Mental Health- Building a Model to Determine Patients with a High Risk for an Emergency Room Visit

Jill Book, Michaela Nolan, & Pengcheng Wu
HINF6400 Intro to Health Data Analytics
Holly Jimison

12/3/2019

Mental Health- Building a Model to Determine Patients with a High Risk for an Emergency Room Visit

I. Introduction

It is estimated that one in five American adults are living with a Mental Health disorder ("Mental Health By the Numbers | NAMI: National Alliance on Mental Illness," n.d.). When these disorders are not managed properly, they can have an impact on the quality of life of those living with the disease ("Primary Care, Behavioral Health, and Public Health: Partners in Reducing Me...: EBSCOhost," n.d.). They also have an impact on the community and local hospitals who have seen an increase in the amount of emergency room visits for mental health patients. In this paper we plan to utilize the tools and theories we have developed by taking Health Data Analytics to analyze a New York State Mental Health Patients data set. Using these data we developed a model using logistic regression that can predict and flag patients who are likely to utilize emergency mental health services. Hospital leaders would benefit from this model by being able to closely monitor and provide additional preventative care support to the high risk patients that are flagged by the model.

II. Background and Motivation

Background:

The mission of the New York State Office of Mental Health (NYS OMH) is to promote the mental health of all New Yorkers, with a particular focus on providing hope and recovery for adults diagnosed with serious mental illness and children diagnosed with serious emotional disturbance ("About OMH," n.d.). The NYS OMH vision focuses on community based and recovery oriented services, consistent with the science of mental health evaluation, diagnosis, and treatment. NYS OMH strives to achieve its mission and vision by leveraging its role as New York State's lead authority of the public mental health system, specifically through its dual responsibilities of 1) setting policy and providing funding for community services, and 2) providing direct inpatient and outpatient services.

The Patient Characteristics Survey (PCS) provides a comprehensive one–week "snapshot" of the population served by New York State's public mental health system ("NYS Patient Characteristics Survey (PCS): 2015 | Kaggle," n.d.). The NYS OMH requires all programs that it licenses or funds (directly or indirectly) to complete the Patient Characteristics Survey. The PCS is conducted every other year, and collects demographic, clinical, and service-related information for each client who receives a mental health service during the one-week period. The 2015 PCS collected data from approximately 4,000 mental health programs providing direct services to approximately 180,000 clients during the survey week. For 2015, this equates to an estimated 1% of the NYS general population.

Our data set contains demographic, clinical, social, and insurance characteristics for each client served by the NYS public mental health system during the week of October 19, 2015. All mental health programs licensed or funded (directly or indirectly) by the NYS Office of Mental Health were required to report. The PCS data were submitted by these providers electronically into a web application via direct entry, file upload, and/or import from other OMH data systems.

Motivation:

Mental health is integral to living a healthy, balanced life. Mental health and physical health together determine a person's health status. According to the National Alliance of Mental Illness (NAMI), nearly one in five U.S. adults live with a mental illness (46.6 million in 2017) ("Mental Health By the Numbers | NAMI: National Alliance on Mental Illness," n.d.). Emotional and mental health is important because it's a vital part of people's lives and impacts their thoughts, behaviors and emotions. Being healthy emotionally can promote productivity and effectiveness in activities like work, school or life. It plays an important part in the health of people's relationships and allows individuals to adapt to changes in their life and cope with adversity.

As the United States shifts to value based care and the Accountable Care Organization model, providers are incentivized to keep patients healthy and coordinate care more efficiently for patients with chronic, complex conditions (Maust, Oslin, & Marcus, 2013). Since the mental health population contributes greatly to the high cost of healthcare, providing preventative care and targeting high risk patients could help keep patients out of the hospital and ultimately reduce costs.

Literature Review:

In 2013, it was reported that New York State had about 250,000 citizens on Medicaid receiving Social Security benefits due to a mental disability. In 2009 the state reported 7 billion dollars in medicaid expenditures on patients with behavioral health and mental health diagnoses (Smith, Erlich et al. 2013). Also in 2013, an evaluation of Personal Health Care Spending in the US ranked Mental and substance abuse disorders fourth most expensive condition category (behind Cardiovascular disease, Diabetes, urogenital, blood, and endocrine diseases, and other noncommunicable diseases). 187.8 billion dollars were spent that year on mental and substance abuse disorders which reflects a 3.7% change from 1996 to 2003(Murray et al., 2016).

Additionally, an analysis performed on 2009 Medicaid claims data on two Western New York State counties examined the relationship between different behavioral health diagnosis and ED utilization. The researchers found that many mental health and patient characteristics, including psychiatric disorders, smoking, and substance abuse, led to frequent ED usage. The researchers concluded their study by emphasizing the need to further research this patient population to understand how smoking, psychiatric disorders, substance abuse, and other mitigating factors impact their care and treatment (Castner, Wu, Mehrok, Gadre, & Hewner, 2015). Also a 2010 study that focused on Veterans Health Administration (VHA) took a closer look at emergency department usage. The researchers found that psychiatric illnesses were common diagnoses for patients who visited the ED frequently. Of the 617 patients with over 25 ED visits per year, 566 (91.7%) patients had a mental health disorder (defined as a psychiatric or substance abuse) (Doran, Raven, & Rosenheck, 2013).

Knowledge Gaps:

According to the literature, many children with Mental Health issues do not receive appropriate services and care. A survey from the National Survey of American Families reported that only 21% of children who need mental health care/services actually receive treatment. (Bringewatt & Gershoff, 2010). This gap in medical care is actively being researched by Mental Health leaders to try and expand services so that all children receive necessary mental health support. Additionally the literature revealed that most people do not have well established mental health literacy and are unable to understand signs of mental health distress. Mental Health advocates urge that all community members must increase their mental health literacy to understand better those with mental health diagnoses. (Jorm 2000)

Stakeholder Analysis:

There are a few different groups of stakeholders who would benefit from the analysis of this mental health data set:

1. **New York State Hospital Leadership:** Hospital leaders in New York State would benefit from taking a closer look at the Mental Health Survey results- especially related to the emergency room visit data. Patients who seek care in the emergency room are usually complex cases that drive up costs. They would be very interested in using a model that can predict which patients are at a high risk for a complicated emergency visit. The hospital leaders would be able to put into place policies and procedures to better support these high risk patients and in turn reduce the emergency room visits from mental health patients.

2. **New York State Office of Mental Health and its affiliated mental service providers:** These stakeholders from the Office of Mental Health of New York State would also benefit from taking a closer look at the Mental Health Survey results.Taking a look at the model and emergency room visit data across the state can help the officials and mental health providers

better understand the mental health patient population. It will help them to understand which areas of New York State need additional support and can use the model and other results of the analysis to try and get more monetary support from the state for mental health patients.

3. **New York State Insurance Companies:** Insurance Companies would benefit from the analysis of the Mental Health survey so they could better understand their policyholders with mental health disorders. Especially insurance companies who have switched to reimbursing for quality measures and keeping patients healthy. These data could help them understand the complexity of patients with mental health disorders and their impact on healthcare costs.

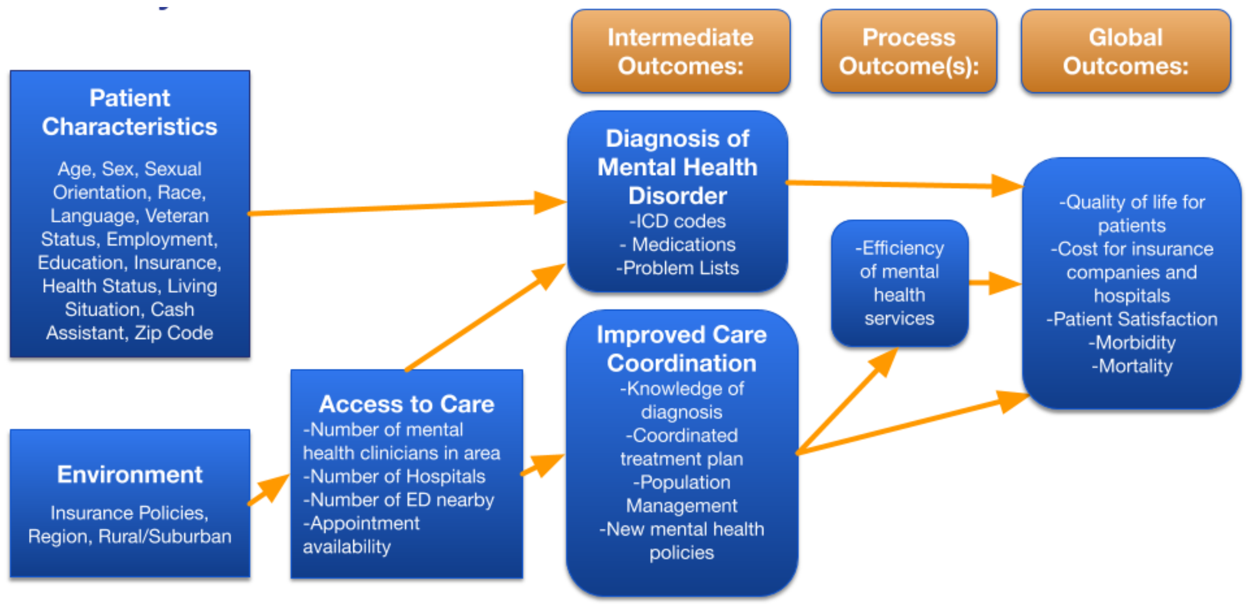
4. **All New York State Residents, Especially Those with Mental Health Disorders:** All New York State residents would benefit from the survey results and analysis to better understand the impact mental health disorders have on their neighbors and fellow residents. Mental Health disorders impact the person with the diagnosis but also the community where they live. Increasing awareness of the impacts of a mental health disorder can help diminish the stigma related to the diagnosis.

Possible Bias & Limitations with Data:

Prior to completing the analysis, there are important possible bias and limitations to note:

1. **Observational Bias-** Provider reported survey data (versus patient reported survey data) and since the reported information is second hand there is a potential for bias responses.
2. **Sampling Bias -** the survey was conducted only over one week (October 19, 2019). The results/sample size may have been different if another time was chosen.
3. **Narrow Scope-** This data set only includes patients at NY state mental hospitals, so results may not be applicable to the general population.
4. **Unknown Responses-** There is no missing/empty data, but many responses are "unknown," which could be considered missing.
5. **Accuracy of Survey Responses-**
    - Some respondents may hide some private disease situations for some personal reasons
    - Some respondents may not know they have the mental disease since it's difficult to be diagnosed
    - Respondents may not be able to accurately state their health status and illness when they fill out the questionnaire because they do not have professional medical knowledge.
6. **External Factors-** When data collectors organize and clean health data, it is easy to ignore the impact of uncontrollable external factors on data accuracy because there is no reliable demographic and sociological knowledge.
7. **Need for medical/clinical knowledge-** Data analysts cannot find the key to the problem quickly and accurately when they do not have rich clinical and scientific experience. They need professional medical researchers to guide the research direction, explore data with researchers with medical backgrounds, and may need to seek medical help when interpreting the results of the data analysis.

Mental Health Analytic Framework:



The mental health analytic framework (*depicted above*) outlines the causes and effects of a mental health disorder diagnosis. It is important for stakeholders and analysts to understand that various patient characteristics, environmental factors, and access to care impact a patient's mental health diagnosis. It is also important to understand that access to important medical care can also lead to improved care coordination between patients, providers, insurance companies, and the community. This improved care coordination can lead to an improved efficiency of mental health services and improved outcomes for the patient. These important outcomes include quality of life, cost for insurance companies and hospitals, patient satisfaction, morbidity, and mortality. According to this framework, stakeholders and other healthcare leaders would improve the lives of their mental health patients by providing support to the diagnosis of mental health, increasing the access to medical care, and improving care coordination.

III. Methods

Focus of the Analysis:

The focus of this analysis on the New York State Mental Health Survey data set is to develop a model built using demographic and diagnostic data related to mental health patients to help predict/flag patients who are likely to be admitted for emergency mental health services. The results of this model could be leveraged by stakeholders to implement better policies and procedures at New York State hospitals so mental health patients can receive well-coordinated preventative and emergency medical care.

The data collected are from the community of patients with mental illnesses, however, it provides much potentially valuable health information of respondents. For instance, hyperlipidemia, diabetes, obesity, liver disease, cancer, heart attack, etc.. The information can assist stakeholders in relating mental problems with other relevant clinical diseases in order to help improve current combination therapy. Usually, the common way to study mental health data is limited in clinical perspective. In this work, we will try to analyze the data, considering patients' background information and environment information to discover some amazing phenomena for further research.

Population:

The data being used in this analysis comes exclusively from the New York State Office of Mental Health (OMH) and includes information on 179,096 patients who visited or were admitted to NYS psychiatric hospitals over the course of one week in 2015. Because of the limitations of the sample population, results from this research likely can not be extrapolated to the general population, and will be mainly relevant to the NYS psychiatric hospitals and the NYS OMH.

A summary of the demographic breakdown of the population included in our sample data is as follows:

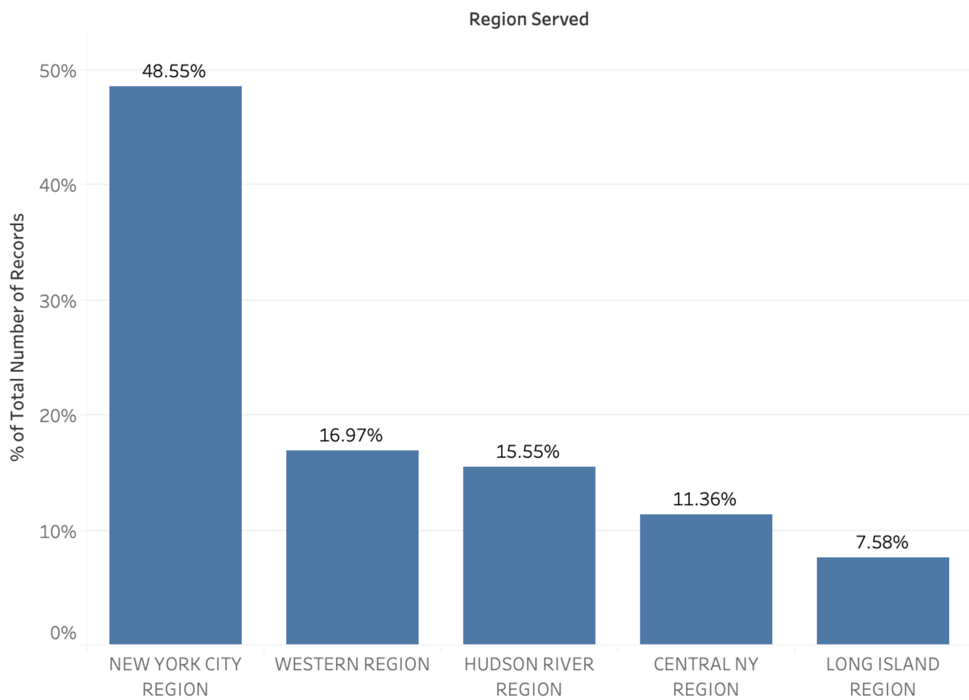
- 80% adults, 20% children
- 50.2% female, 49.6% male, 0.7% transgender
- 5.8% LGBTQ+, 81% heterosexual
- 23% Hispanic
- 26.1% Black, 51.5% White, 2.6% multi-racial, 16.5% other races
- 89.3% English-speaking, 7.1% Spanish-speaking, 2.9% other languages
- 2.5% veterans
- 14.9% employed, 71.2% unemployed and not looking for work, 9.8% unemployed and looking for work
- 13.5% college or graduate degree, 13.4% some college, 54.4% middle school to high school education, 8.7% pre-kindergarten to fifth grade education, 0.4% no formal education
- 92.9% have a diagnosed mental illness, 5.7% do not have a diagnosed mental illness
- 53.2% have one or more chronic medical conditions, 37.2% do not have a chronic medical condition
- 27.4% smokers, 65.6% non-smokers
- 91.2 % insured, 6% uninsured
- 11.4% Central New York region, 15.5% Hudson River region, 7.6% Long Island region, 48.5% New York City region, 17% Western region

- 2.9% experiencing homelessness

Exploratory Data Visualizations:

Below are few data displays and visualizations for key variables in the data set:

1. Percent of Total Number of Records and Region of New York State:



From figure 1 (*above*), it can be seen that the majority of records in the data set comes from the New York City region.

2. **Percent of Total Number of Records and Age Group:**

Age Group	
ADULT	79.96%
CHILD	20.03%
UNKNOWN	0.01%

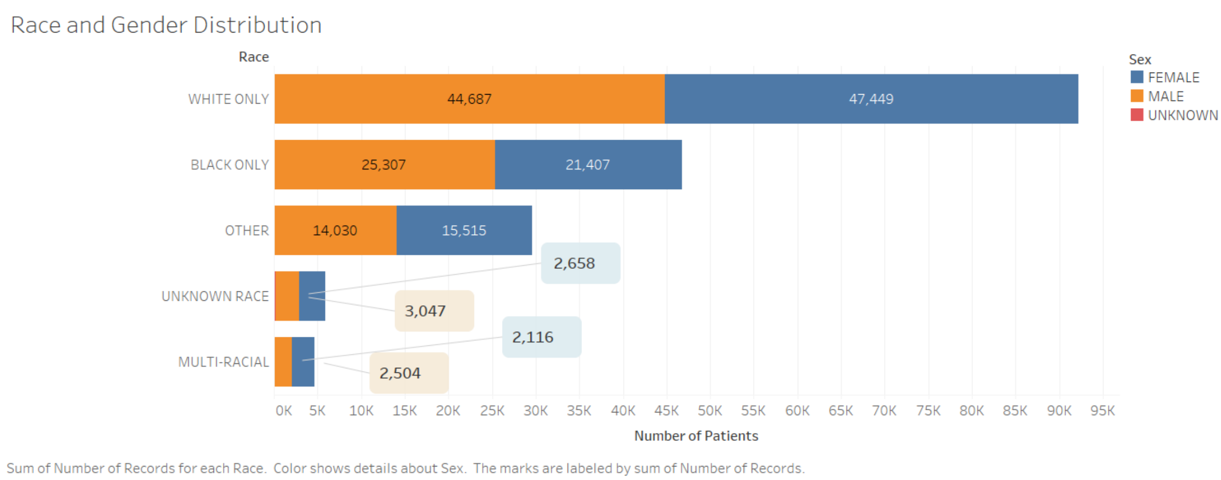
From figure 2 (*above*), it can be seen that a majority of our data set consists of data from adult mental health patients.

3. **Percent of Total Number of Records and Education Level:**

Education Status	
NO FORMAL EDUCATION	0.42%
PRE-K TO FIFTH GRADE	8.70%
MIDDLE SCHOOL TO HIGH SCHOOL	54.36%
SOME COLLEGE	13.45%
COLLEGE OR GRADUATE DEGREE	13.45%
OTHER	1.67%
UNKNOWN	7.96%

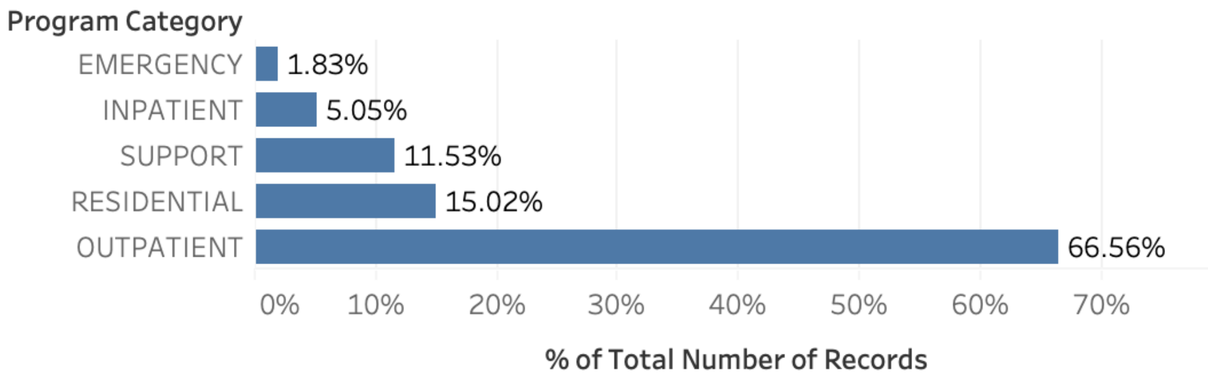
From figure 3 (*above*), it can be seen that in our data set patient population 63.48% of patients have a high school education or less.

4. **Race and Gender Distribution**



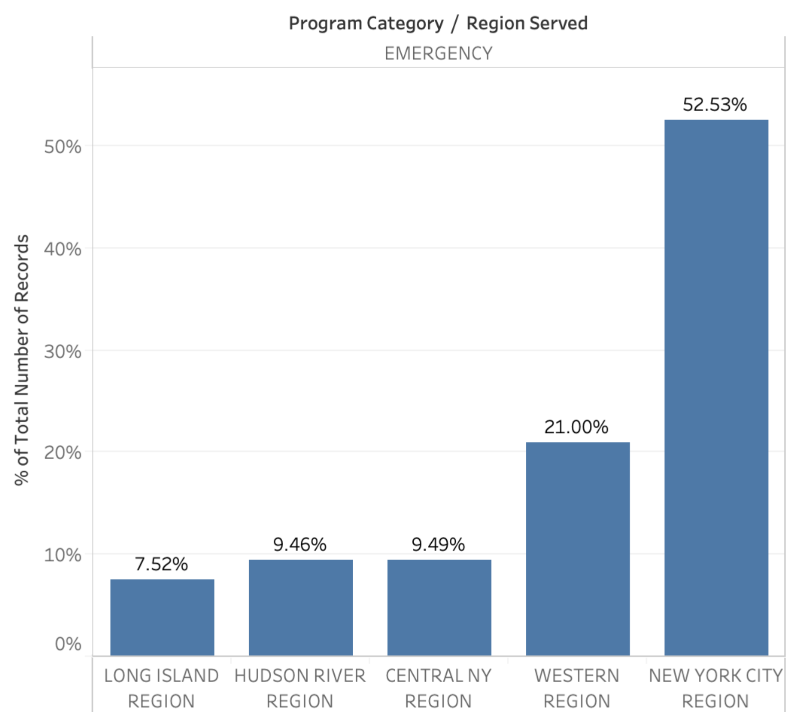
From figure 4 (*above*), it can be seen that in our data set patient population that the majority of patients are white. We are also able to see that across different races, the gender distribution remains consistent (*close to 1:1*).

5. **Percent of Total Number of Records and Program Category/Visit Type:**



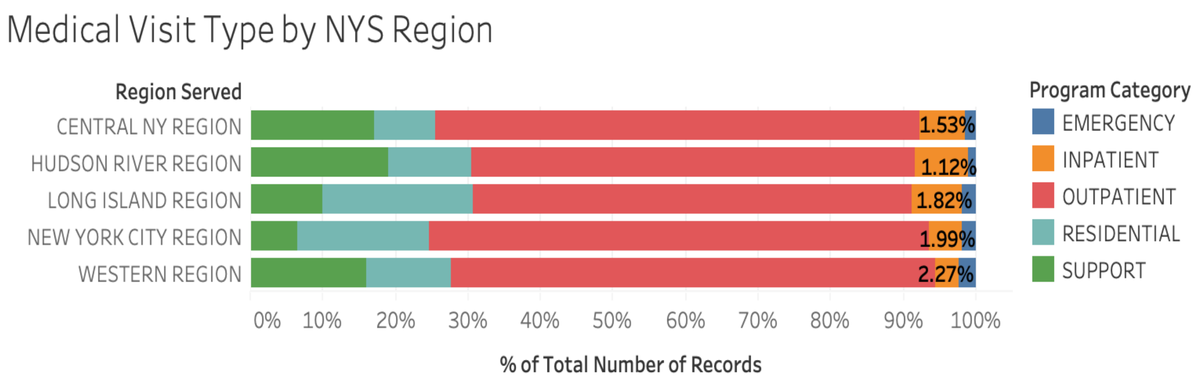
From figure 5 (*above*), it can be seen that the majority, 66.56%, of the records in the data set are related to outpatient services whereas only 1.83% of records are related to emergency services.

6. Percent of Total of Emergency Service Records and Region of New York State



From figure 6 (*above*), it can be seen that for just the Emergency Service related records in the data set, New York City had the largest percentage of total records/visits, 52.53%. This makes sense since as we saw in Figure 1, New York State had the largest percentage of total records.

7. Percent of Emergency Service Records and Region of New York State



From figure 7 (*above*), it can be seen that by looking at the percent of emergency services by the region of New York State that the rates are consistent across the regions (ranging from 1.12-2.27%). Interestingly, Western New York had the highest percentage of emergency services (2.27%) which we did not recognize when looking at the percent of total emergency service broken down by region in Figure 6.

Data Description:

The data set contains 67 variables, almost all of which are nominal data, such as employment status, race, ethnicity, gender identity, and presence of various chronic illnesses. There are also two columns of ordinal data: number of hours worked per week (given in ranges of hours) and education status. There are no continuous variables used in this data set.

The dependent variable we have chosen to look at is the Program Category variable, which describes the type of visit the patient made to the hospital. In this data set, 1.8% of patients were seen for emergency visits, and all other visits were classified as inpatient, outpatient, residential, or support.

Analysis: Chi-Squared Test

First, multiple chi-squared tests are carried out in Stata and Python to verify the rationality of variables the team will choose to analyze. Chi-squared test is commonly used for testing relationships between categorical variables. Considering most variables in this data set are nominal data, and the sample size is large enough to guarantee reliability, chi-squared test is one of the best statistics methods for preheating the analysis..

In the initial work, all the variables whose answers are “YES” or “NO” were extracted from the original data set to reflect respondents' characteristics in both health status and aspects of society and environment. They mainly includes symptom variables like Intellectual Disability, Autism Spectrum, Developmental Disability, Alcohol Related Disorder, Drug Substance Disorder, Mobility Impairment Disorder, Hearing Visual Impairment, High Blood Pressure, Diabetes, Obesity, Heart Attack, Stroke, Pulmonary Asthma, Alzheimer or Dementia, Endocrine Condition, Neurological Condition, Traumatic Brain Injury, Cancer, Smokers society characters like Region, Age, Sex, Transgender, Sexual Orientation, and environment variables like financial assistantship and health plan..

Since one of the team's research questions is Can analyzing demographic and diagnostic data related to mental health patients help predict/flag patients who are likely to be admitted for emergency mental health services? The values of program category were merged to two types, emergency and nonemergency. Although in this data set, number of samples makes frequency result not a problem to worry about, it's still meaningful to merge each variable's values into less types. For example, in the Sexual Orientation variable, patients were divided into bisexual, decline to answer, lesbian or gay, straight or heterosexual, unknown and other. Actually, bisexual, lesbian, gay and other can be categoried as LGBTQ+ type.

According to the test results, many of the variables above are related with the programs patients go to. Age group, based on the p-value returned from the software, indicated a nonsignificant difference among patients in emergency program and nonemergency programs. Transgender, Hispanic Ethnicity, Veteran Status did it, too. For symptom variables, most diseases should be considered potential factors to determine patients' program types. The unrelated diseases are Intellectual Disability, Autism Spectrum, Other Developmental Disability, Alzheimer or Dementia, Liver Disease, Traumatic Brain Injury, and Other Chronic Med Condition. This discovery provided sufficient evidence to the team's research question and inspired them to build a model for patient program type prediction/ classification in next step.

```
. tabulate agegroup programcategory, chi2
```

	Program Category		
Age Group	EMERGENCY	NONEMER..	Total
ADULT	2,662	140,551	143,213
CHILD	624	35,241	35,865
UNKNOWN	0	18	18
Total	3,286	175,810	179,096

Pearson chi2(2) = 2.5882 Pr = 0.274

```
. tabulate sexualorientation programcategory, chi2
```

Sexual Orientation	Program Category EMERGENCY	NONEMER..	Total
LESBIAN OR GAY	**130**	**10,167**	**10,297**
STRAIGHT OR HETEROS..	**2,247**	**142,758**	**145,005**
Total	**2,377**	**152,925**	**155,302**

Pearson chi2(**1**) = **5.2580** Pr = **0.022**

	P_value	Adjusted_p_value	Rejection
Unknown Insurance Coverage	1.576505e-194	8.197826e-193	True
Drug Substance Disorder	1.476089e-120	3.837833e-119	True
Alcohol Related Disorder	3.492363e-94	6.053429e-93	True
Received Smoking Medication	9.495076e-86	1.234360e-84	True
Unknown Chronic Med Condition	4.847614e-79	5.041518e-78	True
Mental Illness	2.035683e-53	1.764259e-52	True
No Insurance	1.184538e-45	8.799426e-45	True
Medicaid Insurance	3.019674e-45	1.962788e-44	True
No Chronic Med Condition	1.685252e-37	9.737010e-37	True
Hyperlipidemia	7.667315e-37	3.987004e-36	True
SSI Cash Assistance	1.308736e-35	6.186751e-35	True
Medicaid and Medicare Insurance	3.121529e-35	1.352663e-34	True
Obesity	1.941719e-34	7.766874e-34	True
Serious Mental Illness	4.849086e-28	1.801089e-27	True
Joint Disease	2.800098e-27	9.707005e-27	True
SSDI Cash Assistance	8.576335e-27	2.787309e-26	True
Medicare Insurance	2.989112e-24	9.143166e-24	True
High Blood Pressure	2.098500e-19	6.062334e-19	True
Smokes	5.994588e-18	1.640624e-17	True
IfMale	2.101983e-17	5.465156e-17	True
Diabetes	5.997728e-14	1.485152e-13	True
Other Cardiac	2.778003e-12	6.566189e-12	True
Private Insurance	7.094948e-12	1.604075e-11	True
Pulmonary Asthma	3.627416e-10	7.859402e-10	True
Endocrine Condition	1.479935e-09	3.078265e-09	True
Received Smoking Counseling	2.306427e-09	4.612854e-09	True
Other Cash Benefits	1.474657e-07	2.840080e-07	True
Public Assistance Cash Program	2.771732e-06	5.147503e-06	True
Mobility Impairment Disorder	3.972035e-06	7.122270e-06	True
Criminal Justice Status	8.662690e-06	1.501533e-05	True
Hearing Visual Impairment	1.828773e-05	3.067619e-05	True
Cancer	8.048305e-04	1.307850e-03	True
Veterans Disability Benefits	2.964634e-03	4.694426e-03	True
Heart Attack	3.069432e-03	4.694426e-03	True
Stroke	3.854834e-03	5.727183e-03	True
Medicaid Managed Insurance	4.063840e-03	5.869991e-03	True
Other Insurance	7.380559e-03	1.037268e-02	True
Neurological Condition	1.073293e-02	1.468717e-02	True
LGBTQ	2.435362e-02	3.247150e-02	True
Kidney Disease	2.719320e-02	3.535116e-02	True
Hispanic Ethnicity	9.103429e-02	[illegible]	False
Alzheimer or Dementia	1.066767e-01	1.320760e-01	False
Other Developmental Disability	1.197411e-01	1.448032e-01	False
Autism Spectrum	2.875198e-01	3.322451e-01	False
Liver Disease	2.841385e-01	3.322451e-01	False
Veterans Cash Assistance	3.288255e-01	3.717158e-01	False
Veteran Status	3.927511e-01	4.345332e-01	False
Transgender	4.387070e-01	4.752659e-01	False
Traumatic Brain Injury	4.727975e-01	5.017443e-01	False
Child Health Plus Insurance	5.540894e-01	5.762530e-01	False
Intellectual Disability	7.810170e-01	7.963310e-01	False
Other Chronic Med Condition	8.204827e-01	8.204827e-01	False

Analysis: Logistic Regression Method 1

In order to determine which patients should be flagged as being at high risk of being admitted for emergency services, the patients will need to be classified as "emergency" or "non-emergency" cases. Because this analysis is only concerned with predicting emergency admissions, all other categories of visits will be grouped into the same category of non-emergency visits. Because of the binary nature of the classification, a logistic regression model is applied to the data. All rows containing missing values are removed from the data, so that they do not cause issues during the creation of the regression model. This limits the data set to about a quarter of its original size.

First, because all of the data is categorical, dummy codes are created in order to convert the categorical data into numeric data. Each response option to the original variables is given its own column of data, and 0 and 1 are then used to indicate the presence or non-presence of each response for each patient. In order to test the accuracy of the logistic regression model, the data is split via random sampling into a training data set and a testing data set. The model is created using the training data set, and then the accuracy is checked by using the model to try to predict the correct type of visit of the testing data set. The amount of cases that were correctly predicted will indicate how accurate the logistic regression model is.

For the first iteration of the logistic regression model, all of the variables are used in the algorithm. Going forward, backward-fitting is used to eliminate variables that do not contribute significantly to the prediction model, which is indicated by a non-significant p-value. Once all non-significant variables have been removed from the model, the remaining variables are used in the logistic regression model as predictors of the type of visit of each patient. The final model is then applied to the testing data set to determine its accuracy.

Analysis: Logistic Regression Method 2

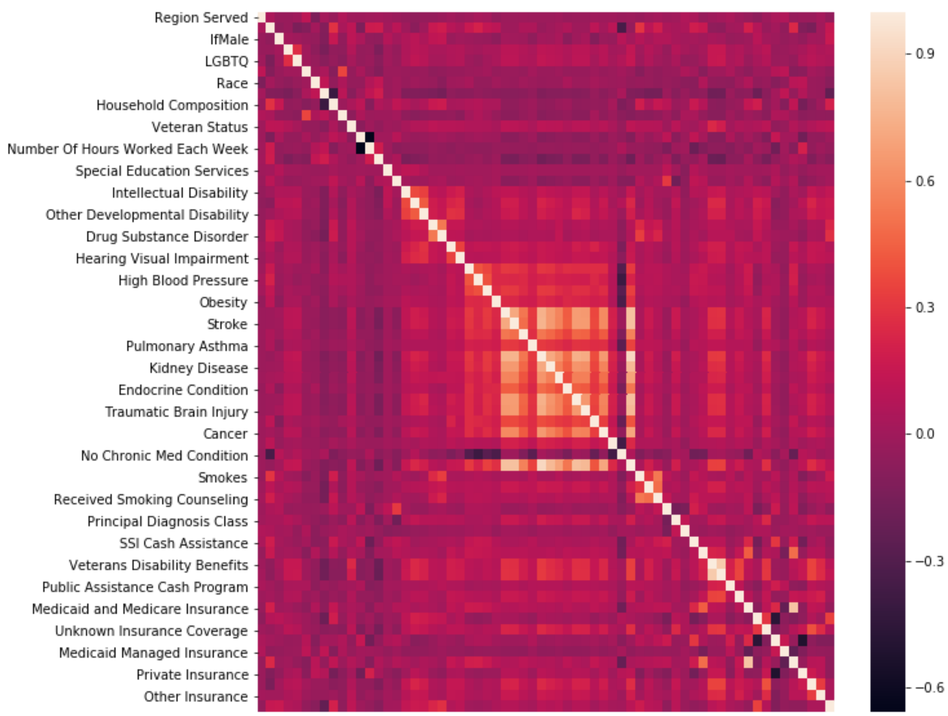
As in Method 1, patients are classified as "emergency" or "non-emergency," and a logistic regression model is used. The key differences are that missing data is left in the data set as "unknown," so that the full data set is used, and instead of creating separate variables for each categorical response, the same variables are maintained. Each categorical response is replaced with a number. For binary responses, 1 is used for "yes," -1 is used for "no," and 0 is used for "unknown." For all variables with multiple responses, "unknown" responses are still assigned to 0, but all other responses are assigned to integers, such as 1, 2, 3, etc., without meaning. The category representing zip codes is removed and replaced with a binary category indicating whether or not the patient is homeless.

The rest of Method 2 follows the same procedure as Method 1, with the data then being split randomly into training and testing data sets, backward-fitting being used to eliminate non-significant variables, and the accuracy being tested against the testing data.

IV. Results

Multicollinearity Analysis:

According to analysis on multicollinearity, we found almost all variables have a variable inflation factor with 0 to 10, more specifically, with range between 0 to 5. This finding illustrates that our model doesn't have serious multicollinearity problem. Only one variable, "Unknown Chronic Medical Condition" has a VIF that is larger than 10 (17.33), indicating this variable may make influence on the accuracy of the model. Generally speaking, since the data set is large enough for logistic regression model, our model avoids the risk of multicollinearity. The correlation heatmap of all independent variables is shown below.



Logistic Regression Method 1 Analysis Results:

Images of the first iteration of the logistic regression analysis are shown below. Using all of the variables in the data set, we see that there are many variables that do not contribute significantly to the model, indicated by the number of asterisks at the end of the row (more asterisks means more significant), as well as multiple NA values. The NA values likely are a result of there being so few emergency cases that several of the variables in the testing data set are never associated with an emergency admission.

Initially, we can see some variables that are highly significant for predicting an emergency admission in the first iteration, including the patient living alone, being a smoker who received smoking medication, having a substance use disorder, and being homeless (zip code 888).

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6955  -0.1280  -0.0718  -0.0256   4.0120

Coefficients: (6 not defined because of singularities)
                                                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                             -1.895e+00  1.840e+00  -1.030 0.303155
`Region.Served_HUDSON RIVER REGION`                                     -5.762e-02  6.960e-01  -0.083 0.934020
`Region.Served_LONG ISLAND REGION`                                       1.697e+00  8.742e-01   1.941 0.052201 .
`Region.Served_NEW YORK CITY REGION`                                     7.530e-01  7.006e-01   1.075 0.282437
`Region.Served_WESTERN REGION`                                           1.776e+00  6.771e-01   2.623 0.008722 **
Age.Group_ADULT                                                         -1.947e-02  4.098e-01  -0.048 0.962108
Sex_MALE                                                                 1.042e-01  1.446e-01   0.720 0.471408
`Transgender_NO, NOT TRANSGENDER`                                       -6.700e-01  6.727e-01  -0.996 0.319274
`Transgender_YES, TRANSGENDER`                                                  NA         NA      NA       NA
Sexual.Orientation_BISEXUAL                                             -1.819e-01  4.773e-01  -0.381 0.703193
`Sexual.Orientation_LESBIAN OR GAY`                                      2.730e-02  4.635e-01   0.059 0.953023
Sexual.Orientation_OTHER                                                -1.691e+01  1.923e+03  -0.009 0.992984
`Sexual.Orientation_STRAIGHT OR HETEROSEXUAL`                           -2.640e-01  3.097e-01  -0.853 0.393920
Hispanic.Ethnicity_YES                                                  -1.881e-01  2.323e-01  -0.810 0.418014
`Race_BLACK ONLY`                                                       -1.741e-02  2.458e-01  -0.071 0.943549
`Race_MULTI-RACIAL`                                                     -1.204e-02  4.169e-01  -0.029 0.976961
`Race_WHITE ONLY`                                                       -4.937e-01  2.292e-01  -2.154 0.031249 *
`Living.Situation_INSTITUTIONAL SETTING`                                -2.775e-01  6.293e-01  -0.441 0.659284
`Living.Situation_PRIVATE RESIDENCE`                                    -3.387e-01  2.724e-01  -1.243 0.213779
`Household.Composition_COHABITATES WITH OTHERS`                          6.218e-01  2.191e-01   2.838 0.004537 **
`Household.Composition_LIVES ALONE`                                             NA         NA      NA       NA
`Preferred.Language_ALL OTHER LANGUAGES`                                -8.729e-02  1.498e+00  -0.058 0.953531
`Preferred.Language_ASIAN AND PACIFIC ISLAND`                           -1.674e+01  2.311e+03  -0.007 0.994222
Preferred.Language_ENGLISH                                              -3.762e-01  1.080e+00  -0.348 0.727637
`Preferred.Language_INDO-EUROPEAN`                                      -2.492e-01  1.307e+00  -0.191 0.848762
Preferred.Language_SPANISH                                               3.990e-02  1.117e+00   0.036 0.971515
Veteran.Status_YES                                                      -1.225e+00  5.723e-01  -2.140 0.032317 *
`Employment.Status_NON-PAID/VOLUNTEER`                                  -1.608e+01  1.324e+03  -0.012 0.990315
`Employment.Status_NOT IN LABOR FORCE:UNEMPLOYED AND NOT LOOKING FOR WORK`  8.387e-02  2.921e-01   0.287 0.773998
`Employment.Status_UNEMPLOYED, LOOKING FOR WORK`                         3.624e-01  3.110e-01   1.165 0.243860
`Number.Of.Hours.Worked.Each.Week_01-14 HOURS`                          -5.473e-01  5.701e-01  -0.960 0.337048
`Number.Of.Hours.Worked.Each.Week_15-34 HOURS`                          -7.316e-01  4.615e-01  -1.585 0.112890
`Number.Of.Hours.Worked.Each.Week_35 HOURS OR MORE`                             NA         NA      NA       NA
`Education.Status_COLLEGE OR GRADUATE DEGREE`                           -7.651e-01  3.941e-01  -1.941 0.052228 .
`Education.Status_MIDDLE SCHOOL TO HIGH SCHOOL`                         -8.175e-01  3.377e-01  -2.420 0.015506 *
`Education.Status_NO FORMAL EDUCATION`                                  -2.369e-01  8.206e-01  -0.289 0.772785
`Education.Status_PRE-K TO FIFTH GRADE`                                 -1.611e+00  4.881e-01  -3.300 0.000965 ***
`Education.Status_SOME COLLEGE`                                         -6.144e-01  3.693e-01  -1.664 0.096138 .
Special.Education.Services_NO                                            6.527e-01  4.272e-01   1.528 0.126517
Special.Education.Services_YES                                           4.086e-01  4.218e-01   0.969 0.332613
Mental.Illness_YES                                                      -8.652e-01  2.435e-01  -3.553 0.000381 ***
Intellectual.Disability_YES                                              4.603e-01  2.630e-01   1.750 0.080101 .
Autism.Spectrum_YES                                                     -8.868e-01  6.481e-01  -1.368 0.171210
Other.Developmental.Disability_YES                                       7.589e-01  2.866e-01   2.648 0.008105 **
Alcohol.Related.Disorder_YES                                             5.514e-01  1.690e-01   3.263 0.001103 **
Drug.Substance.Disorder_YES                                              3.095e-01  1.760e-01   1.759 0.078633 .
Mobility.Impairment.Disorder_YES                                        -1.213e+00  6.150e-01  -1.972 0.048597 *
Hearing.Visual.Impairment_YES                                           -1.977e-01  3.845e-01  -0.514 0.607214
Hyperlipidemia_YES                                                      -4.215e-01  2.818e-01  -1.496 0.134677
High.Blood.Pressure_YES                                                  1.210e-02  2.135e-01   0.057 0.954804
Diabetes_YES                                                            -5.515e-02  2.684e-01  -0.205 0.837182
Obesity_YES                                                             -6.540e-01  2.634e-01  -2.482 0.013047 *
Heart.Attack_YES                                                         8.347e-01  6.214e-01   1.343 0.179176
Stroke_YES                                                              -8.462e-01  1.024e+00  -0.827 0.408433
Other.Cardiac_YES                                                       -1.179e+00  6.039e-01  -1.953 0.050855 .
Pulmonary.Asthma_YES                                                     1.553e-01  2.307e-01   0.673 0.500926
Alzheimer.or.Dementia_YES                                               -1.361e-01  1.092e+00  -0.125 0.900781
Kidney.Disease_YES                                                       2.515e-01  6.109e-01   0.412 0.680598
Liver.Disease_YES                                                       -2.994e-01  3.780e-01  -0.792 0.428304
Endocrine.Condition_YES                                                 -1.177e-01  3.858e-01  -0.305 0.760308
Neurological.Condition_YES                                               2.979e-01  6.253e-01   0.476 0.633780
Traumatic.Brain.Injury_YES                                              -5.329e-01  5.702e-01  -0.935 0.350011
Joint.Disease_YES                                                       -6.468e-01  3.768e-01  -1.716 0.086095 .
Cancer_YES                                                               2.579e-01  4.924e-01   0.524 0.600427
Other.Chronic.Med.Condition_YES                                          1.460e-01  2.070e-01   0.705 0.480563
No.Chronic.Med.Condition_NO                                             -1.140e-01  2.178e-01  -0.523 0.600706
Unknown.Chronic.Med.Condition_YES                                               NA         NA      NA       NA
Smokes_YES                                                               1.176e-01  1.707e-01   0.689 0.490640
Received.Smoking.Medication_YES                                          7.708e-01  2.313e-01   3.332 0.000861 ***
Received.Smoking.Counseling_YES                                         -4.300e-01  2.065e-01  -2.082 0.037358 *
Serious.Mental.Illness_YES                                              -1.047e-02  1.973e-01  -0.053 0.957683
```

```
Serious.Mental.Illness_YES                                               -1.047e-02  1.973e-01  -0.053 0.957683
`Principal.Diagnosis.Class_MENTAL ILLNESS`                               -1.700e+00  7.831e-01  -2.171 0.029927 *
`Principal.Diagnosis.Class_NOT MI - DEVELOPMENTAL DISORDERS`             -1.194e+00  1.093e+00  -1.093 0.274569
`Principal.Diagnosis.Class_NOT MI - ORGANIC MENTAL DISORDER`             -2.742e-01  1.126e+00  -0.244 0.807540
`Principal.Diagnosis.Class_SUBSTANCE-RELATED AND ADDICTIVE DISORDERS`    -1.658e-01  8.178e-01  -0.203 0.839368
`Additional.Diagnosis.Class_MENTAL ILLNESS`                               7.020e-01  4.679e-01   1.500 0.133528
`Additional.Diagnosis.Class_NOT MI - DEVELOPMENTAL DISORDERS`             9.102e-01  5.777e-01   1.575 0.115151
`Additional.Diagnosis.Class_NOT MI - ORGANIC MENTAL DISORDER`             9.542e-01  1.143e+00   0.835 0.403743
`Additional.Diagnosis.Class_SUBSTANCE-RELATED AND ADDICTIVE DISORDERS`    1.455e+00  4.823e-01   3.016 0.002560 **
SSI.Cash.Assistance_YES                                                  -4.016e-01  1.700e-01  -2.362 0.018176 *
SSDI.Cash.Assistance_YES                                                 -3.739e-01  2.465e-01  -1.517 0.129247
Veterans.Disability.Benefits_YES                                          2.062e+00  7.462e-01   2.763 0.005720 **
Veterans.Cash.Assistance_YES                                             -1.731e+01  2.952e+03  -0.006 0.995321
Public.Assistance.Cash.Program_YES                                       -3.669e-01  1.839e-01  -1.994 0.046108 *
Other.Cash.Benefits_YES                                                  -7.939e-01  3.708e-01  -2.141 0.032251 *
Medicaid.and.Medicare.Insurance_YES                                       1.373e-01  5.195e-01   0.264 0.791528
No.Insurance_YES                                                          1.102e-01  4.288e-01   0.257 0.797161
Unknown.Insurance.Coverage_YES                                                   NA         NA      NA       NA
Medicaid.Insurance_YES                                                   -4.025e-01  3.695e-01  -1.089 0.275979
Medicaid.Managed.Insurance_NO                                             1.827e-02  1.732e-01   0.105 0.916010
Medicaid.Managed.Insurance_YES                                                   NA         NA      NA       NA
Medicare.Insurance_YES                                                   -2.114e-01  4.686e-01  -0.451 0.652000
Private.Insurance_YES                                                     4.006e-03  3.562e-01   0.011 0.991029
Child.Health.Plus.Insurance_YES                                           3.387e-01  4.895e-01   0.692 0.488972
Other.Insurance_YES                                                      -3.036e-01  3.549e-01  -0.855 0.392317
Criminal.Justice.Status_YES                                              -4.597e-01  2.117e-01  -2.171 0.029909 *
Three.Digit.Residence.Zip.Code_101                                       -1.615e+01  3.652e+03  -0.004 0.996471
Three.Digit.Residence.Zip.Code_103                                        7.677e-01  4.747e-01   1.617 0.105847
Three.Digit.Residence.Zip.Code_104                                        6.151e-01  3.361e-01   1.830 0.067235 .
Three.Digit.Residence.Zip.Code_105                                       -4.380e-01  1.134e+00  -0.386 0.699316
Three.Digit.Residence.Zip.Code_106                                       -1.523e+01  2.111e+03  -0.007 0.994243
Three.Digit.Residence.Zip.Code_107                                        7.766e-01  8.795e-01   0.883 0.377269
Three.Digit.Residence.Zip.Code_108                                       -1.521e+01  2.688e+03  -0.006 0.995483
Three.Digit.Residence.Zip.Code_109                                        4.349e-01  8.725e-01   0.499 0.618120
Three.Digit.Residence.Zip.Code_110                                       -1.637e+01  2.069e+03  -0.008 0.993688
Three.Digit.Residence.Zip.Code_111                                        2.884e-02  1.067e+00   0.027 0.978438
Three.Digit.Residence.Zip.Code_112                                        9.547e-01  3.129e-01   3.051 0.002280 **
Three.Digit.Residence.Zip.Code_113                                        3.630e-02  5.328e-01   0.068 0.945683
Three.Digit.Residence.Zip.Code_114                                        2.882e-01  4.538e-01   0.635 0.525327
Three.Digit.Residence.Zip.Code_115                                       -1.363e+00  9.307e-01  -1.464 0.143131
Three.Digit.Residence.Zip.Code_116                                       -3.144e-01  1.046e+00  -0.301 0.763719
Three.Digit.Residence.Zip.Code_117                                       -1.332e+00  7.756e-01  -1.717 0.085995 .
Three.Digit.Residence.Zip.Code_118                                        9.034e-01  1.195e+00   0.756 0.449712
Three.Digit.Residence.Zip.Code_119                                       -1.262e-01  8.636e-01  -0.146 0.883819
Three.Digit.Residence.Zip.Code_120                                        1.218e-01  1.148e+00   0.106 0.915491
Three.Digit.Residence.Zip.Code_121                                        2.012e+00  6.762e-01   2.975 0.002928 **
Three.Digit.Residence.Zip.Code_122                                       -1.527e+01  1.599e+03  -0.010 0.992381
Three.Digit.Residence.Zip.Code_123                                       -1.515e+01  1.864e+03  -0.008 0.993512
Three.Digit.Residence.Zip.Code_124                                       -1.504e+01  1.465e+03  -0.010 0.991808
Three.Digit.Residence.Zip.Code_125                                       -1.524e+01  1.179e+03  -0.013 0.989684
Three.Digit.Residence.Zip.Code_126                                       -1.518e+01  1.696e+03  -0.009 0.992855
Three.Digit.Residence.Zip.Code_127                                        1.725e+00  8.907e-01   1.937 0.052750 .
Three.Digit.Residence.Zip.Code_128                                        1.560e+00  7.783e-01   2.004 0.045109 *
Three.Digit.Residence.Zip.Code_129                                       -1.548e+01  1.341e+03  -0.012 0.990787
Three.Digit.Residence.Zip.Code_130                                       -1.510e+01  1.136e+03  -0.013 0.989400
Three.Digit.Residence.Zip.Code_131                                        1.478e+00  7.917e-01   1.867 0.061904 .
Three.Digit.Residence.Zip.Code_132                                        1.384e+00  8.588e-01   1.611 0.107164
Three.Digit.Residence.Zip.Code_133                                        1.202e+00  1.258e+00   0.955 0.339486
Three.Digit.Residence.Zip.Code_134                                       -1.525e+01  1.456e+03  -0.010 0.991639
Three.Digit.Residence.Zip.Code_135                                       -1.510e+01  1.540e+03  -0.010 0.992174
Three.Digit.Residence.Zip.Code_136                                       -1.520e+01  1.414e+03  -0.011 0.991421
Three.Digit.Residence.Zip.Code_137                                       -1.536e+01  1.838e+03  -0.008 0.993333
Three.Digit.Residence.Zip.Code_138                                       -1.538e+01  1.588e+03  -0.010 0.992271
Three.Digit.Residence.Zip.Code_139                                       -1.500e+01  2.156e+03  -0.007 0.994448
Three.Digit.Residence.Zip.Code_140                                       -3.957e-01  6.701e-01  -0.591 0.554839
Three.Digit.Residence.Zip.Code_141                                        5.210e-01  7.112e-01   0.733 0.463825
Three.Digit.Residence.Zip.Code_142                                       -3.326e-01  5.834e-01  -0.570 0.568620
Three.Digit.Residence.Zip.Code_143                                        2.041e-01  8.717e-01   0.234 0.814840
Three.Digit.Residence.Zip.Code_144                                        6.540e-01  6.224e-01   1.051 0.293415
Three.Digit.Residence.Zip.Code_145                                        8.078e-01  6.058e-01   1.333 0.182387
Three.Digit.Residence.Zip.Code_146                                        9.012e-01  5.381e-01   1.675 0.093958 .
Three.Digit.Residence.Zip.Code_147                                       -1.684e+01  1.352e+03  -0.012 0.990061
Three.Digit.Residence.Zip.Code_148                                       -1.762e+00  1.115e+00  -1.580 0.114141
Three.Digit.Residence.Zip.Code_149                                       -1.668e+01  1.675e+03  -0.010 0.992058
Three.Digit.Residence.Zip.Code_888                                        2.196e+00  3.662e-01   5.997 2.01e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The NA variables are removed, and the first iteration of the model is tested against the testing data set. The model is determined to have an accuracy of about 19%, which is very low. The model is then cleaned up iteratively using backwards-fitting to eliminate one-by-one the variables that do not contribute significantly to the model. At the end of this process, 40 variables out of the original 145 dummy variables remain, as shown below. This version of the model is tested against the testing data set and is found to have an accuracy of about 42.5%. While this is a significant improvement, it is still a very low accuracy.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5565  -0.1282  -0.0793  -0.0462   4.0254

Coefficients:
                                                                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                                                                  -4.5469     0.4520 -10.060  < 2e-16 ***
`Region.Served_LONG ISLAND REGION`                                            1.3685     0.3982   3.437 0.000589 ***
`Region.Served_NEW YORK CITY REGION`                                          1.1300     0.3260   3.466 0.000529 ***
`Region.Served_WESTERN REGION`                                                1.7658     0.3356   5.262 1.43e-07 ***
`Race_WHITE ONLY`                                                            -0.5480     0.1503  -3.646 0.000266 ***
`Household.Composition_COHABITATES WITH OTHERS`                               0.4947     0.1666   2.970 0.002976 **
Veteran.Status_YES                                                           -1.3078     0.5688  -2.299 0.021493 *
`Employment.Status_NOT IN LABOR FORCE:UNEMPLOYED AND NOT LOOKING FOR WORK`    0.5445     0.2074   2.626 0.008648 **
`Employment.Status_UNEMPLOYED, LOOKING FOR WORK`                              0.7992     0.2501   3.195 0.001399 **
Mental.Illness_YES                                                           -0.9806     0.2220  -4.418 9.98e-06 ***
Other.Developmental.Disability_YES                                            0.7508     0.2666   2.816 0.004860 **
Alcohol.Related.Disorder_YES                                                  0.5200     0.1638   3.175 0.001499 **
Drug.Substance.Disorder_YES                                                   0.3571     0.1702   2.099 0.035828 *
Mobility.Impairment.Disorder_YES                                             -1.2322     0.5936  -2.076 0.037913 *
Obesity_YES                                                                  -0.7789     0.2441  -3.191 0.001420 **
Other.Cardiac_YES                                                            -1.2057     0.5895  -2.045 0.040820 *
Joint.Disease_YES                                                            -0.8054     0.3671  -2.194 0.028239 *
Received.Smoking.Medication_YES                                               0.8460     0.2257   3.748 0.000178 ***
Received.Smoking.Counseling_YES                                              -0.3903     0.1919  -2.034 0.041992 *
`Principal.Diagnosis.Class_MENTAL ILLNESS`                                   -1.3391     0.2147  -6.238 4.42e-10 ***
`Additional.Diagnosis.Class_SUBSTANCE-RELATED AND ADDICTIVE DISORDERS`        0.7438     0.1621   4.587 4.49e-06 ***
SSI.Cash.Assistance_YES                                                      -0.4712     0.1602  -2.942 0.003262 **
SSDI.Cash.Assistance_YES                                                     -0.5484     0.2136  -2.568 0.010234 *
Veterans.Disability.Benefits_YES                                              1.5728     0.7388   2.129 0.033264 *
Public.Assistance.Cash.Program_YES                                           -0.4775     0.1771  -2.696 0.007016 **
Other.Cash.Benefits_YES                                                      -0.8798     0.3642  -2.416 0.015704 *
Medicaid.Insurance_YES                                                       -0.3938     0.1546  -2.548 0.010831 *
Criminal.Justice.Status_YES                                                  -0.4258     0.2036  -2.091 0.036517 *
Three.Digit.Residence.Zip.Code_103                                            0.8831     0.4229   2.088 0.036792 *
Three.Digit.Residence.Zip.Code_104                                            0.7229     0.2604   2.776 0.005509 **
Three.Digit.Residence.Zip.Code_112                                            1.1074     0.2298   4.818 1.45e-06 ***
Three.Digit.Residence.Zip.Code_121                                            2.5436     0.5112   4.976 6.48e-07 ***
Three.Digit.Residence.Zip.Code_127                                            2.2936     0.7707   2.976 0.002922 **
Three.Digit.Residence.Zip.Code_128                                            2.1777     0.6485   3.358 0.000785 ***
Three.Digit.Residence.Zip.Code_131                                            2.0432     0.6191   3.300 0.000966 ***
Three.Digit.Residence.Zip.Code_132                                            1.9422     0.5386   3.606 0.000311 ***
Three.Digit.Residence.Zip.Code_141                                            1.1905     0.5500   2.165 0.030410 *
Three.Digit.Residence.Zip.Code_144                                            1.2257     0.4371   2.804 0.005041 **
Three.Digit.Residence.Zip.Code_145                                            1.4812     0.4148   3.571 0.000356 ***
Three.Digit.Residence.Zip.Code_146                                            1.4716     0.3070   4.793 1.65e-06 ***
Three.Digit.Residence.Zip.Code_888                                            2.6342     0.2409  10.934  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Initial Model 1

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	2,886	12,787
	Emergency	5	163

Final Model 1

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	6,851	8,822
	Emergency	12	156

Logistic Regression Method 2 Analysis Results:

Images of the first iteration of the model are shown below. This method did not have issues with NA values appearing in the logistic regression model. There are several non-significant variables that need to be removed with backwards-fitting, but far fewer than in Method 1.

Initially, variables that we see as being highly significant include the region of the hospital, the transgender status of the patient, the education level of the patient, the mental illness status of the patient, and whether or not the patient is homeless.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1393  -0.1814  -0.1325  -0.0964   3.7537

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                  -3.372872   0.350735  -9.617  < 2e-16 ***
Region.Served                 0.174192   0.021087   8.261  < 2e-16 ***
Age.Group                     0.432116   0.081120   5.327 9.99e-08 ***
Sex                          -0.056613   0.026547  -2.133 0.032960 *
Transgender                   0.810780   0.066926  12.115  < 2e-16 ***
Sexual.Orientation           -0.010623   0.018186  -0.584 0.559151
Hispanic.Ethnicity           -0.076426   0.032360  -2.362 0.018189 *
Race                          0.004341   0.018250   0.238 0.811970
Living.Situation              0.224744   0.040507   5.548 2.89e-08 ***
Household.Composition        -0.259569   0.037198  -6.978 2.99e-12 ***
Preferred.Language            0.026468   0.034177   0.774 0.438671
Veteran.Status                0.040973   0.063383   0.646 0.517997
Employment.Status             0.131616   0.032011   4.112 3.93e-05 ***
Number.Of.Hours.Worked.Each.Week -0.059669   0.039274  -1.519 0.128685
Education.Status             -0.211296   0.015008 -14.079  < 2e-16 ***
Special.Education.Services    0.018096   0.062509   0.290 0.772199
Mental.Illness               -0.364189   0.042177  -8.635  < 2e-16 ***
Intellectual.Disability       0.155586   0.054770   2.841 0.004501 **
Autism.Spectrum              -0.182660   0.089441  -2.042 0.041127 *
Other.Developmental.Disability -0.015993   0.077394  -0.207 0.836291
Alcohol.Related.Disorder      0.218844   0.041672   5.252 1.51e-07 ***
Drug.Substance.Disorder       0.272920   0.039086   6.983 2.90e-12 ***
Mobility.Impairment.Disorder -0.039624   0.073570  -0.539 0.590168
Hearing.Visual.Impairment    -0.073843   0.067164  -1.099 0.271576
Hyperlipidemia               -0.224293   0.062906  -3.565 0.000363 ***
High.Blood.Pressure          -0.069541   0.047751  -1.456 0.145304
Diabetes                     -0.046722   0.057074  -0.819 0.413008
Obesity                      -0.289993   0.058061  -4.995 5.89e-07 ***
Heart.Attack                 -0.144082   0.185801  -0.775 0.438066
Stroke                       -0.333111   0.209488  -1.590 0.111808
Other.Cardiac                -0.404353   0.111454  -3.628 0.000286 ***
Pulmonary.Asthma             -0.137153   0.056610  -2.423 0.015402 *
Alzheimer.or.Dementia         0.040586   0.182583   0.222 0.824089
Kidney.Disease                0.004469   0.139815   0.032 0.974503
Liver.Disease                -0.228409   0.100399  -2.275 0.022906 *
Endocrine.Condition          -0.043459   0.082978  -0.524 0.600457
Neurological.Condition       -0.013454   0.156774  -0.086 0.931611
Traumatic.Brain.Injury       -0.067955   0.131136  -0.518 0.604318
Joint.Disease                -0.327168   0.075737  -4.320 1.56e-05 ***
Cancer                       -0.179450   0.125401  -1.431 0.152427
Other.Chronic.Med.Condition   0.029658   0.046683   0.635 0.525234
No.Chronic.Med.Condition      0.006460   0.047634   0.136 0.892116
Unknown.Chronic.Med.Condition 1.041081   0.245165   4.246 2.17e-05 ***
Smokes                       -0.010674   0.036438  -0.293 0.769566
Received.Smoking.Medication   0.563089   0.048280  11.663  < 2e-16 ***
Received.Smoking.Counseling  -0.035332   0.045461  -0.777 0.437049
Serious.Mental.Illness       -0.019883   0.033001  -0.603 0.546839
Principal.Diagnosis.Class     0.415281   0.023602  17.595  < 2e-16 ***
Additional.Diagnosis.Class   -0.056868   0.016985  -3.348 0.000813 ***
SSI.Cash.Assistance          -0.172474   0.035937  -4.799 1.59e-06 ***
SSDI.Cash.Assistance         -0.119533   0.046594  -2.565 0.010306 *
Veterans.Disability.Benefits  0.571704   0.120804   4.732 2.22e-06 ***
Veterans.Cash.Assistance      0.226735   0.123370   1.838 0.066085 .
Public.Assistance.Cash.Program -0.181767   0.040891  -4.445 8.78e-06 ***
Other.Cash.Benefits          -0.021043   0.051884  -0.406 0.685050
Medicaid.and.Medicare.Insurance 0.088896   0.091640   0.970 0.332019
No.Insurance                  0.222243   0.071094   3.126 0.001772 **
Unknown.Insurance.Coverage    0.097838   0.061635   1.587 0.112426
Medicaid.Insurance           -0.171305   0.057325  -2.988 0.002805 **
Medicaid.Managed.Insurance    0.016142   0.038847   0.416 0.677767
Medicare.Insurance           -0.097118   0.077202  -1.258 0.208404
Private.Insurance             0.137938   0.053548   2.576 0.009997 **
Child.Health.Plus.Insurance   0.337417   0.074003   4.560 5.13e-06 ***
Other.Insurance              -0.185849   0.063474  -2.928 0.003412 **
Criminal.Justice.Status      -0.116247   0.042703  -2.722 0.006484 **
Homeless                      1.089841   0.050900  21.412  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing this first iteration of the model for accuracy with the testing data set reveals that it is already about 92.2% accurate. This is significantly better than the first iteration of the model in Method 1. In order to continue to improve the accuracy, backwards-fitting is performed to remove non-significant variables. This reduces the model to 36 variables, down from 66 variables, as shown below. This version of the model is tested against the testing data set and is found to have an accuracy of about 95%. This is a small improvement over the first iteration of the model, and a significant improvement over the final model created in Method 1.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1119  -0.1812  -0.1327  -0.0969   3.7624

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -2.86439    0.22915 -12.500  < 2e-16 ***
Region.Served                   0.17575    0.02091   8.404  < 2e-16 ***
Age.Group                       0.37401    0.07464   5.011 5.42e-07 ***
Sex                            -0.05612    0.02616  -2.145 0.031936 *
Transgender                     0.82415    0.06269  13.147  < 2e-16 ***
Hispanic.Ethnicity             -0.06765    0.03118  -2.169 0.030047 *
Living.Situation                0.22279    0.03915   5.690 1.27e-08 ***
Household.Composition          -0.27132    0.03671  -7.391 1.45e-13 ***
Employment.Status               0.14010    0.02784   5.031 4.87e-07 ***
Education.Status               -0.21853    0.01459 -14.978  < 2e-16 ***
Mental.Illness                 -0.37251    0.03878  -9.607  < 2e-16 ***
Intellectual.Disability         0.14246    0.05357   2.659 0.007827 **
Autism.Spectrum                -0.22607    0.07960  -2.840 0.004510 **
Alcohol.Related.Disorder        0.20748    0.04112   5.046 4.52e-07 ***
Drug.Substance.Disorder         0.26953    0.03819   7.057 1.70e-12 ***
Hyperlipidemia                 -0.24711    0.06150  -4.018 5.87e-05 ***
High.Blood.Pressure            -0.09502    0.04310  -2.205 0.027481 *
Obesity                        -0.29746    0.05416  -5.492 3.98e-08 ***
Other.Cardiac                  -0.42766    0.11023  -3.880 0.000105 ***
Pulmonary.Asthma               -0.14725    0.05262  -2.798 0.005140 **
Liver.Disease                  -0.23336    0.09850  -2.369 0.017828 *
Joint.Disease                  -0.34244    0.07407  -4.623 3.77e-06 ***
Unknown.Chronic.Med.Condition   0.72123    0.09276   7.775 7.54e-15 ***
Received.Smoking.Medication     0.53686    0.03925  13.678  < 2e-16 ***
Principal.Diagnosis.Class       0.41577    0.02337  17.787  < 2e-16 ***
Additional.Diagnosis.Class     -0.05497    0.01683  -3.266 0.001090 **
SSI.Cash.Assistance            -0.17096    0.03492  -4.895 9.82e-07 ***
SSDI.Cash.Assistance           -0.12541    0.04239  -2.958 0.003092 **
Veterans.Disability.Benefits    0.73193    0.07502   9.757  < 2e-16 ***
Public.Assistance.Cash.Program -0.17058    0.03981  -4.285 1.82e-05 ***
No.Insurance                    0.30650    0.05475   5.598 2.17e-08 ***
Medicaid.Insurance             -0.11313    0.04271  -2.649 0.008075 **
Private.Insurance               0.17603    0.04558   3.862 0.000113 ***
Child.Health.Plus.Insurance     0.37085    0.07006   5.294 1.20e-07 ***
Other.Insurance                -0.16390    0.06057  -2.706 0.006811 **
Criminal.Justice.Status        -0.09696    0.04152  -2.335 0.019539 *
Homeless                        1.10805    0.05015  22.094  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

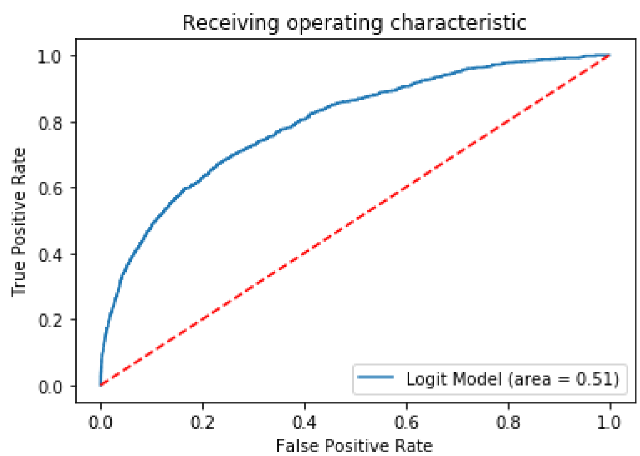
Initial Model 2

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	72,232	5,370
	Emergency	815	679

Final Model 2

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	74,625	2,977
	Emergency	950	544

ROC Curve :



From the figure above, it can be seen that since the curve is smooth in the graph, it indicates that the logistic model obtained isn't overfitting. Meanwhile, it shows that the model can still be improved on the prediction performance of program category, especially on prediction of "EMERGENCY" category.

V. Discussion

Due to its overwhelmingly higher accuracy, it is recommended that the model developed in Method 2 be used, and not the model developed in Method 1. The developed model can be used by hospital leadership, emergency department administration, quality improvement managers, primary care providers, and other interested stakeholders to understand which patients are at a higher risk of having an emergency room visit due to their mental health disorder & other key demographic information. These stakeholders can use this information to properly support those who are flagged as high risk and as a result try to decrease the number of emergency room visits.

A common theme across both models developed with both methods is the significance of many indicators of poverty. While not all indicators of poverty are significant in the model, and not all significant variables are indicators of poverty, many are. This means that overall, if a patient at the hospital exhibits many indicators of poverty, they are more likely to have an increased chance of being high risk for emergency admission in the future. In particular, it is likely that the biggest influence on the risk of emergency admission for patients is if they are experiencing homelessness.

Now that we have provided the stakeholders with the high risk patient model, they need to determine if they would like to use the model in their institution(s). If so, there are a few next steps recommended to make the model operational and useful. First, they must apply the model to the mental health patient population of a selected region/hospital. Next, they must share the data model results with the care providers in a way that is easy to read and follow. This could be accomplished with an electronic dashboard that displays the data. Finally, if the model and data results are well received by the providers, the stakeholders must integrate the model into the EHR system, scheduling system, and hospital workflow. This will guarantee that the model is appropriately used and integrated in order to flag patients in real time so providers can better manage their preventative care.

Along with these implementation recommendations, we urge the stakeholders to keep in mind a few limitations of the data model. The model was built on only one week's worth of patient data in New York State. We are making the assumption that these data points and rates of emergency room visits would be reflective of an average week for this patient population. We are also making the assumption that the reported data is accurate and precise.

With these limitations in mind, it should be noted that there are a few areas of improvement for the data set and data model. If the survey was reflective of a month, quarter, or even a year's worth of data, there would be more data available to build and test the model. This could improve the accuracy and trust by the stakeholders of the model. Also if the survey included data from a wider area, perhaps the Mid-Atlantic states (New York, New Jersey, Pennsylvania, Delaware, and Maryland) the model could impact a wider patient population. The model could also reveal different high risk red flags for patients from different states. Which would be interesting to insurance companies who have policy holders in different states.

General Dashboard Description

A dashboard to implement the model could be quite simple and easily integrated into the existing EHR. The providers would enter the patient information as usual, and the logistic regression model would work in the background to make a prediction of whether or not the patient is at a high risk for emergency admission based on the given characteristics. If the patient is flagged as high risk, a notification will flash at the time of entering the information, and an icon will appear on the patient's EHR indicating their risk level.

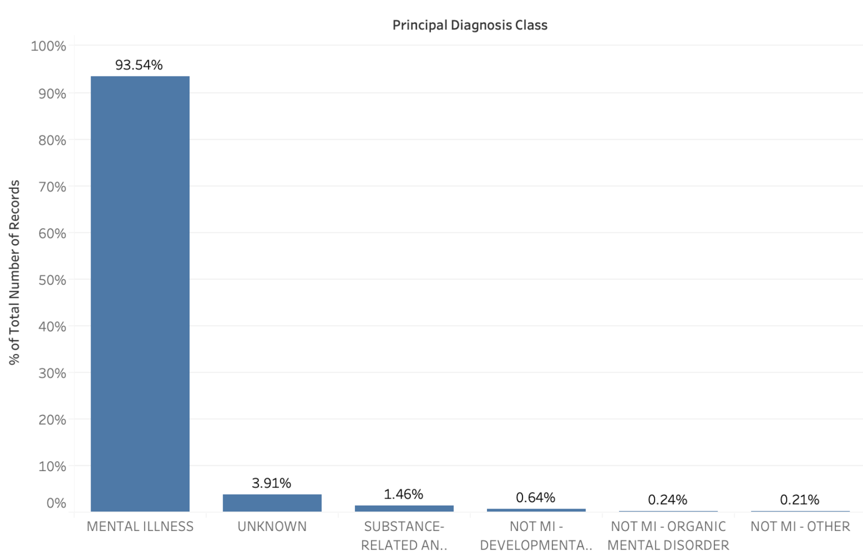
VI. Summary

Mental health disorders have an impact on the quality of life of the patient, the resources of the community, and the costs of the healthcare system. Conducting and analyzing state-wide surveys of the mental health community can help all stakeholders have a better understanding of the factors that make patients at risk for utilizing expensive emergency services. Based on our model, healthcare leaders in New York State should flag patients who have the characteristics indicated by the model, and particularly who are displaying indicators of poverty and/or experiencing homelessness. After gaining buy-in with mental health providers, NYS hospital should integrate this model into their EHR systems as a clinical decision support tool. Providers would then be notified in real-time that they are caring for a high risk patient and coordinate their care properly. After implementing this process, the participating hospitals should create and disseminate an electronic dashboard displaying the patient outcomes for those using the tool. Sharing the dashboard, cost savings, and other results with other hospitals can spread the usefulness of the model and lead to its use in other institutions. In conclusion, using this model as a clinical decision support tool can inspire hospitals and providers to improve the care they are providing to their patients with mental health disorders.

VII. Appendix

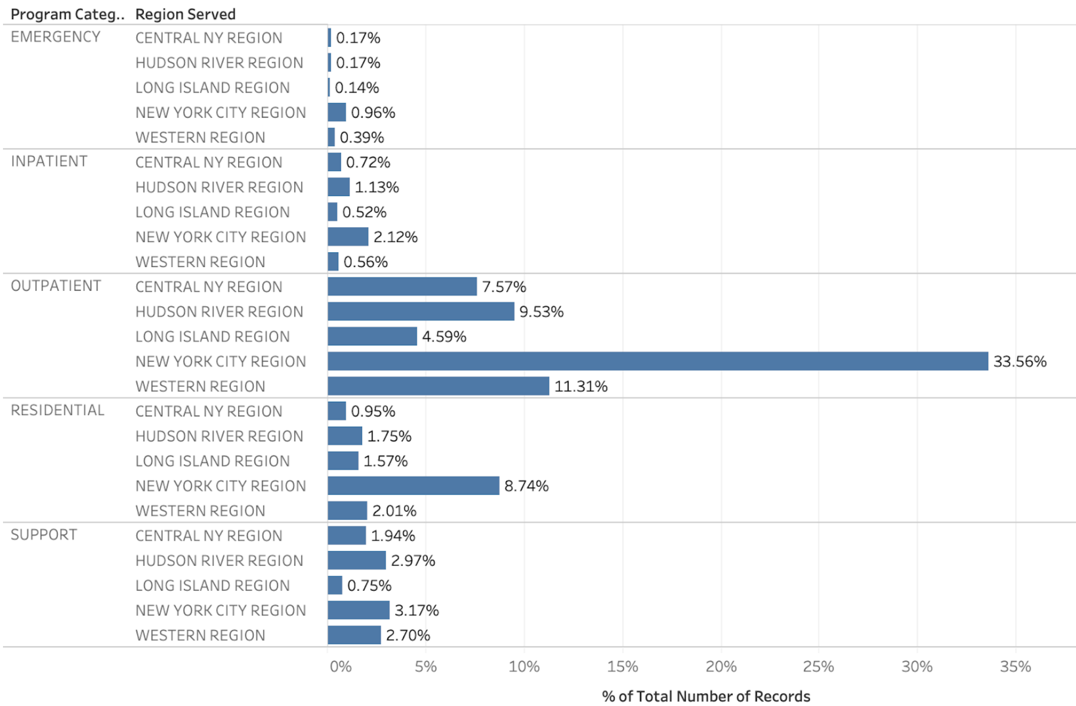
Additional Data Displays for exploring Data Set:

I. Percent of Total Number of Records and Primary Diagnosis:



From figure I (*above*), it can be seen that that Mental Illness was predominantly the primary diagnosis for patients being treated and included in the data set

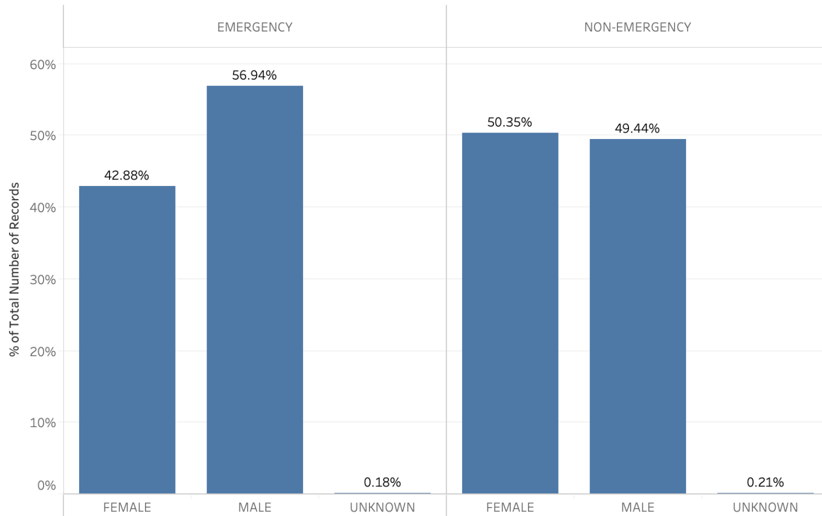
II. Percent of Total Number of Records for Program Category and New York State Region:



From figure II (*above*), it can be seen that the greatest variation in program category/visit type across the New York State Regisions was in the outpatient category.

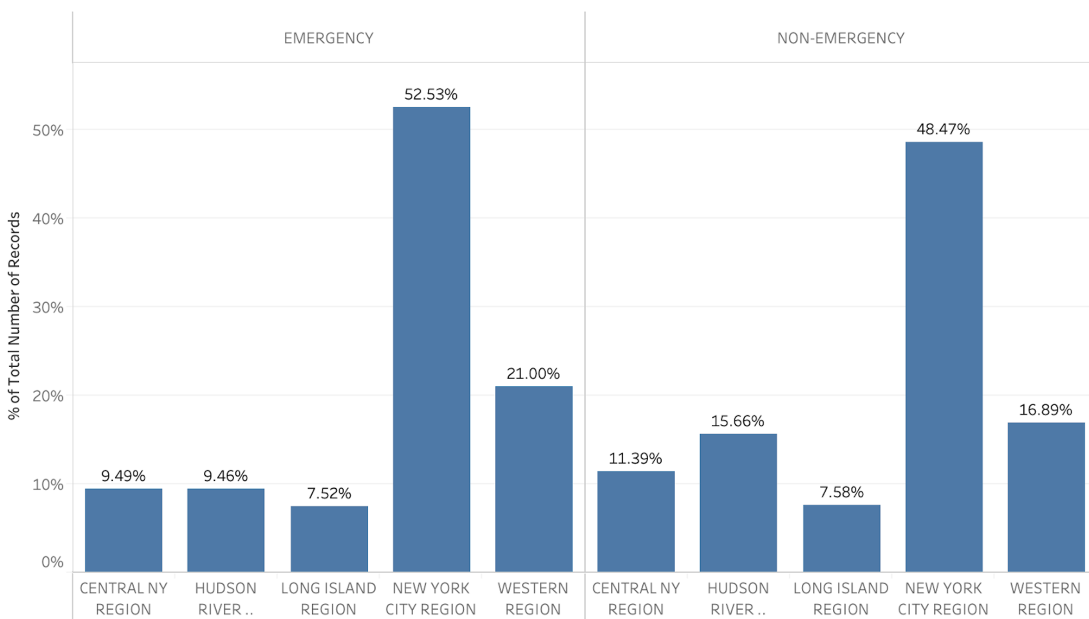
Data Displays for Statistically Significant Variables within the Data Set:

III. Sex and Emergency vs Non-Emergency Visits



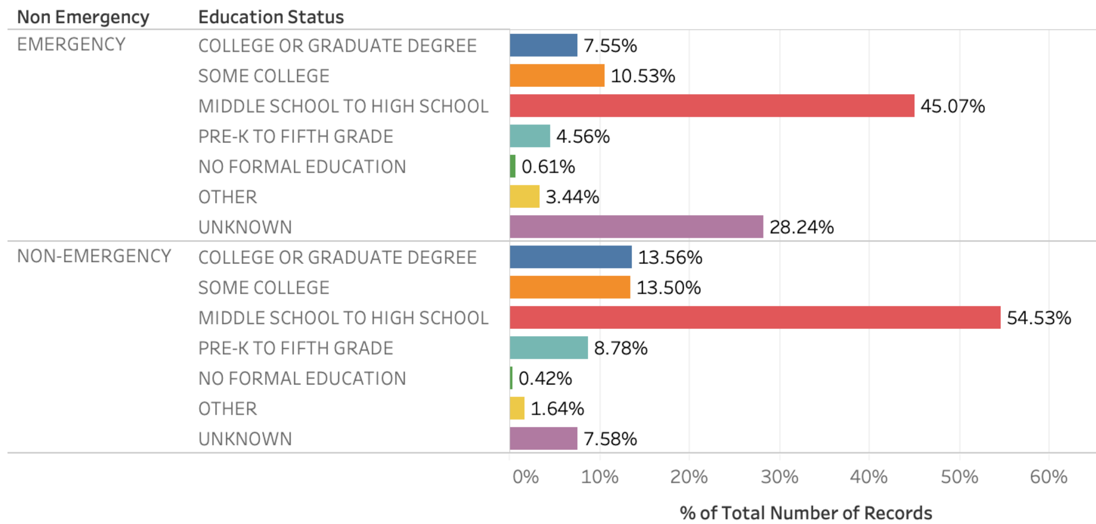
From figure III (*above*), it can be seen that there is a slightly higher percentage of males receiving emergency services than receiving non-emergency services.

IV. New York State Region and Emergency vs Non-Emergency Visits



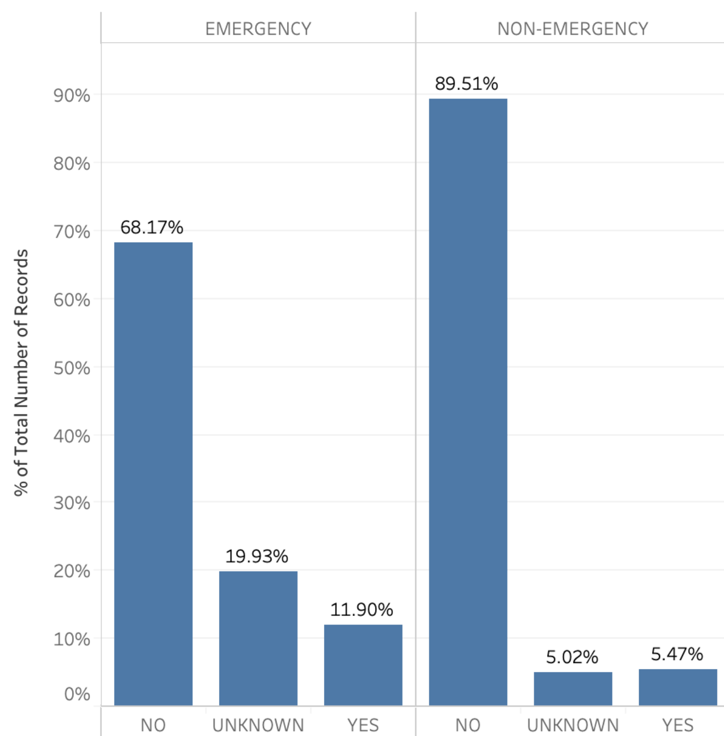
From figure IV (*above*), it can be seen that there are slight variations in the percentage of patients from each region receiving emergency services vs non-emergency services.

V. Education Level and Emergency vs Non-Emergency Visits



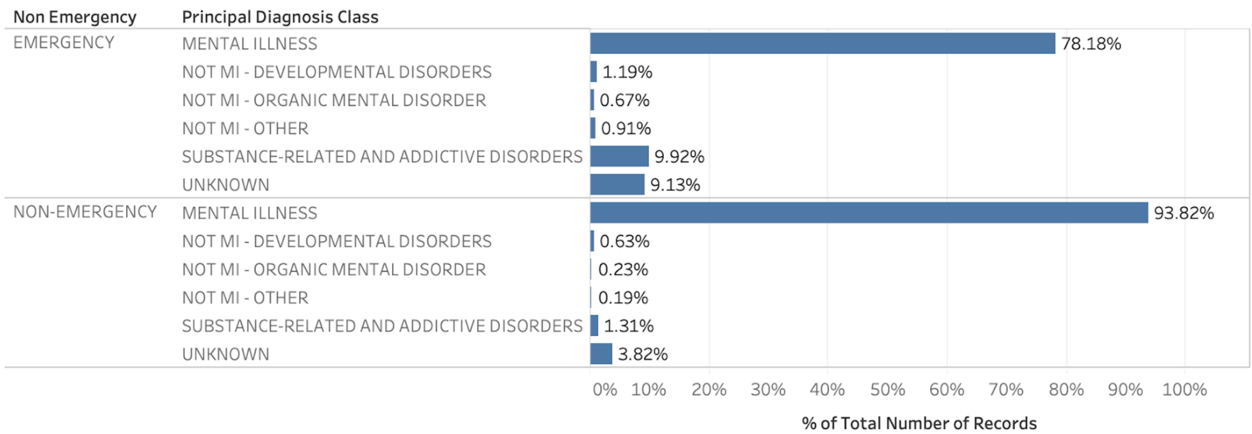
From figure V (*above*), it can be seen that 27.06% of patients with non-emergency services visits had some college education vs 18.1% of patients with emergency service visits had some college education.

VI. Received Smoking Medication and Emergency vs Non-Emergency Visits



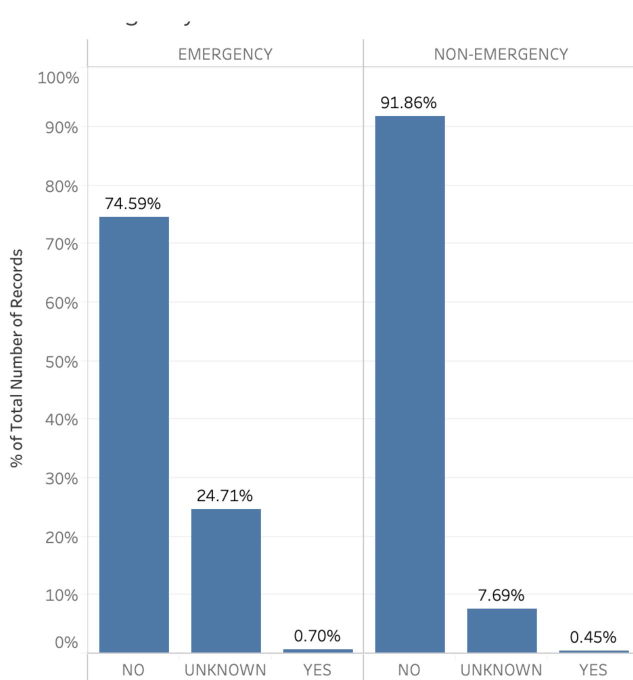
From figure VI (*above*), more patients in the non-emergency services category did not receive any smoking related medication, 89.51% vs 68.17%.

VII. Principal Diagnosis Class and Emergency vs Non-Emergency Visits



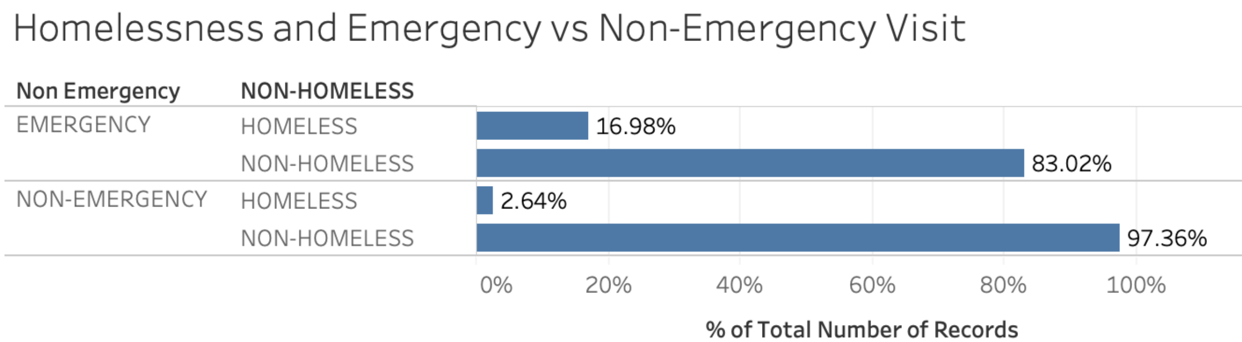
From figure VII (*above*), it can be seen that 93.82% of patients in the non-emergency services group had a principal diagnosis of mental illness. Whereas in the emergency group, 78.81% had a principal diagnosis of mental illness. The next most common diagnosis was Substance-Related and Addictive Disorders.

VIII. Veterans Disability Benefits and Emergency vs Non-Emergency Visits



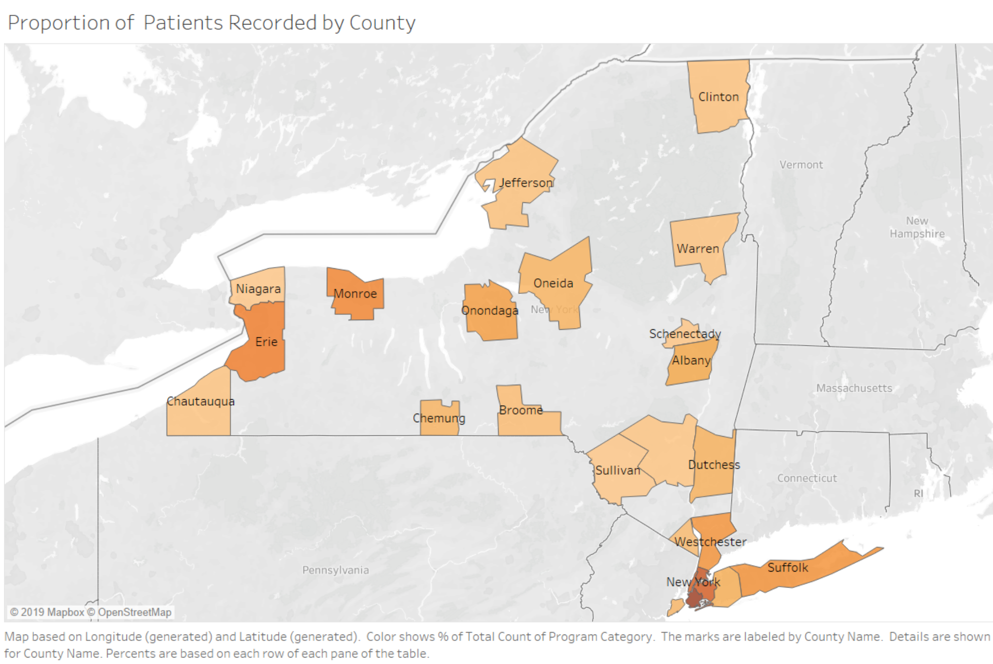
From figure VIII (*above*), it can be seen that more non-emergency visit patients responded "No" to having veterans disability benefits, 91.88% vs 74.59%.

IX. Homelessness and Emergency vs Non-Emergency Visits



From figure IX (*above*), it can be seen that there were a higher number of homeless patients seeking emergency services than non-emergency services, 16.98% vs 2.64%.

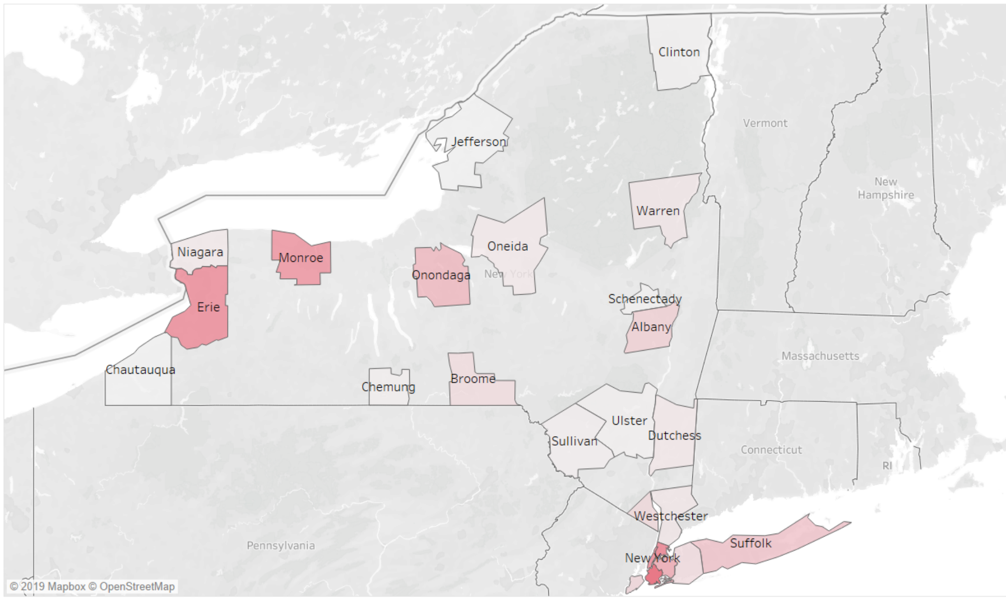
X. Proportion of Patients by County



From figure X (*above*), the counties that have most proportion of emergency patients are: Erie (6.47%), Monroe (5.90%), Suffolk (5.07%), Westchester (4.81%), Queens (9.54%), Bronx (10.67%), Kings (14.02%) and New York (9.17%).

XI. Number of Emergency Patient Visits by County

Number of Emergency Patients by County



Map based on Longitude (generated) and Latitude (generated). Color shows principal diagnal. The marks are labeled by County Name. Details are shown for County Name. The data is filtered on Program Category, which keeps EMERGENCY.

From the figure XI (*above*), through comparison with figure X, it is discovered that counties perform various emergency visit demand. Although the boroughs in NYC contribute close proportion of patients to the data set, emergency visits from the Queens is significantly less than the other boroughs. Westchester has a lower rate of emergency visit, comparing with the total number of patients from there. Besides, Erie and Monroe in western part of the state shows a serious tendency to seek emergency care, that needs our watching.

XII. Results of Multicollinearity Diagnostics

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	.077	.004		18.374	.000		
	RegionServed	.003	.000	.026	11.033	.000	.941	1.063
	IfAdult	.004	.001	.025	8.423	.000	.593	1.687
	IfMale	.000	.000	.003	1.141	.254	.877	1.141
	Transgender	.028	.001	.052	21.056	.000	.883	1.132
	LGBTQ	-.001	.001	-.002	-.986	.324	.881	1.135
	HispanicEthnicity	-.001	.000	-.004	-1.734	.083	.817	1.224
	Race	.000	.000	-.005	-1.842	.065	.900	1.112
	LivingSituation	.003	.001	.014	4.943	.000	.692	1.445
	HouseholdComposition	.002	.000	.015	5.171	.000	.660	1.515
	PreferredLanguage	.000	.001	.001	.583	.560	.824	1.214
	VeteranStatus	.003	.001	.007	2.876	.004	.882	1.134
	EmploymentStatus	.000	.000	.002	.705	.481	.513	1.949
	NumberOfHoursWorkedEachWeek	-.001	.001	-.008	-2.319	.020	.466	2.146
	EducationStatus	-.006	.000	-.067	-25.953	.000	.810	1.234
	SpecialEducationServices	.002	.001	.005	2.209	.027	.946	1.057
	MentalIllness	-.010	.001	-.036	-14.381	.000	.856	1.168
	IntellectualDisability	.000	.001	-.001	-.553	.581	.788	1.268
	AutismSpectrum	-.004	.001	-.012	-4.274	.000	.670	1.493
	OtherDevelopmentalDisability	-.001	.001	-.002	-.778	.436	.698	1.433
	AlcoholRelatedDisorder	.006	.001	.028	9.918	.000	.680	1.471
	DrugSubstanceDisorder	.007	.001	.035	11.974	.000	.643	1.556
	MobilityImpairmentDisorder	-.001	.001	-.004	-1.430	.153	.766	1.306

	HearingVisualImpairment	-.002	.001	-.006	-2.375	.018	.810	1.234
	Hyperlipidemia	-.003	.001	-.013	-4.659	.000	.706	1.416
	HighBloodPressure	-.001	.000	-.005	-1.823	.068	.661	1.512
	Diabetes	-.001	.001	-.003	-.925	.355	.718	1.392
	Obesity	-.002	.000	-.012	-4.633	.000	.746	1.341
	HeartAttack	-.001	.002	-.003	-.765	.444	.316	3.169
	Stroke	-.001	.002	-.003	-.845	.398	.323	3.095
	OtherCardiac	-.002	.001	-.009	-3.000	.003	.637	1.569
	PulmonaryAsthma	-.001	.001	-.004	-1.515	.130	.770	1.299
	AlzheimerorDementia	.003	.002	.007	1.248	.212	.166	6.017
	KidneyDisease	-1.351E-6	.001	.000	-.001	.999	.381	2.625
	LiverDisease	.000	.001	-.001	-.457	.648	.503	1.990
	EndocrineCondition	3.034E-5	.001	.000	.040	.968	.678	1.475
	NeurologicalCondition	-.001	.002	-.002	-.516	.606	.334	2.993
	TraumaticBrainInjury	.000	.001	.000	-.119	.905	.371	2.698
	JointDisease	-.003	.001	-.013	-4.671	.000	.724	1.380
	Cancer	-.001	.001	-.003	-1.027	.304	.486	2.059
	OtherChronicMedCondition	.001	.000	.008	2.791	.005	.713	1.403
	NoChronicMedCondition	.002	.001	.011	3.206	.001	.435	2.297
	UnknownChronicMedCondition	.002	.002	.011	1.088	.276	.058	17.330
	Smokes	-5.668E-5	.000	.000	-.126	.899	.618	1.617
	ReceivedSmokingMedication	.015	.001	.057	20.753	.000	.715	1.398
	ReceivedSmokingCounseling	-.002	.001	-.009	-3.058	.002	.578	1.731
	SeriousMentalIllness	.000	.000	.002	.715	.474	.806	1.241
	PrincipalDiagnosisClass	.007	.001	.035	13.913	.000	.850	1.176
	AdditionalDiagnosisClass	.002	.000	.021	8.637	.000	.930	1.075
	SSICashAssistance	-.003	.000	-.023	-8.364	.000	.743	1.346
	SSDICashAssistance	-.001	.000	-.008	-3.029	.002	.695	1.439
	VeteransDisabilityBenefits	.012	.002	.028	5.997	.000	.252	3.962
	VeteransCashAssistance	.011	.002	.023	5.059	.000	.254	3.941
	PublicAssistanceCashProgram	-.002	.000	-.011	-4.207	.000	.861	1.162
	OtherCashBenefits	-.001	.001	-.006	-2.309	.021	.818	1.223
	MedicaidandMedicareInsurance	.001	.001	.003	.508	.612	.168	5.942
	NoInsurance	.003	.001	.012	3.043	.002	.355	2.816
	UnknownInsuranceCoverage	.008	.001	.020	7.012	.000	.644	1.552
	MedicaidInsurance	-.004	.001	-.023	-4.684	.000	.228	4.383
	MedicaidManagedInsurance	.001	.000	.003	1.216	.224	.873	1.146
	MedicareInsurance	-.001	.001	-.006	-1.054	.292	.160	6.234
	PrivateInsurance	.002	.001	.008	2.348	.019	.425	2.356
	ChildHealthPlusInsurance	.005	.001	.011	4.079	.000	.700	1.428
	OtherInsurance	-.002	.001	-.007	-2.520	.012	.787	1.270
	CriminalJusticeStatus	-.003	.001	-.017	-6.118	.000	.741	1.350

References:

About OMH. (n.d.). Retrieved December 1, 2019, from https://www.omh.ny.gov/omhweb/about/

Bringewatt, E. H., & Gershoff, E. T. (2010). Falling through the cracks: Gaps and barriers in the mental health system for America's disadvantaged children. *Children and Youth Services Review*. https://doi.org/10.1016/j.childyouth.2010.04.021

Castner, J., Wu, Y. W. B., Mehrok, N., Gadre, A., & Hewner, S. (2015). Frequent emergency department utilization and behavioral health diagnoses. *Nursing Research*. https://doi.org/10.1097/NNR.0000000000000065

Doran, K. M., Raven, M. C., & Rosenheck, R. A. (2013). What drives frequent emergency department use in an integrated health system? National data from the veterans health administration. *Annals of Emergency Medicine*. https://doi.org/10.1016/j.annemergmed.2013.02.016

Jorm, A. F. (2000). Mental health literacy: Public knowledge and beliefs about mental disorders. *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.177.5.396

Maust, D. T., Oslin, D. W., & Marcus, S. C. (2013). Mental health care in the accountable care organization. *Psychiatric Services*. https://doi.org/10.1176/appi.ps.201200330

Mental Health By the Numbers | NAMI: National Alliance on Mental Illness. (n.d.). Retrieved November 18, 2019, from https://www.nami.org/learn-more/mental-health-by-the-numbers

Murray, C. J. L., Bulchis, A., Birger, M., Dieleman, J. L., Baral, R., Bui, A. L., … Tobias, M. I. (2016). Us spending on personal health care and public health, 1996-2013.(Report). *JAMA, The Journal of the American Medical Association*, *316*(24), 2627. https://doi.org/10.1001/jama.2016.16885

NYS Patient Characteristics Survey (PCS): 2015 | Kaggle. (n.d.). Retrieved November 18, 2019, from https://www.kaggle.com/new-york-state/nys-patient-characteristics-survey-pcs-2015

Primary Care, Behavioral Health, and Public Health: Partners in Reducing Me...: EBSCOhost. (n.d.). Retrieved December 1, 2019, from http://web.a.ebscohost.com.ezproxy.neu.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=df4999ba-d054-40bc-964c-fcbaba741e64%40sessionmgr4008

Smith, T. E., Erlich, M. D., & Sederer, L. I. (2013). Integrating general medical and behavioral health care: The New York state perspective. *Psychiatric Services*. https://doi.org/10.1176/appi.ps.201300197