

Dear students,

I hope to find you well.

Please find in Moodle the data for the Data Mining Project. Data is regarding a fictional insurance company in Portugal. **Please note that, as discussed in class, the groups may have up to 3 members.** In the ABT (Analytic Based Table) we have data regarding 10.290 Customers. The report be must delivered in PDF format and follow the Nova IMS template and formatting. You may optionally modify the cover page and font colors, although it will not count towards your final grade. The report must contain a maximum of 10 pages of content, excluding the cover page, index and appendices. The code must be submitted as a single file (Python script or Jupyter Notebook).

For each the following variables are available:

Variable	Description	Additional Information
ID	ID	
First Policy	Year of the customer's first policy	May be considered as the first year as a customer
Birthday	Customer's Birthday Year	The current year of the database is 2016
Education	Academic Degree	
Salary	Gross monthly salary (€)	
Area	Living area	No further information provided about the meaning of the area codes
Children	Binary variable (Y=1)	
CMV	Customer Monetary Value	Lifetime value = (annual profit from the customer) X (number of years that they are a customer) - (acquisition cost)
Claims	Claims Rate	Amount paid by the insurance company (€)/ Premiums (€) Note: in the last 2 years
Motor	Premiums (€) in LOB: Motor	Annual Premiums (2016) Negative premiums may manifest reversals occurred in the current year, paid in previous one(s)
Household	Premiums (€) in LOB: Household	
Health	Premiums (€) in LOB: Health	
Life	Premiums (€) in LOB: Life	
Work Compensation	Premiums (€) in LOB: Work Compensations	

As a Data Mining/Analytic Consultant, you are asked develop a Customer Segmentation in such a way that it will be possible for the Marketing Department to better understand all the different Customers' Profiles.

You are expected to define, describe and explain the clusters you chose. Invest time in reasoning how you want to do your clustering, possible approaches, and advantages or disadvantages of different decisions. Simultaneous, you should express the marketing approach you recommend for each cluster.

Good luck!