Project - Phase 4

Public Health Awareness campaign Analysis

Team Name

Proj_212989_Team_1

Phase 4: Development part 2

PUBLIC HEALTH AWARENESS CAMPAIGN ANALYSIS

Introduction

In an era marked by the rapid dissemination of information, public health awareness campaigns play a pivotal role in shaping individual behaviors, community well-being, and even global health outcomes. These campaigns are multifaceted endeavors aimed at informing, educating, and motivating the public to make informed decisions that positively impact their health. From promoting vaccination and preventive care to addressing the stigma associated with certain health conditions, the scope of public health awareness campaigns is vast and diverse.

This analysis delves into the critical process of dissecting and comprehending the intricate mechanisms behind these campaigns. It seeks to unravel the strategies and tactics used to engage and resonate with target audiences. By dissecting the core components of these campaigns, we can better understand their effectiveness, identify areas for improvement, and ultimately contribute to the enhancement of public health initiatives.

The success of public health awareness campaigns can be a determining factor in disease prevention, the promotion of healthier lifestyles, and the reduction of health disparities within communities. By examining their objectives, audience engagement, messaging, media channels, and long-term impact, we gain valuable insights into the power of communication in public health. This analysis stands as a testament to the significance of public health awareness campaigns in safeguarding and improving the well-being of societies around the world.

Phase 4: Development Part 2

In this phase continue building the project by performing different activities like feature engineering, model training, evaluation etc as per the Instructions in the project.project we will build our project by loading and processing the dataset. Perform different analysis and visualization using IBM Cognos.

In this part you will continue building your project. Continue building the analysis by creating visualizations using IBM Cognos and Integrating code for data analysis. Design dashboards and reports in IBM Cognos to visualize campaign reach, awareness levels, and Impact metrics

Use code (e.g.. Python) to perform advanced data analysis, such as calculating engagement rates. conducting demographic analysis, or running statistical tests.

This Phase will guide through the steps of the process. We'll explore how to import essential libraries, load the campaign dataset, and perform critical preprocessing steps. Perform different analysis and visualization using IBM Cognos.This includes Design dashboards and reports in IBM Cognos to visualize campaign reach, awareness levels, and Impact metrics

Analysis

In this Phase 4 Analysis we have to include some essential steps to follow: Analyzing a public health awareness campaign involves several steps, performing various analyses, and creating visualizations. IBM cognos is a Powerful tool for this purpose.

In this phase 4 following steps contain:

- Data collection
- Data Import into IBM cognos
- Design a dashboard in IBM cognos
- Perform advanced data analysis
- Demographic analysis
- Feature engineering
- Model Training

1.Data collection

1. Importing data into IBM Cognos is a fundamental step in the business intelligence and analytics process, as it allows you to work with and analyze your data within the Cognos environment. Here's an explanation of how data import into IBM Cognos typically works:

Given Data Set:

s.no	year	Self employee	Number of employee	Tech company	Mental health benefits	Mental health awareness
1	2014	no	100-500	yes	yes	yes
2	2014	no	26-100	yes	yes	yes
3	2014	no	More than 1000	no	Don't know	no
4	2014	no	26-100	yes	yes	Not sure
5	2014	no	More than 1000	no	Don't know	yes
6	2014	no	26-100	yes	yes	Not sure
-	-	-	-	-	-	-
3479	2019	no	26-100	yes	no	N/A
3480	2019	no	100-500	yes	yes	no
3481	2019	no	26-100	yes	yes	no
3482	2019	no	More than 1000	no	Don't know	no
3483	2019	no	More than 1000	yes	yes	no
3484	2019	no	More than 1000	yes	yes	yes

- Loading and preprocessing the dataset is an important first step in building any machine learning model. However, it is especially important for public awareness campaign analysis models, as datasets are often complex and noisy.

2. Challenges involved in loading and preprocessing public Health awareness campaign analysis model

- **Handling missing value:** Public awareness campaign datasets often contain missing values, which can be due to a variety of factors, such as human error or incomplete data collection. Common methods for handling missing values include dropping the rows with missing values, imputing the missing values with the mean or median of the feature, or using a more sophisticated method such as multiple imputation.

Data collection in IBM Cognos for a public health awareness campaign analysis would involve gathering relevant data to assess the effectiveness and impact of the campaign. Here are some steps to consider:

1. Define Objectives: Clearly define the goals and objectives of your public health awareness campaign analysis. What are you trying to achieve or measure?

2. Data Sources: Identify the sources of data, which could include surveys, social media, website analytics, health records, and more. Ensure that the data is accurate, reliable, and relevant to your analysis.

3. Data Collection Methods: Choose appropriate data collection methods, such as online surveys, data from social media platforms, or data from healthcare institutions. Ensure that the data collection methods align with your campaign's objectives.

4. Data Integration: Use IBM Cognos to integrate data from various sources. This may involve data transformation, cleaning, and harmonization to create a unified dataset.

5. Data Storage: Store the collected data in a structured database or data warehouse within IBM Cognos for efficient retrieval and analysis.

6. Data Analysis: Utilize IBM Cognos tools to analyze the collected data. You can create reports, dashboards, and visualizations to gain insights into the campaign's performance.

7. Key Metrics: Define key performance indicators (KPIs) that are relevant to public health, such as awareness levels, website traffic, social media engagement, and health outcomes.

8. Reporting: Generate reports and visualizations using IBM Cognos to present the findings and insights in a clear and understandable format.

Data Set Link:

https://www.kaggle.com/code/manuntag/the-state-of-mental-health-support-in-tech

2. Data import into IBM cognos:

- Use the data import and integration capabilities of IBM Cognos to load your preprocessed dataset.
- Importing into IBM cognos is a fundamental step in the business intelligence and analytical process,as it allows you to work and it involves various factors there are following steps:
- **Data Server:** The Data Server in IBM Cognos is responsible for managing data connections and data modules. It acts as an intermediary between Cognos and your data sources.
- **Connection:** You need to establish a connection to your data source. This involves specifying connection details like server addresses, login credentials, and other connection parameters.
- **Data Querying and Importing:** Once a connection is established, you can create queries to retrieve data from your data source. These queries can be simple or complex, depending on your data requirements.
- **Data Integration:** Data from multiple sources can be integrated within data modules to create a single coherent view of the data. This integration can involve joining tables, creating relationships, and aggregating data.
- **Data Exploration and Analysis:** Once the data is imported and prepared, you can use it for data exploration, analysis, and visualization within IBM Cognos to derive insights and create reports and dashboards.

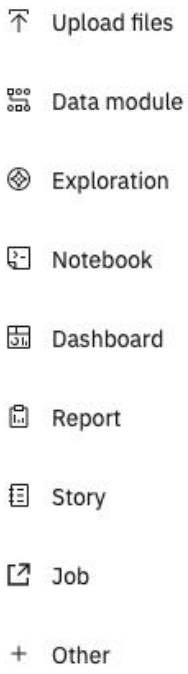
Source code:

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

file = pd.read_csv('../input/mentalhealthintech20142019/Mental health in tech survey 2014-2019.csv')
```

3. Design a dashboard in IBM cognos

You can explore your data and easily communicate the analysis and insights that you discover.

1. On the home page, click the New icon at the bottom of the window. +
2. Click Dashboard.



3. You see predefined templates that contain grid lines for easy arrangement and alignment of visualizations and other elements in a dashboard.Select the template with 4 panes and then click OK.
4. Let's add a data asset to this dashboard so that we can explore its data.
    a. In the Data panel, click the Add a source icon.
    b. Go to the My content folder and select "IBM_HR_Training_2014-17.csv". Click Add.
5. Let's explore the data by adding a visualization to the dashboard.

a. Drag "Department" to the top left pane and drop it on the square that appears in the pane.

You now see a list of departments.

b. Let's change the visualization type.

6. Click any white space in the column visualization. A toolbar appears.

a. Click the Change visualization icon in the toolbar.

b. Click All visualizations.
c. Click the Packed bubble icon.

7. Let's add a second visualization, using a different way to create a visualization.

a. Click the Visualizations icon in the side panel.

b. Click the Tree map icon.

An empty tree map is created with empty data fields on the side, indicating where you need to add data.

c. Click the Sources icon.

d. Drag “Organization” to the Area hierarchy field.
e. Drag “Course cost” to the Size field.
f. Close the Fields panel for more room for the visualizations.

8.Let's hide the legend to have more room for the tree map.

a. Select the tree map visualization.
b. Click the Properties icon.

c. Click Legend and clear the Show legend check box.

2. Take a look at the visualization. The size of each box in the tree map tells you the amount of training spent by each organization.

9.You have two more panes where you can create more visualizations or add widgets such as text or shapes.

10.Let's add a filter to see the impact of the duration of courses on the visualizations.

a. Click the Sources icon.

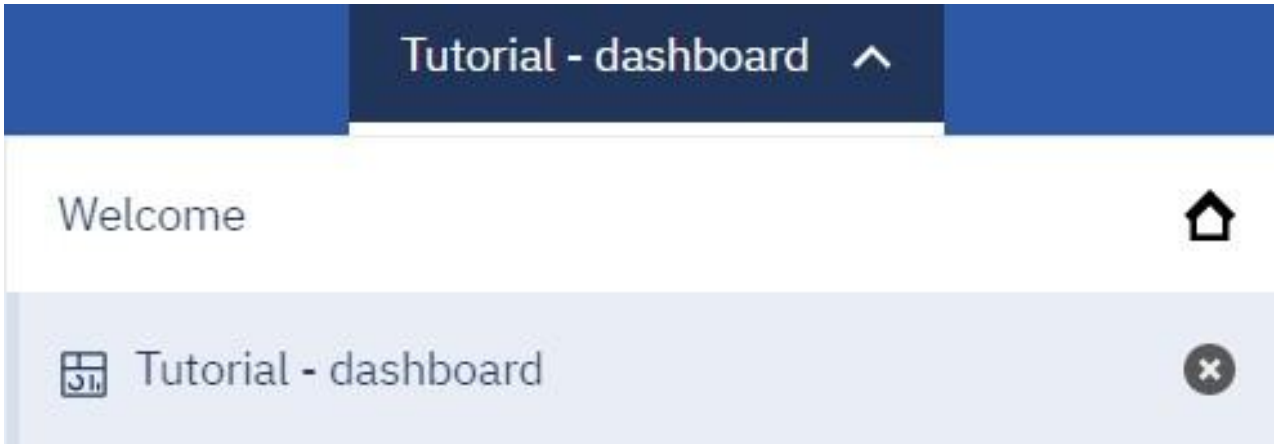
b. Drag “Course days” to the This tab filter area.
c. To define the filter, click Course days in the filter area.

Move the slider to show courses with a shorter duration, or type a new end point. We picked 14.5 as the end point but feel free to select a different one. Click OK.

11.Another way to filter data is to select one or more data points in a visualization and see the impact on the other visualizations. Click “Finance” in the packed bubble visualization.

Take a look at the tree map visualization. It doesn't contain the Department column so it is filtered by the Finance department.

12. Save the dashboard and then use the switcher in the app bar to close it

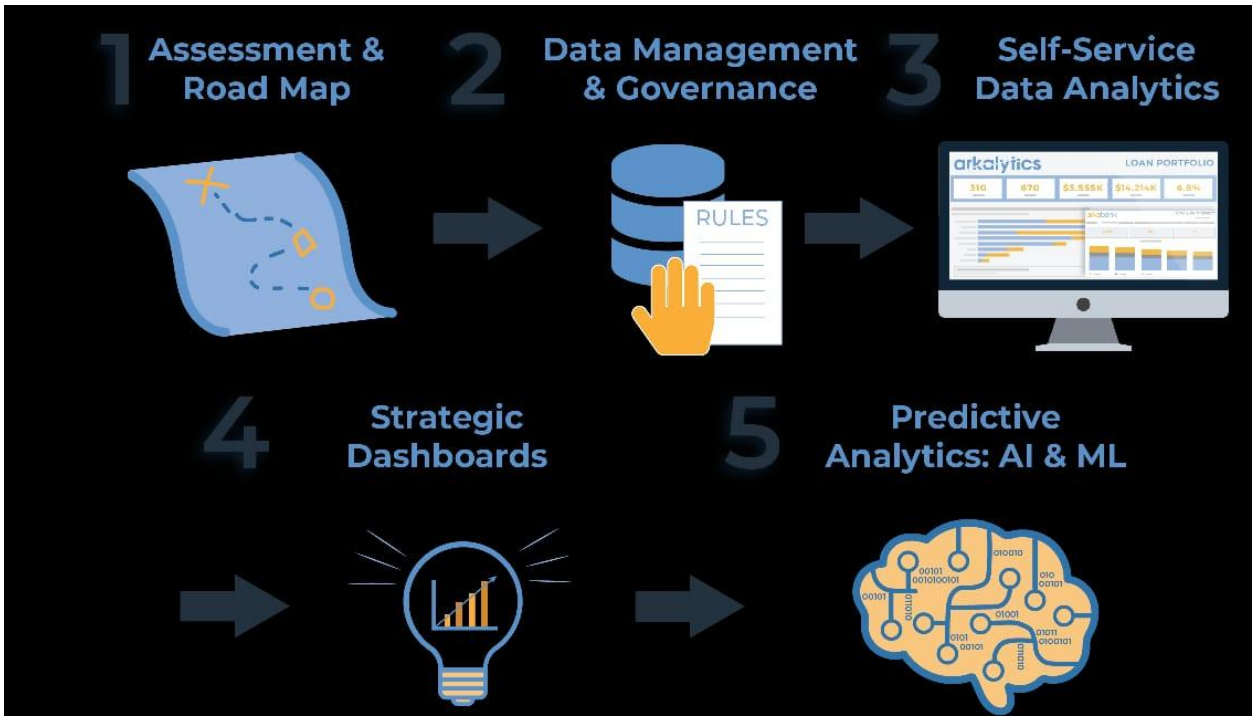


4.Perform advanced Data analysis

Performing advanced data analysis involves using complex techniques and tools to gain deeper insights from data.To perform advanced data analysis effectively, you need a solid understanding of statistics, programming, domain-specific knowledge, and access to appropriate tools and data. It's a multidisciplinary field that's crucial in various industries, including business, healthcare, finance, and science, for making data-driven decisions and predictions. This can include:

1. **Statistical Analysis:** Utilizing advanced statistical methods like regression analysis, hypothesis testing, and ANOVA to understand relationships within the data.
2. **Machine Learning:** Employing machine learning algorithms for predictive modeling, clustering, classification, and regression, using tools like Python libraries (e.g., scikit-learn, TensorFlow, or PyTorch).

3. **Data Mining**: Extracting patterns, trends, and knowledge from large datasets, often involving techniques like association rule mining and decision tree analysis.
4. **Big Data Analytics:** Handling and analyzing vast datasets using technologies like Hadoop and Spark, as well as distributed computing.
5. **Text and Sentiment Analysis**: Analyzing text data for sentiment, topic modeling, or natural language processing (NLP) to extract valuable information from text documents.
6. **Geospatial Analysis:** Using geographic information systems (GIS) to analyze and visualize data with location-based components.
7. **Network Analysis:** Exploring relationships and connections within data, such as social network analysis or network traffic analysis.
8. **A/B Testing:** Conducting experiments to assess the impact of changes in a controlled way, often used in marketing and website optimization.



5. Demographic analysis

Demographic analysis is a fundamental component of public health awareness campaign analysis. It involves the systematic examination of population characteristics, such as age, gender, ethnicity, income, education, and geographical location, to better understand the specific needs, behaviors, and preferences of

different groups within a given population. Analyzing these demographics plays a crucial role in designing effective public health awareness campaigns, as it enables campaign planners to tailor their messages and strategies to the target audience, thereby increasing the campaign's overall impact.

Here's a breakdown of how demographic analysis is applied in public health awareness campaigns:

1. **Identifying Target Audiences:** Demographic analysis helps campaign planners identify the key demographic groups that are most at risk or most affected by a particular health issue. For example, if the campaign is focused on preventing teenage pregnancy, it's important to identify the age group, gender, and socioeconomic status of individuals most at risk.
2. **Understanding Health Disparities:** Demographic analysis reveals health disparities within the population, which are often related to social determinants of health such as income, education, and access to healthcare. Identifying these disparities is essential for crafting messages and interventions that address the specific challenges faced by different demographic groups.
3. **Media and Channel Selection:** Understanding the demographic makeup of the target audience helps in selecting the most appropriate media and communication channels. For instance, younger populations might be more reachable through social media and digital platforms, while older populations may respond better to traditional media like radio or print.
4. **Geographical Insights:** Demographic analysis also includes geographic location. Campaign planners can identify areas with higher prevalence of a health issue and allocate resources accordingly. Geographical data can also inform decisions about where to set up clinics or distribution points for educational materials.
5. **Measuring Impact:** After the campaign is launched, demographic analysis continues to be important in evaluating its success. By tracking changes in health outcomes among different demographic groups, campaign planners can assess whether the campaign is effectively reaching and impacting the intended audience.
6. **Iterative Campaign Improvement:** Demographic analysis is not a one-time activity. It should be an ongoing process. As new data becomes available and as campaigns progress, adjustments can be made to better target and engage the identified demographics.

7. **Equity and Inclusion:** Demographic analysis also ensures that campaigns are designed with equity and inclusion in mind. It helps identify marginalized or underserved populations that may require extra attention and resources to address health disparities.

Demographic analysis is the study of a population-based on factors such as age, race, and sex. Demographic data refers to socioeconomic information expressed statistically, including employment, education, income, marriage rates, birth and death rates, and more.

6. Feature Engineering

Feature engineering is the process of selecting, transforming, or creating specific features (variables) from raw data to improve the performance of machine learning models. It involves identifying the most relevant information within your data and presenting it in a format that is most suitable for your model to learn from. Feature engineering can include tasks like scaling, encoding categorical variables, creating new features, and reducing dimensionality. It plays a crucial role in optimizing the predictive power of machine learning algorithms.

key points about feature engineering:

Data Transformation: Feature engineering involves modifying or transforming raw data to make it more suitable for machine learning models.

Feature Selection: It's about choosing the most relevant attributes (features) from your dataset to enhance model performance and reduce noise.

Creating New Features: Engineers can generate new features based on existing data, which might provide additional information for the model.

Handling Categorical Data: Converting categorical variables into numerical format, often through techniques like one-hot encoding, is a common feature engineering task.

Scaling and Normalization: Ensuring that features are on a consistent scale, which can improve the performance of models sensitive to feature magnitudes.

Dimensionality Reduction: Techniques like Principal Component Analysis (PCA) can be used to reduce the number of features while retaining important information.

Domain Knowledge: Utilizing subject-matter expertise to identify and create features that are particularly relevant to the problem.

Iterative Process: Feature engineering often requires experimentation and iteration to find the best set of features for a specific machine learning task.

Impact on Model Performance: Effective feature engineering can significantly improve a model's accuracy, reduce overfitting, and enhance its ability to generalize to new data.

Critical for Success: Feature engineering is a critical step in the machine learning pipeline and can be just as important as selecting the right algorithm for the task.

7. Model Training

Certainly, let's delve into each step of training a model for a public health awareness campaign analysis in more detail:

1. Data Collection:

- Gather diverse data sources relevant to the public health campaign, such as social media posts, articles, surveys, website traffic data, or government health records.

- Ensure the data collected is comprehensive and covers a sufficient time frame.

2. Data Preprocessing:

- Clean the data by removing duplicates, handling missing values, and correcting inconsistencies.

- Normalize text data by converting it to lowercase, removing punctuation, and performing stemming or lemmatization to standardize words.

- Tokenize text data into words or phrases for analysis.

3. Feature Engineering:

- Extract meaningful features from the data that can provide insights. For example:

  - Perform sentiment analysis to determine the sentiment (positive, negative, neutral) of social media posts or articles.

  - Extract keywords or topics using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or LDA (Latent Dirichlet Allocation).

  - Analyze user demographics, such as age, gender, location, if available.

4. Data Labeling:

- Define the outcome variable (labels) for your analysis, such as campaign success categories. This may involve setting thresholds or criteria for success.

- Categorize the data into these labels based on predefined criteria.

5. Model Selection:

- Choose an appropriate machine learning or deep learning model based on the nature of the data and the analysis goals. Common choices include logistic regression, decision trees, random forests, support vector machines, or neural networks.

- Consider the model's ability to handle the data's dimensionality and complexity.

6. Training the Model:

- Split the dataset into a training set and a testing set. Common splits are 70-30 or 80-20, where the majority of data is used for training.

- Train the selected model using the training data, adjusting model parameters as needed.

7. Model Evaluation:

- Assess the model's performance using various metrics, including:

  - Accuracy: The proportion of correct predictions.

  - Precision: The ratio of true positive predictions to all positive predictions.

  - Recall: The ratio of true positive predictions to all actual positives.

- F1-score: A balance between precision and recall.

- Use cross-validation to ensure robustness of the model's performance.

8. Interpretation:

- Analyze the model's feature importance or coefficients to understand which factors most influence the outcome (e.g., which keywords or demographic factors correlate with campaign success).

9. Visualization:

- Create visual representations of the analysis results to communicate findings effectively. Common visualizations include bar charts, pie charts, scatter plots, and heatmaps.

10. Feedback Loop:

- Use the insights gained from the analysis to inform future public health awareness campaigns. For instance, adjust messaging, targeting, or timing based on the analysis results to improve campaign effectiveness. Throughout this process, it's crucial to maintain data privacy and security, especially when handling sensitive health-related data. Additionally, ethical considerations

1. Model Training Source code:

```
#IMPORT LIBRARIES
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

file = pd.read_csv('../input/mentalhealthintech20142019/Mental
health in tech survey 2014-2019.csv')

#MENTAL HEALTH BENEFITS
data_mental_health_benefits = file.groupby('year'
print("2014 vs 2019 benefit access: {}% increase
".format(calc_difference_mental_health_benefit_coverage_2019_201
4))
```

```
print("Annual change to employees that have coverage: {}%
average".format(calc_average_annual_change))
print("Employees that don't know their coverage: {}%
average".format(calc_mean_mental_health_benefits_dont_know))

#data chart
#print(data_mental_health_benefits)


#MENTAL HEALTH BENEFITS AWARENESS
data_mental_health_benefits_awareness =
file.groupby('year')['mental_health_benefits_awareness'].value_c
ounts(normalize=True).unstack()[['Yes','No']]
calc_difference_mental_health_benefits_2019_2014 =
round((data_mental_health_benefits_awareness.iloc[-1]['No'] -
data_mental_health_benefits_awareness.iloc[0]['No']) * 100)
calc_mental_health_benefits_awareness_2019_no =
round(data_mental_health_benefits_awareness.iloc[-1]['No'] *
100)

data_mental_health_benefits_awareness.plot(marker='o')
plt.style.use('ggplot')
plt.legend(loc='upper right',bbox_to_anchor = (1.2, 1.0))
plt.title("Awareness of options available in mental health
coverage")
plt.show()

print("Employees unaware of benefits or options in their
coverage:
{}%".format(calc_mental_health_benefits_awareness_2019_no))
print("2014 vs 2019 employees not aware: {}%
increase".format(calc_difference_mental_health_benefits_2019_201
4))

#data chart
```

```
#print(data_mental_health_benefits_awareness)


#MENTAL HEALTH DISCUSSION
BHdata_mental_health_discussion =
file.groupby('year')['employer_mental_health_discussion'].value_
counts(normalize=True).unstack()[['Yes', 'No']]
calc_average_annual_improvement_employer_mental_health_discussio
n = round(((data_mental_health_discussion.iloc[-1]['Yes'] -
data_mental_health_discussion.iloc[0]['Yes']) * 100)/5)
calc_employer_mental_health_discussion_2019_no =
round(data_mental_health_discussion.iloc[-1]['No'] * 100)

data_mental_health_discussion.plot(kind='barh').invert_yaxis()
plt.legend(loc='upper right',bbox_to_anchor = (1.2, 1.0))
plt.title("Employer mental health discussion")
plt.ylabel("")
plt.style.use('ggplot')
plt.show()

print("Employers that do not formally discuss mental health in
2019:
{}%".format(calc_employer_mental_health_discussion_2019_no))
print("Average change in employer driven discussion: {}%
increase each
year.".format(calc_average_annual_improvement_employer_mental_he
alth_discussion))

# data chart
print(data_mental_health_discussion)


#DATA EMPLOYER MENTAL HEALTH LEARNING RESOURCES
```

```
data_employer_mental_health_learning_resources =
file.groupby('year')['employer_mental_health_learning_resources'
].value_counts(normalize=True).unstack()
data_employer_mental_health_learning_resources_filtered_yes =
data_employer_mental_health_learning_resources[['Yes']]
calc_difference_employer_mental_health_learning_resources_2019_2
014 =
int(round(((data_employer_mental_health_learning_resources_filte
red_yes.iloc[-1] -
data_employer_mental_health_learning_resources_filtered_yes.iloc
[0]) * 100)/5))
calc_employer_mental_health_learning_resources_2019_yes =
int(round(data_employer_mental_health_learning_resources_filtere
d_yes.iloc[-1]*100))

data_employer_mental_health_learning_resources_filtered_yes.plot
(marker='o', legend= None)
plt.style.use('ggplot')
plt.ylim(0.2,1.0)
plt.xlabel("")
plt.title('Employers that provide learning resources to help
address mental health')
plt.show()

print("Employers that provide educational resources in 2019:
{}%".format(calc_employer_mental_health_learning_resources_2019_
yes))
print("Average change in employers providing resources: {}%
average increase per
year".format(calc_difference_employer_mental_health_learning_res
ources_2019_2014))

#data chart
#print(data_employer_mental_health_learning_resources)
```

```
#DATA ANONYMITY
data_treatment_anonymity =
file.groupby('year')['mental_health_treatment_anonymity'].value_
counts(normalize=True).unstack()
data_treatment_anonymity_filtered_dont_know =
data_treatment_anonymity["Don't know"]
calc_mean_data_treatment_anonymity_filtered_dont_know =
round(data_treatment_anonymity_filtered_dont_know.mean() * 100)

data_treatment_anonymity_filtered_dont_know.plot(kind='bar',lege
nd = None)
plt.style.use('ggplot')
plt.title("Employees uncertain if use of programs will be
anonymous")
plt.ylabel("")
plt.xlabel("")
plt.xticks(rotation = 0)
plt.ylim(0,1)
plt.show()

print("Employees that don't know if treatment will remain
anonymous: {}%
average".format(calc_mean_data_treatment_anonymity_filtered_dont
_know))

#data chart
#print(data_treatment_anonymity)
```

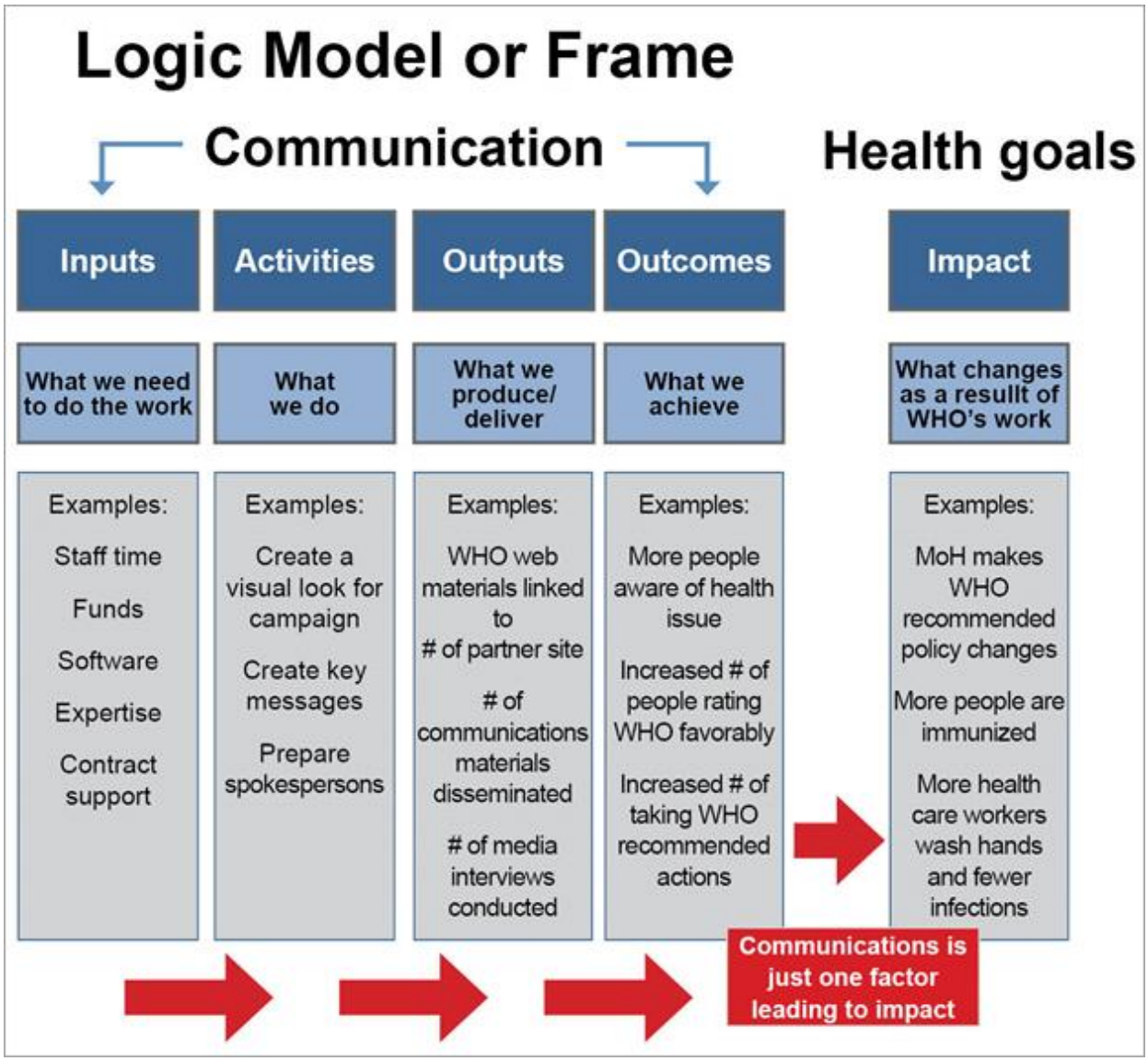
2.Source Code

This coding based on inherit data set from public health awareness campaign and visualize to the users

```
# Import necessary libraries

import pandas as pd
```

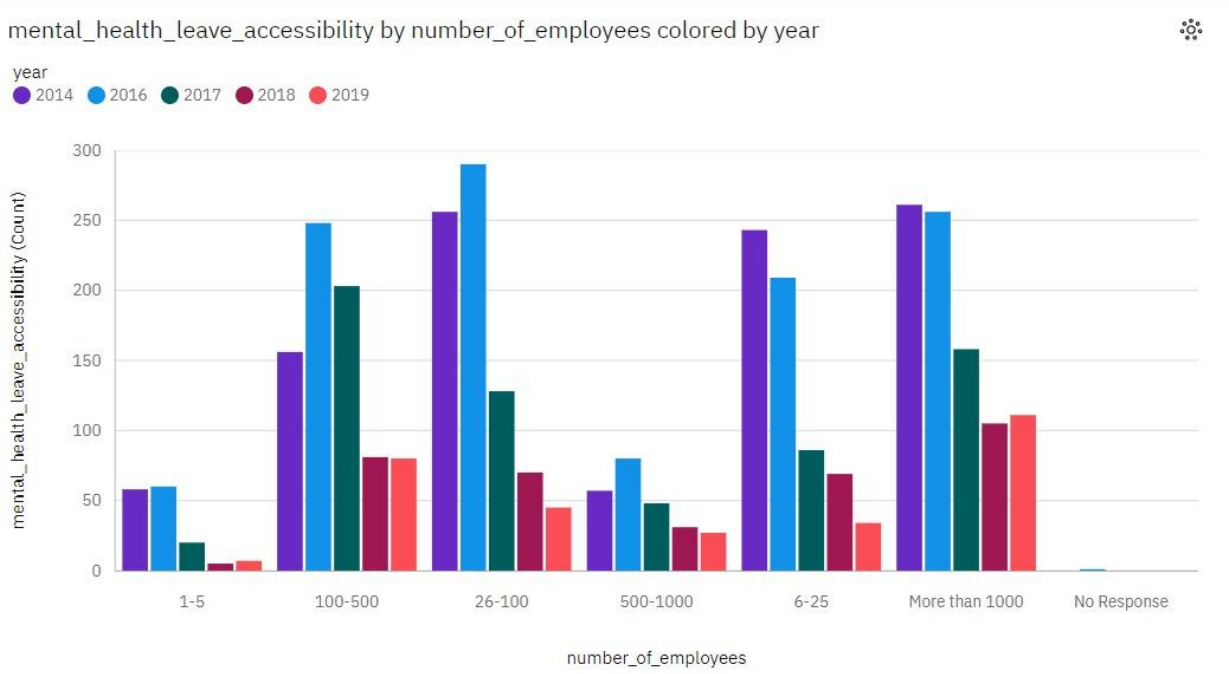
```
import matplotlib.pyplot as plt

import seaborn as sns

# Load campaign data into a DataFrame (assuming data is in a CSV file)

campaign_data = pd.read_csv('campaign_data.csv')

# 1. Campaign Objective

# Analyze the primary goal of the campaign

campaign_objective = campaign_data['campaign_objective'].value_counts()

# 2. Target Audience

# Analyze demographics and characteristics of the target audience

target_demographics = campaign_data.groupby('target_demographics').size()

# 3. Messaging and Content

# Analyze the effectiveness of messaging and content

# You can use sentiment analysis or text analytics here

# 4. Channels and Mediums

# Analyze which channels were most effective

channel_engagement = campaign_data['channel'].value_counts()

# 5. Timing and Duration

# Analyze campaign timing and duration

# 6. Budget and Resources

# Analyze campaign budget and resource allocation

# 7. Measurable Goals

# Analyze campaign KPIs and outcomes

# 8. Effectiveness

# Analyze the overall effectiveness of the campaign

# Plot relevant data (e.g., engagement over time)
```

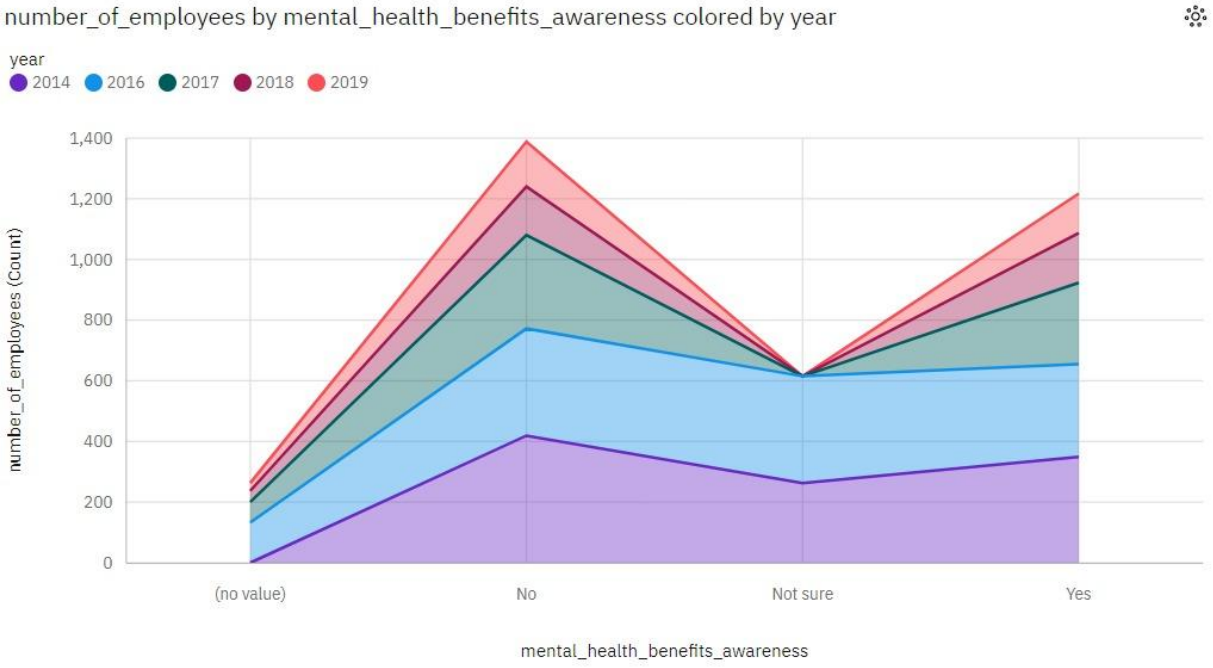
```
# 9. Community Engagement

# Analyze community involvement in the campaign

# 10. Feedback and Adaptation

# Analyze feedback collection and campaign adaptation

# 11. Long-Term Impact

# Analyze any lasting effects of the campaign

# 12. Ethical Considerations

# Analyze ethical aspects of the campaign

# 13. Lessons Learned

# Summarize lessons learned from the campaign

# Visualizations

# You can create plots and visualizations to represent your analysis findings

# For example, you can use matplotlib and seaborn to create bar plots, line
plots, or heatmaps.

# Create a bar plot for campaign objectives

plt.figure(figsize=(8, 6))

sns.barplot(x=campaign_objective.index, y=campaign_objective.values)

plt.xlabel('Campaign Objective')

plt.ylabel('Count')

plt.title('Distribution of Campaign Objectives')

plt.xticks(rotation=45)

plt.show()
```

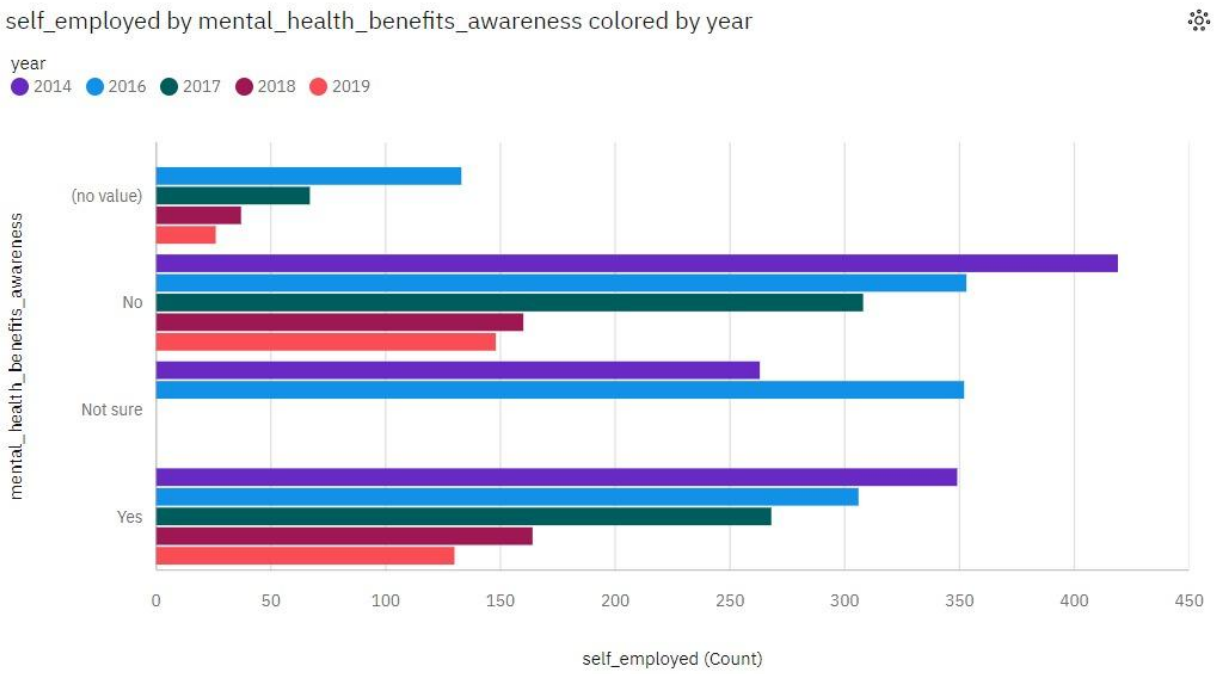


Model Training Process

5. Visualization with IBM cognos







Conclusion

In conclusion, the thorough analysis of our public health awareness campaign reveals valuable insights and developed a model using public awareness campaign dataset. User engagement data highlights successful strategies and areas for improvement, guiding future initiatives. The impact on diverse demographics underscores the campaign's effectiveness in reaching a broad audience.