MWSUG 2019 | PAPER BL-076

Why Admitted Students Opt-out of College Enrollment?

Manasi Murde, Oklahoma State University, Stillwater, Oklahoma

ABSTRACT

Understanding and improving student enrollment has always been important for colleges across universities. It is imperative for universities to have a firm grip on addressing enrollment of prospects, applicants and admitted students. As per research, one out of every three students who are admitted do not enroll in that college. One of the prime reasons why universities must concentrate on increasing their enrollment rate is to avoid loss of revenue. Admitted but non-enrolled students have a negative financial impact on a university due to the loss of tuition fees while expending money on marketing and recruitment. Non-enrollment of students also results in loss of academic potential.

There can be numerous reasons why admitted students fail to enroll within a particular university, which may be specific to one university or common among many schools. Therefore, each school must identify any existing trends within its own unique admitted non-enrolled student population base.

This research concentrates on analyzing data to identify various possible factors and indicators why undergraduate admitted students fail to enroll at Oklahoma State University despite getting an admission. The data for this analysis includes email communications, text messages, demographic, and admission process timestamps of students admitted for Fall 2018. The duration of the data recorded extends through November 2016 to June 2018. All data were obtained from the Marketing and Communications Department within the Undergraduate Admissions office and were de-identified.

Results from this research will help the OSU Marketing and Communications Department by understanding non-enrollment rate and factors that lead to non-enrollment.

INTRODUCTION

Summer melt refers to the phenomenon in which students show an interest in applying and joining a college at the end of high-school, but do not matriculate at a college the following fall.

While summer melt is used to explain why college-intending students fail to enroll in college **at all** in the fall, this research concentrates on why admitted candidates for undergraduate studies at Oklahoma State university fail to enroll.

In this research, the data pertaining to admitted students for undergraduate studies obtained from the Marketing and Communications Department of Oklahoma State University would be analyzed to understand the trends between enrolled versus not enrolled students. It could be that the students who have not enrolled at OSU may have enrolled in some other college. Indicators that highlight the non-enrollment of admitted students will help the department understand how to improve the specific admission process to convert admitted students into enrolled students.

For this research, initial exploratory analysis in SAS Studio 3.71 would be performed on the variables to understand a pattern and distribution of data for admitted students across the several categories. SAS Enterprise Miner 14.3 would be used to run Logistic Regression to identify significant factors affecting the non-enrollment rate of admitted students.

DATA DESCRIPTION

Data Source: Undergraduate Marketing and Communications Department, Oklahoma State University
Data Description: 49 Variables, 10,424 Observations

Every data analysis project begins with gathering and studying the data to be analyzed. Collecting relevant and efficient data is essential for any data analysis project as it establishes a strong foundation for the insights and trends of analysis. Below is a cross-tab of all the variables and description of each variable received.

Sr. No.	Variable Name	Description
1.	UID	Unique Identifier for student
2.	Active City	Name of city of student
3.	Active Region	Name of State of student
4.	Active Postal	Postal Code of student
5.	Gender	Gender of the student (M/F)
6.	Race	Race of the student
7.	Hispanic	Hispanic or not (Yes/No)
8.	First Generation	First in family to go to college (Yes/No)
9.	OSU Legacy	Legacy of OSU (0/1)
10.	OK Promise Eligible	Eligible for OK Promise scholarship or not (Yes/No)
11.	Academic Talent	Excellent Academic Records (Yes/No)
12.	Student Status	Status of student in Application Process
13.	Prospect Date	Date student changed to prospect status
14.	Inquiry Date	Date student changed to inquiry status
15.	Applicant Date	Date Student changed to an applicant
16.	Record Creation date	Date the student record was created
17.	Application Submit Date	Date student submitted application
18.	Admission Date	Date student was admitted
19.	Admitted	Admitted or not (Yes/No)
20.	Enrolled	Enrolled or not (Yes/No)
21.	Orientation Registration Date	Date student registered for Orientation
22.	Housing Application Date	Date student applied for housing
23.	Most Recent Campus Tour Registration	Date of most recent campus tour
24.	Most Recent On-Campus Event Registration	Date if most recent on-campus event registered
25.	Opt-Out Date	Date student opted out from receiving emails
26.	Interaction Date	Last interaction date with student
27.	Event Mail	Date of last event mail sent to student
28.	Social Medial Email	Date of last social media email sent to student
29.	Why OSU Email	Date of “Why OSU” email sent to the student
30.	Apply Email	Date of “Apply to OSU” email sent to the student
31.	Senior Visit Email	Date of Senior Visit email sent o the student
32.	Residential Life Email	Date of Residential Life email sent to the student
33.	Viewbook Email	Date of viewbook email sent to the student
34.	Congratulations Video Email	Date of “Congratulations Video” sent to student
35.	Next Steps #1 Email	Date of first Next Steps to take email sent
36.	Next Steps #2 Email	Date of second Next Steps to take email sent
37.	Next Steps #3 Email	Date of third Next Steps to take email sent

38.	Next Steps #4 Email	Date of fourth Next Steps to take email sent
39.	% Delivered	Percent of emails delivered
40.	% Opened	Percent of emails opened
41.	% Clicked	Percent of emails clicked
42.	Total Delivered	Total number of emails delivered
43.	Total Opened	Total number of emails opened
44.	Total Clicked	Total number of emails clicked
45.	Admission Cancellation Email	Date student cancelled admission
46.	Most Recent Email Opened Subject	Subject of most recent email opened
47.	Most Recent Email Opened Date	Date of most recent email opened
48.	Interactions – Comma Separated	Sequence of email interaction to student
49.	Received Scholarship	Received scholarship or not (Yes/No)

METHODLOGY

DATA CLEANING
In this step data will be cleaned by recoding, imputing missing values and creating variables required for analysis.

EXPLORATORY ANALYSIS:
Initial analysis using descriptive statistics will be performed to understand the distribution of student data across enrolled and non-enrolled students.

FEATURE IMPORTANCE USING LOGISTIC REGRESSION:
Logistic Regression with stepwise selection will be used to identify significant factors influencing the non-enrollment of students.

DATA CLEANING

Data cleaning was performed on the variables provided to maintain a consistent format and relevant set of values throughout.

- Since this research is associated with admitted students who enrolled or did not enroll, 145 student records who were not admitted and were included in this data set were removed. Hence the resulting data set had 10,279 records of admitted students.
- Active Region had more than 50 states (within USA and outside USA states). Based on the frequency top 8 states were renamed from their abbreviation to full form, and the rest of the states were grouped into one category "Others". The top 8 States that were renamed were Oklahoma, Texas, Arkansas, Missouri, California, Colorado, Illinois, Kansas.
- The column Race had more than 10 categories which had redundant values across several categories. These categories were renamed to include similar races into one category. For example, Asian white and Asian, Native Hawaiian are grouped into one category has Asian. The resulting categories for Race variable are "American Indian or Alaska Native", "Asian", "Black or African American", "Native Hawaiian or Other Pacific" and "White".
- OSU Legacy variable indicates whether the student was an OSU Legacy, that is, one or more family members who are an alum of OSU. The initial values of this column were 0, 1, yes or no. The values 1 were replaced with "Yes" and values 0 were replaced with "No".
- Enrolled variable was renamed to "Not Enrolled" which is the target variable and the initial values of Enrolled Yes – were recoded to 0 and No to 1, in order to address the problem statement of factors affecting non-enrollment of admitted students.
- An additional column was created "Days to Admit" which is the difference between the application submission date and the admit date.
- The variables that had more than 50% of missing data were dropped out of analysis. Most of the variables had at least 60% - 70% of missing data.
- The variables Most Recent Email Opened and Interactions – Comma separated are excluded from this analysis and would be used for future text analysis.
- Total of 14 Variables were used for analysis: UID, Active Region, Gender, Race, Hispanic, First Generation, OSU Legacy, Academic Talent, Scholarship, OK Promise Eligibility, Application submission Date, Admission Date, Days to Admit, Not Enrolled (target variable).
- These variables were grouped into further 4 categories:

Demographics	Academics	Family Background	Important Date Variables
Active Region Gender Race Hispanic	Academic Talent Scholarship OK Promise Eligibility	First Generation OSU Legacy	Application Submission Admission Date Days to Admit

- Data imputation node in SAS EM was used to impute missing values in variables Hispanic, First Generation, OSU Legacy, Received Scholarship, days to admit. Mean value for interval and count for categorical variables was used to impute the missing values.

EXPLORATORY ANALYSIS

1) Out of the 10,279 admitted students for fall 2018, 60% did not enroll while only 40% enrolled.

Demographics

2) 61% of admitted female students did not enroll for college, while 60% of admitted male students did not enroll for college. [Refer Figure 1 in Appendix A]

3) 74% of admitted students from other states or regions did not enroll for college, 70% of admitted students from Texas did not enroll for college, while 51% of admitted students from Oklahoma did not enroll for college. [Refer Figure 2 in Appendix A]

4) 60% of admitted White students did not enroll for college, while 63% of admitted students of Other Races did not enroll for college. [Refer Figure 3 in Appendix A]

5) 67% of admitted Hispanic students did not enroll to the college, while 60% of admitted Non-Hispanic students did not enroll to the college. [Refer Figure 4 in Appendix A]

Academics

6) 62% of admitted students with Academic Talent opted out of enrollment, while 60% of admitted students with No Academic Talent opted out of enrollment. [Refer Figure 5 in Appendix A]

7) 43% of admitted students with scholarships opted out of college enrollment, while 80% of admitted students with no scholarships opted out of college enrollment. [Refer Figure 6 in Appendix A]

8) 49% of admitted OK Promise eligible students opted out of college enrollment, while 62% of admitted OK Promise not eligible students opted out of college enrollment. [Refer Figure 7 in Appendix A]

Family Background

9) 62% of admitted first generation students did not enroll to the college, while 60% of admitted non-first-generation students did not enroll to the college. [Refer Figure 8 in Appendix A]

10) 45% of admitted OSU Legacy students opted out of college enrollment while 65% of admitted non-OSU Legacy students opted out of college enrollment. [Refer Figure 9 in Appendix A]

Important Dates

11)65% of admitted students are admitted during August to November, highest being in November.

12) 76% of the students who cancelled their admission, opted-out of college enrollment during March to May, highest being in the month of May.

13) 60% of the admitted students received admits within one month of application.

ANALYSIS OF LOGISTIC REGRESSION

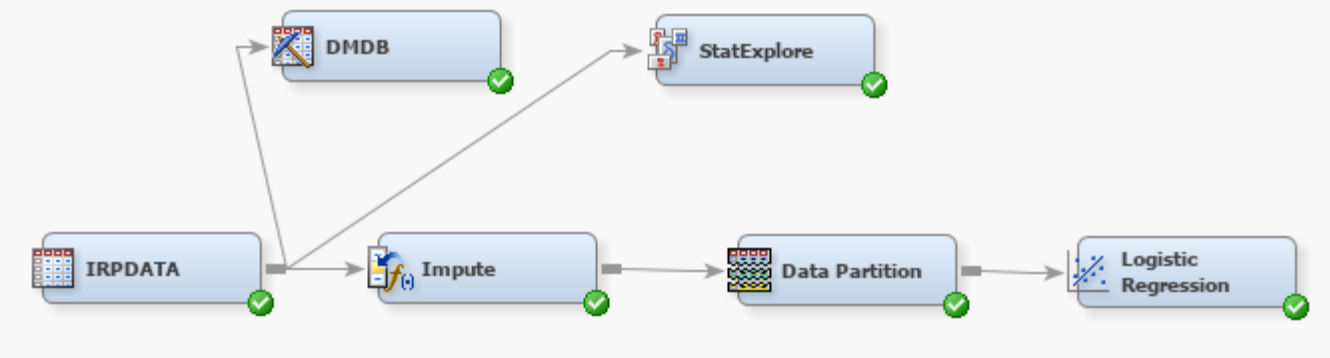


Figure: Logistic Regression in SAS Enterprise Miner 14.3

Logistic regression was used to identify the factors influencing non-enrollment of admitted students. The target variable used was Not Enrolled (1 for students who did not enroll and 0 for students who enrolled).
After the missing values were imputed using the impute node, data was partitioned into 70%-30% ratio for training and validation. Stepwise selection logistic regression with significance level of 0.5 was used to identify the significant factors influencing non-enrollment rate. 9 nominal variables and 1 interval variable were input to logistic regression. [Refer to Figures in Appendix B].

The model ran with a misclassification rate of 29% (Accuracy of 71%) [Refer Figure 3 Appendix B].
Top 4 selected variables were Received Scholarship, Active Region, Academic Talent and OSU Legacy. These variables were significant as per the Maximum likelihood Estimates and Type 3 Analysis of Effects. [Refer to Figure 1 & Figure 2 in Appendix B]
The odds ratio estimates interpretation is as follows: [Refer to Figure 2 in Appendix B]
Students with no scholarship are 86% more likely to not enroll as compared to students with scholarship.
Students from Other States are 55% more likely to not enroll as compared to students from Texas and Students from Texas are 69% more likely to not enroll as compared to students from Oklahoma.
Students with Academic Talent are 65% more likely to not enroll as compared to students with No Academic talent.

CONCLUSION

An article published in 2016, "Top 10 reasons you should attend Oklahoma State University", by Karolyn Bolay, listed great educational value, undergraduate research opportunities, OSU career services, faculty interaction, variety of academic options, great residential life options, great culture, on- campus job opportunities, athletics, and America's greatest homecoming as some of the top 10 reasons to join OSU.
However, the undergraduate enrollment rate at Oklahoma State University was only 40% (60% non-enrollment rate). Based on the research and analysis performed on the data provided by the department, several insights were collected. Based on the factors that influence low enrollment rate below conclusions can be drawn:

Firstly, scholarship plays an important role in a student's education life. Students with no scholarship are more likely to opt out of enrollment as compared to students with scholarship. Getting a scholarship, is very crucial factor in the decision-making process of a student. While the university and undergraduate college could have limited financial aid to provide to the students, it is imperative that the university takes appropriate financial measures to ensure that scholarship can be granted to academically talented students as well as most deserving students.
Students who do not end up getting a scholarship or are not eligible for OK Promise either, may find themselves not able to enroll due to financial reasons.

Secondly, Students with Academic Talent are more likely to opt of enrollment as compared to students with no Academic Talent. One of the reasons could be that students who have high SAT and ACT scores would be hoping to get scholarships and if scholarships were not received, these students would seek out next closest universities where their scores would fetch them a scholarship
Ensuring that the Academic Talent student enroll to the university is beneficial for university to increase the quality of student pool.

Thirdly, students from other states/regions and Texas are more likely to opt-out of college enrollment as compared to students from Oklahoma. One of the reasons that the number of student enrollment is low from outside of Oklahoma could be the proximity of the University from their hometowns. Students prefer to colleges that are closer to their hometown from where they could visit their parents without having to spend a lot on travelling. A reason for students from Texas not enrolling as compared to students from Oklahoma could be the availability and accessibility to competitively good universities such as OSU within Texas.

Many other factors could contribute to the non-enrollment rate. Like the income of the families of students that do not enroll or the communication between the students and department prior to the decision of non-enrollment or students with academic talent have not received sufficient amount of financial aid or scholarship.

Further research can be performed by building predictive models using these factors to predict the students that do not enroll for the college. If email communication data could be made available in future, text analytics can be performed to analyze the communication between students and the college to detect the chances of non-enrollment of a student.

REFERENCES

Benjamin L. CastlemanLindsay C. PageAshley L. Snowdon. Harvard University
https://sdp.cepr.harvard.edu/files/cepr-sdp/files/sdp-summer-melt-handbook.pdf

Alysha Rashid, E-Adviser, College Advising Corps Published in April 2015
https://www.mdrc.org/sites/default/files/Issue%20Focus-CMP_SummerMelt.pdf

The Hidden Brain Podcast, NPR. June 2018
https://www.npr.org/2018/06/18/621023537/summer-melt-why-arent-students-showing-up-for-college

Karolyn Bolay. Blog Published on August 19, 2016
https://newsok.com/article/5513885/top-10-reasons-you-should-attend-oklahoma-state-university

ACKNOWLEDGEMENTS

I am thankful to the undergraduate **Marketing and Communications Department at Oklahoma State University** for giving me this opportunity to work on the data set and providing timely inputs. I would also like to thank my professors **Dr. Goutam Chakraborty** and **Dr. Miriam McGaugh** for their constant support and guidance throughout the project.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Manasi Murde
Graduate Student | Business Analytics
Oklahoma State University
405-612-3156
manasi.murde@okstate.edu

APPENDIX A
Exploratory Analysis

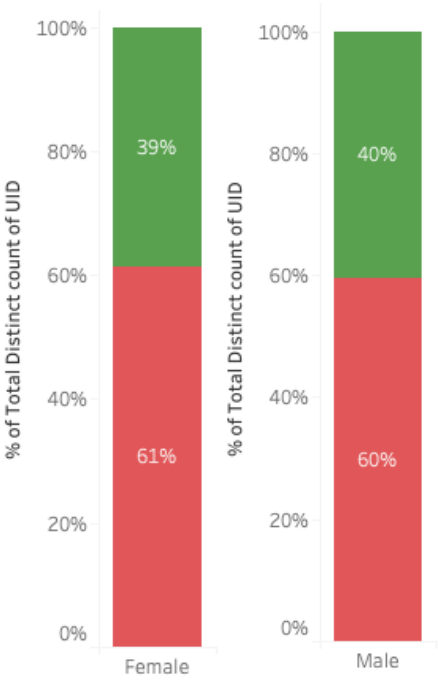


Figure 1: Enrolled vs Not Enrolled Students for Gender

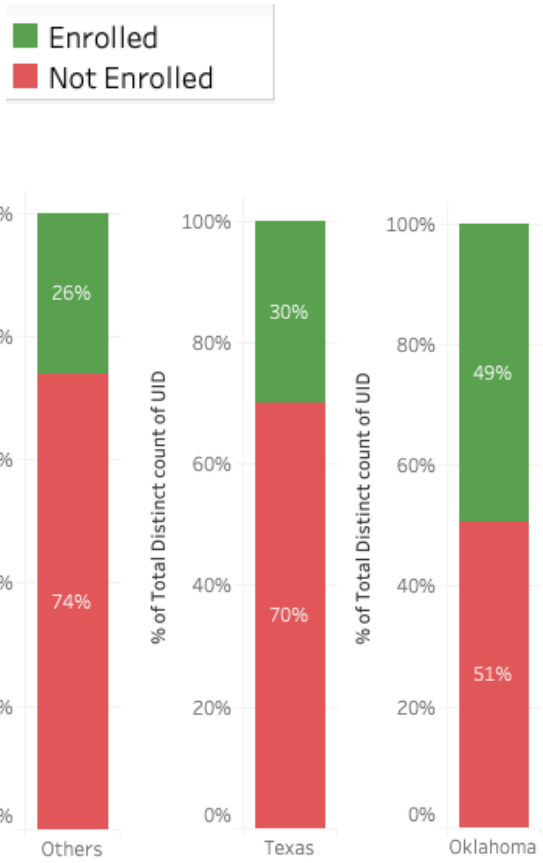


Figure 2: Enrolled vs Not Enrolled Students for States (Active Regions)

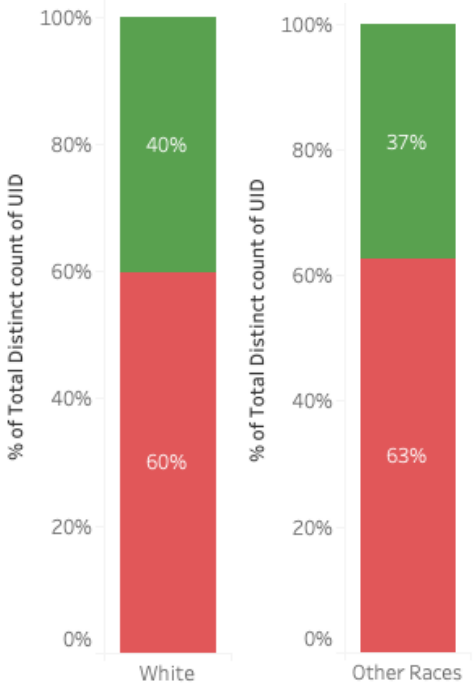
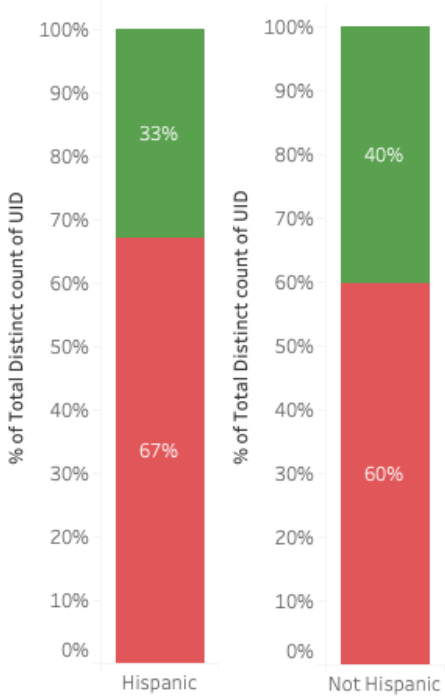


Figure 3: Enrolled vs Not Enrolled Students for Race



Graph 4: Enrolled vs Not Enrolled Students for Hispanic

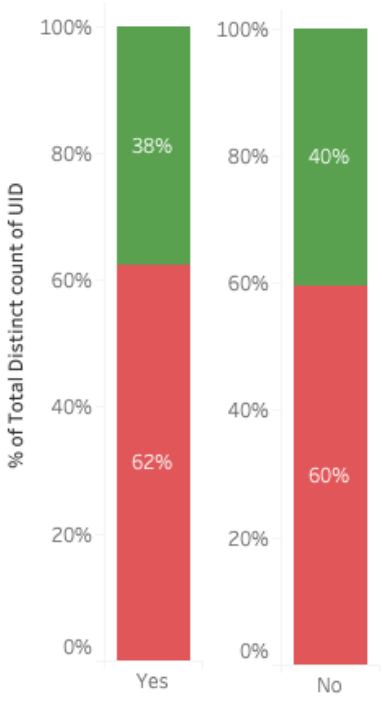


Figure 5: Enrolled vs Not Enrolled Students for Academic Talent

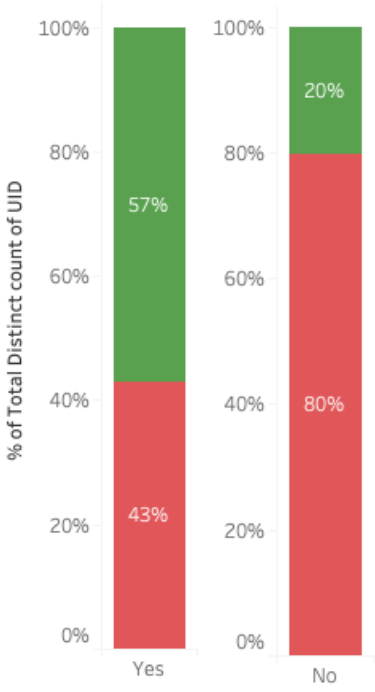


Figure 6: Enrolled vs Not Enrolled Students for Scholarship Received

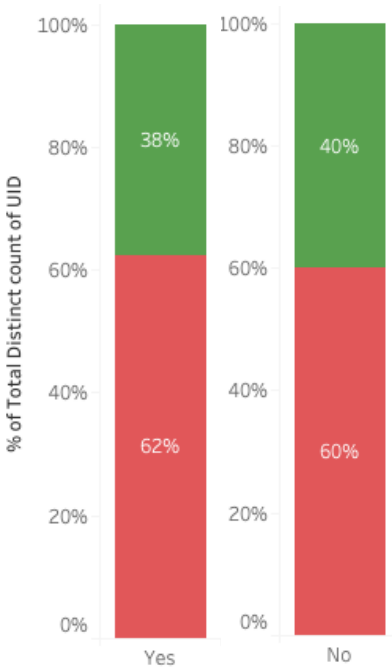


Figure 8: Enrolled vs Not Enrolled Students for First Generation

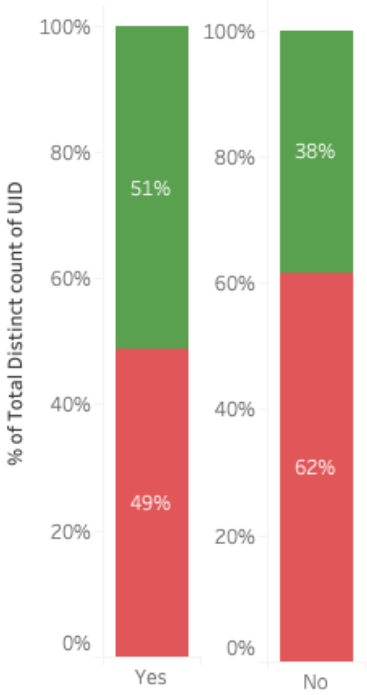


Figure 7: Enrolled vs Not Enrolled Students for OK Promise Eligible

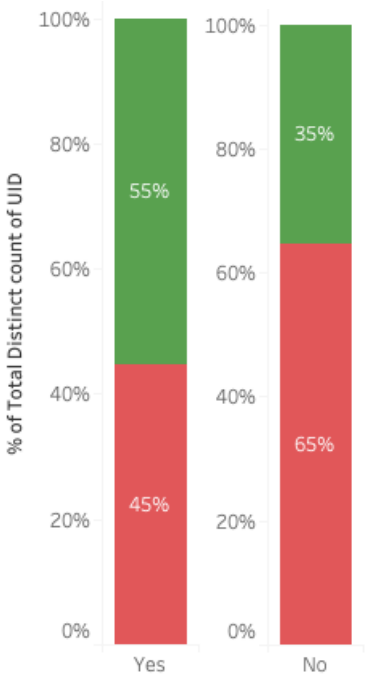


Figure 9: Enrolled vs Not Enrolled Students for OSU Legacy

APPENDIX B
Logistic Regression Analysis

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	IMP_Received_Scholarship		1	1	983.4400		<.0001	0.3182
2	Active_Region		2	2	290.8886		<.0001	0.2959
3	Academic_Talent		1	3	100.5838		<.0001	0.2916
4	IMP_OSU_Legacy		1	4	113.4120		<.0001	0.2939
5	Race		1	5	21.5338		<.0001	0.2981
6	OK_Promise_Eligible		1	6	11.2427		0.0008	0.2997
7	IMP_Days_to_Admit		414	7	510.4391		0.0008	0.3202
8		IMP_Days_to_Admit	414	6		368.9729	0.9454	0.2997

The selected model, based on the misclassification rate for the validation data, is the model trained in Step 3. It consists of the following effects:

Intercept Academic_Talent Active_Region IMP_Received_Scholarship

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Intercept Only	Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
9653.534	8239.393	1414.1402	4	<.0001

Figure 1: Summary of Stepwise Selection

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Academic_Talent	1	99.6437	<.0001
Active_Region	2	256.3045	<.0001
IMP_Received_Scholarship	1	953.6064	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	0.8568	0.0337	645.83	<.0001	2.356
Academic_Talent	No	1	-0.3015	0.0302	99.64	<.0001	0.740
Active_Region	Oklahoma	1	-0.5914	0.0371	253.83	<.0001	0.554
Active_Region	Others	1	0.3942	0.0502	61.60	<.0001	1.483
IMP_Received_Scholarship	No	1	0.8997	0.0291	953.61	<.0001	2.459

Odds Ratio Estimates

Effect		Point Estimate
Academic_Talent	No vs Yes	0.547
Active_Region	Oklahoma vs Texas	0.454
Active_Region	Others vs Texas	1.218
IMP_Received_Scholarship	No vs Yes	6.047

Figure 2: Type 3 Analysis, Maximum Likelihood Estimates, Odds Ratio

Fit Statistics

Target=Not_Enrolled Target Label=' '

Fit Statistics	Statistics Label	Train	Validation
AIC	Akaike's Information Criterion	8249.39	.
ASE	Average Squared Error	0.20	0.19
AVERR	Average Error Function	0.57	0.57
DFE	Degrees of Freedom for Error	7188.00	.
DFM	Model Degrees of Freedom	5.00	.
DFT	Total Degrees of Freedom	7193.00	.
DIV	Divisor for ASE	14386.00	6172.00
ERR	Error Function	8239.39	3510.11
FPE	Final Prediction Error	0.20	.
MAX	Maximum Absolute Error	0.92	0.91
MSE	Mean Square Error	0.20	0.19
NOBS	Sum of Frequencies	7193.00	3086.00
_NW_	Number of Estimate Weights	5.00	.
RASE	Root Average Sum of Squares	0.44	0.44
RFPE	Root Final Prediction Error	0.44	.
RMSE	Root Mean Squared Error	0.44	0.44
SBC	Schwarz's Bayesian Criterion	8283.80	.
SSE	Sum of Squared Errors	2813.34	1194.51
SUMW	Sum of Case Weights Times Freq	14386.00	6172.00
MISC	Misclassification Rate	0.30	0.29

Figure 3: Fit Statistics