

**The Blue Group**

# **Progress Checkpoint**

ITEC 3040 Course Project

Scott Wright, Sophia Maxine Villarosa, Gabrielle Campbell, Amanuel Demeke, Jisuk Kang

# Our Dataset

**Title:** Drug Consumption Dataset

**Source:** UCI Machine Learning Repository

**Some brief information:**

- Contains 1885 people
- 32 attributes spanning from personality traits to specific drug usage.
- Drug usage is described in terms of recency
  - (last day/week/month/decade)
- Other attributes include age, gender, education, country, ethnicity and ID number.




# Problem and Goal

## What we want to solve:


- We want to analyze our drug consumption dataset to determine what kinds of factors in people would determine whether they would be potential users of magic mushrooms.

## Why it's important:

- It has been previously known that magic mushrooms can be used as an antidepressant.
  - On February 15, 2022, a paper was published out of the Johns Hopkins School of Medicine that showed prolonged mushroom use to treat MDD was safe. (MDD = major depressive disorder).
  - This increased the interest in mushrooms in both the medical community and the recreational community.
- 

# Problem and Goal

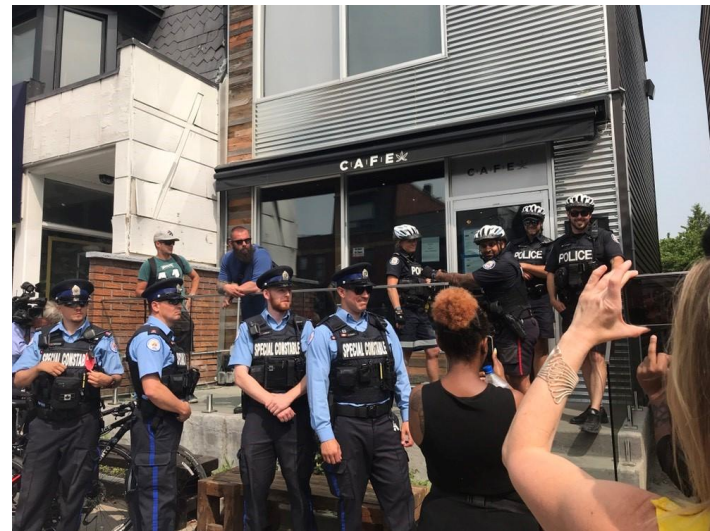
## Why it's important

- As a result, mushroom dispensaries are popping up in major cities in Canada, even though the drug is still illegal.
  - Dispensaries are still being raided and shut down but at a much slower rate than you would think.
  - This is because local police departments are more focused on harmful drugs that fuel organized crime and contribute to the current drug health crisis (like opioids).
- 

# Problem and Goal

## Why it's important

- The current state of mushrooms is reminiscent to how illegal cannabis shops were treated in major Canadian cities a few years ago.
- Police attempted to to shut them down, only for them to pop back up, as demand for cannabis skyrocketed in Canada.
- As of now, you can still find many illegal cannabis shops that police don't care to shut down.
- **Our group predicts that magic mushrooms will follow a similar route to legalization (and normalization).**
- **Because of that, we believe creating a predictor model to determine what kinds of people will want to use magic mushrooms will be very valuable from a marketing perspective.**

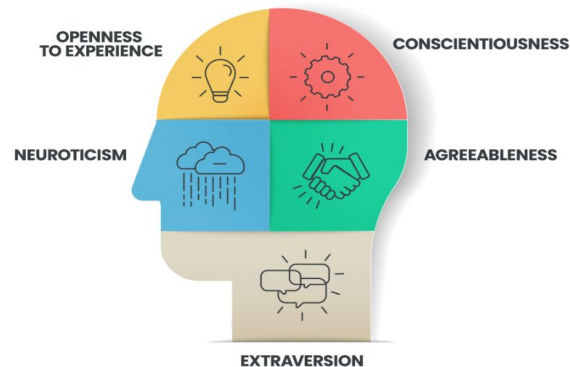


*Police raid of Cafe Dispensary on Harbord street, July 2019*

# Problem and Goal

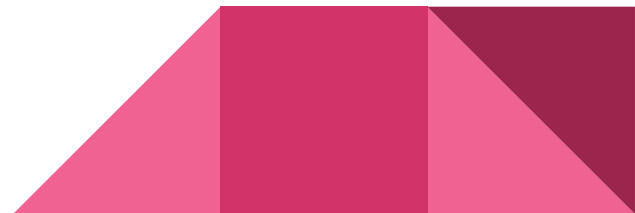
## The results we expect:

- We predict that if a person has used **alcohol**, **cannabis** and **nicotine** they are likely to consume magic mushrooms.
- We also predict that personality traits such as **extraversion** and **openness-to-experience** will be key factors in determining whether someone will use mushrooms.



# Classification/Application Problem

With the dataset, we would like to classify people who use magic mushrooms, based on an individual's personality, age, education, and other drug use.



# Our Preprocess

- Our preprocess for the data centered around *cleaning*.
- We needed to label our attribute labels.
- We needed to transform entity values into legible text.
- We needed to remove attributes we did not want
  - Such as chocolate use, and hardcore drug use.






# Elimination

- Our dataset provided attribute 31, it is a fictitious drug called Semeron which people in the study can claim whether they've used it and how recently
- We decided this attribute would be perfect for detecting and eliminating noise because it indicates that a person is willing to lie about their drug use.
- We do not want these people influencing our class model, so we removed every person that claimed to have used Semeron.



# A Look at our Variables

## Variable identification


- ID - Nominal, Discrete
  - Age - Predictor, Ordinal, Discrete
  - Gender - Predictor, Nominal (Binary), Discrete
  - Education - Predictor, Ordinal, Discrete
  - Nicotine - Predictor, Ordinal, Discrete
  - Alcohol - Predictor, Ordinal, Discrete
  - Cannabis - Predictor, Ordinal, Discrete
  - Semeron - Predictor, Ordinal, Discrete
  - Mushroom - Response, Ordinal, Discrete
- 

# Difficulties and Solutions

## Difficulty 1 : Interpreting Personality Data Values

- The personality measurements are presented in numerical values, but is hard to interpret as there is no explanation of the scale used.

## Solution 1: Look to other data

- Our solution to this is to spend time looking at other data sets and studies which also processed personality data in a similar way. This way we can better incorporate extraversion and openness in our analysis.
  - The other personality values (neuroticism, agreeableness, conscientiousness, impulsivity, and sensation seeking) are not needed, therefore eliminated.
- 

# Difficulties and Solutions

## Difficulty 2: Dataset

- We initially had technical difficulties with importing the data set into Matlab.
- The data file uses the “.data” extension which is an extension we were not familiar with.

```

1,0.49788,0.48246,-0.05921,0.96082,0.126,0.31287,-0.57545,-0.58331,-0.91699,-0.
2,-0.07854,-0.48246,1.98437,0.96082,-0.31685,-0.67825,1.93886,1.43533,0.76096,-
3,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.46725,0.80523,-0.84732,-1.6209
4,-0.95197,0.48246,1.16365,0.96082,-0.31685,-0.14882,-0.80615,-0.01928,0.59042,
5,0.49788,0.48246,1.98437,0.96082,-0.31685,0.73545,-1.6334,-0.45174,-0.30172,1.
6,2.59171,0.48246,-1.22751,0.24923,-0.31685,-0.67825,-0.30033,-1.55521,2.03972,
7,1.09449,-0.48246,1.16365,-0.57009,-0.31685,-0.46725,-1.09207,-0.45174,-0.3017
8,0.49788,-0.48246,-1.7379,0.96082,-0.31685,-1.32828,1.93886,-0.84732,-0.30172,
9,0.49788,0.48246,-0.05921,0.24923,-0.31685,0.62967,2.57309,-0.97631,0.76096,1.
10,1.82213,-0.48246,1.16365,0.96082,-0.31685,-0.24649,0.00332,-1.42424,0.59042,
11,-0.07854,0.48246,0.45468,0.96082,-0.31685,-1.05308,0.80523,-1.11902,-0.76096
12,1.09449,-0.48246,-0.61113,-0.28519,-0.31685,-1.32828,0.00332,0.14143,-1.9259
13,1.82213,0.48246,0.45468,0.96082,-0.31685,2.28554,0.16767,0.44585,-1.6209,-0.
14,1.82213,0.48246,-0.05921,0.24923,-0.31685,-0.79151,0.80523,-0.01928,0.94156,
15,1.82213,0.48246,-0.05921,0.96082,-0.31685,-0.92104,1.45421,0.44585,-0.60633,
16,1.82213,-0.48246,0.45468,0.96082,-0.31685,-2.05048,-1.50796,-1.55521,-1.0753
17,0.49788,0.48246,-0.61113,0.96082,-0.31685,-1.55078,-0.80615,-1.68062,0.28783
18,1.09449,-0.48246,-1.7379,0.96082,-0.31685,0.52135,-1.23177,-0.31776,-0.45321
19,1.82213,-0.48246,0.45468,-0.09765,-0.31685,1.37297,-0.15487,-0.17779,-1.9259
20,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.34799,-1.7625,-2.39883,-1.9259
21,1.09449,-0.48246,-0.05921,0.96082,-0.31685,-0.79151,0.80523,0.7233,1.61108,-
22,2.59171,-0.48246,-2.43591,0.96082,-0.31685,-1.1943,0.47617,-1.11902,-0.60633
23,1.09449,-0.48246,0.45468,0.96082,-0.31685,0.41667,-0.94779,-0.84732,1.11406,

```

*The original .data file in notepad*

# Difficulties and Solutions

## Solution 2: Dataset

- Our first step was transforming the file into an acceptable data type, we went with CSV.
- We then transformed the values into words to make them readable.
  - I.e., data such as age being 18-24 appeared as -0.951.

```

1,0.49788,0.48246,-0.05921,0.96082,0.126,0.31287,-0.57545,-0.58331,-0.91699,-0.
2,-0.07854,-0.48246,1.98437,0.96082,-0.31685,-0.67825,1.93886,1.43533,0.76096,-
3,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.46725,0.80523,-0.84732,-1.6209
4,-0.95197,0.48246,1.16365,0.96082,-0.31685,-0.14882,-0.80615,-0.01928,0.59042,
5,0.49788,0.48246,1.98437,0.96082,-0.31685,0.73545,-1.6334,-0.45174,-0.30172,1.
6,2.59171,0.48246,-1.22751,0.24923,-0.31685,-0.67825,-0.30033,-1.55521,2.03972,
7,1.09449,-0.48246,1.16365,-0.57009,-0.31685,-0.46725,-1.09207,-0.45174,-0.3017
8,0.49788,-0.48246,-1.7379,0.96082,-0.31685,-1.32828,1.93886,-0.84732,-0.30172,
9,0.49788,0.48246,-0.05921,0.24923,-0.31685,0.62967,2.57309,-0.97631,0.76096,1.
10,1.82213,-0.48246,1.16365,0.96082,-0.31685,-0.24649,0.00332,-1.42424,0.59042,
11,-0.07854,0.48246,0.45468,0.96082,-0.31685,-1.05308,0.80523,-1.11902,-0.76096
12,1.09449,-0.48246,-0.61113,-0.28519,-0.31685,-1.32828,0.00332,0.14143,-1.9259
13,1.82213,0.48246,0.45468,0.96082,-0.31685,2.28554,0.16767,0.44585,-1.6209,-0.
14,1.82213,0.48246,-0.05921,0.24923,-0.31685,-0.79151,0.80523,-0.01928,0.94156,
15,1.82213,0.48246,-0.05921,0.96082,-0.31685,-0.92104,1.45421,0.44585,-0.60633,
16,1.82213,-0.48246,0.45468,0.96082,-0.31685,-2.05048,-1.50796,-1.55521,-1.0753
17,0.49788,0.48246,-0.61113,0.96082,-0.31685,-1.55078,-0.80615,-1.68062,0.28783
18,1.09449,-0.48246,-1.7379,0.96082,-0.31685,0.52135,-1.23177,-0.31776,-0.45321
19,1.82213,-0.48246,0.45468,-0.09765,-0.31685,1.37297,-0.15487,-0.17779,-1.9259
20,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.34799,-1.7625,-2.39883,-1.9259
21,1.09449,-0.48246,-0.05921,0.96082,-0.31685,-0.79151,0.80523,0.7233,1.61108,-
22,2.59171,-0.48246,-2.43591,0.96082,-0.31685,-1.1943,0.47617,-1.11902,-0.60633
23,1.09449,-0.48246,0.45468,0.96082,-0.31685,0.41667,-0.94779,-0.84732,1.11406,

```

*The original .data file in notepad*

# Difficulties and Possible Solutions

## Solution 2: Dataset

ID	Age	Gender	Education
1	{'35 - 44'}	{'Female'}	{'Diploma' }
2	{'25 - 34'}	{'Male' }	{'Docterate'}
3	{'35 - 44'}	{'Male' }	{'Diploma' }
4	{'18 - 24'}	{'Female'}	{'Masters' }
5	{'35 - 44'}	{'Female'}	{'Docterate'}
6	{'65+' }	{'Female'}	{'18' }

```
1,0.49788,0.48246,-0.05921,0.96082,0.126,0.31287,-0.57545,-0.58331,-0.91699,-0.
2,-0.07854,-0.48246,1.98437,0.96082,-0.31685,-0.67825,1.93886,1.43533,0.76096,-
3,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.46725,0.80523,-0.84732,-1.6209
4,-0.95197,0.48246,1.16365,0.96082,-0.31685,-0.14882,-0.80615,-0.01928,0.59042
5,0.49788,0.48246,1.98437,0.96082,-0.31685,0.73545,-1.6334,-0.45174,-0.30172,1.
6,2.59171,0.48246,-1.22751,0.24923,-0.31685,-0.67825,-0.30033,-1.55521,2.03972
7,1.09449,-0.48246,1.16365,-0.57009,-0.31685,-0.46725,-1.09207,-0.45174,-0.3017
8,0.49788,-0.48246,-1.7379,0.96082,-0.31685,-1.32828,1.93886,-0.84732,-0.30172
9,0.49788,0.48246,-0.05921,0.24923,-0.31685,0.62967,2.57309,-0.97631,0.76096,1.
10,1.82213,-0.48246,1.16365,0.96082,-0.31685,-0.24649,0.00332,-1.42424,0.59042
11,-0.07854,0.48246,0.45468,0.96082,-0.31685,-1.05308,0.80523,-1.11902,-0.76096
12,1.09449,-0.48246,-0.61113,-0.28519,-0.31685,-1.32828,0.00332,0.14143,-1.9259
13,1.82213,0.48246,0.45468,0.96082,-0.31685,2.28554,0.16767,0.44585,-1.6209,-0.
14,1.82213,0.48246,-0.05921,0.24923,-0.31685,-0.79151,0.80523,-0.01928,0.94156
15,1.82213,0.48246,-0.05921,0.96082,-0.31685,-0.92104,1.45421,0.44585,-0.60633
16,1.82213,-0.48246,0.45468,0.96082,-0.31685,-2.05048,-1.50796,-1.55521,-1.0753
17,0.49788,0.48246,-0.61113,0.96082,-0.31685,-1.55078,-0.80615,-1.68062,0.28783
18,1.09449,-0.48246,-1.7379,0.96082,-0.31685,0.52135,-1.23177,-0.31776,-0.45321
19,1.82213,-0.48246,0.45468,-0.09765,-0.31685,1.37299,-0.15487,-0.17779,-1.9259
20,0.49788,-0.48246,-0.05921,0.96082,-0.31685,-0.34799,-1.7625,-2.39883,-1.9259
21,1.09449,-0.48246,-0.05921,0.96082,-0.31685,-0.79151,0.80523,0.7233,1.61108,-
22,2.59171,-0.48246,-2.43591,0.96082,-0.31685,-1.1943,0.47617,-1.11902,-0.60633
23,1.09449,-0.48246,0.45468,0.96082,-0.31685,0.41667,-0.94779,-0.84732,1.11406
24,0.49788,0.48246,1.16365,0.96082,-0.31685,0.62967,2.57309,-0.97631,0.76096,1.
25,0.49788,0.48246,1.16365,0.96082,-0.31685,0.62967,2.57309,-0.97631,0.76096,1.
```

- The data is much easier to interpret now. At this point we also removed attributes that we thought were not important, and removed personality attributes that we still needed time to understand.

*The original .data file in notepad*

# Schedule (detailed plan of project)

February 24th: Choose our Data Set and set a goal of what we want to do with it.

March 1 - 5th: Summarize the data set and formulate into a classification problem. Discuss about any difficulties and issues encountered.

March 6 - 8th: Find solutions for difficulties and issues. Review the summarized data.

March 9 - March 21st: Use MatLab to analyze the data, developing necessary diagrams and derive the results from the data set. Learn more about interpreting personality data.

March 22 - April 8th: Decide on algorithms/evaluation methods, and finalize our code. Start documenting our final analysis. Complete project report.



**The Blue Group**

**Thank You For Listening**

Scott Wright, Sophia Maxine Villarosa, Gabrielle Campbell, Amanuel Demeke, Jisuk Kang