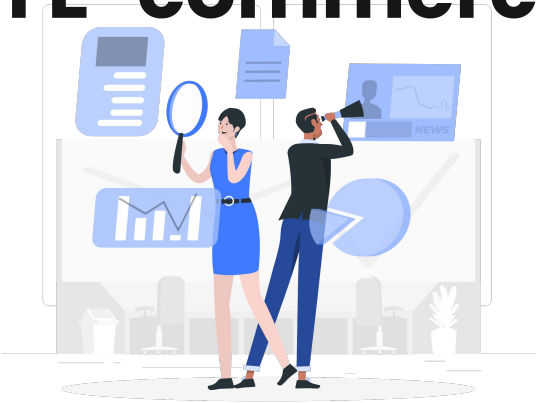
Data-Driven E-commerce Strategies



Leveraging Analytics from Exploration to Segmentation

By Riza Arfiqih

Executive Summary

Business Background

Kestrel E-commerce, a rising star in the e-commerce industry, has experienced rapid growth since its inception in 2018, surpassing 30,000 customers and 50,000 transactions in its inaugural year. They faces challenges in sales consistency, customer acquisition and retention, and customer understanding. To improve performance, the company aims to identify seasonal trends, optimize sales, enhance customer strategies, and personalize marketing.

Problem Definition

The business faces challenges in maintaining consistent sales and profits, acquiring and retaining customers, and understanding customer preferences to optimize strategies.

Objectives

This project aims to leverage customer behavior data to answer key business questions and provide data-driven insights or actionable recommendations for optimizing marketing, enhancing customer retention, and elevate overall Kestrel E-commerce's performance in the coming year.

Key Insights

- Sales performance and profit gained have been quite volatile and has a positive trend throughout the year.
- There is a problem with low rate of customer retention. The average of frequency orders per customer during the year is only 1-2 times.
- 59% of customers contribute 80% of sales and only 36% of products perform well, boosting 80% of profit by year.
- There are 5 Clusters of Customer based on purchased behavior

Methodology

- Use EDA to conduct a thorough exploration of the transaction data to uncover sales patterns and trends.
- Reveal customer cohorts to analyze customer behavior based on their acquisition cohorts to understand customer retention rates.
- Identify the most significant factors (customers and products) driving sales and profit, to prioritize efforts for maximum impact with pareto chart.
- Segment customers based on their Recency, Frequency, and Monetary value using robust K Means clustering model for targeted customer engagement.

Recommendations

- Launch major campaigns, offer discounts, and run festive promotions in a peak seasons to boost sales. Then stimulate demand with targeted promotions and clearance sales.
- Target specific customer segments with tailored campaigns to improve customer retentions.
- Concentrate on the 36% of products that generate 80% of profit and evaluate the low-performing products and consider discontinuing or rebranding them.
- Develop targeted marketing strategies for each customer segment. Focus on At-Risk/Lost Customers, reach out with a sincere apology and offer to resolve any issues. Send targeted emails or WhatsApp highlighting new products or offer limited time discounts and special promotions.

Project Overview

Business Understanding

Business Context: Kestrel E-commerce, a rising star in the e-commerce industry, has experienced rapid growth since its inception in 2018, surpassing 30,000 customers and 50,000 transactions in its inaugural year. During the year the business faces month-to-month variations in sales and profit margins, hindering consistent performance. Struggles to maintain a steady flow of new customers and retain existing ones, impacting long-term growth. And lacks a deep understanding of customers, limiting its ability to tailor strategies and optimize offerings. In today's competitive e-commerce landscape, understanding customer transaction data is crucial for success. To address these challenges, the company aims to identify seasonal trends, optimize sales performance, improve customer acquisition and retention strategies, and tailor personalized marketing campaigns for better customer engagement.

Goals: This project aims to leverage customer behavior data to answer key business questions and provide data-driven insights or actionable recommendations for optimizing marketing, enhancing customer retention, and elevate overall Kestrel E-commerce's performance in the coming year.

Objectives:

EDA

Conduct a thorough exploration of the transaction data to uncover sales patterns and trends.

Pareto Analysis

Identify the most significant factors (customers and products) driving sales and profit, to prioritize efforts for maximum impact.

Cohort Analysis

Analyze customer behavior based on their acquisition cohorts to understand customer retention rates.

Customer Segmentation

Segment customers based on their Recency, Frequency, and Monetary value using robust K Means clustering model for targeted customer engagement. It will provide a foundation for personalized marketing and customer relationship management strategies

The dataset describes the first year of transaction history of Kestrel E-commerce in the United States. Consist of 16 features and > 50.000 trx rows

First Transactions 2018-01-01 | Last Transactions 2018-12-30
Day of Transactions 356 Days

#	Column	Non-Null Count	Dtype
0	Order_Date	51282 non-null	object
1	Time	51282 non-null	object
2	Aging	51282 non-null	float64
3	Customer_Id	51282 non-null	int64
4	Gender	51282 non-null	object
5	Device_Type	51282 non-null	object
6	Customer_Login_type	51282 non-null	object
7	Product_Category	51282 non-null	object
8	Product	51282 non-null	object
9	Sales	51282 non-null	float64
10	Quantity	51282 non-null	float64
11	Discount	51282 non-null	float64
12	Profit	51282 non-null	float64
13	Shipping_Cost	51282 non-null	float64
14	Order_Priority	51282 non-null	object
15	Payment_method	51282 non-null	object

- **Order_Date** : The date the product was ordered.
- **Aging** : The time from the day the product is ordered to the day it is delivered.
- **Customer_Id** : Unique id created for each customer.
- **Gender** : Gender of customer.
- **Device_Type** : The device the customer uses to actualize the transaction (Web/Mobile).
- **Customer_Login_Type** : The type the customer logged in. Such as Member, Guest.
- **Product_Category** : The category of each products.
- **Product** : The description of products.
- **Sales** : Total sales amount (product price * quantity)
- **Quantity** : Unit amount of product.
- **Discount** : Percent discount rate.
- **Profit** : Profit.
- **Shipping_Cost** : Shipping cost.
- **Order_Priority** : Order priority. Such as critical, high etc.
- **Payment_method**: Payment method.

The dataset is clean!

0

Missing Values

No missing values in the dataset

0

Duplicated

No duplicate data

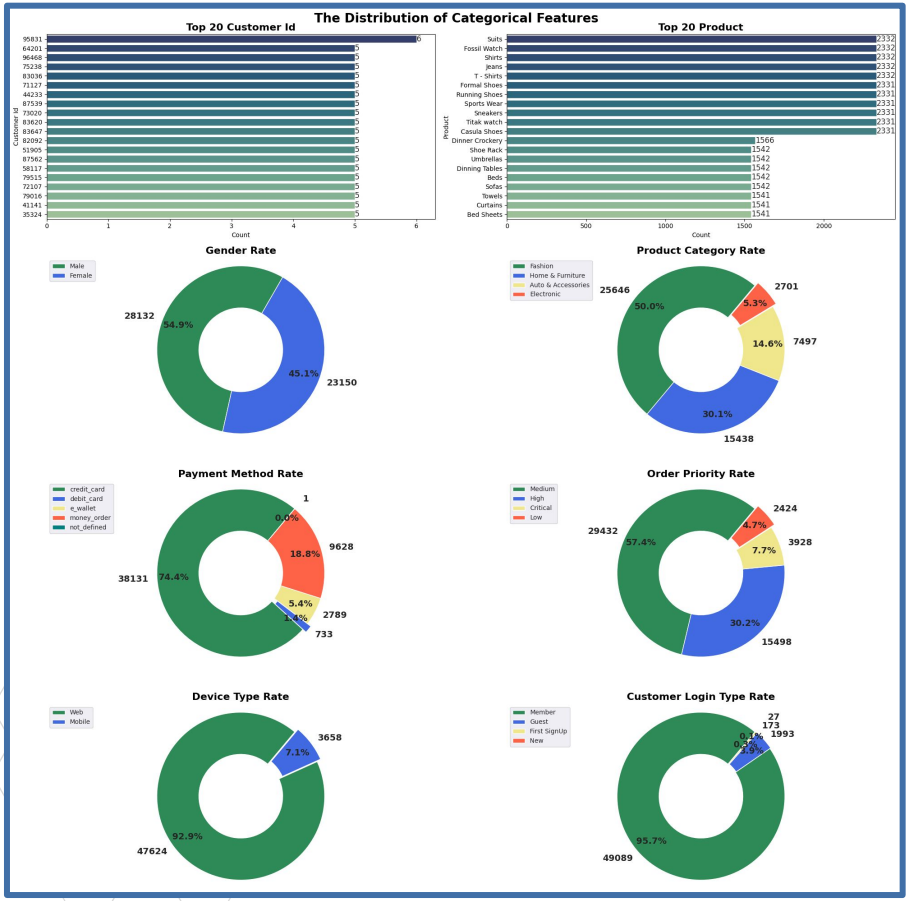
2

Data Transform

Change data type and create new seasonality features, parse the date time into one single feature

Exploratory Data Analysis

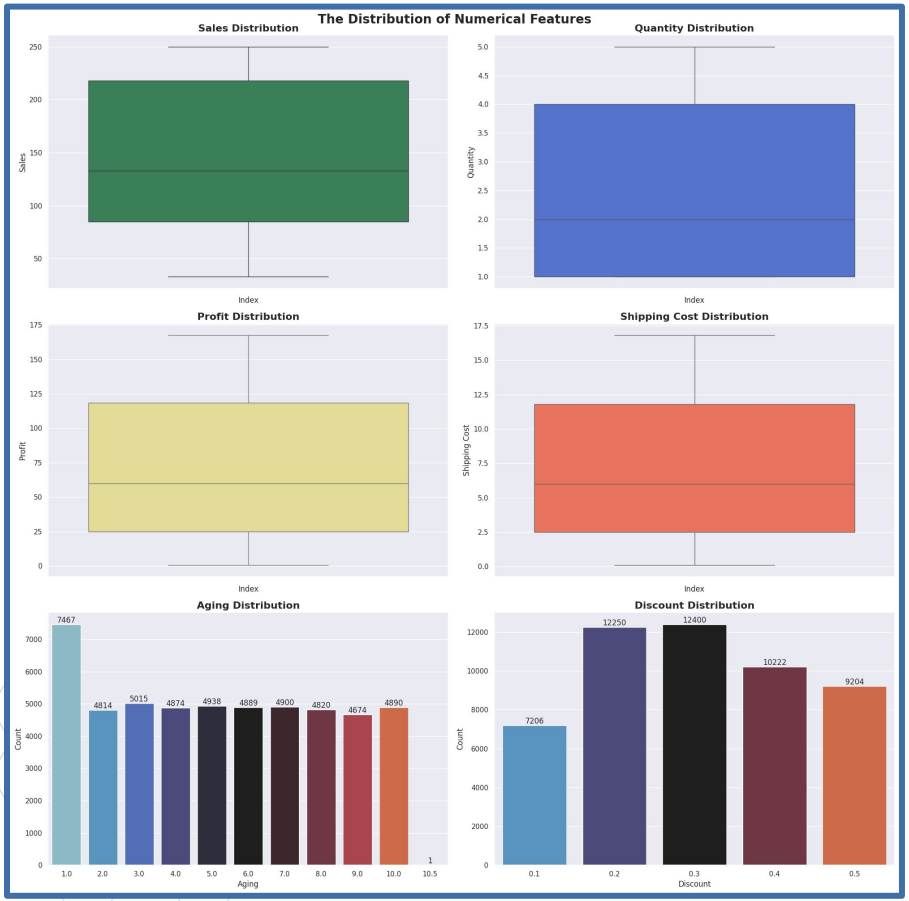
The plot provides insights into the **distribution** of every **categorical features**. It seems almost every feature has dominant values such as Fashion in Product Category.



- **Top 20 Customer** : Shows the top 20 customer IDs based on purchase frequency.
- **Top 20 Product** : Displays the top 20 products. "Suits" and "Fossil Watch" seem to be the top-selling products.
- **Gender Rate** : The distribution of customers by gender. It appears that males make up a slightly larger proportion of customers (54.9%) compared to females (45.1%).
- **Product Category Rate** : The distribution of products across different categories. "Fashion" and "Home & Furniture" are the dominant categories.
- **Payment Method Rate** : The distribution of payment methods used by customers. "Debit Card" is the most popular method, followed by "Credit Card."
- **Order Priority Rate** : The distribution of order priorities. "Medium" and "High" priorities seem to be the most common.
- **Device Type Rate** : The distribution of devices used for purchases."Web" is the most common device type, followed by "Mobile."
- **Customer Login Type Rate** : The distribution of customer login types. "Member" is the most common login type, followed by "Guest."

Exploratory Data Analysis

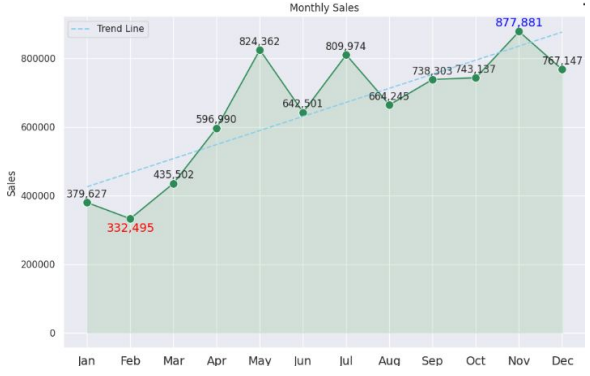
The plot of every numerical feature provides insights into the distribution of sales, quantities, profits, shipping costs, aging, and discounts.

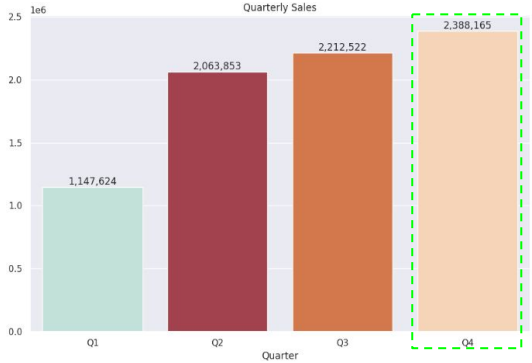


- **Sales Distribution :** Box-plot shows the distribution of sales values. The median sales value appears to be around 150, with the majority of sales falling between 100 and 200.
- **Quantity Distribution** : The distribution of quantities. The median quantity appears to be around 2, with the majority of quantities falling between 1 and 4.
- **Profit Distribution** : The distribution of profit values. The median profit appears to be around 55, with the majority of profits falling between 25 and 120.
- **Shipping Cost Distribution :** The distribution of shipping costs. The median shipping cost appears to be around 5.5, with the majority of costs falling between 2.5 and 12.
- **Aging Distribution** : This bar chart shows the aging distribution. The aging values appear to range from 1 to 10 days, with the highest frequency occurring at 1 day.
- **Discount Distribution** : The distribution of discount values. The discount values appear to range from 0.1 to 0.5, with the highest frequency occurring at 0.3 (30% discount).

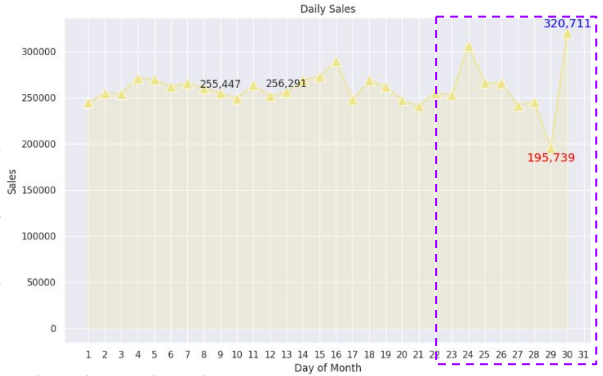
Exploratory Data Analysis

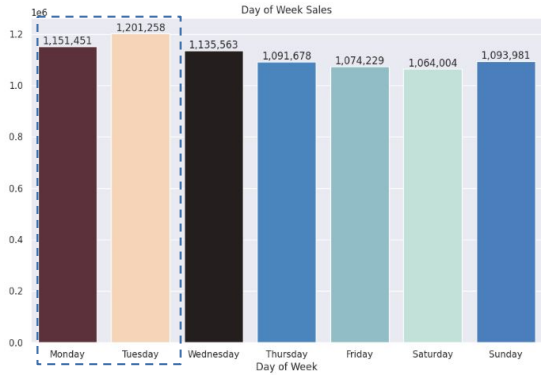
Based on the seasonality sales trend, the **sales are highest in November and May**, with a **dip in February**. Across four quarters, sales are **highest in Q4** and **lowest in Q1**. Then sales appear to fluctuate with **a peak around week 17**.

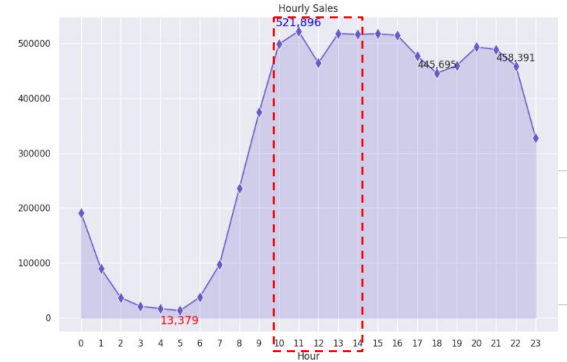




Sales on a **peak around the 24th** and **30th day** of the month, but **dip drastically in 29th**. Also **highest sales on Tuesday** and **Monday**, with a **dip on Saturday**. In the **middle of afternoon** and **late evening are peak of sales**, with a **dip in the early morning**.

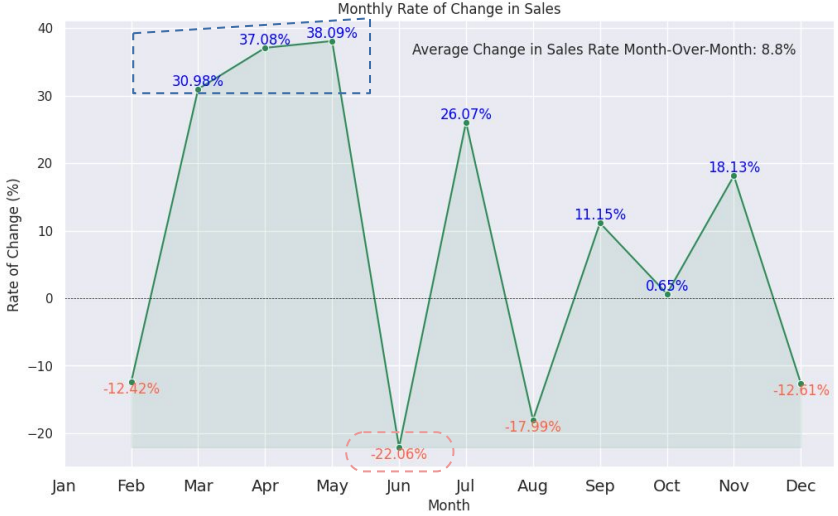




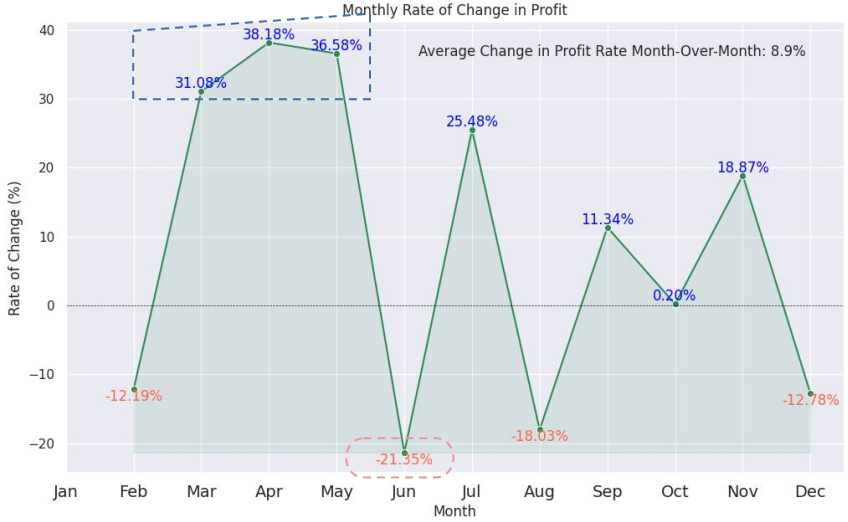


Exploratory Data Analysis

Overall, the sales performance and profit gained have been quite volatile throughout the year.



VS



Strong Growth Periods:

March: Sales experienced a significant increase of 30.98% compared to the previous month.
April: The momentum continued with an even higher growth rate of 37.08%.
May: Sales further accelerated, reaching a peak with a 38.09% increase.
July: After a dip in June, sales rebounded with a 26.07% increase.
November: Sales saw a moderate growth of 18.13%.

Decline Periods:

February: Sales declined by 12.42% compared to the previous month.
June: A significant drop of 22.06% occurred.
August: Sales decreased by 17.99%.
December: The year ended with a decline of 12.61%.

Average Change in Sales Rate Month-Over-Month: 8.8%

Strong Growth Periods:

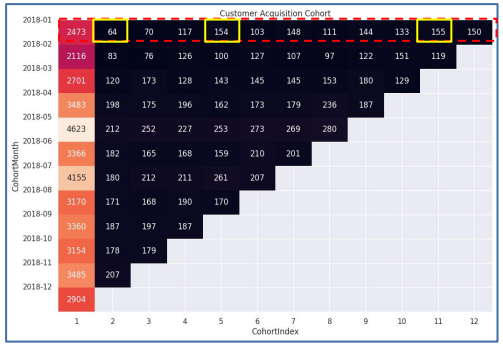
March: Profit significant increase of 31.08% compared to the previous month.
April: Profit further accelerated, reaching a peak with a 38.18% increase.
May: The momentum break with a slightly lower growth rate of 36.58%.
July: After a dip in June, profit rebounded with a 25.48% increase.
November: Profit saw a moderate growth of 18.87%.

Decline Periods:

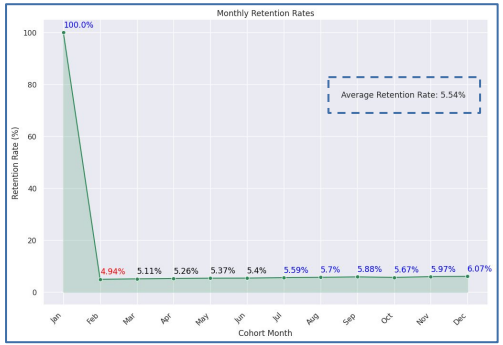
February: Profit declined by 12.19% compared to the previous month.
June: A significant drop of 21.35% occurred.
August: Profit decreased again by 18.03%.
December: The year ended with a decline of 12.78%.

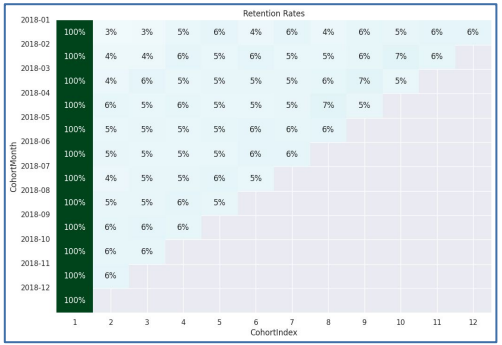
Average Change in Profit Rate Month-Over-Month: 8.9%

In the Cohort Table, we had insights that **2.473 customers made transactions in January**, but **in February, only 64 of the same customers retained and made transactions**. Then had a minor increased again in May (154 customers) and November (155 customers).



The average customer retention rate is 5.54%. This means, **on average, 5.54% of customers acquired in a specific month are still active in the following months.** The rate seems logical to be a low trend because the average of frequency orders per customer during the year is only 1-2 times, and the maximum order is 6 times with only one customer.



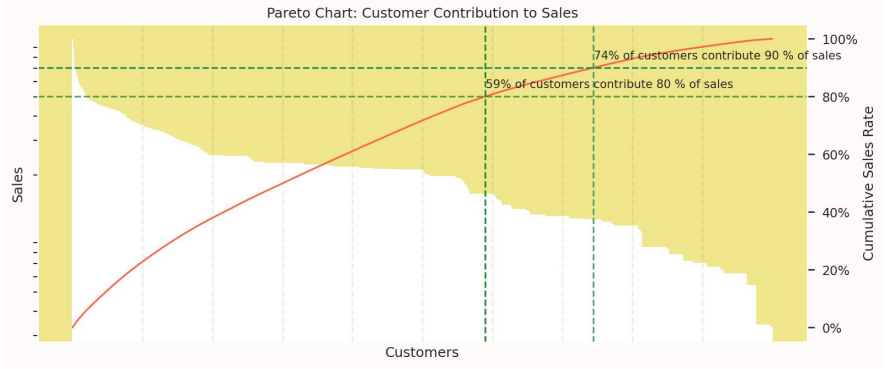


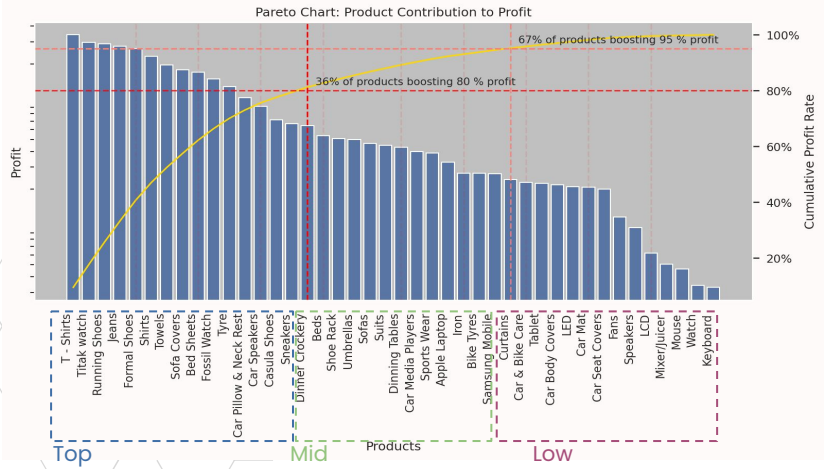
Pareto Analysis

Identify the most significant factors like customers and products driving sales and profit with the pareto charts. Then consider them to prioritize efforts for maximum impact.

Well, from the chart, we can get insights that with a **half amount of customers 59 % (22.994)**, it contribute **80 % of Sales**. While the rest of 41 % (15.996) seem to have a low contribution. That's a good start, we'll further analyze the customer segment based on purchase behavior later.

ParetoCustValue	TotalCustomer	TotalSales
High Value	22.994	6.249.292.00
Low Value	15.996	1.562.872.00





36 % of products can boost 80 % of profit by year. Then **95 % of profit was boosted by 67 % of products**, which means **37 % of products had a poor contribution**. Concentrate on the 36 % of top products because it has a high profit margin and also evaluate the low-performing products when the profit margin is higher than mid-performing, consider not discontinuing them.

ParetoProdValue	TotalProduct	TotalProfit	AvgProfitMargin
Top Performance	15	2.868.470.80	50.49
Mid Performance	13	545.109.10	32.29
Low Performance	14	196.765.00	34.12

Methodology

Step 1 Feature Engineering
Create RFM Metrics

Step 2 Feature Scaling
Standardization

Step 3 K Means Clustering

Step 4 Find the Best K

Step 5 Cluster Visualization
& Analysis

Customer Segmentation

Build the K Means clustering model with 3 features according to RFM.

Why using RFM?

- RFM (Recency, Frequency, Monetary) Analysis is a segmentation technique that uses customer purchase behavior to group them into segments. And very effective as the purchase behavior can be summarized by using a small number of variables.
- It is also very useful for targeted marketing campaigns and other customer relationship management initiatives.

Why using K-means?

- K-means is a effective clustering algorithm for customer segmentation and RFM analysis due to its simplicity, efficiency, and interpretability.
- Computationally efficient algorithm, suitable for clustering large datasets.
- Can be relied on to determine appropriate value for 'k' (the number of clusters) using techniques like the elbow method or silhouette analysis.

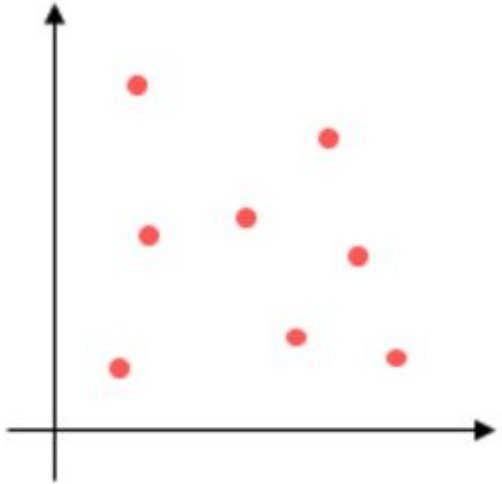
Define the RFM features as follows:

- **Recency (R)**: Using Order_Date feature, calculate the number of days since the customer's last purchase.
- **Frequency (F)**: Using Order_Date feature, calculate the number of total order date for each customer. This represents how often the customer has made purchases.
- **Monetary (M)**: Using Sales feature, calculate the total sales for each customer. This represents the total value of the customer's purchases.

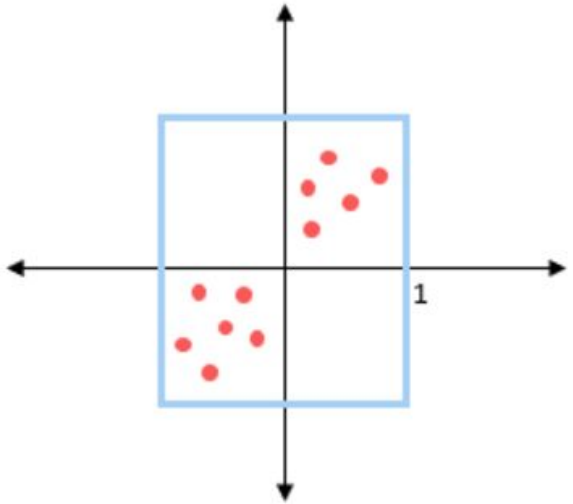
Customer_Id	Recency	Frequency	Monetary
10000	24	2	293.00
10002	132	1	149.00
10004	144	1	85.00
10006	144	2	440.00
10013	227	1	159.00

Feature Scaling

Ensuring that Recency, Frequency, and Monetary (RFM) features are on a similar scale before applying the K Means algorithm. This will help KMeans create more meaningful and accurate customer segments based on their RFM behavior.



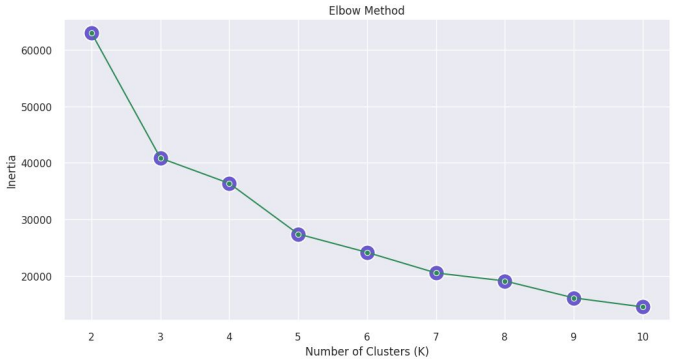
Actual Data

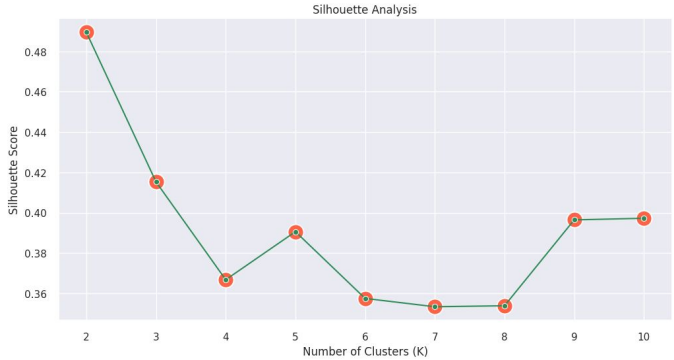


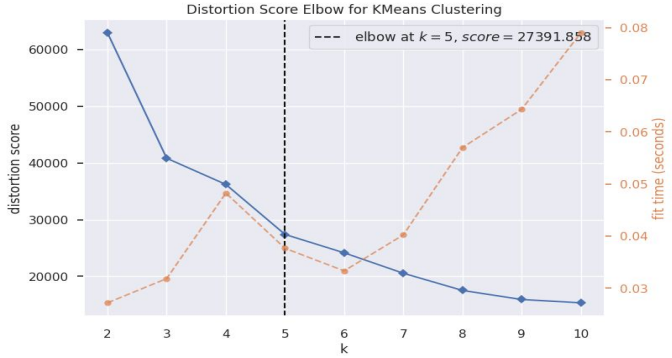
After Standardization

Customer Segmentation

Based on the **Elbow Method** and **Silhouette Analysis** suggest 5 clusters for better interpretation and more effective customer segmentation.

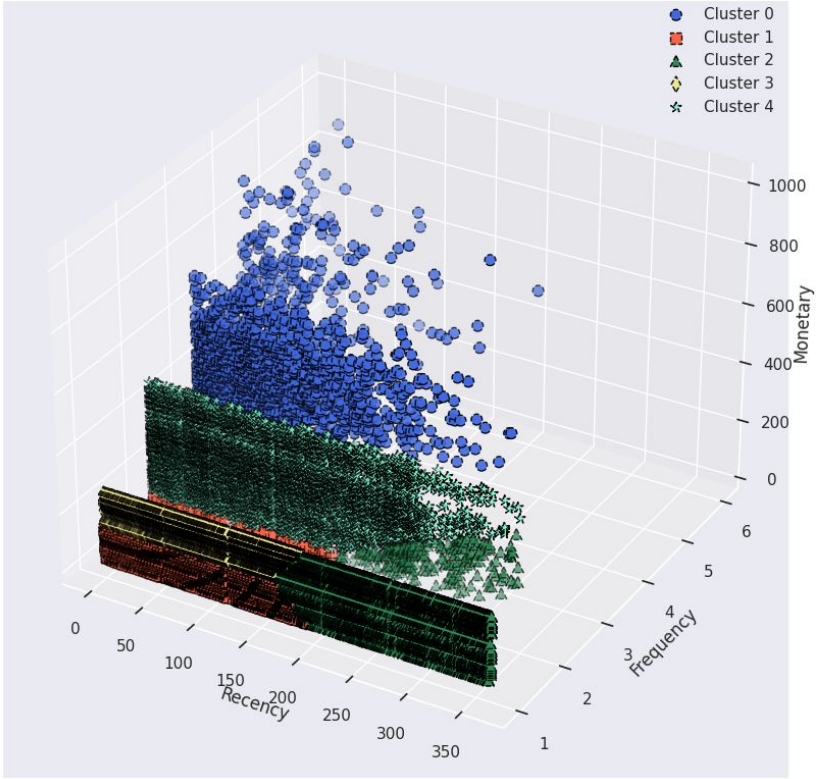




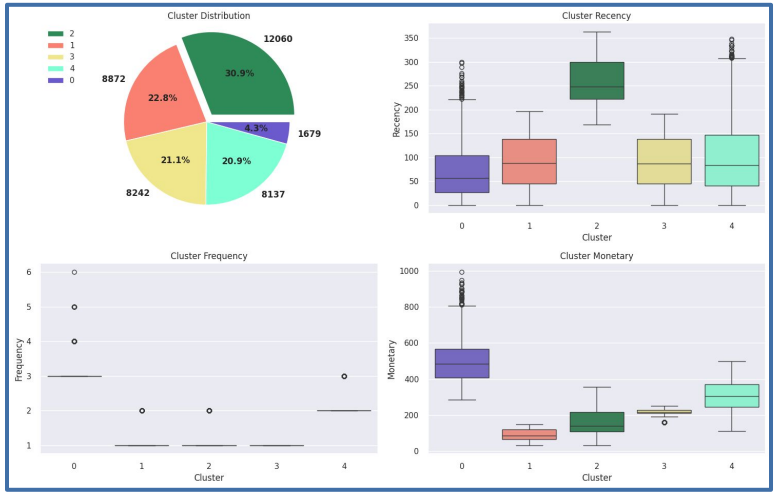


Customer Segmentation

This diagram shows the distribution of customers which is divided into clusters according to the K-Means Clustering algorithm.



Customer Segmentation



Percentile	Recency	Frequency	Monetary
min	0.00	1.00	33.00
10%	23.00	1.00	70.00
20%	48.00	1.00	109.00
30%	72.00	1.00	122.00
40%	99.00	1.00	159.00
50%	133.00	1.00	207.00
60%	163.00	1.00	218.00
70%	198.00	1.00	228.00
80%	236.00	2.00	250.00
90%	287.00	2.00	354.10
max	363.00	6.00	994.00

High <= 40% ←

Moderate <= 50% ←

Low >= 80% ←

→ Low <= 40%

→ Moderate >= 50%

→ High >= 80%

Clusters breakdown and their potential customer segments based on percentiles to rank the customers from the RFM metrics to identify behavior patterns:

Cluster	Customer	High	Moderate	Low
0	1.679	R,F,M	x	x
1	8.872	x	R,F	M
2	12.060	x	x	R,F,M
3	8.242	x	R,F,M	x
4	8.137	F,M	R	x

Customer Segmentation

Cluster 0

High recency, high frequency, high monetary

Champions Customers

Represents the best customers, they are the company's most valuable assets. They purchase frequently, spend a lot of money, and have made purchases very recently.

Cluster 1

Moderate recency, moderate frequency, low monetary

Need Attention Customers

These customers have purchased recently but have only made one purchase. They need encouragement for repeat purchases and more spending.

Cluster 4

Moderate recency, high frequency, high monetary

Loyal Customers

They have a high purchase frequency and high spending, even though their recent activity is moderate.

Cluster 2

Low recency, low frequency, low monetary

At-Risk/Lost Customers

They haven't purchased recently and might be churned. It's the biggest problem for business.

Cluster 3

Moderate recency, moderate frequency, moderate monetary

Promising Customers

They have a moderate purchasing history but haven't purchased recently. Treat them well then I promise they will be loyal.

Conclusion

To summarize, Kestrel E-commerce has identified key areas for improvement, including customer retention, product performance, and targeted marketing strategies.

- Kestrel E-commerce's sales performance and profit gained have been **quite volatile** and has a **positive trend throughout the year**.
- However, they have a problem with **low rate of customer retention**. The average of frequency orders per customer during the year is **only 1-2 times**. It'll be validated by the dominance of At-Risk/Lost Customers in segmentation.
- **59% of customers** contribute 80% of sales where they have high to moderate spending purchase throughout the year and it could be the High-Value customers. And only **36% of products perform well**, boosting 80% of profit by year, meanwhile low-performing products like Mouse, Watch, Keyboard needs evaluation to sell better.
- There are 5 Clusters of Customer based on RFM/purchased behavior:
    - **At-Risk/Lost Customers** 30.93%
    - **Need Attention Customers** 22.75%
    - **Promising Customers** 21.14%
    - **Loyal Customers** 20.87%
    - **Champions Customers** 4.31%

Recommendation

Key Takeaways to Boost Sales:

- **Peak Seasons**: Launch major campaigns, offer discounts, and run festive promotions.
- **Off-Peak Seasons**: Stimulate demand with targeted promotions and clearance sales.
- **Day-of-Week and Time-of-Day**: Optimize marketing efforts and offers for peak times.
- **Seasonal Stock Planning**: Adjust inventory levels to meet demand fluctuations.
- **Demand Forecasting**: Utilize historical data to accurately forecast future demand.

Improve Customer Retention Strategies:

- **Personalized Marketing**: Target specific customer segments with tailored campaigns.
- **Enhanced Customer Experience**: Improve website/apps usability, customer service, and brand interactions.
- **Loyalty Programs**: Reward repeat purchases with points, discounts, or exclusive offers.
- **Product and Service Innovation**: Continuously introduce new offerings and incorporate customer feedback.

Optimize Resource allocation and Improve Profitability:

- **Identify High-Value Customers**: Focus on the top 59% of customers who drive 80% of sales.
- **Personalized Marketing**: Tailor marketing efforts to the needs and preferences of high-value customers.
- **Prioritize Top-Performing Products**: Concentrate on the 36% of products that generate 80% of profit.
- **Product Line Review**: Evaluate the low-performing products and consider discontinuing or rebranding them.

Recommendation

Develop targeted marketing strategies for customer segment:

- **At-Risk/Lost Customers:**
  - **Win-Back Campaigns**: Offer limited-time discounts or special promotions.
  - **Personalized Communication**: Reach out with a sincere apology and offer to resolve any issues.
  - **Re-Engagement**: Send targeted emails or WhatsApp highlighting new products or promotions.
  - **Social Media Engagement**: Interact with them on social media to rekindle interest.
- **Need Attention:**
  - **Also personalized campaigns:** email or WhatsApp highlighting new products or special offers.
  - **Social Media Engagement**: Address campaigns to spark interest.
  - **Offer tiered rewards** based on purchase frequency.
  - **Implement a subscription** model for repeat purchases.
- **Promising:**
  - **Targeted Marketing**: Use email marketing and WhatsApp to promote relevant products.
  - **Cross-Selling and Upselling**: Suggest complementary products or higher-tier options.
  - **Personalized Offers**: Offer discounts or special deals to incentivize repeat purchases.
  - **Customer Surveys**: Gather feedback to understand their needs and preferences.
- **Loyal Customers:**
  - **Personalized Marketing**: Utilize data to tailor promotions and offers.
  - **Loyalty Programs**: Implement tiered loyalty programs with exclusive rewards.
  - **Customer Advocacy**: Encourage customer reviews and referrals.
  - **Early Access**: Provide early access to new products or sales.
- **Champions:**
  - **VIP Treatment:** Provide exceptional customer service and exclusive benefits.
  - **Co-Creation:** Involve them in product development or marketing campaigns.
  - **Personalized Experiences**: Tailor experiences to their individual preferences.
  - **Exclusive Events**: Invite them to special events or product launches.

Thanks!

Let's Collaborate!

 Email

 LinkedIn

 GitHub