Understanding emotional driving factors of vaccination hesitancy and refusal in Bangladesh using Machine Learning approaches

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Bachelor in Computer Science & Engineering

by

Abdullah Al Mamun
CSE07008109
&
Mahfuzur Rahman
CSE07008111

Supervised by: **Adnan Ferdous Ashrafi**
Senior Lecturer

Department of Computer Science and Engineering
STAMFORD UNIVERSITY BANGLADESH

March 2024

Abstract

This thesis investigates the crucial aspects of vaccine decision-making during major health crises, focusing on the lessons learned from the COVID-19 pandemic to anticipate future scenarios. Utilizing machine learning methods, algorithms, and techniques, we delve into the multifaceted factors influencing vaccine decisions, considering both behavioral and emotional dimensions. Key determinants include age, financial situation, and personality traits, which interact to shape diverse perspectives on vaccination. The study particularly emphasizes the role of helping behavior, understanding the altruistic inclination that contributes to collective health and influences vaccine acceptance during crises. Emotions, such as fear, uncertainty, and mistrust, play a pivotal role in decision-making and are addressed through an analysis categorizing individuals into emotional response groups. By incorporating these insights, the research aims to inform proactive strategies that address vaccine hesitancy, enhance community resilience, and safeguard public health in future health emergencies.

Keywords: Vaccination emotional drivers, vaccine hesitancy, vaccine refusal, machine learning, SVM, Random Forest,Linear SVM, Radial SVM, Logistic Regression, K-Neighbour Classifier, Decision tree classifier, Gradient Boosting Classifier, Accuracy, Precision, Recall.

Approval

The Thesis Report Name 'Understanding emotional driving factors of vaccination hesitancy and refusal in Bangladesh using Machine Learning approaches' submitted by Abdullah Al Mamun ID: CSE07008109, Mahfuzur Rahman ID: CSE07008111, to the Department of Computer Science & Engineering, Stamford University Bangladesh, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science (Hons) in Computer Science & Engineering and approved as to its style and contents.

Board of Examiner's Name, Signature, and Date:

....	....	....
(Board Member 1)	**(Board Member 2)**	**(Board Member 3)**
Date:	Date:	Date:

Supervisor's Signature and Date:

....

Adnan Ferdous Ashrafi

Date:

Declaration

We, hereby, declare that the work presented in this Thesis is the outcome of the investigation performed by us under the supervision of Adnan Ferdous Ashrafi, Senior Lecturer, Department of Computer Science & Engineering, Stamford University Bangladesh. We also declare that no part of this Project and thereof has been or is being submitted elsewhere for the award of any degree or Diploma.

Signature and Date:

....
Abdullah Al Mamun
Date:

....
Mahfuzur Rahman
Date:

Dedicated to...
Our parents and our honorable supervisor Adnan Ferdous Ashrafi

Acknowledgments

First, we would like to thank **Almighty Allah** for allowing us to complete this report on *Understanding emotional driving factors of vaccination hesitancy and refusal in Bangladesh using Machine Learning approaches*. We want to thank the **Department of Computer Science and Engineering** for keeping thesis credit in the graduation program's curriculum and giving us a scope to taste the flavour of the research field according to our interests. Now, there is a little effort to show our deep gratitude to our supervisor. We convey our sincere gratitude to our supervisor **Adnan Ferdous Ashrafi**(Senior Lecturer, Department of Computer Science and Engineering), who allowed us to research this project and gain much knowledge in this field. His valuable advice and guidance have always been very helpful. In every phase of this research, his supervision and guidance made this report complete perfectly

Table of Contents

List of Figures

List of Tables

1 Introduction

Vaccines are like shields for our bodies, protecting us from harmful germs that can make us sick. They're tiny doses of weakened or inactive viruses or bacteria that teach our immune systems how to recognize and fight off these invaders. When we get vaccinated, our bodies learn to defend themselves, so if we ever encounter the real germ, we're ready to fend it off without getting sick. It's like giving our immune system a cheat sheet on how to beat the bad guys! Vaccines have been saving lives for decades, preventing diseases like measles, polio, and the flu from spreading and causing harm. They're an essential tool in keeping us healthy and strong. Deciding whether or not to get vaccinated can feel like navigating a maze of emotions, doubts, and external influences that shape our choices. Emotions like fear, uncertainty, or even hope can color our perceptions of vaccines, influenced by past experiences or what we hear from others. Some individuals may feel hesitant due to concerns about side effects or the speed of vaccine development, while others may feel more confident based on their trust in science or healthcare providers. Personality traits like cautiousness or skepticism can also play a role, as can environmental factors such as access to healthcare or cultural beliefs about vaccines. Additionally, demographic factors like age can impact vaccine decisions, with older adults often more eager to get vaccinated compared to younger populations. Moreover, looming threats like climate change can indirectly affect vaccination efforts by altering disease patterns or exacerbating health disparities. Trust issues with local government can further complicate matters, influencing individuals' perceptions of vaccine safety and efficacy. Thus, the challenge of vaccination extends beyond the scientific efficacy of the vaccines themselves; it involves understanding and addressing the complex interplay of emotions, beliefs, and contextual factors—including demographic, environmental, and trust-related influences—that shape individual and collective decisions about immunization. Our study is aimed at classify the individuals emotional situation and delving into the complex landscape of vaccination hesitancy to better understand its underlying emotional,psychological,demographic and behavioural drivers and implications for public health.

1.1 Motivation

Understanding emotional driving factors of vaccination hesitancy and refusal in Bangladesh is motivated after realizing the messy situation in recently passed COVID-19 pandemic. During the COVID-19 pandemic when vaccine reveled and spread from the modern county - they take this advantage in well form. Most of the country got the vaccine as fast as possible. But our country showed a different context and they were not serious about the vaccine. Also when the vaccine came, a certain group of people took the vaccine according to the decision of local government without any planing or research. Early stage of the vaccination, most if the people took the vaccine of first dose. But day by day willingness to take the vaccine was decreased, though the vaccine was available and the corona virus was deadly. Then day by day people were absent-minded about the vaccine. Because they were not serious about vaccine. For all of this we loosed may lives for deadly corona virus. On the other hand, the damage for COVID-19 is less reported where people got vaccinated. After realizing this, this study motivate us to analyze the individual behavioral pattern and other drivers to understand the overall behavioral pattern on that area and then take necessary steps to aware about any vaccination activities that can be any future pandemic situation not only in Bangladesh but also in our subcontinent and worldwide.

1.2 Research Outcome

Analyzing the emotional drivers behind vaccination can provide valuable insights into people's perceptions of vaccines, particularly during critical periods such as the COVID-19 pandemic. By examining these emotional factors, we can better understand individuals' attitudes towards vaccination and their decision-making processes. Utilizing high-quality and comprehensive datasets in our analytical models enables us to accurately gauge the emotional patterns, hesitancy, and refusal rates within a given area. This analysis yields actionable insights that inform targeted interventions to address vaccine hesitancy and refusal effectively. Moreover, this study serves as a valuable resource for researchers interested in delving deeper into the psychology of vaccine acceptance and refusal, offering rich insights into the underlying causes driving these behaviors.

1.3 Types of Tools & methods

In our study, we employed various machine learning libraries, frameworks,classifer and methods like Pandas, NumPy, Seaborn, Matplotlib, DecisionTreeClassifier, accuracy metrics and many other libraries. All our analyses were conducted using Google Colab notebooks.

1.3.1 Google Colab

Google Colab notebook is a cloud-based platform provided by Google for writing and executing Python code in a collaborative environment. Users can create and share notebooks containing code, text, equations, and visualizations, making it a versatile tool for various tasks, including data analysis, machine learning, and education. One of the key benefits of Google Colab notebook is its integration with Google Drive, allowing users to store and access notebooks seamlessly across devices. Additionally, Google Colab provides free access to GPU and TPU resources, enabling faster computation for tasks such as training deep learning models. Moreover, Google Colab offers pre-installed libraries and packages, eliminating the need for manual setup, and supports real-time collaboration, enabling multiple users to work on the same notebook simultaneously. Overall, Google Colab notebook provides a convenient and efficient platform for writing, sharing, and executing Python code, making it an invaluable tool for individuals and teams alike.

1.3.2 Pandas

Pandas is a widely-used Python library for data manipulation and analysis. It provides intuitive data structures and functions designed to facilitate efficient handling of structured data.

- **DataFrame**: A two-dimensional labeled data structure resembling a spreadsheet, enabling easy manipulation and analysis of tabular data.

- **Series**: A one-dimensional labeled array capable of holding any data type, similar to a column in a DataFrame.

- **Data Input/Output**: Supports reading and writing data from various file formats, including CSV, Excel, SQL databases, and JSON.

- **Data Cleaning and Preparation**: Offers functions for handling missing data, removing duplicates, reshaping data, and more.

- **Data Selection and Indexing**: Provides intuitive methods for selecting, slicing, and indexing data based on labels or positions.

- **Aggregation and Grouping**: Allows for efficient data aggregation and grouping operations essential for summarizing and analyzing large datasets.

- **Integration with Other Libraries**: Seamlessly integrates with popular Python libraries like NumPy, Matplotlib, and scikit-learn, enhancing its capabilities for data analysis and visualization.

Overall, Pandas simplifies the process of working with structured data in Python, making it a valuable tool for data scientists, analysts, and developers.

1.3.3 NumPy

Arrays and matrices are supported by the NumPy [1] package for the Python programming language. Many scientific and mathematical applications use basic data structures like arrays and matrices, and NumPy provides a quick and simple way to work with them in Python. NumPy offers a number of crucial features, including:

- The N-dimensional array, sometimes known as an array, is the main data structure in NumPy.

- Numerous mathematical procedures, including trigonometry, statistics, and linear algebra

- Matplotlib integration with additional plotting and visualization tools Support for array operations, such as element-wise and linear algebra-based operations

Using NumPy is significantly faster than working with lists in traditional Python since it is optimized for high performance and saves arrays.

1.3.4 Matplotlib

For the Python programming language, there is a plotting package called Matplotlib [2]. It offers an API for developing a variety of static, animated, and interactive visualizations in Python. Data scientists, engineers, and researchers use Matplotlib to make several types of data visualizations, including line plots, scatter plots, bar charts, histograms, and more. Matplotlib allows for a great deal of customization and can be used to produce intricate representations. It supports a variety of output formats, including PNG, PDF, SVG, and JPG, making it simple to store and share visualizations. To make working with data in Python simple, it can also be connected with other libraries like NumPy and Pandas. There are numerous features in Matplotlib, such as:

- A variety of graphing techniques, including line plots, scatter plots, bar charts, histograms, and more
- Various output formats supported, including PNG, PDF, SVG, and JPG
- Complex visualizations can be made using an object-oriented API.
- Annotation, color, and style customization options
- incorporating NumPy and Pandas as well as other libraries

In general, Matplotlib is a robust and adaptable charting package that is frequently used in the data science field to produce visual representations of data. It is a well-liked option for developing static, animated, and interactive visualizations because of its extensive feature set and interoperability with other libraries.

1.3.5 Seaborn

Seaborn [3] is a Python visualization library built on Matplotlib, known for creating visually appealing statistical graphics with minimal code. It offers a wide range of plot types like scatter plots, bar plots, and heatmaps, and provides default color palettes for aesthetically pleasing plots. With seamless integration with Pandas, it simplifies data manipulation and visualization. Overall, Seaborn streamlines the creation of sophisticated visualizations, aiding data analysts and scientists in communicating insights effectively.

1.4 Chapter Summary

This chapter will provide insights into the feasibility of our study and the technical challenges that must be overcome to make such a system effective and reliable. The study will also provide recommendations for what type of data that should be collected and the best methods for addressing the technical challenges associated with understanding the emotional driving factors of vaccination hesitancy and refusal.

2 Literature Review

The Literature Review presented in this section critically examines and synthesizes existing scholarly works that are central to the research focus of this thesis. This comprehensive review aims to provide a contextual background, offering insights into the current state of knowledge in the field. By delving into a range of scholarly articles, books, and relevant research, the review aims to identify gaps, contradictions, and trends in the existing literature. Moreover, it serves as a foundation for the development of the conceptual framework, guiding the subsequent empirical investigation. Through a meticulous examination of diverse perspectives and theoretical frameworks, this literature review contributes to the theoretical foundation of the study, fostering a deeper understanding of the research problem and setting the stage for the empirical analysis that follows.

2.1 Factors Influencing COVID-19 Vaccine Hesitancy and Acceptance

Walsh et. al.[4] proposed this paper to introduce the factors that influence vaccine hesitancy and acceptance, crucial for successful vaccination campaigns. By synthesizing existing research, it aims to inform targeted interventions to address hesitancy and promote uptake, ultimately aiding in ending the pandemic. Theoretical frameworks like the Health Belief Model (HBM), Social Cognitive Theory, and Theory of Planned Behavior help understand vaccination behavior. The HBM focuses on perceived disease risk and vaccine benefits and barriers. Social Cognitive Theory emphasizes social factors and self-efficacy, while the Theory of Planned Behavior looks at attitudes and perceived control over behavior. Predictors of COVID-19 vaccine hesitancy and acceptance include demographic factors like age and gender, past vaccination behavior, risk perceptions, trust in authorities, social influences, and civic responsibility. Younger individuals and women are more hesitant, while trust in government and healthcare providers increases acceptance. Positive peer pressure and a sense of civic duty also promote vaccine acceptance. Research Gaps and Future Directions: While existing research has identified several predictors of COVID-19 vaccine hesitancy and acceptance, some gaps remain. Future studies should focus on specific attitudes and beliefs regarding COVID-19 vaccination, explore the role of civic responsibility as a predictor, and investigate the complex interplay between demographic factors, risk perceptions, and social influences. Longitudinal studies tracking changes in vaccine attitudes over time and interventions targeting vaccine hesitancy are needed to inform effective public health strategies. Understanding COVID-19 vaccine hesitancy and acceptance is crucial for effective interventions to boost vaccination rates and reach herd immunity. This review identifies key predictors of hesitancy and acceptance, guiding future research and intervention efforts. Addressing vaccine hesitancy is critical for ending the COVID-19 pandemic and safeguarding public health.

2.2 Understanding COVID-19 Vaccine Hesitancy in the United States

According to Rancher et. al.[5] (2023) introduced that the COVID-19 pandemic has prompted urgent efforts to develop and distribute vaccines worldwide. However, vaccine hesitancy remains a significant challenge, especially in the United States. This review examines the factors contributing to COVID-19 vaccine hesitancy in the US, including demographic disparities and regional variations. Understanding these drivers is crucial for designing effective interventions to enhance vaccine acceptance and uptake. The determinants of COVID-19 vaccine hesitancy in the United States encompass a range of individual-level and systemic factors. Demographic characteristics like race, ethnicity, age, gender, and socioeconomic status have consistently influenced vaccine acceptance, with marginalized communities and Black individuals exhibiting higher hesitancy due to historical distrust and inequities. Additionally, psychological factors such as confidence in vaccine safety, complacency about COVID-19 severity, constraints in access, risk calculation, and collective responsibility play significant roles, as indicated by the 5C model. Understanding these determinants is crucial for tailoring interventions to address vaccine hesitancy effectively. Trends in COVID-19 vaccine hesitancy have evolved dynamically, influenced by factors such as vaccine availability, public health messaging, and changing perceptions of the pandemic. Early surveys revealed varying levels of vaccine acceptance, with demographic and geographic disparities. Subsequent research identified the impact of misinformation and political polarization on hesitancy, highlighting the need for targeted communication strategies. Longitudinal studies emphasized changes in vaccine intentions, emphasizing ongoing monitoring and adaptation of public health campaigns to address evolving attitudes towards vaccination. The findings from research on COVID-19 vaccine hesitancy in the United States have significant implications for public health interventions aimed at promoting vaccine acceptance and uptake. Tailored strategies that address the specific concerns and barriers identified within different demographic groups and geographic regions are essential for maximizing the effectiveness of vaccination campaigns. Building trust in the healthcare system, enhancing vaccine literacy, addressing systemic inequities, and leveraging community partnerships are key components of comprehensive interventions to combat vaccine hesitancy. Furthermore, fostering open communication, providing transparent information, and engaging with diverse stakeholders are critical for overcoming resistance to vaccination and fostering a culture of vaccine confidence. COVID-19 vaccine hesitancy in the United States poses significant challenges for public health, equity, and social cohesion. To address this issue effectively, ongoing research is crucial for understanding its underlying determinants and impacts. By implementing evidence-based interventions focused on targeted communication, community engagement, and equitable access to vaccination, we can work towards overcoming vaccine hesitancy and controlling the spread of COVID-19.

2.3 Understanding COVID-19 Vaccine Acceptance and Hesitancy

This paper by Raut et. al [6] aims to comprehensively explore the factors influencing COVID-19 vaccine acceptance and hesitancy, addressing key determinants, trends, and implications for public health interventions. The COVID-19 pandemic has underscored the necessity of vaccination in curbing virus transmission, yet vaccine hesitancy remains a significant obstacle to achieving widespread immunization. Understanding these factors is essential for devising targeted strategies to enhance vaccine uptake and combat the pandemic effectively.Research on COVID-19 vaccine hesitancy highlights various factors influencing acceptance, including demographics (age, gender, race, ethnicity, socioeconomic status), psychological factors (confidence, complacency, constraints, risk perception, collective responsibility), and systemic inequities. Understanding these determinants is essential for developing targeted interventions to address vaccine hesitancy effectively. Studies on COVID-19 vaccine hesitancy reveal changing trends influenced by factors like vaccine availability, public health messaging, and evolving perceptions of the pandemic. Early surveys showed varying acceptance levels, with demographic and regional disparities. Misinformation, conspiracy theories, and political polarization have impacted hesitancy, emphasizing the need for targeted communication strategies. Longitudinal studies highlight shifting vaccine intentions, stressing the importance of ongoing monitoring and adaptation of public health campaigns to address evolving concerns. Addressing COVID-19 vaccine hesitancy requires multifaceted approaches that target individual-level beliefs, systemic barriers, and socio-cultural factors influencing vaccine acceptance. Public health interventions should focus on improving vaccine literacy, building trust in vaccines and healthcare systems, addressing misinformation, and addressing disparities in vaccine access and distribution. Healthcare workers, including clinical pharmacists, play a crucial role in providing accurate information, addressing concerns, and promoting vaccine acceptance within their communities. By implementing evidence-based strategies and tailored interventions, public health authorities can effectively combat vaccine hesitancy and achieve widespread vaccine coverage, contributing to the control of the COVID-19 pandemic.

2.4 A Comprehensive Analysis of COVID-19 Dynamics: Insights from Statistical and Machine Learning Approaches

The COVID-19 pandemic has prompted urgent efforts to develop vaccines and address its impact on global health systems. The paper by Gupta et. al [7] explores research endeavors that integrate statistical analysis and machine learning techniques to comprehensively analyze COVID-19 dynamics. Understanding factors influencing the severity of the pandemic, including vaccine acceptance and safety, is essential for effective public health responses. The COVID-19 pandemic has spurred global vaccination efforts, yet challenges like vaccine hesitancy persist. Researchers are employing varied analytical methods to assess vaccination's impact and identify at-risk individuals. This integrated approach aims to inform effective strategies for combating the pandemic. Researchers utilize diverse datasets, like those from VAERS, to analyze adverse reactions to COVID-19 vaccines, incorporating demographic, medical, and vaccination data. Preprocessing techniques ensure data quality by addressing outliers and missing values. Exploratory data analysis provides insights into variable relationships, guiding subsequent analyses. Statistical tests, including the chi-square test, identify significant factors influencing COVID-19 outcomes. Machine learning models like Logistic Regression and LGBM predict outcomes based on demographic and clinical factors, aiding in risk assessment. Performance evaluation metrics gauge model accuracy, helping identify individuals at risk of vaccination complications. Key findings emphasize early hospitalization, vaccine effectiveness, and debunking vaccine safety myths, guiding targeted vaccination efforts and promoting broader vaccine participation.

Despite significant insights gained from these analyses, limitations exist, particularly regarding the generalized of findings to diverse populations. Future research should focus on expanding datasets and employing advanced analytical techniques, such as recurrent neural networks, to uncover hidden patterns in COVID-19 dynamics and enhance vaccine acceptance and safety. Integrated statistical analysis and machine learning techniques offer a powerful toolset for understanding and addressing the challenges of the COVID-19 pandemic. This approach enables researchers to extract meaningful insights from complex datasets, guiding evidence-based interventions to improve vaccine acceptance and safety globally.

2.5 Navigating the Impact of COVID-19: A Multifaceted Literature Review

The COVID-19 pandemic has brought unprecedented challenges to societies worldwide, affecting various aspects of daily life and posing significant threats to public health, economies, and social structures. A case study of a Bangladeshi researcher Saifuzzaman et. al. [8] synthesizes existing research on the impact of COVID-19, focusing on its implications for education, economics, politics, mental health, and societal well-being. The pandemic has significantly disrupted education globally, with studies revealing challenges in transitioning to remote learning, particularly for marginalized communities due to the digital divide. Efforts are needed to address these disparities and ensure equitable access to education and technology. COVID-19 has triggered significant economic challenges, including supply chain disruptions, reduced consumer spending, and widespread job losses (McKibbin & Fernando, 2020). Alon et al. (2020) emphasize the unequal impact on vulnerable populations like low-income workers and small businesses, accentuating pre-existing inequalities.. The pandemic has led to varied political responses, including lockdowns and debates on healthcare and welfare, with scholars assessing government strategies' effectiveness in curbing transmission and protecting public health. It also impact on global mental health challenges, with increased rates of anxiety, depression, and suicidal ideation, particularly among vulnerable groups, due to lockdowns and social isolation measures. COVID-19 has reshaped societal norms and behaviors, impacting social interaction, community cohesion, and trust in institutions, highlighting the need to understand cultural and social factors in pandemic response for collective well-being. The COVID-19 crisis has diverse impacts across education, economics, politics, mental health, and society.

3 Methodology

The dataset is from Slovakia and the title of dataset is Emotional drivers of the vaccination hesitancy and refusal: A dataset from Slovakia [9]. All materials were distributed in the Slovak language. For better understanding we converted it into English using a online tool. The dataset describe the emotional driver data of random people. It has 195 column with indexing. When we started training, we removed some unnecessary stuff from the dataset and prepared the data for analysis.Then we applied the data on different machine learning models that are well-known. By doing this, we figured out which models work the best for our dataset. To evaluate this we used out self collected dataset.It helps us to understand how to use these models effectively for our research.

3.1 Dataset description

3.1.1 Actual dataset

The dataset titled ”Emotional Drivers of Vaccination Hesitancy and Refusal: A Dataset from Slovakia” [9] consists of emotional responses gathered from a sample of 500 individuals, comprising 250 women and 250 men. It encompasses a total of 195 columns, including an index column. The dataset of Slovakian version (actualDataset.sav file) and csv (actualDataset.csv) format is like this:

Raw Data without labels

	id	StartDate	EndDate	Duration__in_seconds_	infoconsent	sex	age	district	education	education_other	conlib	religion	marital_status
0	00445320-bde3-11eb-ba16-1b6b7e4cc699	2021-05-26 10:06:24	2021-05-26 10:38:23	1,918	1	2	34	65	3		4	4	1
1	005cdfe0-bde3-11eb-b59d-35bad9e128c1	2021-05-26 07:40:25	2021-05-26 08:00:20	1,194	1	1	64	62	5		3	6	1
2	00ab97b0-bde3-11eb-b8ab-51c90806ceff	2021-05-26 08:28:51	2021-05-26 09:05:17	2,185	1	2	38	24	5		3	4	1
3	00b72e40-bde3-11eb-b24c-c1bf9a458e97	2021-05-27 07:35:42	2021-05-27 07:53:56	1,093	1	2	28	13	5		4	3	4
4	013b6a00-bde3-11eb-8a64-01932a52efbb	2021-05-26 11:00:54	2021-05-26 11:14:49	835	1	2	29	59	5		5	2	4
5	014b6cb0-bde3-11eb-8fb4-59b64b5c3c27	2021-05-26 07:54:20	2021-05-26 08:23:43	1,762	1	2	28	35	2		4	5	4
6	01c10760-bde3-11eb-998c-05e40abda4f1	2021-05-28 10:42:57	2021-05-28 11:05:34	1,356	1	2	34	17	5		5	1	4
7	01fae890-bde3-11eb-9798-0b5bc1bf55b2	2021-05-28 23:02:26	2021-05-29 00:03:56	3,689	1	1	40	24	3		4	5	1
8	020e7130-bde3-11eb-a2d2-f1029af5022b	2021-05-26 15:16:00	2021-05-26 15:32:07	966	1	1	49	59	2		4	6	1
9	02d47240-bde3-11eb-837a-472332d34dcd	2021-05-26 19:44:19	2021-05-26 20:03:42	1,162	1	1	81	74	5		4	6	3

Figure 3.1: Dataset in sav format

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	id	StartDate	EndDate	Duration__in_secon	infoconsent	sex	age	district	education	education_other	conlib	religion	marital_status	household
0	00445320-bde3-11e	2021-05-26 10:06:24	2021-05-26 10:38:23	1918	I have read the abov	woman	34	Plešťany	high school		4	4	married	
1	005cdfe0-bde3-11eb	2021-05-26 7:40:25	2021-05-26 8:00:20	1194	I have read the abov	man	64	Dunajská Streda	university (Mgr., Ing.)		3	6	married	
2	00ab97b0-bde3-11e	2021-05-26 8:28:51	2021-05-26 9:05:17	2185	I have read the abov	woman	38	Košice II	university (Mgr., Ing.)		3	4	married	
3	00b72e40-bde3-11e	2021-05-27 7:35:42	2021-05-27 7:53:56	1093	I have read the abov	woman	28	Žiar nad Hronom	university (Mgr., Ing.)		4	3	single	
4	013b6a00-bde3-11e	2021-05-26 11:00:54	2021-05-26 11:14:49	835	I have read the abov	woman	29	Prievidza	university (Mgr., Ing.)		5	2	single	
5	014b6cb0-bde3-11e	2021-05-26 7:54:20	2021-05-26 8:23:43	1762	I have read the abov	woman	28	Nitra	high school without graduation		4	5	single	
6	01c10760-bde3-11e	2021-05-28 10:42:57	2021-05-28 11:05:34	1356	I have read the abov	woman	34	Bratislava IV	university (Mgr., Ing.)		5	1 - religion is not imp	single	
7	01fae890-bde3-11eb	2021-05-28 23:02:26	2021-05-29 0:03:56	3689	I have read the abov	man	40	Košice II	high school		4	5	married	
8	020e7130-bde3-11e	2021-05-26 15:16:00	2021-05-26 15:32:07	966	I have read the abov	man	49	Prievidza	high school without graduation		4	6	married	
9	02d47240-bde3-11e	2021-05-26 19:44:19	2021-05-26 20:03:42	1162	I have read the abov	man	81	Martin	university (Mgr., Ing.)		4	6	widow/widower	
10	02fc96f0-bde3-11eb-	2021-05-26 7:46:33	2021-05-26 8:09:40	1386	I have read the abov	man	72	Košice I	university (PhD.)		6	1 - religion is not imp	married	
11	031f2aa0-bde3-11eb	2021-05-26 13:23:32	2021-05-26 14:15:48	3135	I have read the abov	woman	32	Žilina	high school		4	7 - religion is very in	married	
12	038a7510-bde3-11e	2021-05-26 23:59:46	2021-05-27 0:22:30	1364	I have read the abov	woman	24	Kysucké Nové Mesto	high school		4	7 - religion is very in	single	
13	03db6ac0-bde3-11e	2021-05-26 8:33:24	2021-05-26 8:46:26	782	I have read the abov	man	49	Hlohovec	high school without graduation		5	4	married	
14	03e08650-bde3-11e	2021-05-26 16:54:44	2021-05-26 17:23:57	1753	I have read the abov	woman	29	Medzilaborce	university (Bc.)		5	2	single	
15	044a0c20-bde3-11e	2021-05-26 10:10:35	2021-05-26 10:27:04	989	I have read the abov	woman	31	Levoča	university (Mgr., Ing.)		4	7 - religion is very in	married	
16	04690600-bde3-11e	2021-05-27 18:47:12	2021-05-27 19:24:10	2217	I have read the abov	woman	45	Košice I	university (Mgr., Ing.)		3	4	divorced	
17	04955840-bde3-11e	2021-05-26 17:26:43	2021-05-26 17:48:37	1313	I have read the abov	woman	35	Levice	high school		4	4	single	
18	05188900-bde3-11e	2021-05-26 11:35:51	2021-05-26 11:54:37	1125	I have read the abov	woman	35	Trenčín	university (Mgr., Ing.)		4	3	married	
19	0519d950-bde3-11e	2021-05-26 20:32:51	2021-05-26 20:45:58	786	I have read the abov	woman	18	Poprad	primary school		6	4	single	
20	052184a0-bde3-11e	2021-05-26 8:20:02	2021-05-26 8:27:14	431	I have read the abov	woman	26	Košice III	university (Mgr., Ing.)		4	2	single	
21	05428110-bde3-11e	2021-05-26 11:07:25	2021-05-26 21:23:39	36973	I have read the abov	woman	42	Banská Bystrica	high school without graduation		5	1 - religion is not imp	married	
22	05babee0-bde3-11e	2021-05-26 8:51:57	2021-05-26 9:46:31	3274	I have read the abov	woman	29	Levice	university (Mgr., Ing.)		4	1 - religion is not imp	single	
23	05c8e6a0-bde3-11e	2021-05-26 8:03:46	2021-05-26 8:31:33	1667	I have read the abov	woman	33	Poprad	high school		4	3	married	
24	067777a0-bde3-11e	2021-05-26 9:24:52	2021-05-26 9:36:43	710	I have read the abov	woman	28	Košice-okolie	high school		4	3	single	

Figure 3.2: Dataset in csv format

The dataset columns are structured using various methods or sub-components to form the dataset.The sub-parts are - (1) Socio-demographic questions such as age, gender, education and questions about political orientation and the importance of religion.(2) Individual characteristics (big five personality factors [10], collectivism/individualism [11], consideration for future consequences [12], prosocial motivations [13]), (3) feelings of threat [14] and helplessness related to vaccination [15], COVID-19 pandemic [16], and climate change, and finally (4) prosocial behaviour [17] in three domains: vaccination, helping behaviour during the pandemic and pro-environmental behaviour. All materials were distributed in the Slovak language. The column description and sub-parts methods-

1. **Descriptive info:** Here consists some columns that describe the participants:(`Id`,

survey Start_date, End_date, Duration in second. Info_consent) indicates whether the participants gave their permission. Other columns include (Redirect, Finished, and Status_R). Lets take a look of snap of dataset:-

id	StartDate	EndDate	Duration__in_seconds_	infoconsent	redirect	finished	status_R
00445320-bde3-11eb-ba16-1b6b7e4cc699	5/26/2021 10:06	5/26/2021 10:38	1918	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	29	1
005cdfe0-bde3-11eb-b59d-35bad9e128c1	5/26/2021 7:40	5/26/2021 8:00	1194	I have read the above, I agree and I am interested in participating in the research.	No, I wish to end my participation.	0	1
00ab97b0-bde3-11eb-b8ab-51c90806ceff	5/26/2021 8:28	5/26/2021 9:05	2185	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	40	1
00b72e40-bde3-11eb-b24c-c1bf9a458e97	5/27/2021 7:35	5/27/2021 7:53	1093	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	39	2
013b6a00-bde3-11eb-8a64-01932a52efbb	5/26/2021 11:00	5/26/2021 11:14	835	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	2
014b6cb0-bde3-11eb-8fb4-59b64b5c3c27	5/26/2021 7:54	5/26/2021 8:23	1762	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	30	2
01c10760-bde3-11eb-998c-05e40abda4f1	5/28/2021 10:42	5/28/2021 11:05	1356	I have read the above, I agree and I am interested in participating in the research.	No, I wish to end my participation.	0	2
01fae890-bde3-11eb-9798-0b5bc1bf55b2	5/28/2021 23:02	5/29/2021 0:03	3689	I have read the above, I agree and I am interested in participating in the research.	No, I wish to end my participation.	0	1
020e7130-bde3-11eb-a2d2-f1029af5022b	5/26/2021 15:16	5/26/2021 15:32	966	I have read the above, I agree and I am interested in participating in the research.	No, I wish to end my participation.	0	1
02d47240-bde3-11eb-837a-472332d34dcd	5/26/2021 19:44	5/26/2021 20:03	1162	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	2
02fc96f0-bde3-11eb-9b31-a3a75e28ecb4	5/26/2021 7:46	5/26/2021 8:09	1386	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	33	1
031f2aa0-bde3-11eb-9a63-439e80ba296b	5/26/2021 13:23	5/26/2021 14:15	3135	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	30	1
038a7510-bde3-11eb-87f2-2fea3dff525c	5/26/2021 23:59	5/27/2021 0:22	1364	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	2
03db6ac0-bde3-11eb-b360-4f0234f744e0	5/26/2021 8:33	5/26/2021 8:46	782	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	0	1
03e08650-bde3-11eb-acd7-3386f7261880	5/26/2021 16:54	5/26/2021 17:23	1753	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	35	2
044a0c20-bde3-11eb-a69a-317e663d38c0	5/26/2021 10:10	5/26/2021 10:27	989	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	47	1
04690600-bde3-11eb-93e7-5788a78ca8ea	5/27/2021 18:47	5/27/2021 19:24	2217	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	2
04955840-bde3-11eb-a9c3-8f8bd519ba51	5/26/2021 17:26	5/26/2021 17:48	1313	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	11	2
05188900-bde3-11eb-8ff7-2dffe8e35377	5/26/2021 11:35	5/26/2021 11:54	1125	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	8	1
0519d950-bde3-11eb-b905-ebad4cf144cd	5/26/2021 20:32	5/26/2021 20:45	786	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	2
052184a0-bde3-11eb-95d9-b99a1cd9da42	5/26/2021 8:20	5/26/2021 8:27	431	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	30	2
05428110-bde3-11eb-b235-5f03b60a1860	5/26/2021 11:07	5/26/2021 21:23	36973	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	1
05babee0-bde3-11eb-8f3c-d9807c94a3c2	5/26/2021 8:51	5/26/2021 9:46	3274	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	30	2
05c8e6a0-bde3-11eb-971b-87bc9ca6b165	5/26/2021 8:03	5/26/2021 8:31	1667	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	50	1
067777a0-bde3-11eb-9633-2f8aee7721d9	5/26/2021 9:24	5/26/2021 9:36	710	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	10	2
06c6ec40-bde3-11eb-878f-17b3d3727f89	5/27/2021 13:56	5/27/2021 14:04	478	I have read the above, I agree and I am interested in participating in the research.	Yes, I want to continue.	10	1

Figure 3.3: Descriptive columns in actual dataset

2. **Demographic info:** This part describes some columns: Sex, Age (18-86), District, Religion, Marital status, Education, Education_Other, Conservative & Liberal and Household.

sex	age	district	education	education_other	conlib	religion	marital_status	household
woman	34	PieÅ¡Å¥any	high school		4	4	married	5
man	64	DunajskÃ¡ Streda	university (Mgr., Ing.)		3	6	married	2
woman	38	KoÅ¡ice II	university (Mgr., Ing.)		3	4	married	4
woman	28	Å½iar nad Hronom	university (Mgr., Ing.)		4	3	single	2
woman	29	Prievidza	university (Mgr., Ing.)		5	2	single	3
woman	28	Nitra	high school without graduation		4	5	single	4
woman	34	Bratislava IV	university (Mgr., Ing.)		5	1 - religion is not importa	single	1
man	40	KoÅ¡ice II	high school		4	5	married	3
man	49	Prievidza	high school without graduation		4	6	married	2
man	81	Martin	university (Mgr., Ing.)		4	6	widow/widower	2
man	72	KoÅ¡ice I	university (PhD.)		6	1 - religion is not importa	married	2
woman	32	Å½ilina	high school		4	7 - religion is very impor	married	4
woman	24	KysuckÃ© NovÃ© Mestc	high school		4	7 - religion is very impor	single	4
man	49	Hlohovec	high school without graduation		5	4	married	3
woman	29	Medzilaborce	university (Bc.)		5	2	single	2
woman	31	LevoÄa	university (Mgr., Ing.)		4	7 - religion is very impor	married	4
woman	45	KoÅ¡ice I	university (Mgr., Ing.)		3	4	divorced	3
woman	35	Levice	high school		4	4	single	1
woman	35	TrenÄÃ-n	university (Mgr., Ing.)		4	3	married	4
woman	18	Poprad	primary school		6	4	single	2
woman	26	KoÅ¡ice III	university (Mgr., Ing.)		4	2	single	3
woman	42	BanskÃ¡ Bystrica	high school without graduation		5	1 - religion is not importa	married	4
woman	29	Levice	university (Mgr., Ing.)		4	1 - religion is not importa	single	2
woman	33	Poprad	high school		4	3	married	2
woman	28	KoÅ¡ice-okolie	high school		4	3	single	3
man	54	PreÅ¡ov	high school without graduation		3	4	married	3
man	68	Å½ilina	high school without graduation		3	2	married	2
woman	28	Bardejov	high school		4	4	married	4
woman	18	Martin	primary school		7 - very liberal	1 - religion is not importa	single	4
man	50	Michalovce	high school		2	6	married	4
woman	25	NovÃ© ZÃ¡mky	high school		7 - very liberal	4	single	2
woman	41	HumennÃ©	high school		1 - very conservative	4	divorced	6
woman	32	Å½ilina	university (Mgr., Ing.)		4	4	married	2
man	75	Å½ilina	university (PhD.)		3	7 - religion is very impor	married	3
man	63	Å½ilina	university (Mgr., Ing.)		3	6	married	2

Figure 3.4: Demographic info columns in actual dataset

Also, here is the graphical representation of demographic fields:-

This bar plot illustrates the age distribution of survey participants. Ages are depicted on the X-axis, while the count of participants is on the Y-axis. The columns display the frequency of participants at different ages, arranged in ascending order. The survey indicates that 32-year-olds comprise the largest group, while 78-year-olds represent the smallest.

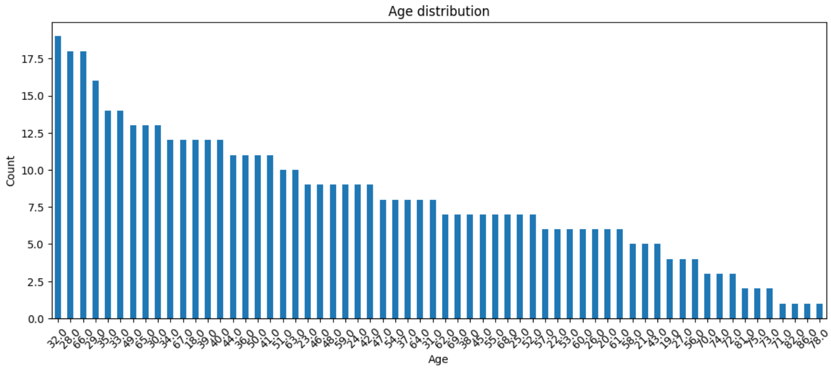


Figure 3.5: Age distribution

This histogram highlights that most participants have a high school education, while ”other” is the least common category.

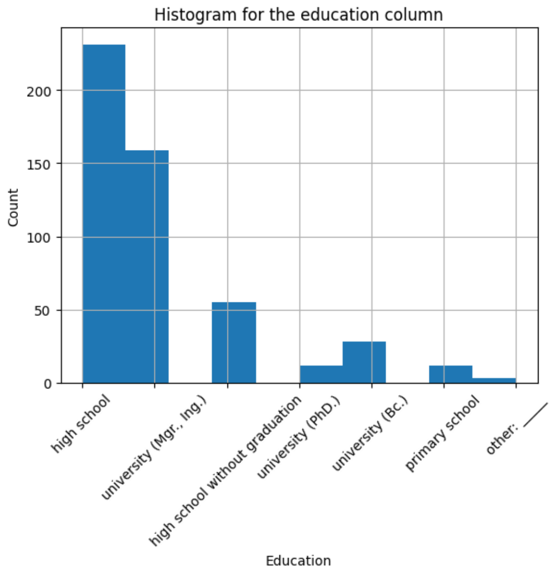


Figure 3.6: Histogram of education column value

This histogram shows the count of males and females. We put the sex variable on the X-axis and the count on the Y-axis. We see both male and female column shows the same value.

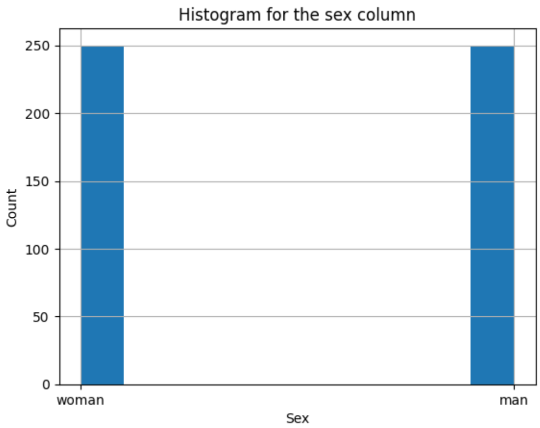


Figure 3.7: Histogram of gander(sex) column value

This histogram represents the religion column. We put the religious importance on the X-axis and the count on the Y-axis. We see that the peak plot is that religion is not important to me at all.

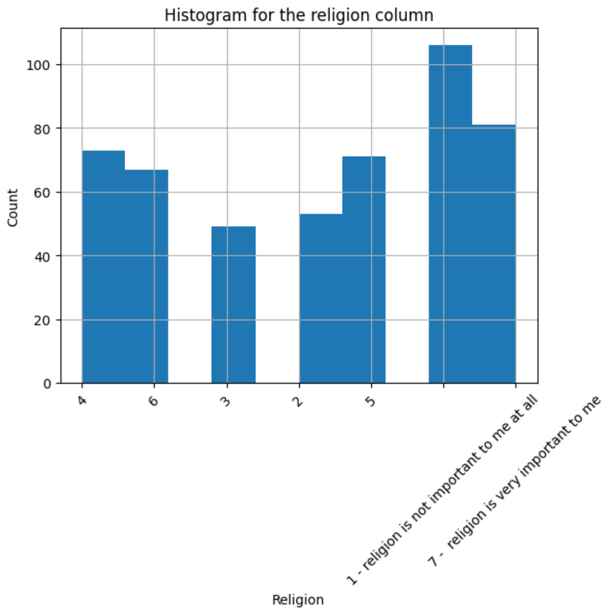


Figure 3.8: Histogram of religion column value

This histogram shows the marital condition of the participants in this survey. We see married person shows more interest in completing the survey.

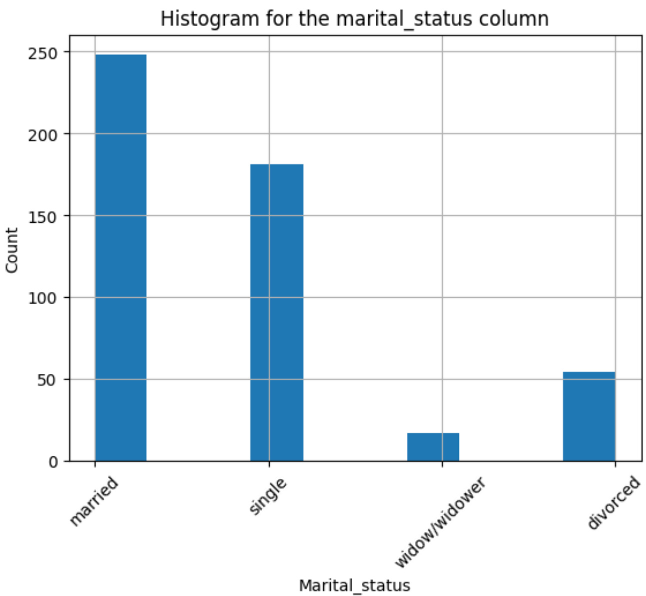
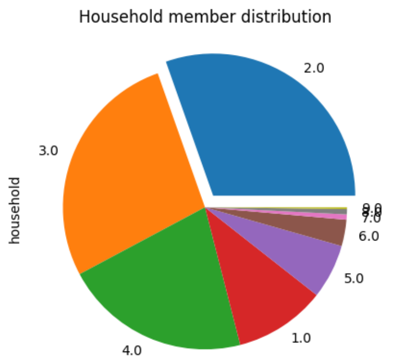
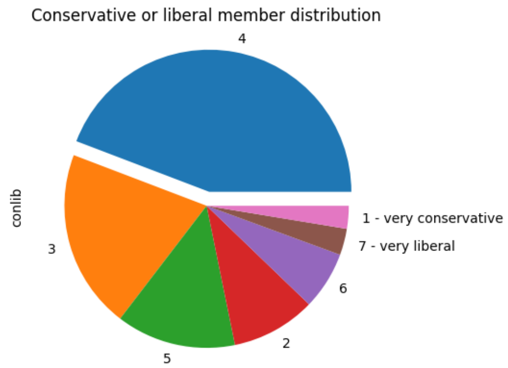


Figure 3.9: Histogram of marital status column value

Figure 3.10 shows two pie charts one is for households and the other is for conlib. In the pie chart (a) we see the number of households 2 is the most counted value. And in pie chart (b) we see the value 4 is the most counted.



(a) Count plot of household column value



(b) Count plot of Conservative & Liberal column value

Figure 3.10: Final Figure

Mixed histogram off age and sex column with target column. Graphical representation is:

This is a stack bar plot where we put the age variables in the X-axis and the Y-axis is their count where males and females are stacked one after another. And color them using the male and female.

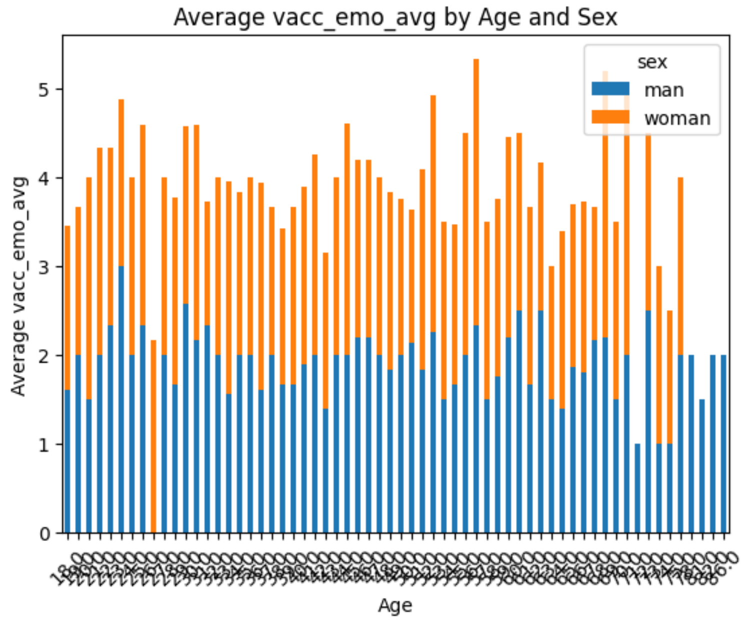


Figure 3.11: Graphical representation of gander and age with target column with stack bar plot

3. **Individual Differences:** To understand the individual differences they used Big five inventory) The Big Five personality traits [10], also known as the Five-Factor Model, provide a framework for understanding human personality. In the dataset to describe this, comprises 30 items/columns forming 5 columns and scale. This 5 average columns are -

   - `BFI Agreeableness(BFI_Agree)`: Agreeableness reflects kindness, cooperativeness, and empathy
   - `BFI OpenMindness(BFI_Open)`: This trait emphasizes imagination, curiosity, and openness to new experiences.
   - `BFI Conscientiousness(BFI_Consc)`: Conscientious individuals are organized, responsible, and disciplined. High conscientiousness is associated with success in work and personal life.
   - `BFI Extraversion(BFI_Extra)`: Extraversion (or extroversion) refers to sociability, assertiveness, and positive emotionality.
   - `BFI NegativeEmotionality(BFI_Negam)`: Negative emotionality involves emotional instability, anxiety, and moodiness.

Dataset looks:

1	bfi1	bfi2	bfi3	bfi4	bfi5	bfi6	bfi7	bfi8	bfi9	bfi10	bfi11	bfi12	bfi13	bfi14	bfi15
2	4	4	neutral	4	totally agr	2	totally agr	4	2	4	neutral	neutral	4	2	neutral
3	2	4	totally agr	I don't agr	4	4	totally agr	2	I don't agr	4	4	totally agr	totally agr	I don't agr	totally agr
4	2	totally agr	totally agr	4	4	2	totally agr	4	2	neutral	4	4	4	2	4
5	2	4	4	neutral	2	2	4	neutral	neutral	neutral	4	4	4	neutral	neutral
6	2	4	4	2	neutral	2	4	4	2	neutral	neutral	neutral	4	2	neutral
7	totally agr	4	totally agr	4	2	4	totally agr	totally agr	I don't agr	neutral	4	4	totally agr	2	4
8	I don't agr	neutral	neutral	neutral	2	2	2	neutral	neutral	totally agr	neutral	neutral	neutral	neutral	2
9	4	4	4	2	neutral	neutral	neutral	4	2	4	4	4	4	2	4
10	2	4	neutral	4	4	neutral	neutral	2	4	2	4	4	4	2	4
11	neutral	neutral	neutral	neutral	neutral	neutral	2	neutral	neutral	2	neutral	neutral	neutral	2	neutral
12	4	totally agr	totally agr	4	4	totally agr	4	totally agr	I don't agr	totally agr	totally agr	4	totally agr	I don't agr	4

1	bfi16	bfi17	bfi18	bfi19	bfi20	bfi21	bfi22	bfi23	bfi24	bfi25	bfi26	bfi27	bfi28	bfi29	bfi30
2	2	totally agr	neutral	2	4	4	4	4	2	4	4	totally agr	totally agr	I don't agr	totally agr
3	4	neutral	totally agr	I don't agr	4	totally agr	totally agr	totally agr	2	totally agr	totally agr	totally agr	4	2	4
4	4	4	4	2	neutral	neutral	4	totally agr	2	4	totally agr	totally agr	neutral	2	totally agr
5	4	4	totally agr	neutral	2	neutral	totally agr	4	neutral	neutral	neutral	4	4	2	neutral
6	4	4	4	2	2	neutral	4	4	2	4	4	4	4	2	neutral
7	4	4	totally agr	2	2	4	4	totally agr	2	neutral	4	neutral	4	4	4
8	2	2	2	neutral	I don't agr	neutral	neutral	neutral	4	2	2	neutral	neutral	neutral	2
9	4	4	4	2	neutral	4	totally agr	totally agr	2	totally agr	4	4	neutral	2	4
10	4	neutral	4	2	neutral	2	4	4	2	4	2	2	2	neutral	neutral
11	4	neutral	neutral	neutral	neutral	neutral	neutral	neutral	2	4	neutral	neutral	neutral	neutral	neutral
12	4	totally agr	totally agr	2	4	totally agr	totally agr	totally agr	2	totally agr	totally agr	4	4	4	totally agr

Figure 3.12: 30 columns of BFI

BFI_extra	BFI_agree	BFI_consc	BFI_negem	BFI_open
3.166666667	4.333333333	3.833333333	2.166666667	4.166666667
4	4.5	4.333333333	1.333333333	4.333333333
3.333333333	4.5	4.166666667	2.333333333	3.833333333
3	4.166666667	4	2.833333333	2.666666667
3	3.833333333	4	2	3
4.166666667	4	4.833333333	2.5	3
2.166666667	2.666666667	2.833333333	3.166666667	2.333333333
3.833333333	4	4	2	3.833333333
2.833333333	3.333333333	3.166666667	2.833333333	3.333333333
3.166666667	2.833333333	3	2.666666667	3
4.666666667	4.5	4.833333333	2.333333333	4.5
4.5	4.333333333	4.5	2.166666667	3.5
3	4.333333333	3.666666667	2.333333333	3.5
4.666666667	3.166666667	4.666666667	1.833333333	2.666666667
4.166666667	4	4.166666667	3.833333333	4.5
4.5	4.166666667	4.666666667	3	4.333333333
4	3.5	4	1.666666667	3.5
3	4	3	3.666666667	4.166666667
3.833333333	3.833333333	3.833333333	2.166666667	3.333333333
4.166666667	3	3.833333333	3.833333333	4.5

Figure 3.13: 5 average columns of 30 columns

Here is the line plot of BFI average columns data along with the date: This line plot represents the BFI value against the day column input. It shows us the frequency of the input of BFI during the whole time.

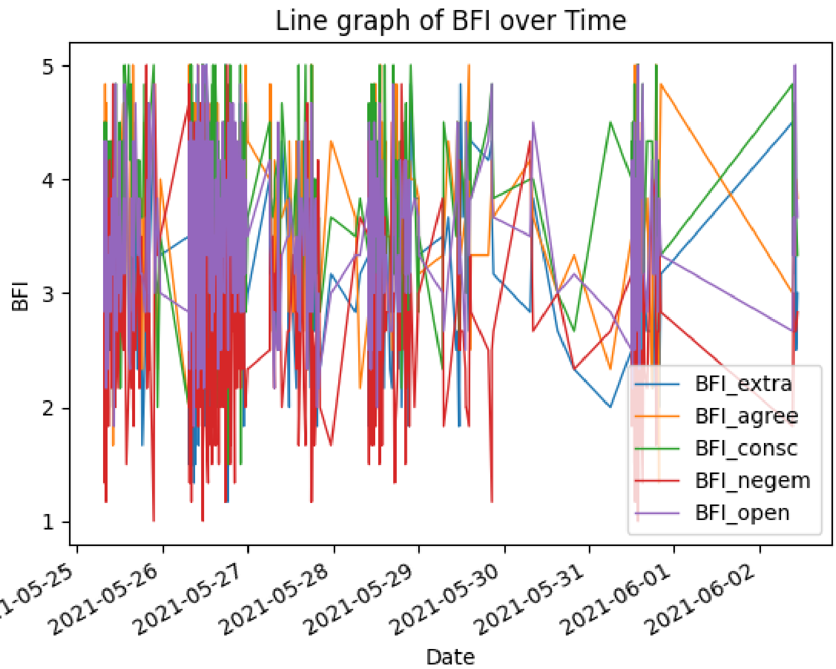


Figure 3.14: Line plot of BFI average columns

To capture Collectivism/Individualism here employed the Horizontal-Vertical Individualism-Collectivism (HVIC) scale [11] [18] , which comprises 14 items forming four sub-scales. The 4 columns of scale are-

- `Horizontal Collectivism (HVIC_HC_avg)`: In HC, the emphasis is on equality, interdependence, and harmony within a group.
- `Vertical Collectivism (HVIC_VC_avg)`: VC highlights hierarchy, status, and conformity to authority.
- `Horizontal Individualism (HVIC_HI_avg)`: HI emphasizes independence, personal autonomy, and equality.
- `Vertical Individualism (HVIC_VI_avg)`: VI combines elements of individualism with a hierarchical perspective.

HVIC_1	HVIC_2	HVIC_3	HVIC_4	HVIC_5	HVIC_6	HVIC_7	HVIC_8	HVIC_9	HVIC_10
5 - completel	4	5 - completel	4	5 - completel	3	3	3	3	5 - completel
4	4	4	4	4	5 - completel	4	4	4	4
4	4	4	3	4	4	3	4	3	4
4	4	3	2	4	3	3	4	3	4
4	4	3	3	4	2	4	3	3	4
4	5 - completel	3	4	4	4	3	4	4	4
2	2	2	3	3	3	3	4	2	3
4	4	3	3	4	4	4	4	3	4
3	3	4	3	3	3	3	3	3	3
3	3	3	3	3	3	5 - completel	3	3	3
4	4	4	4	5 - completel	4	5 - completel	5 - completel	5 - completel	4
3	5 - completel	4	4	4	4	5 - completel	4	5 - completel	5 - completel
5 - completel	4	4	4	4	5 - completel	5 - completel	5 - completel	3	4

HVIC_11	HVIC_12	HVIC_13	HVIC_14	HVIC_15	HVIC_16	HVIC_17	HVIC_18	HVIC_19	HVIC_20
5 - completel	5 - completel	3	3	5 - completel	5 - completel	5 - completel	4	3	3
4	4	4	5 - completel	4	4	3	5 - completel	4	4
4	4	3	4	4	3	4	4	4	3
4	4	3	4	5 - completel	3	2	4	3	3
3	4	3	3	4	4	3	4	3	4
3	5 - completel	4	4	5 - completel	5 - completel	2	4	3	4
4	2	2	2	3	3	3	3	3	4
4	3	3	5 - completel	4	4	3	4	4	4
3	3	3	3	4	4	3	3	4	4
3	4	3	3	3	3	3	3	4	5 - completel
5 - completel	3	5 - completel	5 - completel	4	4	4	5 - completel	4	5 - completel
3	3	5 - completel	4	2	4	3	5 - completel	3	5 - completel
5 - completel	4	3	3	4	4	5 - completel	4	4	4

Figure 3.15: HVIC 20 columns

HVIC_HC_avg	HVIC_VC_avg	HVIC_VI_avg	HVIC_HI_avg
4.5	4.25	3.333333333	3.666666667
4.25	4	4	4.333333333
4	3.75	3	4
4	3.5	2.666666667	3.666666667
3.75	3.75	3	2.666666667
4	4	4	3.666666667
2.5	2.25	2.333333333	3.666666667
4.25	3.5	3	4
3	3.25	3	3
3	3.75	3	3
4.5	4	4.666666667	4.666666667

Figure 3.16: Four average columns of HVIC

Here is the line plot of HVIC average columns over the data collection date:

Here is the line plot of HVIC average columns data along with the date: This line plot represents the HVIC value against the day column input. It shows us the frequency of the input of HVIC during the whole time.

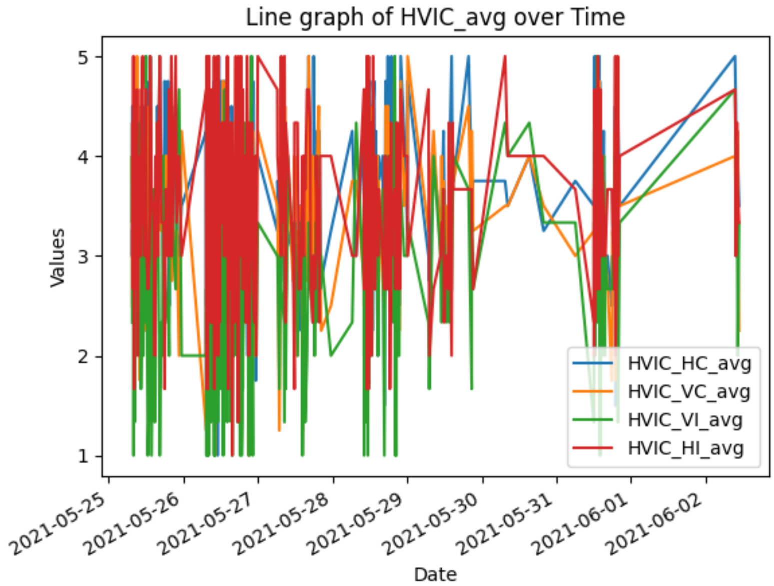


Figure 3.17: Line plot if four average columns of HVIC

To measure whether participants prefer immediate rewards or focus on future and distant outcomes, we used the Consideration of Future Consequences (CFC) [12] scale.The scale has 14 items assigned to two factors: CFC-Future (e.g., "Often I engage in a particular behaviour to achieve outcomes that may not result for many years.") and CFC-Immediate (e.g., "I generally ignore warnings about possible future problems because I think the problems will be resolved before they reach a crisis level."). Par-ticipants responded on a scale from 1 (strongly disagree) to 7 (strongly agree) and the mean scores were calculated. Actual name of the two columns in dataset are -(`CFC_F_avg`, `CFC_I_avg`).

futorient_1	futorient_2	futorient_3	futorient_4	futorient_5	futorient_6	futorient_7	futorient_8	futorient_9	futorient_10	futorient_11	futorient_12	futorient_13	futorient_14
5 - completely	6	3	4	3	4	6	6	2	3	4	4	6	5 - completely
5 - completely	5 - completely	2	2	2	6	6	6	2	2	2	2	6	6
5 - completely	3	5 - completely	4	3	5 - completely	6	5 - completely	2	2	3	3	4	4
5 - completely	4	2	3	4	5 - completely	5 - completely	5 - completely	5 - completely	4	2	3	6	5 - completely
5 - completely	5 - completely	3	4	4	5 - completely	5 - completely	3	2	3	3	4	5 - completely	5 - completely
6	5 - completely	4	4	4	5 - completely	5 - completely	5 - completely	3	2	2	3	5 - completely	4
4	3	4	3	5 - completely	4	4	4	3	3	3	3	4	4
5 - completely	6	2	2	3	3	5 - completely	5 - completely	2	4	2	3	6	5 - completely
5 - completely	5 - completely	5 - completely	5 - completely	5 - completely	5 - completely	6	6	5 - completely	5 - completely	5 - completely	5 - completely	5 - completely	5 - completely
1 - not at all	1 - not at all	1 - not at all	4	4	4	4	4	4	4	7	1 - not at all	4	4
7	7	1 - not at all	2	1 - not at all	7	7	6	1 - not at all	1 - not at all	1 - not at all	2	7	7
5 - completely	5 - completely	3	4	4	7	7	6	5 - completely	7	6	6	6	6
5 - completely	5 - completely	4	4	5 - completely	5 - completely	5 - completely	4	5 - completely	5 - completely	4	5 - completely	4	4

Figure 3.18: CFC 14 columns

CFC_F_avg	CFC_I_avg
5.428571429	3.285714286
5.714285714	2
4.571428571	3.142857143
5	3.285714286
4.714285714	3.285714286
5	3.142857143
3.857142857	3.428571429
5	2.571428571

Figure 3.19: Two average columns of CFC

Here is the data distribution line plot of CFC average columns over the time :

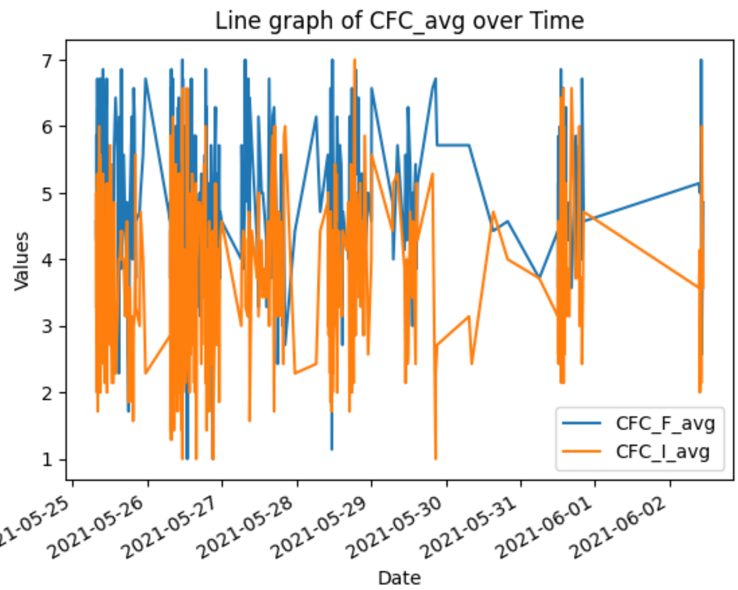


Figure 3.20: Line plot of two average columns of CFC

General Prosocial tendencies were rated using the 23-item Prosocial Tendencies Measure (PTM).The scale ranging from 1 (not at all like me) to 5 (absolutely like me). Finally it form one average item/column in the dataset. The column name is (`prosoc_ALL_avg`) Dataset looks:

prosoc_mot_9	prosoc_mot_10	prosoc_mot_11	prosoc_mot_12	prosoc_mot_13	prosoc_mot_14	prosoc_mot_15	prosoc_mot_16	prosoc_mot_17	prosoc_mot_18	prosoc_mot_19	prosoc_mot_20	prosoc_mot_21	prosoc_mot_22	prosoc_mot_23
4	3	4	3	1 - not at all	5 - completely	3	1 - not at all	3	5 - completely	3	3	3	5 - completely	3
5 - completely	5 - completely	4	3	3	3	4	3	4	4	2	2	4	2	2
4	3	4	3	1 - not at all	4	4	1 - not at all	2	4	4	1 - not at all	3	5 - completely	1 - not at all
4	3	4	3	1 - not at all	3	3	1 - not at all	2	4	3	2	2	3	1 - not at all
3	4	3	3	2	3	3	2	3	4	3	3	3	4	2
4	1 - not at all	5 - completely	4	1 - not at all	5 - completely	4	1 - not at all	3	5 - completely	5 - completely	1 - not at all	4	5 - completely	1 - not at all
2	3	2	2	2	3	2	2	2	2	2	2	2	2	2
4	4	2	2	3	5 - completely	3	3	3	4	3	3	3	3	4
4	4	4	4	4	4	4	3	4	4	4	4	4	4	3
3	3	3	1 - not at all	3	3	3	1 - not at all	1 - not at all	3	3	1 - not at all	1 - not at all	1 - not at all	3
5 - completely	4	3	4	4	4	3	1 - not at all	4	5 - completely	4	3	5 - completely	1 - not at all	5 - completely
5 - completely	2	5 - completely	3	2	4	3	3	2	3	5 - completely	5 - completely	4	4	5 - completely
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	5 - completely	1 - not at all	1 - not at all	4	4	1 - not at all	1 - not at all	1 - not at all	4	1 - not at all	5 - completely	1 - not at all	1 - not at all	5 - completely
4	5 - completely	3	4	3	4	3	2	3	5 - completely	3	3	3	1 - not at all	3
5 - completely	3	4	3	1 - not at all	4	4	1 - not at all	3	4	1 - not at all	1 - not at all	5 - completely	3	3
4	1 - not at all	4	3	1 - not at all	5 - completely	3	1 - not at all	1 - not at all	4	4	1 - not at all	3	1 - not at all	1 - not at all
5 - completely	1 - not at all	5 - completely	5 - completely	1 - not at all	5 - completely	5 - completely	1 - not at all	5 - completely	5 - completely	1 - not at all	1 - not at all	1 - not at all	5 - completely	4
3	3	3	3	3	4	3	3	3	4	3	3	3	3	2
5 - completely	5 - completely	1 - not at all	5 - completely	4	5 - completely	1 - not at all	1 - not at all	3	5 - completely	1 - not at all	1 - not at all	5 - completely	1 - not at all	3
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	1 - not at all	4	2	1 - not at all	3	3	1 - not at all	2	3	3	1 - not at all	2	4	1 - not at all
2	2	2	3	2	3	2	1 - not at all	3	4	3	1 - not at all	2	2	2

Figure 3.21: Prosocial tendencies 23 columns

prosoc_ALL_avg
3.886111111
3.516666667
3.502777778
3.247222222
3.216666667
4.122222222
2.472222222
3.255555556
3.7
2.555555556

Figure 3.22: Single average columns of Prosocial tendencies

Here is the line plot of average column: Here is the line plot of prosocial tendencies average columns data along with the date: This line plot represents the prosocial tendencies value against the day column input. It shows us the frequency of the input of prosocial tendencies during the whole time.

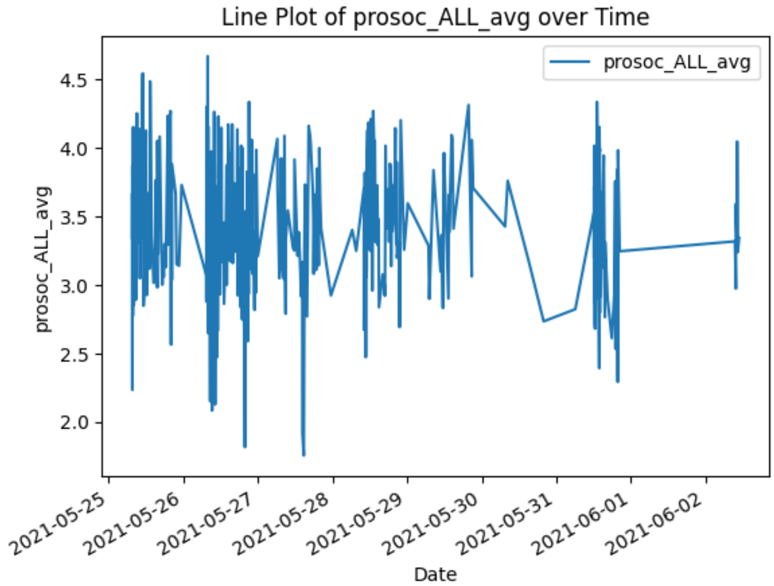


Figure 3.23: Prosocial tendencies average line plot

4. **Emotional Response to Threats:** Here measured feelings of helplessness related to the vaccination, the COVID-19 pandemic, and climate change by four items each. Participants responded on a 7-point scale (1 = completely disagree , 7 = completely agree). Here we got 3 avrage columns/items which comprises 12 columns. The columns name are -

   - `helpless_CLIM_avg`: Feelings of helplessness related to the climate crisis.
   - `helpless_COV_avg`: Feelings of helplessness related to the COVID-19 pandemic.
   - `helpless_VAC_avg`: Feelings of helplessness related to the vaccination.

Bclima_1	Bclima_2	Bclima_3	Bclima_4	Bcovid_1	Bcovid_2	Bcovid_3	Bcovid_4	Bvac_1	Bvac_2	Bvac_3	Bvac_4
3	3	4	4	4	3	2	4	4	4	4	4
6	5	5	2	3	3	3	2	2	3	2	2
5	6	5	5	5	5	5	4	6	5	6	4
3	4	6	5	5	4	5	4	5	5	4	5
5	5	4	4	4	3	3	3	3	3	3	2
5	6	6	5	6	6	6	4	7 - totally agree	6	7 - totally agree	3
4	5	5	5	2	2	2	2	2	2	2	2
5	5	5	4	3	3	3	3	3	3	3	3
4	5	5	5	4	5	6	4	5	6	6	4
4	4	4	4	4	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree
6	5	4	4	6	5	6	5	3	2	5	1 - totally disagree
6	6	5	5	6	3	4	4	5	6	7 - totally agree	7 - totally agree
4	4	4	4	4	4	4	4	4	4	4	4
3	4	4	1 - totally disagree	5	4	4	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree
3	3	4	2	7 - totally agree	1 - totally disagree	7 - totally agree	1 - totally disagree	1 - totally disagree	1 - totally disagree	7 - totally agree	7 - totally agree
7 - totally agree	6	7 - totally agree	5	7 - totally agree	5	7 - totally agree	5	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree
5	5	4	5	3	2	3	3	4	5	5	6
6	4	4	1 - totally disagree	4	2	2	2	2	2	2	1 - totally disagree
3	3	3	3	3	3	3	3	4	5	4	4
3	5	5	3	4	3	3	4	1 - totally disagree	1 - totally disagree	1 - totally disagree	1 - totally disagree

Figure 3.24: Helplessness 12 columns

helpless_CLIM_avg	helpless_COV_avg	helpless_VAC_avg
3.5	3.25	4
4.5	2.75	2.25
5.25	4.75	5.25
4.5	4.5	4.75
4.5	3.25	2.75
5.5	5.5	5.75

Figure 3.25: Three average columns of helplessness

Here is the line plot of helplessness against climate average columns data along with the date: This line plot represents the helplessness against climate value against the day column input. It shows us the frequency of the input of helplessness against climate during the whole time.

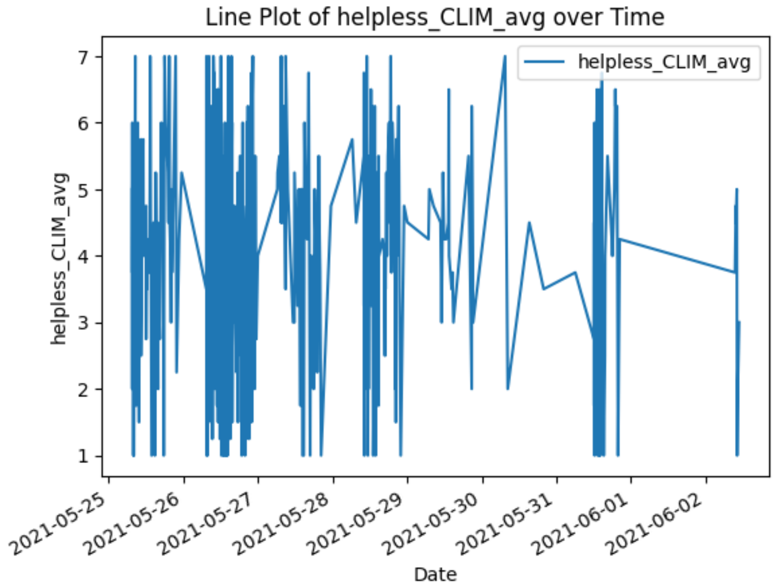


Figure 3.26: Helplessness about climate column value line plot

Feelings of threat [14] related to vaccination against the COVID-19 disease, the disease itself, and climate change were assessed using three questions in each domain. In this part we got 3 final average columns from 9 columns. The 3 columns are-

- `threat_COV_avg`: Average threat level related to coverage (COV) factors.
- `threat_VAC_avg`: Average threat level related to vaccination (VAC) factors.
- `threat_CLIM_avg`: Average threat level related to climate (CLIM) factors.

Ocovid_1	Ocovid_2	Ocovid_3	Ovac_1	Ovac_2	Ovac_3	Oclima_1	Oclima_2	Oclima_3
4	4	4	4	4	4	4	4	4
4	4	4	5	5	4	5	5	5
5	5	5	6	6	5	5	5	5
6	5	5	4	3	3	5	4	4
5	5	5	4	4	4	6	5	6
6	6	5	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	6	7 - I feel very threatened	5
5	4	3	5	4	3	5	4	4
4	5	5	3	3	3	4	5	5
5	5	6	5	5	6	5	5	6
7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	I donâ€™t feel any threat at	I donâ€™t feel any threat at	I donâ€™t feel any threat at	4	4	4
2	2	2	2	2	2	3	3	3
5	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	4	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened
3	3	3	3	3	3	3	3	3
5	5	5	I donâ€™t feel any threat at	I donâ€™t feel any threat at	I donâ€™t feel any threat at	4	5	4
6	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	7 - I feel very threatened	I donâ€™t feel any threat at	I donâ€™t feel any threat at	I donâ€™t feel any threat at
4	5	7 - I feel very threatened	I donâ€™t feel any threat at	I donâ€™t feel any threat at	I donâ€™t feel any threat at	5	5	7 - I feel very threatened
4	5	5	5	5	4	4	4	4

Figure 3.27: Threat 9 columns

threat_COV_avg	threat_VAC_avg	threat_CLIM_avg
4	4	4
4	4.666666667	5
5	5.666666667	5
5.333333333	3.333333333	4.333333333
5	4	5.666666667
5.666666667	7	6
4	4	4.333333333

Figure 3.28: Three average columns of threat

The data about threat about the COVID-19, graphically represented in the line plot:

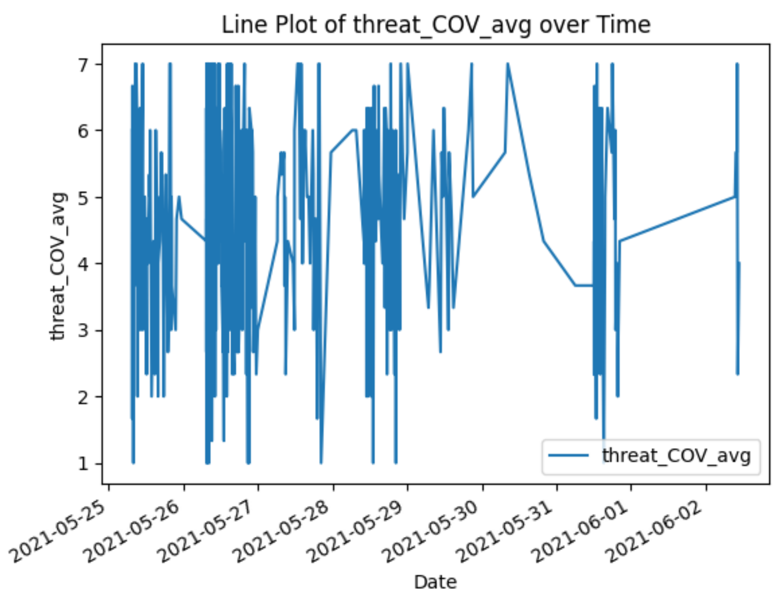


Figure 3.29: Threat about covid-19 column value line plot

5. **Prosocial Intentions and Behaviour [17]:** In this part we got clear knowledge of data which describe about the vaccination hesitancy and behavioral pattern of participants. Here also some sub parts-

   - Prosocial behavior encompasses actions that benefit others, even when there are potential costs to the helper. It includes acts of kindness, cooperation, and support for fellow humans.There are 21 columns where they take response from people. Then they take their average value that are related and put the value in 3 different columns. Columns are- (`procovid_avg`, `proenviro_avg`, `provacc_avg`).

procovid_avg	proenviro_avg	provacc_avg
2.4	2.363636364	1.75
2.8	2.909090909	1.5
2.4	2	1.75
2	1.636363636	1.75
2.2	2.090909091	1.5
2.4	1.818181818	1.75
2.2	1.636363636	2

Figure 3.30: Three average columns of prosocial behavior

Graphical representation of `procovid_avg` using line plot is given below.

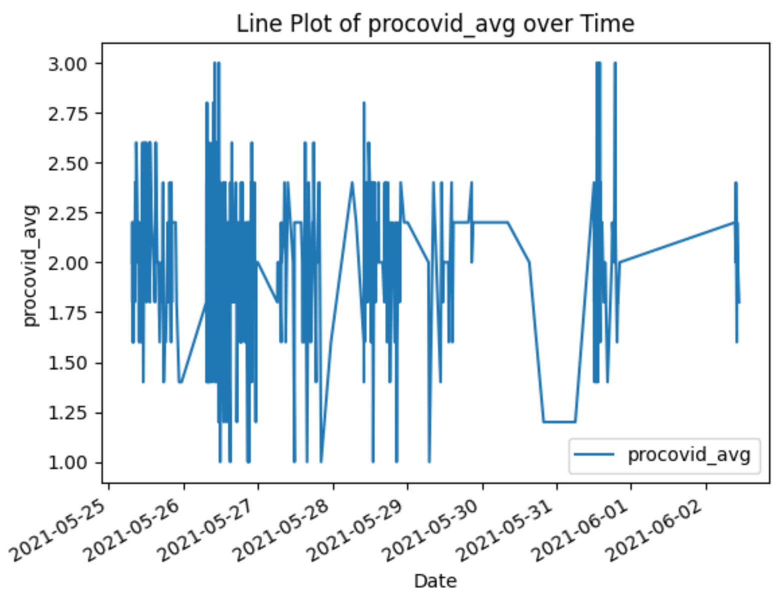


Figure 3.31: Line plot of procovid average column

- Attitudes towards vaccination (*attvac_avg*): There are 10 columns in the dataset to measure attitudes towards vaccination, and 5 average columns are created to aggregate related column values.

- Prosocial Public behavior(prosoc_PUB_avg):Includes behaviors such as wearing masks, social distancing, sharing information, and supporting local businesses.
- Prosocial Emotional behavior(prosoc_EMO_avg): Involves acts driven by empathy and compassion, understanding, and sharing the emotions of others.
- Prosocial Directive behavior(prosoc_DIR_avg): Includes providing clear guidance or instructions to others.
- prosoc_COM_avg: Prosocial Compliant - Refers to following rules, guidelines, and public health recommendations.
- Prosocial Anonymous behavior(prosoc_ANO_avg):Occurs when individuals help others without revealing their identity.

prosoc_PUB_avg	prosoc_EMO_avg	prosoc_DIR_avg	prosoc_COM_avg	prosoc_ANO_avg
3	3.25	4.666666667	5	3.6
3.25	3.75	4	4.5	2.8
1.75	3	3.666666667	4	4.2
2.25	2.5	3.333333333	4	3.4
2.5	3	3	4	3.4
2.5	3.5	4.333333333	5	4.8
2	2	2.333333333	2.5	2.2
3	3	4.333333333	4	2.8
4	4	4	4	3.8
3	1	2.333333333	3	2.2
4.25	4.25	4.333333333	5	2.4

Figure 3.32: Five average columns of attitudes towards vaccination on prosocial intention

prosoc_PUB_avg	prosoc_EMO_avg	prosoc_DIR_avg	prosoc_COM_avg	prosoc_ANO_avg
3	3.25	4.666666667	5	3.6
3.25	3.75	4	4.5	2.8
1.75	3	3.666666667	4	4.2
2.25	2.5	3.333333333	4	3.4
2.5	3	3	4	3.4
2.5	3.5	4.333333333	5	4.8
2	2	2.333333333	2.5	2.2
3	3	4.333333333	4	2.8
4	4	4	4	3.8
3	1	2.333333333	3	2.2
4.25	4.25	4.333333333	5	2.4

Figure 3.33: Five average columns of attitudes towards vaccination on prosocial intention

- Prosocial Altruistic: Altruistic acts involve selflessly helping others without expecting anything in return. During COVID-19, this could include volunteering to deliver groceries to vulnerable individuals or assisting healthcare workers. In this dataset 5 columns for taking response and make a average column for this. The average column name is (`prosoc_ALT_avg`).

prosoc_ALT_avg
3.8
2.8
4.4
4
3.4
4.6
3.8
2.4

Figure 3.34: Single average columns of prosocial altruistic

- Anti vaccination average: In the dataset there are 4 columns where we collect anti vaccination type response and also make a column to store the average value. The column name is (`antivacc_avg`).

antivacc1_R	antivacc7_R	antivacc9_R	antivacc10_R
2	5	4	2
2	4	3	2
2	2	2	2
4	3	3	4
2	3	2	2
2	3	3	2
2	2	2	2
3	4	3	3

Figure 3.35: Anti vaccination 4 columns

antivacc_avg
2.8
3.3
2.3
3
2.3
2.9
1.8
3
2.4

Figure 3.36: Single average columns of anti-vaccination

Line plot of antivacc average data:

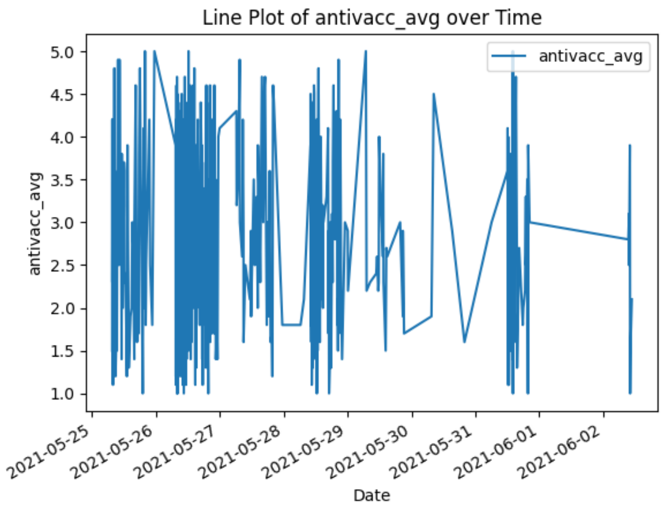


Figure 3.37: Line plot of single average columns of anti-vaccination

Heat map representation of all 30 average columns. From the heat map we can see that most of the columns are not highly co-related or not familiar. So we have to consider most of the columns for further operation.

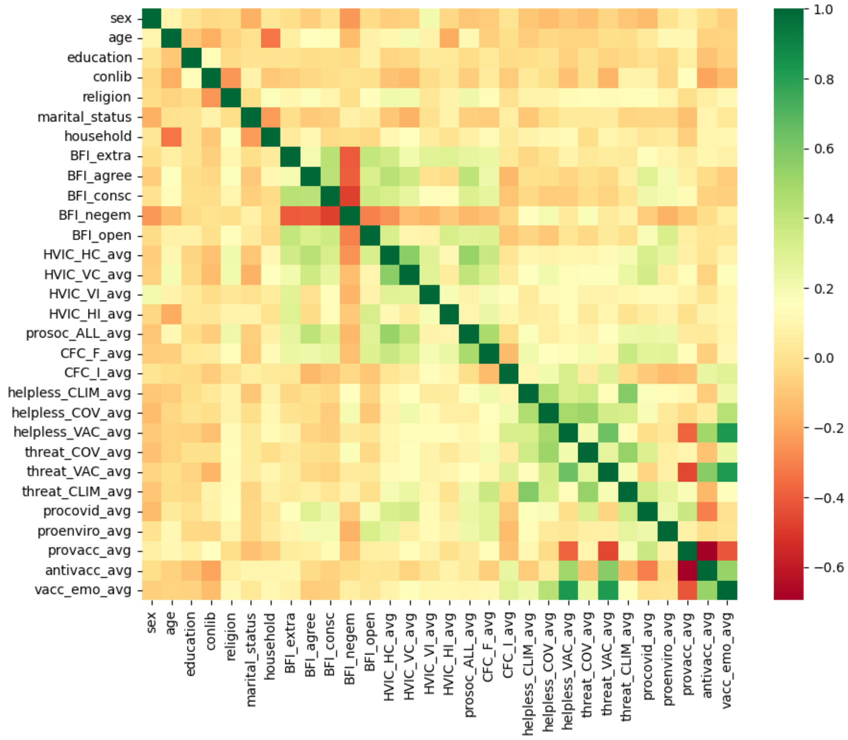


Figure 3.38: Heat map of 30 average columns

3.1.2 Validation dataset

We gathered 48 responses to validate our results. For this dataset, we initially selected 32 items for validation, but in the end, we only utilized 21 columns. Among these, 20 columns represent independent items, while 1 column serves as the dependent column (target column). These 32 columns correspond to the average columns from the original dataset that we described earlier.

L	M	N
16. What do you think "without competition, it is impossible to have a good society"?/ ত	17. Do you feel good when you help others?/আপনি যখন অন্যদেরকে সহযোগিতা করেন	18. Do you sacrifice your own interests for the good of others?/আপনি কি ত
4	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	3	
3	5 - Completely/সম্পূর্ণরূপে	
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	5 - Completely/সম্পূর্ণরূপে	
3	5 - Completely/সম্পূর্ণরূপে	
4	5 - Completely/সম্পূর্ণরূপে	
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	5 - Completely/সম্পূর্ণরূপে	
2	1 = Not at all/মোটেই না	5 - Completely/সম্পূর্ণরূপে
3	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	1 = Not at all/মোটেই না	1 = Not at all/মোটেই না
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	5 - Completely/সম্পূর্ণরূপে	1 = Not at all/মোটেই না
4	5 - Completely/সম্পূর্ণরূপে	
3	5 - Completely/সম্পূর্ণরূপে	
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
3	5 - Completely/সম্পূর্ণরূপে	
5 = Strongly agree/দৃঢ়ভাবে একমত	5 - Completely/সম্পূর্ণরূপে	
1 = Strongly disagree/দৃঢ়ভাবে অসম্মতি	5 - Completely/সম্পূর্ণরূপে	5 - Completely/সম্পূর্ণরূপে
3	4	

Figure 3.39: Snapshot of validation raw dataset

Collected data ratio representations:

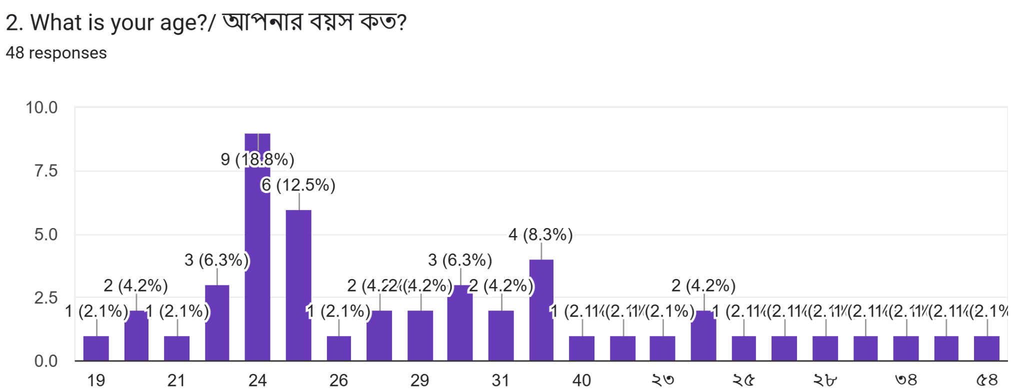


Figure 3.40: Collected age data ratio

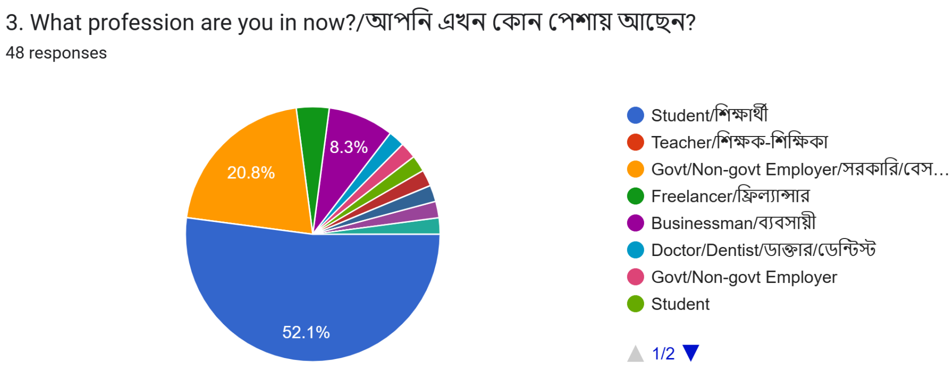


Figure 3.41: Collected profession data ratio

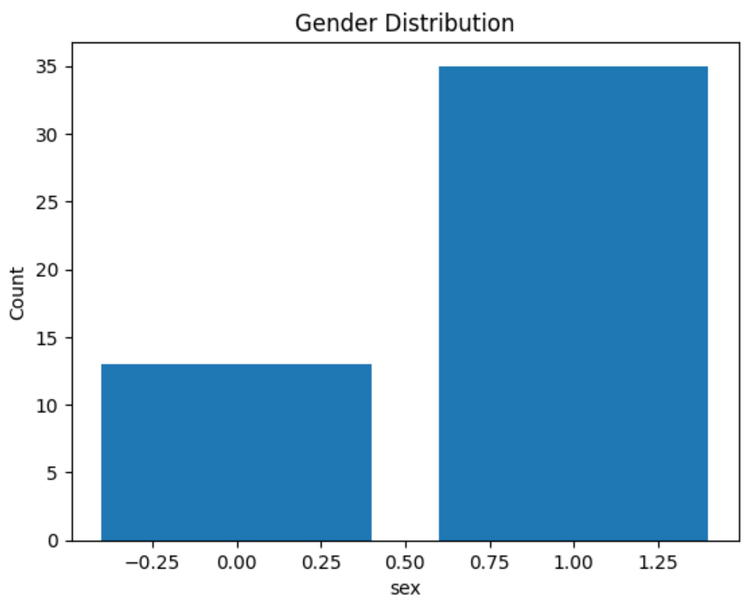


Figure 3.42: Collected gender data ratio

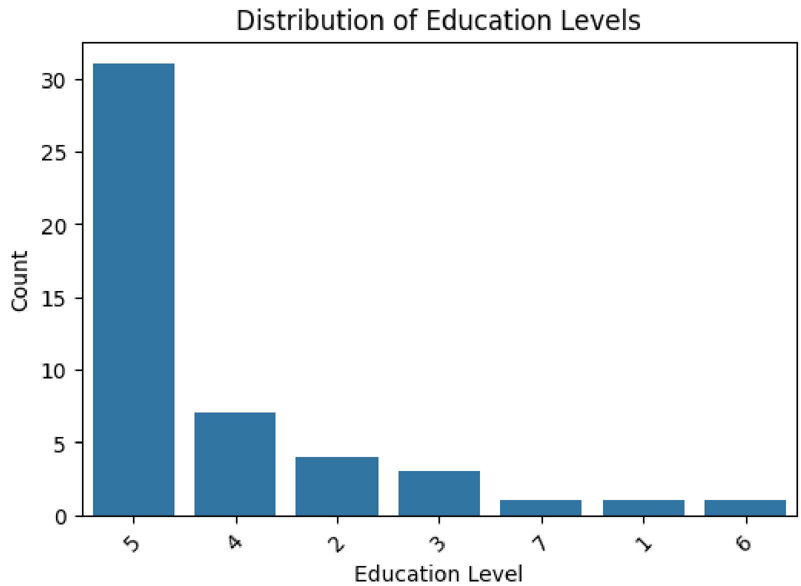


Figure 3.43: Collected education data ratio by count plot

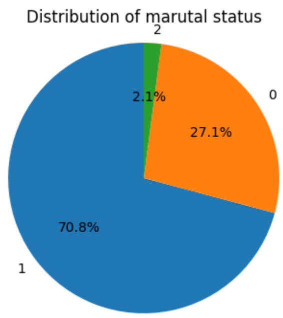


Figure 3.44: Collected marital status data ratio

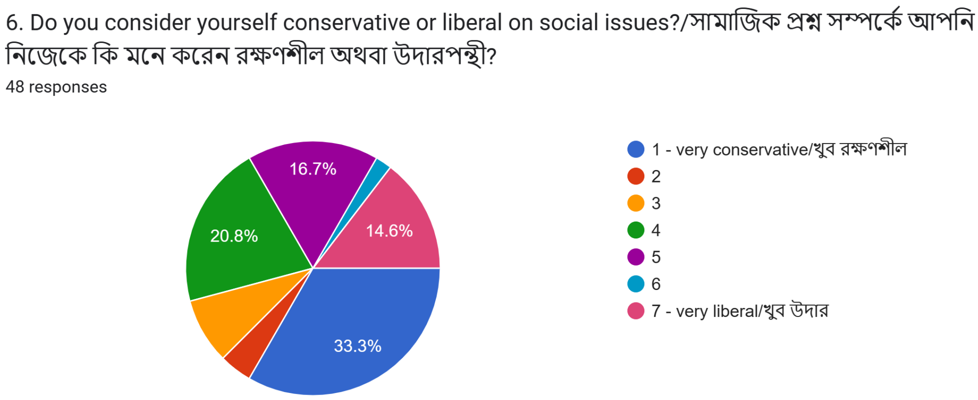


Figure 3.45: Collected conservative & liberal data ratio

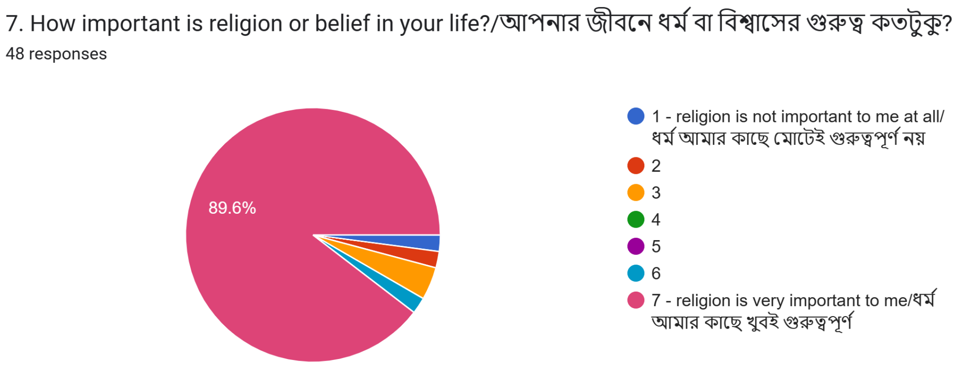


Figure 3.46: Collected religion data ratio

9. How many people live in your family including you?/আপনি সহ আপনার ফ্যামিলিতে কত জন মানুষ বসবাস করেন?
48 responses

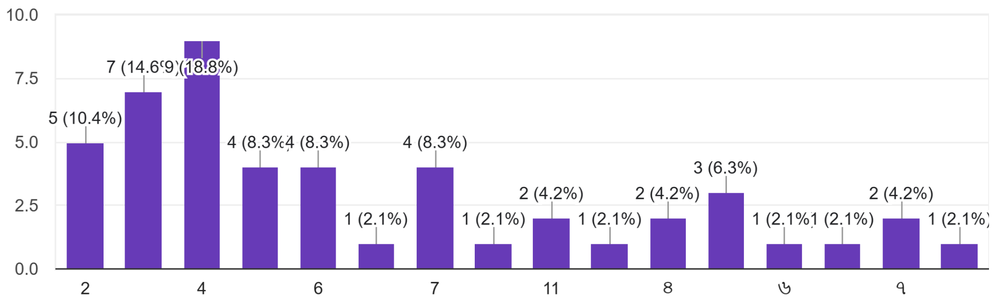


Figure 3.47: Collected house hold data ratio

10. Do you find it hard to influence people around you?/ আপনি কি আপনার চারপাশের লোকেদের প্রভাবিত করা কঠিন বলে মনে করেন? (BFI_extra)
48 responses

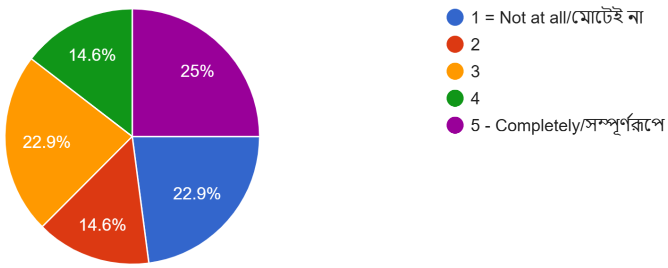


Figure 3.48: Collected BFI extra data ratio

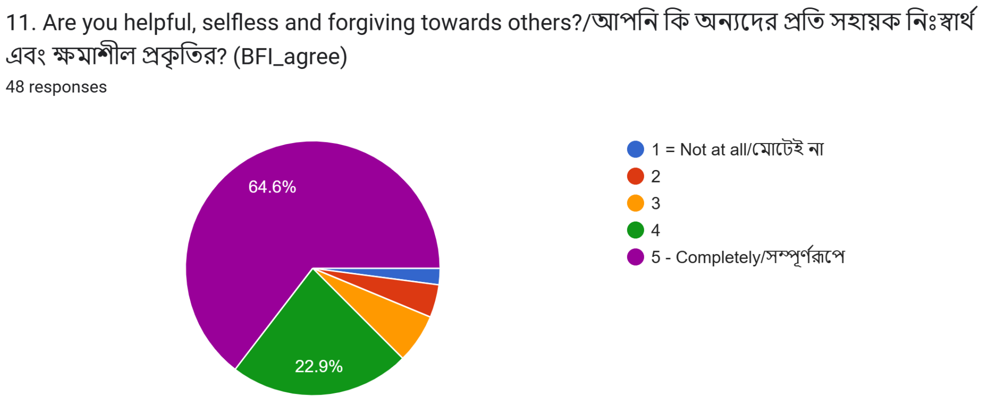


Figure 3.49: Collected BFI agree data ratio

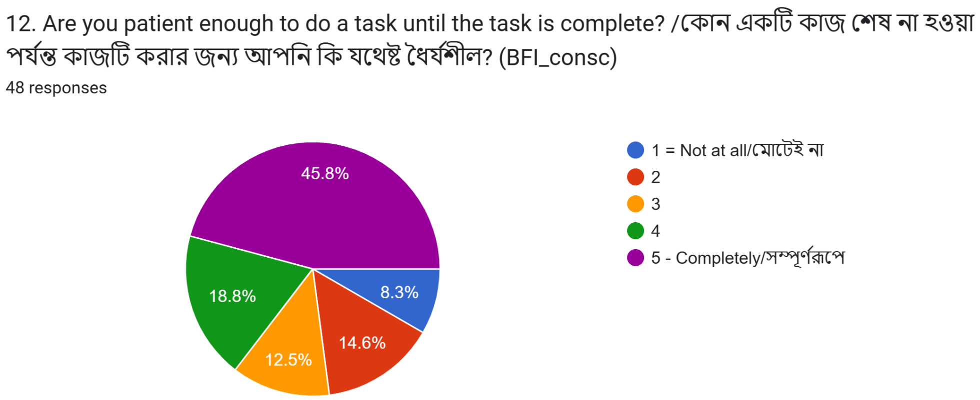


Figure 3.50: Collected BFI consc data ratio

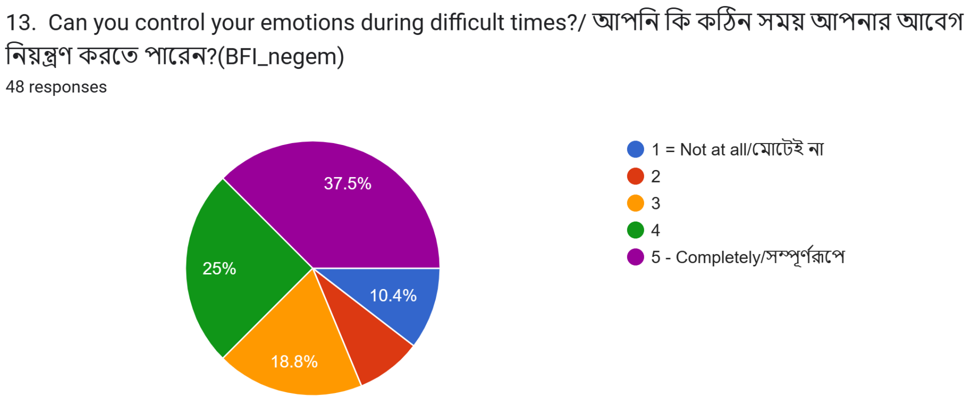


Figure 3.51: Collected BFI negative emotional data ratio

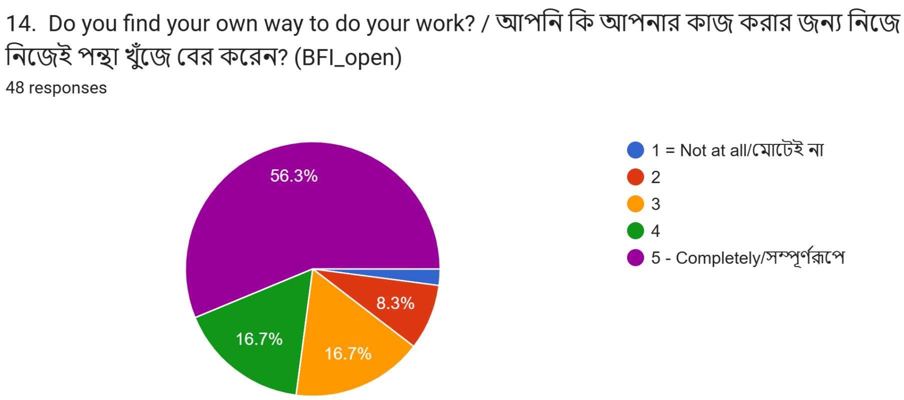


Figure 3.52: Collected BFI open data ratio

15. Do you prefer to do your own task?/আপনি কি আপনার কাজ নিজে করতেই পছন্দ করেন? (HVIC-HI)

48 responses

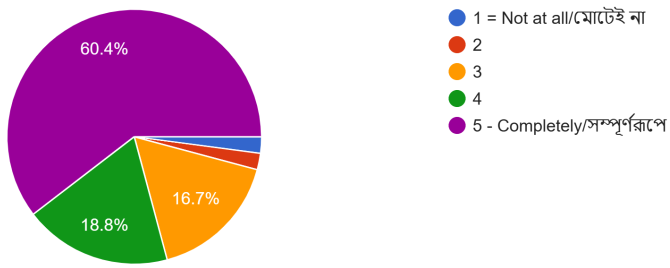


Figure 3.53: Collected HVIC HI data ratio

16. What do you think "without competition, it is impossible to have a good society"?/ আপনি কি মনে করেন "প্রতিযোগিতা ব্যতীত একটি ভাল সমাজ গড়ে তোলা অসম্ভব"? (HVIC-VI)

48 responses

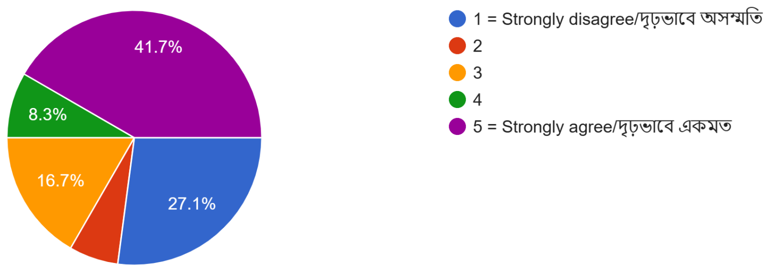


Figure 3.54: Collected HVIC VI data ratio

17. Do you feel good when you help others?/আপনি যখন অন্যদেরকে সহযোগিতা করেন তখন কি ভালো অনুভব করেন? (HVIC-HC)
48 responses

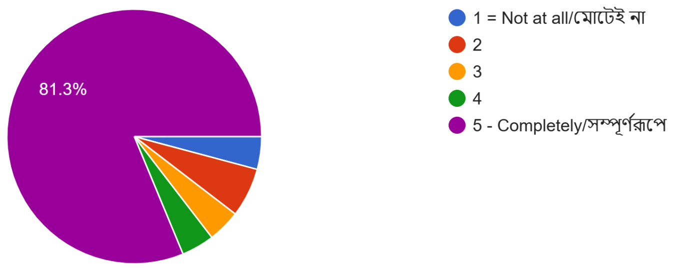


Figure 3.55: Collected HVIC HC data ratio

18. Do you sacrifice your own interests for the good of others?/আপনি কি অন্যের ভালোর জন্য নিজের স্বার্থ বিসর্জন দেন? (HVIC-VC)
48 responses

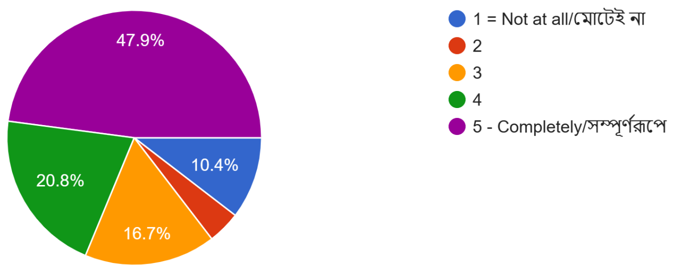


Figure 3.56: Collected HVIC VC data ratio

19. Do you feel that if you help someone, they should help you in the future?/ আপনি কি মনে করেন যে আপনি যদি কাউকে সাহায্য করেন, তারা ভবিষ্যতে আপনাকে সাহায্য করবে? (Prosoc_ALL_avg)
48 responses

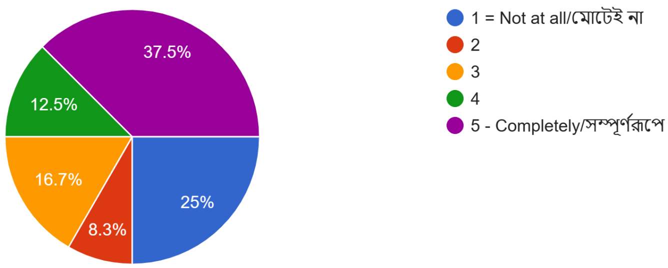


Figure 3.57: Collected prosocial all average data ratio

21. Is your behavior only influenced by the immediate outcomes of your actions?/ আপনার আচরণ কি শুধুমাত্র আপনার কর্মের তাৎক্ষণিক ফলাফল দ্বারা প্রভাবিত হয়?(CFC_I_avg)
48 responses

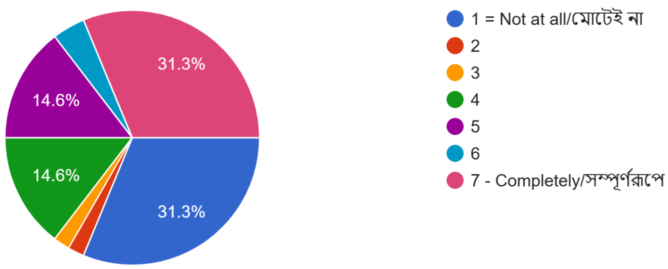


Figure 3.58: Collected CFC_I data ratio

20. Are you willing to sacrifice your immediate happiness or well-being to achieve future outcomes?/ আপনি কি ভবিষ্যতের ফলাফল অর্জনের জন... বা মঙ্গলকে উৎসর্গ করতে ইচ্ছুক? (CFC_F_avg)
48 responses

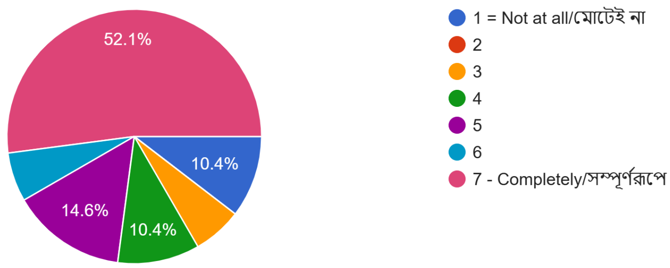


Figure 3.59: Collected CFC_F data ratio

22. Do you think the climate crisis has gone too far?/ আপনি কি মনে করেন যে জলবায়ু সংকট অনেক দূরে চলে গেছে? (helpless_CLIM_avg)
48 responses

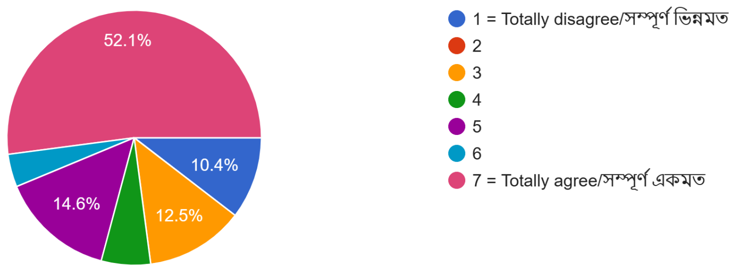


Figure 3.60: Collected helplessness about climate data ratio

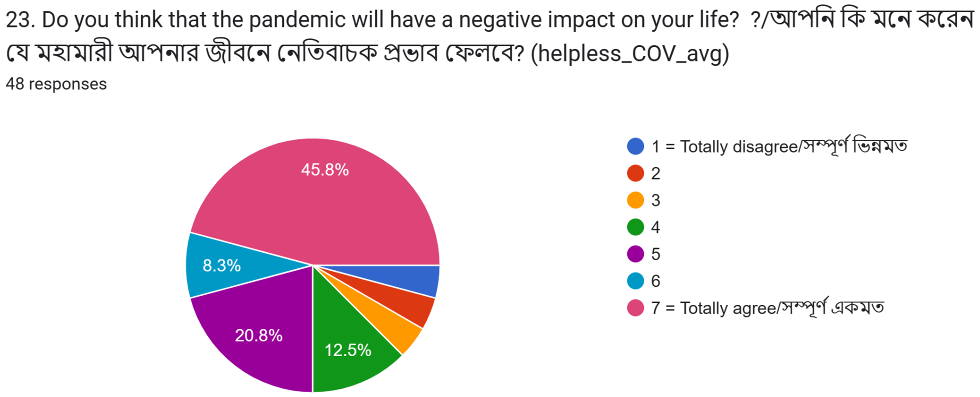


Figure 3.61: Collected helplessness about Covid-19 data ratio

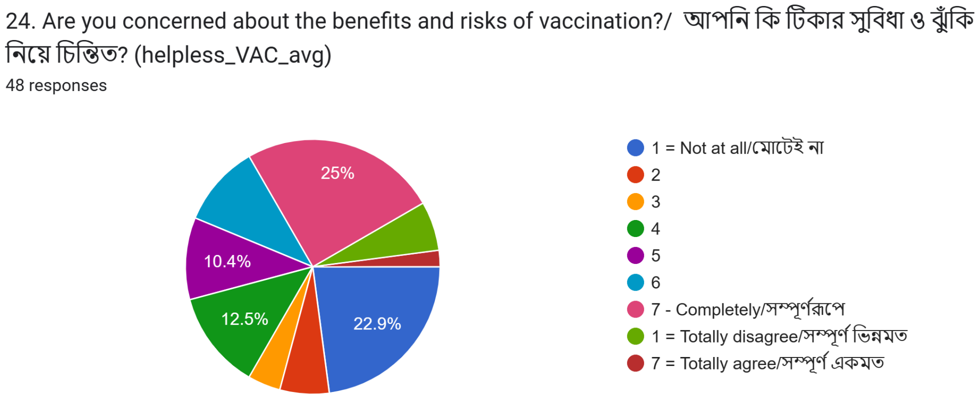


Figure 3.62: Collected helplessness about vaccination data ratio

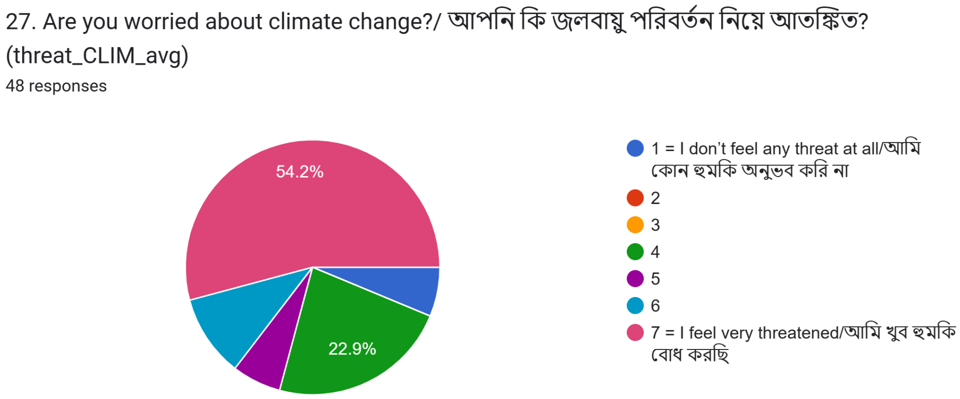


Figure 3.63: Collected threat about climate data ratio

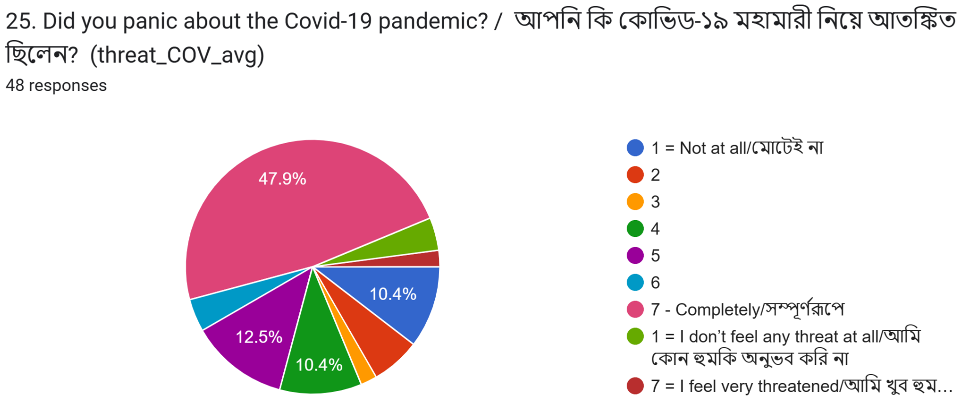


Figure 3.64: Collected threat about Covid-19 data ratio

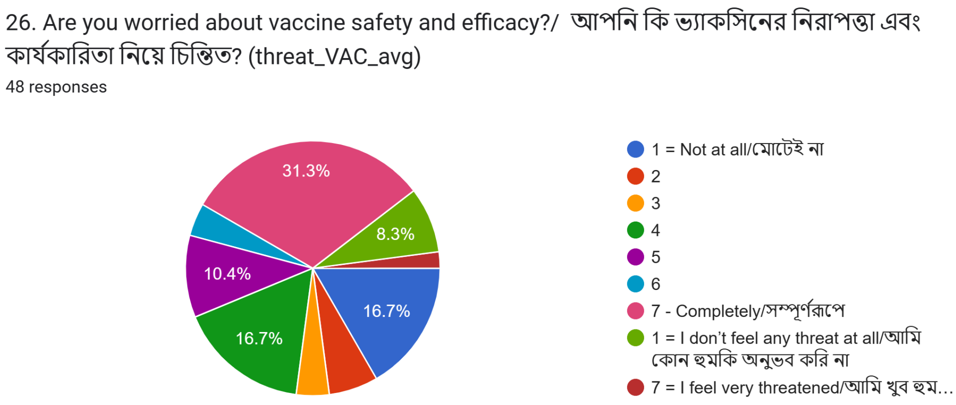


Figure 3.65: Collected threat about vaccination data ratio

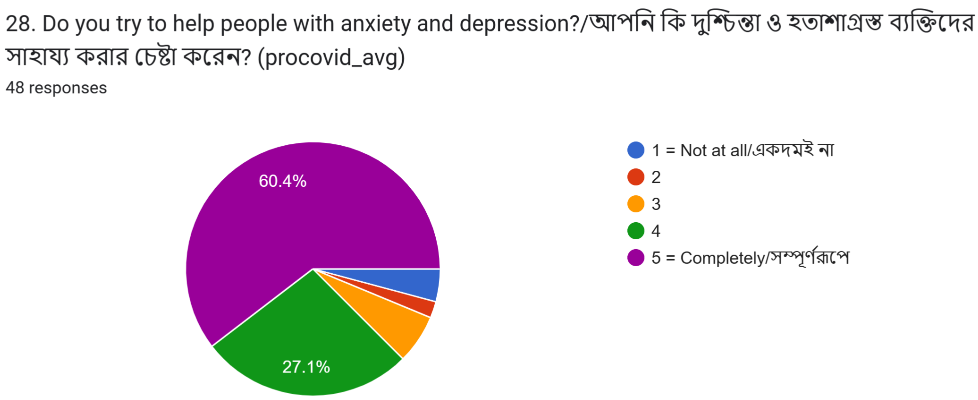


Figure 3.66: Collected procovid data ratio

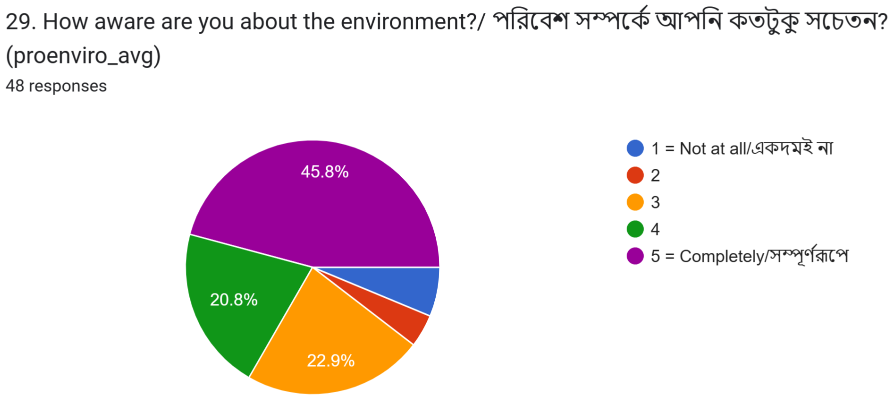


Figure 3.67: Collected pro environment data ratio

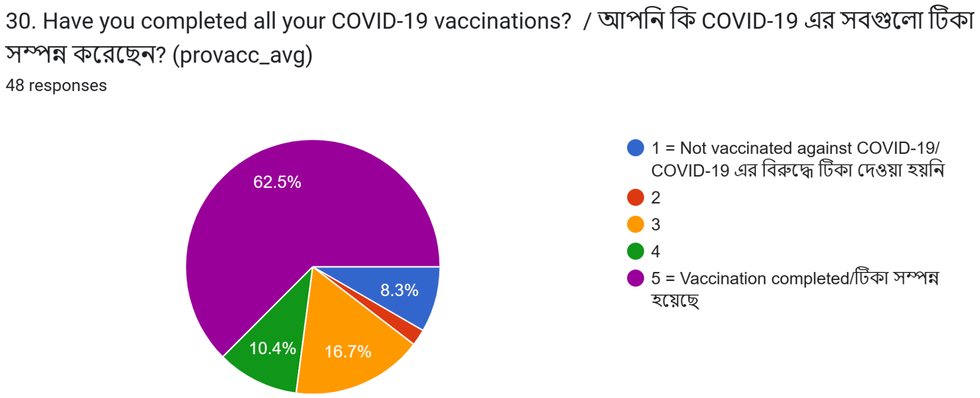


Figure 3.68: Collected pro vaccine data ratio

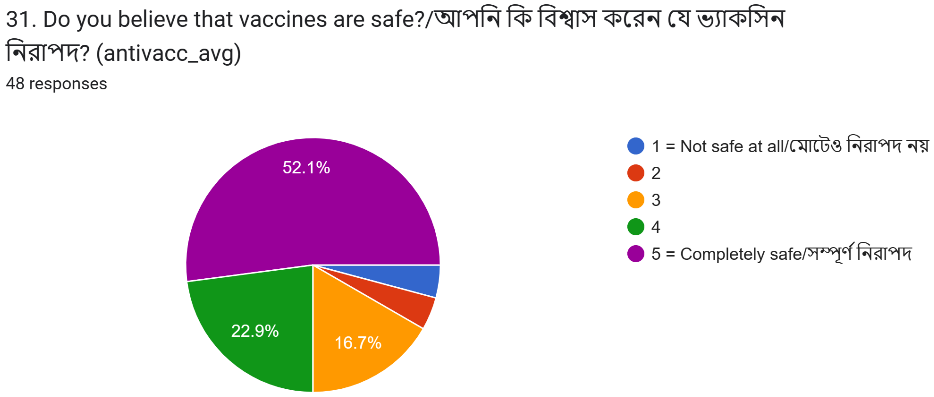


Figure 3.69: Collected anti vaccination variable data ratio

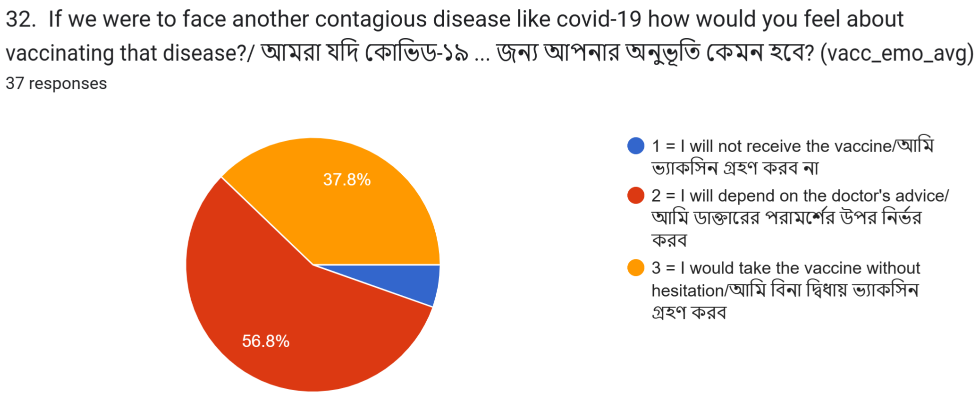


Figure 3.70: Collected target variable data ratio

Here is the validation data set box plot representation samples of some columns which represents about the data diversion and IQR according to data.

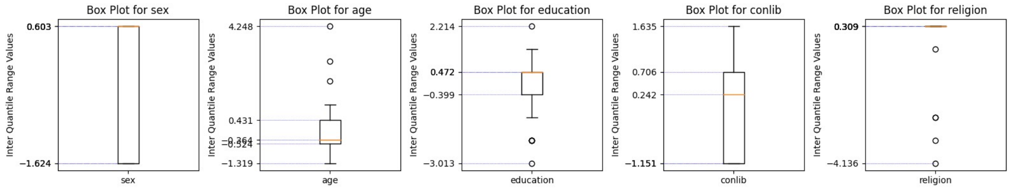


Figure 3.71: Collected target variable data ratio

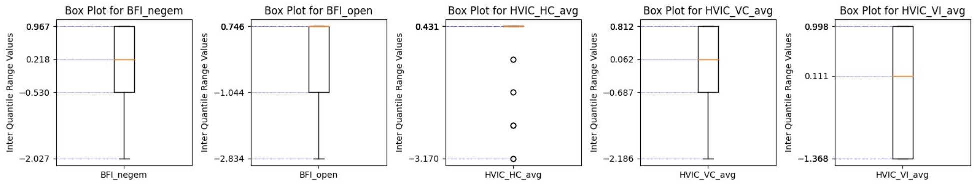
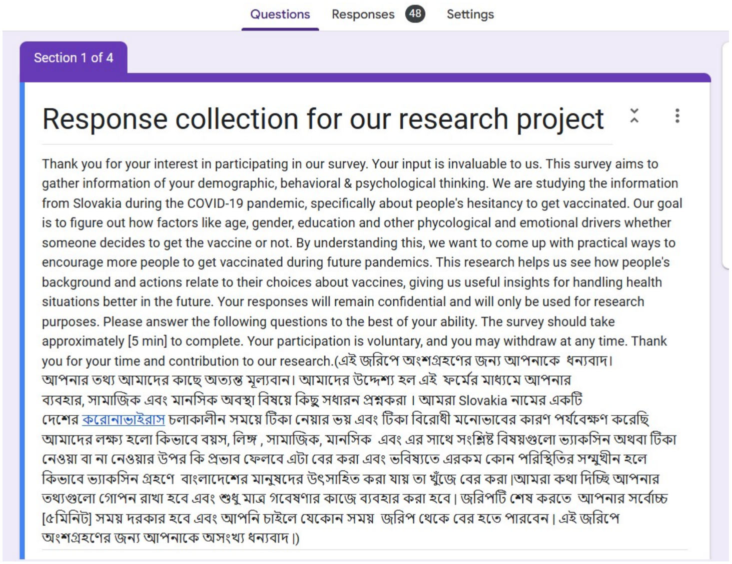


Figure 3.72: Collected target variable data ratio

3.1.3 Dataset Source

Our dataset was compiled by administering pre-defined questions through a Google Form. Ensuring a robust validation set, we secured 48 responses from participants on 32 columns. To obtain a diverse range of insights, we distributed the survey among our university peers and across various social media platforms. This approach facilitated the collection of a comprehensive set of answers for our dataset. The form questions sample are given below. Samples:



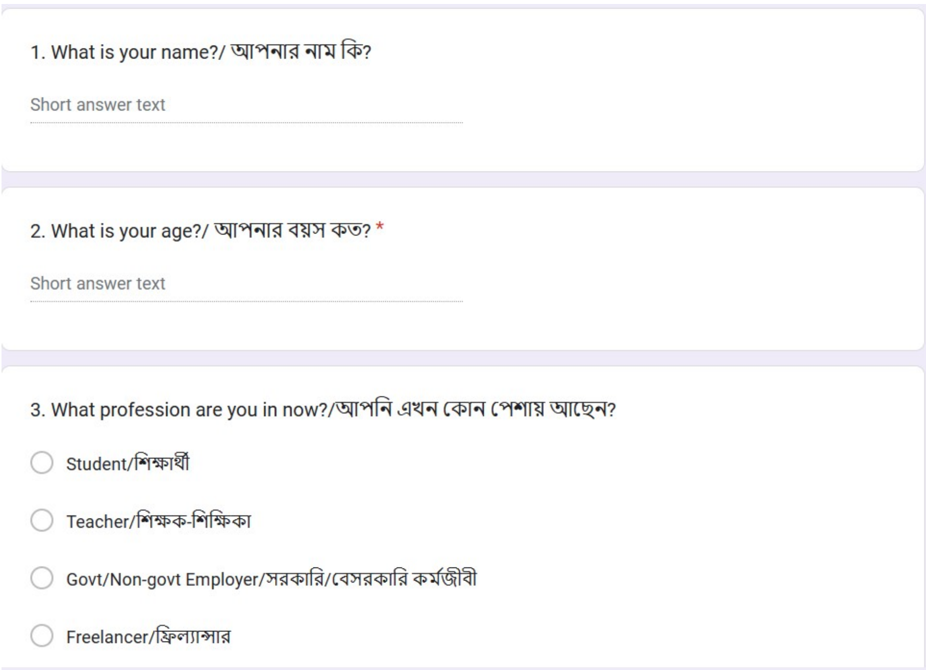
Questions Responses 48 Settings

Section 1 of 4

Response collection for our research project

Thank you for your interest in participating in our survey. Your input is invaluable to us. This survey aims to gather information of your demographic, behavioral & psychological thinking. We are studying the information from Slovakia during the COVID-19 pandemic, specifically about people's hesitancy to get vaccinated. Our goal is to figure out how factors like age, gender, education and other phycological and emotional drivers whether someone decides to get the vaccine or not. By understanding this, we want to come up with practical ways to encourage more people to get vaccinated during future pandemics. This research helps us see how people's background and actions relate to their choices about vaccines, giving us useful insights for handling health situations better in the future. Your responses will remain confidential and will only be used for research purposes. Please answer the following questions to the best of your ability. The survey should take approximately [5 min] to complete. Your participation is voluntary, and you may withdraw at any time. Thank you for your time and contribution to our research.(এই জরিপে অংশগ্রহণের জন্য আপনাকে ধন্যবাদ। আপনার তথ্য আমাদের কাছে অত্যন্ত মূল্যবান। আমাদের উদ্দেশ্য হল এই ফর্মের মাধ্যমে আপনার ব্যবহার, সামাজিক এবং মানসিক অবস্থা বিষয়ে কিছু সধারন প্রশ্নকরা । আমরা Slovakia নামের একটি দেশের করোনাভাইরাস চলাকালীন সময়ে টিকা নেয়ার ভয় এবং টিকা বিরোধী মনোভাবের কারণ পর্যবেক্ষণ করেছি আমাদের লক্ষ্য হলো কিভাবে বয়স, লিঙ্গ , সামাজিক, মানসিক এবং এর সাথে সংশ্লিষ্ট বিষয়গুলো ভ্যাকসিন অথবা টিকা নেওয়া বা না নেওয়ার উপর কি প্রভাব ফেলবে এটা বের করা এবং ভবিষ্যতে এরকম কোন পরিস্থিতির সম্মুখীন হলে কিভাবে ভ্যাকসিন গ্রহণে বাংলাদেশের মানুষদের উৎসাহিত করা যায় তা খুঁজে বের করা ।আমরা কথা দিচ্ছি আপনার তথ্যগুলো গোপন রাখা হবে এবং শুধু মাত্র গবেষণার কাজে ব্যবহার করা হবে। জরিপটি শেষ করতে আপনার সর্বোচ্চ [৫মিনিট] সময় দরকার হবে এবং আপনি চাইলে যেকোন সময় জরিপ থেকে বের হতে পারবেন। এই জরিপে অংশগ্রহণের জন্য আপনাকে অসংখ্য ধন্যবাদ।)

Figure 3.73: Google form sample question(Sample 1)



1. What is your name?/ আপনার নাম কি?

Short answer text

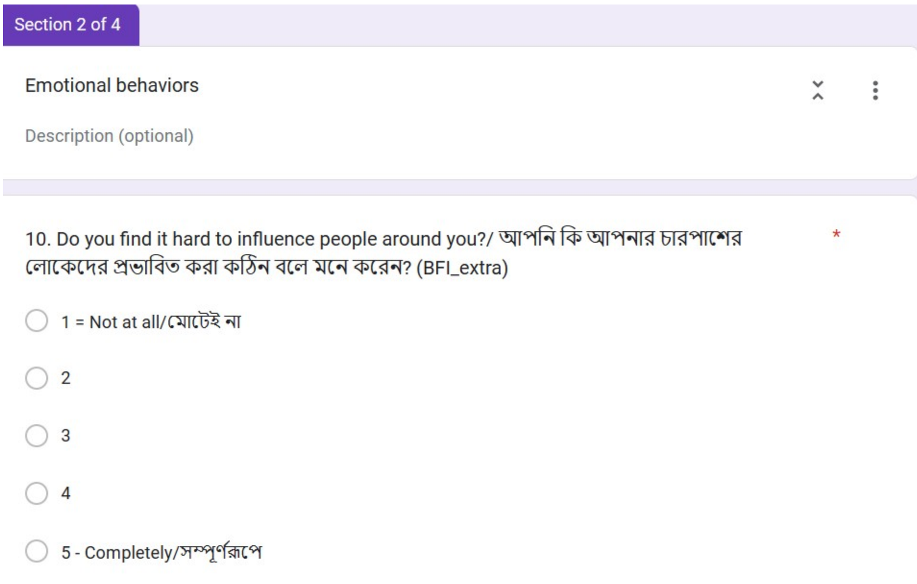
2. What is your age?/ আপনার বয়স কত? *

Short answer text

3. What profession are you in now?/আপনি এখন কোন পেশায় আছেন?

- Student/শিক্ষার্থী
- Teacher/শিক্ষক-শিক্ষিকা
- Govt/Non-govt Employer/সরকারি/বেসরকারি কর্মজীবী
- Freelancer/ফ্রিল্যান্সার

Figure 3.74: Google form sample question(Sample 2)



Section 2 of 4

Emotional behaviors

Description (optional)

10. Do you find it hard to influence people around you?/ আপনি কি আপনার চারপাশের লোকেদের প্রভাবিত করা কঠিন বলে মনে করেন? (BFI_extra) *

1 = Not at all/মোটেই না

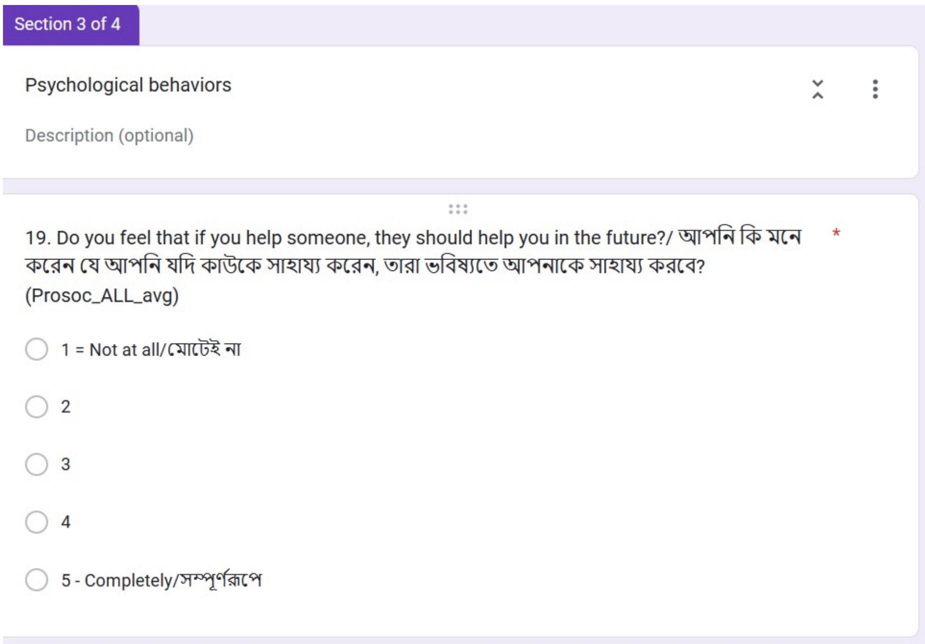
2

3

4

5 - Completely/সম্পূর্ণরূপে

Figure 3.75: Google form sample question(Sample 3)



Section 3 of 4

Psychological behaviors

Description (optional)

19. Do you feel that if you help someone, they should help you in the future?/ আপনি কি মনে করেন যে আপনি যদি কাউকে সাহায্য করেন, তারা ভবিষ্যতে আপনাকে সাহায্য করবে? (Prosoc_ALL_avg) *

1 = Not at all/মোটেই না

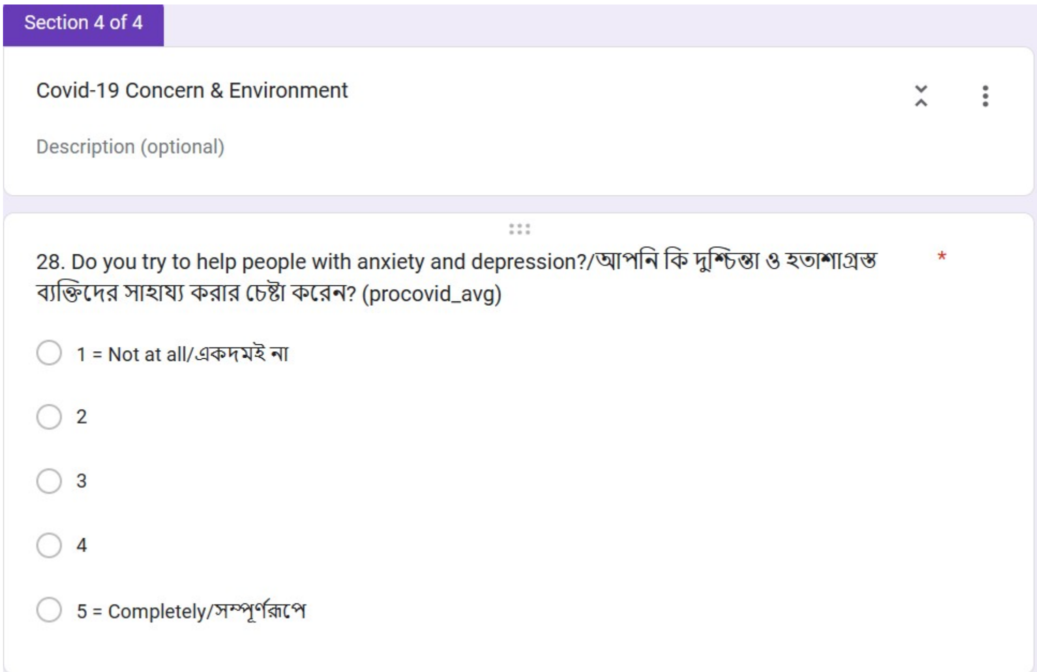
2

3

4

5 - Completely/সম্পূর্ণরূপে

Figure 3.76: Google form sample question(Sample 4)



Section 4 of 4

Covid-19 Concern & Environment

Description (optional)

28. Do you try to help people with anxiety and depression?/আপনি কি দুশ্চিন্তা ও হতাশাগ্রস্ত ব্যক্তিদের সাহায্য করার চেষ্টা করেন? (procovid_avg) *

1 = Not at all/একদমই না

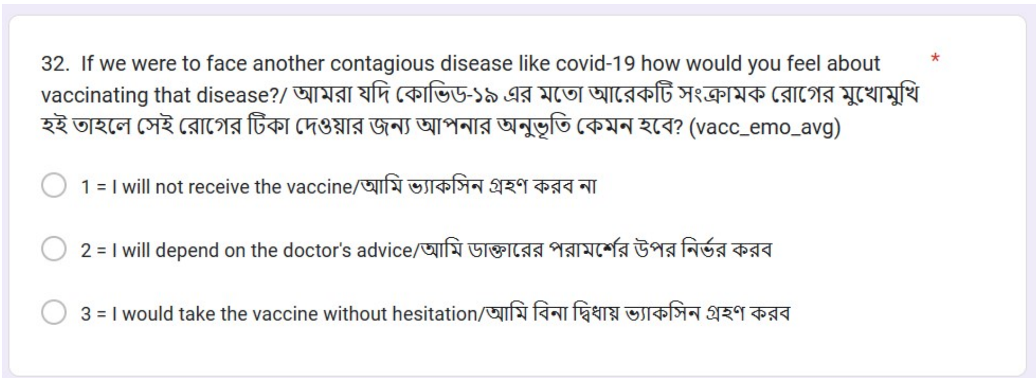
2

3

4

5 = Completely/সম্পূর্ণরূপে

Figure 3.77: Google form sample question(Sample 5)



32. If we were to face another contagious disease like covid-19 how would you feel about vaccinating that disease?/ আমরা যদি কোভিড-১৯ এর মতো আরেকটি সংক্রামক রোগের মুখোমুখি হই তাহলে সেই রোগের টিকা দেওয়ার জন্য আপনার অনুভূতি কেমন হবে? (vacc_emo_avg) *

1 = I will not receive the vaccine/আমি ভ্যাকসিন গ্রহণ করব না

2 = I will depend on the doctor's advice/আমি ডাক্তারের পরামর্শের উপর নির্ভর করব

3 = I would take the vaccine without hesitation/আমি বিনা দ্বিধায় ভ্যাকসিন গ্রহণ করব

Figure 3.78: Google form sample question(For target variable)

3.1.4 Classes

In the dataset from Slovakia, there exists a target column named `vacc_emo_avg`. This column provides an overall indication of a participant's emotional pattern or behavior, derived from the average values of different metrics in the dataset. The average values range from 1 to 7, which presents a wide range for classification purposes and may lead to scattered results, posing challenges in sample classification.To address this issue, we initially applied a ceiling and floor value technique to transform the average values into integer values. Subsequently, we categorized these integer values into three classes:

1. Values 1 and 2 were grouped together as Class 1.

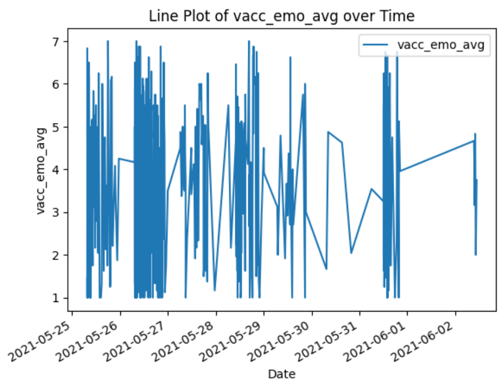
2. Values 3, 4, and 5 were grouped together as Class 2.

3. Values 6 and 7 were grouped together as Class 3.

Based on the nature of the dataset and the scales used, we interpreted Class 1 as indicating low agreement to take the vaccine (labeled as 1), Class 2 as indicating hesitation to take the vaccine (labeled as 2), and Class 3 as indicating agreement to take the vaccine without hesitation (labeled as 3).

vacc_emo_avg
4
3.458333333
5.458333333
4.041666667
3.375
6.375
3
3
5.291666667

(a) Target variable before preprocessing



(b) Target variable line plot

Figure 3.79: Final Figure

```
[ ]  # To check the value of our target column value
     new_df['vacc_emo_avg']

0      2
1      2
2      3
3      2
4      2
      ..
495    3
496    2
497    2
498    2
499    3
Name: vacc_emo_avg, Length: 500, dtype: category
Categories (3, int64): [1 < 2 < 3]
```

Figure 3.80: Target variable after binning

From the count plot we can see that most of the target value is in second class (2) and rest of the target class count ratio is similar.

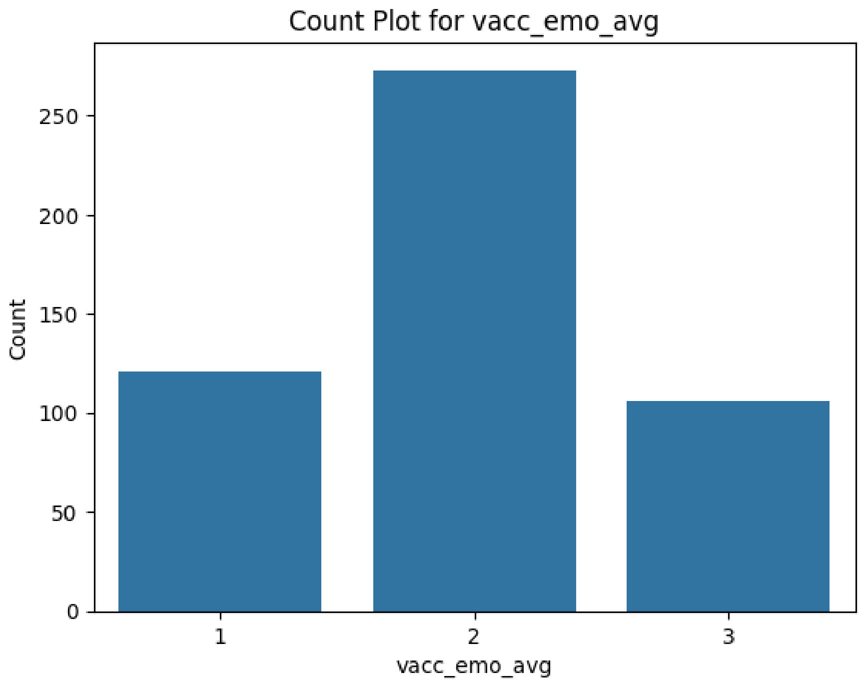


Figure 3.81: Target variable after preprocessing

This modification of the dependent variable allowed us to incorporate three classes into the target column values in our validation data collection.

3.2 Dataset preprocessing

Data preprocessing is an essential preliminary step in the data analysis process, involving the transformation of raw data into a structured format suitable for analysis. It encompasses tasks such as cleaning, transforming, and organizing data to enhance its quality and relevance for subsequent analysis and modeling. By addressing issues such as noise, inconsistency, and incompleteness in the data, preprocessing lays the groundwork for extracting meaningful insights and making accurate predictions. Ultimately, the effectiveness of data analysis and modeling hinges on the quality of the preprocessed data, highlighting the critical importance of this initial phase in any data-driven project.

3.2.1 Actual dataset preprocessing

In the actual dataset has 500 rows and 195 columns. Among them 18 columns contains the average estimated values of the other columns. That's why we took the average columns to reduce data duplication and took demographic and descriptive 12 fields with index which formed a new dataset/data-frame. At the same time we remove the index column from the new data frame. After that we noticed there were 5 columns that hold categorical value. To made a universal data type we use lambda function to make numerical value. Here is the value we used to convert the categorical value into numerical value:

	Sex	
Values	**Man**	**Woman**
Converted values	1	0

Table 3.1: Gander column values

	Marital status			
Values	**Married**	**Single**	**Divorced**	**Other**
Converted values	0	1	2	3

Table 3.2: Marital status column with values

Education	
Dataset value	**Values applied for preprocessing**
High school	0
University(Mrg,Ing)	1
High school without graduation	2
University (BSc)	3
University(PhD)	4
Primary school	5
Others	6

Table 3.3: Education column values

conlib	
Dataset value	**Values applied for preprocessing**
1 - very conservative	0
2	1
3	2
4	3
5	4
6	5
7 - very liberal	6

Table 3.4: conlib column

Religion	
Dataset value	**Values applied for preprocessing**
1 - religion is not important to me at all	0
2	1
3	2
4	3
5	4
6	5
7 - religion is very important to me	6

Table 3.5: Region column values

Table 3.6: Description of 30 columns

Column Name	Description
'sex'	Gender of the individual
'age'	Age of the individual
'education'	Level of education
'conlib'	Conservative & Liberal
'religion'	Religious affiliation
'marital_status'	Marital status
'household'	Household composition
'BFI_extra'	Big Five Inventory - Extraversion
'BFI_agree'	Big Five Inventory - Agreeableness
'BFI_consc'	Big Five Inventory - Conscientiousness
'BFI_negem'	Big Five Inventory - Negative Emotionality
'BFI_open'	Big Five Inventory - Openness
'HVIC_HC_avg'	Household Value Inventory - Health Concern Average
'HVIC_VC_avg'	Household Value Inventory - Vaccination Concern Average
'HVIC_VI_avg'	Household Value Inventory - Vaccination Importance Average
'HVIC_HI_avg'	Household Value Inventory - Health Importance Average
'prosoc_ALL_avg'	Prosocial Behavior Average
'CFC_F_avg'	Cognitive Reflection Test - Faith Average
'CFC_I_avg'	Cognitive Reflection Test - Intuition Average
'helpless_CLIM_avg'	Feelings of Helplessness - Climate Average
'helpless_COV_avg'	Feelings of Helplessness - COVID Average
'helpless_VAC_avg'	Feelings of Helplessness - Vaccination Average
'threat_COV_avg'	Perceived Threat - COVID Average
'threat_VAC_avg'	Perceived Threat - Vaccination Average
'threat_CLIM_avg'	Perceived Threat - Climate Average
'procovid_avg'	Pro-COVID Behavior Average
'proenviro_avg'	Pro-Environmental Behavior Average
'provacc_avg'	Pro-Vaccination Behavior Average
'antivacc_avg'	Anti-Vaccination Behavior Average
'vacc_emo_avg'	Emotional Response to Vaccination Average

Then we applied min-max normalization on all columns without target variable.

```
1 #Check the dataset after normalization
2 new_df.head(10)
```

	sex	age	education	conlib	religion	marital_status	household	BFI_extra	BFI_agree	BFI_consc	...	helpless_COV_avg	helpless_VAC_avg	threat_COV_avg	threat_VAC_avg	threat_
0	0.0	0.2353	0.0000	0.5000	0.5000	0.0000	0.500	0.5217	0.8182	0.6667	...	0.3750	0.5000	0.5000	0.5000	
1	1.0	0.6765	0.1667	0.3333	0.8333	0.0000	0.125	0.7391	0.8636	0.8095	...	0.2917	0.2083	0.5000	0.6111	
2	0.0	0.2941	0.1667	0.3333	0.5000	0.0000	0.375	0.5652	0.8636	0.7619	...	0.6250	0.7083	0.6667	0.7778	
3	0.0	0.1471	0.1667	0.5000	0.3333	0.3333	0.125	0.4783	0.7727	0.7143	...	0.5833	0.6250	0.7222	0.3889	
4	0.0	0.1618	0.1667	0.6667	0.1667	0.3333	0.250	0.4783	0.6818	0.7143	...	0.3750	0.2917	0.6667	0.5000	
5	0.0	0.1471	0.3333	0.5000	0.6667	0.3333	0.375	0.7826	0.7273	0.9524	...	0.7500	0.7917	0.7778	1.0000	
6	0.0	0.2353	0.1667	0.6667	0.0000	0.3333	0.000	0.2609	0.3636	0.3810	...	0.1667	0.1667	0.5000	0.5000	
7	1.0	0.3235	0.0000	0.5000	0.6667	0.0000	0.250	0.6957	0.7273	0.7143	...	0.3333	0.3333	0.6111	0.3333	
8	1.0	0.4559	0.3333	0.5000	0.8333	0.0000	0.125	0.4348	0.5455	0.4762	...	0.6250	0.7083	0.7222	0.7222	
9	1.0	0.9265	0.1667	0.5000	0.8333	1.0000	0.125	0.5217	0.4091	0.4286	...	0.1250	0.0000	1.0000	0.0000	

10 rows × 30 columns

Figure 3.82: Normalized values

Finally, we applied the feature selection technique and we took the top 21 columns with the target column according to the next operations.

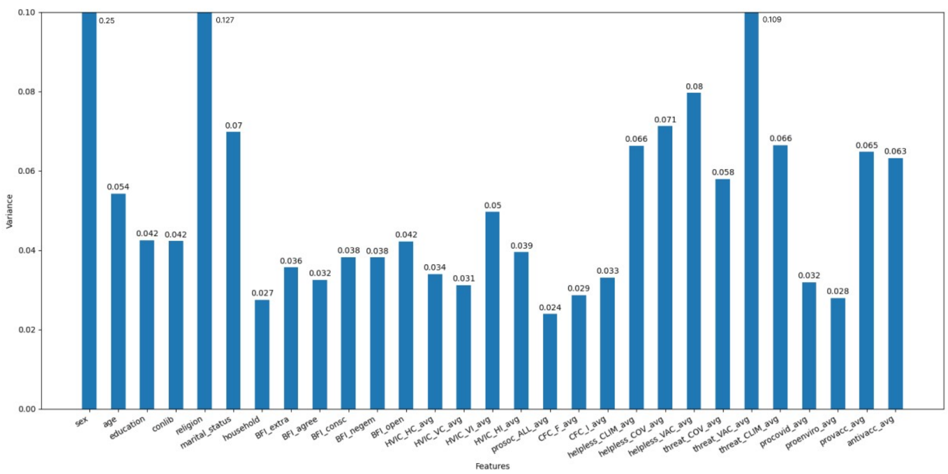


Figure 3.83: Feature importance values

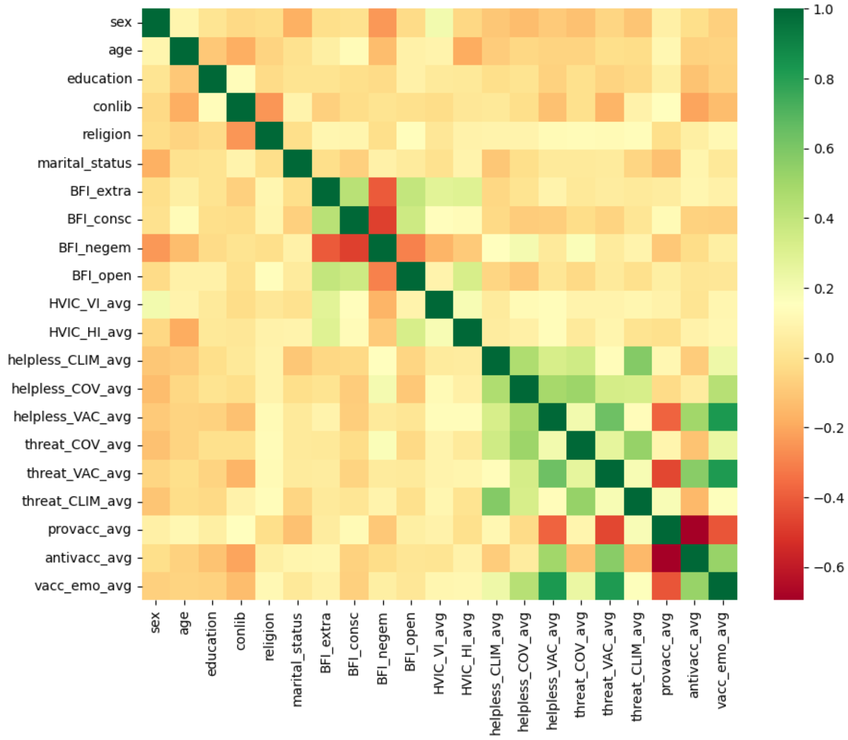


Figure 3.84: Top 20 selected columns heat map with target variable

3.2.2 Validation dataset prepossessing

To create our validation dataset first we give an option to our respondents that they provide us the accurate information. Then we give everyone 34 individual questions after taking their answer we create a .xlsx file. Now we start our prepossessing, first of all, we drop 4 columns that are not related to our actual dataset this is to warm them up to continue the questionnaire. Then rename the columns as they are in the actual dataset. Then re index the columns as in the actual dataset. First, we encode our sex, education, and marital status columns using the lambda function. And for the rest of the column, we use the value mapping function to replace the value. As all the values are categorical so we use as type (int) function to change the type of value into integer. Then we standardize our data and make a boxplot to check the outliers in the dataset. After then we normalize our dataset using the MinMaxNormalization technique. Then create a CSV file to validate our actual dataset.

4 Experimental Setup

In the ”Methods and Techniques Applied” chapter of this thesis, we delve into the systematic framework employed to investigate the research objectives outlined in the preceding chapters. This pivotal section serves as a comprehensive guide, elucidating the methodologies, tools, and analytical techniques meticulously selected to address the research questions posed. By meticulously detailing the approaches utilized for data collection, analysis, and interpretation, this chapter provides readers with a robust understanding of the methodological underpinnings that underlie the study's findings and conclusions. Through a transparent exposition of the research methodology, we aim to enhance the credibility, reliability, and validity of our research outcomes, ultimately contributing to the advancement of knowledge in our field of inquiry.

4.1 Models, Techniques & Algorithms

In the model evaluation, we use two techniques one is k-fold cross-validation and another is the holdout method. In cross-validation, we use five-fold, ten-fold, and fifteen-fold, and in the holdout method we split our dataset in different ratios like 70:30, 80:20, and 85:15 for these different k-fold and splitting datasets we use some machine learning algorithms to see their accuracy. These algorithms are -

1. Min-Max Normalization

2. Round function

3. Linear SVM

4. Radial SVM

5. Logistic Regression

6. KNeighbour Classifier
7. Decisiontree classifier
8. Gradiant Boosting Classifier
9. RandomForest Classifier
10. Naive Bayes

4.1.1 Min-Max Normalization

Min-max normalization [19] is a popular technique used in data preprocessing to scale numerical features to a specific range, typically between 0 and 1. It is particularly useful when the original data has varying scales and the algorithm being used for analysis or modeling is sensitive to scale. Equation is:

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

where x_{norm} is the normalized value. key points are:

- Min-max normalization preserves the shape of the original distribution but scales it to fit within a specified range.
- It is a linear transformation that maps the minimum value to 0 and the maximum value to 1.
- Min-max normalization is sensitive to outliers, as it stretches the range to fit the extremes of the data.
- This technique is widely used in various machine learning algorithms, such as neural networks, support vector machines, and k-nearest neighbors.

4.1.2 Round function

The round() function is like a handy tool in programming that helps tidy up numbers. It's used to make numbers simpler by rounding them to a certain level of accuracy. This makes it easier to work with data or present it neatly. We used round function in data preprocessing section.

4.1.3 Linear SVM:

Linear Support Vector Machine (SVM) [20] is a supervised machine learning algorithm used for classification and regression tasks. Its primary objective is to find a hyperplane that optimally separates data points into different classes. By maximizing the margin between classes, Linear SVM achieves robustness against outliers. Although it assumes linear separability, it can be extended to handle non-linear data using kernel functions. Linear SVM finds applications in text classification, image recognition, and bioinformatics. Its effectiveness lies in handling high-dimensional data and large datasets while avoiding over fitting. Libraries like sci-kit-learn provide convenient implementations of Linear SVM.

We use this model on our actual dataset to see the accuracy. When we use this model we set the kernel as linear for k-fold cross-validation but when we use it for the holdout method we take the parameter kernel as linear, gamma=15, C=7, probability=True. Using this we get a good accuracy score for our actual dataset. Then we apply hyper-parameter tuning on this model to improve its accuracy. That time we use both linear and rbf as kernel, set gamma value 1 to 20, and c value 1, 10, 100, 1000. Then we get the best parameter for is model and improve a little accuracy.

4.1.4 Radial SVM:

The Radial Basis Function (RBF) Support Vector Machine (SVM) [21] is a supervised machine learning algorithm primarily used for classification tasks. Unlike linear SVMs, which operate in finite-dimensional spaces, the RBF SVM works in an infinite-dimensional space by projecting data into a higher-dimensional feature space. The "radial" aspect refers to the circular symmetry around a center point, where the function's value depends solely on the radial distance from the center, disregarding direction. The RBF kernel computes similarity based on this radial distance. The decision boundary, a projection of the hyperplane onto input space, can be curved or even discontinuous, even though the underlying hyperplane is always linear and continuous. RBF SVM finds applications in scenarios where data is not linearly separable.

4.1.5 Logistic Regression:

Logistic Regression is a supervised machine learning algorithm [22] used for classification tasks. Unlike linear regression, which predicts continuous numeric values, logistic regression predicts the probability that an instance belongs to a given class (usually binary: yes/no, true/false). It models the relationship between independent variables (features) and the probability of a binary outcome using a logit function. The output of logistic regression lies between 0 and 1, making it suitable for estimating probabilities.

In the logistic regression model, we take the max iteration 1000 in k-fold cross-validation, on the other hand in the ensemble method we take parameter c=0.1. Its accuracy is also good for the actual dataset.

4.1.6 K-Neighbour Classifier:

The k-Nearest Neighbors (k-NN) algorithm is a supervised machine learning classifier that operates based on the proximity of data points. Given a new data point, k-NN identifies the k nearest neighbors from the training dataset using a chosen distance metric (usually Euclidean distance). For classification, it assigns the majority class among these neighbors to the new point. For regression, it computes the average (or weighted average) of the target values of the neighbors. k-NN's decision boundary is flexible, adapting to the local distribution of data.

We use the K-Neighbour classifier in our actual dataset both in the k-fold cross-validation and holdout method. Here we take 3-neighbours as its parameter to see how it gives the accuracy. It works well in our actual dataset and selected feature dataset.

4.1.7 Decision tree classifier:

The Decision Tree Classifier is a powerful supervised learning algorithm used for both classification and regression tasks. It builds a flowchart-like tree structure where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. Decision trees are easy to interpret and handle both categorical and numerical features. Decision tree classifiers use both k-fold and holdout methods to check the accuracy. We take max depth=7 and random state = 42 as its parameter and its accuracy score is good.

4.1.8 Gradient Boosting Classifier:

The Gradient Boosting Classifier is a powerful ensemble learning technique used for classification tasks. It combines multiple weak learners (usually decision trees) to create a strong predictive model. Starting with an initial model, it improves iterative by adding new models that correct errors made by the previous ones. The final prediction is a weighted sum of all individual models. Gradient Boosting excels in high accuracy, handles missing data well, and is robust against over-fitting. We use Gradient gradient-boosting classifier for both the k-fold and holdout methods. In this model, we take max depth = 7 and random state 42 as the parameter. This model gives us the best result for our dataset.

4.1.9 Random Forest Classifier:

The Random Forest Classifier is a powerful ensemble learning technique used for classification tasks. It combines multiple decision tree classifiers on various sub-samples of the dataset and uses averaging to improve predictive accuracy while controlling over-fitting. Random forests create an ensemble of decision trees, and the final prediction is determined by voting among these trees. It's widely adopted due to its ease of use, flexibility, and ability to handle both classification and regression problems. Random forest classifiers are also used in both k-fold and holdout methods. We also do hyper parameter tuning on this. For this algorithm we take parameter as `n_estimators=100,random_state=42` in k-fold and `n_estimators=100,random_state=0` in holdout. When we apply hyper parameter tuning this time we use `n_estimators=10,20,30, up-to 200`, max feature = 1 to 5 and cv = 5. After this tune, we get little improvement in our accuracy,

4.1.10 Naive Bayes:

Naive Bayes is a probabilistic machine learning algorithm used for classification tasks. It's based on Bayes' theorem and assumes that features are conditionally independent given the class label.

4.1.11 Ensemble method:

In the ensemble method, we use three different algorithms. These are linear SVC, radial SVC, and Logistic Regression. We use some parameters like voting = soft, and weight = 1,2,3 for this ensemble model. The accuracy of this ensemble model is good.

4.2 Evaluation metrics

Precision, recall, F1-score, and accuracy metrics were utilized to evaluate the performance of our models. The abbreviations TP, TS, FP, TN, and FN stand for true positive, total samples, false positive, true negative, and false negative, respectively Here is how we calculated these metrics:

4.2.1 Accuracy

One parameter for assessing classification models is accuracy. The percentage of predictions that our model correctly predicted is known as accuracy. Formally, accuracy has the following definition

$$\text{Accuracy} = \frac{TN+TP}{TP+FP+TN+FN} \tag{4.1}$$

4.2.2 Precision

Precision [**?**] is determined by dividing the total number of positively identified samples by the proportion of accurately classified positive samples.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4.2}$$

4.2.3 Recall

The model's capacity to identify positive samples is gauged by the recall [23]. More positive samples are found when a recall is higher.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4.3}$$

4.2.4 F1-score

The F1 score is a commonly used metric for evaluating the performance of classification models. It is calculated as the harmonic mean of precision and recall and provides a balance between these two metrics. The formula for computing the F1 score is:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

where precision is the ratio of true positive predictions to the total number of positive predictions, and recall is the ratio of true positive predictions to the total number of actual positive instances.

The F1 score is widely used in machine learning research and practice due to its ability to capture both precision and recall simultaneously.

One of the earliest references to the F1 score can be found in the work of Van Rijsbergen (1979) [24].

4.2.5 Root Mean Square Error (RMSE)

RMSE stands for Root Mean Square Error [25]. It's a common metric used to evaluate the performance of a regression model in machine learning.

RMSE measures the average deviation of the predicted values from the actual values in a dataset. It's calculated by taking the square root of the average of the squared differences between the predicted and actual values. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where:

- n is the number of data points.
- y_i is the actual value of the target variable for the ith data point.
- $\hat{y}_i$ is the predicted value of the target variable for the ith data point.

RMSE gives us a single number to quantify the overall performance of the model. Lower values of RMSE indicate better performance, as they represent smaller deviations between predicted and actual values.

It's important to note that RMSE is sensitive to outliers since it squares the differences between predicted and actual values. Therefore, it's often used in conjunction with other evaluation metrics to get a comprehensive understanding of model performance.

4.3 Chapter Summary

The ”Methods Applied & Techniques” chapter of this thesis outlines the systematic framework employed to address the research objectives. This section serves as a comprehensive guide, elucidating the methodologies, tools, and analytical techniques meticulously chosen for investigating the research questions. The chapter discusses the categorization of the `"vacc_emo_avg"` column into three classes, reflecting different levels of agreement towards vaccination. The subsequent section details various machine learning algorithms employed for model evaluation, such as Linear SVM, Radial SVM, Logistic Regression, KNeighbourClassifier, Decision Tree Classifier, Gradient Boosting Classifier, Random Forest Classifier, and Naive Bayes. Each algorithm is described along with their parameters and applications. The chapter concludes with an ensemble method utilizing Linear SVM, Radial SVM, and Logistic Regression. The application of these methods aims to enhance the credibility, reliability, and validity of the research outcomes, contributing to advancements in the field.

5 Results & Analysis

Here We will discuss the results we get after all the models are applied By comparing their Test loss, Test Accuracy, Test categorical Accuracy, Test F1 score, Test Recall, Test precision. Which final results are shown in the different tables and each epoch is shown by graphs.

5.1 K-fold cross validation

K-fold cross-validation is a technique used to evaluate the performance of a machine learning model. It involves splitting the dataset into k equal-sized subsets (or ”folds”). The model is trained on k-1 of these folds and tested on the remaining fold. This process is repeated k times, with each fold used once as the test set. Finally, the performance metrics (e.g., accuracy, precision, recall) from each iteration are averaged to obtain a single estimate of the model's performance. K-fold cross-validation helps in obtaining a more reliable estimate of model performance and reduces the variance in the evaluation process.

5.1.1 K-fold cross validation(30 features):

Model name	For 5-fold	For 10-fold	For 15-fold
Gradient Boosting Classifier	**0.98**	**0.98**	**0.97**
Decision Tree Classifier	0.94	0.96	0.96
Random Forest Classifier	0.93	0.92	0.93
Linear Svm	0.92	0.91	0.92
Logistic Regression	0.91	0.92	0.91
Gaussian Naive Bayes	0.90	0.89	0.89
Radial Svm	0.89	0.90	0.89
KNeighbors Classifier	0.78	0.77	0.77

Table 5.1: Performance of different classifiers using K-fold cross-validation with 30 features

When we first use k-fold cross-validation we use 30 features and we get a good result for the gradient boosting classifier. the accuracy for 5-fold is 0.98, for 10-fold is 0.98 and the 15-fold is 0.97.

5.1.2 K-fold cross-validation(21 features):

Model name	For 5-fold	For 10-fold	For 15-fold
Gradient Boosting Classifier	**0.98**	**0.98**	**0.98**
Decision Tree Classifier	0.95	0.95	0.95
Random Forest Classifier	0.93	0.93	0.93
Linear Svm	0.93	0.93	0.93
Logistic Regression	0.91	0.91	0.91
Gaussian Naive Bayes	0.94	0.93	0.93
Radial Svm	0.90	0.91	0.90
KNeighbors Classifier	0.77	0.79	0.77

Table 5.2: Performance of different classifiers using K-fold cross-validation with 21 features

In our initial use of k-fold cross-validation, we employed 21 features and achieved impressive results with the gradient boosting classifier. Specifically, the accuracy remained consistently high across different fold values: 0.98.

5.2 Data splitting

Data splitting involves dividing a dataset into training, validation, and testing sets to train, fine-tune, and evaluate machine learning models, respectively, ensuring accurate performance assessment and prevention of overfitting.

5.2.1 Spliting data for 30 features

Model name	**Split ratio**		
	85:15	**80:20**	**70:30**
Gradient Boosting Classifier	**0.97**	**0.97**	**0.98**
Linear Svm	0.92	0.91	0.90
Decision Tree Classifier	0.90	0.95	0.93
Random Forest Classifier	0.89	0.89	0.91
Logistic Regression	0.85	0.90	0.88
Radial Svm	0.84	0.88	0.85
Gaussian Naive Bayes	0.81	0.86	0.85
KNeighbors Classifier	0.80	0.84	0.82

Table 5.3: Performance of different classifiers with different data split ratios for 30 features

When we applied the hold-out method we split our dataset into train and test into different ratios and applied different machine learning models and we get also good results in the gradient boosting algorithm. For the 85:15 ratio, the accuracy is 0.97, for the 80:20 ratio the accuracy is 0.97 and for the 70:30 ratio the accuracy is 0.98.

5.2.2 Splitting data for 21 features

Model name	**Split ratio**		
	85:15	**80:20**	**70:30**
Gradient Boosting Classifier	**0.98**	**0.97**	**0.97**
Linear Svm	0.91	0.92	0.90
Decision Tree Classifier	0.90	0.95	0.93
Random Forest Classifier	0.90	0.93	0.93
Logistic Regression	0.88	0.91	0.87
Radial Svm	0.85	0.92	0.87
Gaussian Naive Bayes	0.85	0.89	0.88
KNeighbors Classifier	0.82	0.86	0.80

Table 5.4: Performance of different classifiers with different data split ratios for 21 features

We also get good results when we use 21 features for all the models but all of this the gradient boosting algorithm gives us the best result. For the 85:15 ratio, the accuracy is 0.98, for the 80:20 ratio the accuracy is 0.97 and for the 70:30 ratio the accuracy is 0.97.

5.3 Score for individual classifier

5.3.1 Score for individual classifier using 30 feature

Model name	**F1_score**		**rmse**	**accuracy_score**	
	Train_data	**Test_data**		**Train_data**	**Test_data**
GradientBoostingClassifier	**1.00**	**0.97**	0.16	**1.00**	**0.97**
SupportVectorClassifier	0.98	0.91	0.28	0.98	0.92
DecisionTreeClassifier	0.99	0.90	0.30	0.99	0.90
RandomForestClassifier	0.99	0.89	0.32	0.99	0.89
GaussianNaiveBayes	0.92	0.81	0.43	0.92	0.81
KNeighborsClassifier	0.85	0.76	**0.47**	0.85	0.77

Table 5.5: Performance metrics of classifiers using 30 features

Individually when we apply all the models to see the f1 score, classification report, confusion matrix, rmse, and the accuracy score. In this case, we see the gradient boosting also works well it's train set accuracy is 1.0 and the test set accuracy is 0.97.

5.3.2 *Score for individual classifier using 21 feature*

Model name	F1_score		rmse	accuracy_score	
	Train_data	Test_data		Train_data	Test_data
GradientBoostingClassifier	**1.00**	**0.97**	0.16	**1.00**	**0.97**
SupportVectorClassifier	0.99	0.89	0.32	0.99	0.89
DecisionTreeClassifier	**1.00**	0.91	0.28	**1.00**	0.92
GaussianNaiveBayes	0.95	0.85	**0.38**	0.95	0.85

Table 5.6: Performance metrics of classifiers using 21 features

When we apply 21 features in all the models then we see the gradient boosting again works better. Its train set accuracy is 1.0 and the test set accuracy is 0.97.

5.4 Hyper parameter tuning

Hyperparameter tuning in machine learning is the process of finding the best settings for a model's pre-defined parameters. These settings greatly impact the model's performance, and tuning them optimizes accuracy, convergence speed, and robustness. Techniques like grid search, random search, Bayesian optimization, and gradient-based optimization are used to find the optimal parameter values. Tuning is crucial for maximizing model effectiveness and generalization to new data. According to our model, the comparison table is:

Model	Before Hyper Tune		After Hyper Tune	
	Train data	Test data	Train data	Test data
Support Vector Classifier	0.98	**0.91**	**1.00**	**0.9067**
Random Forest Classifier	**0.99**	0.89	**1.00**	0.8933

Table 5.7: Comparison before and after hyper parameter tuning

Before hyperparameter tuning we get the train set accuracy is .98 but after hyperparameter tuning the accuracy increases to 1.0 in svm. Test set accuracy is also increased a little.

Validation data accuracy		
Model name	**30 features**	**21 features**
K-Neighbors Classifier	**0.45**	**0.50**
Linear Svm	0.43	0.43
Logistic Regression	0.43	0.47
Random Forest Classifier	0.43	0.39
Radial Svm	0.41	0.41
Gradient Boosting Classifier	0.39	0.39
Decision Tree Classifier	0.39	0.37
Gaussian Naive Bayes	0.395	0.45

Table 5.8: Validation dataset accuracy

After applying the validation dataset we see the accuracy is not so bad. For 30 features we get 0.45 and for 21 features we get an accuracy is 0.50 in the K-neighbors classifier.

6 Conclusion

In this report we applied different machine learning algorithms and techniques in the ”Emotional drivers of the vaccination hesitancy and refusal: A dataset from Slovakia” and tried to classified the people hesitancy about the vaccine and vaccination in different situation like COVID-19 pandemic. We visualized the all data and according to nature of dataset we classified the dataset into 3 class that a person is agree, disagree, or in hesitation to take vaccine. After applying the machine learning models we got best training and testing accuracy by using Gradient Boosting Classifier (In K-fold cross validation accuracy is **0.98**. To validate our model accuracy we collected some real data and applied same techniques and models we got the best accuracy score for K-Neighbors classifier, that is **0.50**.

6.1 Limitations

We faced a lot of difficulties in our work. We had some limitations that this paper did not overcome like:

- For just being a two-member team we could not collect more data for validation data that’s why the accuracy of validation data is low
- Could not collect sub parts of each average columns data.
- Demographic fields prevents better accuracy.
- For the above reasons we were not able to make our own model which would have been best suited for this dataset.

6.2 Future Works

Using this dataset, we are attempting to address individual emotional behavior while simultaneously developing a custom model for this purpose.We are also attempting to use the emotional data to identify the psychological state and illness of each individual as well as the population as a whole throughout our subcontinent. In parallel, we are working to get over the obstacles and create a unique model that will support our future efforts in the field globally.

References

[1] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[2] P. Barrett, J. Hunter, J. T. Miller, J.-C. Hsu, and P. Greenfield, "matplotlib–a portable python plotting package," in *Astronomical data analysis software and systems XIV*, vol. 347, 2005, p. 91.

[3] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: https://doi.org/10.21105/joss.03021

[4] J. C. Walsh, M. Comar, J. Folan, S. Williams, and S. Kola-Palmer, "The psychological and behavioural correlates of covid-19 vaccine hesitancy and resistance in ireland and the uk," *Acta Psychologica*, vol. 225, p. 103550, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001691822000658

[5] C. Rancher, A. D. Moreland, D. W. Smith, V. Cornelison, M. G. Schmidt, J. Boyle, J. Dayton, and D. G. Kilpatrick, "Using the 5c model to understand covid-19 vaccine hesitancy across a national and south carolina sample," *Journal of Psychiatric Research*, vol. 160, pp. 180–186, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022395623000778

[6] A. Raut, A. Samad, J. Verma, and P. Kshirsagar, "Acceptance, hesitancy and refusal towards covid-19 vaccination," *Clinical Epidemiology and Global Health*, vol. 21, p. 101283, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213398423000702

[7] H. Gupta and O. P. Verma, "Vaccine hesitancy in the post-vaccination covid-19 era: A machine learning and statistical analysis driven study," *Evolutionary Intelligence*, vol. 16, no. 3, pp. 739–757, 2023.

[8] M. Saifuzzaman, M. M. Rahman, S. F. Shetu, and N. N. Moon, "Covid-19 and bangladesh: situation report, comparative analysis, and case study," *Current Research in Behavioral Sciences*, vol. 2, p. 100034, 2021.

[9] M. Adamus, V. Čavojová, and E. Ballová Mikušková, "Emotional drivers of the vaccination hesitancy and refusal: A dataset from slovakia," *Data in Brief*, vol. 47, p. 108980, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340923000987

[10] S. Roccas, L. Sagiv, S. H. Schwartz, and A. Knafo, "The big five personality factors and personal values," *Personality and Social Psychology Bulletin*, vol. 28, no. 6, pp. 789–801, 2002. [Online]. Available: https://doi.org/10.1177/0146167202289008

[11] S. G. Brown, R. K. Ikeuchi, and D. R. Lucas III, "Collectivism/individualism and its relationship to behavioral and physiological immunity," *Health Psychology and Behavioral Medicine: An Open Access Journal*, vol. 2, no. 1, pp. 653–664, 2014.

[12] A. Strathman, F. Gleicher, D. S. Boninger, and C. S. Edwards, "The consideration of future consequences: Weighing immediate and distant outcomes of behavior." *Journal of personality and social psychology*, vol. 66, no. 4, p. 742, 1994.

[13] T. Besley and M. Ghatak, "Prosocial motivation and incentives," *Annual Review of Economics*, vol. 10, pp. 411–438, 2018.

[14] C. Gordon and A. Arian, "Threat and decision making," *Journal of conflict resolution*, vol. 45, no. 2, pp. 196–215, 2001.

[15] M. Kretchner, U. Lifshin, and M. Mikulincer, "Parental motivated helplessness in vaccinating children against covid-19: Its association with fear, effectiveness and

willingness to vaccinate," *Journal of Health Psychology*, p. 13591053231170055, 2023.

[16] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini, "The covid-19 pandemic," *Critical reviews in clinical laboratory sciences*, vol. 57, no. 6, pp. 365–388, 2020.

[17] H.-W. Bierhoff, *Prosocial behaviour*. Psychology Press, 2005.

[18] F. Booysen, S. Guvuriro, and C. Campher, "Horizontal and vertical individualism and collectivism and preferences for altruism: A social discounting study," *Personality and Individual Differences*, vol. 178, p. 110856, 2021.

[19] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*, 2015.

[20] T. Joachims, "Training linear svms in linear time," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 217–226.

[21] G. L. Prajapati and A. Patle, "On performing classification using svm with radial basis and polynomial kernel functions," in *2010 3rd International Conference on Emerging Trends in Engineering and Technology*. IEEE, 2010, pp. 512–515.

[22] S. C. Lemon, J. Roy, M. A. Clark, P. D. Friedmann, and W. Rakowski, "Classification and regression tree analysis in public health: methodological review and comparison with logistic regression," *Annals of behavioral medicine*, vol. 26, pp. 172–181, 2003.

[23] B. Juba and H. S. Le, "Precision-recall versus accuracy and the role of large data sets," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4039–4048.

[24] C. J. Van Rijsbergen, "Information retrieval," *Butterworth-Heinemann*, 1979.

[25] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)," *Geoscientific model development discussions*, vol. 7, no. 1, pp. 1525–1534, 2014.