	**Features Selected**	**Pseudo-R^2**	**Testing accuracy**
Logistic Regression	15	('Age_Category_Six_Levels', 'Race_Ethnicity','Education_Category', 'Overall_Health', 'Total_Income_Respondent', 'Adult_Employment_Status', 'PDEN10', "Total Active Grant Financial Assistance median", "('HPSA Score', 'median')", 'Has_Medicaid_Or_CHIP','Poverty_Level', 'Received_Treatment_At_Private_Therapist', 'Received_Treatment_At_School', 'Received_Treatment_At_NonClinic_Doctor', 'Num_Days_Skipped_Work_Past_30_Days', 'No_Treatment_Could_Not_Afford')	0.2118	*0.9211* *0.248% above baseline*
Logistic Regression	18	('Gender','Age_Category_Six_Levels','Race_Ethnicity',' Education_Category','Overall_Health', 'Total_Income_Respondent','Adult_Employment_Stat us','Covered_By_Any_Health_Insurance', 'PDEN10', ('HPSA Score', 'median')", 'Has_Medicaid_Or_CHIP','Has_Medicare', 'Poverty_Level', 'Received_Treatment_At_Private_Therapist', 'Num_Days_Skipped_Work_Past_30_Days', 'EAP_Offered','Work_Situation_Past_Week', 'No_Treatment_Where_To_Go', 'No_Treatment_Could_Not_Afford', 'Psychological_Distress_Level_Worst_Month')	0.314	*0.929* *1.08% above baseline*
Logistic Regression	17	('Age_Category_Six_Levels', 'Race_Ethnicity', 'Education_Category', 'Overall_Health', 'Total_Income_Respondent', 'Adult_Employment_Status', 'PDEN10', "Total Active Grant Financial Assistance median", "('HPSA Score',	0.3310	*0.9187* *0.008% above baseline*

		'median')", 'Has_Medicaid_Or_CHIP','Poverty_Level', 'Received_Treatment_At_Private_Therapist','Received_Treatment_At_School', 'Received_Treatment_At_NonClinic_Doctor','Num_Days_Skipped_Work_Past_30_Days', 'No_Treatment_Could_Not_Afford')		

Train test validation from the best model bolded above:

Confusion matrix	
63536	1454
2607	3190
Test accuracy = 0.9426307090284939	

RESULTS AND CONCLUSION

Descriptive statistics are summarized in **Table 1.** The mode age group of the sample was 18 - 25 years old. The majority of the participants were female (54%), White (59%), and had some college education (34%). In addition, most of the participants reported living in areas that have a population density of less than 1 million persons (54%). **Table 2** summarized the bivariate difference between susceptibility to mental health issues and the population density type. Participants from areas that have a population density of less than 1 million persons were more susceptible to mental health issues ($p < 0.01$). Other than that, the odds of being susceptible to mental health issues was 0.2121 (95% CI = 0.2098, 0.2145) ($p<0.01$) time significant [lower] among individuals who came from areas that have less than 1 million persons compared to individuals who came from areas that have more than or equal to 1 million persons.

Recursive Elimination and Forward Selection methods were used to select features and **Table 3** summarized the feature selection methods, selected features, and their respective pseudo-R^2 scores from logistic regression and test accuracy scores. **Table 4** summarized the final logistic regression model for the susceptibility to mental health issues. Population density type was associated with susceptibility to mental health issues after controlling for sociodemographic, treatment type, and mental illness indicator variables (see controlling variables listed in **Table 4**). These indicators (see below) that were included in the Forward Feature Selection Method were dropped because they showed statistical insignificance when fitted in the model.

Dropped variables: ['Moderate_Mental_Illness_Indicator_Past_Year', 'Low_Or_Moderate_Mental_Illness_Indicator_Past_Year', 'Low_Or_Moderate_Mental_Illness_Indicator_Past_Year', 'Race_Ethnicity_Black', 'Race_Ethnicity_Hispanic', 'Race_Ethnicity_Multiple_Races', 'Race_Ethnicity_Native_American']

Table 4. Multivariate Logistic Regression Analysis Between Susceptibility to Mental Health Issues with Population Density Type, and Covariates in NSDUH Survey Participants (2015-2019)

Outcome: Susceptibility to Mental Health Issues	Coefficient	Adjusted Odds Ratio
Overall_Health_Fair_Poor	0.9477***	2.58
Education_Category_Less_than_HS	0.6343***	1.89
Perceived_Unmet_Need	0.5933***	1.81
Adult_Employment_Status_Unemployed	0.5616***	1.75
Education_Category_HS_Grad	0.5268***	1.69
Overall_Health_Good	0.4836***	1.62
Treatment_Type_Past_Year_Inpatient_Only	0.4594***	1.58
Worst_Psychological_Distress_Level	0.3719***	1.45
Adult_Employment_Status_Other	0.3012***	1.35
Education_Category_Some_College_Assoc	0.2479***	1.28
Race_Ethnicity_Black	0.2352***	1.27
Gender_Male	0.1904***	1.21
Overall_Health_Very_Good	0.1414***	1.15
Num_Days_Skipped_Work_Past_30_Days	0.0656***	1.07
Year	-0.0038***	1.00
PDEN10_Less_than_1_Mil	-0.0552**	0.95
Total_Income_Family_Recode_75000orMore	-0.1595****	0.85
Age_Category_Six_Levels_35-49	-0.228***	0.80
Age_Category_Six_Levels_26-34	-0.2291***	0.80
Age_Category_Six_Levels_50-64	-0.2628***	0.77
Age_Category_Six_Levels_65_And_Above	-0.3817***	0.68

*** $P \leq 0.001$
** $P \leq 0.01$

Interpretations

1. Overall perceived health is a key driver in associated psychological distress. When an individual reports fair or poor health, (s)he is 2.58 times more likely to have faced serious psychological distress in the past month compared to an individual who did not report fair/poor health, holding all else constant.
2. Education level has a strong relationship with the outcome variable. Individuals that have less than a high school level of education are 1.8 times more likely to have faced serious psychological distress in the past month than those that have more than a high school level of education, holding all else constant. In addition, individuals with a high school level of education are 1.6 times more likely to have faced serious psychological distress in the past month. Lastly, individuals with some college level of the association have 1.2 times more likely to have faced serious psychological distress in the past month.
3. Individuals that report a perceived unmet need for mental health treatment or counseling are 1.81 times more likely to have faced serious psychological distress in the past month than those that don't have this perception, holding all else constant.
4. The older an individual is, the lesser the likelihood they are to have faced psychological distress in the past month with individuals ages 65+ being 23% less likely to have faced psychological distress in the past month.

LIMITATIONS AND RECOMMENDATION

The original NSDUH dataset contained specific geographic location information on the individual level but was later aggregated/grouped into a 3-level categorical variable. Therefore, the aggregated/grouped data we created using the HPSA and HRSA Grant data could not be related to the NSDUH at the individual level. If more specific geographic information of the participants was available, we would possibly find some associations with our target variable.

Other methods to enhance R^2 score include (1) further feature engineering, (2) adding additional data points, and (3) explore different modeling methods. Creating complementary models would help deepen understanding of certain areas of interest. Rigorous methods should also be performed to adjust for confounding.

REFERENCES

1. Open Minds. (2020). The U.S. Mental Health Market: $225.1 Billion In Spending In 2019: An *OPEN MINDS* Market Intelligence Report. Retrieved From https://openminds.com/intelligence-report/the-u-s-mental-health-market-225-1-billion-in-spending-in-2019-an-open-minds-market-intelligence-report/.
2. Rebecca E.S., David S.M., and Rinad S.B. (2021). Lessons from Maslow: Prioritizing Funding to Improve the Quality of Community Mental Health and Substance Use Services. Retrieved from https://ps.psychiatryonline.org/doi/full/10.1176/appi.ps.202000209.
3. World Health Organization. (2003). Mental Health Policy and Service Guidance Package. Retrieved from https://www.who.int/mental_health/policy/services/6_financing_WEB_07.pdf.
4. Tableau Visualization Software. Tableau Desktop Professional Edition, version (2021.1.5). Available at https://www.tableau.com.
5. Python Software Foundation. Python Language Reference, version 3.8.8. Available at https://www.pythog.org.