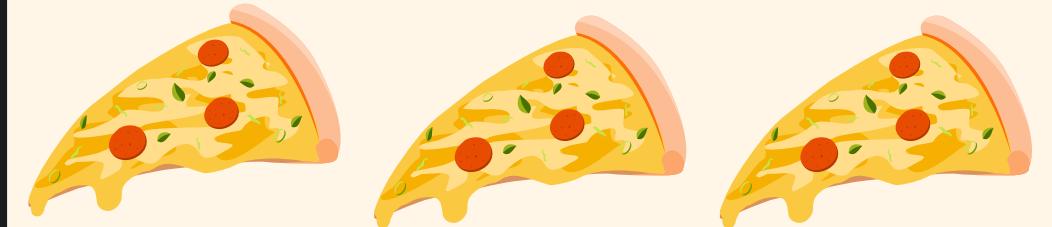


Analysis Report

Customer Portrait

Company: Pizzahut



Contents



METHODOLOGY



DATA ANALYSIS



DATA MAPPING



SUMMARY



OVERVIEW OF ENTERPRISE



OVERVIEW OF ANALYSIS METHOD

Methodology

Data

The data used for analysis in this report is transaction data obtained from a very large chain of pizza stores, specifically sourced from the Pizzahut system in Vietnam. Each piece of data represents a record of individual transactions with customers, including: BillID, Channel, OrderFrom, TransactionDate, SalesAmount, CustomerID, CustomerGender, VoucherStatus, Province.

Time

The data collection period in this report spans over 1 year, from October 1st, 2021, to January 9th, 2023.

Analysis method

The analysis report will utilize a variety of integrated analytical methods to identify the profiles of customers who recover capital and those who do not. The methods employed in the article will include cohort analysis, customer lifetime value, market basket analysis, combined with common calculation methods such as sum and average.

DATA MAPPING

Column name	Meaning	Data type	Example
BillID	The ID of the transaction. This ID is unique in the dataset	Text	1
Channel	The method for delivering food to customers.	Text	Delivery Take away Dine in
OrderFrom	Where do customers place orders from	Text	App Call center Store Website
TransactionDate	When does the transaction take place	Date	2021-10-01

DATA MAPPING

Column name	Meaning	Data type	Example
SalesAmount	The amount of money received from the transaction	Number	296891
CustomerID	ID of the customer	Text	1753863
CustomerGender	The gender of the customers	Text	Male Female Unknown
VoucherStatus	Did the customer use a voucher for their transaction?	Text	Yes No
Province	Province of each customer	Text	Ho Chi Minh City Hanoi

OVERVIEW OF ENTERPRISE



Pizza Hut is a globally renowned restaurant chain specializing in pizza and other Italian-American cuisine. Pizza Hut is celebrated for its diverse menu, which includes various styles of pizzas, pasta, wings, salads, and desserts.

What sets Pizza Hut apart is its commitment to innovation and quality. Beyond its delicious food, Pizza Hut is known for its distinctive dining experience, offering dine-in, takeout, and delivery options to customers. The company prides itself on its friendly service and inviting atmosphere, making it a popular choice for family gatherings, casual outings, and celebrations alike.

OVERVIEW OF ENTERPRISE



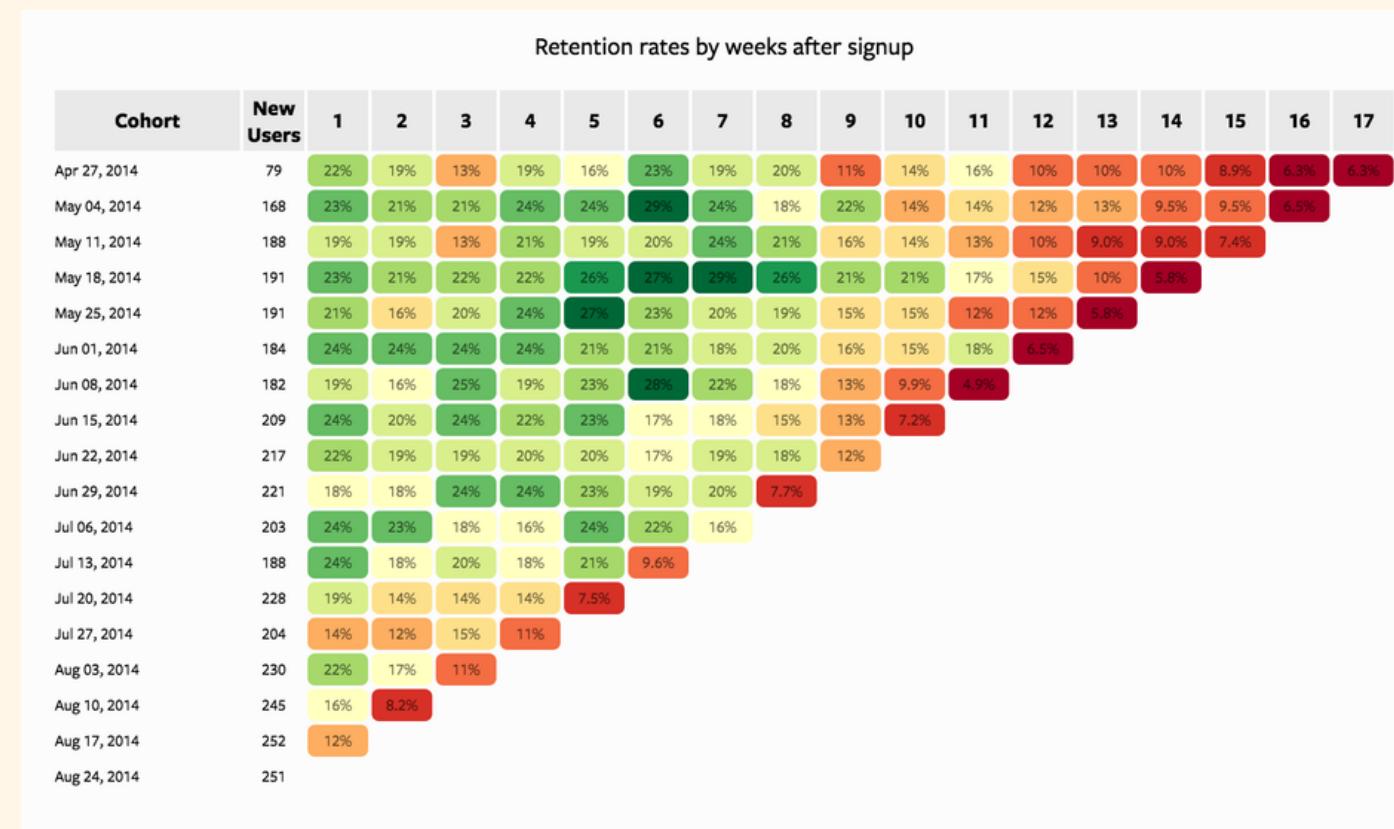
The data is collected from all Pizza Hut store chains in Vietnam over a period of more than 1 year, from October 1st, 2021, to January 9th, 2023. Through this data, the report will clarify the issues that businesses need to pay attention to regarding customer retention, net dollar retention, and customer lifetime value, thereby making recommendations to improve the current issues of the business.

OVERVIEW OF COHORT ANALYSIS

Cohort analysis is a valuable method used in statistics and data analytics to study the behavior of specific groups over time. In business and marketing, it is particularly useful for understanding customer behavior and trends.

Cohorts are groups of individuals who share a common characteristic or experience within a specific time frame.

Cohort analysis involves tracking various metrics or key performance indicators (KPIs) over time within each cohort. Common metrics include customer retention rate, revenue per customer, customer lifetime value, and others relevant to the business goals.



OVERVIEW OF COHORT ANALYSIS

Cohorts are typically analyzed over regular intervals, such as weeks, months, or years, allowing for comparisons and trend identification. This helps in understanding how behaviors change over time within and across cohorts.

By analyzing cohort data, businesses can gain insights into customer behavior patterns, identify trends, and make data-driven decisions. These insights often lead to actionable recommendations aimed at improving customer satisfaction, retention, and overall business performance.

CUSTOMER LIFETIME VALUE

Customer Lifetime Value (CLV) is a metric that represents the total net profit a company expects to earn from a customer throughout their entire relationship with that customer. In essence, it quantifies the value of a customer to a business over the entire duration of their relationship.

By calculating Customer Lifetime Value, businesses can reap several benefits:

1. Performance evaluation: CLV can be used as a key performance indicator (KPI) to evaluate the effectiveness of marketing campaigns, customer service initiatives, and overall business performance. By monitoring changes in CLV over time, businesses can assess the impact of their strategies and make adjustments as needed.
2. Resource allocation: CLV provides insights into the most profitable customer segments, allowing businesses to allocate resources such as marketing budgets, customer service efforts, and product development initiatives accordingly. By investing resources where they are likely to yield the highest returns, businesses can improve overall profitability.
3. Strategic decision making: CLV helps businesses make informed strategic decisions regarding customer acquisition, retention, and relationship management. By understanding the long-term value of different customer segments, businesses can allocate resources more effectively to maximize profitability.

CUSTOMER LIFETIME VALUE

The formula to calculate the Customer Lifetime Value is:

$$CLV = AOV * APF * ACL$$

Customer Lifetime Value Formula:

Average Order Value *
Average Purchase Frequency *
Average Customer Lifespan

CUSTOMER SEGMENTATION

Customer segmentation is the process of dividing a customer base into groups of individuals who are similar in specific ways relevant to marketing, such as demographics, behavior, needs, or preferences. The goal of customer segmentation is to better understand customers and tailor marketing efforts, products, and services to meet their specific needs and preferences more effectively.

Customer segmentation allows businesses to:

- Identify high-value customer segments.
- Tailor marketing messages and offers.
- Improve customer satisfaction and loyalty.
- Optimize product development and innovation.

MARKET BASKET ANALYSIS

Market Basket Analysis (MBA) is a data mining technique used to identify the association between products purchased together by customers. It analyzes transactional data from sales records to discover patterns and relationships among items that tend to be purchased together. The primary goal of market basket analysis is to uncover insights that can be used to improve marketing strategies, optimize product placement, and enhance cross-selling and upselling opportunities.



MARKET BASKET ANALYSIS

The three most common metrics used in market basket analysis are:

- Support: This measures the frequency with which a particular combination of items occurs in transactions. It is calculated as the proportion of transactions containing the combination of items.
- Confidence: This measures the likelihood that if item A is purchased, item B will also be purchased. It is calculated as the proportion of transactions containing item A that also contain item B.
- Lift: This measures the strength of association between items in a transaction. Lift indicates how much more likely it is for two (or more) items to be bought together compared to if they were bought independently of each other.

Market basket analysis is widely used in retail, e-commerce, and other industries to:

1. Identify product affinities
2. Cross-selling and upselling
3. Inventory management
4. Personalized marketing

MARKET BASKET ANALYSIS

However, in addition to the traditional application of identifying items frequently purchased together by each customer to recommend other items, aimed at increasing revenue, market basket analysis can also be applied to determine which characteristics often co-occur with each other, thereby creating a comprehensive picture of the traits of an individual or entity.

In this analytical report, we will also utilize market basket analysis. However, it will not be used to identify items frequently purchased together to recommend products that customers may need. Instead, we will use market basket analysis to determine which characteristics often co-occur within two customer groups: one group comprising customers who recover their capital and the other group comprising customers who do not recover their capital.

DATA ANALYSIS

The problem presented is that before we can analyze the profile of each customer group, we need to segment the customers to determine which ones are potentially profitable and which ones are not profitable. With the assumption that the cost to acquire a new customer is 300,000 VND, we will segment the customers as follows: those with a lifetime value (LTV) below 300,000 VND will be classified as the non-profitable customer group, and those with an LTV greater than 300,000 VND will be classified as the profitable customer group.

First, we will start with calculating the customer lifetime value.

DATA ANALYSIS

The table on the right is calculated based on cohort analysis. The purpose of using cohort analysis is to calculate the customer retention rate. The end date of the data is 2023-01-01, so to calculate the customer retention rate, we will take the average customer retention rate at 2023-01-01 for each index. From there, we will have the average value of the customer retention rate.

From there, we will calculate the retention rate, and finally we will calculate the churn rate. The value of the churn rate is 0.98.

From there, we will calculate the value of ACL (Average customer value), which is a metric used to compute the customer lifetime value (CLV), and it is 1.02.

cohort_index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
cohort_yearmonth																
2021-10-01	1.0	0.15	0.14	0.13	0.12	0.12	0.10	0.11	0.10	0.10	0.10	0.09	0.09	0.07	0.08	0.03
2021-11-01	1.0	0.13	0.11	0.09	0.10	0.09	0.09	0.08	0.08	0.08	0.07	0.07	0.06	0.06	0.02	NaN
2021-12-01	1.0	0.11	0.09	0.09	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.05	0.06	0.02	NaN	NaN
2022-01-01	1.0	0.08	0.08	0.06	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.05	0.02	NaN	NaN	NaN
2022-02-01	1.0	0.08	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.02	NaN	NaN	NaN	NaN
2022-03-01	1.0	0.07	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.02	NaN	NaN	NaN	NaN	NaN
2022-04-01	1.0	0.07	0.06	0.05	0.06	0.05	0.05	0.04	0.04	0.02	NaN	NaN	NaN	NaN	NaN	NaN
2022-05-01	1.0	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.01	NaN						
2022-06-01	1.0	0.07	0.06	0.04	0.04	0.04	0.04	0.01	NaN							
2022-07-01	1.0	0.06	0.05	0.04	0.04	0.04	0.02	NaN								
2022-08-01	1.0	0.06	0.05	0.04	0.05	0.01	NaN									
2022-09-01	1.0	0.06	0.05	0.05	0.02	NaN										
2022-10-01	1.0	0.06	0.05	0.02	NaN											
2022-11-01	1.0	0.06	0.02	NaN												
2022-12-01	1.0	0.03	NaN													
2023-01-01	1.0	NaN														

```
1 retention_rate  
0.01933333333333327
```

```
1 churn_rate = 1 - retention_rate  
2 churn_rate  
✓ 0.0s  
0.9806666666666667
```

```
▷ ▾ 1 ACL = 1/churn_rate  
2 ACL  
[25] ✓ 0.0s  
... 1.0197144799456153
```

DATA ANALYSIS

Next, we will calculate the value of Average Order Value (AOV). This is a metric used to compute the customer lifetime value (CLV). AOV is calculated by dividing the profit earned from each customer by the total number of transactions. The AOV value will vary for each individual customer.

```
1 df1['AOV'] = df1['profit'] / df1['total_transaction']
```



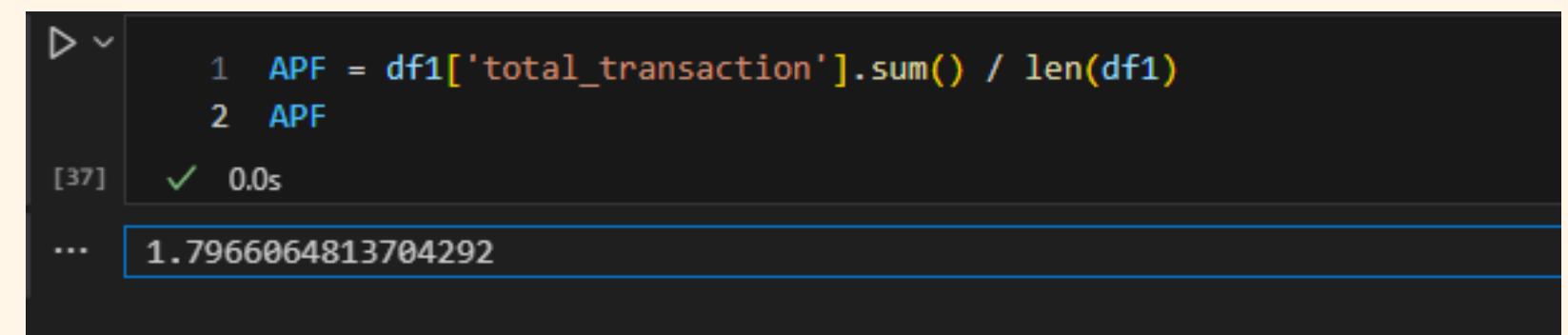
```
1 df1
```

	CustomerID	total_transaction	revenue	profit	AOV
0	0	1	411663	267580.95	267580.950000
1	1	1	105503	68576.95	68576.950000
2	8	1	1098496	714022.40	714022.400000
3	16	1	286558	186262.70	186262.700000
4	18	1	287492	186869.80	186869.800000
...
583637	2174065	1	138943	90312.95	90312.950000
583638	2174066	1	143685	93395.25	93395.250000
583639	2174072	3	953618	619851.70	206617.233333
583640	2174075	1	262489	170617.85	170617.850000
583641	2174083	1	185271	120426.15	120426.150000

583642 rows × 5 columns

DATA ANALYSIS

Next, we will calculate the value of Average Purchase Frequency (APF). This is a metric used to compute the customer lifetime value (CLV). APF is calculated by dividing the total number of transactions by the number of customers. In this dataset, APF will be equal to 1.79.



A screenshot of a Jupyter Notebook cell. The code is:

```
1 APF = df1['total_transaction'].sum() / len(df1)
2 APF
```

The output shows:

[37] ✓ 0.0s

... 1.7966064813704292

DATA ANALYSIS

Finally, we will calculate the value of CLV. To compute CLV, we simply need to multiply AOV * APF * ACL, from which we will obtain the CLV value for each customer as shown in the figure on the right.

```
[41] 1 df1['CLV'] = df1['AOV']*APF*ACL
[41] ✓ 0.0s
[42] 1 df1
[42] ✓ 0.0s
...
   CustomerID  total_transaction  revenue  profit      AOV      CLV
0            0                  1  411663  267580.95  267580.950000  4.902152e+05
1            1                  1  105503  68576.95   68576.950000  1.256347e+05
2            8                  1 1098496  714022.40  714022.400000  1.308107e+06
3           16                  1  286558  186262.70  186262.700000  3.412380e+05
4           18                  1  287492  186869.80  186869.800000  3.423503e+05
...
583637    2174065                  1  138943  90312.95  90312.950000  1.654556e+05
583638    2174066                  1  143685  93395.25  93395.250000  1.711025e+05
583639    2174072                  3  953618  619851.70  206617.233333  3.785281e+05
583640    2174075                  1  262489  170617.85  170617.850000  3.125763e+05
583641    2174083                  1  185271  120426.15  120426.150000  2.206238e+05
583642 rows × 6 columns
```

DATA ANALYSIS

In reality, if the Customer Acquisition Cost (CAC) of a business is 300,000 VND, we will perform customer segmentation as follows: for customers with a CLV less than 300,000 VND, those customers will be classified into the 'unqualified customer' group. Conversely, if a customer has a CLV greater than 300,000 VND, they will be classified into the 'qualified customer' group.

From the results on the right, it can be seen that the ratio between qualified customers and unqualified customers is 50:50.

CAC is 300000

```
[43] 1 df1['customer_segmentation'] = df1['CLV'].apply(lambda x: "unqualified customer" if x<300000 else("qualified customer"))
[43] ✓ 1.8s
```

```
[44] 1 df1
[44] ✓ 0.0s
```

	CustomerID	total_transaction	revenue	profit	AOV	CLV	customer_segmentation
0	0	1	411663	267580.95	267580.950000	4.902152e+05	qualified customer
1	1	1	105503	68576.95	68576.950000	1.256347e+05	unqualified customer
2	8	1	1098496	714022.40	714022.400000	1.308107e+06	qualified customer
3	16	1	286558	186262.70	186262.700000	3.412380e+05	qualified customer
4	18	1	287492	186869.80	186869.800000	3.423503e+05	qualified customer
...
583637	2174065	1	138943	90312.95	90312.950000	1.654556e+05	unqualified customer
583638	2174066	1	143685	93395.25	93395.250000	1.711025e+05	unqualified customer
583639	2174072	3	953618	619851.70	206617.233333	3.785281e+05	qualified customer
583640	2174075	1	262489	170617.85	170617.850000	3.125763e+05	qualified customer
583641	2174083	1	185271	120426.15	120426.150000	2.206238e+05	unqualified customer
583642 rows × 7 columns							

```
[68] 1 len(df_mba[df_mba['customer_segmentation'] == "qualified customer"]) / len(df_mba)
[68] ✓ 0.5s
```

```
[69] ... 0.571636745106454
```

DATA ANALYSIS

This is the final result before we need to analyze the profile of each customer group. In this table, we will analyze the customer profile based on the channel chosen by the person to receive food, the place where the person placed the order (OrderFrom), the value of each order (SalesAmount), the gender of the customer (CustomerGender), whether the customer used a voucher (VoucherStatus), and the province of the customer (Province).

	Channel	OrderFrom	SalesAmount	CustomerID	CustomerGender	VoucherStatus	Province	customer_segmentation
0	Take Away	CALL CENTER	296891	1753863	Unknown	No	Ho Chi Minh City	qualified customer
1	Take Away	STORE	301782	1124050	Unknown	No	Hanoi	qualified customer
2	Take Away	WEBSITE	319792	1626827	Male	No	Hanoi	unqualified customer
3	Take Away	STORE	424762	125643	Male	No	Hanoi	unqualified customer
4	Delivery	STORE	280031	2117237	Unknown	No	Hanoi	qualified customer
...
1048570	Delivery	STORE	178107	1398136	Unknown	No	Southern Provinces	unqualified customer
1048571	Take Away	STORE	331447	2023258	Unknown	No	Hanoi	unqualified customer
1048572	Take Away	WEBSITE	381509	2041510	Female	Yes	Southern Provinces	qualified customer
1048573	Take Away	STORE	508205	1927215	Unknown	Yes	Hanoi	qualified customer
1048574	Delivery	CALL CENTER	349016	324317	Male	No	Hanoi	qualified customer

DATA ANALYSIS

The table on the right shows the profile of each customer group (qualified customer and unqualified customer) based on the mentioned characteristics. Here, we only need to pay attention to three metrics: support, lift, and confidence. The higher these metrics are, the higher the level of confidence. However, this table has also filtered out unreliable results, so most of the results here are reliable.

		antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
17		(Channel_Delivery)	(customer_segmentation_qualified customer)	0.477713	0.571637	0.324854	0.680019	1.189601	0.051776	1.338717	0.305161
32		(Channel_Take Away)	(customer_segmentation_unqualified customer)	0.499869	0.428363	0.270657	0.541456	1.264011	0.056531	1.246634	0.417625
60		(OrderFrom_STORE)	(customer_segmentation_unqualified customer)	0.503430	0.428363	0.268729	0.533795	1.246128	0.053078	1.226150	0.397757
98		(CustomerGender_Unknown)	(customer_segmentation_unqualified customer)	0.604668	0.428363	0.311817	0.515684	1.203846	0.052800	1.180296	0.428322
139		(sales_segmentation_3)	(customer_segmentation_qualified customer)	0.249998	0.571637	0.215623	0.862498	1.508822	0.072715	3.115328	0.449641
140		(sales_segmentation_4)	(customer_segmentation_qualified customer)	0.250002	0.571637	0.243274	0.973088	1.702283	0.100364	15.916898	0.550073
143		(sales_segmentation_1)	(customer_segmentation_unqualified customer)	0.249997	0.428363	0.222479	0.889926	2.077502	0.115389	5.193188	0.691535
269		(VoucherStatus_No, Channel_Delivery)	(customer_segmentation_qualified customer)	0.423056	0.571637	0.278107	0.657376	1.149989	0.036273	1.250244	0.226065
373		(Channel_Take Away, OrderFrom_STORE)	(customer_segmentation_unqualified customer)	0.363542	0.428363	0.217311	0.597761	1.395453	0.061583	1.421136	0.445256
424		(CustomerGender_Unknown, Channel_Take Away)	(customer_segmentation_unqualified customer)	0.349850	0.428363	0.213741	0.610949	1.426241	0.063878	1.469311	0.459672
462		(VoucherStatus_No, Channel_Take Away)	(customer_segmentation_unqualified customer)	0.463942	0.428363	0.263073	0.567039	1.323734	0.064338	1.320296	0.456222
598		(CustomerGender_Unknown, OrderFrom_STORE)	(customer_segmentation_unqualified customer)	0.385801	0.428363	0.226689	0.587580	1.371685	0.061426	1.386053	0.441176
646		(VoucherStatus_No, OrderFrom_STORE)	(customer_segmentation_unqualified customer)	0.495508	0.428363	0.267365	0.539577	1.259626	0.055107	1.241548	0.408556
822		(VoucherStatus_No, CustomerGender_Unknown)	(customer_segmentation_unqualified customer)	0.589821	0.428363	0.309122	0.524095	1.223482	0.056465	1.201157	0.445320
964		(VoucherStatus_No, sales_segmentation_4)	(customer_segmentation_qualified customer)	0.209088	0.571637	0.203675	0.974116	1.704081	0.084153	16.549088	0.522401
969		(VoucherStatus_No, sales_segmentation_1)	(customer_segmentation_unqualified customer)	0.247869	0.428363	0.220907	0.891227	2.080541	0.114730	5.255334	0.690512
1549		(VoucherStatus_No, Channel_Take Away, OrderFro...)	(customer_segmentation_unqualified customer)	0.358388	0.428363	0.216272	0.603456	1.408749	0.062751	1.441547	0.452221
1760		(VoucherStatus_No, CustomerGender_Unknown, Cha...	(customer_segmentation_unqualified customer)	0.342995	0.428363	0.212275	0.618886	1.444769	0.065348	1.499909	0.468562
2074		(VoucherStatus_No, CustomerGender_Unknown, Ord...	(customer_segmentation_unqualified customer)	0.382917	0.428363	0.226136	0.590563	1.378650	0.062109	1.396152	0.445082

DATA ANALYSIS

From the results on the right, we can see that the channel 'OrderFrom,' customer gender, and province typically do not have a significant impact on the customer profile. For unqualified customers, they tend not to use vouchers, have very small transaction values, prefer take away, and their gender is often undefined, with orders typically placed at the store. For qualified customers, it can be observed that they are less likely to use vouchers, their transaction values are typically large, and they tend to use delivery.

SUMMARY

The analysis report has employed various analytical methods to identify the profiles of two customer groups: qualified customers and unqualified customers. Firstly, the report used cohort analysis to calculate customer retention rate and churn rate, leading to the determination of the average customer lifetime (ACL). Then, the article sequentially calculated average order value (AOV) and average purchase frequency (APF). Subsequently, the article utilized customer segmentation methods to classify customers into two main groups. Finally, the report employed market basket analysis to analyze the characteristics that tend to co-occur within each customer group, thereby establishing the customer profiles.

On the ratio between the two customer groups, the qualified customer group has a 50:50 ratio compared to the unqualified customer group.

Unqualified customers are characterized by not using vouchers, having low transaction values, and tending to choose take away. The gender of this customer group is often undefined, and they have a tendency to order at the store. Qualified customers, on the other hand, do not use vouchers, have high transaction values, and typically choose delivery as the delivery method.

SUMMARY

With the business's campaign over the past year, it can be observed that the customer capital recovery rate is 50%, meaning that only one out of every two new customers manages to recover their capital. This is a relatively low rate, indicating that the business is facing a significant challenge in identifying and satisfying customer needs. Therefore, this report has identified customer groups and their profiles to help the business enhance its strengths and address its weaknesses. It is evident that the business's voucher policy has not been effective, possibly due to its inability to attract customers sufficiently. To improve the services for in-store ordering and take away, while also boosting the development of delivery channels.