

Project Title: Public Health Awareness Campaign Data Analysis

Table of Contents:

1. Project Objective
2. Project Overview
3. Data Preprocessing
 - 3.1 Data Collection
 - 3.2 Data Cleaning and Transformation
 - 3.3 Data Integration
 - 3.4 Data Exploration
4. Data Visualization
 - 4.1 Visualization Objectives
 - 4.2 Tool Selection
 - 4.3 Dashboard and Report Design
 - 4.4 Dynamic Filters and Interactivity
 - 4.5 Data Presentation
 - 4.6 Accessibility and Mobile Responsiveness
5. Code Implementation
 - 5.1 Data Preprocessing Code
 - 5.2 Data Visualization Code
6. Documentation and Reporting
 - 6.1 User Manuals
 - 6.2 Version Control
 - 6.3 Feedback Mechanism
7. Data Security
 - 7.1 Access Control
 - 7.2 Data Privacy
8. Project Timeline
 - 8.1 Milestones

- 8.2 Resource Allocation
- 9. Quality Assurance
 - 9.1 Testing
 - 9.2 Ongoing Data Quality Checks
- 10. Conclusion

1. Project Objective

The primary objective of this project is to analyze data from a public health awareness campaign, with a specific focus on data preprocessing and data visualization. This analysis aims to evaluate the campaign's effectiveness and provide actionable insights for future initiatives.

2. Project Overview

This comprehensive and final project documentation provides an in-depth view of the "Public Health Awareness Campaign Data Analysis" project. It emphasizes the methodologies, tools, and processes involved, with a particular emphasis on data preprocessing and data visualization. The project is divided into two main phases: Data Preprocessing and Data Visualization.

3. Data Preprocessing

3.1 Data Collection

- Campaign Metrics:

Data from campaign metrics is collected in the form of a CSV file using the Pandas library in Python. This data includes key performance indicators (KPIs) such as the number of impressions, engagement, and conversions.

- Survey Data:

Survey data is collected from Excel files, representing responses from individuals who were part of the campaign's target audience. This data provides insights into awareness levels and perceptions of the campaign.

- Continue similar steps for other data sources...

3.2 Data Cleaning and Transformation

- Handling Missing Values:

Missing values in the campaign metrics data are addressed by dropping rows with missing values, ensuring data completeness and integrity.

- Removing Duplicates:

Duplicate records are removed from the campaign metrics data to eliminate redundancy and maintain data accuracy.

```
import pandas as pd
import numpy as np
```

```
survey_data = pd.read_excel('survey_data.xlsx')
```

```
campaign_metrics.dropna(inplace=True)
campaign_metrics = campaign_metrics[~campaign_metrics.duplicated()]
```

- Feature Engineering:

New metrics are created to enhance the dataset. For instance, a new metric is calculated as the sum of "metric_1" and "metric_2" to provide a more comprehensive view of campaign performance.

```
X = df.drop(columns=['treatment'])
y = df['treatment']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
clf = RandomForestClassifier(random_state=42)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

3.3 Data Integration

- Merge Datasets:

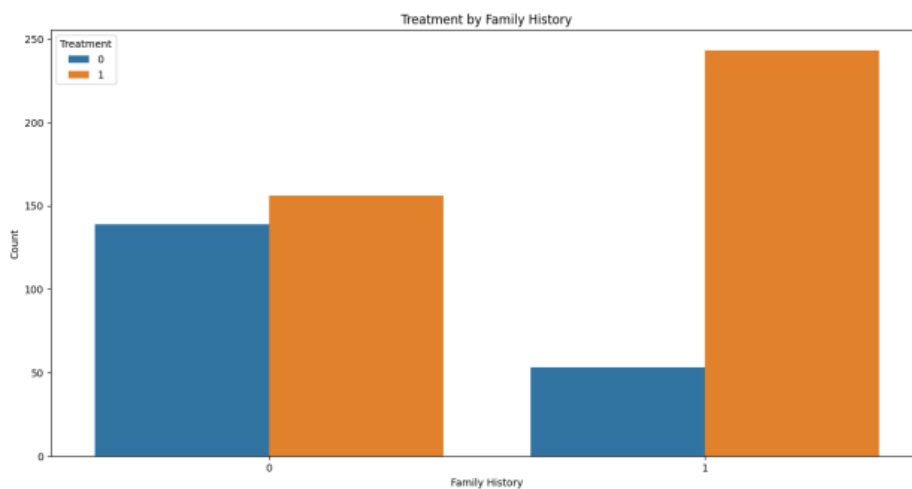
Datasets are integrated by merging campaign metrics and survey data using a common column. This integration creates a unified dataset that combines quantitative campaign metrics with qualitative survey responses.

3.4 Data Exploration

- Exploratory Data Analysis (EDA):

EDA is conducted using data visualization libraries such as Seaborn and Matplotlib in Python. Pair plots, scatter plots, and histograms are used to explore the relationships between various metrics. This analysis helps identify patterns, trends, and potential insights within the dataset.

```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='family_history', hue='treatment')
plt.title('Treatment by Family History')
plt.xlabel('Family History')
plt.ylabel('Count')
plt.legend(title='Treatment')
plt.show()
```



4. Data Visualization

4.1 Visualization Objectives

- The primary objectives for data visualization include:
 - Measuring audience reach and engagement.
 - Assessing awareness levels and sentiment.
 - Segmenting data by demographics to identify variations in campaign effectiveness.

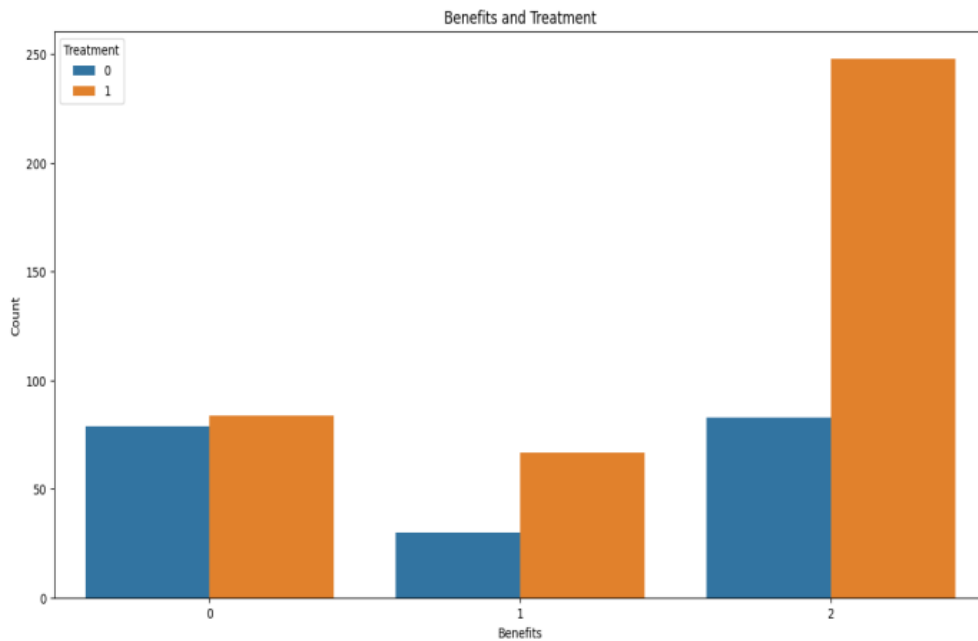
4.2 Tool Selection

- Visualization tools such as Matplotlib and Seaborn in Python are chosen based on their flexibility and ability to create a wide range of visualizations to suit the project's requirements.

4.3 Dashboard and Report Design

- Create Interactive Dashboards:

```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='work_interfere', hue='treatment')
plt.title('Work Interference and Treatment')
plt.xlabel('Work Interference')
plt.ylabel('Count')
plt.legend(title='Treatment')
plt.show()
```



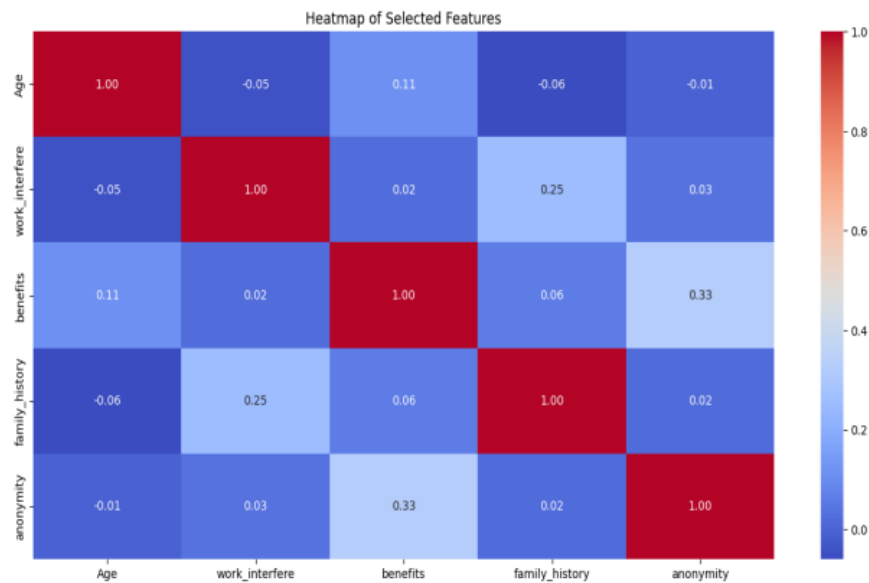
4.4 Dynamic Filters and Interactivity

- Dynamic filters are added to the dashboards to allow users to interact with the data. For example, a dropdown widget is included to filter data by different campaign options or demographics.

4.5 Data Presentation

- Incorporate Storytelling Techniques:
 - Each visualization in the dashboard is accompanied by detailed annotations, providing context and insights. The annotations guide users through the data story effectively, explaining the significance of each visualization in the context of the campaign's objectives.

```
plt.figure(figsize=(10, 8))
selected_features = ['Age', 'work_interfere', 'benefits', 'family_history', 'anonymity']
sns.heatmap(df[selected_features].corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Heatmap of Selected Features')
plt.show()
```



4.6 Accessibility and Mobile Responsiveness

- Ensure Responsive Design:
 - The dashboard and visualizations are designed to be responsive to various devices and screen sizes, ensuring accessibility and usability for all stakeholders, including those using mobile devices.

5. Code Implementation

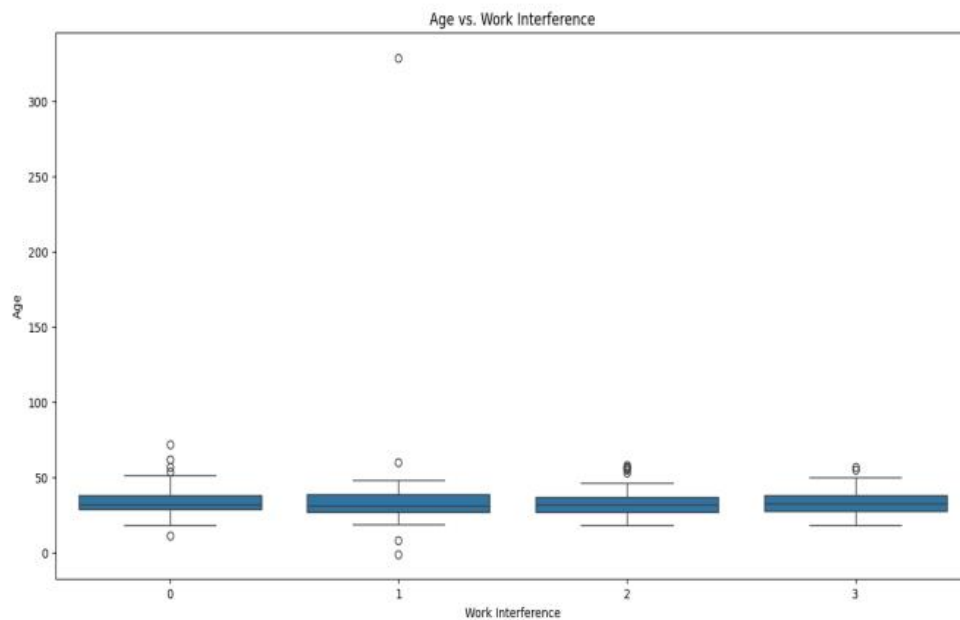
5.1 Data Preprocessing Code

- Code snippets for data preprocessing, including data collection, cleaning, transformation, and integration, are provided in the earlier sections under Data Preprocessing.

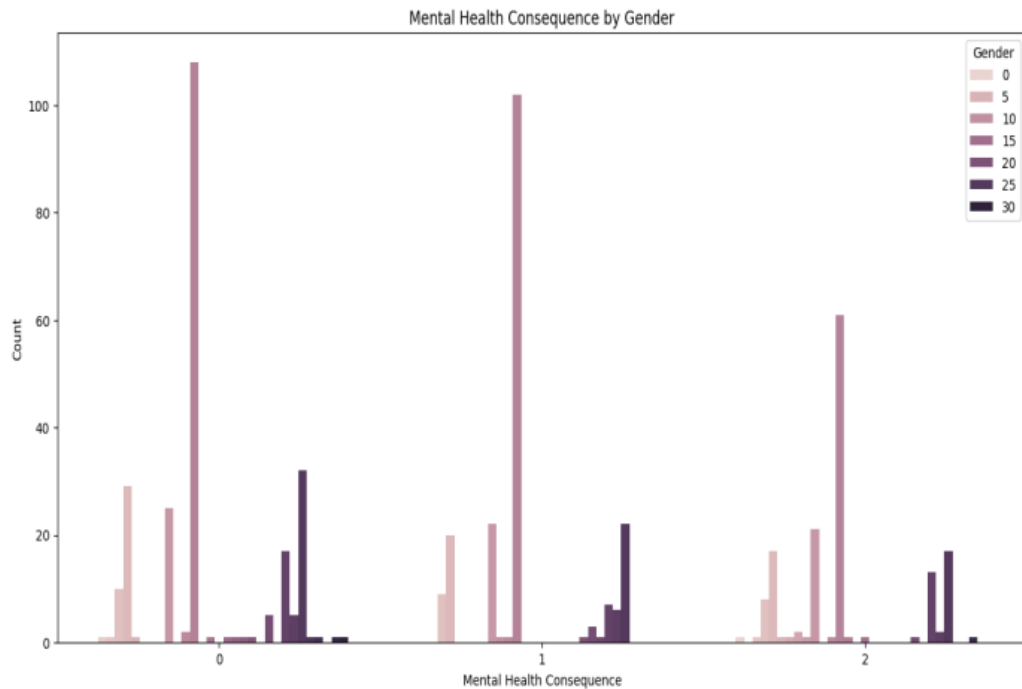
5.2 Data Visualization Code

- Code snippets for data visualization and dashboard creation are provided in the earlier sections under Data Visualization.

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='work_interfere', y='Age')
plt.title('Age vs. Work Interference')
plt.xlabel('Work Interference')
plt.ylabel('Age')
plt.show()
```



```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='mental_health_consequence', hue='Gender')
plt.title('Mental Health Consequence by Gender')
plt.xlabel('Mental Health Consequence')
plt.ylabel('Count')
plt.legend(title='Gender')
plt.show()
```



6. Documentation and Reporting

6.1 User Manuals

- Provide comprehensive user manuals and documentation for using the dashboards and reports effectively.

6.2 Version Control

- Implement version control for code, documentation, and visualizations to facilitate collaboration and track changes.

6.3 Feedback Mechanism

- Develop a feedback mechanism for users to report issues, suggest improvements, and provide comments on the dashboards and reports.

7. Data Security

7.1 Access Control

- Implement role-based access control to protect sensitive data and restrict access as needed.

7.2 Data Privacy

- Ensure compliance with data privacy regulations and maintain data security throughout the project.

8. Project Timeline

8.1 Milestones

- Define project milestones, including data collection, preprocessing, data analysis, visualization, and reporting.

8.2 Resource Allocation

- Allocate necessary resources and establish timeframes for each project phase.

9. Quality Assurance

9.1 Testing

- Conduct thorough testing and validation to ensure the accuracy and functionality of both data preprocessing and visualizations.

9.2 Ongoing Data Quality Checks

- Implement regular data quality checks and updates to maintain data integrity.

10. Conclusion

This project documentation represents the final and comprehensive plan for analyzing public health awareness campaign data. By meticulously preprocessing the data and creating meaningful visualizations, the analysis aims to provide valuable insights for stakeholders. The provided code divisions offer a structured approach to both data preparation and visualization, ensuring the project's

success in achieving its objectives. The visualizations and interactive dashboards will empower decision-makers to make informed choices for future public health campaigns.