# Project Overview

This is data collected from a program designed to help people with substance abuse – addiction to drug and alcohol.

The people who went through the program (clients) filled out a survey asking them about their substance use, their living situations, and their well-being. This survey was conducted once at the start of the program, and once at the end of the program. Occasionally, clients got multiple surveys. This data is in your **Survey Dataset.**

We also collect a few more details about where clients lived, which is in your **Client Dataset.**

During the program, clients got services. These services ranged from counseling to to massage to therapy. The details of these services are in your **Service Dataset**.

*Note, this is all real (though anonymized) data from a project called \*\*\*\*\*\*\*\*\*\*.*

# 

**Our goal is to use this data to help the program managers determine:**

1. Was this program effective? (hint, compare the very first and very last survey)
2. Was it effective for short term outcomes, like helping people decrease their use of alcohol and drug?
3. Was it effective for mid term outcomes, like helping people get arrested less or commit fewer crimes?
4. Was it effective for long term outcomes, like helping find more stable living situations or get employment?
5. Was it effective for everyone, or just some groups of people? If so, who was it effective for, and who wasn’t it? (hint, use demographic and geographic details)
6. Did certain factors predict better outcomes? For example, the cost of a service, the type of a service, a person’s age, etc? (hint, use classic statistics to compare different categories or feature importance in your ML model)

# Data

## Survey Data

**Fields:**

‘ClientId’ - the unique identifier for each client

'SurveyDate’, - when the survey was done

'InterviewType’, was it Intake (before they started the program) or Post (after they finished)

'DOB’, - date of birth

'Gender’, - gender

‘Ethnorace’ - ethnicity and race

‘Veteran’ - are they a veteran?

‘C1LivingSituationId’  - "in the past 30 days, where have you been living most of the time?”

‘D2EducationLevelId’ - "what is the highest level of education you have finished?”

‘D3EmploymentSituationId’ - "are you currently employed?"

‘B1aAlcohol’ - "During the past 30 days, how many days have you used any alcohol?"

‘B1cIllegalDrugs’ - "During the past 30 days, how many days have you used illegal drugs?"

‘C2AddictionStressProblemId’ -  “During the past 30 days, how stressful have things been for you because of your use of alcohol or other drugs?”

‘C4AddictionEmotionalProblemId’ – “During the past 30 days has your use of alcohol or other drugs caused you to have emotional problems?”

‘E1TimesArrested' – “In the past 30 days, how many times have you been arrested?”

'E4CommitedCrime’, “In the past 30 days, how many times have you committed a crime?”

‘F5aDepression’ -  “In the past 30 days, (not due to your use of alcohol or drugs) how many days have you: experienced serious depression”

’F5bAnxiety’ - “In the past 30 days, (not due to your use of alcohol or drugs) how many days have you:  experienced serious anxiety or tension”

**General Coding notes:**

#No = 0

#Yes = 1

#Missing = -9

#Don't know = -8

#Refused = -7

#Skipped = -1

**C1LivingSituationId**

living = {

    "1":"Shelter",

    "2":"Street/Outdoors",

    "3":"Institution",

    "41":"Housed - Own/Rent Apartment",

    "42":"Housed - Someone else's appt",

    "43":"Housed - Halfway house",

    "44":"Housed - Residential Treatment",

    "45":"Housed - Other",

    "46":"Housed - Dormitory/College residence" }

**D2EducationLevelId**

Edu = {0: "Never attended", 1: "1st Grade", 2: "2nd Grade", 3: "3rd Grade", 4: "4th Grade", 5: "5th Grade", 6: "6th Grade", 7: "7th Grade", 8: "8th Grade",

9: "9th Grade", 10: "10th Grade", 11: "11th Grade", 12: "12th Grade/High School Diploma/Equivalent", 13: "College or University / 1st Year Completed",

14: "College or University / 2nd year completed (AA,AS)", 15: "College or University / 3rd Year completed", 16: "Bachelor's Degree (BA, BS) Or Higher",

17: "Voc Tech Program after high school no diploma", 18: "Voc Tech Diploma after high school”}

**D3EmploymentSituationId**

employment = {0: "Other", 1: "Employed full time", 2: "Employed part time", 3: "Unemployed, Looking for work", 4: "Unemployed, Disabled", 5: "Unemployed, Volunteer work",

       6: "Unemployed, retired", 7: "Unemployed, not looking for work”}

**C2AddictionStressProblemId and C4AddictionEmotionalProblemId**

problem = {1: "Not at all", 2: "Somewhat", 3: "Considerably", 4: "Extremely", 5: “Refused”}

## Service Data

'ClientID' – ID for the Client

'ProviderId' – ID for the provider

'ServiceDt' – the date of the service

'ServiceAmt' – the cost of the service

'Service' – the name of the service

'ServiceDesc’ – description of the service

## Client Data

'ClientId' – ID for the Client

‘City’ – the city the client lives

‘State’– the state the client lives

‘Zip’ – the zip code of the client