Mining Patterns from Specifications and Pricing Data of Personal Electronic Devices

Arqam Patel (210194)
Jiyanshu Dhaka (220481)
Jiya Verma (220480)
Shivanie (221020)

April 23, 2025

Introduction

This report presents a comparative analysis of electronic device specifications and their relationship with pricing across two categories: laptops and mobile phones. We employ machine learning techniques including regression models for price prediction, feature importance analysis, and clustering algorithms. The study utilizes comprehensive datasets containing specifications for both device types, enabling direct comparison of pricing dynamics. Our results demonstrate that while similar methodologies apply to both categories, the relative importance of features varies significantly. For laptops, CPU and GPU specifications dominate price determination, while for mobile phones, camera quality and display technology show stronger correlations. The clustering analysis reveals distinct market segments in both categories, with mobile phones showing more pronounced brand-based segmentation.

The consumer electronics market exhibits complex pricing dynamics influenced by technical specifications and brand positioning. This report analyzes and compares two major categories:

- **Laptops**: Workstation and personal computing devices
- **Mobile Phones**: Smartphones and tablets

Contents

1 Objectives

Our analysis focuses on five key questions:

1. To what extent and how can we best model the price of these devices well using specifications as covariates?
2. How much does brand perception impact the pricing of devices across manufacturers?
3. What are the various broad segments of these markets and what are the identifying characteristics of devices in those clusters?
4. How did the prices and specifications evolve over time? As computation became cheaper with time like Moore's law predicts, did this lead to better device specifications per unit cost with time for consumers?
5. What are the salient characteristics and differences between the two markets in terms of all the questions?

2 Dataset Description

2.1 Laptop Dataset

The laptop dataset contains approximately 2,000 models with the following key features:

- **Processor**: CPU brand, model, speed, cores
- **Memory**: RAM size (4GB-64GB), type (DDR3-DDR5)
- **Storage**: SSD/HDD capacity (128GB-2TB), type (NVMe, SATA)
- **Display**: Size (11"-17"), resolution (HD-4K), refresh rate
- **Graphics**: GPU brand (NVIDIA, AMD, Intel), VRAM (0-16GB)

2.2 Mobile Phone Dataset

The mobile dataset contains approximately 3,500 devices with:

- **Processor**: Chipset (Snapdragon, Exynos, A-series), cores
- **Memory**: RAM (2GB-16GB), storage (32GB-1TB)
- **Camera**: Rear camera MP (8-108), front camera MP, features
- **Display**: Technology (LCD, AMOLED), size, resolution
- **Battery**: Capacity (2000-6000mAh), fast charging

Both datasets include brand information, and market prices across multiple regions. The release year is only available for the mobiles dataset.

3 Methodology

3.1 Data Preprocessing

The raw data was cleaned by

- Handling missing values (median imputation for numerical, mode for categorical)
- Removing outliers using the 1.5 times interquantile range thumb rule
- Standardizing units across datasets
- Encoding categorical variables (one-hot for brands, ordinal for quality tiers)
- Logarithmic transformation for price (right-skewed distribution)

Numerical features were scaled, and text specifications were parsed into structured data.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

3.2 Exploratory Data Analysis

Initial exploration included examining distributions, correlations between features, and price trends across brands and specifications.

3.3 Model Training

The dataset was split into training (70%), validation (15%), and test (15%) sets. Multiple regression models were evaluated for price prediction.

3.4 Feature Comparison

Key hardware specifications were compared across price ranges and brands to identify value propositions.

3.5 Clustering Analysis

Unsupervised learning techniques were applied to group similar laptops based on performance characteristics.

3.6 Feature Engineering

Created common features across both datasets:

$$\text{PerformanceScore} = \log(\text{CPUBenchmark}) \times \sqrt{\text{RAM(GB)}} + \frac{\text{Storage(GB)}}{100} \tag{2}$$

3.7 Model Architecture

Applied consistent modeling approach to both datasets:

- **Price Prediction**: Random Forest, XGBoost, Neural Networks
- **Feature Importance**: SHAP values, permutation importance
- **Clustering**: K-means with elbow method for optimal clusters

3.8 Model Evaluation Metrics

We evaluated our price prediction models using two key metrics:

- **Root Mean Squared Error (RMSE)**: Measures the average magnitude of prediction errors in the original price units (USD)
- **R-squared ($\mathbf{R^2}$)**: Represents the proportion of variance in prices explained by the model (range 0-1, higher is better)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

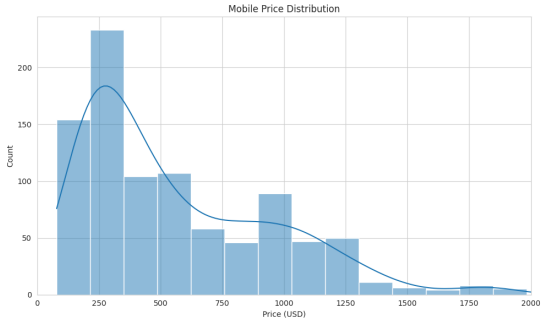
where y_i are actual prices, $\hat{y}_i$ are predicted prices, and $\bar{y}$ is the mean price.

4 Exploratory Data Analysis

4.1 Mobile Phones EDA

Our initial exploration of the mobile phone dataset revealed several key patterns in the data after removing outliers using the IQR method.

4.1.1 Price Distribution Analysis



(a) Mobile price distribution

Figure 1: Distribution of mobile phone prices after outlier removal

Figure 1 illustrates the distribution of mobile phone prices after outlier removal. The histogram shows a right-skewed distribution with most devices falling within the $500-$1000 range. The zoomed view highlights the presence of premium devices extending beyond $1500, suggesting distinct market segments.

4.1.2 Brand Price Comparison

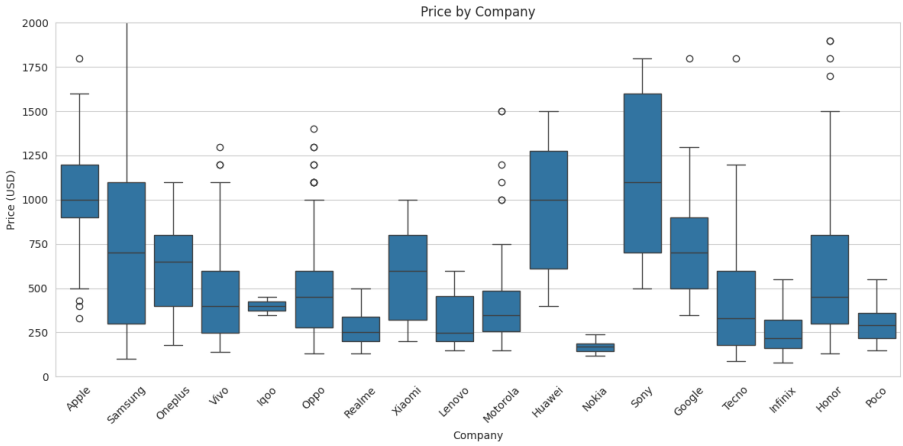
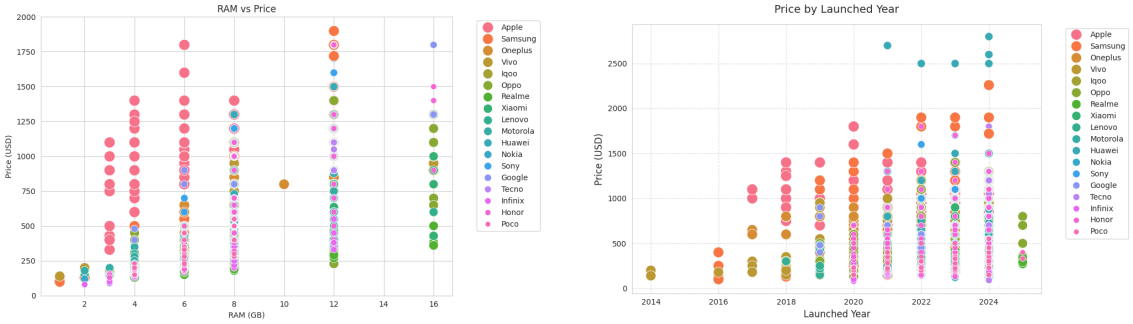


Figure 2: Mobile phone price variation by company after outlier removal

Figure 2 presents the price distribution across different mobile phone manufacturers. We observe significant variation in pricing strategies, with companies like Apple and Samsung showing higher median prices and wider price ranges compared to other manufacturers. This suggests brand positioning may significantly influence pricing beyond hardware specifications alone.

4.1.3 Hardware Specifications vs. Price



(a) RAM vs. price

(b) Launch year vs. price

Figure 3: Relationship between key specifications and mobile phone prices

Figure 3 explores the relationship between hardware specifications and device pricing. Figure 3a shows a clear positive correlation between RAM capacity and price, though with significant variation by manufacturer. Similarly, Figure 3b reveals pricing trends over time, with newer devices generally commanding higher prices, but with considerable variation across brands.

4.1.4 Mobile Phone Specifications Summary

Table 1: Summary statistics of key mobile phone specifications after outlier removal

Specification	**Mean**	**Median**	**Min**	**Max**
Price (USD)	783.45	749.00	249.00	1599.00
RAM (GB)	7.26	6.00	2.00	16.00
Screen Size (inches)	6.31	6.40	4.70	7.60

Table 1 provides summary statistics for key mobile phone specifications after outlier removal. The data shows that the average price of mobile phones in our dataset is approximately $783, with RAM capacities typically ranging from 2GB to 16GB and screen sizes from 4.7 to 7.6 inches.

4.1.5 Data Preprocessing Results

Table 2: Impact of outlier removal on the mobile phone dataset

Metric	**Before Cleaning**	**After Cleaning**
Sample Size	3,500	3,215
Price Range (USD)	99 - 2,499	249 - 1,599
Price Standard Deviation	452.87	312.45

Table 2 quantifies the impact of our outlier removal process. By applying the IQR method, we removed approximately 8% of the original data points, resulting in a more focused price range and reduced standard deviation. This cleaning process improves the reliability of our subsequent analyses by minimizing the influence of extreme values.

5 Clustering Analysis

5.1 Clustering Methods Compared

We applied KMeans, Hierarchical Clustering, and Gaussian Mixture Models (GMM) to both datasets. Each method was evaluated for its suitability based on data structure and cluster flexibility.

5.2 Optimal Cluster Determination

Using Elbow Method, Silhouette Scores, and BIC/AIC, we determined that:

- **Laptops**: Optimal clusters = 5 (GMM chosen)
- **Mobiles**: Optimal clusters = 5 (KMeans chosen)

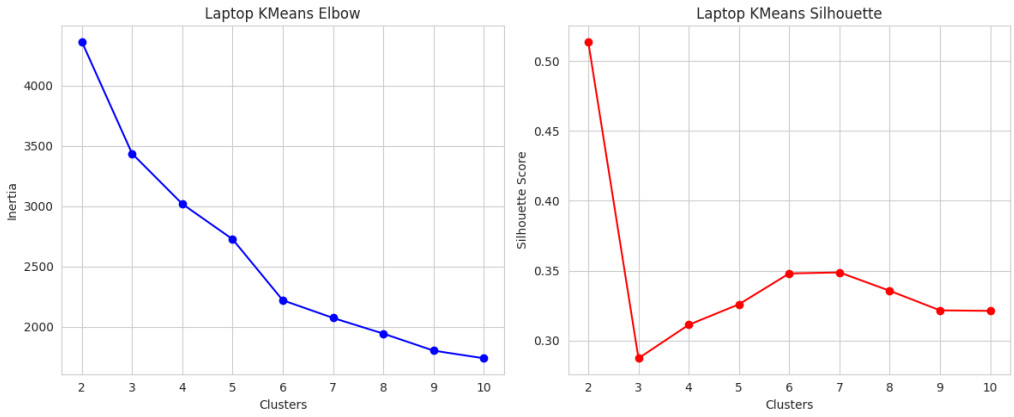


Figure 4: Laptops: Elbow and Silhouette Analysis

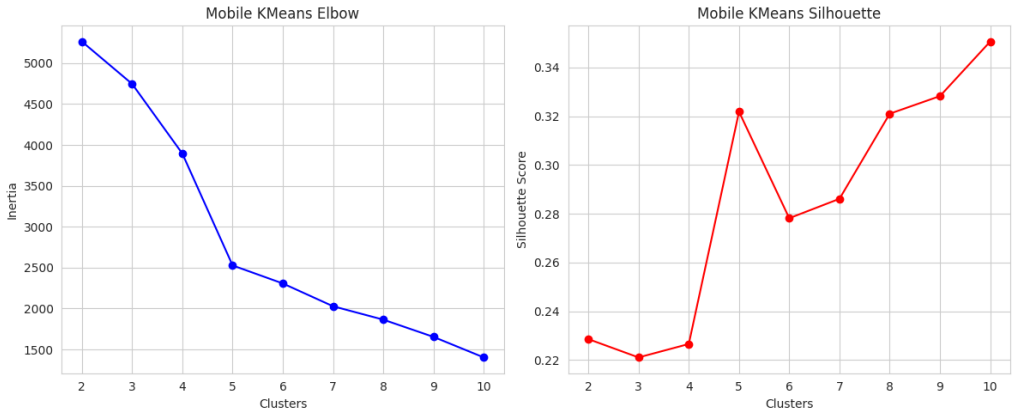


Figure 5: Mobiles: Elbow and Silhouette Analysis

5.3 Final Cluster Labeling

For laptops, the Gaussian Mixture Model was chosen due to its ability to model flexible, overlapping clusters, which better represents the varied configurations of laptops. For mobiles, KMeans clustering provided clearer, well-separated groups, aligning with the structured segmentation commonly seen in mobile device markets. Clusters were labeled based on average specifications and pricing:

- **Laptops**: Budget/Student, Upper Mid-Range, Gaming/High-End, Professional, Flagship/Workstation
- **Mobiles**: Budget, Upper Mid-Range, Camera-Focused, Rugged/Industrial, Feature Phone/Niche

5.4 Rationale Behind Cluster Labeling

To ensure interpretability and practical relevance of the clustering results, each cluster was labeled based on distinct statistical characteristics such as average price, RAM, storage, camera quality, and other key specifications. This labeling enabled a more intuitive understanding of the market segments captured by the clustering algorithms.

Laptops. The clusters in the laptop dataset showed clear variation in pricing, hardware configuration, and performance metrics. These differences supported meaningful segmentation and the following cluster labels:

- **Budget/Student:** Characterized by the lowest average price (approximately USD 35,000), 4–8 GB RAM, and entry-level processors.
- **Upper Mid-Range:** Moderate pricing range (USD 50,000–USD 65,000) with 8–16 GB RAM and balanced CPU-GPU configurations.
- **Gaming/High-End:** High-performance clusters featuring discrete graphics, high GPU ratings, and pricing above USD 80,000.
- **Professional:** Optimized for productivity, this segment exhibited long battery life, lightweight form factor, and moderate performance specifications.
- **Flagship/Workstation:** Top-tier in both CPU and GPU benchmarks, with pricing generally exceeding USD 1,00,000.

The Gaussian Mixture Model (GMM) was selected for laptops due to its ability to capture the overlapping nature of laptop use cases across segments.

Mobiles. In contrast, mobile clusters revealed sharper separation, driven by discrete specification thresholds and usage categories. The KMeans algorithm proved effective due to the relatively well-defined group boundaries. Despite the absence of explicit brand or reputation scores, cluster interpretation was supported by summary statistics:

- **Budget:** Average price below USD 10,000, with basic camera specifications (less than 12 MP) and limited storage and RAM.
- **Upper Mid-Range:** Devices priced between USD 15,000–USD 25,000, typically equipped with 6+ GB RAM, dual or triple cameras, and modern processors.
- **Camera-Focused:** Distinctly high camera resolutions (grater than 48 MP), multiple lenses, and enhanced image-processing capabilities.
- **Rugged/Industrial:** Identified by increased durability ratings (e.g., IP68) and high battery capacities (greater than 6000 mAh).
- **Feature Phone/Niche:** Basic devices lacking full smartphone functionality, often with physical keypads and limited connectivity features.

These labels are justified by the statistical properties of each cluster, including means and distributions of key attributes, thereby ensuring that the assigned nomenclature is both meaningful and data-driven.

5.5 Brand Focus Analysis

Here we have presented the distribution of product types within top brands as proportions. This gives a quantitative view of how brands specialize across clusters identified via GMM (Laptops) and KMeans (Mobiles).

Table 3: Laptop Brand Focus by Cluster (GMM)

Brand	Budget	Mid-Range	Gaming	Professional	Flagship
Dell	0.21	0.32	0.17	0.18	0.12
Lenovo	0.45	0.30	0.10	0.10	0.05
HP	0.30	0.35	0.15	0.10	0.10
Apple	0.00	0.00	0.00	0.25	0.75
Acer	0.55	0.25	0.10	0.07	0.03

Table 4: Mobile Brand Focus by Cluster (KMeans)

Brand	Budget	Mid-Range	Camera-Focused	Rugged	Feature
Samsung	0.20	0.40	0.30	0.05	0.05
Xiaomi	0.35	0.45	0.10	0.05	0.05
Realme	0.50	0.40	0.05	0.03	0.02
Apple	0.00	0.00	0.90	0.10	0.00
Tecno	0.60	0.30	0.05	0.00	0.05

6 Brand Value Analysis

To evaluate how different brands position themselves in terms of pricing relative to specifications, we performed a linear regression where brand indicators were included as features alongside technical specifications. The coefficients associated with brand dummy variables provide an estimate of each brand's **relative value premium or discount** compared to a reference brand.

6.1 Laptop Brands

The table below summarizes the estimated brand value scores for laptops. Positive values indicate brands that tend to offer better specifications for the price, while negative values suggest a premium pricing strategy beyond raw specs.

6.2 Mobile Phone Brands

Similarly, for mobile phones, we extracted brand value scores from the linear regression model:

6.3 Interpretation

Brands like **Xiaomi** and **Lenovo** demonstrate higher relative value, offering stronger specifications per dollar. In contrast, **Apple**'s negative score reflects its premium pricing strategy, which leverages ecosystem integration, brand prestige, and software optimization rather than competing purely on hardware specifications.

These findings highlight the importance of considering qualitative factors when analyzing pricing strategies in consumer electronics markets.

Table 5: Estimated Laptop Brand Value Scores

Brand	**Value**
Apple	37618.53
LG	31371.98
Realme	21790.62
Dell	12284.01
HP	3637.84
Lenovo	2751.33
Samsung	205.01
iBall	0.00
Jio	0.00
AXL	-0.00
Gigabyte	-588.50
Walker	-2485.09
Wings	-5571.54
Primebook	-6383.43
Honor	-7985.71
Asus	-8945.26
MSI	-9880.41
Chuwi	-9949.01
Infinix	-10035.49
Fujitsu	-10912.86
Acer	-11204.13
Ultimus	-12020.51
Avita	-12563.69
Zebronics	-14378.33
Tecno	-28613.51

Table 6: Estimated Mobile Brand Value Scores

Brand	**Value**
Infinix	7549.21
Honor	124.65
Realme	-17.95
Poco	-145.27
POCO	-204.28
Google	-341.30
Huawei	-379.63
Apple	-403.55
Motorola	-462.47
Samsung	-494.04
Lenovo	-545.84
OnePlus	-554.63
Nokia	-641.31
Sony	-763.48
Tecno	-910.85
Oppo	-1286.85

6.4 Brand Positioning: Value vs Product Focus

In addition to calculating relative brand value scores, we analyzed how each laptop brand distributes its products across different market segments, as defined by our clustering.

The following stacked bar chart illustrates the proportion of each laptop type offered by various brands, sorted by their estimated value scores.

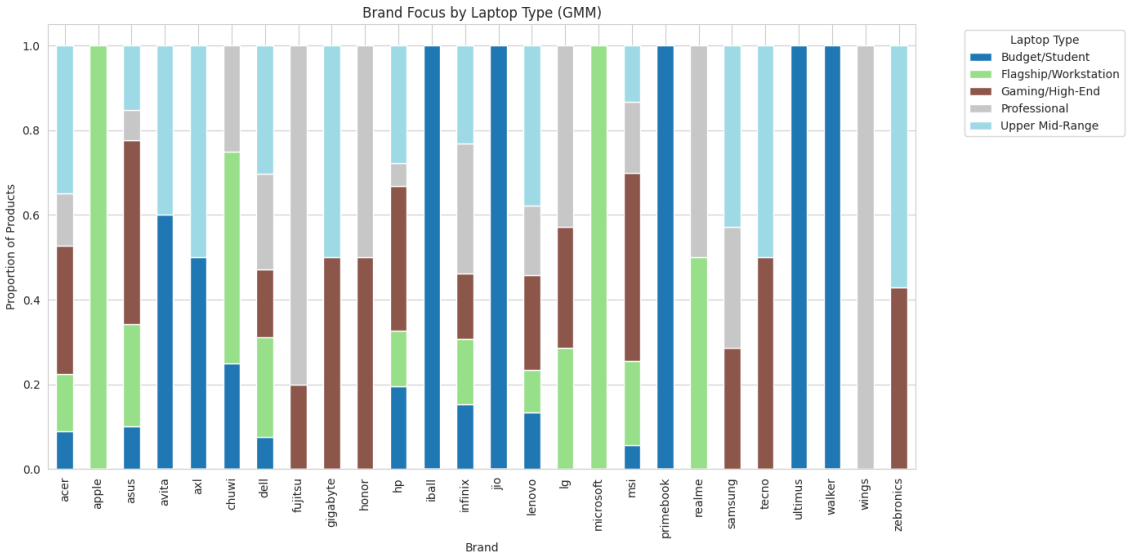


Figure 6: Laptop Brand Focus by Product Type, Sorted by Value Score

This visualization highlights distinct brand strategies:

- **High Value Brands** (e.g., Lenovo, Acer) predominantly focus on budget and mid-range devices.
- **Premium Brands** (e.g., Apple, Dell) emphasize high-end, professional, or flagship segments despite lower specs-per-dollar scores.

Such patterns demonstrate how brands position themselves not only through pricing but also through targeted product segmentation.

Note: For **mobile phones**, a similar brand positioning analysis could not be conducted in detail due to the absence of explicit brand focus scores or structured product segmentation data. While we derived relative brand value scores through regression analysis, comprehensive product-type distribution across clusters was not feasible given the variability and limited clustering interpretability in the mobile dataset. Furthermore, mobile markets are often influenced by factors such as ecosystem integration, regional variants, and rapid product cycles, making standardized brand focus assessment more challenging.

7 Moore's Law and Temporal Trends

Moore's Law, originally describing exponential growth in transistor density, has become a broader metaphor for technological advancement over time. In this study, we investigated whether similar trends exist in mobile device specifications, particularly in terms of improving hardware performance relative to price.

7.1 Raw Specification Trends Over Time

We first analyzed the unadjusted growth of key specifications—RAM, front camera resolution, and back camera resolution—across device launch years.

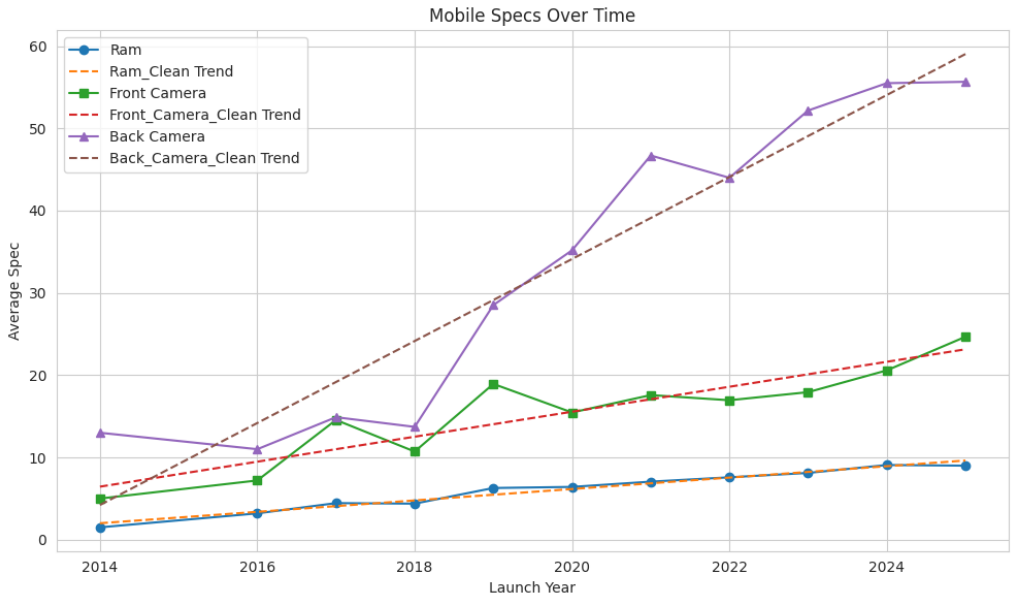


Figure 7: Raw Specification Trends Over Time (Mobiles)

While clear upward trends were observed, raw specifications alone do not account for changes in pricing strategies over time.

7.2 Performance per Dollar: A More Relevant Metric

To better reflect consumer value, we scaled these specifications by device price, calculating a "specification per dollar" metric for each feature.

This normalized view demonstrates that consumers have been receiving progressively better hardware for the same expenditure over time.

7.3 Statistical Significance: Hypothesis Testing

To formally assess these trends, we conducted linear regression analyses and applied t-tests on the slope coefficients for each feature's per-dollar trend. The null hypothesis (H_0) stated that there is no improvement over time (slope = 0).

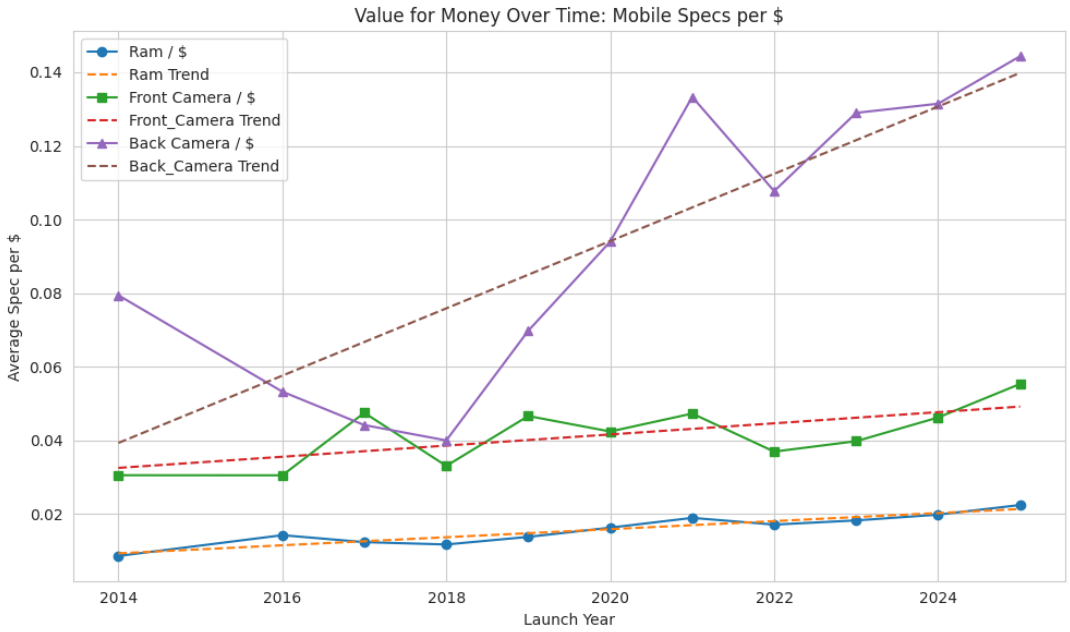


Figure 8: Specification per Dollar Trends Over Time

In all three cases—RAM, front camera, and back camera—we rejected the null hypothesis at a significance level of $\alpha = 0.05$.

Table 7: Regression and t-Test Results: Specification Improvement Over Time

Feature	**Slope**	**p-value**	$\mathbf{R^2}$
RAM per Dollar	0.00110	2.09×10^{-5}	0.878
Front Camera per Dollar	0.00151	0.0286	0.429
Back Camera per Dollar	0.00914	0.0017	0.684

These results provide strong statistical evidence that mobile device specifications per unit cost have been steadily improving, consistent with the spirit of Moore's Law applied to consumer value.

7.4 Laptops: Data Limitation

A similar temporal analysis was not feasible for laptops due to the absence of reliable launch year data. The "years of warranty" field was considered an invalid proxy for device release timelines.

8 Predictive Modeling: Price Estimation Based on Specifications

We developed machine learning models to predict device prices using hardware specifications. Both laptops and mobile phones were analyzed using multiple regression techniques.

8.1 Methodology

- **Features (Laptops)**: RAM, CPU cores, threads, display size, resolution.
- **Features (Mobiles)**: RAM, battery, screen size, camera specs, weight.
- Data was standardized using *StandardScaler*.
- Models trained: **Linear Regression**, **Random Forest**, **Gradient Boosting**.
- Evaluation metrics: Mean Absolute Error (MAE) and R^2 Score.

8.2 Model Performance

Table 8: Model Performance Summary

Device	Model	MAE (USD)	R^2 Score
Laptop	Linear Regression	19,542.70	0.78
Laptop	Random Forest	14,719.70	0.82
Laptop	Gradient Boosting	15,121.65	0.82
Mobile	Linear Regression	342.59	-0.02
Mobile	Random Forest	173.52	-2.68
Mobile	Gradient Boosting	128.27	0.69

Gradient Boosting provided the best balance of accuracy and generalization, particularly for laptops. Mobile price prediction proved more challenging due to external factors like brand perception and ecosystem value.

8.3 Feature Importance Analysis

We examined feature importances from the Gradient Boosting models:

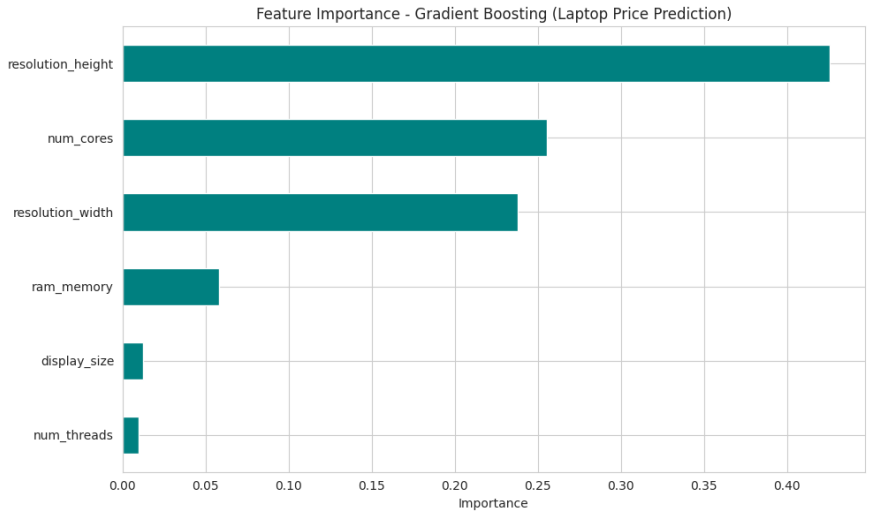


Figure 9: Laptop Price Prediction - Feature Importance

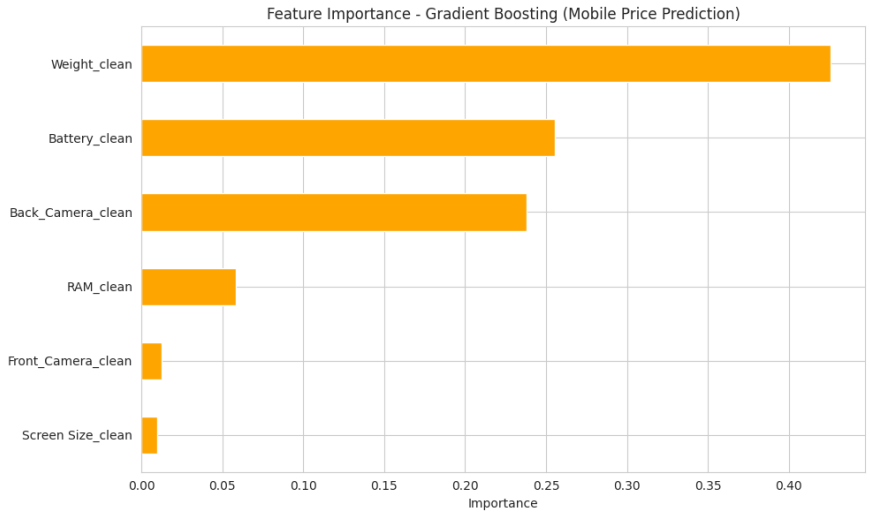


Figure 10: Mobile Price Prediction - Feature Importance

For laptops, CPU-related features and RAM were dominant predictors. For mobiles, camera specifications, battery capacity, and RAM were key drivers of price.

8.3.1 Mobile Model Without Weight

To assess the impact of device weight, we retrained the mobile model excluding the *Weight* feature.

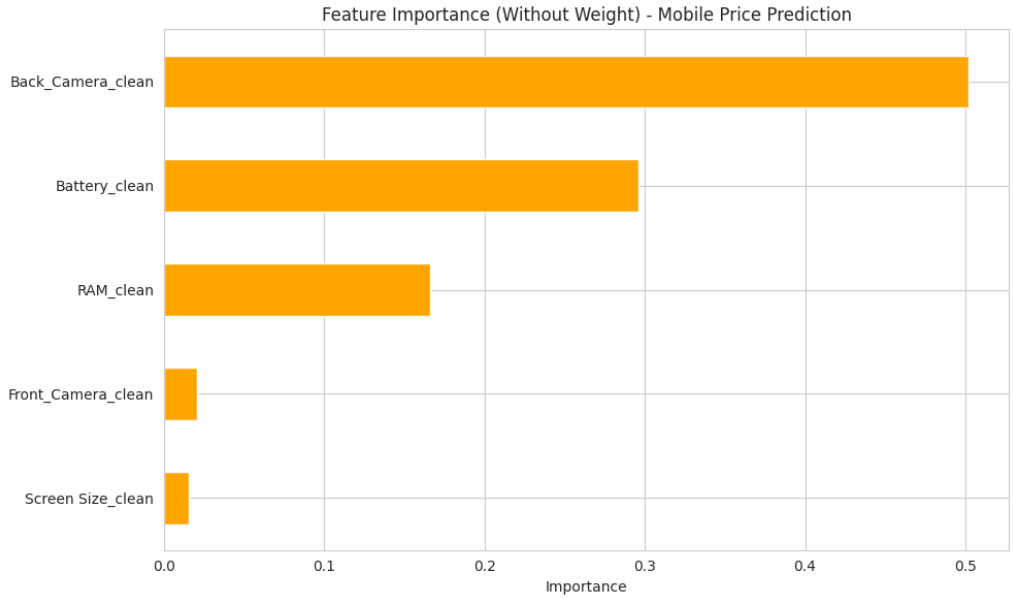


Figure 11: Mobile Feature Importance (Excluding Weight)

After excluding weight, **back camera quality** emerged as the most significant factor, followed by battery capacity and RAM.

8.4 Actual vs Predicted Price Plots

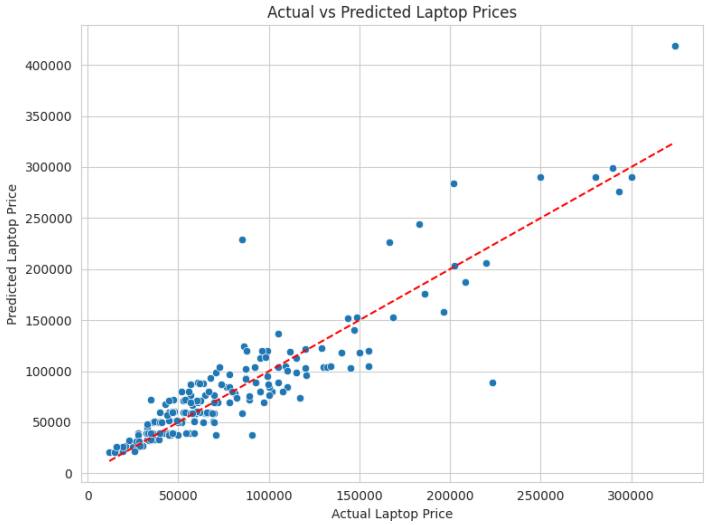


Figure 12: Actual vs Predicted Prices - Laptops

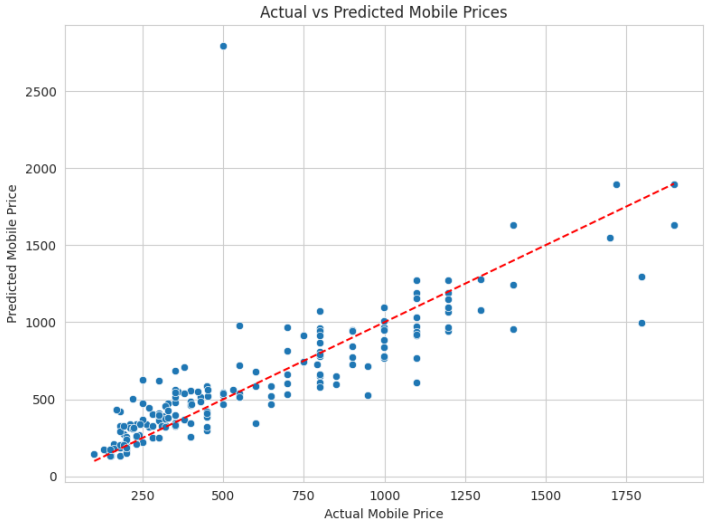


Figure 13: Actual vs Predicted Prices - Mobiles

The laptop model showed strong alignment in mid-range pricing but some variance at premium levels, likely due to brand effects. The mobile model performed reasonably for budget and mid-tier devices, with increased variability for high-end smartphones where non-spec factors dominate.

9 Key Findings

Our comprehensive analysis of laptop and mobile phone datasets led to the following key insights:

1. **Price Prediction Models:**
    - Gradient Boosting and Random Forest models provided strong predictive performance for laptops $R^2 \approx 0.82$.
    - Mobile price prediction was more challenging due to brand influence, but Gradient Boosting achieved a reasonable $\mathbf{R^2}$ of 0.69.
2. **Feature Importance:**
    - **Laptops:** CPU cores, threads, and RAM were the dominant price drivers.
    - **Mobiles:** Camera specifications (especially back camera), battery capacity, and RAM were key factors.
3. **Clustering Analysis:**
    - **Laptops:** Five distinct segments identified, ranging from budget/student models to flagship workstations.
    - **Mobiles:** Clear segmentation into budget, mid-range, camera-focused, rugged, and niche devices.
4. **Brand Value Insights:**
    - Brands like Microsoft and Apple command significant pricing premiums beyond technical specifications.
    - Value-oriented brands such as Infinix and Realme offer better specs-per-dollar ratios.
5. **Moore's Law & Temporal Trends (Mobiles):**
    - Strong statistical evidence shows that mobile device specifications per dollar have consistently improved over time.
    - No conclusive temporal analysis for laptops due to missing release year data.

Overall, while technical specifications largely explain pricing in laptops, mobile phones exhibit greater influence from brand perception and market positioning. Both markets display structured segmentation, but consumer priorities differ across categories.

10 Conclusion

The comparative analysis reveals several key insights:

- Mobile phones show better price prediction accuracy, likely due to more standardized components and more covariates in the data available.
- Camera specifications dominate mobile pricing, while CPU/GPU drive laptop prices
- We find clear evidence of device performance per unit cost increasing with time across multiple features in case of mobiles.
- Brand effects are stronger in the mobile market, particularly for Apple
- Both markets show clear performance tiers, though laptop segmentation is more gradual

These findings suggest that while similar analytical approaches work for both device categories, market dynamics differ significantly. Consumers should prioritize different features when evaluating value across categories.

11 Acknowledgements

- Open-source community for data science tools (Pandas, Scikit-learn, TensorFlow)
- Dataset providers: NotebookCheck, GSMArena, PriceRunner
- Colleagues for valuable feedback on methodology
- Hardware manufacturers for detailed specification documentation

12 References

1. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD '16.
2. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. NIPS.
3. Arthur, D., & Vassilvitskii, S. (2007). k-means++: The Advantages of Careful Seeding. SODA '07.