Marijuana Legalization Impacts

(ETL – Project)

ANALYTICAL ANACONDAS:

Jose Gallegos

Jay Lee

Maria Soria

Jerry Walker

USC Data Analytics Boot Camp

August 24, 2019

INTRODUCTION

Opponents of marijuana legalization claim that there could be negative impacts from legalization. The Analytical Anacondas team wanted to explore these claims. The proposed analysis includes assessing trends associated with some of these claims in states where marijuana has been fully legalized compared to states where it is still completely illegal.

As of August 2019, 11 states and the District of Columbia (D.C.) have fully legalized marijuana use and 12 states where marijuana is still completely illegal. The table below shows those respective states (i.e. D.C.):

Table 1: States Where Marijuana is (1) Fully Legal, and (2) Fully Illegal

Fully Legal	Fully Illegal
Alaska	Alabama
California	Idaho
Colorado	Kansas
District of Columbia	Kentucky
Illinois	Mississippi
Maine	Nebraska
Massachusetts	North Carolina
Michigan	South Carolina
Nevada	South Dakota
Oregon	Tennessee
Vermont	Wisconsin
Washington	Wyoming

The team decided to focus on factors that could potentially have substantial societal impacts. Therefore, the team selected crime rates and drug and alcohol related deaths as the main factors for this assessment. The team proposes a scientific approach to the analysis using the “Fully Illegal” group as the control group, and the “Fully Legal” group as the experimental group. For example, the analysis could compare the crime rate trends in states where marijuana is fully legal compared to the trends in states where it is fully illegal.

- Inverse trends between states where marijuana is fully legal compared to those where it is fully illegal would indicate that there is a correlation between marijuana use and crime rate.
- Similar, or proportional, trends between states where marijuana is fully legal compared to those where it is fully illegal would indicate that there is not a significant correlation between marijuana use and crime rate.

The results could be further evaluated to assess the degree of statistical correlation, should there be an inverse relationship between states where marijuana is fully legal compared to those where it is fully illegal. Additionally, the results could validate whether there is a positive or negative impact from marijuana use.

The same methodology described for assessing the correlation between marijuana use and crime rates would be followed to assess the correlation between marijuana use and drug and alcohol related deaths.

DATA (ETL)

Sources:

Obtained data from the following sources:

- Marijuana legalization by state
  - DISA Global Solutions: https://disa.com/map-of-marijuana-legality-by-state
  - Wikipedia: https://en.wikipedia.org/wiki/Timeline_of_cannabis_laws_in_the_United_States
  - Marijuana Policy Project: https://www.mpp.org/states%20CDC%20website:%20https://wonder.cdc.gov/controller/datarequest/D76
- Crime Rates
  - Uniform Crime Reporting Statistics - UCR Data Online: https://www.ucrdatatool.gov/
- Drug and Alcohol related deaths
  - Center for Disease Control and Prevention (CDC): https://wonder.cdc.gov/controller/datarequest/D76

Formats:

- Pulled data from the following formats:
  - CSV (27 Total Files),

TXT (1 File), (note the .txt file was converter to a .tsv via Jupyter Are we keeping this? If so, need to add a brief statement about what the data is and how we plan to use it.

  - Notebook to read in and create a dataframe)
  - Webscraping (https://inkplant.com/code/state-latitudes-longitudes)

Types of transformation:

- Libraries/Modules
  - Pandas, numpy, sqlalchemy, os, modules
- Procedures
  - Cleaning, joins, groupbys, filtering, combining, removing random data in CSV's, dropped rows/columns/nulls, aggregating data, converted from multiple sources into dataframes, formatted data, added category numbering system
  - Original files were read into dataframes using .read_csv, .read_tsv
  - Large number of csv files were read using a function and combined to create one large dataframe

- Dataframes were converted to SQL tables using SQLALCHEMY

Type of final Database:

- Relational – SQL database

Final tables used in production database:

- LegalizationByState
- DeathTotals
- CrimeStats