ANALYSIS OF NATIONAL SURVEY OF DRUG USE AND HEALTH 2015-2019 IN US

Raminta Misiunaite 2022

This project aims to analyse the NSDUH dataset using statistical analysis methods to test raised hypotheses. The main research questions are focused on substance abuse relationship with social aspects and mental health. This analysis will be examining the use of cigarettes, alcohol, marijuana, and LSD. The first two are the most accepted and commonly used drugs, yet one of the most addictive (americanaddictioncenters.org, 2022) and a big risk factor for the development of various diseases (2019). On the other hand, marijuana and LSD are both illegal and highly stigmatized in society, although having been proven to provide certain health benefits. Cannabis has been suggested to help treat anxiety, depression (Hall, 2019) and LSD research clearly shows that psychedelics have a much greater safety profile than other major addictive drugs (Das, 2016).

Main Hypotheses:

1. People tend to abuse alcohol more often than marijuana.
2. More men have substance abuse problems compared to women.
3. Race and substance abuse problems have no relationship in the sample.
4. Younger age groups have more substance abuse problems.
5. Highest completed education will not influence substance abuse problems.
6. Substance abuse positively correlates with mental struggles.
7. Substance abuse negatively influences health condition.

Additional Hypotheses:

8. People who have tried marijuana are more likely to try LSD.
9. Older people have stronger opinions about adults trying marijuana

The dataset

This dataset, taken from kaggle.com, consists of answers of the comprehensive survey on drug use and demographics in the United States. All column names are identical to the Question Index found in the NSDUH documentation. The values in each column are codes that correspond to a particular answer in the survey. References to each question's meaning can be found in the documentation here. The data is of 2 types: quantitative (such as age, first year of use) and categorical (yes/no questions, rating questions, etc.).

Data preparation

Originally the dataset consisted of almost 3000 columns. For this project not all of them are needed, therefore after reading the documentation, a selection of required columns was listed in separate file. In order to filter only the selected columns, a python script was used:

- At first, as the file was too big to process it all at once, it was divided into 6 smaller .csv files.

```python
import pandas as pd

chunk_size = 50000
batch_no = 1
for chunk in pd.read_csv(r'C:\Users\ramin\OneDrive\Desktop\NSDUH_2015-2019.csv',
                         chunksize=chunk_size, low_memory=False):
    chunk.to_csv('chunk'+str(batch_no)+'.csv', index=False)
    batch_no += 1
```

Figure 1.1

- Then from a smaller batch only the required data was written to a new file.

```python
import pandas as pd

data = pd.read_csv('chunk1.csv', low_memory=False)
my_file = open("final_col_list.txt", "r")
content = my_file.read()
selected_columns = content.split(",")
my_file.close()
new_data = data[selected_columns]
new_data.to_csv('chunk1_selected.csv')
```

Figure 1.2

- Eventually, all 6 selected smaller files were merged together to create the final dataset.

```python
import pandas as pd

# merging csv files
df = pd.concat(
    map(pd.read_csv, ['chunk1_selected.csv', 'chunk2_selected.csv',
                      'chunk3_selected.csv','chunk4_selected.csv',
                      'chunk5_selected.csv','chunk6_selected.csv']), ignore_index=True)

df.to_csv('final_dataset_drugs_2.0.csv')
```

Figure 1.3

Descriptive statistics

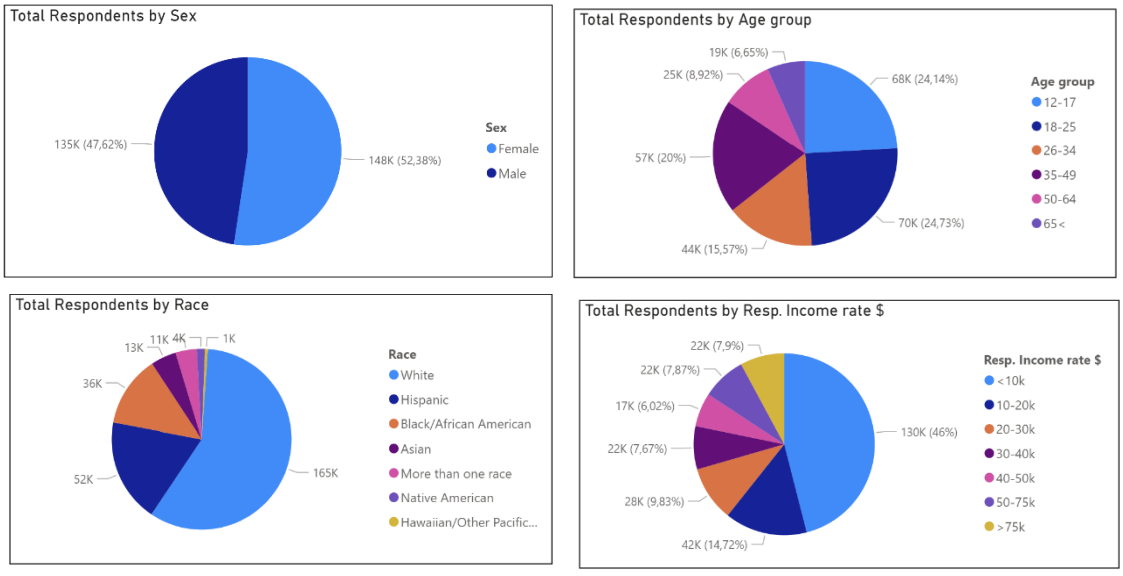

Figure 2.

Firstly, the main characteristics of the dataset are summarised. As it can be observed, the sex distribution is roughly equal. The respondents fall into six age categories, with the biggest group being 18–25-year-olds and 12-17-year-old in the second place. So, it should be noted that almost half of the participants are under 25. Moreover, the respondents come from seven different racial backgrounds and seven yearly income ranges. The most frequent yearly income rate being less than 10 thousand dollars could be explained by a large number of 12–25-year-old respondents.

Abstract Substance Use Statistics

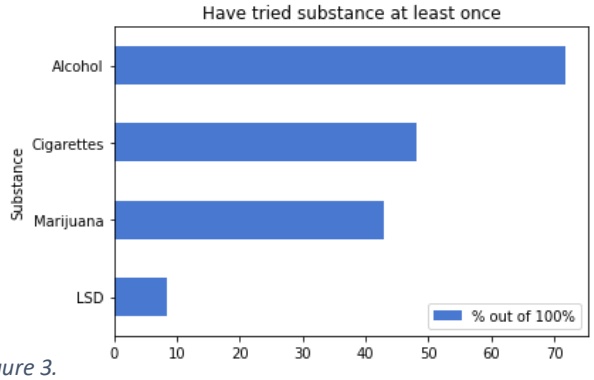

Figure 3.

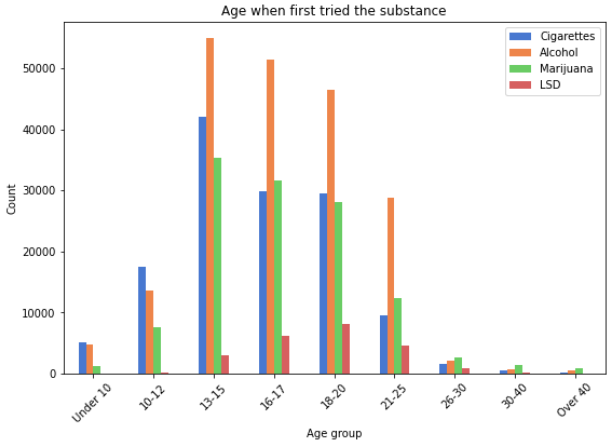

Figure 4.

To start, it was checked how many participants have tried each substance at least once in their life. Not surprisingly, over 70% have had alcohol – as it is very easily accessible and a dominant choice of intoxication in many cultures. Around half of the participants have smoked a cigarette or tried marijuana, while only 8 per cent have taken LSD. The illegality, questionable quality and societal stigma around psychedelics might be the reasons for the low percentage of 'acid' use.

In addition, it was thought that it might be beneficial to check at what age participants have tried the substances first. The ages 13-15 appear to be the time when most people experiment with cigarettes, alcohol, and marijuana for the first time in their lives. Occurrences of "first times with" alcohol or marijuana gradually decrease over the succeeding age groups. Interestingly, more children aged less than 13, tend to try cigarettes first compared with alcohol. Moreover, the substance that the participants tried for the first time, on average, when they were the oldest (18-20), is LSD.

Besides, originally, it was expected that by analysing the decades by the count of the first times, increasing/decreasing trends in substance use could be discovered. However, after plotting the data, the trend that was observed was that the frequency of first substance tries directly correlates with the age of the participants. The more present respondents in that decade – the higher the count. Nevertheless, a few observations can be made about LSD. First records of LSD use show up in the 1960s, which corresponds to the "The Psychedelic Era" of the mid1960s-mid1970s. Moreover, the upward trend of the "acid" popularity is very slight, except for the 1990s and 2010s when a more

significant increase in the count of first tries is noticed.

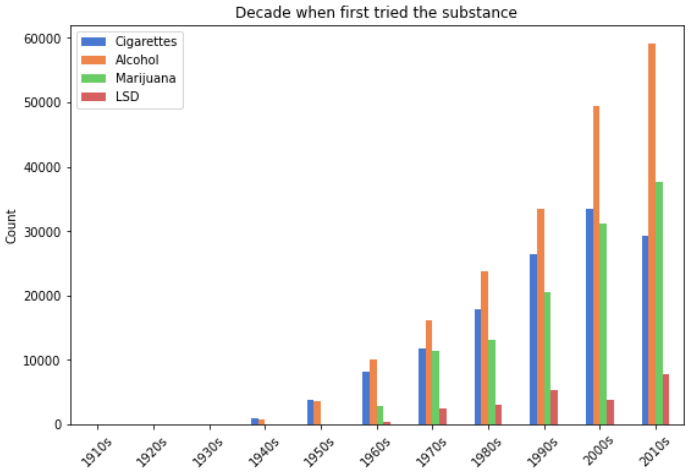

Figure 5.

Testing the hypotheses

1 and 2 hip.

People tend to abuse alcohol more often than marijuana.
More men have substance abuse problems compared to women.

Substances	% of respondents abuse	% men	% women
Alcohol	3.04596	60.7338	39.2662
Marijuana	0.859715	62.4434	37.5566

Figure 6.

After calculating the data, it was found that 3% of total respondents have had/ have alcohol abuse problems, by contrast, less than one percent reported having same problems with marijuana. In both cases of abuse, men make up around two thirds of the abusers. Therefore, both hypotheses were proven right.

3 hip.

Race and substance abuse problems have no relationship in the sample.

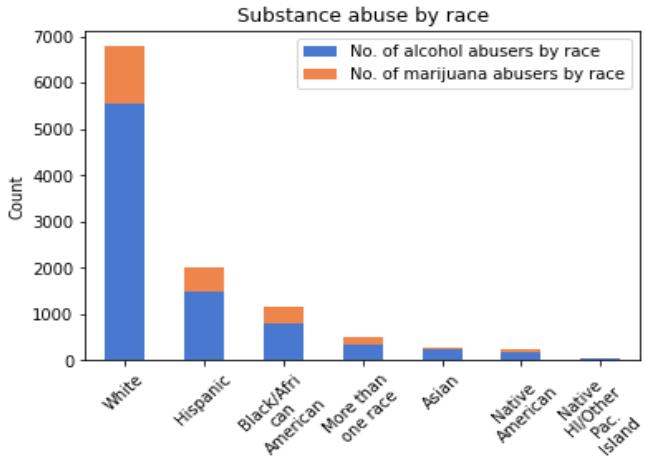

Figure 7.

After categorizing alcohol and marijuana abusers by race, it can be observed that the count of abusers in each race correlates with the number of participants of that race in the sample. Therefore, no direct relationship between the race and abuse is found. The hypothesis was correct.

4 hip.

Younger age groups have more substance abuse problems.

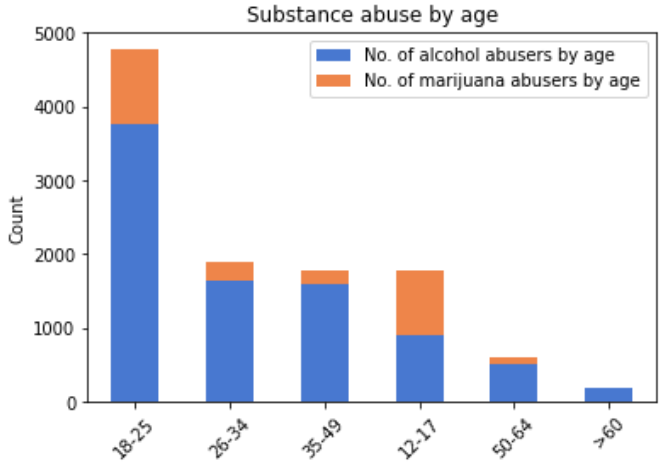

Figure 8.

Looking at the plotted data, it is evident that the age group having the biggest abuse problems is the youth aged 18-25. This could be linked to the youth culture which is strongly associated with alcohol and other illicit drug use. Besides, many young people at that age in the United States go away to study in college or university, in that way gaining a lot of freedom compared to the time of living from their parents, which results in higher rates of . alcohol or drug intake. Surprisingly, another significant group having marijuana abuse problems are 12-17-year-olds. However, the research has shown that teens are way more likely to develop an abuse problem (Winters KC, 2008). In fact, marijuana use disorder accounts for nearly 50% of admission for those ages 12 to 17 years who are receiving substance use disorder treatment, according to the NIDA. So, the hypothesis was correct.

5 hip.

Highest completed education will not influence substance abuse problems.

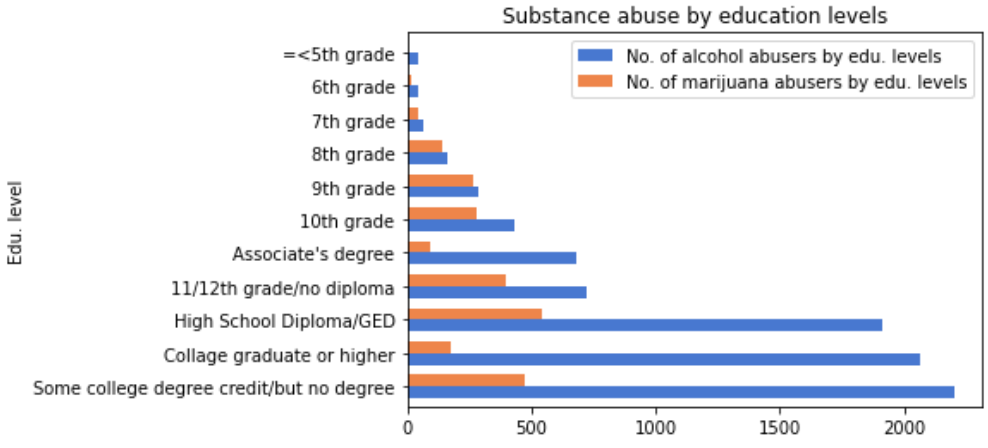

Figure 9.

Observing the substance abuse by education levels chart, it can be seen that alcohol abuse seems to increase sharply within the high school graduate participants and college graduate participants, both with or without having aquired a degree. But that does not necessarily mean that alcohol abuse correlates with higher education. As can be seen in 4 fig, which shows substance abuse by age, alcohol abuse increases within the age range of 18-25, which is the exact age of high school or college graduates. Considering marijuana abuse, similar conclusions can be made. The age group with the most marijuana abuse problems is 12-17 year-olds, whose educational levels span between 7th and 12th grade. The education level chart reinforces this claim because it can be seen that the number of marijuana abusers rises steadily starting from 7th grade. Therefore, the hypothesis was correct, as education levels do not influence abuse problems, but rather the age of the students at that education level.

6 hip

Substance abuse positively correlates with mental struggles.

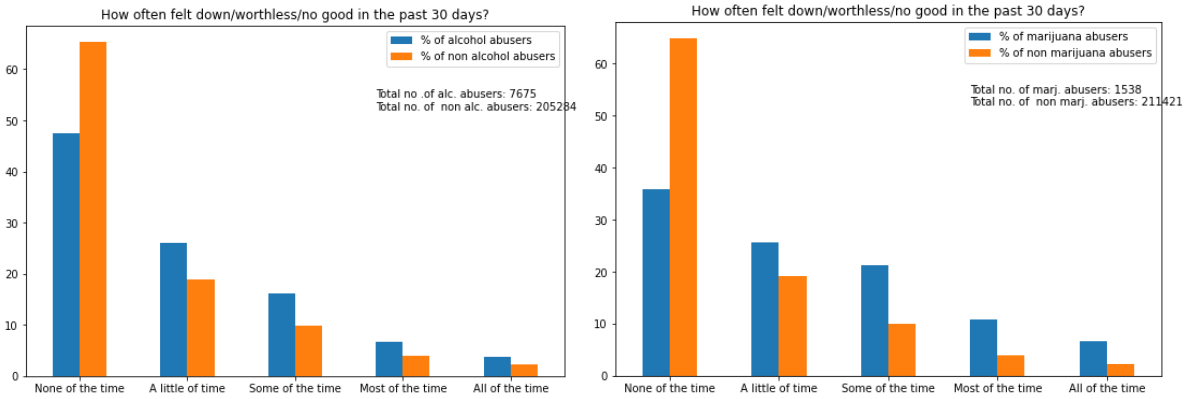

Figure 10.

One of the questions for adult participants was “How often have you felt down/worthless/no good in the past 30 days?”. A correlation between abuse problems and mental state can be observed by comparing the answers to this question from non-abusers and substance abusers. In both charts, the same tendencies are observed. Significantly more non-abusers have stated that they haven’t experienced those feelings in the past 30 days. The other four possible answers indicate that mentioned feelings were experienced, and to what extent. Participants with substance abuse problems have chosen those answers more frequently, therefore, as a result, it can be stated that the hypothesis was correct and there is a relationship between worse mental state and abuse problems.

7 hip.

Substance abuse negatively influences health condition.

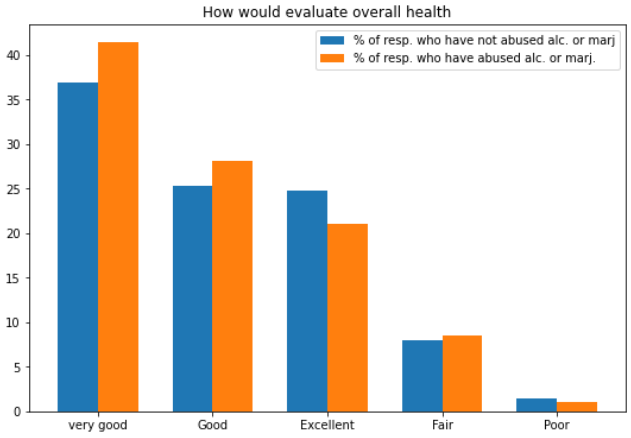

Figure 11.

Analysing the substance abuse influence on health condition graph, it can be seen that almost all participants believe to be in great health. This, however, might not reliable because the participants can have a skewed perception of themselves or have a biased opinion. Although, one significant factor is that the number of healthy substance abusers is higher than that of non-substance abusers. Although it is possible that they are overcompensating. This graph shows that substance abuse does not have a negative impact on overall health, contrarily, it has a positive impact on overall health. Therefore, the hypothesis was wrong, although more thorough research is required to validate this statement.

8 hip

People who have tried marijuana are more likely to try LSD.

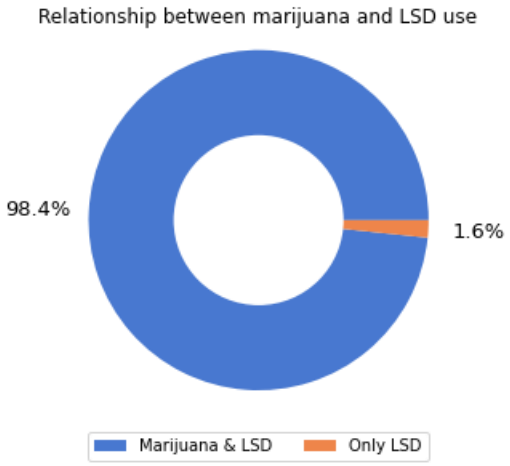

Figure 12.

As the chart visualizes only around 2% of respondents who have taken LSD say, they have never tried marijuana. The relationship between the two might be explained by the fact that marijuana is a ‘easier’, not as strong, more accessible drug, therefore more popular. However, after trying cannabis and getting familiar with its psychoactive properties, people’s ‘interest might be peeked’ and they might go one step further to try stronger – psychedelic drugs. The result proves the hypothesis to be correct.

9 hip

Older people have stronger opinions about adults trying marijuana

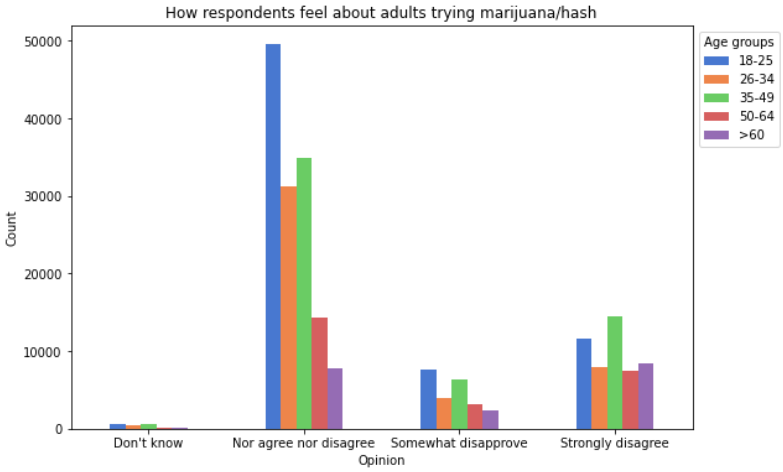

Figure 13.

After categorizing respondents' opinions about adults trying marijuana by age groups, it is obvious that majority of participants do not have a strong stand on the subject. Surprisingly, quite a lot of 18-25, 34–45-year-olds strongly disapprove of cannabis use. Moreover, the opinions of individuals aged more than 60 years old, split almost equally between 'nor agree nor disagree' and 'strongly disagree', with a slightly higher percentage choosing the later. This proves the hypothesis correct, however not to the expected extent.

Results

The insight extracted from the dataset:

- Participants abused alcohol more often than marijuana.
- Men were more likely to have a substance abuse problems.
- None of the seven races has any impact on drug abuse.
- Younger people are more prone to substance abuse problems.
- Highest completed education level itself has no impact on drug abuse, but the age of the students at that education level does.
- Substance abuse correlates positively with mental struggles
- The abuse of alcohol or marijuana seemingly has a positive effect on overall health of the participants.
- Almost all of the participants who have tried LSD, have used cannabis.
- The eldest age group has the biggest number of respondents who strongly disagree with marijuana use in adults.

Future Work

After the completion of the project and its critical evaluation, I would like to reflect on what could be improved. First, I would like to include a correlation heatmap for the variables as an additional evaluation tactic to see if reached conclusions about hypotheses are correct. However, I am not quite sure how to do that with different types of variables yet, in this case, quantitative and categorical data. Also, in the future, I will create a machine learning model to predict alcohol/marijuana use and abuse given some of the person's details (sex, age, mental state, etc.).

Bibliography

americanaddictioncenters.org. (2022, June 30). Retrieved from https://americanaddictioncenters.org/adult-addiction-treatment-programs/most-addictive

Das, S. e. (2016). “Lysergic acid diethylamide: a drug of 'use'?.” . *Therapeutic advances in psychopharmacology vol. 6,3*, 214-28.

Hall, W. H. (2019). Cannabis use and mental health: risks and benefits. *Eur Arch Psychiatry Clin Neurosci*.