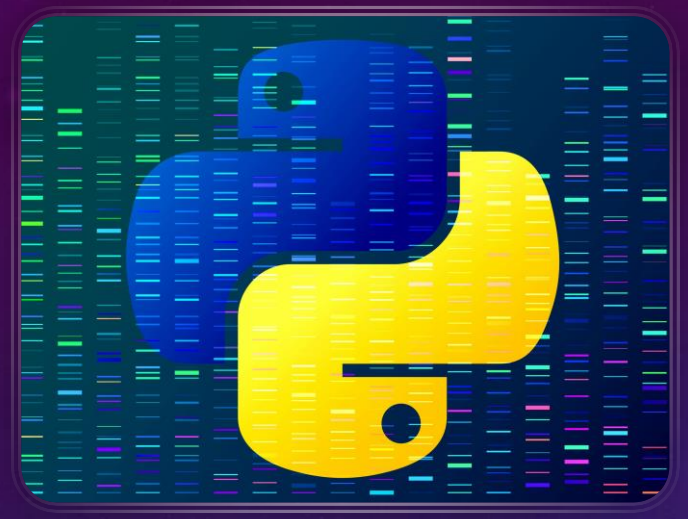

FINAL PROJECT PYTHON FOR DATA ANALYSIS DRUG CONSUMPTION (QUANTIFIED)

- Louis Melchior Giraud
- Etienne Goury
- Aramis Fayolle

OUR THOUGHTS ON THE ASKED QUESTION

- The study is reliable and has been cited several times in other publications.
- The data is clean, no missing values and the data is normalized.
- There are intuitive correlation between the explanatory variables and the target.
- The psychological effects of these drugs are already known and specific, so we believe that the classification models will work well on the dataset.

DATASET

Database contains records for 1885 respondents

For each respondent 12 attributes are known

fictitious drug (Semeron) was introduced to identify over-claimers

Age	Gender	Education	Country	Ethnicity
0.49788	0.48246	-0.05921	0.96082	0.126
-0.07854	-0.48246	1.98437	0.96082	-0.31685

Age	Gender	Education	Country	Ethnicity
35-44	Female	Professional certificate/ diploma	UK	Mixed-White/Asian
25-34	Male	Doctorate degree	UK	White

Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS
0.31287	-0.57545	-0.58331	-0.91699	-0.00665	-0.21712	-1.18084
-0.67825	1.93886	1.43533	0.76096	-0.14277	-0.71126	-0.21575
-0.46725	0.80523	-0.84732	-1.6209	-1.0145	-1.37983	0.40148

Alcohol	Amphet	Amyl	Benzos	Caff	Cannabis	Choc	Coke	Crack	Ecstasy	Heroin	Ketamine	Legalh	LSD	Meth	Mushrooms	Nicotine	Semer	VSA
CL4	CL1	CL1	CL0	CL6	CL3	CL6	CL0	CL0	CL1	CL0	CL0	CL1	CL0	CL0	CL2	CL2	CL0	CL0

Alcohol	Amphet	Amyl	Benzos	Caff	Cannabis	Choc	Coke	Crack	Ecstasy	Heroin	Ketamine	Legalh	LSD	Meth	Mushrooms	Nicotine	Semer	VSA
Used in Last Month	Used over a Decade Ago	Used over a Decade Ago	Never Used	Used in Last Day	Used in Last Year	Used in Last Day	Never Used	Never Used	Used over a Decade Ago	Never Used	Never Used	Used over a Decade Ago	Never Used	Never Used	Used in Last Decade	Used in Last Decade	Never Used	Never Used

PROBLEMS WHICH CAN BE SOLVED

Seven class classifications for each drug separately

Problem can be transformed to binary classification by union of part of classes into one new class
For example, "Never Used", "Used over a Decade Ago" form class "Non-user" and all other classes form class "User"

The best binarization of classes for each attribute

Evaluation of risk to be drug consumer for each drug

LIMITS OF THE DATASET AND THE STUDY

THE AGE OF THE PEOPLE IN THE DATASET ISN'T BALANCED, HALF OF THE PEOPLE ARE BETWEEN 18 AND 35 YEARS.

IT DOESN'T INCLUDE A LARGE DIVERSITY OF ETHNICITIES AND COUNTRIES

The results can be biased because of this