

ReCell – Project

Machine Learning based Dynamic Pricing Strategy Solution

Contents

1. Business Problem Overview and Solution Approach
2. Data Overview
3. Exploratory data analysis
4. Model Performance Summary
5. Business Insights and Recommendations

Business Problem Overview and Solution Approach

- **Core business idea**

Used and refurbished phone market has grown considerably over the past decade, and it continues to provide cost-effective alternatives to both consumers and businesses that are looking to save money when purchasing a smartphone. Used and refurbished devices can be sold with warranties and can also be insured with proof of purchase and with many other lucrative offers. This helps reduce environmental impact and helps in recycling and reducing waste. With all these benefits and growing market there is a huge potential for online retailers to tap on this under the radar market.

- **Problem to tackle**

ReCell, a startup aiming to tap the potential in this market, needs an ML-based solution to develop a dynamic pricing strategy for used and refurbished smartphones. They have gathered data on used smart phones which were sold. Key objective is to analyze the data and draw actionable insights and build a model to predict price for used phones and identify factors which significantly influence it.

- **Financial implications**

Without a right dynamic pricing strategy business either could lose the edge in the market or else could not maximize their profit. Either way it will be a loss of their revenue and potentially lose chance of growing market which forecasted to be \$52.7bn by 2023.

- **How to use ML model to solve the problem**

After analyzing and carefully preparing the data, Machine learning models can use the prepared data to learn and identify significant factors using which it can predict outcomes of unseen data.

For this problem we will analyze and prepare the data, feed into regression models.

- Analyze and prepare the data.
- Create regression model fit the prepared data into it.
- Test the performance of different models.
- Finalize most accurate and stable model which will predict the price for used phones based on factors which significantly influence it

Data Overview

- Data Dictionary

Variables	Description
1. brand_name	Name of manufacturing brand
2. os	OS on which the phone runs
3. Screen size	Size of the screen in cm
4. 4g	Whether 4G is available or not
5. 5g	Whether 5G is available or not
6. main_camera_mp	Resolution of the rear camera in megapixels
7. selfie_camera_mp	Resolution of the front camera in megapixels
8. int_memory	Amount of internal memory (ROM) in GB
9. ram	Amount of RAM in GB
10. battery	Energy capacity of the phone battery in mAh
11. weight	Weight of the phone in grams
12. release_year	Year when the phone model was released
13. days_used	Number of days the used/refurbished phone has been used
14. new_price	Price of a new phone of the same model in euros
15. used_price	Price of the used/refurbished phone in euros

Observations	Variables
3571	15

Notes:

1. Dataset looked consistent with the dictionary provided.
2. There are no duplicate values
3. There are missing values across 6 variables. Counts of missing values across columns are shown in the table beside

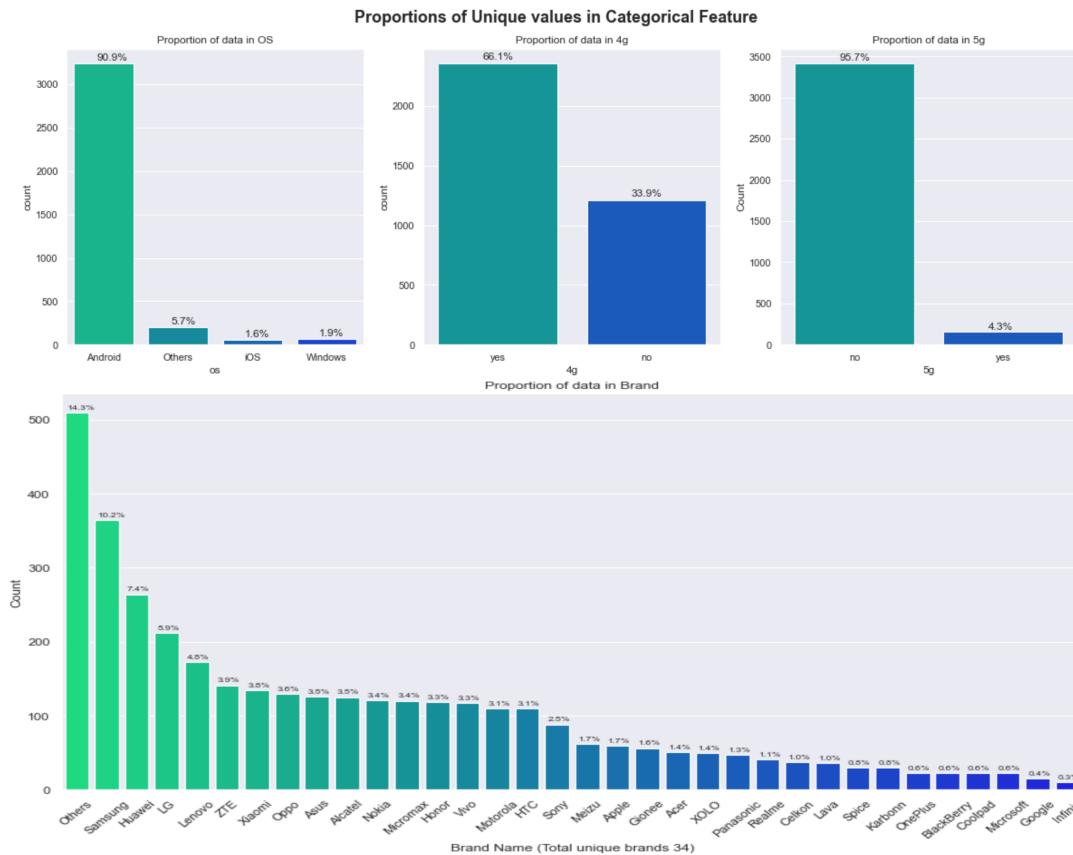
Missing value counts	
brand_name	0
os	0
screen_size	0
4g	0
5g	0
main_camera_mp	180
selfie_camera_mp	2
int_memory	10
ram	10
battery	6
weight	7
release_year	0
days_used	0
new_price	0
used_price	0

Data Overview contd..

- Brief description of significant manipulations made to raw data
 1. Missing values are treated with Median values of the variable
 2. Outliers' treatment was done on the columns except 'ram' as it is a discrete variable.
 3. We have replaced the outliers with their lower and upper whisker value depending on sides of their presence.
 4. We have converted screen size from cm to inch which is more common terms in phone size
 5. We have grouped brands into three categories 'budget', 'midrange', and 'highend' based on their average new price range.

Exploratory Data Analysis

- Data proportions based on categorical features



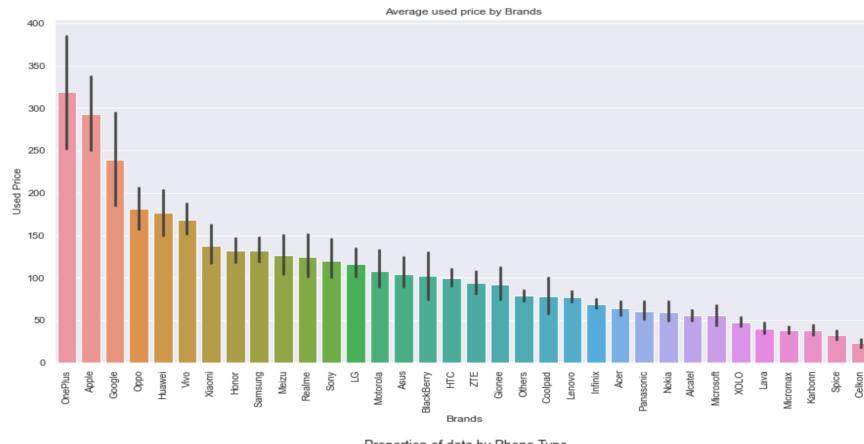
Notes:

1. Android phones are predominant in market, close to ~91%
2. Used phones with 4g capabilities are sold more in the market, close to ~66%
3. 5g is newer technology and not yet has much presence(only ~4%) in used phone market

Notes:

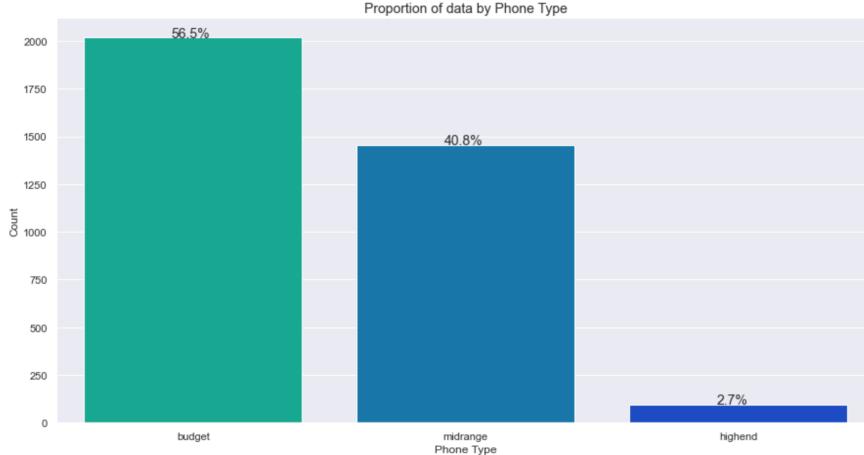
1. Samsung phones, with ~10% of the data are the most popular in used phone market
2. Huawei and LG are the next two popular brands
3. Data with brand as others couldn't be interpreted as there was no possible way to know what brands are considered under this.

Exploratory Data Analysis



Notes:

1. On average used price in market OnePlus having highest average used price, Apple and google are the next two.
2. Celkon has the lowest average price for used phones

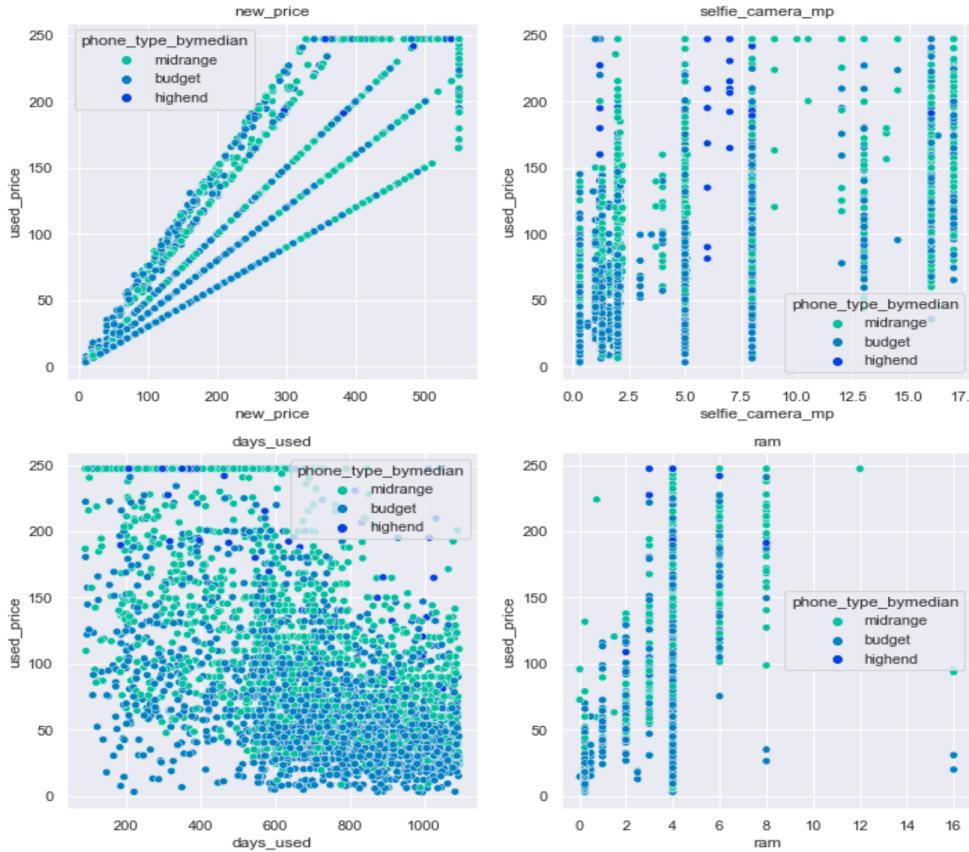


Notes:

1. Data are grouped in three categories based on their average new price. (budget <200 euros, midrange 200 to 400 euros, highend above 400 euros)
2. Budget phones are almost 56% of used phone market.
3. Highend phones are only close to ~3%

Exploratory Data Analysis

Used price vs new price, selfie camera, days used, and ram

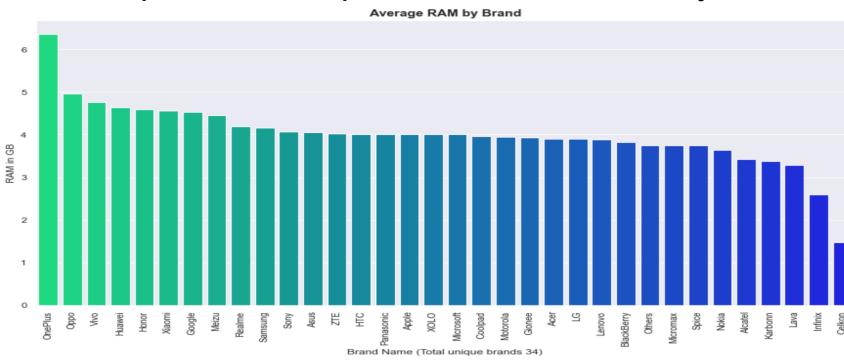


Notes:

1. The higher the new price of the phone the higher the used price will be.
2. The higher the number of days the phone is used the price of the phone in used market will decrease
3. The higher the ram of the phone, price of the used phone increase
4. Statistically Selfie camera shows positive correlation however visually it's not that evident, we will find more during our modeling

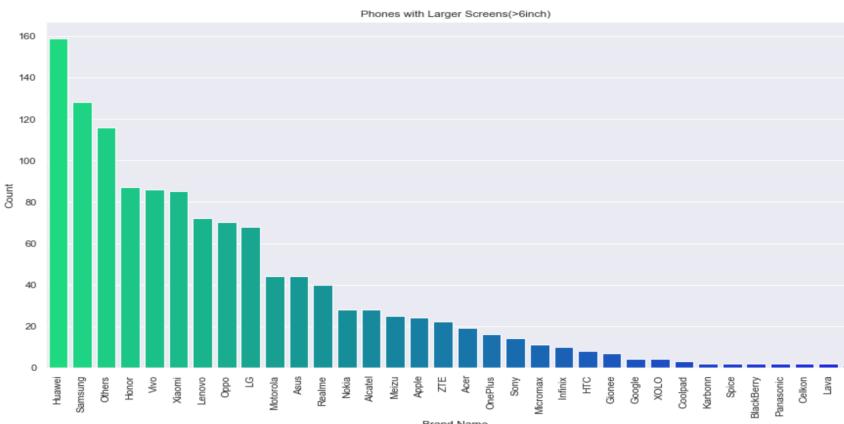
Exploratory Data Analysis

Used price vs new price, selfie camera, days used, and ram



Notes:

1. OnePlus offers the average higher rams in their phones. It has ~6GB of average rams.
2. Celkon offers the lowest average ram.
3. On average most phones have 4GB of ram.



Notes:

1. Huawei and Samsung are market leaders in terms of bigger size screen phones
2. Microsoft shows the least performer when it comes to bigger screens

Model Performance Summary

Approach

- Explored Linear Regression models from sklearn and statsmodels.
- 30% of randomly chosen data are kept for testing model performance.
- Observation count for model training 2499, and testing 1072
- Model performance was checked against RMSE, MAPE, MAE, R squared, Adj. R squared

Model and Parameters

- Initial LinearRegression from sklearn and OLS from statsmodels were prepared with all variables, except brand_name, as we have binned this variable in phone type and we have used the same.
- OLS model is used for further tuning and dropping features based on significance using VIF and p-value.

features	coefficient	p-value
'main_camera_mp',	-0.2056	0.013
'selfie_camera_mp',	0.7288	0.000
'int_memory',	0.0855	0.000
'days_used',	0.0851	0.000
'new_price',	0.3874	0.000
'os_Others',	-5.4782	0.000
'os_iOS',	6.7202	0.006
'4g_yes'	-3.1396	0.000

- All the features having p-value under 0.05(significance level), So all these features are significant for deriving price for used phones
- Positive coefficients denotes increase the price of used phones with the increase in value for the corresponding attributes, and negative coefficients denotes decrease of the price of used phones with increase in value of the corresponding attributes

Performance of the Models

Training set

	Linear Regression sklearn	Linear Regression statsmodels
RMSE	14.212	14.236
MAE	10.392	10.400
R-squared	0.954	0.953
Adj. R-squared	0.953	0.953
MAPE	18.733	18.809

Test set

	Linear Regression sklearn	Linear Regression statsmodels
RMSE	13.778	13.851
MAE	10.121	10.155
R-squared	0.957	0.957
Adj. R-squared	0.957	0.956
MAPE	16.259	16.290

Notes:

- Train and test sets provided almost similar results and both models are comparable.
- MAE and RMSE on the train and test sets are comparable, which shows that the model was not overfitted
- MAE indicates that model was able to predict used phone price within a mean error of ~10 euro on the test data
- MAPE on the test set suggests we can predict within ~16% of the used price.
- Adjusted R squared in test is .956, which means model is able to explain ~95% of the variances.

Business Insights and Recommendations

Insights

- Android phones with close to ~91% are predominant in used phone market.
- Used phones having 4g features are popular in market
- 5g phones are yet to hit used phone market, as it is newer technology.
- Samsung is the market leader in used phone market, Huawei and LG are the next two brands which are popular
- Budget phones (average new prices < 200 euros) are most sold phones in used market, LG is the most popular brand in this category.
- Huawei and Samsung are market leaders in bigger screen phones.
- Infinix is the lowest performing brand in used phone market.
- On average OnePlus phones have higher ram.
- Xiaomi offers best selfie cameras in budget phones.
- Phone's cameras main and selfie, internal memory, how many days the phone have been used, what was the new price of the phone are some of the most important factor which drives the price of used phones

Recommendations

- Inventory of different Android phones, specifically with some must have brands such as Samsung, Huawei, LG, Xiaomi should be ensured.
- Budget phones should be targeted for volume selling.
- Targeted marketing for multimedia use should be done with Huawei, Samsung phones.
- Targeted marketing for gaming use should be done with OnePlus phones.
- Targeted marketing for great selfie phones in budget should be done with Xiaomi phones.
- 5g Phones are yet to hit market in full fledge as it's a new technology, Tapping into this market likely to increase sales and revenue.
- ML solutions should be leveraged to generate the pricing for their used phone inventories, it will help generate dynamic price and give edge over other retailers to increase sales volume and revenue
- Additional data on sales volumes by time, sales geo location, new phone warranty information, used phone warranty provided or not, should be procured and further analyzed to check significance how it affects the pricing. Accordingly, model can be improved to provide more dynamic pricing.

greatlearning
Power Ahead

Happy Learning !

