

Data description

In this section, you can find the description of the different datasets that are provided to address the Challenge's task. At the end of the section, you will find the links to download the data.

Train Dataset

As for training, one file is provided: **train_data.parquet**. The image below illustrates what a row of the train dataset looks like:

sku	date	sold quantity	current price	currency	listing type	shipping logistic type	shipping payment	minutes active
464801	2021-02-01	0	156.78	REA	classic	fulfillment	free_shipping	1440
464801	2021-02-02	0	156.78	REA	classic	fulfillment	free_shipping	1440
464801	2021-02-03	0	156.78	REA	classic	fulfillment	free_shipping	1440
464801	2021-02-04	0	156.78	REA	classic	fulfillment	free_shipping	1440
464801	2021-02-05	1	156.78	REA	classic	fulfillment	free_shipping	1440

This dataset comprises two months (February and March 2021) of sales data at a daily level for a subset of Mercadolibre's *SKUs* (stock keeping units). Each row corresponds to a particular date-*SKU* combination. You can think of a *SKU* as a combination of an Item and a variation. An item could be, for example, a "t-shirt of X brand" and a variation of it could be "size M, color black". So, we are interested in predicting the stock at *SKU* level because you could run out of stock for the color black but you could have plenty of the t-shirt X for color brown. Besides *SKU* and date, for each row, the following fields are available:

Attributes	Description
sold_quantity	number of units of the corresponding <i>SKU</i> that were sold on that particular date.
current_price	point in time correct <i>SKU</i> 's price.
currency	currency in which the price is expressed.
listing_type	type of listing the <i>SKU</i> had for that particular date. Possible values are <i>classic</i> or <i>premium</i> and they relate to the exposure the items have and the fee charged to the seller as a sales comission. Another important advantage for an item listed as premium is its capability to pay in installments without interest rate.
shipping_logistic_type	type of shipping method the <i>SKU</i> offered, for that particular date. Possible values are <i>fulfillment</i> , <i>cross_docking</i> and <i>drop_off</i> .
shipping_payment	whether the shipping for the offered <i>SKU</i> at that particular date was free or paid, from the buyer's perspective.
minutes_active	number of minutes the <i>SKU</i> was available for purchase on that particular date.

Test Dataset

For testing, the following file is provided **test_data.csv**. This file contains only two columns:

Attribute	Description
SKU	indicates the <i>SKU</i> for which you have to make your prediction
target_stock	inventory level (<i>aka</i> number of units of the corresponding <i>SKU</i> for which you have to provide your estimation of inventory days.

Items Data

In the file **items_static_metadata.jl** (notice that the file is actually a jsonline file) there is some extra data related to the *SKUs* characteristics. It comprises a list of dicts where each dict contains a specific *SKU* metadata. The following fields are available:

Attribute	Description
SKU	stock-keeping-unit. This is a unique identifier for each physical different unit of inventory.
item_id	unique identifier of the listing the <i>SKU</i> belongs to. The same listing can be associated with more than one <i>SKUs</i> , for example, the <i>SKUs</i> "t-shirt X, size M color Black" and "t-shirt X, size M, color red" share the same <i>item_id</i> which is "t-shirt X".
item_domain_id	listing's domain id. A domain is a kind of listings clustering within MercadoLibre. For example "t-shirt X" could be in the domain <i>MLB_SPORT_TSHIRTS</i> .
item_title	the listing's title. The title is at item level. So, in the previous example, one possible title could be "t-shirt X".
site_id	the MercadoLibre's site the listing belongs to. The labels <i>MLB</i> , <i>MLA</i> and <i>MLM</i> refer to Brazil, Argentina and Mexico respectively.
product_id	listing product id. Field might be null for some listings. Because an item is listed for the seller, it is often common that different sellers sell the same thing (same product). This could be the case of "t-shirt X" if it is sold for many sellers. So in an intent to catalog the same product, MercadoLibre assigns the same <i>product_id</i> to all those items.
product_id_family	listing product family id. Field might be null for some listings. This is the same as above, but with a high hierarchy catalog product.

Sample submission

We provide a sample submission file (**sample_submission.csv.gz**) for you to visualize the expected format of a submission.

A detailed explanation of the expected format of your submission can be found in the [Evaluation and rules section](#).

Download links

- [test_data.csv](#)
- [train_data.parquet](#)
- [items_static_metadata.jl](#)
- [sample_submission.csv.gz](#)

 License Dataset:



Copyright © 1999-2021 MercadoLibre S.R.L.

[Work with us](#) | [Privacy Policies](#) | [The Licenses](#) | [Terms and Conditions](#) | [Contact us](#)