# Case Study 2: Understanding what Leads to Graduation Retention

## Classification of Assignment

- This is a Category B Assignment - A Group Assignment:
    - Your group may not receive help from anyone outside your group. All questions concerning this assignment should be addressed to your professor. It is an honor code offense to give help to other groups and individuals or receive assistance from other groups and individuals.
    - Groups are assigned by MSBA program.

## Instructions

- A dataset is provided for you labelled Case2.csv. This dataset was retrieved and cleaned from the IPEDS website at https://nces.ed.gov/ipeds/use-the-data. The data provided to you is from the year 2020. Institutions were discarded that had a full-time enrollment under 500 students. You are encouraged to look around IPEDS to get a sense of the data. The variables provided are defined on the last page.

## The main goal of the assignment is to explore relationships affecting the Full-Time Retention Rate.

- In an R script file, load the Case2.csv dataset and using the appropriate functions to inspect the data frame.
- Taking FTRetentionRate as your dependent variable, explore three simple linear relationships with that variable.
- Explain in comments whether you detected a positive relationship, a negative relationship, or no relationship and whether the relationship is considered strong, weak, or not existent.
- Use the techniques you learned in base_graphics to describe both the variables and relationship visually, including a plot of residuals. (you may use ggplot2 instead of base graphics). In doing so, test the assumptions and provide in comments an assessment as to whether each assumption was most likely violated and what that means to the model. Ensure you have the following plus any other visuals you want to include to help describe the relationships:
    - Plot your x and y variable to test for linearity.
    - Provide a histogram of the residuals and describe whether they appear normal or skewed.
    - Conduct a bptest to test for homoskedasticity.
    - Use the plot function on your regression object to look at the remaining assumptions.
- Provide an implication (why do we care) for the relationships you choose, helping to understand what that means for graduation retention.

- In a team RScript file, include the following:
    - A .R file that saves and opens without edits.
    - Code that runs procedurally from top to bottom without error.
    - Your team name along with individual names listed of people that contributed in comments at the start of the script
    - Data loaded properly directly from your working directory (no subfolders) with summary functions helping to understand what all is included in the dataset.
    - 3 simple linear regression models with assessment as to what the statistics in the model refer to.
    - A test of the assumptions for each of the 3 linear regression models and an assessment (in comments) as to whether you think they were violated or not. An assessment about what your findings do to the integrity of the model.
    - Any other visualizations that describe the relationship and the variables along with how those visualizations help you.
    - An implication (in comments) of why we care about the relationship describing what useful information someone can extract from what you found.
- One required R Script file per group should be submitted via blackboard by the due date listed in the system.

Note – splitting the dataset into training and testing groups is not required for this case study.

## Rubric for Paper (80%)

A (100) – This assignment is considered exemplary. The R script file was clean and free from errors and code ran from start to finish. All parts included. Assessments correct and complete.

A- (90) – This assignment is considered well-done. The R script file was free from errors and code ran from start to finish. All parts included. Assessments correct, but missing minor detail.

B (82) - This assignment is considered proficient. The R script file had minor errors preventing it from running without edits. All parts included. Assessments needed clarity and detail.

C (75) - This assignment is considered acceptable. The R script file had errors preventing it from running without edits. Most parts requested were included. Some assessments incorrect or needed clarity and detail.

F (60) – This assignment is considered underwhelming. The R script file was submitted, but major detail was missing and or inaccurate. R file Includes some of the necessary components. Not all assessments correct.

F (0) – No R file submitted.

## Group Contribution and Assessment (20%)

- Everyone in the group must contribute and write a portion of the team paper: introduction, summary and analysis of topic, conclusion, citations.
- Groups will be peer evaluated through teammates. Participation and score through teammates will be incorporated into your final average.

| Variable Name | Description | Definition as listed by IPEDs (Note – Some definitions have been shortened down for brevity). |
|---|---|---|
| unitid | Unit ID of the Institution | |
| institutionName | institution Name | |
| FTRetentionRate | Full-time Students Retention Rate (as a percentage) | The full-time retention rate is the percent of the (fall full-time cohort from the prior year minus exclusions from the fall full-time cohort) that re-enrolled at the institution as either full- or part-time in the current year |
| PTRetentionRate | Part-time Students Retention Rate (as a percentage) | The part-time retention rate is the percent of the (fall part-time cohort from the prior year minus exclusions from the fall part-time cohort) that re-enrolled at the institution as either full- or part-time in the current year |
| TotalEnrollment | Total Enrollment | Total men and women enrolled for credit in the fall of the academic year. CREDIT - Recognition of attendance or performance in an instructional activity (course or program) that can be applied by a recipient toward the requirements for a degree, diploma, certificate, or other formal award. NOTE: Enrollment reported is of the institution's official fall reporting date or October 15. |
| FTEnrollment | Full-time Enrollment | Total men and women enrolled for credit full time in the fall of the academic year. FULL-TIME STUDENT - Undergraduate—A student enrolled for 12 or more semester credits , or 12 or more quarter credits, or 24 or more clock hours a week each term. Graduate—A student enrolled for 9 or more semester credits, or 9 or more quarter credits, or a student involved in thesis or dissertation preparation that is considered full time by the institution. First-professional—As defined by the institution. CREDIT - Recognition of attendance or performance in an instructional activity (course or program) that can be applied by a recipient toward the requirements for a degree, diploma, certificate, or other formal award. NOTE: Enrollment reported is of the institution's official fall reporting date or October 15. |
| PTEnrollment | Part-time Enrollment | Total men and women enrolled for credit part time in the fall of the academic year. PART-TIME STUDENT - Undergraduate—A student enrolled for either 11 semester credits or less, or 11 quarter credits or less, or less than 24 clock hours a week each term. Graduate—A student enrolled for either 8 semester credits or less, or 8 quarter credits or less. |
| S2FRatio | Student-to-Faculty Ratio | Student-to-faculty ratio - Total FTE students not in graduate or professional programs divided by total FTE instructional staff not teaching in graduate or professional programs. Total FTE students is equal to the number of full-time students plus 1/3 the number of part-time students (Fall enrollment component). Graduate or first-professional students enrolled in graduate or professional programs such as medicine, law, veterinary, dentistry, social work, or public health, in which faculty teach virtually only graduate-level students (often referred to as "stand-alone" or "independent" programs) are excluded from both full-time and part-time counts. |
| TuitionAndFees | Tuition And Fees | Published tuition and fees, 2020-21 for academic year reporters only. These data came from the 2020-21 in-district tuition and fee variable(CHG1AY3) on the price of attendance question of the IC component. IF institution does not charge different tuition for in-district students(CHG1AY3) from in-state students(CHG2AY3),the in-district charge would contain the same amount as the in-state (CHG1AY3=CHG2AY3). If the institution does not vary tuition by in-district, in-state, and out-of-state(CHG3AY3), the tuition and fee amount is stored in all 3 variables (CHG1AY3=CHG2AY3=CHG3AY3). Price of attendance for full-time, first-time undergraduate students for the FULL ACADEMIC YEAR: (Tuition and fees are those amounts used by your financial aid office for determining eligibility for student financial assistance) |
| AverageSalary | Average Salary of Professors - All Ranks | Average salary equated to 9 months of full-time non-medical instructional staff - all ranks Derived by multiplying the average weighted monthly salary SAAVMNT for all ranks (ARANK=7) by 9 |
| AllStaff | All Full-time Instructional Staff | Instructional Staff is an occupational category that is comprised of staff who are either: 1) Primarily Instruction or 2) Instruction combined with research and/or public service. The intent of the Instructional Staff category is to include all individuals whose primary occupation includes instruction at the institution. |
| NoTenureSystem | With Faculty Status - No Tenure System | The number of faculty who are not part of the tenure system |
| TenureTrack | With Faculty Status On Tenure-Track | The number of tenure track faculty |
| Tenured | With Faculty Status Tenured | The number of tenured faculty |
| TotalFaculty | With Faculty Status Total | The number of faculty who are either no tenure system, tenure track, or tenured |