MARK5826 Individual Research Project

Peiguo Guan z5143964

**Decreasing Mental Health In Workplace**

My Mission Statement: Assist employees and companies to decrease mental health factors in workplace.

**The Problem**

Mental illness is the leading cause of long-term workplace absence in most developed countries. Mental illness is also associated with working environment like the pressure from supervisor and other employees, which will cause the lower productivity. If we can predict the people who have potential mental health problem and help them to get treatment, this will benefit both the individuals and companies.

**Target Market**

Employees, employers, companies or organisations who want to improve their working environment and decrease the potential elements which may cause mental health diseases.

|  |  |
| --- | --- |
| **Challenges they face** | **How we can help** |
| **The mental health problem.** Employees are facing lots of pressure from their work in work place, which will cause their mental and physical health problem. | Analysis the data and do the survey about what kind of factors may cause mental health problem and help them understand the situation better. |
| **Low productivity.** Companies and employers are facing lower productivity when their employees suffer from mental health problem. | Analysing the factors of mental illness may help leaders or managers in company to take correct strategy and make decision to improve their working environment, and eliminate the potential threatens. |

**The Solution:**

Based on the dataset which from a 2014-2016 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace, I will be basing its research to test who has potential mental health problem and suggest them to get treatment. The data includes information about basic information like age, gender, mental health history country as well as different kinds of attitudes toward workplace, like work remotely, if a person like to work with their friends, if a person can tell the physical or mental situation with their boss and so on.

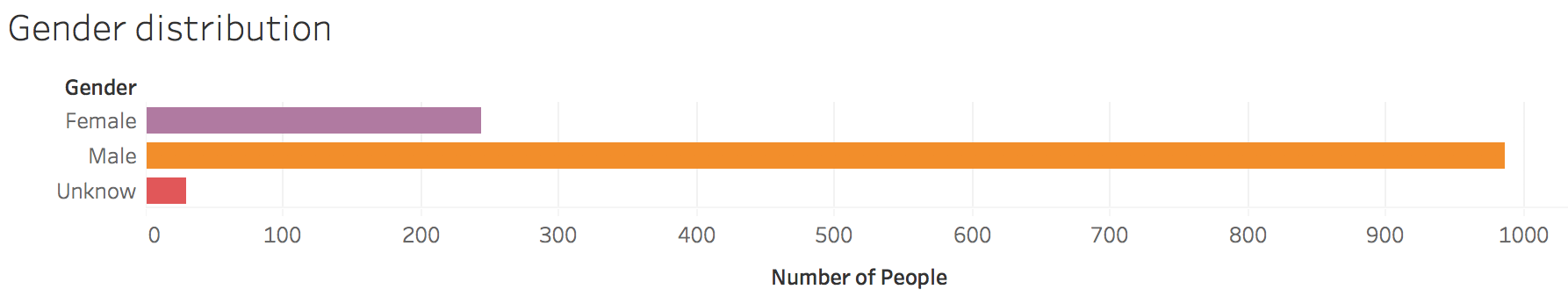
**Marketing Mix**

|  |  |
| --- | --- |
| **Product (Future goal)**  1. National Database (aggregate existing data, allow employees or companies to input new data)  2. Software: Input data traits and generate outcome factors report  3. Analysis Reports: report detailing about situation and improvement strategy. | **Promotion**  Advertising to companies by trade groups and websites, it is also can be done by emailing marketing. Partnerships health institution. We can also raise the public’s mental health caring by these strategies. |

**Data Exploration**

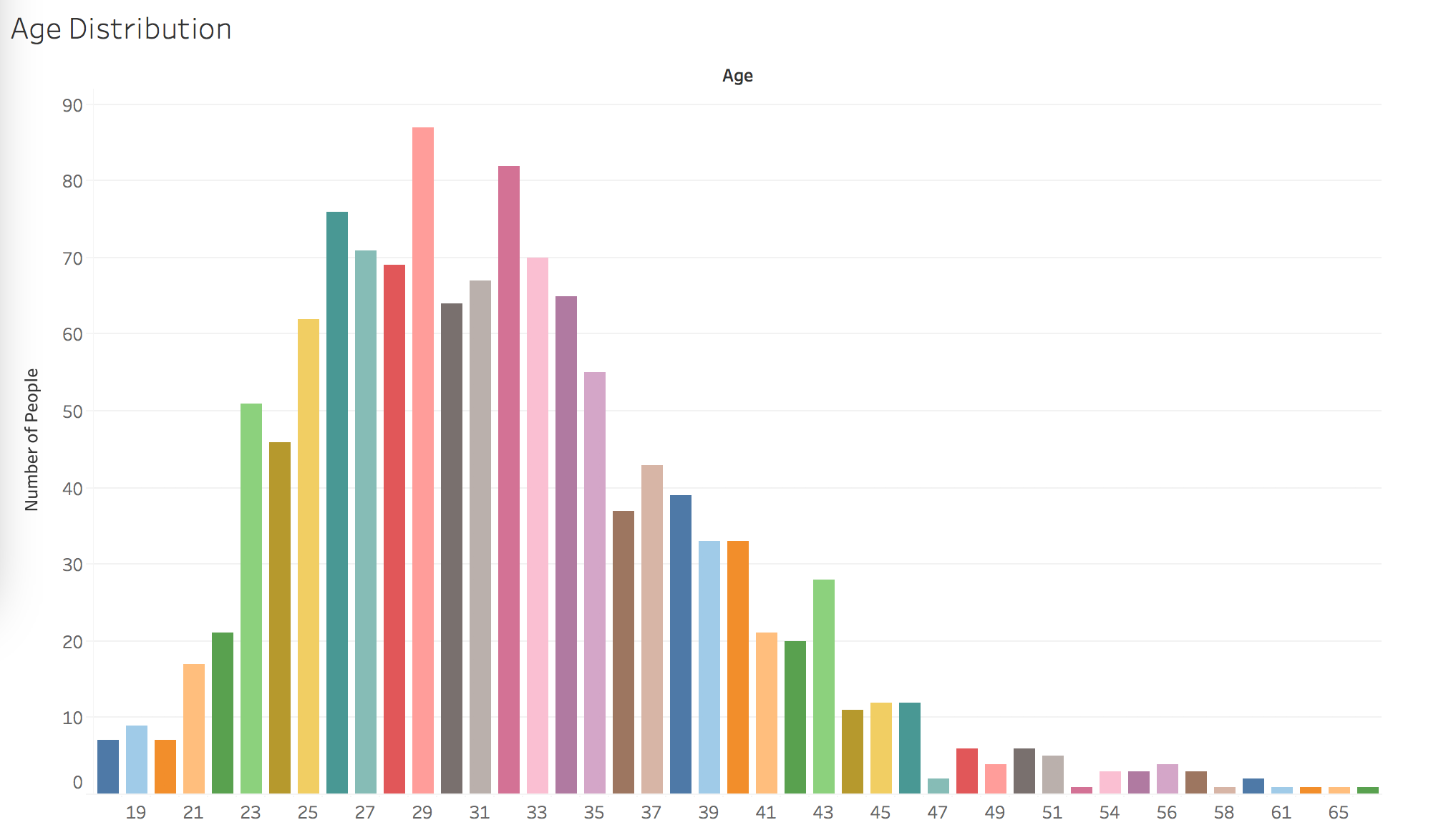
In order to better-understand this project I firstly conduct an exploratory study to identify trends/traits in male and female and their contribution in research, in additional, their attitude toward the mental health diseases in workplace. And then, analysing the personal details like gender, age, family history and their distributions, combined with their working environment information.

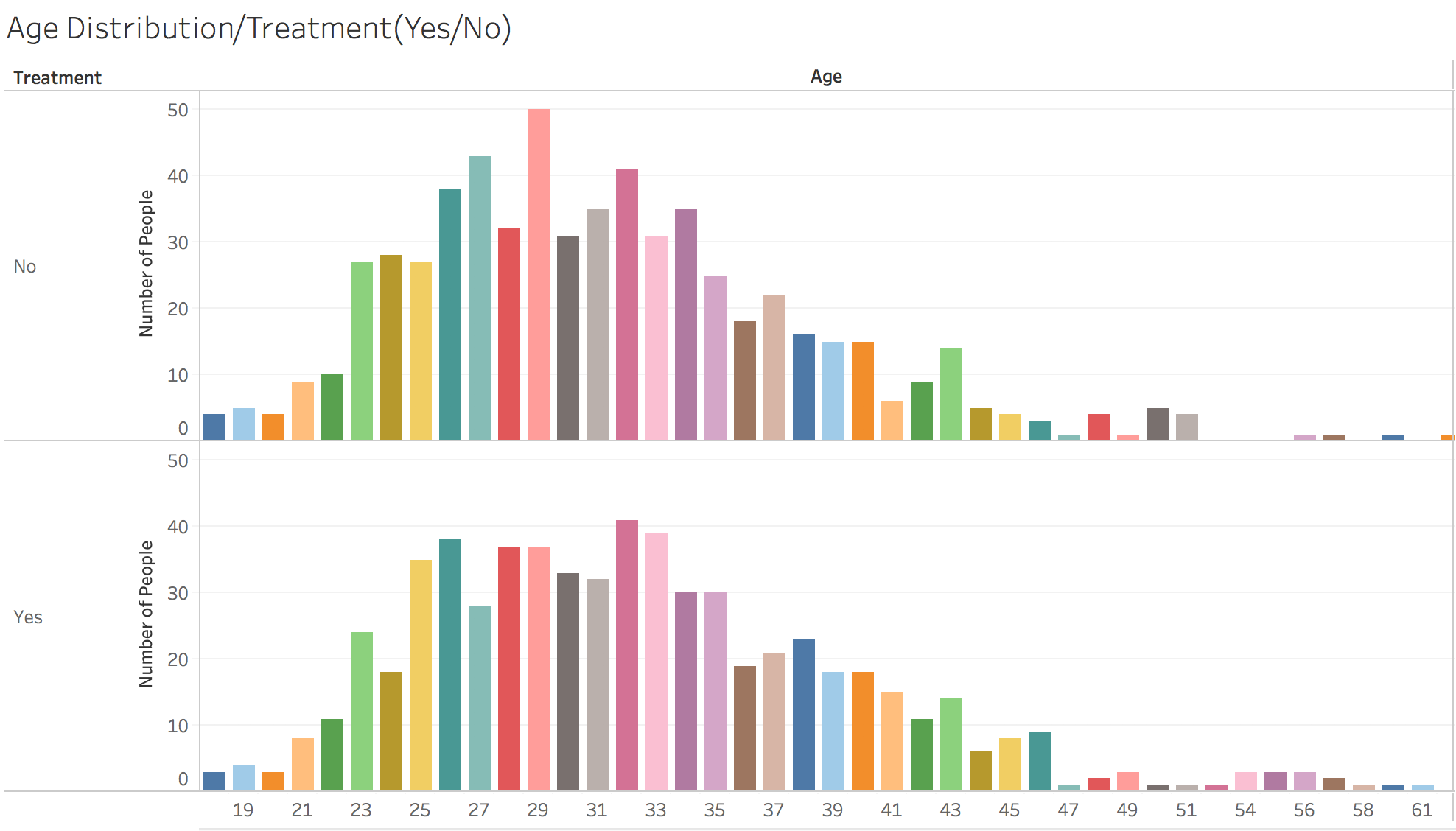
In order to do this, we are using data obtained from the Mental Health in Tech Survey (Kaggle, 2014-2016). This data set contains the data survey from different people and different features about working situation. In the future this data would be required to create a more accurate prediction model, and explain more possibilities to cause mental health in work.



This data set contains 986 male, 244 female and 29 “unknow”(which we can ignore).Unfortunately, most of the surveyors are male and it may cannot be a feature, but it also can tell that man seem to have more consideration about their mental health situation. About 78% of male get mental health disease, and 19% of female.

Age Distribution:

The distribution is from 18 years to 72 years old, and the bar chart shows that the people who are most easy to get mental health disease are from 23 to 40. And 29 years old has the largest number of people over the all. It can also see that young people between 25-33 are more easy suffering mental health in workplace, with the decreasing trends when the age is over 33 years old.  


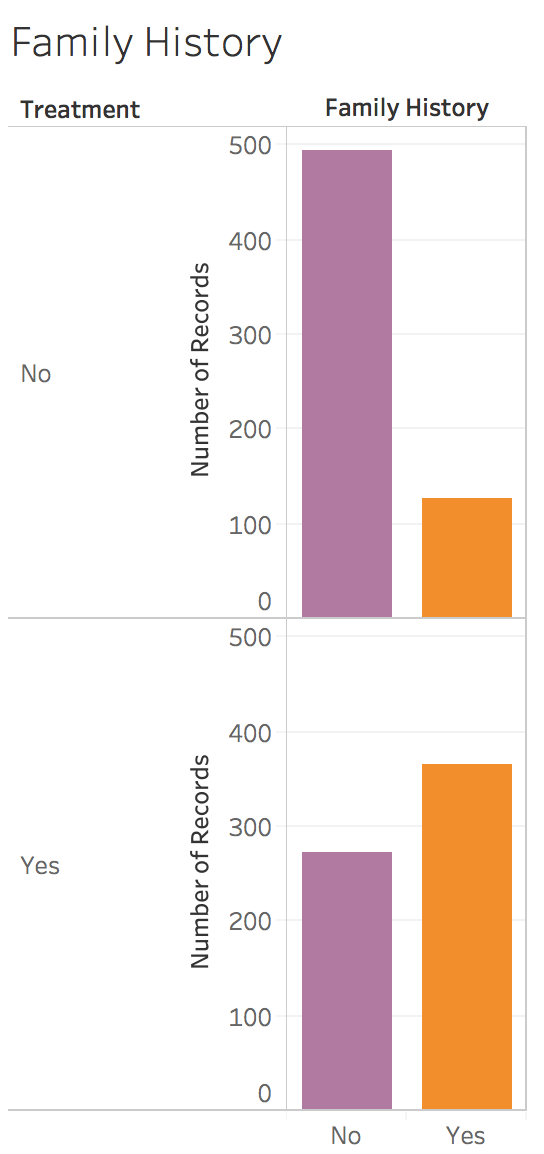


Age Distribution:

The distribution is from 18 years to 72 years old, and the bar chart shows that the people who are most easy to get mental health disease are from 23 to 40. And 29 years old has the largest number of people over the all. It can also see that young people between 25-33 are more easy suffering mental health in workplace, with the decreasing trends when the age is over 33 years old.

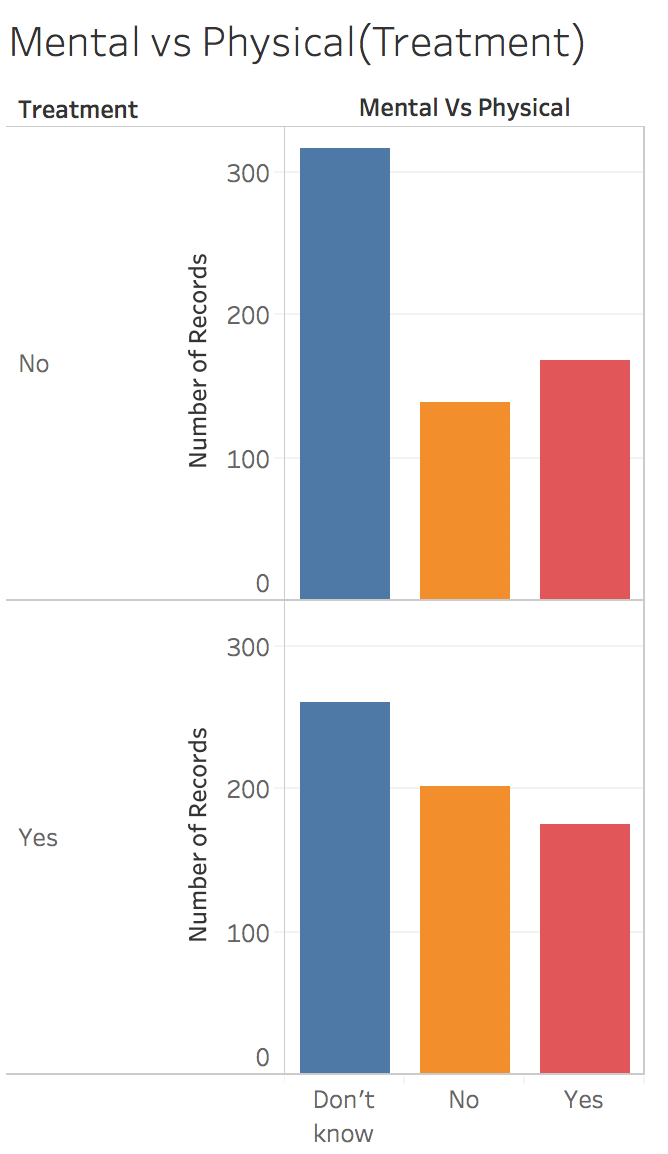
The second graph shows the number of people who have got the mental health treatment before or who are currently getting mental health treatment, compared with the people who hasn’t get any treatment.

Family History(Treatment):

****

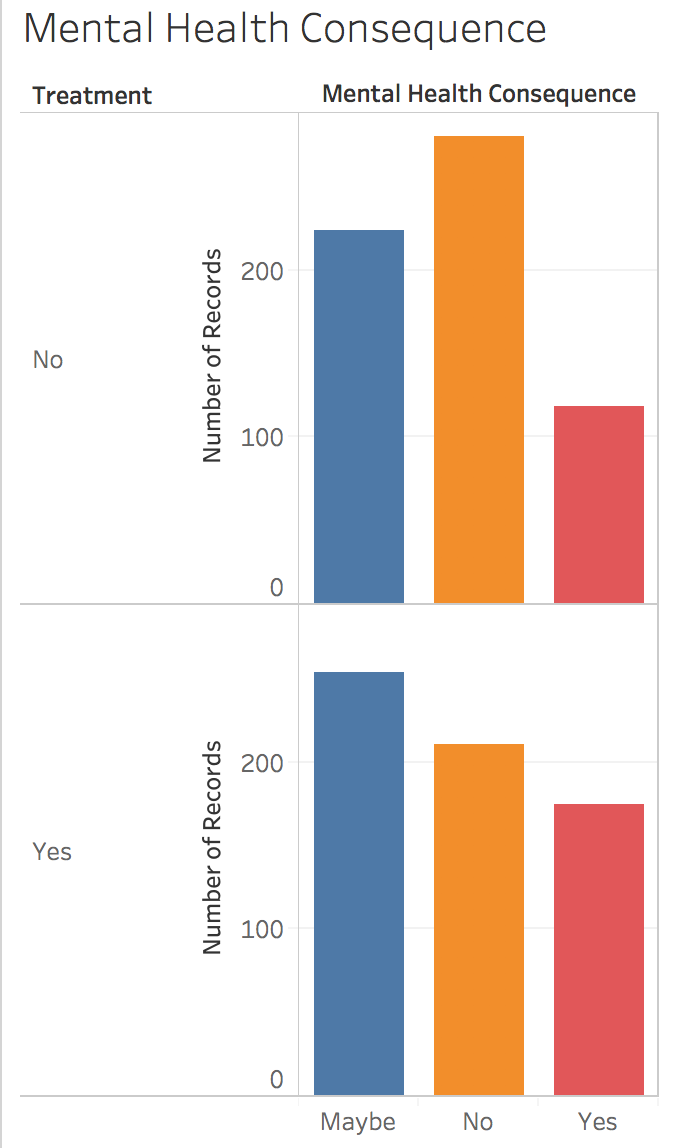
This graph indicates the mental health condition with treatment and family history, it is easy to find that the people who have family history are more likely to get treatment compared with who has not family history.

Mental vs Physical(Treatment):

****

This graph indicates the mental health condition with treatment and the attitude of the employers(If they think the mental health is as important as physical health), it is easy to find that the employers who do not care much about mental health, more employees are likely to get mental health treatment.

Mental Health Consequence(Treatment):

****

This graph indicates the mental health condition with treatment and the consequence of employees by discussing with their employers about their mental health issue. The YES means there will be negative consequence. It is clearly to see that, the number of employees get more treatment than those who talk with their employers and have not negative consequences.

**Data cleaning**

Python was used to clean the data sources to get a dataframe that could be easily processed.

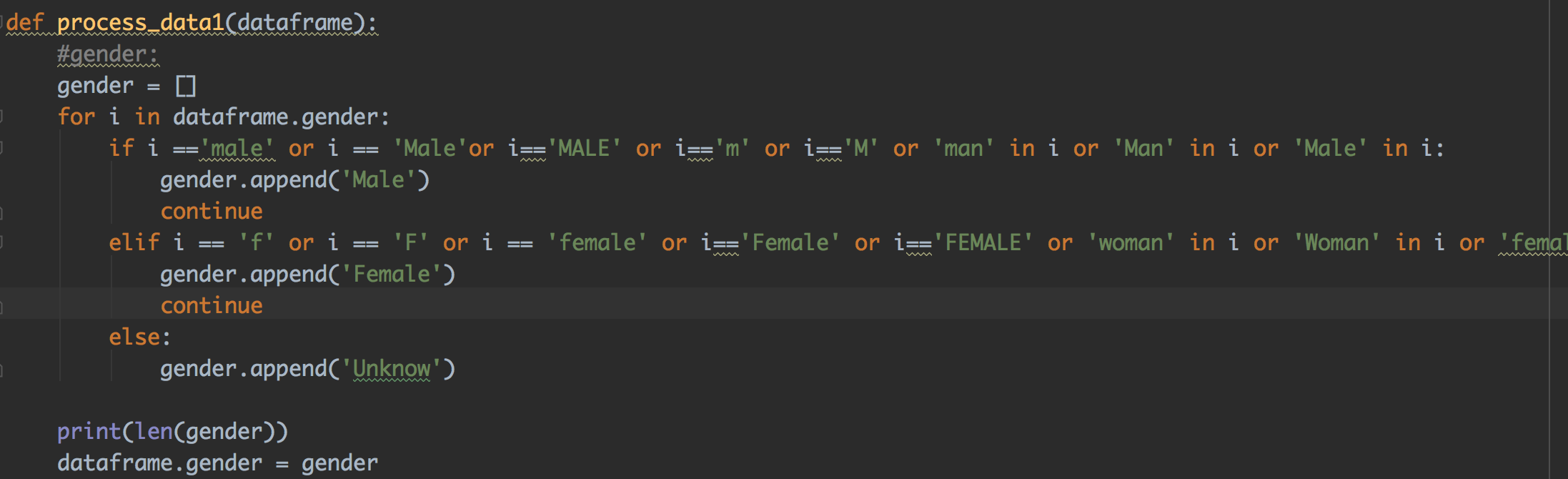
1.Dealing with missing data and get rid of variables, “timestamp”, “unnamed:\_0”, “comments”, “state” to make the dataframe more easier.



2.Cleaning all NaN: Assign default values for each data type; Create lists by data type; Clean the NaN's.



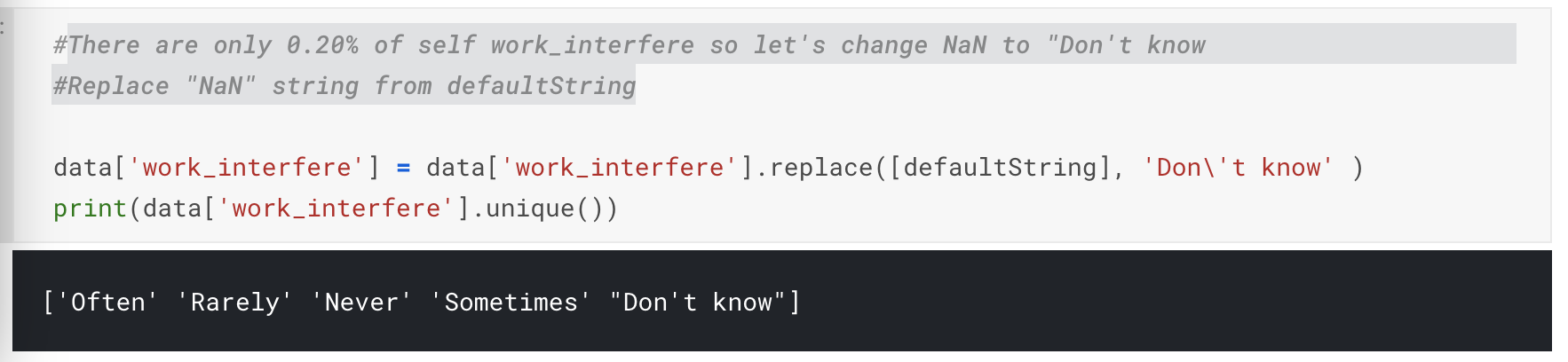
3. Formalize the “Male” and “Female” in gender.



4. There are only 0.014% of self employed so let's change NaN to NOT.

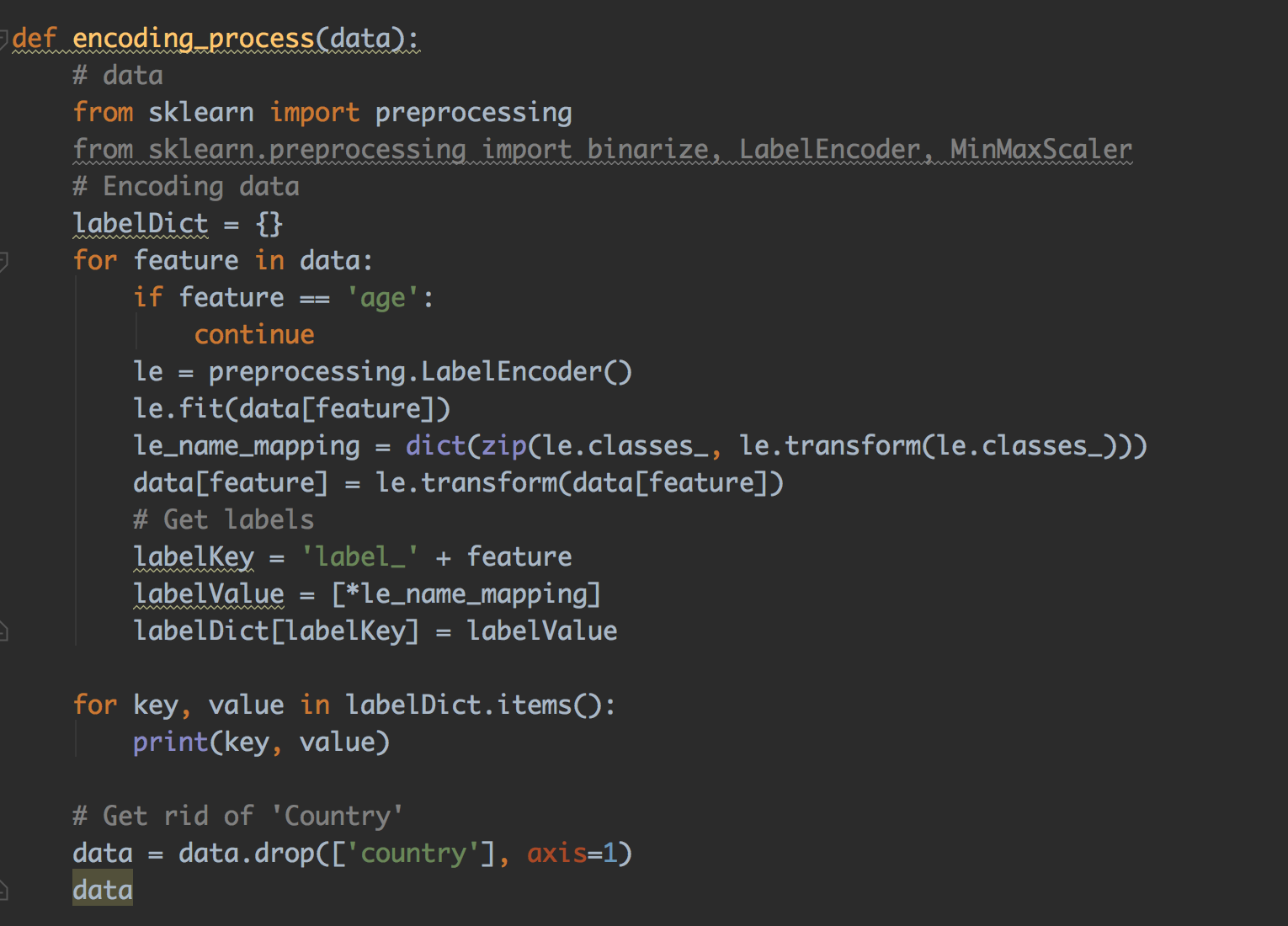


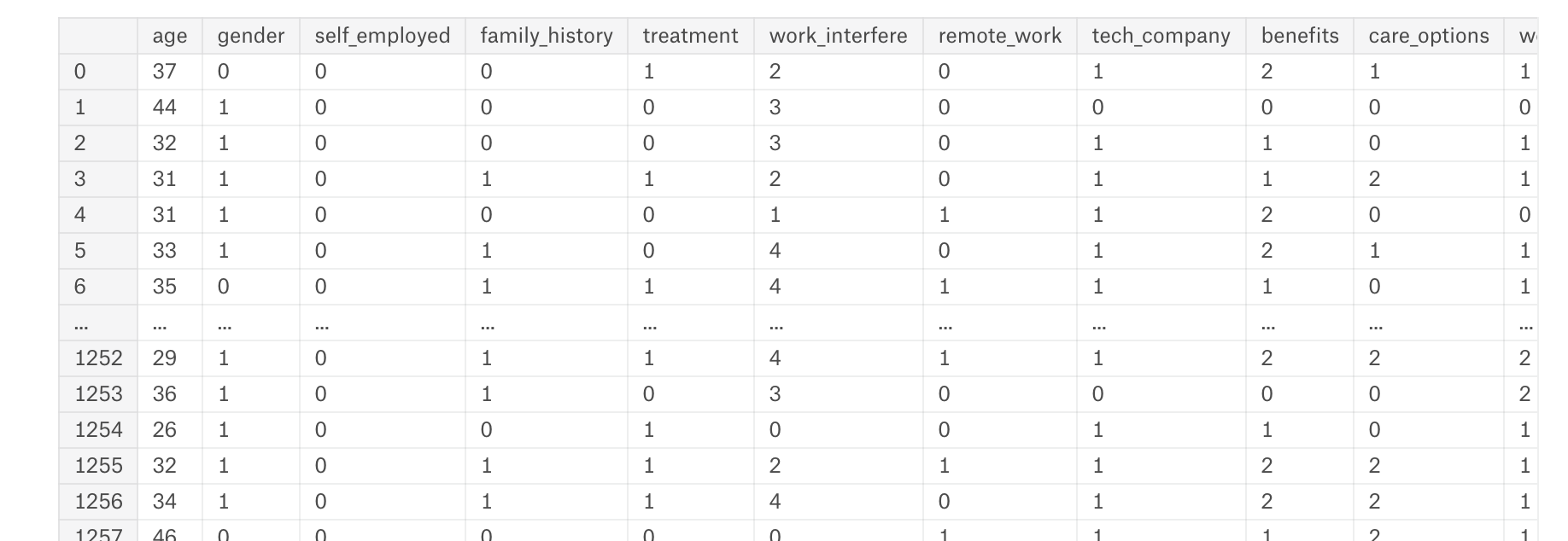
5. There are only 0.20% of self work\_interfere so let's change NaN to "Don't know. Replace "NaN" string from defaultString



**Encoding Data:**

In the encoding data part, I use sklearn preprocessing function to make all of the features have digit value this label used for next training and regression part.

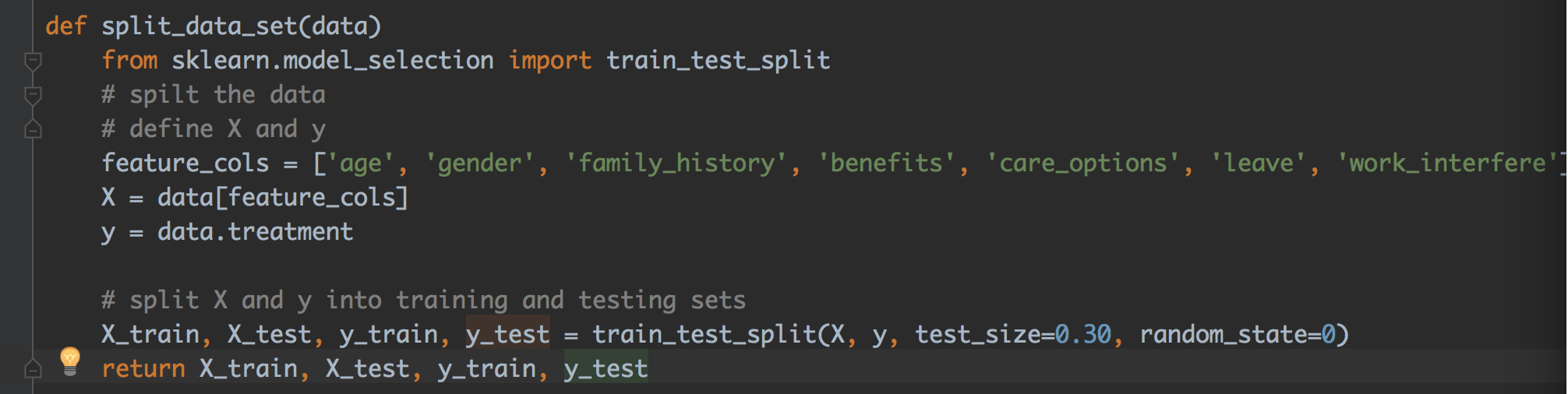
****

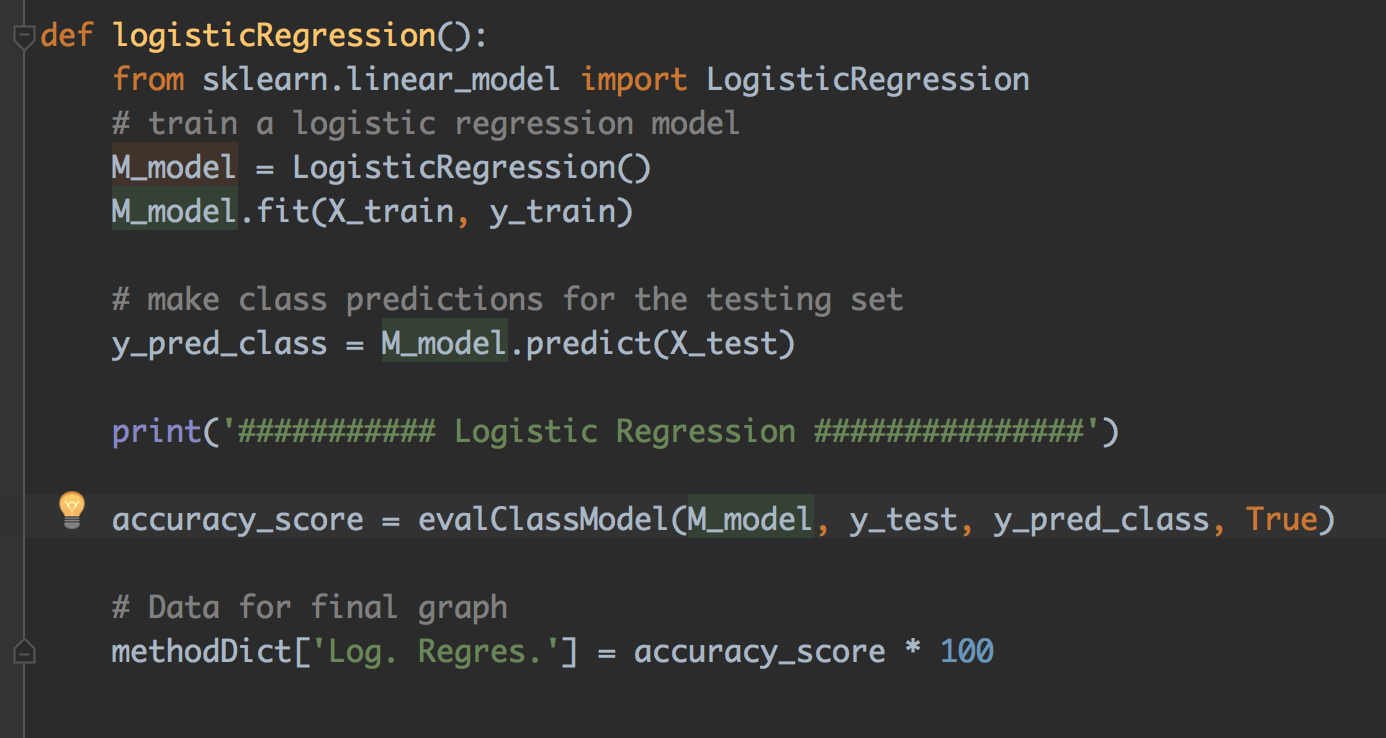
****

**Data Modeling**

The data model I use is Logistic Regression Model.

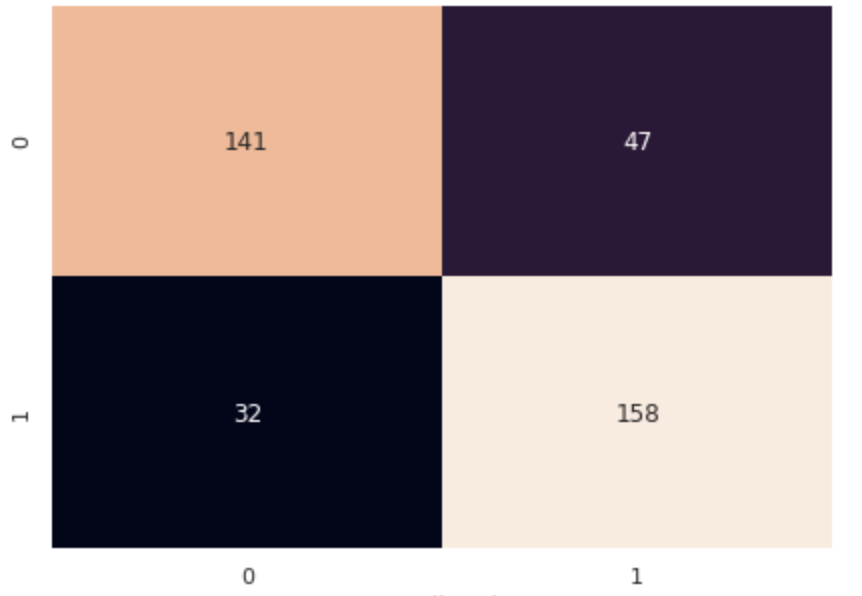
I using logistic model is because it is a binary dependent model(treatment or not-treatment in this project) and several features can be used as varibles.

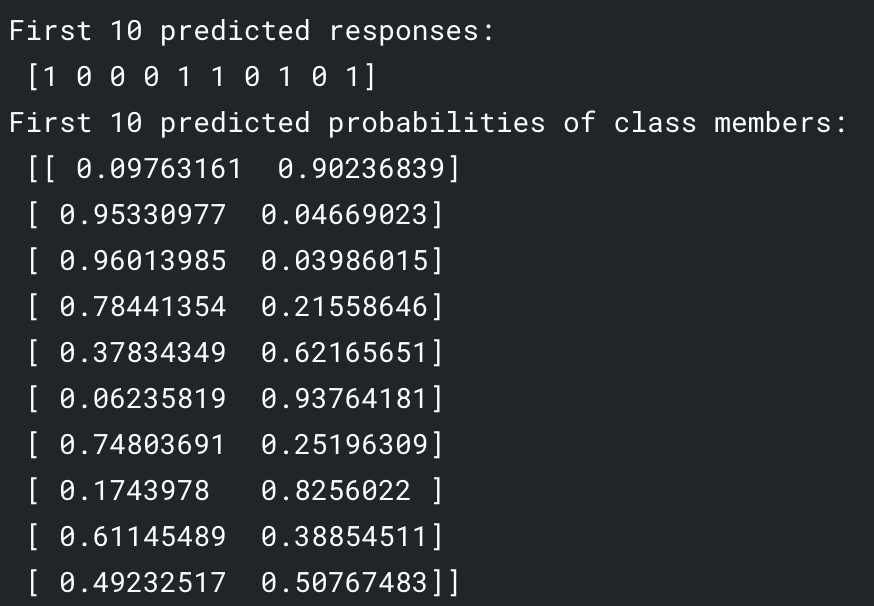


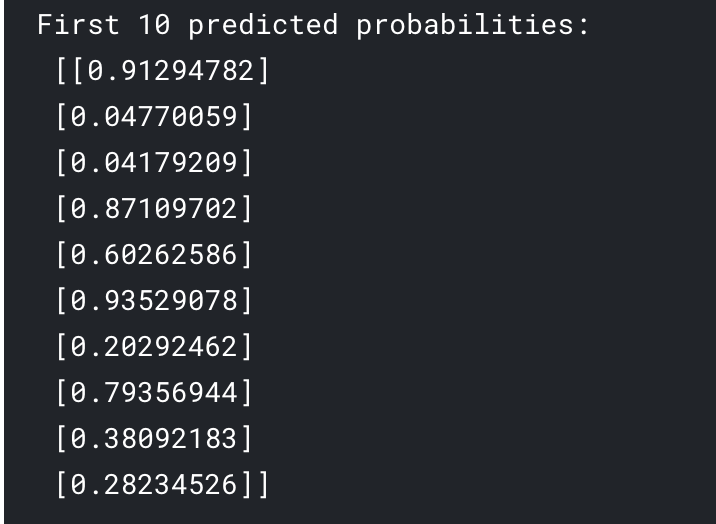


Firstly, I split the dataset as training set and test set. Then I build a logistic regression model on the training set, and train the data set. After that I make a prediction for the test. When the fit process is fininshed, I use the evalClassModel to evaluate the result by Classification accuracy method which is the easiest classification metric to understand. Comparing the true and predicted response values to test the accuracy. Metrics computed from a confusion matrix(visualize Confusion Matrix) and predict the probabilities from the class. And the output is the first ten record of score. It contains predicts response with value 0 and 1 which represents for the “treatment” and “not-treatment”.

Output:







Interpretation:

In the result, 1 means get treatment and 0 means not.

According to the result, the result with 1 is more likely to have mental health problem and need to get treatment. Some of them have already get treatment and the result is apparently is 1, but a person who do not get treatment and has value 1, it means he or she may have potential mental disease and need to get treatment.

**Conclusion and recommendations:**

Based on the survey and the data set, we can predict the employees or employers who has potential mental health problem. According to this we can give some suggestions like getting professional treatment and communicate with the manager in workplace to change management method. By this way, the problem that will affect the productivity and personal healthy will be prevented.

**Business Model**

My aim is to focused on profit organisation and company who want to make company better. By making a survey and test every employees and employers’ situation, this model can help to predict if an individual has potential mental disease or not, so that they can get treatment. At first stage, the profit will come from the charge from the company, and in the future, there will be an expansion plan like give solution and the development to make it more precisely.

**Future Plan**

The result of my research has an important value towards people who work in workspace. If people cannot take care of their mental health disease or potential mental problem, it will cause lots of unnecessary loss. In addition, one of the best way to raise the awareness of the public to focus and care about the mental disease is to share the result of my research. I can help relative department or government to do the research about the public of the living conditions and detect the trend of the problem. By this way, the authority can control the development and find a solution as fast as possible.

Due to the limitation of understanding the problem and the lack of technologies, it needs to take further steps to think of the features and factors about the research, so as to get much precisely research and result.

**Your Future Career Goal**

Before I take this course I just want to be a normal programmer who works for an internet company, but after this course I feel lots of fun in machine learning and it has incredable power which can predict the data and change the future! I feel that I find a new way of my career direction, and i strongly believe that, most of the things in the future are based on the big data, and the next step is to manipulate data through machine learning which can change the digital value to meaningful value in business and marketing. If a person can handle this technology, the trend and business can be exactly predictable. To be more specific, I will study in some internet company and get enough experience. Although I still have not exactly direction of my way, I will try to have my own career. In this two project, I learned lots of things, firstly, I practice to work in a group instead of working alone and I practice communication skills to change ideas with my teammates, and then when we face the projects, there are lots of problem we need to solve and nobody will teach you so I force myself to learn lots of things out of classes like different kinds of machine learning and the more exactly way to present the chart with tableau which help me to extend my knowledgeable. Finally, students like me who major in IT hardly have time to have presentation in class. I was really nervous and scary to stand in front of people and the individual work is really a good chance to challenge myself.

Analysing the data gives me lots of fun and I really want to study more, the next step of my goal is to use this skills to solve reality problems.

**List of References**

Performance Metrics for Classification problems in Machine Learning, 2017, Mohammed Sunasra, < https://medium.com/greyatom/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b>.

West Lafayette, Indiana, United States, *‘Mental Health in Tech Survey*, < https://www.kaggle.com/osmi/mental-health-in-tech-survey >,

< https://www.kaggle.com/osmi/mental-health-in-tech-2016#mental-health-in-tech-2016-neo4j-20161114.zip >

Predictors of mental illness, Kairos, 2018, <https://www.kaggle.com/kairosart/machine-learning-for-mental-health-1>