Spring 4-26-2021

An Exploratory Analysis of the BGSU Learning Commons Student Usage Data

Emily Eskuri
eeskuri@bgsu.edu

An Exploratory Analysis of the BGSU Learning Commons Student Usage Data

Emily Eskuri

Honors Project

Submitted to the Honors College at Bowling Green State University in partial fulfillment of the requirements for graduation with UNIVERSITY HONORS

April 26, 2021

Dr. Umar Islambekov, Department of Mathematics and Statistics, Advisor

Dr. Shuchismita Sarkar, Department of Applied Statistics and Operations Research, Advisor

Dr. Jessica Turos, Office of Academic Assessment, Advisor

Table of Contents

Executive Summary

The purpose of this study was to explore past student usage data in individualized tutoring sessions from the Learning Commons from two academic years. The Bowling Green State University (BGSU) Learning Commons is a learning assistance center that offers various services, such as individualized tutoring, math assistance, writing assistance, study hours, and academic coaching. There have been limited research studies into how big data and analytics can have an impact in higher education, especially research utilizing predictive analytics.

This project applied analytics to individualized tutoring data in the Learning Commons to create a better understanding of why those trends happen in higher education. The visit analysis determined how often students returned to the Learning Commons for individualized tutoring after one visit and when students visited during the semester. The decision tree and random forest model analysis predicted when students would be present at a given time. The random forest analysis also determined what the most important predictor was for the model. The course analysis viewed what type of courses students sought out for individualized tutoring. Students utilized the Learning Commons for individualized tutoring in ten common course types. Finally, the duration of the tutoring session also was analyzed to determine the average duration during a given week in the semester and the day in the week. The information that came from this Honors Project could be useful to the Learning Commons for planning ahead with scheduling tutoring sessions, identifying popular visit times, and planning for student visits during the semester.

Introduction

Most universities and colleges offer peer tutoring services on their campus. These services are typically housed in tutoring centers or academic support labs at the institution (Cooper, 2010). The Bowling Green State University (BGSU) Learning Commons is an environment for students to be supported in their academics. The Learning Commons is located on the first floor of the Jerome Library at BGSU. They offer various services, such as individualized tutoring, math assistance, writing assistance, study hours, and academic coaching. Individualized tutoring is where a student at the University tutors another student to support the student seeking academic assistance (Roscoe & Chi, 2007). These individualized sessions are tracked and recorded by the Learning Commons to gather usage records about the student and their tutoring session. This information is collected when the student signs into the tutoring session and swipes their student ID for checking into the Learning Commons and checking out.

The purpose of this study was to explore past student usage data in individualized tutoring sessions from the Learning Commons from two academic years. Since this project was an exploratory analysis, the preliminary research questions from the original honors proposal were subject to change based on the data that I received from the Learning Commons. One preliminary question regarding time spent and frequency of visits affecting academic performance could not be answered with the given dataset. This research aims to answer the following questions upon viewing the dataset:

- How often do students return to the Learning Commons for individualized tutoring after visiting once?
- When do students typically visit the Learning Commons during the semester for individualized tutoring?

- Can we predict when students will utilize the Learning Commons for individualized tutoring?
- Which courses do students seek for individualized tutoring at the Learning Commons?
- How long do students spend in individualized tutoring sessions during the semester?

The results from this analysis can potentially help the Learning Commons understand when BGSU students seek out tutoring in the semester and be prepared for an increase in tutoring assistance. Furthermore, this project provided insights about the types of courses in which students utilized individualized tutoring. This could help the Learning Commons by assigning more tutors to help students with those courses. These results can give the Learning Commons leadership an overall snapshot of the individualized tutoring sessions over two academic years.

Review of Literature

Peer tutoring is the act of one student at a university tutoring another student an effort to support the learning of the student seeking assistance (Roscoe & Chi, 2007). The student tutor typically understands the academic subject well and has sufficient knowledge to assume the tutor position; however, there is usually not a large disparity in knowledge between either student (Muise, 2016). There are many studies that have provided evidence of the impact that tutoring has on the tutor and tutee (the student being tutored) to improve their learning and understanding of the topic. (Cohen, Kulik, & Kulik, 1982). However, tutoring centers are faced with an overwhelming amount of data in their databases, and not many have had the opportunity to run an analysis to find trends in the data. There have been limited research studies into how big data and analytics can have an impact in higher education, despite there being more interest in exploring the value of the increasing data within higher education (Daniel, 2015).

Big data and analytics could be applied to peer tutoring and tutoring center data to create a better understanding of why those trends happen in higher education. One study has shown that peer tutoring had a positive effect on students' grade point average, performance rate, and success rate (Arco-Tirado, Fernández-Martín, & Hervás-Torres, 2020). Conducting an exploratory data analysis is beneficial because there may be variables or effects that impact the student's experience and if it was positive or negative (Cohen, Kulik, & Kulik, 1982). Conducting a predictive analysis could contribute to all tutoring centers and make scheduling tutors more efficient (Brattin, Sexton, Yin, et al., 2020). One predictive analysis method is a decision tree, which is useful when numeric and categorical predictors are being used. Decision tree models are highly interpretable and are easy to construct using statistical programming. Another useful analysis is a random forest model. Random forest models are made up of many small decision trees that produce predictions, and then combines the estimates and averages the result (Hastie, Tibshirani & Friedman, 2017). Random forest results are beneficial because they can be compared with the decision tree model results. Overall, this study is important because it highlights the value that peer tutoring brings to college-level students at Bowling Green State University.

Methodology

Accessing Dataset

The BGSU Learning Commons has collected data for each student who visited for various services, such as Academic Coaching, Individualized Tutoring, Math Assistance, Study Hours, Writing Assistance, and more. To receive access to the data, I submitted an Institutional Review Board (IRB) Exempt Application because the BGSU student ID numbers would be part of the data. After gaining approval from the IRB, the Learning Commons was able to share their data with me. The Learning Commons shared multiple Student Information and Check-In datasets from the 2016-2017 and 2017-2018 BGSU academic school years. The original plan also included data from the 2018-2019 academic school year; however, there were incomplete data with many Student Information values missing. For the two academic years, there were four Excel workbooks containing Check-In information for each semester, and two Excel workbooks containing Student Information for each academic year.

Cleaning Datasets

Joining the Data

After receiving the Excel workbooks, I created two new columns in each Check-In information sheet, which were named Semester and Academic Year. I then filled the column with their respective information. I did this so I knew where the data were coming from after I had my complete dataset. Then I combined the Fall 2016 and Spring 2017 Check-In information into one sheet by copying all the data from the Spring 2017 semester sheet and pasting it into the Fall 2016 semester sheet. I renamed the sheet to avoid overwriting any data. I used the Microsoft Access application to import the new Excel sheet containing the Fall 2016 and Spring 2017 Check-In data and the Student Information data for the entire 2016-2017 academic year. I did not want a primary key column added to either table, so Access did not include one when the tables finished their import. The primary key for these tables were the BGSU student ID number, so a new column was not necessary. After both tables were imported, I created a new Query Design to join the tables together. I added the tables to the center of the query, and then I linked the Student ID columns to create a primary key connection between the tables. Next, I selected all columns, so everything was included when the data were joined. Finally, I clicked ‘Run’ to display the new table with the data joined together. I exported the table and ensured that the data were joined correctly. I repeated these steps with the Fall 2017 and Spring 2018 Check-In information and the Student Information data for the 2017-2018 academic year. After I completed these steps for the next academic year, I copied all the data from the combined 2017-2018 Learning Commons sheet and pasted it into the 2016-2017 Learning Commons sheet. There were a total of 66,560 entries for the Learning Commons dataset.

De-identifying Student IDs

After the Learning Commons data were combined, I copied the Student ID column and pasted it into a new column at the end of the sheet columns. I removed the duplicates to reveal 5,488 unique Student ID numbers. I created a new column where I created an ID labeled "BGSU00001" and filled the series next to the unique IDs. I created another new column next to the original Student IDs at the beginning of the sheet and called it New BGSU ID. Then I used a VLOOKUP formula to match the unique Student IDs in the original dataset with the duplicates included. I filled the series next to the original Student ID column, and then deleted that column once I confirmed that the New BGSU ID column was correct. I renamed New BGSU ID back to Student ID.

Data Preparation

There were six duplicate columns in the Learning Commons data after the Check-In information and Student Information were joined. These columns were Categories, Tags, Classification, Major, Cumulative GPA, and Assigned Staff. I removed the duplicates of those columns. Since the Learning Commons provides a variety of services to BGSU students, I filtered the Services column to show the entries with individual tutoring only and removed the rest of the rows with other services. There were 26,953 individual tutoring entries in the dataset.

I began to clean individual columns in the dataset. I used the Text to Columns data tool to split information into two columns by using a delimiter in the data. This tool inserted the data to the column to the right, so I made sure to insert a blank column where the data would move so other columns were not being replaced. Since the Classification column included the class standing of the student and their expected graduation year, I used the data tool to separate the class standing from the expected graduation year. The expected graduation year data were separated and placed in a new column to the right of the original Classification column. I labeled the Classification column as Class Standing and the new column as Expected Graduation year. Upon exploration of the data, I found that those columns had inaccurate data. The data in the Classification column represented the known class standing of the student and the last time in a semester that the student visited the Learning Commons. The column for Cumulative GPA also included the last time in a semester that the student visited, so I used the data tool again to separate Cumulative GPA and the semester into two columns. I removed the duplicate semester status column since that column was created once from the Classification column. I did not rename the Cumulative GPA column. After exploring this column, these data were found to be overwritten, so there was one GPA value for the student across the four semesters, so the change in GPA between each semester was not known.

I wanted to view the Course Numbers by Type and Course Code Level, so I duplicated the Course Number column and placed it next to the original. I used the Text to Columns data tool for a third time to separate the Course Type from the Course Code. I created a new column to the right of the Course Code column and grouped the Course Codes by the thousands (e.g., 1000, 2000, 3000, 4000, 5000, 6000, and 7000) to view the code levels. I labeled the column Course Code by Thousands.

I was interested in viewing when the Learning Commons was busy each week during the semester, so I downloaded the BGSU Academic Calendar for both academic years. I created a new column to the left of the Check-In Date column and labeled it Semester Week. Using the academic calendars and normal year calendar, I filtered the Check-In Date column by selecting each day in that semester week and writing the semester week number in the Semester week column. The semester week started on Monday and ended on the following Sunday. The Learning Commons is not open on Saturday, so there were six days included in each semester week. Spring break was counted as semester week 9 in the spring semesters. Furthermore, I wanted to know what time of the day students were visiting the Learning Commons. I inserted a new column to the right of the Check-In Time column and created an if statement to group the time by morning (7:00 AM – 11:59 AM), afternoon (12:00 PM – 4:59 PM), and night (5:00 PM – 10:00 PM). I filled the series of the data, then labeled the column Time of Day.

While viewing the results for the Duration in Minutes column, I noticed that there were extremely large outliers in the data. These could have been from students who swiped into the Learning Commons for individual tutoring and did not remember to swipe out. I removed 213 rows from the dataset because the duration of the visit was longer than 1,000 minutes. The cutoff time followed the hours of operation for the Learning Commons.

Since I wanted to view specific factors for the decision tree and random forest models, I copied the following columns to a new Excel sheet: Check-In Time, Check-In Date, Semester Week, Semester, and Academic Year. I inserted a new column to the right of Check-In Time and labeled the column Hour In. I changed the time format for Check-In Time to HH:MM:SS (a 24-hour digital clock format leading with zero for hours and also displays minutes and seconds). I typed in a formula in the Hour In column to multiply the time by 24, and then I filled the series to the end. While the column was highlighted, I right clicked on the column and selected the Format Cells option. I selected Number from the Category list and specified 5 decimal places. The Hour In column shows the hour from midnight and the decimal value. The Check-In Time column was removed. I inserted another new column to the right of Check-In Date and labeled it Day. I used the text formula to get the first three letters of the day of the week. The Check-In Date column was removed. I created another column to the right of the Hour In column and labeled it Currently Present. For each student visiting the Learning Commons at a particular time (i.e., T), I counted the number visits with Check-In time before T and Check-Out time after T for that day. This column recorded the number of people present at the time a student visited the Learning Commons.

Data Analysis

I used RStudio for the data analysis. I used the ggplot package to create bar charts, histograms, and boxplots of the data. A bar chart was chosen because it allowed me to visually highlight the interaction between a numeric and categorical variable. Histograms are similar to bar charts, but histograms group numbers into ranges to view distributions, while bar charts compare values. Box plots visually display the five-number summary of the values in a data column. The ggplot package in RStudio allowed me group additional values in the charts by color to visualize the data by semesters, time of day, course type, and course code.

I used the tree package to create the decision tree models and used the caret package to create the random forest models. Once I loaded the data into RStudio, I split the data into training and test datasets. The training dataset was the 2016-2017 Academic Year information, and the test dataset was the 2017-2018 Academic Year information. Then I coded and ran the decision tree and random forest models.

Results

Demographics

Out of the 3,524 students who utilized the Learning Commons for individualized tutoring, 56.30% (n=1,984) of students identified as female and 43.70% (n=1,540) identified as male. Roughly 94.78% (n=3,340) of students were undergraduates, 5.05% (n=178) were graduate students, and 0.17% (n=6) of students did not have their degree type specified. The average term GPA for undergraduate students was 3.06, and the average term GPA for graduate students was 3.43. Graduate students had a higher Term GPA than undergraduate students because of the academic requirements in graduate programs.

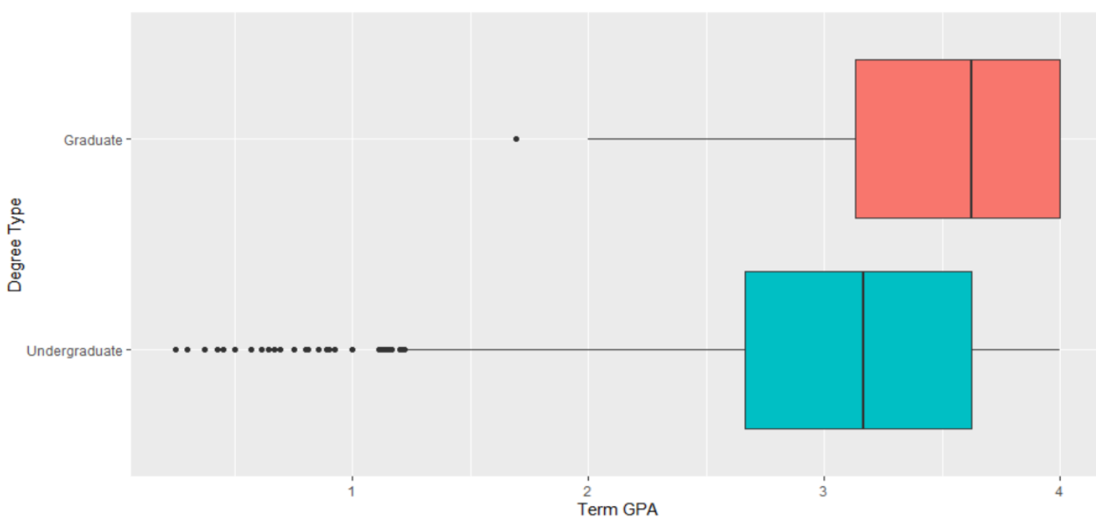

Figure 1: This box plot displays the Term GPA (x-axis) and the Degree Type of students (y-axis) visiting the Learning Commons for individualized tutoring.

Learning Commons Visit Analysis

One area of interest was viewing the number of students who had returned to the Learning Commons after visiting once. About 68.42% (n=2,411) of students who utilized the Learning Commons between Fall 2016 and Spring 2018 had visited more than once. This meant that the majority of students had returned for an individualized tutoring session between those four semesters.

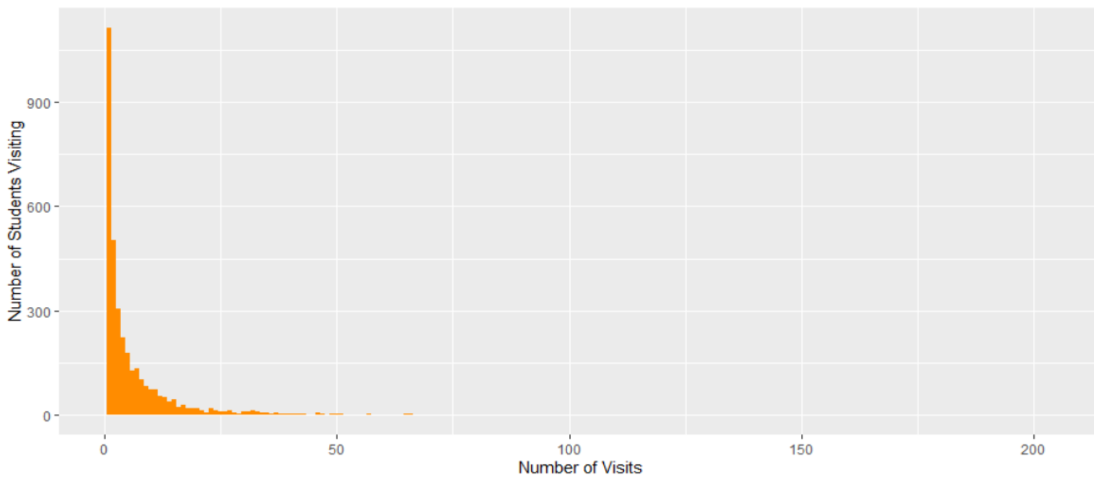

Figure 2: This bar chart displays Number of Visits (x-axis) and Number of Students Visiting the Learning Commons (y-axis).

Since the overall visit data was extremely right-skewed, I wanted to take a closer look at the number of visits between 1 and 10. I found that 80.79% (n=2,847) of students visited the Learning Commons ten times or less between the Fall 2016 and Spring 2018 semesters.

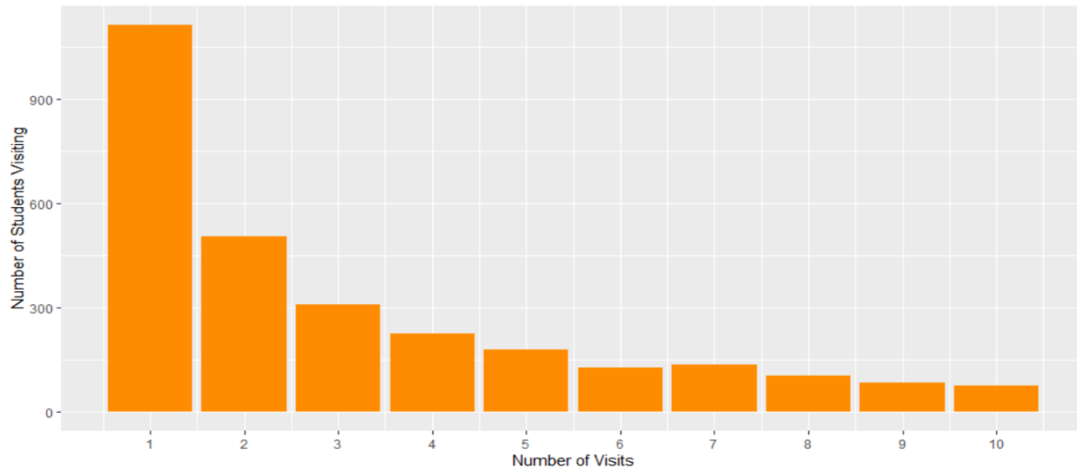

Figure 3: This bar chart displays the Number of Visits between 1 and 10 (x-axis) and Number of Students Visiting the Learning Commons (y-axis).

Since students can visit the Learning Commons for individualized tutoring at any point in the semester, it seemed appropriate to find the count of visits to the Learning Commons each week in the semester. There were 26,828 total visits to the Learning Commons for individual tutoring for the 2016-2017 and 2017-2018 academic years. Both academic years followed 17-week semester schedules, including finals week. The Spring semester schedule included a week off for Spring Break which was week 9. Students visited the Learning Commons more in the Fall semester than the Spring semester, with 57.3% (n=15,364) of the total visits occurring in the Fall semesters. When plotting the chronological order of weeks in the semester, it was evident that students were visiting the Learning Commons more when midterms and finals are typically given. There were peaks in visitation between weeks 5 and 6, then again in weeks 10 and 11 for midterms. Week 16 had the most individual tutoring visits in the semester, which is the week before Final Exams.

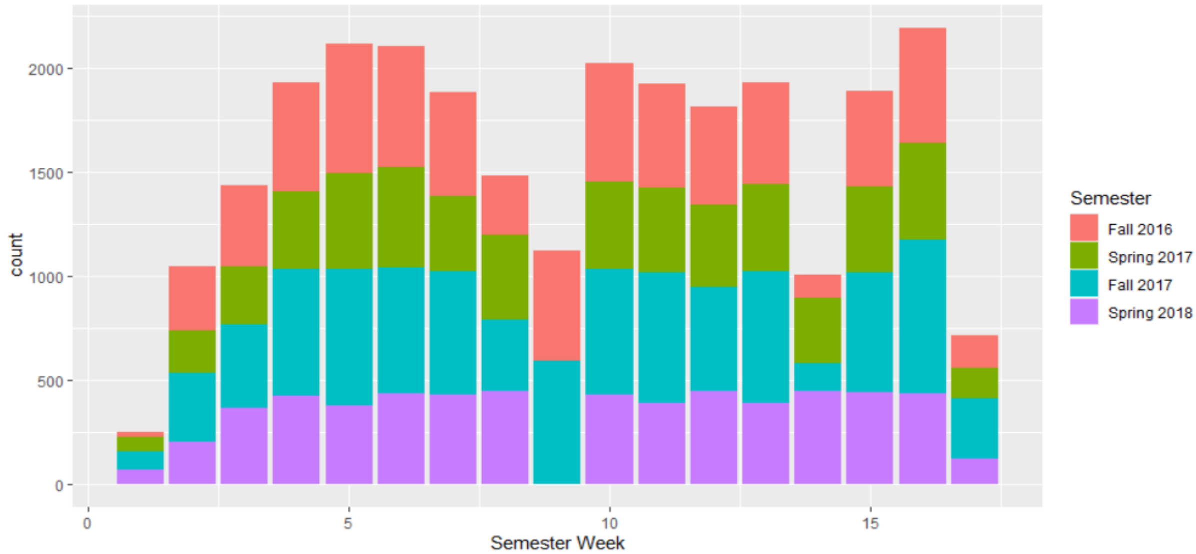

Figure 4: This bar chart displays the week in the semester (x-axis) and the count of student visits to the Learning Commons (y-axis). The bar chart colors are grouped by semester.

Another important analysis was viewing which days in the week and time of day students visited the Learning Commons for individualized tutoring. The busiest days for the Learning Commons were between Monday and Thursday, with 88.77% (n=23,815) of the total individualized tutoring visits. The least busy days were Friday and Sunday; however, the Learning Commons have shorter hours of operation for those two days, closing earlier in the day on Friday and opening later on Sunday. The Learning Commons is closed on Saturday, so there were no visits on that day. This shows that students typically utilize the individualized tutoring services during the week rather than on the weekend. The time of the day also had an impact on student visits for individualized tutoring. Of the total number of individualized tutoring visits to the Learning Commons, 17.63% (n=4,732) of visits were in the morning between 7:00 AM and 11:59 AM, 50.93% (n=13,666) of visits occurred in the afternoon between 12:00 PM and 4:59 PM, and 31.42% (n=8,430) of visits were at night between 5:00 PM and 10:00 PM. About half of the individualized tutoring visits began in the afternoon, which suggests that students may prefer to be tutored in the middle of the day.

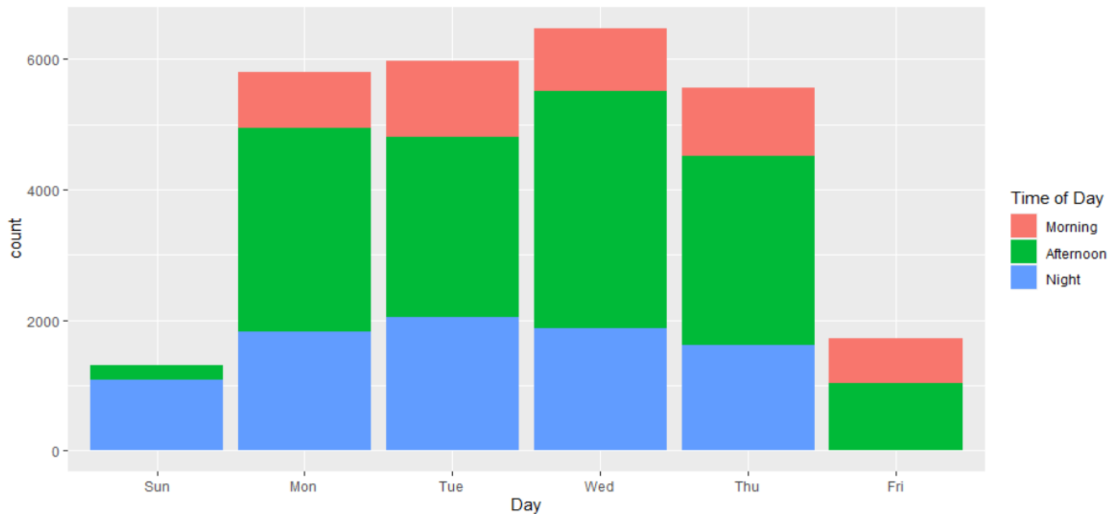

Figure 5: This bar chart displays the day of the week (x-axis) and the count of student visits to the Learning Commons (y-axis). The bar chart colors are grouped by time of the day.

Since students tend to utilize tutoring at different times in the day, it was interesting to determine when the Learning Commons was the busiest in the day. Using the Check-In and Check-Out time data, 5-minute time intervals can be plotted to determine if a student was actively in a tutoring session. This helped determine the peaks of time that the Learning Commons was busiest during the day. There was a steady increase in student traffic through the morning and afternoon until about 4:00 PM when the visits were at its peak. Then there was a decline and dip between 5:00 PM and 6:00 PM, which may be from dinnertime. Then the number of student visits increased again for the evening, until it began to rapidly decrease again.

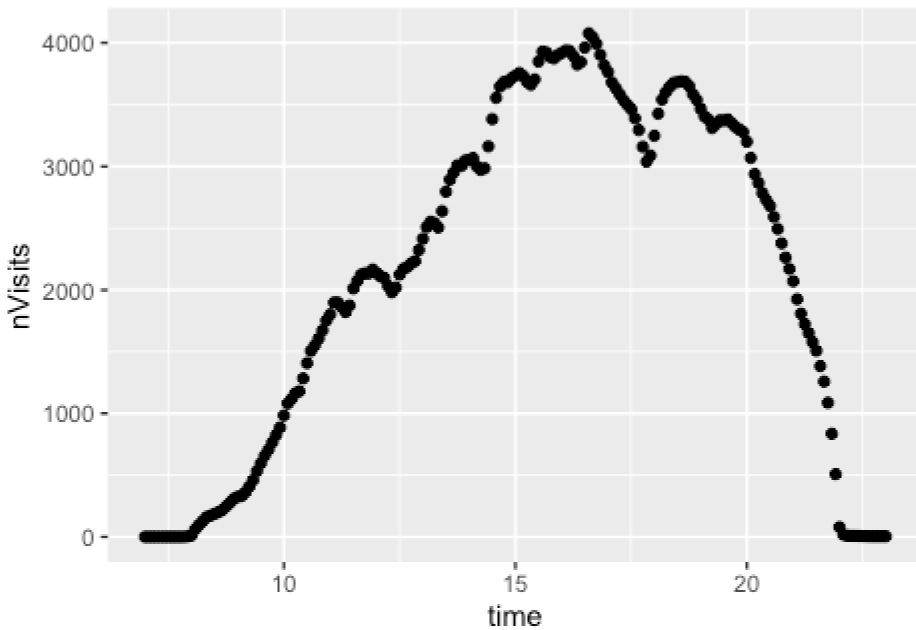

Figure 6: This plot displays the number of students (y-axis) visiting the Learning Commons at a given time in a day (x-axis).

Prediction Analysis for Students Present at a Given Time

Another method of analysis was using predictive models to determine the number of people present at the time a student visits the Learning Commons. The response of Currently Present had the following predictors in the model: Hour In (in hours), Day of Visit (Mon-Sun), Semester, and Semester Week (1-16).

After the data were read in and split into test and training datasets, I ran the decision tree model with the training dataset. The root node was the Hour In predictor, which split when the time was before 13.2917 hours from midnight, and when time was after 13.2917 hours from midnight, which is around 1:17 PM. For hours prior to 10.575, or about 10:34 AM, the predicted number of students currently present at the Learning Commons is 4.113, which is about 4 students. Between 10:34 AM and 1:17 PM, the predicted number of students currently present is 8.883, which is almost 9 students. For time after 13.2917 hours, another decision node was made that views the data by the semester. The left split in the node was for the Spring semester data, and the right split was for the Fall semester data. The largest predicted number of students currently present at the Learning Commons was 30.41, or about 30 students. This value was found in the Fall semester, when the day was Sunday, between 13.2917 hours (about 1:17 PM) and 17.725 hours (about 5:43 PM). The test RMSE was 7.3391 in the predicted model. The RMSE of 7.3391 tells us that the predictions were about 7 students from the true mean number of students present for each time interval. For context, the average number of students currently present at any given time was about 14 students.

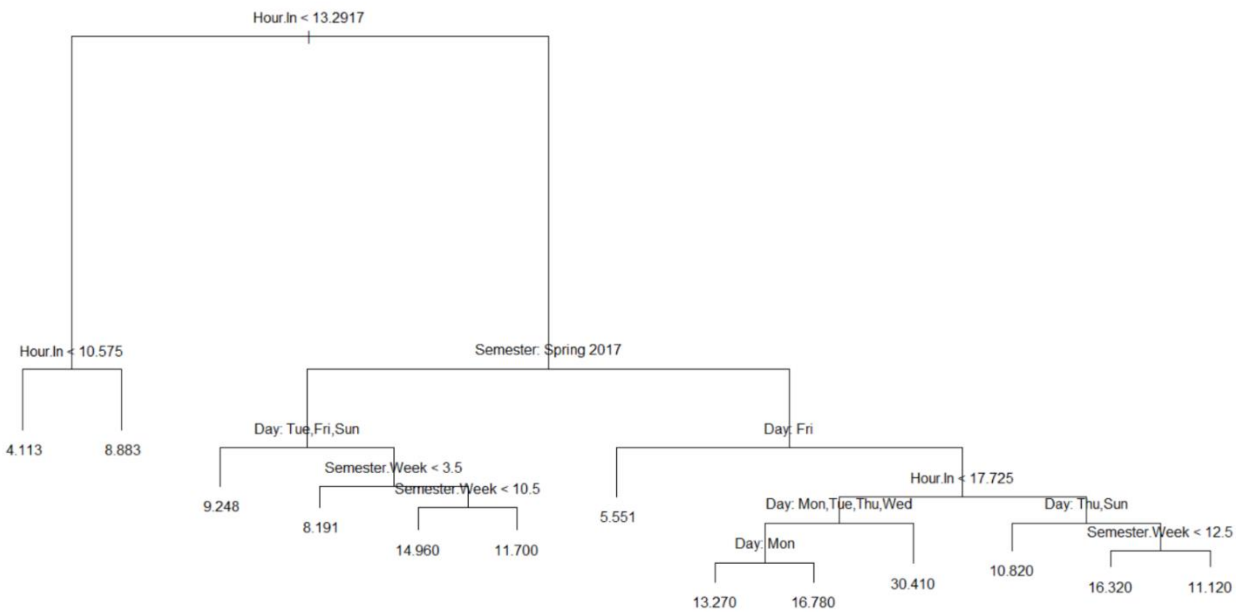

Figure 7: The decision tree node plot for the training dataset.

The results from the random forest prediction model showed that The Hour In predictor had the highest variable importance from the other predictors. The test RMSE was 7.1516 for the random forest model. The RMSE for the test model was similar to the test RMSE of the decision tree model, which means that there was little difference in error in the random forest model.

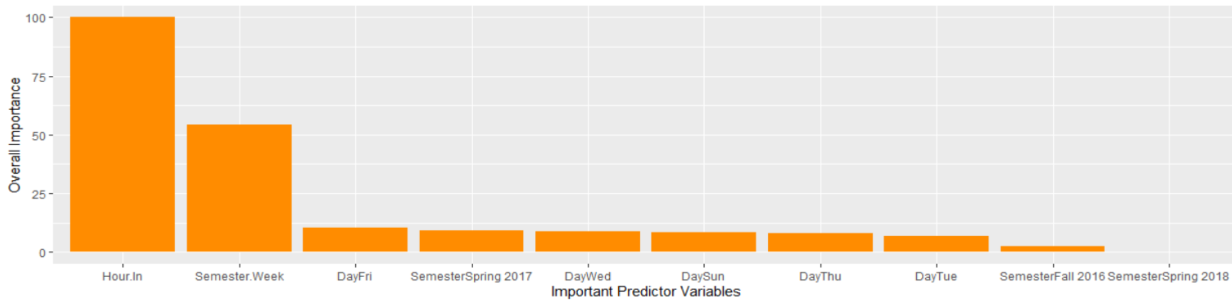

Figure 8: The variable importance plot for the random forest test model.

Learning Commons Course Analysis

Another area of interest was the type of courses that students sought out for individualized tutoring. This course analysis used the data from course type and course code. Course type is the subject of the course, and course code is the level of the course, shown in thousands. Students had individualized tutoring in 92 course types between Fall 2016 and Spring 2018. The ten common course types that students utilized individual tutoring for were Accounting, Business Administration, Biology, Chemistry, Computer Science, Economics, Mathematics, Operations Research, Psychology, and Statistics. Operations Research was the only course type listed that does not have a degree program at BGSU. These ten courses accounted for 83.01% (n=22,270) of the total visits to the Learning Commons for individualized tutoring. The most common course type that students sought tutoring assistance for was Mathematics courses with 44.52% (n=11,945) of the total visits between Fall 2016 and Spring 2018.

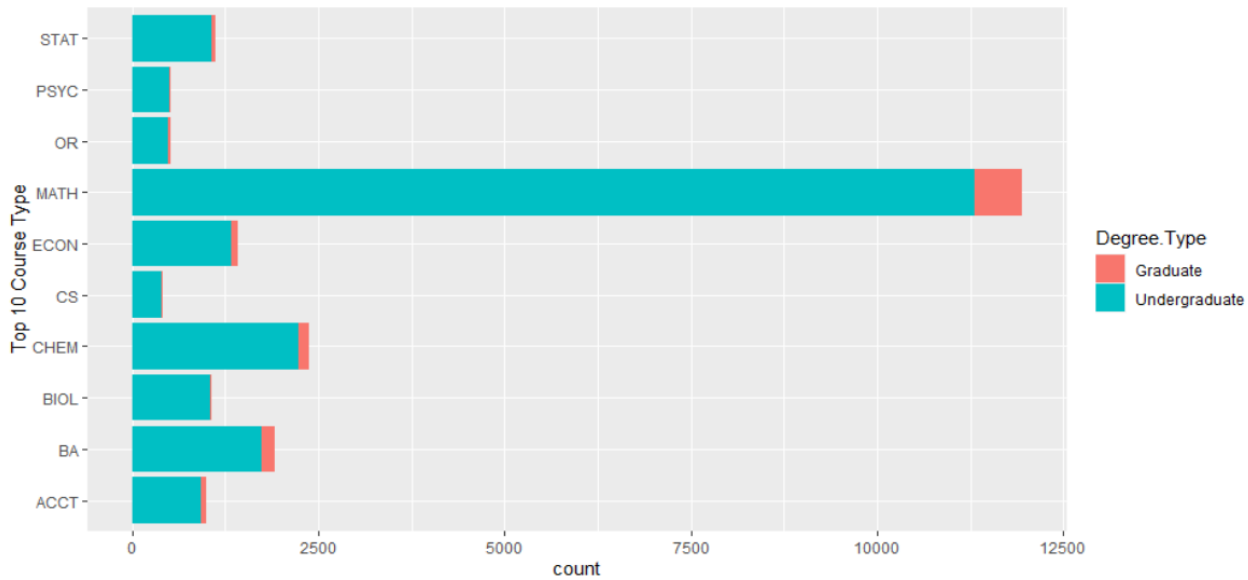

Figure 9: This bar chart displays the count of students (x-axis) and the top 10 course types (y-axis) that students utilize the Learning Commons for individual tutoring. The bar chart colors are grouped by degree type.

The course level of the class was also a factor in student tutoring visits to the Learning Commons. When course levels were grouped in the thousands, about 54.35% (n=13,723) of visits were for a 1000-level class, 29.56% (n=7,464) of visits were for a 2000-level class, 14.55% (n=3,674) of visits were for a 3000-level class, and 1.53% (n=387) were for classes that had a course level of 4000 or higher. Many required BGP courses are 1000 or 2000-level classes, which may explain why there was a higher percentage of those course levels. We can also use the grouped course level information to analyze the ten common course types. Students seek out tutoring at the 2000 level more than other course codes in the following course types: Statistics, Economics, Computer Science, Business Administration, and Accounting. Psychology had a slightly even proportion of students visiting for the 1000, 2000, 3000, and 4000 level courses. Students only sought out tutoring for one class in the Operations Research subject, which is OR 3800.

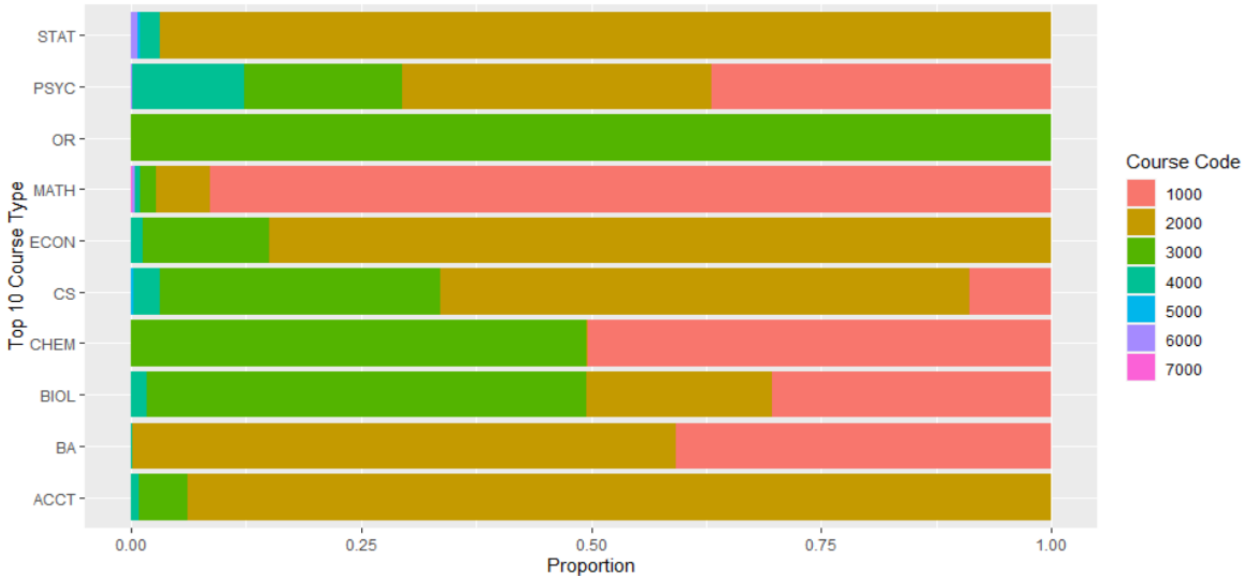

Figure 10: This bar chart displays the proportion of students (x-axis) and the top 10 course types (y-axis) that students utilize the Learning Commons for individual tutoring. The bar chart colors are grouped by the course code by thousands.

<u>Duration of Tutoring Sessions</u>

The Learning Commons collected data on the duration (in minutes) of the individualized tutoring sessions. The duration of a student visit ranged from less than a minute to 830 minutes, which is 13.83 hours. About 1.18% (n=313) of visits were longer than 300 minutes, which means that the duration of the visit was 5 hours or longer. To view the spread of the duration, I filtered the duration of visits to find tutoring sessions that lasted between 1 minute and 300 minutes. The most frequent duration for individualized tutoring sessions were between 70 and 80 minutes. When viewing the data by the course codes grouped by thousands, it showed that all course codes followed a similar right-skewed plot distribution. This means that course level did not have an impact on the duration of a tutoring session at the Learning Commons.

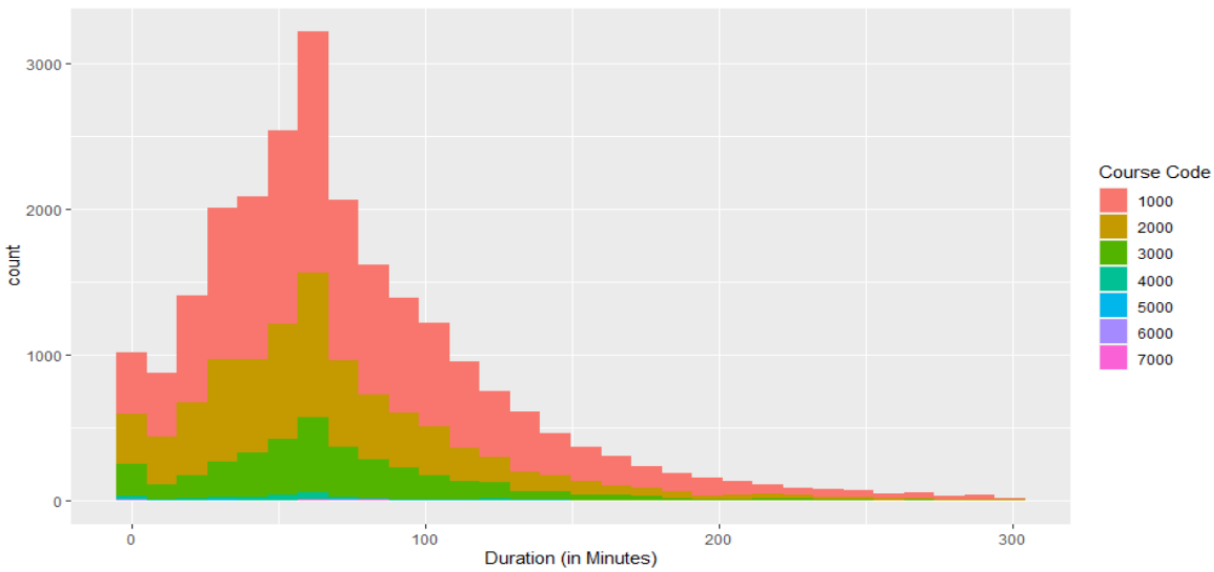

Figure 11: This bar chart displays the duration of a tutoring visit in minutes (x-axis) and the count of tutoring visits that lasted that duration (y-axis). The bar chart colors were grouped by the course code by thousands.

Another method to visualize duration was calculating the average and viewing the tutoring sessions by week in the semester and day of the week. The data included tutoring visits that lasted between 1 minute and 720 minutes, which is 12 hours. The average individualized tutoring session lasted between 70 and 111 minutes each week. Week 17, which is Finals Week, had the highest mean duration for student visits. This suggests that students may spend more time in individualized tutoring sessions to prepare for their final exams.

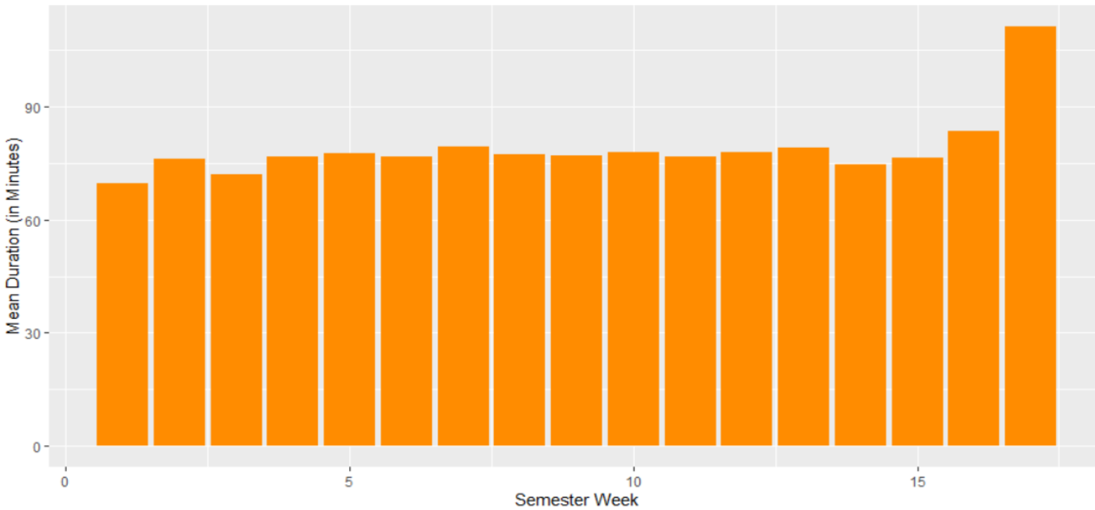

Figure 12: This bar chart displays the semester week (x-axis) and the mean duration of a student visit (y-axis) to the Learning Commons for individual tutoring.

For the day of the week, the average individualized tutoring session lasted between 73 and 93 minutes. The Learning Commons is closed on Saturday, so there were no visits that day. Sunday had the highest mean duration for student visits on that day in the week. The increase in the mean duration on Sunday showed that students were able to spend more time in a tutoring session on a weekend rather than during the weekday.

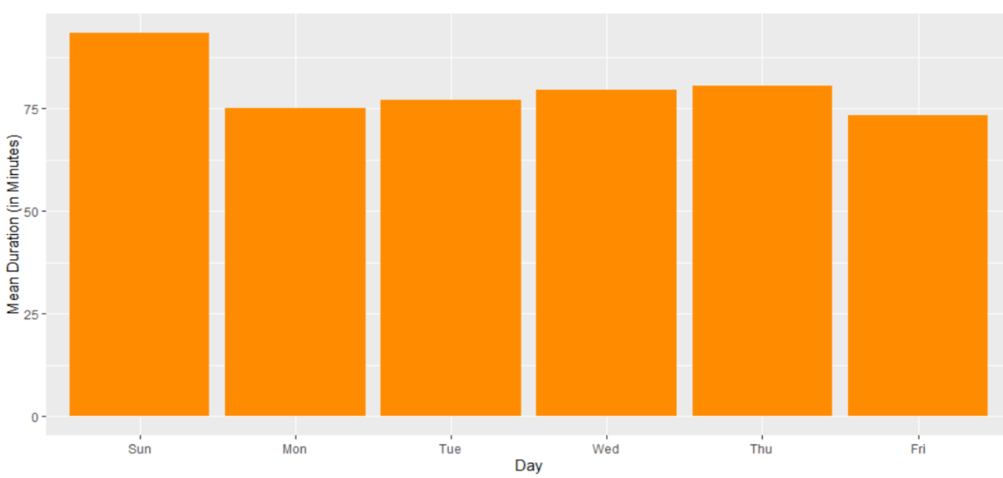

Figure 13: This bar chart displays the day of the week (x-axis) and the mean duration in minutes of a student visit (y-axis) to the Learning Commons for individual tutoring. The Learning Commons is closed on Saturday.

Limitations

The BGSU Learning Commons used specific software to collect and store data from student visits. One limitation to the data were the inaccuracies when the data were collected. There were rows of visits that went well outside the hours of operation for the Learning Commons, sometimes spanning more than a day. Most likely, students forgot to use their student ID to swipe out of the tutoring session, and when that student came back for additional tutoring a different day their new check-in swipe was counted as checking out for the previous tutoring session. Those rows were removed prior to analyzing the data, but the data did not reflect all tutoring visits. There were some missing values in rows of data, which made it difficult for analysis until I found the missing values and added “No Response” to those cells.

Another limitation was the amount of data that I received from the Learning Commons. In my initial proposal, I indicated that I would receive three years of data from the Learning Commons, but the 2018-2019 data had a large amount of missing data for Student Information. This meant that I was working with less data than I had anticipated. There were also inaccuracies in some of the data columns. The Cumulative GPA column had overwritten data, so the GPA shown was the most recent GPA. I could not view the change in Cumulative GPA by each semester, which contradicted one of my original project questions. The Classification column was also unclear. Originally, I had viewed the Classification column as Class Standing and Expected Graduation Year. However, the Classification column showed the students’ last known class standing and semester that they visited the Learning Commons, which included visits from Spring 2021. This would have been difficult to determine what their real class standing was when they initially visited the Learning Commons.

Conclusion

The BGSU Learning Commons provided a large volume of information about the details of student visits to the Learning Commons during the 2016-2017 Academic Year and 2017-2018 Academic Year. This information went through extensive data cleaning and data preparation in order for the information to be analyzed and explained. The results suggest that the student usage data followed similar trends each semester and academic year.

The visit analysis determined how often students returned to the Learning Commons for individualized tutoring after one visit and when students visited in the semester. Many students returned more than once between the two academic years. Student also visited the Learning Commons for individualized tutoring during midterms and on the week before finals were given. Students typically sought out tutoring during the week rather than on the weekend. The time of day also had an impact on student visits, where a larger number of student tutoring sessions occurred in the afternoon between 12:00 PM and 4:59 PM. The peak time for tutoring sessions were around 4:00 PM. Analyzing the past student usage data showed that the four semesters had similar trends for student tutoring visits which could help leadership in the Learning Commons with tutor schedules and identify overall popular times.

The prediction analysis complemented the student usage analysis by determining how many students would be currently present in attendance given the hour the student checked in, the day of the week, semester, and week in the semester. The decision tree analysis predicted that there would be more students present when students checked in sometime between 1:17 PM and 5:43 PM on a Sunday in the Fall semester. This aligned with the time of day that a student visited and involved days during the week. The random forest model predicted that the hour in was the most important predictor of when students would be present for individualized tutoring at the

Learning Commons. This prediction analysis further helps the Learning Commons plan for student visits during the semester.

The course analysis viewed what type of courses students sought out for individualized tutoring. Using the course type and course code information, I found that students utilize the Learning Commons for individualized tutoring in ten common course types. Mathematics was the most common course that students sought out tutoring help. Students also sought help for tutoring at the 1000 level more than other class levels. This could help tutors understand where students need the most assistance in their classes.

Finally, I analyzed the duration of the tutoring session to determine the average duration on a given week in the semester and the day in the week. The average duration for an individualized tutoring session was between 70 to 111 minutes, which shows that the mean duration of a session lasts more than an hour. This information could be useful to the Learning Commons for planning ahead with scheduling tutoring sessions to last longer than an hour.

References

Arco-Tirado, Fernández-Martín, & Hervás-Torres (2020) Solution for forecasting labor demand of drop-in peer tutoring centers with long planning horizons, Studies in Higher Education, 45:11, 2190-2202, DOI: 10.1080/03075079.2019.1597038

Brattin, R., Sexton, R.S., Yin, W. et al. Author Correction: A neural network solution for forecasting labor demand of drop-in peer tutoring centers with long planning horizons. Educ Inf Technol 25, 5897 (2020). https://doi-org.ezproxy.bgsu.edu/10.1007/s10639-019-09967-3

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. American Educational Research Journal, 19(2), 237–248. https://doi.org/10.3102/00028312019002237

Cooper, E. (2010). Tutoring Center Effectiveness: The Effect of Drop-In Tutoring. Journal of College Reading & Learning, 40(2), 21–34.

Daniel, B. (2015). The Value of Big Data in Higher Education. Br J Educ Technol, 46: 904-920. https://doi-org.ezproxy.bgsu.edu/10.1111/bjet.12230

Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second Edition ed., Springer Series in Statistics). Springer.

Muise, A. J. (2016). The importance of college level peer tutoring from the student's perspective (Order No. 10170613). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (1849014071).

Roscoe, R., & Chi, M. (2007). Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors' Explanations and Questions. Review of Educational Research, 77(4), 534-574.