MACHINE LEARNING AND DATA SCIENCE PROJECT

BUSINESS UNDERSTANDING

LAGOSTE STORES

The LaGoste Stores is a well-established company operating in the fashion, sports and luxury sector. Presently they have around 300,000 registered customers and serve more than 1.000.000 consumers a year. They sell products from 5 major categories: Sneakers, Rackets, T-Shirts, Watches and Hats. These five categories can be also be divided in Premium Brand material and more mainstream articles. The Customers can order and acquire those products through 3 channel groups: Physical stores, quaterly catalogs and the companies' website. Globally, the company had solid revenues and a healthy bottom line in the past 5 years, but the profit growth perspectives for the next 2 years are fickle. A few strategic initiatives are being considered to invert the situation. One of those is a Marketing efficiency program to improve marketing activities with a special focus on boosting tremendously the efficiency of the marketing campaigns.

THE PROJECTS

PROJECT "PREDICTIVE MODEL"

Objective

The objective of the team is to build a predictive model that will produce the highest profit for the next direct marketing campaign of the company – the sixth campaign this year that is scheduled for next month. The campaign aims at selling a new product to the customer database (potential of 250,000 customers). To be able to build the predictive model, a pilot campaign was carried out. A sample of 2.500 customers were contacted by mail regarding the acquisition of the product. During the next 2 months, the customers who bought the offer were tagged with a 1 whereas the non-respondents were tagged with a 0. The total cost of the sample campaign was 2.500 contacts * 4€, for a total of 10.000€. Around 12.5% of Customers accepted the offer (which is quite good), each contributing with 20€ of revenue. Overall, the campaign had a negative profit of around -3.750€. (This can vary slightly by group dataset). The idea is to develop a model that predicts customer behavior and apply this model to the rest of the customer base. Hopefully the model will allow the company to cherry pick the customers that are most likely to purchase the offer, while leaving out the non-responders, making the next campaign highly profitable.

As said before, the cost per contact is 4€ and the revenue per accepted offer is 20€.

PROJECT "SEGMENTATION"

Objective

The objective of the team in this project is to identify actionable segments within the company's Customer base. These segments must be identified looking at data available and through the usage of quantitative technics. A priori, two visions are considered important – the customer value segmentation and the product usage segmentation. Nonetheless other perspectives will be valued. The output of this project will be a report identifying the main customer segments and the profile definition of each segment.

THE DATASET

The dataset used for the models development will be "Group_xx_LaGoste.xlsx". This dataset includes a priori information about the 2.500 customers contacted but also a posteriori information about who responded positively (1) or not (0) to the offer (DepVar variable). The dependent variable will be used only for the predictive project. Below you can find a list of variables present in the Dataset:

Variable	Description
AcceptedCmp1	Flag indicating customer accepted offer in campaign 1
AcceptedCmp2	Flag indicating customer accepted offer in campaign 2
AcceptedCmp3	Flag indicating customer accepted offer in campaign 3
AcceptedCmp4	Flag indicating customer accepted offer in campaign 4
AcceptedCmp5	Flag indicating customer accepted offer in campaign 5
Complain	Flag indicating if customer has complained
Custid	Customer ID
DepVar	Binary variable indicating if customer accepted (1) or not (0) a marketing offer from current campaign. Dependent variable of the problem. (Only applicable on the predictive problem)
Dt_Customer	Date of customer's enrolment with the company
Education	Level of education of Customer
Income	Yearly Income of household of Customer
Kidhome	Number of kids in household
Marital_Status	Marital Status of Customer
MntSneakers	Amount spent on Sneakers
MntRackets	Amount spent on Rackets
MntTShirts	Amount spent on Tshirts
MntWatches	Amount spent on Watches
MntHats	Amount spent on Hats
MntPremium_Brand	Amount spent on Premium material
NumCatalogPurchases	Number of purchases made through catalog
NumStorePurchases	Number of purchases made through store
NumDealsPurchases	Number of purchases made with discounts
NumWebPurchases	Number of purchases made through web
NumWebVisitsMonth	Average number of web visits a month to the company site
Recency	Number of days since last purchase
Teenhome	Number of teenagers in household
Year_Birth	Customer's Year of birth

EVALUATION CRITERIA

SEGMENTATION PROBLEM

Segment the Customer database for marketing purposes. Explore and Identify meaningful and actionable segments using Python segmentation algorithms.

PREDICTIVE PROBLEM

Build a predictive model for a dependent variable representing Customers that answer positively to a direct marketing campaign.

DELIVERABLES

A jupyter notebook for each problem, that summarizes the analytical processes and the main conclusions obtained (You should use markdown cells to write the conclusions that you think that should be highlighted during the process). Those files should be sent, by email, to calbuquerque@novaims.unl.pt - deadlines defined in the syllabus.

GUIDELINES FOR PROJECT EVALUATION

a) The quality and synthesis of the observations made in the jupyter notebooks.
b) For both problems, the quality of the pre-processing steps (outliers treatment, fill missing values, quality of the proposed transformations) and the visualizations used.
c) In the segmentation problem, the quality and completeness of the profiles for each segment
d) In the predictive model, the performance of the models proposed (F1 Score and AUC), and the profit associated to the final model chosen by the group as the best model.