DATA SCIENCE PORTOFOLIO

Bayu Triona Asri
Data Scientist | Data Enthusiast

Data engineer Intern - Growlab (March - May 2024)
Administration staff - PPAT (June 2021 - September 2022)
Administration staff - Cafeteria and Internet (April 2015 - May 2021)

Skills

- SQL Database
- Python
- Data Visualization
- Machine Learning

Tools

tableau

jupyter

Google Big Query

python

Looker
PART OF Google Cloud

Data Science Bootcamp
Dibimbing.id

Fullstack Data Analysis
My Skill

Mini Course Data Analytics
RevoU

Internship Certificate
Growlab

01 Beecycle Product Analysis
Tableau

Built Beecycle sales dashboard using Tableau to visualize sales Performance

View on Tableau

02 Superstore sales performance Dashboard
Looker Studio

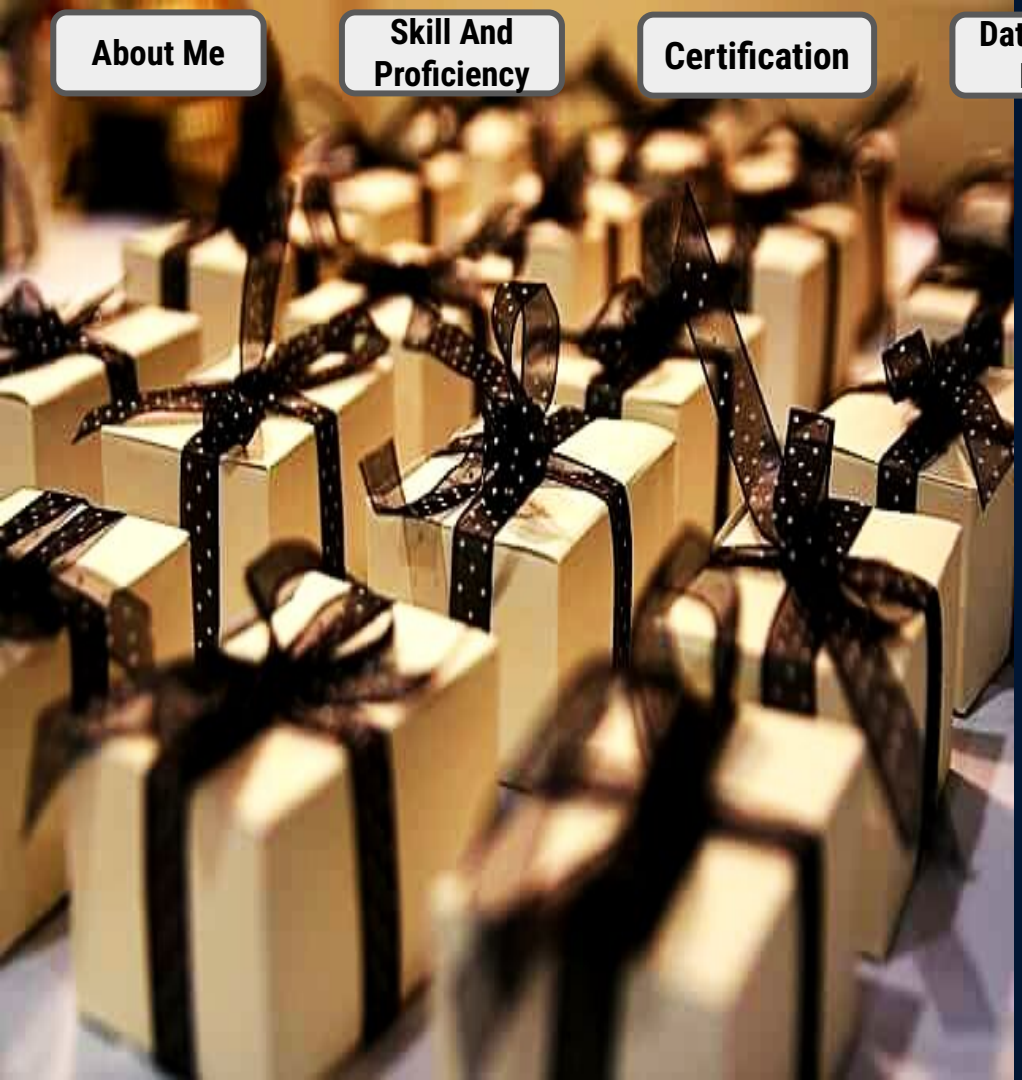
Built Superstore sales Performance dashboard using Looker Studio to visualize sales Performance

View on Looker

03 Hotel Booking Cancelation
Python

This project aim is to understand/identify the pattern or cause of hotel cancelation by using historical data.

View on Github



Customer Segmentation using RFM metrics and K-Means Model Algorithm

By
Bayu Triona Asri

View on Github

Online Retail Dataset

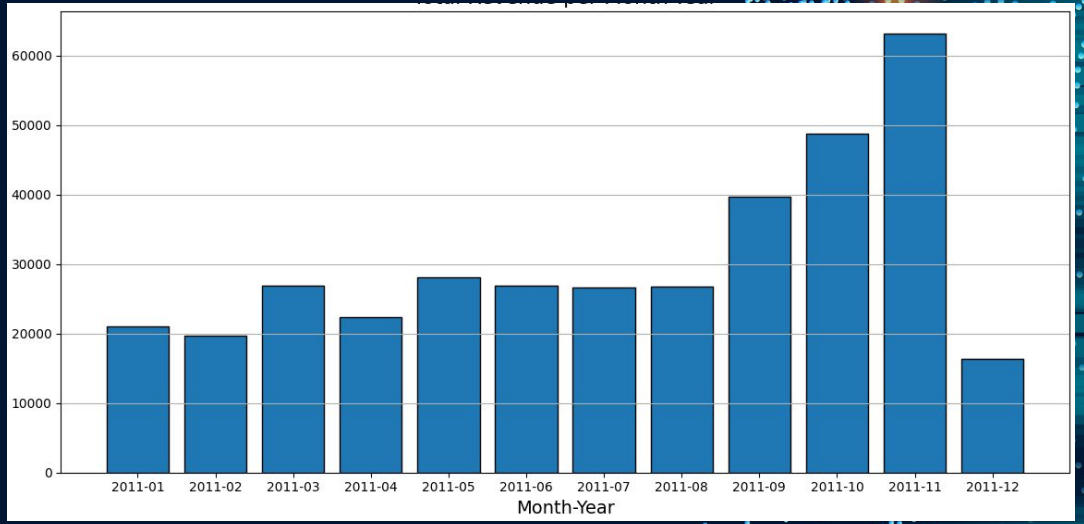
Source :
https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci?select=online_retail_II.csv

Content :

This Online Retail II data set contains all the transactions occurring for a UK-based and registered, **non-store online retail between 01/12/2009 and 09/12/2011**.The company mainly sells unique all-occasion giftware. Many customers of the company are wholesalers.

Dataset
Invoice
StockCode
Description
Quantity
InvoiceDate
Price
Customer ID
Country

The Company is UK-based and registered non-store online retail that selling unique gift for various occasion, such as gift, weddings and holidays. The Company notice the fluctuation in net income throughout the years, and **wanted the different marketing approach to increase their income.**



Goal of this project : to understand customer segments and behavior using RFM (Retency, Frequency, Monetary) and used the result analysis to devise appropriate business approaches and strategies to increase revenue and transaction in the future.

Input

Customer Transaction Behavior

Purchasing History

Process

Customer Segmentation (RFM and K-Means Model)

Output

Tailored Marketing Approach

01 Missing Value Handling

There some missing values found in dataset, 0.41% in Description column and 22.77% in Customer ID column. The rows with the missing values were dropped for the further analysis.

02 Duplicated Value Handling

There are 26.479 duplicated transaction that were dropped from the dataset

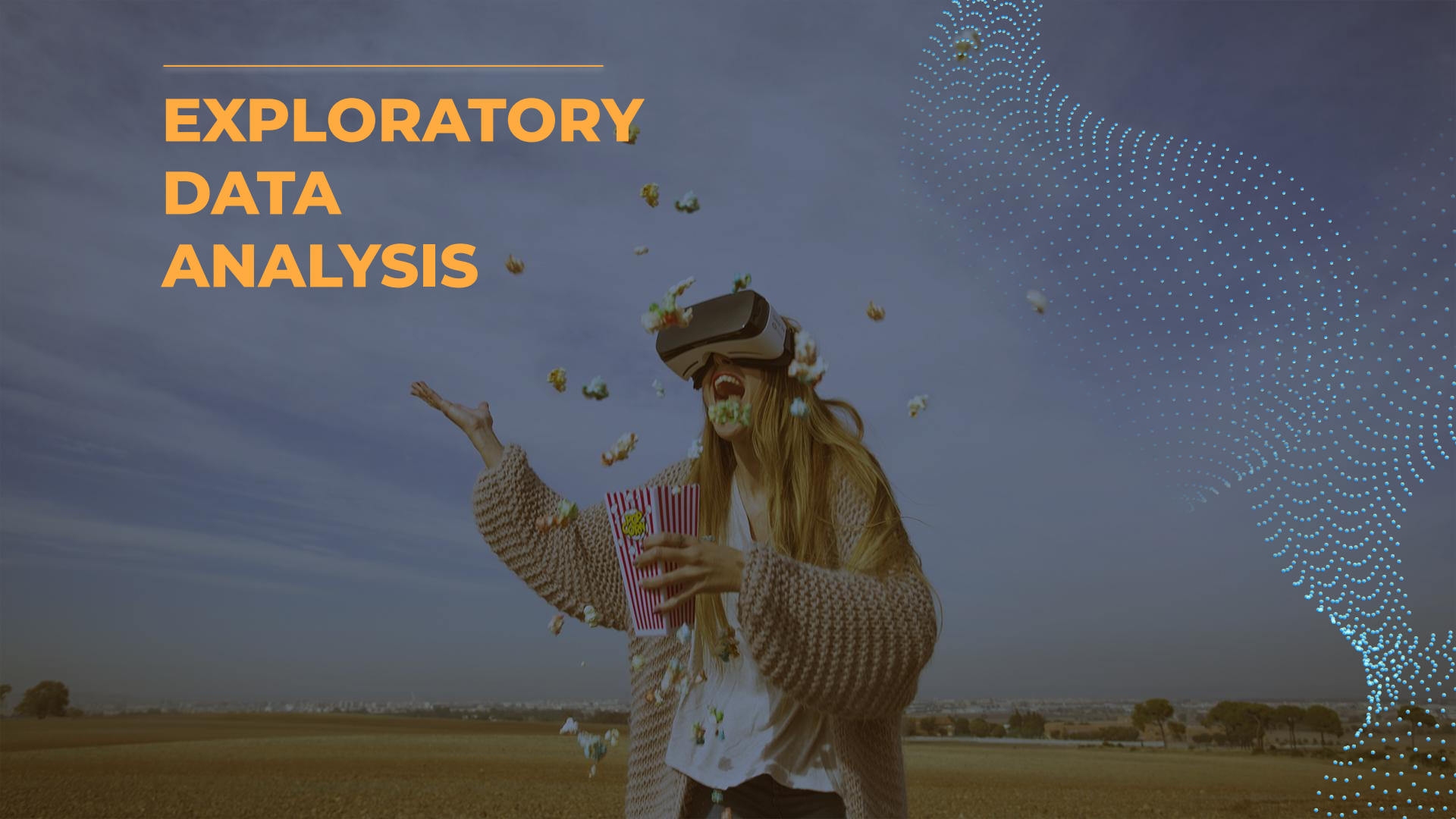
03 Change Data Type

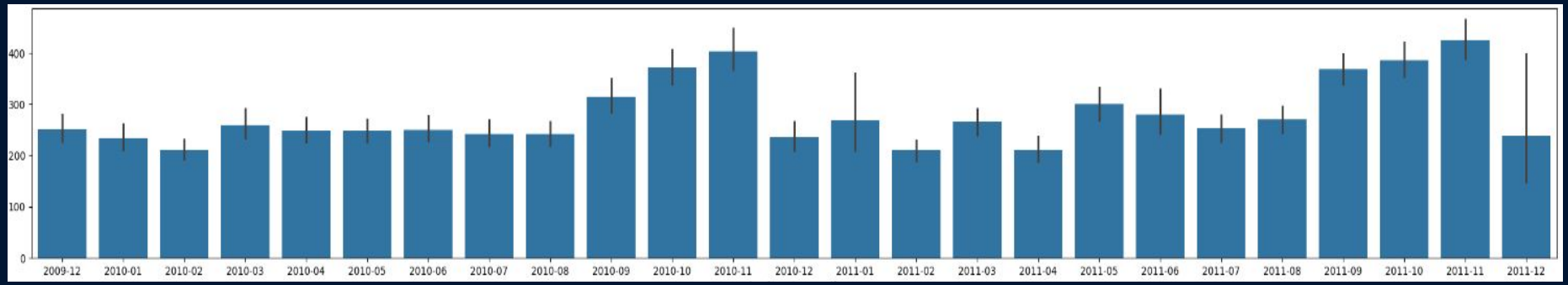
Convert InvoiceDate column into datetime datatype and Customer ID into int datatype.

04 Canceled Order Analysis

We found that there are negative values in quantity and price column, as expected that all the transactions with the negative values are canceled orders and all canceled orders were dropped.

EXPLORATORY DATA ANALYSIS

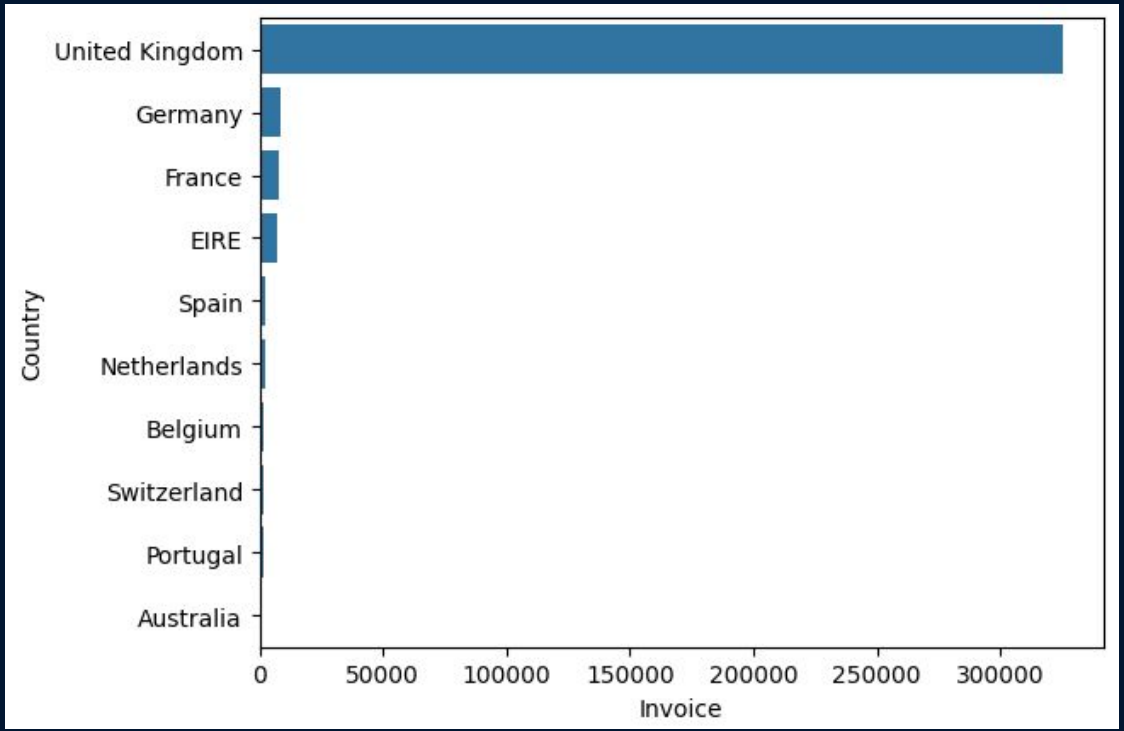




The following graph shows the revenue from this dataset. We cannot clearly and accurately identify the income pattern; therefore, we will take the transaction data from the last year for further analysis.

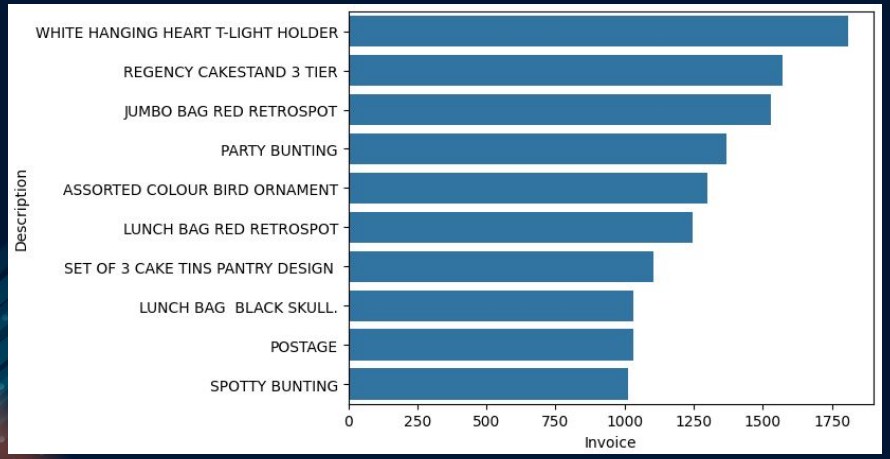
Monthly Income Distribution (January 2011 - December 2011

From the following graph, we can see that quite a fluctuating income at the beginning of the year, which start to increase towards the end of the year, reaching its peaks in November.



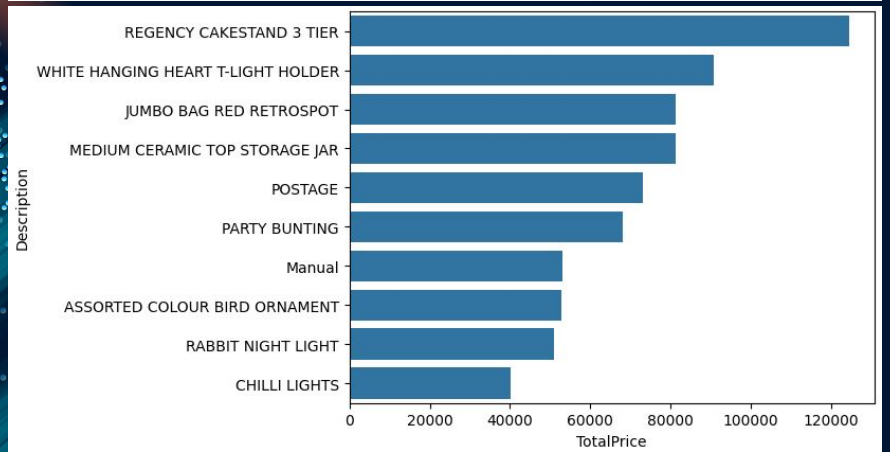
Top 10 Country to Order

Unsurprisingly, the United Kingdom made up the majority of the sales figure given that this is a store based in the UK, following by Germany and France.



Most Ordered Products

The barplot highlights the product with the highest order frequency. And White hanging Heart T-light holder become the most product that ordered by customer

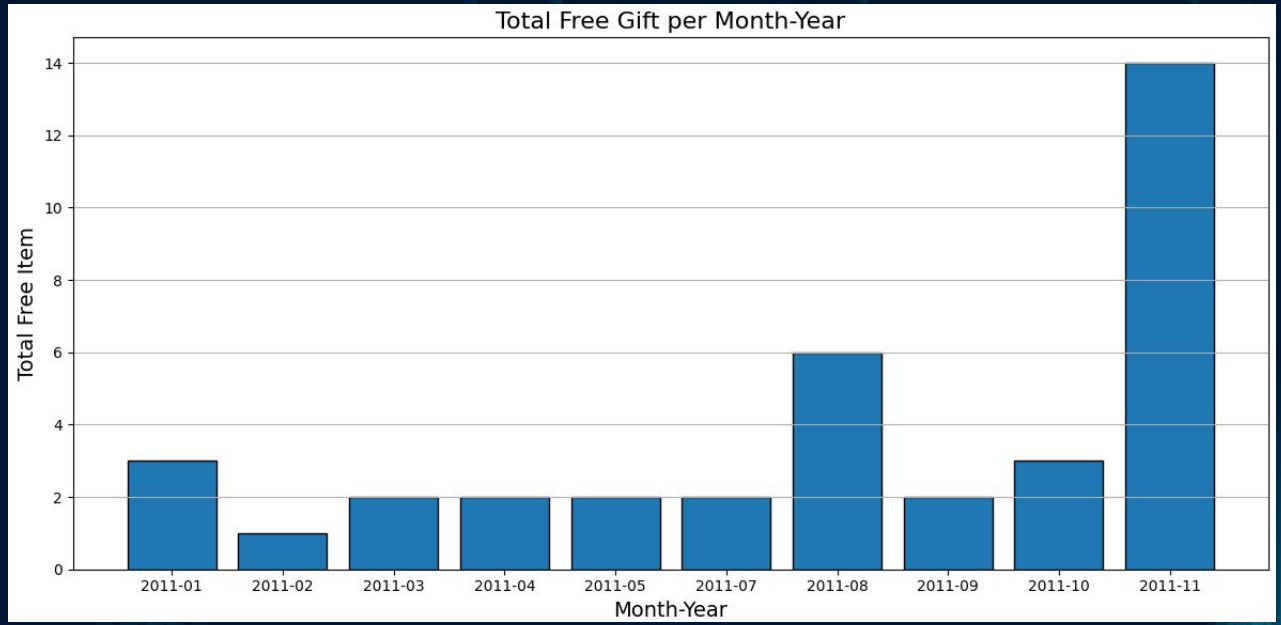


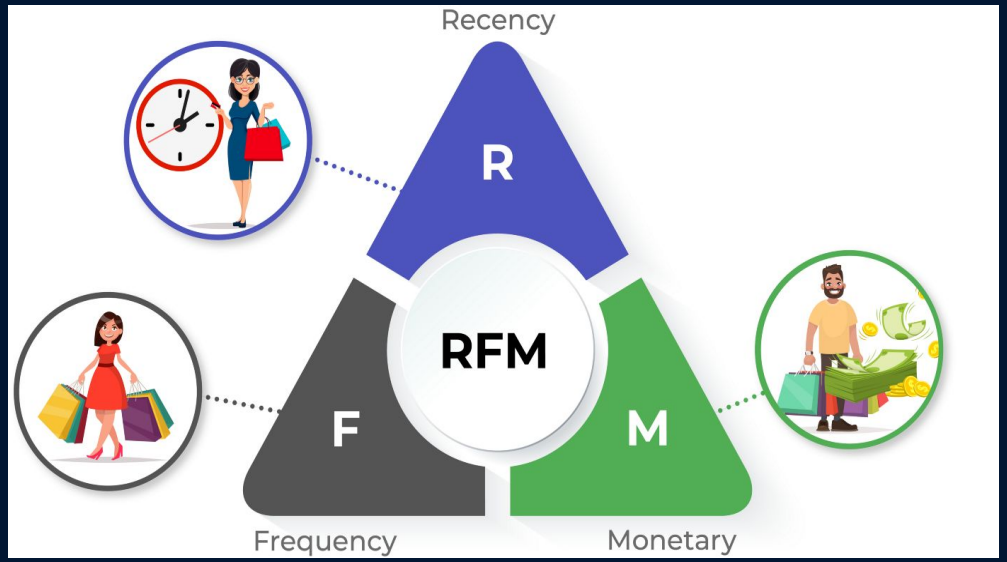
Most Profitable Products

Even though it is not most ordered product , cakestand remains the top product in terms of profitables products.

Monthly Free Items

As we can see that the company consistently giving free items for each month, on average the company giving 2 free items each month. There's significantly increase in November and according to sales data there's also increase.

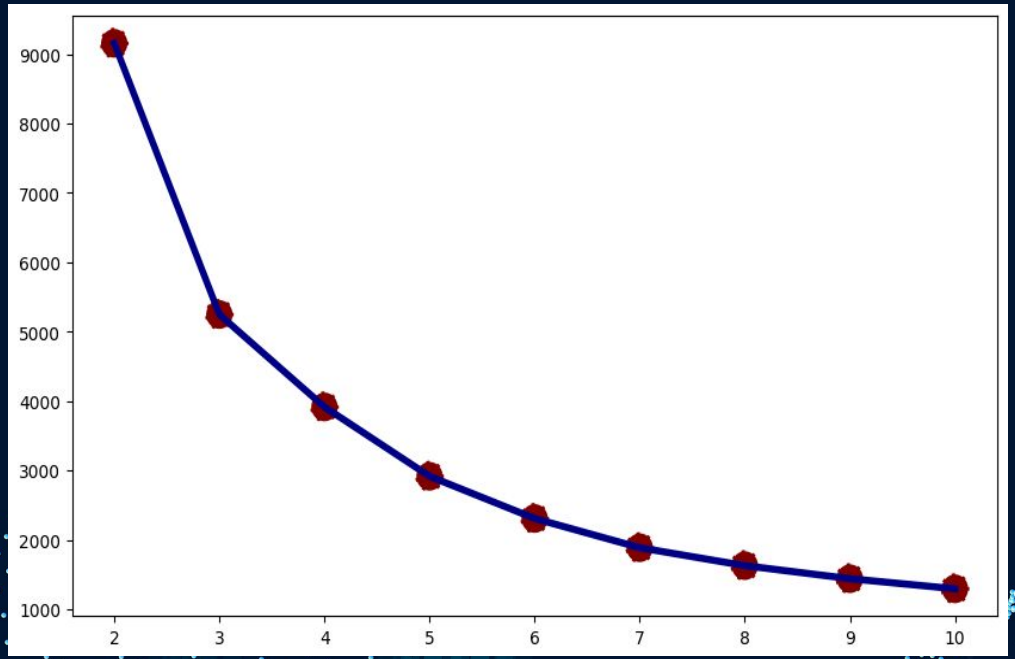




RFM (Recency, Frequency, Monetary) model is a behavior-based model used to analyze the behavior of a customer by measuring when people buy, how often the buy and how much they buy.

Recency (R) : When was the users most recent transaction?
Frequency (F) : How often does the customer transact?
Monetary (M) : How much the customer spend?

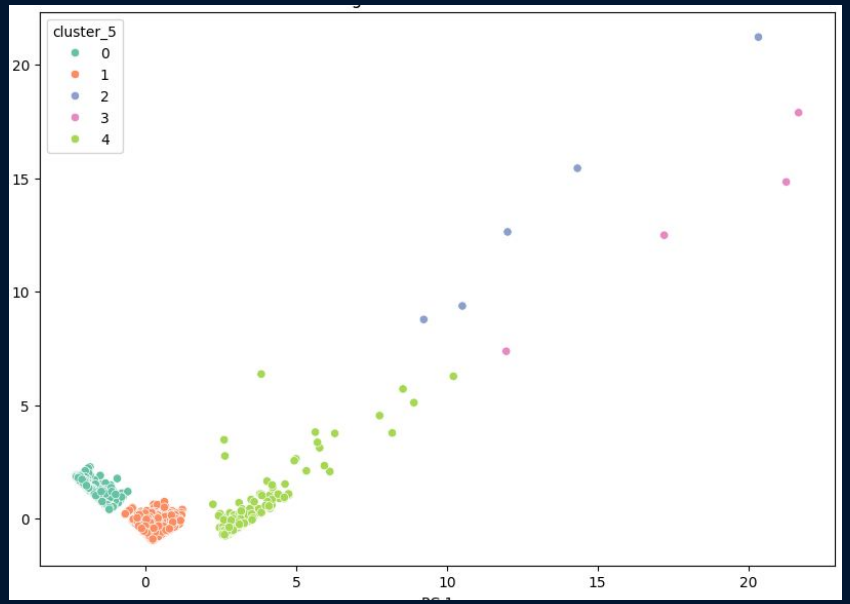
The Elbow Method



The Silhouette Score

```
For 2 The Silhouette Score is : 0.5775611218618004
For 3 The Silhouette Score is : 0.6022456786253381
For 4 The Silhouette Score is : 0.6061576893053512
For 5 The Silhouette Score is : 0.6091110245727241
For 6 The Silhouette Score is : 0.5875642588959503
For 7 The Silhouette Score is : 0.5607066937537608
For 8 The Silhouette Score is : 0.49639743860383184
For 9 The Silhouette Score is : 0.519524410218499
For 10 The Silhouette Score is : 0.4271396980239805
```

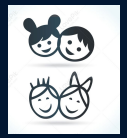
K = 5 obtain the highest silhouette score, for further analysis we're going to use K=5



	cluster_5	recency	frequency	monetary	NumBuyers
0	Hibernating	227.761811	26.482283	474.910985	1016
1	Promising	42.285955	70.563826	1302.946976	2969
2	Champion	9.600000	972.600000	187426.350000	5
3	Loyalist Customer	2.750000	5404.500000	66990.955000	4
4	Potential loyal Customer	17.697778	460.257778	11450.674667	225

Hibernating (Cluster 0)

Hibernating customers are the customers who purchased long time ago, and who spend small amount, and obviously the recency will be high, R > 100, M < 1:000



Promising (Cluster 1)

Promising are customer who have average RFM value, they spend average amount, they also have relatively average recency, R < 50, F < 100, M (1000 - 5000)

Champion (Cluster 2)

Is the cluster of **High Value Customer** who shops frequently, and spend very large amount, M > 150K, R < 15

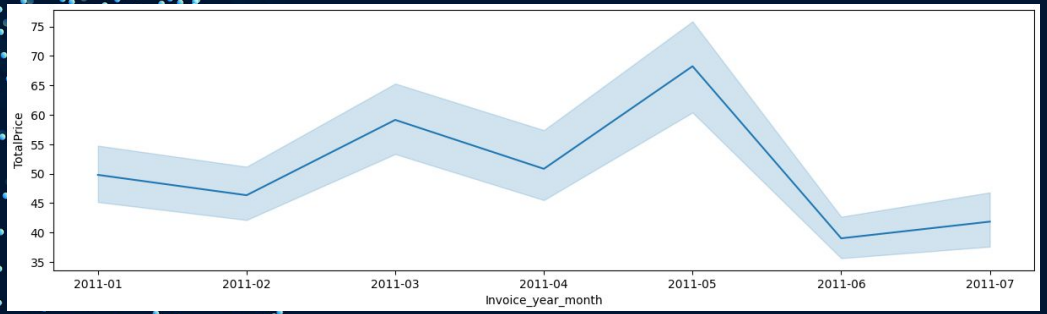
Loyalist Customer (Cluster 3)

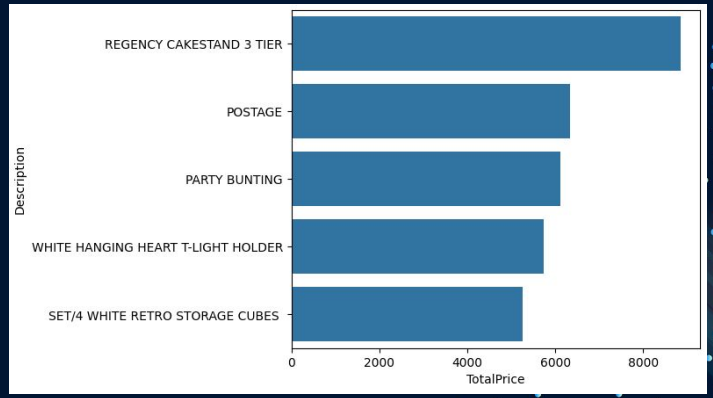
Loyalist are recent customer who frequently bought and spend large amount. R < 15, F > 1000, M > 50.000

Potential Loyal Customer (Cluster 4)

Potential Loyal are recent customer who frequently bought and spend good amount. R < 30, F < 500, M > 10.000

CLUSTER ANALYSIS AND RECOMMENDATIO N



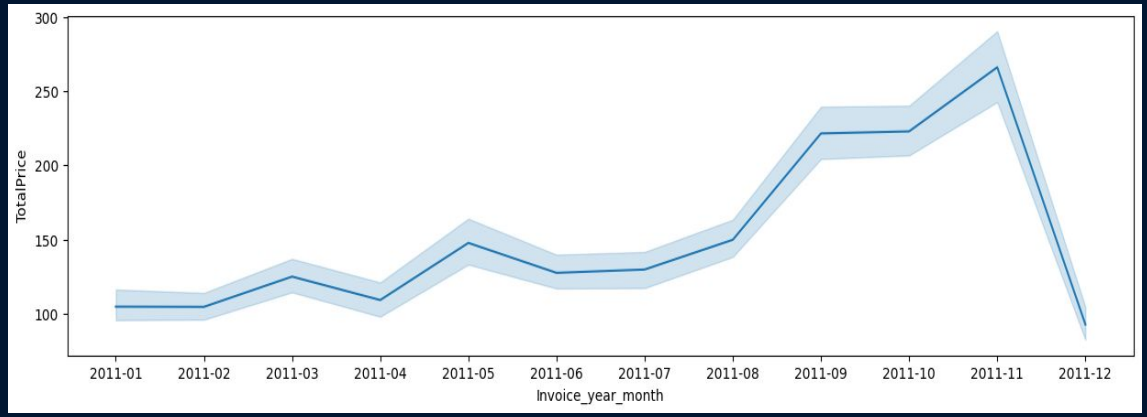


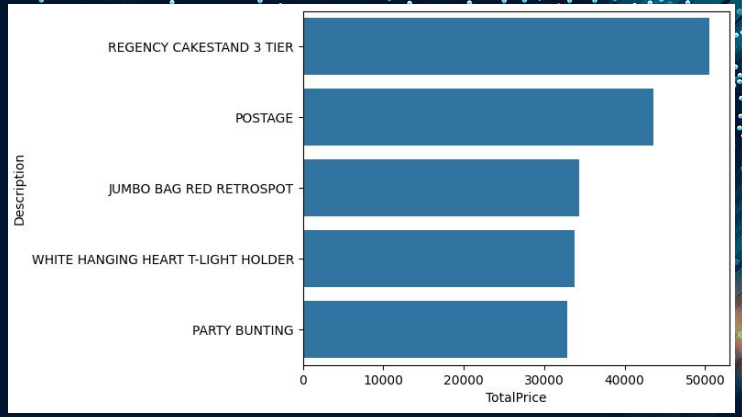
Hibernating (Cluster 0)

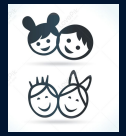
From the graph above we can see that spending in the hibernating cluster appears quite fluctuating, reaching its peak transaction in May, and the last transaction was made in July. From the list product they bought, they tend to buy low budget products, also buying this product in moderate quantities.

Recommendation for this cluster :

1. Sending regular newsletter or highlighting new cheap products, special offers and update about the platform.
2. Reach out to these customer for feedback on why they stopped transaction on our platform.
3. Ensure the platform is user-friendly and efficient, as better shopping experience might motivate them to return.





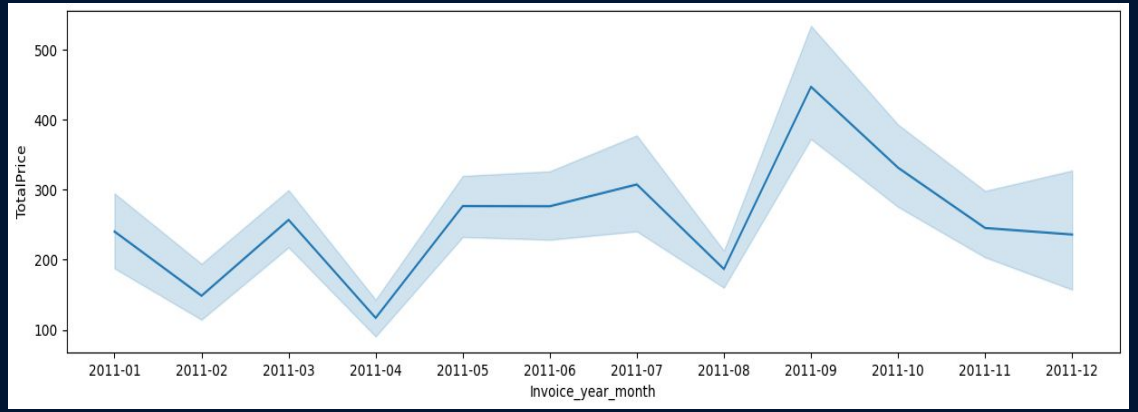


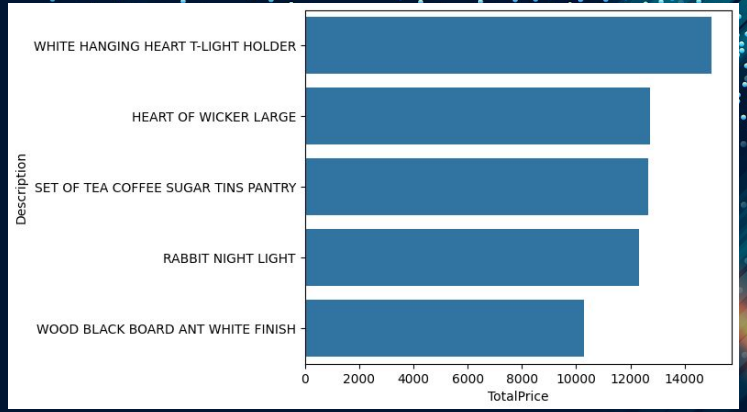
Promising (Cluster 1)

From the graph above we can see that the transaction activity of the promising cluster has been increasing month by month, reaching its peak in November, and from the top list based on income generated by the promising cluster, they tend to buy more low range products, and they make quite a lot of purchase for this products, contributing significantly to the company income.

Recommendation for this cluster :

1. Plan seasonal promotion around key shopping items, suchs as November, to capitalize on peak transaction periods.
2. Offer discounts on bulk purchases to incentivize customer to buy in large quantities, leveraging their tendency to purchase many items.
3. Organize flash sale featuring products they buy most, use limited-time offers to create a sense or urgency.



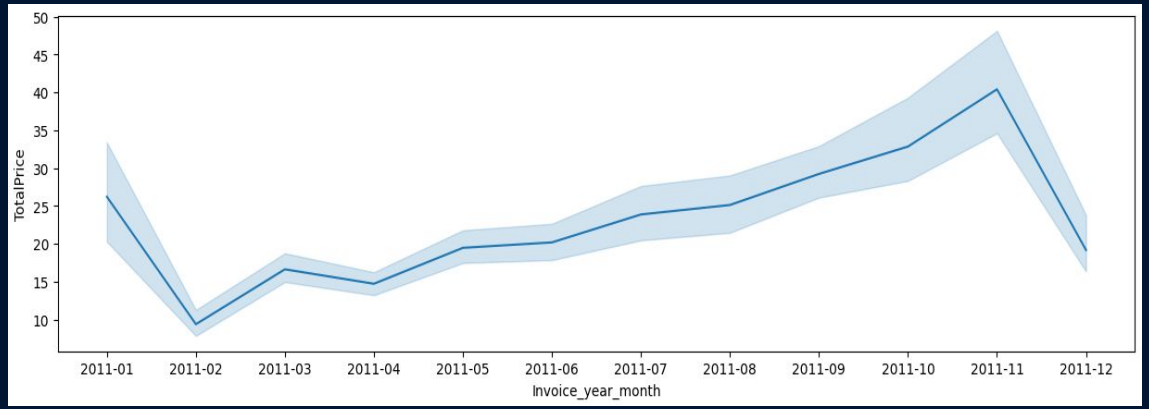


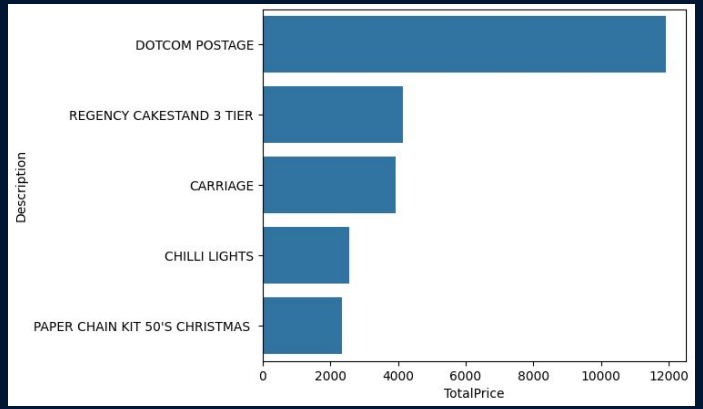
Champion (Cluster 2)

From the transaction graph above, it shows up and down transaction activity for the champion cluster in the early and at the end of the year, with peak spending occurring in September. Champion cluster consistently buying low range products in promising quantities, providing significant profit to the company.

Recommendation for this cluster :

1. Provide premium, responsive customer service exclusively for cluster champion. Ensure any issues or inquiries they have are handled efficiently
2. Provide exclusive access to new products or special offers only for this cluster, make them an early adapter.
3. Ensure a smooth and efficient shopping experience by improving website navigation, checkout process, and customer service.



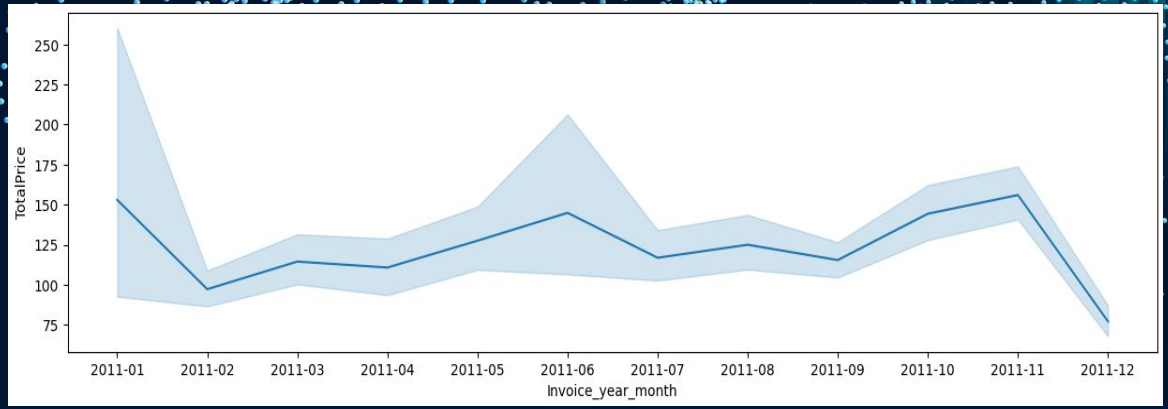


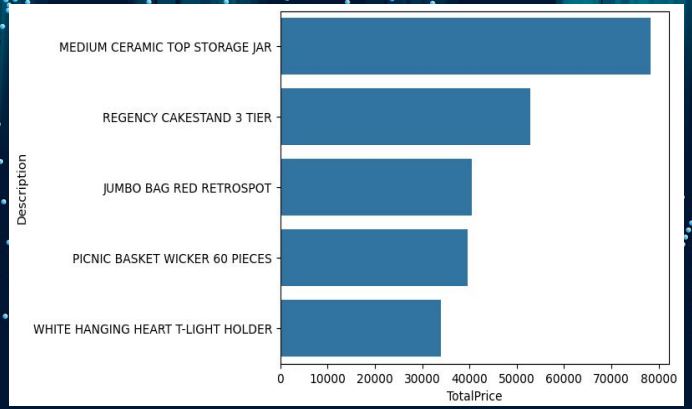
Loyalist Customer (Cluster 3)

Loyalist also experienced an increase in transaction from month to month, although there was a drop at the beginning of the year, and from the list of items purchased by the loyalist cluster, this cluster does not specialize in specific items. The even distribution of income from each type of product is a unique attraction of the loyalist cluster. This cluster generates significant income because their spending on medium and small items is evenly spread.

Recommendation for this cluster :

1. Provide incentives for loyalist cluster members to refer friends or family to our platform. A referral program can help to expand our customer base.
2. Enhance the loyalist cluster's experience by offering VIP or premium membership programs. This could include benefits like free shipping.





Potential Loyal Customer (Cluster 4)

As we can see the potential transaction graphic quite stable until drop at the end of the year, from the list product this cluster bought, like promising cluster, tends to purchase inexpensive items, but they also shows interest in luxury goods, moreover, their high transaction quantities indicate significant potential generating income for the company.

Recommendation for this cluster :

1. Implement a tier reward program based on purchase value. Offer additional discount or exclusive reward for specific purchase.
2. Utilize their shopping behavior data to create more personalized shopping experience. Send highly relevant product recommendation based on their purchase history, including combinations of inexpensive and luxury products that may appeal to them.

Thank You!

Reach me through :

bayu.trias92@gmail.com

Bayu Triona Asri

Triace15