



DRUGS CONSUMPTION

Python for Data Analysis

ESILV A₄

Alexandra CHUVATIN | Axel GUEDJ

THE DATASET

- Our dataset is about drug consumption. It contains records for 1885 respondents for which 12 attributes are known. Addition to this we have their consumption for 18 legal and illegal drugs.
- We do not have missing values.

1. ID is number of record in original database. Cannot be related to participant. It can be used for reference only.
2. Age is age of participant
3. Gender is gender of participant
4. Education is level of education of participant
5. Country is country of current residence of participant
6. Ethnicity is ethnicity of participant
7. Nscore is NEO-FFI-R Neuroticism
8. Escore is NEO-FFI-R Extraversion
9. Oscore is NEO-FFI-R Openness to experience
10. Ascore is NEO-FFI-R Agreeableness
11. Cscore is NEO-FFI-R Conscientiousness

12. Impulsive is impulsiveness measured by BIS-11
13. SS is sensation seeking measured by ImpSS
14. Alcohol is class of alcohol consumption
15. Amphet is class of amphetamines consumption
16. Amyl is class of amyl nitrite consumption
17. Benzos is class of benzodiazepine consumption
18. Caff is class of caffeine consumption
19. Cannabis is class of cannabis consumption
20. Choc is class of chocolate consumption
21. Coke is class of cocaine consumption
22. Crack is class of crack consumption
23. Ecstasy is class of ecstasy consumption

24. Heroin is class of heroin consumption
25. Ketamine is class of ketamine consumption
26. Legalh is class of legal highs consumption
27. LSD is class of LSD consumption
28. Meth is class of methadone consumption
29. Mushrooms is class of magic mushrooms consumption
30. Nicotine is class of nicotine consumption
31. Semer is class of fictitious drug Semeron consumption
32. VSA is class of volatile substance abuse consumption

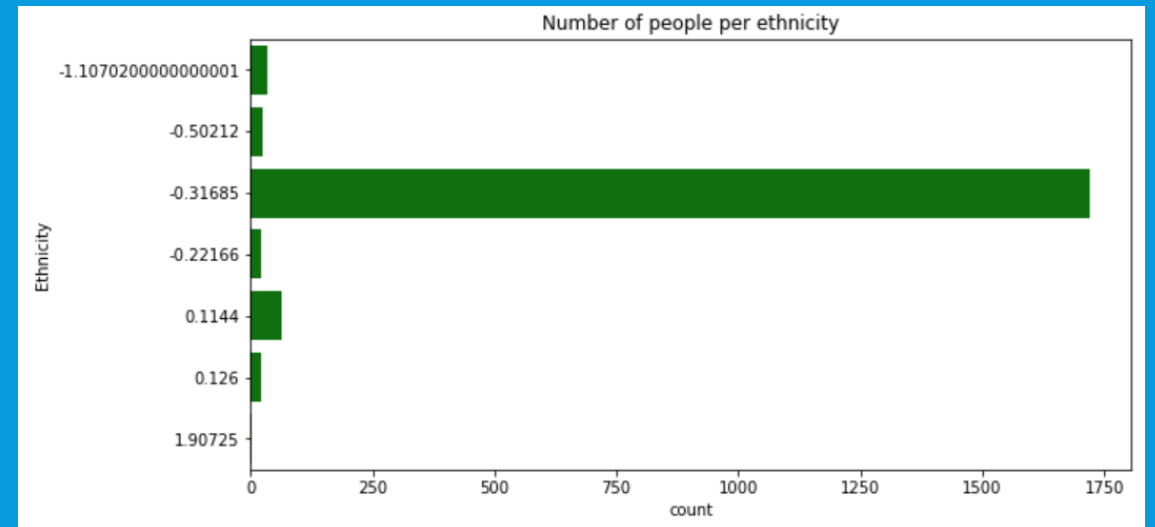
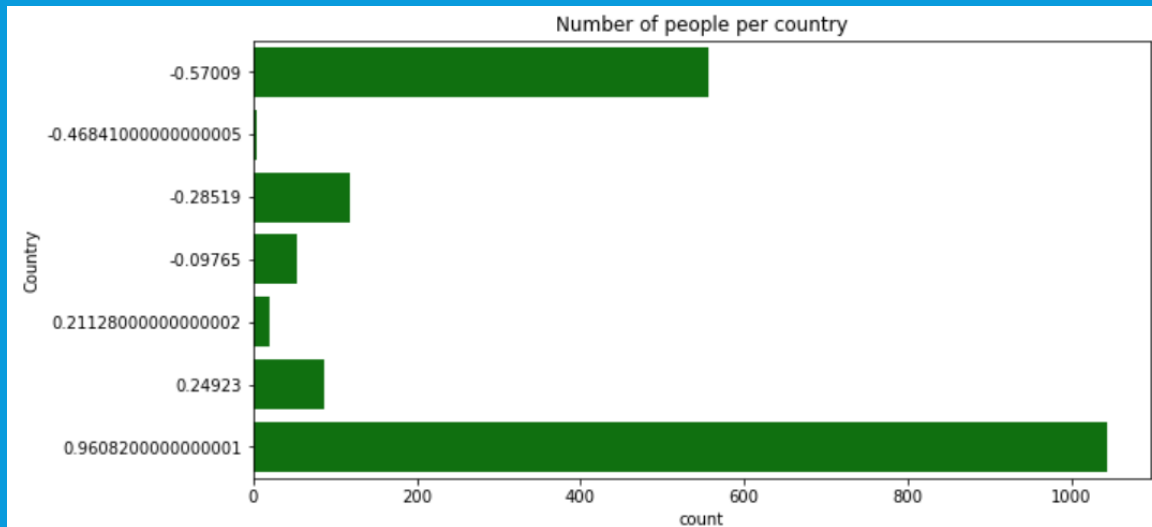
THE GOAL

We want to be able to predict whether the person is a user or not of a chosen drug.

- To do so, we will look at the data and see which features will help us.
- Then, we will class the drugs in categories to help us in our predictions.
- Finally, we will do some encoding for the qualitative features.
- The last task will be to find the model that predicts the best the usage of each drug class !

FIRST OBSERVATIONS

- A good point is that for the **gender** feature we have 50% of male and 50% of female.
- But when we look closer, some features can not help us because of a bad distribution. Thus, we can not use **ethnicity** and **country** in our predictions.



DRUG USERS

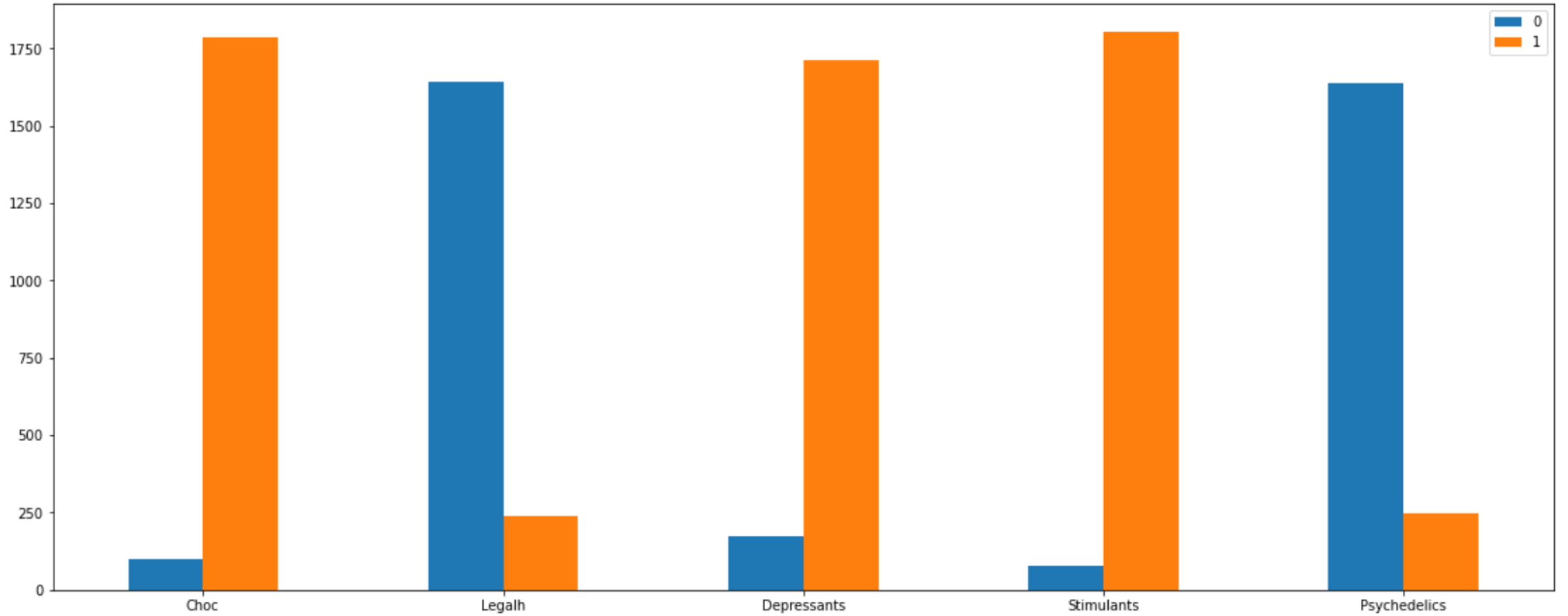
- The first task is to class the participants into 2 groups: “Users” and “Non-Users”. To do so, we decided that a **User** of a drug is someone that gave one of these answers: ‘**Used in last day**’, ‘**Used in last week**’, ‘**Used in last month**’. The rest will be **Non-User**. The goal is to have a one in the dataset if the person is considered as a user and a zero in the other case.
- Next, we put the drugs in 3 different classes: Depressants, Stimulants and Psychedelics according to this site: <https://www.health.gov.au/health-topics/drugs/about-drugs/types-of-drugs>

Unfortunately, we couldn’t class « Choc » and « LegalH » into one of the classes.

NB: first we tried to use the [Drug Wheel](#) to class the drugs, but too many of them weren’t part of a class. We also tried to use the distinction between hard and soft drug, but it appeared it wasn’t a good classification for drugs.

- Finally, we had only one user of the drug « Semer », so we dropped this column because it wasn’t relevant.

Cheat-Sheet: 0 = Non-User, 1 = User



USERS VS NON-USERS FOR EACH DRUG

Decode Age column

18-24 age = 0

25-34 age = 1

35-44 age = 2

45-54 age = 3

55-64 age = 4

65+ age = 5

Decode Education

Left school before 16 years = 0

Left school at 16 years = 1

Left school at 17 years = 2

Left school at 18 years = 3

Some college or university, no certificate or degree = 4

Professional certificate/ diploma = 5

University degree = 6

Masters degree = 7

Doctorate degree = 8

Decode Gender

Female = 0

Male = 1

ENCODING

- The features **Age**, **Gender** and **Education** were already encoded, but we will encode them differently, as seen in class. First we will decode them to easier understand the values.
- Then, to encode them, we use the *get_dummies()* function of the pandas library.

Now, instead of Age (0/1/2/3/4/5) we have *Age_0*, *Age_1*, ...

- Our final columns are:

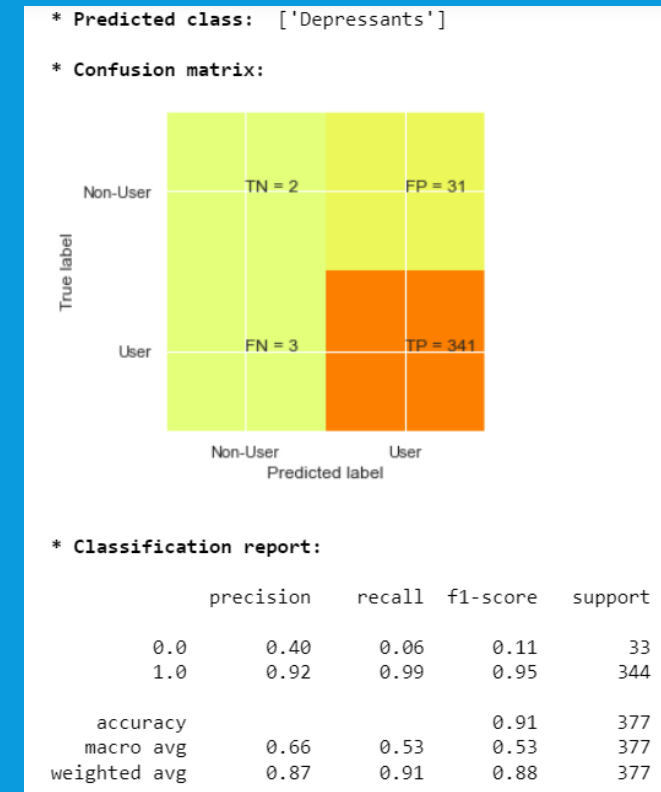
'Nscore', 'Escore', 'Oscore', 'Ascore', 'Cscore', 'Impulsive', 'SS', 'Education_0', 'Education_1', 'Education_2', 'Education_3', 'Education_4', 'Education_5', 'Education_6', 'Education_7', 'Education_8', 'Age_0', 'Age_1', 'Age_2', 'Age_3', 'Age_4', 'Age_5', 'Education_0', 'Education_1', 'Education_2', 'Education_3', 'Education_4', 'Education_5', 'Education_6', 'Education_7', 'Education_8', 'Age_0', 'Age_1', 'Age_2', 'Age_3', 'Age_4', 'Age_5', 'Gender_0', 'Gender_1' + the drugs

PREDICTIONS

- Our model will be a classification model because we want to know whether someone is a « User » or a « Non-User ».
- After standardizing our data, we tried several algorithms such as SVC, KNN, Decision Tree...

The best model we found is Decision Tree with an accuracy of **0.91** !

To compare our models, we used graphics to compute the confusion matrix, and we showed different scores as the accuracy, the recall...



FLASK INTERFACE

- First, to use our model on our Flask application, we saved it using the library pickle.
- Flask helped us creat an internet application thanks to which we can fill some questions about the features we need about a person for our prediction.
- Once all the values are wrote, the model we saved with pickle is used to find if the person is a consumer of the chosen drug (1 if yes, 0 if no).

NB: to load localhost:5000, we need to run the file app.py just before (available in 'predict drugs' on the github)

FLASK INTERFACE

Drugs

How old are you ??

18-24 years ? : ☒ Non ☐ Oui
25-34 years ? : ☒ Non ☐ Oui
35-44 years ☒ Non ☐ Oui
45-54 years ☒ Non ☐ Oui
55-64 years ☒ Non ☐ Oui
65+ years ☒ Non ☐ Oui

Are you ? :

Male ☒ Non ☐ Oui
Female ☒ Non ☐ Oui

Nscore	Escore	Oscore	Ascore
Cscore	Impulsive	SS	

PREDICT IF YOU TAKE DRUGS

Please choose a level of studies :

Left school before 16 years ☒ Non ☐ Oui
Left school at 16 years ☒ Non ☐ Oui
Left school at 17 years ☒ Non ☐ Oui
Left school at 18 years ☒ Non ☐ Oui
Some college or university, no certificate or degree ☒ Non ☐ Oui
Professional certificate/ diploma ☒ Non ☐ Oui
University degree ☒ Non ☐ Oui
Masters degree ☒ Non ☐ Oui
Doctorate degree ☒ Non ☐ Oui

- To use the interface, we need to specify the needed questions. (put a 'yes' for the category the person belongs to for the age, education and gender).
- Push the predict button...
- The answer appears just below !

Depressants consumer ? (1:yes, 0:no) : 1.0

NB: On the flask interface, we predict wether the person will be a Depressant drug consumer or not. To upgrade this, we can add the prediction for all the types of drugs.