# Wine displayed on a log with grapes at the Solis winery close to San Jose California

# Fancy some highbrow wines?

# Introduction

Every year, a retailer launches campaigns for lots of products, including up-selling and cross-selling actions. They send invitation letters to a pre-defined number of clients in which those clients will find a coupon. In this invitation letter, they make them an offer around one specific product. In this case study, you will try to identify the customers that are most likely to buy highbrow wines on the retailers’ web shop.

# Assignment

We will send you two datasets:

* A set, which describes customers and their purchases during 2016. For each customer it is indicated whether they bought highbrow wines on the web shop (=data\_2016\_new).
* The second set contains the exact same information for customers in 2017 but off course, without the information who bought wines online and who did not (=data\_2017\_new).

The marketing department wants to know which customers have the highest chance to buy wines online. We expect you to build a predictive statistical model (propensity to buy) that predicts which customers are most likely to buy wine online.

The choice is up to you which model to build and what kind of technology to use.

# Expected output

1) **Results**

Provide us with the predictions for 2017. This should be a csv file delimited by a "**;**" containing the following columns (in this specific order):

* The customer id
* The predicted class (1 for buyer, 0 for non-buyer)
* The probability of this customer being a buyer

An example of the submission format is provided as "example.csv"

2) **Presentation**

We ask you to prepare a presentation of about 45 min. (max), start by presenting yourself (10min.) and continue with the results of your assignment (focus on what steps you took, less on the result) 30min. (max). During the presentation, we expect you to go through every step of model building that you took and to explain to us why you did it.

Finish with an insight of analytics opportunities in the retail sector – 10min. (max).

There is no right or wrong answer to the case study as such, we are more interested in how you addressed the problem. The tool(s) you use is entirely your choice.

# Field descriptions

|  |  |
| --- | --- |
| Field | Description |
| masked\_customer\_id | Unique customer id |
| jaar | Year of purchase |
| bought\_highbrow\_wines | Has the customer bought wines online |
| Collishop\_customer | Has the customer bought on online on Collishop |
| cat\_AP\_STDR\_PortoONLINE | Customers turnover in category Apperitief sterke drank Porto Online (Porto Online) |
| cat\_AP\_STDR\_WhiskyONLINE | Customers turnover in category Apperitief sterke drank Whiskey (Whiskey Online) |
| cat\_Babyluiers | Customers turnover in category Babyluiers (Diapers) |
| cat\_Ber\_Ger\_DVPortioneerbaar | Customers turnover in category Bereide Gerechten Diepvries Portioneerbaar (Frozen Prepared Meals) |
| cat\_Ber\_Ger\_VersMaaltijdsalades | Customers turnover in category Bereide Gerechten Salades Vers (Fresh Salades Prepared Meals) |
| cat\_Bier\_Genietbieren | Customers turnover in category Genietbieren (Beers) |
| cat\_Bloemen | Customers turnover in category Bloemen (Flowers) |
| cat\_Bot\_Mar\_Boter | Customers turnover in category Boter (Butter) |
| cat\_BroodKorthoudbaar | Customers turnover in category Korthoudbare Broden (Bread with Short Shelf Life) |
| cat\_Chips | Customers turnover in category Chips |
| cat\_ColruytMobile\_Toestellen | Customers turnover in category Mobile Appliances |
| cat\_DeegDV | Customers turnover in category Diepvries Deeg (Frozen Dough) |
| cat\_EleKtroKeuken | Customers turnover in category Electrisch Keuken (Electronical appliances for Kitchen) |
| cat\_Houtpelletskolen\_briketten | Customers turnover in category Houtskool/Pelletes/Brikketen (Charcoal/Pellets/Briquettes) |
| cat\_Incontinentie\_luiers | Customers turnover in category Incontinentie Luiers (Incontinence Diapers) |
| cat\_KaasSeizoenskazen | Customers turnover in category Seizoenskazen (Seasonal Cheeses) |
| cat\_Kauwgum | Customers turnover in category Kauwgum (Gum) |
| cat\_KoudeSauzen | Customers turnover in category Koude Sauzen (Cold sauces) |
| cat\_MelkKarnemelk | Customers turnover in category Karnemelk (Buttermilk) |
| cat\_Notengedroogdfruit\_groenten | Customers turnover in category Droge Noten/Fruit/Groenten (Dried Nuts/Fruits/Vegetables) |
| cat\_Ontbijtgranen\_Volwassenen | Customers turnover in category Ontbijtgranen Volwassen (Cereals Adults) |
| cat\_ParfumerieEHBO | Customers turnover in category Parfum (Perfumes) |
| cat\_Tapas | Customers turnover in category Tapas |
| cat\_Textiel\_Bedlinnen | Customers turnover in category Bedlinnen (Bedding) |
| cat\_Textiel\_Herenondergoed | Customers turnover in category Heren Ondergoed (Male underwear) |
| cat\_Textiel\_Pantys | Customers turnover in category Pantys |
| cat\_VNCBGBereidegerechten | Customers turnover in category Bereide Gerechten (Prepared Meals) |
| cat\_VNCBerBurgers | Customers turnover in category Bereide Burgers (Prepared Burgers) |
| cat\_VNCCharBHWildpasteien | Customers turnover in category Wild pasteien (Wild Pie) |
| cat\_VNCFSalades | Customers turnover in category Salades |
| cat\_VNCGevgeheel | Customers turnover in category Gevogelte (Birds) |
| cat\_VNCKalfStoofvlees | Customers turnover in category Kalfstoofvlees (Calf Beef Stew) |
| cat\_VNCLamSnedenkoteletsteak | Customers turnover in category Lam koteletten (Lamb Chops) |
| cat\_VNCRest | Customers turnover in category Rest |
| cat\_VNCVarkenRest | Customers turnover in category Varkens rest (Pork Rest) |
| cat\_VNCWildSteak | Customers turnover in category Wild steak |
| cat\_VerseKaasFruitkazen | Customers turnover in category Verse Fruit Kaas (Fresh Fruit Cheese) |
| cat\_VisGerookt | Customers turnover in category Gerookte Vis (Smoked Fish |
| cat\_VisVerseSchelpdieren | Customers turnover in category Verse Schelpdieren (Fresh Shell Animals) |
| cat\_Wijn\_Stillewijnen\_RAYON | Customers turnover in category Stille Wijnen in de winkel (Still Wines in store) |
| cat\_Zomerspeelgoed | Customers turnover in category Zomerspeelgoed (Toys for Summer) |
| cat\_bbqfoodevent | Customers turnover in category BBQ |
| cat\_nfokay | Customers turnover in category Near Food okay |
| total\_discount | Total discount given at customer |
| rev\_ticket | Revenue per ticket |
| prod\_ticket | Number of products per ticket |
| n\_cogo | Number of tickets at Collect & Go |
| cogo\_rev | Customers revenue at Collect & Go |
| HOUSEHOLDTYPOLOGY | Householdtypology |
| price\_sens\_colr | Price sensitivity score of the customer (higher score = more price sensitive = tends to buy cheaper articles) |
| SOW\_type\_colr | Share of Wallet category (percentage of household budget spend at Colruyt) |
| SOW\_colr | Share of Wallet (percentage of household budget spend at Colruyt) |

# Special values

* In column Householdtypology, ! means unknown
* In the SOW columns:
  + Outlier\_freq: The customer is an outlier based on his frequency
  + Outlier\_om: The customer is an outlier based on his turnover