A2Z Insurance

Students per group: 3 maximum

Submission date: 06-January-2023 | 23h59

Page limit: 10 pages of content, excluding the cover page, index and appendices.

Note: The submission of all Python code developed is mandatory.

General Context

Finding new customers is vital in every industry. The process for finding new customers begins by learning as much as possible from the existing customers. Understanding current customers allow organizations to identify groups of customers that have different product interests, different market participation, or different response to marketing efforts. With this, we will be able not only to serve better our customers, but also to improve the targeting of prospective customers.

Market segmentation, the process of identifying customers' groups, makes use of geographic, demographic, psychographic, and behavioral characteristics of customers. By understanding the differences between the different segments, organizations can make better strategic choices about opportunities, product definition, positioning, promotions, pricing, and target marketing.

Business Situation

A2Z Insurance (A2Z) is a portuguese long standing insurance company that serves a wide array of insurance services: Motor, Household, Health, Life and Work Compensation. Although A2Z primarily serves portuguese customers, a significant portion of their customer acquisition comes from their web site. Customers can sign up to A2Z services through their branches, by telephone, or on the web site.

In 2016, A2Z became one of the largest insurers in Portugal. However, the lack of a data driven culture in the company ultimately led to poorly maintained databases over the years. A2Z is trying to make better use of the database it has regarding its customers. So far, it has simply mass-marketed everything. All potential and existing customers get the same promotions, and there are no attempts to identify target markets for cross-selling opportunities. Now, A2Z wants start differentiating customers, and developing more focused programs.

A2Z provided you an ABT (Analytic Based Table) with data regarding a sample of 10.290 Customers from its active database. These are customers that had at least one insurance service with the company at the time the dataset was extracted. Your job is to segment the database and find the relevant clusters of customers. To do this, we suggest you segment the customers using different perspectives and approaches, as well as combine and analyze the results. A2Z would like to understand the value and demographics of each customer segment, as well as understand which types of insurance they will be more interested in buying.

Metadata

Variable	Description	Additional Information
ID	ID	
First Policy	Year of the customer's first policy	(1)
Birthday	Customer's Birthday Year	(2)
Education	Academic Degree	
Salary	Gross monthly salary (€)	
Area	Living area	(3)

Variable	Description	Additional Information
Children	Binary variable (Y=1)	
CMV	Customer Monetary Value	(4)
Claims	Claims Rate	(5)
Motor	Premiums (€) in LOB: Motor	(6)
Household	Premiums (€) in LOB: Household	(6)
Health	Premiums (€) in LOB: Health	(6)
Life	Premiums (€) in LOB: Life	(6)
Work Compensation	Premiums (€) in LOB: Work Compensations	(6)

Additional Information

1. May be considered as the first year as a customer
2. The current year of the database is 2016
3. No further information provided about the meaning of the area codes
4. Lifetime value = (annual profit from the customer) X (number of years that they are a customer) - (acquisition cost)
5. Amount paid by the insurance company (€)/ Premiums (€) Note: in the last 2 years
6. Annual Premiums (2016). Negative premiums may manifest reversals occurred in the current year, paid in previous one(s).

Notice the data was stored in a SAS database file. There may be some intricacies when reading this file type using Python. Make sure the data is being imported properly and that you remove (before or after converting it to a pandas data frame) any irrelevant rows/columns you may find.

Expected outcomes

1. Explore the data and identify the variables that should be used to segment customers.
2. Identify customer segments
3. Justify the number of clusters you chose (taking in consideration the business use as well).
4. Explain the clusters found.
5. Suggest business applications for the findings and define general marketing approaches for each cluster.

Note: Invest time into evaluating your preprocessing pipeling, thinking how you want to do your clustering, possible approaches, and advantages or disadvantages of different decisions.