Complete Technical Report

By: Sheridan Burns and Michael McGinn

## Abstract

**Introduction**

In any educational institution, tracking student progression through their degree(s) and major courses is paramount to understanding the overall success of students. With this understanding, a method must be created to track student progress and movement. Thus, through this comprehensive analysis of a student data sample from the ISA department, we will provide such a methodology that can be used to track student progress in the present and moving forward. Through this abstract, we will illustrate a brief overview of our data processing and results, as well as what these results mean in context.

**Data Collection and Preprocessing**

Student information was collected from the department records, which were transformed into structured tables. Each table contained inconsistencies, missing values, and duplicates. Thus, we applied several cleaning methods to ensure that data integrity was maintained throughout all tables. Additionally, we processed the data with techniques such as imputation and column standardization to ensure the dataset was cleaned and ready for analysis. Throughout the entire process, we ensured that each step could be followed and recreated to maintain reproducible analysis and results. Preprocessing the data allowed us to create a foundation to analyze the data properly to classify, track, and obtain results accurately.

**Data Consolidation and Analysis**

Our analysis goals were to create a reproducible way to identify student academic standing and accurately predict student graduation. We determined that observing a student’s completed semesters is a classification system that is clear and reproducible. Thus, our classification methodology will surround keeping a running total of the students’ accumulated term codes. Using this system, we identified patterns and discrepancies in student progression that allowed us to gain insight into anomalies such as traditional, early, and late graduates, as well as students who left the school for some time or switched majors. By following this system, we were able to account for these anomalies a provide a comprehensive view of student progression. Meanwhile, highlighting the dynamic movement of students into and out of the ISA department.

**Conclusion**

The collection, preprocessing, consolidation, and analysis resulted in a framework that simplifies the tracking of students within the ISA department. This system can be used to address the current challenges the data contains as well as help track and identify ways to improve the health of the department. Although this method is effective, there are several limitations and further validations needed to fully refine the usability and accuracy of our analyses.

## Background

This year, our client, Dr. Allison Jones-Farmer has taken on a new role as the interim head of the Information Systems and Analytics (ISA) department. Within this new role, there are many metrics and systems that Dr. Farmer would like to improve upon during her term. In particular, she is interested in formulating an efficient and reproducible system that could be created to classify students as either “Freshmen,” “Sophomores,” “Juniors,” or “Seniors.” Through an efficient methodology to determine a student’s academic standing, Dr. Farmer hopes to find a way to accurately track students' progress over time and predict those who will graduate each semester.

To begin establishing a working definition of a student’s academic standing, the MSBA students have been given various samples of student data snapshots that could be used to formulate a process that can be followed and reproduced to classify both current and future ISA students. Through preprocessing and analysis, one goal is to accurately classify the sample students and document the process.

* Research Question One: What student success metric can be tracked or analyzed to provide a functional definition of a student’s academic standing?

After a functional definition of a student’s academic standing has been established, Dr. Farmer is also looking for an accurate way to track a count of students planning to graduate and earn their degrees each semester. Through a thorough analysis of a sample of student records over time, the goal of establishing what students graduate each semester can be determined. Then, this analysis should be easily reproducible to determine student graduation status moving forward.

* Research Question Two: What student success metrics can be used to establish a procedure to accurately count student graduates each semester?

During the two above analysis steps, thorough documentation of the code and relevant rationale will be provided. The final goal, then, is to improve the organization of the given data samples for easier viewing, analysis, and understanding. This goal will be achieved through a merging and cleaning of the data sets to determine what sets would be best to integrate. Then, the process of data exploration and analysis can be efficiently reproduced by outside users.

* Research Question Three: How can the data table elements be cleaned and combined to add clarity to the content of the student success data and allow for our analyses to be easily reproduced?

In conclusion, the overall goal of this project is to track student progress over time for ISA students. In doing so, a cleaned, consolidated, and comprehensive set of data will be produced for new audiences to determine student progress moving forward.

## Data and Methods

**Data Origination and Integrity Analysis**

Over the past year, Dr Farmer’s ISA department and many other academic departments have begun working on a mission to track student success metrics over time. Within the current tracking software, Bannerweb, each student is assigned a unique “Student ID” identifier, which is how the student success data is tracked for each individual student. For each student, many metrics are tracked including what cohort term each student started in, the term codes each student has participated in, student majors, gender, first-generation indicator, and various GPA-based metrics.

A sample of these metrics has been taken for students in the IS and BA majors and split across five original data tables per major. The sampling method is not known, posing a potential integrity issue and limitation throughout our analyses. Moreover, after a brief discussion with the database overseer, it became increasingly evident that even the database administrators are working to fully understand the data organization. Thus, it will be up to the MSBA students to clarify how the data should be best organized and analyzed to allow for greater data understanding moving forward.

**Preprocessing: Phase One**

During the first phase of preprocessing, we wanted to ensure the correct organization of our data sets. First, it became clear that most of our starting files organized much of the student information in a way that the R platform could not easily comprehend. Thus, our first step was to ensure that a Student ID value was assigned to all the student’s relevant information within the tables and that the data sets could be properly organized. This process was applied to the Majors, Grades, and Degrees Awarded tables. Then, the datasets would allow for proper analyses to be conducted.

With our data sets now properly organized, there were 242 rows and 240 unique values in the IS Major Students table, 1756 rows and 237 unique values in the IS Major Student table, 236 unique values in the IS Major Students Grades table, and 237 unique values in the IS Major Students Attributes table. Then, we found 839 rows and 837 unique values in the BA Major Students table, 4628 rows and 834 unique values in the BA Major Students Majors table, and 747 unique values in the BA Major Students Grades table.

By looking through the Major Students tables, it became evident that various metadata rows at the bottom of the datasets should be removed. After identifying and removing these five bottom rows, the number of rows and unique values became equal for both tables. The Major Students tables were then cleaned.

**Preprocessing: Phase Two**

Before conducting our data analyses, we wanted to clean the data first. Thus, we noticed that there were various blank rows in the Majors tables and that there were many blank major-related rows in the Majors tables. We imputed “None” for all missing values to remedy these issues, since our imputations would be made on categorical or classification-based variables. We then removed the blank majors-related columns from the tables to add more organization and clarity to the overall dataset. The columns removed included the Student ID count in the Students tables and the blank Majors columns in the Majors tables.

**Data Analysis Goal One: Functional Academic Standing**

The first goal of our data analysis is to determine a functional academic standing method that can be used to classify students as Freshmen, Sophomores, Juniors, and Seniors. To begin, we noticed that each student in the BA Major Students table had an array of term codes that could ideally align with the number of semesters that the student had taken at Miami. Thus, a count of term codes became the basis of our determination of a student’s academic standing.

With the establishment of our new methodology, we began the process of classifying each student’s academic standing in the Student Majors table based on how many term codes they had accumulated. With this new classification, a count of each classification title could be determined for each semester. To begin this process, we added a “Term\_Order” column to the Majors tables that would illustrate a numeric representation of what semester the student was on. For example, if a student was in their first semester, they had a value of one, and so on.

With this new Term\_Order column, we were able to easily append a new column onto the Majors tables that would give a classification label for each student record. We determined that all students in the Majors tables were students who had completed at least one semester. Thus, we assigned Freshmen as having 1-2 semesters, 3-4 for Sophomores, 5-6 for Juniors, and 7+ for Seniors. While we thought it could add further insights to our analyses if we would add another classification of Super Senior for those who had accumulated a term count above 8, we noted that this classification was not typical in Miami’s past classification labels. Thus, to maintain this precedence and ensure the simplicity of our data, we refrained from adding this level to our analysis.

Finally, we filtered the Majors tables to include only the instances where students had an IS or BA major on file. This would allow us to focus on only the students of interest and when they were part of the departments of interest. This would also allow for proper tracking, insights, and total calculations for each semester. After completing these steps, Dr. Farmer’s student tracking over time issue became evident. It appears the count of students in both departments varies greatly as the cohorts move from semester to semester. This occurrence would need further research on a person-by-person basis to pinpoint the causes and any trends in these causes.

**Data Analysis Goal Two: Graduate Determination**

As students move through their academic careers, their point of graduation is one of the most important milestones in their journey. At the same time, universities must track student progress over time as a test for the university’s comprehensive success. With this idea in mind, we relied on our previous graduate determination method to track the 202010 cohort as they progressed towards their projected 202320 graduation. We then focused on the 202320 Senior class to determine where each student went. To complete this, we first pulled the student IDs of students in our sorted Majors tables that were classified as “Seniors” in the 202320 semester. With these IDs pulled, we could then pull the rows of the sorted Majors table where these student IDs existed. Doing so showed us what semesters and classifications the student spent in the BA or IS majors.

After pulling this information for the BA 202320 Seniors in particular, we were able to build a story around each student’s academic journey. We saw that many were on track to graduate with 8 semesters in their cohort groups, while others took semesters of school off, decided to graduate a few semesters late, or even graduated a semester early. All of these factors played a role in our count of Seniors we project to graduate in the 202320 term and those who we tracked to potentially graduate a few semesters after 202320.

These counts of potential graduates could then be compared to the Degrees Awarded tables. The Degrees Awarded tables illustrate when and with what degrees each student earned their diplomas. Based on the ease of determining if a student has received their diploma or not, we have given the Degrees Awarded table our highest consideration of data integrity. Thus, we used the data in these tables as a basis for evaluating the validity of our classification and graduate counts. In doing so, we first created three classification columns in these tables to indicate if a student was a Traditional Graduate (graduated with 8 semesters), Late Graduate (took more than 8 semesters to graduate), and Early Graduate (graduated with less than 8 semesters). This would provide more insight and granularity to the data in each Degrees Awarded table. We then compared Student IDs in the 202320 senior class with the Student IDs in the Degrees Awarded table for BA students to see if these values coordinated.

**Data Analysis Goal Three: Coherent Merging and Cleaning**

After analyzing the data, our final goal was to organize the data in the most coherent way possible, including merges and cleaning. We started with 10 original files that contained different information for BA and IS students separately. After preprocessing the data, we determined a count of students based on the distinct number of Student IDs and confirmed there were no NA values in this area. Next, we absorbed the Students tables into the Degrees Awarded tables, as the Students tables only contained a list of Student IDs that could be better represented with the information given in the Degrees Awarded tables.

In the end, we were left with 4 files for each major type (Academic Standing, Attributes, Grades, and Majors) and decided not to merge them further for two reasons: many of the files have different semesterly granularity, and each of the files serve a different specific purpose. For example, the Majors file contains one row for each Student ID, while the Grades file contains one row for each Student ID per semester. Therefore, these tables cannot be merged due to conflicting row counts. These tables also should not be merged due to each table containing different information that should be analyzed on an individual table basis to ensure analysis clarity.

After the merging and cleaning were complete, we wanted to add more analysis methods that would allow future data users to gain more utility from our given data sample. Thus, we used information from the Majors table to create a new pair of tables, called Academic Standing for each major group. The goal of this table creation was to provide a table where one could pull a comprehensive view of a student's most recent academic standing. Thus, these tables list Student IDs, the most recent number of semesters each student has accumulated, as well as the most recent semester that the student was a part of the major, and their coordinating academic standing label. In creating these tables, we also functionally absorbed the Students tables, as these tables only included a distinct list of Student IDs that was now presented by the Academic Standing tables in a more informative format. With the new Academic Standing tables, we hope to allow users the ability to apply a quick query for different Student ID values to see a student’s most recent academic record. Thus, the user would not need to observe a student’s DAR to visualize their most recent academic progress.

## Results

The first concern of our analysis was to determine a clear and reproducible way to classify students according to their academic standing as Freshmen, Sophomores, Juniors, or Seniors. This method produced counts similar to Dr. Farmer’s for each semester. Although, they were slightly different. We hypothesize that this is because the methods previously used to classify students included consideration of a student’s completed academic hours. We did not choose to follow this method, as it can often result in incorrect classifications if a student takes more hours before, during, or after a semester of interest. Furthermore, it may be more tenuous and challenging to observe and track a student’s progress in each department if the student chooses to take some hours in various departments over time.

We also chose to determine a student’s classification by a count of their semesters completed, as it is a quick method to reproduce and can allow for the journey of students to be continually observed in the Majors tables. Although, it is important to note that every student has their own academic journey that can impact their classification. Some may retake semesters, leave the university, and return, drop out of school entirely, or transfer into the university with many semesters already completed. These occurrences all impact a student’s classification level, so it is important to keep in mind that tracking student progression with a singular classification method may never predict student success 100 percent of the time.

The second concern of our analyses was to add more clarity as to why there was a higher or lower count of students as they moved toward the Junior or Senior year classifications. Because we relied on tracking students as they accumulated term codes over time, we were able to observe in what departments these term codes were required and when. Thus, we could paint a clearer picture of where students tend to migrate over time. After observing the student movement of the 20210 BA cohort students over time, we found that only 5 students who started as BA majors remained BA majors until the 202320 semester. We also found that 116 students during this four year period transferred into or out of the department and student cohort. Thus, it is increasingly evident that each student’s academic journey is extremely varying and needs to be observed on a person-by-person basis.

After tracking student classification counts each semester, we proposed that there are 66 Seniors in the 202320 class. After observing each student’s academic journey into and out of the ISA department in the BA Majors table, we propose that 59 students are expected to graduate in the 202320 cohort, and 7 of the 66 students will graduate in the following semesters. This was evident by these seven students having academic records after the 202320 cohort and a count of term codes over the typical 8-semester threshold. However, the Majors table only illustrated to us when these students’ academic journeys ended. We would need to then cross-reference these student counts with those students in the BA Degrees Awarded table to see what students truly graduated. After doing so, we found that 58 of our 59 projected traditional students received their diplomas, and all 7 of the extra semester students received their diplomas. This suggests that only one student in the 66 Senior count never received their diplomas. Thus, providing a graduation rate for this 202010 cohort of 98.48%.

## Discussion and Conclusion

After finding our solutions above, we want to contextualize our findings and discuss the next steps for our findings. First, how does our case study graduation rate compare to the graduation rate of outside departments or universities? According to the Miami University website, the average graduation rate of the most recently reported year, 2020, was 89.2% for all university students (“Graduation and Retention Rates,” n.d.). Using the particular cohort we followed, it appears the BA department is doing quite well in student graduation rates compared to the university at large. But, compared to the national average graduation rate for public university students in 2023 of 63%, it seems Miami University and the ISA departments are highly successful in supporting students through graduation (NCES, 2022).

In total, our findings illustrate how frequently students can move in and out of the university, as well as in and out of various academic departments. In our case study analysis on the 202010 cohort, only 5 of the original BA students in this cohort were found to be seniors in 202320. At the same time, we found that the total number of BA students increased throughout our sample data. Thus, although the BA major appears to be growing in popularity with Miami students, the ISA department may want to take additional measures to improve the retention rate of students over time. According to the national average for public universities, ⅓ of all students change majors over their academic career, and 10% of students change majors more than once (NCES, 2017). These metrics illustrate a much lower major change rate than was seen with the BA major. Therefore, more analysis and possible polling of Miami University students should be conducted to add more understanding as to why so many choose to transfer into and out of the BA major and ISA departments.

Overall, it seems our cohort term counting methodology does a fair job of predicting if and when students will graduate. However, given the limited scope of our data and understanding of the data sourcing, more analyses are needed to fully validate our methods and findings. One overarching trend we found was that many students move into and out of the department, which will need further investigation and efforts to reduce those leaving the BA major in particular. At the same time, the variability of a student’s academic journey also became evident through our analysis of the data. Thus, we propose that with all methods used to determine student classification to predict graduation, all findings must be taken with a grain of salt.

Moving forward, we believe having more information on the data from the data experts could allow us to perform even more accurate analyses to fit Dr. Farmer's and the ISA department’s needs. At the same time, given our experience this these particular data sets, we recommend an alteration of the methods used to gather the student data at its origin. From an outside perspective, the metadata and labels used in this data collection were often incomprehensible by an untrained eye. Thus, if our analyses were to be recreated in the future, having the data collection already organized and labeled for outside viewers to understand could greatly improve analysis efficiency.

## Limitations

There are several limitations to this analysis ranging from the data itself to insights. The first issue encountered was the way that the data was gathered in the first place. The data were collected through Bannerweb, which is a digital platform that contains records for student and faculty information. However, Bannerweb only gives access to certain reports, which is not an all-encompassing way to convey the full picture of student progress over time. Furthermore, the data used in our analyses is only a sample of student metrics over a specific period. Thus, we were limited in our ability to gain a full scope of the student data over time to enrich our analyses.

The second main limitation we ran into was that the data contained many granularity and classification issues that made it difficult to paint a comprehensive picture of each student’s academic journey. For example, in our analysis, we mainly focused on tracking students with a BA or IS major/co-major. Thus, their degrees were indicated as Business Analytics or Information Systems in the Majors tables. However, when we pulled these students from the Degrees Awarded tables, the students ended up graduating with a wide variety of majors that did not correlate between tables. Thus, there appears to be a naming inconsistency that the ISA department should further investigate moving forward to ensure accurate sourcing and counting of student information can be achieved.

The third major limitation we had was that we had a strong lack of background information on our given data sample. A large source of this limitation was that the labeling of data was minimal and often incomprehensible for an outside audience. At the same time, after speaking with the data stewards and scientists, it seemed that there was more background information we were not informed of that could have assisted us in our analyses. For example, Skip informed us after completing our analyses that a student’s cohort term could change over time. This occurrence had not been evident by the datasets we were given, so we think there should be an increase in data communication moving forward for all parties that rely on and use the given data.

## 

## 

## 

## 

## 

## 

## 

## 

## 

## Sources Cited

“Graduation and Retention Rates.” *Miami University Graduation Rates*, Miami University,

nces.ed.gov/pubs2018/2018434/index.asp. Accessed 6 Dec. 2024.

National Center for Education Statistics (NCES). (2017). Percentage of 2011–12 First Time

Postsecondary Students Who Had Ever Declared a Major in an Associate's or Bachelor's Degree Program Within 3 Years of Enrollment, by Type of Degree Program and Control of First Institution: 2014. Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/Datalab/TablesLibrary/TableDetails/11764.

National Center for Education Statistics (NCES). (2022). Undergraduate Retention and

Graduation Rates. Condition of Education. U.S. Department of Education, Institute of Education Sciences. Retrieved May 31, 2022, from https://nces.ed.gov/programs/coe/indicator/ctr.