

Stage 2: Research Design

Joya Wheatfall-Melvin

PPOL 5204: Data Science II

Professor Ioannis Ziogas

February 25, 2025

- **Title Options:**
 - Analyzing the Integrity of Alternative Schooling for Black Girls
 - Focusing on 2020-2021 CRDC Data (Potential COVID influence)
- **Double-Edged Sword:** What is the integrity of the education provided in juvenile education centers? And do they take care of their black female population? (Should I make this state-specific?)

Background

Despite Black women achieving higher education degrees at consistently high rates over the last decade, significant geographic variation in educational attainment persists across the United States (National Center for Education Statistics). The integrity of education provided in juvenile education centers further complicates this issue, particularly for Black girls who face barriers at the K-12 level due to factors like “adultification bias” leading to over-disciplining, underrepresentation in advanced courses, and limited access to gifted programs. While Black women earned 68% of associate’s degrees, 66% of bachelor’s degrees, 71% of master’s degrees, and 65% of doctorate degrees awarded to Black students between 2009 and 2010, these achievements underscore the importance of an equitable K-12 foundation, which remains inconsistent (National Center for Education Statistics).

Research Question:

- Which school characteristics are most predictive of high versus low retention rates for Black female students in alternative and public schools?
 - Sheds light on the integrity of the education provided in juvenile education facilities

Justification:

The integrity of education provided in juvenile education centers has been a contentious issue, however, there is limited research on the treatment and educational outcomes of Black girls. While the overall population of incarcerated youth has decreased since 2000, the reduction of incarcerated disabled Black girls has remained stable (Cabral 2003). This disparity raises questions about the adequacy and equity of educational resources in these settings. The purpose of alternative schooling is to offer specialized care to students with disabilities and students with disciplinary issues a personalized and better educational environment to succeed in school. Alternative schools overrepresent black and brown students. If these alternative schools are not efficient, then we are failing a particular group of students. This study will highlight whether alternative schools serve Black girls effectively or if they contribute to pushing them out of the education system. Additionally, this model will allow us to identify which factors impact the integrity of the education they receive in alternative vs. regular schools for comparison. (Another model could be added later for comparison)

Data

I. **Descriptives:** Metrics that reflect the quality and inclusivity of education for Black female students in Alternative Schools

- **Target Variable / Outcome:** RETN-1: Student Retention Indicator
 - A student is retained if he or she is not promoted to the next grade before the beginning of the following school year. Students are not considered retained if they can proceed to the next grade because they completed a summer school program or for a similar reason. At the high school level, a student who has not accumulated enough credits to be classified as being in the next grade is considered retained.
 - Variable Names for Black Female Retention (K-12) :
 - SCH_RET_KG_BL_F
 - SCH_RET_G01_BL_F
 - SCH_RET_G02_BL_F
 - SCH_RET_G03_BL_F
 - SCH_RET_G04_BL_F
 - SCH_RET_G05_BL_F
 - SCH_RET_G06_BL_F
 - SCH_RET_G07_BL_F
 - SCH_RET_G08_BL_F
 - SCH_RET_G09_BL_F
 - SCH_RET_G10_BL_F
 - SCH_RET_G11_BL_F
 - SCH_RET_G12_BL_F
- **Feature Variables / Predictors:** What features can predict the educational integrity of alternative schooling? Most variables account for black female student population, some indicators account for school resources (e.g. access to dual degree programs) :
 - Student Discipline (Suspension, Expulsion, Corporal Punishment)
 - DISC-7a: Discipline of Students without Disabilities - Corporal Punishment
 - DISC-8a: Transfer to Alternative School for Students without Disabilities
 - DISC-9a: Discipline of Students with Disabilities - Corporal Punishment
 - DISC-9e: Discipline of Students with Disabilities - Expulsions with Educational Services
 - Advanced Placement (AP) & International Baccalaureate Diploma Programme (IB) Enrollment
 - APIB-2: Student Enrollment in the International Baccalaureate (IB) Programme
 - APIB-6: Advanced Placement (AP) Student Enrollment
 - Harassment or Bullying (
 - HIBS-1a: Allegations of Harassment or Bullying - Sex, Race, Color, National Origin, and Disability
 - HIBS-2a: Students Reported as Harassed or Bullied - Sex

- HIBS-3c: Students Disciplined for Harassment or Bullying - Disability
- Restraint & Seclusion
 - OFFN-5: Offenses - Number of Incidents Committed by School Staff
 - OFFN-4: Offenses - Number of Incidents Committed by Student
 - DISC-3: Corporal Punishment Indicator: Does the juvenile center use
- Student Discipline (Referrals to Law Enforcement & School-Related Arrests)
 - ARRS-1b: Discipline of Students with Disabilities - School-Related Arrest
 - ARRS-2a: Discipline of Students with Disabilities - Referred to Law Enforcement Agency
 - OFFN-1: Offenses - Number of Incidents
- Program Enrollment (Gifted & Talented, Dual Enrollment) (2)
 - PENR-1. Gifted and Talented Programs Indicator *if the school has these programs*
 - PENR-2. Gifted and Talented Student Enrollment
 - PENR-4. Student Enrollment in Dual Enrollment Programs
- High School Equivalency Exam (1)
 - EXAM-1: Student Participation in the SAT Reasoning Test or ACT

II. Preprocessing: Explain your approach towards known or detected data issues:

- A. Retention is divided for each grade K-12.** This could be an issue as I am trying to calculate the retention rate for black girls in all alternative schools collectively. During preprocessing, I could create a retention rate for each alternative school to resolve the issue or I could create a retention rate for each type of schooling (e.g. elementary, middle, high). If I make retention a categorical variable based on rates per level of schooling, this could also provide more details into which factors push out more black girls at each level of education; however, this would be more labor intensive.
- B. Feature Selection.** There are a ton of feature variables that are divided by demographics that could be influential predictors of retention. Although that is what the model is for, I want to ensure a valid selection of these variables and careful preprocessing before modeling. To approach this, I need to continue exploring the data and conduct feature selection tests (like ANOVA) to determine which features to include.
- C. Accounting for Reserve Codes.** When merging data specific to black female student retention, there are many -9 values, giving me limited retention data to predict. For a strong model, I want to have ample black female student retention data; I could shift my methodology. Predicting the integrity of alternative schooling could be broader, including a combination of retention rates and other variables.

Reserve Code Value	Definition
-3	Skip Logic Failure
-4	Missing Optional Data ⁶
-5	Action Plan/Quick Plans
-6	Force Certified
-8	EDFacts Missing Data
-9	Not Applicable/Skipped
-11	Suppressed Data
-13	Missing DIND skip logic

Methodology: Discuss your analytical plan.

Alternative Schools SCH_STATUS_ALT: is a public elementary or secondary school that addresses the needs of students that typically cannot be met in a regular school program and is designed to meet the needs of students with academic difficulties, students with discipline problems, or both students with academic difficulties and discipline problems. Alternative education schools may be sited in locations other than a traditional school building such as hospitals, mental health centers, jails, or juvenile detention centers.

Justice (JUST) A justice facility is a public or private facility that confines pre-adjudicated/pre-convicted individuals, postadjudicated/post-convicted individuals, or both. Justice facilities include short-term (90 calendar days or less) and long-term (more than 90 calendar days) facilities, such as correctional facilities, detention centers, jails, and prisons. These facilities may confine juveniles (individuals typically under 21 years of age), adults (individuals typically 21 years of age and older), or both. Some states and jurisdictions include individuals younger than age 21 as adults due to statute/legislation and/or justice procedures. For the CRDC, only individuals up to 21 years of age who are confined in justice facilities are of interest.

- While alternative schools serve as a transitional period in one's educational life (some students stay for days, others for years), it still has a significant effect on one's educational attainment and/or educational perspective. The U.S. Accountability Office finds that “certain groups of students attend alternative schools in greater proportions than they do other schools”, specifically black girls and boys.
1. Compile all Justice and Alternative Schools by COMBOKEY
 2. Preliminary density test of how many alternative schools are in each state.
 3. Preliminary density test of how many black students are in each of those schools.
 - Determine which schools in what state I need to focus on.
 - *Because of the lack of data pertaining to the black female population, I will not be doing a regional dissection of the data.*
 4. All feature and target variables are different datasets. Merge .csv files that account for each variable and demographic into one main dataset.
 5. Preprocess the main dataset.
 - The target Retention Variable is the count variable. Need to turn into a binary variable where:
 - **High Retention:** Schools above a certain retention threshold (e.g., >70%).
 - **Low Retention:** Schools below that threshold (e.g., ≤70%).
- This would be a classification problem:
 - Random Forests

- Information-based classifier
- Most necessary in answering *which* factors most influence retention for black female students
- Logistic Regression
 - Can interpret the data by showing *how* certain factors influence retention for black female students
 - However, could be sensitive to multicollinearity (I believe a lot of my features may be correlated)

References

- Anderson, M. D. (2024, August 13). *How Black Girls Are Harmed in Schools*. Education Writers Association. Retrieved February 2, 2025, from <https://ewa.org/data-research-tips/how-black-girls-are-harmed-in-schools>
- Cabral, B., Annamma, S. A., & Morgan, J. (2023). “When You Carry a Lot”: The Forgotten Spaces of Youth Prison Schooling for Incarcerated Disabled Girls of Color. *Teachers College Record*, 125(5), 95-113. <https://doi.org/10.1177/01614681231181816>
- Hartocollis, A. (2024, December 16). *Black Student Enrollment at Harvard Law Drops by More Than Half*. The New York Times. Retrieved February 2, 2025, from <https://www.nytimes.com/2024/12/16/us/harvard-law-black-students-enrollment-decline.html>
- National Center for Education Statistics. (n.d.). *Fast Facts: Degrees conferred by race/ethnicity and sex (72)*. National Center for Education Statistics. Retrieved February 2, 2025, from <https://nces.ed.gov/fastfacts/display.asp?id=72>
- U.S. Department of Education. (n.d.). *Civil Rights Data Collection Office for Civil Rights*. Civil Rights Data | U.S. Retrieved February 2, 2025, from <https://civilrightsdata.ed.gov/>