Ivey Publishing

VLERICK BUSINESS SCHOOL

W27305

ALLIANZ: OPTIMIZING CUSTOMER ACQUISITION STRATEGY USING MACHINE LEARNING

Bjarne Brié, Tineke Distelmans, Kristof Stouthuysen, Tim Verdonck, Christopher Grumiau, and Thoppan Mohanchandralal Sudaman wrote this case solely to provide material for class discussion. The authors do not intend to illustrate either effective or ineffective handling of a managerial situation. The authors may have disguised certain names and other identifying information to protect confidentiality.

On a rainy evening in October 2019, Thoppan Mohanchandralal Sudaman, the regional chief data and analytics officer at Allianz Benelux (Allianz) was returning home from the office, lost in thought. He could not stop thinking about his two-hour strategy meeting with Allianz 's regional chief executive officer, Anthony Bradshaw. During the meeting, Bradshaw had expressed concerns about Allianz's digitalization strategy. Bradshaw told Thoppan Mohanchandralal that a few days earlier, he had been briefed by the marketing department about online sales channel results having fallen below expectations.

Bradshaw was worried that Allianz could lose market share if it did not react accordingly, which would damage the company's competitive position in the market. Therefore, Bradshaw had asked Thoppan Mohanchandralal to gather a team to investigate why online sales were low and to design an effective customer acquisition strategy. Thoppan Mohanchandralal suggested that in addition to his data office staff, the business transformation unit should be asked for assistance. Bradshaw agreed and proceeded to his next appointment. While walking home, Thoppan Mohanchandralal was thinking about how he should approach this challenging task.

ALLIANZ

Allianz Group, a multinational financial services group, was founded in 1890 in Berlin, Germany. Allianz Group provided life and health insurance, property and liability insurance, and asset management products and services in over seventy countries to more than eighty-eight million customers, making it the third-largest financial services provider in the world.

Striving to embrace the power of innovation and the added value of digitalization, Allianz Group established the Allianz Benelux data office in Belgium. With more than forty data experts, this office was directed by Thoppan Mohanchandralal as regional chief data and analytics officer. He was responsible for transforming the Allianz Benelux data office into a more data-driven company. Thoppan Mohanchandralal had been working at Allianz for four years and reported directly to Allianz's regional chief executive officer.[1]

[1] In July 2021, Joos Louwerier replaced Bradshaw as regional chief executive officer of Allianz.

Thoppan Mohanchandralal and his team at the data office worked closely with the different business departments. In particular, they used business intelligence and machine learning to improve pricing, to detect and prevent fraud, to simplify operations, to handle claims, and to enhance customer and broker centricity. The purpose of focusing on customer and broker centricity was to develop new products and services for the different brokerage channels and their customers. It was also important to optimize the use of currently existing sales channels, with the ultimate goal of providing customers with the best possible advice and services. It was exactly because of the data office's expertise within this segment that Bradshaw had charged Thoppan Mohanchandralal to improve the online sales channel's results.

SALES AT ALLIANZ

Traditionally, Allianz sold insurance through brokers that operated mainly as independent entities, so Allianz rarely worked directly with the customer.[2] However, as many retail businesses were increasingly undergoing transformation into digital sales channels, the company's traditional sales model was also changing. Allianz sold insurance mainly in two ways: through brokers, many of whom set up their own digital sales portal; and through collaborations with third parties known as *affinities*, who benefited financially from insurance sales. Three such affinities with whom Allianz worked closely were Insuro Inc. (Insuro), Seguros International Ltd. (Seguros), and T&B plc (T&B).[3]

T&B sold mostly health insurance, although Allianz also relied on T&B to sell auto insurance policies using an online quote-and-buy system. Potential customers could receive an immediate auto insurance estimate by simply providing all the relevant information using the T&B online portal. Customers were able to immediately and directly receive a quote estimate for a specific insurance policy. More importantly, they could then choose to proceed to purchase the policy online, avoiding the need for time-consuming appointments with the insurance company.

Despite the apparent convenience and attractiveness for the customer, however, Allianz saw a relatively low convergence rate through this sales channel, approximately 12 per cent lower than Insuro and Seguros (see Exhibit 1). In particular, it seemed that although many potential customers were visiting the T&B website and entering all relevant information, most of them dropped out of the transaction process before the final step—purchasing an insurance policy. Clearly, for Allianz to increase sales, it was essential to improve the customer convergence rate on the T&B sales channel. But as Thoppan Mohanchandralal realized, the first step was to learn more about the profile of potential customers on the online channel. What was driving their decision to buy insurance? Equally important, Allianz needed to learn what caused potential customers to abandon the purchase process before reaching the final stage.

DESIGNING A CUSTOMER ACQUISITION STRATEGY

Thoppan Mohanchandralal had worked on customer-centric projects with the data office in the past, but he realized that more knowledge was needed on how to design a customer acquisition strategy. Therefore, the first thing he did the next morning in the office was schedule a meeting with the marketing department.

The vice-president (VP) of marketing at Allianz explained to Thoppan Mohanchandralal and his team that customer conversion was an important part of an organization's customer acquisition strategy. He emphasized that, particularly in a highly competitive online retail environment characterized by low

[2] Only health insurance could be sold through direct sales.
[3] For the purpose of this case, the names of these affinities were disguised.

switching costs, a good customer acquisition strategy could significantly impact an organization's bottom line. Also, an effective customer acquisition strategy should not only help the business identify and reach new potential customers, but should also convince consumers to buy the product or service the business was offering. The VP also stressed that a successful customer acquisition strategy leveraged data to gain insight into customer behaviour and to identify interesting marketing opportunities.

DATA

Taking into account the advice of the VP of marketing, Thoppan Mohanchandralal's team started to collect all available and relevant data that could possibly be of use. In total, the team collected three main data sets: funnel, policy, and regional.

Funnel Data Set

The funnel data set (see Microsoft Excel file "Funnel Data Set," Ivey product no. W27307) contained information about quotes that were given through the various sales channels (i.e., Insuro, Seguros, T&B) starting in October 2018. Each row in the data set represented a quote given by Allianz through one of its sales channels. It contained information about the status of the quoting process, the premium of the policy, and the type of insurance that was requested by the customer (see Exhibit 2).

Policy Data Set

The policy data set (see Microsoft Excel file "Policy Data Set," Ivey product no. W27308) contained all Allianz insurance policies that were still outstanding through T&B (as of October 2019). Each row represented one contract. The data set contained information about the policy holder, the insured vehicle, the type of insurance, and various other specific details (see Exhibit 3).

Regional Data Set

The regional data set (see Microsoft Excel file "Regional Data Set," Ivey product no. W27309) contained sociodemographic information at the postal code (or zip code) level. Each row represented a specific global postal or zip code. For each zip code, information was provided about the level of education, the average income, and many other variables (see Exhibit 4).

After having collected all the data, Thoppan Mohanchandralal and his team began their work on analyzing the available data and all relevant information.[4]

[4] For the purpose of this case, all data sets were encrypted and disguised, with only a random sample of the data included. More specifically, a one-to-one mapping was done between the original names and the names in the spreadsheets. Company names were recoded manually into four groups and zip codes were truncated. Finally, random noise was added to all dates provided in the spreadsheets.

ANALYSIS

Thoppan Mohanchandralal decided that his team had to give some structure to the analytics process, so he decided to use the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework.[5] As the framework recommended, the data science process was split up into six different phases (see Exhibit 5). The CRISP-DM framework was a powerful tool that allowed and emphasized the need to move back and forth between phases, rather than going through the different phases sequentially.

According to the framework, a successful data science project started with developing a good understanding of the objectives and requirements of the business—before even starting to work with the data. The task of working with the data consisted of the second phase in the framework. In this second phase, it was important to develop a deeper understanding of the available data. To do so, data scientists typically used descriptive analytics techniques to gain early insights into the data and discover any potential problems. Therefore, Thoppan Mohanchandralal ordered his team to start creating univariate and bivariate plots, which would offer insight about the customers and their outstanding policies.

The third phase in the framework involved preparing, cleaning, and transforming the data to get it ready for the modelling phase. The team quickly noticed that all three data sets contained many instances of missing and disorderly values, so they spent a significant amount of time cleaning up the data. As an experienced data scientist, Thoppan Mohanchandralal was well aware that these early steps were crucial to the success of the project, even if less exciting than later steps. Therefore, the team members spent up to 80 per cent of their time cleaning up and exploring data, which was a common and highly useful exercise.

The next step—data modelling—was then ready to begin. Thoppan Mohanchandralal decided that it would be best to use both unsupervised and supervised analytics techniques. By using a clustering algorithm, which was an unsupervised machine learning technique, the team could detect how customers within a specific segment were similar and how they were different from other customer groups. Clustering algorithms were often used to personalize marketing campaigns and to develop actions to better serve a customer's needs. To actually predict which potential T&B customers were most likely to be converted to the new business line, the team used supervised learning and applied different classification algorithms.

After the data modelling phase was complete, the models still needed to be evaluated. To do so, the team could use various different metrics including accuracy, recall, and precision rate. Ideally, the metric that would best fit the nature of the business problem would be identified. The final step, according to the CRISP-DM framework, would be to deploy the model. However, the goal of the analysis was to gain insights, rather than to develop a new implementation, so Thoppan Mohanchandralal devoted few resources to the deployment phase.

From the insights generated by the analysis, Thoppan Mohanchandralal and his team identified the types of customers that were most likely to buy insurance online via T&B. In addition, by examining which features were most important to predict customer conversion, the team could derive business actions that were relevant to improve the conversion rate.

Finally, the results and insights could help create an effective and tailored marketing campaign and customer acquisition strategy. Thoppan Mohanchandralal started preparing with his recommendations to Allianz's senior management on how to improve online sales through the T&B online portal.

[5] Rüdiger Wirth and Jochen Hipp, *CRISP-DM: Towards a Standard Process Model for Data Mining*, Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 2000, 5..

EXHIBIT 1: CONVERGENCE RATES FOR THREE MAIN AFFINITIES

	Convergence	
	Yes	*No*
Insuro Inc.	0.25	0.75
Seguros International Ltd.	0.25	0.75
T&B plc	0.13	0.87

Source: Prepared by the case authors.

EXHIBIT 2: DATA DICTIONARY, FUNNEL DATA SET

Variable	Description	Notes
affinity_name	Name of the affinity	
status_report	Outcome of the quoting process	"Policycreated" refers to a policy that followed from the process
offer_number	Identifier for the process	Should be a unique identifier
policy_number	Number of the policy, if a policy followed from the request	Key to link to the policies data set
zipcode_link	Code for small scale regional unit	Key to link to the regional data set
zip4	Postal or zip code (four digits)	
birth_date	Birth date of the policy holder	
brand	Brand of the car	
date_offer	Date of the quoting process	
date_request	Data of request for the policy	Only if requested
policy_start_date	Start date of the policy	Only if requested
premium	Premium offered to customer	Depends on the policy chosen
buildyear_car	Year the car was built	
buildmonth_car	Month the car was built	
wa	Only liability insurance	Mutually exclusive with two options: "wa_bep_ca" and "wa_ca"
wa_bep_ca	Liability + limited casco	Mutually exclusive with "wa" and "wa_ca"
wa_ca	Liability + full casco	Mutually exclusive with two options: "wa" and "wa_bep_ca"
updated_on	Date when data was last updated	

Note: Liability insurance covers injury and property damage to third parties for which the policy holder is legally responsible; limited casco insurance covers common types of vehicle damage (e.g., glass breakage, hail damage); full casco insurance covers comprehensive vehicle damage and loss.
Source: Adapted by the case authors from company documents.

EXHIBIT 3: DATA DICTIONARY, POLICY DATA SET

Name	Description	Notes
policy_number	Policy number, if policy followed from request	Key to link to funnel data set
policy_continuation_date	When the policy should be extended	Should be in the future, taking 16/10/2019 as present day
policy_start_date	When the policy started	
policy_lastchange_date	When the policy was last changed	
premium_wa	Premium for the liability part of the insurance	
premium_other	Premium for additional coverage	
zipcode_link	Postal (or zip) code for small-scale regional unit	Key to link to regional data set
zip4	Postal or zip code (four digits)	
place_residence	Place of residence of the policy holder	
birth_date	Birth date of the policy holder	
gender	Gender of the policy holder	(does not affect pricing)
private_commercial	Use of the vehicle	P = private Z = commercial
bonus_malus_percent	Bonus-malus score derived from the number of claim-free years	Higher score should translate into lower premium
other_cover	Additional cover (in addition to liability)	BEP = limited casco CAS = full casco
premium_other_incl_discount	Premium additional covers including discounts	
worth_car	Value of the vehicle	
brand	Brand of the vehicle	
builddate_car	Month and year the vehicle was built	
weight_car	Weight of the vehicle	
mileage_car	Estimated mileage per year (in kilometres)	
power_car	Power of the vehicle	
fuel_car	Type of fuel used for the vehicle	
chassis	Type of chassis of the vehicle	
drive	Type of drive of the vehicle	
turbo	Whether the vehicle has turbo (yes/no)	
transmission	Gear transmission of the vehicle	

Source: Adapted by the case authors from company documents.

EXHIBIT 4: DATA DICTIONARY, REGIONAL DATA SET*

Name	Description	Notes
zipcode_link	Code for small-scale regional unit	Key to link to funnel and policy data sets
zip4	Postal or zip code (four digits)	
PROVINCE	Province in which postal or zip code is located	
URB	Degree of urbanization (0–7); 0 = unknown, ranges from 1 (= high degree of urbanization) to 7 (= low degree of urbanization).	
INCOME	Most common income(proportionally); ranges from 1 (= high) to 5 (= minimal income). 0 = unknown, 6 = diverse set of incomes within one regional unit.	
SOCCL_X	Percentage of households belonging to social class X (0–5); 0 = unknown, 1 = 0-5%, 2 = 5-25%, 3 = 25-50%, 4 = 50-75% and 5 = 75% or higher.	
EDU_HIGH	Percentage of households with high educational degree (0-5) (0-5); 0 = unknown, 1 = 0-5%, 2 = 5-25%, 3 = 25-50%, 4 = 50-75% and 5 = 75% or higher.	
EDU_MID	Percentage of households with medium educational degree (0–5); 0 = unknown, 1 = 0-5%, 2 = 5-25%, 3 = 25-50%, 4 = 50-75% and 5 = 75% or higher.	Medium educational degree = finished high school
EDU_LOW	Percentage of households with low educational degree (0–5); 0 = unknown, 1 = 0-5%, 2 = 5-25%, 3 = 25-50%, 4 = 50-75% and 5 = 75% or higher.	
DINK	Percentage of households consisting of two earners (0–6); 0 = unknown, 1 = 0-5%, 2 = 5-20%, 3 = 20-35%, 4 = 35-50% and 5 = 50-65%, 6 = 65% or higher.	DINK = Dual Income No Kids
OWN_HOUSE	Ratio of rental to owner-occupied housing (0–5); 0 = unknown, ranges from 1 (= all rental houses) to 5 (= all privately owned houses).	
AVG_HOUSE	Mean value of houses (0–15); 0 = unknown, ranges from 1 (= 0-75000 euros) to 15 (1M euros or more).	
RENT_PRICE	Rental price; 0 = unknown, ranges from 1 (= less than 250 euros) to 11 (=800 euros or more).	
STAGE_OF_LIFE	Most common stage of life (0–9); 0 = unknown, 1 = young singles, 2 = adult singles, 3 = older singles, 4 = families with young children, 6 = families with older children, 7 = Young couples without children, 8 = adult couples without children, 9 = older couples without children.	There is no category 5
SINGLE	Percentage of singles (0–5); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25-50% and 5 = 50% or higher.	

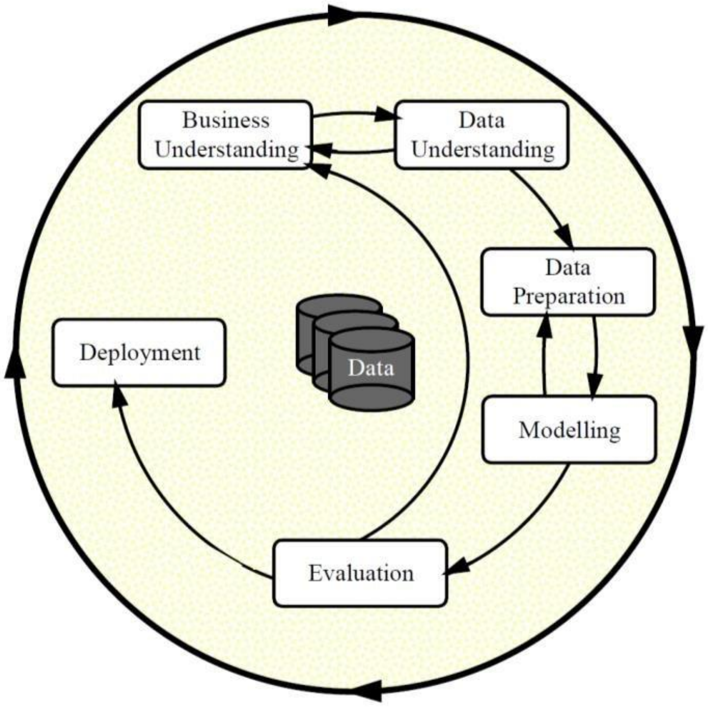
EXHIBIT 4 (CONTINUED)

FAM	Percentage of families (0–5); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25-50% and 5 = 50% or higher.	
FAM_WCHILD	Percentage of couples without children (0–5); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25-50% and 5 = 50% or higher.	
SINGLES_YOUNG	Percentage of young singles (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
SINGLES_MID	Percentage of medium-aged singles (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
SINGLES_OLD	Percentage of older singles (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
FAM_CHILD_Y	Percentage of families with only young children (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
FAM_CHILD_O	Percentage of families with older children (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
FAM_WCHILD_Y	Percentage of young couples without children (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
FAM_WCHILD_MED	Percentage of medium-aged couples without children (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
FAM_WCHILD_OLD	Percentage of older couples without children (0–4); 0 = unknown, 1 = 0%, 2 = less than 10%, 3 = 10-25%, 4 = 25 or higher.	
CIT_HOUSEHOLD	Number of citizens ÷ Number of households (0-6); 0 = unknown, 1 = 1 person, 2 = 1-1.5 persons, 3 = 1.5-2 persons, 4 = 2-2.5 persons, 5 = 2.5-3 persons, 6 = 3 persons or more.	
LOAN	Percentage of households with a loan (0–6); 0 = unknown, 1 = 0-5%, 2 = 5-15%, 3 = 15-30%, 4 = 30-60% and 5 = 60-90%, 6 = 90% or higher.	
SAVINGS	Percentage of households with a savings account (0–6); 0 = unknown, 1 = 0-15%, 2 = 15-50%, 3 = 50-80%, 4 = 80-90% and 5 = 90-95%, 6 = 95% or higher.	
SHOP_ONLINE	Percentage of households that shop online at least five times per year (0–6); 0 = unknown, 1 = 0-5%, 2 = 5-15%, 3 = 15-30%, 4 = 30-45% and 5 = 45-60%, 6 = 60% or higher.	
CAR	Percentage of households with at least one private car (0–5); 0 = unknown, 1 = 0-25%, 2 = 25-50%, 3 = 50-75%, 4 = 75%-90% and 5 = 90% or higher.	

Note: * 0 values have been changed to missing.
Source: Adapted by the case authors from company documents.

EXHIBIT 5: CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING MODEL

The six phases of the current Cross-Industry Standard Process Model for Data Mining



Source: Rüdiger Wirth and Jochen Hipp, *CRISP-DM: Towards a Standard Process Model for Data Mining*, Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 2000, 5. Reprinted with permission.