BIA 650 Final Project:

Coupon Effectiveness Analysis

Team member:

Peng Dong

Rohan Jindal

Ling Ma

Vivek Martin

Tingting Jiang

Abstract: 2

Background 3

1.1History of Coupon 3

1.2 The importance of Coupon 4

1.3 Motivation about the project 5

Company Background: 7

Introduction to the dataset: 8

Customer clustering analysis: 10

Step 1: 11

Step 2: 12

Step 3: 13

Step 4: 13

Calculating Propensities: 14

The Assignment Model: 15

Recommendation: 16

Discriminant Analysis Model: 17

Result 1 Cluster Analysis: 18

Result 2 Propensity rate: 20

Result 3 Revenue, Propensity, Assignment Formula: 20

References: 22

# Abstract:

This paper is about analyzing the effectiveness of coupons and reallocating them to different customers in the most effective way. The goal of the project is to earn the most revenue for the hypothetical company. In this paper , several sections are introduced according to the project progress. In section 1, the background of the project is introduced to illustrate the motivation and importance of the project. In section 2, the information about the data will be introduced ,including the data source and the framework of the data. Section 3 will engage a clustering analysis which cluster the whole dataset into several parts for future work. In section 4, the main part of the project --- the reallocation of the coupons will be engaged. In section 5, conclusions will be drawn from the works and future works will be put out. In the project, several tools are used to accomplish it . MySQL is used to combine the datasets because the original datasets are relational datasets which were separate. Excel is used in the clustering analysis and the coupon reallocation optimization. SAS is used to calculate the propensity.

Keywords: Coupon, Effectiveness, Clustering Analysis, Optimization

# Background

## 1.1History of Coupon

Buying could be very important thing in a daily life and everyone store daily from little thing to massive with quality. Folks always attempt to buy high quality things in a low price. So it comes the “Everyday Coupon”.

It seems that coupon has been involved in our daily life ,but actually the term ”Coupon” derive from the promotion activity held by Coca-Cola in 1887. The first hand-written coupon was created by Atlanta businessman Asa Candler through a brainstorm. Candler’s invention transformed Coca-Cola from an insignificant tonic into a market-dominating drink. His hand-written tickets offered consumers a free glass of Coca-Cola, then priced at five cents. Between 1894 and 1913, an estimated one-in-nine Americans had received a free Coca-Cola, for a total of 8,500,000 free drinks. By 1895, Coca-Cola was being served in every state.

Since then , coupon has become an important way of promotion in business world. the high point of old-fashioned physical coupon was 1992, when 7.9 billion discounts were granted via coupons. Since then, the physical coupon has been in a decline. New-style electrical coupons thrive.

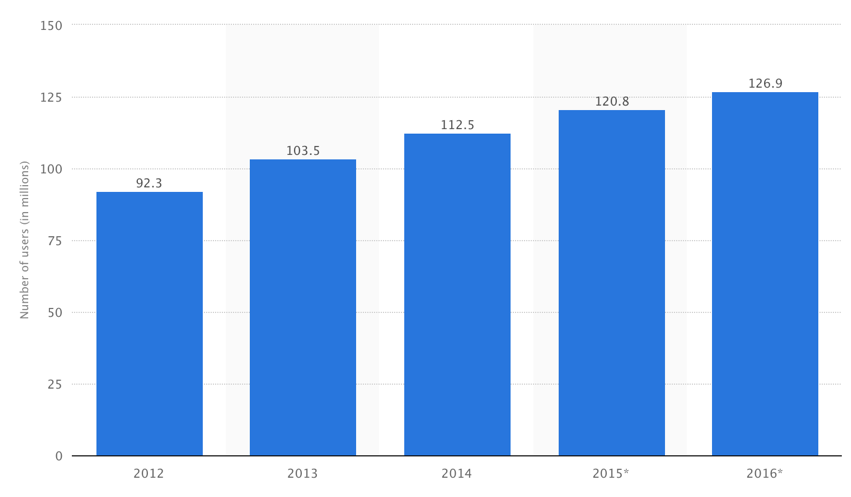


chart 1.1

From chart 1.1 we can see that the usage of digital coupon keep in increase in a high rate. That is because of several reasons: 1.For the consumers, they are becoming more likely to shop online. Digital coupons can become pretty convenient when the customers are shopping online. 2. Because of the development of the mobile devices, a growing number of customers are building a cellphone-based shopping habit, now ,there are a lot of apps allow customers to use coupon through their mobile phone which brings a lot of convenience. 3. For the business holders, digital coupons are more economical .The only cost for them is designing the coupons. They don’t have to pay money for the printing which reduces the risk.

In this project, we choose digital coupons as our object of study because , as discussed above, digital coupons are now in dominant. What’s more, the information about digital coupons are easy to access and store.

## 1.2 The importance of Coupon

Coupons have proven themselves to be highly effective sales tools for every conceivable size and type of business.

Because coupons "pull in the business" they have gained remarkable acceptance and popularity among astute marketing managers. A simple explanation for their acceptance by advertisers is their overwhelming acceptance and use by the consuming public. In fact, Advertising Age (the Bible of the advertising industry) reports that 87% of all shoppers use coupons.

Another independent marketing research firm, the A.C. Nielson Co., reveals that 95% of all shoppers like coupons. And 60% actively look for coupons.

A recent article in the Wall Street Journal entitled, "In a Pinch, Snip.", states that coupon use rises, as the economy in any given area slides. 54% of shoppers surveyed said they had already stepped up use of coupons, and even more are expected to do so.

It's very easy to see why coupon advertising is sweeping the country. Regular use of good couponing strategy will provide a steady stream of new customers and high quality sales leads.

Here are the reasons , why coupon is important

A. Coupons have the effect of expanding or increasing your market area. We know that consumers will travel far to redeem a valuable coupon.

B. Coupons will entice new customers that have been shopping at your competitor. It's a proven fact that consumers will break routine shopping patterns to take advantage of a good coupon offer.

C. Coupons attract new residents when they are actively in the market for products and services.

D. Coupons will re-activate old customers. Those customers that have been lured away by your competitor will start buying from you again when you give them a good reason to do so.

E. Coupon advertising provides the opportunity for additional profits through sale of related items. (Business owners often forget this.) When you offer a special "deal" on a coupon to invite a customer to do business with you, you have to remember that this same customer will probably end up buying additional items that carry a full profit margin.

In addition, you also are being given the opportunity to "sell-up" to a more profitable product or service. You would not have had this opportunity had it not been for the coupon getting the customer through the door in the first place.

F. Coupons build store traffic which results in additional impulse purchases.

G. Coupons are measurable and accountable. Don't overlook that couponing is the most measurable and accountable form of promotion. It's simply a matter of counting the number of coupons redeemed to judge the effectiveness of the offer. Wise use of this consumer feedback will guide you in creating future offers that improve your results.

From the information above , it seems that coupons is an elixir. The fact is not. That’s why we study on the coupons effectiveness.

## 1.3 Motivation about the project

As digital coupons play such an important role in todays’ business world, some companies started to run business based on it. The business mode for them is to allocate the coupons from different business holders to customers. There are several ways for these company distributing coupons, they can put the digital coupon on their webpage , send email or text message to customers’ cellphone or email address. Here we will take Groupon as an example to show how these kind of company making money from their business.

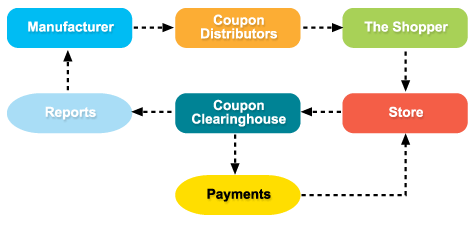


Chart 1.2

Chart 1.2 show how companies like Groupon make money from their business. The manufacturers give the companies coupons, and let them distribute it, then the companies distribute the coupons in their own way. After that , when the shopper bought the goods( or service ),the companies can get a percentage cut of the revenue from the business holders.

In the last several years, companies like Groupon grows rapidly as a new style of advertising companies. Based on huge use groups , they can easily persuade the business holders to give them unlimited numbers of coupons. Business holders were likely to do business with these kind of companies because it seemed that it is a more effective and risk-free way to do advertisement. During that period of time, a lot of “Groupon” thrive from all around the world. But as the business holders became familiar with this new business mod, they found that: Coupon is actually a two-edged sword. It has some advantages , as well as disadvantages.

1. First of all , it cut off the profit.

In average, the profit margin for restaurants and tourism companies is around 9% ,for spa is around 15%. The fact is that , if the business holders can not set proper discount rates, they are definitely losing money. Here I can take one case from my experience: A restaurant holder went to a company like “Groupon” and give them unlimited numbers of coupons. This coupon was a kind of voucher, which means customers could buy this voucher in 55% discount. The result was :With every coupon sold out, the restaurant hold lost about $15. Three months later , this restaurant closed down.

1. Coupons may hurt brand image.

If a customer always receives coupons in a high discount rate , chances are he will think the business holder must has some trouble in running the business and have to use coupons to save it. In long term , in the market ,this company will become a “cheap” one and no one want to go for it.

1. Unlimited coupons may lead to discount – addiction.

Discount – addiction means the customers are addicted to the discount price. If the price go back to the original one ,they won’t buy it any more .In this way , the coupon is totally useless: On one hand, it doesn’t bring new customers. On the other hand , if some of the frequent customers become addicted to the discount, the fact is this business is losing customers.

Nowadays, more and more business holders are realizing the disadvantages of coupons, they don't want to distribute the coupons in an unlimited way. They are not that into the coupons promotion as before any more. We can see it from the change of the cut off price of the “Groupon”-style companies. In 2010, the average cut off rate of “Meituan”(A famous Chinese company similar to Groupon) was around 20%. Now it becomes 5%. The reason is , if the cut off rate remains that high , business holders won’t be likely to do business with them.

Now , if a coupon allocation company go for a business holder, chances are that the business holder is willing to give the company a limited number of coupons to distribute.

In this project , we assume that we are a company like that, and our goal is to allocate the limited number of coupons to the customers in the most effective way, in order to get the biggest profit.

# Company Background:

Recruit Ponpare is Japan's leading joint coupon site, offering huge discounts on everything from hot yoga, to gourmet sushi, to a summer concert bonanza. Ponpare offers valuable coupons to help people find new experiences and discover new opportunities for small luxuries at surprising bargain prices. Ponpare's coupons open doors for customers they've only dreamed of stepping through. They can learn difficult to acquire skills, go on unheard of adventures, and dine like the stars.

Investing in a new experience is not cheap. We fear wasting our time and money on a product or service that we may not enjoy or fully understand. Ponpare takes the high price out of this equation, making it easier for you to take the leap towards your first sky-dive or diamond engagement ring. For example, discount coupons are offered for popular hotels and special products from famous restaurants. These coupons are only valid if a predetermined minimum number of people sign up for the deal within a certain amount of time. The special prices are possible due to discounts that can be offered based on the high volume of business.



Using past purchase and browsing behaviour, this competition asks you to predict which coupons a customer will buy in a given period of time. The resulting models will be used to improve Ponpare's recommendation system, so they can make sure their customers don't miss out on their next favourite thing.

# Introduction to the dataset:

We were provided with a year of transactional data for 22,873 users on the site ponpare.jp. The training set spans the dates 2011-07-01 to 2012-06-23. The test set spans the week after the end of the training set, 2012-06-24 to 2012-06-30. The goal of the competition is to recommend a ranked list of coupons for each user in the dataset (found in user\_list.csv). Your predictions are scored against the actual coupon purchases, made during the test set week, of the 310 possible test set coupons.



The datasets in use were:

**user\_list.csv** - the master list of users in the dataset

**coupon\_list\_test.csv** - the master list of coupons which are considered part of the test set. Your competition predictions should be sourced only from these 310 coupons. You will not receive credit for predicting training set coupons that were purchased during the test set period.

**coupon\_visit\_train.csv** - the viewing log of users browsing coupons during the training set time period. You are not provided this table for the test set period.

**coupon\_detail\_train.csv** - the purchase log of users buying coupons during the training set time period. You are not provided this table for the test set period.

Issues with the dataset:

1.The dataset was initially in Japanese.

The dataset we got from Kaggle was partially in Japanese and so we had to run various codes to translate and standardize the data to get an English dataset.

2. The dataset is too large. The dataset had 22,873 users and 310 coupons.

1. Too many instances for Solver. Solver was unable to process such a large amount of data and hence we had to take a small subset of the entire dataset in order to successfully allocate the coupons to some of the users based on their preferences.
2. Difficult to calculate propensity for the customer. To calculate propensity, we needed to look at the past purchase and viewing history of the consumer. All the data was in different datasets and it was difficult to manage all the datasets to get one dataset with all the information required to calculate the needed propensity.
3. Many-to-many relationship. The above point can be further explained by the presence of the relationship between the different dataset. Each customer had viewed or purchased several coupons in the past. Also, each coupon could have been used by a number of customers. This resulted in a complex relationship between the various datasets. To add to that, all the data related to consumer behaviour was in different datasets which made it difficult to merge all the data together to perform the necessary analysis.

3.Lack of demographic information.

In the data that was provided, there was no information regarding aspects of the customer like the education level, income and marital status. These details are very important to accurately predict the buyer behaviour. The only data we were provided was the age and the sex of the customer. It is very difficult to predict buyer behaviour based on just age and sex of the customer.

This can be further explained by the following examples:

* A person who was recently been engaged or married might be more likely to look at and purchase coupons for vacations and hotel reservations.
* Similarly, a person that is under stress due to work or studies might be more prone to look at spa treatments or leisure activities.

# Customer clustering analysis:

The motivation to do customer clustering analysis before coupon allocation optimization is that marketers often want to group objects into clusters of similar objects.

Cluster analysis is used in a variety of applications. Usually it can be used to identify consumer segments, or competitive sets of products, or groups of assets whose prices co-move, or for geo-demographic segmentation. It is often necessary to split our data into segments and perform any subsequent analysis within each segment in order to develop (potentially more refined) segment-specific insights. This may be the case even if there are no intuitively natural segments in our data.

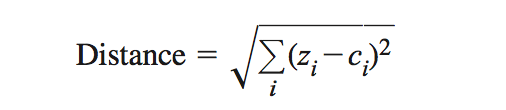
For example, identifying similar customers can help a company identify market segments, in this case, allocate coupons more efficiently and accurately to different customer segments.

Our assumption is that, if we can divide all the customers into proper groups, then we can extract common features of those customer groups and then based on those customer group features, we can allocate different coupon combinations to various customer groups, who share similar purchasing behavior and hobbits.

The objective of customer clustering is to divide customers into four groups based on their purchasing behavior and geographical attributes.

To identify segments in the data we can use statistical techniques broadly called clustering techniques. Based on how we define “similarities” and “differences” between data observations, which can also be defined mathematically using distance metrics, one can find different segmentation solutions. A key ingredient of clustering and segmentation is exactly the definition of these distance metrics (between observations), which need to be defined creatively based on contextual knowledge and not only using “black box” mathematical equations and techniques.

Methodology of our customer clustering analysis is that we will define four centers and calculate the distances to each center. The next step is to see how far each customer is from each of the cluster centers. Let Zi be standardized attribute i for a typical customer, and let Ci be standardized attribute i for a typical cluster center. You can measure the distance from this customer to this cluster center with the usual Euclidean distance formula where the sum is the over all attributes.



The process for our customer clustering analysis is as below:

## Step 1:

Data consolidation:

Since the data for this website is in a relational database, we need to query different tables to get information we want of each customer. And we need to make sure that all the data is numerical in our analysis. This is mainly because that that one needs to define Euclidean distances between observations, which are defined only with metric data. However, one could potentially define distances also for non-metric data.

And then we decide whether or not to scale or standardize the data. Having some variables with a very different range/scale can often create problems: most of the results may be driven by a few large values, more so that we would like. To avoid such issues, we take the approach to has to standardize the data by making all the initial raw attributes being standardized, scaling them by extracting mean and dividing by standard deviation. While this is typically a necessary step, we still should always do it with care: some times we may want our analytics findings to be driven mainly by a few attributes that take large values; other times having attributes with different scales may imply something about those attributes.

## Step 2:

Feature extraction:

The decision about which variables to use for clustering is a critically important decision that will have a big impact on the clustering solution. So we need to think carefully about the variables we will choose for clustering. After good exploratory research, we have a good sense of what variables may distinguish the customers and coupons they used.

Moreover, we often use only a few of the data attributes for segmentation (the segmentation attributes) and use some of the remaining ones (the profiling attributes) only to profile the clusters. For example, in market research and market segmentation, one may use attitudinal data for segmentation (to segment the customers based on their needs and attitudes towards the products/services) and then demographic and behavioral data for profiling the segments found.

After all the feature selection works, we tend to use two groups of feature to do the cluster analysis, Coupon purchase history by category and Coupon location attributes.

|  |  |  |  |
| --- | --- | --- | --- |
| Hotel and Japanese hotel | Leisure | Lesson | Nail and eye salon |
| Kanto | Kyushu-Okinawa | Northeast | Shikoku |
| AGE | Delivery service | Food | Gift card |
| Chugoku | East Sea | Hokkaido | Hokushinetsu |
| Hair salon | Kansai | Other coupon | Relaxation |

## Step 3:

Algorithms implementation:

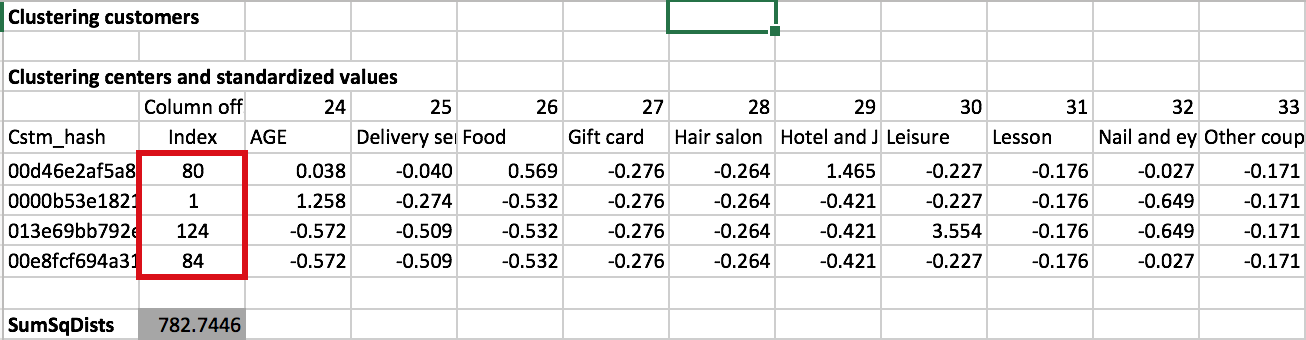
The algorithm we rely on to find the proper cluster is genetic algorithm, which can be roughly explained as below:

Genetic algorithm (GA) provides a method of intelligently searching an optimization model’s feasible region for an optimal solution. Biological terminology is used to describe the algorithm. The objective function is called a fitness function, and a specification of values for all changing cells is called a chromosome. GAs have a particular advantage on non- smooth problems— those with discontinuities, for example. However, they are much less efficient than traditional algorithms such as the simplex method on nice problems.

## Step 4:

Discussion of conclusion:

We use solver to implement GA for us to find four optimal customer groups as below:



Data analytics is used to eventually make decisions, and that is feasible only when we are comfortable (enough) with our understanding of the analytics results, including our ability to clearly interpret them.

We also compare the averages of the profiling feature variable of each segment relative to the average of the variables across the whole customer pools. This can help us better understand whether there are indeed clusters in our data (e.g. if all segments are much like the overall population, there may be no segments). For detailed conclusion summary will stated in the last part of this paper.

And we also did some cross validation analysis to make sure that the cluster we have from analysis is robust. The approach we take is to find out that how much overlap is there between the clusters found using different approaches, changing different cluster numbers. Specifically, for what percentage of our observations the clusters they belong to are the same across different clustering solutions.

The interesting finding for the customer cluster analysis is that:

1. Younger people tend to purchase beauty and hotel related coupons.
2. People who use coupon in Kansai and in Kanto are key indicators of customer groups.

# Calculating Propensities:

Since the required information to calculate the propensities is located in 4 different data sets, we first need to join the tables. This part of the project was carried out using SQL in the SAS SQL procedure. (We use SQL in SAS, because we used SAS to calculate the propensities.)

The code used to join the respective datasets is provided below:

/\*SQL JOIN USERDATA & VISITDATA\*/

**proc** **sql**;

create table work.master as

select userdata.user\_id\_hash,

userdata.age,

userdata.sex,

visitdata.purchase\_flg,

visitdata.view\_coupon\_id\_hash

from work.userdata

inner join work.visitdata

on userdata.user\_id\_hash=visitdata.user\_id\_hash;

**quit**;

/\*SQL JOIN MASTERDATA & COUPONLISTDATA\*/

**proc** **sql**;

create table work.master1 as

select master.user\_id\_hash,

master.age,

master.sex,

master.purchase\_flg,

master.view\_coupon\_id\_hash,

Couplistdata.discount\_price,

Couplistdata.price\_rate,

Couplistdata.en\_genre,

Couplistdata.en\_small\_area,

Couplistdata.en\_large\_area

from work.master

inner join work.Couplistdata

on master.view\_coupon\_id\_hash=Couplistdata.coupon\_id\_hash;

**quit**;

Once the above data preparation steps are carried out we ran the following SAS code to obtain the Propensities.

/\*Calculating Customer Propensities for individual genres\*/

Title "Propensity";

**Proc** **freq** data=master1;

tables USER\_ID\_hash\*en\_genre\*PURCHASE\_FLG /out=Propensity;

**run**;

This code calculates the propensities based on the formula:

Where i represents individual customers.

The code saves the propensities into another file, which is then exported to Excel.

# The Assignment Model:

**Objective:** In this assignment model we want to assign a coupon category to a list of customers based on their propensities, while constraining the model to one coupon category per customer and limiting the total number of coupons in each category according to match the sales revenue quota, in order to maximize Ponpare’s expected revenue.

**Sources:** The datasets for coupon details, customer details, purchase details and visit details were obtained from Ponpare through their competition posted on Kaggle.

**Model Approach:** The expected revenue is maximized using integer programming, by assigning a coupon category to a customer, while restricting the ‘total coupons in each category’ to the ‘available coupons’, and ‘sum of coupons assigned to an individual customer’ to 1.

**Solution and Analysis:** An optimal expected revenue of 618.721 Yen was obtained using solver to maximize expected revenue, against Total Coupons per category ≤Available Coupon per category and Sum assigned to =1 , while the assignment matrix contains binary variables.

Where i represents the customer

j represents the coupon category

The assignment variable is the binary variable determining the coupon to customer assignment.

The average category discounted price is the average price after discount for coupons within the category. (The averages were obtained from SAS code)

% cut per coupon is the % of revenue that is obtained by Ponpare on customers utilizing the coupon.

The results shown in the excel file show that despite a higher discounted price coupons are not assigned to these categories because of the low propensity. While in other cases coupons are allotted by propensity but limited to the total allowable coupons in a category.

Recommendation:

The coupons should be assigned to the individual customers according to the excel sheet. For example customer hash ‘0000b53e182165208887ba65c079fc21’ should be assigned a Relaxation coupon.

****

# Discriminant Analysis Model:

We developed a discriminant analysis model to try to observe a general ‘rule’ from the data. The analysis was carried out for two groups (Purchase and Not purchased) and many explanatory variables.

We chose customer age and sex, discounted price, price rate, individual coupon categories (binary variables) and large area location (binary variables) as explanatory variables. Weights were assigned from -1 to 1 to each of these variables. The product of the weights and explanatory variables was used to obtain the score. The highest and lowest score for the constraint limits were obtained by assigning all weights as 1, and taking the maximum value. The weights and the cutoff value for classification were used as decision variables. According to the cutoff selected, solver will try to classify the purchase as yes or no. The percent correct classifications are calculated by

Running solver to maximize % correct classification against weight and cutoff constraints mentioned earlier an optimal value of 95.36% correct classification was obtained. This value felt too high. On deeper inspection we found that 31479 values were predicted as NO and were actually NO, while 1530 values were predicted as NO and were actually a YES. However no values were predicted as a YES. Yet a 95.36% correct classification was obtained. This is due to the learning data being skewed: Only 4.6% of the Purchase variable records are YES.

The weight’s sign shows the likeliness of the Purchase being a YES or NO. A negative sign would indicate affinity to NO. The greater the magnitude of the weight the close it is towards predicting a YES or a NO. For instance from the excel sheet a rule that the probability of purchasing an item is higher if the coupon belongs to the Food genre and could be used in the Northeast area of Japan.

# Result 1 Cluster Analysis:

According to the clusters analysis result, we find 4 Clusters on customer side filtered by their purchasing behavior and age.

We take reference to those four benchmarks to understand the whole customer dataset. Among them, Cluster 1 makes up 21% of total customers. This type of customers aged at 41 on average and ranged from 19 to 59. They averagely buy other coupon 1.50 times, giftcard 1.158 times and Hotel 1.14 times, as 3 top popular coupons in this cluster. The customers are mainly from Kanto, Kansai.

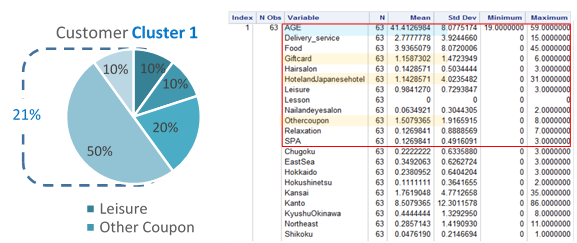


Figure 1 detailed variables in customer cluster 1

Cluster 2 contributes to 24% of total customers. This type of customers aged at 56 on average and ranged from 45 to 77. On average, everyone buys food coupons at 1.4 times, hotel coupons at 0.59 times and other coupon at 0.63 times.

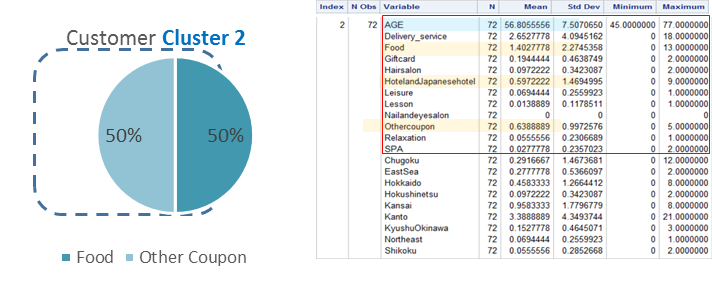


Figure 2 detailed variables in customer cluster 2

Cluster 3 contributes 5% of total customers. Compared to previous cluster, this type of customers are quite young with average 35 year old. The group of cluster 3 purchase food coupons at the average of 1.7333 times, lesson coupon at the average of 1.13333 times and othercoupon at the average of 1.53333.The customers are also domestically from Kansai, and Kanto.

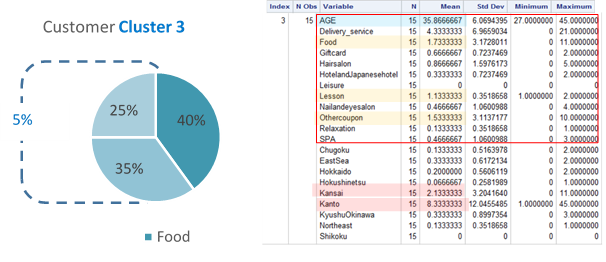


Figure 3 detailed variables in customer cluster 3

Cluster 4 contribute to half of total customers and quite diversified. This type of customers age at the average of 34 and ranged from 20 to 50.They tend to have various preference in purchasing behavior, the top 1 option is other coupon category. And they are mainly from Kanto.

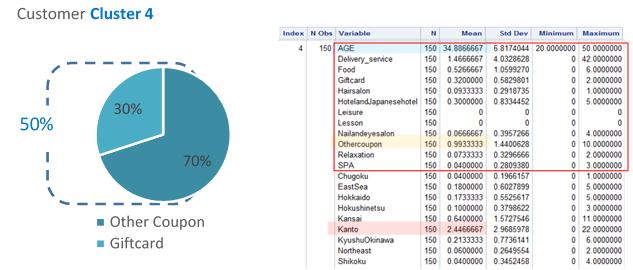


Figure 4 detailed variables in customer cluster 4

# Result 2 Propensity rate:

We calculated the propensity rate of 300 customers and this result will be further used to analyze the accepted rate of recommended coupons and the revenue of Coupon Company through Binary assignment entry using simplex algorithm.



Figure 5 Propensity rate of 300 customers<10 samples>

# Result 3 Revenue, Propensity, Assignment Formula:

Considering the reality,we build up some assumptions to do optimization analysis on our dataset and receive the binary table as below. Through the analysis on customer purchase behaviors, we could better assign coupons to customer, maximize company profits by adding the acceptable possibility of customer to certain amount of coupons.



Figure 6 Binary table of coupon assignment <10 samples>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Should I put this table??if not just delete it  To understand the effectiveness of our classification, we did a matrix to validate the correctness of assigning the coupon to each cluster of customer. And we reach to the result of correct classification at 95.36%.  Table 1 Classification matrix   |  |  |  | | --- | --- | --- | | **Classification matrix** | |  | | **(Actual along side, predicted along top)** | | | | |  | YES | NO | | | YES | 0 | 1530 | | | NO | 0 | 31479 | | |  |  |  | | | **Percent correct classifications** | | | | 95.36% |  |  | | | |
|  |  |

**Conclusion**

In this paper, we evaluate coupon effectiveness, in other words, the possibility of customer accept the recommended coupon based on customer purchasing behaviors and individual background. To achieve this goal, we build up a model to firstly calculate the propensities of customers to various coupon attributes such as coupon genre, discount rate, and category. Specifically, if a customer purchase 4 coupons all together last year, including 2 food, 2 lession. We would assume that this customer maintain continues purchasing behavior next time based on large dataset sample. To enhance our understanding on customer behavior and precise marketing recommendation, we applied cluster analysis in order to classify customers into 4 groups with significantly different characteristics, which helps us to better provide recommendation and promotion strategy to various group. Lastly, we developed another model to assign limited number of coupons based on propensity Linear Programming with Binary entry using simplex algorithm. We imitate the external environment by adding constraints on number of assigned coupons and the number of accepted coupons. Then we used solver to give out the optimal allocation strategy leading to the maximum profits of the company.

In future work, we need to do more research on the assumption that we have already made for this report because of the lacking data. Firstly, we need to specifically look into the discount rate of each coupon and evaluate the ideal discount considering coupon usage date, instead of taking average of discount rate in each category to stand for discount rate. This could be used to make use of the customer surplus and the supply of coupon. Secondly, we need to deeply navigate customer background information, such as income, credit rate, gender, etc. Then we could use a more detailed and distributed cluster analysis to dig deeply on what other factors may truly influence the customer’s final decision. Thirdly, we may take time series into consideration and classify the time into 12 months, 6 months, 3 months. Each time we propose the promotion strategy or provide recommendation to users, we could further record the progress and keep track with sales records. By doing this, we could improve data model step by step and keep company in a competitive status.

# References:

1. Dataset <https://www.kaggle.com/c/coupon-purchase-prediction/data>
2. Genetic Algorithms <650 Course Slides>
3. Cluster Analysis<650 Course Slides>
4. The Business of Coupons-Do coupons lead to repeat, Margaret Ross <http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2605&context=utk_chanhonoproj>
5. Measuring the Success of Your Coupon Campaign,  Matt Winn <http://www.volusion.com/ecommerce-blog/articles/measuring-your-coupon-campaign/>
6. How to Implement an Effective Mobile Coupon Strategy, JT Ripton <http://fitsmallbusiness.com/mobile-coupon-marketing/>
7. A Smarter Way To Reduce Customer Churn <http://www.forbes.com/forbes/welcome/>