INTRODUCTION

1.1 Background

The analysis of Amazon sales data in India focuses on the importance of leveraging data to support strategic decision-making in the e-commerce business. With the rapid growth of the online market in India, this analysis encompasses sales analysis, product performance analysis, and customer analysis. The primary objective is to identify products, regions, and categories that significantly contribute to revenue, as well as to understand customer preferences in the Indian market. By delving into this data, the analysis aims to provide strategic insights to enhance Amazon's sales performance in India, design more effective promotions, and optimize the customer experience.

1.2 Purpose

1. Analyzing Sales:
   Identifying sales trends based on time (daily and monthly) to understand Amazon's sales patterns in India.
2. Evaluating Product Performance:
   Measuring the contribution of products and categories to total revenue to determine the best-performing products/categories.
3. Understanding Customer Behavior:
   Segmenting customers based on total purchases to develop more targeted marketing strategies and assess customer loyalty.
4. Analyzing Regional Performance:
   Identifying regions with the highest and lowest revenue contributions to uncover opportunities for market expansion.
5. Providing Strategic Recommendations:
   Developing data-driven recommendations to boost sales, design effective promotional campaigns, and enhance the customer experience.

1.3 Methodology

1. Data Collection:

   The dataset was sourced from Kaggle under the title “Amazon Sale Report.”

2. Data Preparation:

   The data was cleaned by addressing missing values, normalizing numerical data, and transforming time and category variables.

3. Analysis Methods:

   - Descriptive Analysis: Used to describe sales patterns, regional contributions, and product performance.
   - Clustering (K-Means): Applied to segment customers based on shopping behavior.

EXPLORATION OF DATA

2.1 Dataset Description

This dataset contains detailed information about Amazon sales in India, covering various aspects of transactions. Below is the description of each attribute:

1. Order ID: A unique identifier for each transaction.
2. Date: The date when the transaction occurred.
3. Status: The status of the order, such as Shipped or Cancelled.
4. Fulfilment: The method of order fulfillment, either by Amazon (Fulfilled by Amazon) or a third party (Merchant).
5. Sales Channel: The source of the sales, such as Amazon.in.
6. ship-service-level: The shipping service level, such as Standard or Expedited.
7. Category: The category of the product sold.
8. Qty: The number of units sold in each transaction.
9. Amount: The revenue generated from the transaction.
10. ship-city, ship-state, ship-country: Information about the shipping location (city, state, and country).
11. promotion-ids: Details of promotions applied to specific transactions.
12. B2B: Indicates whether the transaction was for individual customers or businesses (Business-to-Business).

Lastly, the data cleaning process involved imputing missing values. Missing values in the Amount column were addressed using the median, while missing values in ship.postal.code were handled using the mode to optimize the analysis.

2.2 Descriptive Analysis

Descriptive statistics for the key attributes of the dataset, Qty and Amount, are as follows:

1. Qty

Min (0.0000): There are transactions with no recorded units sold, possibly due to order cancellations.

1st Qu. (1.0000): The first 25% of transactions involve the sale of 1 unit.

Median (1.0000): The middle value indicates that more than 50% of transactions involve only 1 unit sold.

Mean (0.9044): The average number of units sold per transaction is slightly below 1, influenced by transactions with zero units sold.

3rd Qu. (1.0000): Up to 75% of transactions involve the sale of 1 unit.

Max (15.0000): There are transactions where 15 units were sold in a single transaction.

2. Amount

Min (0.0): There are transactions with zero revenue, caused by canceled orders.

1st Qu. (459.0): The first 25% of transactions generate revenue below ₹459.

Median (605.0): The median revenue indicates that 50% of transactions generate revenue below ₹605, while the other 50% generate higher revenue.

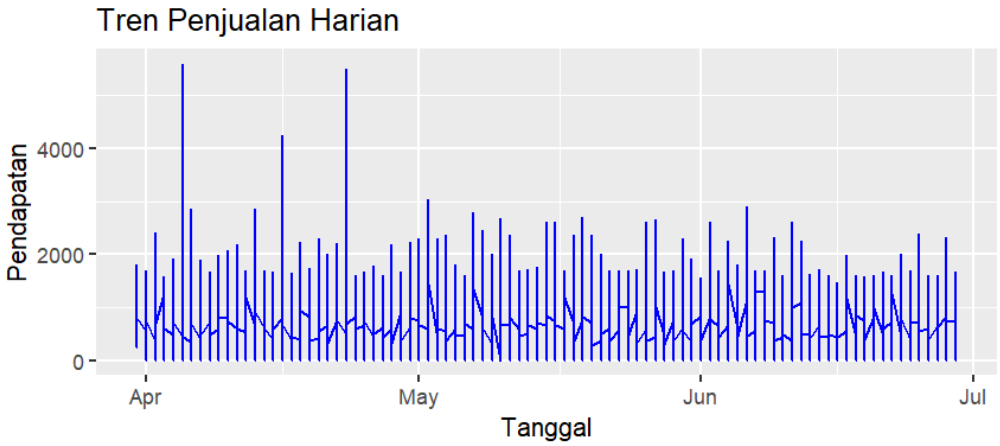
Mean (645.9): The average revenue per transaction is ₹645, suggesting that the distribution is slightly influenced by transactions with higher-than-average revenue.

3rd Qu. (771.0): 75% of transactions generate revenue below ₹771.

RESULTS

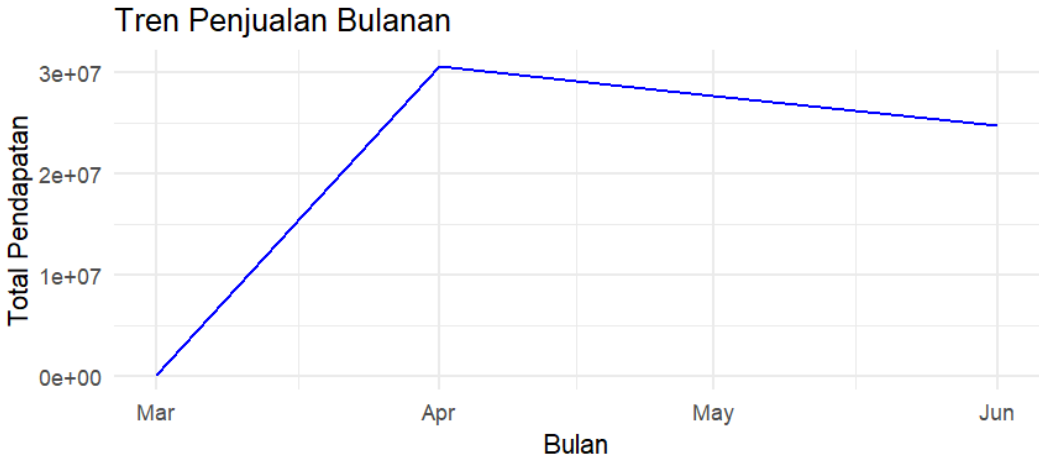
3.1 Sales Trend

Daily Sales Trend



- There are significant daily sales fluctuations, with some revenue peaks exceeding 4000 INR per day.
- Most days have moderate revenue (around 1000–2000 INR), but there are occasional days with sharp spikes.
- Fluctuations appear to be more intense in April and May compared to June.

Monthly Sales Trend

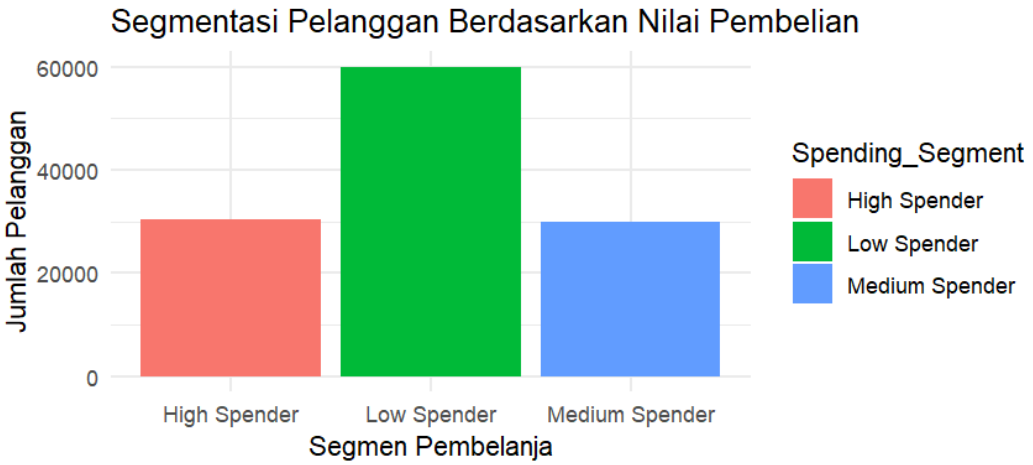


- March to April: There was a significant increase in total revenue. In April, revenue peaked at approximately 3+e07 or over 30 million INR.

- April to June: A gradual decline in total revenue was observed, indicating a downward trend following the peak in April.

3.2 Customer segmentation and loyalty

Customer Segmentation



1. Dominance of the Low Spender Segment

   The Low Spender segment has the largest number of customers, with over 60,000 customers. Customers in this segment tend to make small purchases.

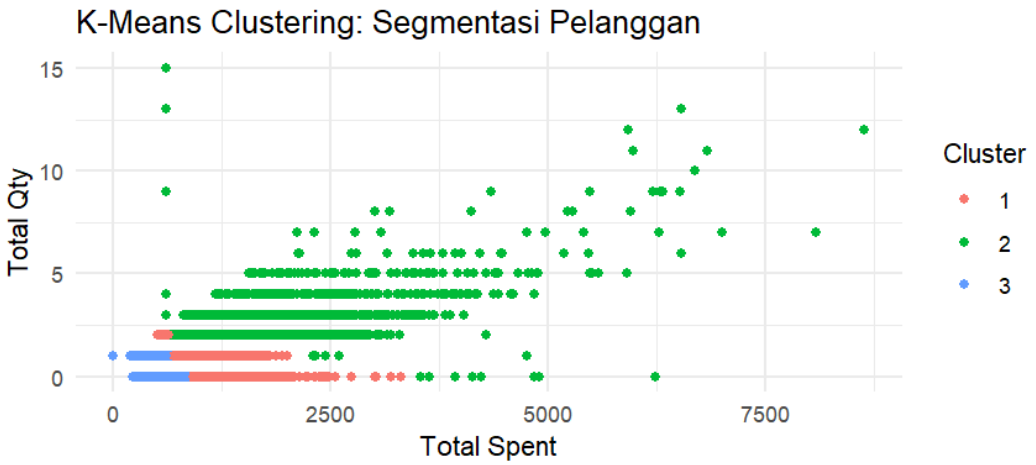
2. Medium Spender Segment

   The Medium Spender segment has fewer customers compared to the Low Spender segment but remains significant, with approximately 30,000 customers.

3. High Spender Segment

   The High Spender segment has the fewest customers, with less than 25,000 customers.

K-Means Clsutering



1. Three Customer Segments

   The chart shows that customers are grouped into three clusters based on Total Qty (Number of Units Purchased) and Total Spent (Purchase Value):

   - Cluster 1 (Red): Customers with very low Total Qty and Total Spent.
   - Cluster 2 (Green): Customers with moderate to high Total Qty and Total Spent.
   - Cluster 3 (Blue): Customers with very low Total Qty and small Total Spent.

2. Characteristics of Each Cluster

   - Cluster 1 (Low Buyers – Red):

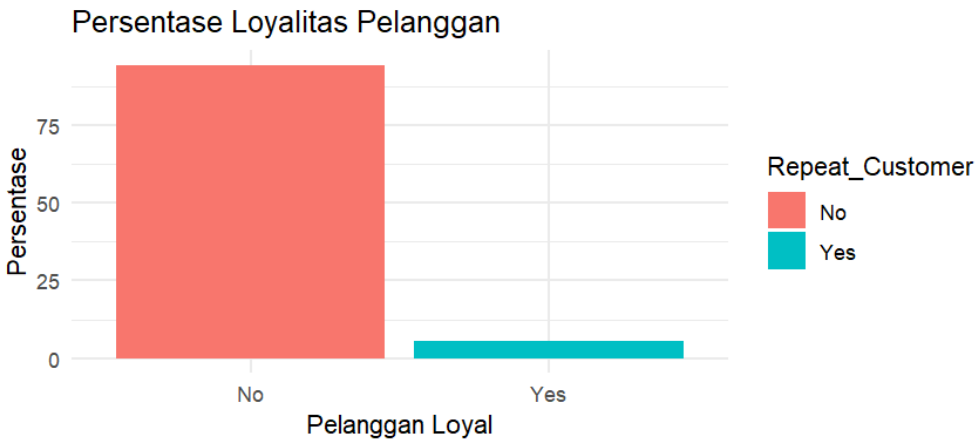
     Customers who rarely make purchases (low Total Qty) and have low Total Spent. These customers may include new buyers or those who only purchase low-value products.
   - Cluster 2 (Moderate to High Buyers – Green):

     Customers with higher purchase volumes (Total Qty) and greater spending. They are likely loyal customers or those making high-value purchases. This cluster significantly contributes to overall revenue.

- Cluster 3 (Low Qty, Low Spent – Blue):
  Customers with very low Total Qty and Total Spent, indicating minimal shopping behavior. They might be passive customers or those who only tried the service once.

3. Customer Distribution
   - Cluster 2 (Green) is the dominant cluster, showing that most customers fall into the moderate to high spending and purchase volume category.
   - Cluster 1 (Red) and Cluster 3 (Blue) have smaller distributions but represent opportunities for improvement.

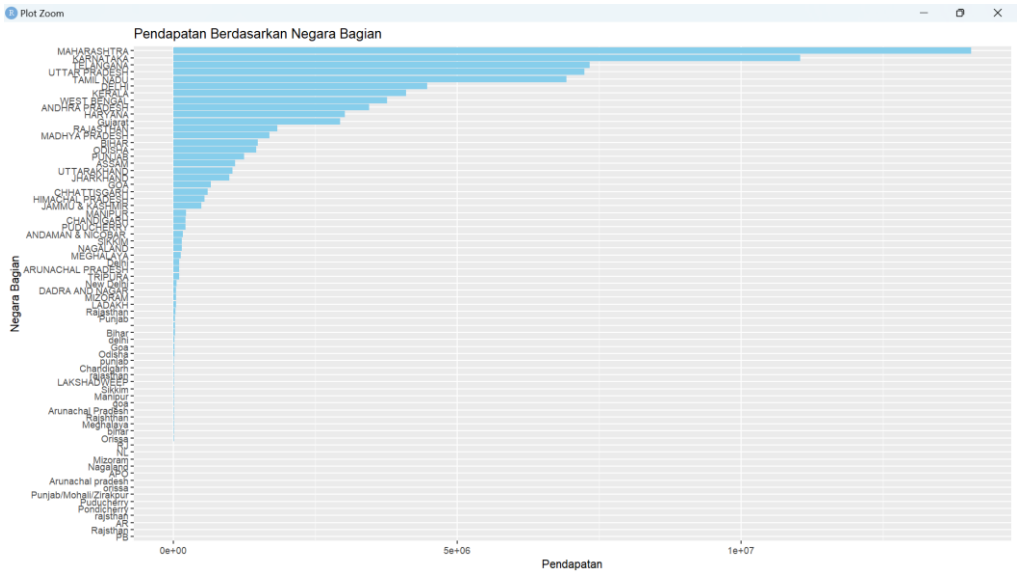
Customer Loyalty



1. Majority of Customers Are Not Loyal
   The majority of customers, over 75%, have only made a single purchase (categorized as "No"). This indicates that most customers have not made repeat purchases and are not considered loyal.
2. Loyal Customers Are Still Very Few
   Only less than 25% of customers fall into the "Yes" category (have made repeat purchases).

3.3 Region Performance

Region Performance



1. Regions with the Highest Revenue

   Maharashtra, Karnataka, Telangana, and Uttar Pradesh are the states with the highest revenue contributions.

2. Regions with Low Revenue

   Nagaland, Mizoram, Sikkim, Tripura, and several smaller states recorded very low revenue. Some regions contributed almost nothing to the total revenue.

3. Revenue Gap Between Regions

   There is a significant gap in revenue between the top-performing region (Maharashtra) and the lowest-performing regions.

3.4 Product Performance Analysis

Output Result

	Category	Total_Amount	Percentage
	<chr>	<dbl>	<dbl>
1	Set	41165534.	49.4
2	kurta	23211347.	27.9
3	Western Dress	11697653.	14.0
4	Top	5624277.	6.75
5	Ethnic Dress	831148.	0.998
6	Blouse	485633.	0.583

1. Dominance of Product Categories

   The "Set" category contributes the largest share to total revenue, accounting for 49.4% of the total.

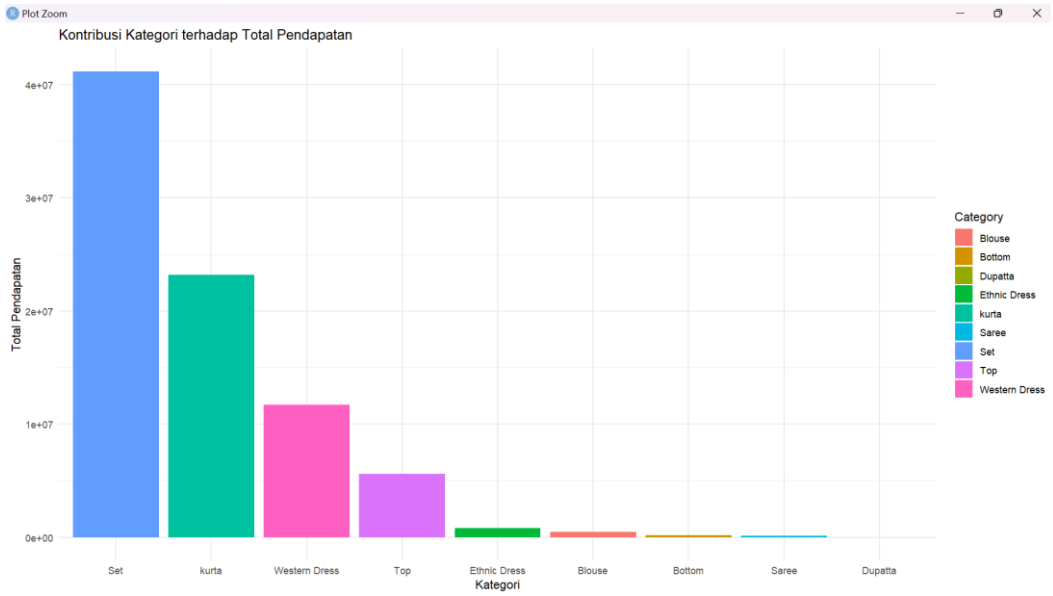
2. Other Popular Categories

   "Kurta" ranks second with a contribution of 27.9%, followed by "Western Dress" in third place with a contribution of 14.0%.

3. Categories with Low Contributions

   "Top" contributes only 6.75%, followed by "Ethnic Dress" (0.998%) and "Blouse" (0.583%).

Categorical Contribution



1. Dominance of Product Categories

   - The "Set" category has the highest revenue contribution, significantly surpassing other categories, with total revenue exceeding 40 million INR.
   - The "Kurta" category ranks second, contributing more than 20 million INR.
   - The "Western Dress" category ranks third, with revenue much smaller than "Set" and "Kurta" but still significant, indicating a demand for modern clothing.

2. Categories with Low Contributions

   "Ethnic Dress," "Blouse," and other categories such as "Saree" and "Dupatta" have very small contributions to total revenue.

INSIGHTS AND RECOMMENDATIONS

4.1 Key Insights

1. Sales by Category:

   The dominance of the "Set" category indicates that customers tend to prefer bundled products due to the added value they offer. Other popular categories, such as "Kurta" and "Western Dress," demonstrate high demand due to cultural relevance in India, the practicality of everyday wear, and the potential demand for modern clothing and global trends in the Amazon India market. Categories like "Blouse" and "Ethnic Dress" show low demand, which may be caused by a lack of product variety or insufficient promotion. There is a significant gap between the top categories ("Set" and "Kurta") and the lower ones like "Blouse" and "Ethnic Dress," highlighting a heavy reliance on a few key categories for total revenue.

2. Regions with the Highest Revenue:

   Maharashtra and Karnataka are key markets that need to be continuously optimized. Revenue contributions from these regions reflect a concentration of customers in metropolitan and advanced areas, likely due to a large customer base, good logistical infrastructure, and a strong preference for online shopping. Regions with lower contributions indicate low adoption of online shopping, limited logistical accessibility, and untapped market potential. There is a significant revenue gap between the top-performing region (Maharashtra) and the lowest-performing regions, indicating that Amazon's market in India is heavily focused on metropolitan and advanced states rather than rural or remote areas.

3. Customer Loyalty:

   Low customer loyalty reflects challenges in retaining customers after their first purchase. Customers may not return due to a lack of promotions, unsatisfactory customer experiences, or irrelevant products.

4. Customer Segmentation:
    - Cluster 2 dominates the customer segments and shows strong potential for further growth through loyalty programs or targeted promotions.
    - Clusters 1 and 3 hold significant potential to be converted into more active customers with the right approaches.

4.2 Recommendations

1. Sales by Category:
    - Enhance promotions for the "Set" category by offering special discounts or exclusive campaigns for bundled purchases.
    - Conduct further observations on categories with low contributions to identify the main causes, such as pricing, product exposure, or availability.
2. Regions with the Highest Revenue:
    - Maximize the potential of high-revenue regions with more aggressive marketing strategies, such as relevant regional campaigns.
    - Launch online shopping campaigns in low-contribution regions to boost adoption and sales.
3. Customer Loyalty:
    - Introduce a loyalty program, such as reward points for repeat purchases.
    - Personalize the customer experience with product recommendations based on purchase history.
4. Customer Segmentation:
    - Focus on Cluster 2 to improve customer retention through premium services, such as free shipping or exclusive promotions.
    - Encourage customers in Clusters 1 and 3 to become more active with incentives like discount vouchers or first-time product offers.

CLOSURE

5.1 Conclusion

1. Product Categories:
    - The "Set" category dominates revenue contributions (49.4%), indicating high customer interest in bundled products.
    - Categories like "Blouse" and "Ethnic Dress" show low contributions, which can be optimized through targeted promotions or product diversification.
2. Regional Performance:
    - States like Maharashtra and Karnataka contribute the most to revenue, highlighting the importance of focusing on well-established metropolitan regions.
    - Low-revenue regions like Mizoram and Sikkim present untapped market potential for expansion.
3. Customer Loyalty:
   Customer loyalty remains low, with more than 75% of customers making only a single purchase. This indicates the need for strategies to improve customer retention.
4. Customer Segmentation:
    - Customers with moderate to high spending (Cluster 2) represent the most promising segment to retain and grow.
    - Customers with low purchase activity (Cluster 1 and 3) require tailored strategies to increase engagement.

5.2 Long-term Strategy Implications

1. Optimization of Dominant Categories:
   Maintain and enhance high-performing categories, such as "Set" and "Kurta," through exclusive promotions, discount campaigns, and product diversification within these categories.
2. Market Expansion by Region:
   Focus on high-revenue regions to maintain market share, while expanding penetration in untapped regions through customer education, improved logistics, and localized promotions.
3. Enhancing Customer Loyalty:
   Invest in loyalty programs designed to encourage repeat purchases, such as reward points, exclusive discounts, and personalized product recommendations.
4. Developing Customer Segmentation Strategies:
    - Focus on moderate to high-spending customers to increase their purchase value through premium offerings.

- Improve engagement with low-spending customers through targeted promotions and education on the benefits of online shopping.

APPENDIX

The R syntax can be accessed through the GitHub link below under the file name "Business Intelligence by Me."

https://github.com/pucuk-yusup/Portofolio-Analisis.git