Project Title: Public Health Awareness

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

Phase 1: Project Definition and Design Thinking

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.

GitHub account Link: https://github.com/IBM-NaanMudhalvan/Public-Health-Awareness

ANALYSIS OBJECTIVES:

- ✓ Demographic Analysis
- ✓ Frequency of activities
- ✓ Engagement by company size
- ✓ Comparison of responses between countries
- ✓ Comparison by gender
- ✓ Time analysis
- ✓ Correlation
- ✓ Geographic analysis

DATA COLLECTION:

Based on the provided data, we can extract metrics, audience demographics, and survey-related information:

Metrics:

Age:

- ✓ Minimum Age: 23
- ✓ Maximum Age: 46
- ✓ Average Age: Approximately 34.2 years (mean)

Audience Demographics:

Gender Distribution:

- ✓ Male: Majority of the respondents
- ✓ Female: Fewer respondents compared to males

Geographic Distribution:

- ✓ Majority of respondents are from the United States, particularly states like Illinois (IL), Texas (TX), and California (CA).
- ✓ Other significant countries include Canada and the United Kingdom.

Surveys:

- ✓ The survey captured responses related to certain behaviors or preferences, indicated by the "Yes," "No," "Often," "Sometimes," and "Rarely" responses for specific questions.
- ✓ Frequency of Engaging in Certain Activities: Respondents were asked about the frequency of engaging in certain activities, ranging from "Never" to "Often."
- ✓ Company Size: Respondents were asked about the size of their respective companies, ranging from "1-5" to "More than 1000."

This information provides an overview of the metrics, audience demographics, and the nature of the survey based on the data provided.

VISUALIZATION STRATEGY:

To visualize the provided data, we'll focus on key aspects such as age distribution, gender distribution, frequency of certain responses, and geographic distribution. We'll use appropriate visualizations for each aspect. Let's define a visualization strategy based on these considerations:

1. **Age Distribution:**

Visualization: Histogram or bar chart.

Description: Show the distribution of ages to understand the age demographics of the respondents.

2. **Gender Distribution:**

Visualization: Pie chart or bar chart.

Description: Display the proportion of respondents for each gender category.

3. **Frequency of Responses ("Often," "Sometimes," etc.):**

Visualization: Bar chart.

Description: Show the frequency of responses for a specific question to identify common patterns.

4. **Geographic Distribution:**

Visualization: Geographic map or bar chart.

Description: Display the number of respondents from each country or state to understand the geographic distribution.

Let's proceed with creating these visualizations using the provided data. We'll start with age distribution and then move on to the other aspects.

CODE INTEGRATION:

To enhance the analysis of the provided data, we can perform various tasks using code, including data cleaning, transformation, and statistical analysis. Here are some aspects of the analysis that can be enhanced using code:

1. **Data Parsing and Structuring**:

Parse the provided data to extract relevant information for each entry, such as age, gender, location, and responses to different questions.

2. **Data Cleaning:**

Remove any inconsistencies, errors, or invalid entries in the data to ensure accuracy and reliability.

3. **Data Transformation:**

Convert categorical variables into numerical representations for further analysis.

Transform the date and time information into a standardized format for easier manipulation.

4. **Statistical Analysis:**

Compute descriptive statistics (e.g., mean, median, mode, standard deviation) for numerical variables (e.g., age).

Generate visualizations (e.g., histograms, pie charts) to illustrate distributions and patterns in the data.

Perform correlations and hypothesis testing to explore relationships between different variables.

5. **Segmentation:**

Group the data based on specific criteria (e.g., age ranges, gender) to analyze patterns within different segments.

6. **Insights and Recommendations:**

Generate insights and recommendations based on the analyzed data, highlighting key trends and patterns.

Let's start by parsing the data and structuring it into a suitable format for further analysis. After that, we can proceed with data cleaning and transformation.