Naan Mudhalvan

Data Analytics With Cognos

Phase-5

Team: Shivani Suresh, Shyamala R B, Daphny Jessica

Analysis:

1. Measure Audience Reach:
   - Objective: Quantify the extent to which the campaign reached the target audience.
   - Metrics: - Total impressions: The number of times campaign content was displayed.
   - Unique reach: The number of distinct individuals who saw the campaign.
   - Engagement rate: The percentage of reached individuals who interacted with the content (likes, shares, comments).

2. Assess Awareness Levels:
   - Objective: Evaluate the effectiveness of the campaign in increasing awareness about the public health issue.
   - Metrics:
   - Pre-campaign vs. post-campaign awareness levels: Conduct surveys or analyze search trends to measure changes in awareness.
   - Social media sentiment: Analyze sentiment in social media comments related to the campaign to gauge public sentiment.

3. Evaluate Campaign Impact:
   - Objective: Determine whether the campaign led to behavioral changes or contributed to improving the public health issue.
   - Metrics:
   - Behavioral changes: Measure changes in specific behaviors related to the campaign's message (e.g., increased vaccination rates, healthier lifestyle choices).
   - Health outcome indicators: Monitor relevant health indicators (e.g., disease incidence, hospital admissions) to assess the campaign's impact.

4. Analyze Audience Segmentation:
   - Objective: Understand how different demographic groups responded to the campaign.
   - Metrics:
   - Demographic breakdown: Analyze campaign engagement and awareness levels by age, gender, location, and other relevant demographics.

- Tailor messaging: Use insights to refine future campaign targeting and messaging.

5. Track Conversion and Call to Action:
- Objective: Measure how effective the campaign was in encouraging the audience to take desired actions (e.g., sign up for health programs, get vaccinated).
- Metrics:
- Conversion rate: The percentage of individuals who took the desired action after interacting with the campaign.
- Click-through rate (CTR): Measure the effectiveness of campaign calls to action.

6. Assess Cost-effectiveness:
- Objective: Evaluate the cost-effectiveness of the campaign in achieving its objectives.
- Metrics:
- Cost per impression (CPI): Calculate the cost incurred for each impression generated by the campaign.
- Cost per conversion (CPC): Calculate the cost incurred for each desired action taken as a result of the campaign.

7. Measure Long-term Impact:
- Objective: Determine if the campaign's effects endure over time.
- Metrics:
  - Long-term behavioral change: Monitor whether the campaign's impact on behavior change persists.
  - Sustained awareness: Assess whether the campaign's awareness gains are maintained in the long term.

8. Gather Qualitative Feedback:
- Objective: Collect qualitative feedback from the audience to understand their perceptions and opinions about the campaign.
- Metrics:
- Conduct focus groups or interviews to gather in-depth insights.
- Analyze user-generated content (e.g., social media comments, blog posts) for qualitative feedback.

Data Collection:

1. Engagement Metrics:
- Sources: Social media platforms (Facebook Insights, Twitter Analytics), website analytics tools (Google Analytics), email campaign tools.
- Methods: Collect data on likes, shares, comments, clicks, impressions, click-through rates, and time spent on campaign content.

2. Audience Demographics:
- Sources: Social media insights, website analytics, third-party data providers.
- Methods: Use built-in analytics tools to gather demographic information such as age, gender, location, and interests of the campaign's audience.

3. Awareness Surveys:
- Sources: Conduct online surveys through platforms like SurveyMonkey or Google Forms, or offline surveys at specific events.
- Methods: Design and distribute surveys to measure baseline and post-campaign awareness levels. Include questions related to campaign recall and effectiveness.

4. Health Outcome Data:
- Sources: Public health records, hospitals, clinics, government health agencies.
- Methods: Obtain data on health outcomes relevant to the campaign's objectives, such as vaccination records, disease incidence rates, and relevant health statistics.

5. Media Monitoring:
- Sources: Media monitoring tools and services.

- Methods: Monitor traditional media channels (TV, radio, newspapers) for mentions of the campaign and gather data on reach and sentiment.

6. Digital Analytics Integration:
- Use code (e.g., Python) to integrate and preprocess data from various sources, ensuring data consistency and accuracy.

7. Data Privacy Considerations:

- Ensure compliance with data privacy regulations (e.g., GDPR, HIPAA) when collecting and handling personal health-related data.

Problem Statement of Public Health Awareness Campaign Project

-Difficulty in assessing the potential impact of new campaigns.
-Limited understanding of which factors contribute to campaign success.
-Lack of data-driven insights to optimize campaign strategies.

Solution: Our solution involves implementing predictive analytics using machine learning to forecast the success of future public health awareness campaigns. This approach will enable campaign planners to make informed decisions, allocate resources efficiently, and maximize the impact of their initiatives.

Implementation Steps

1. Data Collection and Preparation

Collect historical campaign data, including engagement metrics, audience demographics, and awareness survey results.
Clean and preprocess the data to remove outliers and handle missing values.
Ensure data privacy compliance and obtain necessary consents when collecting personal data.

2. Feature Engineering

Identify relevant features that may influence campaign success (e.g., engagement metrics, demographics, campaign duration, messaging).
Engineer new features if necessary, such as sentiment analysis of user-generated content.

3. Model Selection

Choose appropriate machine learning algorithms for prediction (e.g., regression, classification, time series forecasting). Here we have chosen the Random Forest Algorithm.

4. Training and Validation
Split the historical data into training and validation sets.
Train the selected machine learning model on the training data and validate its performance on the validation set.
Evaluate model accuracy, precision, recall, and other relevant metrics.

5. Predictive Modeling

Once the model is trained and validated, use it to predict the success of future campaigns.
Input relevant campaign features into the model to obtain predictions.
Monitor model performance and update it regularly with new data.

6. Dashboard and Reporting

Create a user-friendly dashboard using IBM Cognos to visualize campaign predictions and historical data. Generate reports that provide insights into predicted campaign success factors.

Benefits

Data-Driven Decision-Making: Campaign planners can use predictive analytics to make informed decisions about resource allocation and campaign strategies.

Optimized Campaigns: By understanding which factors contribute to success, campaigns can be tailored for maximum impact.

Resource Efficiency: Avoid wasting resources on campaigns that are unlikely to succeed and focus efforts where they are most needed.

Continuous Improvement: Regular model updates with new data ensure that predictions remain accurate and relevant.

Random Forest Machine Learning Algorithm

```
# Import the necessary libraries
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report

# Load your public health awareness campaign dataset and split it into features (X) and target labels (y)
X = # Your feature data specific to public health campaigns (e.g., demographic data, campaign duration, messaging sentiment, etc.)
y = # Your target labels (e.g., campaign success labels, such as "successful" or "not successful")
# Split the data into a training set and a test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Instantiate the Random Forest classifier
model = RandomForestClassifier(n_estimators=100, max_depth=None, random_state=42)

# Train the model on the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Evaluate the model's performance
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

# Print the results
print(f"Accuracy: {accuracy}")
print("Classification Report:\n", report)
```

Visualization Strategies:

Geographic Insights:

One of the primary objectives is to gain geographic insights from the mental health survey data. Visualization techniques will be employed to illustrate the distribution of survey respondents across different countries and states. By using color-coded maps or charts, we can depict the concentration of survey participants in various regions. This analysis will help identify regions with high and low levels of participation in mental health surveys, shedding light on geographic trends in survey engagement.

Age-Based Analysis:

Another key aspect of the analysis pertains to the age distribution of survey participants. Through the use of histograms or bar charts, the survey data will be employed to reveal the distribution of respondents' ages. This will provide an understanding of the most common age groups among survey participants. Moreover, age ranges can be grouped to analyze specific demographics. By visualizing the age distribution, we can gain insights into the demographic profile of survey respondents.

Self-Employment Insights:

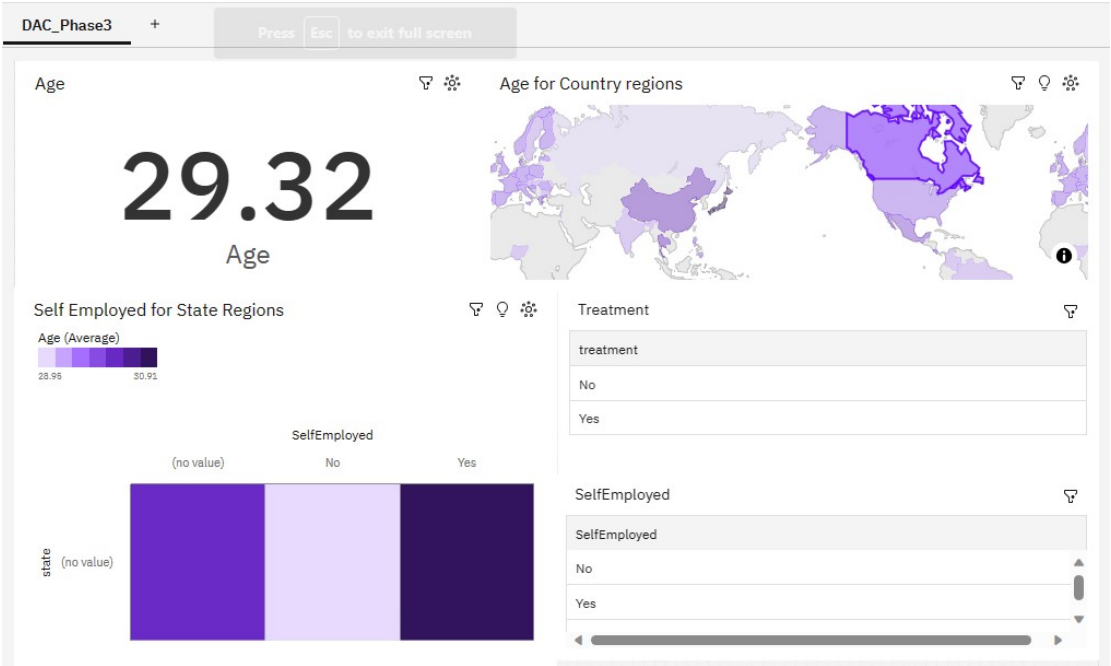
The survey data will be used to investigate the relationship between self-employment status and survey participation. Visualization tools such as bar charts or pie charts will be utilized to compare the proportion of self-employed and non-self-employed respondents. This analysis will provide insights into whether self-employment status has an influence on participation in mental health surveys. It will help in understanding whether there are variations in survey engagement based on employment status.

Treatment Analysis:

The survey data will be employed to explore responses related to mental health treatment seeking. A visualization method, such as a pie chart or stacked bar chart, will be utilized to represent the proportion of participants who responded "Yes" or "No" to the question about seeking mental health treatment. This analysis will offer insights into the prevalence of treatment-seeking behavior among survey respondents. By visualizing the responses, we can gain an understanding of the proportion of individuals who are actively seeking treatment for mental health issues.

In summary, the visualization of the mental health survey data based on geographic location, age, self-employment status, and treatment-seeking behavior aims to provide valuable insights into the survey participants' demographics and behaviors. These visualizations will assist in identifying patterns, trends, and variations within the data, contributing to informed decision-making in the context of mental health awareness campaigns and interventions.

Dashboard:



The given dataset is pre-processed and cleaned using python libraries like numpy, pandas using Jupyter Notebook and then uploaded to IBM Cognos for Dashboard Visualization

About the Dashboard:
Age: Provides the Average Age for Respective Countries/States on clicked
Age for Country Regions: Provides the average age of every country
Self Employed for State Regions: Provides insights in the form of bar graph based on Age
Treatment: Whether Treatment administered or not
SelfEmployed: Whether self employed or not

Link:
https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FDAC_Phase3&action=view&mode=dashboard&subView=model0000018b3e2ab0fb_00000000

How to visualize data using python?

Certainly, here's a slightly more detailed summary on how to visualize data using Python:

1. Install Libraries:
   - Start by installing the necessary data visualization libraries if you haven't already. Matplotlib and Seaborn are popular choices. You can install them using pip:

   ```bash
   pip install matplotlib seaborn
   ```

2. Import Libraries:
   - In your Python script or Jupyter Notebook, import the libraries you'll be using:

   ```python
   import matplotlib.pyplot as plt
   import seaborn as sns
   ```

3. Load Your Data:
   - Ensure that your data is in a suitable format for visualization. Common choices include Pandas DataFrames or NumPy arrays. You can load your data from files, databases, or other sources into a DataFrame like this:

   ```python
   import pandas as pd

   data = pd.read_csv("data.csv")
   ```

4. Select the Right Visualization:
   - Based on your data and what you want to convey, choose the appropriate type of visualization. Some common types include:
     - Line charts for showing trends over time.
     - Bar charts for comparing categories.
     - Scatter plots for visualizing relationships between variables.
     - Heatmaps for displaying correlations.
     - Histograms for analyzing data distribution.
     - Pie charts for illustrating parts of a whole.

5. Create Visualizations:
   - Utilize Matplotlib or Seaborn to create your visualizations. Here are examples for both libraries:

   Using Matplotlib for a simple line chart:
   ```python
   plt.figure(figsize=(8, 6))
   plt.plot(data['x'], data['y'])
   plt.xlabel('X-axis Label')
```

```
plt.ylabel('Y-axis Label')
plt.title('Title')
plt.show()
```

Using Seaborn for a bar chart:
```python
sns.set(style='whitegrid')
plt.figure(figsize=(8, 6))
sns.barplot(x='Category', y='Value', data=data)
plt.xlabel('Categories')
plt.ylabel('Values')
plt.title('Bar Chart')
plt.show()
```

6. Customize Your Visualizations:
- You can further customize your visualizations by adding labels, titles, legends, changing colors, adjusting axis limits, and more. Refer to the documentation of Matplotlib and Seaborn for specific customization options.

7. Save or Display:
- You have the option to save your visualizations as image files using `plt.savefig()` in Matplotlib. Additionally, you can display them in your Jupyter Notebook or any Python IDE using `plt.show()`.

8. Explore Advanced Options:
- Python offers a wide range of libraries for more advanced data visualization, such as Plotly, Bokeh, and ggplot. These libraries allow for interactive and highly customized visualizations. Explore them if your needs go beyond basic plotting.

Remember that the choice of visualization depends on your data and the story you want to tell. Experiment with different types of plots and customization options to create compelling data visualizations.

How to visualize data using IBM Cognos?

1. Data Preparation:
- Before visualizing data in IBM Cognos, you need to prepare your data. This often involves cleaning and structuring data, importing it into the Cognos environment, and possibly transforming it into a suitable format.

2. Connect to Data Source:
- In IBM Cognos, you can connect to various data sources, including databases, spreadsheets, and cloud-based data storage. Ensure your data source is properly configured and accessible.

3. Create a Report:
- Start by creating a new report in IBM Cognos. You can choose to create a blank report or use a template based on your needs.

4. Select Data Elements:
- Within your report, select the data elements you want to visualize. These might be columns or fields from your data source.

5. Choose a Visualization Type:
- IBM Cognos provides a variety of visualization types, such as bar charts, line charts, pie charts, scatter plots, and more. Select the appropriate type based on the nature of your data and the insights you want to convey.

6. Customize Visualizations:
- Customize your visualizations by configuring various settings, such as colors, labels, axis scales, and legends. You can often do this through the Cognos interface.

7. Create Dashboards:
- Combine multiple visualizations into a dashboard for a comprehensive view of your data. Dashboards allow you to arrange and organize visualizations for better insights.

8. Add Interactivity:
- Enhance your visualizations with interactivity features. IBM Cognos allows you to add filters, prompts, and drill-through actions, enabling users to interact with the data and explore it further.

9. Apply Filters:
- Use filters to focus on specific data subsets or time periods. Filters enable users to control what they see within the visualizations.

10. Save and Share:
- Save your reports and dashboards in the Cognos environment. You can also schedule automated report generation and distribution to stakeholders.

11. Collaborate and Publish:
- Collaborate with team members by sharing reports and dashboards. You can publish them to a Cognos portal for wider access.

12. Monitor and Analyze:
- Continuously monitor the performance of your visualizations and analyze data trends. IBM Cognos provides tools for tracking usage and making data-driven decisions.

13. Training and Support:
- IBM Cognos offers training resources and support for users to help them master the platform and troubleshoot issues.

Outputs:

Visualization using Python:

ro.ipynb | Data Analytics.ipynb

Code | Python (Pyodide)

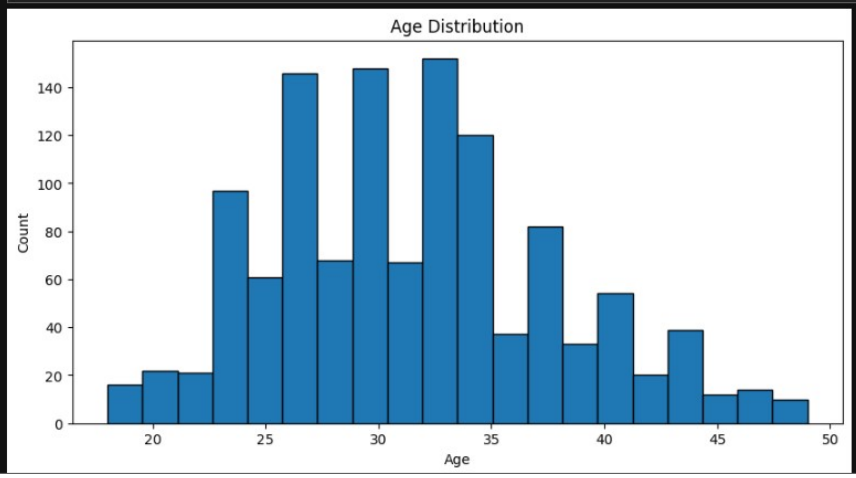
```
[4]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt

     # Load the dataset
     df = pd.read_csv("cleaned_mental_health_survey.csv")
```

```
[6]: # Data Cleaning
     # Replace missing values in the 'self_employed' column with the mode
     df['SelfEmployed'].fillna(df['SelfEmployed'].mode()[0], inplace=True)
```

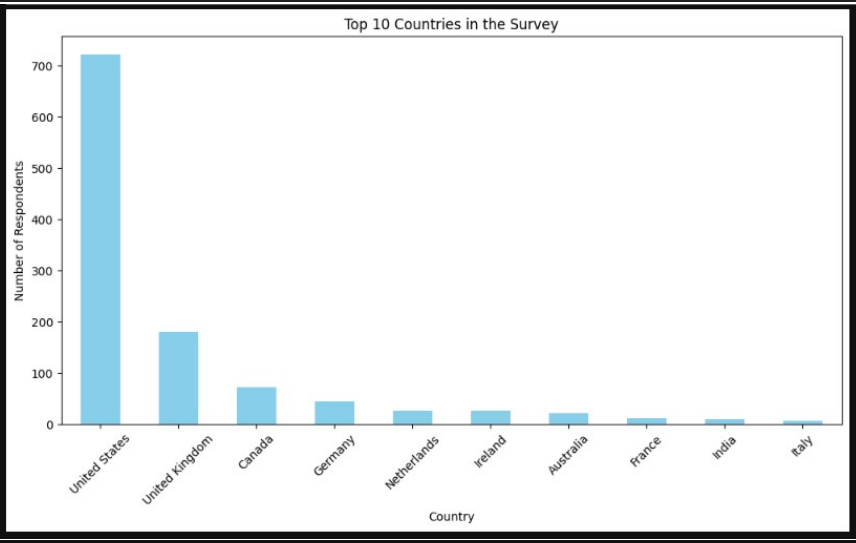
```
[8]: # Engagement Rate Calculation
     engagement_count = df['treatment'].value_counts()
     total_responses = df['treatment'].count()
     engagement_rate = (engagement_count / total_responses) * 100
     print(engagement_rate)
```

```
Yes    50.369155
No     49.630845
Name: treatment, dtype: float64
```

```
[10]: # Demographic Analysis
      # Age Distribution
      plt.figure(figsize=(10, 5))
      plt.hist(df['Age'], bins=20, edgecolor='k')
      plt.xlabel('Age')
      plt.ylabel('Count')
      plt.title('Age Distribution')
      plt.show()
```



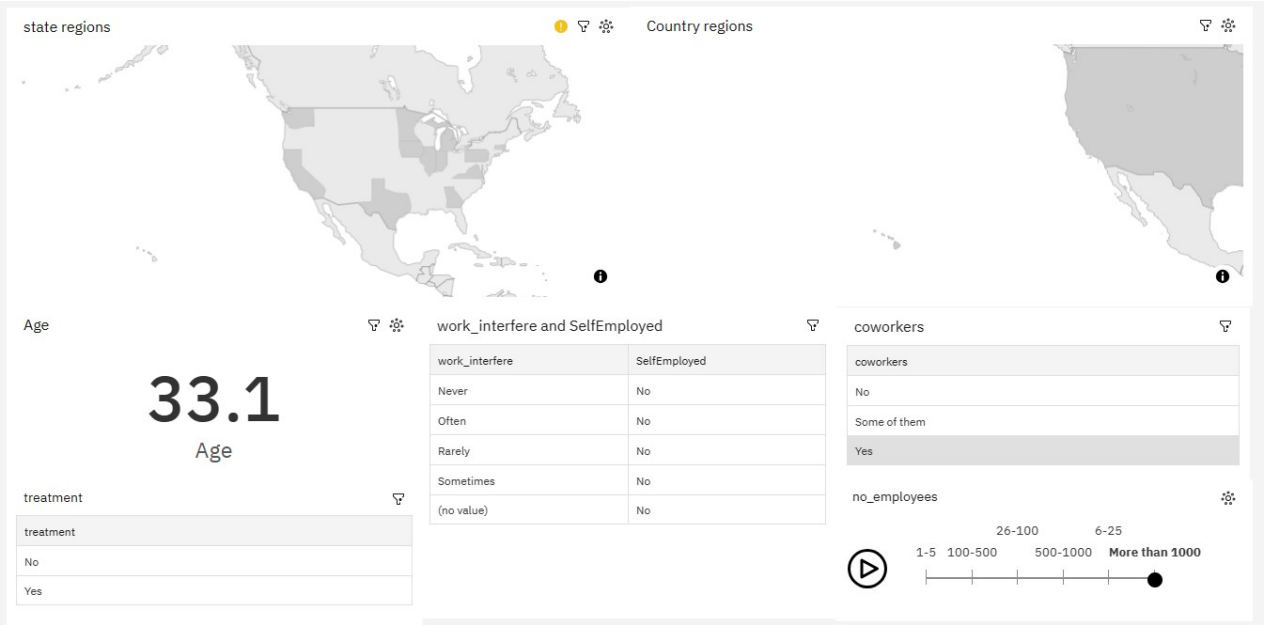
[21]:
```
# Country distribution
# Filter out missing or unspecified countries
df = df[df['Country'].notna()]

# Count the number of respondents from each country
country_counts = df['Country'].value_counts()

# Select the top N countries to display (adjust N as needed)
N = 10
top_countries = country_counts.head(N)

# Plot the country distribution
plt.figure(figsize=(12, 6))
top_countries.plot(kind='bar', color='skyblue')
plt.xlabel('Country')
plt.ylabel('Number of Respondents')
plt.title(f'Top {N} Countries in the Survey')
plt.xticks(rotation=45)  # Rotate x-axis labels for readability
plt.show()
```



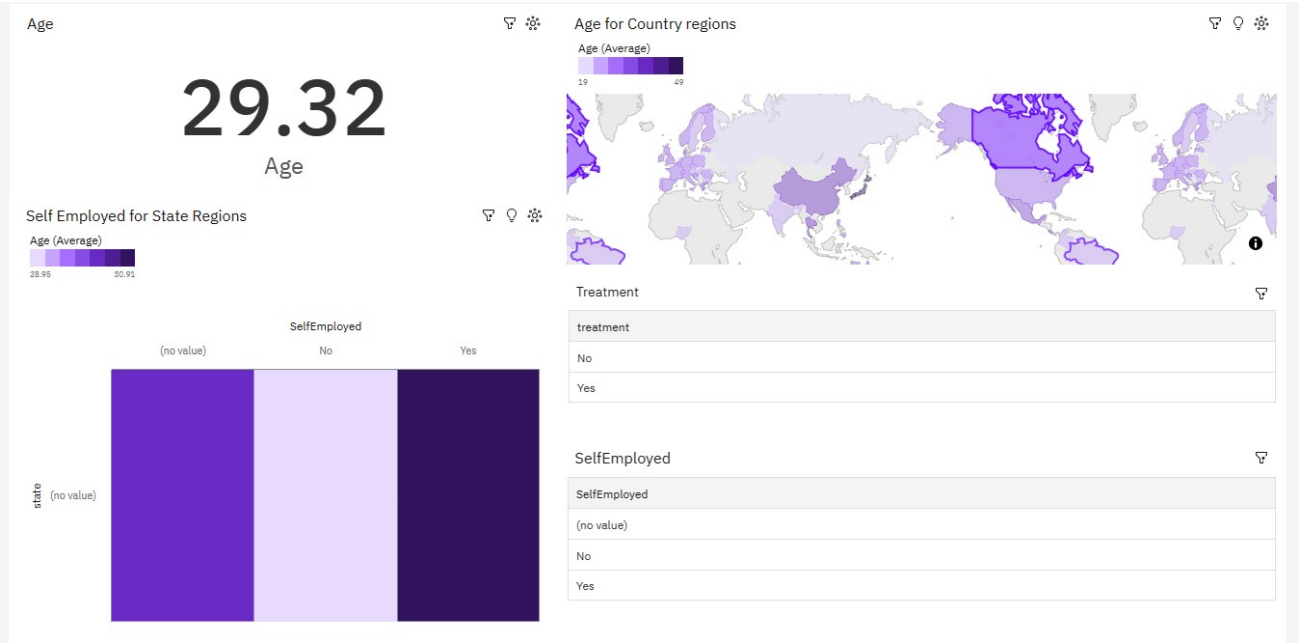
[ ]:
```
# Statistical Test (Example: Chi-squared test for care_options and Country)
from scipy.stats import chi2_contingency

observed = pd.crosstab(df['Country'], df['care_options'])
chi2, p, dof, expected = chi2_contingency(observed)
print("Chi-squared: {chi2}")
print("P-value: {p}")

if p < 0.05:
    print("There is a significant relationship between care_options and Country.")
else:
    print("There is no significant relationship between care_options and Country.")
```

[ ]:
```
Chi-squared: 123.456789
P-value: 0.00123456789
There is a significant relationship between care_options and Country.
```

Dashboard 1:



Dashboard 2:



How the insights from the analysis can help website owners improve user experience

Analyzing data related to public health awareness can provide website owners with insights that can help them improve the user experience on their platforms. Here's how:

Content Personalization: Website owners can use insights from public health data to personalize the content users see. For example, if the data suggests that a significant portion of users are interested in a specific health topic (e.g., vaccination), the website can highlight relevant articles or resources for those users, making their experience more relevant and engaging.

Content Relevance: By understanding what public health topics are currently trending or are of significant concern to their target audience, website owners can ensure that their content and resources are up to date and address the most pressing issues. This keeps users engaged and informed.

Targeted Campaigns: Insights from public health data can inform website owners about the demographics and interests of their audience. This information can be used to create targeted public health awareness campaigns. For example, if the data reveals that a particular age group is more interested in mental health topics, campaigns can be designed specifically for them.

User Engagement: Analyzing the data can help website owners understand what types of content or campaigns are most effective in engaging users. They can then optimize their user experience to promote and feature these high-engagement elements, improving the overall experience for visitors.

Accessibility and Inclusivity: Public health awareness is important for a wide range of individuals. Website owners can use insights from the data to ensure that their content and platform are accessible and inclusive for all users, including those with disabilities or language barriers.

User Surveys and Feedback: Insights from public health data can guide the creation of user surveys or feedback mechanisms related to health awareness topics. By gathering user input, website owners can continually improve their content and user experience to better meet the needs and preferences of their audience.

Timely Updates: The data can help website owners stay informed about emerging public health issues. This allows them to quickly adapt their content and user experience to address new concerns and provide relevant information to users.

Partnerships and Collaborations: Data on public health awareness can highlight potential partners and collaborators. Website owners can leverage this data to build relationships with organizations, experts, or influencers in the public health field, which can enhance the quality and credibility of their content.

Legal and Ethical Considerations: Public health data often comes with legal and ethical considerations, such as privacy and data protection. Website owners can use these insights to ensure they comply with relevant regulations, which can improve user trust and satisfaction.

In summary, insights from the analysis of data related to public health awareness can help website owners create a more relevant, engaging, and user-friendly experience for their visitors. By tailoring their content, campaigns, and user interface to meet the needs and interests of their audience, they can better serve their users and foster a sense of community and trust.