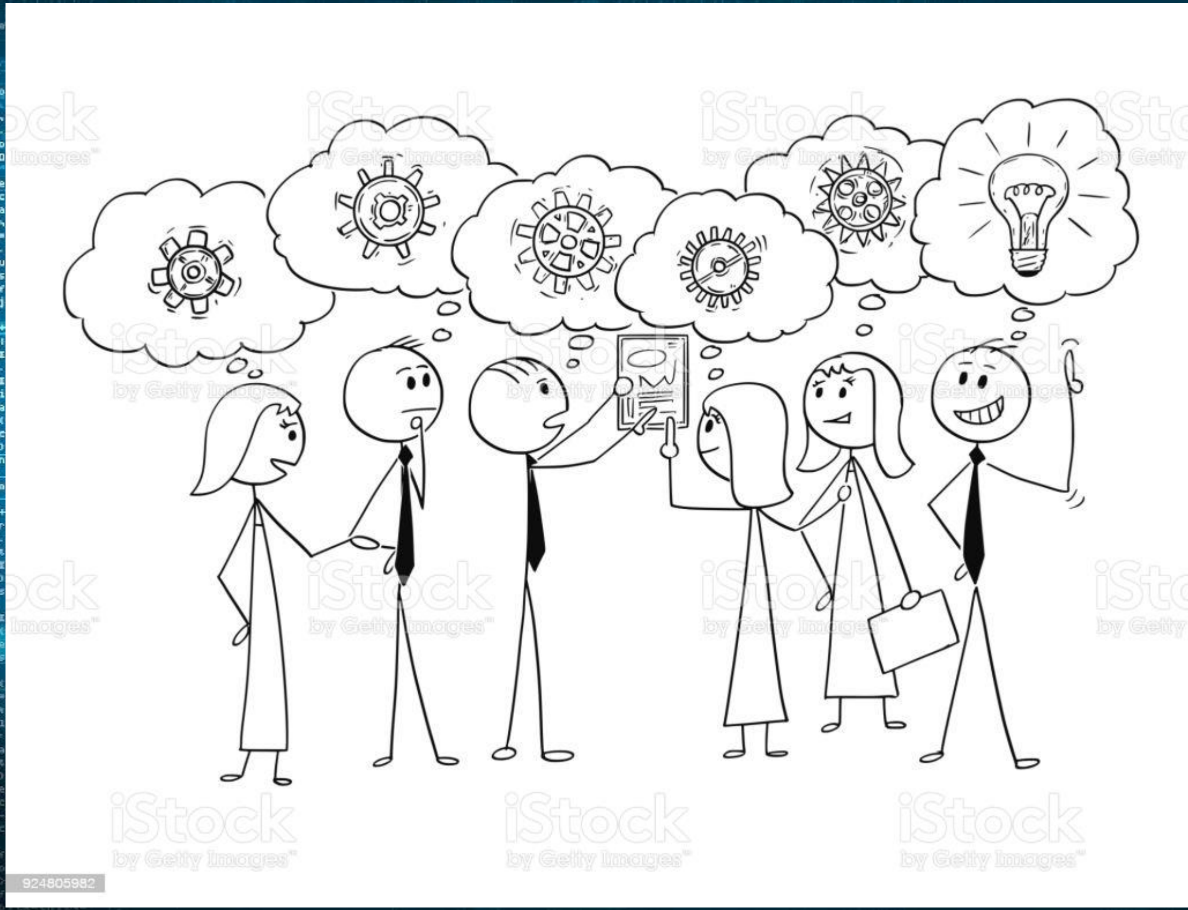
Drug Consumption Exploratory Data Analysis

By Mathis REZZOUK & Thibaud PREMAOR

Issue: Assess the risks of using nicotine, mushrooms, LSD and methamphetamine



What were the difficulties encountered ?

```
[ ] # Re-définition des colonnes et des données
    df=df.rename(columns={0:'ID',1:'Age',2:'Gender',3:'Education',4:'Country',5:'Ethnicity',6:'Nscore',7:'Escore',8:'Oscore',9

    #On regarde si il y'a des 'null' dans les données
    df.isna().sum()

    # L'ID n'a pas d'intérêt ici, l'indexation est suffisante
    df=df.drop(columns=['ID'])
```

```
#Gender remise en forme
func=lambda x : 'M' if (x>0) else 'F'
df['Gender'] = df['Gender'].apply(func)
```

```
[ ] #Education remise en forme
    def Educ(x):
      if x<-1:
        x='Left school before 18 years old'   # On a fait le choix de regrouper toutes les personnes ayant arrêté leur scolari
      elif x==-0.61113:
        x='At college or university, no degree'
      elif x==-0.05921:
        x='Professional diploma'
      elif x==0.45468:
        x='University degree'
      elif x==1.16365:
        x='Master degree'
      else:
        x='Doctor degree'
      return(x)
```

Data cleansing is something very delicate because we have to observe each column of the dataframe and the data they contain

Figuring out what to show and predict afterwards is the centerpiece of the project. So we need to compare the data and ask ourselves multiple questions about it

What were our accomplishments during this project ?

We have succeeded in creating a dataframe that is very easy to visualize in order to better understand and appropriate it

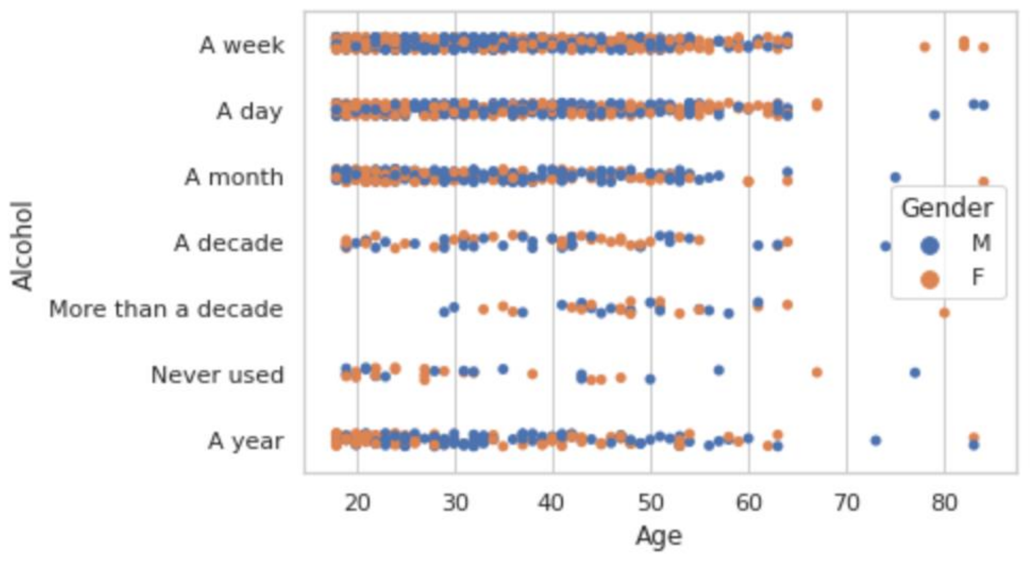
	0	1	2	3	4
0	1	0.49788	0.48246	-0.05921	0.96082
1	2	-0.07854	-0.48246	1.98437	0.96082
2	3	0.49788	-0.48246	-0.05921	0.96082
3	4	-0.95197	0.48246	1.16365	0.96082
4	5	0.49788	0.48246	1.98437	0.96082
...	...	...	...	...	...
1880	1884	-0.95197	0.48246	-0.61113	-0.57009
1881	1885	-0.95197	-0.48246	-0.61113	-0.57009
1882	1886	-0.07854	0.48246	0.45468	-0.57009
1883	1887	-0.95197	0.48246	-0.61113	-0.57009
1884	1888	-0.95197	-0.48246	-0.61113	0.21128

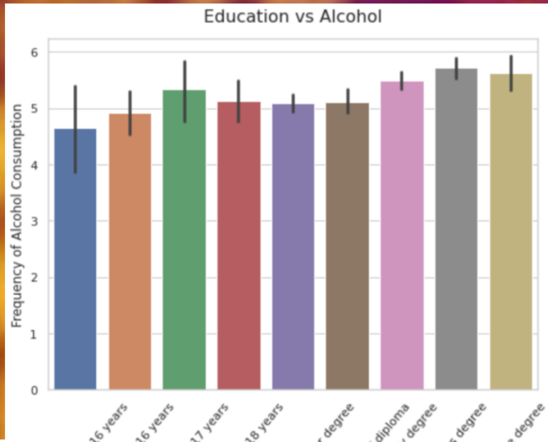
Dataframe before cleaning

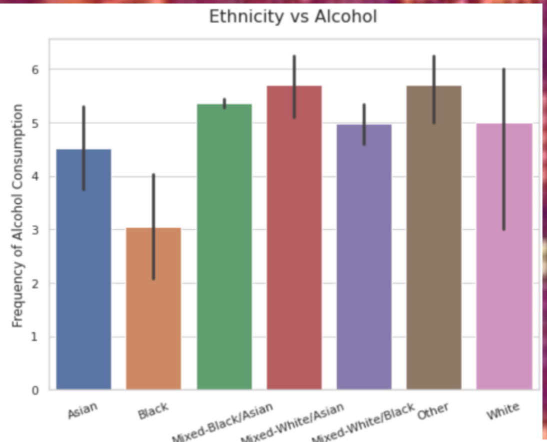
	Age	Gender	Education	Country
0	37	M	Professional diploma	UK
1	28	F	Doctor degree	UK
2	42	F	Professional diploma	UK
3	19	M	Master degree	UK
4	44	M	Doctor degree	UK

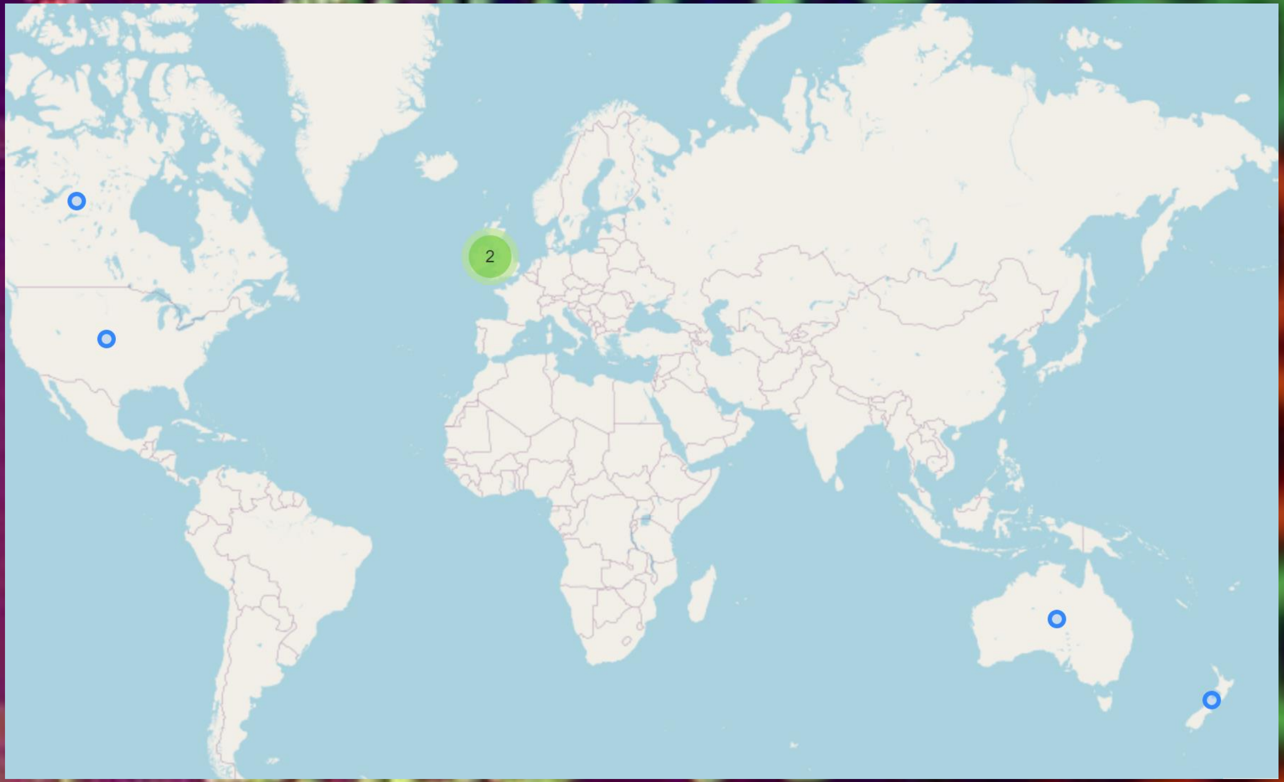
Dataframe after cleaning

We have done some very convincing visualizations that have allowed us to see more clearly into the subject









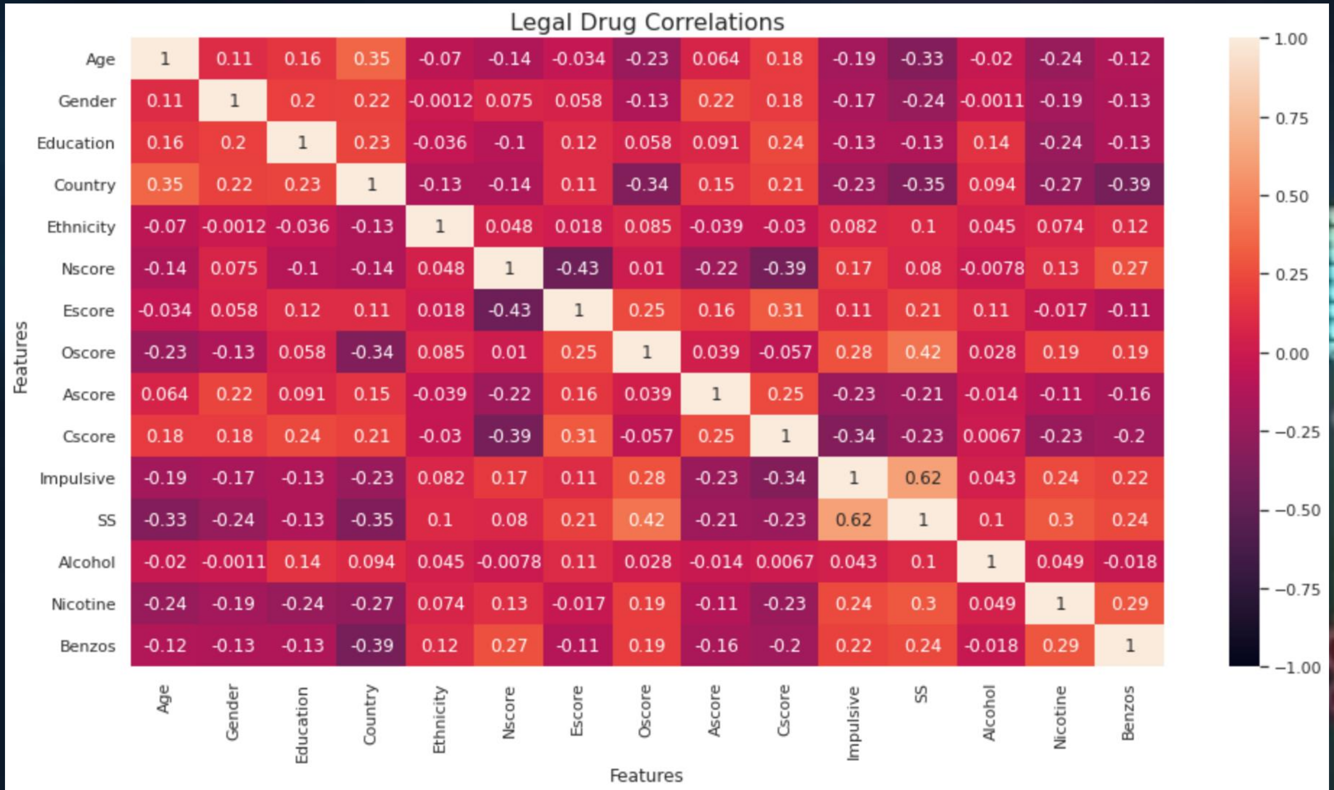
We were able to make predictions of different types in order to compare the prediction models

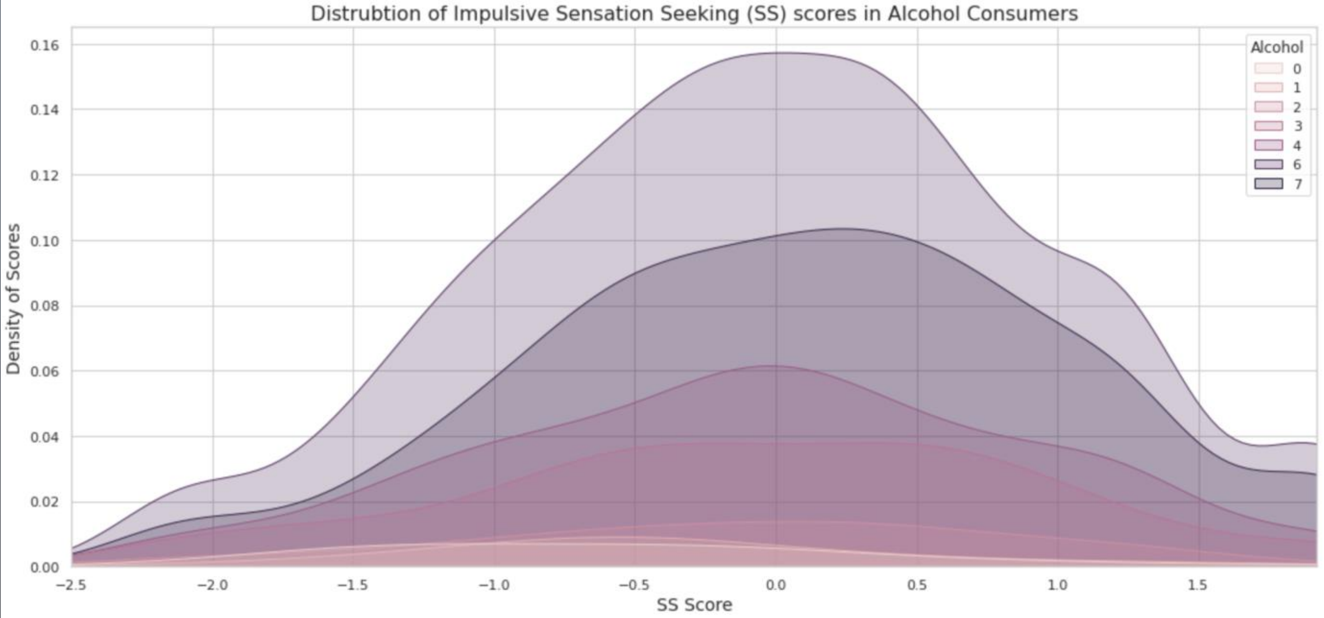
```
     Logisitc Regression trained.
        Ridge Classifier trained.
 Support Vector Machines trained.
Random Forest Classifier trained.
```

We will talk about it soon…

What were our thoughts on the subject ?

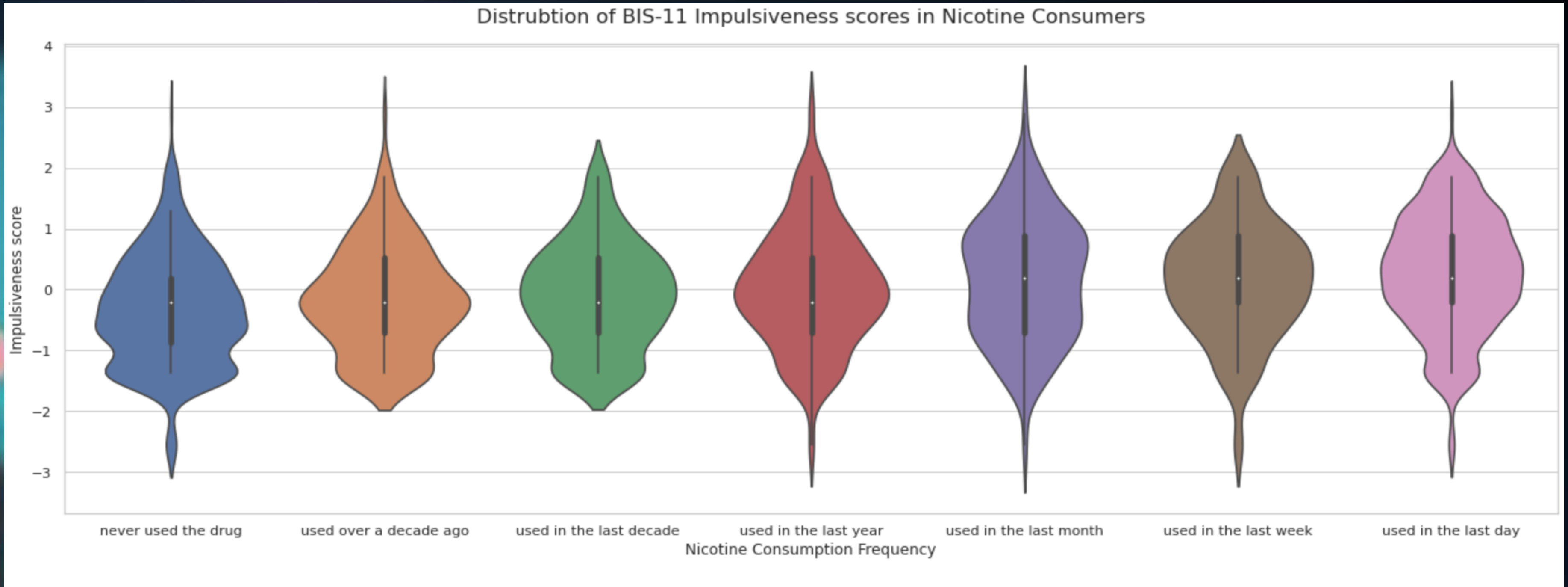
We decided to separate illegal drugs from legal drugs to establish visualizations by type. Let's start with the legal drugs...

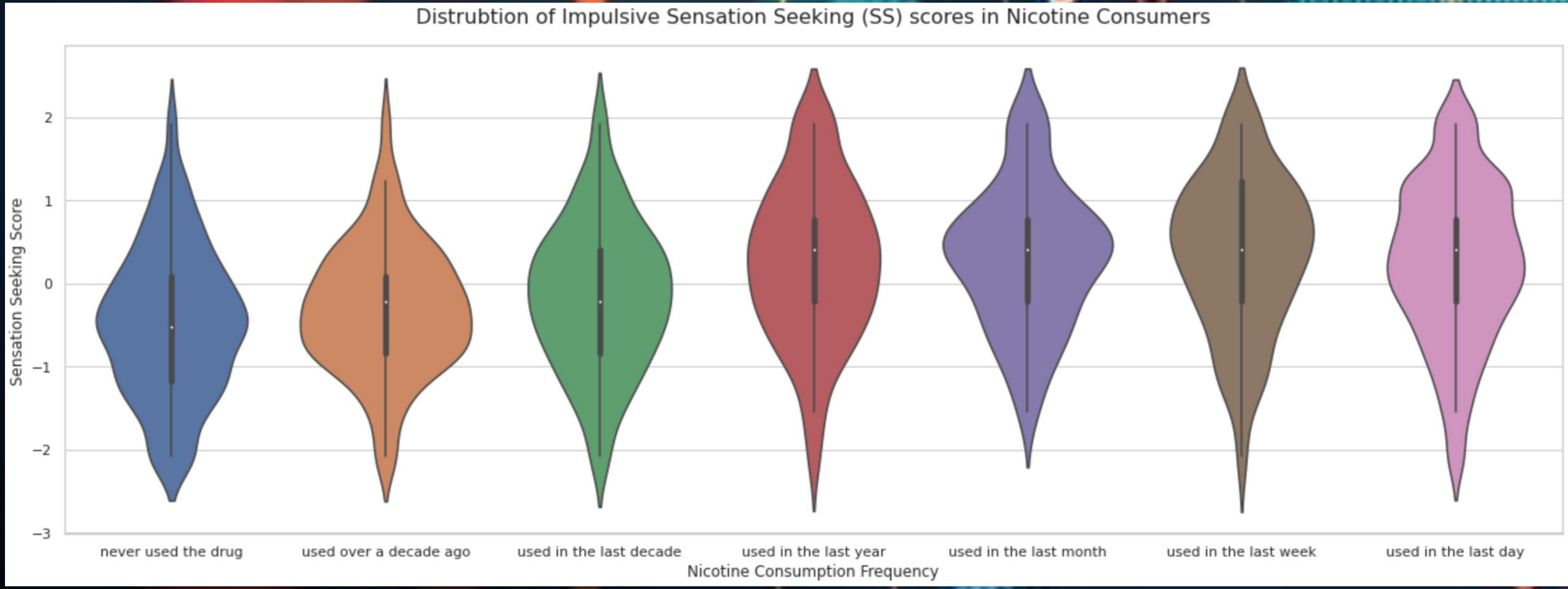




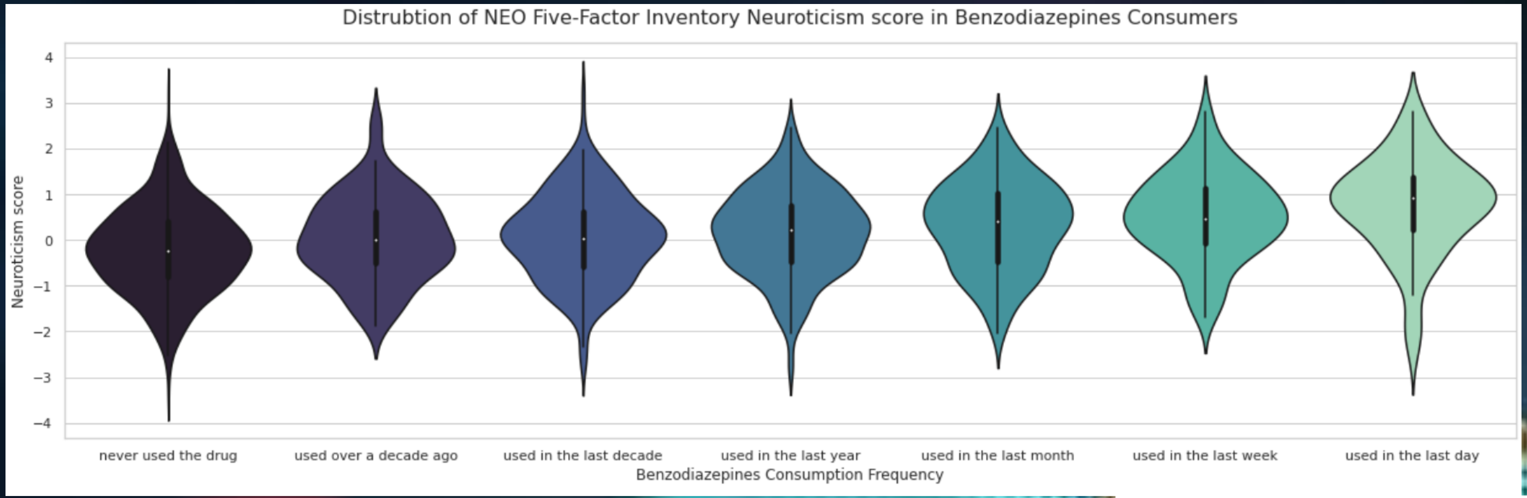
Alcohol use showed a slight positive correlation with education level, such that higher education was associated with more frequent alcohol use. Specifically, those with a university degree or higher (i.e., master's or doctorate) drank the most alcohol. In addition, frequent drinkers tended to have higher impulsive sensation seeking (SS) scores, such that a majority of non-drinkers (never drank or drank more than a decade ago) had an SS score close to -1.0 while frequent drinkers (drank one week or one day ago) tended to have a score between 0 and 0.25.

Unlike alcohol, nicotine showed a marked difference in consumption between the sexes, with men consuming more nicotine than women.



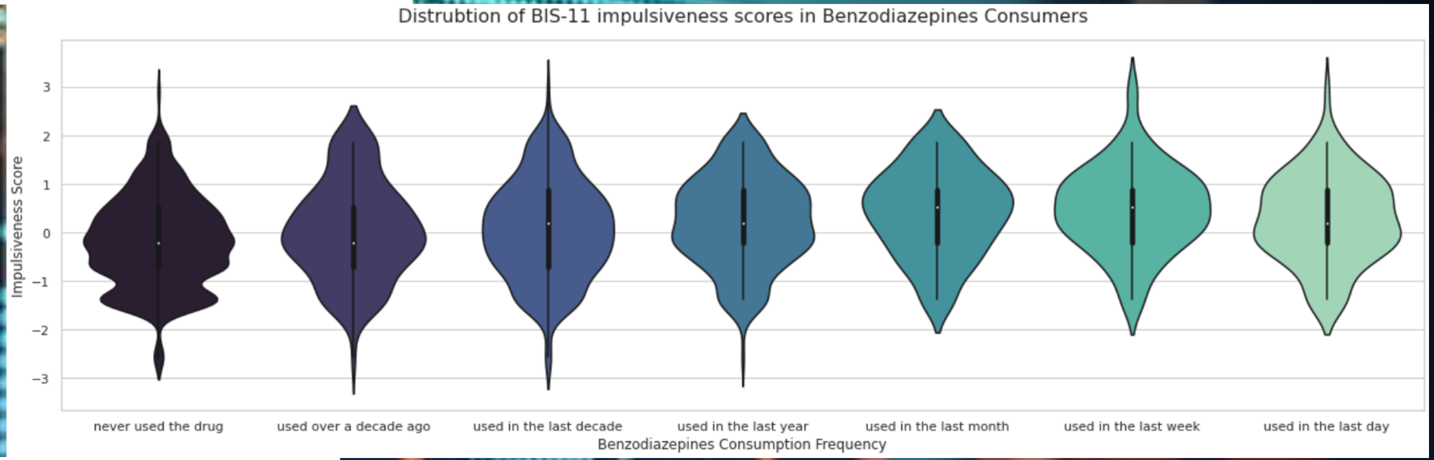


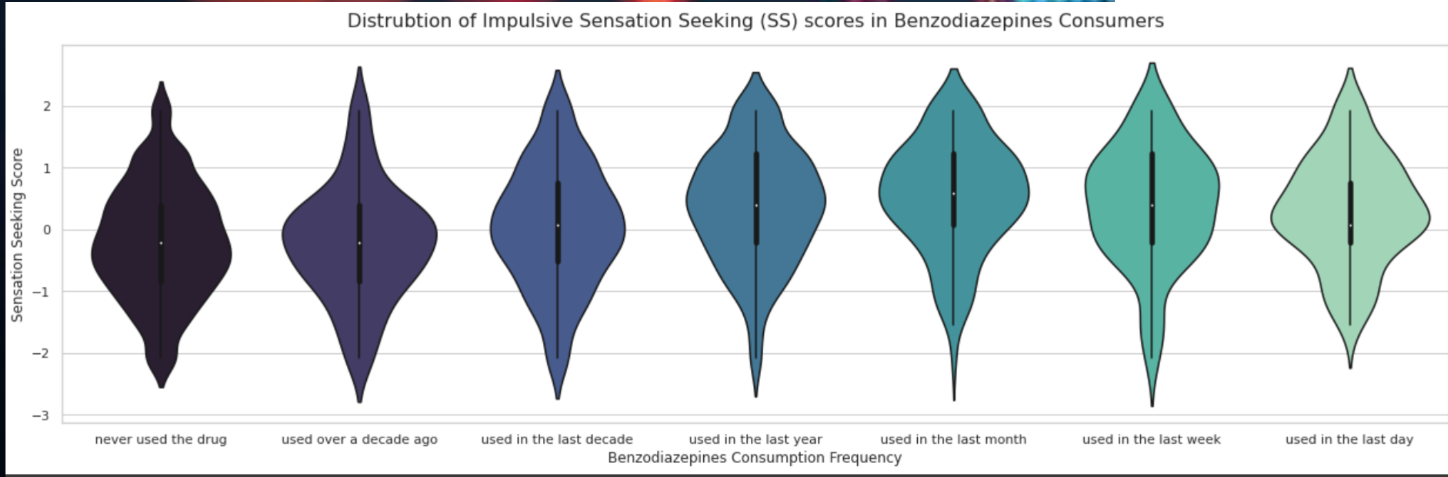
Furthermore, nicotine use was positively correlated with both the BIS-11 impulsivity score and SS. However, this relationship was slightly more pronounced in SS.



We examined the relationship between Benzodiazepine and various personality measures.

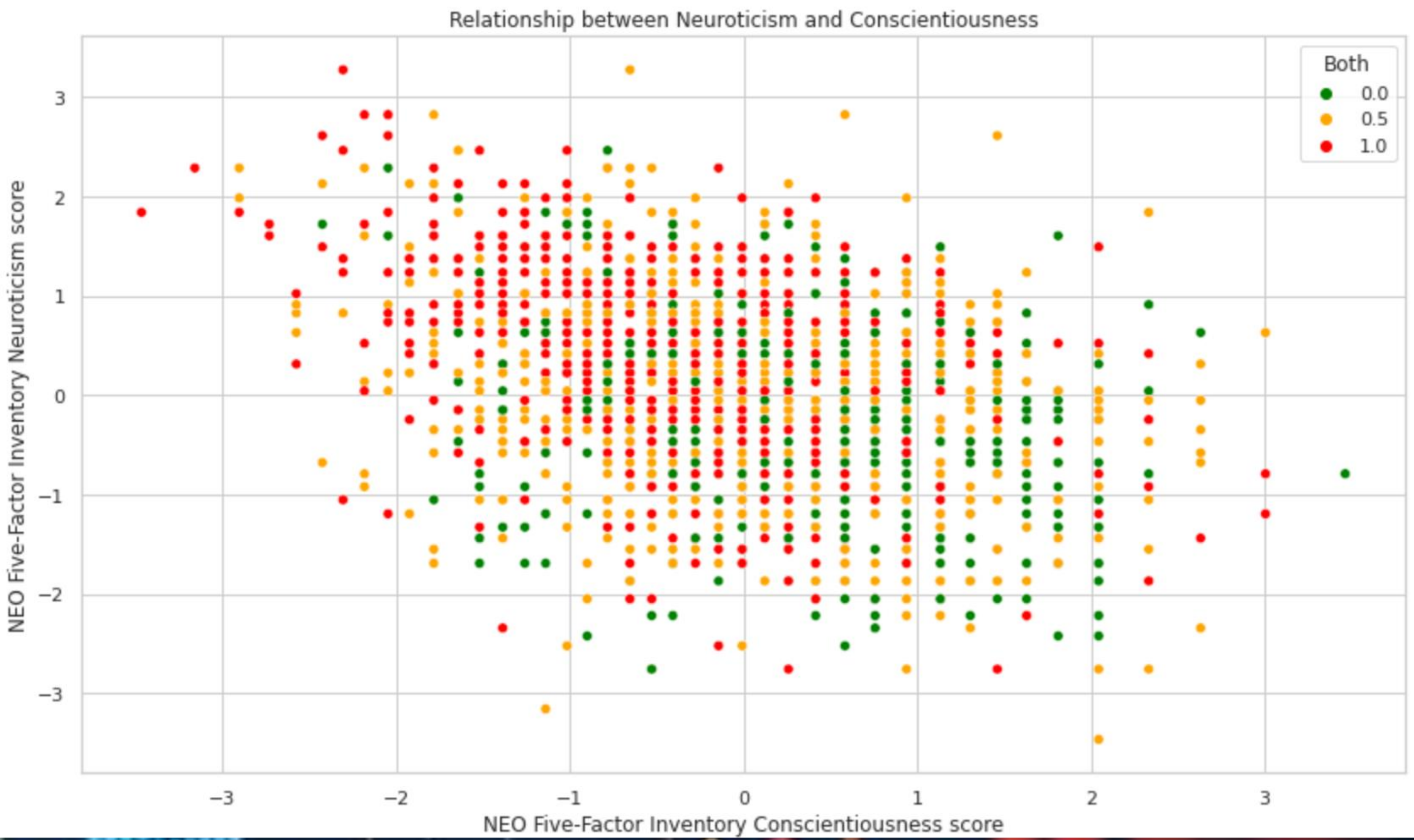
Benzodiazepine use frequency showed a positive correlation with neuroticism (Nscore), impulsivity, and SS scores. This correlation was strongest in the Nscores.



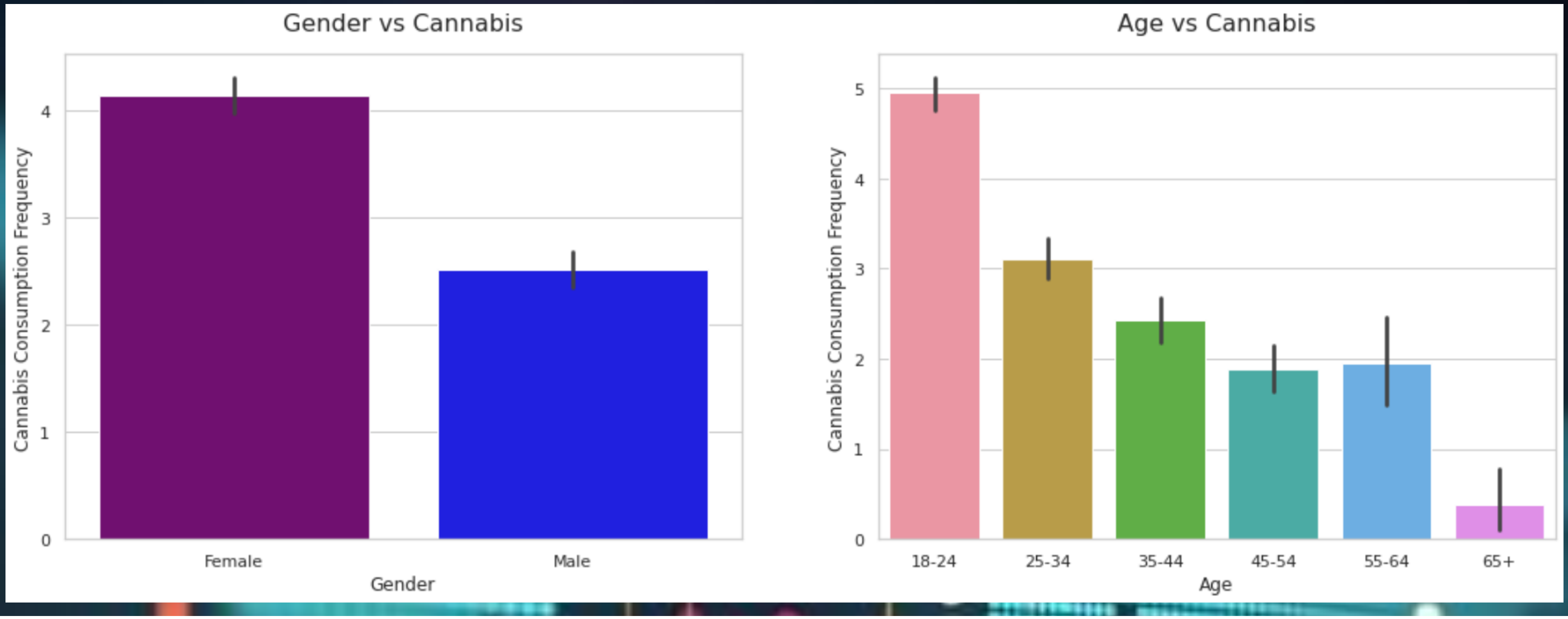


These results align with other research that has suggested a relationship between benzodiazepine sensitivity and neuroticism with those with higher neuroticism showing higher sensitivity to Benzos

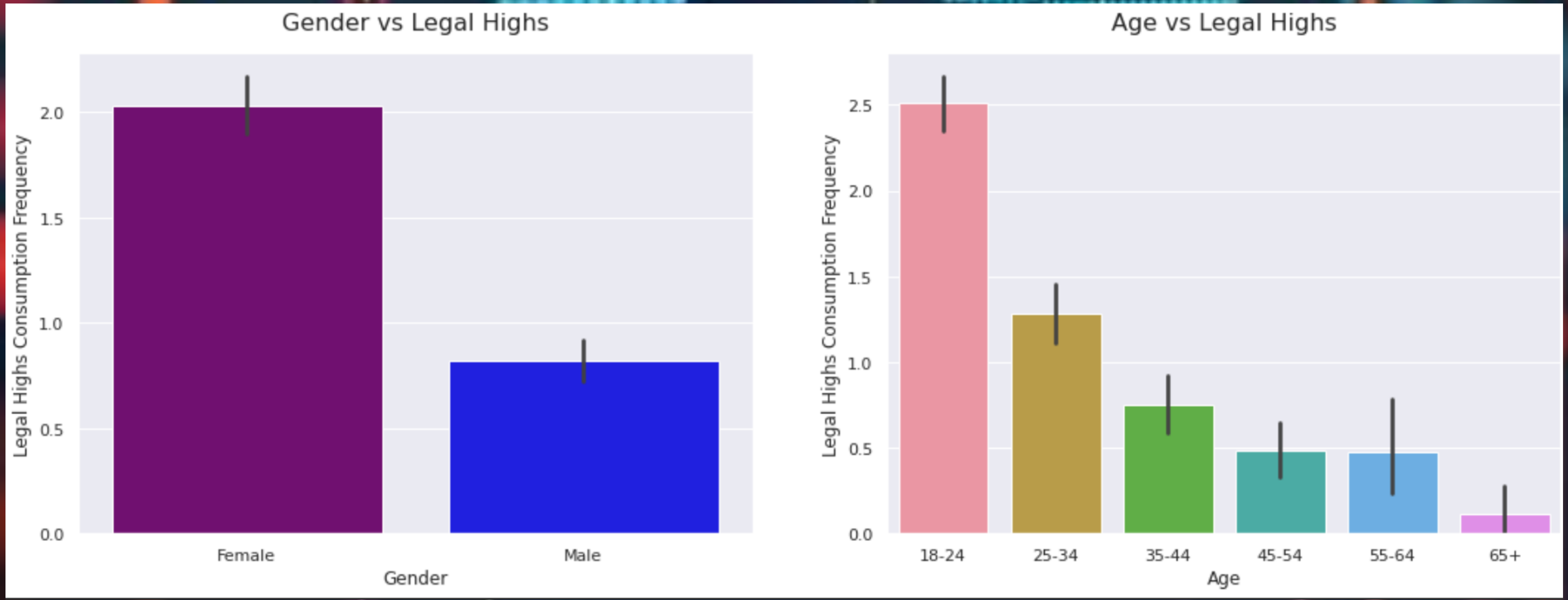
Interestingly, there was a strong negative correlation between Nscore and Conscientiousness (Cscore). To better understand how this relationship was related to drug use, we examined nicotine and benzo use. Not surprisingly, those who used both nicotine and benzo tended to have higher Nscores and lower Cscores, while those who used no drugs had high Cscores and low Nscores.

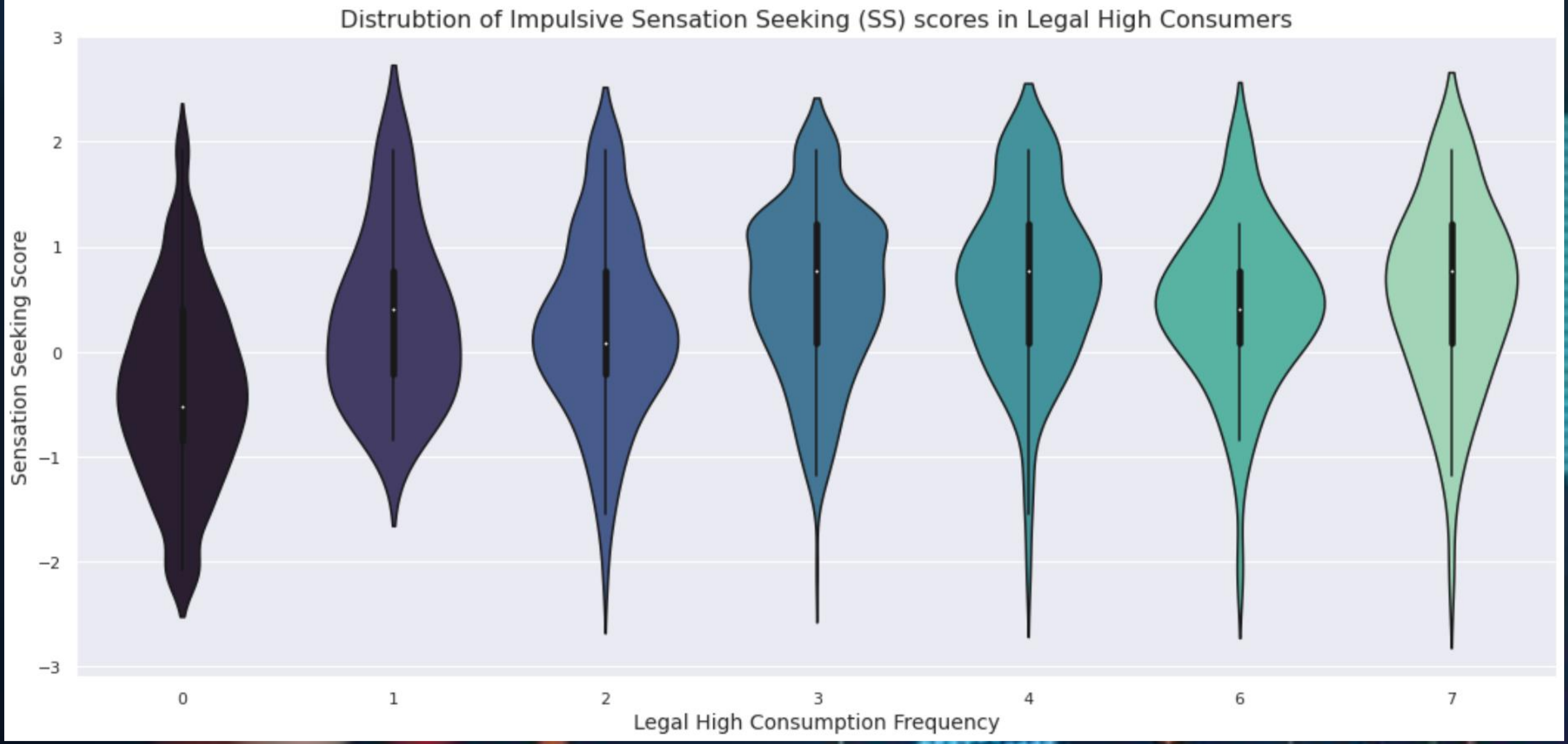


Now let's look at illegal drugs..

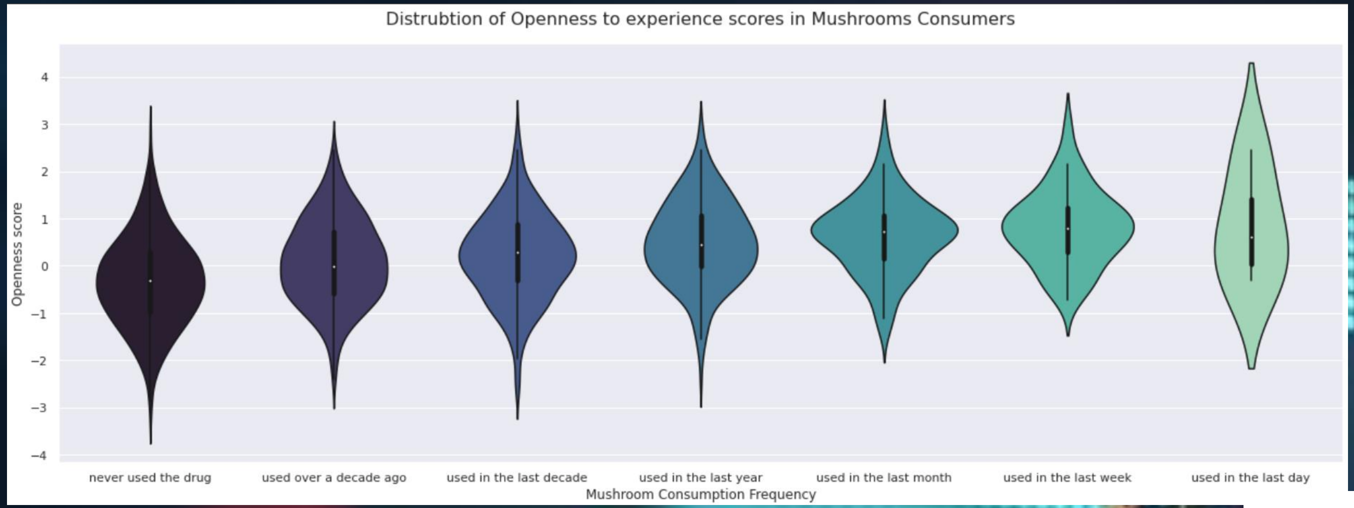


In both cannabis and legal highs, men used the drug more frequently than women. In addition, the frequency of use of cannabis, legalh and ecstasy was negatively correlated with age, so that younger individuals used them more frequently than older individuals.

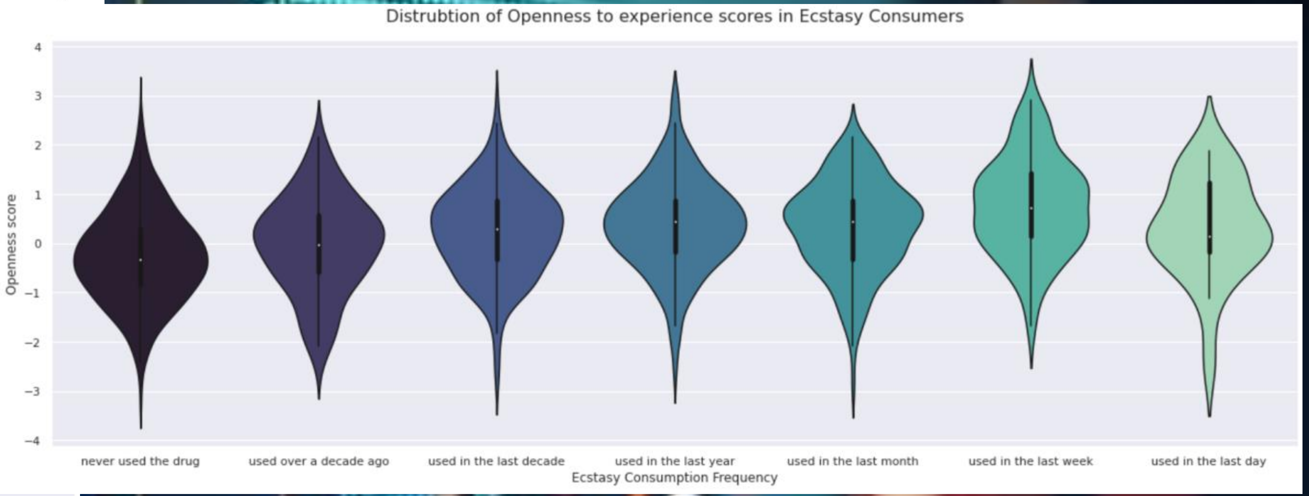


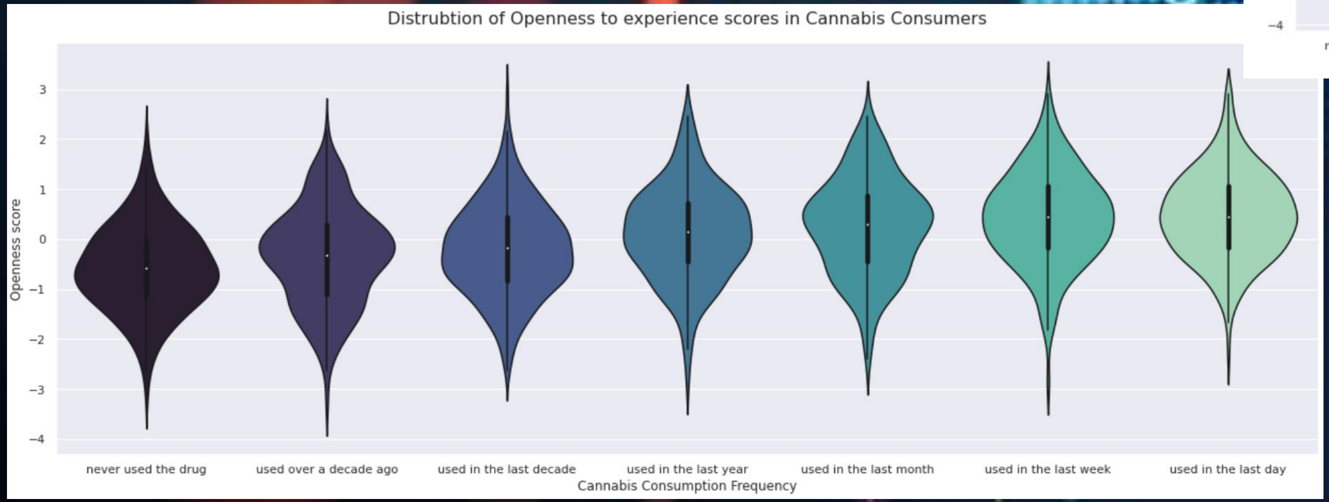


All illegal drugs were positively correlated with SS. Individuals who never or rarely used any of the illegal drugs showed SS scores of -0.5 to -1.0, whereas frequent users' scores ranged from 0.5 to 1.5. This relationship was most pronounced among legalh.



Cannabis, ecstasy, and mushrooms also showed a positive correlation with openness to experience (Oscore), with most occasional and infrequent users' scores ranging from 0 to -0.5 while the most frequent users had scores between 0.5 and 1.0





Have we created variables in our project ?

In order to make our predictions, we need to determine whether or not each case in the dataframe is addictive or not in order to make a prediction afterwards

In the case of nicotine, for example, we have created a last column that presents by 1 or 0 the addiction or not of the person. This column is called Nicotine_User

Nicotine_User
1
1
0
1
1
...
0
1
1
1
1

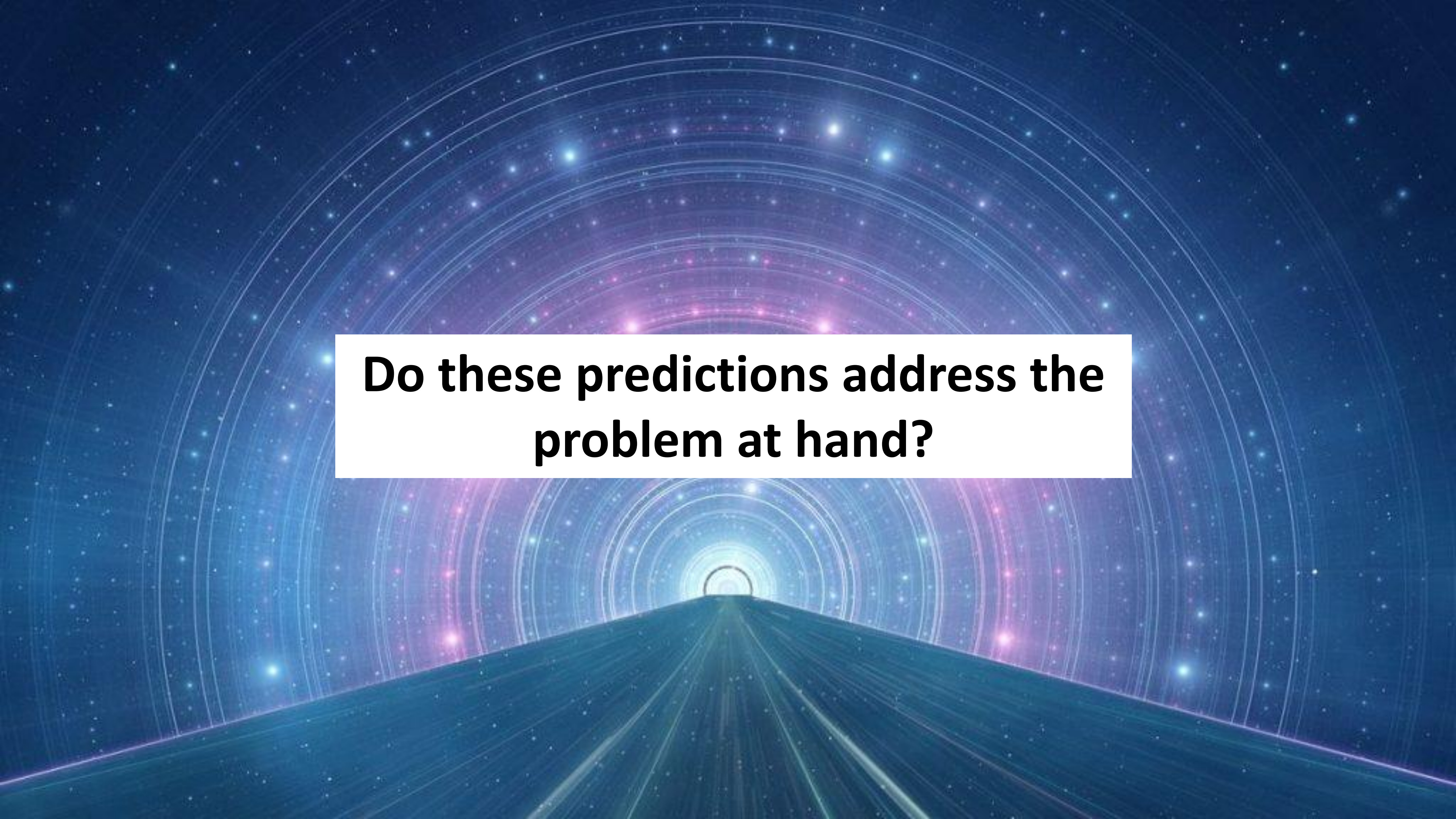
We therefore create these four variables in the form of columns for each drug whose addicts we want to predict, i.e. for LSD, mushrooms, Nicotine and methamphetamine.

Meth_User
0
1
0
0
0
...
0
1
0
0
0

Nicotine_User
1
1
0
1
1
...
0
1
1
1
1

Mushrooms_User
0
0
0
0
1
...
0
1
1
1
1

LSD_User
0
1
0
0
0
...
1
1
1
1
1

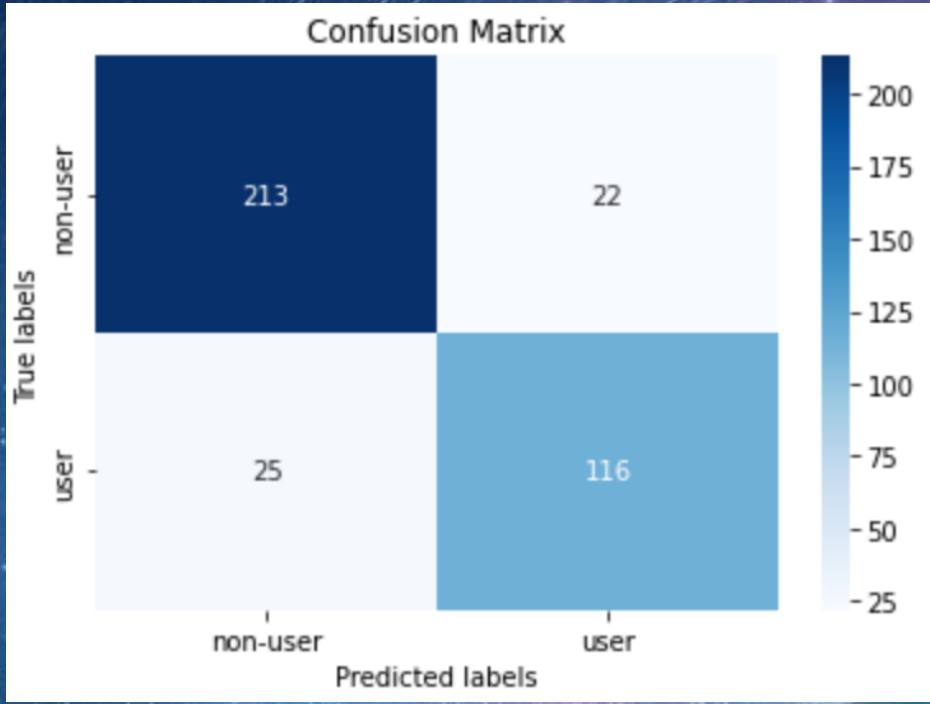

Do these predictions address the problem at hand?

Accuracy of the prediction algorithms for LSD

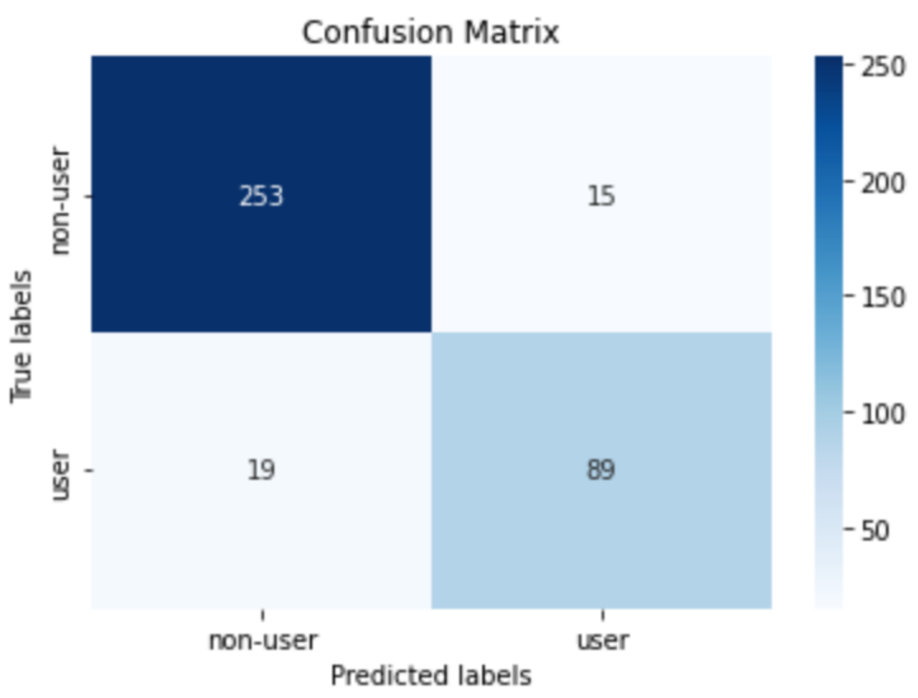
```
                  ACCURACY
     Logisitc Regression Accuracy: 89.63%
        Ridge Classifier Accuracy: 90.96%
 Support Vector Machines Accuracy: 90.43%
Random Forest Classifier Accuracy: 88.83%
---------------------------------------------
                  F1 SCORES
     Logisitc Regression F1-Score: 0.8186
        Ridge Classifier F1-Score: 0.83962
 Support Vector Machines F1-Score: 0.84071
Random Forest Classifier F1-Score: 0.81416
```

Yes, because our predictions effectively evaluate the addictive case of each patient to a type of drug as we can see on this table that shows the accuracy of four predictors: Logistic regression, Random Forest Classifier, Ridge Classifier and the Support Vector Machines. To classify the prediction algorithms we will take as a parameter the accuracy.

Random Forest confusion matrix for mushrooms

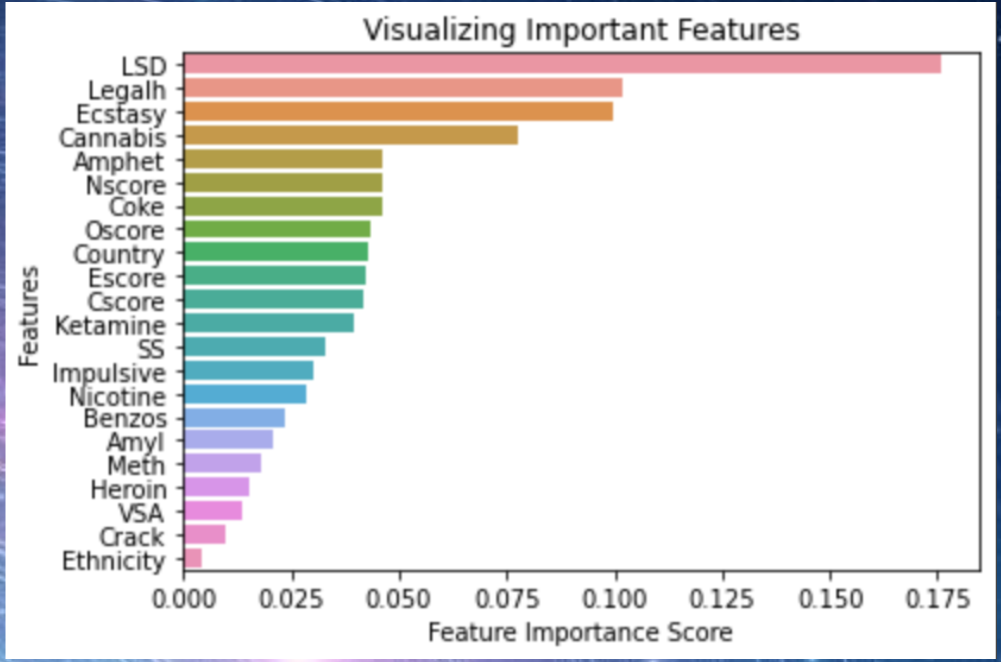


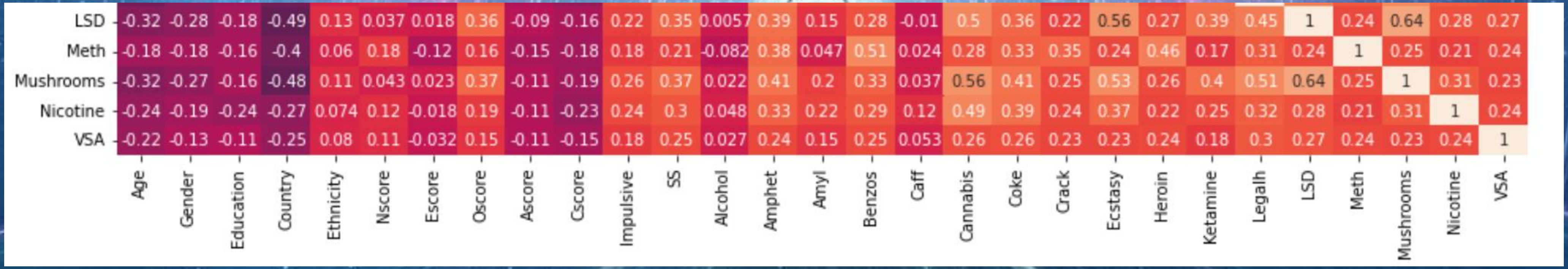
Ridge Classifier confusion matrix for LSD



In the end, we see that the logistic regression method and the Random Forest Classifiers are the best performing prediction methods.

Some visualizations of these predictions show information that is consistent with our initial dataframe visualizations. Indeed, we notice in the graph on the right that Random Forest relies the most on the LSD parameter to make these predictions of mushroom addicts which is consistent with the correlation matrix of the dataframe where we notice that LSD and mushrooms are highly correlated.





	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS	Alcohol	Amphet	Amyl
LSD	-0.32	-0.28	-0.18	-0.49	0.13	0.037	0.018	0.36	-0.09	-0.16	0.22	0.35	0.0057	0.39	0.15
Meth	-0.18	-0.18	-0.16	-0.4	0.06	0.18	-0.12	0.16	-0.15	-0.18	0.18	0.21	-0.082	0.38	0.047
Mushrooms	-0.32	-0.27	-0.16	-0.48	0.11	0.043	0.023	0.37	-0.11	-0.19	0.26	0.37	0.022	0.41	0.2
Nicotine	-0.24	-0.19	-0.24	-0.27	0.074	0.12	-0.018	0.19	-0.11	-0.23	0.24	0.3	0.048	0.33	0.22
VSA	-0.22	-0.13	-0.11	-0.25	0.08	0.11	-0.032	0.15	-0.11	-0.15	0.18	0.25	0.027	0.24	0.15

	Benzos	Caff	Cannabis	Coke	Crack	Ecstasy	Heroin	Ketamine	Legalh	LSD	Meth	Mushrooms	Nicotine	VSA
LSD	0.28	-0.01	0.5	0.36	0.22	0.56	0.27	0.39	0.45	1	0.24	0.64	0.28	0.27
Meth	0.51	0.024	0.28	0.33	0.35	0.24	0.46	0.17	0.31	0.24	1	0.25	0.21	0.24
Mushrooms	0.33	0.037	0.56	0.41	0.25	0.53	0.26	0.4	0.51	0.64	0.25	1	0.31	0.23
Nicotine	0.29	0.12	0.49	0.39	0.24	0.37	0.22	0.25	0.32	0.28	0.21	0.31	1	0.24
VSA	0.25	0.053	0.26	0.26	0.23	0.23	0.24	0.18	0.3	0.27	0.24	0.23	0.24	1