|  |  |  |
| --- | --- | --- |
| NicePng_national-emblem-of-india_9441929.png | **PUBLIC HEALTH CAMPAIGN AWARENESS** |  |

**NAAN MUDHALVAN PROJECT REPORT**

Submitted by

**DHIVYA M (730321243006)**

**NITISRI S (730321243017)**

**RETHANYA C (730321243021)**

**THRISHNA S (730321243031)**

**FIFTH SEMESTER**

**ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

**BUILDERS ENGINEERING COLLEGE , KANGEYAM**

**ANNA UNIVERSITY :: CHENNAI 600 025**

**NOVEMBER 2023**



**BONAFIDE CERTIFICATE**

Certified that this is a Bonafide record of work done by **DHIVYA M(730321243006) , NITISRI S(730321243017)** and **RETHANYA C(730321243021) , THRISHNA S(730321243031)** in NAAN MUTHALVAN during the Academic year 2023- 24 for fifth Semester.

|  |  |
| --- | --- |
|  |  |
| **STAFF -INCHARGE** | **HEAD OF THE DEPARTMENT** |
|  |  |
| Submitted for the Naan mudhalvan | viva voce held on\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |  |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER** |

**ABSTRACT**

Public health campaigns are crucial in addressing pressing health issues, raising awareness, and promoting positive behavioural changes among individuals and communities. Childhood obesity is a complex and multifaceted problem with far-reaching consequences, including increased risk of chronic diseases and reduced quality of life. This campaign is designed to increase awareness and encourage healthy lifestyle choices among children and their families.The campaign employs a multi-faceted approach, combining traditional and digital marketing, community engagement, and educational initiatives to reach a broad audience. It leverages social media platforms, community events, and partnerships with local schools and healthcare providers to disseminate key messages and resources.

Key campaign components include:

Educational Resources: Providing easily accessible materials and resources to help parents and caregivers make informed decisions regarding their children's nutrition and physical activity.

Social Media Engagement: Leveraging popular social media platforms to share health tips, success stories, and healthy recipes, while also creating a sense of community and support for those striving to make healthier choices.

Evaluation and Metrics: Implementing measurable goals and key performance indicators (KPIs) to assess the campaign's impact, ensuring that it is effectively contributing to reducing childhood obesity rates and improving overall public health.

This campaign aims to create a sustainable culture of health and wellness within the community, with the ultimate goal of reducing childhood obesity rates and improving long-term health outcomes.

|  |
| --- |
| **TABLE OF CONTENT** |
| **CHAPTER TITLE PAGE NO.** |
| **NO.** |
| 1. **PROBLEM STATEMENT 5** |
| **2 DESIGN THINKING 6** |
| **3 DATASET DEFINITION 8** |
| **4 DATA PREPROCESSING AND 10**  **FEATURE EXTRACTION** |
| **5 PROPOSED ALGORITHM 14**    **6 PROPOSED INNOVATION 18**  **TECHNIQUE**  **7 CONCLUSION AND 22**  **FUTURE SCOPE** |
|  |

**CHAPTER 1**

**PROBLEM STATEMENT**

A lack of widespread awareness hinders the effectiveness of public welfare programs and initiatives, preventing them from reaching their full potential. This issue arises from inadequate communication, limited outreach, and insufficient education about the existence, purpose, and benefits of various public welfare programs among the target audience.

**Key Challenges:**

**1.Low Awareness Levels:**

Many individuals, especially those in marginalized or underserved communities, lack awareness of available public welfare programs. Limited understanding of eligibility criteria and application procedures for welfare programs.

**2. Ineffective Communication Channels:**

Current communication channels may not effectively reach the target audience, leading to information gaps. Insufficient utilization of modern communication tools and technologies.

**3. Diverse Target Audience:**

Public welfare programs cater to a diverse demographic, and one-size-fits-all approaches may not effectively engage different segments of the population. Tailoring communication strategies to diverse cultural, linguistic, and socioeconomic backgrounds is a challenge.

**4**. **Stigma and Mistrust:**

Stigma associated with seeking assistance and mistrust in public institutions can discourage individuals from accessing welfare programs. Overcoming negative perceptions and building trust is crucial for effective outreach.

**5. Limited Resources for Awareness Campaigns:**

Budget constraints may limit the resources allocated for public awareness campaigns, affecting the reach and frequency of messaging. Finding cost-effective strategies for maximum impact is a significant challenge.

**CHAPTER 2**

**DESIGN THINKING**

**ANALYSIS OBJECTIVES**

When applying design thinking to public welfare awareness, the analysis objectives play a crucial role in understanding the effectiveness of the designed solutions. Here are key analysis objectives aligned with each phase of the design thinking process:

1. **Empathize:**

**Objective**: Understand the perspectives, needs, and challenges of the target audience to inform the design process.

**Metrics**: Conduct surveys and interviews to gather qualitative data on current awareness levels, perceptions, and barriers.

1. **Define:**

**Objective:** Clearly define the problem of low public welfare awareness and identify key stakeholders.

**Metrics**: Develop a problem statement that includes measurable indicators, such as the percentage of the population unaware of specific welfare programs.

1. **Ideate:**

**Objective:** Generate creative and diverse ideas for raising awareness.

**Metrics**: Evaluate the quantity and diversity of ideas generated during brainstorming sessions. Focus on ideation sessions that lead to innovative and feasible concepts.

**4. Prototype:**

**Objective:** Develop and test prototypes of potential awareness campaigns.

**Metrics:** Assess the effectiveness of prototypes in small-scale pilot programs. Measure participant engagement, feedback, and any observable changes in awareness or behavior.

**5**. **Test:**

**Objective:** Collect feedback and iterate based on real-world insights.

**Metrics:** Evaluate the success of the pilot programs by analyzing participant feedback, survey data, and any observable impact on awareness levels or community engagement.

**6. Implement:**

**Objective:** Scale up successful awareness strategies for broader implementation.

**Metrics:** Measure the reach and impact of the scaled-up campaign. Track participation rates, changes in awareness, and collaboration with partners**.**

**7. Evaluate:**

**Objective:** Assess the impact of the awareness campaign on a larger scale.

**Metrics**: Use quantitative and qualitative data to evaluate the success of the campaign. Measure changes in public awareness, behavior, and participation in welfare programs.

**8. Learn and Iterate:**

**Objective:** Document lessons learned and iterate for continuous improvement**.**

**Metrics:** Analyze feedback from all stages of the design process. Document insights, successes, and challenges. Track changes in awareness and participation over time to inform future iterations.

**CHAPTER 3**

**DATASET DEFINITION**

**SURVEY DATASET:**

Survey datasets are collections of data that are gathered through the administration of surveys or questionnaires. These datasets are commonly used in various fields, including social sciences, market research, healthcare, and more, to gather information about a particular topic, population, or research question. Here are some key points about survey datasets:**Data Collection:** Survey datasets are created by collecting responses to a set of questions or prompts. These questions can be open-ended or structured (multiple choice, Likert scales, etc.), and they are designed to elicit specific information from the respondents.**Population:** The dataset typically represents a specific population or group of individuals who are the target of the survey. This could be a general population, a specific demographic, or individuals with certain characteristics.**Sampling:** Survey datasets may be collected from the entire population (a census) or from a sample of the population. Sampling methods can include random sampling, stratified sampling, or convenience sampling, depending on the research goals and available resources.**Variables:** Survey datasets consist of variables, which are the characteristics or measures of interest. These variables can be categorical (e.g., gender, education level) or continuous (e.g., age, income).**Data Format:** Survey data can be collected in various formats, including paper-based surveys, online surveys, telephone interviews, or in-person interviews. The data can be stored in spreadsheets, databases, or other digital formats.**Data Cleaning:** Survey datasets often require extensive data cleaning and preprocessing to handle missing values, outliers, and ensure data consistency.**Analysis:** Survey datasets are used for statistical analysis, which can include descriptive statistics, inferential statistics, and data visualization. Researchers may seek to identify patterns, correlations, or trends in the data.**Privacy and Ethics:** Data privacy and ethical considerations are important when dealing with survey datasets. Researchers must often obtain informed consent from participants and ensure that the data is handled securely and anonymized to protect individual identities.**Documentation:** Good documentation is crucial for survey datasets. This includes details about the survey design, sampling method, question wording, and any transformations or recoding of variables.**Sharing and Accessibility:** Some survey datasets are publicly available for research purposes, while others are private or proprietary. Researchers may share their datasets through repositories or data archives.Popular sources of survey datasets include government agencies (e.g., U.S. Census Bureau), academic research projects, market research firms, and non-governmental organizations. These datasets play a significant role in empirical research and can provide valuable insights into various aspects of society, behavior, and economics.

**CHAPTER 4**

**DATA PREPROCESSING AND FEATURE EXTRACTION**

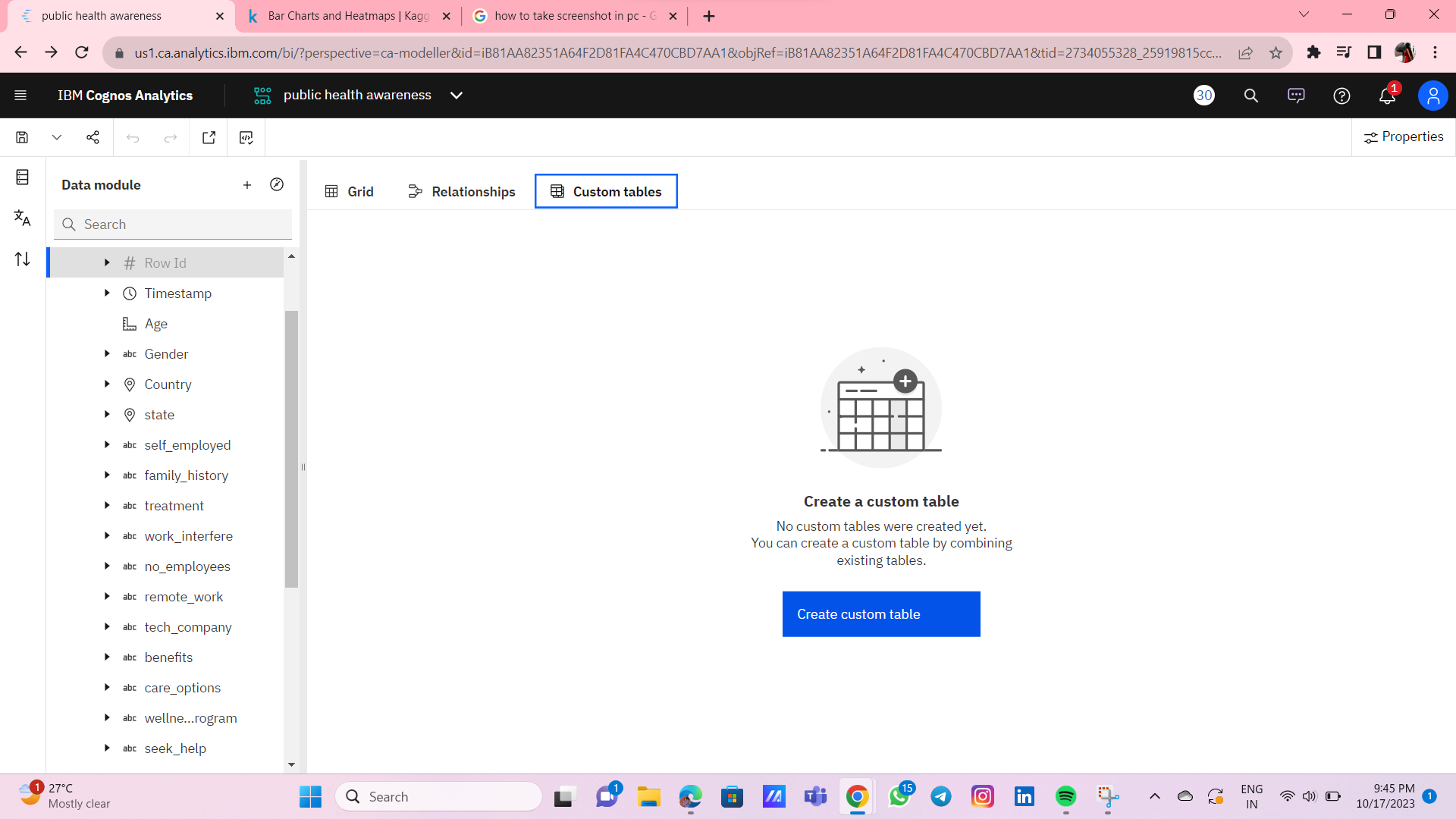
**DATA PREPROCESSING**

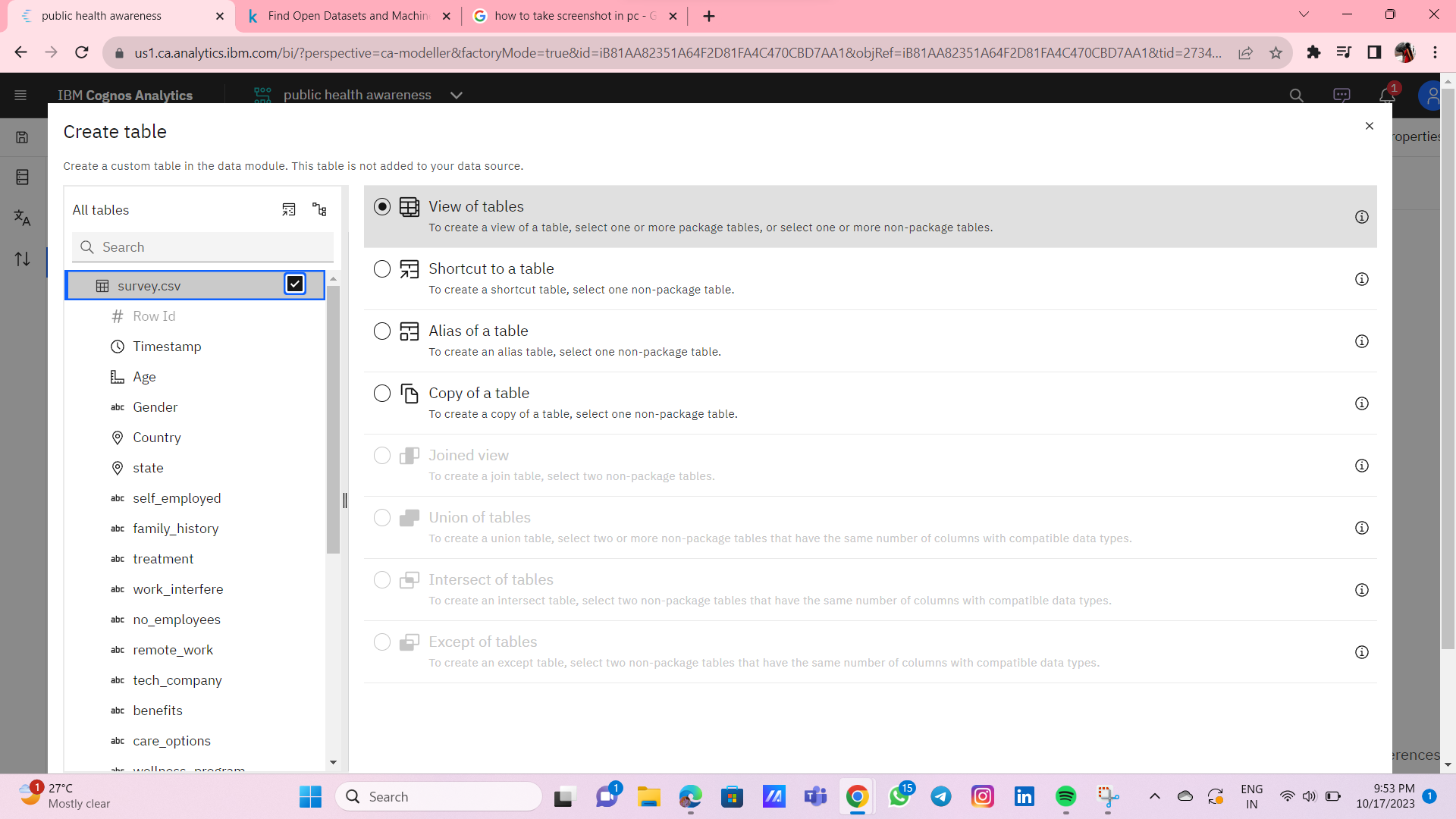
Data preprocessing is a crucial step in working with survey datasets. It involves cleaning and transforming the data to ensure it is accurate, complete, and suitable for analysis. Here are some common data preprocessing tasks specific to survey datasets:**Data Cleaning:Handling Missing Values:** Survey datasets often contain missing or incomplete responses. You can choose to impute missing values, remove records with missing data, or analyze missing data patterns.**Outlier Detection**: Identify and address outliers that may skew your analysis. Outliers can result from errors in data collection or be genuine but extreme values.**Data Validation:** Check for data entry errors, such as invalid responses, and correct or remove them as needed.**Consistency Checks:** Ensure that responses are consistent within and between variables. For example, check that age and birth year are consistent.Variable Transformation:**Categorical Encoding**: Convert categorical variables into numerical format, often using one-hot encoding or label encoding, depending on the variable's nature.**Normalization or Scaling:** Scale numerical variables if they have different ranges to ensure they have similar importance in the analysis.

**Feature Extraction:**

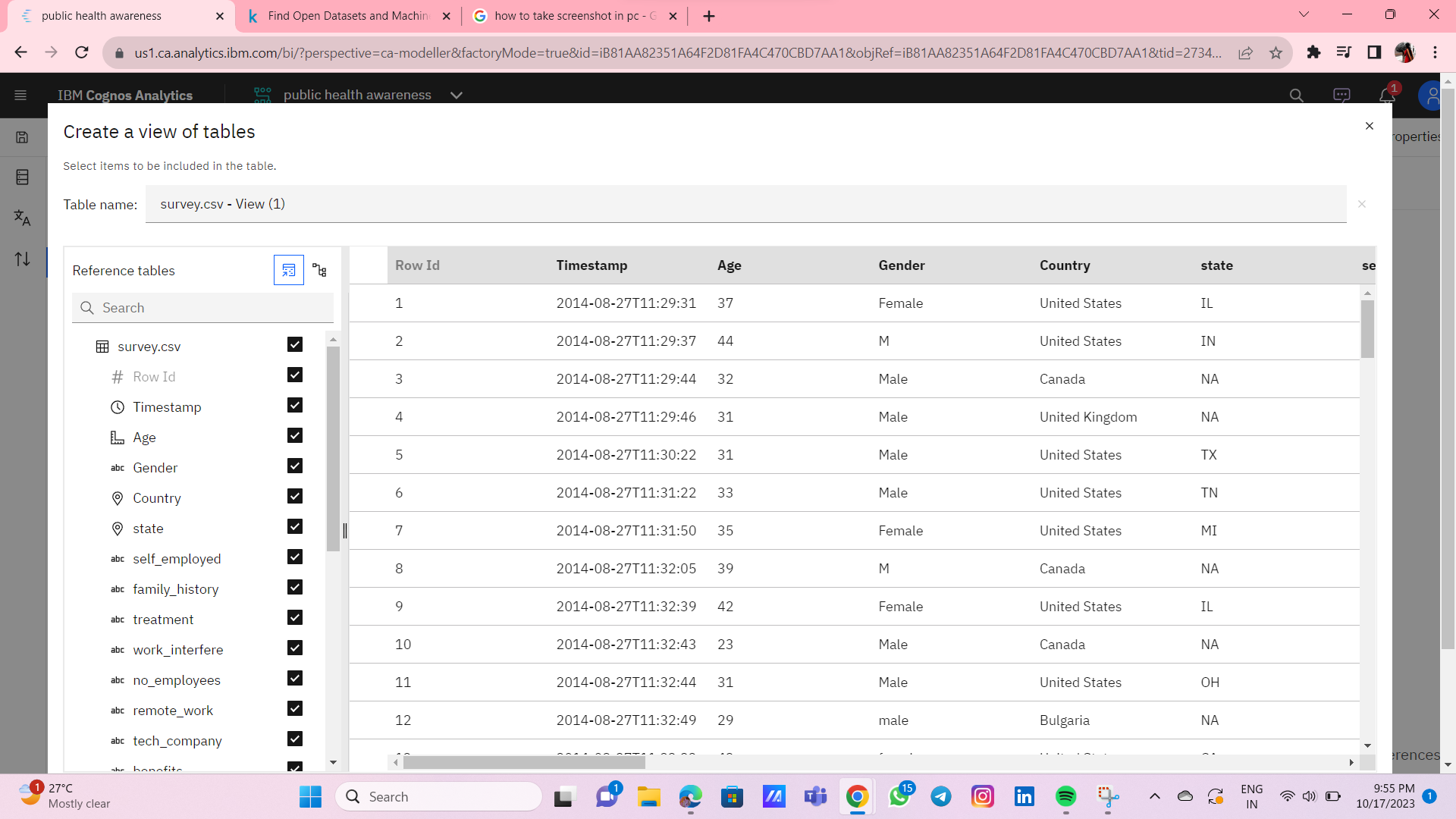
Decide which variables are relevant for your analysis and remove irrelevant or redundant ones. Feature selection methods can help you make these decisions.**Dimensionality Reduction:** Use techniques like principal component analysis (PCA) to reduce the dimensionality of your dataset while preserving important information.**Handling Categorical Variables:Dummy Variables:** Create dummy variables for categorical variables, especially if you plan to use them in regression or other modeling techniques.**Ordinal Encoding:** Encode ordinal categorical variables in a way that preserves their order and meaningfulness.**Dealing with Survey Weights:**Some survey datasets include sample weights to account for survey design and non-response bias. Researchers need to consider these weights in their analyses to ensure that the results are representative of the target population.**Data Imputation:**If missing data is prevalent and you decide not to remove records with missing values, you may use data imputation techniques to estimate missing values, such as mean imputation, median imputation, or machine learning-based imputation methods.**Data Standardization:**Standardize data if necessary, particularly for variables with different units or measurement scales.**Text Data Processing:**If the survey dataset contains text responses, text data preprocessing may involve tasks such as text cleaning, tokenization, stemming, and sentiment analysis.

**Documentation:**Maintain clear documentation of all preprocessing steps and the reasons for each transformation. This documentation is essential for transparency and reproducibility.**Privacy and Security:**Ensure that personally identifiable information (PII) is handled appropriately and that data privacy and security guidelines are followed**.**

****

**Fig 4.1**

**Fig 4.2**

****

**Fig 4.3**

**CHAPTER 5**

**PROPOSED ALGORITHM**

**SUPPORT VECTOR MACHINES (SVM)**

Support Vector Machines (SVM) are a class of supervised machine learning algorithms used for classification and regression tasks. They are powerful tools for both linear and nonlinear data analysis. SVMs work by finding a hyperplane that best separates data points into different classes, while maximizing the margin between the two classes. Here are the key concepts and components of Support **Vector Machines:Linear Separation:**SVMs are originally designed for binary classification problems, aiming to find a hyperplane that best separates two classes of data points.**Hyperplane:**In a two-dimensional space, a hyperplane is a line, and in higher-dimensional spaces, it's a flat affine subspace.**Margin:**The margin is the distance between the hyperplane and the nearest data point from either class. SVM aims to maximize this margin.**Support Vectors:**Support vectors are the data points that are closest to the hyperplane. These data points play a critical role in defining the margin and the decision boundary.**Kernel Trick:**SVMs can handle nonlinear data by transforming the input data into a higher-dimensional space. This transformation is often done using a kernel function (e.g., polynomial, radial basis function) that allows SVM to find nonlinear decision boundaries.**C Parameter:**The C parameter in SVM controls the trade-off between maximizing the margin and minimizing the classification error. A small C value prioritizes a wider margin (potentially allowing some misclassification), while a large C value prioritizes correctly classifying as many points as possible (potentially leading to a smaller margin).**Soft Margin SVM:**In practical applications, data may not be linearly separable. In such cases, a soft margin SVM allows for a certain degree of misclassification to find a balance between maximizing the margin and minimizing errors.**Multi-Class Classification:**SVMs can be extended to handle multi-class classification problems. Common techniques include one-vs-one and one-vs-all strategies**.**

**Regression (Support Vector Regression):**SVMs can also be used for regression tasks by fitting a hyperplane to predict a continuous output value. In this case, the margin represents an ε-insensitive tube around the predicted values.

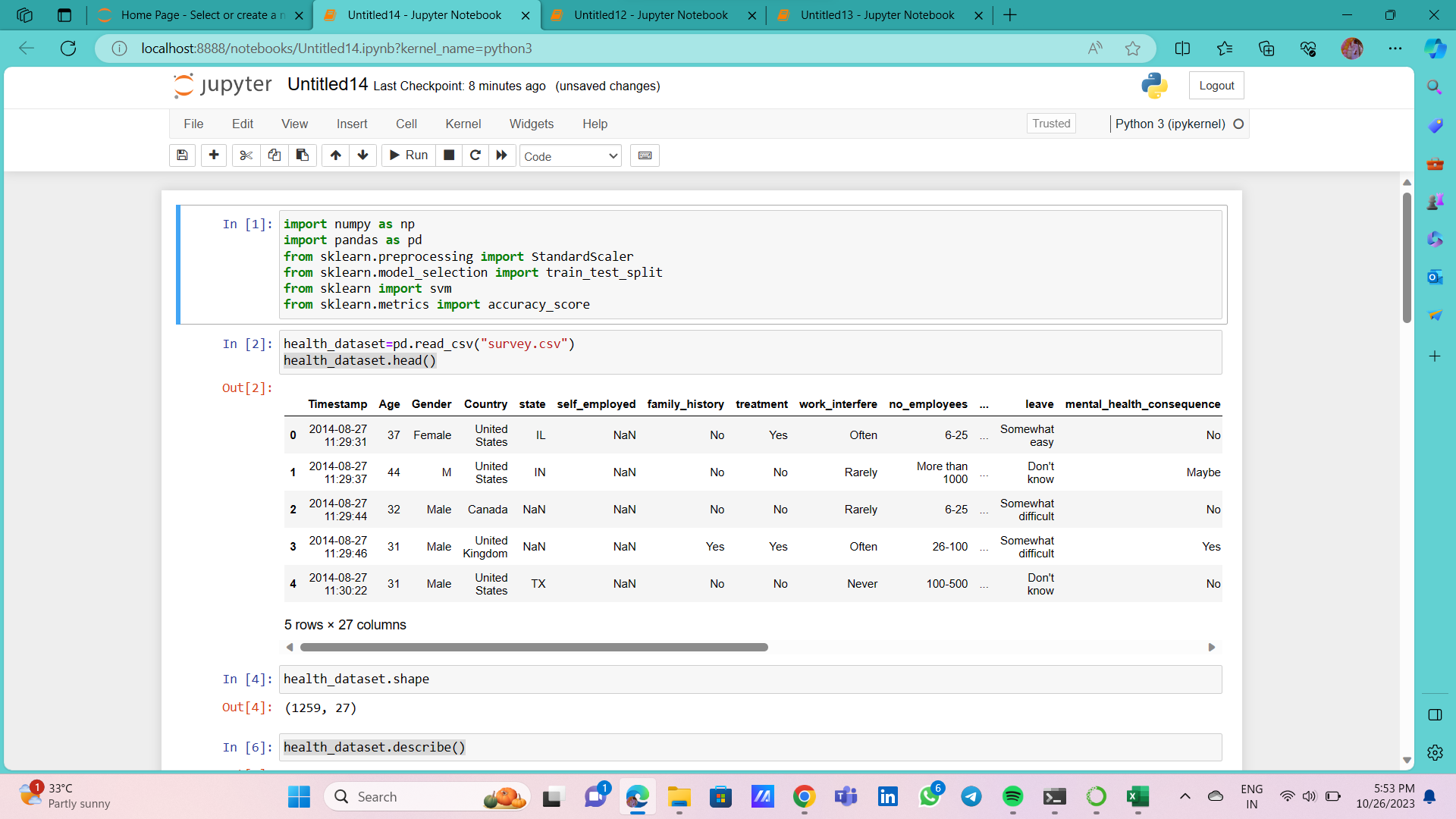


Fig 6.1

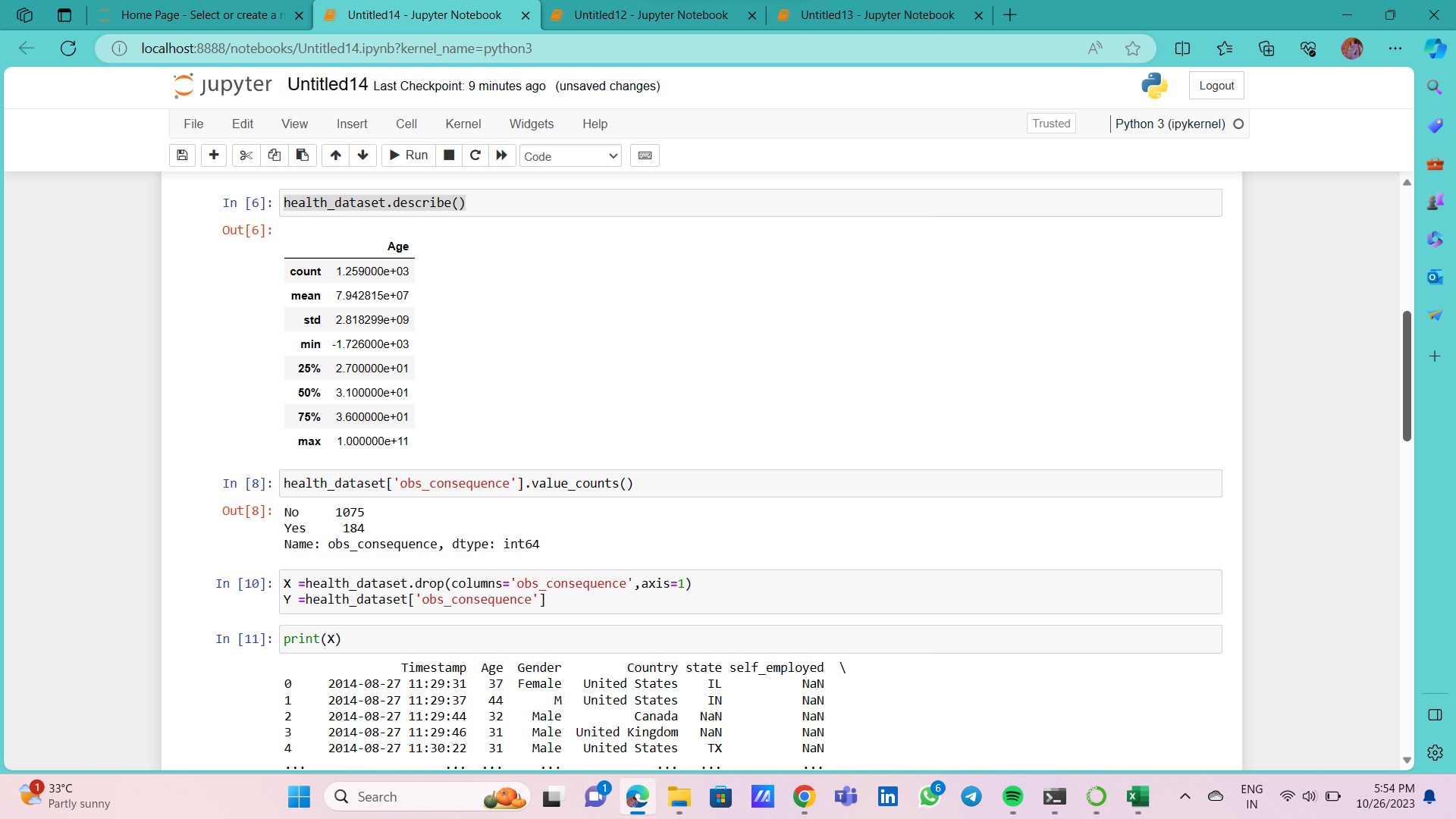


Fig 6.2

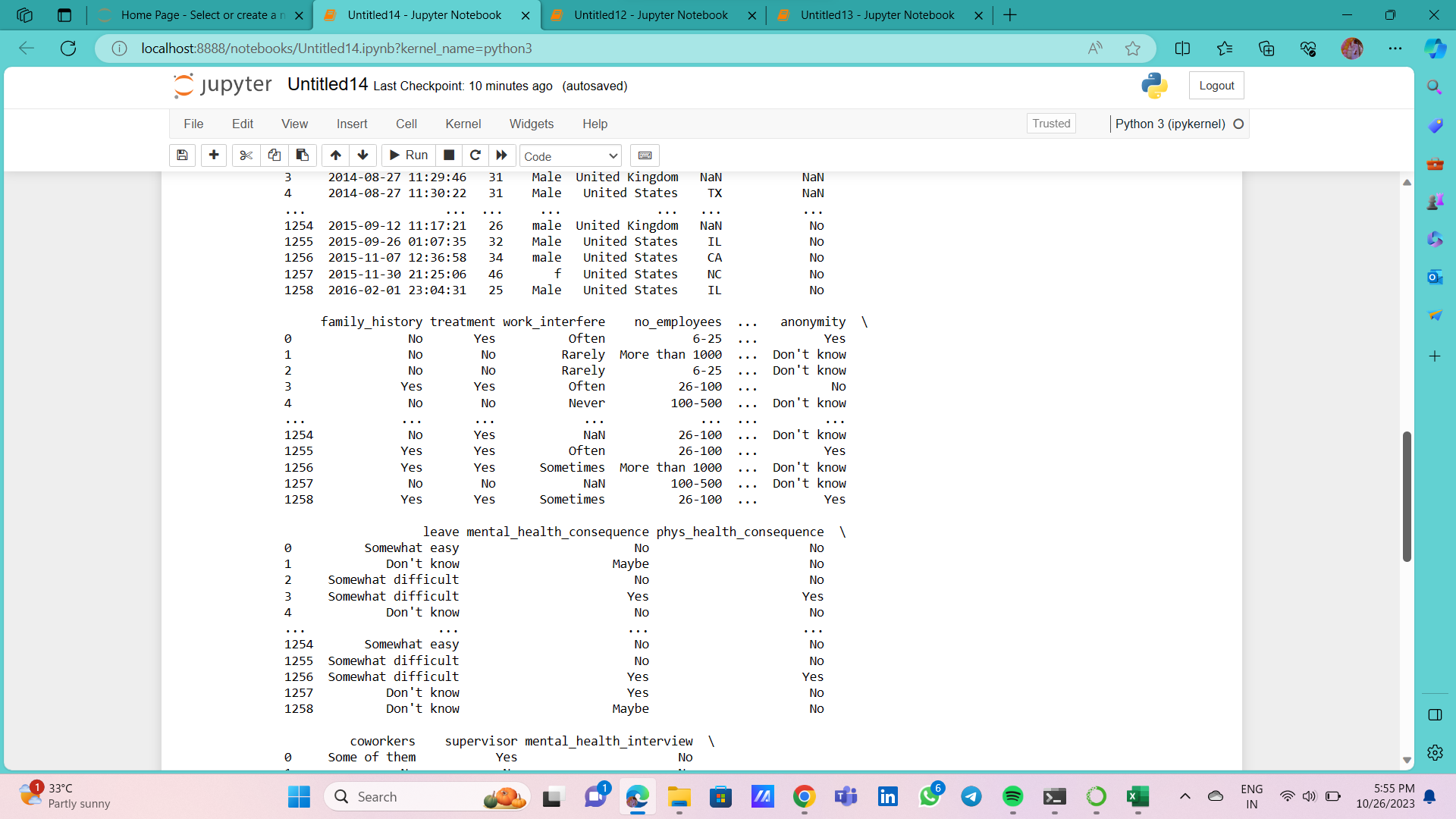


Fig 6.3

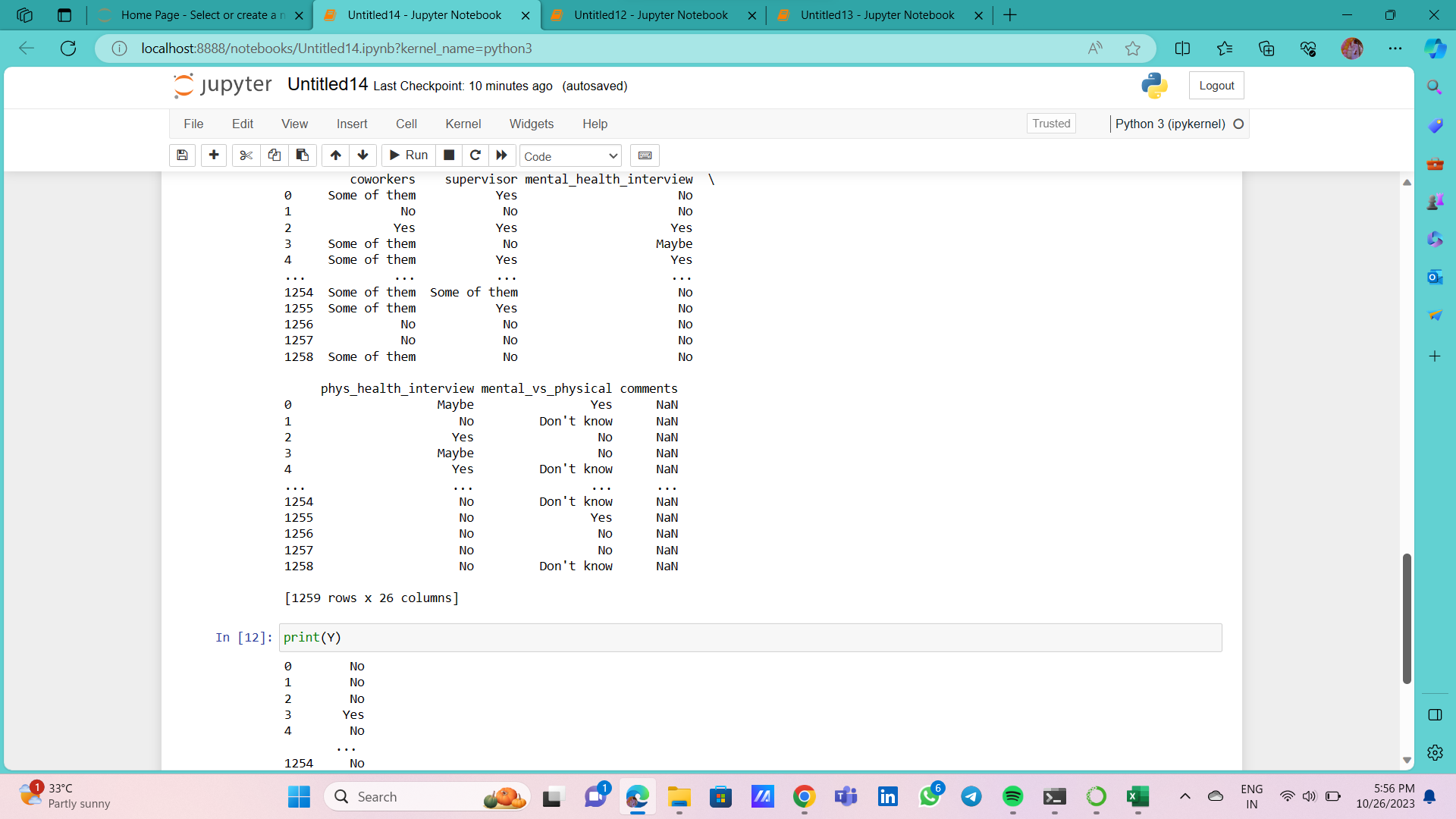


Fig 6.4

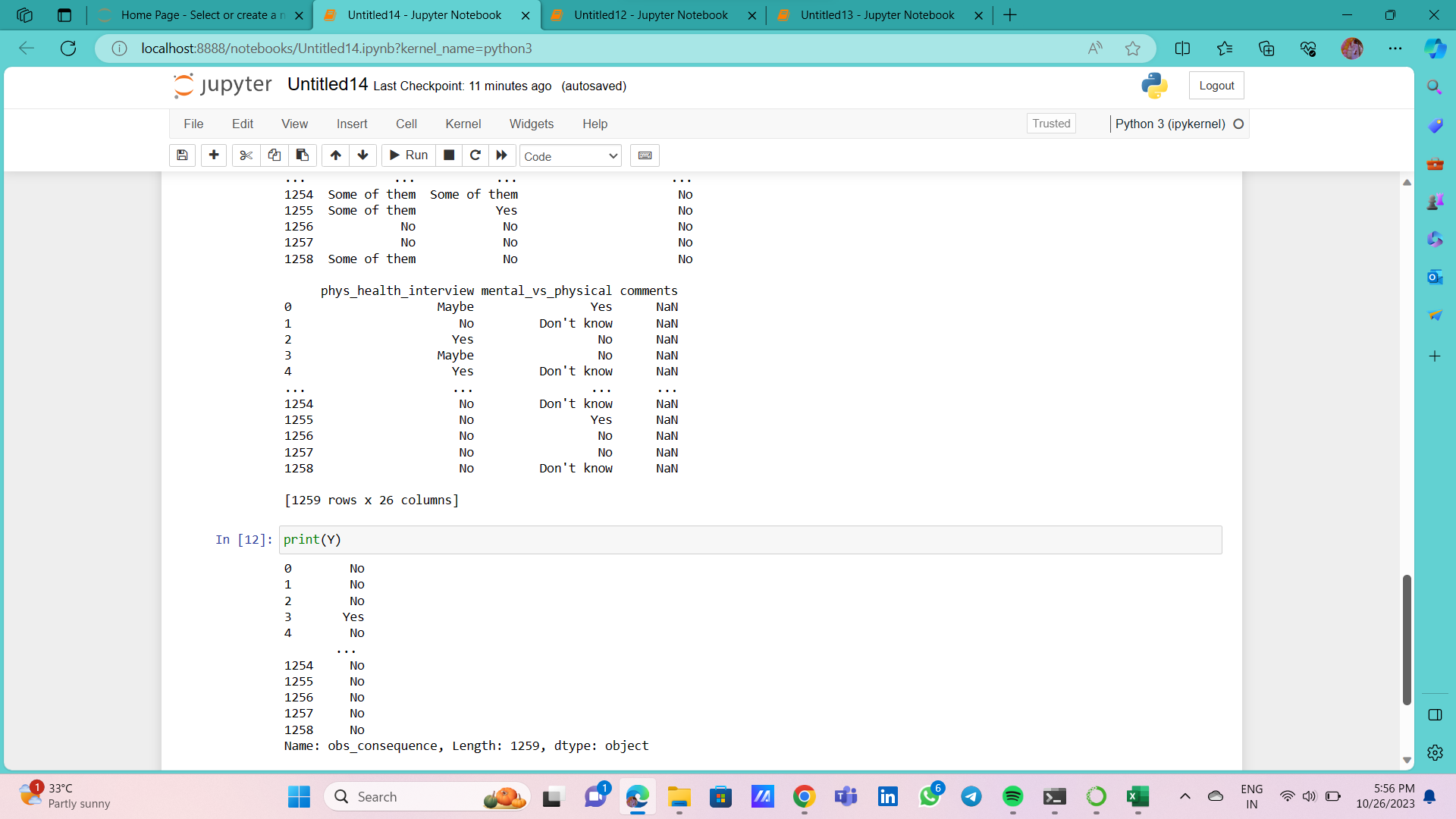


Fig 6.5

**CHAPTER 7**

**PROPOSED INNOVATION TECHNIQUE**

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it’s best suited for classification. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three. Let’s consider two independent variables x1, x2, and one dependent variable which is either a blue circle or a red circle.



Fig 7.1

### How does SVM work?

One reasonable choice as the best hyperplane is the one that represents the largest separation or margin between the two classes.



Fig 7.2

we choose the hyperplane whose distance from it to the nearest data point on each side is maximized. If such a hyperplane exists it is known as the **maximum-margin hyperplane/hard margin**. we choose L2. Let’s consider a scenario like shown below



Fig 7.3

we have one blue ball in the boundary of the red ball. The blue ball in the boundary of red ones is an outlier of blue balls. The SVM algorithm has the characteristics to ignore the outlier and finds the best hyperplane that maximizes the margin. SVM is robust to outliers.



Fig 7.4

So in this type of data point what SVM does is, finds the maximum margin as done with previous data sets along with that it adds a penalty each time a point crosses the margin. So the margins in these types of cases are called **soft margins**. When there is a soft margin to the data set, the SVM tries to minimize *(1/margin+∧(∑penalty))*. Hinge loss is a commonly used penalty. If no violations no hinge loss.If violations hinge loss proportional to the distance of violation.

Till now, we were talking about linearly separable data(the group of blue balls and red balls are separable by a straight line/linear line). What to do if data are not linearly separable?



Fig 7.5

Say, our data is shown in the figure above. SVM solves this by creating a new variable using a **kernel**. We call a point xion the line and we create a new variable yi as a function of distance from origin o.so if we plot this we get something like as shown below



Fig 7.6

In this case, the new variable y is created as a function of distance from the origin. A non-linear function that creates a new variable is referred to as a kernel.

**CHAPTER 8**

**CONCLUSION AND FUTURE SCOPE**

**CONCLUSION**

Concluding a public health campaign awareness effort is an important step in assessing the impact and effectiveness of the campaign. The conclusion typically involves summarizing key findings, outcomes, and lessons learned. Here's how you might conclude a public health campaign awareness initiative:**Objectives:Data and Metrics:**Share data and metrics that demonstrate the campaign's impact. This can include statistics on the reach of the campaign, audience engagement, and changes in behavior or awareness among the target population.**Key Findings:**Highlight the most important findings from the campaign. Discuss any trends, patterns, or insights that emerged from the data.**Success Stories and Testimonials:**Share success stories or testimonials from individuals or communities that were positively affected by the campaign. These personal accounts can provide a powerful testament to the campaign's impact.**Challenges and Lessons Learned:**Acknowledge any challenges or obstacles encountered during the campaign. Discuss the lessons learned and how they can inform future campaigns.**Community Involvement:**Reflect on the level of community involvement and engagement in the campaign. Discuss whether the campaign succeeded in fostering a sense of ownership and empowerment among the target population.**Recommendations for Future Campaigns**:Provide recommendations for future public health awareness campaigns. These can include suggestions for improving outreach, communication, or data collection methods.**Call to Action:**End with a call to action. Encourage individuals and organizations to continue supporting public health initiatives and to stay engaged in health-promoting behaviors.**Next Steps:**Provide information on any follow-up actions, events, or campaigns that will build on the awareness and momentum generated by the concluded campaign.A well-structured conclusion to a public health campaign awareness effort not only summarizes the campaign's impact but also sets the stage for ongoing efforts to improve public health and well-being. It's an opportunity to celebrate achievements and inspire further action in the interest of the community's health.

**FUTURE SCOPE**

**Emerging Health Threats**: Public health campaigns will continue to address emerging health threats, such as new infectious diseases, antimicrobial resistance, and the effects of climate change on health. Campaigns will need to educate the public on preventive measures and behavioral changes to mitigate these threats.**Mental Health Awareness:** There is a growing recognition of the importance of mental health. Future campaigns will aim to reduce stigma, increase awareness, and promote mental well-being. This includes campaigns related to stress management, anxiety, depression, and other mental health issues.**Vaccine Promotion:** The COVID-19 pandemic has highlighted the importance of vaccination. Future campaigns will focus on increasing vaccine acceptance, not only for COVID-19 but also for other preventable diseases. These campaigns will target vaccine hesitancy and provide accurate information.**Health Equity**: Addressing health disparities and promoting health equity will be a central theme in public health campaigns. These campaigns will aim to reduce disparities in access to healthcare, health outcomes, and social determinants of health.**Chronic Disease Prevention:** Non-communicable diseases (NCDs) like diabetes, heart disease, and obesity are a major public health concern. Future campaigns will emphasize lifestyle changes, healthy eating, physical activity, and early detection to prevent and manage these conditions.**Environmental Health:** With increasing awareness of the impact of environmental factors on health, campaigns will focus on issues like air quality, water safety, and climate change. They will encourage individuals and communities to take actions that reduce their environmental health risks.**Health Technology:** The integration of technology, such as telemedicine and health apps, into healthcare systems will necessitate campaigns that educate the public on how to use these tools effectively and securely.**Global Health Issues:** Public health campaigns will continue to address global health challenges, including infectious diseases, access to clean water and sanitation, and maternal and child health. They will work to build awareness and drive action on these international issues.**Health Literacy**: Improving health literacy will be an ongoing priority. Campaigns will aim to make health information more accessible and understandable for the general public.**Research and Data-Driven Campaigns:** Public health campaigns will increasingly rely on data analytics and research to tailor messages and interventions to specific populations and needs.