

---

---

# Predicting Drug Consumption Patterns Using Machine Learning

— Sandip Patel —

---

---

## A collection of hand-drawn illustrations of various pharmaceuticals and substances. It includes a small box with a cross symbol, a bottle, a syringe, several pills and capsules, a blister pack, and three cigarettes.

# Target Users

***SAMHSA***

Substance Abuse and Mental Health  
Services Administration

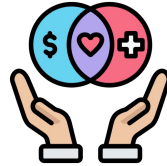


National Institute  
on Drug Abuse

- ❖ Substance Abuse and Mental Health Services Administration (SAMHSA)
- ❖ National Institute on Drug Abuse (NIDA)
- ❖ Educational Institutions
- ❖ Family Support Organizations

# The Multi-Faceted Impact of Addressing Drug Abuse

## The Impact of Substance Abuse on Society



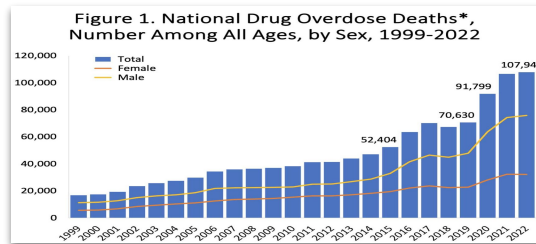
## Improved Public Health:

- Early intervention prevents drug abuse escalation, reducing healthcare burdens and deaths.
- Nearly **107,941** drug overdose deaths occurred in the **U.S. in 2022**.

Source : <https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates>




## Enhanced Quality of Life:

- Personalized treatment boosts recovery rates, benefiting individuals, families, and communities.
- Effective interventions can significantly lower the **\$740 billion** annual cost of drug-related expenses in the **U.S.**



# Dataset Overview

## Features:

- **Demographic** 
  - Gender
  - Age group
  - Countries
  - Educational background
- **Personality Scores** 
  - Oscore
  - Cscore
  - Escore
  - Ascore
  - Nscore
- **Physiological Trait** 
  - Impulsiveness
  - Sensation Seeking (SS)

**Targets:** (18 types of legal and illegal drugs)

### Stimulants

- Amphetamines (Amphet)
- Cocaine (Coke)
- Crack
- Ecstasy (Ecstasy)
- Methamphetamine (Meth)
- Caffeine (Caff)
- Cannabis
- Chocolate (Choc)
- Nicotine

### Depressants

- Alcohol
- Benzodiazepines (Benzos)
- Amyl nitrite (Amyl)

### Hallucinogens

- LSD
- Mushrooms

### Opioids

- Heroin
- Legal highs (Legalh)
- Semeron (Semer)

### Dissociatives

- Ketamine (Ketamine)

### Other

- Volatile substances (VSA)  
i.e Laughing Gas, Ether,  
Gasoline etc.

**Data Source:** <https://www.kaggle.com/datasets/mexwell/drug-consumption-classification>

# Data Processing and Predictive Modeling Workflow

## Data Acquisition

Data was acquired from Kaggle, which used codes instead of human-readable values for each feature.

| Alcohol $\sum_A$ |       |
|------------------|-------|
| CL6              | 505   |
| CL5              | 759   |
| CL4              | 287   |
| CL3              | 198   |
| CL2              | 68    |
| CL1              | 34    |
| CL0              | 34    |
| Grand Total      | 1,885 |

## Data Transformation

The original data was then mapped to human-readable values.

| Alcohol     |       |
|-------------|-------|
| Last Day    | 505   |
| Last Week   | 759   |
| Last Month  | 287   |
| Last Year   | 198   |
| Last Decade | 68    |
| Decade Ago  | 34    |
| Never       | 34    |
| Grand Total | 1,885 |

## Feature Engineering

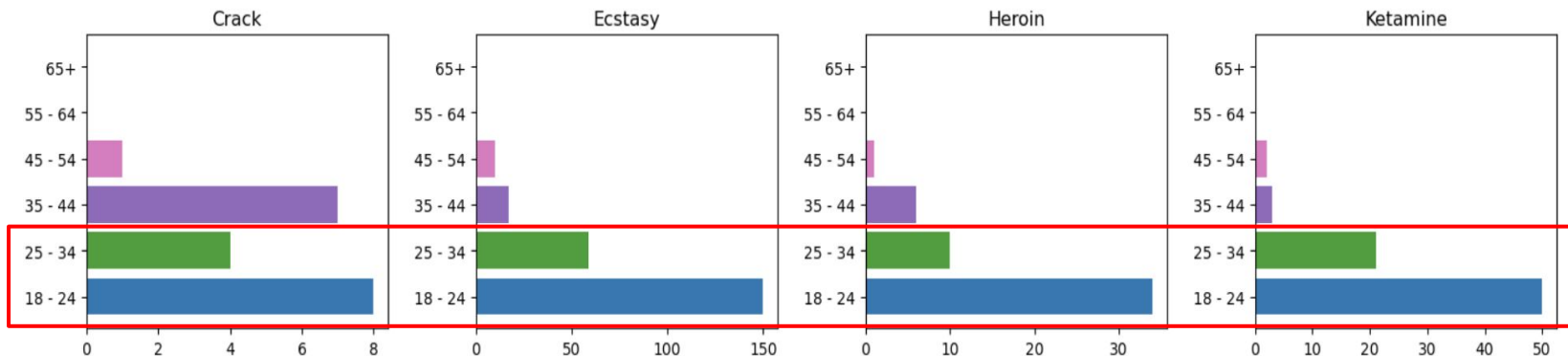
It involves converting categorical variables into numerical forms and ensuring all data is consistently numeric i.e. Active users (who used substance in last day, last week or last month).

| Alcohol     | Final Mapped |
|-------------|--------------|
| Last Day    | 505          |
| Last Week   | 759          |
| Last Month  | 287          |
| Last Year   | 0            |
| Last Decade | 0            |
| Decade Ago  | 0            |
| Never       | 0            |
| Grand Total | 1,551        |

## Model Development

- Logistic Regression
- Decision Tree
- Random Forest

# Preliminary EDA findings



**Younger individuals (18 to 34)** are more likely to engage in drugs consumption.

**Gender** influences substance use patterns, with **males** possibly more inclined towards certain substances.

| Heroin |          |            |           |
|--------|----------|------------|-----------|
| Gender | Last Day | Last Month | Last Week |
| Female | 3        | 8          | 3         |
| Male   | 10       | 16         | 13        |

| Coke   |          |           |            |
|--------|----------|-----------|------------|
| Gender | Last Day | Last Week | Last Month |
| Female | 5        | 14        | 37         |
| Male   | 14       | 27        | 62         |

# Preliminary EDA findings

| Nscore (bin) | Heroin   |           |            |
|--------------|----------|-----------|------------|
|              | Last Day | Last Week | Last Month |
| 10           |          |           | 1          |
| 20           |          |           | 3          |
| 30           | 3        | 4         | 5          |
| 40           | 7        | 9         | 11         |
| 50           | 3        | 3         | 4          |

| Nscore (bin) | Coke     |           |            |
|--------------|----------|-----------|------------|
|              | Last Day | Last Week | Last Month |
| 10           |          |           | 1          |
| 20           | 3        | 8         | 14         |
| 30           | 7        | 9         | 43         |
| 40           | 6        | 21        | 27         |
| 50           | 3        | 2         | 15         |

| Escore (bin) | Heroin   |           |            |
|--------------|----------|-----------|------------|
|              | Last Day | Last Week | Last Month |
| 10           |          | 2         |            |
| 20           | 3        | 1         | 5          |
| 30           | 8        | 6         | 10         |
| 40           | 2        | 6         | 7          |
| 50           |          | 1         | 2          |

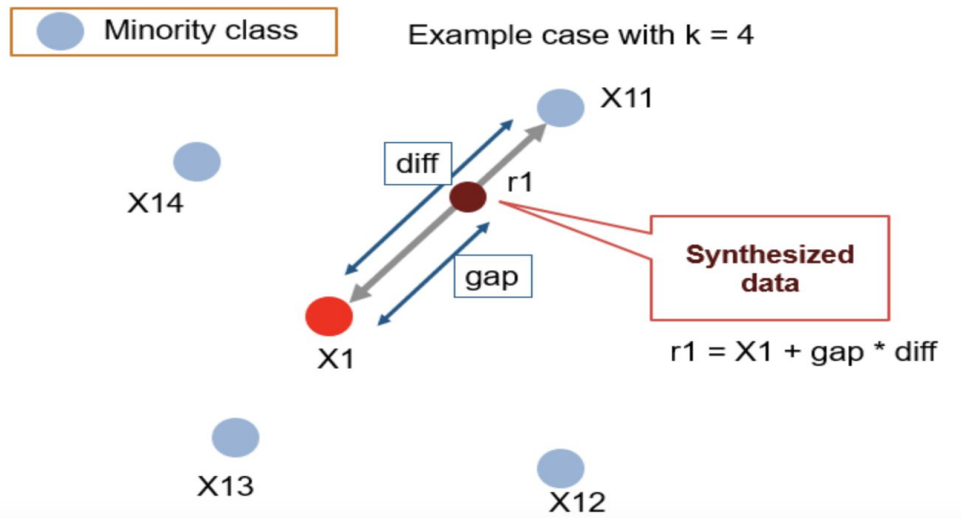
| Escore (bin) | Coke     |           |            |
|--------------|----------|-----------|------------|
|              | Last Day | Last Week | Last Month |
| 10           |          | 1         | 2          |
| 20           | 1        | 1         | 4          |
| 30           | 8        | 16        | 41         |
| 40           | 10       | 19        | 39         |
| 50           |          | 4         | 13         |

Personality traits like higher **Neuroticism (Nscore)**, **Extraversion (Escore)**, and **Sensation Seeking (SS)** values are strongly related to substance use behaviors.



# Modeling

- ❖ The dataset was heavily imbalanced, requiring alternate training methods.
- ❖ I used SMOTE (Synthetic Minority Oversampling Technique) to oversample the minority class and improve model learning.



My best model was **Logistic Regression** with SMOTE (**ADASYN**), achieving **83%** accuracy for predicting heroin abuse on the test set. Recall (**58%**) and Precision (**9%**)

# Drug Usage Patterns Across Age Groups

| Target Drugs | 18 - 24 | 25 - 34 | 35 - 44 | 45 - 54 | 55 - 64 | 65+  |
|--------------|---------|---------|---------|---------|---------|------|
| Alcohol      | 1.50    | 1.01    | 1.00    | 0.90    | 0.81    | 0.72 |
| Amphet       | 1.17    | 0.97    | 1.02    | 1.16    | 0.73    | 0.60 |
| Amyl         | 1.42    | 2.08    | 1.35    | 0.51    | 0.25    | 0.61 |
| Benzos       | 0.77    | 1.00    | 1.11    | 1.18    | 0.96    | 1.34 |
| Caffeine     | 2.94    | 2.82    | 3.88    | 3.12    | 2.04    | 1.58 |
| Cannabis     | 1.40    | 0.99    | 0.89    | 0.91    | 0.86    | 0.71 |
| Chocolate    | 3.09    | 2.93    | 2.79    | 3.14    | 1.86    | 1.26 |
| Coke         | 0.99    | 1.26    | 0.92    | 0.82    | 1.16    | 0.79 |
| Ecstasy      | 1.64    | 1.18    | 0.96    | 0.80    | 0.44    | 0.77 |
| Heroin       | 3.71    | 4.39    | 1.30    | 0.09    | 0.14    | 0.59 |
| Ketamine     | 1.87    | 1.39    | 0.54    | 0.80    | 0.66    | 0.87 |
| Legalh       | 1.30    | 0.92    | 1.05    | 0.81    | 0.99    | 0.76 |
| LSD          | 1.93    | 1.17    | 0.83    | 0.70    | 0.52    | 0.73 |
| Meth         | 1.45    | 1.22    | 0.75    | 0.73    | 0.95    | 0.82 |
| Mushrooms    | 1.24    | 0.90    | 1.32    | 0.98    | 0.62    | 0.57 |
| Nicotine     | 1.09    | 1.05    | 1.01    | 0.89    | 0.94    | 0.92 |
| VSA          | 4.38    | 2.32    | 0.68    | 0.13    | 0.53    | 0.78 |

Based on the provided coefficients, here are the four strongest associations:

- ❖ **18 - 24, 25-34** and **Heroin** with coefficients of **3.71, 4.39**.
- ❖ **18 - 24** and **VSA** with a coefficient of **4.38**.
- ❖ **35 - 44** and **Caffeine** with a coefficient of **3.88**.

# Analyzing Personality Traits



## OPENNESS

Imagination  
Curiosity  
Creativity  
Intellectual curiosity  
Sensation Seeking

(Oscore)

1.16



## CONSCIENTIOUS

Organization  
Diligence  
Responsibility  
Self-discipline

(Cscore)

0.94



## EXTRAVERSION

Sociability  
Assertiveness  
Energy level  
Positive emotions

(Escore)

1.01



## AGREEABLE

Compassion  
Politeness  
Cooperation  
Trust

(Ascore)

0.96



## NEUROTICISM

Anxiety  
Emotional instability  
Moodiness  
Vulnerability  
Impulsiveness

(Nscore)

1.06

A **coefficient** of **1.16** means that for every 1-unit increase in the **Oscore**, the odds of drug consumption increase by approximately **16%**

# Impact of Personality Traits on Drug Usage Coefficients

| Target Drugs | Oscore | SS   | Nscore | Escore | Impulsive | Ascore | Cscore |
|--------------|--------|------|--------|--------|-----------|--------|--------|
| Alcohol      | 1.00   | 1.74 | 1.00   | 1.03   | 0.97      | 0.99   | 1.00   |
| Amphet       | 0.78   | 1.09 | 1.09   | 1.09   | 1.04      | 1.01   | 0.71   |
| Amyl         | 1.10   | 0.82 | 0.36   | 0.87   | 1.14      | 0.91   | 0.58   |
| Benzos       | 1.21   | 0.83 | 2.29   | 1.26   | 0.94      | 1.01   | 1.17   |
| Caffeine     | 1.01   | 1.23 | 1.65   | 1.37   | 1.32      | 1.51   | 1.06   |
| Cannabis     | 1.71   | 1.11 | 0.84   | 0.95   | 1.03      | 1.02   | 0.85   |
| Chocolate    | 1.16   | 0.99 | 1.44   | 0.94   | 1.31      | 1.31   | 1.58   |
| Coke         | 0.65   | 1.72 | 1.28   | 1.14   | 1.06      | 0.97   | 1.14   |
| Ecstasy      | 1.19   | 1.06 | 0.76   | 1.21   | 0.77      | 0.86   | 0.80   |
| Heroin       | 0.61   | 1.02 | 0.89   | 0.52   | 0.87      | 0.53   | 1.68   |
| Ketamine     | 1.86   | 0.78 | 0.85   | 1.06   | 0.63      | 0.60   | 0.89   |
| Legalh       | 1.22   | 0.94 | 1.09   | 0.94   | 0.97      | 0.92   | 0.75   |
| LSD          | 1.46   | 0.90 | 1.15   | 1.00   | 0.71      | 1.01   | 0.72   |
| Meth         | 1.04   | 1.31 | 1.00   | 0.97   | 0.93      | 0.89   | 0.86   |
| Mushrooms    | 1.53   | 1.02 | 0.60   | 0.64   | 0.92      | 1.01   | 0.84   |
| Nicotine     | 1.06   | 1.13 | 1.06   | 1.01   | 1.18      | 1.03   | 0.86   |
| VSA          | 1.07   | 1.67 | 0.70   | 1.24   | 0.70      | 0.79   | 0.56   |

**Nscore for Benzos:**

**2.29**

**Oscore for Ketamine:**

**1.86**

**Cscore for Chocolate:**

**1.58**

**Nscore for Caffeine:**

**1.65**

**Oscore for**

**Mushrooms: 1.53**

# Exploring Drug Associations

| Target Drugs | Caffeine | Benzos | Cannabis | Coke | Nicotine | Legalh | Ecstasy | Amphet | LSD  | Ketamine | Amyl | Alcohol | Mushro.. | Chocola.. | Semer | Meth | VSA  | Heroin |
|--------------|----------|--------|----------|------|----------|--------|---------|--------|------|----------|------|---------|----------|-----------|-------|------|------|--------|
| Alcohol      | 1.85     | 1.49   | 1.25     | 1.50 | 1.02     | 0.78   | 1.49    | 0.78   | 1.23 | 1.27     | 1.29 | 0.00    | 1.34     | 1.41      | 1.01  | 0.72 | 1.01 | 0.78   |
| Amphet       | 1.57     | 2.00   | 1.44     | 1.17 | 1.07     | 1.58   | 1.25    | 0.00   | 0.84 | 1.36     | 1.14 | 0.80    | 0.94     | 1.17      | 1.35  | 1.01 | 0.96 | 0.88   |
| Amyl         | 5.30     | 2.63   | 0.84     | 2.20 | 2.12     | 1.94   | 1.18    | 1.14   | 0.23 | 1.31     | 0.00 | 2.27    | 1.77     | 0.89      | 0.76  | 1.30 | 1.11 | 0.48   |
| Benzos       | 1.34     | 0.00   | 1.08     | 1.36 | 1.50     | 1.17   | 1.05    | 1.62   | 0.74 | 1.43     | 1.14 | 1.30    | 0.97     | 1.01      | 1.44  | 1.82 | 1.21 | 1.13   |
| Caffeine     | 0.00     | 1.17   | 1.10     | 1.27 | 1.58     | 1.14   | 1.02    | 1.33   | 1.18 | 1.11     | 1.28 | 1.29    | 1.04     | 1.22      | 1.01  | 1.11 | 1.23 | 1.02   |
| Cannabis     | 1.01     | 1.06   | 0.00     | 1.22 | 1.88     | 1.25   | 1.40    | 1.02   | 0.98 | 1.03     | 0.89 | 1.02    | 1.24     | 1.07      | 0.84  | 0.99 | 1.08 | 0.93   |
| Chocolate    | 1.44     | 1.05   | 1.44     | 1.00 | 0.91     | 1.46   | 1.51    | 1.15   | 1.21 | 0.97     | 1.05 | 0.97    | 1.21     | 0.00      | 1.03  | 1.00 | 1.02 | 1.02   |
| Coke         | 1.24     | 1.54   | 1.90     | 0.00 | 1.24     | 0.78   | 2.10    | 1.41   | 1.02 | 0.91     | 1.40 | 1.33    | 1.09     | 1.11      | 1.34  | 1.17 | 1.07 | 1.46   |
| Ecstasy      | 0.80     | 1.04   | 1.81     | 2.09 | 1.36     | 1.30   | 0.00    | 1.22   | 1.51 | 1.52     | 1.18 | 1.27    | 0.98     | 1.57      | 1.25  | 0.67 | 0.75 | 1.05   |
| Heroin       | 1.96     | 1.41   | 0.85     | 2.02 | 1.31     | 0.97   | 0.71    | 0.95   | 0.78 | 1.07     | 0.92 | 0.49    | 0.89     | 0.86      | 0.77  | 1.00 | 1.02 | 0.00   |
| Ketamine     | 0.89     | 1.60   | 1.04     | 0.70 | 2.05     | 1.45   | 1.64    | 1.90   | 1.90 | 0.00     | 1.39 | 1.48    | 1.16     | 1.00      | 1.34  | 0.76 | 0.82 | 1.14   |
| Legalh       | 1.23     | 1.07   | 1.84     | 0.86 | 1.11     | 0.00   | 1.18    | 1.56   | 0.94 | 1.26     | 1.20 | 0.83    | 1.00     | 1.42      | 0.89  | 0.94 | 1.01 | 0.87   |
| LSD          | 1.26     | 0.85   | 1.31     | 1.07 | 0.89     | 1.14   | 1.85    | 0.78   | 0.00 | 1.38     | 0.76 | 1.29    | 1.82     | 1.16      | 0.84  | 0.93 | 1.25 | 0.93   |
| Meth         | 1.54     | 2.46   | 0.81     | 0.97 | 1.10     | 0.92   | 0.83    | 1.10   | 0.93 | 0.77     | 1.17 | 0.80    | 0.99     | 0.82      | 0.85  | 0.00 | 1.05 | 1.19   |
| Mushrooms    | 1.18     | 0.97   | 2.38     | 1.21 | 0.75     | 1.20   | 0.99    | 1.13   | 1.99 | 1.41     | 1.03 | 1.37    | 0.00     | 0.85      | 0.81  | 0.98 | 0.87 | 0.91   |
| Nicotine     | 1.28     | 1.14   | 1.90     | 1.12 | 0.00     | 1.00   | 1.07    | 1.01   | 0.93 | 1.17     | 1.12 | 1.02    | 0.90     | 0.93      | 1.13  | 0.98 | 0.99 | 0.95   |
| VSA          | 3.95     | 2.69   | 1.38     | 1.35 | 0.88     | 2.00   | 0.58    | 0.80   | 2.44 | 0.75     | 1.59 | 0.81    | 0.87     | 1.07      | 0.92  | 1.84 | 0.00 | 1.02   |

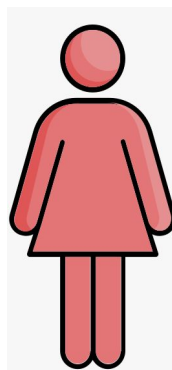
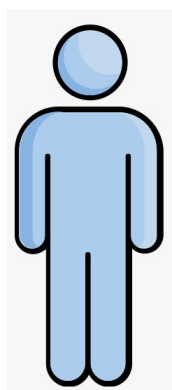
Based on the coefficients, the two strongest associations among substance uses are:

1. **Caffeine** and **Amyl** with a coefficient of **5.30**, **Caffeine** and **VSA** with a coefficient of **3.95**
2. **Benzos** and **Amyl** with a coefficient of **2.63**, **Benzos** and **VSA** with a coefficient of **2.69**

These high values indicate a significant likelihood that individuals using **caffeine** are also likely to use **amyl** and **VSA**.

# Gender and Drug Usage Patterns

| Target Drugs | Male | Female |
|--------------|------|--------|
| Alcohol      | 0.83 | 0.96   |
| Amphet       | 1.12 | 0.90   |
| Amyl         | 1.16 | 0.86   |
| Benzos       | 1.06 | 0.94   |
| Caffeine     | 1.69 | 1.31   |
| Cannabis     | 1.15 | 0.87   |
| Chocolate    | 1.70 | 1.97   |
| Coke         | 1.10 | 0.91   |
| Ecstasy      | 1.02 | 0.98   |
| Heroin       | 1.10 | 0.91   |
| Ketamine     | 1.27 | 0.79   |
| Legalh       | 1.15 | 0.87   |
| LSD          | 1.10 | 0.91   |
| Meth         | 1.08 | 0.93   |
| Mushrooms    | 1.16 | 0.86   |
| Nicotine     | 1.07 | 0.93   |
| VSA          | 0.98 | 1.02   |



## ❖ Substances with Highest Association:

➤ Chocolate (**1.70**)

➤ Caffeine (**1.68**)

## ❖ Significant Substances:

➤ **Ketamine (1.27)**

➤ **Amphetamines (1.11)**

## ❖ Substances with Highest Association:

➤ Chocolate (**1.97**)

➤ Caffeine (**1.31**)

## ❖ Significant Substances:

➤ **VSA (Volatile Substance Abuse) (1.01)**

➤ **Ecstasy (0.98)**

**Men** appear to be more at risk for substance abuse



# Regional Variations in Drug Consumption

| Target Drugs | USA   | Australia | New Zealand | Republic of Ireland | Canada | UK   | Others |
|--------------|-------|-----------|-------------|---------------------|--------|------|--------|
| Alcohol      | 0.81  | 0.92      | 1.03        | 0.83                | 0.79   | 1.60 | 0.99   |
| Amphet       | 1.82  | 1.39      | 0.76        | 0.96                | 1.18   | 0.57 | 0.75   |
| Amyl         | 0.71  | 1.67      | 0.72        | 0.91                | 0.45   | 1.83 | 0.82   |
| Benzos       | 1.46  | 1.16      | 3.76        | 0.83                | 0.90   | 0.66 | 1.01   |
| Cannabis     | 1.45  | 1.09      | 1.13        | 1.03                | 0.96   | 0.67 | 1.08   |
| Coke         | 0.84  | 0.88      | 0.88        | 1.07                | 1.14   | 1.11 | 1.10   |
| Ecstasy      | 1.01  | 1.23      | 0.83        | 1.09                | 1.03   | 0.94 | 0.93   |
| Heroin       | 10.02 | 0.51      | 0.81        | 1.66                | 1.80   | 0.26 | 0.18   |
| Ketamine     | 0.52  | 1.12      | 0.84        | 0.89                | 1.06   | 1.83 | 0.96   |
| Legalh       | 1.08  | 0.97      | 1.27        | 0.99                | 1.31   | 0.79 | 1.10   |
| LSD          | 1.81  | 1.56      | 0.84        | 0.84                | 1.09   | 0.42 | 1.48   |
| Meth         | 1.83  | 0.89      | 0.79        | 0.80                | 0.97   | 0.57 | 1.27   |
| Mushrooms    | 1.74  | 1.16      | 0.79        | 0.82                | 1.26   | 0.53 | 1.10   |
| Nicotine     | 0.98  | 0.93      | 0.96        | 1.10                | 1.02   | 1.03 | 0.98   |
| VSA          | 1.18  | 1.12      | 1.04        | 1.77                | 0.33   | 1.55 | 0.55   |

- ❖ **USA: Heroin - 10.02**
- ❖ **Australia: Amphet - 1.39**
- ❖ **New Zealand: Benzos - 3.76**
- ❖ **Republic of Ireland: VSA - 1.76**
- ❖ **Canada: Heroin - 1.80**
- ❖ **UK: Ketamine - 1.83**
- ❖ **Others: LSD - 1.47**

These coefficients represent the strongest relationships between the features (listed drugs) and the likelihood of substance use in each respective country

# So What???

- ❖ Initiate immediate support and education programs for young males in the **USA**, focusing on **heroin** use (coefficient: **10.02**) and other high-risk substances.
- ❖ Implement evidence-based strategies tailored to males aged **18-34**, addressing critical substances like **heroin, amphetamines** and **ketamine**.
- ❖ Stress the importance of ongoing data collection to monitor intervention effectiveness.
- ❖ Aim to reduce substance dependence over the next several decades.



## Next Steps

- ❖ Explore using these models on different datasets.
- ❖ Plan to enhance model robustness with cross-validation.
- ❖ Compare results with similar projects for benchmarking and insights.