James Santarpio
Mardi Elings
ECNS 460

Oregon Drug Decriminalization Analysis

In 2021, almost 100,000 people died of preventable drug overdoses, a situation that has been increasing exponentially since the late 1990's. Measure 110 in Oregon to decriminalize schedule I and II drugs passed and was implemented in 2021. The law made personal non-commercial possession of a controlled substance no more than a Class E violation, with a max of a $100 fine. Another aspect is to establish drug addiction treatment and recovery programs funded in part by the state's marijuana tax revenue and state prison savings. Now, the measure is on the ballot to be repealed. We decided to do a sentiment analysis using a social media API to determine if people feel the law is beneficial or should be overturned. Besides the sentiment analysis, we will use crime and drug use data from Oregon and similar states (WA, CA, etc..) to compare the effectiveness of measure 110.

For the sentiment analysis data, we will use API's from social media (Facebook). We plan to collect data that has key words relating to measure 110 and drug decriminalization with this API. Additionally, other social media sites such as reddit or instagram could be used to gather data.

Sentiment analysis algorithms we are familiar with to be used include naive bayes, logistic regression, and support vector machines. More research will go into these three, other available algorithms, and which algorithms fit the best. We will split the data into a training set and testing set and train and compare different models with different classification algorithms.

It will be interesting to see what people's opinion on social media sites will have. This is not generalizable to the entire public. It will be important to find demographic information for facebook and other social media sites as well to see if a different demographic has differing opinions on the issues.

Data:

From the FBI crime data reporter we can get the number of arrests by crime and number of drug arrests. There are lots of options for this data, but what stood out to me were states with populations and estimated crimes over several decades. It can be easily merged by states or, depending on other data, drug use. They also have a breakdown of crimes by city.
https://cde.ucr.cjis.gov/LATEST/webapp/#

Number of overdose deaths by state from the CDC reports on deaths and rate. Can be merged by the state and county variables (or fips codes) and year.
https://www.cdc.gov/nchs/pressroom/sosmap/drug_poisoning_mortality/drug_poisoning.htm
https://www.cdc.gov/nchs/nvss/vsrr/prov-county-drug-overdose.htm

The substance abuse and mental health data archive (SAMHSA) from 2002-2019. This national survey on drug use and health has data on if someone used illicit drugs in the past year, if they approach someone selling drugs, and information about health treatment and recovery frequency.
https://www.datafiles.samhsa.gov/data-sources

DOSE covers non-fatal overdose data by state, which could be merged by state variable. Other interesting variables are the change in overdoses for each drug category month to month.
https://www.cdc.gov/drugoverdose/nonfatal/dose/surveillance/dashboard/index.html#stateInfo

Social Media data:
(to be continued)

Facebook demographics:
https://www.statista.com/statistics/187549/facebook-distribution-of-users-age-group-usa/