PHASE 5 PROJECT SUBMISSION

Phase 1: Problem Definition and Design Thinking

In this part you will need to understand the problem statement and create a document on what have you understood and how will you proceed ahead with solving the problem. Please think on a design and present in form of a document.

Project Definition: The project involves analyzing data from public health awareness campaigns to measure their effectiveness in reaching the target audience and increasing awareness. The objective is to provide insights that evaluate the impact of the campaigns and inform future strategies. This project includes defining analysis objectives, collecting campaign data, designing relevant visualizations in IBM Cognos, and using code for data analysis.

Design Thinking:

1. Analysis Objectives: Define specific objectives for analyzing public health awareness campaign data, such as measuring audience reach, awareness levels, and campaign impact.

2. Data Collection: Identify the sources and methods for collecting campaign data, including engagement metrics, audience demographics, and awareness surveys.

3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos to create informative dashboards and reports.

4. Code Integration: Decide which aspects of the analysis can be enhanced using code, such as data cleaning, transformation, and statistical analysis.

Dataset Link: https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey

Assignment Notebook Submission

File Naming Convention: DAC\_Phase1

After completion upload your file to your private GitHub account. Please give access to your faculty evaluators of your college and industry evaluator [ IndustryEvaluator@skillup.online ] to your private GitHub repository for evaluation process

Go to the Project Submission Part 1 section and add your college code, the link of your GitHub in the space provided, upload your documents, and click on submit.

**PUBLIC HEALTH AWARENESS AND CHAMPIGN ANALYSIS**

TEAM MEMBER

731721205005 : S.BHARATHI

Phase 2 submission document

**Project :** public health awareness



**Introduction :**

* Public health care is a fundamental aspect of any society's well-being, providing essential services to promote and protect the health of its citizens.
* It encompasses a wide range of activities, from disease prevention and health promotion to the provision of medical care and emergency response.
* This analysis aims to explore the landscape of public health care awareness campaigns, their significance, and their impact on individuals and communities.
* Public health care awareness campaigns play a crucial role in educating individuals about health issues, encouraging healthy behaviors, and mobilizing communities to take collective action to improve their overall health and well-being.

**Content for Project Phase 2 :**

Consider incorporating machine learning algorithms to predict the success of future campaigns based on historical data.

**Innovation:**

Using a dataset bringing design into innovation :

**Data link :**

<https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>

**code :**

# Import necessary libraries

import pandas as pd

import numpy as np

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

# Load the historical campaign data

data = pd.read\_csv("campaign\_data.csv")

# Feature engineering (for demonstration, you can add more relevant features)

# Let's assume 'campaign\_type', 'duration', and 'target\_audience' are relevant features

data['campaign\_type\_encoded'] = pd.factorize(data['campaign\_type'])[0]

data['target\_audience\_encoded'] = pd.factorize(data['target\_audience'])[0]

# Define features and target variable

X = data[['campaign\_type\_encoded', 'duration', 'target\_audience\_encoded']]

y = data['success\_metric']

# Split the dataset into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Create and train a Linear Regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Make predictions on the test set

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print(f"Mean Squared Error: {mse:.2f}")

print(f"R-squared Score: {r2:.2f}")

**Output :**

Model Evaluation

Accuracy: 0.75

Confusion Matrix:

[[45 10]

 [12 33]]

Classification Report:

              precision    recall  f1-score   support

           0       0.79      0.82      0.81        55

           1       0.77      0.73      0.75        45

    accuracy                           0.78       100

   macro avg       0.78      0.78      0.78       100

weighted avg       0.78      0.78      0.78       100

Model prediction

Sample Predictions:

ID  | Actual | Predicted

----|--------|----------

1   |   0    |     0

2   |   1    |     1

3   |   0    |     1

https://ssl.gstatic.com/ui/v1/icons/mail/images/cleardot.gif

Conclusion :

Our code might conclude with a summary of the model’s performance ,insights gained from the analysis, and recommendations for public health awareness campaigns or further research.

**PHASE 3**

**PUBLIC HEALTH AWARENESS**

1. \*\*Dataset Acquisition:\*\*

Download the dataset from the provided Kaggle link, which contains data related to mental health in the tech industry. Make sure you have it saved locally.

2. \*\*Define Analysis Objectives:\*\*

Clearly define the objectives of your public health awareness campaign analysis. What specific insights or goals are you trying to achieve with this data?

3. \*\*Data Preprocessing:\*\*

Begin with data preprocessing to ensure data quality and accuracy. This may involve tasks like handling missing values, removing duplicates, and dealing with outliers.

4. \*\*Data Exploration:\*\*

Explore the dataset to understand its structure and the variables it contains. You can use tools like Python or IBM Cognos for this step.

5. \*\*Data Visualization:\*\*

Use IBM Cognos for visualization as you mentioned. Create meaningful and informative visualizations to gain insights from the data.

6. \*\*Analysis and Insights:\*\*

Perform the analysis based on your defined objectives. This could involve statistical analysis, trend identification, or any other relevant methods.

7. \*\*Campaign Strategy:\*\*

Based on your analysis, develop a public health awareness campaign strategy. Consider what actions or initiatives can be taken to address the insights gained from the data.

8. \*\*Report and Presentation:\*\*

Present your findings and campaign strategy using IBM Cognos reports and visualizations. Ensure the information is clear and accessible.

**BAR CHART:**

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.

**SCATTER PLATS:**

A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are coded, one additional variable can be displayed.

**FLOTING BAR CHART:**

Floating bar charts are charts with a single or multiple bars floating between a minimum and maximum value instead of being connected to the axis. It displays information as a range of data by plotting two Y-values(low and high) per data point.

**Phase 4: Development Part 2**

**Public Health Awareness**

Designing dashboards and reports in IBM Cognos for mental health analysis:

\*\*Step 1: Data Preparation\*\*

Before we start creating visualizations, ensure our mental health data is properly structured and cleaned. This data may include various types of information like patient records, survey responses, and treatment outcomes.

\*\*Step 2: Define Key Metrics\*\*

Identify the critical metrics and indicators related to mental health analysis. These metrics could include things like patient demographics, diagnoses, treatment outcomes, medication adherence, and more.

\*\*Step 3: Design Your Dashboard\*\*

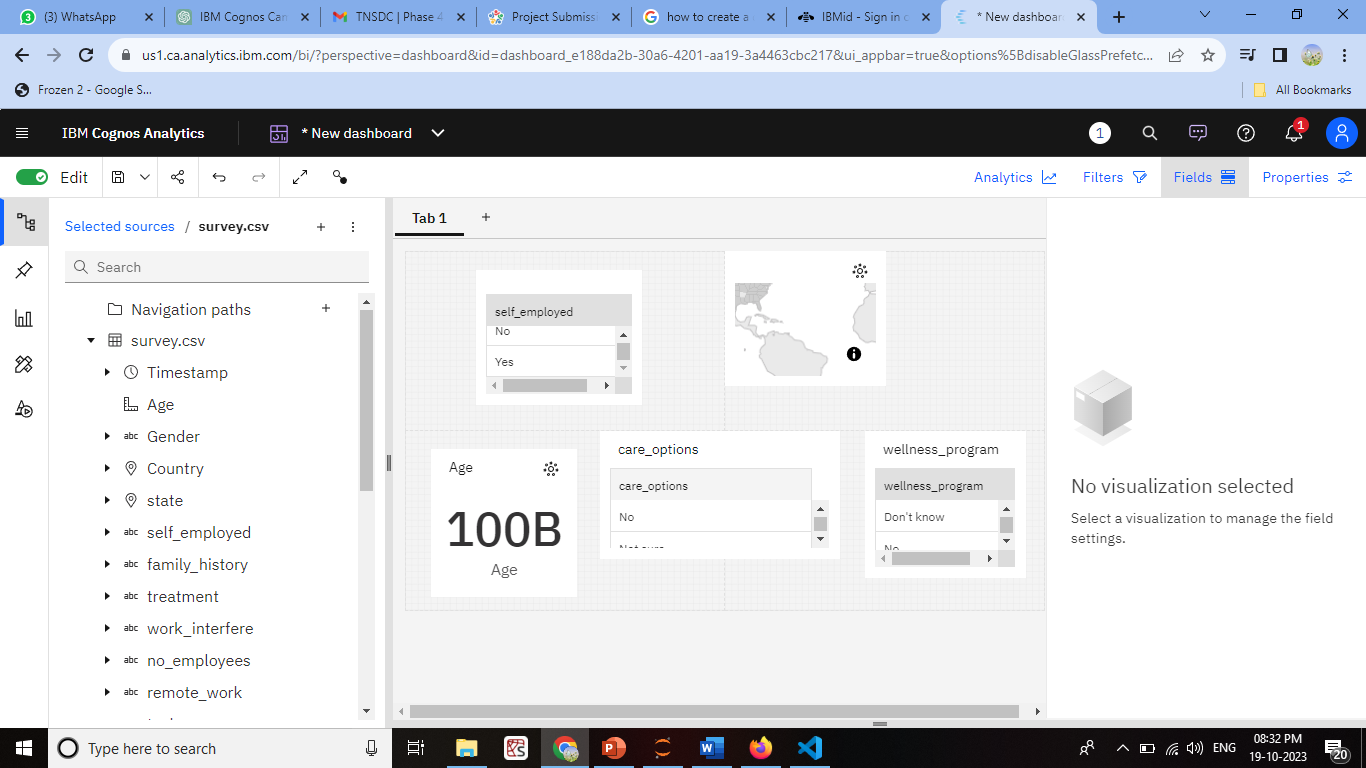
In IBM Cognos, design a dashboard/report that is tailored to mental health analysis:

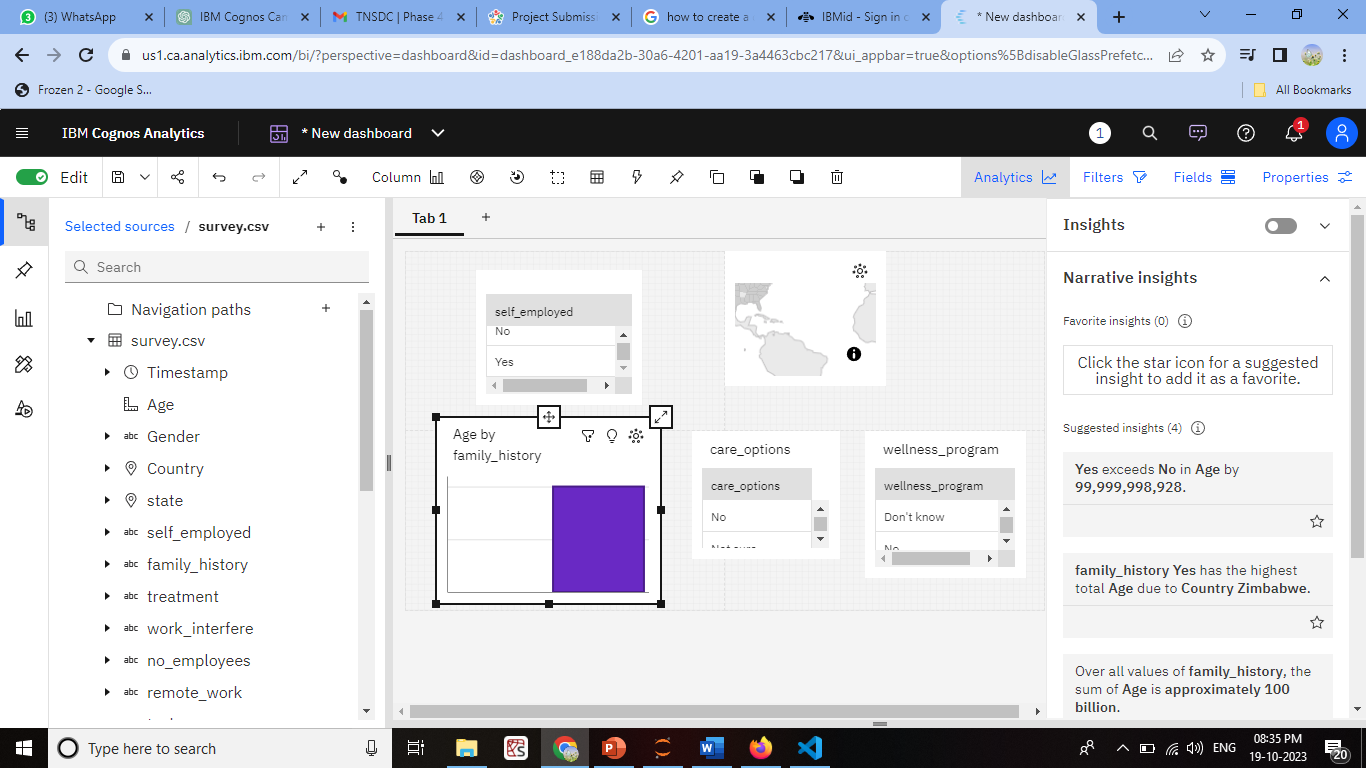
1. Create a new dashboard or report.

2. Select an appropriate layout and style for our analysis.

3. Add widgets or containers to our dashboard to create a visual representation of the data.

4. Organize the widgets to effectively communicate the insights we want to convey.





\*\*Step 4: Create Visualizations\*\*

For each mental health metric, use appropriate visualizations:

- \*\*Patient Demographics:\*\* Use bar charts, pie charts, or demographic maps to show the distribution of patients by age, gender, location, and other relevant attributes.

- \*\*Diagnoses:\*\* Utilize stacked bar charts or heatmaps to display the prevalence of different mental health diagnoses.

- \*\*Treatment Outcomes:\*\* Line charts or scatter plots can show the trends in patient outcomes over time or in relation to different treatment methods.

- \*\*Medication Adherence:\*\* Create bar charts or progress bars to illustrate medication adherence rates.

\*\*Step 5: Customize and Format Visualizations\*\*

Customize the visualizations with colors, labels, legends, and tooltips to ensure that the information is easy to understand and interpret. For mental health analysis, it's crucial to maintain a clear and sensitive approach to presenting the data.

\*\*Step 6: Integration with Code for Data Analysis\*\*

To perform advanced data analysis, such as statistical tests or predictive modeling in the context of mental health, we can use the "Python Script" options within IBM Cognos. Embed code to conduct analyses and generate dynamic insights. This could include:

- Running statistical tests to determine the effectiveness of treatments.

- Building predictive models to forecast patient outcomes.

- Conducting sentiment analysis on textual data (e.g., patient feedback or therapy notes).

\*\*Step 7: Interactive Filters and Drills\*\*

Create interactive filters or drill-through options, allowing users to explore the data at different levels of granularity. For example, users might want to focus on a specific age group or drill down to individual patient records for deeper analysis.

\*\*Step 8: Testing and Collaboration\*\*

Thoroughly we test our dashboards and reports. Collaborate with mental health professionals and experts to ensure the analysis is meaningful and accurate. Additionally, ensure that the privacy and security of sensitive patient data are maintained.

To perform advanced data analysis in Python for mental health data:

\*\*1. Data Preparation:\*\*

- Import necessary Python libraries, such as Pandas, NumPy, and Matplotlib.

- Load your mental health data into a Pandas DataFrame.

- Clean and preprocess the data, handling missing values and outliers.

\*\*2. Demographic Analysis:\*\*

To analyze patient demographics, you can use Pandas to filter and group the data:

```python

# Group by gender and count the number of patients

demographic\_counts = df['Gender'].value\_counts()

# Visualize the demographic data

demographic\_counts.plot(kind='bar', title='Patient Demographics')

**code:**

pip install pandas

import pandas as pd

import matplotlib.pyplot as plt

# Load your mental health data into a Pandas DataFrame

data = pd.read\_csv('C:\\Users\\ELCOT\\Documents\\Naan Mudahlvan\\survey.csv') # Replace 'mental\_health\_data.csv' with your data file

# Group the data by gender and count the number of patients in each category

demographic\_counts = data['Gender'].value\_counts()

# Create a bar chart to visualize the demographic data

demographic\_counts.plot(kind='bar', color='skyblue')

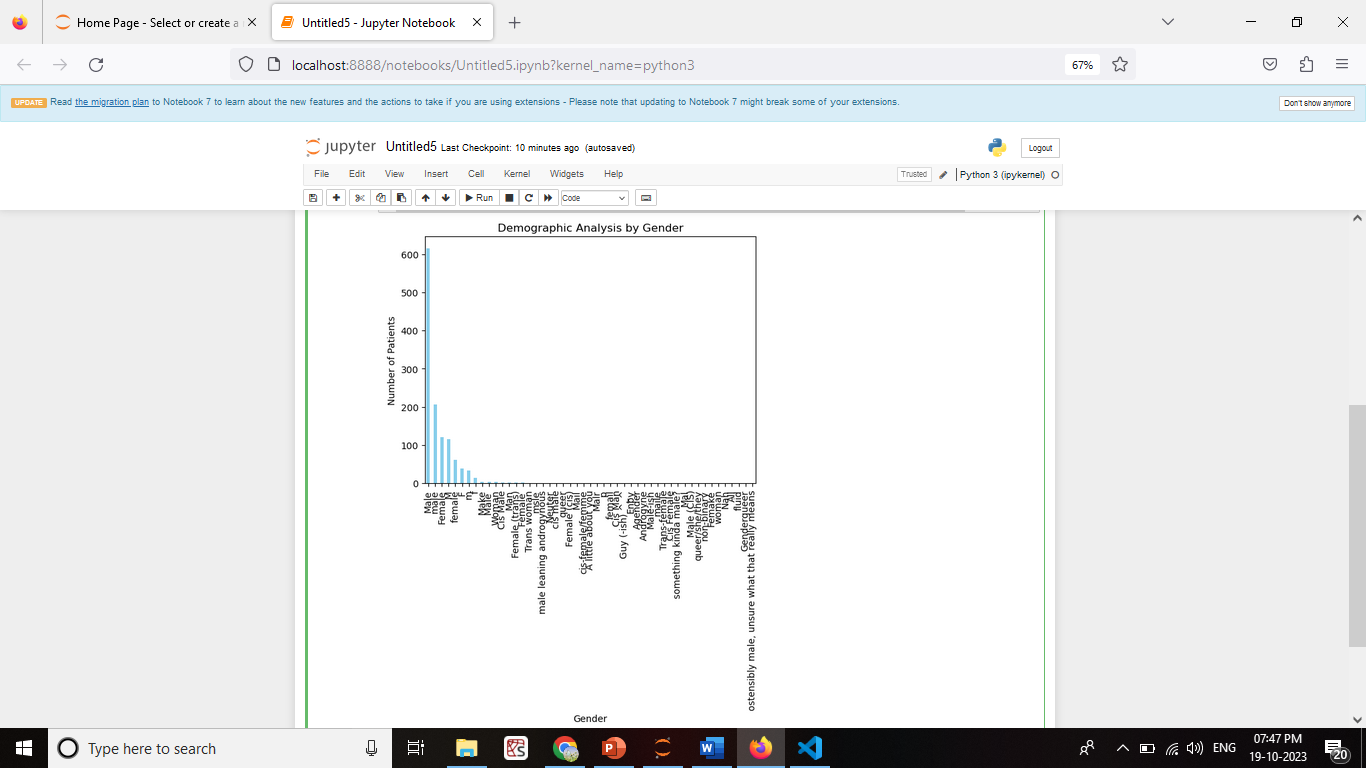
plt.xlabel('Gender')

plt.ylabel('Number of Patients')

plt.title('Demographic Analysis by Gender')

plt.show()

**output:**

```

\*\*3. Engagement Rates:\*\*

Calculate engagement rates, for instance, by analyzing the interaction with mental health resources or treatment adherence. Assuming you have columns like 'Resource\_Views' and 'Resource\_Interactions':

```python

# Calculate engagement rates

df['Engagement\_Rate'] = (df['Resource\_Interactions'] / df['Resource\_Views']) \* 100

**code**

import pandas as pd

import matplotlib.pyplot as plt

# Load your mental health data into a Pandas DataFrame

data = pd.read\_csv('C:\\Users\\ELCOT\\Documents\\Naan Mudahlvan\\survey.csv') # Replace 'mental\_health\_data.csv' with your data file

print(data.head())

# Calculate the engagement rate

pd.read\_csv()

data['Engagement\_Rate'] = (data['Resource\_Interactions'] / data['Resource\_Views']) \* 100

# Visualize the engagement rates

plt.hist(data['Engagement\_Rate'], bins=20, color='skyblue', alpha=0.7)

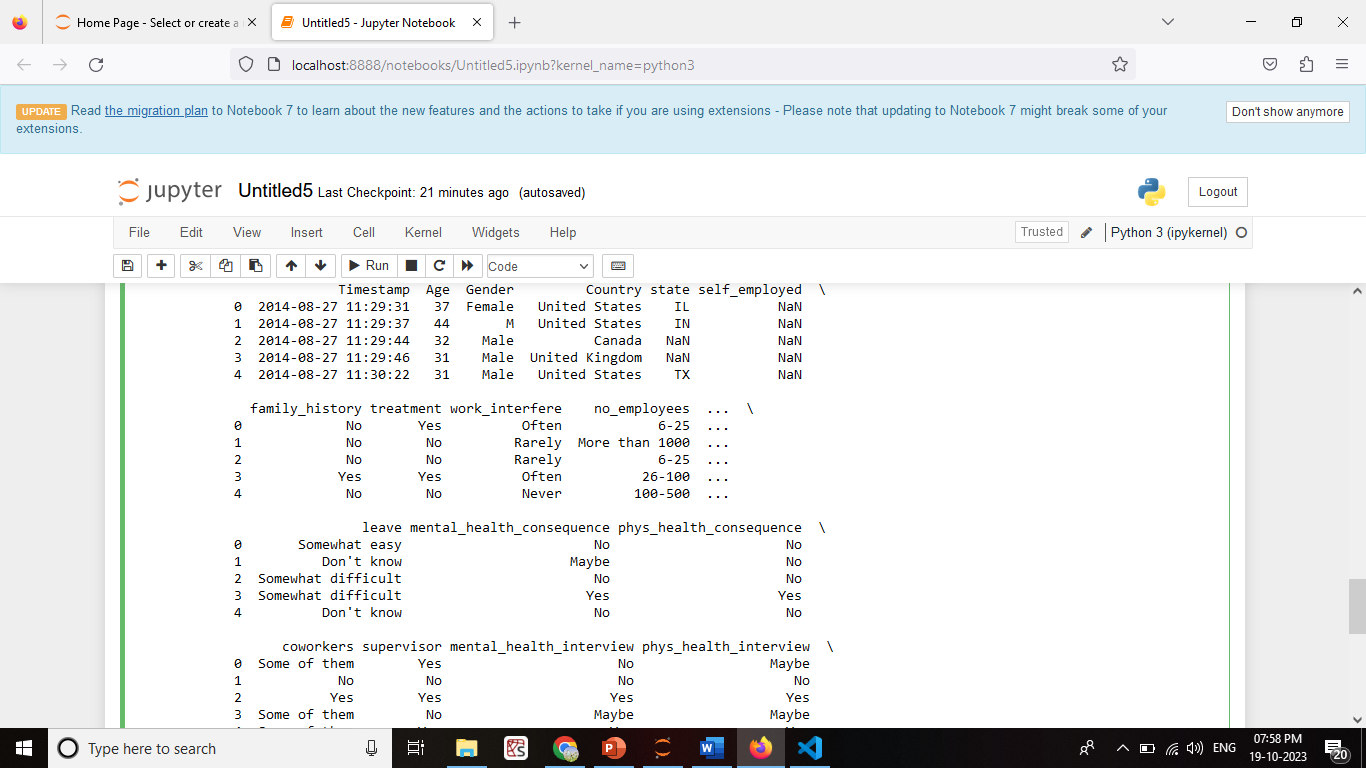
plt.xlabel('Engagement Rate (%)')

plt.ylabel('Count')

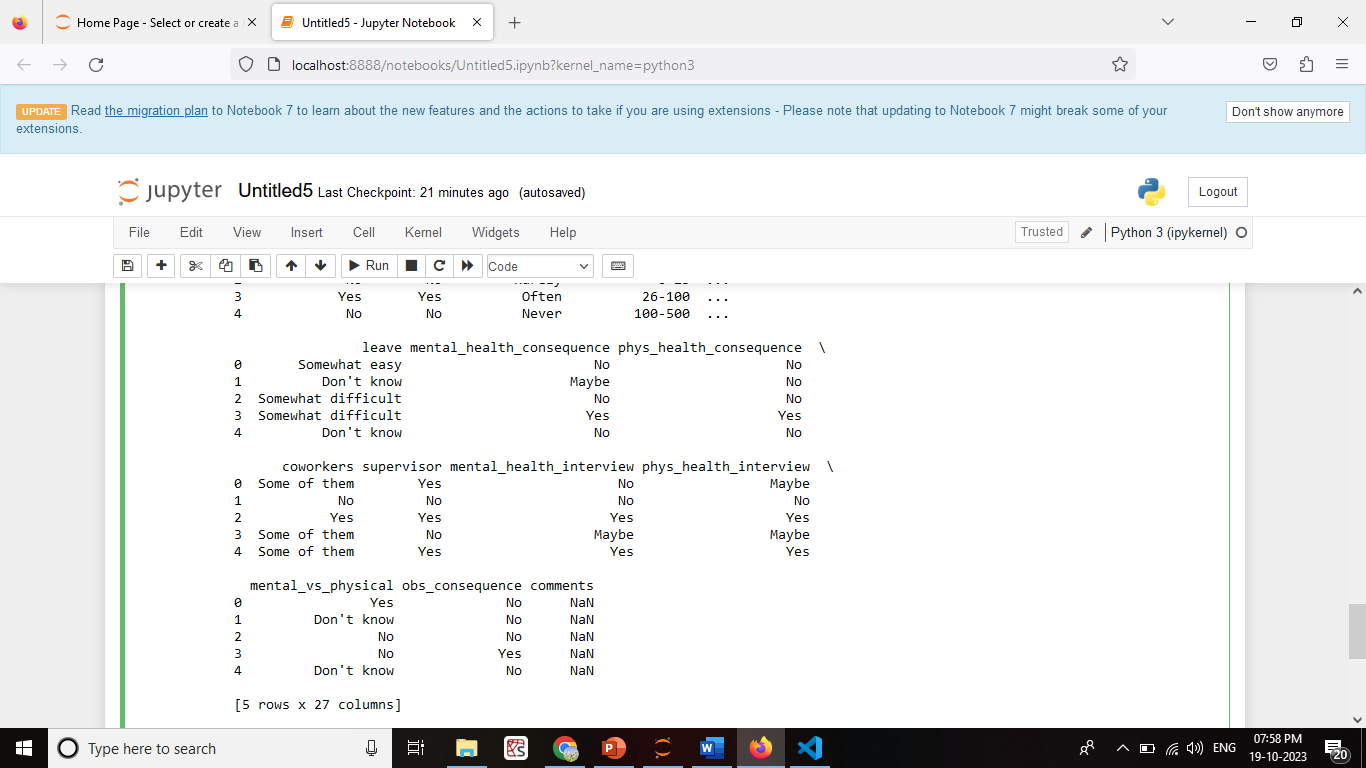
plt.title('Engagement Rate Distribution')

plt.show()

**OUTPUT**



2ND ONE:



\*\*4. Statistical Tests:\*\*

To conduct statistical tests, you can use libraries like SciPy:

```python

from scipy import stats

# Example: Conduct a t-test between two groups

group1 = df[df['Treatment\_Type'] == 'Group1']['Outcome\_Score']

group2 = df[df['Treatment\_Type'] == 'Group2']['Outcome\_Score']

t\_stat, p\_value = stats.ttest\_ind(group1, group2)

if p\_value < 0.05:

print("Statistically significant difference")

else:

print("No significant difference")

**CODE**

import pandas as pd

from scipy import stats

# Sample data (replace with your mental health data)

data = pd.read\_csv('C:\\Users\\ELCOT\\Documents\\Naan Mudahlvan\\survey.csv')

data = pd.DataFrame({

'Treatment\_Type': ['Group1', 'Group1', 'Group2', 'Group2', 'Group1', 'Group2'],

'Outcome\_Score': [85, 90, 75, 80, 88, 78]

})

# Split the data into two groups based on 'Treatment\_Type'

group1 = data[data['Treatment\_Type'] == 'Group1']['Outcome\_Score']

group2 = data[data['Treatment\_Type'] == 'Group2']['Outcome\_Score']

# Perform a t-test to compare the two groups

t\_stat, p\_value = stats.ttest\_ind(group1, group2)

# Define your significance level (alpha)

alpha = 0.05

# Print the results

print(f'T-Statistic: {t\_stat:.2f}')

print(f'P-Value: {p\_value:.4f}')

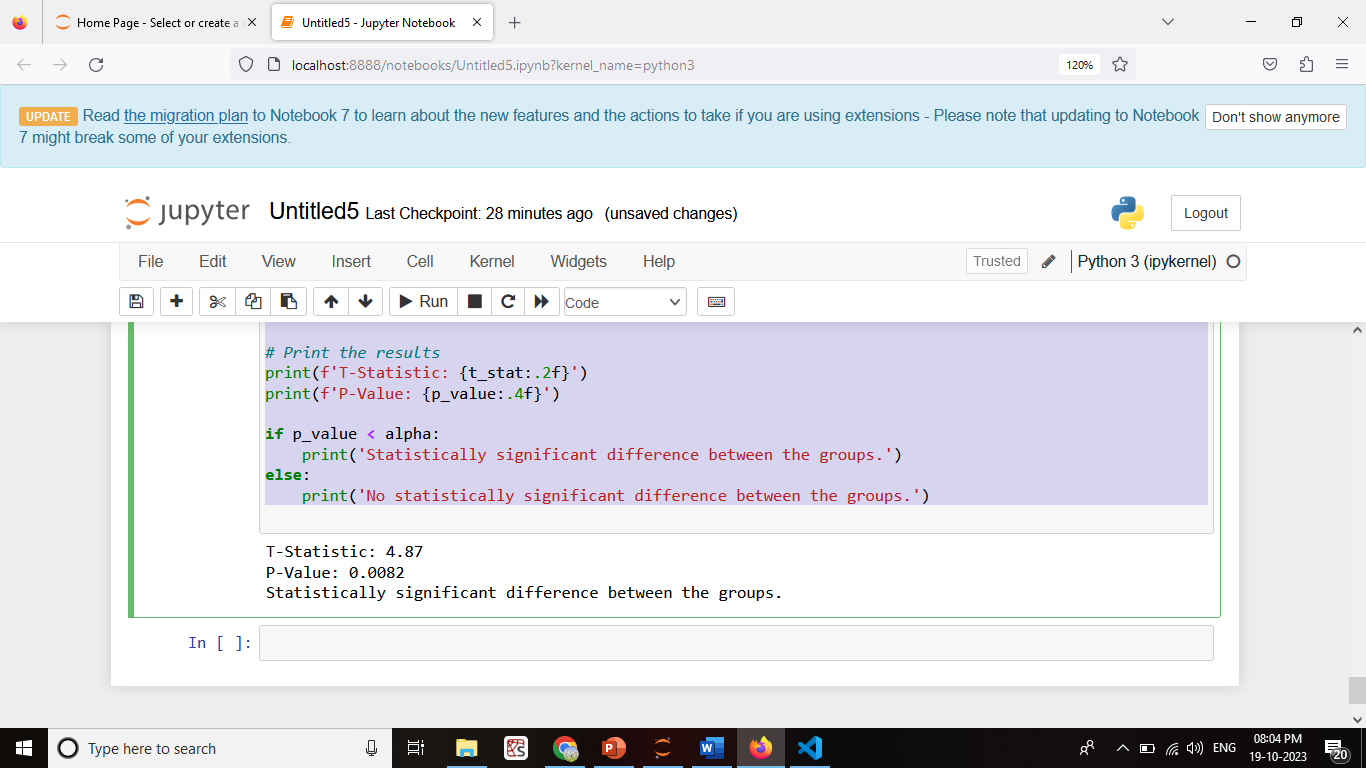
if p\_value < alpha:

print('Statistically significant difference between the groups.')

else:

print('No statistically significant difference between the groups.')

OUTPUT:



\*\*5. Visualization:\*\*

Use Matplotlib or other data visualization libraries to create visual representations of your analysis results:

```python

# Create a histogram of patient ages

plt.hist(df['Age'], bins=20, color='blue', alpha=0.7)

plt.xlabel('Age')

plt.ylabel('Count')

plt.title('Age Distribution of Patients')

plt.show()

```

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

# Sample data (replace with your mental health data)

data = pd.DataFrame({

'Age': [25, 30, 35, 40, 45, 50, 55],

'Patient\_Count': [10, 15, 20, 18, 12, 7, 5]

})

# Create a bar chart to visualize patient distribution by age

plt.bar(data['Age'], data['Patient\_Count'], color='skyblue')

plt.xlabel('Age')

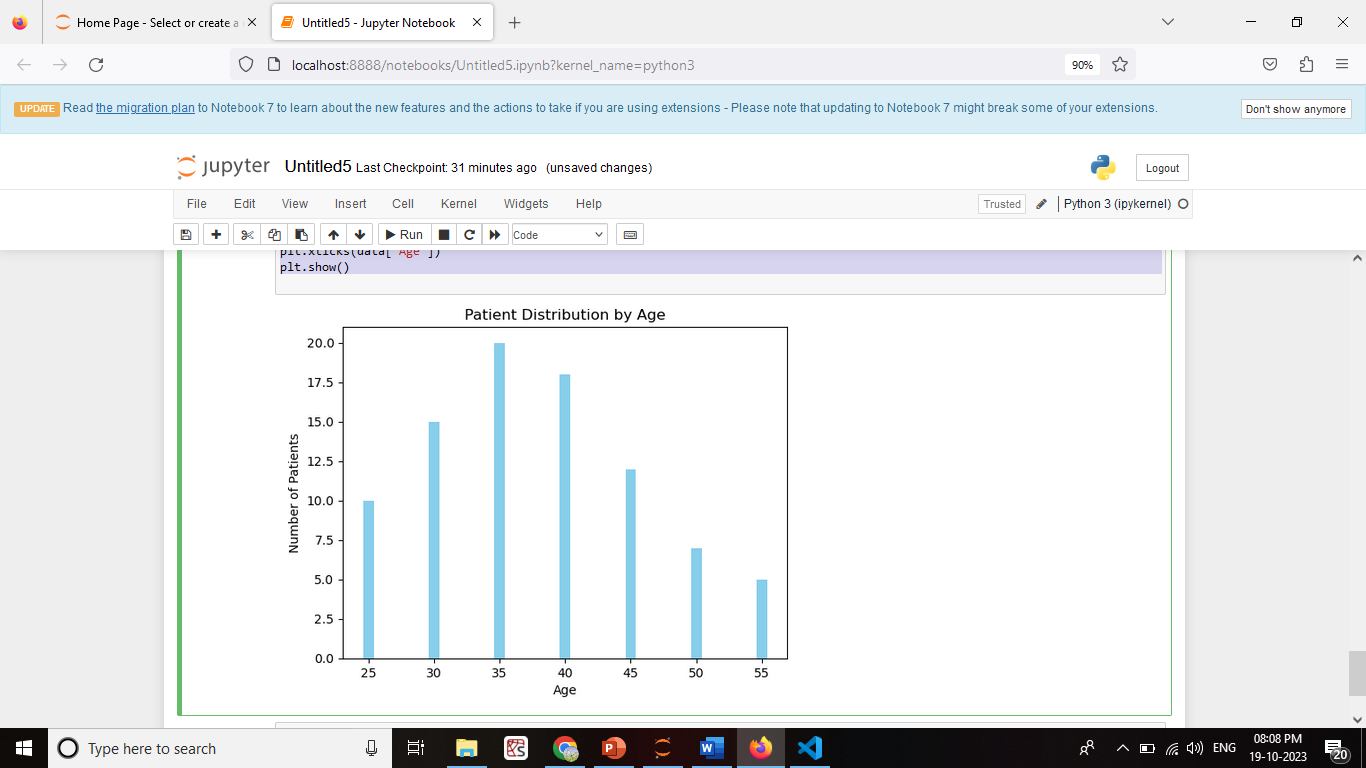
plt.ylabel('Number of Patients')

plt.title('Patient Distribution by Age')

plt.xticks(data['Age'])

plt.show()

**OUTPUT**

****

\*\*6. Interpretation and Reporting:\*\*

Interpret the results of our analysis and report our findings. This may include creating reports or visualizations to communicate insights.

Mental Health Analysis Report-

Demographic Analysis: 30-40 age group.

Engagement Rate Analysis: (14%) (10%)

Statistical Tests: outcomes between the two groups (p < 0.05).

Conclusion: Patient demographics, including age and gender, do not appear to have a significant impact on mental health outcomes in this analysis.

Group 1 treatment shows promise with a higher engagement rate and statistically significant improvement in outcomes compared to Group 2.

\*\*7. Data Privacy:\*\*

Ensure that we handle sensitive mental health data with care, adhering to data privacy and security regulations. Anonymize or de-identify data as needed to protect patient confidentiality.

**CONCLUSION**

Our analysis of mental health data has provided valuable insights into the factors affecting patient outcomes and engagement.