

# ReCell

## - Trusetd Reseller

**Data based prediction model of prices for used/refurbished smartphones  
phones**

# Background

1. ReCell, a startup dealing with refurbished and used mobile devices is aiming to tap the potential in the used/refurbished market.
2. ReCell wants to analyze the data provided and build a linear regression model to predict the price of a used phone
3. Also identify factors that significantly influence it.



# Business Problem Overview and Solution Approach

- Financial implications

A new IDC (International Data Corporation) forecast predicts that the used phone market would be worth \$52.7bn by 2023 with a compound annual growth rate (CAGR) of 13.6% from 2018 to 2023.

The rising potential of refurbished and used mobile devices market fuels the need for an ML-based solution.

The solution will provide ReCell the ability to develop a dynamic pricing strategy for used and refurbished smartphones.

ReCell has a good collection of data on the various attributes of the phone that might impact the price of the used and refurbished phones.

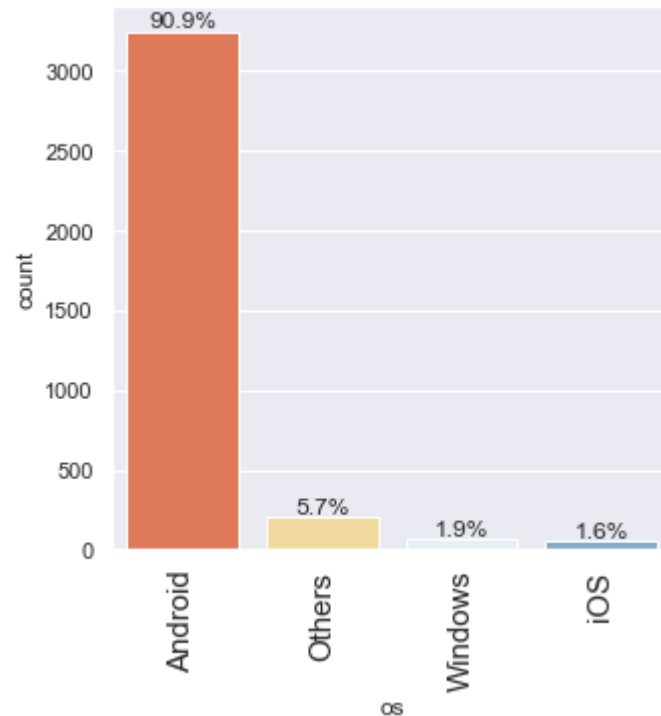
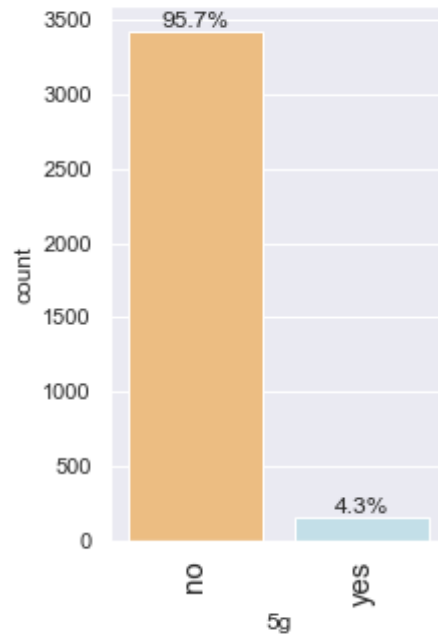
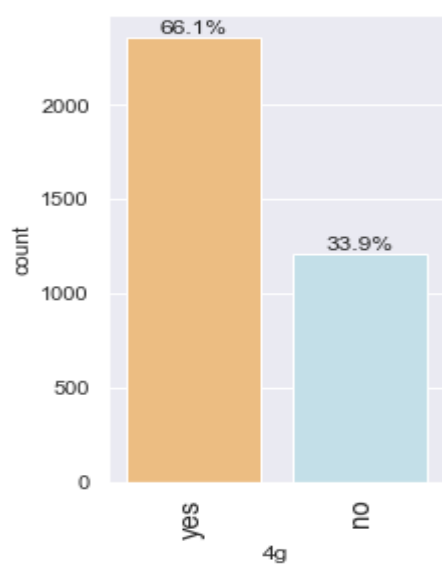
# Data Overview

Variable	Description
brand_name	Name of manufacturing brand
os	OS on which the phone 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 phone battery in mAh
weight	Weight of the phone in grams

Variable	Description
release_year	Year when the phone model was released
days_used	Number of days the used/refurbished phone has been used
new_price:	Price of a new phone of the same model in euros
used_price	Price of the used/refurbished phone in euros

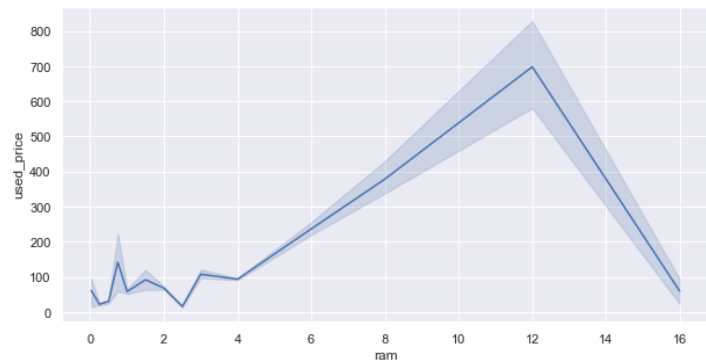
Observations	Variables
3571	15

# Exploratory Data Analysis



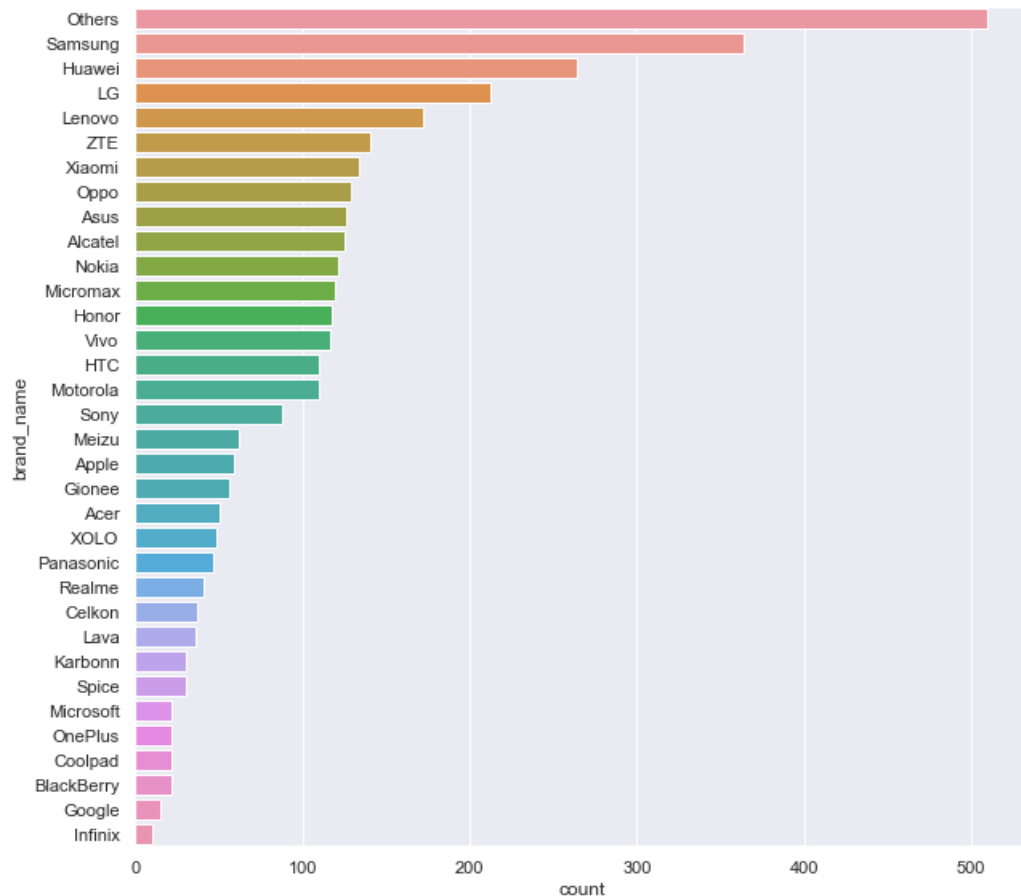
- \* Android has almost 91% share
- \* Most of the phones have 4G
- \* Most of the phones do not have 5G

# EDA

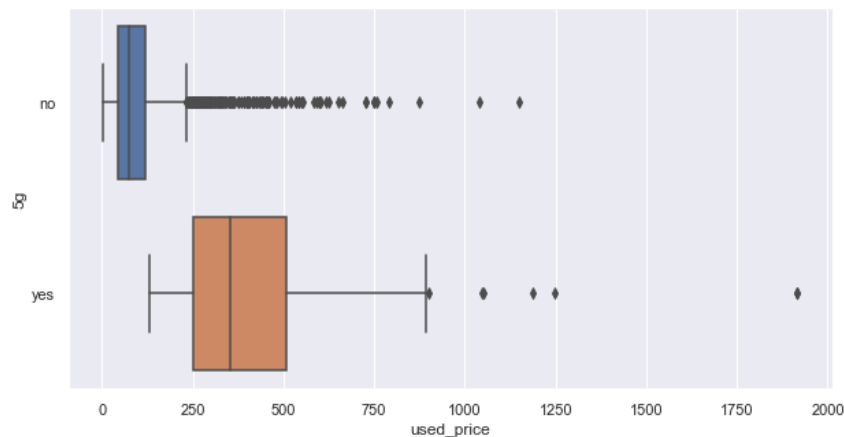


- Increase in ram leads to higher used\_price,
- Interesting that price decreases after ram size reached 12 GB

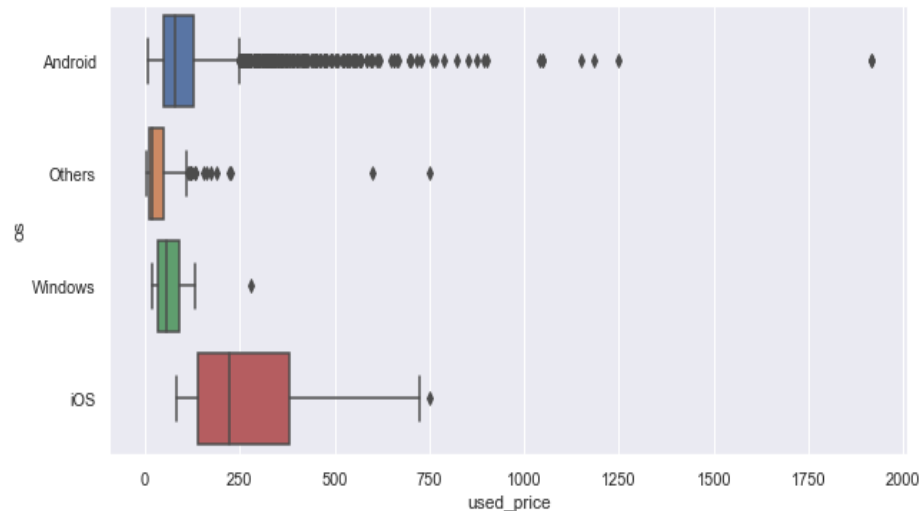
1. Samsung, Huawei and LG are top 3 individual brands
2. There are 33 brands in total
3. Most of the phones are not from the 33 brands



# EDA

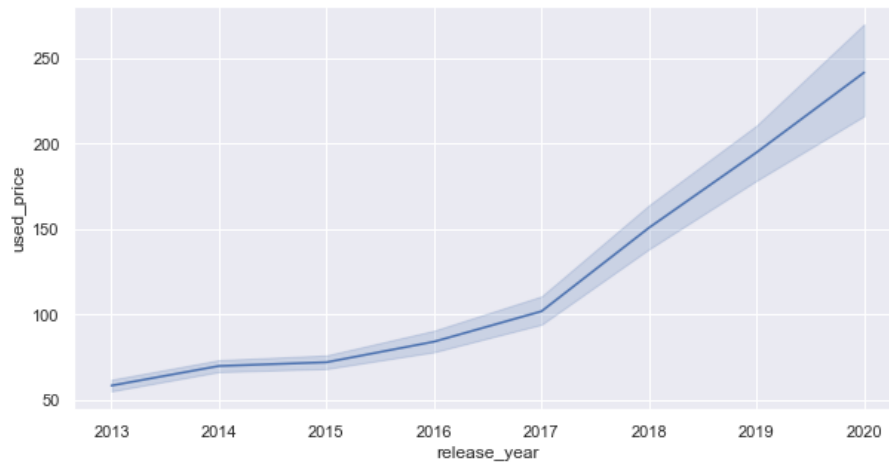
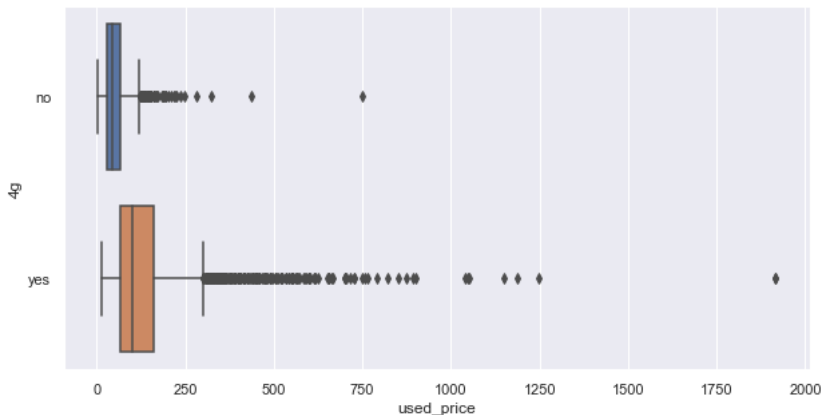


5G phones are in general higher priced



Among the brands 'iOS' has the highest priced phones  
Android phones have large outliers , indication wide variety of phones

# EDA

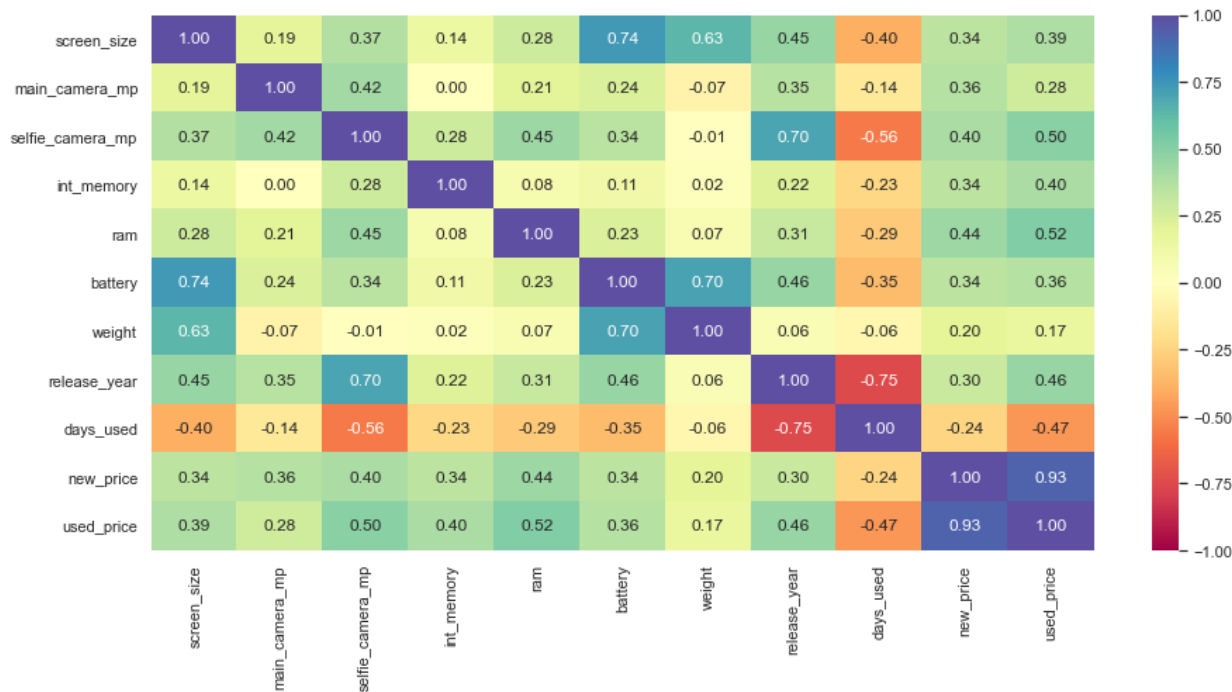


Unlike 5G, there is not considerable difference in price for phones with or without 4G, also there are a lot of outliers, indicating a wide range of 4G phones.

The later the release\_year, the higher the price.



# EDA – Contd.



- Battery and weight are correlated , as expected.
- Screen\_size and battery are correlated , indicating large screen size needs large batteries.
- selfie\_camera\_mp and release\_year are correlated , indicating new models have better selfie cameras.
- Days\_used and used\_price are negatively correlated.

# Overview of ML model and its parameters

- Using the ordinary Least Square Model for regression model
- Some of the variable used in the model  
os, screen\_size, 4g, 5g, main\_camera\_mp, selfie\_camera\_mp, int\_memory, ram, battery, weight, release\_year, days\_used, new\_price and few others
- Total observations: 3571 . Observation taken after dropping nulls: 3368
- There were null values , the treatment for null values was to drop the null rows.
- The columns that were having null values did not have perfect normal distribution
- Some of the missing values were key factor for the dependent variable

# Model Performance Summary

- Key model metrics with on the training data set ( 2357 observation), for the initial Model:
- R-squared: 0.956
- Adj. R-squared: 0.955
- Some of the variables exhibited strong **Multicollinearity**
- **Multicollinearity** was treated through **Variance Inflation Factor (VIF)**
- **VIF: VIF** value of less than **10** was taken as the benchmark for **Multicollinearity**
- The final list of variable taken into consideration for Final model:
- ['selfie\_camera\_mp', 'int\_memory', 'new\_price', 'Alcatel', 'Apple', 'Asus', 'Celkon', 'Coolpad', 'Gionee', 'Google', 'HTC', 'Honor', 'Lava', 'Lenovo', 'Meizu', 'Micromax', 'OnePlus', 'Oppo', 'Others', 'Panasonic', 'Samsung', 'Sony', 'Xiaomi', 'ZTE', 'Others', 'Windows', '5G\_dum']

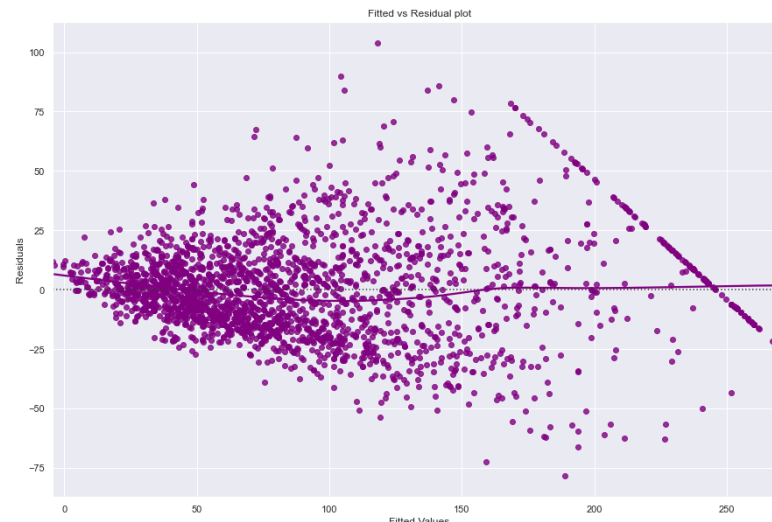
# Model Performance Summary

- Summary of key performance metrics for training from the OLS Model:

Model	R-squared	Adj. R-squared
Initial Model	0.956	0.955
Pre-Final Model – Training Data	0.966	0.965
Final Model – Training Data	0.965	0.965

# Linear Regression Assumptions

- The variable in the final models was tested for **Linear Regression Assumptions**
1. No Multicollinearity : Variables with high VIF scores was removed recursively after checking its impact on the model. VIF score of less than 10 was taken as benchmark.
  2. Linearity of variables:  
The plot of Residuals (errors) Vs fitted values (predicted values) did not show any pattern.
  3. Independence of error terms:  
The plot of Residuals (errors) Vs fitted values (predicted values) did not show any pattern.



# Linear Regression Assumptions – Contd.

## 4. Normality of error terms:

The histogram of residuals does have a bell shape.

The residuals more or less follow a straight line except for the tails.

### Results of the Shapiro-Wilk test.

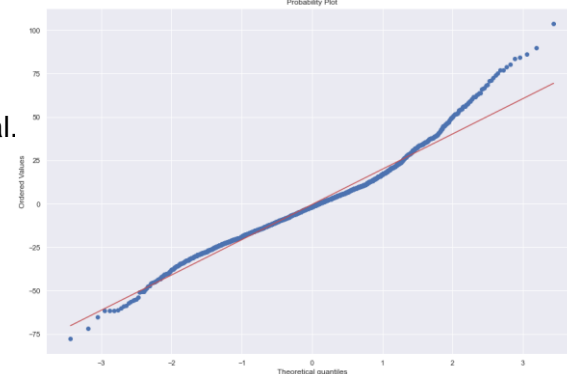
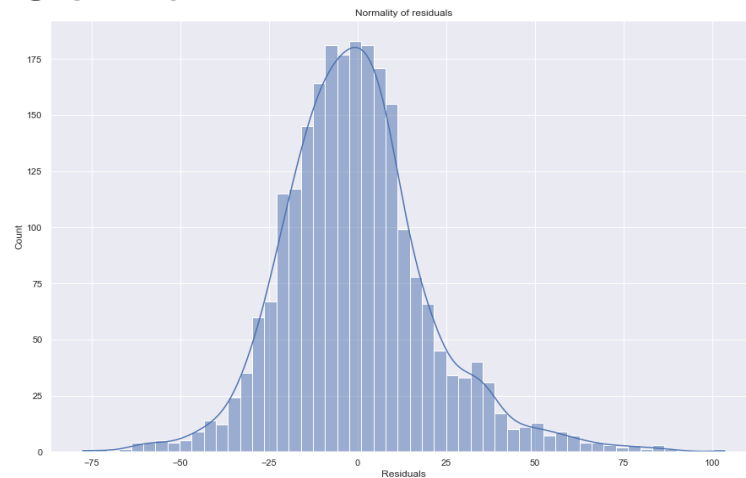
(statistic=0.9716421365737915, pvalue=3.3734527662670446e-21)

Since p-value < 0.05, the residuals are not normal as per the Shapiro-Wilk test.

Strictly speaking, the residuals are not normal.

However, as an approximation, we can accept this distribution as close to being normal.

So, the assumption is satisfied



## Linear Regression Assumptions – Contd.

### 5. No Heteroscedasticity

[('F statistic', 1.0654310933381863), ('p-value', 0.14112269394353816)]

Since  $p\text{-value} > 0.05$ , we can say that the residuals are homoscedastic.  
So, this assumption is satisfied.

### Sample Actual Vs Predicted:

Actual	Predicted
93.1700	118.618383
74.7500	90.585870
247.1925	190.634999
109.8000	80.149991
83.1600	59.411924
247.1925	236.210764
120.8400	111.376061
152.8100	130.988516
60.0200	84.506598
150.3500	110.070237

# Key Model Performance Metrics

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
Training Performance	20.595608	15.388272	0.890274	0.888907	21.080383
Test Performance	21.630428	16.175851	0.893233	0.890077	22.921057

1. The model is able to explain ~89% of the variation in the data, which is very good.
2. The train and test RMSE and MAE differences are low and comparable. So, our model is not suffering from overfitting.
3. The MAPE on the test set suggests we can predict within 23% of the used\_price.
4. Hence, we can conclude the model is good for prediction as well as inference purposes.



# Key Model Performance Metrics - Comparison

- Training Performance Comparison:

	Linear Regression sklearn	Linear Regression statsmodels
RMSE	0.094446	20.595608
MAE	0.076640	15.388272
R-squared	0.983244	0.890274
Adj. R-squared	0.982895	0.888907
MAPE	1.836847	21.080383

- Test Performance Comparison:

The R-squared and Adj. R-squared are comparable between the sklearn learn model and Linear Regression statsmodels

	Linear Regression sklearn	Linear Regression statsmodels
RMSE	0.101140	21.630428
MAE	0.080199	16.175851
R-squared	0.983241	0.893233
Adj. R-squared	0.982405	0.890077
MAPE	1.976830	22.921057

# Business Insights and Recommendations

1. **Android** phones dominate the used phone market. **Nearly 81% of the market share.**
2. 'selfie\_camera\_mp', 'int\_memory', 'new\_price' and 5G have a positive coefficients. So, as they increase, used\_price of the phone also increases.
3. Among the brands 'Apple', 'Google' and 'Honor' have a positive coefficients. So, as these brands have higher used\_price.
4. new\_price: Mean is around 185 Euors used\_price: Mean is around 65 Euors.
5. OnePlus, Apple, Google have highest mean used\_price
6. Among the brands 'iOS' has the highest priced phones
7. Android phones have large outliers , indication wide variety of phones

# Business Insights and Recommendations

8. OnePlus, Oppo and Vivo offer the max RAM
9. Huawei, Samsung offer the most number of phones with screen size greater than 6 inches.
10. Xiaomi, Realme and Vivo offer many models with selfie camera with more than 8Mp on budget price (less than 100 Euros) Sony, Samsung, Oppo, Honor and Huawei also good number of models with selfie camera more than 8Mp on budget.
11. The latest the release\_year the higher the price.

# Business Insights and Recommendations

- Additional data sources for model improvement:

Geography and user attributes.

- Model implementation in real world:

The model can be implemented in real world and feedback taken to improve the model.

- Potential business benefits from model:

ReCell can offer dynamic pricing and offer competitive rates.

**greatlearning**  
*Power Ahead*

Happy Learning !

