

RECELL PROJECT

BY: DR. CYNTHIA OFFOHA

AGENDA

- EXECUTIVE SUMMARY
- BUSINESS PROBLEM OVERVIEW & SOLUTION APPROACH
- DATA OVERVIEW
- EDA - UNIVARIATE ANALYSIS & KEY QUESTIONS
- EDA – BIVARIATE ANALYSIS & KEY QUESTIONS
- DATA PREPROCESSING
- DATA PREPARATION FOR MODELING
- MODEL BUILDING – LINEAR REGRESSION
- LINEAR REGRESSION AND MODEL SUMMARY
- RECOMMENDATIONS & CONCLUSIONS

EXECUTIVE SUMMARY

3

- Buying and selling used electronics such as phones and tablets used occur prominently within online marketplace sites
- Over the years, used and refurbished device market has progressed in trend and new IDC (International Data Corporation) predicts used phone market would be valued at \$52.7bn by 2023 with CAGR of 13.6% from 2018-2023
- Refurbished and used devices continue to generate cost-effective alternatives for both consumers and businesses
- Refurbished and used devices are sold with warranties and insured with proof of purchase
- Third party vendors are seen to offer attractive offers to customers for refurbished or used devices
- Thus, it is shown that increasing the longevity of devices through second-hand trade decreases their environmental impact and assist in recycling and reducing waste
- The outbreak from COVID-19 was shown to decreased the amount in spending leading to buying phones and tablets for urgent needs
- It is observed that selfie camera, screen size, and internal memory added value to phones which are very significant leading to (+) coefficients
- More refurbished phones with larger screens and efficient better selfie camera are seen with expensive brands whereas lower number of refurbised phones were observed with cheaper brands
- Moreover, minimal (+) correlation were observed between phone weights and battery

BUSINESS PROBLEM OVERVIEW & SOLUTION APPROACH⁴

- Business Problem
 - The rising potential of this market fuels the need for an ML-based solution to develop a dynamic pricing strategy for used and refurbished devices
 - Recell wants to examine the provided data and develop a linear regression model to predict the cost of used phone/tablet and identify the important influence factors
- Solution Approach
 - To determine phone prices utilizing regression model for strategic pricing to be certain that targets are met
 - To conduct Exploratory Data Analysis in which Univariate and Bivariate Analysis will be analyzed into to perform a linear model to further verify the business problem

DATA OVERVIEW

5

VARIABLE	DESCRIPTIONS
brand_name	Name of manufacturing brand
os	OS on which the device 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 device battery in mAh
weight	Weight of the device in grams
release_year	Year when the device model was released
days_used	Number of days the used/refurbished device has been used
normalized_new_price	Normalized price of a new device of the same model in euros
normalized_used_price	Normalized price of the used/refurbished device in euros

- 3454 rows and 15 columns
 - 3454 entries
- Release_year and days_used are of int64 data type
- 4 object type columns and remaining (11) are floats
- Non-null count -> some columns consist of missing values
 - Such as: selfie_camera_mp, int_memory, ram, battery and weight
- No duplicate rows in dataset
- main_camera_mp column -> max.179 missing values
- 34 various brand names
- 4 different operating systems
- Release_year -> consist of 8 different values
- Thus: brand name consist of many categories while OS consist of fewer categories

DATA OVERVIEW

6

Analyzing Missing Values

- main_camera_mp column
 - Android devices --> missing main camera pixels but different brands
- selfie_camera_mp column
 - Selfie_camera_mp column --> consist of two missing values
- RAM column
 - 4 missing values of Ram
 - All missing values are Nokia brand
- weight column
 - Missing values --> XOLO, Windows, and Android OS devices
- NOTE
 - Android devices consist of most missing values
 - Most of the brand names varies for most missing values

DATA STRUCTURE

	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_year
0	Honor	Android	14.50	yes	no	13.0	5.0	64.0	3.0	3020.0	146.0	2020
1	Honor	Android	17.30	yes	yes	13.0	16.0	128.0	8.0	4300.0	213.0	2020
2	Honor	Android	16.69	yes	yes	13.0	8.0	128.0	8.0	4200.0	213.0	2020
3	Honor	Android	25.50	yes	yes	13.0	8.0	64.0	6.0	7250.0	480.0	2020
4	Honor	Android	15.32	yes	no	13.0	8.0	64.0	3.0	5000.0	185.0	2020

	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_year
3449	Asus	Android	15.34	yes	no	NaN	8.0	64.0	6.0	5000.0	190.0	201
3450	Asus	Android	15.24	yes	no	13.0	8.0	128.0	8.0	4000.0	200.0	201
3451	Alcatel	Android	15.80	yes	no	13.0	5.0	32.0	3.0	4000.0	165.0	202
3452	Alcatel	Android	15.80	yes	no	13.0	5.0	32.0	2.0	4000.0	160.0	202
3453	Alcatel	Android	12.83	yes	no	13.0	5.0	16.0	2.0	4000.0	168.0	202

	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_year
59	Infinix	Android	17.32	yes	no	NaN	8.0	32.0	2.0	6000.0	209.0	2020
60	Infinix	Android	15.39	yes	no	NaN	8.0	64.0	4.0	5000.0	185.0	2020
61	Infinix	Android	15.39	yes	no	NaN	8.0	32.0	2.0	5000.0	185.0	2020
62	Infinix	Android	15.39	yes	no	NaN	16.0	32.0	3.0	4000.0	178.0	2019
63	Infinix	Android	15.29	yes	no	NaN	16.0	32.0	2.0	4000.0	165.0	2019

	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_year
114	Nokia	Others	5.18	no	no	0.3	0.0	0.06	NaN	1020.0	91.3	2020
335	Nokia	Others	5.18	no	no	0.3	0.0	0.10	NaN	1200.0	88.2	2020
2059	Nokia	Others	5.18	no	no	0.3	0.0	0.06	NaN	NaN	82.6	2016
2090	Nokia	Others	7.62	no	no	5.0	0.0	0.06	NaN	1200.0	111.4	2013

DATA STRUCTURE

8

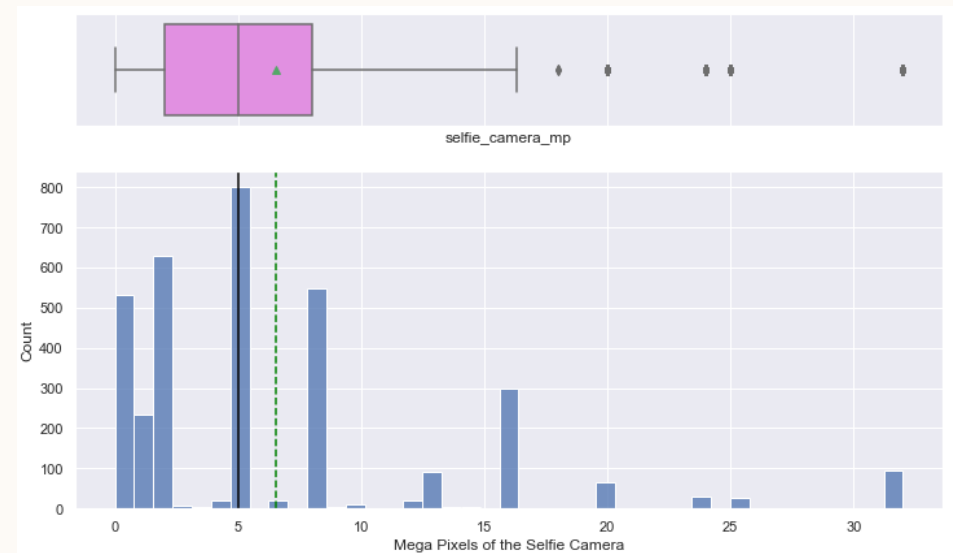
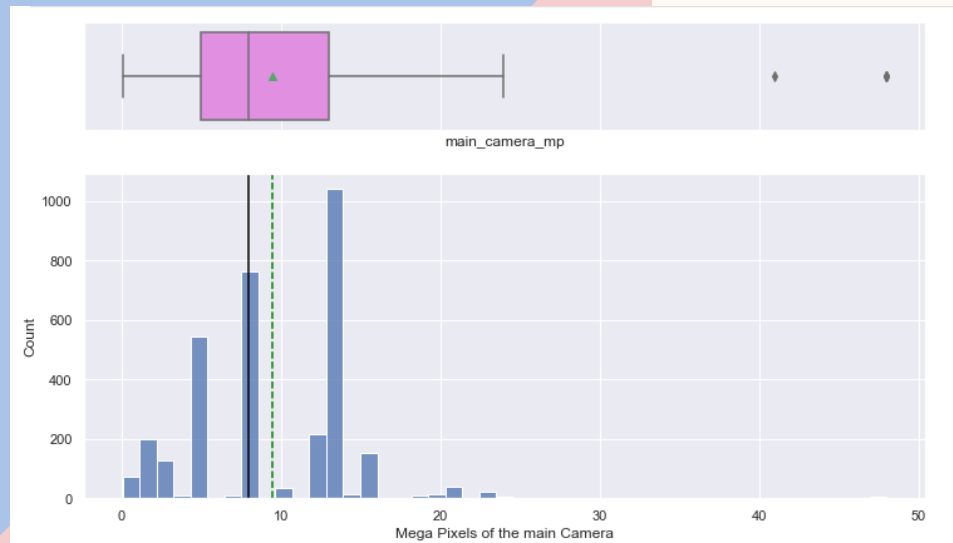
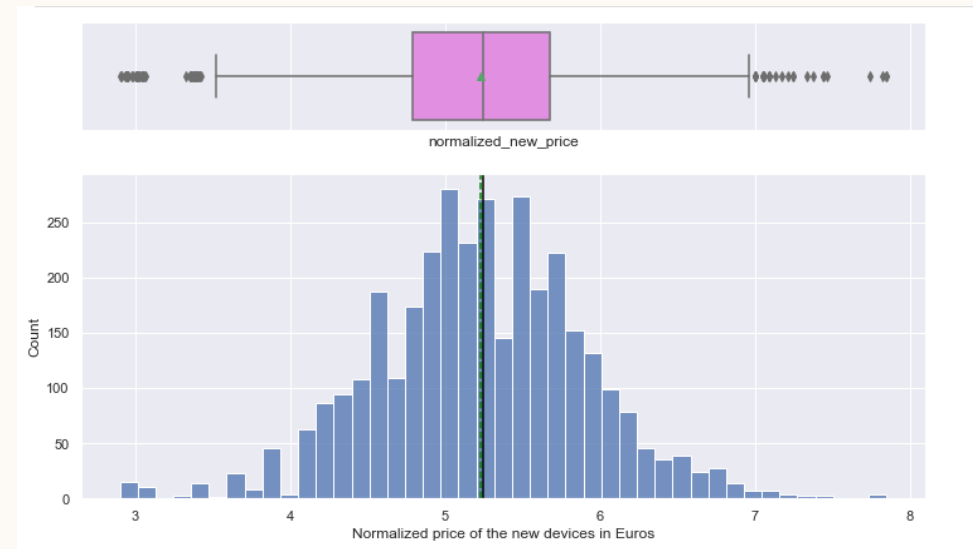
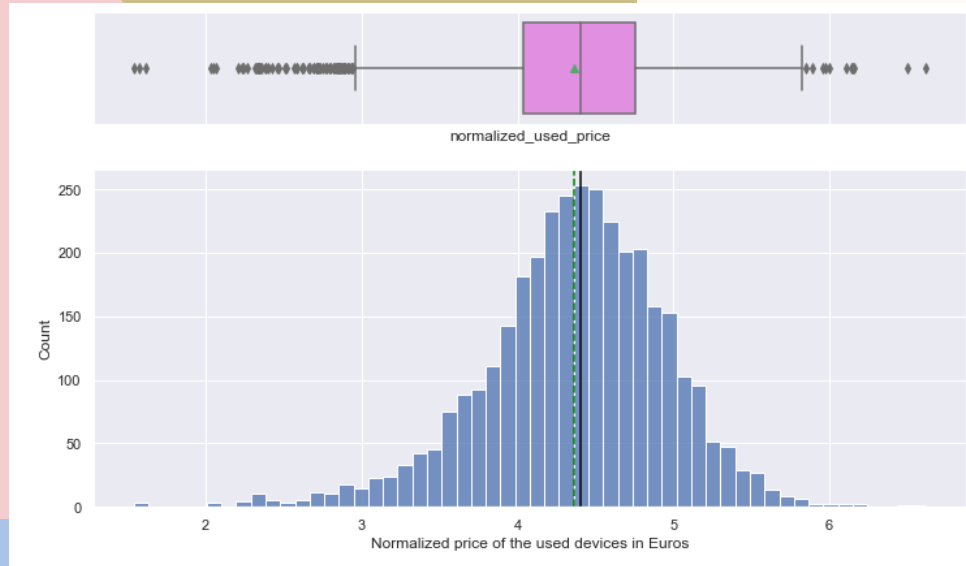
	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_ye
3002	XOLO	Android	12.70	yes	no	13.0	5.0	32.0	4.0	2400.0	NaN	20
3003	XOLO	Android	12.83	yes	no	8.0	5.0	16.0	4.0	3200.0	NaN	20
3004	XOLO	Android	12.70	no	no	8.0	2.0	32.0	4.0	2100.0	NaN	20
3005	XOLO	Android	10.29	no	no	5.0	0.3	32.0	4.0	1800.0	NaN	20
3006	XOLO	Android	12.70	no	no	5.0	0.3	16.0	4.0	2500.0	NaN	20
3007	XOLO	Windows	12.70	no	no	8.0	2.0	32.0	4.0	2200.0	NaN	20
3008	XOLO	Android	12.70	no	no	8.0	5.0	32.0	4.0	2500.0	NaN	20

	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_ye
1829	Meizu	Android	12.83	yes	no	13.0	5.0	16.00	4.00	NaN	145.0	20
1831	Meizu	Android	12.83	yes	no	20.7	5.0	16.00	4.00	NaN	158.0	20
1832	Meizu	Android	13.61	yes	no	20.7	2.0	16.00	4.00	NaN	147.0	20
1962	Microsoft	Windows	25.55	no	no	5.0	3.5	32.00	4.00	NaN	675.9	20
2058	Nokia	Others	5.18	no	no	0.3	0.0	0.06	0.02	NaN	81.0	20

	count	mean	std	min	25%	50%	75%	max
screen_size	3454.0	13.713115	3.805280	5.080000	12.700000	12.830000	15.340000	30.710000
main_camera_mp	3275.0	9.460208	4.815461	0.080000	5.000000	8.000000	13.000000	48.000000
selfie_camera_mp	3452.0	6.554229	6.970372	0.000000	2.000000	5.000000	8.000000	32.000000
int_memory	3450.0	54.573099	84.972371	0.010000	16.000000	32.000000	64.000000	1024.000000
ram	3450.0	4.036122	1.365105	0.020000	4.000000	4.000000	4.000000	12.000000
battery	3448.0	3133.402697	1299.682844	500.000000	2100.000000	3000.000000	4000.000000	9720.000000
weight	3447.0	182.751871	88.413228	69.000000	142.000000	160.000000	185.000000	855.000000
release_year	3454.0	2015.965258	2.298455	2013.000000	2014.000000	2015.500000	2018.000000	2020.000000
days_used	3454.0	674.869716	248.580166	91.000000	533.500000	690.500000	868.750000	1094.000000
normalized_used_price	3454.0	4.364712	0.588914	1.536867	4.033931	4.405133	4.755700	6.619433
normalized_new_price	3454.0	5.233107	0.683637	2.901422	4.790342	5.245892	5.673718	7.847841

EDA: UNIVARIATE ANALYSIS

9



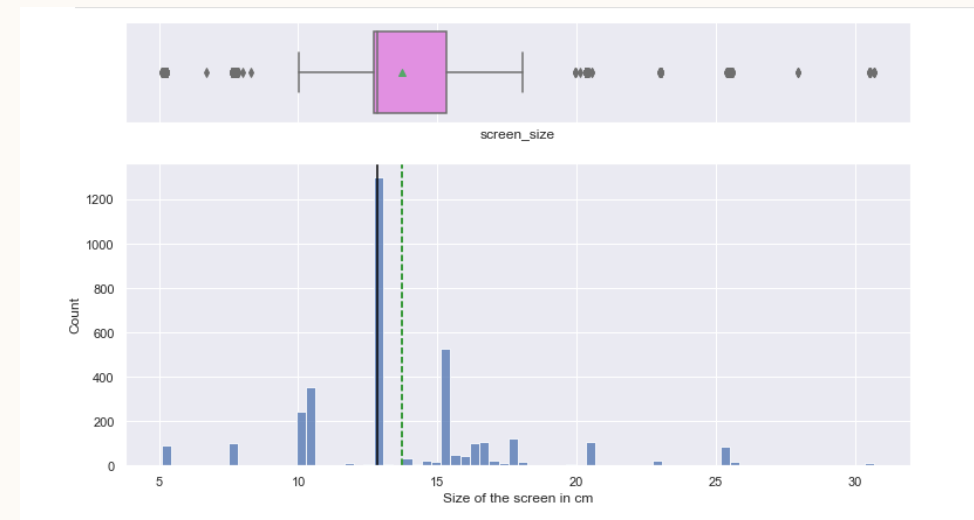
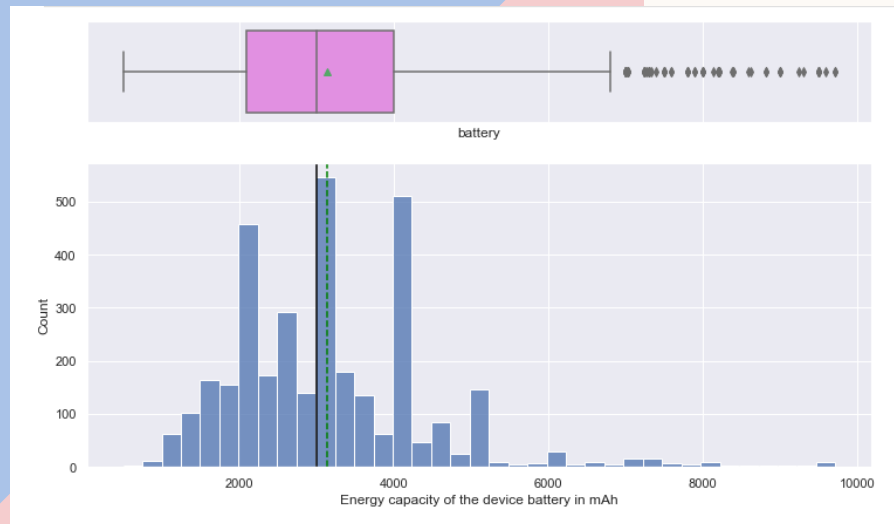
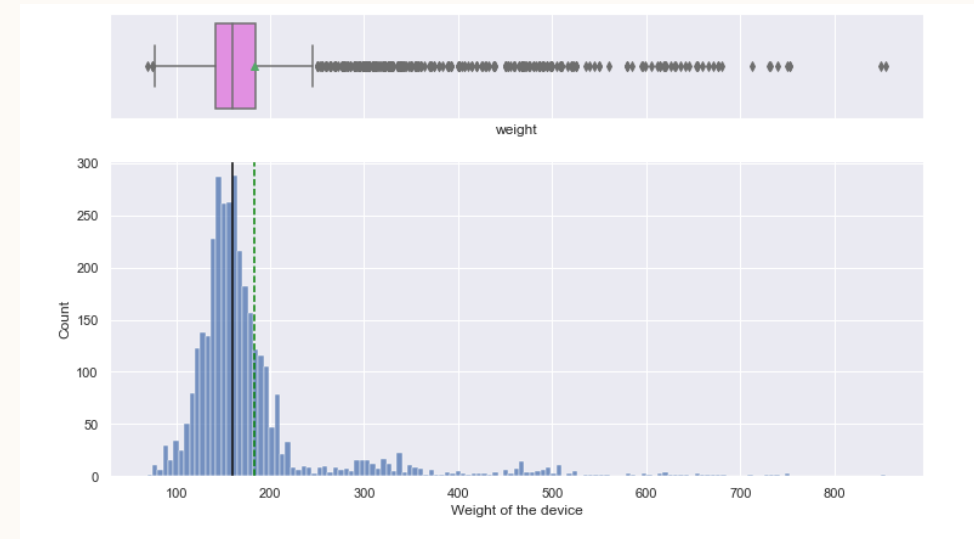
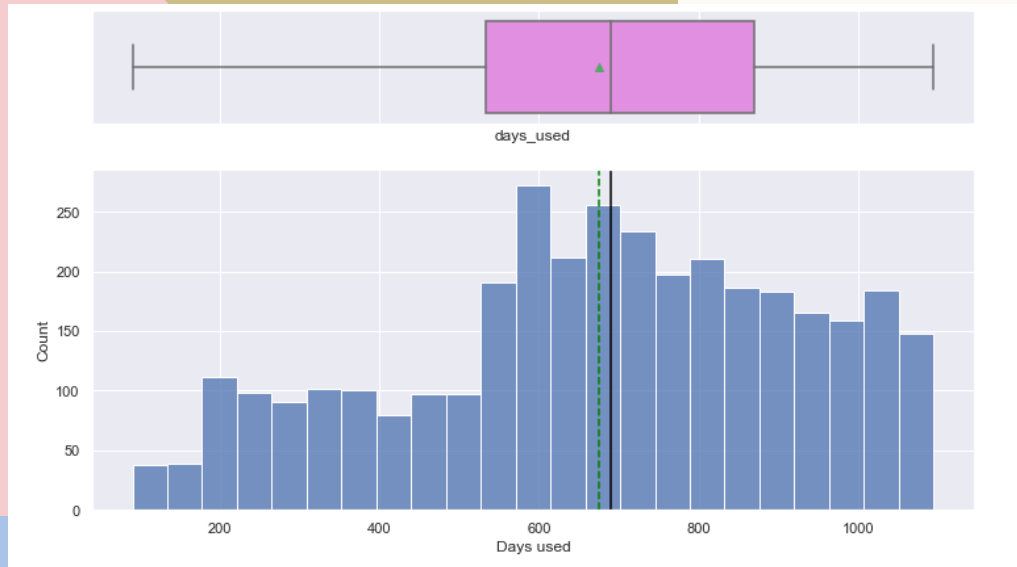
OBSERVATIONS

10

- Normalized Used Price
 - Used phones price showed normal distribution
 - Outlier seen at the box_plot
- Normalized price of New Devices
 - Used phones prices illustrated normal distribution
- Main camera pixels of the devices
 - Plot illustrated normal distribution
- Selfie Camera Resolution
 - Plot illustration of selfie camera distribution is right skewed

EDA: UNIVARIATE ANALYSIS

11



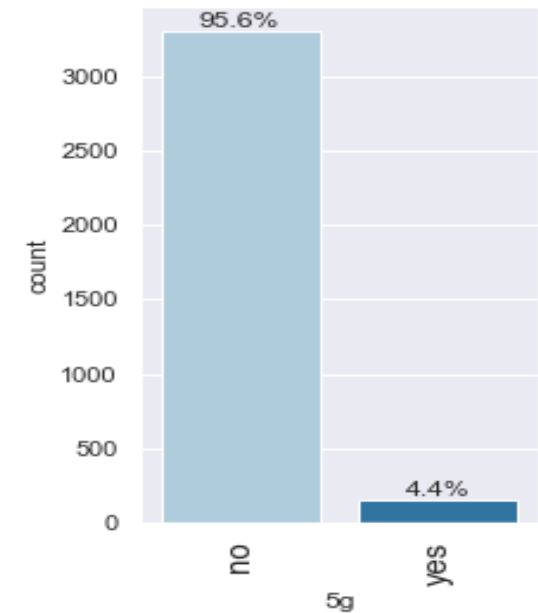
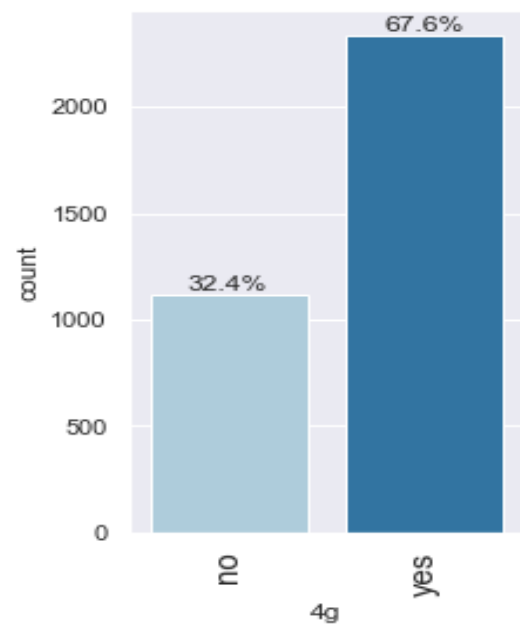
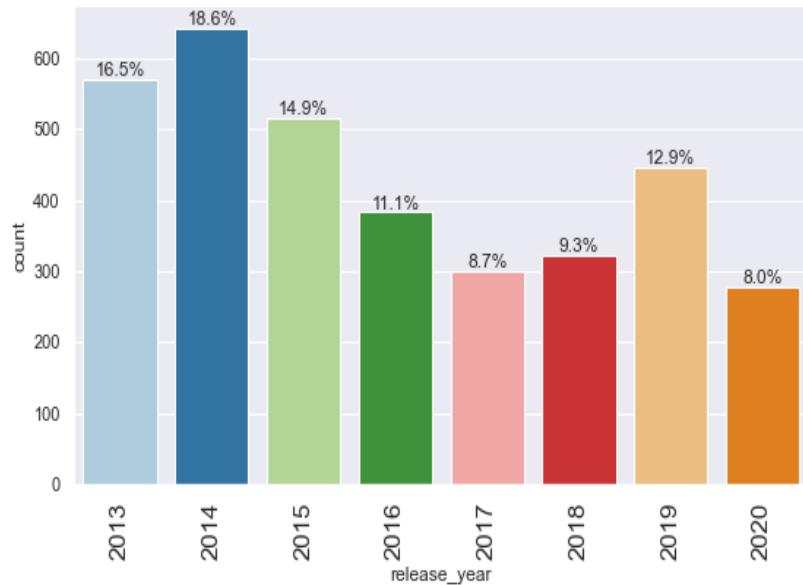
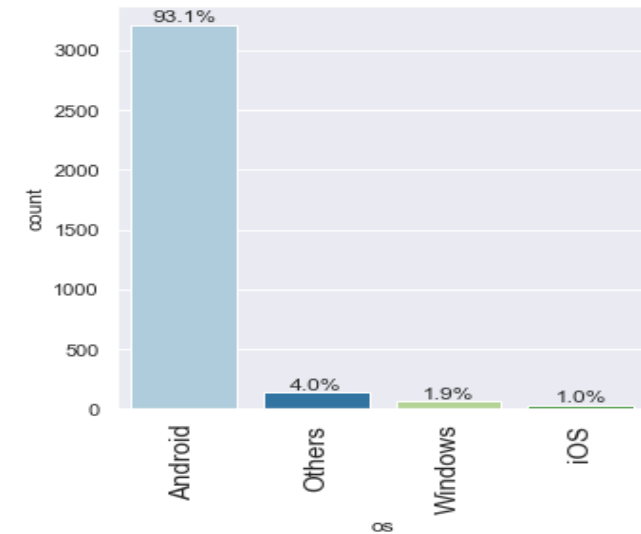
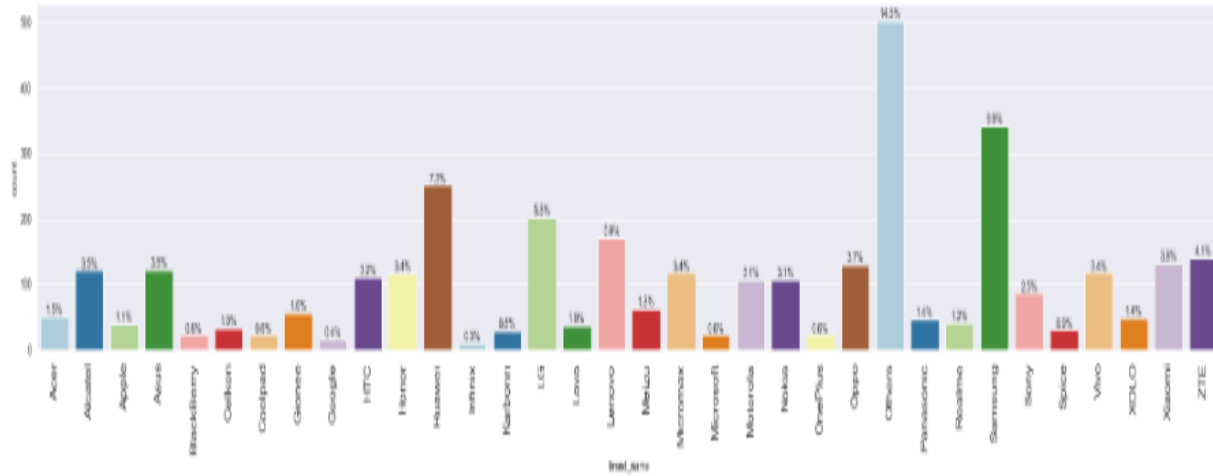
OBSERVATIONS

12

- Days Used
 - Plot showed a normal distribution
- Weight of Devices
 - Plot showed a right skewed demonstration
- Battery
 - Right-skewed was illustrated on the plot
- Device Screen Size
 - The plot showed a normal distribution

EDA: UNIVARIATE ANALYSIS

13



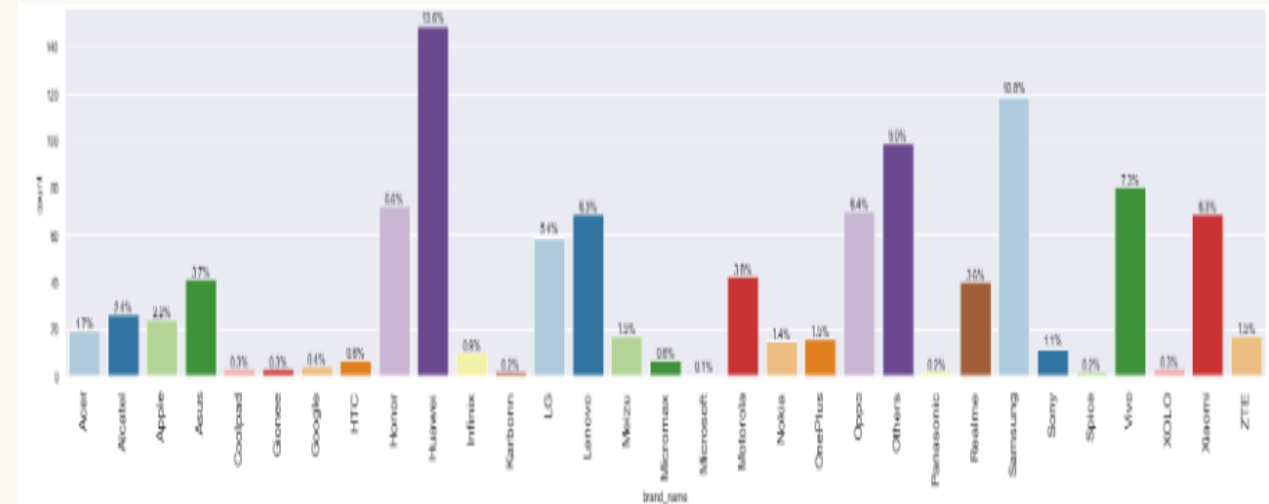
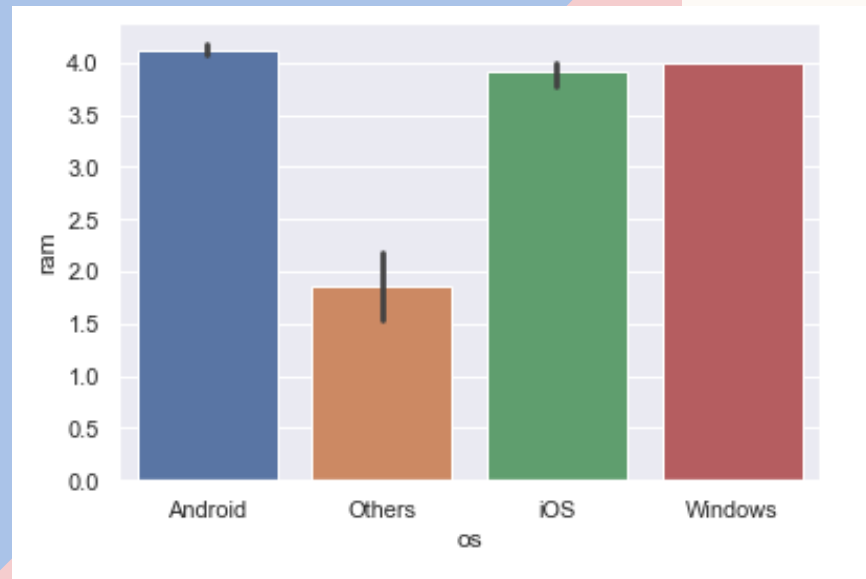
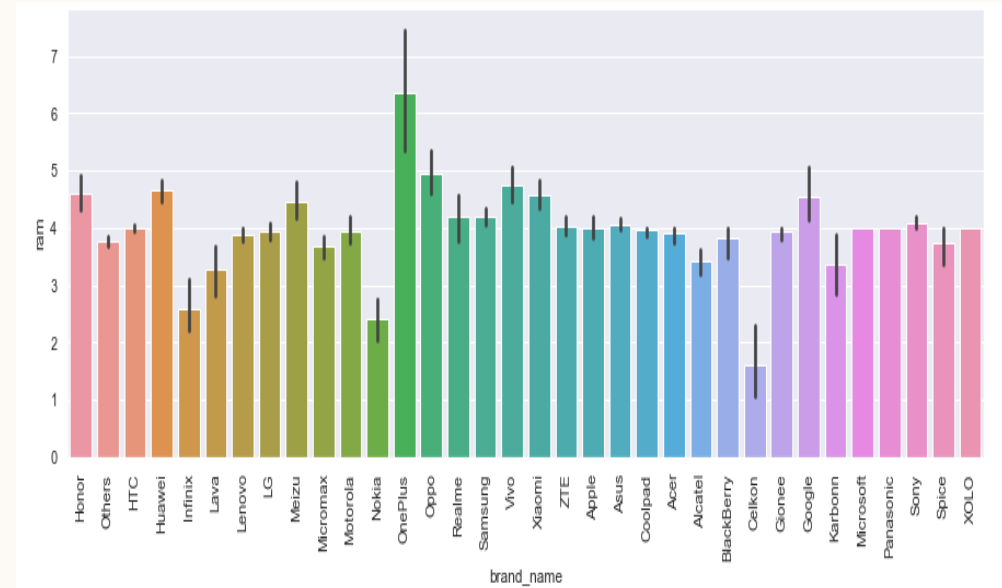
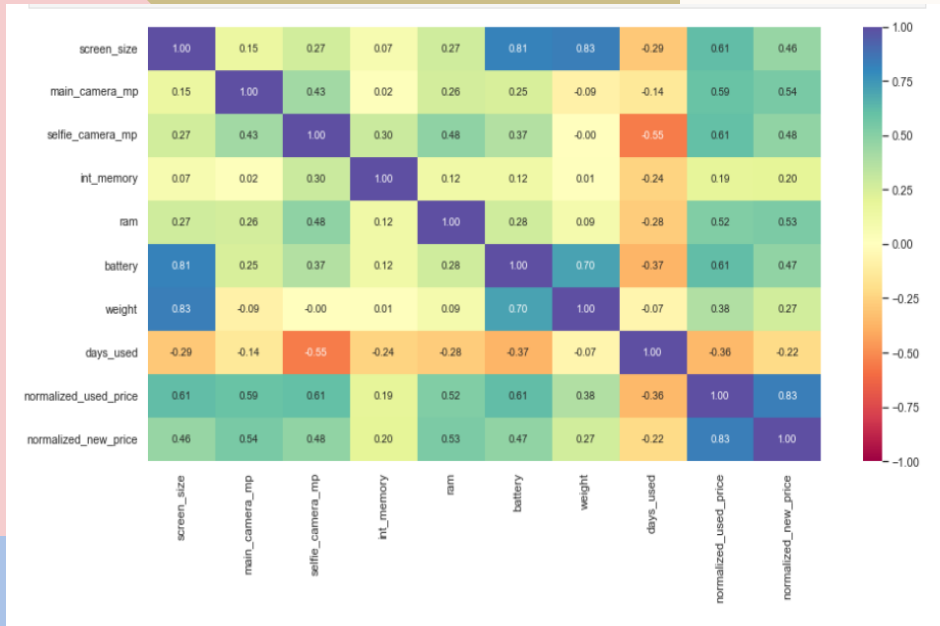
OBSERVATION

14

- Brand Name
 - Samsung consist higher percentage of customers purchase of refurbished phones
- Operating Systems
 - Android devices -> most refurbished with 93.1%
 - IOS devices -> least refurbished with 1.0%
- Release Year
 - 2014 released devices were seen to be the most refurbished ones
- 4G and 5G devices
 - 4G devices were more refurbished compared to 5G devices

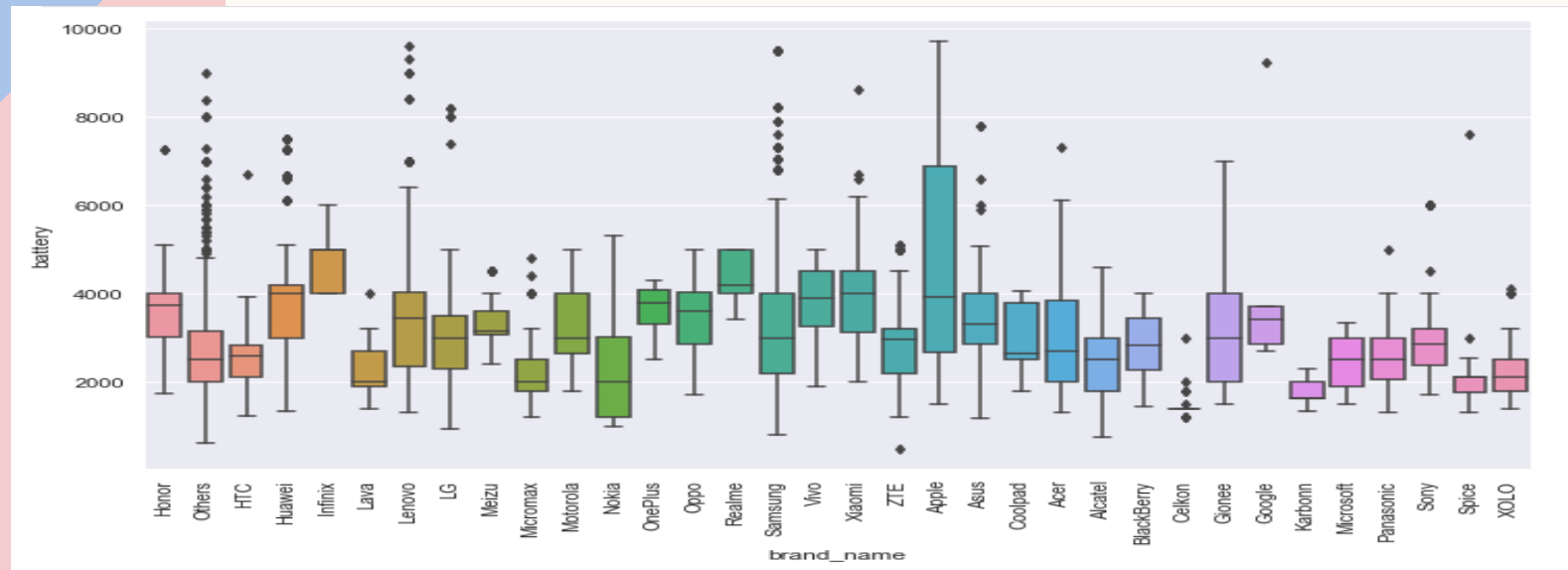
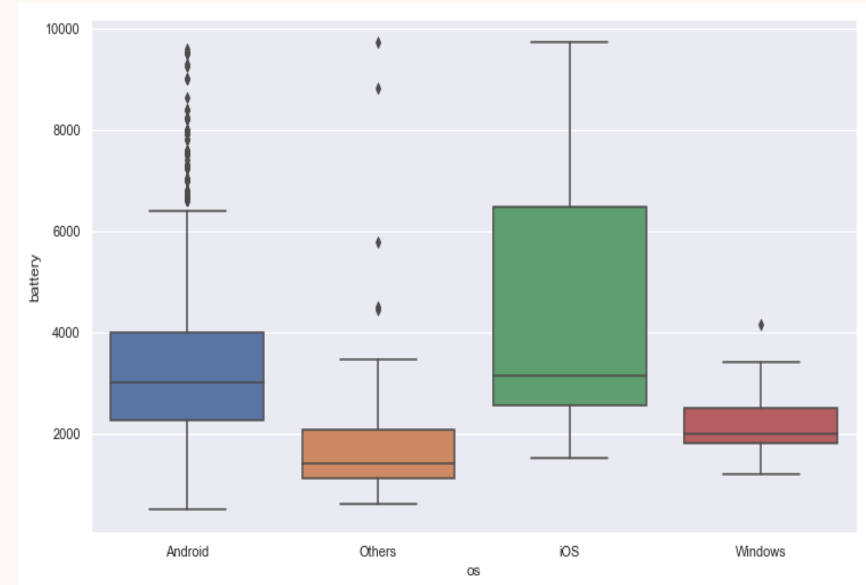
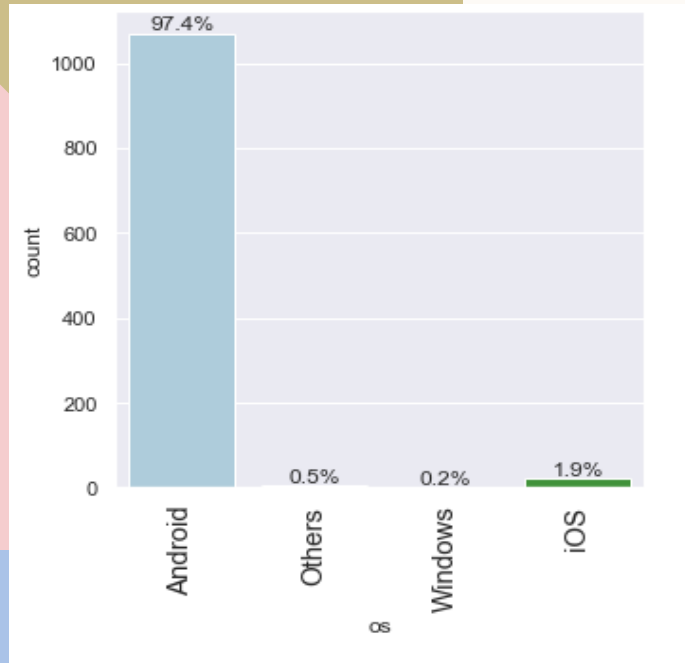
EDA: BIVARIATE ANALYSIS

15



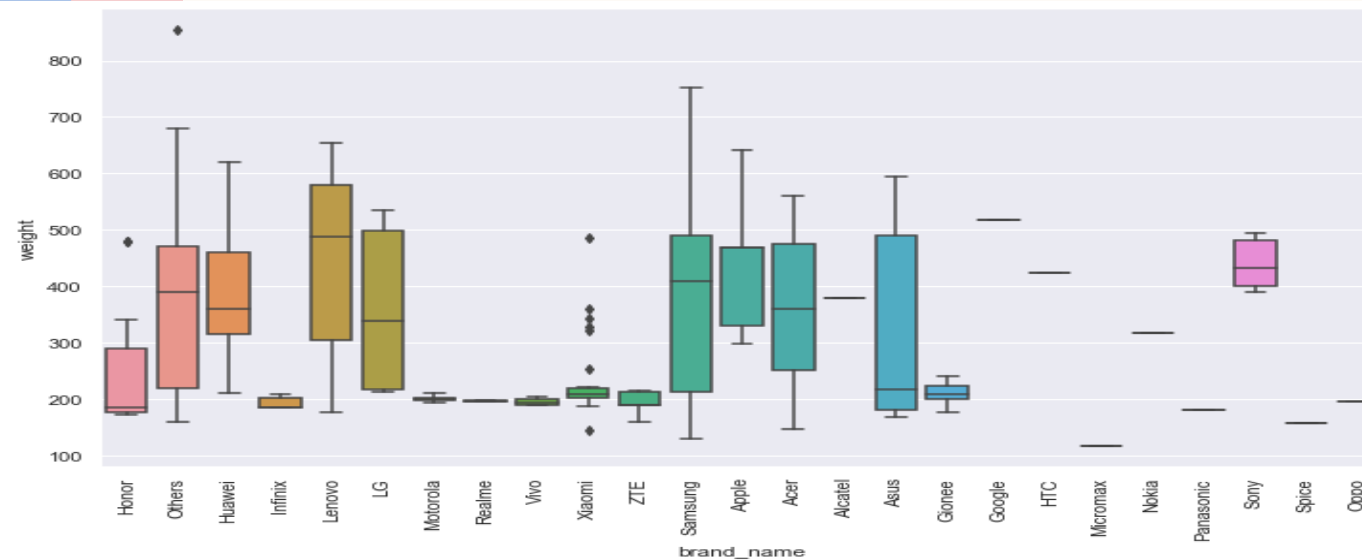
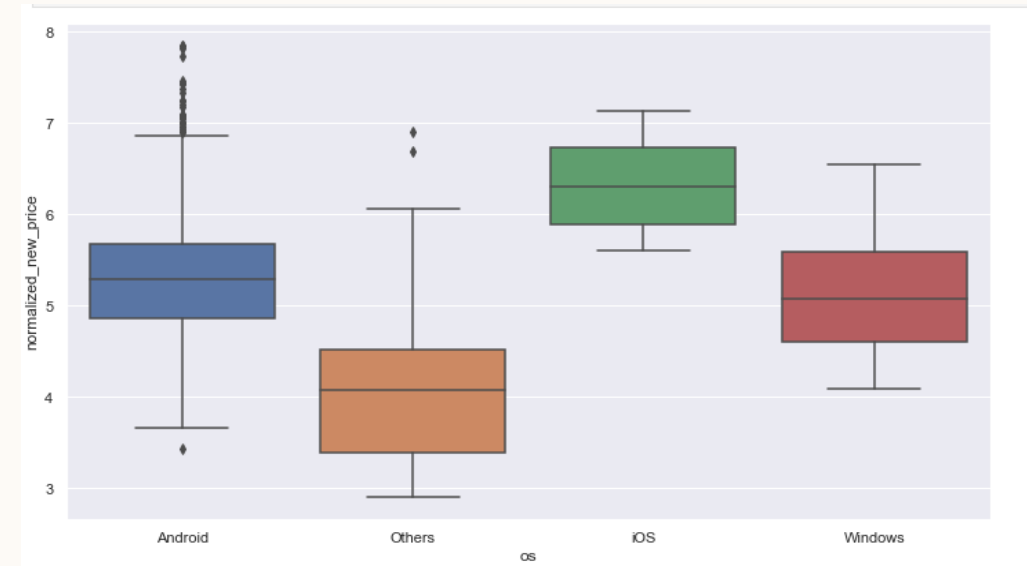
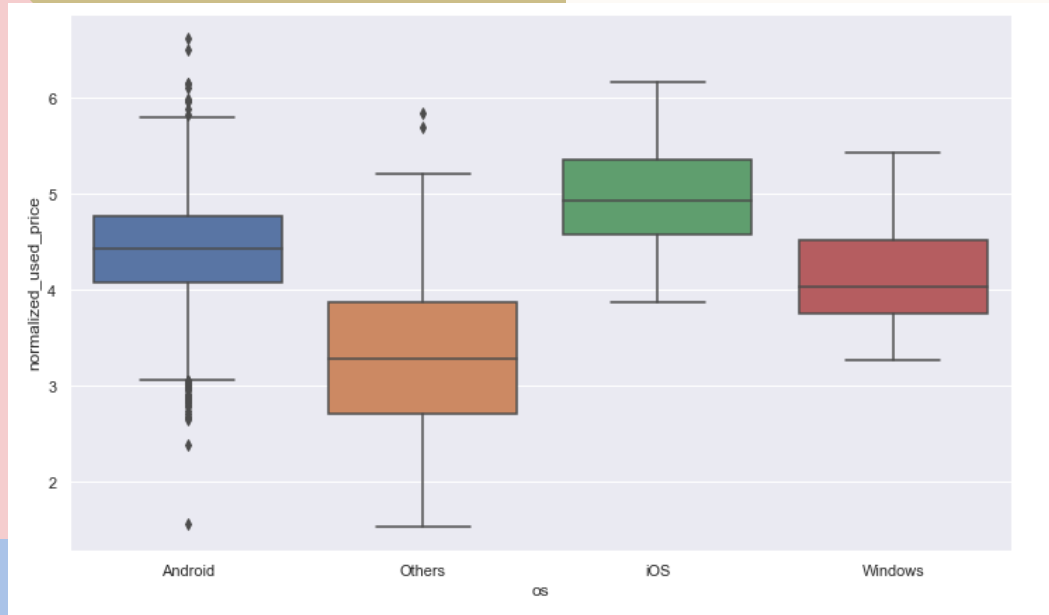
EDA: BIVARIATE ANALYSIS

16



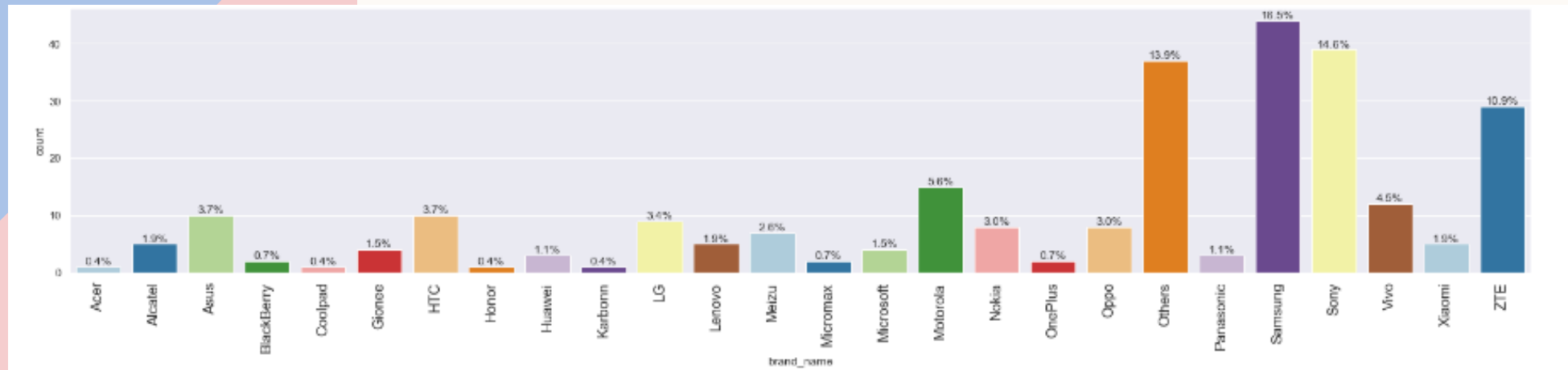
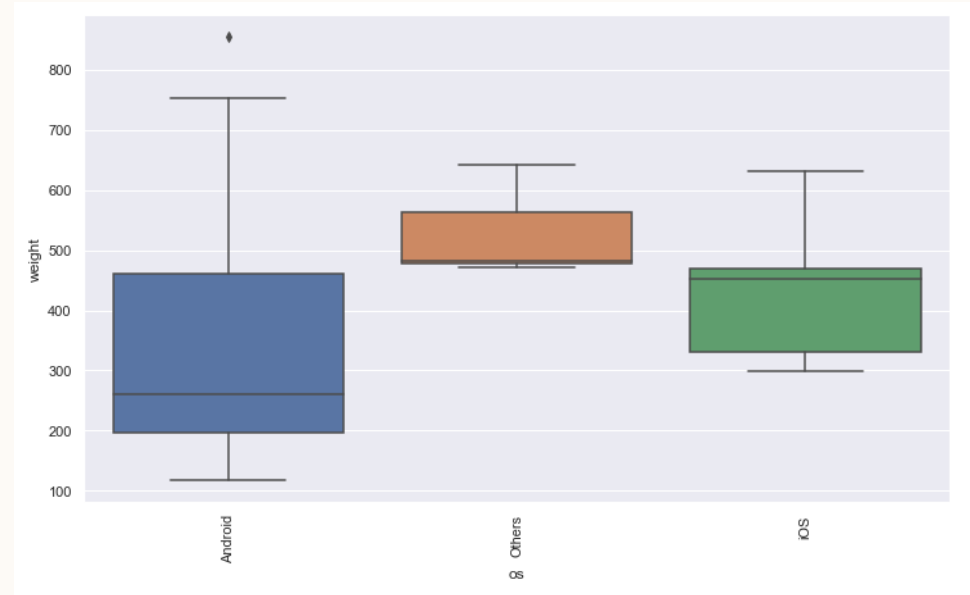
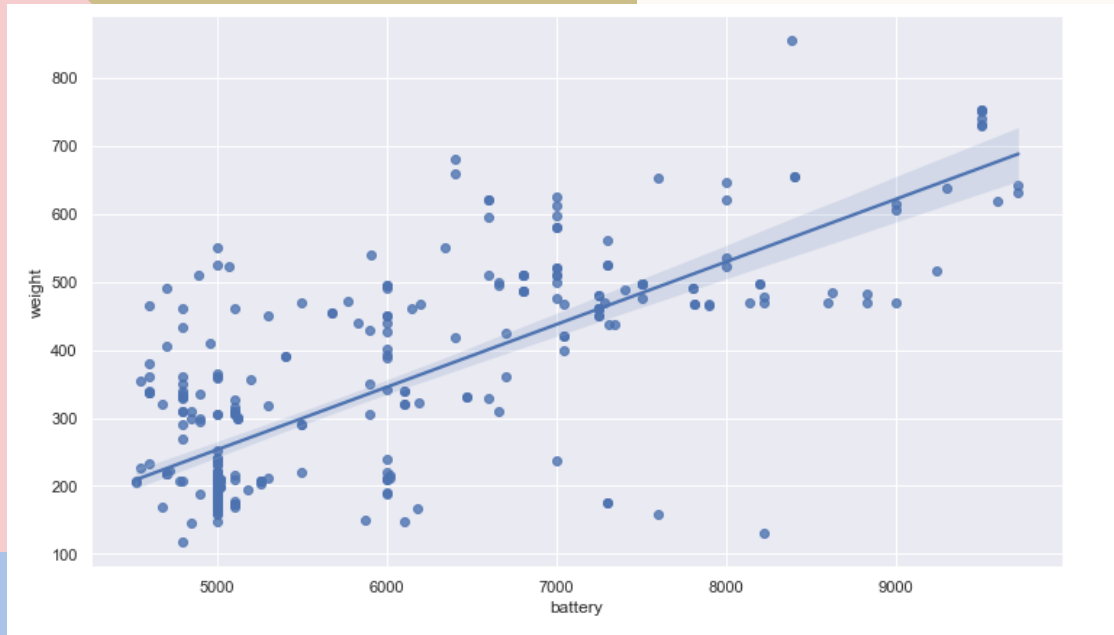
EDA: BIVARIATE ANALYSIS

17



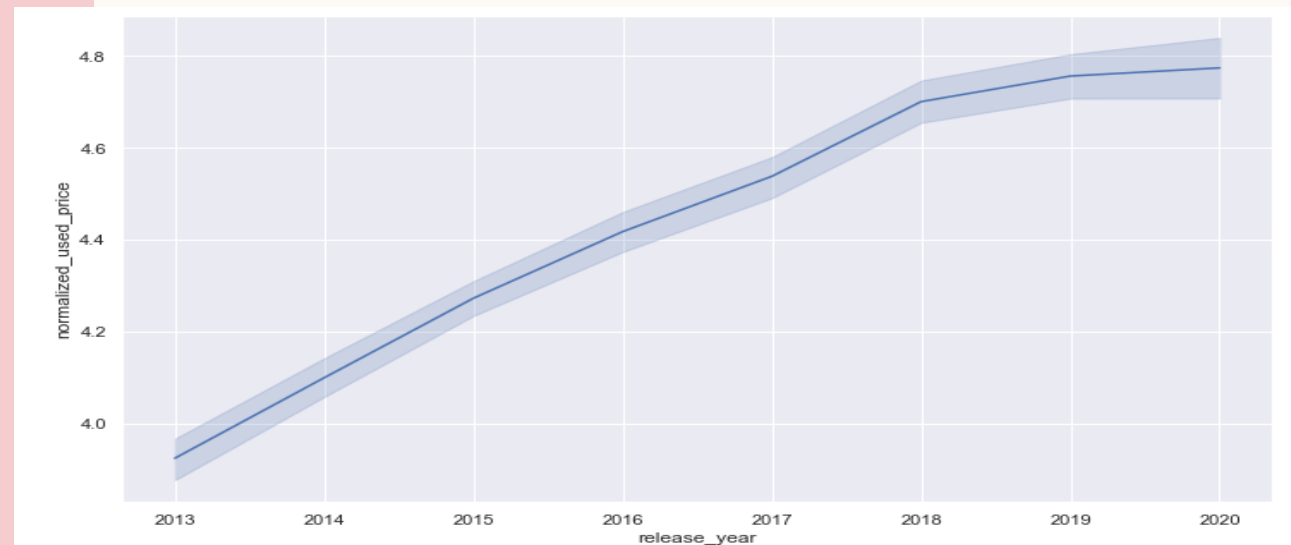
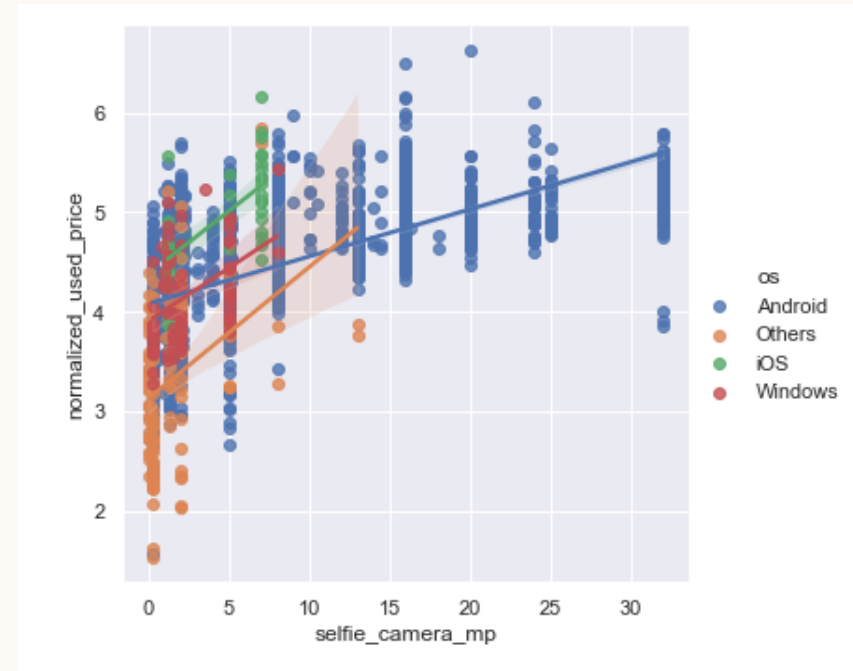
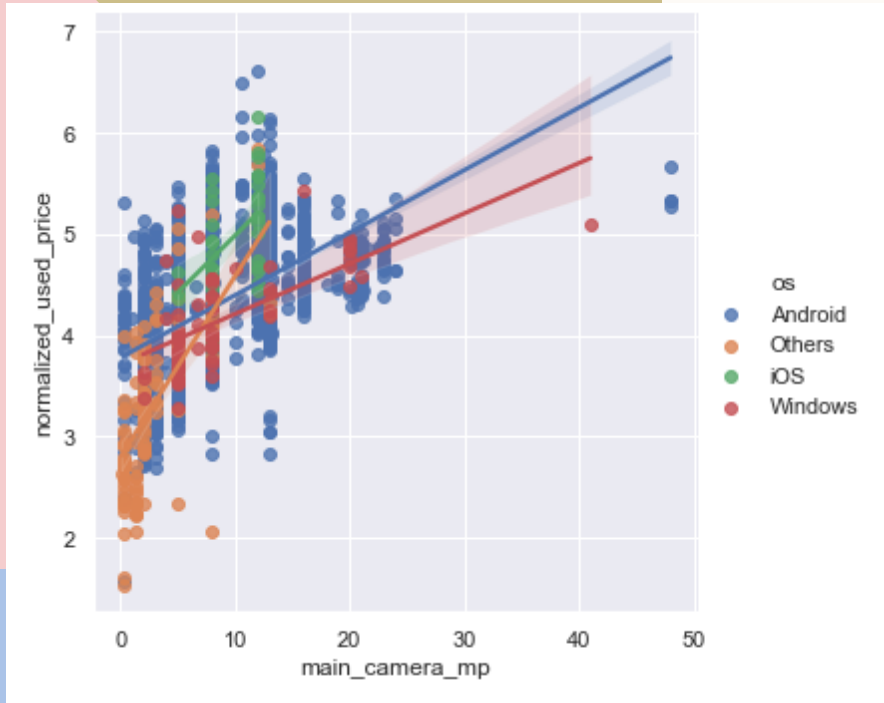
EDA: BIVARIATE ANALYSIS

18



EDA: BIVARIATE ANALYSIS

19



OBSERVATIONS

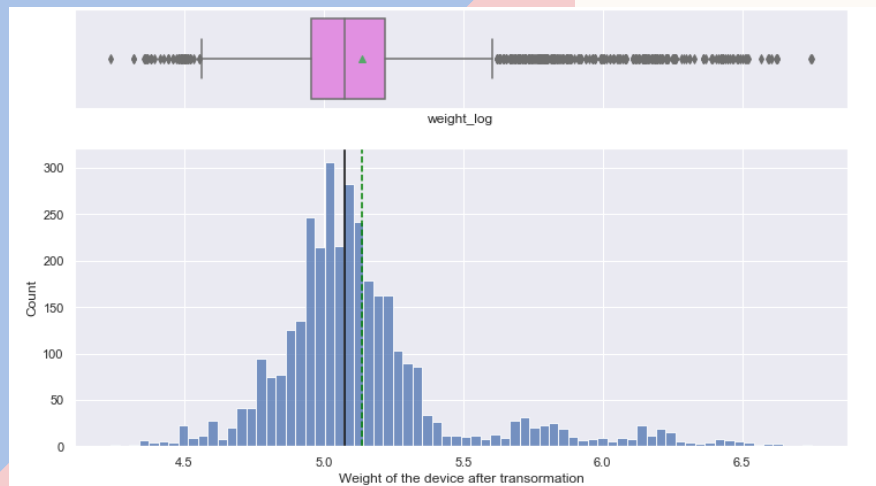
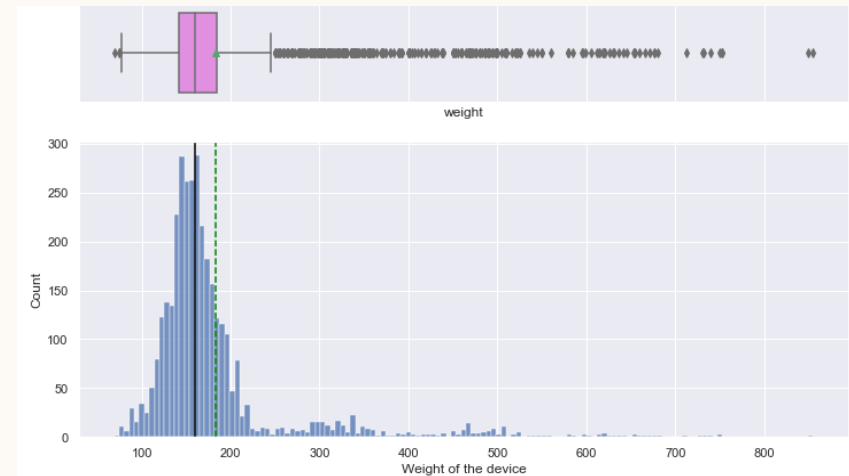
20

- Highly (+) correlation shown between:
 - Battery and screen_size
 - Weight and screen size
 - Normalized_new_price and normalized_used_price
- (-) correlation between:
 - Days_used and Selfie_camera
 - Days_used and normalized_used_price
- Android devices yields most RAMS
- OnePlus -> gives more RAM
- Huawei brand -> Consist of highest % of devices with screen size > 6in
- Android -> offers >6in size for most devices
- Apple brand -> offers greater energy capacity in mAh
- IOS devices -> shown to have higher normalized used and new price
- Samsung brand: produces most devices with large batteries > 4500mAh
 - Leading to heavier phones
 - Thus, android phones produces heavier phones
 - Note: As battery capacity increases so as the Weight
- Samsung devices: consist of greater count >13 main camera pixels
 - Thus, android consist of most % devices > 13 camera mega pixels
- Selfie camera mega pixels
 - Android devices: >8
 - Offers the greatest number
 - Huawei devices: >8
 - IOS devices: <8

DATA PREPROCESSING

21

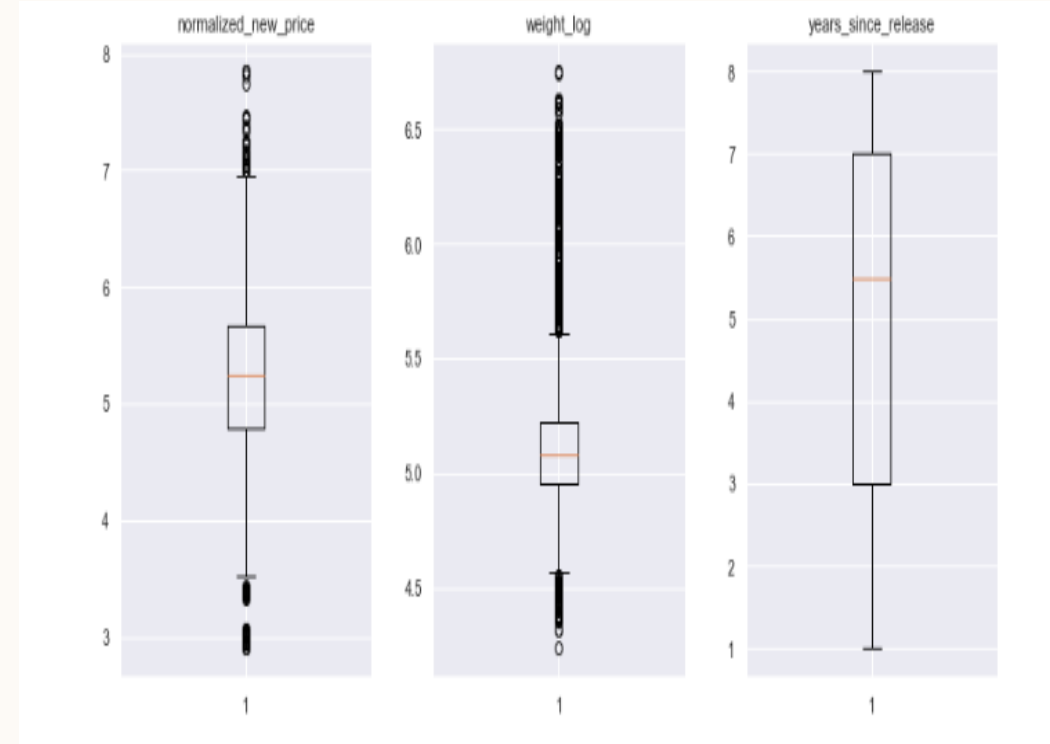
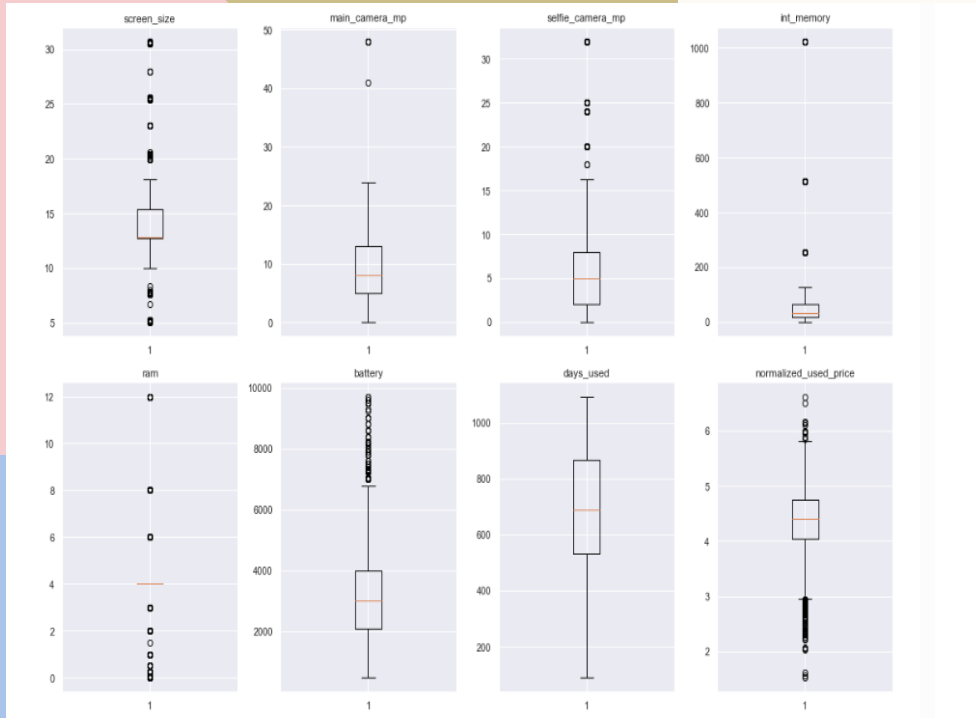
	brand_name	os	screen_size	4g	5g	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	weight	release_year
0	Honor	Android	14.50	yes	no	13.0	5.0	64.0	3.0	3020.0	146.0	2020
1	Honor	Android	17.30	yes	yes	13.0	16.0	128.0	8.0	4300.0	213.0	2020
2	Honor	Android	16.69	yes	yes	13.0	8.0	128.0	8.0	4200.0	213.0	2020
3	Honor	Android	25.50	yes	yes	13.0	8.0	64.0	6.0	7250.0	480.0	2020
4	Honor	Android	15.32	yes	no	13.0	8.0	64.0	3.0	5000.0	185.0	2020



- 3454 duplicates in Brand Name
- Main camera, Selfie Camera, Int Memory, RAM, Battery, and weight consist duplicate missing values
 - Weight column depicts better representation after log transformation compared to previous

OUTLIER CHECK

22



- No outlier in days used clolumn
- Few outliers in main camera, selfie camera, int memory and battery
- Greater than few lower and upper outliers in screen size and weight columns
- Note: All outliers were treated

DATA PREPARATION FOR MODELING

23

```
os screen_size 4g 5g main_camera_mp selfie_camera_mp \
0 Android 14.50 yes no 13.0 5.0
1 Android 17.30 yes yes 13.0 16.0
2 Android 16.69 yes yes 13.0 8.0
3 Android 19.30 yes yes 13.0 8.0
4 Android 15.32 yes no 13.0 8.0

int_memory ram battery days_used normalized_new_price weight_log
0 64.0 3.0 3020.0 127 4.715100 4.983607
1 128.0 8.0 4300.0 325 5.519018 5.361292
2 128.0 8.0 4200.0 162 5.884631 5.361292
3 64.0 6.0 6850.0 345 5.630961 5.617149
4 64.0 3.0 5000.0 293 4.947837 5.220356

years_since_release
0 1
1 1
2 1
3 1
4 1

0 4.307572
1 5.162097
2 5.111084
3 5.135387
4 4.389995
Name: normalized_used_price, dtype: float64
```

	const	screen_size	main_camera_mp	selfie_camera_mp	int_memory	ram	battery	days_used	normalized_new_price	weight_log	
0	1.0	14.50	13.0	5.0	64.0	3.0	3020.0	127.0	4.715100	4.983607	
1	1.0	17.30	13.0	16.0	128.0	8.0	4300.0	325.0	5.519018	5.361292	
2	1.0	16.69	13.0	8.0	128.0	8.0	4200.0	162.0	5.884631	5.361292	
3	1.0	19.30	13.0	8.0	64.0	6.0	6850.0	345.0	5.630961	5.617149	
4	1.0	15.32	13.0	8.0	64.0	3.0	5000.0	293.0	4.947837	5.220356	

of rows in train data: 2417
of rows in test data: 1037

MODEL BUILDING – LINEAR REGRESSION

24

```
=====
                        OLS Regression Results
=====
Dep. Variable:    normalized_used_price    R-squared:                0.835
Model:            OLS                    Adj. R-squared:           0.834
Method:           Least Squares          F-statistic:             812.6
Date:             Fri, 01 Apr 2022        Prob (F-statistic):       0.00
Time:             01:53:37               Log-Likelihood:          52.372
No. Observations: 2417                  AIC:                    -72.74
Df Residuals:     2401                  BIC:                    19.90
Df Model:         15
Covariance Type:  nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0729	0.169	-0.430	0.667	-0.405	0.259
screen_size	0.0408	0.004	10.693	0.000	0.033	0.048
main_camera_mp	0.0208	0.002	13.350	0.000	0.018	0.024
selfie_camera_mp	0.0163	0.002	9.478	0.000	0.013	0.020
int_memory	0.0004	0.000	2.279	0.023	5.78e-05	0.001
ram	0.0231	0.005	4.616	0.000	0.013	0.033
battery	1.133e-06	7.38e-06	0.153	0.878	-1.33e-05	1.56e-05
days_used	6.431e-05	3.11e-05	2.068	0.039	3.34e-06	0.000
normalized_new_price	0.4116	0.012	33.725	0.000	0.388	0.436
weight_log	0.2584	0.040	6.418	0.000	0.179	0.337
years_since_release	-0.0115	0.005	-2.420	0.016	-0.021	-0.002
os_Others	-0.0582	0.028	-2.050	0.040	-0.114	-0.003
os_Windows	0.0565	0.037	1.529	0.127	-0.016	0.129
os_iOS	0.0147	0.046	0.322	0.747	-0.075	0.104
4g_yes	0.0289	0.016	1.846	0.065	-0.002	0.060
5g_yes	-0.0467	0.031	-1.491	0.136	-0.108	0.015

```
=====
Omnibus:            214.589    Durbin-Watson:           1.937
Prob(Omnibus):      0.000     Jarque-Bera (JB):         431.386
Skew:               -0.579     Prob(JB):                 2.12e-94
Kurtosis:           4.715     Cond. No.                 1.22e+05
=====
```

Observations:

- Adjusted R-squared = 0.834
 - model not underfitting
- Y-int = constant coefficient (-0.0729)
- Normalized_new_price coef. --> 0.4116
- Train, test RMSE, and MAE are comparable --> model not overfitting
- MAE --> model can predict used device price w/n mean error of 0.187
- MAPE --> 4.58: suggests can predict w/n 4.6% of used device prices

LINEAR REGRESSION ASSUMPTIONS

25

	feature	VIF
0	const	1227.232818
1	screen_size	5.020059
2	main_camera_mp	2.130616
3	selfie_camera_mp	3.613245
4	int_memory	2.149691
5	ram	2.061785
6	battery	3.511445
7	days_used	2.579919
8	normalized_new_price	2.795831
9	weight_log	4.297022
10	years_since_release	5.073806
11	os_Others	1.328570
12	os_Windows	1.023320
13	os_iOS	1.094783
14	4g_yes	2.294751
15	5g_yes	1.709624

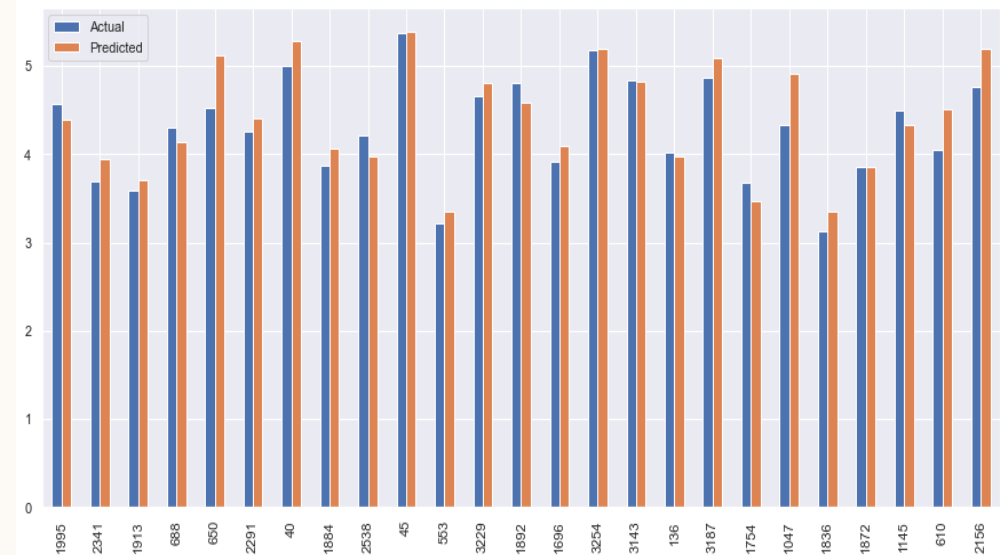
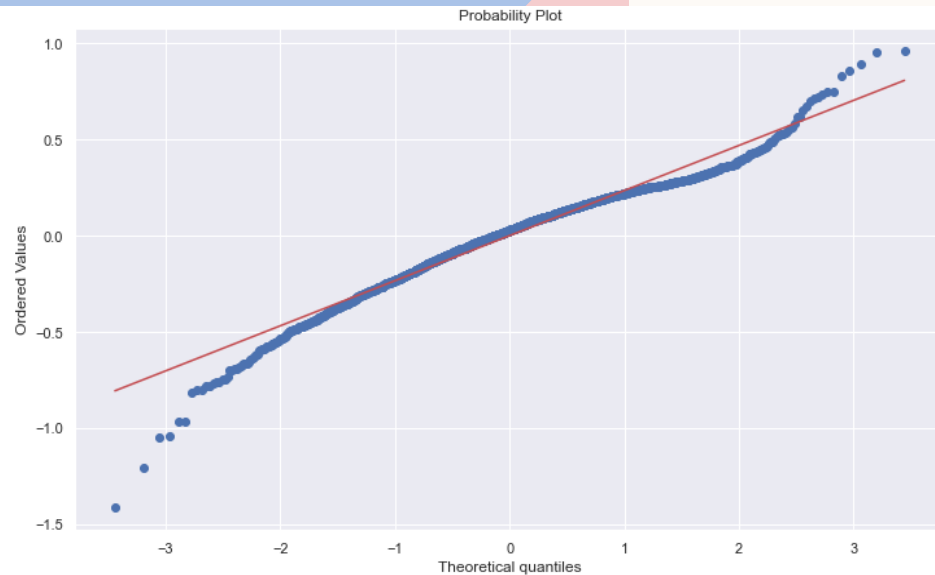
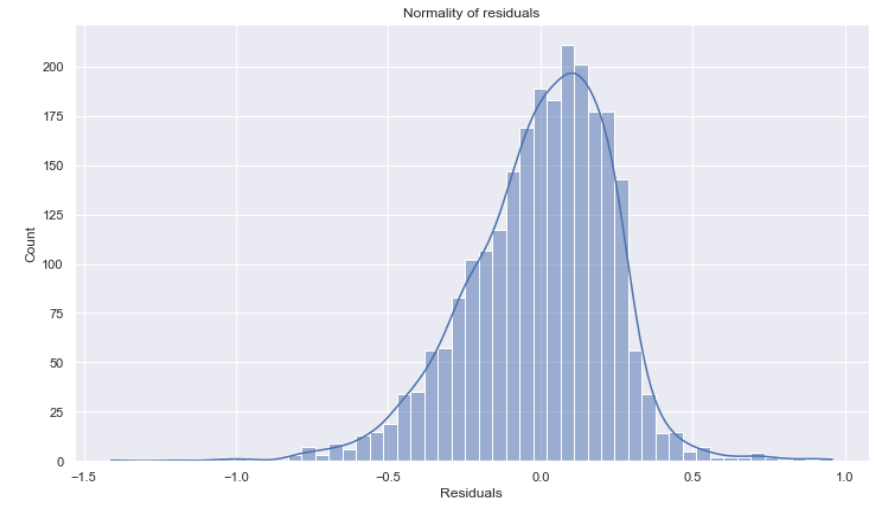
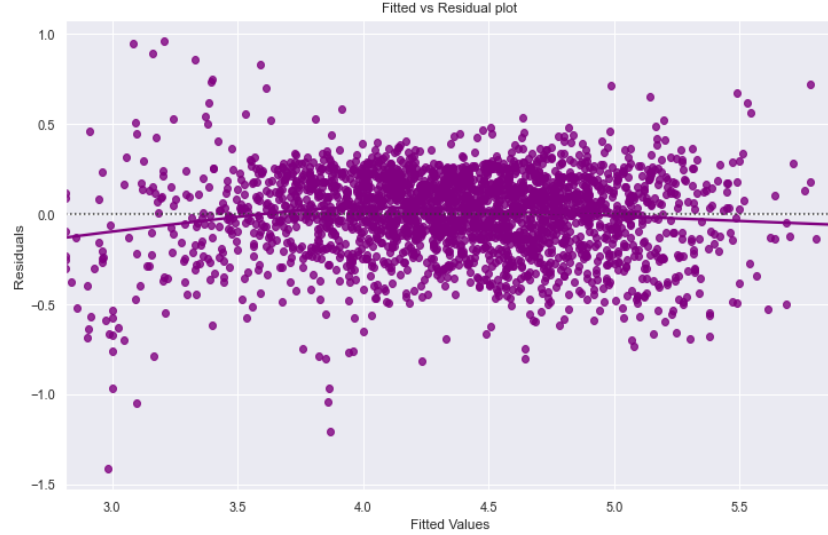
	feature	VIF
0	const	1227.210923
1	screen_size	4.947946
2	main_camera_mp	2.120467
3	selfie_camera_mp	3.104717
4	int_memory	2.104077
5	ram	2.057198
6	battery	3.436401
7	days_used	1.857689
8	normalized_new_price	2.488419
9	weight_log	4.283882
10	os_Others	1.328478
11	os_Windows	1.023297
12	os_iOS	1.086010
13	4g_yes	1.922239
14	5g_yes	1.686187

- If VIF
 - Between 1 & 5: low multicollinearity
 - Between 5 & 10: moderate multicollinearity
 - >10: high multicollinearity
- Removing screen_size: shows great impact on predictive power compared to years_since_release
- Removing years_since_release: shows VIF of screen_size below 5
 - Thus; no multicollinearity and assumption is satisfied
- No changes on adj. R-squared w/ dropping column
 - Indicates dropping column has little impact on the model
- Dropping high p-value variables
 - Adj. R-squared --> 0.834: shows model can explain ~83% variance
 - Shows model is ok
 - Olsmodel adj. R-squared --> 0.834
 - Indicated variables dropped has no impact on the mmodel

	col	Adj. R-squared after_dropping col	RMSE after dropping col
0	years_since_release	0.834071	0.237812
1	screen_size	0.826593	0.243112

LINEAR REGRESSION ASSUMPTIONS

26



OBSERVATIONS

27

- Test for Linearity and Independence
 - Scatter plot
 - Indicates distribution of residuals (errors) vs fitted values (predicted values)
 - No pattern in the plot → linearity and independence assumptions satisfied
- Test for Normality
 - Histogram → shows bell shape illustration
 - The residuals approx. followed a straight line except the tails
 - Shapiro-Wil test
 - P-value < 0.05 → residuals are not normal
 - Thus: can accept distribution as close to being normal so assumption is satisfied
- Test for Homoscedasticity
 - P-value $(0.62) > 0.05$ → residuals are homoscedastic; thus, assumption is satisfied
 - Homoscedasticity: variance of residuals is symmetrically distributed
 - Heteroscedastic: variance is unequal for residuals across the regression line
 - Model demonstrated good results showing actual and predicted values to be comparable

MODEL SUMMARY

28

```
=====
Omnibus:                217.210   Durbin-Watson:           1.936
Prob(Omnibus):           0.000   Jarque-Bera (JB):        431.648
Skew:                    -0.589   Prob(JB):                1.86e-94
Kurtosis:                4.702   Cond. No.                2.04e+03
=====
```

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.23735	0.183576	0.834647	0.83396	4.424958

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.244273	0.188165	0.834565	0.832952	4.607551

- Model defined ~83% variations in the data
 - Model is not underfitting
- Train and test RMSE (0.23, 0.23) & MAE (0.17, 0.18)
 - Comparable; indicating model did not suffer from overfittig
- MAPE --> indicated prediction of used devices w/n 4.6% error
- Thus; model olsmod2 showed good predictions and inferences

INSIGHTS & CONCLUSION

29

- Normalized new price showed strong (+) correlation with used price indicating the greater the price of new device, the greater the price of same device in refurbished market
- An increase in weight correlated with an increase in price for refurbished device
- Screen size, main camera, selfie camera, int. Memory, RAM, normalize new price, weight, and 4G have (+) coefficients indicating as the variable increases, the price of used devices increases as well
- Operating systems of devices except Android, iOS, and Windows possessed (-) coefficients indicating as the variables increases, the price of used devices decreases
- EDA demonstrated that expensive brands have maximum number of cell phones with larger screen size and better selfie camera
- Release year, days used, new price, and brand name had great impact on the final used price whereas 5G comes with new phones which will lead to increase in sales
- Some factors that demonstrated significant impact on refurbished device prices are screen size, main camera mega pixels, selfie camera pixels, internal memory, RAM, price of similar new device, weight of devices, whether device is 4G or not, and operating systems

RECOMMENDATION

30

- Most retailers run checks repair on used phones to meet certain standards in refurbished market then certain factors such as # of days of used device, battery, and release years should not affect the price on refurbished device
 - Therefore: Recell should pay less attention to those factors because as long the device is repaired, it will gain market
- This model can be utilized for predicting used price of phones and company should try to purchase newly released phones due to the resale price will be greater
- Company should try to avoid purchasing phones from Apple because the phones can be changed for the latest ones which is great option for customers
 - Thus, Recell should focus its operations or have more priority on phones with OS Android

The background features a large, light cream-colored circle on the left and a large, light pink circle on the right. These two circles overlap in the center. The area where they overlap is filled with a series of thin, white, concentric circular lines that radiate from the center of the pink circle. The top and bottom edges of the image are framed by a solid dark blue color.

THANK YOU!!!