# 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 which are most likely to buy highbrow wines on the retailers webshop.

# Assignment

We will send you 2 datasets:

* A set which describes customers and their purchases during 2016. For each customer it’s indicated whether or not they bought highbrow wines on the webshop.
* The second set contains the exact same information for customers in 2017 but, off course, without the information who bought wines online and who didn’t.

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

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.

For questions or remarks you are free to contact:

Dries De Dauw – dries.dedauw@colruytgroup.com - +32(0)494 84 12 26

# 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 cat\_AP\_STDR\_PortoONLINE |
| cat\_AP\_STDR\_WhiskyONLINE | Customers turnover in category cat\_AP\_STDR\_WhiskyONLINE |
| cat\_Babyluiers | Customers turnover in category cat\_Babyluiers |
| cat\_Ber\_Ger\_DVPortioneerbaar | Customers turnover in category cat\_Ber\_Ger\_DVPortioneerbaar |
| cat\_Ber\_Ger\_VersMaaltijdsalades | Customers turnover in category cat\_Ber\_Ger\_VersMaaltijdsalades |
| cat\_Bier\_Genietbieren | Customers turnover in category cat\_Bier\_Genietbieren |
| cat\_Bloemen | Customers turnover in category cat\_Bloemen |
| cat\_Bot\_Mar\_Boter | Customers turnover in category cat\_Bot\_Mar\_Boter |
| cat\_BroodKorthoudbaar | Customers turnover in category cat\_BroodKorthoudbaar |
| cat\_Chips | Customers turnover in category cat\_Chips |
| cat\_ColruytMobile\_Toestellen | Customers turnover in category cat\_ColruytMobile\_Toestellen |
| cat\_DeegDV | Customers turnover in category cat\_DeegDV |
| cat\_EleKtroKeuken | Customers turnover in category cat\_EleKtroKeuken |
| cat\_Houtpelletskolen\_briketten | Customers turnover in category cat\_Houtpelletskolen\_briketten |
| cat\_Incontinentie\_luiers | Customers turnover in category cat\_Incontinentie\_luiers |
| cat\_KaasSeizoenskazen | Customers turnover in category cat\_KaasSeizoenskazen |
| cat\_Kauwgum | Customers turnover in category cat\_Kauwgum |
| cat\_KoudeSauzen | Customers turnover in category cat\_KoudeSauzen |
| cat\_MelkKarnemelk | Customers turnover in category cat\_MelkKarnemelk |
| cat\_Notengedroogdfruit\_groenten | Customers turnover in category cat\_Notengedroogdfruit\_groenten |
| cat\_Ontbijtgranen\_Volwassenen | Customers turnover in category cat\_Ontbijtgranen\_Volwassenen |
| cat\_ParfumerieEHBO | Customers turnover in category cat\_ParfumerieEHBO |
| cat\_Tapas | Customers turnover in category cat\_Tapas |
| cat\_Textiel\_Bedlinnen | Customers turnover in category cat\_Textiel\_Bedlinnen |
| cat\_Textiel\_Herenondergoed | Customers turnover in category cat\_Textiel\_Herenondergoed |
| cat\_Textiel\_Pantys | Customers turnover in category cat\_Textiel\_Pantys |
| cat\_VNCBGBereidegerechten | Customers turnover in category cat\_VNCBGBereidegerechten |
| cat\_VNCBerBurgers | Customers turnover in category cat\_VNCBerBurgers |
| cat\_VNCCharBHWildpasteien | Customers turnover in category cat\_VNCCharBHWildpasteien |
| cat\_VNCFSalades | Customers turnover in category cat\_VNCFSalades |
| cat\_VNCGevgeheel | Customers turnover in category cat\_VNCGevgeheel |
| cat\_VNCKalfStoofvlees | Customers turnover in category cat\_VNCKalfStoofvlees |
| cat\_VNCLamSnedenkoteletsteak | Customers turnover in category cat\_VNCLamSnedenkoteletsteak |
| cat\_VNCRest | Customers turnover in category cat\_VNCRest |
| cat\_VNCVarkenRest | Customers turnover in category cat\_VNCVarkenRest |
| cat\_VNCWildSteak | Customers turnover in category cat\_VNCWildSteak |
| cat\_VerseKaasFruitkazen | Customers turnover in category cat\_VerseKaasFruitkazen |
| cat\_VisGerookt | Customers turnover in category cat\_VisGerookt |
| cat\_VisVerseSchelpdieren | Customers turnover in category cat\_VisVerseSchelpdieren |
| cat\_Wijn\_Stillewijnen\_RAYON | Customers turnover in category cat\_Wijn\_Stillewijnen\_RAYON |
| cat\_Zomerspeelgoed | Customers turnover in category cat\_Zomerspeelgoed |
| cat\_bbqfoodevent | Customers turnover in category cat\_bbqfoodevent |
| cat\_nfokay | Customers turnover in category cat\_nfokay |
| n\_tickets | Number of tickets the customer has |
| total\_discount | Total discount given at customer |
| total\_revenue | Total revenue of 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