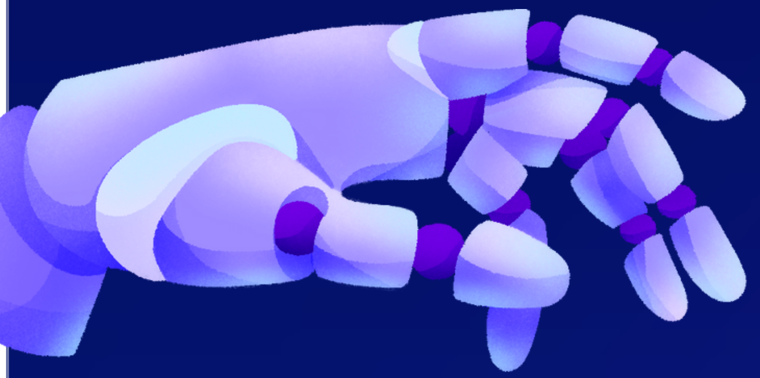
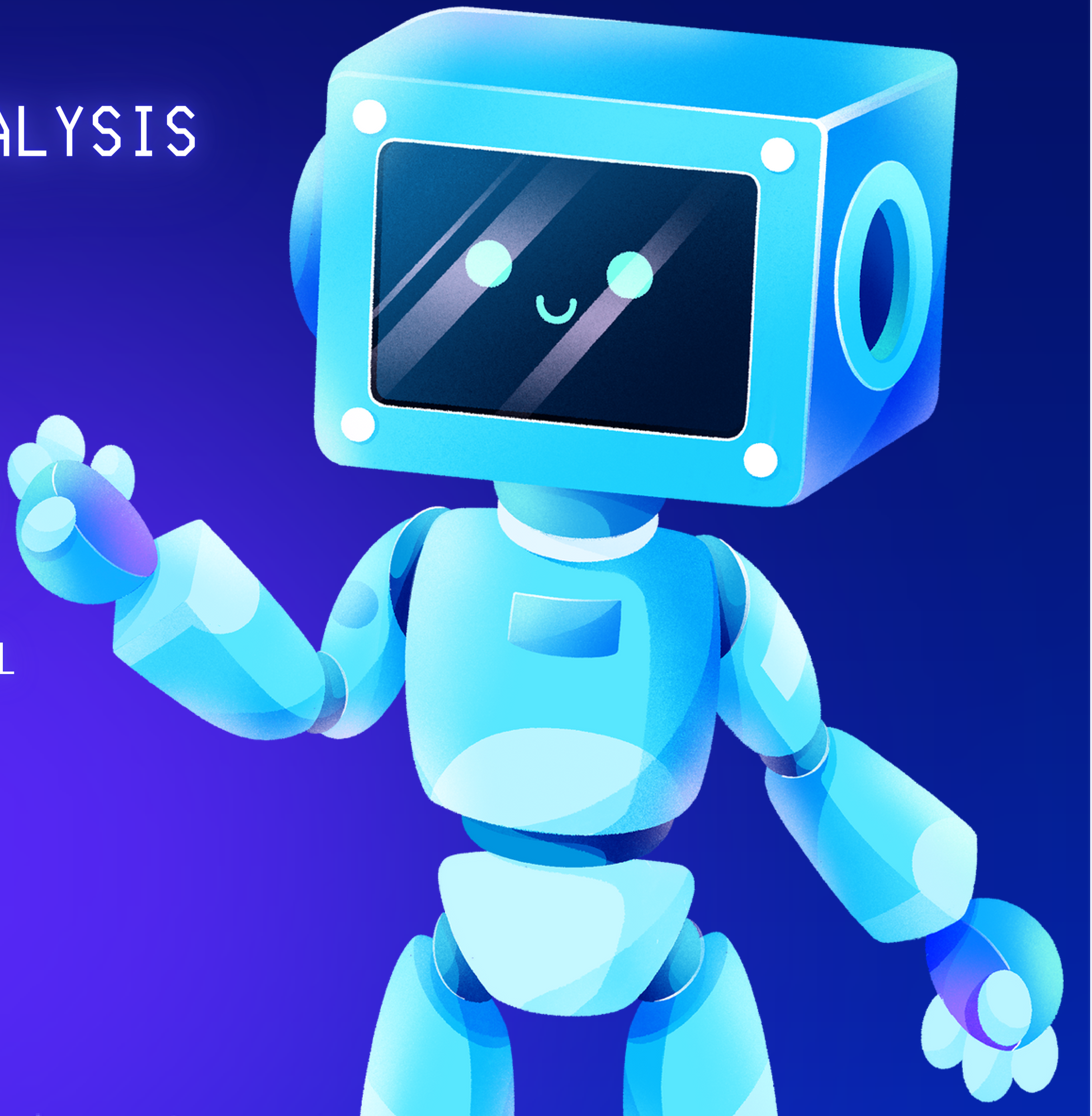


PYTHON FOR DATA ANALYSIS

DRUG CONSUMPTION ANALYSIS AND PREDICTION



By Adam NASSIR, Gaspard LUCIEN and Malvin NOËL

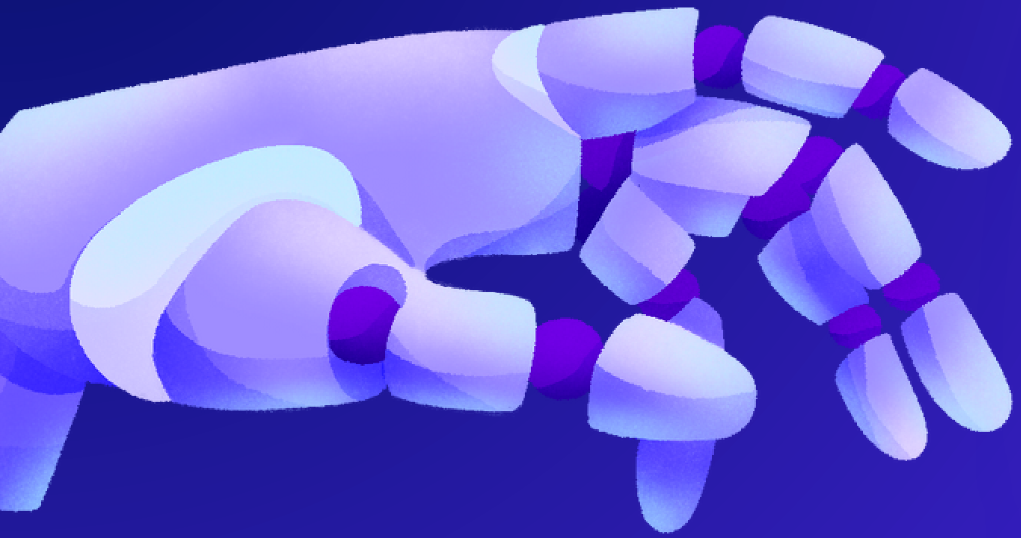
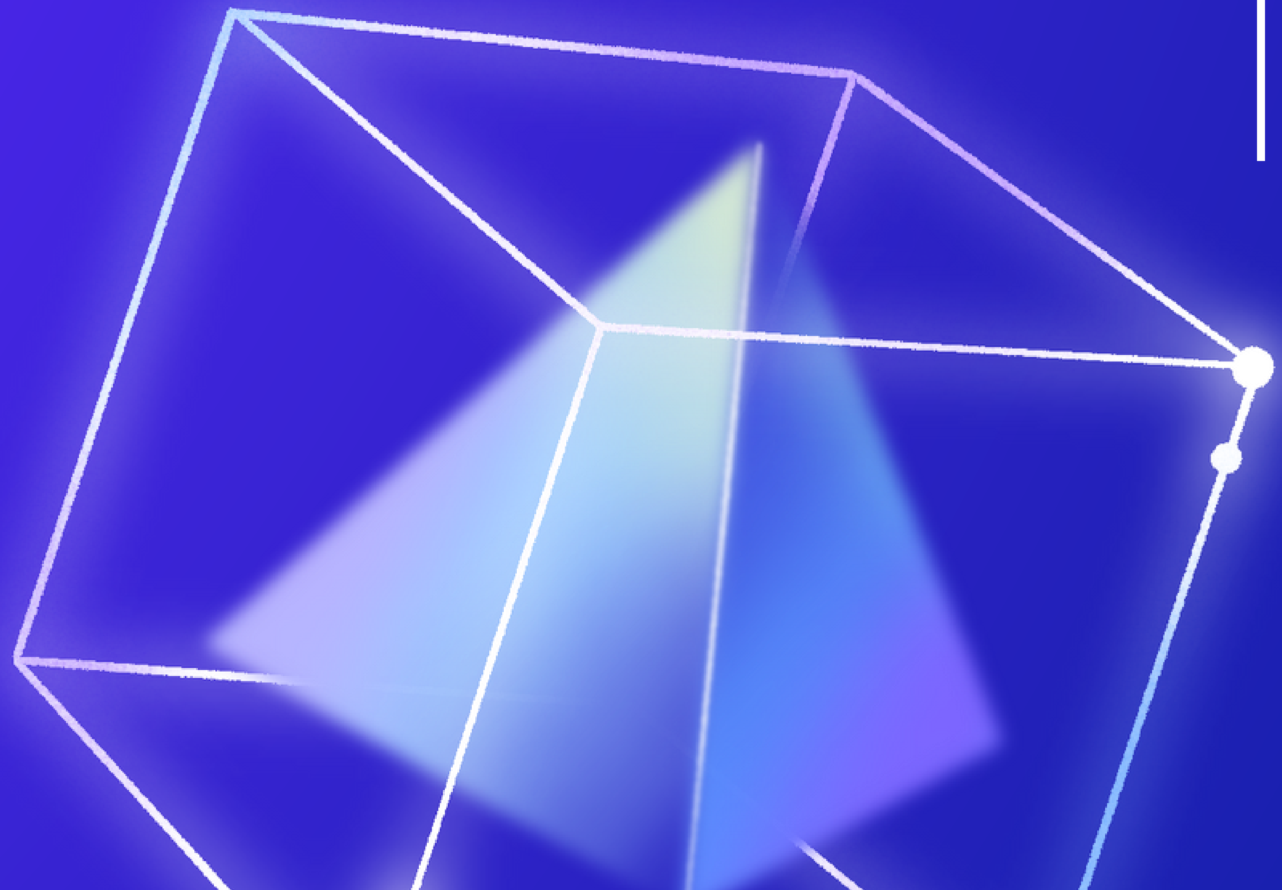


TABLE OF CONTENTS



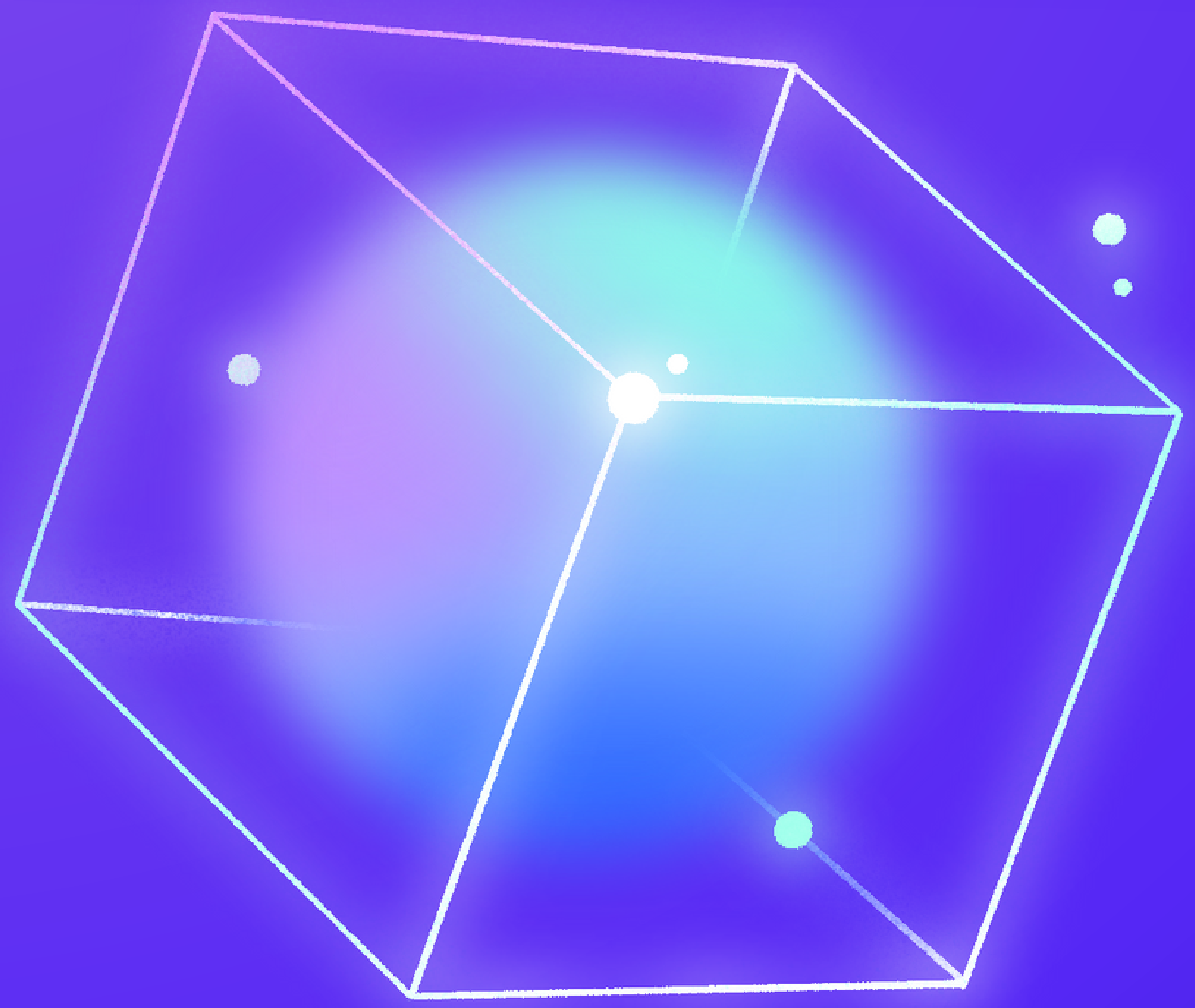
INTRODUCTION

- Developing a model to forecast drug consumption patterns

- Examining demographic, psychological, physical, and educational attributes to understand their influence on drugs use

- Demonstration of statistical relationships between contributing factors and drug use.

DATA PREPROCESSING



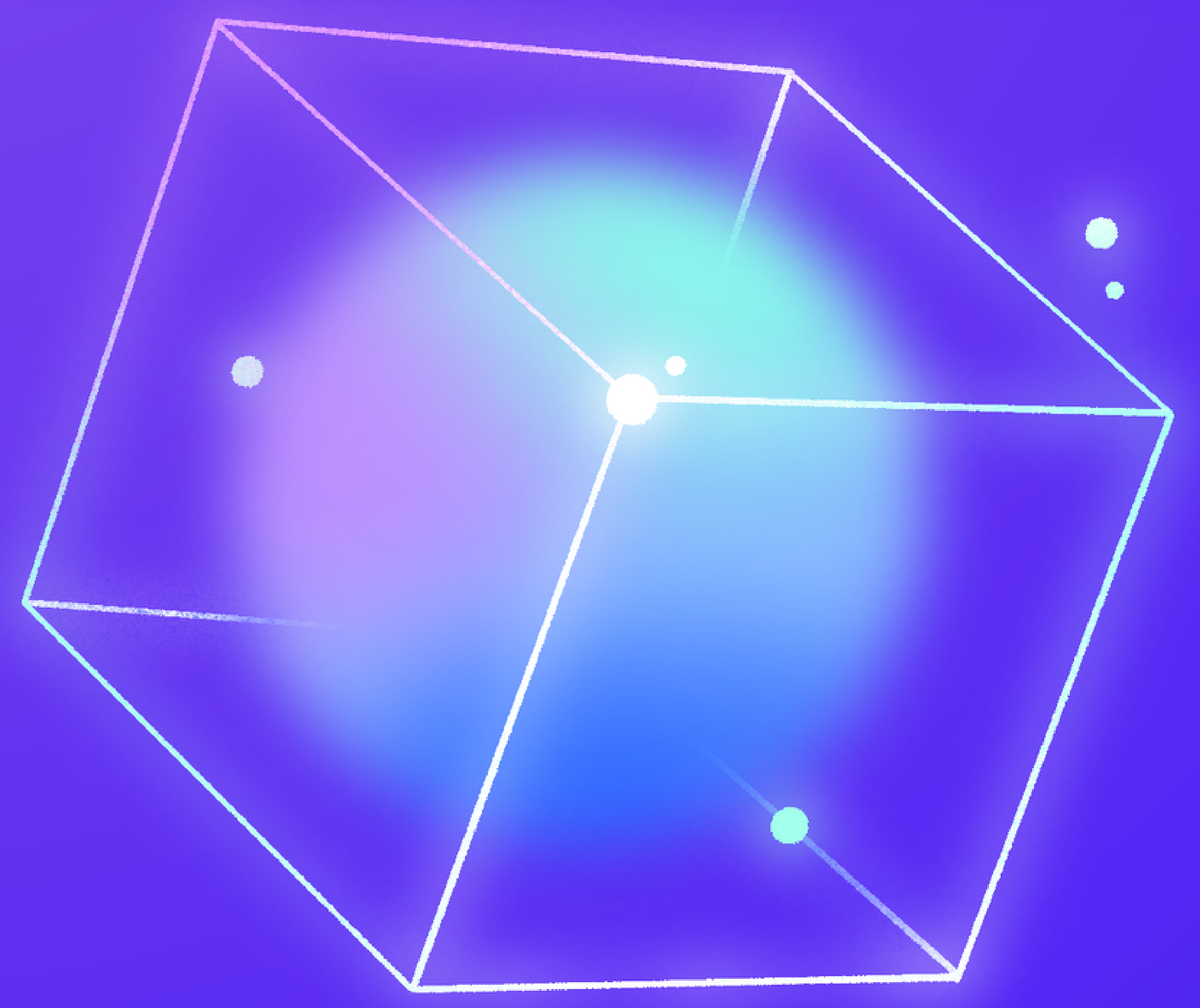
Objective

To prepare and refine a comprehensive dataset for analysis.

Approach

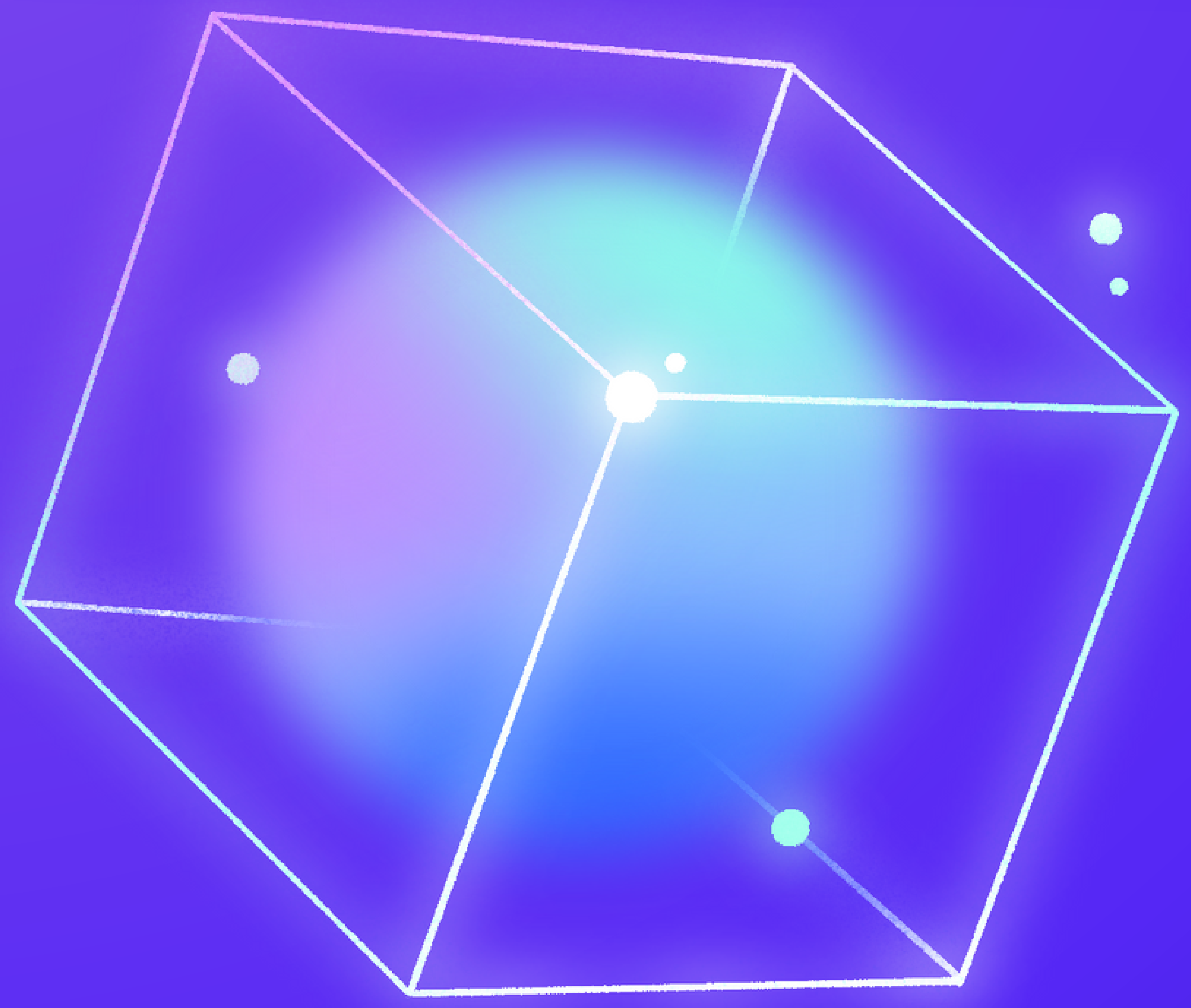
La

DATA PREPROCESSING



Feature	Description
ID	Identification
Age	Age range of participant
Gender	Male or Female
Education	Level of education
Country	Country of orgin
Ethnicity	Ethnicity/Race of participant
Nscore	Quantified NEO Five-Factor Inventory Neuroticism score
Escore	Quantified NEO Five-Factor Inventory Extraversion score
Oscore	Quantified NEO Five-Factor Inventory Openness to experience score
Ascore	Quantified NEO Five-Factor Inventory Agreeableness score
Cscore	Quantified NEO Five-Factor Inventory Conscientiousness score
Impulsive	Quantified BIS-11 impulsiveness score
SS	Quantified Impulsive Sensation Seeking score
Drug	Various drugs were examined and measured in terms of frequency of use

DATA PREPROCESSING



ACTIONS REALIZED ON THE DATA

- SET THE COLUMN'S TITLES
- CHECK FOR MISSING VALUES
- CHECK FOR DUPLICATE VALUES
- SET THE DATA LABELS

DATA VISUALISATION

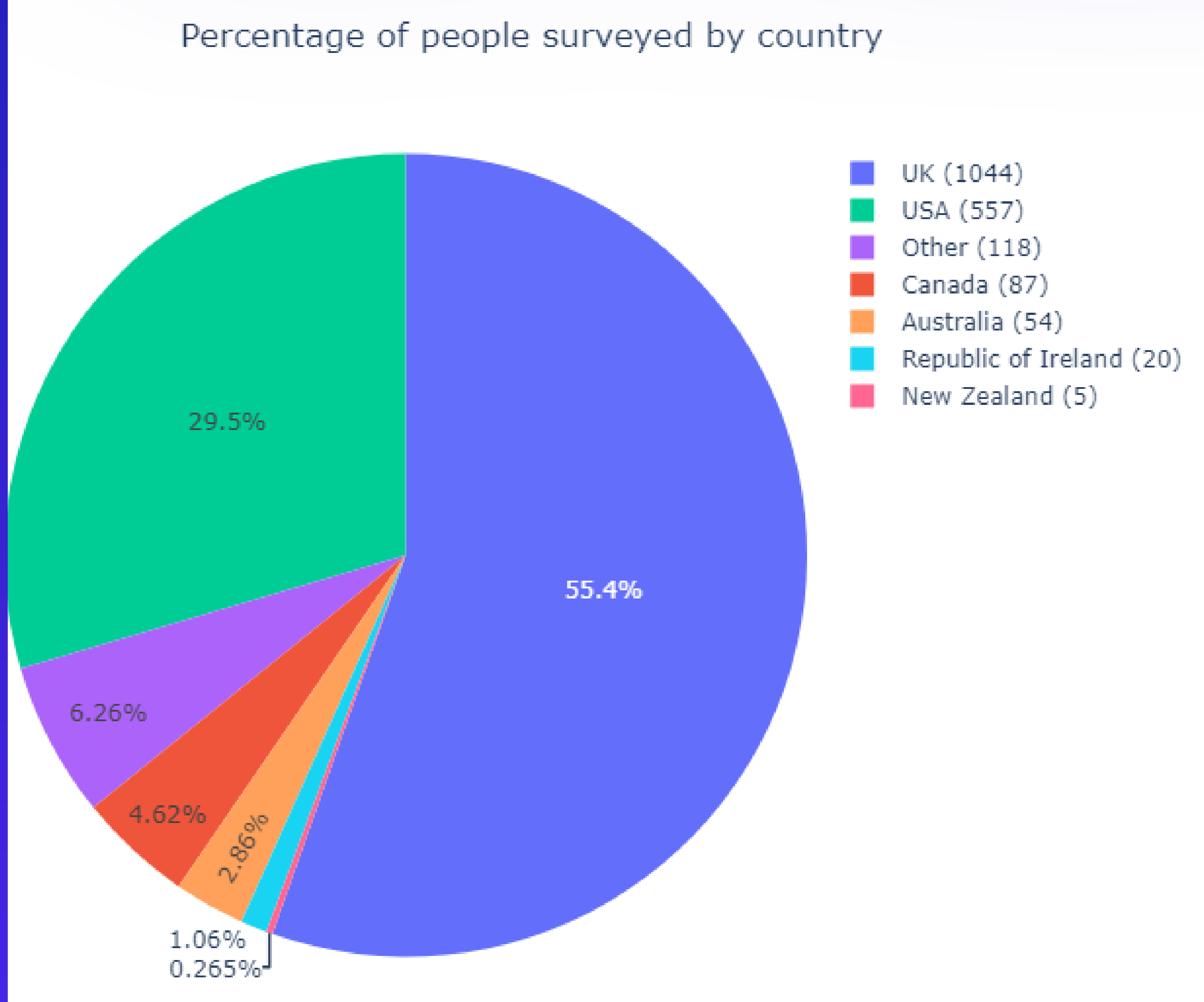
Objective

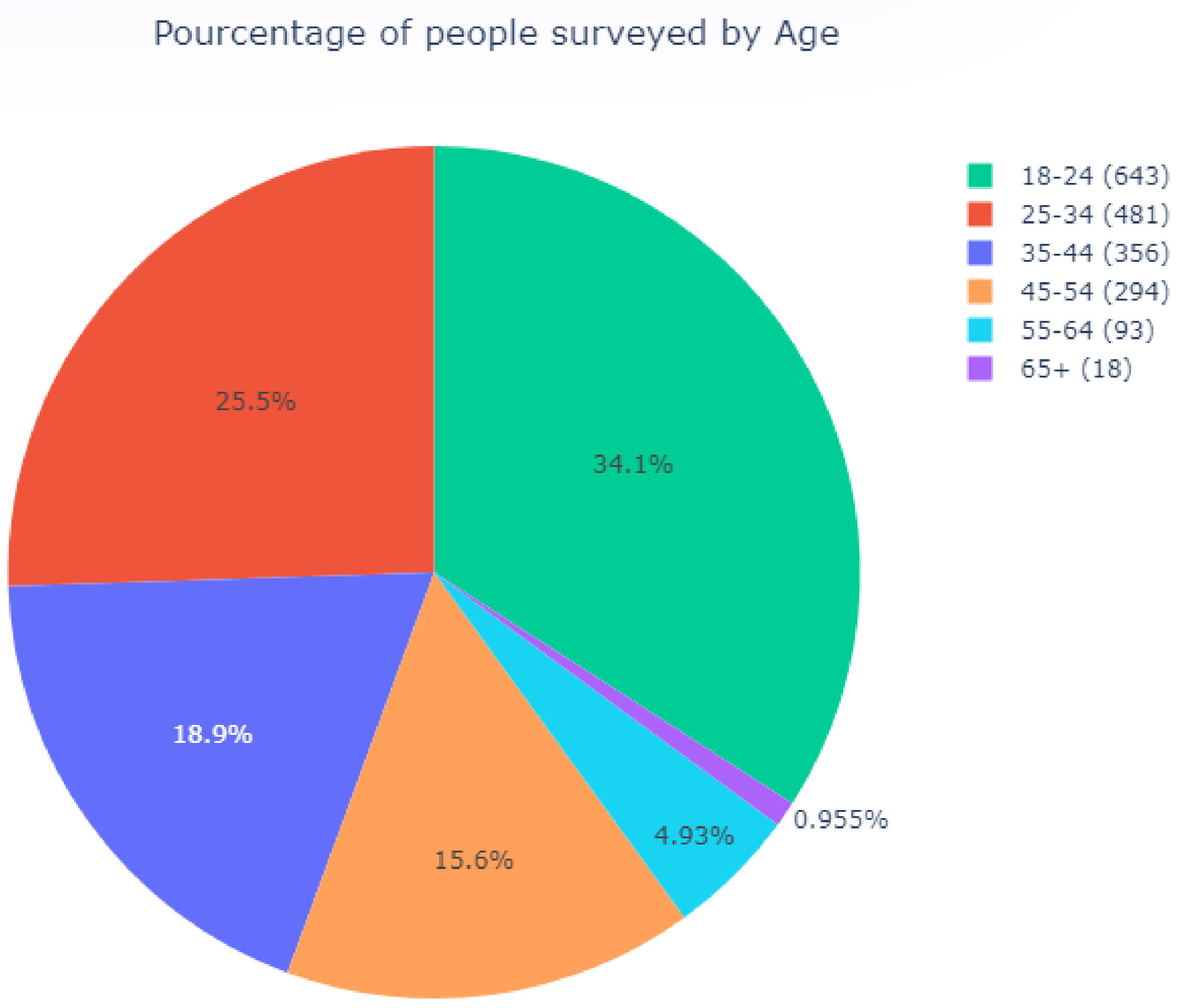
To translate the dataset into understandable and insightful visual formats.

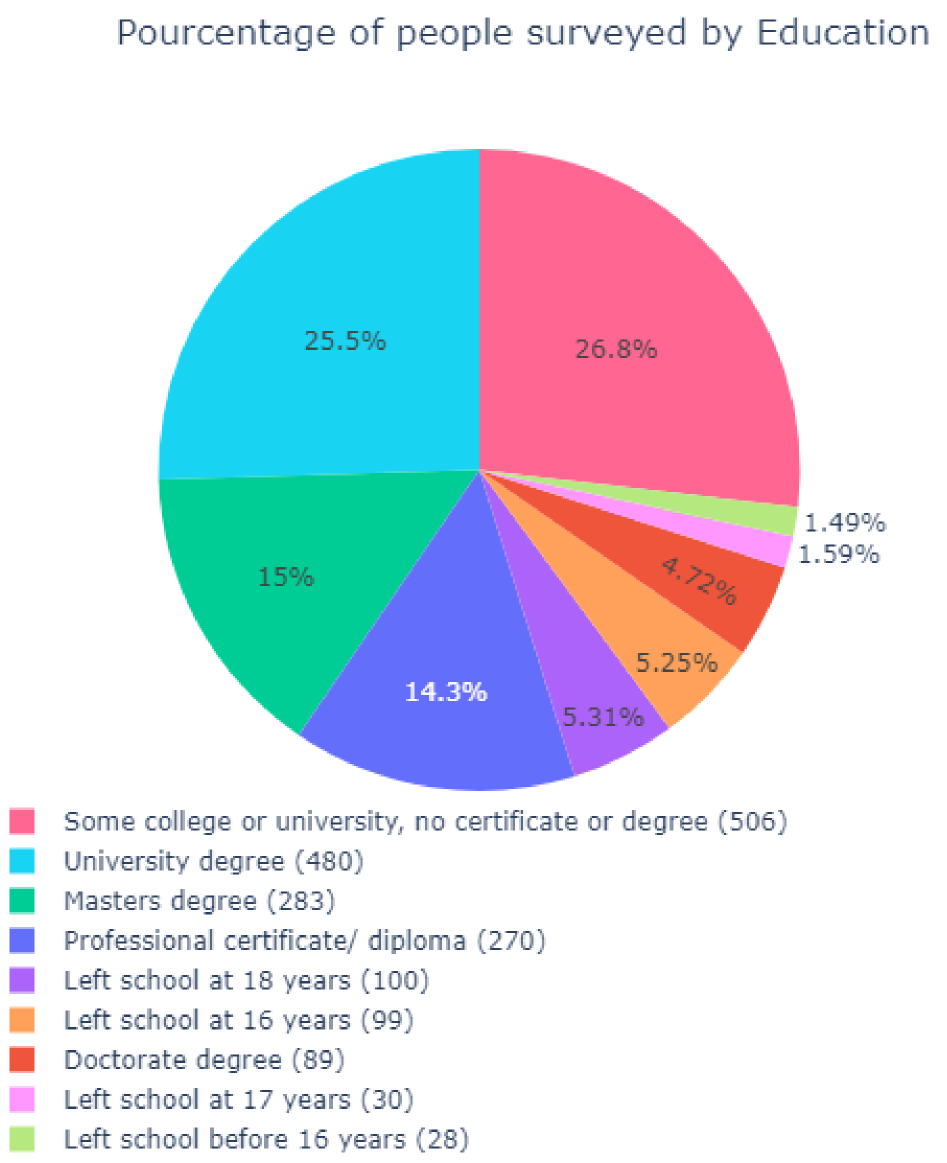
Approach

We aim to uncover hidden trends and patterns, making the data accessible and interpretable for both technical and non-technical audiences.

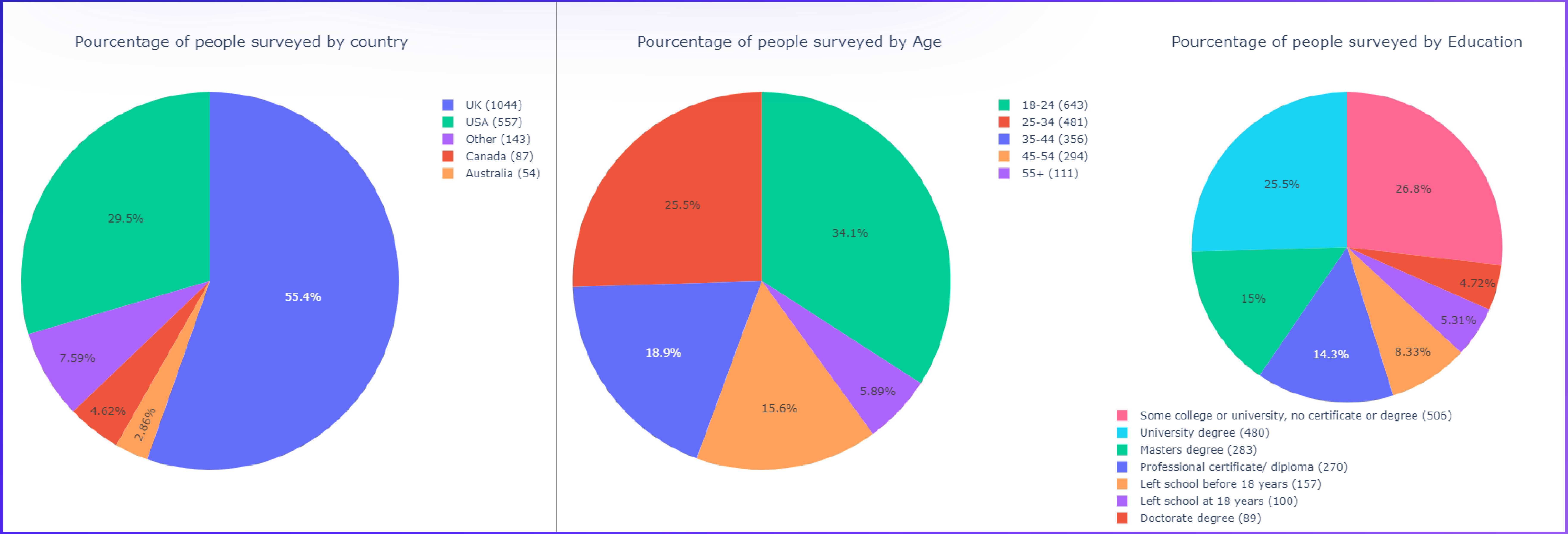
DATA VISUALISATION





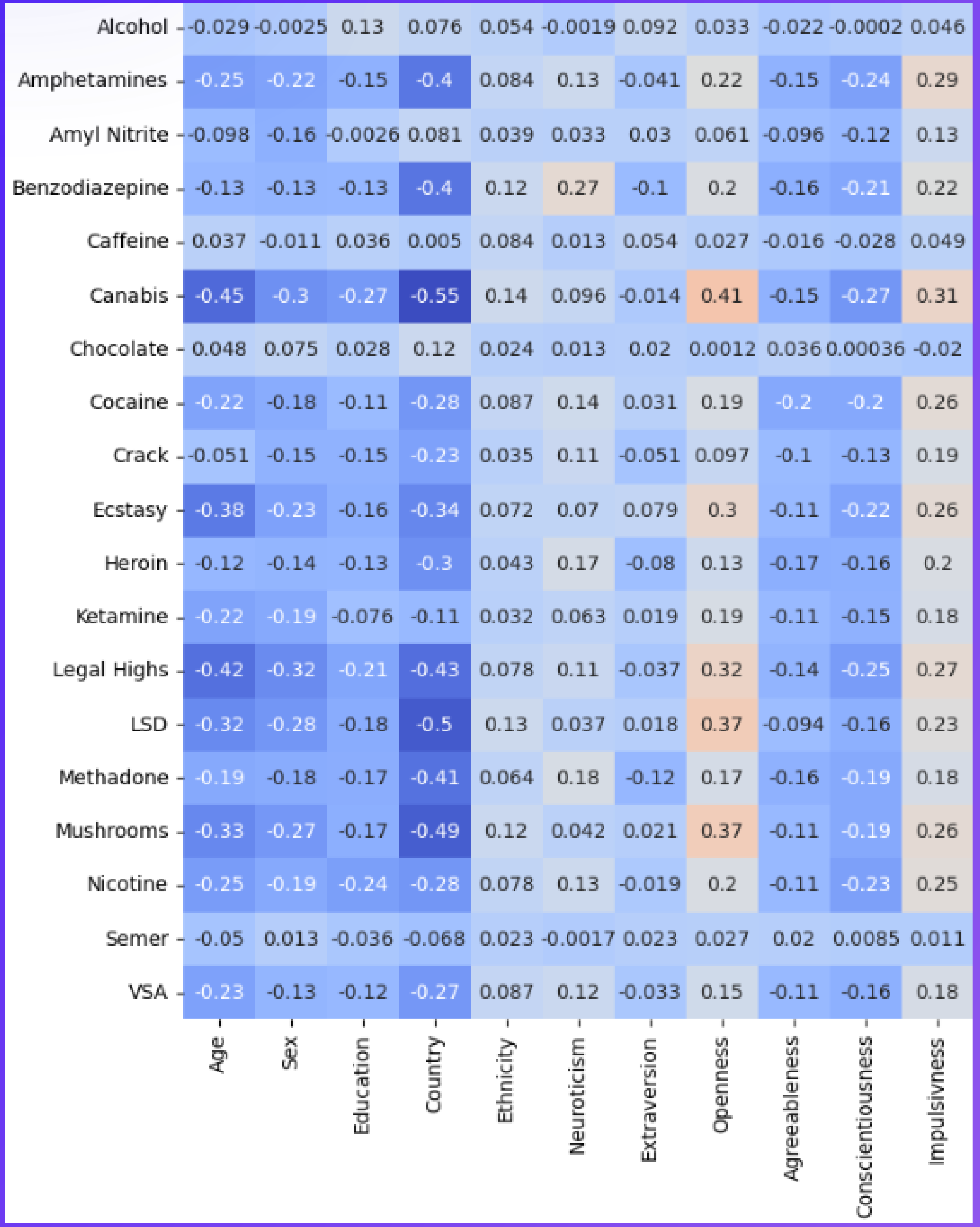


DATA VISUALISATION AFTER REARRENGEMENT

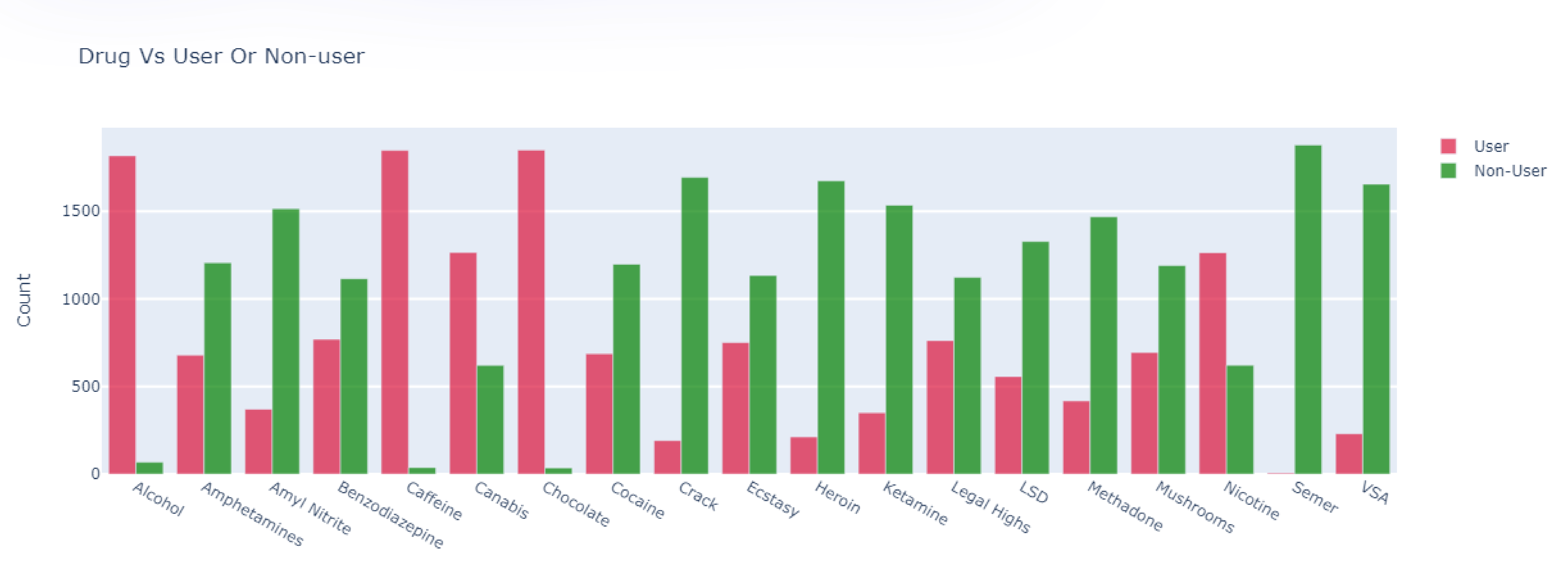


DATA VISUALISATION

Heat Map Showing the relations between the social situation data and the drugs consumption data



DATA VISUALISATION



DATA PREDICTION

Objective

To predict future drug consumption patterns using machine learning.

Approach

We apply RandomForestClassifier, LabelEncoder and binary encoding.

DATA PREDICTION :

DRUG USE PREDICTION MODEL BASED ON DEMOGRAPHIC, PSYCHOLOGICAL, PHYSICAL AND EDUCATIONAL FACTORS.

Profile features :

Age
Sex
Education
Country
Ethnicity
Neuroticism
Extraversion
Openness
Agreeableness
Conscientiousness

Prediction of drug consumption :

Alcohol
Amphetamines
Amyl Nitrite
Benzodiazepine
Caffeine
Canabis
Chocolate
Cocaine
Crack
Ecstasy
Heroin
Ketamine
Legal Highs
LSD Methadone
Mushrooms
Nicotine Semer
VSA

DATA PREDICTION : EXAMPLE

Profile features

Profile: Age 35-44
Sex Female
Education Professional diploma
Country UK
Ethnicity Mixed-White/Asian
Neuroticism Average
Neuroticism
Extraversion Average
Openness Average Openness
Agreeableness Average
Conscientiousness Average

Probability of having take at least one :

Alcohol: 99.50%
Amphetamines: 87.57%
Amyl Nitrite: 8.16%
Benzodiazepine: 81.85%
Caffeine: 100.00%
Canabis: 16.32%
Chocolate: 98.83%
Cocaine: 4.00%
Crack: 0.00%
Ecstasy: 2.36%
Heroin: 1.00%
Ketamine: 1.53%
Legal Highs: 2.95%
LSD: 1.83%
Methadone: 0.00%
Mushrooms: 2.30%
Nicotine: 86.50%
Semer: 0.00%
VSA: 4.25%

DATA PREDICTION :

DRUG USE PREDICTION MODEL BASED ON DEMOGRAPHIC, PSYCHOLOGICAL, PHYSICAL AND EDUCATIONAL FACTORS.

Precision for Alcohol: 99.46%
Precision for Amphetamines: 86.76%
Precision for Amyl Nitrite: 84.77%
Precision for Benzodiazepine: 88.10%
Precision for Caffeine: 99.62%
Precision for Canabis: 93.40%
Precision for Chocolate: 99.52%
Precision for Cocaine: 84.54%
Precision for Crack: 90.24%
Precision for Ecstasy: 87.12%

Precision for Heroin: 92.34%
Precision for Ketamine: 86.88%
Precision for Legal Highs: 90.34%
Precision for LSD: 88.35%
Precision for Methadone: 83.49%
Precision for Mushrooms: 87.69%
Precision for Nicotine: 90.86%
Precision for Semer: 100.00%
Precision for VSA: 84.85%

Average Precision: 90.44%

DATA PREDICTION :

DRUG USE PREDICTION MODEL WITH COMMON DRUGS (ALCOHOL, CAFFEINE, CHOCOLATE, NICOTINE)

Profile features

Age
Sex
Education
Country
Ethnicity
Neuroticism
Extraversion
Openness
Agreeableness
Conscientiousness

Nicotine
Chocolate
Caffeine
Alcohol

Prediction of drug consumption

Amphetamines
Amyl Nitrite
Benzodiazepine
Canabis
Cocaine
Ecstasy
Heroin
Ketamine
Legal Highs
LSD Methadone

Mushrooms Nicotine
Semer VSA
Crack

DATA PREDICTION : EXAMPLE

DRUG USE PREDICTION MODEL WITH COMMON DRUGS (ALCOHOL, CAFFEINE, CHOCOLATE, NICOTINE)

Profile features

Age 35-44
Sex Female
Education Professional diploma
Country UK
Ethnicity Mixed-White/Asian
Neuroticism Average
Neuroticism
Extraversion Average
Openness Average Openness
Agreeableness Average
Conscientiousness Average
Alcohol Used in Last Week Caffeine Used in Last Day Nicotine Used in Last Decade Chocolate Used in Last Week

Probability of having take at least one time :

Amphetamines: 84.00%
Amyl Nitrite: 9.00%
Benzodiazepine: 86.00%
Canabis: 15.00%
Cocaine: 11.00%
Crack: 1.00%
Ecstasy: 9.00%
Heroin: 1.00%
Ketamine: 10.00%
Legal Highs: 5.00%
LSD: 8.00%
Methadone: 0.00%
Mushrooms: 7.00%
Semer: 0.00%
VSA: 8.00%

DATA PREDICTION :

DRUG USE PREDICTION MODEL WITH COMMON DRUGS (ALCOHOL, CAFFEINE, CHOCOLATE, NICOTINE)

Precision for Amphetamines: 92.70%
Precision for Amyl Nitrite: 94.24%
Precision for Benzodiazepine: 93.87%
Precision for Canabis: 97.00%
Precision for Cocaine: 92.40%
Precision for Crack: 95.59%
Precision for Ecstasy: 92.11%
Precision for Heroin: 96.30%

Precision for Ketamine: 94.63%
Precision for Legal Highs: 94.17%
Precision for LSD: 92.87%
Precision for Methadone: 92.94%
Precision for Mushrooms: 93.44%
Precision for Semer: 100.00%
Precision for VSA: 94.81%

Average Precision: 94.47%

DATA PREDICTION :

DETERMINING THE BEST MODEL AND FEATURES

After standardizing the values in the heatmap and not taking into account specific features

Alcohol	
Age	-0.493926
Sex	-0.370167
Education	0.364151
Country	0.086219
Ethnicity	-0.053963
Neuroticism	-0.366489
Extraversion	0.161909
Agreeableness	-0.478270
Conscientiousness	-0.357074
Sensation Seeing	0.252624
Alcohol	5.254320

Amphetamines	-0.411966
Amyl Nitrite	0.130753
Benzodiazepine	-0.413322
Caffeine	0.356214
Cocaine	0.146659
Crack	-0.468818
Ecstasy	0.056900
Heroin	-0.545480
Ketamine	0.013752
LSD	-0.287919
Methadone	-0.799265
Nicotine	-0.006619
Semer	-0.558521

DATA PREDICTION :

THREE MODELS

Random Forest Classifier

rf_params = {'n_estimators': [50, 100, 150],
'max_depth': [None, 10, 20]
,'min_samples_split': [2, 5, 10],
'min_samples_leaf': [1, 2, 4]}

Scalable Linear Support Vector Machine

svc_params = {'C': [0.1, 1, 10],
'kernel': ['linear', 'rbf'],
'gamma': ['scale', 'auto']}

K-Neighbors Classifier

knn_params = {'n_neighbors':[3, 5, 7], 'weights':['uniform', 'distance']}

Drug	RFF	Best_params_rff
Alcohol	0.94164456	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
Amphetamine	0.9204244	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 100}
Amyl Nitrite	0.93633952	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
Benzodiazepir	0.9469496	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
Caffeine	0.95490716	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
Canabis	0.93633952	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 150}
Chocolate	0.94429708	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
Cocaine	0.92307692	{'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 50}
Crack	0.98143236	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 100}
Ecstasy	0.95225464	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 150}
Heroin	0.96816976	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
Ketamine	0.95490716	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
Legal Highs	0.90981432	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
LSD	0.9602122	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 150}
Methadone	0.94960212	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 100}
Mushrooms	0.96816976	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
Nicotine	0.92572944	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
Semer	0.99734748	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
VSA	0.9469496	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 50}

Drug	KNN	Best_params_knn
Alcohol	0.59681698	{'n_neighbors': 5, 'weights': 'uniform'}
Amphetamine	0.71618037	{'n_neighbors': 7, 'weights': 'uniform'}
Amyl Nitrite	0.76392573	{'n_neighbors': 7, 'weights': 'distance'}
Benzodiazepir	0.75331565	{'n_neighbors': 3, 'weights': 'distance'}
Caffeine	0.76657825	{'n_neighbors': 3, 'weights': 'distance'}
Canabis	0.63925729	{'n_neighbors': 7, 'weights': 'distance'}
Chocolate	0.60212202	{'n_neighbors': 7, 'weights': 'uniform'}
Cocaine	0.72148541	{'n_neighbors': 5, 'weights': 'distance'}
Crack	0.89655172	{'n_neighbors': 3, 'weights': 'distance'}
Ecstasy	0.71618037	{'n_neighbors': 5, 'weights': 'distance'}
Heroin	0.90981432	{'n_neighbors': 3, 'weights': 'distance'}
Ketamine	0.85676393	{'n_neighbors': 5, 'weights': 'distance'}
Legal Highs	0.73474801	{'n_neighbors': 7, 'weights': 'distance'}
LSD	0.73740053	{'n_neighbors': 5, 'weights': 'distance'}
Methadone	0.85676393	{'n_neighbors': 5, 'weights': 'distance'}
Mushrooms	0.65517241	{'n_neighbors': 7, 'weights': 'distance'}
Nicotine	0.6259947	{'n_neighbors': 7, 'weights': 'distance'}
Semer	0.99734748	{'n_neighbors': 3, 'weights': 'uniform'}
VSA	0.80901857	{'n_neighbors': 5, 'weights': 'distance'}

Drug	SVC	Best_params_svc
Alcohol	1	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Amphetamine	1	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
Amyl Nitrite	0.98143236	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Benzodiazepir	1	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
Caffeine	0.99469496	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Canabis	1	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
Chocolate	0.99734748	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Cocaine	0.99734748	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Crack	0.98938992	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
Ecstasy	1	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Heroin	1	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Ketamine	0.99204244	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Legal Highs	1	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
LSD	0.99469496	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Methadone	1	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}
Mushrooms	1	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
Nicotine	1	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
Semer	0.99734748	{'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
VSA	0.9867374	{'C': 1, 'gamma': 'scale', 'kernel': 'linear'}

DATA PREDICTION :

CLASS REPORTS COMPARISON ON ALCOHOL

RF VS KN

	Model 1	Model 2	Difference
precision	0.94	0.58	0.36
recall	0.94	0.6	0.33999999999999997
f1-score	0.93	0.58	0.35000000000000001

RF VS SVC

	Model 1	Model 2	Difference
precision	0.94	0.92	0.019999999999999907
recall	0.94	0.94	0.0
f1-score	0.93	0.93	0.0

KN VS SVC

	Model 1	Model 2	Difference
precision	0.58	0.92	-0.34000000000000001
recall	0.6	0.94	-0.33999999999999997
f1-score	0.58	0.93	-0.35000000000000001

FLASK

Objective

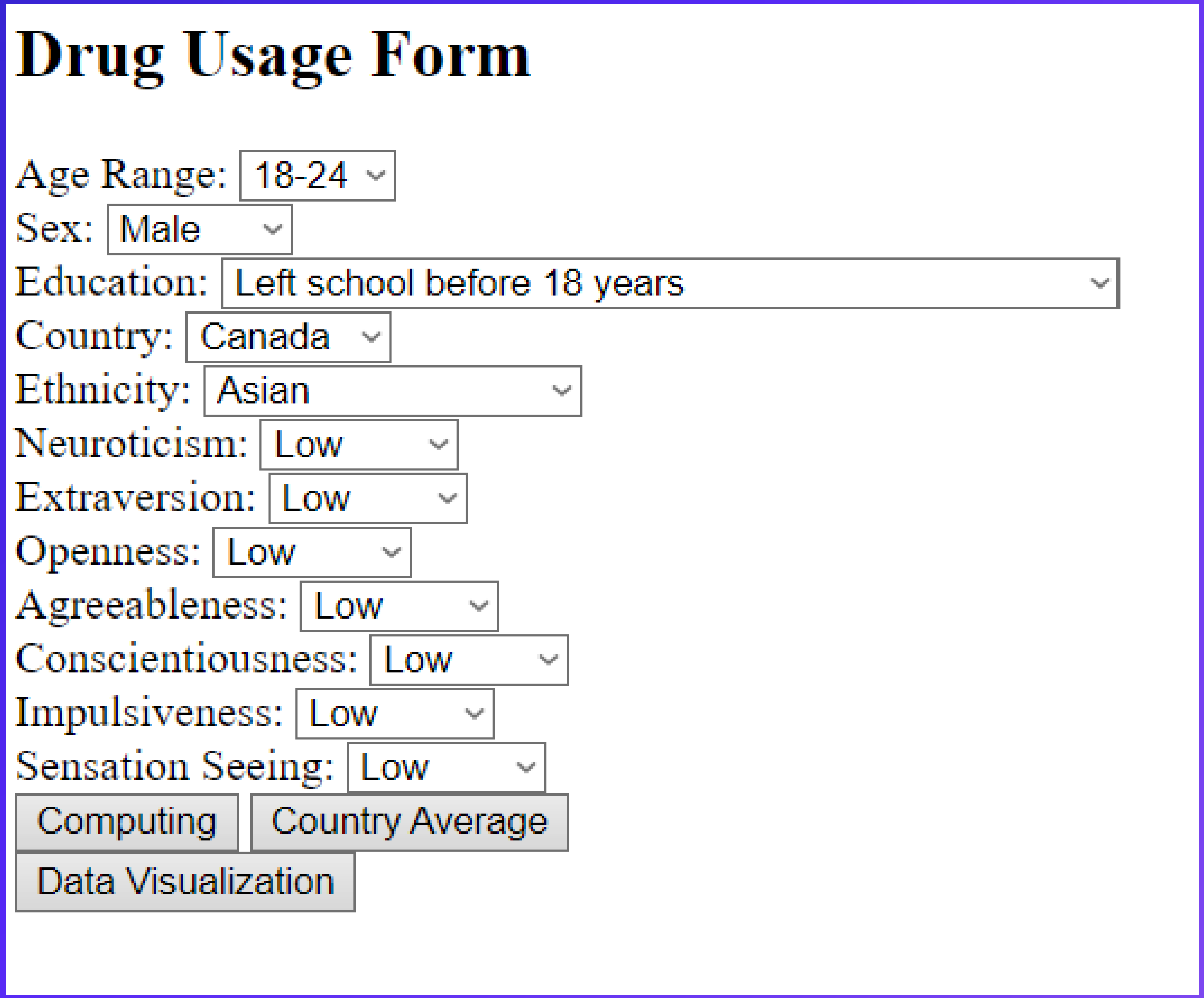
Create an intuitive application to use our predictive model and display our visualizations.

Approach

We redid our preprocessing and modifying dataframes steps to be able to properly re-use our visualizations and percentage prediction model.

FLASK :

APPLICATION FOR PREDICTING DRUG USE BASED ON DEMOGRAPHIC, PSYCHOLOGICAL, PHYSICAL AND EDUCATIONAL FACTORS.



Drug Usage Form

Age Range: 18-24

Sex: Male

Education: Left school before 18 years

Country: Canada

Ethnicity: Asian

Neuroticism: Low

Extraversion: Low

Openness: Low

Agreeableness: Low

Conscientiousness: Low

Impulsiveness: Low

Sensation Seeing: Low

Computing | Country Average

Data Visualization

FLASK :

RESULTS OF PREDICTION

Prediction Result

Alcohol: 71.0%

Amphetamines: 42.0%

Amyl Nitrite: 20.0%

Benzodiazepine: 36.0%

Caffeine: 84.73333333333333%

Canabis: 68.0%

Chocolate: 84.0%

Cocaine: 33.333333333333336%

Crack: 18.0%

Ecstasy: 40.0%

Heroin: 13.0%

Ketamine: 42.0%

Legal Highs: 51.0%

LSD: 31.0%

Methadone: 40.0%

Mushrooms: 36.0%

Nicotine: 50.0%

Semer: 14.000000000000002%

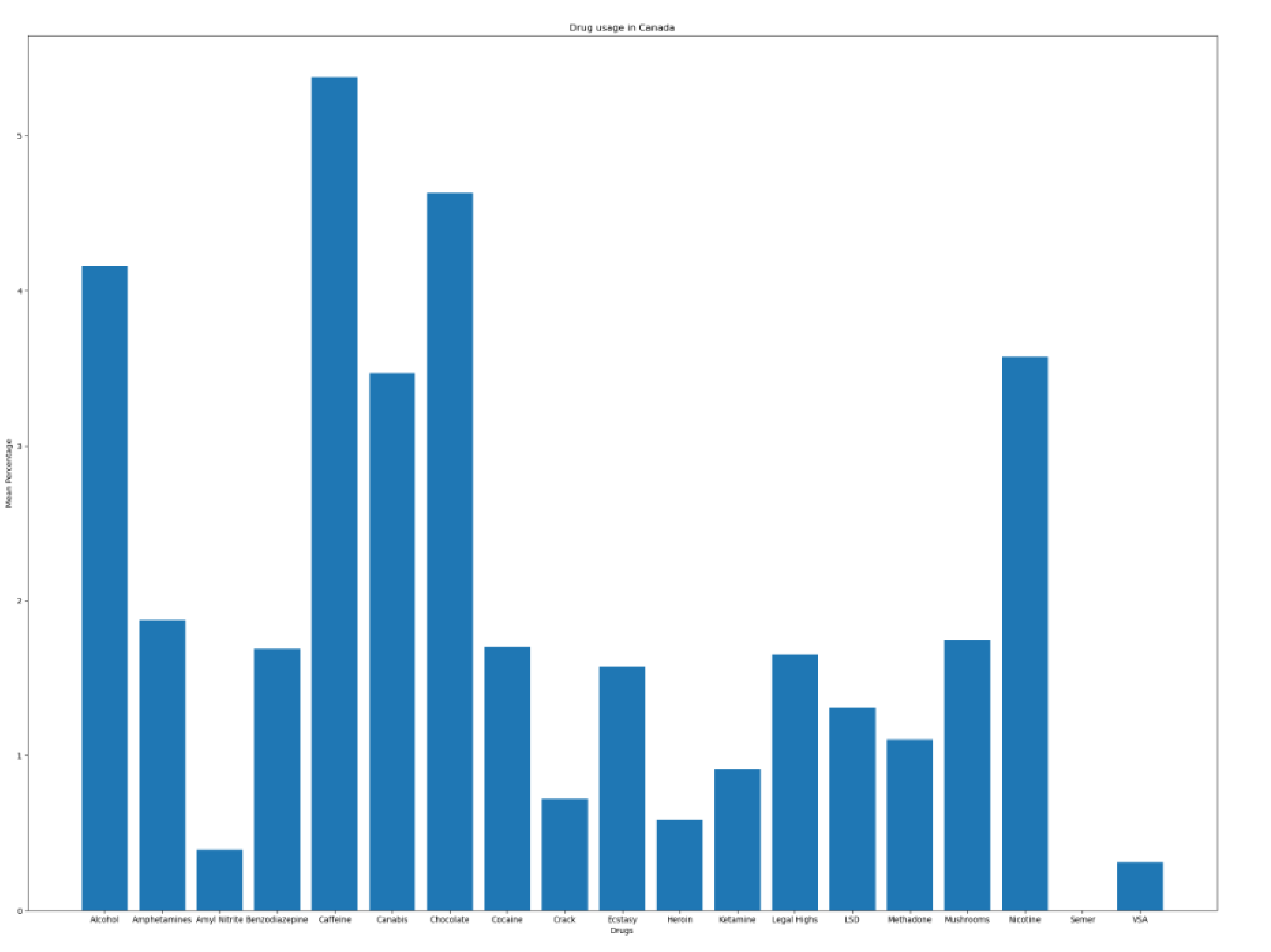
VSA: 34.0%

FLASK :

VISUALIZATIONS

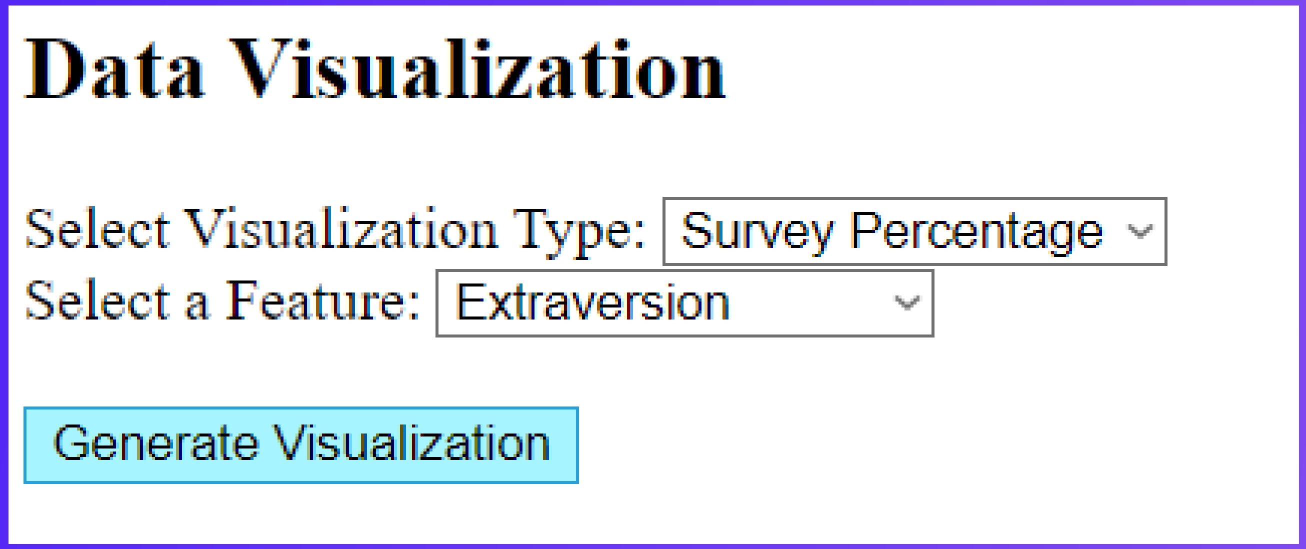
Drug Usage Graph

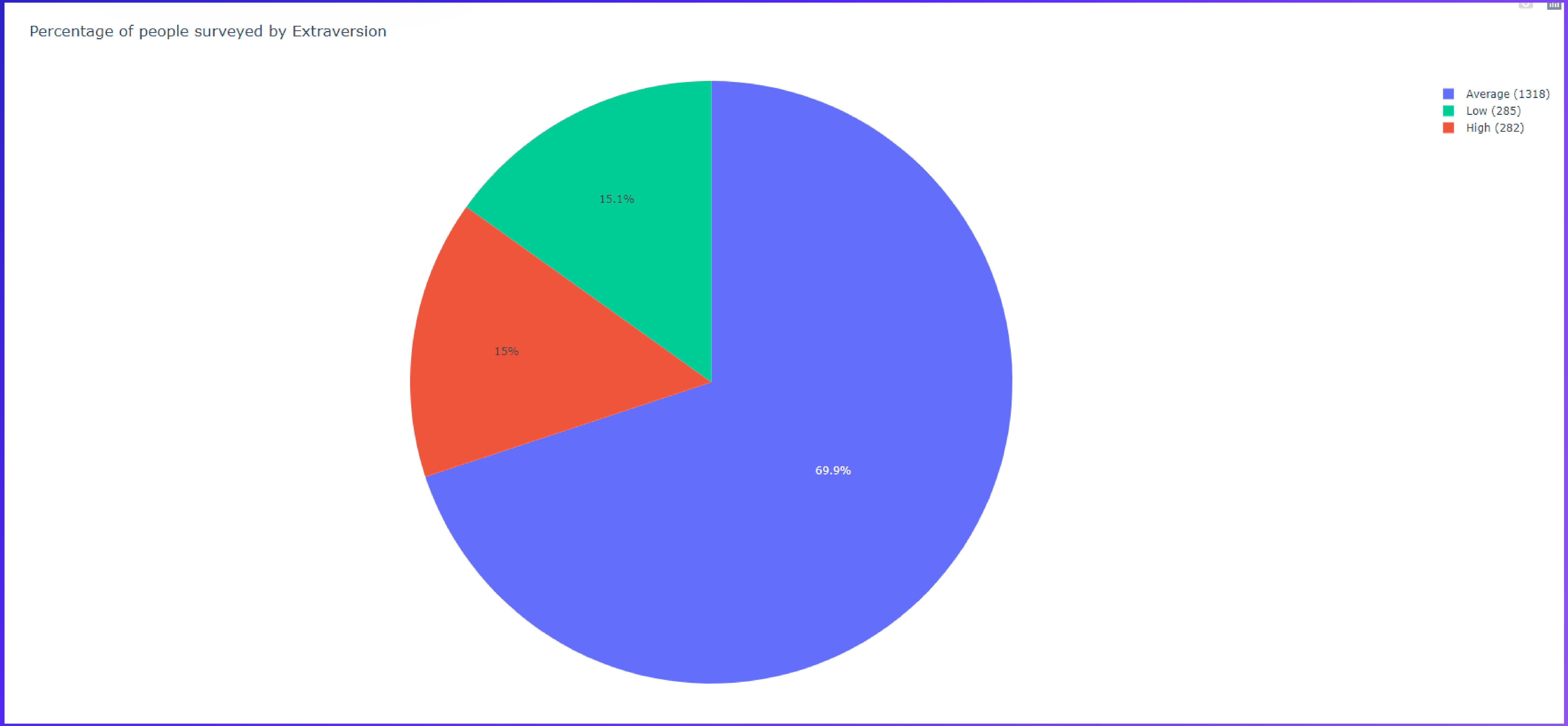
Select a country: Canada Show Graph



FLASK :

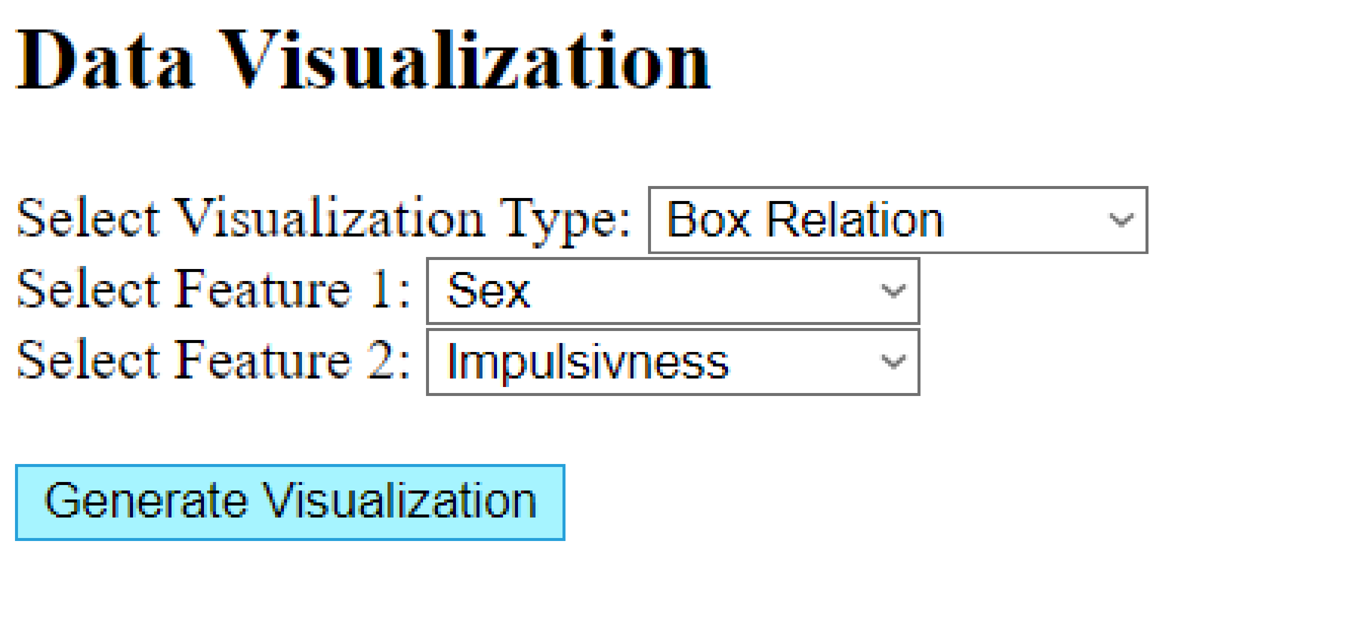
VISUALIZATIONS

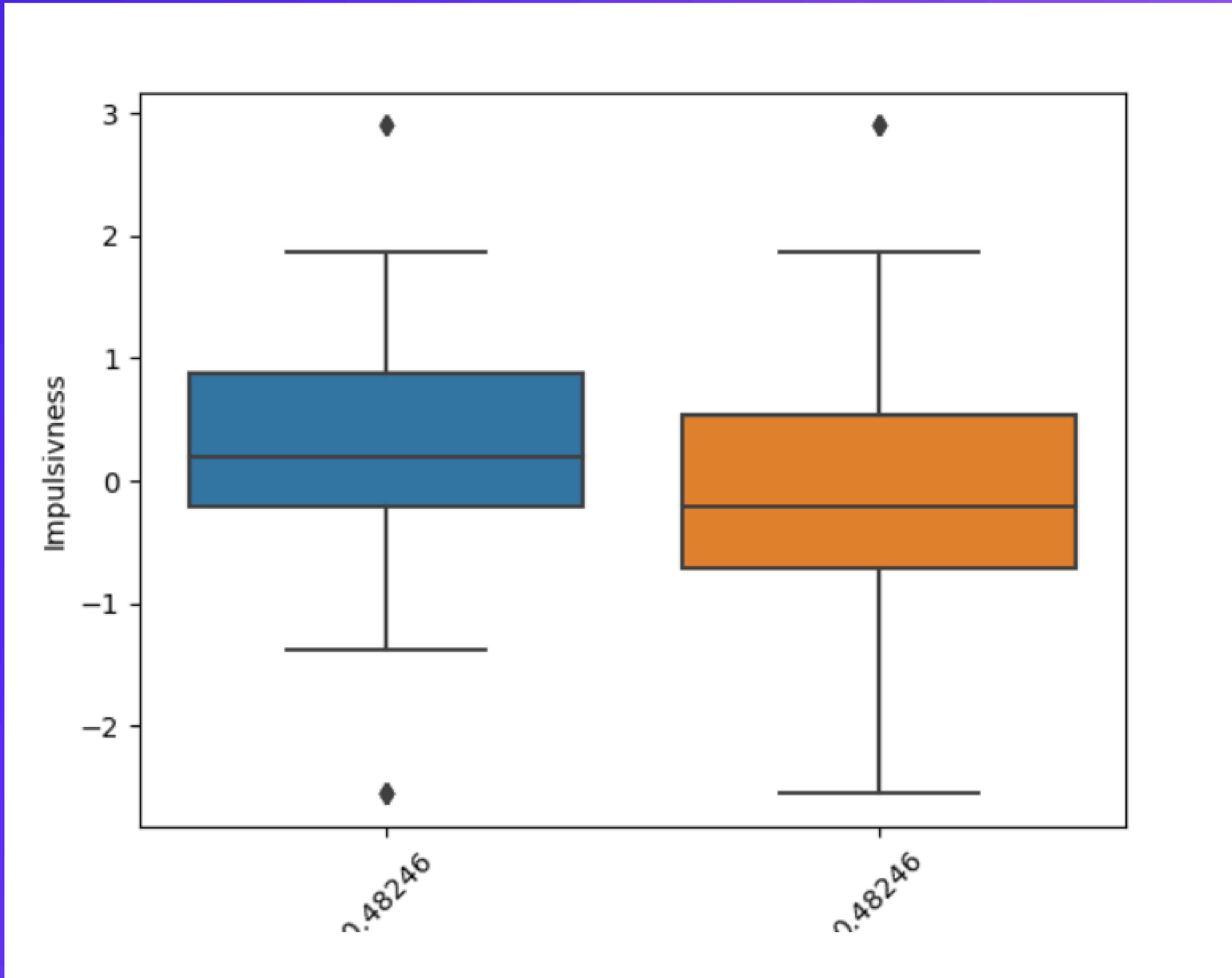




FLASK :

VISUALIZATIONS





CONCLUSION

THANK YOU!