MENTAL HEALTHCARE

ASSIGNMENT 10.3

Siddhartha Bhaumik

Introduction

- Mental health disorders are one of the leading health issues in the United States and it affects almost 10% of the population. This has significantly increased since the beginning of Covid19 pandemic and is not just seen in adults but children and young adults as well.
- Mental health awareness is another significant problem as lot of people are unaware of their problems until those turn into severe health issues like anxiety, depression, and other life-threatening risks.
- Also, many people don't know how to or from where to get help. Or they feel ashamed talking about their condition as that may impact their relationship both personally and professionally as well as their growth in life and at workplace.
- Another important thing to note is that in U.S there is a shortage in certified mental health professionals.
- Data Science/Artificial Intelligence can play an important part here by bridging some of the current gaps in Mental Healthcare sector.

Research Questions

1. With so much patient data now available digitally like health reports, lab reports, social media interactions, etc.,different AI tools and techniques can analyze patient's data and flag physical and mental states. This can help in early detection and remedies.
2. Lot of people are hesitant to open in front of doctors and therapists because of stigma or fear of being judged.
   People tend to trust a robot more since it won't judge, is unbiased and can provide instant answers to health-related questions. Several fitness gadgets are in market which can track your sleep, heart rate, blood pressure, etc. and can share that information through apps which can further evaluate and predict your overall health. I see some AI desktop/mobile apps in market for self-assessment and therapy which can be very useful if enhanced further and marketed properly.
3. Machine learning and Deep learning can provide greater accuracy in diagnosing mental health conditions and predicting patient outcomes. So, they can assist doctors and therapists in providing better treatment.
4. A major issue which I see with Data Science/AI in mental health sector is privacy. All sensitive information related to a particular person is available to an AI software and if misused or breached can cause greater damage physically, mentally, and financially.
5. Mental health is often overlooked which many times leads up to serious health issues. This is not specific to any age group as all are vulnerable and not specific to any particular location as it can happen anywhere from home,school,workplace, etc. So, self awareness as well as guidance,support and counselling is needed at the earliest possible stage.

Approach

I plan to focus most on awareness because that's what I think is lacking globally and specially more in third world nations.

Timely treatment is the key to success.
Every person has a mobile device now with access to internet. With the help of Data Science/AI, mental health apps can be created which can help a person with self-assessment, morale boost with positive conversations, cognitive therapy, mind games and other related stuffs.
While doing initial research I found some AI apps already in the market which is a good sign but these can be enhanced further with additional features/capabilities.

Approach Outcome

Awareness is the key but there are other aspects as well like proper medical treatments and therapies for mental patients. Data Science can definitely help in these areas but my focus is more towards awareness.
So, my approach partially addresses this problem.

Datasets/Citations

- "COVID-19 and Mental Health Search Terms" dataset from Kaggle. https://www.kaggle.com/datasets/luckybro/mental-health-search-term The search interest of mental health related terms on Google before and after the outbreak of COVID-19 pandemic
  reveals how public's concern is affected by the pandemic, and its impact to mental health of people around the world.
- "Mental Health in Tech Survey" dataset from Kaggle https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey This dataset is from a 2014 survey that measures attitudes towards mental health and frequency of mental health
  disorders in the tech workplace.
- "Any Mental Illness in the Past Year among Adults Aged 18 or Older, by State: 2018-2019" dataset from SAMHDA.gov https://pdas.samhsa.gov/saes/state This dataset is maintained by 'Substance Abuse & Mental Health Data Archive' government agency and contains any type of mental health related issues in adults aged 18 and older for the year 2018-2019.

Required Libraries

```
library(ggplot2)
library(pastecs)
library(dplyr)
library(purrr)
library(stringr)
library(lm.beta)
library(tidyverse)
library(corrplot)
library(car)
theme_set(theme_minimal())
```

Plots and Tables

I believe histograms and box plots will be useful in visualizing the data.
Regarding tables, I plan to explore 'gt' package as it looks simple yet powerful.

```
## Load required package
library(ggplot2)
library(pastecs)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:pastecs':
##
##     first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(purrr)
library(stringr)
theme_set(theme_minimal())

# Q1 Data importing and cleaning steps are explained in the text and follow a
# logical process.Outline your data preparation and cleansing steps.

## Load 'Any Mental Illness in the Past Year Data' from 51 US states to
state_any_mental_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/map_data.csv")

## Viewing Sample data from 'Any mental illness/past year' dataset
head((state_any_mental_df))
```

```
##                              outcome   age_group year_pair      state estimate
## 1 Any Mental Illness in the Past Year 18 or Older   2018-19    Alabama 0.212901
## 2 Any Mental Illness in the Past Year 18 or Older   2018-19     Alaska 0.214692
## 3 Any Mental Illness in the Past Year 18 or Older   2018-19    Arizona 0.200635
## 4 Any Mental Illness in the Past Year 18 or Older   2018-19   Arkansas 0.203352
## 5 Any Mental Illness in the Past Year 18 or Older   2018-19 California 0.194866
## 6 Any Mental Illness in the Past Year 18 or Older   2018-19   Colorado 0.231950
##   ci_lower ci_upper
## 1 0.190994 0.236585
## 2 0.194543 0.236314
## 3 0.179126 0.224021
## 4 0.182782 0.225598
## 5 0.184540 0.205624
## 6 0.209919 0.255546
```

```
## Load 'Major Depressive Episode in the Past Year Data' from 51 US states to
state_dep_mental_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/map_data_rcrd.csv

## Viewing Sample data from 'Major depressive episode/past year' dataset
head((state_dep_mental_df))
```

```
##                                  outcome   age_group year_pair      state
## 1 Major Depressive Episode in the Past Year 18 or Older   2018-19    Alabama
```

```
## 2 Major Depressive Episode in the Past Year 18 or Older   2018-19     Alaska
## 3 Major Depressive Episode in the Past Year 18 or Older   2018-19    Arizona
## 4 Major Depressive Episode in the Past Year 18 or Older   2018-19   Arkansas
## 5 Major Depressive Episode in the Past Year 18 or Older   2018-19 California
## 6 Major Depressive Episode in the Past Year 18 or Older   2018-19   Colorado
##   estimate ci_lower ci_upper
## 1 0.081327 0.069224 0.095330
## 2 0.085115 0.073037 0.098976
## 3 0.078124 0.066463 0.091629
## 4 0.078977 0.067765 0.091862
## 5 0.071717 0.065501 0.078473
## 6 0.084633 0.072943 0.097998
```

```
## Load 'Received Mentalhealth services in Past Year Data' from 51 US states to
state_rcvd_mental_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/map_data_dep.csv

## Viewing Sample data from 'Received Mental health services/past year' dataset
head((state_rcvd_mental_df))
```

```
##                                         outcome   age_group year_pair
## 1 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 2 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 3 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 4 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 5 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 6 Received Mental Health Services in the Past Year 18 or Older   2018-19
##        state estimate ci_lower ci_upper
## 1    Alabama 0.156306 0.137093 0.177657
## 2     Alaska 0.170857 0.151458 0.192176
## 3    Arizona 0.136570 0.118184 0.157306
## 4   Arkansas 0.167128 0.148040 0.188134
## 5 California 0.130828 0.121788 0.140431
## 6   Colorado 0.187090 0.167010 0.208979
```

```
## Its a clean dataset with relevant information from each US state.
## So, Nothing to filter out.

## Load the 'COVID-19 and Mental Health Search Terms data'  to
covid19_world_df <- readxl::read_excel("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/search_ter

covid19_us_df <- readxl::read_excel("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/search_term_u
## The search interest of mental health related terms on Google before and
## after the outbreak of COVID-19 pandemic reveals how public's concern is
## affected by the pandemic, and its impact to mental health of people around the world.
## The mental health related search terms are "mental health", "depression",
## "anxiety", "ocd", "obsessive compulsive disorder", "insomnia", "panic attack",
## "counseling", "psychiatrist".
## Search interest is indicated by a number between 0 and 100, where 100 means
## the most popular point of time(by week), 1 means the least, and 0 no enough data.

## Viewing Worldwide Sample data
head((covid19_world_df))
```

```
## # A tibble: 6 x 10
##   Week                depression anxiety `obsessive compulsive ~`   ocd insomnia
##   <dttm>                   <dbl>   <dbl>                    <dbl> <dbl>    <dbl>
```

```
## 1 2019-06-16 00:00:00         81      87                        61    58       75
## 2 2019-06-23 00:00:00         76      85                        75    60       76
## 3 2019-06-30 00:00:00         73      83                        63    57       70
## 4 2019-07-07 00:00:00         80      90                        74    62       77
## 5 2019-07-14 00:00:00         80      90                        72    65       76
## 6 2019-07-21 00:00:00         79      92                        74    59       77
## # ... with 4 more variables: `panic attack` <dbl>, `mental health` <dbl>,
## #   counseling <dbl>, psychiatrist <dbl>
```

```
## Viewing US Sample data
head((covid19_us_df))
```

```
## # A tibble: 6 x 10
##   Week                depression anxiety `obsessive compulsive ~`   ocd insomnia
##   <dttm>                   <dbl>   <dbl>                     <dbl> <dbl>    <dbl>
## 1 2019-06-16 00:00:00         70      89                        37    69       77
## 2 2019-06-23 00:00:00         70      91                        51    73       83
## 3 2019-06-30 00:00:00         63      87                        41    70       74
## 4 2019-07-07 00:00:00         74      92                        60    74       84
## 5 2019-07-14 00:00:00         70      92                        70    77       81
## 6 2019-07-21 00:00:00         75      93                        42    72       82
## # ... with 4 more variables: `panic attack` <dbl>, `mental health` <dbl>,
## #   counseling <dbl>, psychiatrist <dbl>
```

```
## Its a clean dataset with relevant information. So, Nothing to filter out.

## Load the 'Mental Health in Tech Survey' to
tech_survey_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/survey.csv")

## This dataset is from a 2014 survey that measures attitudes towards mental
## health and frequency of mental health disorders in the workplace.

## Viewing Sample data
head((tech_survey_df))
```

```
##             Timestamp Age Gender        Country state self_employed
## 1 2014-08-27 11:29:31  37 Female  United States    IL          <NA>
## 2 2014-08-27 11:29:37  44      M  United States    IN          <NA>
## 3 2014-08-27 11:29:44  32   Male         Canada  <NA>          <NA>
## 4 2014-08-27 11:29:46  31   Male United Kingdom  <NA>          <NA>
## 5 2014-08-27 11:30:22  31   Male  United States    TX          <NA>
## 6 2014-08-27 11:31:22  33   Male  United States    TN          <NA>
##   family_history treatment work_interfere   no_employees remote_work
## 1             No       Yes          Often           6-25          No
## 2             No        No         Rarely More than 1000          No
## 3             No        No         Rarely           6-25          No
## 4            Yes       Yes          Often         26-100          No
## 5             No        No          Never        100-500         Yes
## 6            Yes        No      Sometimes           6-25          No
##   tech_company   benefits care_options wellness_program  seek_help  anonymity
## 1          Yes        Yes     Not sure               No        Yes        Yes
## 2           No Don't know           No       Don't know Don't know Don't know
## 3          Yes         No           No               No         No Don't know
## 4          Yes         No          Yes               No         No         No
## 5          Yes        Yes           No       Don't know Don't know Don't know
## 6          Yes        Yes     Not sure               No Don't know Don't know
```

```
##                leave mental_health_consequence phys_health_consequence
## 1      Somewhat easy                         No                       No
## 2         Don't know                      Maybe                       No
## 3 Somewhat difficult                         No                       No
## 4 Somewhat difficult                        Yes                      Yes
## 5         Don't know                         No                       No
## 6         Don't know                         No                       No
##      coworkers supervisor mental_health_interview phys_health_interview
## 1 Some of them        Yes                      No                 Maybe
## 2           No         No                      No                    No
## 3          Yes        Yes                     Yes                   Yes
## 4 Some of them         No                   Maybe                 Maybe
## 5 Some of them        Yes                     Yes                   Yes
## 6          Yes        Yes                      No                 Maybe
##   mental_vs_physical obs_consequence comments
## 1                Yes              No     <NA>
## 2         Don't know              No     <NA>
## 3                 No              No     <NA>
## 4                 No             Yes     <NA>
## 5         Don't know              No     <NA>
## 6         Don't know              No     <NA>
```

```
tech_survey_upd_df <- tech_survey_df %>% filter(Country == "United States") %>% filter(Age > 12) %>% sel

head(tech_survey_upd_df)
```

```
##   Age Gender family_history treatment remote_work work_interfere   benefits
## 1  37 Female             No       Yes          No          Often        Yes
## 2  44      M             No        No          No         Rarely Don't know
## 3  31   Male             No        No         Yes          Never        Yes
## 4  33   Male            Yes        No          No      Sometimes        Yes
## 5  35 Female            Yes       Yes         Yes      Sometimes         No
## 6  42 Female            Yes       Yes          No      Sometimes        Yes
##   wellness_program  seek_help  anonymity mental_health_consequence
## 1               No        Yes        Yes                       No
## 2       Don't know Don't know Don't know                    Maybe
## 3       Don't know Don't know Don't know                       No
## 4               No Don't know Don't know                       No
## 5               No         No         No                    Maybe
## 6               No         No         No                    Maybe
##   obs_consequence
## 1              No
## 2              No
## 3              No
## 4              No
## 5              No
## 6              No
```

```
# Standardize Gender with Male, Female, Other
tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "M" | tech_survey_upd_df["Gender"] == "m"
                             | tech_survey_upd_df["Gender"] == "male" | tech_survey_upd_df["Gender"] ==
                             | tech_survey_upd_df["Gender"] == "Cis male" | tech_survey_upd_df["Gender"]
                             | tech_survey_upd_df["Gender"] == "Male-ish" | tech_survey_upd_df["Gender"]
                             | tech_survey_upd_df["Gender"] == "Man" | tech_survey_upd_df["Gender"] ==
                             | tech_survey_upd_df["Gender"] == "Malr" | tech_survey_upd_df["Gender"] ==
```

```
                              | tech_survey_upd_df["Gender"] == "Mal" | tech_survey_upd_df["Gender"] == 
                              | tech_survey_upd_df["Gender"] == "maile" | tech_survey_upd_df["Gender"] ==


tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "F" | tech_survey_upd_df["Gender"] == "f"
                              | tech_survey_upd_df["Gender"] == "female" | tech_survey_upd_df["Gender"] =
                              | tech_survey_upd_df["Gender"] == "Cis Female" | tech_survey_upd_df["Gender
                              | tech_survey_upd_df["Gender"] == "Woman" | tech_survey_upd_df["Gender"] ==
                              | tech_survey_upd_df["Gender"] == "Femake" | tech_survey_upd_df["Gender"] =
                              | tech_survey_upd_df["Gender"] == "Female (trans)"] <- "Female"

tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "Female (trans)" | tech_survey_upd_df["Gen
                              | tech_survey_upd_df["Gender"] == "non-binary" | tech_survey_upd_df["Gender
                              | tech_survey_upd_df["Gender"] == "Genderqueer" | tech_survey_upd_df["Gend
                              | tech_survey_upd_df["Gender"] == "Trans woman" ] <- "Others"

head(tech_survey_upd_df)
```

```
##   Age Gender family_history treatment remote_work work_interfere   benefits
## 1  37 Female             No       Yes          No          Often        Yes
## 2  44   Male             No        No          No         Rarely Don't know
## 3  31   Male             No        No         Yes          Never        Yes
## 4  33   Male            Yes        No          No      Sometimes        Yes
## 5  35 Female            Yes       Yes         Yes      Sometimes         No
## 6  42 Female            Yes       Yes          No      Sometimes        Yes
##   wellness_program  seek_help  anonymity mental_health_consequence
## 1               No        Yes        Yes                        No
## 2       Don't know Don't know Don't know                     Maybe
## 3       Don't know Don't know Don't know                        No
## 4               No Don't know Don't know                        No
## 5               No         No         No                     Maybe
## 6               No         No         No                     Maybe
##   obs_consequence
## 1              No
## 2              No
## 3              No
## 4              No
## 5              No
## 6              No
```

```
## Only considering survey results from United States as its the majority.
## Noticed some negative numbers under 'Age' column which I will filter out.
## Under 'Gender' column, I see lot of variation and spelling error like Male,
## Mail,maile, M, Cis Male, Female, Cis Female, etc. I will make it consistent as Male, Female, Other.
## Dropped some columns like State, No of Employee, Tech company, etc. as
## I don't think they add much value.
## With a clean dataset, show what the final data set looks like.
## sHowever, do not print off a data frame with 200+ rows; show me the data in the most condensed form 
## Basically I am looking for how many people opted for 'Treatment'.

# Q2 With a clean dataset, show what the final data set looks like.
# However, do not print off a data frame with 200+ rows; show me the data in the
# most condensed form possible.
```

```
## Viewing Sample data from 'Any mental illness/past year' dataset
head(state_any_mental_df)
```

```
##                               outcome   age_group year_pair      state estimate
## 1 Any Mental Illness in the Past Year 18 or Older   2018-19    Alabama 0.212901
## 2 Any Mental Illness in the Past Year 18 or Older   2018-19     Alaska 0.214692
## 3 Any Mental Illness in the Past Year 18 or Older   2018-19    Arizona 0.200635
## 4 Any Mental Illness in the Past Year 18 or Older   2018-19   Arkansas 0.203352
## 5 Any Mental Illness in the Past Year 18 or Older   2018-19 California 0.194866
## 6 Any Mental Illness in the Past Year 18 or Older   2018-19   Colorado 0.231950
##   ci_lower ci_upper
## 1 0.190994 0.236585
## 2 0.194543 0.236314
## 3 0.179126 0.224021
## 4 0.182782 0.225598
## 5 0.184540 0.205624
## 6 0.209919 0.255546
```

```
## Viewing Sample data from 'Major depressive episode/past year' dataset
head(state_dep_mental_df)
```

```
##                                  outcome   age_group year_pair      state
## 1 Major Depressive Episode in the Past Year 18 or Older   2018-19    Alabama
## 2 Major Depressive Episode in the Past Year 18 or Older   2018-19     Alaska
## 3 Major Depressive Episode in the Past Year 18 or Older   2018-19    Arizona
## 4 Major Depressive Episode in the Past Year 18 or Older   2018-19   Arkansas
## 5 Major Depressive Episode in the Past Year 18 or Older   2018-19 California
## 6 Major Depressive Episode in the Past Year 18 or Older   2018-19   Colorado
##   estimate ci_lower ci_upper
## 1 0.081327 0.069224 0.095330
## 2 0.085115 0.073037 0.098976
## 3 0.078124 0.066463 0.091629
## 4 0.078977 0.067765 0.091862
## 5 0.071717 0.065501 0.078473
## 6 0.084633 0.072943 0.097998
```

```
## Viewing Sample data from 'Received Mental health services/past year' dataset
head(state_rcvd_mental_df)
```

```
##                                       outcome   age_group year_pair
## 1 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 2 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 3 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 4 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 5 Received Mental Health Services in the Past Year 18 or Older   2018-19
## 6 Received Mental Health Services in the Past Year 18 or Older   2018-19
##        state estimate ci_lower ci_upper
## 1    Alabama 0.156306 0.137093 0.177657
## 2     Alaska 0.170857 0.151458 0.192176
## 3    Arizona 0.136570 0.118184 0.157306
## 4   Arkansas 0.167128 0.148040 0.188134
## 5 California 0.130828 0.121788 0.140431
## 6   Colorado 0.187090 0.167010 0.208979
```

```
## Viewing US Sample data related to 'Covid19 & Mental health effect/awareness'
head(covid19_us_df)
```

```
## # A tibble: 6 x 10
##   Week                depression anxiety `obsessive compulsive ~`   ocd insomnia
##   <dttm>                   <dbl>   <dbl>                     <dbl> <dbl>    <dbl>
## 1 2019-06-16 00:00:00         70      89                        37    69       77
## 2 2019-06-23 00:00:00         70      91                        51    73       83
## 3 2019-06-30 00:00:00         63      87                        41    70       74
## 4 2019-07-07 00:00:00         74      92                        60    74       84
## 5 2019-07-14 00:00:00         70      92                        70    77       81
## 6 2019-07-21 00:00:00         75      93                        42    72       82
## # ... with 4 more variables: `panic attack` <dbl>, `mental health` <dbl>,
## #   counseling <dbl>, psychiatrist <dbl>
```

```
## Viewing Mental health in US Tech industry
head(tech_survey_upd_df)
```

```
##   Age Gender family_history treatment remote_work work_interfere   benefits
## 1  37 Female             No       Yes          No          Often        Yes
## 2  44   Male             No        No          No         Rarely Don't know
## 3  31   Male             No        No         Yes          Never        Yes
## 4  33   Male            Yes        No          No      Sometimes        Yes
## 5  35 Female            Yes       Yes         Yes      Sometimes         No
## 6  42 Female            Yes       Yes          No      Sometimes        Yes
##   wellness_program  seek_help  anonymity mental_health_consequence
## 1               No        Yes        Yes                        No
## 2       Don't know Don't know Don't know                     Maybe
## 3       Don't know Don't know Don't know                        No
## 4               No Don't know Don't know                        No
## 5               No         No         No                     Maybe
## 6               No         No         No                     Maybe
##   obs_consequence
## 1              No
## 2              No
## 3              No
## 4              No
## 5              No
## 6              No
```

```
# Q3 What do you not know how to do right now that you need to learn to import
# and cleanup your dataset?

## I am not able to replace the values in Gender column for Male, Female and
## Others in a simple and short steps.

# Q4 Discuss how you plan to uncover new information in the data that is not self-evident.

## First I want to see if Covid19 pandemic has a direct impact on mental health cases.
## Also, if its related to other datasets like Major depressive episode or people
## received treatment in past year.
## Tech survey data is older and pre-pandemic, so there won't be any relationship
## with Covid19 but I want to see how many people took any treatment, how much
## workplace and work environment is responsible for the mental stress,
## does people with family history in mental health tend to be more aware and
## willing to take treatment compared to rest, Is there a gender bias and other stuffs.
```

```
# Q5 What are different ways you could look at this data to answer the questions you want to answer?

## I am planning to use generalized linear model or Correlation coefficient to
## explore the data and relationship between different factors and the outcome
## of mental health treatment and awareness.

# Q6 Do you plan to slice and dice the data in different ways, create new variables,
# or join separate data frames to create new summary information? Explain.

## Yes, I plan to join dataframes 'state_any_mental_df' and 'state_rcvd_mental_df'
## to see if I can find out the percentage of people who reported some kind of
## mental illness and out of which how much percentage actually went for the treatment.
## I also plan to see the US states with highest vs lowest cases reported.
## For the tech survey data, I want to slice based on Gender and other columns
## like family history, company wellness program, mental health consequences, etc.
## and see if that has any relationship with number of people went for treatment.

# Q7 How could you summarize your data to answer key questions?

## I will use the summary function and other slice/dice
## and join methods as mentioned above.

# Q8 What types of plots and tables will help you to illustrate the findings to your questions?
# Ensure that all graph plots have axis titles, legend if necessary, scales are appropriate,
# appropriate geoms used, etc.).

## Box plot and histogram.

# Q9 What do you not know how to do right now that you need to learn to answer your questions?

## I need more data exploration to answer this.

# Q10 Do you plan on incorporating any machine learning techniques to answer your research questions? E

## Yes, I plan to split the dataset, 70:30 into training and validation,
## use Generalized linear model to predict and check the model accuracy.
```