

Comparing Drug Use Between Age Groups: A Predictive Analysis

Group 7: Jade Chen, Jessica Luo, Heidi Lantz, Nazia Edroos

Outline

Summary	2
Introduction:	2
Methods & Results:	2
Data Cleaning	3
EDA	3
Analysis	3
Discussion:	3
Impact of our findings	4
Future directions	4
References:	4

Summary

(heidi will enter information here)

Introduction:

- provide some relevant background information on the topic so that someone unfamiliar with it will be prepared to understand the rest of your report
- clearly state the question you tried to answer with your project
- identify and describe the dataset that was used to answer the question

This dataset is sourced from [FiveThirtyEight's repository](#), which is derived from the [2013 National Survey on Drug Use and Health](#), conducted by the United States Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality.

The dataset covers self-reported drug use trends across 17 age groups in the United States, examining 13 different substances.

Disclaimer: This analysis is distinct from the FiveThirtyEight article [How Baby Boomers Get High](#). Our study focuses on a different research question, examining whether adolescent alcohol and marijuana use (ages 12-19) can predict hard substance use in early adulthood (ages 20-29) using predictive modeling.

Predictive Question

Can patterns of alcohol and marijuana use in early adolescence predict hard substance use behaviors in early adulthood (ages 20-29)?

Methods & Results:

(will be done by Jessica and Jade)

- describe in written english the methods you used to perform your analysis from beginning to end that narrates the code the does the analysis.
- your report should include code which:
 - loads data from the original source on the web
 - wrangles and cleans the data from it's original (downloaded) format to the format necessary for the planned classification or clustering analysis
 - performs a summary of the data set that is relevant for exploratory data analysis related to the planned classification analysis

- creates a visualization of the dataset that is relevant for exploratory data analysis related to the planned classification analysis
 - performs classification or regression analysis
 - creates a visualization of the result of the analysis
- note: all tables and figure should have a figure/table number and a legend

Data Cleaning

1. Handle missing values by removing rows with excessive missing data or imputing reasonable values.
2. Check format of variable and transform as needed.
3. Check for outliers in age-related variables.
4. Standardize or normalize predictor variables as needed.

```
# read in the data
data <- read.csv("data/drug-use-by-age.csv")
```

EDA

Talk about EDA yada yada.

Analysis

Split into test/training set.

Use linear regression, blah blah blah. Perform correlation analysis to identify significant predictors before modeling. Evaluate model performance using accuracy or RMSE.

Discussion:

(will be done by Nazia)

What we found:

summarize what you found

Does this align with our expectations?

discuss whether this is what you expected to find?

Impact of our findings

discuss what impact could such findings have?

Future directions

discuss what future questions could this lead to?

References:

at least 4 citations relevant to the project. (heidi will insert this)