Python for data analysis: Final project Drug Consumption

Maryam Dollet,

Lucas Artaud,

Iswarya Sivasubramaniam

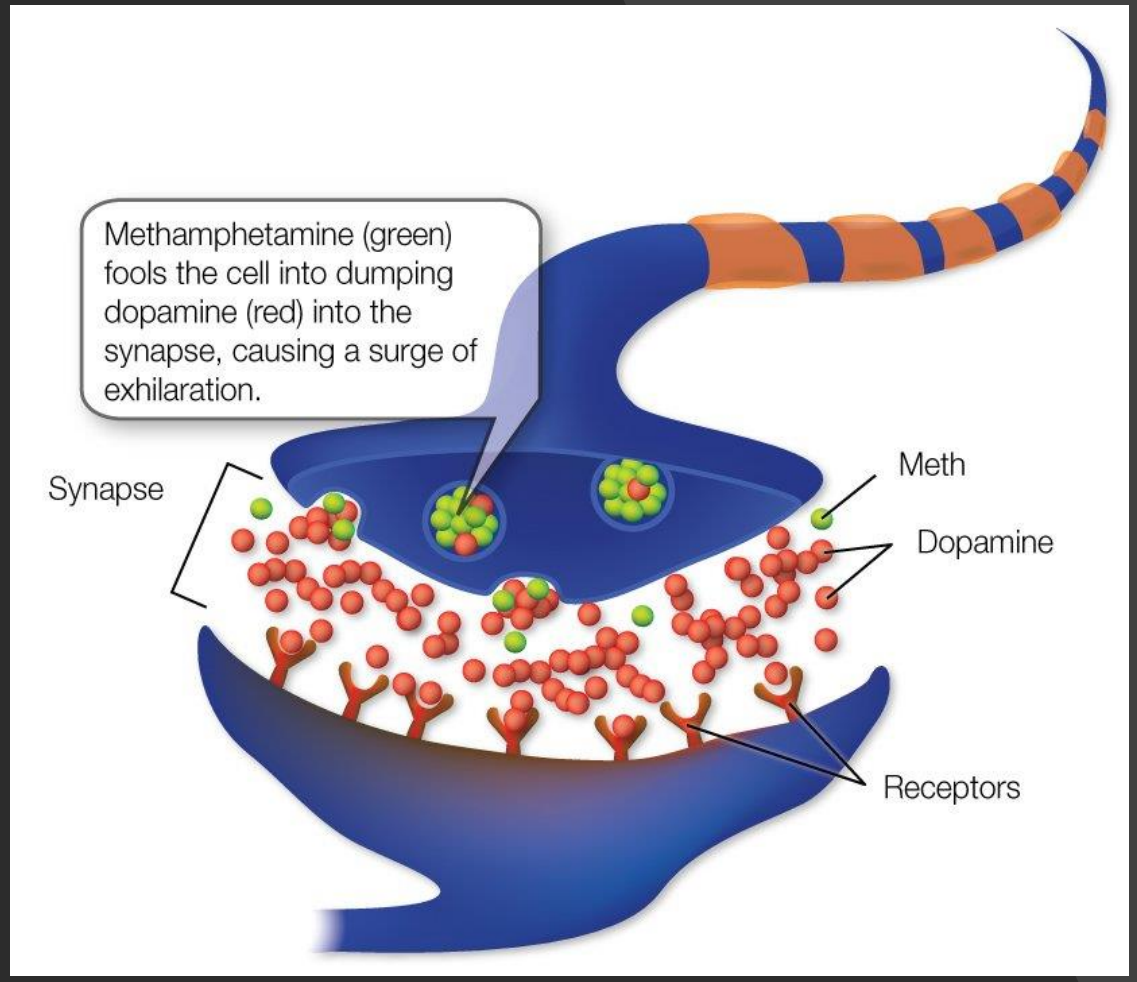
DIA1

What is a drug?

A drug is a chemical that interacts with proteins in the body to affect a physiological function.

Globally, we estimate that 11,200,000 people worldwide consume drugs.

Collectively, smoking, alcohol and illicit drug use kill 11.8 million people each year. This is more than the number of deaths from all cancers, which is about 10 million.



The Problem

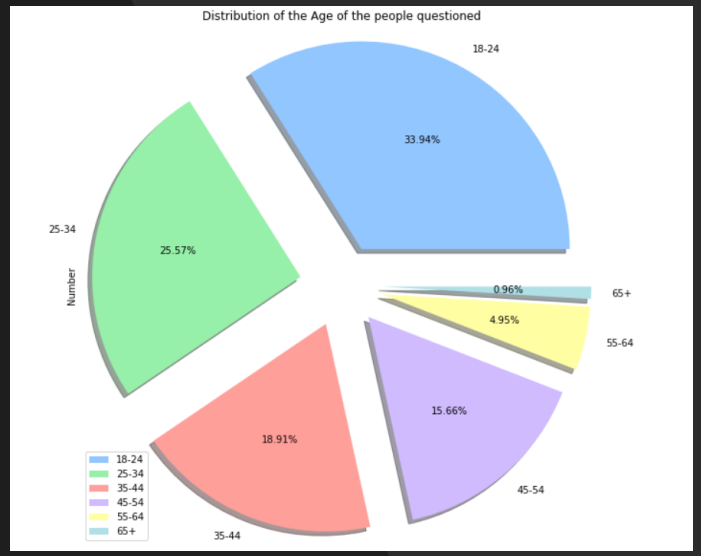
- We should prevent the usage of drugs among the population that is why we are here, members of the drug control committee, to show the extent of this problem. But also, to show how you, the government, can make a difference.

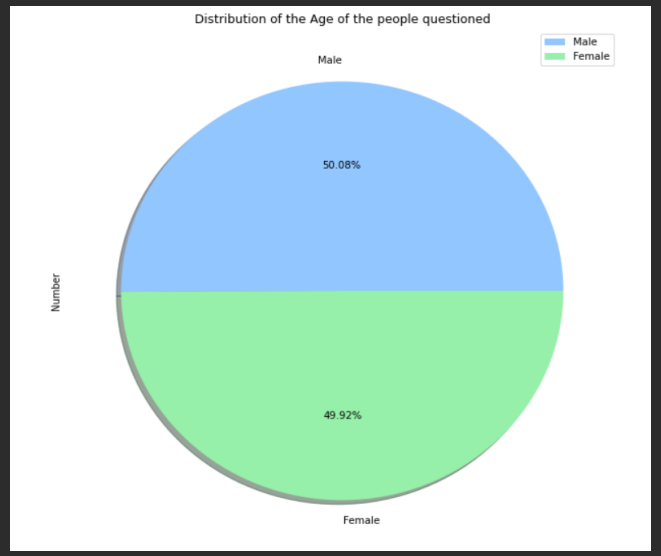
- You chose our committee to make this conduct this study and to help you make a decision on this issue.

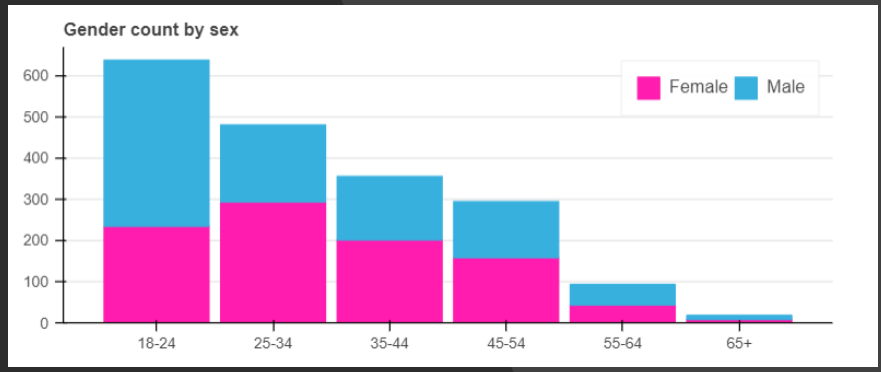
- We made this study using the data from a survey made in 2015 in the UK.

- **How can we target drug users to limit drug consumption ?**

Study on the data set



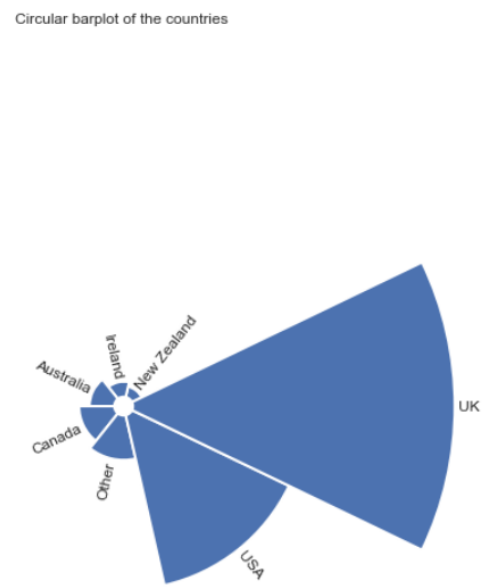


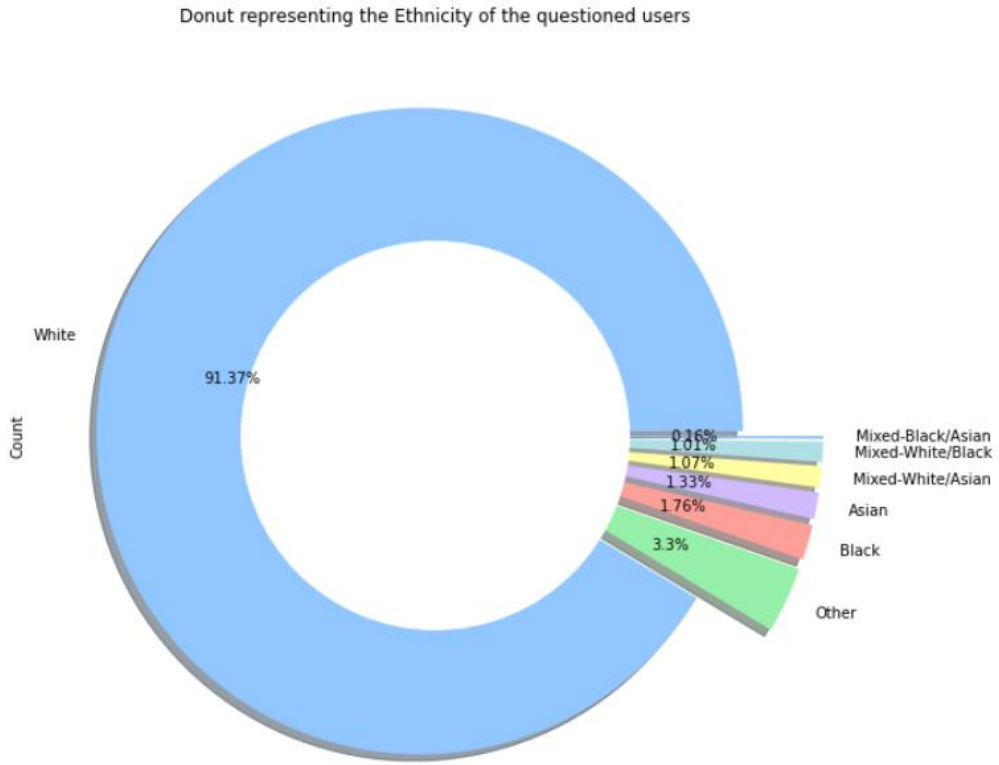


The dataset contains 32 columns (with 18 legal and illegal drugs and 7 personality scores).
We have one fictive drug Semer and we have chocolate, which medically not considered as drugs.
We have decided to not using them in our study.

Study on the data set

- We can see that the majority of the questioned people are white and from the UK,
- We can conclude that the dataset is biased in term of age and ethnicity.

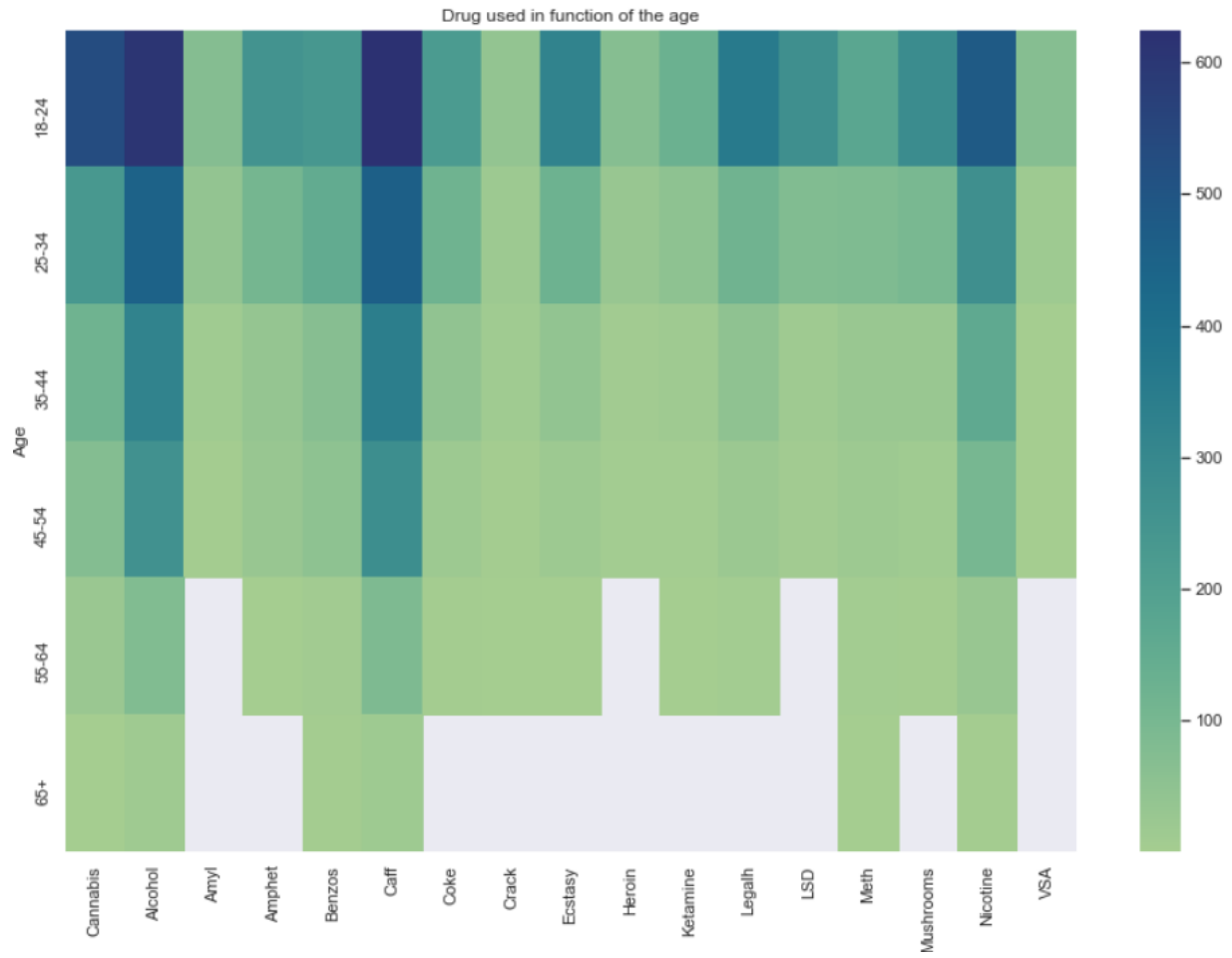




Create a typical user

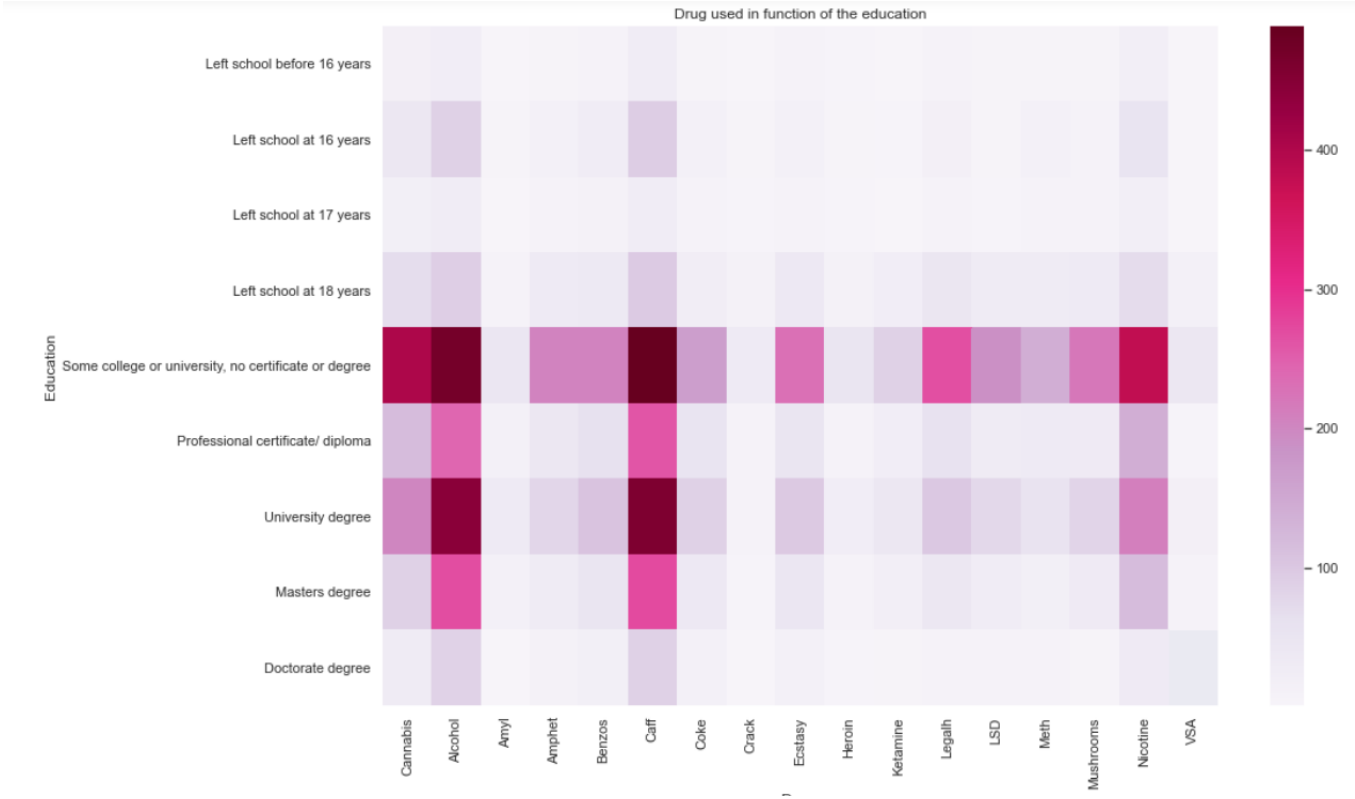
- We created some maps representing the age and the education of the respondents and the drug the used to know which drugs are used widely.

- Here we can see that most of the drug consumers have between 18 and 34 years.



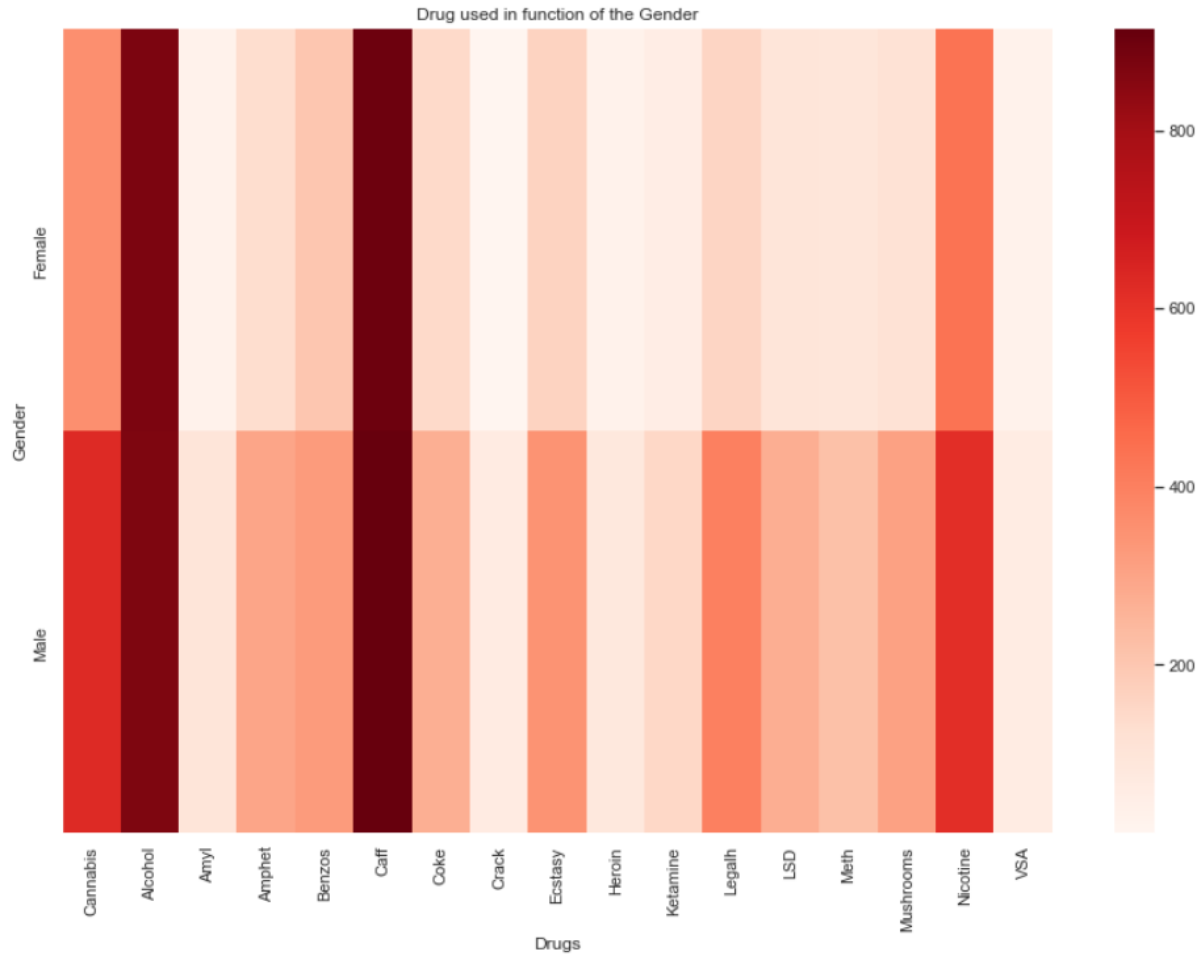
Create a typical user

- Most of the drug users are from some college or university.

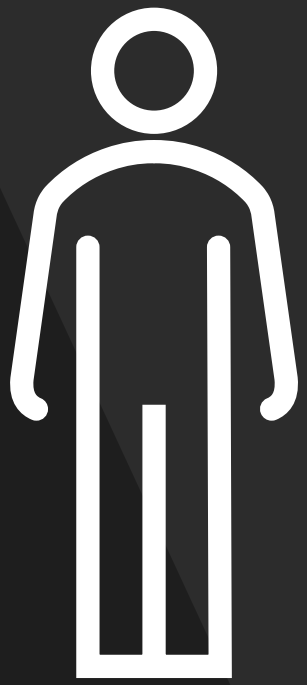


Create a typical user

- Most of the drug users are male



Persona of the drug user



Age : 18-34

Country : UK

Ethnicity : White

Gender : Male

Education : Some college or university, no certificate or degree

Used drug : legal → Alcohol, Nicotine
Illegal → Cannabis ,Ecstasy, Amphet

Machine Learning

- We wanted to develop a machine learning model which will predict the use of a drug by a person with his personal information (age, ethnicity, education, country, gender and the use other drugs)
- We binarized the data because it’s easier to predict 0 or 1 than a value between 0 and 7
- It’s 1 if the drug was used during the last year, or 0 otherwise

Streamlit representing our study

Main Menu

- Home
- Data Analysis
- **Machine Learning**

Risk of being a certain drug's consumer

We will now test 6 different models of Classification and try to predict, if the individual has consumed the drug in question

First we will try without binarizing the drug we want to predict. When we binarize, the model will predict if the individual has consumed the drug in question over the last 12 months (1 year)

It might take a while to perform 6 models on 17 drug colums.

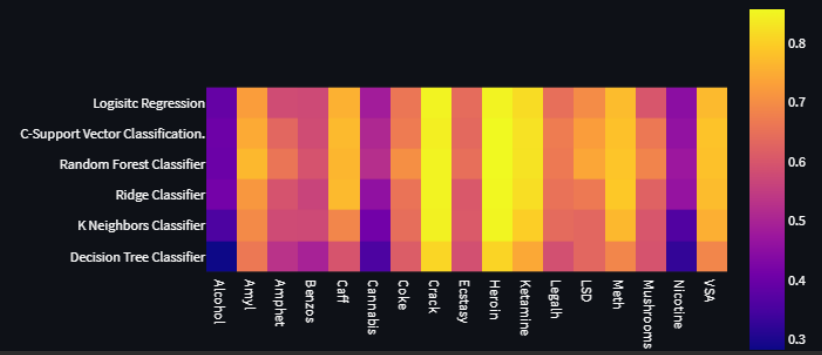
The purpose is to find the best model for each drug, to evaluate this we will calculate the accuracy of each model

Click on the button to start the analysis !

Start

General Analysis : Heatmaps

Accuracy of each model tested on each drug



Conclusion

- It’s important to focus on cannabis, which is the first illicit drug and on young people especially on men.

- Moreover, to have a detailed report, we should do another survey and diversify the people we questioned (especially the ethnicity and age) in order to have a more precise model to predict the drug used from the details given.

- We have to do some preventions and also implement more drug rehabilitation centers near the schools and colleges.

Thank you for your attention