UNIVERSITY of VIRGINIA | DARDEN Business Publishing

UVA-QA-0864
Nov. 17. 2017

Retention Modeling at Scholastic Travel Company (A)

On a sunny Monday afternoon in early spring 2013, David Powell entered his new office and took a deep breath. He pondered his first few days as the new data analyst for Scholastic Travel Company (STC), an educational tourism firm. Powell had filled his first week of employment meeting the firm's departmental leadership and attending a company-wide new-employee-orientation program, and was eager to get started on his first project.

Just a few hours earlier, at the weekly marketing strategy meeting, Powell's new supervisor, Stephen Blackford, stressed the urgency of a new data initiative centered on customer retention. As Blackford outlined, in less than two weeks, contract renewal opportunities would begin for customers who went on an STC trip in 2012. During the meeting, he presented a dataset with all of the known information about the previous year's client base (see **Exhibits 1** and **2**). From his past experience, Blackford was confident that models could be constructed to predict whether or not a customer would book again in 2013. With such a model, he hoped to design a more nuanced marketing strategy that targeted certain subsets of the client population to save cost and improve yield. With multiple plausible methodologies in mind, Powell knew he needed to get to work immediately so he could give Blackford an accurate prediction model before the end of the week.

Company Background

STC was not a particularly young company. Founded in the 1960s, it grew from a single-person operation to a multi-million-dollar business, was bought and sold twice (first to the management when the founder retired, then to a private equity firm), and survived almost having to declare bankruptcy post-9/11. Yet as of 2013, it was one of the premium providers of cultural and educational trips: history and science trips to middle- and high-school students, exchange trips for university students, cultural immersion, artistic destination trips, and other tours worldwide.

Customers chose STC because of its superb ability to coordinate the numerous details associated with taking a large group of primarily young people on a far-away journey. These included the procedures related to obtaining proper documents and permits (e.g., visas), logistical details (bus, plane, train, and other tickets), meet-and-greet at the transfer points and destinations; hotel, meal, and entertainment bookings; taking care of safety concerns (chaperons and accompanying security guards to ensure physical safety and eliminate the possibility of sexual, emotional, and substance abuse); insurance; and accident "resolution" (e.g., searching for missing travelers[1] or replacing a lost passport). All were critical elements of a successful trip, which the school teachers,

[1] As a company executive put it, "We never lost a student…permanently."

This disguised case was prepared by Anton Ovchinnikov, Distinguished Faculty Professor of Management Science and Operations Management and Scotiabank Scholar of Customer Analytics at the Smith School of Business, Queen's University, Canada.

university administrators, parents, and students themselves were glad to outsource to trusted professionals, and STC was a prime example.

The majority of the trips STC managed were of the "teacher organized, parent paid" type. This meant that the teacher (or university administrator) determined the itinerary, desired duration, and activity schedule, but the parents (or students) paid for the trip. For that purpose, it was not uncommon for the teachers/administrators held meetings with parents/students prior to the trip. STC typically kept track of those meetings, as they often revealed important information about the upcoming trips. STC representatives would often attend such meetings, either in person or virtually.

STC also collected and carefully tracked multiple types of data about the travel group and the organizing teacher/administrator, and it sought feedback after the trip. This was all recorded in the STC cloud-based database and was easily accessible to Powell.

Prediction Task and Available Data

Powell's ultimate task was to predict which customers would book with STC in the 2013–14 school year (fall 2013 to spring 2014). He decided to build a model that took the data available as of spring 2013 to make this prediction. To build such a model, however, Powell would need to replicate the 2013–14 prediction task on the available data. This meant that for training his model, he would use the data from the 2012–13 school year—which showed whether a certain group had been retained or not—and try to predict based on the client-profile information as of the end of the 2011–12 school year. He was lucky that STC took snapshots of customer-profile data once per year, so this historical data was available (see **Exhibit 1**).

Powell knew from the marketing meeting that the company used post-trip surveys in order to track performance and get feedback from the teachers. These responses, coupled with trip data such as trip revenue, trip length, and school size were what he would use to construct his retention model. A comprehensive list of data fields was included with the spreadsheet that Blackford sent to Powell, and is listed in **Exhibit 2**. With a sample size of nearly 2,400 groups, Powell was hopeful that he could have a model that would make reasonably accurate predictions for Blackford before the end of the week, so that the new marketing strategy could be deployed before the sales season started later in the spring.

Exhibit 1

Retention Modeling at Scholastic Travel Company (A)

Snapshot of the Data

(First five and last five entries shown; to fit the snapshot on a single page, several data fields are hidden)

ID	Program.Code	From.Grade	To.Grade	Group.State	Is.Non.Annual	Days	Travel.Type	Departure.Date	Return.Date	Deposit.Date	Special.Pay	Tuition	.	Retained.in.2012
1	HS	4	4	CA	0	1	A	14/01/2011	14/01/2011	30/08/2010	NA	424	.	1
2	HC	8	8	AZ	0	7	A	14/01/2011	21/01/2011	15/11/2009	CP	2350	.	1
3	HD	8	8	FL	0	3	A	15/01/2011	17/01/2011	15/10/2010	NA	1181	.	1
4	HN	9	12	VA	1	3	B	15/01/2011	17/01/2011	07/01/2011	NA	376	.	0
5	HD	6	8	FL	0	6	T	16/01/2011	21/01/2011	30/09/2010	NA	865	.	0
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2385	HC	7	8	CA	0	5	A	28/06/2011	02/07/2011	15/12/2010	NA	1892	.	0
2386	HD	8	8	CA	0	5	A	29/06/2011	03/07/2011	15/10/2010	FR	1699	.	1
2387	HD	10	12	CA	0	6	A	29/06/2011	05/07/2011	18/01/2011	SA	2149	.	1
2388	HS	4	4	CA	0	1	A	30/06/2011	30/06/2011	17/12/2010	NA	449	.	1
2389	HD	8	8	WA	0	6	A	30/06/2011	05/07/2011	29/10/2010	NA	2135	.	1

Note: The full dataset is available in the accompanying student spreadsheet, UVA-QA-0864X.

Data source: Company data adjusted by author using unspecified constants.

Exhibit 2

Retention Modeling at Scholastic Travel Company (A)

Data Dictionary

Data Field Name	Example	Description
ID	1	Self-explanatory.
Program.Code	HN	This is a very granular code that describes where the trip went and what it did. HN, for instance, is a history program that runs in New York.
From.Grade	8	This is the lowest grade in school of a participant on that program.
To.Grade	8	This is the highest grade in school of a participant on that program.
Group.State	IN	This is the two-letter designator for the state in which the originating school is located. OTHER stands for rare geographies that appear in the data only once.
Is.Non.Annual.	1	1/0 indicating if the group from this school typically skips a year in between programs. These will rarely repeat the very next year.
Days	3	The number of days the group was on the program and with one of the instructors.
Travel.Type	A	Mode of travel from the originating school location to the starting location of the program (A = Air, B = Bus, T = Train).
Departure.Date	19/02/2011	The date that the group left its originating school.
Return.Date	21/02/2011	The date the group returned to its originating school.
Deposit.Date	20/10/2010	The date by which registrants are supposed to have at least an initial deposit in prior to departure. The time in the school year when certain events occur can be important; for instance, there are no deposit dates in the summer since no one would be around to act on them.
Special.Pay	NA	The most important of these are school accounts (SA). That means that, contrary to the usual practice, the teacher collects all of the money and then remits it in bulk to STC. The normal arrangement is STC handling all of the cash collection from parents/students.
Tuition	1174	This is the price it costs each full-paying participant (FPP) to go on the program. West-coast air trips are more expensive per person than midwestern bus groups.
FRP.Active	72	FRP is the full refund program. This is the number of FPPs on the trip who bought trip-cancellation insurance.

Exhibit 2 (continued)

FRP.Cancelled	13	This is the number of FPPs on the trip who bought trip-cancellation insurance, but then cancelled it.
FRP.Take.up.percent.	0.6857	This is the percentage of the FPPs who bought the FRP and ended up paying for it.
Early.RPL	02/03/2010	This is the date that the first communication went out to the group. Often this can be 12 to 18 months before the trip actually departed.
Latest.RPL	10/08/2010	This is the date that the last communication inviting people to join the group went out. Often this can be 6 to 9 months before the trip actually departed.
Cancelled.Pax	15	This is the number of passengers who signed up with a $100 deposit but then cancelled before the group departed.
Total.Discount.Pax	7	This is the total number of extra passengers who went along without paying full price (or typically anything). These would be the chaperones and the teachers.
Initial.System.Date	02/03/2010	This is the date when the teacher first agreed to get this trip organized. It is typically the earliest of the dates relative to group activities.
Poverty.Code	A	Poverty code for the area in which the originating school (and by extension, most of the parents who will be paying for the trip) resides based on estimated percentage below the poverty line. A is 0 to 5.9, B is 6 to 15.9, C is 16 to 30.9, D is 31 or more, E is unclassified, Space if DISTCLASS = U (Supervisory Union).
Region	Other	This is a larger aggregation of state areas. Some large states, like California, are their own region. Others are combined.
CRM.Segment	1	This is a type of school code used in the customer-relationship-management (CRM) system to describe the school. The codes are numbered 1–11 but are in no particular order; proprietary, but it is a designation of a customer type that may be helpful.
School.Type	PUBLIC	Public or private.
Parent.Meeting.Flag	1	1/0 indicating whether a parent meeting was held. These are typically strong indicators of parent engagement and of a teacher who understands that these can be important to successfully organizing one of these out-of-school programs.
MDR.Low.Grade	7	This is the lowest grade in the originating school.
MDR.High.Grade	8	This is the highest grade in the originating school.
Total.School.Enrollment	955	This is the total enrollment of the school (to differentiate big schools from little ones).

Exhibit 2 (continued)

Income.Level	P	Like poverty code, an indication of ability of parents to pay for these programs. A is lowest, Q is highest, Z is unclassified.
EZ.Pay.Take.Up.Rate	0.2286	This is a % of the FPPs that sign up for an automatic bank draft installment plan.
School.Sponsor	0	This is an indication (1/0) of whether or not the school is officially sponsoring the trip. Mostly, though these programs draw from the same school, they are typically run independently.
SPR.Product.Type	East Coast	A high level of aggregation of the very granular tour types.
SPR.New.Existing	EXISTING	EXISTING means that the group has traveled with STC before—most often the year before. NEW, with few exceptions, means that the school has never traveled before with STC.
FPP	105	This is the actual number of FPPs who went on the trip.
Total.Pax	112	This is the actual number of total passengers (including chaperones and teachers) who went on the trip.
SPR.Group.Revenue	125735.4	This is the total amount paid for all of the participants to go on the program from that group.
NumberOfMeetingswithParents	0	Number of meetings with parents prior to the trip.
FirstMeeting	18/11/2010	The date of the first meeting with parents (NA if none held).
LastMeeting	28/11/2010	The date of the last meeting with parents (NA if none he.ld, may be same as the first meeting if only one meeting was held)
DifferenceTraveltoFirstMeeting	93	The number of days from the first parent meeting to travel date.
DifferenceTraveltoLastMeeting	103	The number of days from the last parent meeting to travel date.
SchoolGradeTypeLow	Elementary	The lowest grade type in the school.
SchoolGradeTypeHigh	Elementary	The highest grade type in the school.
SchoolGradeType	Elementary->Elementary	Combination of the above denoting the type of school.
DepartureMonth	January	Month of departure.
GroupGradeTypeLow	K	The lowest grade type in the group that travels.
GroupGradeTypeHigh	Elementary	The highest grade type in the group that travels.
GroupGradeType	K->Elementary	Combination of the above denoting the type of the group that travels.

Exhibit 2 (continued)

MajorProgramCode	H	Aggregation of the granular program code; the first letter of the program code.
SingleGradeTripFlag	1	Indicator for the trip taken by a group comprising students from the same grade.
FPP.to.School.enrollment	0.06364617	The ratio of FPP to school enrollment.
FPP.to.PAX	0.93650794	The ratio of FPP to total PAX on the trip.
Num.of.Non_FPP.PAX	4	The number of PAX who are not FPP.
SchoolSizeIndicator	L	A label for the size of the school (S, M, L, S-M, M-L), by quintiles of sizes.
Retained.in.2012.	1	THIS IS THE 1/0 SUCCESS METRIC WE ARE TRYING TO PREDICT—DID THE GROUP ACTUALLY RETURN THE NEXT YEAR?

Data source: Company data, adjusted by author.