ReCell – Dynamic Pricing for Used Devices

Esther Joseph

Contents / Agenda

Executive Summary

The used device market is projected to reach Σ52.7 billion by 2023, growing at a 13.6% CAGR.

OBJECTIVE: Develop a **linear regression model** to predict used device prices and identify key factors influencing them.

KEY FINDINGS:

- The price of used devices is significantly influenced by **camera quality (main and selfie megapixels)**, **RAM**, **weight**, and **it's new price**.
- **4G capability increases resale value**, while **5G capability currently decreases it**, likely due to market adoption.
- **Devices from brands like Nokia, Xiaomi, and Asus** have a positive impact on resale prices, making them more profitable in second-hand markets.
- **Devices with higher "years since release" show a predictable decline in price**, highlighting the importance of device age in pricing models.

ACTIONABLE INSIGHTS & RECOMMENDATIONS:

- Optimize pricing strategies for devices with **superior camera features**, **higher RAM**, and **lighter weight**, as these factors drive value.
- Promote 4G-capable devices to maximize revenue; monitor 5G market trends for future pricing strategy adjustments.
- Focus on **sourcing devices from high-profit brands** (e.g., Xiaomi, Nokia) to improve profit margins.
- Implement a dynamic pricing strategy that accounts for **device age** and depreciation curves to remain competitive.

Business Problem Overview and Solution Approach

ReCell, needs a dynamic pricing strategy to accurately price used devices and remain competitive in the market.

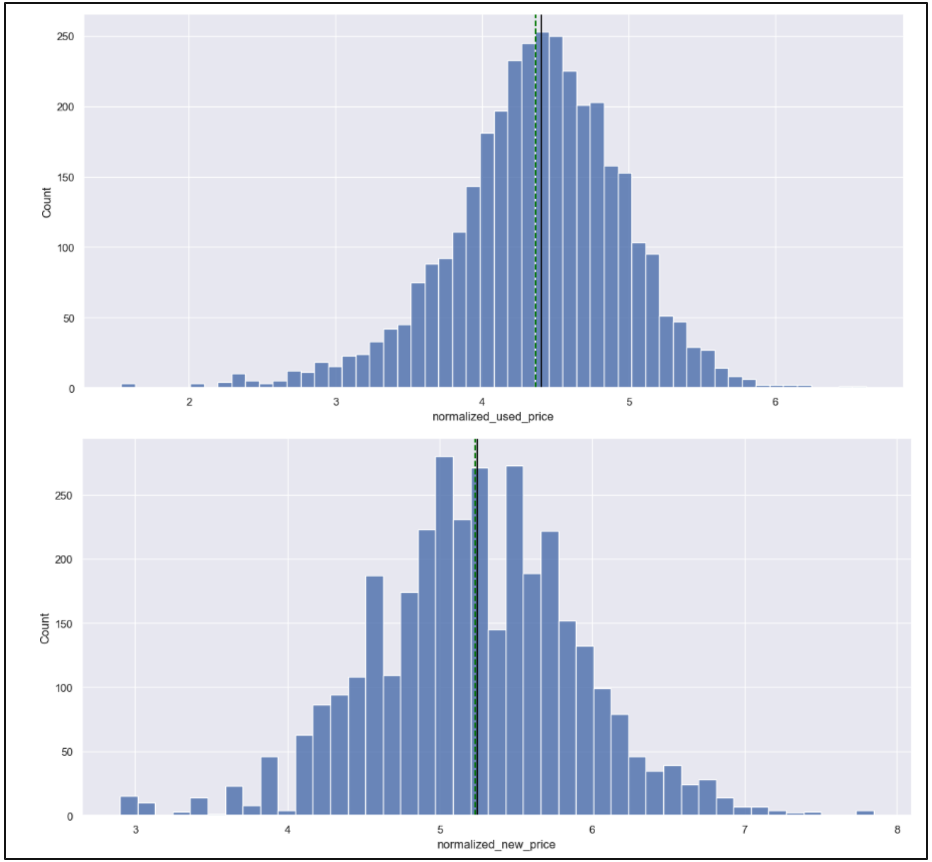
CHALLENGE: Identifying the factors influencing the resale price of devices to set competitive yet profitable pricing.

GOAL: Build a predictive model to estimate the normalized price of a used devices, identifying factors that significantly influence its value.

SOLUTION APPROACH

1) **Data Understanding & EDA**: Analyze the dataset to uncover key patterns and correlations.
2) **Data Preprocessing**: Handle missing values, create new features, and address outliers.
3) **Modeling**: Use **Linear Regression** with regular checks for multicollinearity and assumption violations.
4) **Model Evaluation**: Assess goodness-of-fit (R^2, Adj. R^2) and error metrics (RMSE, MAE, MAPE) on train/test splits.
5) **Insights & Recommendations**: Translate model findings into actionable strategies for ReCell's pricing system.

EDA Results (Univariate Highlights - Price Distributions)



Most used device prices are concentrated **between 4 and 5,** while new device prices cluster **between 5 and 6**.

Insights from Price Distributions:

- There’s a clear price gap between new and used devices, with used devices predominantly priced around **20-25% lower than new devices** in normalized terms.
- This emphasizes the cost-effectiveness of refurbished devices, aligning with ReCell's target market of budget-conscious buyers.

EDA Results (Univariate Highlights - Screen Size & Camera (Main and Selfie))

Observation:

Most devices feature **screen sizes between 13-15 inches**, with the main camera resolution peaking at **10 MP** and selfie cameras at **5 MP**.

Insight:

- The predominant screen size range suggests that users favor devices that strike an optimal balance between portability and visual clarity, ideal for everyday tasks.
- The camera specifications indicate a focus on functional photography rather than high-end capabilities. This aligns with the typical requirements of the used phone market where affordability and basic functionality are prioritized over advanced features.

Market Implications

These insights reflect a market trend where cost-effectiveness and essential functionality are key selling points, making devices with these characteristics more appealing to the used phone buyers who prioritize value over luxury.

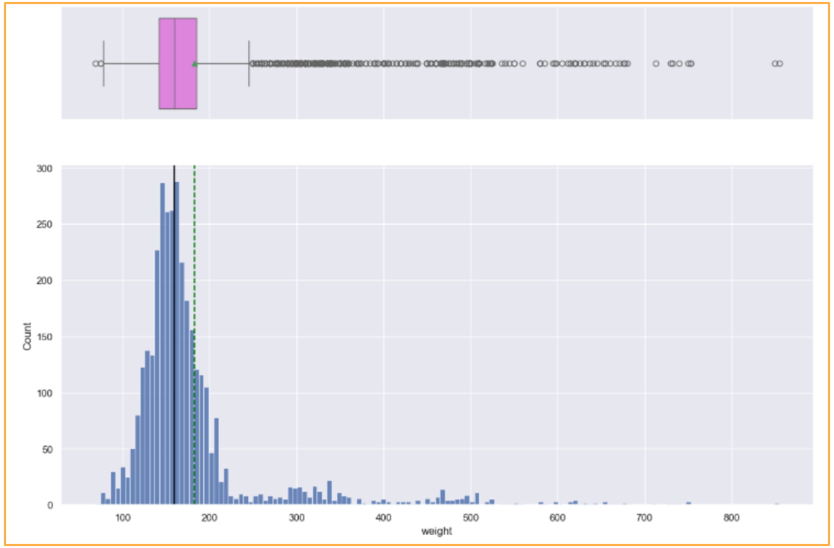
EDA Results (Univariate Highlights – Internal Memory, RAM & Weight

Observation:

Peaks are evident at **54 GB** for internal memory, **4 GB** for RAM, and **150-200 grams** for weight.

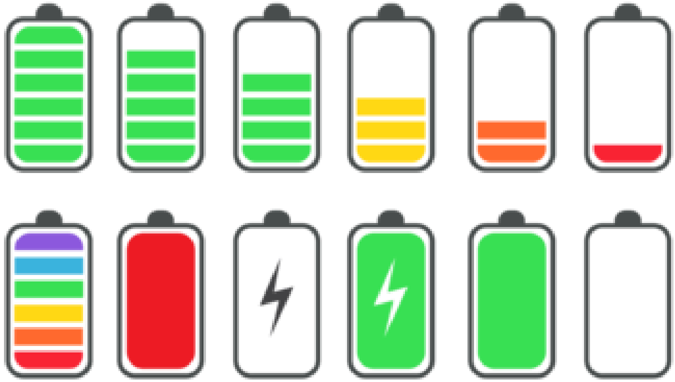
Insight and Market Implications:

The dominance of moderate specifications such as **54 GB internal memory** and **4 GB RAM** suggests that consumers of used devices prefer practicality and affordability over high-end features. The preferred weight range supports a demand for portable yet durable devices. This trend indicates an opportunity for ReCell to optimize its pricing strategy by stocking devices that meet these common criteria, effectively addressing consumer preferences and enhancing market competitiveness.



EDA Results (Univariate Highlights - Battery and Days Used)

Devices primarily feature batteries in the **2000-4000 mAh range**, which typically offer **5-10hours** from research. This range suits the needs of most users for daily tasks, providing a reliable gauge of device usability in a refurbished state.

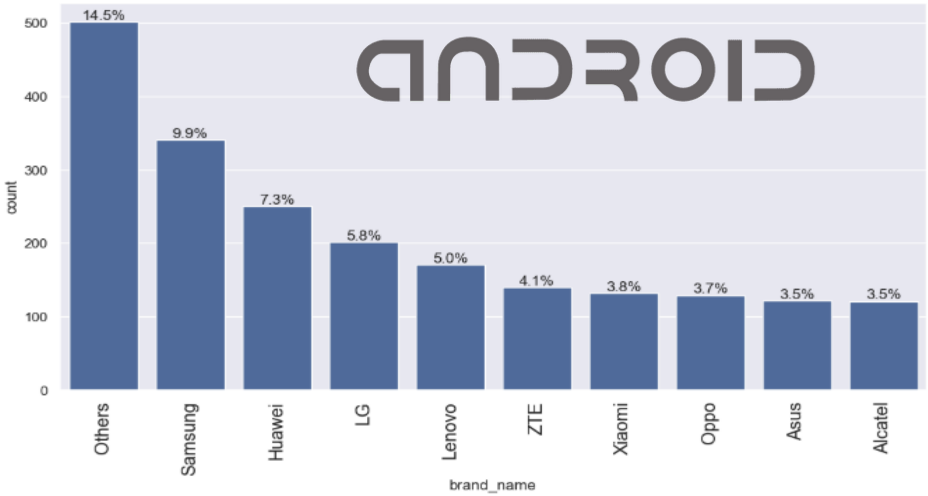


The spread in the number of days used indicates a market with varied device life spans, suggesting that many consumers are open to using devices for extended periods, provided the battery life and device performance are adequate.

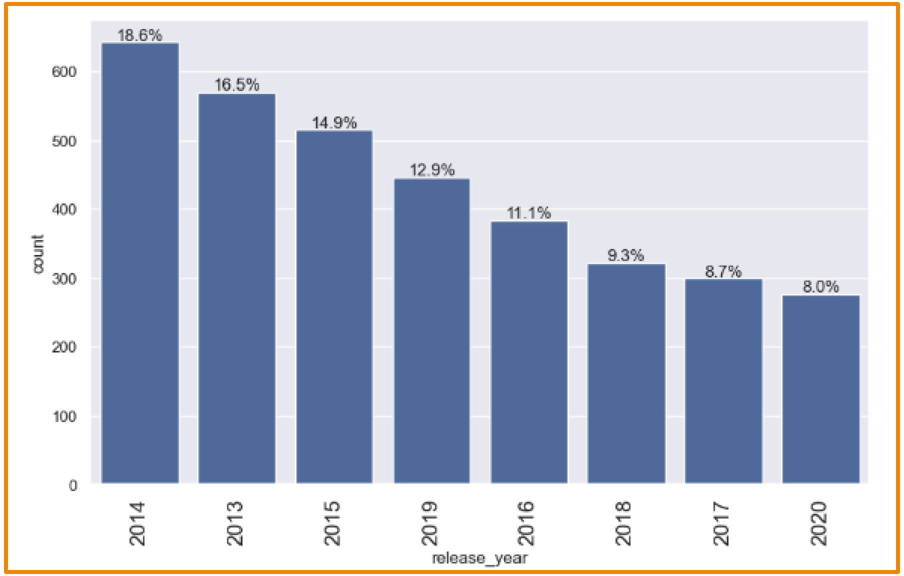
EDA Results (Univariate Highlights – Brand Name, OS)

The majority of devices are from brands categorized as "Others", followed by Samsung and Huawei.
Most devices operate on Android OS.

This suggests that consumers purchasing used devices might be brand-agnostic, focusing more on other factors. And the dominance of Android in the used device market could allow ReCell to emphasize the flexibility and customization options available with these devices, appealing to consumers looking versatile technology.



EDA Results (Univariate Highlights – Connectivity & Release Year)



Devices release years from 2013 to 2020, with a slight peak in 2014 and a gradual decline in newer years.
While 67.6% of devices support 4G, indicating strong adoption of this standard, only 4.4% of devices support 5G, highlighting its limited presence in the used device market.

Insight:

The gradual decline in newer devices suggests it might take several years for phones to transition from the primary to the secondary market.

The low presence of 5G highlights a significant future growth opportunity as more 5G-enabled devices eventually reach the used market.

EDA Results (Bivariate Analysis: Part 1)

- Larger screen sizes contribute to higher device weight. High positive correlation (~0.83).
- Resale value closely tied to the original retail price. Strong correlation (~0.83)
- Higher RAM configurations drive slightly better resale value. Moderate correlation (~0.52).
- Used prices for newer devices (e.g., 2020 models) are significantly higher than older devices.

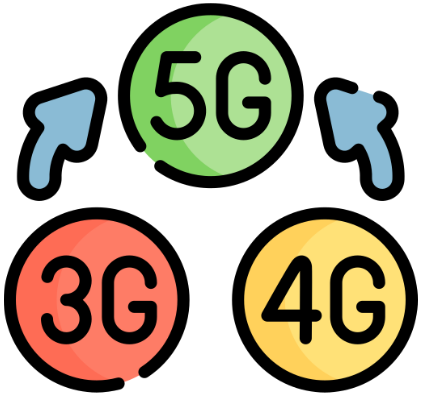
<u>**Business Recommendations**</u>

- Use screen size and weight as key differentiators, especially for lightweight models.
- Create marketing campaigns emphasizing the newer devices as recent model to maximize profitability.
- Highlight high-RAM models in advertising to emphasize performance.

EDA Results (Bivariate Analysis: Part 2)

Key Findings & Insights

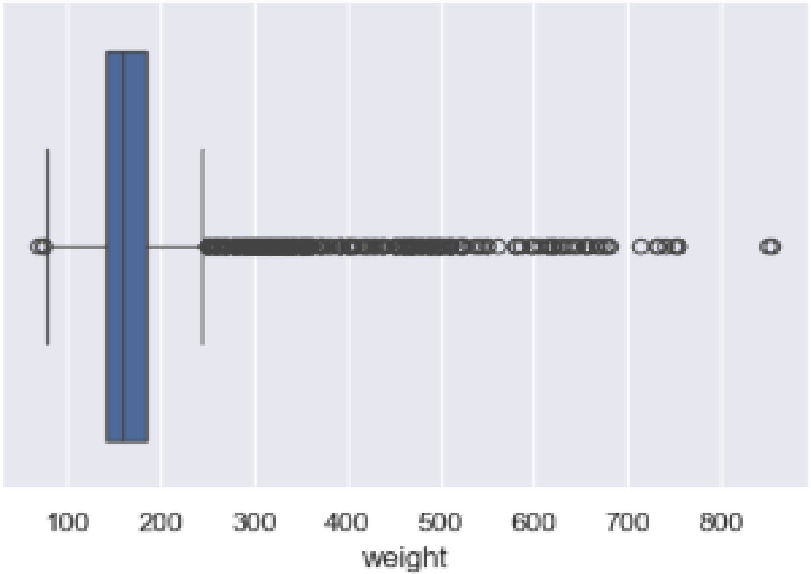
- Devices with 4G or 5G connectivity are priced significantly higher in the resale market.
- Camera resolution positively impacts resale prices. **Sony** dominates with ~39.4% of devices offering rear cameras above 16MP. Brands like Motorola and Samsung also have strong camera specifications.
- **Huawei** and **LG** devices are heavier, likely due to larger screens and batteries.
- Lightweight brands like **Xiaomi** and **Vivo** cater to portability-focused buyers.



Business Recommendations

- **Capitalize on Connectivity Trends**: Invest in 5G-compatible devices to attract future-forward buyers.
- Provide upgrade incentives for older models to 4G/5G.
- **Leverage Camera Features**: Highlight premium photography capabilities of brands like Sony and Motorola.
- Use targeted ads for camera-centric buyers.
- **Innovate on Design**: Introduce lightweight options for high-performing devices to expand the appeal of larger models.

Data Preprocessing - Overview



- **No duplicates** were found in the dataset, ensuring data integrity and preventing model bias.
- Missing values in critical columns (e.g., main_camera_mp, selfie_camera_mp, ram) were handled through hierarchical imputation using medians grouped by release_year and brand_name.
- Final imputation for residual missing values was done using the column median, ensuring consistent and reliable data.
- **Outliers:** Key variables like screen_size, battery, and `weight showed noticeable outliers.
- Outliers Action Taken: Outliers were identified but retained as they may represent premium devices or niche categories. These may impact pricing strategies but provide valuable insights into high-end products.

Data Preprocessing - Feature Engineering & Data Preparation for Modeling

Feature Engineering:

- years_since_release created from release_year — the years_since_release feature is critical for price depreciation models, helping to set accurate expectations for used device prices.

- Dropped the original release_year column to avoid redundancy.

Data Preparation:

- Categorical features (e.g., brand_name, os, 4g, 5g) were converted into dummy variables for modeling.
- Split the dataset into 70% training and 30% test data for model evaluation, ensuring the model is tested on unseen data.

Model Performance Summary

Model Overview: The model used is **Linear Regression (OLS)**, trained to predict **normalized used prices** of phones and tablets based on various features.

Most Important Factors for Prediction:

- **Normalized new price** has a significant impact on predicting the used price, as indicated by its large positive coefficient.
- **Main camera megapixels** and **selfie camera megapixels** are also important predictors, with both showing significant positive correlations with used price.
- **Years since release** is negatively correlated with the used price, as expected (the older the device, the lower the price).
- **Brand name** plays a notable role, with **Xiaomi**, **Nokia**, **Asus**, and **Celkon** being key variables for higher or lower used prices.

Actionable Insight: Understanding the key features influencing the used price can help businesses prioritize or market phones based on their higher resale value. This can inform decisions around stock rotation, trade-in programs, and pricing strategies.

Model Performance Summary (Key Performance Metrics)

Model Evaluation:

- The **R-squared** value for the model is **0.843** for the training data and **0.829** for the test data, suggesting that the model explains a substantial amount of variance in the used price.

- The **Adjusted R-squared** for both training and test data is close to the R-squared, indicating that the inclusion of additional variables does not significantly harm the model's generalizability.

- The **RMSE (Root Mean Squared Error)** is **0.2297** for training and **0.2391** for test data, demonstrating that the model is relatively accurate.

- The **MAPE (Mean Absolute Percentage Error)** is around **4.3%**, showing that the model's predictions are on average within 4.3% of the true values.

THESE PERFORMANCE METRICS SUGGEST THAT THE MODEL IS HIGHLY RELIABLE AND CAN BE USED FOR ACCURATE PRICING PREDICTIONS OF USED DEVICES.

APPENDIX

Data Background and Contents

Data Description

- The data contains the different attributes of used/refurbished phones and tablets. The data was collected in the year 2021. The detailed data dictionary is given below.

Data Dictionary

- brand_name: Name of manufacturing brand
- os: OS on which the device runs
- screen_size: Size of the screen in cm
- 4g: Whether 4G is available or not
- 5g: Whether 5G is available or not
- main_camera_mp: Resolution of the rear camera in megapixels
- selfie_camera_mp: Resolution of the front camera in megapixels
- int_memory: Amount of internal memory (ROM) in GB
- ram: Amount of RAM in GB
- battery: Energy capacity of the device battery in mAh
- weight: Weight of the device in grams
- release_year: Year when the device model was released
- days_used: Number of days the used/refurbished device has been used
- normalized_new_price: Normalized price of a new device of the same model in euros
- normalized_used_price: Normalized price of the used/refurbished device in euros

Model Assumptions

Multicollinearity Check (VIF): Variance Inflation Factor (VIF) was calculated for all features to check for multicollinearity. Variables such as **brand_name_Others**, **screen_size**, and **brand_name_Samsung** had high VIFs (above 5), indicating multicollinearity. After dropping variables with high VIF, the model improved without losing performance, with all VIFs dropping below 5.

Linearity and Independence of Errors: The **residuals vs fitted plot** showed no significant patterns, confirming that the relationship between predictors and the response is linear.

The **Durbin-Watson statistic** was close to 2, indicating no major issues with autocorrelation, which supports the independence of errors.

Normality of Residuals:
The **Q-Q plot** and **Shapiro-Wilk test** confirmed that the residuals are approximately normally distributed, which is an important assumption for linear regression.

The **Shapiro-Wilk test p-value** was very small (close to zero), indicating non-normality in the residuals.

Homoscedasticity Check:

The **Goldfeld-Quandt test** showed a p-value of **0.812**, indicating that the residuals are homoscedastic (i.e., have constant variance).

Technical Data Analysis & Linear Modeling

- For a more detailed view of the data analysis, feature engineering, and linear regression modeling, please refer to the full Jupyter notebook.

- Click here to view the Jupyter notebook

Thank You!