

# Exploratory Data Analysis and Statistical Testing of Amazon Mobile Phones Under ₹30,000

## 1. Introduction

The rapid growth of e-commerce platforms such as Amazon has significantly influenced consumer purchasing behavior in the smartphone market. Customers often rely on price and user ratings to make informed decisions. This project aims to perform **Exploratory Data Analysis (EDA)** and apply **appropriate statistical tests** to understand the relationship between mobile phone prices and customer ratings for smartphones listed under ₹30,000 on Amazon.

---

## 2. Objectives of the Study

The main objectives of this study are:

1. To explore the distribution of mobile phone prices and ratings.
  2. To examine whether the average customer rating is significantly high.
  3. To study the relationship between price and rating.
  4. To compare prices across major brands using statistical methods.
  5. To test whether brand and rating category are statistically independent.
- 

## 3. Dataset Description

The dataset consists of **157 mobile phone listings** collected from Amazon. The variables included are:

- **Product Name:** Name of the mobile phone
  - **Price (INR):** Listed price of the phone
  - **Rating:** Customer rating (out of 5)
  - **Specs / Description:** Additional specifications or offers
- 

## 4. Data Cleaning

Before analysis, the dataset required cleaning:

- Prices were stored as text with commas and were converted into numeric values.
- Ratings were extracted from text (e.g., "4.4 out of 5 stars") and converted into numeric form.
- Rows with missing price or rating values were removed to maintain accuracy.

After cleaning, **153 valid observations** remained for analysis.

---

## 5. Exploratory Data Analysis (EDA)

### 5.1 Price Analysis

- The mean price of mobiles is approximately ₹31,743.
- The median price is around ₹27,999.
- The price distribution is **right-skewed**, indicating the presence of a few high-priced outliers.

### 5.2 Rating Analysis

- The average rating is approximately 4.25.
- Ratings show very low variability, with most values between 4.2 and 4.4.
- This indicates generally high customer satisfaction across listed products.

### 5.3 Price vs Rating

A scatter plot of price against rating shows **no strong linear relationship**, suggesting that higher-priced mobiles do not necessarily receive better ratings.

---

## 6. Statistical Tests Applied

### 6.1 One-Sample t-Test

**Objective:** To test whether the average rating is significantly greater than 4.2.

- Null Hypothesis ( $H_0$ ):  $\mu = 4.2$
- Alternative Hypothesis ( $H_1$ ):  $\mu > 4.2$

The one-sample t-test was applied since the population variance is unknown and the sample size is sufficiently large.

**Conclusion:** The test result supports rejecting the null hypothesis, indicating that the average rating is significantly greater than 4.2.

---

### 6.2 Pearson Correlation Test

**Objective:** To examine the relationship between price and rating.

- Null Hypothesis ( $H_0$ ):  $\rho = 0$  (no correlation)
- Alternative Hypothesis ( $H_1$ ):  $\rho \neq 0$

The correlation coefficient was found to be weak, and the p-value was not statistically significant.

**Conclusion:** There is no strong relationship between price and customer rating.

---

### 6.3 One-Way ANOVA

**Objective:** To test whether mean prices differ among major brands.

- Null Hypothesis ( $H_0$ ): Mean prices are equal across brands
- Alternative Hypothesis ( $H_1$ ): At least one brand has a different mean price

ANOVA was applied to the top three brands based on frequency.

**Conclusion:** The ANOVA test indicates a statistically significant difference in prices across brands.

---

### 6.4 Chi-Square Test of Independence

**Objective:** To test whether brand and rating category are independent.

Ratings were categorized into Low, Medium, and High. A contingency table was created between brand and rating category.

- Null Hypothesis ( $H_0$ ): Brand and rating category are independent
- Alternative Hypothesis ( $H_1$ ): Brand and rating category are not independent

**Conclusion:** The chi-square test suggests that brand and rating category are not completely independent.

---

## 7. Key Findings

- Most mobiles are priced between ₹25,000 and ₹30,000.
  - Customer ratings are consistently high across brands.
  - Price does not significantly influence customer ratings.
  - Different brands follow distinct pricing strategies.
- 

## 8. Limitations of the Study

- The data is limited to Amazon listings only.
  - Ratings are subjective and may contain bias.
  - Some extreme price outliers may influence summary statistics.
- 

## 9. Conclusion

This study demonstrates that while mobile phone prices vary significantly across brands, customer ratings remain consistently high. Statistical analysis confirms that higher prices do not guarantee better customer

satisfaction. The combination of EDA and hypothesis testing provides meaningful insights into consumer trends in the mid-range smartphone market.

---

## 10. References

- Amazon India (product listings)
- Statistical Methods for Business and Economics
- Pandas and SciPy documentation