

Data Preprocessing

Project Description

2021/2022



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. Thus, the company The Market (a company that sells a wide variety of products) intends to take advantage of the data produced by its information systems and subsequently carry out various useful analyzes that will help the growth and success of the company.



The data that the company has on the transactions that occurred reflect the situation of the business, as well as allow to outline the profile of the company's customers. Thus, the company managers hire a team of data scientists to transform data into value, and more specifically to monitor the business, as well as to 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 has a big lack of information on their activity and mainly about their customers shopping behavior.

Hence, The 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.

About the data

The data represent day-to-day transactions between the years 2018 and 2020.

Description of the transactional table variables:

Variable	Description
<i>transaction_id</i>	Transaction ID
<i>cust_id</i>	Customer ID
<i>tran_date</i>	Transaction's date
<i>DOB</i>	Customer date of birth
<i>Kidhome</i>	Customer has a kid: 1=yes, 0=no
<i>Nationality</i>	Customer nationality
<i>Gender</i>	Customer gender (M/F)
<i>Address</i>	Customer city
<i>Channel</i>	Sales channel name
<i>Type_payment</i>	Type of payment (mbway, credit card, paypal, cash)
<i>Product_line</i>	Name of the product line
<i>Unit_price</i>	Price of the product
<i>Quantity</i>	Quantity bought on the transaction
<i>Tax</i>	Tax amount of the transaction
<i>Total_amt</i>	Amount spent on the transaction
<i>Cogs</i>	Costs of goods sold
<i>Rating</i>	Customer rating of the product (0-10)

Requirements:

Thus, as consultants it is required:

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

Note the following:

- The transactional 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.
- You do not need to create time series

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 17st (11:59pm)**.

