Data Preprocessing

Project Description

2023/2024

Joana Neves
jneves@novaims.unl.pt

Project Description

Nowadays, the retail market sector is competitive, with customer service and satisfaction being the main priority in this industry. As such, it becomes crucial for companies to find some differentiating factors and understand the customers' behavior. Thus, Mega Market (a company that sells a wide variety of products) intends to take advantage of the data stored by its information systems and develop different analyses that will help the growth and success of the company.

The data that the company has represents the customers' transactions. These transactions reflect the business's situation and allow the outline of the profile of the company's customers. Thus, Mega Market managers hire a team of data scientists to monitor the business and segment the customers. Within this team of data scientists, there is a subgroup dedicated exclusively to data preprocessing (DP Team).

The DP Team should prepare the data to advanced analysis methods and already give some insights from the business since this company lacks information on their activity and mainly about their customers' shopping behavior.

Hence, Mega Market needs an exploratory analysis that will allow to answer some simple business questions and an analytic-based table (ABT) to serve the purpose of a descriptive analysis - segmentation of customers. Basically, the DP Team, aims to use data from the company's IS and create an ABT to deliver to the next team that will finish the job.

Description of the transactional table variables:

Variable	Description
transactionNo	Transaction ID
date	Transaction's date
Product Id	Product ID
Product category Id	Product category ID
Product category name	Name of the product category
Product Name	Name of the product
Unit Price	Product unit price
Quantity	Number of items bought
total_payed	Amount spent by the customer
customerNo	Customer ID
Nationality	Customer nationality
Gender	Customer gender (M/F/O)
Monthly Income	Customer monthly income
Age	Customer age
Kids	Customer have kids (1-yes;0-no)
Reviews	Customer let a review about the product (1-yes;0-no)
Payment	Customer payment type
Channel	Sales channel name

Requirements:

Thus, as consultants it is required:

1. Build an ABT
2. Withdraw some insights using visualization tools

Note the following:

- The transactional table will be given by the professor
- The software that can be used are Excel, SAS (Enterprise Guide and/or Miner), PowerBI or any other you may want to use.

Suggestion (guidelines):

1. Perform some initial descriptive statistics (SAS Miner)
2. Treat outliers (SAS Miner)
3. Treat missing values (SAS Miner)
4. Check the coherence (SAS Miner/SAS Guide)
5. Transform and create derived variables (SAS Guide/Miner)

6. Create the final ABT (SAS Guide)
7. Create some visualizations (PowerBI) (you can use the transactional table provided by SAS Miner)

Deliverables:

- Enhanced customer-signature table;
- PDF Report (reporting all steps of your project);
- Document with visualizations (in PowerBI).

All documents must be **submitted to moodle** until **December 15 (23h59m)**.