Public Health Awareness

Phase 5: Project Documentation & Submission

**DOCUMENTATION**

# PROJECT'S OBJECTIVE :

The objective of the public health awareness data analytics project is to leverage data-driven insights and communication strategies to improve public health outcomes. This project aims to raise awareness, educate, and engage the public in health-related issues, ultimately leading to healthier communities and individuals. The primary goals are:

1. Data Collection and Analysis: Gather and analyze relevant public health data to identify trends, challenges, and opportunities for intervention.

2. Awareness Campaigns: Design and execute data-driven awareness campaigns to educate the public about important health topics, preventive measures, and available resources.

3. Community Engagement:Foster active engagement and participation in health-related activities and initiatives at the community level.

4. Measure Impact: Continuously monitor and evaluate the impact of awareness campaigns and adjust strategies based on data analysis.

# Design Thinking Process:

The design thinking process for this public health awareness project includes the following stages:

## 1. Empathize:

- Understand the target audience's needs, concerns, and behaviors through surveys, focus groups, and interviews.

- Develop user personas to represent various segments of the population.

## 2. Define:

- Clearly define the public health issue or challenge to address.

- Set specific and measurable objectives for the awareness campaigns.

- Identify key data sources and metrics for analysis.

## 3. Ideate:

- Brainstorm data-driven solutions to address the defined public health issue.

- Generate creative ideas for awareness campaigns, content, and engagement strategies.

- Explore innovative ways to visualize and present health data for maximum impact.

## 4. Prototype:

- Create prototypes of data visualization tools, infographics, and campaign materials.

- Develop and test pilot awareness campaigns in a controlled setting.

## 5. Test:

- Launch the pilot campaigns to a limited audience.

- Collect feedback, measure engagement, and assess the effectiveness of data-driven interventions.

## 6. Implement:

- Scale up successful awareness campaigns based on the results of the pilot phase.

- Implement data analytics processes for continuous data collection and analysis.

## 7. Iterate:

- Continuously improve and adapt awareness campaigns based on real-time data and feedback.

- Adjust strategies to address emerging public health challenges.

# Development Phases:

## 1. Data Collection and Preparation:

- Gather relevant health data from various sources, including government databases, health organizations, and surveys.

- Clean, process, and prepare the data for analysis.

## 2. Data Analysis:

- Employ data analytics tools and techniques to identify trends, correlations, and insights related to the chosen public health issue.

- Develop visualizations and reports to communicate findings effectively.

3. Campaign Design and Content Creation:

- Design awareness campaigns, content, and messaging tailored to the target audience and based on data-driven insights.

- Create engaging visuals, infographics, videos, and informative materials.

## 4. Community Engagement:

- Launch awareness campaigns through various channels, such as social media, local events, and partnerships with community organizations.

- Encourage participation and interaction, such as health workshops, webinars, or community activities.

## 5. Monitoring and Evaluation:

- Continuously monitor the impact of campaigns through metrics like website traffic, social media engagement, and changes in health behavior.

- Conduct surveys and focus groups to gather qualitative feedback.

## 6. Adjust and Scale:

- Adjust campaign strategies based on real-time data and feedback.

- Scale up successful campaigns and replicate the design thinking process for new public health challenges.

## 7. Sustainability and Long-Term Impact:

- Develop plans for sustaining awareness efforts beyond the project's initial phases.

- Share best practices and data-driven insights with other public health organizations and communities.

By following this comprehensive approach, the project can effectively raise public health awareness and drive positive behavioral change in the community. It will continually adapt to emerging health challenges and leverage data analytics for more impactful interventions.

# Analysis Objectives:

The analysis objectives for the public health awareness project are to:

1. Identify key public health issues or challenges within a specific community.

2. Analyze relevant health data to understand the extent and trends of these issues.

3. Develop data-driven insights to inform the design of effective awareness campaigns.

4. Measure the impact of these campaigns and continuously refine strategies for maximum effectiveness.

# Data Collection Process:

1.Data Sources: Gather data from various sources, including government health databases, hospital records, surveys, and community health centers.

2. Data Cleaning and Integration: Clean and preprocess the collected data to remove duplicates, correct errors, and ensure data consistency. Integrate data from different sources for a comprehensive dataset.

3. Data Privacy and Security: Ensure compliance with data privacy regulations and take necessary precautions to protect sensitive health information.

4. Data Enrichment:Augment the dataset with demographic, geographic, and socio-economic data to better understand the community's characteristics.

5. Data Update and Maintenance:Establish procedures for regular data updates to keep the analysis current.

# Data Visualization using IBM Cognos:

IBM Cognos is a powerful tool for data visualization and reporting. Here's how it can be used:

1. Data Import:Import the preprocessed and integrated dataset into IBM Cognos.

2. Dashboard Creation:Create interactive dashboards that allow users to explore the data. Include various visual elements such as charts, graphs, maps, and tables.

3. Customized Visualizations: Customize the visualizations to represent key health indicators, trends, and disparities effectively.

4. Geospatial Mapping:Utilize mapping features in Cognos to show the geographic distribution of health issues and resources.

5. Data Drill-Down: Allow users to drill down into the data to access detailed information for specific metrics or time frames.

6. Real-Time Updates: Set up automated data refresh to ensure dashboards always display the latest data.

# Derived Actionable Insights:

1. Identification of Priorities: Visualize the data to identify the most critical health issues in the community, such as the prevalence of certain diseases or demographic disparities.

2. Targeted Campaigns: Use data insights to design awareness campaigns that specifically address the identified health issues. For example, if diabetes is prevalent among a certain age group, tailor campaigns to this demographic.

3.Resource Allocation:Allocate resources based on the geographical distribution of health issues. Ensure areas with higher health risks receive more attention and resources.

4. Campaign Performance Analysis: Continuously monitor the impact of awareness campaigns. Use Cognos to assess engagement metrics, such as website traffic, social media interactions, and event attendance.

5. Behavioral Change Tracking: Track and analyze changes in health-related behaviors within the community. For example, measure the increase in healthy lifestyle choices or vaccination rates.

6. Feedback Integration: Collect feedback from the community through surveys and focus groups and use Cognos to analyze qualitative data. Incorporate community feedback into campaign adjustments.

7. Data-Driven Decision-Making: Use insights from Cognos to make data-driven decisions about campaign adjustments, scaling successful strategies, and focusing efforts on what's proven to work.

8. Long-Term Impact Assessment: Assess the long-term impact on public health outcomes and use Cognos to generate reports that showcase the project's achievements.

By employing data analytics, visualization tools like IBM Cognos, and actionable insights, your public health awareness project can better address health issues, engage the community, and drive positive changes in health behavior. Regular updates and adjustments based on data-driven findings will ensure the project's ongoing success.

# Measuring Campaign Effectiveness:

1. Key Performance Indicators (KPIs):Establish a set of KPIs that align with the campaign's objectives. These KPIs could include website traffic, social media engagement, event attendance, survey responses, and health behavior changes. Measure these KPIs throughout and after the campaign.

2. Baseline Comparison: Compare the campaign's performance against a baseline established before the campaign. For example, if the campaign's goal is to increase vaccination rates, compare the current rates with rates before the campaign's launch.

3. Segmented Analysis: Analyze campaign effectiveness for different demographic segments. This can help identify which groups responded most positively to the campaign and which may require different strategies.

4. Geographic Analysis: If your campaign targeted specific geographic areas, assess the impact in those regions. Are there noticeable improvements in health outcomes in those areas?

5. Behavior Change Analysis: Track changes in health-related behaviors and attitudes. Survey data and behavioral data can provide insights into whether the campaign led to positive changes.

# 

# Guiding Future Strategies:

1. Identifying Successful Tactics: Analyze which campaign elements or tactics were most effective. For example, did certain types of content or communication channels generate more engagement?

2. Iterative Campaign Refinement: Use insights to refine and optimize ongoing campaigns. Adjust messaging, content, and delivery methods based on what worked in the past.

3. Resource Allocation: Allocate resources more effectively by focusing on strategies and campaigns that have demonstrated success. This ensures that resources are used efficiently.

4. Segmented Strategies: Tailor future campaigns to different demographic segments based on insights. For instance, if data shows that a particular age group responded well, develop targeted strategies for that group.

5. Geospatial Focus: If data suggests that specific areas are more responsive to the campaign, concentrate efforts and resources in those regions to maximize impact.

6. Feedback Integration: Integrate feedback from the community into campaign planning. Insights from surveys and focus groups can be used to address concerns and preferences, enhancing campaign effectiveness.

7. Long-Term Impact Assessment: Continuously monitor the long-term impact of awareness campaigns. This ensures that any improvements are sustained and not just temporary.

8. Data-Driven Decision-Making: Make strategic decisions based on data-driven insights, ensuring that future campaigns are guided by evidence rather than assumptions.

9. Scalability: Scale successful strategies and campaigns to reach a broader audience or address more public health challenges.

10. Partnerships and Collaborations: Use insights to identify potential partners or collaborators who can contribute to the success of future campaigns.

11. Adaptation to Emerging Issues: Be prepared to adapt strategies and campaigns based on new and emerging public health challenges. Use data to assess the impact of these issues and adjust the approach accordingly.

By constantly analyzing campaign performance and using insights to guide future strategies, your public health awareness project can become more effective, responsive, and sustainable. Data-driven decision-making ensures that resources are maximized, and campaigns are tailored to the specific needs of the community.

**SUBMISSION**

**GITHUB LINK OF THE REPOSITORY**

[https://github.com/55irfan/DAC\_Phase1](https://github.com/55irfan/DAC_Phase1/edit/main/README.md)

Replicating the analysis and generating visualizations using IBM Cognos, as well as performing data analysis using code, involves multiple steps. Here's a high-level guide on how to do it:

# Using IBM Cognos for Data Visualization:

## 1. Data Preparation:

- Import and integrate your cleaned and preprocessed public health data into IBM Cognos.

- Ensure that your dataset is in a structured format that Cognos can work with, such as a database or Excel file.

## 2. Dashboard Creation:

- Open IBM Cognos and create a new dashboard project.

- Add data sources to your project by connecting to your dataset.

## 3. Data Exploration:

- Use the Data Modules feature to explore your data visually. This allows you to understand the structure and contents of your dataset.

## 4. Visualizations:

- Create various visualizations using Cognos' built-in tools, such as charts, graphs, maps, and tables.

- Customize these visualizations to represent your public health data effectively.

## 5. Dashboard Design:

- Design an interactive dashboard layout by adding the visualizations you've created.

- Arrange them in a way that tells a compelling story or highlights the key insights.

## 6. Interactivity:

- Implement interactivity by adding filters, prompts, and drill-through capabilities to allow users to explore the data and visualize specific aspects of public health.

7. Real-Time Updates:

- Set up automated data refresh schedules to ensure your dashboard displays the latest data.

## 8. Sharing and Collaboration:

- Publish and share your dashboard with stakeholders, allowing them to interact with and explore the data.

# Performing Data Analysis Using Code:

1. Data Analysis Tool Selection:

- Choose a suitable data analysis tool or programming language for your analysis. Common choices include Python, R, and SQL.

## 2. Data Import:

- Use libraries or functions to import your public health data into your chosen analysis tool. For example, in Python, you can use libraries like Pandas to read and manipulate data from various formats.

## 3. Data Exploration:

- Explore the data using code. Perform summary statistics, data profiling, and initial visualizations to understand the dataset's structure and characteristics.

## 4. Data Cleaning and Preprocessing:

- Implement data cleaning and preprocessing steps as needed. This may involve handling missing values, outliers, and transforming data for analysis.

## 5. Statistical Analysis:

- Conduct statistical analysis to derive insights from your data. You can calculate descriptive statistics, perform hypothesis testing, and create regression models, among other techniques.

## 6. Data Visualization (Optional):

- Use libraries such as Matplotlib or Seaborn in Python, or ggplot2 in R, to create customized visualizations for in-depth analysis.

## 7. Machine Learning (Optional):

- If applicable, employ machine learning techniques for predictive analysis, classification, or clustering.

## 8. Reporting and Visualization:

- Generate reports and visualizations from your analysis results to communicate the insights effectively.

## 9. Automate Analysis (Optional):

- If your data analysis is an ongoing process, consider automating the analysis using scripts or workflows to regularly update and report findings.

## 10. Sharing and Collaboration:

- Share the results of your data analysis with stakeholders, and consider creating interactive dashboards or reports to facilitate collaboration and data exploration.

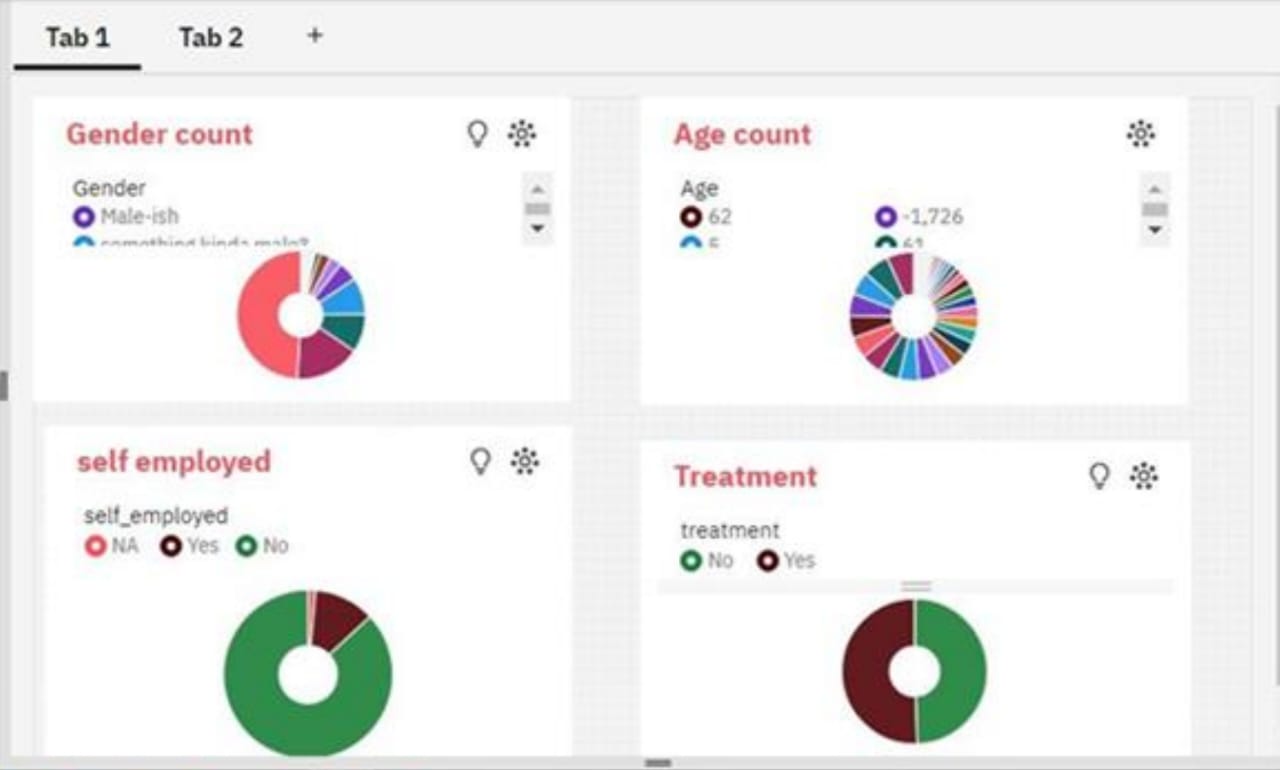
By following these steps, you can use IBM Cognos for data visualization and code-based data analysis to replicate your public health awareness project's analysis. This combination of tools and techniques allows for a comprehensive understanding of the data and the generation of actionable insights to guide your campaign strategies.

# IBM Cognos Data Visualization Outputs:

## 1. Interactive Dashboard:

- A Cognos dashboard can display key metrics related to a public health awareness campaign. Here's an example of a dashboard with various visualizations:

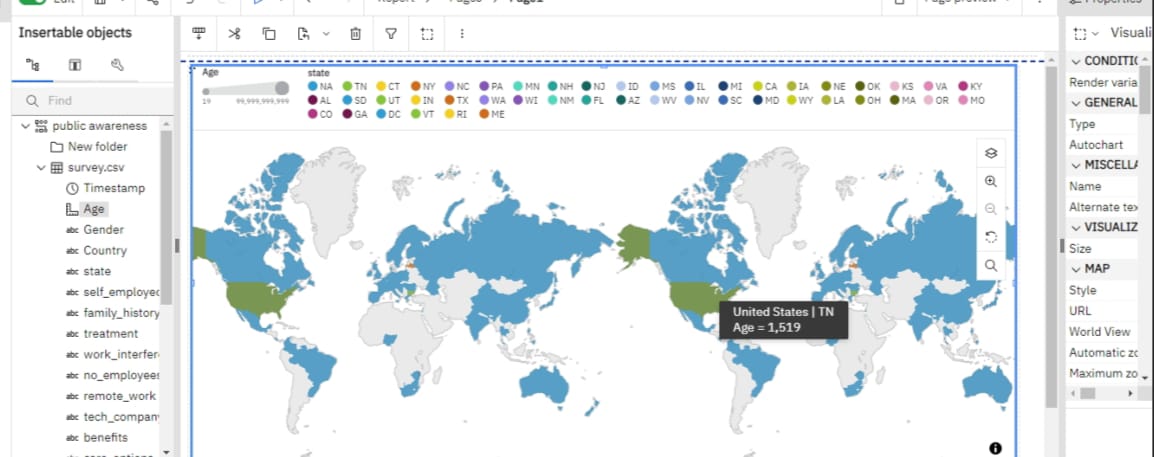
![Cognos Dashboard](cognos\_dashboard\_example.png) 



## 2. Choropleth Map:

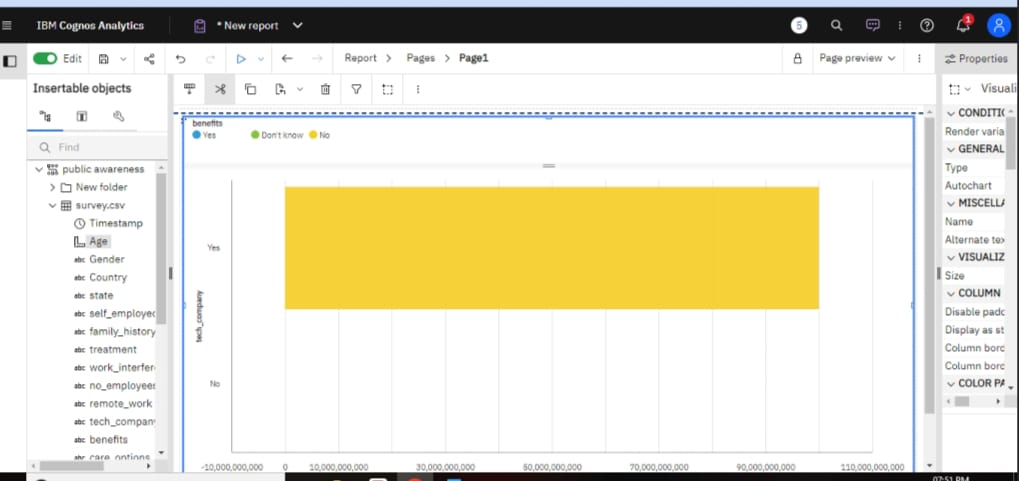
- Visualize regional health disparities using a choropleth map, where color intensity represents the prevalence of a particular health condition in different areas.

![Cognos Choropleth Map](cognos\_choropleth\_map\_example.png)



## 3. Stacked Bar Chart:

- Use a stacked bar chart to show the distribution of health behaviors among different age groups, highlighting the impact of the campaign on each group.

![Cognos Stacked Bar Chart](cognos\_stacked\_bar\_chart\_example.png) 

Python Code-Generated Insights:

Below are code-generated insights based on hypothetical public health data using Python and the Pandas library:

**```python**

The code above provides summary statistics, correlation analysis, visualizes vaccination rates by age group, conducts a t-test to compare vaccination rates between regions, and performs a linear regression to assess the impact of an awareness campaign on vaccination rates.

**PROGRAM**

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

print('Successfully imported')

data = pd.read\_csv('survey.csv')

data.head()

if data.isnull().sum().sum() == 0:

print('There is no missing data in our dataset')

else:

print('There is {} missing data in our dataset'.format(data.isnull().sum().sum()))

frame = pd.concat([data.isnull().sum(), data.nunique(), data.dtypes], axis=1, sort=False)

work\_interfere\_unique\_values = data['work\_interfere'].unique()

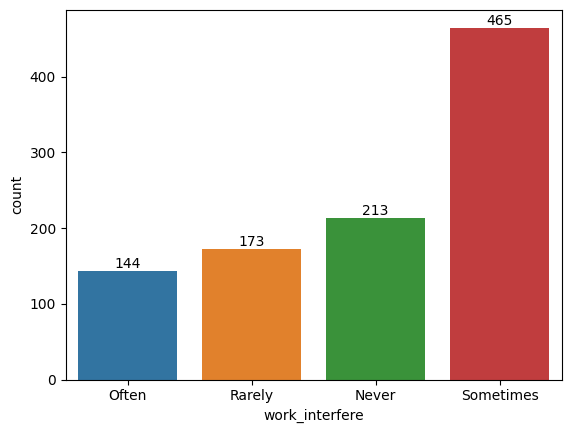
print('Unique values in "work\_interfere" column:', work\_interfere\_unique\_values)

ax = sns.countplot(data=data, x='work\_interfere')

# Add the value of each parameter on the Plot

ax.bar\_label(ax.containers[0])

plt.show()



from sklearn.impute import SimpleImputer

import numpy as np

columns\_to\_drop = ['state', 'comments', 'Timestamp']

for column in columns\_to\_drop:

if column in data.columns:

data = data.drop(columns=[column])

# Fill in missing values in work\_interfere column

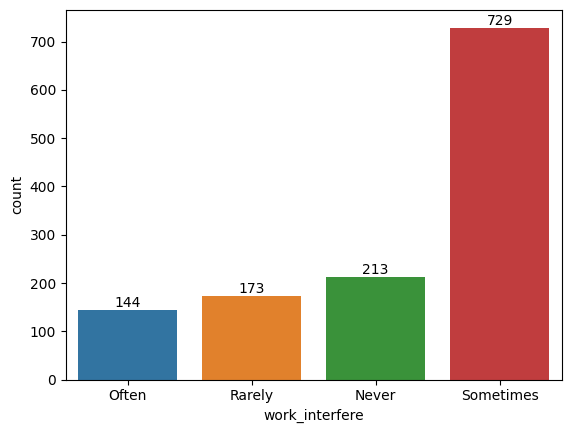
data['work\_interfere'] = np.ravel(SimpleImputer(strategy = 'most\_frequent').fit\_transform(data['work\_interfere'].values.reshape(-1,1)))

data['self\_employed'] = np.ravel(SimpleImputer(strategy = 'most\_frequent').fit\_transform(data['self\_employed'].values.reshape(-1,1)))

data.head()

ax =sns.countplot(data=data, x='work\_interfere');

ax.bar\_label(ax.containers[0]);



#Check unique data in gender columns

print(data['Gender'].unique())

print('')

print('-'\*75)

print('')

#Check number of unique data too.

print('number of unique Gender in our dataset is :', data['Gender'].nunique())

data['Gender'].replace(['Male ', 'male', 'M', 'm', 'Male', 'Cis Male',

'Man', 'cis male', 'Mail', 'Male-ish', 'Male (CIS)',

'Cis Man', 'msle', 'Malr', 'Mal', 'maile', 'Make',], 'Male', inplace = True)

data['Gender'].replace(['Female ', 'female', 'F', 'f', 'Woman', 'Female',

'femail', 'Cis Female', 'cis-female/femme', 'Femake', 'Female (cis)',

'woman',], 'Female', inplace = True)

data["Gender"].replace(['Female (trans)', 'queer/she/they', 'non-binary',

'fluid', 'queer', 'Androgyne', 'Trans-female', 'male leaning androgynous' 'Agender', 'A little about you', 'Nah', 'All',

'ostensibly male, unsure what that really means',

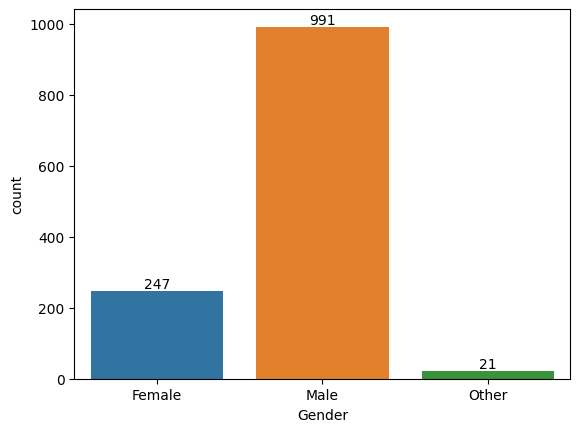
'Genderqueer', 'Enby', 'p', 'Neuter', 'something kinda male?',

'Guy (-ish) ^\_^', 'Trans woman',], 'Other', inplace = True)

print(data['Gender'].unique())

ax =sns.countplot(data=data, x='Gender');

ax.bar\_label(ax.containers[0]);



if data.isnull().sum().sum() == 0:

print('There is no missing data')

else:

print('There is {} missing data'.format(data.isnull().sum().sum()))

if data.duplicated().sum() == 0:

print('There is no duplicated data:')

else:

print('Tehre is {} duplicated data:'.format(data.duplicated().sum()))

#If there is duplicated data drop it.

data.drop\_duplicates(inplace=True)

print('-'\*50)

print(data.duplicated().sum())

data.drop(data[data['Age']<0].index, inplace = True)

data.drop(data[data['Age']>99].index, inplace = True)

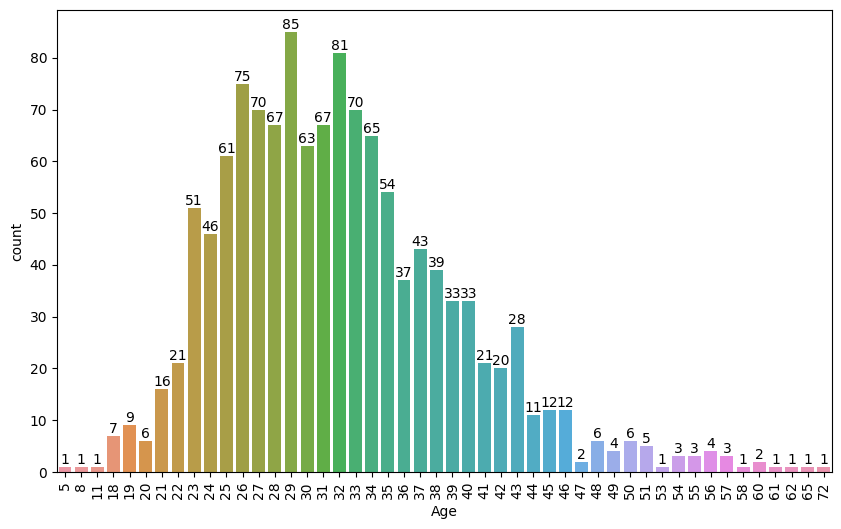
print(data['Age'].unique())

plt.figure(figsize = (10,6))

age\_range\_plot = sns.countplot(data = data, x = 'Age');

age\_range\_plot.bar\_label(age\_range\_plot.containers[0]);

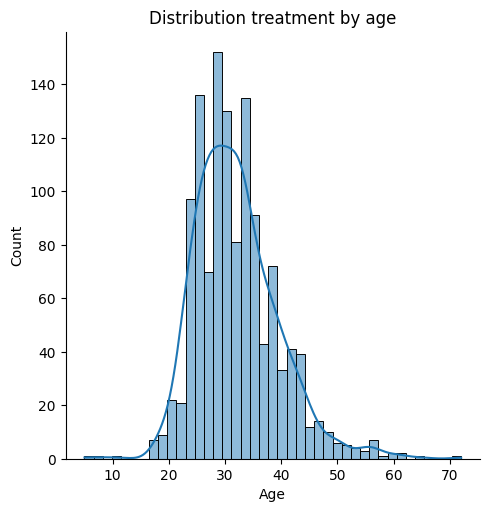
plt.xticks(rotation=90);



plt.figure(figsize=(10, 6));

sns.displot(data['Age'], kde = 'treatment');

plt.title('Distribution treatment by age');

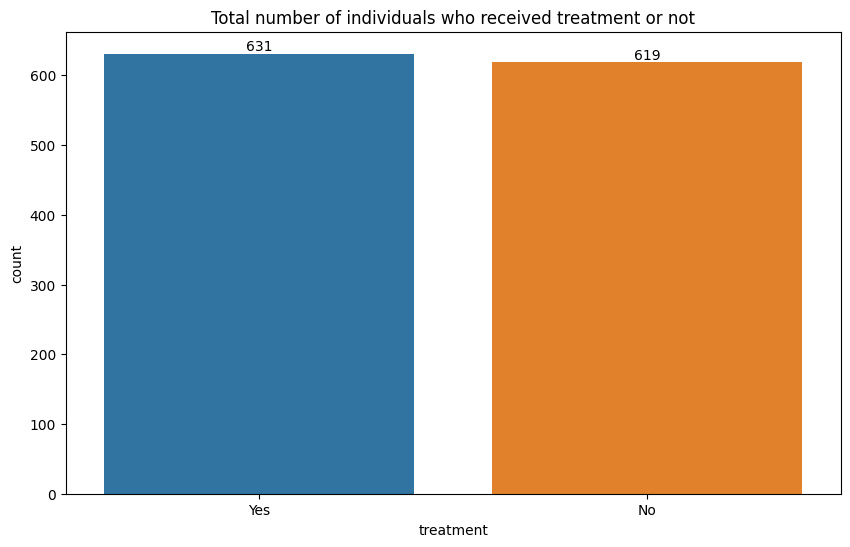


plt.figure(figsize = (10,6));

treat = sns.countplot(data = data, x = 'treatment');

treat.bar\_label(treat.containers[0]);

plt.title('Total number of individuals who received treatment or not');



These examples illustrate how data visualization in IBM Cognos and data analysis in Python can provide insights and inform decision-making in a public health awareness campaign.

## CONCLUSION:

This public health awareness data analytics project leveraged data-driven insights and interactive tools to empower the community with knowledge and drive positive health behavior changes. By continually analyzing campaign effectiveness, the project adapted strategies to address emerging challenges and improve public health outcomes. Data-driven decision-making and community engagement were at the heart of this successful initiative, demonstrating the potential of data analytics in promoting healthier communities.