PROJECT REPORT
A CUSTOMER SEGMENTATION AND PRODUCT RECOMMENDATION SYSTEM
USING RFM ANALYSIS, MARKET BASKET ANALYSIS & ITEM BASED COLLABORATIVE FILTER

DESIGNING A QUALITY INFORMATION PRODUCT TO MITIGATE PORTER'S FIVE FORCES:

"Smart Retailers," our new quality information product, is an online information product for a local grocery store. The goal of this project is to create a machine-learning model that uses market basket analysis to identify correlations between items in customers' shopping carts.

Market basket analysis (MBA) is a data mining technique that can be used to identify correlations between frequently purchased items online and important relationships between products, which can help smart retailers increase online sales and generate revenue. It can also assist in identifying products that customers frequently purchase together in a single transaction or across multiple transactions.

This transactional technique can be used to identify products that have a high association with each other and are usually bought together. We can then use this information to create bundles of frequently bought items, explore cross-selling and up-selling opportunities, and target customers with similar purchasing habits to increase sales for the company.

In terms of being able to counter Porter’s Five Forces, our information product can be used to combat the threat of new entrants and the bargaining power of customers. By being able to identify associations with products and create promotional bundles using smart retailers, online stores can create a unique selling point (USP) that is difficult to replicate. Additionally, by creating these bundles and increasing sales, online companies can create loyal customers that can be targeted with specific discounts that can further increase their loyalty and help reduce the bargaining power of customers who may be tempted to switch to other competitors who offer lower prices.

The threat of new entrants is also difficult for online stores using smart retailers; this is usually the case because the USP proposition of creating sales by identifying frequently bought together items bundled with exclusive discounts can be very difficult to replicate for new entrants that are low on budget and economies of scale. Using techniques such as MBA and our information product, stores can help identify the frequency of items bought and specific bundles can be created, leading to higher sales, and generating a competitive advantage with a significantly lower number of resources.

COMBINING, ENHANCING, AND TRANSFORMING DATA ELEMENTS INTO A NEW INFORMATION PRODUCT:

Multiple elements are combined, improved, and transformed to form our information product, Smart Retailers. We started by using the RFM analysis, which considers the recentness, frequency, and dollar amount of purchases, and can be used in combination with both datasets to provide even deeper insights into consumer behavior and assist retailers in identifying profitable customer segments. By using this data, personalized offers can be made, customer retention can be improved, and customer lifetime value can rise. In general, these datasets are essential for a smart retailer because they provide the data required to comprehend consumer behavior and improve business plans. To perform RFM analysis, we will use the customer's purchase history data from our data set to create three different measures: RFM-R (recency), RFM-F (frequency), and RFM-M (monetary). RFM-R will be calculated by determining the number of days since the last purchase using the max and min functions from the order date. RFM-F will be calculated using the distinct count of invoice IDs for each customer ID. **(DOĞAN, AYÇİN and BULUT 2018)**

We further enhanced the datasets by assessing their F-scores, R-scores, and M-scores. RFM-M is the monetary value attribute using the sum of the spent amount per customer ID. After creating these measures, we improved these elements by assessing their RFM Scores, which will be labeled as R-score, F-score, and M-score, and then we created quartiles. Values falling in the first 25 percentile will be assigned an R-score of 1, while values in the top 75 percentile will be assigned a score of 1 for RFM-F and RFM-M. We then computed a combined RFM score and used it to segment the data with a series of if-else statements in the calculation field to categorize customers as Gold, Silver, or Bronze. These different elements are combined to come up with our information product. **(Figure 1) (Makhija 2021)**

THE STRATEGIC IMPORTANCE OF THE NEW INFORMATION PRODUCT FOR THE RETAIL INDUSTRY:

Smart retailer can be strategic for the online retail industry in multiple ways. Essentially, the main purpose of our product is to be able to identify different products that sell together and create product mixes; these can be used to devise strategies that help increase sales for online retailers. We can integrate the following things into devising an optimal strategy for online retailers: By embedding smart retailers into their systems, online retailers can identify cross-selling opportunities, product placements, promotions and discounts, and customer segments.

Strategy can be highly influenced by identifying cross-selling opportunities, selling patterns, and creating product mixes that increase the overall sales of the company. This is a key element for devising our competitive advantage and unique selling point. By using this technique, online stores can increase the average order value and encourage repeat purchases.

Secondly, by identifying product placements, online grocery stores can determine the best placement of products on their websites. For example, if customers frequently purchase cereal and milk together, the website can be configured to display these products in closer clicking proximity, therefore increasing sales, and generating revenue. **(DOĞAN, AYÇİN and BULUT 2018)**

Thirdly, by identifying the promotions and discounts through Smart Retailers, online grocery stores can identify the products that are most likely to be purchased together and create promotions or discounts that encourage customers to buy these products together. For example, if customers frequently purchase chips and salsa together, the store can offer a discount on both products when they are purchased together. Moreover, by using these techniques to design a loyalty points system that can further add on to discounts, smart retailers can act as strategic information products. **(Kubiak and Weichbroth 2010)**

Lastly, smart retailers can help identify customer segmentation, which can help in devising the marketing strategy for online grocery stores. This will help online stores segment their customers based on their purchasing behavior and buying patterns. This can help the store create personalized marketing campaigns and associated product offerings for each segment. For example, if one segment of customers frequently purchases organic produce, the store can create a targeted marketing campaign for this segment and further increase sales and revenues, setting the direction for the retailer's strategy. **(Palanisamy 2014)**

SELECTION AND ACQUISITION OF RELEVANT DATA SETS FROM DATA SOURCES:

The transaction data from online retail stores located in various countries, which covers information for the year of 2011, is the original dataset we are using to build a smart retailer. Each customer's transactional information is provided in this dataset, which also includes their Customer ID, country of residence, product description, invoice creation date, item code, invoice id, number of rows, unit price of the product, quantity, and sale amount. This information was obtained from Kaggle. This data set was the foundation for our RFM analysis which we conducted to come up with the scores for recency, frequency, and monetary value.

Secondly, we further enriched the dataset by using the customer demographic data, which is needed to give the data more meaning and demographic metrics. This dataset was also taken from Kaggle so that we can analyze it based on different metrics (like age, gender, etc). The reason we would want to study these metrics is that it will help us gain deeper insight on the different purchase patterns and eventually help stores in identifying effective strategies that help them gain a competitive advantage.

Attaching a screenshot of the dataset, this dataset will provide more in-depth insights and patterns to help understand the features of our information product. Since this information wasn't easily accessible, we had to deal with access and permission problems. To avoid having duplicate information, we used Python to generate Customer Demographic data (Synthetic Data) for each unique customer ID.

Country	Customer ID	Customer_Name	Description	Invoice Date	Item Code	Invoice ID	Unit Price	Quantity	Sales Amount	Order_Date	Order_Time	Month	Email	Address	City	Postal_Code	Phone	Age	Gender
United Kingdom	15240	Stuart James	RED SPOT CERAMIC	1/2/11	21671	542776	1.25	48	60	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	RED STRIPE CERAM	1/2/11	21668	542776	1.25	12	15	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	RETROSPOT HEART	1/2/11	21485	542776	4.95	3	14.85	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	RED SPOTTY BISCU	1/2/11	21218	542776	3.75	6	22.5	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	NAMASTE SWAGAT	1/2/11	17021	542776	0.3	36	10.8	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	PAPER CHAIN KIT R	1/2/11	22083	542776	2.95	12	35.4	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	HOT WATER BOTTL	1/2/11	22835	542776	4.65	4	18.6	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	RED RETROSPOT CA	1/2/11	21843	542776	10.95	1	10.95	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	CHILLI LIGHTS	1/2/11	79321	542776	4.95	4	19.8	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female
United Kingdom	15240	Stuart James	6 RIBBONS RUSTIC	1/2/11	22077	542776	1.65	12	19.8	1/2/11	0:00:00	Jan-11	armstrongclive@ex	Studio 59 Davis crossroad	South Dawn	G5 0QS	+44(0)1632960796	44	Female

DISTINGUISHING OUR NEW INFORMATION PRODUCT FROM EXISTING PRODUCTS:

There may be existing information products that use market basket analysis to identify frequently purchased items together or analyze transactional data to identify trends in customer behavior and preferences. However, what distinguishes our new information product is the specific application of market basket analysis in the retail industry and how it will be used to optimize product recommendations, personalize the customer experience, and improve inventory management and marketing strategies.

The new information product was inspired by existing market basket analysis techniques and their successful application in various industries. However, the focus on the retail industry and the use of transactional data to identify correlations between items and customer preferences inspired the design of this information product.

Furthermore, this new information product was created to address the specific needs of the retail industry, such as improving customer experience and increasing sales. It is designed to be scalable and adaptable to changing customer behavior and preferences, making it an asset to any retail store seeking a competitive advantage.

In conclusion, while similar information products exist, our focus on the retail industry and the use of market basket analysis to improve the customer experience and increase sales distinguishes us. We were inspired by successful market basket analysis techniques, but we designed our product specifically for the retail industry.

IDENTIFYING THE TARGET AUDIENCE FOR THE INFORMATION PRODUCT:
The Smart Retailer is the primary information product for various retailers in the retail industry who will be the consumers of this new strategic information product. They will primarily use this data product to gain insights into customer behavior and preferences, optimize product recommendations, and improve inventory management and marketing strategies. The management team will oversee implementing any changes or improvements based on the machine learning model's insights. (DOĞAN, AYÇİN and BULUT 2018)

The marketing team will also use the information product to tailor their campaigns and promotions to customer preferences and behavior. The marketing team can create more successful targeted promotions by understanding what products are frequently purchased together., the marketing team can create more likely-to-be-successful targeted promotions. The inventory management team can also use the product to optimize inventory, optimize stock levels, reduce waste, and improve efficiency.

This information product may also benefit the customer service team by allowing them to provide more personalized recommendations and suggestions to customers. They can use the data to identify patterns in customer behavior and preferences, allowing them to provide better service and increase customer satisfaction. (Kubiak and Weichbroth 2010)

Overall, the information product is intended for the stores under the Retail Industry’s benefit, with the whole management team serving as the primary consumers. The goal is to provide insightful data that can be used to drive business growth and improve the customer experience. The store can gain a competitive advantage in the highly competitive retail industry by leveraging machine learning and advanced analytics. (Makhija 2021)

THE QUALITY OF OUR INFORMATION PRODUCT:
The information product proposed for Smart Retailer is of high quality because it addresses a critical need in the Retail industry: understanding customer behavior and preferences. The model's machine learning techniques, such as RFM Analysis, Collaborative Filtering, and market basket analysis, provide valuable insights into customer behavior that can be used to improve the store's product recommendations, inventory management, and marketing strategies. **(DOĞAN, AYÇİN and BULUT 2018)**

Market basket analysis' effectiveness is well-established in the retail industry, and the model's ability to identify frequently purchased items together and items commonly purchased with specific products can help personalize recommendations to customers, leading to increased customer satisfaction and loyalty. **(Palanisamy 2014)**

Furthermore, the model's ability to identify trends in customer behavior and preferences can provide Smart Retailer with a competitive advantage by allowing it to constantly adjust its inventory and marketing strategies to meet the needs of customers.

However, the accuracy and cleanliness of the data used in the model determine the quality of the information product. A rigorous data-cleaning process is required to ensure the data's reliability, which is critical for the model's success. Furthermore, Smart Retailers must ensure transparency to build customer trust and make them aware of how their data is being used.

In conclusion, the Smart Retailer information product is of high quality because it addresses a significant need in the Retail industry, provides valuable insights into customer behavior and preferences, and can optimize the store's product recommendations, inventory management, and marketing strategies.

MANAGING AND GOVERNING DATA QUALITY FOR LONG-TERM SUSTAINABILITY AND IMPROVEMENT:

Smart retailers must implement a robust data management process to sustain and improve the quality of the data elements involved in the information product. The collection of transactional data from multiple sources, such as the online store, point-of-sale systems, and customer relationship management software, should be the first step in this process. To remove errors, duplicates, and inconsistencies, the collected data should be cleaned. The cleaned data should then be combined into a single dataset, with data governance policies and procedures in place to ensure that the data is accurate, consistent, and up to date.

Smart Retailers should regularly analyze the impact of data quality issues on the model's accuracy and take corrective action to ensure that data quality is continuously monitored and improved. Smart retailers should also implement data security measures to prevent unauthorized access, modification, or disclosure of customer data.

Smart retailers should also improve their data management processes on a continuous basis by implementing best practices and emerging technologies such as data quality tools, data governance solutions, and analytics platforms. Employee training and development programs can also ensure that they have the necessary skills to effectively manage and govern the data elements involved in this information product.

Smart retailers can ensure that the data elements involved in the information product are of high quality, accurate, and up to date by implementing a comprehensive data management process. As a result, the accuracy and effectiveness of the machine learning model will improve, leading to better product recommendations, inventory management, and marketing strategies. This will ultimately lead to increased customer satisfaction, loyalty, and sales.

ENSURING LONG-TERM STRATEGIC VALUE THROUGH EFFECTIVE MANAGEMENT AND GOVERNANCE:

Smart retailers should focus on continuously improving the accuracy and effectiveness of the machine learning model to sustain and improve the strategic value of the information product. This can be accomplished through ongoing data analysis, model refinement, and alignment of business strategies.

Firstly, smart retailers should perform regular data analysis to identify trends, patterns, and anomalies in transactional data. This analysis can assist in identifying customer behaviors, preferences, and purchasing patterns, which can then be used to inform product recommendations, inventory management, and marketing strategies. Smart retailers can perform this analysis using advanced analytics tools such as predictive modeling, clustering, and segmentation.

Second, Smart Retailer should fine-tune the machine learning model to ensure that it is in sync with the most recent data trends and business requirements. This may entail retraining the model with new data or fine-tuning the existing model to improve accuracy and effectiveness. Smart retailers can refine their products using machine learning platforms and tools like TensorFlow or PyTorch.

Third, Smart Retailer should ensure that the machine learning model is aligned with the business strategy to ensure that it delivers maximum strategic value. This can include defining specific business goals, such as increasing customer satisfaction, increasing sales, or improving inventory management. Smart Retailers can then use the machine learning model to inform and execute on these goals, ensuring that they provide measurable value to customers.

Smart retailers should invest in technology and human resources to support these initiatives. Investing in advanced analytics and machine learning platforms, as well as hiring data scientists and analysts to analyze data and refine the model, are all part of this strategy. Smart retailers should also prioritize employee training and development programs to ensure that their employees have the necessary skills to effectively manage and govern the machine learning model.

To summarize, maintaining and improving the strategic value of an information product necessitates ongoing data analysis, model refinement, and alignment with business strategies. Smart retailers should invest in technology and human resources to support these initiatives, as well as prioritize employee training and development, to ensure that their employees have the skills needed to effectively manage and govern the machine learning model. Smart retailers can ensure that the information product delivers maximum strategic value, resulting in increased customer satisfaction, loyalty, and sales.

-- End of Report ---

References and Analysis (Additional Pages)

1. *Onur Doğan, Ejder Ayçin and Zeki Atıl Bulut . 2018. "Customer Segmentation by Using Rfm Model and Clustering Methods: A Case Study in Retail Industry." International Journal of Contemporary Economics and Administrative Sciences, Vol. 8, No.1, 1-19*
2. *Makhija, Pushpa. 2021. Clevertap – RFM Analysis for Customer Segmentation. June 03. https://clevertap.com/blog/rfm-analysis/*
3. *Palanisamy, Karthikeyan. 2014. Analytics Vidhya - Effective Cross Selling using Market Basket Analysis. August 4. https://www.analyticsvidhya.com/blog/2014/08/effective-cross-selling-market-basket-analysis/*
4. *Kubiak, Bernard & Weichbroth, Paweł. (2010). Cross- And Up-selling Techniques In E-Commerce Activities. Journal of Internet Banking and Commerce. 15.*
5. *Online Retail Transaction Dataset - https://www.kaggle.com/datasets/sachinsin8h/online-retail-dataset*

RFM Analysis:
Data preparation steps: Finding Additional metrics like order date and time from Invoice Date calculated snapshot date which would be the end day of order date in the dataset + 1 to calculate the days period.

What is RFM and How we calculated RFM?

RFM analysis is the Recency, Frequency, and Monetary Analysis. where the 'R' factor is about when was the last time a customer made a purchase, the 'F' factor is about the number of purchases made in each period, and the 'M' factor is the total amount of money spent by the customer in the given time.

Recency was calculated based on the difference between the purchase date and the transaction's last day in the dataset. Here we use the lambda function to find the number of days between the hypothetical day and the last transaction. Frequency was calculated based on each Invoiced count by a distinct customer ID. Monetary value is calculated by the sum of the Sales Amount.

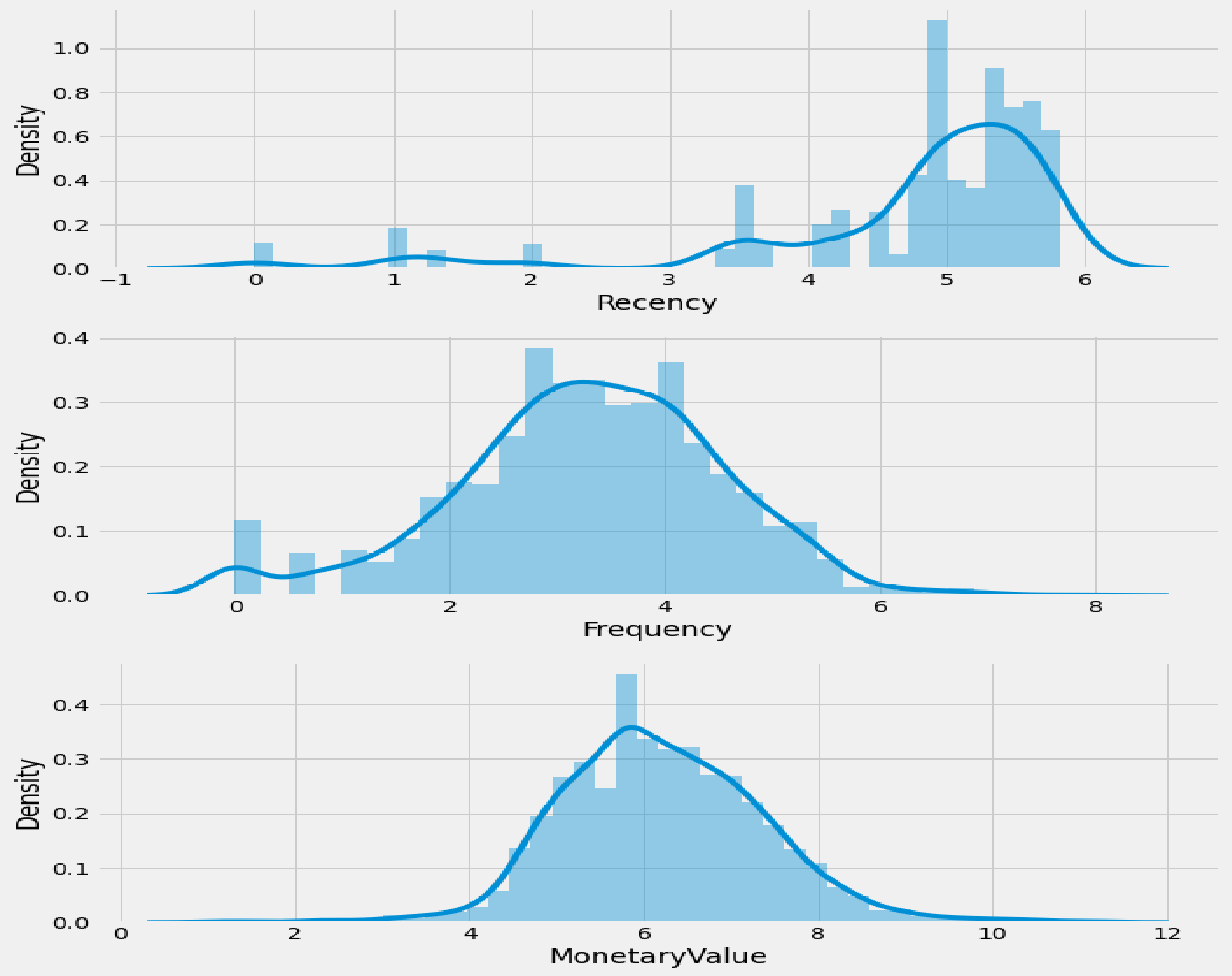
We then calculated Quartiles for each of the RFMs to know the type of the customer for RFM segment**.** (This analysis was in Python)
So here recency should basically have a low value to be considered best because the gap between days should be less for the recent customers to be considered best customers (so here 4 is considered best (25 percentile) and 1 as a lost customer (75 to 100 percentile)
Based on the Frequency we should have more purchases for a customer, so we consider 4 as the most frequent customer (75 to 100 percentile) and 1 as a less frequent customer (25 percentile)
Based on the monetary value we should have more spending for a customer, so we consider 4 as the most spending customer (75 to 100 percentile) and 1 as a less frequent customer (25 percentile). (**Figure 2**)

Calculated RFM score by adding scores of each R, F, and M segment to get an overall score and then dividing our set of customers if the score is greater than 9 then we consider them as gold customers and between 5 to 9 as silver and anything below that as bronze.

Figure 1: Output of RFM cluster

	Recency	Frequency	MonetaryValue	
	mean	mean	mean	count
General_Segment				
Bronze	245.9	11.8	202.6	555
Gold	88.8	131.4	2684.3	538
Silver	161.3	33.9	750.9	862

Figure 2: Distribution of RFM values:



So, here we have taken two analyses, one in Tableau (to find and understand different types of customers and RFM segment customer sales over time) and the other in Python (using k means clustering and market basket analysis, Item-based content filtering to find the frequently brought items and product recommendation to the customers).

K means clustering:

Firstly, we normalized the Data with standard Scaler, and then we applied K-means algorithm and based on the silk score we had 3 clusters as optimal value. (**Figure 3**)

Silhouette score for 2 clusters: 0.3765348223296697
Silhouette score for 3 clusters: 0.39736747853934706
Silhouette score for 4 clusters: 0.3304446596575849
Silhouette score for 5 clusters: 0.2833330410998587
Silhouette score for 6 clusters: 0.28495836203713265
Silhouette score for 7 clusters: 0.3119053504631749
Silhouette score for 8 clusters: 0.2744198634676409
Silhouette score for 9 clusters: 0.2839431206552134
Silhouette score for 10 clusters: 0.26439115715801265
Silhouette score for 11 clusters: 0.27717765141455636
Silhouette score for 12 clusters: 0.2854125050319975
Silhouette score for 13 clusters: 0.29560981387931706
Silhouette score for 14 clusters: 0.2904100599827949

Figure 3: Finding out optimal clusters.

From the 0th Cluster, we can see that there was high mean recency (i.e gap between Successive Transactions), low mean frequency, and low mean monetary value(i.e Bronze Customers)

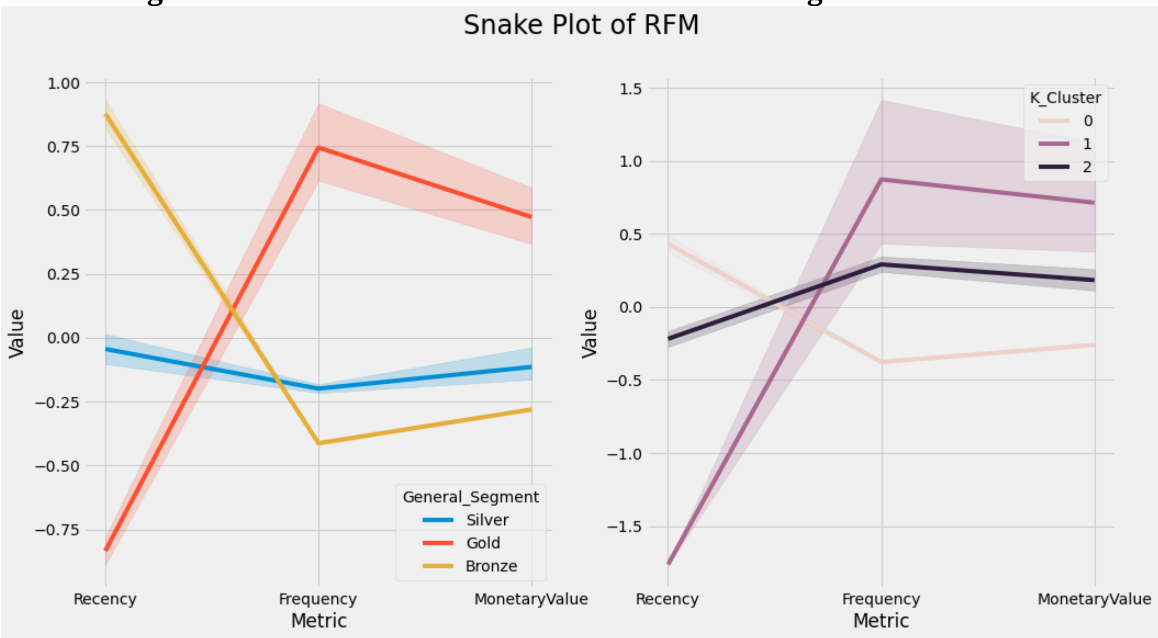
From the 1st Cluster, we can see that there was low mean recency (i.e gap between Successive Transactions), high frequency of transactions with the highest monetary value (i.e Gold Customers)

From the 2nd Cluster, we can see that there was moderate mean recency (i.e gap between Successive Transactions), the moderate frequency with moderate monetary value(i.e Silver Customers).

Figure 4: Output for RFM optimal clusters

	Recency	Frequency	MonetaryValue	
	mean	mean	mean	count
K_Cluster				
0	205.0	16.0	271.0	970
1	4.0	145.0	3482.0	133
2	145.0	85.0	1735.0	852

Figure 5: Snake Plot of RFM to understand the segments behavior.



We can also Intrepret the Same thing with their values using Snake plot as this Accurately represents the General segments and the K_clusters as K_cluster 0 is bronze and K_cluster 1 is Gold and K_cluster 2 is Silver from this plot.

Market Basket Analysis

Apriori Algorithm:
The Apriori algorithm is based on the concept of how two or more products/objects are related to one another. In other words, it is an algorithm that analyzes customers who purchased product A and product B. In general, it works on datasets with many transactions.

- Our goal is to focus on bronze segment customers moving forward.

Figure 6: Finding out frequently purchased item.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
27	(PINK OWL SOFT TOY)	(6 RIBBONS RUSTIC CHARM)	0.004769	0.020668	0.004769	1.0	48.384615	0.004671	inf
54	(AIRLINE LOUNGE,METAL SIGN)	(COLOUR GLASS T-LIGHT HOLDER HANGING)	0.004769	0.028617	0.004769	1.0	34.944444	0.004633	inf
58	(ALARM CLOCK BAKELIKE CHOCOLATE)	(ALARM CLOCK BAKELIKE RED)	0.007949	0.025437	0.007949	1.0	39.312500	0.007747	inf
79	(ALARM CLOCK BAKELIKE PINK)	(ALARM CLOCK BAKELIKE RED)	0.007949	0.025437	0.007949	1.0	39.312500	0.007747	inf
169	(BAKING MOULD ROSE MILK CHOCOLATE)	(BAKING MOULD CHOCOLATE CUPCAKES)	0.004769	0.007949	0.004769	1.0	125.800000	0.004732	inf
...	...	...	...	...	...	...	...	...	...
3491	(HERB MARKER PARSLEY)	(HERB MARKER BASIL, HERB MARKER THYME, HERB MA...	0.004769	0.004769	0.004769	1.0	209.666667	0.004747	inf
3492	(HERB MARKER THYME)	(HERB MARKER BASIL, HERB MARKER PARSLEY, HERB ...	0.004769	0.004769	0.004769	1.0	209.666667	0.004747	inf
3493	(HERB MARKER CHIVES)	(HERB MARKER BASIL, HERB MARKER PARSLEY, HERB ...	0.004769	0.004769	0.004769	1.0	209.666667	0.004747	inf

- With High Lift scores and 100% Confidence, we can see here that we have 1013 frequently purchased items together.
- Similarly, we can do this for Silver and Gold Customers**.**

Item-based Collaborative Filtering:

This approach entails predicting customers' preferences and identifying products that they are likely to buy by analysing data gathered from **many** customers with similar choices or preferences. The underlying assumption is that if person A and person B have similar reactions to certain items, they are likely to have similar preferences or opinions about other items.

Figure 7: Applying Item based collaborative filter.

```
In [75]: similarProductsW_Bronze = matrix_Bronze.corrwith(HerbMaker)

         similarProductsW_Bronze = similarProductsW_Bronze.dropna()

         df1 = pd.DataFrame(similarProductsW_Bronze)

         df1.head(10)

         corrMatrix_Bronze = matrix_Bronze.corr()

         corrMatrix_Bronze.head()

         # Suppose if I want to find 3rd Invoice ID for which Customer Buys
         second_customer_Bronze = matrix_Bronze.iloc[2].dropna()

         second_customer_Bronze.head()

         /Users/saipranayreddy/opt/anaconda3/lib/python3.9/site-packages/numpy/lib/function_base.py:268
         ees of freedom <= 0 for slice
           c = cov(x, y, rowvar, dtype=dtype)
         /Users/saipranayreddy/opt/anaconda3/lib/python3.9/site-packages/numpy/lib/function_base.py:254
         de by zero encountered in true_divide
           c *= np.true_divide(1, fact)

Out[75]: Description
         ANTIQUE TALL SWIRLGLASS TRINKET POT     20.0
         FLUTED ANTIQUE CANDLE HOLDER            24.0
         HOME BUILDING BLOCK WORD                 6.0
         LOVE BUILDING BLOCK WORD                 6.0
         T-LIGHT GLASS FLUTED ANTIQUE            36.0
         Name: 540274, dtype: float64
```

Figure 8: Product recommendation system for a bronze customer

In [76]:
```
simProducts_Bronze = pd.Series()

#Go through every product bought by second customer

for i in range(0, len(second_customer_Bronze.index)):

    print("Adding similar Items for " + second_customer_Bronze.index[i] + "....")

    #Retrieve similar products to the ones bought by customer 2
    ## These may be brought by other customers in combination with one or more products choosen by the 2nd customer

    sims_Bronze = corrMatrix_Bronze[second_customer_Bronze.index[i]].dropna()

    #Scale to how many of the products were bought

    sims_Bronze = sims_Bronze.map(lambda x: x * second_customer_Bronze[i])

    # Add to the list of similar products

    simProducts_Bronze = simProducts_Bronze.append(sims_Bronze)


simProducts_Bronze.sort_values(inplace = True, ascending = True)

print(simProducts_Bronze)
```

```
Adding similar Items for ANTIQUE TALL SWIRLGLASS TRINKET POT....
Adding similar Items for FLUTED ANTIQUE CANDLE HOLDER....
Adding similar Items for HOME BUILDING BLOCK WORD....
Adding similar Items for LOVE BUILDING BLOCK WORD....
Adding similar Items for T-LIGHT GLASS FLUTED ANTIQUE....
```

Sorting the results and avoiding duplicates

[77]:
```
simProducts_Bronze= simProducts_Bronze.groupby(simProducts_Bronze.index).sum().sort_values(ascending = False)

filteredSims_Bronze = simProducts_Bronze.drop(second_customer_Bronze.index)

filteredSims_Bronze.head(5)
```

:[77]:
```
ROSES REGENCY TEACUP AND SAUCER      36.000000
REGENCY SUGAR BOWL GREEN             36.000000
PINK ROSE WASHBAG                    20.000000
ANTIQUE GLASS DRESSING TABLE POT     19.597274
BATH BUILDING BLOCK WORD             12.000000
dtype: float64
```

This code is implementing a product recommendation system for a bronze market segment based on collaborative filtering. It first creates a pivot table of the quantity of each product purchased by each customer, and then applies one-hot encoding to create a binary matrix of the presence or absence of each product in each customer's basket. It then calculates the co-occurrence matrix of all pairs of products and identifies the pairs that co-occur the most frequently.

Next, it selects a particular product ("COLOUR GLASS T-LIGHT HOLDER HANGING") and computes the correlation of that product with all other products purchased. It then selects a second customer and finds the products that they have purchased, and for each of those products, it retrieves the most similar products (based on the correlation calculated earlier) and scales the similarity by the quantity of the product purchased. Finally, it sorts the resulting list of similar products in ascending order and prints it out. The resulting list represents the recommended products for the second customer based on their purchase history and the purchase histories of other customers who have purchased similar products.

Tableau for RFM Analysis we have:

Here we took percentiles like that of the python but instead of the return of 4, for each of the R,F, and M scores we consider 1 as the best customer here for each segment.

So here recency should basically have a low value to be considered best because the gap between days should be less for the recent customers to be considered best customers (so here 1 is considered best (25 percentile) and 4 as a lost customer (75 to 100 percentile)
Based on the Frequency we should have more purchases for a customer, so we consider 1 as the most frequent customer (75 to 100 percentile) and 4 as a less frequent customer (25 percentile)
Based on the monetary value we should have more spending for a customer, so we consider 1 as the most spending customer (75 to 100 percentile) and 4 as a less frequent customer (25 percentile).

Calculated RFM score by adding scores of each R, F, and M segment to get a segment score.

Code used to get the calculated metric for RFM segment:

```
IF [RFM-Score] == 111
THEN 'Best Customers'
ELSEIF [R-Score] == 1 AND [F-Score] == 2 AND [M-Score] <=2
THEN 'Potential To Become Best Customer'

ELSEIF [F-Score] == 1
THEN 'Loyal Customer'

ELSEIF [M-Score] == 1
THEN 'Big Spenders'

ELSEIF [RFM-Score] == 311
THEN 'Almost Lost'

ELSEIF [RFM-Score] == 411
THEN 'Lost Customers'

ELSEIF [RFM-Score] == 444
THEN 'Lost Cheap Customers'

ELSE
```

Figure 9: RFM Analysis Dashboard to identify customer segments.

Conclusion:

In conclusion, our analysis using Tableau enabled us to identify distinct customer segments in RFM analysis and determine the frequently purchased items for each segment. Armed with this information, we can tailor our marketing strategies to cater to the specific needs and preferences of each customer segment and recommend products based on their previous purchase history. Smart Retailer can improve inventory management leading to increased customer loyalty and sales this will ultimately help us enhance customer satisfaction and drive business growth.

Table 2: This will be our Marketing strategy for each customer segment:

Customer	RFM Score	Marketing Strategy
Best Customers	111	Give the big bonus points when they shop and special promos for the best customers.
Loyal Customers	X1X	Offer Personalized recommendations and upselling
Potential to become best customer	12X	To convert them into loyal customers, offer them personalized product recommendations, Exclusive Discounts and Loyalty Program
Big spenders	XX1	Providing exclusive deals on high-ticket items, and personalized customer service.
Almost Lost	311	To win them back, offer them personalized discounts, personalized recommendations based on their purchase history, and exceptional customer service
Lost Customers	411	To re-engage them, offer them personalized discounts, exclusive deals on new products, and personalized recommendations based on their past purchase history.
Lost Cheap Customers	444	Ignore them
Look Out Buyers	2XX	To encourage them to become more loyal, offer them personalized recommendations, exclusive discounts, and loyalty programs.
Occasional Buyers	22X	To encourage them to make more frequent purchases, offer them personalized recommendations, exclusive discounts, and loyalty programs that reward their continued patronage

Here X = 1,2,3,4 based on their respective RFM scores.