**Public health care awareness**

**Phase 1**

**Problem Definition**

The problem we aim to address is the need to assess the effectiveness of public health awareness campaigns in reaching their intended audience, increasing awareness, and making data-driven decisions to enhance their impact. This project recognizes the challenges faced by public health campaigns, including low awareness levels, difficulty in reaching diverse audiences, and the need for improved communication strategies. Design thinking is a human-centred approach to innovation that focuses on understanding the needs of users and developing solutions that meet those needs. Data science is a field of study that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data. The goal is to create distinct customer segments based on relevant attributes and behaviours, allowing for personalized interactions and targeted campaigns.

**Design Thinking**

**Data Collection and Integration:**

Gather and integrate comprehensive and relevant data is crucial for analyzing the effectiveness of public health awareness campaigns. Here, we identify sources and methods for gathering campaign data, including engagement metrics, audience demographics, and awareness surveys Ensure data accuracy, consistency, and compliance with privacy regulations.

**Data Pre-processing and Cleansing:**

Cleanse and pre-process the collected data, handling missing values, outliers, and data quality issues. Normalize and standardize the data for consistent analysis. Modifying the data to be suitable for the model we about to use to segregate the customers.

**Feature Engineering:**

Identify relevant features and create new variables that may enhance the model's predictive capabilities. Incorporate external data sources and demographic information, if applicable.

**Clustering Algorithms:**

Develop a preliminary framework based on the identified data points. Code can be used to segment the audience or campaigns into distinct groups based on various attributes. This segmentation can help identify patterns and target specific subgroups more effectively.Utilize data science tools and techniques such as clustering algorithms (e.g., k-means, hierarchical clustering) to create initial customer segments. Validate the prototype with a subset of customer data to ensure its effectiveness.

**Visualization:**

While visualization tools like IBM Cognos are excellent for creating dashboards, code can help automate the creation of dynamic and interactive visualizations. You can generate plots, charts, and graphs programmatically.

**Model Validation and Performance Metrics:**

Establish validation procedures to assess the accuracy and reliability of the predictive models. Define appropriate performance metrics, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), or R-squared, for model evaluation.

**Real-time Data Updates:**

Implement mechanisms for real-time or periodic data updates to ensure that the predictive models stay current and relevant.

**User Interface and Reporting:**

Create a user-friendly interface for stakeholders to access and interact with the predictive system.

Generate reports and visualizations to communicate predictions and insights effectively.

**Scalability and Efficiency:**

Design the system to handle large volumes of data efficiently and accommodate scalability as data grows over time.

**Feedback Loop:**

Establishing a feedback loop with users to collect ongoing input and make iterative improvements to the system.

**Interpretation and Analysis:** Observe patterns, clusters, or groupings in the reduced-dimensional space. Analyze the relationships between data points and identify any trends or anomalies. Optionally, color-code points based on known customer attributes (e.g., demographics) to gain further insights.

**Phase 1 Deliverables:**

For Phase 1, the following deliverables are expected:

**Clean and Preprocessed Dataset:**

Provide a dataset where duplicates have been removed, missing values have been addressed, and categorical variables have been encoded. This dataset should be ready for use in machine learning models.

**List of Engineered Features:**

Document all the newly created features along with their descriptions and how they were calculated. This list should make it clear how each engineered feature adds value to the predictive model.

**Data Preprocessing and Feature Engineering Report:**

Create a detailed report that outlines the steps taken during data preprocessing and feature engineering. This report should cover the methods used, challenges encountered, and the rationale behind the decisions made at each step.

**Conclusion:**

In Phase 1, By integrating code into these aspects of your analysis, you can improve efficiency, accuracy, and the depth of insights obtained from the public health awareness campaign data. It allows for more advanced analysis techniques and automation, which are particularly beneficial when dealing with large and complex datasets.