Recell

Used Cell Phone Price Prediction

Contents

Business Problem Overview and Solution Approach

- **Core business idea**

  - Refurbished and used devices continue to provide cost-effective alternatives to both consumers and businesses that are looking to save money when purchasing a smartphone.
  - Recell aims to tackle the rising potential of this comparatively under-the-radar market fuels the need for an ML-based solution to develop a dynamic pricing strategy for used and refurbished smartphones.

- **Problem to tackle**

  - The objective is to analyze the data provided and build a linear regression model to predict the price of a used phone and identify factors that significantly influence it.

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

  - A linear regression model can be used to train and test on the available data. This model could then predict with price of used phone with more than 90% accuracy

Data Overview

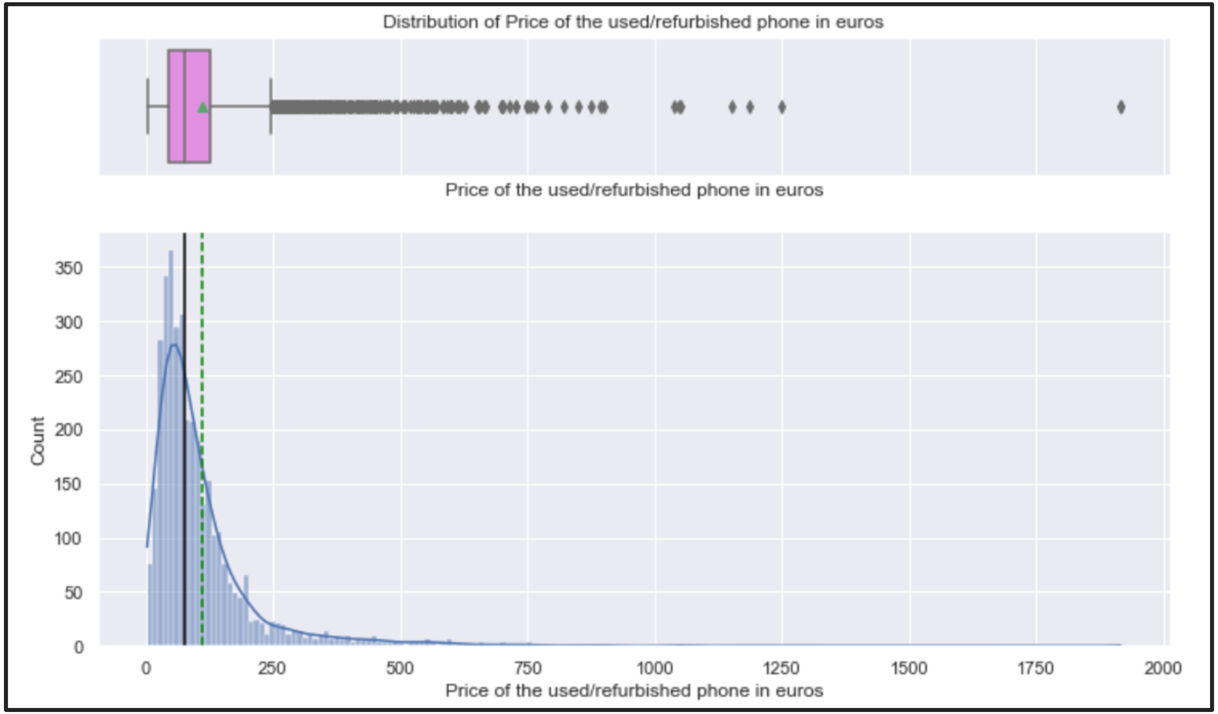
Variable	Description
brand_name:	Name of manufacturing brand
Os	OS on which the phone runs
screen_size	Size of the screen in cm
4g	Whether 4G is available or not
5g	Whether 5G is available or not
main_camera_mp	Resolution of the rear camera in megapixels
selfie_camera_mp	Resolution of the front camera in megapixels
int_memory	Amount of internal memory (ROM) in GB
Ram	Amount of RAM in GB

Variable	Description
battery	Energy capacity of the phone battery in mAh
weight	Weight of the phone in grams
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	Columns
3571	15

Exploratory Data Analysis

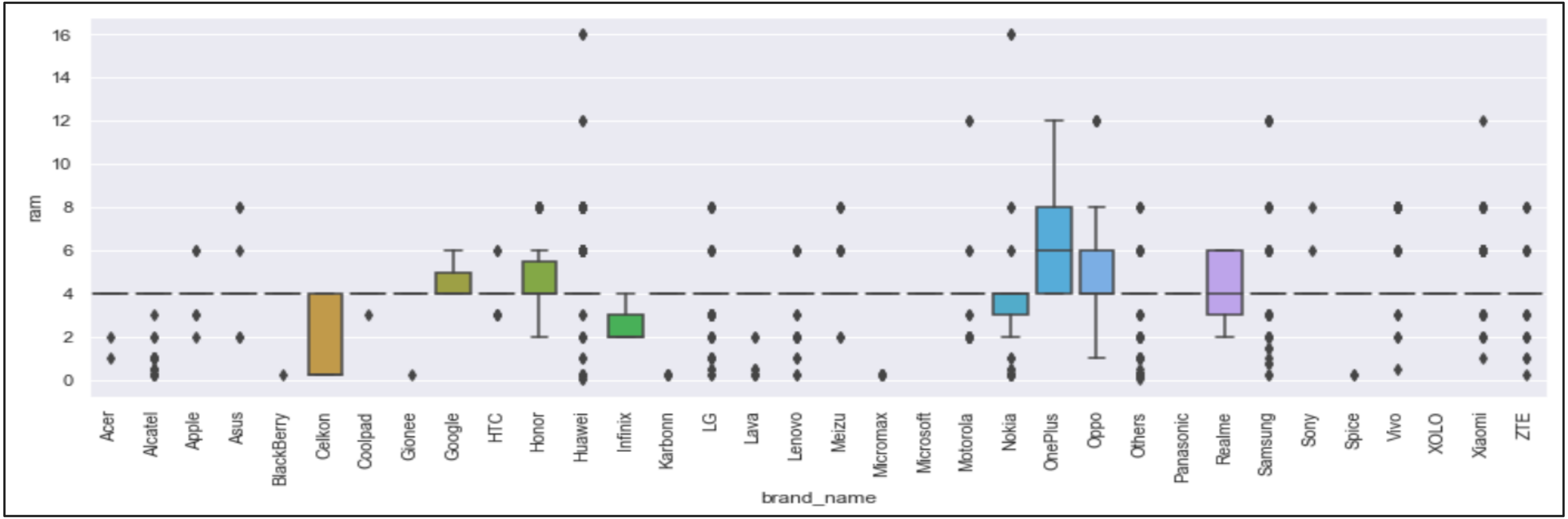
Distribution of Price of the used/refurbished phone



- The distribution of price of used phone is right skewed.
- The boxplot indicates the presence of outliers which means some used phones are expensive.
- Mean, which is 109.9, is much greater than median due to extreme values towards higher end.

Exploratory Data Analysis

RAMs by Brand



Each brand has on an average 4GB of RAM in their phones.
OnePlus is the brand with a mean around 6GB

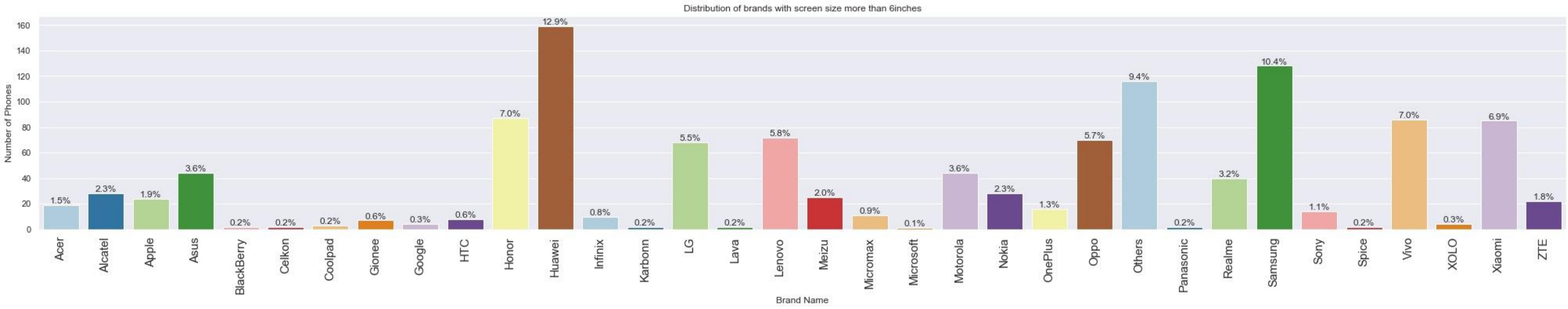
Exploratory Data Analysis

Distribution of Weight for Phones with more than 4500mAh battery

- For phones with large battery, many phones weigh around or above 200gms.
- However, there are outliers from as small as 23gms to 950gms.

Exploratory Data Analysis

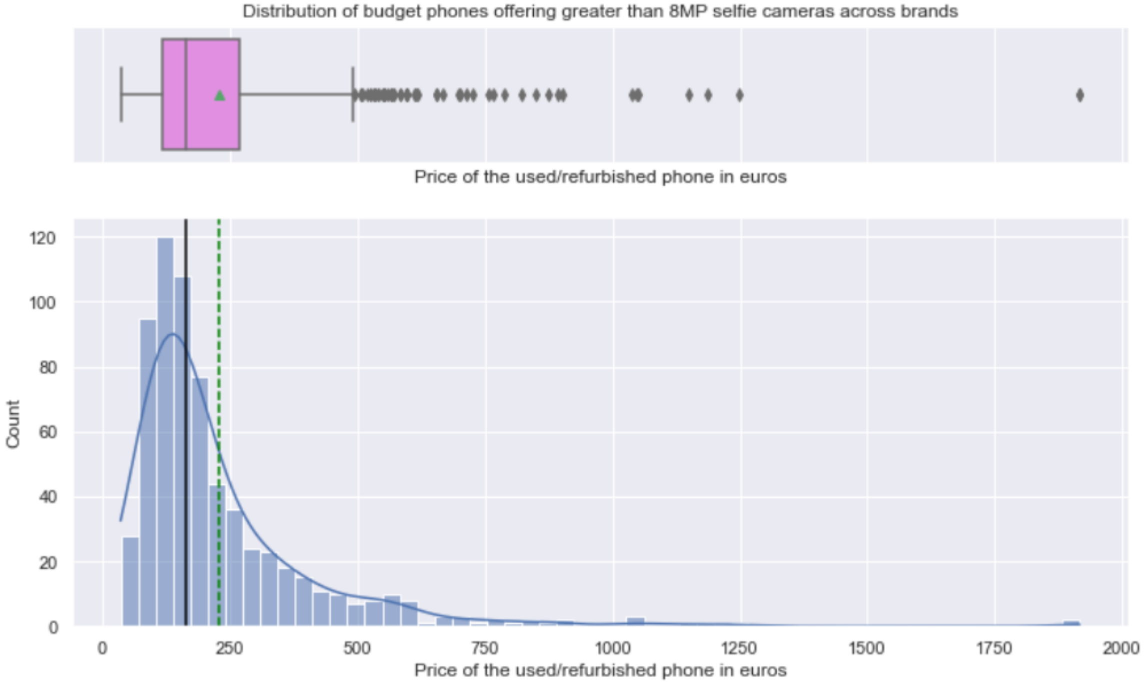
Screen size greater than 6 inches by Brands



- Huwaei has the highest number of phones with more than 6inches of screen size.
- Microsoft is the brand with least number of phones having screen size more than 6 inches.

Exploratory Data Analysis

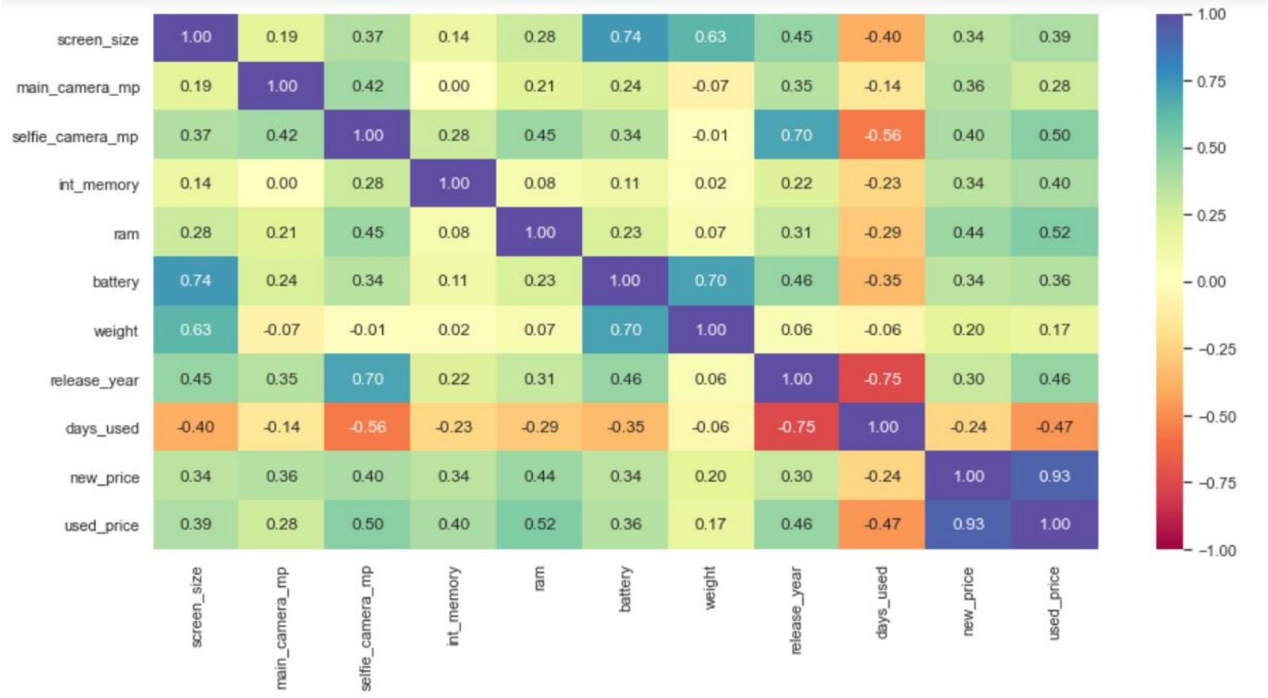
Distribution of Budget Phones with a selfie camera greater than 8MP



- Distribution of used price (untransformed) is right skewed for budget phones with selfie camera more than 8MP
- Huwaei tops the list in this category

Exploratory Data Analysis

Correlation of different variables with Used Price of a Phone



- used_price is highly correlated with new_price. The higher new price could point to a higher used price as we saw in their bivariate analysis.

- RAM and selfie camera does also have a very close correlation with used price.

- With days used, used price has a negative correlation, which might be pointing to the fact that the longer a phone has been used the lesser the used price becomes.

Pre-Processing of Data

Category	Variables	Method
Fixing the data types	brand_name Os,4g,5g	Conversion of object data type to category
Missing value Treatment	main_camera_mp int_memory,ram, weight ,battery, selfie_camera_mp	Replacing missing values with median for each variable
Feature Engineering	Screen_size	Conversion of units from cm to inches
Log transformation	Used_price, New_price	Transformation to remove skewness
Outlier treatment	Weight,Screen_size	

Model Performance Summary

- **Overview of ML model and its parameters**

  - Multiple Linear Regression Model was built to

    - Find dependency of target variable on predictors
    - Predict fitted values and compare them to actual values
    - Dependent variable : ‘used_price’
    - Independent variables : all other variables including dummy variables

  - Data was split into train and test data at a ratio 70:30

    - Number of rows in train data =2499
    - Number of rows in test data = 1072

Model Performance Summary

- **Performance matrices for the first ML Model**

Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.081557	0.067535	0.990146	0.989953	1.650466

Test Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.080874	0.069081	0.989808	0.989329	1.650876

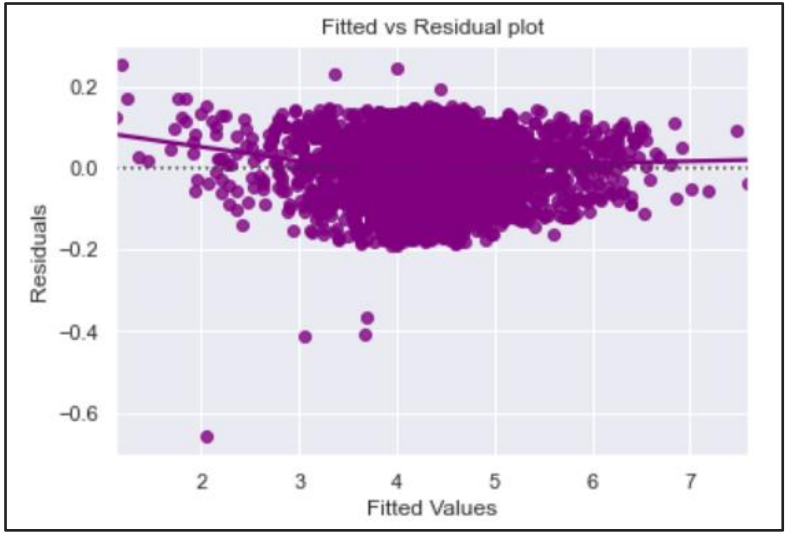
- The training R^2 is 99%, indicating that the model explains 99% of the variation in the train data. So, the model is not underfitting.
- MAE and RMSE on the train and test sets are comparable, which shows that the model is not overfitting.
- MAE indicates that our current model is able to predict used price within a mean error of 0.07 euros on the test data.
- MAPE on the test set suggests we can predict within 1.65% of the used price.

Model Performance Summary

- **Assumptions of Linear Regression**

- Linearity and Independence of Variables

	Actual Values	Fitted Values	Residuals
844	4.619862	4.560641	0.059220
1539	4.724552	4.771691	-0.047139
3452	4.743975	4.637224	0.106751
1727	4.175771	4.259117	-0.083346
1926	4.233382	4.190340	0.043041

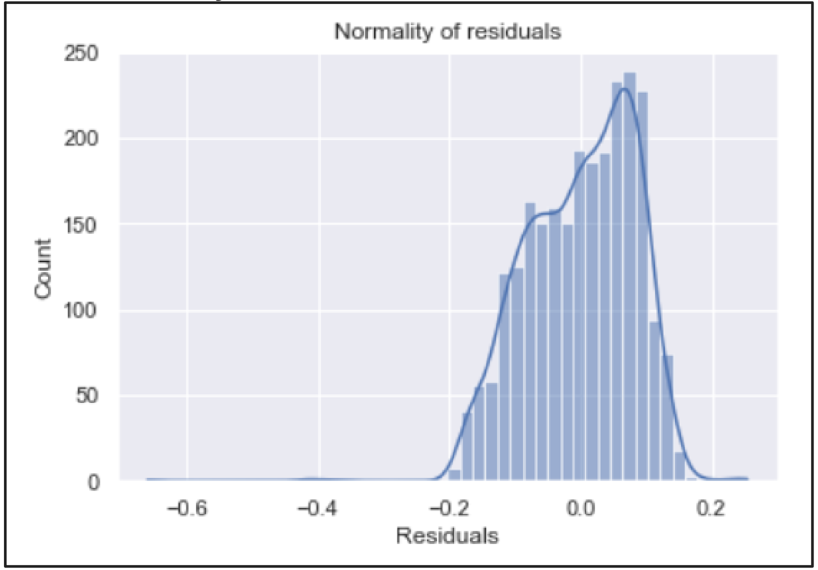


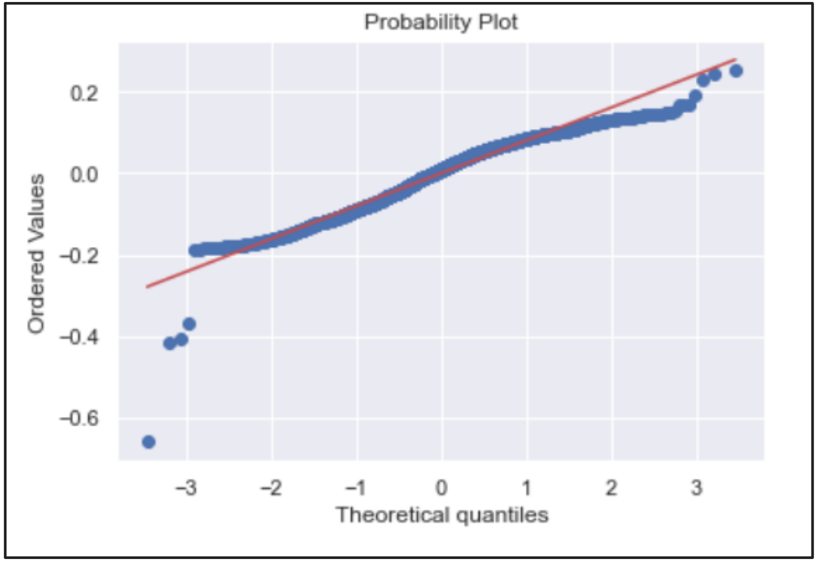
- The scatter plot shows the distribution of residuals (errors) vs fitted values (predicted values).
- If there exist any pattern in this plot, we consider it as signs of non-linearity in the data and a pattern means that the model doesn't capture non-linear effects.
- **We see no pattern in the plot above. Hence, the assumptions of linearity and independence are satisfied.**

Model Performance Summary

- **Assumptions of Linear Regression**

- Normality





- Even though there is a slight skewness to the left, the curve considered to be normal.
- The residuals more or less follow a straight line except for the tails.

Model Performance Summary

- **Assumptions of Linear Regression**

- No Heteroscedasticity:

❖ The goldfeldquandt test was used. If we get a p-value > 0.05 we can say that the residuals are homoscedastic. Otherwise, they are heteroscedastic.

❖ - Null hypothesis: Residuals are homoscedastic

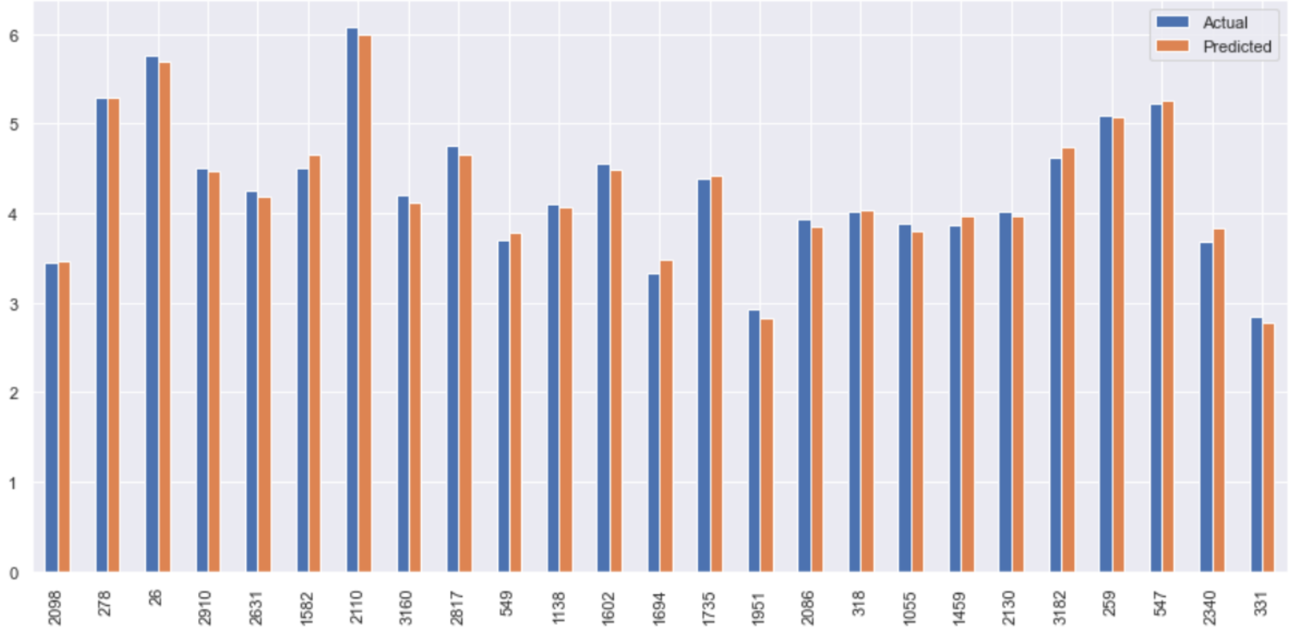
❖ - Alternate hypothesis: Residuals have heteroscedasticity

```
[('F statistic', 1.0518720430356685), ('p-value', 0.18653840792817605)]
```

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

Model Performance Summary

- **Predictions on the test set**



	Actual	Predicted
2098	3.450622	3.458192
278	5.281527	5.297606
26	5.754666	5.697975
2910	4.510530	4.473533
2631	4.251348	4.183939
1582	4.506233	4.651058
2110	6.070230	5.985111
3160	4.194793	4.117877
2817	4.760206	4.656976
549	3.696103	3.782893

- We can observe here that our model has returned pretty good prediction results, and the actual and predicted values are comparable.

Model Performance Summary

- **Performance Matrices**

Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.08214	0.068326	0.990004	0.989972	1.670585

Test Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.080509	0.06902	0.989899	0.989823	1.65104

- The model can explain about 99% of the variation in the data, which is very good.
- The train and test RMSE and MAE are low and comparable. So, our model is not suffering from overfitting.
- The MAPE on the test set suggests we can predict within 1.65% of the used price.
- Hence, we can conclude this model is good for prediction as well as inference purposes.

Model Performance Summary

- **Comparison of the initial model created with sklearn and the final statsmodels model**

Training performance comparison:

	Linear Regression sklearn	Linear Regression statsmodels
RMSE	0.081557	0.082140
MAE	0.067535	0.068326
R-squared	0.990146	0.990004
Adj. R-squared	0.989953	0.989972
MAPE	1.650466	1.670585

Test performance comparison:

	Linear Regression sklearn	Linear Regression statsmodels
RMSE	0.080874	0.080509
MAE	0.069081	0.069020
R-squared	0.989808	0.989899
Adj. R-squared	0.989329	0.989823
MAPE	1.650876	1.651040

- The performance of the two models is close to each other.

Conclusion:

- Linear regression model was used to predict the future price of a used phone from the given data.
- The model is showing 99% accuracy in predicting the used phone prices.
- The comparison between the predicted price and actual price also shows that the two are on average quite close.
- This method of predicting the used phone prices will be very helpful for the company to predict the future prices of used phones.
- Phones with a higher newer price will be getting a higher used price in the market.
- The longer the phone has been used the lower the used price.
- These are some of the characteristics that contribute to used price: 'selfie_camera_mp', 'release_year', 'days_used', 'new_price_log', '4g' and '5g']

Business Insights and Recommendations

- Higher the new price, higher will be the used price. The company could latch on to offers that come for new phones and then sell them as used phones for higher prices.

- A good selfie camera in a phone can fetch a good used price.

- If a used phone has only been in the hands of the customer for lesser number of days, the company can sell it for good resale value.

- Phone prices are varying within the data. However, we could state with certainty that phones with a price range of 80 to 120 euros sell in higher numbers. The higher the sales the better would be the profit.

- Android phones are dominating the market as seen in the data. The company could invest more in these to get more profit by selling more of Android phones. However, an Apple or Google phone, even though they are not selling in higher numbers, as their value is high in the market, they could also bring in more profit for each phone sold.

- A good RAM can fetch good sale price for phones. This can be seen from the moderate correlation between the used and new price of phone against RAM.

greatlearning

Power Ahead

Happy Learning !