Final Project

Customer Segmentation

from Everything Plus
transaction data

Outline

Project Background

The Data

Exploratory Data Analysis

Project Background

We're helping *Everything Plus,* an online store that sells household goods, make personalized offers for their users. Our goal is to analyze the transaction history data to identify patterns and trends that can help us make informed decisions. This project is a unique opportunity as Everything Plus hasn't conducted this kind of research before. We will provide detailed insights and recommendations to help personalize offers for users.

The Data

We started by cleaning and preparing the data. This included various tasks such as renaming columns, converting data types, removing duplicate entries, handling missing values, eliminating outliers, and discarding irrelevant data.

As a result, around 8,174 rows, which represented 0.985% of the original 541,909 rows, were eliminated from the datasets.

Order Information

order_id

timestamp

customer_id

quantity

Item Information

item_id

item_price

item_name

Exploratory Data Analysis

Daily

73 Customers

83 Transactions

1,750 Products

19,000 Items

$33,572 Revenues

Weekly

408 Customers

494 Transactions

10,465 Products

113,630 Items

$200,771 Revenues

Monthly

1394 Customers

2099 Transactions

44,478 Products

482,928 Items

$853,277 Revenues

Outline

- RFM Analysis
- K-Means & Evaluation Metrics
- Segmentation
- Insights and Recommendations

RFM Analysis

What is it?

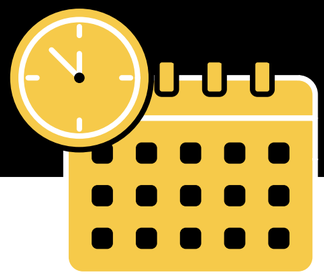
- RFM refers to three significant customer traits: ***Recency, Frequency, and Monetary*** value. These metrics play a vital role in understanding a customer's behavior as frequency and monetary value determine the customer's lifetime value, while recency measures engagement and affects retention.

RFM factors illustrate these facts:

- the more recent the purchase, the more responsive the customer is to promotions.
- the more frequently the customer buys, the more engaged and satisfied they are monetary value differentiates heavy spenders from low-value purchasers.

RFM Analysis

Metrics



Recency

The freshness of the customer activity, be it purchases or visits

E.g. Time since last order or last engaged with the product

Frequency

The frequency of the customer transactions or visits

E.g. Total number of transactions or average time between transactions

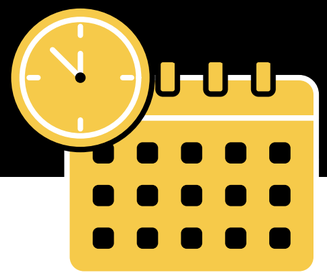
Monetary

The intention of customer to spend or purchasing power of customer

E.g. Total or average transactions value

RFM Analysis

with our Data



Recency

This data was obtained by grouping the data by `customer_id` and then calculate the difference between max `timestamp` of each customer and max `timestamp` of the whole data.

Frequency

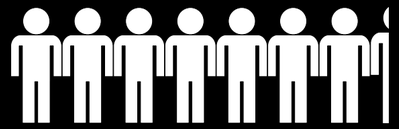
This data was obtained by grouping the data by `customer_id` and aggregating the count of `order_id`.

Monetary

This data was obtained by grouping the data by `customer_id` and aggregating the product of `item_price` and `quantity`.

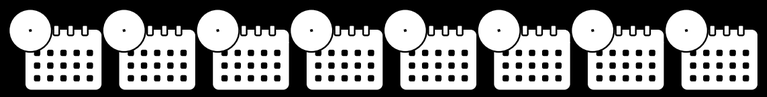
RFM Analysis

with our Data



7370 Unique
Customer ID

Median



80 Days from last purchase

23.5 Orders

$358 of Revenues

Average

123 Days from last purchase

72.4 Orders

$1389 of Revenues

K-Means Clustering

K-means clustering is a type of unsupervised machine learning algorithm that groups unlabeled data into k number clusters, where k is a user- defined integer.

The algorithm starts by randomly selecting k points as centroids of k clusters. It then calculates the Euclidean distance for each remaining data point from each of those centroids and assigns each data point to its closest cluster.

The algorithm then recalculates the centroid for each cluster by taking the mean of all the vectors inside the group, and repeats this process until all data points are assigned to a cluster. To optimize the performance of K-means clustering, it is important to determine the optimal value for k, which can be achieved through simple data visualization or other techniques.

K-Means Clustering Evaluation Metrics

Silhouette Score

- The silhouette score is a metric used to evaluate the performance of a clustering algorithm and to determine an optimal value of k.
- It is calculated by averaging the distances from intra-cluster and nearest cluster samples.
- A score of 0 indicates overlapping clusters, while a score of 1 indicates well-separated and dense clusters.
- A higher score indicates better cluster separation.

Calinski Harabaz Index

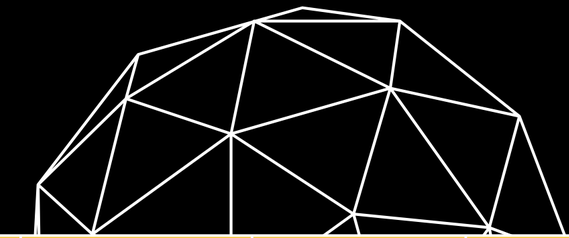
- It is also known as the Variance Ratio Criterion.
- Calinski Harabaz Index is defined as the ratio of the sum of between-cluster dispersion and of within-cluster dispersion.
- The higher the index the more separable the clusters.

Davies Bouldin Index

- The Davies Bouldin index is defined as the average similarity measure of each cluster with its most similar cluster, where similarity is the ratio of within- cluster distances to between- cluster distances.
- The minimum value of the DB Index is 0, whereas a smaller value (closer to 0) represents a better model that produces better clusters.

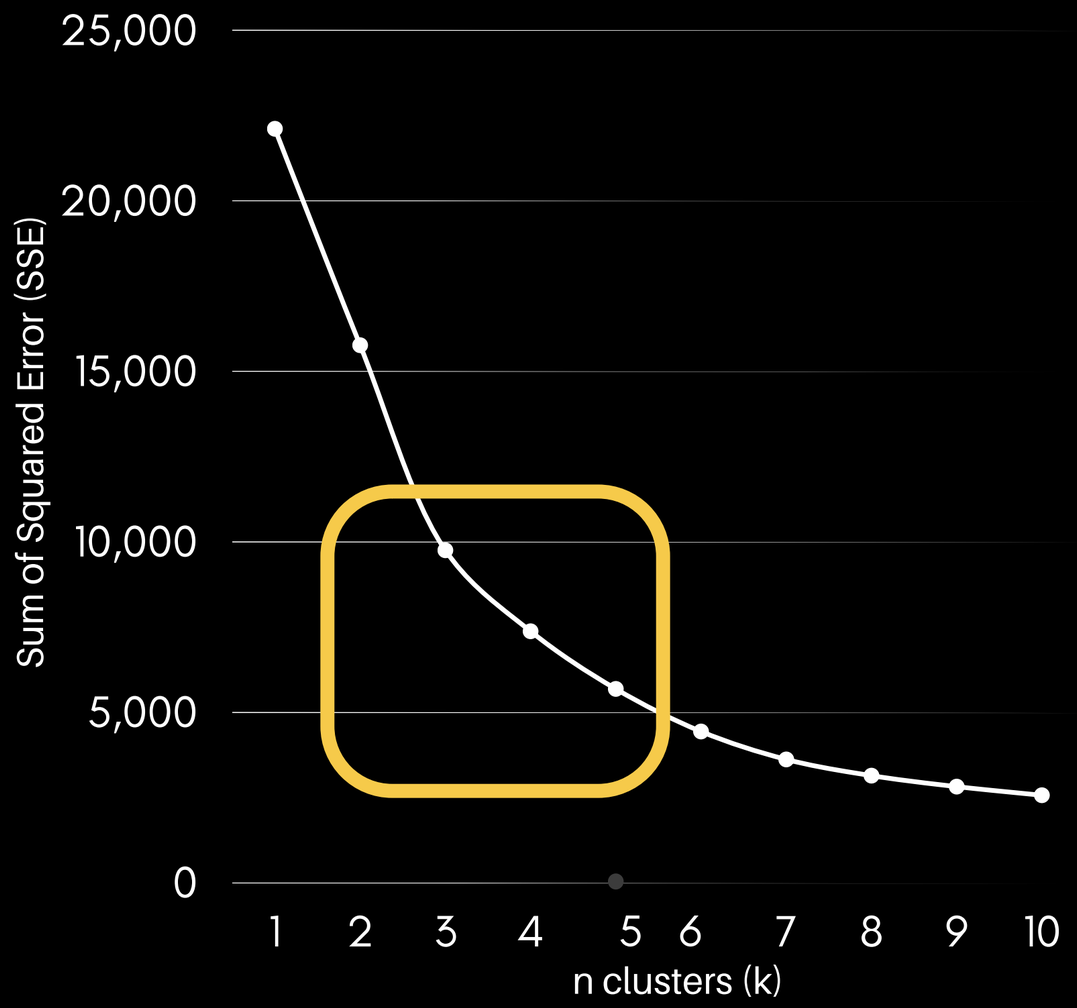
Segmentation

1. After comparing Sillhouette Score, Calinski Harabaz Index, and Davies Bouldin Index, of each `k` value from 2 to 10. We obtain that the best `k` value for our model to perform with our data is 7.
2. However, ***we will choose for `k`*** *value to be 5* instead as we were able to obtain a higher Silhouette Score and Calinski Harabaz index, indicating that the clusters were better separated.
3. However, we had to accept a higher Davies Bouldin index in the process, which was a reasonable trade-off.



Calinski n clusters (k)	Silhouette Score	Harabaz Index	Davies Bouldin Index	Rank
2	0.943753	2964.507537	0.730553	4.0
3	0.539512	4664.744834	0.712943	7.0
4	0.539058	4523.797471	0.696567	5.0
5	***0.56906***	***5106.421061***	***0.664354***	***2.0***
6	*7* 0.469337	5017.520620	0.669621	3.0
7	0.483411	6256.275645	0.640617	1.0
8	0.391012	5875.310103	0.721913	6.0
9	0.395776	5774.781187	0.779827	9.0
10	0.395619	5310.436881	0.736784	8.0

Elbow Method

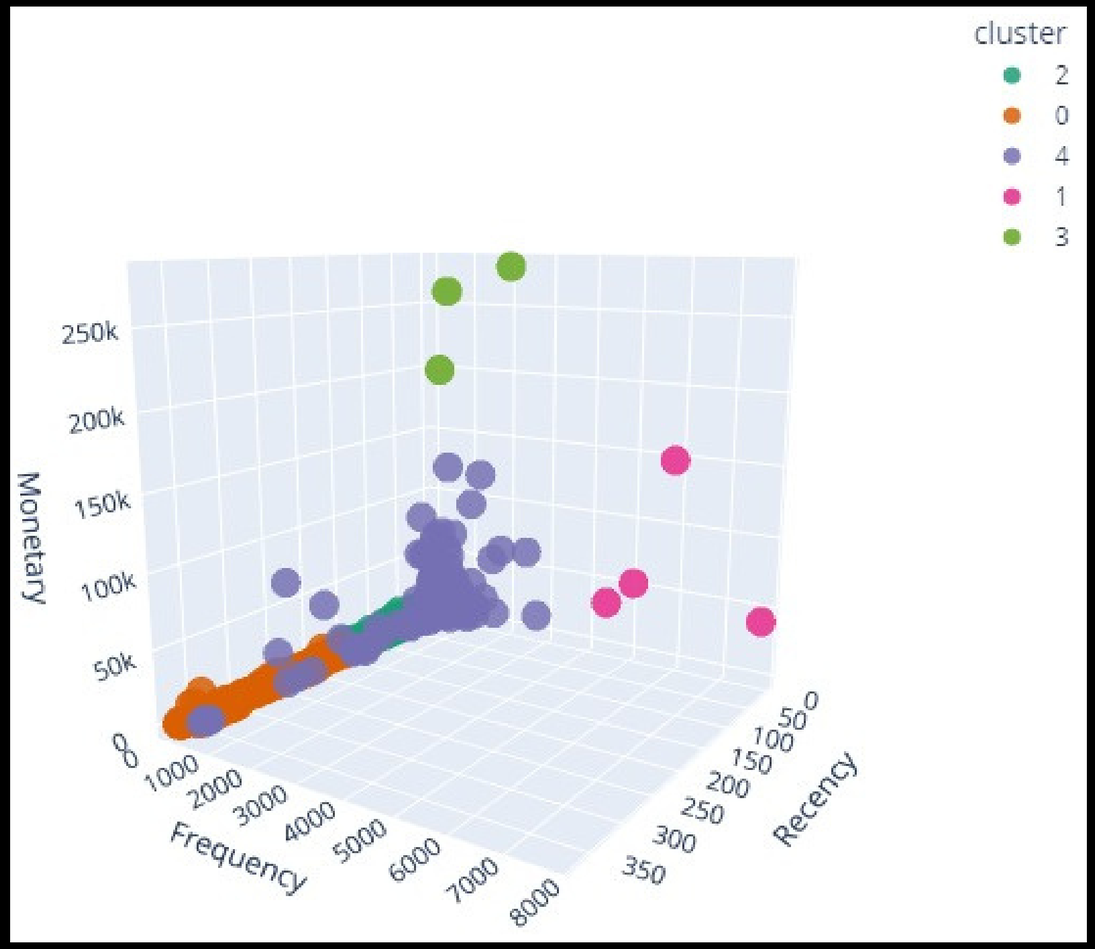


Segmentation

It is debatable that whether the `elbow point` fall in the `k` value of 3 or 5

Since `k` value of 3 ranked 7 out of 9, we decided to go through with `k` value of 5

Segmentation Result



Metrics	Statistics	Cluster				
		0	*1*	*2*	*3*	*4*
Recency	count	2655	4	4383	3	325
	mean	254.79	1.5	51.22	3	31.56
	std	64.2	1.73	41.58	4.36	68.24
	min	151	0	0	0	0
	25%	200	0.75	17	0.5	3
	50%	249	1	40	1	10
	75%	308	1.75	77	4.5	24
	max	373	4	158	8	373
Frequency	count	2655	4	4383	3	325
	mean	30.25	5807	60.84	945.33	494.39
	std	60.37	1455.46	71.45	971.62	301.36
	min	1	4440	1	339	4
	25%	1	4931.25	6	385	343
	50%	7	5494	34	431	434
	75%	31	6369.75	88.5	1248.5	567
	max	535	7800	357	2066	2755
Monetary	count	2655	4	4383	3	325
	mean	380.49	68912.73	1036.96	246120.46	11292.61
	std	849.15	53205.23	1407.66	41822.66	16153.7
	min	0	33481.57	0	199206.05	1070.47
	25%	0	39218.26	104	229431.68	3997.47
	50%	118.95	47194.46	554.85	259657.3	6105.37
	75%	409.78	76888.93	1429.86	269577.66	10892
	max	18745.86	147780.42	14545.85	279498.02	125490.88

Segmentation Result

0 Churned Customers

- Have not make purchases for a long time
- Rarely make purchases
- Generates low revenues individually 36% of total customers

2 New Customers

- Purchase recently
- Low amount of purchases
- Generates low revenues 59% of total customers

4 Potential Loyalist

- Purchase recently
- Moderate to frequent purchases
- Low to moderate revenues 4.4 % of total customers

1 Loyal Customers

- Purchase recently Frequent purchases
- Moderate to high revenues 0.054% of total customers

3 Big Spender

- Purchase recently
- Low amount but heavy purchases
- Generates huge revenues 0.047% of total customers

Insights and Recommendations

General Recommendations

Personalized Offers

General Recommendations

Create Customer Membership Card

Creating a membership ID can bring significant benefits to Everything Plus, such as improving customer experience, building stronger relationships with customers, and increasing revenue.

By capturing demographic data such as location, gender, age, income, and interests, we can gain insights into the preferences and behaviors of our customers, which can help us tailor our offers and promotions to their needs and interests.

With this additional data, we can:

Personalize offers even further

With demographic data, we can tailor offers and promotions based on location, age, income, and interests.

For example, we can offer discounts on products that are popular in a specific location or offer promotions on products that are popular with a specific age group.

Increase customer engagement and loyalty

With demographic data, we can send targeted emails, social media posts, or ads that resonate with specific customer groups.

For instance, we can create ads on social media platforms that target young adults who live in the city and are interested in eco-friendly products.

Improve customer experience

With demographic data, we can identify and address pain points in the customer journey.

For example, if we notice that a particular demographic group is not responding well to our marketing efforts, we can investigate and improve the quality of our offerings or tailor them better to their needs.

Personalized Offers

Cluster 0 - Churned Customers

Since these customers have not made a purchase for an extended period, it's unlikely that they will become active customers again. However, there are still some ways to generate revenue from this group

- Offer them a ***special promotion*** or discount to encourage them to make a purchase.
- Use their data to understand ***why they churned*** and use that information to improve the user experience for future customers.
- Consider using ***re-engagement*** tactics such as email campaigns or retargeting ads to remind them of the benefits of shopping with Everything Plus.

Personalized Offers

Cluster 2 - New Customers

Since these customers are new to Everything Plus, it's important to provide them with a great first impression to encourage repeat purchases

- Offer them a ***personalized discount*** or promotion as a welcome gesture.
- Provide them with a seamless onboarding experience to make their first purchase as easy and enjoyable as possible.
- ***Follow up*** with them after their first purchase to ask for ***feedback*** and offer further assistance if needed.
- Consider using ***retargeting ads*** to remind them of the benefits of shopping with Everything Plus and encourage repeat purchases.

Personalized Offers

Cluster 4 - Potential Loyalist

This cluster has the potential to become loyal customers, so it's important to provide them with the right incentives to encourage repeat purchases

- Offer them a ***personalized discount*** or promotion to encourage them to make more purchases.
- Use their data to understand their preferences and personalize their shopping experience.
- Consider introducing a ***loyalty program*** to incentivize them to make more purchases and become loyal customers.

Personalized Offers

Cluster 1 - Loyal Customer

It's important to maintain a strong relationship with this group of customers to keep them satisfied and encourage repeat purchases

- Offer them ***personalized rewards*** or discounts as a way of thanking them for their loyalty.
- Consider introducing a ***loyalty program*** to incentivize them to make more purchases.
- Provide them with ***early access*** to new products or ***exclusive offers*** to make them feel valued.
- Use their data to understand their ***preferences*** and ***personalize*** their shopping experience.

Personalized Offers

Cluster 3 - Big Spender

This group of customers generates a lot of revenue individually, so it's important to keep them satisfied and encourage repeat purchases

- Offer them ***personalized rewards*** or discounts as a way of thanking them for their high level of spending.
- Consider introducing a ***VIP program*** with exclusive perks such as early access to sales or free shipping.
- Use their data to understand their preferences and personalize their shopping experience.
- Offer them a wide range of ***high-end products and services*** that cater to their needs and interests.

Thank You!