IBM PROJECT NAAN MUDHALVAN

PHASE 4: DEVELOPMENT PART 2

TOPIC: PUBLIC HEALTH AWARENESS CAMPAIGN ANALYSIS

INTRODUCTION :

In this phase , we have taken a leap forward in transforming our data into actionable information. During this phase, we have harnessed the capabilities of IBM Cognos and integrated advanced code for data analysis to elevate our project's analytical depth and visual impact. This part allows to understand how to handle data, explore it visually, and evaluate the performance of a machine learning model using various metrics. It serves as a practical example for data analysis and predictive modeling tasks.

STEP 1 : DATA SUMMARIZATION

```
desc_stats = df.describe()
print(desc_stats)
```

This step provides a condensed overview of numerical data in the project. It helps assess data quality, offers initial insights into data characteristics, supports decision-making, and facilitates clear communication of key findings..

STEP 2 : CREATE NEW COLUMN

```
df['Grouped_Gender'] = df['Gender'].apply(lambda x: 'Others' if x not in
['Male', 'Female', 'Non-binary'] else x)
```

This step creates a new column which can help in having more visually appealing and easier to understand charts and graphs..

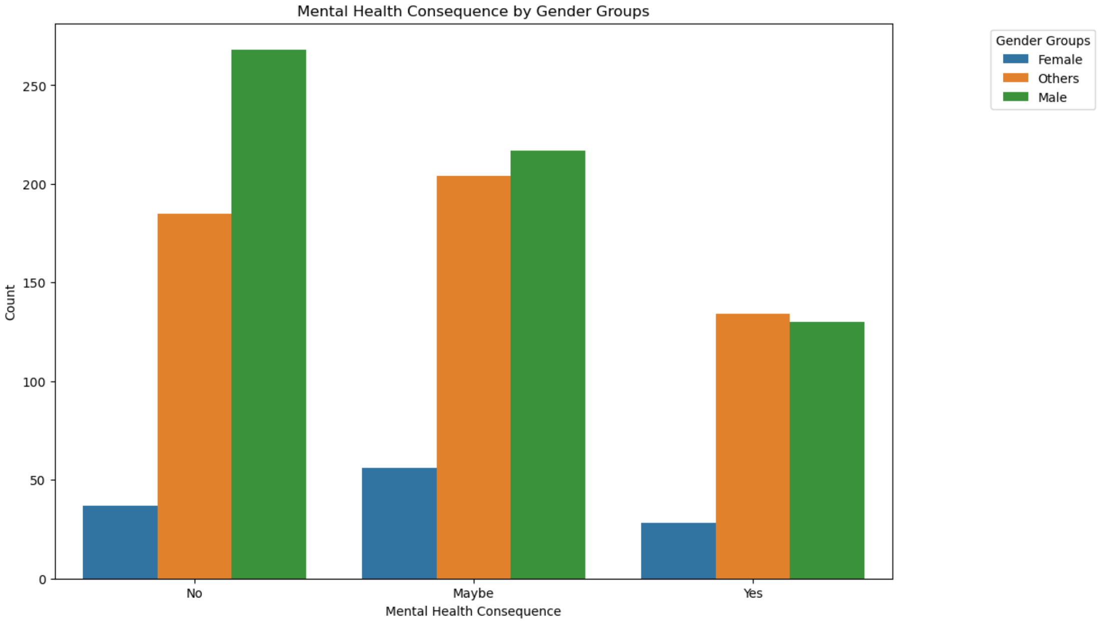
STEP 3 : DATA VISUALIZATION

This section includes various data visualization steps using matplotlib and seaborn for understanding the dataset.

Mental Health Consequence by Gender Groups

```
plt.figure(figsize=(12, 8))
sns.countplot(x='mental_health_consequence', hue='Grouped_Gender', data=df)
plt.title('Mental Health Consequence by Gender Groups')
plt.xlabel('Mental Health Consequence')
plt.ylabel('Count')
plt.legend(title='GenderGroups',loc='upperright', bbox_to_anchor=(1.25, 1))
plt.show()
```

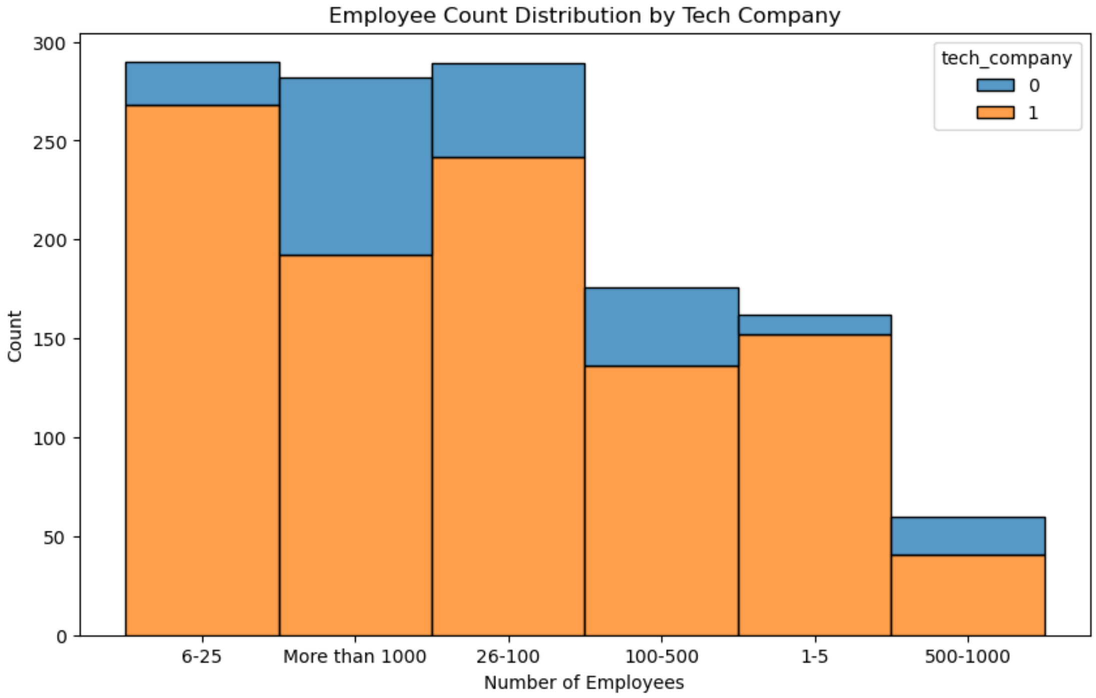
This step creates a bar graph to visualize the distribution of Mental Health Consequence by Gender Groups.



Employee Count Distribution by Tech Company

```
plt.figure(figsize=(10, 6))
sns.histplot(data=df, x='no_employees', hue='tech_company', multiple='stack')
plt.title('Employee Count Distribution by Tech Company')
plt.xlabel('Number of Employees')
plt.ylabel('Count')
plt.show()
```

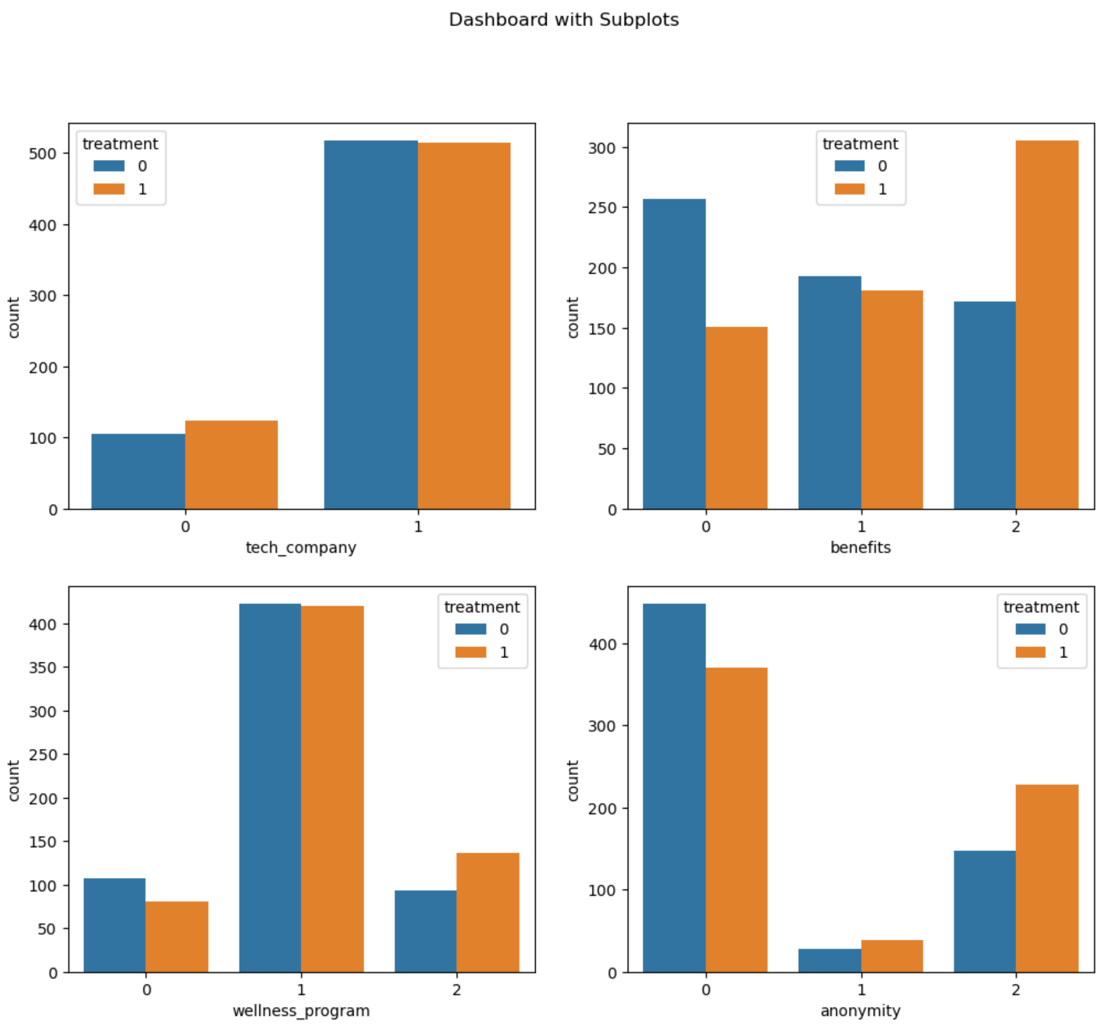
This step creates a histogram to show the Employee Count Distribution by Tech Company.



Dashboard with Subplots

```
fig, axs = plt.subplots(2, 2, figsize=(12, 10))
sns.countplot(x='tech_company', hue='treatment', data=df, ax=axs[0, 0])
sns.countplot(x='benefits', hue='treatment', data=df, ax=axs[0, 1])
sns.countplot(x='wellness_program', hue='treatment', data=df, ax=axs[1,
0])
sns.countplot(x='anonymity', hue='treatment', data=df, ax=axs[1, 1])
plt.suptitle('Dashboard with Subplots')
plt.show()
```

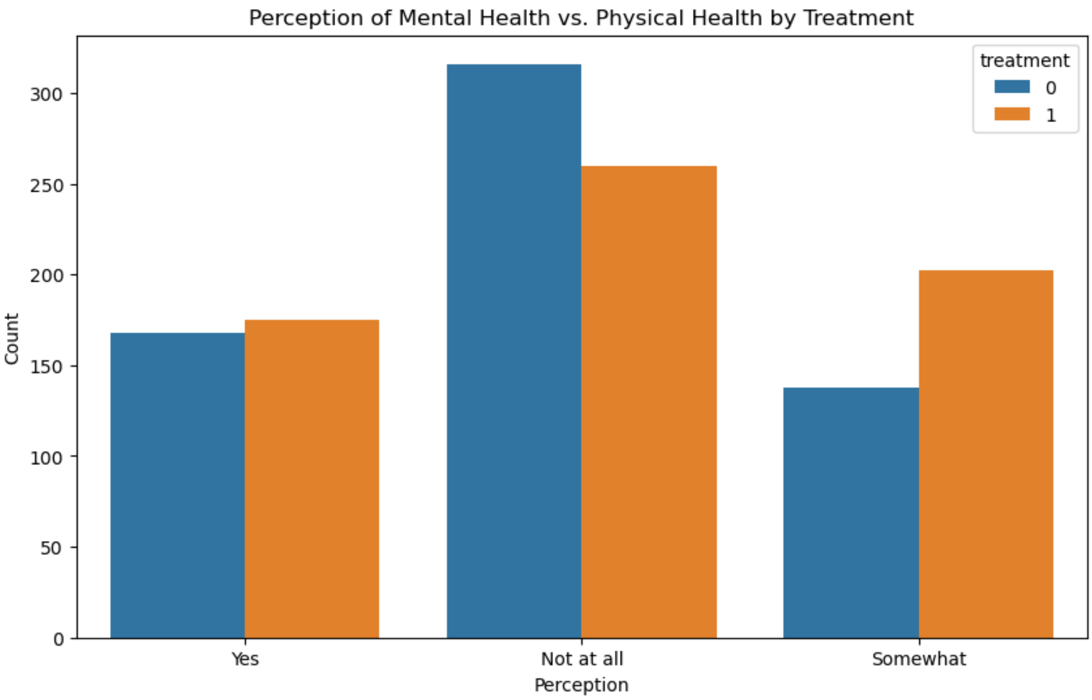
This step creates a Dashboard with Subplots of count plots for distribution of treatment based on tech company, benefits , wellness_program and anonymity.



Disparate Analysis between Mental Health and Perceived Physical Health

```
perception_labels = {0: 'Not at all', 1: 'Somewhat', 2: 'Yes'}
df['Mental_vs_Physical_Label']=df['mental_vs_physical']   .map(perception_labels)
plt.figure(figsize=(10, 6))
sns.countplot(x='Mental_vs_Physical_Label', data=df, hue='treatment')
plt.title('Perception of Mental Health vs. Physical Health by Treatment')
plt.xlabel('Perception')
plt.ylabel('Count')
plt.show()
```

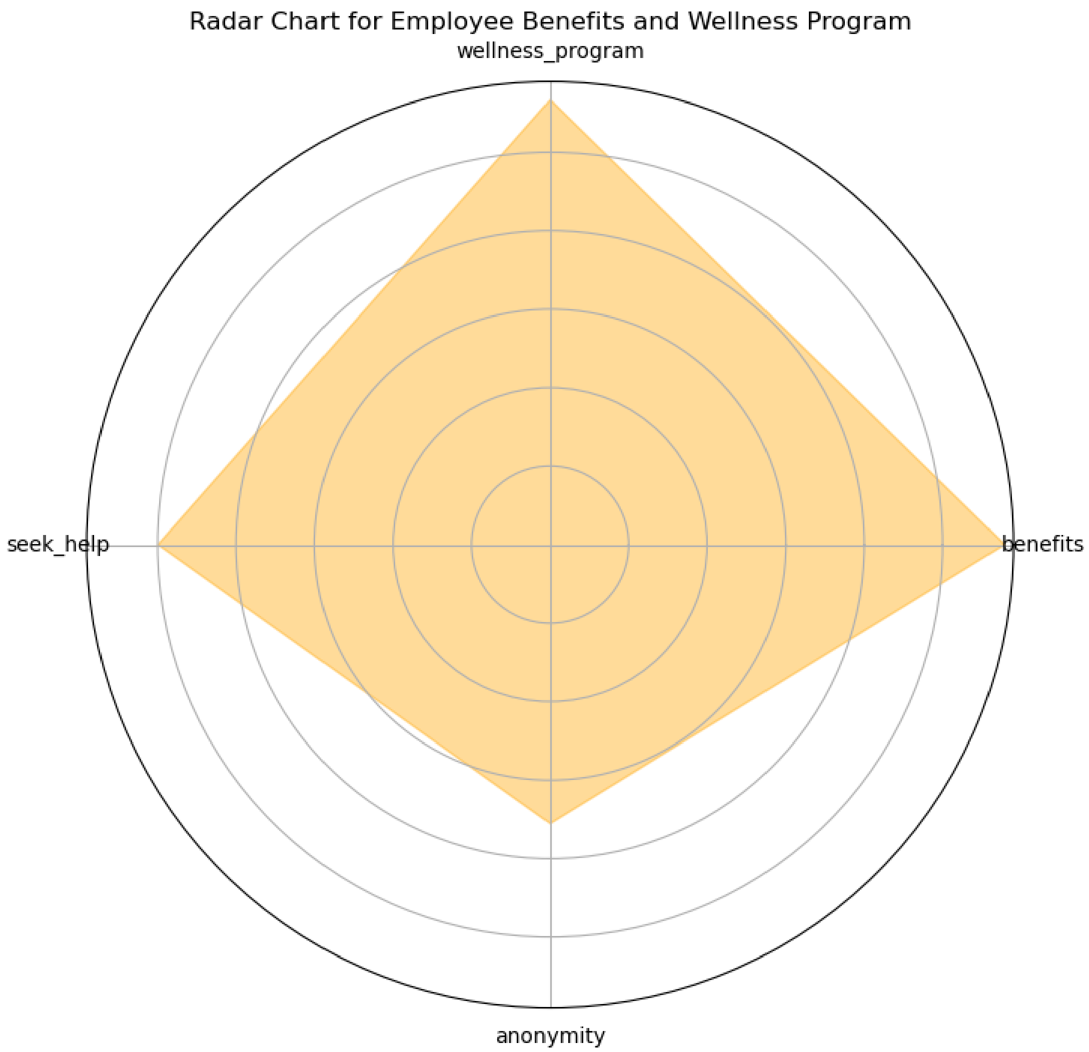
This step creates a bar graph to show the comparison between mental health and perceived physical health.



Employee Benefits and Wellness Program Evaluation by Dimension

```
benefits_cols = ['benefits', 'wellness_program', 'seek_help', 'anonymity']
benefits_values = df[benefits_cols].mean()
benefits_values = benefits_values / benefits_values.sum()
plt.figure(figsize=(8, 8))
angles = [n / len(benefits_cols) * 2 * np.pi for n in range(len(benefits_cols))]
ax = plt.subplot(111, polar=True)
ax.fill(angles, benefits_values, color='orange', alpha=0.4)
ax.set_yticklabels([])
ax.set_xticks(angles)
ax.set_xticklabels(benefits_cols)
plt.title('Radar Chart for Employee Benefits and Wellness Program')
plt.show()
```

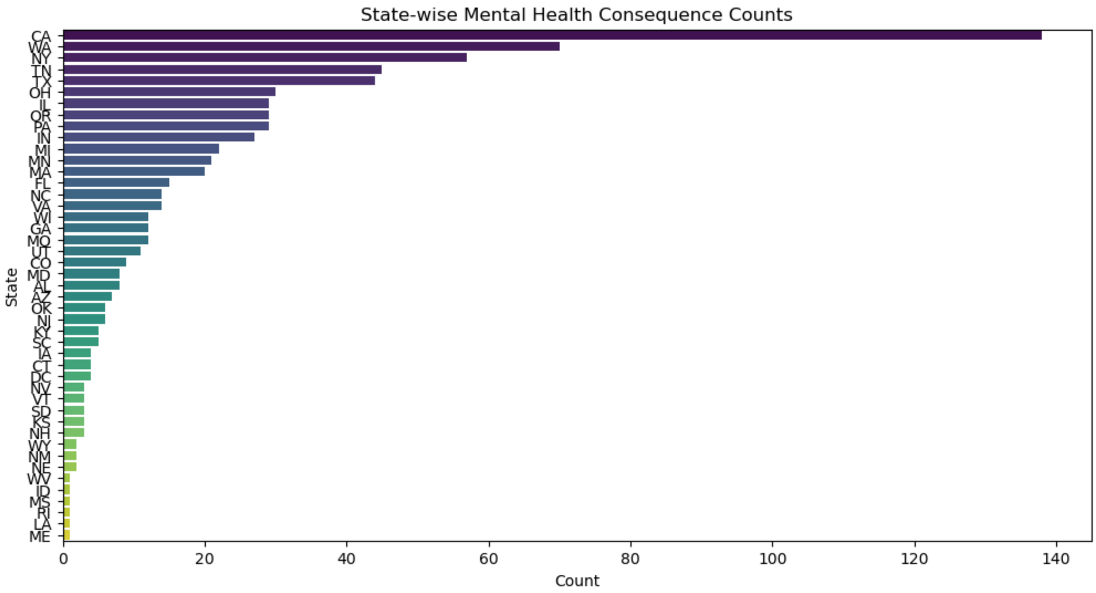
This step generates a radar chart to visualize the correlation between employee benefits and wellness program.



Distribution of State-wise Mental Health Consequence

```
state_counts = df['state'].value_counts()
plt.figure(figsize=(12, 6))
sns.barplot(x=state_counts, y=state_counts.index, palette='viridis')
plt.title('State-wise Mental Health Consequence Counts')
plt.xlabel('Count')
plt.ylabel('State')
plt.show()
```

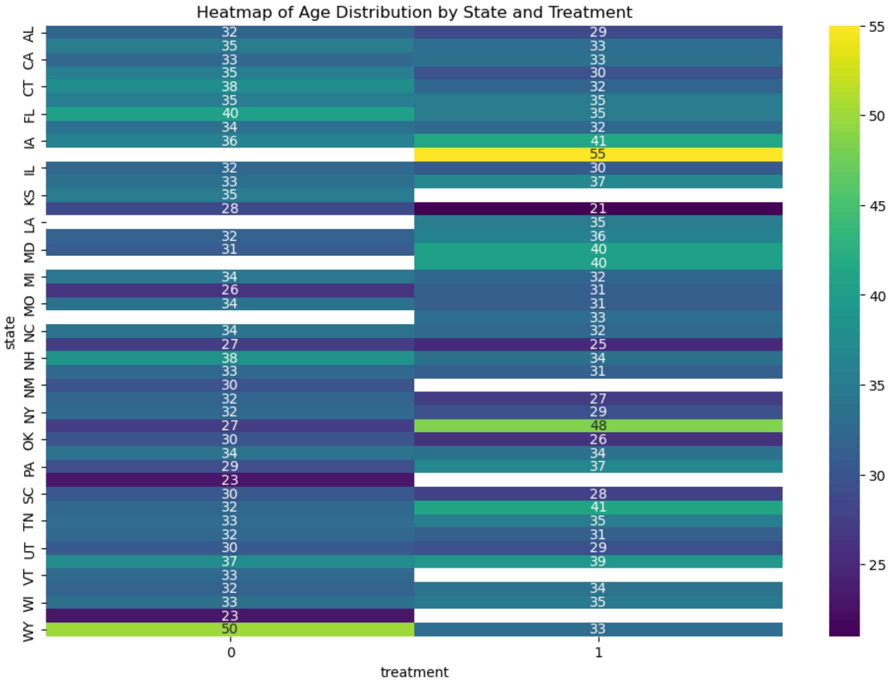
This step creates a Horizontal Bar Chart to visualize the State-wise Mental Health Consequence Counts.



Distribution of Age by State and Treatment

```
state_age_pivot = df.pivot_table(values='Age', index='state',
columns='treatment', aggfunc='mean')
plt.figure(figsize=(12, 8))
sns.heatmap(state_age_pivot, annot=True, cmap='viridis')
plt.title('Heatmap of Age Distribution by State and Treatment')
plt.show()
```

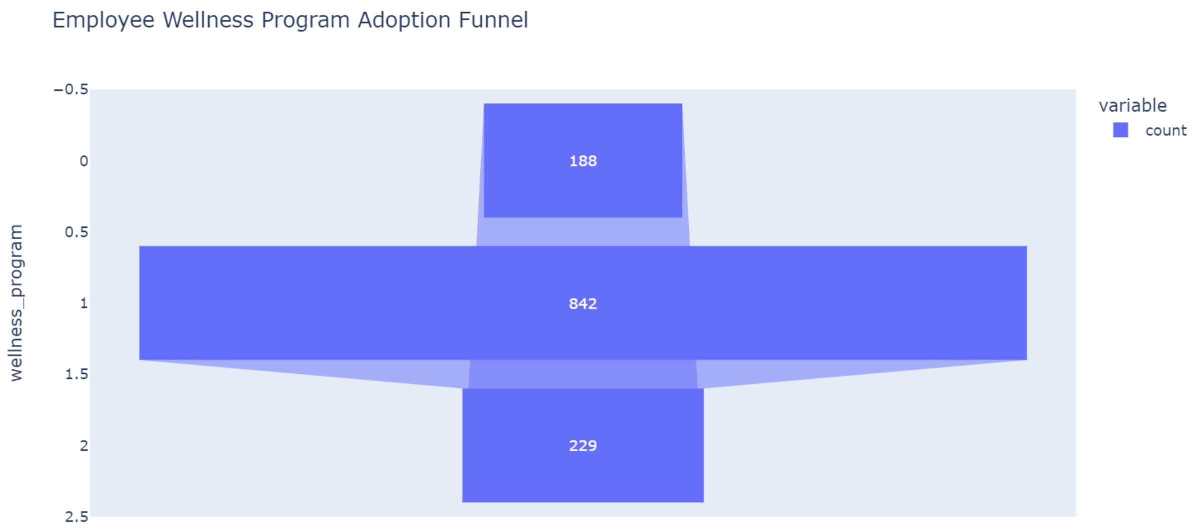
This step creates a heatmap to show the Age Distribution by State and Treatment.



Progress of Employee Wellness Program Enrollment

```
import plotly.express as px
wellness_program_counts = df['wellness_program'].value_counts()
fig = px.funnel(wellness_program_counts, title='Employee Wellness
Program Adoption Funnel', labels={'index': 'Adoption Status', 'value':
'Count'})
fig.show()
```

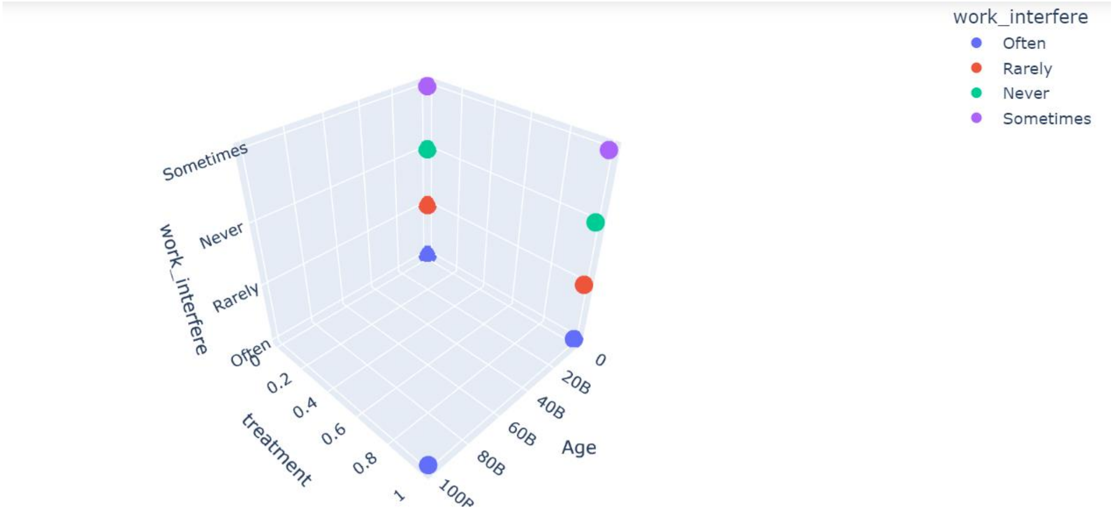
This step creates a Funnel Chart for Employee Wellness Program Adoption by importing plotly.express .



Distribution between Age, Treatment, and Work Interference using 3D

```
fig = px.scatter_3d(df, x='Age', y='treatment', z='work_interfere',
color='work_interfere')
fig.update_layout(title='3D Scatter Plot')
fig.show()
```

This step creates a 3D scatter plot to show the distribution between Age, treatment and work interference.



STEP 7 : PERFORMING STATISTICAL TEST

```
from scipy.stats import ttest_ind
t_stat, p_value = ttest_ind(df[df['treatment'] == 1]['Age'], df [df
['treatment'] == 0]['Age'])
print(f"T-Statistic: {t_stat}\nP-value: {p_value}")
```

```
OUTPUT:
T-Statistic: 0.9881466556871645
P-value: 0.32327099499937273
```

In this step, we perform an independent two-sample t-test. It calculates the t-statistic and p-value to assess whether there is a statistically significant difference in the ages of two groups: the treatment group and the control group, as represented in a DataFrame (df).

TEAM MEMBERS:

SHYAM KUMAR M	- 2021115107
SUMATHI K	- 2021115112
SIVA GANESH S	- 2021115108
SUBASHREE S R	- 2021115110
YUKESH S	- 2021115331