Business Presentation

Contents

In this presentation, we are going to get deep into the Refurbished and used cell phone market.
Buying and selling used smartphones used to be something that happened on a handful of online marketplace sites. But the used and refurbished phone market has grown considerably over the past decade, and 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. This growth can be attributed to an uptick in demand for used smartphones that offer considerable savings compared with new models.

Business Problem Overview and Solution Approach

- **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.**

- **We as data scientist are going to analyze the data which is provided by ReCell and build a linear regression model to predict the price of a used phone and identify factors that significantly influence it.**

Data Overview

- We have a data with 3571 rows and 15 columns
- Some columns have missing values.
- Also, we have outliers in our data

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
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

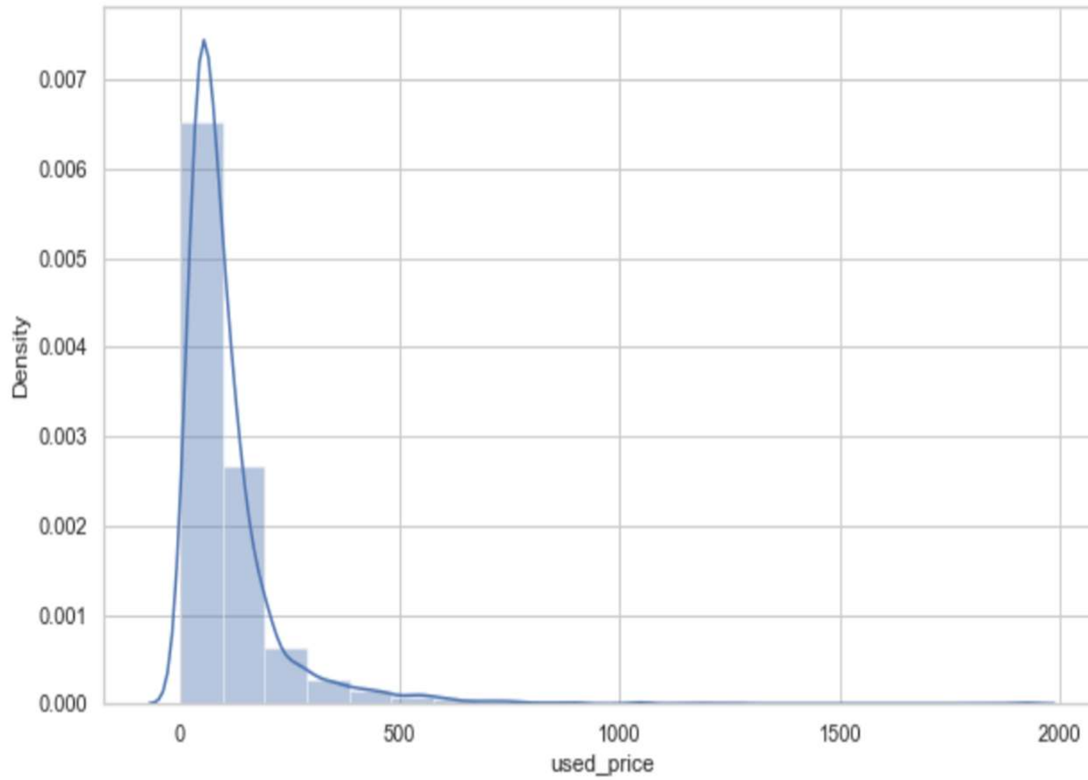
- os column has 4unique values
- 4g and 5g have 2 unique values
- brand_name has 34 unique values
- release_year rang from 2015 to 2020
- The average of used_price is 109.880 euro

- Columns brand_name, os, 4g and 5g need to be dummy variable.

- Also, 4g and 5g have overlapping so, I tried to apply a feature engineering to have in a new column called "4_5g" containing 3 possible values: 5g, 4g, and Other.

EDA

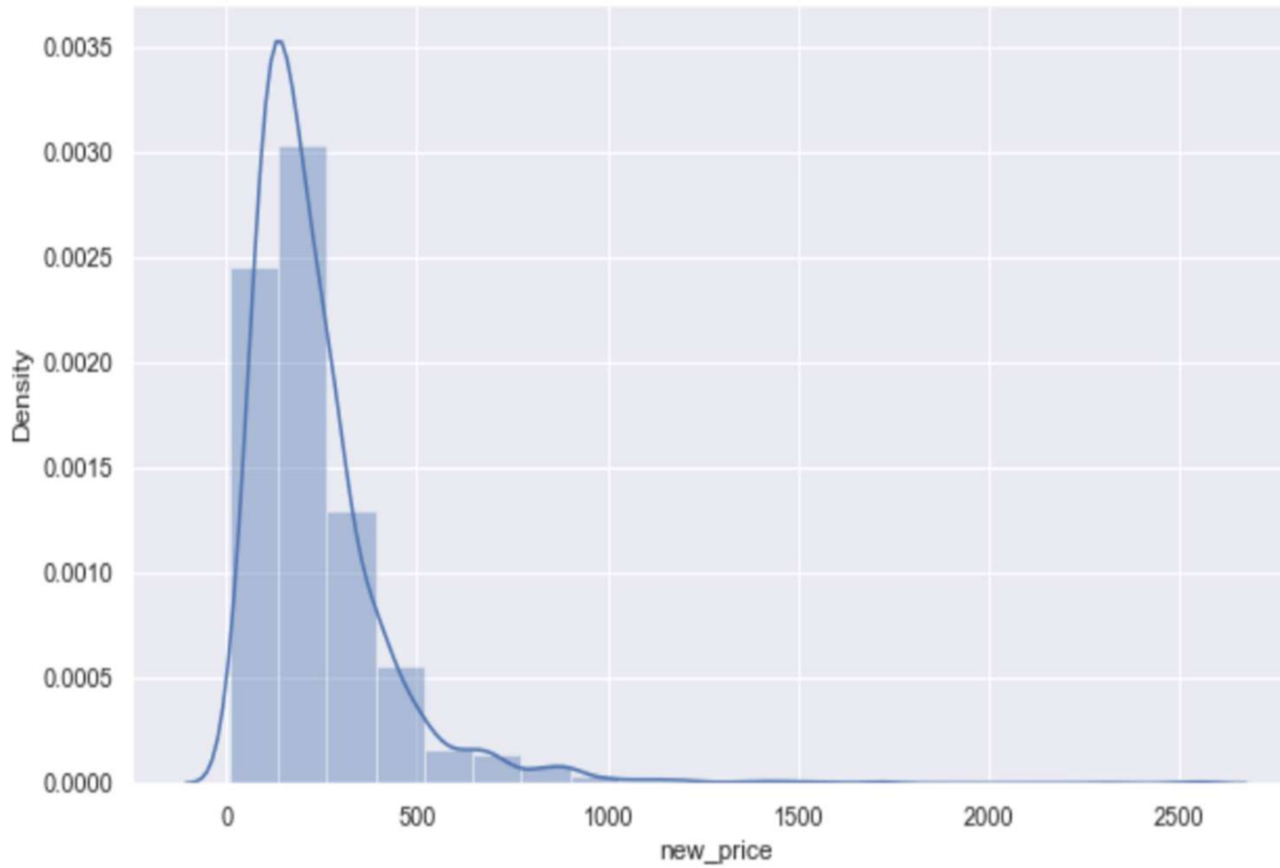
- Used_price and new_price are highly skewed to the right so, I used log transformation to make their shape better and look like a normal distribution.



EDA

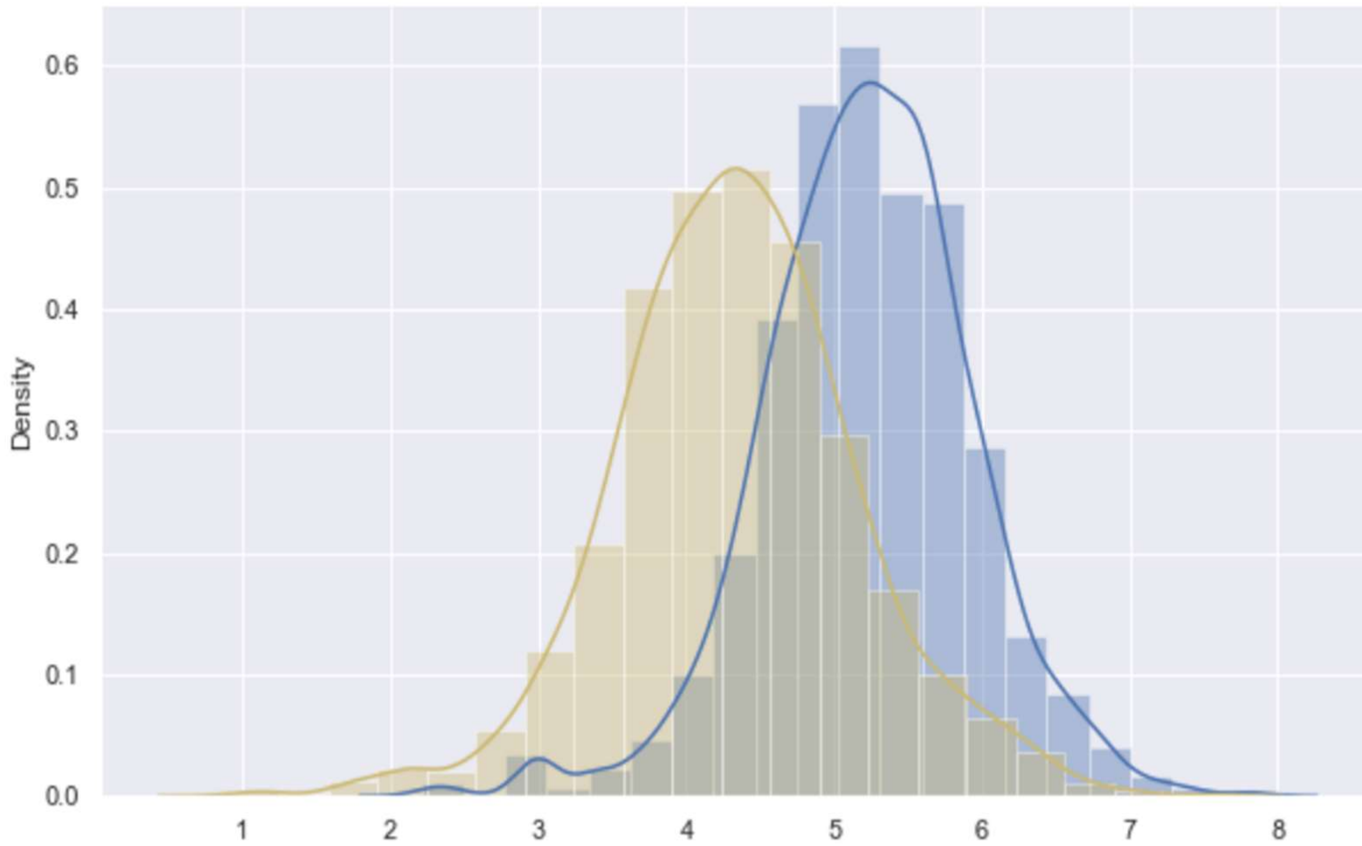
- As I said before new_price is highly skewed to the right

- This plot is about distribution of this column before log transformation.



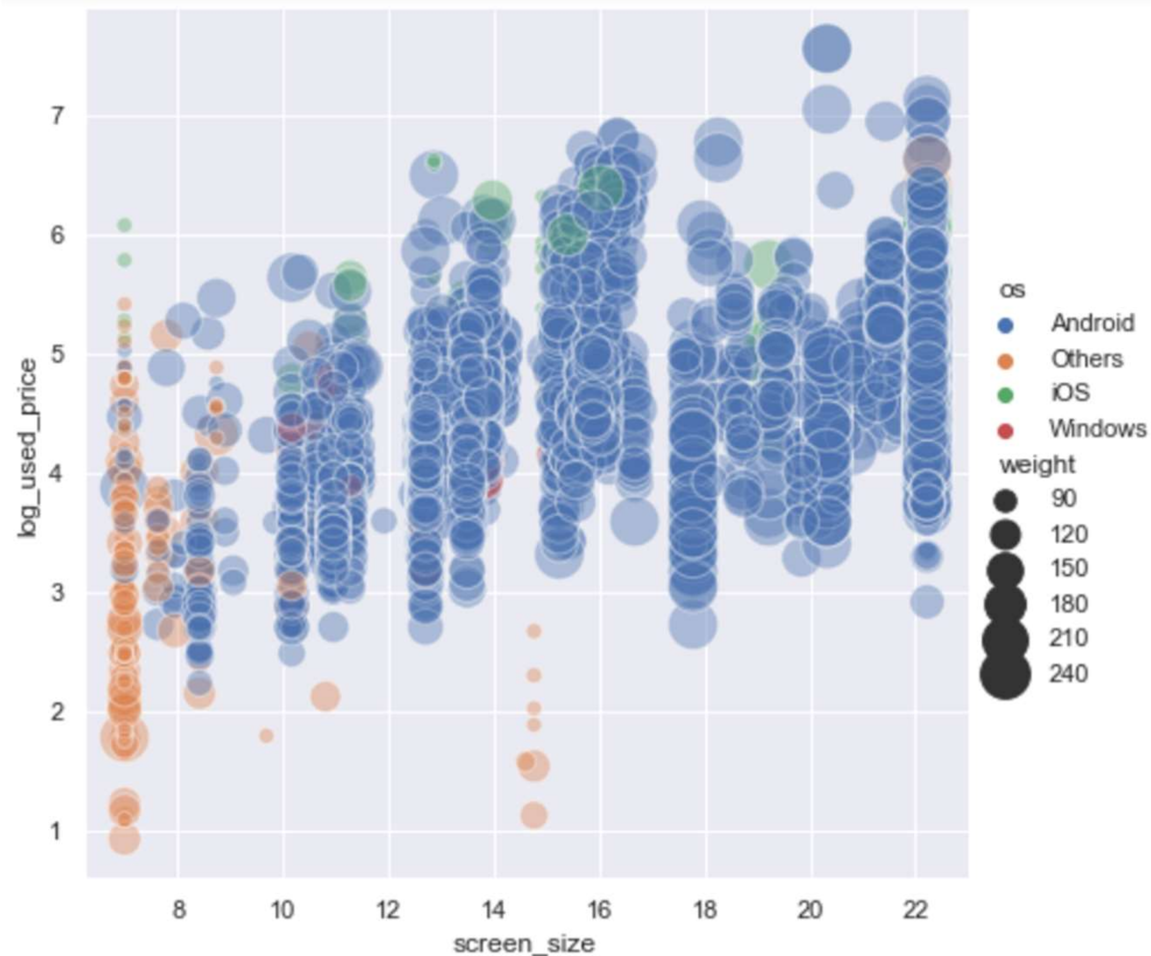
EDA

- Distribution of columns new_price and used_price after log transformation
- it seems that Log transformation is helpful in reducing the skewness. the blue one is related to new_price and yellow is related to used_price.



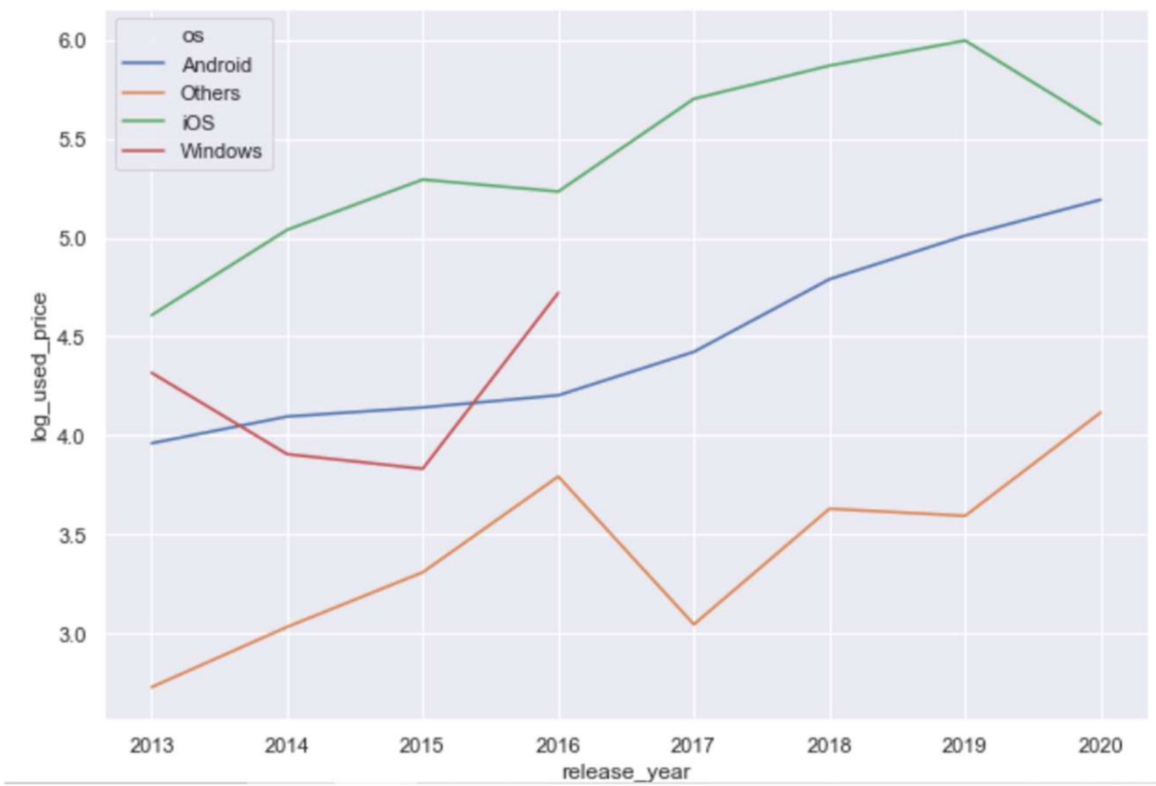
EDA

- it seems that heavy cell phones have a bigger screen, and a bigger screen does affect the price.
- Android is much than other [os] which is showing very good in this plot.



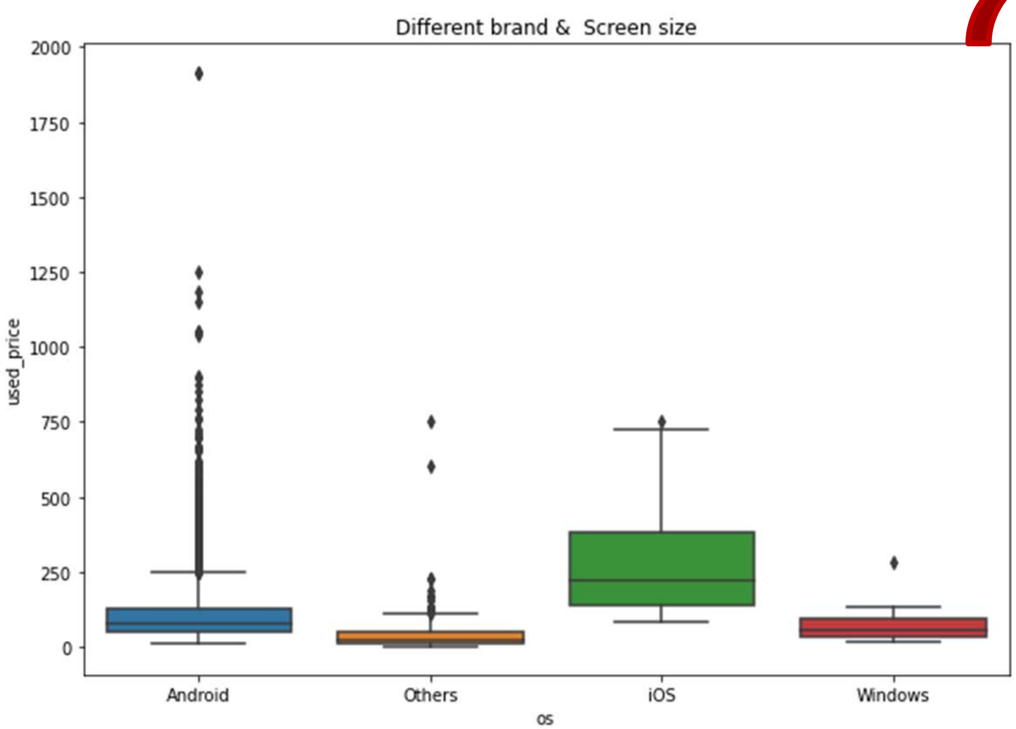
EDA

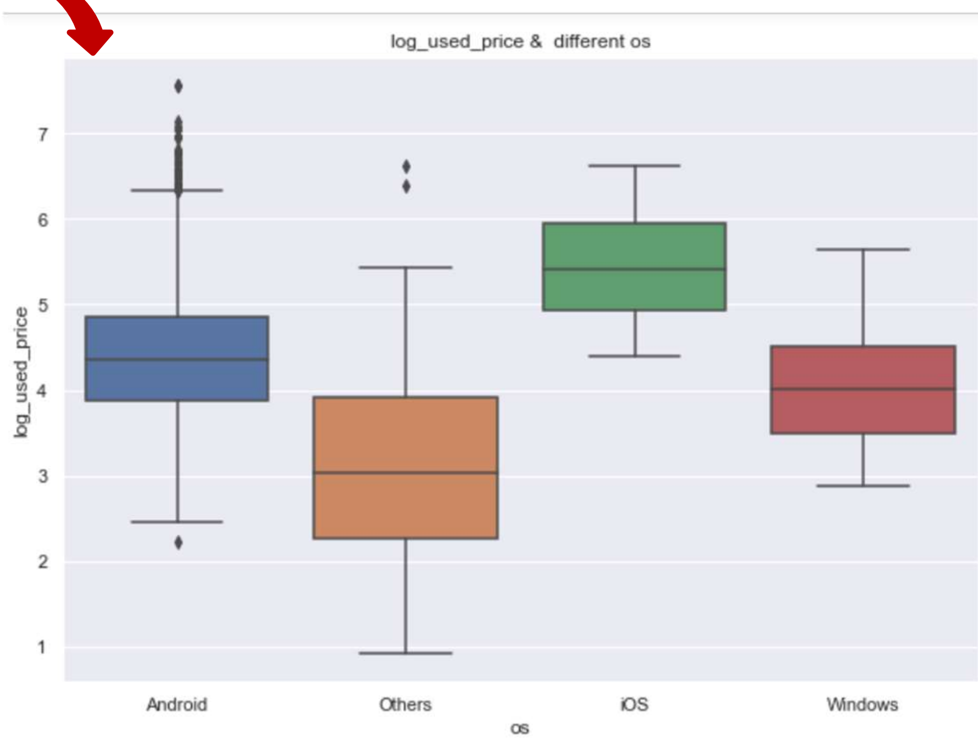
- The price of ios and Android were going up till 2019 but then ios is coming down. Also, the Others [os] are going up.
- Windows stop producing in 2016.
- Totally it is acceptable that the recently released cell phone has a price higher than the old released.



EDA

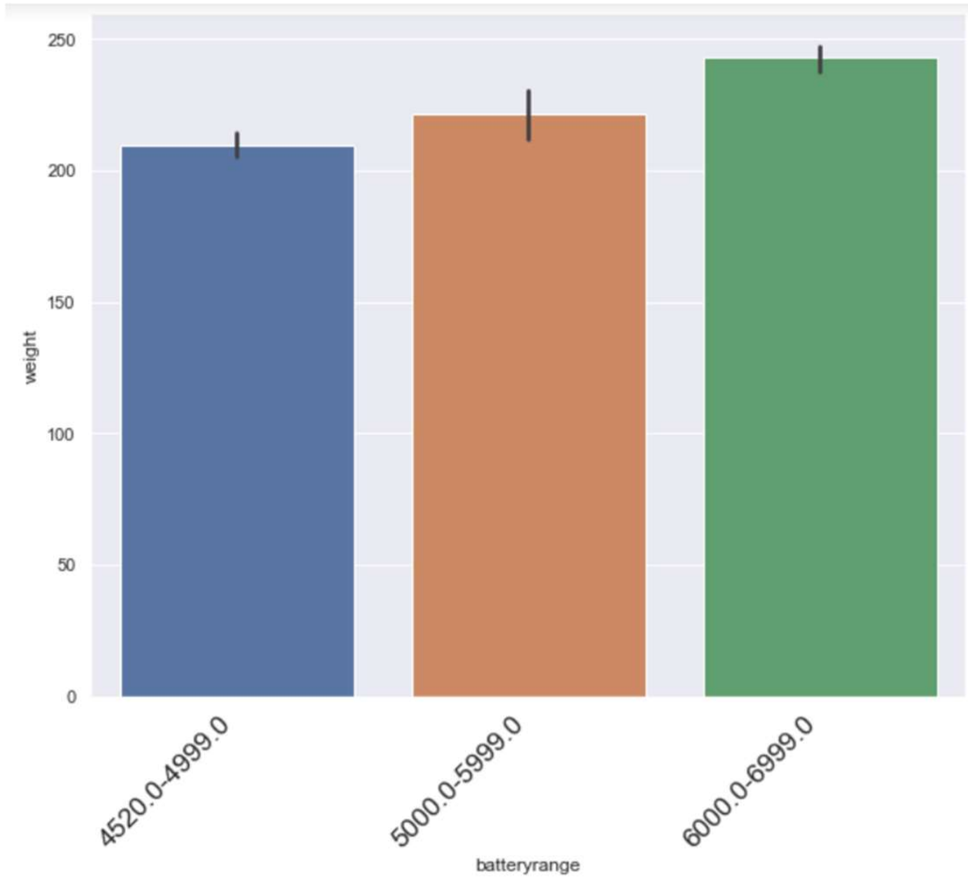
- after dealing with missing values and outliers they seem better than before. their skewness is less with lower outliers which are not shaped the distribution.





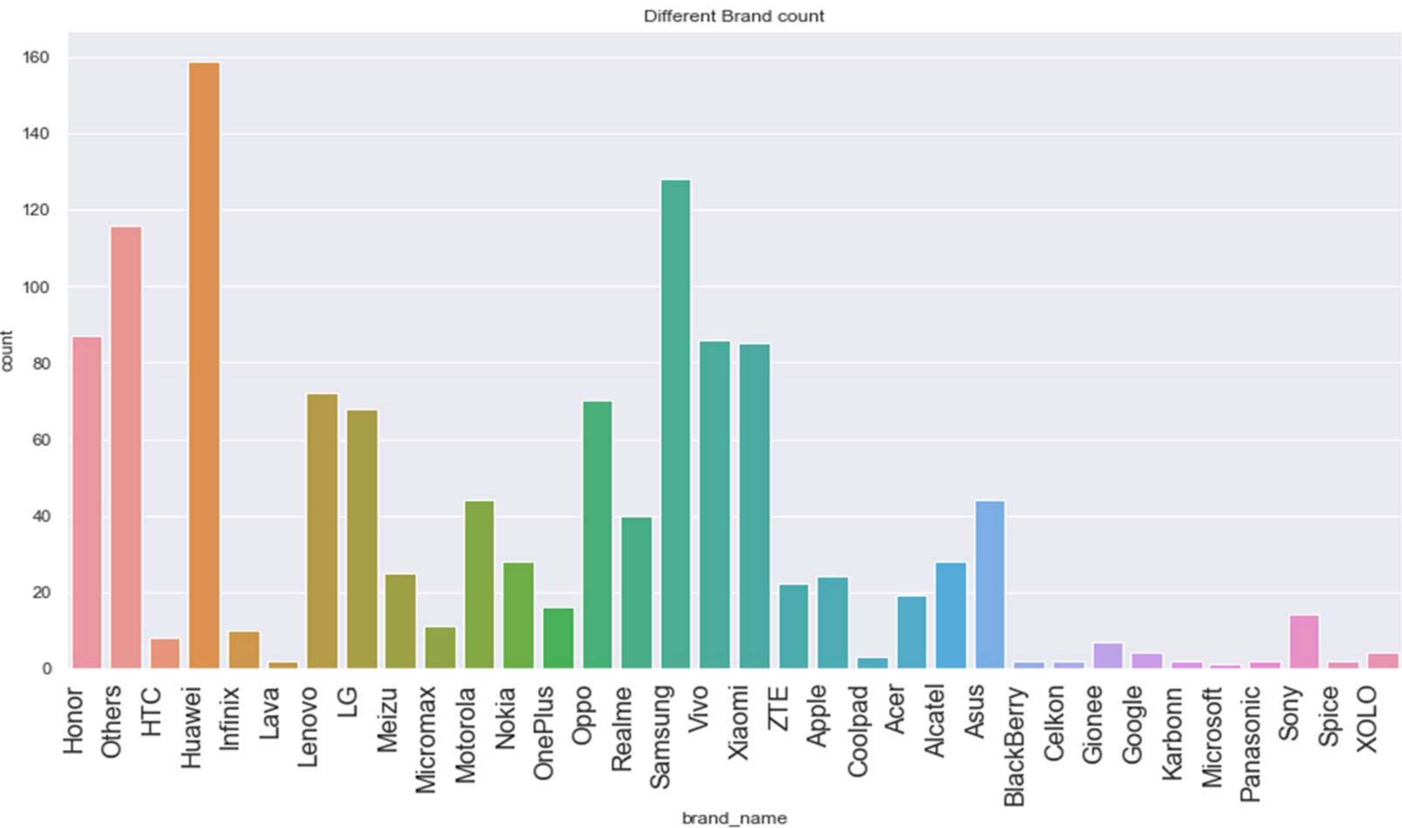
EDA

- How does the weight vary for phones offering large batteries (more than 4500 mAh)?
- I started with filtering the data base on mAh >4500 and get their means
- for best understanding I categorized battery column to 3 groups and then got a bar plot
- cellphones with highest weight have larger batteries



EDA

- How many phones are available across different brands with a screen size larger than 6 inches?

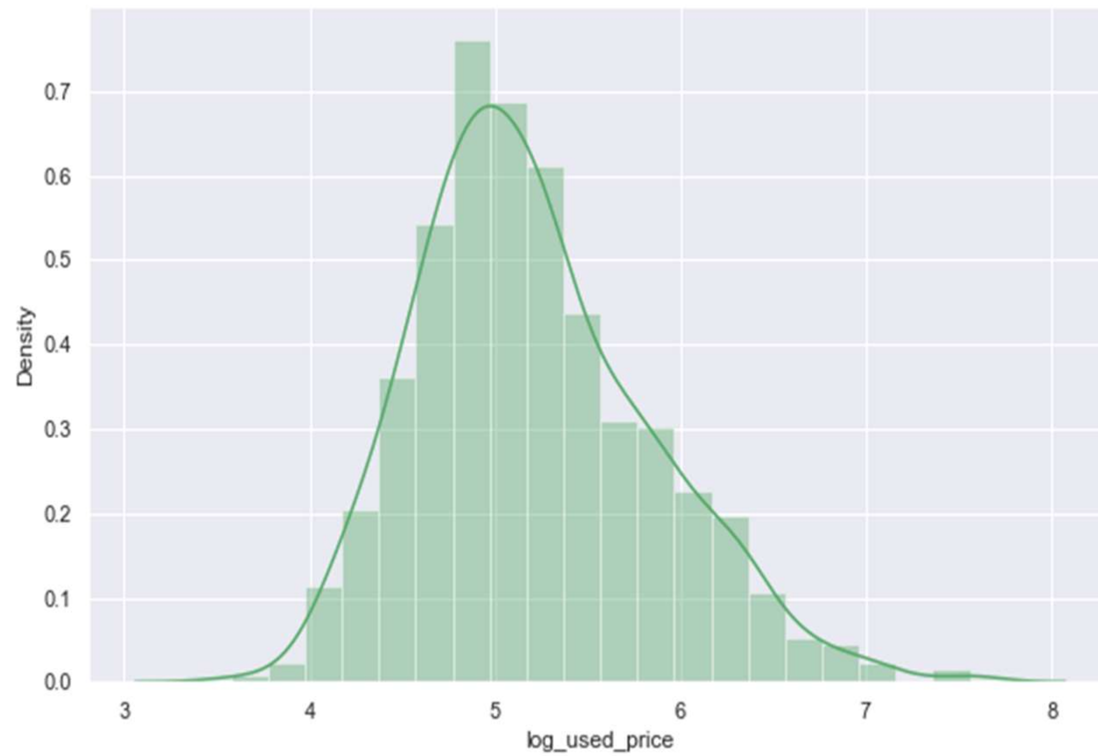


- between all brands, Huawei and Samsung cell phones are more than other brands which have screen sizes bigger than 6 inches (159 and 128) also, we have Other category which is located after them(116)

EDA

- What is the distribution of budget phones offering greater than 8MP selfie cameras across brands?

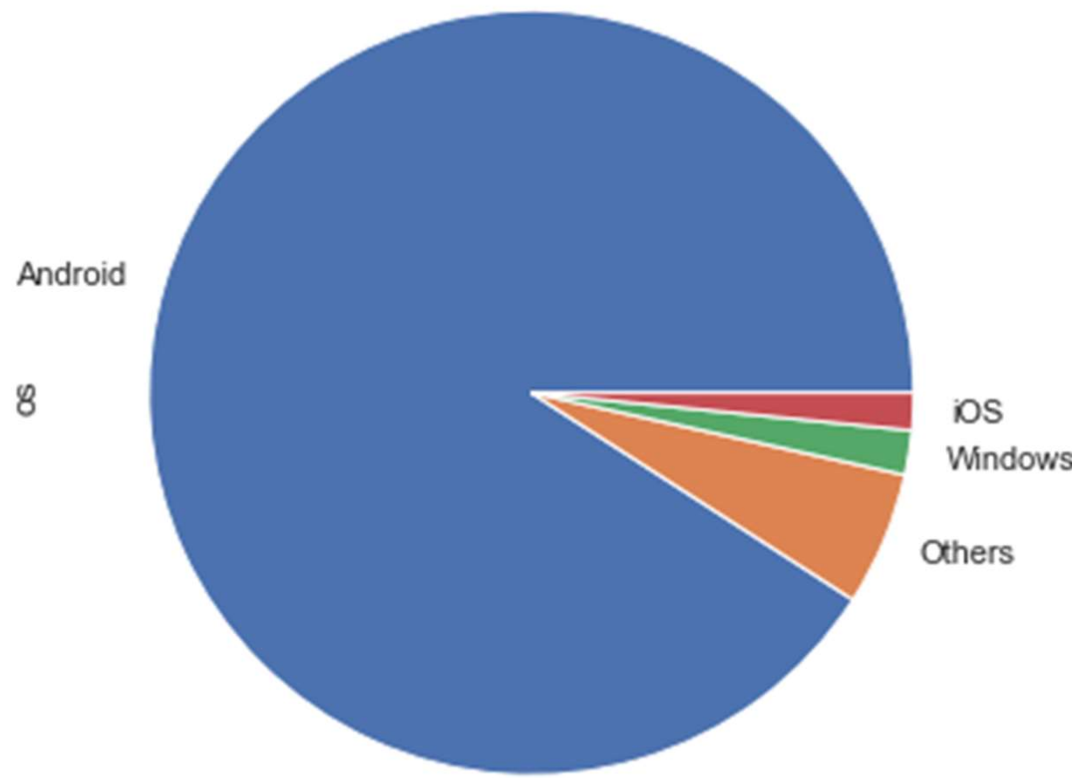
- the distribution of budget phones offering greater than 8MP selfie cameras across brands is normal in compere to previews plot it is not very skewed to the right



EDA

- What percentage of the used phone market is dominated by Android devices?

- around 90% of the used phone market is dominated by Android devices



Model Performance Summary

- Overview of ML model and its parameters

1. Create Dummy Variables for 3 columns: 4_5g , os and brand_name
2. since ram column has just one value(4.000) I will delete it.
3. Then I split the data:

independent variables
x = all my data but: log_used_price, 4_5g_other, os_Others, brand_name_Xiaomi and ram

dependent variable
y = log_used_price

Model Performance Summary

1. There is **negative** relationship between used_ price, days_used, some brand name like Gionee, Panasonic and Lenovo. It means for example, for every unit increase in days_used (one day) there is a 0.0011€ decrease in used_price.

2. 1 unit increase in int_memory(GB) leads to an increase in used_price by 0.001€.

3. 1 unit increase in screen_size (cm) leads to an increase in used_price by 0.001788€.

4. 1 EURO increase in log_new_price (price of new cell phone) leads to increase in used_price(used cell phones) by 0.99€.

	coef
screen_size	0.0017
int_memory	0.0001
days_used	-0.0011
log_new_price	0.9977
brand_name_Gionee	-0.0372
brand_name_Lenovo	-0.0238
brand_name_Panasonic	-0.0310

Model Performance Summary

- Training Performance

RMSE	MAE	R-squared	Adj. R-squared	MAPE	variance_score
11.473	7.136	0.991	0.991	0.07	0.991

- Test Performance

RMSE	MAE	R-squared	Adj. R-squared	MAPE	variance_score
11.487	7.322	0.99	0.99	0.071	0.99

- The model can explain ~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 0.07% of the used cell phone price.

- Hence, we can conclude the model olsmod2 is good for prediction as well as inference purposes.

Business Insights and Recommendations

- All the assumptions of linear regression have been checked but two tests for checking the normality are not satisfied. since I did set the model on log transformation for **Y** I believe we should apply non-parametric tests e.g. chi-square in place of correlation. Although I believe there should be some complicated transformations to solve this problem which I don't know them yet.

- Anyway, as we were asked, I did summary the final model and closest solution although the assumptions were not met.

Business Insights and Recommendations

- The model can be used for predictive purposes as it can make predictions within ~7% of the actual price.
- ReCell should look to attract people who want to sell used phones which have been released in recent years and have not been used for many days.
- They should also try to gather, and put-up phones having a high price for new models to try and increase revenue.
- They can focus on volume for the budget phones and offer discounts during festive sales on premium phones.
- Additional data regarding customer demographics (age, gender, income, etc.) can be collected and analyzed to gain better insights into the preferences of customers across different segments.

greatlearning

Power Ahead

Happy Learning !