The Effect of Drug Abuse on Mental Health

Milestone Report

Raj Singh

Table of Contents

Problem Statement

We still don't know how exactly certain drugs are related to development of mental health-related issues. Drugs such as alcohol, marijuana, cocaine, and hallucinogens are frequently used by college students. Use of these substances is sometimes carried into their young adult lives after they graduate. At the same time, mental health disorders (such as anxiety and depression) begin to develop for certain individuals. If we could establish a strong correlation between psychological distress and substance use, we could alert young adults that use of these substances could have deleterious effects on their mental health. This could have applications in the treatment of mental health disorders as well; people may be delighted to hear that they could alleviate their depression simply by quitting their use of drugs. On the other hand, if we discover that use of substances has little correlation with development of these disorders, we could choose to focus our treatment efforts elsewhere.

The Data

The data comes from the Substance Abuse and Mental Health Services Administration. Specifically, it comes from the 2015 National Survey on Drug Use and Health. Because we are mostly interested in the effect that drugs have on young adults, we will focus on individuals between the ages of 18 and 25. There are several factors that the survey uses to delineate substance use, such as the number of days that the individual has used the substance in the past year. There are two dependent variables that we will focus on in particular: the survey has devised its own sophisticated metrics to indicate whether an individual is psychologically distressed, in addition to whether the individual has had a major depressive episode in the past year. The data comes with a codebook that describes how the data was collected and how certain metrics were calculated; it also clarifies the meaning of all ambiguous numbers within the dataset. Hopefully, this will make data cleaning a relatively simple process.

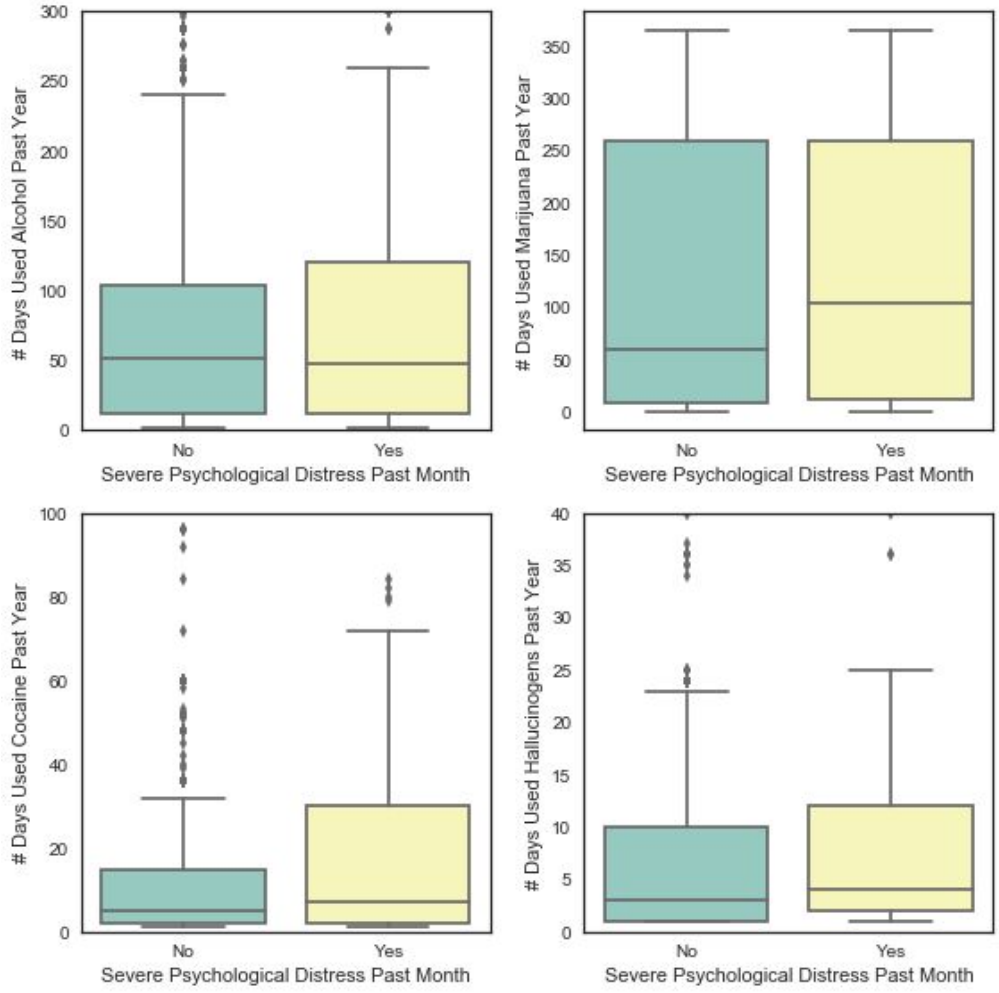
Data Cleaning

First, we must import the data by using tabs as the delimiter. Then, we will make a Python dictionary corresponding the column names of the .tsv file that we are interested in with the meaning of the code according to the codebook provided by the Substance Abuse and Mental Health Services Organization.We use the keys of the dictionary to boolean index the DataFrame so we get a new DataFrame with the columns that we are interested in. We then use a renaming function to rename the columns so that we can see what each column name really means. Then, we filter the age to make sure that we are only getting the observations inside the age range of 18-25. What follows was the most time intensive step thus far: we must to replace the values in the DataFrame based on their meaning according to the codebook. We have to do this for all 47 columns. Then, we inspect the numerical data for outliers that would definitely be bad data. All of the outliers that we find make sense in the context of the problem, and so there is no real need to eliminate any of the outliers. Finally, we write the cleaned data to a new .csv file. The details of these cleaning steps can be found here

Exploratory Data Analysis

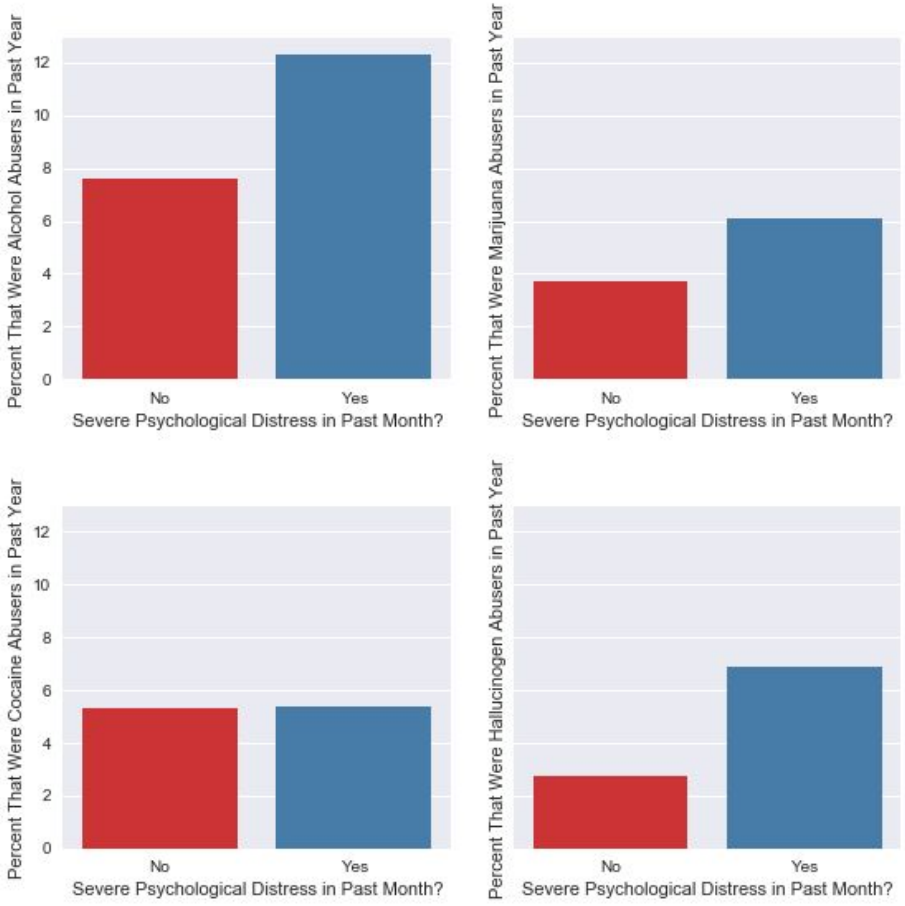
Here, we will create some visualizations to try and investigate any trends and/or patterns that may be useful to look into during our analysis. Keep in mind, our goal is to find out if excessive substance abuse has a correlation with development of mental illnesses and psychological distress. Whether one causes the other is yet to be seen, but we can certainly learn if the two are associated by exploring the data further. Once we establish this correlation, we can further investigate (perhaps using whether or not first use of the drug was at a young age) whether this drug use caused their psychological distress. Then, perhaps, we can use this as a reason to caution adolescents against using drugs. Or, if we find no definitive correlation/contribution, we can choose to focus our efforts to improve worldwide mental health elsewhere.

Severe Psychological Distress (Past Month) and # of Days that Substance Was Used (Past Year)



As we can see from the above visualization of boxplots, marijuana is the only substance where there was a noticeably higher median '# Days Used in Past year'. Given that we did not find much correlation in the remaining substances, we will choose to move along with our analysis without doing any hypothesis testing. Much of the exploratory analysis on fields involving the number of days that the substance was used in the past month or year yielded insignificant results. Thus, we will choose not to highlight any more of these findings.

Severe Psychological Distress (Past Month) and Substance Abuse (Past Year)



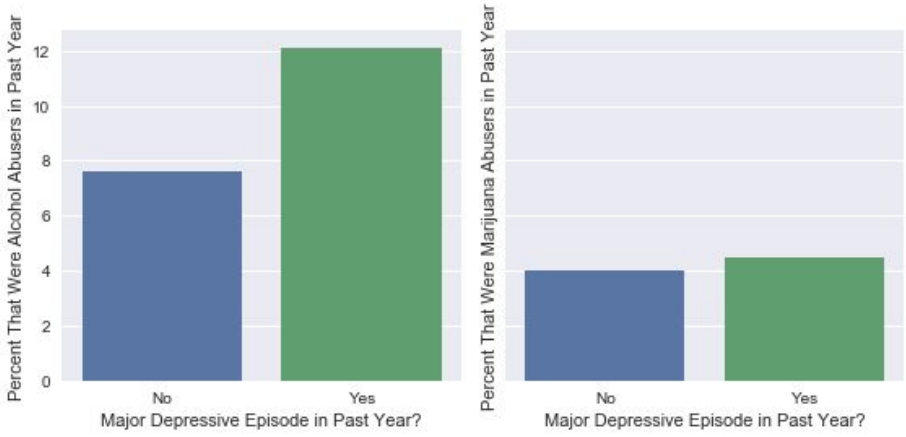
Here, we took the subsets of people with and without severe psychological distress in the past month, and compared the percent of them that were abusers of the substance. The results are shown in the above barplot.

As we can see, there is a higher percentage of abusers for every substance in the subset of people with severe psychological distress in the past month. This seems like a trend worth investigating.

So, we set up two-sample proportion z-tests for each substance. The details of performing these tests can be found here under the heading "Applying Inferential Statistics." In the cases of alcohol and marijuana abuse, we reject the null hypothesis at the $\alpha = 0.01$ level. We can conclude that the proportion of alcohol and marijuana abusers with severe psychological distress in the past month is significantly greater than that of people without severe psychological distress in the past month. With cocaine abuse, we fail to reject the null hypothesis and conclude that there is no significant difference. When performing this test on the proportion of hallucinogen abusers, we get a p-value of 0.016. So, we fail to reject the null hypothesis at the $\alpha = 0.01$ level, but we reject it at the $\alpha = 0.05$ level. There does indeed seem to be a correlation between substance abuse and severe psychological distress.

Since we got clear statistical significance with alcohol and marijuana, let's investigate how abuse of these substances relates to major depressive episodes in the past year.

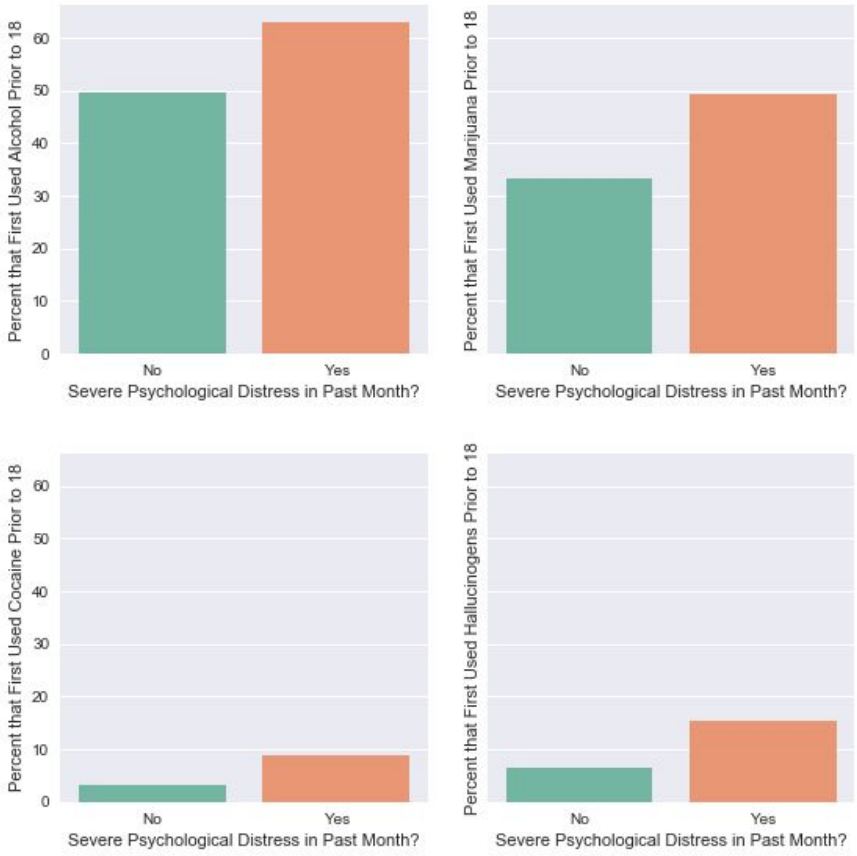
Major Depressive Episodes (Past Year) and Alcohol/Marijuana Abuse (Past Year)



In both cases, there is a higher percentage of individuals that have abused the substance amongst those who have had a major depressive episode in the past year.

We will take the sample approach that we did earlier and perform two-sample proportion z-tests. With alcohol abuse, we reject the null hypothesis at the $\alpha = 0.01$ level and conclude that there is a significant difference between the proportions of alcohol abusers amongst those with and without MDE's in the past year. In the case of marijuana abuse, however, we fail to reject the null hypothesis. This is an interesting finding because we rejected the null hypothesis with high certainty in regards to those with severe psychological distress. The details of this analysis can be found here under the heading "Alcohol Abuse on Major Depressive Episode." Let's continue to investigate more trends in the data.

Severe Psychological Distress (Past Month) and Trying Substance Prior to 18



As we can see from the above barplots, there is a higher percentage of individuals that have tried the substance prior to 18 amongst those with severe psychological distress in the past month.

We will use a two-sample proportion z-test again. With every substance, we reject the null hypothesis at the $\alpha = 0.01$ level: thus, We can conclude that the proportion of those who have used each substance prior to 18 amongst those who have had severe psychological distress in the past year is significantly greater than that of individuals that have not had psychological distress in the past month. The details of this analysis can be found here under the heading "Substance Use before 18 - Effect on Psychological Distress."

We are seeing that excessive substance use (as measured by a variety of metrics) does have a strong correlation with severe psychological distress.

Conclusion

The above findings are only a part of the many visualizations and extensive statistical analysis performed in this exploratory data analysis notebook. There were other findings, including the fact that the mean psychological distress score (those with scores greater than or equal to 13 were classified as having severe psychological distress) was significantly higher for those who had abused all substances and for those who had first used all substances prior to 18. We chose not to focus on those findings because we will treat this as a supervised classification problem, rather than a regression problem. Nonetheless, many of the findings corroborate the fact that excessive substance use can be used to predict whether or not someone is experiencing severe psychological distress. If we can create a model to successfully predict whether or not someone is experiencing severe psychological distress based on their substance use, we can focus our efforts on getting these individuals the help that they need. Furthermore, we can improve our understanding of the relationship between substance use and psychological distress.