INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PRESENTATION ON

Amazon Sneakers Sales Analysis

Presented by:
Sk. Sameer
Batch no: 456

Agenda

- Introduction / Sneakers Market Overview
- Business Problem
- Data Source
- Tools & Libraries Used
- Dataset Description
- Data Cleaning & Manipulation
- Exploratory Data Analysis (EDA)
- Visuals
- Challenges
- Conclusion

INNOMATICS RESEARCH LABS

Introduction:-

- Online shopping has become very popular, and Amazon is one of the biggest platforms for buying products.
- In this project, we collected sneakers data from Amazon using web scraping.
- The dataset includes information like product name, brand, price, rating, Usage Types.
- We used Exploratory Data Analysis (EDA) to study and understand this data.
- This analysis helps us see which brands are popular and what price ranges are common.
- It also shows how customer ratings are related to product prices.
- The main aim of this project is to find useful patterns and trends in sneaker sales.
- These insights can help in better business and marketing decisions.

Business Problem:-

Amit Sharma is a Student planning to buy a pair of sneakers for daily use. He visits Amazon and searches for sneakers, only to find thousands of options with similar-looking designs but huge price differences.

Despite similar features like:

Material quality, Brand type, Customer ratings, Usage type (casual, running, sports)

He observes: Some sneakers cost ₹999 – ₹1,499 Others cost ₹3,000 – ₹8,000+ This creates confusion for Amit and millions of online shoppers.

Business Problem Statement

Online sneaker buyers face difficulty in selecting the right product due to overwhelming choices and unclear pricing logic. The objective is to analyze Amazon sneaker data using Exploratory Data Analysis (EDA) to understand pricing patterns, feature impact, and value-for-money insights.

Data Source

- Data collected from Amazon India using web scraping.
- Focused on the [sneakers] product category.
- Extracted structured data from multiple product pages.
- **Key fields included:**
  - Product Name
  - Price
  - Rating
  - Brand
- Data cleaned and organized into a tabular format for analysis.
- Used to analyze pricing trends, product features, and customer preferences.

Tools and Libraries used :

- Python - core programming language
- Selenium - for browser automation and scraping dynamic content
- Requests - for sending HTTP requests to Amazon pages
- Pandas – For storing and cleaning Data
- Matplotlib & Seaborn – For Visualization

Dataset Description :

Data was collected from the Amazon website using web scraping techniques. Since Amazon pages load content dynamically, Selenium WebDriver was used.

- 400 + Brands
- Columns – Brand Name, Gender, MRP, Selling Price, Discount Percent, Rating, etc….
- Brands Included – Puma, Campus, US polo ASSN, Cruiser, Centrino, Bucca Buchi, Sketchers, etc….

	Brand Name	Gender	MRP	Selling Price	Discount Percent	Rating	Discount Amount	Price_Category	Rating_Category	Usage_Type
0	Boldfit	Men	1599	1499	56	3.8	100	Low-Mid	Average	Casual
1	Boldfit	Men	1599	1199	56	3.8	400	Low-Mid	Average	Casual
2	Campus	Men	1499	999	58	4.2	500	Budget	Good	Unknown
3	Boldfit	Men	999	1199	58	3.7	-200	Low-Mid	Average	Casual
4	Campus	Men	3499	699	38	3.8	2800	Budget	Average	Unknown

Column Description :

- Brand Name – Name of the sneaker brand.
- Gender – Target user category (Men, Women, Unisex).
- MRP – Original listed price of the product.
- Selling Price – Current discounted selling price.
- Discount Percent – Percentage of discount offered on MRP.
- Rating – Average customer rating of the product.
- Discount Amount – Amount saved (MRP – Selling Price).
- Price Category – Price range group (Budget, Low-Mid, Mid, Upper-Mid, Premium).
- Rating Category – Rating level group (Poor, Average, Good, Excellent).
- Usage Type – Intended use of the sneaker (Casual, Sports, Running, Sneaker).

Data Cleaning & Manipulation :

1. Handling Data Types
- Converted MRP and Selling Price columns to integer format.
- Converted Rating column to numeric (float) for analysis.
- Converted categorical columns like Price Category, Rating Category, Usage Type to category datatype

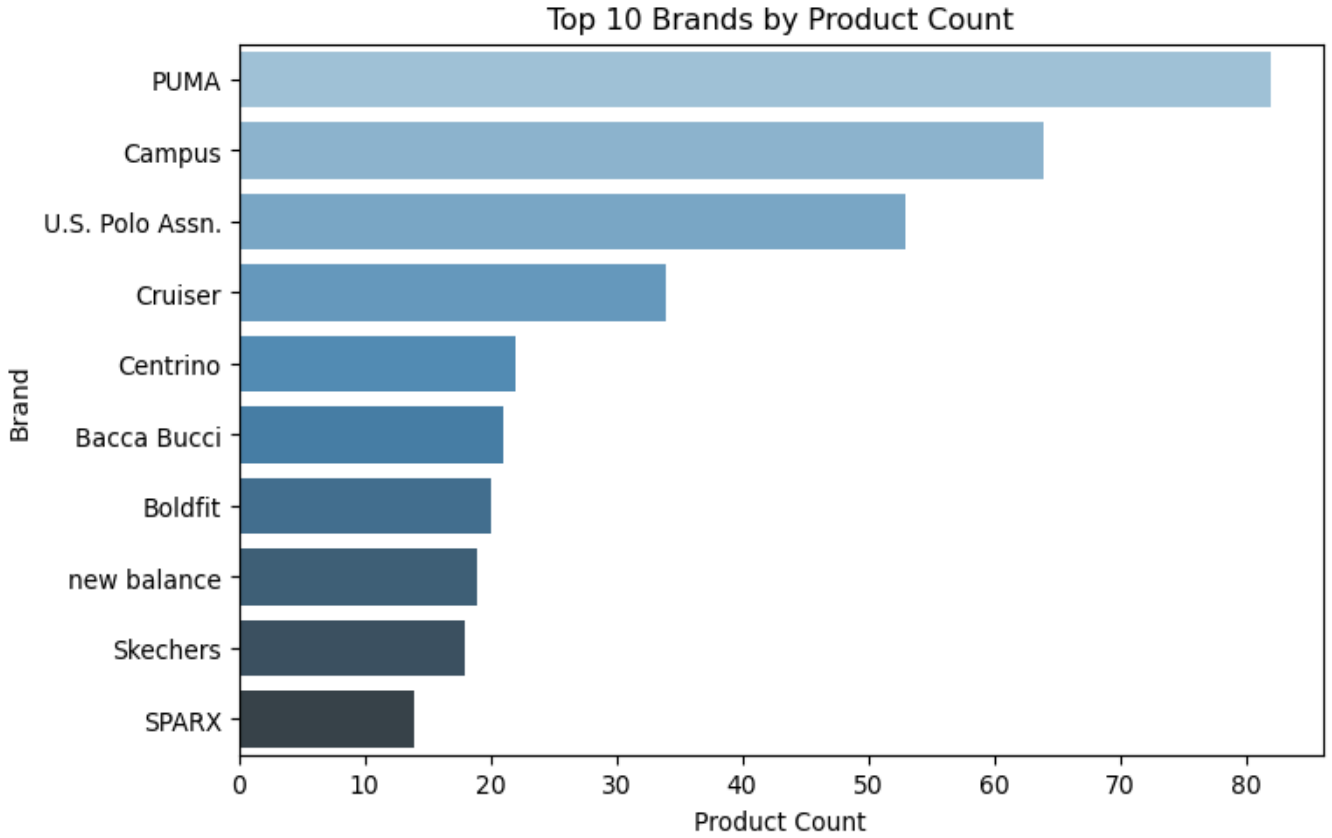
2. Feature Engineering (New Columns Created)
- Discount Amount - Difference between MRP and Selling Price.
- Price Category - Grouped prices into: Budget, Low-Mid, Mid, Upper-Mid, Premium.
- Rating Category - Classified ratings as: Poor, Average, Good, Excellent

3. Handling Missing Values
- Filled missing values in Rating & Rating Category using mode.

4. Outlier & Data Consistency Check
- Used boxplot to detect outliers, Found negative values in Discount Amount.
- Converted negative values to positive using absolute value.

EDA – Exploratory Data Analysis

Univariate Analysis :



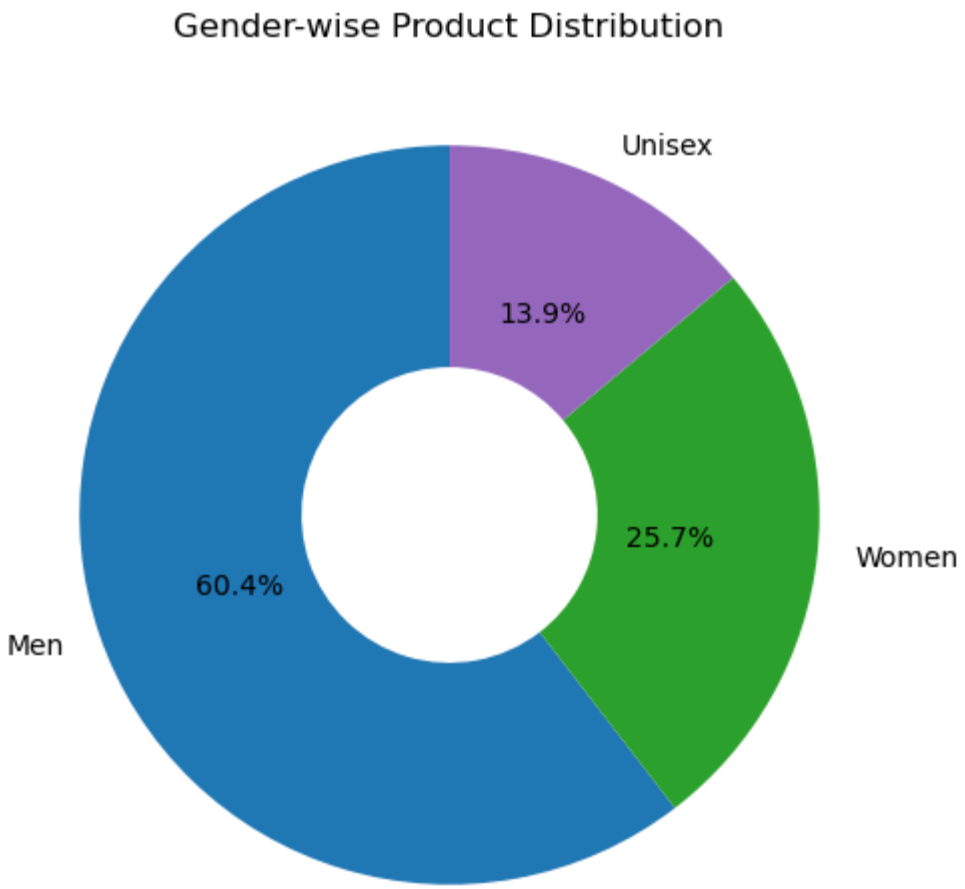
Plot Type: Horizontal Bar Chart

Key Insights:

- A few brands dominate the platform in terms of product availability.
- The remaining brands form a long tail with limited listings.

INNOMATICS RESEARCH LABS

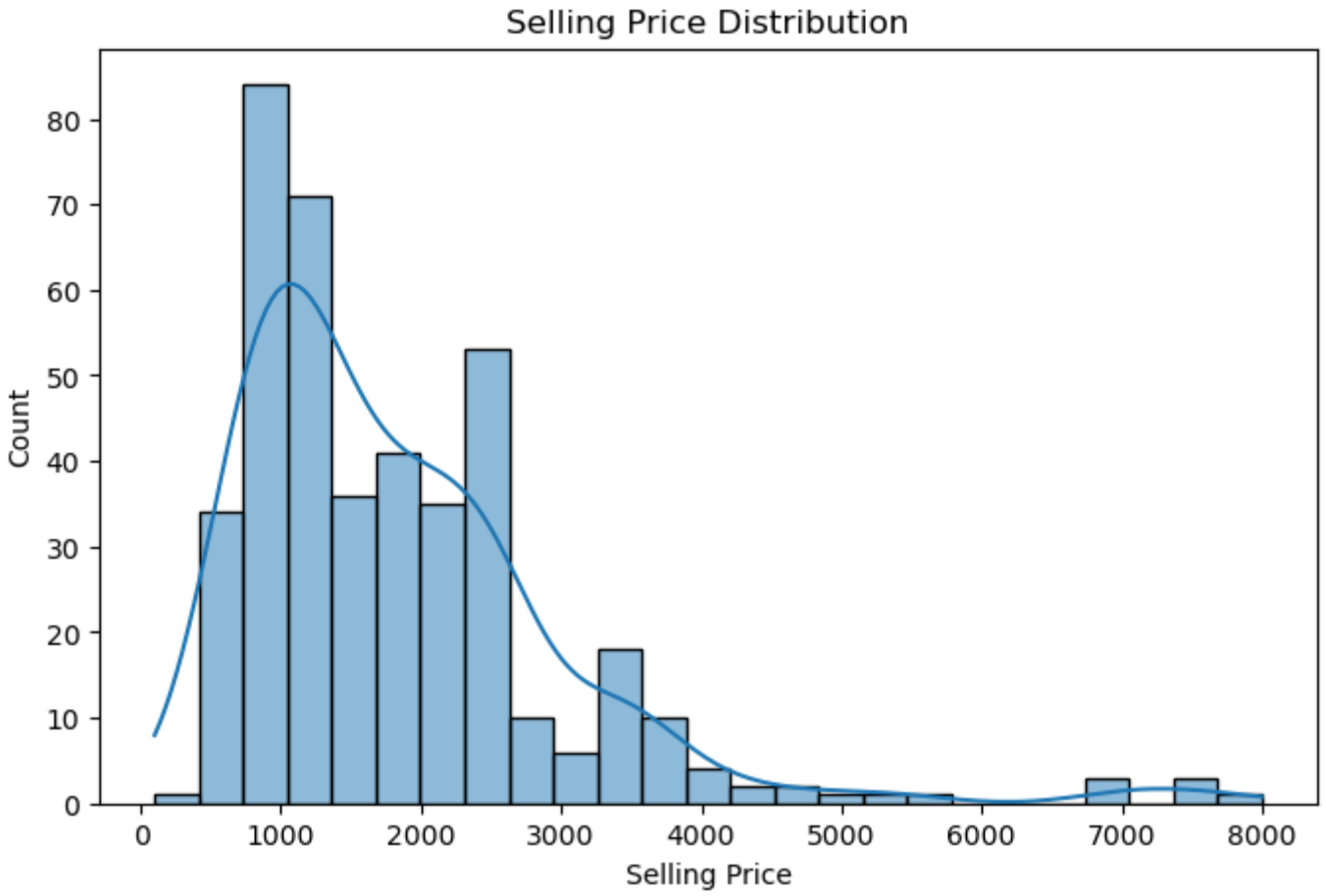
Gender-Wise Product Distribution :



Plot Type: Donut Chart

- Key Insights:
- Men's footwear clearly dominates the dataset.
- Women and unisex categories have relatively lower representation.

Selling Price Distribution :

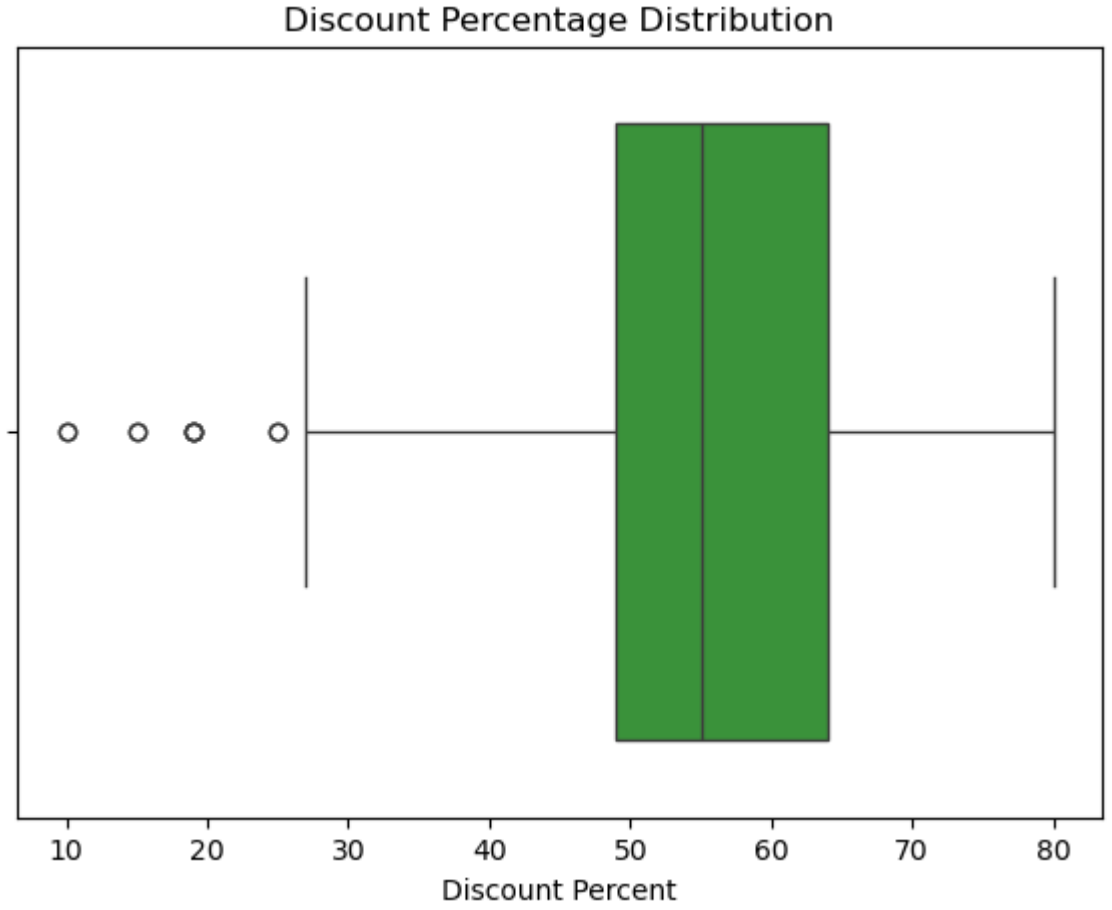


Plot Type: Histogram with KDE

Key Insights:

- Majority of products are concentrated in the lower price range.
- Few high-priced products create a right-skewed distribution.

Discount Price Distribution :



Plot Type: Box Plot

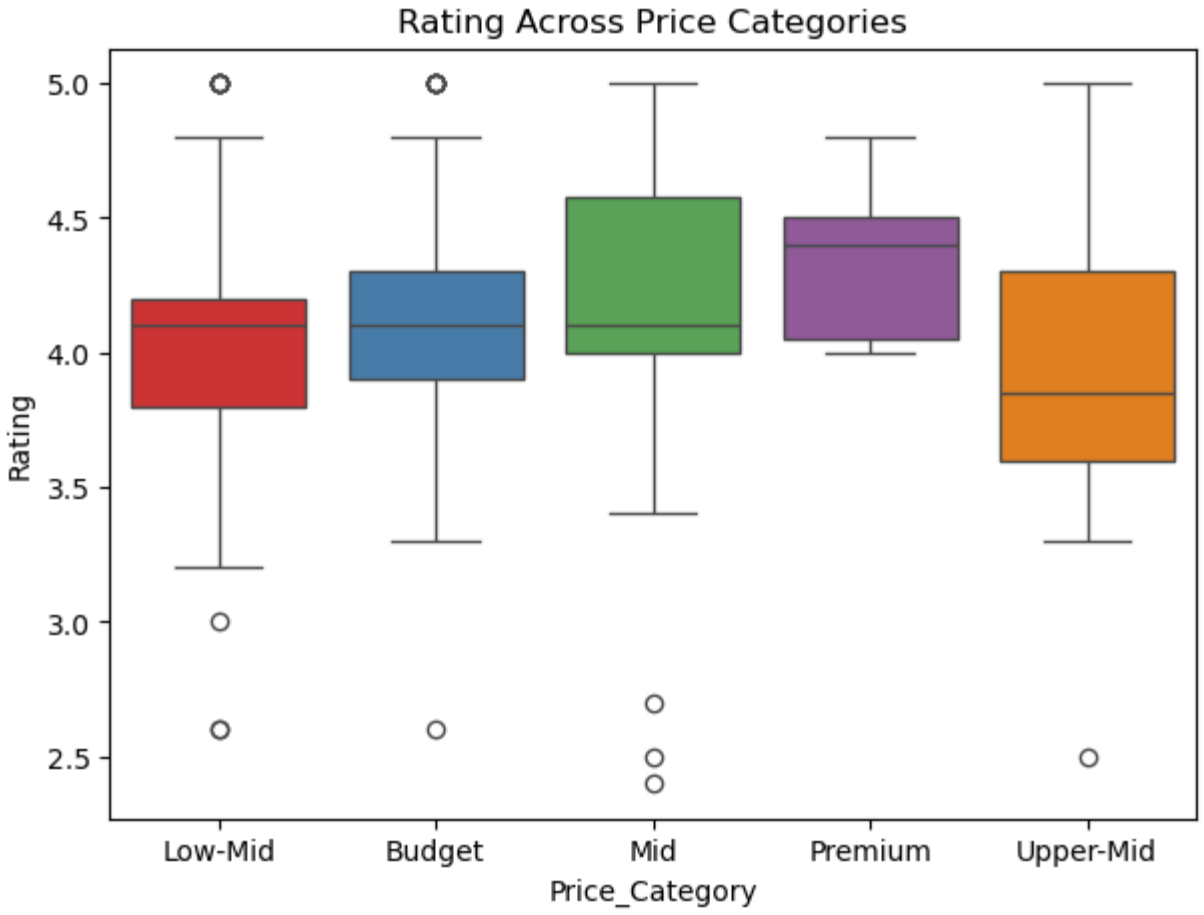
Key Insights:

- High discount percentages are common across most products.
- Outliers indicate aggressive promotional pricing strategies.

Most sneakers are sold with high discounts (around 50–60%), and only a few products have very low or extremely high discounts.

Bivariate Analysis

Rating Across Price Categories :



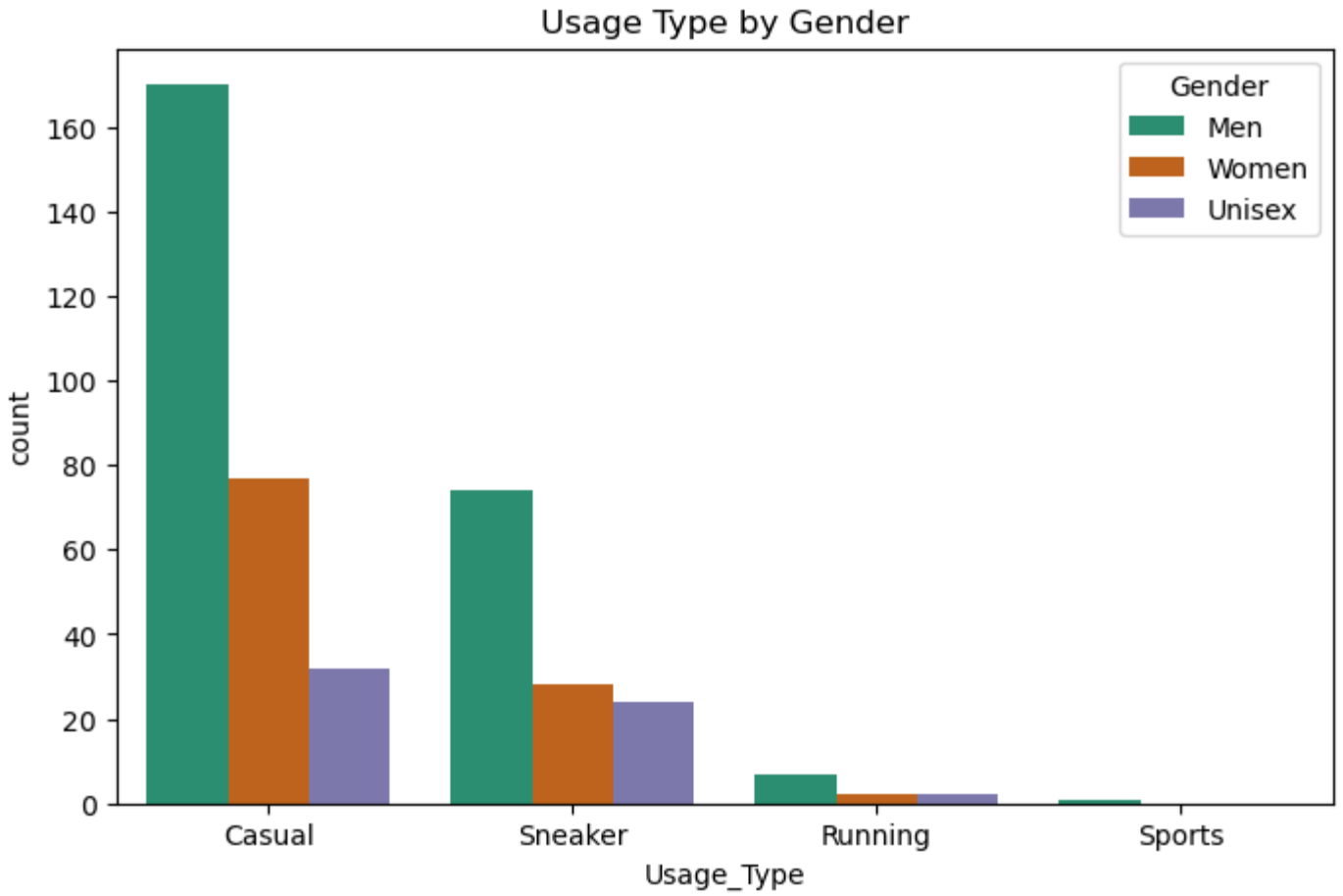
Plot Type: Box Plot

Key Insights:

- Higher price does not consistently translate into higher ratings.
- Value perception plays a stronger role than price alone.
- Ratings are similar across all price categories, so expensive shoes are not always rated better than cheaper ones.

INNOMATICS RESEARCH LABS

Usage Type by Gender:

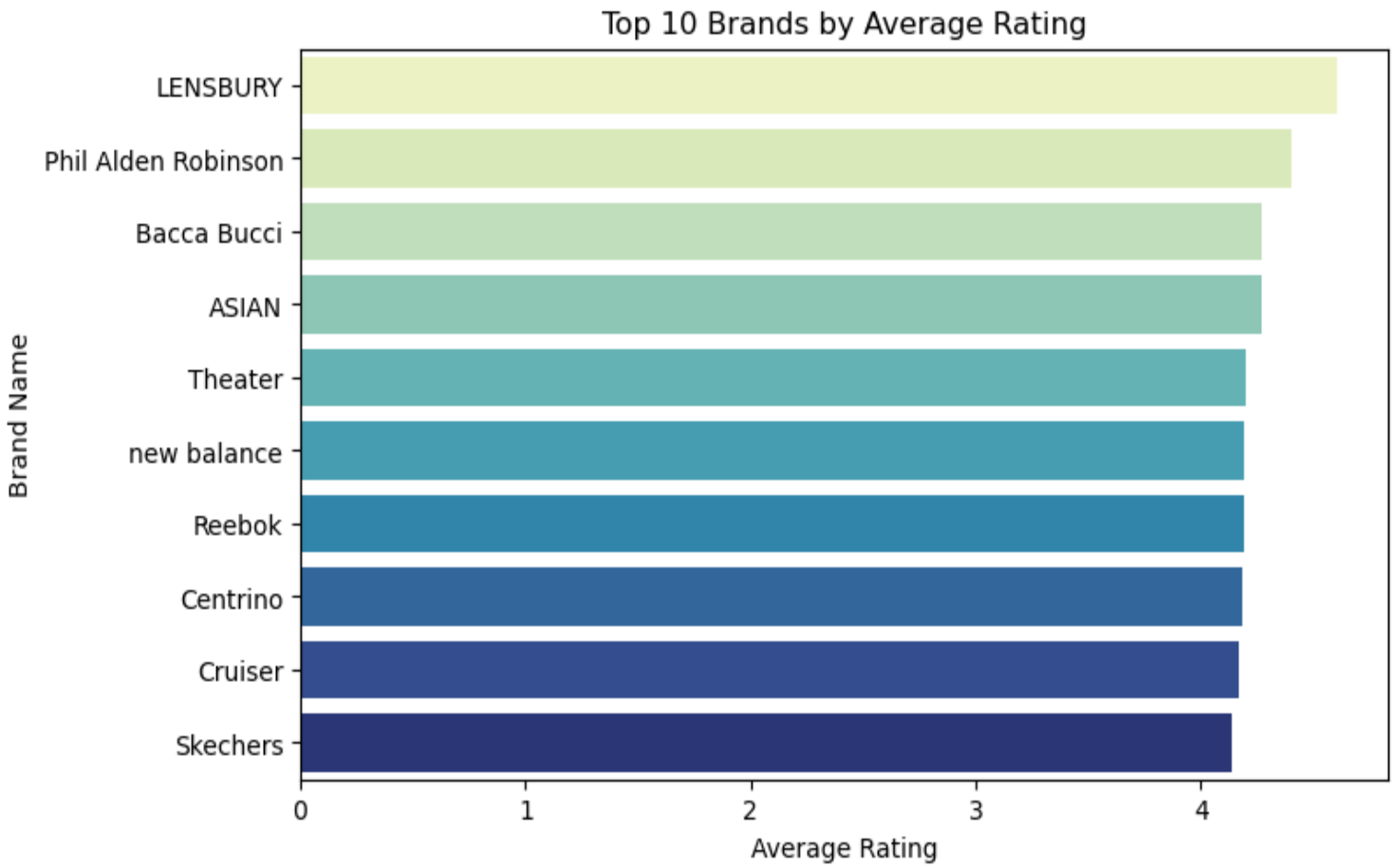


Plot Type: Grouped Count Plot

Key Insights:

- Casual footwear dominates across all genders.
- Men show higher product availability across usage types.

Top 10 Brands by Average Rating:



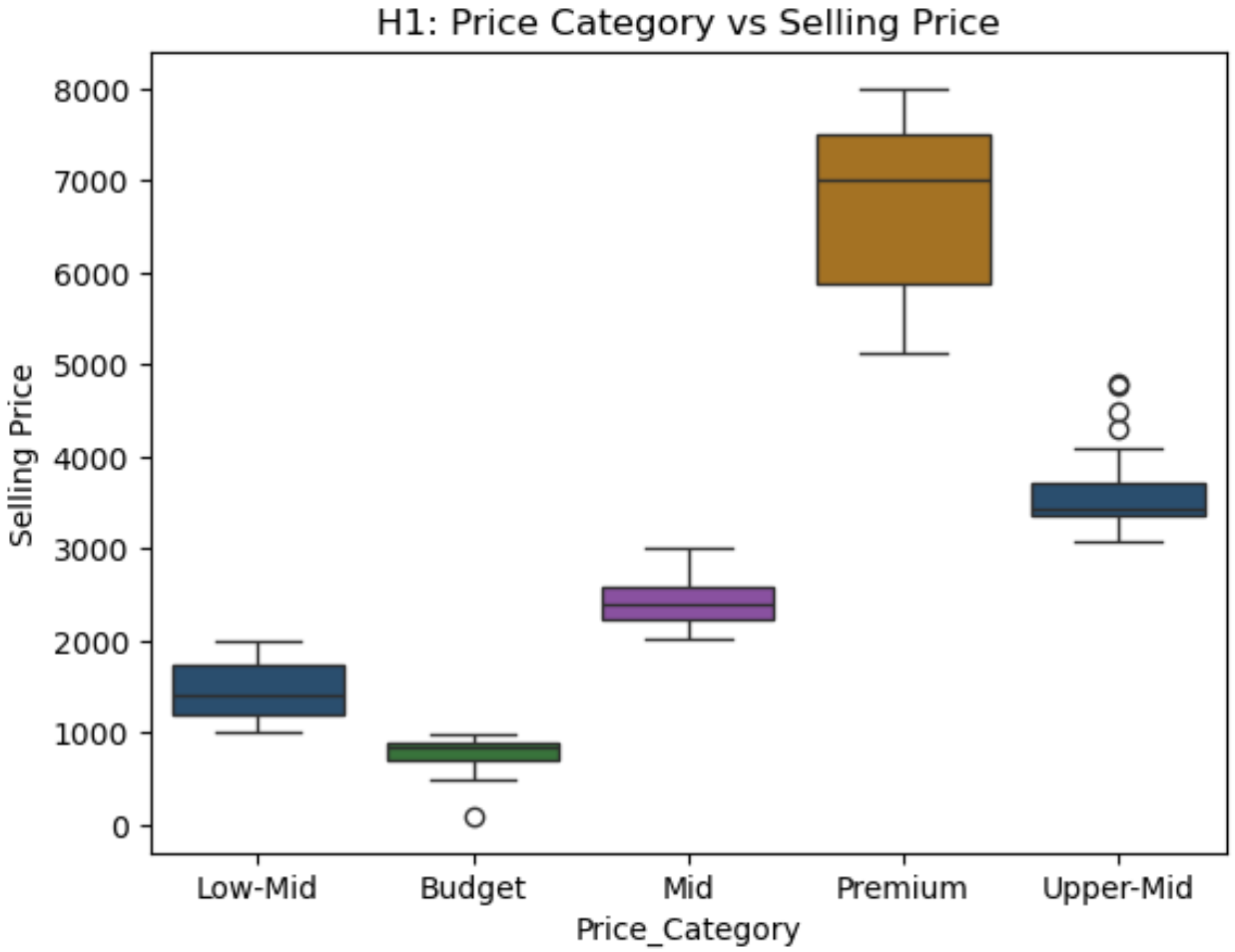
Plot Type: Aggregated Bar Chart

Key Insights:

- Some less frequent brands outperform popular brands in ratings.
- Product volume does not guarantee better customer satisfaction.

INNOMATICS RESEARCH LABS

Hypothesis Testing = Using data to prove or disprove an assumption.



Plot Type: Box Plot

Is there a significant difference in Selling Price across Price Categories? (One – Way ANOVA)

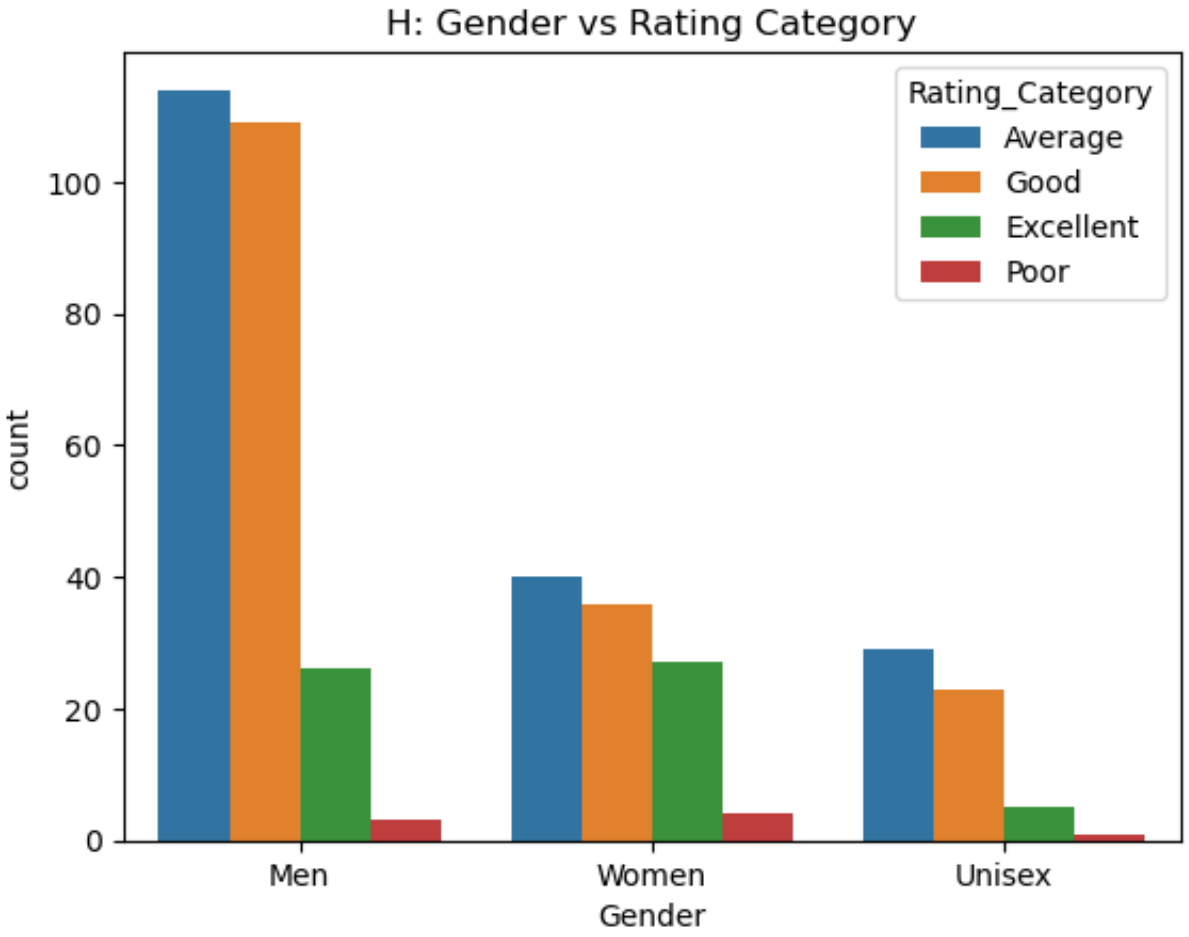
Conclusion

$P < 0.05$, Reject Null Hypothesis

Selling prices vary significantly across different price categories.

INNOMATICS RESEARCH LABS

Gender vs Rating Category:



Plot Type: Grouped Bar Chart

Does Rating Category depend on Gender? (Chi-Square Test)

Conclusion

$P < 0.05$, Reject Null Hypothesis

Customer rating category is significantly associated with gender.

INNOMATICS RESEARCH LABS

Difficulties :

- **Dynamic web pages** – Amazon loads content dynamically, so elements were not available immediately and required delays (time.sleep).
- **Missing values (NaN)** – Some products had no ratings or categories, which required filling using mode and category mapping.
- **Comma-separated prices** – Values like 1,499 caused type conversion issues and had to be cleaned manually.
- **Regex Mismatch** – I got inconsistent data even after using proper regex.
- **Extra Brand Names** – I got extra brand names which are extracted from sponsored charts.
- **Missing product information** – Some listings did not contain ratings or discount values.

INNOMATICS RESEARCH LABS

Conclusion :

- Men's sneakers are the most common products.
- Most sneakers are in the Budget and Low-Mid price range.
- Ratings are generally good (around 3.5 to 4.5).
- Casual and sneaker types are used the most.
- Prices and ratings change by brand and gender.
- Discounts are mostly similar across usage types.

Overall: Affordable, well-rated sneakers are the most popular, and gender-based marketing can improve sales.

THANK YOU

INNOMATICS RESEARCH LABS