Term Project Step 2

Burden Joshua

2022-05-22

Legalization of Marijuana

Introduction

With more and more states adding medical and recreational cannabis to their ballots, questions are now being asked more openly about the impact of cannabis on public health, crime statistics, and popularity of legalization. The data out there is sparse, but the data that has been collected could show trends and insights on the future of legislation and public sentiment of legalization, along with showing the potential for lowered crime and raised tax revenue in states and at the federal level.

Research questions

I want to know, based on the data that is out there, if legalization of marijuana:

- Reduces the number of arrests of all nonviolent offenses significantly
- Would increase the tax income for states and government entities
- Would drug usage go down by making it legal
- What is the state tax collections for legal rec states vs med only vs no rec/med
- What is the public support for legalization

Approach

I will look at the data that is out there for sales, crime, and public sentiment and see if there is anything that stands out to answer any of the questions above.

How your approach addresses (fully or partially) the problem.

The data will tell me whether or not the any or all of my questions can be answered and if there is discrepancies in the data, more research might be conducted to see if there is correlation.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

- https://data.world/denver/marijuana-related-crime
- https://www.kaggle.com/datasets/tunguz/drug-use-by-age
- https://www.kaggle.com/datasets/mykeysw/marijuana-sales-forecasting-in-tx
- https://data.world/sya/marijuana-laws-by-state
- https://data.world/denver/marijuana-gross-sales
- https://data.world/health/support-for-legal-marijuana
- https://data.world/opensavannah/cannabis-justice
- https://www.kaggle.com/datasets/terenceshin/historical-prices-for-biggest-weed-stocks

Required Packages

- Gplot2

Plots and Table Needs

- Histograms
- Scatter plots
- CDF
- Linear Regressions

Questions for future steps

Is there more data out there that would help with this research, and how deep should these models go in terms of covering the questions?

Step 2

At this point you should have framed your problem/topic, described the data, and how you plan to solve the problem. Now you need to move on to the next step of analyzing and preparing the data.

- Data importing and cleaning steps are explained in the text and follow a logical process. Outline your data preparation and cleansing steps.
- With a clean dataset, show what the final data set looks like. However, do not print off a data frame with 200+ rows; show me the data in the most condensed form possible.
- What do you not know how to do right now that you need to learn to import and cleanup your dataset?
- Discuss how you plan to uncover new information in the data that is not self-evident.
- What are different ways you could look at this data to answer the questions you want to answer?
- Do you plan to slice and dice the data in different ways, create new variables, or join separate data frames to create new summary information? Explain.
- How could you summarize your data to answer key questions?
- What types of plots and tables will help you to illustrate the findings to your questions? Ensure that all graph plots have axis titles, legend if necessary, scales are appropriate, appropriate geoms used, etc.).
- What do you not know how to do right now that you need to learn to answer your questions?
- Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.
- Some additional questions you may want to consider asking yourself as you work through this section of the project:
- What features could you filter on?
- How could arranging your data in different ways help?
- Can you reduce your data by selecting only certain variables?
- Could creating new variables add new insights?
- Could summary statistics at different categorical levels tell you more?
- How can you incorporate the pipe (%>%) operator to make your code more efficient?

You can use the following template for Step 2:

- How to import and clean my data
- What does the final data set look like?
- Questions for future steps.
- What information is not self-evident?
- What are different ways you could look at this data?
- How do you plan to slice and dice the data?

- How could you summarize your data to answer key questions?
- What types of plots and tables will help you to illustrate the findings to your questions?
- Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.
- Questions for future steps.