Exploratory Analysis of the Relationship between Personality Factors and Drug Consumption using R

Kelsey Charbeneau

Math 340

12/1/2023

Abstract

This project aims to to explore the predictive power of personality traits as measured by the NEO-FFI-R model, for drug consumption risk assessment. For this study, I will be using a dataset retrieved through the UC Irvine Machine Learning Repository, which includes personality measurements (NEO-FFI-R, BIS-11, ImpSS), demographic factors (age, gender, country of residence, and ethnicity), and respondents' use of 18 legal and illegal drugs across a sample of 1885 respondents. I aim to transform the data in order to yield insights into the predictive capacity of the Hexaco personality model for assessing drug consumption risk.

Background

Big Five (Costa & McCrae, 1992):
-Five personality traits (underlying facets) fully describing one's personality

- Neuroticism : the tendency to experience negative emotions and psychological distress in response to stressors
- Extraversion : the degree of sociability, positive emotionality, and general activity
- Openness : levels of curiosity, independent judgment, and conservativeness
- Agreeableness : altruistic, sympathetic, and cooperative tendencies
- Conscientiousness : one's level of self-control in planning and organization

The Dataset at a Glance

```
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,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,-0.71126,-0.21575,CL5,CL2,CL2,CL0,CL6,CL4,CL6,CL3,CL0,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,-1.37983,0.40148,CL6,CL0,CL0,CL0,CL6,CL3,CL4,CL0,CL0,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,-1.37983,-1.18084,CL4,CL0,CL0,CL3,CL5,CL2,CL4,CL2,CL0,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,-0.21712,-0.21575,CL4,CL1,CL1,CL0,CL6,CL3,CL6,CL0,CL0,CL1
,CL0,CL0,CL1,CL0,CL0,CL2,CL2,CL0,CL0
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,CL0,CL0,CL0,CL6,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,CL0,CL0,CL0,CL6,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,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,CL0,CL0,CL0,CL6,CL0,CL0
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,CL0,CL0,CL0,CL6,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,CL0,CL0,CL0,CL2,CL0,CL1
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,CL1,CL0,CL2,CL6,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,CL1,CL1,CL6,CL0,CL0
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,CL0,CL0,CL0,CL1,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,CL0,CL0,CL0,CL6,CL0,CL0
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
```

1:1 Text file

Console

Organizing the Dataset

```
drug_consumption <- read.table("C:\\Users\\charb\\Desktop\\drug_consumption.data"
, sep = ",", header = FALSE)
View(drug_consumption)

# Rename the columns of the "drug_consumption" dataset
colnames(drug_consumption) <- c("ID", "Age", "Gender", "Education", "Country",
"Ethnicity", "Nscore", "Escore", "Oscore", "Ascore", "Cscore", "Impulsive", "SS",
"Alcohol", "Amphet", "Amyl", "Benzos", "Caff", "Cannabis", "Choc", "Coke", "Crack",
"Ecstasy", "Heroin", "Ketamine", "Legalh", "LSD", "Meth", "Mushrooms", "Nicotine",
"Semer", "VSA")
drug_consumption$Impulsive <- NULL # Delete column
drug_consumption$SS <- NULL # Delete column
```

Codebook for Demographics

```
Education (Real) is level of education of participant and has one of the values:
    Value     Meaning                                                  Cases  Fraction
    −2.43591  Left school before 16 years                                 28     1.49%
    −1.73790  Left school at 16 years                                     99     5.25%
    −1.43719  Left school at 17 years                                     30     1.59%
    −1.22751  Left school at 18 years                                    100     5.31%
    −0.61113  Some college or university, no certificate or degree       506    26.84%
    −0.05921  Professional certificate/ diploma                          270    14.32%
     0.45468  University degree                                          480    25.46%
     1.16365  Masters degree                                             283    15.01%
     1.98437  Doctorate degree                                            89     4.72%
```

Recoding Demographic Factors

Repeated for all demographic variables in the dataset: age, gender, education, country, and ethnicity

```
#   GENDER
# Create a new variable "Gender" in drug_consumption
drug_consumption$Gender <- ifelse(drug_consumption$Gender == 0.48246, "Female",
                                  ifelse(drug_consumption$Gender == -0.48246, "Male", NA))
# Convert the "Gender" column to a factor for categorical representation
drug_consumption$Gender <- as.factor(drug_consumption$Gender)
```

Codebook for Personality Factors

Variable Information According to the Database

Nscore (Real) is NEO–FFI–R Neuroticism. Possible values are presented in table below:

Nscore	Cases	Value	Nscore	Cases	Value	Nscore	Cases	Value
12	1	−3.46436	29	60	−0.67825	46	67	1.02119
13	1	−3.15735	30	61	−0.58016	47	27	1.13281
14	7	−2.75696	31	87	−0.46725	48	49	1.23461
15	4	−2.52197	32	78	−0.34799	49	40	1.37297
16	3	−2.42317	33	68	−0.24649	50	24	1.49158
17	4	−2.34360	34	76	−0.14882	51	27	1.60383
18	10	−2.21844	35	69	−0.05188	52	17	1.72012
19	16	−2.05048	36	73	0.04257	53	20	1.83990
20	24	−1.86962	37	67	0.13606	54	15	1.98437
21	31	−1.69163	38	63	0.22393	55	11	2.12700
22	26	−1.55078	39	66	0.31287	56	10	2.28554
23	29	−1.43907	40	80	0.41667	57	6	2.46262
24	35	−1.32828	41	61	0.52135	58	3	2.61139
25	56	−1.19430	42	77	0.62967	59	5	2.82196
26	57	−1.05308	43	49	0.73545	60	2	3.27393
27	65	−0.92104	44	51	0.82562			
28	70	−0.79151	45	37	0.91093			

Recoding Personality Factors

Repeated for all personality factors in the dataset: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness.

```
# Recoding Personality Factors
#  Nscore------------------------------------------------------------------
values <- c(-3.46436, -3.15735, -2.75696, -2.52197, -2.42317, -2.34360,...
corresponding_Nscores <- 12:60
drug_consumption$Nscore <- corresponding_Nscores[match(drug_consumption$Nscore, values)]

#  ESCORE------------------------------------------------------------------
values <- c(-3.27393, -3.00537, -2.72827, -2.53830, -2.44904, -2.32338,...
corresponding_Escores <- 16:59
drug_consumption$Escore <- corresponding_Escores[match(drug_consumption$Escore, values)]

#  OSCORE------------------------------------------------------------------
values <- c(-3.27393, -2.85950, -2.63199, -2.39883, -2.21069, -2.09015,...
corresponding_Oscores <- 24:60
drug_consumption$Oscore <- corresponding_Oscores[match(drug_consumption$Oscore, values)]

#  ASCORE------------------------------------------------------------------
values <- c(-3.46436, -3.15735, -3.00537, -2.90161, -2.78793, -2.70172,...
corresponding_Ascores <- 12:60
drug_consumption$Ascore <- corresponding_Ascores[match(drug_consumption$Ascore, values)]

#  CSCORE------------------------------------------------------------------
values <- c(-3.46436, -3.15735, -2.90161, -2.72827, -2.57309, -2.42317,...
corresponding_Cscores <- 17:59
drug_consumption$Cscore <- corresponding_Cscores[match(drug_consumption$Cscore, values)]
```

Map for Drug Factors

Value	Class	Alcohol Cases	Alcohol Fraction	Amphet Cases	Amphet Fraction	Amyl Cases
CL0	Never Used	34	1.80%	976	51.78%	1305
CL1	Used over a Decade Ago	34	1.80%	230	12.20%	210
CL2	Used in Last Decade	68	3.61%	243	12.89%	237
CL3	Used in Last Year	198	10.50%	198	10.50%	92
CL4	Used in Last Month	287	15.23%	75	3.98%	24
CL5	Used in Last Week	759	40.27%	61	3.24%	14
CL6	Used in Last Day	505	26.79%	102	5.41%	3

Recoding Drug Factors

Repeated for all drug factors in the dataset: Alcohol, Amphetamine, Amyl, Benzos, Caffeine, Cannabis, Chocolate, Coke, Crack, Ecstasy, Heroin, Ketamine, Legal Highs, LSD, Methadone, Mushrooms, Nicotine, Semeron, and Volatile Substance Abuse

```
#  Recode drug factors---------------------------------------------

recode_map <- c(
  "CL0" = 0,
  "CL1" = 1,
  "CL2" = 2,
  "CL3" = 3,
  "CL4" = 4,
  "CL5" = 5,
  "CL6" = 6
)

# ALCOHOL----------------------------------------------------------
drug_consumption <- drug_consumption %>%
  mutate(Alcohol = recode(Alcohol,!!!recode_map))
drug_consumption <- drug_consumption %>%
  mutate(Alcohol = as.numeric(Alcohol))
```

New Library: Tidyverse

```
library(tidyverse)
```

- dplyr: package for data manipulation, including functions for filtering, selecting specific columns, mutating, and arranging data frames.
- tidyr: data manipulation package for restructuring data (e.g., separating and combining)
- ggplot2: data visualization package

Preparing the Dataset for Analysis

```
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,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,-0.71126,-0.21575,CL5,CL2,CL2,CL0,CL6,CL4,CL6,CL3,CL0,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,-1.37983,0.40148,CL6,CL0,CL0,CL0,CL6,CL3,CL4,CL0,CL0,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,-1.37983,-1.18084,CL4,CL0,CL0,CL3,CL5,CL2,CL4,CL2,CL0,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,-0.21712,-0.21575,CL4,CL1,CL1,CL0,CL6,CL3,CL6,CL0,CL0,CL1,CL0,CL0,CL1,CL0,CL0,CL2,CL2,CL0,CL0
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,CL0,CL0,CL0,CL6,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,CL0,CL0,CL0,CL6,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,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,CL0,CL0,CL0,CL6,CL0,CL0
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,CL0,CL0,CL0,CL6,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,CL0,CL0,CL0,CL2,CL0,CL1
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,CL1,CL0,CL2,CL6,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,CL1,CL1,CL6,CL0,CL0
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,CL0,CL0,CL0,CL1,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,CL0,CL0,CL0,CL6,CL0,CL0
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
```

1:1 Text file

Console

(b) Raw Dataset

ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Alcohol	Amphet
1	35-44	Female	Professional certificate/diploma	UK	Mixed-White/Asian	39	41	40	29	41	5	2
2	25-34	Male	Doctorate degree	UK	White	29	44	53	40	40	5	2
3	35-44	Male	Professional certificate/diploma	UK	White	31	30	38	24	33	6	0
4	18-24	Female	Masters degree	UK	White	34	37	44	39	45	4	0
5	35-44	Female	Doctorate degree	UK	White	43	27	41	33	49	4	1
6	65+	Female	Left school at 18 years	Canada	White	29	45	33	47	51	2	0
7	45-54	Male	Masters degree	USA	White	31	33	41	33	47	6	0
8	35-44	Male	Left school at 16 years	UK	White	24	44	38	33	51	5	0
9	35-44	Female	Professional certificate/diploma	Canada	White	42	50	37	40	48	4	0
10	55-64	Male	Masters degree	UK	White	33	49	34	39	42	6	1
11	25-34	Female	University degree	UK	White	26	30	36	30	52	5	0
12	45-54	Male	Some college or university, no certificate or degree	Other	White	24	49	45	22	37	5	1
13	55-64	Female	University degree	UK	White	56	51	47	24	35	5	1
14	55-64	Female	Professional certificate/diploma	Canada	White	28	30	44	41	57	1	0

(c) Revised Dataset

Figure: Results of Recoding

Correlations between Drug and Personality Factors

NEUROTICISM

```
cor(drug_consumption$Nscore,drug_consumption$Alcohol)= −0.003623107
cor(drug_consumption$Nscore,drug_consumption$Amphet)= 0.1350988
cor(drug_consumption$Nscore,drug_consumption$Amyl)= 0.03482094
cor(drug_consumption$Nscore,drug_consumption$Benzos)= 0.2771312
cor(drug_consumption$Nscore,drug_consumption$Caff)= 0.01400058
cor(drug_consumption$Nscore,drug_consumption$Cannabis)=0.1026491
cor(drug_consumption$Nscore,drug_consumption$Choc)=0.01070876
cor(drug_consumption$Nscore,drug_consumption$Coke)=0.1430317
cor(drug_consumption$Nscore,drug_consumption$Crack)= 0.116194
cor(drug_consumption$Nscore,drug_consumption$Ecstasy)= 0.07311829
cor(drug_consumption$Nscore,drug_consumption$Heroin)=0.1781161
cor(drug_consumption$Nscore,drug_consumption$Ketamine)=0.06377705
cor(drug_consumption$Nscore,drug_consumption$Legalh)=0.1193922
cor(drug_consumption$Nscore,drug_consumption$LSD)=0.04142275
cor(drug_consumption$Nscore,drug_consumption$Meth)=0.1887433
cor(drug_consumption$Nscore,drug_consumption$Mushrooms)=0.04511796
cor(drug_consumption$Nscore,drug_consumption$Nicotine)=0.1317899
cor(drug_consumption$Nscore,drug_consumption$VSA)=0.1164636
```

Correlations between Drug and Personality Factors

EXTRAVERSION

```
cor(drug_consumption$Escore, drug_consumption$Alcohol)=0.0236585
cor(drug_consumption$Escore, drug_consumption$Amphet)=-0.04399795
cor(drug_consumption$Escore, drug_consumption$Amyl)=-0.01303947
cor(drug_consumption$Escore, drug_consumption$Benzos)=-0.0416334
cor(drug_consumption$Escore, drug_consumption$Caff)=0.01798573
cor(drug_consumption$Escore, drug_consumption$Cannabis)=-0.04618776
cor(drug_consumption$Escore, drug_consumption$Choc)=0.018859
cor(drug_consumption$Escore, drug_consumption$Coke)=-0.02706438
cor(drug_consumption$Escore, drug_consumption$Crack)=0.01753025
cor(drug_consumption$Escore, drug_consumption$Ecstasy)=-0.01022163
cor(drug_consumption$Escore, drug_consumption$Heroin)=-0.04810981
cor(drug_consumption$Escore, drug_consumption$Ketamine)=-0.02090958
cor(drug_consumption$Escore, drug_consumption$Legalh)=-0.07163388
cor(drug_consumption$Escore, drug_consumption$LSD)=-0.008244115
cor(drug_consumption$Escore, drug_consumption$Meth)=-0.07930321
cor(drug_consumption$Escore, drug_consumption$Mushrooms)=-0.01990358
cor(drug_consumption$Escore, drug_consumption$Nicotine)=-0.01256289
cor(drug_consumption$Escore, drug_consumption$VSA)=-0.04541033
```

Correlations between Drug and Personality Factors

OPENNESS

```
cor(drug_consumption$Oscore,drug_consumption$Alcohol)=0.03517981
cor(drug_consumption$Oscore,drug_consumption$Amphet)=0.2177496
cor(drug_consumption$Oscore,drug_consumption$Amyl)=0.06349917
cor(drug_consumption$Oscore,drug_consumption$Benzos)=0.2009715
cor(drug_consumption$Oscore,drug_consumption$Caff)=0.0293603
cor(drug_consumption$Oscore,drug_consumption$Cannabis)=0.4166981
cor(drug_consumption$Oscore,drug_consumption$Choc)=-0.001564351
cor(drug_consumption$Oscore,drug_consumption$Coke)=0.1875638
cor(drug_consumption$Oscore,drug_consumption$Crack)=0.09955328
cor(drug_consumption$Oscore,drug_consumption$Ecstasy)=0.2955959
cor(drug_consumption$Oscore,drug_consumption$Heroin)=0.1331485
cor(drug_consumption$Oscore,drug_consumption$Ketamine)=0.1838828
cor(drug_consumption$Oscore,drug_consumption$Legalh)=0.3167716
cor(drug_consumption$Oscore,drug_consumption$LSD)=0.3658541
cor(drug_consumption$Oscore,drug_consumption$Meth)=0.1683063
cor(drug_consumption$Oscore,drug_consumption$Mushrooms)=0.3699001
cor(drug_consumption$Oscore,drug_consumption$Nicotine)=0.1934488
cor(drug_consumption$Oscore,drug_consumption$VSA)=0.151577
```

Correlations between Drug and Personality Factors

AGREEABLENESS

```
cor(drug_consumption$Ascore,drug_consumption$Alcohol)=-0.02196795
cor(drug_consumption$Ascore,drug_consumption$Amphet)=-0.1512892
cor(drug_consumption$Ascore,drug_consumption$Amyl)=-0.09584465
cor(drug_consumption$Ascore,drug_consumption$Benzos)=-0.1692687
cor(drug_consumption$Ascore,drug_consumption$Caff)=-0.01482282
cor(drug_consumption$Ascore,drug_consumption$Cannabis)=-0.1499492
cor(drug_consumption$Ascore,drug_consumption$Choc)=0.03858875
cor(drug_consumption$Ascore,drug_consumption$Coke)=-0.2001018
cor(drug_consumption$Ascore,drug_consumption$Crack)=-0.1065712
cor(drug_consumption$Ascore,drug_consumption$Ecstasy)=-0.1157888
cor(drug_consumption$Ascore,drug_consumption$Heroin)=-0.1732953
cor(drug_consumption$Ascore,drug_consumption$Ketamine)=-0.1136285
cor(drug_consumption$Ascore,drug_consumption$Legalh)=-0.1419472
cor(drug_consumption$Ascore,drug_consumption$LSD)=-0.09593415
cor(drug_consumption$Ascore,drug_consumption$Meth)= -0.1602879
cor(drug_consumption$Ascore,drug_consumption$Mushrooms)=-0.1139185
cor(drug_consumption$Ascore,drug_consumption$Nicotine)=-0.1126732
cor(drug_consumption$Ascore,drug_consumption$VSA)=-0.1157866
```

Correlations between Drug and Personality Factors

CONSCIENTIOUSNESS

```
cor(drug_consumption$Cscore,drug_consumption$Alcohol)=−0.0001288821
cor(drug_consumption$Cscore,drug_consumption$Amphet)=−0.2430689
cor(drug_consumption$Cscore,drug_consumption$Amyl)=−0.1148207
cor(drug_consumption$Cscore,drug_consumption$Benzos)=−0.2078345
cor(drug_consumption$Cscore,drug_consumption$Caff)−0.02254394
cor(drug_consumption$Cscore,drug_consumption$Cannabis)=−0.2791208
cor(drug_consumption$Cscore,drug_consumption$Choc)=0.002738498
cor(drug_consumption$Cscore,drug_consumption$Coke)=−0.1967217
cor(drug_consumption$Cscore,drug_consumption$Crack)=−0.1357951
cor(drug_consumption$Cscore,drug_consumption$Ecstasy)=−0.2220136
cor(drug_consumption$Cscore,drug_consumption$Heroin)=−0.1633221
cor(drug_consumption$Cscore,drug_consumption$Ketamine)=−0.1582593
cor(drug_consumption$Cscore,drug_consumption$Legalh)=−0.2619945
cor(drug_consumption$Cscore,drug_consumption$LSD)=−0.16464
cor(drug_consumption$Cscore,drug_consumption$Meth)−0.1972478
cor(drug_consumption$Cscore,drug_consumption$Mushrooms)=−0.1947422
cor(drug_consumption$Cscore,drug_consumption$Nicotine)=−0.2305804
cor(drug_consumption$Cscore,drug_consumption$VSA)=−0.1642802
```

Correlations between Drug and Personality Factors

IMPULSIVITY

```
cor(drug_consumption$Impulsive,drug_consumption$Alcohol)=0.04597158
cor(drug_consumption$Impulsive,drug_consumption$Amphet)=0.2894382
cor(drug_consumption$Impulsive,drug_consumption$Amyl)=0.1262638
cor(drug_consumption$Impulsive,drug_consumption$Benzos)=0.2233744
cor(drug_consumption$Impulsive,drug_consumption$Caff)=0.04938775
cor(drug_consumption$Impulsive,drug_consumption$Cannabis)=0.3105287
cor(drug_consumption$Impulsive,drug_consumption$Choc)=−0.02017834
cor(drug_consumption$Impulsive,drug_consumption$Coke)=0.2600421
cor(drug_consumption$Impulsive,drug_consumption$Crack)=0.1857307
cor(drug_consumption$Impulsive,drug_consumption$Ecstasy)=0.260864
cor(drug_consumption$Impulsive,drug_consumption$Heroin)=0.1977012
cor(drug_consumption$Impulsive,drug_consumption$Ketamine)=0.1776646
cor(drug_consumption$Impulsive,drug_consumption$Legalh)=0.2675788
cor(drug_consumption$Impulsive,drug_consumption$LSD)=0.2292051
cor(drug_consumption$Impulsive,drug_consumption$Meth)=0.1815242
cor(drug_consumption$Impulsive,drug_consumption$Mushrooms)=0.2636839
cor(drug_consumption$Impulsive,drug_consumption$Nicotine)=0.2462988
cor(drug_consumption$Impulsive,drug_consumption$VSA)=0.1810185
```

Correlations between Drug and Personality Factors

SENSATION SEEKING

```
cor(drug_consumption$SS, drug_consumption$Alcohol)0.1084723
cor(drug_consumption$SS, drug_consumption$Amphet)=0.3311052
cor(drug_consumption$SS, drug_consumption$Amyl)=0.1952802
cor(drug_consumption$SS, drug_consumption$Benzos)=0.2479033
cor(drug_consumption$SS, drug_consumption$Caff)=0.05204936
cor(drug_consumption$SS, drug_consumption$Cannabis)=0.4561366
cor(drug_consumption$SS, drug_consumption$Choc)=−0.03983596
cor(drug_consumption$SS, drug_consumption$Coke)=0.3433521
cor(drug_consumption$SS, drug_consumption$Crack)=0.1902012
cor(drug_consumption$SS, drug_consumption$Ecstasy)=0.3881862
cor(drug_consumption$SS, drug_consumption$Heroin)=0.2136844
cor(drug_consumption$SS, drug_consumption$Ketamine)=0.2436086
cor(drug_consumption$SS, drug_consumption$Legalh)=0.4055779
cor(drug_consumption$SS, drug_consumption$LSD)=0.3655358
cor(drug_consumption$SS, drug_consumption$Meth)=0.2188839
cor(drug_consumption$SS, drug_consumption$Mushrooms)=0.3782854
cor(drug_consumption$SS, drug_consumption$Nicotine)=0.3056346
cor(drug_consumption$SS, drug_consumption$VSA)=0.2505994
```

First attempt at Visualization

	Alcohol	Amphet	Amyl	Benzos	Caff	Cannabis	Choco	Coke	Crack	Ecstasy	Heroin	Ketamine	Legalh	LSD	Meth	Mushrooms	Niotine	VSA
Nscore		0.1		0.3		0.1		0.1	0.1	0.1	0.2	0.1	0.1		0.2		0.1	0.1
Escore													-0.1		-0.1			
Oscore		0.2	0.1	0.2		0.4		0.2	0.1	0.3	0.1	0.2	0.3	0.4	0.2	0.4	0.2	0.2
Ascore		-0.2	-0.1	-0.2		-0.2		-0.2	-0.1	-0.1	-0.2	-0.1	-0.1	-0.1	-0.2	-0.1	-0.1	-0.1
Cscore		-0.2	-0.1	-0.2		-0.3		-0.2	-0.1	-0.2	-0.2	-0.2	-0.3	-0.2	-0.2	-0.2	-0.2	-0.2
Impulsive		0.3	0.1	0.2		0.3		0.3	0.2	0.3	0.2	0.2	0.3	0.2	0.2	0.3	0.2	0.2
SS	0.1	0.3	0.2	0.2	0.1	0.5		0.3	0.2	0.4	0.2	0.2	0.4	0.4	0.2	0.4	0.3	0.3

Creating a loop for all Drug/Personality Correlations

```
# personality vs drug factors

personality_columns <- c("Nscore", "Escore", "Oscore", "Ascore", "Cscore")
drug_columns <- colnames(drug_consumption)[12:ncol(drug_consumption)]

correlation_results <- matrix(
  NA,
  nrow = length(personality_columns),
  ncol = length(drug_columns),
  dimnames = list(personality_columns, drug_columns)
)

for (personality in personality_columns) {
  for (drug in drug_columns) {
    correlation <- cor.test(drug_consumption[[personality]], drug_consumption[[drug]])
    correlation_results[personality, drug] <- correlation$estimate
  }
}
correlation_results
```

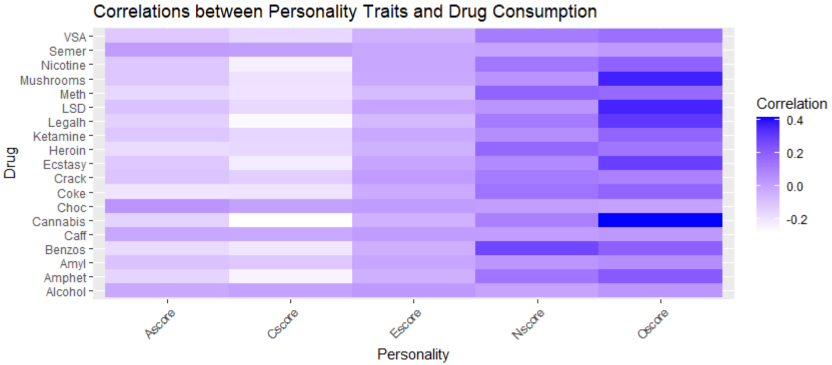
Code for Visualizing Drug/Personality Correlations

```
# Convert the correlations to a data frame
correlation_df <- as.data.frame(correlation_results)
correlation_df <- cbind(rownames(correlation_df), correlation_df)
names(correlation_df) <- c("Personality", colnames(correlation_results))

# Reshape data for plotting
correlation_long <- pivot_longer(correlation_df, cols = -Personality, names_to = "Drug", valu

# Visualize the correlations
ggplot(data = correlation_long, aes(x = Personality, y = Drug, fill = Correlation)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Correlations between Personality Traits and Drug Consumption") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Visualization of Drug/Personality Correlations



Stratifying the Distribution of given Personality Factors by the Interval of Drug Consumption

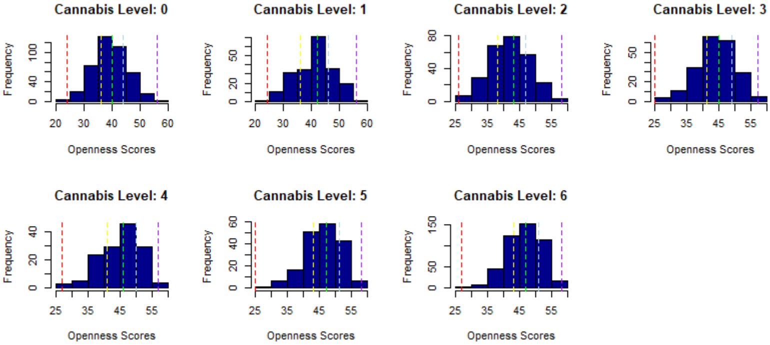
```
# Distribution of Openness Scores by Cannabis Consumption
# Set up the plotting layout
par(mfrow = c(2, 4))

for (cannabis_level in 0:6) {
  # Subset data for the specific cannabis level
  subset_data <- drug_consumption[drug_consumption$Cannabis == cannabis_level, ]

  # Create the histogram for Oscore at each Cannabis level
  hist(subset_data$Oscore,
       col = "darkblue",
       main = paste("Cannabis Level:", cannabis_level),
       xlab = "Openness Scores",
       ylab = "Frequency")

  # Adding quantiles to the histograms
  quantiles <- quantile(subset_data$Oscore, probs = c(0, 0.25, 0.5, 0.75, 1))
  abline(v = quantiles, col = c("red", "yellow", "green", "lightblue", "purple"), lty = 2)
}
```

Comparative Histograms of Openness Scores by levels of Cannabis Consumption



Replacing individual Drug Factor Code with a Loop

```
# Recode map for drug factors
recode_map <- c(
  "CL0" = 0,
  "CL1" = 1,
  "CL2" = 2,
  "CL3" = 3,
  "CL4" = 4,
  "CL5" = 5,
  "CL6" = 6
)

# Loop through each drug factor and apply recode operation
for (drug_factor in drug_factors) {
  drug_consumption <- drug_consumption %>%
    mutate_at(vars(drug_factor), list(~recode(., !!!recode_map))) %>%
    mutate_at(vars(drug_factor), as.numeric)
}
```

Recap of the Data Set

- 5 Nominal Variables: identifier, age, gender, education, country, ethnicity
- 7 Interval Variables (Personality Factors)
- 18 Ordinal Variables (Drug Factors)
- 1855 respondents (hence 1855 rows across 31 variables)

Creating a loop for all Correlations among Drug Factors

```
drug_columns <- colnames(drug_consumption)[12:ncol(drug_consumption)]

correlation_results <- matrix(
  NA,
  nrow = length(drug_columns),
  ncol = length(drug_columns),
  dimnames = list(drug_columns, drug_columns)
)

for (i in 1:length(drug_columns)) {
  for (j in 1:length(drug_columns)) {
    # Compute correlation
    correlation <- cor.test(drug_consumption[[drug_columns[i]]], drug_consumption[[drug_colum
    correlation_value <- correlation$estimate

    # Omit correlations between the same drugs (where r=1)
    if (i == j) {
      correlation_value <- NA
    }

    correlation_results[i, j] <- correlation_value
  }
}
```

Code for Visualizing Correlations among Drug Factors

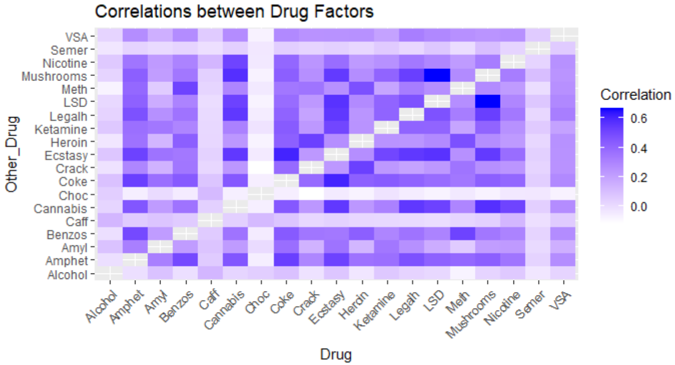
```
# Convert the correlations to a data frame
correlation_df <- as.data.frame(correlation_results)
correlation_df <- cbind(rownames(correlation_df), correlation_df)
names(correlation_df) <- c("Drug", colnames(correlation_results))

correlation_long <- pivot_longer(correlation_df, cols = -Drug, names_to = "Other_Drug", valu

# Remove NA values (where r=1) for the heatmap
correlation_long <- na.omit(correlation_long)

# Visualize the correlations
ggplot(data = correlation_long, aes(x = Drug, y = Other_Drug, fill = Correlation)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Correlations between Drug Factors") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Visualization of the Correlations among Drug Factors



Constructing a Multiple Linear Regression Model

```
# Optimization
# Create a list of drug factors
drug_factors_2 <- c("Alcohol", "Cannabis", "Mushrooms", "LSD", "Legalh")

# Create an empty data frame to store results
optimized_table <- data.frame(
  Drug_Factor = character(),
  Nscore = numeric(),
  Escore = numeric(),
  Oscore = numeric(),
  Ascore = numeric(),
  Cscore = numeric()
)

# Create a function to perform optimization for each drug factor
perform_optimization <- function(drug_factor) {
  # Objective function to minimize consumption for the given drug factor
  objective_function <- function(personality_traits) {
    # Perform linear regression for the drug factor based on personality traits
    model <- lm(as.formula(paste(drug_factor, "~ Nscore + Escore + Oscore + Ascore + Cscore"
```

Linear Regression Model Cont

```
    # Predict consumption using the updated model and given personality traits
    new_data <- data.frame(t(personality_traits))
    names(new_data) <- c("Nscore", "Escore", "Oscore", "Ascore", "Cscore")
    predicted <- predict(model, newdata = new_data)

    return(sum(predicted^2))  # Objective is to minimize the sum of squared consumption
  }

  # Set initial values for personality traits
  initial_personality_traits <- c(0, 0, 0, 0, 0)

  # Define lower and upper bounds for personality traits
  lower_bounds <- c(12, 16, 24, 12, 17)  # Adjusted lower bounds for personality traits
  upper_bounds <- c(60, 59, 60, 60, 59)  # Adjusted upper bounds for personality traits

  # Use optim() function to minimize the objective function
  optimized_traits <- optim(
    par = initial_personality_traits,
    fn = objective_function,
    method = "L-BFGS-B",  # Choose an appropriate optimization method
    lower = lower_bounds,  # Use defined lower bounds for personality traits
    upper = upper_bounds   # Use defined upper bounds for personality traits
  )$par

  # Return the optimized values of personality traits
  return(round(optimized_traits))  # Return rounded values
}
```

Linear Regression Model Cont

```
# Perform optimization for each drug factor using a loop
for (drug_factor in drug_factors_2) {
  # Get the optimized traits for the current drug factor
  optimized_values <- perform_optimization(drug_factor)

  # Add the drug factor and optimized values to the table
  optimized_table <- rbind(optimized_table, c(drug_factor, optimized_values))
}

# Assign column names to the table
colnames(optimized_table) <- c("Drug_Factor", "Nscore", "Escore", "Oscore",...

# Print the table with optimized values for each drug factor
print(optimized_table)
```

Linear Regression Model Results

	Drug_Factor	Nscore	Escore	Oscore	Ascore	Cscore
1	Alcohol	60	16	24	60	17
2	Cannabis	15	20	24	29	50
3	Mushrooms	18	19	24	26	40
4	LSD	16	18	24	24	35
5	Legalh	12	21	24	25	43

Figure: The Optimal Composition of Personality Factors for Limiting Consumption of Each Drug Factor.

R Code for Visualizing our Multiple Linear Regression Model

```
# Your data
data <- data.frame(
  Drug_Factor = c("Alcohol", "Cannabis", "Mushrooms", "LSD", "Legalh"),
  Nscore = c(60, 15, 18, 16, 12),
  Escore = c(16, 20, 19, 18, 21),
  Oscore = c(24, 24, 24, 24, 24),
  Ascore = c(60, 29, 26, 24, 25),
  Cscore = c(17, 50, 40, 35, 43)
)

# Reshape data for plotting
data_long <- data %>%
  pivot_longer(cols = -Drug_Factor, names_to = "Personality", values_to = "Value")

# Plotting grouped bar chart
ggplot(data_long, aes(x = Drug_Factor, y = Value, fill = Personality)) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7) +
  labs(
    title = "Optimized Personality Traits across Drug Factors",
    subtitle = "Grouped bar chart showcasing optimized personality trait values...
    x = "Drug Factor",
    y = "Optimized Value",
    fill = "Personality"
  ) +
  theme(legend.position = "top")
```

Visualization of our Multiple Linear Regression Model

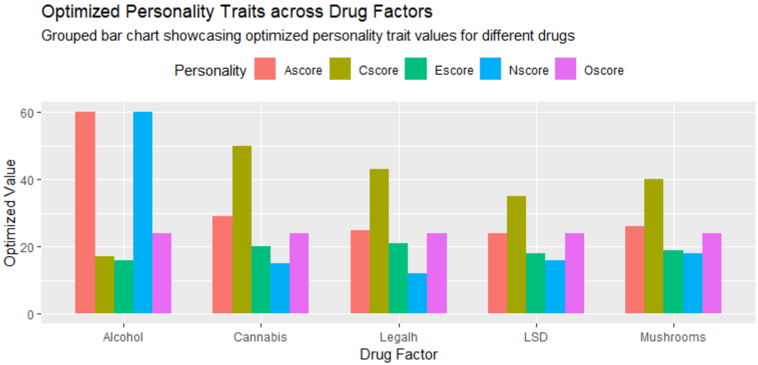


Figure: The Composition of Personality Factors that Minimize Drug Consumption