1. **Kelly**
   - Within an illness and age group, can we look at geographic distribution of treatment settings
   - Quality Readings related to Outcomes?
   - Are Quality Ratings related to Settings?
   - Ways to combine client level data with quality level data
2. **Gretchen**
   - Larger Scale Approach
   - States - all cases - against Poverty Dataset
3. **Samy**
   - Multinomial Logistic Regression
   - Correlation Matrix/Heat Map
   - K Nearest Neighbors Classification
   - Analysis over multiple years (2013-2018)
   - Binary Cross entropy classification
   - Advanced Visualization with Website

Report

Introduction: Describe your project scenario. Starting out, what did you hope to accomplish/learn?

At the beginning of 2021, the COVID-19 pandemic took the world by storm, exposing humans' vulnerability to a viral spillover event years in the making. Although forewarned by many, few took heed of the recommendations and before the magnitude of the problem was acknowledged, it was far too late. However, COVID-19 also uncovered another insidious disease that has been looming quietly for many years: that of mental health. Compounded with the isolation brought by quarantine, in addition to economic strife, illness, death, and a dearth of support systems, the incidence of mental and behavioral disorders has increased and is only projected to worsen. In light of these findings, and to take prophylactic action against another potential global catastrophe, our research team sought to investigate the trajectory of mental illness in recent years and identify factors that demonstrate statistically significant relationships with mental and behavioral disorders. The datasets evaluated were produced from the Substance Abuse & Mental Health Services Administration of the U.S Department of Health and Human Services. The specific datasets utilized included Mental health Client-level Data (described in the Data Section), ****, and ***.

The Data: Describe your data set and its significance. Where did you obtain this data set from? Why did you choose the data set that you did? Indicate if you carried out any preprocessing/data cleaning/outlier removal, and so on to sanitize your data.

The 2013-2018 Mental Health Client Level Data Report is a large scale effort who inception dates back to 2002. At that time, URS data reporting began to initially collect system performance measures, which was later followed by client outcomes at an aggregate level. As the datasets on an annual basis are each very large and , in order to narrow the scope of this assignment, we focused on the 2018 Year Dataset. However, as is discussed later, we did do some comparative analysis of particular trends for each year from 2013-2018. We chose our dataset because of its large size, relevance to the topic. In terms of data cleaning, the dataset was already precleaned. Yet, for sake of thoroughness, we dropped all null and NaN values, in addition to duplicates, from the dataset using Pandas and compared the size, which demonstrated no change. This indicates that each row in the Client Level Data most likely represents a unique patient. In terms of outliers, we used the exploratory data analysis process as a means to get a sense of whether. However, as most data points were either binary or ordinal in nature, there was an inherent lack of outliers.

Our data was somewhat limited in the following facets. Although the year was reported, months and days were excluded. These details would provide supplemental benefit due to the seasonality of depression. For example, in a study by Ayers et al. in 2013, Google mental health queries monitored from 2006 to 2010 revealed seasonal patterns for all mental health queries, with winter peaks and summer troughs (14% difference in the United States; 11% difference for Australia).

Experimental Design: Describe briefly your process, starting from where you obtained your data all the way to means of obtaining results/output.

Data was obtained from the https://wwwdasis.samhsa.gov/dasis2/mhcld.htm, in delimited form (comma separated values). There were six datasets available for analysis regarding client level data, and while each dataset was used, the focus for this investigation was on the most recent one available, 2018.

Beyond the original specifications: Highlight clearly what things you did that went beyond the original specifications. That is, discuss what you implemented that would count toward the extra-credit portion of this project (see section below).

Beyond the original specifications, our group pursued multiple means to enhance data visualization, user experience and complexity and robustness of analysis. In terms of visualization, in order to explore python web frameworks, we built a web application using the Flask library. For advanced statistical analysis, we implemented machine learning algorithms to . Using the sk-learn library, we built a prediction classifier using the K-Nearest Neighbors algorithm. K nearest neighbors works to classify unlabeled observations by assigning them to the clo with the most similar labeled examples. We allowed users to experiment with feature selection towards their desire label prediction, in addition providing the option modulate the number of neighbors to assess the model accuracy.

Testing: Describe what testing you did. Describe the unit tests that you wrote. Show a sample run of 1 or 2 of your tests (screen captures or copy-and-paste is fine).

Conclusions: Summarize your findings, explain how these results could be used by others (if applicable), and describe ways you could improve your program. You could describe ways you might like to expand the functionality of your program if given more time.

Future studies could involve additional research on the historical c. . Moreover, it will be interesting to