SMU
SINGAPORE MANAGEMENT UNIVERSITY

KNOW YOUR CUSTOMER

—Based on OCBC Credit Card Data

ISFS613-Analytics in Financial Service
Team 8
Bai Jiacheng | Jiang Nan| Lin Bei | Sun Qiansheng | Yao Yu

Contents

Business Scenario

Despite Singapore having the second highest credit card penetration rate in Asia Pacific, credit cards in circulation are increasingly declining. This is not only a trend from the macro-perspective of the nation, but also reflects a reduction in the average number of credit cards a Singaporean owns.

OCBC has rich credit card data and excellent reputation. However, competitive market campaigns by companies worldwide, means OCBC needs to know more about their customers, to ensure retaining their valuable customers. At the same time, OCBC analysts needs to balance the revenue and the risk.

The bottlenecks OCBC faced now are how to identify the valuable customers and what effective measures should be taken to reach them.

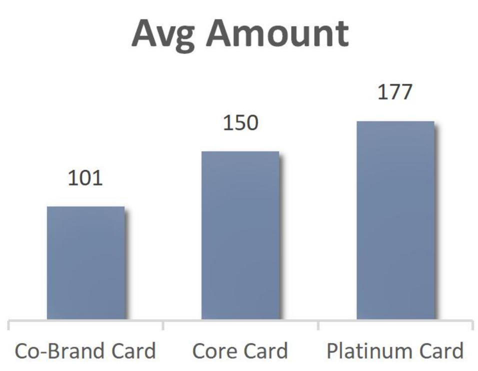
We want to find out the most valuable customers based on the 80/20 rule, which small percentage of the overall population contributes more than half the transaction amounts and who are the revolver customers to extract more revenue from in the long run.

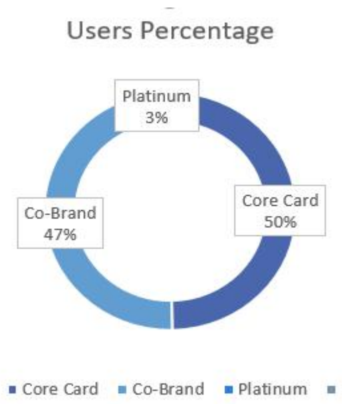
In order to retain these customers and generate more revenue for the bank, we want to find out the profiles of them and do marketing targeted promotions.

Data Preparation and Exploration

The data set includes two small sub-sets: The demographic data and the transaction data. The first one described the basic profile of each customers and the other is the two months transaction data of them. Before building the model, we start cleaning the data set, where we replace the null in district and remove the people with unknown credit limit. After that, we merge the demographic data with the transaction data.

Based on our findings, the overall distribution of people has several features, such as a higher proportion being male users and people using Co-Brand Card and Core Card predominantly as compared to Platinum card.





Analytical Solutions

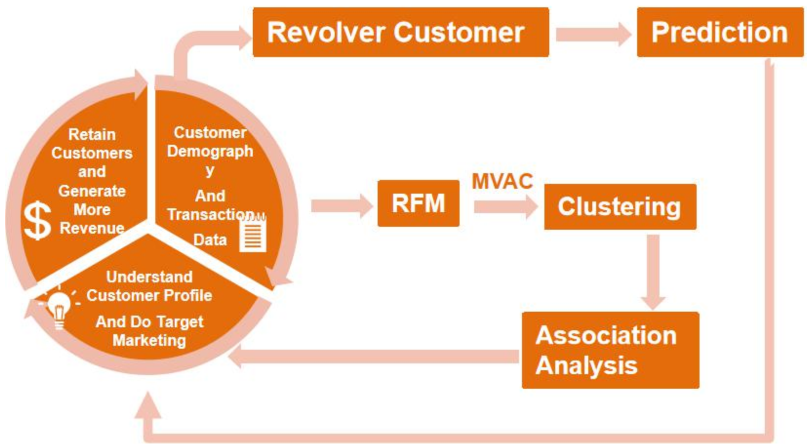


Figure 1 The Overall Solution

Our two target customers will be MVAC (median value and active customers) and revolvers. For extracting MVAC, we will use RFM model to generate the group with monetary score and frequency score higher than 3. After we simply explored some insights of MVAC, we use K-means clustering to separate them into several groups, based on their buying pattern and recommend the appropriate merchants that bank should do more promotions on. For revolver, we hope to predict accurately the new customers' flag. Since the pattern of hard revolver is different from the other two revolver types, we only predict the hard revolvers. We will use dummy variables like how many times the revolver buying from which merchant type, how many times they buy each month and the spending divided by credit limit ratio. After that, we use ensemble learning to train the model and do testing on new customers data set. Furthermore, we can know the profile of the revolvers which help us locate them.

Target customer

MVAC

RFM MODEL

We use the demographic data merged with the transaction data to generate the frequency, average amount and the recency of the latest transaction (set the last date of the data set as the target). Then, we use the SAS Enterprise Guide to run the result.

The result is shown below. We can observe the heat map where the group in red indicates a group of customers are recently frequent high value buyers.

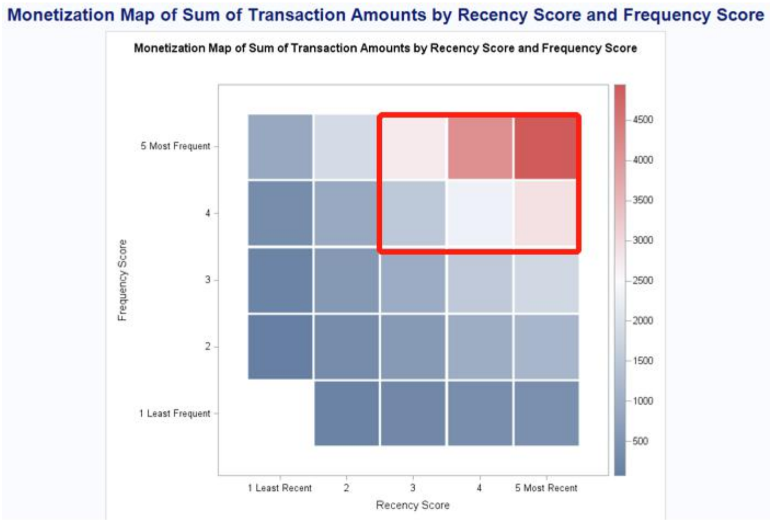


Figure 2 RFM Result

MVAC profile

After we extract the median value and active customers from the overall dataset, we find out that this group only occupies 15% of the overall customers (3163 customers), but they contribute more than 50% of the transactions (average $975 per customer per month).

Most of the customers are middle-age males. They spend their money mostly on life needs: supermarket, food and utility. For better understanding of what is their spending pattern and what is their buying sequence, we decided to do clustering in this group. However, we find out the real data with merchant type has bad performance in clustering. We decide to use the dummy variables to calculate each category in credit card type and merchant type of each MVAC. Hence, the variables we use in clustering are listed as below:

Variables
PartyID
AgeBand
CreditLimitBand
District
Gender
Cardtype=Co-Brand/Core Card/ Platinum
Category = merchant type
Avg Transaction Amount

Table 1 Clustering Variables Table

MVAC clusters

In clustering, we set the clustering method to Average, which can generate more equal size clusters, set the initialization method as First and set the maximum number of clusters into 10 as if we have too many clusters, it will be less representative.

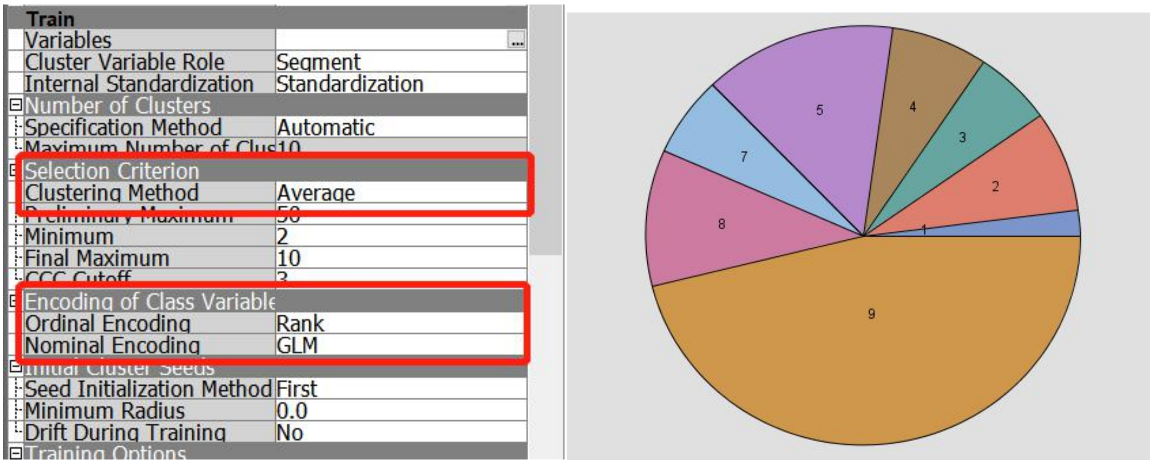


Figure 3 Clustering method and result

After we go through all the clusters, we summarise the buying pattern of each kind of cluster into one table:

Clusters	Number of people	Description
2	248	Gas Station Spending, Old Uncles
3	184	High Tech and healthy lifestyle guys
4	230	Enjoyer, young age, spend more on recreation
5	450	Co-Brand Card users, mainly spend on food and drug store.
7	198	More spend on clothes store and others.
8	329	Crazy Robinson Co-brand Card Users, high amount spending
9	1461	Life Needs Housewife, Core Card Users
Others	126	Hard Ware Users, New house owners

Table 2 Cluster Description

Compared to the existing solution which OCBC provided, they have offered **TWO Co-Brand Cards: Robinson and NTUC cards.** The other core card contains: 365 Card and FRANK card etc.

Therefore, we think the customer groups 2 and 4 will be the likely potential customers which OCBC is targeting at since these groups' buying pattern have not been served well, based on the existing solution.

That is why we used the association analysis and group profile to suggest what OCBC can do in subsequent steps.

Enjoyers

This group of people is different from the average customer profile: they are nearly younger than 30 years old. Their spending pattern is more on recreation, travelling and hotel booking. Average spending amount per person per month is more than 1200 dollars, which can contribute nearly 5 million dollars increase annually. Therefore, compared to others, this group of people are more likely to spend on high value products and frequently purchase then.

For better understanding their buying patter and do some cross-sell, we use MBA analysis on them. The results are shown below:

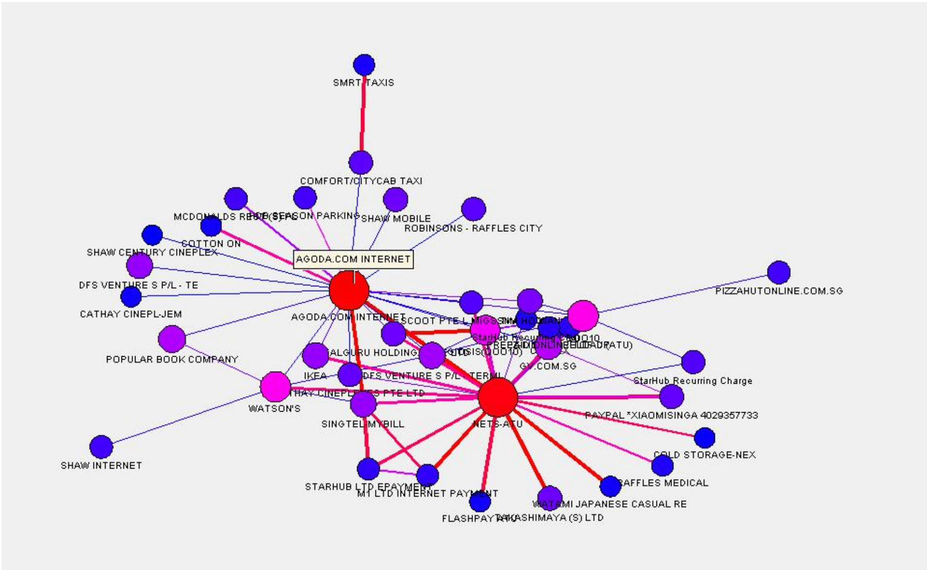


Figure 4 Link Graph of Enjoyer

We can find out that the famous merchants they always purchase is Agoda and NETS.
However, based on the research of OCBC credit card, there are several cards already covering the function of the travelling or buying tickets discounts. Therefore, we only do some recommendations here.

Car Owners

These group of people use 90% of their all transactions on gas station and food. They are almost 50-70 years old. They only occupied 20% of the MVAC, but their average spending amount is over $700 per customer per month, which can generate more than 2million dollars increase. These statistics result convince that these car owners are potential valuable customers.
The main merchants these customers will pay for are: ESSO, FAIRPRICE, CALTEX. Therefore, we want to know which kinds of merchant the car owners will head to after their cars fill up the gas.

MBA analysis

The method here we used is Association Analysis. We label the transaction of each customers by the oldest to newest transaction date and remove the transactions of food as 80% of the overall transactions are food. We want to search for some other opportunities. After that, we put the PartyID, merchant_name and label into Sas Enterprise Miner, set the Confidence to more than 50 and Maximum items as 4. The results are shown as follows:

General	
Node ID	Assoc
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Maximum Number of Iten	100000
Rules	...
Association	
Maximum Items	4
Minimum Confidence Leve	50
Support Type	Percent
Support Count	.
Support Percentage	5.0

Support(%)	Confidence(%)	Rule
1.673640167	100	CALTEX ==> ISETAN-SERANGOON (NEX) ==> CALTEX - SERANGOON AVE
1.673640167	33.33333333	QOO10 ==> CALTEX - TAMPINES

Figure 5 MBA setting and sample results

For the support of the chains, we think more than 10% is acceptable, as sometimes the link between two merchants may not be common but the relationship between them is quite fixed and stable. Therefore, our main conclusion is the Supermarket (NTUC Fairprice,Giant), Online retailer(Qoo10) and Department store(Isetan) will be the most likely shopping destinations for them.

Recommendations based MVAC Special Groups

For Enjoyers

The link graph (Figure 4) shows that the buying pattern of enjoyers is quite unique and simple. They will buy twice or more from the same merchants. Therefore, we highly encourage to add in some rules like providing more discounts if they purchase on the second or third times. Also, we conclude that there are four merchants OCBC can have opportunity to cooperate with in the future: Watson's, Qoo10, Agoda and NETS.

For Car Owners:

Premium Car Owners Card

This Core Card is specially designed for car owners. The applicants enjoy double credits for purchasing in gas stations (CALTEX, ESSO) and also enjoy one year free 'Autowise' Car Insurance. The reason why we promote this card is because the car owners who are past middle age, will more focus on their car quality. Therefore, they may have high probability to buy car related products.

Besides, we also include some supermarket, department store and online shopping promotions on the cards. For example:

Rewards	Merchant Name
Online Shopping 10% off Discount for first users	QOO10
$20 Voucher at ISETAN	ISETAN
Extra 5% Discount for Fair Price and Giant	FAIRPRICE GIANT

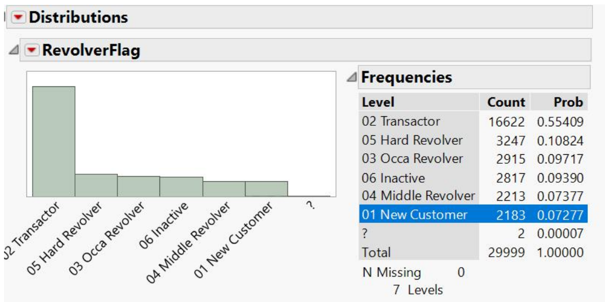
Table 3 Rewards for Premium Car Owner Card

Revolver Customer Prediction

Revolvers are consumers who carry credit card balances from one month to the next. Revolvers as a group are important sources of revenue since they pay interest on their balances. By contrast, transactors are consumers who pay their credit card balance in full and on time every month. Transactors do not carry a balance from month to month so they do not pay interest or late fees. Thus, we try to predict revolver labels of new customers and segment revolver groups to improve revenues of bank.

Exploratory Data Analysis

In the dataset, there is a RevolverFlag variable including categories such as transactor, occa revolver, middle revolver, hard revolver and new customer. We combine the occa, middle and hard revolver into REVOLVER category first and define customers with REVOLVER label as revolver customers.



Level	Count	Prob
02 Transactor	16622	0.55409
05 Hard Revolver	3247	0.10824
03 Occa Revolver	2915	0.09717
06 Inactive	2817	0.09390
04 Middle Revolver	2213	0.07377
01 New Customer	2183	0.07277
?	2	0.00007
Total	29999	1.00000

N Missing 0
7 Levels

Figure 6 Distribution of RevolverFlag feature

There are over 8 000 revolver customers accounting for 28% of total customers and about 120 000 transactions are made by revolver customers. Approximately one-third of the customers and transactions are related to revolver customers.

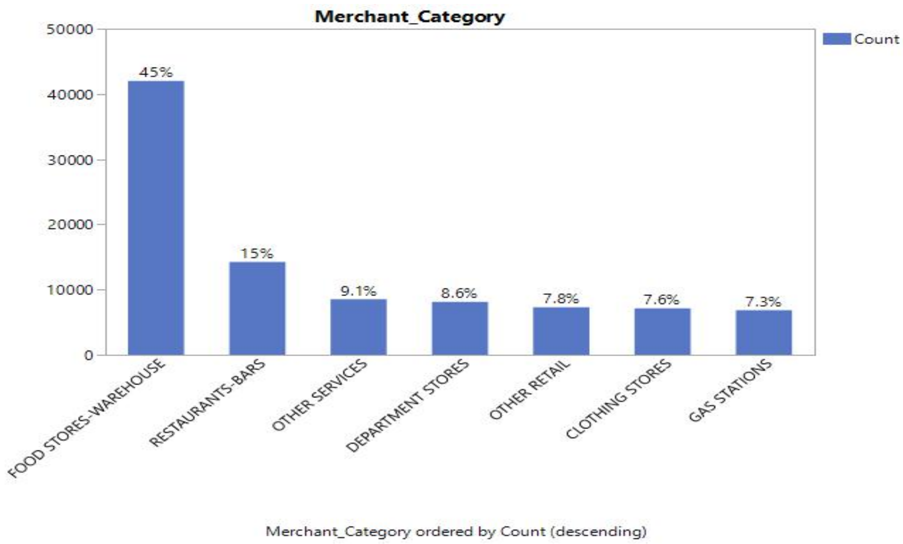


Figure 7 Merchant category distribution of revolver customers

From the charts, we can conclude that revolver customers spend mainly on daily expenses such as food, clothes and gas.

Most revolver customers have low credit limits, most of which are under 20 000 and the most active customers based on the most transaction frequencies and amounts, range from 30 to 60 years old

Predictive Models

We have learnt more about the characteristics of revolvers after EDA. Then we try to build predictive models to predict the revolver label. Since we need to do much data transformation, we choose Python as the tool.

Revolver Label Selection

Initially, we try to combine occa revolver, middle revolver and hard revolver into one REVOLVER category and use REVOLVER as our revolver label. However, the model performances are not pleasant since most of the models obtain accuracy under 70%.

Hence, we just take hard revolver as our revolver label and make it into a binary classification task where "1" means hard revolver. "0" means transactors.

Feature Engineering

We then conduct the following data transformation:

a. Merge tables: Merge Demographics and Transaction table and select rows with transactor and hard revolver labels.
b. Get dummy variables for Gender, CardType and Merchant_Category columns
c. Group by customer ID and aggregate by mean amount
d. Important features: Average daily spending amount of merchant categories and card types, Average daily spending amount for all transactions

The dataframe looks as follows, after feature engineering.

```
train_dt.head()
```

	ID	Age	Credit	GenderF	GenderM	RECREATION_Amount	UTILITIES_Amount	VEHICLES_Amount	FOOD STORES-WAREHOUSE_Amount	EDUCATION_Amount	...
0	2359312	45	65	1	0	0.0	0.0	0.0	0.0	0.0	...
1	3407976	45	5	0	1	0.0	0.0	0.0	0.0	0.0	...
2	2097356	35	5	0	1	0.0	0.0	0.0	0.0	0.0	...
3	2883793	55	5	1	0	0.0	0.0	0.0	0.0	0.0	...
4	40632550	35	15	1	0	0.0	0.0	0.0	0.0	0.0	...

5 rows × 41 columns

Figure 8 data sample

Training models

The models we have used are supporting vector machines, decision tree, random forest and AdaBoost. We split the training data and test data as 90:10. For hyperparameter tuning, we apply RandomizedSearchCV in most cases. We train and test the above models using Python Scikit-learn package and then measure the performance by confusion matrix and ROC curve.

Model Performance

The best model is random forest with the following hyperparameters.

```
pdict = {'criterion': ['gini', 'entropy'], 'bootstrap':[True, False], 'max_depth': list(range(10, 30, 2)),
         'min_samples_split': list(range(2, 10)), 'max_features': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6]}
rfm = get_compare(ensemble.RandomForestClassifier(n_estimators = 50, random_state = 2019), pdict)
```

```
best para:
{'min_samples_split': 2, 'max_features': 0.4, 'max_depth': 22, 'criterion': 'entropy', 'bootstrap': False}

        The evaluation report is:
        Accuracy = 0.8579439252336448, Precision = 0.3170731707317073
```

Figure 9 Best parameters of Random Forest Model

The accuracy rate is 85.8% and the area under ROC curve is 0.9965. The performance is quite good with random forest model for hard revolver prediction.

```
AUC: 0.9964610499197188
```

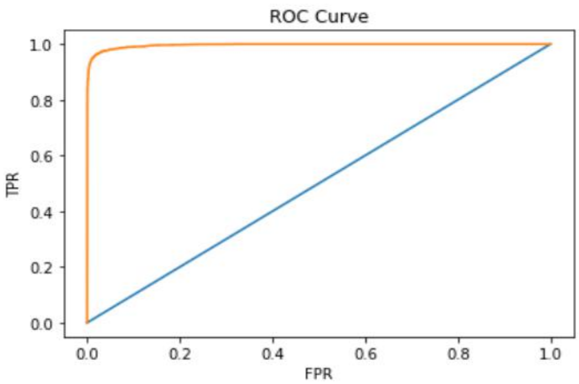


Figure 10 ROC curve of Random Forest model

Business recommendation

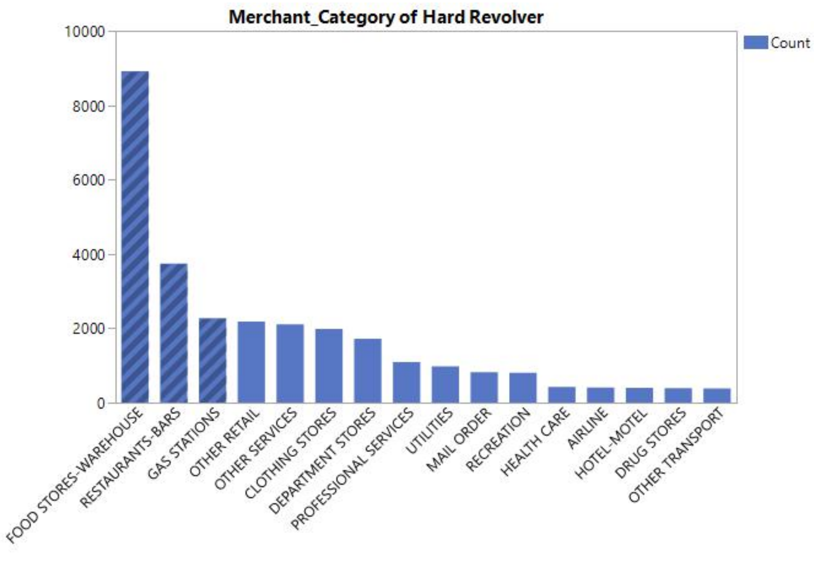


Figure 11 Merchant category distribution of hard revolver customers

From the graph, we can see that hard revolvers spend most in food, restaurant and gas station. Thus, we propose some recommendations based on their spending pattern.

Customer Acquisition

After we predict hard revolver labels of new customers using predictive modelling, the bank can do target marketing for these customers. The bank can set up exclusive promotions for the predicted hard revolvers in popular restaurants and gas stations so that the customers spend more by credit cards and then generate more revenues.

Customer Retention

For existing hard revolver customers, the bank may develop some marketing activities. For those who spend more than 15 times or 3000 dollars in a month, there will be cashback rewards. Besides, to dig deeper into the customers, further analysis such as clustering and market basket analysis can be done to conduct precision marketing.

Summary

According to the data given, we have carried out the analysis of customers and merchant category from two dimensions. We narrowed the scope of analysis and improved the depth of analysis by targeting MVAC and cluster analysis. Based on customer portfolio and consumption habits of different clusters, combined with merchant types and consumption rules, we give targeted suggestions. As for revolvers, which can bring huge revenue to the bank, we intend to carry out revolver label for new customers of the bank by building a prediction model so as to better target marketing and issue credit cards.

The analysis showed that OCBC credit card holders spent most of their money on lifestyle scenarios, including gas station, entertainment, travelling, and more. Therefore, when we gave suggestions, we considered the buying pattern which comes from Market Basket Analysis of sub-groups of our MVAC. Furthermore, we also conclude the characteristics of the hard revolvers which will help OCBC identify their valuable customer easily.

References

[1] OCBC Credit Card Introduction / https://www.ocbc.com/personal-banking/cards/index.html

[2] Bosnjak, Z. , and O. Grljevic . "Credit users segmentation for improved customer relationship management in banking." *IEEE International Symposium on Applied Computational Intelligence & Informatics* IEEE, 2011.

[3] Chen, Daqing , S. L. Sain , and K. Guo . "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining." *Journal of Database Marketing & Customer Strategy Management*19.3(2012):197-208.