EXPLORING AND EVALUATING THE PERSONALITY TRAITS ASSOCIATED WITH SUBSTANCE ABUSE IN YOUNG PEOPLE

Submitted by – Theekshitha Varatharajasarma (COADDS221F-009)

BSc(Hons) Data Science

National Institute of Business Management(NIBM) - NIC

Colombo, Sri Lanka

Date of Submission - 30th of April 2023

A project report submitted to the national Institute of Business management for the partial fulfillment of requirement of the Advanced Diploma in Data Science.

DECLARATION

I hereby declare that the work presented in this project report was carried out independently by myself and have cited the work of others and given due reference diligently.

……………………………….

Theekshitha Varatharajasarma

……………………………..

Date

I certify that the above student carried out his/her project under my supervision and guidance.

……………………………..

Supervisor

: Ms. Chamilanka Wanigasekara

……………………………..

Date

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my lecturers and especially to my project supervisor, Ms. Chamilanka Wanigasekara, for their guidance and feedback and advice. I would like to extend my gratitude to my fellow classmates with whom I was able to work together. Lastly, I would be remiss in not mentioning my family, especially my mother without whose support I could not have undertaken the journey of my academics.

ABSTRACT

The purpose of this study is to investigate relationships between substance usage, personality traits and IQ scores. Patterns of usage of drugs, inter-dependencies between different drugs and the risk of being prone to drug abuse in different people are to be explored as well.

The design of this study includes an introduction, discussion about the research theme, literature review, objectives, hypotheses, a guide about the variables and key words and then moving into the methodology used to perform the research and diving into the data analysis and interpretation for each objective and finally concluding the report based on the results and patterns identified with additional recommendations.

The study has found a significant relationship between substance usage and intentionality bias. However, such a relationship was not found for sense of agency, unpredictability, free will as well as IQ scores. However, some factors relating to substance abuse have been identified to have impacted these scores. A prediction model could not be drawn from the data present. No any relationship between IQ scores and personality trait scores has been found and no significant relationship between existing personality disorders, calculated IQ scores and the risk of using substances could be shown as well.

The study performed well and gathered strong evidence about inter-dependencies between the usage of one drug and another. It also shows that drug usage affects people of different ages and genders differently. The research shows that almost all the drugs are used by men more than women and the most used drugs are cannabis, ecstasy, and amphetamine.

All of these processes and findings are discussed in a detailed manner with the limitations present. Overall, it would include the use of secondary data, answers for questionnaires that may not be completely dependable even though anonymous, data constrained to a specific country/region and inability to construct an accurate prediction model. Nonetheless, this research contributes to understanding the effects of drug usage on young individuals and highlights the importance of addressing substance abuse in society.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

Substance abuse is a pressing issue in the current world with a 26 percent increase over the previous decade (UNODC -2020). It is also a growing concern that more young people are using drugs at increasing frequency levels. The use of various drugs has been identified to have negatively impacted one's physical and psychological health, including effects in addiction, personality, and behavioral problems.

For effective prevention, awareness, and treatments to be put in place, it is important to understand how substance abuse impacts personality and behavioral patterns. In depth, it is also important to understand the various substances, their risks of them being used, and the relationship they have with young people of both genders in different stages of life. Having that on mind, this research seeks to explore and analyze substance abuse, especially in youngsters, investigating and differentiating between different drugs and the effects they impose with the focus on personality traits and intelligence quotient.

1.2 Research Problem

With increasing levels of substance usage especially in youngsters, it is importance to analyse the status of substance usage of different drugs in different people for facilitating the needed attention in an effective way, to the groups that need most help with an understanding of the effects of substance abuse. As there are multiple studies about the medical conditions and mental health disorders created by substance use, this study gives a special focus as to how drug usage affects one's day-to-day personality, his intentions, and activities. As drugs negatively affect one's brain function, knowing the relationship between related personality trait issues and drug consumption could facilitate effective awareness to people in a way lay-people could understand as well as the prevention strategies.

1.3 Research Questions

i. What is the substance usage situation currently and how does it affect a person's personality and behavior?
ii. What are the risks young people tending to face because of substance usage?
iii. What drugs do different people are most exposed to?
iv. What are the dependencies among drug usage, age, gender, personality traits, personality disorders as well as IQ scores?

1.4 Objectives of the Project

The objectives of this project are to:

i. Identifying whether there is a relationship between substance usage and personality traits.
ii. Finding whether the IQ level of a person is affected by drugs and related personality traits.
iii. Identifying the most exposed and frequently used drug and the demographic groups of people most affected.
iv. Analyzing whether usage of one drug enhances the chances of using other ones.
v. Exploring the difference in impacts caused by different drugs on different types of people.

1.5 Scope of the Research

The main scope of research is to identify usages of different drugs and understanding the effects and risks by analyzing associated changes in personality traits. Even though the research focuses on young people it covers, around people from teens to mid-thirties and divides people according to gender to identify the risks better and to increase the diversity of the sample. The research focuses on people using drugs or have used drugs in the past and evaluate their personality from the tests conducted and the results got. The study will analyze the effects of different types of drugs and frequencies of usage on personality traits.

1.6 Justification of the Research

This research is justified by the growing percentage of people using drugs, the growing concerns in mental health and the impacts of these substance use that affects a very importance part of people in the society that are of youngsters.

The research seeks to fill a gap in the literature by mainly focusing the effects of substance abuse on personality traits which is a sub-area that is comparatively unexplored that would facilitate preventive measures and awareness on this specific area.

1.7 Expected Limitations

The main limitation of this research study is not being able to collect data on my own due to lack of time and having to use secondary data and hence not being able to focus specially on the region of the world we reside in which would be another limitation. Another one would be even though the details are collected anonymously by professionals in an effective way, we cannot be sure about people giving a hundred percent honest answers which might have an impact on the study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to the Research Theme:

Substance abuse is a growing concern especially among young people worldwide. It not only affects their physical and mental health but also has a considerable effect on their day-to-day activities, personality traits and intelligence. The theme of the research aims to focus and explore the relationship between the consumption and frequency of difference substances as well as its impact on personality traits. This study also explores whether different people divided by age sex or, pre-existing personality disorders such as narcissism have higher risks of experiencing changes in personality traits such as sense of agency and intentionality bias and the relationship these factors have with our focus variable- substances/drugs.

2.2 Theoretical Explanation About the Key Words in the Topic:

Table 2.2 – Key Words

Key Words	Definition/ Explanation	Source
Substance Abuse	The harmful or hazardous use of psychoactive substances, including alcohol and illicit drugs, which can lead to physical and mental health problems, social and legal consequences, and addiction.	American Psychiatric Association. (2013).
Personality Disorder	A mental health condition that affects an individual's thoughts, feelings, and behaviors, leading to patterns of behavior that deviate significantly from cultural norms and expectations	
Personality Traits	Enduring patterns of thoughts, feelings, and behaviors that differentiate individuals from one another	McCrae, R. R., & Costa, P. T. (1987)
Sense of Agency	The feeling of control over one’s own actions	Render A, Jansen P(2019):

Intentionality Bias	"Automatic tendency to judge other people's actions to be intentional." The tendency to overestimate the role of intentional causes and underestimate the role of situational or environmental factors when explaining other people's behavior.	Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002).
IQ Scores	Measures of an individual's cognitive abilities, particularly their problem-solving and reasoning skills	Wechsler Adult Intelligence Scale (WAIS)
Dopamine	A neurotransmitter that plays a crucial role in the regulation of movement, motivation, reward, and emotional arousal in the brain	Yamamoto T, Fujieda Y, Miura H, et al. (2016).

2.3 Findings by Other Researchers:

This section seeks to explore the studies conducted by previous researchers and the findings by them in related topics to our research theme.

In 2006, a study that focused on frequent substance dependent people and pathological gamblers, the findings included that frequent alcohol consummators have showed changes in their personality traits in a way it increased impulsivity, anxiety as well as depression. They also showed decreased planning abilities, cognitive flexibility, and time management. It showed that these impaired functions were minimal in nicotine users(smokers) compared to the other ones. (Goudriaan et al., 2006).

A study that researched about dopamine in drug usage and dependance has used imaging studies to identify the effects of drug abuse in the dopaminergic system and in the human brain to find effective treatment strategies. It states that as brain on drug produces elevated levels of dopamine, to restore behavioral patterns and brain functions affected by drug use on a chronic level, one should provide alternative activities that produces dopamine and reinforces the system for drug addicts to control taking drugs. (Volkow N D et al., 2007)

Another study has stated that alcohol use has been linked to significant cognitive impairment, including memory and decision-making difficulties. It says that alcoholism depends on internal(genetic) as well as external/environmental variables. And studying alcoholics' neuropathology and neuroimages, their brain is more vulnerable to damage which results in the cognitive impairments and problems stated above. (Oscar-Berman & Marinkovic, 2007)

When analyzing the changes in brain and neurological performances of drug users based on brain images, research found that substance use creates negative mood, a weakening of the control circuits. It also concluded that even though in the initial stages drug use mostly affects voluntary behavior, frequent and long-term usage can cause impairment on neuronal circuits and affects the free will of a person making them to act in a different way caused by the drug. It also showed that compared to Cannabis, Cocaine had severe effects on cognitive abilities. And users of the drug have showed impairment on attention, decision making and working memory abilities. (Volkow et al., 2010).

Genetically and hormonally males and females have many differences. A research focused on how gender affects difference on the effects caused by selected substances and chemicals and got the results which is that males are more likely to use drugs and that they may experience severe cognitive impairment than females as well. (Fattore & Fratta, 2010).

By revising the National Epidemiological Survey on Alcohol and Related Conditions (NESARC), a study in 2010 has analyzed the prevalence of it with substance dependance. It has been found that substance abuse is linked to the development of several personality disorders such as borderline personality disorder, narcissistic personality disorder, antisocial personality disorder etc. (Trull et al., 2010).

Another study researched substance use and its effects on personality in middle-aged people. Levels of personality scores were used to predict the chances of longitudinal substance use. The results were that higher levels of neuroticism, extraversion and openness and lower levels of conscientiousness and agreeableness have predicted increased levels of substance usage. Neuroticism, extraversion, openness, conscientiousness, and agreeableness are the five tested personality traits in the study. (Turiano N A et al., 2012)

By analyzing the role of cognitive control in drug dependance, researchers found that drugs affect cognitive control and impairment in task performance in a significant level in some instances. They also stated that people who use cocaine were identified with lower self-control and sense of agency which eventually makes it harder for them to quit and make them dependent. (Luijten et al., 2014)

A researcher proved positive results when analyzing whether cannabis usage in a frequent manner impacts neuropsychological functions. Heavy Cannabis (Marijuana) users have had greater impairment than light users on attentive and executive functions. For these heavy users, even a day of abstinence from the drug showed significant differences in these functions. (Pope & Yurgelun-Todd, 2016)

In a study of determining factors in drug abuse and personality traits, it has found that drug usage affects a person's sense of agency and especially people with narcissism experienced a reduced sense of agency. However, it did not show a correlation between sense of agency and intentionality Bias. (Render A, Jansen P 2019)

Drug abuse has been found to be negatively affect our brain and frequent usage can result on addiction as the brain gets used to drugs and keeps wanting more and more. Neuroplasticity is the brain's ability to change and adapt according to its experiences. It is what makes it possible for us to learn new languages, skills etc. It can adapt to habitual functions. However, in a negative shade, the same thing happens due to habitual drug usage resulting in addiction. And this is the reason continued therapy often helps drug users to learn new, healthier habits, practice them and adopt to them instead. (Dr. Mavrikaki M 2020).

Considering the use of finding the association between drug usage and its negative impacts, data-prevention strategies have been identified to have developed to address the issue of substance abuse, including education programs, counseling, and treatment. Research and data driven solutions have shown that early intervention and prevention efforts can be effective in reducing drug use and its negative consequences (National Institute on Drug Abuse, 2021).

2.4 The research gap:

While previous studies have explored about relationships between substance abuse with brain functions, cognitive functions, and physical and mental health, the ones exploring the relationship substance abuse has with personality traits is comparatively less. In the ones that explore this topic, there is still a lack of understanding about how it affects especially young people and the relationships between substance abuse, personality traits, intelligence quotient (IQ), and existing personality disorders. This study seeks to fill the research gap by analyzing the above factors and also adding how gender difference affects these substances as well as dependency of the usage of one drug to another.

2.5 Table for Variables, Their Definitions, and Sources:

Table 2.5 – Variables.

Variable	**Definition**	**Source**
Participant	The participant number/id	Render A, Jansen P (2019)
Age	Age of the participant	
Sense of Agency	The feeling of control over one's own actions	
Intentionality Bias	The tendency to attribute intentional agency to events or entities, even when there is no evidence of such intent	Knobe & Nichols, (2007)
Narcissism	A personality disorder characterized by a grandiose sense of self-importance, a lack of empathy for others, and a preoccupation with fantasies of power, success, and attractiveness.	American Psychiatric Association. (2013).
Free will	"The power of acting without the constraint of necessity or fate; the ability to act at one's own discretion." This means the ability to act freely without getting affected by external factors.	Oxford English Dictionary
Unpredictability	The quality of being impossible to know about or predict"	Cambridge Dictionary

<table>
<tr><td>Cannabis
Ecstasy
LSD
Mushrooms
Amphetamine
Cocaine
Ketamine</td><td>"Substances that, when taken into the body, alter the normal functioning of the body, its mental processes or behavior."

Illegal drugs: "Controlled substances that have no recognized medical use and a high potential for abuse."</td><td>World Health Organization(WHO)
&
United States Drug Enforcement Administration (DEA)</td></tr>
<tr><td>FREQUENCIES</td><td rowspan="6">This gives the frequency of usage of legal and illegal dangerous drugs.

Psychotropics are a type/group of drugs that affect a person’s mental state.

Tranquilizers are a type/group of drugs that is used to reduce anxiety tension etc. Some of these when prescribed can be used for medical purposes in the right amount.</td><td rowspan="6">Render A, Jansen P(2019)</td></tr>
<tr><td>Legal Drugs
i. Fre_Alcohol</td></tr>
<tr><td>Drugs legal for medical purposes only
i. Fre_Cannabis
ii. Fre_Ketamine</td></tr>
<tr><td>Illegal Drugs
i. Fre_Cocaine
ii. Fre_Amphetamine
iii. Fre_Ecstasy
iv. Fre_LSD
v. Fre_Mushrooms</td></tr>
<tr><td>Fre_Psychotropics</td></tr>
<tr><td>Fre_Tranquilizers</td></tr>
<tr><td>More_than_cannabis</td><td>Whether people have used or are using at least one drug experimented about, except cannabis</td><td rowspan="2">Render A, Jansen P(2019)</td></tr>
<tr><td>More_than_cannabis =2(FILTER)</td><td>Whether people have used or are using two or more drugs experimented about, except cannabis</td></tr>
</table>

2.6 Chapter Conclusion

In conclusion, substance abuse is a global issue that is becoming a growing concern, especially among young people. It has a significant negative impact on intelligence, personality traits, daily activities, physical and mental health, and daily activities. Prior research has demonstrated that heavy substance use can reduce planning skills, cognitive flexibility, and time management while increasing impulsivity, anxiety, and depression. It leads the brain to addiction. Affecting cognitive control and task performance, substance use makes it more difficult for people to stop using and encourages dependency. Additionally, a number of personality disorders, including borderline personality disorder, narcissistic personality disorder, and antisocial personality disorder, have been linked to substance abuse. The effects of specific drugs and chemicals have also been found to differ between the sexes, with men being more likely to use drugs and suffering from more severe cognitive impairment. However, there is still a lack of understanding about how substance abuse affects young people and the relationships between substance abuse, personality traits, intelligence quotient (IQ), and existing personality disorders which this study seeks to explore.

CHAPTER 3
METHODOLOGY

3.1 Introduction

This chapter focuses on the types of data to be analyzed, data collection tools, the conceptual framework for doing the research, hypothesis to be tested, operationalization table giving additional details about the variables, and finally the data analysis methods that are to be used. As the research aims to find the usage and frequency usage of various drugs and its effect on personality traits especially on younger people, this chapter gives a clear idea about the procedures used to collect the data and the methods that are to be used to analyze the data collected.

3.2 Population, Sample, and Sampling Technique

I have downloaded the data from data.mendeley.com. This dataset was originally created by Anna Render, Petra Jackson, at Faculty of Psychology, Pedagogic, & Sport Science, University of Regensburg, Bavaria, Germany in 2019. The population for this study is young generation who have exposure to drugs. The sample size is 210 participants aged from 17 to 34 years consisting of 84 males and 126 females in Germany and they were tested under different experiments.

The majority of the participants were students who studied in the fields of psychology, sport sciences, arts and humanities, criminology, law, natural sciences, and other disciplines. Some of the participants were working people and not students. For the data to be mor accurate, data was analyzed anonymously without the need of participants informing about their personal lives and they were also informed about the procedures, and they have consented to the experiments conducted with a written format before starting the procedures. The sample size consisted of 210 people.

The sampling was selected by a statistical technique “power analysis” which is used to estimate the smallest size of sample required to get proper results and required level of significance. To estimate the sample size necessary to explore the effects power analysis was used.

3.3 Type of Data to be Collected and Data Sources

By the original researchers, data relating to substance usage, personality traits, and IQ scores were collected through anonymous questionnaires and experiments were conducted in which the participants had to participate (with their consent) from which various scores have been calculated.

Through anonymous questionnaires:

- Age (Numerical)
- Gender (Categorical)
- Whether participants have consumed different drugs or not (Categorical-Nominal)
- If yes, the frequency of consumption (Categorical-Ordinal)

Through Experiments (All Numerical):

- Personality trait scores
- Narcissism score
- IQ scores

3.4 Data Collection Tools and Plan

The survey was designed in a way to collect data using anonymous questionnaires to get the information about demographic and drug usage related data as well as physical experiments conducted using the participants with which the score related data were finalized. It was anonymous for the purpose of increasing honest replies and to ensure people's privacy and confidentiality about their personal information from which they would be able encouraged to answer honestly. For ensuring the accuracy, the sample size was confirmed using a statistical power analysis and was pre-tested. 210 people residing in Germany, especially young people from the ages of 17 to 34 with a majority consisting of students were tested to get a more precise data about drug usage and its effects on youngsters.

"The study took part at the University of Regensburg, and the duration was roughly 50 minutes. Sessions started with the ZVT, followed by the computer experiment of intentionality bias and intentional binding. The second part included the online questionnaires generated with social survey. The ZVT and two computer-based experiments took about 30 minutes, and the questionnaires 20 minutes." (Render A, Jansen P 2019) `

3.5 Conceptual Framework

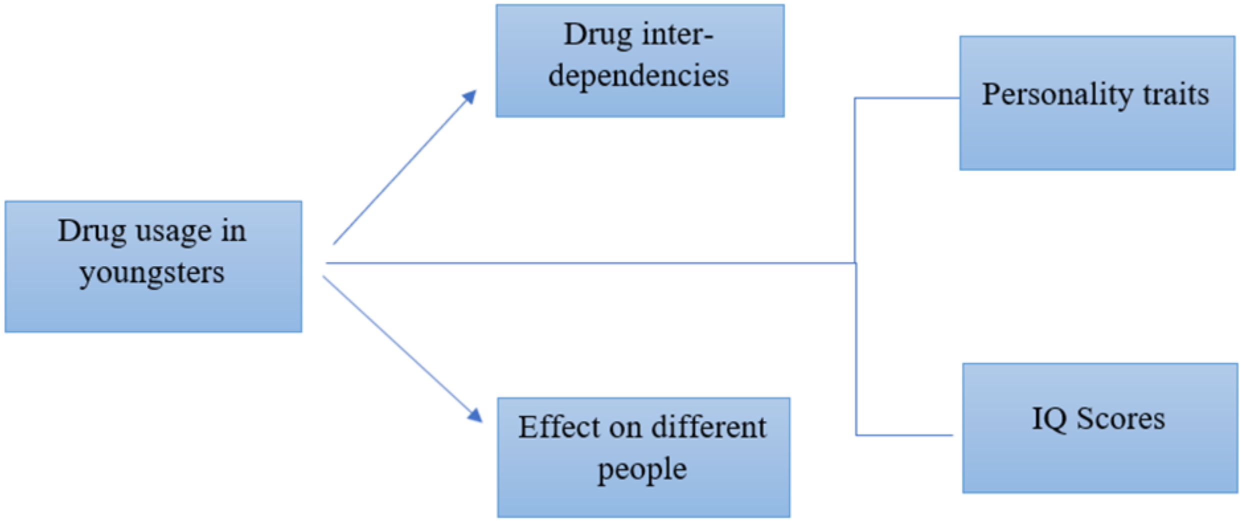

Figure 3.1 – Conceptual Framework

First when identifying whether there is a relationship between substance usage and personality traits, the independent variables will be the variables relating to drug usage(whether people use drugs or not, if yes, frequency of usage of drugs, whether people have used at least one drug except cannabis, whether people have used at least two drugs except cannabis.) The dependent variables will be sense of agency, intentionality bias, unpredictability, and free will. Each dependent variable will be tested individually with the independent variables.

After this, in finding whether the IQ level of a person is affected by drugs and related personality traits, we test the relationship between the independent variables: usage, frequency of usage of drugs and dependent variable IQ score. After this, we test the relationship between the independent variable personality trait scores and the dependent variable IQ scores.

Graphical Representations like pie-charts, bar-charts, box plots, scatter plots will be used to identify the most exposed and frequently used drug and the demographic groups of people most affected.

Moving onto analyzing whether usage of one drug enhances the chances of using other ones, usage and Frequency of usage of different drugs will be tested with one another (when one is dependent, the other is independent). This will be tested on most used drugs.

Finally, while exploring the difference in the impacts caused by different of drugs on different types of people, independent variables such as narcissism, IQ scores, age and sex will be tested against the dependent variable drug usage to check if there is a relationship between the two.

3.6 Hypothesis

The hypothesis is that the independent variables and their relative dependent variables have no significant relationship. This will be the null hypothesis and the alternative hypothesis will be that there is indeed a relationship. Our hypothesizes will be:

i. Identifying whether there is a relationship between substance usage and personality traits.
   - Null Hypothesis - Substance usage has no significant relationship with personality scores.
   - Alternative Hypothesis - Substance usage has a significant relationship with personality scores.

ii. Finding whether the IQ level of a person is affected by drugs and related personality traits.
   - Null Hypothesis - Personality trait scores have no significant relationship with IQ scores.
   - Alternative Hypothesis - Personality trait scores have a significant relationship with IQ scores.

iii. Analyzing whether usage of one drug enhances the chances of using other ones.
   - Null Hypothesis - Usage and/or frequency of usage of one drug does not affect the usage and/or frequency of usage of another drug.
   - Alternative Hypothesis - Usage and/or frequency of usage of one drug does affect the usage and/or frequency of usage of another drug.

iv. Exploring the difference in the impacts caused by different drugs on different types of people.
   - Null Hypothesis - Whether or not people have any existing personality disorders, their age and gender do not impact risk they have of using drugs.
   - Alternative Hypothesis - Whether or not people have any existing personality disorders, their age and gender impact risk they have of using drugs.

3.7 Operationalization Table

Table 3.7 – Operationalization Table

Variable	**Indicators**	**Measures**
Participant	Unique id	-
Age	Age	17 - 34 (Numerical)
Gender	Male or Female	Categorical 1 – Male 2 - Female
Sense of Agency	Score of sense of agency got by experiment.	By Scores (Numerical)
Intentionality Bias	Score of intentionality bias got by experiment	By scores (Numerical)
Narcissism	Scores of narcissisms got by experiment.	By scores (Numerical)
Free will	Scores of free will got by experiment.	By scores (Numerical)
Unpredictability	Scores of unpredictability got by experiment.	By scores (Numerical)
Cannabis Ecstasy LSD Mushrooms Amphetamine Cocaine Ketamine	Whether participants have used or are using these drugs or no.	Categorical Binary (For every Drug) 1 – No 2- Yes
Fre_Alcohol	The frequency of usage of the alcohol	0- Not at all 1- Once a month 2- 2-3 times a month 3- Once a week
Fre_Amphetamine Fre_Ecstasy Fre_Ketamine Fre_LSD Fre_Cocaine		0- Not at all 1- Every 3 months 2- Up to once a month 3- 2-3 times a month

Fre_Mushrooms Fre_Tranquilizers Fre_Psychotrops	The frequency of usage of all the other dugs AND The frequency of usage of tranquilizers and psychotropics which are types of drugs	4- 1-2 times a week 5- 3-4 times a week 6- Almost daily
Fre_Cannabis		0- Not at all 1- Every 3 months 2- Up to once a month 3- 2-3times a month 4- 1-2times a week 5- 3-4 times a week 6- Almost daily 7- Daily 8- Several times daily
More_than_ cannabis	Whether people have used or are using at least one drug experimented about, except cannabis	Categorical Binary 1 – No 2 - Yes
More_than_ cannabis =2(FILTER)	Whether people have used or are using at least one drug experimented about, except cannabis	Categorical Binary 1 – No 2 - Yes

3.8 Methods of Data Analysis

The data collected are analyzed using Python and RStudio. All the objectives stated analyzed statistically. Graphical representation as well as statistical tools are used.

The collected data will be analyzed with the use of,

i. Descriptive statistics (mean, median, mode, minimum, maximum, percentiles, quantiles, standard deviation, etc.)

ii. Graphical representations (pie-charts, histogram, bar charts, scatter plots, box plots, pair plots, etc.)
iii. Inferential statistics (ANOVA (Analysis of Variance) Test, T Test, chi-squared test, Multiple and Simple Linear Regression)

For the objectives finalized for the research the following methods, techniques, and tools.

i. Identifying whether there is a relationship between substance usage and personality traits.
   - ANOVA Test
ii. Finding whether the IQ level of a person is affected by drugs and related personality traits.
   - ANOVA Test & Multiple Linear Regression
iii. Identifying the most exposed and frequently used drug and the demographic groups of people most affected.
   - Graphical Representations like pie-charts, bar-charts, box plots, scatter plots.
iv. Analyzing whether usage of one drug enhances the chances of using other ones.
   - Chi-Squared Test
v. Exploring the difference in the impacts caused by different drugs in different types of people.
   - Logistic Regression

The null hypothesis is that there is no significant relationship between the variables tested. After the processes, hypotheses have been tested, and the null hypothesis has been accepted or declined accordingly. Relationships, results, and conclusions with relevant prediction models have been drawn.

CHAPTER 4

DATA ANALYSIS & INTERPRETATION

4.1 Introduction

In this chapter, I will discuss the data analysis I have performed by listing out the steps I have undertaken by explanations, results for the tests performed and visualization. I will first explain the data preparation process(preprocessing and wrangling) and then dive into the data analysis performed for each objective of my research.

4.2 Data Preprocessing and Data Wrangling

The original dataset I downloaded was in SAV file format. I have performed data transformation by converting it into an excel file for me to access and read it. After understanding After understanding about my dataset, I deleted the extra sheets about the variables and sources in order to make it into one sheet with my selected data.

As the dataset was not of an excessively big size, I did not have to do much data reduction or compressing to do. However as mentioned before I changed the downloaded dataset into one sheet in a csv format comprising of one sheet of data. As there were two header rows, I removed one.

I proceeded to continue this step by selecting columns according to my research objectives and removing other unwanted, extra space consuming data and finalized after naming the file according to my research and the selected data and columns. I performed data integration by changing the raw dataset to match my objectives of this research. For this purpose, I have changed some column names in an easily understandable manner and corrected minor spelling mistakes and then, finalized the personalized dataset. I have done these steps in python. I have exported the python data frame into a csv file to show the dataset in here.

I performed data cleaning. There were a considerable number of missing values which I had removed. I have checked the dataset for outliers and removed them as well. There were no duplicates in the dataset. I then used Boxplots to the numerical variables to check and visualize whether there are any outliers.

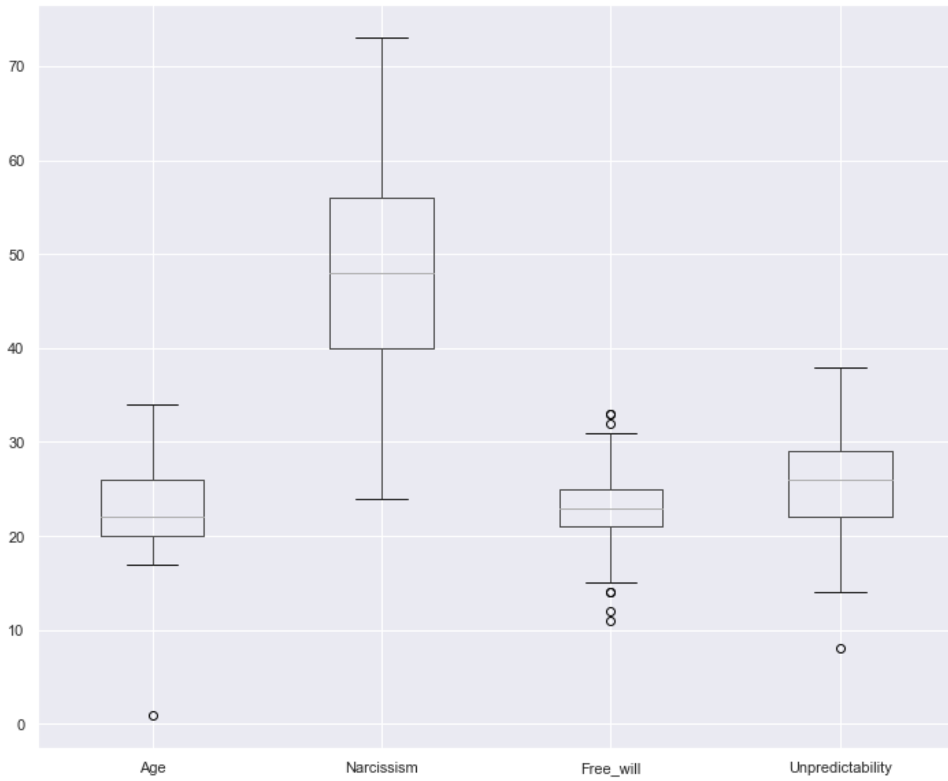

Figure 4.1 – Boxplot 1

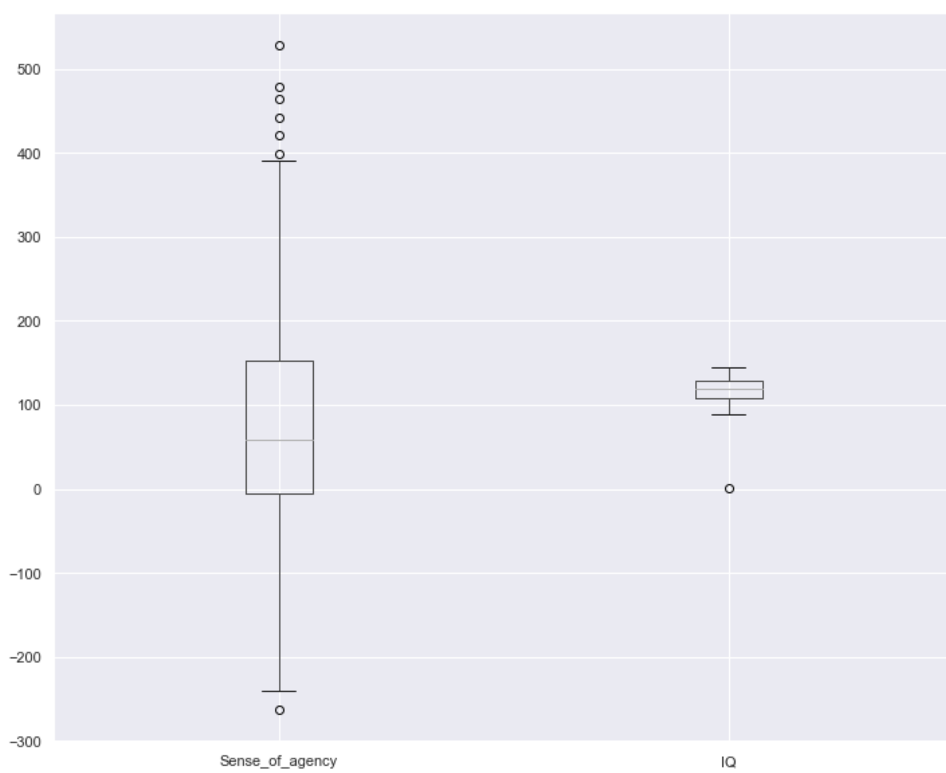

Figure 4.2 – Boxplot 2

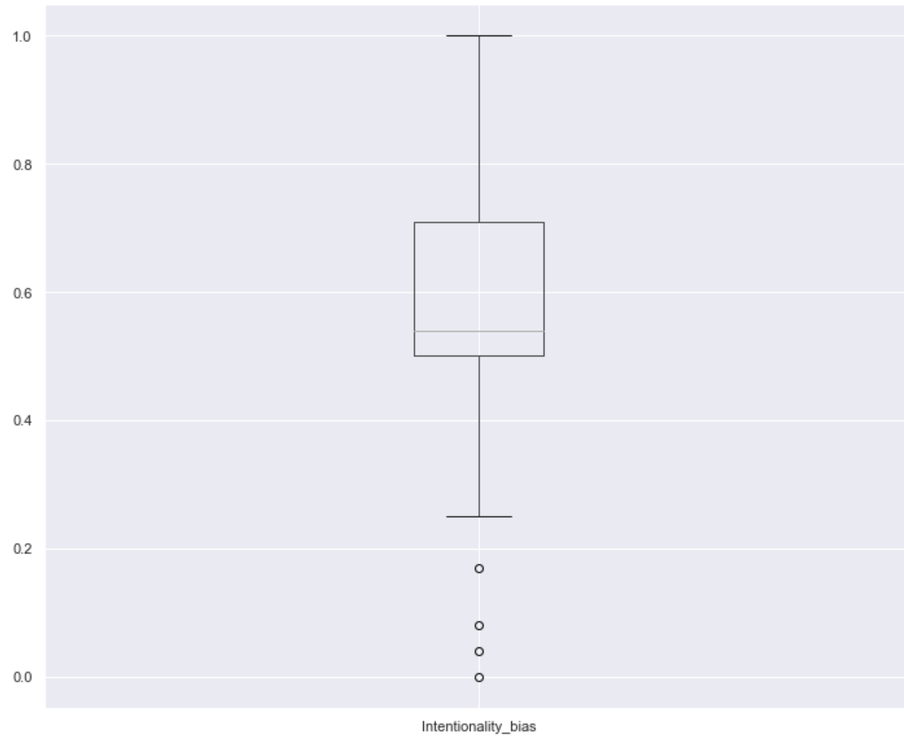

Figure 4.3 – Boxplot 3

Then, I removed the outliers present. Boxplots after removing the outliers are represented below.

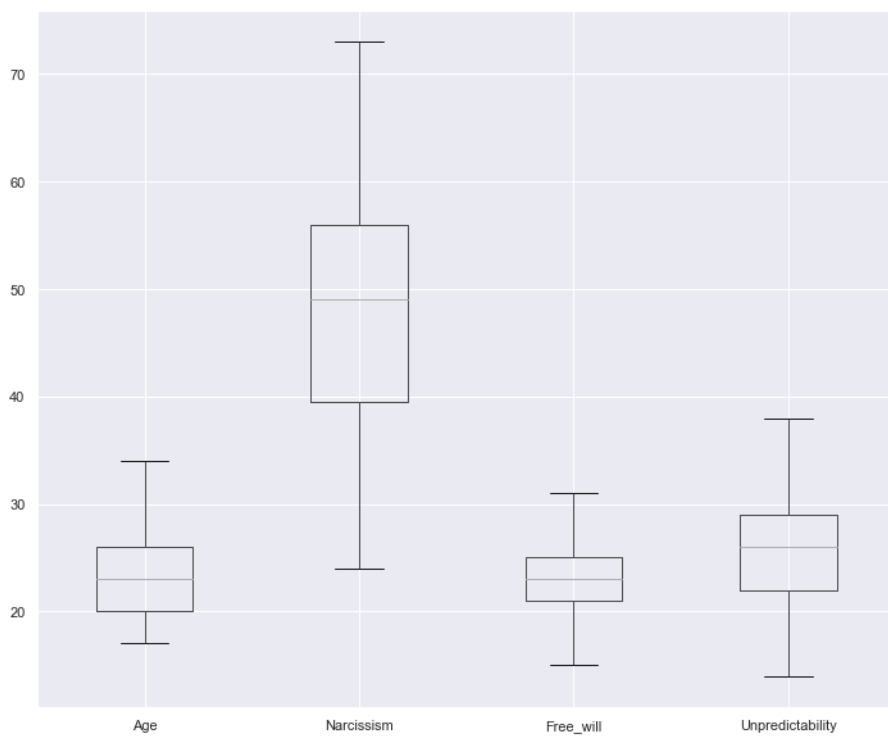

Figure 4.4 – Cleaned boxplot 1

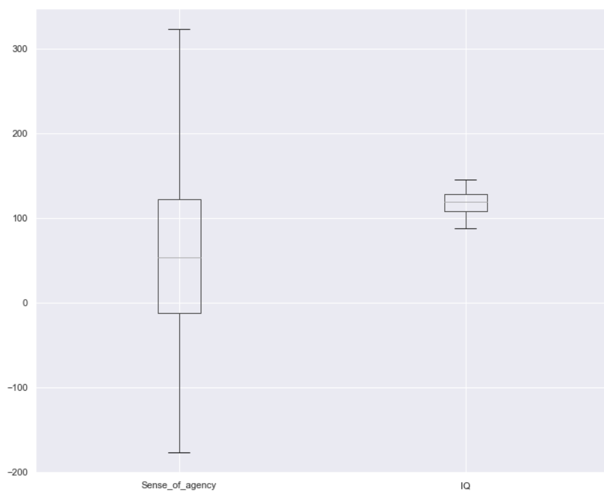

Figure 4.5 – Cleaned boxplot 2

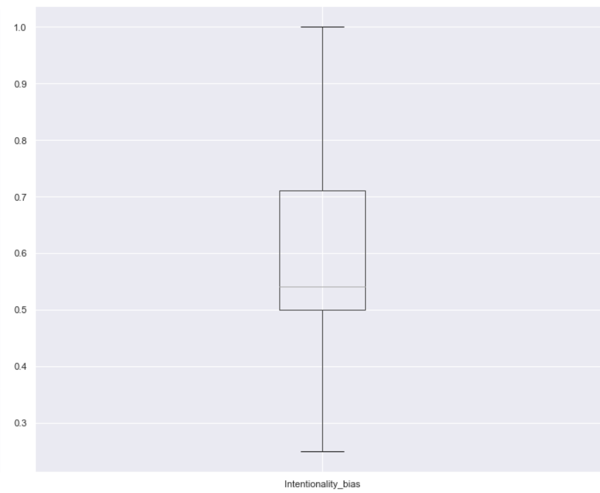

Figure 4.6– Cleaned boxplot 3

I then reshaped my dataset by combining or ordering the columns in a way that makes more sense as I have changed the order of the columns in a way where related variables will be near each other according to my research theme and objectives in a grouped manner for people to understand better and for me to work with the dataset easier. Now after double-checking I have finalized the dataset file and saved it after changing the file name to an easier, short, and related name (‘drug_and_personality’). I have also normalized the data before fitting it into a regression model.

4.3 Descriptive Analysis and Data Spread

Results of descriptive analysis consisting of count, mean, standard deviation, minimum, first quartile, median, third quartile and maximum.

	Age	IQ	Sense_of_agency	Intentionality_bias	Free_will	Unpredictability	Narcissism
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000
mean	23.329341	118.553892	67.893772	0.611257	23.131737	25.754491	47.976048
std	3.571040	15.779450	107.383373	0.177089	3.452868	4.989709	10.206868
min	17.000000	88.000000	-176.660000	0.250000	15.000000	14.000000	24.000000
25%	20.000000	107.500000	-12.100000	0.500000	21.000000	22.000000	39.500000
50%	23.000000	119.500000	53.450000	0.540000	23.000000	26.000000	49.000000
75%	26.000000	128.500000	122.000000	0.710000	25.000000	29.000000	56.000000
max	34.000000	145.000000	322.780000	1.000000	31.000000	38.000000	73.000000

Figure 4.7 – Descriptive statistics

As for the data spread, all the numerical variables have been represented using boxplots above. Data spread, patterns and relationships between personality traits and IQ scores according to gender are visualized below.

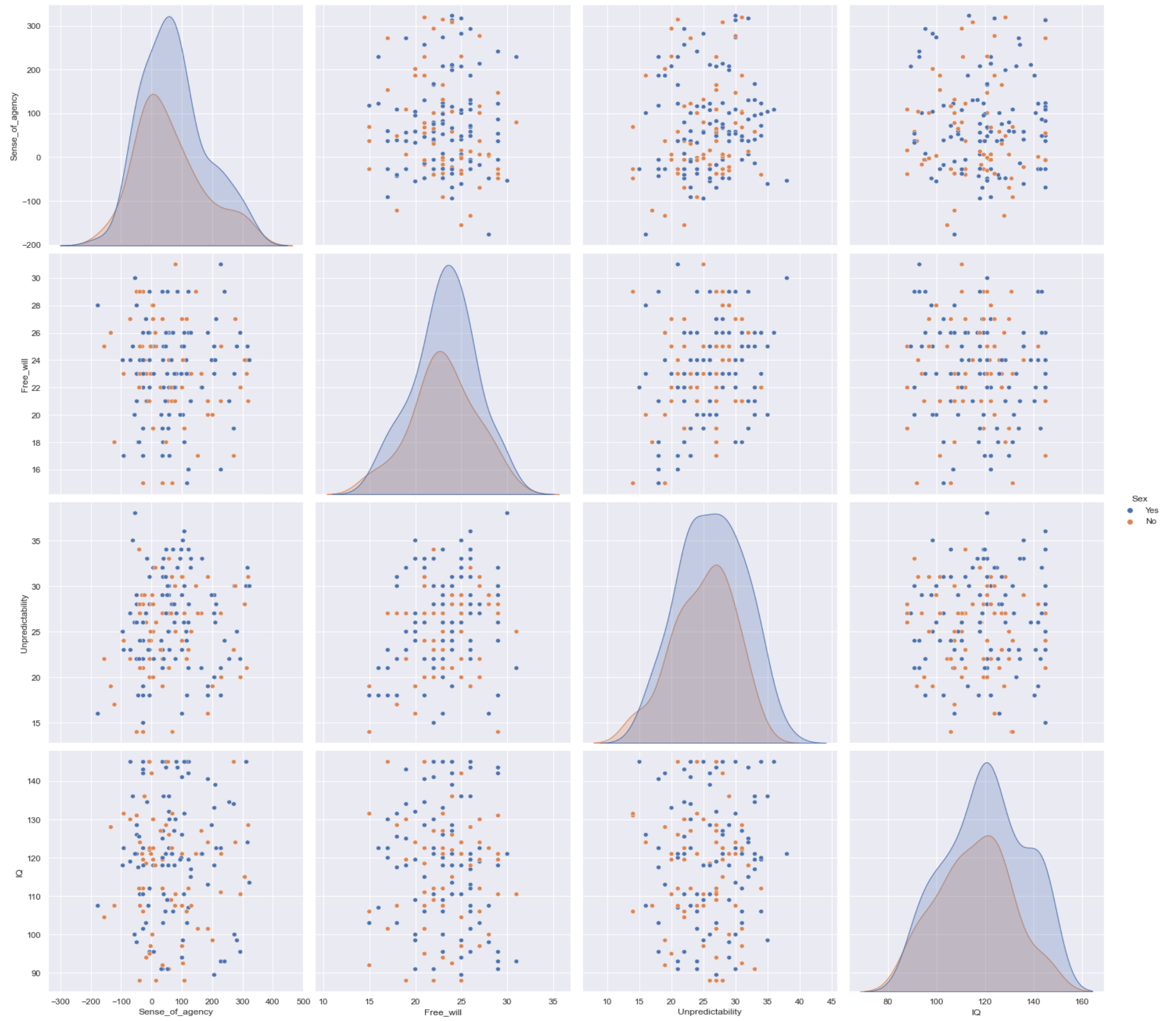

Figure 4.8 – Data spread

Categorical variables are described as it was shown in figure 3.7.

4.4 Substance Usage & Personality Traits

In this section, I have tested whether there is a relationship between substance usage and personality traits employing ANOVA test for each selected personality trait that are,

i. Sense of Agency
ii. Intentionality Bias
iii. Unpredictability
iv. Free Will

The four personalities have been used as Y variables. All of these are numerical variables given in the format of scores (Scores calculated by making the participants involved in various experiments).

For all four sections, same X variables have been used that are,

i. Drug Usage
   - Cannabis, Amphetamine, Ecstasy, LSD, Mushrooms, Ketamine, Cocaine
   - More than cannabis, Two more than cannabis
ii. Frequency of Drug Usage
   - Cannabis, Amphetamine, Ecstasy, LSD, Mushrooms, Ketamine, Cocaine
   - Alcohol
   - Psychotropics, Tranquilizers

All the X variables ae categorical collected from anonymous questionnaires by the same participants. The categories for each variable are listed in figure 3.7.

For all four tests, the hypothesis is,

- Null Hypothesis - Substance usage & frequency of usage have no significant relationship with personality scores.
- Alternative Hypothesis - Substance usage & frequency of usage have significant relationships with personality scores.

Two ANOVA tests were performed for each personality trait, one to find if there is a relationship between whether people use substances and the personality trait, another to find if there is a relationship between the frequency of substance usage and the personality trait.

4.4.1 Sense of Agency

Sense of Agency is defined as the feeling of control over one's own actions (Render A, Jackson P 2019) ANOVA test was performed to identify whether there is a relationship between sense of agency and drug usage. The results were,

	sum_sq	df	F	PR(>F)
C(Cannabis)	2.663362e+03	1.0	0.231115	0.631363
C(Amphetamine)	2.834264e+03	1.0	0.245945	0.620634
C(Ecstasy)	3.125701e+03	1.0	0.271235	0.603234
C(LSD)	1.516652e+04	1.0	1.316086	0.253030
C(Mushrooms)	4.686423e+03	1.0	0.406668	0.524589
C(Ketamine)	5.681461e+04	1.0	4.930131	0.027815
C(Cocaine)	2.257723e+04	1.0	1.959156	0.163564
C(Q("more_than_cannabis"))	1.187773e+03	1.0	0.103070	0.748601
C(Q("more_than_cannabis=2"))	1.187773e+03	1.0	0.103070	0.748601
Residual	1.820785e+06	158.0	NaN	NaN

Figure 4.9 – ANOVA - Sense of Agency & Drug Usage

	sum_sq	df	F	PR(>F)
C(fre_alcohol)	5.644453e+04	5.0	1.012327	0.413753
C(fre_tranquilizers)	1.636801e+04	4.0	0.366948	0.831745
C(fre_psychotropics)	4.754173e+04	3.0	1.421094	0.240238
C(fre_cannabis)	1.088958e+05	8.0	1.220647	0.293187
C(fre_amphetamine)	2.346531e+04	5.0	0.420848	0.833406
C(fre_Ecstasy)	3.762914e+04	6.0	0.562396	0.759441
C(fre_LSD)	2.755145e+04	5.0	0.494132	0.780080
C(fre_mushrooms)	1.467560e+04	4.0	0.329007	0.857995
C(fre_ketamine)	2.251787e+04	6.0	0.336546	0.916328
C(fre_cocaine)	4.135650e+04	5.0	0.741725	0.593762
Residual	1.282416e+06	115.0	NaN	NaN

Figure 4.10 – ANOVA - Sense of Agency & Frequency of Drug Usage

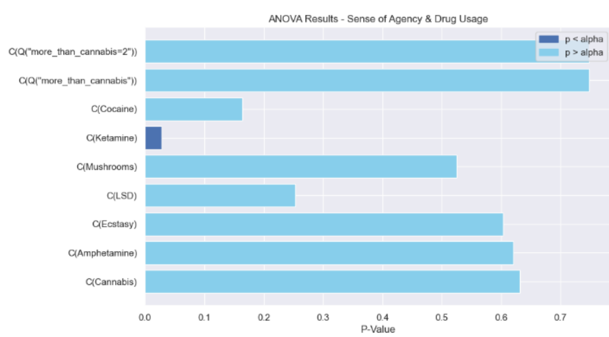

Figure 4.11 – Bar Plot 1 for ANOVA Results – Sense of Agency

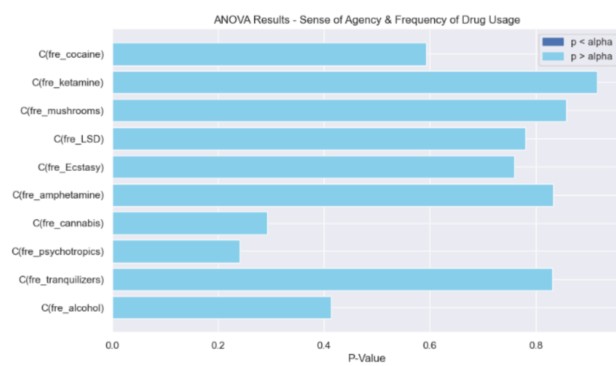

Figure 4.12 – Bar Plot 2 for ANOVA Results – Sense of Agency

In this case, the results show that there is a significant effect ketamine usage has in people's sense of agency scores (p value < alpha for ketamine usage). Hence, we can reject the null hypothesis for ketamine usage. However, the usage or the frequency of usage of any other drugs do not have any significant effect in people's sense of agency scores, indicating no evidence against the null hypothesis. Hence, we can accept the null hypothesis for all the other drugs tested. The large amount of unexplained variance is indicated by the huge residual sum of squares for both ANOVA tests, tested for drug usage and frequency of usage.

4.4.2 Intentionality Bias

Intentionality bias is the tendency to attribute intentional agency to events or entities, even when there is no evidence of such intent (Knobe& Nicholas 2007). It was tested whether there is a relationship between intentionality bias and drug usage by performing an ANOVA test. The results were,

	sum_sq	df	F	PR(>F)
C(Cannabis)	0.056172	1.0	1.804630	0.181080
C(Amphetamine)	0.001239	1.0	0.039797	0.842135
C(Ecstasy)	0.051451	1.0	1.652976	0.200435
C(LSD)	0.001034	1.0	0.033226	0.855597
C(Mushrooms)	0.000134	1.0	0.004302	0.947785
C(Ketamine)	0.047327	1.0	1.520468	0.219380
C(Cocaine)	0.096635	1.0	3.104586	0.080006
C(Q("more_than_cannabis"))	0.060991	1.0	1.959471	0.163530
C(Q("more_than_cannabis=2"))	0.060991	1.0	1.959471	0.163530
Residual	4.917987	158.0	NaN	NaN

Figure 4.13 – ANOVA – Intentionality Bias & Drug Usage

	sum_sq	df	F	PR(>F)
C(fre_alcohol)	0.190502	5.0	1.500341	0.195147
C(fre_tranquilizers)	0.068342	4.0	0.672802	0.612160
C(fre_psychotropics)	0.107525	3.0	1.411396	0.243070
C(fre_cannabis)	0.446650	8.0	2.198560	0.032422
C(fre_amphetamine)	0.102312	5.0	0.805778	0.547809
C(fre_Ecstasy)	0.177237	6.0	1.163228	0.330916
C(fre_LSD)	0.143718	5.0	1.131883	0.347537
C(fre_mushrooms)	0.257197	4.0	2.532021	0.044116
C(fre_ketamine)	0.389574	6.0	2.556816	0.023189
C(fre_cocaine)	0.125431	5.0	0.987860	0.428366
Residual	2.920363	115.0	NaN	NaN

Figure 4.14 – ANOVA – Intentionality Bias & Frequency of Drug Usage

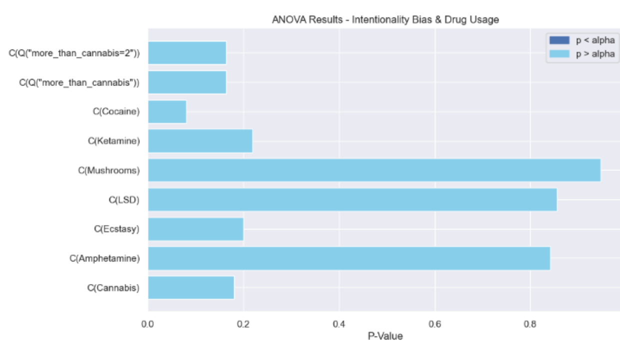

Figure 4.15 – Bar Plot 1 for ANOVA Results – Intentionality Bias

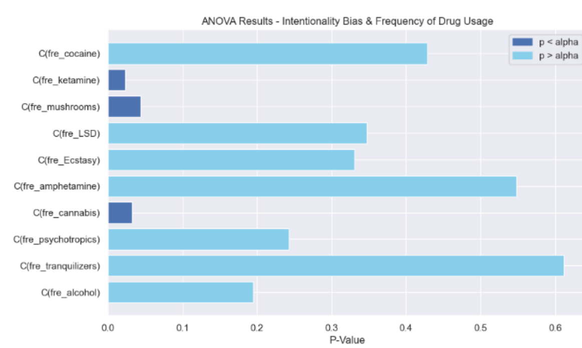

Figure 4.16 – Bar Plot 2 for ANOVA Results – Intentionality Bias

In this case, the results show that there is a significant effect frequency of ketamine, mushroom and cannabis usages have in people's intentionality bias scores (p value < alpha for ketamine usage). Hence, we can reject the null hypothesis for these factors and conclude that there is a significant relationship between the factors - frequency of ketamine usage, frequency of mushroom usage and frequency of cannabis usage and intentionality bias scores. However, the usage or the frequency of usage of any other drugs do not have any significant effect in people's intentionality bias scores, indicating no evidence against the null hypothesis. Hence, we can accept the null hypothesis for all the other factors/drugs tested.

4.4.3 Unpredictability

It is the quality of being impossible to know about or predict. The fact whether there is a relationship between unpredictability and drug usage has also been tested by implementing ANOVA tests. The results were,

	sum_sq	df	F	PR(>F)
C(Cannabis)	20.853008	1.0	0.829893	0.363691
C(Amphetamine)	28.984371	1.0	1.153499	0.284455
C(Ecstasy)	2.166229	1.0	0.086210	0.769437
C(LSD)	39.637983	1.0	1.577484	0.210977
C(Mushrooms)	3.662866	1.0	0.145772	0.703122
C(Ketamine)	17.718837	1.0	0.705162	0.402324
C(Cocaine)	2.786669	1.0	0.110902	0.739561
C(Q("more_than_cannabis"))	0.247966	1.0	0.009868	0.920994
C(Q("more_than_cannabis=2"))	0.247966	1.0	0.009868	0.920994
Residual	3970.119702	158.0	NaN	NaN

Figure 4.17 – ANOVA - Unpredictability & Drug Usage

	sum_sq	df	F	PR(>F)
C(fre_alcohol)	41.885684	5.0	0.342775	0.885954
C(fre_tranquilizers)	48.534715	4.0	0.496485	0.738332
C(fre_psychotropics)	46.235796	3.0	0.630625	0.596700
C(fre_cannabis)	190.653081	8.0	0.975142	0.459033
C(fre_amphetamine)	157.769917	5.0	1.291125	0.272564
C(fre_Ecstasy)	103.938837	6.0	0.708828	0.643127
C(fre_LSD)	122.494791	5.0	1.002447	0.419611
C(fre_mushrooms)	117.952040	4.0	1.206589	0.311873
C(fre_ketamine)	121.111837	6.0	0.825942	0.552119
C(fre_cocaine)	65.402131	5.0	0.535224	0.749217
Residual	2810.501860	115.0	NaN	NaN

Figure 4.18 – ANOVA - Unpredictability & Frequency of Drug Usage

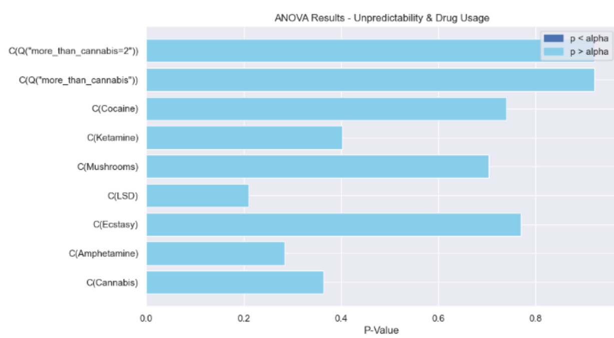

Figure 4.19 – Bar Plot 1 for ANOVA Results – Unpredictability

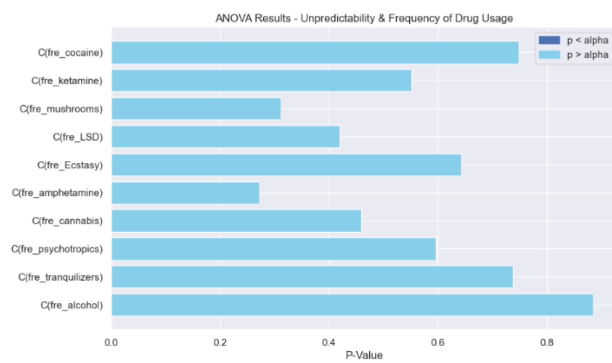

Figure 4.20 – Bar Plot 2 for ANOVA Results – Unpredictability

The results show that the usage and frequency of usage of any drugs do not have any significant relationship with people’s unpredictability scores as the p values are greater than alpha for all the variables(drugs) tested. Hence, the null hypothesis cannot be rejected for any drugs.

4.4.4 Free Will

Free will means the ability to act freely without getting affected by external factors. ANOVA was performed to identify whether there is a relationship between free will and drug usage. The results were,

	sum_sq	df	F	PR(>F)
C(Cannabis)	2.175209	1.0	0.180981	0.671110
C(Amphetamine)	29.218061	1.0	2.430996	0.120958
C(Ecstasy)	3.766336	1.0	0.313366	0.576415
C(LSD)	7.870993	1.0	0.654881	0.419590
C(Mushrooms)	7.202655	1.0	0.599274	0.440012
C(Ketamine)	0.136743	1.0	0.011377	0.915191
C(Cocaine)	25.108581	1.0	2.089079	0.150336
C(Q("more_than_cannabis"))	11.533502	1.0	0.959608	0.328783
C(Q("more_than_cannabis=2"))	11.533502	1.0	0.959608	0.328783
Residual	1898.997081	158.0	NaN	NaN

Figure 4.21 – ANOVA – Free Will & Drug Usage

	sum_sq	df	F	PR(>F)
C(fre_alcohol)	83.508182	5.0	1.546156	0.181040
C(fre_tranquilizers)	74.678765	4.0	1.728349	0.148473
C(fre_psychotropics)	93.609099	3.0	2.888625	0.038588
C(fre_cannabis)	67.024536	8.0	0.775601	0.624946
C(fre_amphetamine)	229.624282	5.0	4.251500	0.001389
C(fre_Ecstasy)	102.282003	6.0	1.578128	0.159715
C(fre_LSD)	18.482672	5.0	0.342207	0.886314
C(fre_mushrooms)	29.219728	4.0	0.676255	0.609774
C(fre_ketamine)	81.952530	6.0	1.264461	0.279277
C(fre_cocaine)	96.616903	5.0	1.788865	0.120565
Residual	1242.234183	115.0	NaN	NaN

Figure 4.22 – ANOVA – Free Will & Frequency of Drug Usage

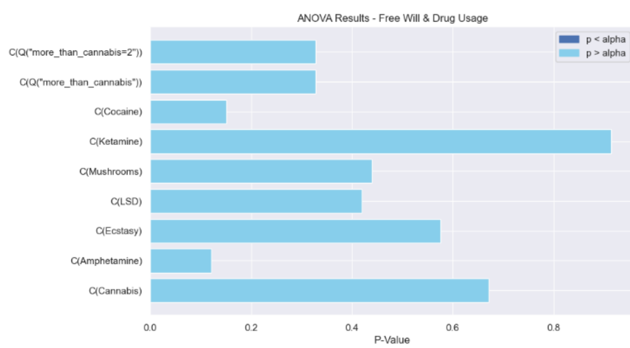

Figure 4.23 – Bar Plot 1 for ANOVA Results – Free Will

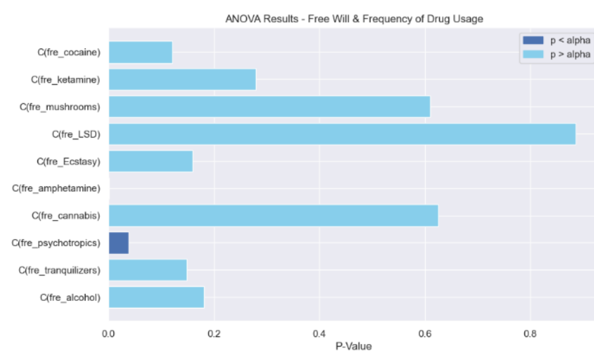

Figure 4.24 – Bar Plot 2 for ANOVA Results – Free Will

Only one p value is less than alpha that is of the frequency of usage of psychotropics. Hence, null hypothesis can be rejected for only this factor proving a significant relationship between the frequency of psychotropics usage and the free will scores. The usage or frequency of usage of any other drugs do not have any significant relationship with people's free will scores as the p values are greater than alpha for all these other variables. Hence, the null hypothesis cannot be rejected for other drugs.

4.5 IQ Level, Drug Usage and Personality Traits

In this section, it was explored whether the IQ level of a person is related to substance usage and the personality traits of said person.

The objective is divided into two parts. The first is to test the relationship with IQ Scores and Substance usage, the latter is to test the relationship between IQ scores and personality trait scores.

4.5.1 IQ Level & Drug Usage

The hypothesis for this objective,

- Null Hypothesis – Drug usage and frequency of drug usage have no significant relationship with IQ scores.
- Alternative Hypothesis – Drug usage and frequency of drug usage have a significant relationship with IQ scores.

To identify whether there is a relationship between IQ level and drug usage, an ANOVA test was performed.

The results for the ANOVA are,

	sum_sq	df	F	PR(>F)
C(Cannabis)	147.662504	1.0	0.610981	0.435588
C(Amphetamine)	270.864547	1.0	1.120753	0.291372
C(Ecstasy)	30.409266	1.0	0.125824	0.723275
C(LSD)	51.027113	1.0	0.211134	0.646511
C(Mushrooms)	489.873869	1.0	2.026945	0.156503
C(Ketamine)	5.998566	1.0	0.024820	0.875017
C(Cocaine)	46.669436	1.0	0.193104	0.660946
C(Q("more_than_cannabis"))	476.120504	1.0	1.970038	0.162406
C(Q("more_than_cannabis=2"))	476.120504	1.0	1.970038	0.162406
Residual	38185.583991	158.0	NaN	NaN

Figure 4.25 – ANOVA – IQ & Drug Usage

	sum_sq	df	F	PR(>F)
C(fre_alcohol)	869.570658	5.0	0.668069	0.648429
C(fre_tranquilizers)	655.791854	4.0	0.629786	0.642223
C(fre_psychotropics)	1062.364901	3.0	1.360314	0.258507
C(fre_cannabis)	1659.933836	8.0	0.797053	0.606312
C(fre_amphetamine)	614.826156	5.0	0.472356	0.796209
C(fre_Ecstasy)	478.176743	6.0	0.306143	0.932640
C(fre_LSD)	1859.324336	5.0	1.428472	0.219245
C(fre_mushrooms)	3283.814547	4.0	3.153591	0.016824
C(fre_ketamine)	787.453639	6.0	0.504151	0.804148
C(fre_cocaine)	269.323673	5.0	0.206915	0.958998
Residual	29937.198792	115.0	NaN	NaN

Figure 4.26 – ANOVA – IQ & Frequency of Drug Usage

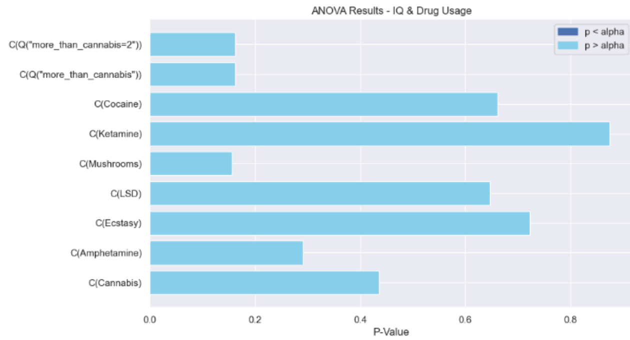

Figure 4.27 – Bar Plot 1 for ANOVA Results – IQ

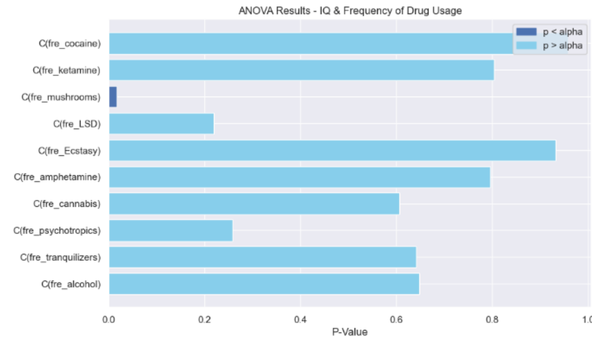

Figure 4.28 – Bar Plot 2 for ANOVA Results – IQ

As per the results, only one p value is less than alpha that is of the frequency of usage of mushrooms. Hence, null hypothesis can be rejected for only this factor proving a significant relationship between the frequency of mushrooms usage and the IQ scores. The usage or frequency of usage of any other drugs do not have any significant relationship with people's IQ scores as the p values are greater than alpha for all these other variables. Hence, the null hypothesis cannot be rejected for other drugs.

4.5.2 IQ Level & Personality Traits

The hypothesis for this objective,

- Null Hypothesis –Personality trait scores have no significant relationship with IQ scores.
- Alternative Hypothesis – Personality trait scores have a significant relationship with IQ scores.

To identify whether there is a relationship between IQ level and personality trait scores, multiple linear regression continued by ridge regression have been performed.

The results for the multiple linear regression are,

```
The Accuracy  on the training dataset is:  0.008028166553156635
The Accuracy  on the testing dataset is:  -0.003440296719137148
The RMSE  on the training dataset is:  0.2615635062258528
The RMSE  on the testing dataset is:  0.3236016989904633
The MAE  on the training dataset is:  0.2112399375836739
The MAE  on the testing dataset is:  0.2703447048954109
Coefficients: [ 0.00822228 -0.09582034  0.01291711 -0.02524624]
Intercept:  0.5835233415134617
```

Figure 4.29 – Results of Multiple Linear Regression

The results are very poor. The model only correctly predicts the target variable of about 0.8% of the time. The model complexity might need to be increased as the accuracy is very poor, and the model is likely not identifying the patterns correctly. The negative accuracy of -0.34% on the testing dataset raises concerns about the model's ability to generalize to unseen data. Also, the unusual negative value might be due to potential issues like overfitting. The low root mean squared error (RMSE) and mean squared error(MSE) suggests decent performance.

The picture below shows the correlation between IQ scores and each personality trait score. No correlation patterns could be seen in the pictures suggesting a very low / no correlation.

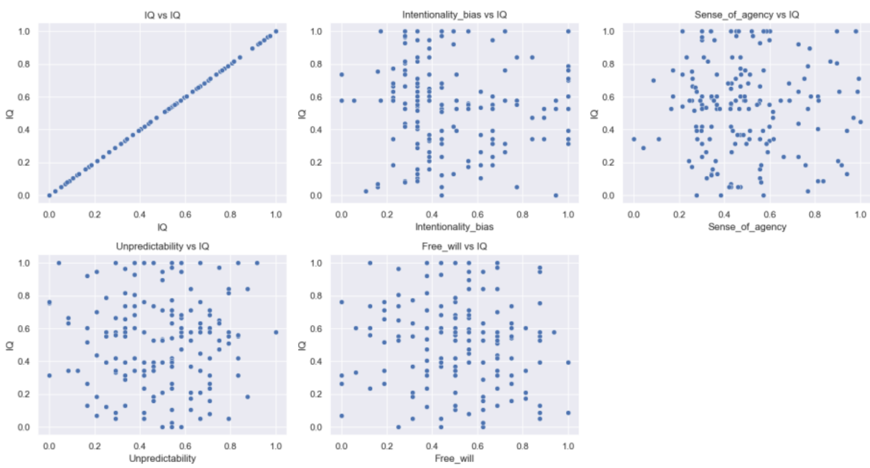

Figure 4.30 – Correlation maps between IQ & personality scores

The correlation heatmap below shows the correlation between IQ score and each personality trait score.

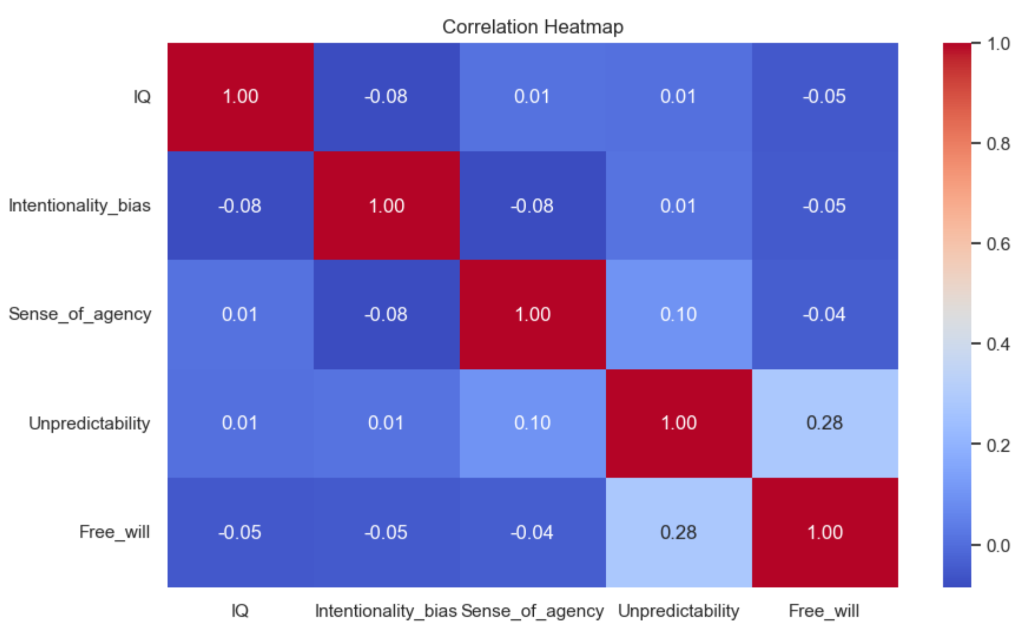

Figure 4.31 – Correlation heatmap between IQ & personality scores

As the correlation heatmap shows, the correlations between any personality score with IQ score is very poor.

Moving on from the multiple linear regression, to increase the model complexity, polynomial features to a degree of 5^{th} have been used. To avoid overfitting on the test data due to over complexity of the model, ridge regression has been used with polynomial features.

The results for the multiple linear regression with polynomial features,

```
The R^2 on the training dataset is: 0.8861418959275826
The R^2 on the testing dataset is: -54942.86397203198
The RMSE on the training dataset is: 0.0886154513666248
The RMSE on the testing dataset is: 75.7224441203341
The MAE on the training dataset is: 0.0627555731433848
The MAE on the testing dataset is: 29.4291402621904
```

Figure 4.32 – Results of Multiple Linear Regression with polynomial features

The very low negative test accuracy with a high train accuracy suggests overfitting. Hence, ridge regression has been used. The results are,

```
The R^2 on the training dataset is: 0.25585397050196634
The R^2 on the testing dataset is: 0.015688817339647776
The RMSE on the training dataset is: 0.22654608657388373
The RMSE on the testing dataset is: 0.3205023615103252
The MAE on the training dataset is: 0.17803695365243968
The MAE on the testing dataset is: 0.2711892933263969
```

Figure 4.33 – Results of Ridge Regression.

Here, we can finally see an increased positive test accuracy.

The model explains approximately 25.6% of the variance in the training data showing a moderate fit to the training set. However, the model’s explanatory power is low when tested on new data (1.6%). Considerably reasonable MSE and RMSE values are shown. However, there is not enough evidence to reject the null hypothesis based on the low R2 scores. Hence, we accept the null hypothesis and conclude that there is no significant relationship between personality trait scores and IQ scores (Personality trait scores do not affect IQ scores.).

4.6 Most Exposed Drugs and People Most Exposed To.

In this section, I have used graphical representations to differentiate between the usage of different drugs by which I could rank the tested drugs from most used to least used. The results are,

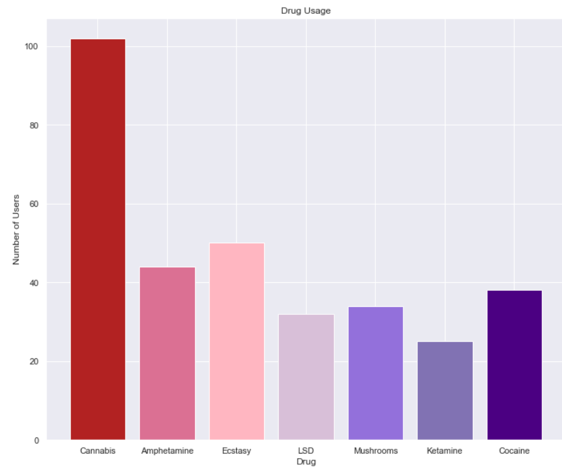

Figure 4.34 – Drug Users

According to the graph above, the drug that most people have used or are using is cannabis. After this, Ecstasy and amphetamine are used by a higher number of people as well. The drug used by the least amount of people is ketamine.

After this, I have further divided the usage of each drug according to gender. The results are,

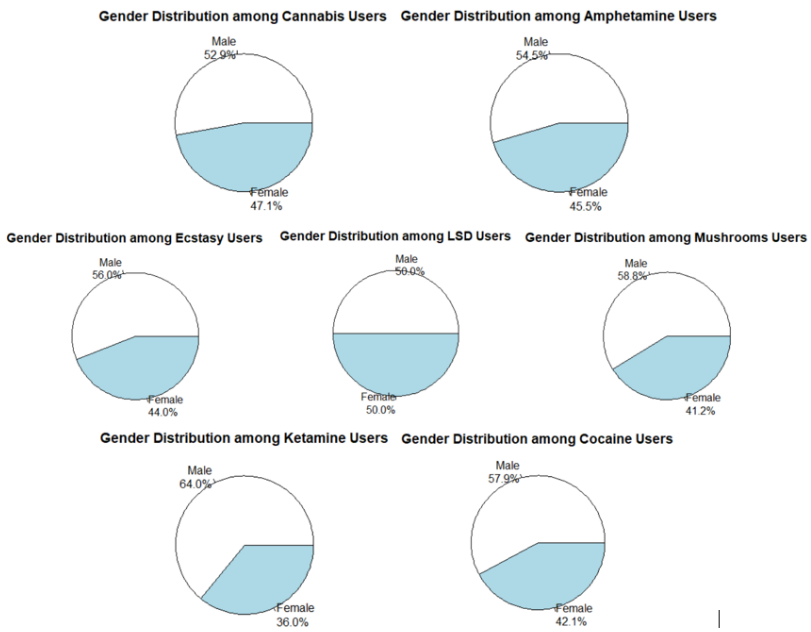

Figure 4.35 – Drug Usage & Gender Classification

According to the pie charts, it could be seen that all the drugs except LSD are used by a higher percentage of men when compared to women. LSD is used by an equal percentage of men and women(50-50). The drug that has a higher difference between percentage of male and female users is ketamine which has 64 percent of male users and only 36 percent of female users.

I have also checked the different frequency levels of the drugs. Here every drug is divided by the number of people with different frequencies of usage. The users are further divided by gender as well. The graphs

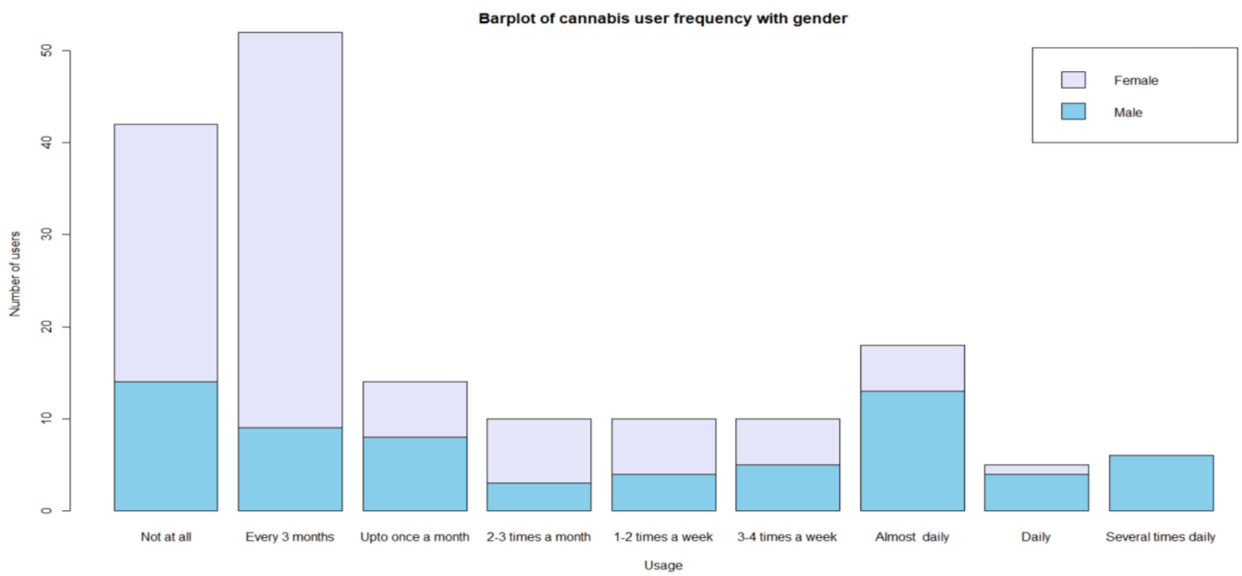

Figure 4.36 – Cannabis User Frequency

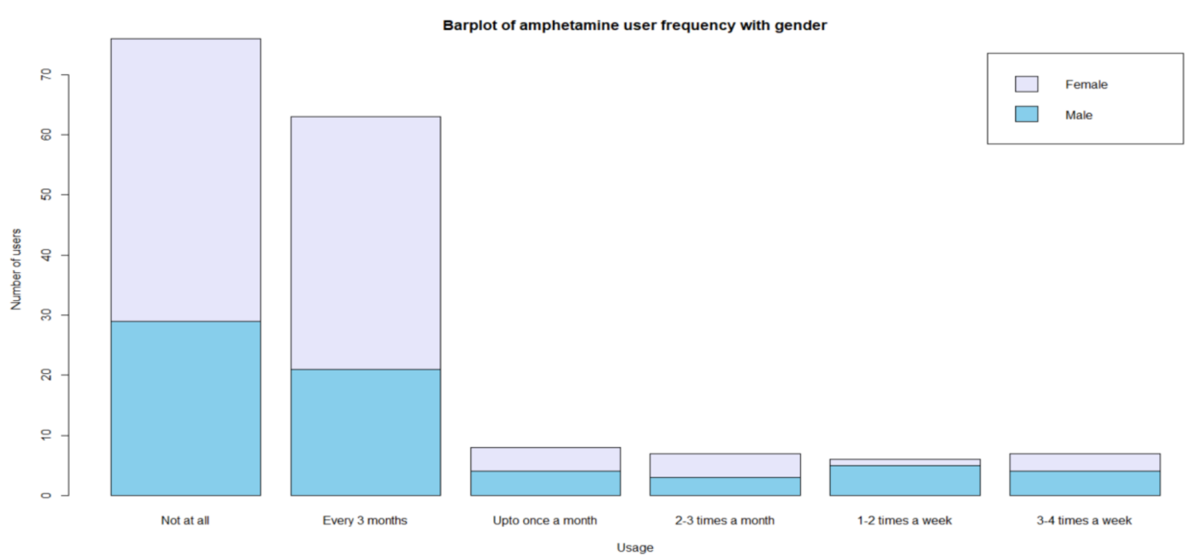

Figure 4.37 – Amphetamine User Frequency

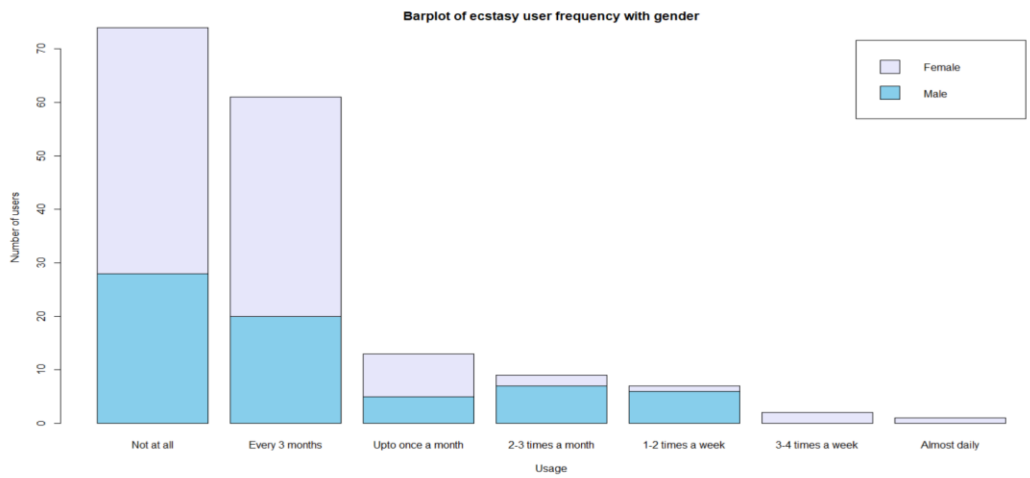

Figure 4.38 – Ecstasy User Frequency

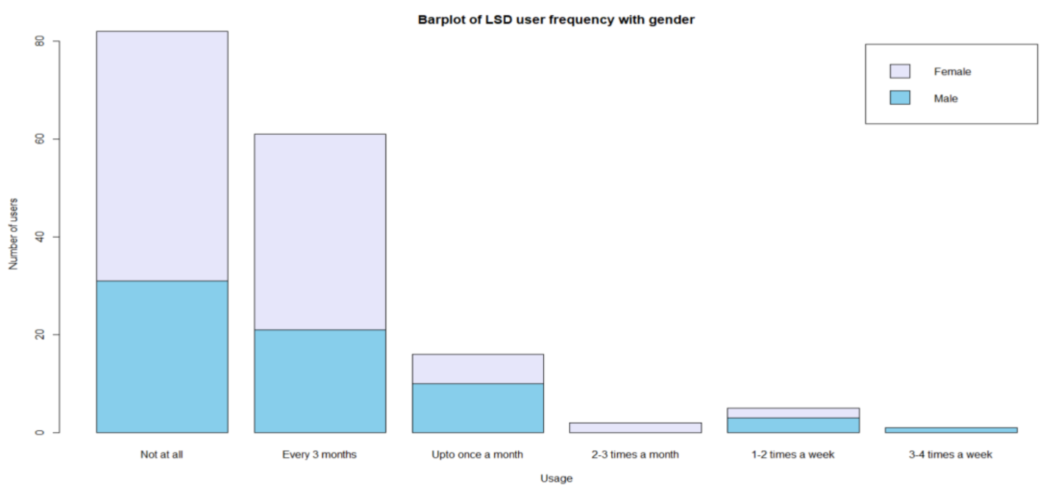

Figure 4.39 – LSD User Frequency

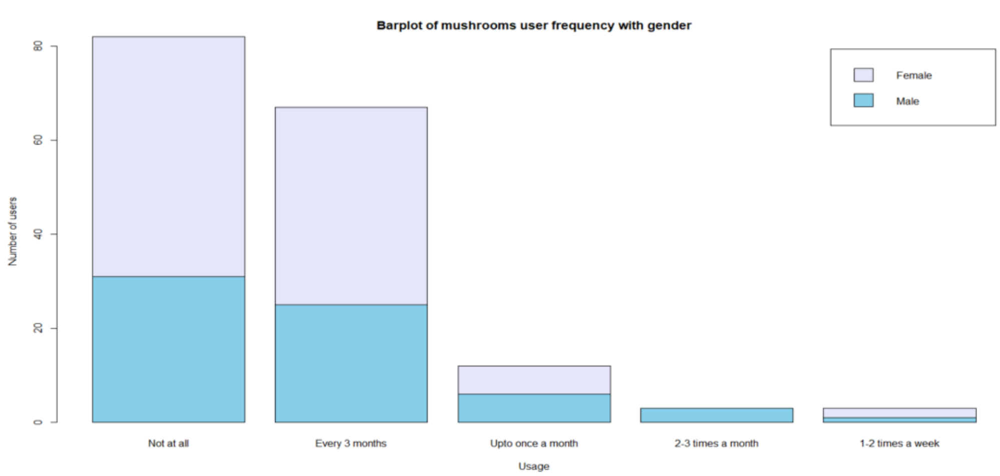

Figure 4.40 – Mushrooms User Frequency

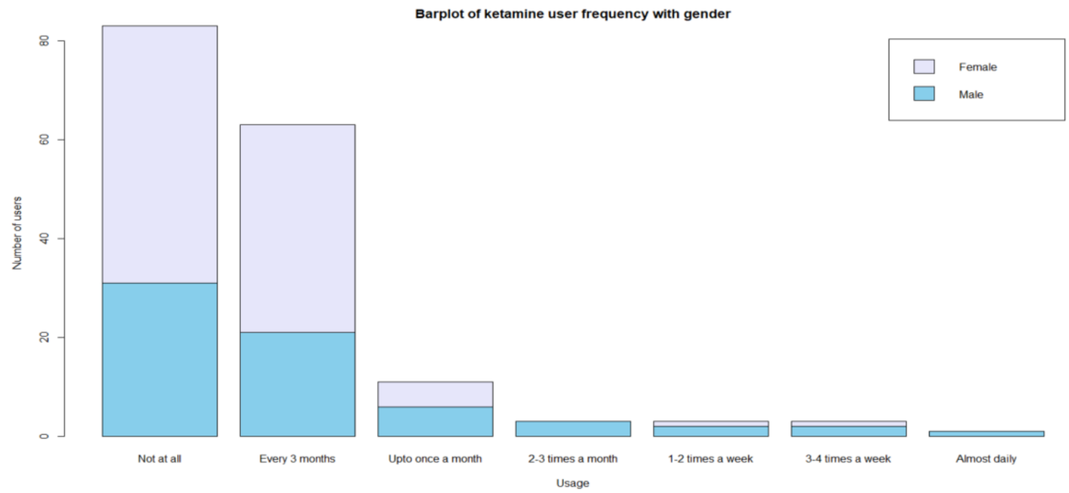

Figure 4.41 – Ketamine User Frequency

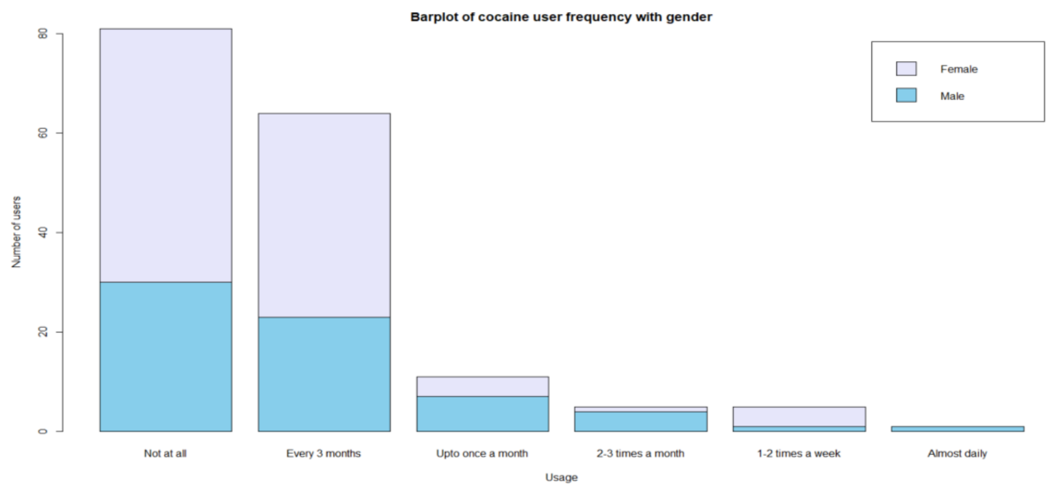

Figure 4.42 – Cocaine User Frequency

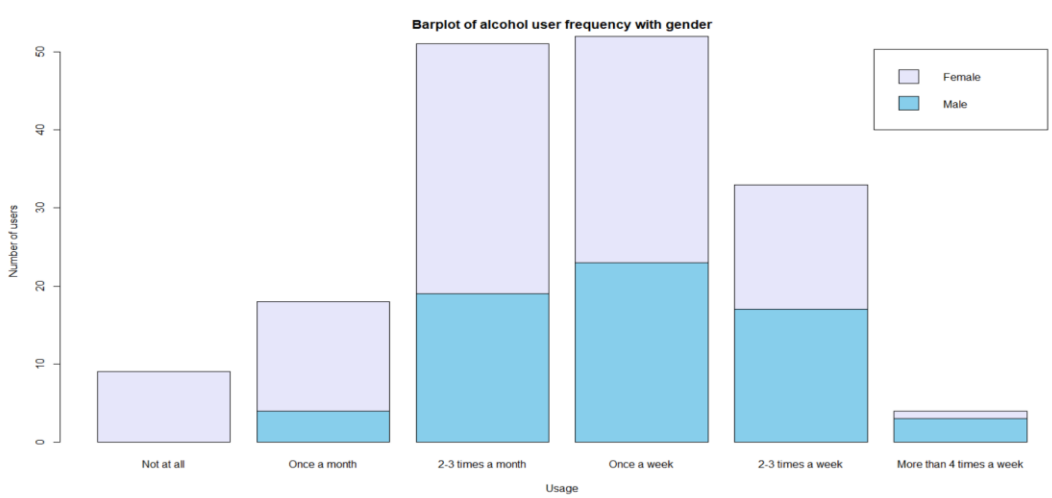

Figure 4.43 – Alcohol User Frequency

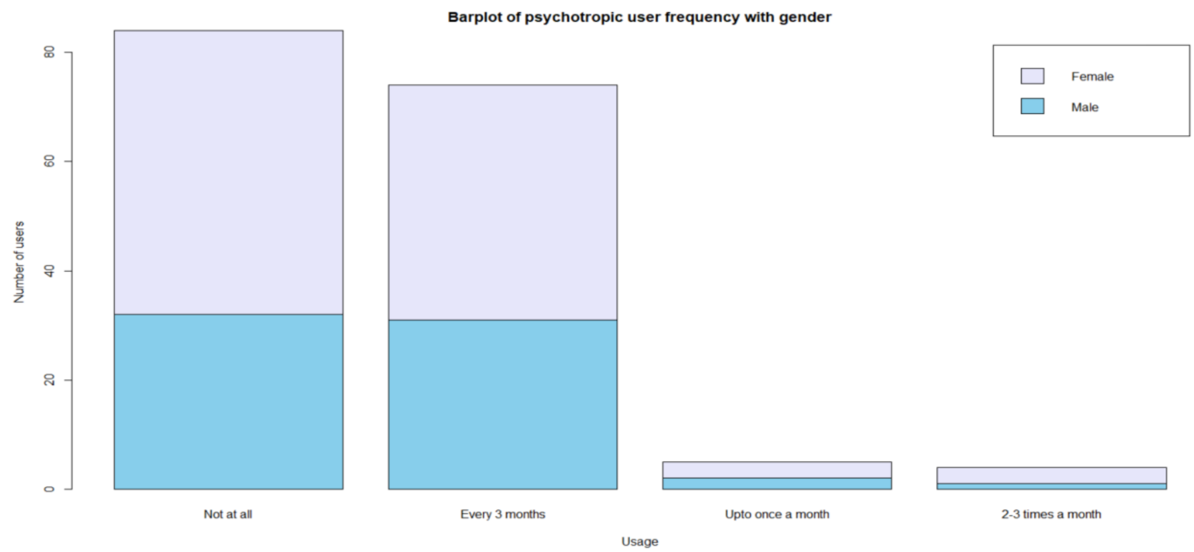

Figure 4.44 – Psychotropics User Frequency

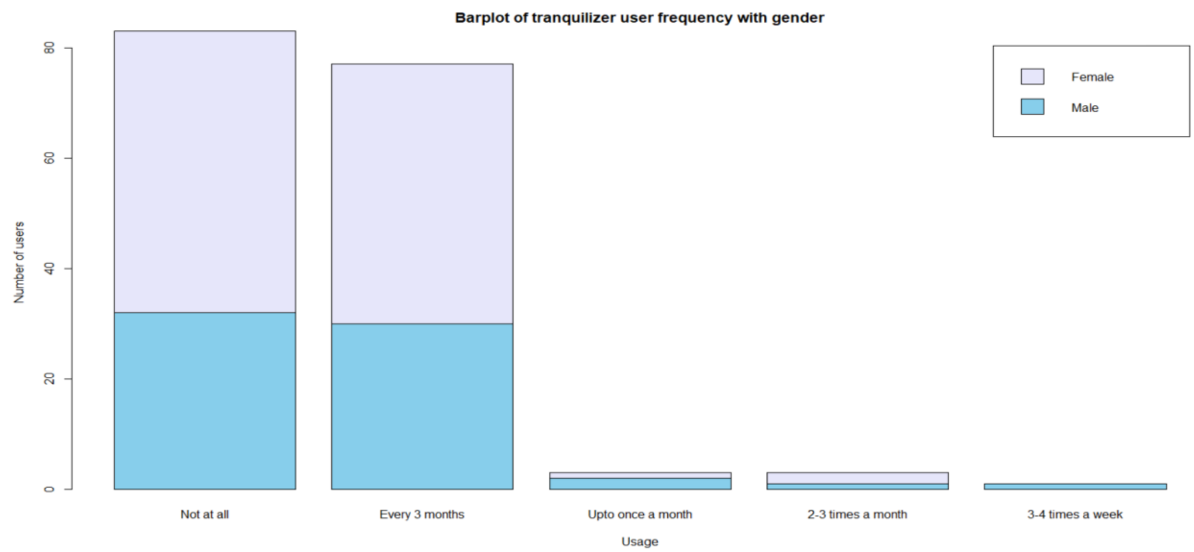

Figure 4.45 – Tranquilizer User Frequency

4.7 Dependencies Between Various Drugs.

In this section, I have used a chi-squared test to check whether the usage of one drug improves or enhances the chances of the usage of another drug.

- Null Hypothesis - Usage and/or frequency of usage of one drug does not affect the usage and/or frequency of usage of another drug.
- Alternative Hypothesis - Usage and/or frequency of usage of one drug does affect the usage and/or frequency of usage of another drug.

The top used three drugs have been selected that are cannabis, ecstasy, and amphetamine. After that, their dependencies with every other drug have been analyzed. The results are,

For Cannabis,

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Cannabis, Ecstasy)
X-squared = 43.172, df = 1, p-value = 5.012e-11


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Cannabis, Amphetamine)
X-squared = 35.879, df = 1, p-value = 2.1e-09


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Cannabis, LSD)
X-squared = 23.241, df = 1, p-value = 1.429e-06


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Cannabis, Mushrooms)
X-squared = 25.189, df = 1, p-value = 5.199e-07


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Cannabis, Ketamine)
X-squared = 16.86, df = 1, p-value = 4.024e-05


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Cannabis, Cocaine)
X-squared = 29.265, df = 1, p-value = 6.312e-08
```

Figure 4.46 – Chi-Squared Test for Cannabis

In each case, the p-value is very small (less than 0.05), indicating strong evidence against the null hypothesis. Therefore, we can conclude that there is a significant association between Cannabis use and each of the other drugs (Ecstasy, Amphetamine, LSD, Mushrooms, Ketamine, Cocaine) under consideration. Hence, cannabis use influences the use of all the other drugs.

For Ecstasy,

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Ecstasy, Cannabis)
X-squared = 43.172, df = 1, p-value = 5.012e-11


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Ecstasy, Amphetamine)
X-squared = 94.358, df = 1, p-value < 2.2e-16


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Ecstasy, LSD)
X-squared = 52.755, df = 1, p-value = 3.778e-13


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Ecstasy, Mushrooms)
X-squared = 52.817, df = 1, p-value = 3.66e-13


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Ecstasy, Ketamine)
X-squared = 50.561, df = 1, p-value = 1.155e-12


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Ecstasy, Cocaine)
X-squared = 79.487, df = 1, p-value < 2.2e-16
```

Figure 4.47 – Chi-Squared Test for Ecstasy

Again, in each case, the p-value is very small (less than 0.05), indicating strong evidence against the null hypothesis. Therefore, we can conclude that there is a significant association between Ecstasy use and each of the other drugs (Cannabis, Amphetamine, LSD, Mushrooms, Ketamine, Cocaine) under consideration. Hence, use of ecstasy influences the use of all the other drugs.

For Amphetamine,

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Amphetamine, Ecstasy)
X-squared = 94.358, df = 1, p-value < 2.2e-16


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Amphetamine, Cannabis)
X-squared = 35.879, df = 1, p-value = 2.1e-09


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Amphetamine, LSD)
X-squared = 58.039, df = 1, p-value = 2.57e-14


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Amphetamine, Mushrooms)
X-squared = 58.562, df = 1, p-value = 1.97e-14


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Amphetamine, Ketamine)
X-squared = 77.787, df = 1, p-value < 2.2e-16


        Pearson's Chi-squared test with Yates' continuity correction

data:  table(Amphetamine, Cocaine)
X-squared = 88.781, df = 1, p-value < 2.2e-16
```

Figure 4.48 – Chi-Squared Test for Amphetamine

In this case also, the p-value is very small (less than 0.05) in each case, indicating strong evidence against the null hypothesis. Therefore, we can conclude that there is a significant association between amphetamine use and each of the other drugs (Ecstasy, cannabis, LSD, Mushrooms, Ketamine, Cocaine) under consideration. Hence, amphetamine use influences the use of all the other drugs.

In conclusion, there is strong evidence to reject the null hypothesis and accept the alternative hypothesis that the usage of one drug does affect the usage of another drug. All the highest used three drugs influence the usage of other drugs.

4.8 Impacts of the Drugs on Different People

In this section, I have tested whether the impacts on drug usage are different from one type of people to another type of people.

- Null Hypothesis – People’s narcissism score, their age and IQ level do not impact risk they have of using drugs.
- Alternative Hypothesis - People’s narcissism score, their age and IQ level impact risk they have of using drugs.

For this case, logistic regression has been used where independent variables are IQ score, narcissism score and age and the dependent variable is drug usage (whether people use the said drug or not). Logistic regressions have been performed for each drug separately. The results are shown and explained below,

Cannabis:

```
Metrics for Cannabis
Precision: 0.4615
Recall: 0.8571
F1-Score: 0.6000
Specificity: 0.5000
AUC-ROC: 0.7143
Accuracy: 0.6190

P-values for Cannabis
Constant: 0.5356
Age: 0.0005
IQ: 0.0415
Narcissism: 0.6685
Optimization terminated successfully.
         Current function value: 0.621778
         Iterations 5
```

Figure 4.49 – Effects on People by Cannabis

Cannabis shows a moderate precision and high recall, indicating that the model is better at identifying individuals who use cannabis than those who do not. The model considers Age and IQ as significant predictors, while Narcissism does not significantly contribute.

The results show that age and IQ significantly influence the likelihood of cannabis use as their p values are less than alpha (p value < 0.05). However, narcissism does not significantly contribute to the likelihood of the person being a cannabis user. The model has moderate performance with a relatively high recall, indicating its ability to people using cannabis but a lower precision indicating struggles in identifying the non-users. Also, the lower specificity shows the challenges in identifying non-users.

LSD:

```
Metrics for LSD
Precision: 0.8182
Recall: 0.6000
F1-Score: 0.6923
Specificity: 0.8333
AUC-ROC: 0.8111
Accuracy: 0.7037

P-values for LSD
Constant: 0.8305
Age: 0.0
IQ: 0.0042
Narcissism: 0.2397
Optimization terminated successfully.
         Current function value: 0.613213
         Iterations 5
```

Figure 4.50 – Effects on People by LSD

The results show that age and IQ have significant relationships with LSD usage ($p < 0.05$) , while Narcissism does not ($p > 0.05$). The comparatively high precision and moderate recall suggests a good balance between correctly identifying users and avoiding the non-users. Considerably high AUC-ROC (81%) and accuracy (70%) shows that the model has a good overall performance.

Cocaine:

```
Metrics for Cocaine
Precision: 1.0000
Recall: 0.6429
F1-Score: 0.7826
Specificity: 1.0000
AUC-ROC: 0.9167
Accuracy: 0.8077

P-values for Cocaine
Constant: 0.9353
Age: 0.0
IQ: 0.6959
Narcissism: 0.4177
Optimization terminated successfully.
         Current function value: 0.564876
         Iterations 5
```

Figure 4.51 – Effects on People by Cocaine

The model suggests that age has a significant relationship with the LSD usage (p value < 0.05). However, the model suggests that IQ and narcissism do not have a significant relationship with

LSD usage (p value > 0.05). The perfect precision, considerably good recall, a good accuracy of 80.7%, and AUC-ROC score of 0.9167 shows really good performance of the overall model.

Ecstasy:

```
Metrics for Ecstasy
Precision: 0.9167
Recall: 0.6875
F1-Score: 0.7857
Specificity: 0.8750
AUC-ROC: 0.8203
Accuracy: 0.7500

P-values for Ecstasy
Constant: 0.6261
Age: 0.0
IQ: 0.1668
Narcissism: 0.0675
Optimization terminated successfully.
         Current function value: 0.597238
         Iterations 5
```

Figure 4.52 – Effects on People by Ecstasy

According to the results, age is a significant predictor of Ecstasy, while IQ and narcissism do not significantly contribute to predicting narcissism. As indicated by the high precision, the model correctly identifies the true positives among predicted positives. The F1 score gives a balanced tradeoff between precision and recall. A high specificity shows the model's ability to correctly identify true negatives. A high AUC-ROC score shows good discriminatory power and there is a considerably good overall accuracy of 75%.

Ketamine:

```
Metrics for Ketamine
Precision: 0.8000
Recall: 0.6667
F1-Score: 0.7273
Specificity: 0.7273
AUC-ROC: 0.8586
Accuracy: 0.6897

P-values for Ketamine
Constant: 0.6583
Age: 0.0
IQ: 0.2407
Narcissism: 0.3626
Optimization terminated successfully.
         Current function value: 0.582155
         Iterations 5
```

Figure 4.53 – Effects on People by Ketamine

According to the p values, age has a significant relationship with ketamine usage meanwhile IQ and narcissism do not. With a precision of 80%, the model accurately identifies a significant proportion of true positives among the predicted positives. Specificity shows the model correctly identifies true negatives around 73% of the time. A strong discriminatory power can be seen from the 85.86% AUC-ROC score. However, the model has a limited overall accuracy of 68.97%.

Amphetamine:

```
Metrics for Amphetamine
Precision: 0.7778
Recall: 0.4667
F1-Score: 0.5833
Specificity: 0.8000
AUC-ROC: 0.6200
Accuracy: 0.6000

P-values for Amphetamine
Constant: 0.7911
Age: 0.0
IQ: 0.6317
Narcissism: 0.017
Optimization terminated successfully.
         Current function value: 0.601169
         Iterations 5
```

Figure 4.54 – Effects on People by Amphetamine

According to the p values, age and narcissism significantly influence Amphetamine usage (p values < 0.05) meanwhile IQ does not (p value > 0.05). AUC-ROC score of 62% shows moderate discriminatory power and there is a moderate overall accuracy of%.

Mushrooms:

```
Metrics for Mushrooms
Precision: 0.7273
Recall: 0.6154
F1-Score: 0.6667
Specificity: 0.7857
AUC-ROC: 0.6703
Accuracy: 0.7037

P-values for Mushrooms
Constant: 0.9677
Age: 0.0
IQ: 0.0002
Narcissism: 0.0836
```

Figure 4.55 – Effects on People by Mushrooms

Age and IQ strongly influence mushroom usage as indicated by very low p values. However, narcissism cannot be concluded to significantly influence mushroom usage as its p value is greater than alpha. Model has a moderate discriminatory power as addressed by the 67 percent of AUC-ROC score. There is a reasonable overall accuracy of 70 percent. With a precision of 72.73%, the model accurately identifies a substantial proportion of true positives among predicted positives. A recall of 61.54% indicates the model's ability to capture positive cases, though with some trade-offs in precision.

In summary, age has a significant relationship with the usage of all the drugs. In other words, age influences drug usage that being said for all the drugs tested. IQ score significantly influences the usage of cannabis, LSD and mushroom meanwhile narcissism has a significant relationship with only amphetamine usage.

CHAPTER 5
DISCUSSION AND RECOMMENDATIONS

5.1 Discussion

This research report explores and analyzes substance abuse, especially in youngsters, investigating and differentiating between different drugs and the effects they impose with the focus on personality traits and intelligence quotient. It analyses whether people use various drugs and if yes, the frequency of their usage. It explores whether these substance usages impact a person's personality traits and IQ scores. It also explores whether drugs have different effects on different people. The inter-dependencies of one drug to another, as well as the usage pattern of different drugs based on gender, age and frequency levels have also been examined and visualized.

Considering drug usage and personality traits, the following patterns are identified. Usage of ketamine has been found to have a significant relationship between a person's sense of agency scores. The frequency of usage of ketamine, mushrooms and cannabis showed to have a significant effect on people's intentionality bias scores. None of the drugs tested either the usage or the frequency of usage have been found to impact a person's unpredictability scores. And finally, the frequency of usage of psychotropics has been found to affect free will scores.

Moving on to IQ level and drug usage, only the frequency of usage of mushrooms has been found to have a significant relationship with IQ scores of people. Considering IQ level and personality trait scores, the scores of no personality traits tested (sense of agency, intentionality bias, unpredictability, and free will) has been found to affect IQ scores. Hence, it has been concluded that there is no significant relationship between personality trait scores and IQ scores.

The drugs used by the highest number of people is cannabis and the drug used by the lowest number of people is ketamine. The most used three drugs are cannabis, ecstasy, and amphetamine, respectively. After this, cocaine, mushrooms, and LSD are used in that specific order. All the drugs except for LSD are used by a higher percentage of men more than women while men and women use LSD equally. The drug that has a higher difference between percentage of male and female users is ketamine. The frequency levels and gender divisions for all the above mentioned drugs have been visualized using graphs as well.

While checking the dependancies between one drug to another for the top three most used drugs, it has been found that all the three drugs(cannabis, ecstasy and amphetamine) influence the usage of all the other drugs examined. Hence, a strong evidencehas been found that a person who uses one drug is more prone to try another drug as well.

While examining the impacts of the drugs on different people, for all the drugs tested, age has been a factor that affects the usage of them. IQ score significantly influences the usage of cannabis, LSD and mushroom meaning people with different IQ scores tend to have been influenced to use the said drugs differently. Lastly the narcissism scores of people have been found to have a significant relationship with only amphetamine usage.

The limitations of this research would be that the data used is secondary, and the questionnaires used to get data, even though anonymous could not be concluded to have perfect reliability. It is also worth noticing that it could not be shown how and why exactly different drugs have different impacts on different types of people. Nonetheless, this research contributes to understanding the effects of drug usage on young individuals and highlights the importance of addressing substance abuse in society.

5.2 Recommendations

As the research shows strong evidence about the higher risk of people who already consume or have consumed one drug, to consume or try another drug, prevention measures could be focused on people who have a history of drug usage.

Awareness campaigns: Given the high prevalence of substance abuse among young adults, it is recommended that awareness campaigns be conducted to educate young adults on the dangers of substance abuse. As our study shows, almost all the drugs are used my men more than women, proper increased guidance to male students while not neglecting female students as well could be helpful. More focus could be directed towards the highest used drugs.

As it was shown that people of different ages have different impacts caused by substance usage, relevant awareness could be spread. There is a gap for future research to focus more on the divisions of these people and why they have different impacts. To address the limitations in this research, data from different regions around the world will be helpful in identifying geographical patterns as well. The sample size could also be increased to include diverse people.

As most of our participants(subject matters) are students, students at high school and university could be guided and provided with awareness as to the risks they are prone to face because of growing substance usage. Proper counseling options in academic places could be helpful as well.

Collaboration between stakeholders such as educators, healthcare providers, parents and law enforcement agencies are essential in the fight against substance abuse. A concerted effort should be made to develop and implement effective strategies to combat substance abuse among young adults.

5.3 Conclusions

In conclusion, this research delves into the complex interplay between substance abuse, personality traits, and intelligence quotient, particularly among young individuals. The findings reveal distinctive patterns associated with various drugs, shedding light on their impact on personality traits and IQ scores. Notably, the study identifies significant relationships, such as ketamine influencing sense of agency scores and the frequency of psychotropic usage affecting free will scores and the frequency of usage of ketamine, mushrooms and cannabis influencing people’s intentionality bias scores.

The exploration of drug usage and its correlation with intelligence quotient exposes a noteworthy association, specifically with the frequency of mushroom usage. However, no substantial link between personality trait scores and IQ scores is identified. Cannabis emerges as the most widely used drug, with ketamine being the least used, and gender differences in drug usage are evident, that all the drugs tested except LSD are used by men more than women whereas LSD is used equally between the genders.

Examining dependencies between top-used drugs demonstrates a strong likelihood that individuals experimenting with one drug are prone to trying others. Age emerges as a significant factor influencing drug usage across all substances studied, and intelligence quotient plays a role in shaping the patterns of cannabis, LSD, and mushroom usage. Additionally, a unique connection is established between narcissism scores and amphetamine usage.

Despite the insightful findings, the research acknowledges its limitations, such as reliance on secondary data and potential questionnaire reliability concerns. The study highlights the necessity of addressing substance abuse in society and contributes valuable insights into understanding the nuanced effects of drug usage on young individuals.

APPENDICES

Importing data, libraries, data preprocessing, descriptive statistics

```
# Packages / libraries
import os #provides functions for interacting with the operating system
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, explained_variance_score, mean_absolute_error, mean_squared_error
from math import sqrt

%matplotlib inline


# Increases the size of sns plots
sns.set(rc={'figure.figsize':(12,10)})
```

```
data = pd.read_csv('data_senseofagency.csv')
data
```

```
data = data.drop(0)
data
```

```
data.columns
```

```
data = data[['VPN', 'Sex', 'Age', 'IQ', 'Intentionality_bias','score_NA', 'score_FW_FW','score_FW_U', 'Cannabis',
             'Amphetamin', 'Ecstasy', 'LSD', 'mushrooms', 'Ketamine','cocaine', 'more_than_cannabis', 'fre_alcohol',
             'fre_tranquilizers', 'fre_psychotrops', 'fre_Cannabis','fre_Amphetamin', 'fre_Ecstasy',
             'fre_LSD', 'fre_mushrooms','fre_Ketamine', 'fre_cocaine', 'diff_binding', 'filter_$']]
data
```

```
data1 = data.rename(columns ={'VPN':'Participant',
                              'score_NA':'Narcissism',
                              'score_FW_FW':'Free_will',
                              'score_FW_U':'Unpredictability',
                              'Amphetamin':'Amphetamine',
                              'mushrooms':'Mushrooms',
                              'cocaine':'Cocaine',
                              'fre_Cannabis':'fre_cannabis',
                              'fre_psychotrops':'fre_psychotropics',
                              'fre_Amphetamin':'fre_amphetamine',
                              'fre_Ectsasy':'fre_ecstasy',
                              'fre_Ketamine':'fre_ketamine',
                              'diff_binding':'Sense_of_agency',
                              'filter_$':'more_than_cannabis=2'
                              })
data1
```

```
data1.columns
```

```
data1.isnull().sum()
```

```
data1 = data1.dropna()
data1
```

```
data1.boxplot(column=['Age','Narcissism','Free_will','Unpredictability'])
data1.boxplot(column=['Free_will'])
data1.boxplot(column=['Sense_of_agency','IQ'])
data1.boxplot(column=['Intentionality_bias'])
```

```
data2 = data1[(data['Age'] > 10) &
              (data['Unpredictability'] > 10) &
              (data['Intentionality_bias'] > 0.2) &
              (data['Free_will'].between(15, 31)) &
              (data['Sense_of_agency'].between(-240, 345)) &
              (data['IQ']>1)]
data2
```

```
data2.boxplot(column=['Age','Narcissism','Free_will','Unpredictability'])
data2.boxplot(column=['Intentionality_bias'])
data2.boxplot(column=['Sense_of_agency','IQ'])
```

```
df = data2.reindex(columns=['Participant', 'Sex', 'Age', 'IQ','Sense_of_agency','Intentionality_bias','Free_will',
                            'Unpredictability','Narcissism','Cannabis', 'Amphetamine', 'Ecstasy','LSD', 'Mushrooms', 'Ketamine',
                            'Cocaine', 'more_than_cannabis','more_than_cannabis=2','fre_alcohol', 'fre_tranquilizers',
                            'fre_psychotropics','fre_cannabis', 'fre_amphetamine', 'fre_Ecstasy', 'fre_LSD',
                            'fre_mushrooms', 'fre_ketamine', 'fre_cocaine'])
```

```
data = df.to_csv('drug_and_personality.csv', index=False)
```

```
data = pd.read_csv('drug_and_personality.csv')
data
```

```
data[['Age', 'IQ', 'Sense_of_agency',
       'Intentionality_bias', 'Free_will', 'Unpredictability', 'Narcissism']].describe()
```

```
data.replace({1: 'No', 2: 'Yes'}, inplace=True)
g = sns.pairplot(data[['Intentionality_bias','Sense_of_agency','Free_will','Unpredictability','IQ','Sex']],
                 hue = 'Sex', height = 5)
```

ANOVA Tests for personality traits

```
# Performing ANOVA
import statsmodels.api as sm
from statsmodels.formula.api import ols

formula =
'Intentionality_bias ~ C(Cannabis) + C(Amphetamine) + C(Ecstasy) + C(LSD) + C(Mushrooms) + C(Ketamine) + C(Cocaine) + C(Q("more_t

model = ols(formula ,data=data).fit()
anova_results = sm.stats.anova_lm(model, typ=2)
anova_results
```

```
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches

p_values = anova_results['PR(>F)']
variables = p_values.index

alpha = 0.05
colors = ['b' if p < alpha else 'skyblue' for p in p_values]

plt.figure(figsize=(10, 6))
plt.barh(variables, p_values, color=colors)
plt.xlabel('P-Value')
plt.title('ANOVA Results - Intentionality Bias & Drug Usage')
plt.legend(handles=[mpatches.Patch(color='b', label='p < alpha'),
                    mpatches.Patch(color='skyblue', label='p > alpha')],
           loc='upper right')

plt.show()
```

Same format has been followed for all the drugs

IQ & Drug Usage

```
# Performing ANOVA
import statsmodels.api as sm
from statsmodels.formula.api import ols

formula = 'IQ ~ C(Cannabis) + C(Amphetamine) + C(Ecstasy) + C(LSD) + C(Mushrooms) + C(Ketamine) + C(Cocaine) + C(Q("more_than_car
model = ols(formula, data=data).fit()

anova_results = sm.stats.anova_lm(model, typ=2)
anova_results
```

```
# Visualizing the results with a bar plot
p_values = anova_results['PR(>F)']
variables = p_values.index

alpha = 0.05
colors = ['b' if p < alpha else 'skyblue' for p in p_values]

plt.figure(figsize=(10, 6))
plt.barh(variables, p_values, color=colors)
plt.xlabel('P-Value')
plt.title('ANOVA Results - IQ & Drug Usage')
plt.legend(handles=[mpatches.Patch(color='b', label='p < alpha'),
                    mpatches.Patch(color='skyblue', label='p > alpha')],
           loc='upper right')

plt.show()
```

IQ & Personality Traits

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

numeric_data = data.select_dtypes(include=['float64', 'int64'])

scaler = MinMaxScaler()
normalized_data = scaler.fit_transform(numeric_data)
normalized_df = pd.DataFrame(normalized_data, columns=numeric_data.columns)
```

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

numeric_data = data.select_dtypes(include=['float64', 'int64'])

scaler = MinMaxScaler()
normalized_data = scaler.fit_transform(numeric_data)
normalized_df = pd.DataFrame(normalized_data, columns=numeric_data.columns)

data1 = normalized_df[['IQ','Sense_of_agency','Intentionality_bias','Free_will','Unpredictability']]

X = data1.drop('IQ', axis = 1).values
Y = data1['IQ']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, train_size=0.80, test_size = 0.20, random_state=42)

#Fitting the linear regression model
lm = LinearRegression(fit_intercept = True)
lm.fit(X_train, Y_train)
Y_pred = lm.predict(X_train)

print('The Accuracy  on the training dataset is: ',
      lm.score(X_train, Y_train) )
#print('The Accuracy n2  on the training dataset is: ',
#      r2_score(Y_train,Y_pred) )

# Model Accuracy on testing dataset
print('The Accuracy  on the testing dataset is: ',
      lm.score(X_test, Y_test) )
```

```python
# The Root Mean Squared Error (RMSE)
print('The RMSE  on the training dataset is: ',
      sqrt(mean_squared_error(Y_train,Y_pred)))
print('The RMSE  on the testing dataset is: ',
      sqrt(mean_squared_error(Y_test,lm.predict(X_test))))

# The Mean Absolute Error (MAE)
print('The MAE  on the training dataset is: ',
      mean_absolute_error(Y_train,Y_pred))
print('The MAE  on the testing dataset is: ',
      mean_absolute_error(Y_test,lm.predict(X_test)))

# Coefficients
print('Coefficients:', lm.coef_ )

# The Intercept
print('Intercept: ', lm.intercept_)
```

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures, StandardScaler
from sklearn.linear_model import Ridge
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
from math import sqrt
from sklearn.model_selection import cross_val_score

# Assuming you have your original data loaded and normalized_df as in your original code

# Select relevant features
data1 = normalized_df[['IQ', 'Intentionality_bias', 'Sense_of_agency', 'Unpredictability', 'Free_will']]

# Features and target variable
X = data1.drop('IQ', axis=1)
Y = data1['IQ']

# Polynomial Features
poly = PolynomialFeatures(degree=5)
X_poly = poly.fit_transform(X)

# Feature Scaling
scaler = StandardScaler()
X_poly_scaled = scaler.fit_transform(X_poly)

# Split the dataset into training and testing sets
X_train_poly, X_test_poly, Y_train, Y_test = train_test_split(X_poly_scaled, Y, train_size=0.80, test_size=0.20,
                                                              random_state=42)

# Ridge Regression
ridge = Ridge(alpha=1.0)
ridge.fit(X_train_poly, Y_train)

# Model Accuracy on training dataset
Y_pred_ridge_train = ridge.predict(X_train_poly)
print('The R^2 on the training dataset is:', r2_score(Y_train, Y_pred_ridge_train))


# Model Accuracy on testing dataset
Y_pred_ridge_test = ridge.predict(X_test_poly)
print('The R^2 on the testing dataset is:', r2_score(Y_test, Y_pred_ridge_test))

# The Root Mean Squared Error (RMSE)
print('The RMSE on the training dataset is:', sqrt(mean_squared_error(Y_train, Y_pred_ridge_train)))
print('The RMSE on the testing dataset is:', sqrt(mean_squared_error(Y_test, Y_pred_ridge_test)))

# The Mean Absolute Error (MAE)
print('The MAE on the training dataset is:', mean_absolute_error(Y_train, Y_pred_ridge_train))
print('The MAE on the testing dataset is:', mean_absolute_error(Y_test, Y_pred_ridge_test))

# Coefficients
print('Coefficients:', ridge.coef_)

# The Intercept
print('Intercept:', ridge.intercept_)
```

```python
import seaborn as sns
import matplotlib.pyplot as plt

# Target Variable Distribution
plt.figure(figsize=(10, 6))
sns.histplot(data['IQ'], kde=True)
plt.title('Distribution of IQ')
plt.show()

# Relationship Between Features and Target
plt.figure(figsize=(15, 8))
for i, feature in enumerate(data1.columns):
    plt.subplot(2, 3, i + 1)
    sns.scatterplot(x=data1[feature], y=data1['IQ'])
    plt.title(f'{feature} vs IQ')
plt.tight_layout()
plt.show()

# Correlation Heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(data1.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```

Graphical Representations

````
```{r}
Gen.drug.freq=table(Sex,fre_cannabis)
Gen.drug.freq

rownames(Gen.drug.freq) <- c("Male","Female")
bar_colors <- c("skyblue", "lavender")

barplot(Gen.drug.freq, beside = FALSE, col = bar_colors,
        legend = rownames(Gen.drug.freq),
        main = "Barplot of cannabis user frequency with gender",
        xlab = "Usage",
        ylab = "Number of users",
        names.arg = c("Not at all","Every 3 months","Upto once a month","2-3 times a month","1-2 times a week","3-4 times a week","Almost  daily","Daily","Several times daily")
)
```
````

````
```{r}
# function to format labels with percentage
format_pct <- function(x) {
  paste0(formatC(100*x, format = "f", digits = 1), "%")
}

# cannabis users
cannabis_users <- subset(data, Cannabis == 2)
gender_table <- table(cannabis_users$Sex)
gender_labels <- c("Male", "Female")
pie(gender_table, labels = paste0(gender_labels, "\n", format_pct(prop.table(gender_table))),
    main = "Gender Distribution among Cannabis Users")
````

Same procedure was repeated for all the other drugs.
````

```python
import pandas as pd
import matplotlib.pyplot as plt

#Replace 1 with "No" and 2 with "Yes"
data.replace({1: 'No', 2: 'Yes'}, inplace=True)

# Calculate the total number of users for each drug
users_cannabis = data['Cannabis'].value_counts()
users_amphetamin = data['Amphetamine'].value_counts()
users_ecstacy = data['Ecstasy'].value_counts()
users_lsd = data['LSD'].value_counts()
users_mushrooms = data['Mushrooms'].value_counts()
users_ketamine = data['Ketamine'].value_counts()
users_cocaine = data['Cocaine'].value_counts()

# Create a bar chart
labels = ['Cannabis','Amphetamine', 'Ecstasy', 'LSD', 'Mushrooms', 'Ketamine', 'Cocaine']
users = [users_cannabis['Yes'], users_amphetamin['Yes'], users_ecstacy['Yes'], users_lsd['Yes'],
         users_mushrooms['Yes'],users_ketamine['Yes'],users_cocaine['Yes']]
colors = ['firebrick','palevioletred','lightpink','thistle','mediumpurple','m','indigo']

plt.bar(labels, users, color=colors)
plt.xlabel('Drug')
plt.ylabel('Number of Users')
plt.title('Drug Usage')
x = plt.show()
x
```

Chi-squared test

```r
```{r}
attach(data)

test1 <- chisq.test(table(Cannabis,Ecstasy))
test2 <- chisq.test(table(Cannabis,Amphetamine))
test3 <- chisq.test(table(Cannabis,LSD))
test4 <- chisq.test(table(Cannabis,Mushrooms))
test5 <- chisq.test(table(Cannabis,Ketamine))
test6 <- chisq.test(table(Cannabis,Cocaine))

test1
test2
test3
test4
test5
test6
```
```

Same process was repeated to other drugs.

Logistic Regression

```python
from imblearn.over_sampling import SMOTE
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import precision_score, recall_score, f1_score, confusion_matrix, roc_auc_score, accuracy_score
import statsmodels.api as sm
import numpy as np

# Assuming your data is stored in a DataFrame called 'data'

# List of drugs
drugs = ['Cannabis', 'LSD', 'Cocaine', 'Ecstasy', 'Ketamine', 'Amphetamine', 'Mushrooms']
```

```python
for drug in drugs:
    # Extract predictors (X) and target variable (Y) for the current drug
    X = data[['Age', 'IQ', 'Narcissism']]
    y = data[drug]

    # Convert target variable to numeric format (0 or 1)
    label_encoder = LabelEncoder()
    y_numeric = label_encoder.fit_transform(y)

    # Apply SMOTE over-sampling
    smote = SMOTE(sampling_strategy='auto', random_state=42)
    X_resampled, y_resampled = smote.fit_resample(X, y_numeric)  # Use the numeric version of y

    # Split the resampled data into training and testing sets
    X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.1, random_state=42)

    # Standardize numerical variables
    scaler = StandardScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)

    # Fit a logistic regression model
    model = LogisticRegression()
    model.fit(X_train_scaled, y_train)

    # Use statsmodels to get p-values
    X_train_sm = sm.add_constant(X_train_scaled)
    logistic_model = sm.Logit(y_train, X_train_sm)
    result = logistic_model.fit()

    # Predictions on the test set
    y_pred = model.predict(X_test_scaled)

    # Calculate evaluation metrics
    precision = precision_score(y_test, y_pred, pos_label=1, zero_division=0)
    recall = recall_score(y_test, y_pred, pos_label=1, zero_division=0)
    f1 = f1_score(y_test, y_pred, pos_label=1, zero_division=0)

    tn, fp, fn, tp = confusion_matrix(y_test, y_pred).ravel()
    specificity = tn / (tn + fp)

    auc_roc = roc_auc_score(y_test, model.predict_proba(X_test_scaled)[:, 1])
    accuracy = accuracy_score(y_test, y_pred)

    # Print results for each drug
    print(f"\nMetrics for {drug} ")
    print(f"Precision: {precision:.4f}")
    print(f"Recall: {recall:.4f}")
    print(f"F1-Score: {f1:.4f}")
    print(f"Specificity: {specificity:.4f}")
    print(f"AUC-ROC: {auc_roc:.4f}")
    print(f"Accuracy: {accuracy:.4f}")

    # Print only p-values from the logistic regression summary
    print(f"\nP-values for {drug}")
    coefficient_labels = ['Constant', 'Age', 'IQ', 'Narcissism']
    for i, p_value in enumerate(result.pvalues):
        print(f"{coefficient_labels[i]}: {round(p_value, 4)}")
```

REFERENCES

- Volkow, N.D., Fowler, J.S., Wang, G.-J., Swanson, J.M. and Telang, F. (2007). Dopamine in Drug Abuse and Addiction. *Archives of Neurology*, 64(11), p.1575.
- Goudriaan, A.E., Oosterlaan, J., de Beurs, E. and van den Brink, W. (2006). Neurocognitive functions in pathological gambling: a comparison with alcohol dependence, Tourette syndrome and normal controls. *Addiction*, 101(4), pp.534–547.
- Oscar-Berman, M. and Marinković, K. (2007). Alcohol: Effects on Neurobehavioral Functions and the Brain. *Neuropsychology Review*, 17(3), pp.239–257.
- Volkow, N.D., Wang, G.-J., Fowler, J.S., Tomasi, D., Telang, F. and Baler, R. (2010). Addiction: Decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. *BioEssays*, [online] 32(9), pp.748–755.
- Fattore, L. and Fratta, W. (2010). How important are sex differences in cannabinoid action? *British Journal of Pharmacology*, 160(3), pp.544–548.
- Trull, T.J., Jahng, S., Tomko, R.L., Wood, P.K. and Sher, K.J. (2010). Revised NESARC Personality Disorder Diagnoses: Gender, Prevalence, and Comorbidity with Substance Dependence Disorders. *Journal of Personality Disorders*, 24(4), pp.412–426.
- Turiano, N.A., Whiteman, S.D., Hampson, S.E., Roberts, B.W. and Mroczek, D.K. (2012). Personality and substance use in midlife: Conscientiousness as a moderator and the effects of trait change. *Journal of Research in Personality*, 46(3), pp.295–305.
- Luijten, M., Machielsen, M., Veltman, D., Hester, R., de Haan, L. and Franken, I. (2014). Systematic review of ERP and fMRI studies investigating inhibitory control and error processing in people. *Journal of Psychiatry & Neuroscience*, 39(3), pp.149–169.
- Pope, H.G. and Yurgelun-Todd, D. (1996). The residual cognitive effects of heavy marijuana use in college students. *JAMA*, [online] 275(7), pp.521–527. Available at: https://pubmed.ncbi.nlm.nih.gov/8606472/ [Accessed 18 Mar. 2023].
- Render, A. and Jansen, P. (2019). Dopamine and sense of agency: Determinants in personality and substance use. *PLOS ONE*, 14(3), p.e0214069.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association.

- psycnet.apa.org. (n.d.). *APA PsycNet*. [online]
- World Health Organization (2022). *World Health Organization*. [online] Who.int.
- DEA (2019). *Home | DEA*. [online] Dea.gov.
- Cambridge University Press (2022). *Cambridge Dictionary | English Dictionary, Translations & Thesaurus*. [online] Cambridge.org.
- NIDA (2022). *NIDA.NIH.GOV | National Institute on Drug Abuse (NIDA)*. [online] National Institute on Drug Abuse. Available at: https://nida.nih.gov/.
- Pearsonassessments.com. (2019). *Wechsler Adult Intelligence Scale | Fourth Edition*. [online]
- Jaber, M., Robinson, S.W., Missale, C. and Caron, M.G. (1996). Dopamine receptors and brain function. *Neuropharmacology*, 35(11), pp.1503–1519.
- NICHOLS, S. and ULATOWSKI, J. (2007). Intuitions and Individual Differences: The Knobe Effect Revisited. *Mind & Language*, 22(4), pp.346–365.
- United Nations : World Drug Report 2020. (n.d.). *World Drug Report 2020*. [online]
- PhD, M.M. (2020). *Brain plasticity in drug addiction: Burden and benefit*. [online] Harvard Health Blog.