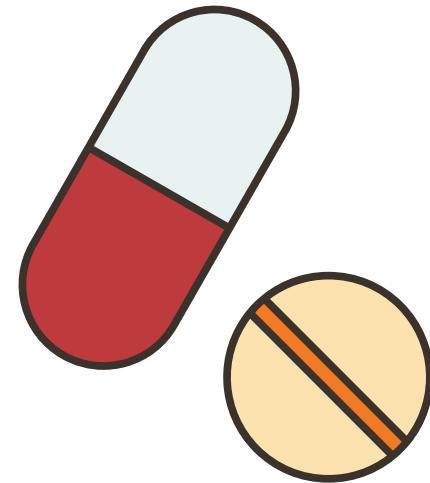


# DRUG CONSUMPTION PREDICTION & EDA

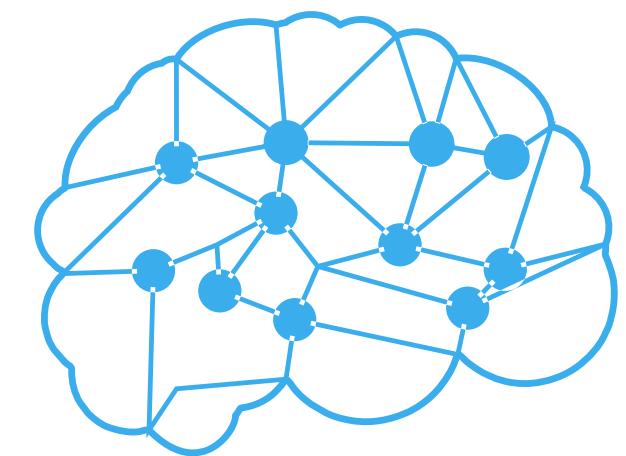
by EMRE ÖZYÜREK

# Data Set Overview

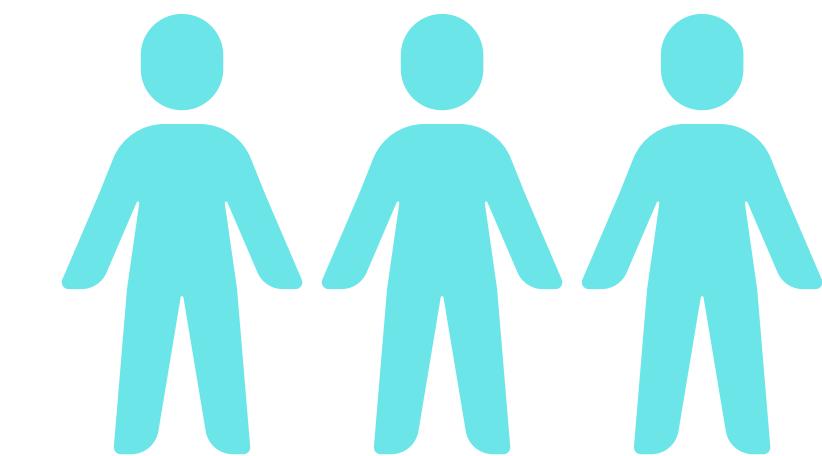
## Drug Consumption Data Set



19 drugs were examined  
and  
measured in terms of  
frequency of use



7  
Personality measurements

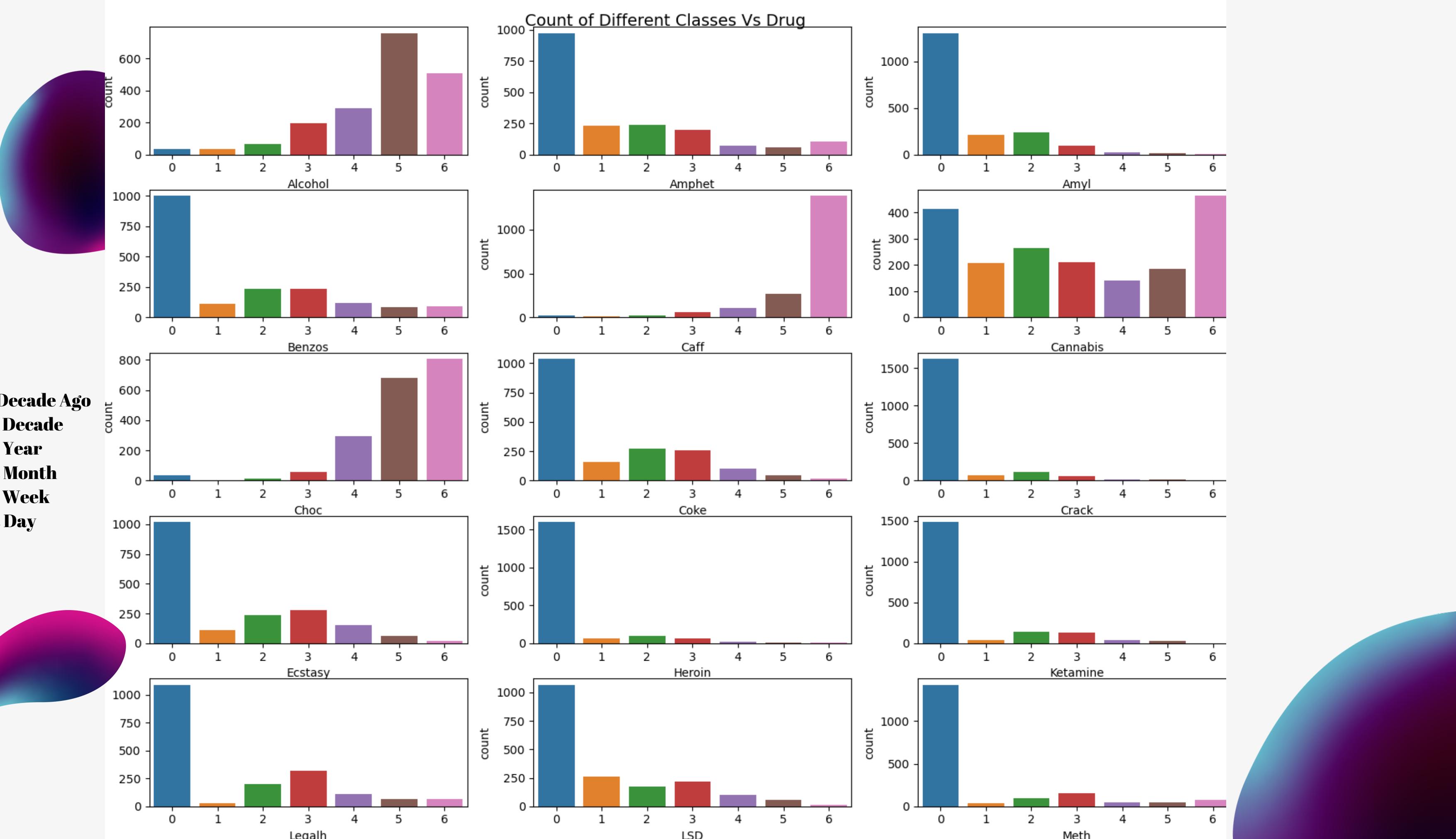


Database contains records  
for 1884  
respondents  
.



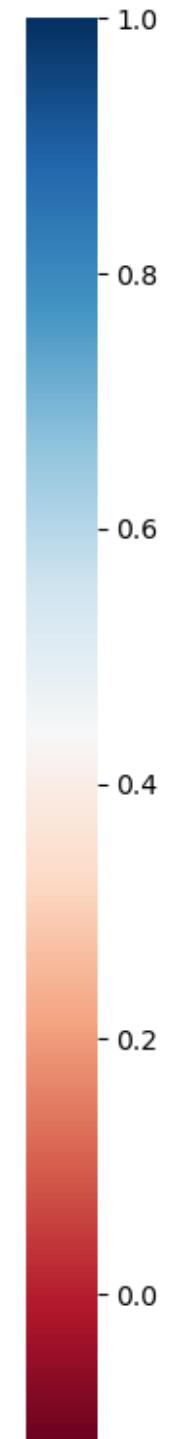
Other Features (Age, Gender,  
Education, Country of  
residence and Ethnicity)

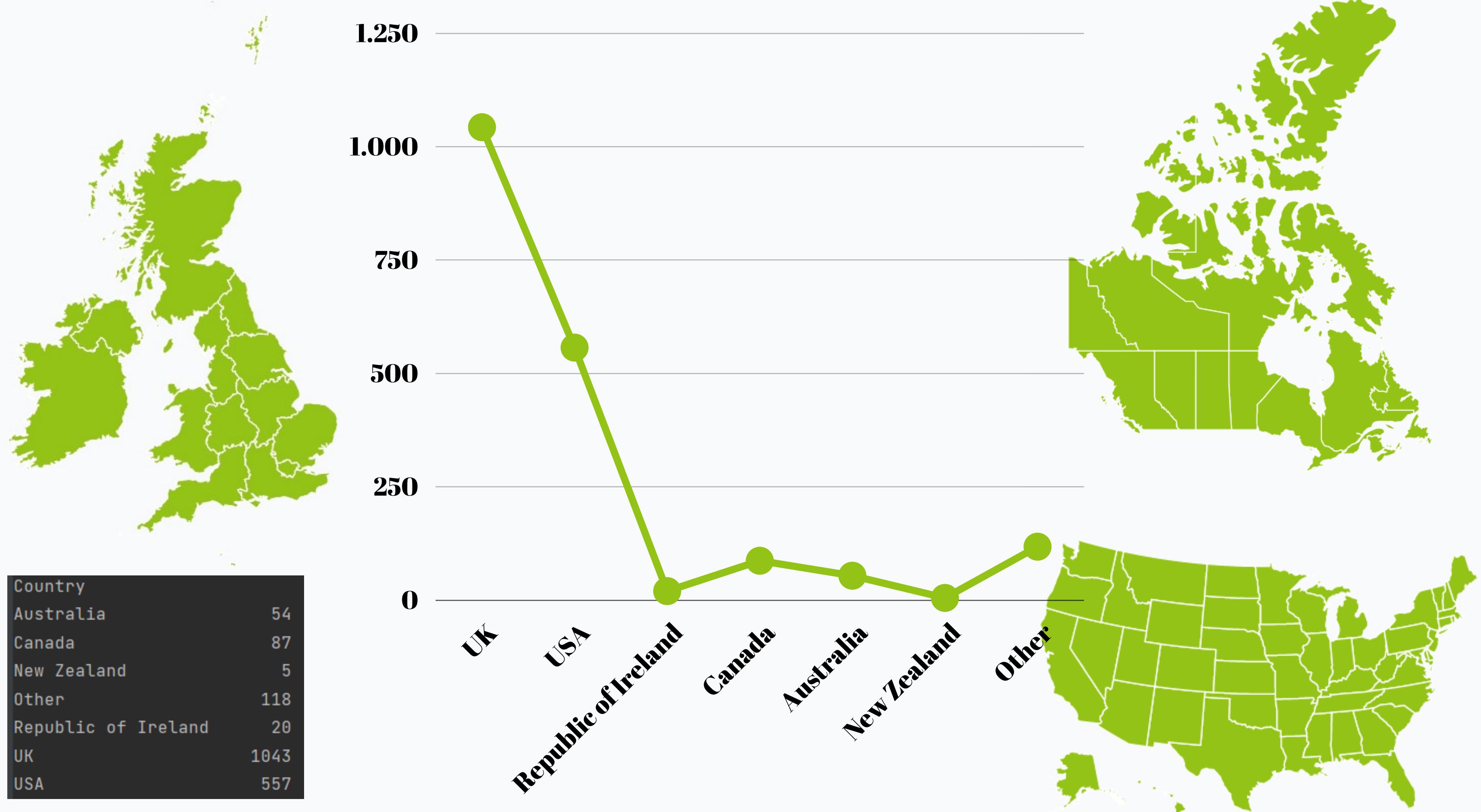
Participants were questioned concerning their use of 18 legal and illegal drugs -- alcohol, amphetamines, amyl nitrite, benzodiazepine, cannabis, chocolate, cocaine, caffeine, crack, ecstasy, heroin, ketamine, legal highs, LSD, methadone, mushrooms, nicotine and volatile substance abuse and one fictitious drug (Semeron) which was introduced to identify over-claimers.



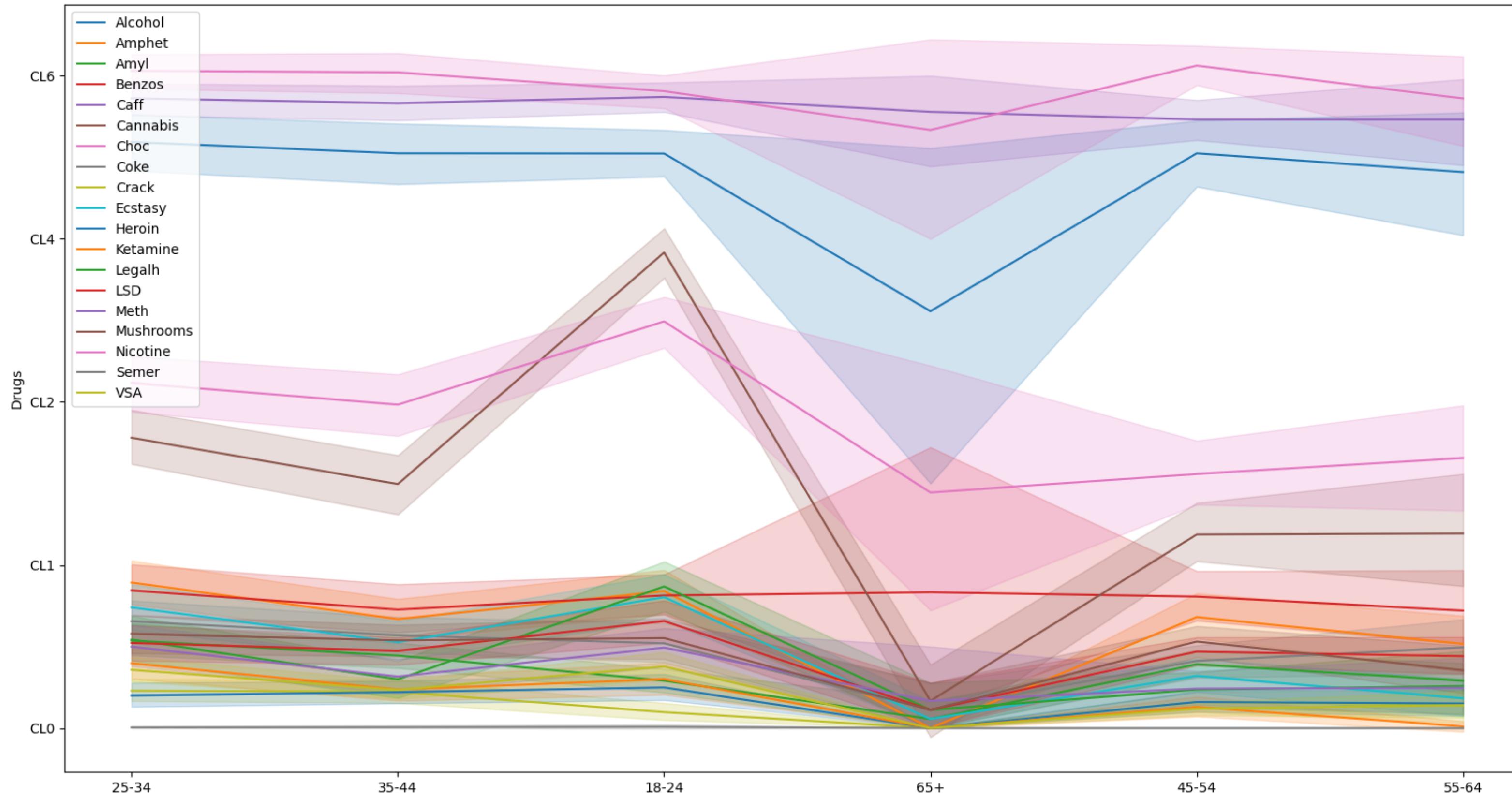
# Correlation of Drugs

|           | Alcohol | Amyl    | Amphet | Benzos | Caff    | Cannabis | Coke   | Crack  | Ecstasy | Heroin | Ketamine | Legalh | LSD    | Meth    | Mushrooms | Nicotine | VSA    |       |
|-----------|---------|---------|--------|--------|---------|----------|--------|--------|---------|--------|----------|--------|--------|---------|-----------|----------|--------|-------|
| Alcohol   | 1       | 0.09    | -0.022 | -0.018 | 0.12    | 0.037    | 0.0089 | 0.084  | -0.018  | 0.065  | -0.036   | 0.058  | 0.017  | 0.0031  | -0.082    | 0.018    | 0.049  | 0.026 |
| Amyl      | 0.09    | 1       | 0.3    | 0.21   | 0.08    | -0.0015  | 0.2    | 0.37   | 0.14    | 0.35   | 0.12     | 0.34   | 0.25   | 0.15    | 0.048     | 0.2      | 0.22   | 0.15  |
| Amphet    | -0.022  | 0.3     | 1      | 0.49   | 0.052   | -0.064   | 0.43   | 0.5    | 0.28    | 0.49   | 0.34     | 0.35   | 0.45   | 0.39    | 0.38      | 0.4      | 0.33   | 0.25  |
| Benzos    | -0.018  | 0.21    | 0.49   | 1      | 0.049   | -0.055   | 0.34   | 0.42   | 0.33    | 0.32   | 0.41     | 0.28   | 0.35   | 0.28    | 0.51      | 0.33     | 0.29   | 0.25  |
| Caff      | 0.12    | 0.08    | 0.052  | 0.049  | 1       | 0.11     | 0.034  | 0.077  | 0.014   | 0.035  | 0.017    | 0.014  | 0.0011 | -0.0079 | 0.024     | 0.037    | 0.12   | 0.054 |
| Choc      | 0.037   | -0.0015 | -0.064 | -0.055 | 0.11    | 1        | -0.067 | -0.067 | -0.12   | -0.06  | -0.077   | -0.029 | -0.059 | -0.082  | -0.045    | -0.079   | -0.044 | -0.08 |
| Cannabis  | 0.0089  | 0.2     | 0.43   | 0.34   | 0.034   | -0.067   | 1      | 0.43   | 0.23    | 0.53   | 0.22     | 0.29   | 0.52   | 0.5     | 0.28      | 0.56     | 0.49   | 0.26  |
| Coke      | 0.084   | 0.37    | 0.5    | 0.42   | 0.077   | -0.067   | 0.43   | 1      | 0.38    | 0.59   | 0.42     | 0.42   | 0.39   | 0.36    | 0.33      | 0.4      | 0.39   | 0.26  |
| Crack     | -0.018  | 0.14    | 0.28   | 0.33   | 0.014   | -0.12    | 0.23   | 0.38   | 1       | 0.22   | 0.52     | 0.22   | 0.18   | 0.21    | 0.35      | 0.25     | 0.24   | 0.23  |
| Ecstasy   | 0.065   | 0.35    | 0.49   | 0.32   | 0.035   | -0.06    | 0.53   | 0.59   | 0.22    | 1      | 0.25     | 0.49   | 0.53   | 0.56    | 0.24      | 0.52     | 0.37   | 0.23  |
| Heroin    | -0.036  | 0.12    | 0.34   | 0.41   | 0.017   | -0.077   | 0.22   | 0.42   | 0.52    | 0.25   | 1        | 0.24   | 0.23   | 0.26    | 0.46      | 0.26     | 0.22   | 0.24  |
| Ketamine  | 0.058   | 0.34    | 0.35   | 0.28   | 0.014   | -0.029   | 0.29   | 0.42   | 0.22    | 0.49   | 0.24     | 1      | 0.4    | 0.4     | 0.17      | 0.4      | 0.25   | 0.18  |
| Legalh    | 0.017   | 0.25    | 0.45   | 0.35   | 0.0011  | -0.059   | 0.52   | 0.39   | 0.18    | 0.53   | 0.23     | 0.4    | 1      | 0.45    | 0.31      | 0.5      | 0.32   | 0.3   |
| LSD       | 0.0031  | 0.15    | 0.39   | 0.28   | -0.0079 | -0.082   | 0.5    | 0.36   | 0.21    | 0.56   | 0.26     | 0.4    | 0.45   | 1       | 0.24      | 0.65     | 0.28   | 0.28  |
| Meth      | -0.082  | 0.048   | 0.38   | 0.51   | 0.024   | -0.045   | 0.28   | 0.33   | 0.35    | 0.24   | 0.46     | 0.17   | 0.31   | 0.24    | 1         | 0.25     | 0.21   | 0.24  |
| Mushrooms | 0.018   | 0.2     | 0.4    | 0.33   | 0.037   | -0.079   | 0.56   | 0.4    | 0.25    | 0.52   | 0.26     | 0.4    | 0.5    | 0.65    | 0.25      | 1        | 0.31   | 0.23  |
| Nicotine  | 0.049   | 0.22    | 0.33   | 0.29   | 0.12    | -0.044   | 0.49   | 0.39   | 0.24    | 0.37   | 0.22     | 0.25   | 0.32   | 0.28    | 0.21      | 0.31     | 1      | 0.24  |
| VSA       | 0.026   | 0.15    | 0.25   | 0.25   | 0.054   | -0.08    | 0.26   | 0.26   | 0.23    | 0.23   | 0.24     | 0.18   | 0.3    | 0.28    | 0.24      | 0.23     | 0.24   | 1     |





Drugs vs Age



# Personality measurements

**Nscore:** *Neuroticism* score

**Escore:** *Extraversion* score

**Oscore:** *Openness* to experience score

**Ascore:** *Agreeableness* score

**Cscore:** *Conscientiousness* score

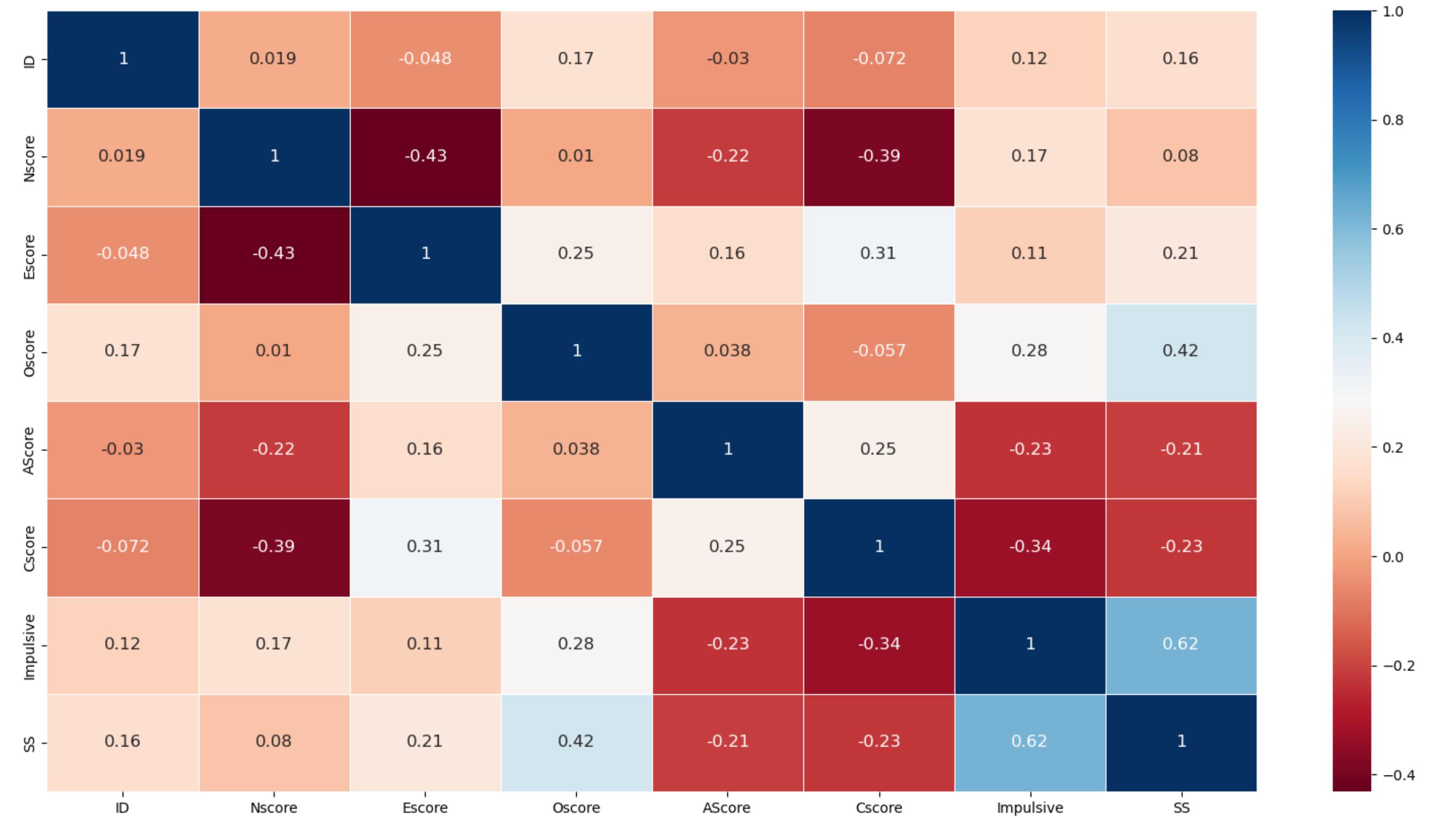
**Impulsive:** *impulsiveness* score

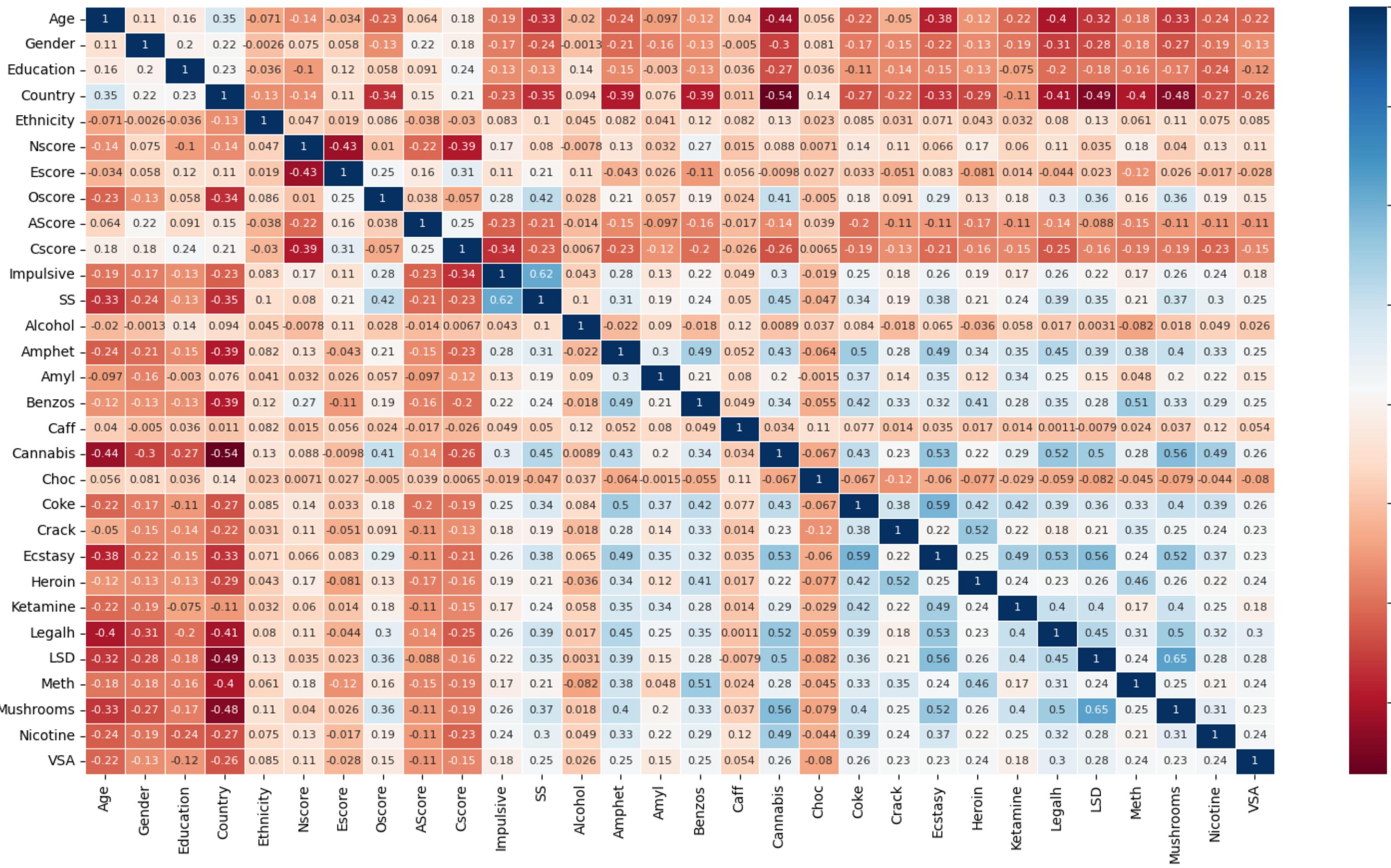
**SS:** *Impulsive Sensation Seeking* score

the profile of users of illicit drugs has a characteristic form:

N ↑, O ↑, A ↓, C ↓, Imp ↑, SS ↑

# Numeric Columns Correlation



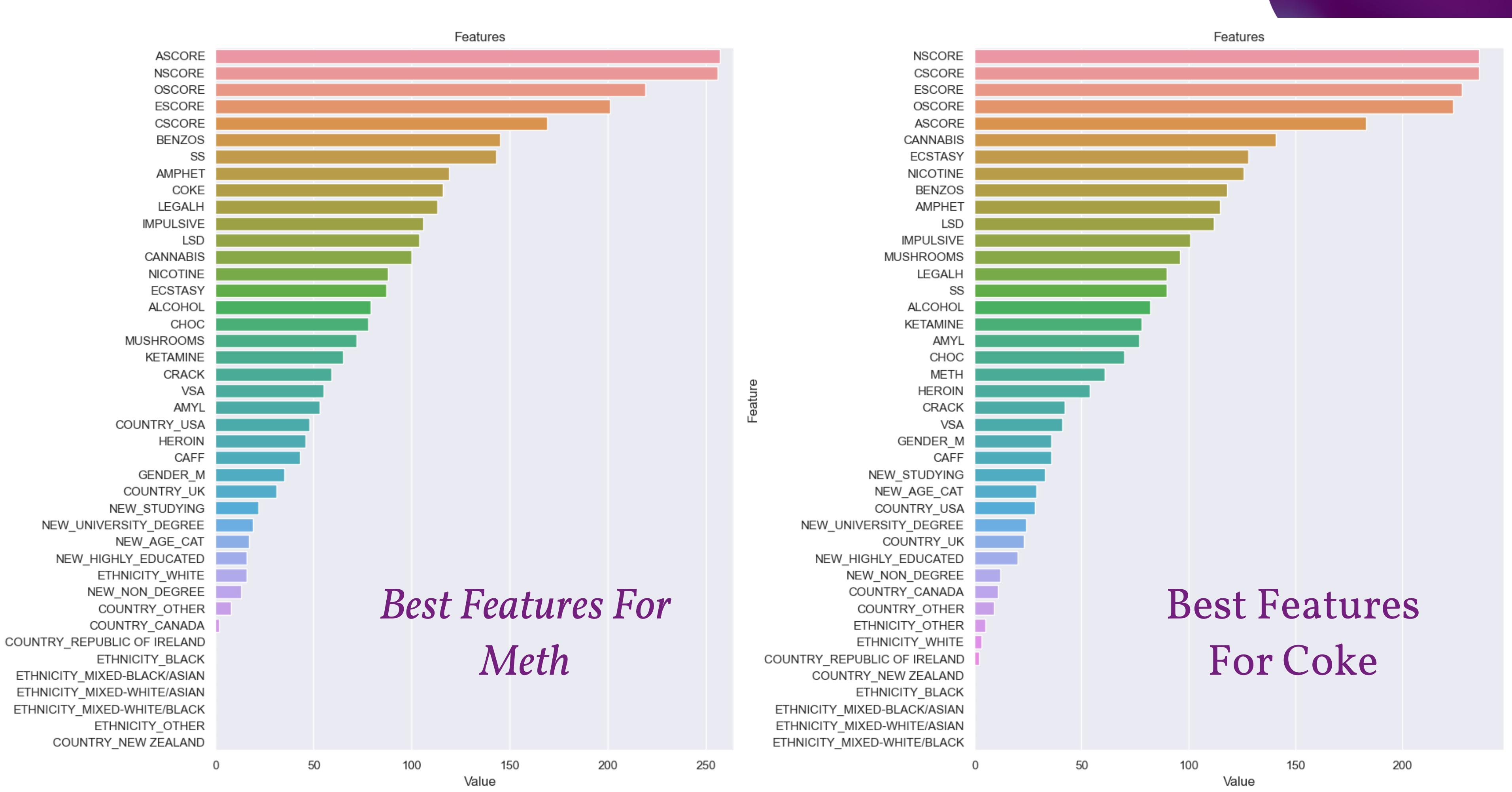


# ML

## Feature Engineering Effects *Before - After*

```
In [27]: base_models(X, y, scoring="accuracy")
Base Models....  
accuracy: 0.8161 (LR)
accuracy: 0.8065 (KNN)
accuracy: 0.8454 (SVC)
accuracy: 0.7884 (CART)
accuracy: 0.8561 (RF)
accuracy: 0.8443 (Adaboost)
accuracy: 0.8475 (GBM)
accuracy: 0.8438 (XGBoost)
accuracy: 0.847 (LightGBM)
```

```
In [32]: base_models(X, y, scoring="accuracy")
Base Models....  
accuracy: 0.823 (LR)
accuracy: 0.8097 (KNN)
accuracy: 0.8342 (SVC)
accuracy: 0.7649 (CART)
accuracy: 0.8507 (RF)
accuracy: 0.8342 (Adaboost)
accuracy: 0.8427 (GBM)
accuracy: 0.8374 (XGBoost)
accuracy: 0.8374 (LightGBM)
```



# Hyperparameter Optimization

```
In [43]: best_models = hyperparameter_optimization(X, y, cv=5, scoring="accuracy")

Hyperparameter Optimization.....
##### KNN #####
accuracy (Before): 0.8097
accuracy (After): 0.8129
KNN best params: {'n_neighbors': 7}

##### CART #####
accuracy (Before): 0.8294
accuracy (After): 0.8374
CART best params: {'max_depth': 3, 'min_samples_split': 2}

##### RF #####
accuracy (Before): 0.8513
accuracy (After): 0.8497
RF best params: {'max_depth': 15, 'max_features': 5, 'min_samples_split': 15, 'n_estimators': 300}

##### XGBoost #####
accuracy (Before): 0.8347
accuracy (After): 0.8422
XGBoost best params: {'learning_rate': 0.1, 'max_depth': 8, 'n_estimators': 100}

##### LightGBM #####
accuracy (Before): 0.8406
accuracy (After): 0.8417
LightGBM best params: {'learning_rate': 0.01, 'n_estimators': 300}
```