Project Description

The data from the Pennsylvania Department of Education is distributed across various segments of the official website. Unfortunately, there does not exist a centralized repository that consolidates all information about individual schools or facilitates seamless comparisons among schools, stratified by levels (elementary, middle, high), or across various districts. This lack of a unified data resource hampers the ability of service providers, residents, parents, and social impact funders to gather the most comprehensive view of what is going on in the community.

In this project, you will need to combine and make searchable data points located in different Pennsylvania Department of Education datasets for a more holistic view of individual schools and school districts in the Lehigh Valley (and possibly Pennsylvania). The goal of this project is for you to download all the data sets listed below, design a database for them, clean up and upload the data, and build meaningful data visualization applications.

The following data about schools in Pennsylvania should be included in the overall project.

1. School facts from school year 2017-2018 to the present.
2. District facts from school year 2017-2018 to the present.
3. School fiscal data for school year 2018-2019.
4. District fiscal data from the school year 2016-2017 to the school year 2018-2019.
5. Future ready performance data from school year 2017-18 to the present.
6. Low income data for public schools from school year 2005-2006 to the present.
7. Low income data for private and non-public schools from school year 2005-2006 to the present.
8. Keystone Exam state level/school level data from year 2015 to 2022..
9. PSSA school level/state level data from year 2015 to year 2023.
10. Academic performance data from the school year 2012-13 to 2016-17.
11. Graduates Public by school from school year 2004-2005 to the present.
12. Dropout data for public schools from school year 2004-2005 to the present.
13. Cohort graduation rates from 2010-2011 to the present.
14. Enrollment reports for public schools from 2004-2005 to the present.
15. Enrollment reports for private schools from 2005-2006 to the present
16. Financial data:
    - Annual Financial Report (AFR) data (Minimum requirement is Summary Level: Expenditure and Revenue Data)
    - Financial Data Elements

- General Fund Budget (GFB) data
- Historical Subsidy Files data

17. Professional and Support Personnel data

For this group project, each team will have 3-4 members. In the last week of class, each group will present the application that they develop.

Team Project Phase I

Things to complete before the due date:

Due 09/06 at midnight

1. Send me the members of your team and your team name. Each team will have 3 to 4 members. Make sure to create a GitHub repository and add me to it.

Due 09/12

2. Collect data from links provided. Five extra points will be given to students that write a crawler instead of downloading files manually. Make sure that you add that file to your repository.

GitHub

3. Pick a web framework to develop this project. The database system that we will use is SQLite. For the team project, each team will have a shared account in the server so all team members can access and manipulate the database. You need to create a simple LOCAL project using your framework. The project must:
    a. Read, Write, Update from a Database
    b. Have at least one screen that displays information from the database.

Some examples:
Django: https://docs.djangoproject.com/en/5.1/intro/tutorial01/
Ruby on Rails: https://guides.rubyonrails.org/getting_started.html
Flask: https://flask.palletsprojects.com/en/3.0.x/tutorial/

GitHub

4. Define your research question.
    a. What application do you want to develop?
    b. What is the focus of your application?

c. Who are your potential users?

Note: You will have the opportunity to revisit this part one more time.

PDF

For your reference, Future Ready PA Index already provides some visualization that may be helpful for you to think about.