

ANALYZING THE RELATIONSHIP BETWEEN PERSONALITY TRAITS AND DRUG CONSUMPTION

Stive Romarique

Hugo Verley

Raphaël Veron

DATA DESCRIPTION



Data Source: Our dataset is sourced from [archive.ics](#).




Dataset Overview: The dataset includes 1886 records, detailing age, gender, education level, and specific personality scores like extraversion, neuroticism, etc. It also records consumption patterns for various drugs.



Significance of the Data: This dataset provides a unique opportunity to uncover patterns and relationships between personality and drug use, offering insights that could be valuable for psychological and public health research.

DATASET CHARACTERISTICS



Drug consumption (quantified)

Donated on 10/16/2016

Classify type of drug consumer by personality data

Dataset Characteristics	Subject Area
Multivariate	Social Science
Associated Tasks	Feature Type
Classification	Real
# Instances	# Features
1885	-

DATA PREPROCESSING

Initial Cleaning: Our initial step involved cleaning the data. We addressed missing values, removed outliers, and filtered irrelevant entries to ensure data quality and relevance.

Initial Cleaning :

[illegible]

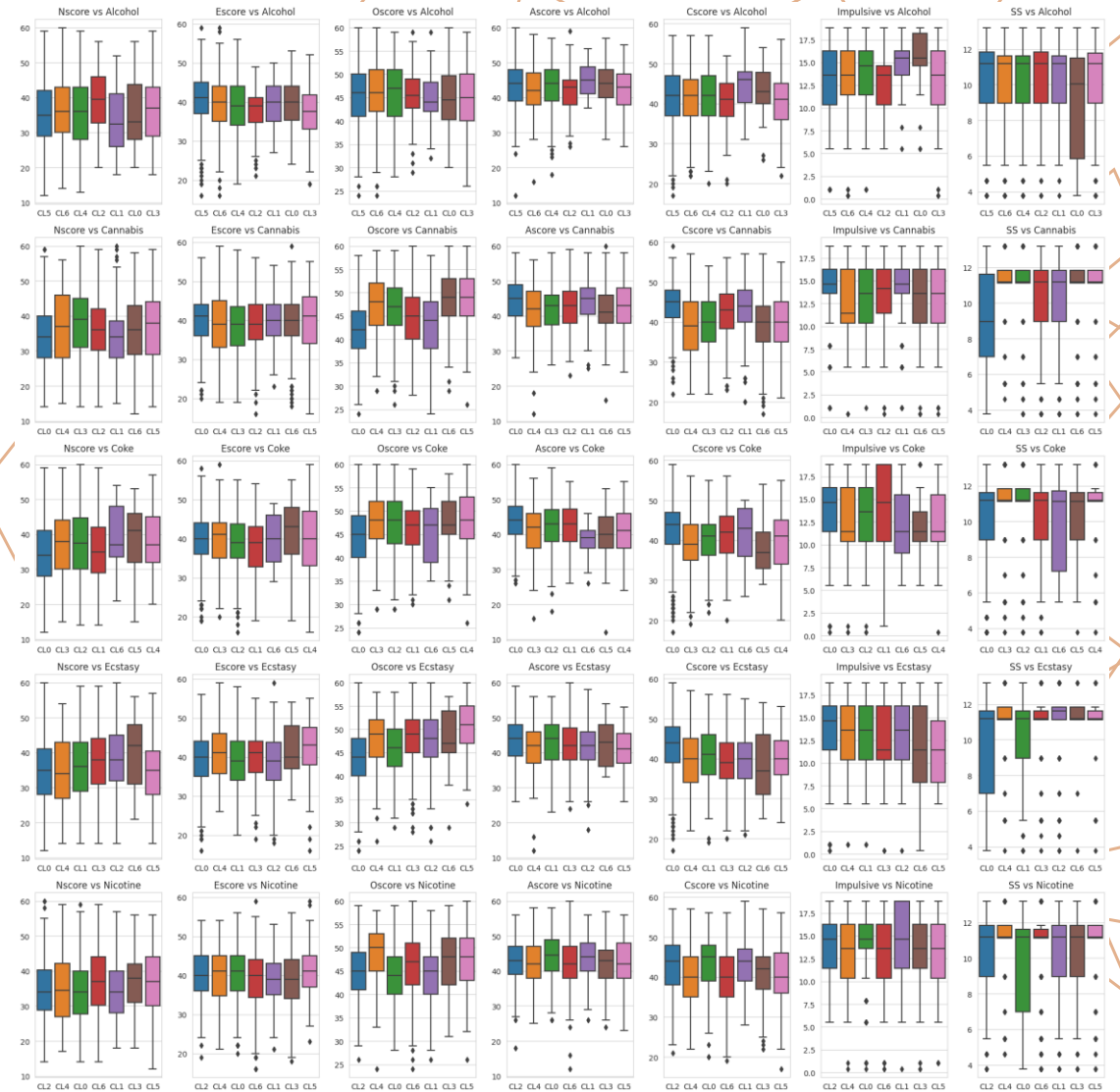
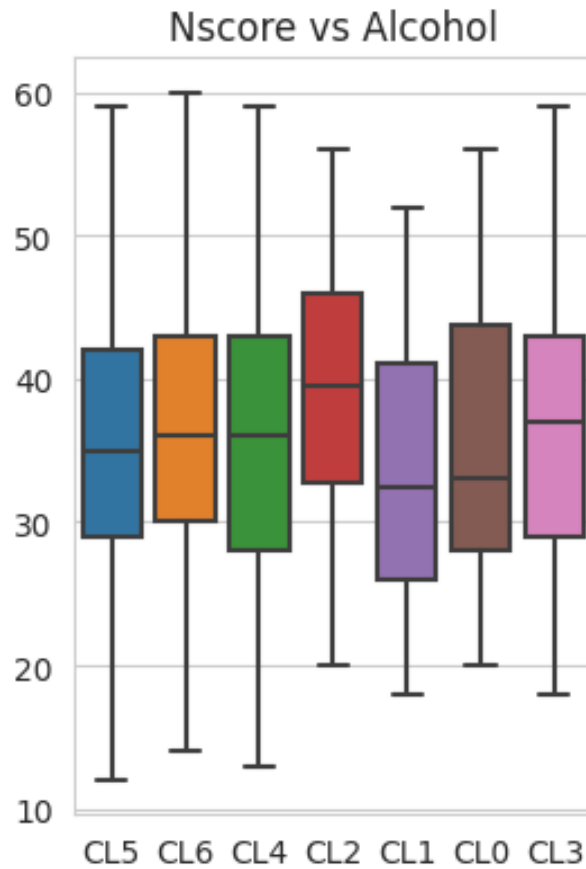
Data Transformation: We then transformed the data, binarizing continuous variables and encoding categorical ones to prepare them for analysis and modelling.

Data Transformation:

	ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore
0	1	2	0	6	5	3	39	36	42	37
1	2	1	1	0	5	6	29	52	55	48
2	3	2	1	6	5	6	31	45	40	32
3	4	0	0	5	5	6	34	34	46	47
4	5	2	0	0	5	6	43	28	43	41
5 rows × 38 columns										

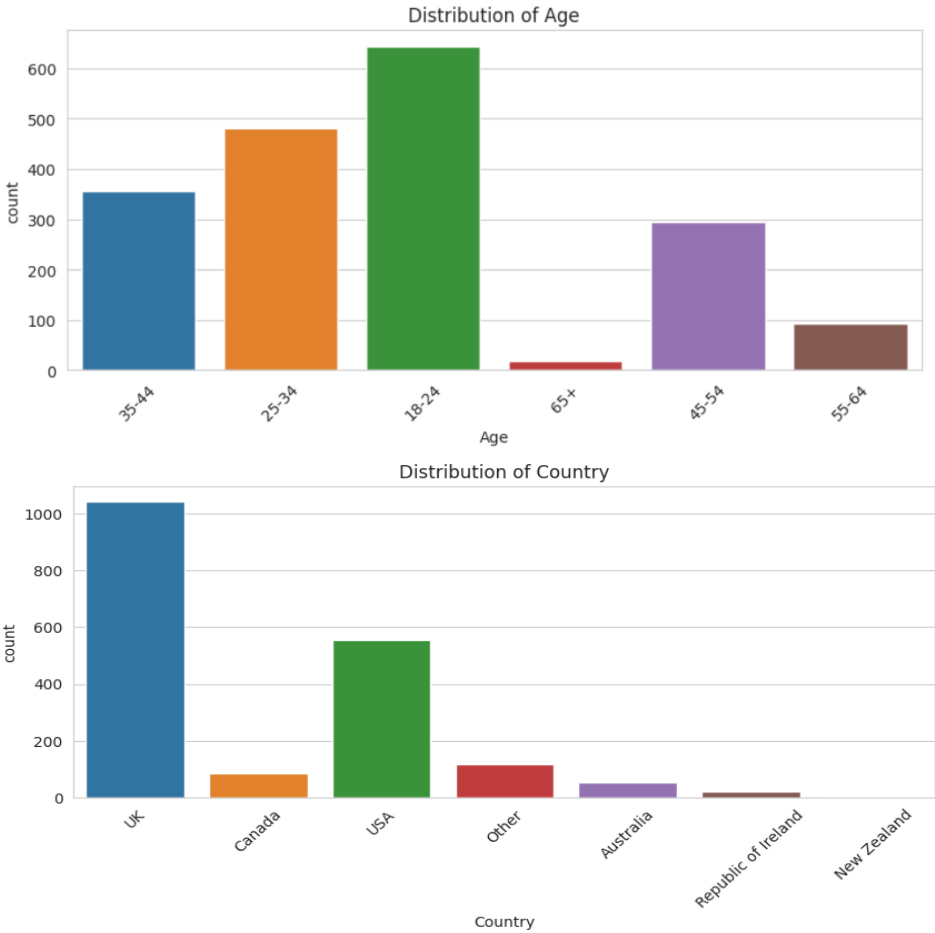
EXPLORATORY DATA ANALYSIS (EDA)

Statistical Summary: Our exploratory analysis began with a statistical overview, examining distributions, means, and variances of different variables.

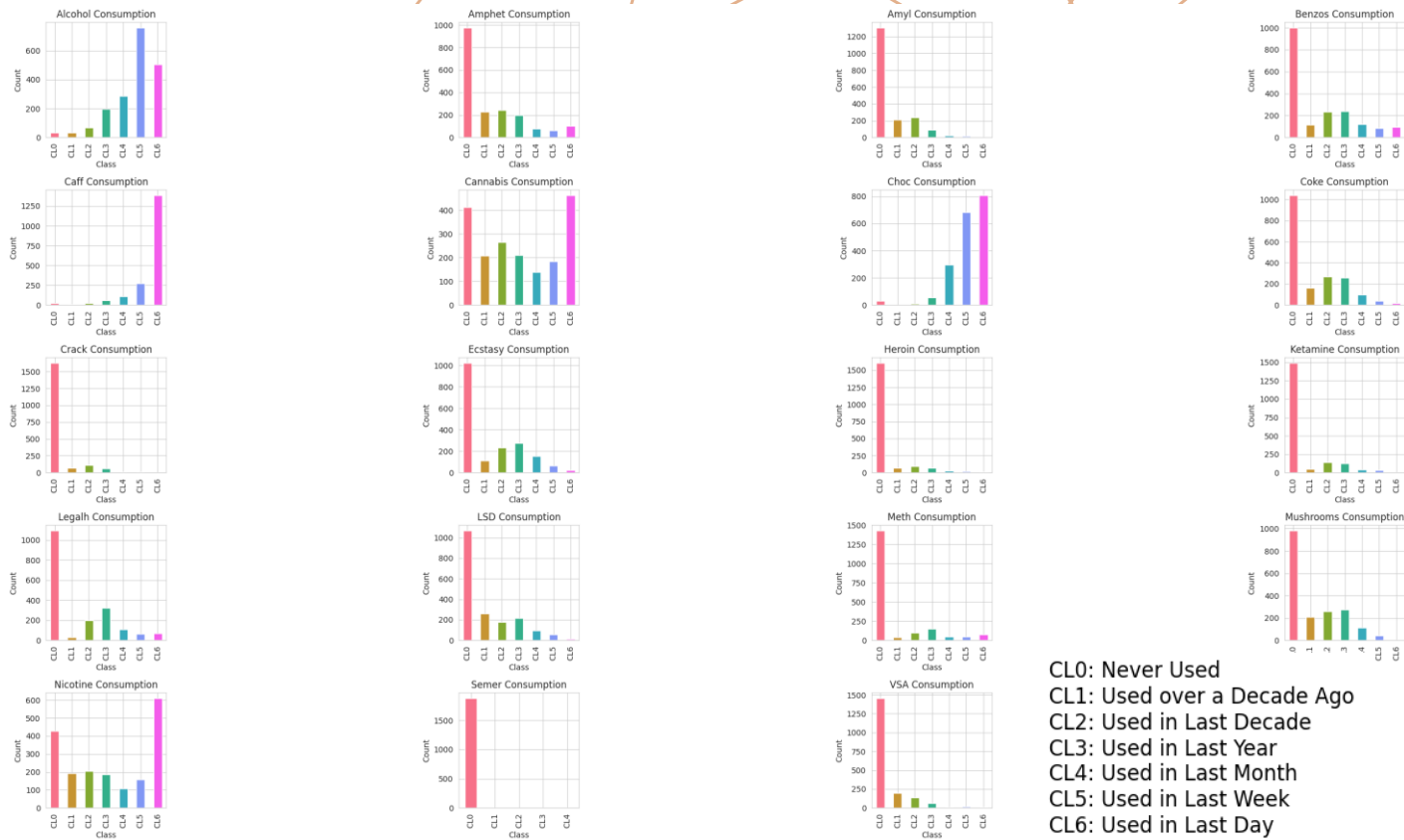


EXPLORATORY DATA ANALYSIS (EDA)

Visualizations: Visualizations like histograms, scatter plots, and box plots were used to uncover initial patterns and relationships in the data, such as the distribution of personality traits across different drug consumption levels.

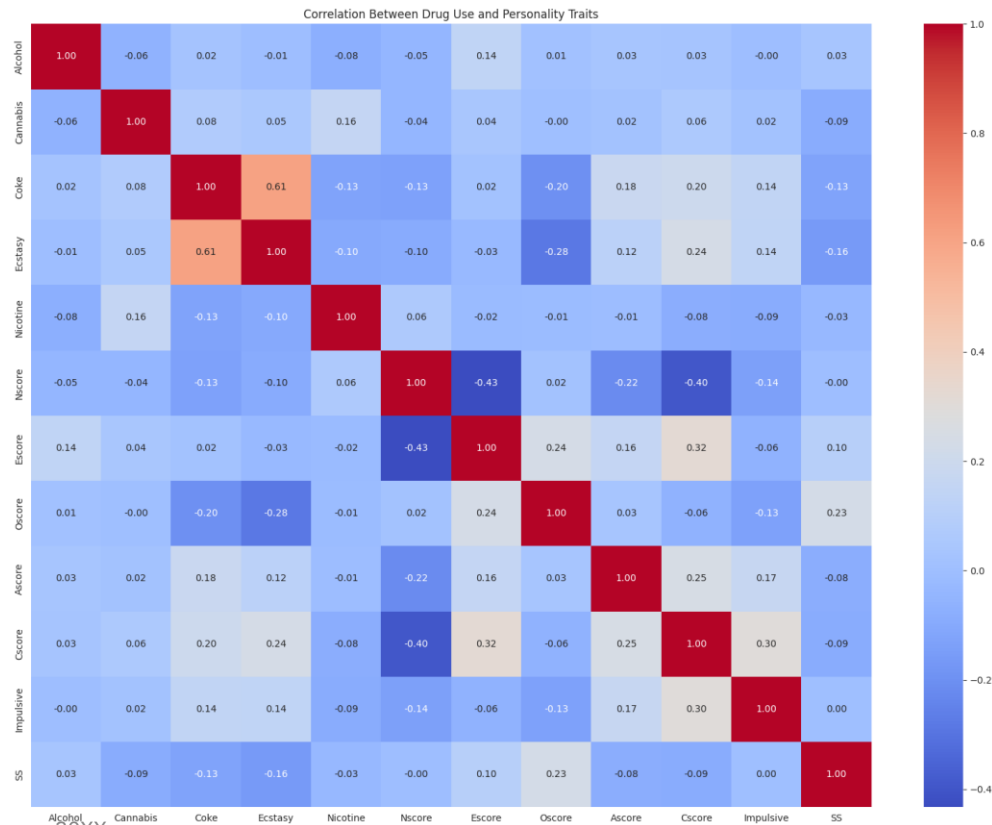


Drug Consumption Bar Plot

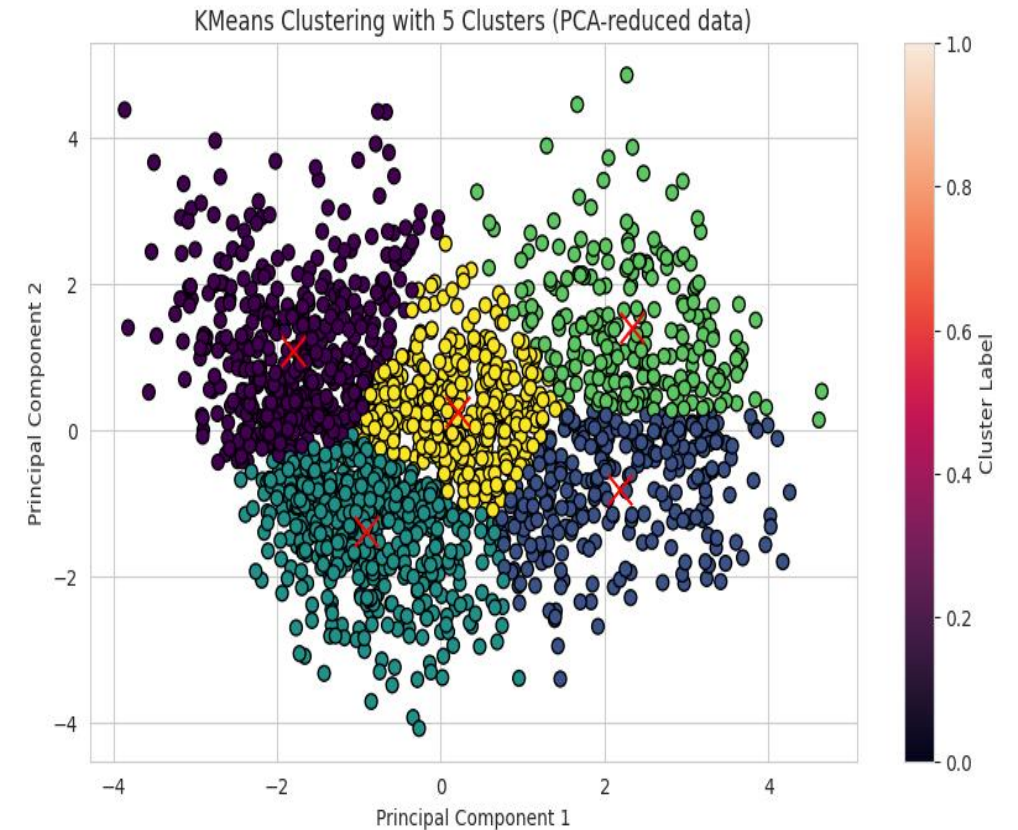


FEATURE ENGINEERING

Feature Selection: Feature selection was conducted using techniques like correlation analysis and principal component analysis to identify the most predictive variables.



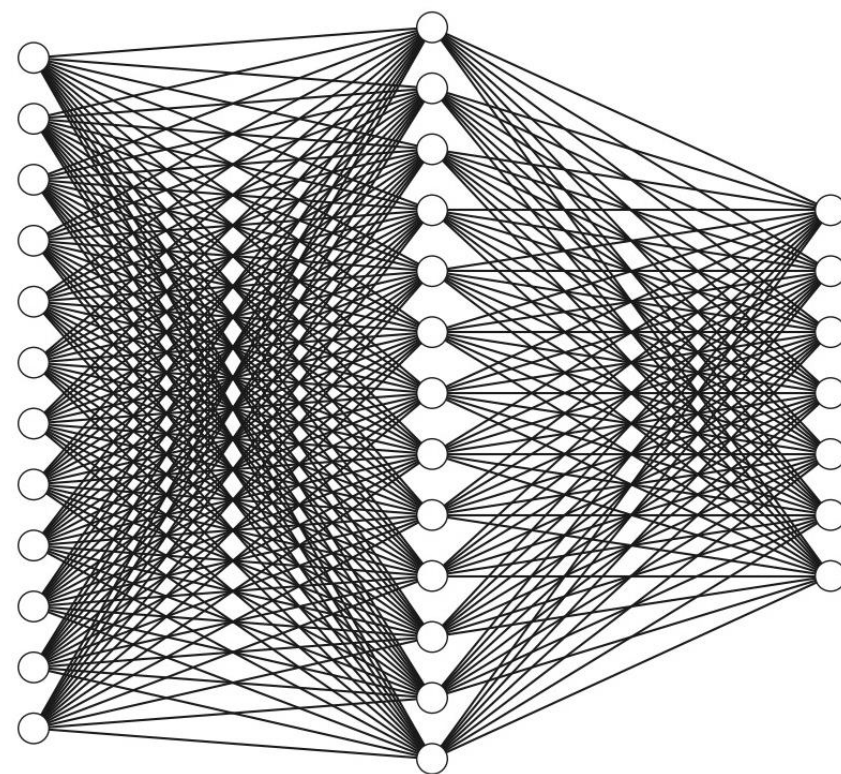
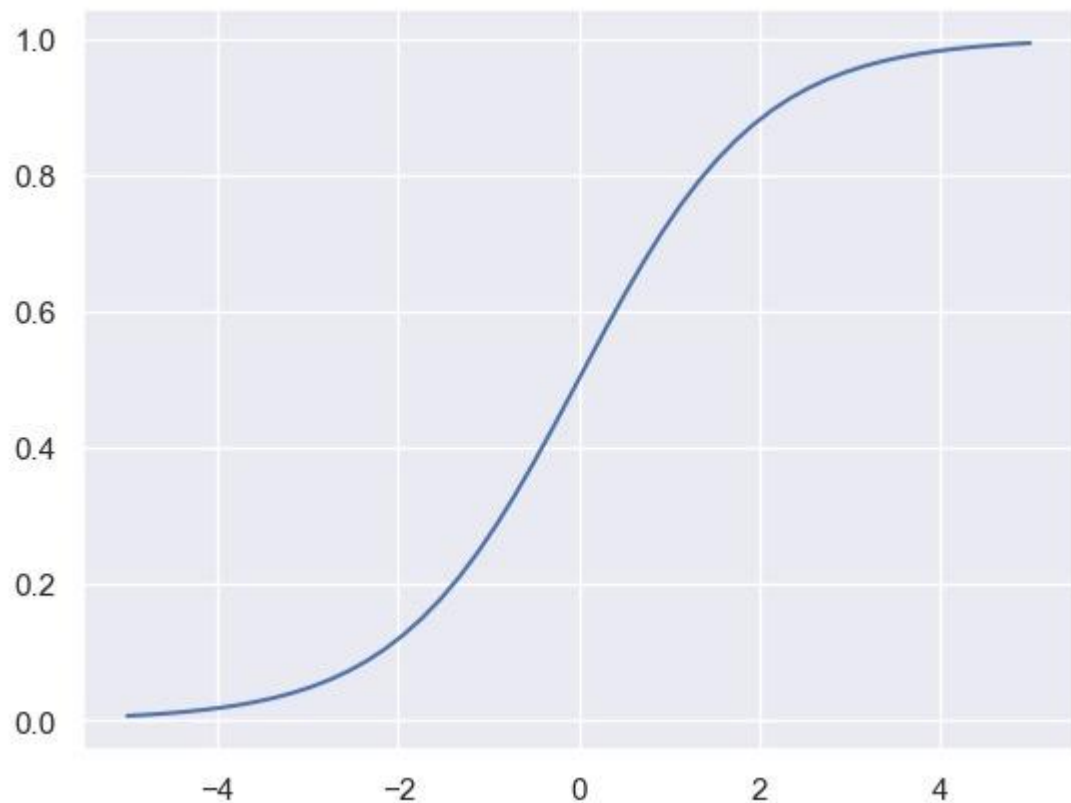
Creating New Variables: We engineered new features that combined different personality traits, hypothesizing that certain trait combinations might better predict drug use.



MODEL SELECTION AND TRAINING

Choice of Model(s): We evaluated several models, including **logistic regression** and **Neural Network**, for their suitability in handling our data and predicting outcomes.

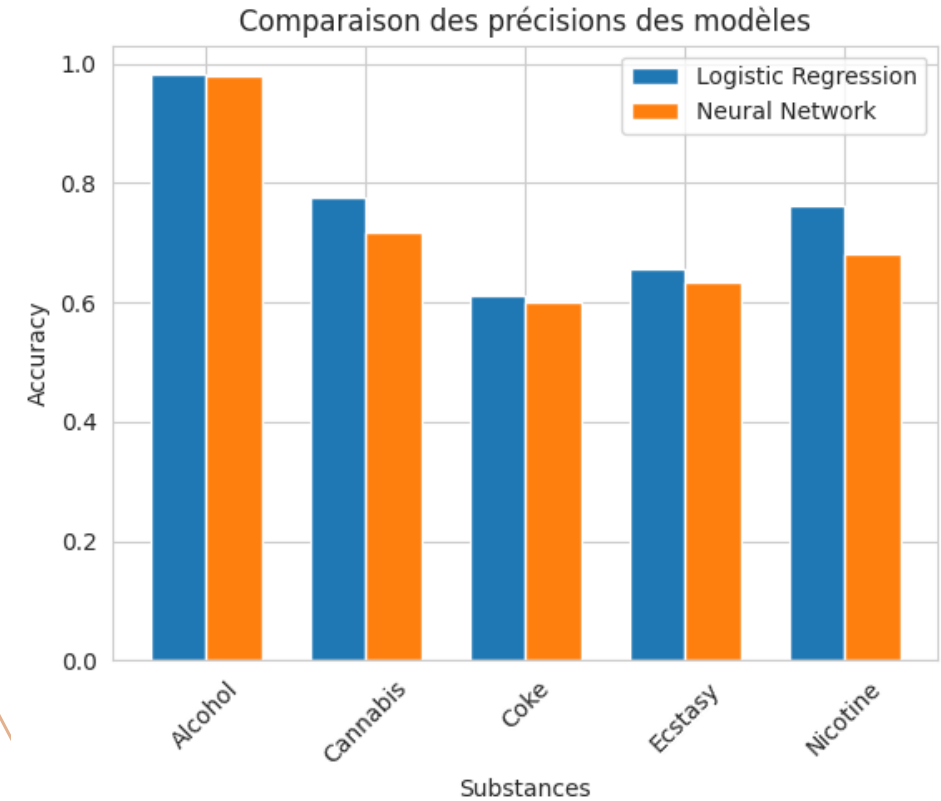
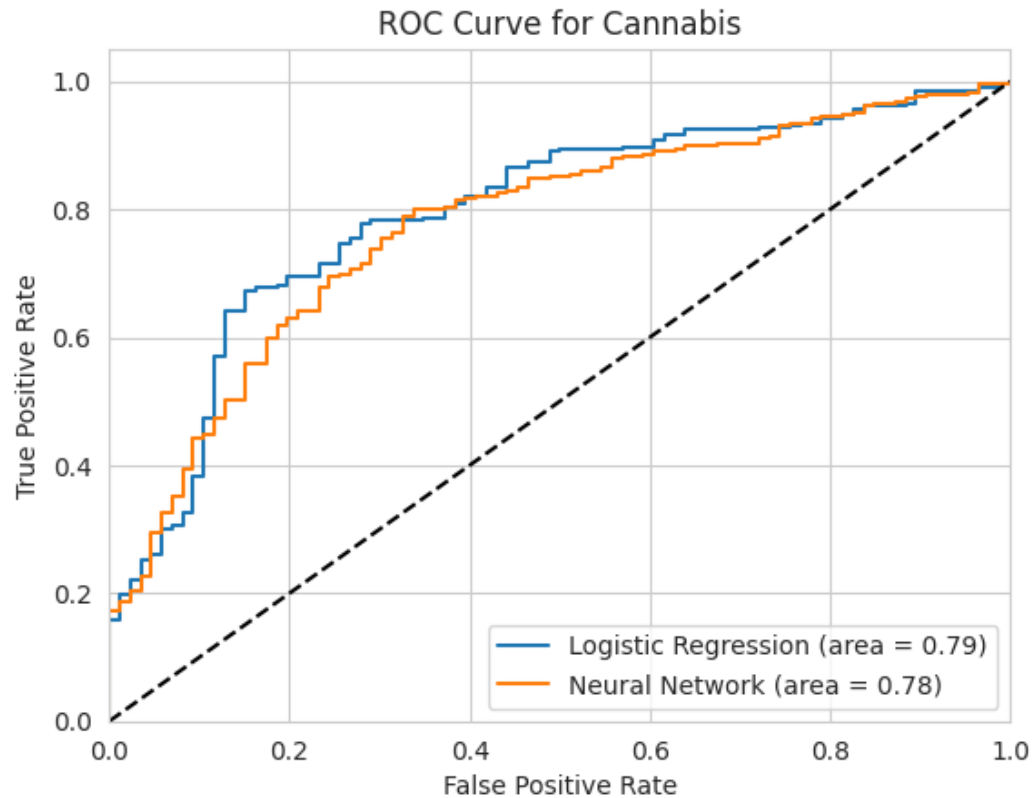
Training Process: The selected models were trained on a subset of the data, using cross-validation to ensure robustness and prevent overfitting.



MODEL EVALUATION

Evaluation Metrics: Model performance was assessed using metrics like accuracy, precision, recall, and the AUC score.

Results: The results indicate that **Logistic Regression** performed the best, accurately predicting drug consumption in 0,98% of cases.



INSIGHTS AND INTERPRETATIONS

Key Findings: Our analysis revealed significant relationships between certain personality traits and the likelihood of drug consumption.

Interpretations: For example, traits like Cscore were found to be strong predictors of ecstasy use. This suggests that personality profiling could be a valuable tool in preventive health strategies.

Conclusion: We've chosen neural networks for substance use prediction due to their ability to handle complex relationships, offer advanced modeling, ensure high accuracy, and mitigate overfitting.

CHALLENGES AND LIMITATIONS

Challenges Faced: Challenges in this project included dealing with imbalanced data and interpreting complex relationships between variables.

Limitations of the Study: It's important to note the limitations of our approach, including potential biases in the dataset and the generalizability of our findings.

CONCLUSION AND FUTURE WORK

Concluding Thoughts: In conclusion, our study provides valuable insights into the relationship between personality traits and drug consumption, with potential applications in psychology and public health.

Suggestions for Future Research: Future research could explore deeper into subtypes of personality traits, incorporate longitudinal data, or examine the influence of external factors like socio-economic status.