

HUỲNH NGỌC DUNG

PHÂN TÍCH DỮ LIỆU VỚI PYTHON
**MÔ HÌNH
HỒI QUY TUYẾN TÍNH
TẬP DỮ LIỆU USED DEVICE PRICING**

GROUP 6



TẬP DỮ LIỆU USED DEVICE PRICING

Đề tài: Thị trường thiết bị đã qua sử dụng và tái chế đã phát triển đáng kể trong thập kỷ qua vì nó cung cấp các lựa chọn tiết kiệm chi phí cho cả người tiêu dùng và doanh nghiệp đang tìm cách tiết kiệm tiền khi mua một chiếc máy. Tối ưu hóa tuổi thọ của các thiết bị thông qua giao dịch thiết bị đã qua sử dụng cũng giảm thiểu tác động môi trường của chúng và giúp tái chế và giảm thiểu rác thải. Đây là một tập dữ liệu mẫu về giá cả chuẩn hóa của thiết bị đã qua sử dụng và mới của thiết bị tái chế / đã qua sử dụng.

Mục tiêu: thực hiện phân tích dữ liệu và áp dụng các mô hình dự đoán để có thể **định giá các thiết bị đã qua sử dụng.**

Tập dữ liệu như sau:

	device_brand	os	screen_size	4g	5g	rear_camera_mp	front_camera_mp	internal_memory	ram	battery	weight	release_year	days_used	normalized_used_price	normalized_new_price
0	Honor	Android	14.50	yes	no	13.0	5.0	64.0	3.0	3020.0	146.0	2020	127	4.307572	4.715100
1	Honor	Android	17.30	yes	yes	13.0	16.0	128.0	8.0	4300.0	213.0	2020	325	5.162097	5.519018
2	Honor	Android	16.69	yes	yes	13.0	8.0	128.0	8.0	4200.0	213.0	2020	162	5.111084	5.884631
3	Honor	Android	25.50	yes	yes	13.0	8.0	64.0	6.0	7250.0	480.0	2020	345	5.135387	5.630961
4	Honor	Android	15.32	yes	no	13.0	8.0	64.0	3.0	5000.0	185.0	2020	293	4.389995	4.947837
...
3449	Asus	Android	15.34	yes	no	Nan	8.0	64.0	6.0	5000.0	190.0	2019	232	4.492337	6.483872
3450	Asus	Android	15.24	yes	no	13.0	8.0	128.0	8.0	4000.0	200.0	2018	541	5.037732	6.251538
3451	Alcatel	Android	15.80	yes	no	13.0	5.0	32.0	3.0	4000.0	165.0	2020	201	4.357350	4.528829
3452	Alcatel	Android	15.80	yes	no	13.0	5.0	32.0	2.0	4000.0	160.0	2020	149	4.349762	4.624188
3453	Alcatel	Android	12.83	yes	no	13.0	5.0	16.0	2.0	4000.0	168.0	2020	176	4.132122	4.279994

3454 rows x 15 columns

TẬP DỮ LIỆU USED DEVICE PRICING

Tập dữ liệu gồm
3454 dòng
15 cột.

```
RangeIndex: 3454 entries, 0 to 3453
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   device_brand     3454 non-null    object  
 1   os                3454 non-null    object  
 2   screen_size       3454 non-null    float64
 3   4g                3454 non-null    object  
 4   5g                3454 non-null    object  
 5   rear_camera_mp   3275 non-null    float64
 6   front_camera_mp  3452 non-null    float64
 7   internal_memory  3450 non-null    float64
 8   ram               3450 non-null    float64
 9   battery           3448 non-null    float64
 10  weight            3447 non-null    float64
 11  release_year     3454 non-null    int64   
 12  days_used        3454 non-null    int64   
 13  normalized_used_price 3454 non-null    float64
 14  normalized_new_price 3454 non-null    float64
dtypes: float64(9), int64(2), object(4)
```

Tên cột	Loại dữ liệu	Mô tả
device_brand	object	Tên hãng sản xuất
os	object	Hệ điều hành mà thiết bị chạy trên đó
screen_size	float64	Kích thước màn hình tính bằng cm
4g	object	Có 4G hay không
5g	object	Có 5G hay không
rear_camera_mp	float64	Độ phân giải của camera sau tính bằng megapixel
front_camera_mp	float64	Độ phân giải của camera trước tính bằng megapixel
internal_memory	float64	Dung lượng bộ nhớ trong (ROM) tính bằng GB
ram	float64	Dung lượng RAM tính bằng GB
battery	float64	Dung lượng năng lượng của pin thiết bị tính bằng mAh
weight	float64	Trọng lượng của thiết bị tính bằng gam
release_year	int64	Năm mẫu thiết bị được phát hành
days_used	int64	Số ngày thiết bị đã qua sử dụng/tân trang đã được sử dụng
normalized_used_price	float64	Giá bình thường của thiết bị đã qua sử dụng/tân trang
normalized_new_price	float64	Giá bình thường của một thiết bị mới cùng kiểu máy

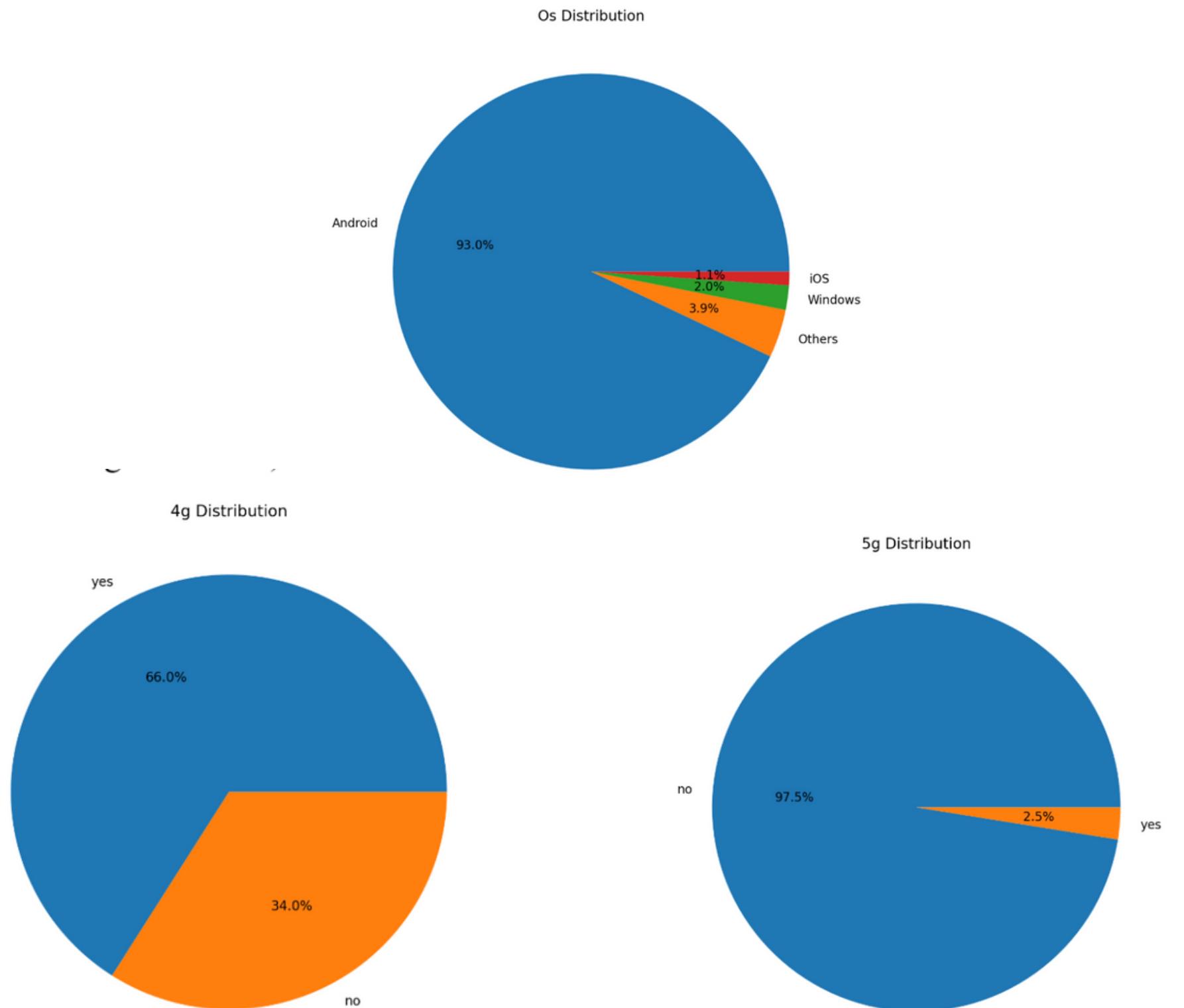
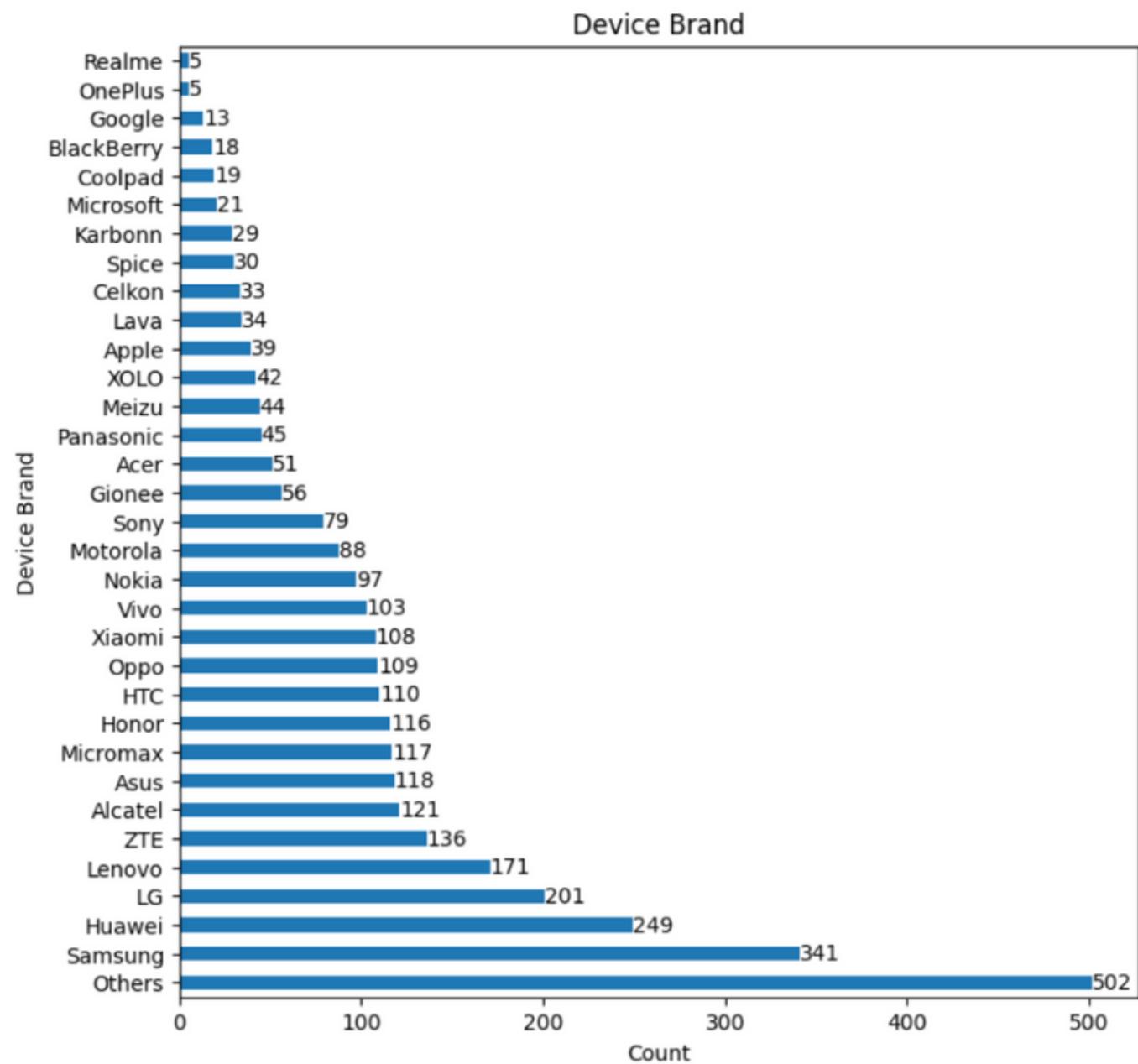
TẬP DỮ LIỆU USED DEVICE PRICING

Sau khi xử lý các giá trị rỗng và ngoại lệ, bộ dữ liệu còn 3250 dòng và 15 cột

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3250 entries, 0 to 3453
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   device_brand    3250 non-null    object  
 1   os               3250 non-null    object  
 2   screen_size     3250 non-null    float64
 3   4g               3250 non-null    object  
 4   5g               3250 non-null    object  
 5   rear_camera_mp  3250 non-null    float64
 6   front_camera_mp 3250 non-null    float64
 7   internal_memory 3250 non-null    float64
 8   ram              3250 non-null    float64
 9   battery          3250 non-null    float64
 10  weight           3250 non-null    float64
 11  release_year    3250 non-null    int64  
 12  days_used       3250 non-null    int64  
 13  normalized_used_price 3250 non-null    float64
 14  normalized_new_price 3250 non-null    float64
dtypes: float64(9), int64(2), object(4)
memory usage: 535.3+ KB
```

TẬP DỮ LIỆU USED DEVICE PRICING

Biến phân loại



TẬP DỮ LIỆU USED DEVICE PRICING

Biến liên tục

	count	mean	std	min	25%	50%	75%	max
screen_size	3250.0	13.611351	3.848012	5.080000	12.700000	12.830000	15.320000	30.710000
rear_camera_mp	3250.0	9.474948	4.804145	0.080000	5.000000	8.000000	13.000000	48.000000
front_camera_mp	3250.0	6.083415	6.593530	0.000000	2.000000	5.000000	8.000000	32.000000
internal_memory	3250.0	51.674108	84.330528	0.010000	16.000000	32.000000	64.000000	1024.000000
ram	3250.0	3.958003	1.214194	0.020000	4.000000	4.000000	4.000000	12.000000
battery	3250.0	3073.750000	1296.961103	500.000000	2100.000000	3000.000000	3950.000000	9720.000000
weight	3250.0	182.403323	90.348620	69.000000	140.450000	158.000000	181.375000	855.000000
release_year	3250.0	2015.762769	2.185467	2013.000000	2014.000000	2015.000000	2018.000000	2020.000000
days_used	3250.0	696.074462	235.190458	91.000000	564.000000	706.000000	879.000000	1094.000000
normalized_used_price	3250.0	4.340524	0.576362	1.536867	4.018992	4.377328	4.724973	6.158969
normalized_new_price	3250.0	5.215134	0.667324	2.901422	4.788970	5.242170	5.637696	7.741043

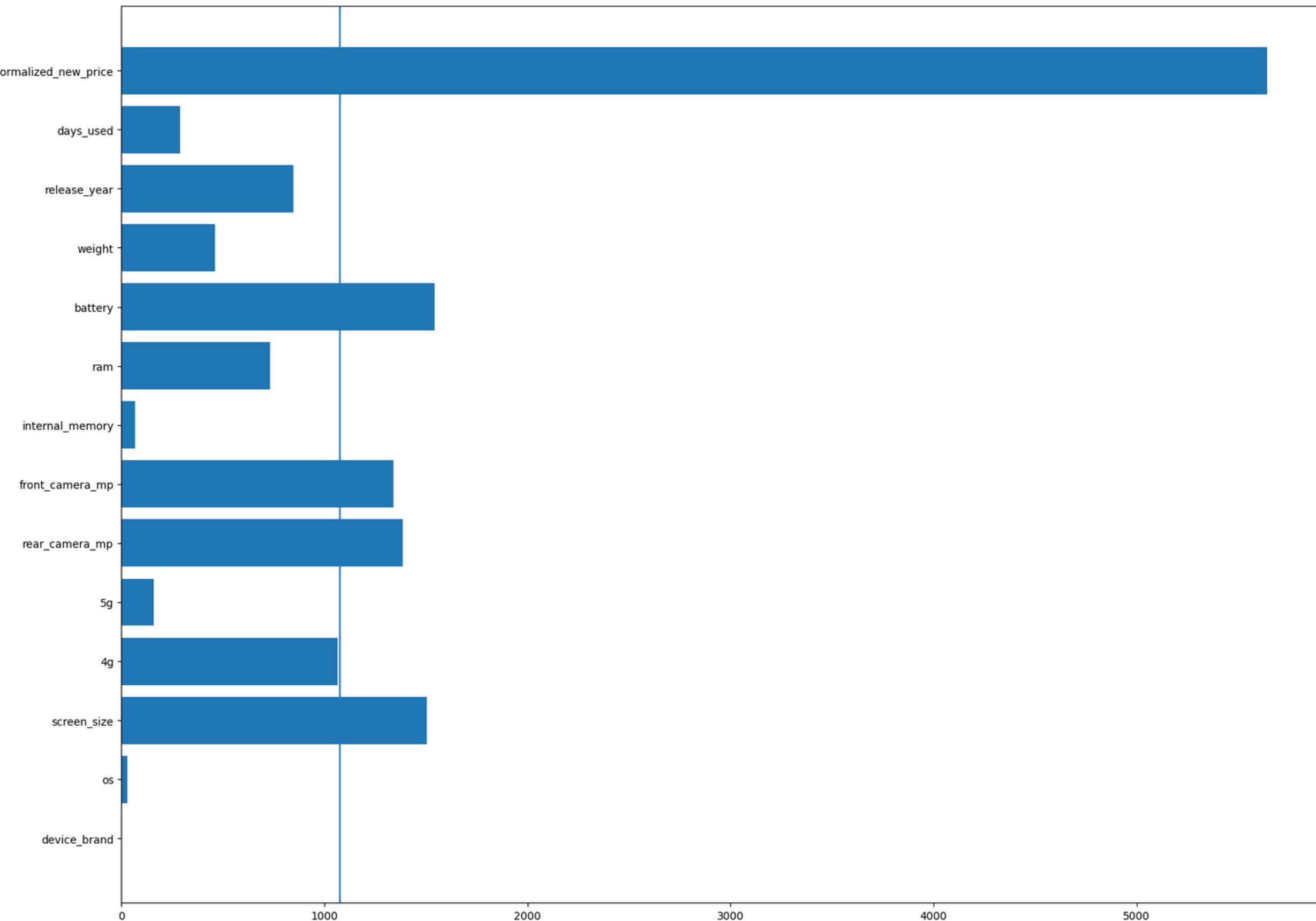
TẬP DỮ LIỆU USED DEVICE PRICING

Tổng quan về tiêu chuẩn của các thiết bị cũ được mua nhiều nhất:

- Thương hiệu: Samsung (Vì Others là tổng hợp của nhiều loại thương hiệu khác)
- Hệ điều hành: Android
- 4G: Có
- 5G: Có
- Kích thước màn hình: 12,8cm
- Camera sau: 8 MP
- Camera trước: 5MP
- Bộ nhớ trong: 32GB
- RAM: 4GB
- Pin: 3000 mAh
- Trọng lượng: 158 g
- Năm phát hành: 2015
- Số ngày sử dụng: 706 ngày

TẬP DỮ LIỆU USED DEVICE PRICING

Sử dụng
phương pháp
SelectKBest để
chọn ra các đặc
trưng quan
trọng từ một
tập dữ liệu.



TẬP DỮ LIỆU USED DEVICE PRICING

CÁC BIẾN ĐƯỢC CHỌN

Mô hình đơn biến

“normalized_new_price”

Mô hình đa biến

“normalized_new_price”

“battery”

“screen_size”

“rear_camera_mp”

“front_camera_mp”

MÔ HÌNH HỒI QUY TUYẾN TÍNH

Mô hình hồi quy tuyến tính đơn biến

OLS Regression Results						
Dep. Variable:	normalized_used_price	R-squared:	0.685			
Model:	OLS	Adj. R-squared:	0.685			
Method:	Least Squares	F-statistic:	5641.			
Date:	Sat, 09 Dec 2023	Prob (F-statistic):	0.00			
Time:	05:05:49	Log-Likelihood:	-720.34			
No. Observations:	2600	AIC:	1445.			
Df Residuals:	2598	BIC:	1456.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.6178	0.050	12.372	0.000	0.520	0.716
normalized_new_price	0.7132	0.009	75.109	0.000	0.695	0.732
Omnibus:	125.793	Durbin-Watson:	2.009			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	149.193			
Skew:	-0.520	Prob(JB):	4.01e-33			
Kurtosis:	3.543	Cond. No.	43.4			

MÔ HÌNH HỒI QUY TUYẾN TÍNH

Mô hình hồi quy tuyến tính đa biến

OLS Regression Results

Dep. Variable:	normalized_used_price	R-squared:	0.825			
Model:	OLS	Adj. R-squared:	0.824			
Method:	Least Squares	F-statistic:	2442.			
Date:	Sat, 09 Dec 2023	Prob (F-statistic):	0.00			
Time:	05:05:49	Log-Likelihood:	43.575			
No. Observations:	2600	AIC:	-75.15			
Df Residuals:	2594	BIC:	-39.97			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.1868	0.042	28.551	0.000	1.105	1.268
normalized_new_price	0.4297	0.010	44.611	0.000	0.411	0.449
battery	2.169e-05	6.43e-06	3.372	0.001	9.08e-06	3.43e-05
screen_size	0.0395	0.002	18.271	0.000	0.035	0.044
rear_camera_mp	0.0217	0.001	17.980	0.000	0.019	0.024
front_camera_mp	0.0170	0.001	20.217	0.000	0.015	0.019
Omnibus:	241.434	Durbin-Watson:	1.921			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	551.739			
Skew:	-0.564	Prob(JB):	1.55e-120			
Kurtosis:	4.955	Cond. No.	3.02e+04			

MÔ HÌNH HỒI QUY TUYẾN TÍNH

Dự đoán sử dụng mô hình hồi quy tuyến tính đa biến

```
normalized_new_price = 10  
battery = 3000  
screen_size = 10  
rear_camera_mp = 8  
front_camera_mp = 5  
  
# Thực hiện dự đoán  
predicted_price = a + b * normalized_new_price + c * battery + d * screen_size + e * rear_camera_mp + f * front_camera_mp  
  
print("Dự đoán giá điện thoại cũ:", predicted_price)
```

Dự đoán giá điện thoại cũ: 6.202999948142564

Giả sử thiết bị đã qua sử dụng với

- giá hiện hành là 10 USD
- dung lượng pin là 3000mAh
- kích cỡ màn hình là 10 inch
- Camera sau là 8 MP
- Camera trước là 5MP

=>

**KẾT QUẢ DỰ ĐOÁN
THIẾT BỊ ĐÓ SẼ ĐƯỢC
BÁN LẠI VỚI GIÁ 6 USD**