NJ Title 1 Schools Analysis - Process Documentation

Project: Title 1 Schools Analysis (Grades 5-8)
Analyst: Swastik Bhatnagar
Date: November 2025
Tools: Python (Pandas, PyPDF2), Google Looker Studio

1. Process Overview

a. How I Extracted the Data

I worked with three data sources to complete this analysis:

Enrollment Data (Excel):

- Downloaded the 2024-2025 enrollment file from NJ Department of Education website
- The file contained multiple sheets; I loaded the "School" sheet which had enrollment data by grade level
- Key challenge: Column headers were in row 3 instead of row 1, so I specified `header=2` when reading the Excel file

Title 1 Schools List (PDF):

- Extracted Title 1 district names from the federal PDF document using PyPDF2 library
- The PDF had an unstructured format with district names mixed with codes and funding amounts
- Manually cleaned and created a list of 47 Title 1 district names from the extracted text

School Directory:

- The enrollment Excel file already contained school names, districts, and counties, so I didn't need to separately access the online directory

b. How I Merged the Data

The merging process had three steps:

1. **Filtered for grades 5-8:** I identified columns for Fifth Grade, Sixth Grade, Seventh Grade, and Eighth Grade, then calculated the total enrollment across these four grades for each school

2. **Filtered for Title 1 schools:** I created a matching function that checked if each school's district name contained any of the 47 Title 1 district names I extracted from the PDF. This

identified 102 Title 1 schools with 22,731 students in grades 5-8

3. **Created two ranking datasets:**

    - **County rankings:** Grouped schools by county and summed total students, then ranked counties from highest to lowest
    - **School rankings:** Ranked schools within each district based on grades 5-8 enrollment

c. Language and Tools Used

- **Python** with **Pandas** library for all data processing and analysis
- **PyPDF2** for extracting text from the Title 1 schools PDF
- **Google Colab** as the coding environment
- **Google Looker Studio** for creating interactive visualizations
- **CSV exports** for sharing clean data with stakeholders

d. Challenges and Solutions

Challenge 1: Excel headers in wrong row

- *Problem:* Column names appeared in row 3, causing initial load to show "Unnamed" columns
- *Solution:* Used `pd.read_excel(file, sheet_name='School', header=2)` to correctly identify headers

Challenge 2: Extracting Title 1 districts from messy PDF

- *Problem:* PDF had unstructured text with district names mixed with codes (e.g., "3400660 ABSECON CITY 90,622")
- *Solution:* Used PyPDF2 to extract all text, then manually cleaned and created a curated list of 47 district names. While I attempted automated parsing, manual verification ensured accuracy given the importance of correctly identifying Title 1 schools

Challenge 3: District name matching

- *Problem:* District names in the enrollment data didn't always exactly match PDF names (e.g., "Absecon Public Schools District" vs "ABSECON CITY")
- *Solution:* Used partial string matching with `.apply(lambda x: any(t1_dist.upper() in str(x).upper()))` to catch variations

2. Key Takeaways for Jersey STEM

Takeaway 1: Geographic Concentration in Coastal Counties

Atlantic County has the highest concentration of Title 1 students in grades 5-8 with 4,202 students across 19 schools. The top three counties (Atlantic, Hudson, and Monmouth) represent over 45% of all Title 1 students in these grades. This suggests Jersey STEM should prioritize partnerships in coastal regions where Title 1 schools are most concentrated, potentially maximizing impact per outreach effort.

Takeaway 2: Large Variation in School Sizes Creates Different Program Needs

Title 1 schools vary dramatically in size, from small schools with fewer than 50 students in grades 5-8 to large schools with 400+ students. For example, Emma C Attales in Atlantic County serves 392 students while nearby H Ashton Marsh serves only 2 students in these grades. This means Jersey STEM should develop flexible program models—some schools need comprehensive programs for hundreds of students, while others need targeted interventions for smaller cohorts. A one-size-fits-all approach won't work effectively.

---

3. Deliverables

a. Code

- Google Colab notebook with complete Python code for data extraction, cleaning, merging, and analysis
- All code is documented with comments explaining each step
- Available at: [https://colab.research.google.com/drive/1OFa3RjirxS40xCM2m0Bd5_YvM6QnVLX5?usp=sharing]

b. Visualization

- Interactive Google Looker Studio dashboard with two views:
    1. **County Rankings:** Bar chart showing counties ranked by total Title 1 students (grades 5-8)
    2. **School Rankings:** Detailed table showing all 102 schools with their enrollment and rank within their district
- Dashboard link: [https://lookerstudio.google.com/s/lgwviowAYV8]

c. Data Files

- `county_rankings_title1.csv` - County-level aggregated data
- `school_rankings_title1.csv` - School-level detailed data with district rankings

Summary

This analysis successfully identified and ranked 102 Title 1 schools across 19 NJ counties serving 22,731 students in grades 5-8. The data reveals significant geographic concentration in coastal counties and wide variation in school sizes, providing actionable insights for Jersey STEM's program planning and resource allocation.

Total Title 1 Schools Analyzed: 102
Total Students (Grades 5-8): 22,731
Counties Represented: 19
Data Sources: 3 (NJ DOE Enrollment, Federal Title 1 List, School Directory)