

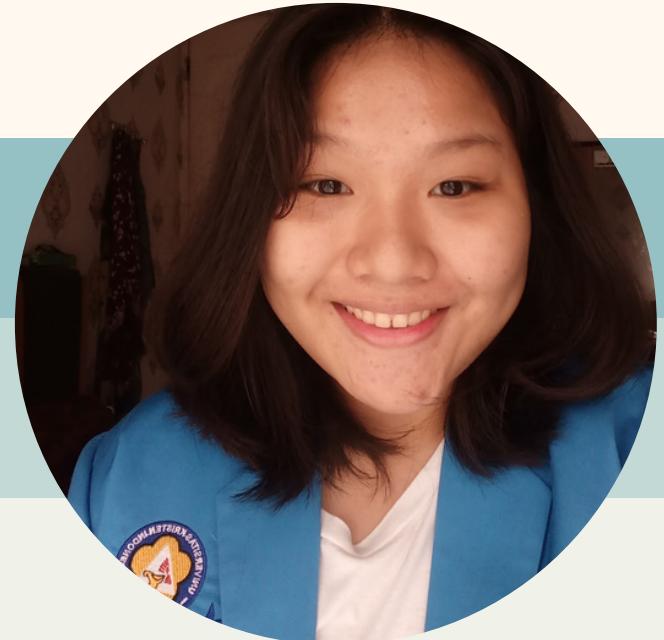
SUPERSTORE

RFM

ANALYSIS

By Team 2

MEMBER OF TEAM 2



**Eva Priscilla
Simorangkir**



**Liana Indah
Sari**



**Nisrina Syifa
Syauqiyah**



**Putri Nur
Alifah**



**Rizky Eka
Purnama**

Outline

- **The Objective**
- **Data Preparation**
- **Analysis & Finding**
- **Conclusion**



THE OBJECTIVE

BUSINESS UNDERSTANDING

This is a sales transaction data set of UK-based e-commerce (online retail) for one year. This London-based shop has been selling gifts and homewares for adults and children through the website since 2007. Their customers come from all over the world and usually make direct purchases for themselves. There are also small businesses that buy in bulk and sell to other customers through retail outlet channels.



BUSINESS UNDERSTANDING

- How was the sales trend over the months?
- What are the most frequent purchased products?
- How many products does the customer purchase in each transaction?
- How does each cluster/group contribute to the company's revenue?
- What are the interval date between transactions?
- Based on your findings, what strategy could you recommend to the business to gain more profit?

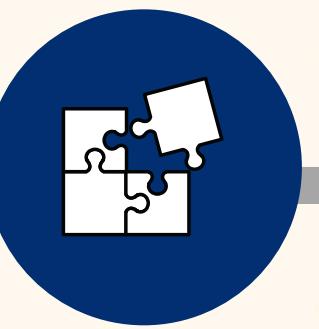


METHODOLOGY & CHALLENGES



METHODOLOGY

- Data Preparation
- Data Preprocessing
- RFM Model Building
- Clustering
- Predictive Model Building
- Exploration Data Analysis



CHALLENGES

- Huge amount of data
- There is a missing value
- Ambiguous meaning of columns of features

A photograph showing a black electronic calculator, a white notepad, and a silver and black pen resting on a light-colored wooden desk. The calculator has a red 'ON/C' button and a grey '0' button. The notepad is open to a blank page.

DATA PREPARATION

Initial Data Quality Check

Total Data



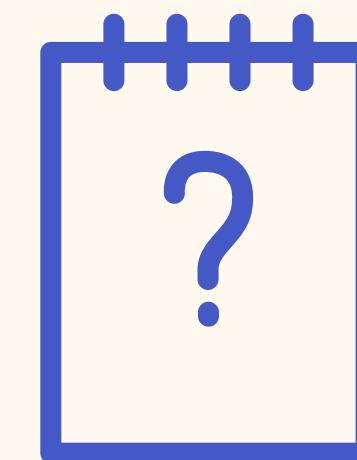
541,909
Data Entries

Total Columns



8 Initial
Columns

Missing Values

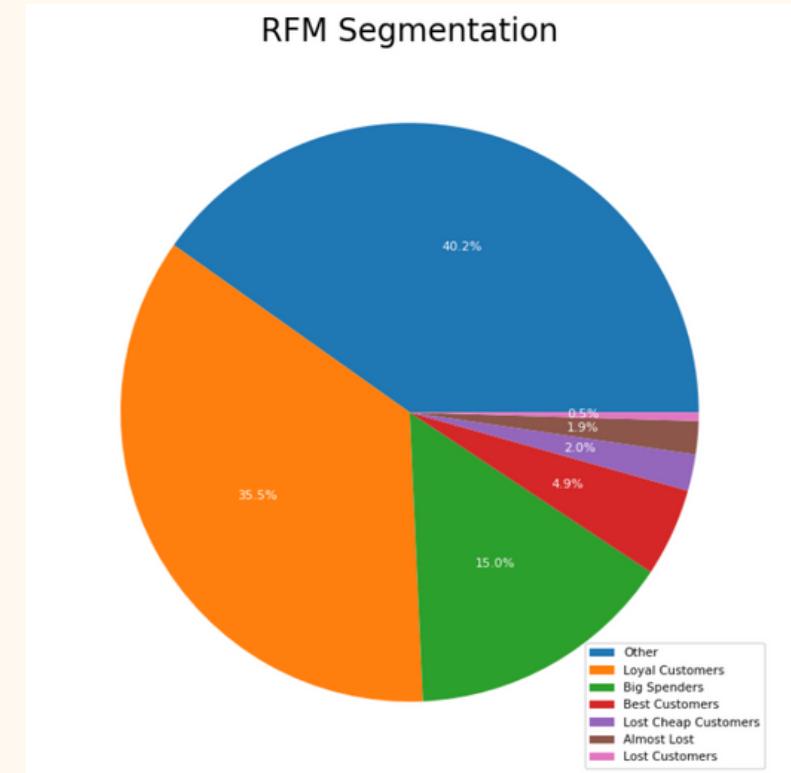
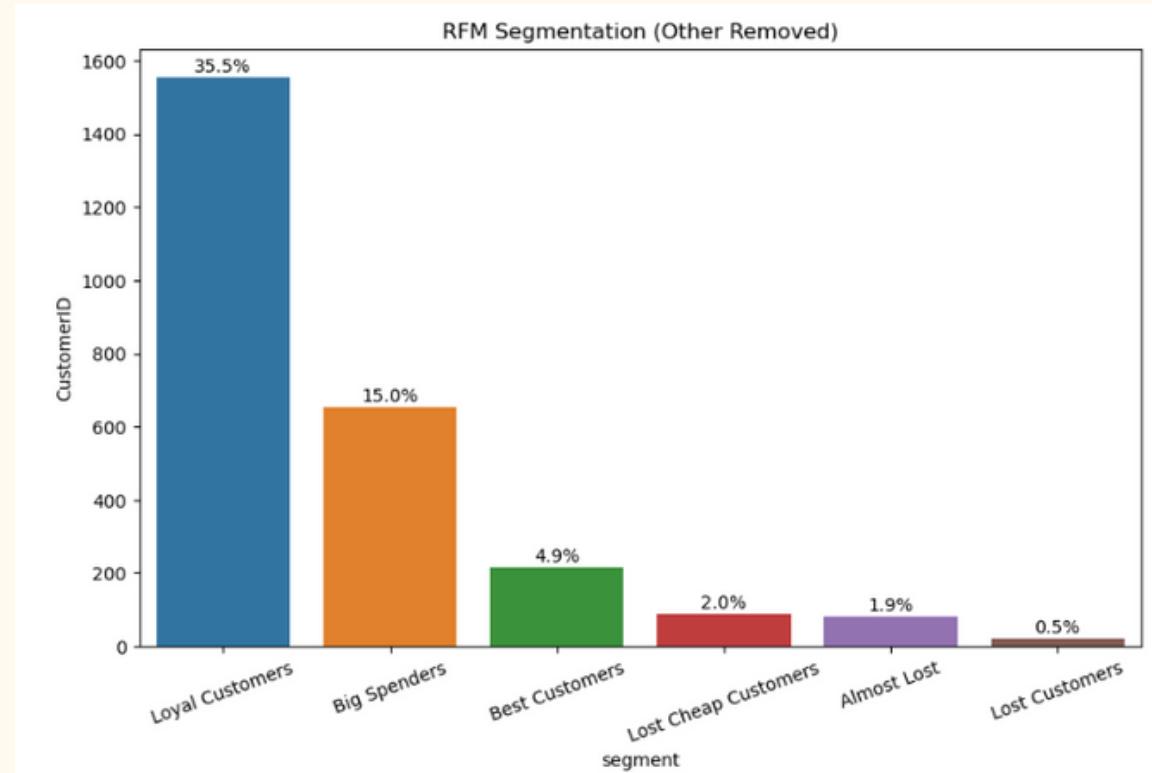


135,080
missing entries
on customer ID

1,454 missing
entries on
description

R F M

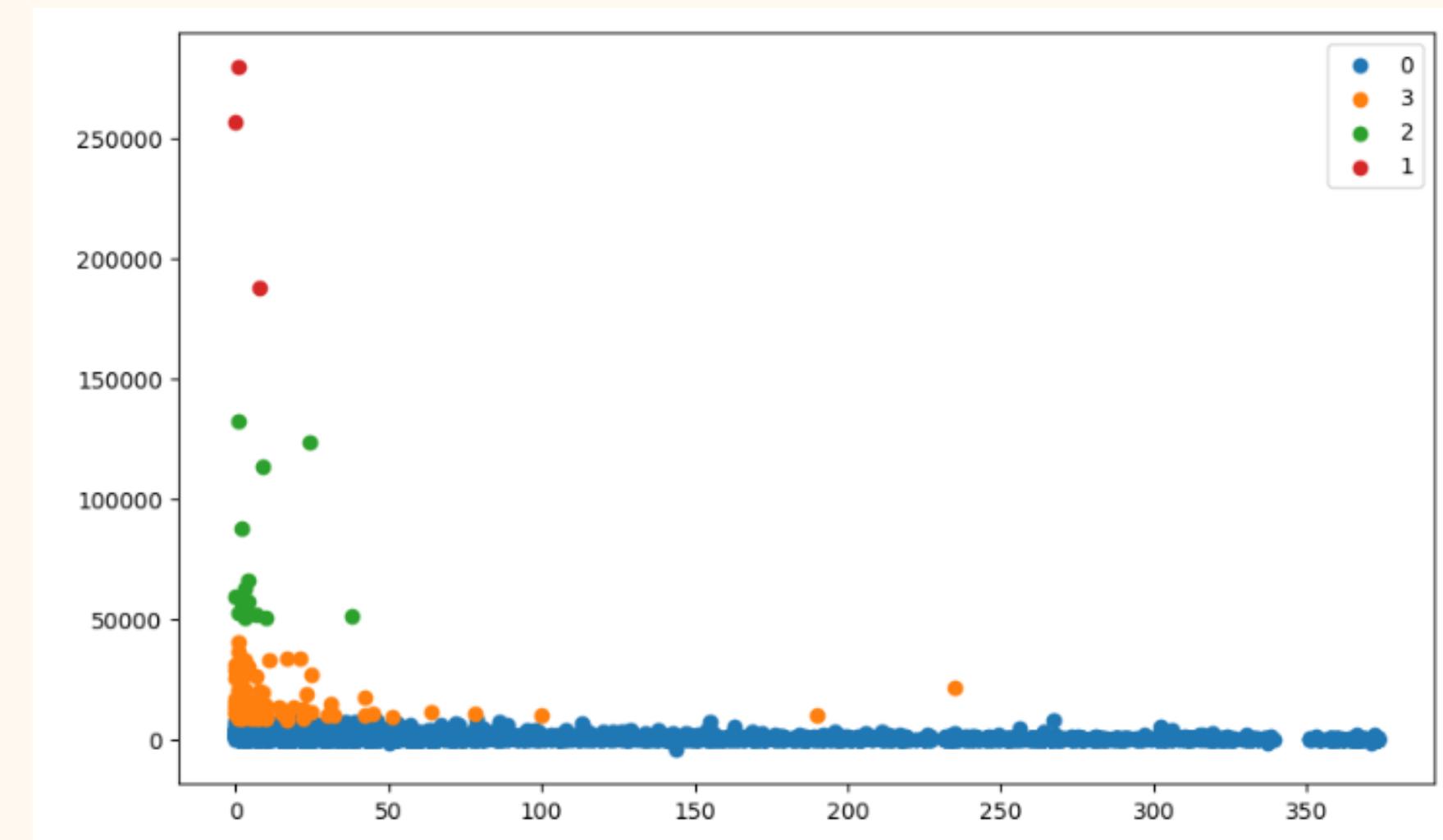
M O D E L



The rfm score's number and segments are visualized, and the results are as follows:

1. The percentage of "Other" segments is still about 40%; segmentation can still be used to gain more segment insights.
2. The 'Big Spender' and 'Loyal Customers' groups have the highest numbers following that, indicating that there are still many customers, with a total percentage of 50%. The number of customers in this group can be increased with the use of special promotional strategies.
3. The percentage of "Best Customers" is at 4.9%. In order to safeguard this customer, it is preferable to implement a loyalty program.
4. If "Lost Cheap Customers," "Lost Customers," and "Almost Lost" are included as churn customers, then 4.4% of all customers have been lost, which is still a respectable percentage. Retargeting for these clients is an option if necessary.

C L U S T E R I N G



Clustering based on recency and monetary

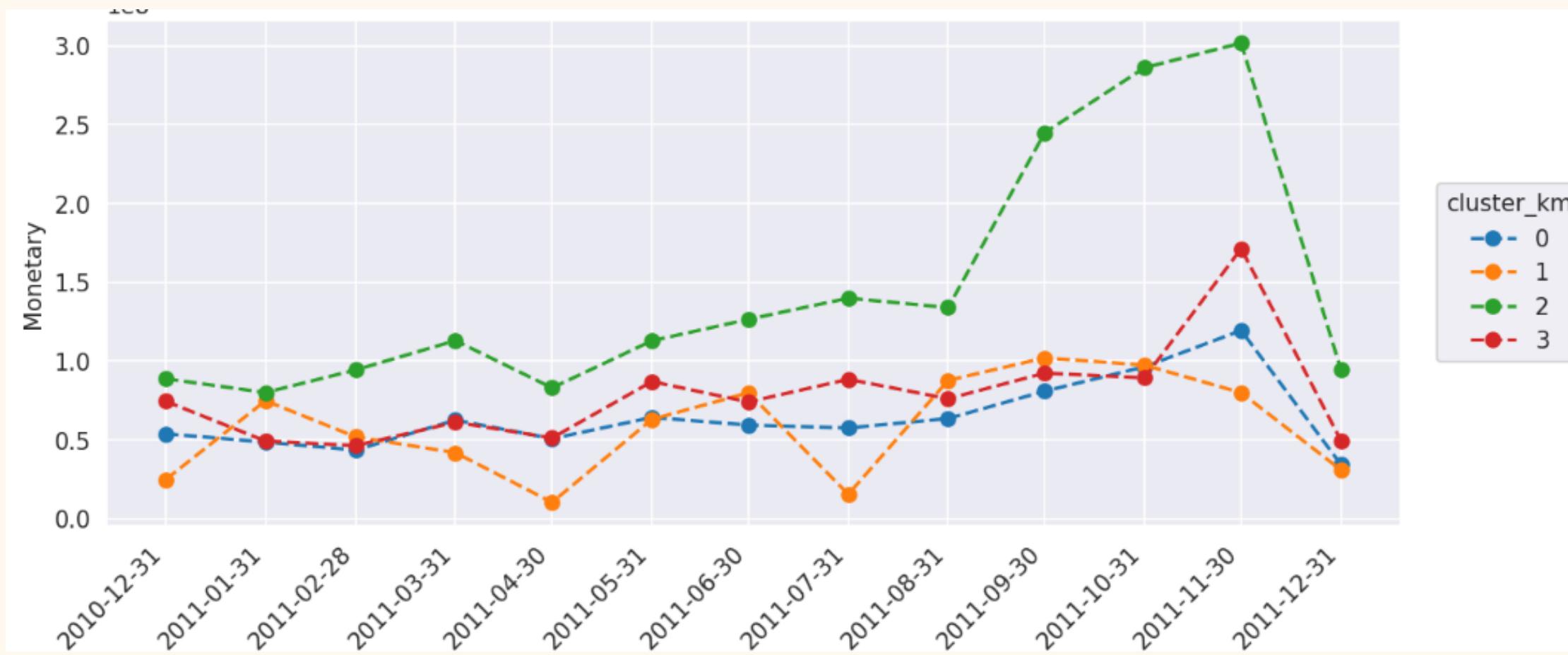
From the graph above it can be concluded that:

- Cluster 0 is a type of customer that has low monetary and various recencies.
- Cluster 1 is a type of customer who has moderate monetary and recency is between 0-250.
- Cluster 2 is a type of customer who has a fairly high monetary level of around 50000 - 140000 and recency is below 50.
- Cluster 3 is a type of customer who has high monetary and recency below 50.

A photograph of a black electronic calculator with a red 'ON/C' button, a silver pen, and a white notebook with a yellow cover, all resting on a light-colored wooden desk.

ANALYSIS & FINDING

HOW WAS THE SALES TREND OVER THE MONTHS?



There are some interesting insights from month and totalcost table, the highest total cost occurred at 1132407.74 Poundsterling at November 2011 that's probably caused by 11.11.11 promos. and then followed by next month lowest total cost occurred at 342506.38 Poundsterling at December 2011

HOW DOES EACH CLUSTER/GROUP CONTRIBUTE TO THE COMPANY'S REVENUE?

From the graph and previous clustering data, we could assume:

- cluster 1 : as a total has the lowest total cost throughout time change
- cluster 2: as a total has the highest spend total cost throughout time change
- cluster 3 as a total has the second highest spend total cost throughout time change
- cluster 0 as a total has the third highest spend total cost throughout time change.

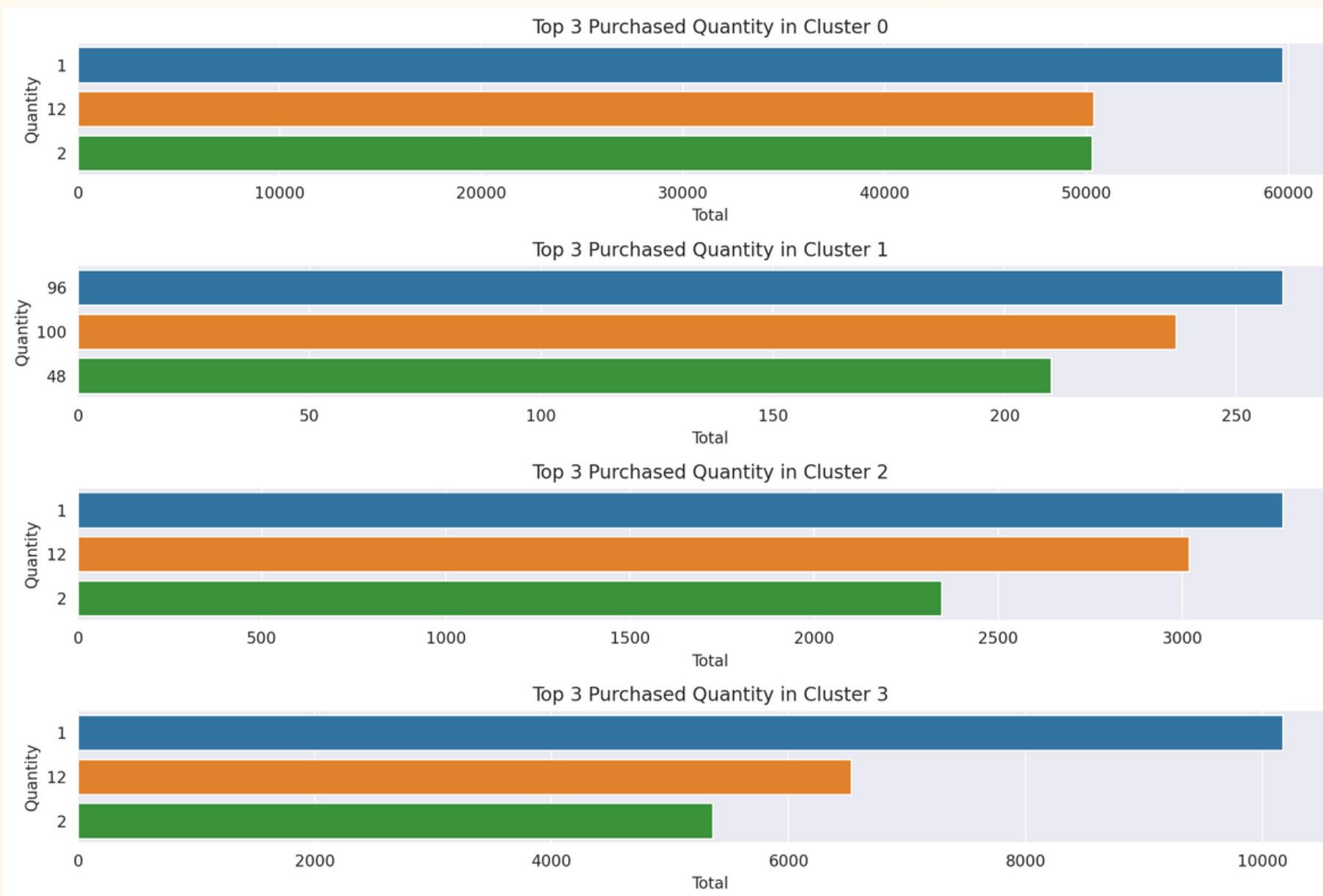
And if we relate the totalcost with the numberof person of each cluster, we could see 3 customers in cluster 1 has critical value to the company

WHAT ARE THE MOST FREQUENT PURCHASED PRODUCTS?

cluster_km	Quantity	Description
0	80995	PAPER CRAFT , LITTLE BIRDIE
0	74215	MEDIUM CERAMIC TOP STORAGE JAR
0	60	PICNIC BASKET WICKER 60 PIECES
2	1	POSTAGE
1	1412	SET OF TEA COFFEE SUGAR TINS PANTRY
...
2	-1	POSTAGE
0	-2	Manual
0	-1	Manual
0	-74215	MEDIUM CERAMIC TOP STORAGE JAR
0	-80995	PAPER CRAFT , LITTLE BIRDIE

The top 3 of the product that ordered are paper craft, little birdie; medium ceramic top storage jar; picnic basket wicker 60 pieces. But 2 of them menium ceramic top storga jar anda papaer craft little birdie also the most cancelled and done by same cluster 0 customer that bought a whole bunch of the same products

HOW MANY PRODUCTS DOES THE CUSTOMER PURCHASE IN EACH TRANSACTION ?



From the graph we could identify

- cluster 0 tend to purchase item in low quantities (1,2,12), but with very high repeats.
- cluster 1 tend to buy product quantity in high amount (48, 100, 96) with low number of transaction.
- Cluster 2 tend to buy product in low quantities and high repeats (1,2,12)
- Cluster 3 tend to but in low quantities and high repeats (1,2,12)

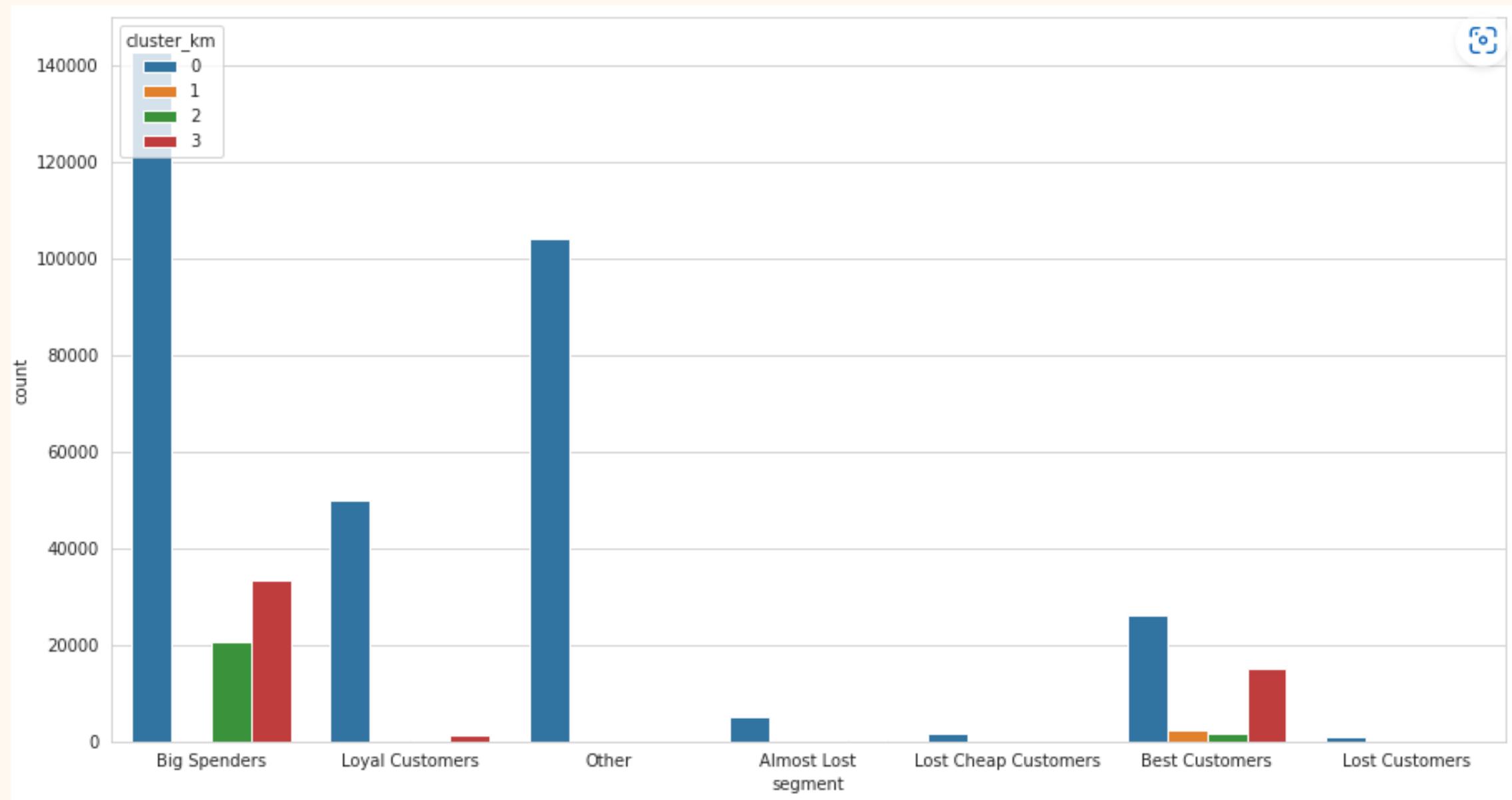
WHAT ARE THE INTERVAL DATE BETWEEN TRANSACTIONS?

InvoiceNo	CustomerID	cluster_km	recency
540247	15464.0	0	53066
540458	12501.0	0	46935
542107	16222.0	0	43566
543040	17337.0	0	39109
537144	15880.0	0	38745
...
572885	15311.0	2	0
572881	12748.0	3	0
572835	17428.0	3	0
572834	12748.0	3	0
C581569	17315.0	0	0

cluster_km
0 373
1 8
2 38
3 235

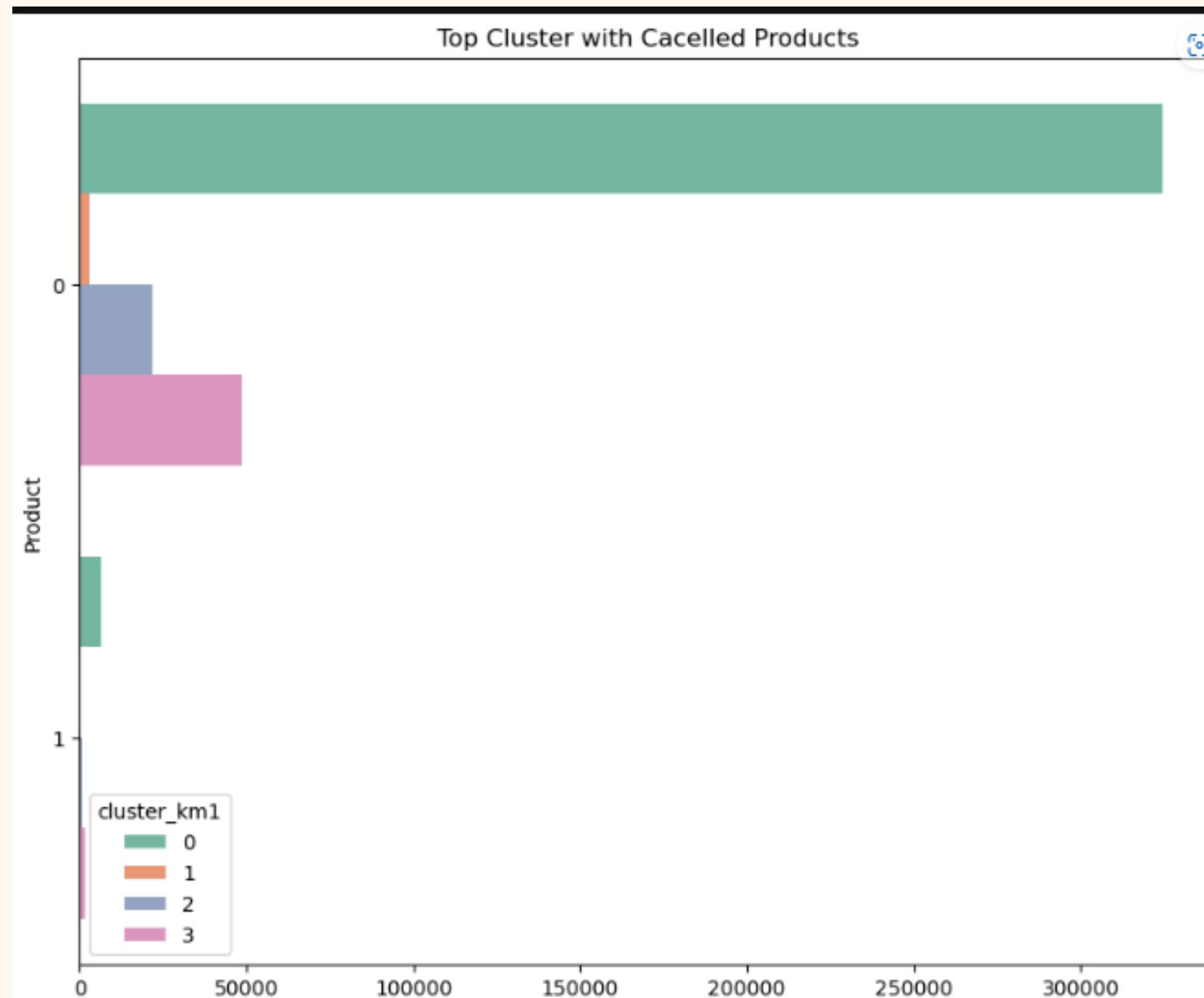
From the calculation above it can be seen that cluster 0 has the largest interval value, namely 373, the second order is cluster 3 of 235, the third order is cluster 2 of 38, the last is cluster 1 of 8.

Clusters and Customer Segmentation



- cluster 0: distributed along all segmented region, but the tops segment are big spender, loyal customers, and others
- cluster 1: mostly best customer and a bit of big spenders
- cluster 2 :mostly big spenders and a bit best customer, but with more count of people
- cluster 3: mostly big spenders and best customer, but with more count of people

Clusters and Cancellation



The cancellation concentrated on cluster 0 followed by clusters 3,2, and 1.

Cluster	Cancelled_sum
0	6634
3	1571
2	679
1	21

WHAT STRATEGY COULD YOU RECOMMEND TO THE BUSINESS TO GAIN MORE PROFIT?

Cluster 0

We could offer a discount on the top 3 most popular items, such as the T-Heart Holder, Regency cakestand 3 tier, and big bag retrospot, for cluster 0 customers who have a high cancellation rate and make up the majority of the population.

Cluster 1

For cluster 1 (The Best Customers) with high quantity purchases, high total costs, and high recency, we could create bundle promotions to motivate them and loyalty points for their regular purchases and large spends; the bundling/discount includes the top 3 products, which are the spaceboy lunch box, round snack box set of 4 boxes, and postage.

Cluster 2

The top 3 things for cluster 2 (the high-spending and a few of the best customers) that we could discount are the regency cakestand with 3 tiers, the gigantic bag in red retrospot, and the carriage.

Cluster 3

Regency Cakestand 3 Tier, Jumbo Bag Red Retrospot, and Lunch Bag Red Retrospot are the top 3 products we could discount for Cluster 3 (The Big Spenders and Some of the Best Customers).

CONCLUSION

372

Cluster 0 has the largest date interval value

6634

Cluster 0 is the cluster that cancels the most

263

based on the average recency per segment, the lost cheap customer segment is the segment with the highest average recency

Top 3

The most purchased products are White Hanging Heart T-Light Holder, Recency Cakestand 3 Tier, Jumbo Bag Retrospot

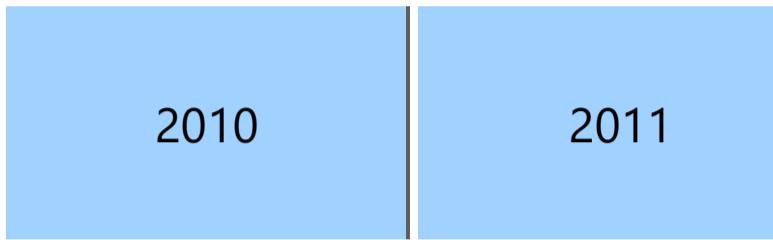


**THANK
YOU**



SUPERSTORE DASHBOARD

Year



Month

Select all

January

February

March

April

May

June

July

August

September

October

November

December

Cluster KM

All

Segment

All

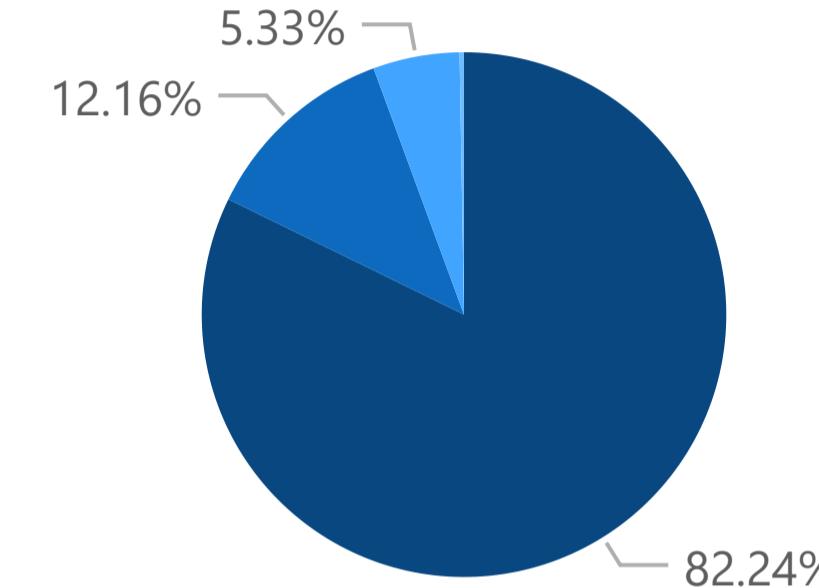
Frequency

1285M

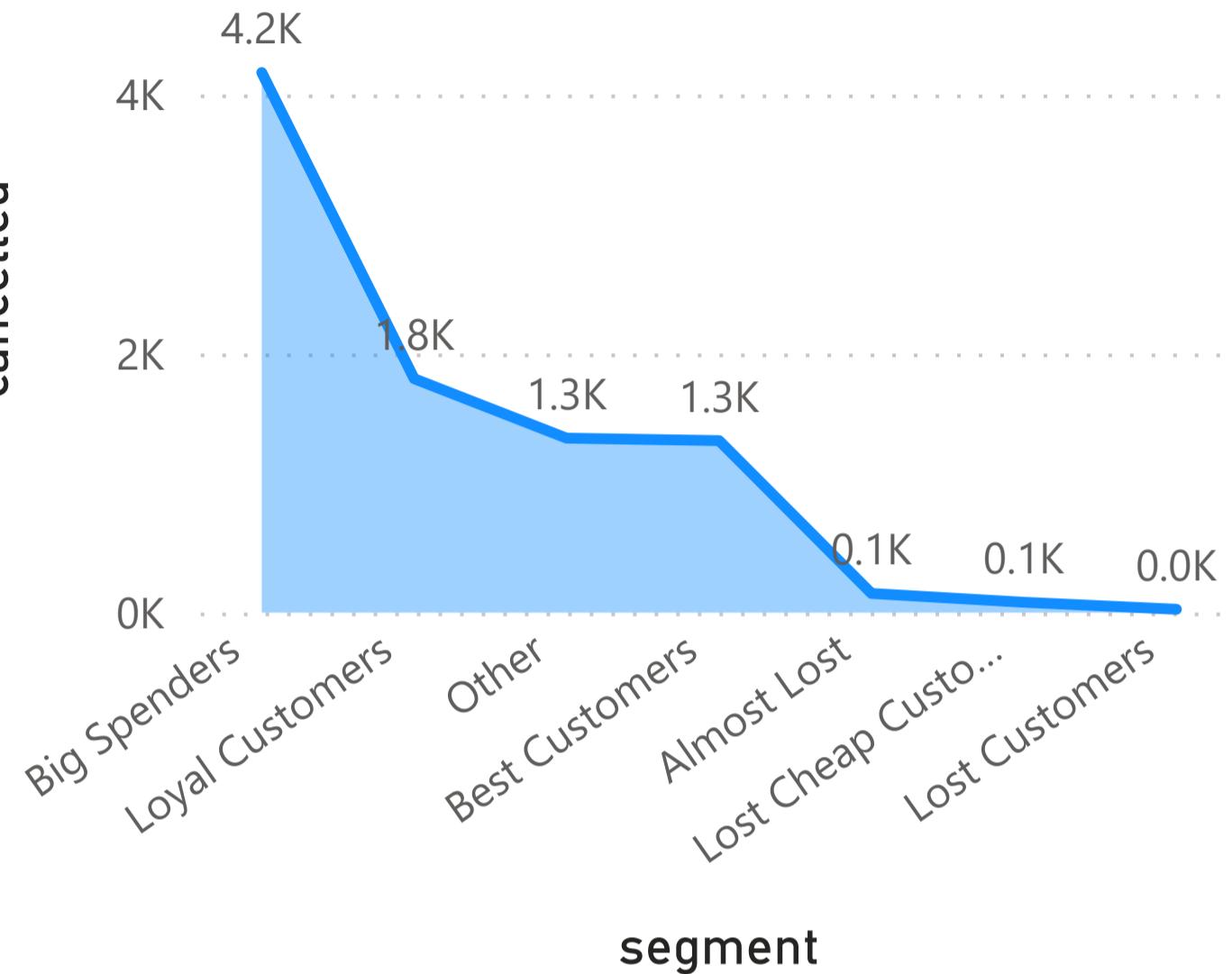
Quantity

4907K

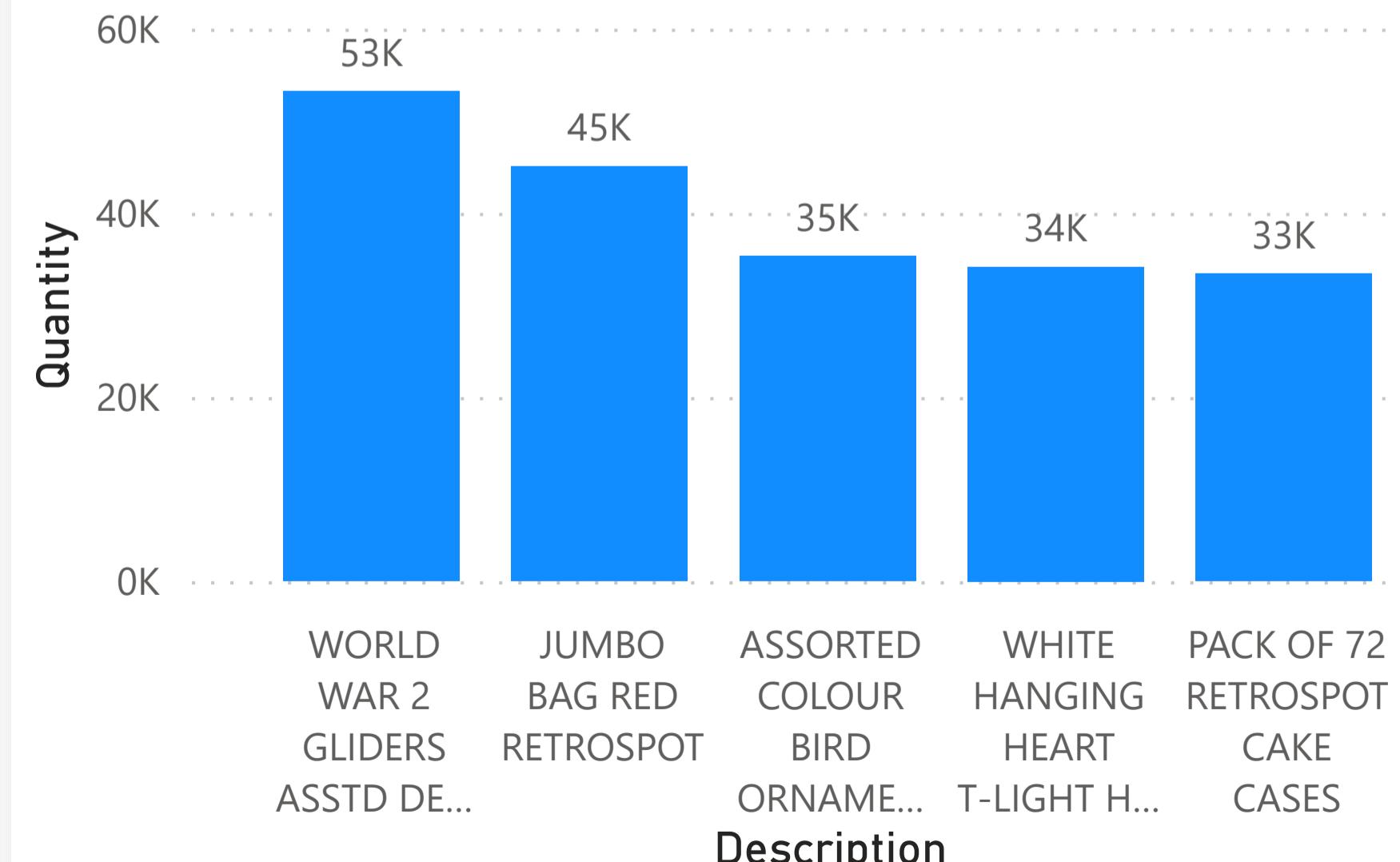
frequency by cluster



cancelled by segment



Top 5 Most Purchased Product



Average of recency by segment

