Understanding Drug Use: Insights and Key Factors of Drug Behaviour

Abstract: In the report, we utilize a dataset containing measures of different personality traits and frequency alongside use frequencies of multiple drugs for 1885 participants. We found that not only are there distinct differences between users and non-users of drugs, but also significant differences with the drug user category. Models were constructed to determine the relationship between each personality trait and a particular drug use risk behaviour, with differences being inferred using model outputs.

1. INTRODUCTION

Drug abuse has consistently been a problem in society regardless of region and time period, with the global drug situation constantly increasing in severity and complexity[1] e.g. the ongoing Opioid epidemic in the US. Current methods of prevention set by global policymakers largely revolve around the prohibition of these substances which was heavily influenced by President Nixon during his War on Drugs campaign. However, as we can now see, this global undertaking was not only a massive failure in preventing drug abuse domestically -contributing to the increased rate of drug overdoses and drug-related illness-, but also helped nurtured and sustained powerful drug cartels and criminal organizations across the world[2].

To develop better drug policies, a foundational understanding of why people abuse drugs must be established. This report will attempt to characterize and understand the underlying psychological relationships that influence drug behaviours using different personality trait measures. (150)

2. ANALYTICAL QUESTIONS AND DATA

2.1 Analytical Questions

To answer our research question, differences between users/ non-users and users of different drugs were investigated. Listed below are the questions that will be answered in this paper:

1. What are the differences in psychological profiles between users and non-users?
2. How do the psychological profiles change between daily, weekly and monthly drug users?
3. How do psychological profiles differ between users of different drug categories?

2.2 Dataset

Drug consumption data for our analysis were obtained from the UCI Repository (Drug consumption(quantified) 2015). The dataset contains 1885 rows, 12 input features and drug usage measurements of 18 legal and illegal drugs. All input features were originally categorical and were later quantified by the author using different methods based on the original data type (Ordinal/Nominal)[3]. Post-quantification, all features can be treated as real values; therefore, data transformation is not needed when inputting them into models. **Table 1** provides a brief summary of the 12 features. Drug usage was categorized into 7 classes based on most recent use.

Features	Data Type	Values
Age	Categorical (Ordinal)	6 Age Groups
Education	Categorical (Ordinal)	9 Levels of Education
Gender	Categorical (Nominal)	[0.4826(Female), -0.4826(Male)]
Country	Categorical (Nominal)	7 Different Countries
Ethnicity	Categorical (Nominal)	7 Different Ethnicities
NScore (N)	Continuous	[-3.46436……3.27393]
EScore (E)	Continuous	[-3.27393……3.27393]
OScore (O)	Continuous	[-3.27393……2.90161]
AScore (A)	Continuous	[-3.46436……3.46436]
CScore (C)	Continuous	[-3.46436……3.46436]
Impulsive	Continuous	[-2.55524……2.90161]
Sensation Seeking (SS)	Continuous	[-2.07848……1.92173]

Table 1. Feature Summary

As the dataset has a wide resolution of drug usage across different drugs, drug users can thus be aggregated based on drug frequency usage (monthly, weekly, daily) as well as their preferred drugs CNS Classification (depressants, stimulants, hallucinogens) and legality status. (281)

3. ANALYSIS

3.1 Data Preparation

The dataset came relatively clean with the absence of missing values, requiring only the addition of column headers and the removal of the first row, which contained corrupted drug usage frequency measures.

3.2 Data Derivation

Before new data can be derived, the minimum use frequency of a drug user must be defined. For this paper, a minimum use frequency of any of the drugs: Benzodiazepine, Heroin, Methadone, Amyl Nitrate, Amphetamines, Ecstasy, Cocaine, Crack, Mushrooms, LSD, Ketamine, Legal Highs, Volatile Substance Abuse in the last month is required to be considered as a drug user – "month-based". Participants who have used any of the drugs above outside of the last month or longer were considered as non-drug users.

It should be noted that cannabis, alcohol, caffeine, chocolate and nicotine were excluded from this list due to the common use in the general population. With a baseline use frequency established, users were further classified into "week-based" and "day-based" users, resulting in the creation of three new columns "monthly_u", "weekly_u" and "daily_u".

Using use frequency as a metric of drug preference, each participant's preferred drug was identified. Users were then grouped into 3 separate drug classes based on their preferred drug's effect on the central nervous system. As in the previous step, commonly used drugs were excluded from consideration and participants were thus classified as non-users. As Legal

Highs and Volatile Substance Abuse do not fall into any of the three categories, they were also classified as non-users. If a participant has an equal use frequency for two or more drugs in separate classification, their classification label was chosen at random. **Figure 1** shows the proportion of drug users categorized in each drug class.

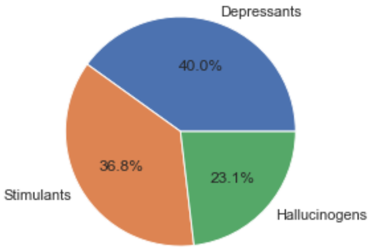

Figure 1. Proportion of users in each drug class

Lastly, participants were also grouped into legal and illegal drug users based on the legal status of their preferred drug. Cannabis was classified as a legal drug due to the shift in perception of its prohibition alongside its common use.

To determine if a high overlap exists between our newly derived data, Jaccard's Index was used. Frequency-based grouping saw the highest degree of overlap within the different groups, with "monthly_u" and "weekly_u" producing a Jaccard's Index of 0.673. This however, should not affect our work as we are determining differences based on changes in coefficient value using the "monthly_u" as a baseline measure.

3.3 Model Reconstruction

Logistic Regression (LR) was the model of choice for our modelling work as its coefficients are easily interpreted and can be directly compared to coefficients of other models to infer the change in feature importance to the classification. Python's "sklearn.linear_model.LogisticRegression" package was used for this analysis.

Before any modelling work was conducted, key assumptions made by LR was first verified. Only features regarding personality traits were considered as those were the only factors included in our model. Feature multicollinearity was determined using the variance inflation factor (VIF). All features had a VIF of <4, with the "SS" having the highest value of 2.1 (**Table 2.**) .This indicates that all our features have little to no multicollinearity present satisfying the assumption[4].

Features	VIF
NScore	1.488
EScore	1.571
OScore	1.438
AScore	1.198
CScore	1.468
Impulsive	1.803
SS	2.099

Table 2. Features and their associated VIF scores

To determine if independent variables are linearly related to the log odds of our dependent variable, empirical logit plots were constructed. Across all tested features, a linear fitted line was seen, satisfying the assumption.

As there are generally fewer drug users than there are non-users, the dependent variables were consistently imbalanced across the different groups. Therefore, to circumvent this "SMOTENC" from the "imblearn" package was utilized. The features "Impulsivity" and "SS" were treated as categorical variables in "SMOTENC" as these features did not follow a normal distribution; therefore, synthetic data generated for these variables had to fall within one of the predefined discrete classes. All other features were treated as continuous variables.

The 7 personality measures were used as the independent variables across all our modelling work, with scaling being done prior using "scale" from the "sklearn.preprocessing" module. A minor decrease in model performance was observed; however, the benefits of being able to directly compare feature importance between models outweigh the marginal decrease in accuracy (~0.01-0.03).

3.4 Model Validation

As class imbalances were mitigated using random oversampling, we were able to use ROC as one of our model performance metrics alongside accuracy (misclassification rate). Both train-test split and 10 K-Fold cross-validation were used to determine the presence of overfitting and determine the generalization error.

The effect sizes of our claims inferred from our models were determined using each model's AUC score. According to Harris et al., for a phenomenon with a 25% base rate, an AUC of 0.622 or greater is required for medium effect size[5]. According to the NHS, 9.4% of adults (16 to 59) have taken an illicit drug in the last year. Assuming under-reporting is present, the real base rate of drug use would still be significantly lower than 25% of the adult population in the UK. Therefore, the AUC score required for a medium effect size would be lower. Our lowest-scoring model had an AUC Score of 0.613, therefore differences seen between groups in each model would have an effect size of medium or higher. (885)

4. FINDINGS, REFLECTIONS AND FURTHER WORK

4.1 What are the differences in psychological profiles between users and non-users?

Features	*Coefficient Values*
N	0.043
E	-0.314
O	0.468
A	0.163
C	-0.309
I	0.041
SS	0.892

Table 3. Features and their associated coefficient values

With reference to **Table 3**, high degrees of N, O, Impulsive and Sensation Seeking traits have a positive relationship with drug use status, highlighting them as possible determinants of drug use behaviour. Sensation seeking seems to be the largest determinant seeing a coefficient value of 0.892. Whereas non-users have a higher degree of E, A, and C. These differences between the two populations are also reflected in the distribution graphs below **Figure 2**. As a pleasant surprise, the differences seen between users and non-users from our models were also observed in another study conducted by Torriano et al. Their study found a positive correlation between drug use and increasing N and O scores and increasing C. A decreases the risk of drug use[6], providing confidence of our modelling techniques.

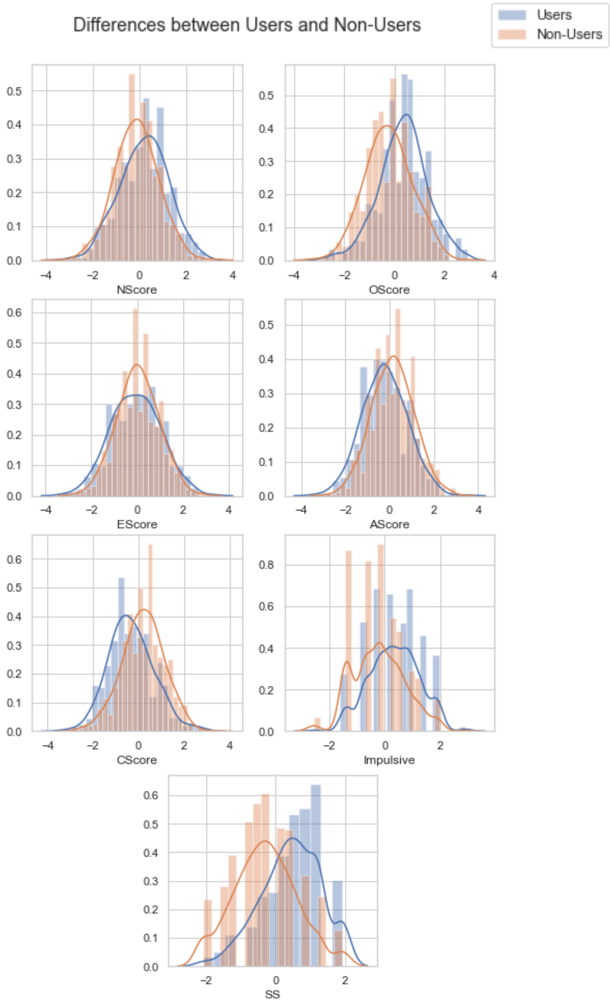

Figure 2. Differences in feature distribution between users and non-users.

4.2 How do the psychological profiles change between daily, weekly and monthly drug users?

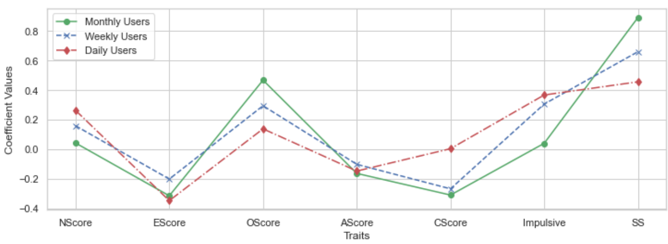

Figure 3. Differences in feature coefficient values between the 3 different use frequency models.

Figure 3. shows that N, C and Impulsiveness are positively correlated with increased drug use. Sensation Seeking, even though still identified as the most important feature in our daily model, sees a decrease in feature importance as drug use frequency decreases with a similar trend being seen in O. This indicates that neuroticism, conscientiousness and impulsiveness are the main determinants of a user's degree of addiction, with high degrees of neuroticism having been documented to have a positive correlation with many other forms of addiction i.e. shopping and internet addiction[7].

4.3 How do psychological profiles differ between users of different drug categories?

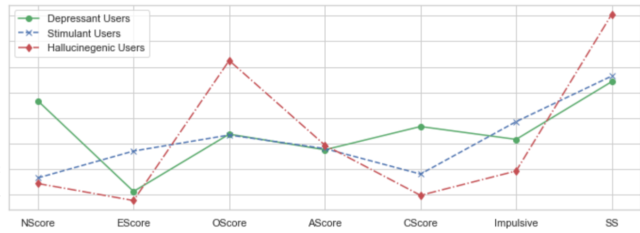

Figure 4. Differences in feature coefficients between the 3 different drug class models.

Depressants, stimulants and hallucinogens each affect the central nervous system in distinctly different ways. We would thus expect users to differ greatly in several psychological features, providing some insight for their preferences—**figure 4**. We see the traits N, O, C, Impulsivity and SS as good determinants of drug preferences. Depressant and hallucinogen users have the greatest difference in psychological profiles within this grouping. Users tend to be on opposite ends for 4 out of 7 traits, whereas stimulant users fall somewhere between the two. Stimulant and depressant users, on the other hand, have the least difference, with only 2 out of the 7 features being distinctly different. The degree of A seems to be the least important feature in drug preference with a coefficient of ~0 across all three drug classes.

Results from this analysis show that drug behaviour can and is heavily influenced by psychological profiles, with significant differences between users and non-users and users of different drugs. We see a strong relationship between high degrees of neuroticism, impulsivity and sensation-seeking with increased risk of drug consumption. Drug preference being largely determined by a user's degree of continuousness, impulsivity and sensation-seeking personality traits.

4.4 Reflections and Future Work

Overall, the dataset has been suitable for the analysis conducted by this paper due to the number of features and a wide range of drugs and drug frequencies recorded. With the use of data derivation, all analytical questions were answered without the use of any additional features from other datasets. A strong limitation of this analysis, however, revolves around the size of the dataset itself. With 1884 usable observations, inferences can be made from the data however, the precision of such inferences will be limited. Mode of reporting frequency usage may also be another limitation as it results in nested classes. Future work would include investigating the differences in drug use behaviour of participants in different age groups, country of origin, ethnicity and degree of education. (640)

Reference

1. United Nations : Office on Drugs and Crime. 2020. World Drug Report 2019: 35 Million People Worldwide Suffer From Drug Use Disorders While Only 1 In 7 People Receive Treatment. [online] Available at: <https://www.unodc.org/unodc/en/frontpage/2019/June/world-drug-report-2019_-35-million-people-worldwide-suffer-from-use-disorders-while-only-1-in-7-people-receive-treatment.html> [Accessed 19 December 2020].

2.Coyne, C. and Hall, A., 2020. Four Decades And Counting: The Continued Failure Of The War On Drugs. [online] Cato Institute. Available at: <https://www.cato.org/publications/policy-analysis/four-decades-counting-continued-failure-war-drugs> [Accessed 19 December 2020].

3.Fehrman, E., Muhammad, A., Mirkes, E., Egan, V. and Gorban, A., 2017. The Five Factor Model of personality and evaluation of drug consumption risk. ARXIV, [online] Available at: https://arxiv.org/abs/1506.06297v2 [Accessed 19 December 2020].

4.O'brien, R., 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity, [online] 41(5), pp.673-690. Available at: <https://www.researchgate.net/publication/226005307_A_Caution_Regarding_Rules_of_Thumb_for_Variance_Inflation_Factors> [Accessed 15 December 2020].

5.Rice, M. and Harris, G., 2005. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law and Human Behavior, 29(5), pp.615-620.

6. Your Bibliography: Turiano, N., Whiteman, S., Hampson, S., Roberts, B. and Mroczek, D., 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.

7. Andreassen, C., Griffiths, M., Gjertsen, S., Krossbakken, E., Kvam, S. and Pallesen, S., 2013. The relationships between behavioral addictions and the five-factor model of personality. Journal of Behavioral Addictions, 2(2), pp.90-99.