NEHRU INSTITUTE OF ENGINEERING AND TECHNOLOGY COIMBATORE-641105

Pravin Raj B ,

Pulla Reddy Y,

Vamshi krishna K,

Thirumaran G.K,

Adithyan E.

Title :- Public Health Awareness

Project Definition:

The project involves analyzing data from public health awareness campaigns to measure their effectiveness in reaching the target audience and increasing awareness. The objective is to provide insights that evaluate the impact of the campaigns and inform future strategies.

This project includes defining analysis objectives, collecting campaign data, designing relevant visualizations in IBM Cognos, and using code for data analysis. Good health is essential for every human being. Health information is essential for maintaining good health, preventing diseases as well as making sound health decisions. People can only be able to access, utilize, and benefit from healthcare services if they have proper information about these services. It is here that health information literacy comes into play.

The information related to every aspect of health is easily available today, but the main problem here lies in finding, selecting, and using relevant health information and preventing misinformation. Libraries have a pivotal role to play here.

This chapter is mainly concerned with identifying the gaps in the provision of health information to the general public and the role of health information literacy in paving the way of filling up these gaps. It will be helpful in knowing the current standing of public and medical libraries in providing health information resources and services. It will also suggest the role of these libraries of India in promoting health information literacy among their respective user communities.

Design Thinking:

1. Analysis Objectives: Define specific objectives for analyzing public health awareness campaign data, such as measuring audience reach, awareness levels, and campaign impact.
2. Data Collection: Identify the sources and methods for collecting campaign data, including engagement metrics, audience demographics, and awareness surveys.
3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos to create informative dashboards and reports.
4. Code Integration: Decide which aspects of the analysis can be enhanced using code, such as data cleaning, transformation, and statistical analysis.
5. Data abstraction: Data were collected on target users, health conditions, objective of the intervention, details on the Design Thinking process, study design and sample, and reported health outcomes. If information was not reported in the article, we contacted the study authors. Studies were also evaluated to determine whether the intervention improved all targeted outcomes (successful), at least one targeted outcome (mixed success), or no targeted outcomes (not successful). Data quality was assessed using the National Institutes of Health’s (NIH’s) National Heart, Lung, and Blood Institute Study Quality Assessment Tools (12).
6. Tools Used: E-mail marketing,Graphic design software,video editing software,data visualization tools

Algorithm:-

Machine Learning & Linear Regression:-

The device analysis of the illness based on the prediction made through the algorithm used by the device. The Naïve Bayes Classifier is used to make intelligent health predictions, and it is implemented in this way.

The increase of the health problems and the need for effective medical health care have led to an investigation of machine learning that can be applied in health problems. This paper presents a recent systematic review of machine learning approaches in predicting health problems. Furthermore, we will discuss the challenges, limitations, and future directions for the application of machine learning in the health field.

We collect research articles and studies that are related to the machine learning approaches in predicting mental health problems by searching reliable databases. Moreover, we adhere to the PRISMA methodology in conducting this systematic review. We include a total of 30 research articles in this review after the screening and identification processes.

Then, we categorize the collected research articles based on the mental health problems such as schizophrenia, bipolar disorder, anxiety and depression, posttraumatic stress disorder, and mental health problems among children. Discussing the findings, we reflect on the challenges and limitations faced by the researchers on machine learning in health problems. Additionally, we provide concrete recommendations on the potential future research and development of applying machine learning in the health field.

Data Set :-

<https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>

PROCEDURE:

FOR PREDICT THE FUTURE CHANGES IN PUBLIC HEALTH AWARENESS USING RANDOM FOREST.

STEPS INVOLVED IN RANDOM FOREST:

1. DATA COLLECTION AND PREPARATION:
   * Data collection is the primary step in preparation for analysis. Gather historical data that includes both predictor variable (features) and the target variables (what you want to predict). In this we predict the future awareness that a company is about to make.

Clean and preprocess the data by handling the process

1. DATA SPLITTING:
   * Split the dataset into two parts (training and testing) or (validation set). The training set used to train the Random Forest model and testing set will be used to evaluate its performance
2. FEATURE SELECTION:
   * In large number of Datas. It includes such as Timestamp, Age, Reviews, Country, etc.
3. RANDOM FOREST MODEL TRANING:
   * Train the random forest model using the Training data. This involves building an ensemble of decision trees.

Random Forest is an ensemble of multiple decision trees. Each tree is trained on random subset of the features.

This randomness reduces overfitting.

1. MODEL EVALUATIONN:
   * Use the testing set to evaluate the model’s performance. Common evaluation metrics include MAE, MSE, or R-squared for regression tasks and accuracy, precision, recall.
2. PREDICT FUTURE VALUES:
   * Once the model is trained and evaluated, you can use it to make predictions on new, unseen data for future time periods.
3. MONITERING AND DELPOYMENT:
   * Moniter the real-time changes to continuous predictions. And deploy the model in real-time if your satisfied with the performance and requirements.

# Display the pivot table

Print(“\nPivot Table (Cross-tabulation):”)

Print(pivot\_table)

Average count per route and Stop names

DataFrame with daily aggregation:

Timestamp Age

1. 2014-08-27 100000019373
2. 1 2014-08-28 9105

2 2014-08-29 5258

3 2014-08-30 744

4 2014-08-31 229

.. … …

519 2016-01-28 0

520 2016-01-29 0

521 2016-01-30 0

522 2016-01-31 0

523 2016-02-01 25

[524 rows x 2 columns]

Average Passenger Count per Route:

Age Timestamp

1. -1726 2014-08-28 10:07:53.000000000
2. 1 -29 2014-08-27 12:39:14.000000000
3. -1 2014-08-30 20:55:11.000000000
4. 3 5 2014-08-28 10:35:55.000000000
5. 8 2014-08-29 09:10:58.000000000
6. 5 11 2014-08-29 17:26:15.000000000
7. 18 2014-08-27 13:55:27.142856960
8. 7 19 2014-08-28 09:29:48.666666496
9. 20 2014-08-31 16:34:02.166666496
10. 9 21 2014-08-28 03:44:11.500000000
11. 22 2014-09-08 20:08:46.857142784
12. 11 23 2014-09-18 11:51:20.137254912
13. 24 2014-09-16 12:07:56.804347904
14. 13 25 2014-09-15 03:30:46.016393472
15. 26 2014-09-03 00:32:46.706666496
16. 15 27 2014-09-03 03:34:02.774648064
17. 28 2014-09-17 15:33:52.264706048
18. 17 29 2014-09-04 03:27:23.741176320
19. 30 2014-09-11 10:39:39.365079296
20. 19 31 2014-09-10 18:57:21.029850624
21. 32 2014-09-15 10:09:05.621951232
22. 21 33 2014-09-02 07:35:54.114285824
23. 34 2014-09-20 12:39:00.076922880
24. 23 35 2014-09-01 06:00:31.163636480
25. 36 2014-09-23 11:33:12.891891968
26. 25 37 2014-09-01 14:22:31.046511616
27. 38 2014-09-09 23:20:08.384615424
28. 27 39 2014-09-08 11:58:44.121212160
29. 40 2014-09-02 07:01:47.696969728
30. 29 41 2014-09-20 03:15:48.000000000
31. 42 2014-09-06 06:32:48.100000000
32. 31 43 2014-09-06 11:23:24.000000000
33. 44 2014-09-09 08:50:29.818181888
34. 33 45 2014-09-12 04:36:31.000000000
35. 46 2014-11-04 19:02:18.833333248
36. 35 47 2014-08-28 04:35:47.500000000
37. 48 2014-08-28 12:05:29.833333504
38. 37 49 2014-08-28 15:44:40.249999872
39. 50 2014-08-28 08:28:13.833333504
40. 39 51 2014-08-29 14:22:40.600000000
41. 53 2014-08-28 18:21:23.000000000
42. 41 54 2014-08-27 19:48:30.666666752
43. 55 2014-08-28 04:13:39.333333248
44. 43 56 2014-09-14 17:23:32.500000000
45. 57 2014-08-27 19:59:27.333333248
46. 45 58 2014-08-27 16:13:40.000000000
47. 60 2014-11-26 00:25:36.500000000
48. 47 61 2014-08-29 01:20:32.000000000
49. 62 2014-08-27 17:12:01.000000000
50. 49 65 2014-08-27 19:17:07.000000000
51. 72 2014-10-02 21:25:16.000000000
52. 51 329 2014-08-27 15:05:21.000000000
53. 99999999999 2014-08-27 15:24:47.000000000
54. <ipython-input-9-9f484f111fe9>:4: FutureWarning: The default value of numeric\_only in DataFrameGroupBy.sum is deprecated. In a future version, numeric\_only will default to False. Either specify numeric\_only or select only columns which should be valid for the function.

Data\_daily\_aggregated = data.groupby(pd.Grouper(key=’Timestamp’, freq=’D’)).sum().reset\_index()

<ipython-input-9-9f484f111fe9>:11: FutureWarning: The default value of numeric\_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric\_only will default to False. Either specify numeric\_only or select only columns which should be valid for the function.

Data\_route\_avg = data.groupby(‘Age’).mean().reset\_index()

Pivot Table (Cross-tabulation):

Country \

Age -1726 -29 -1 5

Timestamp

2014-08-27 11:29:31 NaN NaN NaN NaN

2014-08-27 11:29:37 NaN NaN NaN NaN

2014-08-27 11:29:44 NaN NaN NaN NaN

2014-08-27 11:29:46 NaN NaN NaN NaN

2014-08-27 11:30:22 NaN NaN NaN NaN

… … … … …

2015-09-12 11:17:21 NaN NaN NaN NaN

2015-09-26 01:07:35 NaN NaN NaN NaN

2015-11-07 12:36:58 NaN NaN NaN NaN

2015-11-30 21:25:06 NaN NaN NaN NaN

2016-02-01 23:04:31 NaN NaN NaN NaN

\

Age 8 11 18 19

Timestamp

2014-08-27 11:29:31 NaN NaN NaN NaN

2014-08-27 11:29:37 NaN NaN NaN NaN

2014-08-27 11:29:44 NaN NaN NaN NaN

2014-08-27 11:29:46 NaN NaN NaN NaN

2014-08-27 11:30:22 NaN NaN NaN NaN

… … … … …

2015-09-12 11:17:21 NaN NaN NaN NaN

2015-09-26 01:07:35 NaN NaN NaN NaN

2015-11-07 12:36:58 NaN NaN NaN NaN

2015-11-30 21:25:06 NaN NaN NaN NaN

2016-02-01 23:04:31 NaN NaN NaN NaN

… work\_interfere \

Age 20 21 … 56

Timestamp …

2014-08-27 11:29:31 NaN NaN … NaN

2014-08-27 11:29:37 NaN NaN … NaN

2014-08-27 11:29:44 NaN NaN … NaN

2014-08-27 11:29:46 NaN NaN … NaN

2014-08-27 11:30:22 NaN NaN … NaN

… … … … …

2015-09-12 11:17:21 NaN NaN … NaN

2015-09-26 01:07:35 NaN NaN … NaN

2015-11-07 12:36:58 NaN NaN … NaN

2015-11-30 21:25:06 NaN NaN … NaN

2016-02-01 23:04:31 NaN NaN … NaN

\

Age 57 58 60 61

Timestamp

2014-08-27 11:29:31 NaN NaN NaN NaN

2014-08-27 11:29:37 NaN NaN NaN NaN

2014-08-27 11:29:44 NaN NaN NaN NaN

2014-08-27 11:29:46 NaN NaN NaN NaN

2014-08-27 11:30:22 NaN NaN NaN NaN

… … … … …

2015-09-12 11:17:21 NaN NaN NaN NaN

2015-09-26 01:07:35 NaN NaN NaN NaN

2015-11-07 12:36:58 NaN NaN NaN NaN

2015-11-30 21:25:06 NaN NaN NaN NaN

2016-02-01 23:04:31 NaN NaN NaN NaN

\

Age 62 65 72 329

Timestamp

2014-08-27 11:29:31 NaN NaN NaN NaN

2014-08-27 11:29:37 NaN NaN NaN NaN

2014-08-27 11:29:44 NaN NaN NaN NaN

2014-08-27 11:29:46 NaN NaN NaN NaN

2014-08-27 11:30:22 NaN NaN NaN NaN

… … … … …

2015-09-12 11:17:21 NaN NaN NaN NaN

2015-09-26 01:07:35 NaN NaN NaN NaN

2015-11-07 12:36:58 NaN NaN NaN NaN

2015-11-30 21:25:06 NaN NaN NaN NaN

2016-02-01 23:04:31 NaN NaN NaN NaN

Age 99999999999

Timestamp

2014-08-27 11:29:31 NaN

2014-08-27 11:29:37 NaN

2014-08-27 11:29:44 NaN

2014-08-27 11:29:46 NaN

2014-08-27 11:30:22 NaN

… …

2015-09-12 11:17:21 NaN

2015-09-26 01:07:35 NaN

2015-11-07 12:36:58 NaN

2015-11-30 21:25:06 NaN

2016-02-01 23:04:31 NaN

[1246 rows x 1325 columns]

DATA REDUCTION

In [10]:

Print(“\nSampled DataFrame:\n”)

Print(data)

# Technique 2: Aggregating data (e.g., weekly aggregation)

Data[‘coworkers’] = pd.to\_datetime(data[‘Age’])

Data\_weekly\_aggregated = data.groupby(pd.Grouper(key=’coworkers’, freq=’W’)).sum().reset\_index()

# Display the dataframe with weekly aggregation

Print(“\nDataFrame with weekly aggregation:”)

Print(data\_weekly\_aggregated)

Sampled DataFrame:

Timestamp Age Gender Country state self\_employed \

0 2014-08-27 11:29:31 37 Female United States IL NaN

1 2014-08-27 11:29:37 44 M United States IN NaN

2 2014-08-27 11:29:44 32 Male Canada NaN NaN

3 2014-08-27 11:29:46 31 Male United Kingdom NaN NaN

4 2014-08-27 11:30:22 31 Male United States TX NaN

… … … … … … …

1254 2015-09-12 11:17:21 26 male United Kingdom NaN No

1255 2015-09-26 01:07:35 32 Male United States IL No

1256 2015-11-07 12:36:58 34 male United States CA No

1257 2015-11-30 21:25:06 46 f United States NC No

1258 2016-02-01 23:04:31 25 Male United States IL No

Family\_history treatment work\_interfere no\_employees … \

0 No Yes Often 6-25 …

1 No No Rarely More than 1000 …

2 No No Rarely 6-25 …

3 Yes Yes Often 26-100 …

4 No No Never 100-500 …

… … … … … …

1254 No Yes NaN 26-100 …

1255 Yes Yes Often 26-100 …

1256 Yes Yes Sometimes More than 1000 …

1257 No No NaN 100-500 …

1258 Yes Yes Sometimes 26-100 …

Leave mental\_health\_consequence phys\_health\_consequence \

0 Somewhat easy No No

1 Don’t know Maybe No

2 Somewhat difficult No No

3 Somewhat difficult Yes Yes

4 Don’t know No No

… … … …

1254 Somewhat easy No No

1255 Somewhat difficult No No

1256 Somewhat difficult Yes Yes

1257 Don’t know Yes No

1258 Don’t know Maybe No

Coworkers supervisor mental\_health\_interview \

0 Some of them Yes No

1 No No No

2 Yes Yes Yes

3 Some of them No Maybe

4 Some of them Yes Yes

… … … …

1254 Some of them Some of them No

1255 Some of them Yes No

1256 No No No

1257 No No No

1258 Some of them No No

Phys\_health\_interview mental\_vs\_physical obs\_consequence comments

0 Maybe Yes No NaN

1 No Don’t know No NaN

2 Yes No No NaN

3 Maybe No Yes NaN

4 Yes Don’t know No NaN

… … … … …

1254 No Don’t know No NaN

1255 No Yes No NaN

1256 No No No NaN

1257 No No No NaN

1258 No Don’t know No NaN

[1259 rows x 27 columns]

DataFrame with weekly aggregation:

Coworkers Age

0 1970-01-04 100000038724

<ipython-input-10-3dfb27f44d71>:6: FutureWarning: The default value of numeric\_only in DataFrameGroupBy.sum is deprecated. In a future version, numeric\_only will default to False. Either specify numeric\_only or select only columns which should be valid for the function.

Data\_weekly\_aggregated = data.groupby(pd.Grouper(key=’coworkers’, freq=’W’)).sum().reset\_index()

In [11]:

Plt.figure(figsize=(12,7))

Sns.heatmap(data.corr(),annot=True,cmap=’pink’)

Plt.plot()

<ipython-input-11-000db9519e9d>:2: FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.

Sns.heatmap(data.corr(),annot=True,cmap=’pink’)

CONCLUSION:-

The preprocessing of the data for Public health awareness is a critical step that lays the foundation for meaningfull insights and decision making.Through Data cleaning, Transformation and Integration.we ensure that the data accurate , complete and relevant. Proper preprocessing not only enhances the quality of analysis but also paves the way for the development of effective strategies to improve public transportation system and ultimately benefit communities and the environment.

# importing the required libraries

Import matplotlib.pyplot as plt

Import numpy as np

# define data values

X = np.array([3,2,5,3,1,4]) # X-axis points

Y = np.array([1,1,3,2,1,2]) # Y-axis points

Plt.plot(x, y) # Plot the chart

Plt.show() # display

In [15]:

Import pandas as pd

#create dataset

Df = pd.DataFrame({‘age’: [1, 2, 4, 5, 5, 6, 6, 7, 8, 10, 11, 11, 12, 12, 14],

‘review’: [64, 66, 76, 73, 74, 81, 83, 82, 80, 88, 84, 82, 91, 93, 89]})

In [8]:

Import matplotlib.pyplot as plt

Plt.scatter(df.hours, df.score)

Plt.title(‘Age vs. Review’)

Plt.xlabel(‘Hours’)

Plt.ylabel(‘Score’)

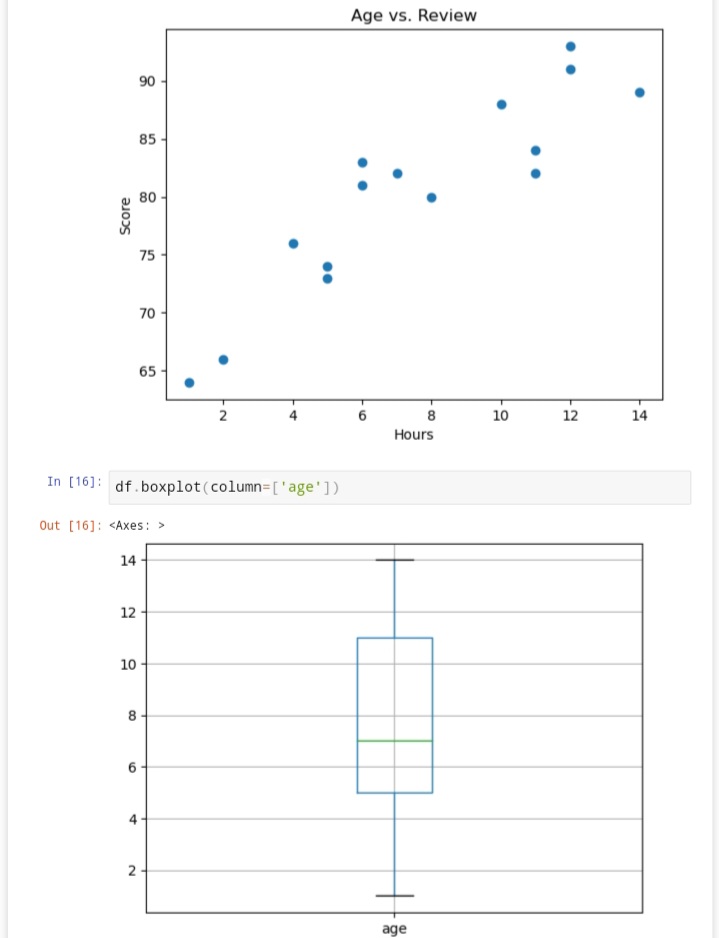
Plt.show()

In [16]:

Df.boxplot(column=[‘age’])

Out [16]:

<Axes: >



In [18]:

Import statsmodels.api as sm

#define response variable

Y = df[‘age’]

#define explanatory variable

X = df[[‘review’]]

#add constant to predictor variables

X = sm.add\_constant(x)

#fit linear regression model

Model = sm.OLS(y, x).fit()

#view model summary

Print(model.summary())

OLS Regression Results

Dep. Variable: age R-squared: 0.831

Model: OLS Adj. R-squared: 0.818

Method: Least Squares F-statistic: 63.91

Date: Wed, 25 Oct 2023 Prob (F-statistic): 2.25e-06

Time: 21:10:43 Log-Likelihood: -27.941

No. Observations: 15 AIC: 59.88

Df Residuals: 13 BIC: 61.30

Df Model: 1

Covariance Type: nonrobust

Coef std err t P>|t| [0.025 0.975]

Const -26.1024 4.238 -6.160 0.000 -35.257 -16.947

Review 0.4192 0.052 7.995 0.000 0.306 0.532

Omnibus: 0.328 Durbin-Watson: 1.415

Prob(Omnibus): 0.849 Jarque-Bera (JB): 0.448

Skew: 0.262 Prob(JB): 0.799

Kurtosis: 2.335 Cond. No. 792.

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

C:\Users\ELCOT\Anaconda\Lib\site-packages\scipy\stats\\_stats\_py.py:1806: UserWarning: kurtosistest only valid for n>=20 … continuing anyway, n=15

Warnings.warn(“kurtosistest only valid for n>=20 …

In the conclusion,

Objective: The objective of the public health awareness project is to raise awareness about a specific health issue within a target population, ultimately promoting healthier behaviors and improving overall community health.Design Thinking Process:Empathize:Understand the target audience, their needs, concerns, and existing knowledge about the health issue.Conduct surveys, interviews, and observations to gather insights

Define:Clearly define the health Issue and its significance.Identify key messages and behaviors that need to be promoted.Ideate:Brainstorm creative and effective ways to convey the messages to the audience.Generate a range of ideas for awareness campaigns and interventions.

Prototype:Develop a prototype or pilot program based on selected ideas

This could include creating visual materials, educational content, or interactive tools.Test:Test the prototype with a small group from the target population.Gather feedback and make necessary adjustments based on their responses.Development Phases:

1. Research and Planning:

Gather data and information about the health issue.Identify key stakeholders and partners.Develop a project plan, including goals, budget, and timeline.

1. Message and Content Creation:

Craft compelling and culturally sensitive messages.Create visual and written content, such as pamphlets, videos, websites, and social media posts.

1. Distribution and Outreach:

Determine the best channels to reach the target audience (e.g., community events, social media, local clinics).

Develop an outreach strategy to disseminate information effectively

d.Implementation:

Execute the awareness campaign or intervention as planned.Monitor its progress and make real-time adjustments as needed.

1. Evaluation:

Assess the impact of the project through various metrics (e.g., increased knowledge, changed behaviors, reduced disease incidence).Collect feedback from the target population to measure effectiveness.

1. Sustainability and Scaling:

Develop a plan to sustain the project’s impact over time.If successful, consider scaling the project to reach a wider audience or expand to other communities.

Insights from data analysis can play a crucial role in measuring campaign effectiveness and guiding future strategies for marketing and advertising efforts. Here’s how these insights can be used in this context:

KPI Monitoring:

Establish key performance indicators (KPIs) that align with campaign goals, such as click-through rates (CTR), conversion rates, return on investment (ROI), and customer acquisition cost (CAC).Regularly monitor these KPIs during and after the campaign to assess its performance.

Comparative Analysis:

Compare campaign performance against historical data or benchmarks from previous campaigns. This provides context for evaluating the current campaign’s success.

Segmentation Analysis:

Analyze the campaign’s impact on different audience segments. Are certain demographics or customer groups responding better to the campaign? Use this information to tailor future campaigns to specific segments.

Conversion Funnel Analysis:

Track users’ journey through the conversion funnel (e.g., awareness, consideration, conversion).Identify drop-off points and bottlenecks in the funnel to optimize the user experience.

Attribution Analysis:

Use attribution models to understand which touchpoints or channels are driving conversions.Determine the role of each channel (e.g., social media, email, search) in the customer’s decision-making process.A/B Testing:Conduct A/B tests to compare different campaign elements (e.g., ad copy, visuals, landing pages).

Analyze which variations perform better and implement those findings in future campaigns.

Customer Lifetime Value (CLV):

Calculate the CLV of customers acquired through the campaign. This helps assess the long-term impact and profitability of the campaign.

Sentiment Analysis:

Analyze social media and customer feedback sentiment to understand public perception of the campaign.Adjust strategies based on the sentiment and address any negative feedback.

Modeling:

Use predictive analytics to forecast future campaign outcomes based on historical data.

Adjust strategies based on these forecasts.By analyzing these aspects, you can measure the effectiveness of a campaign and gain valuable insights for future strategies:

Optimization:

Identify areas that need improvement and refine strategies for better results.

Using IBM Cognos for Data Visualization:

Data Preparation:

Ensure your data is accessible to IBM Cognos. This can be through a database connection, data files, or other supported data sources.

Connect to Data:

Open IBM Cognos and establish a connection to your data source.

Data Modeling:

Define the data model by selecting the relevant tables and fields for your analysis.

Report Design:

Create a new report.Add data items to your report by dragging and dropping fields from the data model onto the report canvas.

Visualization Creation:

Build visualizations by selecting the appropriate chart types (e.g., bar charts, line charts, pie charts) and configuring them with the data items you’ve added.Filtering and Interactivity:Add filters, prompts, and interactive elements to allow users to explore the data dynamically.

Dashboard Design:

If needed, create a dashboard by combining multiple visualizations and reports into a single view.Scheduled Reports:Set up automated report generation and distribution to relevant stakeholders.

Data Analysis:

Use pandas for data analysis tasks, such as aggregating data, filtering, and performing calculations.

Data Visualization:

Create data visualizations using matplotlib or Seaborn. For example, you can create bar charts, scatter plots, and line graphs to visualize your data.

Statistical Analysis:

Utilize Python libraries like NumPy and SciPy for statistical analysis if needed.Machine Learning (Optional):If your analysis involves predictive modeling or machine learning, you can use libraries like scikit-learn to build and evaluate models.

Generate Reports:

Use Jupyter Notebook or other reporting tools to create reports summarizing your analysis.

In this project,

We analysis the public health awareness up and downs