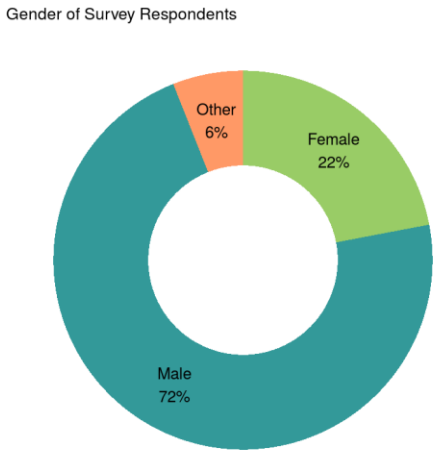
MENTAL HEALTH IN TECH SURVEY

Presented By:
BARANEETHARAN S
DHANUSH T
MAHESH KUMAR M
BARATH S
JEYANTH P



PROBLEM DEFINITION

The problem in the context of mental health in the tech industry is multifaceted. It includes issues such as high levels of stress, burnout, imposter syndrome, work-life imbalance, and stigma surrounding mental health. Employees in the tech sector often face intense workloads, tight deadlines, and a competitive environment, which can lead to mental health challenges.

Design Thinking Approach

Empathize:

Conduct surveys and interviews with tech industry professionals to understand their mental health challenges, triggers, and coping mechanisms. Gather data on the prevalence of mental health issues in the tech industry and their impact on productivity and well-being.

Define:

Define specific problem areas within mental health, such as stress management, work-life balance, or stigma reduction. Create user personas to represent different segments of the tech workforce, considering factors like role, experience level, and personal circumstances.

Ideate:

Brainstorm potential solutions and interventions to address the identified problems. Encourage diverse perspectives from tech employees, mental health experts, and HR professionals.

Prototype:

Develop prototypes of mental health support programs or initiatives, such as employee assistance programs, peer support networks, or stress reduction workshops. Test these prototypes with a small group of tech industry professionals to gather feedback.

Test:

Collect feedback from the pilot tests and iterate on the prototypes. Assess the effectiveness of the interventions in improving mental health outcomes and overall job satisfaction.

Implement:

Roll out the refined solutions on a larger scale, making them accessible to a broader range of tech employees. Develop communication strategies to promote mental health awareness and encourage participation.

Evaluate:

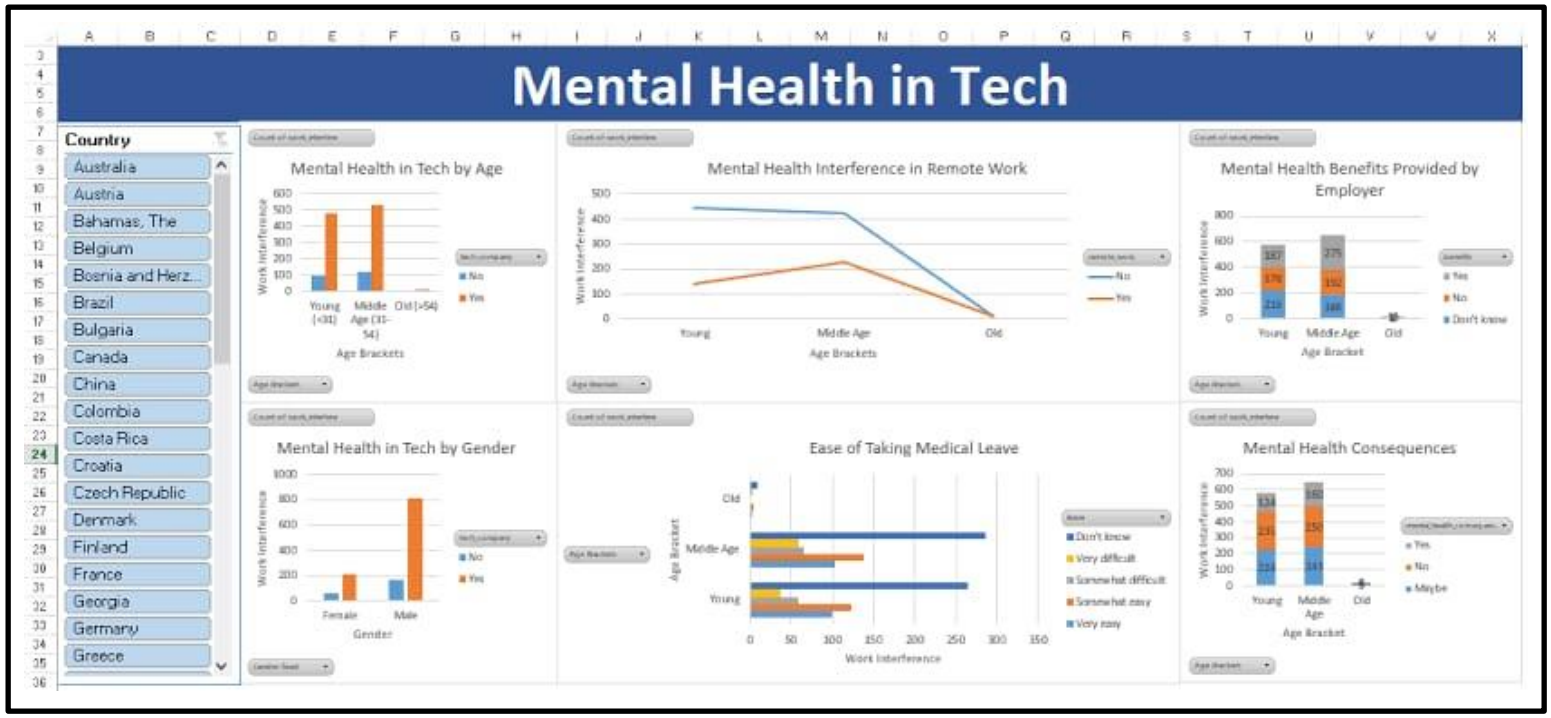
Continuously monitor and evaluate the impact of the implemented solutions on mental health metrics, such as stress levels, absenteeism, and employee turnover. Make necessary adjustments and improvements based on ongoing feedback and data analysis.

Iterate:

Use a cyclical design thinking approach to continuously refine and enhance mental health support initiatives in the tech industry.

Remember that mental health is a complex and evolving issue, so an ongoing commitment to design thinking and adaptability is crucial to address the unique challenges faced by tech professionals effectively.

MENTAL HEALTH IN TECH SURVEY



- **Anonymous Peer Support Platform:**

  Create a platform where tech professionals can anonymously connect with peers facing similar mental health challenges, fostering a supportive community

Mindfulness and Meditation Apps:

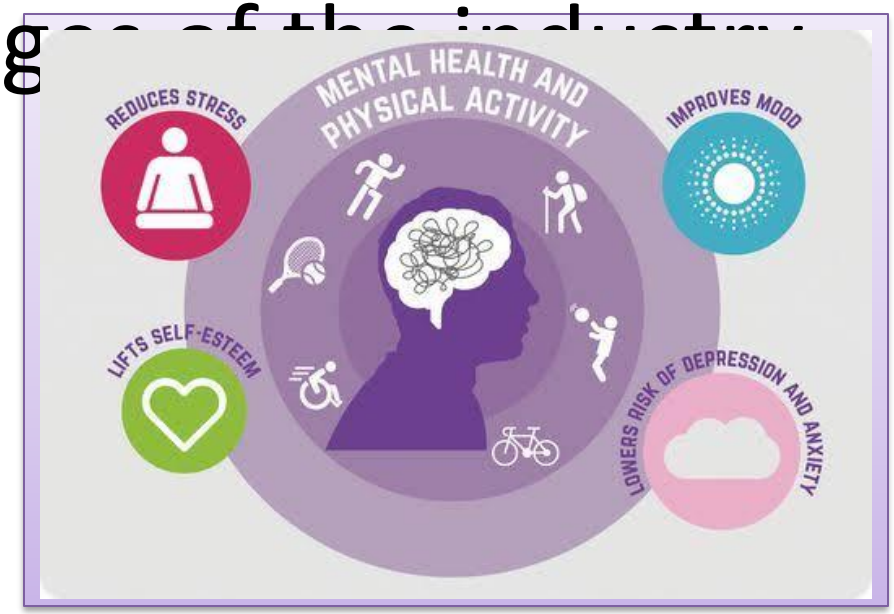
Integrate brief mindfulness and meditation exercises within the tech environment through apps, promoting mental well-being during work hours.

Flexible Work Hours:

Implement flexible work hours or remote work options to reduce stress and accommodate different personal schedules, contributing to a healthier work-life balance.

Virtual Mental Health Workshops:

Host virtual workshops on stress management, mindfulness, and resilience, providing tech employees with tools to cope with the unique challenges of the industry.



Regular Mental Health Check-ins :

Conduct regular, anonymous mental health check-ins to gauge the well-being of tech professionals and identify areas where additional support may be needed.

Educational Campaigns:

Launch campaigns that raise awareness about mental health in the tech industry, helping to destigmatize seeking help and fostering a

Mental Health Training for Managers :

Provide training to managers on recognizing signs of mental health issues, fostering a supportive work environment, and effectively addressing mental health concerns within their teams.

Employee Assistance Programs (EAP):

Offer comprehensive EAPs that include counseling services, workshops, and resources to support mental health needs dential manner.

Personalized Wellness Plans:

Develop personalized mental health and wellness plans for employees, considering individual preferences and needs, and prov... accordingly.

DATA SECURITY AND USER PRIVACY SHOULD BE PRIORITIZED BEFORE IMPLEMENTING THESE ABOVE IDEAS

Data Preprocessing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

Step 1: Remove duplicate or irrelevant observations

Remove unwanted observations from your dataset, including duplicate observations or irrelevant observations. Duplicate observations will happen most often during data collection. When you combine data sets from multiple places, scrape data, or receive data from clients or multiple departments, there are opportunities to create duplicate data.

Step 2: Fix structural errors

Structural errors are when you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization. These inconsistencies can cause mislabeled categories or classes. For example, you may find “N/A” and “Not Applicable” both appear, but they should be analyzed as the same category.

Step 3: Filter unwanted outliers

Often, there will be one-off observations where, at a glance, they do not appear to fit within the data you are analyzing. If you have a legitimate reason to remove an outlier, like improper data-entry, doing so will help the performance of the data you are working with. However, sometimes it is the appearance of an outlier that will prove a theory you are working on.

Step 4: Handle missing data

You can't ignore missing data because many algorithms will not accept missing values. There are a couple of ways to deal with missing data. Neither is optimal, but both can be considered.

Step 5: Validate and QA

At the end of the data cleaning process, you should be able to answer these questions as a part of basic validation:

Does the data make sense?

Does the data follow the appropriate rules for its field?

Importing the python libraries in jupyter noetbook(pandas,numpy)

Using pandas to read the dataset(Survey.csv):

[6]:
```
import numpy as np
import pandas as pd
```

[7]:
```
df=pd.read_csv('survey.csv')
```

[8]:
```
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
```

Using dataframe to verify the header of dataset

```
 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
```

[17]:
```
df.head
```

[17]:
```
<bound method NDFrame.head of              Timestamp  Age  Gender        Country state self_employed  \
0   2014-08-27 11:36:48   33    male  United States    CA            No
1   2014-08-27 11:37:08   35    male  United States    TN            No
2   2014-08-27 11:39:36   42    Male  United States    WA            No
3   2014-08-27 11:43:36   38  Female  United States    TX            No
4   2014-08-27 11:44:43   30    male  United States    IL            No
..                  ...  ...     ...            ...   ...           ...
81  2014-11-06 11:24:38   56  female  United States    OR           Yes
82  2015-01-03 03:38:30   44       M  United States    OH            No
83  2015-02-21 09:48:13   38    Male  United States    TX            No
84  2015-02-21 10:45:51   46    male  United States    MD            No
85  2015-07-22 18:57:54   30       M  United States    CA            No

   family_history treatment work_interfere    no_employees  ...  \
0             Yes       Yes         Rarely          26-100  ...
1             Yes       Yes      Sometimes  More than 1000  ...
2             Yes       Yes      Sometimes          26-100  ...
3             Yes       Yes      Sometimes          26-100  ...
4             Yes       Yes         Rarely          26-100  ...
..            ...       ...            ...             ...  ...
81             No        No         Rarely             1-5  ...
82            Yes       Yes      Sometimes         100-500  ...
83             No       Yes      Sometimes  More than 1000  ...
84            Yes       Yes      Sometimes         100-500  ...
85            Yes       Yes      Sometimes          26-100  ...
```

```
       coworkers    supervisor mental_health_interview phys_health_interview  \
0            Yes           Yes                      No                    Yes
1   Some of them           Yes                      No                    Yes
2   Some of them  Some of them                   Maybe                    Yes
3   Some of them           Yes                      No                     No
4   Some of them           Yes                      No                     No
..           ...           ...                     ...                    ...
81           No            No                       No                     No
82  Some of them  Some of them                      No                     No
83           Yes           Yes                      No                     No
84  Some of them           Yes                     Yes                    Yes
85           Yes           Yes                   Maybe                  Maybe

   mental_vs_physical obs_consequence  \
0         Don't know              No
1                 No              No
2         Don't know              No
3                Yes              No
4         Don't know              No
..               ...             ...
81        Don't know              No
82        Don't know              No
83        Don't know              No
84        Don't know              No
85               Yes              No

                                        comments
0                Relatively new job. Ask again later
1   Sometimes I think  about using drugs for my me...
2   I selected my current employer based on its po...
3   Our health plan has covered my psychotherapy a...
4   I just started a new job last week hence a lot...
..                                                ...
81  I'm self-employed on contract with small start...
82  My mental health issues were the direct result...
```

Cleaning the dataset using dropna method and saved the dataset in 'cleaned_data.csv'

```
84  Just starting a new job hence the numerous I d...
85                                Bipolar disorder

[86 rows x 27 columns]>
```

[10]:
```
df.tail()
```

[10]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequence	ph
1254	2015-09-12 11:17:21	26	male	United Kingdom	NaN	No	No	Yes	NaN	26-100	...	Somewhat easy	No	
1255	2015-09-26 01:07:35	32	Male	United States	IL	No	Yes	Yes	Often	26-100	...	Somewhat difficult	No	
1256	2015-11-07 12:36:58	34	male	United States	CA	No	Yes	Yes	Sometimes	More than 1000	...	Somewhat difficult	Yes	
1257	2015-11-30 21:25:06	46	f	United States	NC	No	No	No	NaN	100-500	...	Don't know	Yes	
1258	2016-02-01 23:04:31	25	Male	United States	IL	No	Yes	Yes	Sometimes	26-100	...	Don't know	Maybe	

5 rows × 27 columns

[11]:
```
df=df.dropna()
```

[12]:
```
df=df.drop_duplicates()
```

[13]:
```
df.to_csv('cleaned_data.csv',index=False)
```

[14]:
```
df=pd.read_csv('cleaned_data.csv')
```

[15]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

Importing the python libraries like (pandas, numpy, matplotlib)

[5]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

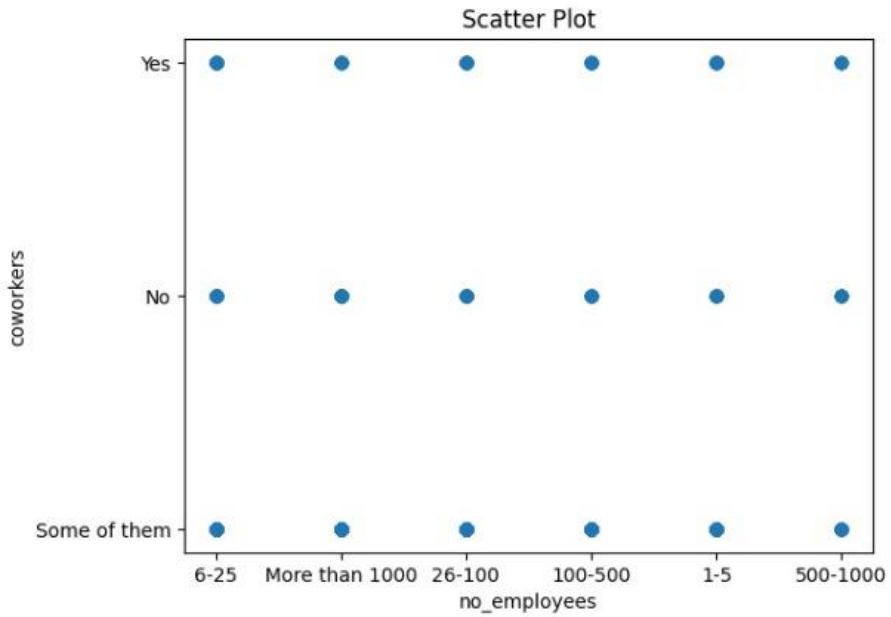
[6]:
```
df=pd.read_csv("survey.csv")
```

[10]:
```
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)
```

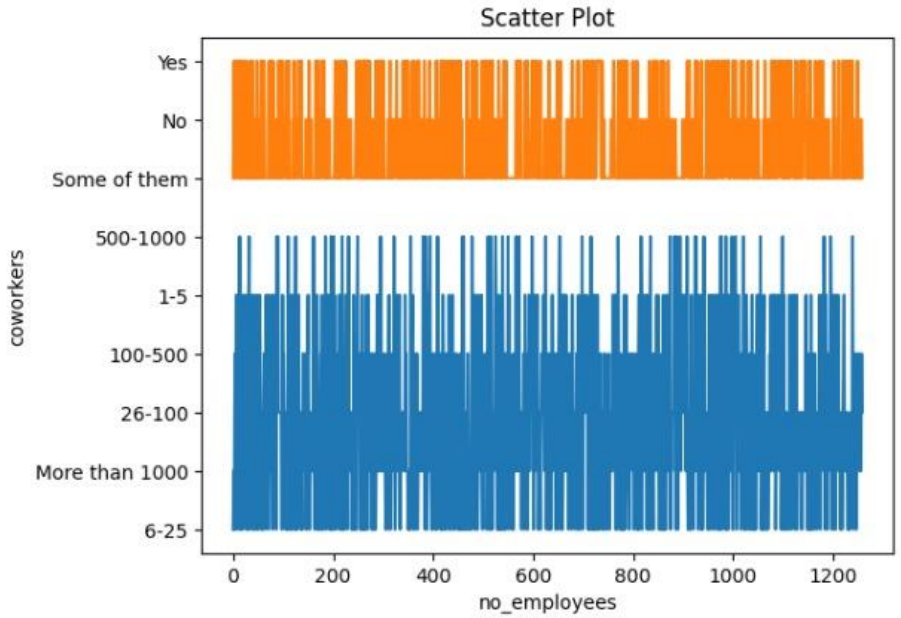
Using matplotlib library to visualize the dataset in scatter plot type:

[17]:
```
import pandas as pd
import matplotlib.pyplot as plt


# reading the database
data = pd.read_csv("survey.csv")

# Scatter plot with day against tip
plt.scatter(data['no_employees'], data['coworkers'])

# Adding Title to the Plot
plt.title("Scatter Plot")

# Setting the X and Y labels
plt.xlabel('no_employees')
plt.ylabel('coworkers')

plt.show()
```



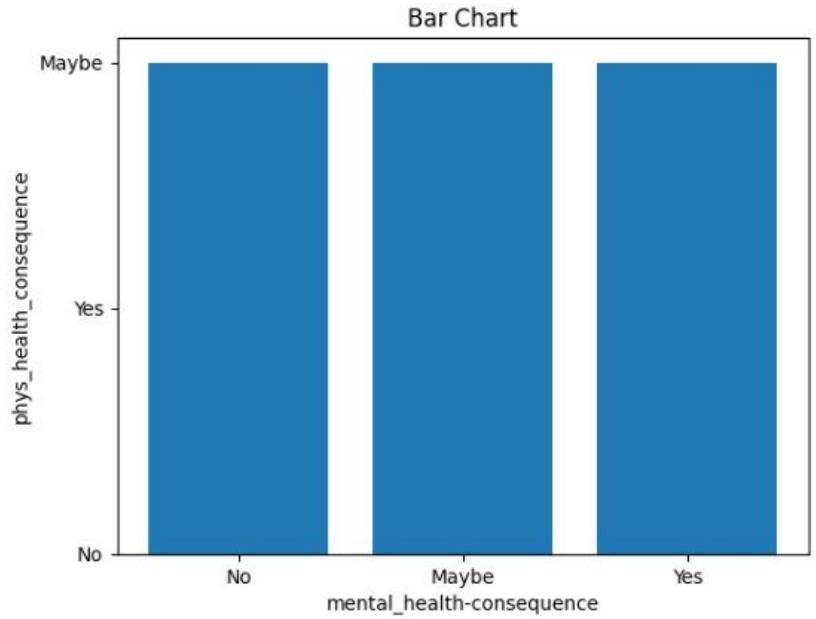
Analysis the dataset plotting using matplotlib.pyplot - library

[20]:
```
import pandas as pd
import matplotlib.pyplot as plt


# reading the database
data = pd.read_csv("survey.csv")

# Scatter plot with day against tip
plt.plot(data['no_employees'])
plt.plot(data['coworkers'])

# Adding Title to the Plot
plt.title("Scatter Plot")

# Setting the X and Y labels
plt.xlabel('no_employees')
plt.ylabel('coworkers')

plt.show()
```



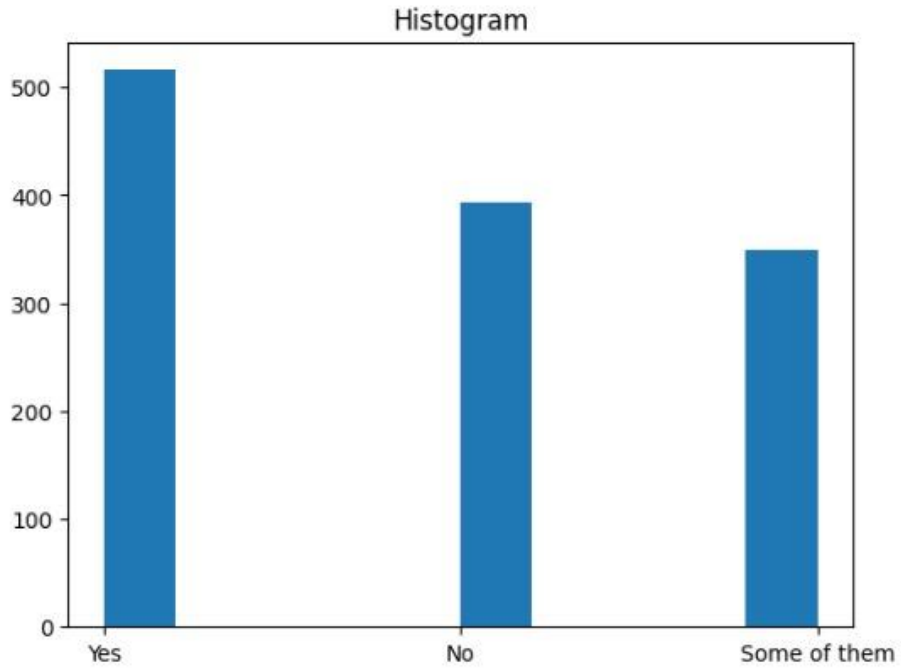
Taken the two columns given datset to visualize the data's in bar plot:

[21]:
```
import pandas as pd
import matplotlib.pyplot as plt


# reading the database
data = pd.read_csv("survey.csv")

# Bar chart with day against tip
plt.bar(data['mental_health_consequence'], data['phys_health_consequence'])

plt.title("Bar Chart")

# Setting the X and Y labels
plt.xlabel('mental_health-consequence')
plt.ylabel('phys_health_consequence')

# Adding the legends
plt.show()
```



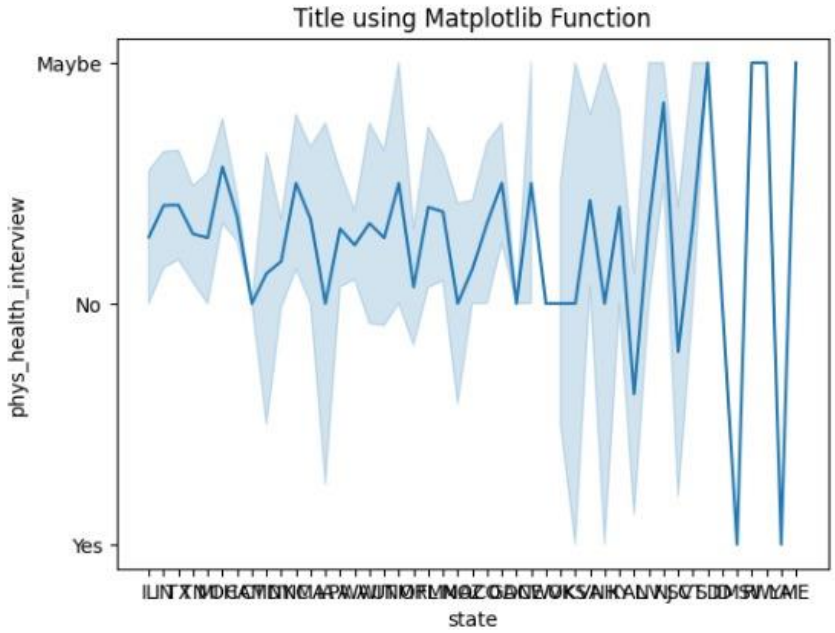
Visualize the dataset in histrogram :

[22]:
```
import pandas as pd
import matplotlib.pyplot as plt


# reading the database
data = pd.read_csv("survey.csv")

# histogram of total_bills
plt.hist(data['supervisor'])

plt.title("Histogram")

# Adding the legends
plt.show()
```



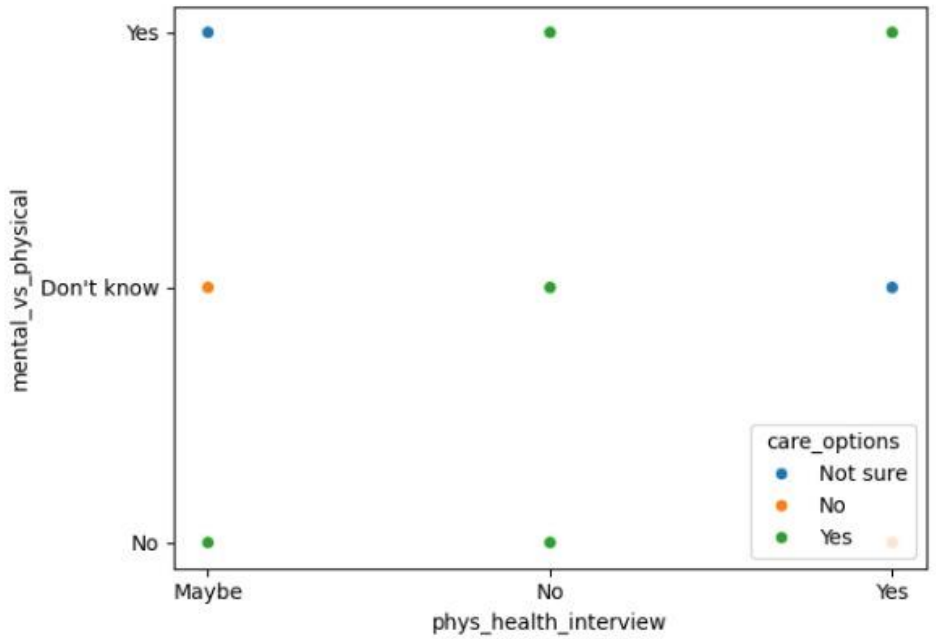
Importing the seaborn to analysis and visualize the graph in given dataset :

[26]:
```
# importing packages
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd


# reading the database
data = pd.read_csv("survey.csv")

# draw lineplot
sns.lineplot(x="state", y="phys_health_interview", data=data)

# setting the title using Matplotlib
plt.title('Title using Matplotlib Function')

plt.show()
```



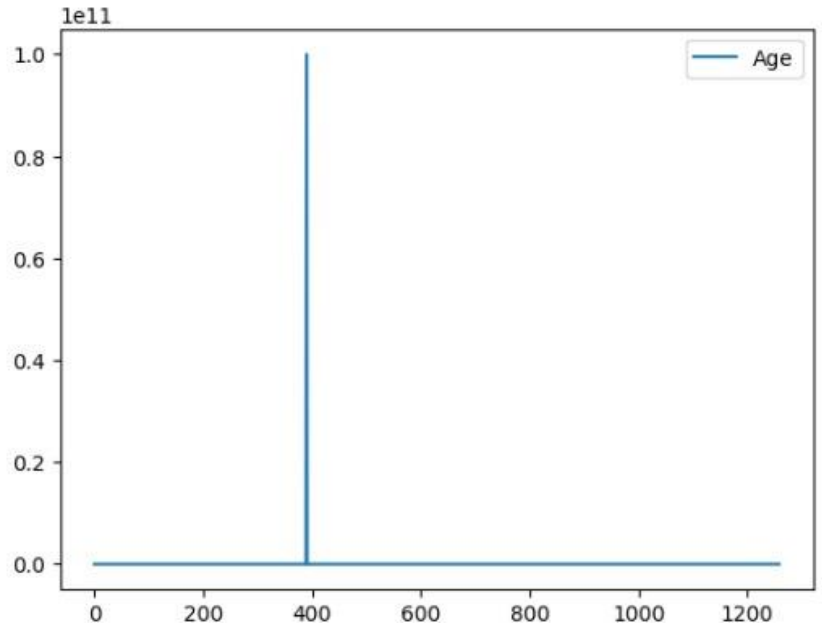
Visualize the data's in scatter plot using sns(seaborn plotitng):

[33]:
```
# importing packages
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

# reading the database
data = pd.read_csv("survey.csv")

sns.scatterplot(x='phys_health_interview', y='mental_vs_physical', data=data,
            hue='care_options')
plt.show()
```



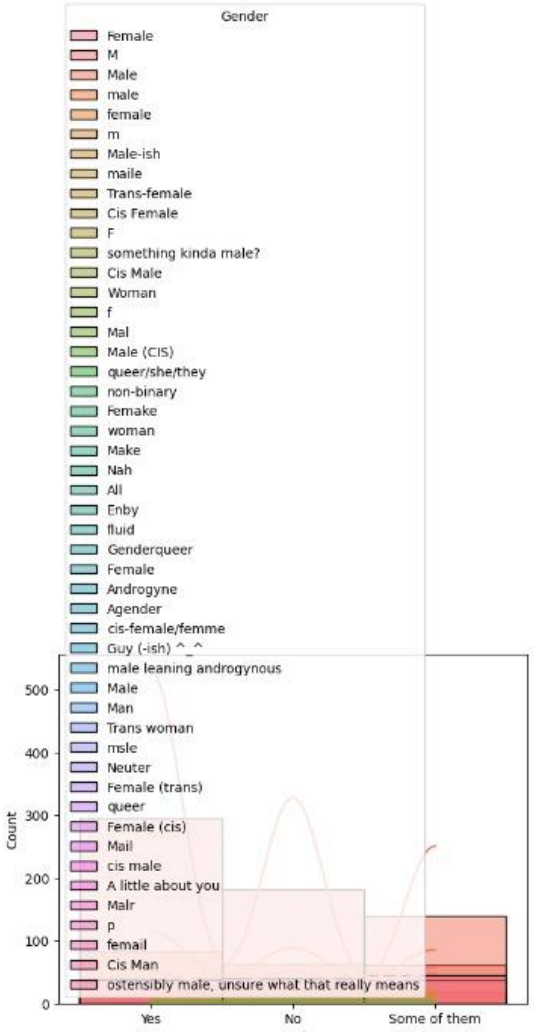
Visualize the dataset in Line Plot using seaborn library:

[36]:
```
# importing packages
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd


# reading the database
data = pd.read_csv("survey.csv")

# using only data attribute
sns.lineplot(data=data.drop(['benefits'], axis=1))
plt.show()
```



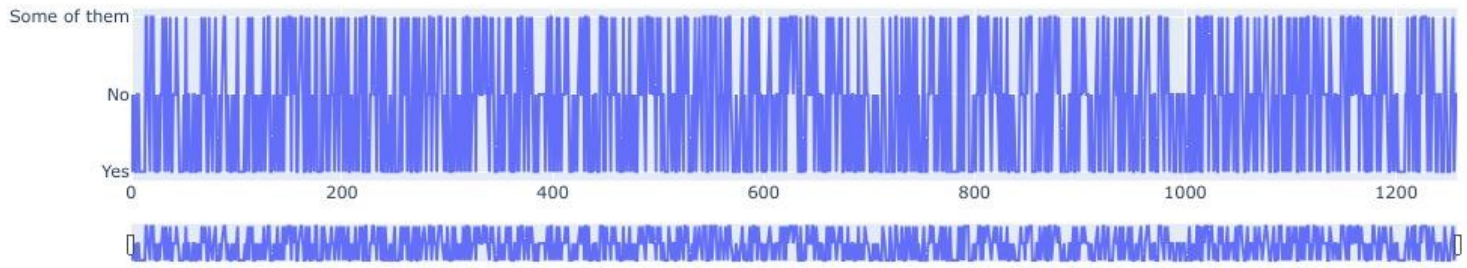
Taken the ‘supervisor’ column in given dataset to visualize the histplot :

```
[39]: # importing packages
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd


# reading the database
data = pd.read_csv("survey.csv")

sns.histplot(x='supervisor', data=data, kde=True, hue='Gender')

plt.show()
```



Import plotly .graph_objects to visualize the data:

[53]:
```
import plotly.graph_objects as px
import pandas as pd

# reading the database
data = pd.read_csv("survey.csv")

plot = px.Figure(data=[px.Scatter(
    y=data['supervisor'],
    mode='lines',)
])

plot.update_layout(
    xaxis=dict(
        rangeselector=dict(
            buttons=list([
                dict(count=1,
                    step="day",
                    stepmode="backward"),
            ])
        ),
        rangeslider=dict(
            visible=True
        ),
    )
)

plot.show()
```



Thank you…!