TEAM NAME
Tom'N Jerry

TEAM MEMBER
Xia Feng
Boxue Li
Peiqing Lian
Xinrong Lian
Ruoxuan Zhou

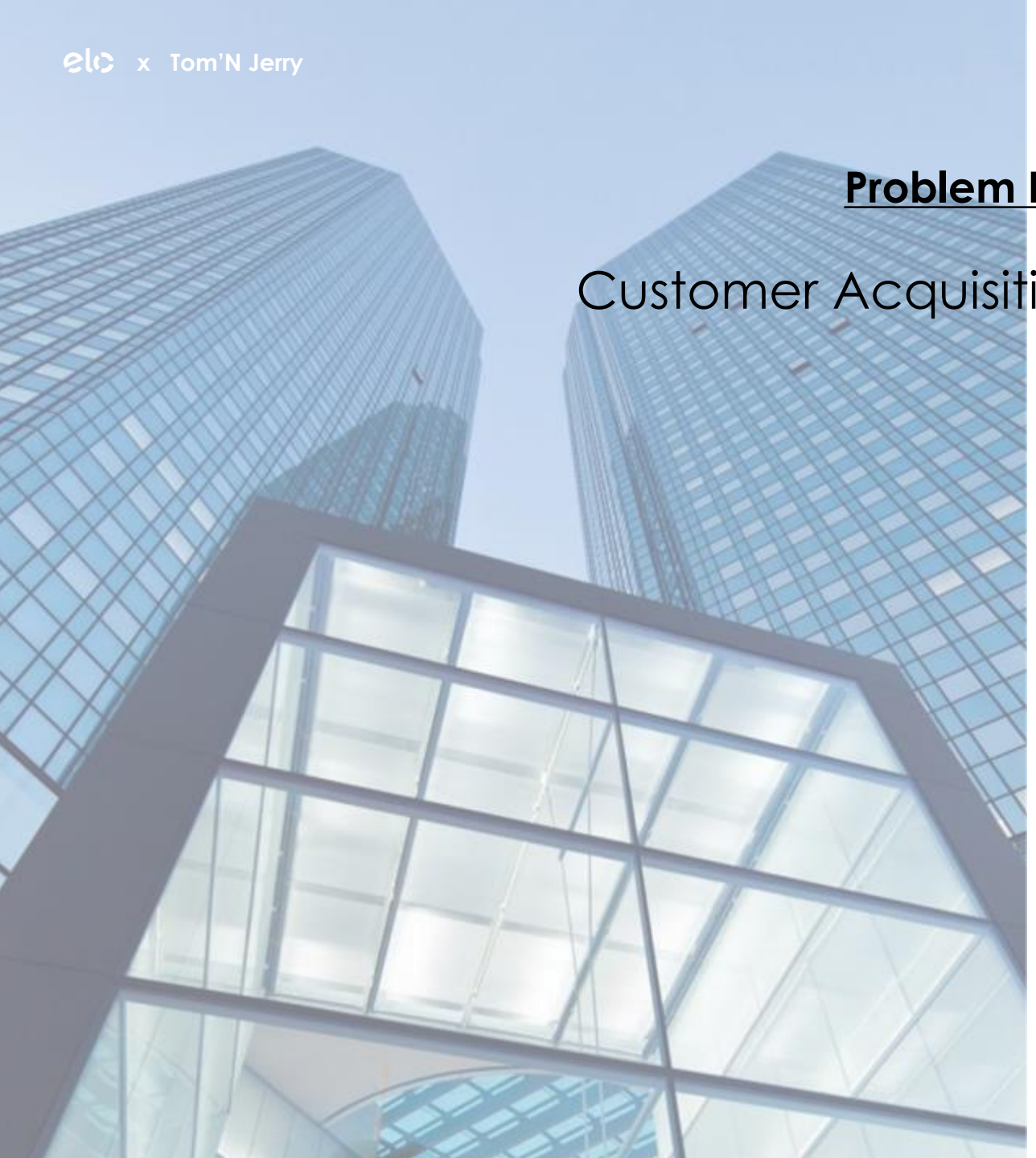
Brazilian Financial Association

ELO

Customer Relationship Promotion Project

CONTENTS

Part 01 Problem Definition



Problem Definition

Customer Acquisition and Retention.

1. Analyze the customers' data from the Elo and understand the buying habits based on different clusters' customers.

2. Provide personalized strategies for Elo to improve the retention rate and acquire more customers

Part 02 Data & Model Description

2.1 3 Technical Approach

Data cleaning
Data aggregation
Train prediction model
Build k-means cluster model

Data visualization
Cluster analysis

Data visualization

2.2 Data Manipulation

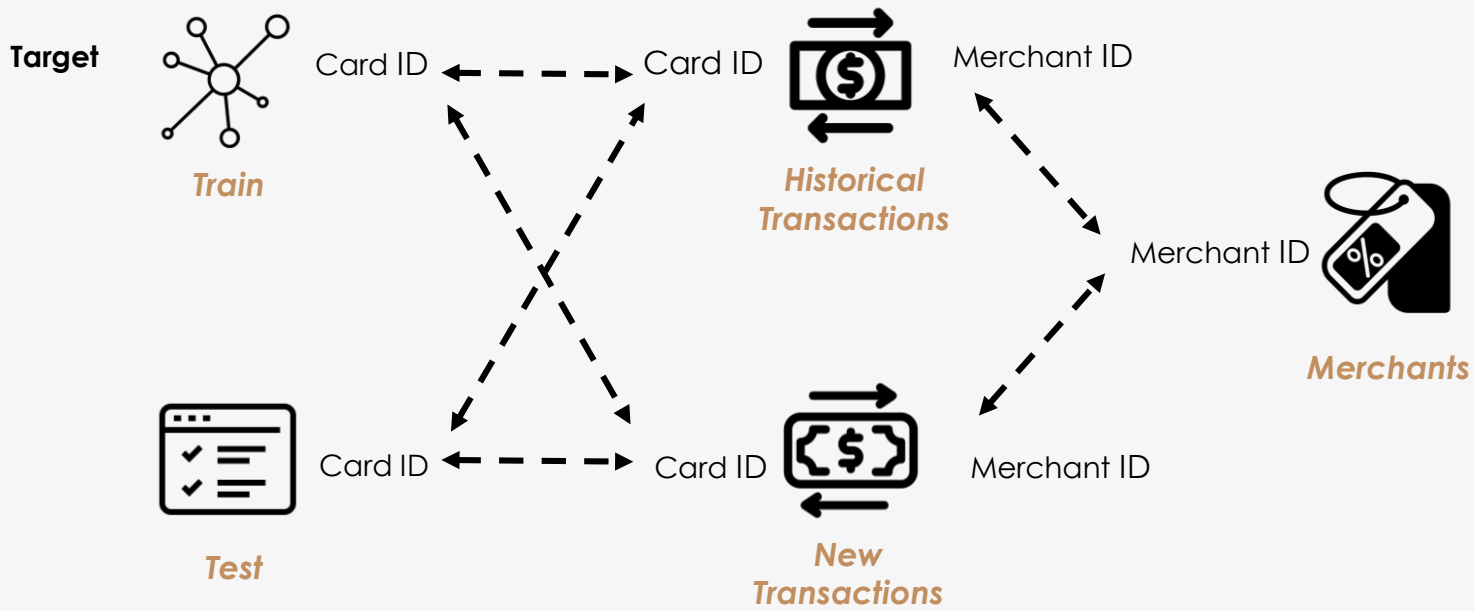
The competition involves 5 datasets.

Historical before new merchants were recommended

New after new merchants were recommended

The datasets are largely anonymized, and the meaning of the features are not elaborated.

[Visualizing Dataset Relationship]



Step 1

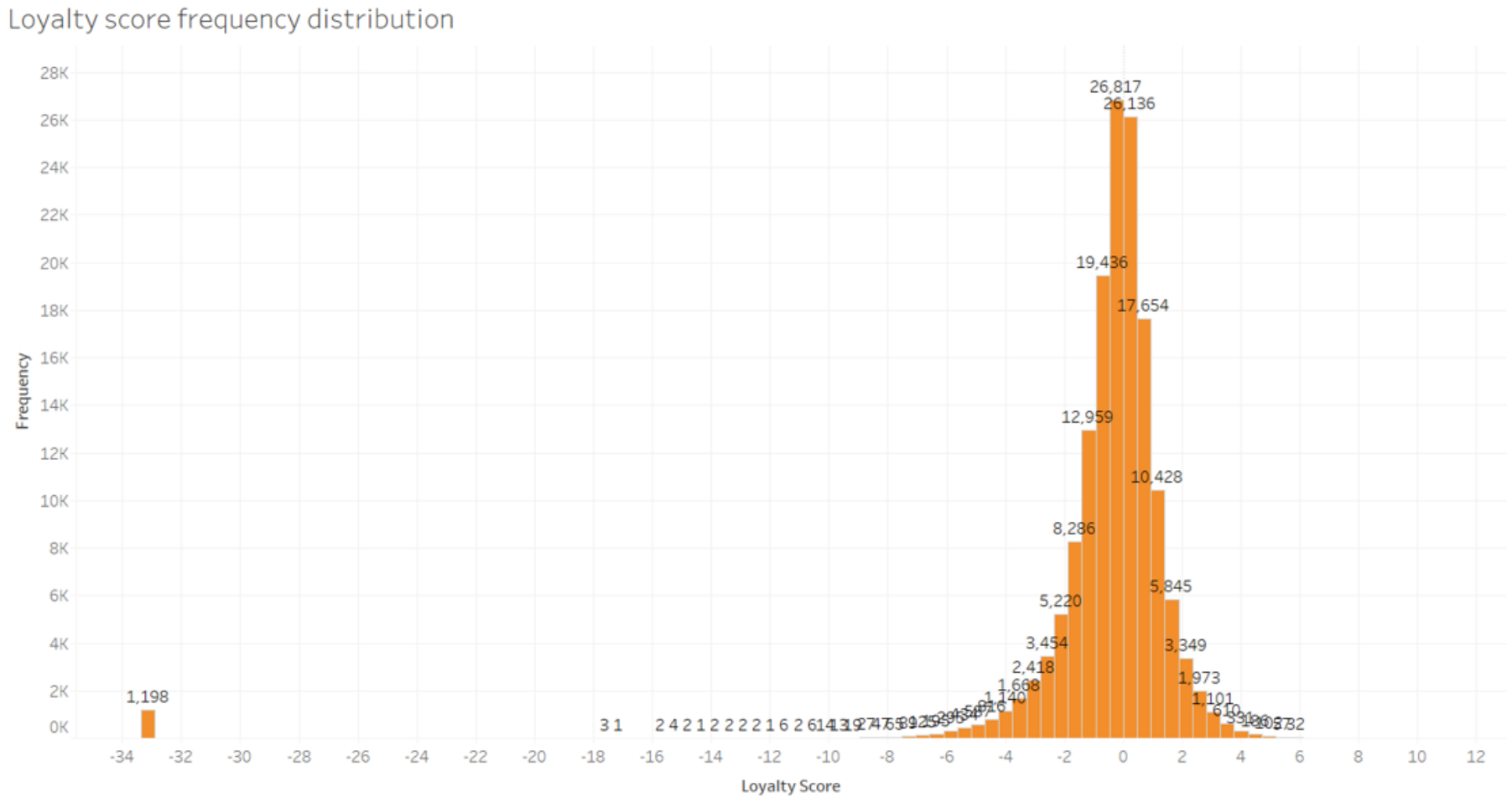
Aggregate data to one dataset, remove all missing value.

[New Features]

Date-based	▶ Purchase date, quarter, month, day, hour, week of year, weekday, weekend; ▶ Price (purchase amount / installments); ▶ Elapsed time: the difference between the date of a card's first active month and Feb 01, 2018.
Statistics	Count, sum, max, min, mean, std, difference between max and min, number of unique elements, etc.

Step 2

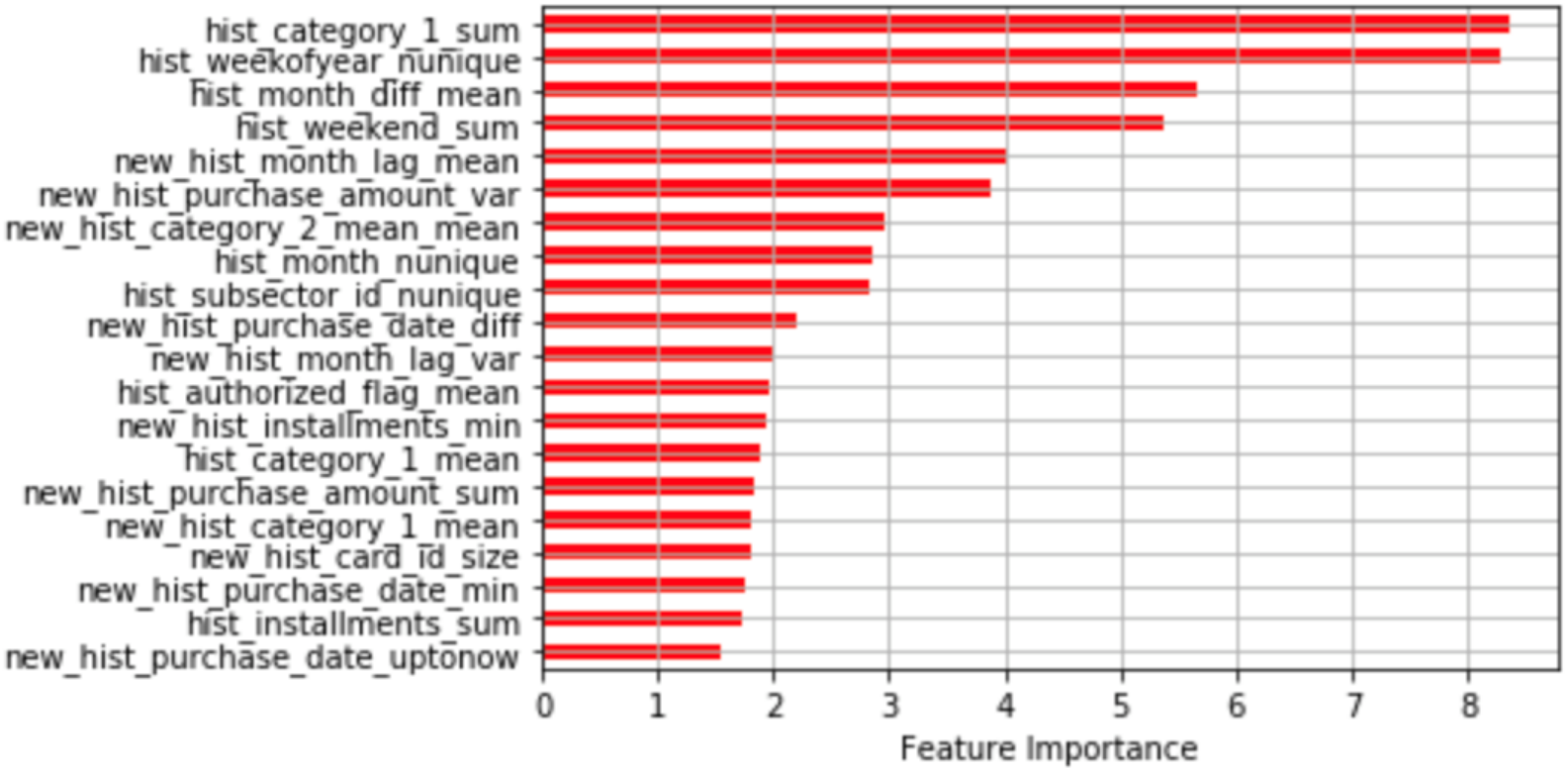
Drive the data (loyalty score distribution, correlation about features).



The plot of count of target for target (bin).

2.3 K-means Cluster

First build a XGBoost model(RMSE for test:3.7).
Then generate a Feature importance rank.

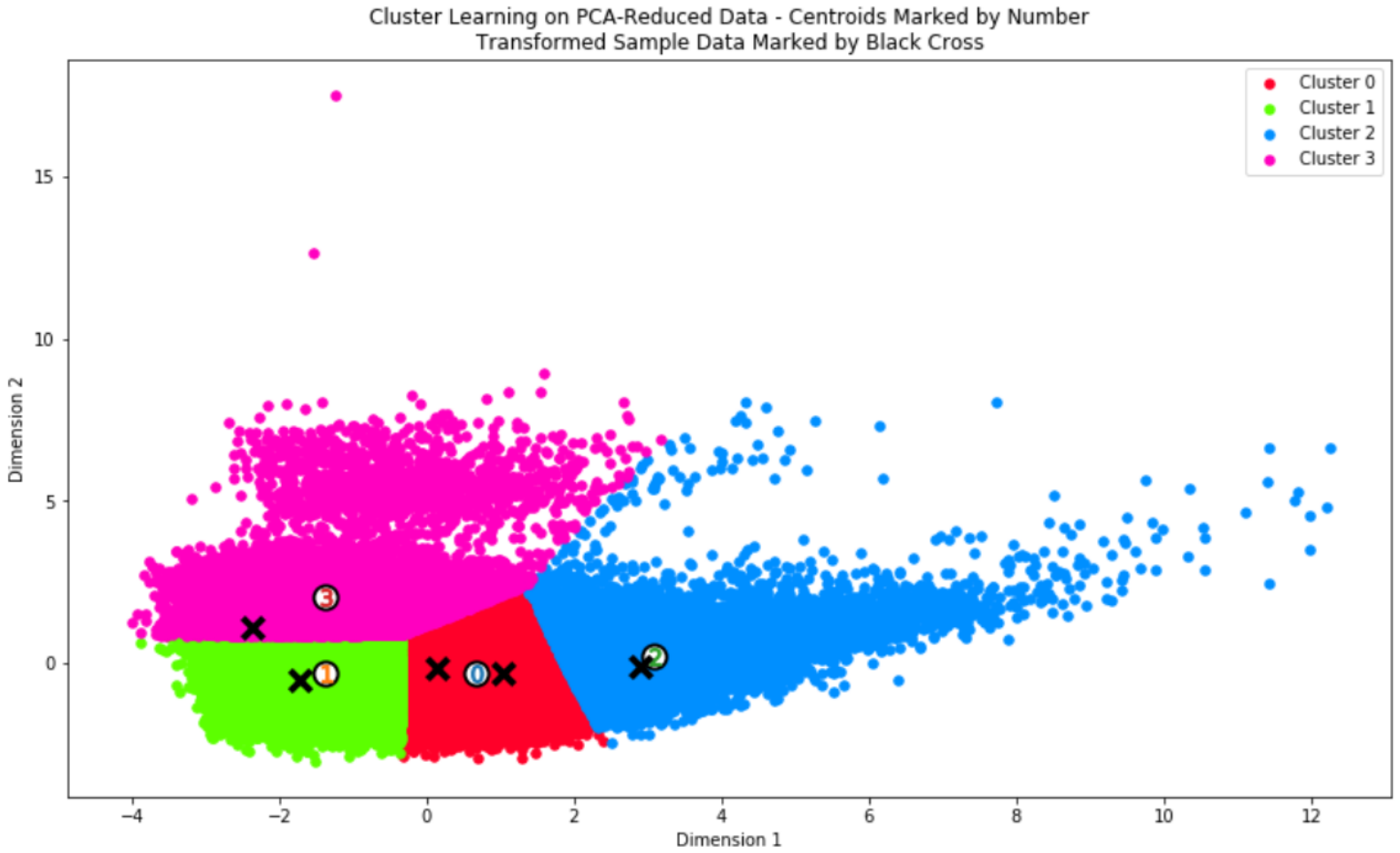


Use standard scaler and PCA to process the data.
Calculate the silhouette score to find the optimal *k* number.

```
# Loop through clusters
for n_clusters in range_n_clusters:
    # TODO: Apply your clustering algorithm of choice to the reduced data
    clusterer = KMeans(n_clusters=n_clusters).fit(reduced_data)

    # TODO: Predict the cluster for each data point
    preds = clusterer.predict(reduced_data)

    # TODO: Find the cluster centers
    centers = clusterer.cluster_centers_

    # TODO: Predict the cluster for each transformed sample data point
    sample_preds = clusterer.predict(pca_samples)

    # TODO: Calculate the mean silhouette coefficient for the number of clusters chosen
    score = silhouette_score(reduced_data, preds, metric='euclidean')
    print("For n_clusters = {}. The average silhouette_score is : {}".format(n_clusters, score))
```

```
For n_clusters = 3. The average silhouette_score is : 0.3575765101483685
For n_clusters = 4. The average silhouette_score is : 0.3861450509511866
For n_clusters = 5. The average silhouette_score is : 0.33938691788864134
For n_clusters = 6. The average silhouette_score is : 0.3512653104151818
```

Train the model(k = 4).
Analyze the characteristic for the different cluster.



2.4 Descriptive Statistics

hist_category_1_sum

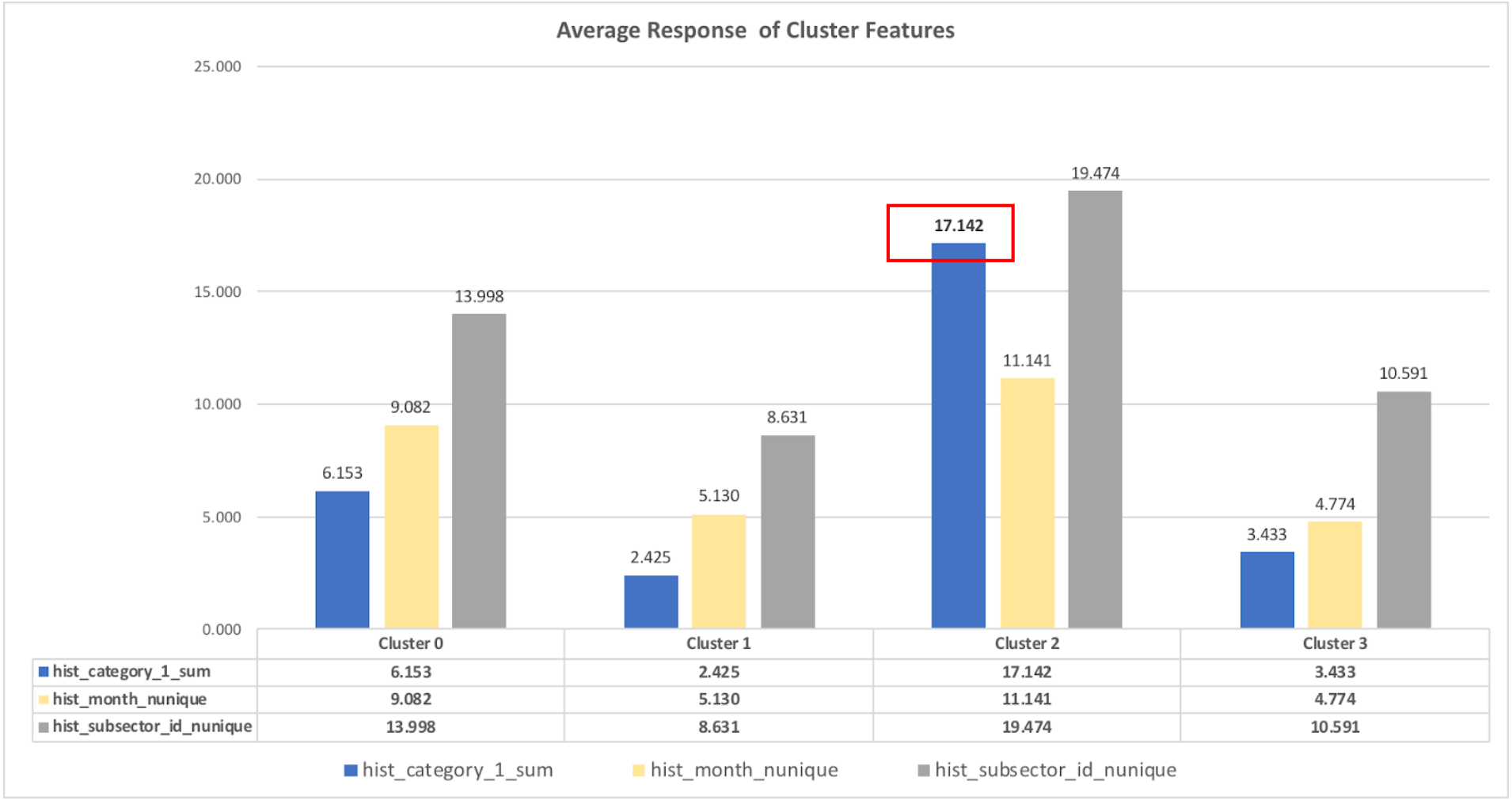
sum transactions of anonymized category 1

hist_month_nunique

the number of unique months in a year in which transaction occurs

hist_subsector_id_nunique

number of unique merchant category group



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
hist_category_1_sum	6.153	2.425	17.142	3.433
hist_month_nunique	9.082	5.130	11.141	4.774
hist_subsector_id_nunique	13.998	8.631	19.474	10.591

2.4 Descriptive Statistics

hist_category_1_sum

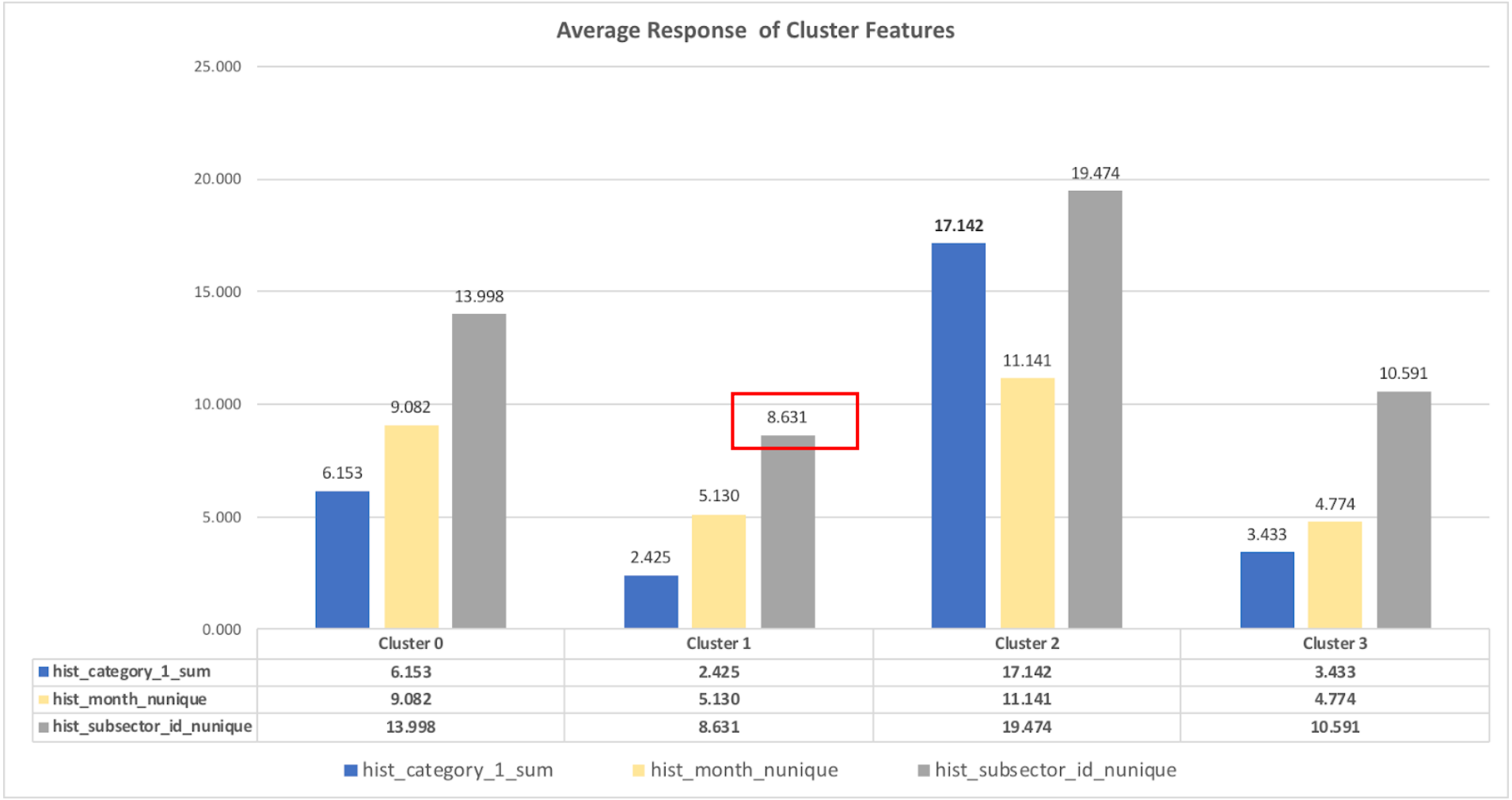
sum transactions of anonymized category 1

hist_month_nunique

the number of unique months in a year in which transaction occurs

hist_subsector_id_nunique

number of unique merchant category group



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
hist_category_1_sum	6.153	2.425	17.142	3.433
hist_month_nunique	9.082	5.130	11.141	4.774
hist_subsector_id_nunique	13.998	8.631	19.474	10.591

target

loyalty numerical score

new_hist_month_lag_mean

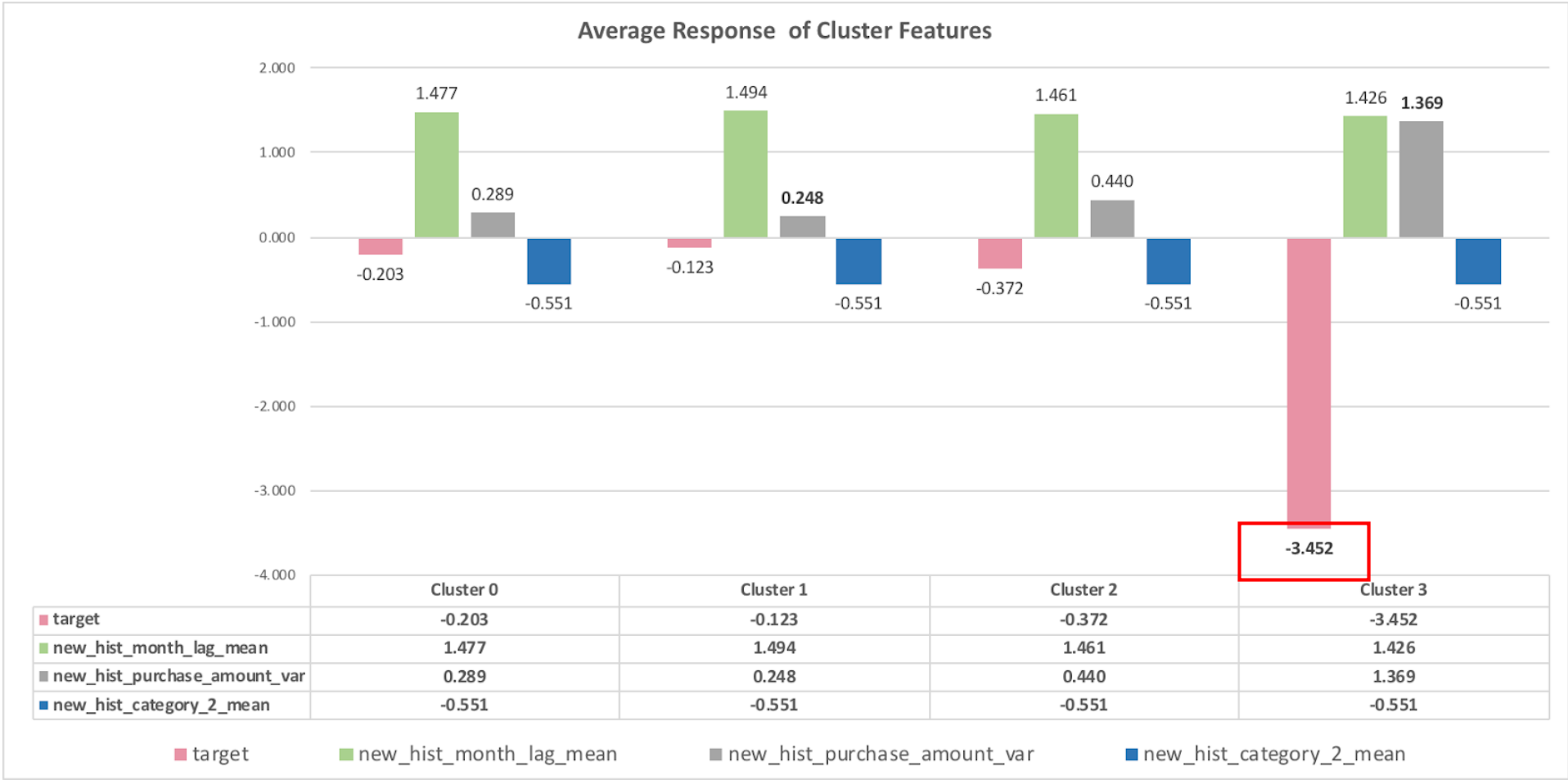
mean of the day before each customers' regular payment time of the installment （1 means the customer will pay the money for the installment before 1 day of the regular payment time)

new_hist_purchase_amount_var

variance of each customers' purchase amount for new merchants

new_hist_category_2_mean

mean of purchase amount of anonymized category 2 for new merchants(normalized)



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
target	-0.203	-0.123	-0.372	-3.452
new_hist_month_lag_mean	1.477	1.494	1.461	1.426
new_hist_purchase_amount_var	0.289	0.248	0.440	1.369
new_hist_category_2_mean	-0.551	-0.551	-0.551	-0.551

target

loyalty numerical score

new_hist_month_lag_mean

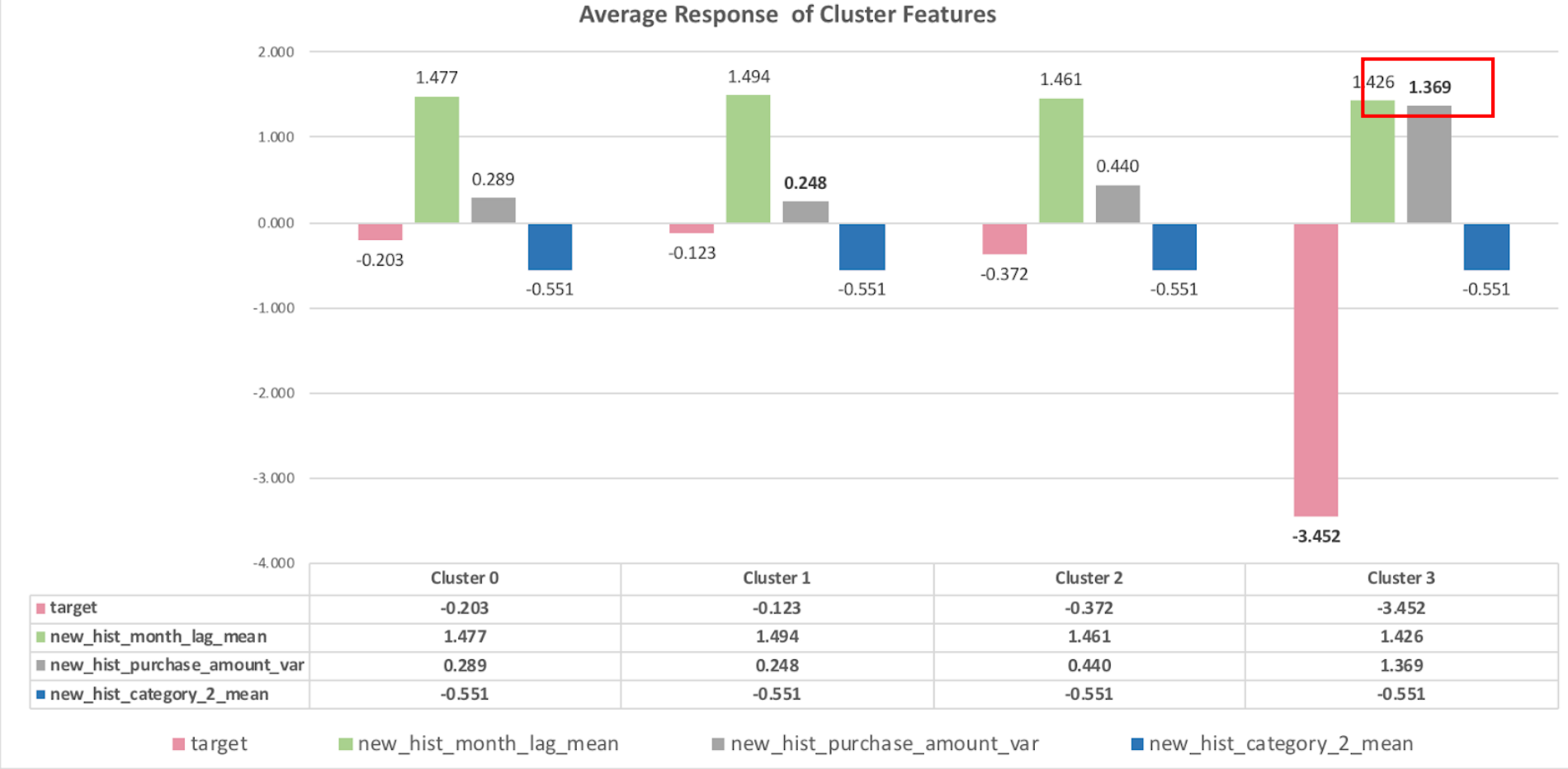
mean of the day before each customers' regular payment time of the installment （1 means the customer will pay the money for the installment before 1 day of the regular payment time)

new_hist_purchase_amount_var

variance of each customers' purchase amount for new merchants

new_hist_category_2_mean

mean of purchase amount of anonymized category 2 for new merchants(normalized)



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
target	-0.203	-0.123	-0.372	-3.452
new_hist_month_lag_mean	1.477	1.494	1.461	1.426
new_hist_purchase_amount_var	0.289	0.248	0.440	1.369
new_hist_category_2_mean	-0.551	-0.551	-0.551	-0.551

target

loyalty numerical score

new_hist_month_lag_mean

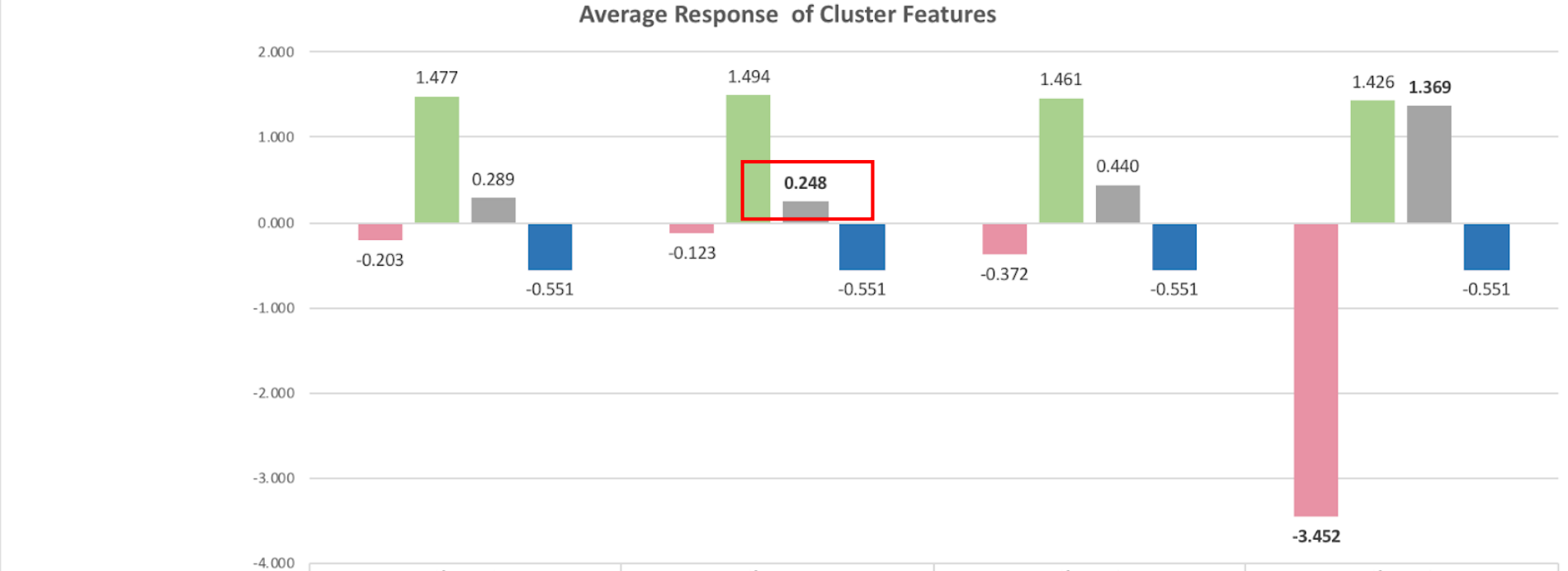
mean of the day before each customers' regular payment time of the installment （1 means the customer will pay the money for the installment before 1 day of the regular payment time)

new_hist_purchase_amount_var

variance of each customers' purchase amount for new merchants

new_hist_category_2_mean

mean of purchase amount of anonymized category 2 for new merchants(normalized)



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
target	-0.203	-0.123	-0.372	-3.452
new_hist_month_lag_mean	1.477	1.494	1.461	1.426
new_hist_purchase_amount_var	0.289	0.248	0.440	1.369
new_hist_category_2_mean	-0.551	-0.551	-0.551	-0.551

hist_weekofyear_nunique

the number of unique weeks in a year in which transaction occurs

hist_month_diff_mean

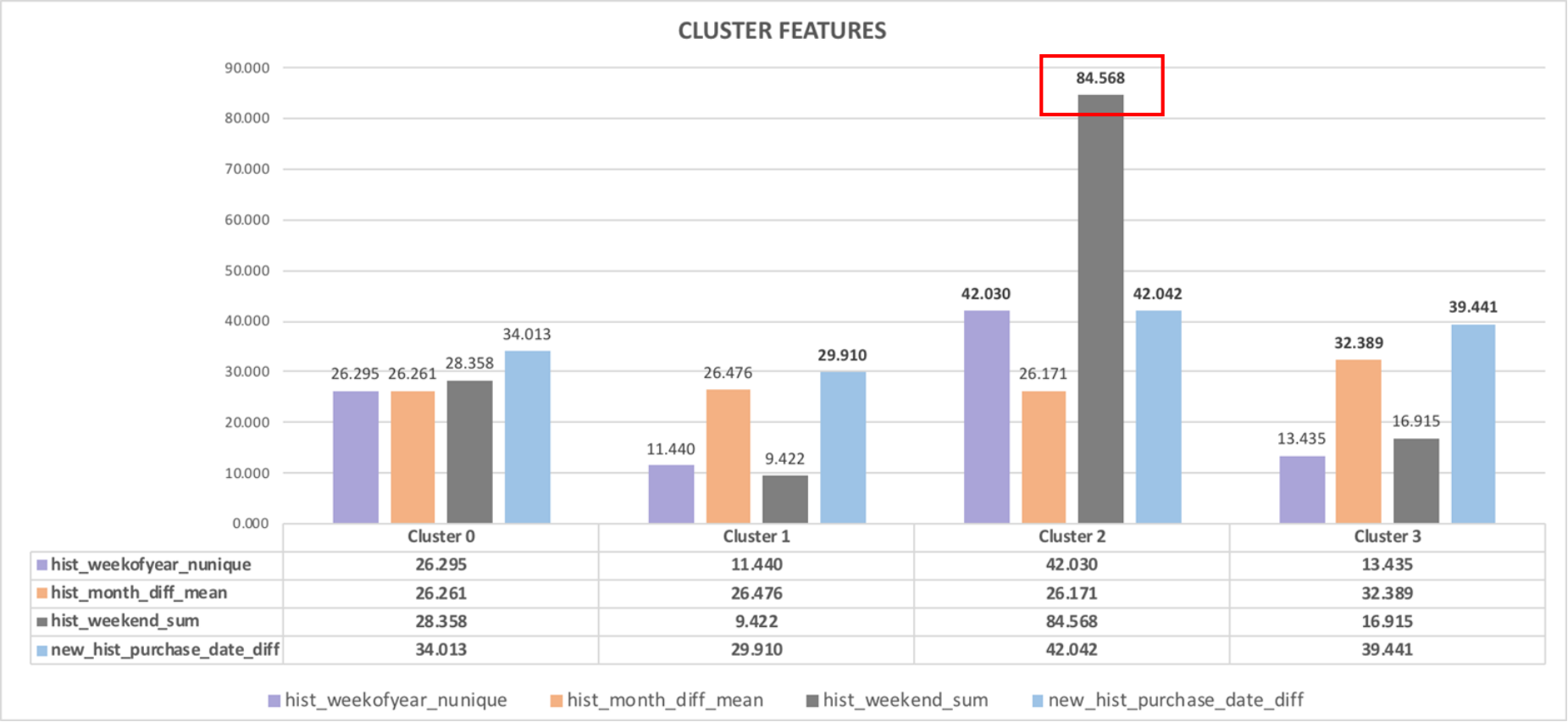
the average of the difference between the date of a card's first active month and the month Elo made merchants' recommendation to this card

hist_weekend_sum

total transactions happened in weekend time for each card id

new_hist_purchase_date_diff

the difference between two latest transactions' date of new merchant



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
hist_weekofyear_nunique	26.295	11.440	42.030	13.435
hist_month_diff_mean	26.261	26.476	26.171	32.389
hist_weekend_sum	28.358	9.422	84.568	16.915
new_hist_purchase_date_diff	34.013	29.910	42.042	39.441

hist_weekofyear_nunique

the number of unique weeks in a year in which transaction occurs

hist_month_diff_mean

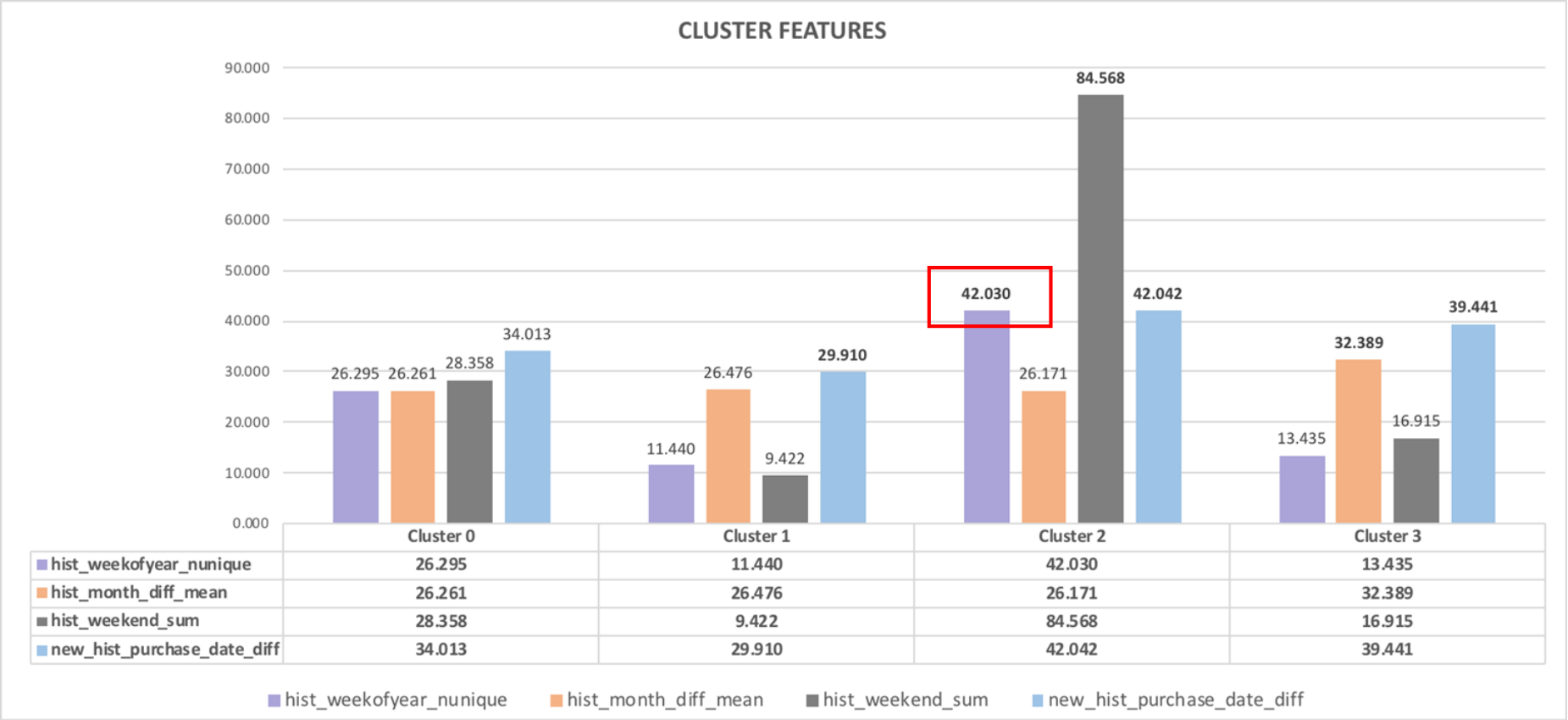
the average of the difference between the date of a card's first active month and the month Elo made merchants' recommendation to this card

hist_weekend_sum

total transactions happened in weekend time for each card id

new_hist_purchase_date_diff

the difference between two latest transactions' date of new merchant



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
hist_weekofyear_nunique	26.295	11.440	42.030	13.435
hist_month_diff_mean	26.261	26.476	26.171	32.389
hist_weekend_sum	28.358	9.422	84.568	16.915
new_hist_purchase_date_diff	34.013	29.910	42.042	39.441

hist_weekofyear_nunique

the number of unique weeks in a year in which transaction occurs

hist_month_diff_mean

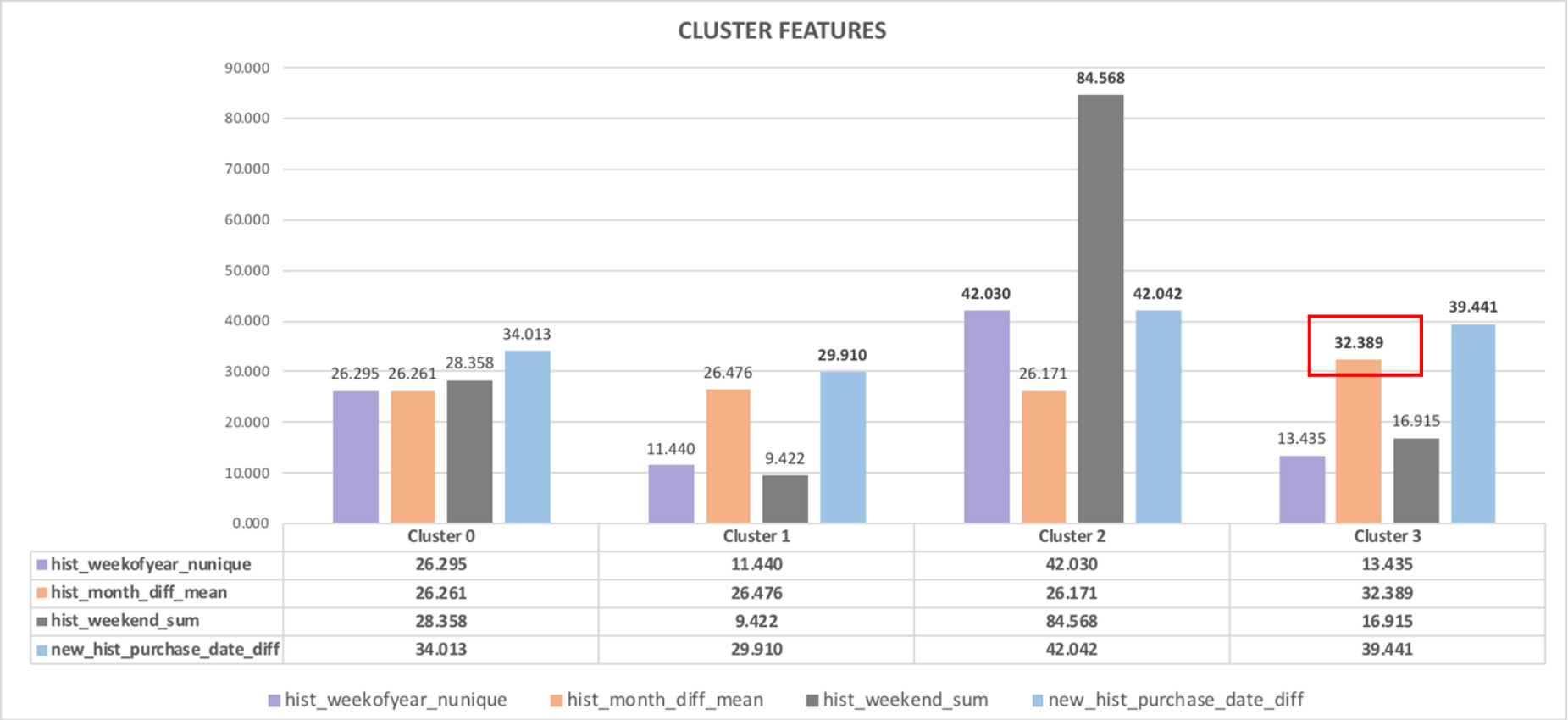
the average of the difference between the date of a card's first active month and the month Elo made merchants' recommendation to this card

hist_weekend_sum

total transactions happened in weekend time for each card id

new_hist_purchase_date_diff

the difference between two latest transactions' date of new merchant



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
hist_weekofyear_nunique	26.295	11.440	42.030	13.435
hist_month_diff_mean	26.261	26.476	26.171	32.389
hist_weekend_sum	28.358	9.422	84.568	16.915
new_hist_purchase_date_diff	34.013	29.910	42.042	39.441

hist_weekofyear_nunique

the number of unique weeks in a year in which transaction occurs

hist_month_diff_mean

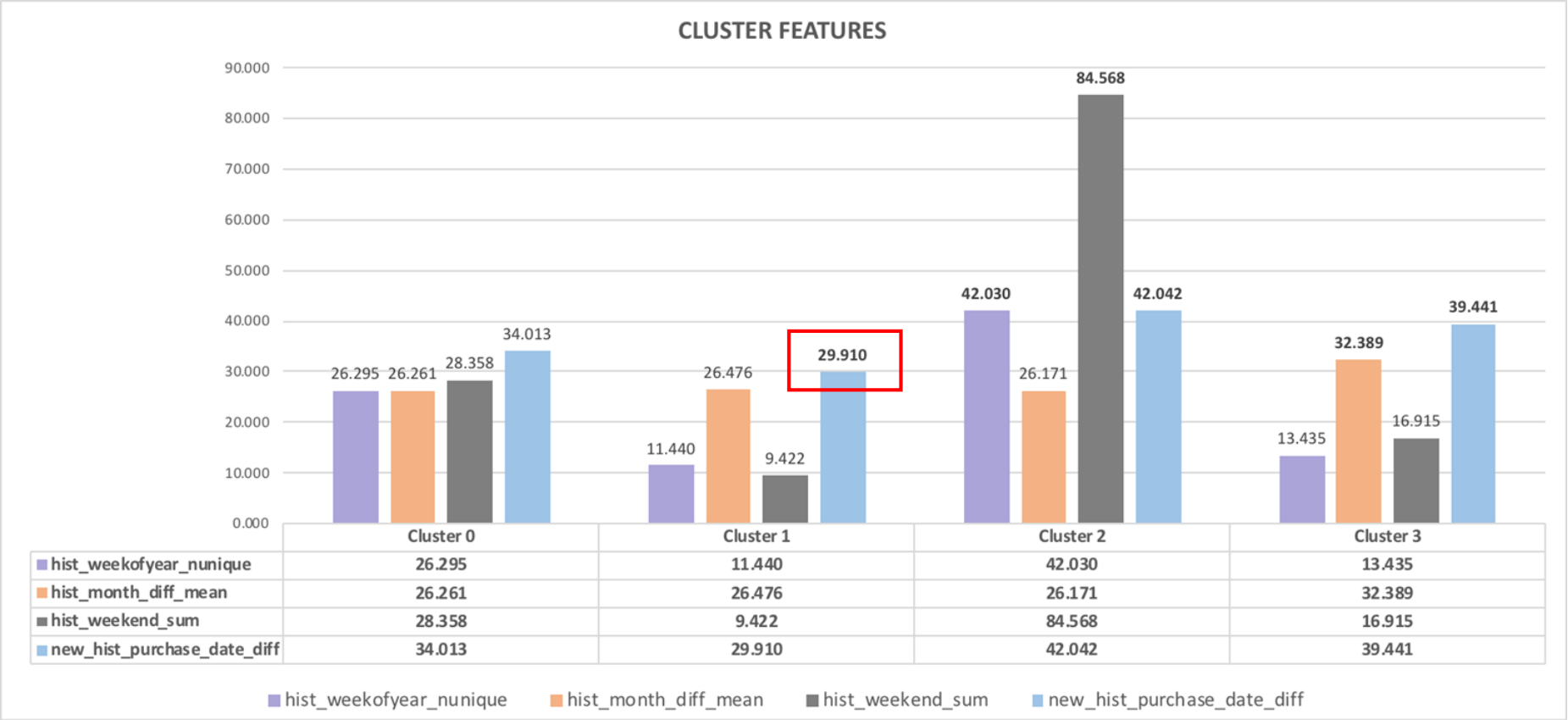
the average of the difference between the date of a card's first active month and the month Elo made merchants' recommendation to this card

hist_weekend_sum

total transactions happened in weekend time for each card id

new_hist_purchase_date_diff

the difference between two latest transactions' date of new merchant



	Cluster 0	Cluster 1	Cluster 2	Cluster 3
hist_weekofyear_nunique	26.295	11.440	42.030	13.435
hist_month_diff_mean	26.261	26.476	26.171	32.389
hist_weekend_sum	28.358	9.422	84.568	16.915
new_hist_purchase_date_diff	34.013	29.910	42.042	39.441

Part 03 Summary & Insights

Customers of Elo can be divided into four clusters based on their data performance

(Combining with STP Strategy)

	[Cluster 2]	[Cluster 3]	[Cluster 1]	[Cluster 0]
Importance Ranking	1st	2nd	3rd	Last
Features	▶ Have a strong preference for category 1 product; more categories ▶ Transactions happen in most weeks and months of year——high consumption frequency ▶ Be sensitive to Elo recommendation	▶ Have a big variation of purchasing amount; ▶ Relatively low consumption frequency new merchants. ▶ largest time interval between activating the card and the first purchase.	▶ Be sensitive to new merchants' recommendation; ▶ Have a special liking for several categories, dislike spending on weekend; ▶ Low consumption frequency	▶ Relatively high frequency consumption; ▶ Steady consumption.
Segmentation	Primary customers, high frequency, versatile purchase types and time, sensitive to ads.	High or low loyalty, big variance, but lowest consumption frequency.	Small variance, would like to try new merchants, but generally low consumption frequency.	Steady consumption, maintain the status quo.

Part 04 Recommendations

5.2 Recommendations

Marketing plans can be made differently in various clusters

Cluster 2 - Important Development Customers

Recommendations
Expand the recommended categories

- Increase the advertising amount; Provide advertising with diverse categories and high quality
- Recommend complementary products or premium alternatives of category 1

Cluster 3 - Important Retaining Customers

Recommendations
Raise consumption frequency, prolong customer life cycle

- Increase mail advertise from Elo to recall customers;
- Give additional coupons to attract them.

Cluster 1 - Promising Customer

Recommendations
Encourage consumption when the company has sufficient resources

- Launch incentive programs, such as reward points system, give reward when the customer has the first purchase of each day, or reach a certain amount of credit;
- Introduce hierarchical system, encourage customers to unlock more concessions by reaching certain amount of credits, reset every year;
- If the average age of consumers is relatively low, consider building a younger brand line or sponsor campus activities

Cluster 0 - Steady Customer

Recommendations
If the company's resource is limited, this group can be paid less attention

- Maintain the status

Part 05 References & Citations

References & Citations

- ***Analytics***

  https://github.com/ritchieng/machine-learning-nanodegree/tree/master/unsupervised_learning/customer_segments

  https://www.kaggle.com/c/elo-merchant-category-recommendation/overview

  https://xgboost.readthedocs.io/en/latest/python/python_api.html

- ***Research***

  https://towardsdatascience.com/clustering-algorithms-for-customer-segmentation-af637c6830ac

Tom'N Jerry
Xia Feng / Boxue Li / Peiqing Lian / Xinrong Lian / Ruoxuan Zhou

THANKS

Q&A