Substance Use & Effect in Public Health – Data Analysis (NSDUH 2019)

Introduction

This project leverages the 2019 National Survey on Drug Use and Health (NSDUH) dataset — a comprehensive, nationwide survey administered by the Substance Abuse and Mental Health Services Administration (SAMHSA). The survey gathers data on substance use, mental health, and behavioral health indicators across the U.S. population.

For this analysis, I focused on a subset of 10,915 individual responses from two counties in Oklahoma. The data includes demographic details, substance use patterns (alcohol, marijuana, cocaine, tobacco), and self-reported mental and physical health statuses.

Approach

To manage and analyze the large NSDUH dataset, I used Excel, NoSQL, and Power BI, each playing a key role in the process:

- Used **excel** it to break the massive dataset into smaller, manageable sub-sets, clean up duplicates, and filter out irrelevant or missing data.
- Used **NoSQL (SQL Server)** then, based the size and structure of the dataset (over 1,000 columns), I used NoSQL techniques within SQL Server to: Decode survey codes into human-readable values using the official NSDUH codebook
- Joined relevant tables (e.g., alcohol use, mental health, physical health, demographics) using respondent IDs Transform and restructure the data into star schema-like formats for smoother integration into Power BI
- **Power BI** was used to build interactive dashboards. This helped me visualize patterns — like how substance use varies across different age groups and genders, and how it links to mental health conditions and chronic illnesses.

Challenges

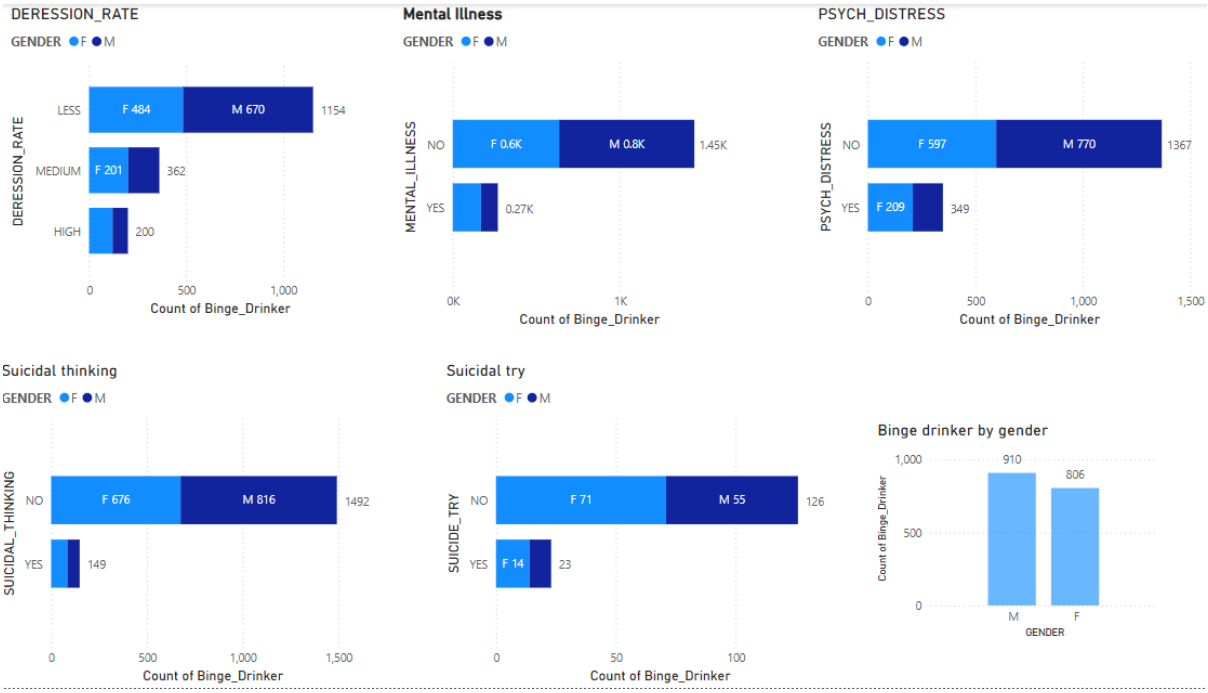
- One of the biggest challenges was the size and complexity of the dataset. With over 10,000 columns and thousands of records, initial data cleaning was time-consuming — especially when dealing with duplicate entries, inconsistent formatting, and numerous invalid or error values.
- Another major hurdle was that all responses were coded numerically, meaning I had to continuously refer to the official NSDUH codebook to decode each variable into

meaningful, readable text. This step was crucial to ensure the accuracy of insights, but it added extra layers to the cleaning process.

- Since the data is based on self-reported survey responses, some fields appeared incomplete or potentially unreliable, which meant making informed decisions about what to keep or discard was necessary before moving forward with analysis.

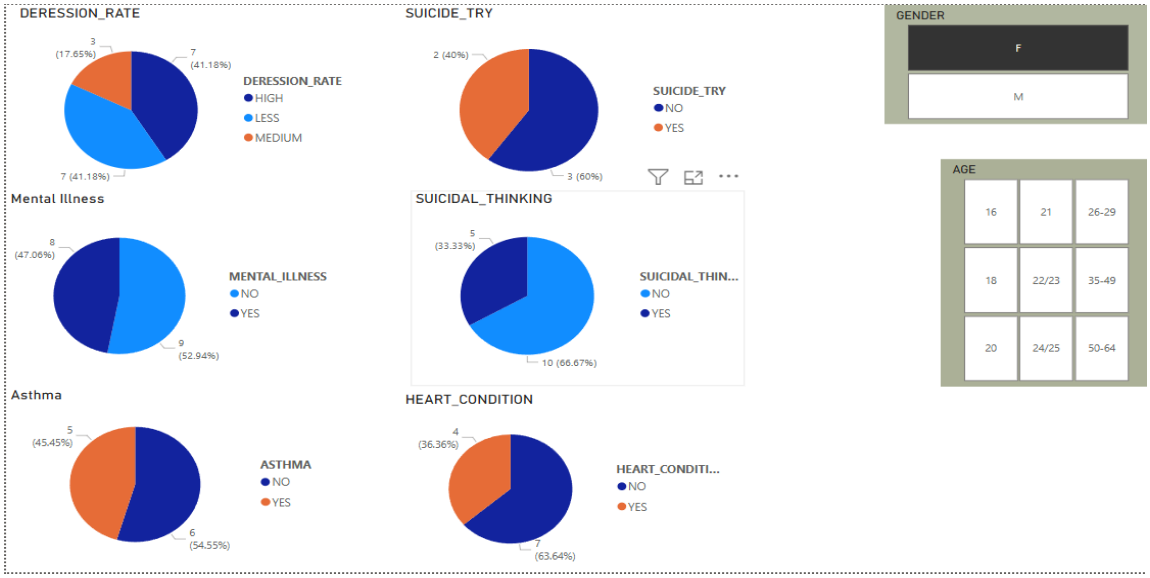
Result

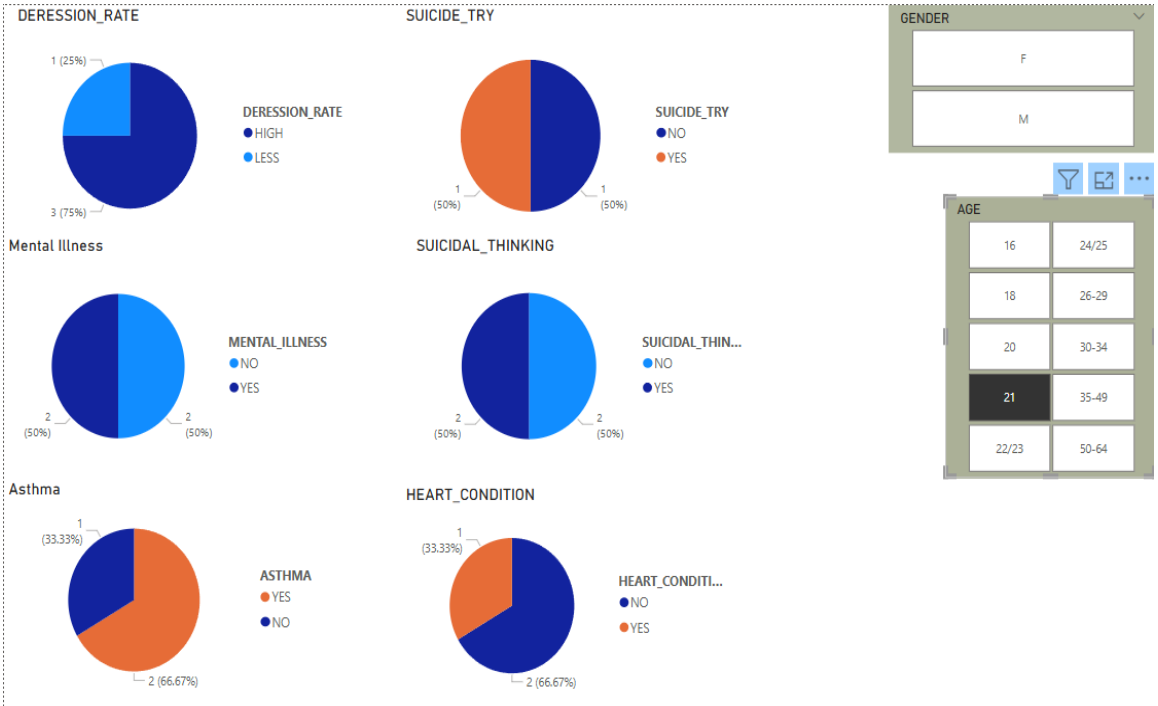
Binge drinking# mental disorders



- Female binge drinkers showed 10% higher rates of depression, serious mental illness, and suicidal thoughts compared to males — despite being fewer in number.
- Binge Drinking & Age: The highest number of binge drinkers fall within the 35–49 age group, with males making up the majority by approximately 5%.

Cocaine use# health





Among frequent cocaine users (more than 2 days/week), women reported higher health impacts:

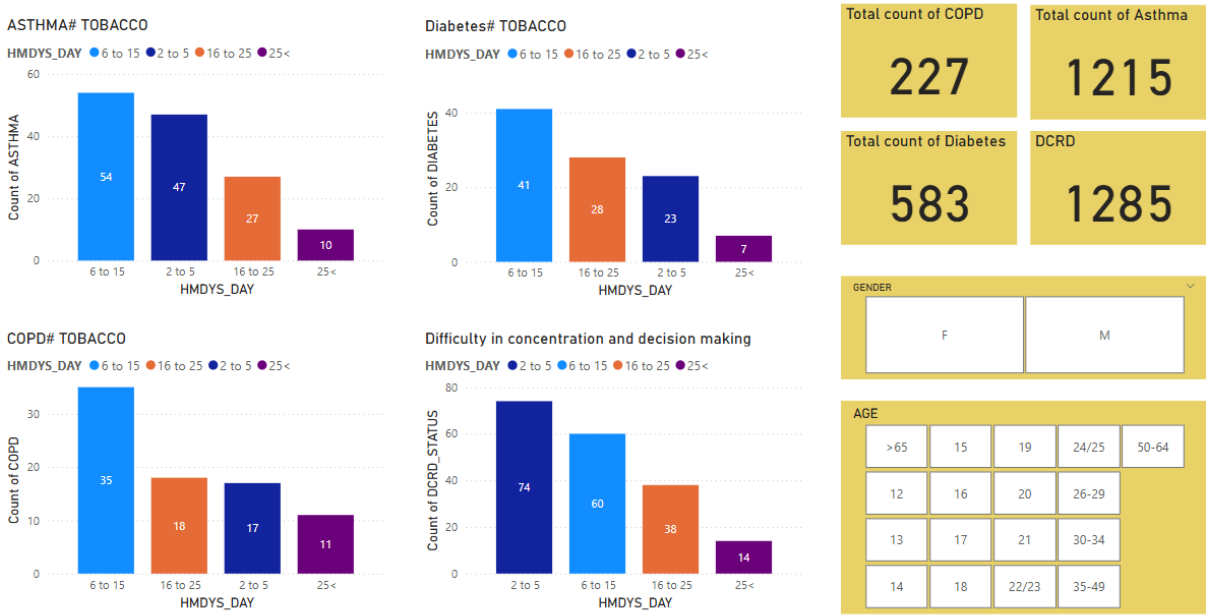
- Over 50% of users with depression were female.
- 47% of female users showed signs of mental illness.

- Suicidal ideation was notably higher among women.

Physical health effects like asthma and heart conditions were reported among both male and female users.

The 21–25 age group showed the highest concentration of cocaine users affected by both mental and physical health issues.

Tobacco#health



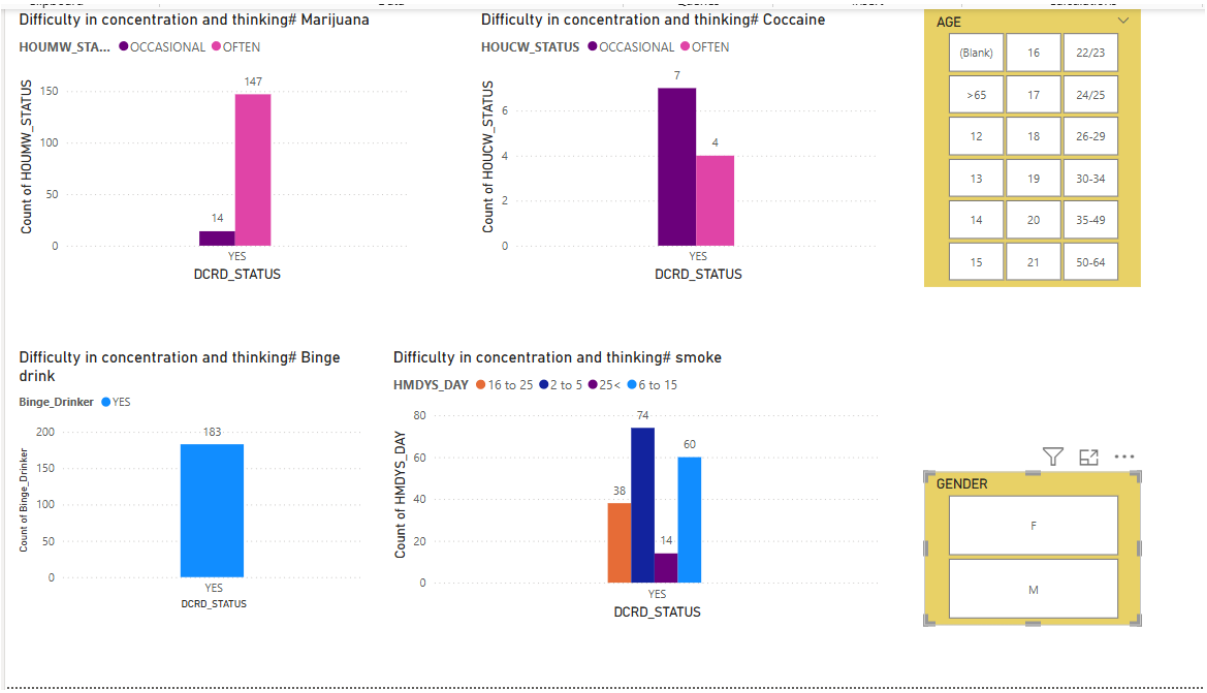
Among people who smoke more than 2 cigarettes a day (up to 35):

- Over 10% of asthma patients fell into this heavy-smoking category.
- Around 20% of individuals with diabetes reported similar smoking habits.
- Nearly 30% of those with COPD were heavy smokers.

While men were more likely to smoke over 25 cigarettes a day, women experienced more severe health consequences from tobacco use.

The 35–49 age group had the highest number of people affected by heavy tobacco intake.

Difficulty in thinking and concentration



- Difficulty with concentration and thinking was reported by:
    - 28% of frequent cocaine users
    - 17% of marijuana users

Most of these individuals fall within the 21–29 age group.

- In contrast, those who reported similar cognitive issues while also binge drinking and smoking more than 2 cigarettes daily were mostly in the 35–49 age group — indicating potential compounding effects of multiple substances at an older age.

Conclusion

This project demonstrates the real-world value of combining public health data with SQL and Power BI to uncover actionable insights.
These findings can support targeted interventions, public awareness, and better policy design for behavioral and mental health.