

Drugs most closely correlated to age

Tyler Caudy

CMSE381

Final Project

Project Category: Drug use by age

<https://github.com/fivethirtyeight/data>

Final Report

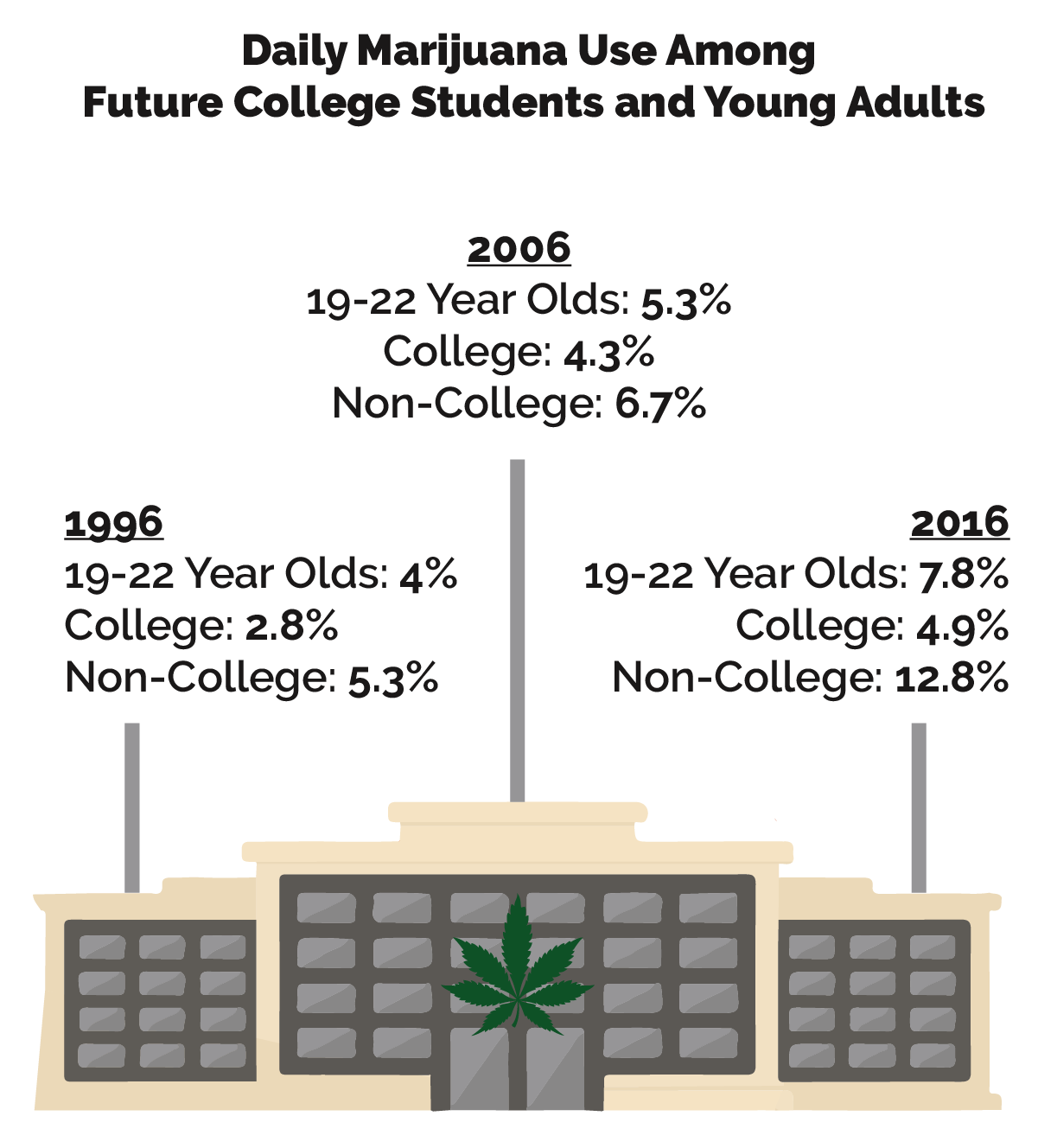
1 Introduction

Whether it’s a prescription, medical card, or brown bag with a bottle, this world is becoming more and more dependent on drugs. Pharmaceutical companies have boomed in the last century and prison time for a joint is now considered unjust. The general consensus of society is shifting, or in other words, the perception on drugs of the average person is changing. This means that all age groups are shifting, younger generation included, some more volatile than others.

Which drugs in the given dataset have no significance to age? Are there drugs that have increased usage and frequencies as someone grows up? At around which age does this habit start? Is there even enough data to conclude reasonable assumptions? These are some of the important questions to consider when assessing the results of the data.

2 Related Work

Studies done by the Mission Harbor Behavioral Health suggest that marijuana use in college students has nearly doubled in the last twenty years. Stimulants such as e-cigarettes and hookahs are becoming more transparent in the younger generation as the connotation behind cigarettes hinders compared to older-younger generations.



3 Methods

Since this data is using age as the grouping, an unsupervised learning approach would be most suitable. My original focus was to incorporate popular unsupervised algorithms such as a support vector machines, tree clusters, and even principal component analysis. However, after hours of try and fail I realized that these algorithms are not suitable for a dataset such as this. My understanding is that there either isn’t enough data with the dataset we were given or there were too many groups to consider when trying to analyze through methods such as the listed above.

The support machine vector is a generalization of the maximal margin classifier. This requires that the classes be separable by two groups. This is why SVM’s are popular in image processing models and explains how the classifier would be unsuccessful with our age grouped data. A boundary separating the age groups in half would only distort age significance. Tree clusters would perform the same way as after each iteration the root node is branched into two.

4 Results

Deviating from the required project classifiers which I was unable to perform, I decided to visualize the statistical significances to the best of my ability.

Background pattern

Description automatically generated

Figure 1: Spread of frequency correlation

Chart

Description automatically generated with medium confidence

Figure 2: Spread of usage correlation

For the initial data analysis, we can see that alcohol and marijuana tendencies are the strongest in the increasing age groups. Other drugs with lower correlation to age suggest exactly that. This includes pain relievers, OxyContin, and meth. Either there isn't enough data to conclude a significance, or there is no correlation whatsoever with respect to age.

Research of inhalants online suggest that they are drugs pertaining to aerosol sprays and gases. We can see that that the frequency has a strong negative correlation with respect to age, which suggests that as these participants get older, they are much less likely to get involved in these kinds of acts. With this data in mind, one suggestion can be that inhalants are one of those drugs that kids stupendously decide to do at a young age, and consequently regret. This sort of data is not predictable in any way and is a good example of one variable that has no significance to age. Other drugs include meth and pain killers.

Chart, scatter chart

Description automatically generated

Figure 3: Alcohol vs Age

Chart, line chart, scatter chart

Description automatically generated

Figure 4: Marijuana vs Age

Figures four and five look at the alcohol and marijuana usages over each age. The younger age groups appear to be more closely related to one another whereas a separation occurs in the older age groups. This suggests that as people reach their retirement age, alcohol users are less likely to subside in habits compared to marijuana users.

The graphs which look at the frequencies for the two drugs can be seen in the code markdown. These graphs show that the frequency for marijuana and alcohol users increases until age ~twenty, then starts to consolidate.

Diagram

Description automatically generated with medium confidence

Figure 5: Drug classifiers vs Age

When looking at the groupings of the various drugs we can see that hallucinogens (marijuana), sedatives (alcohol), and stimulants (cigarettes) are most correlational to age. This confirms our observation from the previous figures when looking at the statistics of the single drug itself.

5 Conclusions and Future Work

This project’s goal was to identify any statistics significances with the data set given and find any correlations of drugs and their categorical counterparts with respect to age.

On class of the 16th, the professor had used the drugs dataset to tell us that unsupervised learning approaches would be most suitable for this task. I stood up to this task but was quite unsuccessful. The number of support vectors I was able to obtain were insignificant. My trees had no branches due to the already predetermined age groupings. Even then, a binary age classifier would’ve only furthered the age distinction between the groups.

Improvements for the future:

* A larger dataset with more studies.
* Yearly studies in dataset to measure changes in habits over time. We could then predict future changes.
* Give myself more time to account for failed classifiers.

Conclusions:

* Alcohol and marijuana are most closely attributed to age.
* This dataset was lackluster.

Bibliography

[1] “Substance Abuse Addiction and Age Groups Guide.” *Mission Harbor Behavioral Health*, sbtreatment.com/addiction-and-age-groups-guide/.

[2] Fivethirtyeight. “Fivethirtyeight/Data.” *GitHub*, github.com/fivethirtyeight/data.