

Josh Chan, Matthew A. Lanham, James C. Reeder
 Purdue University Krannert School of Management
 chan151@purdue.edu; lanhamm@purdue.edu; jreeder@purdue.edu

Abstract

This study focuses on understanding how various institutions can incorporate data analytics into recruitment practices in order to convert more admitted students to matriculating students that would ultimately accept their offers.

While students are trying to get into their desired colleges, institutions are trying to recruit the right students in order to remain competitive in the education market.

I developed a solution that integrates clustering and predictive modeling that allows an institution to identify its target segment and understand the underlying needs of those potential students.

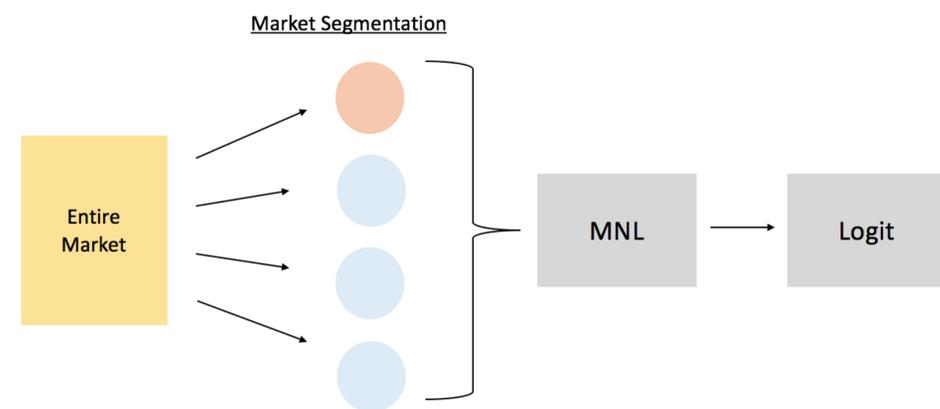
Introduction

The aim of this study is to provide data-driven recommendations for school recruitment offices to tailor their marketing expenditures on potential students that are most likely to attend their institutions.

Two internal datasets from Krannert were used in order to predict what would actually happen in reality. I first used cluster analysis to understand what are the factors that were separating potential students into different groups and identified the target segment. I then used a multinomial logit model to understand the unique characteristics of different segments and strived to understand the concerns of each group. I finally used a logit model to predict how likely an individual would attend the institution based on given characteristics of the target segment.

By focusing on those significant features from the causal relationship model, the business recruitment office can now package its marketing approaches with elements that are meaningful and valuable to those students.

Methodology



Methodology

Data Sources

Data sets were collected by the Admission Office and distributed to Marketing Analytics students in the Fall of 2017

- The Project Dataset contained 541 rows and 44 columns
- The Choice Dataset contained 4479 rows and 42 columns

1) Data Cleansing

- Remove Irrelevant Variables
- Missing Entries: Imputation with SAS Enterprise Miner (SAS leveraged all the data in our dataset and made educational guesses for the missing entries)

2) Market Segmentation—Cluster Analysis

As marketing analyst, I tried to locate the target segment before conducting any analysis. From the Project Dataset, I realized students were being asked two types of questions: broad and specific. For example, students were first asked to rate the “Career Preparation” aspect of Krannert. Then, the broad aspect would be broken down into several subcategories such as “Career Job Opportunity,” “Career Alumni,” and “Career Internship,” and students were asked to rate them accordingly.

I utilized two cluster analysis to break down the population into 4 segments. First we conducted a cluster analysis on the broad categories to discover the 3 most important separating factors. After that, I drilled down to the subcategories of those 3 most important separating factors and conducted another cluster analysis with the intention to identify the true causes of separation within the population.

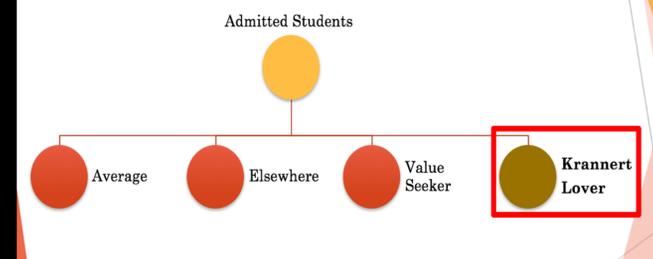
3) Differentiating Characteristics – Multinomial Logit Model

There were reasons why individuals belonged to certain groups and cluster analyses were used to identify the factors that drove students into various segments. After that, I picked a selection of variables from the Choice Dataset such as GPA Quartiles, Scholarship, Gender, and Legacy to feed into the multinomial logit model and tried to discover the differentiating characteristics of the target group comparing to the other segments.

4) Logit Model

Up to this point, I had been mainly performing the analysis on students from the Project Dataset (n=540) since it was smaller in size compared to the Choice Dataset (n=4478). After knowing the differentiating characteristics of the target group, I came up with several profiles with different demographics, and predicted the probability of an individual attending Krannert School of Management by using the Choice Dataset (larger number of observations). The result from the logit analysis not only allowed the office to have a numeric percentage in terms of enrollment, but also showed officers the variables that were significant to enrollment.

Results

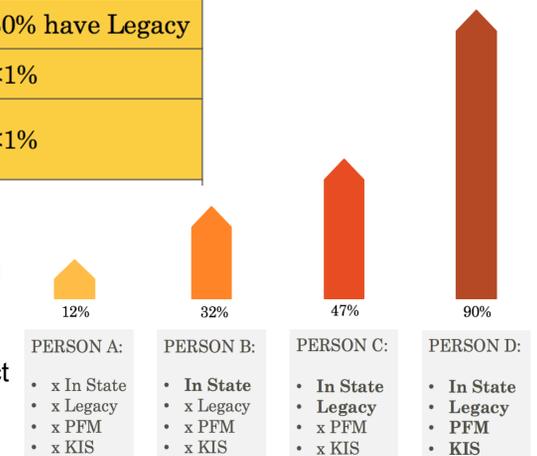


Based on the results from cluster analysis and pivot tables, I picked “Krannert Lover” as the target segment.

| Variable | Outcome |
|-----------------------------------|------------------|
| Gender | 50% |
| In State (+) | 35% are In State |
| Average GPA Quartiles (-) | 2.838 |
| Average Math GPA Quartiles | 2.405 |
| Average English GPA Quartiles | 2.706 |
| Legacy (+) | 30% have Legacy |
| Purdue For Me (+) | <1% |
| Krannert Information Sessions (+) | <1% |

With known target segment, I used pivot table to understand the characteristics of “Krannert Lover” and utilized those features as inputs for logit and multinomial logit model.

Four student profiles were created to test the strength of the logit model. As we can see, orientation programs have a significant impact on whether students would eventually enroll in Krannert.



Conclusions

Given the results, I recommended the business recruitment office to target the group “Krannert Lover.” The business recruitment office is recommended to focus on students that are both in-state and have legacy in their families. Moreover, they should motivate students to attend both “Purdue For Me” and “Krannert Information Sessions. Messages such as Career Job Opportunity, Alumni, Physical Safety, and Affordable Financial Aids should be incorporated and reinforced in all recruiting materials.

Acknowledgements

Great thanks to Professor James Reeder and Matthew Lanham