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
Multivariate

Subject Area
Social Science

Associated Tasks
Classification

Feature Type
Real

 # Instances
1885

 # Features
-

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID,Age,Gender,Education,Country,Ethnicity,Nscore,Escore,Oscore,Ascore,Cscore,Impulsive,SS,Alcohol,Amphet,Amyl,Benzos,Caff,Cannabis,Choc,Coke,Crack,Ecstasy,Heroin,Ketamine,Legalh,LSD,M													
2	1,0.49788,0.48246,-0.05921,0.96082,0.126,0.31287,-0.57545,-0.58331,-0.91699,-0.00665,-0.21712,-1.18084,CL5,CL2,CL0,CL2,CL6,CL0,CL5,CL0,CL0,CL0,CL0,CL0,CL0,CL0,CL0,CL0,CL2,CL0,CL0													
3	2,-0.07854,-0.48246,1.98437,0.96082,-0.31685,-0.67825,1.93886,1.43533,0.76096,-0.14277,-0.71126,-0.21575,CL5,CL2,CL2,CL0,CL6,CL4,CL6,CL3,CL0,CL4,CL0,CL2,CL0,CL2,CL3,CL0,CL4,CL0,CL0													
4	3,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.46725,0.80523,-0.84732,-1.6209,-1.0145,-1.37983,0.40148,CL6,CL0,CL0,CL0,CL6,CL3,CL4,CL0,CL0,CL0,CL0,CL0,CL0,CL0,CL0,CL1,CL0,CL0,CL0													
5	4,-0.95197,0.48246,1.16365,0.96082,-0.31685,-0.14882,-0.80615,-0.01928,0.59042,0.58489,-1.37983,-1.18084,CL4,CL0,CL0,CL3,CL5,CL2,CL4,CL2,CL0,CL0,CL0,CL2,CL0,CL0,CL0,CL0,CL2,CL0,CL0													

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.

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

Initial Cleaning :

	ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore
0	1	35-44	Female	Professional certificate/ diploma	UK	Mixed-White/Asian	39	36	42	37
1	2	25-34	Male	Doctorate degree	UK	White	29	52	55	48
2	3	35-44	Male	Professional certificate/ diploma	UK	White	31	45	40	32
3	4	18-24	Female	Masters degree	UK	White	34	34	46	47
4	5	35-44	Female	Doctorate degree	UK	White	43	28	43	41

5 rows × 32 columns

Data Transformation:

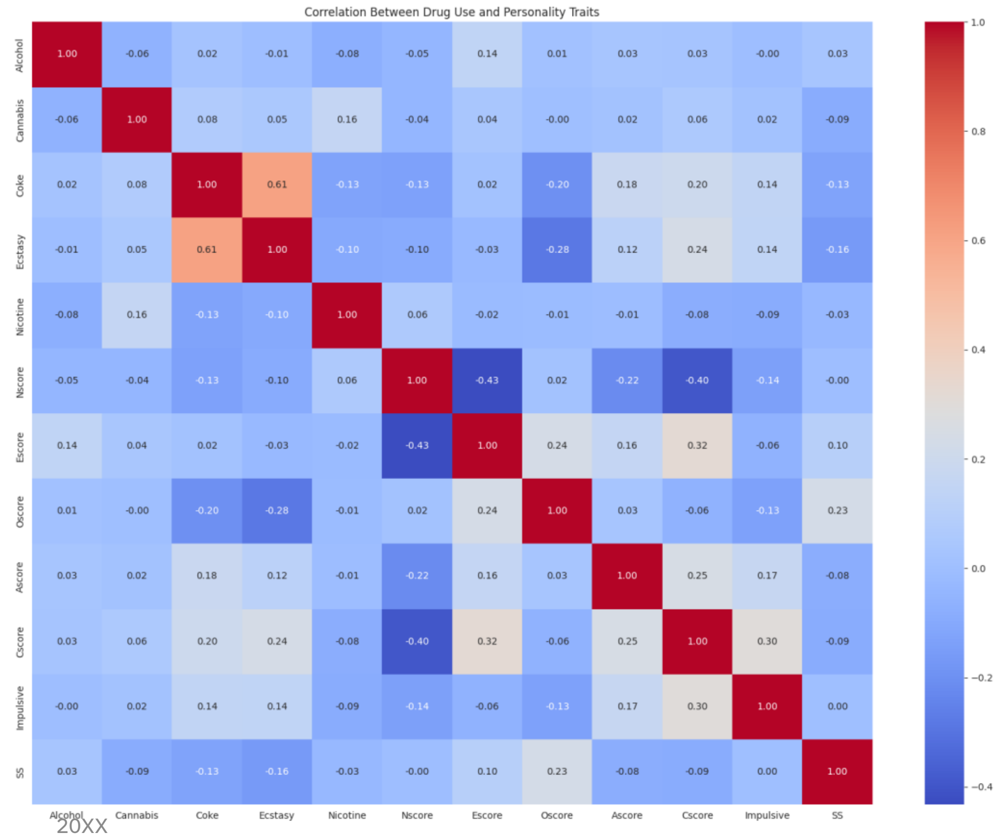
	ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore
0	1	2	0	6	5	3	39	36	42	37
1	2	1	1	0	5	6	29	52	55	48
2	3	2	1	6	5	6	31	45	40	32
3	4	0	0	5	5	6	34	34	46	47
4	5	2	0	0	5	6	43	28	43	41

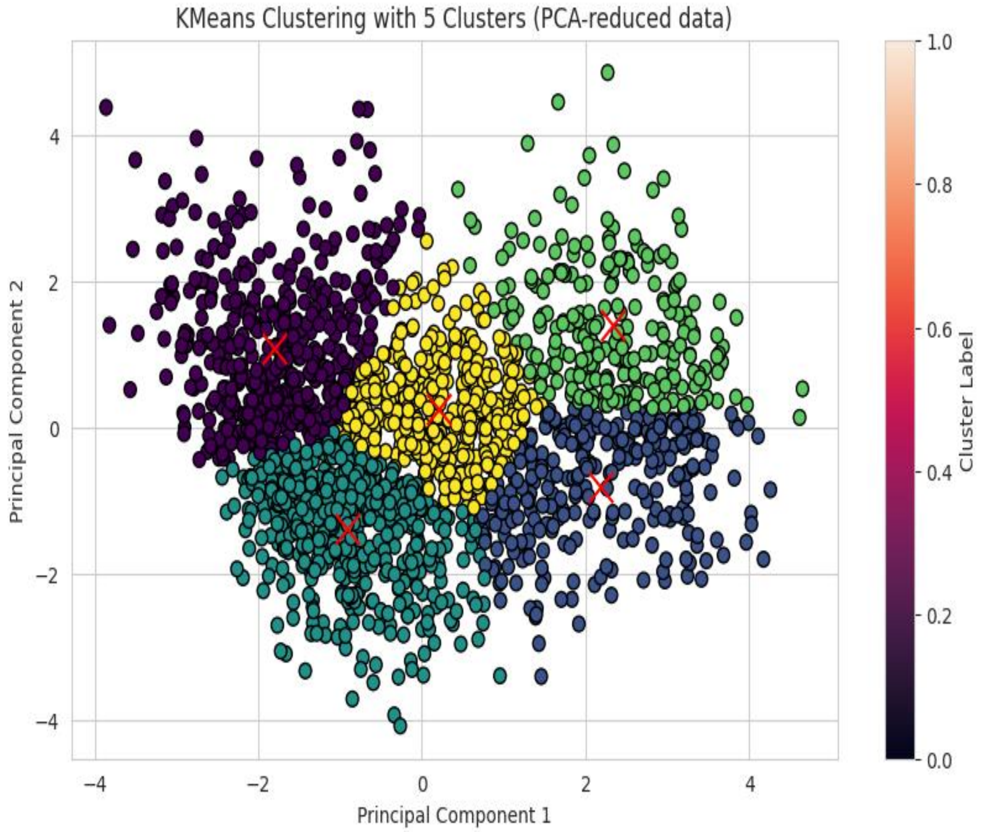
5 rows × 38 columns

FEATURE ENGINEERING

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

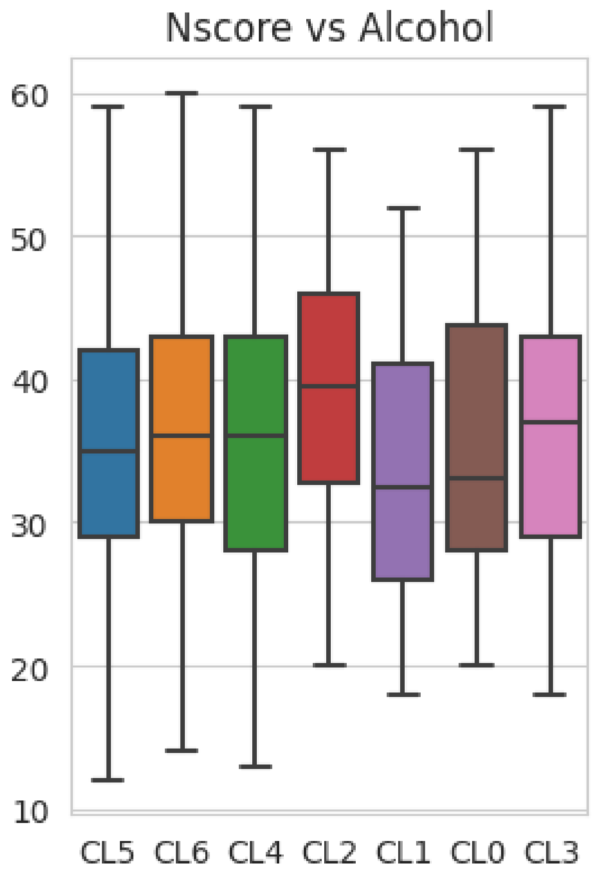
Creating New Variables: We engineered new features that combined different personality traits, hypothesizing that certain trait combinations might better predict drug use.

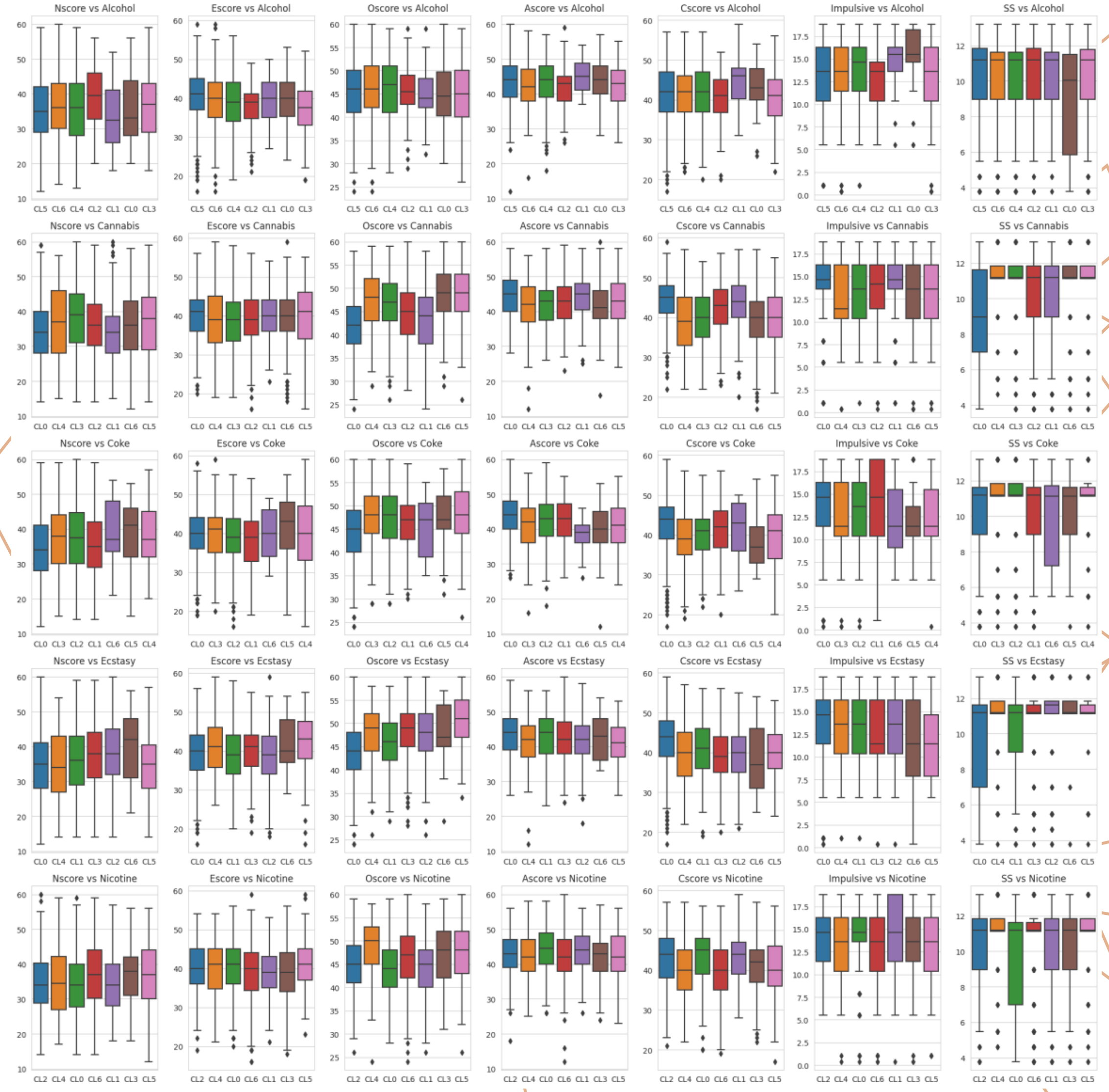




EXPLORATORY DATA ANALYSIS (EDA)

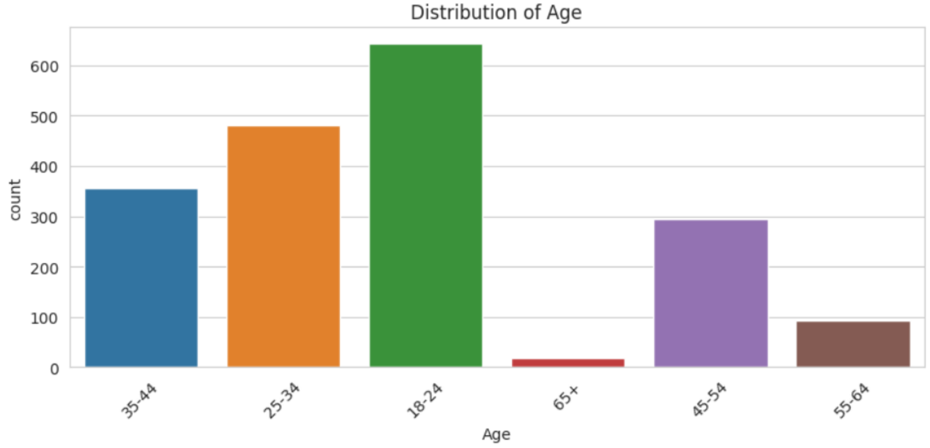
Statistical Summary: Our exploratory analysis began with a statistical overview, examining distributions, means, and variances of different variables.

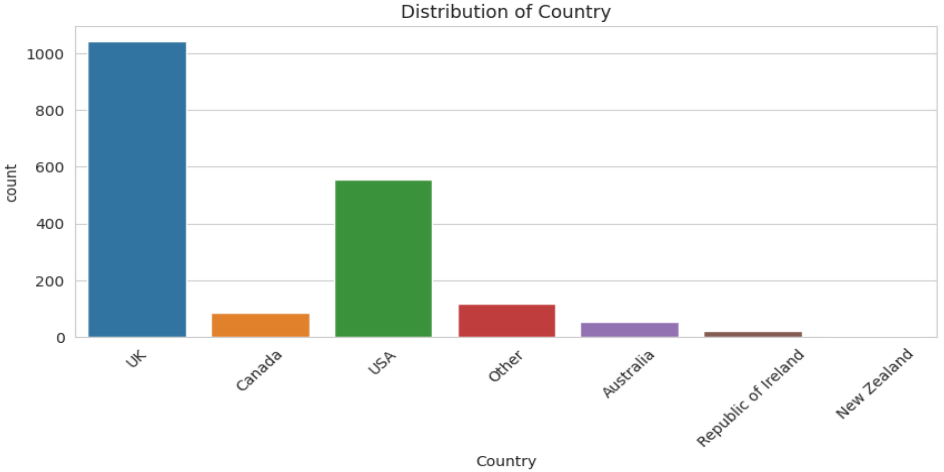




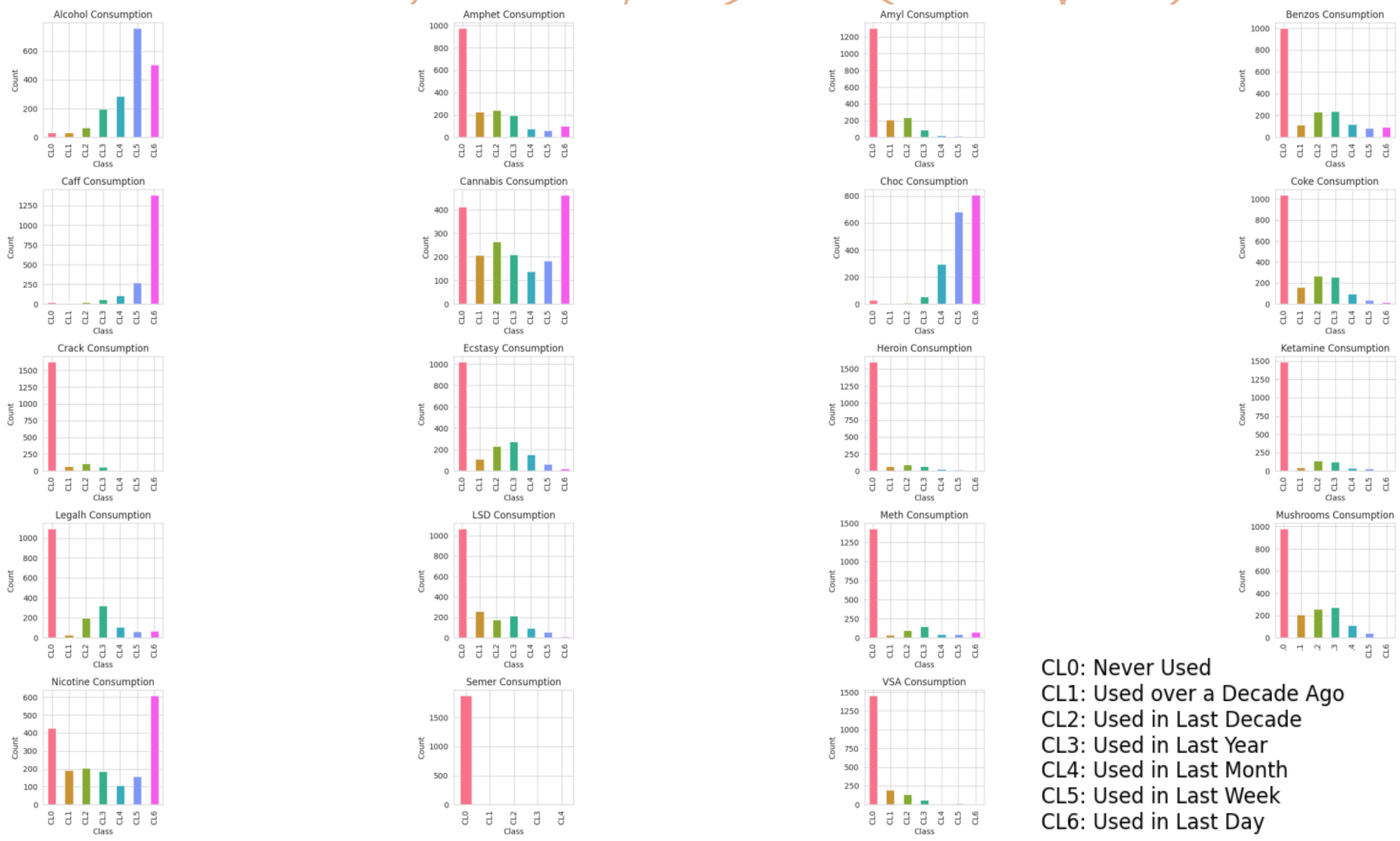
EXPLORATORY DATA ANALYSIS (EDA)

<u>Visualizations:</u> 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

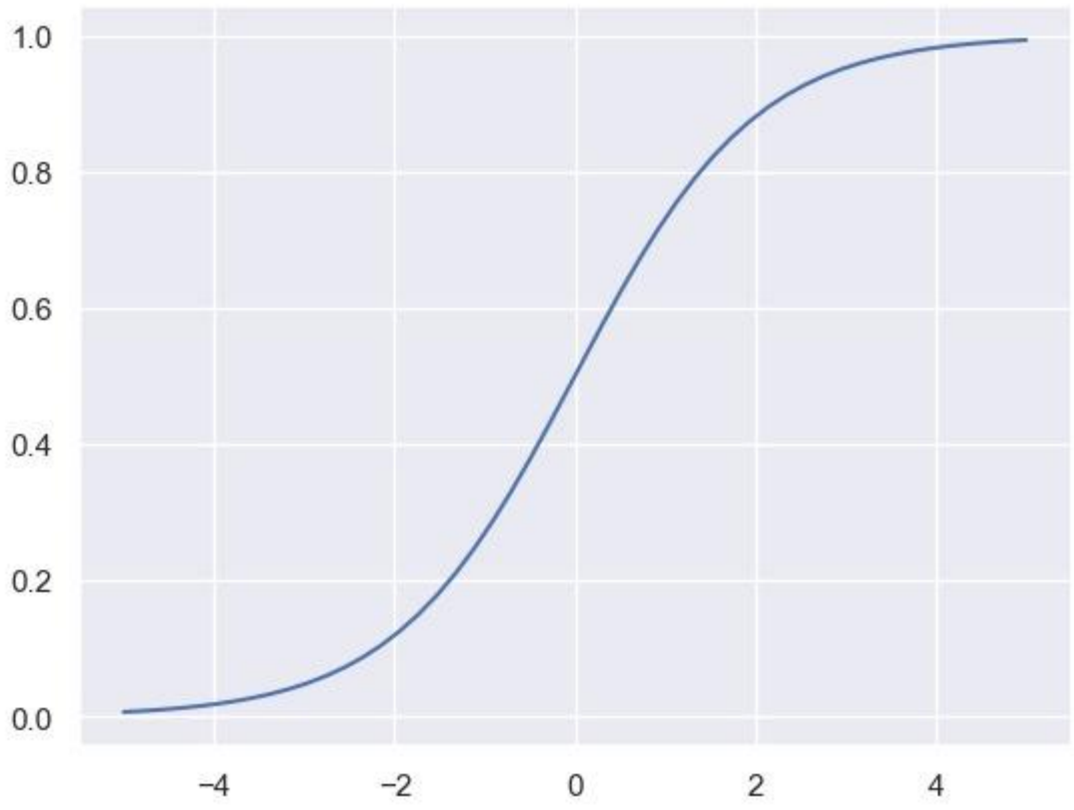


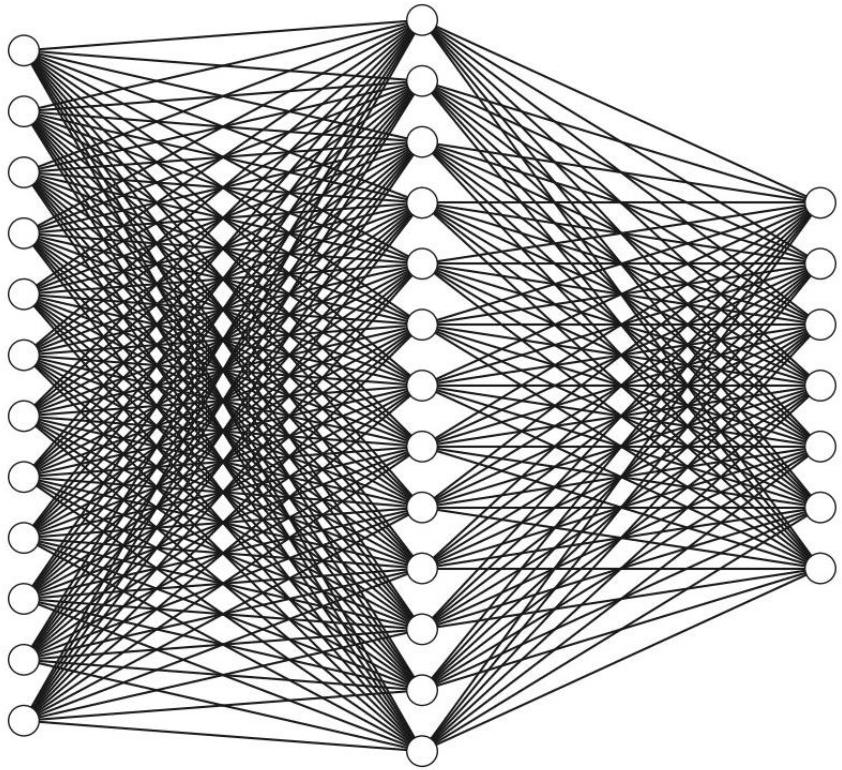
CL0: Never Used
CL1: Used over a Decade Ago
CL2: Used in Last Decade
CL3: Used in Last Year
CL4: Used in Last Month
CL5: Used in Last Week
CL6: Used in Last Day

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.

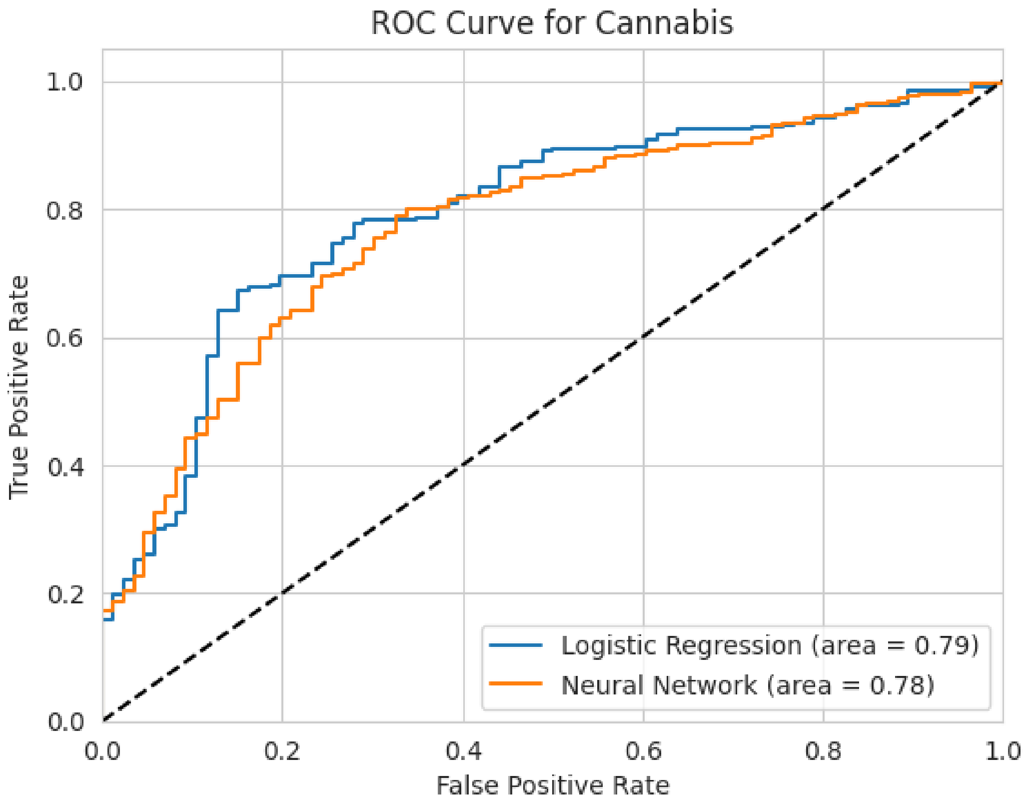


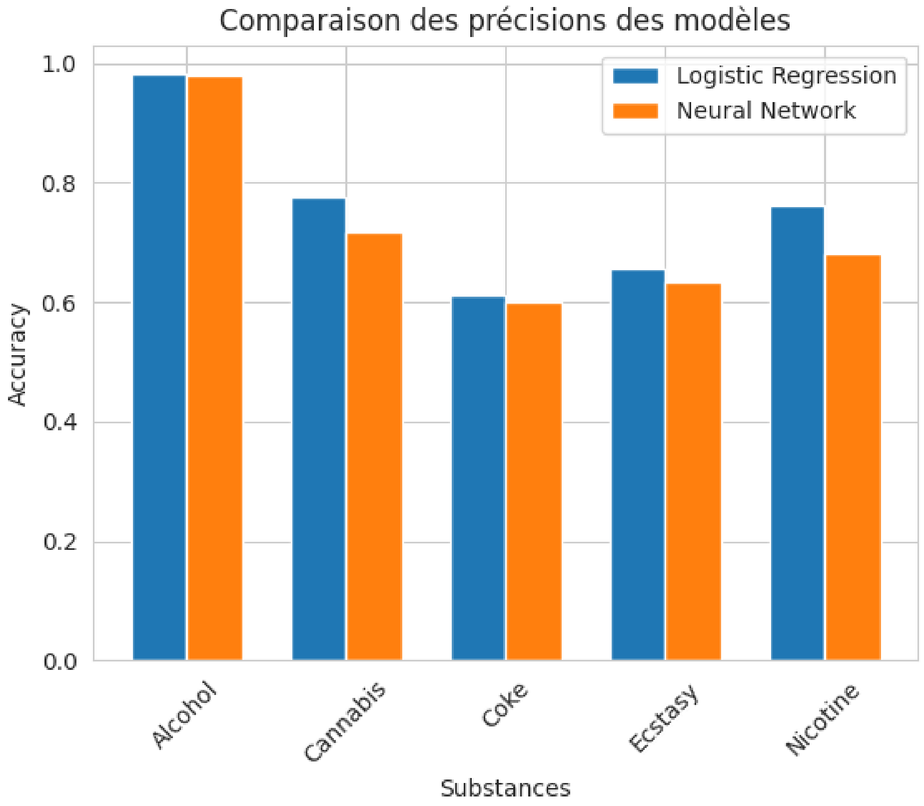


MODEL EVALUATION

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.

PRESENTATION OF STREAMLIT

Streamlit

CONCLUSION AND FUTURE WORK

- **<u>Concluding Thoughts:</u>** In conclusion, our study provides valuable insights into the relationship between personality traits and drug consumption, with potential applications in psychology and public health.
- **<u>Suggestions for Future Research:</u>** Future research could explore deeper into subtypes of personality traits, incorporate longitudinal data, or examine the influence of external factors like socio-economic status.