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 over time. Thus, through this comprehensive analysis of student progress data, 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 provide a brief overview of our data processing and results, as well as what these results mean in context.

**Data Collection and Preprocessing**

**Data Consolidation and Analysis**

**Conclusion**

## 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. With 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.

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 the 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 integrating student success data into a new software called Workday. Within Workday, 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, there are many metrics that 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 departments and split across five original data tables per department. 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 new Workday data organization. Thus, it will be up to the MSBA analysts to provide clarity into 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 Major Students, Major Students Grades, Major Students Majors, and Major Students Attributes tables for each department. Then, the datasets can be analyzed to determine what columns or rows are not necessary.

**Preprocessing: Phase Two**

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 there were various metadata rows at the bottom of the datasets that 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 are then cleaned.

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

The first goal of our data analysis is to determine a method by which to determine the functional academic standing method that can be used to classify students as Freshmen, Sophomores, Juniors, and Seniors. To begin, we first noticed that each student in the IS 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. After doing so, Dr. Farmer’s student tracking over time issue became evident. The count of students in both departments seemed to dwindle as they approached their junior and senior years. The next phase of our analysis will discuss one main source of this occurrence.

**Data Analysis Goal Two: Graduate Determination**

The second goal of our analysis was to establish an accurate way to determine the count of students graduating or not graduating each year. For this analysis step, we brought in a new data table, Major Students Degrees Awarded, that showed what students graduated with what degrees and in what term code. This table offers very accurate results of students who graduated each term. Thus, if a student in our academic standing table also has a record in the Major Students Degrees Awarded, it is indisputable that the student has graduated with their degree. With this understanding, we could begin working backward to compare the two tables to justify our student graduates each semester.

To perform this analysis, we created three new indicator columns in our academic standing tables that would show if the student was a traditional graduate, took extra semesters to graduate, or graduated early. Traditional graduates have a record in the Major Students Degrees Awarded table and have a count of 8 term codes in the Academic Standing table. Extra semesters students have a record in the Degrees Awarded table, but have a count of term codes greater than 8. Lastly, if a student has a record in the Degrees Awarded table and has a count of term codes less than 8, they would be considered an early graduate.

Finally, we added a column to the Academic Standing table that pulls in a student’s max term code. This will be the semester that the student graduated in and will allow us to track the number of graduates each semester. Then, we were able to compare a count of students who received their degrees each semester to a count of how many students we fit into our three classifications. Doing so will allow us to determine if our graduate classification method would result in the true number of graduates that had received their diplomas each semester.

* Detail preprocessing steps
* Describe methods used to determine the outcome
* Justifies the credibility of the methods and data used

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

* Each table has a specific purpose and should not be merged
  + Makes it more coherent, for example if you're looking for grades of a given student you would look in BA\_Grades and not BA\_Academic\_Standing
* Many tables also have different row counts and could not be merged regardless
* BA\_Students/IS\_Students have been absolved
* Leaves us with 4 tables for each BA and IS, Academic Standing, Attributes, Grades, Majors
* Each table provides a specific insight for the department to be able to analyze
* As for cleaning, the only table that needed attention was the Majors table
* we imputed NA values with None in the Majors table
* Removed Major4, Major5, Major6 as every student had NA values for this
* Changed the names of the tables to a more concise format, for example BA Major Students - Major becomes BA\_Majors

## 

## Results

* Neutral description of results
* Unbiased justification based on answers to the research questions
* Solid justification for qualitative or quantitative results
* Do not illustrate implications

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. According to our table of student classifications each semester, it can be seen that this methodology provides a reasonable way to track students over time (Appendix 1). This is indicated by a majority of the student counts by classification being consistent over time. However, this method is not perfect, as students often change departments or take a non-traditional amount of time to graduate. Thus, as students move toward the junior or senior classification, the tracking approach of these students has to become more granular.

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. For this analysis, we looked at the Degrees Awarded tables to see what and how many students graduated each term code. In doing so, we assumed that these tables had the most correct information out of all of our tables, as it is quite simple to determine if a student has received their degree or not. Thus, we used our findings from these tables in combination with a count of term codes per student to determine what students received their degrees after completing 8 semesters, after completing more than 8 semesters, or after completing less than 8 semesters.

We then created a matrix table that showed how many students graduated on time, early, or late (Appendix 2). In doing so, we identified that many students were graduating early or taking a few more semesters before being awarded their degrees. Thus, if the count of student classifications each semester does not logically add up, it may be caused by some students taking a non-traditional graduation path. Thus, further proving that providing students with a typical freshman, sophomore, junior, or senior classification may not predict when the student will graduate. Furthermore, searching for a waterproof methodology to predict or track student success over time with a blanket approach may not always produce assumable results, as a student’s academic journey can be inherently unpredictable.

## Discussion and Conclusion

* Insights and interpretation of the analysis
* Comparison to other schools (graduation rates?)
* Use external resources to support the conclusion
* Next steps and implications

After seeing how a student’s academic journey can be highly unpredictable, we wanted to see if the unpredictability of our sample data was reflected in national trends. That is, are most American college students as unpredictable in their academic journey as those in our analysis?

## Limitations

* Clearly articulates limitations to the analysis

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 were not an all-encompassing way to convey the full picture. Furthermore, the data used in our analyses is only a sample of student metrics over time.

* Not possible to confirm through the data whether or not students had transferred out of the department. Students who had a major in IS or BA at any point in their time at Miami were included in the dataset for all 8+ semesters.
* Another limitation is that we aren’t able to determine which students are MSBA students as we don’t have access to graduate student information. This will impact the graduation count of students.