

Presented by Muhammad Fikri Riyanto

E-COMMERCE SALES ANALYSIS

<https://github.com/muhammadfikririyanto/Data-Analyst-Portofolio>



INTRODUCTION

This project aims to analyze and visualize sales data from a fictional e-commerce platform covering the period from January 2024 to April 2025. The goal is to uncover insights into sales patterns across various dimensions, such as time-based trends, top-performing product categories and items, customer preferences in payment methods, and cities with the highest transaction volumes. The dataset used is AI-generated and consists of 500 transaction entries with 8 key columns: Order_ID, Date, Product, Category, Price, Quantity, Customer_Location, and Payment_Method.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

MAIN CHALLENGES IDENTIFIED

01

MONTHLY SALES FLUCTUATIONS

Are there specific months showing a significant increase or drop in sales?

02

UNEVEN PRODUCT CATEGORY PERFORMANCE

Which product categories contribute the most to total sales, and which are underperforming?

03

SALES DISTRIBUTION BY LOCATION

Which regions generate the highest and lowest sales performance?



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

STEPS USED IN THIS ANALYSIS



- 01** Import Dataset
- 02** Data Understanding
- 03** Data Cleaning & Preparation
- 04** Exploratory Data Analysis (EDA)
- 05** Conclusion

RESULT OF IMPORT DATASET

IMPORT DATASET

	Order_ID	Date	Product	Category	Price	Quantity	Customer_Location	Payment_Method
0	1001	2024-01-01	Tablet	Accessories	4380993	4	Medan	Bank Transfer
1	1002	2024-01-02	Smartwatch	Accessories	13620110	2	Yogyakarta	E-Wallet
2	1003	2024-01-03	Headphones	Accessories	14467111	2	Jakarta	E-Wallet
3	1004	2024-01-04	Smartwatch	Electronics	3068103	1	Yogyakarta	E-Wallet
4	1005	2024-01-05	Smartwatch	Accessories	8969872	1	Bandung	E-Wallet

The loaded dataset contains 8 columns, which include: Order_ID, Date, Product, Category, Price, Quantity, Customer_Location, Payment_Method, Total_Sales, Month_Year, and Day_of_Week. In the next stage, I will apply techniques such as derivation, deduplication, imputation (if applicable), and transformation.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

DATA UNDERSTANDING

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Order_ID         500 non-null    int64  
 1   Date             500 non-null    object  
 2   Product          500 non-null    object  
 3   Category         500 non-null    object  
 4   Price            500 non-null    int64  
 5   Quantity         500 non-null    int64  
 6   Customer_Location 500 non-null    object  
 7   Payment_Method   500 non-null    object  
dtypes: int64(3), object(5)
memory usage: 31.4+ KB
```

The table provides a descriptive statistical summary of three numeric columns: Order_ID, Price, and Quantity. The Order_ID values range from 1001 to 1500, with a mean of 1250.5, indicating a sequential and evenly distributed order ID system. For the Price column, values vary widely, starting from approximately 564,000 to almost 20 million. The average price is around 10.4 million, but the high standard deviation suggests significant variability in product pricing across the dataset. As for Quantity, the number of items per order ranges from 1 to 4, with a mean of 2.49. The data shows that most orders consist of 1 to 4 items, with the median at 2, indicating a fairly balanced distribution.

This image shows a summary of a DataFrame containing 500 rows and 8 columns. All columns are fully populated, meaning there are no missing values.

The columns include details like order ID, date, product, category, price, quantity, customer location, and payment method. Three of the columns (Order_ID, Price, and Quantity) contain numerical data, while the rest hold text-based information like product names or payment types.

	Order_ID	Price	Quantity
count	500.000000	5.000000e+02	500.000000
mean	1250.500000	1.044409e+07	2.490000
std	144.481833	5.383152e+06	1.126249
min	1001.000000	5.643230e+05	1.000000
25%	1125.750000	6.065718e+06	1.000000
50%	1250.500000	1.043067e+07	2.000000
75%	1375.250000	1.464505e+07	4.000000
max	1500.000000	1.999648e+07	4.000000



DATA CLEANING & PREPARATION

CONVERTING THE DATE COLUMN TO DATETIME FORMAT

Order_ID	int64
Date	datetime64[ns]
Product	object
Category	object
Price	int64
Quantity	int64
Customer_Location	object
Payment_Method	object
dtype: object	

FILTER DATE (JANUARI 2024-APRIL 2025)

Order_ID	Date	Product	Category	Price	Quantity	Customer_Location	Payment_Method
0	2024-01-01	Tablet	Accessories	4380993	4	Medan	Bank Transfer
1	2024-01-02	Smartwatch	Accessories	13620110	2	Yogyakarta	E-Wallet
2	2024-01-03	Headphones	Accessories	14467111	2	Jakarta	E-Wallet
3	2024-01-04	Smartwatch	Electronics	3068103	1	Yogyakarta	E-Wallet
4	2024-01-05	Smartwatch	Accessories	8969872	1	Bandung	E-Wallet
...
481	2025-04-26	Tablet	Accessories	8695810	3	Yogyakarta	Credit Card
482	2025-04-27	Smartphone	Accessories	7685197	4	Surabaya	Debit Card
483	2025-04-28	Smartwatch	Accessories	7214301	4	Yogyakarta	E-Wallet
484	2025-04-29	Laptop	Accessories	11692304	4	Surabaya	Bank Transfer
485	2025-04-30	Headphones	Electronics	15504124	2	Medan	Credit Card

CHECKING & REMOVING DUPLICATE DATA DATETIME FORMAT

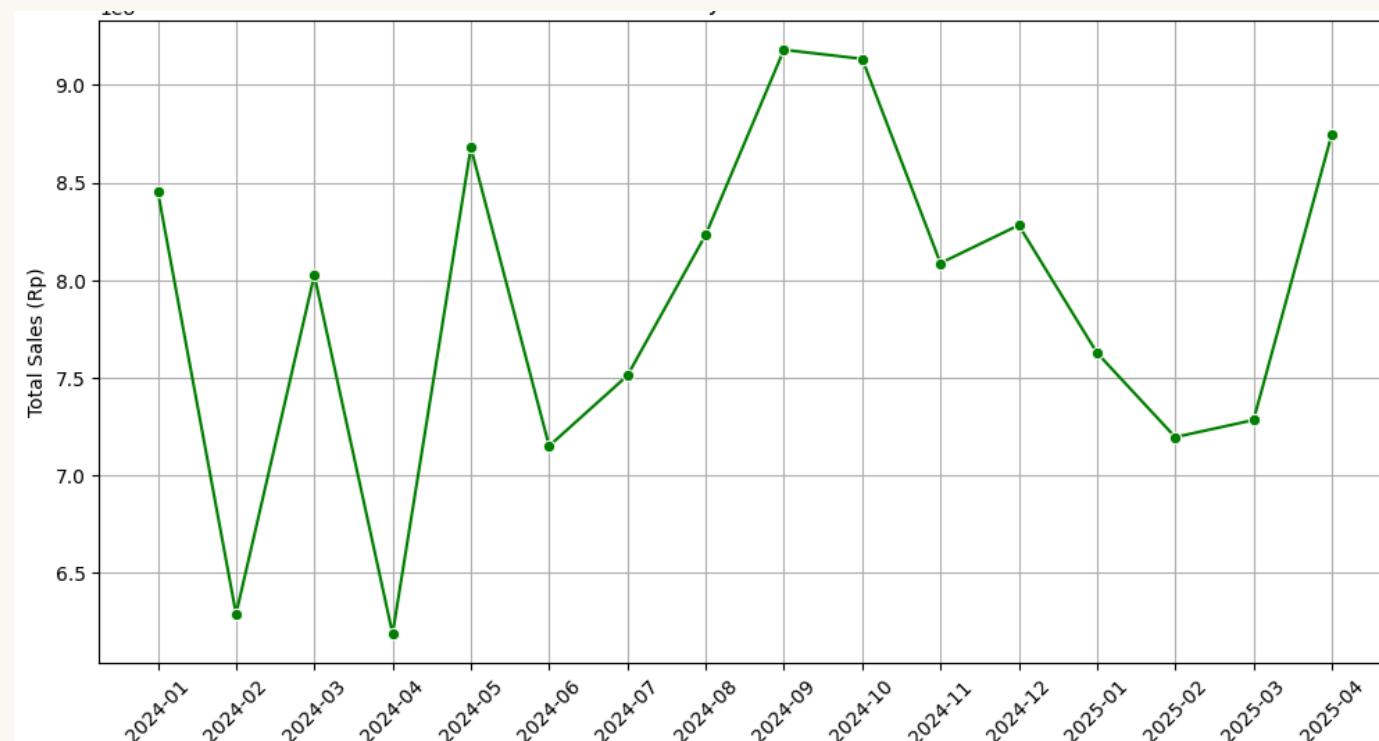
	Price	Quantity	Total_Sales
0	4380993	4	17523972
1	13620110	2	27240220
2	14467111	2	28934222
3	3068103	1	3068103
4	8969872	1	8969872
...
481	8695810	3	26087430
482	7685197	4	30740788
483	7214301	4	28857204
484	11692304	4	46769216
485	15504124	2	31008248
486 rows x 3 columns			

As part of the data cleaning process, the remove duplicated step is performed to eliminate rows with identical values across all columns—namely Price, Quantity, and Total_Sales. This ensures that duplicate entries do not skew analytical results, such as total sales calculations or aggregated metrics. Removing duplicates helps maintain data integrity and reliability prior to further analysis.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)



SALES TREND

From the line chart above, the period from January 2024 to June 2024 shows fluctuations or instability in sales. There is an increase from June 2024 to September 2024, followed by a decline until March 2025. However, in April 2025, there is an increase in sales. Since the data for May 2025 is incomplete, it is not displayed.

	Month_Year	Total_Sales
0	2024-01	845,500,887
1	2024-02	628,861,143
2	2024-03	802,565,969
3	2024-04	618,733,138
4	2024-05	867,949,543
5	2024-06	715,010,176
6	2024-07	751,078,487
7	2024-08	823,219,217
8	2024-09	918,100,508
9	2024-10	913,458,596
10	2024-11	808,733,681
11	2024-12	828,172,204
12	2025-01	762,747,265
13	2025-02	719,536,831
14	2025-03	728,303,869
15	2025-04	874,820,484

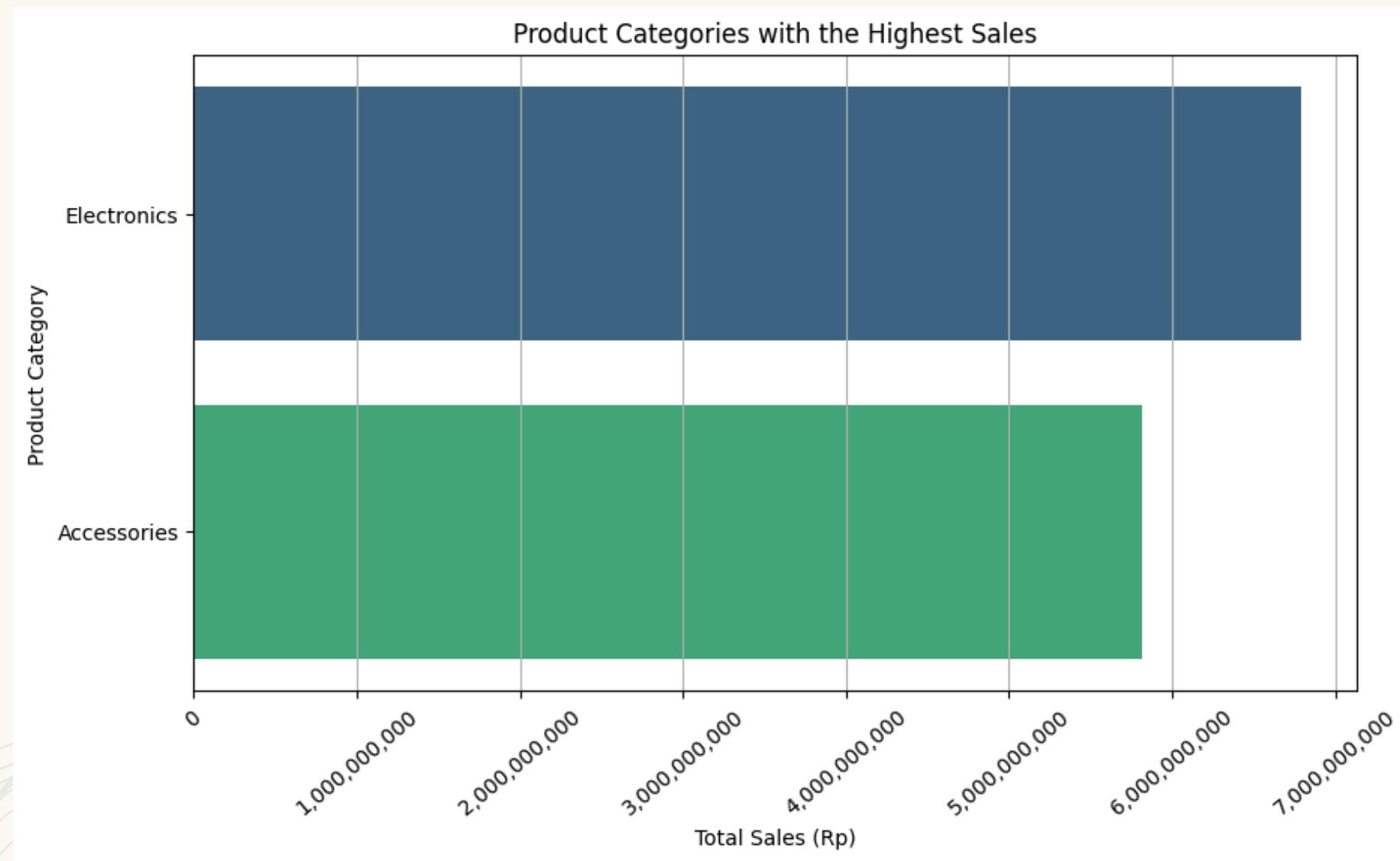
SALES FROM JANUARY 2024 TO APRIL 2025

Based on the monthly sales data from January 2024 to April 2025, the total accumulated sales amounted to approximately 12.61 billion. A time series analysis reveals notable month-to-month fluctuations, with the lowest sales recorded in April 2024 (618.7 million) and the peak in September 2024 (918.1 million). Despite a decline during Q2 2024, the overall trend indicates a recovery and stable growth in the second half of the year. Early 2025 shows a slight dip compared to late 2024 but remains relatively consistent, with monthly sales staying above 700 million. These patterns suggest the presence of seasonal effects and potential external factors influencing performance, providing valuable insights for future sales forecasting and strategic planning.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)

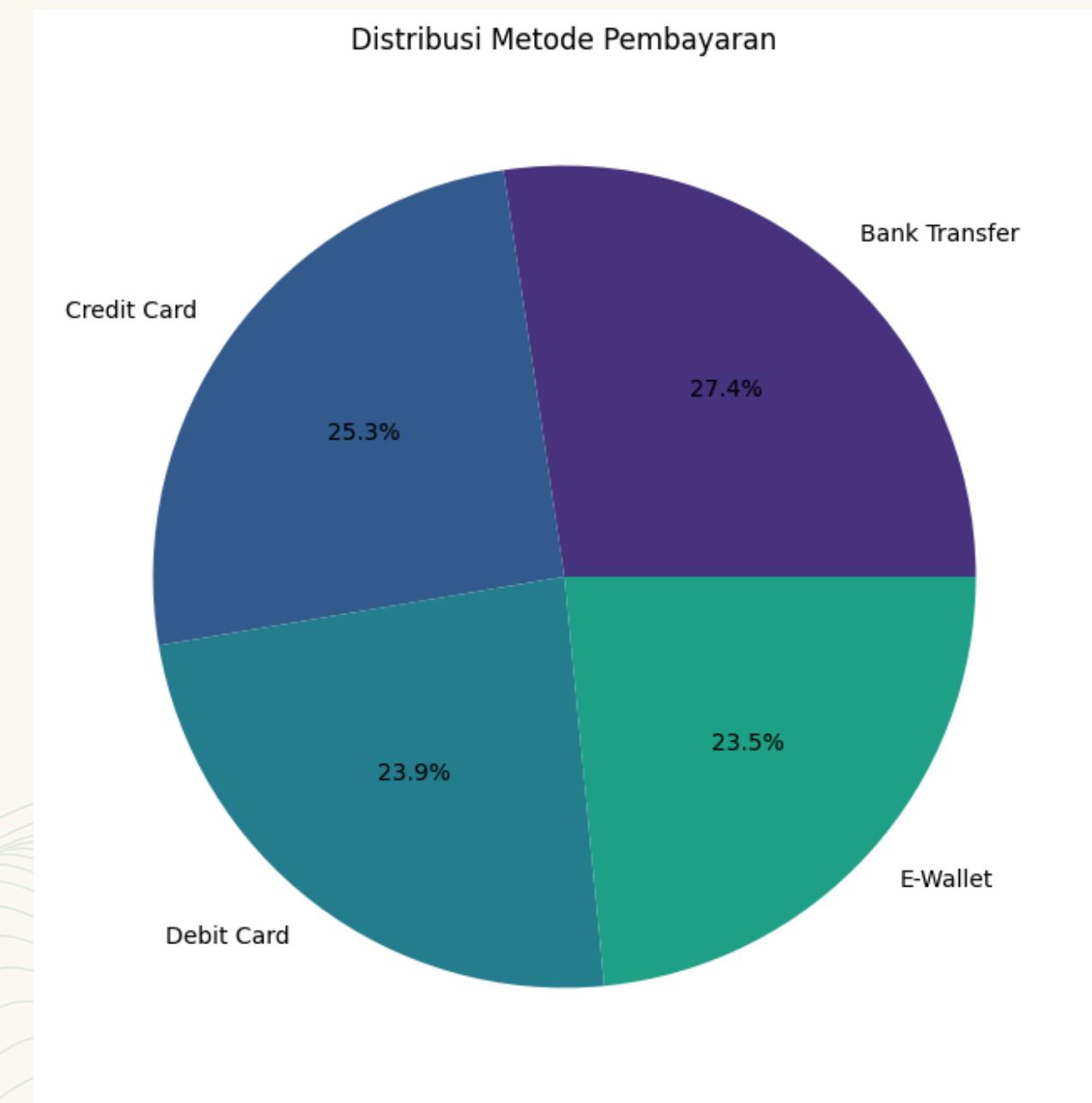


ANALYSIS OF THE BEST-SELLING PRODUCT CATEGORIES

Based on the available sales data, products in the Electronics category recorded the highest total sales, reaching approximately Rp 7 billion. In comparison, the Accessories category ranked second with total sales of nearly Rp 6 billion. The sales difference between the two categories is around Rp 1 billion, indicating that electronic products have higher demand or transaction value compared to accessories. This insight can be a valuable consideration for shaping future business strategies, particularly in terms of inventory management, promotions, and product development.



EXPLORATORY DATA ANALYSIS (EDA)



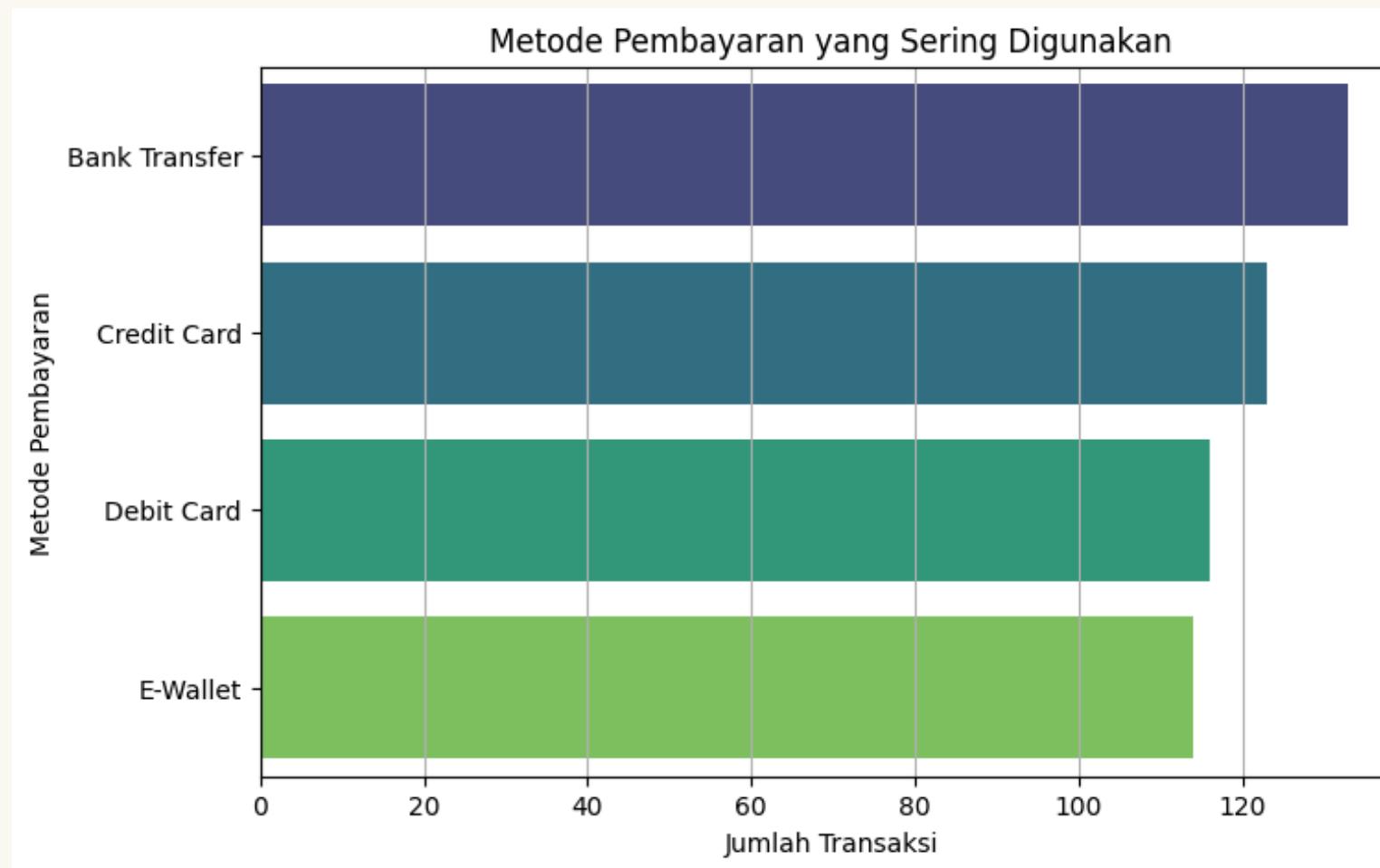
ANALYSIS OF THE MOST POPULAR PAYMENT METHODS

Based on the data presented, the most commonly used payment method among customers is Bank Transfer, accounting for 27.4% of total transactions. Coming in second is Credit Card with 25.3%, followed by Debit Card at 23.9%, and E-Wallet at 23.5%. Although the differences are relatively small, the dominance of Bank Transfer suggests it remains the preferred option for many customers.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)



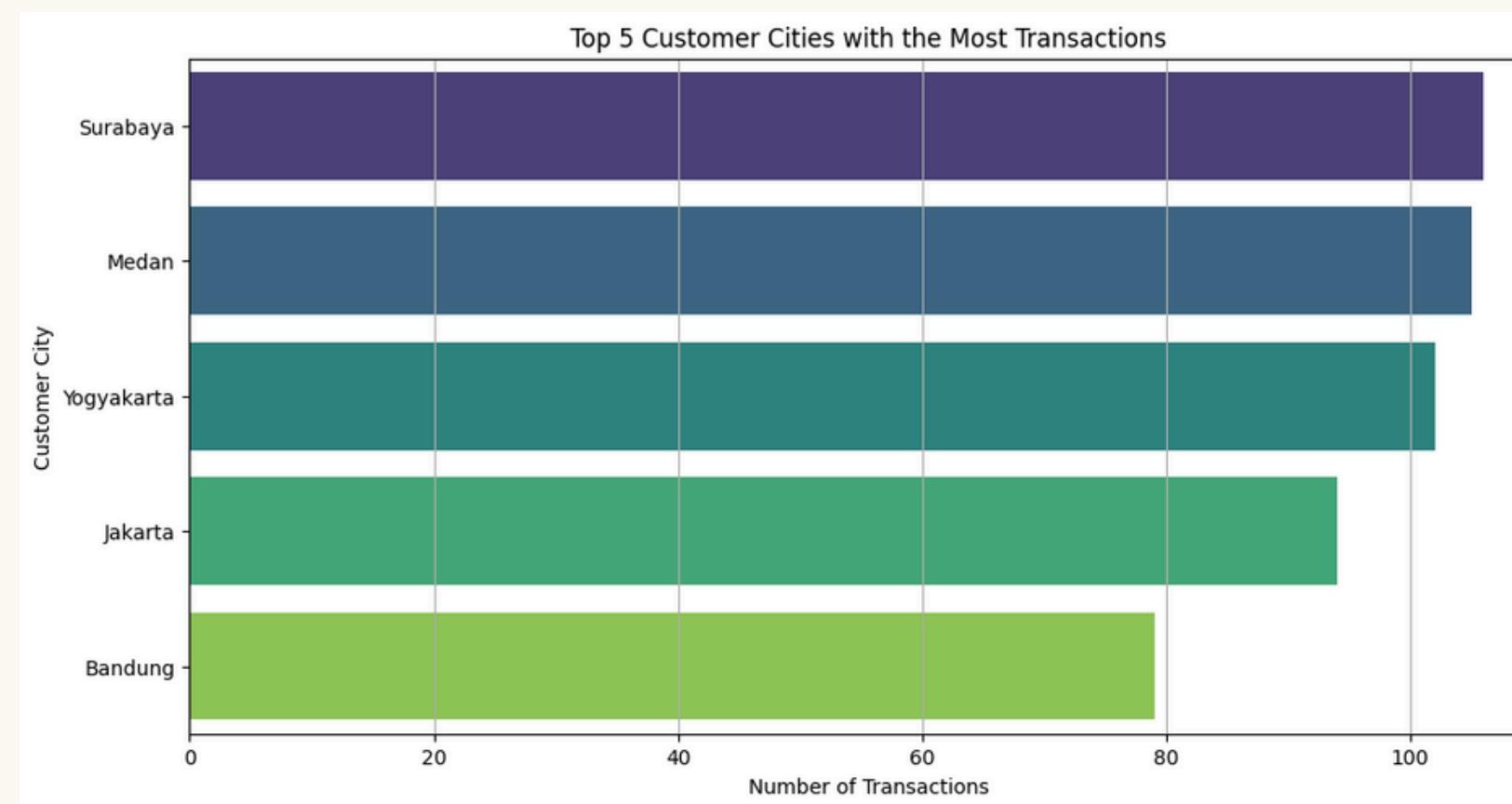
ANALYSIS OF POPULAR PAYMENT METHODS

Similar to the previous pie chart, this bar chart shows the total usage of the most frequently used payment methods, with bank transfer being the highest and e-wallet the lowest.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)

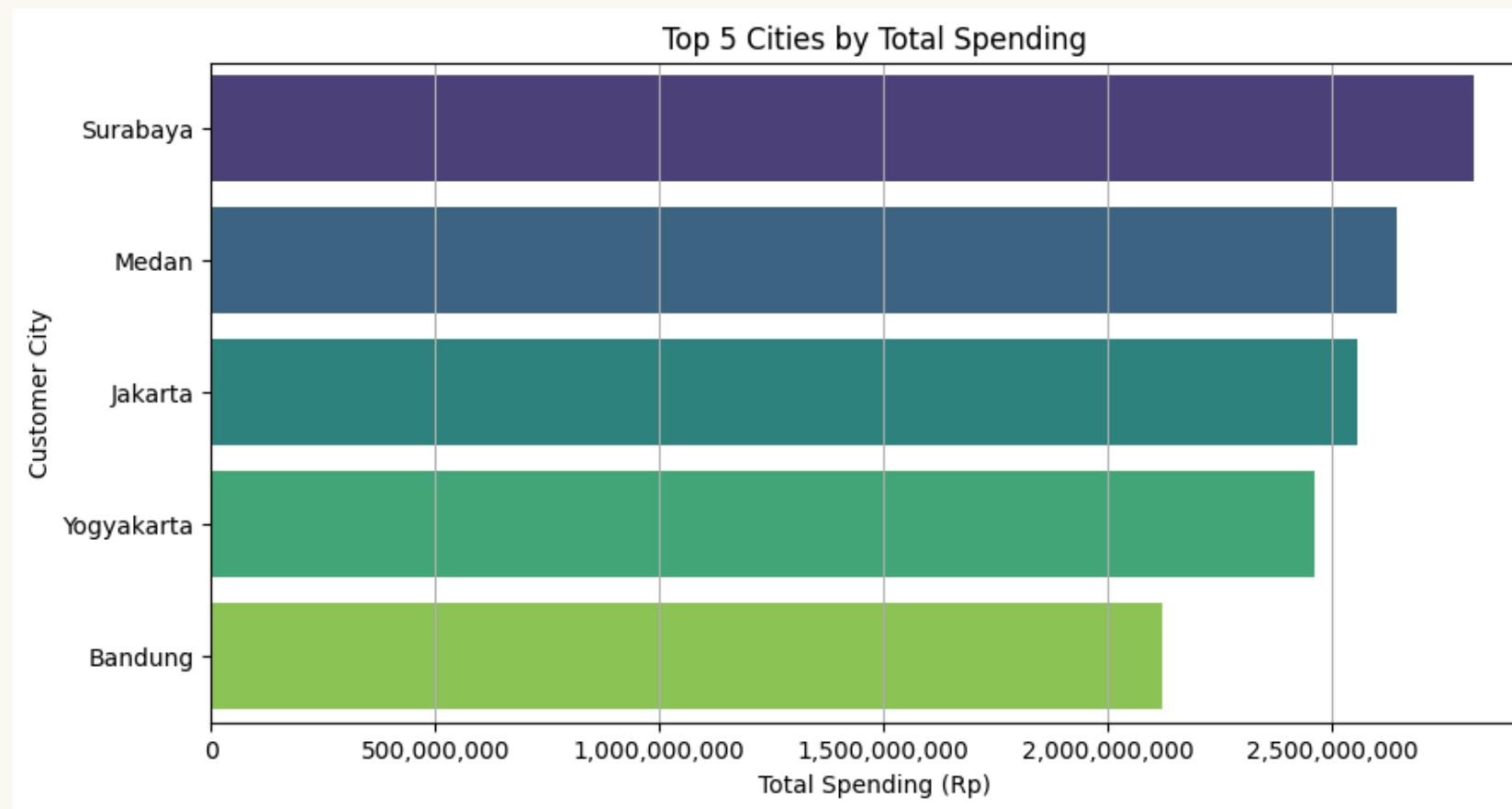


ANALYSIS OF CUSTOMER LOCATIONS WITH THE MOST TRANSACTIONS

Based on the chart, the city with the highest number of customer transactions is Surabaya, followed by Medan, Yogyakarta, Jakarta, and Bandung. Surabaya leads in transaction volume, indicating it is a highly active market for e-commerce sales. Medan and Yogyakarta also show strong transaction levels, slightly below Surabaya. Meanwhile, Jakarta and Bandung rank fourth and fifth, respectively.



EXPLORATORY DATA ANALYSIS (EDA)



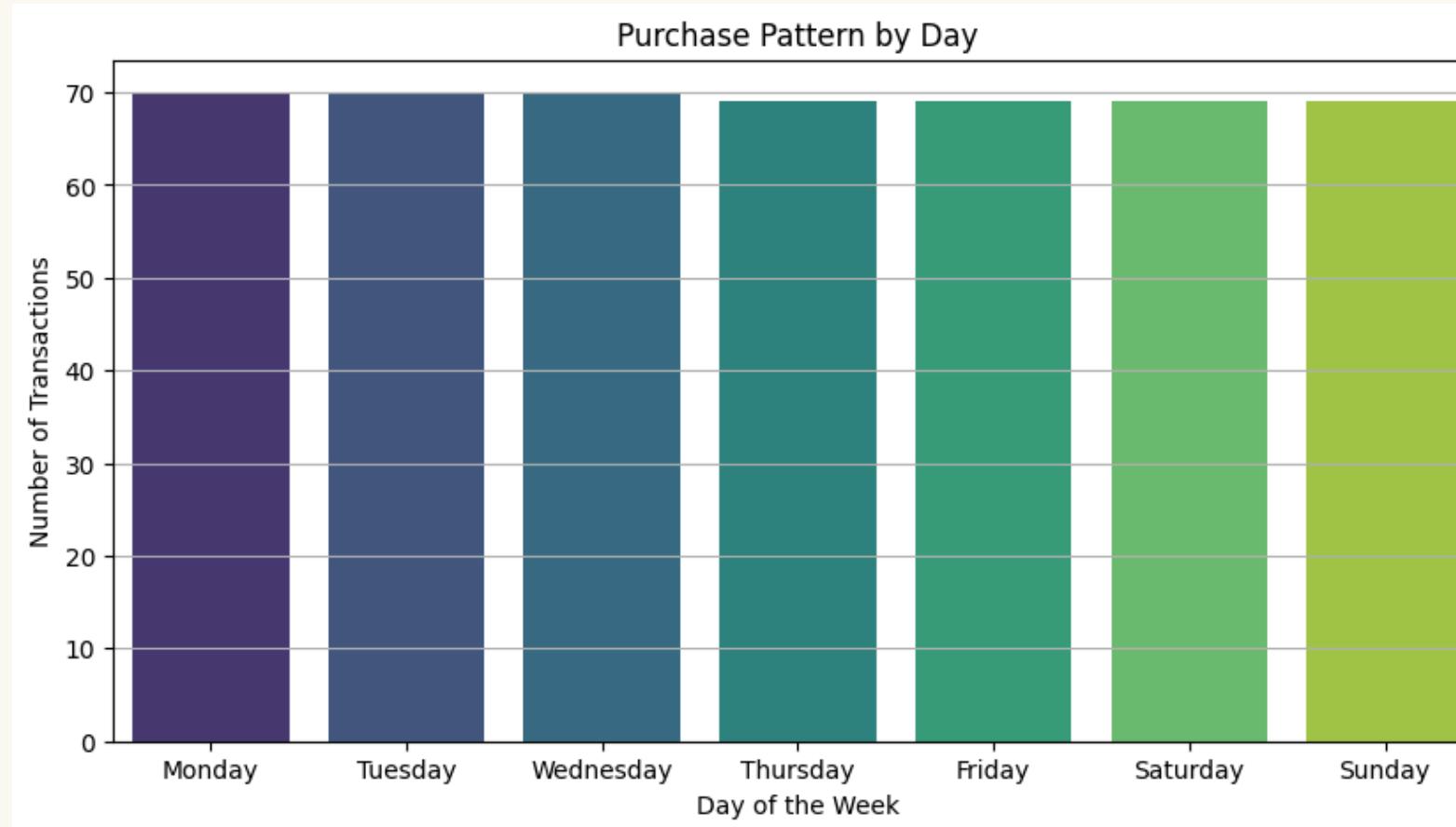
CUSTOMER SEGMENTATION BASED ON TOTAL PURCHASES

Based on cities, Surabaya recorded the highest sales with a total of over 2.8 billion Rupiah, while Bandung had the lowest with total sales of 2.1 billion Rupiah.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)



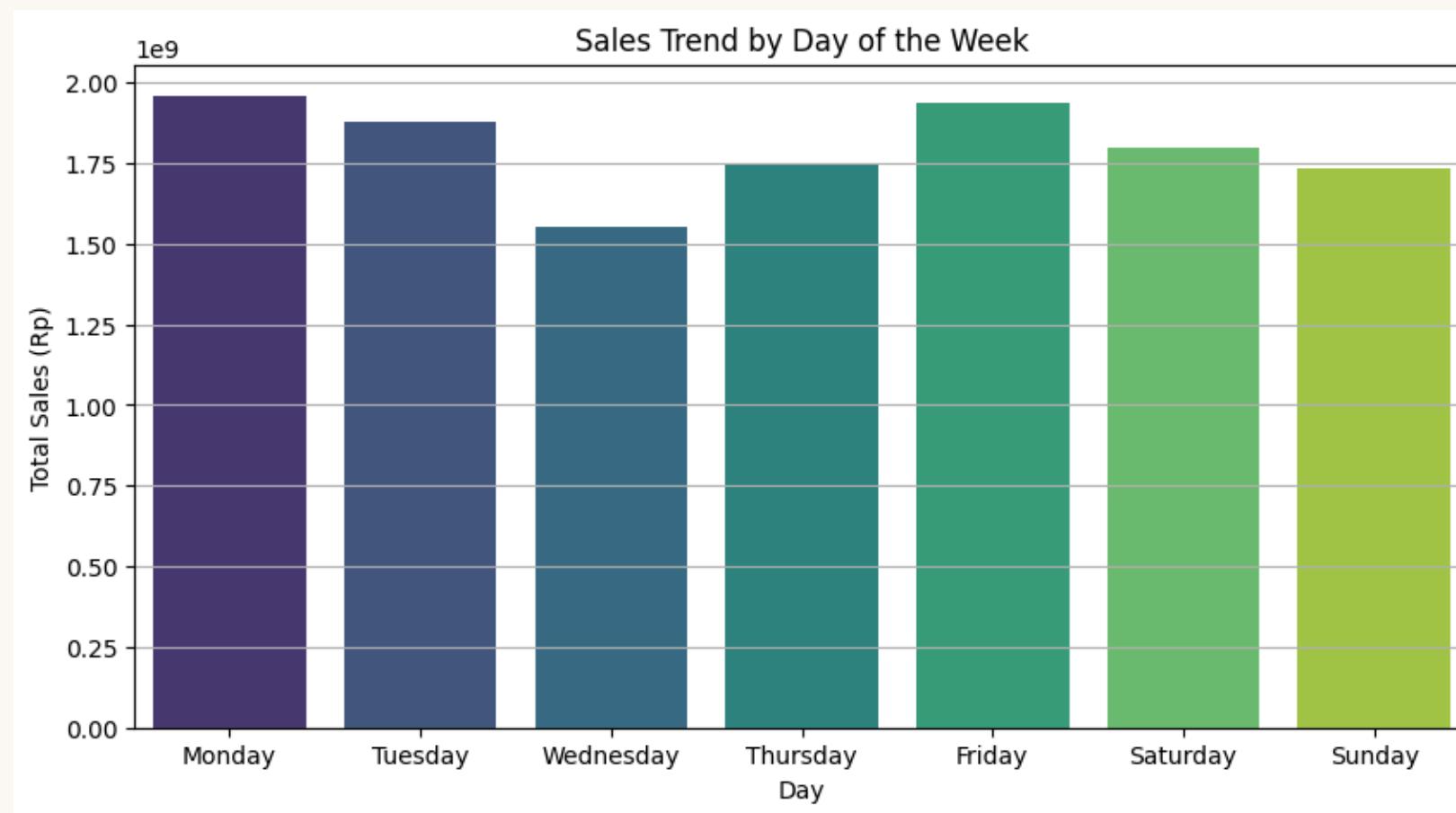
PURCHASE PATTERN BASED ON DAY OF THE WEEK

The bar chart illustrates the purchasing pattern throughout the week. Transaction volumes are slightly higher on Monday, Tuesday, and Wednesday compared to the other days. Nevertheless, the overall number of transactions remains relatively consistent across all days of the week.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)



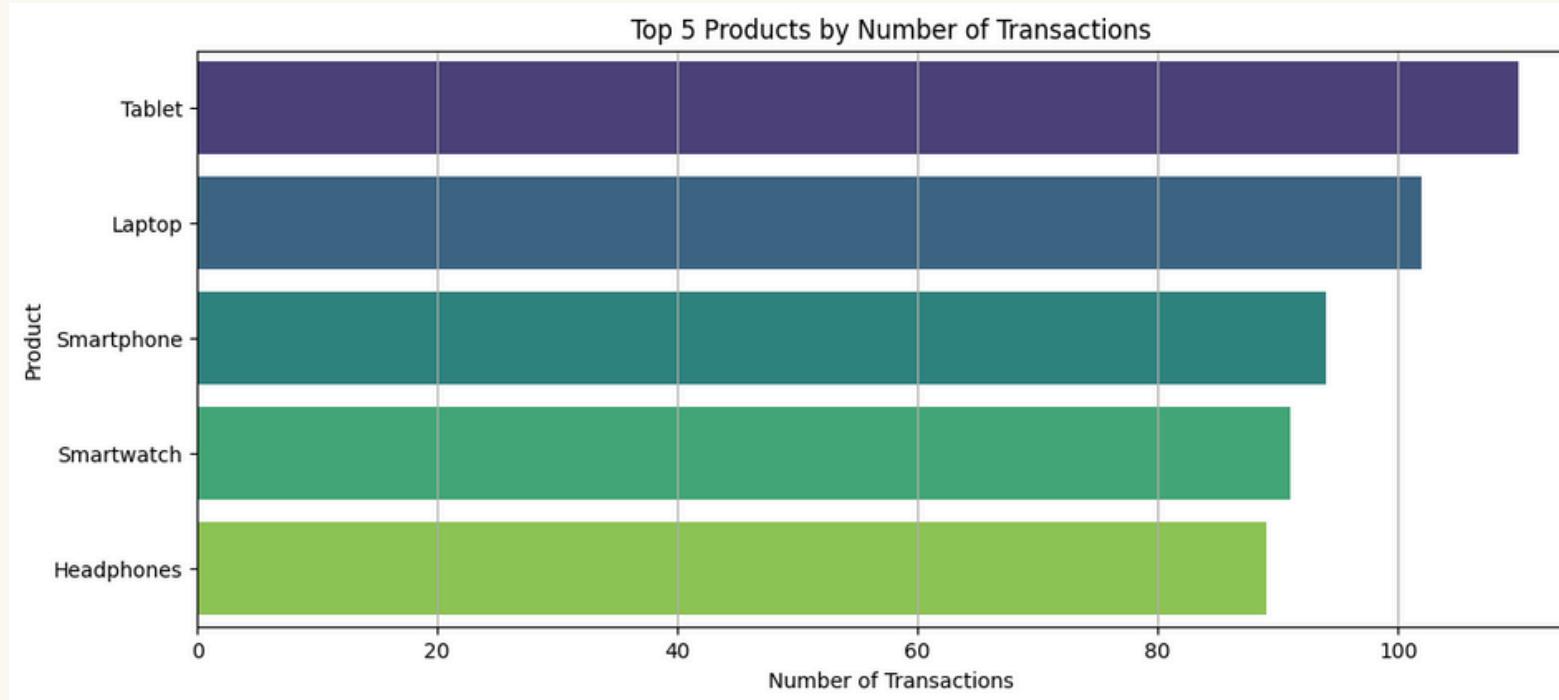
ANALYSIS OF SALES TRENDS OVER TIME

Total sales by day of the week. Monday recorded the highest sales with a total of nearly 2 billion Rupiah, while Wednesday had the lowest sales with a total of just over 1.5 billion Rupiah



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

EXPLORATORY DATA ANALYSIS (EDA)



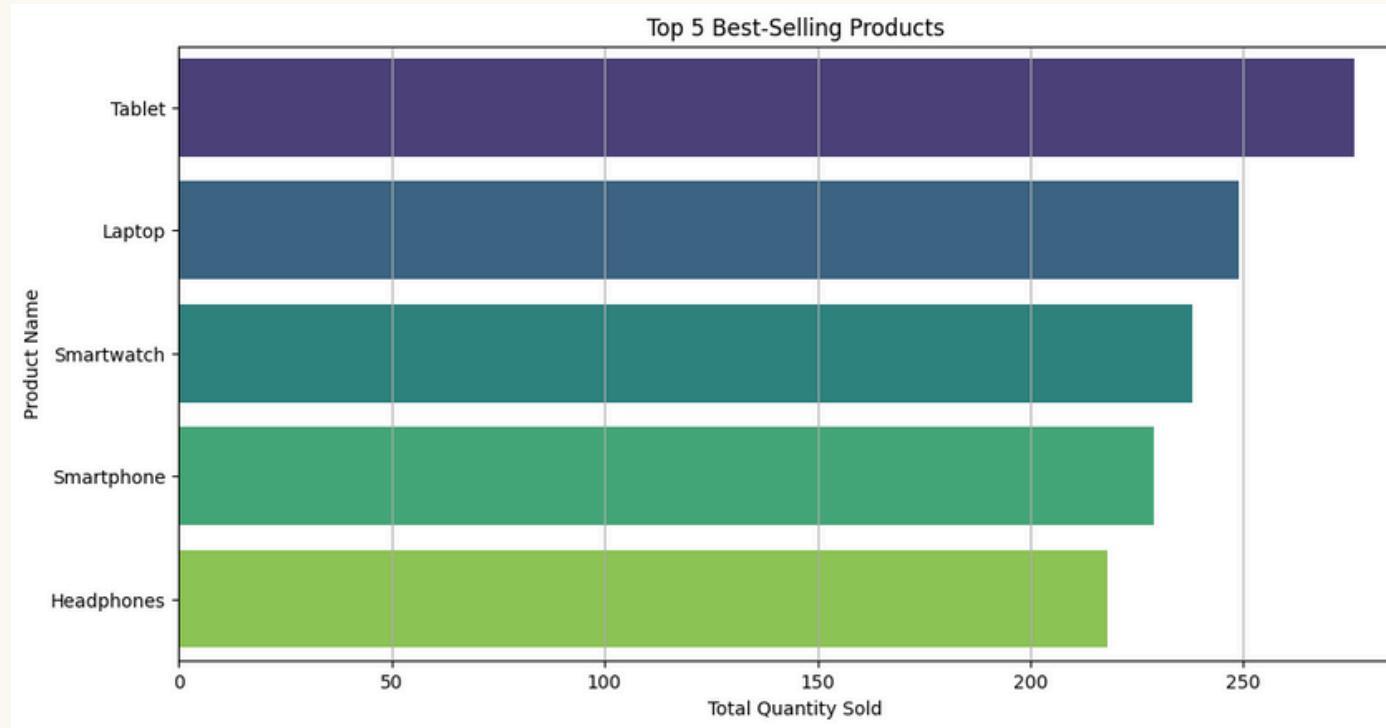
ANALYSIS OF TOP AND BOTTOM SELLING PRODUCTS

The bar chart illustrates the top five products with the highest number of transactions. Among these, tablets lead with approximately 110 transactions, making them the most frequently purchased item. Laptops follow closely behind with around 102 transactions. Smartphones take the third spot with about 94 transactions, slightly ahead of smartwatches, which recorded roughly 92 transactions. Headphones complete the list with approximately 90 transactions. Overall, the data suggests that tablets and laptops are the most popular products, indicating strong customer interest and demand in these categories.



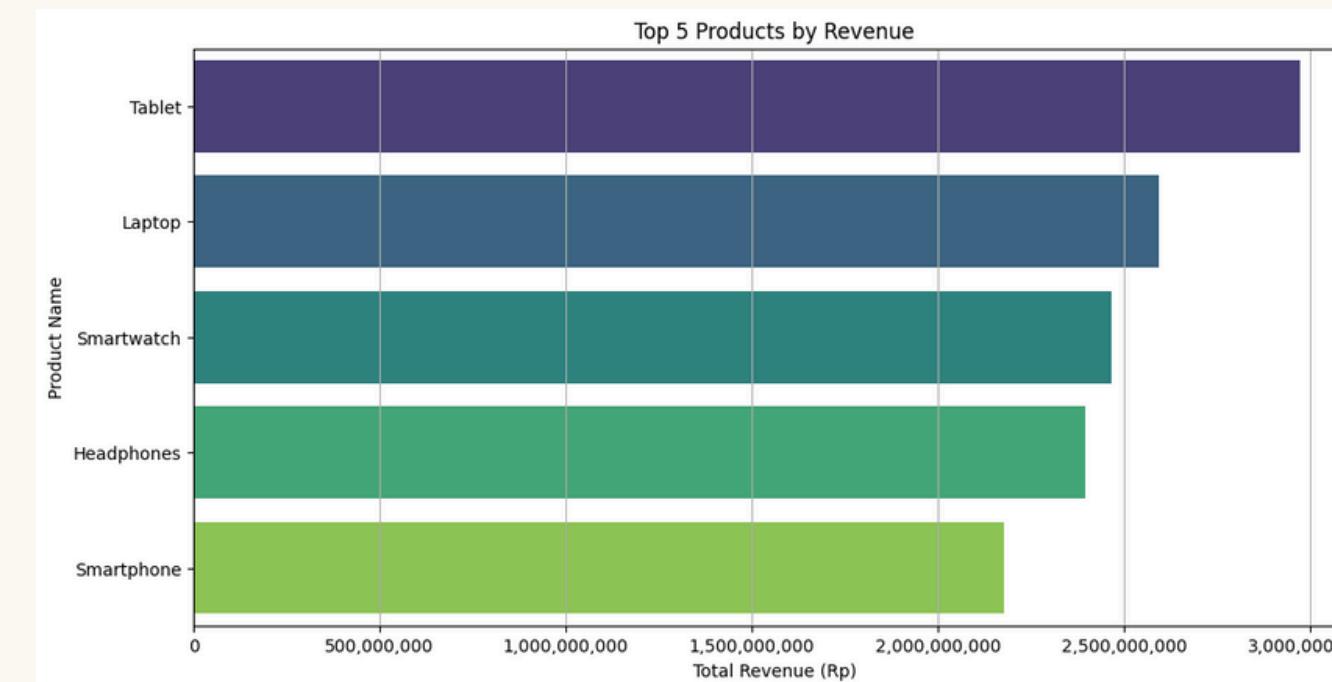
EXPLORATORY DATA ANALYSIS (EDA)

PRODUCT PERFORMANCE ANALYSIS



The bar chart illustrates the top five products with the highest number of transactions. Among these, tablets lead with approximately 110 transactions, making them the most frequently purchased item. Laptops follow closely behind with around 102 transactions. Smartphones take the third spot with about 94 transactions, slightly ahead of smartwatches, which recorded roughly 92 transactions. Headphones complete the list with approximately 90 transactions. Overall, the data suggests that tablets and laptops are the most popular products, indicating strong customer interest and demand in these categories.

PRODUCT PERFORMANCE ANALYSIS

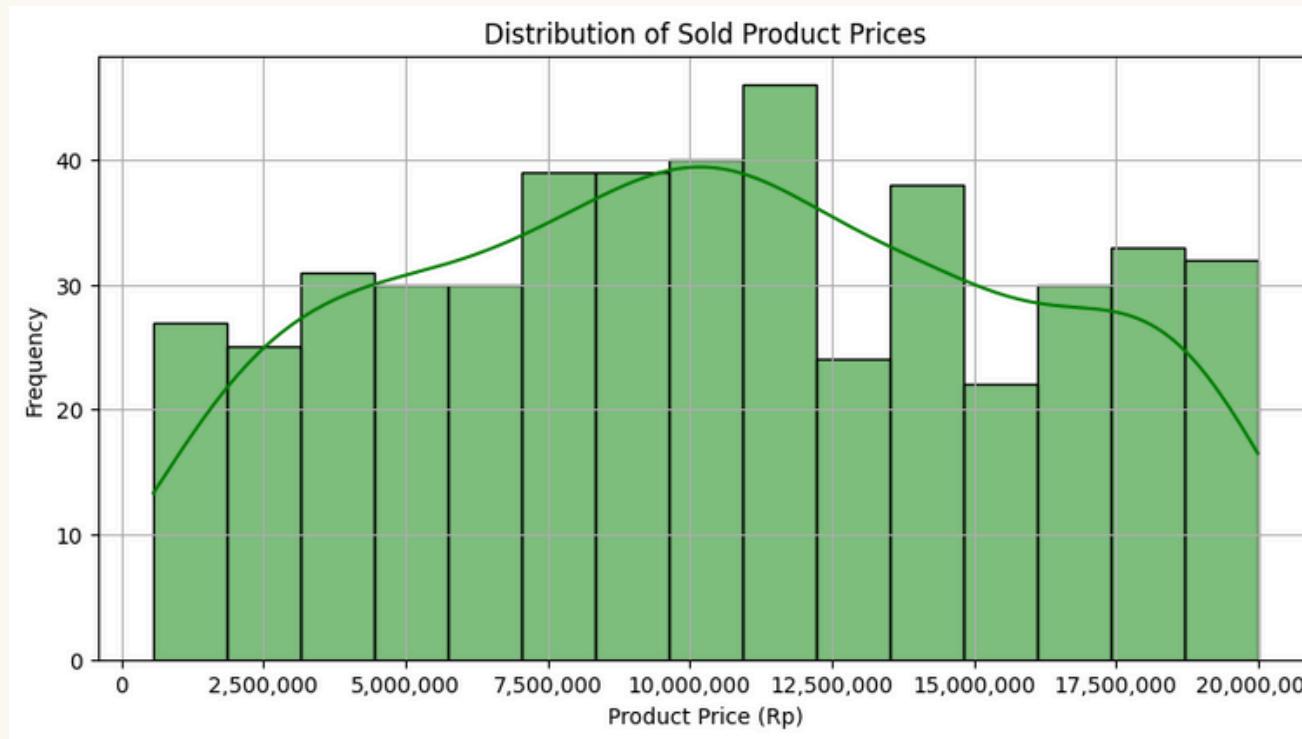


The chart shows that tablets are the leading product in terms of revenue, contributing almost Rp 3 billion. Laptops also generate significant income, exceeding Rp 2.6 billion. Smartwatches and headphones follow closely, with earnings near Rp 2.5 billion and Rp 2.4 billion, respectively. Despite being a commonly used device, smartphones recorded the lowest revenue among the top five, with around Rp 2.1 billion. This suggests that tablets and laptops not only sell well but may also have higher price points, contributing to greater total revenue.



EXPLORATORY DATA ANALYSIS (EDA)

ANALYZING THE DISTRIBUTION OF SOLD PRODUCT PRICES



The histogram illustrates the distribution of sold product prices. The most frequently sold products fall within the price range of Rp 10,000,000 to Rp 12,500,000, with over 45 sales in that category. Additionally, the ranges of Rp 7,500,000–10,000,000 and Rp 12,500,000–15,000,000 also show high sales frequencies, indicating strong consumer interest in mid-to-high priced products. Although sales also occur at lower (below Rp 2,500,000) and higher price points (above Rp 17,500,000), their frequencies are relatively lower. Overall, the price distribution is nearly normal but slightly right-skewed, suggesting a concentration of sales in the mid-price range.

CORRELATION ANALYSIS BETWEEN VARIABLES



The most influential factor on Total Sales is Price, followed by Quantity. Since both have positive correlations with total sales, increasing either can lead to higher sales. However, price and quantity appear to be independent of each other.



<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>

CONCLUSION

From the analysis above, it can be concluded that the total sales from January 2024 to April 2025 amounted to IDR 12,606,791,998 (Twelve billion, six hundred six million, seven hundred ninety-one thousand, nine hundred and eight Rupiah). The highest sales occurred in September 2024, exceeding 900 million Rupiah. The category with the highest sales was Electronics, with sales of 7 billion Rupiah. The most frequently used payment method was bank transfer, accounting for 27.4%. The city with the highest number of transactions was Surabaya, with more than 100 transactions and total sales exceeding 2.8 billion Rupiah. In a week, the most transactions occurred on Monday, Tuesday, and Wednesday, with 70 transactions. The most purchased product was a tablet, with over 100 units sold and total sales exceeding 3 billion Rupiah.



CONTACT ME !



E-MAIL

muhammadfikririyanto90@gmail.com



LINKEDIN

<https://www.linkedin.com/in/muhammad-fikri-riyanto-365737223/>