Pontificia Universidad Javeriana
Facultad de Ingeniería
Departamento de Ingeniería Industrial

Métodos de Analítica II
Caso de Negocio

En el curso de Métodos de Analítica II de la Maestría en Analítica para la Inteligencia de Negocios, los estudiantes aprenderán herramientas analíticas valiosas que pueden ser aplicadas en un contexto empresarial para resolver problemas reales. Para enriquecer su aprendizaje, se llevará a cabo un caso de negocio a lo largo del semestre, que permitirá a los estudiantes aplicar los conceptos teóricos y las técnicas analíticas a un problema específico de un cliente.

El proceso de solución de un caso de negocio incluye la identificación del problema del cliente, la exploración y pre-procesamiento de datos, la evaluación de diferentes modelos analíticos y la selección de la mejor alternativa de solución, así como su monitoreo y control continuo. Estos pasos se corresponden con el enfoque en el ciclo de vida de datos de CRISP-DM y demuestran la importancia de un enfoque sistemático y riguroso para la solución de problemas de negocios a través de la analítica.

Después de solucionar un caso de negocio a través de un modelo analítico, es esencial generar un informe que resuma la solución para presentarla al cliente. Algunos de los elementos que debería incluir el informe son:

- **Resumen Ejecutivo:** Resumen del problema del cliente y la solución propuesta.
- **Descripción de la Necesidad:** Explicación de la necesidad o problema que se busca resolver.
- **Datos y Exploración:** Descripción de los datos obtenidos, su procesamiento y partes relevantes del Análisis Exploratorio de Datos.
- **Metodología Analítica:** Una sección que describa las metodologías utilizadas, incluyendo las métricas de evaluación.
- **Solución Propuesta:** Una vez finalizado, se debe detallar la solución que permite resolver la necesidad del cliente, sus métricas, su implementación, forma de uso.
- **Conclusiones y Anexos**

Pontificia Universidad Javeriana
Facultad de Ingeniería
Departamento de Ingeniería Industrial

Retention Modeling at Scholastic Travel Company

Cliente

La *Scholastic Travel Company* (STC) es una compañía de turismo educativo con más de 50 años de experiencia en la organización de viajes culturales y educativos. La agencia ofrece una amplia gama de viajes, incluidos viajes de historia y ciencia para estudiantes de secundaria y universitarios, inmersión cultural y viajes artísticos en destinos de todo el mundo. Los clientes eligen STC por su habilidad en coordinar los detalles logísticos y de seguridad de los viajes. La mayoría de los viajes son organizados por profesores y pagados por padres, y STC recopila y registra información sobre el grupo de viajeros y el profesor organizador.

Problema de Negocio

STC desea mejorar su eficacia en el Marketing y reducir costos, ya que sus campañas actuales están teniendo una baja tasa de conversión. El objetivo principal de STC es determinar cuáles son los clientes más propensos a no organizar un nuevo viaje, con el fin de crear campañas de retención efectivas.

Datos disponibles

STC cuenta con una base de datos histórica que incluye información sobre todos los grupos que han viajado con ellos desde su fundación. Para el desarrollo del modelo de predicción, se proporcionará una tabla con información sobre los viajes realizados en el año 2011, junto con los datos de retención en 2012, con un total de 2389 registros.

Adaptado de: Ovchinnikov, A., Scholar, S., Ovchinnikov, A., & Scholar, S. (2017). Retention Modeling at Scholastic Travel Company (A). *Darden Business Publishing Cases, 1-7*.

Exhibit 1

Retention Modeling at Scholastic Travel Company (A)

Snapshot of the Data

(First five and last five rows of data 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	HD	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)

Data Dictionary

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)

Data Dictionary

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 held, 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)

Data Dictionary

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.