Spring 2018

Admissions Events & Prospective Student Conversion

Joel Klein
St. Cloud State University

Admissions Events & Prospective Student Conversion

St. Cloud State University

Joel Klein

Spring 2018

Table of Contents

Executive Summary:

St. Cloud State University (SCSU) is the largest state university in Minnesota in terms of student enrollment. The main vision of St. Cloud State is *"Through active discovery, applied knowledge and creative interaction, we positively transform our students and the communities where they live and work."* In order to protect such a vision, St. Cloud State must continue to attract prospective new-entering first-year students to the institution. However, in the recent decade, enrollment numbers for St. Cloud State University have fallen marginally. As reported in the *Star Tribune* in March 2016, "[St. Cloud State] is dealing with falling enrollment and nagging deficits. SCSU enrollment stands at 15,461, down from 18,650 in the fall of 2010 … It is currently battling a $6 million budget gap." Admissions events play a tremendous role in attracting new incoming students and, in turn, reducing the budget gap.

This study's main focus is to analyze recent available admissions event data to determine which set of events are most effective in acquiring prospective incoming students. This research will further develop the SCSU Admissions Office's understanding of prospective student behaviors in efforts to more effectively budget available resources in attracting and targeting prospective students. In addition, through a full breakdown and analysis of SCSU admissions event and application records from 2016 to 2018, this report provides insight to predict the likelihood a prospective student will apply to and enroll in St. Cloud State University. This understanding will effectively add insight to university campaign advertising and marketing strategies to acquire higher levels of student applications and enrollments in the future.

The most important insights resulting from this analysis are:

- Although the number of prospects increased by almost 34,000 from the fall 2016 cycle to the fall 2017 cycle, the number of admissions event-attending prospects decreased by almost 1,400. There were also fewer prospects who applied (1,032 less), were admitted (504 less), and eventually enrolled (297 less).
- Attending an admissions event drastically increases the probability a student will apply (26 percentage points higher), get admitted (21 percentage points higher), and enroll (nine percentage points higher).
- 70 percent of those who attend a Preview Day apply. Other events with higher conversion rates are Discover Red & Black (59 percent) and Daily Campus Visit (58 percent).
- It is difficult to predict the probability a prospect will apply and enroll based on simply the number of and types of admissions events attended. More information must be known in order to develop a more accurate prediction model.

These insights will help guide and better influence important decisions regarding the continual improvement in enrollment levels at St. Cloud State University.

Data and Demographics:

There are two main data queries available for this analysis. The SCSU Department of Analytics and Institutional Research provided admissions event data which contains records showing the admissions event(s) a prospective student was invited to/registered for/attended along with application information such as if a prospect applied, when he/she applied, if he/she were admitted. This table also contains demographic information such as class rank and gender for the prospective students who applied. A second data query contained all students who were enrolled in the fall 2016 and fall 2017 semesters. The students whose ID's in this table mapped to the admissions event data were then identified as prospects who later enrolled. A full list of the information these data sources contain is listed in the appendix. These table's contents allow for analysis and solutions to the main research questions listed below. In total there are roughly 400,000 event records spanning from 2009 to 2018 for 299,975 prospective new-entering first-year students for fall term cycles 2016, 2017, and 2018. Throughout this report, these anticipated cohorts are referred to as 20173 (fall 2016), 20183 (fall 2017), and 20193 (fall 2018). All reporting is done at the student level, and not the event level meaning there is one record per prospect. A total of 1977 different admissions event names listed were categorized into 20 distinct event types of interest (listed in appendix). The main admissions event types of interest throughout the analysis are: Admitted Student Days, Daily Campus Visits, Discover Red & Black, High School Visits, Preview Days, Minnesota Education Fair, and Wisconsin Education Fair. The research attempts to determine how effective these events are at converting attending prospects.

Research Problems:

The SCSU Admissions Office is searching for insights to describe the prospective student populations as it relates to admissions event attendance. The major questions of interest addressed in this report are:

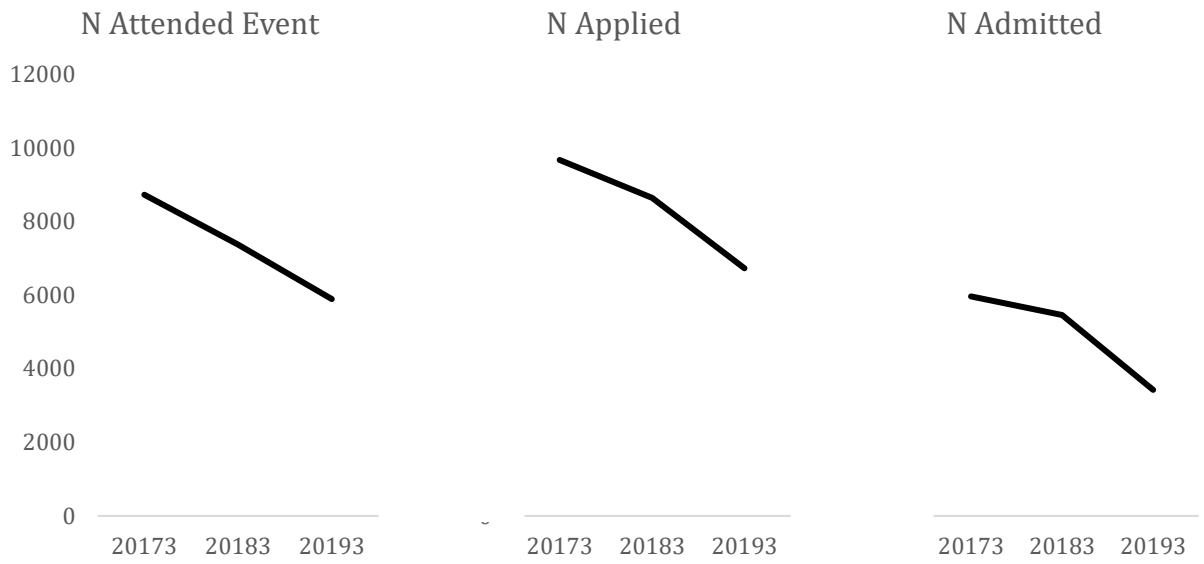
- Did we see fewer prospects along the way leading up to the two most recent cycles (fall 2017 and fall 2018) in comparison to the fall 2016 cycle?
- Which admissions events are most successful in acquiring prospects to apply? How much do conversion rates differ?
- What admissions events are more likely to gain admitted students? More likely to result in enrolled students?
- Is there any admissions event for the fall 2016 cohort that stands out as giving a high likelihood of a student eventually being both enrolled and retained?
- What is the best event order for producing the highest application conversion rates?
- Can a reliable predictive model be created to predict the probability a student will apply to SCSU with known information such as event attendance and demographic information?

Findings:

Table 1: *Number of Total Prospects, Event Attending Prospects, Applying Prospects, and Admitted Prospects by Anticipated Year-Term*

Anticipated Year/Term	N Prospects	N Attended Event	N Applied	N Admitted	N Enrolled
Fall 2016 (20173)	77864	8723	9667	5962	1806
Fall 2017 (20183)	111616	7366	8635	5458	1509
Fall 2018 (20193)	110245	5889	6725	3423	

Figure 1: *Number of Total Prospects, Event Attending Prospects, Applying Prospects, and Admitted Prospects by Anticipated Year-Term*

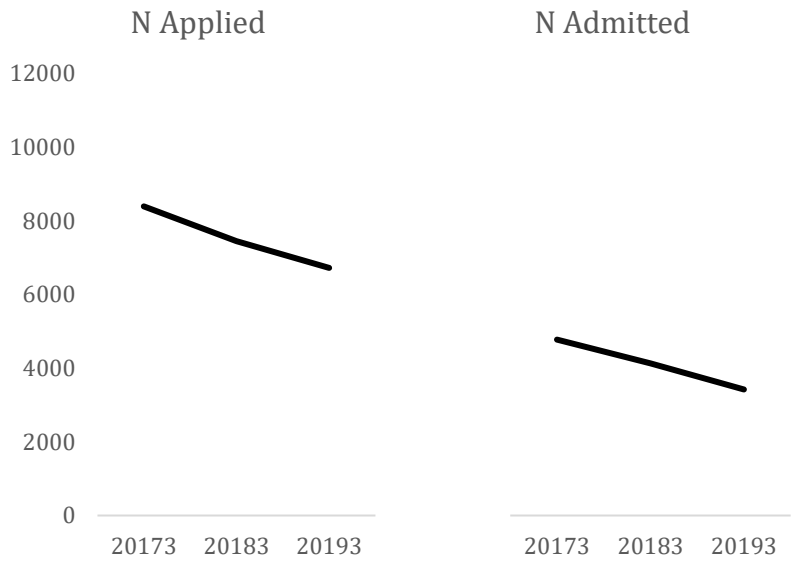


A general trend analysis shows increases or decreases in recent years of the number of total, event-attending, applicant, and admitted prospects. It is important to note that this analysis is conducted during the fall 2018 cycle. The data is current up to February 26. Because the cycle is in process, the number of prospects applying and admitted will be lower than the fall 2016 and fall 2017 cycles. Also, because the fall 2018 semester hasn't happened, there is no enrollment information available for the fall 2018 cycle. Although the number of prospects increased by almost 34,000 from the fall 2016 cycle to the fall 2017 cycle, the number of admissions event-attending prospects decreased by almost 1,400. There were also fewer prospects who applied (1,032 less), were admitted (504 less), and eventually enrolled (297 less). This is cause for concern as likely the lower number of prospects attending events led to lower number of prospects who enrolled. An additional assessment of how this year's cycle is performing compared to the last two cycles will give better insight on how many applications, admissions, and enrolled students we might expect from prospective students next fall. A trend analysis of the number of prospects, number of applicants, and number of admissions at the same time as February 26 for the last two cycles will provide the information to generate these estimates.

Table 2: *Number of Total Prospects, Applying Prospects, and Admitted Prospects at this Time (Feb. 26) of the Fall 2018 Cycle Last 2 Years by Anticipated Year-Term*

Anticipated Year/Term	N Prospects	N Applied	N Admitted
Fall 2016 (20173)	74320	8396	4777
Fall 2017 (20183)	78003	7448	4127
Fall 2018 (20193)	110245	6725	3423

Figure 2: *Number of Applying Prospects and Admitted Prospects at this Time (Feb. 26) of the Fall 2018 Cycle Last 2 Years by Anticipated Year-Term*



At this time during the last two cycles, the number of prospects who applied was much higher. Since the fall 2016 cycle the number of applicants by February 23 have declined 11 percent from 8396 to 7448 in the fall 2017 cycle and 10 percent from 7448 to 6725 in the fall 2018 cycle. In the last two cycles, the total number of prospects who applied on or before February 23 accounted for 86 percent of all applicants. Based on this trend of the last two cycles, we would expect the total number of applicants to be roughly 7770 at the end of the cycle (down 10 percent from fall 2017 cycle). Admissions have also declined at this time of the last two cycles. The number of admissions have declined from 4777 to 4127 (14 percent) in the fall 2017 cycle and declined from 4127 to 3423 (17 percent) in the fall 2018 cycle. In the last two cycles, the total number of prospects admitted on or before February 23 accounted for 78 percent of all admissions. Based on this trend of the last two cycles, we would expect the total number of admissions to be roughly 4390 at the end of this year's cycle. Based on the last two cycles, about 29 percent of admissions enroll. This implies that this year's cycle will result in 1275 enrolled students (down 15.5 percent from the fall 2017 cycle). The number of prospects who attended an admissions event can not be compared at the time of February 23 to the previous cycles. However the decrease in the number of prospects who attended an event from the fall 2016 cycle to the fall 2017 cycle may potentially explain the decrease in the number of applicants, admissions, and enrollments from these prospects. This analysis will further examine the effect that admissions events have on the likelihood a prospect will apply, get admitted, and enroll at SCSU.

**Note: The analysis for all figures, tables, and summaries below are only reported for the fall 2016 and fall 2017 cycles. The fall 2018 cycle is excluded because it is currently in progress.*

Figure 3: *Application, Admission, & Enrollment Rates by Whether a Prospect Attended Event*

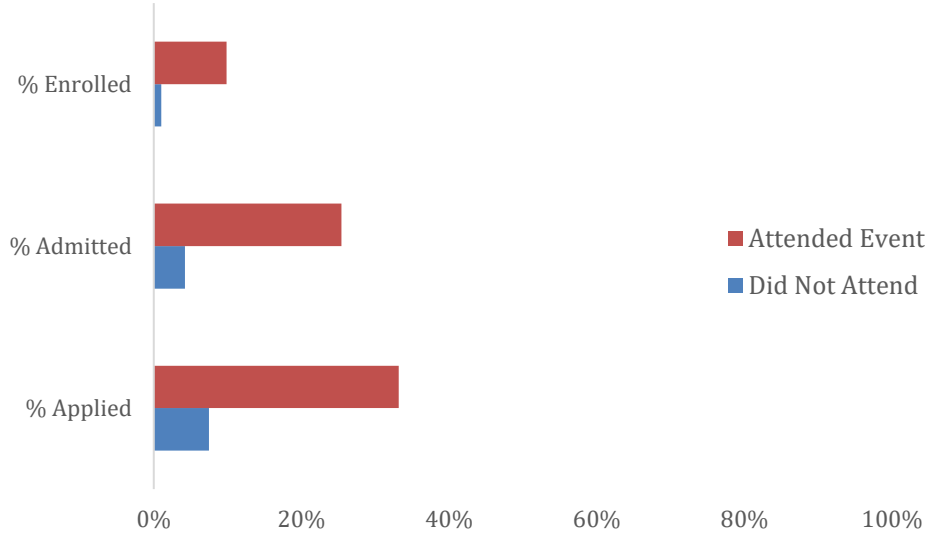


Table 3: *Application, Admission, & Enrollment Rates by Whether a Prospect Attended Event*

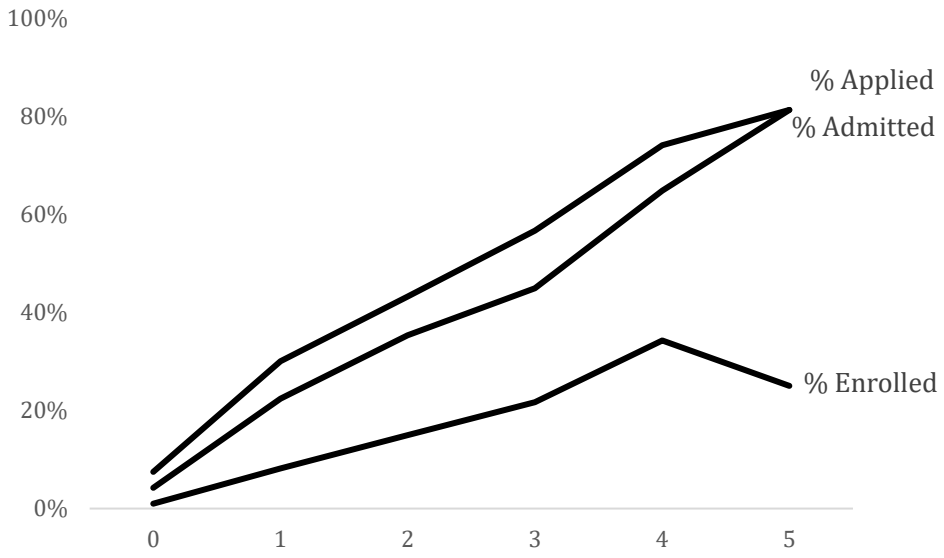
	N Prospects	% Applied	% Admitted	% Enrolled
Did Not Attend	173391	7%	4%	1%
Attended Event	16089	33%	25%	10%

Attending an admissions event drastically increases the probability a student will apply (26 percentage points), get admitted (21 percentage points), and enroll (nine percentage points). Three statistical z-tests evaluated if the proportion of prospects applying, getting admitted, and enrolling is greater for those who attend an event. For all three tests, the p-value was almost zero, meaning that there is strong evidence that suggests that the proportion of prospects applying, getting admitted, and enrolling are higher for those who attend admissions events.

Table 4: *Application, Admission, & Enrollment Rates by the Number of Events a Prospect Attended*

Number of Events Attended	N Prospects	% Applied	% Admitted	% Enrolled
0	173391	7%	4%	1%
1	13026	30%	22%	8%
2	2471	43%	35%	15%
3	461	57%	45%	22%
4	108	74%	65%	34%
5	16	81%	81%	25%
6	7	100%	100%	71%

Figure 4: *Application, Admission, & Enrollment Rates by the Number of Events a Prospect Attended*



Although most do not attend multiple admissions events, for those that do, the probability the prospect applies, gets admitted, and enrolls increases when they attend additional events. The application rate increases from seven to 30 percent when a prospect attends one event and increases 13 more percentage points when they attend a second. This trend is similar for admission and enrollment rates.

Figure 5: *Application Rates by Whether a Prospect Attended a Particular Event*

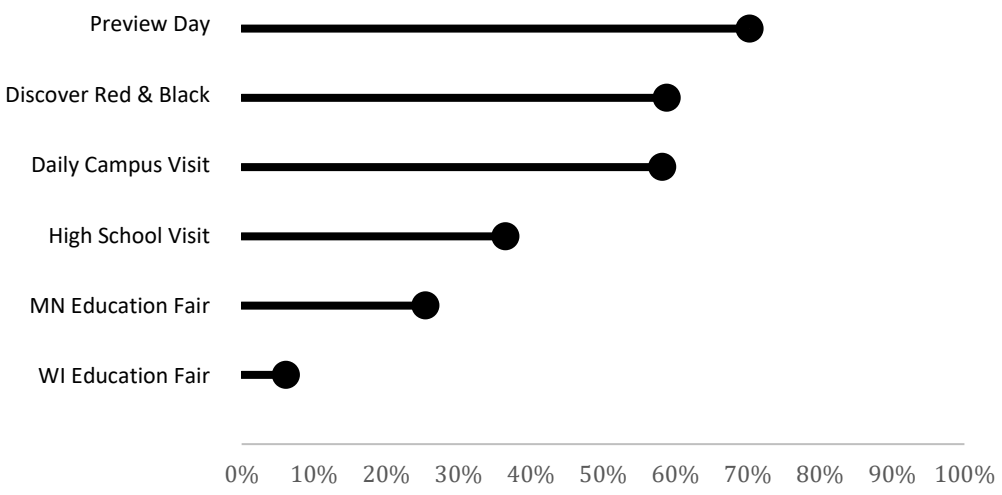


Table 5: *Application Rates by Whether a Prospect Attended a Particular Event*

Event Type	Did Not Attend % Applied	Attended % Applied
Daily Campus Visit	9%	58%
Discover Red & Black	9%	59%
High School Visit	9%	37%
Preview Day	9%	70%
Minnesota Education Fair	10%	25%
Wisconsin Education Fair	10%	6%

The main research question for this analysis is discovering which admissions events lead to the highest application, admission, and enrollment rates. The main admissions events of interest are listed in the Data and Demographics section as well as above in Table 5. For almost all events, of the prospects who do not attend, roughly ten percent apply. The admissions event with the highest application conversion rate is a Preview Day. 70 percent of those who attend a Preview Day apply. Other events with higher conversion rates are Discover Red & Black (59 percent) and Daily Campus visit (58 percent). The lowest conversion rates were for the Minnesota (25 percent) and Wisconsin Education fairs (6 percent), events that occur off campus.

Figure 6: *Admission & Enrollment Rates by Whether a Prospect Attended a Particular Event*

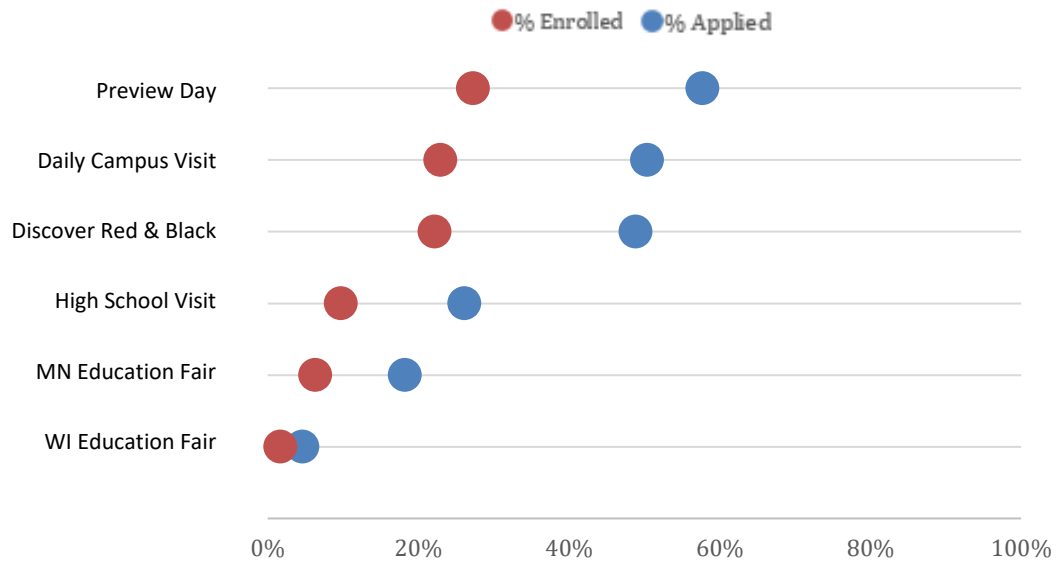


Table 6: *Admission Rates by Whether a Prospect Attended a Particular Event*

Event Type	Did Not Attend % Admitted	Attended % Admitted
Daily Campus Visit	5%	50%
Discover Red & Black	6%	49%
High School Visit	6%	26%
Preview Day	6%	58%
Minnesota Education Fair	6%	18%
Wisconsin Education Fair	6%	5%

Table 7: *Enrollment Rates by Whether a Prospect Attended a Particular Event*

Event Type	Did Not Attend % Enrolled	Attended % Enrolled
Admitted Student Day	2%	79%
Daily Campus Visit	1%	22%
Discover Red & Black	2%	23%
High School Visit	2%	10%
Preview Day	2%	27%
Minnesota Education Fair	2%	6%
Wisconsin Education Fair	2%	2%

Admission rates are highly associated with application rates. For almost all events, of the prospects who do not attend, roughly six percent are admitted. The admissions event with the highest admission conversion rate is a Preview Day. 58 percent of those who attend a Preview Day are admitted. Other events with higher admit conversion rates are Discover Red & Black (49 percent) and Daily Campus visit (50 percent). The lowest conversion rates were for the Minnesota (18 percent) and Wisconsin Education fairs (5 percent), events that occur off campus.

For almost all events, of the prospects who do not attend, roughly one to two percent enroll. The admissions event with the highest enrollment conversion rate is to no surprise an Admitted Student Day who attracts 79 percent to enroll. Events other than an Admitted Student Day with higher conversion rates are again a Preview Day (27 percent), Discover Red & Black (23 percent) and Daily Campus visit (22 percent).

Overall, it is clear that the highest converting admissions events in promoting applications, admissions, and enrollment are the Preview Day, Discover Red & Black, and Daily Campus Visits.

**Note: The analysis for all figures, tables, and summaries below are only reported for the fall 2017, fall 2018 cycles, and fall 2019 cycles.*

Table 8: *Application, Admission, & Enrollment Rates by First Event Attended*

First Event	N	% Applied	% Admitted	% Enrolled
Did Not Attend an Event	288738	7%	4%	1%
High School Visit	4677	29%	19%	6%
Daily Campus Visit	4224	51%	43%	16%
Discover Red & Black	1109	53%	43%	15%
Preview Day	**977**	**63%**	**51%**	**18%**

Further analysis allows for understanding which order of events is the most effective in achieving the highest conversion rates. There is an issue however since most prospects do not attend multiple events. As a result, the analysis will only examine which event is the most effective on its own and which two events are most effective and in which order. Prospects that attend only a Preview Day are more likely to apply (63 percent), get admitted (51 percent), and enroll (18 percent) compared to prospects who only attend one of the other events. The next best stand-alone events are Discover Red & Black and Daily Campus Visits. This follows the same pattern discovered earlier when examining which events are the most effective in converting prospective students.

Table 9: *Application, Admission, & Enrollment Rates by Order of First Two Events Attended*

First Event	Second Event	N	% Applied	% Admitted	% Enrolled
Did Not Attend an Event		288738	7%	4%	1%
High School Visit		4071	24%	15%	4%
Daily Campus Visit		4026	49%	41%	15%
Discover Red & Black		987	52%	42%	14%
Preview Day		919	62%	49%	16%
High School Visit	High School Visit	314	36%	28%	8%
High School Visit	Daily Campus Visit	190	78%	68%	28%
Daily Campus Visit	**High School Visit**	**86**	**90%**	**81%**	**41%**
High School Visit	**Preview Day**	**61**	**84%**	**62%**	**38%**
Daily Campus Visit	**Daily Campus Visit**	**58**	**74%**	**71%**	**41%**
Discover Red & Black	High School Visit	54	80%	61%	28%
Discover Red & Black	Discover Red & Black	49	37%	29%	14%
High School Visit	Discover Red & Black	41	80%	73%	27%
Daily Campus Visit	Discover Red & Black	32	81%	72%	41%
Preview Day	*High School Visit*	*26*	*92%*	*81%*	*46%*
Daily Campus Visit	*Preview Day*	*22*	*91%*	*91%*	*45%*

Finding the order of which two events are the most effective in converting prospective students is more difficult. Because very few prospects attend multiple events, the sample size for each combination of two events is relatively small. Based on the small sample size, it is difficult to decipher which two events are clearly the best. Early, this analysis determined that the Preview Day, Discover Red & Black, and Daily Campus Visits are the most effective events. The three combinations of events highlighted in black in Table 9 show three of the most effective orders. Because the sample sizes of the event combinations italicized are relatively small, it is difficult to statistically conclude those combinations are more effective. However, there is still value in knowing that the Preview Day occurs in both of these combinations and together with another event, generates high application, admission, and enrollment rates.

The analysis and summary information discovered in the exploratory analysis above provides insight for building two predictive models: one to determine the likelihood or probability a prospective student will apply and another to determine the probability a prospective student will enroll.

The model for predicting the probability a prospective student will apply considers predictors such as the types of admissions events attended, the number of admissions events attended, and when the admissions events were attended. Unfortunately, demographic information of prospects is unknown prior to a student applying. To build the model the prospects in the fall 2017 and fall 2018 cycles will be split into two groups: 60 percent will be used to train the model and 40 percent will be used to validate the model. The fall 2019 cycle will be the test set once the model is built.

Table 10: *Logistic Regression Model to Predict the Probability a Prospective Student will Apply*

Variable	Estimate	P-value	Effect Size	Power (Decrease)	Power (Increase)
Intercept	-0.48	<0.01			
Number of Events Attended	0.77	<0.01	42%	100%	100%
Attended a Preview Day	1.86	<0.01	31%	67%	47%
Attended Discover Red & Black	1.49	<0.01	22%	72%	51%
Attended a Daily Campus Visit	1.31	<0.01	18%	89%	70%
Attended a High School Visit	0.58	<0.01	6%	92%	75%
Days Before Term When Attended 1st Event is Missing	-0.24	<0.01	-2%	92%	76%
Days Before Term When Attended 1st Event	-0.01	<0.01	-6%	100%	100%

** Note: The model estimates are reported for a one-unit change in the variable. The effect size is the change in probability of the response from the mean probability when the predictor is present. The effect sizes are calculated over the range of the predictor variable. For instance, if a prospect attended a Preview Day, they probability of applying, on average, increases by 31 percentage points from the mean. The effect size for days before term when the first event was attended is reported over a range of 100 days. The effect size for number of events attended is reported over a range of 3 events. Power (Decrease/Increase) is the probability of correctly*

detecting a 2.5 percentage point decrease/increase in the average response probability for the given variable, with all other variables held constant.

Table 11: *Fit Statistics for Logistic Model to Predict Application Probability*

Statistic	Training	Validation
Entropy R Square	0.10	0.09
Area Under ROC Curve	0.62	0.62
Lift for Predicting Top 20%	3.20	3.15

The logistic model explanatory terms are listed in Table 10. All of the variables in the model are significant at the five percent level. There is no need for p-value cutoff adjustment procedures to minimize the probability of committing a type one error because the p-values are all extremely small. Although all add value in predicting the probability a student will attend an event, some have a larger effect relative to others. If a prospect attended a Preview Day and if a prospect attended Discover Red & Black have the largest effect on the outcome. Their respective effect sizes across their entire range are 31 and 22 percent.

The average probability a prospective student applies is 9.5 percent. This analysis is interested in detecting a change of at least 2.5 percentage points in either direction of the mean response for a particular variable. So if the prediction probability is estimated to be 12 percent (or higher) or 7 percent (or lower), the variable, with all other variables held constant, changes the probability at least 2.5 percentage points. The power (probability that the test will accurately detect such a difference) for each variable is reported in Table 10. For example, if a prospect attended a Preview Day, the probability of correctly detecting a 2.5 percentage point increase in the probability a student applies is only 47 percent. Because of the small standard errors due to large sample sizes, the power of detecting a 2.5 percentage point difference for some predictor variables is 100 percent. In general, statisticians aim for a percentage of 80 percent or higher.

Further descriptions of the types of statistical error, p-values, and power are summarized in Appendix C.

For all provided information, the model performs relatively mediocre in predicting the outcome. Good accuracy measures for comparing logistic models are R-squared, misclassification rate, area under the receiver operating curve (ROC) curve, and lift. Generally, statisticians consider a model with 0.62 area under the ROC curve a relatively poor performing model. It is difficult to predict the probability a prospect will apply based on simply the number of and types of admissions events attended. More information must be known in order to develop a more accurate prediction.

The model for predicting the probability a prospective admitted student will enroll considers predictors such as the types of admissions events attended, the number of admissions events attended, when the admissions events were attended, and demographic information. To build the model the admitted prospects in the fall 2017 and fall 2018 cycles will be split into two groups: 60 percent will be used to train the model and 40 percent will be used to validate the model. The fall 2019 cycle will be the test set the once the model is built.

Table 12: *Logistic Regression Model to Predict the Probability a Prospective Admitted Student will Enroll*

Variable	Estimate	P-value	Effect Size	Power (Decrease)	Power (Increase)
Intercept	-1.03	<0.01			
Attended an Admitted Student Day	2.22	<0.01	50%	14%	13%
Attended a Daily Campus Visit	0.62	<0.01	14%	30%	27%
Attended a Preview Day	0.58	<0.01	13%	14%	13%
Attended Discover Red & Black	0.58	<0.01	13%	14%	13%
N Events Attended before Admitted Student Day	0.18	<0.01	12%	78%	74%
Student of Color	-0.15	0.02	-3%	51%	47%
High Achiever	-0.54	<0.01	-10%	57%	52%

** Note: The model estimates are reported for a one-unit change in the variable. The effect size is the change in probability of the response from the mean probability when the predictor is present. The effect sizes are calculated over the range of the predictor variable. For instance, if an admitted prospect attended an admitted student day, their probability of applying, on average, increases by 50 percentage points from the mean. The effect size for number of events attended is reported over a range of 3 events. Power (Decrease/Increase) is the probability of correctly detecting a 2.5 percentage point decrease/increase in the average response probability for the given variable, with all other variables held constant.*

Table 13: *Fit Statistics for Logistic Model to predict Enrollment Probability*

Statistic	Training	Validation
Entropy R Square	0.08	0.08
Area Under ROC Curve	0.66	0.68
Lift for Predicting Top 20%	1.75	1.80

The logistic model explanatory terms are listed in Table 12. All of the variables in the model are significant at the five percent level. There is no need for p-value cutoff adjustment procedures to minimize the probability of committing a type one error because the p-values are all extremely small. Although all add value in predicting the probability a student will attend an event, some have a larger effect relative to others. If an admitted prospect attended an Admitted Student Day, as expected, has the largest effect (50 percentage points) on the probability of enrolling.

The average probability an admitted prospective student enrolls is 28 percent. This analysis is interested in detecting a change of at least 2.5 percentage points in either direction of the mean response for a particular variable. So if the prediction probability is estimated to be 30.5 percent (or higher) or 25.5 percent (or lower), the variable, with all other variables held constant, changes the probability at least 2.5 percentage points. The power (probability that the test will accurately detect such a difference) for each variable is reported in Table 12. For example, if an admitted prospect attended an Admitted Student Day, the probability of correctly detecting a 2.5 percentage point increase in the probability a student applies is a mere 13 percent.

For all provided information, the model performs relatively mediocre in predicting the outcome. Good accuracy measures for comparing logistic models are R-squared, misclassification rate, area under the receiver operating curve (ROC) curve, and lift. Generally, statisticians consider a model with 0.66 area under the ROC curve a relatively poor performing model. It is difficult to predict the probability an admitted prospect will enroll based on simply the number of and types of admissions events attended. More information must be known in order to develop a more accurate prediction model.

Conclusion:

This analysis leverages St. Cloud State University's vision of *"discovery, applied knowledge and creative interaction*" through student research to provide the Admissions Office at St. Cloud State with the insights necessary to strategically acquire prospective students through more effective event strategies. The main findings in this research are:

- Although the number of prospects increased by almost 34,000 from the fall 2016 cycle to the fall 2017 cycle, the number of admissions event-attending prospects decreased by almost 1,400. There were also fewer prospects who applied (1,032 less), were admitted (504 less), and eventually enrolled (297 less).
- Attending an admissions event drastically increases the probability a student will apply (26 percentage points higher), get admitted (21 percentage points higher), and enroll (nine percentage points higher).
- 70 percent of those who attend a Preview Day apply. Other events with higher conversion rates are Discover Red & Black (59 percent) and Daily Campus visit (58 percent).
- It is difficult to predict the probability a prospect will apply and enroll based on simply the number of and types of admissions events attended. More information must be known in order to develop a more accurate prediction model.

Not only did this study determine which admissions events are most effective in attracting applications and admissions, it also reported the most effective order and built two models: one to predict if a prospective student, based on their admissions event attendance, is likely to apply and the other to predict if an admitted prospect, based on their characteristics and event attendance, will enroll. This research provides insights that can potentially improve the way the university more effectively attracts prospective students to enroll in the institution.

Appendix A: Means of Analysis

Methods and Solution:

Problem: Did we see fewer prospects along the way leading up to the two most recent cohorts (20183 and 20193) in comparison to 20173?

- Prepare a time series analysis of the number of prospects each year from 2016 to 2018. The analysis will determine if this data validates the recent fall in prospects to SCSU, which began in 2016.

Problem: Which admissions events are most successful in acquiring prospects to apply? How much do conversion rates differ?

- Comparative analyses of the application conversion rates for students who attend given Admissions Events. This will determine if particular admissions events are more effective in acquiring applications and how much more effective.

Problem: What admissions events are more likely to gain admitted students? More likely to result in enrolled students?

- Comparative analyses of the admission and enrollment conversion rates for students who attend given Admissions Events. This will determine if particular admissions events are more effective in acquiring admissions and enrolled prospects and how much more effective.

Problem: Is there any admissions event for the fall 2016 cohort that stands out as giving a high likelihood of a student eventually being both enrolled and retained?

- Comparative analyses of enrollment and retention rates for students who attend different admissions events. This will determine if particular admissions events are more effective in both acquiring enrolled prospects and retaining students in terms 2 and 3.

Problem: What is the optimal event timeline for producing the highest application conversion rates?

- Comparative analyses of the application conversion rates for students who attend given Admissions Events in a particular order during a specified timeline. This will determine if attending particular recruiting events in a given order throughout a given time interval is more effective in acquiring applicants.

Problem: Can a reliable predictive model be created to predict the probability a student will apply to SCSU with known information such as event attendance and demographic information? Prepare a predictive model using the anticipated fall 2016 – fall 2017 prospects' event attendance, application, and demographic information to predict the probability students will apply in fall 2018. The main demographics that will be used for the prediction models are: Anticipated Year of Enrollment, International/Non-International, Student-of-color/Non-Student-of-Color, if high achieving, if among Top 10 and 25 of high school graduating class, if an ACE admit or referral, and if applied for FAFSA.

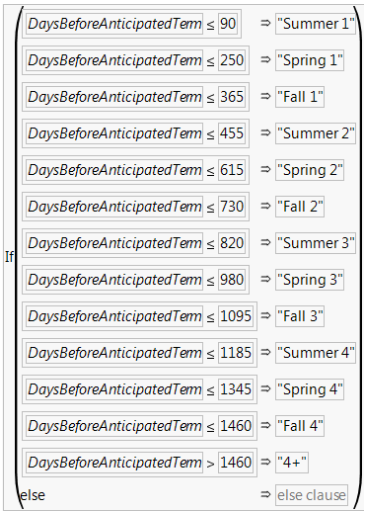
Data Manipulation:

A series of data manipulation procedures are required in order to conduct the appropriate analysis. The manipulation, cleaning, and joining procedures are outlined by each research question and their respective procedures to prepare for proper analysis.

Prior Preparation:

1. Begin with 2,624,353 Observations of Prospective student's with anticipated YRTR of enrollment between fall 2016 and fall 2018
2. Remove 258,779 observations w/ APP_CAT = Undergrad Transfer
3. Remove duplicate values in data set. 415,013 Observations Remain.
4. No prospects have more than one application.
5. Recode all event names into one of 19 categories listed in the appendix.
6. Sort by Event Date and create a formula column which is a counter for the number of times a student appears.
7. Create a formula column which is a counter for the number of events a student attends.
8. Create a formula to find the number of days before the anticipated term of the event date. Remove records where the number of days before the anticipated start date is negative. (570 Observations)
9. Remove observations where prospect had previously applied for a different term. This will place the attention on students who are applying one time and not taint the results with prospect event records duplicated for prospects who applied again at a later time for a different YRTR. Removes 328 observations from 108 students. (414,115 Observations Remain)

10. Some prospectIDs are duplicated in the file because they are in two different programs. Remove 2502 duplicated event observations for these duplicated prospectIDs.
11. Remove observations with the same Name, Event Name, and Log but multiple UIDs. (408,199 Observations Remain)
12. Create a new variable to assign the time period of the event using the # of days before term the event occurred. The formula is:



Problem: Did we see fewer prospects along the way leading up to the two most recent cohorts (20183 and 20193) in comparison to 20173?

1. Calculate the number of unique prospects with a pivot table in Excel by anticipated year and cohort.

Problem: Which Admissions Events are most successful in acquiring prospects to apply? How much do conversion rates differ?

1. Create 19 Columns for each event type. Code these columns 1 if the event was in the category and they attend, 0 if otherwise.

2. Summarize the data at the prospect level. Group by ProspectID and sum the 19 Columns created in number 1. Then create 19 dummy variables for the 19 columns created in 1 to be 1 if the prospect attended the event, 0 otherwise.
3. Create a Pivot Table in Excel to calculate the proportion of students who applied by each event. Also get a count of the total sample size. This allows for a z-test to be performed to test if there is a difference in proportions. Examine by year as well.

Problem: What admissions events are more likely to gain admitted students? More likely to result in enrolled students?

1. Create a dummy variable column to code if a student was admitted: 1 = Admitted, 0 = Not Admitted.
2. Create a dummy variable column to code if a student enrolled: 1 = Enrolled, 0 = Not Enrolled.
3. Follow the same procedure in parts 1 and 2 for problem 1.
4. Create a Pivot Table in Excel to calculate the proportion of students who were admitted by each event. Also get a count of the total sample size. This allows for a z-test to be performed to test if there is a difference in proportions. Examine by year as well.
5. Create a Pivot Table in Excel to calculate the proportion of students who enrolled by each event. Also get a count of the total sample size. This allows for a z-test to be performed to test if there is a difference in proportions. Examine by year as well.

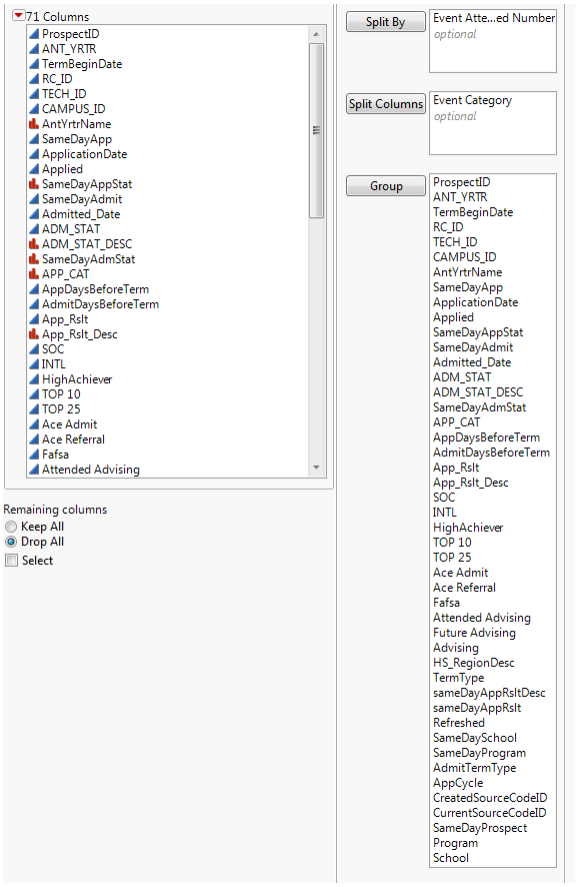
Problem: Is there any admissions event for the fall 2016 cohort that stands out as giving a high likelihood of a student eventually being both enrolled and retained?

1. Left outer join the enrollment and retention data to the prospects by Tech ID.

2. Create dummy variables for Term 2 and Term 3 retention for these students: 1 = Retained, 0 = Not Retained.
3. Create a Pivot Table in Excel to calculate the proportion of students who were retained in Term 2 and in Term 3 by each event. Also get a count of the total sample size. This allows for a z-test to be performed to test if there is a difference in proportions.

Problem: What is the optimal event timeline for producing the highest application conversion rates?

1. Subset Observations where Log = "Event Attended"
2. In JMP go to Tables > Split and enter the columns as shown below in their respective dialog box:



Problem: Can a reliable predictive model be created to predict the probability a student will apply to SCSU with known information such as event attendance and demographic information?

1. Manipulate the data so that there is one record per prospect with demographic information and event attendance information. Use such variables to build a regression or random forest model to predict the likelihood a student will apply, get admitted, and enroll. Further information will become available for this process throughout the model building process.

Appendix B: Main Event Categories

Admitted Student Day
AVID
Be the Broadcaster
College Access Group - Transfer
College Fair (Other)
Community College Event
Daily Campus Visit
Discover Red & Black
Experience STEM @ SCSU
High School Visit
Honors Visit Day
Junior Visit Day
Minnesota Education Fair
MNACC College Fair
NACAC Fair
Other
Preview Day
Senior to Sophomore Class
Wisconsin Education Fair

Reference Spreadsheet: EventCategories.xlsx to observe which event names were assigned to each main event category.

Appendix C: Types of Statistical Error, P-Values, and Power

Types of Statistical Error:

When statistician performs a hypothesis test, a decision is made whether to reject or not reject the null hypothesis in favor of an alternative. In either instance, there is potential the conclusion is incorrect. There are two types of incorrect conclusions a hypothesis test is susceptible to: false positive and false negative.

In the instance of a false positive, the null hypothesis is wrongly rejected in favor of the alternative. This is known in statistics as a type one error. The probability of making a type one error (α level) is predetermined before the test and is the p-value cutoff the statistician uses as a guideline whether to reject the null. To avoid type one errors, a statistician is advised to reduce the α level for a more conservative test. For instance, typically the α level is set low during medical trials involving the effectiveness of a particular drug with high side affects. This is to assure that the probability of incorrectly observing this result is low.

The second type of error is a false negative conclusion, which occurs if the null hypothesis is not rejected when the alternative is true. This false negative is referred to as a type two error. Unlike the probability of a type one error, the probability of a type two error cannot generally be calculated without knowing the population parameter of interest. The probability of a type two error is merely 1 minus the power. Typically the probability of performing a type two error decreases as the sample size increases. In the case of testing for effective drugs, a larger sample size will lower the probability of falsely not detecting a difference. This will further assure that the truly effective drugs will not go undetected.

It is important for statisticians to understand the impact a wrong decision will have on the business, health, etc. This is why determining and understanding type I and type II errors is essential before conducting analysis and reporting results.

P-Values:

P-values are often misused and misinterpreted, leading to many inaccurate results. The definition of a p-value is simply the probability, under the assumption that there is no difference, of collecting data that shows a difference equal to or more extreme than what was observed. A p-value is calculated under that assumption that there is no difference. The p-value only shows the probability the data is more extreme than the null hypothesis. A p-value does not show any meaning of effect size. For instance, in statistical hypothesis tests, p-values and the specified cutoff only tells whether or not there is a likely difference in means, proportions, etc. It does not tell how big the difference is, or give any other insight as to how to practically use the information.

P-values are misused in a variety of other ways such as determining significance in multiple comparison tests, selecting explanatory variables in models, and determining the false discovery rate of an experiment. Statisticians can follow a number of best practices to minimize misinterpretations and misuse of the p-value.

When performing a lot of multiple comparison tests for significance of regressors, procedures such as Bonferoni's method and the Benjamani-Hochberg method can adjust the p-value threshold to assure the false discovery rate does not increase. Another practice when determining

explanatory variables in models is to incorporate the effect size of a variable through the selection process. Although a p-value may be less than .05, the size of the effect could still explain a good chunk of the variability in the response. There is also the possibility of committing type II error if a variable is excluded from a model. This is also why determining power is important. It was shown during multiple presentations that p-values are very susceptible to change as the sample size varies. A statistician does not want to make the mistake of eliminating a predictor with a large effect size in the early stages of the analysis due to simply evaluating a p-value. As a general practice, statisticians should always report the effect size and the power of a test whenever reporting p-values.

Power:

Statistical power is inversely related to the probability of a type two error. A type two error is a false negative, or incorrectly failing to reject the null in favor of the alternative. Power is equal to 1 minus the probability of a type two error. Power is defined to be the probability of correctly rejecting the null when the alternative is true. In other words, if there is a true difference in means, proportions, etc., power is the probability we detect it. Power is an important calculation because it determines the probability we correctly detect the alternative hypothesis when it is true. Although power cannot be directly chosen like that of a p-value, it is possible to calculate using mathematics. Power calculation is dependent on three parameters: effect size, sample size, and the alpha level. Greater effect sizes are easier to detect and larger sample sizes decrease the test sensitivity. If any three out of these four parameters are known, the fourth can be calculated.

There are a couple of reasons why a power analysis is useful. The most frequent use of power is to determine the number of trials or the sample size to find a certain sized effect (if the power is

known). This is important as it tells the statistician how many trials must be run to avoid conducting a type two error. The second use is to actually calculate the power of a test. If you only have a certain sample size to work with what will the probability be that you actually detect the true alternative. If the power in this instance is low and the null is rejected, there is uncertainty that the result is true.

Calculating power is fairly difficult and often requires software. Because most of the data is already collected for analysis in STAT 381, the power analysis will be useful to determining the power of a given statistical test. The desired power level is often 80%.

Appendix D: Variable List

Variable Name	Description
ProspectID	Prospect Identification (Primary Key)
ANT_YRTR	ANT_YRTR: The year term associated with a given application. Rightmost character designates the term type: 1 = summer, 3 = Fall, and 5 = Spring. Converts summer to Fall yrtrs.
RC_ID	System Maintained Reporting Center
TECH_ID	Technical Identification Number: System-generated technical ID for this
CAMPUS_ID	The campus ID for the unique information being created.
AntYrtrName	Term name of the ant_yrtr attribute
SameDayApp	Indicator for if application record was from on or before the compareable same day of the week for each cycle
MinAppDate	Minimum application date on record for a given student per term.
SameDayAppStat	Derived Application Status from on or before the compareable same day of the week for each cycle
SameDayAdmit	
Begin_Date	Formatted as date version of the begin_date string from st_adm_stat
ADM_STAT	Admission Status: The admission status code of student (i.e. undergraduate, PSEO)
ADM_STAT_DESC	Long Description: Long description of the adm_stat code.
SameDayAdmStat	
APP_CAT	Derived Type of Applications. Determined NEF or NET based on HS Graduation Year. If the observation has an admission status, then that value is represented.
AppDaysBeforeTerm	Number of days before the term begin date where the given application was submitted
AdmitDaysBeforeTerm	Number of days before the term begin date for a given admission begin date
App_Rslt	The result code of the application (Accepted, Denied, Waiting List, Accpt Appeal, Pending)
App_Rslt_Desc	Long Description: Long description of the app_rslt code.
SOC	Student of Color Identifier
INTL	International Student Identifier
HighAchiever	
TOP 10	Graduated in Top 10 Percent of Class Indicator
TOP 25	Graduated in Top 25 Percent of Class Indicator
Ace Admit	
Ace Referral	
Fafsa	
Attended Advising	
Future Advising	
Advising	
HS_RegionDesc	Description associated with the HS_Region
TermType	
sameDayAppRsltDesc	Long Description: Long description of the sameDayAppRslt code.

sameDayAppRslt	The result code of the same day application (Accepted, Denied, Waiting List, Accpt Appeal, Pending)
Refreshed	
SameDaySchool	
SameDayProgram	
AdmitTermType	
AppCycle	String value for each of the cycles: FallThsCyc = Fall This Cycle SpringThsCyc = Spring This Cycle Fall1Cyc = Fall 1 Cycle Ago Spring1Cyc = Spring 1 Cycle Ago Fall2Cyc = Fall 2 Cycles Ago Spring2Cyc = Spring 2 Cycles Ago
CreatedSourceCodeID	Source code associated with prospect record that was designated during its initial import into Hobsons. Source codes are updated for prospects records everytime the prospect record is updated.
CurrentSourceCodeID	Source code associated with the most recent update to a prospect record in Hobsons. Source codes are updated for prospects records everytime the prospect record is updated.
SameDayProspect	
Program	Program associated with the Major attribute in this table
School	School associated with the major attribute in this table
Created	Prospect created date from Hobsons

Resources:

Walsh, P. (2016, March 3). In all-sport meeting, St. Cloud State athletes told 6 programs being eliminated. *Star Tribune.* Retrieved from http://www.startribune.com/st-cloud-state-dropping-6-sports-programs/370796061/