The Relationship Between Iowa Public School Budgets and Student Proficiency

Nicolas J. Vetter

Northwest Missouri State University, Maryville MO 64468, USA
`S567397@nwmissouri.edu` or `nvetter73@gmail.com`

Abstract. The writer will complete the Abstract last.

Keywords: Data Analytics · Iowa Public School · Budgets · Proficiency

Helpful Links:

– GitHub Capstone Repository:
  https://github.com/VetterNic2/msda-capstone

– Public Overleaf:
  https://www.overleaf.com/read/bhvmwnchhksz#8526a8

1 Introduction

Whether you grew up in the house of a teacher/parent or not, you probably have a memory or two about your high school life. Whether your school was rich or poor, all Iowa Public Schools have a duty to give their students the best chance to succeed after they graduate. In order to give their students the opportunity to be successful, the school system needs to invest their resources into their pupils. This project is going to show which Iowa Public Schools invested the most/least in their students. This report will also give the reader an idea of how much of an investment a school should spend to make their student the most successful they can possibly be, and whether there is a correlation between proficiency and investment per pupil. Success will be measured by proficiency scores in this report. These scores will be cleaned in PostgreSQL, analyzed through machine learning in Python and visualized in Tableau. The data sources and references are shown in their respective sections.

Limitation: Iowa Public schools have several different budgets for all kinds of expenditures. Because of the short report time-frame, the writer will not be able to analyze every budget within the Iowa Public School system. The writer is going to keep it simple and analyze the budget every Iowa Public school used in the 2017 academic year: general instruction.

1.1 Goals of this Project

2 Data Sources

– Math and Reading Proficiency in Iowa by School Year:
  https://data.iowa.gov/Primary-Secondary-Ed/Math-And-Reading-Proficiency-in-Iowa-by-School-Yea/f3h8-mnxi/about_data

– Iowa School District Expenditures by Fiscal Year:
  https://data.iowa.gov/School-Finance/Iowa-School-District-Expenditures-by-Fiscal-Year/uutu-bzs3/about_data

2.1   Data Collection

– **Math and Reading Proficiency in Iowa by School Year:**

This data set was collected from the Iowa.gov website as public information. The Iowa.gov website has an "Action Query" function that helps the writer filter the necessary data from the dataset. This is called preliminary cleaning of the dataset and the dataset will be cleaned again in section 3 of this report. The writer will only be using the reading proficiency rating of the 2017 11th-grade students in each respective district. This is to shorten the report for the tight time window.

– **Iowa School District Expenditures by Fiscal Year:**

This data set was collected from the Iowa.gov website as public information. The Iowa.gov website has an "Action Query" function that helps the writer filter the necessary data from the dataset. This is called preliminary cleaning of the dataset and the dataset will be cleaned again in section 3 of this report. The writer will only be using one year's worth of data from the 2017 fiscal/academic year. This is to shorten the report for the tight time window.

2.2   Data Description

– **Math and Reading Proficiency in Iowa by School Year:**

The total storage space of this structured dataset is 48KB. It contains 610 records and 14 attributes. However, the writer won't be using all of the 14 attributes in this report. Some were deemed irrelevant to the analysis and report. With that in mind, the attributes to be used for analysis are as follows with the datatype shown in parenthesis behind the attribute: School Year(Number), Topic(Text), Grade(Number), District(Text), District Name(Text), Percent Proficient(Number), Proficient Category(Text).

– **Iowa School District Expenditures by Fiscal Year:**

The total storage space of this structured dataset is 366KB. It contains 3509 records and 13 attributes. However, the writer won't be using all of the 13 attributes in this report. Some were deemed irrelevant to the analysis and report. With that in mind, the attributes to be used for analysis are as follows with the datatype shown in parenthesis behind the attribute: Year(Number), Dist(Text), District Name(Text), Fund(Text), Expenditures Per Pupil(Number), Amount(Number), Enrollment Category(Text), Enrollment Category Number(Number).

3   Data Cleaning/Manipulation using PostgreSQL

3.1   Cleaning Process

– **Process and Tools of Data Cleaning:**

The preprocessing of data was very minimal for these data sources. This is because of the lack of missing values within the CSV files. The only missing values were related to the geographic location of school districts, and for this report, that was deemed irrelevant. Because of the irrelevance, those missing values were dropped from the SQL tables. There were no relevant missing values because Iowa.gov/data pre-cleans data so it is accurate and ready for third-party analysis.

Since the data came from two different CSV files, SQL needed to be used to clean and join the two files together. Specifically, PgAdmin and PostgreSQL were selected because of the ease of table creation and import/exporting of CSV data. First, the tables needed to be created using the PgAdmin interface

and copy statements for both CSV files. Second, the unnecessary attributes needed to be dropped from the tables to ensure simplicity and accuracy. Along with the unnecessary attributes, this report is only going to analyze reading proficiency. This decision was made because of the tight timeline of the report and the need for only one dependent variable. Third, the two revised tables needed to be joined with a JOIN statement. After the JOIN statement ran successfully, the new data table was saved to a CSV file for further analysis.

3.2 JOIN Statement and Figures

– **PostgreSQL JOIN:**

The two CSV files described in the previous section were joined to only show relevant attributes and information. This will make it much simpler to analyze if there is only one CSV file to run through the machine learning model and create visualizations. The JOIN statement can be seen at the bottom of Fig. 2.

– **Cleaning/Manipulation Figures**

Fig. 1. Create/Clean District Expenditures
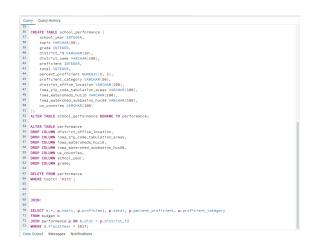

Fig. 2. Create/Clean Proficiency and JOIN

3.3 Clean Data Overview

– **Attributes and Variables:**

The attributes after the cleaning process was completed are as follows:
"fiscalyear","dist","district name","source","expenditures per pupil","amount","enrollment category","enrollment category number","topic","proficient","total","percent proficient","proficient category".

Below is a sample of the CSV's first line of the cleaned data:
2017,"0009","AGWSR","Instruction",7989,4997256,"600-999","3","Reading",32,41,"78.00","70.1 - 80".

If the data shown in the line above is surrounded by "", it is a string/text data type. However, if the data shown in the line above is surrounded by nothing, it is an integer data type.

To align with the goal of the report, the dependent variable of the project is "percent proficient" and the independent variables are all of the other attributes contained within the cleaned CSV file(outlined above). However, the main independent variable that will be analyzed is "expenditures per pupil".

4   Exploratory Data Analysis

5   Machine Learning in Python - Correlation

6   Tableau Visualization of Results

7   Conclusion

[]

References

1. Dhaliwal, T.K., Bruno, P.: The rural/nonrural divide? K–12 district spending and implications of equity-based school funding. aera Open **7**, 2332858420982549 (2021), DOI: 10.1177/2332858420982549
2. Johnson, J.: More doesn't mean better: Larger high schools and more courses do not boost student achievement in iowa high schools. Rural School and Community Trust (2006), ERIC Document: ED491173