DATA ANALYTICS

Public health awareness campaign analysis

Team members:

1) K.Harini - harinikannan1502@gmail.com

2) V. karpaga Idaya - keccse21053@kingsedu.ac.in

3) JP.Jeshma - keccse21051@kingsedu.ac.in

4) U.Lavanya - keccse21062@kingsedu.ac.in

Problem statement:

"Analyze data from public health awareness compaigns to measure their effectiveness in reaching the target audience and increasing awareness"

MOTIVATION:

At the beginning of every year, I pledge to contribute to 3 social issues that I deeply care about to give back to the community. In January 2020, I committed to spend time on learning about and sensitizing my network of people about **Mental Health and Emotional Wellbeing**.

People entering or already working in the tech industry are leading increasingly stressful lives and working in highly competitive environments. With this lifestyle, self-care and mental well-being often takes a backseat until it becomes a major health issue.

Hence, I strongly feel mental health issues are a growing concern in our community and spreading awareness this topic is a pressing need in the current times. This research will help me in supplementing my posts on social media forums with scientific research and data-driven insights.

DATA SOURCE:

*Are you s	How many	Is your em	Is your prin	Does your	Do you kn	Has your e
FALSE	26-100	TRUE	TRUE	I don't kno	No	Yes
FALSE	26-100	TRUE	TRUE	Yes	No	No
FALSE	26-100	TRUE	TRUE	I don't kno	No	No
FALSE	100-500	TRUE	TRUE	I don't kno	No	Yes
FALSE	26-100	TRUE	TRUE	I don't kno	No	I don't kno
FALSE	100-500	TRUE	FALSE	Yes	No	No
FALSE	26-100	TRUE	TRUE	Yes	Yes	No
TRUE						
FALSE	26-100	TRUE	TRUE	Yes	No	Yes

METHODOLOGY & FINDINGS:

Each of the research questions listed above is addressed using descriptive graphical analysis in this section below and the insights are documented below the respective visualizations.

Bar charts, pivot tables, funnel visualization, grouped column charts and tabular summaries are some of the tools used for making comparisons and drawing inferences. Each graph aggregates the survey responses to summarize the characteristics of the data that are most suited to answer the research questions.

IMPORTING DEPENDENCIES:

The following libraries are required to run this Jupyter Notebook successfully.

Code:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
```

DATA QUALITY ASSESSMENT:

This section contains preliminary analysis to understand data structure, survey questions, responses and overall quality of the data.

Code:

```
survey_df = pd.read_csv('../1-
Data/OSMI_2019_MentalHealth_in_Tech_Survey_Results.csv')
total_participants = survey_df.shape[0]
print("Total Participants in 2019: ", total_participants)
display(survey_df.head(3))
```

1) Support for Mental Health in Tech

- **84%** of the participants have responded with a **score between 1 to 3** to indicate inadequate support
- **40%** participants have voted for **score 3** indicating moderate support
- Only **16%** of the participants have responded with a **score between 4 and 5** to indicate good support

Implications

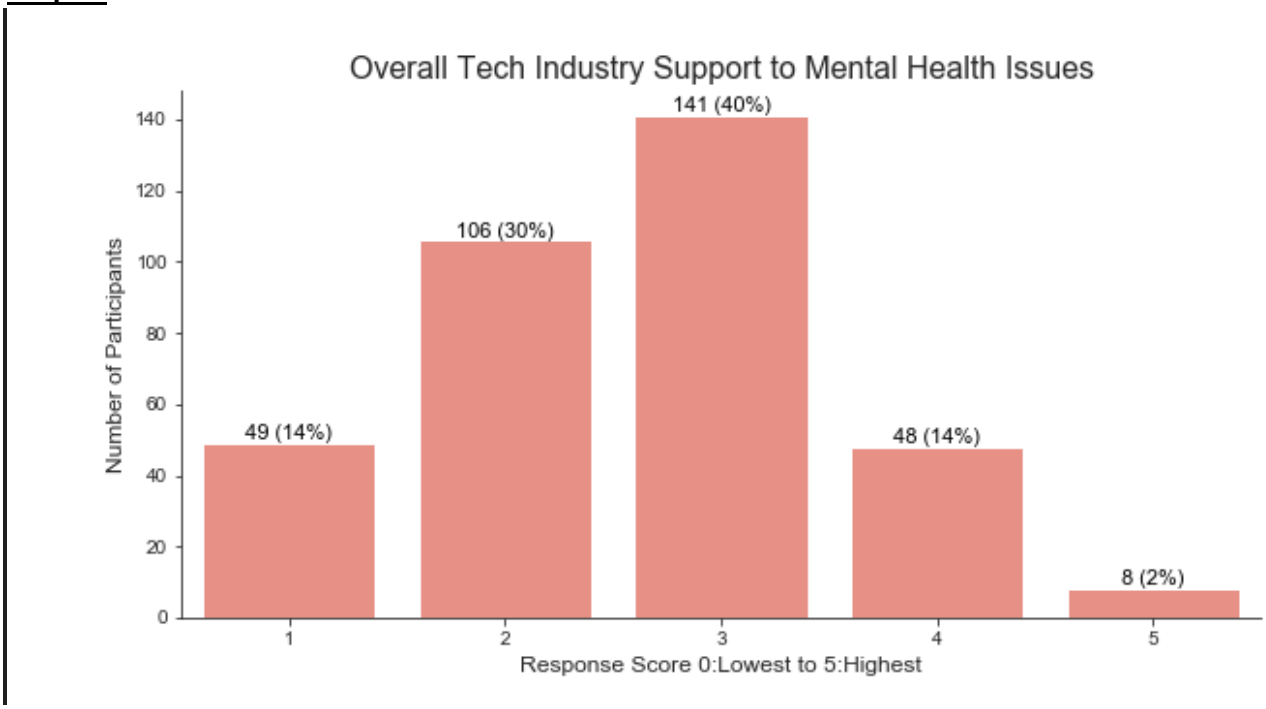
Overall perception of support for mental health in tech is **not adequate**

Code:

```
industry_support_mh = pd.DataFrame(survey_df["Overall, how well do you think the tech
industry supports employees with mental health issues?"])
industry_support_mh = industry_support_mh.rename(columns=
                                {"Overall, how well do you think the tech industry supports
employees with mental health issues?":
                                 "Tech Industry Support to Mental Health Issues"})
plt.figure(figsize=(10,5))
```

```
sns.set_style("ticks")
ax1 = sns.countplot(x="Tech Industry Support to Mental Health Issues",
data=industry_support_mh,
                color = 'salmon', saturation = 0.7)
sns.despine()

plt.title('Overall Tech Industry Support to Mental Health Issues', size = 16)
plt.xlabel('Response Score 0:Lowest to 5:Highest', size = 12)
plt.ylabel('Number of Participants', size = 12)

total = float(len(industry_support_mh))
for p in ax1.patches:
   height = p.get_height()
   ax1.text(p.get_x()+p.get_width()/2.,
         height + 1,
         '{:1.0f}'.format(height) + ' (' + '{:1.0f}'.format(100*(height/total)) + '%' + ')',
         ha="center", size=11, color = 'black')
plt.savefig("../3-Outputs/1-Support_Tech_Mental_Health.png")

plt.show()
```

output:



2)Mental Health Issues Experienced

- **42%** respondents (~147 / 352) **have experienced** a mental disorder in the past/at present
- **30%** participants (~107 / 352) have **never experienced** any mental illness

- **29%** participants (~100 / 352) are **unsure** if they have ever suffered from a mental disorder

Implications

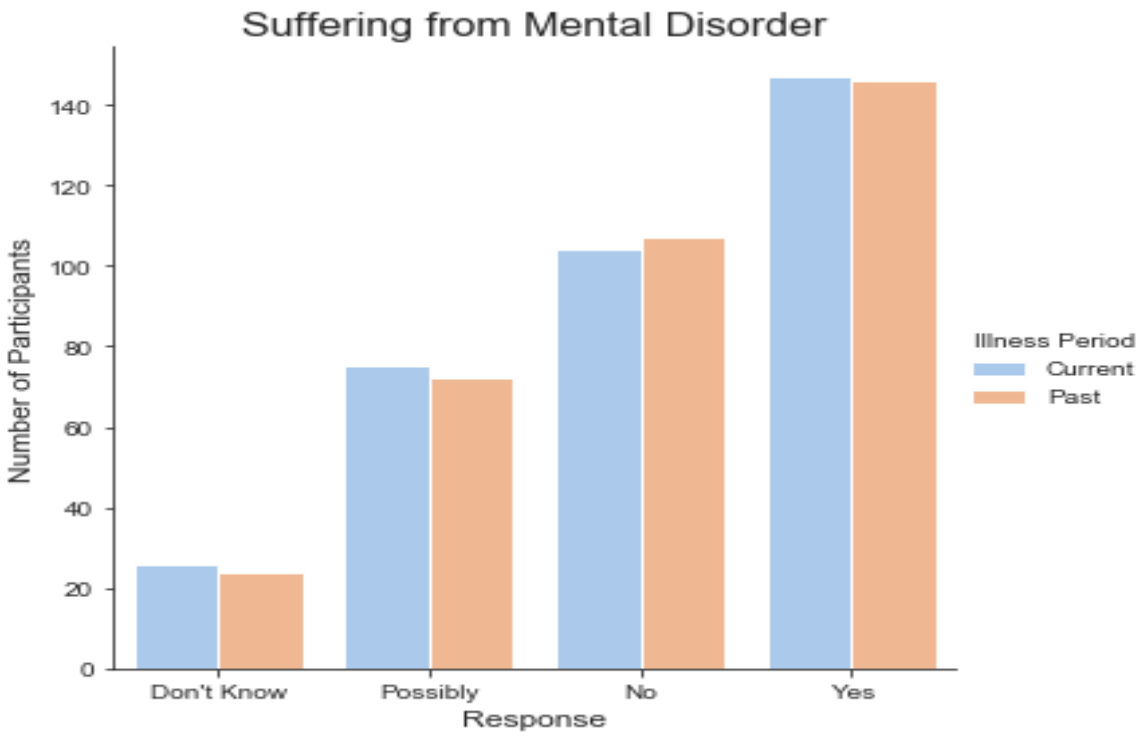
More participation from employees who have faced a mental health issue/unsure could lead to **bias in data.**

Only 30% respondents confidently claim no mental health issues - **alarmingly high number of affected employees.**

Code:

```
suffered = pd.DataFrame(survey_df.filter(items=[ 'Do you *currently* have a mental health disorder?',
                                                 'Have you had a mental health disorder in the past?' ]))

suffered = suffered.rename(columns= {'Do you *currently* have a mental health disorder?': 'Currently',
                                     'Have you had a mental health disorder in the past?': 'Past'})

suffered_current = pd.DataFrame(suffered[['Currently']])
suffered_current['Illness Period'] = 'Current'
suffered_current = suffered_current.rename(columns= {'Currently': 'Response'})

suffered_past = pd.DataFrame(suffered[['Past']])
suffered_past['Illness Period'] = 'Past'
suffered_past = suffered_past.rename(columns= {'Past': 'Response'})

suffered_all = suffered_current.append(suffered_past)
suffered_all['Participants'] = 1

suffered_table = pd.pivot_table(suffered_all, values='Participants', index=['Response'],
                columns=['Illness Period'], aggfunc=np.sum, fill_value=0)
display(suffered_table)

g = sns.catplot(x="Response", hue="Illness Period", data=suffered_all, kind="count", palette = "pastel");
sns.set_style("ticks")
sns.despine()

plt.title('Suffering from Mental Disorder', size = 16)
plt.xlabel('Response', size = 12)
plt.ylabel('Number of Participants', size = 12)

plt.savefig("../3-Outputs/2-Experienced_MentalHealthIssues.png")

plt.show()
```

output:



3)Medical Healthcare & Help Resources

- **Decreasing respondents** at stages coverage **(47%),** awareness **(37%),** discussions **(30%)** and help resources **(29%)**
- **Less than 50%** of the participants have **healthcare coverage f**or mental health issues and **merely 30%** end up getting additional help and resources
- The code below creates a table, funnel visualization and a bar chart to compare and measure the positive response of participants for the following 4 stages:

- **Coverage**
- **Awareness**
- **Discussions**
- **Help Resources**

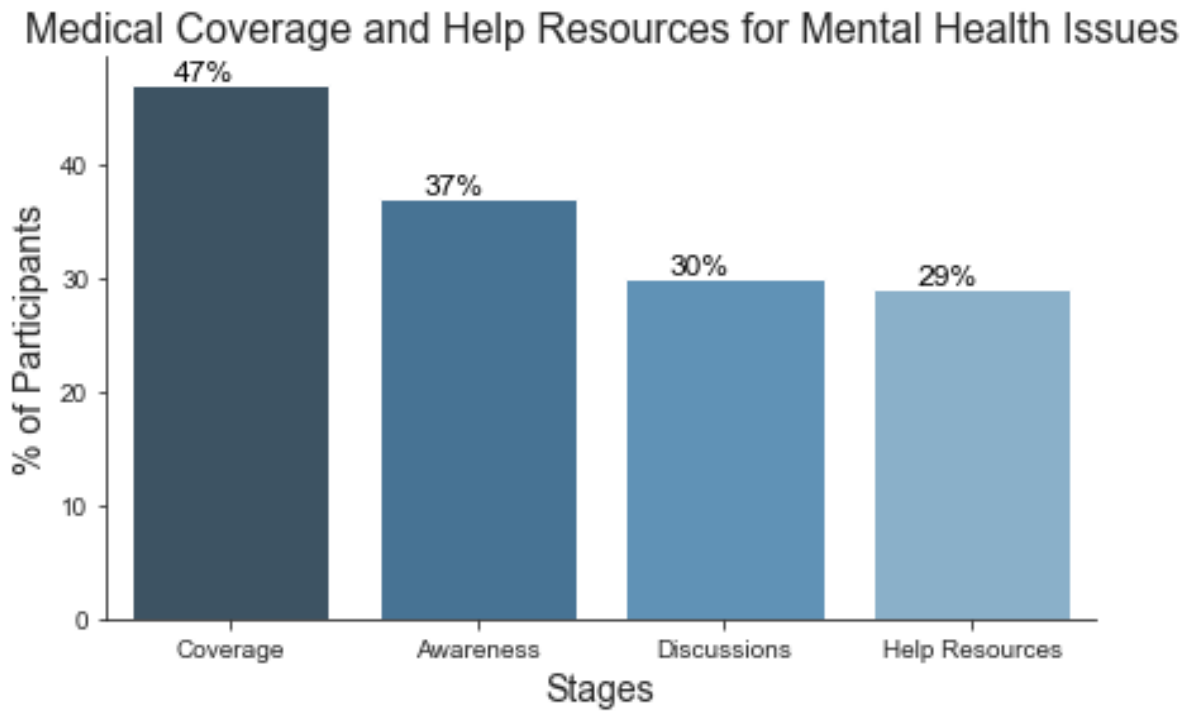
Code:

```
coverage = pd.DataFrame(survey_df["Does your employer provide mental health benefits as part of healthcare coverage?"])
coverage['Participants'] = 1
coverage = coverage.rename(columns= {"Does your employer provide mental health benefits as part of healthcare coverage?":
                                     "Coverage"})
awareness = pd.DataFrame(survey_df["Do you know the options for mental health care available under your employer-provided health coverage?"])
```

```
awareness['Participants'] = 1
awareness = awareness.rename(columns=
                             {"Do you know the options for mental health care available under your
employer-provided health coverage?":
                              "Awareness"})

discussions = pd.DataFrame(survey_df["Has your employer ever formally discussed mental
health (for example, as part of a wellness campaign or other official communication)?"])
discussions['Participants'] = 1
discussions = discussions.rename(columns=
                             {"Has your employer ever formally discussed mental health (for
example, as part of a wellness campaign or other official communication)?":

resources = pd.DataFrame(survey_df["Does your employer offer resources to learn more
about mental health disorders and options for seeking help?"])
resources['Participants'] = 1
resources = resources.rename(columns=
                             {"Does your employer offer resources to learn more about mental
health disorders and options for seeking help?":
                              "Resources"})

coverage = coverage.groupby(['Coverage']).Participants.agg('sum').to_frame('Participants
Count').reset_index()
awareness =
awareness.groupby(['Awareness']).Participants.agg('sum').to_frame('Participants
Count').reset_index()
discussions =
discussions.groupby(['Discussions']).Participants.agg('sum').to_frame('Participants
Count').reset_index()
resources = resources.groupby(['Resources']).Participants.agg('sum').to_frame('Participants
Count').reset_index()

coverage = coverage[coverage['Coverage'] == 'Yes']
coverage = coverage.rename(columns= {"Coverage" : "Response"})
coverage['Stage'] = "Coverage"

awareness = awareness[awareness['Awareness'] == 'Yes']
awareness = awareness.rename(columns= {"Awareness" : "Response"})
awareness['Stage'] = "Awareness"

discussions = discussions[discussions['Discussions'] == 'Yes']
discussions = discussions.rename(columns= {"Discussions" : "Response"})
discussions['Stage'] = "Discussions"

resources = resources[resources['Resources'] == 'Yes']
resources = resources.rename(columns= {"Resources" : "Response"})
resources['Stage'] = "Help Resources"
```

```
all_stages = coverage.append(awareness)
all_stages = all_stages.append(discussions)
all_stages = all_stages.append(resources)
all_stages['Total Participants'] = survey_df.shape[0]
all_stages['% Participants'] = round(100 * all_stages['Participants Count']/all_stages['Total
Participants'],0)

all_stages = all_stages[['Stage', 'Response', 'Participants Count', 'Total Participants', '%
Participants']]
display(all_stages)

data = dict(
   Participants=all_stages['Participants Count'],
   Stage=all_stages['Stage'])
fig = px.funnel(data, x='Participants', y='Stage', width=800, height=500)


fig.write_image("../3-Outputs/3-Medical_HelpResources_Mental_Health_Funnel.png")
fig.show()

plt.figure(figsize=(7,4))
graph = sns.barplot(x="Stage", y="% Participants", data=all_stages, palette="Blues_d",
saturation = 0.7)
sns.set_style("ticks")
sns.despine()

plt.title('Medical Coverage and Help Resources for Mental Health Issues', size = 16)
plt.xlabel('Stages', size = 14)
plt.ylabel('% of Participants', size = 14)

for p in graph.patches:
      graph.annotate('{:.0f}'.format(p.get_height()) + '%', (p.get_x()+0.3, p.get_height()),
            ha='center', va='bottom',
            color= 'black', size = 12)

plt.savefig("../3-Outputs/3-Medical_HelpResources_Mental_Health.png")
 plt.show()
```

Output:



4)Importance for Physical & Mental Health

- Importance given by employers to **physical health** has **higher distribution** between scores **5 to 10** with the **peak at 5.**
- Importance given by employers to **mental health** has **higher distribution** between scores **3 to 7** with the **peak at 5.**

Implications

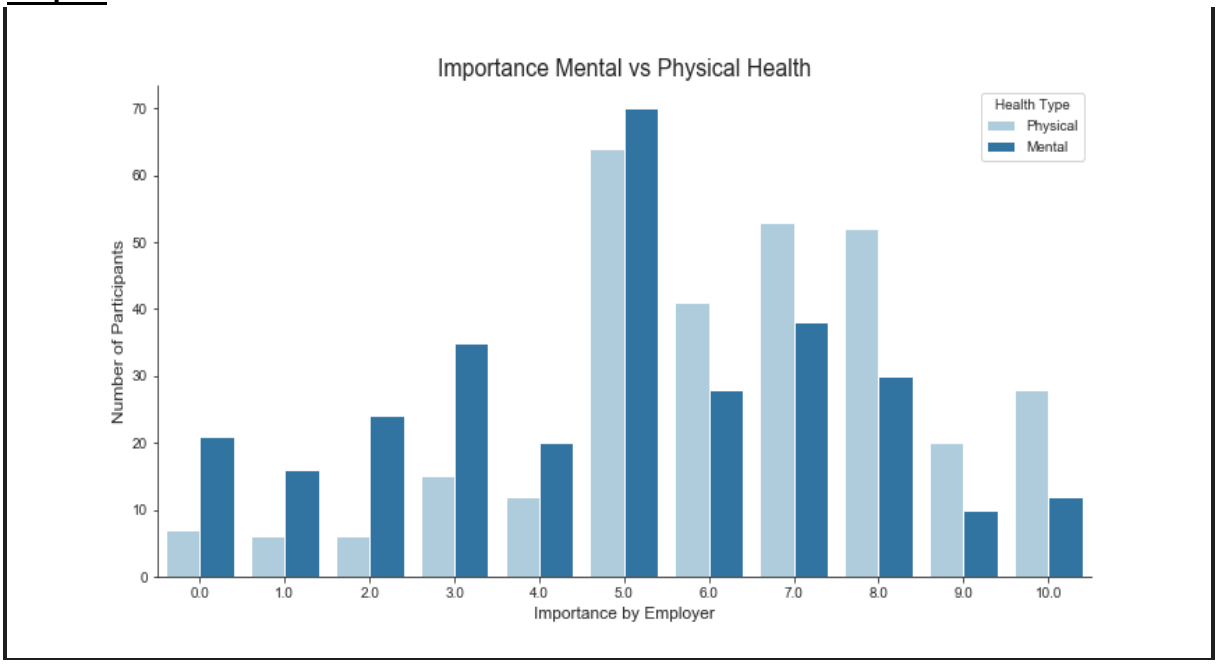
Employers **do not give enough importance** to mental health compared to physical health

Code:

```
physical_importance = pd.DataFrame(survey_df["Overall, how much importance does your employer place on physical health?"])
physical_importance['Participants'] = 1
physical_importance['Health Type'] = 'Physical'
physical_importance = physical_importance.rename(columns= {"Overall, how much importance does your employer place on physical health?":
                                "Importance Level"})

physical_importance = physical_importance.groupby(['Importance Level', 'Health Type']).Participants.agg('sum').to_frame('Participants Count').reset_index()

mental_importance = pd.DataFrame(survey_df["Overall, how much importance does your employer place on mental health?"])
mental_importance['Participants'] = 1
mental_importance['Health Type'] = 'Mental'
```

```
mental_importance = mental_importance.rename(columns= {"Overall, how much
importance does your employer place on mental health?":
                                "Importance Level"})

mental_importance = mental_importance.groupby(['Importance Level', 'Health
Type']).Participants.agg('sum').to_frame('Participants Count').reset_index()

importance = physical_importance.append(mental_importance)
compare_importance = pd.merge(physical_importance, mental_importance, how='outer',
on=['Importance Level'])
compare_importance = compare_importance.rename(columns= {'Participants Count_x':
'Physical Health', 'Participants Count_y': 'Mental Health'})
display(compare_importance[['Importance Level', 'Physical Health', 'Mental Health']])

plt.figure(figsize=(12,6))
sns.set_style("ticks")
ax = sns.barplot(x="Importance Level", y="Participants Count", hue="Health Type",
data=importance, palette = "Paired")
sns.despine()
plt.title('Importance Mental vs Physical Health', size = 16)
plt.xlabel('Importance by Employer', size = 12)
plt.ylabel('Number of Participants', size = 12)

plt.savefig("../3-Outputs/4-Physical_vs_Mental_Health.png")

plt.show()
```

output:



5)Leave Policy for Mental Health Issues

- **44%** respondents find it **relatively easy** to ask for leaves for mental health while **19%** respondents find it **relatively difficult**
- **30%** respondents are **neutral/unaware** of the difficulty they may pose while asking for time off.

Implications

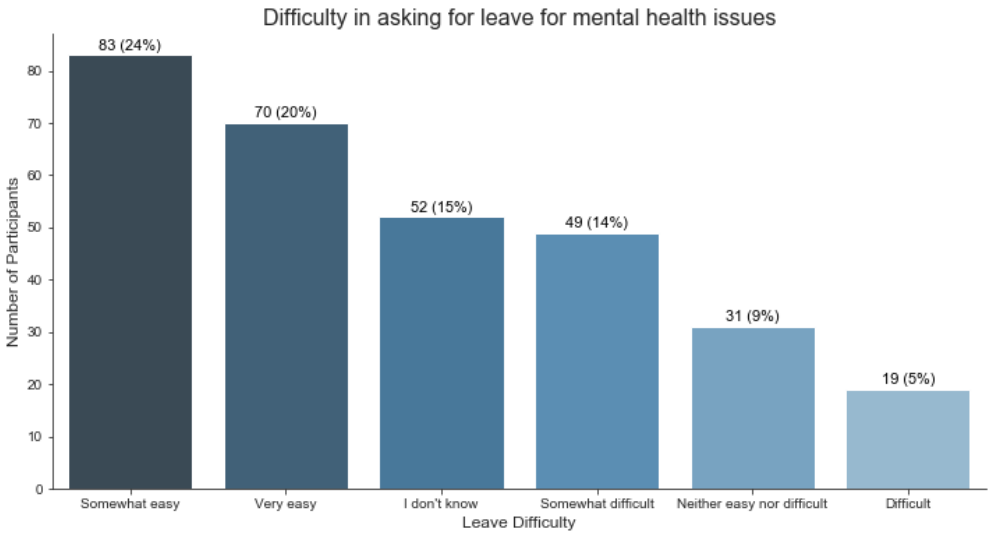
Leave policies around mental health issues **may be in favor of the employees** at many workplaces but employers may want to take **more effort** to improve and create awareness among employees.

Code:

```
leave = pd.DataFrame(survey_df["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?"])
leave['Participants'] = 1
leave = leave.rename(columns= {"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?":
                               "Leave Difficulty"})

leave = leave.groupby(['Leave Difficulty']).Participants.agg('sum').to_frame('Participants Count').reset_index()
leave['% Participants'] = round(100 * leave['Participants Count']/total_participants,0)
display(leave)

plt.figure(figsize=(12,6))
sns.set_style("ticks")
ax1 = sns.countplot(x="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?",
                    data=survey_df,
                    palette="Blues_d", saturation = 0.7,
                    order = survey_df['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?'].value_counts().index)
sns.despine()

plt.title('Difficulty in asking for leave for mental health issues', size = 16)
plt.xlabel('Leave Difficulty', size = 12)
plt.ylabel('Number of Participants', size = 12)

total = total_participants
for p in ax1.patches:
   height = p.get_height()
   ax1.text(p.get_x()+p.get_width()/2.,
         height + 1,
         '{:1.0f}'.format(height) + ' (' + '{:1.0f}'.format(100*(height/total)) + '%' + ')',
         ha="center", size=11, color = 'black')

plt.savefig("../3-Outputs/5-LeavePolicy_Mental_Health.png")
```

plt.show()

output:



CONCLUSION

The above exploration validates the initial hypothesis of growing concerns regarding mental health issues in the tech industry. The findings and insights of this research are limited to the participants who responded to the survey and may not be a generalization of the entire tech industry and workplaces.

However, based on this ethnographic study, we observe that a majority of the participants (84%) feel that the support to mental health in workplaces in tech is not adequate. A large majority (70%) of employees have experienced mental disorders to some extent in the past or at present. While many participants shared that the importance to physical health by their employers was high, the relative importance to mental health has been much lower.

The leave policies seem to be in favor of employees, but employers may to take more efforts to make employees aware and comfortable of seeking time off for mental wellbeing. A large number of participants (50%) do not have medical healthcare coverage and benefits for mental health issues. A vast majority (70%) do not have access to open discussions and resources to learn more about mental health issues.