
MARKETING ANALYTICS

# Database description

ARNAUD DE BRUYN

# Overview

All assignments will be based on the same database, named “ma_charity” (for *Marketing Analytics Charity*), which we will load as a MySQL database.

The database will be provided in two sizes:

- ma_charity_small (a 10% random extract, for assignment 1)
- ma_charity_full (the full-sized database, for assignments 2 and 3)

# Loading the data

Do NOT open it in MySQL Workbench directly, the files are too voluminous.

In MySQL Workbench, go to **Server** > **Data Import** > **Import from Self-Contained File**, and select the file you need, then import. Do it for both (it will create two separate databases).

# Table format

## Charity tables

```
CREATE TABLE `actions` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `contact_id` int(11) NOT NULL,
  `campaign_id` char(4) CHARACTER SET utf8 DEFAULT NULL,
  `message_id` char(8) CHARACTER SET utf8 DEFAULT NULL,
  `action_date` date DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=MyISAM;

CREATE TABLE `acts` (
  `id` int(10) unsigned NOT NULL AUTO_INCREMENT,
  `contact_id` int(10) unsigned NOT NULL,
  `amount` decimal(9,2) DEFAULT NULL,
  `act_date` date DEFAULT NULL,
  `act_type_id` char(2) CHARACTER SET utf8 DEFAULT NULL,
  `payment_method_id` char(2) CHARACTER SET utf8 DEFAULT NULL,
  `channel_id` char(2) CHARACTER SET utf8 DEFAULT NULL,
  `campaign_id` char(4) CHARACTER SET utf8 DEFAULT NULL,
  `message_id` char(8) CHARACTER SET utf8 DEFAULT NULL,
  PRIMARY KEY (`id`),
  KEY `idx_contact` (`contact_id`),
  KEY `idx_amount` (`amount`),
  KEY `idx_act_date` (`act_date`),
  KEY `idx_act_type` (`act_type_id`),
  KEY `idx_payment_method` (`payment_method_id`),
  KEY `idx_channel` (`channel_id`),
  KEY `idx_campaign` (`campaign_id`),
  KEY `idx_message` (`message_id`)
```

```
) ENGINE=MyISAM;

CREATE TABLE `channels` (
  `id` char(2) CHARACTER SET utf8 NOT NULL,
  `label` varchar(45) CHARACTER SET utf8 NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=MyISAM;

CREATE TABLE `contacts` (
  `id` int(10) unsigned NOT NULL AUTO_INCREMENT,
  `external_id` varchar(20) CHARACTER SET utf8 DEFAULT NULL,
  `prefix_id` varchar(5) CHARACTER SET utf8 DEFAULT NULL,
  `first_name_clean` varchar(32) CHARACTER SET utf8 DEFAULT NULL,
  `zip_code` varchar(5) CHARACTER SET utf8 DEFAULT NULL,
  `town_clean` varchar(60) CHARACTER SET utf8 DEFAULT NULL,
  `code_geo` varchar(5) CHARACTER SET utf8 DEFAULT NULL,
  `active` tinyint(3) unsigned DEFAULT '1',
  PRIMARY KEY (`id`)
) ENGINE=MyISAM;

CREATE TABLE `payment_methods` (
  `id` char(2) CHARACTER SET utf8 NOT NULL,
  `label` char(60) CHARACTER SET utf8 DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1;

CREATE TABLE `prefixes` (
  `id` char(5) CHARACTER SET utf8 NOT NULL,
  `label` char(45) CHARACTER SET utf8 NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=MyISAM;
```

## Assignment #2

The following table is only provided for the “ma_charity_full” database, and will be used exclusively for assignment #2.

```
CREATE TABLE `assignment2` (
  `contact_id` int(11) unsigned NOT NULL,
  `calibration` tinyint(3) unsigned DEFAULT NULL,
  `donation` tinyint(3) unsigned DEFAULT NULL,
  `amount` decimal(9,2) DEFAULT NULL,
  `act_date` date DEFAULT NULL,
  PRIMARY KEY (`contact_id`),
  KEY `idx_contact_calibration` (`contact_id`,`calibration`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1;
```

# Explanations

## Contacts

The contacts table contains the list of individuals who have made a donation to the charity in the past or have been in relationship with the charity. Some of these individuals may have no corresponding donations in the database.

Most of the fields are self-explanatory. Please note that:

- The field “id” in the “contacts” table corresponds to the field “contact_id” in the “acts” and “actions” tables.
- The meaning of the field “prefix_id” can be found in the reference table “prefixes”.
- the suffix “_clean” means that the text has been cleaned, stripped of accents, and correctly formatted.
- “code_geo” corresponds to the geographical code that you might find in INSEE databases. It might have missing values.
- “active” means that the address is still active in the database (=1). Contacts become inactive (=0) if they died, moved without known address, or have asked the charity to be ignored in future campaigns.

## Actions

The charity will send solicitations to its donors to generate further donations. Every time a donor is solicited, it triggers an entry in the “actions” table.

Each solicitation campaign has a code (e.g., campaign_id=C124), and within the same campaign, solicitations may be divided into multiple messages (e.g., message_id=C124-12).

The charity might create different message ids within the same campaign to track:

- The behavior of various segments
- Marketing tests (e.g., with or without a brochure)
- Various versions of the same messages
- Etc.

Suppose you launch a direct marketing campaign in November 2020. You give that campaign a code, say "C140". You then decide who you are going to send that message to. If you decide to send a letter to 100,000 people, then there will be 100,000 rows in the "actions" table, with the code campaign_id = C140, and the contact_id of the people whom you have solicited.

If some of those people make a donation on that campaign, then their donation will appear on the "acts" table with the corresponding campaign_id.

Note that you may find some entries in the table "acts" with no corresponding codes in the "actions" table, as if some people donated on solicitations they never received. This is the case for:

- Acquisition campaigns (you don't know how many people have been solicited, only those who responded).
- Spontaneous donations (a donor makes a donation without being asked to do so, or without using the coupon, so no link can be made between the donation and the solicitation).
- Etc.

In theory, the action_date should be when the solicitation has been sent, and the act_date when the donation was made in return.

However, action_date may also be set to when the list of people who would receive the solicitation letter has been tagged in the database, which might be weeks before the actual solicitation campaign happens. My suggestion would be not to rely too heavily on that field.

## Acts

The table "acts" corresponds to donations.

Note that a donor may make a donation without a corresponding action. For instance:

- He decides to spontaneously send a check
- He responds to a solicitation, but without using the coupon that accompanies the letter (in which case the charity cannot assign the donation to a specific solicitation)
- He makes a donation on the Internet
- Etc.

Most fields have corresponding reference tables:

- "act_type_id " indicates whether the donation is an automatic deduction (PA) or a one-off donation (DO).
- "payment_method_id" indicates how the donation has been made. The corresponding values can be found in the "payment_methods" table.
- "channel_id" indicates which channel has been used for the donation. Please refer to the "channels" reference table.

## Assignment2

Please refer to the document about that specific assignment to understand the content of that table.