DIA-1

Drug Consumption Analysis & Predictions

Realized by:

Hamza HALINE

Joshua BORNET

YAHYA BOUAYAD

Table of contents

01

Introduction

General Introduction:

-In today's dynamic world, understanding the factors influencing drug consumption patterns is of paramount importance for public health, education, and policy-making.

-The selection of this dataset stems from a deliberate choice to explore the intersection of personality traits, education, demographics, and drug usage patterns.

-Our dataset name is : drug+consumption+quantified.csv
-Link : https://archive.ics.uci.edu/dataset/373/drug+consumption+quantified

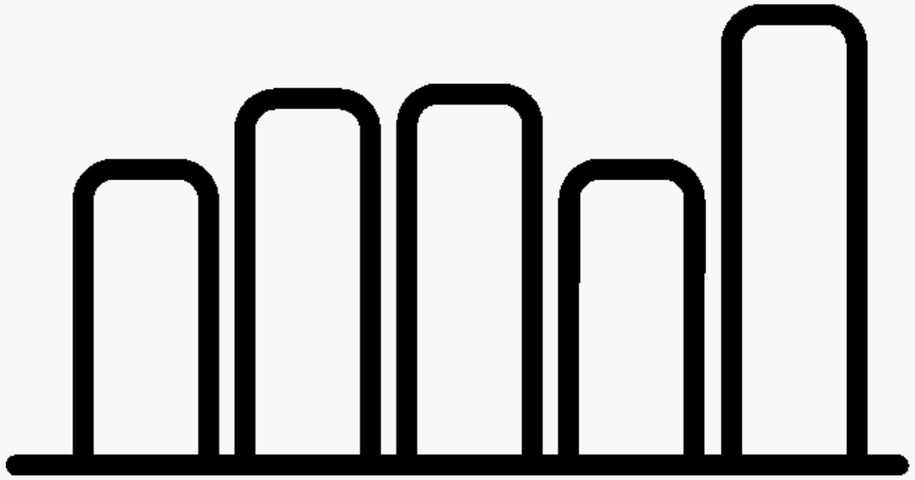
Dataset Presentation:

-Our chosen dataset is a comprehensive repository of information, comprising records for 1885 respondents, each characterized by 30 attributes.

-it contains:

- an ID column
- 5 demographic columns (features)
- 7 personality traits (features)
- 18 drugs with their usage frequency (target)
- a fake drug called Semeron to verify reliability of answers

-



-Each drug variable can take 6 different values:

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

-

Dataset Presentation:

Feature	Description
ID	Identification
Age	Age range of participant
Genre	Male or Female
Education	Level of education
Pays	Country of orgin
Ethnie	Ethnicity/Race of participant
Neuroticisme	Emotional instability, Anxiety, Sensitivity
Extraversion	Sociable, Outgoing, Energetic
Ouverture à l'expérience	Curiosity, Creativity, Open-mindedness
Amicalité	Agreeableness : Compassionate, Cooperative, Tolerant
Conscience	Conscientiousness: Organized, Responsible, Detail-oriented
Impulsivité	Impulsivity: Spontaneous, Rash, Impatient
Recherche de sensations	Sensation Seeking: Thrill-seeking, Adventurous, Risk-taking

02

Data Pre-Processing and Data visualisation

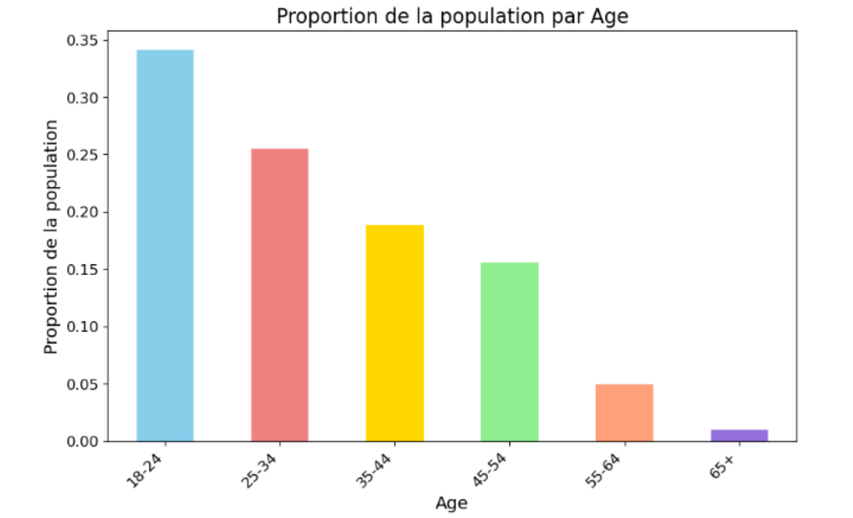
Data loading, exploring and pre-processing:

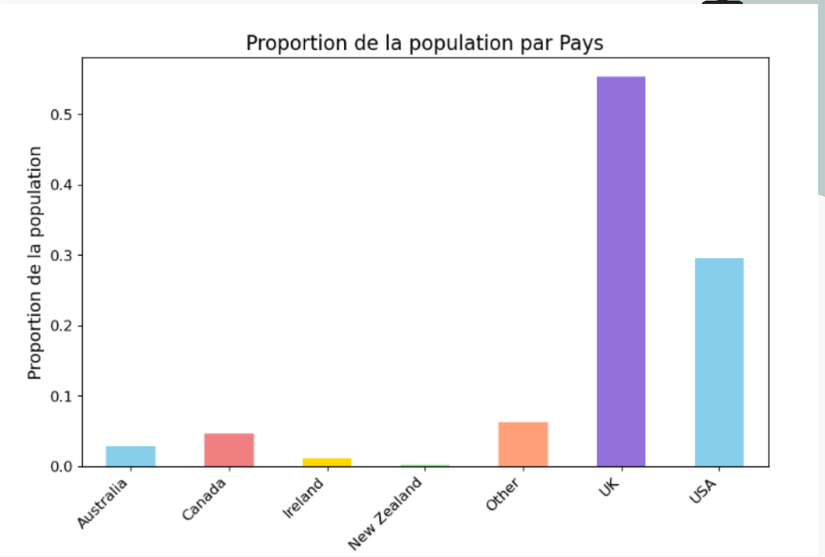
-Remapper les valeurs par leur vraie signification
-Supprimer toutes les personnes qui prétendent avoir consommé de la semeron
-Créer des listes pour faciliter la gestion des données que nous voulons étudier
-Corrélation entre toutes les données

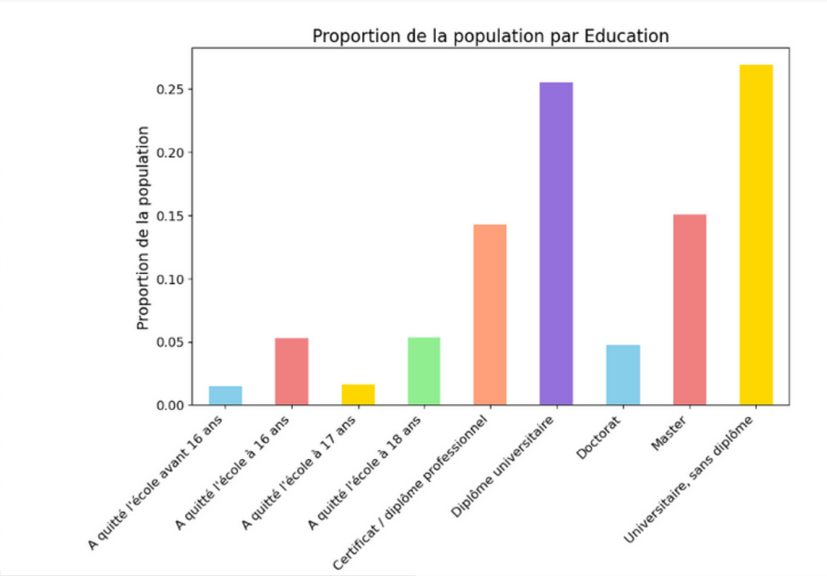
```python
import pandas as pd

fichier_data = 'drug_consumption.data'

# Lire le fichier .data avec pandas
# Vous devez spécifier le délimiteur approprié, par exemple '\s+' pour les espaces
data = pd.read_csv(fichier_data, delimiter=',', header=0)

# Spécifier les noms des colonnes
noms_colonnes = ['ID', 'Age', 'Genre', 'Education', 'Pays', 'Ethnie', 'Neuroticisme', 'Extraversion', 'Ouverture à l\'expéri

# Ajouter les noms de colonnes au DataFrame
data.columns = noms_colonnes

data = data.set_index('ID')
# Afficher le DataFrame avec les noms de colonnes
print(data)
```

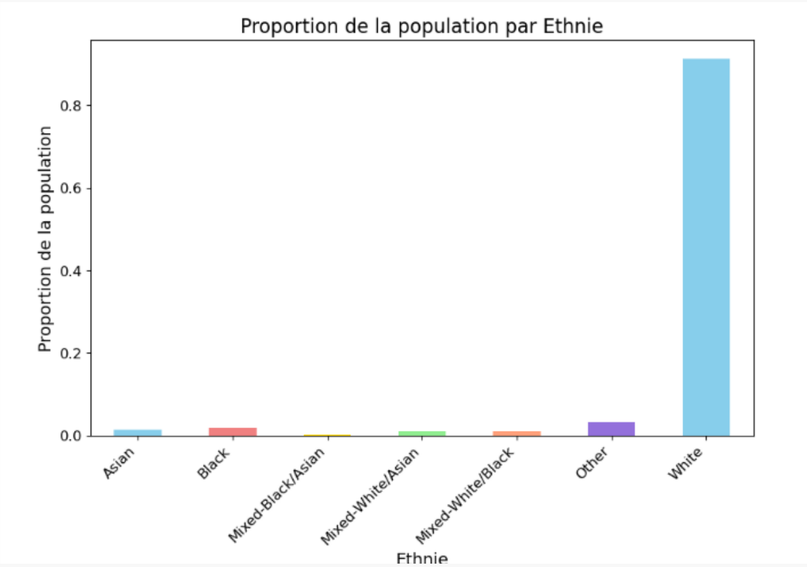
```python
pers_data = data.copy()
age = ['18-24' if i <= -0.9 else
       '25-34' if i >= -0.5 and i < 0 else
       '35-44' if i > 0 and i < 1 else
       '45-54' if i > 1 and i < 1.5 else
       '55-64' if i > 1.5 and i < 2 else
       '65+'
       for i in pers_data['Age']]

genre = ['Femme' if i > 0 else "Homme" for i in pers_data['Genre']]

education = ['A quitté l\'école avant 16 ans' if i <-2 else
             'A quitté l\'école à 16 ans' if i > -2 and i < -1.5 else
             'A quitté l\'école à 17 ans' if i > -1.5 and i < -1.4 else
             'A quitté l\'école à 18 ans' if i > -1.4 and i < -1 else
             'Universitaire, sans diplôme' if i > -1 and i < -0.5 else
             'Certificat / diplôme professionnel' if i > -0.5 and i < 0 else
             'Diplôme universitaire' if i > 0 and i < 0.5 else
             'Master' if i > 0.5 and i < 1.5 else
             'Doctorat'
             for i in pers_data['Education']]

pays = ['USA' if i < -0.5 else
          'New Zealand' if i > -0.5 and i < -0.4 else
          'Other' if i > -0.4 and i < -0.2 else
          'Australia' if i > -0.2 and i < 0 else
          'Ireland' if i > 0 and i < 0.23 else
          'Canada' if i > 0.23 and i < 0.9 else
          'UK'
          for i in pers_data['Pays']]

ethnie = ['Black' if i < -1 else
            'Asian' if i > -1 and i < -0.4 else
            'White' if i > -0.4 and i < -0.25 else
            'Mixed-White/Black' if i >= -0.25 and i < 0.11 else
            'Mixed-White/Asian' if i > 0.12 and i < 1 else
            'Mixed-Black/Asian' if i > 1.9 else
            'Other'
            for i in pers_data['Ethnie']]


pers_data['Age'] = age
pers_data['Genre'] = genre
pers_data['Education'] = education
pers_data['Pays'] = pays
pers_data['Ethnie'] = ethnie
pers_data[info_col].head()
```

Bar Plots: Proportion of Population by a given Demographic Feature

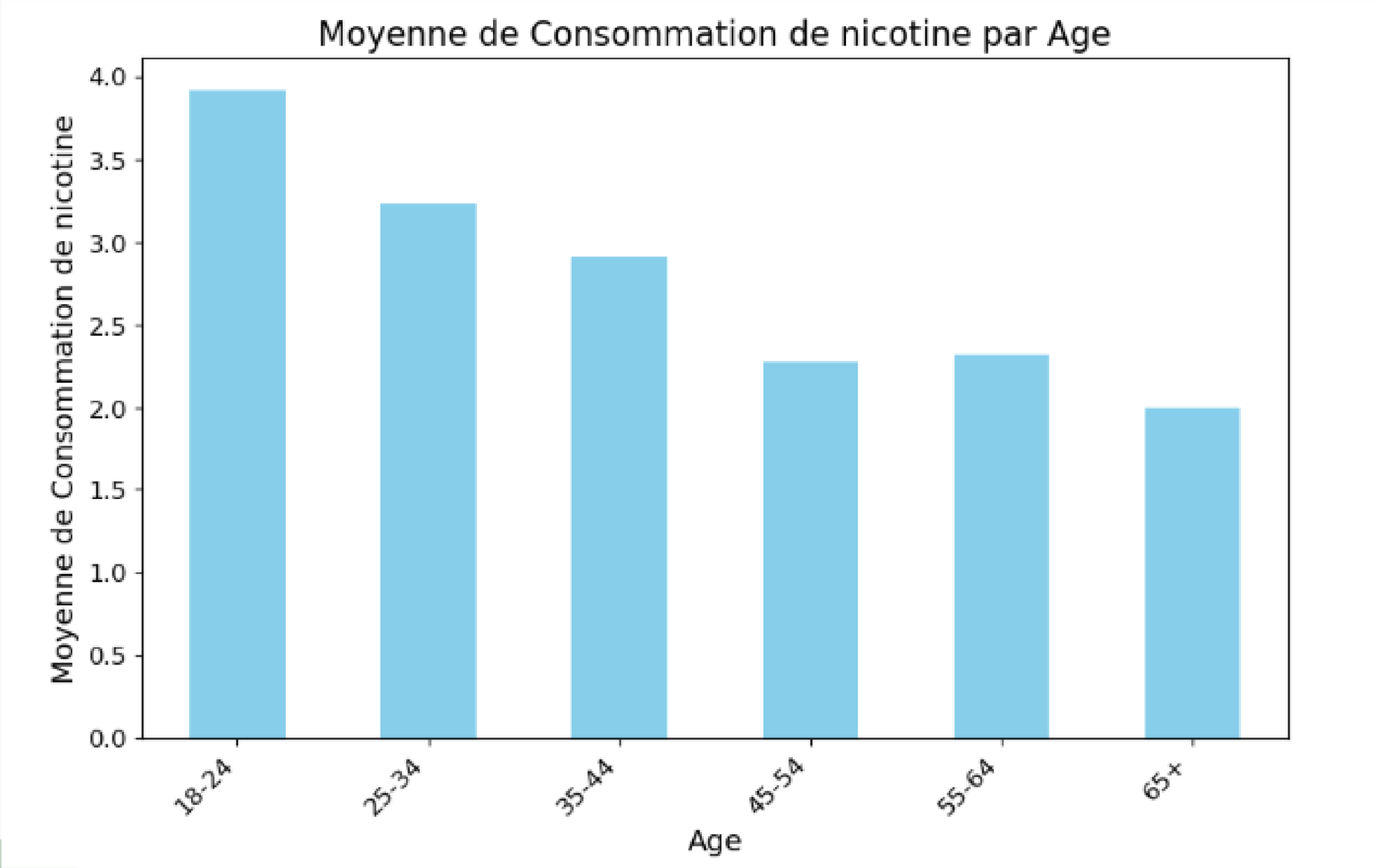




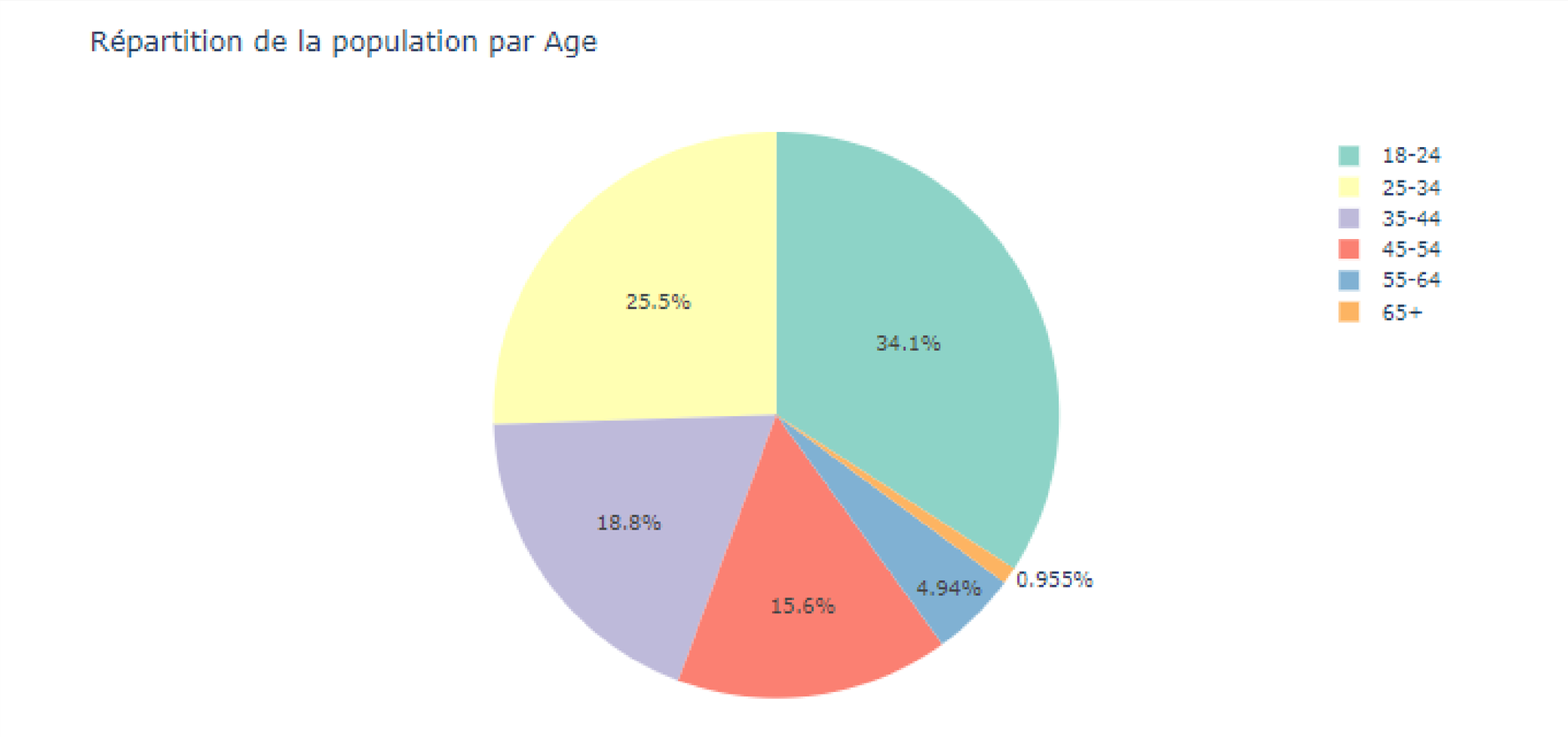




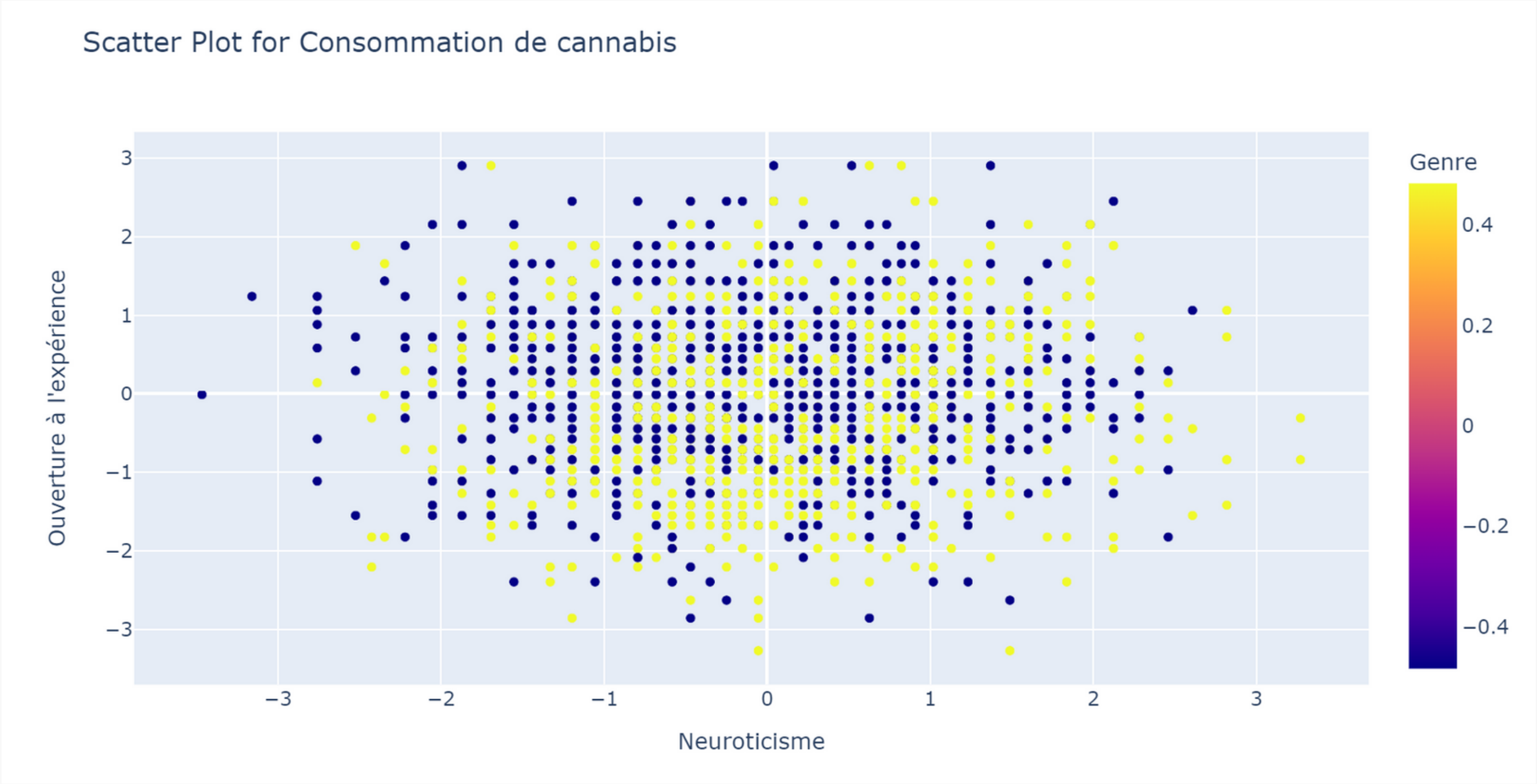
Bar Plot of Average Drug Consumption by Age

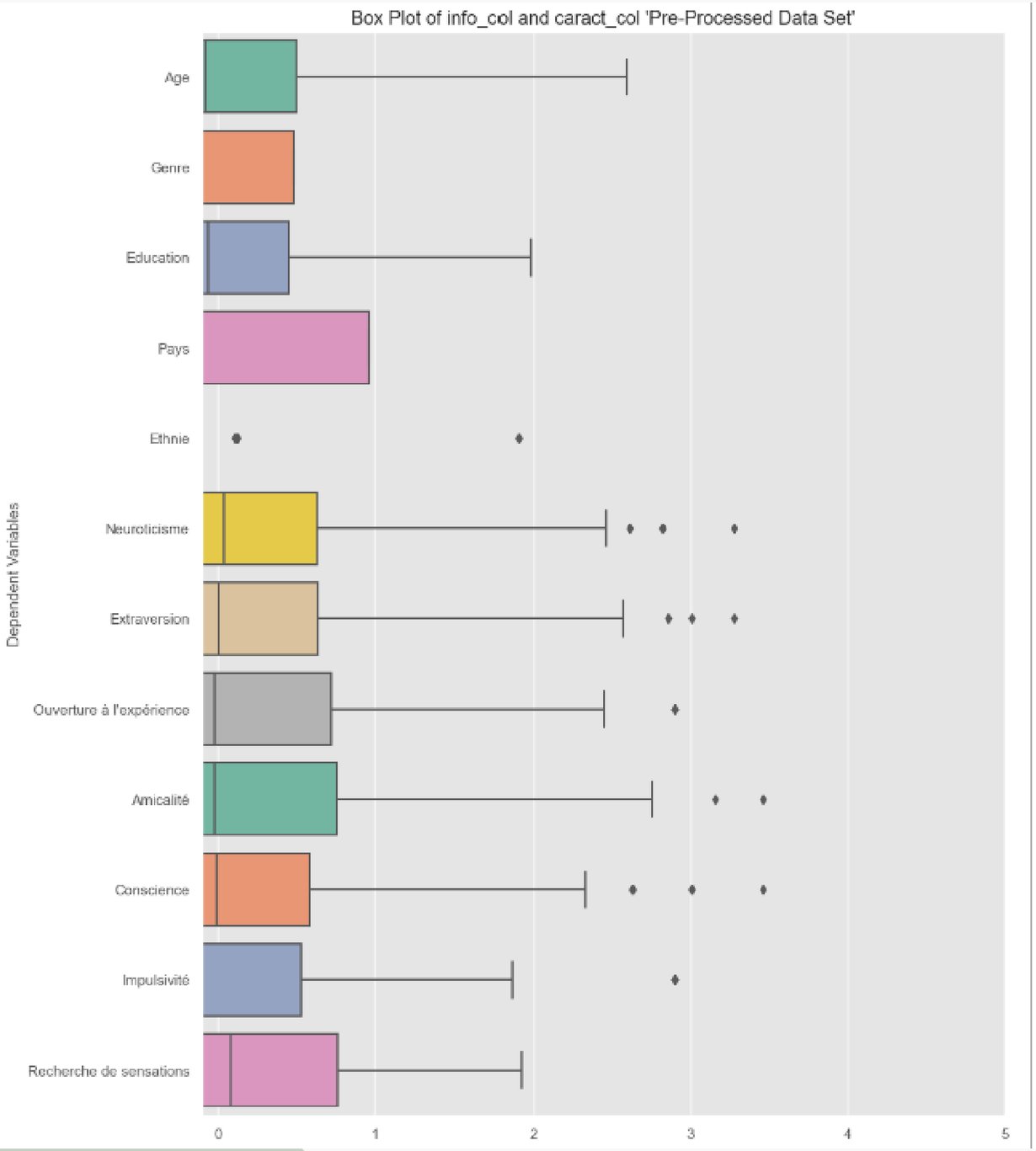


Pie chart of Population Distribution by Demographic Features



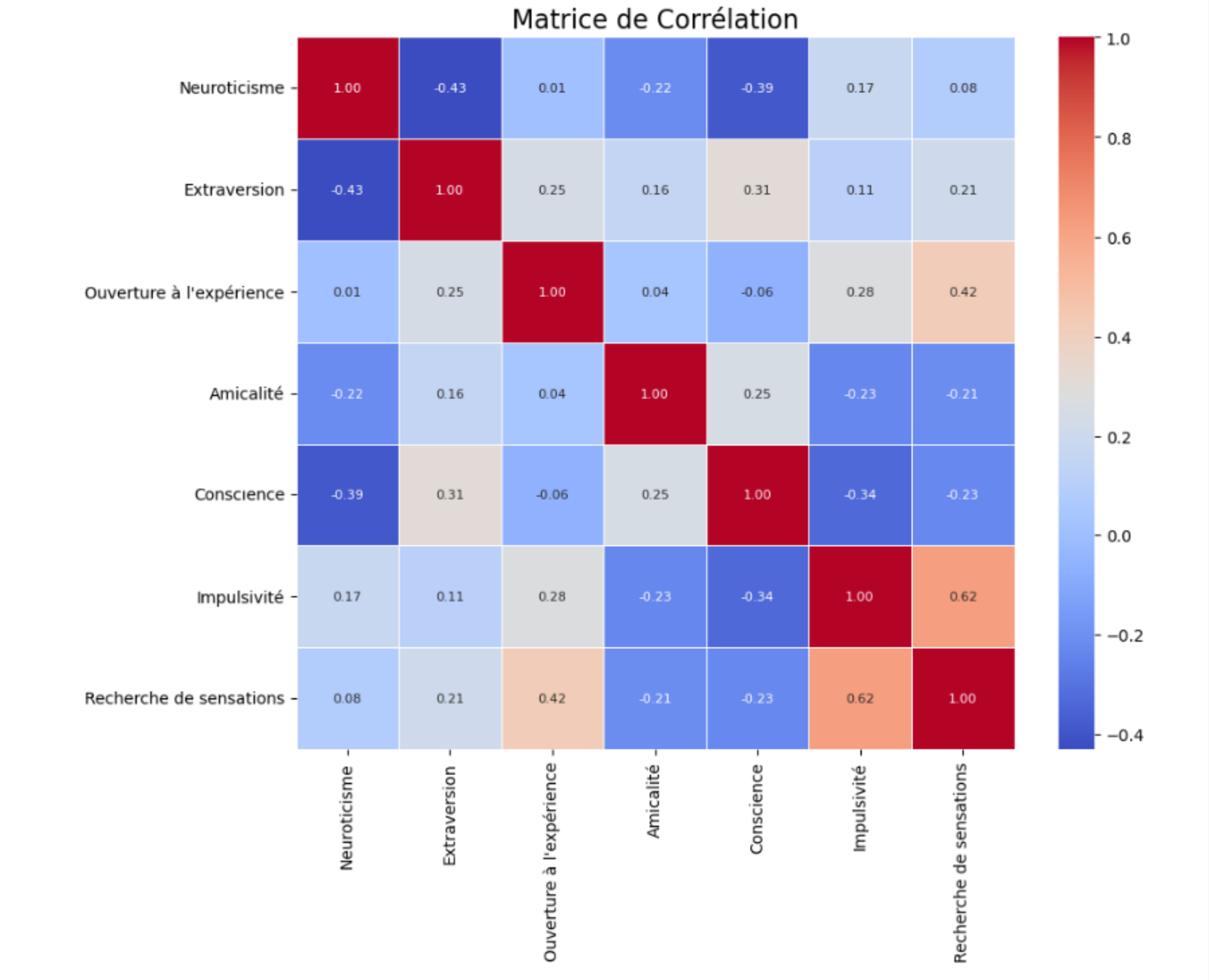
Scatter Plot for Specific Drug type

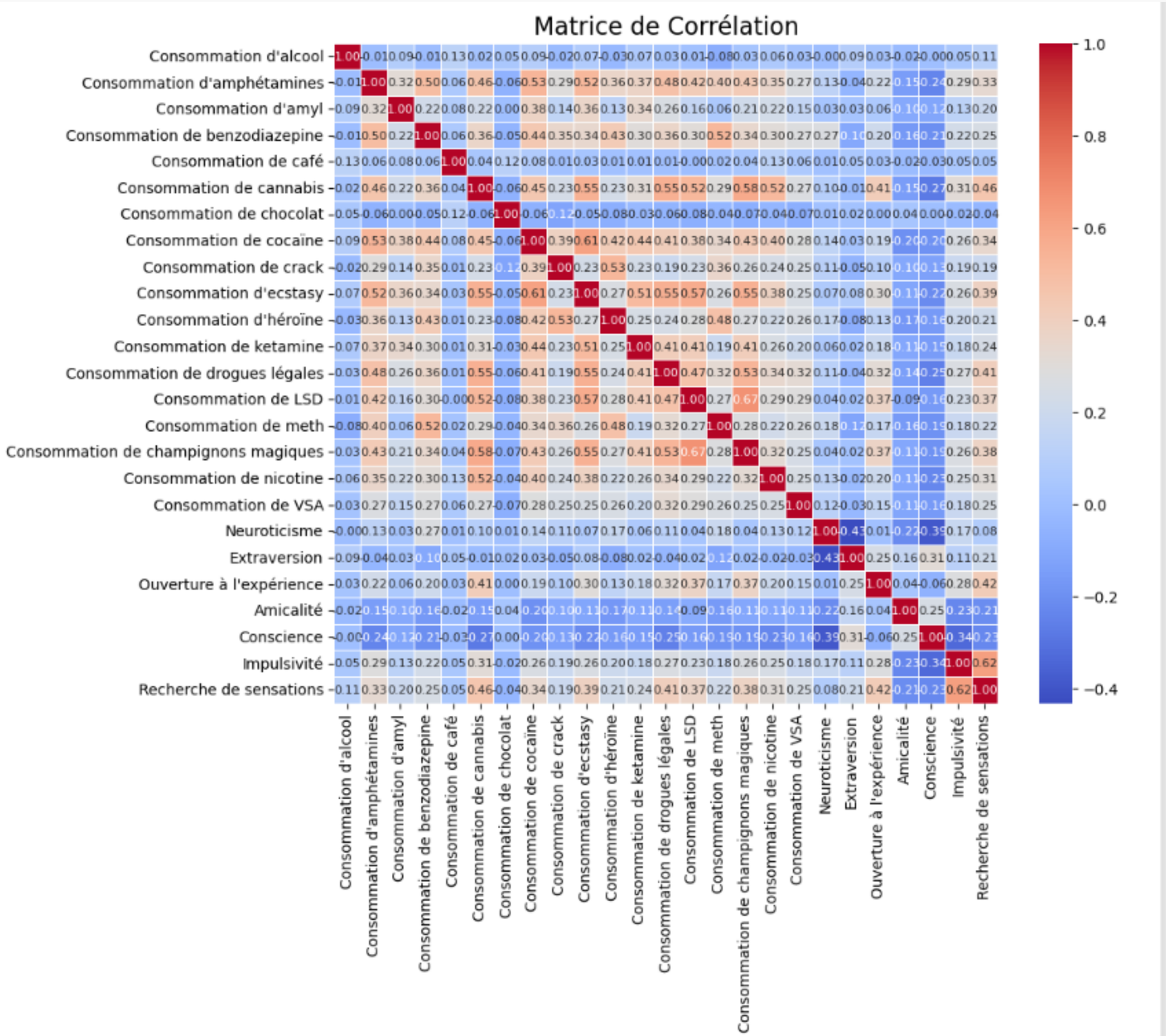




Box Plot of Pre-Processed Data features

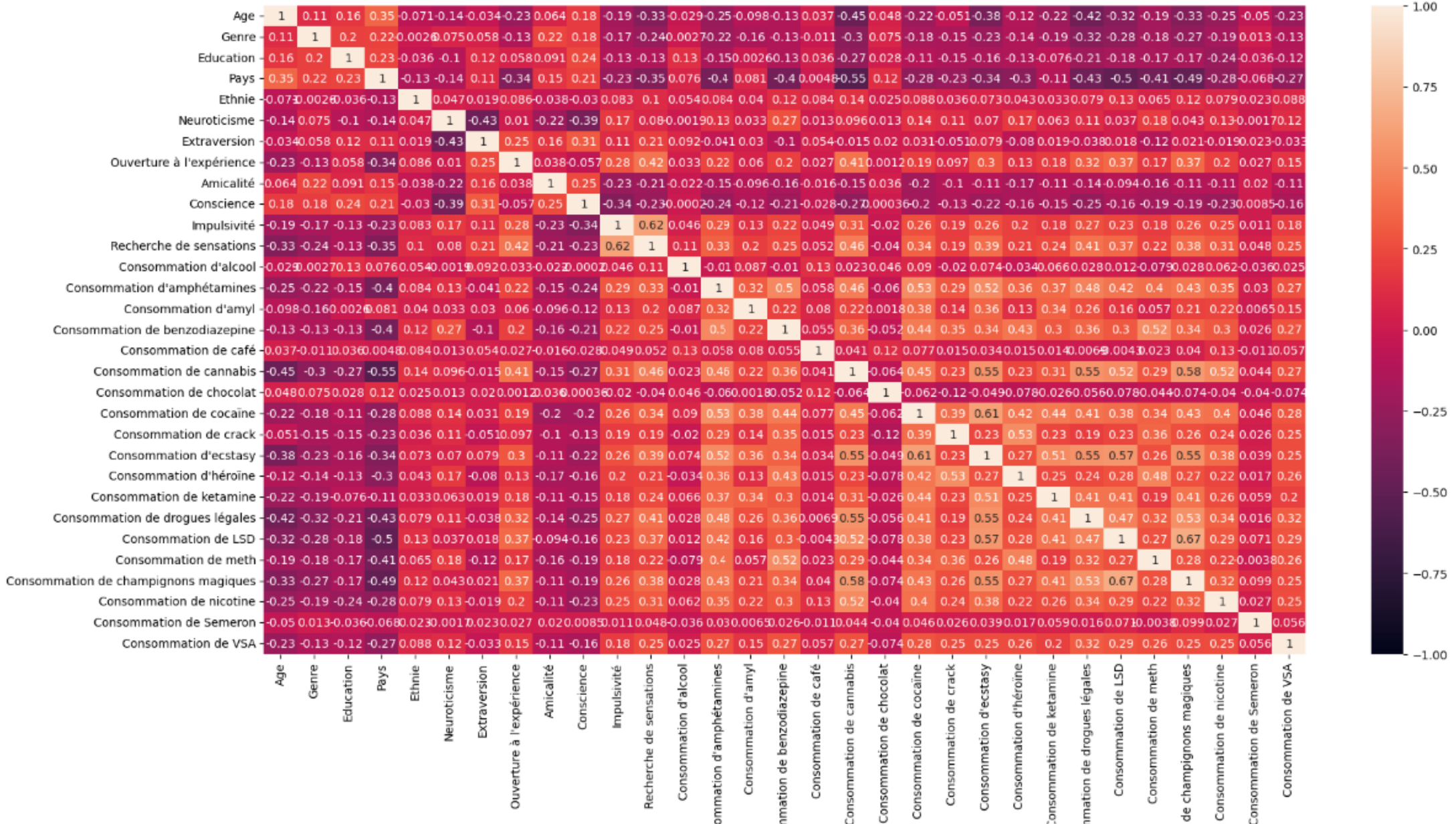
Correlation Matrice of personal traits



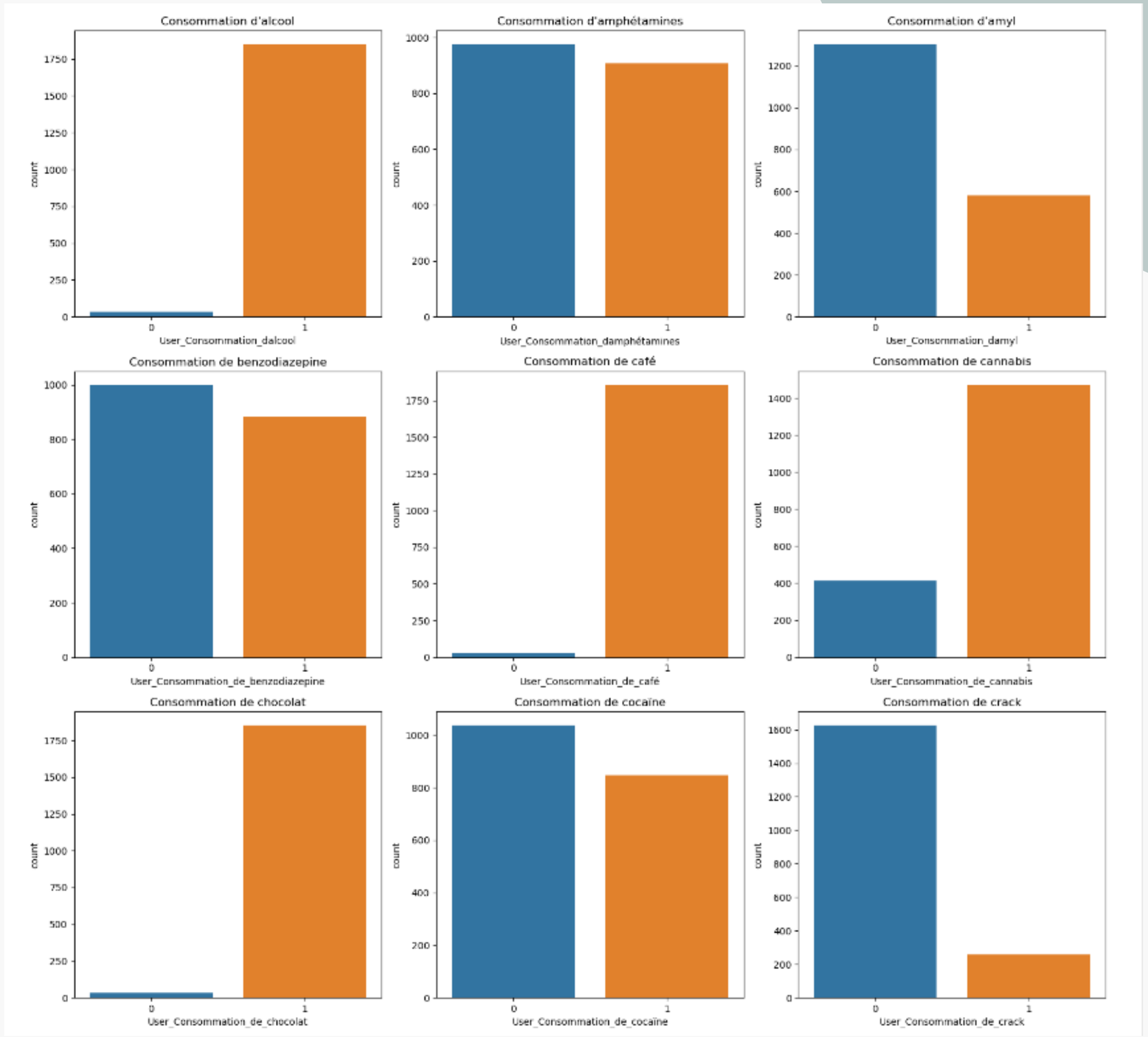


Correlation Matrice of drugs types and personal traits

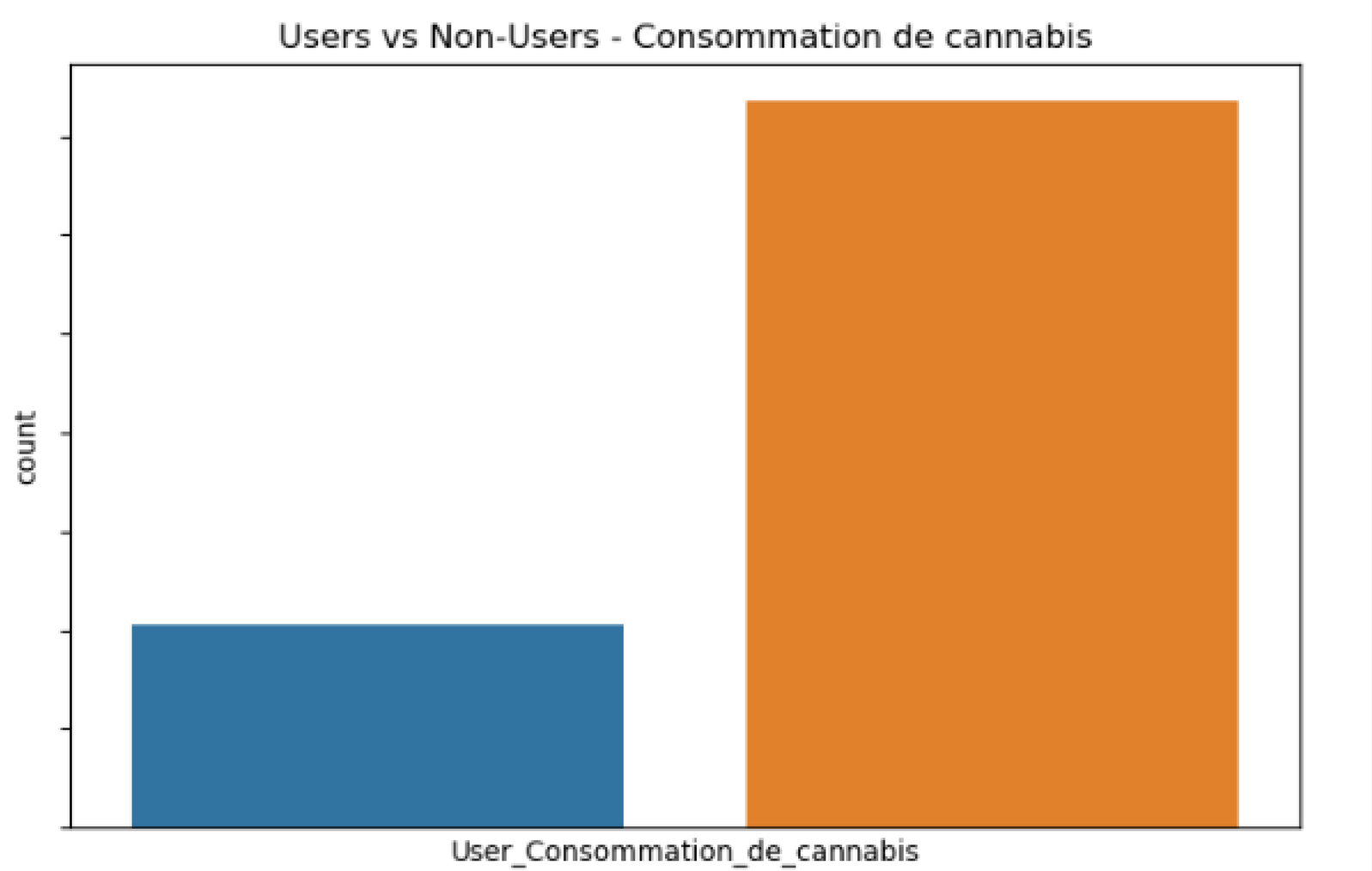
Heatmap of all features



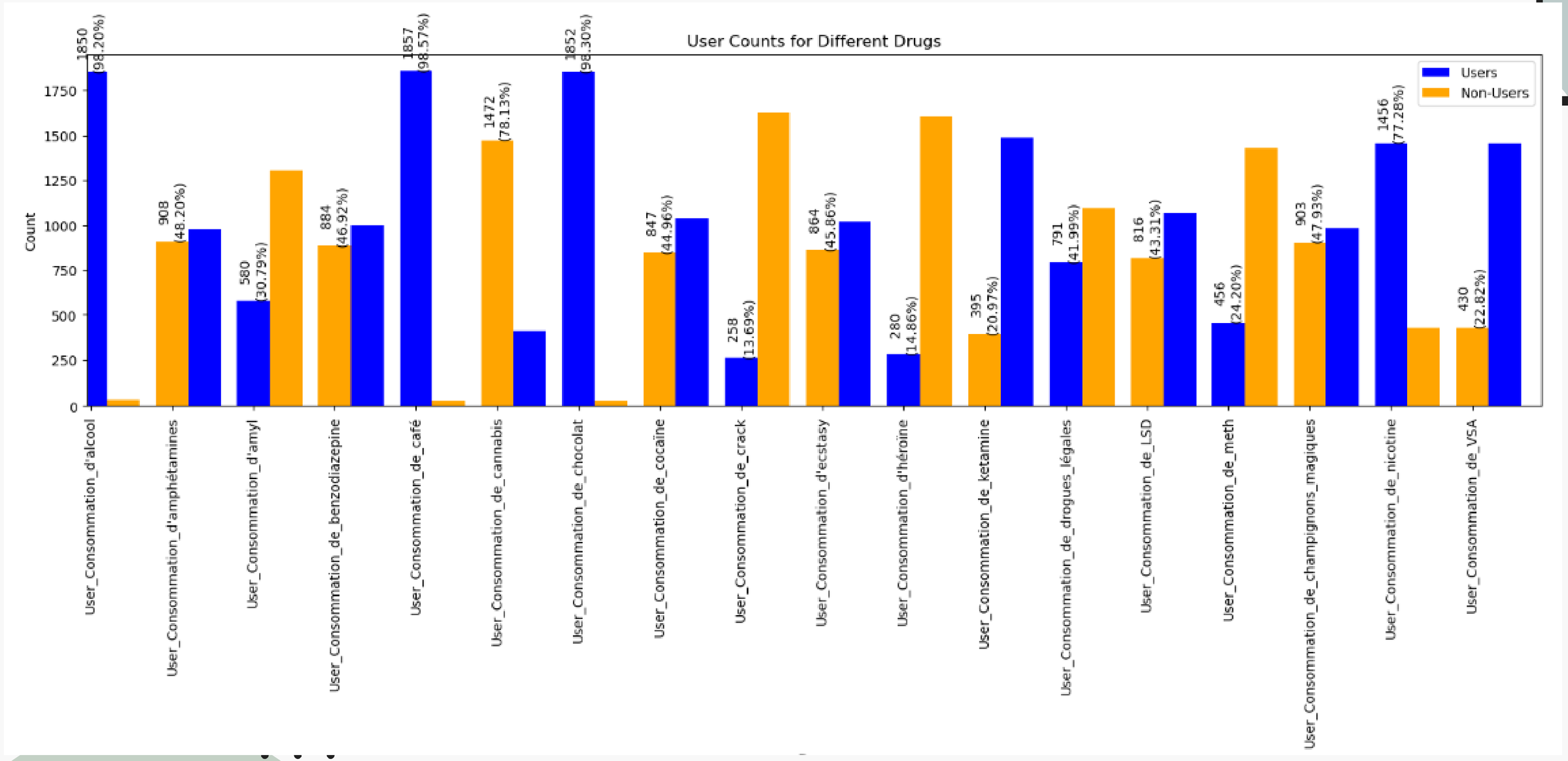
Function that returns a plot of Users vs Non-Users of all drugs each on it’s own plot



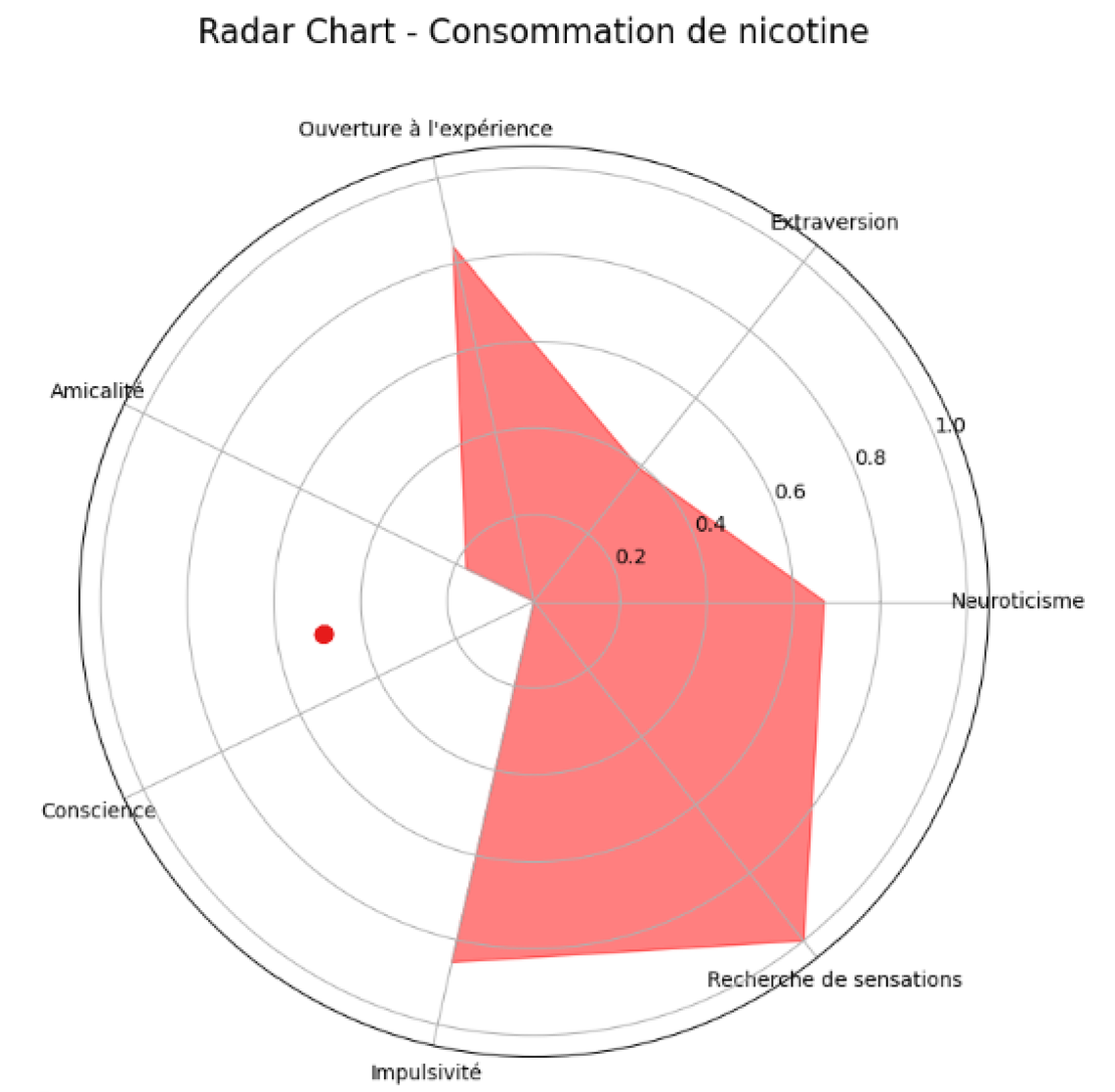
Function that returns a plot of Users vs Non-Users of a chosen drug

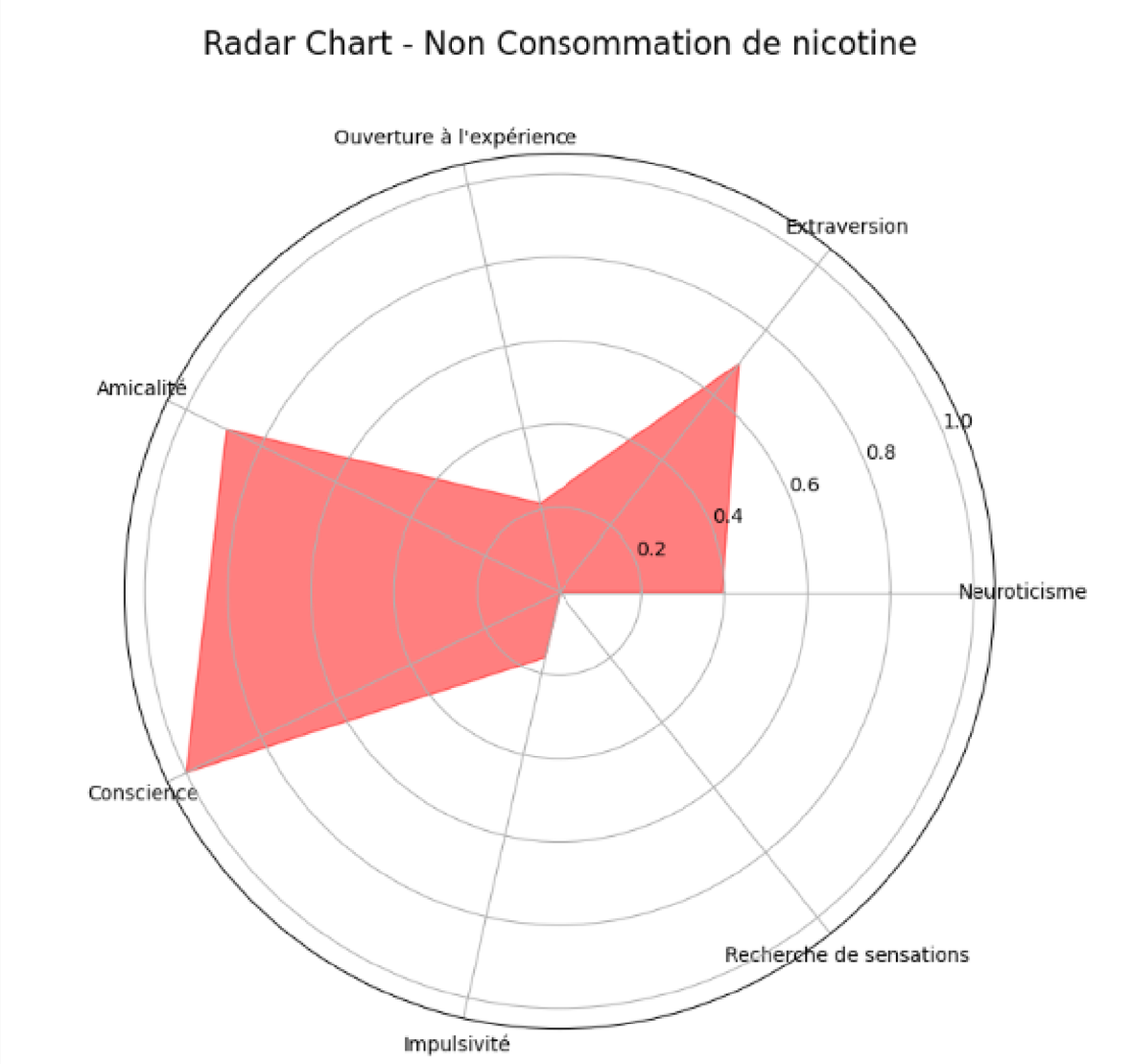


User vs Non-Users Counts for Different Drugs

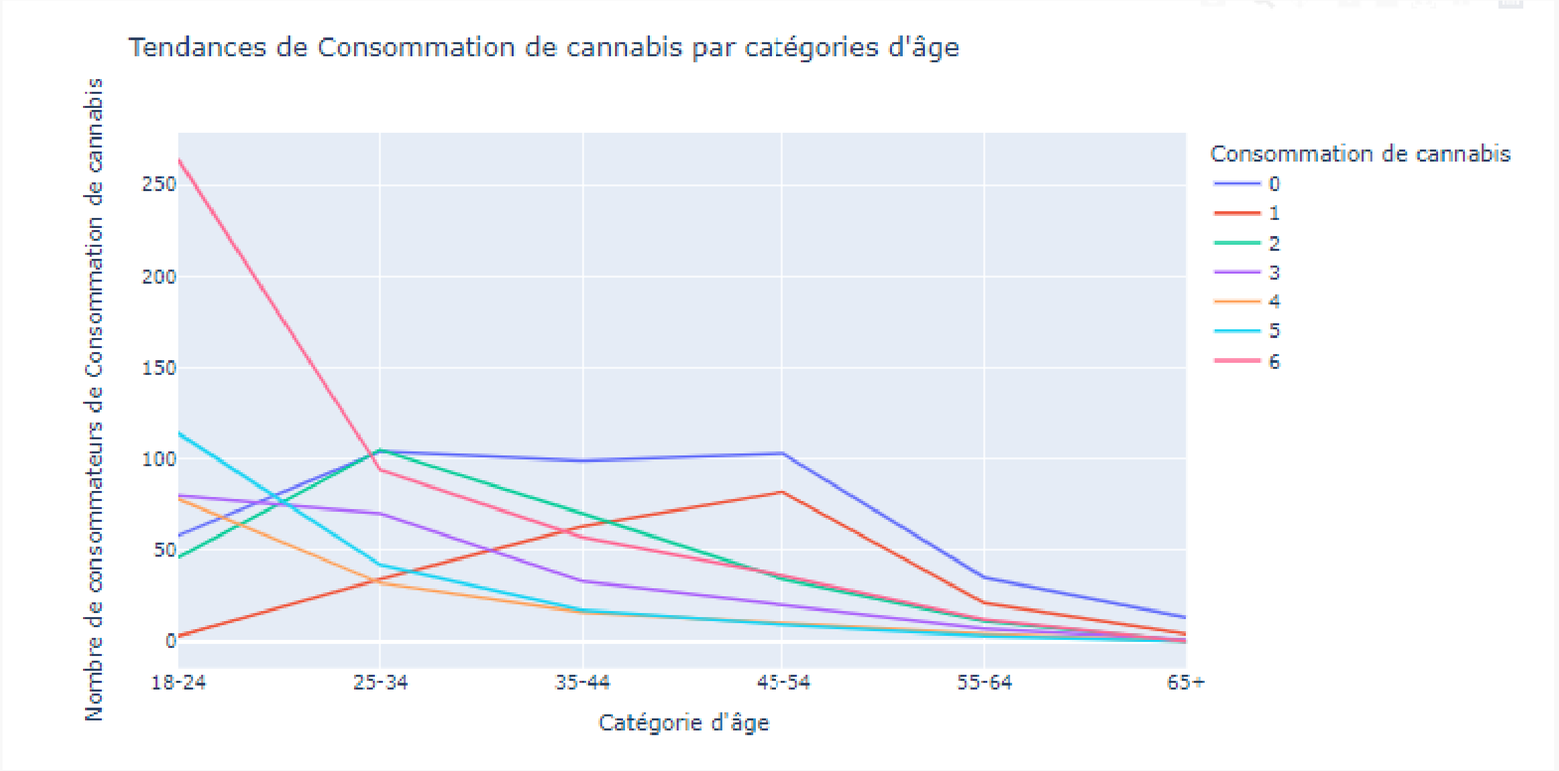


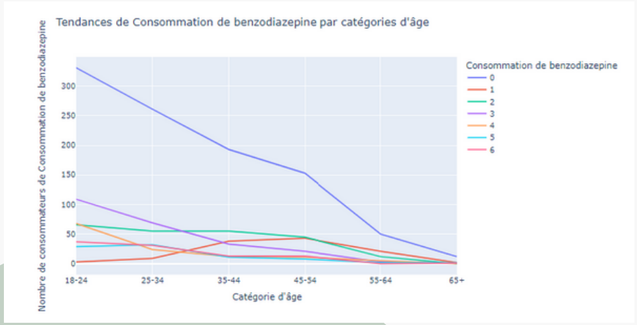
Radar Chart: Substance Consumption vs. Non-Consumption Based on Personality Traits

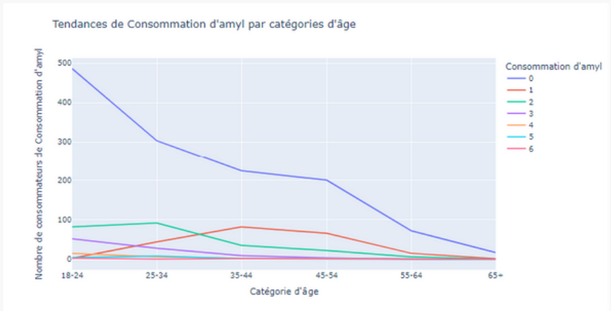


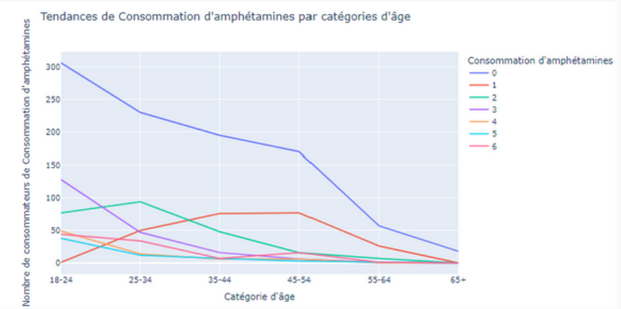


Consumption Trends of each drug Across Age Categories

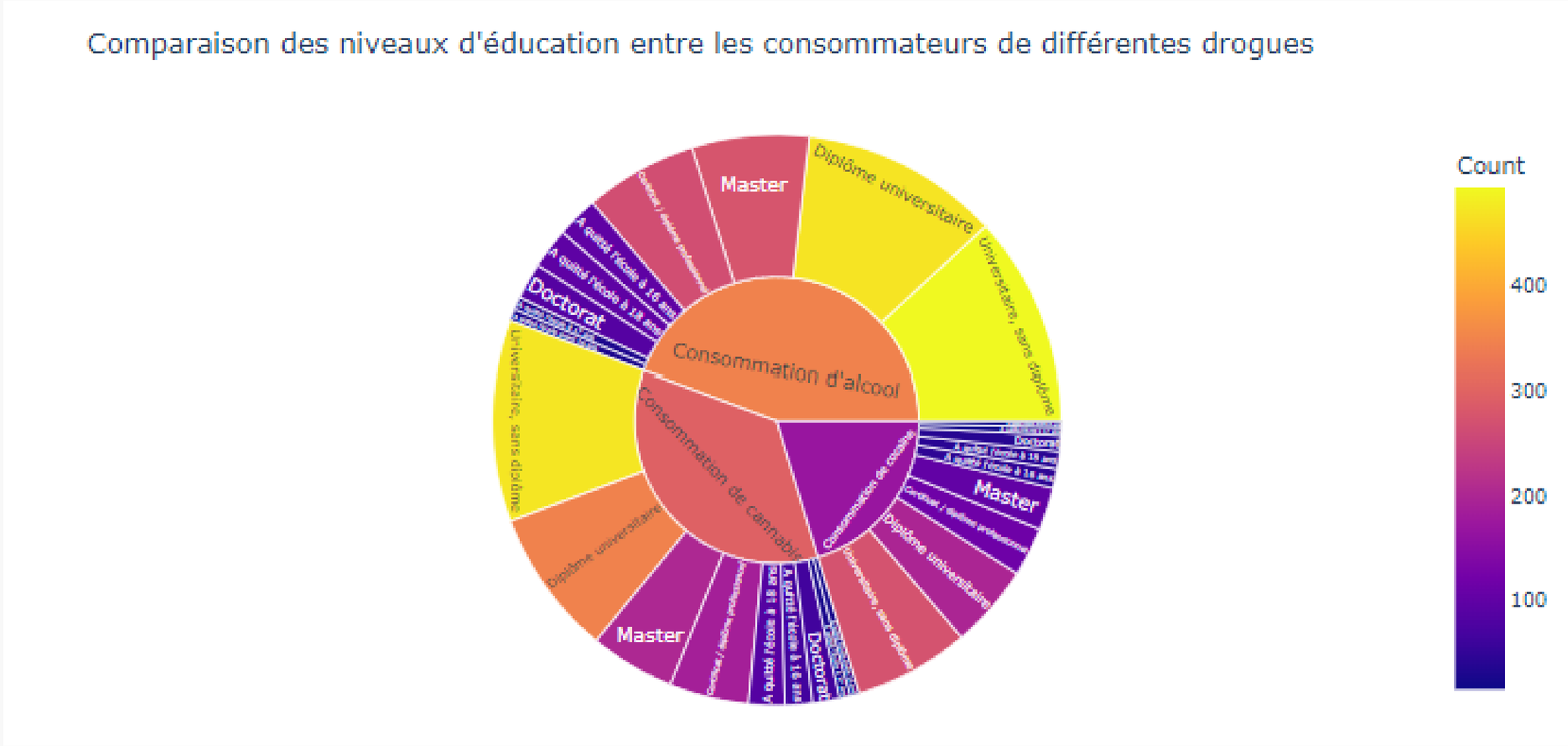




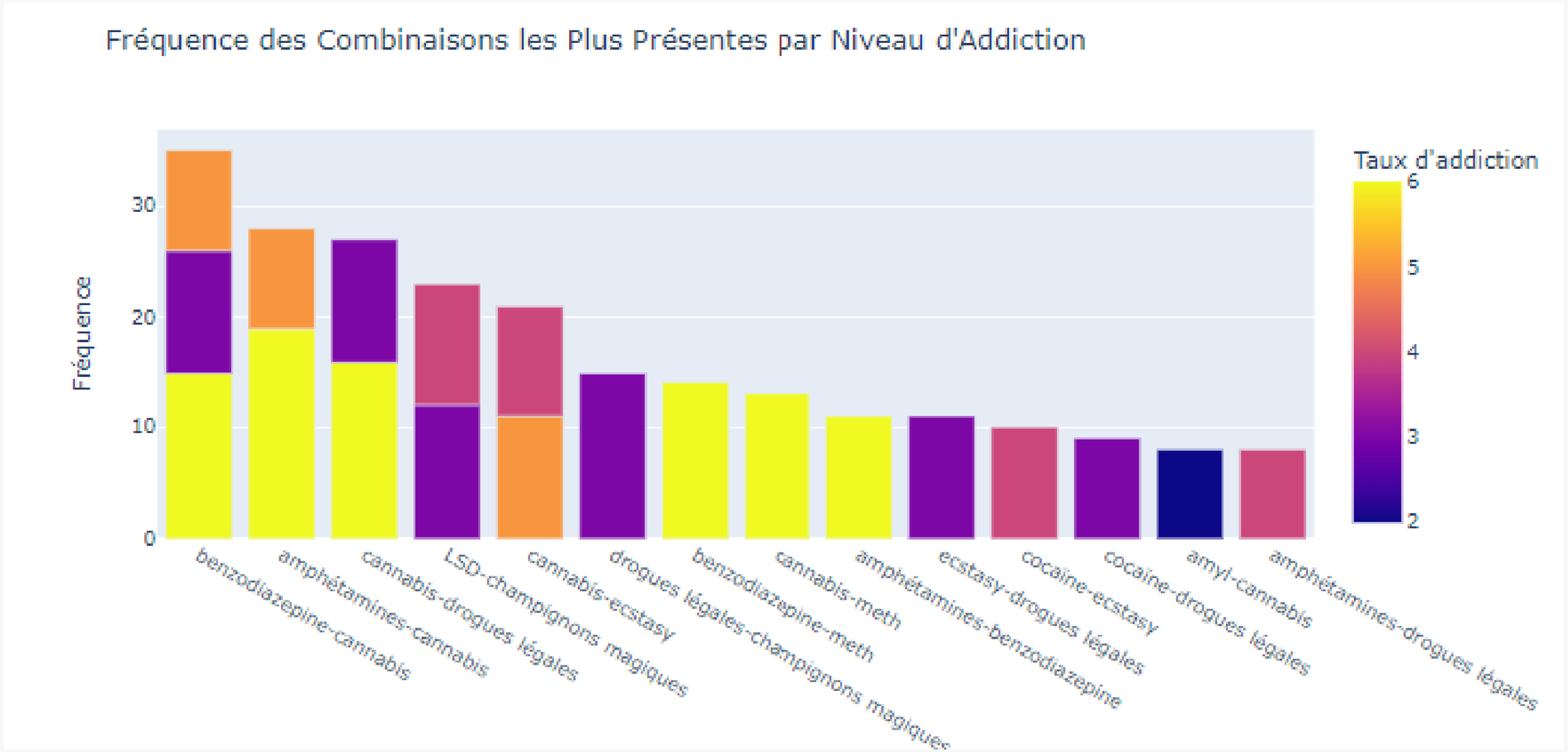




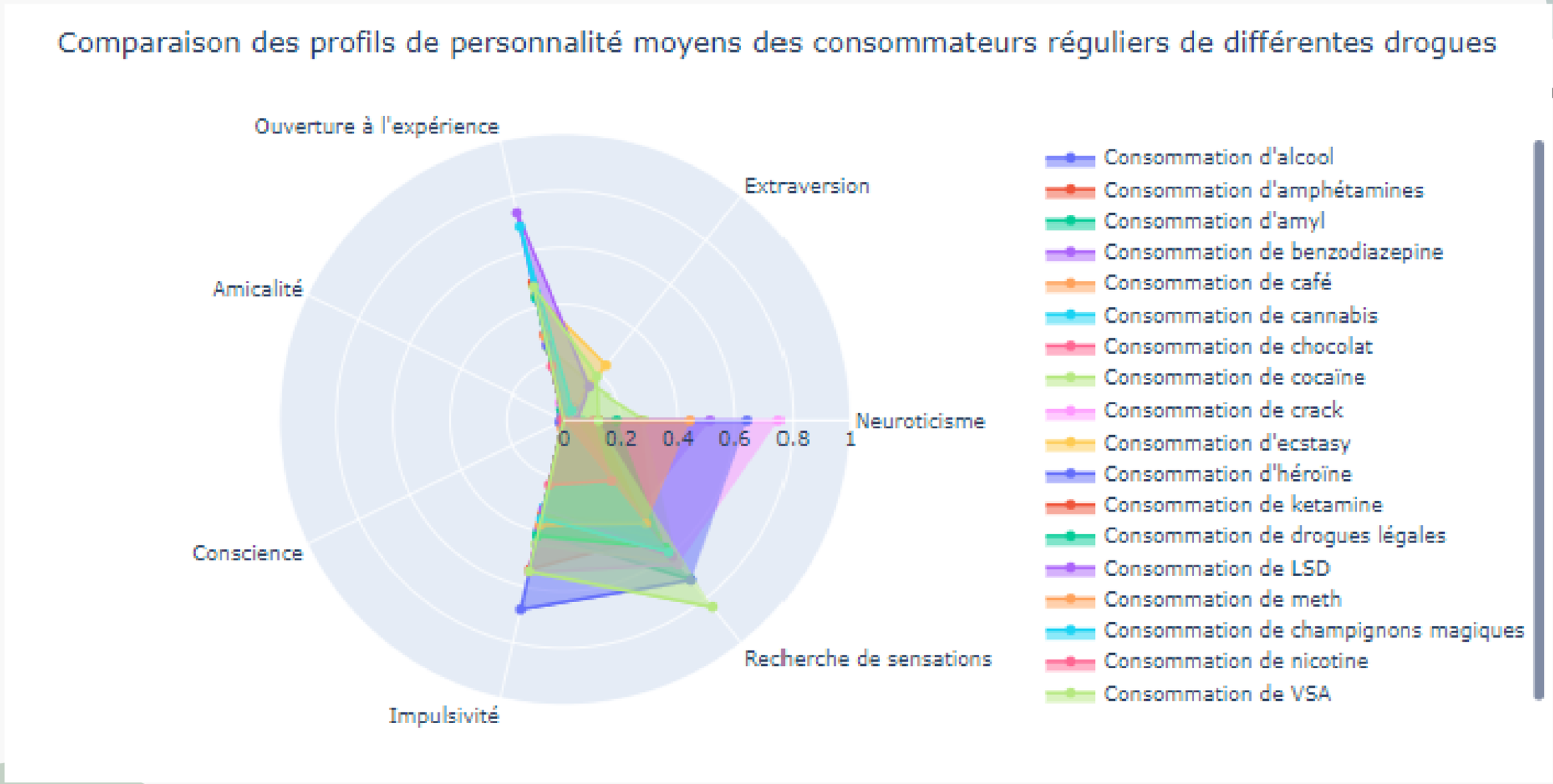
Comparison of Education Levels Among Consumers of Different Drugs



Frequency of Most Common Combinations by Level of Addiction



Comparison of Average Personality Profiles Among Regular Consumers of Different Drugs



03

Modeling

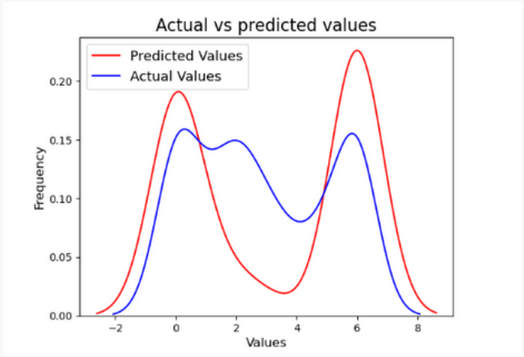
Strategy 1:

We trained our model only on demographic features and personnal traits features and we tried to predict the target feature

01

SVM:

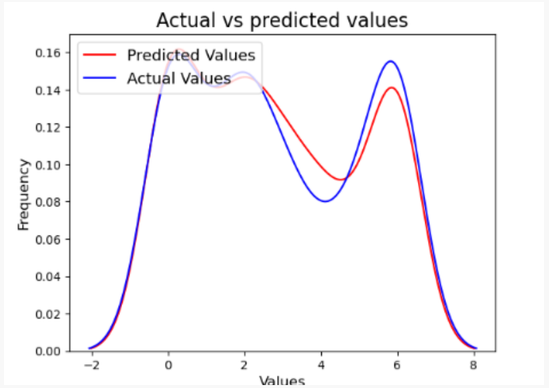
Accuracy: 40.58%



02

Decision Tree:

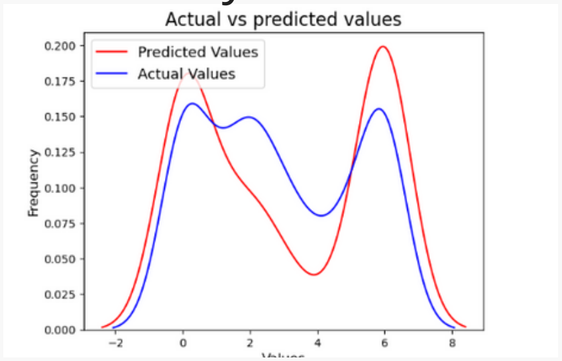
Accuracy: 29.18%



03

Random Forest:

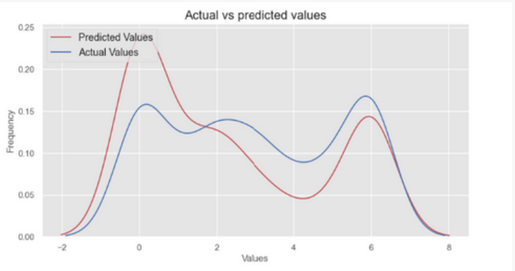
Accuracy: 39.26%



04

K-NN:

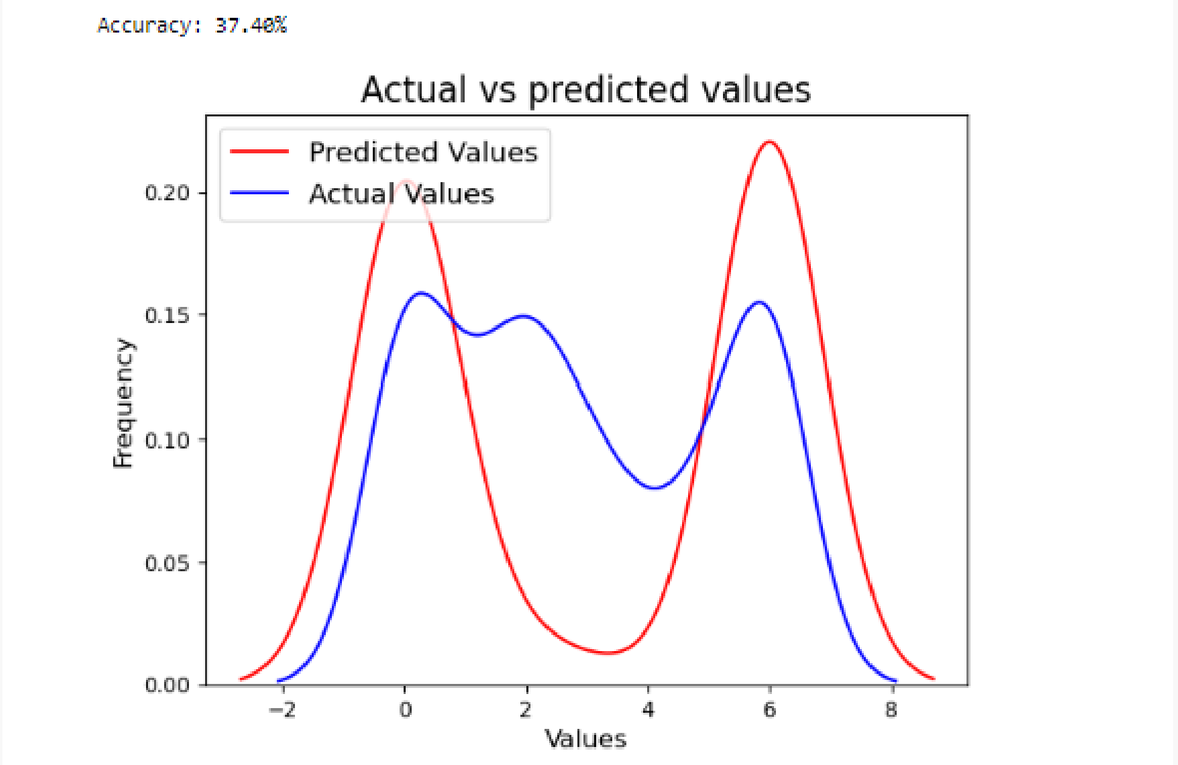
Accuracy: 33,56%



Strategy 2:

We trained our model with GridSearch by changing t he hyperparametres:

Meilleurs paramètres : {'C': 10, 'gamma': 0.01, 'kernel': 'rbf'}



Strategy 3:

We trained our model on all features except one target features that we want

```
features_cannabis = data_cannabis.drop(["Target","Consommation de Semeron"], axis=1)
target_cannabis = data_cannabis['Target']

for name, model in models.items():
    print(name + ' statistique :')
    prediction_training(features_cannabis,target_cannabis,model)
```

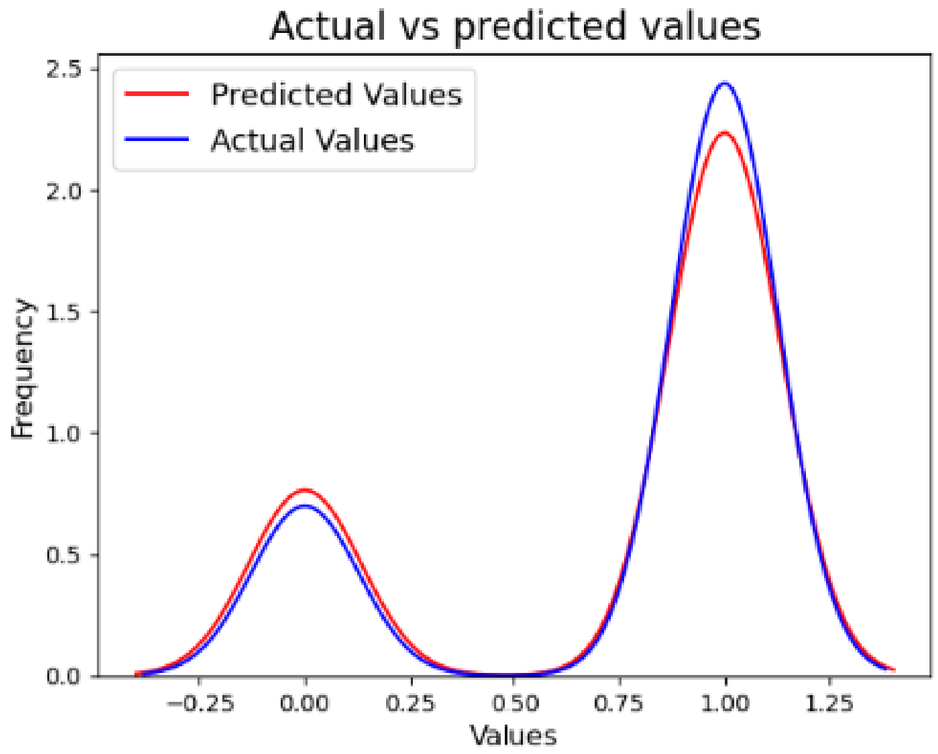
Algorithmes used:

SVM

Support Vector Machines statistique :

Accuracy: 89.92%

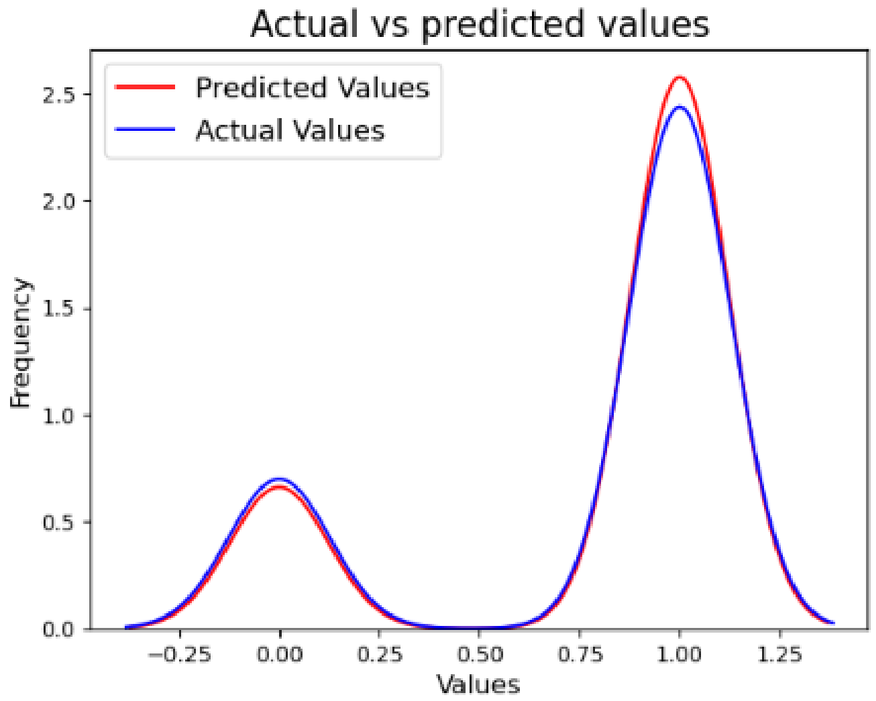
	precision	recall	f1-score	support
0	0.74	0.85	0.79	84
1	0.95	0.91	0.93	293
accuracy			0.90	377
macro avg	0.85	0.88	0.86	377
weighted avg	0.91	0.90	0.90	377



Algorithmes used:

Decision Tree Classifier statistique :
Accuracy: 81.17%

	precision	recall	f1-score	support
0	0.58	0.54	0.56	84
1	0.87	0.89	0.88	293
accuracy			0.81	377
macro avg	0.73	0.71	0.72	377
weighted avg	0.81	0.81	0.81	377



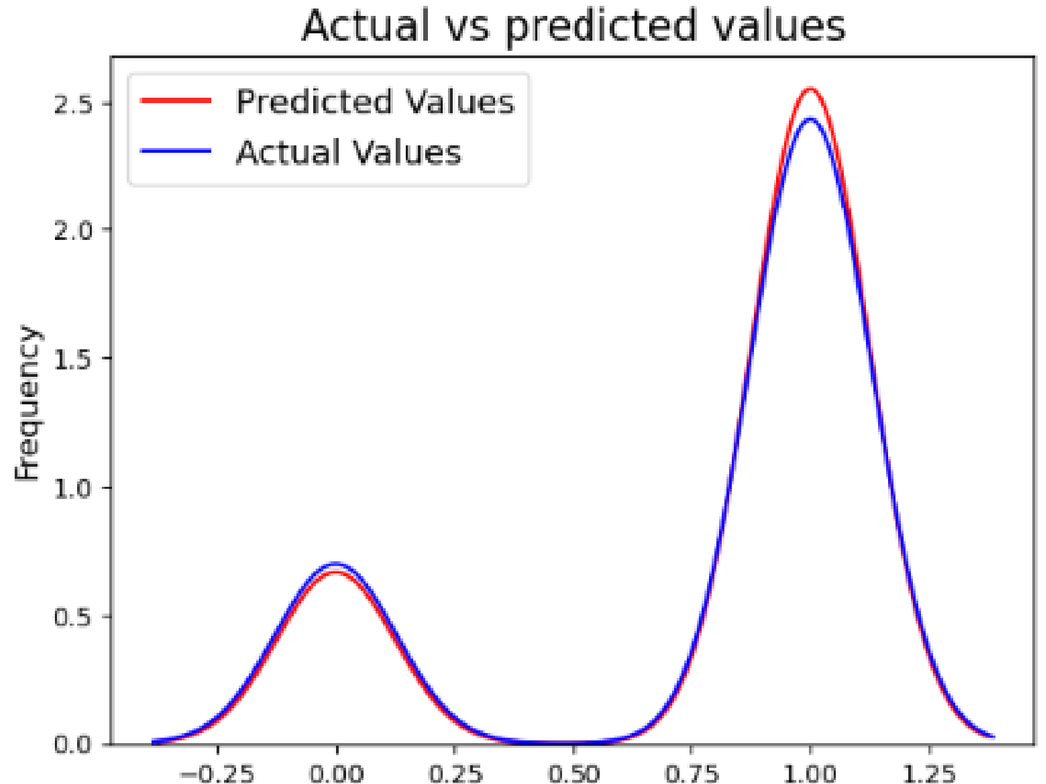
Decision Tree Classifier

Algorithmes used:

Random forest classifier

Random Forest Classifier statistique :
Accuracy: 87.80%

	precision	recall	f1-score	support
0	0.74	0.69	0.72	84
1	0.91	0.93	0.92	293
accuracy			0.88	377
macro avg	0.83	0.81	0.82	377
weighted avg	0.88	0.88	0.88	377



04

Bonus touch

New Data added:

- Created a Google Form to systematically collect data on drug addiction among students at ESILV school.
- Included questions covering demographic information, academic experiences, and various aspects of drug usage patterns to ensure a comprehensive analysis.
- Implemented measures to guarantee participant anonymity and encouraged honest responses to enhance the reliability of the collected data.

Link:
https://docs.google.com/forms/d/e/1FAIpQLSdciFpJZGCfMeNfpObE1PAqkZ1xhai-3oQ-YtlxN-QbyVXzig/viewform?usp=pp_url

Enquête sur la Consommation de Substances

Merci de prendre quelques instants pour participer à notre enquête sur la consommation de substances. Les informations collectées seront utilisées à des fins éducatives dans le cadre d'un projet scolaire. Votre participation est volontaire et anonyme.

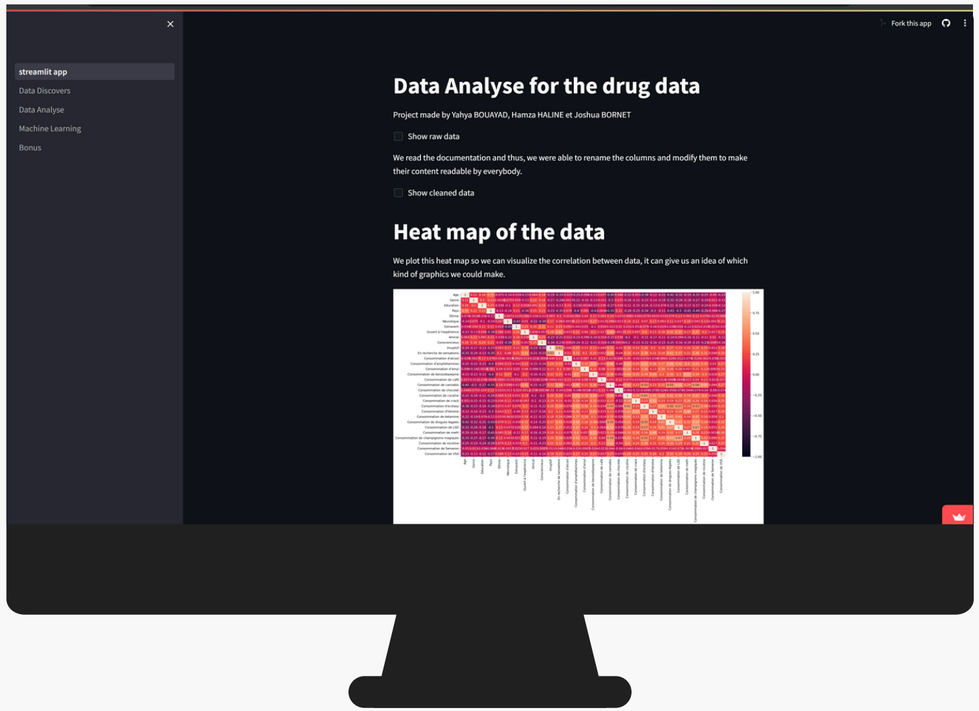
* Indique une question obligatoire

Quel est votre âge ? *

Votre réponse

05

API


Fork this app
streamlit app
Data Discovers
Data Analyse
Machine Learning
Bonus
Data Analyse for the drug data
Project made by Yahya BOUAYAD, Hamza HALINE et Joshua BORNET
Show raw data
We read the documentation and thus, we were able to rename the columns and modify them to make their content readable by everybody.
Show cleaned data
Heat map of the data
We plot this heat map so we can visualize the correlation between data, it can give us an idea of which kind of graphics we could make.

Data Analyse for the drug data

Project made by Yahya BOUAYAD, Hamza HALINE et Joshua BORNET

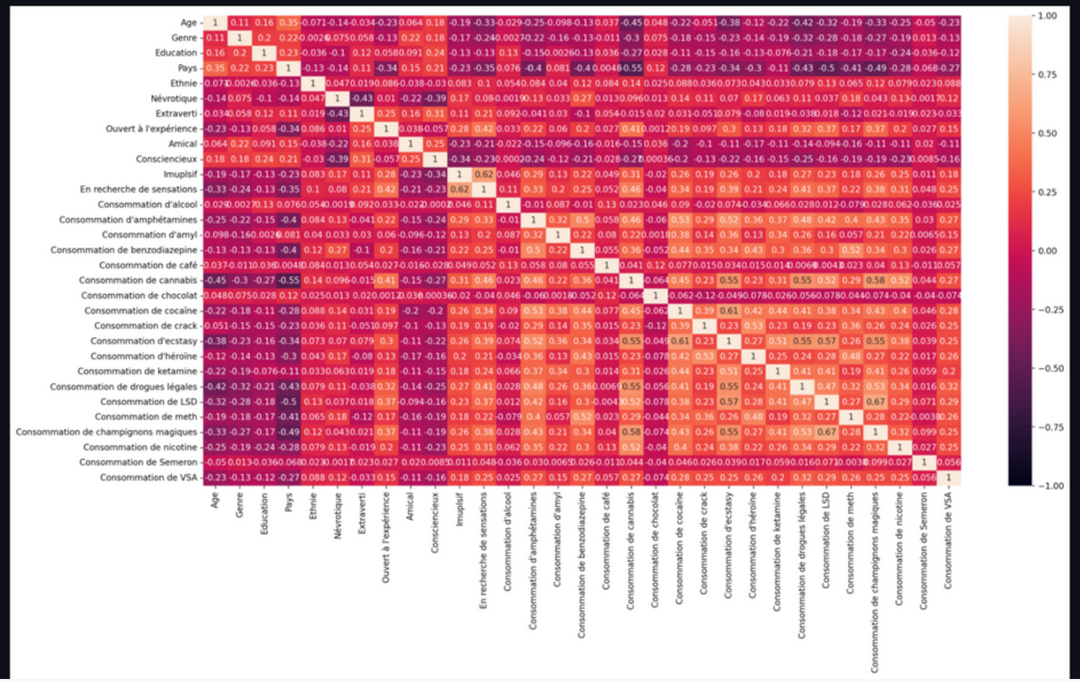
☐ Show raw data

We read the documentation and thus, we were able to rename the columns and modify them to make their content readable by everybody.

☐ Show cleaned data

Heat map of the data

We plot this heat map so we can visualize the correlation between data, it can give us an idea of which kind of graphics we could make.



Streamlit:

Link:
https://docs.google.com/forms/d/e/1FAIpQLSdciFpJZGCfMeNfpObE1PAqkZ1xhai-3oQ-YtlxN-QbyVXzig/viewform?usp=pp_url

06

Conclusions


A picture is worth a thousand words
"Don't consume drugs! Opt for coffee and chocolate instead.”

Do you have any questions?

PS: In case this is your question
No, i don’t sell any kind of drugs