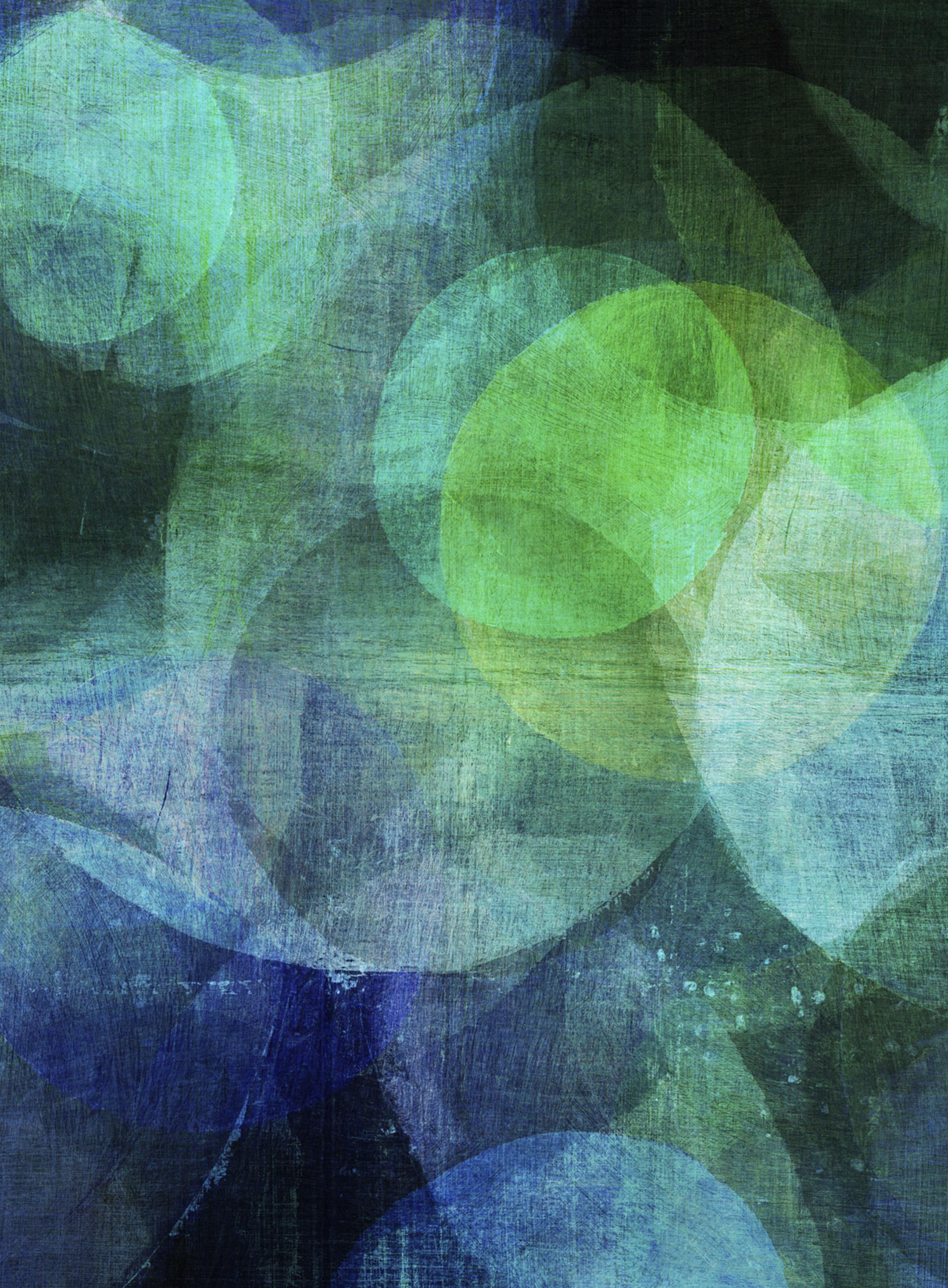


# RECELL VALUE PREDICTION

---

*Presented by*  
*Sathya*



# **Problem Summary**

*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.*

*The ML model to predict the resell values of the mobile based on the features including the brand.*

# Data Set Overview

*The recall Dataset has 3571 rows with 15 columns.*

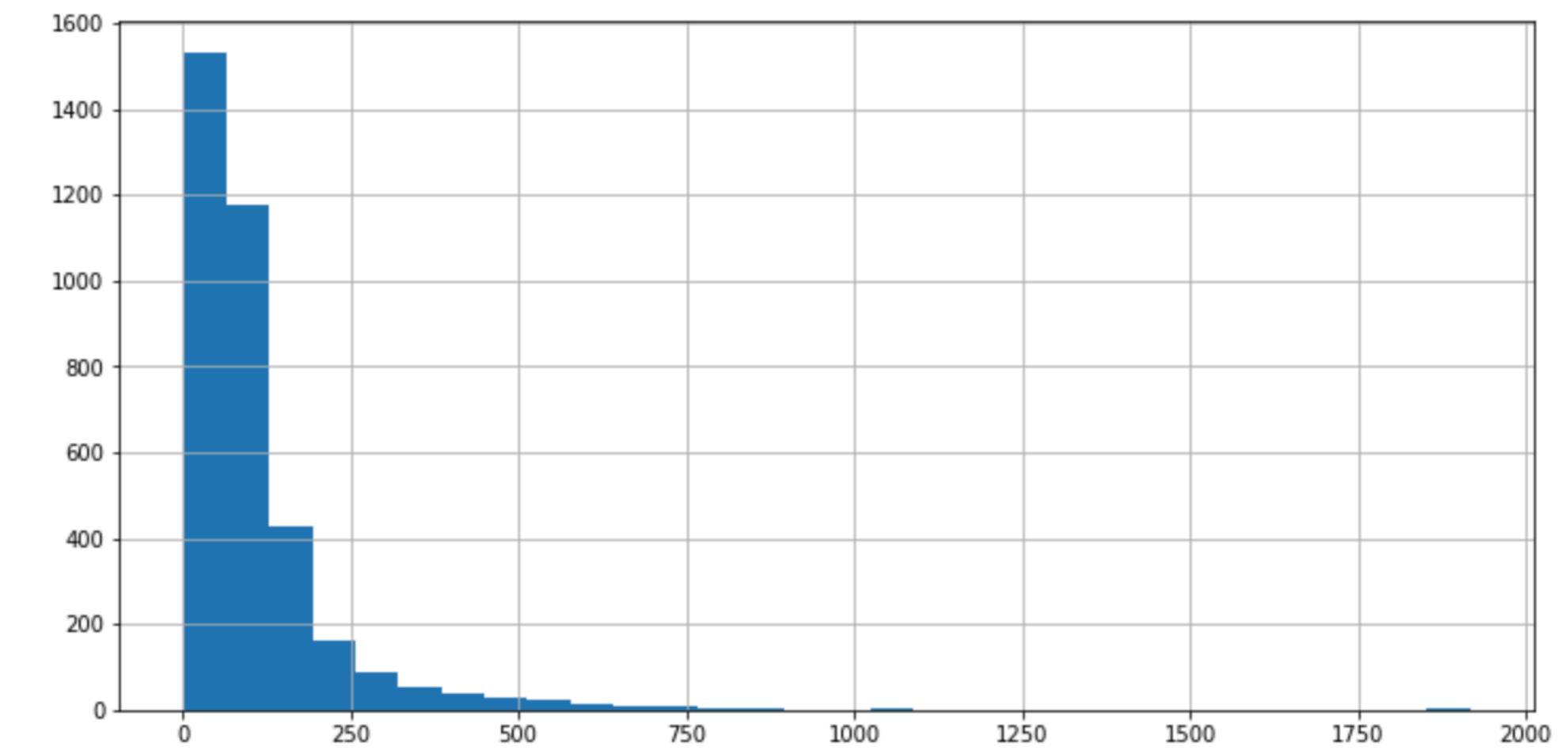
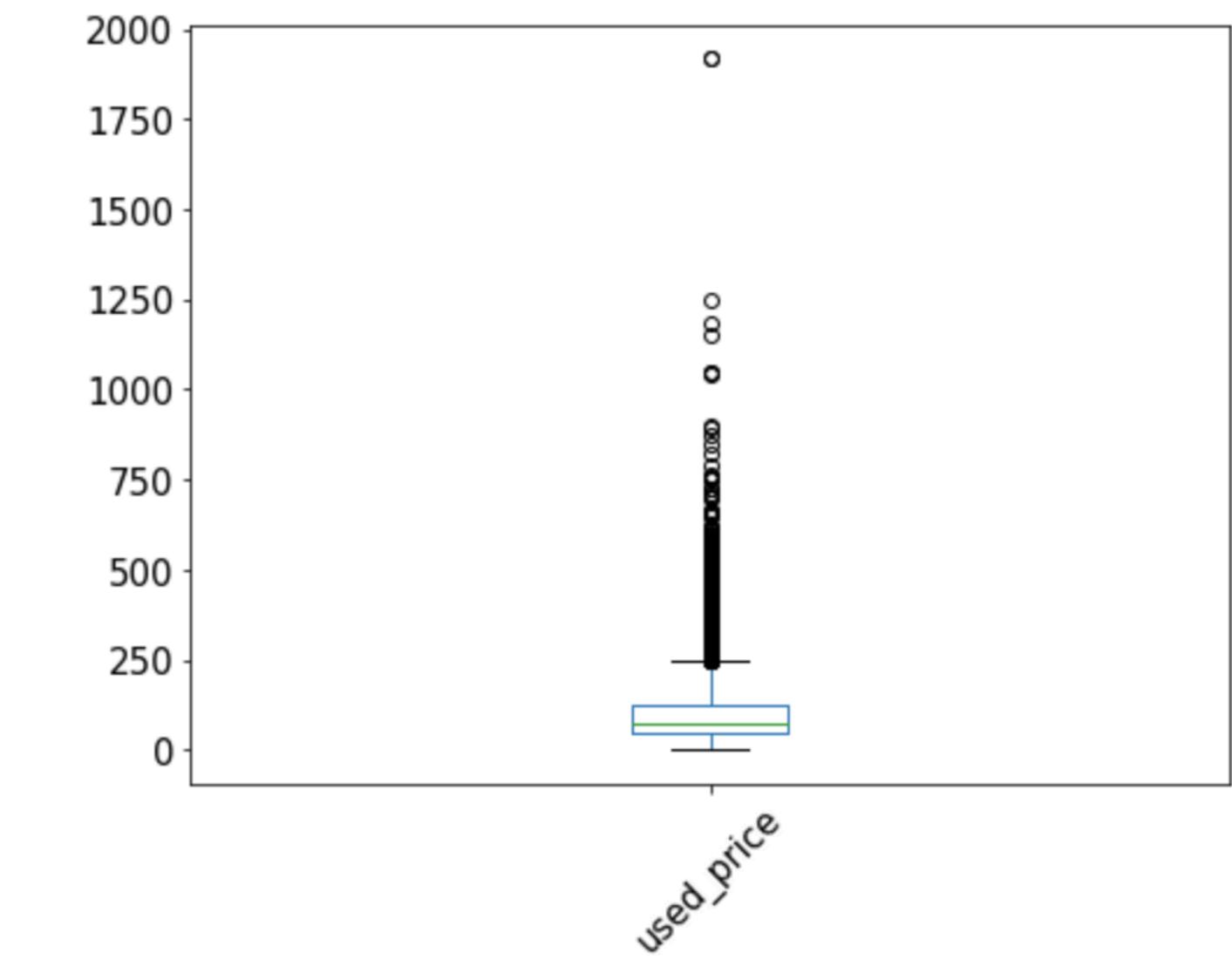
*There are four categorical columns and remaining 11 columns are numerical*



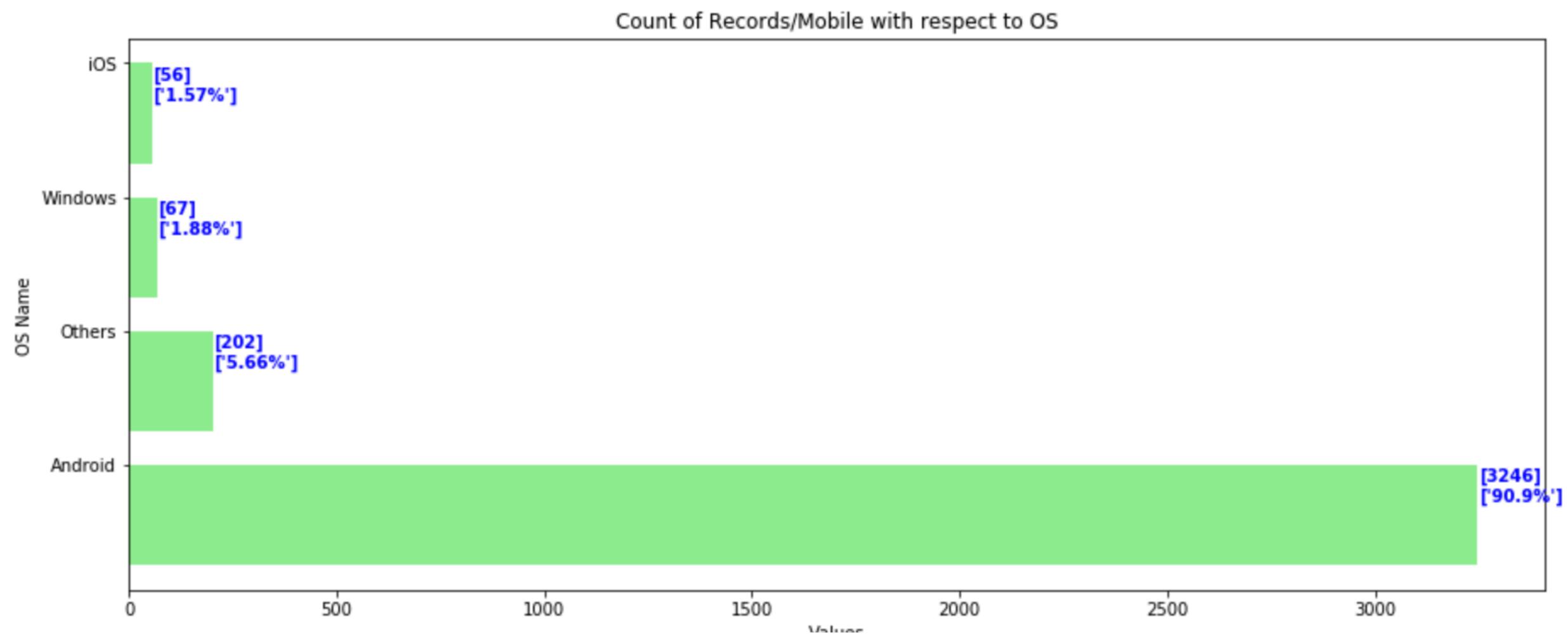
# EDA -Used Price - Distribution

*There are outliers which are having greater value of used\_price which makes the mean value greater than median.*

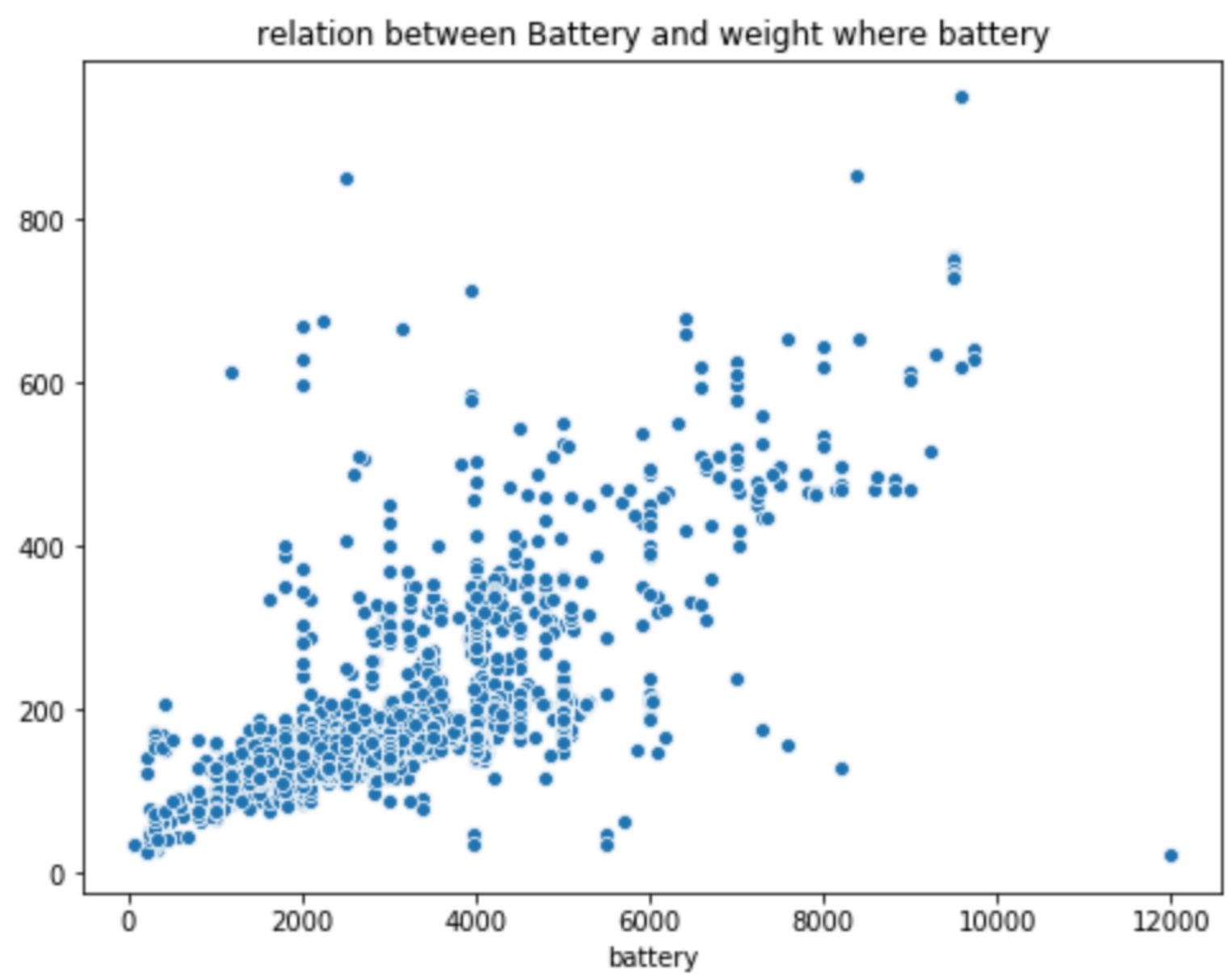
*Where the mean is 109 and median as 75 and mode is 59.78.*



# EDA

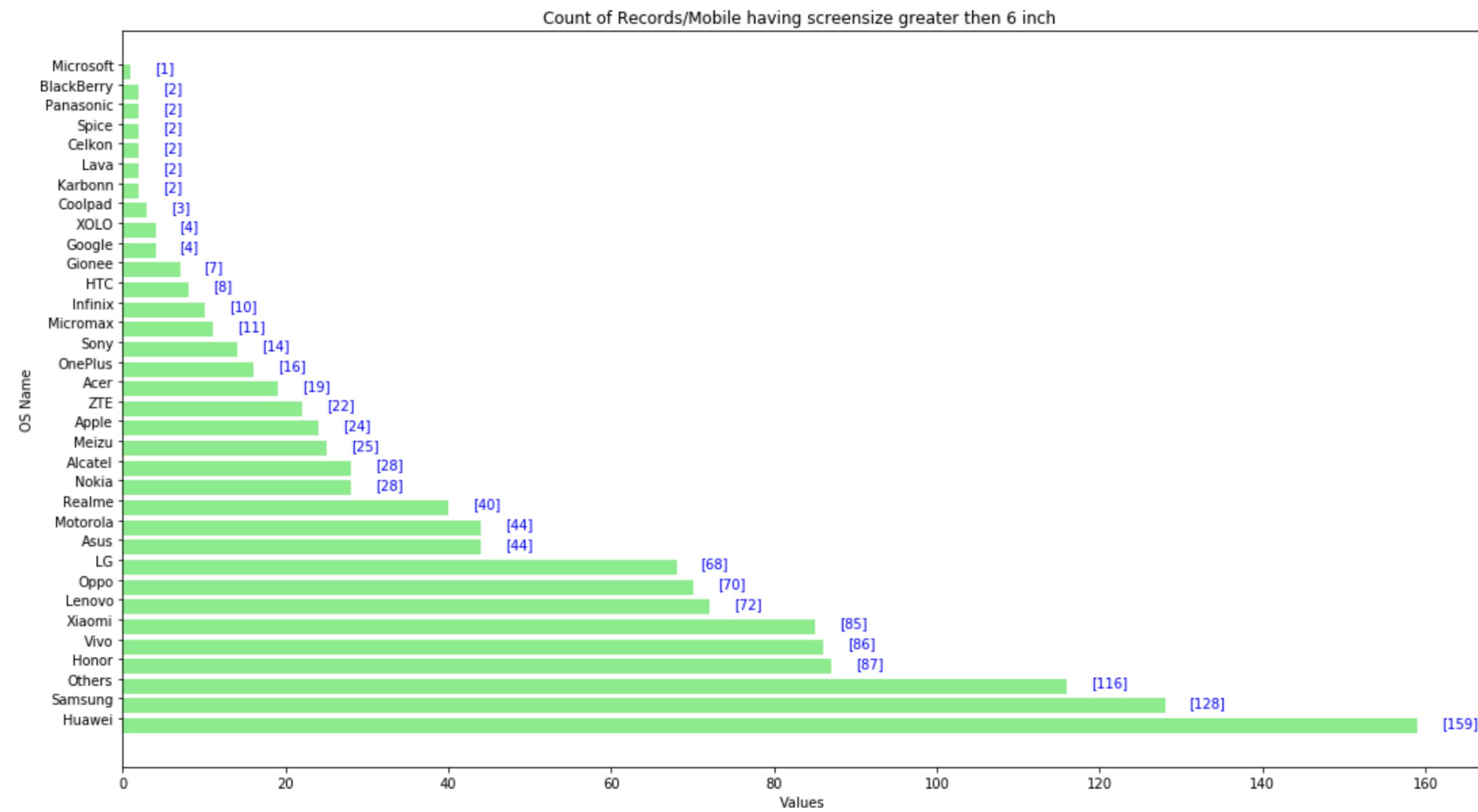


*The Android OS is dominating the market with 90% of the total*



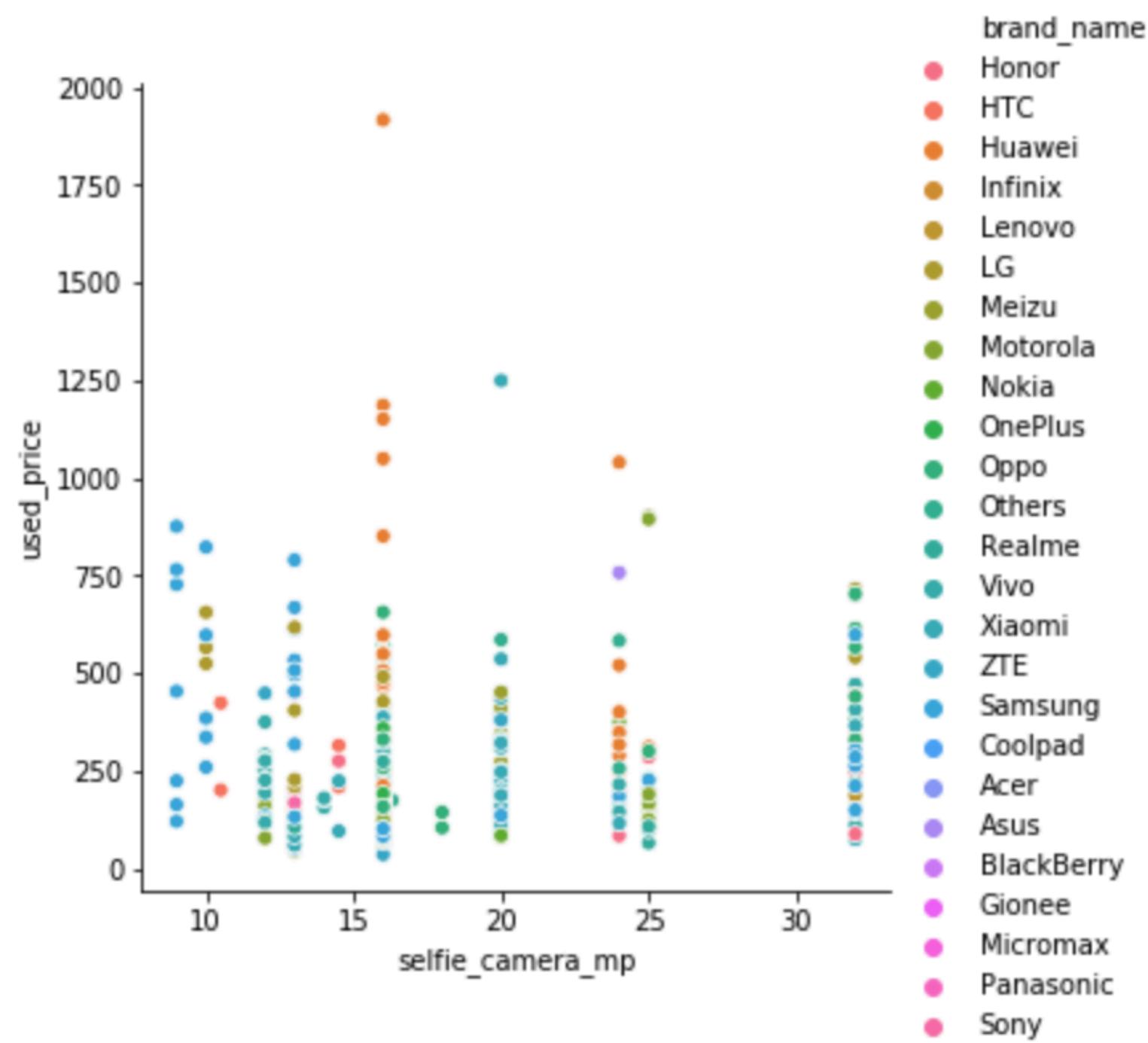
*The weight of the mobile and battery in mAh is highly correlated. as the battery power increases the weight the mobile also increase*

# EDA



*There are 1235 phones greater than 6 inch out of 3571 phones. and the Huawei brand most dominant brand with bigger screen size*

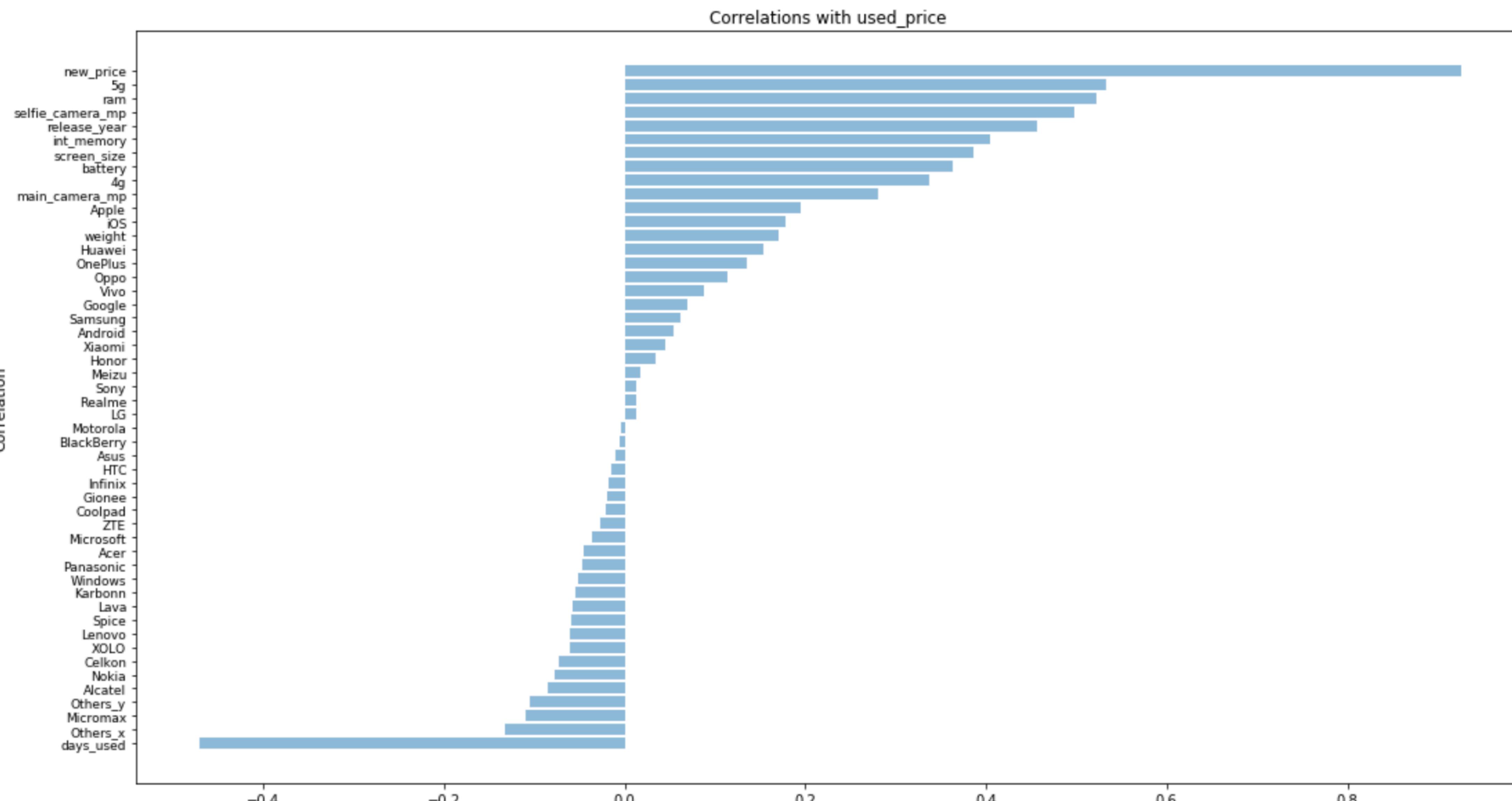
# EDA



*Budget phones nowadays offer great selfie cameras, allowing us to capture our favorite moments with loved ones. What is the distribution of budget phones offering greater than 8MP selfie cameras across brands?*

*Ans: not any particular pattern , it is distributed across the price, there are other*

# Co-Relation of Features - With Used price



# Data Preprocessing

*Outliers : Removed the data from the dataset which has the used price high value*

*Filling Null Values :*

*null values in main\_camera\_mp, selfie\_camera\_mp which are filled based on the mean value of the respective brand*

*Null values of battery and weight is filled with mean value of the respective brand*

*The OS with Others is filled with the Similar value in the brand*

*The brand which has the value as ‘Others’ is dropped form the dataset*

*Converting the categorical to numeric :*

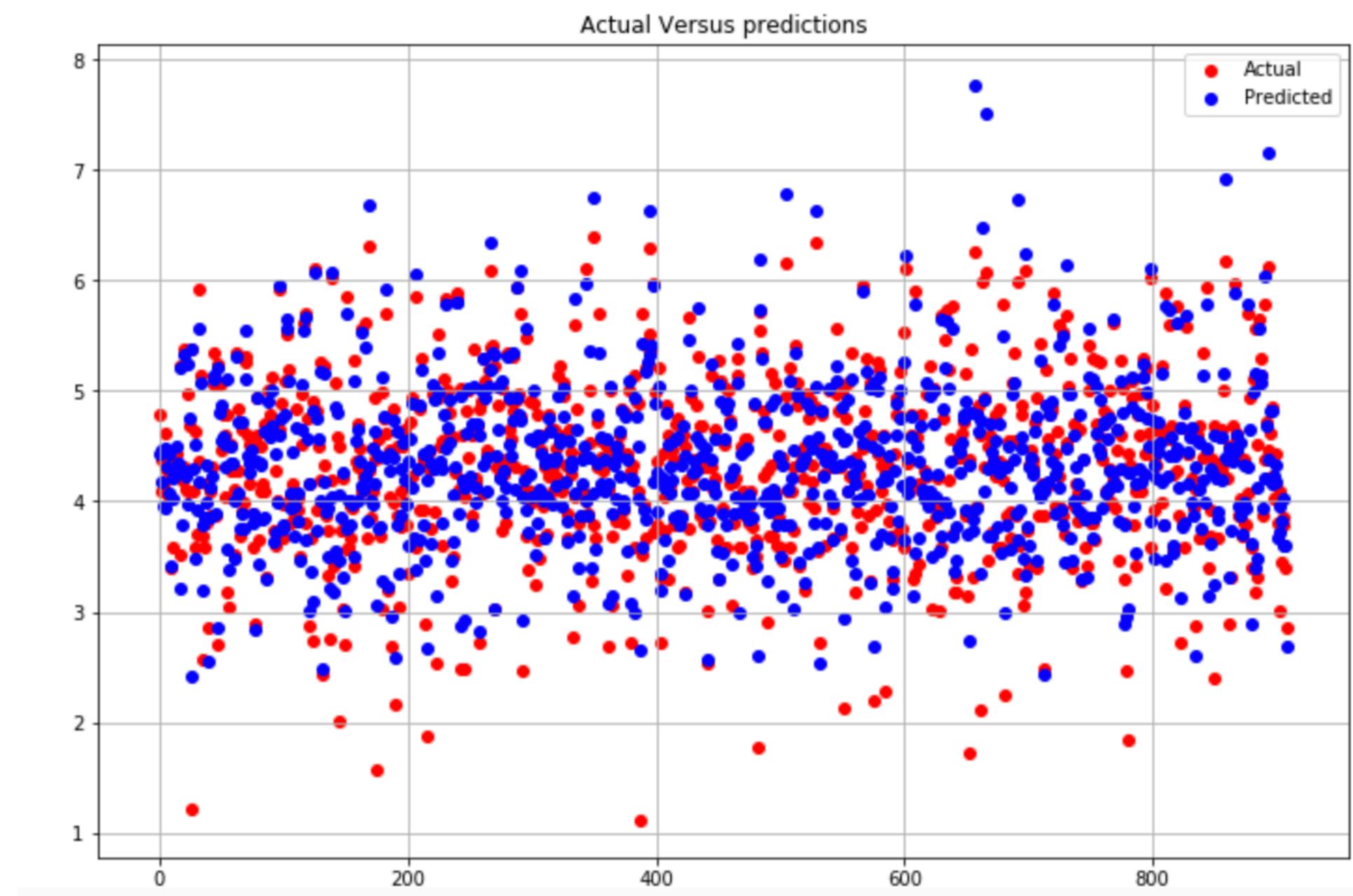
*1) Converted the 4g , 5g from yes/no to 1/0*

*2) created the get\_dummies for both brand\_name & OS*

# Model #1

*Model #1 : Without any feature engineering. Selected all the features*

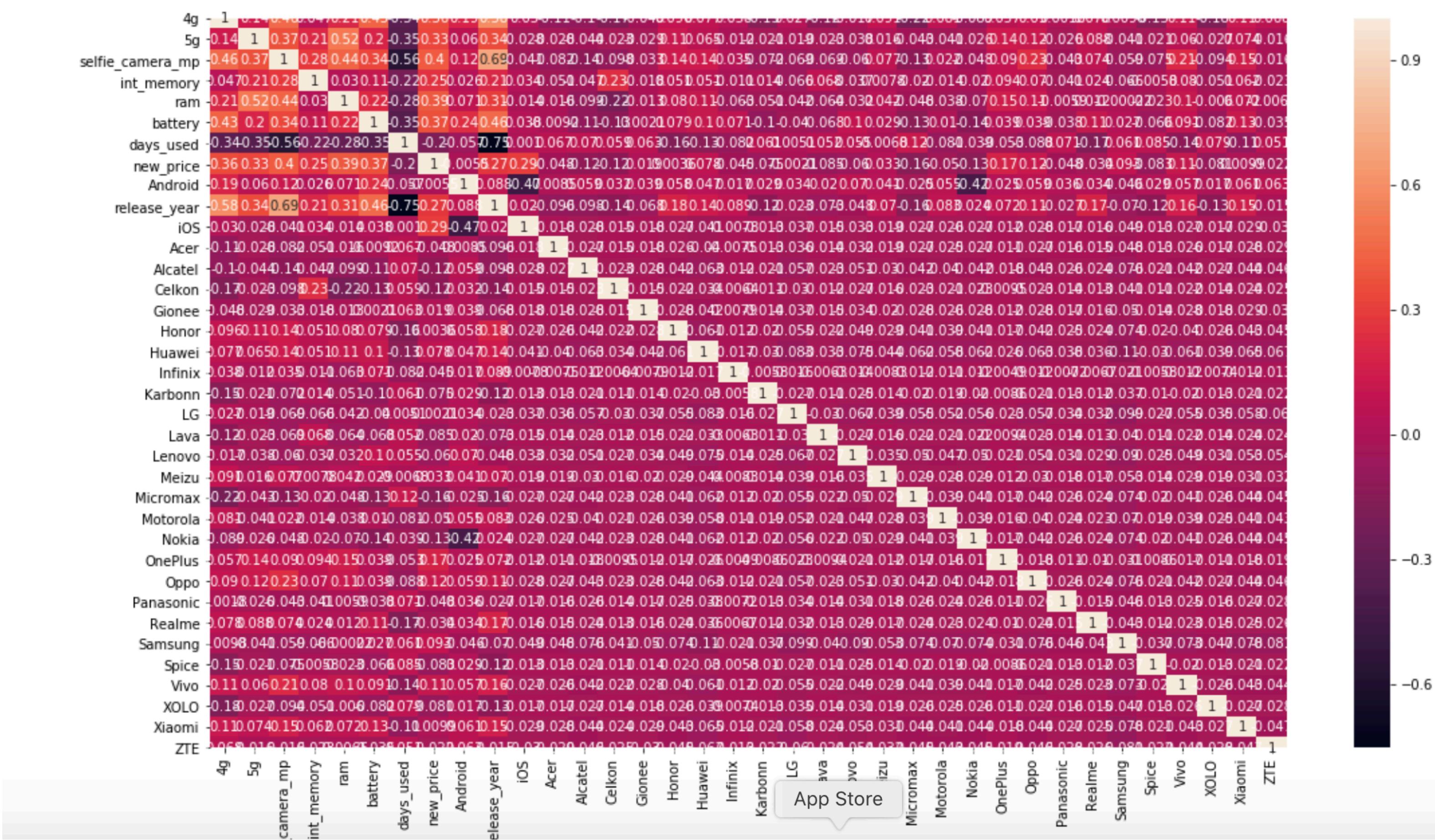
- > Both the training and test score 0.94
- > the RMSE of the test predicted value against the actual value : 21.35
- > The R2 value of the predicted value : 0.941



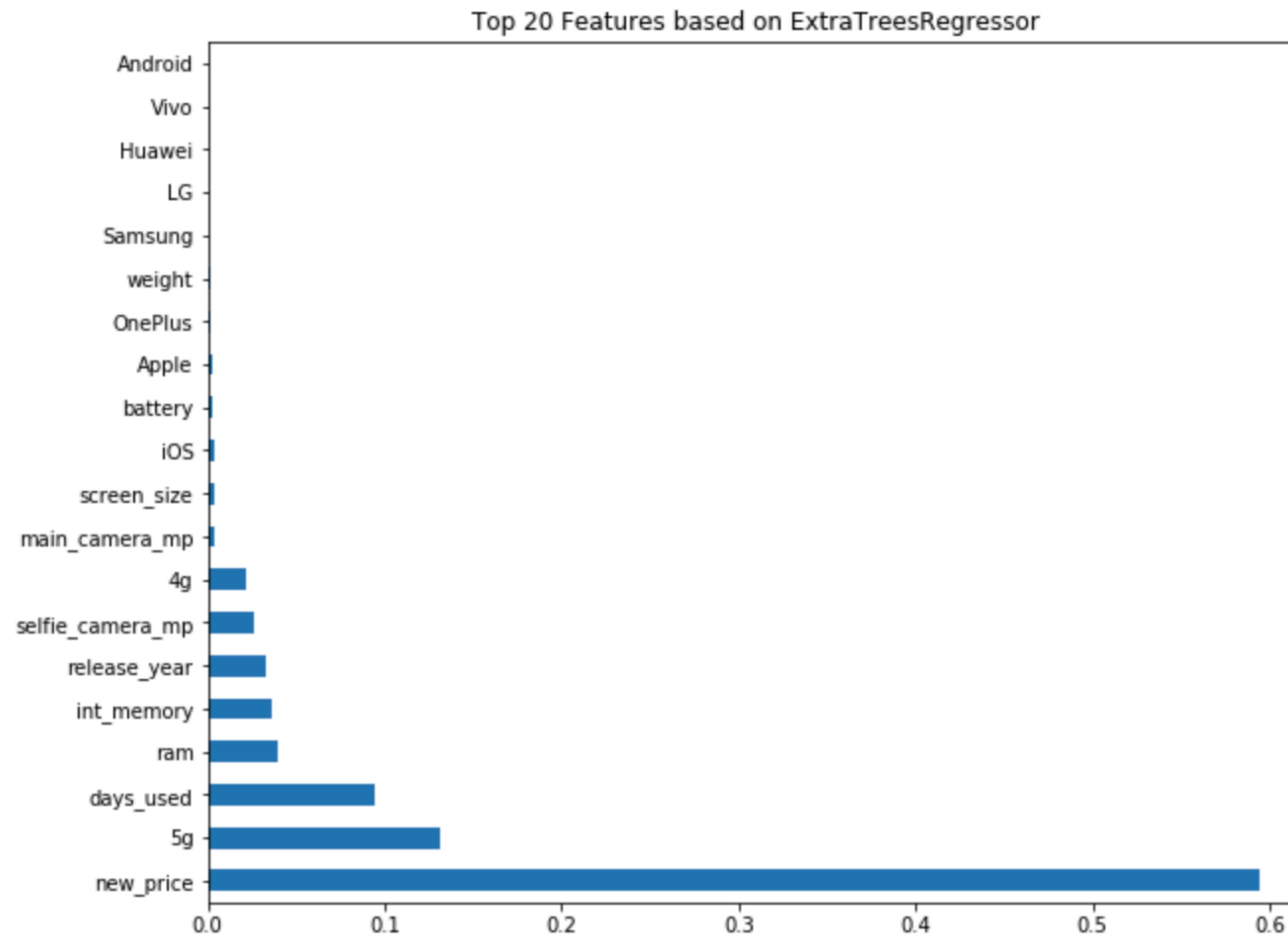
```
explained_variance: 0.8969  
mean_squared_log_error: 0.0035  
r2: 0.8966  
MAE: 0.1756  
MSE: 0.066  
RMSE: 0.2569
```

# Feature Engineering

Based on the heat map the features which are having high co-relation with each other (Multi Collinearity) , in thane variable is selected based on the high co-effecient with the used price column



# Top 20 features - Based on the ExtraTreesRegressor



# New Model on the selected feature

## *New Prediction Model Summary*

```
explained_variance: 0.8736  
mean_squared_log_error: 0.0042  
r2: 0.8731  
MAE: 0.1899  
MSE: 0.081  
RMSE: 0.2846
```

## *Old Prediction Model Summary*

```
explained_variance: 0.8969  
mean_squared_log_error: 0.0035  
r2: 0.8966  
MAE: 0.1756  
MSE: 0.066  
RMSE: 0.2569
```

*The old model without feature engineering works better than the feature engineered prediction model.*

**THANK  
YOU**