Summary Team name: Tomato303

Presented by Jaekang Lee, Alexander MacKenzie, Chang Sun, Tzuying Yu.

Introduction & Background: The growing non-medical use of prescription drugs is a global health concern. The Canadian's Non-Medical Use of Prescription Drugs Survey data collected by RADARS® System gives an opportunity to understand the situation. Our team explored the raw survey data by visualizing, and came up with three research questions for the report.

Data Exploration:
There are two interesting findings with regard to demographic features. By plotting the bar plot for misuse among age groups, as people get older, the chance of misuse decreases. Moreover, for misuse among income groups, we find that lower income households have a higher chance in misusing prescription drugs. In our preliminary data explorations, we created four preliminary sub datasets. They are the substance usage among all types of substances provided, mental disorders, substance nonmedical use, and the frequency of some of the substances used. Our goal is to understand the correlation between each individual variable and whether they provide us with any social group information. We used R studio's corrplot function to plot out correlation heatmaps for our four datasets. The biggest finding is that opioid use is highly positively correlated with codeine use. This could suggest most substance users would prefer codeine as their opioid of choice, and most substance misuses also happen with prescription/non-prescription codeine products.

Research Question1: There are 185 variables in total to consider. How do we explore all of these?
We used principal components and mutual information to compute which variables to keep. We found out that Opioid, Codeine and Cocaine best explained our response variable 'NMU' with highest mutual information. To continue investigating which specific groups best explained 'NMU', we used principal components to reduce dimensionality and capture social groups.(The process is shown in video) Finally, we took the top components then calculated mutual information to find which component best explained 'NMU' hence we got our top NMU social groups.
Research Question 2: What types of social groups best explain non-medical use?
Opioid, Codeine, Cocaine were top three NMU drugs and Quebec was top NMU social group.
Research Question 3: How does frequency of smoking and drinking impact drug use, specifically drug misuse and drug use recency?
Given prior evidence for a correlation between smoking and non-medical use of drugs, we specifically compared drug use between different smoking and drinking groups based on their smoking and drinking frequency. Generally, our plots show that increased frequency of smoking and alcohol consumption produces higher non-medical drug use percentages. In the context of drinking, one striking result is that non-pharmaceutical amphetamines, fentanyl, and steroids are used much more by males who drink 4-5 times per week compared to any other drinking group, including daily drinkers. We used "glms" to quantify this relationship between the binomial response, 'NMU', and both smoking and drinking.

Results found that the odds of an occasional smoker misusing any drug in their lifetime is 2.34 times the odds of a non-smoker misusing any drug. The odds of a daily smoker misusing any drug in their lifetime is 1.96 times the odds of a non-smoker misusing any drug. A final model was fit between ‘NMU’ and opioid, codeine, fentanyl, ghb, heroin, and stimulant use. After model comparisons proved this model to be the best fit based on drop-in-deviance tests, coefficients determined that there is a 1030.6% increase in odds of misusing drugs in one's lifetime going from individuals who've never used prescription opioids to individuals who have. Additionally, the odds of lifetime codeine users having misused any drug in their life is 1.78 times the odds of lifetime non-codeine users having misused drugs at any point in their life. The model also found that compared to non-GHB/GBL lifetime users, people who used GHB or GBL in their life had 2.31 times the odds of misusing drugs in their life. For heroin, individuals who used heroin in the last year had a 12238.8% increase in odds of misusing drugs in their life compared to individuals who've never done heroin. Lastly, those who used prescription stimulants at any point in their life had 3.29 times the odds of misusing drugs compared to those who didn't ever use prescription stimulants.

Limitations:

- Difficulty in building certain models (random forest) due to existence of only binary and categorical variables. Causes bias towards heavier proportions.
- No way to know beyond the set of answers. (Ex. Reason for NMU was removed by the data provider)
- No other timeline make it hard to make strong claims.