ReCell Case Study

ReCelland Supervised Learning Course

April 2023

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Appendix

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What and How

Executive Summary

- The goal is to find a dynamic pricing strategy for used and refurbished phones and tablets
- Utilized data to build a linear regression model that will predict the price of used phone/tablet
- Identify factors that influence the price
- Focused on data from
  - A variety of most popular manufacturers
  - Phone attributes
  - Operating System
  - If the phone had 5G/4G abilities
  - New and used prices
  - Correlation between these factors

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Conclusions

Executive Summary

- The Model
  - Does a good job of explaining the variation in the data
  - 84% of the variation is explained
  - Is within 4.4% of the ratings on the test data
  - Is good for prediction as well as inference

- As the new price goes up by one unit, so does the used price by .4348 units, all other variables held constant
- If the ram increases by one unit, the used price goes up by .0208 units, all other variables held constant
- 5g capabilities lowers by one unit, the used price also lowers by .0609 units, all other variables held constant
- 4g capabilities raises by one unit, the used price raises by .0456 units, all other variables held constant
- The older the phone gets the used price drops by .0297 units, all other variables held constant
- Karbonn, Lenovo and Xiaomi have positive correlations with the used price

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Recommendations

Executive Summary

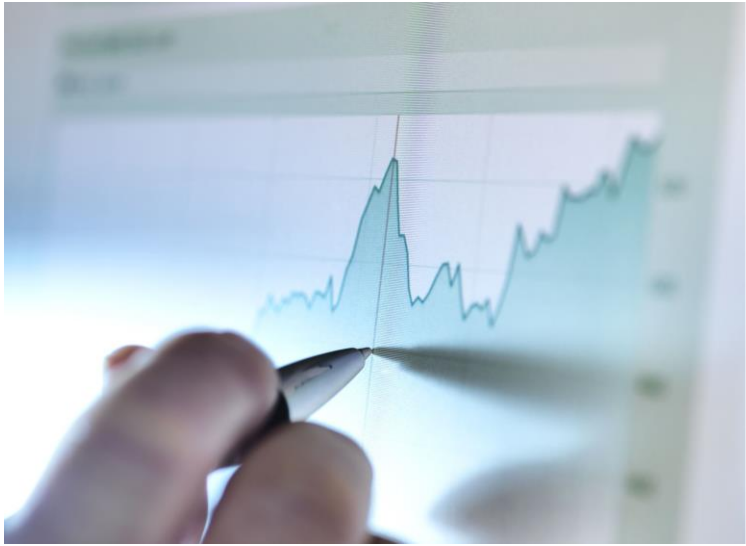
- Focus on buying and reselling phones/tablets that are
  - Newer
  - More expensive when new
  - With 4g capabilities
  - More ram
  - Higher main and selfie camera megapixels
- Concentrate on Karbonn, Lenovo and Xiaomi brands
- Continue to utilize the data and the model to notice trends and changes in customers preferences
- Pay attention to the phone/tablets with 5g abilities as this feature becomes more prominent in the market

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

How can we discover the best price

Business Problem Overview and Solution Approach

- Find the best phone/tablet attributes and characteristics that impact the price to capitalize on trends in the market
- What does the data tell us?
- The Approach
  - Developed the questions to explore data with
  - Perfome data overview
  - Exploratory Data Analysis
  - Data Preprocessing
  - Modelings
  - Check for linear regression assumptions
  - Finalize model summary
  - Developed recomendations



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Overview

- 3454 Rows
- 15 Columns
  - Brand Name (object)
  - Operating System (object)
  - Screen Size (float64)
  - 4G (object)
  - 5G (object)
  - Main Camera MP (float64)
  - Selfie Camera MP (float64)
  - Internal Memory (float64)

- Columns Continued
  - Ram (float64)
  - Battery (float64)
  - Weight (float64)
  - Release Year (int64)
  - Days Used (int64)
  - Normalized Used Price (float64)
  - Normalized New Price (float64)

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Overview

- No Duplicates
- Missing Values
  - Main Camera MP: 179
  - Selfie Camera MP: 2
  - Internal Memory: 4
  - Ram: 4
  - Battery: 6
  - Weight: 7
  - All other columns had all their values

- Float (9), Int64(2), Object(4)

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data – Average, Max, Min

Exploritory Data Analysis

Average

- Screen Size: 13.71
- Main Camera MP: 9.46
- Selfie Camera MP: 6.55
- Internal Memory: 54.57
- Ram: 4.04
- Battery: 3,133.4
- Weight: 182.75
- Release Year: 2016
- Days Used: 674.87
- Normalized Used Price: 4.36
- Normalized New Price: 5.23

Min

- Screen Size: 5.08
- Main Camera MP: 0.08
- Selfie Camera MP: 0
- Internal Memory: 0.01
- Ram: 0.02
- Battery: 500
- Weight: 69
- Release Year: 2013
- Days Used: 91
- Normalized Used Price: 1.54
- Normalized New Price: 2.9

Max

- Screen Size: 30.71
- Main Camera MP: 48
- Selfie Camera MP: 32
- Internal Memory: 1,024
- Ram: 12
- Battery: 9720
- Weight: 855
- Release Year: 2020
- Days Used: 1,094
- Normalized Used Price: 6.62
- Normalized New Price: 7.85

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data – Average, Max, Min

Exploritory Data Analysis

	screen_size	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_year	days_used	normalized_used_price	normalized_new_price
count	3454.000000	3275.000000	3452.000000	3450.000000	3450.000000	3448.000000	3447.000000	3454.000000	3454.000000	3454.000000	3454.000000
mean	13.713115	9.460208	6.554229	54.573099	4.036122	3133.402697	182.751871	2015.965258	674.869716	4.364712	5.233107
std	3.805280	4.815461	6.970372	84.972371	1.365105	1299.682844	88.413228	2.298455	248.580166	0.588914	0.683637
min	5.080000	0.080000	0.000000	0.010000	0.020000	500.000000	69.000000	2013.000000	91.000000	1.536867	2.901422
25%	12.700000	5.000000	2.000000	16.000000	4.000000	2100.000000	142.000000	2014.000000	533.500000	4.033931	4.790342
50%	12.830000	8.000000	5.000000	32.000000	4.000000	3000.000000	160.000000	2015.500000	690.500000	4.405133	5.245892
75%	15.340000	13.000000	8.000000	64.000000	4.000000	4000.000000	185.000000	2018.000000	868.750000	4.755700	5.673718
max	30.710000	48.000000	32.000000	1024.000000	12.000000	9720.000000	855.000000	2020.000000	1094.000000	6.619433	7.847841

Link to Appendix slide on data background check

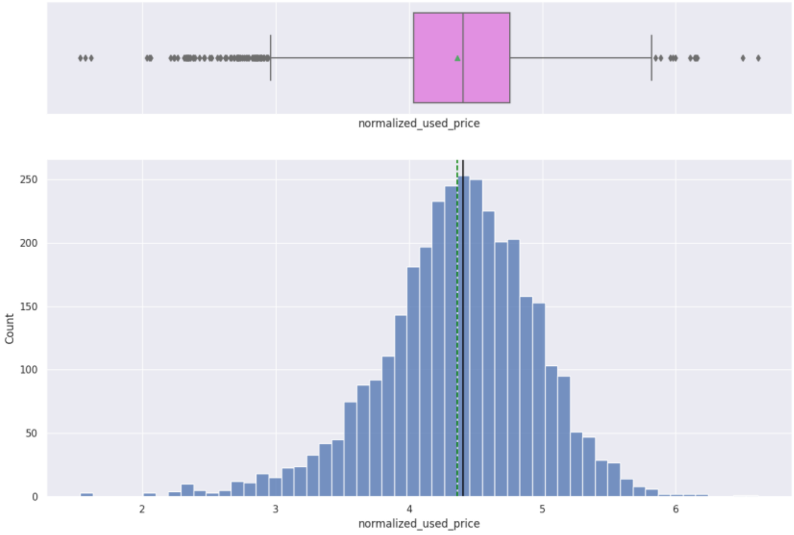
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

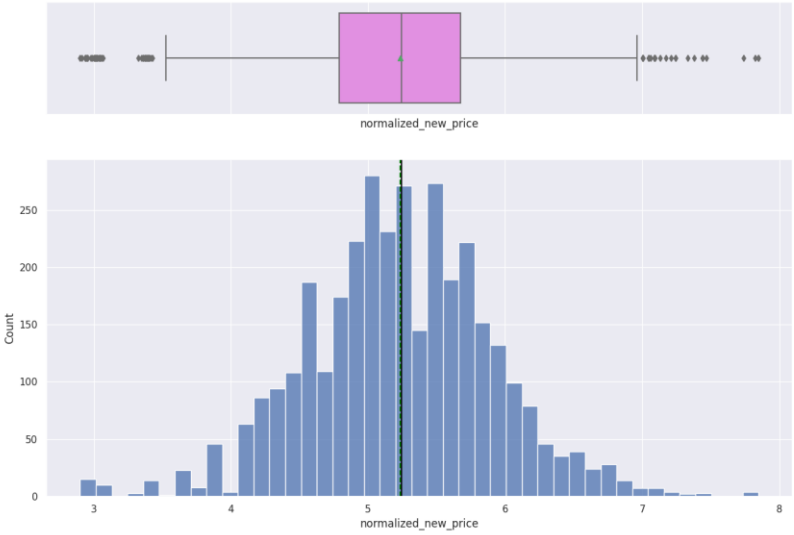
Prices

Exploritory Data Analysis

- The used price is normally distributed
- Used Price mean is 4.36

- The new price is normally distributed
- New Price mean is 5.23





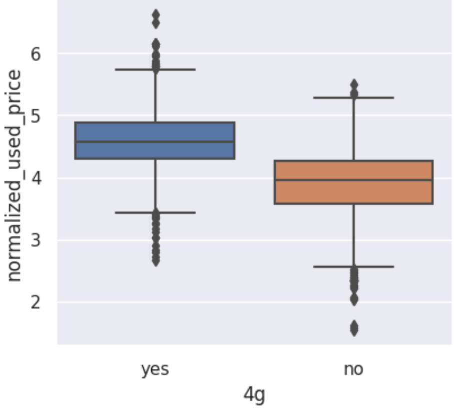
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

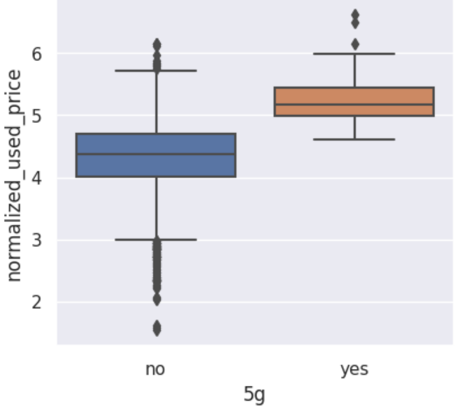
Prices

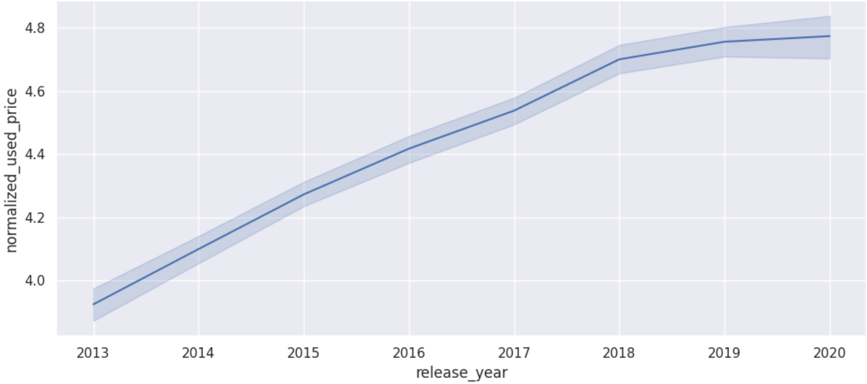
Exploritory Data Analysis

- The used price is greater when the phone has 4g and/or 5g capabilities
- The used price is highest when it has 5g abilities

- Normalized Used Price rises the newer the phone





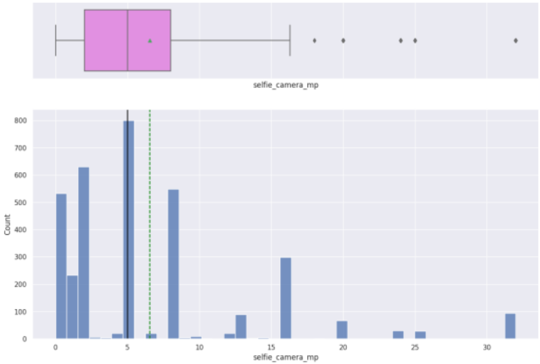


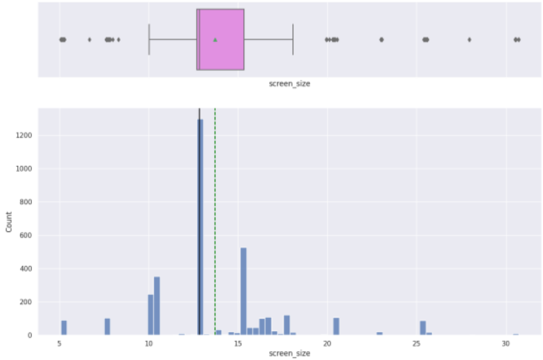
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

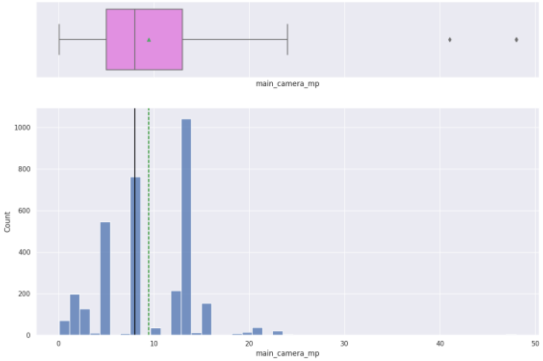
Phone Attributes

Exploritory Data Analysis

- Selfie Camera data is right skewed
- Selfie Camera megapixels mean is 6.6
- 14 cm is the average screen size
- Main Cameras megapixels mean is 9.46





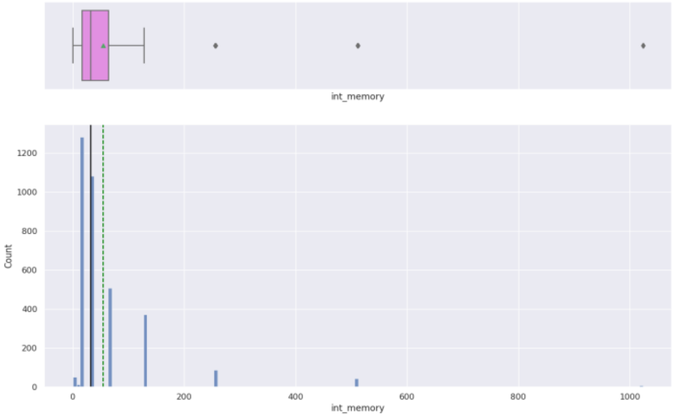


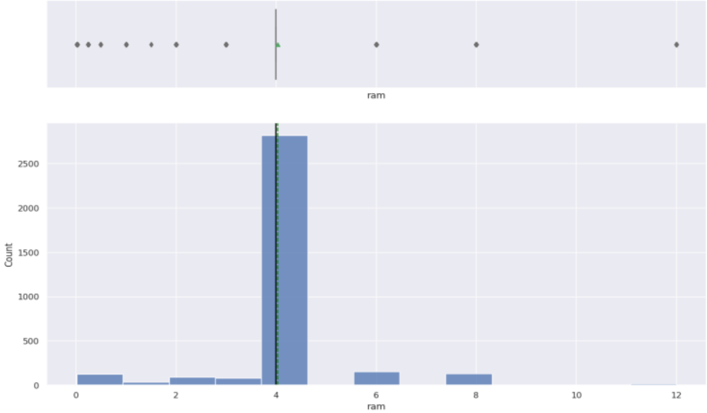
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Phone Attributes

Exploritory Data Analysis

- Internal Memory is right skewed
- Almost all of the Internal memory is between 0 and 54 GB
- Internal Memory mean is 54 GB





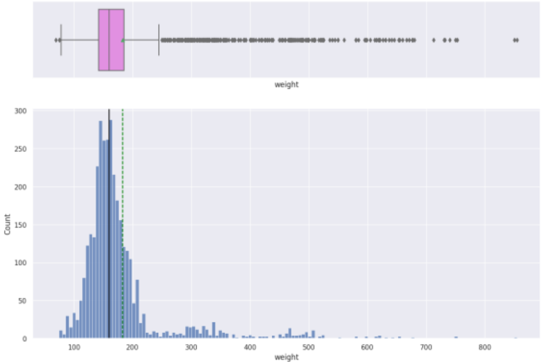
- 4 Gb of Ram is the by far the most popular

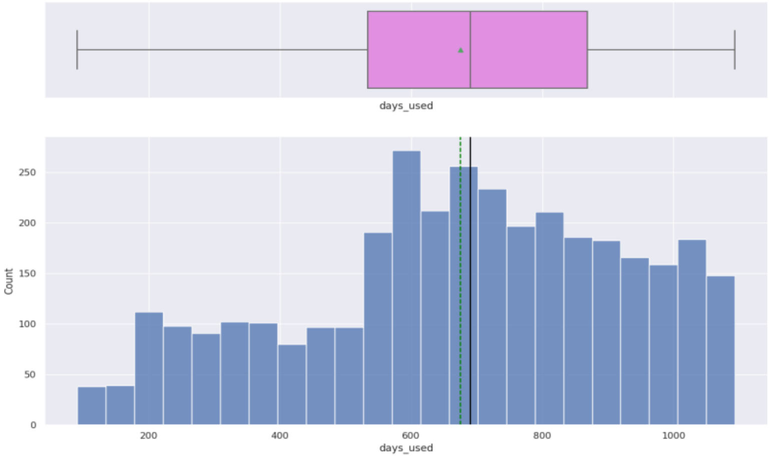
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

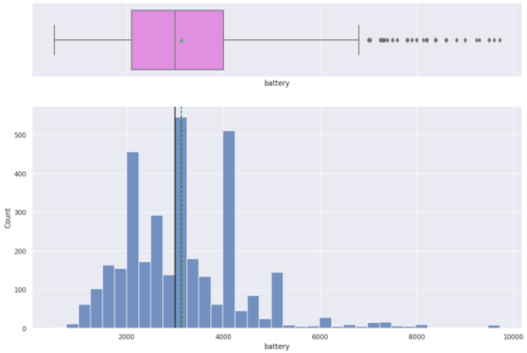
Phone Attributes

Exploritory Data Analysis

- Weight normally distribution with a small right skew
- Weight has a large amount of outliers
- Days used is left skewed
- Battery has a large amount of outliers





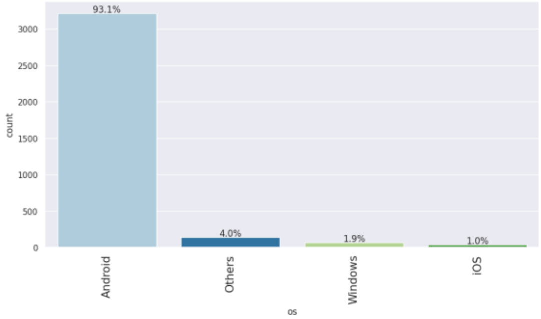


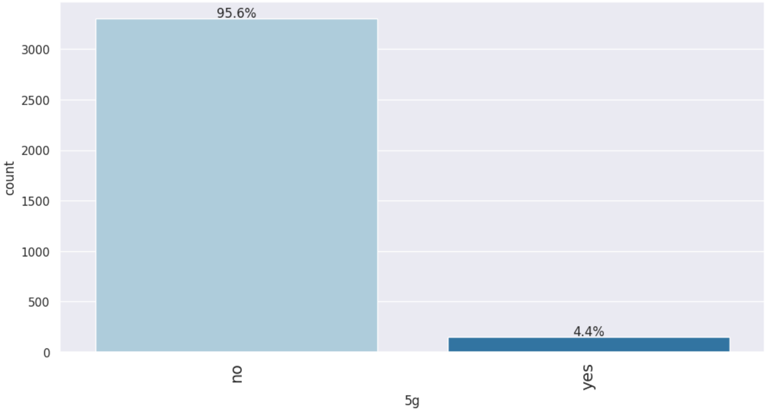
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

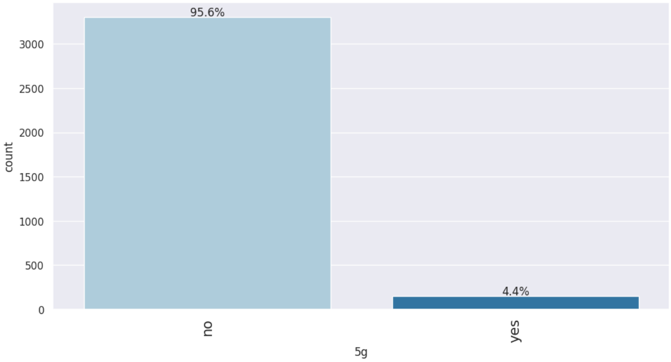
Phone Attributes

Exploritory Data Analysis

- Android was by far the most used Operating System
- More phones had 4g availability than 5g
- 3,3616 phones had 4g capabilities
- 2,535 phones had 5g capabilities





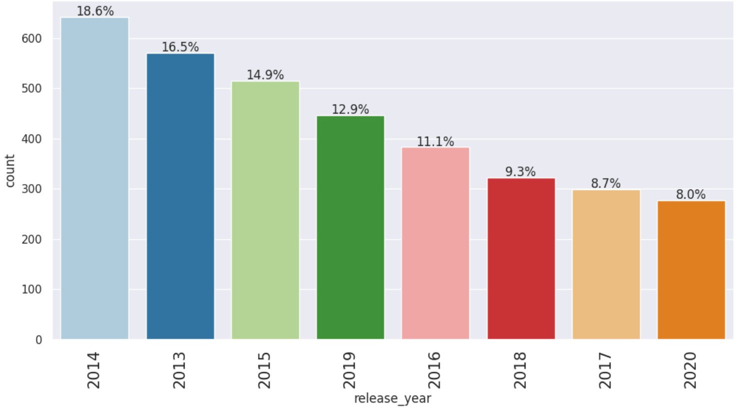


Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

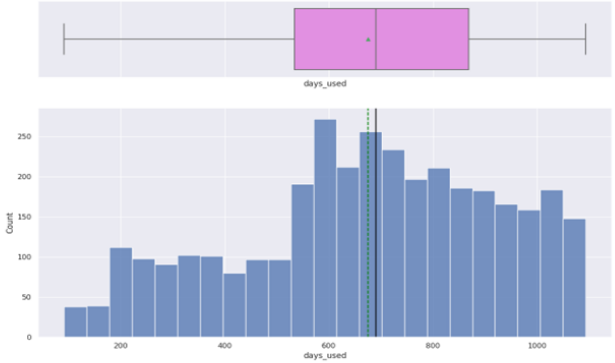
Phone Attributes

Exploritory Data Analysis

- 2014, 13 and 15 make up 50% of the phones/tablets



- The mean count of Days Used was 675 days
- The distribution of Days Used was left skewed

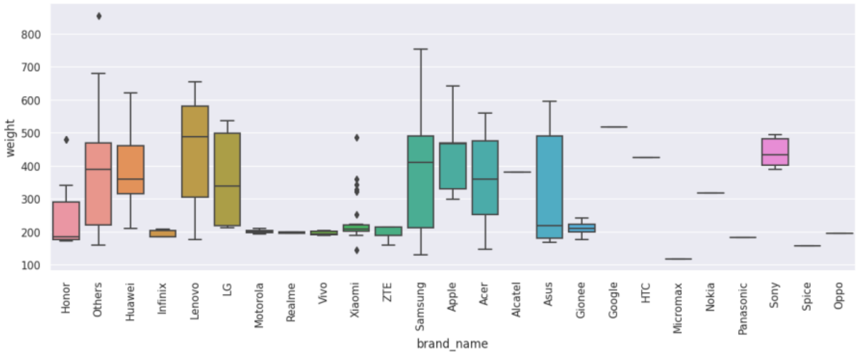


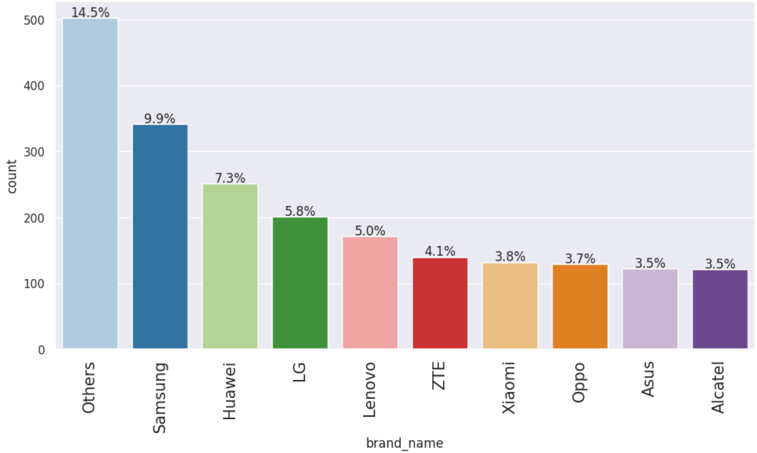
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

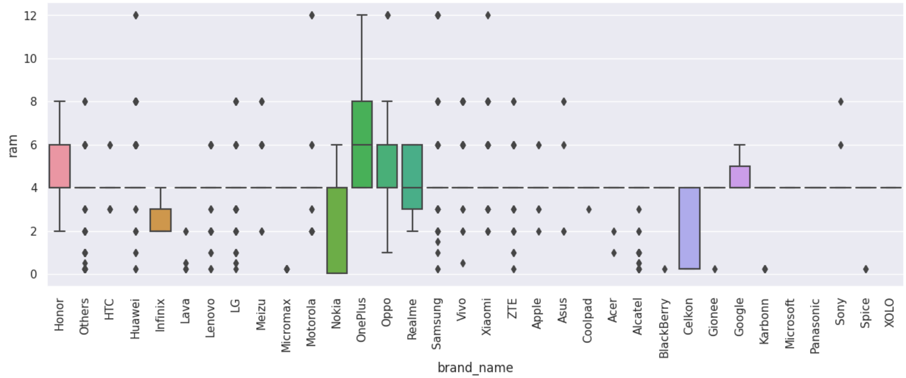
Brands

Exploritory Data Analysis

- Others is the biggest brand represented
- Nokia, Honor, Infinix, OnePlus, Realme, Celkon and Google did not have Ram outliers
- Majority of the brands did not have weight outliers







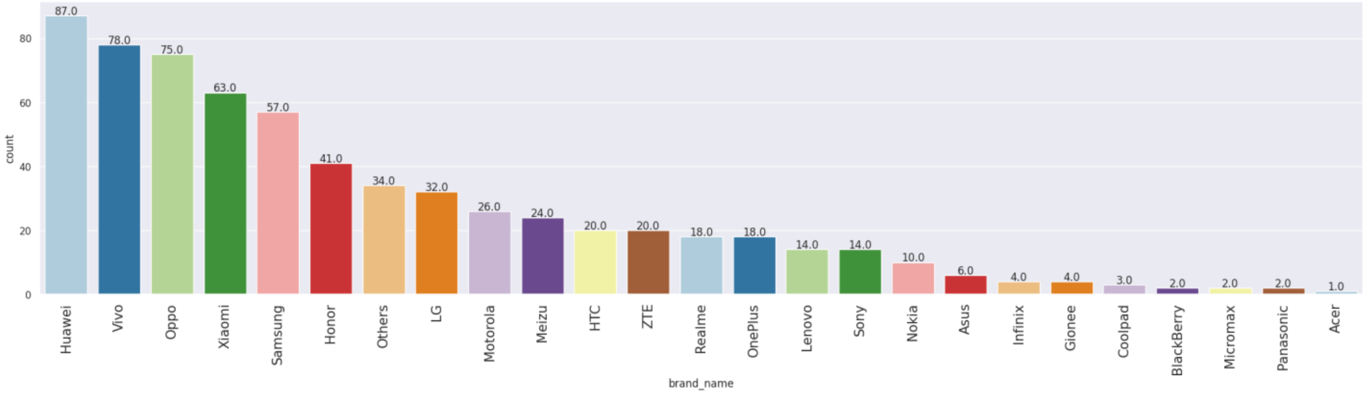
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Brands – Selfie Camera

Exploritory Data Analysis

- Huawei, Vivo, Oppo, Xiaomi, Samsung are the brands with the most selfie cameras with megapixels greater than eight
- Acer, Panasonic, Micromax and Blackberry are the brands with the least selfie cameras greater than eight megapixels

Selfie Camera MP >8



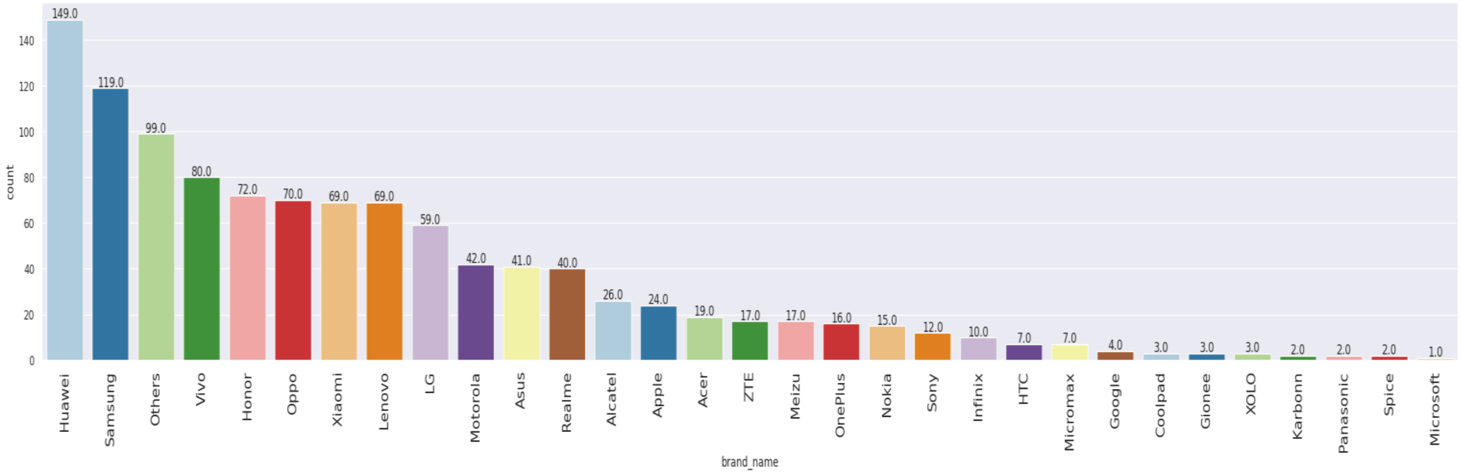
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Brands

Exploritory Data Analysis

- Huawei, Samsung and others have the most large screens
- Microsoft, Spice, and Panasonic have the least amount of large screens

Large Screen Size by Brand

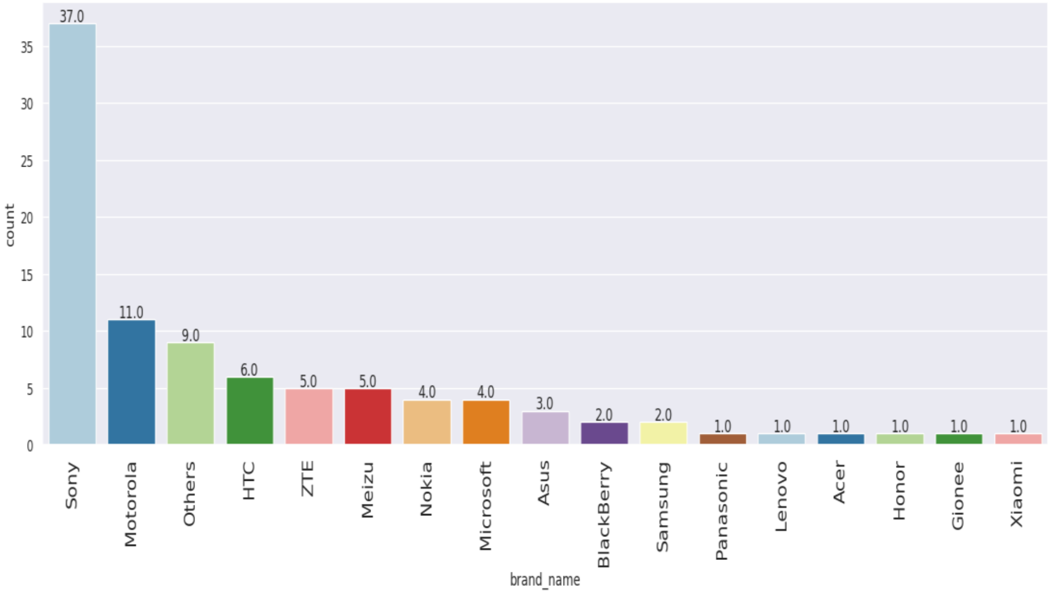


Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Brands

Exploritory Data Analysis

Main Camera MP >16



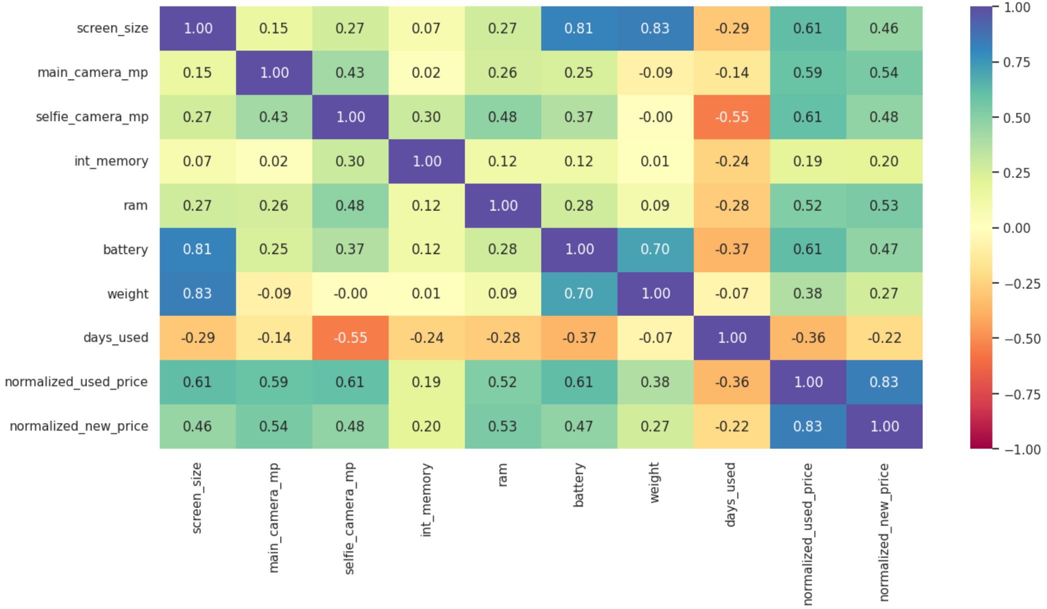
- Sony has by far the most cameras with main cameras above 16 megapixels
- Motorola, Others, HTC, ZTE, Meizu, Nokia and Microsoft are the closest to Sony
- There are less than three of each of the rest of the brands

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Correlation Check

Exploritory Data Analysis

- Correlation >0.5
  - Screen Size and
    - Batter and Weight
  - Batter and Weight
  - New Price and
    - Both Cameras, Ram, Used Price
  - Used Price and
    - Screen Size, Both Cameras, Ram, Battery, New Price
- Correlation <-0.5
  - Days Used and Selfie Camera



	screen_size	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	days_used	normalized_used_price	normalized_new_price
screen_size	1.00	0.15	0.27	0.07	0.27	0.81	0.83	-0.29	0.61	0.46
main_camera_mp	0.15	1.00	0.43	0.02	0.26	0.25	-0.09	-0.14	0.59	0.54
selfie_camera_mp	0.27	0.43	1.00	0.30	0.48	0.37	-0.00	-0.55	0.61	0.48
int_memory	0.07	0.02	0.30	1.00	0.12	0.12	0.01	-0.24	0.19	0.20
ram	0.27	0.26	0.48	0.12	1.00	0.28	0.09	-0.28	0.52	0.53
battery	0.81	0.25	0.37	0.12	0.28	1.00	0.70	-0.37	0.61	0.47
weight	0.83	-0.09	-0.00	0.01	0.09	0.70	1.00	-0.07	0.38	0.27
days_used	-0.29	-0.14	-0.55	-0.24	-0.28	-0.37	-0.07	1.00	-0.36	-0.22
normalized_used_price	0.61	0.59	0.61	0.19	0.52	0.61	0.38	-0.36	1.00	0.83
normalized_new_price	0.46	0.54	0.48	0.20	0.53	0.47	0.27	-0.22	0.83	1.00

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Missing Value Treatment

Data Preprocessing

- No duplicate values
- There were 194 missing values
- Grouped the missing values by Release Year and Brand Name and imputed the column median
- Selfie Camera MP, Battery, and Weight had missing values remaining
  - Group by Brand Name and transformed them with the column median
- There were ten Main Camera MP values still missing
  - Applied the median for the column to those missing values

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Feature Engineering

Data Preprocessing

- Created a new column called years_since_release from the release_year column
  - Dropped the release_year column

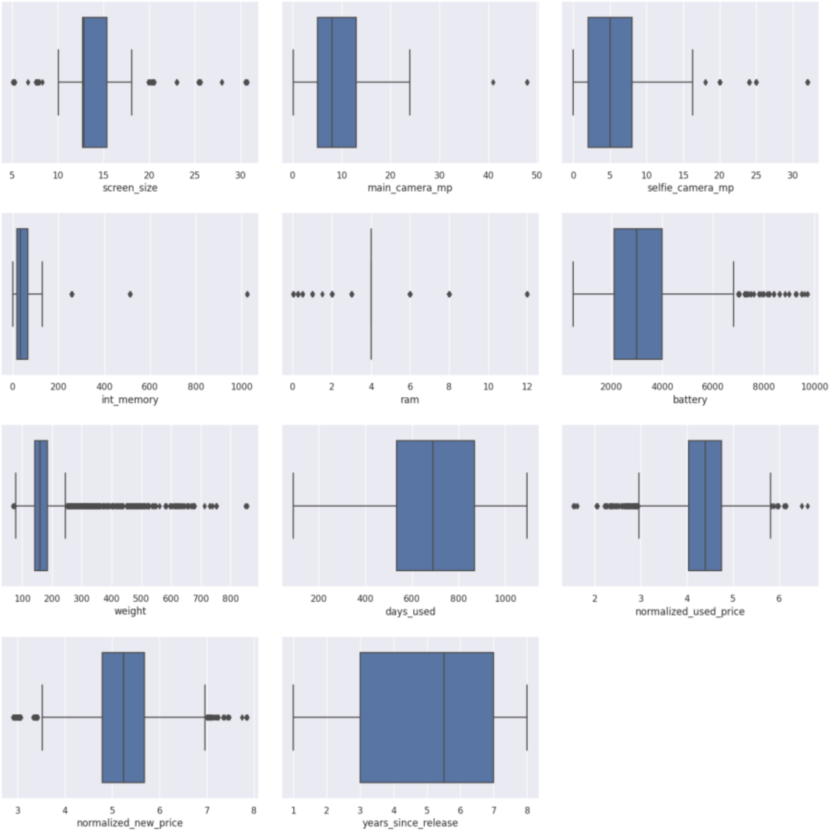
Count	3454
Mean	5.035
STD	2.298
Min	1
25%	3
50%	5.5
75%	7
Max	8

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Outlier Check

Data Preprocessing

- Attributes with 5+ outliers
  - Screen Size
  - Selfie Camera
  - Ram
  - Battery
  - Weight
  - Used Price
  - New Price

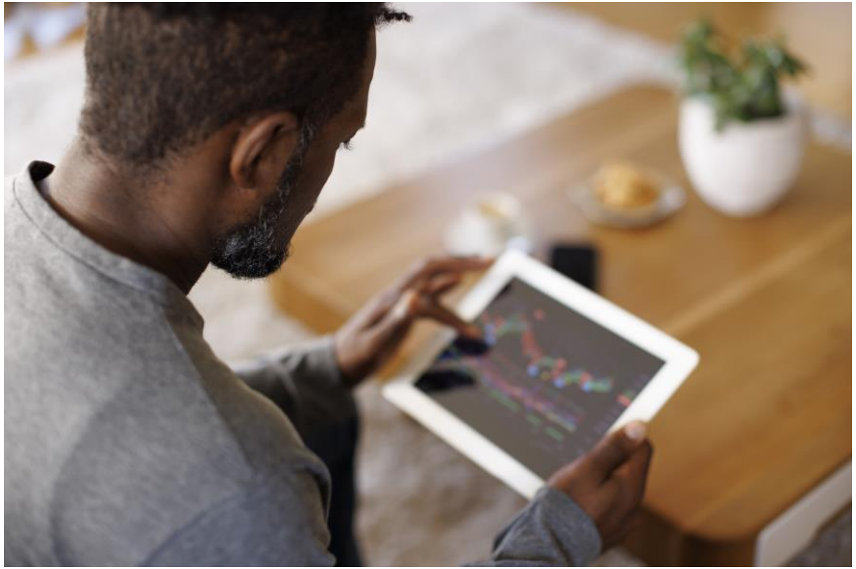


Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Prep for Modeling

Data Preprocessing

- Predict the normalized price of used devices
- Defined dependent and independent variables
  - X dropped normalized
- Added a constant to the data
- Created dummy variables
- Split the data into
  - Train: 2,417 (70%)
  - Test: 1,037 (30%)
- Built a linear regression model with the training data



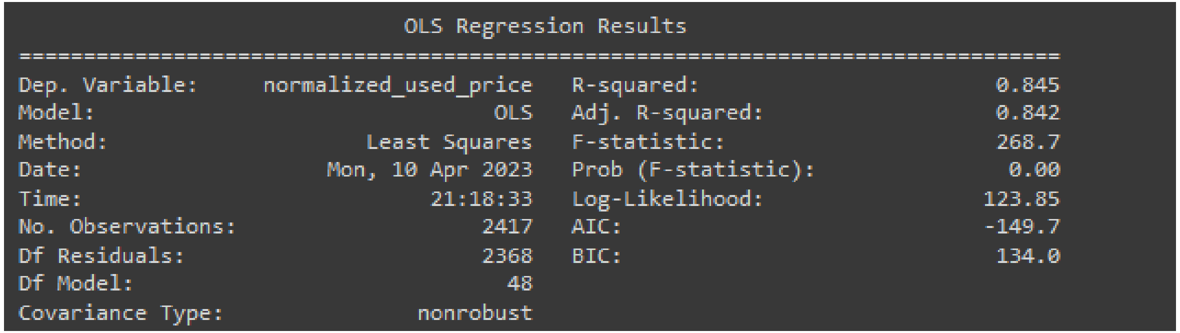
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Linear Regression Modeling

Data Preprocessing

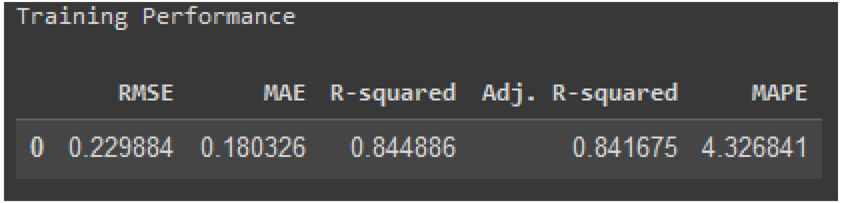
- Built a model using OLS on the training set
- Adj R-Squared is .842 which is good
- Const Coefficient is 1.3156
- Training and Test performance data are acceptably close
- 38 P-Values are over 0.05

Checked Model



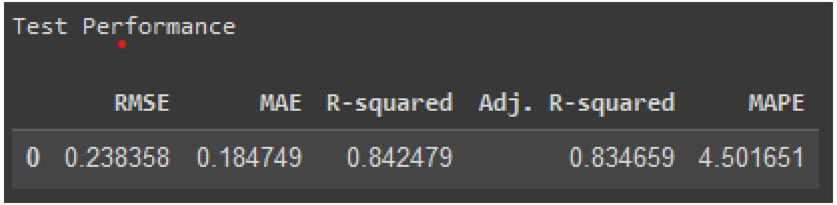
```
                            OLS Regression Results
==============================================================================
Dep. Variable:     normalized_used_price   R-squared:                       0.845
Model:                               OLS   Adj. R-squared:                  0.842
Method:                    Least Squares   F-statistic:                     268.7
Date:                   Mon, 10 Apr 2023   Prob (F-statistic):               0.00
Time:                           21:18:33   Log-Likelihood:                 123.85
No. Observations:                   2417   AIC:                            -149.7
Df Residuals:                       2368   BIC:                             134.0
Df Model:                             48
Covariance Type:               nonrobust
```

Checked Training Performance



```
Training Performance

       RMSE       MAE  R-squared  Adj. R-squared      MAPE
0  0.229884  0.180326   0.844886        0.841675  4.326841
```

Checked Testing Performance



```
Test Performance

       RMSE       MAE  R-squared  Adj. R-squared      MAPE
0  0.238358  0.184749   0.842479        0.834659  4.501651
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

What assumptions were checked

Model Assumptions

- Checked for
  - No multicollinearity
  - Linearity of variables
  - Independence of error terms
  - Normality of error terms
  - No heteroscedasticity
- All assumptions were validated

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Multicollinearity Checks

Model Assumptions

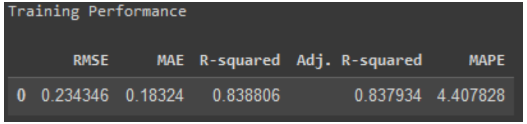
- Checked VIF
  - Had Screen Size(7.67), Weight(6.39), Apple brand(13.05), Huawei brand(5.98), Other brands(9.71), and Samsung brand(7.53) over the VIF threshold (5)
- Dropped Apple brand
  - Still had Screen Size(7.64), Weight(6.39), Huawei brand(5.58), Other brands(9.07) and Samsung brand(6.99) over the VIF threshold
- Dropped Other brands
  - Still had Screen Size(7.57) and Weight(6.36) over the VIF threshold
- Dropped Screen Size
  - Did not have any columns over the VIF threshold

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Dropped High P-Value Variables

Model Assumptions

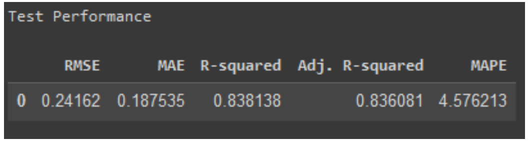
- Dropped high p-value variables
- Training Performance



Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.234346	0.18324	0.838806	0.837934	4.407828

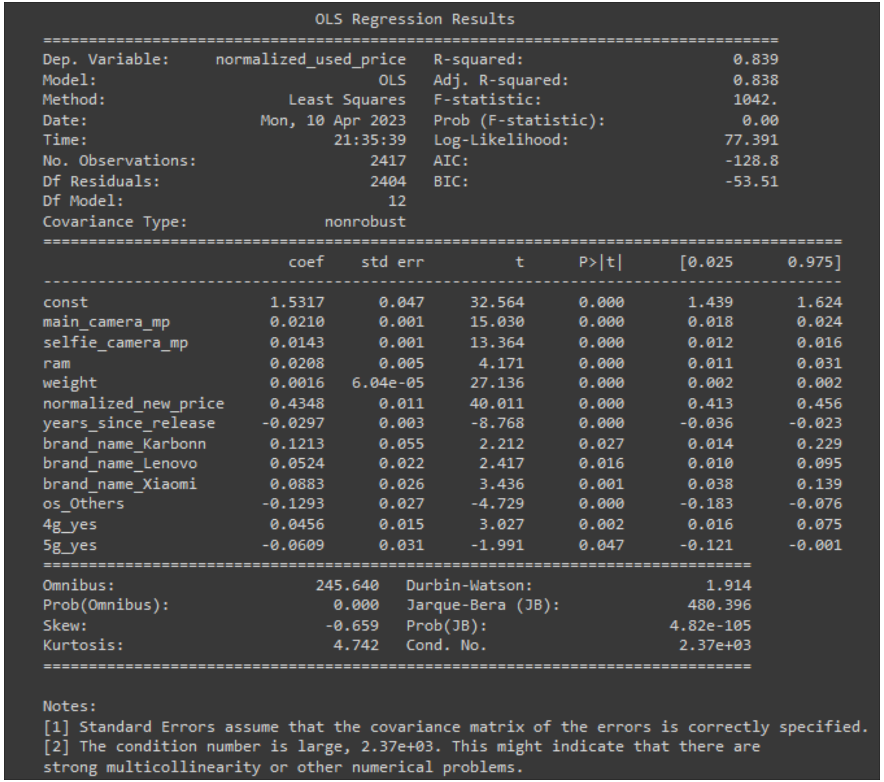
- Test Performance



Test Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.24162	0.187535	0.838138	0.836081	4.576213

- Adj R-Squared is .838 which is good
- Const Coefficient is 1.3156
- Training and Test performance data are acceptably close



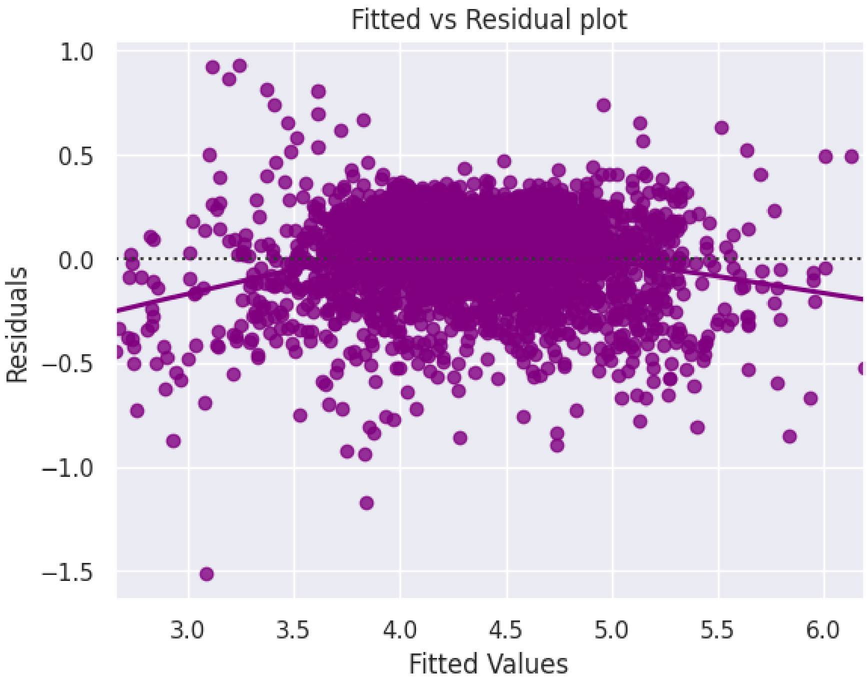
```
                            OLS Regression Results
==============================================================================
Dep. Variable:     normalized_used_price   R-squared:                       0.839
Model:                               OLS   Adj. R-squared:                  0.838
Method:                    Least Squares   F-statistic:                     1042.
Date:                   Mon, 10 Apr 2023   Prob (F-statistic):               0.00
Time:                           21:35:39   Log-Likelihood:                 77.391
No. Observations:                   2417   AIC:                            -128.8
Df Residuals:                       2404   BIC:                            -53.51
Df Model:                             12
Covariance Type:               nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     1.5317      0.047     32.564      0.000       1.439       1.624
main_camera_mp            0.0210      0.001     15.030      0.000       0.018       0.024
selfie_camera_mp          0.0143      0.001     13.364      0.000       0.012       0.016
ram                       0.0208      0.005      4.171      0.000       0.011       0.031
weight                    0.0016   6.04e-05     27.136      0.000       0.002       0.002
normalized_new_price      0.4348      0.011     40.011      0.000       0.413       0.456
years_since_release      -0.0297      0.003     -8.768      0.000      -0.036      -0.023
brand_name_Karbonn        0.1213      0.055      2.212      0.027       0.014       0.229
brand_name_Lenovo         0.0524      0.022      2.417      0.016       0.010       0.095
brand_name_Xiaomi         0.0883      0.026      3.436      0.001       0.038       0.139
os_Others                -0.1293      0.027     -4.729      0.000      -0.183      -0.076
4g_yes                    0.0456      0.015      3.027      0.002       0.016       0.075
5g_yes                   -0.0609      0.031     -1.991      0.047      -0.121      -0.001
==============================================================================
Omnibus:                      245.640   Durbin-Watson:                   1.914
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              480.396
Skew:                          -0.659   Prob(JB):                    4.82e-105
Kurtosis:                       4.742   Cond. No.                     2.37e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.37e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Tested for Linearity and Independence

Model Assumptions

- No Pattern
- Model is linear
- Residuals are independent

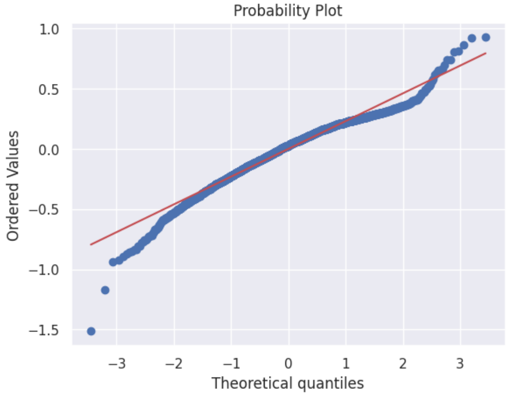


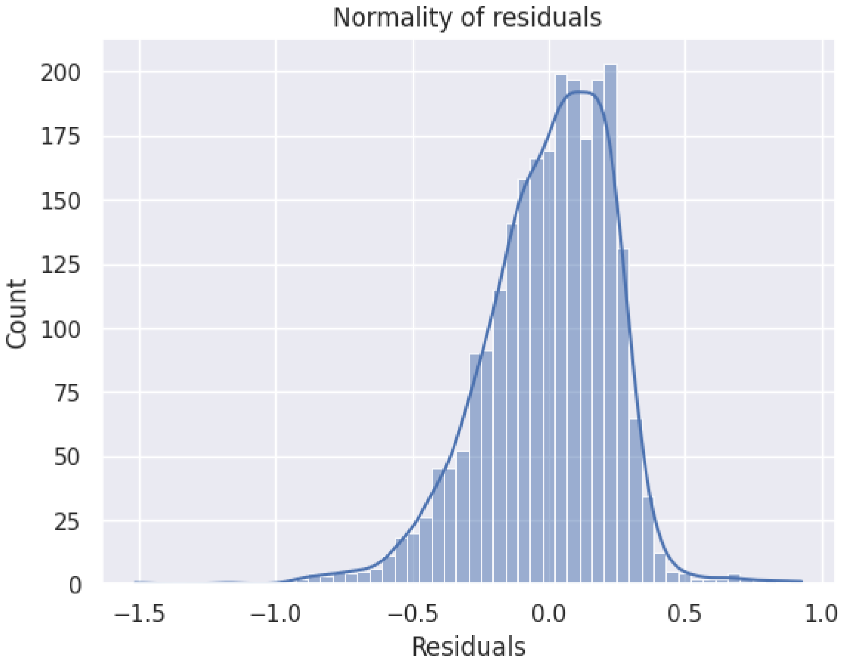
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Tested for Normality

Model Assumptions

- Data is normally distributed
- Shapiro Results
  - 0.968
- It is greater than 0.05 = Normal Distribution





Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Tested for Homoscedasticity

Model Assumptions

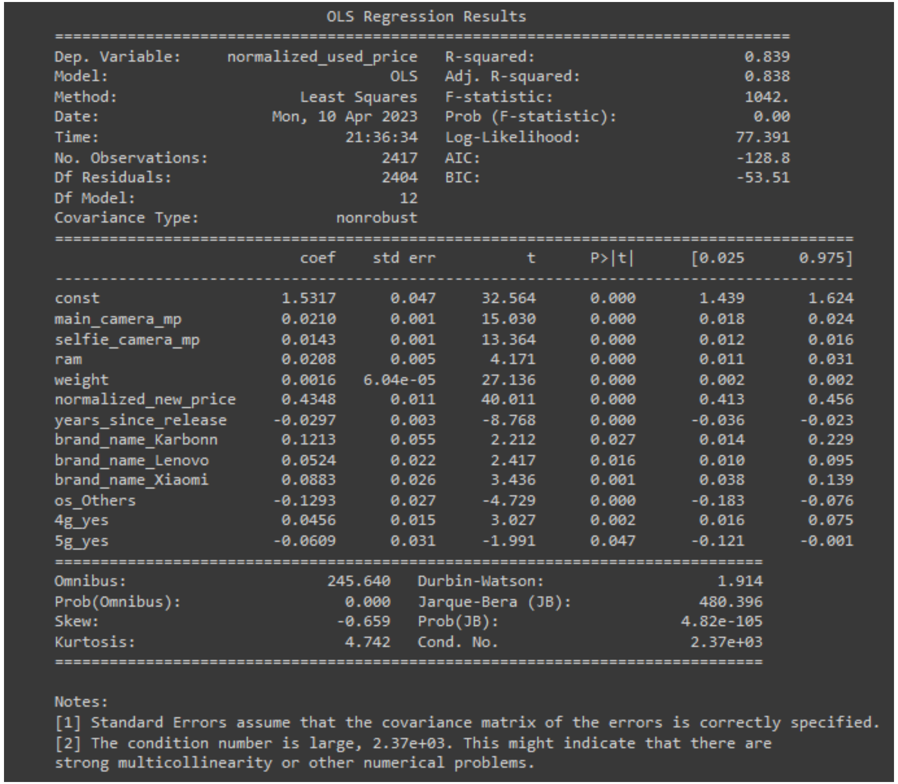
- P-Value of 0.411
- Greater than 0.05
- Residuals are homoscedastic

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

OLS Regression Results

Model Performance Summary

- The r-squared and adjusted r-squared reasonably close
- 84% of the data is explained
- There are no P-Values that are over .05
- Used the least squares method



```
                              OLS Regression Results
==================================================================================
Dep. Variable:     normalized_used_price   R-squared:                        0.839
Model:                               OLS   Adj. R-squared:                   0.838
Method:                    Least Squares   F-statistic:                      1042.
Date:                   Mon, 10 Apr 2023   Prob (F-statistic):                0.00
Time:                           21:36:34   Log-Likelihood:                  77.391
No. Observations:                   2417   AIC:                             -128.8
Df Residuals:                       2404   BIC:                             -53.51
Df Model:                             12
Covariance Type:               nonrobust
=========================================================================================
                           coef    std err          t      P>|t|      [0.025       0.975]
-----------------------------------------------------------------------------------------
const                    1.5317      0.047     32.564      0.000       1.439        1.624
main_camera_mp           0.0210      0.001     15.030      0.000       0.018        0.024
selfie_camera_mp         0.0143      0.001     13.364      0.000       0.012        0.016
ram                      0.0208      0.005      4.171      0.000       0.011        0.031
weight                   0.0016   6.04e-05     27.136      0.000       0.002        0.002
normalized_new_price     0.4348      0.011     40.011      0.000       0.413        0.456
years_since_release     -0.0297      0.003     -8.768      0.000      -0.036       -0.023
brand_name_Karbonn       0.1213      0.055      2.212      0.027       0.014        0.229
brand_name_Lenovo        0.0524      0.022      2.417      0.016       0.010        0.095
brand_name_Xiaomi        0.0883      0.026      3.436      0.001       0.038        0.139
os_Others               -0.1293      0.027     -4.729      0.000      -0.183       -0.076
4g_yes                   0.0456      0.015      3.027      0.002       0.016        0.075
5g_yes                  -0.0609      0.031     -1.991      0.047      -0.121       -0.001
==============================================================================
Omnibus:                      245.640   Durbin-Watson:                   1.914
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              480.396
Skew:                          -0.659   Prob(JB):                    4.82e-105
Kurtosis:                       4.742   Cond. No.                     2.37e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.37e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

OLS Regression Results

Model Performance Summary

Test Performance

- RMSE: 0
- MAE: 0.188
- R-Squared: 0.838
- Adj. R-Squared: 0.836
- MAPE: 4.576

Training Performance

- RMSE: 0.234
- MAE: 0.183
- R-Squared: 0.839
- Adj. R-Squared: 0.838
- MAPE: 4.408

The training and test performance were well within the acceptance range of 2 from each other.

Because of this, the model is not overfitted and should be used.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Great Learning

Happy Learning !