1. **Elo Merchant Category Recommendation**

This project is intended to help understand customer loyalty and build a recommendation engine with discount from credit card provider.

**Problem-**

Build machine learning model to predict loyalty score for card id’s in test dataset. Training dataset contains loyalty score for each card id, historical transactions and new merchant transactions contain information about each card's transactions, and merchants.csv contains aggregate information for each merchant\_id

**Client-**

Elo one of the largest payment brands in Brazil.

Why do they care about the problem?

Elo 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.

**Data**

<https://www.kaggle.com/c/elo-merchant-category-recommendation/data>

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

**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
* merchants.csv - additional information about all merchants / merchant\_ids in the dataset.
* 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*.

**Data fields**

Data field descriptions are provided in Data Dictionary.xlsx.

**Modeling Approach**

Linear regression model will be used to predict loyalty score for each card id.

Feature engineering methods will be applied on dataset. These features shall be used for modelling loyalty score. Also feature importance shall be generated.

**Deliverables**

1. Code (Jupyter notebook)
2. Report on finding of the project