Explaining and Predicting First Year Student Retention via Card Swipe Systems

Emergent Research Forum (ERF) Paper

Caleb Bradberry  
Radford University  
cbradberry@radford.edu  
Matt Wayman  
Radford University  
mwayman@radford.edu  
Jonathan Charnok  
Radford University  
jcharnock1@radford.edu  
Andrew Ray  
Radford University  
aaray@radford.edu  
Jaget Dhami  
Radford University  
jdhami@radford.edu  
Jeff Pittges  
Radford University  
jpittges@radford.edu

Abstract

Organizations know that retaining a customer is more cost effective than recruitment. Data analytics provide mechanisms for explaining and predicting customer retention. Challenges inherent to customer retention are even more pervasive within higher education, with freshman attrition being a significant problem. However, an untapped opportunity exists within most institutions’ existing systems infrastructure: identification card systems used to access campus facilities and campus services. Using knowledge from information systems, this research examines the potential of card swipe data to predict student retention. These ‘card swipes’, bound to a single student primary key in the database, provide a profile of student interaction in the campus environment that is rarely investigated by institutional researchers. This provides a new opportunity for furthering organizational intelligence within higher education by providing new dimensions of a student’s behavior. This research examines retention by social factors in addition to traditional academic factors.

Keywords

Student retention, point of sale systems, retention analytics

Introduction

First year student attrition in higher education is a challenge for many institutions (Thammasiri et al. 2014; Tinto 1975, 2007). There are ongoing efforts to address the problem of understanding the academic, social, and commercial factors that influence student retention (Bogard et al. 2011; Bukralia et al. 2012; Delen 2010). Within this body of work, efforts have been made to apply different forms of modeling and machine learning to understand the vast amount of data to which a university has access (Delen 2010; Palmer 2013). Considerable focus has been placed on the academic performance side of retention, but there remains a need to understand the social and environment factors (Aguiar et al. 2014; Tinto 2007).

One of the data points that most institutions of higher education have available is that of identification card (ID cards) information. Specifically, institutions use ID cards to allow students access to dorm and academic buildings. These cards are also used to make purchases, such as books at the beginning of a semester, or for campus meals. Effectively, every time a student’s ID card is swiped on campus, a new data point about that student is created. This data point captures data around: where they were on campus, what time, what door was used, was something purchased? Inherently, a university ID card swipe represents a social interaction on campus if that card is in-bound to a building.
This presents an opportunity for the business of higher education, for institutional researchers, and for any organization that may have a card swipe system: to what extent can the data points around a card swipe system be used to explain and predict retention? While this research examines card swipes within the context of higher education, it is reasonable that large firms that employ an identification card system would also find utility in leveraging the information stored in their ID card system database (Ma et al. 2013).

To answer this question, this research presents a brief review of the retention literature in terms of academic measures and social measures. Next, this emergent research paper shows a preliminary statistically significant result that retention can be explained and predicted by card swipe counts. The paper concludes by presenting the next steps in the research.

**Review of Literature**

Institutional research is a broad aspect of higher education that focuses on gaining organizational intelligence around the institution (Wilensky, 2015). While traditional organization analytics functions focus on transforming data into insight that further shareholder value, institutional research can leverage these analytical functions to gain organizational intelligence of each business unit of the institution (Fincher 1978).

Student attrition becomes a two-fold problem, with a loss of external funding directly tied to a loss of students, which also provide a source of internal funding(Tinto 2007). This becomes a very practical problem: schools losing students are also losing funding. It then becomes of paramount importance from an organizational perspective to be able to explain historical student attrition, and predict future attrition; information systems presents a set of tools for explaining and predicting data phenomena and positioning it within organizational insight with the work conducted in business intelligence and analytics (Basole et al. 2013; Chen et al. 2012; Phillips-wren et al. 2015). These tools allow institutions of higher education: to explain factors that led to historical attrition, as well as predict future attrition. To utilize business analytics, the organization must first know their data and have a general understanding of some of the factors that contribute to the phenomena they want to explain and predict. These factors are examined below.

**Performance Factors from Literature**

A performance factor for retention relates to a student’s performance in their academic pursuits in higher education. These are quantifiable and known post facto of attrition. Prior work in student retention typically focuses on academic performance factors. Table 1 presents a summary of the most common academic performance factors that are examined regarding student retention. Additionally, Table 1 includes a summary of the results of the research if that factor was supported, partially support, or not supported, and a summary list of references for the purposes of familiarizing the reader with some of the work being conducted. Partially supported is indicated as a promising factor, but needs further examination, per the author of the original work.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Supported</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School GPA; Mid semester College GPA</td>
<td>Yes.</td>
<td>(Bogard et al. 2011; Bukralia et al. 2012; Delen 2010)</td>
</tr>
<tr>
<td>SAT Composite Score;</td>
<td>Yes.</td>
<td>(Aguiar et al. 2014; Murtaugh et al. 1999; Thammasiri et al. 2014)</td>
</tr>
<tr>
<td>SAT Discrete Scores</td>
<td>Partially Supported.</td>
<td>(Aguiar et al. 2014; Murtaugh et al. 1999; Thammasiri et al. 2014)</td>
</tr>
<tr>
<td>Transfer Hours</td>
<td>Partially Supported.</td>
<td>(Thammasiri et al. 2014)</td>
</tr>
<tr>
<td>Declared Major</td>
<td>Partially Supported.</td>
<td>(Murtaugh et al. 1999)</td>
</tr>
</tbody>
</table>
Predicting Student Retention via Card Swipe Systems

<table>
<thead>
<tr>
<th>Factor</th>
<th>Supported</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman Orientation</td>
<td>Partially Supported</td>
<td>(Bogard et al. 2011; Delen 2010)</td>
</tr>
<tr>
<td>Financial Aid Amount</td>
<td>Partially Supported</td>
<td>(Delen 2010)</td>
</tr>
<tr>
<td>(Student Loans, Scholarship)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission Year</td>
<td>No.</td>
<td>(Delen 2010)</td>
</tr>
<tr>
<td>Fall Grant/Work Study</td>
<td>Partially Supported</td>
<td>(Delen 2010)</td>
</tr>
<tr>
<td>Admission Type</td>
<td>Yes.</td>
<td>(Aguiar et al. 2014)</td>
</tr>
</tbody>
</table>

Table 1. Common Performance Predictors of Retention.

Social, Commercial, Demographic Factors from Literature

There is a body of literature that suggests social factors are as important as academic factors in retaining students (Gerdes et al. 1994; Pritchard and Wilson 2003; Tinto 1975). This body of work examines how students are setup to succeed in terms of social support, social well-being, and having a network on which they can rely for help. Table 2 presents some of the most common factors from the seminal work on retention as a function of social factors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Supported</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>On campus resident</td>
<td>Partially Supported</td>
<td>(Aguiar et al. 2014; Gerdes et al. 1994)</td>
</tr>
<tr>
<td>First generation college student</td>
<td>Partially Supported</td>
<td>(Aguiar et al. 2014; Gerdes et al. 1994; Pritchard and Wilson 2003)</td>
</tr>
<tr>
<td>Age</td>
<td>Yes.</td>
<td>(Aguiar et al. 2014; Bogard et al. 2011; Delen 2010; Palmer 2013)</td>
</tr>
<tr>
<td>Ethnicity, Gender, Martial Status, In-State Student</td>
<td>Partially Supported</td>
<td>(Aguiar et al. 2014; Bogard et al. 2011; Delen 2010; Murtaugh et al. 1999)</td>
</tr>
</tbody>
</table>

Table 2. Common Social/Demographic Predictors of Retention.

What’s New?

Traditional retention research is focused on data easily accessible to higher education administrators and institutional researchers: academic and demographic data. However, this research proposes an alternative approach with on the prevalence of ID cards. Student retention should be examined through the lens environmental interaction. By pairing the ID card swipe data with commonly available student data, higher education organizations can gain insight to both sides of the retention problem. This kind of insight has been demonstrated successfully with ‘reward cards’ from grocery stores and retailers (Esposti 2014). Specifically, how can student retention be explained through data, as well as how can student attrition be explained through data. The next section presents preliminary results of the card swipe data.

Population, Datasets, and Preliminary Results

The population being analyzed is first year freshmen at a university of roughly 10,000 students. This dataset accounts for 5,692 college freshmen, of which 728 were not retained after their first semester at the university. Two datasets were used to study student retention based on card swipes: card swipe data itself, and traditional retention data described in the brief literature review.
The card swipe dataset contains the following columns: anonymized student ID, building ID, door ID, door description, and time and date. For the card swipe data, there were 102,187 card swipes of students who were not retained and 1,762,467 of students who were retained. A new column was created for aggregate count of swipes per student, called CardCount. CardCount was added to the second dataset, which contains all the fields described in the literature review.

The data was then modeled with student retention as a function of swipe count. With the scale of the data presenting variance inflation, the data was then sampled taking 300 rows of data at a time. The results are presented in Table 3. What is found is a statistically significant p-value on swipe count using a logit model. This tells us that the total number of swipes a student had on campus is a statistically significant predictor of student retention.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.3730</td>
<td>0.078</td>
<td>30.1720</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>CardCount</td>
<td>0.0001</td>
<td>0.0001</td>
<td>3.3620</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Table 3. Logit Results

![Figure 1. Card Swipes by Retention and Hour of Day.](image)

Additionally, the time of day that the card swipe was recorded was modeled. Figure 1 shows the result of modeling card swipes for retained and non-retained students. Figure 1 illustrates the spike in non-retained students and the times in which they are using their university ID card. The non-retained students pattern is activity between 1-4AM. This figure motivates further investigation of card swipe data as a mechanism by which universities, and organizations in general can leverage their existing ID card system data into organizational intelligence.

**Next Steps**

The initial statistics provide promising results. The results presented were only a small representation of the total amount of card swipe data available to the institution. The next step of this research is to add academic building data to the dataset. Point of sale data for student book purchases and meal plans is also contained in the card swipe database, and will be examined.

For future research and theoretical grounding, two more data sources are going to be added to this project to extend beyond card swipe systems to gain a more robust view of the social metrics of predicting student retention: learning management system (LMS) data, and WiFi usage data. LMS data will provide a more robust data set around when, what, and how the students access course content, and the analytics will provide metrics for predicting retention. The WiFi provides information on students’ network utilization. WiFi data usage will be used to provide information about where and what a student as doing on campus, e.g. if a student was signed onto a router in a lecture hall at 9:30am, and they are scheduled for a 9:30am
class, then that student was probably doing school work. Whereas a student at 2:00am at night using multiple gigabytes of data was probably focused on entertainment and not study. The goal of these new datasets is to provide more dimensions to predict student retention.

This paper provides promising results on a new dimension of predicting student retention. Due to the robustness of the initial results, more dimensions will be examined with these new data sets. While this paper frames the question of the efficacy of card swipes, future work looks to identify new dimensions for explaining and predicting student attrition and retention.

REFERENCES


