

CUSTOMER SEGMENTATION ANALYSIS USING RFM MODEL

Project made by: Mai Huy Khang

**Notes: Because the dataset obtained can only be used for research and simulation purposes and has no commercial value. There will be errors; please forgive the shortcomings.*



TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
1.1. Definition.....	1
1.1.1. Customer segmentation.....	1
1.1.2. Customer Segmentation using RFM analysis.	1
1.2. Dataset.....	2
CHAPTER 2: RFM SCORES TRANSACTION.....	3
2.1. Data preprocessing.....	3
2.2. Calculate the Recency, Frequency and Monetary score for each customer..	4
2.2.1. Determining Recency, Frequency, Monetary values.....	4
2.2.2. Computing Quantile of RFM score	4
CHAPTER 3: RFM INSIGHT ANALYTIC	7
3.1. Distribution analyst	7
3.1.1. Customer segmentation distribution.....	7
3.1.2. Trending of customer segmentation throughout the fiscal year.....	8
3.2. General business performance	9
3.2.1. Detail information by each customer segmentation.....	11
CHAPTER 4: RECOMMENDATIONS	13
4.1. For each group of customer segmentation:.....	13
4.2. For business performances:	13

NAME OF TABLES AND PICTURES

Picture 1.1 Types of Segmentation	1
Picture 2.1 Features in dataset.....	3
Picture 2.2 Detail statistics	3
Picture 2.3 Calculating elements for RFM model.....	4
Picture 2.4 Computing Quantile of RFM score	5
Table 2.1 Customer Segmentation definition	6
Picture 3.1 Customer segmentation proportion distribution.....	7
Picture 3.2 Customer segmentation quantity distribution	7
Picture 3.3 Trending of customer segmentation throughout the fiscal year	8
Picture 3.4 Sale performance weekly, monthly and quarterly	10
Table 3.1 Detail information about customer segmentation groups	11
Table 4.1 Recommendation for each group of customer segmentation	13

CHAPTER 1: INTRODUCTION

1.1. Definition

1.1.1. Customer segmentation

Customer segmentation is a method of dividing customers into groups or clusters on the basis of common characteristics. The market researcher can segment customers into the B2C model using various customer's demographic characteristics such as occupation, gender, age, location, and marital status. Psychographic characteristics such as social class, lifestyle and personality characteristics and behavioral characteristics such as spending, consumption habits, product/service usage, and previously purchased products. In the B2B model using various company's characteristics such as the size of the company, type of industry, and location.

Types of Segmentation



Picture 1.1 Types of Segmentation

1.1.2. Customer Segmentation using RFM analysis.

RFM (Recency, Frequency, Monetary) analysis is a behavior-based approach grouping customers into segments. It groups the customers on the basis of their previous purchase transactions. How recently, how often, and how much did a customer buy.

- **Recency (R):** Who purchased recently? Number of days since last purchase (least recency)
- **Frequency (F):** Who has purchased frequently? It means the total number of purchases. (High frequency)
- **Monetary Value(M):** Who has a high purchase amount? It means the total money customer spent (high monetary value)

In this report, each of the three variables (Recency, Frequency, and Monetary) consists of five equal groups, which create 125 (5x5x5) different customer segments

and in this circumstance, I will separate into 11 big groups of customer segmentation (detail in Chapter 2)

The RFM Model assigns a score of 1 to 5 to criterion. Recency assigns 1 point for the longest period since the last time a consumer utilized the service, and 5 points for the shortest time. The higher the score for Frequency and Monetary, the better the service utilization density and total income.

1.2. Dataset

The dataset of 541,909 records was created by customers from 38 countries. Taking place from December 2010 – 2011 with total sales over 9.7 million dollars and around 4070 products.

CHAPTER 2: RFM SCORES TRANSACTION

2.1. Data preprocessing

The data table I used has 8 features about customers' transactions information and their countries.

```
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   InvoiceNo    541909 non-null   object
1   StockCode    541909 non-null   object
2   Description  540455 non-null   object
3   Quantity     541909 non-null   int64
4   InvoiceDate   541909 non-null   datetime64[ns]
5   UnitPrice    541909 non-null   float64
6   CustomerID   406829 non-null   float64
7   Country      541909 non-null   object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)

#Check duplicated values
df.duplicated().value_counts()

False      536641
True         5268
dtype: int64
```

Picture 2.1 Features in dataset

Overall, all features fit their right datatype and there are some duplicates records. On top of that, Description and CustomerID variances have less quantity due to NULL values.

	Quantity	UnitPrice	CustomerID
count	541909.000000	541909.000000	406829.000000
mean	9.552250	4.611114	15287.690570
std	218.081158	96.759853	1713.600303
min	-80995.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13953.000000
50%	3.000000	2.080000	15152.000000
75%	10.000000	4.130000	16791.000000
max	80995.000000	38970.000000	18287.000000

Picture 2.2 Detail statistics

For further information, we can see that there are negative values in the sale context. Through the returns, refunds and discounts abilities checking, there is no clue for those cases and turns out there are fault data that need to be removed from the dataset.

After summarizing errors, data preprocessing was performed as following:

- Removing duplicated and negative values
- Removing NULL values in Description and CustomerID variances
- Calculating for Sales column (Price units*Quantity)

2.2. Calculate the Recency, Frequency and Monetary score for each customer

2.2.1. Determining Recency, Frequency, Monetary values

As given definition above, I calculate the elements for RFM model as:

```
import datetime
current_date = max(df['InvoiceDate']) + datetime.timedelta(days=1)

df_RFM = df_notna.groupby('CustomerID').agg({
    'InvoiceDate': lambda x: (current_date - max(x)).days,
    'InvoiceNo': 'count',
    'Sales': 'sum'
}).round(2)
```

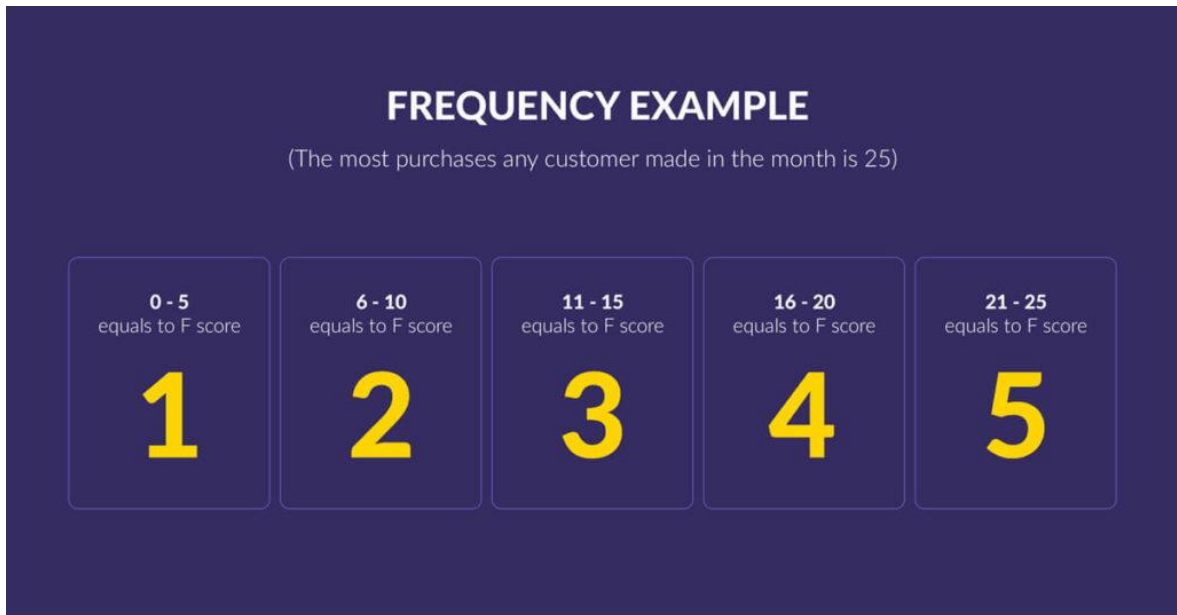
Picture 2.3 Calculating elements for RFM model

- Recency: calculated by the max days in dataset plus 1 day, so the shortest period for purchasing to current day would be 1 unit.
- Frequency: calculated by counting the quantity of invoices for each customer.
- Monetary values: calculated by summing the total payments of each customer.

2.2.2. Computing Quantile of RFM score

Customers with the lowest recency, highest frequency and monetary amounts considered as top customers. Specifically, the qcut(), quantile-based discretization

function, would be used in this transform step to bin the data based on sample quantiles.



Picture 2.4 Computing Quantile of RFM score

After scoring each criterion in RFM model, I group the dataset in following segmentation below:

Segment name	Definition	Score list
<i>Champions</i>	Bought recently, order often and spend the most.	555, 554, 544, 545, 454, 455, 445
<i>Loyal</i>	Orders regularly. Responsive to promotions.	543, 444, 435, 355, 354, 345, 344, 335
<i>Potential Loyalists</i>	Recent customers who spent good amounts.	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
<i>New Customers</i>	Bought most recently.	512, 511, 422, 421 412, 411, 311
<i>Promising</i>	Potential loyalist a few months ago. Spends frequently and a good amount. But the last purchase was several weeks ago.	525, 524, 523, 522, 521, 515, 514, 513, 425, 424, 413, 414, 415, 315, 314, 313
<i>Need attention</i>	Core customers whose last purchase happened more than one month ago.	535, 534, 443, 434, 343, 334, 325, 324
<i>About to sleep</i>	Made their last purchase a long time ago but in the last 4 weeks either visited the site or opened an email.	331, 321, 312, 221, 213, 231, 241, 251

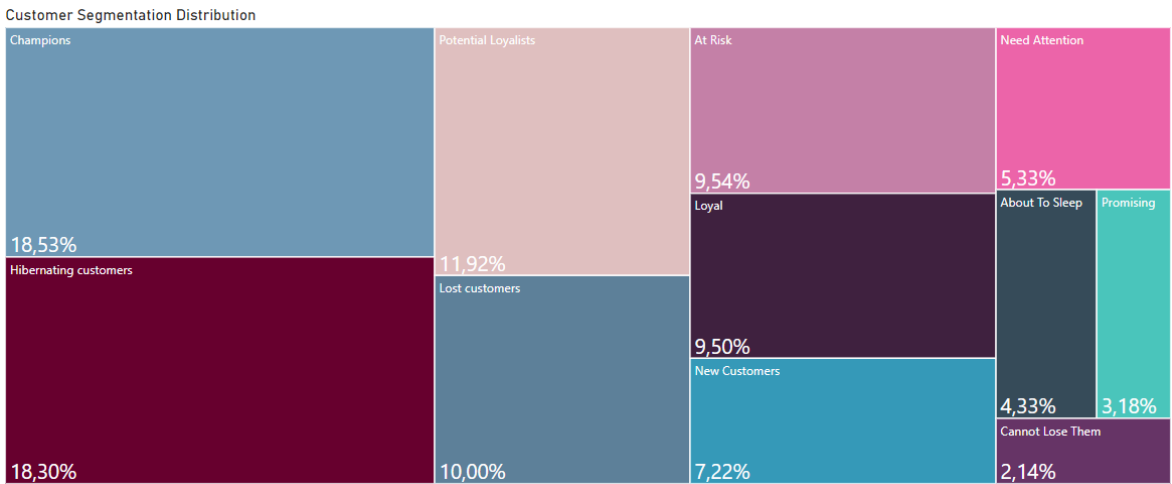
<i>Cannot Lose Them</i>	Made the largest orders, and often. But haven't returned for a long time.	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
<i>At Risk</i>	Similar to 'Cannot Lose Them' but with smaller monetary and frequency value.	155, 154, 144, 214, 215, 115, 114, 113
<i>Hibernating customers</i>	Customers who made smaller and infrequent purchases before but haven't purchased anything in a long time.	332, 322, 233, 232, 223, 222, 132, 123, 122, 212, 211
<i>Lost</i>	Made last purchase long time ago and didn't engage at all in the last 4 weeks.	111, 112, 121, 131, 141, 151

Table 2.1 Customer Segmentation definition

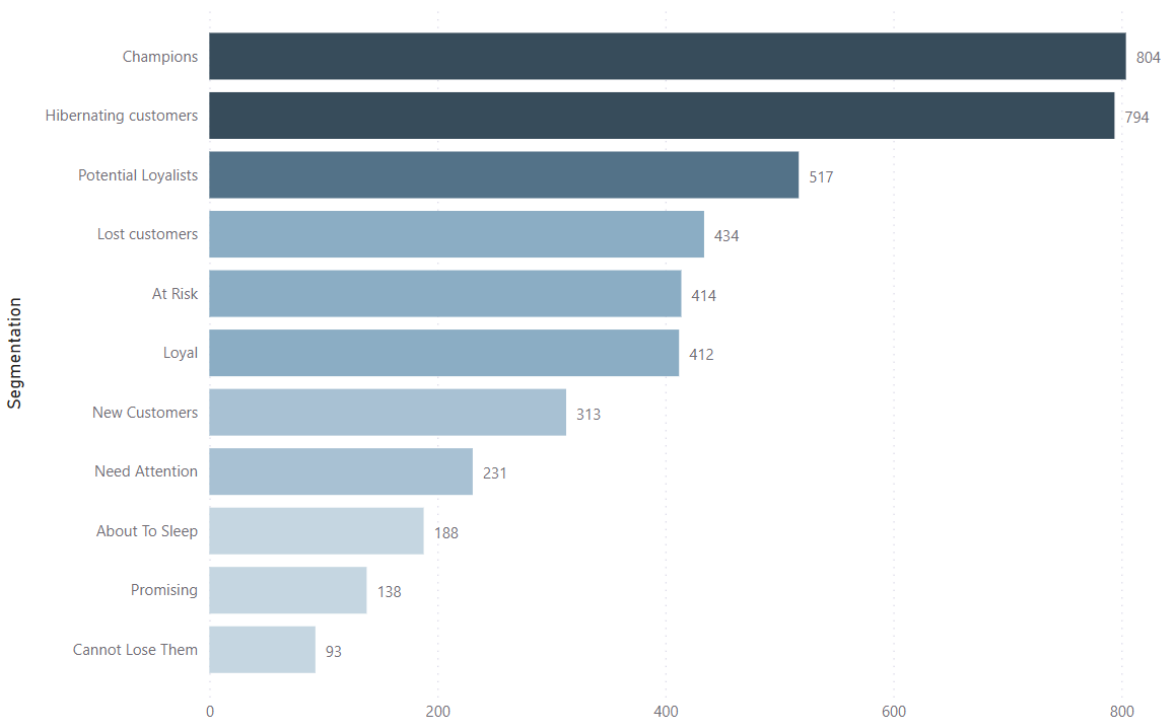
CHAPTER 3: RFM INSIGHT ANALYTIC

3.1. Distribution analyst

3.1.1. Customer segmentation distribution



Picture 3.1 Customer segmentation proportion distribution



Picture 3.2 Customer segmentation quantity distribution

❖ **Warning that negative groups of customers still made up a large proportion.**

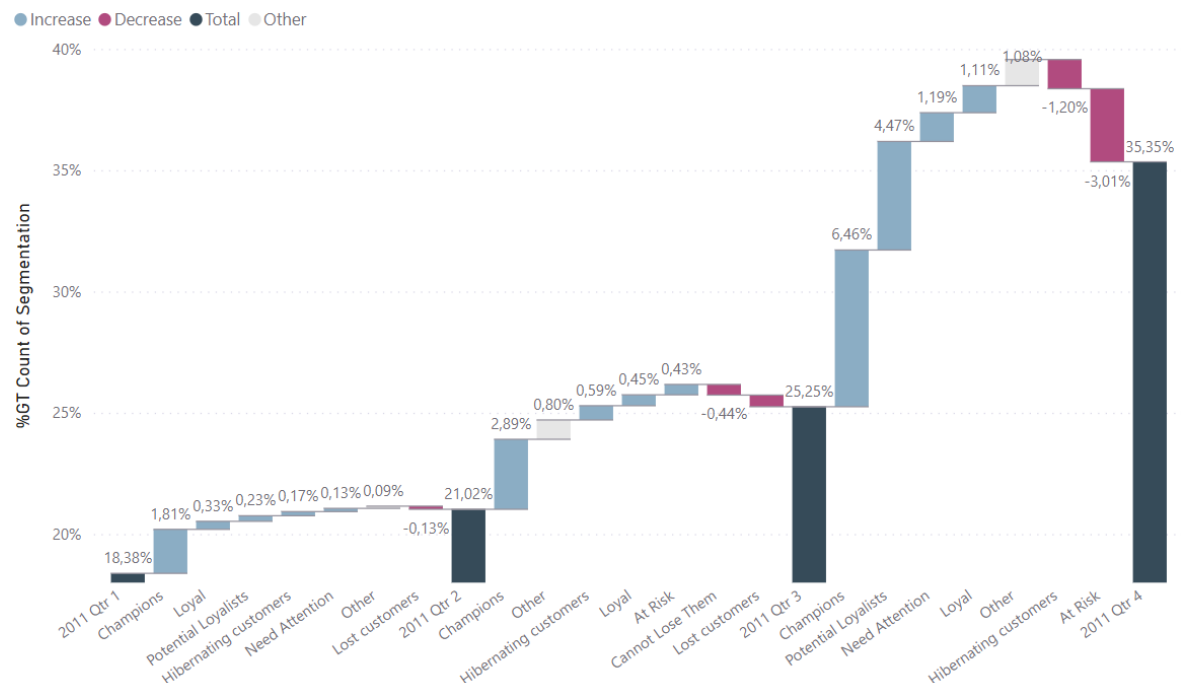
In the negative group stage, there are concerns that hibernating, lost and at risks groups remained at very high proportions and total of them accounted for almost 40% of all customers. As those groups can lead to lost or churn customers, which drawback to company revenue and wealth, we will need some solutions to plumb and convert them to regular customers.

❖ **Maintain champions group and take care more of the groups which in conversion progress such as new customer, need attention, about to sleep and promising group.**

Champions group is vital to our company, which accounted for 18,53% and brought the highest revenue to the business. On top of that, other in conversion progress such as new customer, need attention, about to sleep and promising group of customers were only stay from 3-7%, so we need more attention for those to move them to loyal and champions' type.

3.1.2. Trending of customer segmentation throughout the fiscal year

As the distance from each group varies most of the time, we tend to analyze the trend in quarter period to see how they grow and fall clearly.



Picture 3.3 Trending of customer segmentation throughout the fiscal year

❖ **The champions, potential loyalists, loyal and need attention groups tended to increase consistency.**

At first look, we can recognize that customers in the champions group represented the growth in customer quantities, which accounted for around 50 to 71% of total

growth and always took the top of list. Followed by potential loyalists and loyal groups with a slight upward almost 10% of total grand. Those were good signs of the company's effort in attracting and maintaining their customers.

❖ **Effort to plumb cannot lose and lost customer groups.**

Every customer that falls to those groups shows a reduction in company wealth. In fact, the company had made an effort to prevent that damage and the result could not be happier when the proportion dipped from 0,13% to 0,48% for the lost customer group and 0,07% to 0,44% for the cannot lose them group at the end of the second quarter.

❖ **Fluctuating in hibernating and risk groups of customers.**

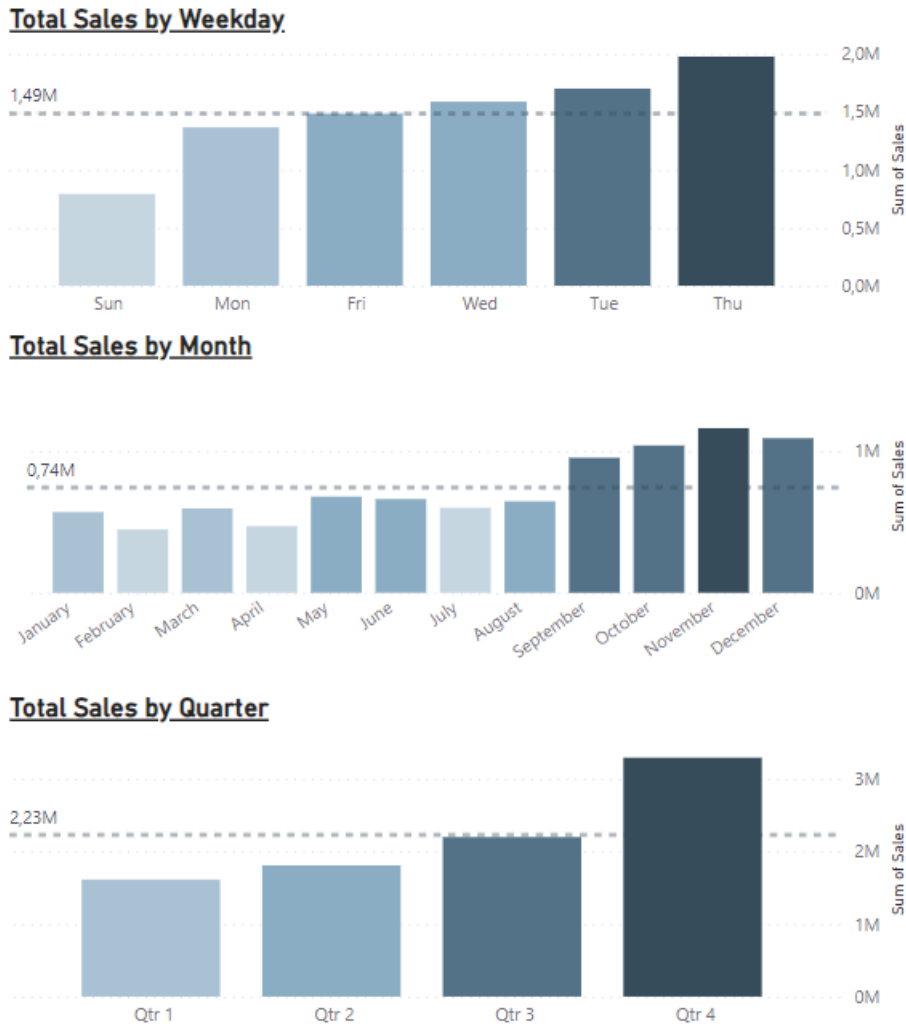
Hibernating and risk groups of customers are dangerous stages before our customers turn into lost and churn, that's why those groups need to take care wisely.

The statistics showed that hibernating groups literally surged from the beginning of the year at 0.17% and continued to go up to the peak itself at 0.59% during the second quarter. Finally, downtrend was seen at the final quarter which dropped for 1,2%

Besides, the risk groups appeared to be rising in the middle quarter and was the deepest downtrend in the final quarter by 3,01%.

3.2. General business performance

Time-series analysis is a method of analyzing a collection of data points over a period of time. Instead of recording data points intermittently or randomly, time series analysts record data points at consistent intervals over a set period of time. Below are graphs that show sale performances through different catalogs including weekly, monthly and quarterly.



Picture 3.4 Sale performance weekly, monthly and quarterly

In general, we can see that the company had their high peaks during weekdays (Monday to Friday). In addition, the busies months in the year started from September to December, which was the fourth quarter with the revenue accounting for almost 37% in year grand total percentage.

On the other hand, the low volume of sales regularly dropped almost twice at the weekend by 700.000 dollar and gradually rose in next days. On top of that, the performances were also not outstanding in the first 3 quarters of the year period, which went up and down several times between 1,3 and 2 million dollars and obviously below the revenue average of 2,23 million dollars.

As the main products of the company are household and personal goods, being purchased during weekdays and at the end of the year is an understatement. Our customers might make the conversion whenever they get the chance and a good deal for their need.

3.2.1. Detail information by each customer segmentation

Segment name	Revenue (thousand)	Top markets	Purchasing habits (Top down)		
			<i>Weekday</i>	<i>Month</i>	<i>Time</i>
<i>Champions</i>	5.494	UK, Germany, French	Tues to Wed	Sep to Dec	11:30 AM – 1:30 PM
<i>Loyal</i>	990	UK, Germany, French, Belgium	Tues, Mon	Oct	12:00 PM – 1:30 PM
<i>Potential Loyalists</i>	303	UK, Germany, French	Sun, Thus	Oct, Nov	11:30AM – 1:30 PM
<i>New Customers</i>	66	UK, Germany, French	Mon to Thus	Oct, Nov	12:30 PM – 1:30 PM
<i>Promising</i>	328	UK, French	Mon to Thus	Nov, Dec	12:30 PM – 1:00 PM
<i>Need attention</i>	342	UK, French, Spain	Thus, Tues	Sep to Dec	12:00 PM – 1:00 PM
<i>About to sleep</i>	54	UK	Mon, Sun, Thus	Oct	02:00 PM – 02:30 PM
<i>Cannot Lose Them</i>	301	UK, Germany, Channel Island, Japan, Switzerland	Thus, Web	Jan	12:00 PM – 12:30 PM
<i>At Risk</i>	635	UK, Germany, French, Spain	Thus, Wed, Sun	Mar, May, Sep	11:30 AM – 2:30 PM
<i>Hibernating customers</i>	320	UK, Germany, French	Thus	Sep	11:30 AM – 2:30 PM
<i>Lost</i>	77	UK, French	Thus, Tues, Sun, Mon	Mar, Dec	1:30 PM

Table 3.1 Detail information about customer segmentation groups

❖ **At risk customers needed the highest taking care level.**

At risk group not only gained high revenue for the company, but also contained rare markets including Spain. Moreover, those customers were one of the widest periods for purchasing, which was still potential and needed to take care more.

❖ **Thursday is the best day to take care of customers.**

The statistics have proven that all of the customer groups that need to take care have purchasing habits on Thursday. The company needs to offer discounts and other benefits for those groups on Thursday to maintain and attract customers.

❖ **The company's golden hours occurred from 11:30AM – 03:00PM**

This period of time is marked as the most popular time for making purchases for every group of customers. Normally, people have their break and lunch during this period, so that the company needs to do activities and also special offers for old and new customers at those times.

❖ **The positive groups made purchases in the fourth quarter mostly while the negative one made it in the first quarter.**

We all know that nearest to the end of the year, the volume of sales goes up sharply, there is no strange that regular and core customers wait for special offers for that period while the negative one just bought by chance.

❖ **UK, Germany, French are our main markets in this fiscal year.**

UK, Germany, French, especially UK is the most attractive markets of the company throughout the year, spread out every segment and accounted for almost 94% of total grand.

CHAPTER 4: RECOMMENDATIONS

4.1. For each group of customer segmentation:

Segment	Actionable Tip
<i>Champions</i>	<ul style="list-style-type: none"> ○ Reward them. ○ Can be early adopters of new products. ○ These customers have proven to have a higher willingness to pay, so don't use discount pricing to generate incremental sales. ○ Most likely to send referrals.
<i>Loyal</i>	<ul style="list-style-type: none"> ○ Upsell higher value products. ○ Advocacy programs and reviews. ○ Giveaway gift with point's marks.
<i>Potential Loyalists</i>	<ul style="list-style-type: none"> ○ Offer membership / loyalty program. ○ Keep them engaged. ○ Offer personalized recommendations.
<i>New Customers</i>	<ul style="list-style-type: none"> ○ Provide on-boarding support. ○ Give them early access. ○ Start building relationships through invitation and points events. ○ Special discounts with occasions (first-purchasing, first invitation...)
<i>Promising</i>	<ul style="list-style-type: none"> ○ Offer coupons. ○ Bring them back to the platform and keep them engaged. ○ Offer personalized recommendations.
<i>Need attention</i>	<ul style="list-style-type: none"> ○ Make limited time offers. ○ Offer personalized recommendations.
<i>About to sleep</i>	<ul style="list-style-type: none"> ○ Make subject lines of emails very personalized. ○ Revive their interest by giving a specific discount on a specific product. ○ Voucher old membership returns
<i>Cannot Lose Them</i>	<ul style="list-style-type: none"> ○ Win them back via renewals or newer products, don't lose them to competition. ○ Talk to them if necessary. ○ Spend time on the highest possible personalization. ○ Voucher old membership returns.
<i>At Risk</i>	<ul style="list-style-type: none"> ○ Provide helpful resources on the site. ○ Send personalized emails.
<i>Hibernating customers</i>	<ul style="list-style-type: none"> ○ Include them in your standard email communication but regularly check if they don't flag your content as spam. ○ Do not overspend on this segment.
<i>Lost</i>	<ul style="list-style-type: none"> ○ Revive interest with reach out campaign. ○ Ignore otherwise.

Table 4.1 Recommendation for each group of customer segmentation

4.2. For business performances:

❖ **Low performance at weekday**

Leverage discounts for old inventories at weekend:

Make full use of discount mindsets and selling-off inventories in long term storage or extremely low demands to attract customers and gain more revenue and recognition on the weekend.

Targeted Marketing Campaigns:

Create targeted marketing campaigns to specifically address weekdays. Offer weekday-specific promotions, discounts, or incentives to attract customers during those times. Utilize email marketing, social media, and other advertising channels to reach the target audience effectively.

Expand Product Offerings:

Consider introducing new products or services that are appealing specifically during weekdays. For example, we could offer lunch specials or introduce limited-time weekday-only items. Diversifying our offerings can attract new customers and encourage existing ones to visit on weekdays.

Customer Loyalty Programs:

Implement a customer loyalty program that rewards frequent weekday shoppers. Offer exclusive discounts, early access to promotions, or additional benefits for loyal customers. This can incentivize customers to choose our business on weekdays.

❖ **Leverage the potential volume of Thursday and activities during golden hours.**

Special offers for products in period frames

Depending on the high volume of specific hours in a day, we can make discounts and limited offers in a period of 2 or 3 hours for some products that need to be upselling or in most concerns by customers. Those will help to give out our inventory and approach to more customers due to their habits.

Livestream activities on Thursday

There is no doubt that our customers nowadays like to watch online shopping in order to witness the products in real-life and take special offers of any. Leveraging that, we should make efforts to livestreams in platform that may approach not only old but new customers on this high-volume day.

❖ **Enforce UK and Germany markets exploitation.**

Separate service and on-site agencies for those countries

As customers in those countries contribute huge amounts of revenue and impacts. We need to make sure the customer services for them are the best and broaden the presence of the company in their area will be a great opportunity to communicate and research further to their communities.

Develop more brick & mortal facilities in those countries.

This will improve the recognition of the brand and make more comforts for old and new customers. On top of that, we can advance our competitiveness in those markets and build trust for potential customers in the future.