ReCell

Richard Chukwu

12/08/2024

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Appendix

Executive Summary

- The used and refurbished device market has grown significantly over the past decade. An IDC forecast predicts it will be worth $52.7 billion by 2023, with a compound annual growth rate(CAGR) of 13.6% from 2018 to 2023. This growth is driven by the demand for cost-effective alternatives to new devices, offering significant savings. The market is also boosted by the availability of warranties, insurance , options, and environmental benefits from reduced waste and recycling. The COVID- 19 d pandemic may further enhance this trend as consumers limit discretionary spending.

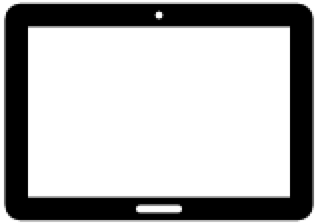
INSIGHTS AND RECOMMENDATIONS

Focus on High-Value Features: Manufacturers and sellers should emphasize and market devices with superior camera capabilities, larger memory capacities, and higher RAM to maximize resale potential.

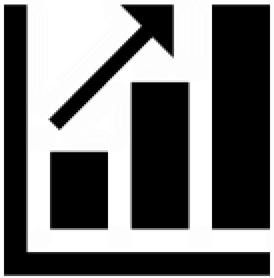
Monitor Market Trends: Stay informed about consumer preferences and market trends, especially regarding device specifications and release years, to optimize inventory and pricing strategies

Continuous Model Evaluation: Regularly evaluate and update the Linear Regression model to incorporate new data and adjust predictions based on evolving market dynamics.

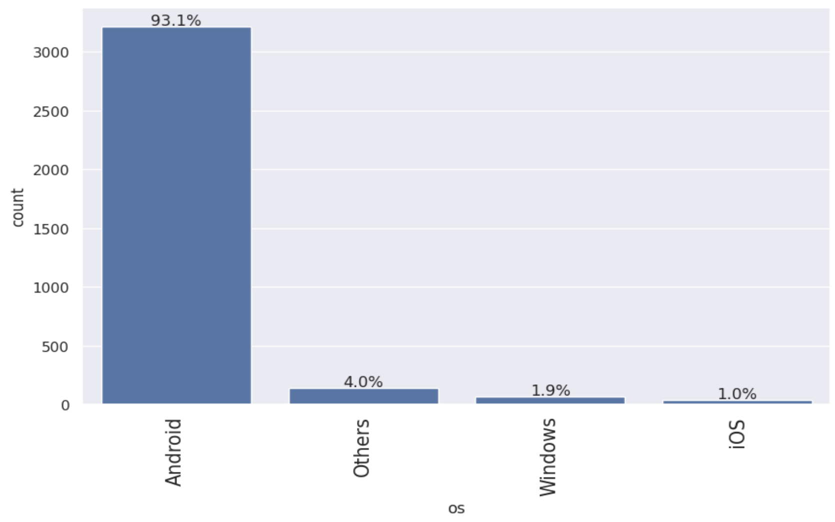
Business Problem Overview and Solution Approach

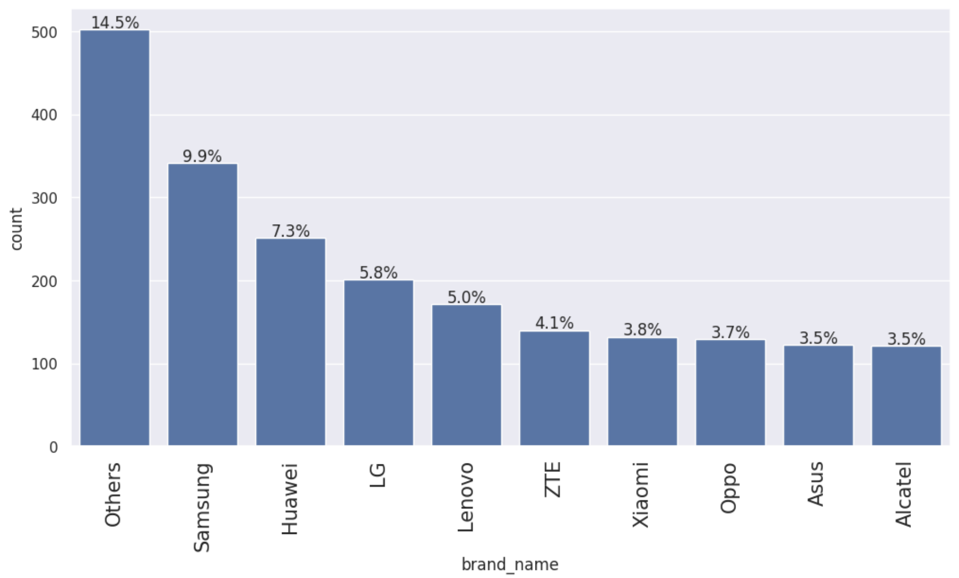


Develop a machine learning-based solution to create a dynamic pricing strategy for used and refurbished phones and tablets.



Build a linear regression model to predict the price of these devices and identify the key factors influencing their prices.





- Samsung,Huawei, and LG exhibit the highest usage percentages among all brands, indicating their strong market presence in the dataset. Conversely, brands like Asus and Alcatel show comparatively lower usage rates. Understanding these brand preferences can provide valuable insights into market dynamics and consumer preferences within the used and refurbished device sector.

- Among users, 93.1% prefer Android, followed by 40.% using other operating systems, with iOS being the least popular at 1.0%. The distribution underscores the widespread adoption of Android compared to other operating systems in the dataset.

EDA

Univariate Analysis

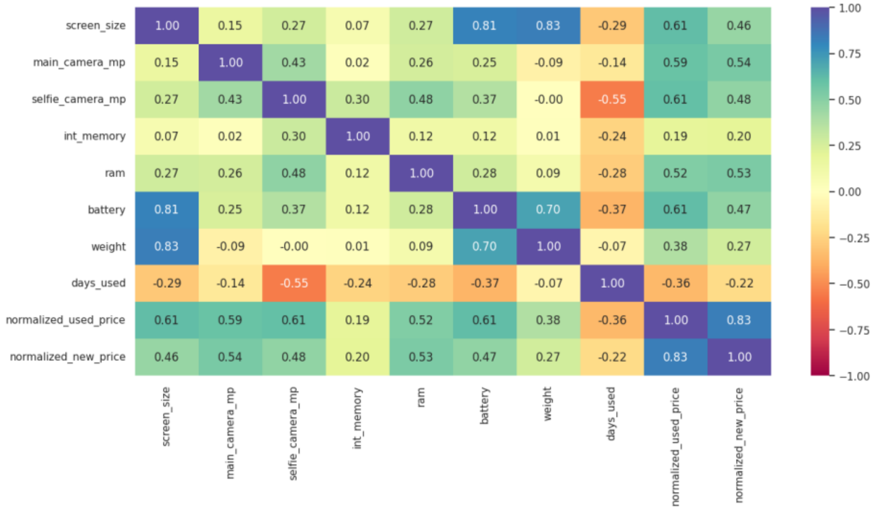
The normalized price of used devices is lower compared to the normalized price of new devices of the same models.

Most of the used phones were released in 2014, followed by 2013, 2015, 2019, 2018, 2017, and 2020 in descending order of frequency.

EDA

Bivariate Analysis

- Larger screens increase phone weight but don't change significantly with usage. Higher camera resolutions and RAM size strongly influence phone prices. Internal memory links to selfie camera quality but not usage time. Battery capacity correlates with screen size. Phone weight and condition are minimally affected by usage days. Used prices closely follow new prices, indicating consistent depreciation. Initial prices are influenced more by specifications than age.



Data Preprocessing

- No duplicated value
- We addressed missing values by first imputing column medians grouped by release_year and brand_name, then by brand_name alone, and finally using the overall column median for any remaining missing values. This method ensured accurate and complete data across all features

- Duplicate value check
- Missing value treatment

Data Preprocessing

- We created a new column, years_since_release, by subtracting release_year from the baseline year, 2021. The original release_year column was dropped to avoid redundancy, as years_since_release provides the same information in a more useful format for analysis.
- To predict the normalized price of used devices, we first defined the dependent variable (y) as normalized_used_price and the independent variables (X) excluding it. After adding an intercept to the data, we created dummy variables for categorical features to encode them properly. We then split the dataset into training (70%) and testing (30%) sets to build and evaluate a Linear Regression model. This preparation ensures the model is trained and evaluated effectively, leveraging all relevant features and properly handling categorical data.

- Outlier check (treatment if needed)
- Feature engineering
- Data preparation for modeling

Model Performance Summary

- **Overview of ML Model and Parameters:**

  We built a Linear Regression model to predict the normalized price of used devices. The model was trained using Ordinary Least Squares (OLS) method.

- **Most Important Factors:**

  The normalized_used_price was the most significant predictor, with a strong positive relationship to the normalized_new_price.

Model Performance Summary

Key Performance Metrics:

Metric	Train Data	Test Data
RMSE	0.375626	0.379049
MAE	0.296987	0.294222
R-squared	0.695109	0.699165
Adj. R-squared	0.694857	0.698583
MAPE	5.732845%	5.715602%

The model shows strong predictive performance with similar metrics on both train and test sets.
High R-squared values (around 0.695) indicate a good fit.
Low RMSE and MAE values demonstrate accurate predictions.
The MAPE values, both below 6%, suggest that the model is reliable for practical use.

APPENDIX

Data Background and Contents

The dataset, collected in 2021, focuses on used and refurbished phones and tablets. It includes various attributes such as brand name, operating system, screen size, camera specifications, memory, battery capacity, weight, release year, days used, and normalized prices. The goal is to analyze factors influencing device prices and build predictive models for the normalized used price based on these features.

Great Learning

Happy Learning !