Week 03: Final Project Proposal
Section 2, Team 4

Names and Emails
- Andres Salcedo (asalcedo@ischool.berkeley.edu)
- Cassidy Bruner (cassidybruner@berkeley.edu)
- Eyup Agtepe (eyupagtepe@ischool.berkeley.edu)
- Gabrielle Doran (gabrielle_doran@ischool.berkeley.edu)

GitHub Repository: https://github.com/cassidy-bruner/mids-207-final-project-team-4.git

Project Description
This project aims to develop a predictive model to assess whether a survey respondent is likely to struggle with alcohol abuse based on their reported adverse childhood experiences and socioeconomic background. Key predictors include early-life stressors such as family disruption (like parental separation or incarceration) or experiences of poverty (like food insecurity).

Motivation
Alcohol abuse has significant public health, social, and economic costs. Understanding how early-life adversity shapes risk is crucial for timely prevention. This project will identify key factors linked to alcohol misuse, offering predictive insights and highlighting vulnerable populations to help policymakers, providers, and community organizations target resources more effectively.

Data
We used data from the CDC's 2023 Behavioral Risk Factor Surveillance System (BRFSS), a nationally representative cross-sectional survey of U.S. adults that has been conducted since 1984. The survey is administered via telephone interviews and collects information on a wide range of health-related risk behaviors, chronic health conditions, and preventive service use. For our analysis, we plan to focus on variables that capture Adverse Childhood Experiences (ACE) and psychosocial factors, such as, parental violence, physical or sexual abuse, parental divorce, incarceration of household members, and perceived safety during childhood. We also include behavioral and health-related indicators such as age of smoking initiation, education level, unmet basic needs, number of days with poor mental health, and self-reported poor physical or mental health. These predictors will be used to model risky alcohol use, with outcome variables defined as average alcoholic drinks per day in the past 30 days and binge drinking frequency.
Features for modeling (Adverse Childhood Experiences): Feature label, column#

1. How Often Did Your Parents Beat Each Other Up, 596
2. How Often Did A Parent Physically Hurt You In Any Way, 597
3. How Often Did A Parent Swear At You, 598
4. How Often Did Anyone Ever Touch You Sexually, 599
5. Did an adult make you feel safe and protected, 602
6. Were Your Parents Divorced/Separated?, 595

7. Live With Anyone Who Served Time in Prison or Jail, 594
8. Education Level, 187
9. How old when you first started smoking?, 276-278
10. During the past 30 days, on how many days did you use marijuana or hashish?
11. Did you smoke marijuana or cannabis, 585
12. Live With Anyone Depressed, Mentally Ill, Or Suicidal?, 591
13. Did an adult make sure basic needs were met, 603
14. Live With a Problem Drinker/Alcoholic?, 592
15. Live With Anyone Who Used Illegal Drugs or Abused Prescriptions?, 593
16. How do other people usually classify you in this country?, 623-624
17. Number of Days Mental Health Not Good, 104-105
18. Poor Physical or Mental Health, 106-107

- Features to create target column: Feature label, column#
    1. Avg alcoholic drinks per day in past 30, 232-233
    2. Binge Drinking, 234-235

Related Work

Recent studies have shown that there is an increasing interest in using machine learning methods to better understand the connection between substance use disorders and traumatic childhood experiences. To predict substance abuse risk, studies have employed a variety of data driven techniques, with Jing et al. (2020) using Random Forest algorithms to achieve 74-86% accuracy, while Leza et al. (2021) conducted systematic reviews that found strong associations between substance use later in life and childhood adversity. Song et al. (2023) demonstrated the efficacy of analyzing health outcomes associated with adverse childhood experiences using BRFSS data. Building on these studies, our project will investigate the predictive relationship between risky adult drinking behaviors and adverse childhood experiences by applying machine learning methods to BRFSS data.

Jing, R., et al. (2020). Analysis of substance use and its outcomes by machine learning I. Childhood evaluation of liability to substance use disorder. https://pmc.ncbi.nlm.nih.gov/articles/PMC6980708/

Leza, L., et al. (2021). Adverse childhood experiences (ACEs) and substance use disorder (SUD): A scoping review. https://www.sciencedirect.com/science/article/abs/pii/S0376871621000582

Song, E., et al. (2023). The association between adverse childhood experiences and mental health wellbeing during adulthood. https://jphe.amegroups.org/article/view/9233/html