Public Health Awareness Project

Phase 2 Submission

In this phase, we aim to ponder upon and explore the innovative steps that will be taken in order to implement the project. This project mainly focuses on analysing public health campaigns in order to measure their effectiveness in reaching the target audience and increasing awareness.

Listed below is a step-by-step approach as to how the project will be implemented.

1.Design Thinking and Data Collection

Design thinking involves research on analysis objectives and clearly defining the goals of the analysis. This overview has been briefly explained in the Phase 1 document. Apart from the data collection strategies discussed in the Phase 1 document, we will be mainly referring to the Kaggle dataset provided to use along with the project assignment.

(https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey)

2.Data cleaning and processing

Data cleaning and preprocessing are crucial steps which involve preparing the raw data for analysis by addressing issues such as missing values, outliers, inconsistencies, and formatting problems. In this step, it is made sure that all data collected are formatted, uniformly standardised and documented so that they can be used for prediction and analysis in the forthcoming steps.

3.Data Analysis

Data analysis involves the exploration, transformation, and interpretation of data to uncover patterns, trends, and insights. In this project, this step aims to analysis objects actually affect the effectiveness of a particular campaign, which factor contributes the most to success of a campaign and

much more. Listed below are detailed explanation of the key steps in the data analysis process.

Descriptive Statistics:

Descriptive statistics involves understanding of the basic characteristics of the dataset. This includes measures like mean, median, mode, range, and standard deviation.

Data Visualization:

Visualizations (histograms, box plots, scatter plots, etc.) are used to explore the distribution of variables and relationships between variables.

Based on this exploration, a hypothesis about relationships and patterns in the data is formulated.

Machine Learning Strategies

In this project ML is used for analysing large datasets. ML can contribute to predictive modelling to assess the impact of different campaign factors on effectiveness. It can help identify key features that correlate with increased awareness.

ML algorithms, such as *regression* and *classification models* are used to provide insights into which factors contribute most to campaign success. For the above implementation, we make use of machine learning algorithms as well as Python libraries such as Pandas, matplotlib and seaborn.

4.Visualisation Design

Visualisation design involves dashboard design, temporal analysis, demographics breakdown and comparison of visuals. (All four aspects a=have been discussed in the Phase 1 document)

Line charts, bar charts, pie charts, histograms are used for effective representation of data. Factors such as Clarity and simplicity are kept in

mind in order to convey the information effectively to the target audience. IBM Cognos is used for competent implementation of this step.

5. Code for Data Analysis:

We use coding languages like Python (with libraries such as Pandas, NumPy, and Matplotlib/Seaborn) or R for in-depth data analysis. This might be necessary for advanced statistical modeling or custom analysis not easily achievable with point-and-click tools.

Exploratory Data Analysis (EDA)

EDA is the initial phase of data analysis. It involves getting to know your data, identifying any initial patterns, and understanding its structure. Key tasks in EDA include:

➢ Displaying the first few rows of data to understand its format.

➢ Getting an overview of the dataset to check for missing values and data types.

➢ Computing summary statistics to understand the distribution of data.

➢ Creating data visualizations to identify relationships and trends within the data.

Statistical Analysis

➢ Hypothesis testing: Determining if there are statistically significant differences between groups or variables.

➢ Regression analysis: Establishing relationships between variables and making predictions.

➢ Clustering or classification: Grouping data points based on similarity or assigning labels to data points.

6. Feedback and Iteration:

➢ Encourage stakeholders to ask questions, express concerns, and share their thoughts on the analysis and its implications.

➤ Pay close attention to the questions and concerns raised by stakeholders.

➤ Share the revised analysis and visualizations with stakeholders, highlighting the changes and improvements made based on their feedback.

7. Documentation:

Document your methodology, data sources, analysis techniques, and key findings. This documentation will be crucial for sharing insights with stakeholders and for future reference.

Tools and Technologies:

1. IBM Cognos:

IBM Cognos is a business intelligence and analytics platform that allows organizations to create interactive dashboards and reports.

➤ Data Integration: Begin by connecting Cognos to your data sources, whether they are databases, spreadsheets, or other data stores.

➤ Dashboard Creation: Cognos provides a user-friendly interface for building interactive dashboards.

➤ Data Exploration: Cognos offers tools for exploring and visualizing data, making it easier to identify patterns and trends in your dataset.

2. Python or R:

Python and R are powerful programming languages used for advanced data analysis and statistical modeling.

Using these tools and technologies effectively enhance your data analysis and project collaboration, making the entire process more efficient and productive.

TEAM MEMBERS:

1. RUKSHANA J - 2021115086

2. SWATHI K - 2021115116

3. THIRUMURUGAN K - 2021115117

4. VASUMATHY V - 2021115118

5. KAVIRAM R - 2021115332