Carnegie Mellon University

PPS Retention/Mobility Research

Client: Steven Greene from Pittsburgh Public Schools
Project Members: Huiyi Guo, Jenny Luo, Yuhang Ying
Project Advisor: Zach Branson

Contents

Introduction

- Client: Pittsburgh Public Schools, Steven Greene
- Project Overview:
  - The Pittsburgh Promise funds scholarship for post-secondary education.
    - GPA >= 2.5, attendance rate >= 90%
  - How promise scholarship influences post-secondary retention/mobility
  - New criteria on choosing eligible students to fund
  - IMRAD paper
- Research questions:
  - What is the relationship between the Promise scholarship use and students' post-secondary retention and completion?
  - What are the factors that would affect students' retention and mobility among post-secondary institutions?

Data Sets

11 Data Sets:

- School Enrollment
- Course Enrollment
- **Attendance**
- Demographics
- NSC
- SAT
- AP
- **GPA**
- Keystone
- CTE
- **Scholarship**

Data Description

- GPA
  - Cumulative GPA for each school year
  - We focus on Senior students and their Senior GPA
  - GPA>=2.5 is required for Promise Scholarship
- AP
  - AP course name and AP exam score
  - We summarize students by # of AP courses taken, avg AP score
- Course Enrollment
  - Courses taken by each student per semester
  - Credit earned by each course and the grades
  - We summarize students by number of AP and IB courses taken

Data Description

- Scholarship
  - Information about eligibility of the students who applied and their acceptance of the Promise Scholarship, one row for each student
  - Including student ID, graduation year, eligibility to core and extension Promise, status of receiving scholarship, and high school name
    - *QualifiedforCorePromise* = binary variable for eligibility
    - *EverReceivedPromiseAward* = binary variable for receiving scholarship
  - Join the scholarship data with other data sets to conduct EDA

Data Description

- School Enrollment
  - Enrollment records to and from PPS schools before 9th grade
  - Non-PPS schools in Pittsburgh are not included
- Attendance
  - Grades 9-12
  - The number of days attributed to a student for each unique school year, school, and attendance code.
  - Attendance Status: "Absent Excused", "Absent Unexcused", "Present", "Present Excused", "Present Unexcused".
  - We calculate the attendance rate by: 1- (“Absent Unexcused” / ”totalDays”)
- SAT
  - The highest SAT totals for the 4 years in high school
  - Multiple rows for one student if student has earned the same score on multiple trials
- Keystone
  - Test scores for the Keystone Assessment

Carnegie Mellon University

Data Description

- Demographics
  - Demographic information for students' each year in high schools
  - Including student ID, grade level, race, gender, economic status, special education
- NSC
  - Semester enrollment information for students in college institutions
    - Each row = every semester a student enrolled in a college institution
  - Including student ID, high school graduation date, enrollment date, enrollment status, school type
  - Not available for 2019-2020 enrollment
- CTE
  - Career and Technical Education (CTE) certifications earned by each student
    - Each row = information of a certification earned by a student
  - Including student ID, certification category, credential information, and certification earned date
  - We calculate the number of certifications earned by each student

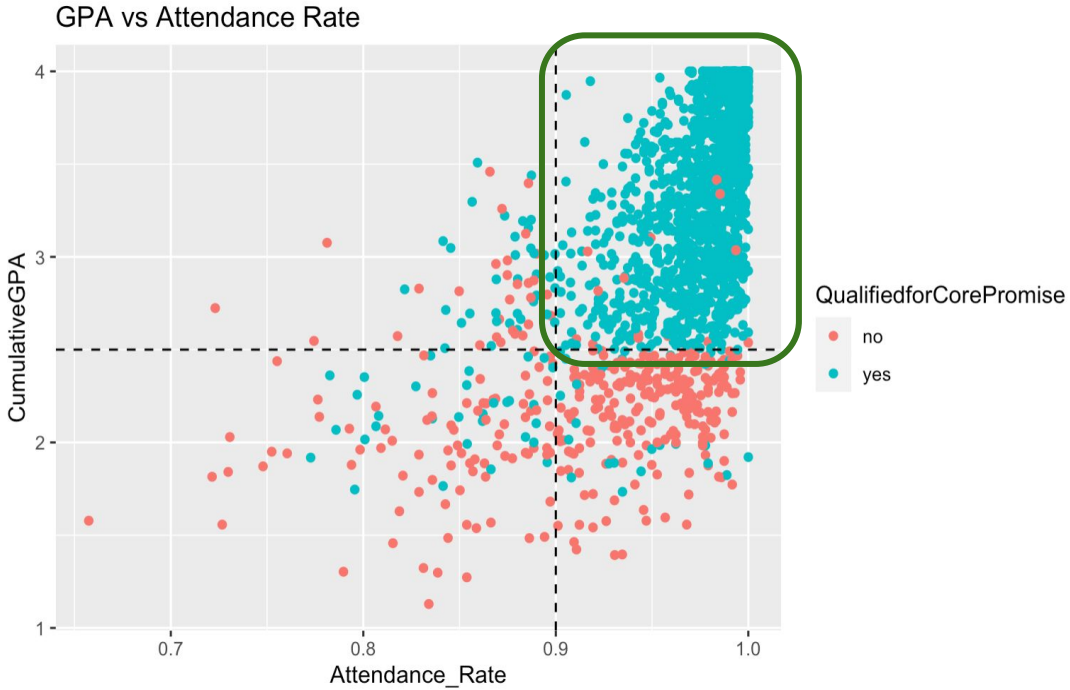
Methods

- Preliminary EDA
  - Students receive scholarship after they graduate
    - Summarize all metrics for senior students, 1 row per student
  - GPA x Attendance x Scholarship analysis
    - 3846 senior students with their GPA and attendance rates, among them 2187 applied
    - Eligibility cutoff: GPA >= 2.5, Attendance rate >= 90%(PPS website)
      - Analyze students application results around the cutoff
  - Demographics x Scholarship analysis
    - Combine demographics data of different cohorts together → overall demographics data
    - Join demographics and scholarship data by RandomID and GradYear
      - Obtain 2223 senior students
    - Explore eligibility vs. gender, race, economic status, and IEP group

Methods

- Preliminary EDA
  - NSC x Scholarship analysis
    - Filter first semester enrollment records in colleges → filtered NSC data
    - Join filtered NSC data and scholarship data by RandomID and GradYear
    - Explore eligibility & acceptance of scholarship vs. school type
  - CTE x Scholarship analysis
    - Join CTE and scholarship data by RandomID
    - Explore eligibility vs. #certifications/category, and acceptance of scholarship vs. #certifications/category
- Modeling
  - Explore relationship between scholarship and other factors
  - Compare enrollment status for those who received scholarship vs. who didn't
  - Account for the effect from Scholarship eligibility cutoff (GPA & attendance)

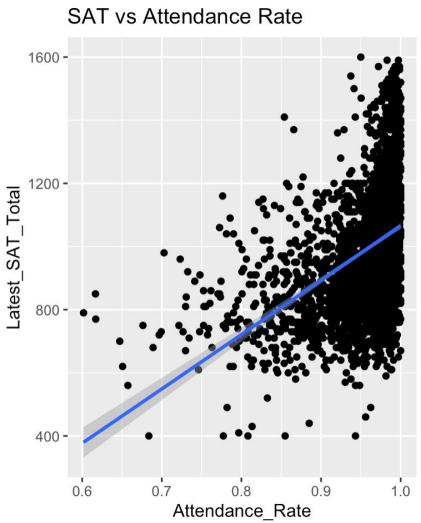
Results

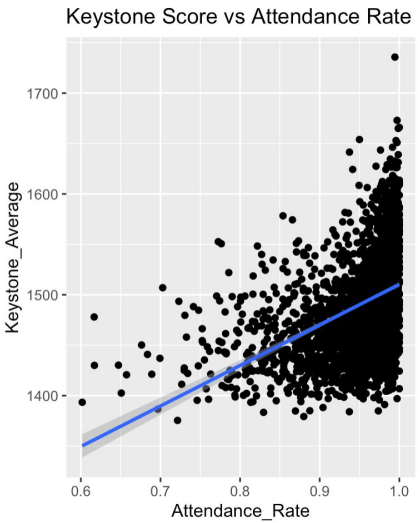


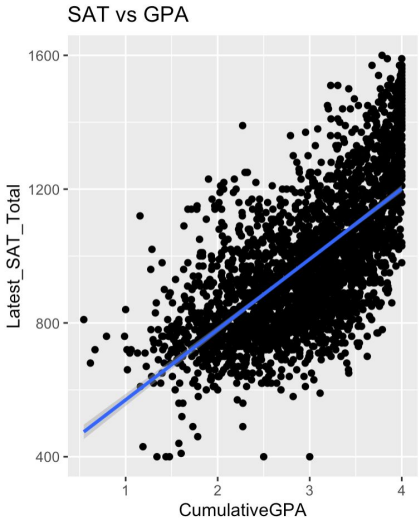
- Black Dashed Lines:
  - GPA = 2.5
  - Attendance Rate = 0.9
- Green square:
  - Students who qualified for Core Promise

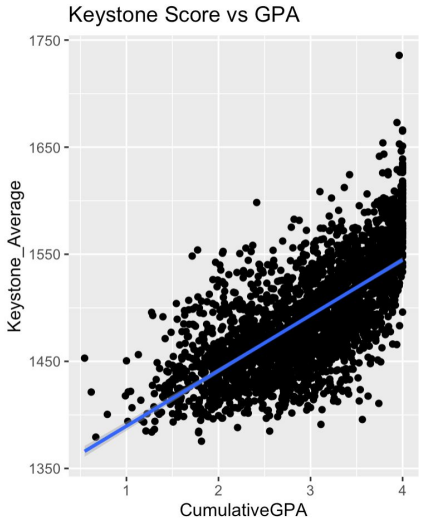
Carnegie Mellon University

Results







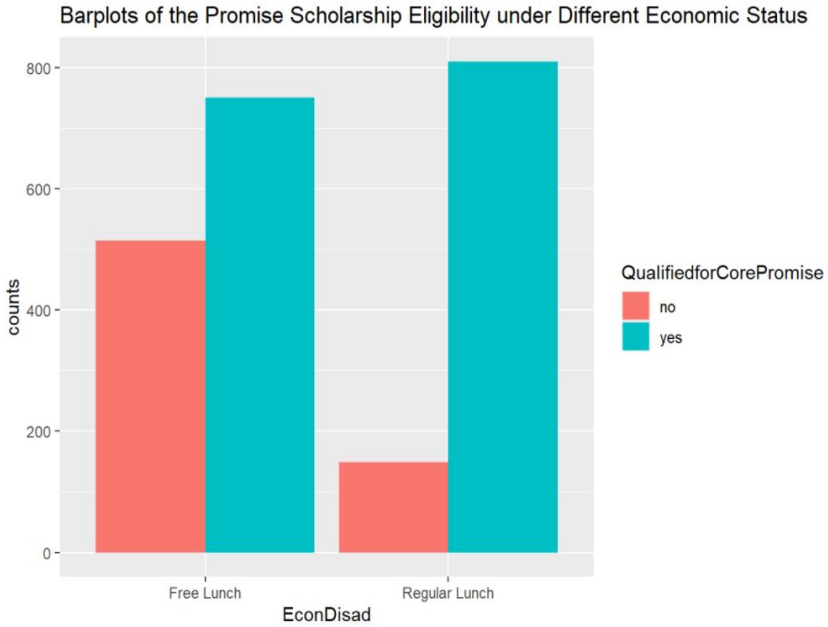


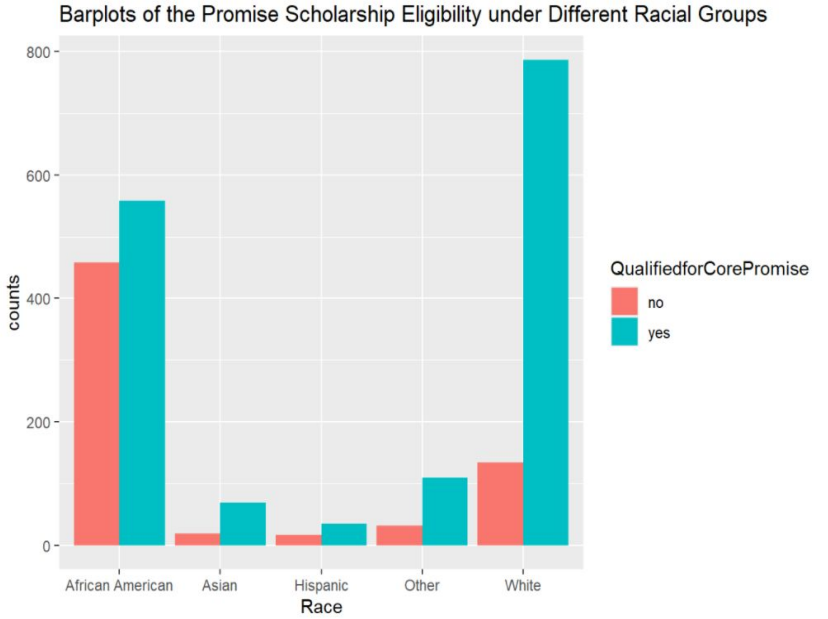
Correlation: 0.43 0.42 0.68 0.69

- Collinearity?

Carnegie Mellon University

Results

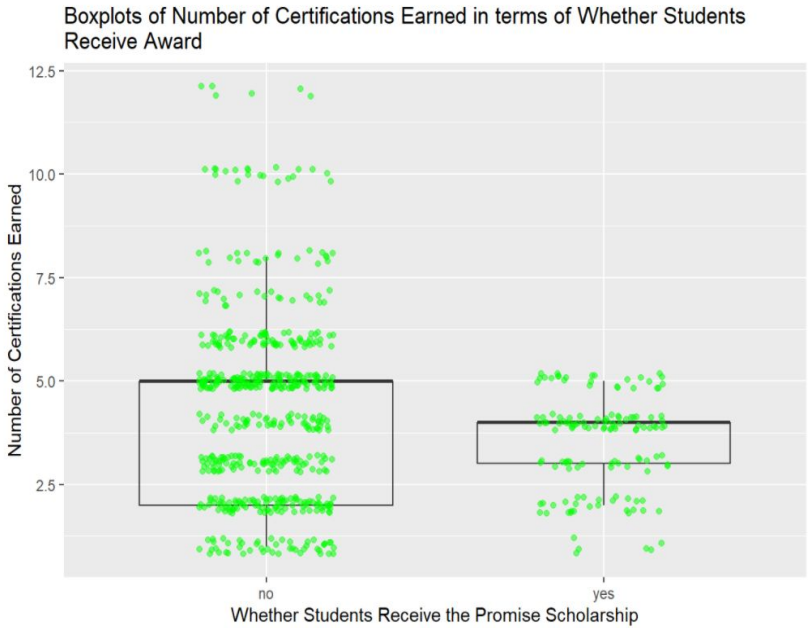


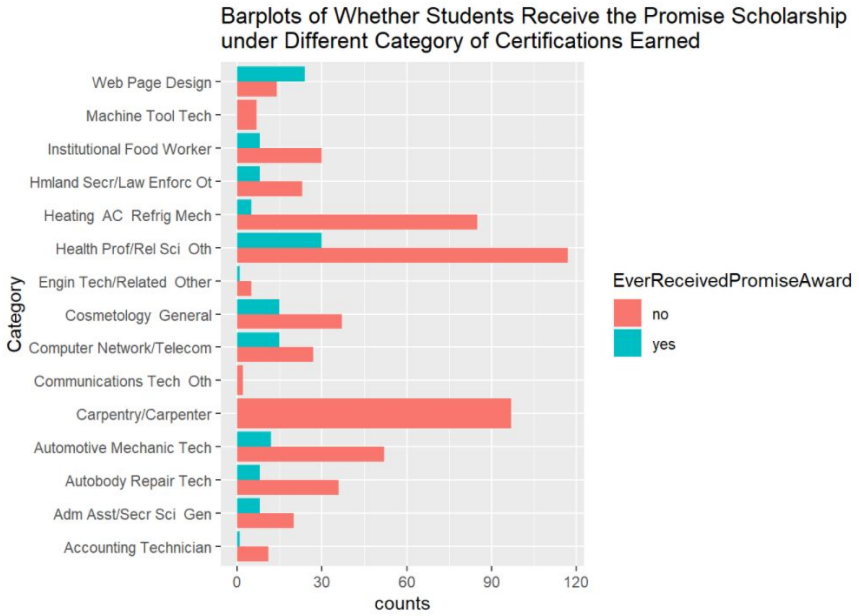


- Better economic status → more likely to be qualified
- The proportion of eligibility seems highest among white people.

Carnegie Mellon University

Results





- More #certifications earned → less likely to receive
- For the majority, the proportion of not receiving > the proportion of receiving

Carnegie Mellon University

Next Steps & Roadblocks

- Meeting with our client next week
  - Report preliminary EDA findings
  - Confirm our understanding about the scholarship data
  - Discuss the feedbacks from client
  - Discuss next steps
- Intermediate goals to solve research questions

Carnegie Mellon University

Next Steps & Roadblocks

- Roadblocks that prevent progress
  - Lack understanding about data
    - Need to confirm our understanding about the scholarship data
    - No specific qualification requirements about SAT & AP
  - Unsure about how to treat missing/repetitive/weird observations in data
  - Non-immediate reply from the client regarding our questions, rescheduled meeting next week.

Questions?