

# Title : 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.

## Data Cleaning and Preprocessing:

In [1]:

```
import numpy as np import pandas as pd

import osfor dirname, _, filenames in os.walk('/kaggle/input'):

    for filename 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.shape
data.head(10)
```

Load the Dataset

Timestamp	Age	Gender	Country	state	self_empt	family_his	treatment	work_inte
27-08-2014 11:29	37	Female	United States	IL	NA	No	Yes	Often
27-08-2014 11:29	44	M	United States	IN	NA	No	No	Rarely
27-08-2014 11:29	32	Male	Canada	NA	NA	No	No	Rarely
27-08-2014 11:29	31	Male	United Kingdom	NA	NA	Yes	Yes	Often
27-08-2014 11:30	31	Male	United States	TX	NA	No	No	Never
27-08-2014 11:31	33	Male	United States	TN	NA	Yes	No	Sometime
27-08-2014 11:31	35	Female	United States	MI	NA	Yes	Yes	Sometime
27-08-2014 11:32	39	M	Canada	NA	NA	No	No	Never
27-08-2014 11:32	42	Female	United States	IL	NA	Yes	Yes	Sometime
27-08-2014 11:32	23	Male	Canada	NA	NA	No	No	Never
27-08-2014 11:32	31	Male	United States	OH	NA	No	Yes	Sometime
27-08-2014 11:32	29	male	Bulgaria	NA	NA	No	No	Never
27-08-2014 11:33	42	female	United States	CA	NA	Yes	Yes	Sometime
27-08-2014 11:33	36	Male	United States	CT	NA	Yes	No	Never
27-08-2014 11:33	27	Male	Canada	NA	NA	No	No	Never

```
# Step 2: Drop duplicates and Check data types of columns
data = data.drop_duplicates()
import seaborn as sns
print(data.dtypes)
```

```
Age          int
Gender       String
Country      String
Physical     String
Health
```

Mental	String
Health	
Benefits	String

```
# Step 3: Check data types of columns
print("\nCheck 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 = data.dropna()
print("\nHandle missing values")
print(data.shape)
```

Handle missing values

```
(6414906, 6)
```

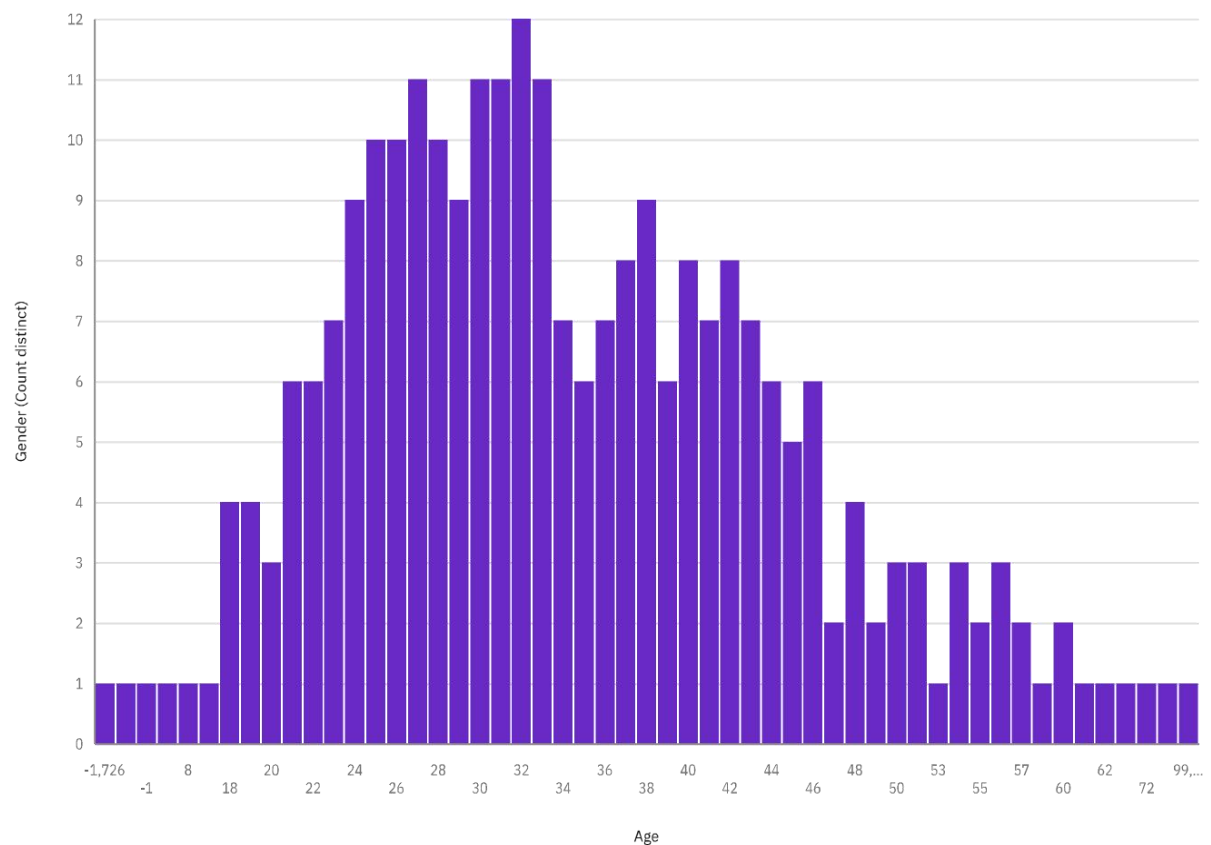
```
# Step 6 : Unique values for each column in the DataFrame
print(data.nunique())
```

Age	1646
Gender	49
Country	49
Physical	49
Health	
Mental	49
Health	
Benefits	49

## Visualization on IBM Cognos:

1

Gender by Age



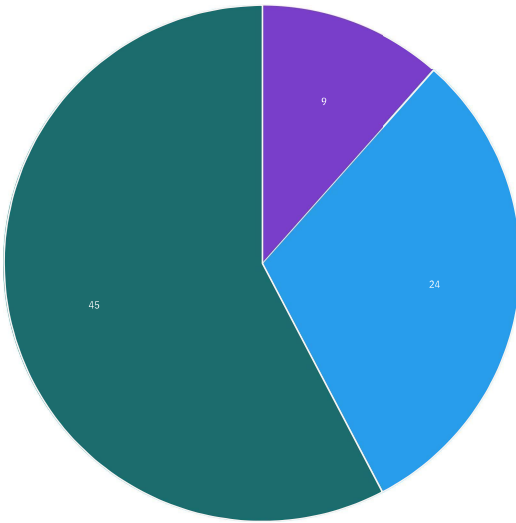
**Gender** by **Age** is the bar chart which to take the people to survey in certain age to find male or female

2

state by self\_employed

self\_employed

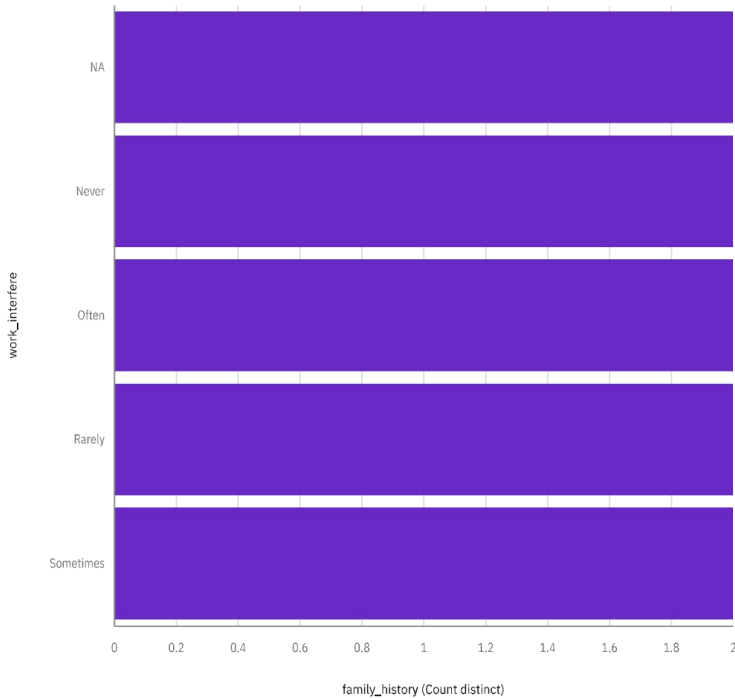
NA Yes No



**State by self\_employed** which means how the people are self employed by statewise

3

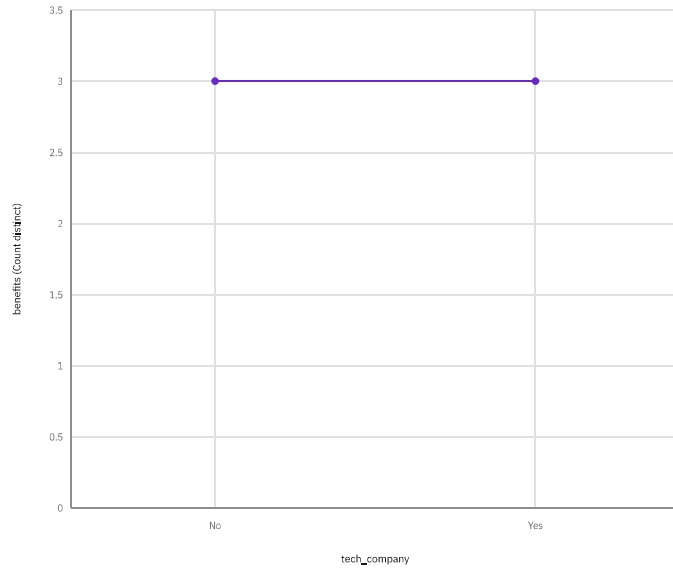
family\_history by work\_interfere



**Family history by work interfere** which people to make there doing the job

4

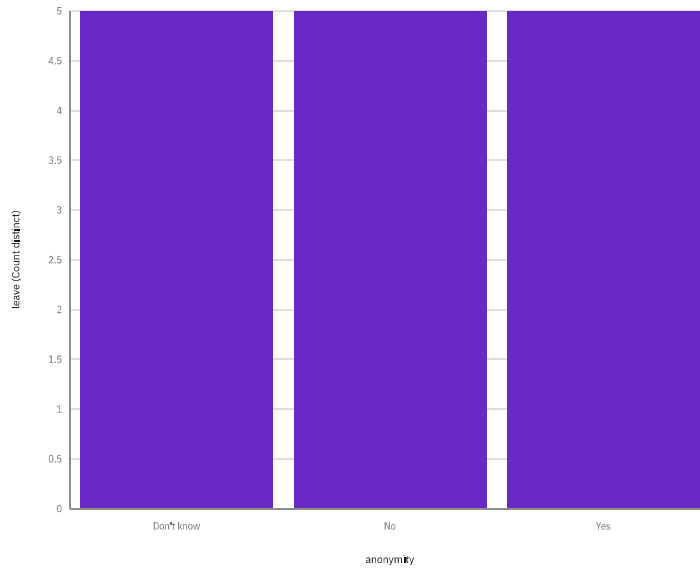
benefits by tech\_company



The person who benefit  
who by tech companys

5

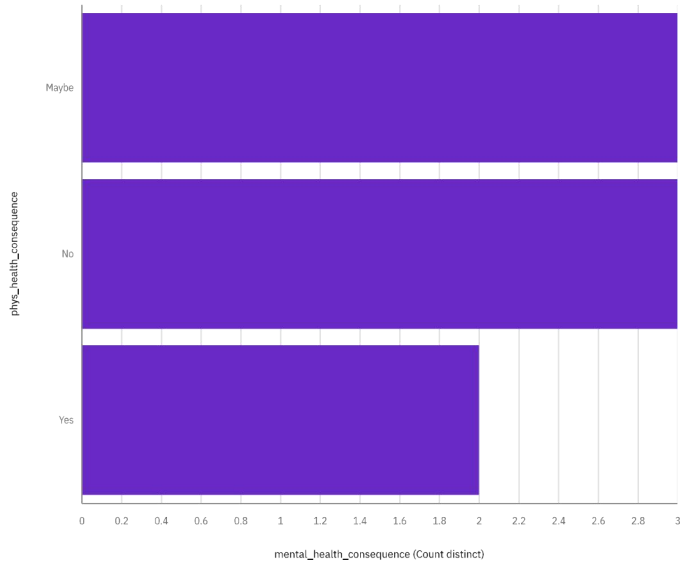
leave by anonymity



**Leave by anonymly** who can take  
leave on sick or illnes

6

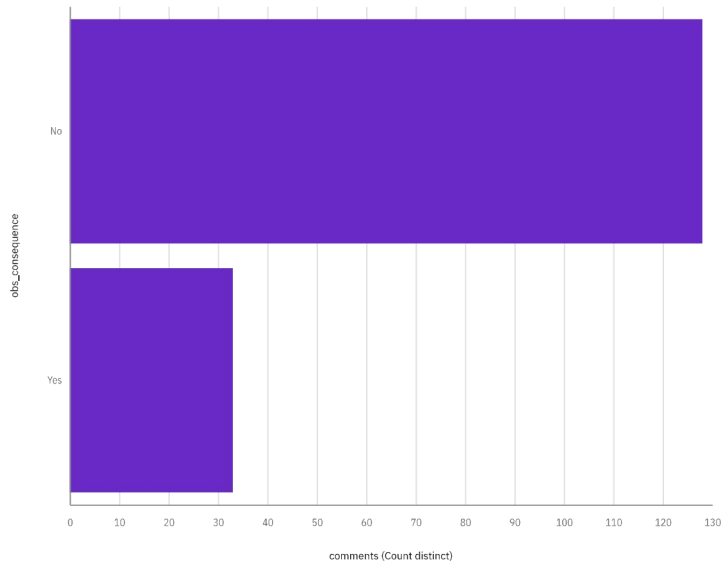
mental\_health\_consequence by phys\_health\_consequence



Person who can by **Mental\_health** and **Physical\_health**

8

comments by obs\_consequence



**Comments** by **obs\_consequence** which the who see some consequences on the commented on that