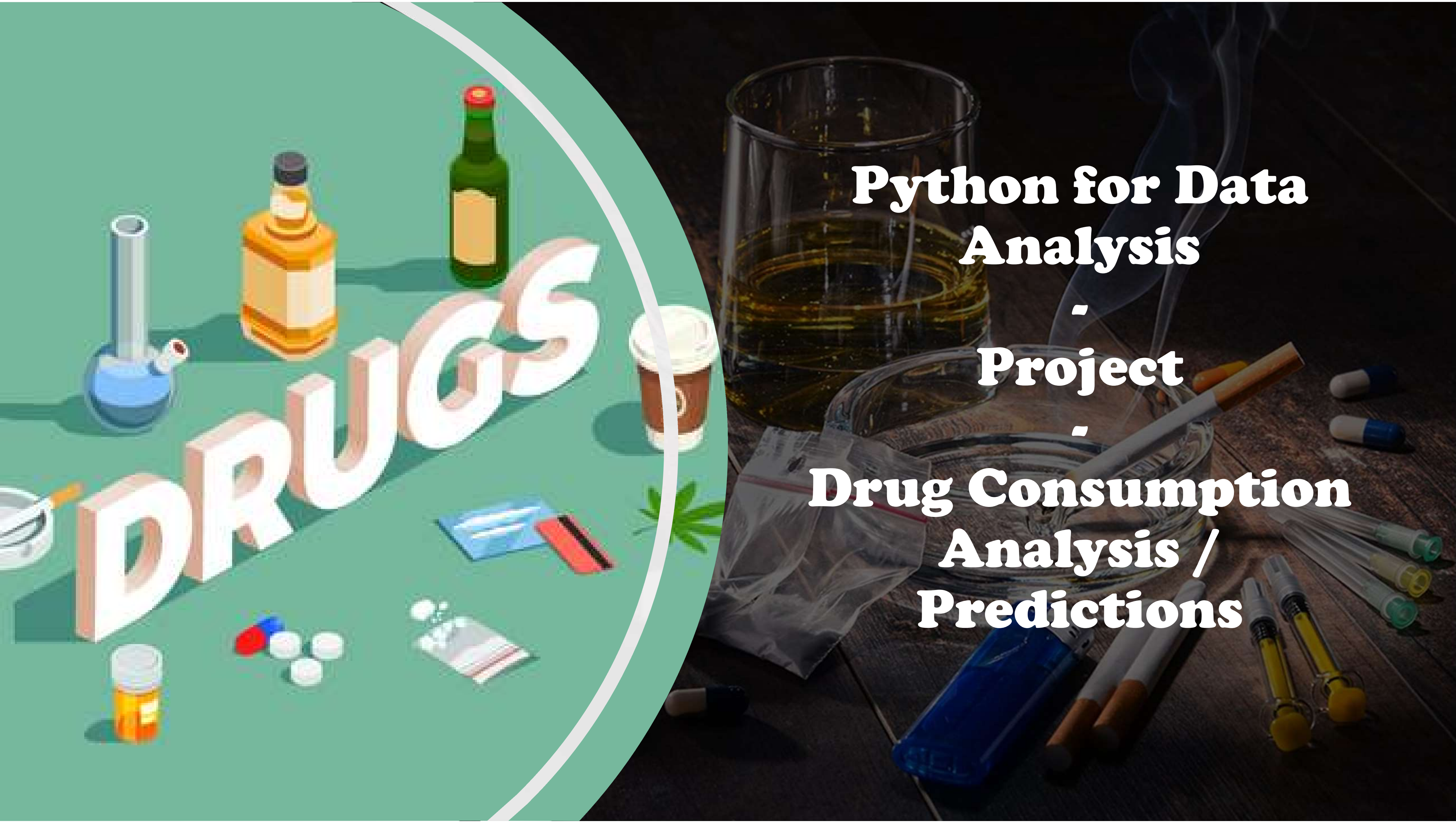


Python for Data Analysis
-
Project
-
Drug Consumption Analysis / Predictions

Data Set

More informations : https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29

5 Demographic Features

- Age
- Gender
- Education
- Country
- Ethnicity

7 Personality Features

- Neuroticism
- Extraversion
- Openness to experience
- Agreeableness
- Conscientiousness
- Impulsiveness
- Sensation seeking

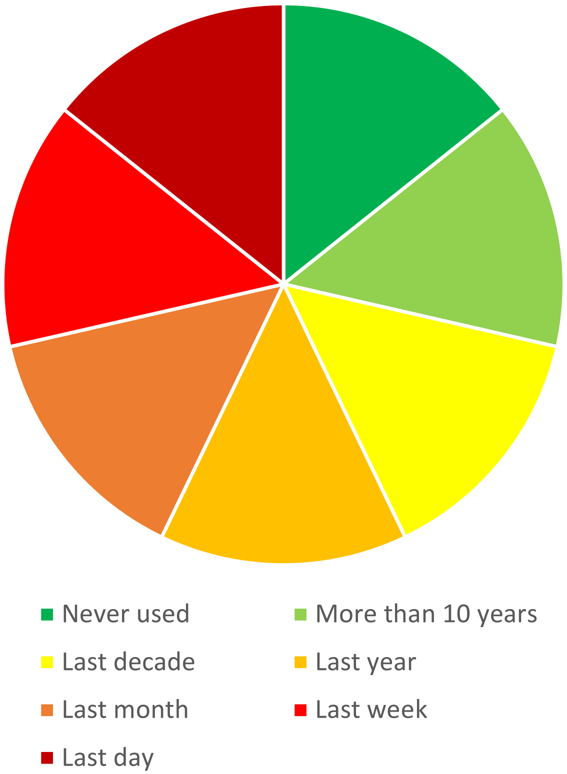
19 Drug Consumption targets

- Alcohol
- Amphetamines
- Amyl nitrite
- Benzodiazepine
- Caffeine
- Chocolate
- Cocaïne
- Crack
- Ecstasy
- Heroin
- Ketamine
- Legal highs
- Lysergic acid diethylamide
- Methadone
- Magic mushrooms
- Nicotine
- Volatile substance abuse
- Fictitious drug Semeron*

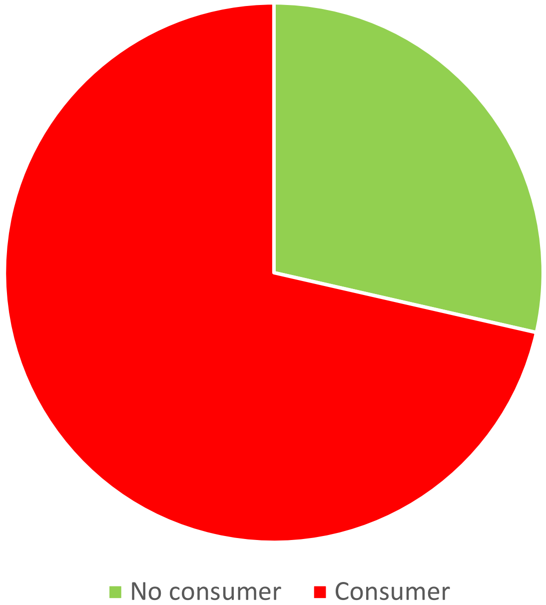
*note : that last drug is ficticious : it will be removed from the dataset as well as the participants who lied about it

Problems we'll try to solve

Find a model to predict between 7 classes for each drug



Find a model to predict between 2 classes for each drug



Variables created

Dataframes

- **df** : contains all the dataset from the csv file
- **df_drugs** : contains all the 7 definitions of the classes (replacing 0 by 'Never used', etc.)
- **df_corr :** correlation matrix of the dataframe
- **Y_drugs** : contains a binary value for each drug column : 0 for 'not user' and 1 for 'user'

Lists

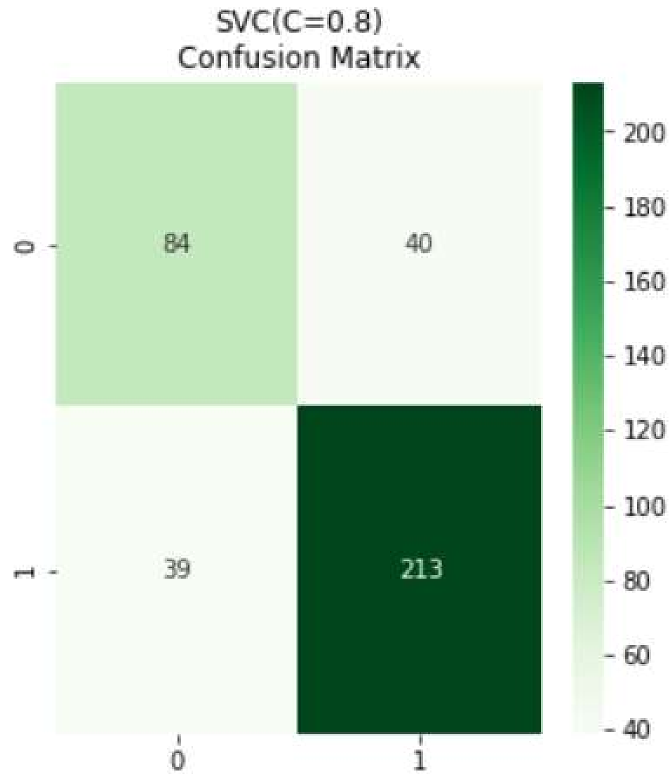
- **demographic** : names of the demographic columns
- **personality** : names of the personality columns
- **drugs** : names of the drug consumption columns
- **age_cat** : scaled values of ages converted to category values
- **gender_cat** : scaled values of genders converted to category values
- **education_cat** : scaled values of educations converted to category values
- **country_cat** : scaled values of countries converted to category values
- **ethnicity_cat** : scaled values of ethnicities converted to category values
- **algos** : dictionnary containing classification algorithmes we will use, and their parameters we will try
- **ests** : result of the function *predict_all_consumptions*

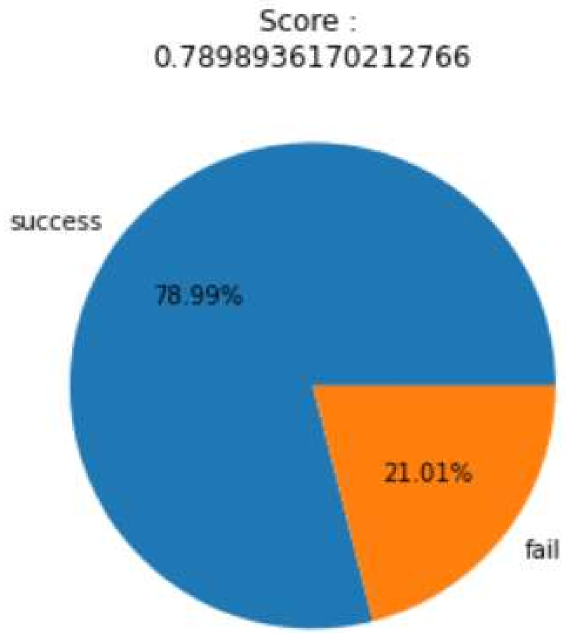
Functions

- **frequency_plots** : function to give for each column of a dataframe, a pie chart and a barplot of value counts
- **plot_and_table** : function to give for a given column of a dataframe, a pie chart, a barplot and a tab of value counts
- **test_algo** : applies a gridsearchcv and returns the best estimator and best score
- **classifiers_test** : returns a list of estimators for each algorithmes
- **disp_scores** : plot a confusion matrix and a pie chart of the success rate
- **predict_all_consumptions** : returns list of the best estimator for each target of the problem

Predicting consumer / not consumer

```
Cannabis consumption
KNeighborsClassifier fitted, with score : 0.7961461794019933
SVC fitted, with score : 0.8127906976744186
LogisticRegression fitted, with score : 0.8108017718715393
RidgeClassifier fitted, with score : 0.8074706533776302
Best estimator is (0.8127906976744186, SVC(C=0.8))
```

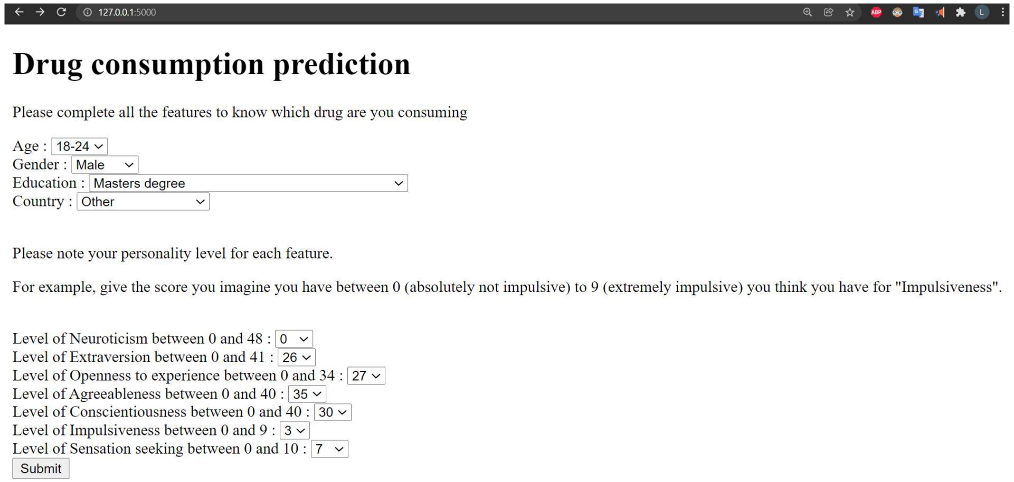




→ For each drug, we take the 6 most correlated features with the target and we try some algorithmes with different parameters, and we display the best estimator for each algorithm, the the best of them, and we display the confusion matrix and success rate between real test values and predicted test values. Then we save the estimator.

API : Flask

Remarks : In the home page, you can choose some features and give yourself a score of your personality characters. When you submit, the API predicts if you are a drug consumer for each drug with the best model we saved in the notebook.
If there is an error when you press submit button, try to run app.py with anaconda


127.0.0.1:5000

Drug consumption prediction

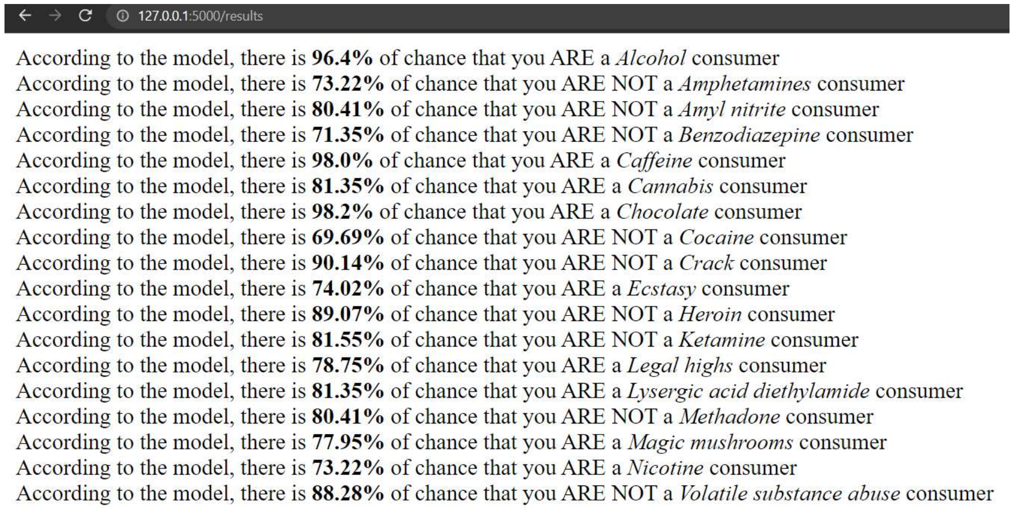
Please complete all the features to know which drug are you consuming

Age : 18-24
Gender : Male
Education : Masters degree
Country : Other

Please note your personality level for each feature.

For example, give the score you imagine you have between 0 (absolutely not impulsive) to 9 (extremely impulsive) you think you have for "Impulsiveness".

Level of Neuroticism between 0 and 48 : 0
Level of Extraversion between 0 and 41 : 26
Level of Openness to experience between 0 and 34 : 27
Level of Agreeableness between 0 and 40 : 35
Level of Conscientiousness between 0 and 40 : 30
Level of Impulsiveness between 0 and 9 : 3
Level of Sensation seeking between 0 and 10 : 7
Submit


127.0.0.1:5000/results

According to the model, there is **96.4%** of chance that you ARE a *Alcohol* consumer
According to the model, there is **73.22%** of chance that you ARE NOT a *Amphetamines* consumer
According to the model, there is **80.41%** of chance that you ARE NOT a *Amyl nitrite* consumer
According to the model, there is **71.35%** of chance that you ARE NOT a *Benzodiazepine* consumer
According to the model, there is **98.0%** of chance that you ARE a *Caffeine* consumer
According to the model, there is **81.35%** of chance that you ARE a *Cannabis* consumer
According to the model, there is **98.2%** of chance that you ARE a *Chocolate* consumer
According to the model, there is **69.69%** of chance that you ARE NOT a *Cocaine* consumer
According to the model, there is **90.14%** of chance that you ARE NOT a *Crack* consumer
According to the model, there is **74.02%** of chance that you ARE a *Ecstasy* consumer
According to the model, there is **89.07%** of chance that you ARE NOT a *Heroin* consumer
According to the model, there is **81.55%** of chance that you ARE NOT a *Ketamine* consumer
According to the model, there is **78.75%** of chance that you ARE a *Legal highs* consumer
According to the model, there is **81.35%** of chance that you ARE a *Lysergic acid diethylamide* consumer
According to the model, there is **80.41%** of chance that you ARE NOT a *Methadone* consumer
According to the model, there is **77.95%** of chance that you ARE a *Magic mushrooms* consumer
According to the model, there is **73.22%** of chance that you ARE a *Nicotine* consumer
According to the model, there is **88.28%** of chance that you ARE NOT a *Volatile substance abuse* consumer