

# Business Value Analysis of Catering Shops from ‘Dian Ping’ Platform with Python Software

Wenshuo Zhang

BSc(Hons) Business Information Technology

## Introduction

In the age with information trend, product has not been the only focus of enterprise marketing and customers has played a more important role.

*Dianping.com* was established in April 2003, is one of China's largest consumer guide website of citizens life. It is now facing business crisis such as loss of merchants and declining competitiveness.

## Main Aim

Applying the *SRFMP* model and *K-Means* method to cluster and classify different merchants according to the data of the Dian Ping Website and analyse the characteristics of different merchant categories with personalized strategies.

## Materials and methods

- The creation process of customer's value is to attain a specific goal through customer activities. The way for suppliers to improve its competitive ability is to add more resources into the total pool of customers' resources. [1]
- *RFM* model is a customer value analysis method proposed by *Hughes* to recognize valuable customers from big data by three attributes including:  
R: Recency of the last purchase;  
F: Frequency of the purchases;  
M: Monetary value of the purchases. [2]
- A definition of *K-Means* Clustering algorithm is that the degree of similarity based on n objects is divided into K groups such that objects in the same group have higher similarities, while objects in different groups have lower. [3] [4]

## Implementation

### 1. Data Exploration

Exploratory analysis is the analysis of missing data, data rules and outliers. **Figure 1** shows the results of data exploration.

|             | null | max    | min  |
|-------------|------|--------|------|
| dist_id     | 0    | 4551   | 110  |
| avgPrice    | 0    | 521    | 0    |
| shopPower   | 0    | 5      | 0    |
| picTotal    | 0    | 9687   | 1    |
| popularity  | 0    | 188    | 0    |
| hits        | 0    | 511163 | 1198 |
| monthHits   | 0    | 9956   | 0    |
| voteTotal   | 0    | 6375   | 1    |
| wishTotal   | 0    | 7725   | 0    |
| branchTotal | 0    | 133    | 0    |

Figure 1 Data Exploration

### 2. Data Cleaning

The specific method of data cleaning is to discard the records whose average price is empty and discard the records whose popularity and recommended index are zero at the same time.

### 3. Data Reduction

According to *SRFMP* model of *DianPing.com* on the value of shops, five indicators related to them are selected as the indicators to identify the value of shops. **Figure 2** shows the comparison of the selected attributes.

| Attributes | Notes                      | Abbreviation |
|------------|----------------------------|--------------|
| avgPrice   | Based on customers' review | M            |
| avgScore   |                            | S            |
| picTotal   |                            | P            |
| hits       |                            | F            |
| voteTotal  |                            | R            |

Figure 2 Data Attributes Reduction

### 4. Data Transformation

Data transformation transforms data into appropriate formats to meet the needs of appropriate mining tasks and algorithms. **Figure 3** shows the data after standardization of standard deviation.

| Index | P         | R          | F          | M         | S         |
|-------|-----------|------------|------------|-----------|-----------|
| 1     | -0.355192 | -0.418619  | -0.345411  | 0.118922  | -0.491276 |
| 2     | 0.342288  | 0.510176   | 0.761828   | 0.446688  | 1.84952   |
| 3     | 0.0077739 | 0.009725   | 0.0338742  | 0.228817  | -0.983889 |
| 4     | 1.00991   | 3.36782    | 0.74285    | 1.16592   | 1.33983   |
| 5     | 1.38753   | 0.841819   | 3.44886    | 1.4954    | 1.66875   |
| 6     | -0.263986 | -0.0635111 | 0.498919   | -0.526472 | -1.73718  |
| 7     | -0.483335 | -0.48528   | -0.353395  | 0.37487   | -1.00933  |
| 8     | 1.47952   | 4.25471    | 7.77828    | 1.21185   | 2.89489   |
| 9     | 0.415495  | -0.0054767 | -0.0632278 | 0.778294  | 0.962284  |
| 10    | 0.007794  | 0.122791   | 0.113784   | 0.558983  | 0.528184  |

Figure 3 Data Transformation Table

### 5. Outlier Detection

The task of outlier detection is to find objects that are different from most other objects. **Figure 4** shows the distance error graph.

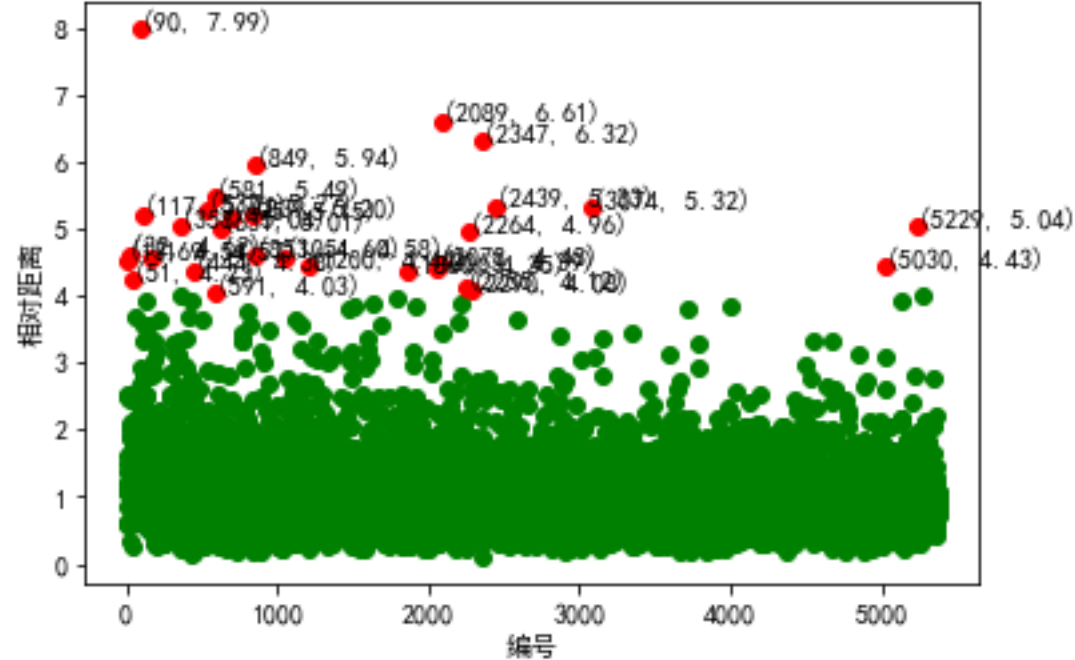


Figure 4 Results of discrete point test

### 6. K-Means Clustering

K-Means clustering algorithm is used to cluster shops' data from Dian Ping website into five categories.

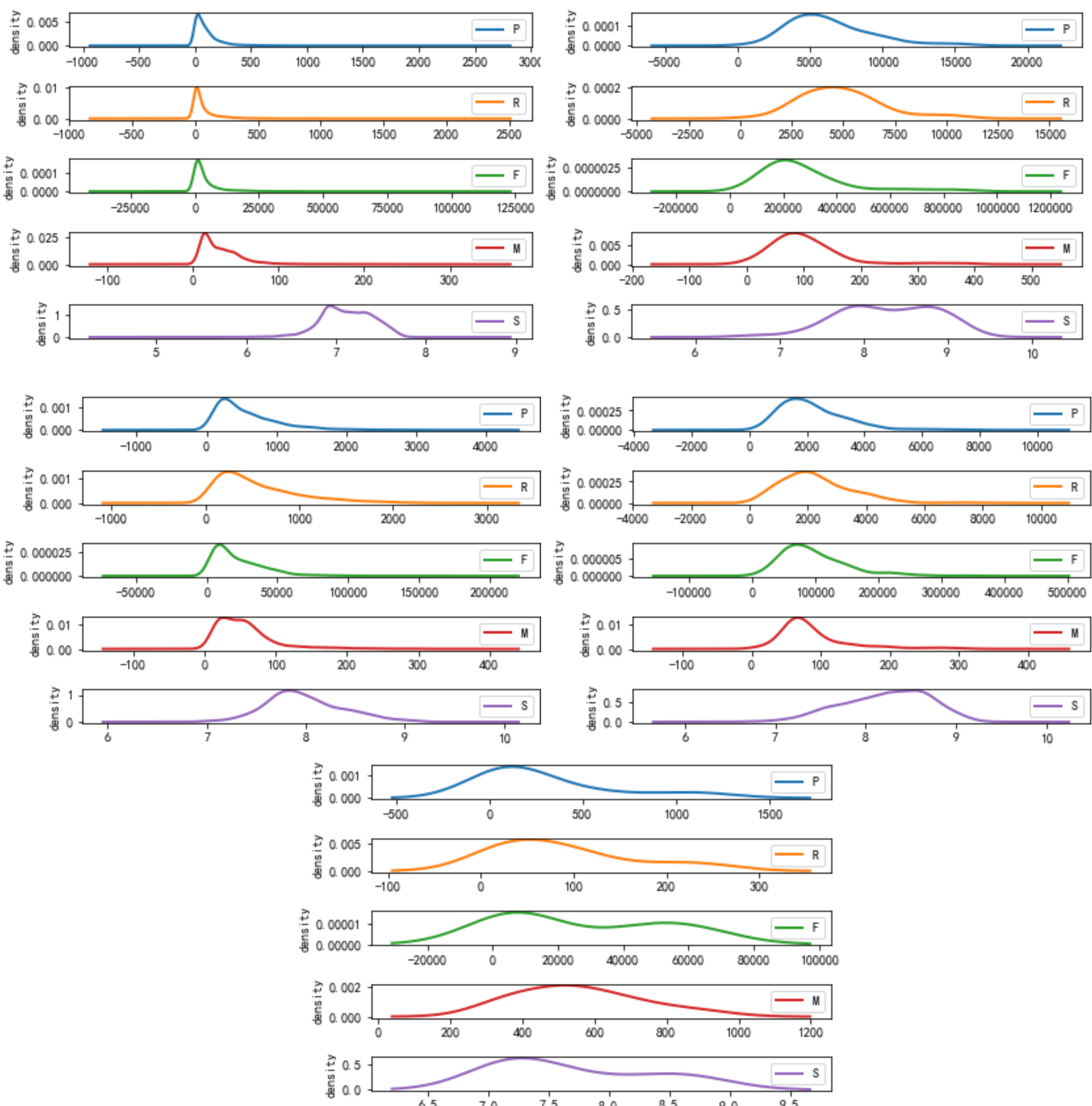


Figure 5 Probability Density Function Graph

## Results

According to the results of clustering, features analysis has been done as shown in **Figure 6**.

| Category | P | R | F | M | S |
|----------|---|---|---|---|---|
| 0        | P | R | F | M | S |
| 1        | P | R | F | M | S |
| 2        | P | R | F | M | S |
| 3        | P | R | F | M | S |
| 4        | P | R | F | M | S |

Tips: Red and bold represent maximum, red represents second maximum, normal font represents medium, green and bold represent second minimum and green represents minimum.

Figure 6 Shop features description table

Shop groups have five ranks with corresponding meaning.

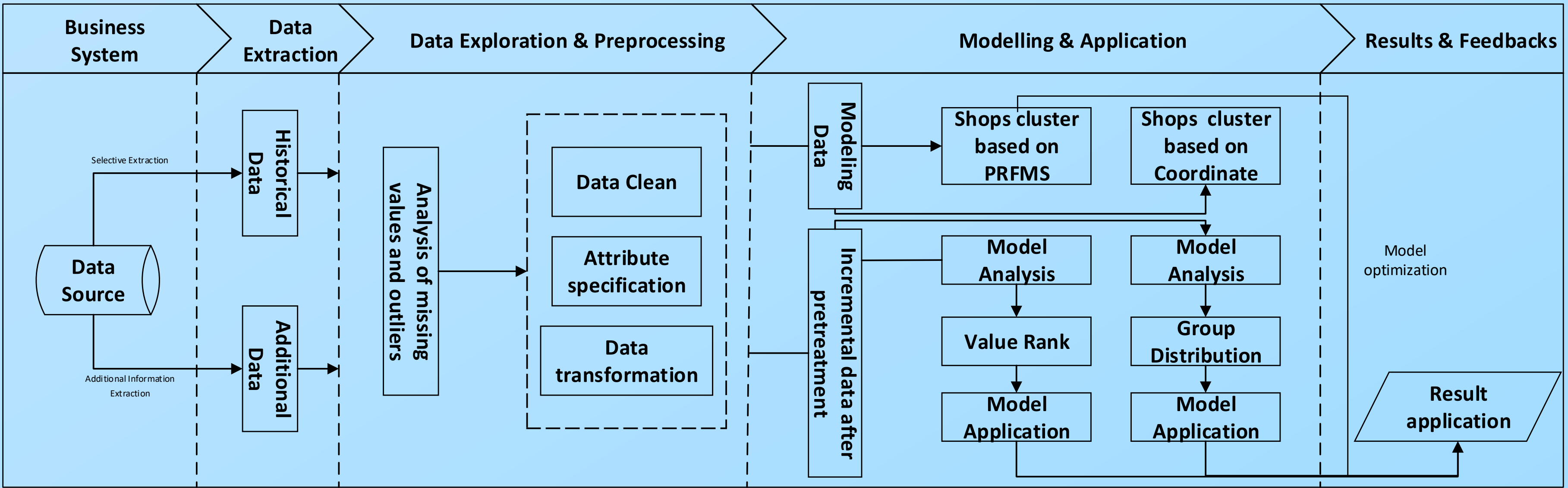
| Groups | Ranks | Meaning               |
|--------|-------|-----------------------|
| 1      | 1     | Important Maintenance |
| 3      | 2     | Important Development |
| 2      | 3     | General               |
| 4      | 4     | Important Retention   |
| 0      | 5     | Low Value             |

Figure 7 Value ranking of shop groups

## Conclusions

- The shop data from Dian ping platform can be classified into **five groups** with different features which can be taken as basis for platform to make **strategies** including member's preferential policy, credits exchange and cross selling.
- The analysis of the clustering results, individual shop **case study** and TripAdvisor **contrast experiment** have been done to verify the model and help the platform manage well.
- Further research is needed to conduct the analysis with **incremental data** at the same interval. **Surveys** or **interviews** can be collected to help the platform explore the reasons behind the data.

## Research Design



## Reference

- [1] J. Peppard, "Customer Relationship Management (CRM) in financial services," European Management Journal, vol. 18, (3), pp. 312-327, 2000.
- [2] C. Cheng and Y. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," Expert Systems with Applications, vol. 36, (3), pp. 4176-4184, 2009.
- [3] Fabian Pedregosa et al, "Scikit-learn: Machine Learning in Python". Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
- [4] 26. K. Kusriani, "Grouping of Retail Items by Using K-Means Clustering," Procedia Computer Science, vol. 72, pp. 495-502, 2015.



Prifysgol Cymru  
Y Drindod Dewi Sant  
University of Wales  
Trinity Saint David

