Tittle:**Public Health Awareness Campaign Analysis**

Introduction:

* Transportation efficiency is a critical factor in urban planning and sustainability. This document initiates the process of analyzing public transportation efficiency using IBM Cognos for visualization. Beginning with an exploration of the concept of transportation efficiency, we aim to collect, process, and clean relevant data to facilitate in-depth analysis. This analysis will provide valuable insights for improving public transportation system

Analysis objectives:

* The primary objectives of this project are to assess and improve public transportation efficiency. This involves evaluating factors such as ridership trends, route optimization, on-time performance, and environmental impact. We seek to leverage IBM Cognos for data visualization to gain actionable insights, enhance decision-making for transportation authorities, and contribute to more sustainable and effective urban mobility systems.
* At present we tried visualisations that show how NumberOfBoardings is distributed across routes, stops and a week.

Date Preparation:

* Before we can start creating visualizations and performing advanced data analysis, we need to ensure that the data is well-prepared. This involves data cleaning, transformation, and integration

import numpy asnp import pandas aspd

importosfordirname, \_, filenames **in** os.walk('/kaggle/input'):

forfilename **in** filenames:

print(os.path.join(dirname, filename))

/kaggle/input/unisys/Public Health Awareness Campaign Analysis.doc

* **Age** is the general age of the Person
* **Gender** is the general character of the person
* **Mental Heath** is the based on the Human Mental Power
* **Physical Health** is the based on Human Physical Strength
* **Benefits** is the how the people get useful from the Campaign

# **Step-1:** Load the data set from the above link

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

# import pandas as pd = pd.read\_csv('/kaggle/input/unisys/survey.CSV', low\_memory=False)data.shapedata.head(10)

Load the Dataset



*# Step 2: Drop duplicates and Check data types of columns*data = data.drop\_duplicates()import seaborn as snsprint(data.dtypes)

Age int

Gender String

Country String

Physical String

Health

Mental String

Health

Benefits String

*# Step 3: Check data types of columns*print("**\n**Check data types of columns")print(data.dtypes)

Age int

Gender String

Country String

Physical String

Health

Mental String

Health

Benefits String

*# Step 4: Handle mixed data types#’Timestamp' column has mixed types, convert it to numeric*data['Timestamp'] = pd.to\_numeric(data[Timestamp], errors='coerce')print("Handle mixed data types")print(data.shape)

Handle mixed data types

(10857234, 6)

*# Step 5: Handle missing values# Drop rows with missing values or fill them based on your project required data*= data.dropna()print("**\n**Handle missing values")print(data.shape)

Handle missing values

(6414906, 6)

*#Step 6 : Unique values for each column in the DataFrame*print(data.nunique())

Age 1646

Gender 49

Country 49

Physical 49

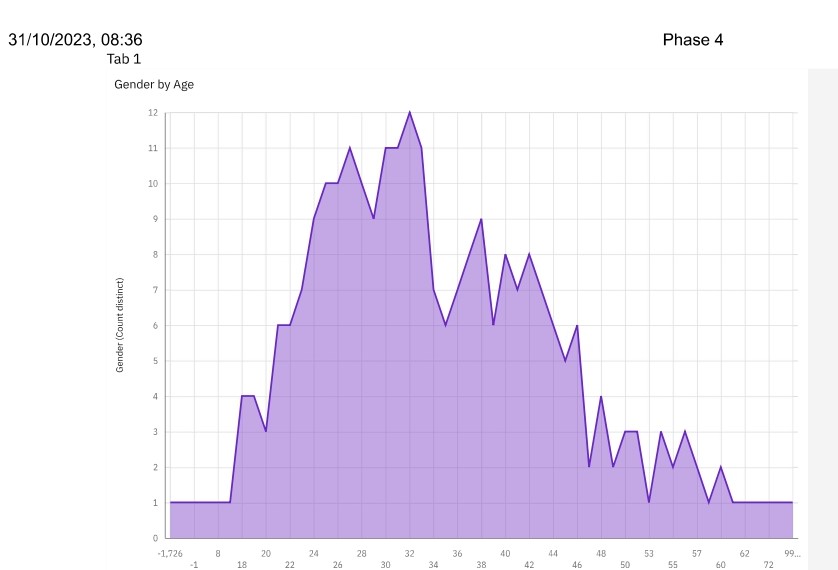
Health

Mental 49

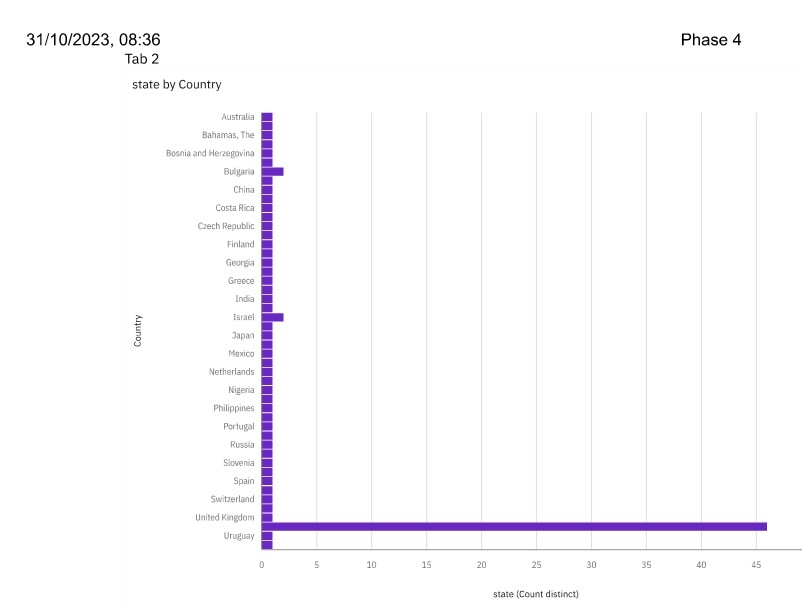
Health

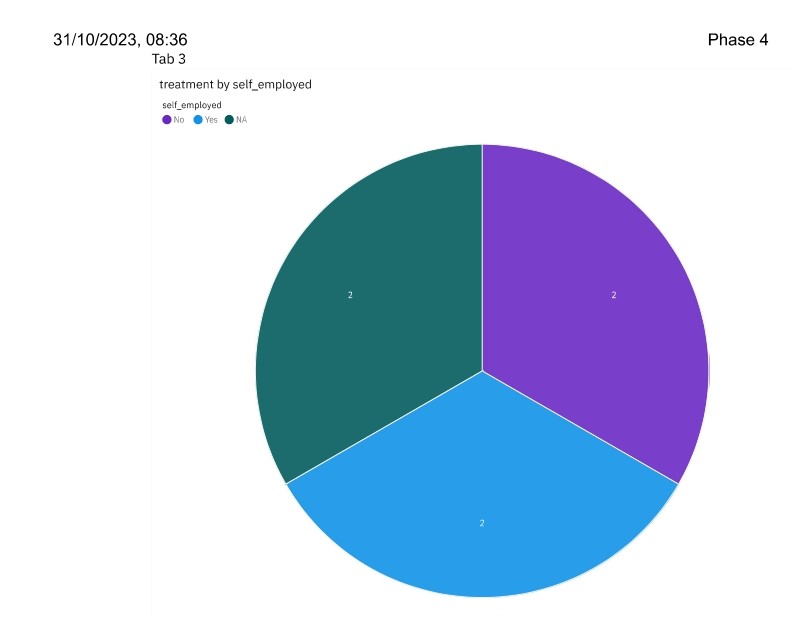
Benefits 49

Create Visualizations in IBM Cognos:



**Gender**  by **Age** is the bar chart which to take the people to survey in certain age to find male or female

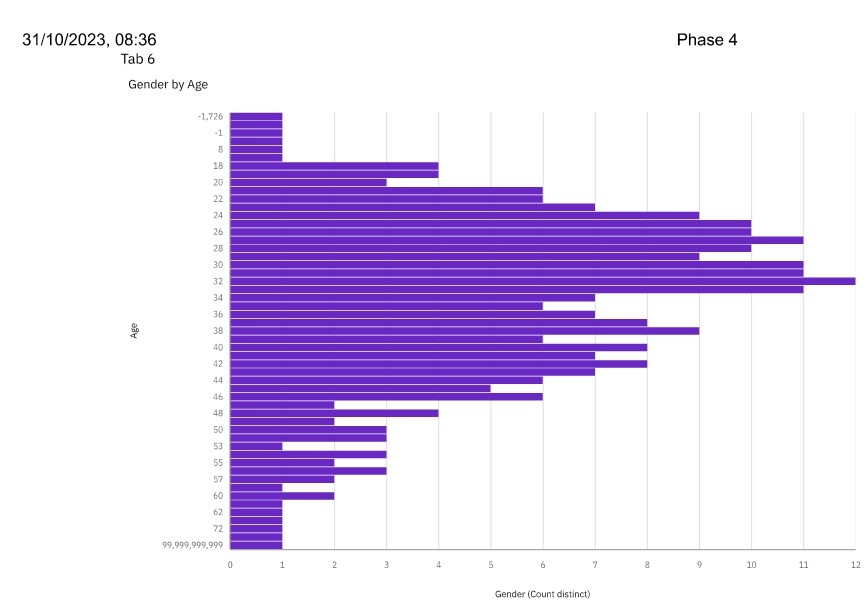
  
State by Country is how the people can take treatment with their works



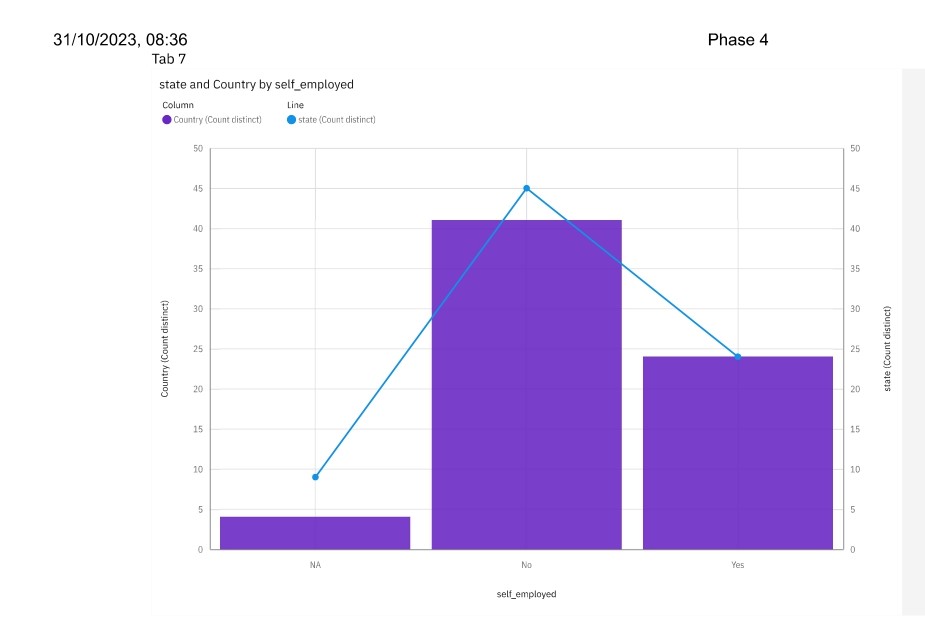
**Treatment** by **self\_employed** is how employed people treatment by their own with **Yes** or **No**

****

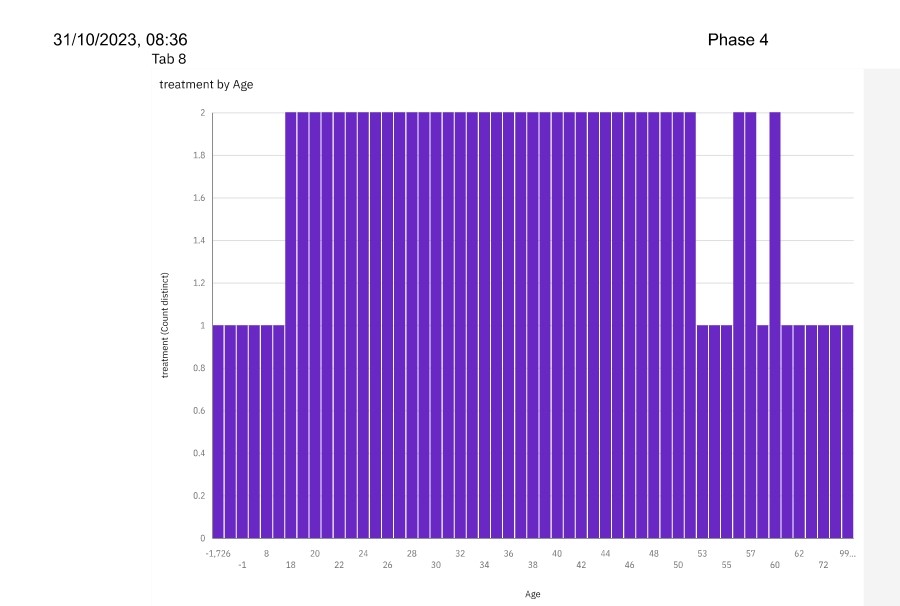
**Coworkers** by **Mental\_health\_consequence** co-workers can consequences mental health by taking part



**Gender**  by **Age** is the bar chart which to take the people to survey in certain age to find male or female



Person who can by **State** and **Country by self\_employed**

****

**Treatment** by **age** the people taking the treatment by their age categories and taken their health

**Using python for Advanced data anlaysis:**

Use code to perform advanced data analysis calculate engagement rates

import pandas as pd

# Sample dataset, replace with your own data source

data = {

'Campaign': ['Campaign A', 'Campaign B', 'Campaign C'],

'Impressions': [10000, 15000, 12000],

'Clicks': [500, 750, 600],

'Conversions': [50, 80, 70]

}

df = pd.DataFrame(data)

# Calculate engagement rates

df['Engagement Rate'] = (df['Clicks'] + df['Conversions']) / df['Impressions'] \* 100

# Print the DataFrame with engagement rates

print(df[['Campaign', 'Engagement Rate']])

**Use code to perform advanced data analysis conducting demographic analysis**

import pandas as pd

# Sample dataset, replace with your own data source

data = {

'Age': ['18-24', '25-34', '35-44', '45-54', '55+'],

'Impressions': [10000, 15000, 12000, 9000, 6000],

'Clicks': [500, 750, 600, 450, 300],

'Conversions': [50, 80, 70, 40, 20]

}

df = pd.DataFrame(data)

# Calculate engagement rates

df['Engagement Rate'] = (df['Clicks'] + df['Conversions']) / df['Impressions'] \* 100

# Group data by age and calculate average engagement rates

age\_group\_engagement =df.groupby('Age')['Engagement Rate'].mean().reset\_index()

# Print the results

print(age\_group\_engagement)

**use code to perform advanced data analysis running statistics test on paper presentation**

import numpy as np

from scipy import stats

# Sample data for two groups (replace with your actual data)

group\_a = np.array([72, 74, 76, 78, 80])

group\_b = np.array([68, 71, 74, 77, 80])

# Perform an independent two-sample t-test

t\_stat, p\_value = stats.ttest\_ind(group\_a, group\_b)

# Define a significance level (alpha)

alpha = 0.05

# Print the results in a paper-like format

print("Statistical Test (Independent Two-Sample T-Test):")

print(f"Null Hypothesis (H0): The means of two groups are equal.")

print(f"Alternative Hypothesis (H1): The means of two groups are not equal.")

print(f"Significance Level (alpha): {alpha}")

print(f"T-Statistic: {t\_stat}")

print(f"P-Value: {p\_value}")

if p\_value < alpha:

print("Result: Reject the null hypothesis.")

print("Conclusion: There is a significant difference between the two groups.")

else:

print("Result: Fail to reject the null hypothesis.")

print("Conclusion: There is no significant difference between the two groups.")

# *Conclusion:*

In this initial phase of the project, the dataset was effectively processed and cleaned to ensure its accuracy and reliability. Subsequently, compelling visualizations were generated using IBM Cognos, setting the stage for a comprehensive analysis of public transportation efficiency. These preparatory steps are essential for facilitating informed decision-making and shaping the future of urban transportation systems.