Time of Day vs Course Performance

Darin Young

Applied Data Science – DSC 680

August 2025

Business Problem

My project is centered the idea that students of Midwestern State University (and other colleges, for that matter) may have more struggles in courses that are during times of the day or days of the week that are problematic for their personal schedule. For example, there may be core courses (courses that every student must complete to obtain a degree from MSU) that are only offered at 8AM on weekdays. Could these courses see higher completion/success rates if more sections were offered during times that are more fit for a typical college student? My hypothesis for this project is:

Courses that are offered on days/times that are inconvenient for the typical college student result in lower completions/success rates than courses that are offered at convenient times.

If my hypothesis is true, there may be reason to encourage leadership to adjust course scheduling to see if outcomes improve based on the change.

Background/History

As part of the completion of the capstone course of my Master's Program at Bellevue University, I was asked to complete three totally open-ended research projects. Considering my role as a Data Analyst at Midwestern State, I felt it was an excellent opportunity to answer some institutional research questions regarding our students. My supervisor, Dr. Eboneigh Harris, had already been doing some exploratory research into this idea and seemed to be discovering some interesting trends. With that said, I felt it was an excellent problem to explore further. Additionally, Dr. Harris seemed to believe that course performance was just generally worse in the early hours of the day (8am classes are the worst). My goal was to either prove or disprove this feeling.

Data Explanation

The data for this project is real student data from the Fall and Spring semesters from Fall 2022 to Spring 2024. I used three datasets for this project, and they are as follows:

Courses Dataset

Each row in the courses dataset represents a course that a student attempted in a given term and each column is an attribute for that course. The columns included are described below:

- TERM
    - Term code for the given course
- CRSEID
    - Unique ID for the given course
- SUBJ
    - Four letter subject code for the given course
    - ACCT, ENGL, POLS, etc.
- NUMB
    - Number for the given course
    - 1133, 1433, etc.

- SECT
    - Section number for the given course
    - 101, 102, 103, etc.
- Merge
    - Manually created column which combines SUBJ, NUMB, and SECT to join with the Grades dataset.
- Online Course
    - Manually created column that identifies online courses with a 1/0 flag.
    - Courses with an "X" at the beginning of SECT are online courses.
- TITLE
    - Name of the given course
    - Financial Accounting, Graphic Design II, etc.
- INSTTYPE
    - Instruction type for the given course
    - Lecture, lab, internship, etc.
- MTNGDAYS
    - Days when the course met
    - MWF, TR, MTRF, etc.
- BEGTIME
    - Time the course began in military time
    - 0700 = 7:00am, 1500 = 3:00pm, etc.
- ENDTIME
    - Time the course ended in military time.

Grades Dataset

Each row in the grades dataset represents a grade that a student received in a given term and each column is an attribute for that course. The columns included are described below:

- MID
    - Unique student identifier
- TERM
    - Same as above, term value for the given course
- SUBJECT
    - Four letter subject code for the given course
- SECTION
    - Section number for the given course
- TITLE
    - Name of the course
- GRADE
    - Letter grade for the course
    - A, B, C, D, F, W, I, etc.
    -

- Merge
    - Same as above, used to join the courses and grades datasets

Demographics Dataset

- MID
    - Unique student identifier
- TERM
    - Most recent term that this data derives from
- Gender
    - Student gender
    - M, F, N (if unknown)
- Ethnicity
    - Coded values for the student's race.
    - 0 = Missing
    - 1 = Non-Resident Alien
    - 2 = Hispanics of Any Race
    - 3 = Black or African American
    - 4 = White
    - 5 = American Indian Alaskan Native
    - 6 = Asian
    - 7 = Native Hawaiian Pacific Islander
    - 8 = Two or More Races
    - 9 = Unknown
- Classification
    - Freshman, Sophomore, Junior, etc.
- Age

Data Preparation

This dataset was prepared partially using Excel and partially using Jupyter notebook. The first thing I did in Excel was create the “Merge” column in the Courses and Grades datasets. Next, I created the “Online Courses’ column (1 = Online Course, 0 = not) in the courses dataset to easily identify online courses. Both of these steps could’ve been done in Jupyter, but for some reason I was having difficulty doing so. So, I opted to just create them in Excel before importing to Jupyter.

After the data was loaded to Jupyter, I made a few more changes to the data:

Only include lecture courses

After discussing with Dr. Harris, we both agreed that the analysis would be most powerful if we only looked at a subset of courses in order to best compare how time impacts performance. So, I limited the dataset to include only lecture courses, which are the majority of courses taught each semester and are the “typical” courses that a traditional student might take. In essence, this excluded courses that don’t have regular start times (internships, clinicals, etc.) and courses that simply are for credit only and don't have start times at all (independent studies, thesis, etc.).

Only include undergraduate courses

Second, I limited the dataset to only include undergraduate courses (where NUMB is less than or equal to 4999) because this analysis is mostly focused on the traditional student, not graduate students.

Exclude courses for credit and courses marked "Incomplete"

Then, I excluded all courses from the grades dataset that had a grade value of "NC", "CR" or "I". NC and CR represent no credit and credit given, respectively. Courses that yield this type of grading are typically non-traditional courses and also do not allow for any analysis as it relates to performance (DFW rates and/or GPA points). Most courses with this grading system were removed when I filtered out all courses not named "lecture", but some still remained so they needed to be removed manually. Courses with the grade "I" were also excluded because they represent courses in which a student was removed from a course for non-performance related reasons (family emergency, health, etc.).

Categorize time of day values

Then, I felt it would be best to group courses with similar start times into categories. The categories I determined were:

- 7:00-9:00am = Early Morning
- 9:00-11:00am = Morning
- 11:00am-1:00pm = Midday
- 1:00-5:00pm = Afternoon
- 5:00pm and after = Evening

Categorizing each time value made it easier to visualize and compare across groups when it comes to performance overall.

Removing Evening courses

Then, I removed all evening courses from the dataset. I did this because I felt that the primary goal of the analysis was to compare performance across times of the day for the **traditional student**. Generally speaking, the traditional student does not take courses that start after 5pm. In fact, it's likely that students that take these courses are probably working adults or individuals who are returning to school after not completing their degree – both of which are not traditional students.

As an added note, there were around 158,000 courses in the dataset over the 3-year span. Only around 3,000 of them (less than 2%) started after 5pm. So, a very small minority that likely contains non-traditional students. The breakdown of overall performance/enrollment separated by time category be seen in Appendix A.

Merging Datasets

Finally, I merged each of the datasets using the "Merge" and MID columns mentioned above. All three datasets were combined into one large dataset, so my analysis was ready to begin.

Methods

For my analysis, I used the following methods:

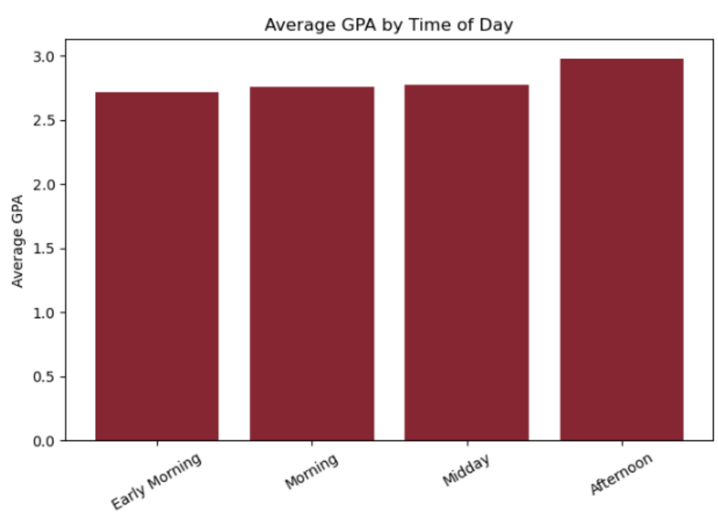
- Exploratory Data Analysis
  - I first looked at overall performance trends for each time-of-day category.
    - Do grades get better in courses as the day goes on?
    - Do students drop, fail or withdraw (DFW) more often in earlier courses?
- ANOVA Testing
  - Is there a significant difference between GPA/DFW rates for each time category?
- Logistic Regression
  - Then, I created a logistic regression to predict whether a student would DFW from a course.
  - I used Time Category as the primary predictor with other demographic variables as controls (Ethnicity, Classification, Gender, etc.)
- Linear Regression
  - Then, I created a linear regression to predict GPA, using the same predictors as above.

Analysis

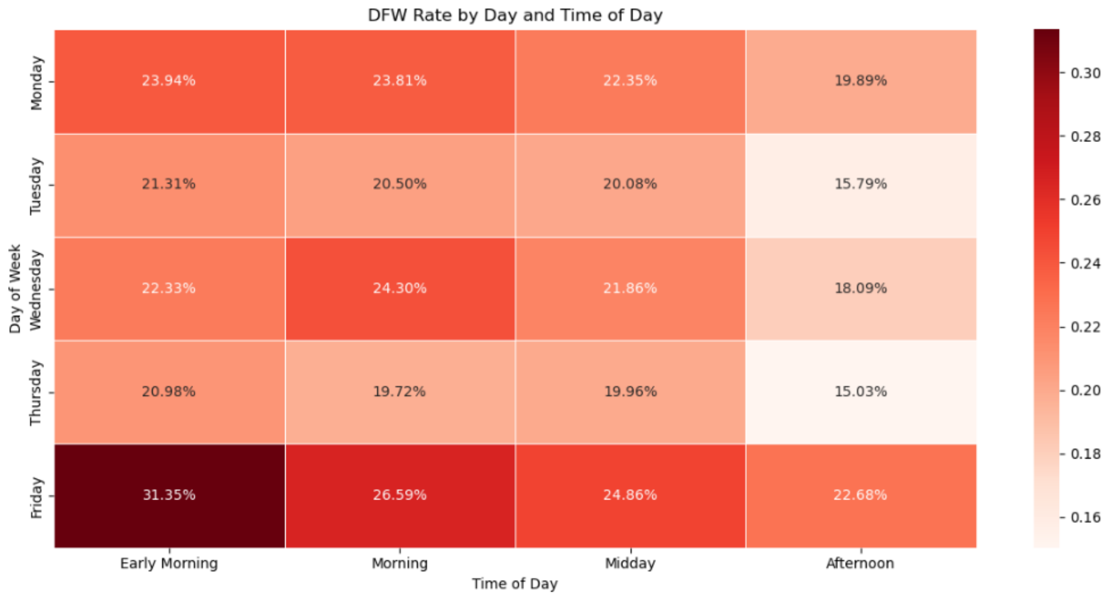
Exploratory Analysis

As mentioned above, the first step of my analysis was to explore the dataset related to the hypothesis and see what trends emerged. Generally speaking, I found that most courses over the three-year time span were offered during the morning or midday. As mentioned previously, a full breakdown of enrollment (and performance statistics) for each time category can be found in Appendix A.

In my initial exploration, I found that GPA slightly increased later in the day. Additionally, DFW rates were at their highest in the early morning and morning and declined later in the day – both of which aligned with my original hypothesis. I also found that when I viewed start time as a continuous variable rather than a categorical one (using the time category created earlier), the general trend held for DFW rates. All of these visuals can be seen in Appendix B.

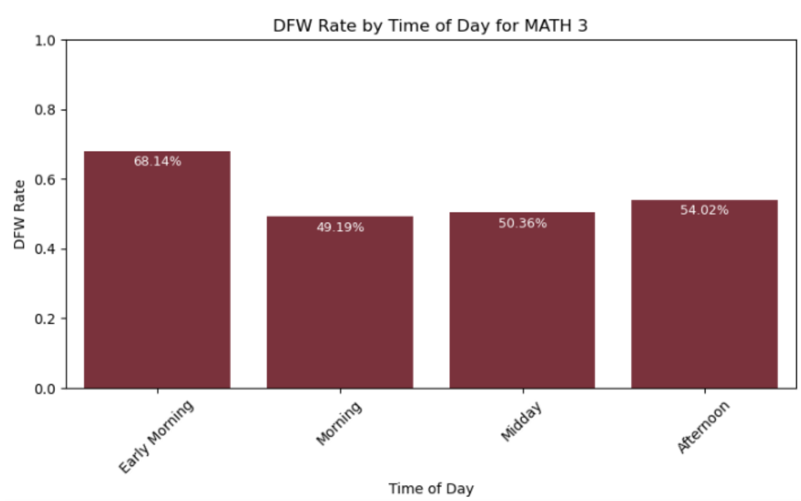


Next, I was interested to see where the highest DFW rates and lowest GPA occurred across days of the week vs time of day. So, I created a few heatmaps to visualize these relationships. Initially, I created a heatmap that was course schedule (all combinations of MWF, TR, MWRF, MF, etc.) vs time category. This initially seemed very interesting, but I quickly learned that the visual relationship was essentially meaningless, because the categories with the highest values were generally high because they had a small sample size. For example, MTRF in the early morning had a 58% DFW rate over the three-year span. This was only because there were a total of three courses offered using this schedule during the early morning category. So, I instead broke out the day of the week variable so if a course had Monday in its schedule, it would be counted. And then additionally, any course that had Tuesday would be counted in Tuesday. In essence, I made non-exclusive categories to see if there were any visual relationships between time of day and day of week overall. With this change, I made a new heatmap, which is shown below. As you can see, DFW rates are simply higher in the earlier times of the day - the



most severe instance being on Fridays. In other words, Friday classes at 8AM are likely not going to do well for most students. The overall heatmap mentioned above and the heatmap for GPA (which shows similar trends) can be seen in Appendix C.

My next point of interest was observing specific courses to see if their individual performance might improve later in the day. So, I found the top 10 courses in terms of DFW rate and the bottom 10 courses in terms of GPA (with a minimum overall enrollment of 200 over the three-year span, to ensure relevant sample), and visualized their performance over the various time categories. Generally speaking, the trend was present. However, it wasn't as clear as the overall picture, mostly because some courses are only offered at specific times each term. Take Beginning Algebra, for example – it had the third highest overall DFW rate (53%) over the three-year span and was taken by over 800 students. But when you break it down by time of day, the trend is shocking (shown to the right). As you can see, the highest



category for this course, by far, is early morning. However, this is only one example – and in the interest of fairness, there are counter examples. MATH 1233 (College Algebra) had the second highest DFW rate (56%). But, the lowest time category was in fact early morning (54%) as compared to the highest category being afternoon (60%). What I mean to say is, the trend we are observing – students perform better in courses when they are offered later in the day – can be true for specific courses while untrue for others. With this in mind, I refocused my analysis at this point to view overall trends rather than that of individual courses.

In any case, the visuals for each of the courses mentioned above are not included in this report – so as to not bloat the document. They are available upon request.

Statistical Significance Testing (ANOVA & Games Howell)

My next step was to test whether the differences between each time category were statistically significant, to ensure that the analysis could actually result in a meaningful conclusion.

My first test was an ANOVA test for GPA vs time category, which resulted in a p-value of 1.69e-155 (essentially zero). This implies that the relationship between GPA and time of day is not due to random chance – **i.e., time of day impacts GPA**. Next, I did a Games Howell test, which compares each time category against each other and tests individually for statistical significance. In every case, the difference between GPA for the various time categories was significant at the 5% threshold. This again implies that time of day impacts GPA and is not due to random chance.

I then did the same tests for DFW rates vs time category. The p-value for the ANOVA test was 5.75e-78, again essentially zero, which again implies that DFW rates are impacted by time of day. The same also held true for the Games Howell test for all but one comparison – early morning vs morning. The p-value for this comparison group was .974, which did not meet the 5% threshold. This implies that there is not a statistically significant difference between DFW rates for early morning classes vs morning classes. Practically, this means that courses offered before 11am are basically the same as it relates to DFW rates (when you don't consider specific days offered).

The results of all of these tests can be seen in Appendix D.

Regression Analyses

Once I had determined confidently that time of day had a relevant impact on success (GPA/DFW rates), I felt it would be valuable to build a model to predict the success metrics while controlling for demographic variables, to see if the relationship held.

First, I ran a logistic regression to predict DFW rates. I used time category, gender, ethnicity and classification as my predictors and yielded a .09 R^2 value, implying that 9% of the variance in DFW rates can be explained by the predictors. This R^2 value is quite low and can surely be improved.

Second, I ran a linear regression to predict GPA. I used the same predictors as above – time category, gender, ethnicity and classification – and yielded another R^2 value of .09. Again, quite low.

Considering the low performance of the models, the implication is that there are many other factors that contribute to student success in courses. This should be obvious – student preparedness, instructor, difficulty of course, difficulty of program, student engagement are all realistic factors that likely contribute to a student's level of success overall as well within specific courses. However, this does not mean that time of day does not contribute, as we have seen there is some relationship via the other steps in the analysis.

The results of both regression analyses can be viewed in Appendix E.

Conclusion

Overall, it seems clear that there is at least some link between course performance overall and when the course occurs during the day. With that said, if any relationship exists, it is that **students perform worse overall in morning courses as compared to those offered later in the day,** when controlling for other potential contributing factors.

Limitations

The main limitations for this analysis are as follows:

- Confounding Variables
    - Students are not randomly assigned to courses – they choose when to take a course. With that in mind, the performance in the course could be more directly related to the student's individual engagement, the instructor, and/or course difficulty and not related to the actual time of day.
- Statistical Significance vs Practical Significance
    - The dataset is quite large, so differences in GPA that are statistically significant (via p-value) might not be practically significant – i.e., a difference in GPA of .1 might be significant statistically, but not relevant as it relates to real world application.
- Correlation, not Causation
    - We can say that "GPA tends to be higher in afternoon classes" but we cannot and should not say that "Afternoon classes are a main cause for higher GPA."

Assumptions

The main assumptions for this analysis are as follows

- Time of Day Categories are meaningful and realistic
    - When creating the time-of-day categories, I created them in a way that felt realistic. But it is certainly possible that my categorization was inaccurate and unmeaningful.
- No interaction effects
    - My analysis assumes that each variable are additive in nature and not multiplicative in nature. For example, the impact of time-of-day is the same for freshman as it is for seniors, when it could be the case that the impact is greater for a given classification.

Challenges

The main challenge I faced during this analysis was likely the same challenge that most data professionals face - cleaning and preparing the data. One of my professors during my undergrad at Wilmington University stated that data professionals spend about 80% of their time cleaning and preparing data and only 20% actually analyzing the data. I can attest to that. Cleaning and preparing is an iterative process that requires fine-tuning and adjusting. As you move further through your analysis, there may be new things to add or change to prepare your data appropriately. It can be extremely tedious but it's what makes this process so fun and rewarding.

Future Uses/Additional Applications

I believe this analysis should be revisited in the future with the primary consideration being how specific instructors might impact course performance. It could be the case that certain instructors only teach at certain times (Professor Doe really likes the 8AM remedial math course, for example). If this is the case, time of day is just a mask for the real issue and should be resolved via allowing more instructors to teach courses with low performance. In any case, this would be my next step if I were to perform this analysis again in the future.

Recommendations

My recommendation based on this analysis would be to **adjust scheduling for specific, low-performing courses on an experimental basis for a few terms** to see if the performance of the course increases. Specifically, I would experiment with the Beginning Algebra course mentioned earlier in the analysis and open more sections of the course later in the day to see if the trend observed continues.

Implementation Plan

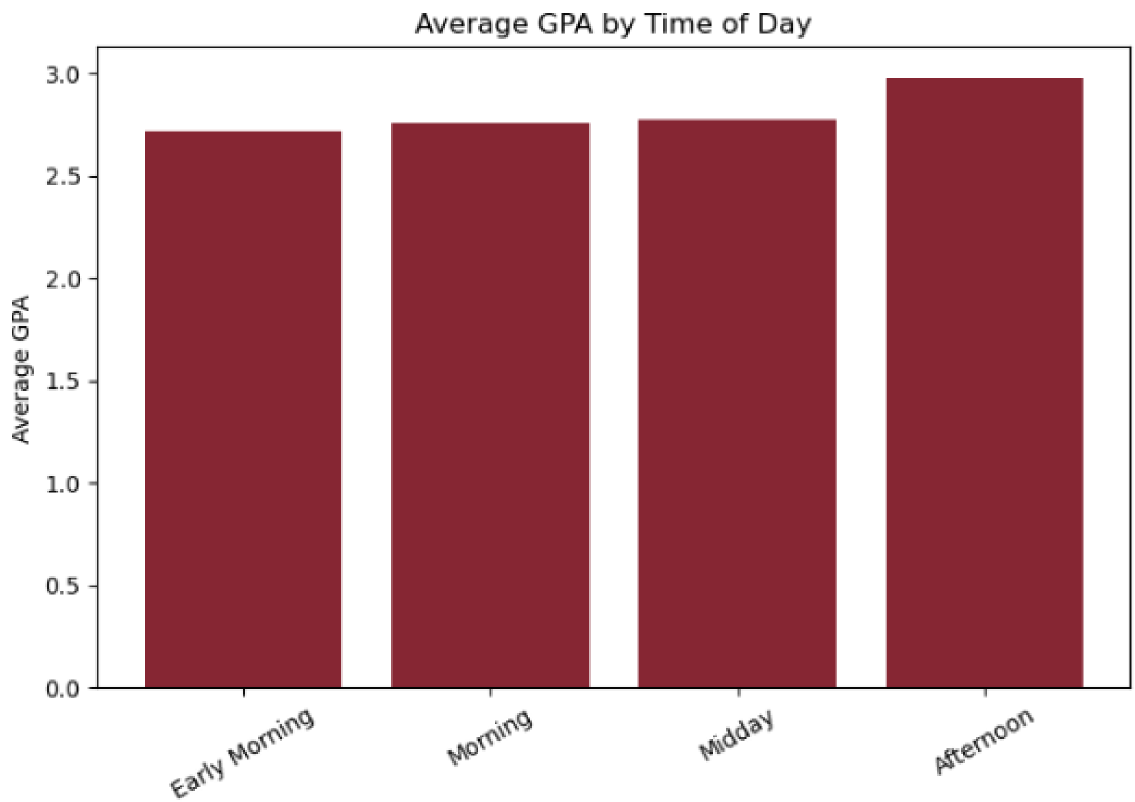
My implementation plan would be as follows:

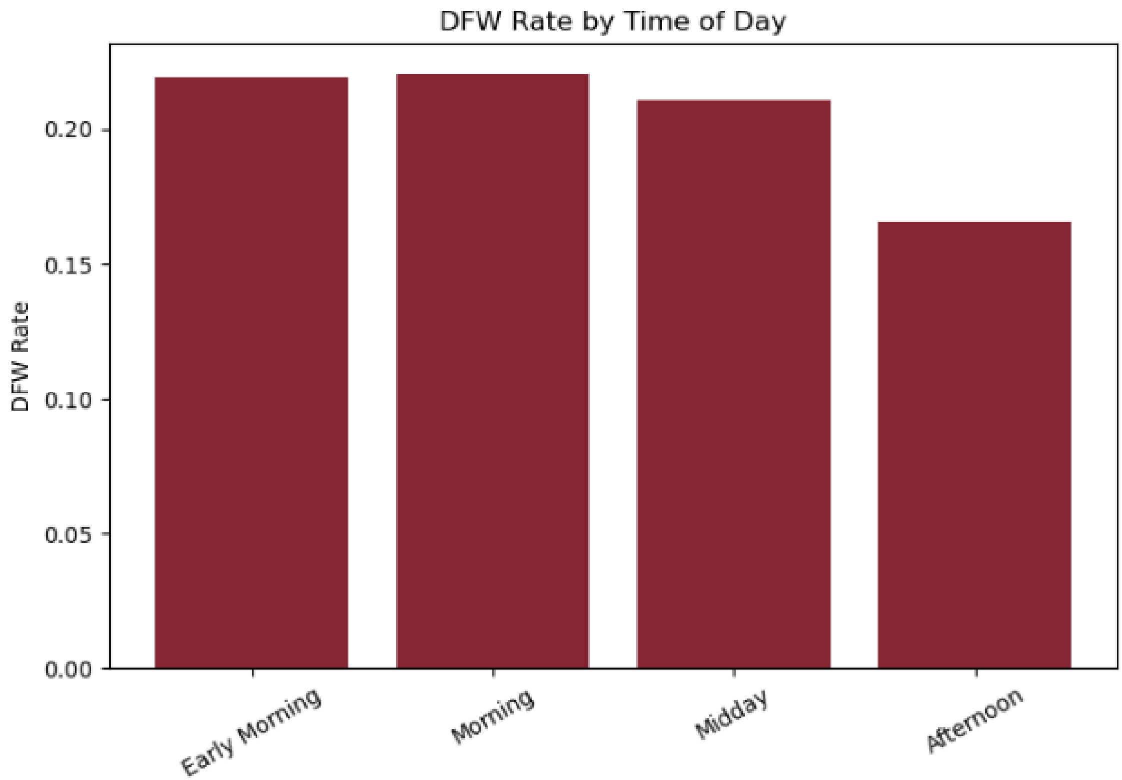
1. Adjust course scheduling to allow for more sections later in the day, particularly for low performing courses.
2. Observe impact for courses with adjusted schedule and revisit analysis.
3. If course performance increases, expand scheduling standards to more programs and observe impact.
4. If course performance stays the same or the difference is negligible, evaluate other factors like instructors, program difficulty, etc.

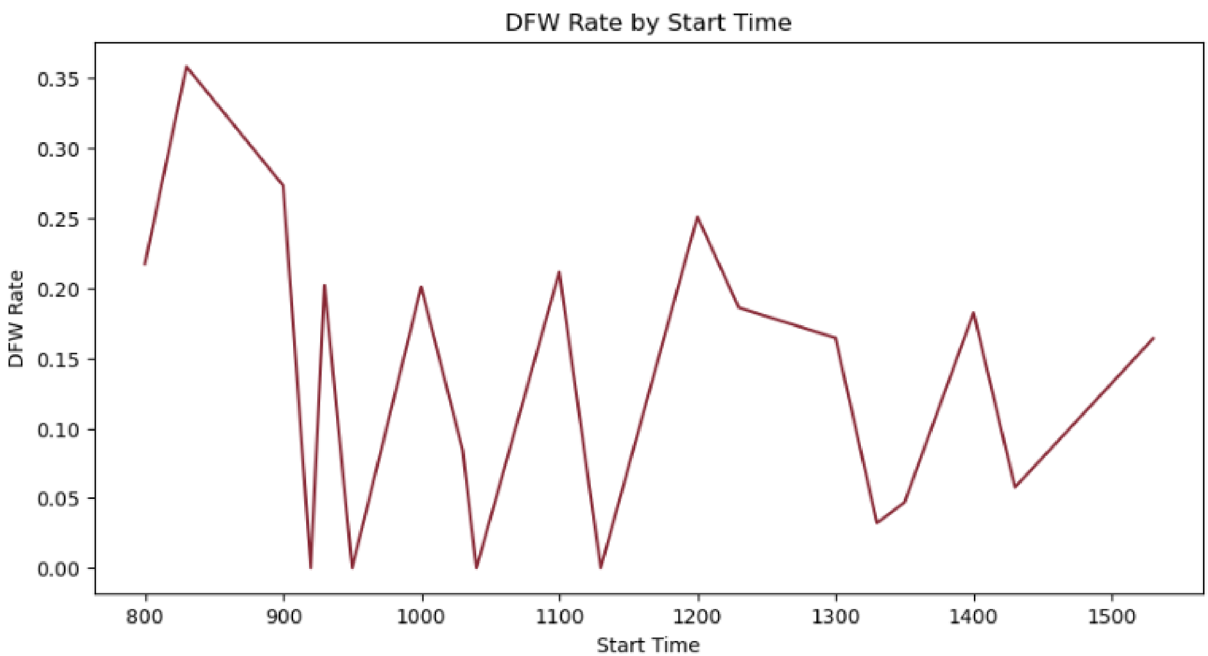
Appendix A – Overall Performance and Enrollment

	Time_Category	Avg_GPA	DFW_Rate	Enrollment_Count
0	Early Morning	2.716898	0.219250	23585
1	Morning	2.756887	0.220636	53867
2	Midday	2.777769	0.210500	51240
3	Afternoon	2.980995	0.165627	26439
4	Evening	3.384405	0.069080	3011

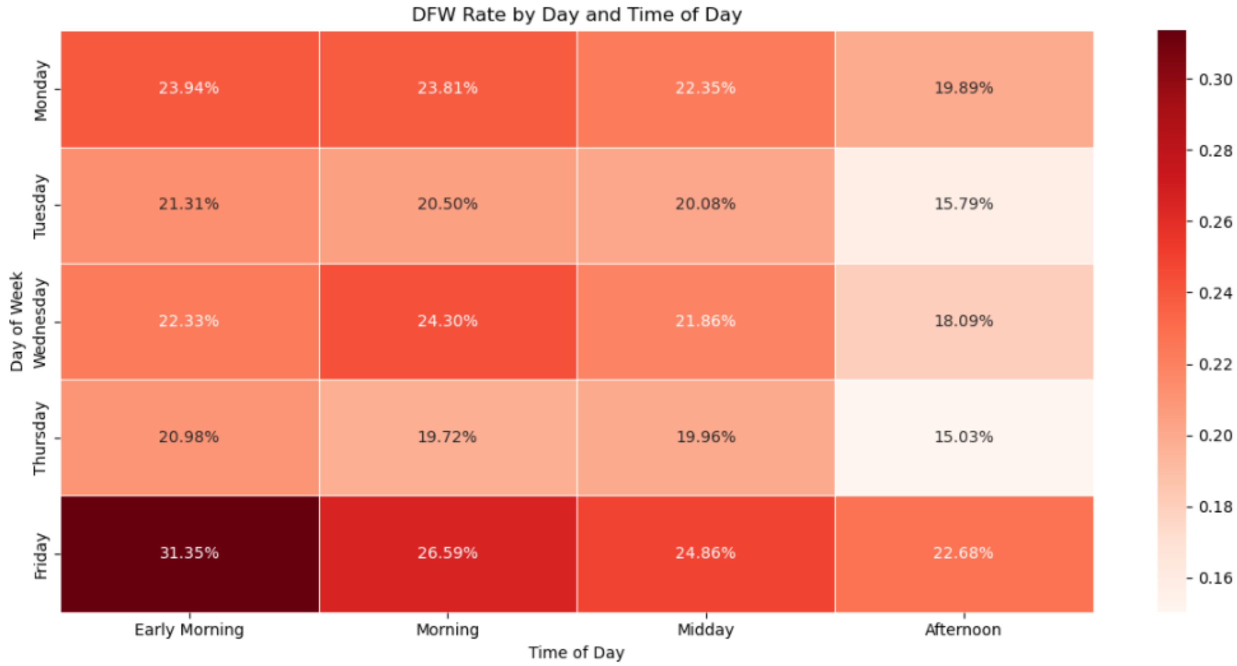
Appendix B – Success Metrics over Time

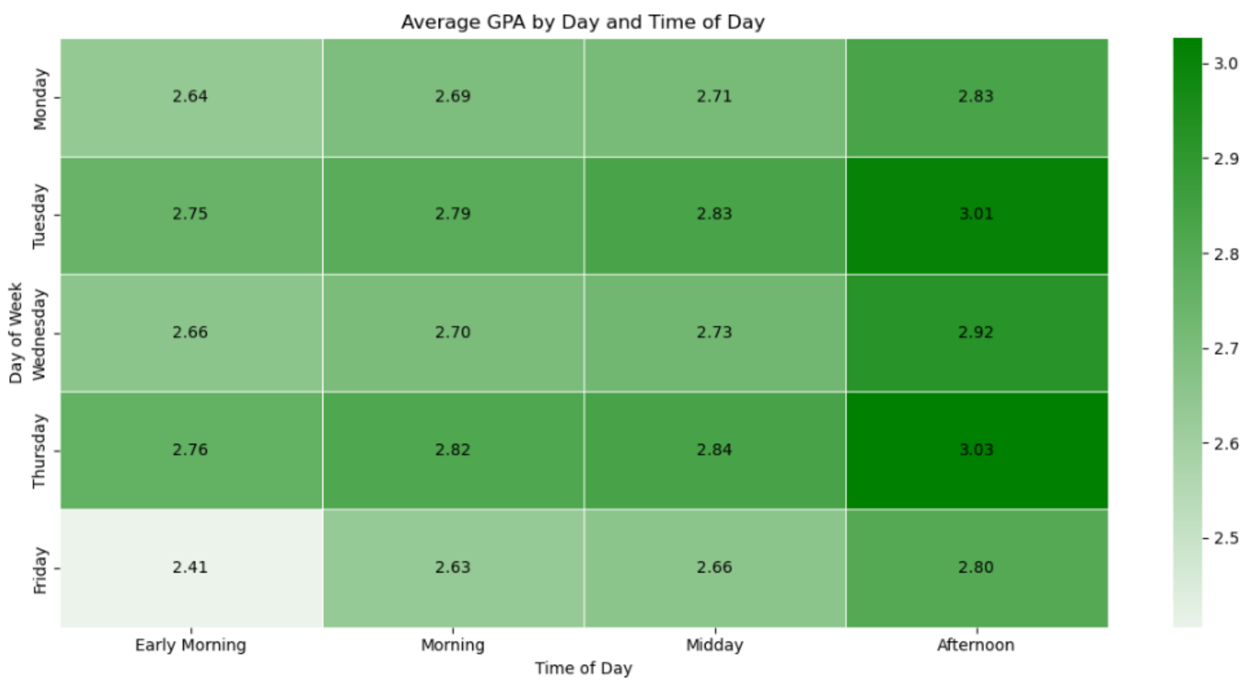


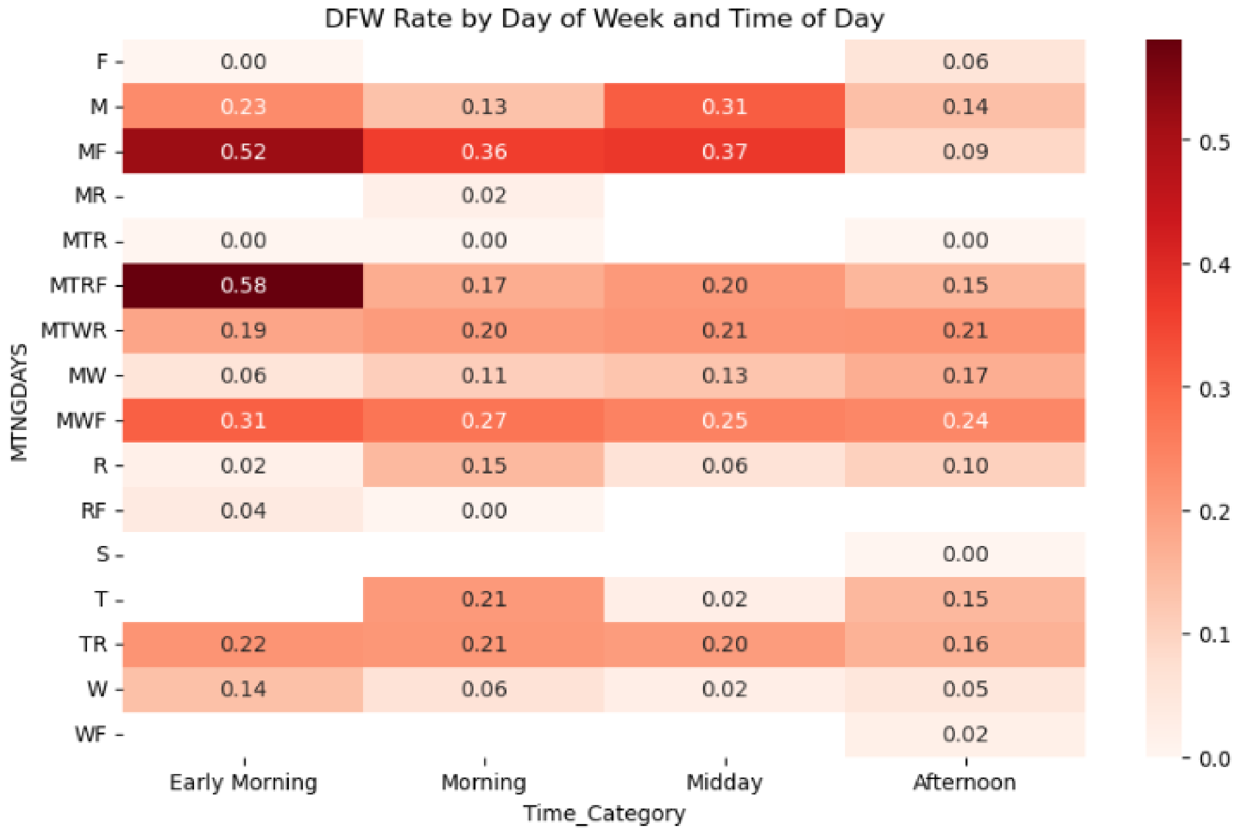

DFW Rate by Time of Day
DFW Rate
0.00
0.05
0.10
0.15
0.20
Early Morning
Morning
Midday
Afternoon


DFW Rate by Start Time
DFW Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
800
900
1000
1100
1200
1300
1400
1500
Start Time

Appendix C – Heatmaps (Time vs Day of Week)







Appendix D – Statistical Significance Testing

GPA

	Sum of Squares	Degrees of Freedom	F Statistic	p-value
C(Time_Category)	1,087.77	3	240.22	1.69e-155
Residual	219,411.33	145362	nan	nan

	Time A	Time B	Mean GPA (A)	Mean GPA (B)	Difference	Std. Error	T-statistic	Degrees of Freedom	p-value	Hedge's g
0	Early Morning	Morning	2.717	2.757	-0.040	0.010	-4.02	42727	3.41e-04	-0.032
1	Early Morning	Midday	2.717	2.778	-0.061	0.010	-6.07	43458	7.53e-09	-0.049
2	Early Morning	Afternoon	2.717	2.981	-0.264	0.011	-23.93	45568	0.00e+00	-0.222
3	Morning	Midday	2.757	2.778	-0.021	0.008	-2.63	98094	4.30e-02	-0.017
4	Morning	Afternoon	2.757	2.981	-0.224	0.009	-24.36	53355	0.00e+00	-0.184
5	Midday	Afternoon	2.778	2.981	-0.203	0.009	-21.91	54086	0.00e+00	-0.167

DFW Rates

	Sum of Squares	Degrees of Freedom	F Statistic	p-value
C(Time_Category)	59.36	3	120.52	5.75e-78
Residual	25,469.27	155127	nan	nan

	Time A	Time B	Mean DFW (A)	Mean DFW (B)	Difference	Std. Error	T-statistic	Degrees of Freedom	p-value	Hedge's g
0	Early Morning	Morning	0.219	0.221	-0.001	0.003	-0.43	45074	9.74e-01	-0.003
1	Early Morning	Midday	0.219	0.210	0.009	0.003	2.70	45215	3.50e-02	0.021
2	Early Morning	Afternoon	0.219	0.166	0.054	0.004	15.18	47711	8.29e-12	0.137
3	Morning	Midday	0.221	0.210	0.010	0.003	4.00	104991	3.75e-04	0.025
4	Morning	Afternoon	0.221	0.166	0.055	0.003	18.96	57978	0.00e+00	0.137
5	Midday	Afternoon	0.210	0.166	0.045	0.003	15.42	57920	0.00e+00	0.113

Appendix E – Regression Analyses

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                    DFW   No. Observations:               155130
Model:                          Logit   Df Residuals:                   155109
Method:                           MLE   Df Model:                           20
Date:                Fri, 08 Aug 2025   Pseudo R-squ.:                 0.09099
Time:                        09:53:24   Log-Likelihood:                -72043.
converged:                       True   LL-Null:                       -79254.
Covariance Type:            nonrobust   LLR p-value:                     0.000
===============================================================================================
                                  coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Intercept                      -1.4773      0.373     -3.956      0.000      -2.209      -0.745
C(Time_Category)[T.Morning]    -0.1240      0.020     -6.226      0.000      -0.163      -0.085
C(Time_Category)[T.Midday]     -0.1967      0.020     -9.746      0.000      -0.236      -0.157
C(Time_Category)[T.Afternoon]  -0.3946      0.024    -16.428      0.000      -0.442      -0.348
C(Gender)[T.M]                  0.2073      0.013     15.548      0.000       0.181       0.233
C(Gender)[T.N]                  1.5586      0.358      4.356      0.000       0.857       2.260
C(Ethnicity)[T.1.0]             0.7108      0.374      1.900      0.057      -0.023       1.444
C(Ethnicity)[T.2.0]             1.3611      0.373      3.646      0.000       0.629       2.093
C(Ethnicity)[T.3.0]             1.6219      0.373      4.343      0.000       0.890       2.354
C(Ethnicity)[T.4.0]             1.0645      0.373      2.852      0.004       0.333       1.796
C(Ethnicity)[T.5.0]             1.1621      0.380      3.062      0.002       0.418       1.906
C(Ethnicity)[T.6.0]             0.8842      0.376      2.354      0.019       0.148       1.620
C(Ethnicity)[T.7.0]             1.6971      0.393      4.314      0.000       0.926       2.468
C(Ethnicity)[T.8.0]             1.3185      0.375      3.520      0.000       0.584       2.053
C(Ethnicity)[T.9.0]             1.1775      0.375      3.139      0.002       0.442       1.913
C(Classification)[T.GM]        -2.2803      0.079    -28.980      0.000      -2.434      -2.126
C(Classification)[T.JR]        -1.1387      0.019    -60.484      0.000      -1.176      -1.102
C(Classification)[T.PB]        -2.3685      0.245     -9.668      0.000      -2.849      -1.888
C(Classification)[T.SB]        -1.4712      0.064    -22.876      0.000      -1.597      -1.345
C(Classification)[T.SO]        -0.8219      0.019    -43.529      0.000      -0.859      -0.785
C(Classification)[T.SR]        -1.7668      0.018   -100.669      0.000      -1.801      -1.732
===============================================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             GPA_Points   R-squared:                       0.111
Model:                            OLS   Adj. R-squared:                  0.111
Method:                 Least Squares   F-statistic:                     904.2
Date:                Fri, 08 Aug 2025   Prob (F-statistic):               0.00
Time:                        09:53:25   Log-Likelihood:            -2.2802e+05
No. Observations:              145366   AIC:                         4.561e+05
Df Residuals:                  145345   BIC:                         4.563e+05
Df Model:                          20
Covariance Type:            nonrobust
===================================================================================================
                                      coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------
Intercept                           2.4899      0.179     13.875      0.000       2.138       2.842
C(Time_Category)[T.Morning]         0.0990      0.009     10.525      0.000       0.081       0.117
C(Time_Category)[T.Midday]          0.1271      0.009     13.414      0.000       0.109       0.146
C(Time_Category)[T.Afternoon]       0.2702      0.011     25.137      0.000       0.249       0.291
C(Gender)[T.M]                     -0.1472      0.006    -23.684      0.000      -0.159      -0.135
C(Gender)[T.N]                     -1.1747      0.224     -5.251      0.000      -1.613      -0.736
C(Ethnicity)[T.1.0]                -0.2016      0.180     -1.122      0.262      -0.554       0.151
C(Ethnicity)[T.2.0]                -0.4919      0.179     -2.740      0.006      -0.844      -0.140
C(Ethnicity)[T.3.0]                -0.7155      0.180     -3.985      0.000      -1.067      -0.364
C(Ethnicity)[T.4.0]                -0.2408      0.179     -1.342      0.180      -0.593       0.111
C(Ethnicity)[T.5.0]                -0.3450      0.182     -1.893      0.058      -0.702       0.012
C(Ethnicity)[T.6.0]                -0.2238      0.180     -1.242      0.214      -0.577       0.129
C(Ethnicity)[T.7.0]                -0.7931      0.190     -4.167      0.000      -1.166      -0.420
C(Ethnicity)[T.8.0]                -0.4262      0.180     -2.367      0.018      -0.779      -0.073
C(Ethnicity)[T.9.0]                -0.3548      0.180     -1.967      0.049      -0.708      -0.001
C(Classification)[T.GM]             1.1265      0.024     47.924      0.000       1.080       1.173
C(Classification)[T.JR]             0.6868      0.010     69.391      0.000       0.667       0.706
C(Classification)[T.PB]             1.2298      0.074     16.550      0.000       1.084       1.375
C(Classification)[T.SB]             0.9240      0.026     35.351      0.000       0.873       0.975
C(Classification)[T.SO]             0.5038      0.010     48.301      0.000       0.483       0.524
C(Classification)[T.SR]             0.8991      0.009    104.180      0.000       0.882       0.916
==============================================================================
Omnibus:                     9742.438   Durbin-Watson:                   0.606
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11767.865
Skew:                          -0.693   Prob(JB):                         0.00
Kurtosis:                       2.849   Cond. No.                         263.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```