# Data Understanding

## Initial Data Collection Report

The first step of our approach to this project was to load the dataset to a jupyter notebook where we would be able to analyze all the data information.

The data that we will use during was supplied by the company's administration council, it is a xlxs file with 10000 entries with 30 column each entry.

## Data Description Report

Customer features:

* **CUSTID: numeric – customer identification number**
* **DAYSWUS: numeric - number of days as a customer**
* **AGE**: numeric - age of the customer
* **EDUC: numeric - years of education**
* **INCOME: numeric - household income**
* **KIDHOME: binary – 1 = child under 13 lives at home**
* **TEENHOME: binary – 1 = child 13-19 years lives at home**
* **FREQ: numeric - number of purchases in past 18 months**
* **RECENCY: numeric - number of days since last purchase**
* **MONETARY: numeric - total sales to this person in 18 months**
* **LTV: numeric - Lifetime value of the customer**

**Wine type:**

* **PERDEAL: percentage - % purchases bought on discount**
* **DRYRED: percentage - % of wines that were dry red wines**
* **SWEETRED: percentage - % sweet or semi-dry reds**
* **DRYWH: percentage - % dry white wines**
* **SWEETWH: percentage - % sweet or semi-dry white wines**
* **DESSERT: percentage - % dessert wines (port, sherry, etc.)**
* **EXOTIC: percentage - % very unusual wines**

**Accessories**

* **~~SMRACK: binary – 1 = bought the small wine rack $50~~**
* **~~LGRACK: binary – 1 = bought the large wine rack $100~~**
* **~~HUMID: binary – 1 = bought wine cellar humidifier $75~~**
* **~~SPCORK: binary – 1 = silver-plated cork extractor $60~~**
* **~~BUCKET: binary – 1 = silver wine bucket $150~~**
* **ACCESS: numeric - number of accessories (not SPCORK)**

**General Sales Information:**

* **WEBPURCH: percentage - % of purchases made on website**
* **WEBVISIT: numeric - average # visits to website per month**
* **~~COMPLAIN: binary – 1 = made a complaint in last 18 mo.~~**
* **~~MAILFRND: binary – 1 = appears on a purchased list of “mail friendly” customers~~**
* **~~EMAILFRD: binary – 1 = on a purchased list of “e-mail friendly” customers~~**

In this dataset we have 3 types of variables numeric, percentage and binary.

After the ana

## Data Exploration Report

## Data Quality Report