Quiz_Solution_-_Drug_Consumption

1 Quiz - Drug Consumption

Solution to Drug Consumption quiz on Moodle.

[1]:
```
import pandas as pd

df = pd.read_pickle("https://setu-datamining2.github.io/live/files/ICPSR_36536.
 ↪pkl")
pd.set_option('display.max_columns', None)
print(df.shape)
df.head()
```

```
(1877, 31)
```

[1]:

ID	Age	Gender	Education	Country	Ethnicity	Neuroticism	Extraversion	Openness_to_Experience	Agreeableness	Conscientiousness	Impulsiveness	Sensation_Seeking	Alcohol	Amphetamines	Amyl_Nitrite
1	35-44	Female	Professional	UK	Mixed-White/Asian	0.31287	-0.57545	-0.58331	-0.91699	-0.00665	-0.21712	-1.18084	CL5	CL2	CL0
2	25-34	Male	PhD	UK	White	-0.67825	1.93886	1.43533	0.76096	-0.14277	-0.71126	-0.21575	CL5	CL2	CL2
3	35-44	Male	Professional	UK	White	-0.46725	0.80523	-0.84732	-1.62090	-1.01450	-1.37983	0.40148	CL6	CL0	CL0
4	18-24	Female	Masters	UK	White	-0.14882	-0.80615	-0.01928	0.59042	0.58489	-1.37983	-1.18084	CL4	CL0	CL0
5	35-44	Female	PhD	UK	White	0.73545	-1.63340	-0.45174	-0.30172	1.30612	-0.21712	-0.21575	CL4	CL1	CL1

ID	Benzodiazepine	Caffeine	Cannabis	Chocolate	Cocaine	Crack	Ecstasy	Heroin
1	CL2	CL6	CL0	CL5	CL0	CL0	CL0	CL0
2	CL0	CL6	CL4	CL6	CL3	CL0	CL4	CL0
3	CL0	CL6	CL3	CL4	CL0	CL0	CL0	CL0
4	CL3	CL5	CL2	CL4	CL2	CL0	CL0	CL0
5	CL0	CL6	CL3	CL6	CL0	CL0	CL1	CL0

ID	Ketamine	Legal_Highs	LSD	Methadone	Mushrooms	Nicotine	Semeron	Volatiles
1	CL0	CL0	CL0	CL0	CL0	CL2	CL0	CL0
2	CL2	CL0	CL2	CL3	CL0	CL4	CL0	CL0
3	CL0	CL0	CL0	CL0	CL1	CL0	CL0	CL0
4	CL2	CL0	CL0	CL0	CL0	CL2	CL0	CL0
5	CL0	CL1	CL0	CL0	CL2	CL2	CL0	CL0

[2]:
```
legal = 'Alcohol Caffeine Chocolate Nicotine'.split()
demo_columns = list(df.columns[:5])
personality_columns = list(df.columns[5:12])
drugs = list(df.columns[12:])
illegal = [d for d in drugs if d not in legal]
features = demo_columns + personality_columns
targets = legal + illegal
df = df[features + targets].copy()
```

Group columns - not needed but nice to do.

[3]:
```
def highlight_cols(s):
  colors =  (['white'] * len(demo_columns) +
    ['azure'] * len(personality_columns) +
    ['bisque'] * len(legal) + ['thistle'] * len(illegal))
  return [f'background-color: {c}' for c in colors]
df.head().style.apply(highlight_cols, axis=1)
```

Using style to colour groups of columns.

[3]: <pandas.io.formats.style.Styler at 0x7fcc712fcd00>

Output does not work when generating PDF copy of notebook.

1.1 Q1

What is the max value of `Neuroticism`?

[4]:
```
df.Neuroticism.max()
```

[4]: 3.27393

For all questions the output is only the required value

1.2 Q2

What is the range of `Extraversion`?

[5]:
```
df.Extraversion.max() - df.Extraversion.min()
```

[5]: 6.54786

1.3 Q3

How many people never used an illegal drug?

[6]:
```
df['Illegal_Drug_Count'] = (df[illegal]!="CL0").sum(axis=1)
df['Never_Ilegal_Drug'] = df.Illegal_Drug_Count == 0
```

[7]:
```
df.Never_Ilegal_Drug.sum()
```

[7]: 299

1.4 Q4

How many people never used a legal drug?

[8]:
```
df['Legal_Drug_Count'] = (df[legal]!="CL0").sum(axis=1)
df['Never_Legal_Drug'] = df.Legal_Drug_Count == 0
```

[9]:
```
df.Never_Legal_Drug.sum()
```

[9]: 1

1.5 Q5

How many people used exactly one illegal drug?

[10]:
```
(df.Illegal_Drug_Count==1).sum()
```

[10]: 213

1.6 Q6

What is the (spearman) correlation between the number of different illegal drugs taken and `Openness_to_Experience`?

[11]:
```
columns = ['Openness_to_Experience','Illegal_Drug_Count']
df[columns].corr(method='spearman').values[0,1]
```

[11]: 0.37124154467902964

1.7 Q7

How many people who have recently used an illegal drug are from the UK?

[12]:
```
df['Recent_Illegal_Drug_Count'] = (df[illegal]>="CL3").sum(axis=1)
df['Recent_Illegal_Drug'] = df.Recent_Illegal_Drug_Count>0
```

[13]:
```
df.loc[df.Recent_Illegal_Drug, "Country"].value_counts()['UK']
```

[13]: 411

1.8 Q8

What is the probability that a person is male given than they are from the USA?

[14]:
```
df.loc[df.Country=="USA"].Gender.value_counts(normalize=True)["Male"]
```

[14]: 0.6243194192377496

1.9 Q9

What is the probability that a person is a recent illegal drug user given than they are from the USA?

[15]:
```
df.loc[df.Country=="USA", "Recent_Illegal_Drug"].
↪value_counts(normalize=True)[True]
```

[15]: 0.9655172413793104

[15]:
```

```