Used Device Price Prediction

ReCell: Supervised Learning Foundations

Contents

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

Executive Summary

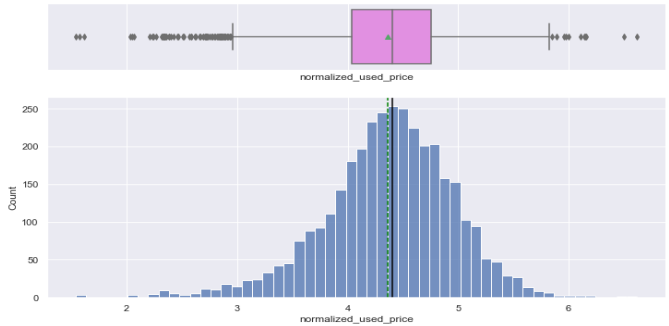
- The model explains ~84% of the variation in the data and can predict the normalized used device price within ~4.5% and can be used for predictive purposes
- Based on the most significant predictors of the normalized used device price, ReCell can consider the following steps to improve the business:
    - Attract people who want to sell used phones and tablets which were originally released in recent years and have good front and rear camera resolutions
    - Sell devices with more RAM and 4G connectivity as they are good candidates for reselling to certain customer segments
    - Try to gather and put up phones having a high price for new models to try and increase revenue
- 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 price segments
- ReCell can also look to sell other used gadgets, like smart watches, which might attract certain segments of customers

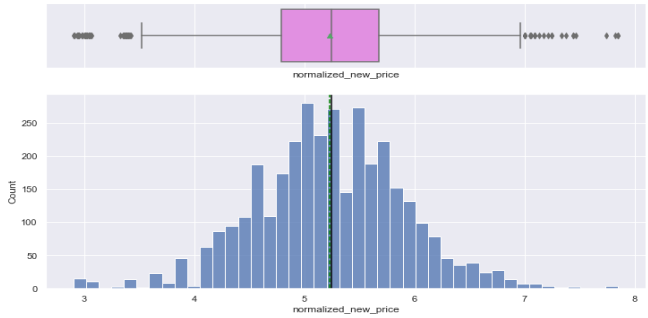
Business Problem Overview and Solution Approach

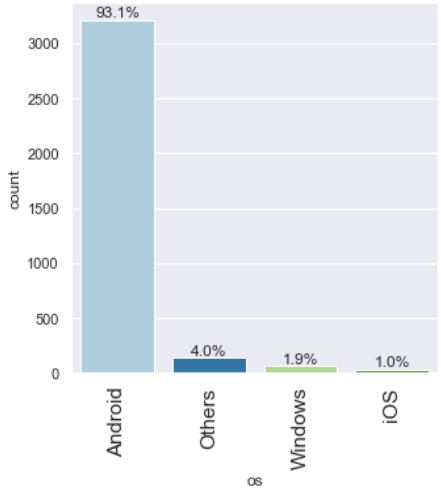
- ReCell is a startup aiming to tap the rising potential of the comparatively under-the-radar used phone and tablet market, which is predicted to be worth $52.7bn by 2023 by the International Data Corporation
- Refurbished and used devices provide cost-effective alternatives to both consumers and businesses that are looking to save money when purchasing a phone or tablet. They can be sold with warranties and can also be insured with proof of purchase. Maximizing the longevity of devices through second-hand trade also reduces their environmental impact and helps in recycling and reducing waste
- With third-party vendors/platforms, such as Verizon, Amazon, etc., providing attractive offers for refurbished phones and tablets, along with the impact of the COVID-19 outbreak, the used and refurbished device market has a good chance of getting a significant boost
- The task at hand is to analyze the data provided and develop a dynamic pricing strategy for used and refurbished phones and tablets using a linear regression model and identify factors that significantly influence the price of a used device

EDA Results

- The normalized used device prices and the normalized prices of new device models are almost normally distributed
- Android devices dominate ~93% of the used device market



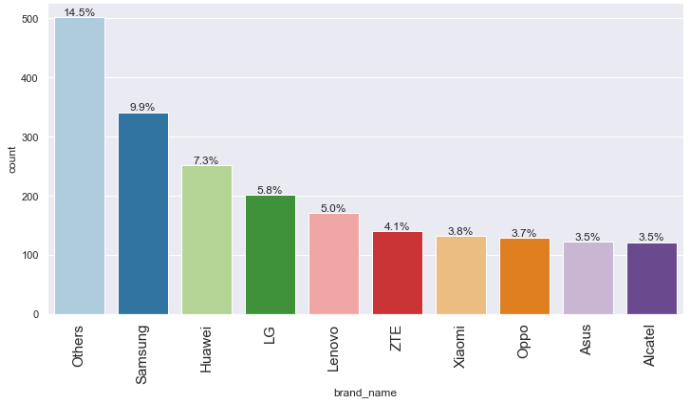




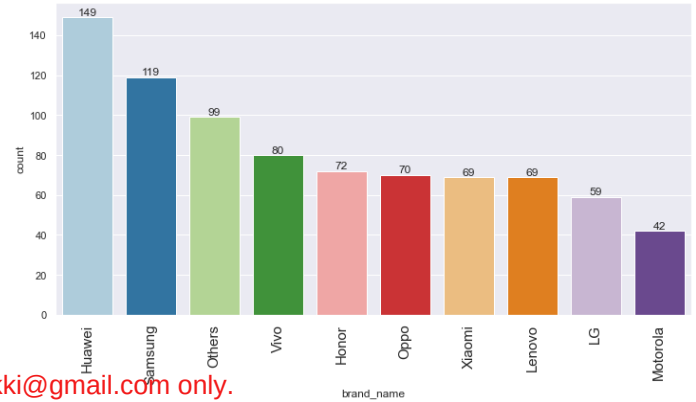
Link to Appendix slide on data background check

EDA Results

- Samsung has the most number of devices in the data, followed by Huawei and LG
- 14.5% of the devices in the data are from brands other than the listed ones

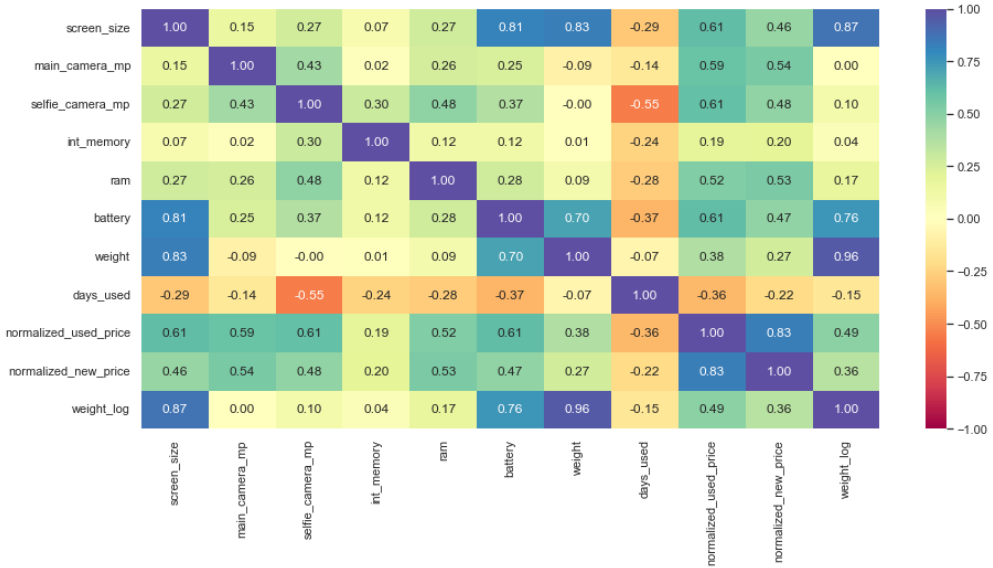


- Huawei and Samsung offer a lot of devices with large screens suitable for customers buying phones and tablets for entertainment purposes



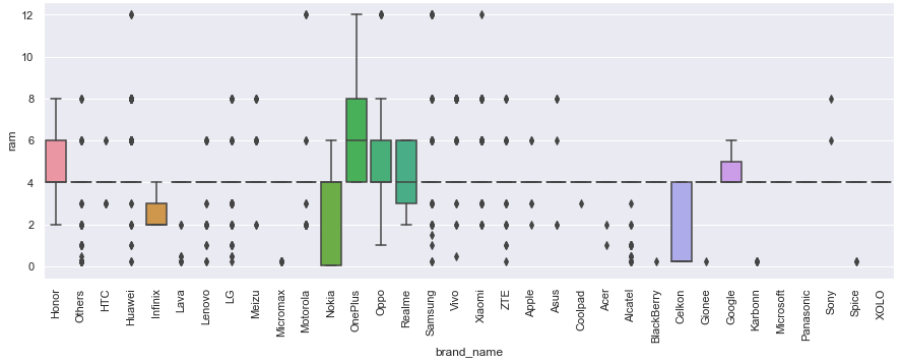
EDA Results

- The normalized used device price is highly correlated with the normalized price of a new device model
    - This makes sense as the price of a new model is likely to affect the used device price
- The normalized used device price is also moderately correlated with the resolution of the cameras provided, the size of the screen, and the battery capacity
- Weight, screen size, and battery capacity of a device show a good amount of correlation
    - This makes sense as larger battery capacity requires bigger space, thereby increasing screen size and weight

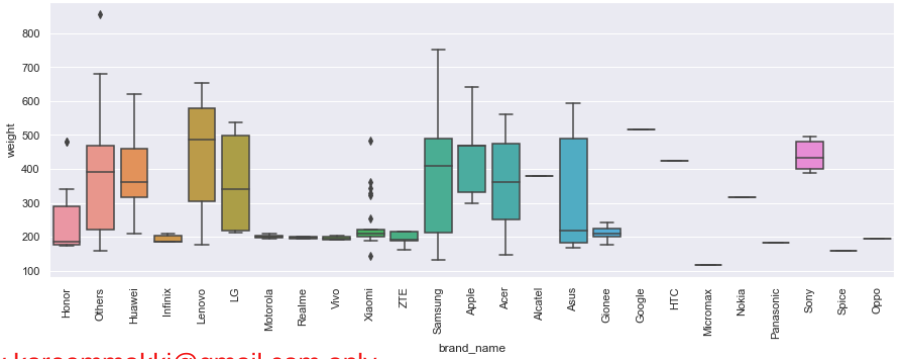


EDA Results

- 50% of the devices offered by most of the companies have 4GB of RAM
- 50% of OnePlus devices have 6GB or more RAM, indicating that OnePlus devices offer more RAM in general

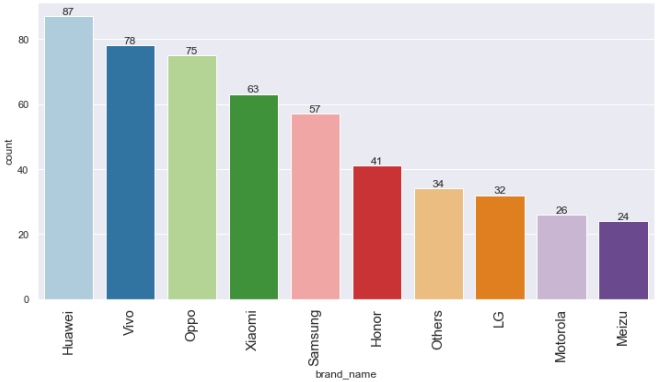


- Some devices offered by brands like Vivo, Realme, Motorola, etc. weigh just about 200g but offer great batteries
- Some devices offered by brands like Huawei, Apple, Sony, etc. offer great batteries but are heavier

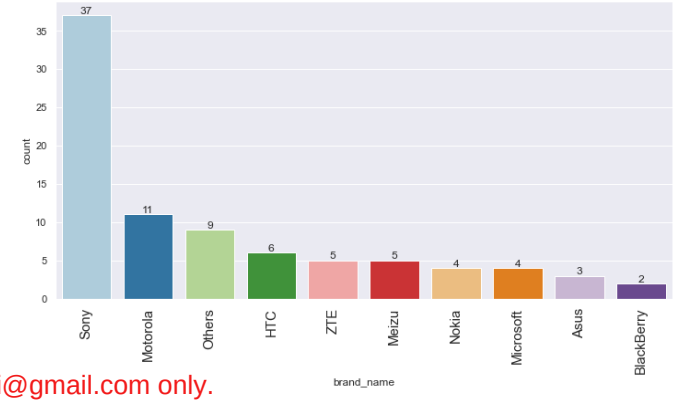


EDA Results

- Huawei is the go-to brand for the customers specifically looking for good front cameras to click cool selfies as they offer many devices with powerful front cameras
- Oppo and Vivo also offer a lot of devices suitable for this customer segment

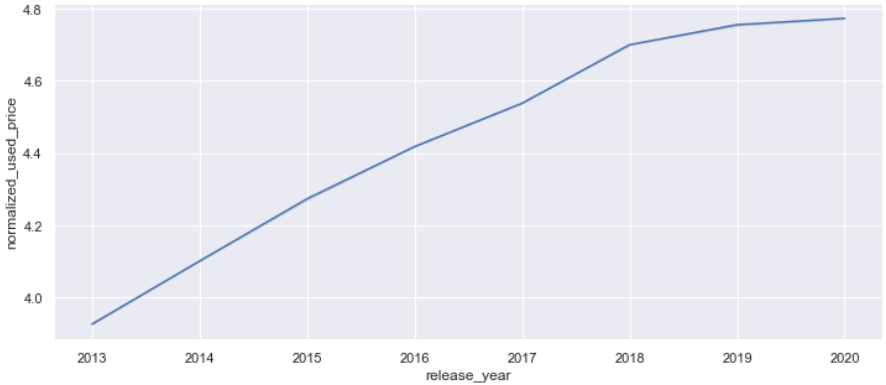


- Sony is the go-to brand for great rear cameras as they offer many devices under this criteria

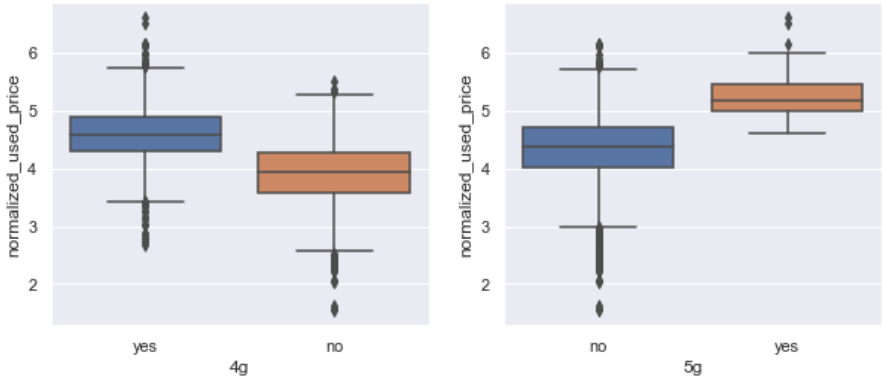


EDA Results

- The price of used devices has increased over the years



- Used devices offering 4G and 5G networks are generally priced higher than the ones without 4G and 5G networks



Data Preprocessing

- There are no duplicate values in the data
- The missing values in the data were imputed as follows:
    - Initially, missing values will be imputed by the median grouped by release year and brand
    - The remaining missing values will be imputed by the median grouped by brand
    - Any leftover missing values will be imputed by the column median
- There are outliers in a few columns but they are not treated as they are proper values
- A column years_since_release was created from the release_year column considering the year of data collection, 2021, as the baseline

Model Performance Summary

- A linear regression model was built with the normalized used device price as the target variable
- The most significant predictors of the normalized used device price are the normalized price of a new device of the same model, the weight of the devices screen, the resolution of the rear and front cameras, the years since the original release of the device, the amount of RAM, the availability of 4G and 5G network
- The key metrics for assessing model performance have been summarized in the table below

Data	R^2	Adj. R^2	RMSE	MAE	MAPE
Train	0.839	0.838	0.234	0.183	4.395
Test	0.838	0.836	0.241	0.187	4.556

Link to Appendix slide on model assumptions

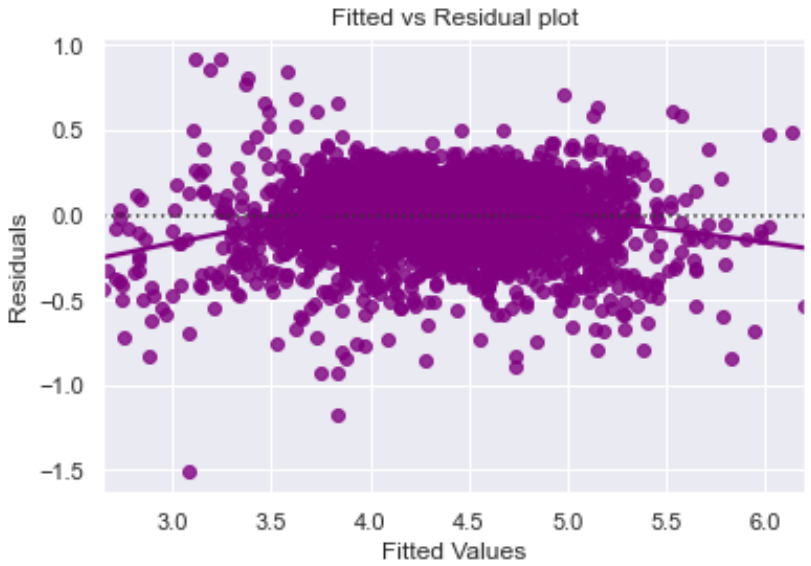
APPENDIX

Data Background and Contents

- The data contains information about 3454 phones and tablets and their characteristics
- The characteristics include brand name, availability of 4G and 5G, screen size, weight, storage capacity, amount of RAM, normalized used model price, normalized new model price, and more
- There are a few missing values in some numerical columns
- A few notable points from the statistical summary of the data:
    - The weight ranges from 69g to 855g, which does not seem incorrect as the data contains feature phones and tablets too
    - There are a few unusually low values for the internal memory and RAM of used devices, but those are likely due to the presence of feature phones in the data
    - The average value of the normalized price of a used device is approx. 4.4 euros

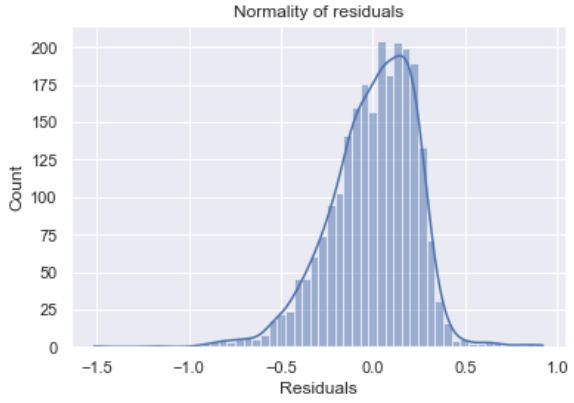
Model Assumptions

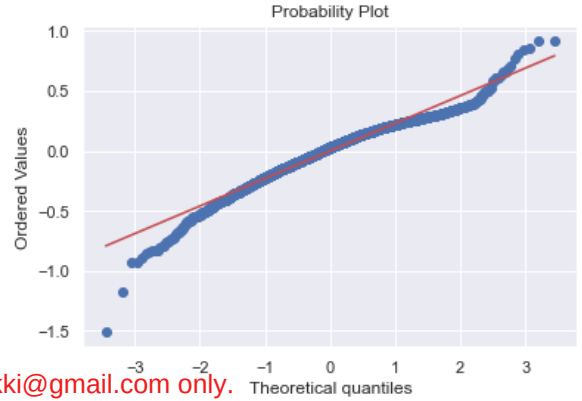
- Multicollinearity was checked for using VIF
  - All numerical variables with VIF > 5 were dropped
  - VIF for the constant and dummy variables were ignored
- For linearity and independence tests, the residuals vs fitted values plot was checked
  - The plot does not show any pattern, so the assumption is satisfied



Model Assumptions

- For normality test, the distribution of residuals, the Q-Q plot, and also applied the Shapiro-Wilk test were checked
    - The distribution does seem close to normal as per the plots
    - The Shapiro-Wilk test does suggest that the distribution is not strictly normal, but we will accept this an approximately normal distribution
- For homoscedasticity test, the goldfeldquadt test was applied and a p-value of 0.4402 was obtained, which shows that the residuals are homoscedastic





Final Model Summary

OLS Regression Results

Dep. Variable:	normalized_used_price	R-squared:	0.839
Model:	OLS	Adj. R-squared:	0.838
Method:	Least Squares	F-statistic:	895.7
Date:	Thu, 03 Mar 2022	Prob (F-statistic):	0.00
Time:	01:37:23	Log-Likelihood:	80.645
No. Observations:	2417	AIC:	-131.3
Df Residuals:	2402	BIC:	-44.44
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.5000	0.048	30.955	0.000	1.405	1.595
main_camera_mp	0.0210	0.001	14.714	0.000	0.018	0.024
selfie_camera_mp	0.0138	0.001	12.858	0.000	0.012	0.016
ram	0.0207	0.005	4.151	0.000	0.011	0.030
weight	0.0017	6e-05	27.672	0.000	0.002	0.002
normalized_new_price	0.4415	0.011	39.337	0.000	0.419	0.463
years_since_release	-0.0292	0.003	-8.589	0.000	-0.036	-0.023
brand_name_Karbonn	0.1156	0.055	2.111	0.035	0.008	0.223
brand_name_Samsung	-0.0374	0.016	-2.270	0.023	-0.070	-0.005
brand_name_Sony	-0.0670	0.030	-2.197	0.028	-0.127	-0.007
brand_name_Xiaomi	0.0801	0.026	3.114	0.002	0.030	0.130
os_Others	-0.1276	0.027	-4.667	0.000	-0.181	-0.074
os_iOS	-0.0900	0.045	-1.994	0.046	-0.179	-0.002
4g_yes	0.0502	0.015	3.326	0.001	0.021	0.080
5g_yes	-0.0673	0.031	-2.194	0.028	-0.127	-0.007

Omnibus:	246.183	Durbin-Watson:	1.902
Prob(Omnibus):	0.000	Jarque-Bera (JB):	483.879
Skew:	-0.658	Prob(JB):	8.45e-106
Kurtosis:	4.753	Cond. No.	2.39e+03

Great Learning

Happy Learning !