Drug Consumption

Analysis and prediction

Can we predict your drug consumption ?

By Janany Jegatheeswaran
and Ilies Gourri.

Sommaire

01
Dataset presentation

What are the features ? Their proportion ?

Data set description

- 1885 respondents
- 12 personals 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

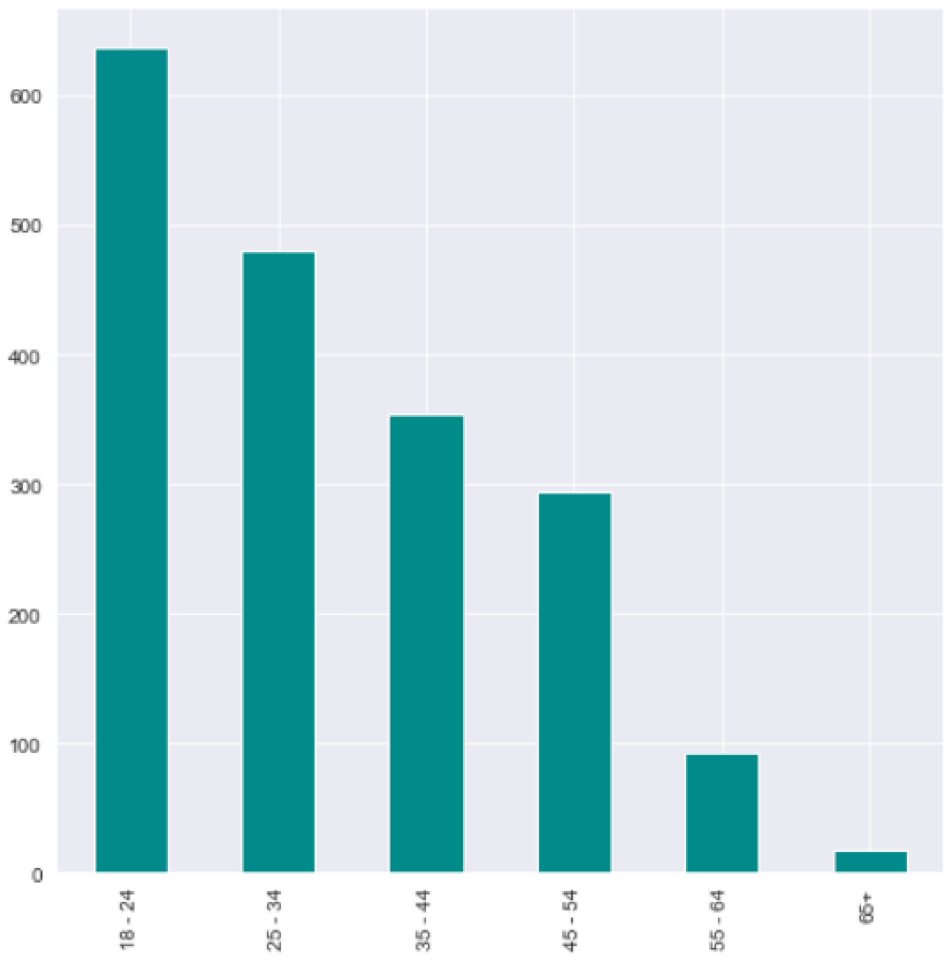
Data set description

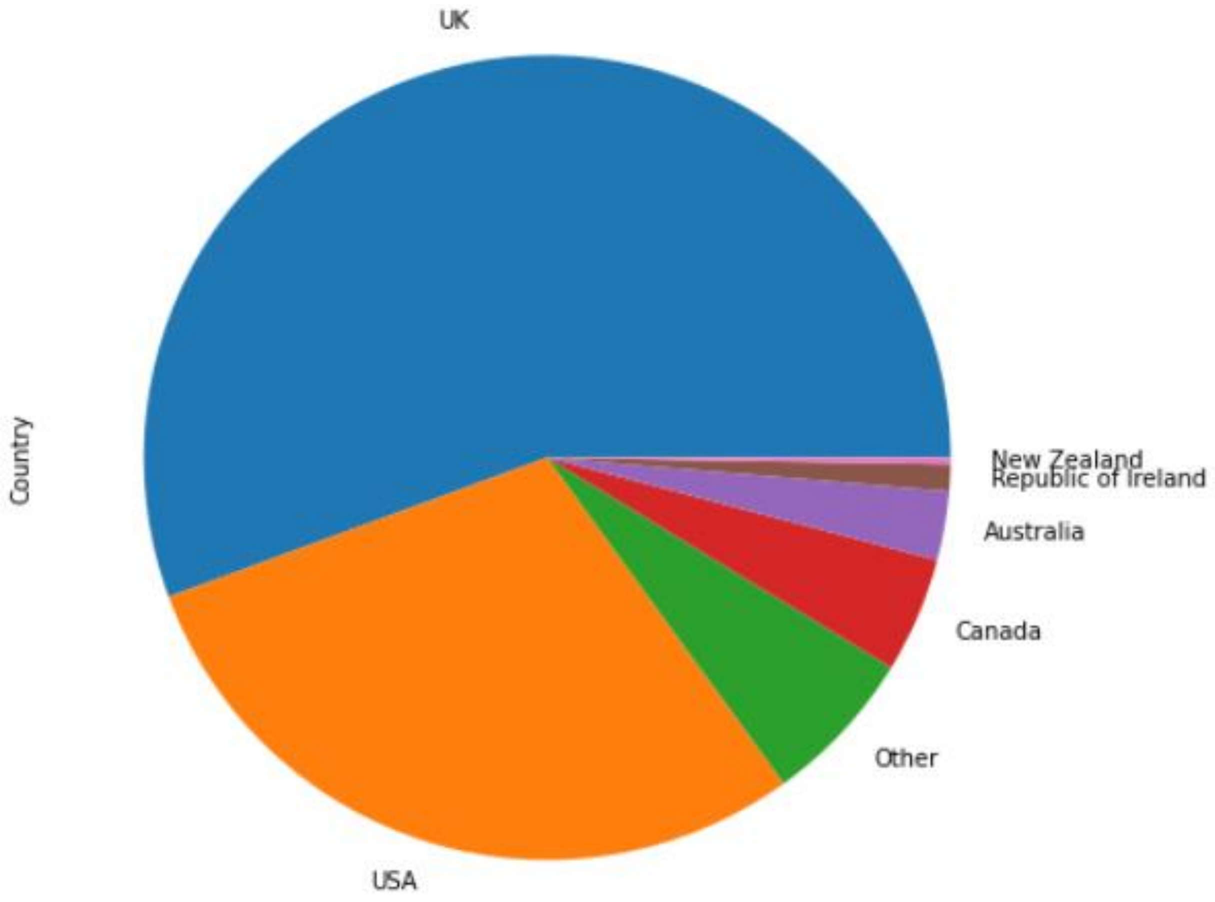
- 1885 - 8 = 1877 respondents
- Rename columns
- 17 drugs, legal and illegal, rated by consumption frequency

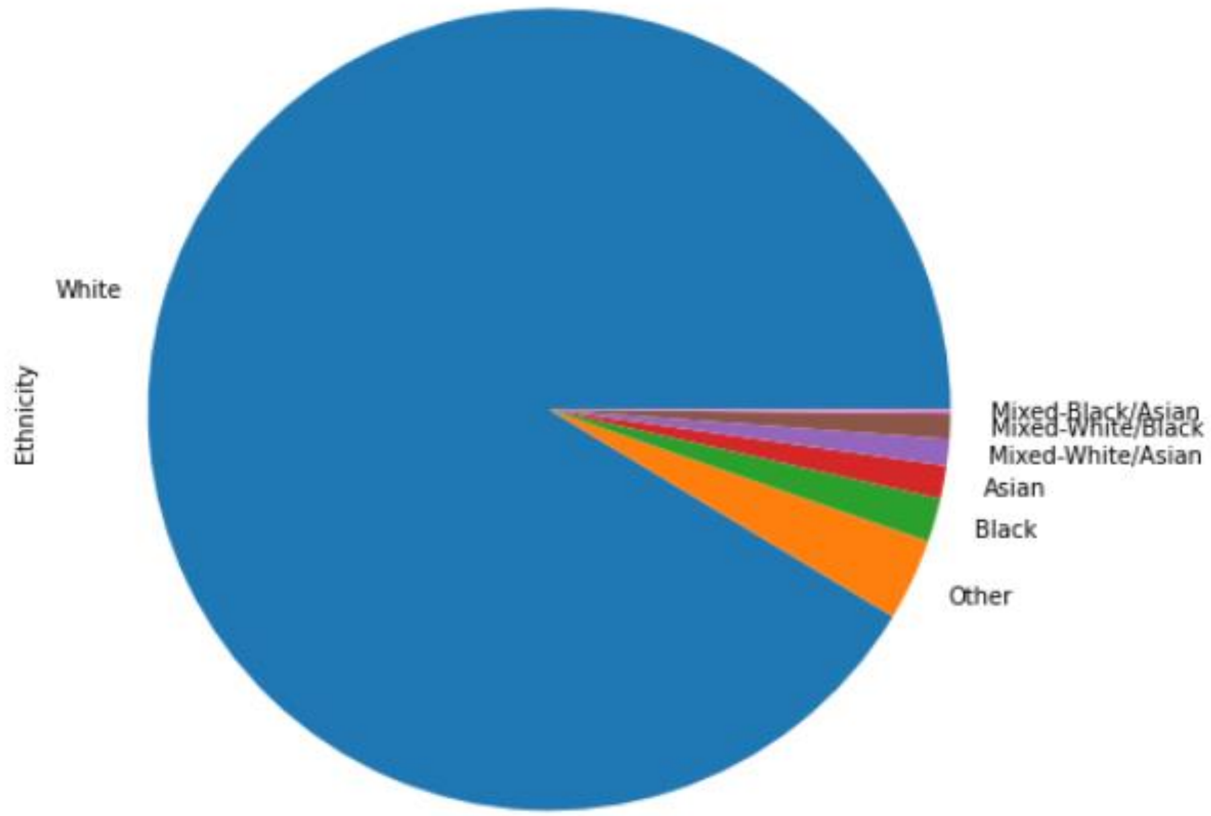
ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS	Alcohol	Amphet	Amyl	Benzos
1	35 - 44	Female	Professional diploma	UK	Mixed-White/Asian	0.31287	-0.57545	-0.58331	-0.91699	-0.00665	-0.21712	-1.18084	5	2	0	2
2	25 - 34	Male	Doctorate degree	UK	White	-0.67825	1.93886	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

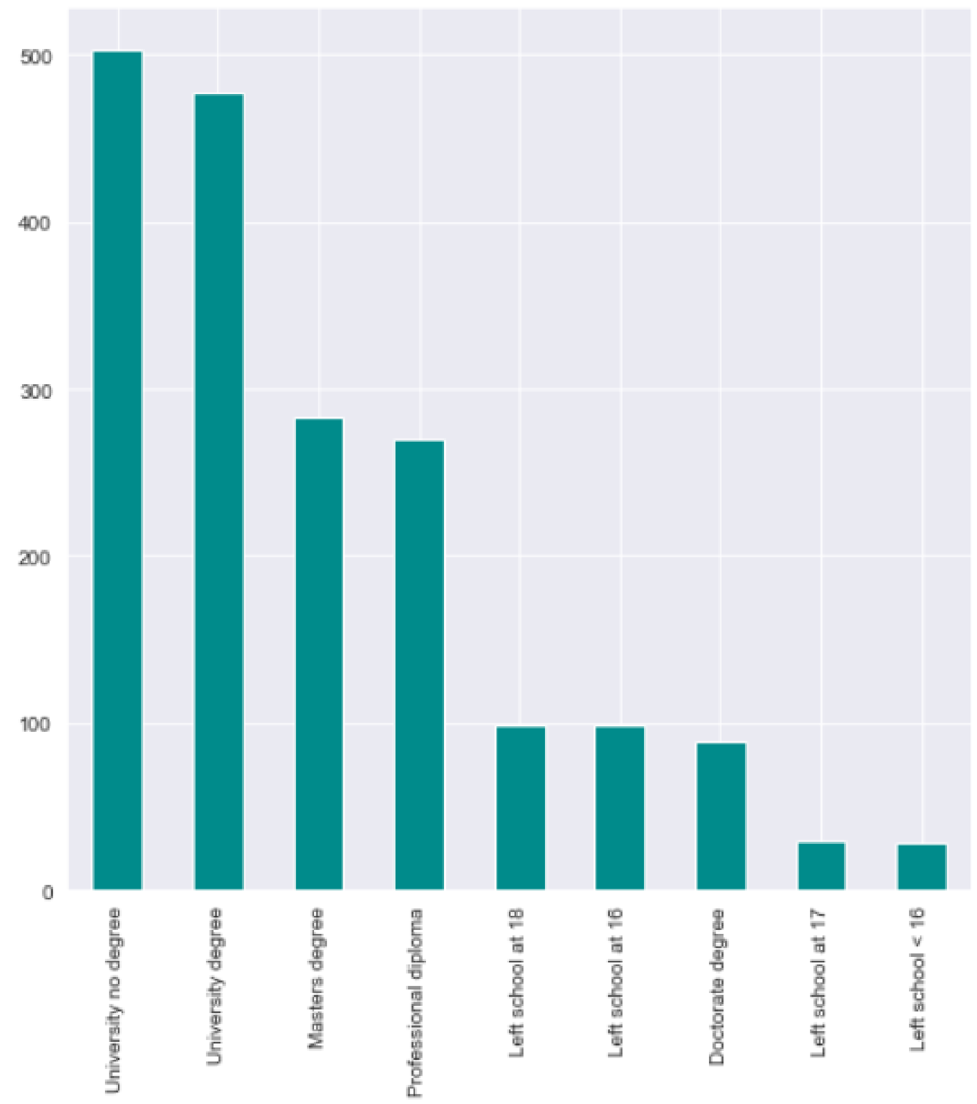
02
Data exploration

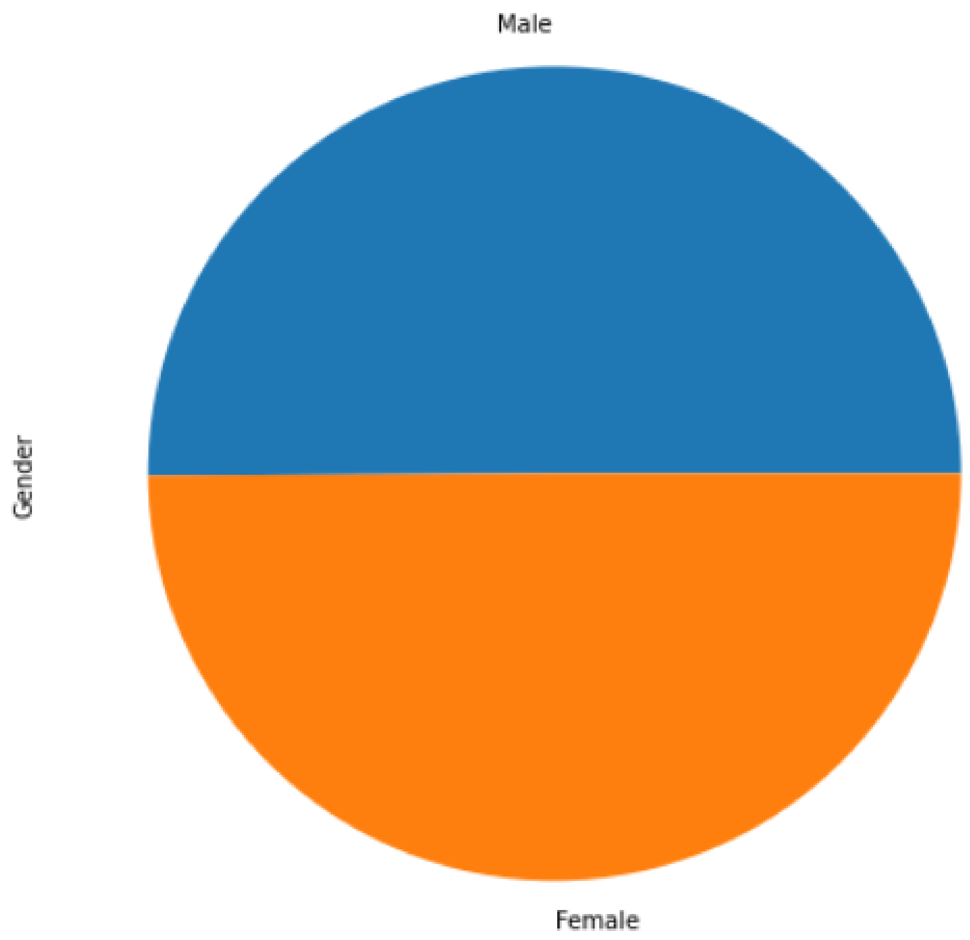
Is there some features more relevant than the other ?







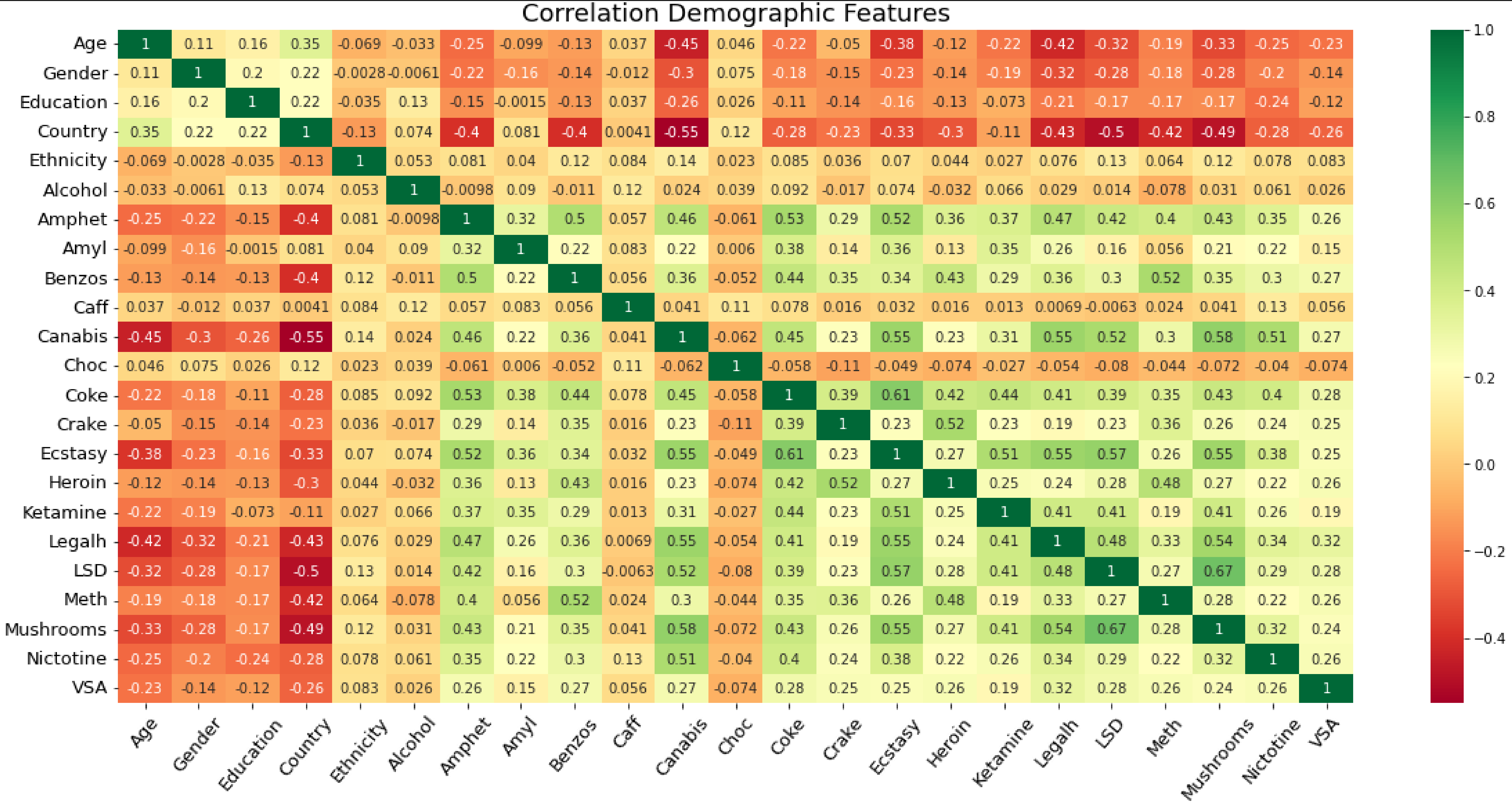


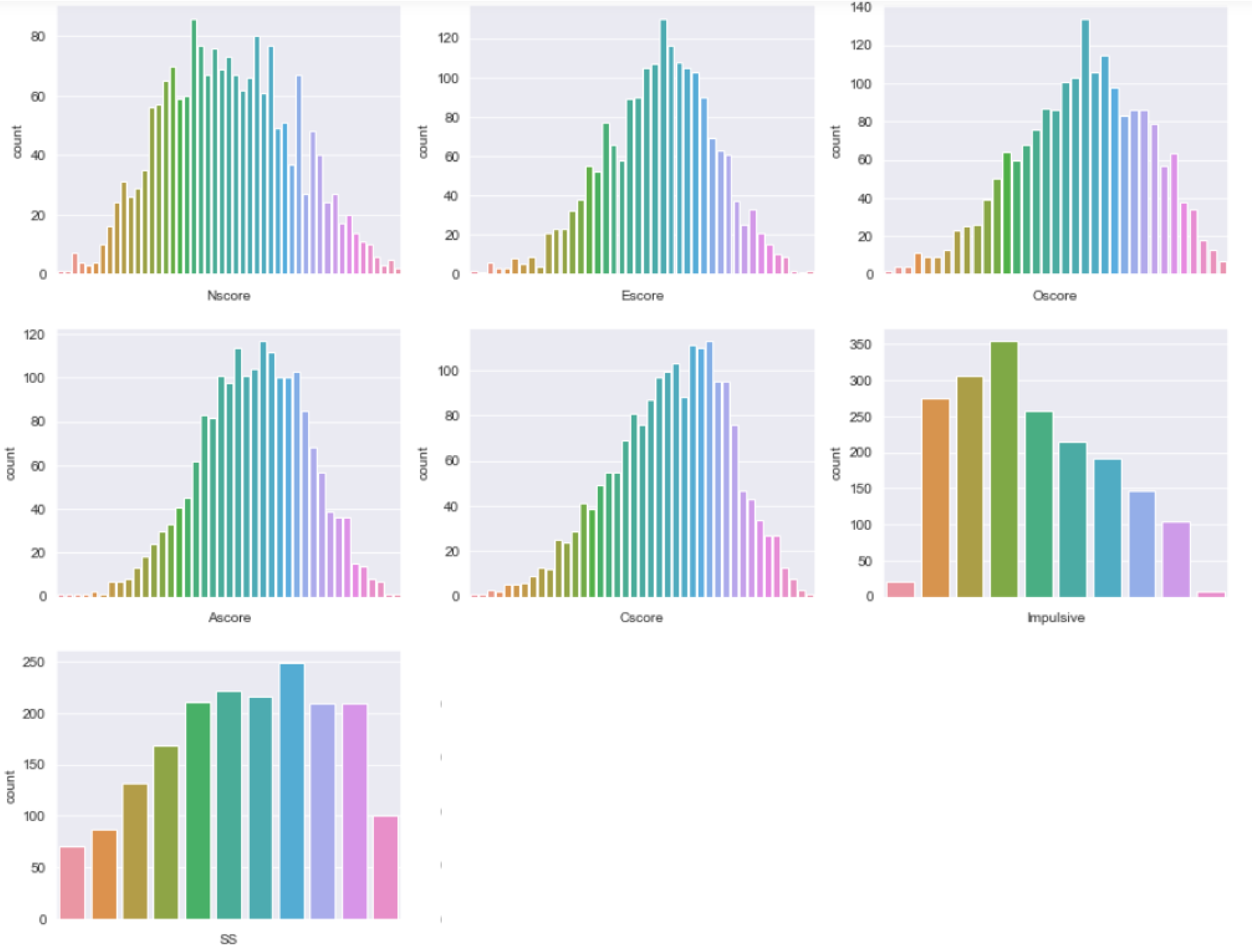


Proportion

Demographic

Correlation :
drugs and demographic.

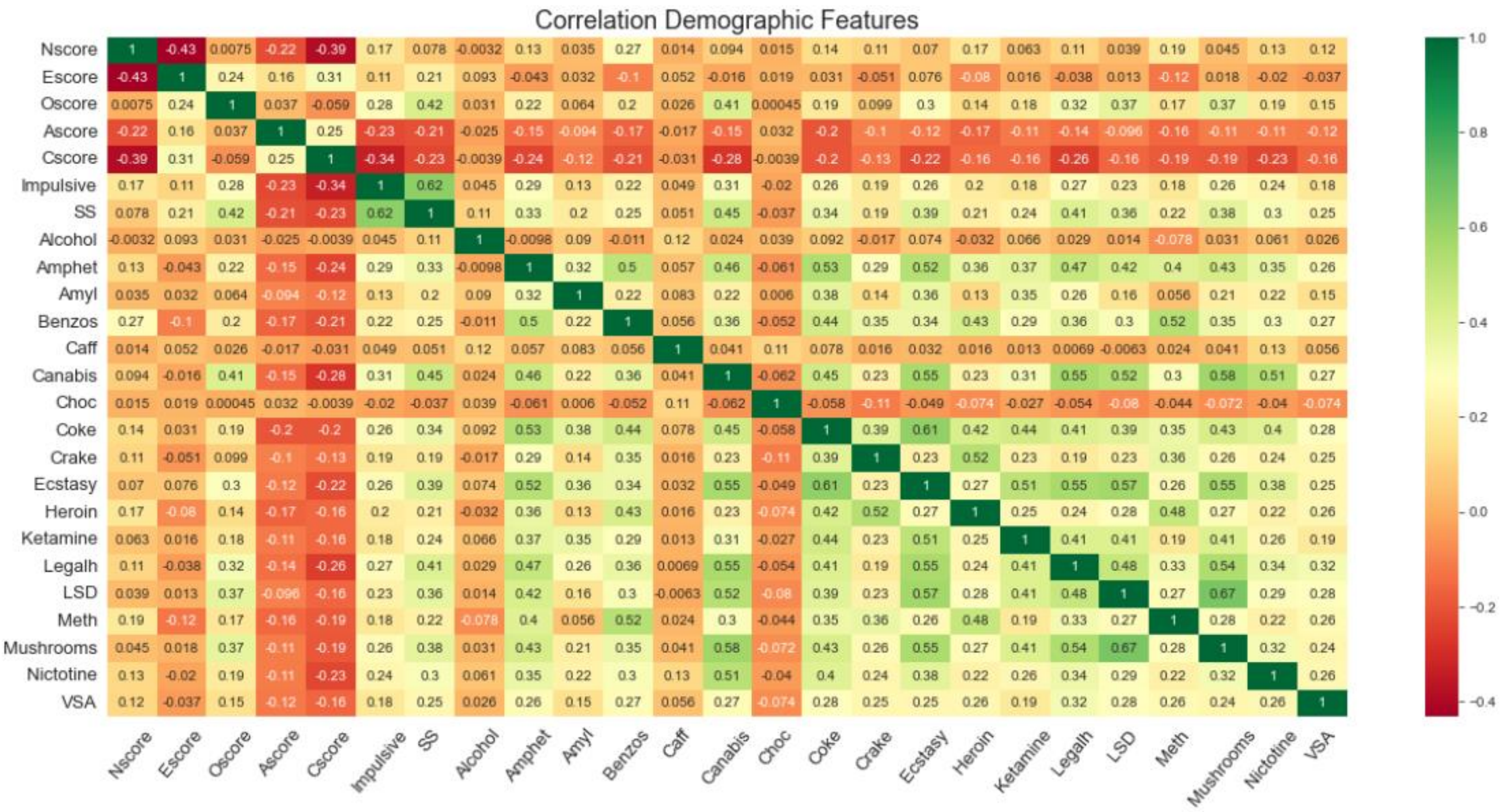




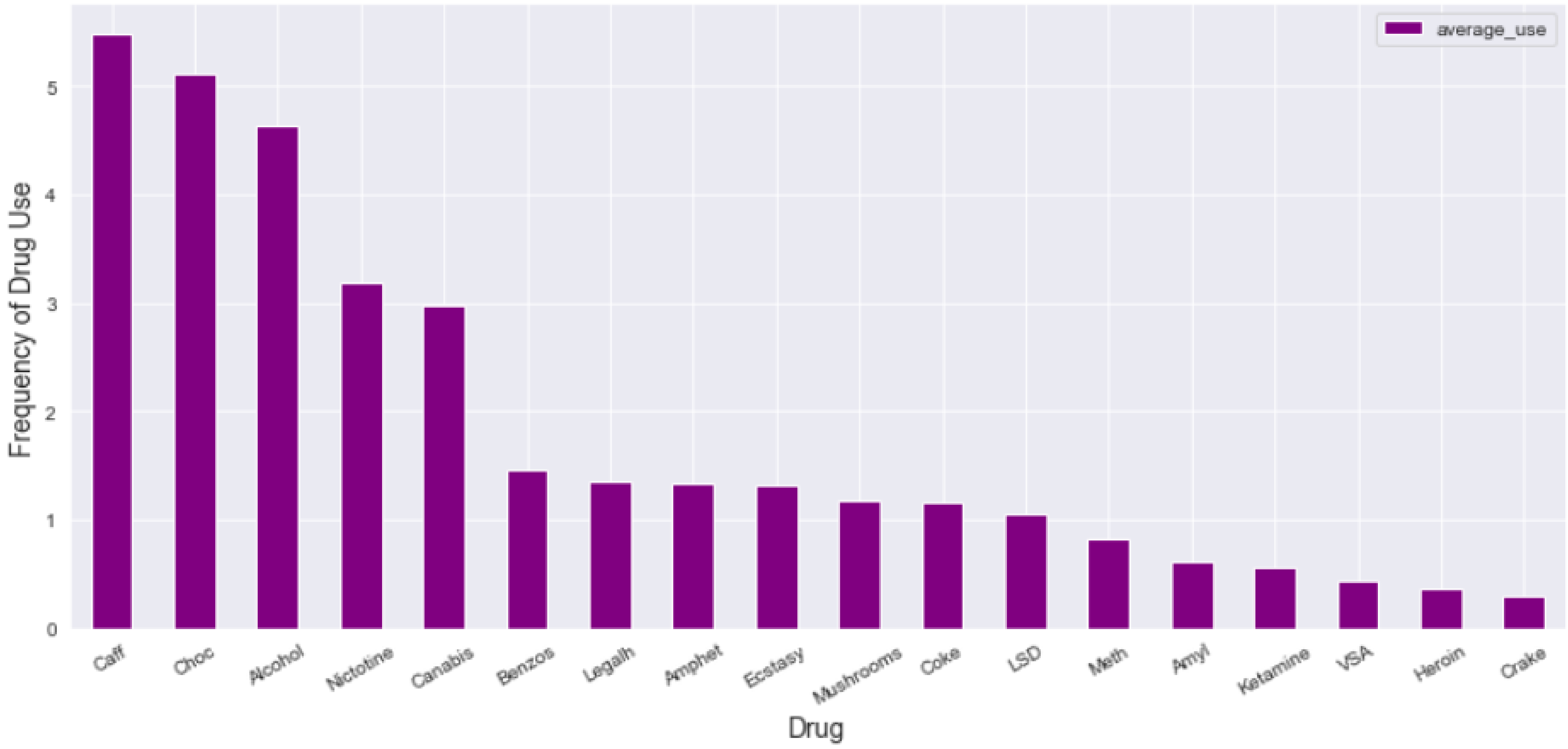
Proportion

Personality features.

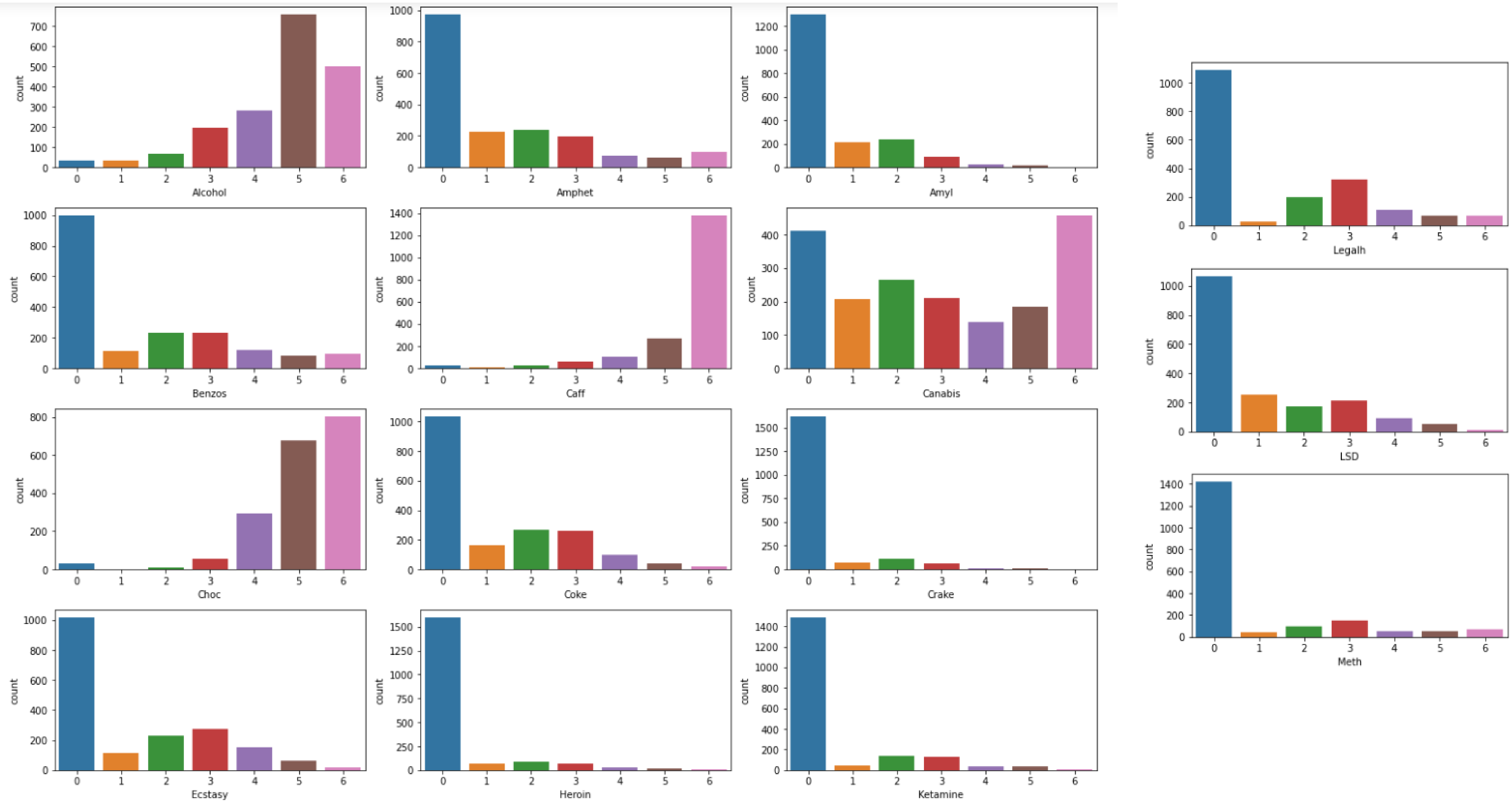
Correlation :
drugs and personality.



Average use of each drug



Consumption of Drug by Different Classes



03
Data modeling

Can we predict if you are a drug consumer ?

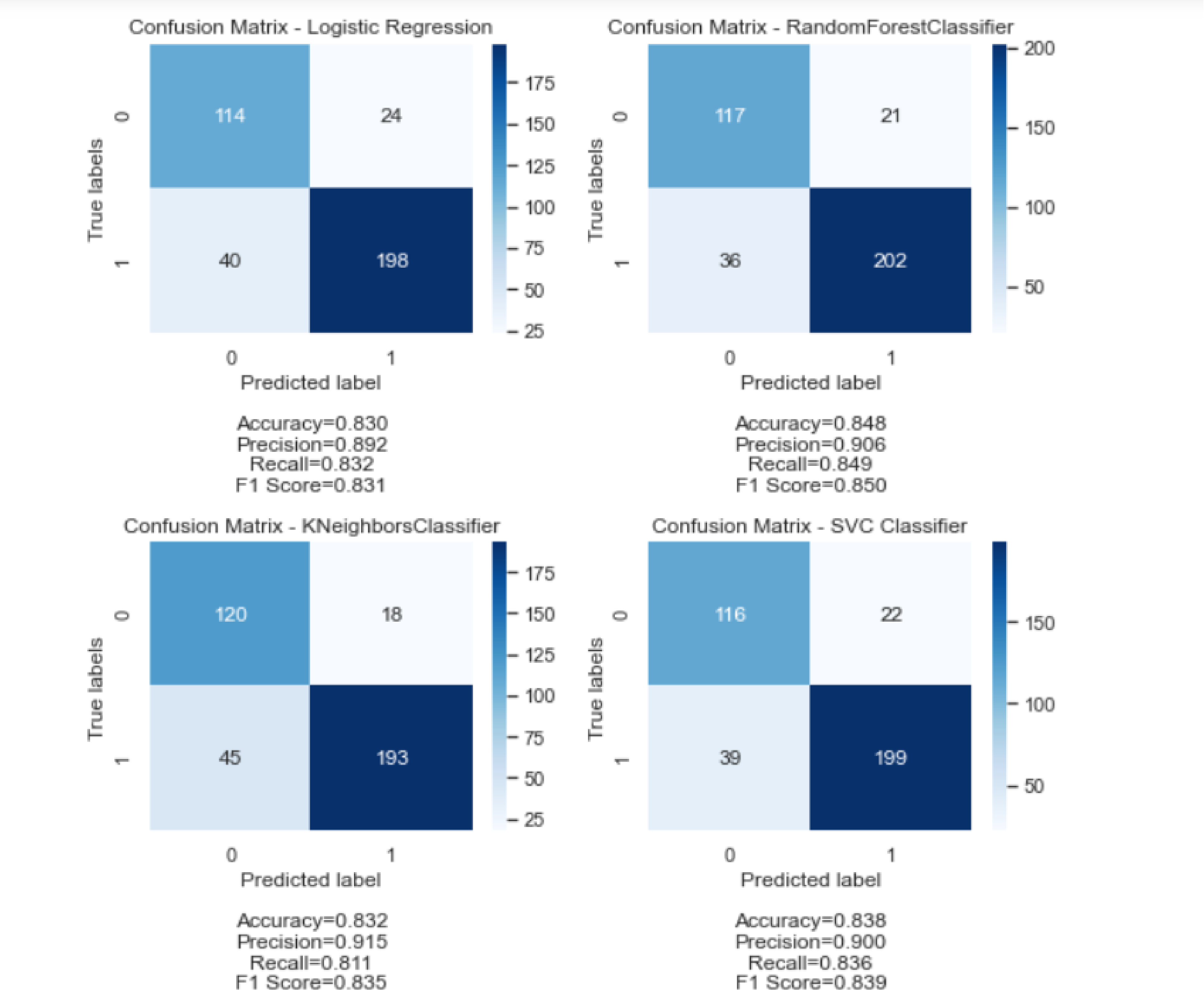
Algorithms

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

```
                    ACCURACY
      Logisitc Regression Accuracy: 82.98%
  Support Vector Machines Accuracy: 83.78%
Random Forest Classifier Accuracy: 84.84%
            KNN Classifier Accuracy: 83.24%
---------------------------------------------
                    F1 SCORES
      Logisitc 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



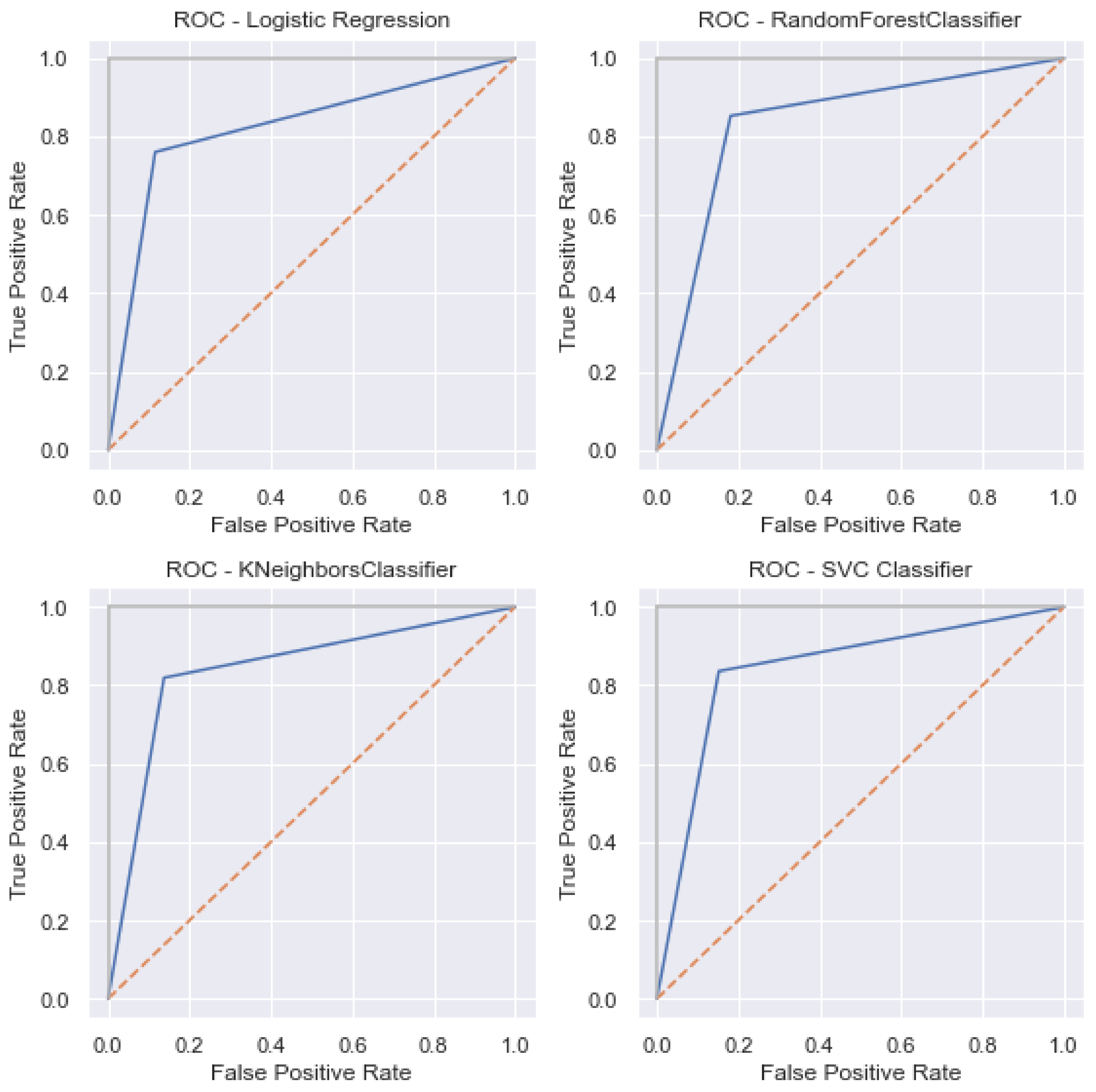
Grid search

```
______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_esti
mators': 256}
 F1-Score: 0.85224
Wall time: 1min 12s
```

ROC curve :

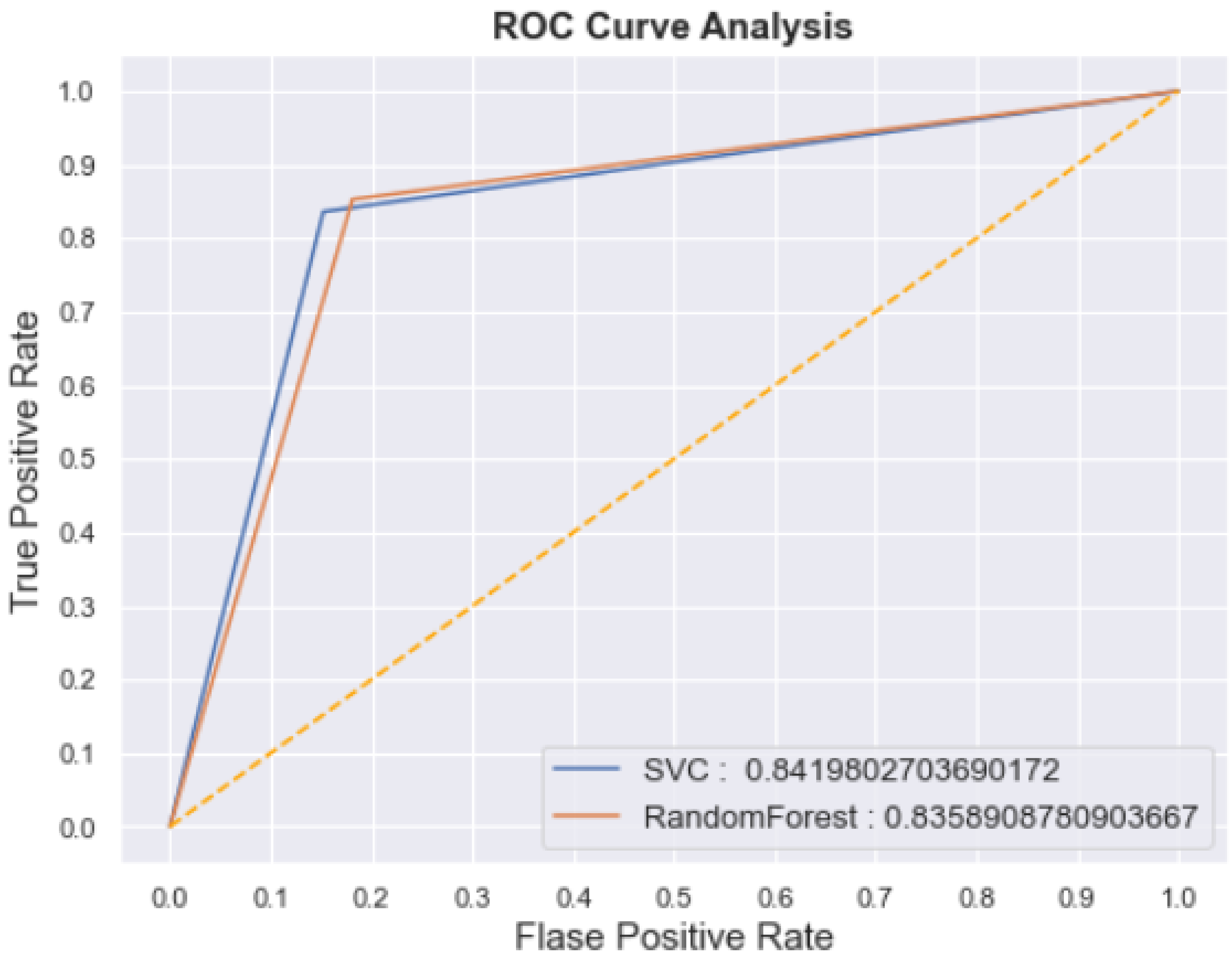
AUC Score

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



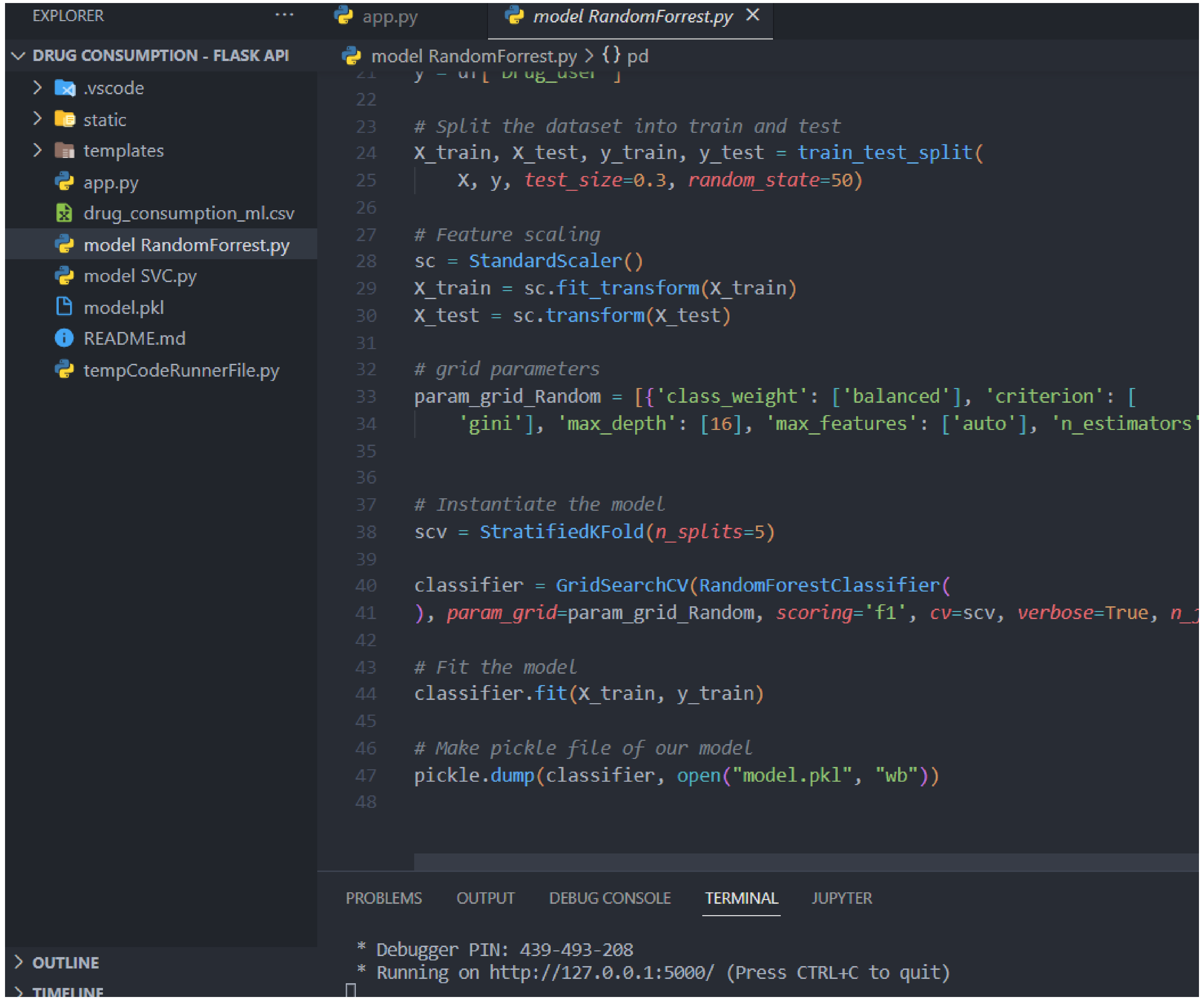
Final model choice :

SVC



04
API

How can we deploy our model ?


EXPLORER
app.py
model RandomForrest.py
DRUG CONSUMPTION - FLASK API
.vscode
static
templates
app.py
drug_consumption_ml.csv
model RandomForrest.py
model SVC.py
model.pkl
README.md
tempCodeRunnerFile.py
model RandomForrest.py > {} pd
Split the dataset into train and test
X_train, X_test, y_train, y_test = train_test_split(
X, y, test_size=0.3, random_state=50)
# Feature scaling
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
# grid parameters
param_grid_Random = [{'class_weight': ['balanced'], 'criterion': [
'gini'], 'max_depth': [16], 'max_features': ['auto'], 'n_estimators'
# Instantiate the model
scv = StratifiedKFold(n_splits=5)
classifier = GridSearchCV(RandomForestClassifier(
), param_grid=param_grid_Random, scoring='f1', cv=scv, verbose=True, n_
# Fit the model
classifier.fit(X_train, y_train)
# Make pickle file of our model
pickle.dump(classifier, open("model.pkl", "wb"))
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
JUPYTER
* Debugger PIN: 439-493-208
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
OUTLINE
TIMELINE

Flask API

API Exemple

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

Neuroticism High

Extraversion High

Openness to experience High

Agreeableness Medium

Conscientiousness Medium

Impulsiveness High

Sensation seeking High

Fill out more information here :

Age 25 - 34

Gender Male

Education University Degree

Country UK

Ethnicity White

Predict

You have high risk of drug consumption :(

Thank you !

JANANY JEGATHEESWARAN

ILIES GOURRI