# PYTHON PROJECT 4A DRUG CONSUMPTION

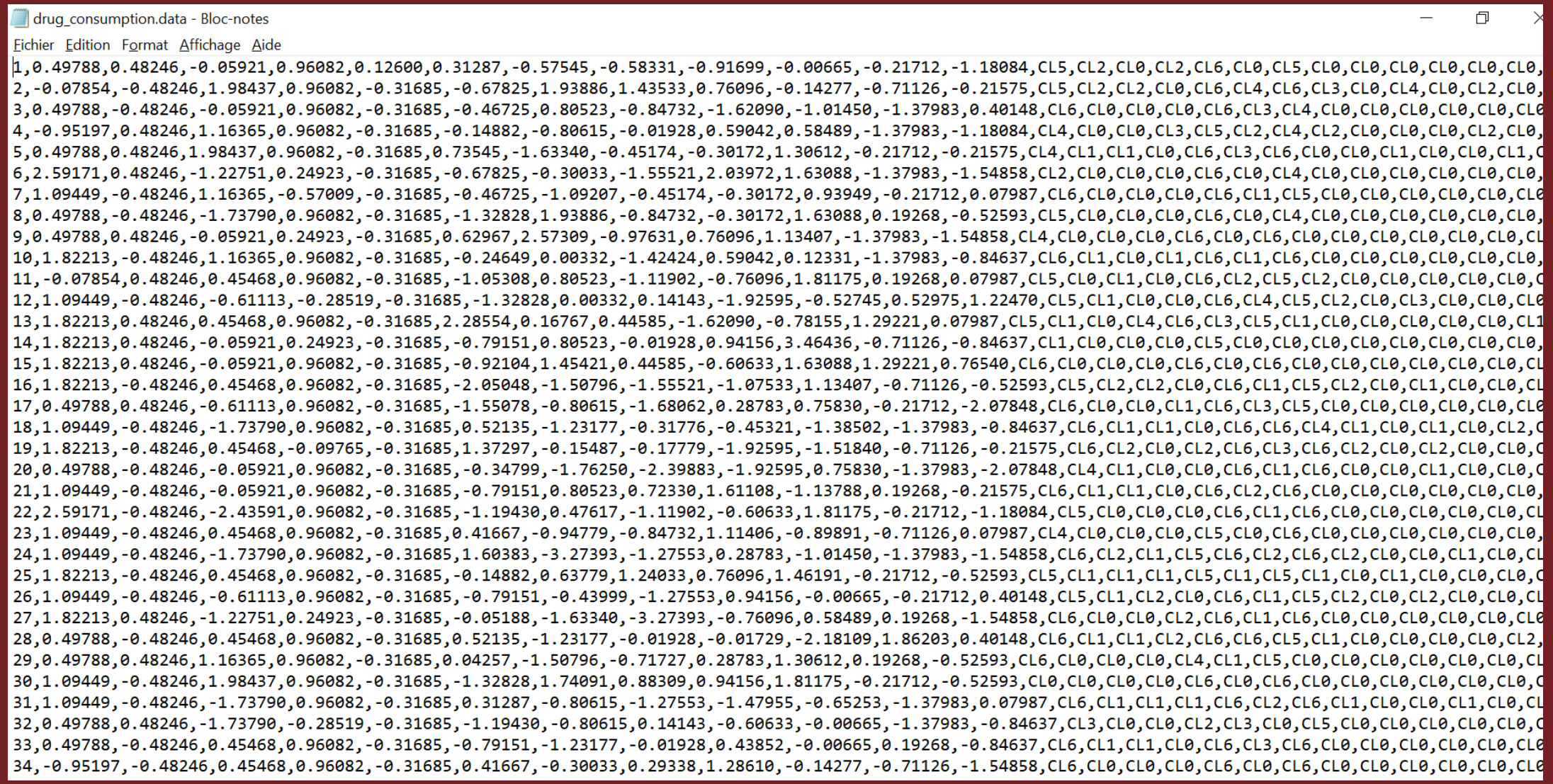
*Jules griaud*

# DO YOUR PERSONALITY TRAITS INDICATE HOW MANY DRUGS YOU ARE LIKLY GOING TO TRY ONCE IN YOUR LIFE?

- This data set has 32 columns, 19 of them being the different drugs that people have tried or not and the 6 of them being the personality scores of each individual and the rest being variables about their country of origin, age , ethnicity ect. Having seen this, I put out a hypothesis that there has to being a certain personality score that leeds people to try more drugs. In this study i will be looking at the comparaison of the personality scores to the number of drugs an individual has tested to see if I can predict how many drugs an individual has tried in his life.

# CLEANING OF THE DATA SET

# BEFORE:


drug_consumption.data - Bloc-notes

Fichier Edition Format Affichage Aide

```
1,0.49788,0.48246,-0.05921,0.96082,0.12600,0.31287,-0.57545,-0.58331,-0.91699,-0.00665,-0.21712,-1.18084,CL5,CL2,CL0,CL2,CL6,CL0,CL5,CL0,CL0,CL0,CL0,CL0,CL0,
2,-0.07854,-0.48246,1.98437,0.96082,-0.31685,-0.67825,1.93886,1.43533,0.76096,-0.14277,-0.71126,-0.21575,CL5,CL2,CL2,CL0,CL6,CL4,CL6,CL3,CL0,CL4,CL0,CL2,CL0,
3,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.46725,0.80523,-0.84732,-1.62090,-1.01450,-1.37983,0.40148,CL6,CL0,CL0,CL0,CL6,CL3,CL4,CL0,CL0,CL0,CL0,CL0,CL0
4,-0.95197,0.48246,1.16365,0.96082,-0.31685,-0.14882,-0.80615,-0.01928,0.59042,0.58489,-1.37983,-1.18084,CL4,CL0,CL0,CL3,CL5,CL2,CL4,CL2,CL0,CL0,CL0,CL2,CL0,
5,0.49788,0.48246,1.98437,0.96082,-0.31685,0.73545,-1.63340,-0.45174,-0.30172,1.30612,-0.21712,-0.21575,CL4,CL1,CL1,CL0,CL6,CL3,CL6,CL0,CL0,CL1,CL0,CL0,CL1,C
6,2.59171,0.48246,-1.22751,0.24923,-0.31685,-0.67825,-0.30033,-1.55521,2.03972,1.63088,-1.37983,-1.54858,CL2,CL0,CL0,CL0,CL6,CL0,CL4,CL0,CL0,CL0,CL0,CL0,CL0,
7,1.09449,-0.48246,1.16365,-0.57009,-0.31685,-0.46725,-1.09207,-0.45174,-0.30172,0.93949,-0.21712,0.07987,CL6,CL0,CL0,CL0,CL6,CL1,CL5,CL0,CL0,CL0,CL0,CL0,CL0
8,0.49788,-0.48246,-1.73790,0.96082,-0.31685,-1.32828,1.93886,-0.84732,-0.30172,1.63088,0.19268,-0.52593,CL5,CL0,CL0,CL0,CL6,CL0,CL4,CL0,CL0,CL0,CL0,CL0,CL0,
9,0.49788,0.48246,-0.05921,0.24923,-0.31685,0.62967,2.57309,-0.97631,0.76096,1.13407,-1.37983,-1.54858,CL4,CL0,CL0,CL0,CL6,CL0,CL6,CL0,CL0,CL0,CL0,CL0,CL0,CL
10,1.82213,-0.48246,1.16365,0.96082,-0.31685,-0.24649,0.00332,-1.42424,0.59042,0.12331,-1.37983,-0.84637,CL6,CL1,CL0,CL1,CL6,CL1,CL6,CL0,CL0,CL0,CL0,CL0,CL0,
11,-0.07854,0.48246,0.45468,0.96082,-0.31685,-1.05308,0.80523,-1.11902,-0.76096,1.81175,0.19268,0.07987,CL5,CL0,CL1,CL0,CL6,CL2,CL5,CL2,CL0,CL0,CL0,CL0,CL0,C
12,1.09449,-0.48246,-0.61113,-0.28519,-0.31685,-1.32828,0.00332,0.14143,-1.92595,-0.52745,0.52975,1.22470,CL5,CL1,CL0,CL0,CL6,CL4,CL5,CL2,CL0,CL3,CL0,CL0,CL0
13,1.82213,0.48246,0.45468,0.96082,-0.31685,2.28554,0.16767,0.44585,-1.62090,-0.78155,1.29221,0.07987,CL5,CL1,CL0,CL4,CL6,CL3,CL5,CL1,CL0,CL0,CL0,CL0,CL0,CL1
14,1.82213,0.48246,-0.05921,0.24923,-0.31685,-0.79151,0.80523,-0.01928,0.94156,3.46436,-0.71126,-0.84637,CL1,CL0,CL0,CL0,CL5,CL0,CL0,CL0,CL0,CL0,CL0,CL0,CL0,
15,1.82213,0.48246,-0.05921,0.96082,-0.31685,-0.92104,1.45421,0.44585,-0.60633,1.63088,1.29221,0.76540,CL6,CL0,CL0,CL0,CL6,CL0,CL6,CL0,CL0,CL0,CL0,CL0,CL0,CL
16,1.82213,-0.48246,0.45468,0.96082,-0.31685,-2.05048,-1.50796,-1.55521,-1.07533,1.13407,-0.71126,-0.52593,CL5,CL2,CL2,CL0,CL6,CL1,CL5,CL2,CL0,CL1,CL0,CL0,CL
17,0.49788,0.48246,-0.61113,0.96082,-0.31685,-1.55078,-0.80615,-1.68062,0.28783,0.75830,-0.21712,-2.07848,CL6,CL0,CL0,CL1,CL6,CL3,CL5,CL0,CL0,CL0,CL0,CL0,CL0
18,1.09449,-0.48246,-1.73790,0.96082,-0.31685,0.52135,-1.23177,-0.31776,-0.45321,-1.38502,-1.37983,-0.84637,CL6,CL1,CL1,CL0,CL6,CL6,CL4,CL1,CL0,CL1,CL0,CL2,C
19,1.82213,-0.48246,0.45468,-0.09765,-0.31685,1.37297,-0.15487,-0.17779,-1.92595,-1.51840,-0.71126,-0.21575,CL6,CL2,CL0,CL2,CL6,CL3,CL6,CL2,CL0,CL2,CL0,CL0,C
20,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.34799,-1.76250,-2.39883,-1.92595,0.75830,-1.37983,-2.07848,CL4,CL1,CL0,CL0,CL6,CL1,CL6,CL0,CL0,CL1,CL0,CL0,C
21,1.09449,-0.48246,-0.05921,0.96082,-0.31685,-0.79151,0.80523,0.72330,1.61108,-1.13788,0.19268,-0.21575,CL6,CL1,CL1,CL0,CL6,CL2,CL6,CL0,CL0,CL0,CL0,CL0,CL0,
22,2.59171,-0.48246,-2.43591,0.96082,-0.31685,-1.19430,0.47617,-1.11902,-0.60633,1.81175,-0.21712,-1.18084,CL5,CL0,CL0,CL0,CL6,CL1,CL6,CL0,CL0,CL0,CL0,CL0,CL
23,1.09449,-0.48246,0.45468,0.96082,-0.31685,0.41667,-0.94779,-0.84732,1.11406,-0.89891,-0.71126,0.07987,CL4,CL0,CL0,CL0,CL5,CL0,CL6,CL0,CL0,CL0,CL0,CL0,CL0,
24,1.09449,-0.48246,-1.73790,0.96082,-0.31685,1.60383,-3.27393,-1.27553,0.28783,-1.01450,-1.37983,-1.54858,CL6,CL2,CL1,CL5,CL6,CL2,CL6,CL2,CL0,CL0,CL1,CL0,CL
25,1.82213,-0.48246,0.45468,0.96082,-0.31685,-0.14882,0.63779,1.24033,0.76096,1.46191,-0.21712,-0.52593,CL5,CL1,CL1,CL1,CL5,CL1,CL5,CL1,CL0,CL1,CL0,CL0,CL0,C
26,1.09449,-0.48246,-0.61113,0.96082,-0.31685,-0.79151,-0.43999,-1.27553,0.94156,-0.00665,-0.21712,0.40148,CL5,CL1,CL2,CL0,CL6,CL1,CL5,CL2,CL0,CL2,CL0,CL0,CL
27,1.82213,0.48246,-1.22751,0.24923,-0.31685,-0.05188,-1.63340,-3.27393,-0.76096,0.58489,0.19268,-1.54858,CL6,CL0,CL0,CL2,CL6,CL1,CL6,CL0,CL0,CL0,CL0,CL0,CL0
28,0.49788,-0.48246,0.45468,0.96082,-0.31685,0.52135,-1.23177,-0.01928,-0.01729,-2.18109,1.86203,0.40148,CL6,CL1,CL1,CL2,CL6,CL6,CL5,CL1,CL0,CL0,CL0,CL0,CL2,
29,0.49788,0.48246,1.16365,0.96082,-0.31685,0.04257,-1.50796,-0.71727,0.28783,1.30612,0.19268,-0.52593,CL6,CL0,CL0,CL0,CL4,CL1,CL5,CL0,CL0,CL0,CL0,CL0,CL0,CL
30,1.09449,-0.48246,1.98437,0.96082,-0.31685,-1.32828,1.74091,0.88309,0.94156,1.81175,-0.21712,-0.52593,CL0,CL0,CL0,CL0,CL6,CL0,CL6,CL0,CL0,CL0,CL0,CL0,CL0,C
31,1.09449,-0.48246,-1.73790,0.96082,-0.31685,0.31287,-0.80615,-1.27553,-1.47955,-0.65253,-1.37983,0.07987,CL6,CL1,CL1,CL1,CL6,CL2,CL6,CL1,CL0,CL0,CL1,CL0,CL
32,0.49788,0.48246,-1.73790,-0.28519,-0.31685,-1.19430,-0.80615,0.14143,-0.60633,-0.00665,-1.37983,-0.84637,CL3,CL0,CL0,CL2,CL3,CL0,CL5,CL0,CL0,CL0,CL0,CL0,C
33,0.49788,-0.48246,0.45468,0.96082,-0.31685,-0.79151,-1.23177,-0.01928,0.43852,-0.00665,0.19268,-0.84637,CL6,CL1,CL1,CL0,CL6,CL3,CL6,CL0,CL0,CL0,CL0,CL0,CL0
34,-0.95197,-0.48246,0.45468,0.96082,-0.31685,0.41667,-0.30033,0.29338,1.28610,-0.14277,-0.71126,-1.54858,CL6,CL0,CL0,CL0,CL6,CL0,CL6,CL0,CL0,CL0,CL0,CL0,CL0
```

# AFTER

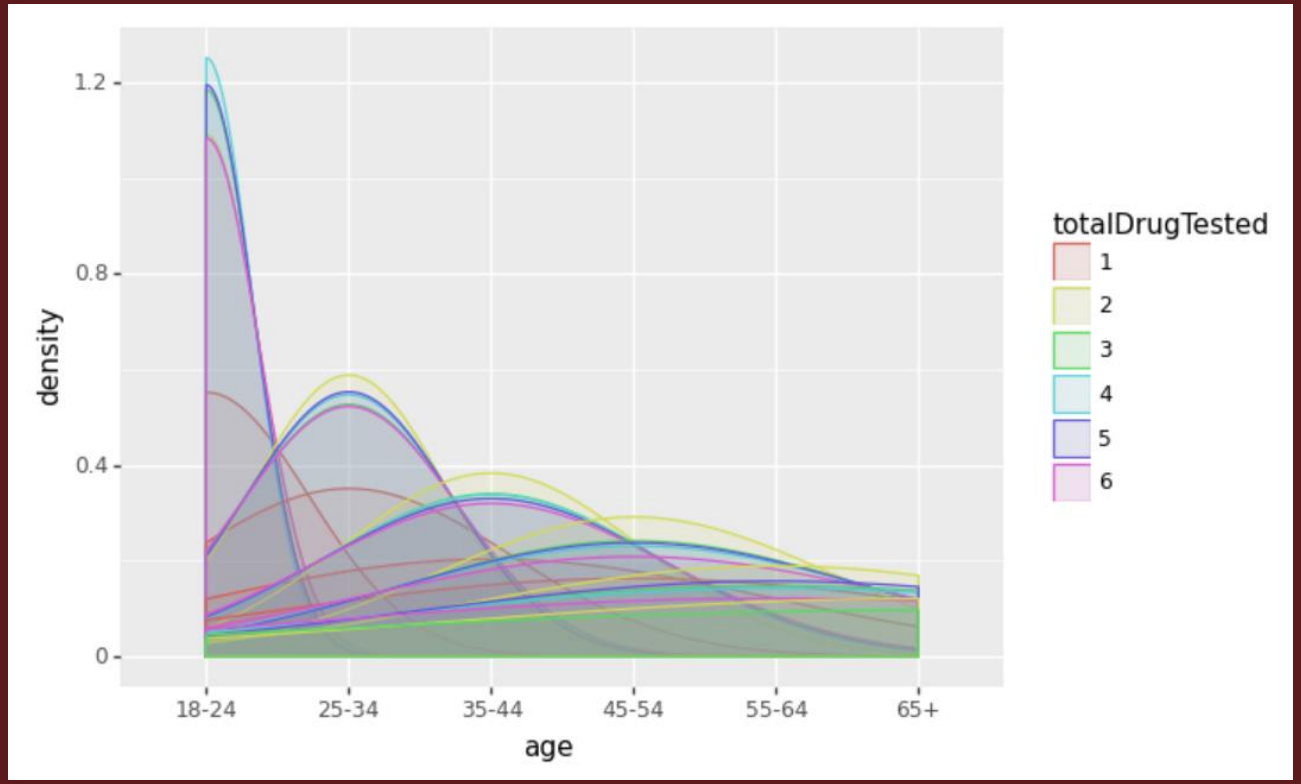
| | ID | age | gender | education | country | ethnicity | Nscore | Escore | Oscore | Ascore | ... | ecstasy | heroin | ketamine | legalh | LSD | Meth | mushrooms | nicc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 35-44 | 0.48246 | -0.05921 | 0.96082 | 0.12600 | 0.31287 | -0.57545 | -0.58331 | -0.91699 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **1** | 2 | 25-34 | -0.48246 | 1.98437 | 0.96082 | -0.31685 | -0.67825 | 1.93886 | 1.43533 | 0.76096 | ... | 1 | 0 | 1 | 0 | 1 | 1 | 0 | |
| **2** | 3 | 35-44 | -0.48246 | -0.05921 | 0.96082 | -0.31685 | -0.46725 | 0.80523 | -0.84732 | -1.62090 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| **3** | 4 | 18-24 | 0.48246 | 1.16365 | 0.96082 | -0.31685 | -0.14882 | -0.80615 | -0.01928 | 0.59042 | ... | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| **4** | 5 | 35-44 | 0.48246 | 1.98437 | 0.96082 | -0.31685 | 0.73545 | -1.63340 | -0.45174 | -0.30172 | ... | 1 | 0 | 0 | 1 | 0 | 0 | 1 | |

the variables that interested me to answer my hypothesis where the 7 personality traits and the drugs they had tested. To transforme the data on the drugs i changed all the CL1-7 to 1 and CL0 to 0. if they had tried the drug the string was replaced by 1 if not it was replaced by 0. I aslo changed the age variable because i beleived that age would influence how many drugs you have tried.I equally removed all the people having said that they took the semer drug which was a fictitious drug.

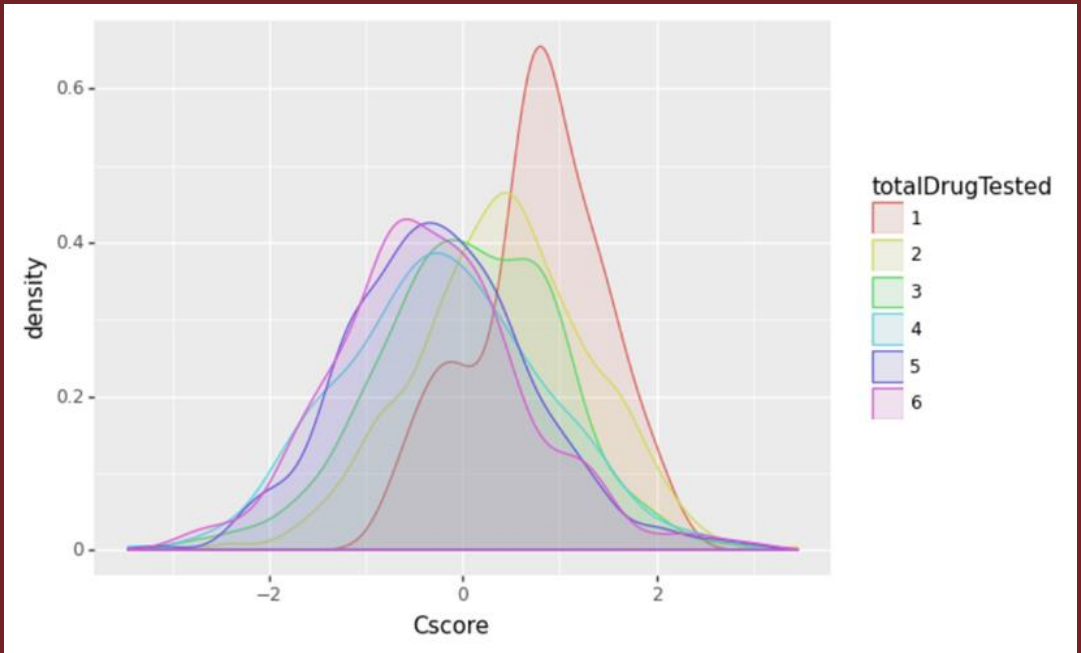
For the rest of the analysis I decided to add a column called totalnumberdrugs which can take a value from 1-6, 1 being having tried 1-3 types of drugs and 6 being 16-19 types of drugs.

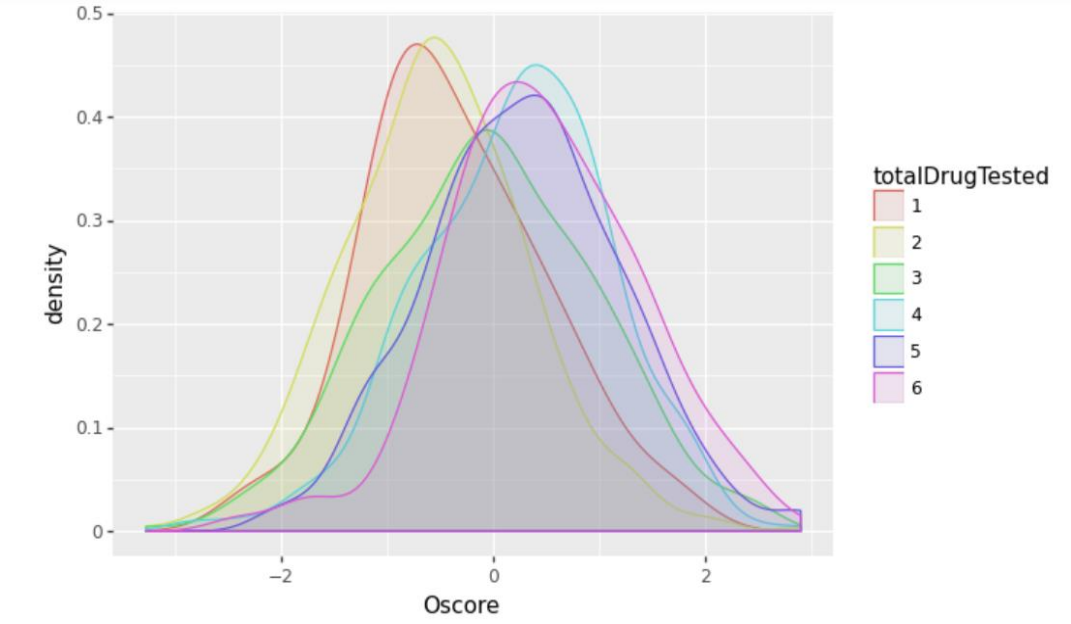
# ANALYSIS OF THE EFFECT OF AGE

I expected to see that the older generation would have tested more drugs since it seemed logic to me that the more time you have spent alive the more opportunities you will have to try new drugs. By analysing this chart we can see that the highest density for the people having tested the most drugs are actually between 18 and 24 years old. This means that being older doesn't mean that you have tried more drugs so i don't have to take age into account for the correlation between personality and the number of drugs tested.



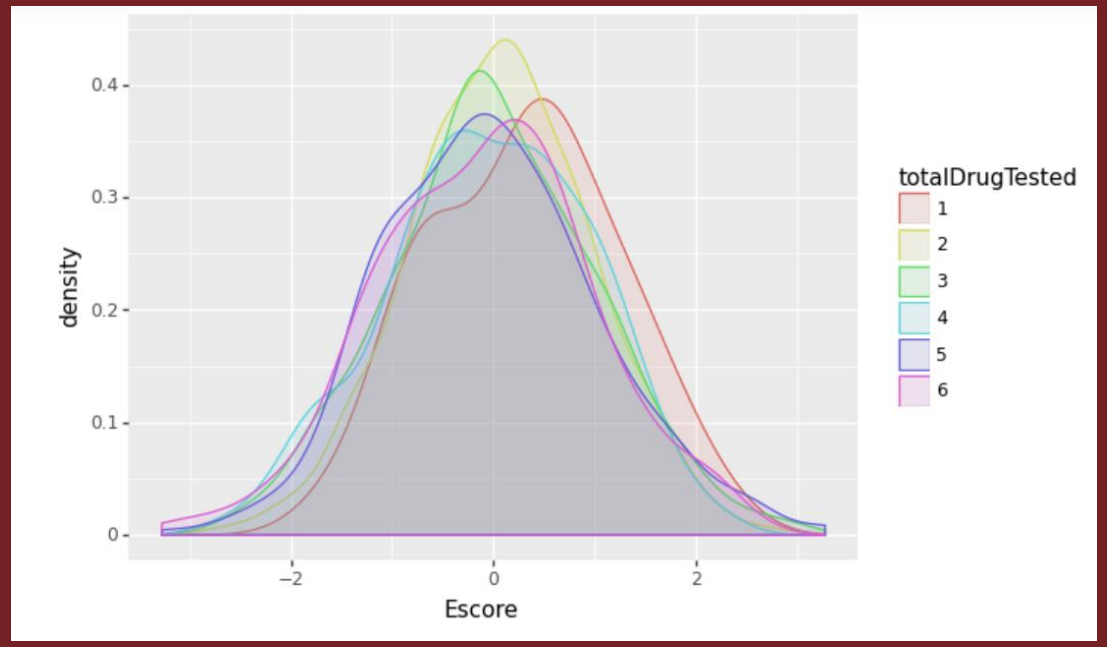
# THE LINK BETWEEN THE NUMBER OF DRUGS TESTED AND A CERTAIN PERSONALITY TRAIT

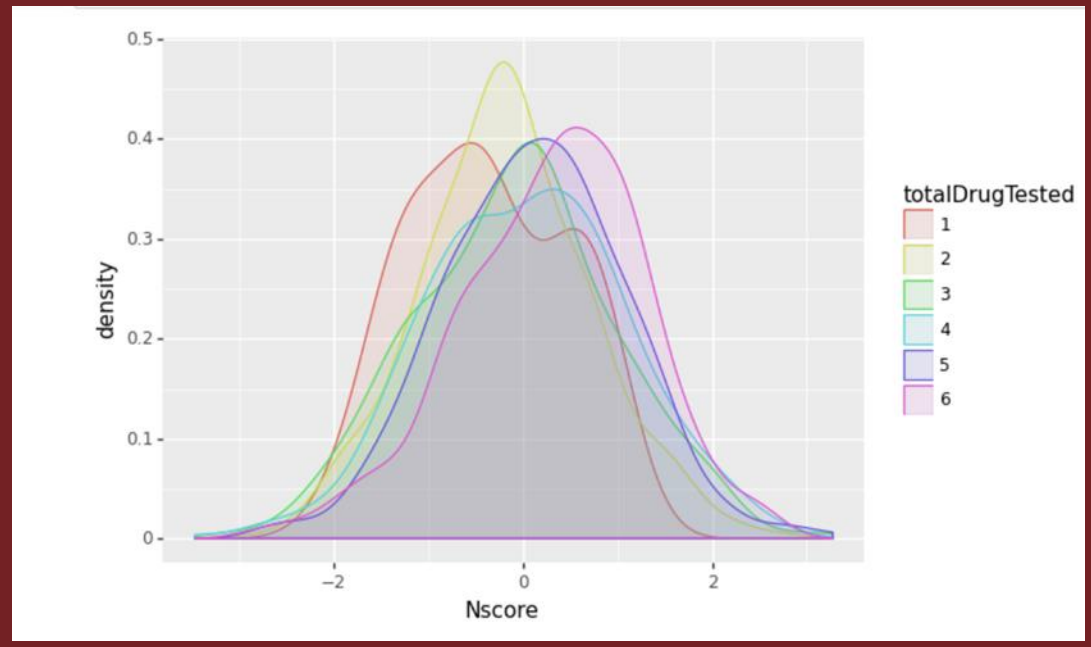


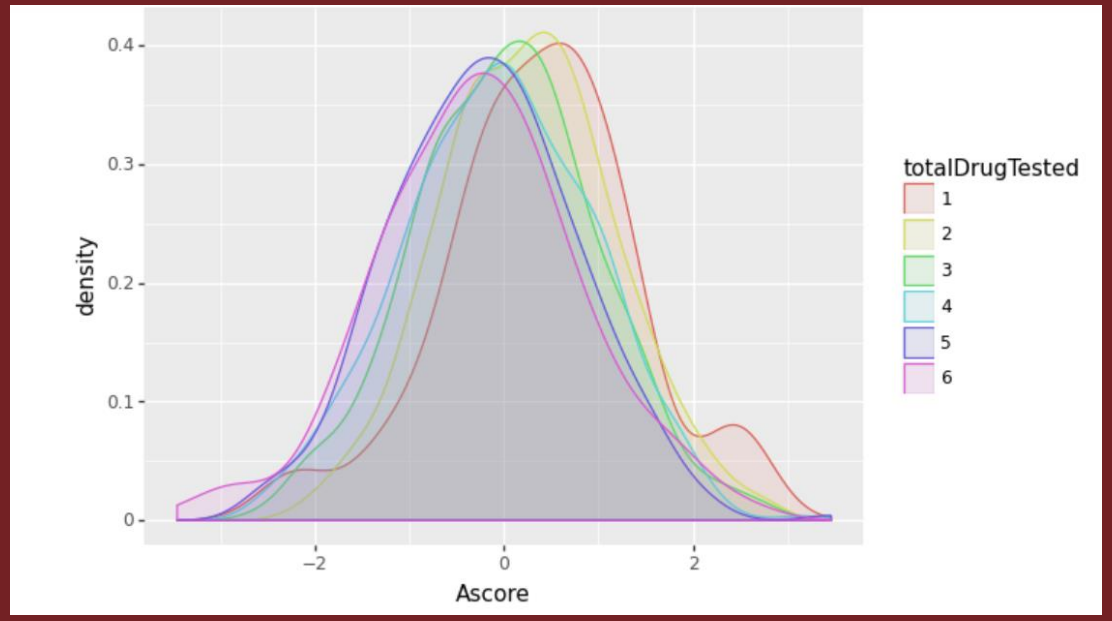




Here we have 2 graphs representing the Cscores and Oscores of individuals to the number of drugs they have tested. As we can see, there is a slight difference between people who have tested many drugs and those who have tried less. In the case of the Cscore we can see that people having tried 1-6 drugs are majorily with a positive cscore and about 70% of people that have tried more have a negative Cscore.

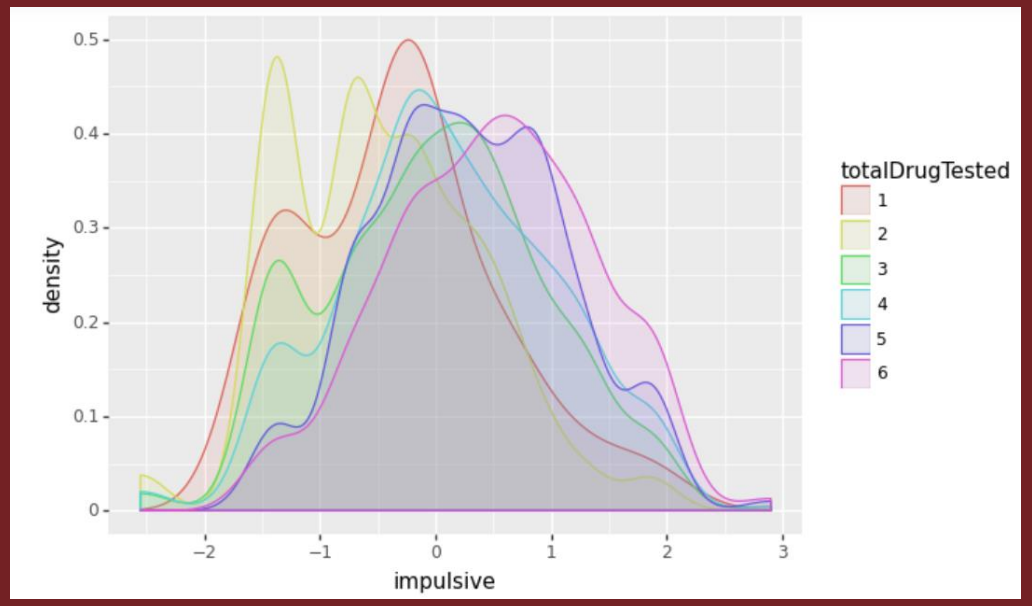
For the Oscrore 70% of people having tried 1-6 drugs have a negative Oscore. About 60% of people having tried 13 or more drugs have a positive Oscore. With these two graphs we can see that people with a high cscore and a low oscore have most likly tried less drugs than average.

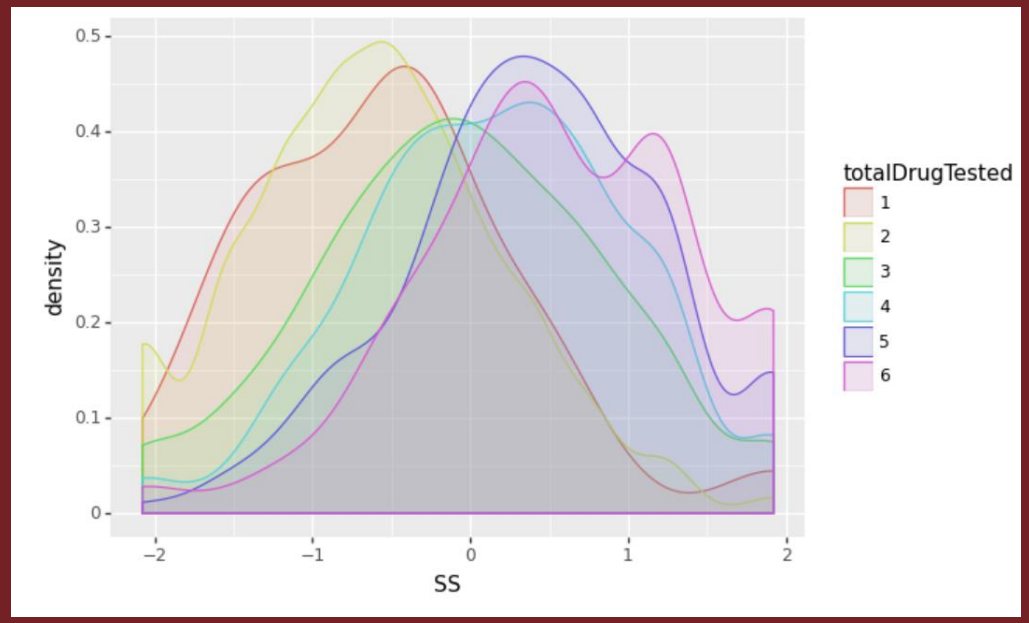



- In these 3 graphs the seperations between number of drugs tested and there personality score wern't far enough to be able to conclude anything. They each have different peaks on opposites sides of the x=0 mark but they spread out too much to be able to determin a correlation.




- Here we can see that there is a clear separation between the people who have tested more drugs and those who have teste very few. In the impulsivness graph we can determin that people with a score inferieur to -1 has a very high probability to have tested less than 6 drugs and about 66% of the people with a score higher than 0.5 have tested more than 9 drugs.
- For the sensation seeking personality trait we can see that above 0 the people tend to have tested more drugs, with about 80 percent of the people above 0.5 having tested more than 12 drugs. 70 percent of people with a score lower than -0.5 have tested less than 6 drugs equally.

# CALCULATING THE CORRELATION R TO VERIFY

| | Nscore | Escore | Oscore | Ascore | Cscore | impulsive | SS |
|---|---|---|---|---|---|---|---|
| totalDrugTested1 | 0.181 ** | -0.057 ** | 0.366 ** | -0.199 ** | -0.313 ** | 0.351 ** | 0.462 ** |

when I calculated  r i was glad to see that my interpretations of th graphs were pretty correct. The Nscore, Escore and Ascore have a near to 0 correlation to the number of drugs tested. The 4 others have a moderate correlation, meaning that by combining the 4 of them we should be able to predict the number of drugs you have tested with a certain marge of error. Sensation seeking has the biggest correlation.

# PREDICTION

- We aplying different statical models to the dataset in order to predict the amount of drugs you have tried in your life, the predicition that the computer had to make wasn't binary and have 19 different possible values. This meant the highest I could get was 20% especially since we saw that the variables wern't very well correlated. That being said when i tried predicting if someone used and above average amount of drugs i ended up finding a model that allowed me to accuratly predict 72% of the time. Which is not too bad. I can can conclude that there is a moderate correlation between the personality traits and the amount of drugs tested in your life time.

THANK YOU FOR YOU TIME