

MindAble

Project Objectives

Within the workplace, we would like to

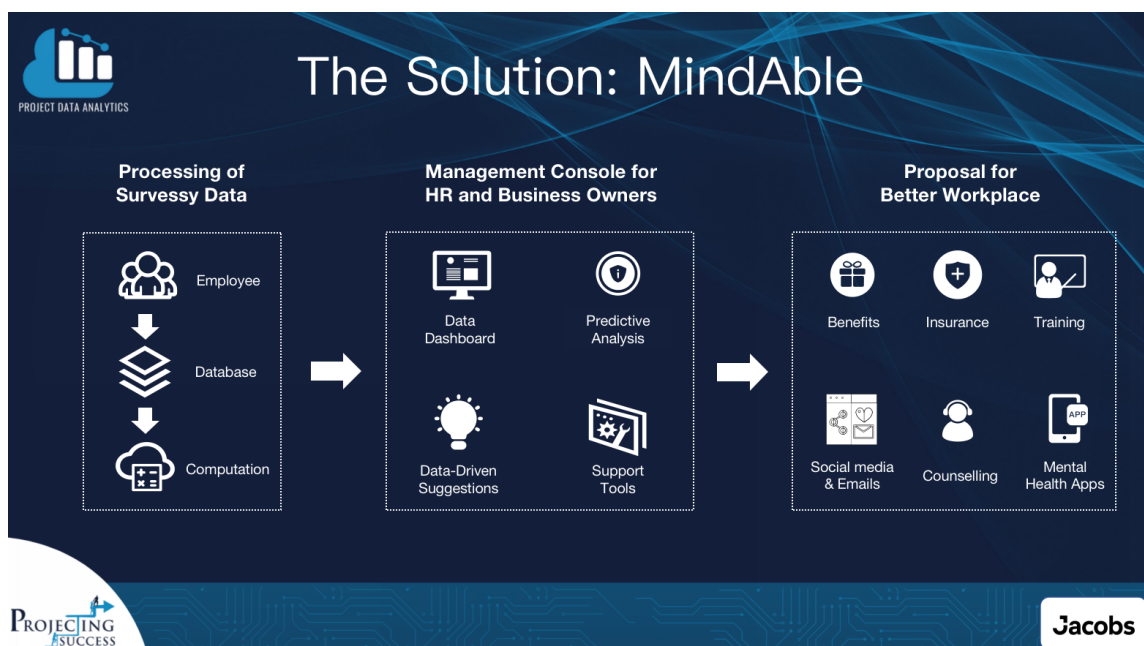
- Analyse the differing attitudes towards mental health and frequency of mental health disorders.
- Insight in how each employee would like to work to optimise their productivity whilst ensuring a positive workplace.
- Identify what factors need to be highlighted based on survey data and use data mining to create various proposals for implementing workplace improvements.

We face the difficulties of:

- How do we reduce the stigma surrounding mental health in the workplace?
- How do we identify what business can do to create a mentally resilient environment?
- Can we identify what factors play the biggest part in employee mental health?
- Can we create clear and concise action proposals based on data?

Introduction

The promoted solution 'MindAble' is designed for Challenge 4 'Healthy Minds, Healthy Lives' of Project Hack 16 by Team WorkMind. MindAble is a cross-platform web application. It is able to collect anonymous survey data from employees and perform automatic data mining and visualization techniques that help HR and business owners to understand the survey data, make a data-driven decision and create a better mentally resilient and productive environment.



Import Package and Data

```
In [53]: import pandas as pd
import pylab as pl
import matplotlib.pyplot as plt
import numpy as np
from sklearn.ensemble import RandomForestRegressor, RandomForestClassifier
```

```
In [2]: xls = pd.ExcelFile('./data/Mental Health Data.xlsx')
```

```
In [3]: # read multiple sheets
year = ['2014', '2016', '2017', '2018', '2019', '2020', '2021']

all_dfs = []
for y in year:
    df = pd.read_excel(xls, y)
    all_dfs.append(df)
```

```
In [4]: all_data = pd.concat(all_dfs)
```

```
In [5]: all_data.head()
```

Out[5]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	wo
0	2014-12-15 00:43:49	40.0	Male	United States	AL	No	Yes	Yes	
1	2014-12-01 09:58:28	31.0	Female	Belgium	NaN	No	No	No	
2	2014-11-16 08:42:35	38.0	Male	United States	AL	No	Yes	Yes	
3	2014-11-06 11:24:38	56.0	female	United States	OR	Yes	No	No	
4	2014-11-05 10:08:44	27.0	femail	United States	OK	No	No	No	

5 rows × 220 columns

```
In [6]: # print the column index
for c in all_data.columns:
    print (c)
```

Timestamp
 Age
 Gender
 Country
 state
 self_employed
 family_history
 treatment
 work_interfere
 no_employees
 remote_work
 tech_company
 benefits
 care_options
 wellness_program
 seek_help
 anonymity
 leave
 mental_health_consequence
 phys_health_consequence
 coworkers
 supervisor
 mental_health_interview
 phys_health_interview
 mental_vs_physical
 obs_consequence
 comments
 Are you self-employed?
 How many employees does your company or organization have?
 Is your employer primarily a tech company/organization?
 Is your primary role within your company related to tech/IT?
 Does your employer provide mental health benefits as part of healthcare coverage?
 Do you know the options for mental health care available under your employer-provided coverage?
 Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?
 Does your employer offer resources to learn more about mental health concerns and options for seeking help?
 Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?
 If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:
 Do you think that discussing a mental health disorder with your employer would have negative consequences?
 Do you think that discussing a physical health issue with your employer would have negative consequences?
 Would you feel comfortable discussing a mental health disorder with your coworkers?
 Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?
 Do you feel that your employer takes mental health as seriously as physical health?
 Have you heard of or observed negative consequences for co-workers who have been open about mental health issues in your workplace?
 Do you have medical coverage (private insurance or state-provided) which includes treatment of mental health issues?
 Do you know local or online resources to seek help for a mental health disorder?
 If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to clients or business contacts?
 If you have revealed a mental health issue to a client or business contact, do you believe this has impacted you negatively?
 If you have been diagnosed or treated for a mental health disorder, do you ever

ver reveal this to coworkers or employees?

If you have revealed a mental health issue to a coworker or employee, do you believe this has impacted you negatively?

Do you believe your productivity is ever affected by a mental health issue?

If yes, what percentage of your work time (time performing primary or secondary job functions) is affected by a mental health issue?

Do you have previous employers?

Have your previous employers provided mental health benefits?

Were you aware of the options for mental health care provided by your previous employers?

Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)?

Did your previous employers provide resources to learn more about mental health issues and how to seek help?

Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?

Do you think that discussing a mental health disorder with previous employers would have negative consequences?

Do you think that discussing a physical health issue with previous employers would have negative consequences?

Would you have been willing to discuss a mental health issue with your previous co-workers?

Would you have been willing to discuss a mental health issue with your direct supervisor(s)?

Did you feel that your previous employers took mental health as seriously as physical health?

Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?

Would you be willing to bring up a physical health issue with a potential employer in an interview?

Why or why not?

Would you bring up a mental health issue with a potential employer in an interview?

Why or why not?.1

Do you feel that being identified as a person with a mental health issue would hurt your career?

Do you think that team members/co-workers would view you more negatively if they knew you suffered from a mental health issue?

How willing would you be to share with friends and family that you have a mental illness?

Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?

Have your observations of how another individual who discussed a mental health disorder made you less likely to reveal a mental health issue yourself in your current workplace?

Do you have a family history of mental illness?

Have you had a mental health disorder in the past?

Do you currently have a mental health disorder?

If yes, what condition(s) have you been diagnosed with?

If maybe, what condition(s) do you believe you have?

Have you been diagnosed with a mental health condition by a medical professional?

If so, what condition(s) were you diagnosed with?

Have you ever sought treatment for a mental health issue from a mental health professional?

If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?

If you have a mental health issue, do you feel that it interferes with your work when NOT being treated effectively?

What is your age?

What is your gender?

What country do you live in?

What US state or territory do you live in?

What country do you work in?

What US state or territory do you work in?

Which of the following best describes your work position?

Do you work remotely?

Does your employer provide mental health benefits as part of healthcare coverage?

Do you know the options for mental health care available under your employer-provided health coverage?

Does your employer offer resources to learn more about mental health disorders and options for seeking help?

If a mental health issue prompted you to request a medical leave from work, how easy or difficult would it be to ask for that leave?

Would you feel more comfortable talking to your coworkers about your physical health or your mental health?

Would you feel comfortable discussing a mental health issue with your direct supervisor(s)?

Have you ever discussed your mental health with your employer?

Describe the conversation you had with your employer about your mental health, including their reactions and what actions were taken to address your mental health issue/questions.

Would you feel comfortable discussing a mental health issue with your coworkers?

Have you ever discussed your mental health with coworkers?

Describe the conversation with coworkers you had about your mental health including their reactions.

Have you ever had a coworker discuss their or another coworker's mental health with you?

Describe the conversation your coworker had with you about their mental health (please do not use names).

Overall, how much importance does your employer place on physical health?

Overall, how much importance does your employer place on mental health?

Do you have medical coverage (private insurance or state-provided) that includes treatment of mental health disorders?

Do you know local or online resources to seek help for a mental health issue?

If you have revealed a mental health disorder to a client or business contact, how has this affected you or the relationship?

If you have revealed a mental health disorder to a coworker or employee, how has this impacted you or the relationship?

Was your employer primarily a tech company/organization?

Did your previous employers provide resources to learn more about mental health disorders and how to seek help?

Would you have felt more comfortable talking to your previous employer about your physical health or your mental health?

Would you have been willing to discuss your mental health with your direct supervisor(s)?

Did you ever discuss your mental health with your previous employer?

Describe the conversation you had with your previous employer about your mental health, including their reactions and actions taken to address your mental health issue/questions.

Would you have been willing to discuss your mental health with your coworkers at previous employers?

Did you ever discuss your mental health with a previous coworker(s)?

Describe the conversation you had with your previous coworkers about your mental health including their reactions.

Did you ever have a previous coworker discuss their or another coworker's mental health with you?

Describe the conversation your coworker had with you about their mental health (please do not use names)..1

Overall, how much importance did your previous employer place on physical health?

Overall, how much importance did your previous employer place on mental health?

Have you ever been diagnosed with a mental health disorder?

Anxiety Disorder (Generalized, Social, Phobia, etc)

Mood Disorder (Depression, Bipolar Disorder, etc)
 Psychotic Disorder (Schizophrenia, Schizoaffective, etc)
 Eating Disorder (Anorexia, Bulimia, etc)
 Attention Deficit Hyperactivity Disorder
 Personality Disorder (Borderline, Antisocial, Paranoid, etc)
 Obsessive-Compulsive Disorder
 Post-Traumatic Stress Disorder
 Stress Response Syndromes
 Dissociative Disorder
 Substance Use Disorder
 Addictive Disorder
 Other
 Anxiety Disorder (Generalized, Social, Phobia, etc).1
 Mood Disorder (Depression, Bipolar Disorder, etc).1
 Psychotic Disorder (Schizophrenia, Schizoaffective, etc).1
 Eating Disorder (Anorexia, Bulimia, etc).1
 Attention Deficit Hyperactivity Disorder.1
 Personality Disorder (Borderline, Antisocial, Paranoid, etc).1
 Obsessive-Compulsive Disorder.1
 Post-traumatic Stress Disorder
 Stress Response Syndromes.1
 Dissociative Disorder.1
 Substance Use Disorder.1
 Addictive Disorder.1
 Other.1
 Anxiety Disorder (Generalized, Social, Phobia, etc).2
 Mood Disorder (Depression, Bipolar Disorder, etc).2
 Psychotic Disorder (Schizophrenia, Schizoaffective, etc).2
 Eating Disorder (Anorexia, Bulimia, etc).2
 Attention Deficit Hyperactivity Disorder.2
 Personality Disorder (Borderline, Antisocial, Paranoid, etc).2
 Obsessive-Compulsive Disorder.2
 Post-traumatic Stress Disorder.1
 Stress Response Syndromes.2
 Dissociative Disorder.2
 Substance Use Disorder.2
 Addictive Disorder.2
 Other.2

Have you ever sought treatment for a mental health disorder from a mental health professional?

If you have a mental health disorder, how often do you feel that it interferes with your work **when being treated effectively?**

If you have a mental health disorder, how often do you feel that it interferes with your work **when *NOT* being treated effectively (i.e., when you are experiencing symptoms)?**

Have your observations of how another individual who discussed a mental health issue made you less likely to reveal a mental health issue yourself in your current workplace?

Would you bring up your mental health with a potential employer in an interview?

Are you openly identified at work as a person with a mental health issue?

Has being identified as a person with a mental health issue affected your career?

How has it affected your career?

If they knew you suffered from a mental health disorder, how do you think that team members/co-workers would react?

Describe the circumstances of the badly handled or unsupportive response.

Have you observed or experienced supportive or well handled response to a mental health issue in your current or previous workplace?

Describe the circumstances of the supportive or well handled response.

Overall, how well do you think the tech industry supports employees with mental health issues?

Briefly describe what you think the industry as a whole and/or employers could do to improve mental health support for employees.

If there is anything else you would like to tell us that has not been covered by the survey questions, please use this space to do so.

Would you be willing to talk to one of us more extensively about your experiences with mental health issues in the tech industry? (Note that all interview responses would be used *anonymously* and only with your permission.)

What country do you **live** in?

What US state or territory do you **live** in?

What is your race?

Other.3

What country do you **work** in?

What US state or territory do you in?

Start Date (UTC)

Submit Date (UTC)

Network ID

#

Are you self-employed?

If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to clients or business contacts?

If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to coworkers or employees?

Do you have previous employers?

Have your previous employers provided mental health benefits?

Were you aware of the options for mental health care provided by your previous employers?

Would you have been willing to discuss your mental health with your coworkers at previous employers?

If you have a mental health disorder, how often do you feel that it interferes with your work **when** *NOT* **> being treated effectively (i.e., when you are experiencing symptoms)?**

If they knew you suffered from a mental health disorder, how do you think that team members/co-workers would react?

Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?

Have you observed or experienced supportive or well handled response to a mental health issue in your current or previous workplace?

Would you be willing to talk to one of us more extensively about your experiences with mental health issues in the tech industry? (Note that all interview responses would be used *anonymously* and only with your permission.)

What US state or territory do you **work** in?

Are you self-employed?

Do you have previous employers?

Do you *currently* have a mental health disorder?

Have you ever been *diagnosed* with a mental health disorder?

What disorder(s) have you been diagnosed with?

If possibly, what disorder(s) do you believe you have?

If so, what disorder(s) were you diagnosed with?

If you have a mental health disorder, how often do you feel that it interferes with your work *when being treated effectively*?

If you have a mental health disorder, how often do you feel that it interferes with your work *when* **NOT** *being treated effectively (i.e., when you are experiencing symptoms)?*

Would you bring up your *mental* health with a potential employer in an interview?

If they knew you suffered from a mental health disorder, how do you think that at your team members/co-workers would react?

Have you observed or experienced an *unsupportive or badly handled response* to a mental health issue in your current or previous workplace?

Have you observed or experienced a *supportive or well handled response* to a mental health issue in your current or previous workplace?

Would you be willing to talk to one of us more extensively about your experiences with mental health issues in the tech industry? (Note that all interview responses would be used _anonymously_ and only with your permission.)
 What country do you **live** in?
 What US state or territory do you **live** in?
 What country do you **work** in?
 What US state or territory do you **work** in?
 Have you been diagnosed with COVID-19?

Processing of Surveys Data

Data Cleaning and Preprocessing

Since the columns are based on the survey question, there is no label of people who has mental health issue or not. We need to find out the target of people suffering the mental health problems. In the datasheet, we find out there are too many hypothetical questions. Therefore, we define TWO levels of mental health condition for employees completing their survey and they are:

- Type I: Potentially suffer from mental health disorders
- Type II: Already suffer from mental health disorders

Find employees who potentially or already suffer mental health problems

```
In [7]: # effective questions for determining the types
target_col = [
    'treatment', # Have you sought treatment for a mental health condition?
    'work_interfere', # If you have a mental health condition, do you feel t
    'mental_health_consequence', # Do you think that discussing a mental hea
    'coworkers', # Would you be willing to discuss a mental health issue wit
    'mental_health_consequence', # Do you think that discussing a mental hea
    'mental_health_interview', # Does the individual currently have mental h
    'Would you feel comfortable discussing a mental health disorder with you
    'If you have been diagnosed or treated for a mental health disorder, do
    'If you have revealed a mental health issue to a client or business cont
    'Do you believe your productivity is ever affected by a mental health is
    'Would you feel comfortable discussing a mental health issue with your d
    'Have you ever discussed your mental health with your employer?',
    'If you have been diagnosed or treated for a mental health disorder, do
    'Do you currently have a mental health disorder?',
    'If you have a mental health disorder, how often do you feel that it int
    'Have you ever sought treatment for a mental health disorder from a ment
]

# visuliaz the survey answers subject to questions
```

```

for t in target_col:
    print ('=====')
    print (t)
    print(all_data[t].unique())

```

```

=====
treatment
['Yes' 'No' nan]
=====
work_interfere
['Sometimes' nan 'Rarely' 'Never' 'Often']
=====
mental_health_consequence
['Maybe' 'No' 'Yes' nan]
=====
coworkers
['Some of them' 'No' 'Yes' nan]
=====
mental_health_consequence
['Maybe' 'No' 'Yes' nan]
=====
mental_health_interview
['No' 'Maybe' 'Yes' nan]
=====
Would you feel comfortable discussing a mental health disorder with your coworkers?
[nan 'Maybe' 'Yes' 'No']
=====
If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to clients or business contacts?
[nan 'Sometimes, if it comes up' "No, because it doesn't matter"
 'No, because it would impact me negatively' 'Not applicable to me'
 'Yes, always']
=====
If you have revealed a mental health issue to a client or business contact, do you believe this has impacted you negatively?
[nan "I'm not sure" 'Yes' 'No']
=====
Do you believe your productivity is ever affected by a mental health issue?
[nan 'Yes' 'Not applicable to me' 'No' 'Unsure']
=====
Would you feel comfortable discussing a mental health issue with your direct supervisor(s)?
[nan 'Yes' 'Maybe' 'No']
=====
Have you ever discussed your mental health with your employer?
[nan 0. 1.]
=====
If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to coworkers or employees?
[nan 'Sometimes, if it comes up'
 'No, because it would impact me negatively' 'Not applicable to me'
 'Yes, always' "No, because it doesn't matter"]
=====
Do you currently have a mental health disorder?
[nan 'No' 'Yes' 'Maybe' 'Possibly' "Don't Know"]
=====
If you have a mental health disorder, how often do you feel that it interferes with your work <strong>when being treated effectively?</strong>
[nan 'Sometimes' 'Not applicable to me' 'Often' 'Rarely' 'Never']
=====
Have you ever sought treatment for a mental health disorder from a mental health professional?
[nan 1 0]

```

```
In [8]: def match_type_I(index, df):
    matched_df = []

    if index in ['coworkers',
                 'Would you feel comfortable discussing a mental health disc
    ]:
        matched_df = df[ (df[index]=='Some of them') | (df[index]=='No') | (

    if index == 'mental_health_consequence':
        matched_df = df[ (df[index]=='Sometimes') | (df[index]=='Often')]

    if index in ['mental_health_interview',
                 'work_interfere',
    ]:
        matched_df = df[ (df[index]=='Maybe')]

    if index in ['If you have been diagnosed or treated for a mental health
                 'If you have been diagnosed or treated for a mental health d
    ]:
        matched_df = df[ (df[index]=='No, because it would impact me negativ
                        | (df[index]=='Yes, always') | (df[index]=='Sometime
                        ]

    if index in ['If you have revealed a mental health issue to a client or
                 'Would you feel comfortable discussing a mental health issu
    ] :
        matched_df = df[ (df[index]=='Yes') | (df[index]=='No') ]

    if index == 'Do you currently have a mental health disorder?' :
        matched_df = df[ (df[index]=='Maybe') | (df[index]=='Possibly')]

    return matched_df
```

```
In [9]: def match_type_II(index, df):
    matched_df = []
    if index in ['treatment',
                 'Have you ever sought treatment for a mental health disorde
    matched_df = df[ (df[index]=='Yes') | (df[index]=='No')]

    if index in ['mental_health_interview',
                 'Do you currently have a mental health disorder?',
                 'Do you believe your productivity is ever affected by a men
    ]:
        matched_df = df[ (df[index]=='Yes')]

    if index in ['Have you ever discussed your mental health with your emplo
                 'Do you currently have a mental health disorder?'
    ]:
        matched_df = df[ (df[index]==1) | (df[index]==0)]

    if index == 'If you have a mental health disorder, how often do you feel
        matched_df = df[ (df[index]=='Sometimes') | (df[index]=='Often') | (d

    return matched_df
```

```
In [10]: type_I_dfs = []
    type_II_dfs = []

    for t in target_col:
        matched_df = match_type_I(t, all_data)
        if len(matched_df)>0:
            type_I_dfs.append(matched_df)
```

```

    matched_df = match_type_II(t, all_data)
    if len(matched_df)>0:
        type_II_dfs.append(matched_df)

type_I_dfs = pd.concat(type_I_dfs)
type_II_dfs = pd.concat(type_II_dfs)

```

```
In [11]: len(type_I_dfs)
```

```
Out[11]: 4279
```

```
In [12]: len(type_II_dfs)
```

```
Out[12]: 3204
```

Get the rows where people are likely mental healthy

```

In [13]: target_dfs = pd.merge(type_I_dfs, type_I_dfs)
         # get the rows where people are likly mental healthy
         df_merge = pd.merge(all_data, target_dfs, how='inner')
         all_data = all_data.append(df_merge)

         all_data['Duplicated'] = all_data.duplicated(keep=False) # keep=False marks
         df_final = all_data[~all_data['Duplicated']] # selects only rows which are n
         del df_final['Duplicated'] # delete the indicator column

```

```

/var/folders/dt/_85w7mvd1h5ff3x6kqgzbd3w0000gn/T/ipykernel_8076/3767860516.p
y:4: FutureWarning: The frame.append method is deprecated and will be remove
d from pandas in a future version. Use pandas.concat instead.
    all_data = all_data.append(df_merge)

```

```

In [14]: healthy_df = df_final
         healthy_df

```

Out[14]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	v
0	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	
116	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
123	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
127	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
129	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
130	NaT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

995 rows × 220 columns

```
In [15]: healthy_df['labels'] = '0'
         type_I_dfs['labels'] = '1'
         type_II_dfs['labels'] = '2'
         all_dfs = pd.concat([healthy_df, type_I_dfs, type_II_dfs])
```

/var/folders/dt/_85w7mvd1h5ff3x6kqwzbd3w0000gn/T/ipykernel_8076/2711298567.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

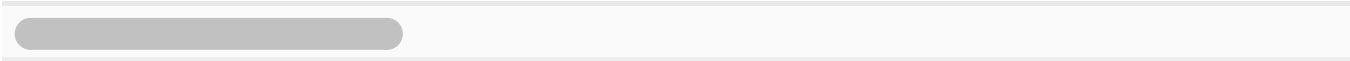
```
         healthy_df['labels'] = '0'
```

```
In [16]: all_dfs.describe()
```

Out[16]:

	Age	Are you self-employed?	Is your employer primarily a tech company/organization?	Is your primary role within your company related to tech/IT?	Do you have medical coverage (private insurance or state-provided) which includes treatment of a mental health issues?	Do you have a mental health condition?
count	2.619000e+03	3646.000000	4567.000000	3582.000000	730.000000	3646
mean	1.145476e+08	0.273999	0.737902	0.929369	0.657534	0.657534
std	3.383194e+09	0.446070	0.439824	0.256243	0.474860	0.474860
min	-1.726000e+03	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.700000e+01	0.000000	0.000000	1.000000	0.000000	1.000000
50%	3.100000e+01	0.000000	1.000000	1.000000	1.000000	1.000000
75%	3.600000e+01	1.000000	1.000000	1.000000	1.000000	1.000000
max	1.000000e+11	1.000000	1.000000	1.000000	1.000000	1.000000

8 rows x 47 columns



Data Dashboard

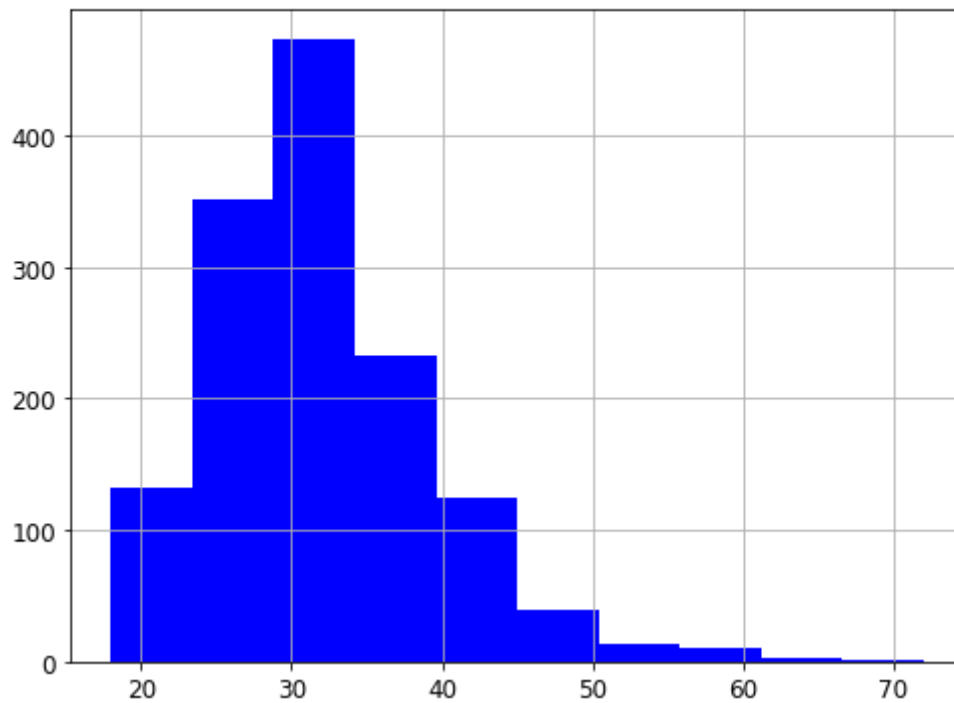
Insight I: Employees' Background

Age

```
In [17]: plt.rcParams.update({'font.size': 12})
tmp1 = type_I_dfs[ (type_I_dfs['Age']>=18) & (type_I_dfs['Age']<= 99)]
pl.suptitle("Age Distribution for Type I: Potentially suffer from mental health issues")
tmp1 ['Age'].hist(figsize=(8, 6), color = "Blue")
```

Out[17]: <AxesSubplot:>

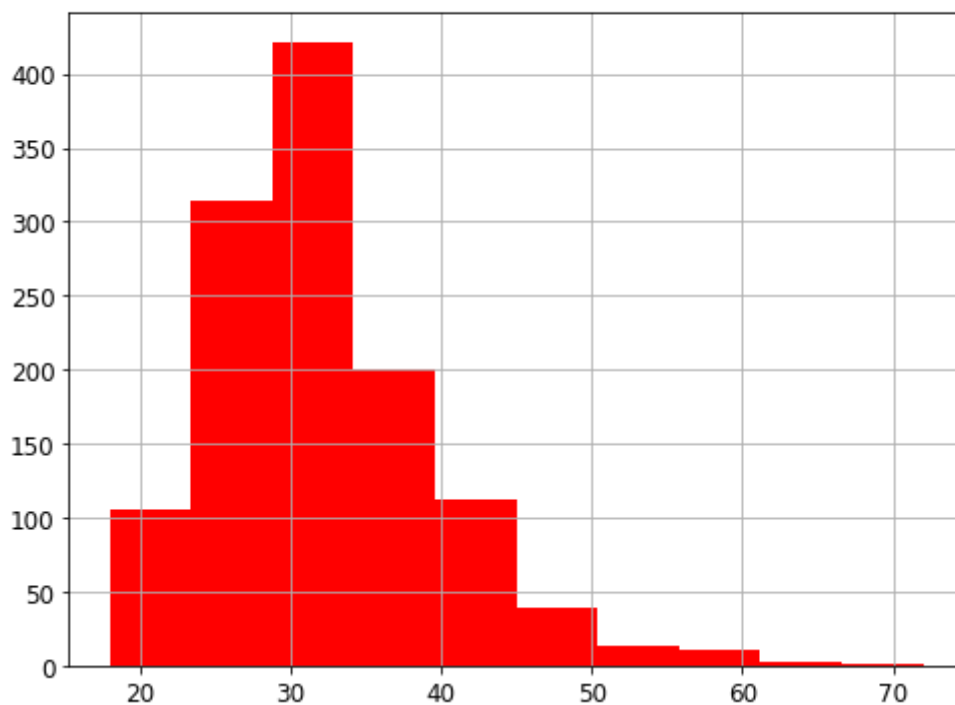
Age Distribution for Type I: Potentially suffer from mental health disorders



```
In [18]: tmp2 = type_II_dfs[ (type_II_dfs['Age']>=18) & (type_II_dfs['Age']<= 99)]  
pl.suptitle("Age Distribution for Type II: Already suffer from mental health  
tmp2 ['Age'].hist(figsize=(8, 6), color = "red")
```

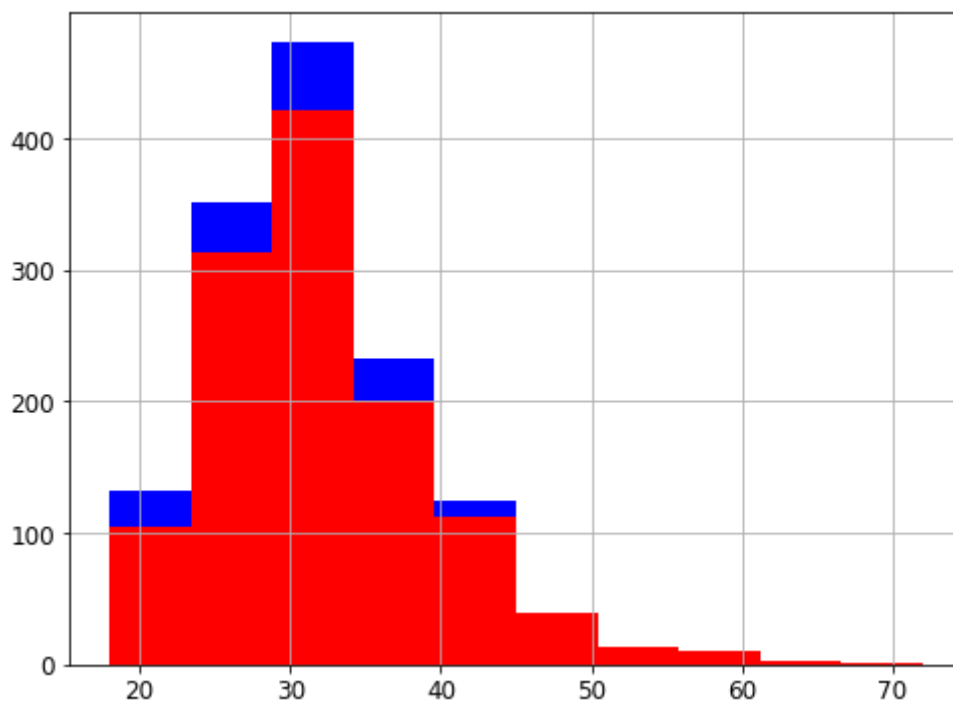
Out[18]: <AxesSubplot:>

Age Distribution for Type II: Already suffer from mental health disorders



```
In [19]: tmp1 ['Age'].hist(figsize=(8, 6), color = "Blue")  
tmp2 ['Age'].hist(figsize=(8, 6), color = "red")
```

Out[19]: <AxesSubplot:>

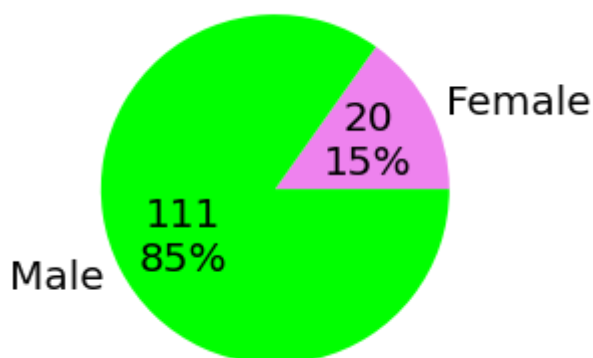


The age distributions of two types of mental health condition groups are similar

Gender

```
In [20]: def label_function(val):
          return f'{val / 100 * len(df):.0f}\n{val:.0f}%'
fig = plt.figure()
ax = fig.add_subplot(111)
tmp1 = type_I_dfs [ (type_I_dfs['Gender']=='Male') | (type_I_dfs['Gender']=='Female') ]
tmp1.groupby('Gender').size().plot(kind='pie', autopct=label_function, textp
ax.set_ylabel('')
```

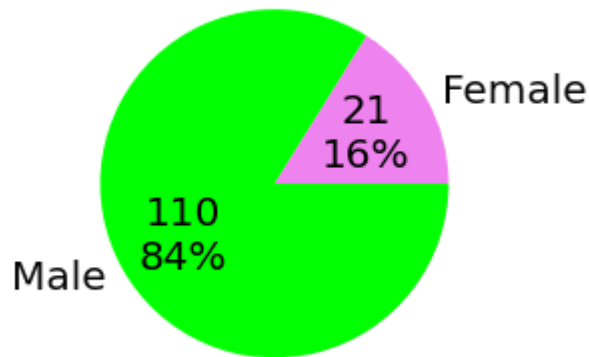
Out[20]: Text(0, 0.5, '')



```
In [21]: fig = plt.figure()
          ax = fig.add_subplot(111)
          tmp2 = type_II_dfs [ (type_II_dfs['Gender']=='Male') | (type_II_dfs['Gender']=='Female') ]
          tmp2.groupby('Gender').size().plot(kind='pie', autopct=label_function, textp
          ax.set_ylabel('')
```



```
Out[21]: Text(0, 0.5, '')
```

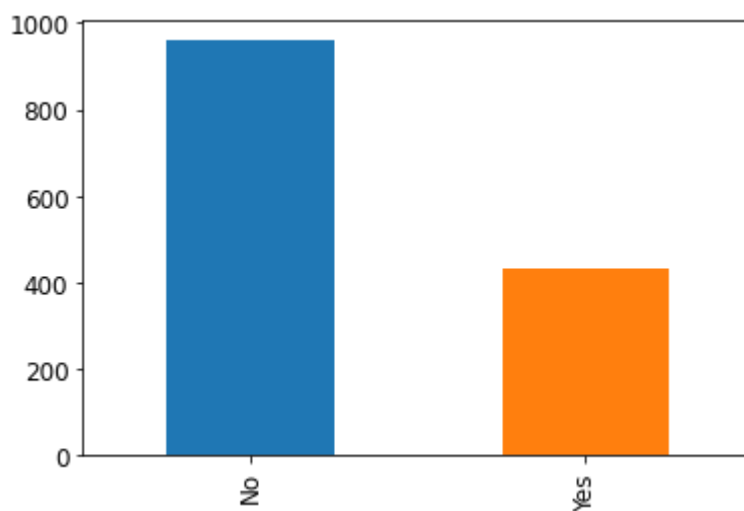


The gender distributions of two types of mental health condition groups are similar

Insight II: Work Environment

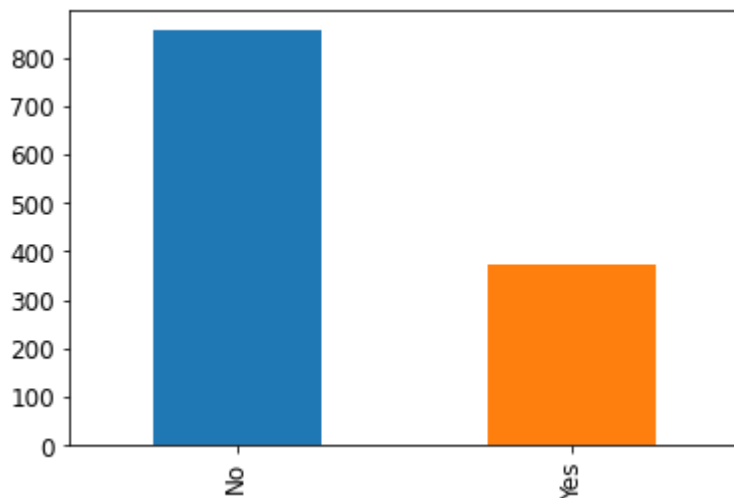
```
In [22]: tmp1 = type_I_dfs[(type_I_dfs['remote_work']=='Yes') | (type_I_dfs['remote_'  
tmp1['remote_work'].value_counts().plot(kind='bar',color=['C0', 'C1'])
```

```
Out[22]: <AxesSubplot:>
```



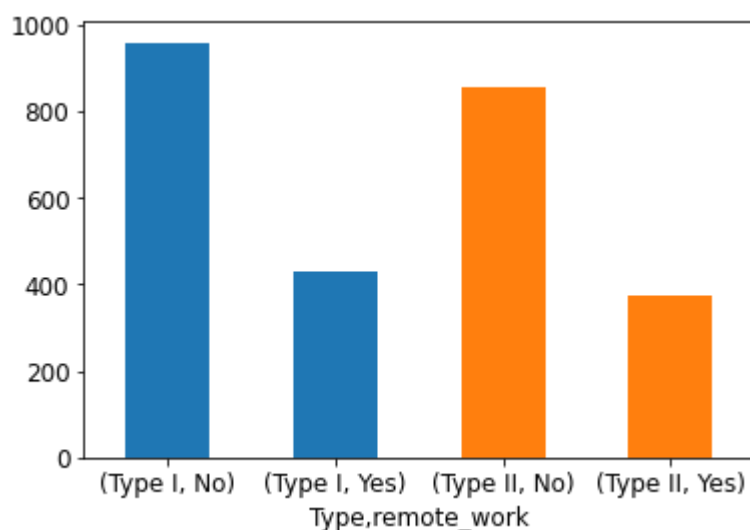
```
In [23]: tmp2 = type_II_dfs[(type_II_dfs['remote_work']=='Yes') | (type_II_dfs['remo  
tmp2['remote_work'].value_counts().plot(kind='bar',color=['C0', 'C1'])
```

```
Out[23]: <AxesSubplot:>
```



```
In [24]: tmp = pd.concat([tmp1, tmp2])
tmp['Type'] = (len(tmp1)*('Type I',) + len(tmp2)*('Type II',))
tmp.groupby(['Type']).remote_work.value_counts().plot(kind='bar',color=['C0'
```

```
Out[24]: <AxesSubplot:xlabel='Type,remote_work'>
```



Insight III: Mental Health Domain

Association with Physical Health

```
In [25]: # Find the questions related to physical health
keyword = 'physical'
key_cols = []
for col in all_data.columns:
    if keyword in col:
        key_cols.append(col)
print(key_cols)
```

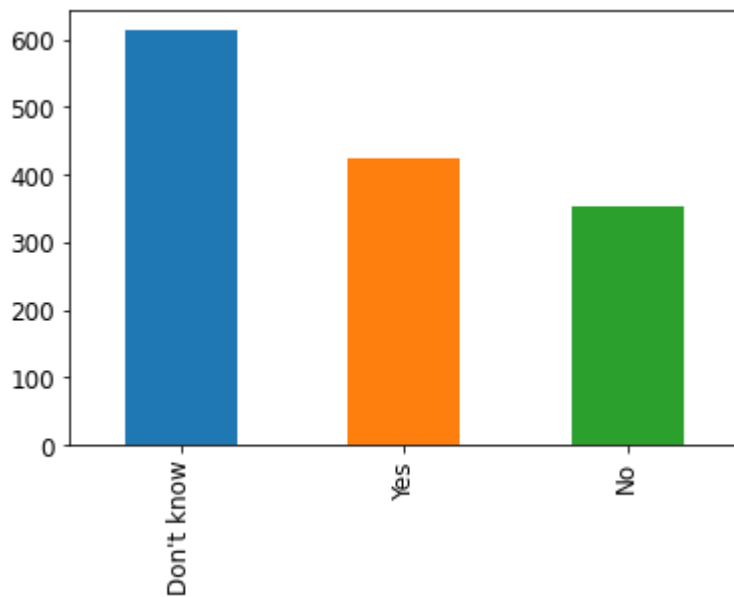
```
['mental_vs_physical', 'Do you think that discussing a physical health issue
with your employer would have negative consequences?', 'Do you feel that you
r employer takes mental health as seriously as physical health?', 'Do you th
ink that discussing a physical health issue with previous employers would ha
ve negative consequences?', 'Did you feel that your previous employers took
mental health as seriously as physical health?', 'Would you be willing to br
ing up a physical health issue with a potential employer in an interview?',
'Would you feel more comfortable talking to your coworkers about your physic
al health or your mental health?', 'Overall, how much importance does your e
mployer place on physical health?', 'Would you have felt more comfortable ta
lking to your previous employer about your physical health or your mental he
alth?', 'Overall, how much importance did your previous employer place on ph
ysical health?']
```

```
In [26]: for index in key_cols:
          print ('=====')
          print (index)
          print (all_data[index].unique())

=====
mental_vs_physical
['Yes' "Don't know" 'No' nan]
=====
Do you think that discussing a physical health issue with your employer woul
d have negative consequences?
[nan 'No' 'Maybe' 'Yes']
=====
Do you feel that your employer takes mental health as seriously as physical
health?
[nan "I don't know" 'Yes' 'No']
=====
Do you think that discussing a physical health issue with previous employers
would have negative consequences?
[nan 'None of them' 'Some of them' 'Yes, all of them']
=====
Did you feel that your previous employers took mental health as seriously as
physical health?
[nan "I don't know" 'Some did' 'None did' 'Yes, they all did']
=====
Would you be willing to bring up a physical health issue with a potential em
ployer in an interview?
[nan 'Maybe' 'Yes' 'No']
=====
Would you feel more comfortable talking to your coworkers about your physica
l health or your mental health?
[nan 'Same level of comfort for each' 'Physical health' 'Mental health']
=====
Overall, how much importance does your employer place on physical health?
[nan 6. 7. 0. 9. 10. 1. 5. 3. 8. 4. 2.]
=====
Would you have felt more comfortable talking to your previous employer about
your physical health or your mental health?
[nan 'Physical health' 'Same level of comfort for each' 'Mental health']
=====
Overall, how much importance did your previous employer place on physical he
alth?
[nan 3. 5. 8. 7. 10. 9. 2. 4. 1. 6. 0.]
```

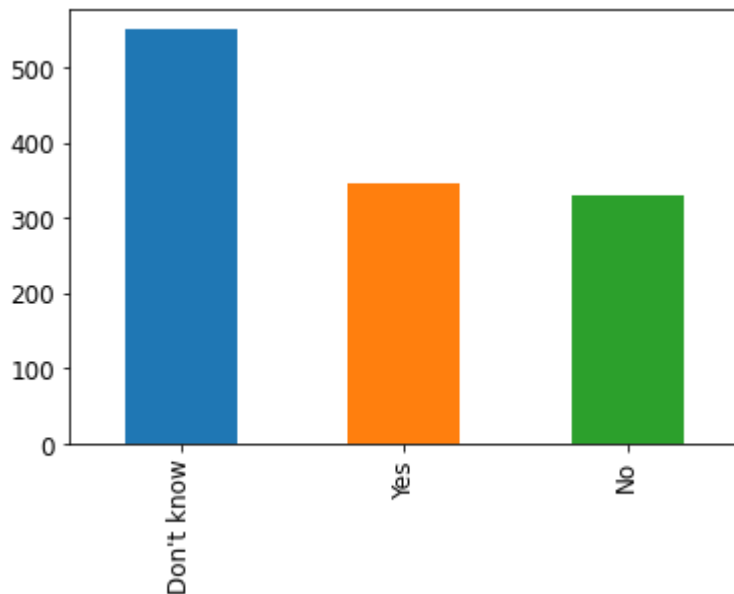
```
In [27]: tmp1 = type_I_dfs[(type_I_dfs['mental_vs_physical']== 'Yes') | (type_I_dfs['
tmp1['mental_vs_physical'].value_counts().plot(kind='bar',color=['C0', 'C1'
```

```
Out[27]: <AxesSubplot:>
```



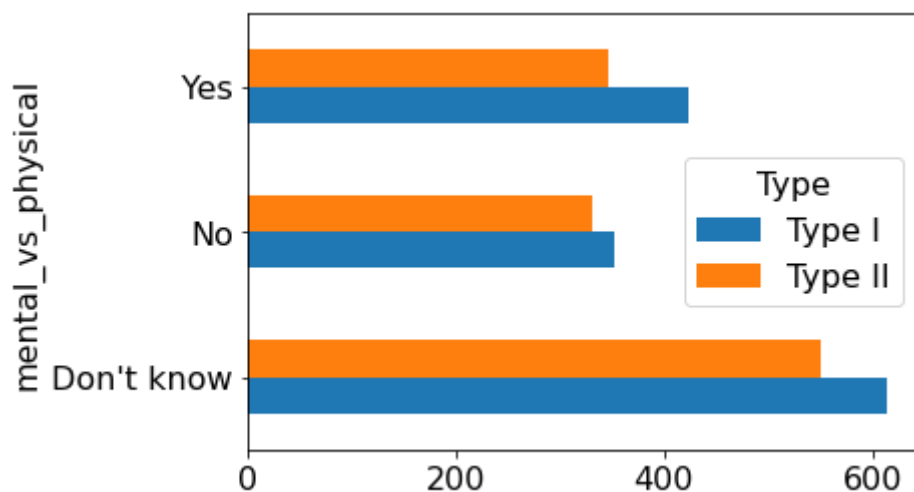
```
In [28]: tmp2 = type_II_dfs[(type_II_dfs['mental_vs_physical']=='Yes') | (type_II_dfs['mental_vs_physical']=='No')]
tmp2['mental_vs_physical'].value_counts().plot(kind='bar',color=['C0', 'C1'])
```

Out[28]: <AxesSubplot:>



```
In [29]: plt.rcParams.update({'font.size': 16})
tmp = pd.concat([tmp1, tmp2])
tmp['Type'] = (len(tmp1)*('Type I',) + len(tmp2)*('Type II',))
tmp.groupby(['Type']).mental_vs_physical.value_counts().unstack(0).plot.barh
```

Out[29]: <AxesSubplot:ylabel='mental_vs_physical'>

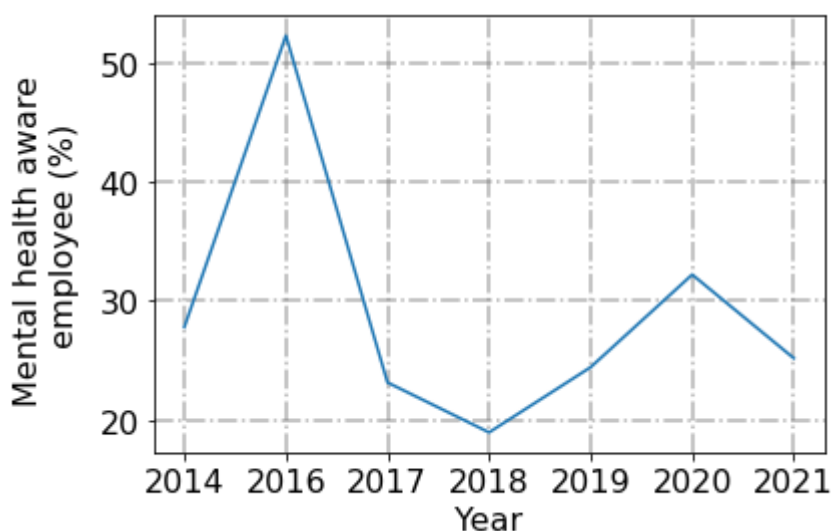


Insight IV: Data Changes

```
In [30]: # read multiple sheets
year = ['2014', '2016', '2017', '2018', '2019', '2020', '2021']

count_by_year = []
for y in year:
    df = pd.read_excel(xls, y)
    count = 0
    for index in key_cols:
        try:
            count += len(df[(df[index] == 'Yes')])
        except:
            pass
    count_by_year.append(count/len(df)*100)
```

```
In [31]: plt.plot(year, count_by_year)
plt.xlabel("Year")
plt.ylabel("Mental health aware \n employee (%) ", labelpad=10)
plt.grid(visible=True, which='major', color='gray', alpha=0.6, linestyle='dashed')
plt.show()
```



Predictive Analysis

Data Preparation

```
In [32]: # Filling missing values
all_dfs = all_dfs.replace(np.nan, 0)

# Encoding categorical variables
all_dfs_enconde = pd.get_dummies(all_dfs)

# Filling missing values
all_dfs_enconde = all_dfs_enconde.replace(np.nan, 0)

/Users/rc/miniforge3/envs/ds/lib/python3.10/site-packages/pandas/core/algorithms.py:798: FutureWarning: In a future version, the Index constructor will not infer numeric dtypes when passed object-dtype sequences (matching Series behavior)
    uniques = Index(uniques)
```

```
In [33]: all_dfs_enconde
```

Out[33]:

	Age	Are you self-employed?	Is your employer primarily a tech company/organization?	Is your primary role within your company related to tech/IT?	Do you have medical coverage (private or state-provided) which includes treatment of a mental health issues?	Do you have previous employers?	Have you ever sought treatment for a mental health issue from a professional?
0	0.0	0.0	1.0	0.0	0.0	1.0	0.0
1	0.0	0.0	1.0	0.0	0.0	1.0	0.0
2	0.0	0.0	1.0	0.0	0.0	1.0	0.0
4	0.0	0.0	0.0	1.0	0.0	1.0	0.0
5	0.0	0.0	1.0	0.0	0.0	1.0	0.0
...
122	0.0	0.0	0.0	1.0	0.0	0.0	0.0
124	0.0	0.0	1.0	1.0	0.0	0.0	0.0
126	0.0	0.0	1.0	1.0	0.0	0.0	0.0
127	0.0	0.0	1.0	1.0	0.0	0.0	0.0
130	0.0	0.0	1.0	1.0	0.0	0.0	0.0

8478 rows x 17533 columns

```
In [34]: x = all_dfs_enconde.iloc[:, 0:-3]
y = all_dfs_enconde.iloc[:, -3:]
```

Training

```
In [45]: # create regressor object
regressor = RandomForestRegressor(n_estimators = 100,max_depth=2, random_state=0)

# fit the regressor with x and y data
regressor.fit(X, y)
```

```
Out[45]: RandomForestRegressor(max_depth=2, random_state=0)
```

Inferencing

```
In [50]: regressor.predict(X)
```

```
Out[50]: array([[0.80046471, 0.19953529, 0.          ],
                [0.80046471, 0.19953529, 0.          ],
                [0.80046471, 0.19953529, 0.          ],
                ...,
                [0.02720948, 0.5838874 , 0.38890312],
                [0.47875669, 0.02001161, 0.5012317 ],
                [0.47875669, 0.02001161, 0.5012317 ]])
```