

E-commerce Furniture Sales Analysis & Prediction

Project Overview

This project analyzes 2,000 furniture product listings scraped from AliExpress to uncover key business drivers influencing sales performance. The study explores pricing, shipping strategies, and product characteristics, followed by predictive modeling to estimate units sold.

Dataset Summary

- 1 Total Records: 2,000
- 2 Key Features: productTitle, price, sold, tagText
- 3 Target Variable: sold (Units Sold)
- 4 Original price column removed due to 76% missing values

Data Cleaning Steps

- 1 Converted price from string to numeric
- 2 Handled missing values in shipping category
- 3 Grouped rare shipping categories into 'others'
- 4 Dropped originalPrice due to high missing data

Key Business Insights

- 1 Price distribution is right-skewed (majority low-to-mid priced products)
- 2 Sales distribution is right-skewed (few high-performing products dominate sales)
- 3 Lower-priced items generally sell more (price sensitivity observed)
- 4 Free shipping significantly increases both sales volume and total revenue
- 5 Total Revenue: \$2,181,048.07
- 6 Average Price: \$156.56
- 7 Average Units Sold: 23.49

Predictive Modeling Results

- 1 Baseline Model: Linear Regression ($R^2 \approx -0.01$)
- 2 Random Forest Model: $R^2 \approx 0.66$
- 3 Random Forest significantly outperformed Linear Regression
- 4 Sales behavior shows complex, non-linear relationships

Conclusion

The analysis reveals that pricing strategy and free shipping are major drivers of sales performance. Random Forest modeling successfully captured non-linear relationships, explaining approximately 66% of sales variance. These findings provide actionable insights for revenue optimization and pricing strategy improvement.