Predictive Pricing Strategy for Refurbished Devices

Supervised Learning - Foundations Project: ReCell

January 20, 2024

Contents / Agenda

Executive Summary

- In our project, we leveraged machine learning to create a dynamic pricing model for used and refurbished electronic devices, a crucial tool for staying competitive in a rapidly evolving market. The model, centered around a linear regression framework, analyzed various factors such as brand, camera quality, memory size, and device age, to predict resale values. This approach offers ReCell a data-driven strategy to optimize pricing, manage inventory effectively, and align with market trends, thereby enhancing profitability and market position.

  - **Insight 1:** The used phone market is growing rapidly, with an expected value of $52.7bn by 2023. ReCell can capitalize on this by optimizing its pricing strategy.

  - **Recommendation 1:** Implement a dynamic pricing model based on the ML model insights to stay competitive.

  - **Insight 2:** Key features like brand, camera quality, memory, and age significantly impact used device pricing.

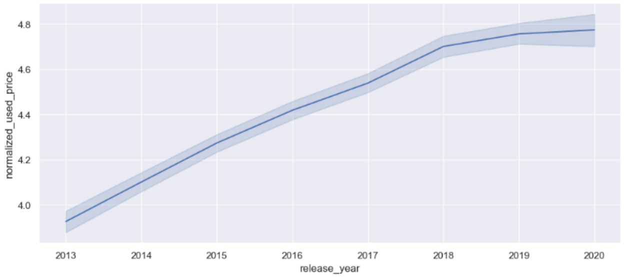
  - **Recommendation 2:** Focus marketing and stock on devices with high-demand features and consider offering promotions on less popular features to boost sales.

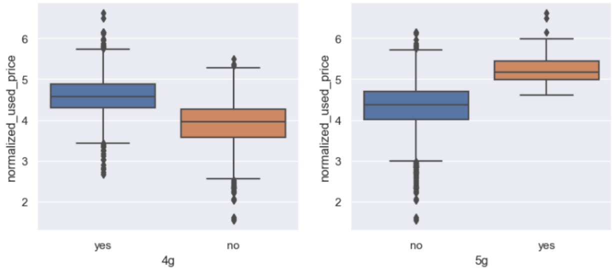
Business Problem Overview and Solution Approach

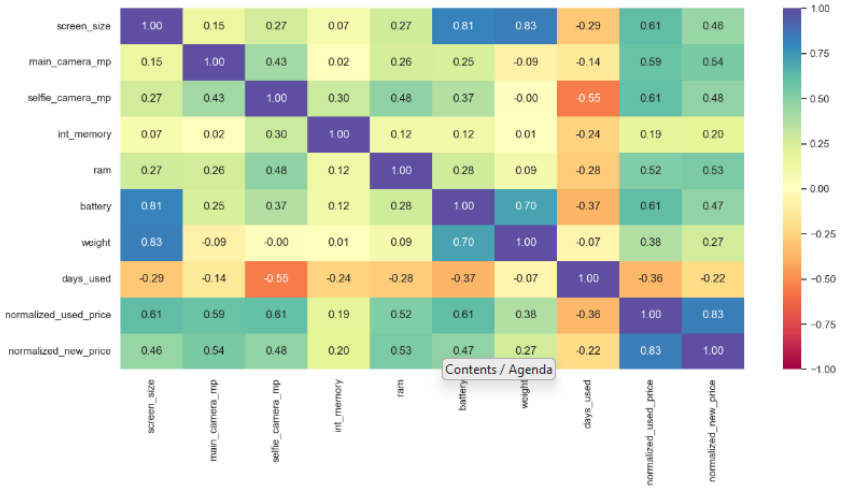
- Problem Statement:
    - The task is to develop a dynamic pricing strategy for used and refurbished devices in the growing market. The goal is to build a linear regression model to predict the price of a used phone or tablet and identify the factors that significantly influence the pricing.
- Solution Approach/Methodology:
    - The approach involves analyzing the provided data to understand the factors affecting the pricing of used devices. Feature selection and engineering will be performed to identify significant predictors. A linear regression model will be built, and its performance will be evaluated to provide insights for ReCell's dynamic pricing strategy in the used device market.

EDA Results

- Identified strong correlations between device features like camera quality, memory, age, and their resale value.
  - The price of used phones steadily increased from 2013 to 2020.
  - Phones that offer either 4g or 5g typically resell for more compared to phones without it.
  - Features that have the strongest correlation with used phone prices are screen size, main and selfie camera megapixels, battery life, and the original price of the phone when new.







[Link to Appendix slide on data background check](#)

EDA Results

- Observed trends indicating certain brands and newer models retain value better.
    - OnePlus phones typically have the most ram, while Nokia phones provide the least.
    - Apple phones offer the largest battery.
    - Huawei and Samsung makes phones with the largest screens compared to its competitors.
    - Huawei makes the most phones with a selfie camera capable of shooting >8 megapixels, with Vivo and Oppo following closely behind.
    - Sony leads the competition by a large margin in producing phones with a main camera that can shoot >16 megapixels.

Data Preprocessing

- Duplicate value check
  - There were no duplicate values.
- Missing value treatment
  - Imputed missing values by column medians grouped by release year and brand name.
  - Remaining missing values were imputed by column medians grouped by brand name.
  - The last of the missing values in “main_camera_ep” were filled by the column median.
- Outlier check (treatment if needed)
  - Detected outliers using boxplot
  - Columns weight, battery, normalized used price, and normalized new price appear to have the greatest number of outliers.

Data Preprocessing

- Feature engineering
  - Created a new column “years_since_release” from the “release_year” column
  - Release year is subtracted from the baseline year 2021
  - The original “release_year” column was dropped
- Data preparation for modeling
  - The dependent variable was set to the normalized price of used devices
  - Categorical features were encoded using dummy variables
  - The data was then split into train and test sets with a ratio of 70/30

Model Performance Summary

- Overview of ML model and its parameters
  - **Model Type:** Linear Regression
  - **Key Parameters:** Regularization (L1/L2), learning rate, and number of iterations
  - **Model Objective:** To predict the price of used/refurbished devices based on various features

- Summary of most important factors used by the ML model for prediction
  - Brand name
  - Camera resolution (main and selfie)
  - Internal memory and RAM
  - Battery capacity
  - Device age (days used)

Link to Appendix slide on model assumptions

Model Performance Summary

- Summary of key performance metrics for training and test data in tabular format for comparison

	RSME	MAE	R-Squared	Adj. R-Squared	MAPE
OLS Model 1 (Train)	0.229761	0.178533	0.848887	0.845785	4.293664
OLS Model 2 (Train)	0.231114	0.179533	0.847102	0.846082	4.322087

	RSME	MAE	R-Squared	Adj. R-Squared	MAPE
OLS Model 1 (Test)	0.239062	0.188692	0.832176	0.823844	4.513288
OLS Model 2 (Test)	0.238225	0.188703	0.833349	0.830734	4.513062

After comparing both models, we can see Model 2 performs slightly better than Model 1 after removing multicollinearity and high p-values.

APPENDIX

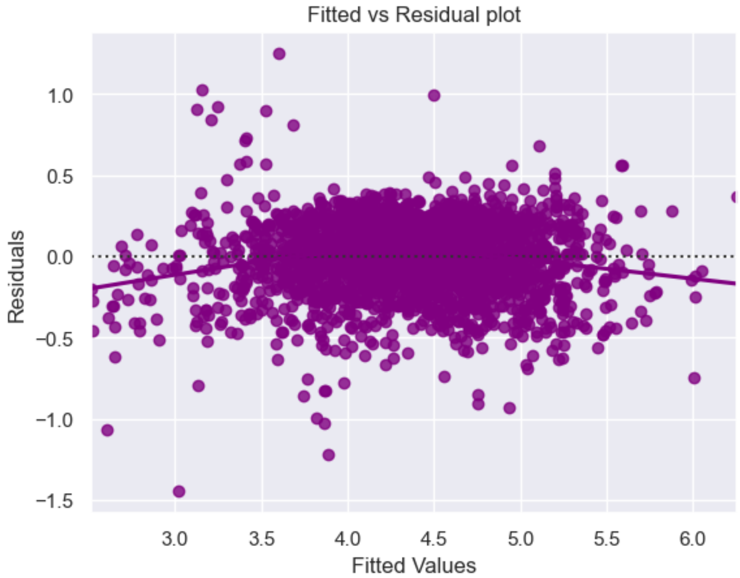
Data Background and Contents

- The dataset, curated in 2021, mirrors the latest trends in the used and refurbished electronic device market, encompassing key features like technical specifications, usage metrics, and market-driven attributes vital for assessing devices' resale values. This ensures the data's contemporary relevance and practical utility for market analysis.
- **Contents:**
  - **Brand Name:** Influence of manufacturer on resale value.
  - **Operating System:** iOS, Android, etc., affecting marketability and value.
  - **Screen Size:** In centimeters, indicating display size preference.
  - **Network Compatibility:** 4G and 5G support, relevant to technology evolution.
  - **Camera Quality:** Resolution of main and selfie cameras in megapixels.
  - **Memory Specs:** Internal memory (ROM) and RAM in GB, affecting performance and storage.
  - **Battery Capacity:** Measured in mAh, indicating usage duration.
  - **Weight:** Device portability factor.
  - **Release Year:** Model's original release year, influencing technology relevance.
  - **Days Used:** Measure of wear and tear, directly impacting resale value.
  - **Normalized Prices:** Comparisons across models/brands with normalized new and used prices.

Model Assumptions

- Linearity: Checked using scatter plots of fitted values vs residuals
    - No pattern indicates the model is linear and residuals are independent.

	Actual Values	Fitted Values	Residuals
1744	4.261975	4.311553	-0.049578
3141	4.175156	3.860832	0.314324
1233	4.117410	4.433246	-0.315836
3046	3.782597	3.848331	-0.065733
2649	3.981922	3.920911	0.061011



Model Assumptions

- Multicollinearity: Assessed with VIF scores
    - VIF score = 1 no correlation
    - VIF score > 5 or close shows signs of moderate multicollinearity
    - VIF score >10 show signs of high multicollinearity

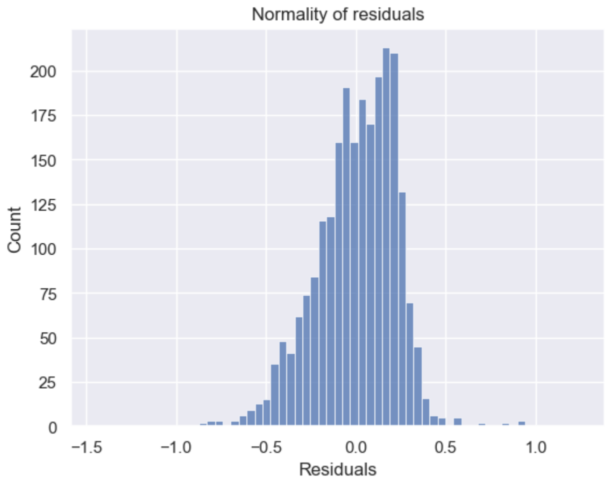
	feature	VIF
0	const	232.676933
1	screen_size	8.262147
2	main_camera_mp	2.418167
3	selfie_camera_mp	2.872720
4	int_memory	1.363390
5	ram	2.283507
6	battery	4.066126
7	weight	6.417982
8	days_used	2.580338
9	normalized_new_price	3.218722
10	years_since_release	4.878548
11	brand_name_Alcatel	3.458576
12	brand_name_Apple	11.195090
13	brand_name_Asus	3.652764
14	brand_name_BlackBerry	1.623330
15	brand_name_Celkon	1.873667
16	brand_name_Coolpad	1.575445
17	brand_name_Gionee	2.076683
18	brand_name_Google	1.388001
19	brand_name_HTC	3.460299
20	brand_name_Honor	3.559332
21	brand_name_Huawei	6.395779
22	brand_name_Infinix	1.191800
23	brand_name_Karbonn	1.628707
24	brand_name_LG	5.354573
25	brand_name_Lava	1.826002
26	brand_name_Lenovo	4.705199
27	brand_name_Meizu	2.417090
28	brand_name_Micromax	3.779514
29	brand_name_Microsoft	2.092692
30	brand_name_Motorola	3.487485
31	brand_name_Nokia	3.755089
32	brand_name_OnePlus	1.586224
33	brand_name_Oppo	4.285316
34	brand_name_Others	10.833481
35	brand_name_Panasonic	1.890592
36	brand_name_Realme	1.978470
37	brand_name_Samsung	8.013571
38	brand_name_Sony	2.891092
39	brand_name_Spice	1.638484
40	brand_name_Vivo	3.734375
41	brand_name_XOLO	2.163880
42	brand_name_Xiaomi	4.082217
43	brand_name_ZTE	4.343056
44	os_Others	1.885820
45	os_Windows	1.742679
46	os_iOS	10.037221
47	4g_yes	2.543525
48	5g_yes	1.808205

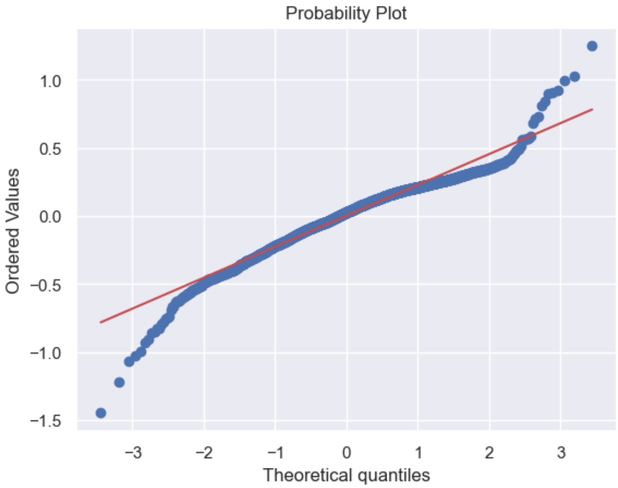
Model Assumptions

- Homoscedasticity: Tested using the goldfeldquandt test.
  - P-value = 0.871
  - P-value > 0.05 means residuals are homoscedastic.

Model Assumptions

- Normality: Validated using Q-Q plots and the Shapiro-Wilk test





- Shapiro-Wilk test: P-value = 2.99 (> 0.05 means residuals are normally distributed)

Additional Recommendations

- Inventory Management Based on Predictive Analysis
  - Utilize the ML model to forecast future trends in device popularity and price depreciation. This predictive insight can guide inventory purchases, ensuring that ReCell stocks up on devices that are expected to retain value or are predicted to become popular.
- Customized Customer Recommendations
  - Implement a recommendation system for customers on the ReCell platform that suggests devices based on their past searches, preferences, and budget. This personalized approach can enhance customer experience and potentially increase sales.
- Sustainable Practices and Marketing
  - Given the environmental benefits of refurbishing and reusing electronic devices, ReCell could emphasize its commitment to sustainability in its marketing campaigns. This not only appeals to environmentally conscious consumers but also aligns with global efforts towards reducing electronic waste.
- Training and Development
  - Invest in continuous training for the data science team to keep up with the latest trends and techniques in machine learning and data analysis. This ensures that ReCell's pricing strategy remains cutting-edge and data-driven.

Great Learning

Happy Learning !