CAPP 122 Final Project Winter 2023
March 8, 2023
Amari Bauer – baueramari
Eshan Prashar – eshanprashar
Sarah Walker – swalker10

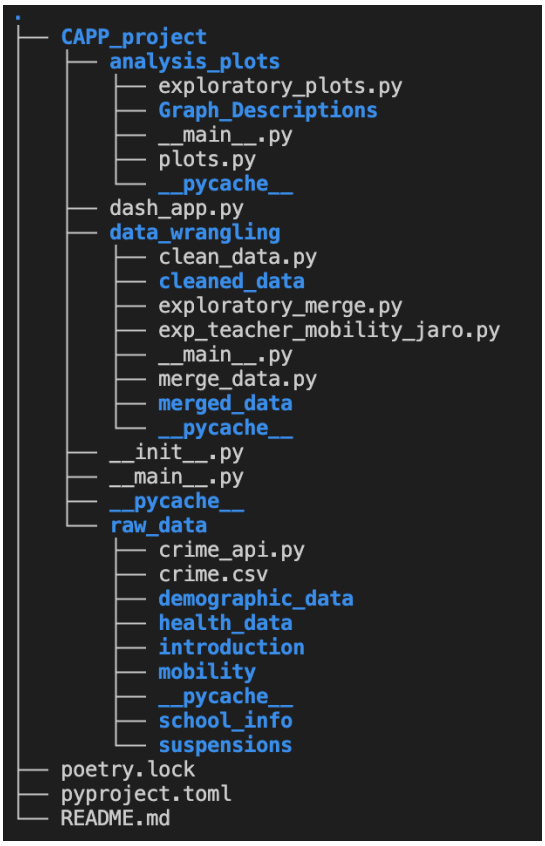
Panic At The School

Abstract

Our aim of this project is to assess factors which may be contributing to low attendance rates in Chicago Public Schools (CPS). We focus specifically on high schools (grades 9-12), which have shown the lowest attendance rates out of all grade levels in the last decade. Moreover, there was a noticeable drop in high school attendance post-COVID, with rates falling from around 86.51% before 2019 to an average of about 78.79% in 2021 and 2022 following the pandemic. In our investigation we use data from CPS (cps.edu), an API from the City of Chicago Data Portal, and other publicly available data sources to investigate factors in attendance. Based on news sources, we divided our study into two focus areas: the relationship between school misconduct and community crime rates as an influence on the criminalization of students, and the impact of COVID in schools. During this process, we used pandas to clean our data, plotly to create data visualizations, and dash to launch our web page/application. While our findings suggest a few variables which show some, or no, relationship to attendance, further analysis would be necessary to fully understand what has contributed to its declining rate in CPS high schools.

Overall Structure of Software

Within the PanicAtTheSchools repository, our package is titled CAPP_project. The bulk of our package is split into three folders here, organized in a data pipeline. All raw data is collected, including the file to access the API, in the **raw_data folder**. From there, the data heads to the **data_wrangling folder**. Within the data_wrangling folder, all data is read in and cleaned, before being merged with other data files and exported for analysis and plotting. The **analysis_plots** folder loads in the merged datasets and contains all code to visualize those datasets, as well as final graph analysis descriptions, to be used in our Dash app. Outside of those three main folders is dash_app.py, the file to create and launch our Dash app, as well as our _init__ and __main__ files to operate the files from the terminal and make it interactive. Sub-packages data_wrangling and analysis_plots also have their own __main__ files.



Member Responsibilities

In the initial stages of the project, every member was responsible for researching and investigating data sources. Once we gained an idea about the scope of our project plan, we roughly divided the responsibilities as such: Eshan would complete a majority of the data cleaning and merging, Sarah would complete a majority of the data visualization, and Amari would complete a majority of the web application. However, since we worked together for many tasks, we each had a hand in, or made additional contributions to, every section of the process.

Amari sourced the crime data through the City of Chicago's Data Portal API. He handled and cleaned this data as well as the general attendance data downloaded from the Chicago Public Schools website. He also built the base structure for our repository by creating all the folders and files we would work with throughout the process. Next, he merged necessary columns of these two sources and saved the output in our repository as merged data files. He used plotly to create an attendance vs. crime data visualization (line graph). Furthermore, he can be credited for writing all of the dash_app.py file which created our dash web page. In the final stages of the project, he collaborated with Eshan to ensure our pathways and files were integrated properly for our package to run as a single command line, and have files speaking to each other correctly.

Sarah downloaded and cleaned misconduct/suspension data from the CPS website. She merged columns from her suspensions data frame with columns from Amari's crime and attendance data frames, and saved these merged files in the repository. She used plotly to create visualizations for this data, visualizations for the merged data files Eshan had completed, and

visualizations to be included on the introduction page of our web application. In the final stages of the project, she spent time formatting the graphs that would be used in the final output, and adjusted code in the plots.py file as needed. She performed analysis on half the graphs that would be included on the webpages and wrote descriptions for them.

Eshan downloaded and cleaned data files from the CPS website, and from other publicly available sources which captured information on neighborhood demographics, school demographics, health and COVID statistics, school finance data, and teacher data. He analyzed and merged the necessary variables, and saved these merged files in our repository. He also conducted a lot of extra work behind the scenes to investigate these different factors of attendance, including plotting his own data to search for any relationships and using a jaro-winkler function to categorize/clean teacher data files, but much of this was omitted from the final output. He performed analysis on, and wrote descriptions for, the other half of graphs which were included on our webpage. In the final stages of the project, he was responsible for writing the main.py files which allow a user to run the package from the terminal. He and Amari collaborated to structure the sub-packages and understand how all our files interacted.

Guide on How to Interact With the Application

These are the steps to interact with the dash application (also mentioned in README):

1. ~/PanicAtTheSchool$ **poetry install**
2. ~/PanicAtTheSchool$ **poetry shell**
3. (in the shell) ~/PanicAtTheSchool$ **python3 -m CAPP_project**
4. From here, you can choose to directly launch the dash app by entering '1', or you may examine each step in the data pipeline by entering '2.'
5. If you choose '1', select the pop-up message "Open in Browser" appearing on the bottom right of the screen to launch the web page.
6. If you choose '2', you have several options:
   - fetch: runs the API data collection process to create or update our crime.csv
   - clean: loads in our data sets, filtering and managing unusal or unknown values
   - merge: loads in our cleaned data sets, and merges them for further analysis and plotting.
   - exploratory: fetches teacher and mobility, matches with school data using Jaro-Winkler and merges files; this data is not used for final analysis
   - plot: runs all the code to create visualizations
   - all: runs the entire data pipeline, cleaning, merging, plotting and ultimately launching our Dash appplication.
7. When you have fully explored the webpage and are back in your terminal, run ctrl+c to close the application.

Note 1: Pulling data from the Chicago Data Portal API (as would happen if you selected 'fetch' or 'all') requires a private token to be assigned to TOKEN at line 16 of the 'crime_api.py' file, instead of using the environmental variable as discussed in class. Professor Turk has been provided this private token for grading purposes. Pulling this data takes approximately 10 minutes.

Note 2: If you do not wish to pull data from the API or have not been given the private token, you can access the static crime data file at the below link. This file can be downloaded as a csv at the link, placed in the 'CAPP_project/raw_data' subdirectory, and renamed 'crime.csv.'

Link to crime CSV:
https://uchicago.app.box.com/file/1154307867116?s=3tpb8dibaad6i1usrc5dr96cymzdv7m4

What Did We Try to Accomplish, and What Was Actually Accomplished

From the early stages of this project, we were interested in assessing declining attendance rates in Chicago Public Schools (CPS) over the last decade. Originally, we planned to investigate numerous variables categorized as either in-school, or out-of-school factors which influenced a student's decision, or ability, to attend school. The purpose was to visualize how these factors differed across schools, and if plotting them against attendance rates in those schools would show a negative or positive relationship.

After reading news sources, finding and discussing available data, choosing packages to use, and imagining how we wanted to visualize the output, we modified the scope of our project so that the output would be more focused and attainable. Upon completion, our project was able to accomplish the cleaning, merging, and analysis of CPS attendance data, CPS suspension data, school finance data, demographic data, and City of Chicago crime data. We produced a web application which displays our graphs and organizes the analysis of our information onto four webpages.

The results of the project accomplished the following. On the introduction page, we show that grades 9-12 have the lowest attendance rates over the last decade, with the greatest decline occurring in the years following the COVID pandemic.

On the second page of our web application, Misconduct, we relate community crime levels (ranked as low, medium, or high) to attendance and school misconduct data. Our graphs show that high schools located in low crime neighborhoods have the lowest attendance rates and highest suspension rates, and high schools in high crime neighborhoods have the opposite. In-school suspensions are assigned at higher rates than out-of-school suspensions across all high schools. The percentage of misconducts leading to a suspension showed no relationship to high school attendance rates, and the percentage of police notifications made were relatively the same across schools in all neighborhoods.

On the third page of our web application, Impact of COVID in Schools, we categorize schools into their pre vs. post COVID attendance levels. We visualize the change in pre and post COVID attendance rates for each category. Moreover, we show that high schools with high levels of high-income students are also schools with historically high attendance rates, and high schools with historically low attendance rates are receiving the most dollars per student.

On our conclusion page, we speak to the additional data sources cleaned within our repository (but omitted from the final analysis) that can be used in further research to create more detailed regressions of in-school and out-of-school elements which could further explain declining attendance in Chicago Public High Schools.