Drugs consumers

Rayane Badji

Arthur Baudier

Ins & Outs

Facts to remember

- Nearly half of college students use illicit drugs.

- In 2020,
- 91,799 drug overdose deaths occurred in the United States

- In 2021,
- 107,622 drug overdose deaths in the US

Our take on the issue

- Find a way to lower the number of drugs consumers by assuming or predicting drugs consumption based by using IA aptitudes.

Dataset

Our dataset is composed of 3 main types of columns :

1. Those describing the profile of the individuals
2. Those describing the personality of the individuals
3. Those describing the frequency of drug use of these individuals

Dataset size :1885 x 32

profile

ID	age	gender	education	country	ethnicity
1	[35, 45]	Female	Professional certificate/ diploma	UK	White-Asian
2	[25, 34]	Male	Doctorate degree	UK	White
3	[35, 45]	Male	Professional certificate/ diploma	UK	White

personality

ID	Stress_score	escore	oscore	ascore	csore	impulsive_bis11	sensation_ImpSS
1	56.06	41.21	43.57	36.77	49.90	42.85	22.44
2	41.35	79.61	76.26	60.98	47.94	33.79	46.57
3	44.48	62.30	39.29	26.61	35.36	21.54	62.00

Frequency of drug use

| ID	alcohol	amphet	benzos	cannabis	coke	crack	ecstasy	heroin	ketamine	legalhigh	LSD	meth
1	LastWeek	LastDecade	LastDecade	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed
2	LastWeek	LastDecade	NeverUsed	LastMonth	LastYear	NeverUsed	LastMonth	NeverUsed	LastDecade	NeverUsed	LastDecade	LastYear
3	LastDay	NeverUsed	NeverUsed	LastYear	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed	NeverUsed
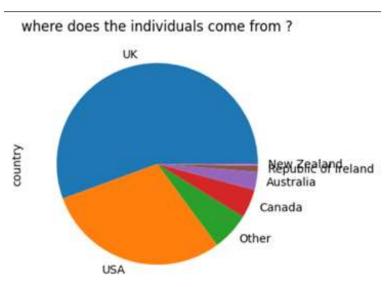
Zoom on the profile of our population
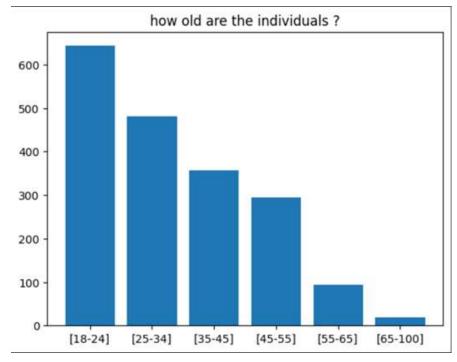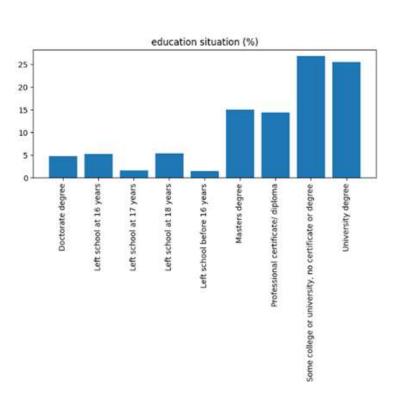
localisation	age	education level
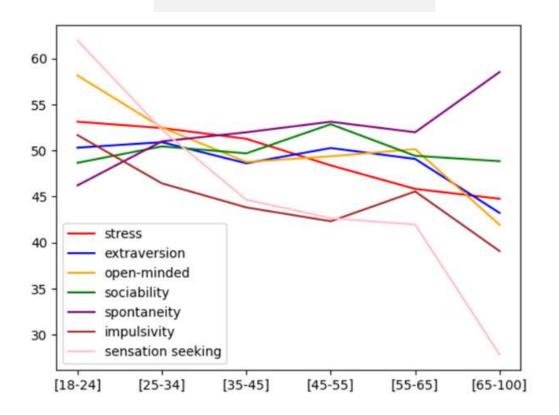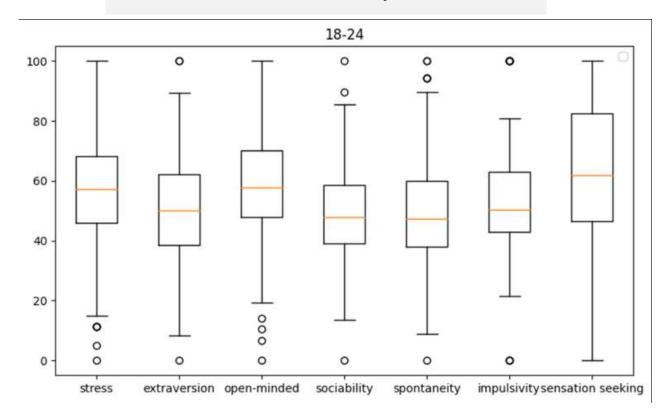
Zoom on the personality of our population
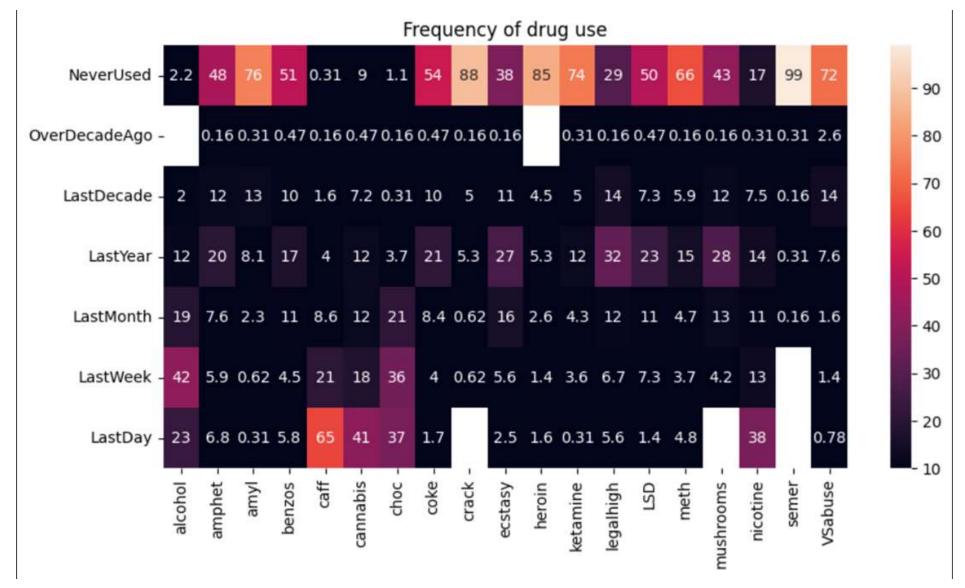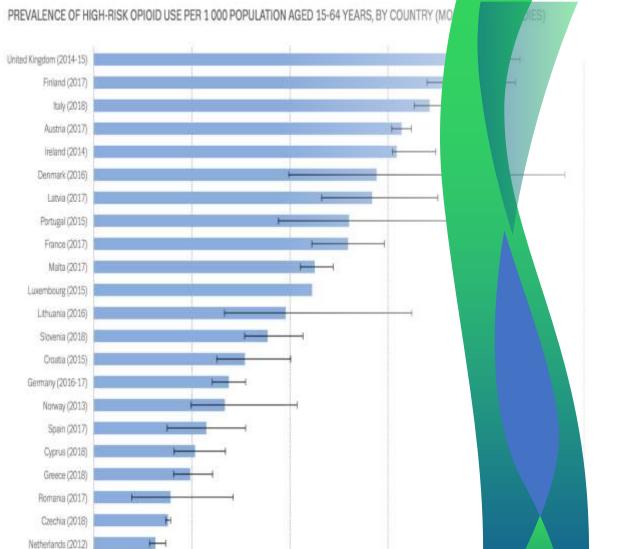
| Personality by age | Focus on the 18-24 years olds |

Zoom on the drugs taken by our population

Frequency of drug use

	alcohol	amphet	amyl	benzos	caff	cannabis	choc	coke	crack	ecstasy	heroin	ketamine	legalhigh	LSD	meth	mushrooms	nicotine	semer	VSabuse
NeverUsed	2.2	48	76	51	0.31	9	1.1	54	88	38	85	74	29	50	66	43	17	99	72
OverDecadeAgo		0.16	0.31	0.47	0.16	0.47	0.16	0.47	0.16	0.16		0.31	0.16	0.47	0.16	0.16	0.31	0.31	2.6
LastDecade	2	12	13	10	1.6	7.2	0.31	10	5	11	4.5	5	14	7.3	5.9	12	7.5	0.16	14
LastYear	12	20	8.1	17	4	12	3.7	21	5.3	27	5.3	12	32	23	15	28	14	0.31	7.6
LastMonth	19	7.6	2.3	11	8.6	12	21	8.4	0.62	16	2.6	4.3	12	11	4.7	13	11	0.16	1.6
LastWeek	42	5.9	0.62	4.5	21	18	36	4	0.62	5.6	1.4	3.6	6.7	7.3	3.7	4.2	13		1.4
LastDay	23	6.8	0.31	5.8	65	41	37	1.7		2.5	1.6	0.31	5.6	1.4	4.8		38		0.78

PREVALENCE OF HIGH-RISK OPIOID USE PER 1 000 POPULATION AGED 15-64 YEARS, BY COUNTRY (MO... ...IES)

Prevalence per 1 000 population aged 15-64 years (with uncertainty intervals)

Problematic

How can we correlate our dataset and what we know about drug consumption

How can we assume or predict the drugs consumptions of young consumers to prevent drog abuse ?

Smoking, alcohol and drug use is an important risk factor for early death: 11.4 million die prematurely as a result each year. Over 350,000 die from overdoses (alcohol and illicit drug use disorders) each year.

Prevention by prediction

Consumption abuse

> In order to highlight the information that interests us and have a coherent result at the end of the analysis, we must follow specific steps.

Supervised learning

Choose the model depending on its accuracy

Unsupervised learning

PCA
Kmeans

Visualization

- Getting an overview of our data

Pre-Processing

- Preparing the data for our problematic

Steps we followed
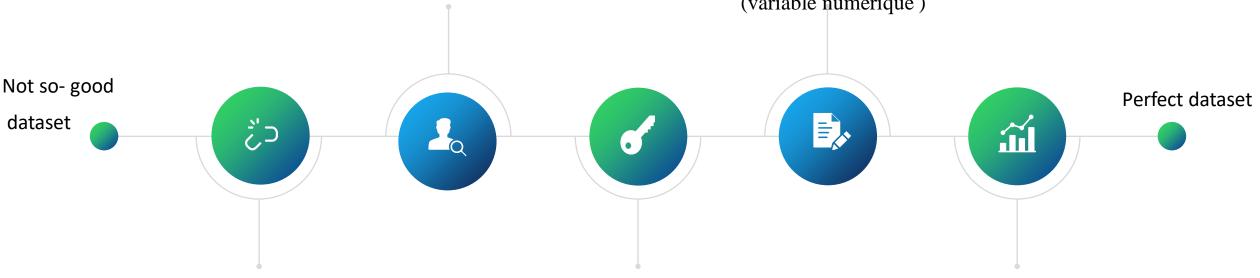
Important parts of the Pre-processing

Outliers

- Example : Incoherent value CL1 for young consumers

Modified variables

- Example : age - > distribution gaussienne (variable numérique)

Not so- good
dataset

Perfect dataset

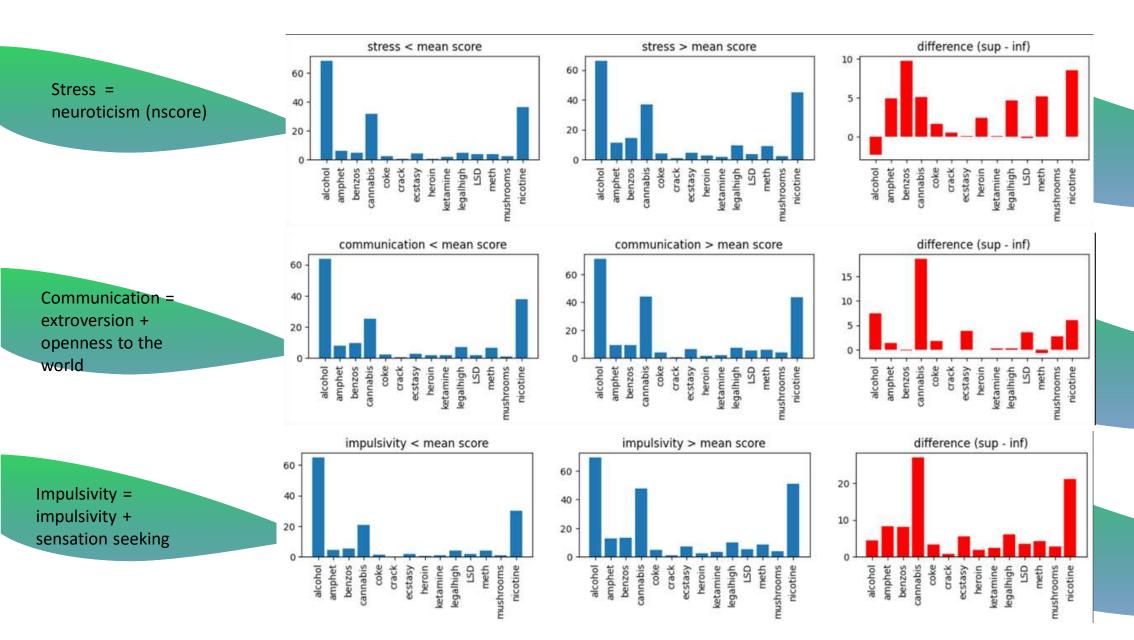## Data not correlated with our problematic

- Our target is not stopping people from eating chocolate or drinking coffee

New variables

- Number of drugs taken

- Global scores

Encoding and normalizing

Link between personality and frequent drug use (at least once a week)

Stress =
neuroticism (nscore)

Communication =
extroversion +
openness to the
world

Impulsivity =
impulsivity +
sensation seeking
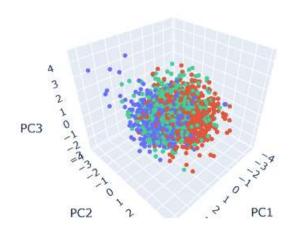
Unsupervised models

PCA

K-means

Total Explained Variance: 22.86%

target_cannabis
- NeverUsed
- UsedThisYear
- UsedOverYear
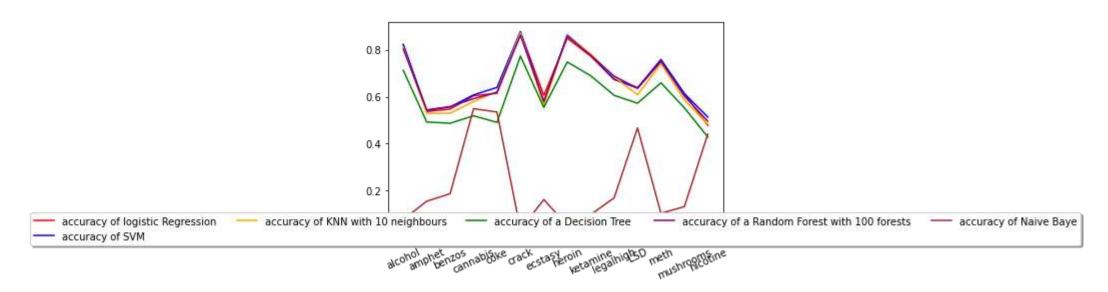
Cluster	Stress_score	cscore	communicative_score	impulsive_score	age	cannabis
0	46.0	58.0	48.0	33.0	39.0	1.0
1	44.0	52.0	62.0	65.0	27.0	2.0
2	66.0	38.0	47.0	56.0	28.0	1.0

```
Result: 750 out of 1869 samples were correctly labeled.
Accuracy score: 0.40
```
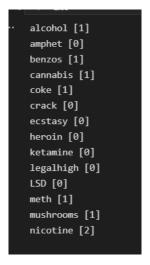
3 types of consumers -> we can reduce the number of labels for the drugs columns

Supervised models

accuracy of logistic Regression — accuracy of KNN with 10 neighbours — accuracy of a Decision Tree — accuracy of a Random Forest with 100 forests — accuracy of Naive Baye
accuracy of SVM

alcohol amphet benzos cannabis coke crack ecstasy heroin ketamine legalhigh LSD meth mushrooms nicotine
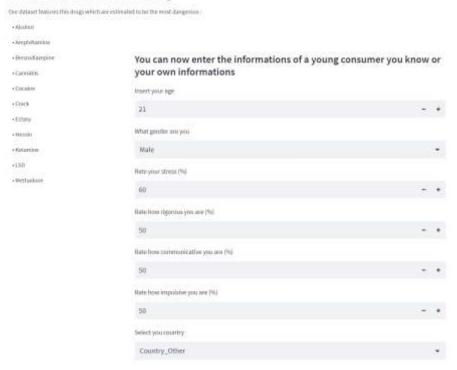
	age	gender	Stress_score	cscore	communicative_score	impulsive_score
0	21	1	60	70	60	40

alcohol [1]
amphet [0]
benzos [1]
cannabis [1]
coke [1]
crack [0]
ecstasy [0]
heroin [0]
ketamine [0]
legalhigh [0]
LSD [0]
meth [1]
mushrooms [1]
nicotine [2]

How the API looks

We will explore the dataset "drugs consumers" to predict if a young consumer is taking drugs currently or will use it eventually

Our dataset features this drugs which are estimated to be the most dangerous :

- Alcohol
- Amphetamine
- Benzodiazepine
- Cannabis
- Cocaine
- Crack
- Ectasy
- Heroin
- Ketamine
- LSD
- Methadone

You can now enter the informations of a young consumer you know or your own informations

Insert your age

21

What gender are you

Male

Rate your stress (%)

60

Rate how rigorous you are (%)

50

Rate how communicative you are (%)

50

Rate how impulsive you are (%)

50

Select you country

Country_Other

You will now have to choose the model you want to use to predict the frequence of consumptions of a consumer you know

You can see here the accuracy of each model

Select the model you want to use to guess and predict

Naive Baye

	age	gender	stress_score	cocore	communicative_score	impulsive_score	Country_Australia	Cou
0	21	1	60	50		50	50	0

alcohol : You used it this year and you may use it again 1

amphet : Never Used it in your life 0

benzos : You used it before this year 1

cannabis : You used it this year and you may use it again 2

coke : You used it this year and you may use it again 2

crack : Never Used it in your life 0

ecstasy : You used it before this year 1

heroin : Never Used it in your life 0

ketamine : Never Used it in your life 0

legalhigh : You used it before this year 1

LSD : Never Used it in your life 0

meth : Never Used it in your life 0

mushrooms : You used it before this year 1

nicotine : You used it this year and you may use it again 2

The end

Thank you for listening