

Drug Consumption

Analysis and prediction

Can we predict your drug consumption ?

By Janany Jegatheeswaran
and Ilies Gourri.

Sommaire

01

Dataset presentation

02

Data exploration

03

Data modeling

04

API

01

Dataset presentation

What are the features ? Their proportion ?

Data set description

Value	Description
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

- 1885 respondents
- 12 personal features (demographic and personality traits)
- 18 drugs, legal and illegal, rated by consumption frequency

	1	2	3	4	5	6	7	8	9	10	...	22	23	24	25	26	27	28	29	30	31
0																					
1	0.49788	0.48246	-0.05921	0.96082	0.12600	0.31287	-0.57545	-0.58331	-0.91699	-0.00665	...	CL0	CL0	CL0	CL0	CL0	CL0	CL0	CL2	CL0	CL0
2	-0.07854	-0.48246	1.98437	0.96082	-0.31685	-0.67825	1.93886	1.43533	0.76096	-0.14277	...	CL4	CL0	CL2	CL0	CL2	CL3	CL0	CL4	CL0	CL0
3	0.49788	-0.48246	-0.05921	0.96082	-0.31685	-0.46725	0.80523	-0.84732	-1.62090	-1.01450	...	CL0	CL0	CL0	CL0	CL0	CL0	CL1	CL0	CL0	CL0
4	-0.95197	0.48246	1.16365	0.96082	-0.31685	-0.14882	-0.80615	-0.01928	0.59042	0.58489	...	CL0	CL0	CL2	CL0	CL0	CL0	CL0	CL2	CL0	CL0
5	0.49788	0.48246	1.98437	0.96082	-0.31685	0.73545	-1.63340	-0.45174	-0.30172	1.30612	...	CL1	CL0	CL0	CL1	CL0	CL0	CL2	CL2	CL0	CL0

Data set description

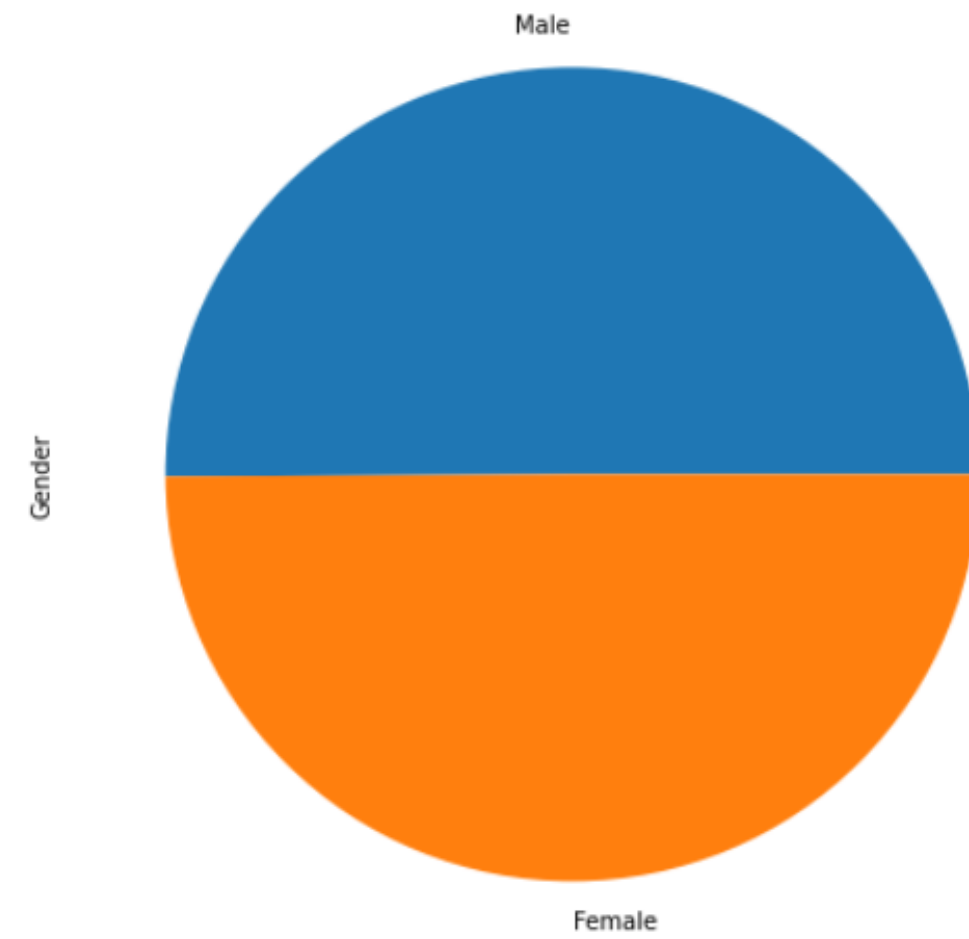
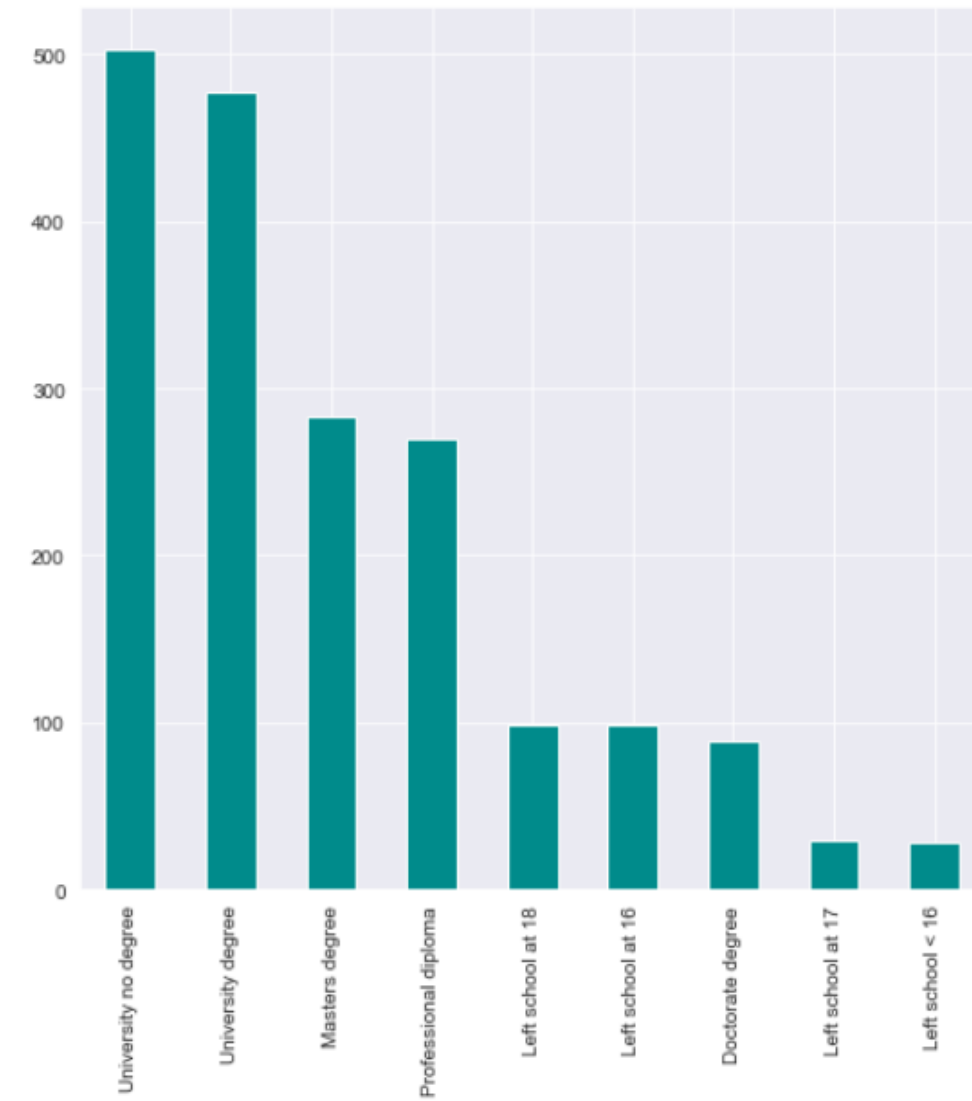
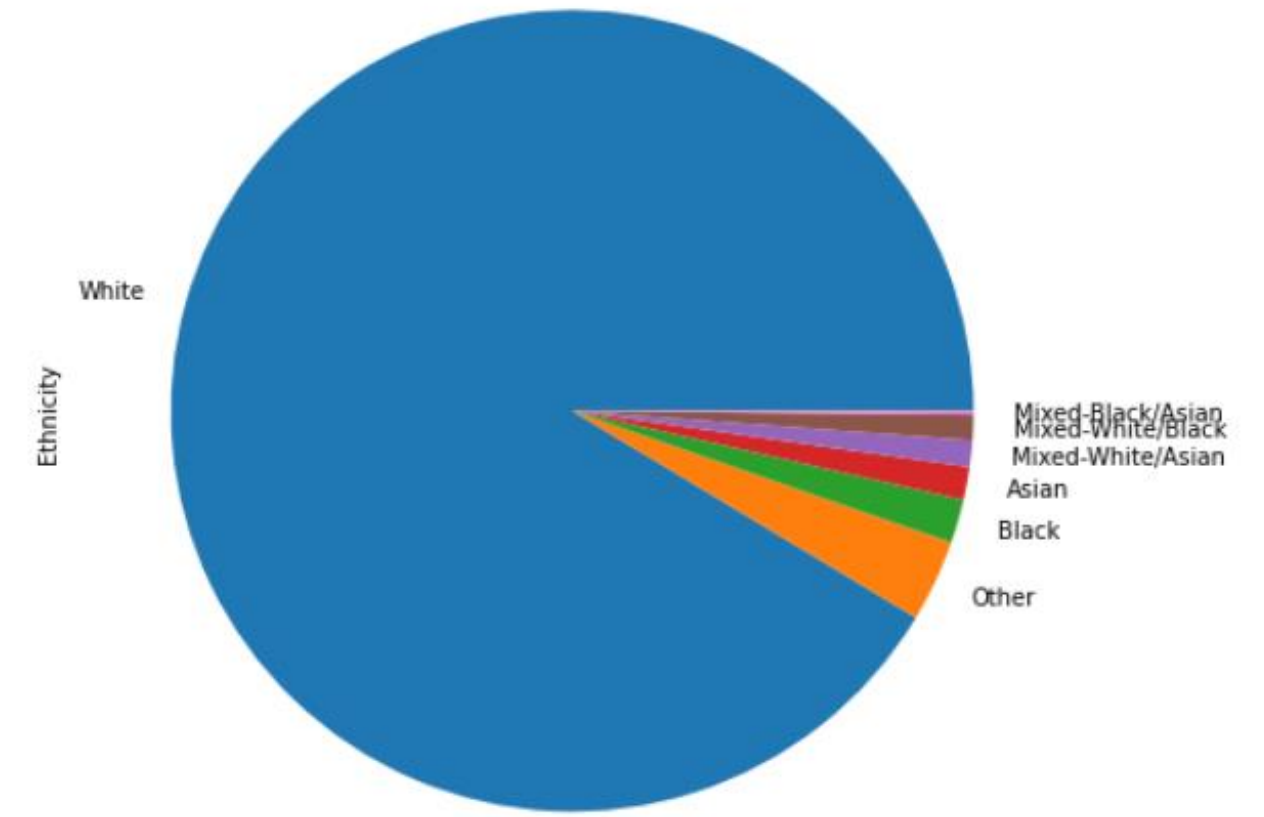
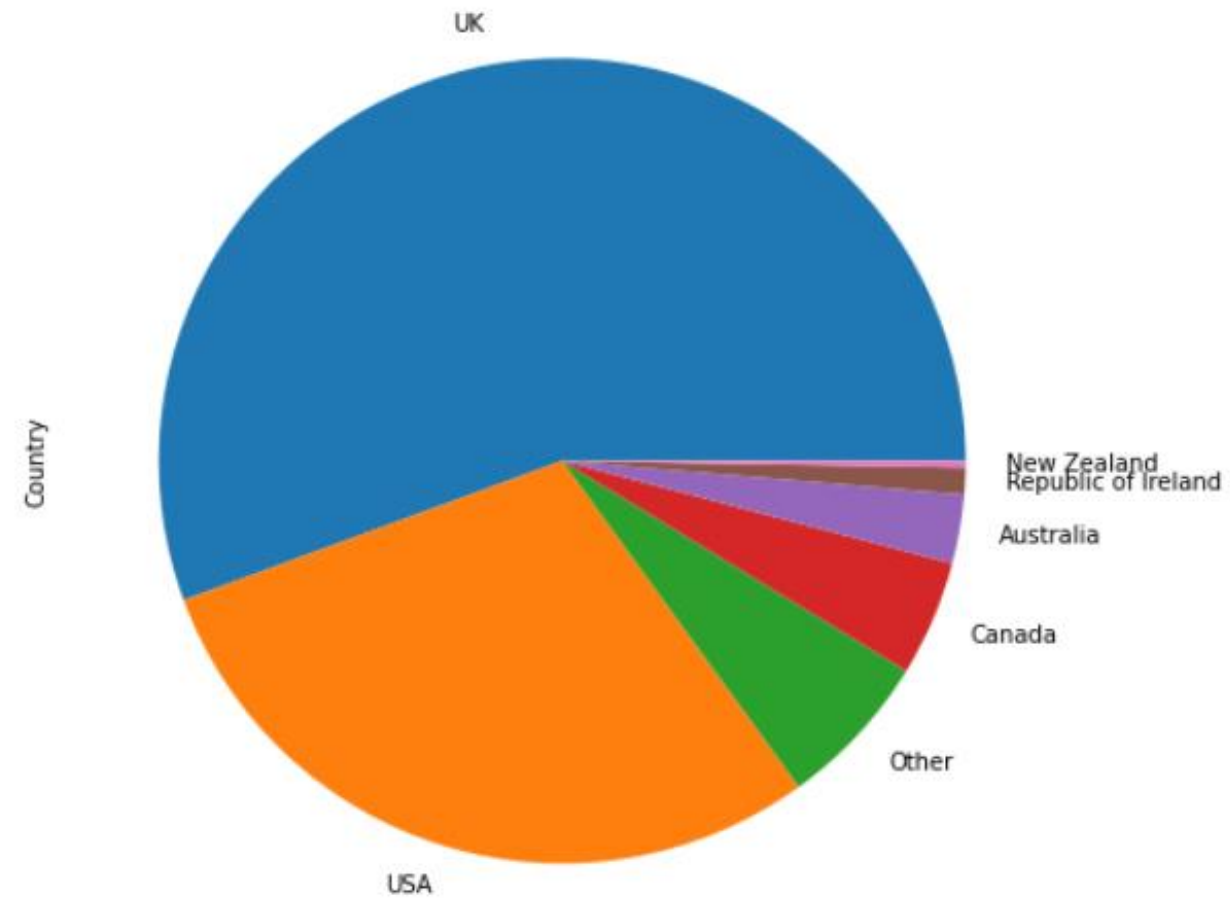
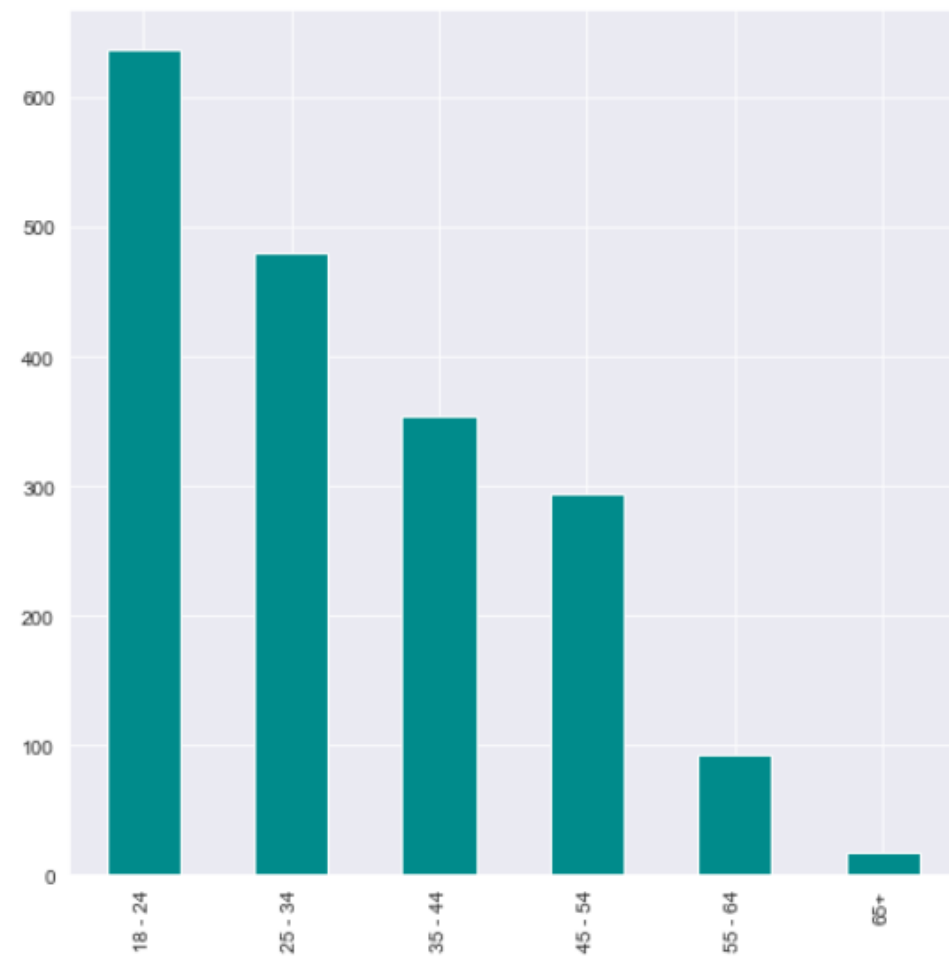
- 1885 - 8 = 1877 respondents
 - Rename columns
- 17 drugs, legal and illegal, rated by consumption frequency

ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS	Alcohol	Amphet	Amyl	Benzos
1	35 - 44	Female	Professional diploma	UK	Mixed-White/Asian	0.31287	-0.57545	-0.58331	-0.91699	-0.00665	-0.21712	-1.18084	5	2	0	2
2	25 - 34	Male	Doctorate degree	UK	White	-0.67825	1.93886	1.43533	0.76096	-0.14277	-0.71126	-0.21575	5	2	2	0
3	35 - 44	Male	Professional diploma	UK	White	-0.46725	0.80523	-0.84732	-1.62090	-1.01450	-1.37983	0.40148	6	0	0	0
4	18 - 24	Female	Masters degree	UK	White	-0.14882	-0.80615	-0.01928	0.59042	0.58489	-1.37983	-1.18084	4	0	0	3
5	35 - 44	Female	Doctorate degree	UK	White	0.73545	-1.63340	-0.45174	-0.30172	1.30612	-0.21712	-0.21575	4	1	1	0

02

Data exploration

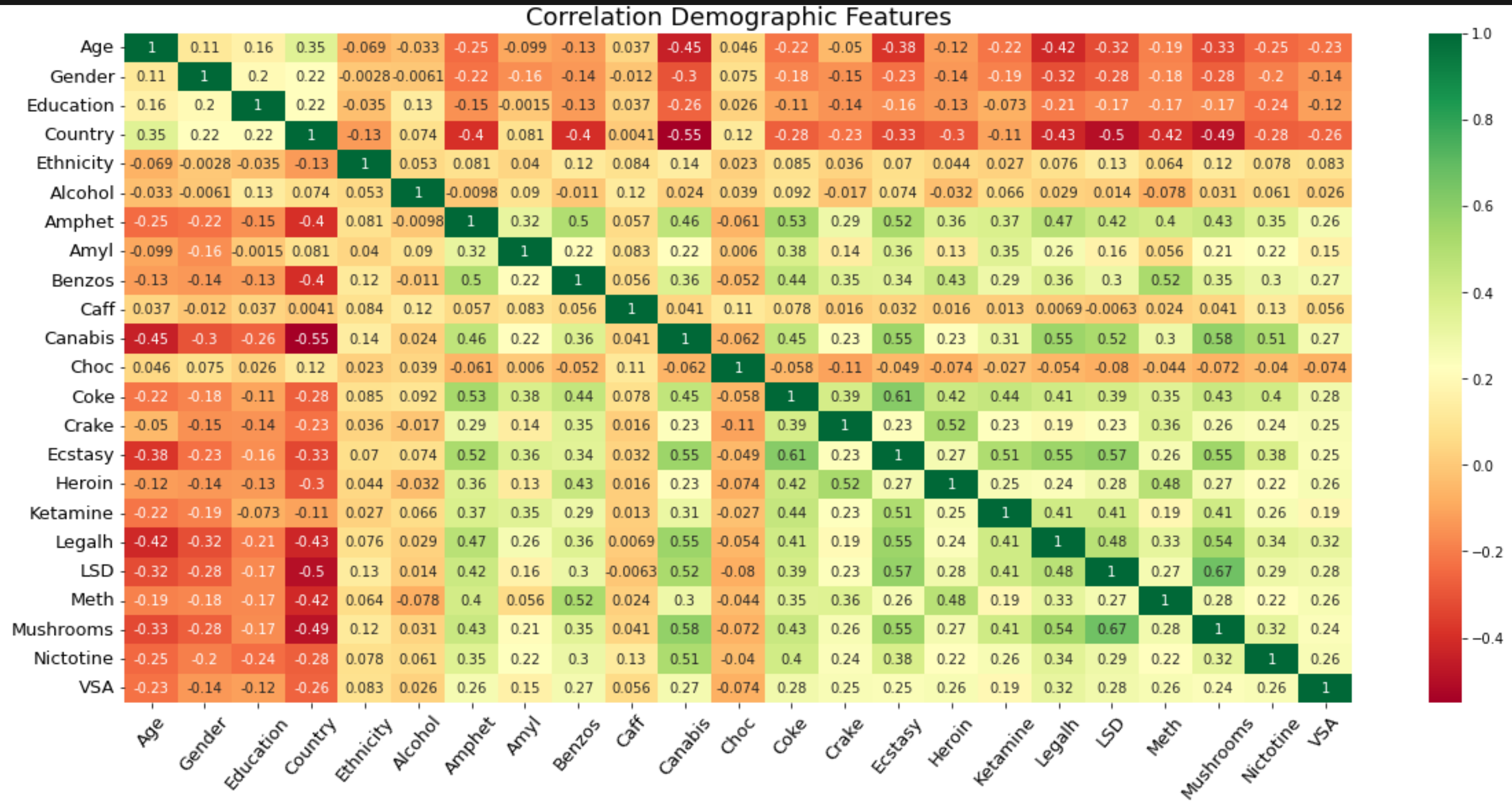
Is there some features more relevant than the other ?

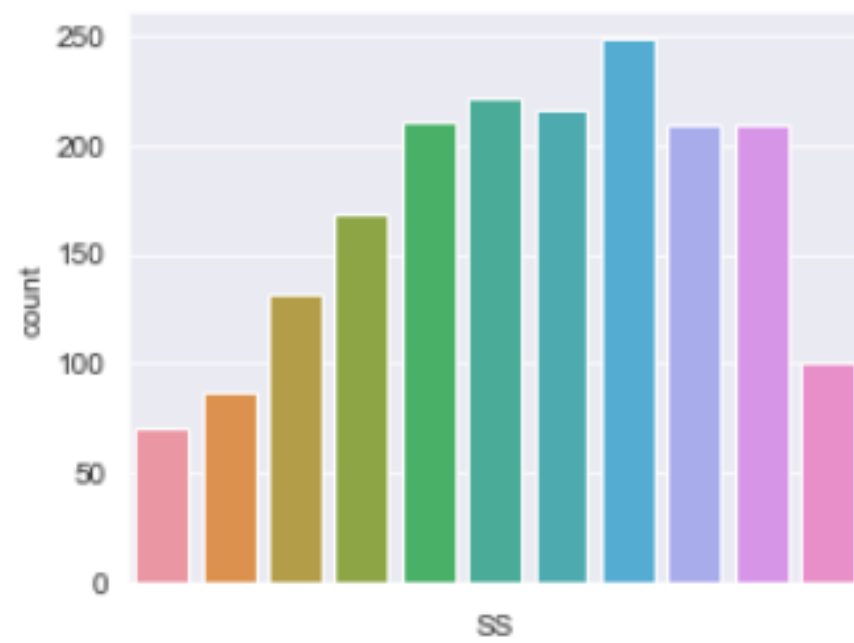
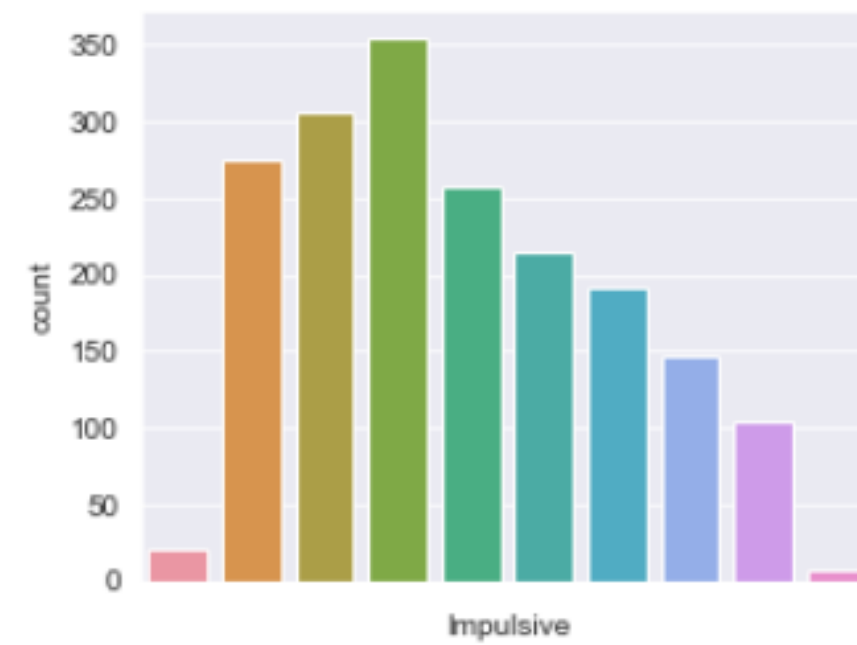
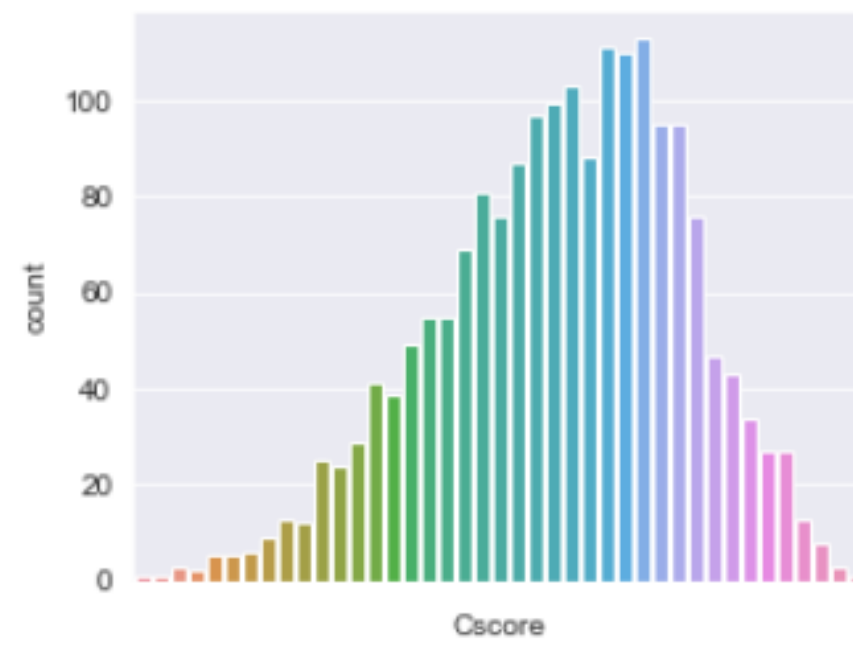
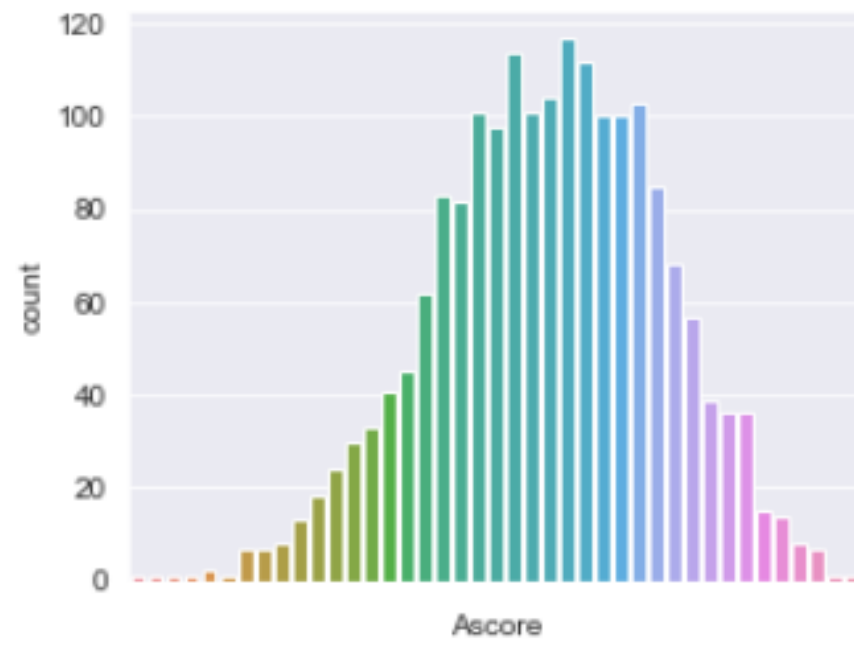
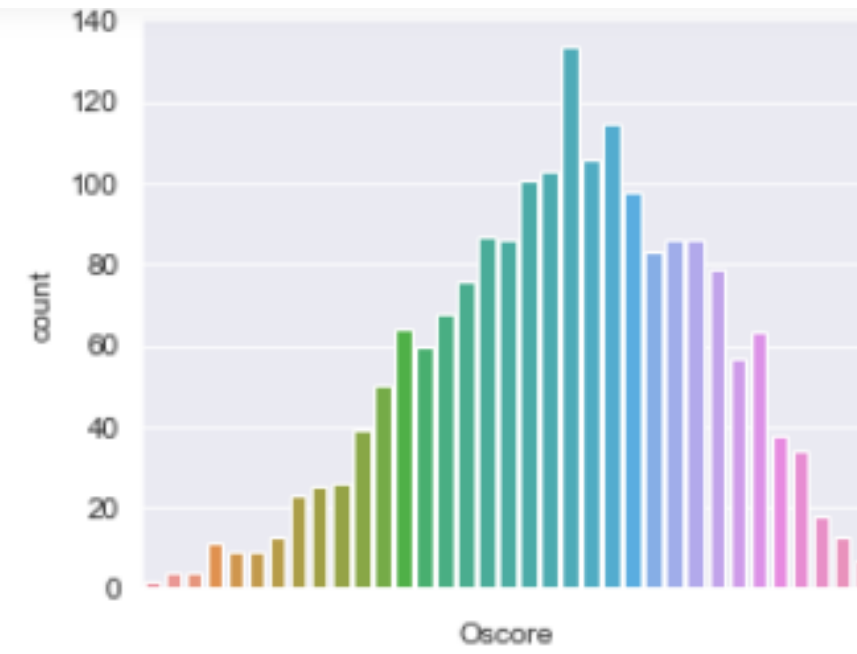
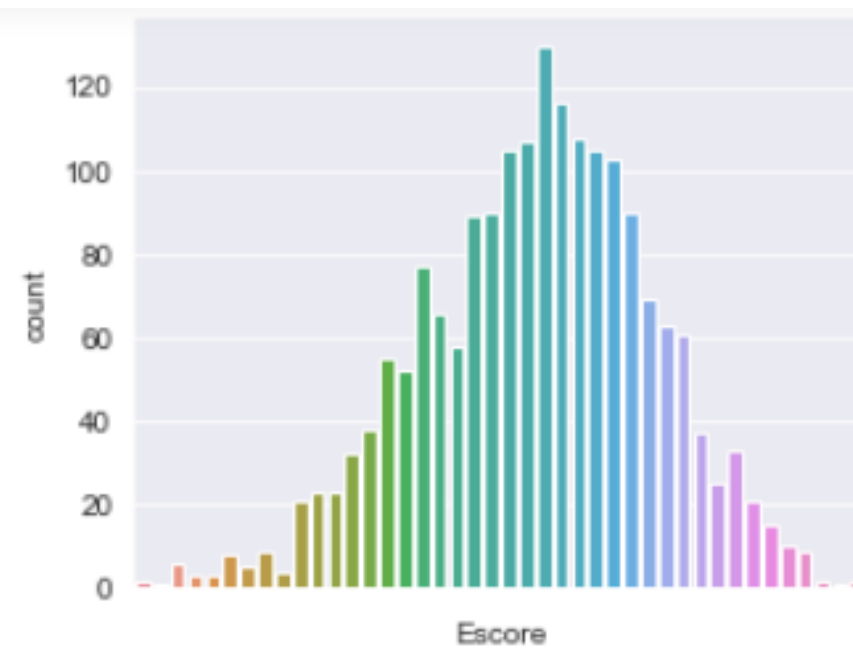
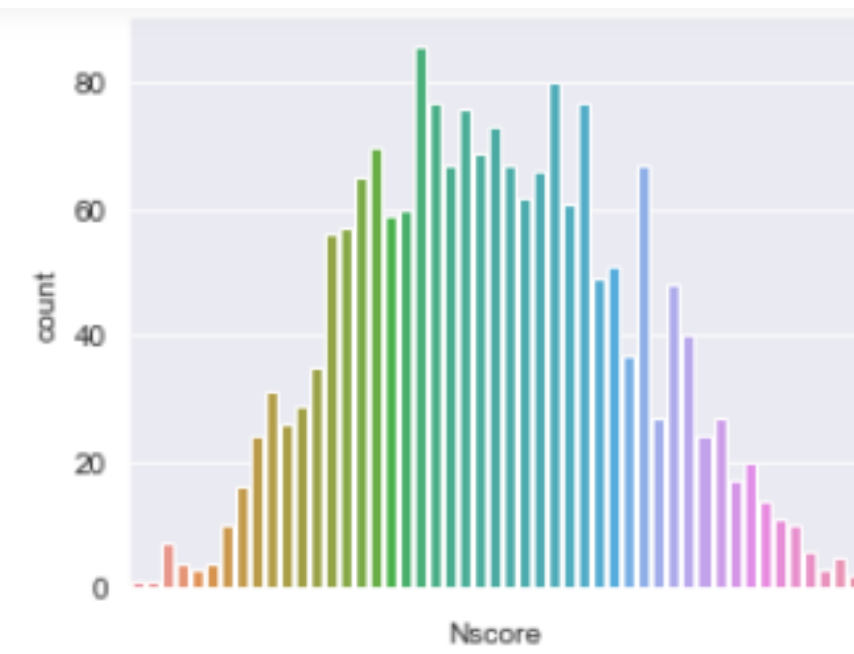


Proportion

Demographic

Correlation : drugs and demographic.

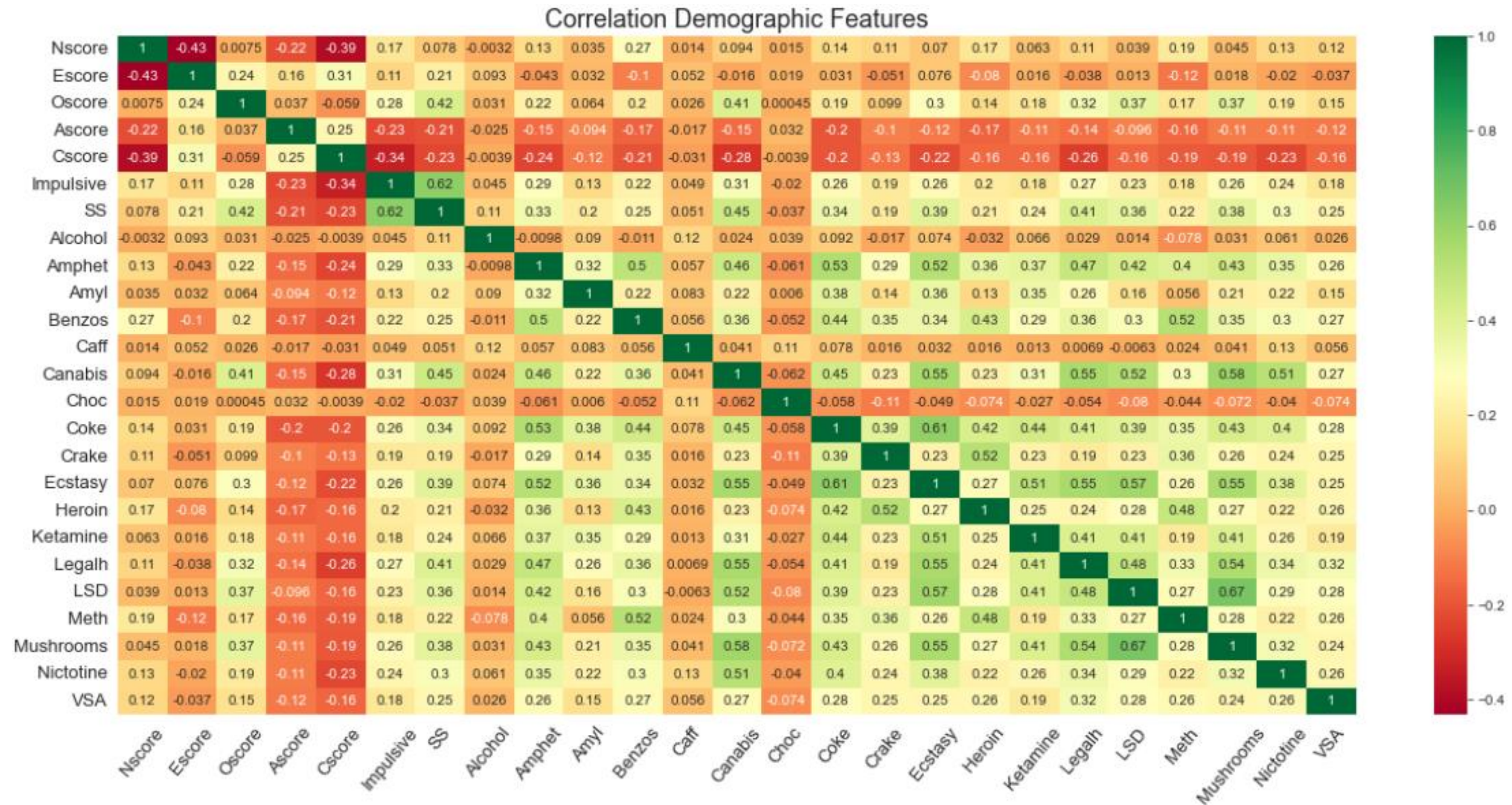




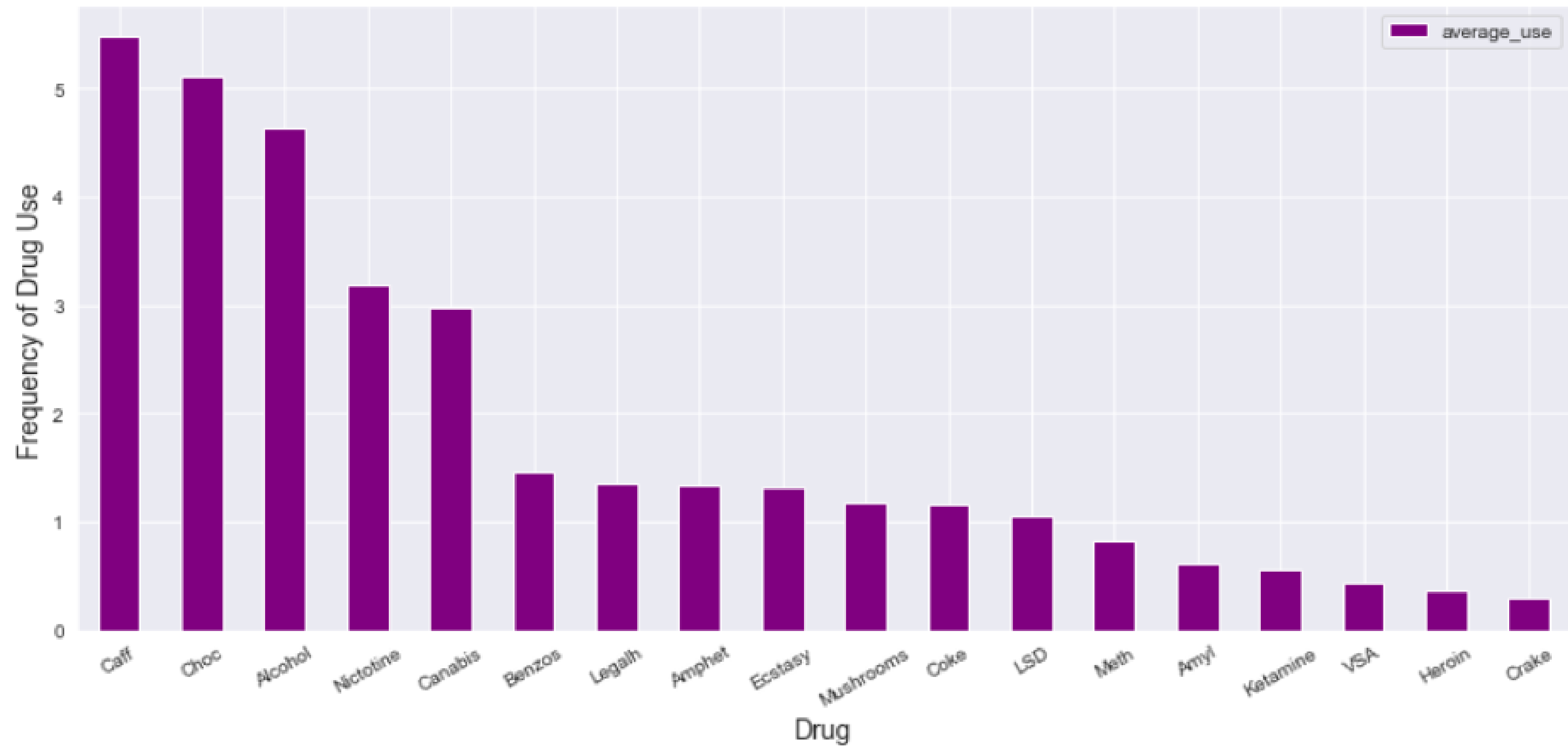
Proportion

Personality features.

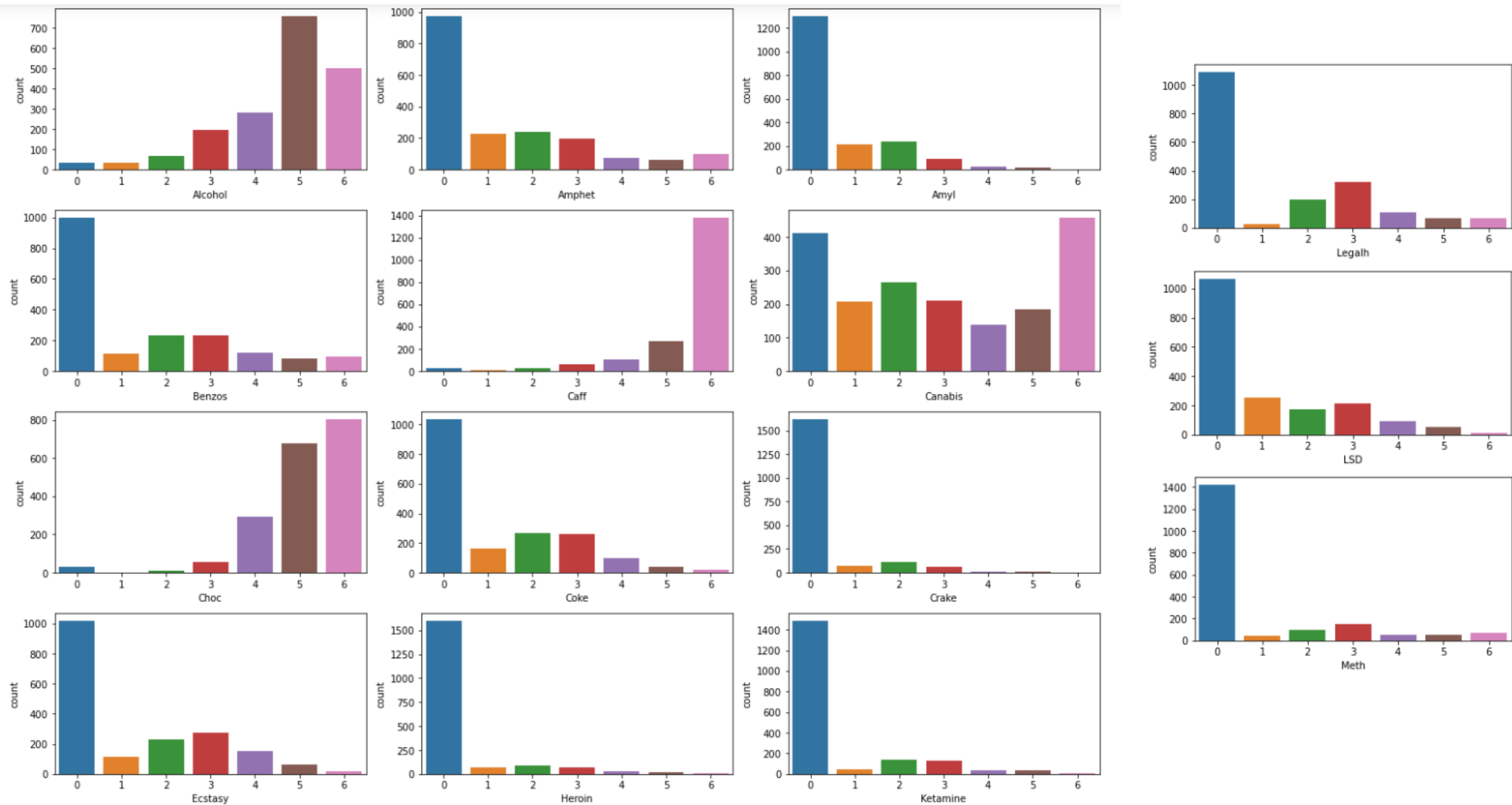
Correlation : drugs and personality.



Average use of each drug



Consumption of Drug by Different Classes



03

Data modeling

Can we predict if you are a drug consumer ?

- Logistic Regression
- Support Vector Machines
- Random Forest Classifier
 - KNN Classifier

Algorithms

ACCURACY

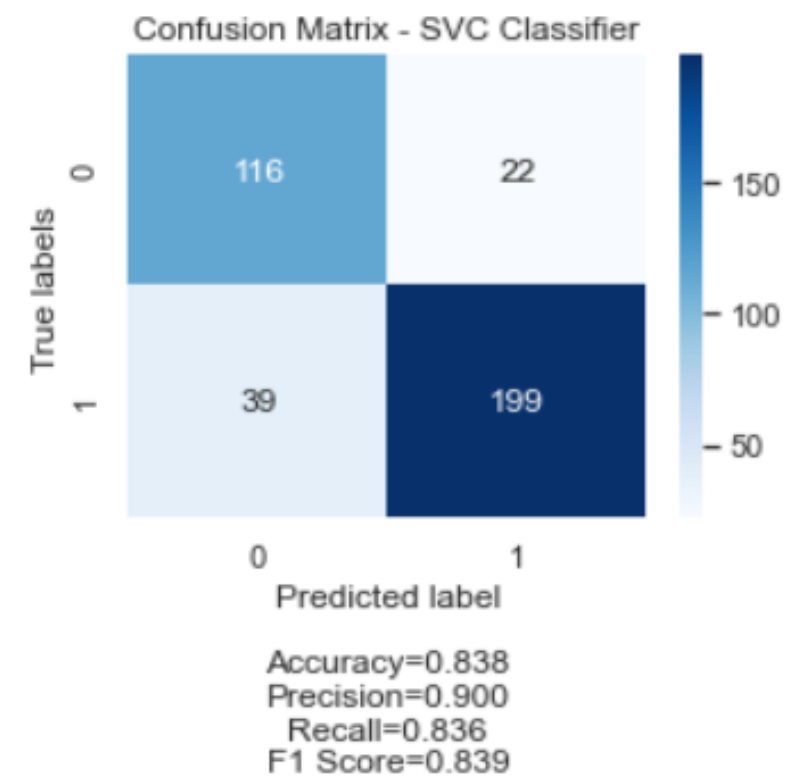
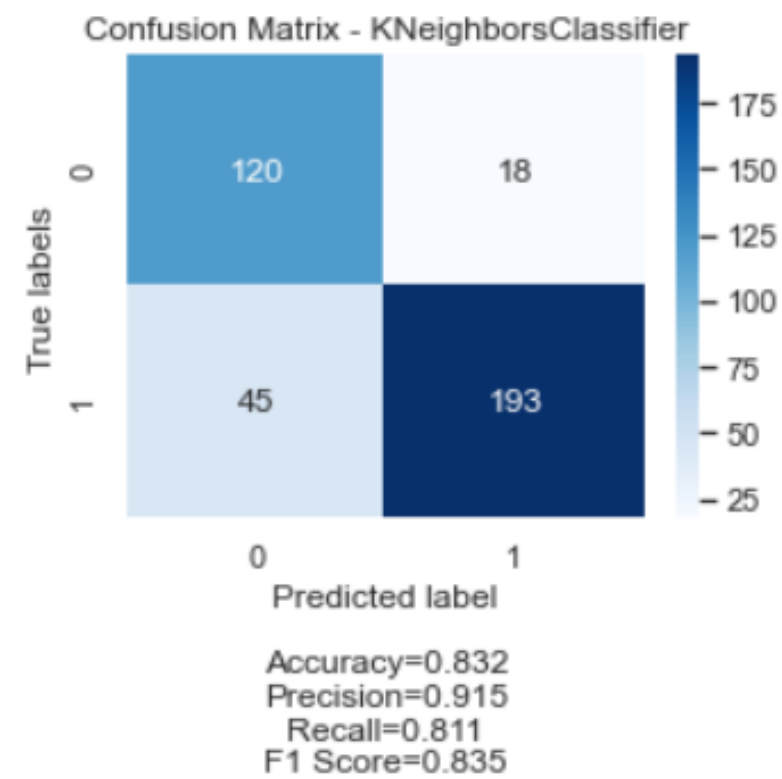
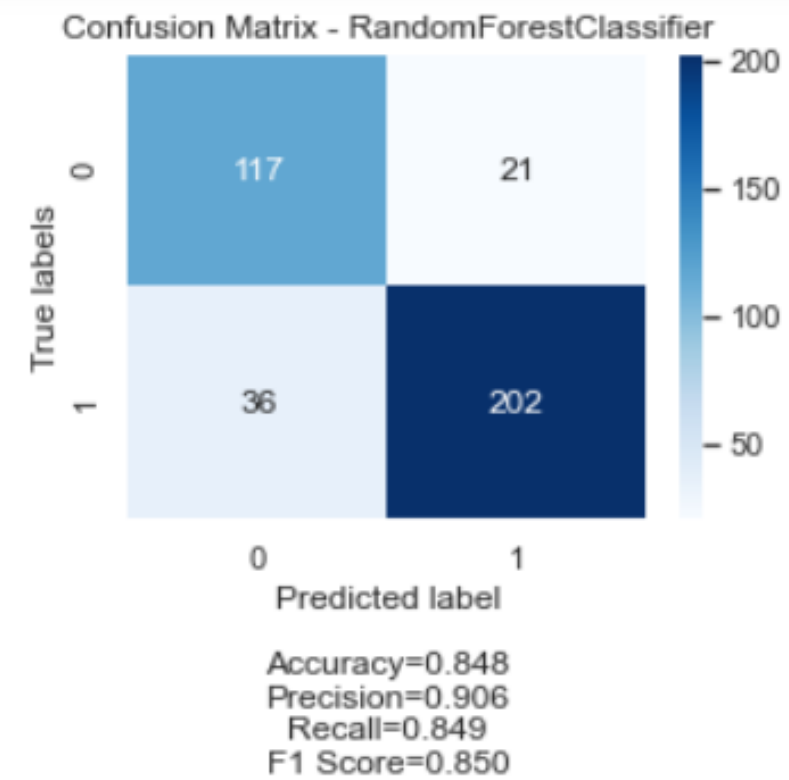
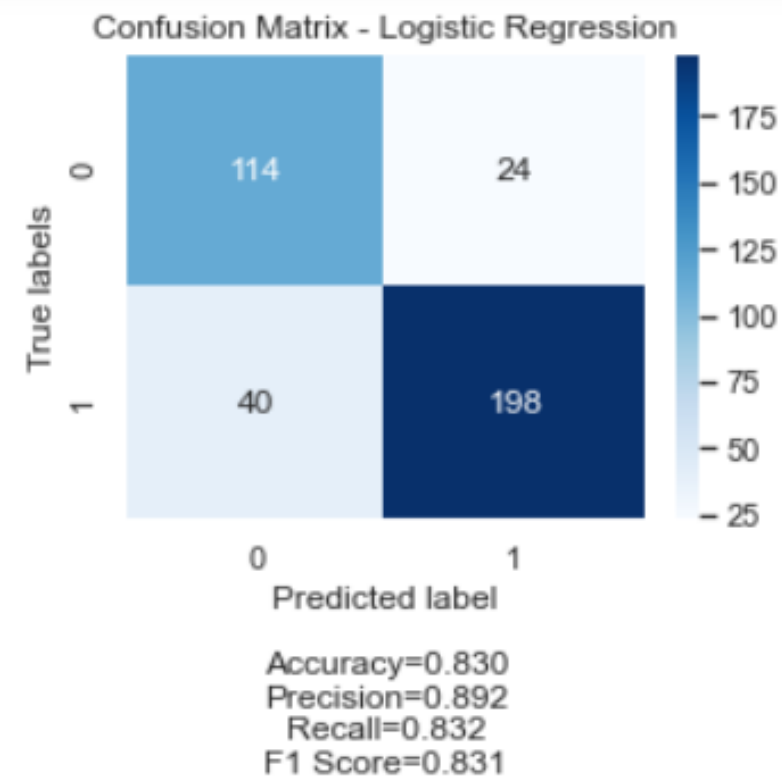
Logistic Regression Accuracy: 82.98%
Support Vector Machines Accuracy: 83.78%
Random Forest Classifier Accuracy: 84.84%
KNN Classifier Accuracy: 83.24%

F1 SCORES

Logistic Regression F1-Score: 0.83149
Support Vector Machines F1-Score: 0.83947
Random Forest Classifier F1-Score: 0.84985
KNN Classifier F1-Score: 0.83487

Model predictions and score

Confusion matrix



Grid search

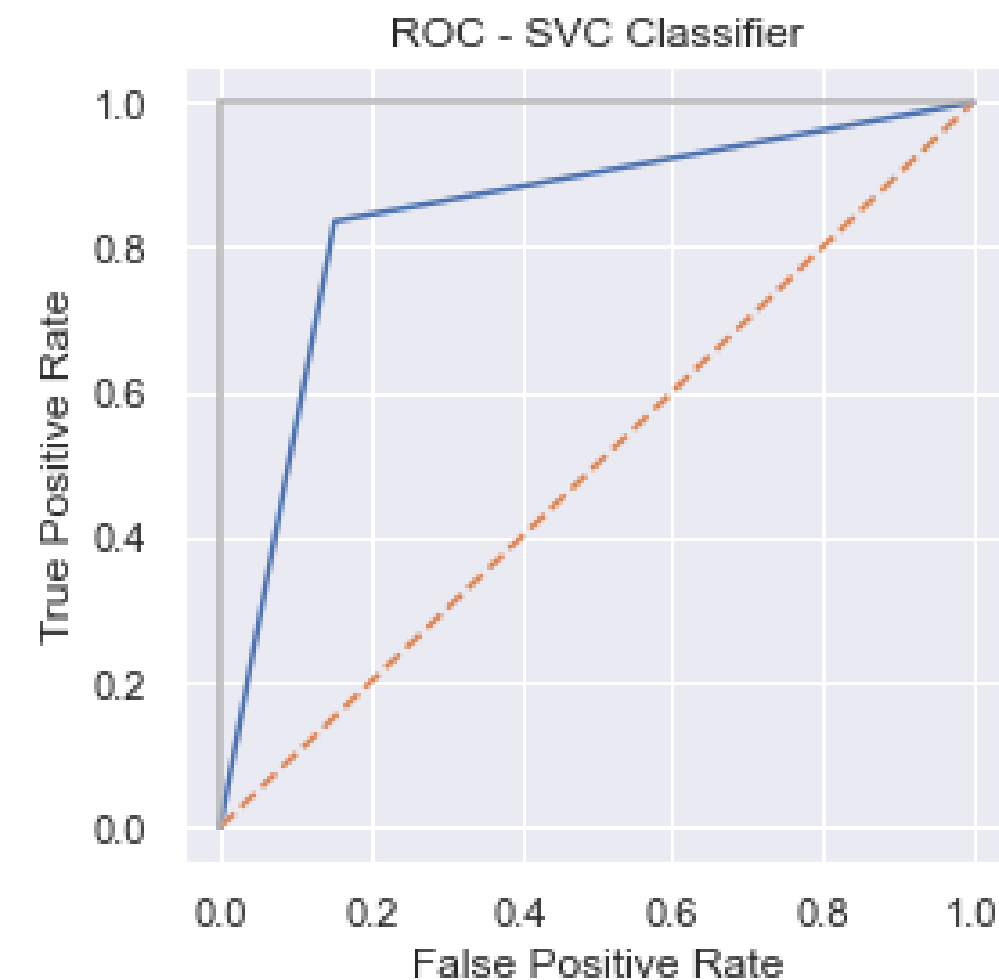
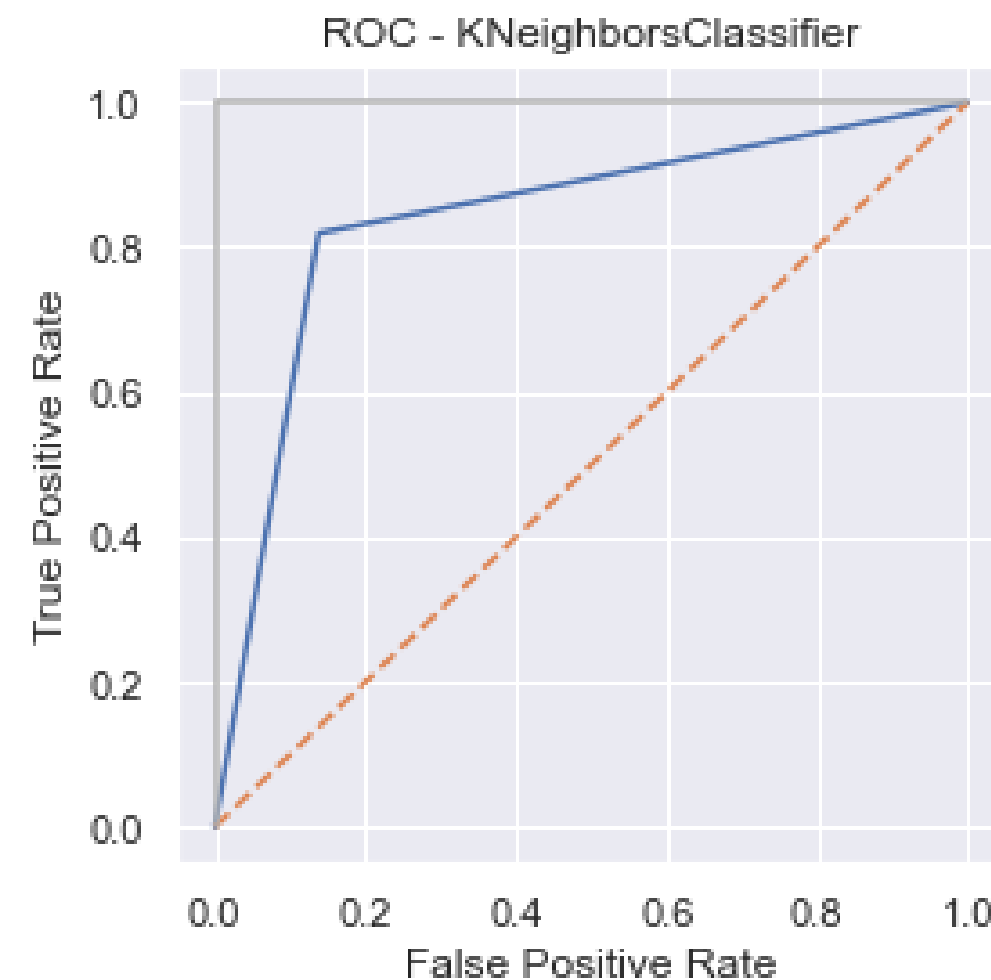
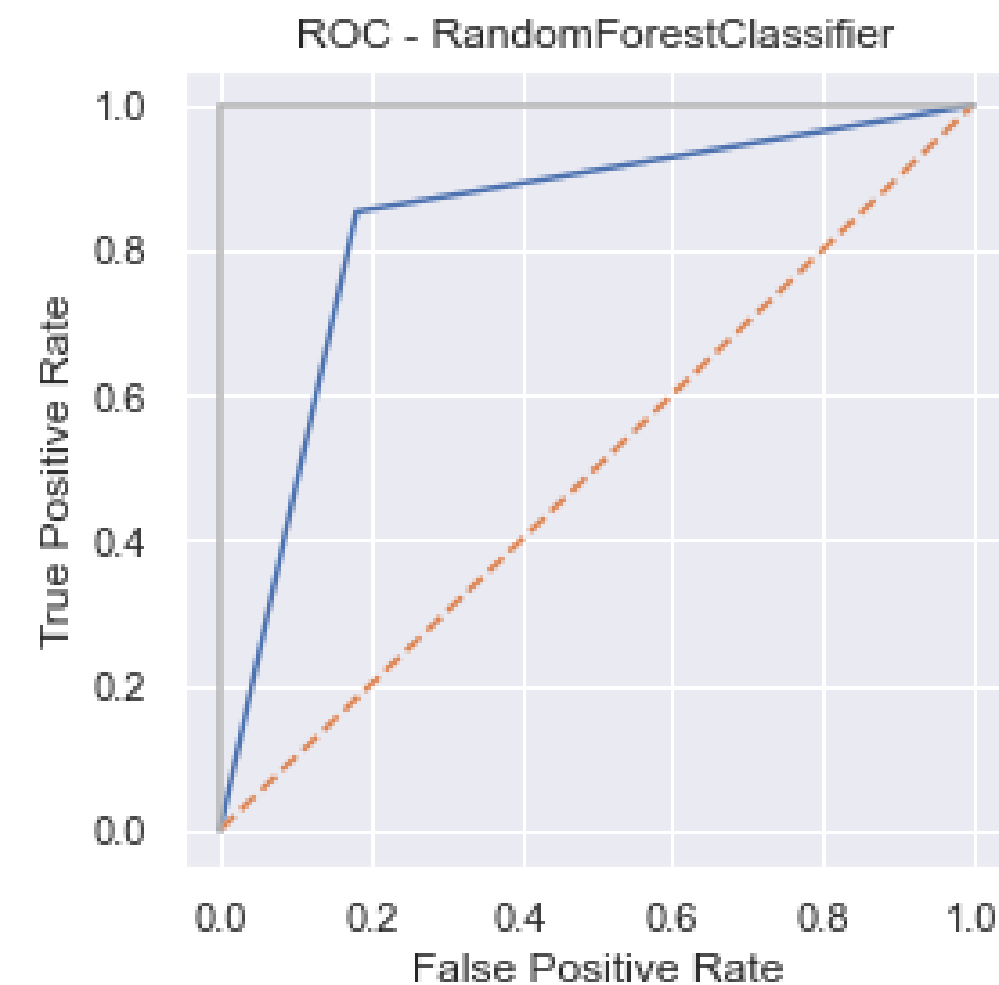
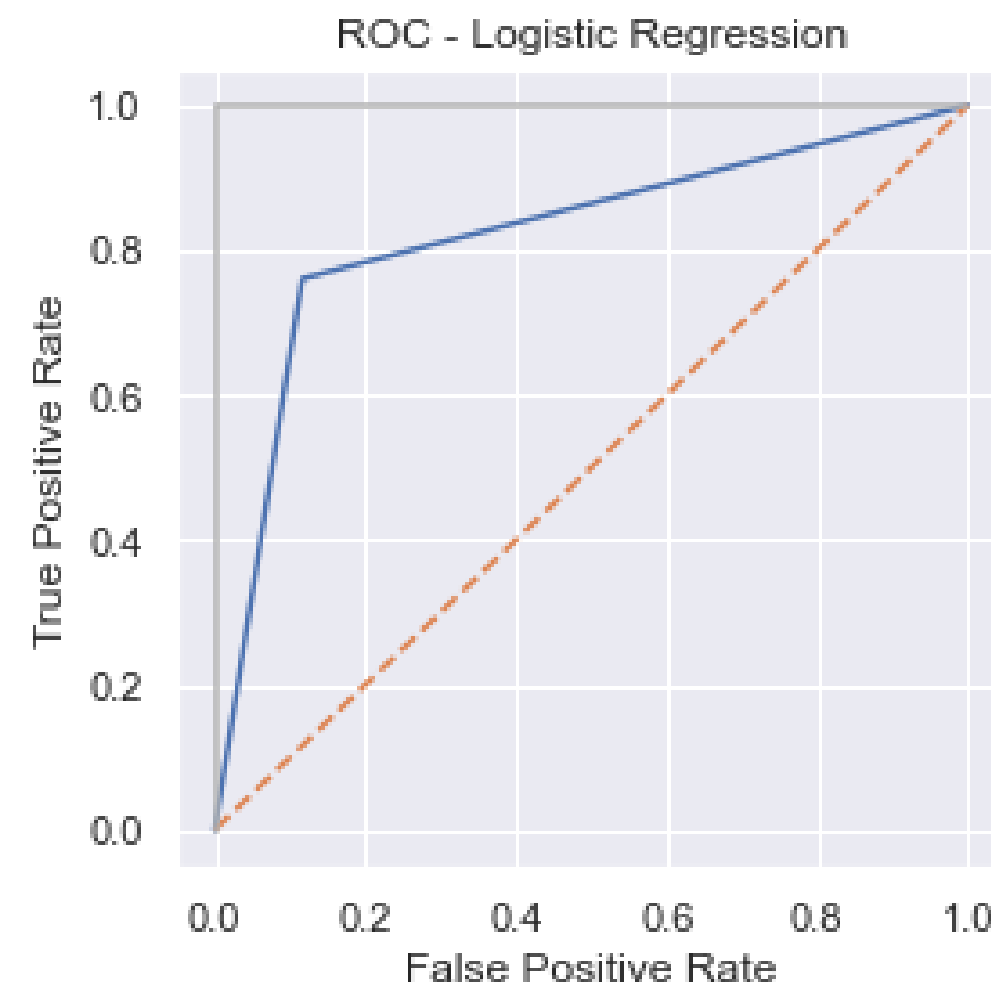
```
____LOGISTIC REGRESSION____
Fitting 5 folds for each of 14 candidates, totalling 70 fits
Best score :
  0.8439834617587707
Best parameters :
  {'C': 0.1, 'class_weight': 'balanced', 'penalty': 'l2', 'solver': 'liblinear'}
  F1-Score: 0.80926
____SVC____
Fitting 5 folds for each of 25 candidates, totalling 125 fits
Best score :
  0.8639716224304694
Best parameters :
  {'C': 1, 'gamma': 0.1, 'kernel': 'rbf'}
  F1-Score: 0.84217
____KNN____
Fitting 5 folds for each of 116 candidates, totalling 580 fits
Best score :
  0.8494959542301407
Best parameters :
  {'metric': 'manhattan', 'n_neighbors': 29, 'weights': 'distance'}
  F1-Score: 0.83732
____RANDOM FOREST____
Fitting 5 folds for each of 108 candidates, totalling 540 fits

Best score :
  0.8630818822275999
Best parameters :
  {'class_weight': 'balanced', 'criterion': 'gini', 'max_depth': 36, 'max_features': 'auto', 'n_estimators': 256}
  F1-Score: 0.85224
Wall time: 1min 12s
```

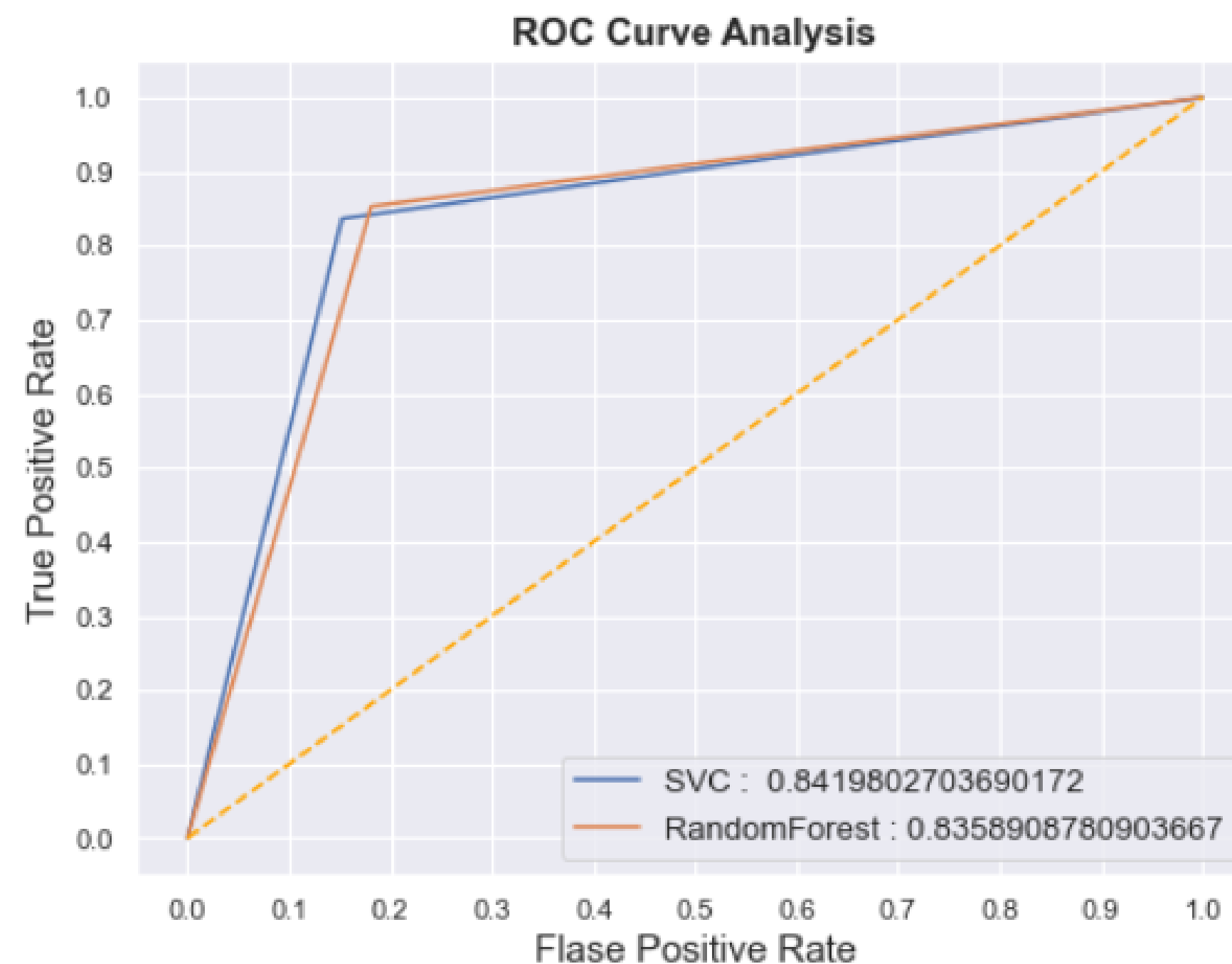
ROC curve :

AUC Score

```
Logistic Regression : 0.8222810863475826  
RandomForestClassifier : 0.8358908780903667  
KNeighborsClassifier : 0.8408232858360736  
SVC Classifier : 0.8419802703690172
```



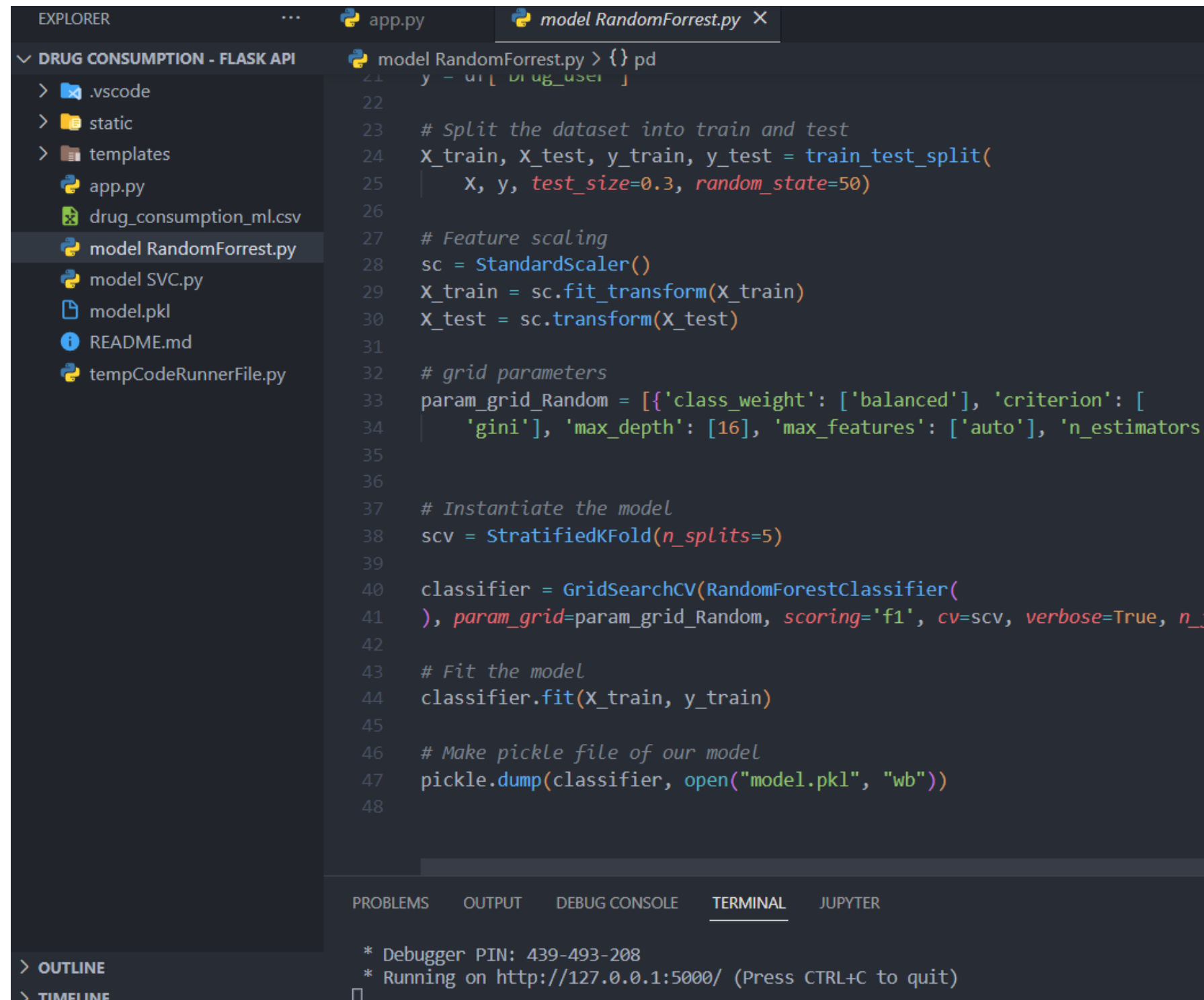
**Final model
choice :
SVC**



04

API

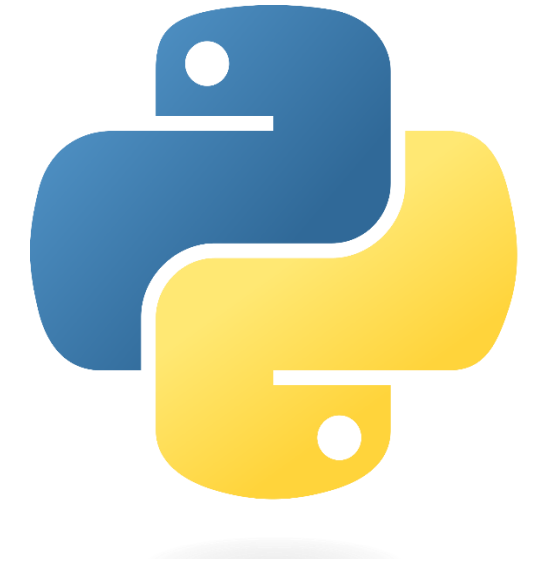
How can we deploy our model ?



```
21 y = util.load_csv('drug_consumption_ml.csv')
22
23 # Split the dataset into train and test
24 X_train, X_test, y_train, y_test = train_test_split(
25     X, y, test_size=0.3, random_state=50)
26
27 # Feature scaling
28 sc = StandardScaler()
29 X_train = sc.fit_transform(X_train)
30 X_test = sc.transform(X_test)
31
32 # grid parameters
33 param_grid_Random = [{'class_weight': ['balanced'], 'criterion': ['gini'], 'max_depth': [16], 'max_features': ['auto'], 'n_estimators': [100]}]
34
35
36
37 # Instantiate the model
38 scv = StratifiedKFold(n_splits=5)
39
40 classifier = GridSearchCV(RandomForestClassifier(
41 ), param_grid=param_grid_Random, scoring='f1', cv=scv, verbose=True, n_jobs=-1)
42
43 # Fit the model
44 classifier.fit(X_train, y_train)
45
46 # Make pickle file of our model
47 pickle.dump(classifier, open("model.pkl", "wb"))
48
```

PROBLEMS OUTPUT DEBUG CONSOLE **TERMINAL** JUPYTER

```
* Debugger PIN: 439-493-208
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```



```
<label for="Escore">Extraversion</label>
<select id="Escore" name="Escore" required>
  Select length of the course
  <option disabled selected value>-- select an option --</option>
  <option value="-3.27393">Low</option>
  <option value="0">Medium</option>
  <option value="3.27393">High</option>
</select>
```

Flask API

API Exemple

Neuroticism

Extraversion

Openness to experience

Agreeableness

Conscientiousness

Impulsiveness

Sensation seeking

Age

Gender

Education

Country

Ethnicity

Predict

Fill out your personality traits to predict your potential attraction for drugs

Fill out more information here :

You have high risk of drug consumption :(

Thank you !

Drug Consumption | Analysis and prediction



**JANANY
JEGATHEESWARAN**



**ILIES
GOURRI**

DIA 3 | Python for data analysis