Mental Health In Tech Industry

PROJECT 1

By:
Priyanka Chandramohan

PROBLEM STATEMENT:

- Develop a Model that can predict whether a employee seek treatment or not
- Also identify the key Features that lead to mental Health Problems in tech space.

1. Introduction

- Nearly 86% of employees report improved work performance and lower rates of absenteeism after receiving treatment for depression, according to an April 2018 article in the Journal of Occupational and Environmental Medicine. This means big gains in retention and productivity for employers. By providing employees access to mental health benefits, the company can begin to create a culture of understanding and compassion at the tech company. And having employees who feel cared for and happy isn't just good, it's good business.

- Companies can use this model to know better about employee mental health issues and provide benefits for the needful employees, thus making efficient use of companies resources. This model can help in cutting off extra cost of providing mental health benefits for people who don't seek and use that money for other benefits of that employee. This will eventually increase employee satisfaction leading employee retention overall.

Main Goal

- Create an analytical framework to understand
  * Key factors impacting Mental Health.
- Develop a modeling framework
  * Increase employee satisfaction leading employee retention overall.

Features Summary

- This dataset is from a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace.
- **Timestamp: Tells us the time when the person is being surveyed.**
- **Age : age of the person**
- **Gender : Gender of the person (1 – Female 2 – Male 3 - Other)**
- **Country : Country of the Person.**
- **State**: If you live in the United States, which state or territory do you live in?
- **self_employed**: Are you self-employed?
- **family_history**: Do you have a family history of mental illness?
- **treatment**: Have you sought treatment for a mental health condition?

- **work_interfere**: If you have a mental health condition, do you feel that it interferes with your work?
- **no_employees**: How many employees does your company or organization have?
- **remote_work**: Do you work remotely (outside of an office) at least 50% of the time?
- **tech_company**: Is your employer primarily a tech company/organization?
- **benefits**: Does your employer provide mental health benefits?
- **care_options**: Do you know the options for mental health care your employer provides?
- **wellness_program**: Has your employer ever discussed mental health as part of an employee wellness program?
- **seek_help**: Does your employer provide resources to learn more about mental health issues and how to seek help?
- **anonymity**: Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
- **leave**: How easy is it for you to take medical leave for a mental health condition?
- **mental***health***consequence**: Do you think that discussing a mental health issue with your employer would have negative consequences?
- **phys***health***consequence**: Do you think that discussing a physical health issue with your employer would have negative consequences?
- **coworkers**: Would you be willing to discuss a mental health issue with your coworkers?
- **supervisor**: Would you be willing to discuss a mental health issue with your direct supervisor(s)?
- **mental***health***interview**: Would you bring up a mental health issue with a potential employer in an interview?
- **phys***health***interview**: Would you bring up a physical health issue with a potential employer in an interview?
- **mental***vs***physical**: Do you feel that your employer takes mental health as seriously as physical health?
- **obs_consequence**: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
- **comments**: Any additional notes or comments

This is what the dataset looks like,

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	benefits
188	2014-08-27 12:53:05	23	male	United Kingdom	NaN	No	Yes	No	NaN	1-5	Yes	Yes	Don't know
1009	2014-08-29 09:46:56	44	F	United States	CA	No	No	No	NaN	100-500	No	Yes	Don't know
302	2014-08-27 14:20:43	26	Male	United States	MA	No	No	No	Rarely	100-500	No	Yes	Yes
18	2014-08-27 11:34:53	46	male	United States	MD	Yes	Yes	No	Sometimes	1-5	Yes	Yes	Yes
210	2014-08-27 13:06:00	40	Female	United States	DC	No	No	No	NaN	26-100	No	Yes	Yes

Table1 - Dataset

It has lots of rows and columns , which is 1259 rows and 27 columns. Here is the list of columns this dataset has,

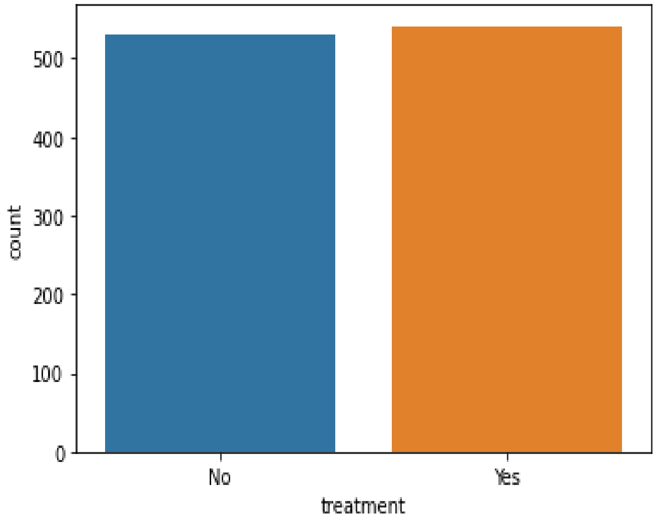
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 27 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Timestamp                  1259 non-null   object
 1   Age                        1259 non-null   int64
 2   Gender                     1259 non-null   object
 3   Country                    1259 non-null   object
 4   state                      744 non-null    object
 5   self_employed              1241 non-null   object
 6   family_history             1259 non-null   object
 7   treatment                  1259 non-null   object
 8   work_interfere             995 non-null    object
 9   no_employees               1259 non-null   object
 10  remote_work                1259 non-null   object
 11  tech_company               1259 non-null   object
 12  benefits                   1259 non-null   object
 13  care_options               1259 non-null   object
 14  wellness_program           1259 non-null   object
 15  seek_help                  1259 non-null   object
 16  anonymity                  1259 non-null   object
 17  leave                      1259 non-null   object
 18  mental_health_consequence  1259 non-null   object
 19  phys_health_consequence    1259 non-null   object
 20  coworkers                  1259 non-null   object
 21  supervisor                 1259 non-null   object
 22  mental_health_interview    1259 non-null   object
 23  phys_health_interview      1259 non-null   object
 24  mental_vs_physical         1259 non-null   object
 25  obs_consequence            1259 non-null   object
 26  comments                   164 non-null    object
dtypes: int64(1), object(26)
memory usage: 265.7+ KB
```

Table2 - Columns in

2. Exploratory data analysis

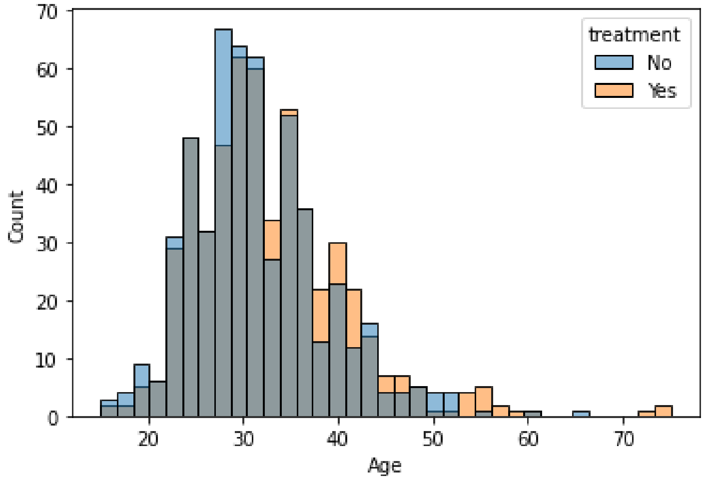
People who opt for Mental Treatment



Plot1 – Analysis on treatment

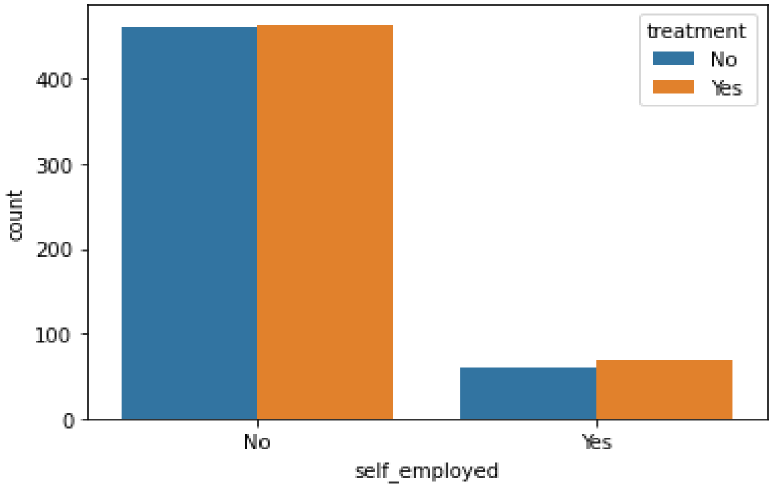
•Little more than 50% people are taking treatment , this is our target varible and there is no such class imbalance we are reday to go with this column.

People taking mental health treatment according to the Age



Plot2 – Analysis on Age

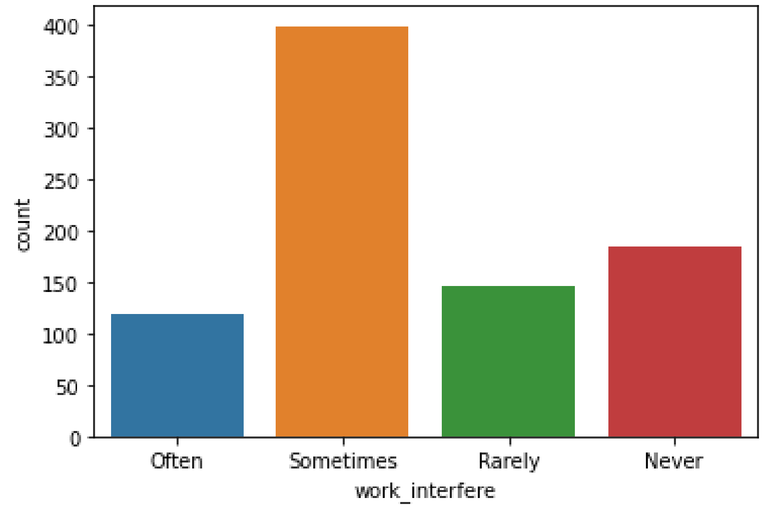
- Both the curve are merging so we can take any key insight from here.



Plot3 – Analysis on self_employed

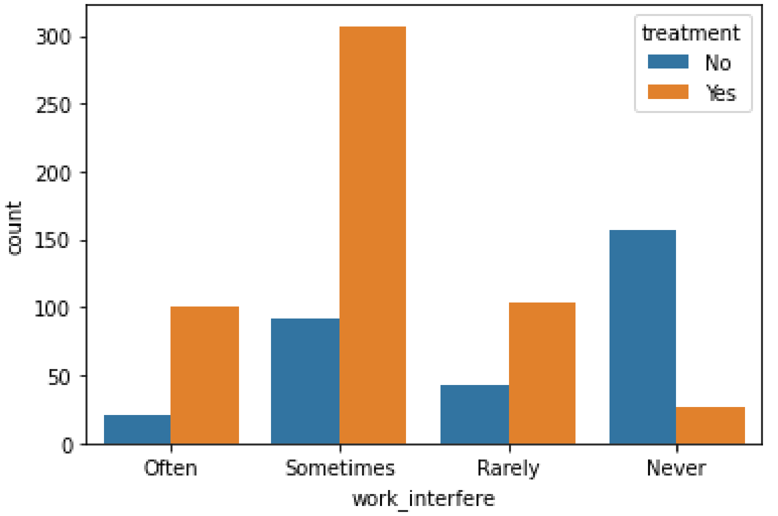
- More than 80% people(approx) are not self_employed. Despite being the large difference, people taking mental treatment are almost same. So it does not Matter whether the person is self_employed or not, person is taking mental help. There is imbalance but the class distribution is same so no difference at all.

Is your mental health affecting your work power.



Plot4 – Analysis on self_employed

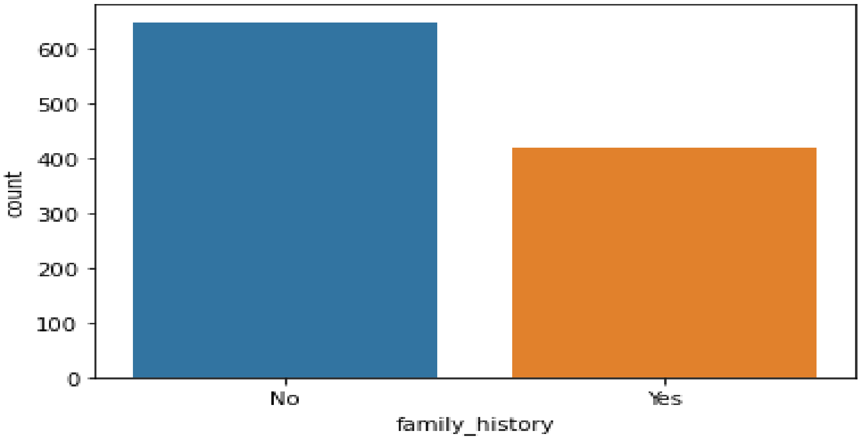
Almost 80% people feel that their mental health conditions affect their work sometimes,rarely and frequently. Mental health conditions sometimes become an interfere while working about 45%.



Plot5 – Analysis on work_interfare

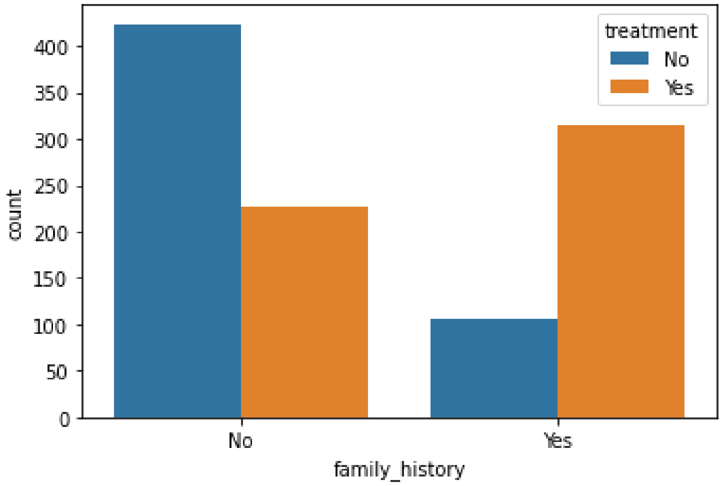
The plots prove that almost 80% want to get treatment. But it's surprising to know even mental health never has interfered at work, there is a little group that still want to get treatment before it become a job stress. It can be triggered by the requirements of the job do not match the capabilities, resources or needs of the worker If you are running a tech organization , you should consider providing resources for employees seeking treatment and it will help in boosting employee experience and will definitely increase their productivity**.**

Family history of mental illness



Plot6 – Analysis on family_history

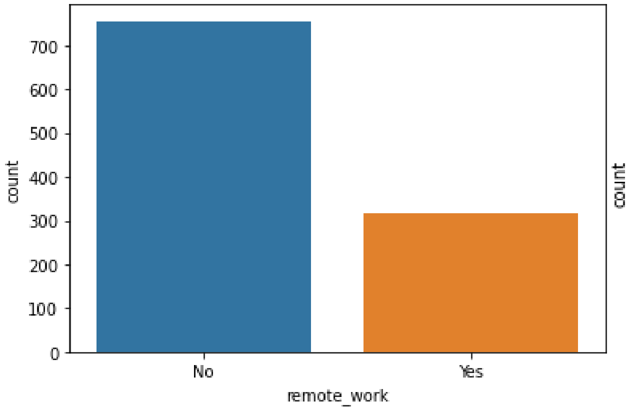
There are less than 40 % people who have family history of mental health problem.

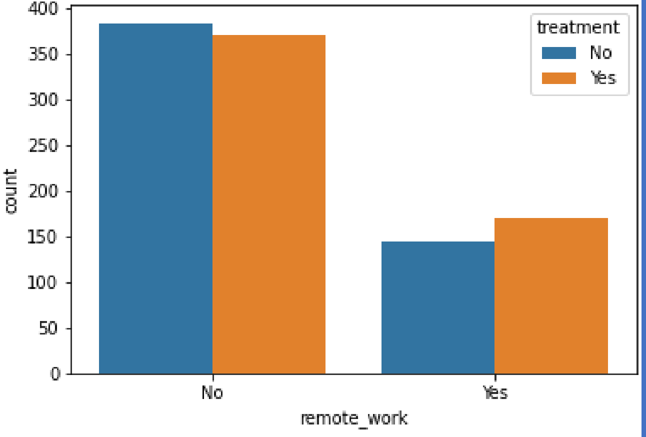


Plot7 – Analysis on family_history and Target

It clearly shows people who have mental health problems in their Family are more bound to take treatment in comparison of the people who dont have mental health problems in their family. This will be an important feature.

People doing remote work(more than 50% of their time)

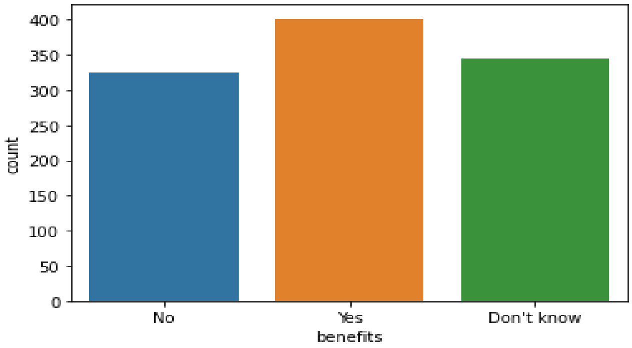




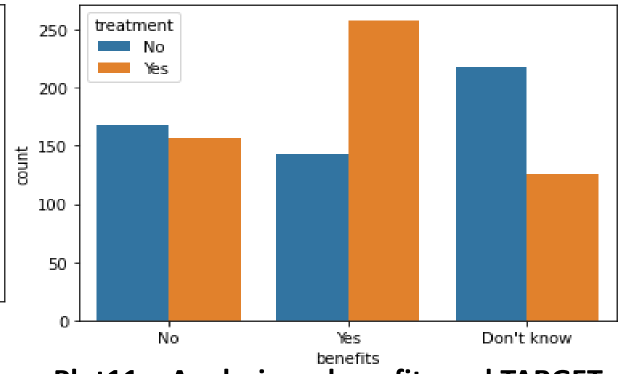
Plot8 – Analysis on remote_work **Plot9 – Analysis on family_history and Target**

Doesn't matter remote work or not, almost 50% of people in both categories seek treatment. People in remote work are sligtly more in number who seek treatment. It might be due to lack of social interaction in remote mode.

does employer provide mental health benefits



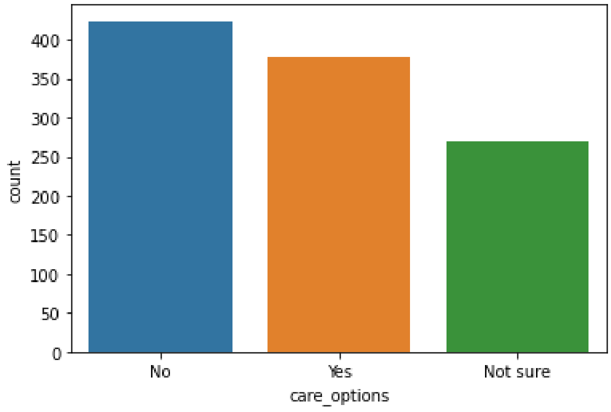
Plot10 – Analysis on benefits



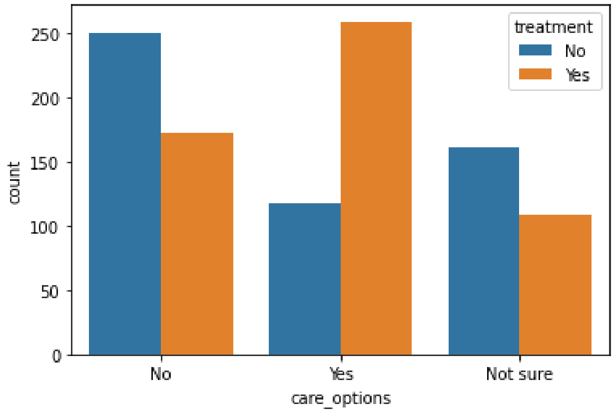
Plot11 – Analysis on benefits and TARGET

We see that around 38% of the respondents said that their employer provided them mental health benefits, whereas a significant number (32%) of them didn't even know whether they were provided this benefit. Coming to the second graph,we see that for the people who YES said to mental health benefits, around 63% of them said that they were seeking medical help. So we can see the employer resources are utilized to a larger extent. Even if you think about the cost , you should definitely go for it because it is efficiently utilized by the employees. Surprisingly, the people who said NO for the mental health benefits provided by the company, close to 45% of them who want to seek mental health treatment.

Mental healthcare options provided by Employer



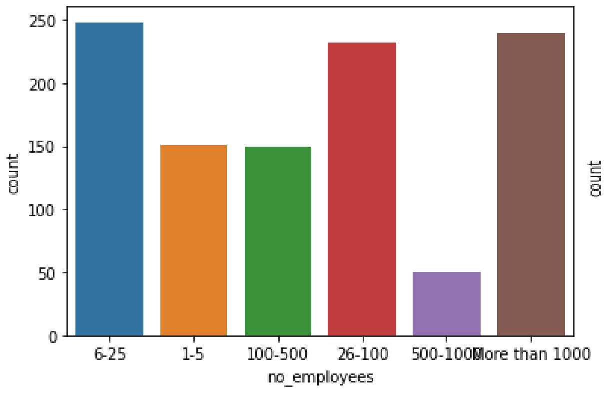
Plot12 – Analysis on care_options



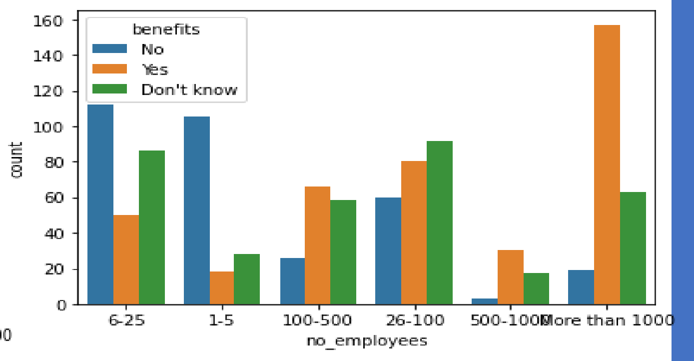
Plot13 – Analysis on care_options and Target

40% of employees are not provided any care options and 25% are not sure whether care options exist in company. We can see 60% of employees whose organization don’t have care options are seeking treatment. These organizations need to address this issue. People who have care options are actually seeking treatment; this can validate our claim to have care options.

Number of Employees in the Company



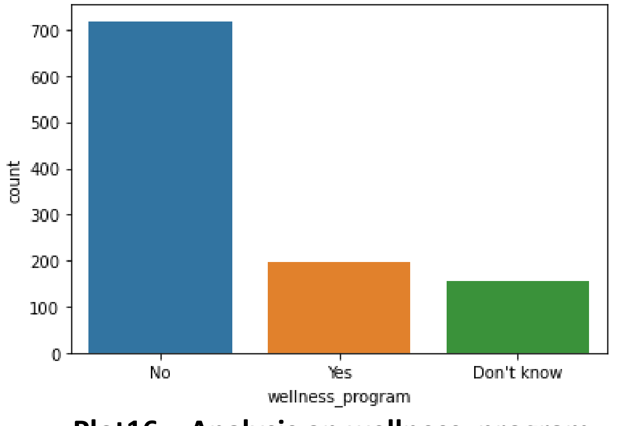
Plot14 – Analysis on no_employees



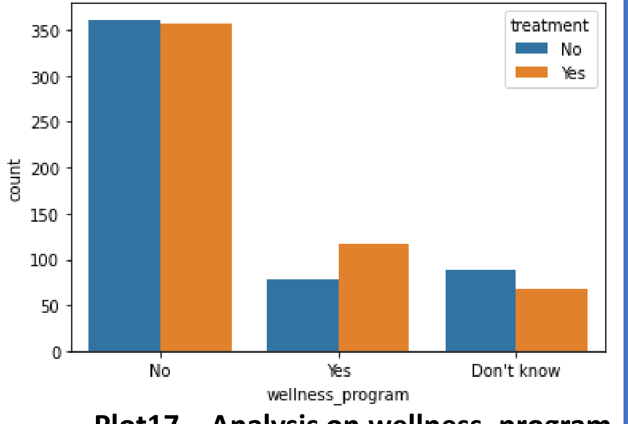
Plot15 – Analysis on no_employees and Target

Smaller companies are providing less benefits than the larger one. Size of the is not affecting whether people are seeking mental help or not.

Does Mental Health comes in your employee wellness program



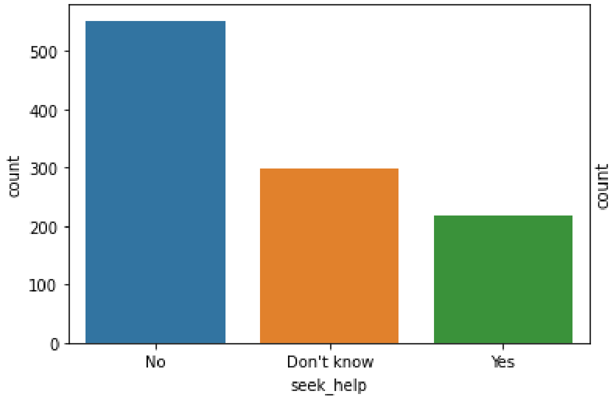
Plot16 – Analysis on wellness_program



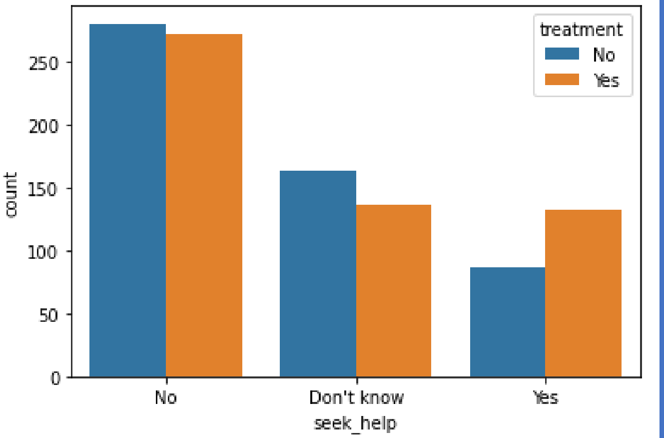
Plot17 – Analysis on wellness_program and Target

70% employers haven't discussed mental health as a employee wellness program. Around 50% People who don't know about the program are seeking help. This means organizations should explain the mental health benefits provided by the company Companies should include mental health in the employee wellness program. This shouldn't be overlooked

Does employer provide ways to seek help in case of Mental Health



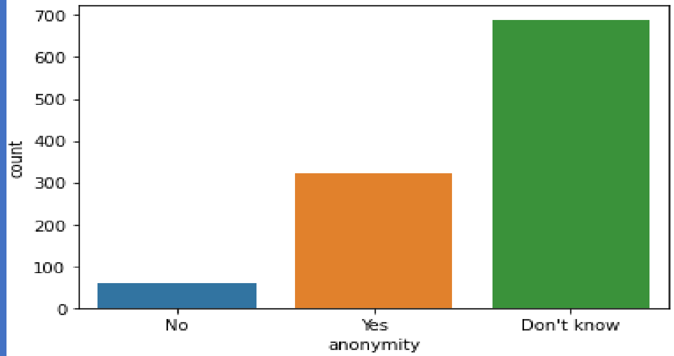
Plot18 – Analysis on seek_help



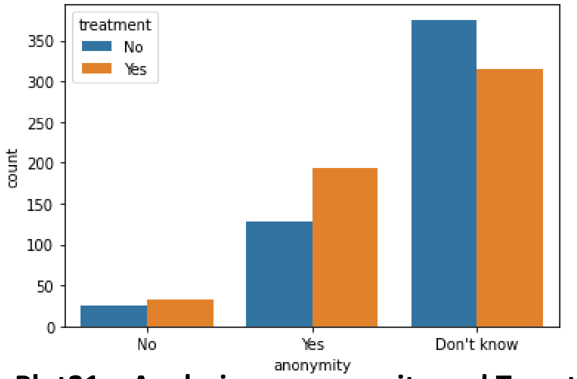
Plot19 – Analysis on seek_help and Target

More Than 51% respondents said they have not been provided ways for seeking in case of mental health by the employers. It can be clearly seen that the places where they provided help, more percentage of people took treatment.

Is your Identity be remained anonymous if you use resources



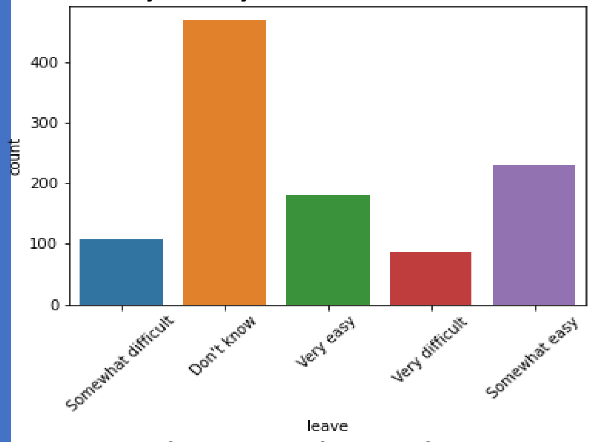
Plot20 – Analysis on anonymity



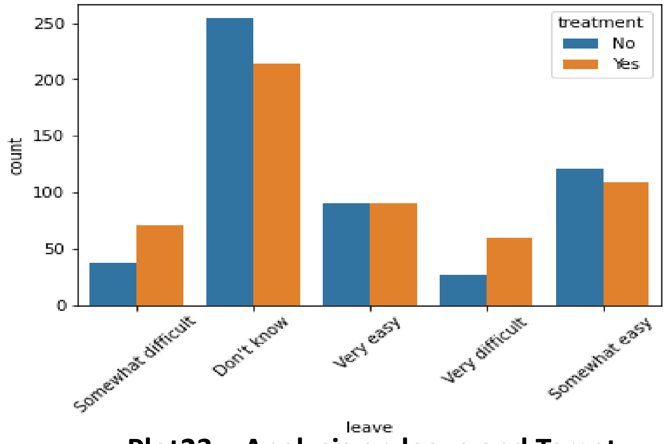
Plot21 – Analysis on anonymity and Target

Almost 65% of the respondents have no idea about it. Companies should create a safe environment for people to feel safe to share their problems. We can establish the fact that people who think their anonymity is protected are more willing to seek treatment.

How easy it for you take Medical Leave for a Mental Health Condition



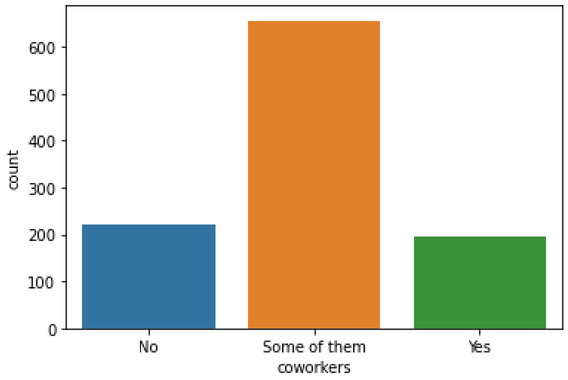
Plot22 – Analysis on leave



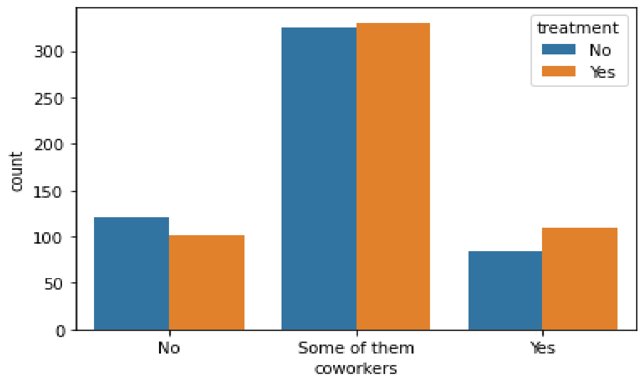
Plot23 – Analysis on leave and Target

- 45% have no idea about it.
- Places where Leave is Difficult to get they are more into Treatment.
- Important feature for the Model.

Discussing Mental Health with the co-worker



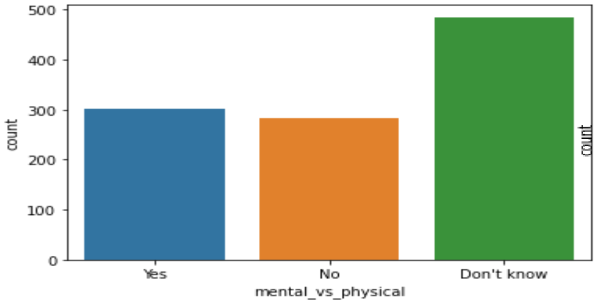
Plot24 – Analysis on co-workers

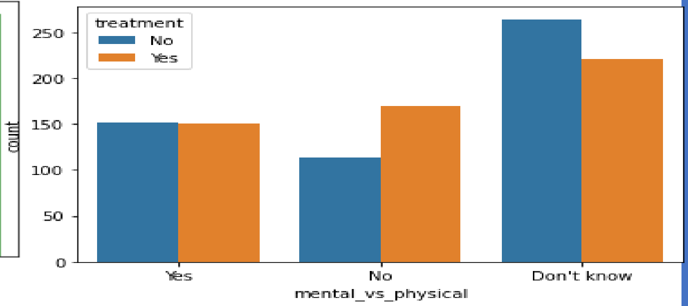


Plot25 – Analysis on co-workers and Target

Good sign that most people have at least some people (co-workers) to talk to about the mental health issues.

Is Mental Health Taken Seriously?

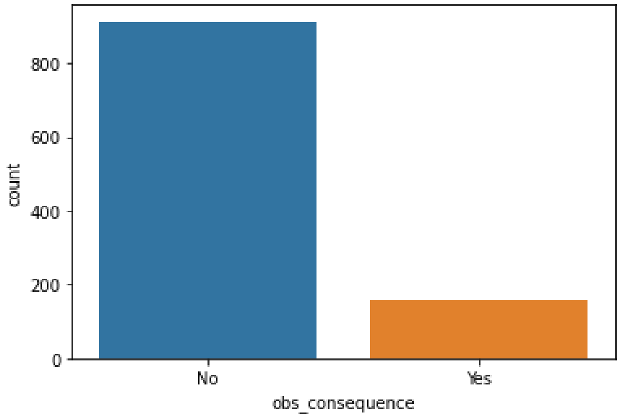


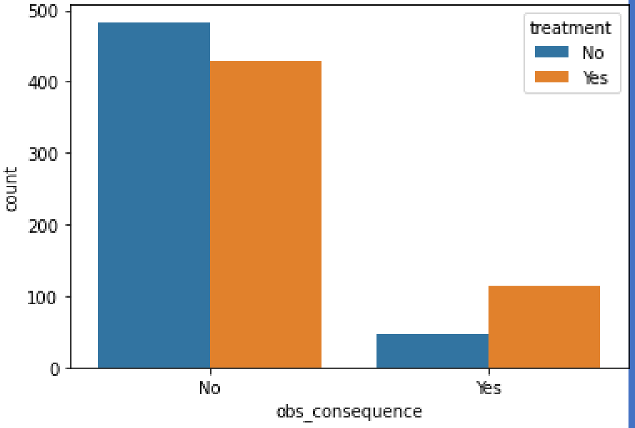


Plot26 – Analysis on mental vs physical **Plot27 – Analysis on mental vs physical and Target**

Places where Mental health is not treated as same as Physical Health , people are more inclined for Treatment.

Does Colleagues Have negative consequences about Mental Health?





Plot28 – Analysis on obs_consequence **Plot29 – Analysis on obs_consequence and Target**

Splitting Data

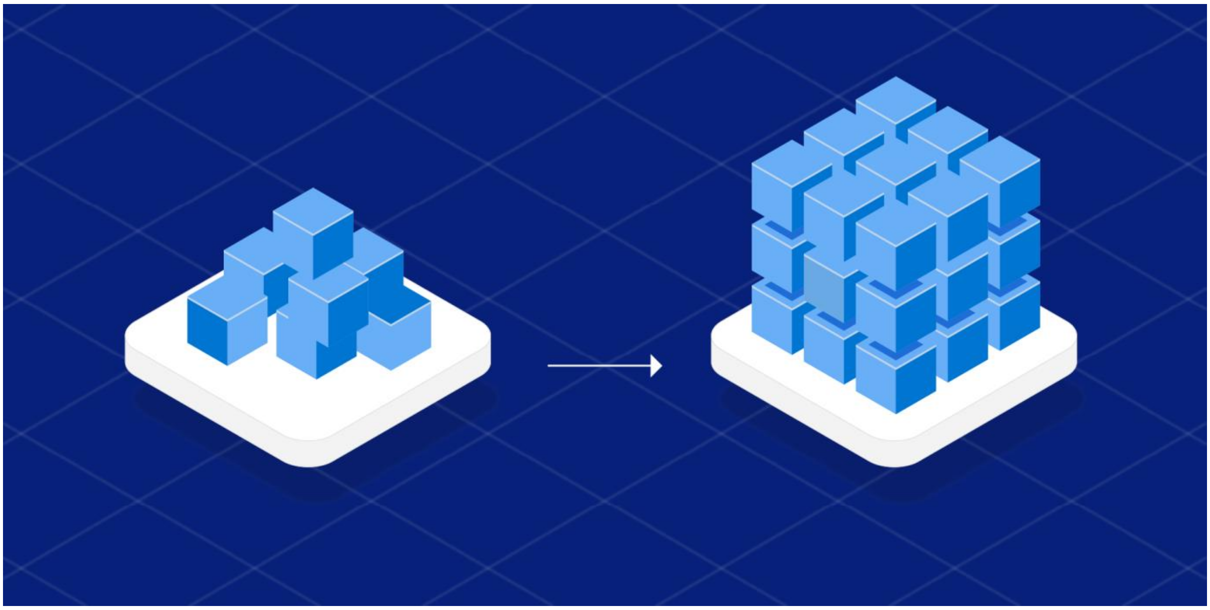
- Data splits into training dataset and testing dataset.
- Training dataset is used to fit the machine learning model.
- Test dataset is used to evaluate the fit machine learning model.
- Here 85% of the data taken as training dataset and remaining 15% of dataset used for testing purpose

Data Pre-processing

- Null Values treatment.
- Duplicate Values treatment.
- Label Encoding
- Ordinal Encoding

Transformation of Data

- **T**o scale data into a uniform format that would allow us to utilize the data in a better way.
- For performing fitting and applying different algorithms to it.
- The basic goal was to enforce a level of consistency or uniformity to dataset.



Fitting Different Model

Following classifiers are used for predicting whether employee seek mental treatment or not:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine
- Gradient Boosting
- Xgboost
- Adaboost

Cross Validation & Parameter Tuning

- It is a resampling procedure used to evaluate machine learning models on a limited data sample.
- Basically,Cross validation is a technique using which Model is evaluated on the Dataset on which it is not trained that is it can be a test data or can be another set as per availability .
- Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting.

Different models

1. Logistic Regression

- Logistic Regression is a machine learning algorithm for classification problem.
- It is a predictive analysis algorithm and based on the concept of probability.
- It is most useful for understanding the influence of several independent variables on a single output variable.

	Name	F1_score_trainset	F1_score_validationset
0	Logistic Regression	0.825023	0.818548

2. Decision Tree Classifier

- Given a data of attributes together with its classes, a decision tree produces tree produces a sequence of rules that can be used to classify the data.
- Decision Tree is simple to understand and visualize , and can handle both numerical and categorical data.

	Name	F1_score_trainset	F1_score_validationset
0	DECISION TREE	0.845771	0.845624

3. Random Forest Classifier

- Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of the dataset.
- The sub sample size is always the same as the original input sample size but the samples drawn with replacement.

	name	F1_Score_trainset	F1_score_validationset
0	RANDOM FOREST	1.0	0.836386

4. Support Vector Machine

- SVM is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible.
- The objective of the SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

	Name	F1_score_trainset	F1_score_validationset
0	SVM	0.883392	0.824692

5. Gradient Boosting

- It is a group of ML algorithms that combine many weak learning models together to create a strong predictive model.
- It is a sequential Ensemble learning technique where the performance of the model improves over iterations.

	name	F1_Score_trainset	F1_score_validationset
0	Gradient Boosting	0.903863	0.837939

6. XG Boosting

- XG boost is a decision tree based ensemble ML algorithm that uses a gradient bossting framework.
- It is a perfect combination of software and hardware optimization techniques to yield superior results using less computing resources in the shortest amount of time.

	name	F1_Score_trainset	F1_score_validationset
0	XG Boost	1.0	0.804219

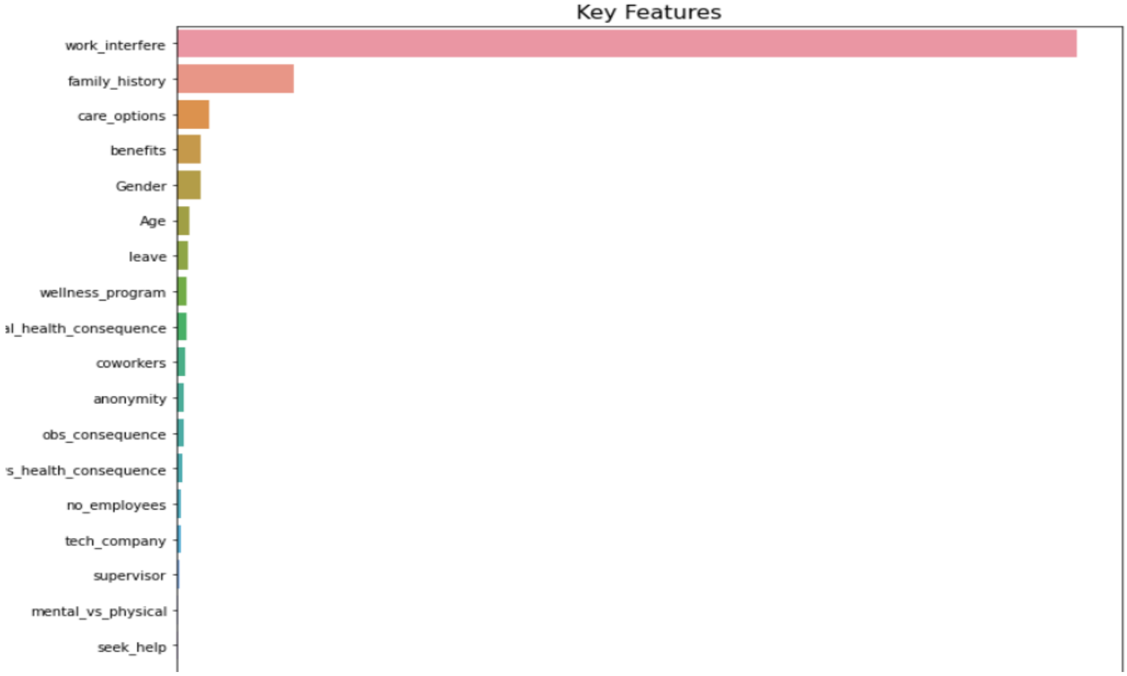
Comparison of Model

- After comparing f-1 score of all the models, the f-1 score of Gradient Boosting and Random Forest comes out to be highest.

- After doing the Final analysis with the test set we choose the Gradient Boosting Algorithm as it is providing the highest f-1 Score of 87 %.

Feature importance

- Feature selection is the process of reducing the number of features when developing a predictive model.
- It is desirable to reduce the number of input variables to both reduce the computational cost of the modelling and , in some cases , to improve the performance of the model.



Conclusion

- Work interference has the largest contribution.

- Whether the employee's mental health issues interfering with the work is the thing that the company should ask for its employees.

- Family history and care options(programs and benefits) provided by company is also influential in employees who want to get treatment.

- For all the remaining features, there has been a little contribution. noticing/knowing some of these features beforehand can even help support an individual who may be experiencing a mental health issues and and connect them with the appropriate employee resources.