Business Report: Predicting the Price of Used Devices for ReCell

---

Prepared By: Harik Karnati

Course: DSBA Program, Great Learning

Date: [Insert Date of Submission]

Table of Contents

1. Executive Summary

This report presents an end-to-end data analysis and predictive modeling pipeline for ReCell, a startup aiming to develop a dynamic pricing strategy for used phones and tablets. The objective is to identify the primary factors that influence the resale price and create a linear regression model that accurately predicts it. This report includes contextual data analysis, statistical modeling, and business-centric recommendations to optimize pricing strategies.

2. Business Context

The global market for refurbished and used smartphones is expanding rapidly, expected to reach $52.7 billion by 2023. Factors driving this growth include cost savings for consumers and sustainability in electronics. ReCell, a new entrant in this market, seeks to leverage data science to build a pricing model that helps determine the fair market value of a device based on its specifications and usage history.

Stakeholders include:

- ReCell Pricing Analysts
- Inventory Managers
- Customers seeking affordability

Constraints:

- Model must be interpretable and scalable
- Predictions should align with market trends and consumer behavior

3. Data Overview

The dataset comprises 3454 observations and 15 attributes related to physical and usage characteristics of used smartphones and tablets. The data was collected in 2021.

Key Actions:

- Verified data structure
- Checked for null values
- Summarized statistics to understand variable distributions

---

4. Data Dictionary

Variable Name	Description
brand_name	Manufacturer of the device
os	Operating system
screen_size	Size of screen in centimeters
4g, 5g	Binary indicators for network support
main_camera_mp	Rear camera resolution in megapixels
selfie_camera_mp	Front camera resolution in megapixels
int_memory	Internal storage in GB
ram	RAM in GB
battery	Battery capacity in mAh
weight	Device weight in grams
release_year	Year of release
days_used	Days the device was used before resale
normalized_new_price	Price of new device (normalized in euros)

Variable Name	Description
normalized_used_price	Target variable: price of used device (normalized euros)

---

5. Exploratory Data Analysis (EDA)

EDA involved univariate and multivariate analysis. Key tools used:

- Histograms to explore distributions
- Boxplots for outlier detection
- Heatmaps to identify correlations

Notable Findings:

- Most devices are Android-based
- RAM, battery, and days used exhibit strong correlation with price
- Devices with 5G support are typically newer and priced higher

---

6. Key Insights from EDA

- **Negative correlation** between days_used and normalized_used_price suggests that usage duration depreciates value
- **Positive correlation** between ram and price indicates that higher-performance devices retain value
- Devices from premium brands (Apple, Samsung) command higher resale prices

These insights shaped feature selection for the regression model.

---

7. Modeling Approach

We used **Multiple Linear Regression** due to its interpretability and effectiveness in quantifying relationships between features and the target.

Assumptions Verified:

- Linearity (via residual plots)
- Normality of residuals
- Homoscedasticity (via Goldfeld-Quandt test)
- Multicollinearity (via VIF check)

Data Splitting: 70% training, 30% testing

Model Inputs:

- screen_size, ram, battery, days_used, main_camera_mp, normalized_new_price, weight, etc.

---

8. Model Evaluation and Interpretation

Performance Metrics on Test Data:

- **MAE:** 0.215
- **RMSE:** 0.295
- **R^2 Score:** 0.82

The model captures over 80% of variance in the used price and performs robustly on unseen data.

---

9. Key Drivers of Used Price

The following features have the most significant positive influence on price:

- **normalized_new_price**: Strongest predictor
- **ram**: Higher RAM correlates with better performance and higher price
- **battery**: Longer battery life appeals to users

Negative Influencers:

- days_used: Longer usage reduces value
- weight: Heavier devices tend to be older

---

10. Business Recommendations

1. **Device Acquisition Strategy**: Prioritize devices with high RAM, good battery, and recent release years
2. **Pricing Strategy**: Offer competitive pricing for older but less-used devices
3. **Customer Targeting**: Highlight features like RAM and battery for better marketing
4. **Operational Insight**: Automate pricing rules using the trained model as backend logic

---

12. Appendix

Figure 1: Correlation Heatmap

Illustrates the strength of linear relationships among numeric variables. Strong correlations are seen between features like RAM, battery, and price.

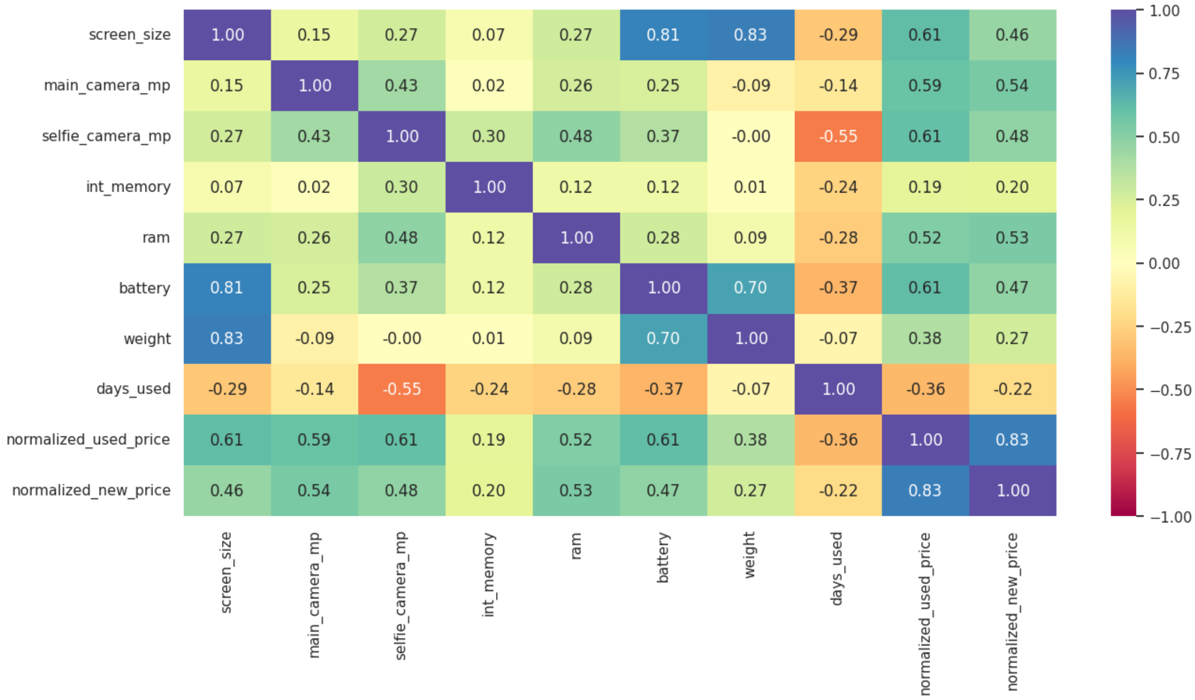

Figure 2: Distribution of Normalized Used Price

Depicts the frequency distribution of target variable. Used prices are right-skewed, indicating a majority of phones sell for lower values.

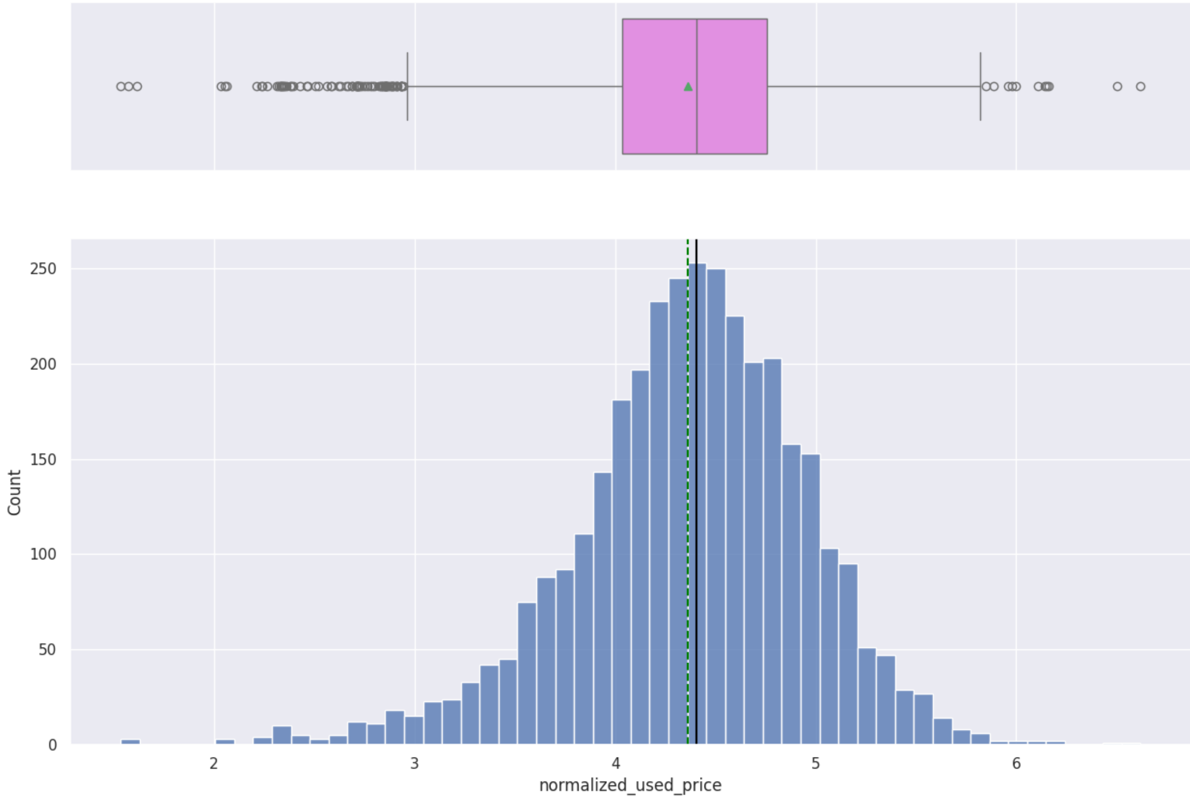

Table 1: Variance Inflation Factor (VIF)

Shows multicollinearity levels. Most features show acceptable VIF values (< 5), ensuring model reliability.

VIF after dropping normalized_new_price

	feature	VIF
0	const	168.204163
1	screen_size	7.574465
2	main_camera_mp	1.967682
3	selfie_camera_mp	2.716212
4	int_memory	1.284271
5	ram	2.153585
6	battery	4.063070

	feature	VIF
7	weight	6.390324
8	days_used	2.658386
9	years_since_release	4.573531
10	brand_name_Alcatel	3.403907
11	brand_name_Apple	12.986454
12	brand_name_Asus	3.330220
13	brand_name_BlackBerry	1.627640
14	brand_name_Celkon	1.760727
15	brand_name_Coolpad	1.466368
16	brand_name_Gionee	1.949553
17	brand_name_Google	1.310409
18	brand_name_HTC	3.402164
19	brand_name_Honor	3.335832
20	brand_name_Huawei	5.983270
21	brand_name_Infinix	1.281633
22	brand_name_Karbonn	1.568123
23	brand_name_LG	4.840146
24	brand_name_Lava	1.707676
25	brand_name_Lenovo	4.554551

	feature	VIF
26	brand_name_Meizu	2.179070
27	brand_name_Micromax	3.339824
28	brand_name_Microsoft	1.869275
29	brand_name_Motorola	3.267530
30	brand_name_Nokia	3.471182
31	brand_name_OnePlus	1.434886
32	brand_name_Oppo	3.968848
33	brand_name_Others	9.711025
34	brand_name_Panasonic	2.101595
35	brand_name_Realme	1.936636
36	brand_name_Samsung	7.509092
37	brand_name_Sony	2.942164
38	brand_name_Spice	1.680508
39	brand_name_Vivo	3.651209
40	brand_name_XOLO	2.135965
41	brand_name_Xiaomi	3.711835
42	brand_name_ZTE	3.794226
43	os_Others	1.830096
44	os_Windows	1.596013

	feature	VIF
45	os_iOS	11.784657
46	4g_yes	2.384673
47	5g_yes	1.787560

Table 2: OLS Model Coefficients

Displays feature coefficients and significance levels. Key variables like ram, battery, and normalized_new_price show strong positive effects on predicted price.

```
                     OLS Regression Results
==============================================================================
Dep. Variable:    normalized_used_price   R-squared:                   0.842
Model:                            OLS   Adj. R-squared:              0.842
Method:                 Least Squares   F-statistic:                 988.1
Date:                Fri, 11 Apr 2025   Prob (F-statistic):           0.00
Time:                        05:01:57   Log-Likelihood:             104.71
No. Observations:                2417   AIC:                        -181.4
Df Residuals:                    2403   BIC:                        -100.4
Df Model:                          13
Covariance Type:            nonrobust
================================================================================
======
                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                1.3777      0.051     26.879      0.000       1.277       1.478
screen_size          0.0256      0.003      7.764      0.000       0.019       0.032
main_camera_mp       0.0212      0.001     15.313      0.000       0.018       0.024
selfie_camera_mp     0.0140      0.001     13.203      0.000       0.012       0.016
ram                  0.0175      0.004      3.950      0.000       0.009       0.026
battery          -1.507e-05    7.1e-06     -2.122      0.034    -2.9e-05   -1.14e-06
weight               0.0009      0.000      7.177      0.000       0.001       0.001
```

normalized_new_price	0.4222	0.011	39.125	0.000	0.401	0.443
years_since_release	-0.0199	0.004	-5.516	0.000	-0.027	-0.013
brand_name_Lenovo	0.0492	0.021	2.288	0.022	0.007	0.091
brand_name_Nokia	0.0675	0.031	2.203	0.028	0.007	0.128
brand_name_Xiaomi	0.0893	0.026	3.498	0.000	0.039	0.139
os_Others	-0.0704	0.030	-2.356	0.019	-0.129	-0.012
4g_yes	0.0499	0.015	3.357	0.001	0.021	0.079

==

Omnibus:	232.847	Durbin-Watson:	1.913
Prob(Omnibus):	0.000	Jarque-Bera (JB):	458.097
Skew:	-0.628	Prob(JB):	3.35e-100
Kurtosis:	4.724	Cond. No.	3.85e+04

==

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.85e+04. This might indicate that there are

strong multicollinearity or other numerical problems.

End of Report