# Shopee Sales Data Analysis Report

## 1. Project Background

Shopee is one of the leading e-commerce platforms in Southeast Asia, with a large user base and a diverse range of products.  
The objective of this analysis is to study Shopee’s sales data, identify the characteristics of best-selling products, and explore key factors affecting sales to develop more effective sales strategies.

## 2. Data Overview and Preprocessing

\*\*Data Source\*\*: Shopee E-commerce Sales Dataset  
  
\*\*Example Data Fields\*\*:  
- `title`: Product name  
- `price\_actual`: Product price  
- `total\_sold`: Total sales  
- `rating`: Customer rating  
- `reviews`: Number of reviews  
- `location`: Seller location  
  
\*\*Data Cleaning Steps\*\*:  
- \*\*Handling Missing Values\*\*: Removed rows with excessive missing values and filled in certain empty fields.  
- \*\*Data Type Conversion\*\*: Converted `total\_sold` from `object` to `int` to ensure compatibility for analysis.  
- \*\*Outlier Removal\*\*: Filtered out abnormally low or high-priced products to ensure representative analysis results.

## 3. Data Analysis and Visualization

## (1) Best-Selling Product Analysis

Visualized the \*\*TOP 10 best-selling products\*\*.  
Found that best-selling products are mainly concentrated in \*\*electronics, home essentials, and beauty care\*\*.

## (2) Price vs. Sales Relationship

Analyzed the correlation between price and sales using a scatter plot.  
Found that \*\*products priced between $10-$50 tend to have higher sales\*\*, while extremely high or low-priced products have relatively lower sales.

## (3) Customer Ratings and Sales Impact

Calculated the average sales across different rating intervals (1-5 stars).  
Found that \*\*products with ratings above 4.5 tend to have significantly higher sales\*\*, indicating a positive relationship between customer ratings and sales performance.

## (4) K-Means Clustering Analysis

Selected `price\_actual` and `total\_sold` as feature variables.  
Applied \*\*K-Means clustering\*\* to categorize products into \*\*high-sales low-price, low-sales high-price, and balanced sales\*\* groups.  
Found that \*\*high-value-for-money products are more likely to become best-sellers\*\*.

## 4. Key Insights

✅ \*\*Best-Selling Categories\*\*: Electronics, home essentials, and beauty care are the most popular.  
✅ \*\*Price Impact on Sales\*\*: Products priced between \*\*$10-$50\*\* achieve the best sales, requiring strategic pricing adjustments.  
✅ \*\*Importance of Customer Ratings\*\*: Products with \*\*ratings above 4.5\*\* perform better, emphasizing the need for positive customer feedback.  
✅ \*\*Characteristics of High-Sales Products\*\*: Affordable pricing, high ratings, appropriate pricing, and active customer engagement.

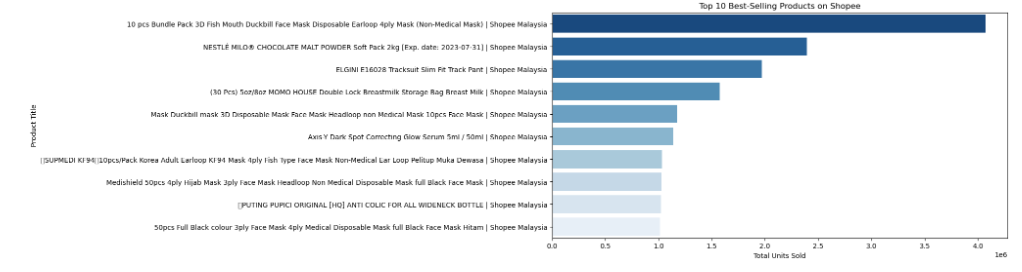
## 5. Conclusions and Recommendations

📌 \*\*Optimize Product Pricing\*\*: Set mainstream product prices between \*\*$10-$50\*\* to maximize sales.  
📌 \*\*Improve Customer Ratings\*\*: Encourage customer reviews and address negative feedback to boost ratings and increase sales.  
📌 \*\*Implement Targeted Marketing Strategies\*\*: Optimize promotional tactics for different product categories, such as offering discounts on electronics.  
📌 \*\*Enhance Inventory Management\*\*: Ensure sufficient stock for best-selling products (e.g., home essentials and beauty care) to avoid stockouts and revenue loss.

## Next Steps

1. \*\*Apply these analytical insights to Shopee’s platform\*\* to test the impact of pricing optimization and customer rating management.  
2. \*\*Further segment product categories\*\* and develop tailored marketing strategies for different types of products.  
3. \*\*Develop a predictive model\*\* using machine learning to forecast future best-selling products.  
  
🚀 Through this analysis, we provide actionable sales strategy recommendations for Shopee merchants, paving the way for deeper insights and optimization in the future!

Best-Selling Product Analysis:



Price vs. Sales Relationship:

