CSPB 4502: Data Mining Project Proposal

Jarryd Allison

Overview

- Title
- Team
- Description
- Prior Work
- Datasets
- Proposed work
- List of tools
- Evaluation

Title and Team

Title:

Colorado Public School Academic Performance

Team:

- Jarryd Allison

Description

I will be analyzing the academic performance of public school students from 2014 - 2023 in order to assess how performance has changed, what might have reasonably affected performance, and how public funds have been spent in relation to performance. I hope to better understand school performance and financing, and derive interesting and novel insights into how public policy might change to better Colorado schools.

Prior Work

I’ve done extensive research over the past two years into the public school system, notably focusing on pre-COVID statistics to better understand performance prior to those outlier years. I’ve researched multiple datasets, and also done some preliminary excel work to view the data.

Datasets

- CMAS - Mathematics, English Language Arts, Science and Social Studies Data and Results (2014 - 2023)
  - https://www.cde.state.co.us/assessment/cmas-dataandresults
  - Provided by the Colorado Department of Education
  - This is downloaded in csv format onto my personal computer
- Financial data
  - https://www.cde.state.co.us/schoolview/financialtransparency/homepage
  - Provided by the Colorado Department of Education
  - This is a queryable database only from the provided website. I will be researching more datasets to augment this.

Proposed work

- Data cleaning:
  - Much of the data is incomplete and/or inconsistent. I will need to both infer missing data as well as identify discrepancies and develop a way to process these so that it makes sense.
- Data preprocessing:
  - I will most likely conduct some form of data reduction, as I have access to many data points that may not be relevant to the topic.
  - I will also conduct some data normalization to identify relevant trends amongst various factors, such as test scores across subjects, aggregate pass/fail percentages, and so on.
- Data integration:
  - I will need to find a way to integrate financial and other relevant data into my dataset to better understand how finances and performance are related. To do this, I need to identify better sources, or develop a web scraper, to match financial data with each school code in my dataset.

List of tools we intend to use

- Python
- Jupyter Notebooks
- Possibly javascript
- Looker Studio
- Numpy/Pandas

Evaluation

I plan on evaluating my reports by comparing the results to other publicly available comparisons provided by third parties. There are several in Colorado. I can also compare the test results and finances nationally to determine if Colorado falls within the statistical bounds of other studies. Finally, I hope to be challenged by my classmates to ensure that I can explain the reasoning and steps taken such that the results are repeatable.