DATS 599 – Practicum
Final Report
Spring 2019

Eric Dong

Contents

1 Introduction **1**

2 Data **1**
2.1 Data Description 1
2.2 Data Preprocessing 2

3 Visualizations **2**
3.1 Expanding Previous Dashboards 2
3.2 Submatriculation Enrollment 3
3.3 One Deadline Versus Two Deadlines 3

4 Discussion **8**

5 Usage Instructions **9**

6 Future Work **10**

7 Acknowledgments **13**

1 Introduction

Over the course of the past decade, there has been a tremendous rise in interest towards STEM education, largely in response to rapid technological advancements and shifts in relevant skills desired by employers. The University of Pennsylvania School of Engineering and Applied Science (SEAS) has also experienced this trend. In 2007, SEAS received a total of 1,121 applications for its master's programs. That number exploded to 6,280 during the 2019 application cycle, marking a 38.6% year-over-year growth rate. Furthermore, since 2007, SEAS has expanded its offerings by adding 6 majors and increasing the number of engineering courses available. Such extreme growth can pose many logistical challenges. For example, the unprecedented increase in applications and available programs makes it difficult to predict future application numbers as well as enrollment rates, especially in the face of rising competition from other institutions. This information is essential to the various SEAS departments in terms of determining who to admit for the next year and allocating teaching staff and classroom space for these students. Additionally, visualizing past admissions and enrollment patterns can help the admissions office quantitatively measure the impact of policy changes. This is particularly relevant because in the past year alone, the online MCIT program was introduced and many programs shifted from two application deadlines to a single spring deadline. These changes will undoubtedly affect SEAS as a whole. Understanding admissions patterns and accurately predicting future results is crucial to managing the growth at SEAS appropriately.

For my DATS 599 Practicum, I attempt to address this problem by building upon the Tableau dashboard Jiadi Xiong previously created to visualize and predict admissions statistics in an automated manner. My objective is to expand upon the existing tool's capabilities so that it can easily scale and incorporate data beyond external applicants. In particular, I generated additional Tableau-based visualizations and predictions that included submatriculants as well as the ability to compare having two application deadlines against one. The final goal was to have an easy-to-use tool for a wide range of stakeholders. The SEAS admissions office staff for each department may use this dashboard to predict how many applications they will receive and how many students will matriculate once offered admissions in order to allocate spots. Furthermore, they can examine the impact of past admissions policy and program changes, such as the shift from two deadlines to one and the introduction of the online MCIT program, in order to guide future policy. Ideally, administrative staff and professors in each department can also use this tool to predict demand for specific majors and classes and use this information to determine each semester how many professors should be teaching, which classes should be offered, and how restrictive should class enrollment be.

2 Data

2.1 Data Description

Detailed admissions data were provided for each department from 2007 to 2019 by the Penn Engineering Graduate Admissions Office. The data includes information on the number of people at each stage of the admissions pipeline for both the master's and PhD programs, with the stages being applied, admitted, matriculated, and enrolled. In the case of a program using two deadlines, each deadline was treated as a separate pipeline with their own applied, admitted, and matriculated statistics as well as an initial and final enrollment breakdown. There were also supplemental data in the form of high level demographics (i.e. gender split) and more granular metrics (i.e. attribution of not enrolling from not accepting the offer, deferring in or out, and not responding). Additionally, data were given for the number of master's students who enrolled each year via transfer and submatriculation. All these data were updated periodically at intervals throughout each year. For the purposes of this practicum, the data of primary interest were the finalized applicant, admitted, enrolled, and submatriculated numbers of each deadline for each year for every program. As of the submission of this report, the statistics for the 2019 admissions cycle were partially incomplete.

2.2 Data Preprocessing

The provided data were clean and outside of the occasional update, were ready to use. However, preprocessing in terms of formatting was required in order to ensure compatibility with Tableau. Tableau requires a time series data source to contain fields as columns and timestamps as rows. The data explained in Section 2.1 were thus reformatted into `Tableau_Input.xlsx`, which as the name suggests, serves as the data source for my Tableau dashboards. Table 1 contains the schema for visualizing and predicting general admissions data and submatriculant enrollment while Table 2 contains the schema for the Tableau dashboard comparing the switch from two deadlines to one deadline. These correspond to the `FINAL` and `2DDL` tabs in `Tableau_Input.xlsx`, respectively.

Year	Program	Major	Status (AAE)	Number
2007-2019	{MS, PHD}	All departments (i.e. DATS)	Applicant, Admitted, Enrolled, Submatriculate	int

Table 1: Schema for general admissions and submatriculation data

Year	Program	Major	Status (AAE)	Deadline (DDL)	Number
2007-2019	{MS}	All departments (i.e. DATS)	Applicant, Admitted, Enrolled (Inital), Enrolled (Final)	First Deadline, Second Deadline, One Deadline	int

Table 2: Schema for two versus one deadline comparison data

The data file, `Tableau_Input.xlsx` also contains a few additional tabs. Specifically, `Submatrics` contains only the data relevant to submatriculations while `Trans` is an empty placeholder for formatted transfer data. Currently, these tabs do not play any role in the Tableau dashboards. The last tab, `Prediction` contains the validations of predictions from Tableau's forecasting feature for recent years in the form of comparing predictions with their actual realizations and the resulting mean absolute percentage error (MAPE). The `Prediction` tab will be discussed further in Section 4.

For the sake of readability, some fields and values were aliased, or renamed, once imported into Tableau. Additionally, in order to generate forecasts for all the provided time series, null or empty values were set to 0. This padding was done because Tableau's built in forecasting function, which uses exponential smoothing requires at least five data points in each time series to perform predictions.

3 Visualizations

Using Tableau, I generated a series of visualizations. As mentioned earlier, my primary contribution was expanding the dashboard from just external applicants to also include submatriculants and simultaneously display multiple deadline formats. The visualizations summarize historic data as well as, when appropriate, project the data out two years using Tableau's forecasting option.

3.1 Expanding Previous Dashboards

The previous dashboards built by Jiadi, `Admission Statistics.twbx` and `Admission Statistics - 2DDL.twbx` highlighted the admissions statistics for external applicants. The visualizations consisted of color-coded tables and multi-element charts that explored applicant numbers, admissions rates, and enrollment rates across years and majors. By augmenting the input data source, I added more components to each graph and included the 2019 application cycle. For example, the sheet `Major vs (Recent) Year` in `Admission Statistics.twbx` is now capable of comparing submatriculant numbers on top of the other admissions statistics, as highlighted in Figures 1 and 2.

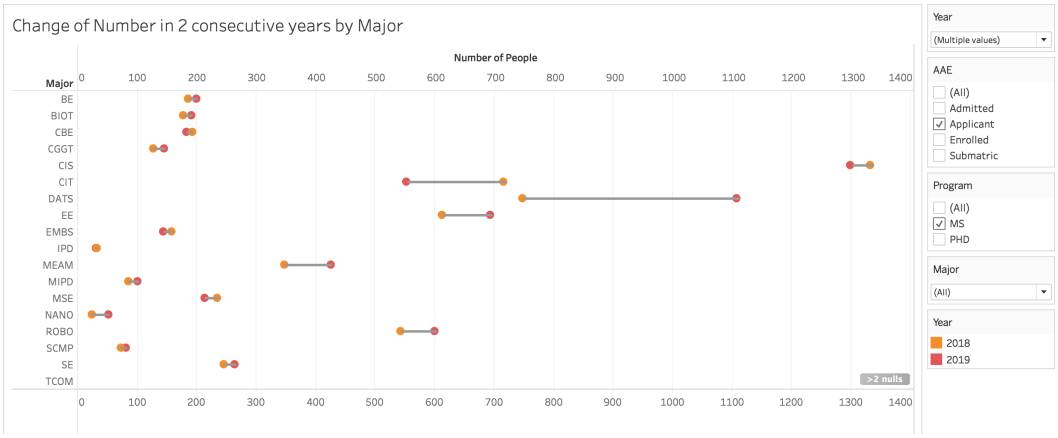

Figure 1: 2019 versus 2018 applicant numbers by major

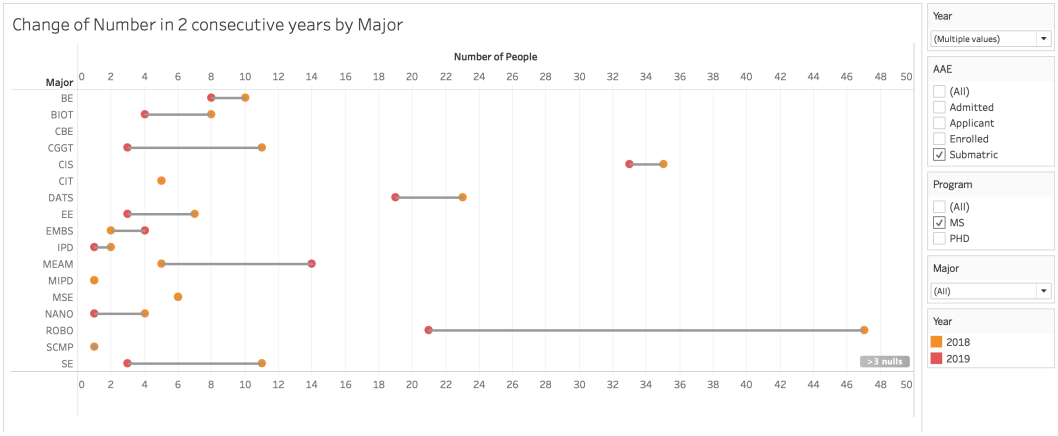

Figure 2: 2019 versus 2018 submatriculation numbers by major

3.2 Submatriculation Enrollment

The dashboard I created containing graphs on general admissions and submatriculation data is given in `Admission Statistics Submat Pred.twbx`. This file draws from the `FINAL` tab from `Tableau_Input.xlsx`. The first two worksheets of `Admission Statistics Submat Pred.twbx` are `Enrolled Breakdown` and `Enrolled Breakdown Total`. They visualize the attribution and breakdown of overall master's student enrollment from external applicants and submatriculants. Figures 3 and 4 show these two worksheets.

The worksheets `Submatric Pred` and `Total Submatric Pred`, which are given by Figures 5 and 6 respectively, select and display the submatriculant series from Figures 3 and 4 in order to better visualize the scale of the temporal changes.

The last element is `Submatric Pred Summary` and is shown in Figure 7. This table summarizes each of the series from Figure 5 in order to simultaneously compare across majors.

3.3 One Deadline Versus Two Deadlines

I created the dashboard `Admission Statistics - 2DDLv1DDL.twbx` as the analog to `Admission Statistics - 2DDL.twbx` in order to handle the logic for department admissions switching from having two deadlines to one deadline after 2018. This Tableau file draws from the `2DDL` tab from `Tableau_Input.xlsx` and allows for the comparison of multiple deadline types. The first two worksheets, `Ad Stats` and `Ad Rate`

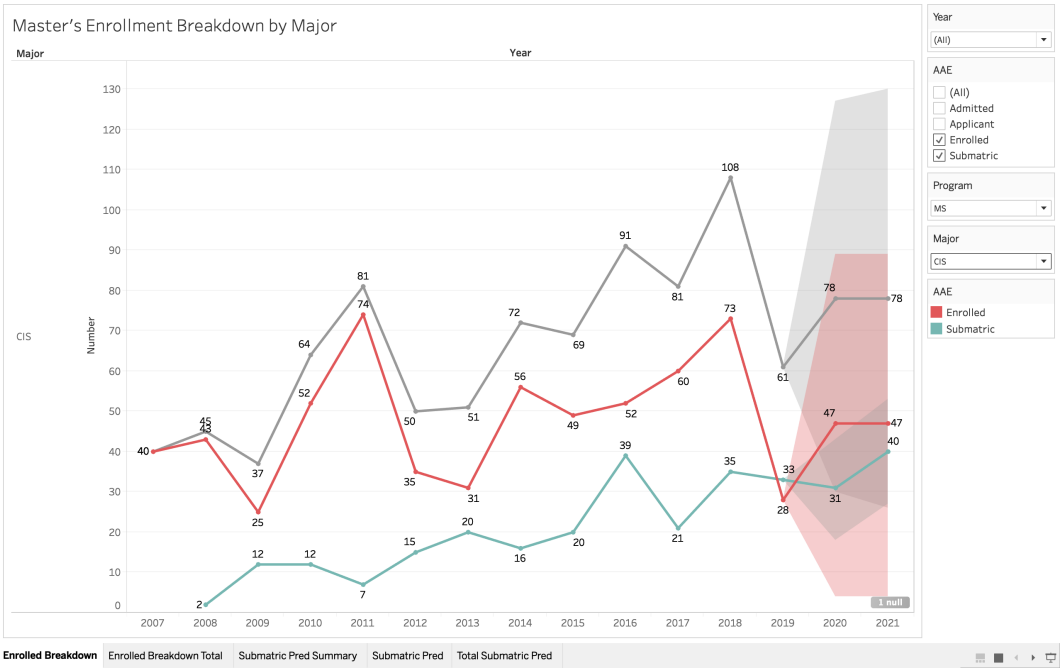

Figure 3: Master's enrollment breakdown by major

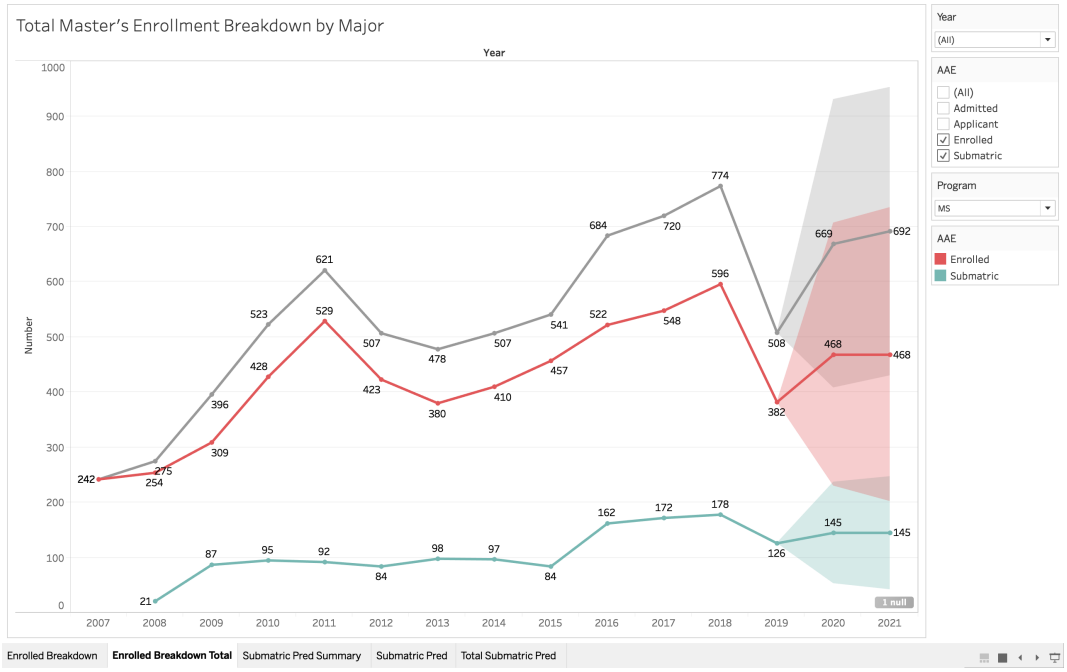

Figure 4: Total master's enrollment breakdown by major

`& Enrolled Rate` tabulate all of the admissions statistics later visualized in the rest of the dashboard. For example, `Ad Rate & Enrolled Rate`, given in Figure 8 displays the color coded admissions and enrollment rates across each year and major for the relevant deadline structure.

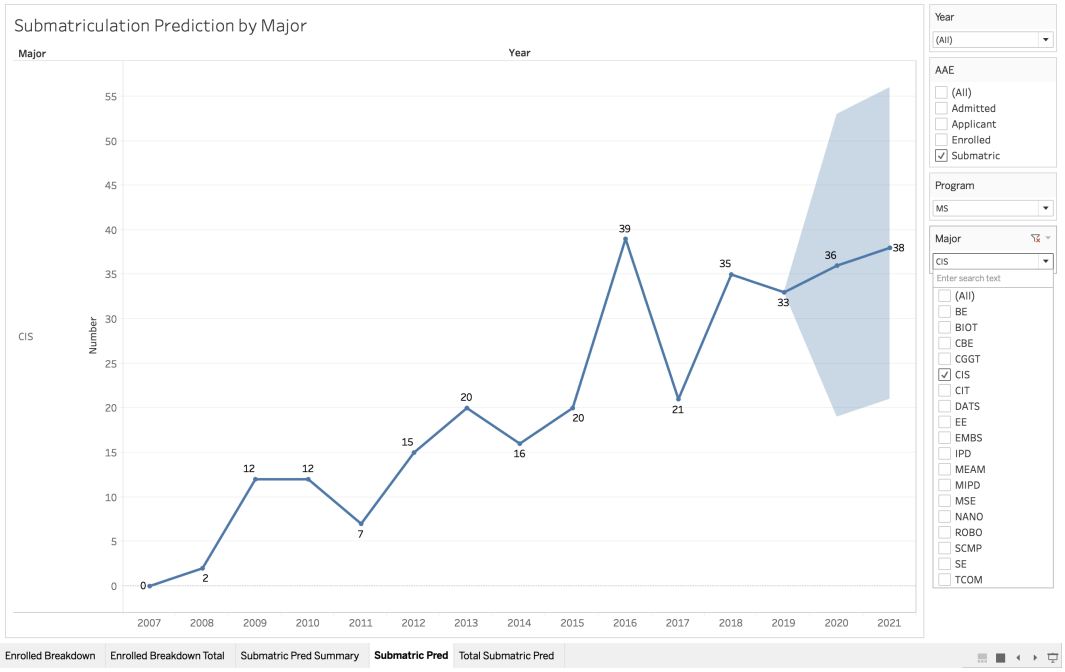

Figure 5: Submatriculation prediction by major

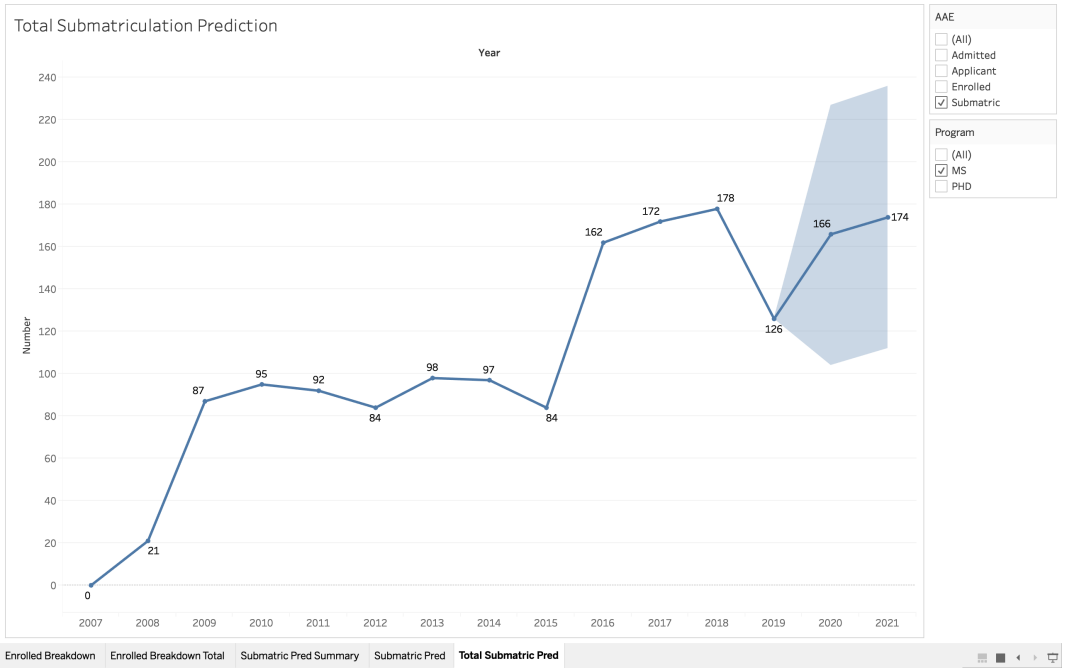

Figure 6: Total submatriculation prediction

This dashboard also contains worksheets that directly compare how each department fares as they shifted from two deadlines in 2018 to one deadline in 2019. `Applicants 2DDL` juxtaposes the number of applicants and enrolled students for each deadline in 2018 against that of 2019. This is shown in Figure 9. Note that

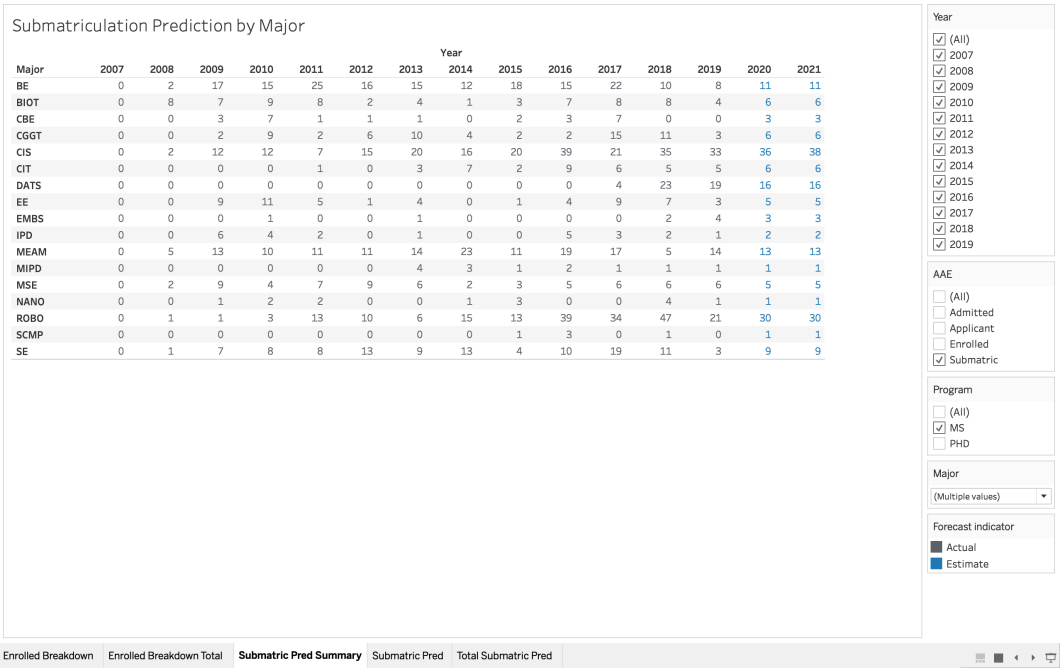

Figure 7: Submatriculation prediction by major

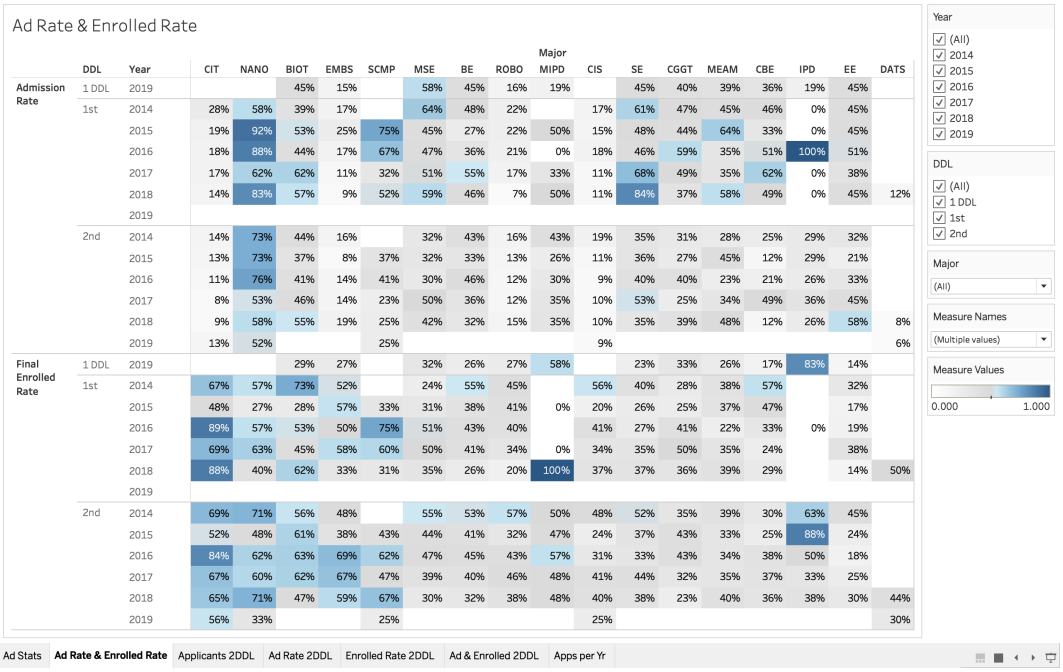

Figure 8: Admissions rates and enrollment rate between deadlines

CIS, DATS CIT, SCMP, and NANO retained the two deadline structure in 2019.

`Ad Rate 2DDL` and `Enrolled Rate 2DDL` are similar but instead of comparing levels, these sheets represent the admissions rate and enrollment rate, respectively in two separate graphs to highlight the

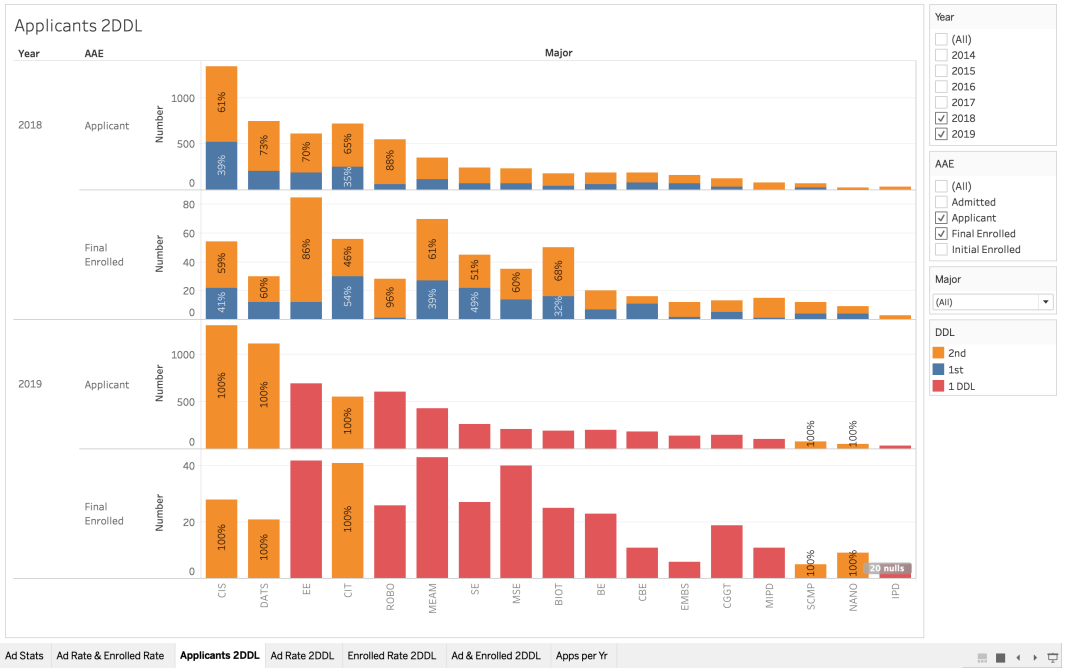

Figure 9: Applicant and enrollment numbers between 2019 and 2018

magnitude of change from the different deadline types, as shown by Figures 10 and 11.

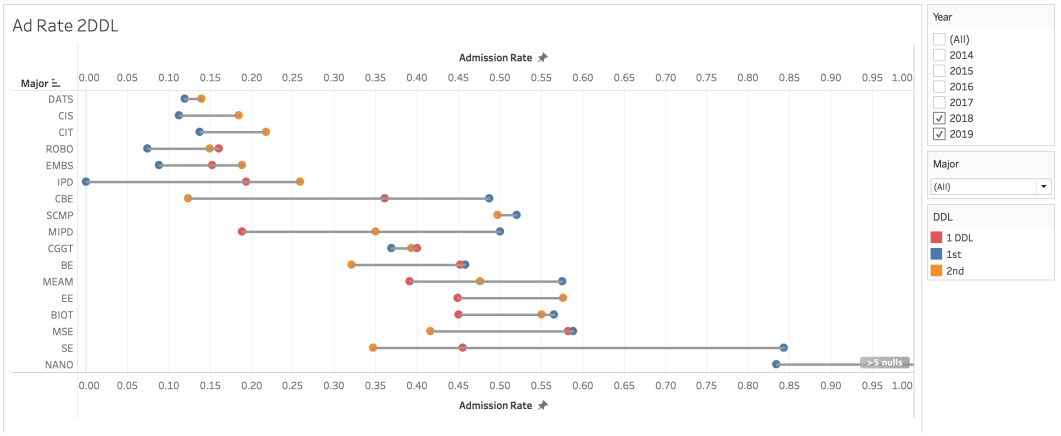

Figure 10: Admissions rate between two deadlines versus one deadline

`Ad & Enrolled 2DDL` in Figure 12 combines the information from Figures 9, 10, and 11 in order to visualize at the same time the impact of switching deadline types on applicant numbers, admissions rates, and enrollment rates.

The last worksheet is `Apps per Yr`, shown in Figure 13. This visualization shows the total number of applications received each year for each relevant deadline. The numbers are further broken down by the number of applications for each major.

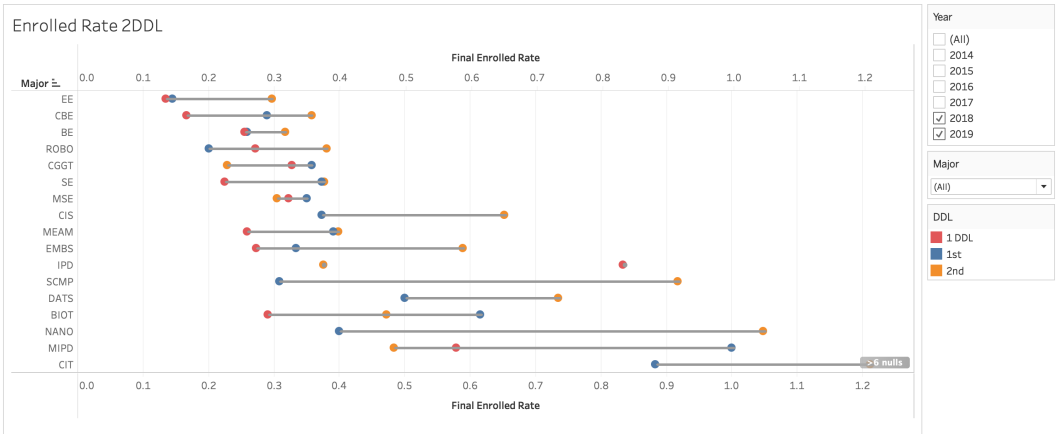

Figure 11: Enrollment rate between two deadlines versus one deadline

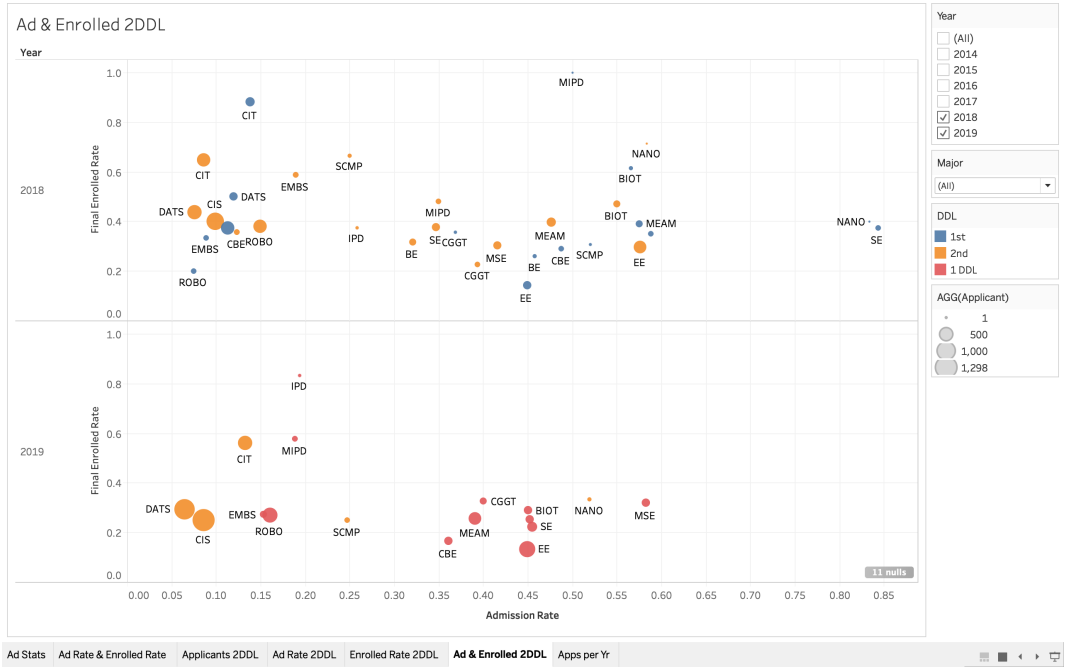

Figure 12: Applicant and enrollment numbers between 2019 and 2018, denoted by size

4 Discussion

Visualizing past admissions data yields many interesting insights. For example, from Figure 4, we can see that the rise in overall master's enrollment can be attributed to increases in both external students and submatriculants. However, while the number of submatriculants have increased steadily, there was a sharp dip in external student enrollment in 2012 and 2013. This drop was large enough to significantly decrease the total enrollment number. Another interesting trend is from the introduction of the DATS program which began accepting applications in 2018. Since its inception, the DATS program has already become the second most popular major at SEAS in terms of applications received. And if this pattern continues, DATS will surpass CIS next year as the leader in applications received. Watching the DATS program's rapid development relative to other, more established majors will be interesting to observe in the coming years.

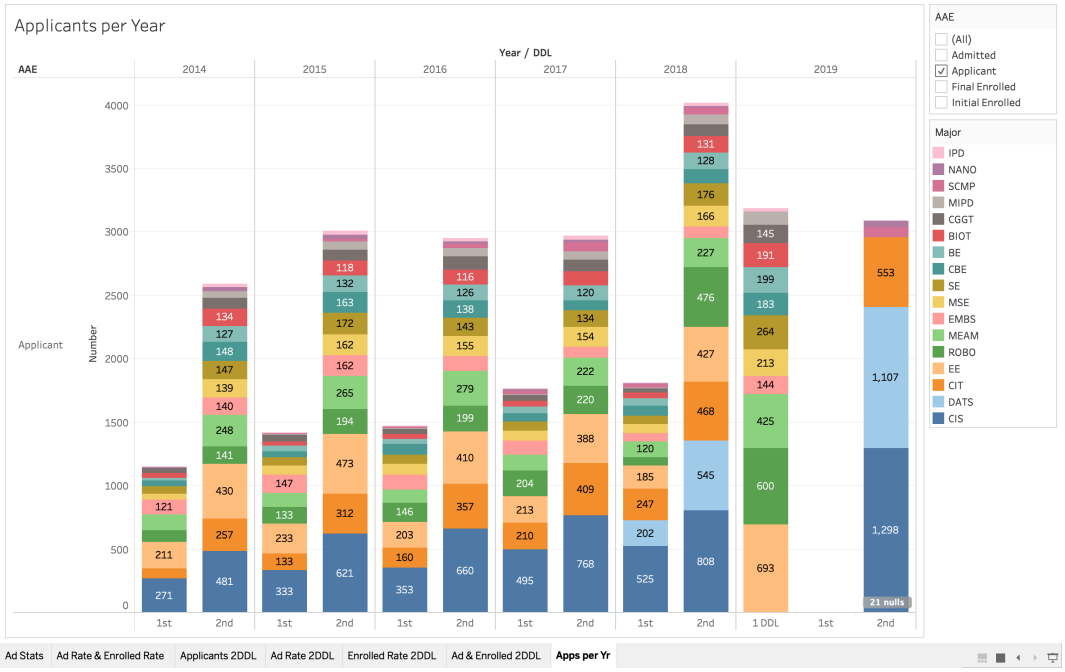

Figure 13: Applicants each year for each relevant deadline

As mentioned, it is also of interest to examine the impact of moving from two deadlines to one deadline. From initial analysis, it seems that the deadline policy switch has not adversely affected applicant numbers, although the largest departments such as CIS still use two deadlines. There is also no clear pattern of change in admissions rate from this shock and it is still too early to examine enrollment rates as not all decisions have been received yet.

Other interesting results came from validating prediction accuracy, which are summarized in Tables 3, 4, 5, and 6. We can see that generally Tableau's exponential smoothing forecasting function works well and regularly acheives a low average MAPE. There are a couple outliers though. First note from Table 5 that since DATS is such a new program and therefore has many zero-padded values, its forecasts are inaccurate and severely under-predicted. And since not all enrollment results have been received yet, the 2019 enrollment yield in Table 6 is systematically overestimated by the forecast. The 2019 application predictions in Table 5 are on average less accurate than those of 2018 in Table 3 which may be caused by the admissions deadline policy change during that period. Lastly, for example from Table 5 we can see a significant over-prediction of CIT applications, indicating that the recently launched online MCIT program has partially cannibalized applications for the on campus version.

5 Usage Instructions

The objective of this practicum was to not only build a developed visualization tool but to also make it accessible to the relevant stakeholders. To facilitate this, I created the file, `README` which documents how to use these dashboards in detail. Using screenshots and step-by-step instructions, I mapped how out to update and change the visualizations and even how to create new ones.

Major	Actual	Predicted	MAPE
BE	186	192	3.2%
BIOT	177	162	8.5%
CBE	192	174	9.4%
CGGT	127	162	27.6%
CIS	1332	1293	2.9%
CIT	715	804	12.4%
EE	612	623	1.8%
EMBS	158	228	44.3%
IPD	32	27	15.6%
MEAM	347	410	18.2%
MIPD	85	67	21.2%
MSE	234	261	11.5%
NANO	24	34	41.7%
ROBO	543	535	1.5%
SE	246	246	0.0%
		Average MAPE	14.6%

Table 3: Validation of 2018 applicant number prediction

Major	Actual	Predicted	MAPE
BE	35.3%	50.3%	42.5%
BIOT	49.0%	55.9%	14.1%
CBE	28.8%	39.3%	36.2%
CGGT	28.6%	48.2%	68.7%
CIS	52.5%	40.9%	22.1%
CIT	60.8%	67.6%	11.2%
EE	26.1%	18.8%	28.1%
EMBS	60.9%	67.9%	11.6%
MEAM	45.2%	37.0%	18.1%
MIPD	43.3%	47.9%	10.5%
MSE	31.2%	41.3%	32.4%
NANO	58.8%	54.3%	7.7%
ROBO	55.3%	56.7%	2.6%
SE	37.5%	39.8%	6.1%
		Average MAPE	22.3%

Table 4: Validation of 2018 enrollment yield prediction

6 Future Work

Going forward, it would be of interest to observe how the impact of certain shocks play out. For example, there were many policy changes made recently on the rules of submatriculating. This visualization tool could be used to understand the impact of these changes starting next year. Additionally, Tables 7 and 8 provide predictions for the 2020 application cycle which can also be verified after the next cycle.

The choice to maintain the dashboard in Tableau was largely due to its clean interface and its interactiveness (i.e. being able to filter views). However, those unfamiliar with the software may experience a bit of a learning curve which the `README` file attempts to address. Even then, it may still be difficult to get buy in from the people who would truly benefit from this tool. As a result, it would be helpful to expand upon this project by allowing concise, subsections of these files to be reported to relevant individuals so that they

Major	Actual	Predicted	MAPE
BE	199	200	0.5%
BIOT	191	164	14.1%
CBE	183	230	25.7%
CGGT	145	150	3.4%
CIS	1298	1405	8.2%
CIT	553	802	45.0%
DATS	1107	37	96.7%
EE	693	617	11.0%
EMBS	144	193	34.0%
IPD	31	38	22.6%
MEAM	425	400	5.9%
MIPD	101	91	9.9%
MSE	213	289	35.7%
NANO	52	29	44.2%
ROBO	600	655	9.2%
SCMP	81	70	13.6%
SE	264	267	1.1%
		Average MAPE w/ DATS	22.4%
		Average MAPE w/o DATS	17.8%

Table 5: Validation of 2019 applicant number prediction

Major	Actual	Predicted	MAPE
BE	25.6%	49.6%	94.3%
BIOT	29.1%	55.9%	92.4%
CBE	16.7%	39.3%	136.1%
CGGT	32.8%	48.2%	47.2%
CIS	25.2%	47.5%	88.2%
CIT	56.2%	63.5%	13.0%
DATS	29.6%	3.1%	89.6%
EE	13.5%	16.4%	21.2%
EMBS	27.3%	72.6%	166.1%
IPD	83.3%	39.3%	52.9%
MEAM	25.9%	41.1%	58.7%
MIPD	57.9%	46.7%	19.3%
MSE	32.3%	41.3%	27.9%
NANO	33.3%	54.8%	64.5%
ROBO	27.1%	56.0%	106.8%
SCMP	25.0%	41.3%	65.4%
SE	22.5%	30.9%	37.5%
		Average MAPE	69.5%

Table 6: Validation of 2019 enrollment yield prediction

can more easily digest this material.

There are also a couple improvements that can be made regarding the dashboard itself. While I was able to include submatriculants into this project, transfer students are still excluded. Adding them in would be a logical next step. Furthermore, zero-padding null data points leads to under-prediction for shorter time

Major	Predicted
BE	208
BIOT	168
CBE	183
CGGT	154
CIS	1426
CIT	669
DATS	740
EE	655
EMBS	169
IPD	37
MEAM	446
MIPD	105
MSE	248
NANO	40
ROBO	748
SCMP	75
SE	284

Table 7: Prediction for 2020 applicant numbers

Major	Predicted
BE	42.8%
BIOT	55.9%
CBE	33.7%
CGGT	48.2%
CIS	37.5%
CIT	60.8%
DATS	29.6%
EE	13.4%
EMBS	45.8%
IPD	61.3%
MEAM	34.6%
MIPD	67.5%
MSE	41.3%
NANO	48.3%
ROBO	43.6%
SCMP	33.2%
SE	29.9%

Table 8: Prediction for 2020 enrollment yield

series. It would be ideal to be able to limit the amount of padding done to whatever is absolutely necessary to create a forecast.

The hope is that this visualization tool will be able to help the SEAS community. By elegantly visualizing past admissions data and generating accurate forecasts in a compact setup, this practicum project attempts to alleviate some of the logistical headaches experienced by admissions.

7 Acknowledgments

Special thanks to Professor Boon Thau Loo, who served as the advisor to this DATS 599 Practicum.