

Exploratory Data Analysis of an E-Commerce Dataset

Name: Venkata Mohana Krishna Gude | Student Number: 25039889

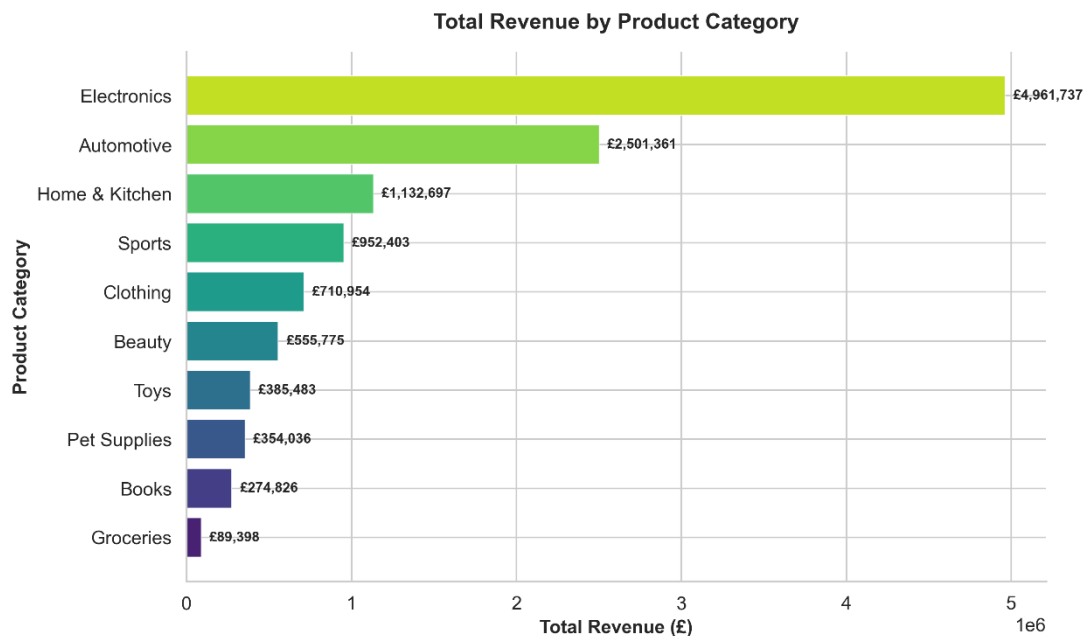
1. Introduction:

This exploratory analysis is based on an e-commerce data set containing product price, customer ratings, quantity sold, and revenue, and seeks to explore revenue by product category, understand the relationship between price and customer ratings, and explore the relationship of important operational metrics using statistical tests and data visualizations.

2. Revenue by Product Category:

Referring back to the bar chart below for total revenue, Electronics earnings at £4,961,737 are over twice that of Automotive at £2,501,361. The lowest grossing category of the 34 is Groceries at £89,398.

The distribution indicates that a few categories (e.g., Electronics) contribute considerably to sales while items with lower average prices that tend to be used daily have a lesser effect on overall revenue. The distribution is thus heavily skewed towards a few high-earning categories.



3. Product Price vs Customer Rating:

The scatter plot shows the price of product against the customer satisfaction rating. The Pearson correlation coefficient $r = -0.005$ between the price and the customer satisfaction indicates no linear correlation.

This indicates that higher prices may not necessarily guarantee higher ratings or indicate customer satisfaction, as other factors such as quality, service, or brand reputation could play a more meaningful role.



4. Correlation Analysis:

The heatmap below shows the Pearson correlations among the most important variables. Among the predictors, Product Price and Unit Price have a perfect correlation ($r = 1.000$) since they represent the same indicator of price.

There was a strong positive correlation between Product Price and Item Total ($r = 0.871$), wherein increases in the prices paid for each product were expected to considerably increase Item Total. There was also a moderate correlation between Quantity and Item Total ($r = 0.268$).

Price variables did not correlate to Product Rating, confirming that customer satisfaction is essentially price independent.

