EdgeCon ANNUAL CONFERENCE 2020

Creating a Culture of Data-Informed Decision Making - Recruitment, Admissions, Retention and Graduation strategies based on Predictive Models

Bharathwaj Vijayakumar
Associate Director of Analytics
Rowan University

About Rowan University

- Founded in 1923 as a teacher preparation college

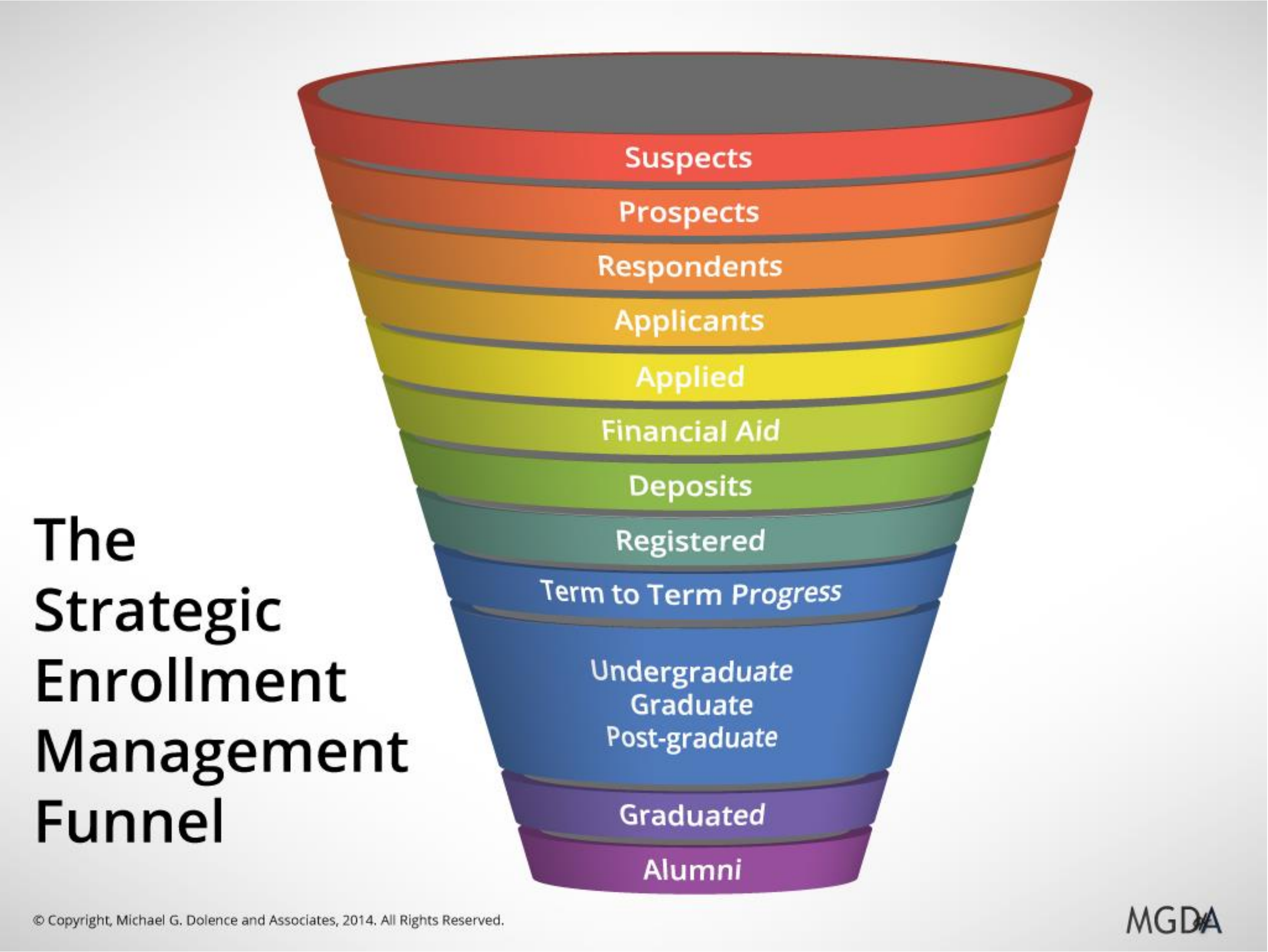
- Academic degree programs include:
  85 Bachelor’s
  46 Master’s
  2 Professional
  6 Doctoral

- Fall 2019 student body
  Total: 19,618
  Undergraduate: 16,011
  Graduate: 2,417
  Professional/Medical: 1,190

Models at Rowan University

- Visual & Predictive analytics was first implemented at Rowan in 2013
- Total enrollment has increased by 47% since 2013
- Incoming First-Time UG class has grown to 67% from 1,618 to 2,695 students since 2013
- Different models developed at Rowan University
  - ➢ Recruitment
  - ➢ Admissions and scholarships
  - ➢ Retention
  - ➢ Graduation
  - ➢ Various visual analytical tools to track student progress

Higher Education Funnel



EdgeCon ANNUAL CONFERENCE 2020

Presentation Overview

1. Recruitment Model – Identify the target market to recruit the students and convert them to applicants

2. Admissions Model Quality & Enrollment – Determine the probability of enrollment for each applicant

3. Bottleneck Course Analysis – Determine the bottleneck courses in a program to help the students graduate on time

4. Attrition Model – Determine the probability of attrition from the first semester to the second semester

1.Recruitment Model

Recruitment Model

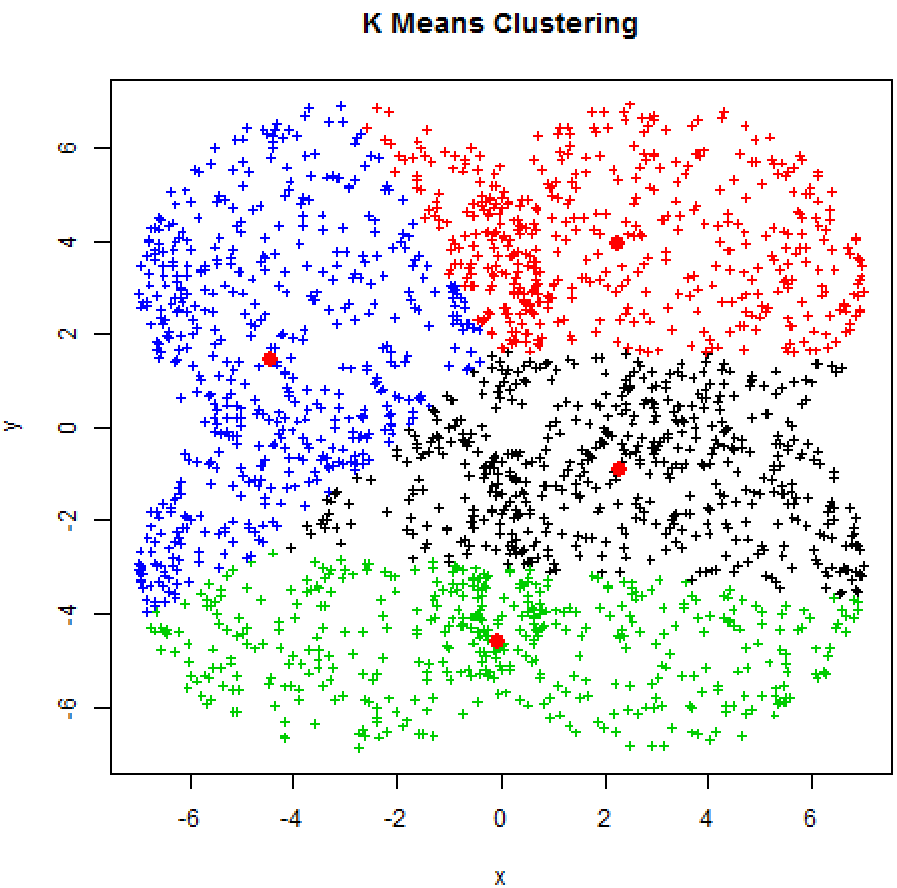
Goal:
Identify clusters or segments based on historical data and use those to determine our targeting strategies for prospective students

Methodology used:

- Use visual analytics to understand historical trends and patterns
- Use K-means clustering
- Create clusters or segments and use those to buy names from different sources

Recruitment Model

- K-means Cluster analysis



EdgeCon ANNUAL CONFERENCE 2020

Recruitment Model

- K-means Cluster analysis
- Classified transfer students enrolled and not enrolled into 3 groups using k-means clustering

- Variables used for clustering
  - Transfer GPA
  - Distance

- Variables used for understanding the clusters
  - Credits transferred in
  - Age
  - Days between application to enrollment
  - Major
  - Transfer Institution
  - Scholarships offered
  - Number of first generation students

Recruitment Model

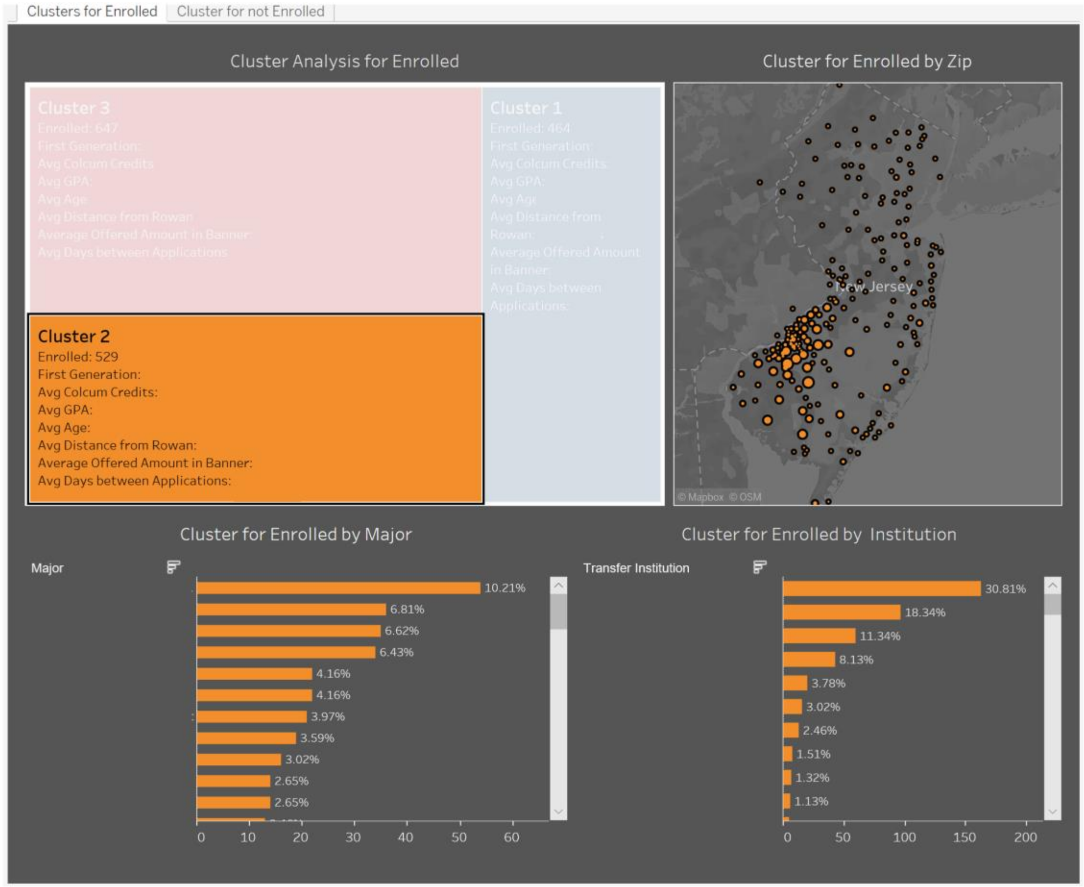
Cluster Analysis for Enrolled

Cluster 3
Enrolled: 647
First Generation:
Avg Colcum Credits:
Avg GPA:
Avg Age:
Avg Distance from Rowan:
Average Offered Amount in Banner:
Avg Days between Applications:

Cluster 2
Enrolled: 529
First Generation:
Avg Colcum Credits:
Avg GPA:
Avg Age:
Avg Distance from Rowan:
Average Offered Amount in Banner:
Avg Days between Applications:

Cluster 1
Enrolled: 464
First Generation:
Avg Colcum Credits:
Avg GPA:
Avg Age:
Avg Distance from Rowan:
Average Offered Amount in Banner:
Avg Days between Applications:

Recruitment Model



- Narrow down the target segments
- Classify the enrolled students in high, medium and low based on GPA
- Write a target population statement for each cluster, and have different marketing campaigns for each cluster based on distance, major, schools etc.
- Similarly, analyze for students not enrolled, and understand the segments

2.Admissions – Quality & Enrollment Model

Admissions Quality & Enrollment Model

Goals

- To identify the right transfer students
- Offer them right amount of scholarships
- Yield them

Methodology Used

- Use visual analytics to understand historical trends and patterns
- Use logistic regression for modeling
- Identify probability of enrollment for individual students and develop communication strategies

Admissions Quality & Enrollment Model

Enrollment Model

Outcome Variable: Chances of a student enrolling at Rowan University. Will receive a probability score

Some of the input variables

- Distance
- Number of credits brought in
- Institution transferring from
- Previous degree
- Gender
- Major
- Some financial aid variables

Admissions Quality & Enrollment Model

Quality Model

Outcome Variable: Likelihood of a student graduating on time at Rowan University. Will get a probability score. Calculated based on the number of credits brought in, expected graduation time and actual graduation time

Examples of input variables

- Department
- Institution transferring from
- Gender
- College
- Credits transferring
- Days between app to enroll
- Age
- Transfer GPA
- Major
- Need

Admissions Quality& Enrollment Model

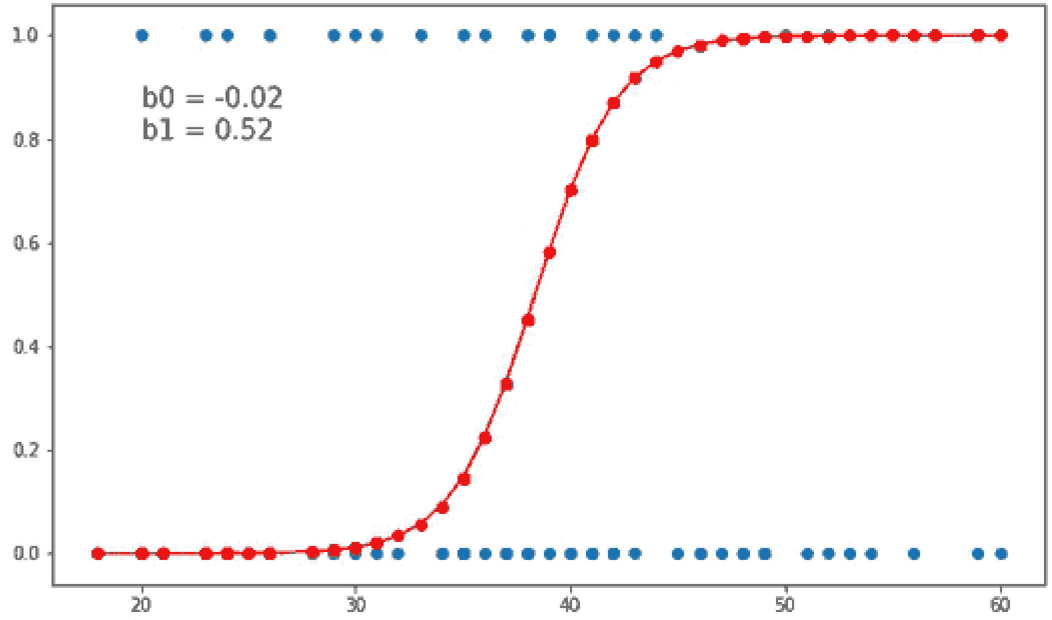
To identify the right transfer students – Achieved by analyzing historical data by reviewing student yield, and graduation rates and building regression models

Offer students the correct amount of scholarship funds –
Review the current distribution of scholarships, and fit them to the model. Identify the correct amount based on the distribution of students in each decile and the given budget. Decide which deciles to target.

Yield them – Monitor the yield and improve communication in the highly ranked deciles to improve the yield

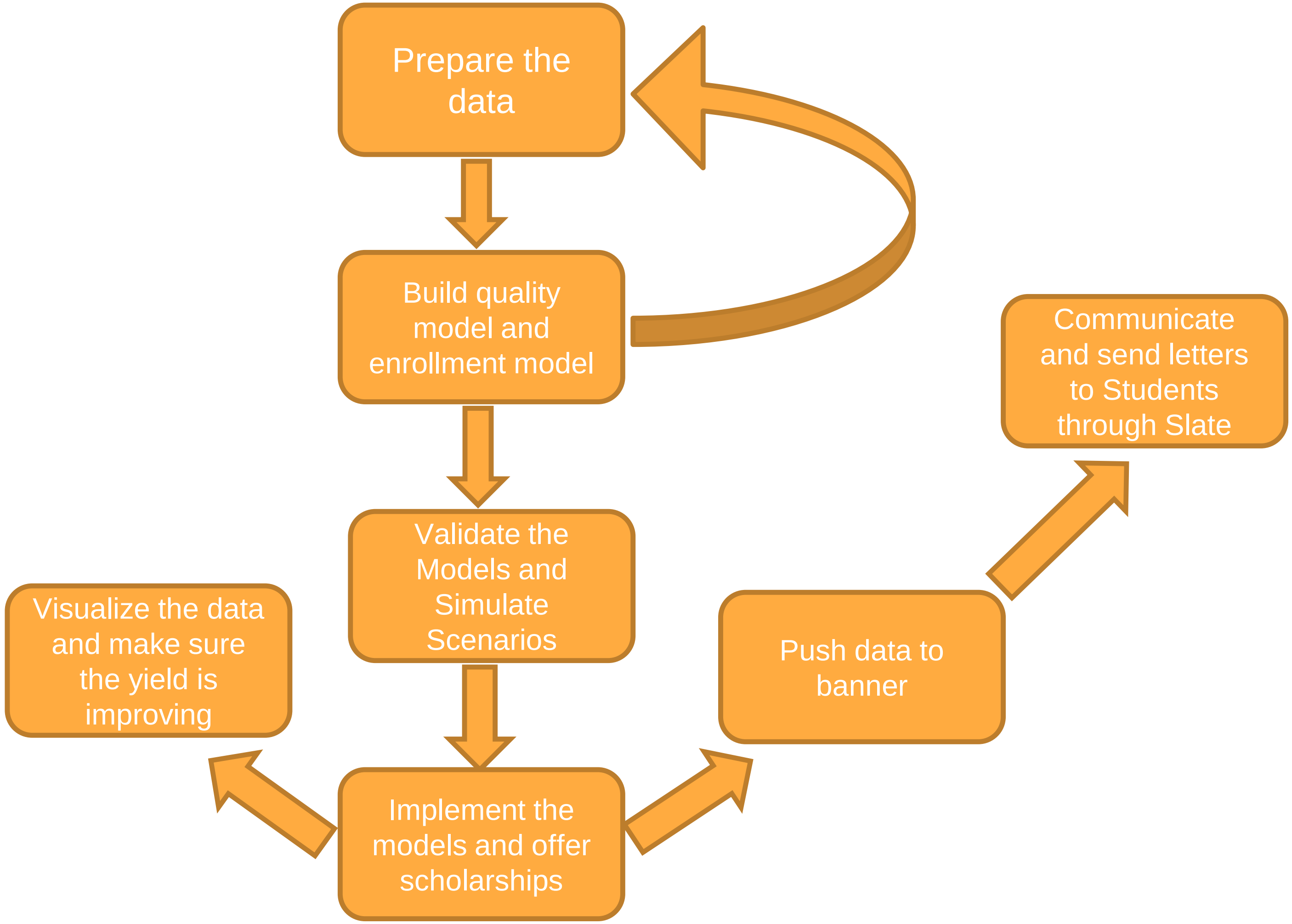
Admissions Quality & Enrollment Model

Logistic Regression



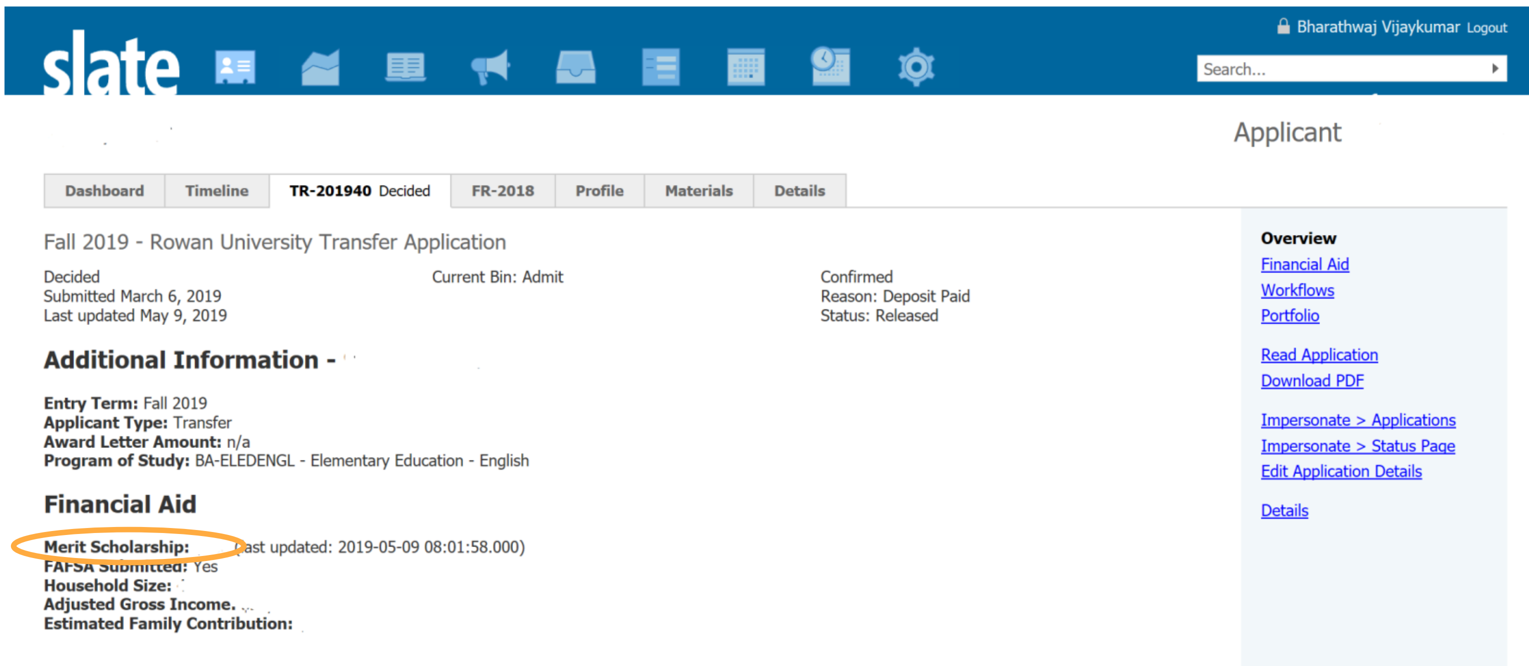
Predicted Y lies within the 0 or 1 range

Admissions Quality & Enrollment Model



Admissions Quality & Enrollment Model

Implementation



Admissions Quality & Enrollment Model

Implementation

Transfer Scholarships

	Predicted Enrollment Decile									
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	Grand Total
Number of Admitted Students with 12 credits or above	2	10	46	123	285	469	433	414	579	2,361
% of Total Admitted Transfer Apps w 12 or more credits	0.08%	0.42%	1.95%	5.21%	12.07%	19.86%	18.34%	17.53%	24.52%	100.00%
Predicted Enrollment	0	3	17	56	159	304	325	353	545	1,762
Deposited	2	2	24	70	188	334	329	326	482	1,757
Actual Enrolled	1	2	22	66	178	321	310	292	446	1,638
New Scholarship										
Offered Amount in Banner										
Predicted Spending										
Paid Amount in Banner										
Avg. High School GPA										
Avg. Distance from Rowan										
First Generation										
Avg. Yield by Zip in the Past	75.35%	61.35%	57.03%	65.36%	66.67%	68.99%	73.68%	73.56%	74.76%	71.40%

3. Bottleneck Course Analysis

Bottleneck Course Analysis

Goals

- Visualize the sequence of courses that students take in a specific program
- Visualize the effect of grades in one course, on the courses that follow that
- To identify the courses that effect the graduation time of students
- To tighten, or relax the prerequisite requirements based on the model recommendations
- To provide enough information to the advisors about the courses that the student has taken in the past, and the chances of that student getting a bad/good grade in the courses yet to be taken

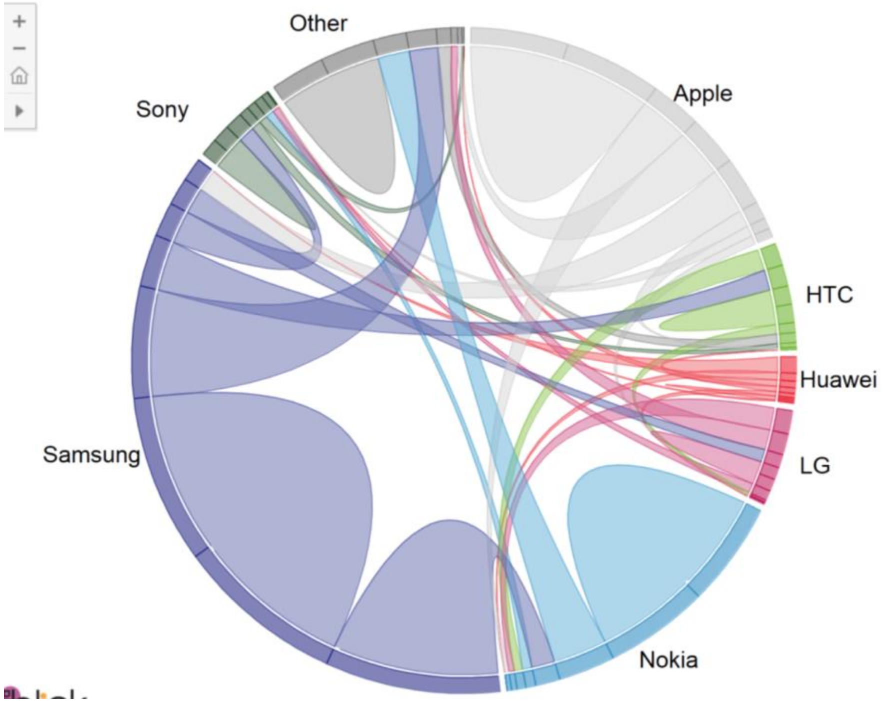
Bottleneck Course Analysis

Methodology Used

- Chord Diagrams
- Logistic Regression
- Apriori Model

Bottleneck Course Analysis

Chord Diagram



- Look at magnitude of a flow from point A to point B
- Visually pleasing
- Can handle cyclic bidirectional flows A to B to A

Bottleneck Course Analysis

Chord Diagram

- Chord Diagram used to show the sequence of courses that students take and grades they get.

- Visualize the effect of grades in one course, on the courses that follow that

- Since students can take courses out of order, there can be bidirectional flows.

https://rowan.shinyapps.io/Major_Pathways_Chord_Diagram/

Bottleneck Course Analysis

Logistic Regression

- To identify the courses that affect the graduation time of students

- To tighten or relax the prerequisite requirements based on the model recommendations

Bottleneck Course Analysis

Logistic Regression

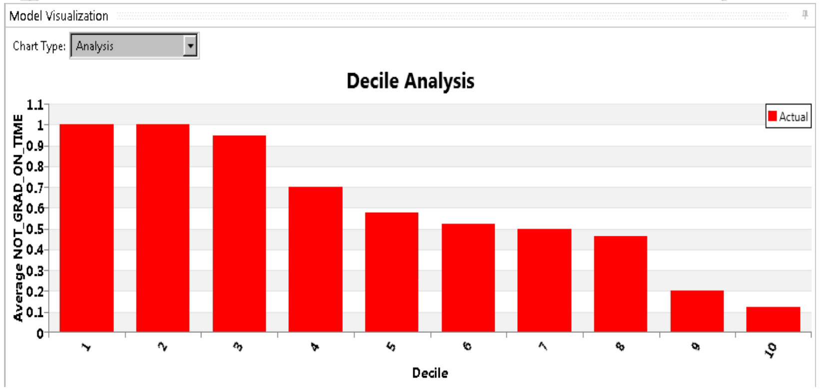
Suggested variables by analytics

Variables:
Binary(FINAL_GRADE_Bins_CS01205,Bad Grade)
Binary(FINAL_GRADE_Bins_CS04114,Bad Grade)
Binary(FINAL_GRADE_Bins_CS04222,Bad Grade)
Binary(FINAL_GRADE_Bins_CS04305,Good Grade)
Binary(FINAL_GRADE_Bins_CS04315,Good Grade)
Binary(FINAL_GRADE_Bins_CS04390,Good Grade)
Binary(FINAL_GRADE_Bins_CS07321,Good Grade)
Binary(FINAL_GRADE_Bins_CS07340,Bad Grade)
Binary(FINAL_GRADE_Bins_MATH01122,Bad Grade)
Binary(FINAL_GRADE_Bins_MATH01130,Bad Grade)
Binary(FINAL_GRADE_Bins_MATH01131,Bad Grade)
Binary(FINAL_GRADE_Bins_MATH01210,Bad Grade)
Binary(FINAL_GRADE_Bins_MATH03160,Good Grade)
Binary(FINAL_GRADE_Bins_PHIL09130,Bad Grade)
Binary(FINAL_GRADE_Bins_STAT02290,Bad Grade)
Binary(FINAL_GRADE_CS04113,A)
Binary(FINAL_GRADE_CS04113,F)
Binary(FINAL_GRADE_CS04114,A)
Binary(FINAL_GRADE_CS04222,A)
Binary(FINAL_GRADE_CS04400,A)
Binary(FINAL_GRADE_CS06311,A)
Binary(FINAL_GRADE_CS07321,A)
Binary(INSTRUCTOR_LAST_NAME_CS04113,[
Binary(INSTRUCTOR_LAST_NAME_CS07210,
Binary(INSTRUCTOR_LAST_NAME_CS07321,
TOOK_COURSE_CS01205
TOOK_COURSE_CS04305
TOOK_COURSE_CS07320
TOOK_COURSE_MATH01122

Response Variable:
Not Graduate on time – 1;
Graduate on time - 0

Variables included in the final model

Included Variables:
Binary(FINAL_GRADE_Bins_CS04113,Bad Grade)
Binary(FINAL_GRADE_Bins_CS04400,Good Grade)
Binary(FINAL_GRADE_Bins_CS06205,Bad Grade)
Binary(FINAL_GRADE_Bins_CS06310,Bad Grade)
Binary(FINAL_GRADE_Bins_CS06311,Good Grade)
Binary(FINAL_GRADE_Bins_CS07210,Good Grade)
Binary(FINAL_GRADE_Bins_CS07320,Good Grade)
Binary(FINAL_GRADE_CS04315,A)
Binary(FINAL_GRADE_MATH03160,F)
Binary(INSTRUCTOR_LAST_NAME_STAT02290,)



<B- is bad grade
>=B- Is good grade

EdgeCon ANNUAL CONFERENCE 2020

Bottleneck Course Analysis

Logistic Regression

Percentage contribution of variables

Variable	Percentage Model Contribution
Binary(FINAL_GRADE_Bins_CS04113,Bad Grade)	20.13 %
Binary(FINAL_GRADE_MATH03160,F)	16.80 %
Binary(FINAL_GRADE_Bins_CS06205,Bad Grade)	11.60 %
Binary(FINAL_GRADE_Bins_CS07210,Good Grade)	10.19 %
Binary(FINAL_GRADE_CS04315,A)	10.11 %
Binary(FINAL_GRADE_Bins_CS06310,Bad Grade)	9.73 %
Binary(INSTRUCTOR_LAST_NAME_STAT02290,)	6.59 %
Binary(FINAL_GRADE_Bins_CS07320,Good Grade)	5.70 %
Binary(FINAL_GRADE_Bins_CS06311,Good Grade)	4.91 %
Binary(FINAL_GRADE_Bins_CS04400,Good Grade)	4.23 %

Coefficients and P-values

Variable	Coef	S.E.	Wald chi-sqr	p-value
Intercept	0.4212	0.1602	6.911	0.00857
Binary(FINAL_GRADE_Bins_CS04400,Good Grade)	-1.226	0.4266	8.262	0.00405
Binary(FINAL_GRADE_Bins_CS04113,Bad Grade)	2.014	0.3908	26.56	0.000000
Binary(FINAL_GRADE_Bins_CS06205,Bad Grade)	1.441	0.4179	11.89	0.000565
Binary(FINAL_GRADE_Bins_CS07210,Good Grade)	-1.278	0.4290	8.877	0.00289
Binary(FINAL_GRADE_MATH03160,F)	2.276	1.092	4.345	0.03712
Binary(FINAL_GRADE_Bins_CS07320,Good Grade)	-1.161	0.4678	6.164	0.01304
Binary(INSTRUCTOR_LAST_NAME_STAT02290,)	-0.8914	0.5096	3.059	0.08027
Binary(FINAL_GRADE_Bins_CS06310,Bad Grade)	1.288	0.5092	6.401	0.01141
Binary(FINAL_GRADE_CS04315,A)	1.396	0.5428	6.611	0.01014
Binary(FINAL_GRADE_Bins_CS06311,Good Grade)	-0.8920	0.4927	3.277	0.07024

EdgeCon ANNUAL CONFERENCE 2020

Bottleneck Course Analysis

Apriori Algorithm

- If a student earns a poor grade in a course, what is the chance of that person getting a good or poor grade in the following courses?
- How are the course grades affecting the future courses?
- Does it match the prerequisites that the program has set?
- Compare with their current required courses to see if the pre requisite grades has to be lowered or increased

Bottleneck Course Analysis

Apriori Algorithm

- Used generally to predict which items are bought together

For ex: Beer-Chips

Laundry detergent-Dryer sheets

- Can we adapt that to courses to see which courses and grades usually occurs together for a person?

Support {🍎} = $\frac{4}{8}$

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

This says how popular an item is, as measured by the proportion of transactions in which an item appears.

Confidence {🍎 → 🍺} = Support {🍎, 🍺} / Support {🍎}

This says how likely item Beer is purchased when Apple is purchased? This does not account for how popular the item is. It may misrepresent the relationship.

Lift {🍎 → 🍺} = Support {🍎, 🍺} / (Support {🍎} x Support {🍺})

This says how likely beer is purchased when item apple is purchased, while controlling for how popular beer is. Lift should be greater than 1

Source: http://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html

Bottleneck Course Analysis

- Apriori will give an antecedent and consequent along with confidence, support and lift
- It will tell us which courses-grades combination occurs often and likely to occur

Consequent	Antecedent	Instances	Support %	Confidence %	Lift
MATH03160B = T	CS04113B = T MATH01130B = T	82	21.189	80.488	1.519
MATH03160B = T	CS04113B = T MATH01131B = T	59	15.245	81.356	1.536
MATH03160B = T	CS04114B = T MATH01130B = T MATH01131B = T	45	11.628	82.222	1.552
MATH03160B = T	CS04113B = T CS04114B = T	46	11.886	82.609	1.559
MATH03160B = T	CS04113B = T MATH01130B = T MATH01131B = T	39	10.078	87.179	1.646
CS04113G = T	MATH01130G = T MATH03160G = T	65	16.796	80.0	1.749
CS04113G = T	MATH01130G = T CS04114G = T	65	16.796	80.0	1.749
CS04113G = T	CS06310G = T CS04222G = T	75	19.38	80.0	1.749
CS04113G = T	CS04390G = T CS04114G = T	80	20.672	80.0	1.749
CS04113G = T	STAT02290G = T MATH01130G = T CS04315G = T	40	10.336	80.0	1.749

4. Fall to Spring Attrition Model

Fall to Spring Attrition Model

Goals

1. Determine the probability of attrition from the first Fall to Spring semester
2. Increase the retention rates and help students graduate on time

Methodology Used : Logistic Regression

Variables Used

1. Unmet need
2. SAT
3. Scholarship received
4. High School GPA
5. Days between app to enroll & App to deposit
6. Residency
7. Program
8. Cost of attendance

Fall to Spring Attrition Model

Actual Vs Predicted Attrition

Probability Decile	Enrolled in Fall	Predicted Not Returned	Sum of Actual Not Returned
1	245	63	38
2	244	45	34
3	244	29	33
4	244	21	26
5	244	17	17
6	244	14	13
7	244	12	16
8	244	11	8
9	244	8	10
10	244	4	4
Grand Total	2,441	225	199

Fall to Spring Attrition Model

Attrition by Admit Type

Admit Code	Enrolled in Fall	Predicted Not Returned	Sum of Actual Not Returned
EO	158	14	18
FR	1,009	66	52
IN	9	1	0
MA	51	4	5
SA	1,213	140	124
TR	1	0	0
Grand Total	**2,441**	**225**	**199**

Questions?