DATA DICTIONARY

EUREKATHON

Challenging Data for Zero Hunger

Powered by

November 2020

PURPOSE OF THE DOCUMENT

This document contains all the details about the data available for the competition, from meaning of data fields to how the datasets were created.
Additionally, if you find other relevant datasets, publicly available, you can request the ingestion of that data into your machines. Please consult the Participant Guide for the details on how to do this.

We would like to remind you of the terms and conditions of the 2020 edition of EUREKATHON.

Confidential Information

1. The term Confidential Information (CI) means any and all information disclosed by Organizers and/or Mentors to the Participant in any form (including written, verbally, electronically, visually, or in any other tangible form, and all copies of any such information), which is identified as, notified as being, or which would be reasonably expected to be confidential or proprietary to the Organizers;

2. CI includes any data, software, software documentation, source code, documentation for functional specifications, development guidelines, inventions, training materials, third party confidential information, and any information given by the Organizers which is disclosed to the Participant in connection with the Event;

3. The Participant may only use CI disclosed to it under these regulations for the sole purpose of the Event;

4. The Participant must not disclose any CI to any third party without the express and prior written consent of the Organizers;

5. The Participant must not reverse engineer or decompile any software included in any CI or make any unauthorized copies of any CI;

6. In case the Participant is authorized by the Organizers to make copies of any CI, he/she shall do it exclusively for the purposes of the competition. All the copied CI, as well as the respective supports, must be eliminated as soon as the competition ends. The Participant can keep no copies of the information;

7. The Participant must not attempt to re-identify any personal information that has been de-identified in the CI.

Contents

BANCO ALIMENTAR DATA

REA (Rede de Emergência Alimentar) receives food assistance requests on a national level. The requests can be made either by filling in an online form or by calling a helpline answering the same question as in the online form. REA staff then briefly analyze the request, check for duplicates and identify the relevant institution that operates in the area of the residence of the applicant. Then, they forward the request to this institution if basic criteria are met (such as there exists an institution operating in the area that still has capacity to accept assistance requests). The institution receives the request and validates whether it meets their criteria for food assistance. The approval decision is based on the details of the form, as well as the data from Social Security database on income and expenditure of the applicant.

You will be given anonymized data from 20,699 food assistance requests, the information obtained from filling in the online form/answering questions during a call. Some questions in the form are multiple choice while others have an open answer. Approximately 30% of requests have registered outcomes (information whether they were approved or not).

Security/anonymization measures

In order to preserve privacy of the applicants for food assistance, some privacy measures were enforced in this challenge, such as:

1. Fields/variables containing personal identification are not present in the dataset;
2. Fields/variables that might contain any personal or sensitive information are not present. Relevant information from these fields was converted to flag variables as "is_reason_for_asking_unemployment";
3. Geographical location of the applicants was aggregated to ensure that 4-digit postal codes (CP4) with low frequency are not present;
4. Counties with low representativity were anonymized.
5. Field/variable age of the applicant was discretized into 10 bins (new values are from 0 to 9), where higher values correspond to higher quantiles;
6. Name of the institution to which the request was forwarded has been hashed.

Table: BA_requests_data_eurekathon

Description: Data from forms for food assistance requests.

Field	Description	Data type	Value example
zinf	Code of the branch of BA alimentar that operates in the county (concelho) of the request	string	01- BACF LISBOA

county	County (concelho) of residence, as indicated by the applicant	string	Lisboa
age_bin	Applicant age, as indicated by the applicant (binned)	numeric	1-10
cp4	CP4 of the applicant (is some cases the 3rd or 4th digit is replaced with x)	string	2735
lives_alone	Indicator does the applicant live alone, as indicated by the applicant	string	Sim
number_people_household	Number of people in the household of the applicant, as indicated by the applicant	string	1
number_kids_less_10yr	Number of kids younger than 10 in the applicant's household, as indicated by the applicant	string	1
food_support	Indicator whether the applicant already receives food support, as indicated by the applicant	string	Sim
first_time_asking_help	Indicator whether this is the first time that the applicant asks for food assistance, as indicated by the applicant	boolean	True
has_fridge	Indicator whether the applicant owns a fridge, as indicated by the applicant	string	Sim
has_freezer	Indicator whether the applicant owns a freezer, as indicated by the applicant	string	Sim
can_cook_home	Indicator whether the applicant has the conditions to cook at home, as indicated by the applicant	string	Sim
can_pickup_products_near_home	Indicator whether the applicant can pick up food supplies near home, as indicated by the applicant	string	Sim
request_reason	Reason for requiring food assistance (reason can be chosen from a predefined set of options), as indicated by the applicant	string	unemployed
is_reason_for_asking_haschildren	Additional justification of the request, as indicated by the applicant: Indicator whether the applicant specified that the reason for requiring assistance is "children"	boolean	True
is_reason_for_asking_unemployment	Additional justification of the request, as indicated by the	boolean	True

	applicant: Indicator whether the applicant specified that the reason for requiring assistance is "unemployment"		
is_reason_for_asking_pregnancy	Additional justification of the request, as indicated by the applicant: Indicator whether the applicant specified that the reason for requiring assistance is "pregnancy"	boolean	True
is_reason_for_asking_layoff	Additional justification of the request, as indicated by the applicant: Indicator whether the applicant specified that the reason for requiring assistance is "layoff"	boolean	True
is_reason_for_asking_sickness	Additional justification of the request, as indicated by the applicant: Indicator whether the applicant specified that the reason for requiring assistance is "health issues"	boolean	True
is_reason_for_asking_low_income	Additional justification of the request, as indicated by the applicant: Indicator whether the applicant specified that the reason for requiring assistance is "low income"	boolean	True
district	District of residence of the applicant	string	Lisboa
forwarding_month_id	Month when the request was forwarded to the institution operating in the area of the applicant	string	202004
forwarded_to	Institution operating in the area of the applicant to which the request was forwarded to	string hashed	9b2d5b4678781e53038e91ea5324530a03f27dc1d0e5f6c9bc9d493a23be9de0
forwarded_y_n	Indicator whether the request for food assistance was forwarded from the central point (REA) to the institution operating in the area of the applicant	string	Sim
reason_not_forwarded	Reason for not forwarding the request from the central point (REA)	string	Duplicated request

	to the institution operating in the area of the applicant (reason can be chosen from a predefined set of options)		
request_approved	Indicator whether the request for food assistance was approved	string	Sim
reason_not_approved	Reason for not approving the request (reason can be chosen from a predefined set of options)	string	No justification for support
outcome_observations_ incompleterequest	Additional clarification of the request outcome: Indicator whether the request was identified as incomplete	bool	True
outcome_observations_ duplicaterequest	Additional clarification of the request outcome: Indicator whether the request was identified as duplicate	bool	True
outcome_observations_lowprior ity	Additional clarification of the request outcome: Indicator whether the request was analysed and classified as low priority	bool	True
outcome_observations_ fooddelivered	Additional clarification of the request outcome: Indicator whether the applicant received food support as a result of this request	bool	True
month_id	Month when the request was made	string	202004

File: Eurekathon_Data_Campaigns.xlsx

Description: Food collected in the Lisbon region, per supermarket and time slot, in the two campaigns (May and November) organized by Banco Alimentar in 2019.

Field	Description	Data type	Value example
Data_Hora	Date and hour when food was collected	String	2020-05-25 12
nome_entidade	Name of the store where food was collected	string	Auchan Cascais
id_super	Index representing the store	integer	1
total	Total weight (in kg) of food collected in that store, date and hour	integer	1
media	Average weight (in kg) of food collected per container in that store, date and hour	integer	1
nrbx	Number of containers with food collected	integer	1

campanha	ID of the campaign where the collection took place (a campaign usually lasts 2 days)	integer	1
dia	Day of the campaign (a campaign usually lasts 2 days)	integer	1

File: Eurekathon_Data_GoogleAnalytics.xlsx

Description: Traffic data from websites bancoalimentar.pt and alimentestaideia.pt, extracted from the Google Analytics platform.

Field	Description	Data type	Value example
Website	The visited property (website).	string	bancoalimentar.pt
Days Since Last Session	The number of days elapsed since users last visited the website, used to calculate user loyalty.	integer	1
User Bucket	Randomly assigned users tag to allow A/B testing and splitting of remarketing lists. Ranges from 1-100.	integer	1
Session Duration	The length (returned as a string) of the session measured in seconds and reported in increments in seconds.	integer	1
Source / Medium	The source and type of referral (separated by a "/"). If users arrived without a referrer, its value is "(direct) / (none)".	string	(direct) / (none)
Date Hour and Minute	Combines the date, the hour and the minute in which the session started, formatted as YYYYMMDDHHMM.	integer	202008281425
Page	The visited page. Use this with hostname to get the page's full URL.	string	/quem-somos/pagina-noticias/noticias-federacao/campanha-de-papel-por-alimentos/
Previous Page Path	The page visited immediately before. Use this with hostname to get the page's full URL.	string	(entrance)
City	Users' city, derived from their IP addresses or Geographical IDs.	string	Funchal

Device	The brand of the mobile device used to access the page. If users did not access from a mobile device, its value is "Not Mobile".	string	Google Nexus 5
Users	The number of users that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
Sessions	The number of sessions that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
Bounces	The number of single page sessions that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
Time on Page	Average time (in seconds) spent by users on the page.	time	0.0
Exits	The number of exits (last pageview on the session) that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
Pageviews	The number of pageviews that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
Entrances	The number of entrances (first pageview on the session) that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
New Users	The number of new users (first session on the website) that match the given key (Days Since Last Session / User Bucket / Session Duration / Source / Medium / Date Hour and Minute / Page / City / Device).	integer	1
Hits	Total number of hits for the view (profile). This metric sums all hit types, including pageview, custom event, ecommerce, and other types.	integer	1

NOS CLIENT PROFILES DATA

You will be provided with 4 tables containing anonymized monthly information on characteristics of a sample of NOS clients with TV and mobile service (~460K). The sample is taken across a 6-month period, and for each client, data is provided for two consecutive months. Hence, for the first sample of clients, information from March and April is provided, for the second sample, information from April and May is available, for the third information is available from May and June, and for the fourth the information is for months June and July. For each sample of clients, in the month following the period for which their data is available, additional information is present regarding whether the client has performed an altruistic action. Therefore, for clients from the first sample, with their data available from March and April, additional variable is available in May with value equal to "1" if the client has shown altruistic behaviour in May. Similarly, for the second sample of clients, whose data is available for months April and May, information on whether they performed an altruistic action is available for the month of June. Additionally, fifth table will be available that contains characterization of NOS clients summarized at the geographical level.

Security/anonymization measures

As we care about the clients' privacy, several security and anonymization measures were applied in this challenge, such as:

1. Variables containing personal identification were hashed;
2. Parishes (Freguesias) with low population were anonymized;
3. 4-digit postal codes from the areas with low population were aggregated;
4. Continuous variables were discretized into 10 bins (new values are from 0 to 9), where higher values correspond to higher quantiles;
5. Details on types of service contracted, type of content and channel seen by clients, as well as any personal characteristics such as age and gender estimates are coded into generic categories (A,B,C,…) or replaced with integer IDs.

Table: client_list_eurekathon

Description: A list of all the clients that might appear the other tables: *client_tv_data_eurekathon, client_mobile_data_eurekathon and client_profile_data_eurekathon*. If the information for a client is available in this table for month_id= x, then the same client appears in the other tables for months x-1 and x-2.

Field	Description	Data type	Value example
sa	Identifier of the client	string	4329560142D21690C211BCA3EE65C83B
month_id	Identifier of the month	numeric	202005-202008

is_donor	Identifier whether the client performed an altruistic action during the designated month	numeric	1

Table: client_tv_data_eurekathon

Description: Monthly information regarding clients' tv consumption and some basic client characteristics. Only information extracted from TV Boxes is being measured (e.g. there is no online consumption).

Field	Description	Data type	Value example
sa	Identifier of the client	string	E68FC63E14BA6D8F58C8D37F414CE403
month_id	Identifier of the month	numeric	202003-202007
segment	Type of service that the client has. Coded: A,B,C…	string	A
package_type	Characterization of a bundle type that a client has (number of bundled services)	string	3P
fiscal_num_bin	Client's NIF (binned)	Numeric (1-10)	1-10
gender	Client's gender (coded 0,1)	numeric	0-1
subcategory_1	ID of the 1st most common category seen by that SA per month	numeric	1
subcategory_2	ID of the 2nd most common category seen by that SA per month	numeric	1
subcategory_3	ID of the 3rd most common category seen by that SA per month	numeric	1
subcategory_4	ID of the 4th most common category seen by that SA per month	numeric	1
subcategory_5	ID of the 5th most common category seen by that SA per month	numeric	1
subcategory_6	ID of the 6th most common category seen by that SA per month	numeric	1
subcategory_7	ID of the 7th most common category seen by that SA per month	numeric	1
subcategory_8	ID of the 8th most common category seen by that SA per month	numeric	1
subcategory_9	ID of the 9th most common category seen by that SA per month	numeric	1

subcategory_10	ID of the 10th most common category seen by that SA per month	numeric	1
subcategory_qtd_seconds_mon_1_bin	Number of seconds of the 1st most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_2_bin	Number of seconds of the 2nd most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_3_bin	Number of seconds of the 3rd most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_4_bin	Number of seconds of the 4th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_5_bin	Number of seconds of the 5th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_6_bin	Number of seconds of the 6th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_7_bin	Number of seconds of the 7th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_8_bin	Number of seconds of the 8th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_9_bin	Number of seconds of the 9th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
subcategory_qtd_seconds_mon_10_bin	Number of seconds of the 10th most common category seen by that SA per month (binned)	Numeric (1-10)	1-10
qtd_seconds_mon_bin	Total amount of seconds seen per month (binned)	Numeric (1-10)	1-10
nr_devices	Number of different TV Devices (boxes) belonging to the same SA (client)	numeric	1
nr_seconds_seen_0h2h_bin	Total number of seconds per month seen in the 2h interval between hours 0 and 2 (binned)	Numeric (1-10)	1-10
nr_seconds_seen_3h5h_bin	Total number of seconds per month seen in the 2h interval between hours 3 and 5 (binned)	Numeric (1-10)	1-10

nr_seconds_seen_6h8h_bin	Total number of seconds per month seen in the 2h interval between hours 6 and 8 (binned)	Numeric (1-10)	1-10
nr_seconds_seen_9h11h_bin	Total number of seconds per month seen in the 2h interval between hours 9 and 11 (binned)	Numeric (1-10)	1-10
nr_seconds_seen_12h14h_bin	Total number of seconds per month seen in the 2h interval between hours 12 and 14 (binned)	Numeric (1-10)	1-10
nr_seconds_seen_15h17h_bin	Total number of seconds per month seen in the 2h interval between hours 15 and 17 (binned)	Numeric (1-10)	1-10
nr_seconds_seen_18h20h_bin	Total number of seconds per month seen in the 2h interval between hours 18 and 20 (binned)	Numeric (1-10)	1-10
nr_seconds_seen_21h23h_bin	Total number of seconds per month seen in the 2h interval between hours 21 and 23 (binned)	Numeric (1-10)	1-10
channel_diversity_sa_bin	Number of channels seen by that SA per month (binned)	Numeric (1-10)	1-10

Table: client_mobile_data_eurekathon

Description: Monthly information regarding clients' mobile usage. A client may have one or more mobile phones numbers, and usage of each mobile phone is detailed separately. Hence, a client who has 2 mobile phone numbers in March, has two entries in this table for the month of March, each one for a different phone number.

Field	Description	Data type	Value example
sa	Identifier of the client	string	E68FC63E14BA6D8F58C8D37F414CE403
month_id	Identifier of the month	numeric	202003-202007
telephone_num	Identifier of the client's mobile phone number	string	1FEA52CFD37DD8900F028B73430402AB
nr_calls_sent_month_bin	Total number of calls sent per month (binned)	Numeric (1-10)	1-10

mean_call_sent_dur_bin	Average duration of the sent calls per month (binned)	Numeric (1-10)	1-10
min_call_sent_dur_bin	Minimum duration of the sent calls per month (binned)	Numeric (1-10)	1-10
max_call_sent_dur_bin	Maximum duration of the sent calls per month (binned)	Numeric (1-10)	1-10
nr_calls_sent_distinct_numbers_bin	Number of calls sent to distinct numbers (binned)	Numeric (1-10)	1-10
nr_distinct_cells_sent_bin	Number of different cell towers from which the calls where made (binned)	Numeric (1-10)	1-10
nr_distinct_freguesia_sent_bin	Number of different parishes from which the calls where made (binned)	Numeric (1-10)	1-10
nr_distinct_concelho_sent_bin	Number of different counties from which the calls where made (binned)	Numeric (1-10)	1-10
nr_calls_rec_month_bin	Number of received calls per month (binned)	Numeric (1-10)	1-10
mean_call_rec_dur_bin	Average duration of the received calls per month (binned)	Numeric (1-10)	1-10
min_call_rec_dur_bin	Minimum duration of the received calls per month (binned)	Numeric (1-10)	1-10
max_call_rec_dur_bin	Maximum duration of the received calls per month (binned)	Numeric (1-10)	1-10
nr_calls_rec_distinct_numbers_bin	Number of calls received from distinct numbers (binned)	Numeric (1-10)	1-10
nr_distinct_cells_rec_bin	Number of different cell towers on which the calls were received (binned)	Numeric (1-10)	1-10
nr_distinct_freguesia_rec_bin	Number of different parishes where the calls were received (binned)	Numeric (1-10)	1-10
nr_distinct_concelho_rec_bin	Number of different counties where the calls were received (binned)	Numeric (1-10)	1-10

freguesia_sent	Parish matching the most frequent cell tower when the number sends calls.	string	BENFICA
concelho_sent	County matching the most frequent cell tower when the number sends calls.	string	LISBOA
concelho_rec	County matching the most frequent cell tower when the number receives calls.	string	LISBOA
freguesia_rec	Parish matching the most frequent cell tower when the number receives calls.	string	BENFICA
cost_recharge_bin	Amount spent on buying data/ data refills (binned)	Numeric (1-10)	1-10
marketing_name	Make and model of most frequent equipment in that month associated with this phone number	string	Galaxy J4+
manufacturer	Manufacturer of the phone associated with this phone number	string	Apple Inc
rate_plan	Client's rate plan associated with this phone number *(tarifário)*: coded A, B, C,..	string	A
bands_supported	Supported frequency bands of the mobile phone associated with this phone number	string	GSM 1800GSM 900

Table: client_profile_data_eurekathon

Description: Monthly information regarding clients' profile, portfolio, churn, payment and collection process events. A collections (dunning) process is started every time a client misses a payment due date and consists of a timeline of warnings/actions to follow up on past due bills.

Field	Description	Data type	Value example
sa	Identifier of the client	string	E68FC63E14BA6D8F58C8D37F414CE403
month_id	Identifier of the month	numeric	202003-202007

cp4	4-digit postal code (CP4) of client's residence. In some cases, 3rd or 4th digit is replaced with x.	string	1000
dsc_class_churn	Identifier of client churn event (type 1 or type 2)	numeric	0-2
geom_distrito_dsc	District of client's residence	string	LISBOA (DISTRITO)
geom_municipio_dsc	Municipality of client's residence	string	LISBOA (MUNICIPIO)
geom_freguesia_dsc	Parish of client's residence	string	BENFICA
geom_nuts_i_dsc	Name of NUTS statistical regions of Portugal, level I	string	CONTINENTE
geom_nuts_ii_dsc	Name of NUTS statistical regions of Portugal, level II	string	LISBOA (NUTS II)
geom_nuts_iii_dsc	Name of NUTS statistical regions of Portugal, level III	string	GRANDE LISBOA
after_due_days_qty_bin	Number of days after invoice payment due date that led the client to a collections/dunning process. (binned)	Numeric (1-10)	1-10
step_entry_open_amt_bin	Debt in € at the time of entering collections process (binned).	Numeric (1-10)	1-10
pack_tem_a_flg	Flag indicating if the client pack has A type add on	numeric	0-1
pack_tem_b_flg	Flag indicating if the client pack has B type add on	numeric	0-1
pack_tem_c_flg	Flag indicating if the client pack has C type add on	numeric	0-1
typeA_card_qty	Quantity of mobile SIM cards of type A that the client has	numeric	1
typeB_card_qty	Quantity of mobile SIM cards of type B that the client has	numeric	1
typeC_card_qty	Quantity of mobile SIM cards of type C that the client has	numeric	1
typeD_card_qty	Quantity of mobile SIM cards of type D that the client has	numeric	1
typeE_card_qty	Quantity of mobile SIM cards of type E that the client has	numeric	1
typeF_card_qty	Quantity of mobile SIM cards of type F that the client has	numeric	1
typeG_card_qty	Quantity of mobile SIM cards of type G that the client has	numeric	1
typeH_card_qty	Quantity of mobile SIM cards of type H that the client has	numeric	1

typeI_card_qty	Quantity of mobile SIM cards of type I that the client has	numeric	1
premium_qty	Quantity of premium channels that the client has	numeric	1
premium_flg	Flag indicating if client has at least one premium channel	numeric	0-1
pack_dsc	Description of a bundle (pacote) that the client has	string	NOS3i_100Mb (042)
premium_A_qty	Quantity of premium channels of category A that the client has	numeric	1
premium_B_qty	Quantity of premium channels of category B that the client has	numeric	1
premium_C_qty	Quantity of premium channels of category C that the client has	numeric	1
premium_D_qty	Quantity of premium channels of category D that the client has	numeric	1
premium_A_amt_bin	Amount paid on premium channels of category A (binned)	Numeric (1-10)	1-10
premium_B_amt_bin	Amount paid on premium channels of category B (binned)	Numeric (1-10)	1-10
premium_C_amt_bin	Amount paid on premium channels of category C (binned)	Numeric (1-10)	1-10
premium_D_amt_bin	Amount paid on premium channels of category D (binned)	Numeric (1-10)	1-10

Table: indicators_data_eurekathon

Description: Monthly information regarding clients' profile, portfolio, churn, payment and collection process events characterized at the level of a parish. A collections (dunning) process is started every time a client misses a payment due date and consists of a timeline of warnings/actions to follow up on past due bills.

Field	Description	Data type	Value example
month_id	Identifier of the month	numeric	202003-202007
cp4	4-digit postal code (CP4) of client's residence. In some cases, 3rd or 4th digit is replaced with x.	string	1000
geom_distrito_dsc	District of client's residence	string	LISBOA (DISTRITO)

geom_municipio_dsc	Municipality of client's residence	string	LISBOA (MUNICIPIO)
geom_freguesia_dsc	Parish of client's residence	string	BENFICA
geom_nuts_i_dsc	Name of NUTS statistical regions of Portugal, level I	string	CONTINENTE
geom_nuts_ii_dsc	Name of NUTS statistical regions of Portugal, level II	string	LISBOA (NUTS II)
geom_nuts_iii_dsc	Name of NUTS statistical regions of Portugal, level III	string	GRANDE LISBOA
churn_rate_1_bin	Percentage of clients labelled as type 1 churn (binned)	Numeric (1-10)	1-10
churn_rate_2_bin	Percentage of clients labelled as type 2 churn (binned)	Numeric (1-10)	1-10
stepA_rate_bin	Percentage of clients in step A of collections process (binned)	Numeric (1-10)	1-10
stepB_rate_bin	Percentage of clients in step B of collections process (binned)	Numeric (1-10)	1-10
stepC_rate_bin	Percentage of clients in step C of collections process (binned)	Numeric (1-10)	1-10
stepD_rate_bin	Percentage of clients in step D of collections process (binned)	Numeric (1-10)	1-10
total_debt_bin	Total debt of clients in collections (binned)	Numeric (1-10)	1-10
mean_debt_bin	Average debt of clients in collections (binned)	Numeric (1-10)	1-10
avg_premium_flg	Average number of clients with flag indicating if client has any premium channel	numeric	0.2
avg_pack_tem_a_flg	Average number of clients with flag indicating if client pack has A add on	numeric	0.2
avg_pack_tem_b_flg	Average number of clients with flag indicating if client pack has B add on	numeric	0.2
avg_pack_tem_c_flg	Average number of clients with flag indicating if client pack has C add on	numeric	0.2
avg_typeA_card_qty	Average number of mobile SIM cards of type A per client	numeric	0.2
avg_typeB_card_qty	Average number of mobile SIM cards of type B per client	numeric	0.2
avg_typeC_card_qty	Average number of mobile SIM cards of type C per client	numeric	0.2
avg_typeD_card_qty	Average number of mobile SIM cards of type D per client	numeric	0.2
avg_typeE_card_qty	Average number of mobile SIM cards of type E per client	numeric	0.2
avg_typeF_card_qty	Average number of mobile SIM cards of type F per client	numeric	0.2

avg_typeG_card_qty	Average number of mobile SIM cards of type G per client	numeric	0.2
avg_typeH_card_qty	Average number of mobile SIM cards of type H per client	numeric	0.2
avg_typeI_card_qty	Average number of mobile SIM cards of type I per client	numeric	0.2
avg_arput1_bin	Average of ARPUT (average revenue per user) type 1 per client (binned)	Numeric (1-10)	1-10
avg_arput2_bin	Average of ARPUT (average revenue per user) type 2 per client (binned)	Numeric (1-10)	1-10
avg_premium_A_qty	Average quantity of premium channels of category A	numeric	0.2
avg_premium_B_qty	Average quantity of premium channels of category B	numeric	0.2
avg_premium_C_qty	Average quantity of premium channels of category C	numeric	0.2
avg_premium_D_qty	Average quantity of premium channels of category D	numeric	0.2
avg_premium_A_amt_bin	Average amount paid on premium channels of category A (binned)	Numeric (1-10)	1-10
avg_premium_B_amt_bin	Average amount paid on premium channels of category B (binned)	Numeric (1-10)	1-10
avg_premium_C_amt_bin	Average amount paid on premium channels of category C (binned)	Numeric (1-10)	1-10
avg_premium_D_amt_bin	Average amount paid on premium channels of category D (binned)	Numeric (1-10)	1-10

EXTERNAL DATA SOURCES

File: Eurekathon_Data_GeoMappingPT.xlsx

Description: Mapping between the different statistical subdivisions of Portugal's territory, complemented with demographic data (from the last census – 2011).

Field	Description	Data type	Value example
codseccao	Primary key and ID of the "secção estatística", subdivision of Portugal employed for statistical purposes.	integer	100105006

dtccfr	ID of the "freguesia" (parish), subdivision of Portugal employed for statistical purposes, to which the "secção estatística" belongs.	integer	100105
freguesia	Name of parish	string	Benfica
dtcc	ID of the "concelho" (county), subdivision of Portugal employed for statistical purposes, to which the "freguesia" belongs.	integer	1106
concelho	Name of county	string	Lisboa
dt	ID of the "distrito" (district), subdivision of Portugal employed for statistical purposes, to which the county belongs.	integer	11
distrito	Name of district	string	Lisboa
cod_nutsiii	ID of the "NUTS III", subdivision of Portugal employed for statistical purposes, to which the "concelho" belongs.	string	PT170
nutsiii	Name of the "NUTS III".	string	Area Metropolitana de Lisboa
cod_nutsii	ID of the "NUTS II", subdivision of Portugal employed for statistical purposes, to which the "NUTS III" belongs.	string	PT17
nutsii	Name of the "NUTS II".	string	Area Metropolitana de Lisboa
cod_nutsi	ID of the "NUTS I", subdivision of Portugal employed for statistical purposes, to which the "NUTS II" belongs.	string	PT1
nutsi	Name of the "NUTS I".	string	Portugal Continental
residentes_total	Total number of inhabitants of the "secção estatística", according to the last census (2011).	integer	100
edificios	Total number of buildings in the "secção estatística", according to the last census (2011).	integer	100
residentes_homens	Total number of male inhabitants of the "secção estatística", according to the last census (2011).	integer	100
alojamento	Total number of residential units in the "secção estatística", according to the last census (2011).	integer	100
familias	Total number of family units living in the "secção estatística", according to the last census (2011).	integer	100
long	Longitude of the centroid of the "secção estatística".	float	-9,19749
lat	Latitude of the centroid of the "secção estatística".	float	38,74982

Table: geo_mapping_eurekathon

Description: Mapping between CP4 4-digit postal code and the territorial divisions county, municipality and parish.

Field	Description	Data type	Value example
CP4	4-digit postal code (CP4)	string	4485
county	County of country (Distrito)	string	PORTO (DISTRITO)
municipality	Municipality name (Município)	string	VILA DO CONDE (MUNICIPIO)
parish	Parish name (Freguesia)	string	GIÃO

Pordata source

Description: Files extracted from PORDATA's official website (https://www.pordata.pt/) containing information regarding the different indicators related to population density, minimum wage per location, unemployment, job offers or Social Security subsidies, amongst others. These contain location identifiers such as *municipality* that can be crossed with other datasets (use Latin-1 encoding when reading these fields). Each file is self-explanatory and contains a brief data description.

Pordata files

- PORDATA_Diferença-entre-o-salário-mínimo-nacional-e-a-remuneração-base-média-mensal.xlsx
  Difference between the national minimal wage and average monthly basic remuneration

- PORDATA_Estimativas-a-31-12.xlsx
  Resident population, estimates at December 31st

- PORDATA_Inactivos-por-100-activos-segundo-os-Censos.xlsx
  Inactive population per 100 active persons, according to the Census

- PORDATA_Inscrita-nos-centros-de-emprego-em-percentagem-da-população-residente.xlsx
  Unemployment registered at the public employment office in total of resident population aged 15 to 64 (%)

- PORDATA_Inscrita-nos-centros-de-emprego-por-sexo.xlsx
  Unemployment registered at the public employment office (yearly average): total and by sex

- PORDATA_Inscrita-nos-centros-de-emprego-por-tempo-de-inscrição.xlsx

Unemployment registered at the public employment office (yearly average): total and by length of registration

- PORDATA_Inscrita-nos-centros-de-emprego-por-tipo-de-desemprego.xlsx
  Unemployment registered at the public employment office (yearly average): total and by unemployment type

- PORDATA_Ofertas-de-emprego-por-grandes-sectores-de-actividade-económica-(2001-2014).xlsx
  Job offers (yearly average) available at the public employment office: total and by large sectors of economic activity (2001-2014)

- PORDATA_Pensões-da-Segurança-Social-e-CGA-em-percentagem-da-população-residente.xlsx
  Pensions as a % of resident population aged 15 and over: total, Social Security and Public Administration Retirement Fund

- PORDATA_Pensões-de-sobrevivência-invalidez-e-velhice.xlsx
  Social Security pensions: total, survivors pensioners, disability pensioners, old-age pensioners

- PORDATA_Por-condição-perante-o-trabalho-segundo-os-Censos.xlsx
  Inactive population, according to the Census: total and by work situation

- PORDATA_Remuneração-por-sexo.xlsx
  Average monthly wage of employees: basic remuneration and earnings by sex

- PORDATA_RMG-e-RSI-em-percentagem-da-população-residente.xlsx
  Social Security beneficiaries of Guaranteed Minimum Income and Social Integration Benefit in total of resident population aged 15 and over

- PORDATA_RMG-e-RSI-em-percentagem-dos-beneficiários-activos.csv
  Social Security beneficiaries of Guaranteed Minimum Income and Social Integration Benefit in total of active beneficiaries (%)

- PORDATA_RMG-e-RSI-em-percentagem-dos-beneficiários-activos.xlsx
  Social Security beneficiaries of Guaranteed Minimum Income and Social Integration Benefit in total of active beneficiaries (%)

- PORDATA_RMG-e-RSI-por-grupo-etário.xlsx
  Social Security beneficiaries of Guaranteed Minimum Income and Social Integration Benefit: total and by age group

- PORDATA_RMG-e-RSI-por-sexo.xlsx
  Social Security beneficiaries of Guaranteed Minimum Income and Social Integration Benefit: total and by age group

- PORDATA_Subsídio-de-bonificação-por-deficiência.xlsx

Disability allowance of Social Security

- PORDATA_Subsídio-de-desemprego-em-percentagem-da-população-residente.xlsx
  Social Security unemployment benefits in total of resident population aged 15 and over (%)

- PORDATA_Subsídio-de-desemprego-em-percentagem-dos-beneficiários-activos.xlsx
  Social Security beneficiaries of unemployment benefit in total of active beneficiaries (%)

- PORDATA_Subsídio-de-desemprego-por-sexo.xlsx
  Social Security beneficiaries of unemployment benefit: total and by sex

- PORDATA_Subsídio-por-assistência-à-3ª-pessoa.xlsx
  Tertiary care allowance of Social Security

- PORDATA_Subsídio-por-doença-por-sexo.xlsx
  Social Security beneficiaries of sickness benefit: total and by sex

- PORDATA_Subsídio-social-de-desemprego-em-percentagem-dos-beneficiários-activos.xlsx
  Social Security beneficiaries of supplementary unemployment benefit in total of active beneficiaries (%)

- PORDATA_Subsídio-social-de-desemprego-por-sexo.xlsx
  Social Security beneficiaries of supplementary unemployment benefit: total and by sex

DSSG source

Description: Files extracted from Data Science for Social Good DSSG github (https://github.com/dssg-pt/dados-MTSSS) containing information from MTSSS (Ministério do Trabalho, Solidariedade e Segurança Social) on the impact of COVID19 on the job market regarding sick leave, layoffs, collective dismissals and activity reduction requests per location identifier. Some processing was performed on the original datasets. More detailed documentation can be found on DSSG github.

File: mtss_timeline.csv

Field	Description	Data type	Value example
date	Date of data	date	01/03/2020
nr_sickleave_requests	Number of sick leave requests per day	integer	100
cum_nr_sickleave_requests	Cumulative number of sick leave requests	integer	1000

cum_nr_workers_layoff	Cumulative number of layoff workers	integer	1000
cum_total_salary	Cumulative total salary of layoff workers	integer	100000
total_collective_dismissal_requests	Collective dismissals - total number of collective dismissal requests (cumulative counts with restart every month)	integer	100
total_collective_dismissal_requests_micro	Collective dismissals - total number of collective dismissal requests in micro companies (1 to 9 employees) (cumulative counts with restart every month)	integer	100
total_workers_dismissed	Collective dismissals - total number of dismissed workers (cumulative counts with restart every month)	integer	100
total_workers_dismissed_micro	Collective dismissals - total number of dismissed workers in micro companies (1 to 9 employees) (cumulative counts with restart every month)	integer	100
nr_requests_fullstop_independentworkers	Activity reduction - number of full stop requests for independent workers	integer	100
nr_requests_reduction_independentworkers	Activity reduction - number of reduction requests for independent workers	integer	100
total_requests_independentworkers	Activity reduction - total full stop requests for independent workers	integer	100
nr_requests_fullstop_independentworkers_extension	Activity reduction - number of activity full stop extension requests for independent workers	integer	100
nr_requests_reduction_independentworkers_extension	Activity reduction - number of activity full stop extension requests for independent workers	integer	100
total_requests_independentworkers_extension	Activity reduction - total full stop extension requests for independent workers	integer	100
nr_requests_fullstop_moe	Activity reduction - number of activity full stop requests for MOE workers (membro de orgão estatutário)	integer	100
nr_requests_reduction_moe	Activity reduction - number of activity reduction requests for MOE workers (membro de orgão estatutário)	integer	100
total_requests_moe	Activity reduction - total full stop requests for MOE workers (membro de orgão estatutário)	integer	100
nr_requests_fullstop_moe_extension	Activity reduction - number of activity full stop extension requests for MOE workers (membro de orgão estatutário)	integer	100
nr_requests_reduction_moe_extension	Activity reduction - number of activity full stop extension requests for MOE workers (membro de orgão estatutário)	integer	100

total_requests_moe_extension	Activity reduction - total full stop extension requests for MOE workers (membro de orgão estatutário)	integer	100

File: mtsss_company_sector_layoff.csv

Field	Description	Data type	Value example
company_sector	Company industry sector	string	Electricidade
date	Date of data	date	2020-04-17
nr_companies_layoff	Number of companies in layoff	integer	100
total_workers	Total number of workers in layoff	integer	100
nr_workers_f	Number of female workers in layoff	integer	100
nr_workers_m	Number of male workers in layoff	integer	100

File: mtsss_layoff_geo.csv

Field	Description	Data type	Value example
region	Region of country	string	Lisboa
nr_companies_layoff	Number of companies in layoff	integer	100
perc_companies_layoff_region	Percentage of companies in layoff per region	float	25

File: mtsss_sickleave_geo.csv

Field	Description	Data type	Value example
county	County	string	LISBOA
total	Total number of requested sick leaves	integer	100

File: mtsss_company_layoff_sickleave_sector.csv

Field	Description	Data type	Value example
company_sector	Company industry sector	string	Electricidade

total_workers	Total number of workers	integer	100
nr_workers_f	Total number of female workers	integer	100
nr_worker_m	Total number of male workers	integer	100
nr_companies	Total number of companies	integer	100
nr_workers_layoff	Total number of workers in layoff	integer	100
nr_workers_layoff_f	Total number of female workers in layoff	integer	100
nr_workers_layoff_m	Total number of male workers in layoff	integer	100

GoogleMaps source

File: Eurekathon_Data_StoreFeatures.csv

Description: Location and Google Maps' ratings of supermarkets located in the Lisbon region.

Field	Description	Data type	Value example
store	Name of the store	string	Auchan Cascais
lat	Latitude of the store location	floating point	38.7020315
lng	Longitude of the store location	floating point	-9.4163201
ratings	Number of total user ratings on Google Maps platform	integer	5
rating	Average rating score (on a scale 1-5) of the place on Google Maps platform reviews	floating point	4

File: Eurekathon_Data_DistanceMatrix.csv

Description: Travelling distances and times between locations in the Lisbon region (locations considered include every parish and a comprehensive list of supermarkets).

Field	Description	Data type	Value example
origin	Place of origin, when calculating a distance, that can be a supermarket or a parish (freguesia)	string	Aguas Livres
type	Distance or time: type of value that is registered, which measures the distance in km or min between origin and destination	string	distance

destination	Place of destination, when calculating a distance, that can be a supermarket	string	Auchan Cascais
Value	Distance (in km) or time (in min) of the route between origin and destination	string	10
local_type	Type of origin place, a supermarket or a parish (freguesia)	string	freguesia

File: Eurekathon_Data_Footfall.csv

Description: Hourly distribution of visits for supermarkets located in the Lisbon region, per day of the week (source: Google Maps – "Popular times").

Field	Description	Data type	Value example
store	Name of the store which footfall data is registered	string	Auchan Cascais
day	Day of the week which footfall data is registered	string	Monday
0 - 23	Footfall level (number of people in that place) in that store, day and hour of the day. The name of the column specifies the hour interval. Footfall is measured as a relative value from 0-100, where 100 represents the maximum number of people in that place in the entire week	integer	1