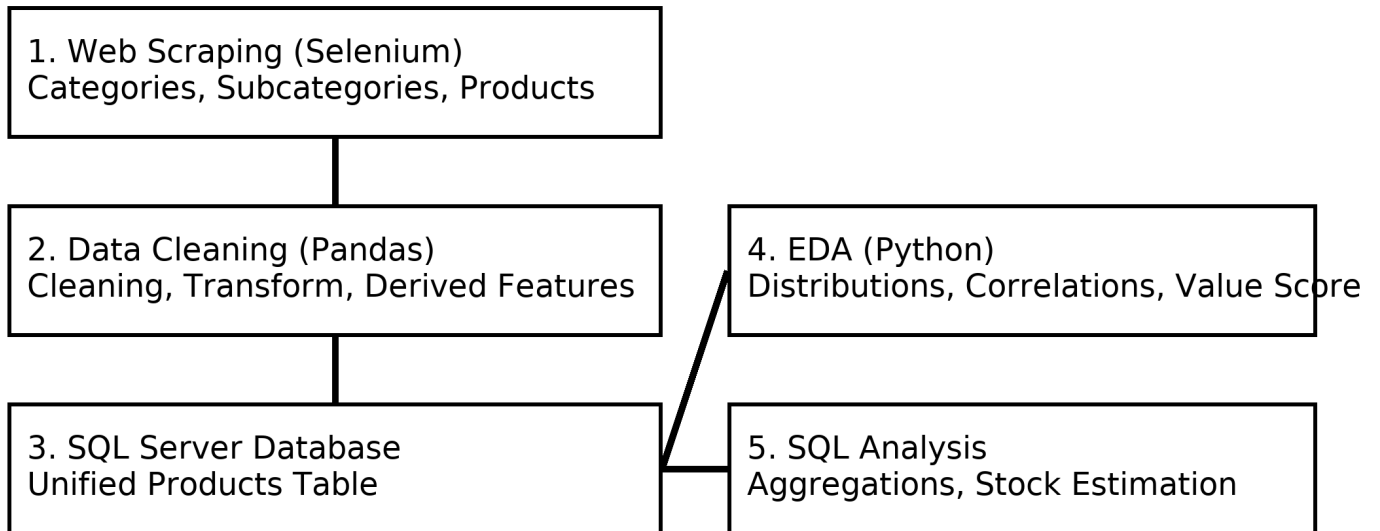


Banggood Product Trend Analysis – Final Report

Architecture Diagram



Executive Summary

This report documents the complete ETL pipeline applied to Banggood product data across three categories. Data scraping, cleaning, SQL loading, exploratory analysis, and final insights were developed to identify pricing behavior, customer engagement, and product value across categories.

Scraping Methodology

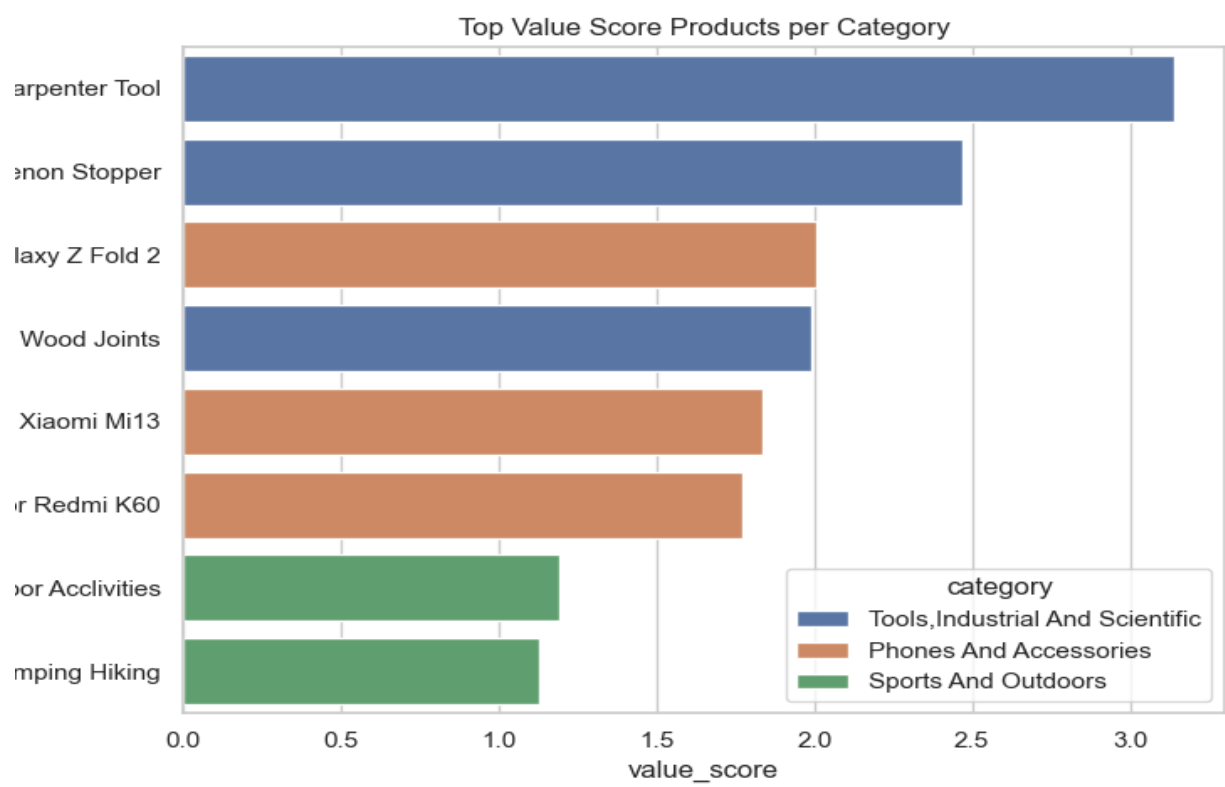
Selenium automation was used to handle Banggood’s dynamic content. Each category page was scanned for subcategories, followed by infinite scrolling to ensure all products were captured.

Data Cleaning & Transformation

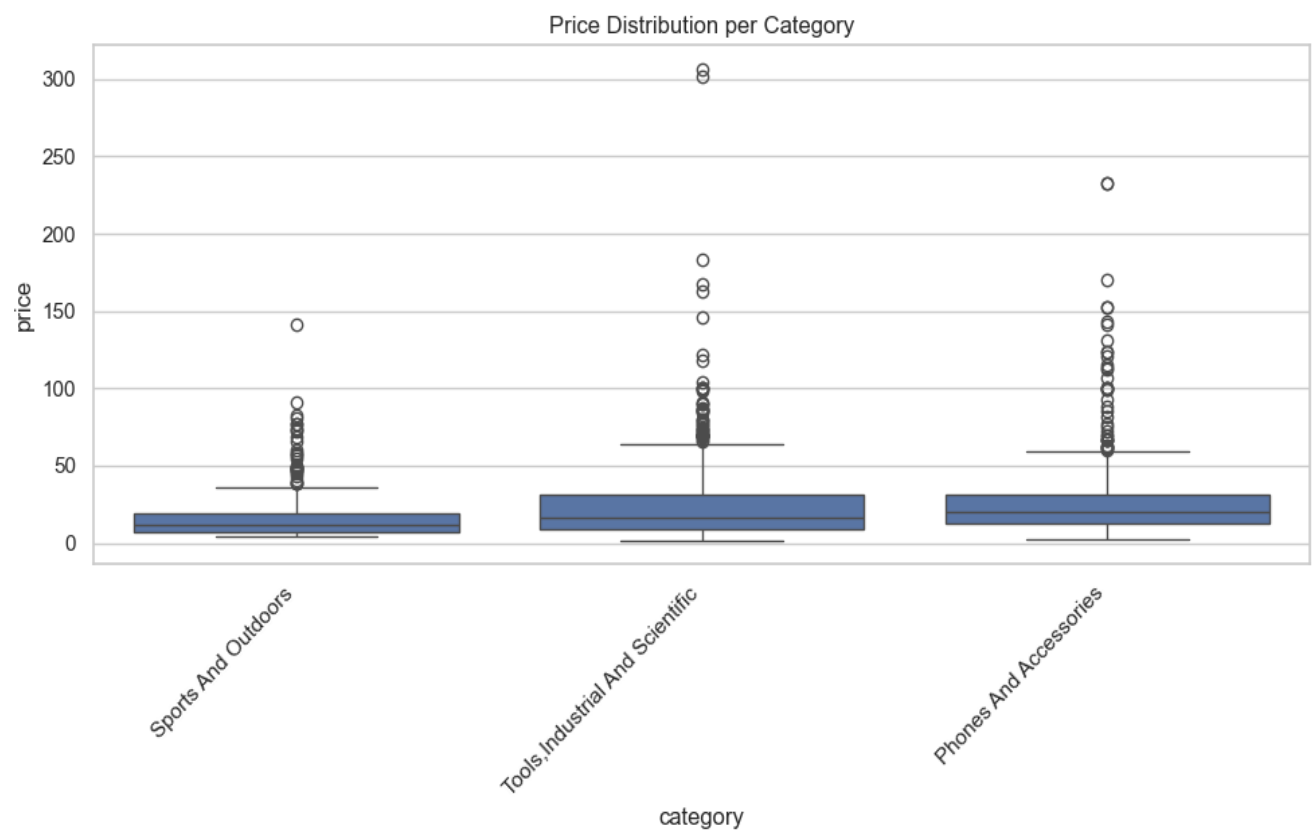
Price, rating, and review data were normalized to numeric formats. Missing values were handled and derived features such as value_score, popularity_score, and price_bucket were generated. Category and subcategory names were extracted from URLs.

Exploratory Data Analysis (with Graphs)

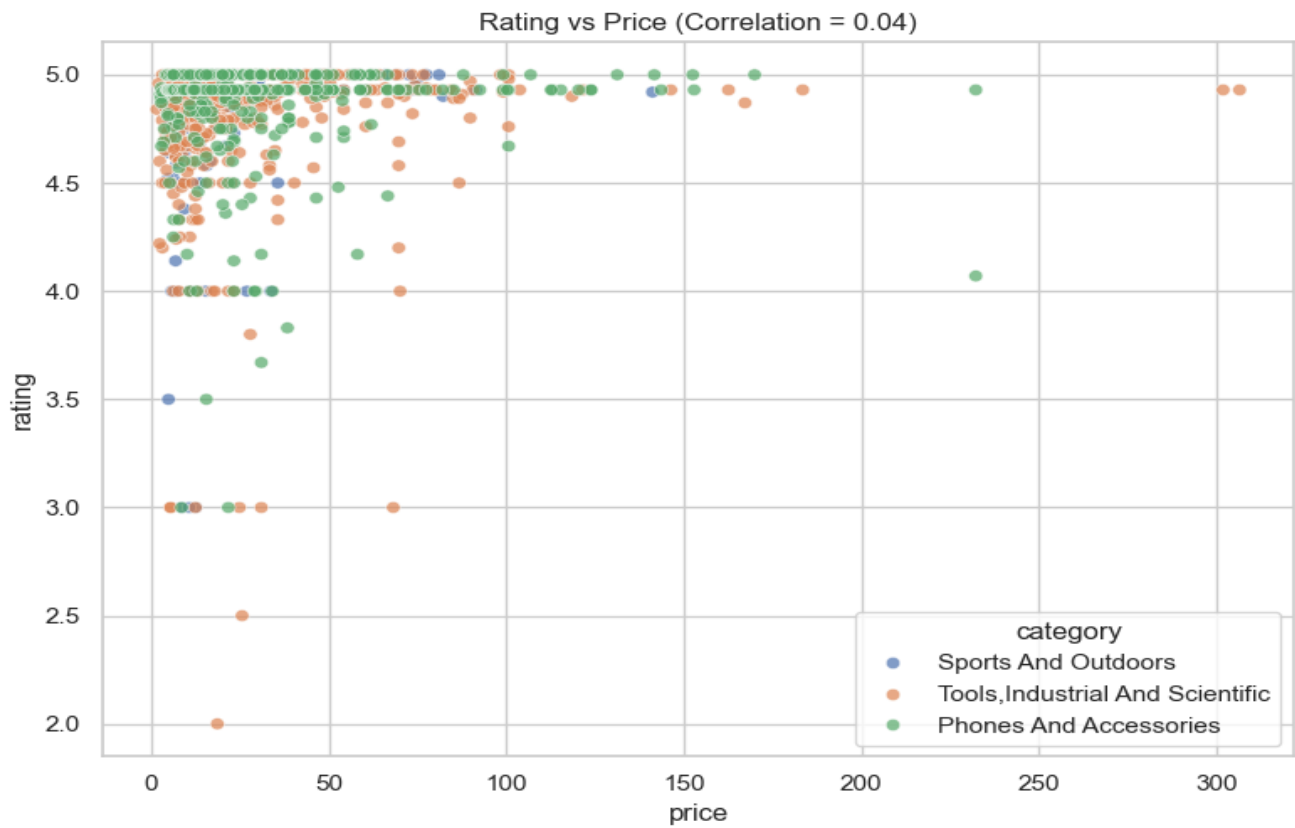
Top Value Score Products per Category



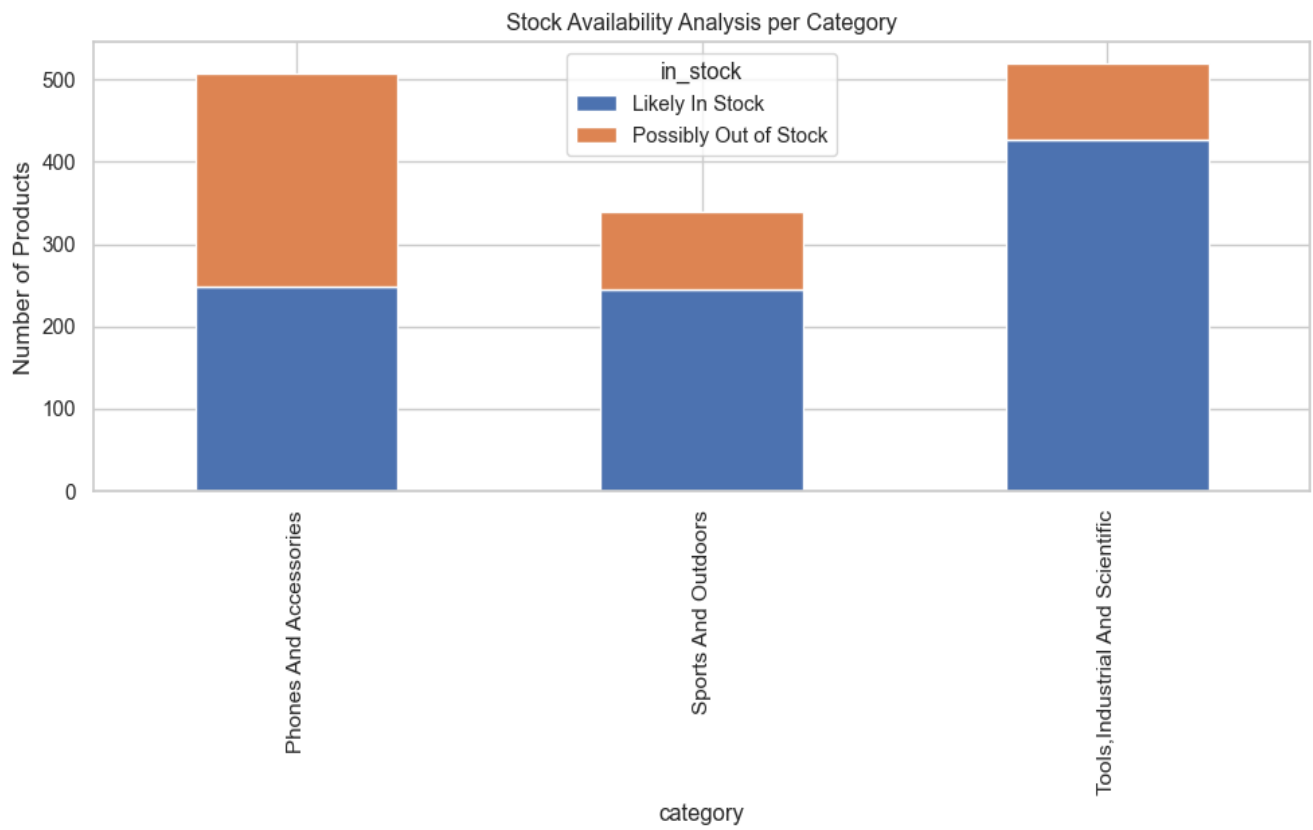
Price Distribution per Category



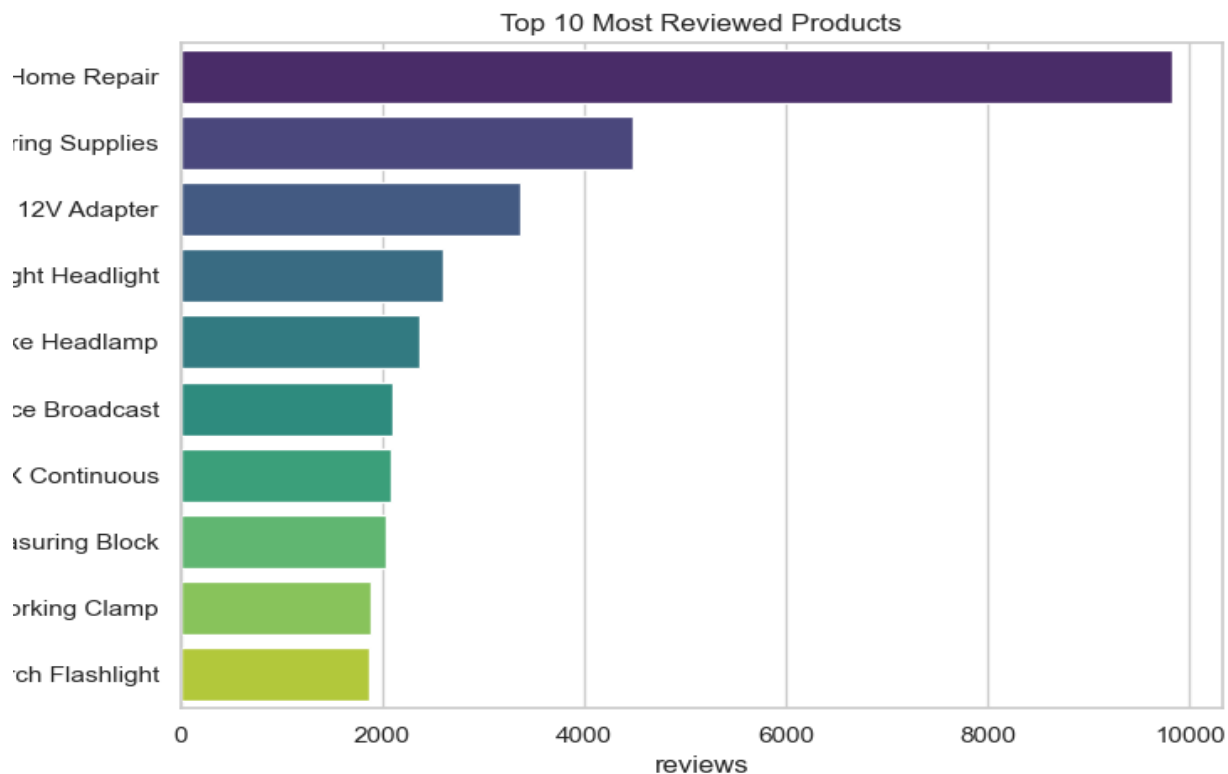
Rating vs Price Correlation



Stock Availability Analysis per Category



Top 10 Most Reviewed Products



SQL Aggregated Insights

SQL Server was used to compute analytical metrics including average price per category, rating summaries, product counts, top-reviewed items, and estimated stock availability across categories.

Final Conclusions & Recommendations

The ETL pipeline successfully produced a consolidated, clean dataset suitable for both Python and SQL analysis. Findings suggest strong consumer engagement in specific subcategories, pricing inconsistencies across categories, and opportunities for improvement in product visibility and promotional targeting.