

Python for Data Analysis

PROJECT : DRUG CONSUMPTION

BY,
Manisha Priyadarshini RAWLA





Dataset Description



The dataset from the UCI Machine Learning repository is the result of an online survey conducted between 2011 and 2012 among 1,885 respondents aged 18 and over, from English-speaking countries.

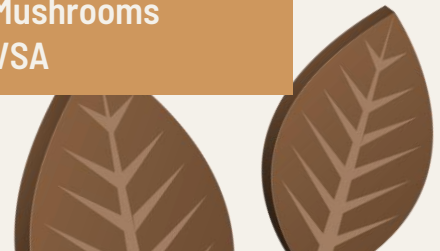
The survey collected data including Big Five personality traits (NEO-FFI-R), impulsivity (BIS-11), sensation seeking (ImpSS), and demographic information.

The data set contained information on the consumption of 18 central nervous system psychoactive drugs legal and illegal. For each of these drugs, each individual had to choose his level of consumption:

- CL0 Never Used
- CL1 Used over a Decade Ago
- CL2 Used in Last Decade
- CL3 Used in Last Year
- CL4 Used in Last Month
- CL5 Used in Last Week
- CL6 Used in Last Day



Demographic Features	Personality Traits	Legal Drugs	Illegal Drugs
<ul style="list-style-type: none">• Age• Gender• Education• Country• Ethnicity	<ul style="list-style-type: none">• Nscore• Escore• Oscore• Ascore• Impulsive• SS	<ul style="list-style-type: none">• Alcohol• Caff• Chocolate• Nicotine	<ul style="list-style-type: none">• Amphet• Amyl• Benzos• Cannabis• Crack• Ecstasy• Heroin• Ketamine• Legalh• LSD• Meth• Mushrooms• VSA



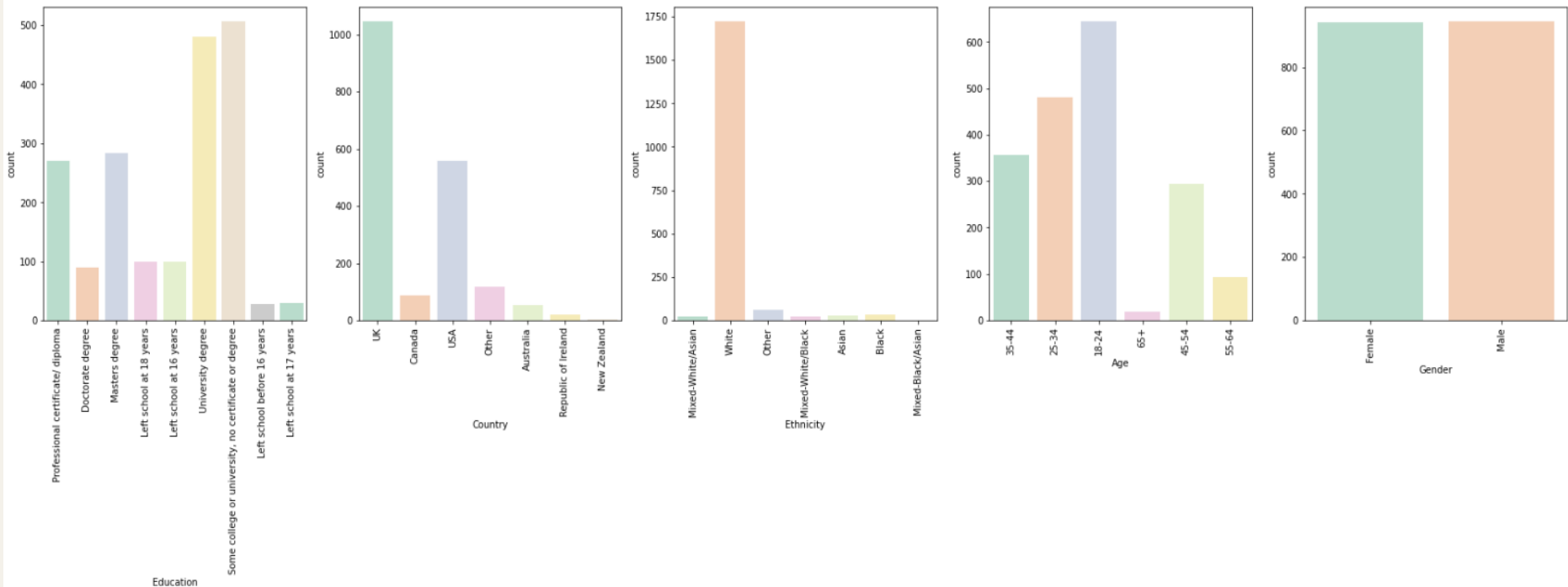


Dataset Analysis



DEMOGRAPHIC DATA

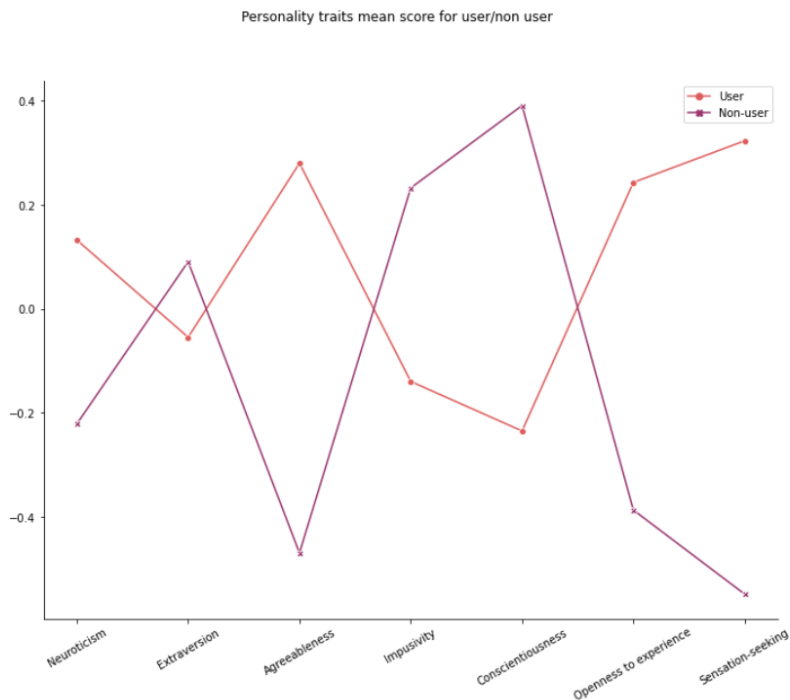
Some categories are almost uniformly disparate such as Education and Age or even equally represented which Gender.



Within the Data frame's columns, we can see certain unequally represented categories, such as Ethnicity, which has up to 90% white participants, and Country, which has predominantly UK participants.

Country	
UK	55.384615
USA	29.549072
Other	6.259947
Canada	4.615385
Australia	2.864721
Republic of Ireland	1.061008
New Zealand	0.265252

OBSERVATIONS



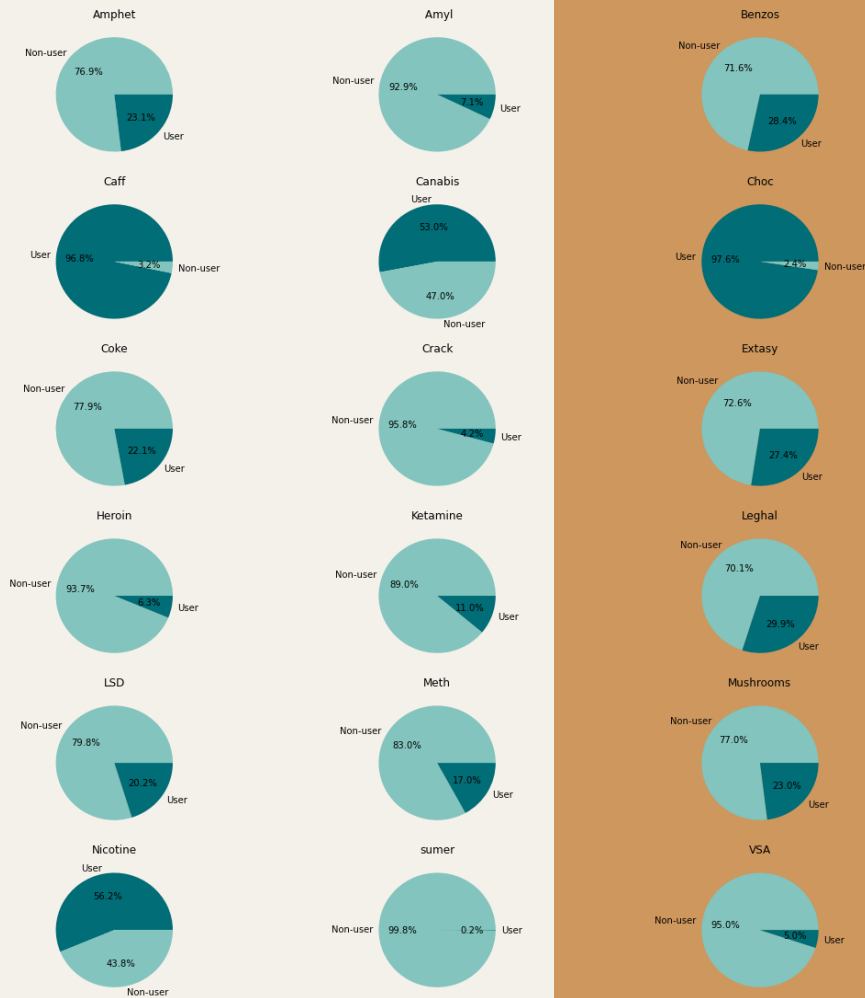
By creating a new column of user/non-user based on the annual consumption of illegal drugs, plotted the mean score for each personality traits for consumer or not consumers.

We can see a strong correlation between the consumption and the changing behavior, the personality traits are diametrically opposed.



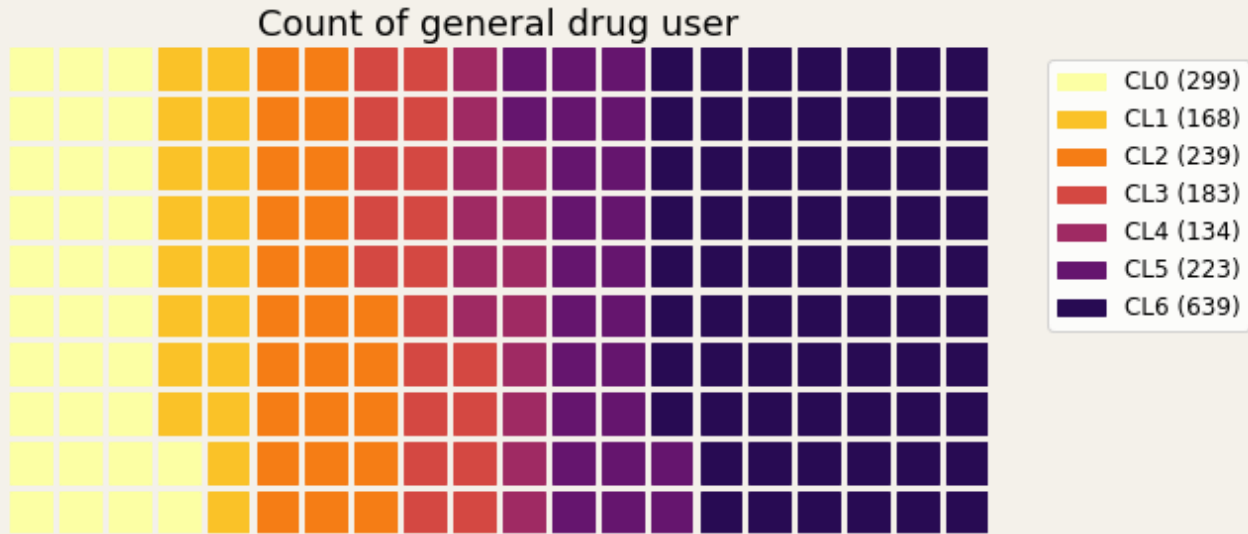
We can observe a widely disparate representation of user/non-user for illegal and legal drugs.

Usage of different drugs



Normal Use

We can see a widely disparate representation of those classes, with a majority of users up to a year (rf Machine Learning part).

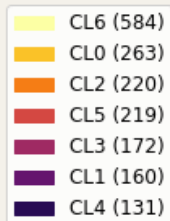
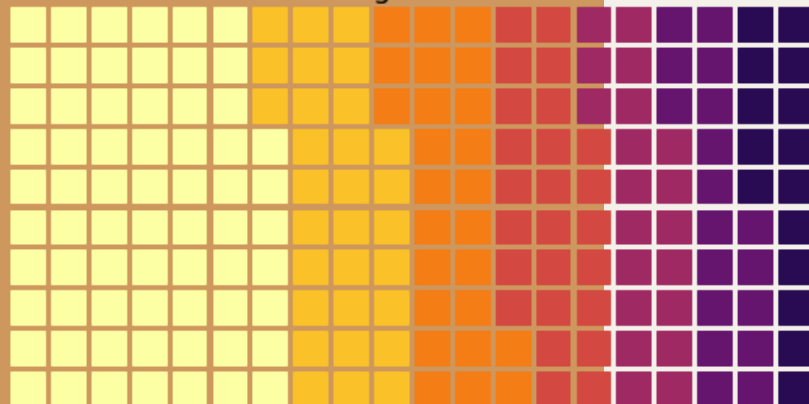




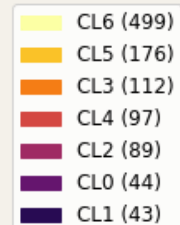
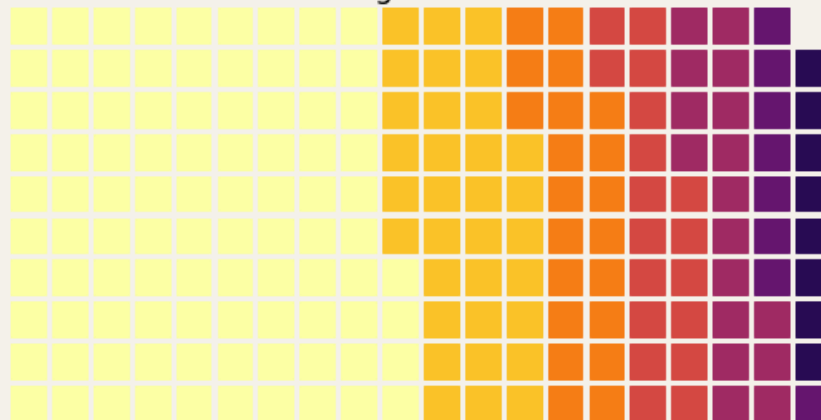
Getaway drugs

We can observe here that the Getaway drugs Nicotine & Alcohol seems to have a more highly count of illegal drug user.

Count of drug user for Alcohol

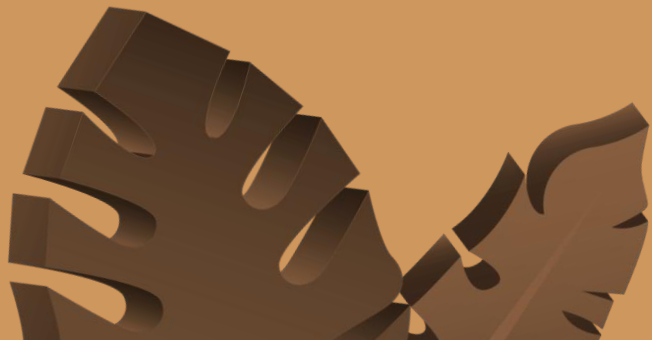


Count of drug user for Nicotine





Algorithms Used



Multiple Machine Learning Algorithms from the scikit-learn library with the « Illegal drug user/nonuser » as an output, by checking for each individual if he used an illegal drug up to a year before the survey. Here is a little overview of the algorithms tested :

- Logistic Regression
- Linear SVC
- Random Forest
- Decision Tree
- KNN
- SVM





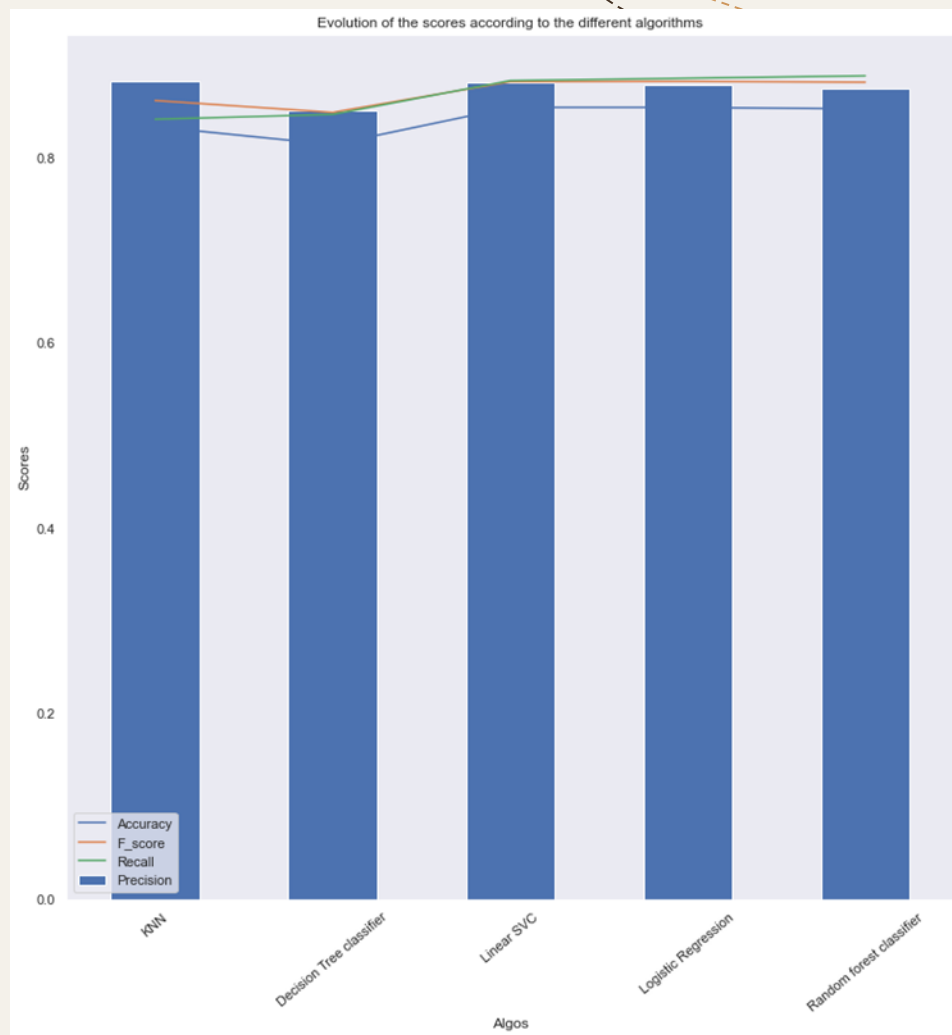
CONCLUSION





THE BEST MODEL??

The goal of these projections is to anticipate whether or not someone will take drugs. Particularly focusing on the system to enable specialists such as psychologists and others to detect current or future users. To accomplish so, we want to decrease the rate of false positives while increasing recall.



The background is split horizontally. The top half is light gray, and the bottom half is a solid brown color. On the left side, two brown leaves with lighter brown veins are partially visible. On the right side, two more brown leaves with lighter brown veins are partially visible. In the top right corner, there are two curved lines: one dashed black line and one dashed orange line.

THANK YOU!