# Business Understanding

## Description

Imagine being hungry in an unfamiliar part of town and getting restaurant recommendations served up, based on your personal preferences, at just the right moment. The recommendation comes with an attached discount from your credit card provider for a local place around the corner!

Right now, [Elo](https://www.cartaoelo.com.br/), one of the largest payment brands in Brazil, has built partnerships with merchants in order to offer promotions or discounts to cardholders. But do these promotions work for either the consumer or the merchant? Do customers enjoy their experience? Do merchants see repeat business? Personalization is key.

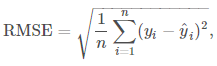
Elo has built machine learning models to understand the most important aspects and preferences in their customers’ lifecycle, from food to shopping. But so far none of them is specifically tailored for an individual or profile. This is where you come in.

In this competition, Kagglers will develop algorithms to identify and serve the most relevant opportunities to individuals, by uncovering signal in customer loyalty. Your input will improve customers’ lives and help Elo reduce unwanted campaigns, to create the right experience for customers.

## Evaluation

**1. RMSE:**

Submissions are scored on the root mean squared error. RMSE is defined as:



where y^  is the predicted loyalty score for each card\_id, and y is the actual loyalty score assigned to a card\_id.

**2.Submission File:**

card\_id, target

C\_ID\_9e86007114,0

C\_ID\_1c9f77086c,0.5

C\_ID\_07b20e9908,0

C\_ID\_63d6bac69a,0

C\_ID\_bbc26a86eb,0

C\_ID\_f749aad790,0

C\_ID\_7b5c15ff41,-0.25

C\_ID\_ec6b0f2d30,0

C\_ID\_0a11e759c5,0

# Data Understanding

## What files do I need?

The **train.csv** and **test.csv** files contain the card\_ids that we'll be using for training and prediction.

The **historical\_transactions.csv** and**new\_merchant\_transactions.csv** files contain information about each card's transactions.**historical\_transactions.csv** contains up to 3 months' worth of transactions for every card at any of the provided merchant\_ids. **new\_merchant\_transactions.csv** contains the transactions at new merchants (merchant\_ids that this particular card\_id has not yet visited) over a period of two months.

**merchants.csv** contains aggregate information for each merchant\_id represented in the data set.

## What should I expect the data format to be?

The data is formatted as follows:

**train.csv** and **test.csv** contain card\_ids and information about the card itself - the first month the card was active, etc. **train.csv** also contains the target.

**historical\_transactions.csv** and **new\_merchant\_transactions.csv** are designed to be joined with **train.csv**, **test.csv**, and **merchants.csv**. They contain information about transactions for each card, as described above.

**merchants.csv** can be joined with the transaction sets to provide additional merchant-level information.

## What am I predicting?

You are predicting a loyalty score for each card\_id represented in **test.csv** and **sample\_submission.csv**.

## File Descriptions

* train.csv - the training set
* test.csv - the test set
* sample\_submission.csv - a sample submission file in the correct format - contains all card\_ids you are expected to predict for.
* historical\_transactions.csv - up to 3 months' worth of historical transactions for each card\_id
* new\_merchant\_transactions.csv - two months' worth of data for each card\_id containing ALL purchases that card\_id made at merchant\_ids that were *not visited in the historical data*.
* merchants.csv - additional information about all merchants / merchant\_ids in the dataset.

## Data Fields

Data field descriptions are provided in **Data Dictionary.xlsx**.

## Data Volume

Train.csv: 201917

Test.csv : 123623

historical\_transactions.csv : 29112361

new\_merchant\_transactions.csv: 1963031

merchants.csv: 334696

## Null values

**key=historical\_transactions:**

merchant\_id 138481

category\_3 178159

category\_2 2652864

dtype: int64

**key=new\_merchant\_transactions:**

[2018-12-05 23:17:20,824] data\_provider.py:171:INFO - null counts:

merchant\_id 26216

category\_3 55922

category\_2 111745

dtype: int64

## Others

Target: They have 1 unique value: -33.21928095. This seems to be a special case. Maybe it would be reasonable to simply exclude these samples. We'll try later.