Python for Data Analysis final Project

Victoria Garnil

Adrien Forestier

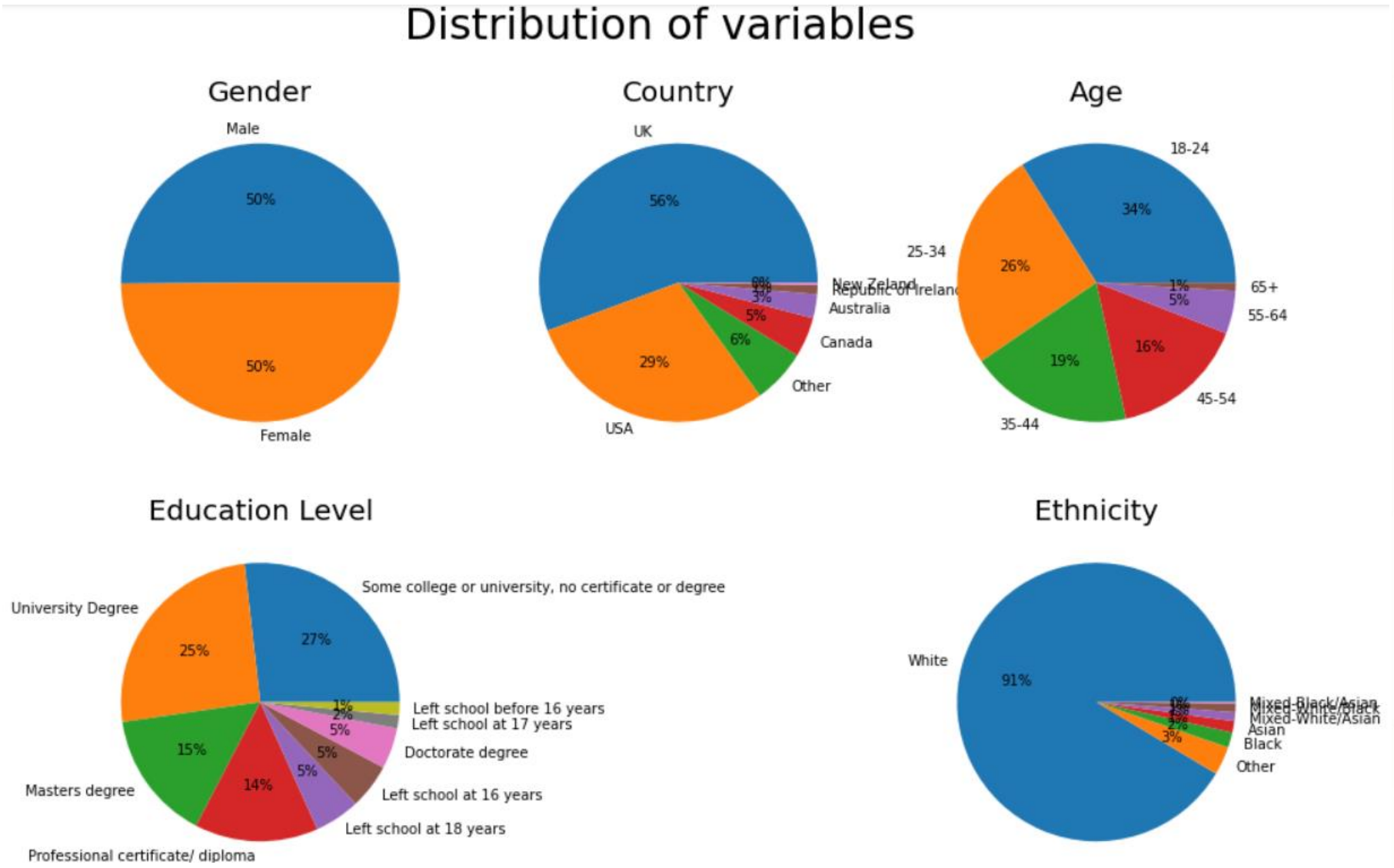
Maxence Ferreira

The Dataset

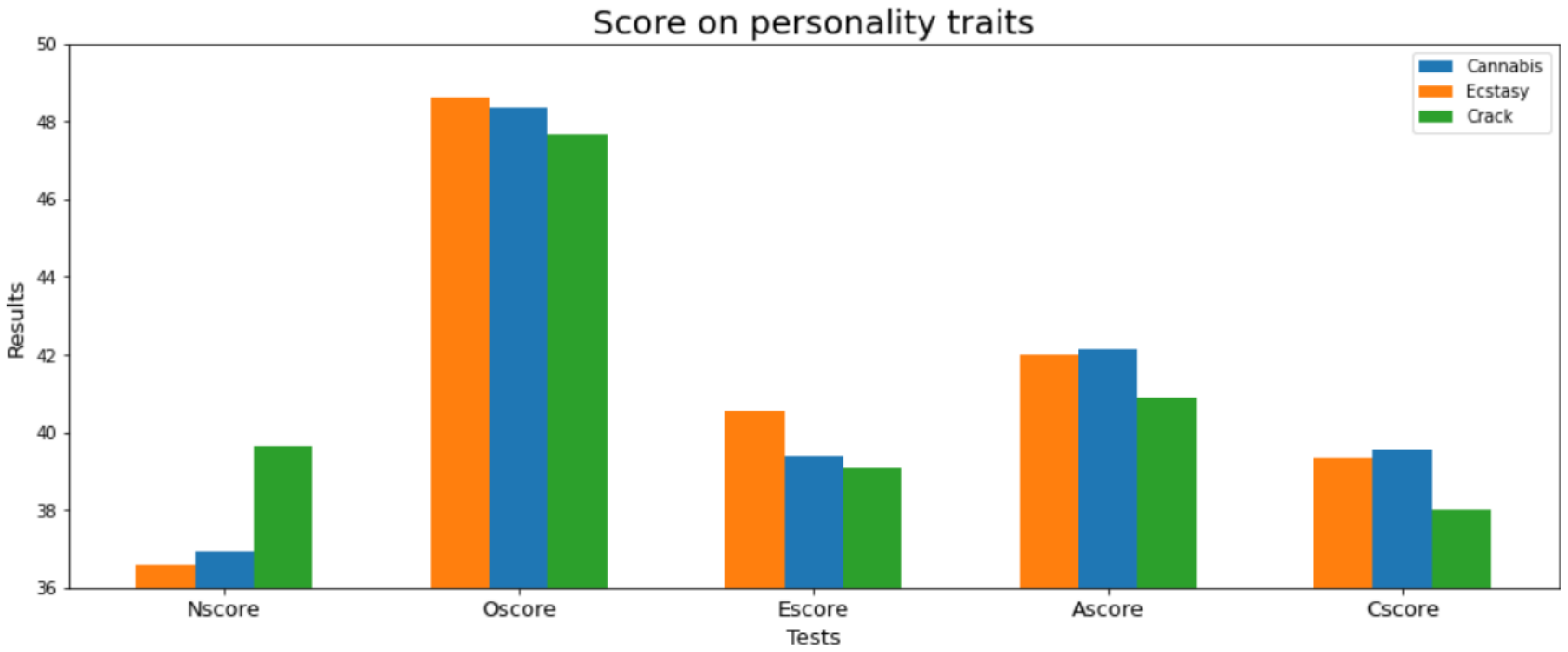
	Age	Gender	Education Level	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsiveness	Sensation Seing
0	-0.07854	-0.48246	1.98437	0.96082	-0.31685	-0.67825	1.93886	1.43533	0.76096	-0.14277	-0.71126	-0.21575
1	0.49788	-0.48246	-0.05921	0.96082	-0.31685	-0.46725	0.80523	-0.84732	-1.62090	-1.01450	-1.37983	0.40148
2	-0.95197	0.48246	1.16365	0.96082	-0.31685	-0.14882	-0.80615	-0.01928	0.59042	0.58489	-1.37983	-1.18084
3	0.49788	0.48246	1.98437	0.96082	-0.31685	0.73545	-1.63340	-0.45174	-0.30172	1.30612	-0.21712	-0.21575
4	2.59171	0.48246	-1.22751	0.24923	-0.31685	-0.67825	-0.30033	-1.55521	2.03972	1.63088	-1.37983	-1.54858

Alcohol	Amphetamines	Amyl nitrite	Benzodiazepine	Caffeine	Cannabis	Chocolate	Cocaine	Crack	Ecstasy	Heroin	Ketamine	Legal highs	LSD	Methadone	Magic mushrooms	Nicotine	Volatile substance abuse
5	2	2	0	6	4	6	3	0	4	0	2	0	2	3	0	4	0
6	0	0	0	6	3	4	0	0	0	0	0	0	0	0	1	0	0
4	0	0	3	5	2	4	2	0	0	0	2	0	0	0	0	2	0
4	1	1	0	6	3	6	0	0	1	0	0	1	0	0	2	2	0
2	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	6	0

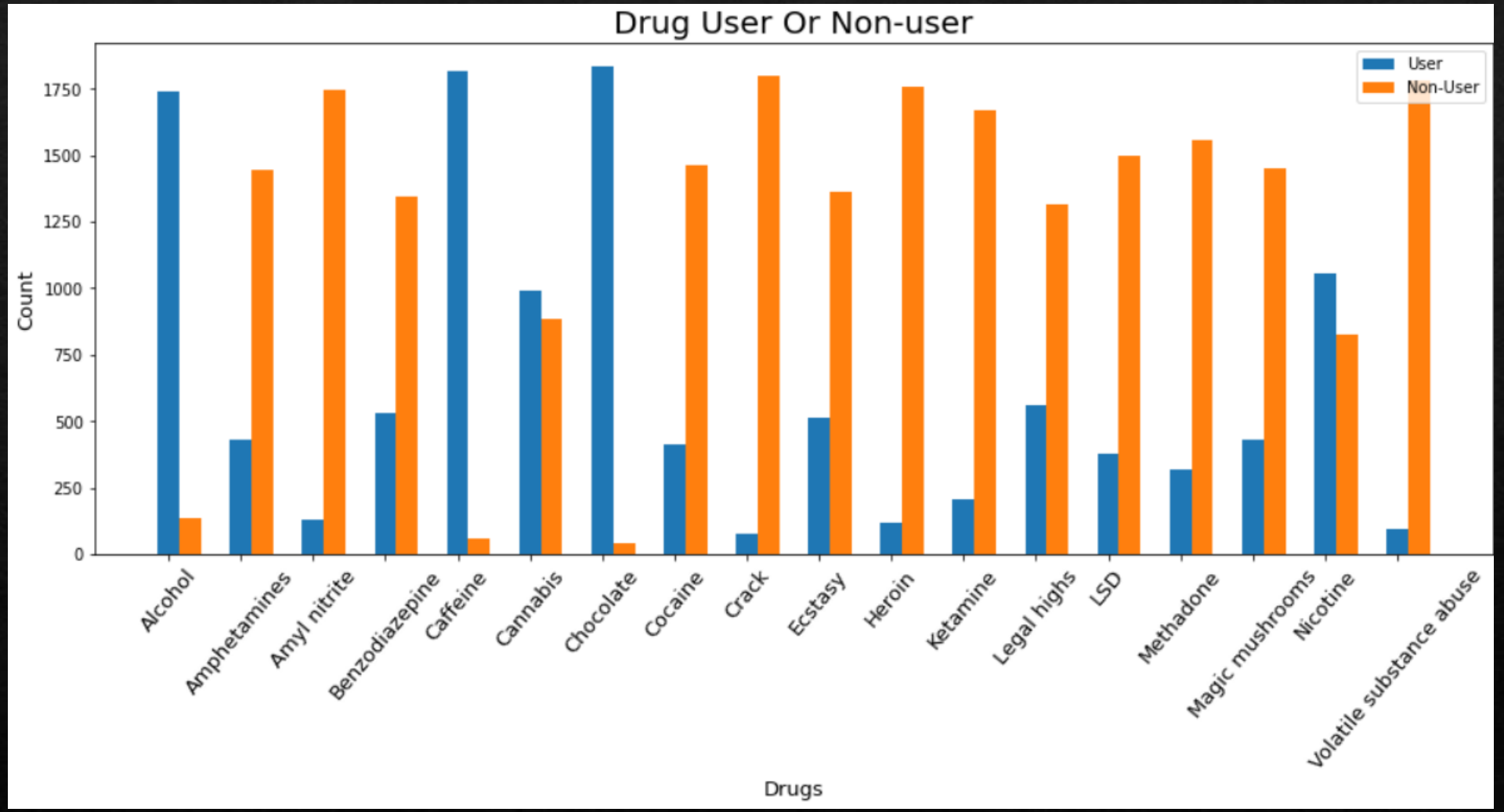
Visualizations



Visualizations



Our focus:


Drug User Or Non-user
User
Non-User
Count
0
250
500
750
1000
1250
1500
1750
Alcohol
Amphetamines
Amyl nitrite
Benzodiazepine
Caffeine
Cannabis
Chocolate
Cocaine
Crack
Ecstasy
Heroin
Ketamine
Legal highs
LSD
Methadone
Magic mushrooms
Nicotine
Volatile substance abuse
Drugs

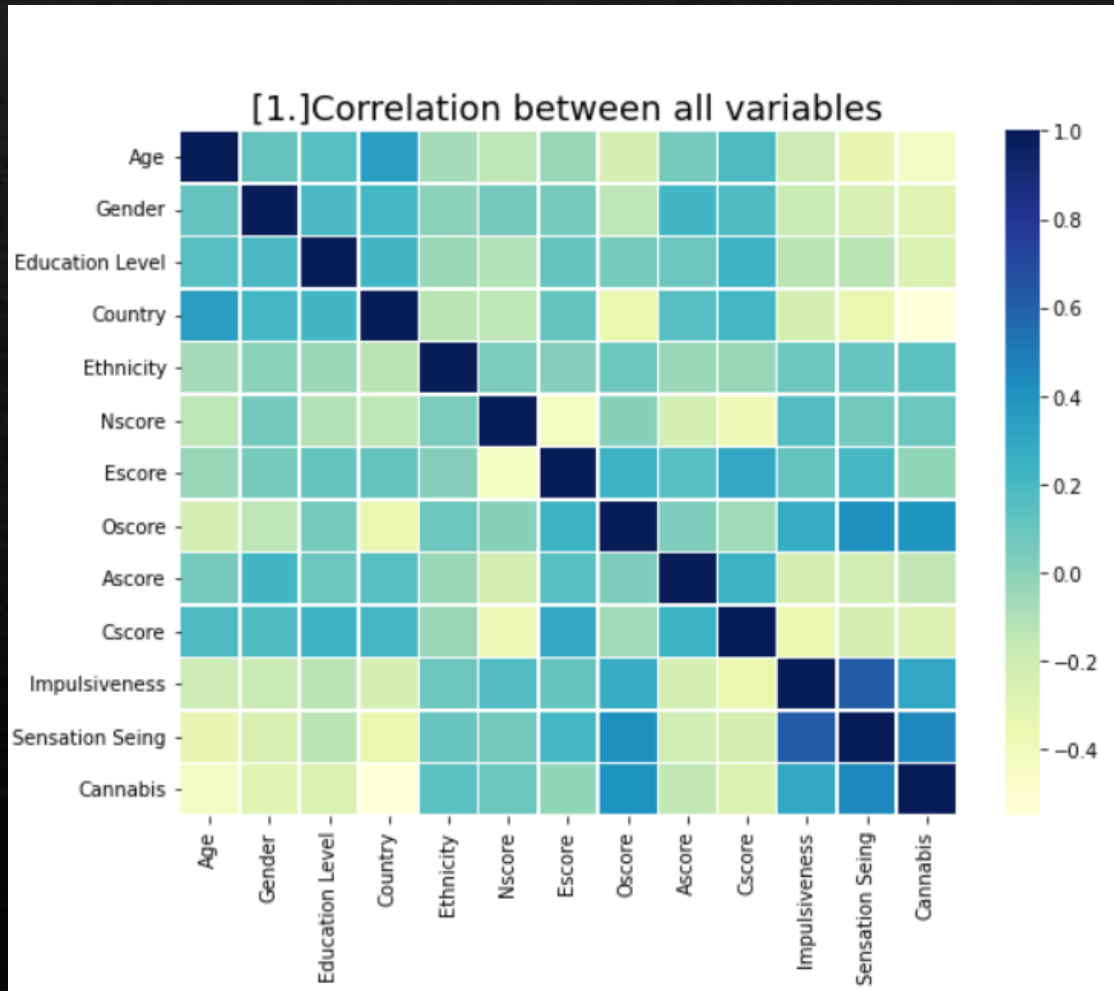
How to predict drug use, and more specifically Cannabis use in a population according to certain parameters?

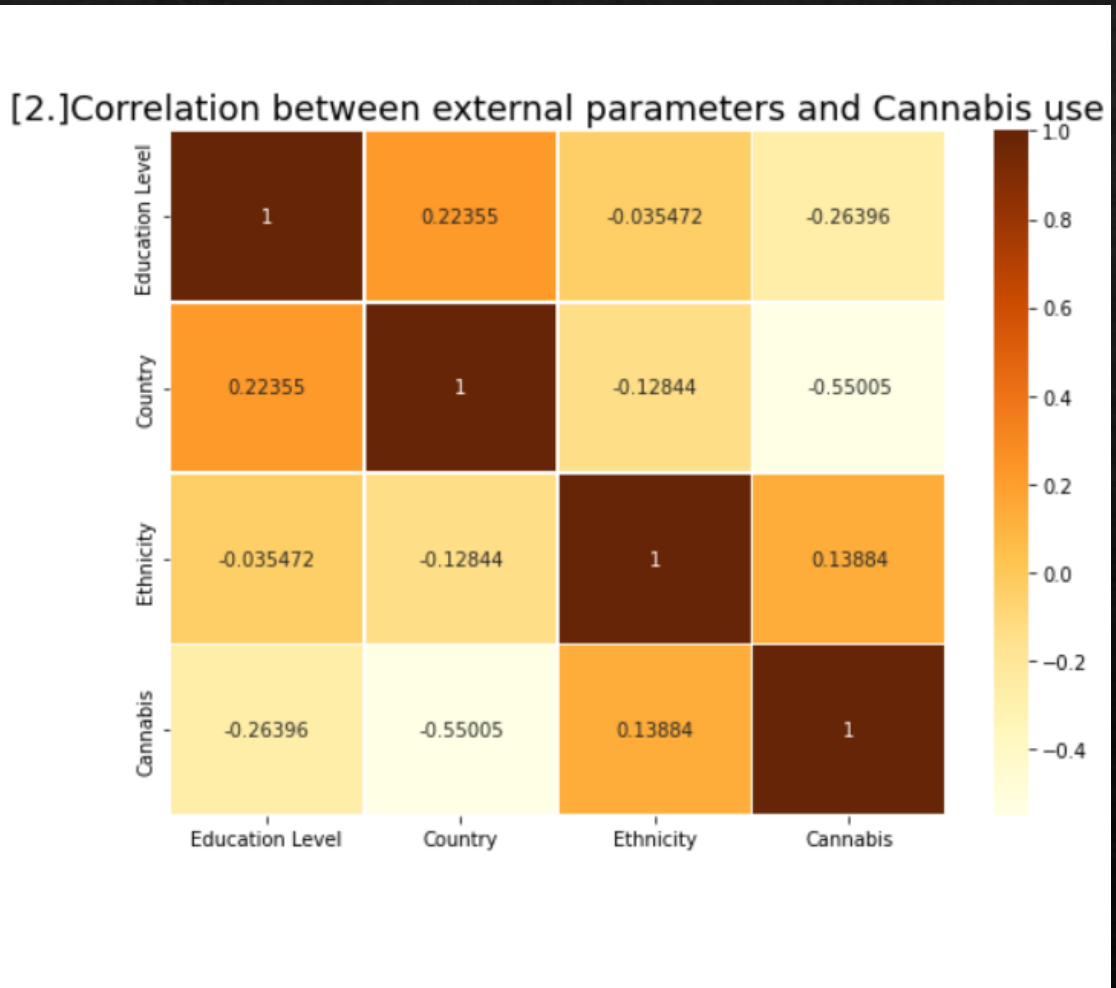
User vs non-user

Alcohol	Amphetamines	Amyl nitrite	Benzodiazepine	Caffeine	Cannabis
5	2	2	0	6	4
6	0	0	0	6	3
4	0	0	3	5	2
4	1	1	0	6	3
2	0	0	0	6	0

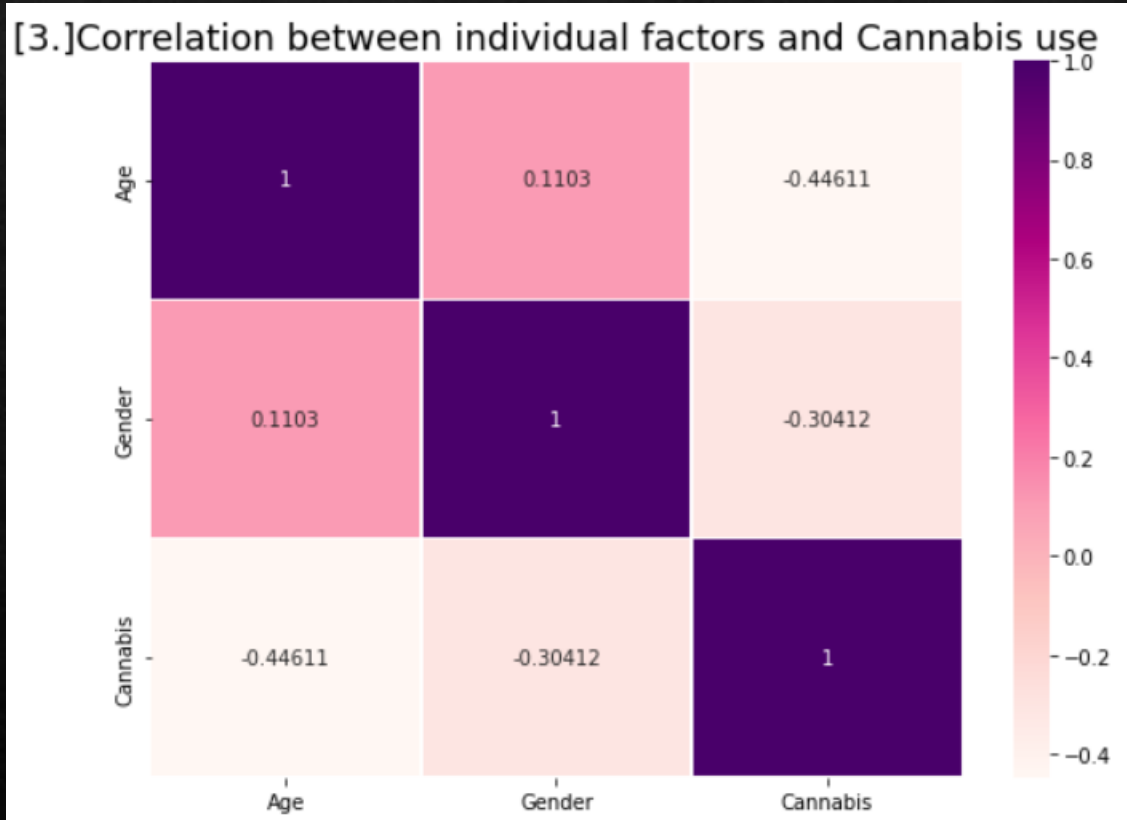
	Alcohol_user	Amphetamines_user	Amyl nitrite_user	Benzodiazepine_user	Caffeine_user	Cannabis_user
0	User	Non-user	Non-user	Non-user	User	User
1	User	Non-user	Non-user	Non-user	User	User
2	User	Non-user	Non-user	User	User	Non-user
3	User	Non-user	Non-user	Non-user	User	User
4	Non-user	Non-user	Non-user	Non-user	User	Non-user

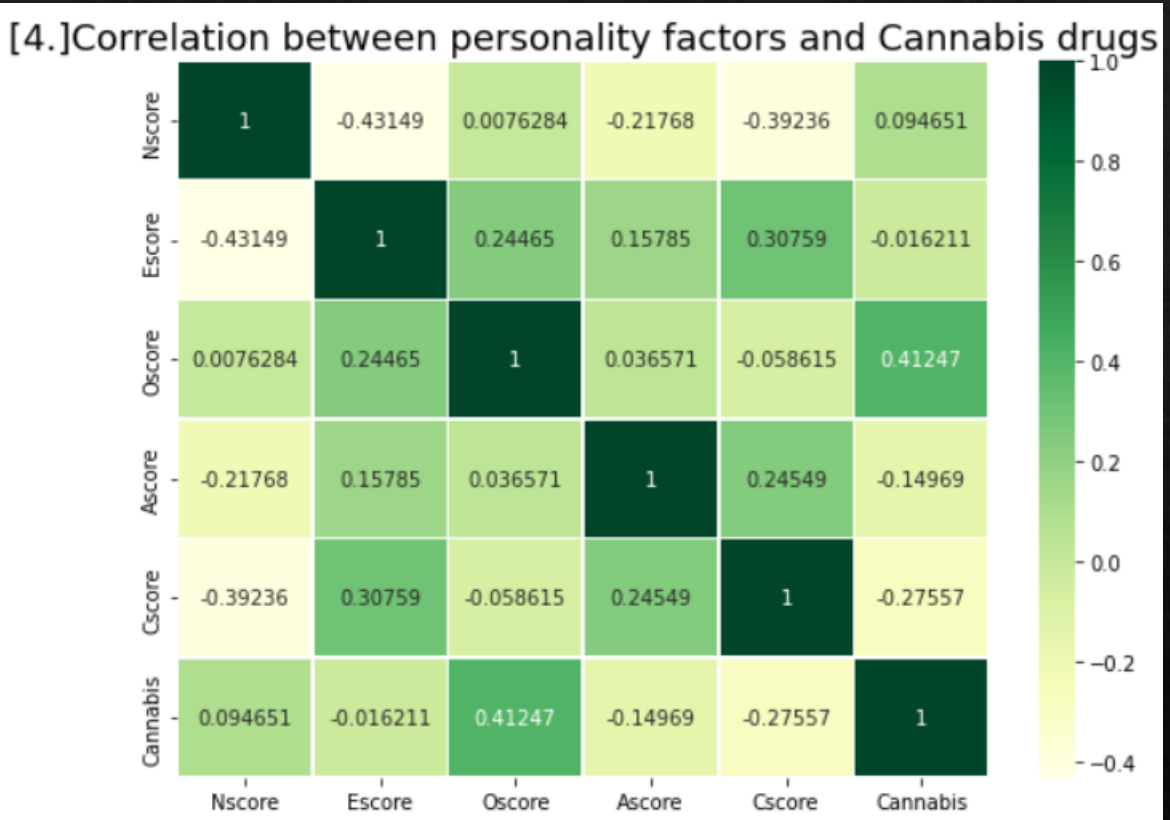
Correlations:





Correlations:





Machine Learning:

```
svr = svm.SVR()
cross_val_score(svr, X_train, y_train)

#instantiation
model_SVC = SVC( kernel = 'linear', gamma = 'scale', shrinking = False,)

#training
model_SVC.fit(X_train, y_train)

#precision calculation
print(model_SVC.score(X_test, y_test))

#Prediction
prediction = model_SVC.predict(X_test)

#display of results
result = "Result : "
if prediction[0] == 0:
  result = result + "Non-User"
if prediction[0] == 1:
  result = result + "User"
print(result)
```

```
0.8016129032258065
Result : Non-User
```

```
GB=GradientBoostingClassifier()
GB.fit(X_train,y_train)

#precision calculation
print(GB.score(X_test, y_test))

#Prediction
prediction = GB.predict(X_test)

#display of results
result = "Result : "
if prediction[0] == 0:
  result = result + "Non-User"
if prediction[0] == 1:
  result = result + "User"
print(result)
```

```
0.8209677419354838
Résultat : Non-User
```

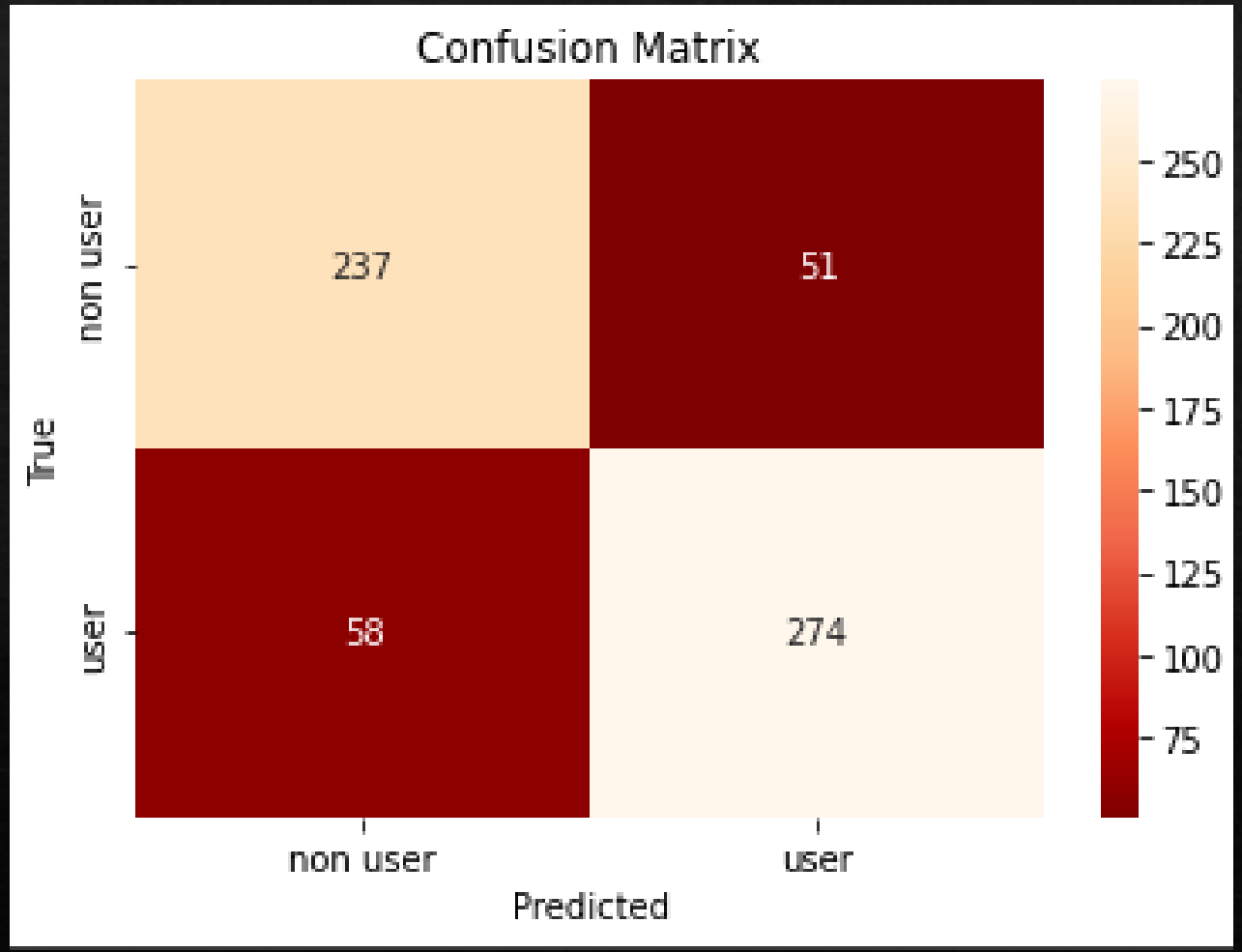
```
RFC=RandomForestClassifier()
RFC.fit(X_train,y_train)

#precision calculation
print(RFC.score(X_test, y_test))

#Prediction
prediction = RFC.predict(X_test)

#display of results
result = "Result : "
if prediction[0] == 0:
  result = result + "Non-User"
if prediction[0] == 1:
  result = result + "User"
print(result)
```

```
0.8080645161290323
Result : User
```

```
# Best model
best_algorithm, best_perf
```

```
(ExtraTreesClassifier(), 0.8451612903225807)
```

Best model:



Conclusion:

How to predict drug use, and more specifically Cannabis use in a population according to certain parameters?