Flipkart Mobile Phones Exploratory Data Analysis

This document walks through the exploratory data analysis performed on Flipkart smartphone listings, covering cleaning, feature engineering, descriptive insights, key business questions and actionable conclusions.

0 Objective

Provide data driven answers to:

1 What specification attributes correlate most with user ratings?

2 Which price bands are overcrowded or underserved?

3 Which models deliver the best performance per rupee?

4 Are there data quality issues that need remediation before predictive modelling?

1 Dataset snapshot

Source: flipkart_mobile.csv (240 raw rows). After deduplication-ready cleaning, 234 unique phone variants with full numeric specs remain. Key fields: Product Name, Description, Price, Rating, RAM_GB, Battery_mAh, Camera_MP.

2 Business questions & data-backed answers

Q1. Which specs drive higher ratings?

Price, RAM and Battery show the strongest positive correlations with Rating (r=0.75, 0.68, 0.59 respectively). Camera megapixels correlate slightly negative (-0.22).

Q2 What price band is most crowded?

A 63 percent of models fall between Rs 8 000 and Rs 18 000, with a sharp drop in offerings beyond Rs 25 000.

Q3 Which phones offer the highest value? A Normalising (RAM+Battery+Camera+Rating) and dividing by price identified Infinix Note 40X 5G, vivo T4x 5G and Motorola g05 (budget) as top value picks (Value >0.34).

Q4 Any immediate data quality problems?

A 213 duplicate listings, 6 phones missing camera spec, and a few extreme price outliers (beyond 99th percentile Rs 29k). These should be addressed before modelling.

3 Descriptive statistics

Price: Rs 5999-28999 (mean 13591)

RAM: 3.0-12.0 GB (median 7.0)

Battery: 5000-7000 mAh (median 5160)

Camera: 32-108 MP (median 50)

4 Data quality summary

- 213 duplicate names (likely multi-listings)

- 6 missing Camera_MP

- Price outliers outside Rs 6k-29k range

Proposed fixes: deduplicate by lowest price, impute/drop missing camera, winsorise price extremes.

5 Conclusion & next steps

Mid-tier segment (Rs 8k-18k) is highly competitive; spec increments in RAM and Battery translate to higher user satisfaction, while camera megapixel marketing has limited impact. A handful of phones deliver standout value; these could be highlighted in merchandising.

Data cleanup (duplicates, imputation, outlier handling) is recommended before building a predictive pricing or recommendation model.