Smartphone Price Analysis Report

Joshua Ndala

Table of Content

Main Objective

The goal of this analysis was to develop a predictive model for smartphone pricing, with a dual focus on both prediction accuracy and interpretability. This analysis provides valuable insights for:

- Manufacturers: Understanding key price drivers for competitive positioning
- Retailers: Optimizing pricing strategies for different market segments
- Consumers: Understanding value proposition of different phone features
- Market Analysts: Identifying pricing trends and brand positioning effects

Dataset Description

The dataset has 1,512 smartphone models with 22 attributes.

Table 1

Dataset Attribute Information

Category	Attribute	Data Type	Description
Device Identification	phone_name	object	Model name of the smartphone
	brand	object	Manufacturer of the phone
Base Specifications	os	object	Operating system version
	ram(GB)	int64	RAM capacity in gigabytes
	storage(GB)	int64	Storage capacity in gigabytes
Display	inches	float64	Screen size diagonal measurement
	resolution	object	Screen resolution (width x height)
Power	battery	int64	Battery capacity in mAh
	battery_type	object	Type of battery (Li-Po or Li-Ion)
Physical	weight(g)	float64	Weight in grams
Video Capabilities	video_720p	bool	Support for 720p video
	video_1080p	bool	Support for 1080p video
	video_4K	bool	Support for 4K video
	video_8K	bool	Support for 8K video
Frame Rates	video_30fps	bool	Support for 30fps recording
	video_60fps	bool	Support for 60fps recording
	video_120fps	bool	Support for 120fps recording

	video_240fps	bool	Support for 240fps recording
	video_480fps	bool	Support for 480fps recording
	video_960fps	bool	Support for 960fps recording
Time	announcement_date	object	Phone release date
Price	price(USD)	float64	Price in US dollars

The goal of the analysis is to understand price determinants in the smartphone market and create an accurate pricing model.

Data Exploration and Engineering

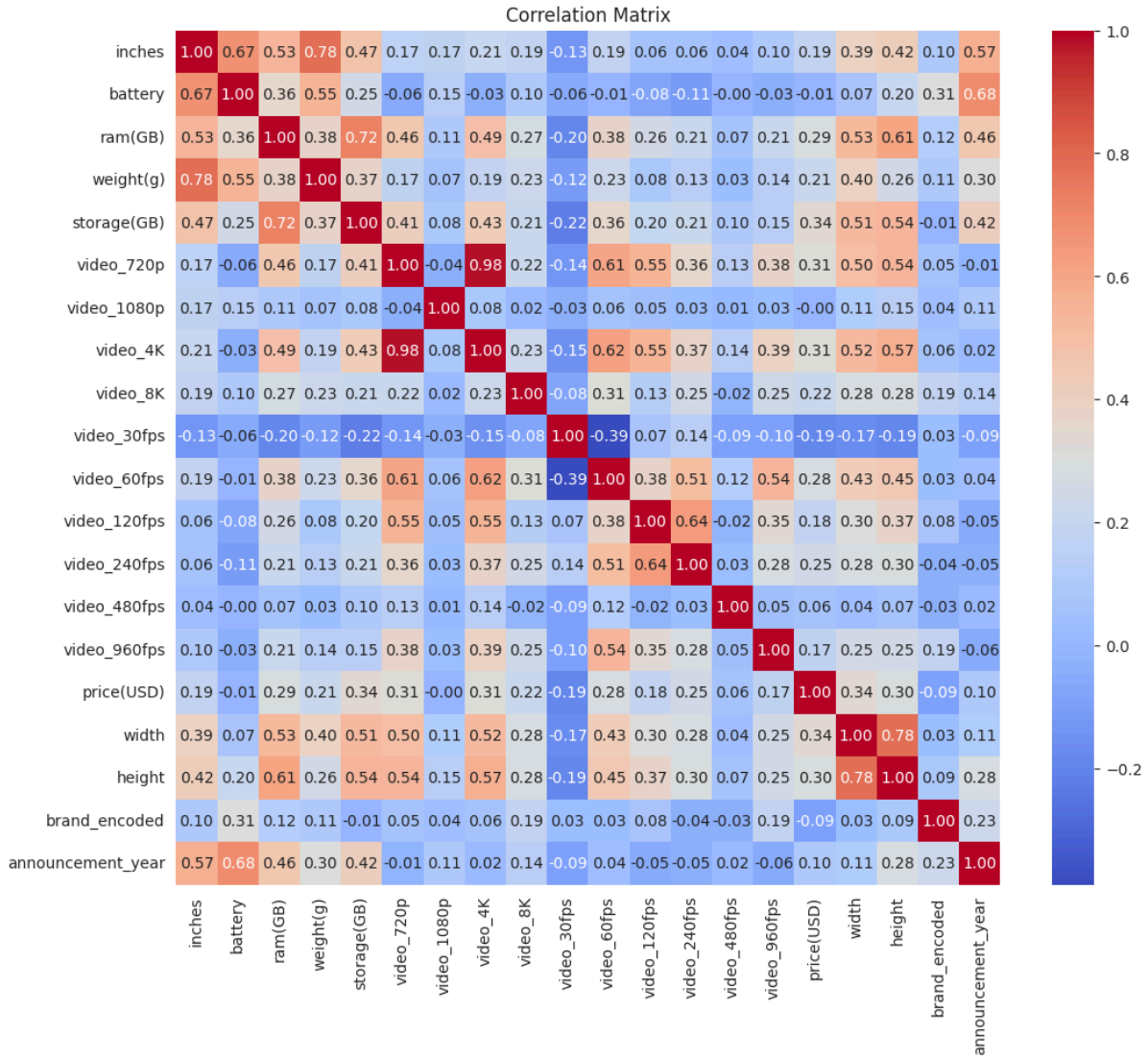
Data Cleaning:

- Standardized storage values (converted 1TB to 1024GB)
- Split resolution into width and height components
- Encoded categorical variables (brands)
- Extracted year from announcement dates
- Removed irrelevant non-numerical columns

Initial correlation analysis showed relatively weak direct relationships between individual features and price, with most correlations below 0.4. However, strong correlations between various technical specifications (e.g., screen dimensions, RAM, and storage) were found. This suggested that while individual features might not strongly predict price alone, their combined effects and interactions could be more informative.

Figure 1

Correlation matrix showing relationships between features



Feature Engineering:

Based on these observations, combined features that could better capture the overall value proposition of each device were developed:

1. Screen Metrics:
    - Screen area calculation (width × height)
    - Pixel density (pixels per square inch)
    - Screen-to-body ratio

2. Performance Scores:
    - Comprehensive video capability score (combining resolution and frame rate capabilities)
    - Memory and storage interactions

    - Device generation score
    - Battery efficiency metrics

3. Brand Positioning:
    - Brand-specific price positioning
    - Price segment categorization

This engineering approach proved effective, as our composite features showed stronger relationships with price than individual specifications. Particularly, the brand positioning features emerged as crucial predictors, suggesting that market positioning plays a more significant role in pricing than raw technical specifications.

Model Training and Evaluation

Three different models were trained and evaluated using a 80-20 train-test split with 5-fold cross-validation:

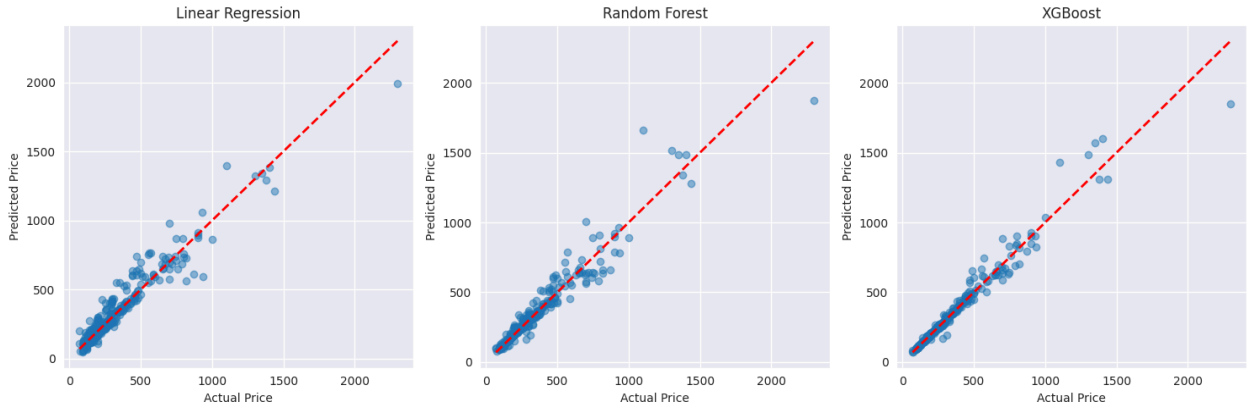
1. Linear Regression (Baseline):
   - R^2 Score: 0.925 ± 0.006
   - RMSE: $75.99
   - MAE: $47.69
   - Highly interpretable but less accurate

2. Random Forest:
   - R^2 Score: 0.937 ± 0.013
   - RMSE: $68.84
   - MAE: $34.72
   - Better accuracy with moderate interpretability

3. XGBoost:
   - R^2 Score: 0.961 ± 0.012
   - RMSE: $51.30
   - MAE: $21.80
   - Best accuracy but least interpretable

Figure 2

Model performance comparison showing predicted vs actual prices for Linear Regression, Random Forest, and XGBoost



Model Recommendation

XGBoost is recommended as the primary predictive model due to its superior performance (96.1% accuracy). After feature optimization, the model maintained its performance while using fewer features, demonstrating its robustness. For situations requiring more interpretability, the Random Forest model provides a good balance between accuracy (94.2%) and explainability.

Key Findings and Insights

1. Brand Positioning is Crucial:
    - Brand positioning relative to market average is the strongest price predictor
    - This suggests pricing is more influenced by brand strategy than raw specifications

Figure 3

Average smartphone prices by brand, showing significant price variation across manufacturers

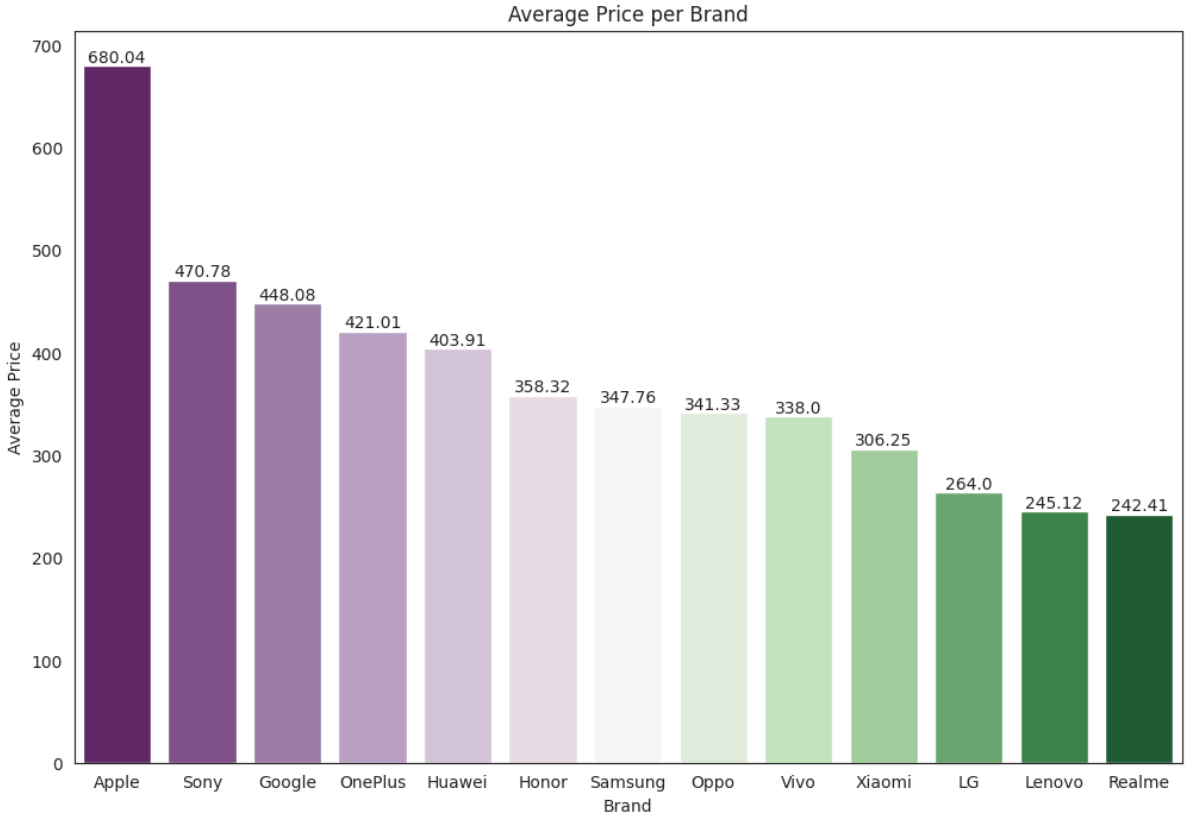
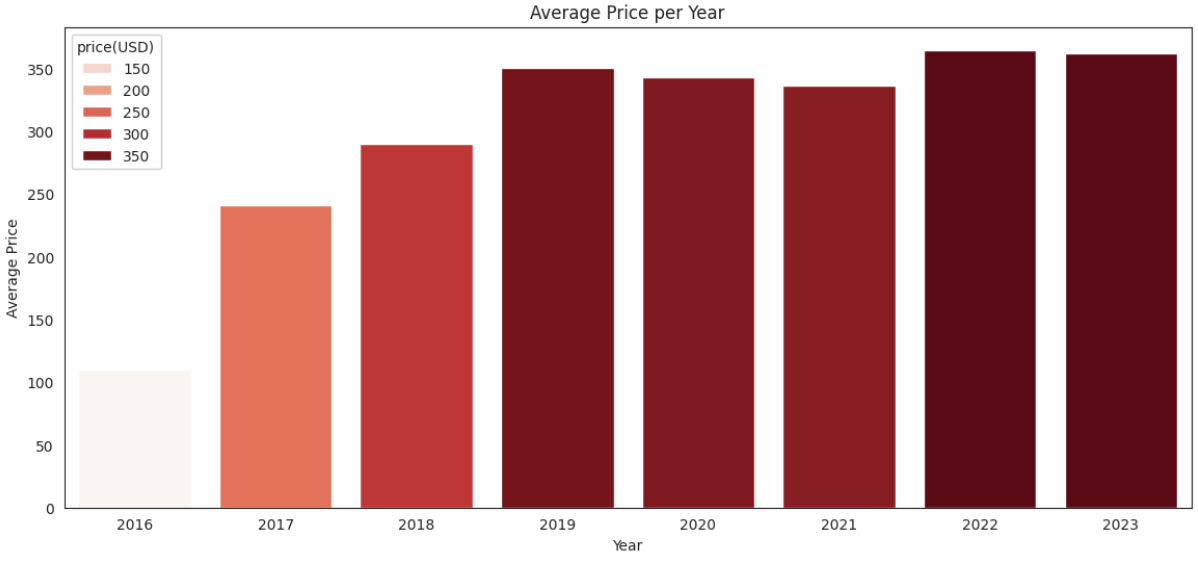


Figure 4

Average price trends over time, demonstrating market evolution

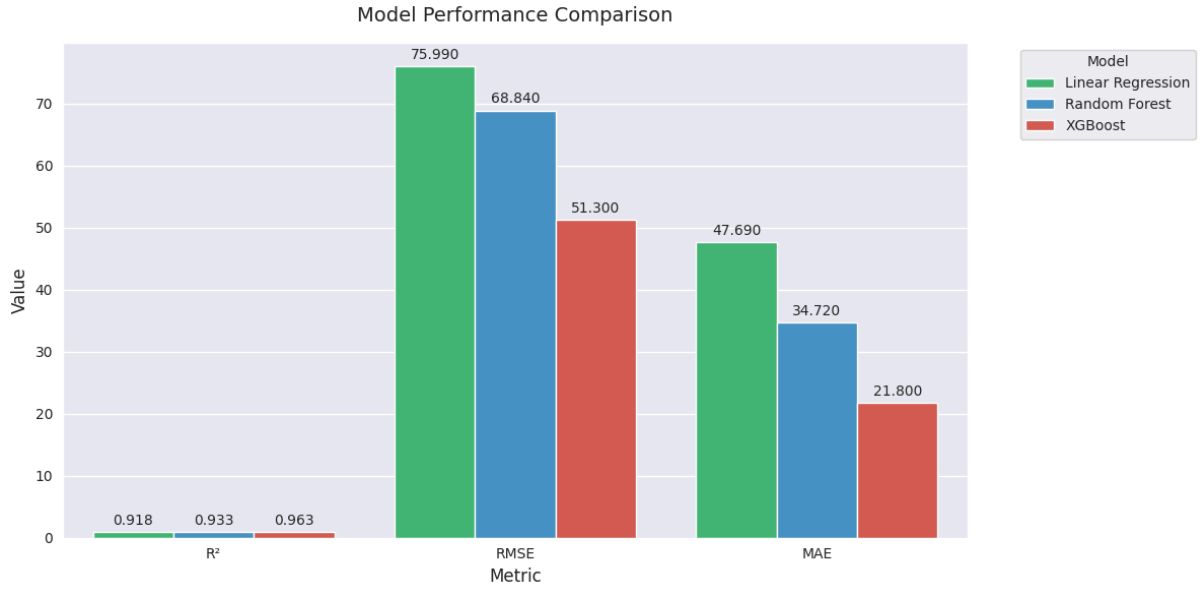


2. Feature Importance:
    - Most important features were brand-related and screen metrics
    - Traditional specs (RAM, storage) had less impact than expected
    - Battery and weight relationships showed minimal individual impact

3. Model Performance:
    - All models performed well ($R^2 > 0.92$)
    - Simplified feature set maintained or improved performance
    - Price predictions are most accurate in mid-range segment

Figure

Performance metrics comparison across models



4. Market Dynamics:
    - Strong correlation between brand positioning and price
    - Technical specifications have less direct impact on price than brand value
    - Premium segment shows more price variation than budget segment

Next Steps and Recommendations

Data Enhancement

- Include software features and OS version information
- Add camera quality metrics beyond video capabilities
- Collect data on market share and sales volumes
- Include regional pricing variations

Model Improvements

- Develop separate models for different price segments
- Implement time-series analysis for price trends
- Create brand-specific sub-models
- Include competitor pricing information

Business Applications

- Develop dynamic pricing recommendations
- Create market positioning analysis tools
- Build competitive analysis framework
- Design automated price optimization system

Further Research

- Investigate regional pricing differences
- Analyze the impact of technology generations
- Study the relationship between features and customer satisfaction
- Examine price depreciation over time