

Customer segmentation – demo

The present report is obtained from an open dataset regarding clients of a supermarket.

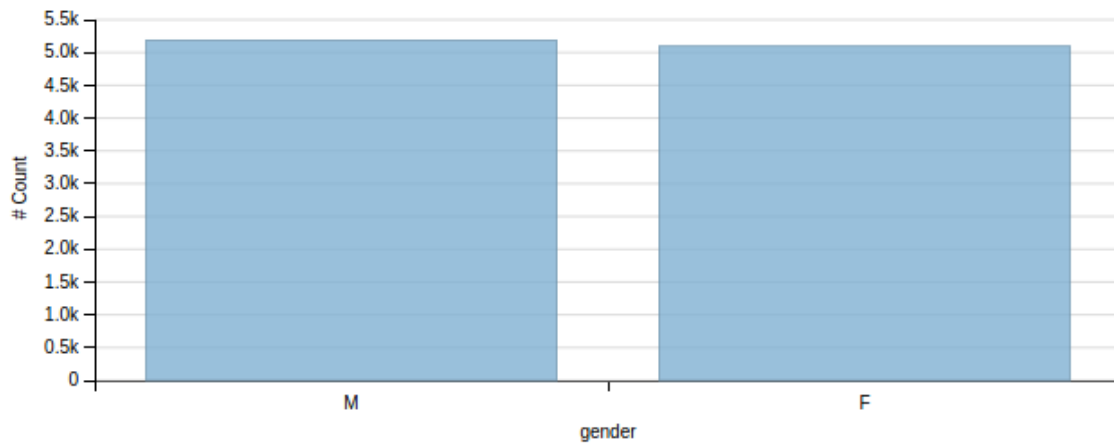
With Segmentation is possible to derive a group of categories starting from a set of observations. The method stems from machine learning and allows a scientific, data-driven approach to marketing and decision-making.

In the first part we will see the distribution of the most important variables in the dataset. Segmentation's output comes in the second part.

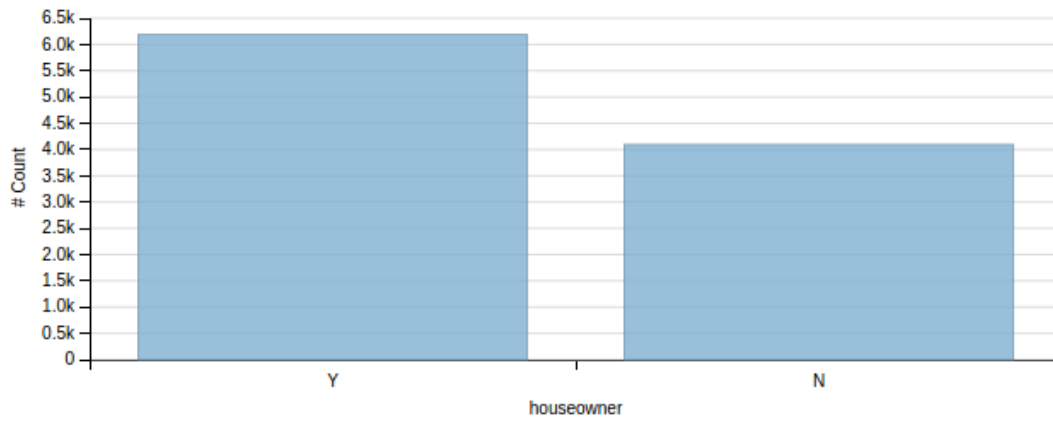
1 – Variable distributions

Variables taken into considerations to build the segments are:

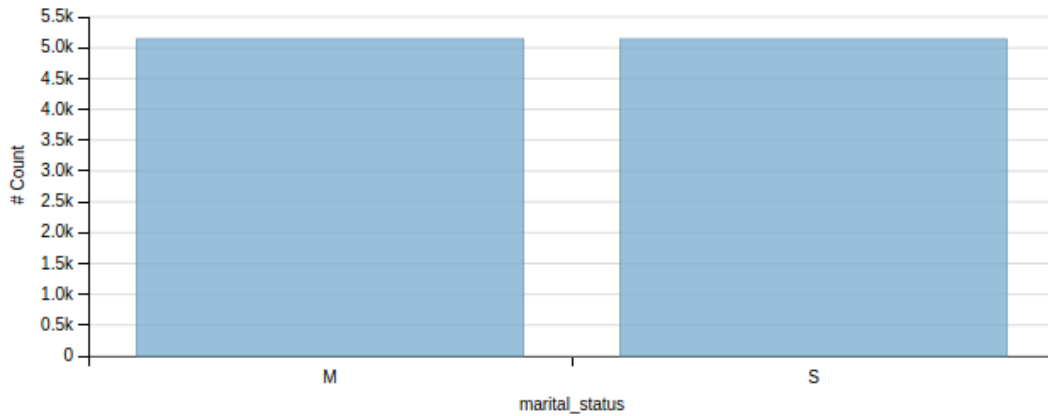
- *gender*: M vs. F in the original data, this variable is here vectorized in order to occupy only one dimension in the model, with range [0,1]. 0 = female, 1 = male. There is no prevalent sex in the distribution.



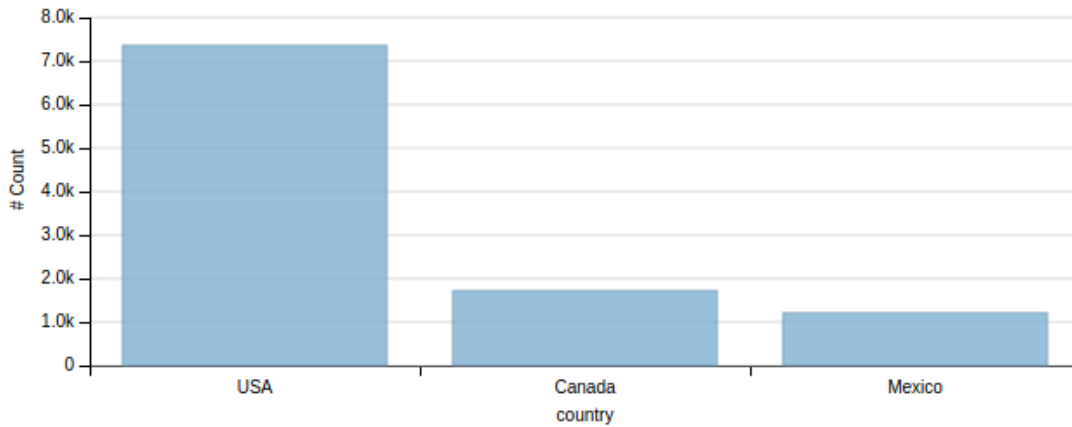
- *houseowner*: Y vs N in the original data, here the variable is vectorized in order to occupy only one dimension in the model, with range [0,1]. 0 = does not own an house, 1 = owns an house. 60% of clients owns an house.



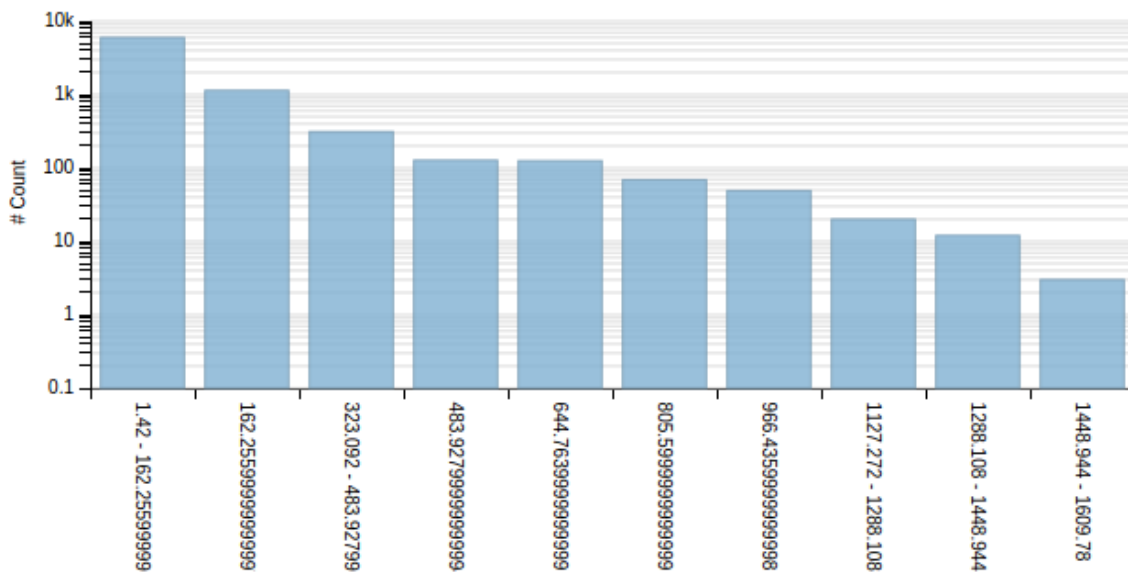
- *marital_status*: In the original dataset M vs S, the variable is vectorized to occupy only one dimension in the model, with range [0,1]. 0 = single, 1 = married. There is no prevalence of one value respect the other.



- country: Values are USA, Mexico e Canada. 70% of clients are Americans, the rest is made of Canadians and Mexicans.

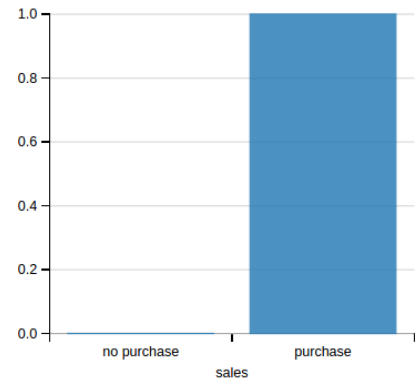
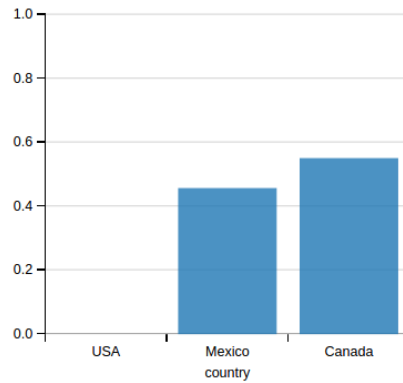
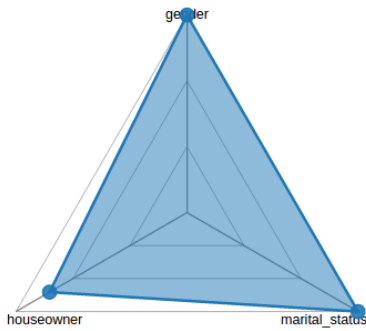


- sales: In the original dataset this is the value in dollars of the total money spent by a customer in the supermarket. This variable is discretized in two values (purchase vs. no purchase) for the Segmentation, to distinguish occasional customers from regular ones. The majority of customers spent less than 300\$ (see logarithmic scale on the y axis).

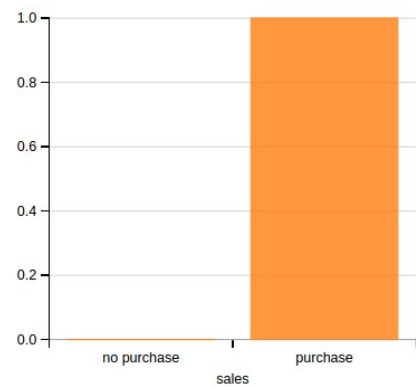
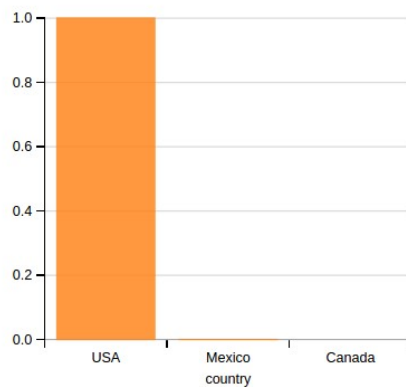
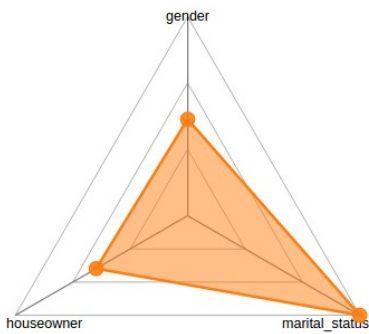


2 – Customer segments

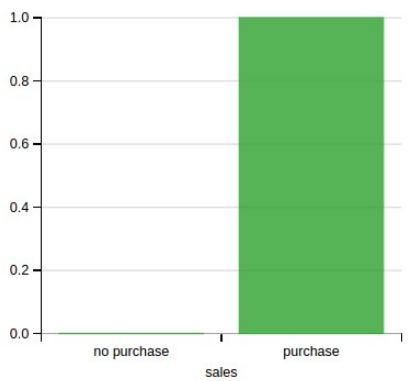
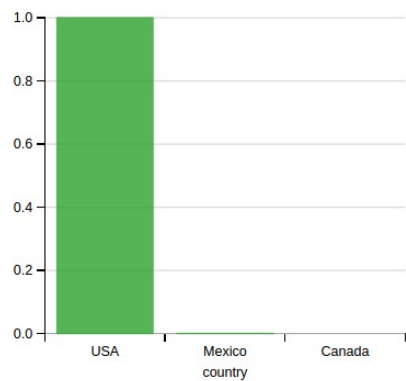
1 – **Non-american accomplished husband.** Clients in this segment live outside the USA, are male, married and likely owns an house. While this segment is profitable, the other most similar one (segment 4) is made of non-americans and is not profitable.



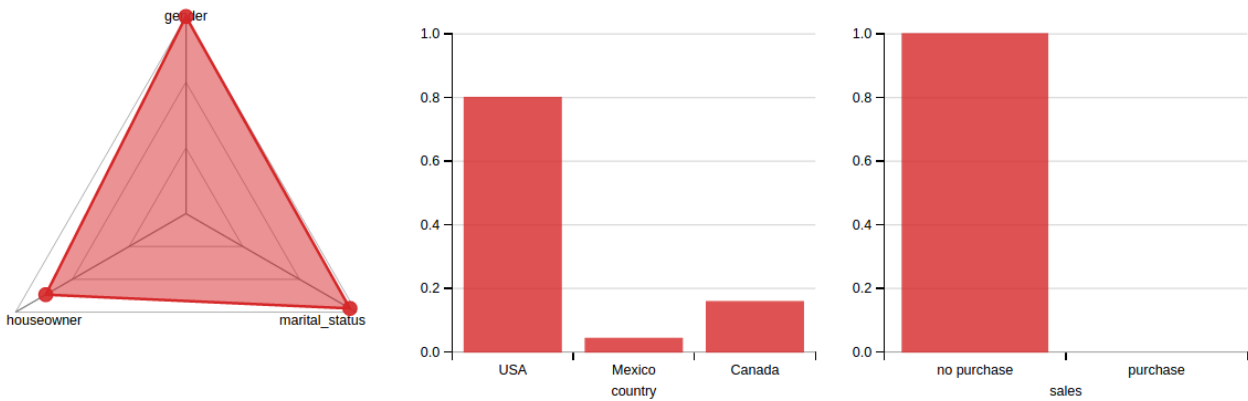
2 – **Married american couples.** Segment 2 is made of married americans, both genders. Ownership of an house is not a differentiating factor.



3 – **American single male.** Segment 3 is made of american males, not married.



4 – **Accomplished american husband.** Similar to segment 1 for gender, house ownership and marital status. The two differences are 1) these are american customers, and 2) they don't buy.



Consultancy

- Differentiate marketing and services among 3 channels, corresponding to segments 1, 2, 3:
 1. Out of the USA the target is male, married and owns an house. The same profile is not profitable in the USA (see segment 4).
 2. Married american couples, without assuming they already bought an house.
 3. Single american male, without assuming he already bought an house.