TITLE: Cyber Security Awareness Survey

Abstract:   
In the current digital era, where people and organizations encounter increasingly complex cyberthreats, cybersecurity awareness is essential. The purpose of this report is to assess public awareness of cyber security as it stands . It looks at a number of variables that affect the effectiveness and drawbacks of campaigns to raise awareness of cyber security. It also offers suggestions for raising awareness in order to successfully reduce cyber risks.

Introduction:

The speed at which digital technologies have developed in recent years has drastically changed our everyday lives, workplaces, and communication styles. The transition to digital has greatly improved connectivity and streamlined procedures. It has, however, also brought about a variety of cyberthreats that can impact both people and businesses. The increasing prevalence of cybersecurity risks, including malware, phishing attacks, data breaches, and identity theft, highlights the necessity of heightened awareness. Because digital data is so widely used, cybercriminals find it to be an attractive target. They take advantage of security system flaws and human error. As such, it is imperative to cultivate a strong culture of cybersecurity awareness. It is critical that people and organizations recognize potential cyberthreats and put strategic countermeasures in place. Reduced risk exposures can be achieved by educating people about safe online practices and the value of regular software updates.

Cyber threats are becoming more and more common as digital technology becomes more and more integrated into every aspect of our lives, from social media and personal banking to enterprise management and government infrastructure. While increasing efficiency and connectivity is one benefit of this technological integration, data security is a major vulnerability that must be addressed. Since so much private data is being digitalized, hackers are always coming up with inventive and complex ways to compromise systems. Typical cyberthreats include identity theft, which can have disastrous financial and personal consequences, malware, which can interfere with or damage systems, phishing attempts, which deceive people into disclosing sensitive information, and data breaches, which expose private and corporate information.

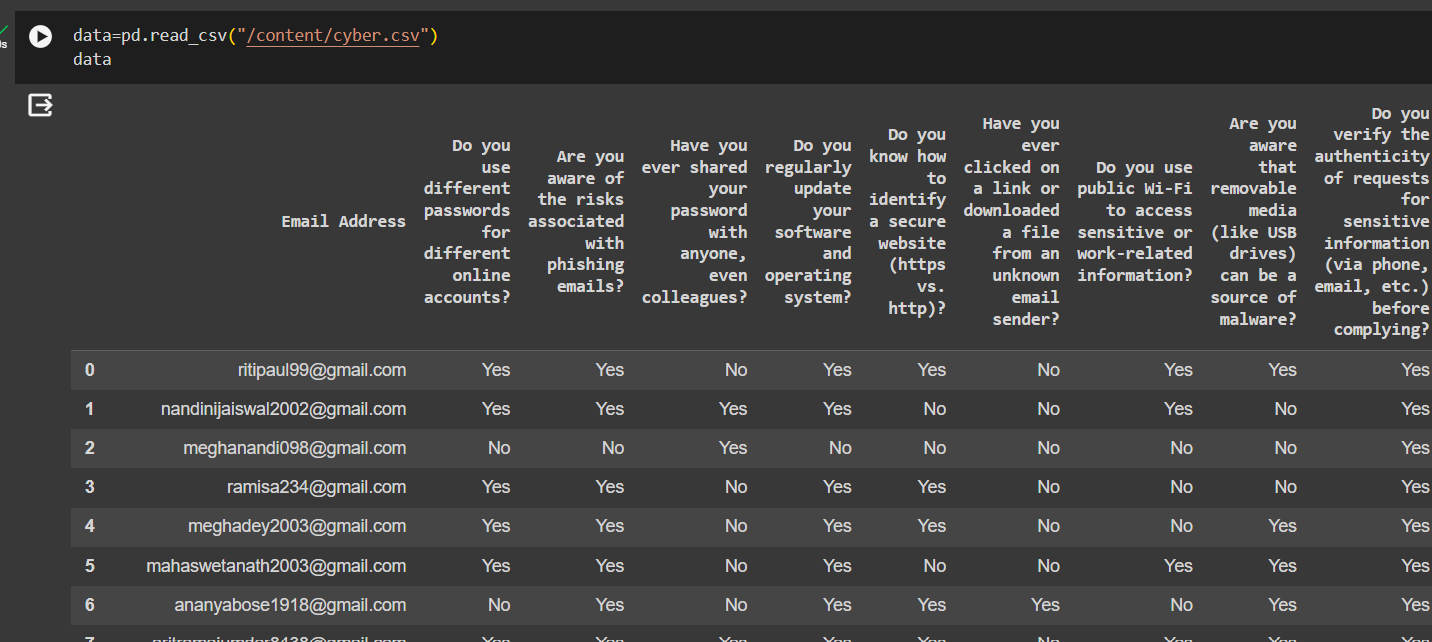
It is impossible to overestimate the significance of fostering a culture of cybersecurity awareness as we continue to navigate this digital era—it is a vital part of defending our increasingly digital way of life against the constantly changing landscape of cyber threats.

Methods:

As a survey was conducted through gforms to actually understand the awareness among people, near by 50 records were collected.

To assess the data firstly exploratory data analysis is performed then to classify whether people are aware about cyber security or not , a logistic regression model is fitted.

Data overview:



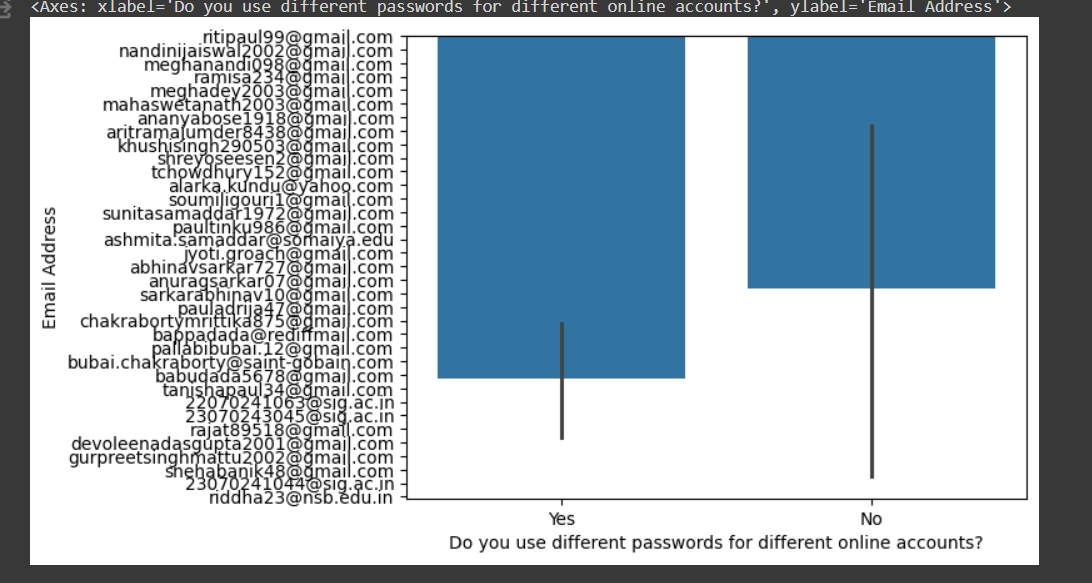
Exploratory Data Analysis:

The primary goal of exploratory data analysis (EDA), a crucial first step in data analysis, is to learn as much as possible about the dataset. In exploratory data analysis (EDA), key features of the data are summarized, patterns are found, anomalies are detected, and hypotheses are developed for more research. EDA frequently uses methods like data visualization, statistical modeling, and summary statistics to find patterns and relationships in the data, which helps with hypothesis generation and decision-making for deeper analysis. In general, EDA helps analysts and data scientists better understand their data before delving into more intricate analyses or modeling, acting as a vital foundation for ensuing data-driven tasks.

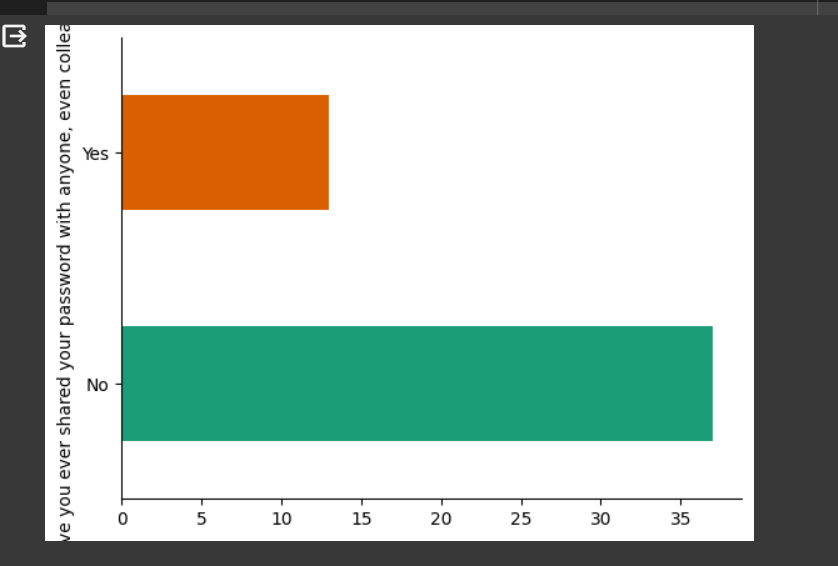
Logistic regression model:

When there are only two possible outcomes for a categorical outcome variable, such as in binary classification tasks, logistic regression is a statistical model that is employed. It uses one or more predictor variables to estimate the likelihood that a given input belongs to one of the two classes. In contrast to linear regression, logistic regression maps the input variables to the output using the logistic function, guaranteeing that the predicted probabilities lie within the range of 0 and 1. Maximum likelihood estimation is used to estimate the model parameters. The resulting model can offer insights into the relationships between the predictors and the probability of the outcome, making it a widely used tool for making predictions and comprehending the factors influencing a binary decision in a variety of fields, including marketing, finance, and medicine.

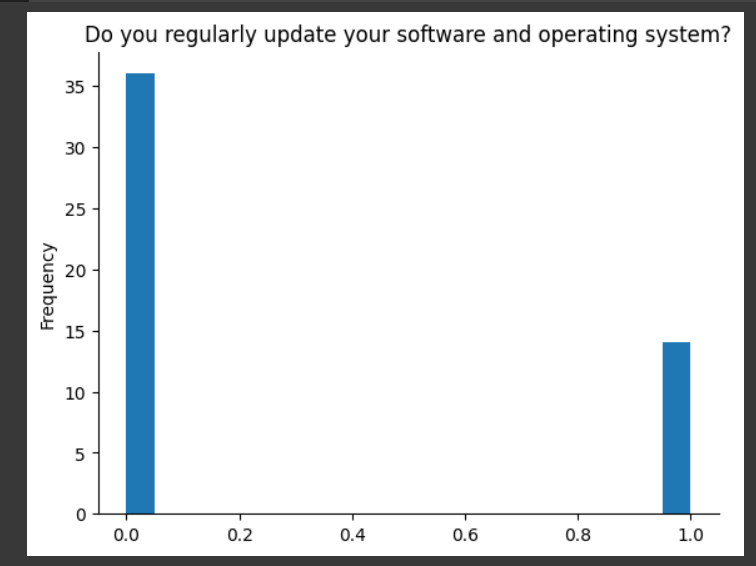
Results and Analysis:



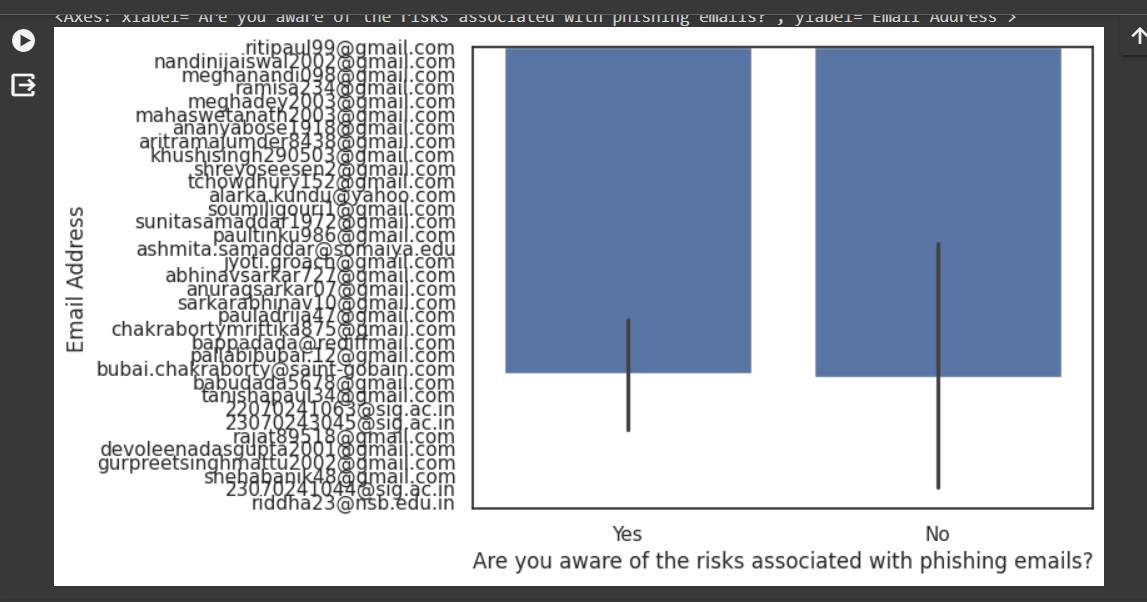
This bar chart clearly shows that people keep different passwords for online accounts , sufficing the needs of account security.



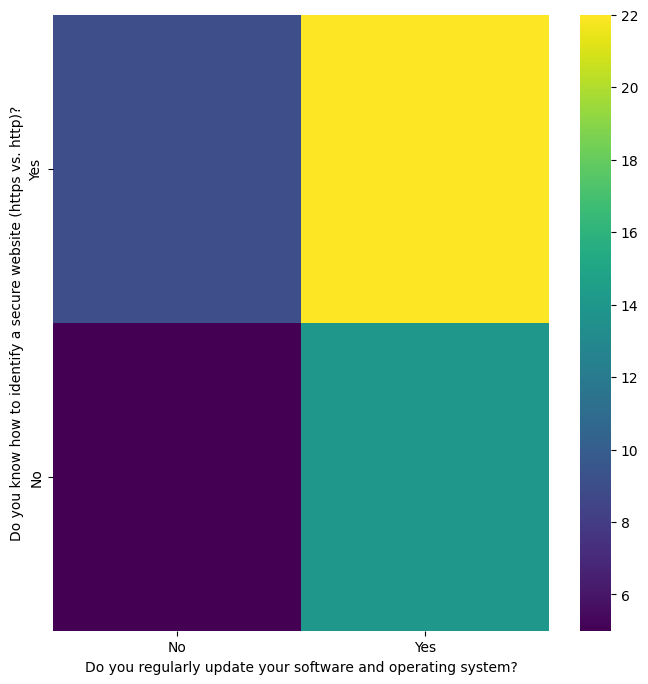
This bar chart shows that people are unlikely to share their passwords with anyone , thereby restricting their account information.



People do update their operating system regularly, and use antivirus thereby having a sense of protecting the system from malware.



The population is not much aware of the phishing emails , thereby clearly depicting that some concepts of cyber awareness are not much known.

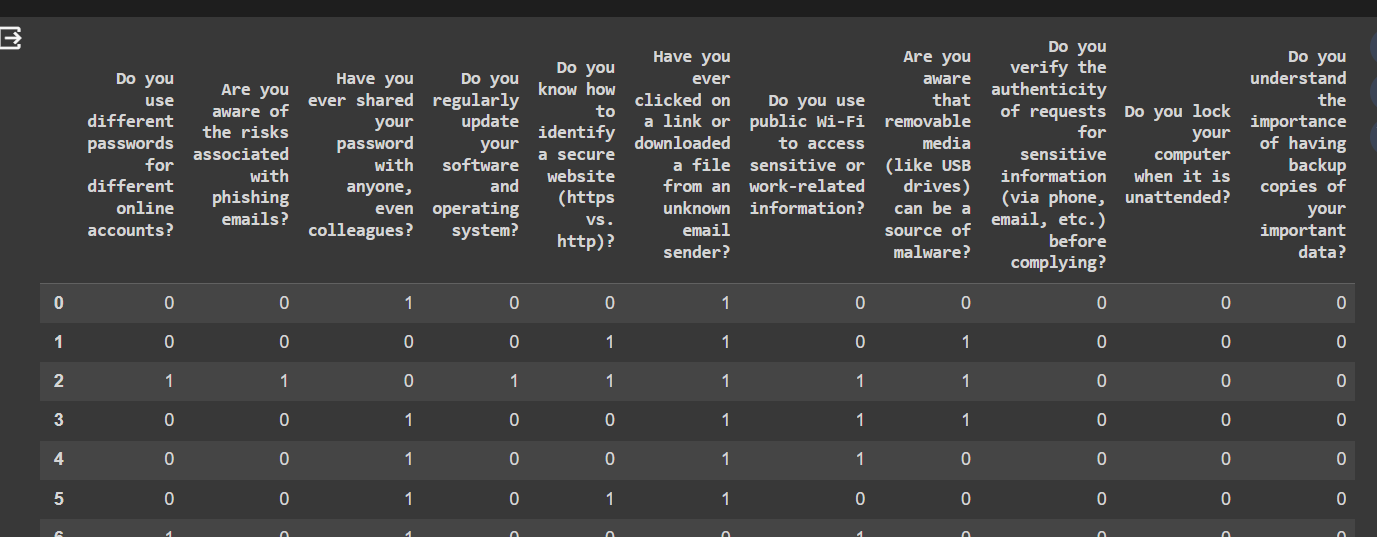


Through this grid chart we can see around 22 people among 50 people knows how to identify a secure website as well as regularly update the OS, whereas around 6 people don’t do any of them. Also around 10 people know how to identify a secure website but doesn’t update their OS.

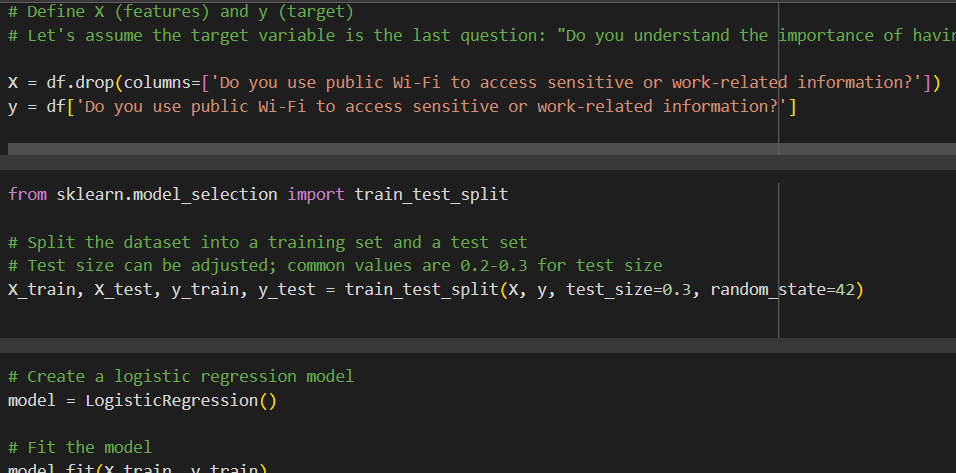


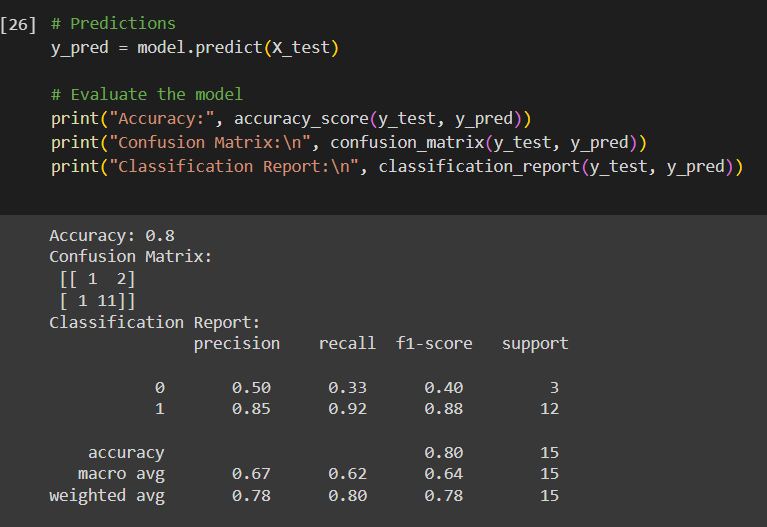
To understand about the cyber awareness , a logistic regression model is fit on where people are aware or not .

The responses are replaced by 0 or 1, to fit the training and testing data.



Now, model is fitted in the training data which is 80% of the data and rest 20% is used for testing data.





With high precision, recall, and F1-score for class 1 (Yes), the model performs reasonably well overall, suggesting that it can successfully identify positive cases. The reduced recall and precision for class 0 (No), however, point to potential for improvement in the detection of negative cases. Furthermore, the evaluation metrics could have been affected by the imbalance in the class distribution, underscoring the significance of taking class distribution into account when evaluating models.   
This suggests that Based on people's answers to various cybersecurity awareness questions, the logistic regression model performs reasonably well overall in predicting people's understanding of the significance having knowledge of cyber security and backup of important data. The evaluation metrics are probably impacted by the class distribution discrepancy, where fewer people (Class 0) may not understand the importance of cyber awareness and threat , than do (Class 1) people. The model's performance metrics may be distorted by this imbalance, especially for the minority class.

Conclusion:

Thereby , we can say that approximately a large population of people are aware of the importance of cyber security . They know about having backup of important data, using passwords , not using public wifi to open important data , whereas there is lack of knowledge about how to secure their operating systems, not click on unknown links , about phishing emails or how to detect a possible corrupted site.

Hence cyber security awareness among people has reached half way among people and a lot more study needs to be done