

TRỰC QUAN

DỮ LIỆU

**THÔNG TIN ĐIỆN THOẠI VÀ
NHU CẦU NÂNG CẤP**

NỘI DUNG

I

Giới thiệu đề tài

II

**Trực quan hóa dữ
liệu**

III

**Phân tích tương
quan**

IV

Thử nghiệm phân lớp

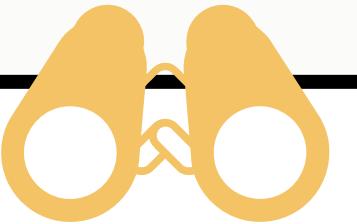
V

Kết luận

I

GIỚI THIỆU ĐỀ TÀI VÀ MỤC TIÊU

BỐI CẢNH



- Thị trường smartphone toàn cầu bước vào giai đoạn bão hòa, chu kỳ thay thế thiết bị kéo dài
- Thách thức: duy trì doanh số, lòng trung thành thương hiệu và động lực nâng cấp thiết bị

MỤC TIÊU



Ứng dụng Data Visualization để:

- Xác định ai có xu hướng nâng cấp
- Phân tích thời điểm nâng cấp
- Tìm hiểu yếu tố ảnh hưởng đến quyết định nâng cấp

PHẠM VI DỮ LIỆU VÀ NGHIÊN CỨU

PHẠM VI & DỮ LIỆU

- Dữ liệu khảo sát mô phỏng từ nhiều thị trường
- Phân tích các nhóm yếu tố:
- Nhân khẩu học
- Kinh tế – thu nhập
- Tình trạng thiết bị & hành vi sử dụng
- Động lực và định hướng nâng cấp

Ý NGHĨA NGHIÊN CỨU

Hỗ trợ doanh nghiệp:

- Đưa ra chiến lược marketing phù hợp
- Cá nhân hóa chương trình nâng cấp & chăm sóc khách hàng
- Nhận diện phân khúc khách hàng tiềm năng

PHƯƠNG PHÁP NGHIÊN CỨU

- Exploratory Data Analysis (EDA)
- Kỹ thuật trực quan hóa
- Phân tích đơn biến (Univariate)
- Phân tích hai biến (Bivariate)
- Tổng hợp insight & đề xuất chiến lược

GIỚI THIỆU BỘ DỮ LIỆU

Lưu Tự động • iphone_upgrade_dirty... • Đã lưu vào PC này

Tìm kiếm

Tệp Trang đầu Chèn Vẽ Bố trí Trang Công thức Dữ liệu Xem lại Xem Tự động hóa Trợ giúp

Chú thích Chia sẻ

Dán Phông chữ Căn chỉnh Số Kiểu Ô Chính sửa Mức nhạy cảm Phản bối... Mindjet

Sensitivity Sắp xếp Tim & Lựa & Lọc Định dạng chọn

Chèn Xóa bỏ Định dạng

Aptos Narrow 11 A^ A^ Chung Định dạng Có điều kiện Định dạng như Bảng Kiểu Ô

B I U \$ % , .00 .00

Bảng tạm

W10 fx

1	responden	age	gender	country	city_tier	income_us	current_br	device_age	battery_he	is_storage	ecosystem	ad_recall	peer_upgr	interest_ne	cracked_s	launch	expected_i	trade_in_v	budget_str	most_inter	upgrade_within_90days	
2	31070	18	Female	VN	metropolit	1105.62	Xiaomi	102	94	not full		1	many	1	Pro Max	0	#####	1655.83	51.05	145.1	camera	1
3	145765	32	Male	SG	rural	8522.67	Apple	34	100	not full		0	never	1	Pro Max	1	9/1/2022	1335.03	397.87	11	camera	0
4	104273	26	Female	VN	urban	615.43	Samsung	72	97	not full		1	many	2	Pro Max	1	#####	1303.38	161.74	185.5	AI	1
5	188667	44	Male	IN	rural	467.28	Apple	43	95	not full		3	never	2	Pro Max	0	9/3/2021	1160.22	469.1	147.9	display	0
6	62829	29	Female	VN	urban	1390.84	Apple	78	100	not full		1	once	0	Pro Max	0	#####	1437.06	330.98	79.5	camera	1
7	45940	19	Male	US	metropolit	10624.77	Other	69	100	not full		1	once	3	Pro Max	0	6/9/2019	1356.2	60.15	12.2	battery	1
8	56920	37	Male	JP	metropolit	5206.39	Other	3	100	not full		1	once	0	Base	0	6/1/2025	927.71	249.99	13	zoom	0
9	51707	38	Female	KR	metropolit	1585.41	Other	11	99	not full		3	never	1	Base	0	6/2/2024	944.52	175.22	48.5	AI	0
10	187916	42	Male	VN	rural	1131.35	Xiaomi	16	98	not full		1	once	2	Pro Max	1	#####	1400.27	167.71	108.9	zoom	1
11	104022	31	Female	US	urban	3105.73	Apple	20	99	not full		5	once	0	Pro	0	8-Sep-23	1189.53	545.22	20.7	battery	1
12	183652	24	Male	OTHER	rural	687.46	Samsung	96	93	not full		1	once	1	Pro Max	1	#####	1223.89	128.34	159.4	camera	0
13	193283	37	Male	IN	metropolit	1094.33	Vivo	109	95	not full		1	many	1	Pro Max	1	#####	1367.04	47.35	120.6	camera	1
14	179447	42	Male	US	urban	2162.25	Other	74	94	not full		0	once	2	Pro Max	1	7/5/2019	1501.26	66.37	66.4	battery	0
15	35496	27	Female	AU	metropolit	3863.06	Apple	47	97	not full		1	once	0	Pro Max	0	#####	1224.19	415.32	20.9	camera	0
16	75535	35	Female	KR	metropolit	2246.53	Xiaomi	74	95	not full		0	never	3	Pro Max	0	#####	1292.03	91.04	53.5	camera	1
17	23095	36	Male	US	metropolit	7924.21	Samsung	25	97	not full		0	many	1	Pro Max	1	#####	1418.38	276.62	14.4	battery	0
18	162820	21	Male	KR	metropolit	3718.06	Other	14	100	not full		0	never	2	Pro	0	#####	1207.34	201.98	27	display	1
19	196630	37	Male	US	metropolit	2149.29	Apple	93	98	not full		4	many	1	Pro Max	0	#####	1407.64	225.16	55	camera	0
20	29015	33	Female	JP	rural	3396.22	Samsung	95	97	not full		0	never	0	Pro Max	0	#####	1113.18	123.24	29.1	camera	0
21	24696	33	Male	VN	urban	1128.24	Xiaomi	101	93	not full		0	never	0	Pro Max	0	#####	1368.29	90.24	113.3	battery	0
22	186017	23	Female	SG	metropolit	4542.15	Apple	97	92	full		1	once	2	Pro Max	1	#####	1428.49	199.33	27.1	camera	1
23	152016	26	Female	VN	urban	510.59	Apple	44	97	not full		1	once	0	Pro Max	0	#####	1219.6	410.43	158.5	camera	0
24	81045	37	Female	JP	urban	2248.57	Samsung	26	100	FALSE		1	once	1	Pro	0	2/5/2023	1138.65	350.96	35	camera	0
25	21135	23	Male	JP	metropolit	12353.17	Other	3	94	full		0	never	4	Base	0	6/8/2025	886.36	230.09	5.3	AI	0
26	142399	23	Female	VN	urban	741.28	Samsung	49	100	not full		2	once	5	Pro Max	0	#####	1254.92	280.35	131.5	battery	0

THUỘC TÍNH

Tên thuộc tính	Mô tả	Ghi chú	
respondent_id	Mã định danh người tham gia khảo sát	Không trùng lặp	peer_upgrades_count
age	Tuổi người dùng	Dạng số, có thể xuất hiện giá trị bất thường	interest_new_model_bucket
gender	Giới tính	Male, Female	cracked_screen
country	Quốc gia sinh sống	VN, SG, IN	launch
city_tier	Mức độ phát triển khu vực	metropolitan, urban, rural	expected_price_usd
income_usd_monthly	Thu nhập trung bình tháng (USD)	Chênh lệch tùy quốc gia	trade_in_value_usd
current_brand	Thương hiệu điện thoại	Apple, Samsung, Xiaomi...	budget_stretch_pct
device_age_months	Tuổi đời thiết bị đang dùng (tháng)	Có giá trị rất lớn hoặc chưa hợp lý	most_interested_function
battery_health_pct	Phần trăm độ chai pin	0 đến 100	upgrade_within_90days
is_storage_full	Tình trạng bộ nhớ đầy	full, not full	
ecosystem_count	Số lượng thiết bị chung hệ sinh thái	0, 1, 2,...	
ad_recall	Mức độ nhớ quảng cáo về sản phẩm mới	none, few, many	

KIỂM TRA DỮ LIỆU

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200000 entries, 0 to 199999
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   respondent_id   200000 non-null   int64  
 1   age              200000 non-null   int64  
 2   gender            200000 non-null   object  
 3   country           200000 non-null   object  
 4   city_tier          200000 non-null   object  
 5   income_usd_monthly 196840 non-null   float64 
 6   current_brand      200000 non-null   object  
 7   device_age_months 200000 non-null   int64  
 8   battery_health_pct 196862 non-null   float64 
 9   is_storage_full    200000 non-null   object  
 10  ecosystem_count    200000 non-null   int64  
 11  ad_recall           200000 non-null   object  
 12  peer_upgrades_count 200000 non-null   int64  
 13  interest_new_model_bucket 198020 non-null   object  
 14  cracked_screen      200000 non-null   int64  
 15  launch              200000 non-null   object  
 16  expected_price_usd 196706 non-null   float64 
 17  trade_in_value_usd 196819 non-null   float64 
 18  budget_stretch_pct 196834 non-null   float64 
 19  most_interested_function 198852 non-null   object  
 20  upgrade_within_90days 200000 non-null   int64  
dtypes: float64(5), int64(7), object(9)
memory usage: 32.0+ MB
```

- **Dataset gồm 200,000 quan sát và 21 biến**
- **Bao gồm 12 biến định lượng và 9 biến định danh**
- **Điều chỉnh kiểu dữ liệu:**
 - **respondent_id → chuyển về object (mã định danh)**
 - **cracked_screen & upgrade_within_90days → chuyển về object (biến nhị phân)**

KIỂM TRA BIẾN ĐỊNH LƯỢNG

	count	mean	std	min	25%	50%	75%	max
age	200000.0	33.172600	9.442468	18.0	27.0000	33.000	39.00	100.0
income_usd_monthly	196840.0	3333.364152	3241.069544	50.0	1026.6900	2271.775	4788.77	35000.0
device_age_months	200000.0	59.790805	37.627435	0.0	26.0000	59.000	93.00	150.0
battery_health_pct	196862.0	96.817425	3.330598	55.0	95.0000	97.000	99.00	105.0
ecosystem_count	200000.0	1.117580	1.138434	0.0	0.0000	1.000	2.00	5.0
peer_upgrades_count	200000.0	1.353525	1.210943	0.0	0.0000	1.000	2.00	11.0
expected_price_usd	196706.0	1314.059193	201.605388	50.0	1237.5725	1346.245	1436.22	2500.0
trade_in_value_usd	196819.0	279.572826	193.571468	-50.0	139.3800	235.850	387.40	2000.0
budget_stretch_pct	196834.0	72.241962	75.957402	-10.0	20.0000	42.700	102.70	600.0

KIỂM TRA BIẾN ĐỊNH LƯỢNG

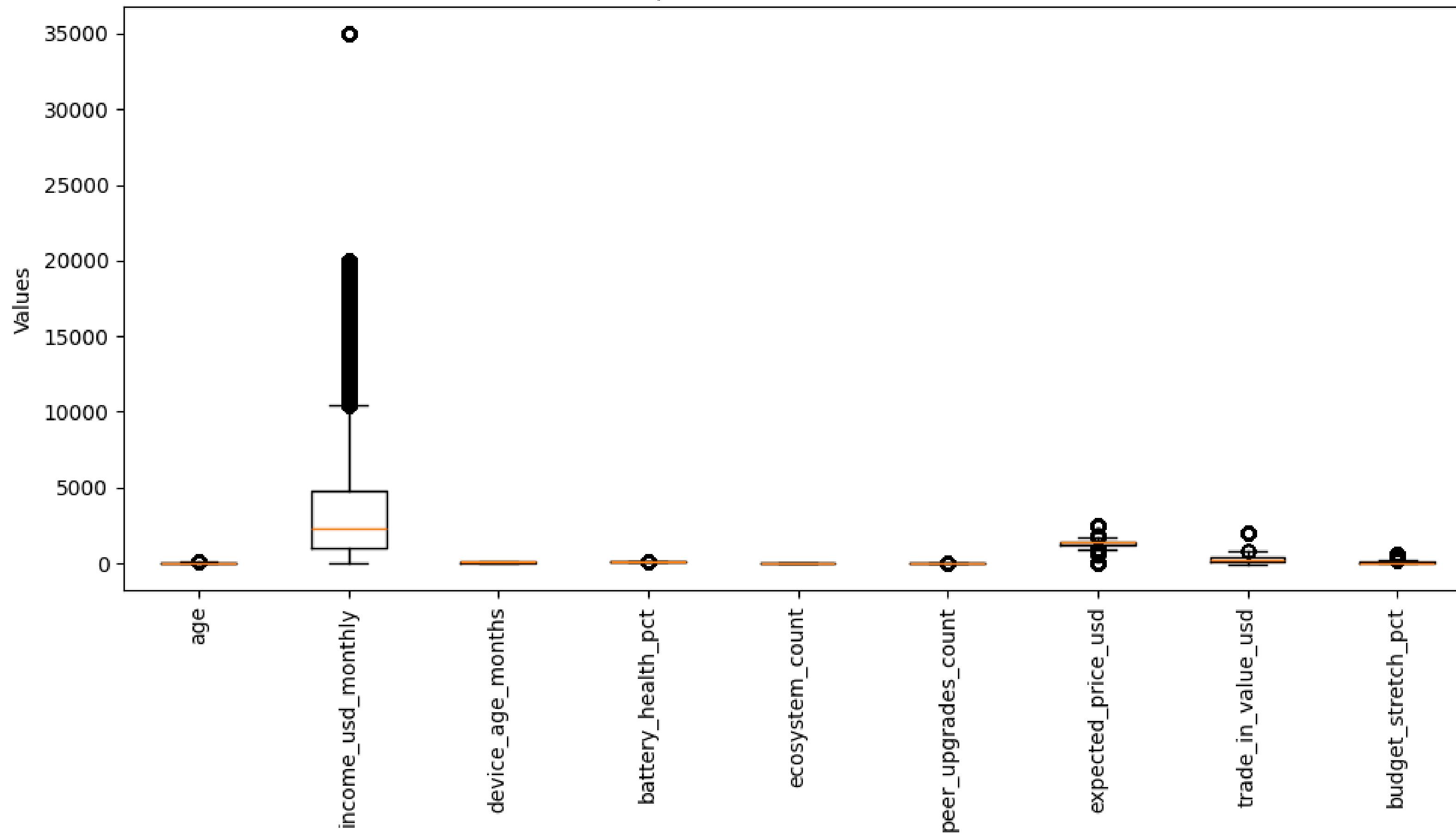
- **Thu nhập:** trung bình ~3.000 USD/tháng và độ lệch chuẩn lớn, chênh lệch thu nhập cao giữa nhóm người dùng
- **Tuổi thiết bị:** trung bình ~5 năm, nhiều thiết bị sử dụng lâu năm
- **Sức khỏe pin:** trung bình ~96,8%, hầu hết 95–99% → nhiều thiết bị pin còn tốt
- **Budget stretch:** trung bình ~72% → người dùng sẵn sàng chi vượt ngân sách
 - **Bất thường:** giá trị -10%

KIỂM TRA BIẾN ĐỊNH DANH

		count	unique	top	freq
	respondent_id	200000	192000	150316	2
	gender	200000	11	Female	94589
	country	200000	41	VN	66086
	city_tier	200000	3	metropolitan	89859
	current_brand	200000	7	Apple	88370
	is_storage_full	200000	9	not full	148471
	ad_recall	200000	3	once	78382
	interest_new_model_bucket	198020	3	Pro Max	162233
	launch	200000	5933	2025-06-02	8766
	most_interested_function	198852	10	camera	74732

KIỂM TRA OUTLIER

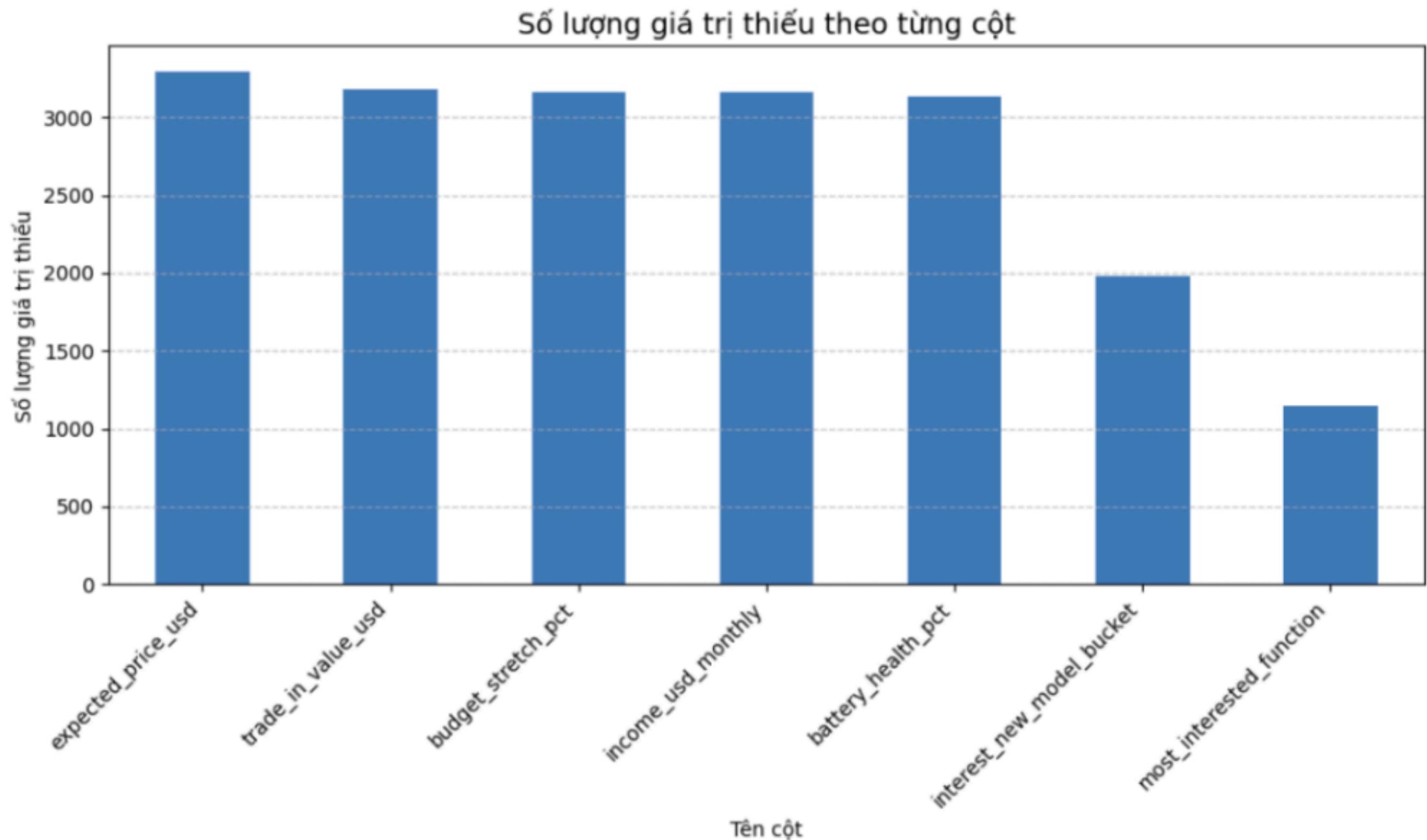
Boxplot of Numerical Features



KIỂM TRA OUTLIER

- **income_usd_monthly** có nhiều ngoại lai nhất cho thấy thu nhập biến thiên rất lớn, xuất hiện nhóm thu nhập rất cao
→ ta cần xử lý (capping/transform) trước khi phân tích & mô hình
- Các biến còn lại phân bố ổn định, ngoại lai ít và không đáng kể
- **battery_health_pct** ổn định quanh 95–100%
- **device_age_months** chủ yếu <100 tháng
- Một số giá trị bất thường nhỏ trong **trade_in_value_usd** & **budget_stretch_pct**

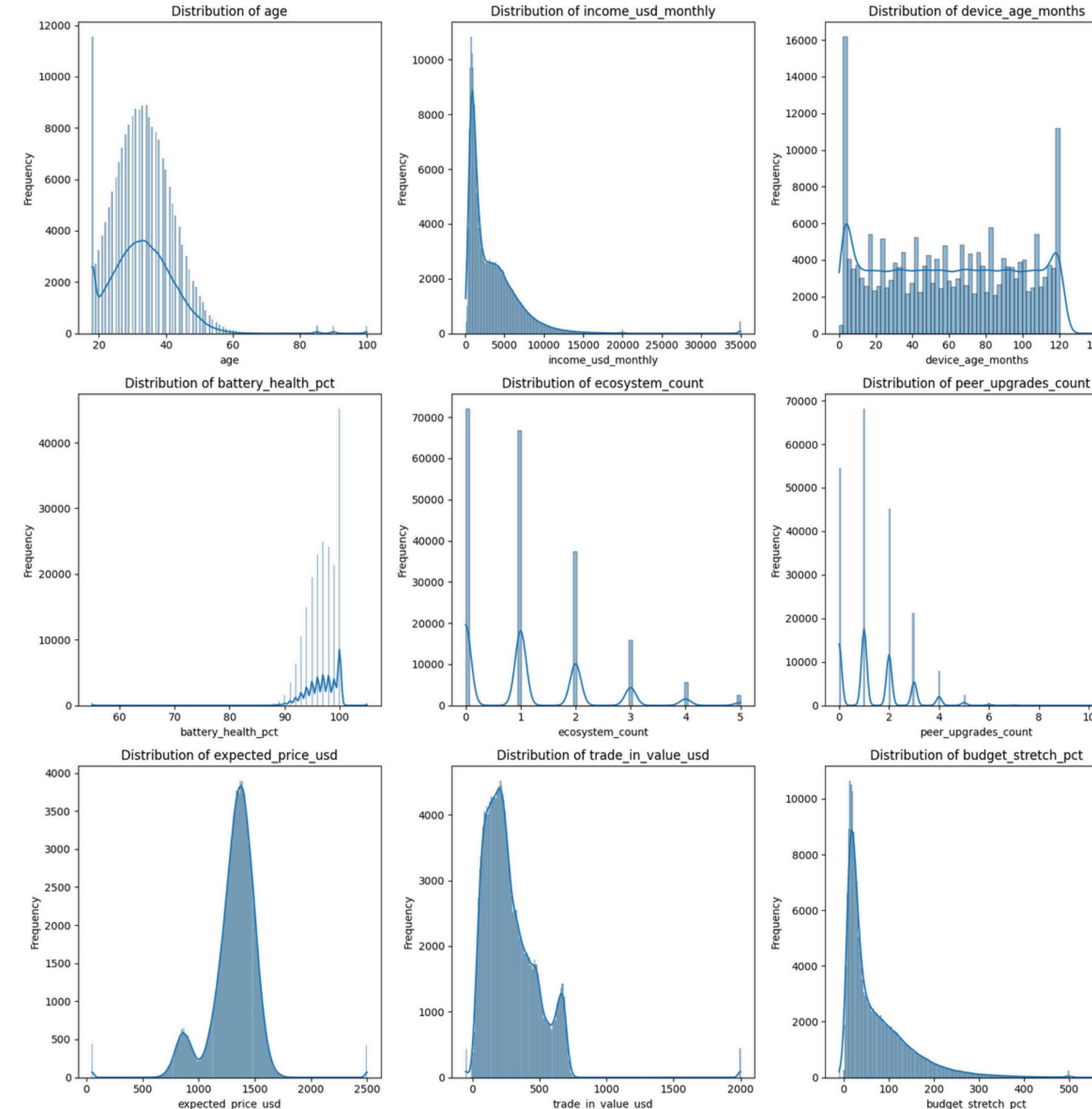
KIỂM TRA MISSING VALUE



Trong quá trình kiểm tra, các giá trị trống trong bộ dữ liệu được phát hiện như sau:

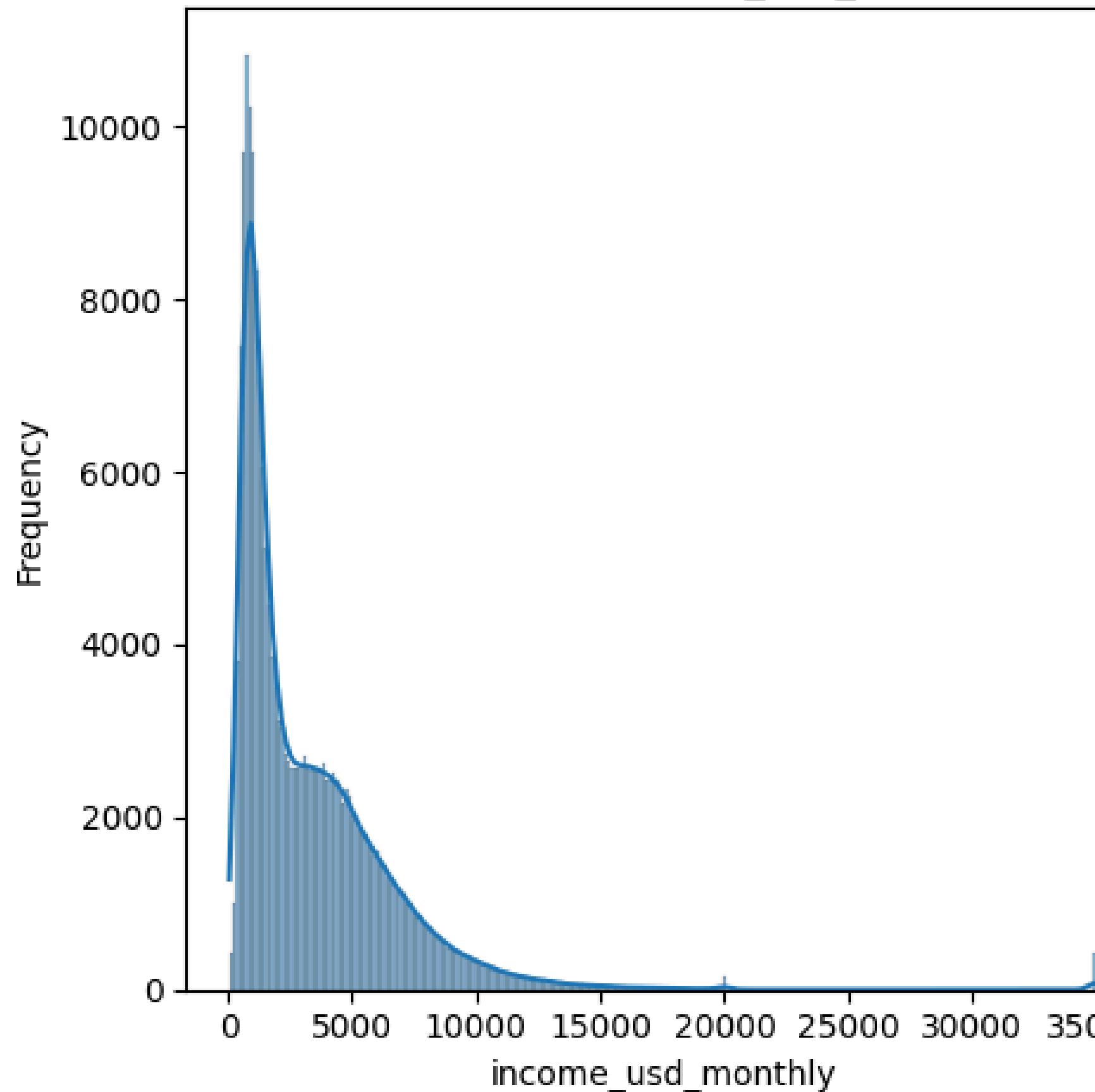
- **income_usd_monthly: 3.160 bản ghi (1,58%).**
- **battery_health_pct: 3.138 (1,57%).**
- **expected_price_usd: 3.294 (1,65%).**
- **trade_in_value_usd: 3.181 (1,59%).**
- **budget_stretch_pct: 3.166 (1,58%).**
- **interest_new_model_bucket: 1.980 (0,99%).**
- **most_interested_function: 1.148 (0,57%).**

KIỂM TRA PHÂN PHỐI



KIỂM TRA PHÂN PHỐI

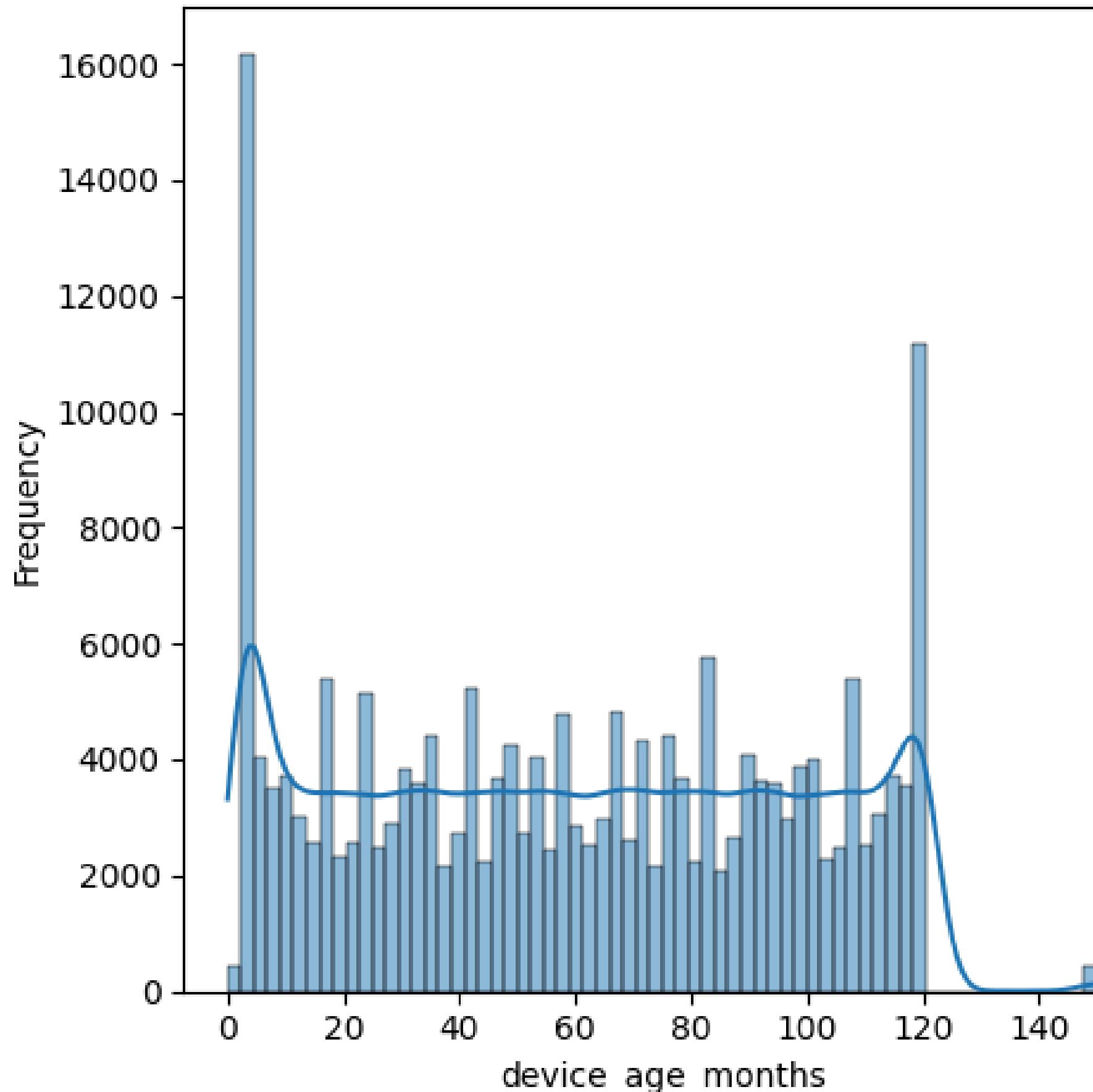
Distribution of income_usd_monthly



- Lệch phải mạnh
 - Đa số thu nhập < 5.000 USD
 - Ít người > 10.000 USD
- Chênh lệch thu nhập lớn

KIỂM TRA PHÂN PHỐI

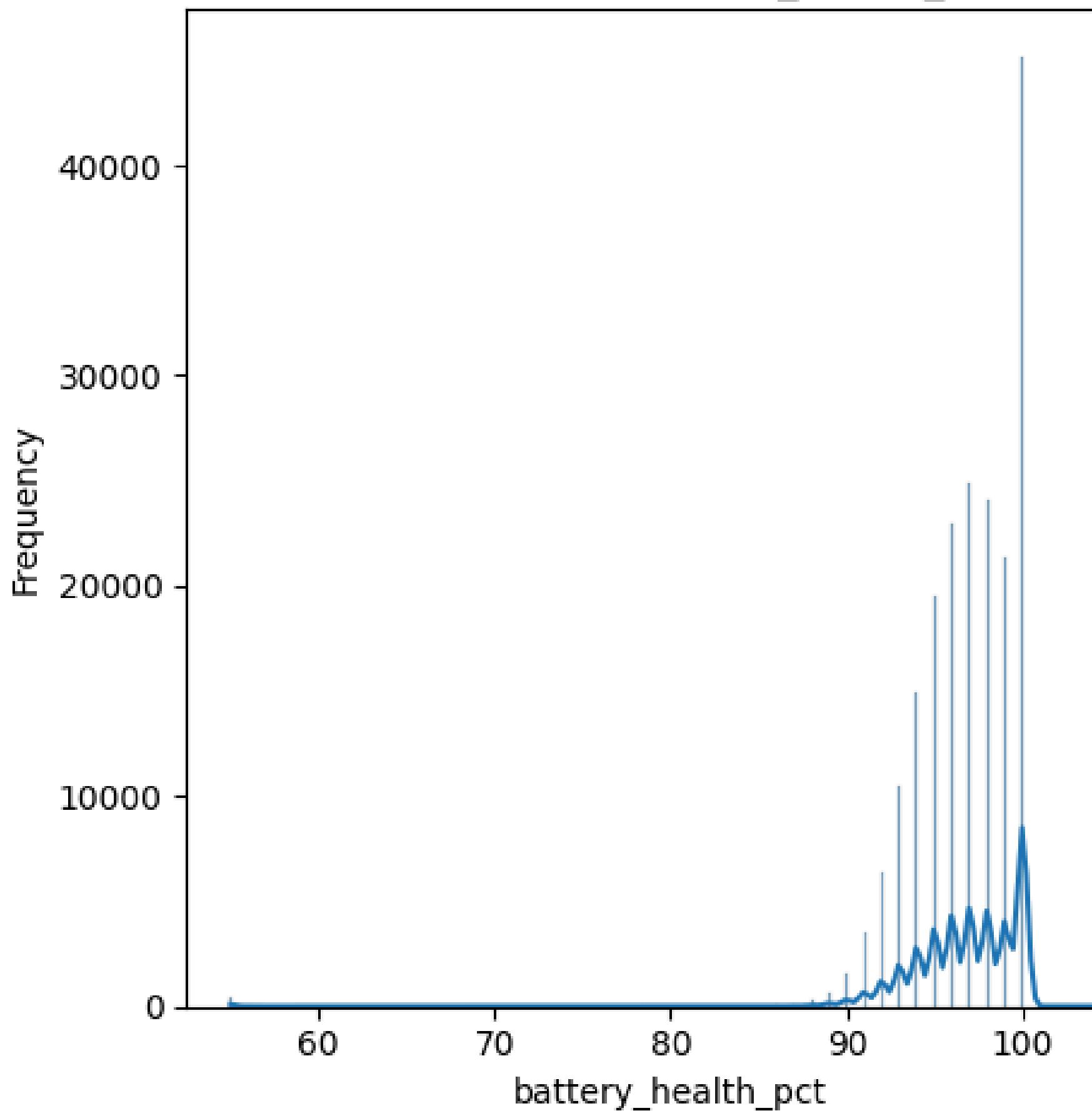
Distribution of device_age_months



- Phân bố rộng, nhiều mức từ vài tháng đến >10 năm
→ mức độ sử dụng thiết bị rất đa dạng

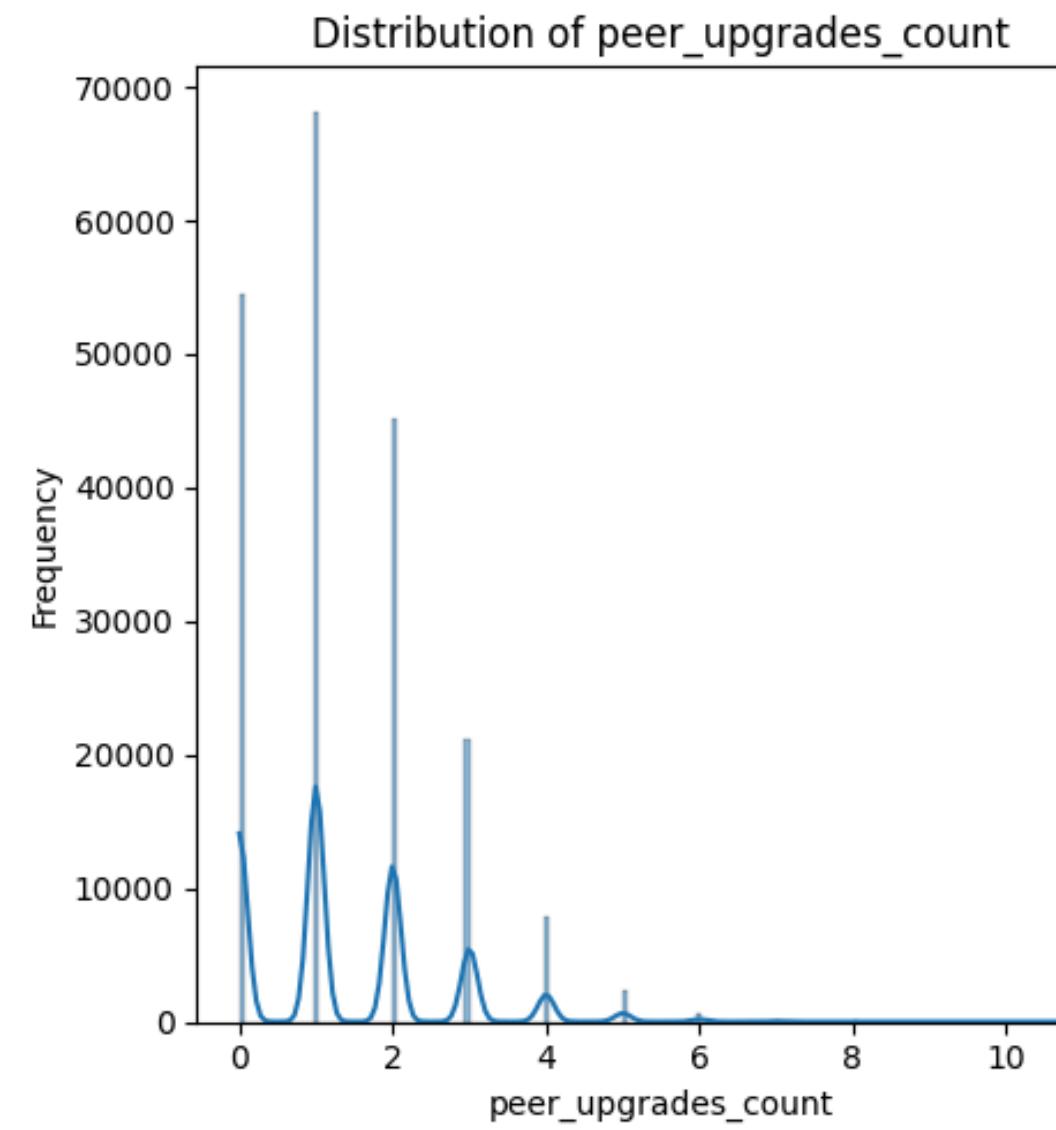
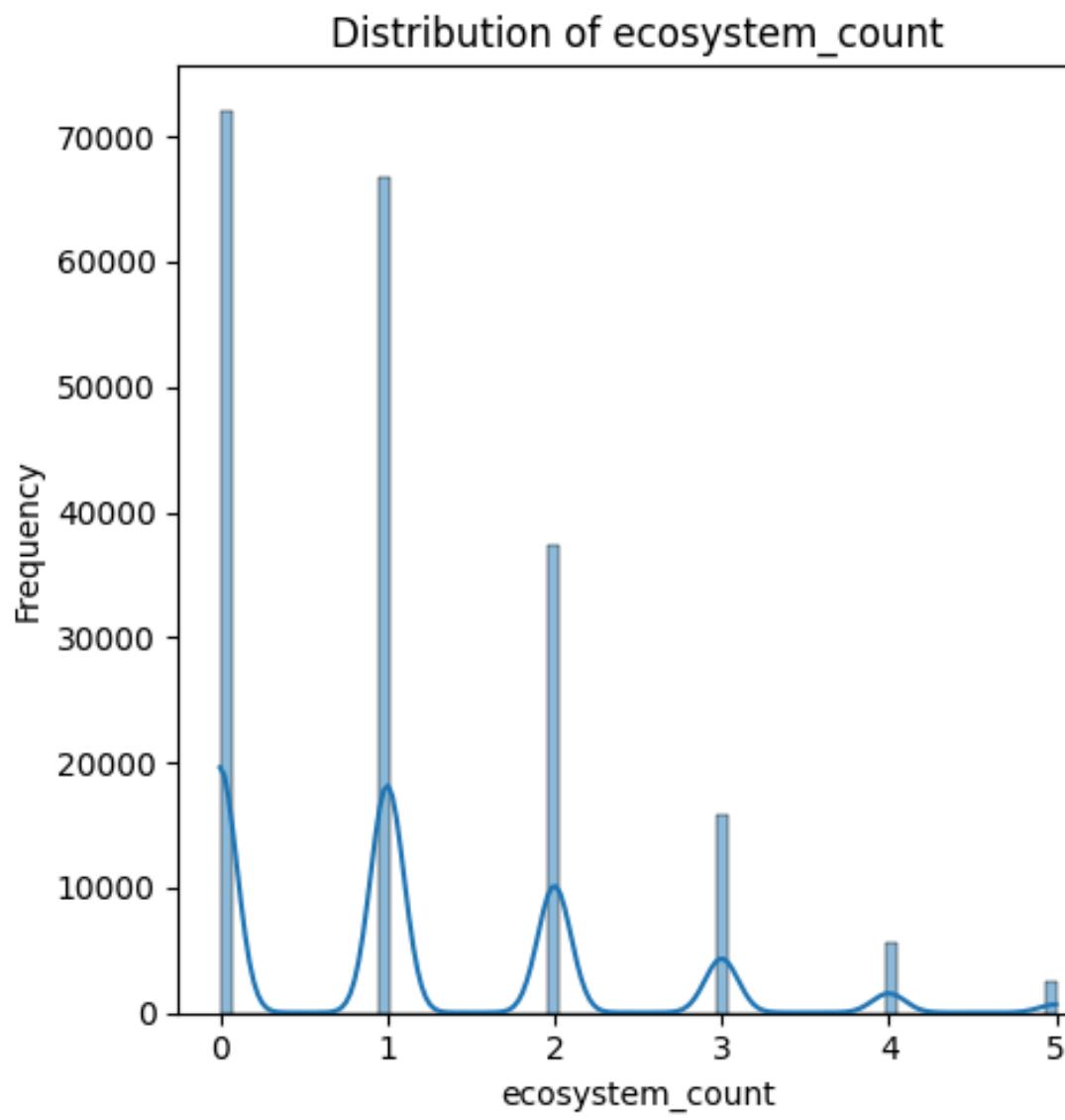
KIỂM TRA PHÂN PHỐI

Distribution of battery_health_pct



- Tập trung 95–100%
- Tuổi thiết bị phần lớn còn tốt

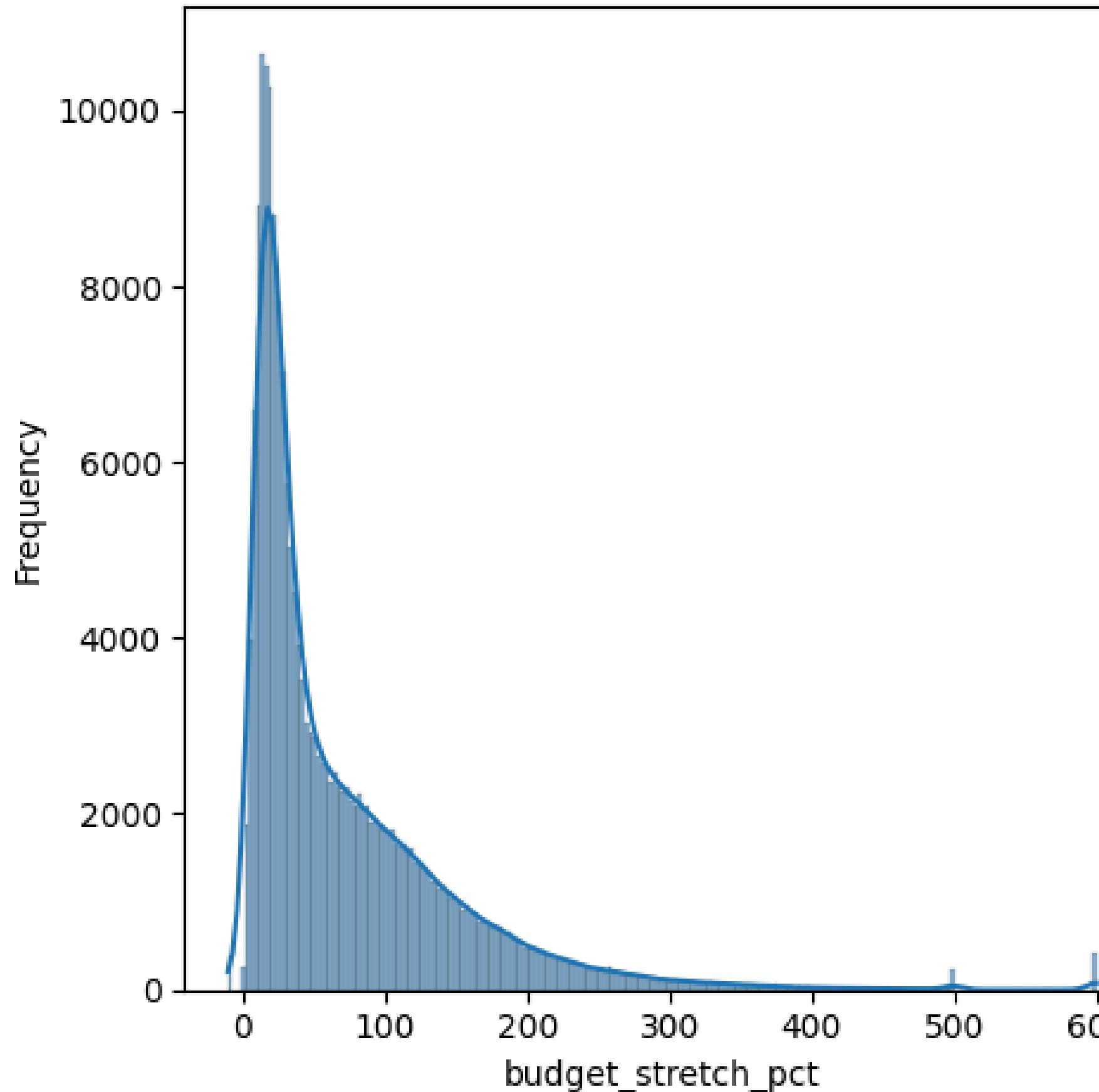
KIỂM TRA PHÂN PHỐI



- Lệch về 0
- Người dùng thường sở hữu 1-2 thiết bị
- Ít người quen nâng cấp

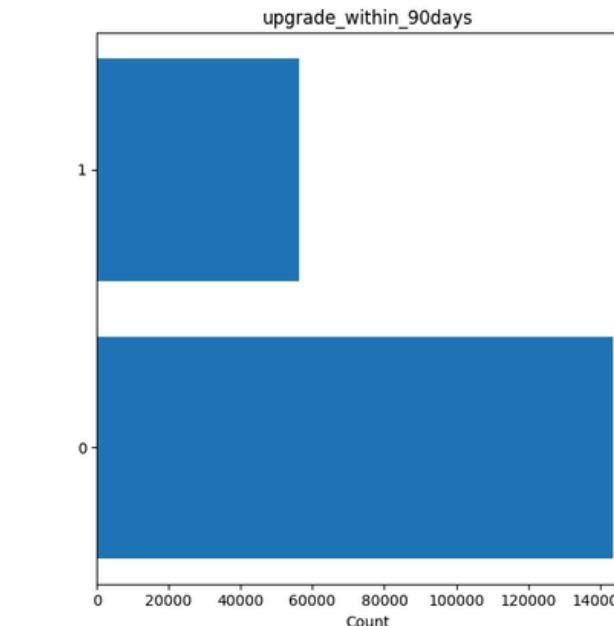
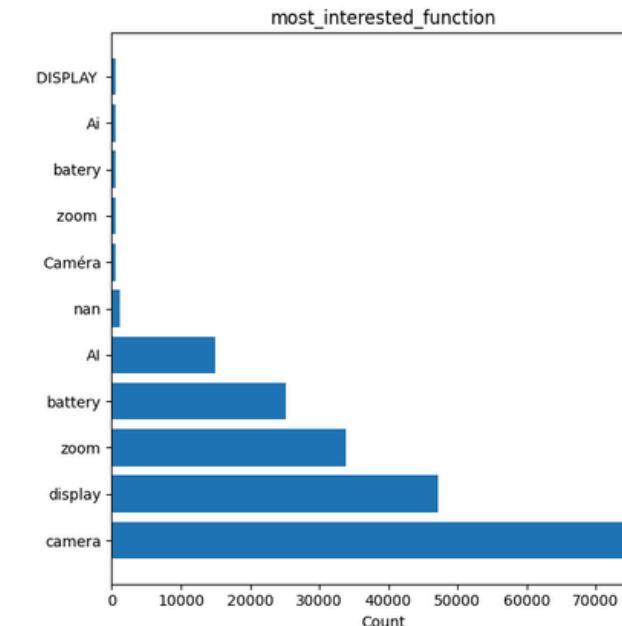
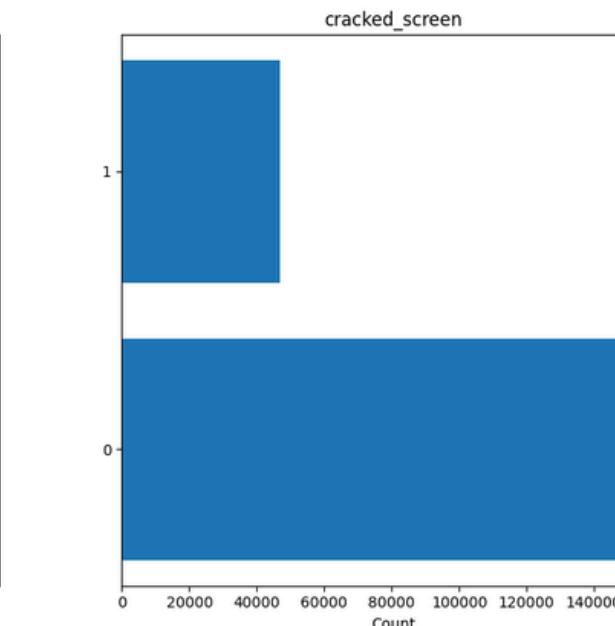
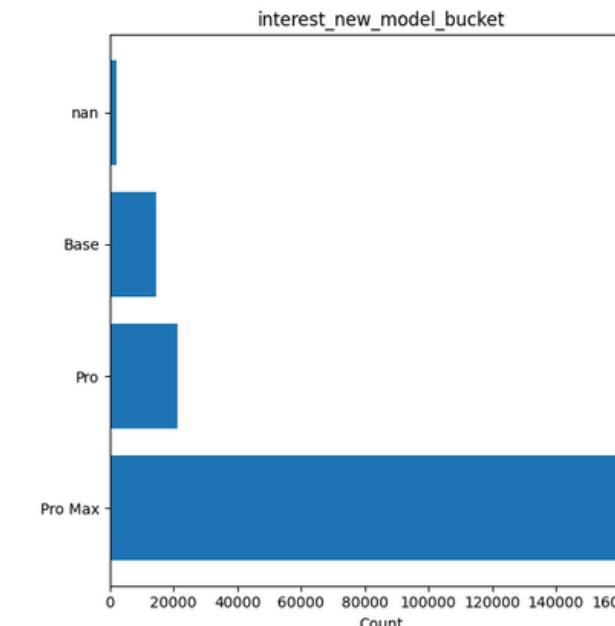
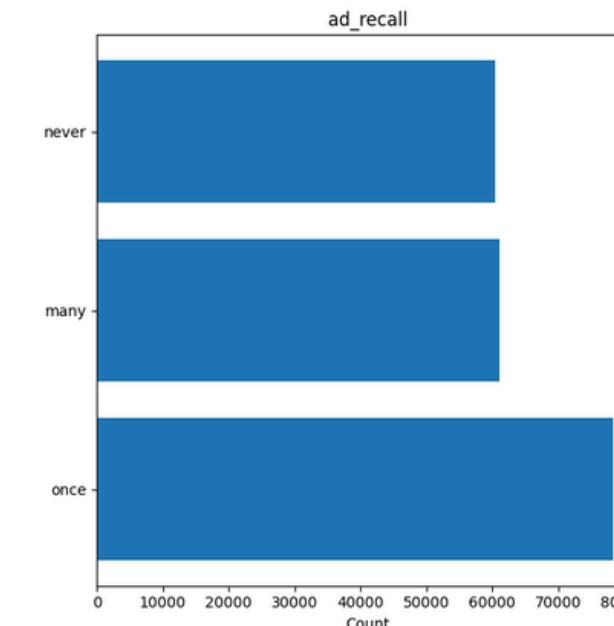
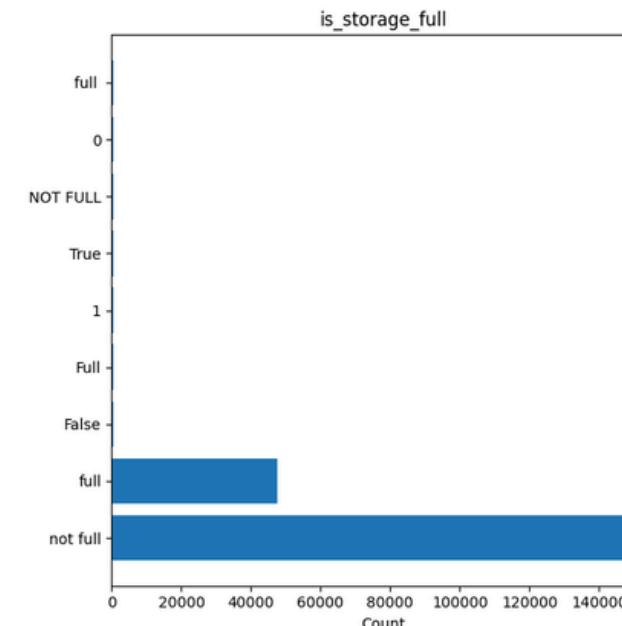
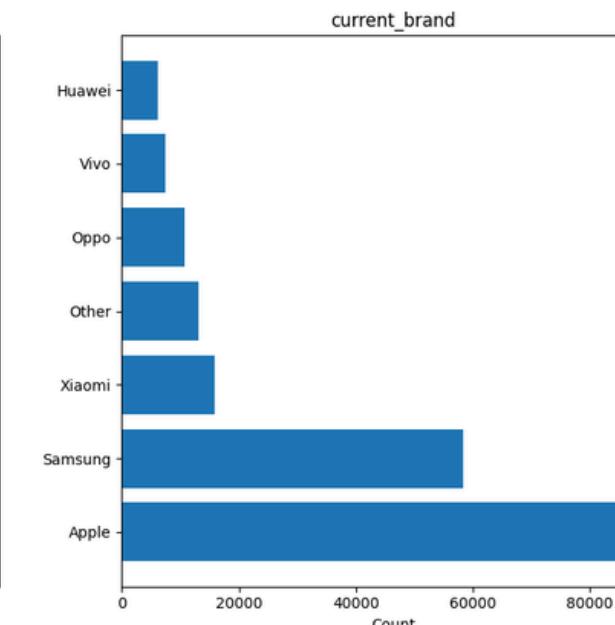
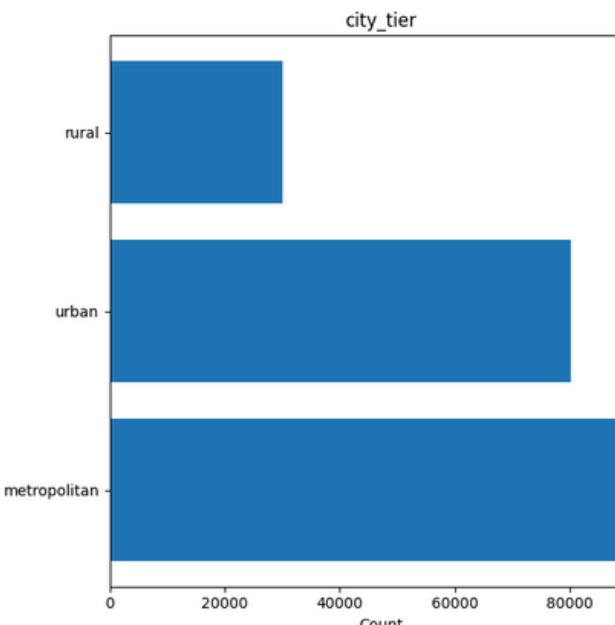
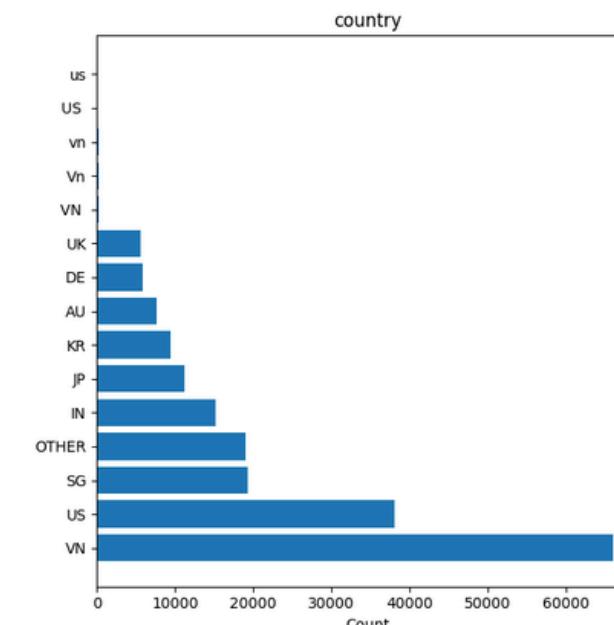
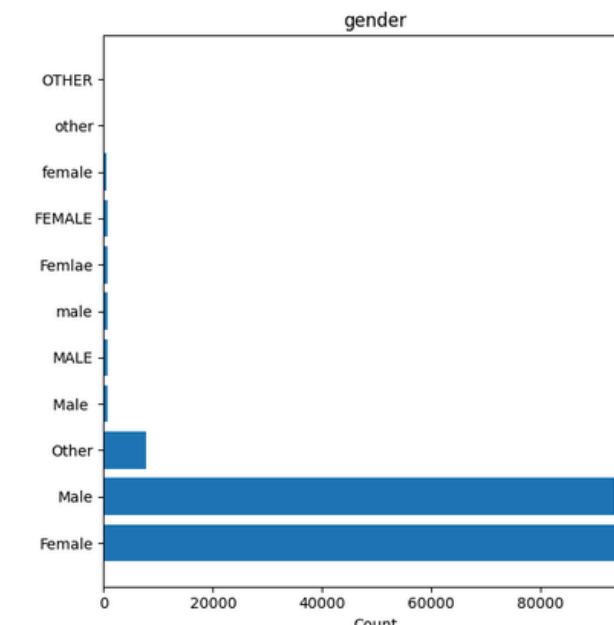
KIỂM TRA PHÂN PHỐI

Distribution of budget_stretch_pct

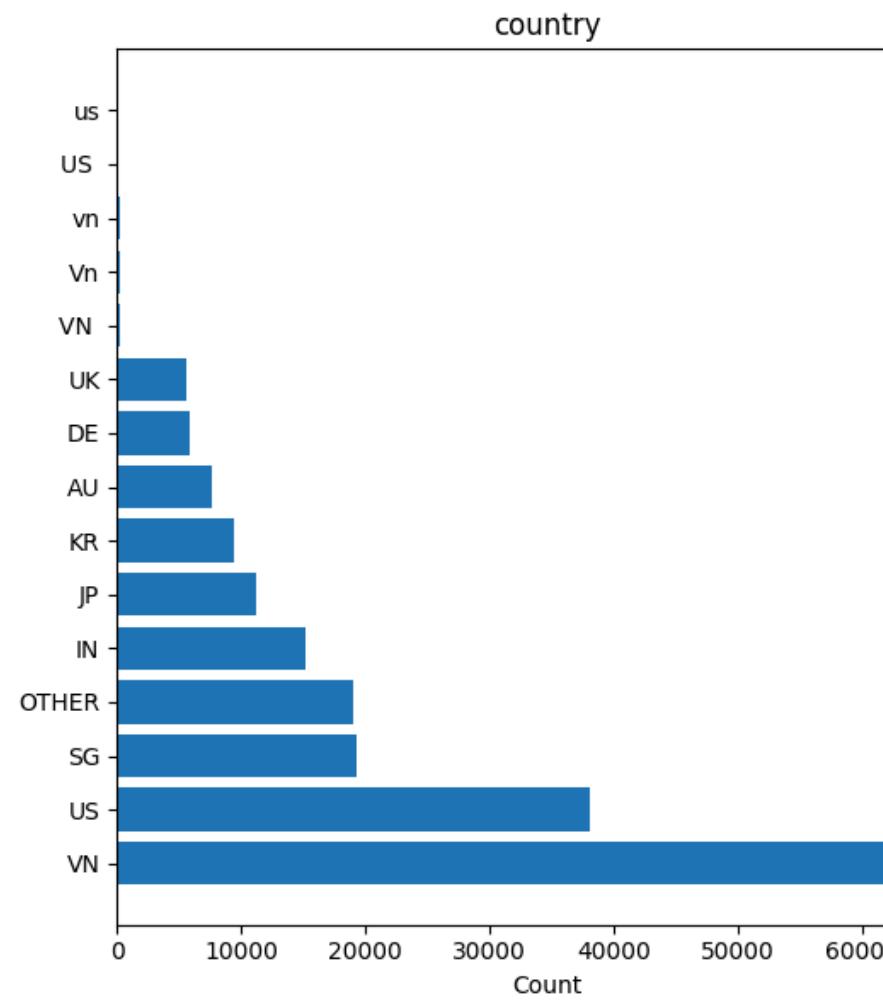
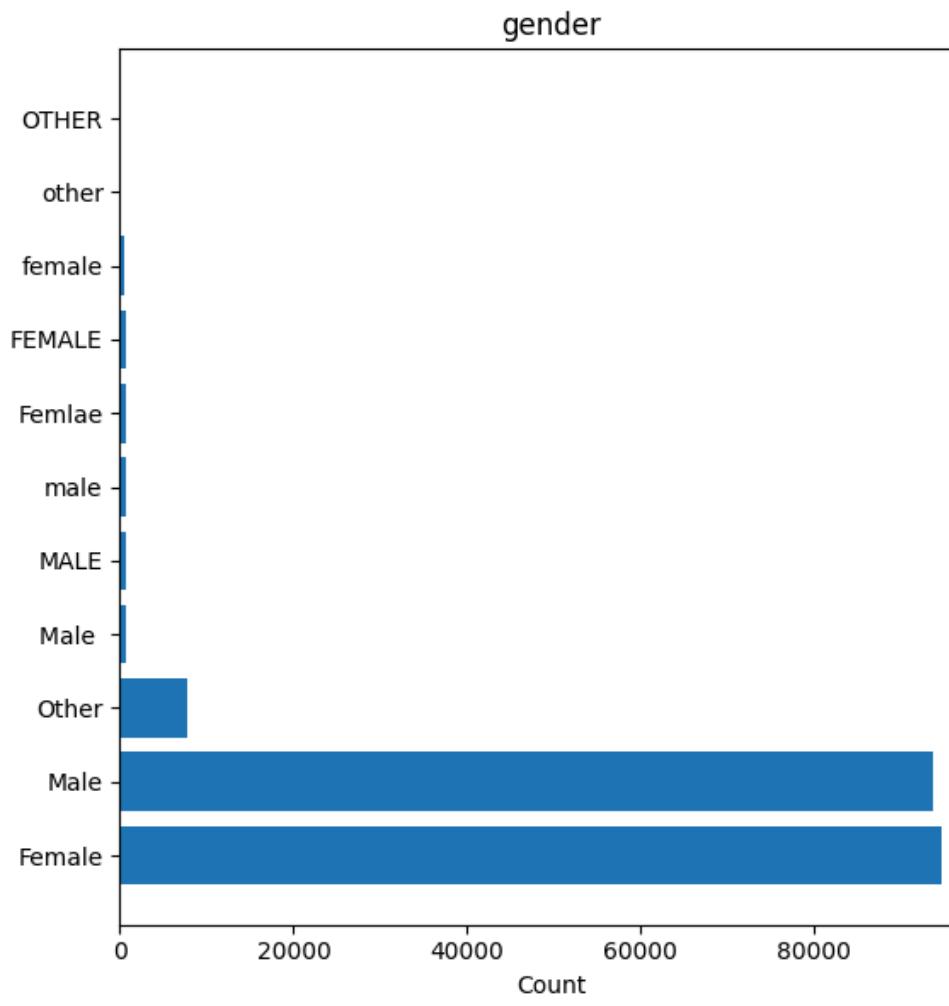


- Lệch phải, đa số < 100%
 - Có ngoại lai cao
- Tồn tại nhóm người săn sòng
chi vượt trội

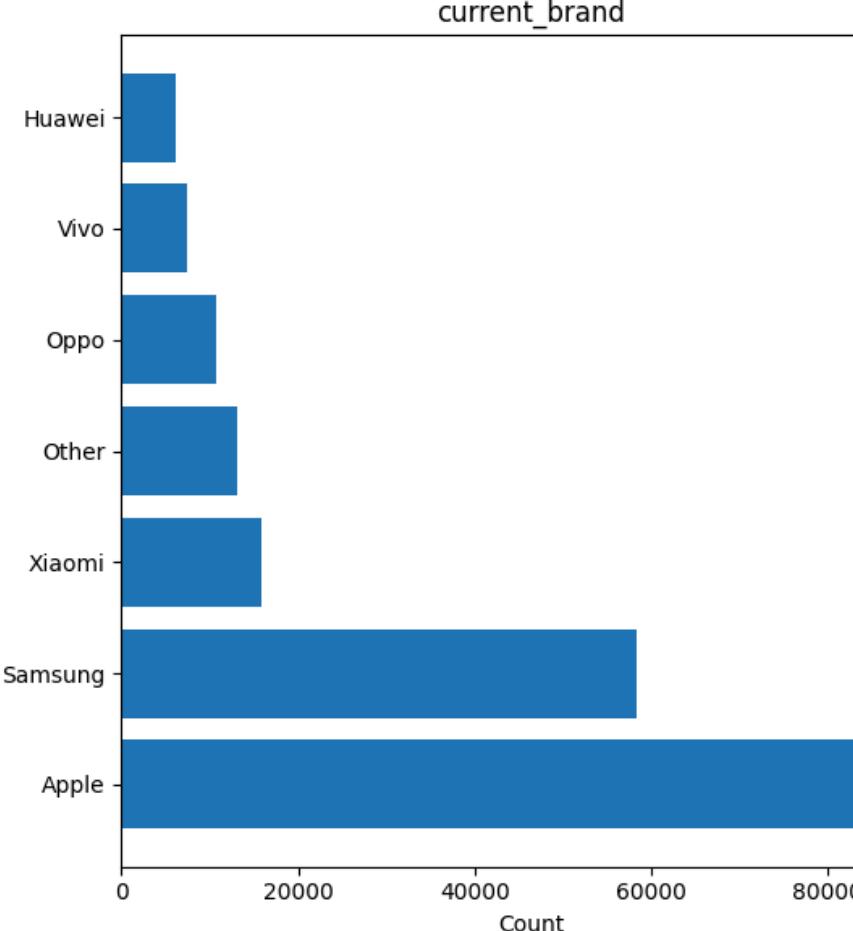
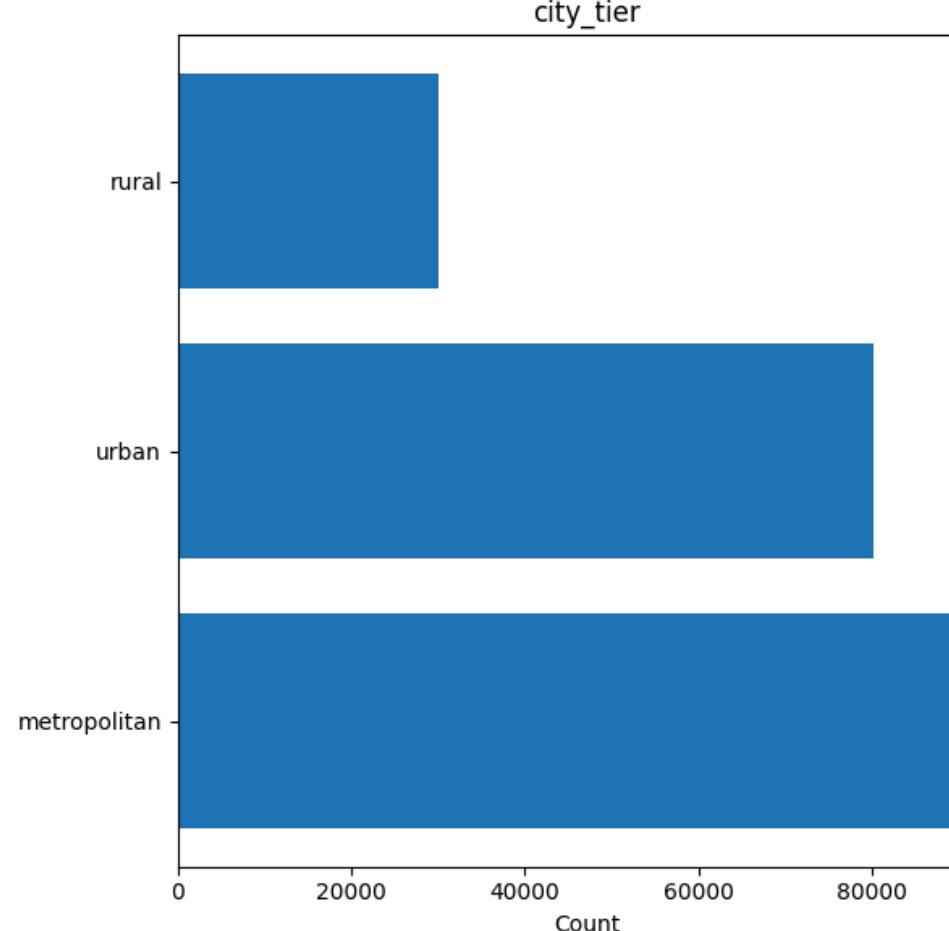
KIỂM TRA PHÂN PHỐI



KIỂM TRA PHÂN PHỐI

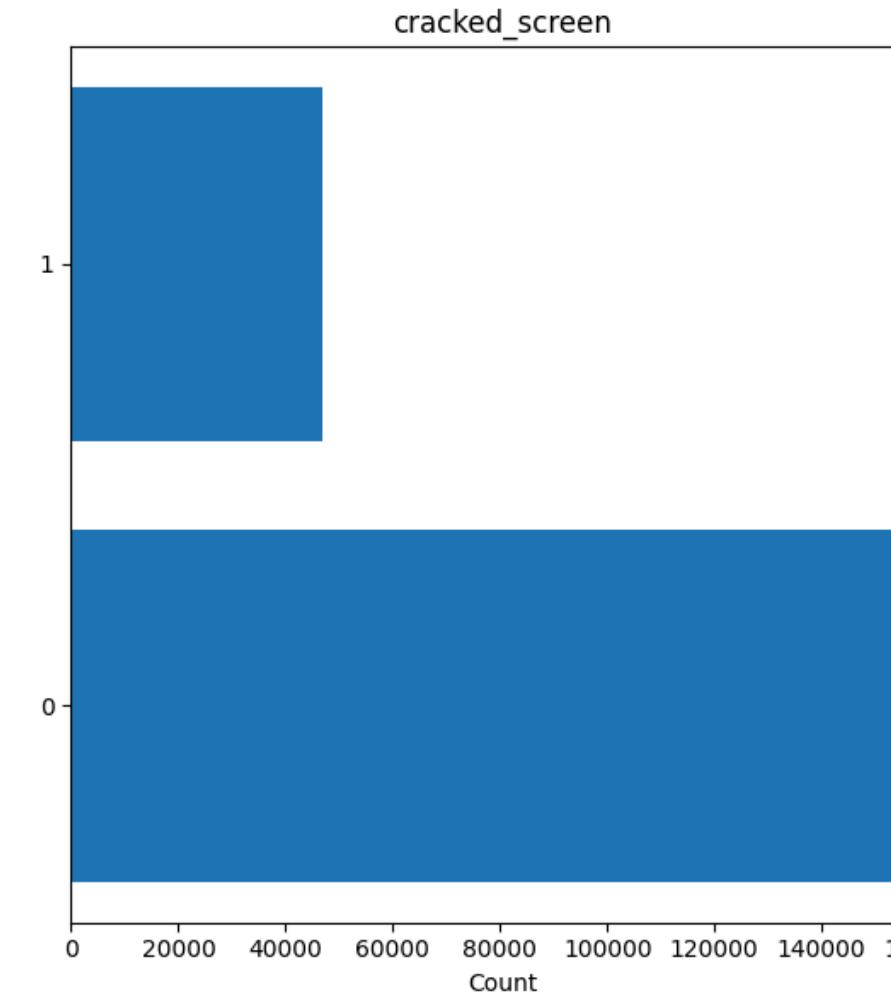
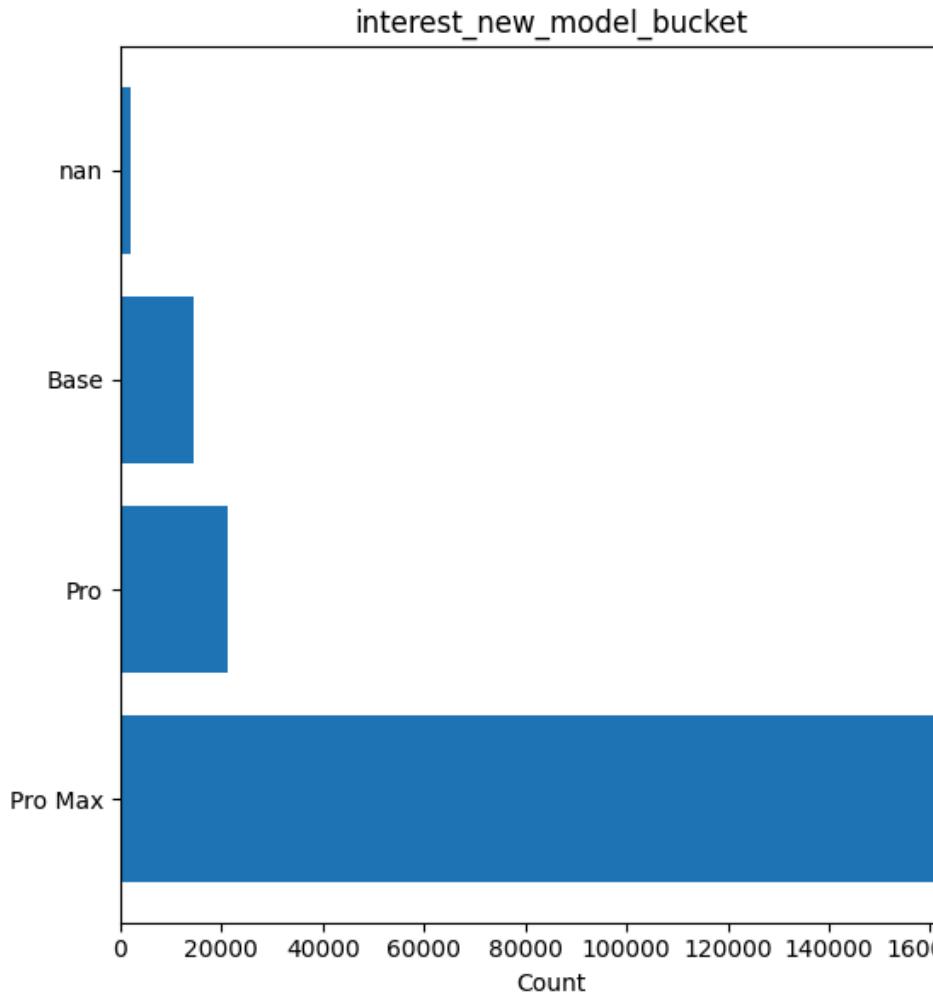


- **Gender:** Nghiêng về Female; nhẫn không thống nhất (Female/female/...)
- **Country:** Thực tế chỉ khoảng 10 quốc gia nhẫn không thống nhất (VN, Vn, vn,...)

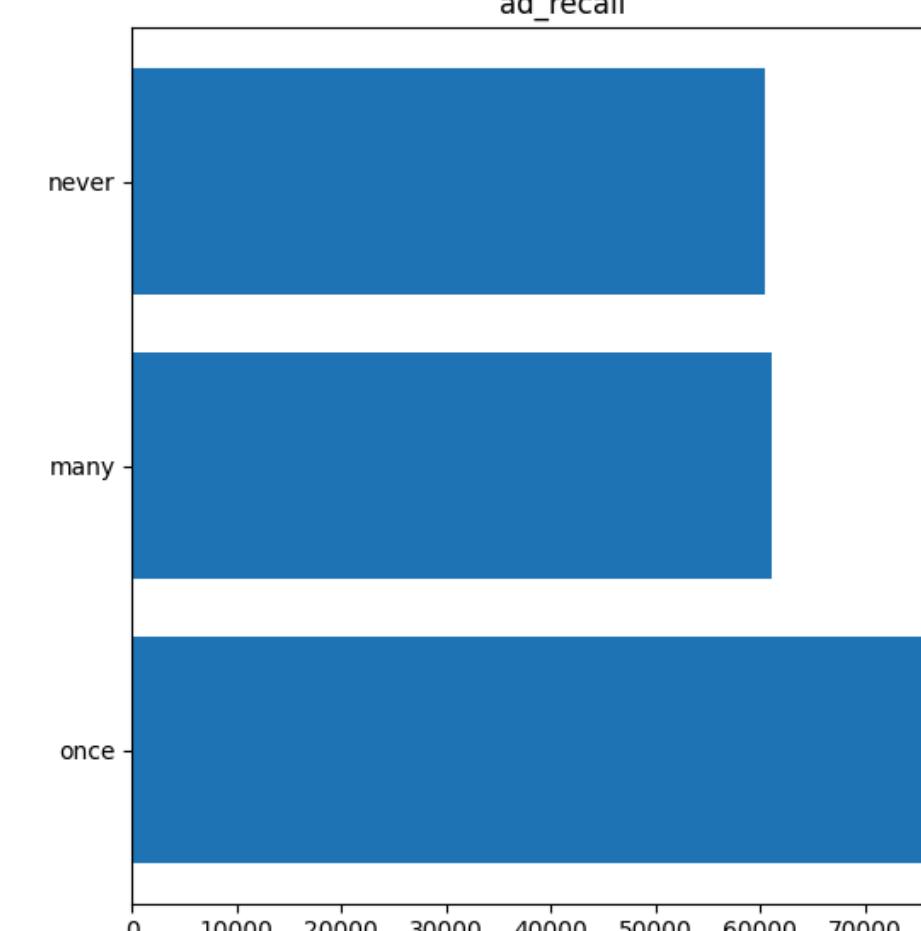
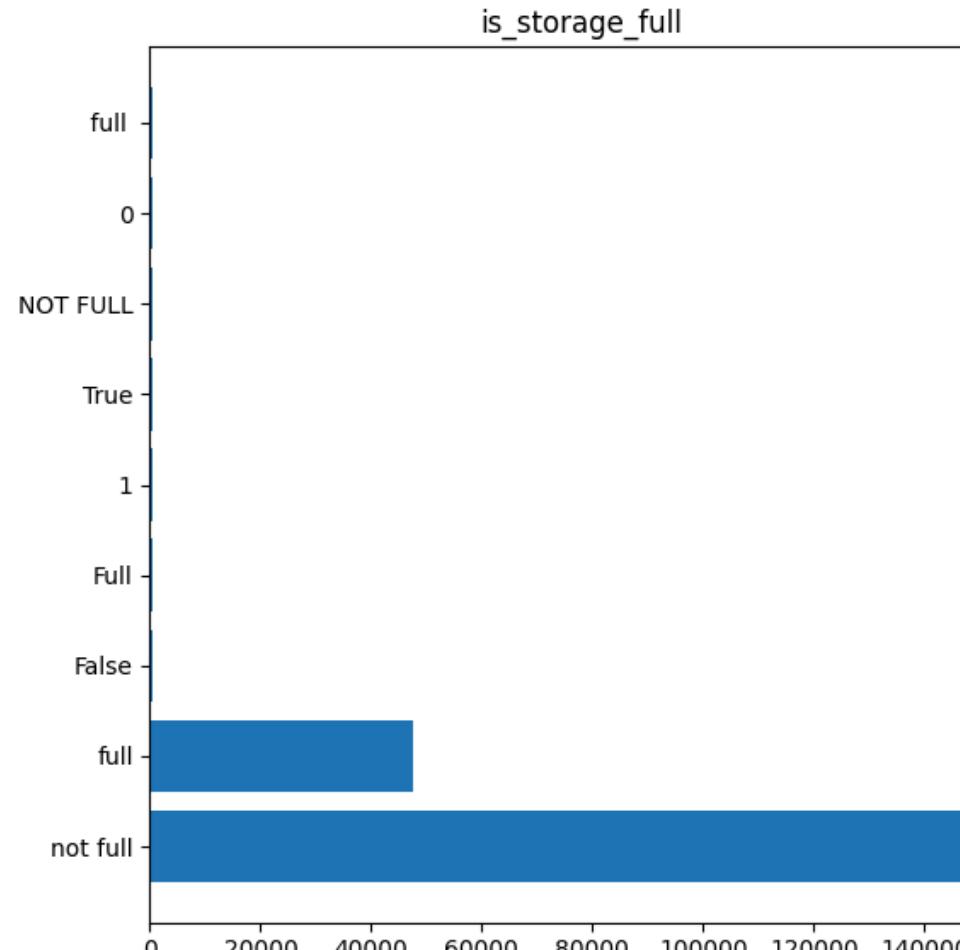


- **City tier:** Tập trung metropolitan (45%), người dùng đô thị, tiếp cận công nghệ cao
- **Brand hiện tại:** Apple dẫn đầu (44%), sau đó Samsung, Xiaomi, Oppo

KIỂM TRA PHÂN PHỐI

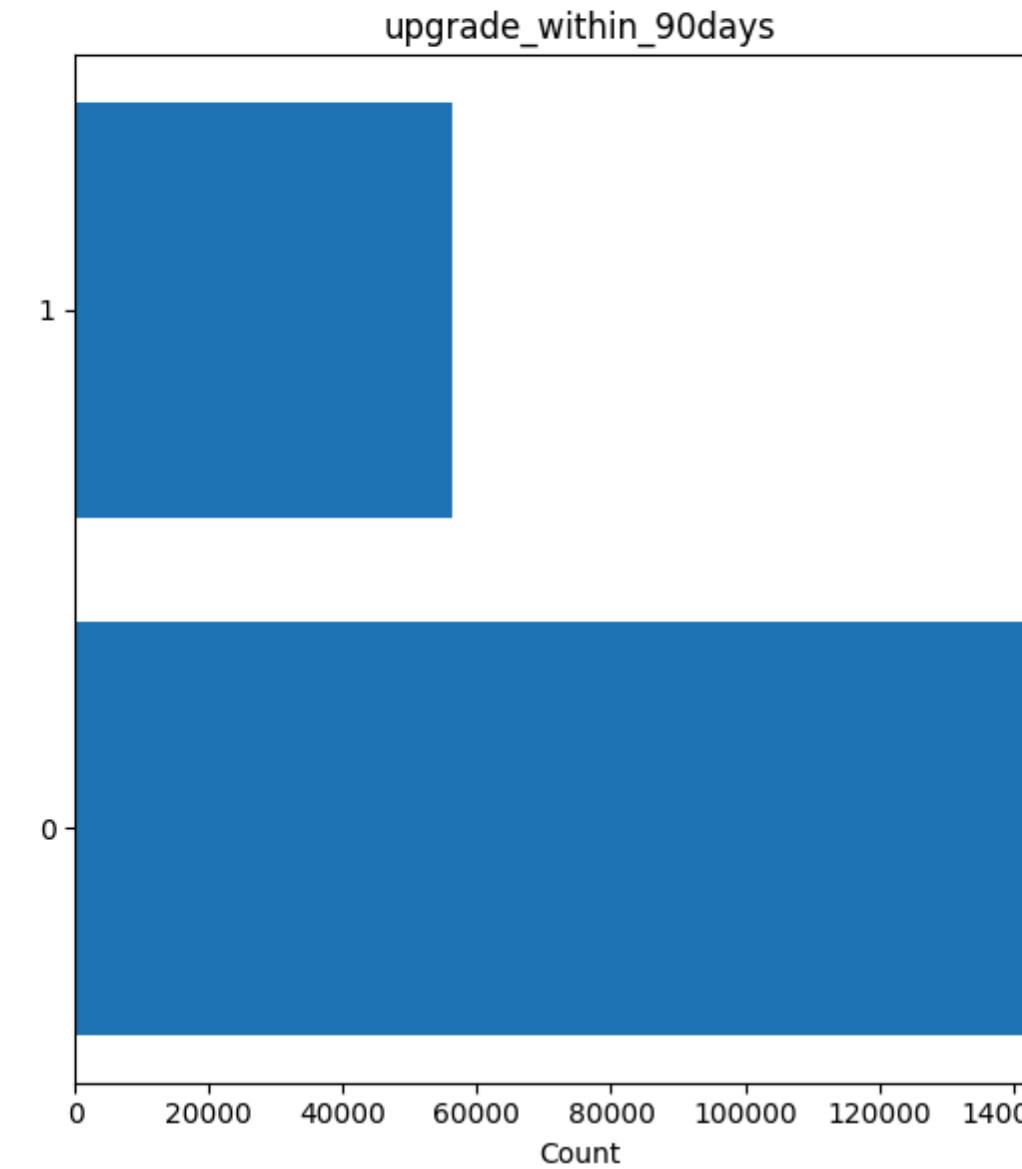
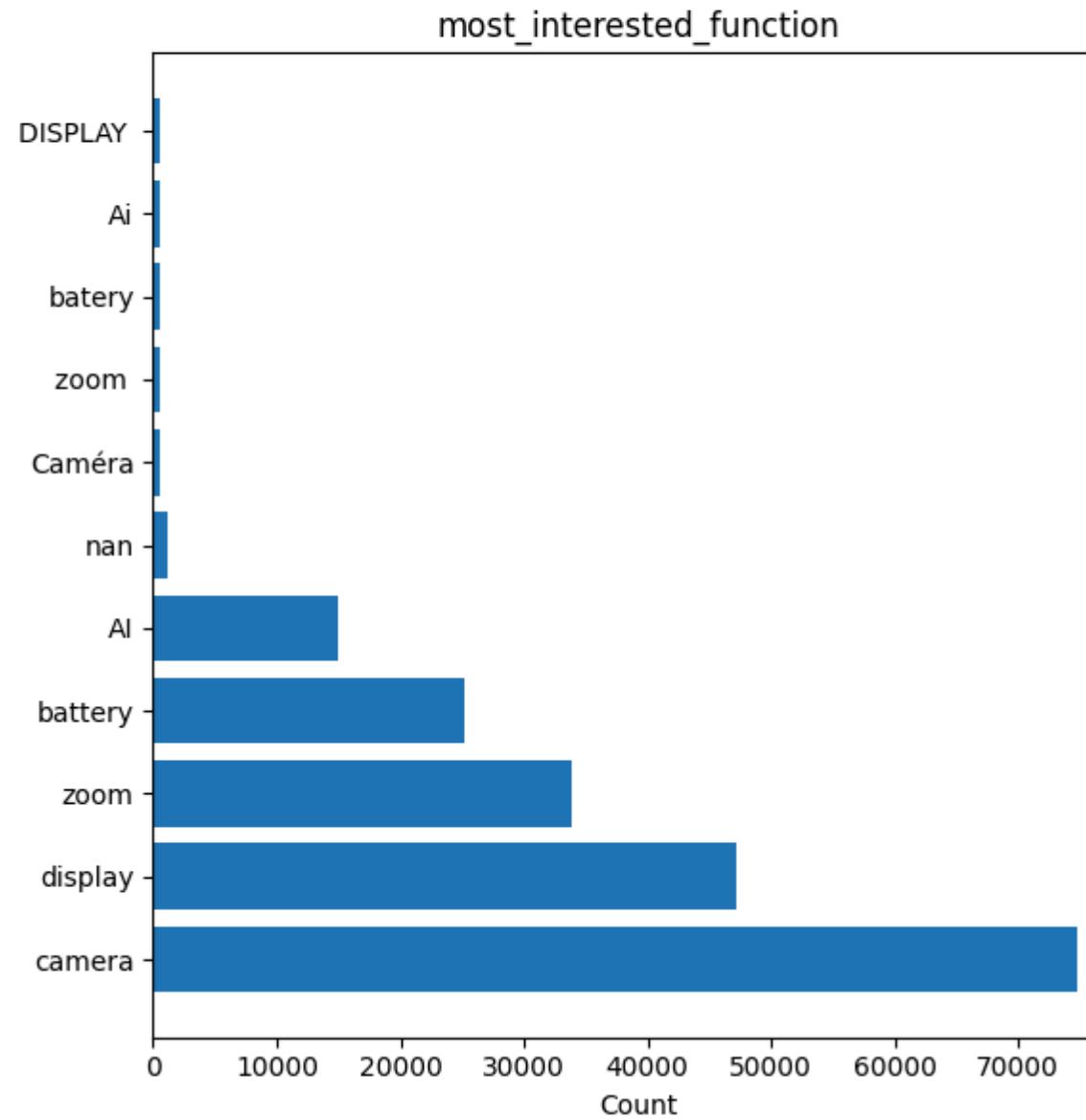


- Quan tâm mẫu mới: Pro Max >80%
- Màn hình nứt (cracked_screen): 23%



- Bộ nhớ (is_storage_full): khoảng 74% chưa đầy
- Ghi nhớ quảng cáo (ad_recall): Chủ yếu once (39%)

KIỂM TRA PHÂN PHỐI



- Tính năng ưu tiên: Camera (37%) đứng đầu, sau đó màn hình & pin
- Quyết định nâng cấp trong vòng 90 ngày: ~28%

TRỰC QUAN HÓA DỮ LIỆU

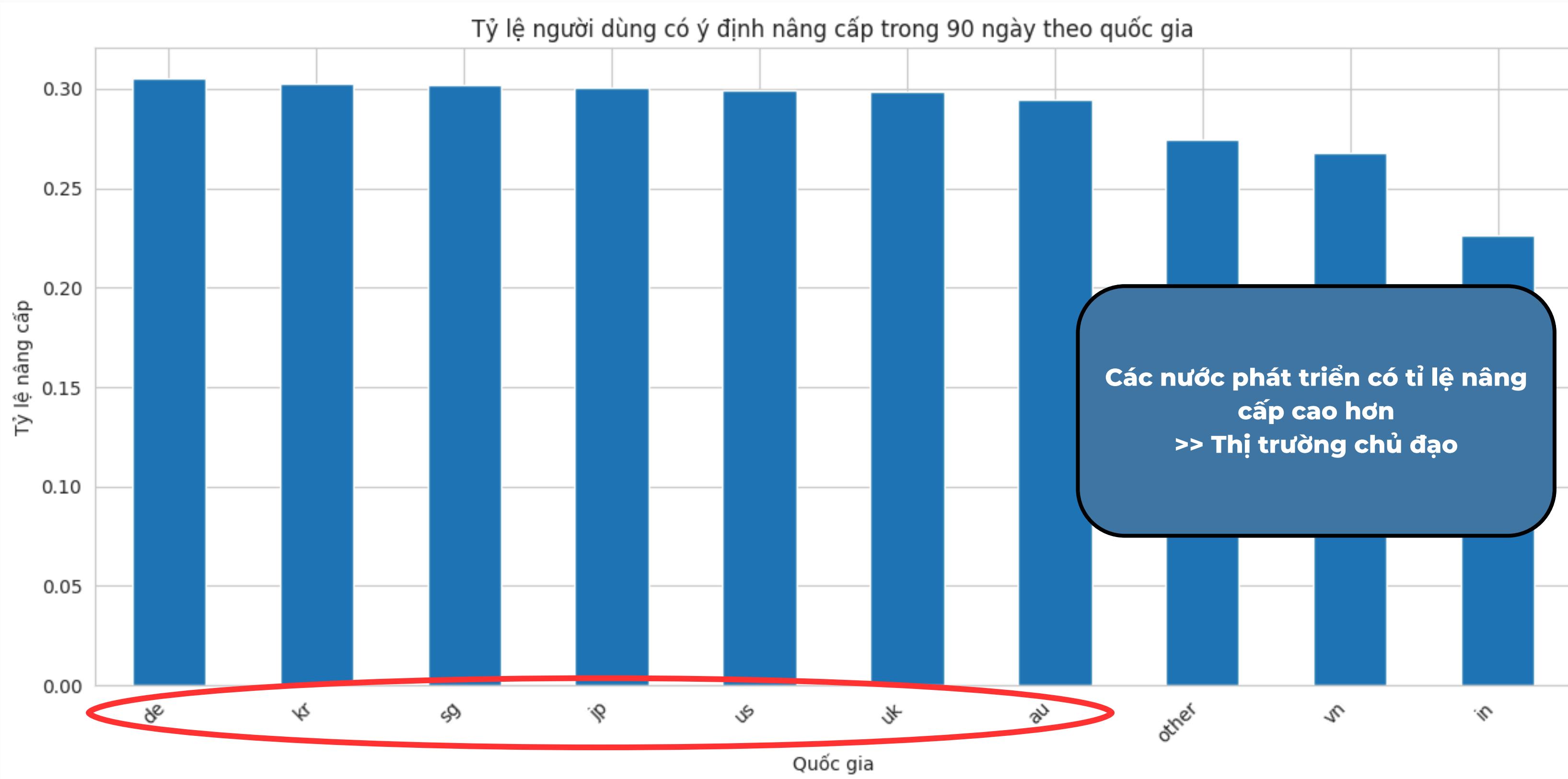
1. THEO QUỐC GIA

2. THEO THƯƠNG HIỆU

3. THEO NHÂN KHẨU HỌC

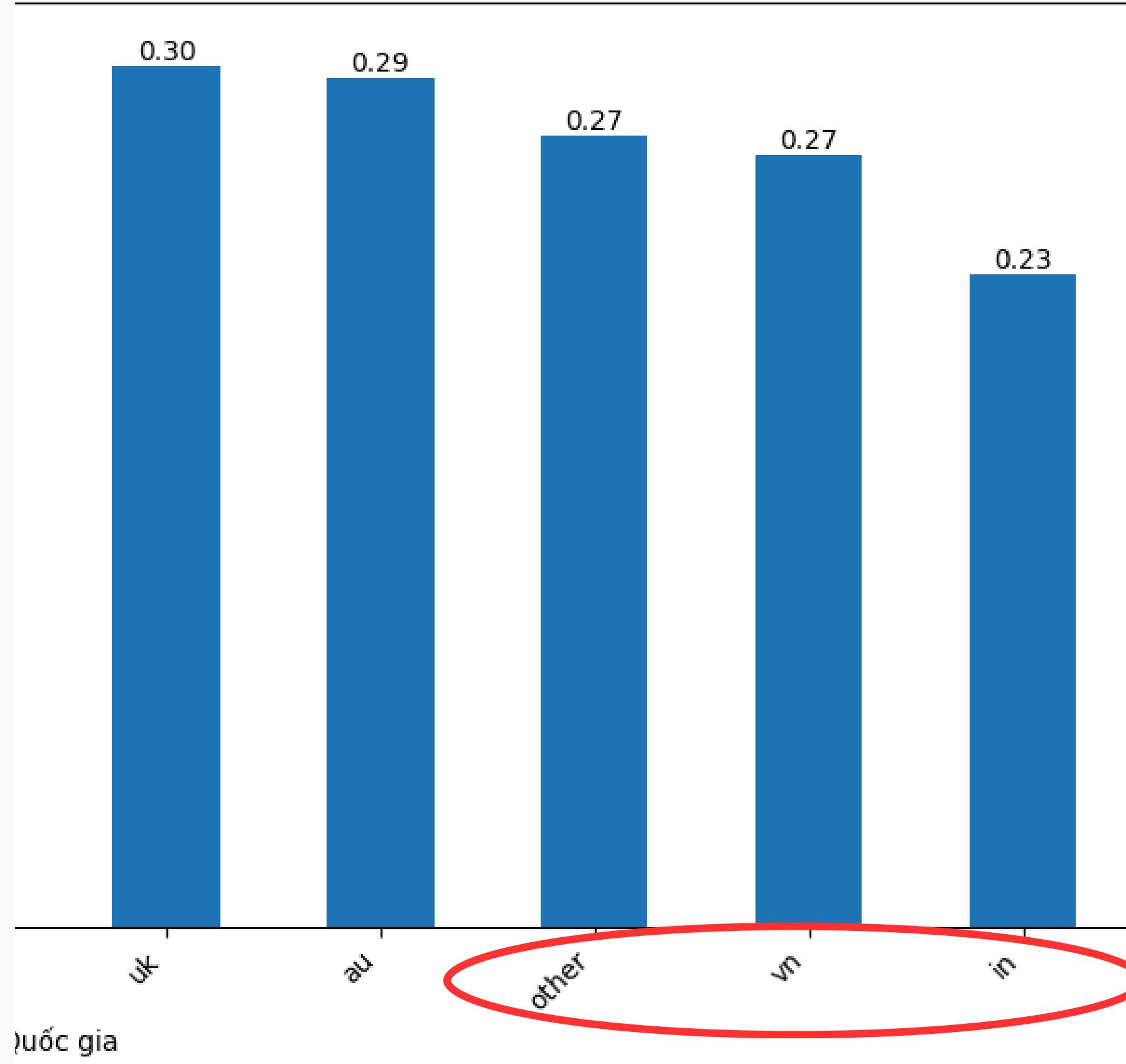
**4. THEO HÀNH VI, TÂM LÝ
NGƯỜI TIÊU DÙNG**

THEO QUỐC GIA



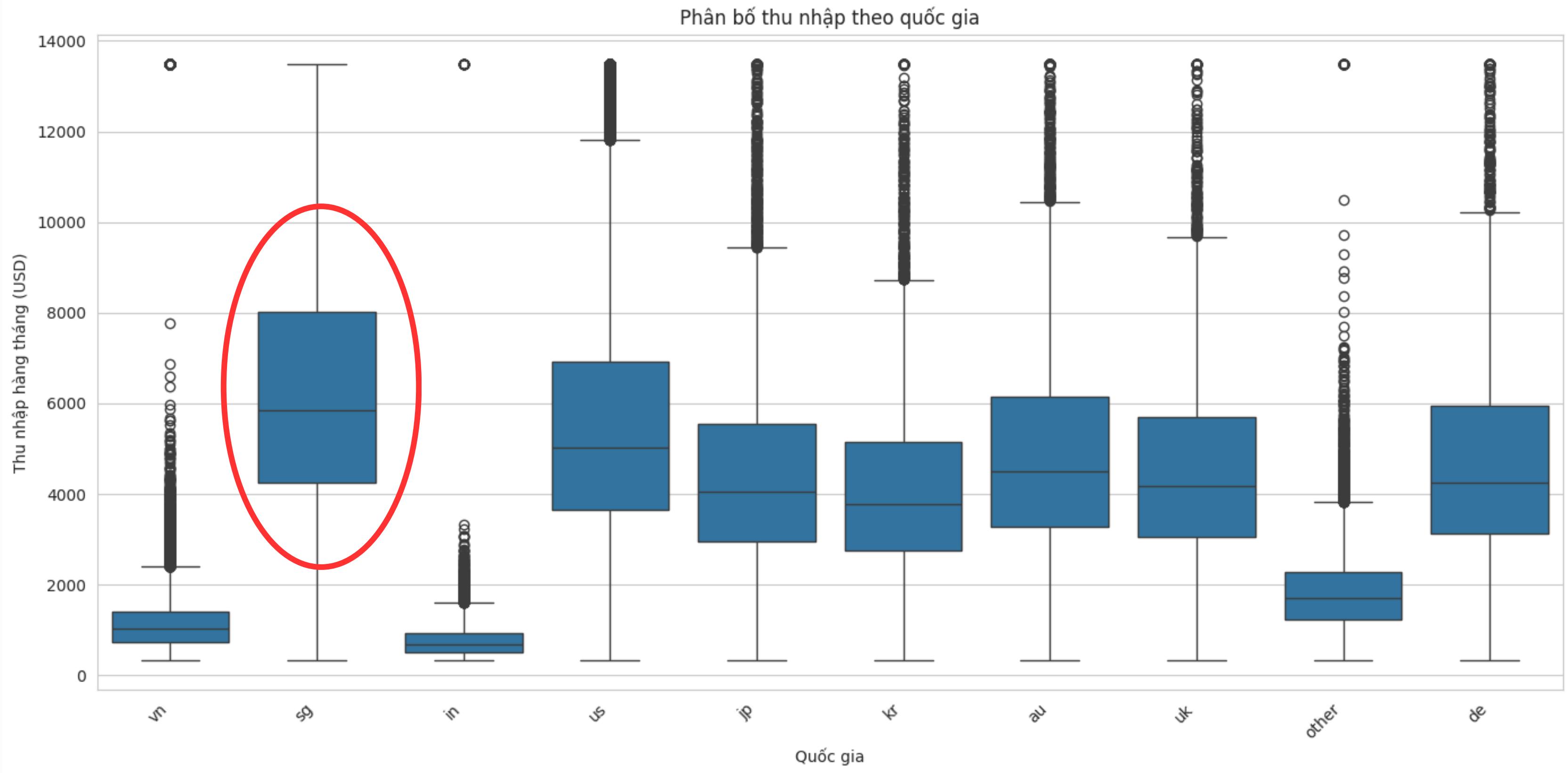
THEO QUỐC GIA

tỷ lệ nâng cấp trong 90 ngày theo quốc gia

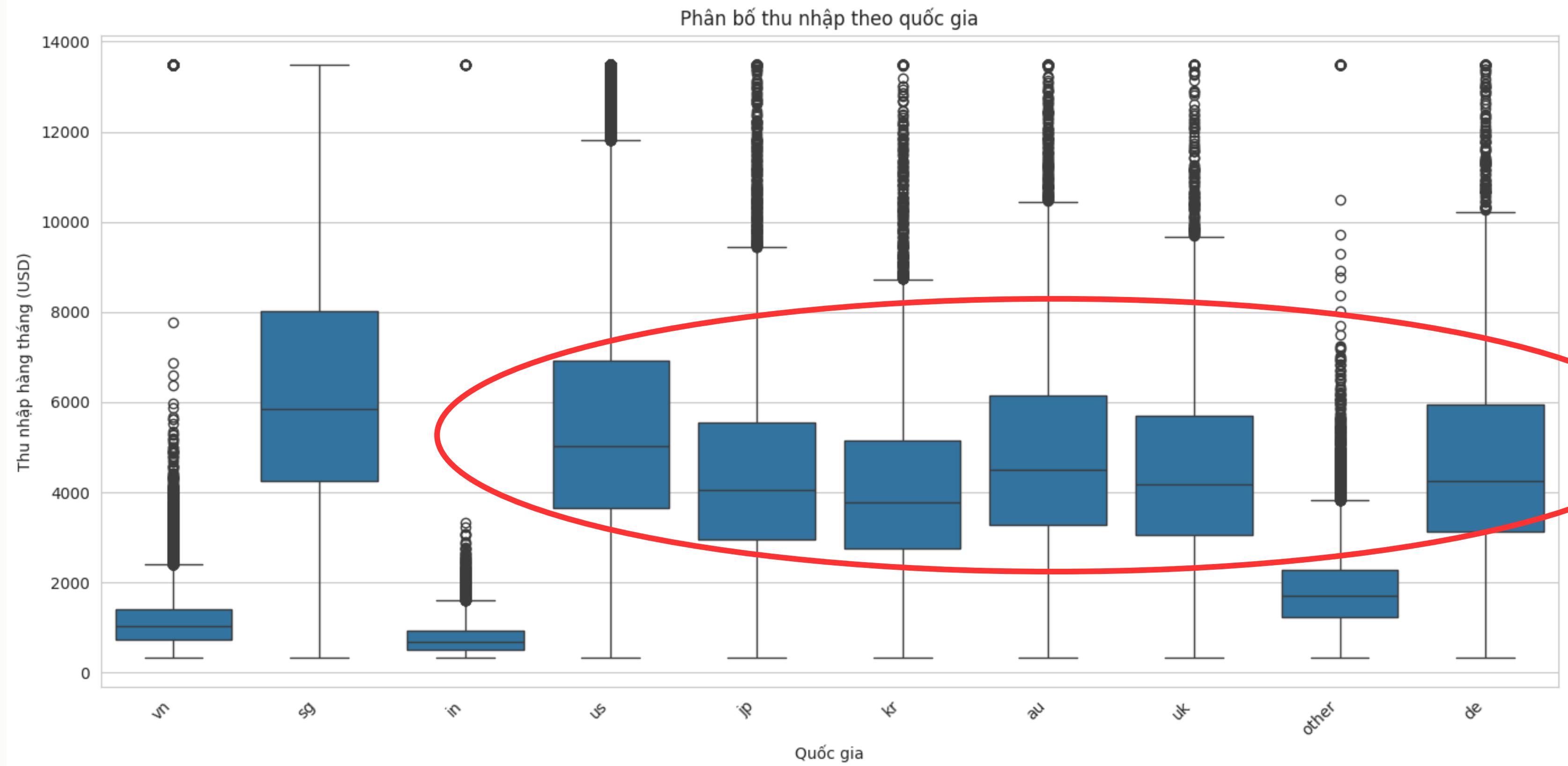


**Việt Nam và Ấn Độ nằm trong
nhóm có tỉ lệ nâng cấp thấp nhất
>> Có sự nhạy cảm về giá**

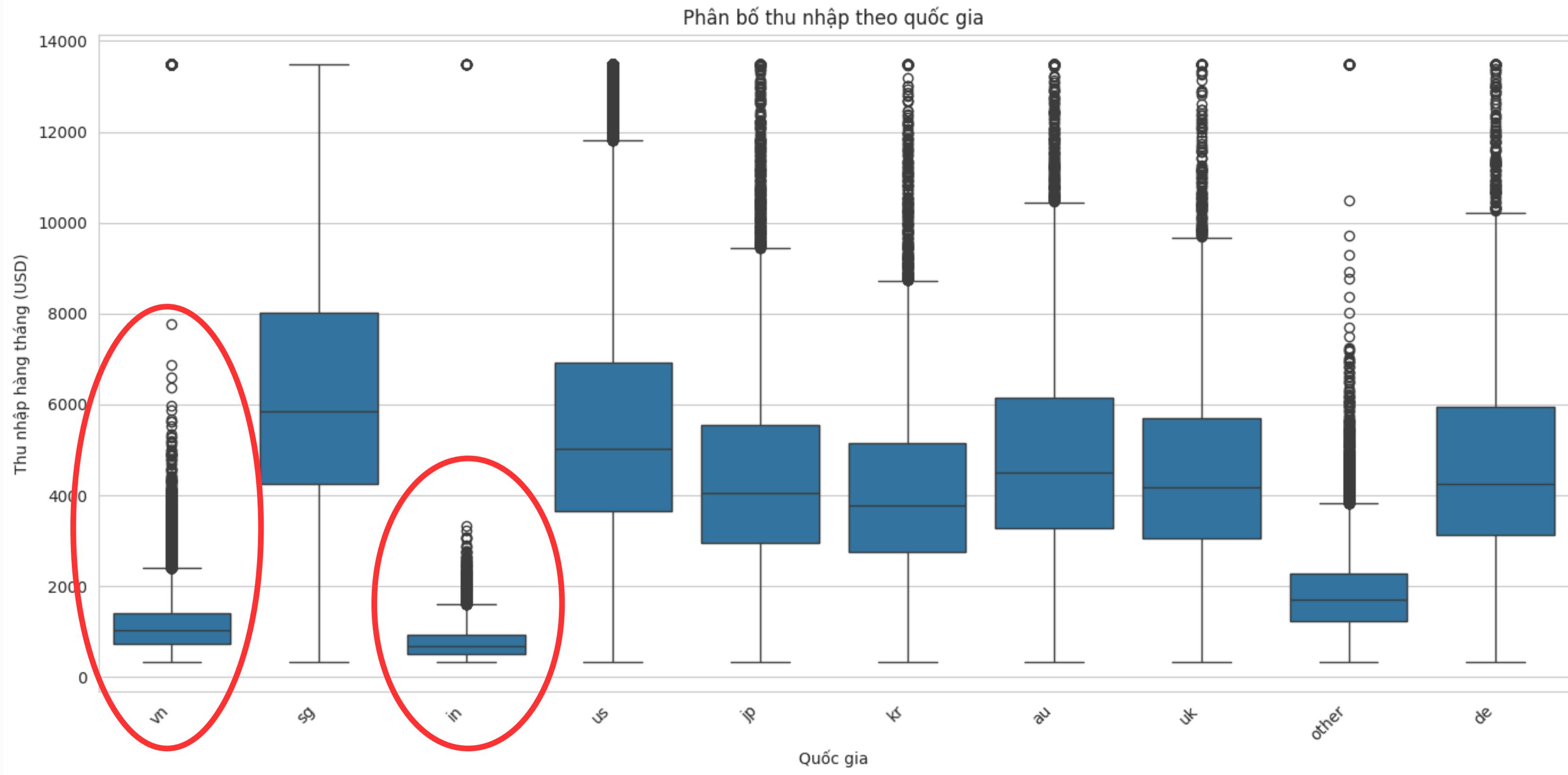
THEO QUỐC GIA



THEO QUỐC GIA

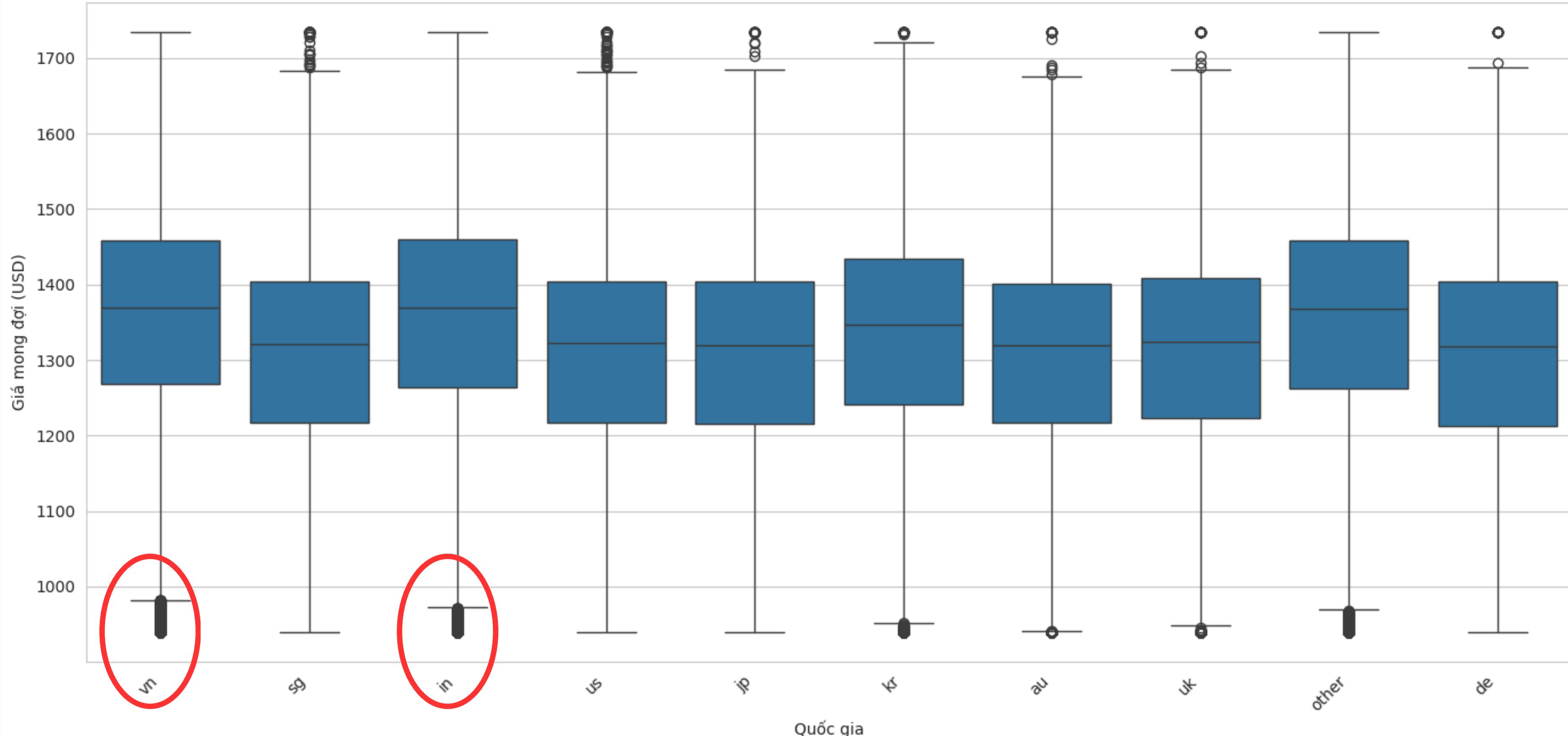


THEO QUỐC GIA



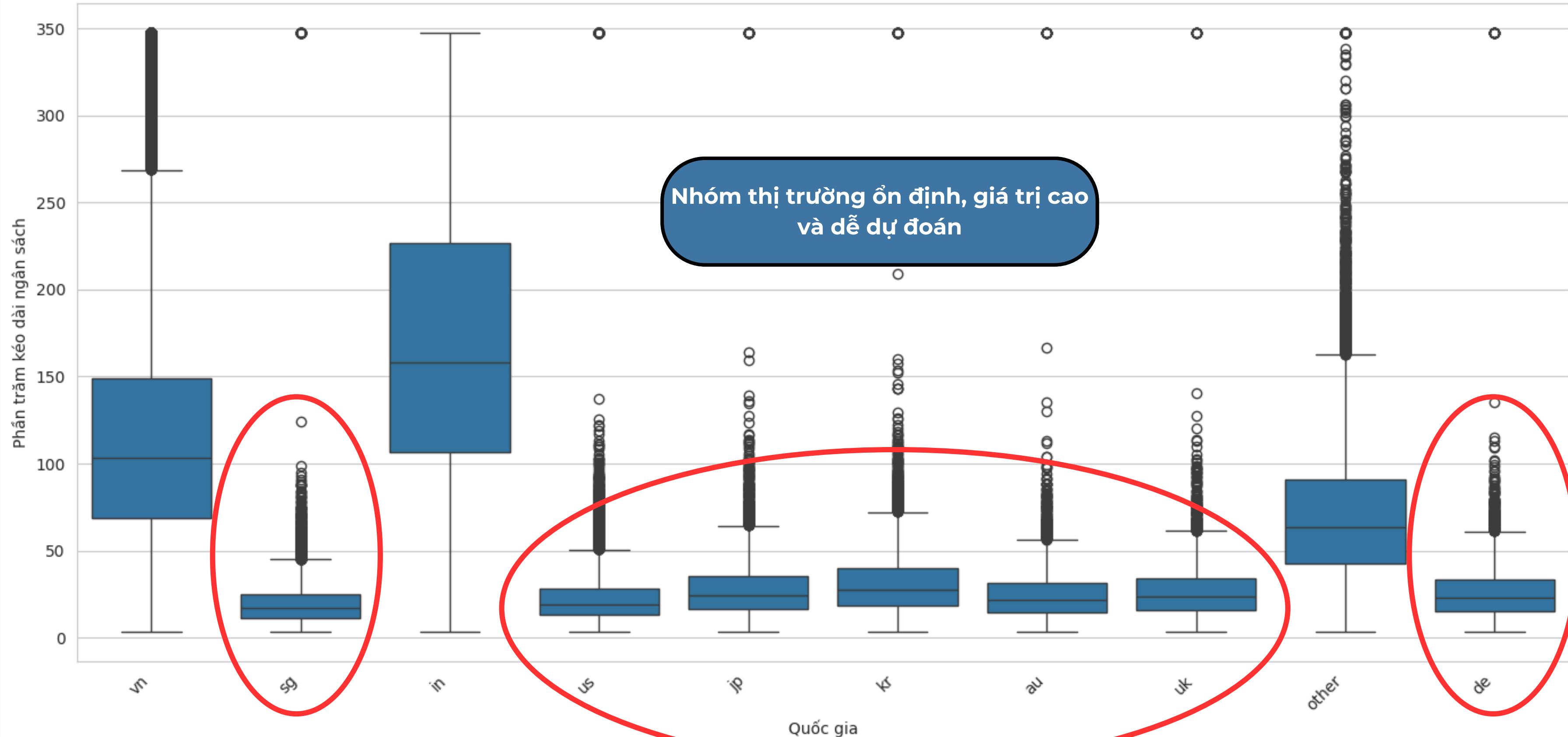
THEO QUỐC GIA

Phân bố giá mong đợi theo quốc gia



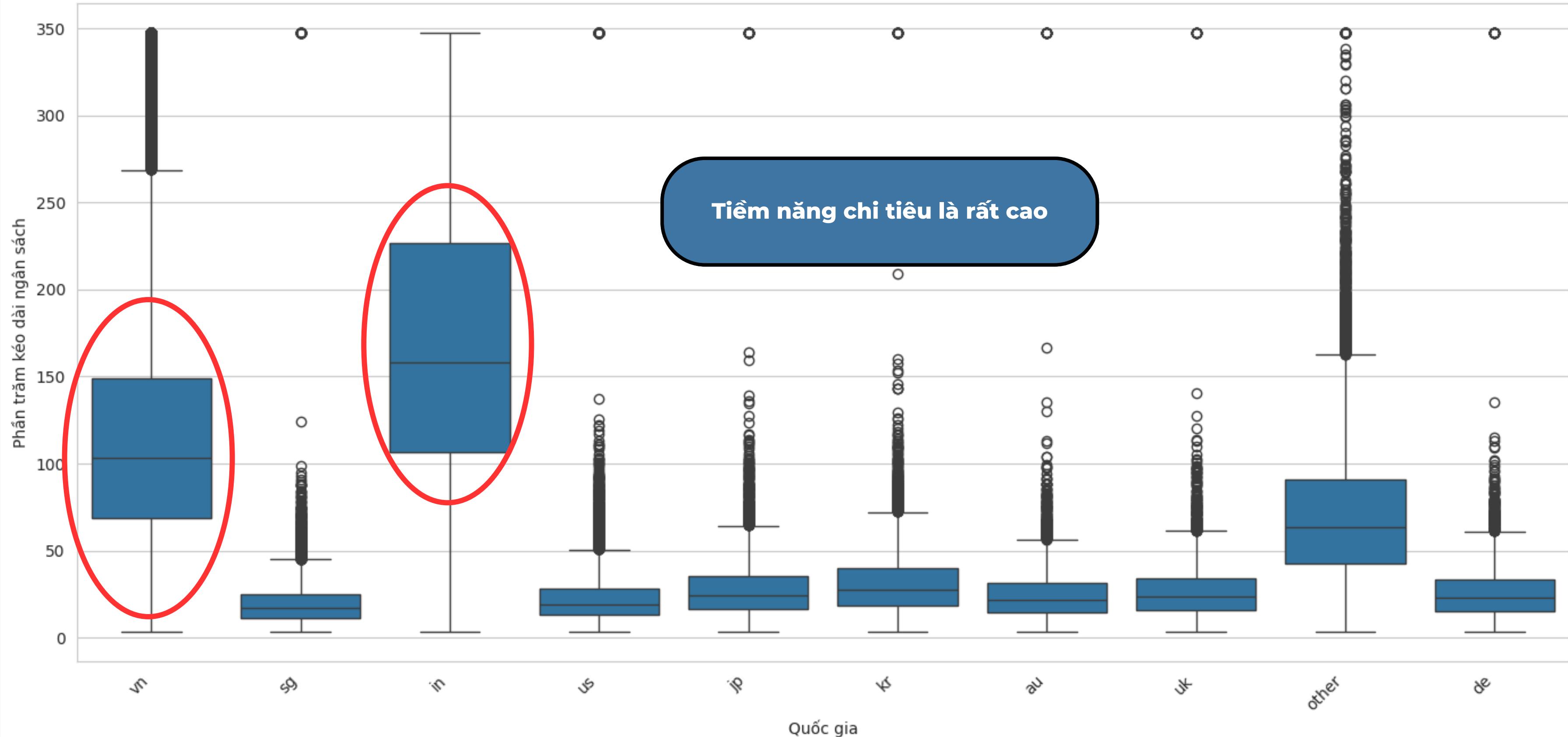
THEO QUỐC GIA

Phân bố phần trăm chi trả vượt ngân sách theo quốc gia



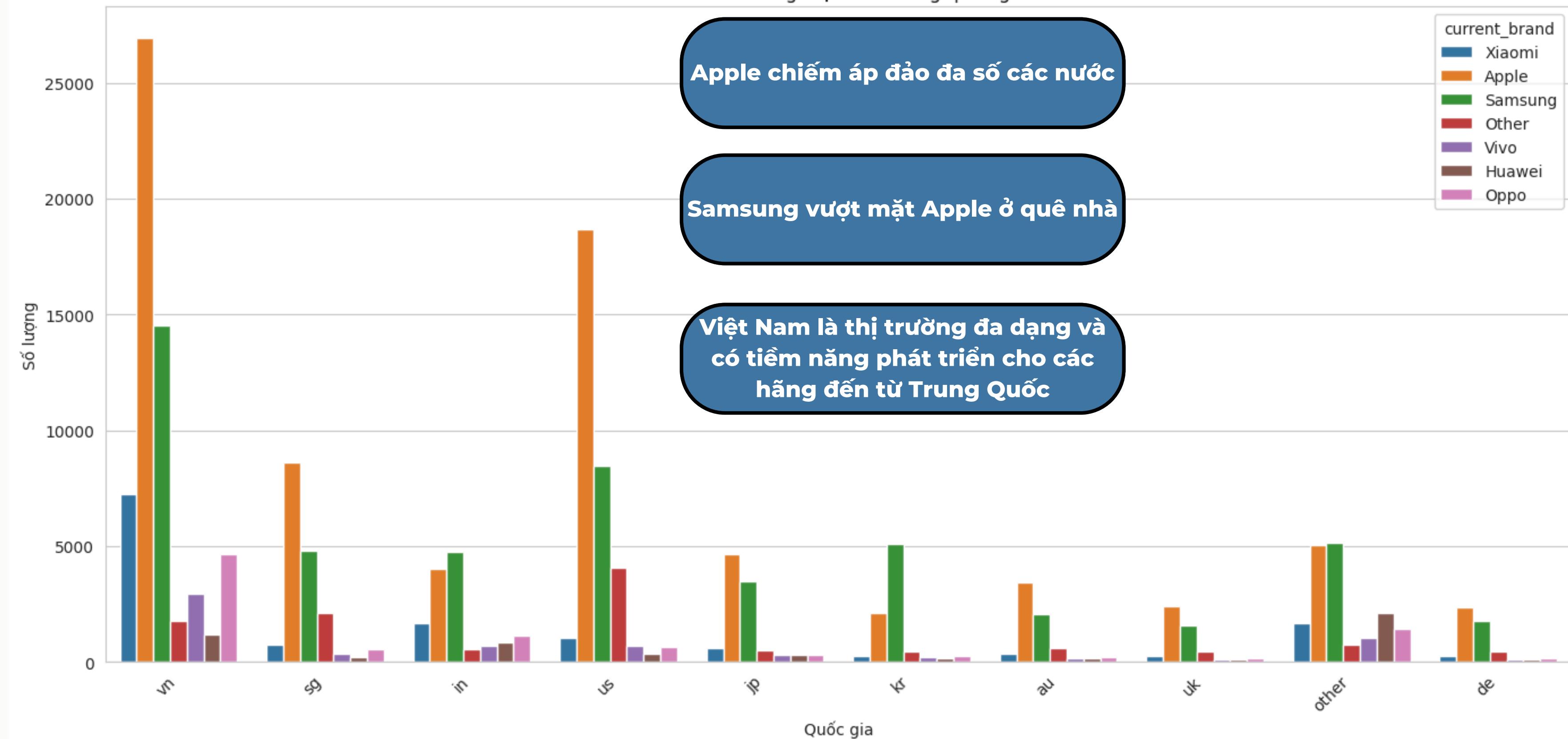
THEO QUỐC GIA

Phân bố phần trăm chi trả vượt ngân sách theo quốc gia



THEO QUỐC GIA

Phân bổ thương hiệu theo từng quốc gia



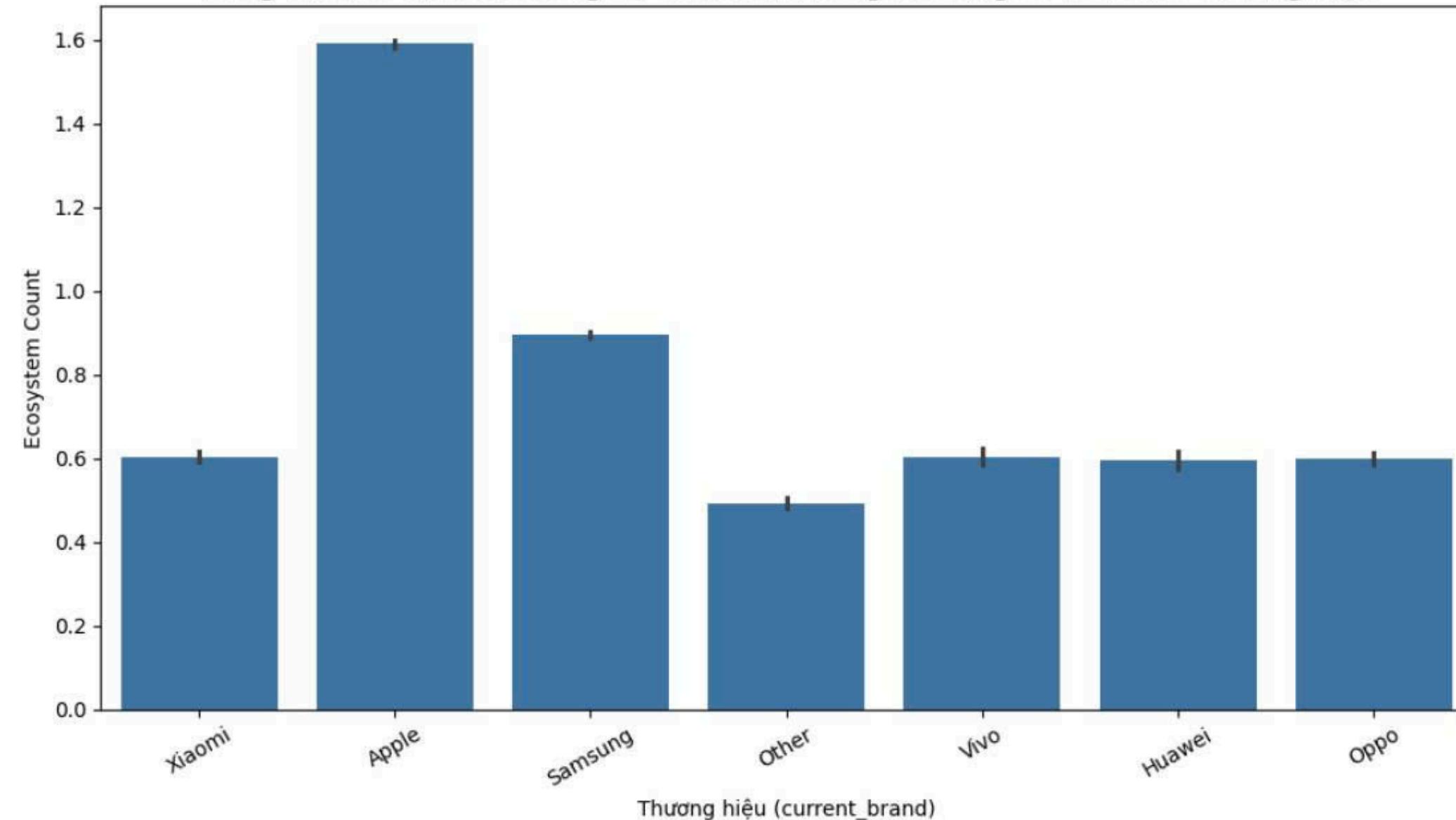
THEO THƯƠNG HIỆU

Xác định xu hướng tiêu dùng, mức độ trung thành, cũng như đánh giá sự khác biệt trong trải nghiệm thiết bị giữa các nhóm người dùng



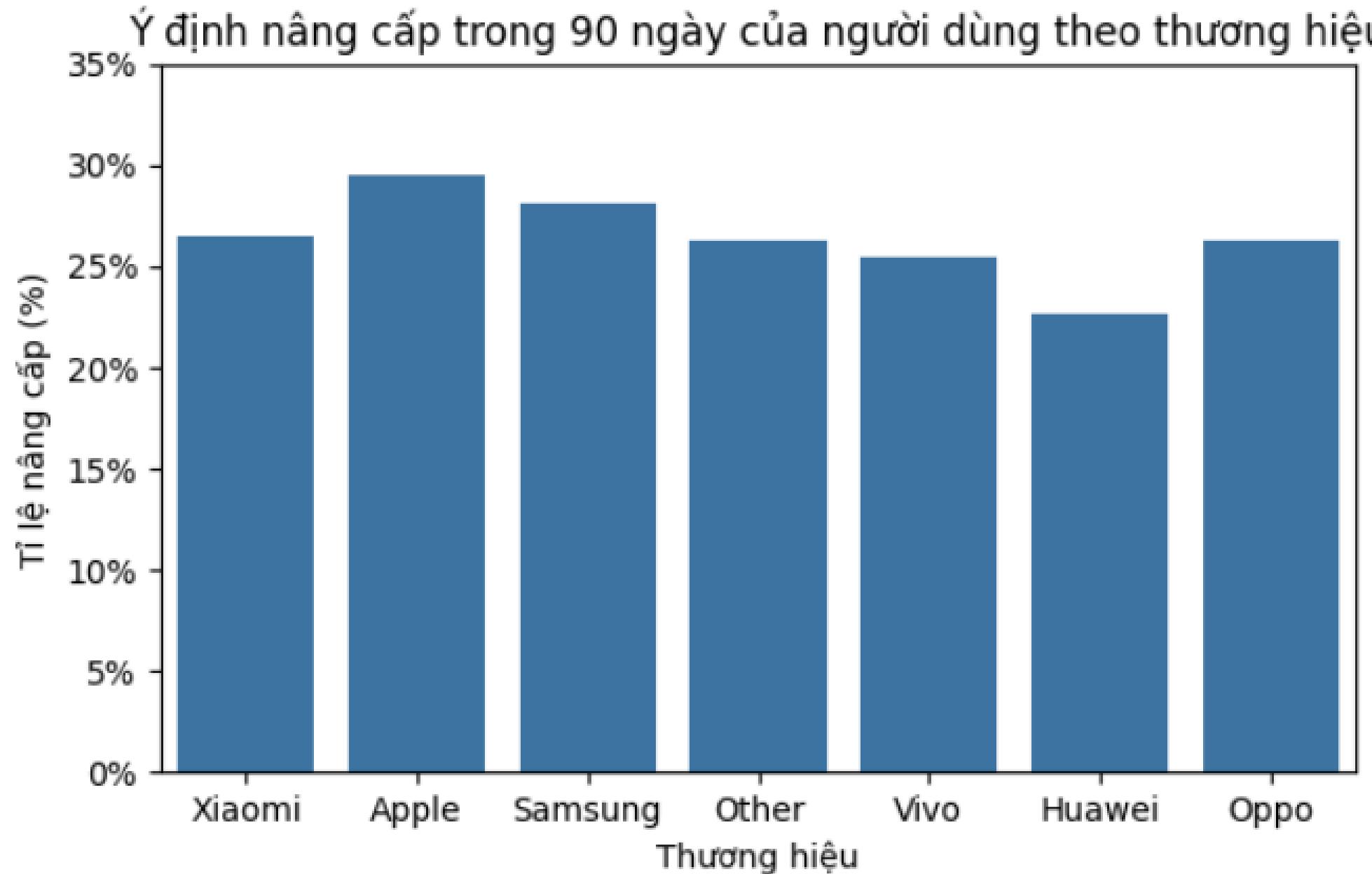
TRUNG BÌNH SỐ THIẾT BỊ CHUNG HỆ SINH THÁI MÀ NGƯỜI DÙNG SỞ HỮU

Trung bình số thiết bị chung hệ sinh thái mà người dùng sở hữu theo thương hiệu



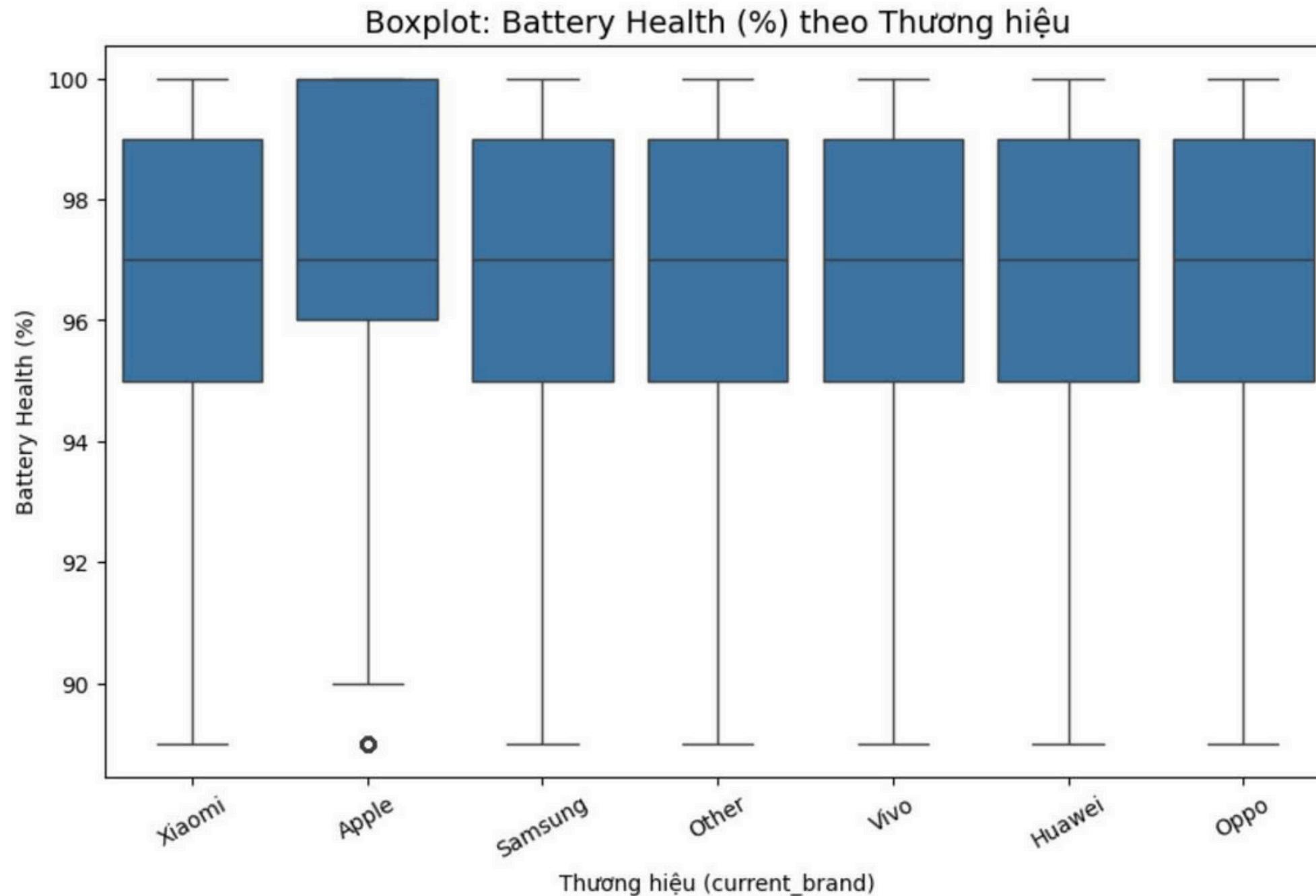
- Người dùng Apple sở hữu trung bình nhiều thiết bị nhất với ~1,6 sở hữu thêm thiết bị
- Samsung xếp thứ hai (~0.9), thể hiện hệ sinh thái đang phát triển nhưng vẫn kém Apple đáng kể.
- Các thương hiệu còn lại (Xiaomi, Vivo, Huawei, Oppo, Other) thấp hơn đáng kể, phản ánh hệ sinh thái chưa đủ mạnh hoặc ngại trải nghiệm các thiết bị chung hệ sinh thái

Ý ĐỊNH NÂNG CẤP



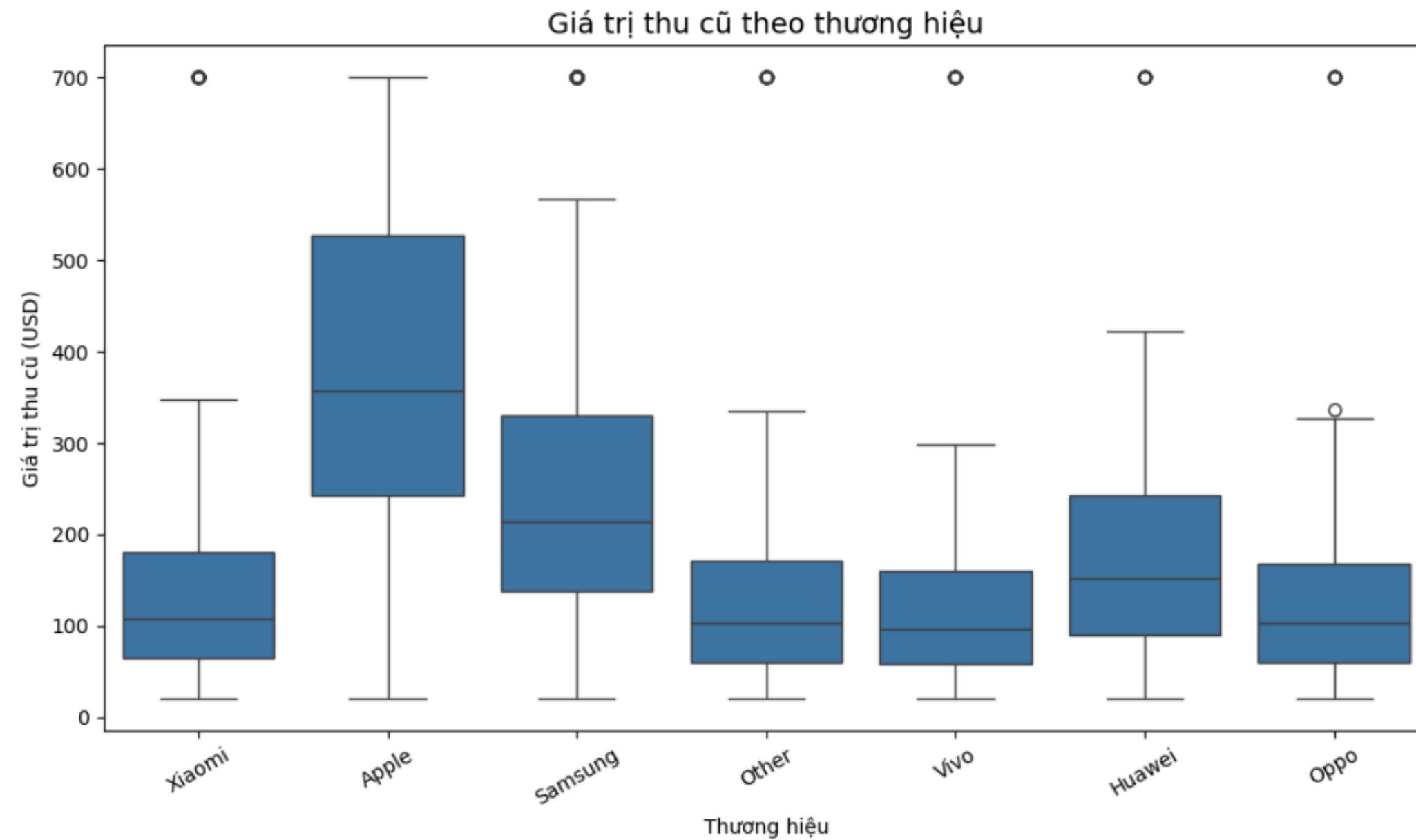
- Người dùng Apple có tỷ lệ dự định nâng cấp cao nhất (~30%)
- Samsung cũng thể hiện tỷ lệ nâng cấp cao
- Ngược lại, người dùng Huawei và Vivo có tỷ lệ nâng cấp thấp nhất,
=> Các thương hiệu cao cấp với hệ sinh thái mạnh kích thích nhu cầu nâng cấp cao hơn
Nhóm thương hiệu phổ thông có xu hướng giữ thiết bị lâu hơn.

PHẦN TRĂM ĐỘ CHAI PIN



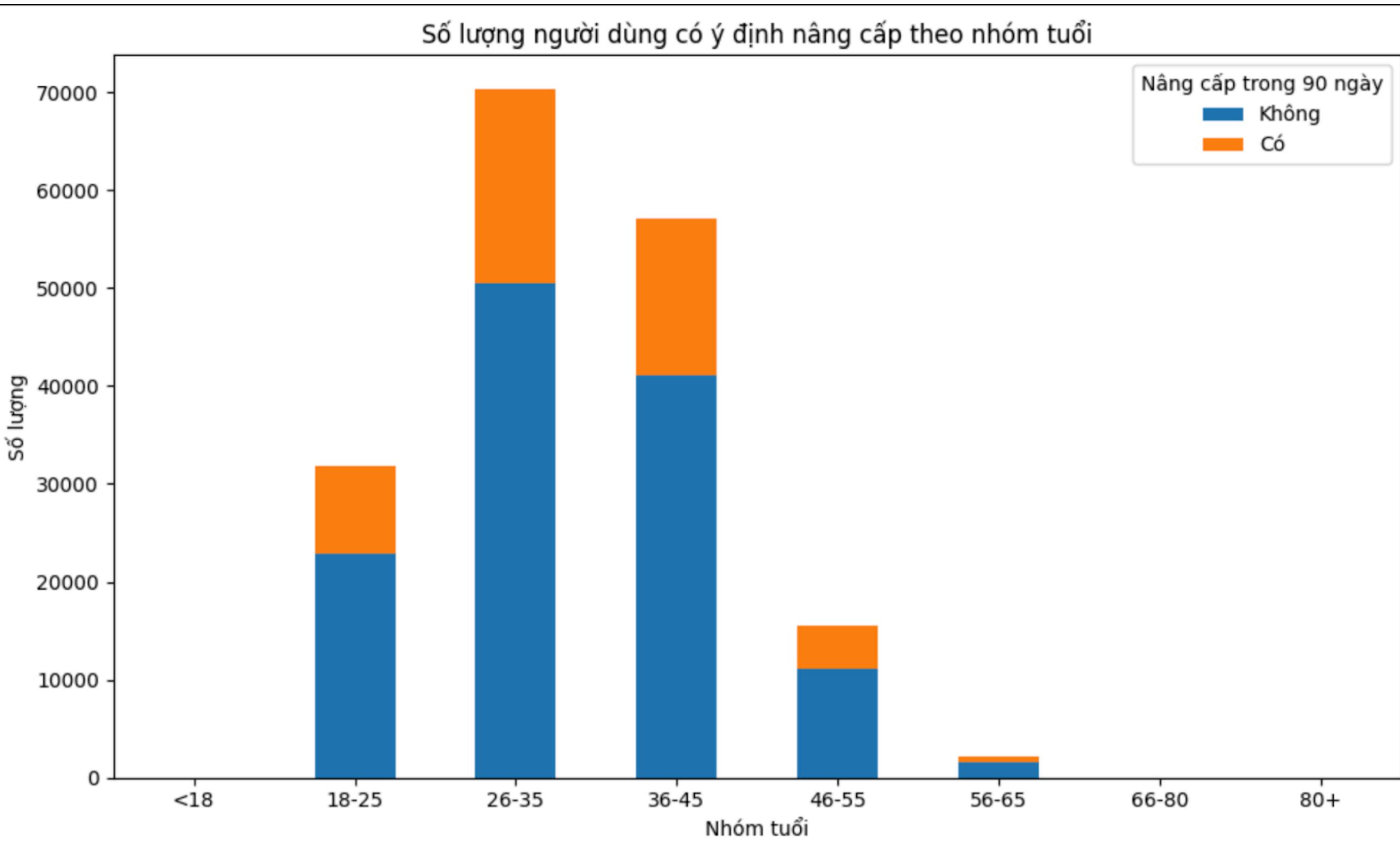
- Sức khỏe pin giữa các thương hiệu không khác biệt đáng kể, đa số thiết bị duy trì mức 95%–100%. Chỉ có Apple là nhỉnh hơn một chút cho thấy khả năng tối ưu năng lượng và quản lý pin tốt của hãng.
- Một số ngoại lệ pin dưới 90% xuất hiện chủ yếu ở Apple nhưng không ảnh hưởng đến xu hướng chung.

GIÁ TRỊ THU CŨ



- Apple có giá trị thu cũ cao nhất và ổn định nhất
 - Samsung đứng thứ hai nhưng chênh lệch đáng kể so với Apple.
 - Các thương hiệu khác (Xiaomi, Oppo, Vivo, Huawei, Other) có giá thu cũ thấp và biến động nhỏ
- Có thể giá trị thương hiệu và hệ sinh thái là yếu tố quan trọng quyết định khả năng giữ giá thiết bị.

THEO NHÂN KHẨU HỌC

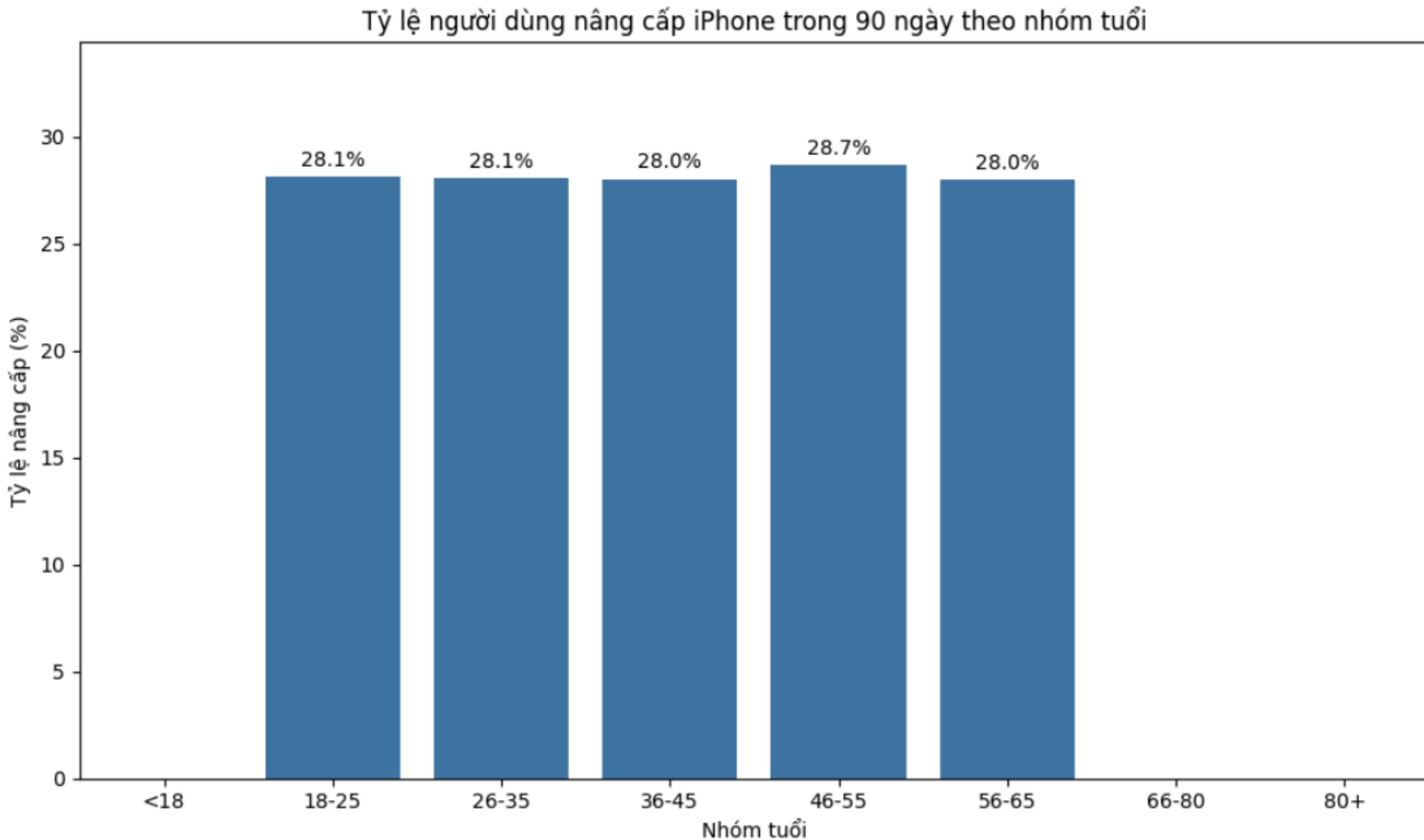


Nhóm 26-35 và 36-45 chiếm phần lớn mẫu của bộ dữ liệu
>> Nhóm tuổi năng động trong quyết định nâng cấp

Nhóm <18 và >56 chiếm rất nhỏ
>> Càng lớn tuổi, tốc độ chấp nhận công nghệ càng chậm

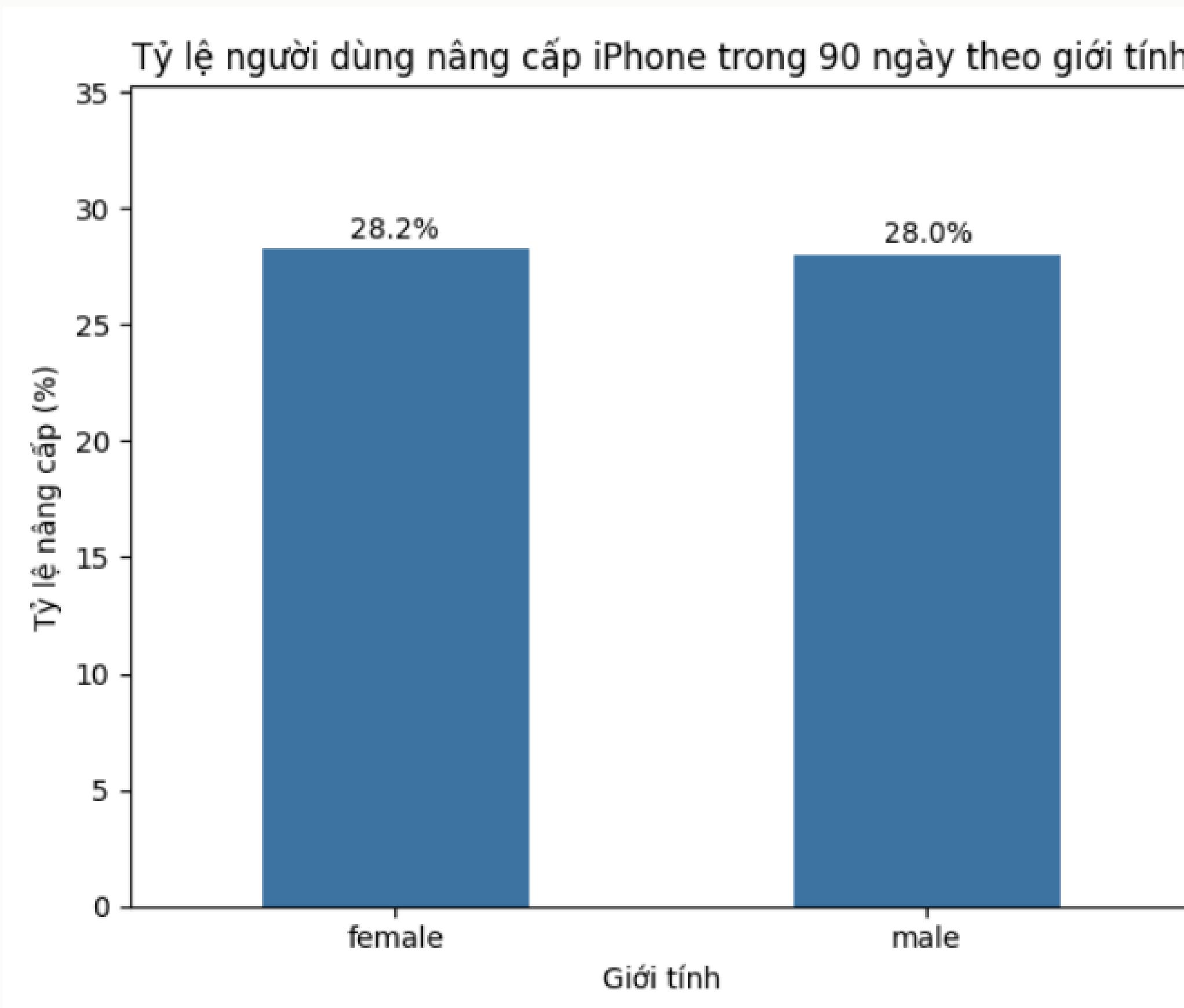
Nhóm 18-25 đáng chú ý mặc dù số lượng thấp hơn
>> Nhạy cảm với công nghệ, bắt xu hướng tốt

THEO NHÂN KHẨU HỌC

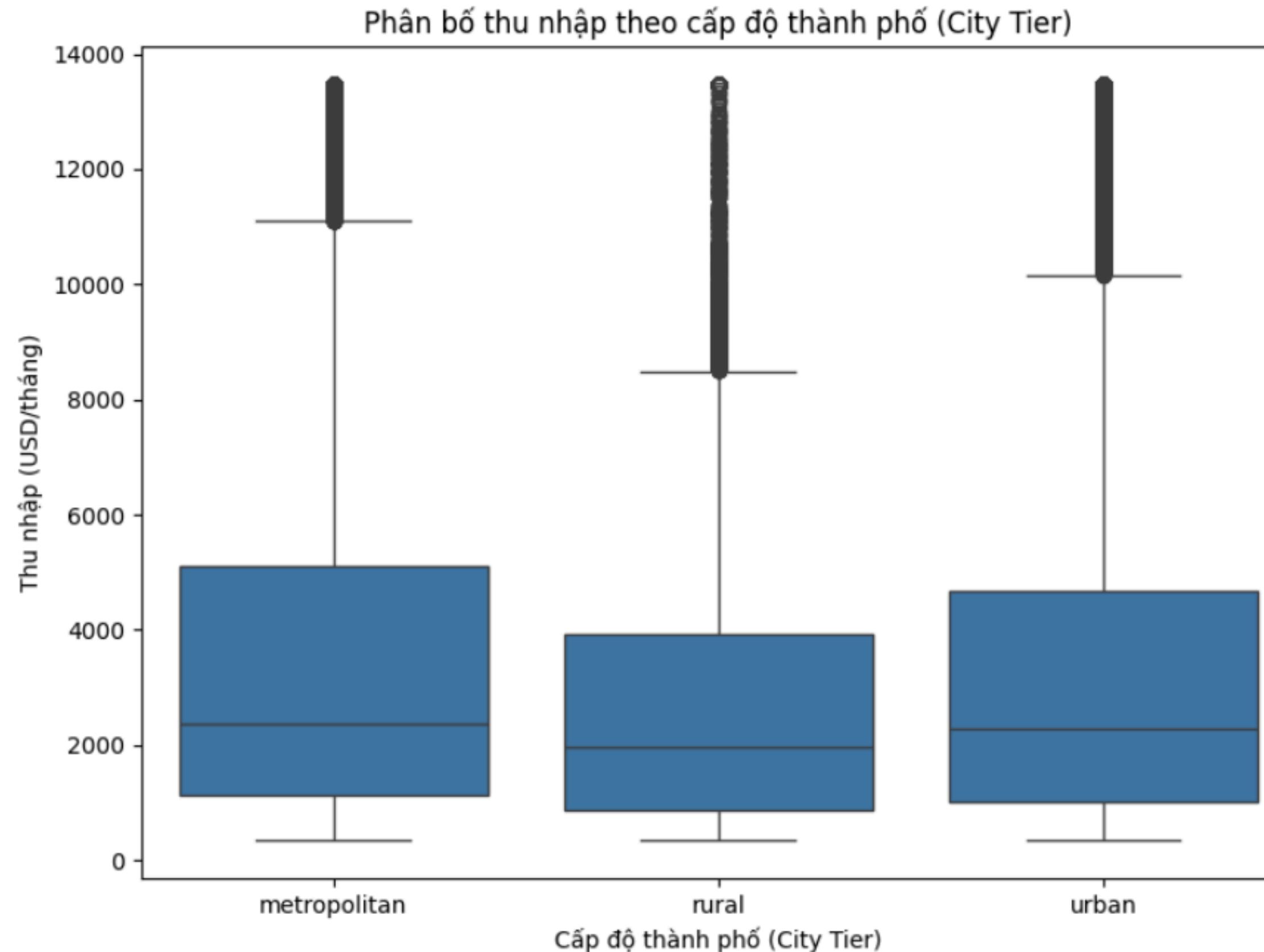


- **Tỷ lệ nâng cấp giữa các nhóm tuổi gần như ngang nhau, ~28%**
- **Tuy nhiên, sự chênh lệch giữa số lượng người trong từng nhóm là rất lớn**

THEO NHÂN KHẨU HỌC



THEO NHÂN KHẨU HỌC



Nhóm Metropolitan có:

- Thu nhập trung vị cao
- Phạm vi dao động rộng hơn
- Outliers >> có nhiều cá nhân thu nhập > 11.000USD

Nhóm Urban có:

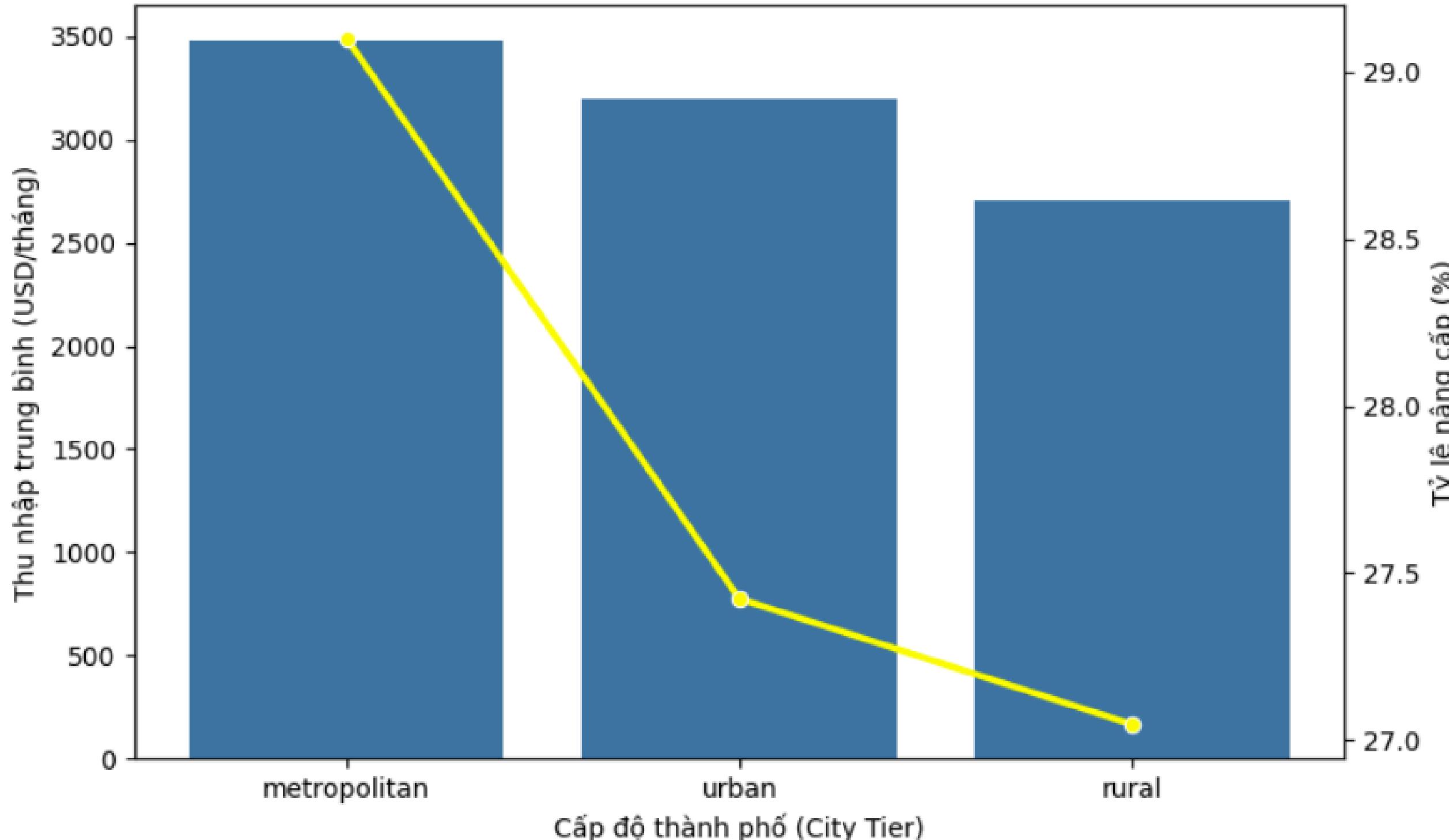
- Thu nhập trung vị tương đối cao
- Phạm vi dao động nhỏ hơn Metropolitan
- Outliers >> có nhiều cá nhân thu nhập > 10.000 USD

Nhóm Rural có:

- Thu nhập trung vị thấp nhất
- Phạm vi dao động nhỏ nhất

THEO NHÂN KHẨU HỌC

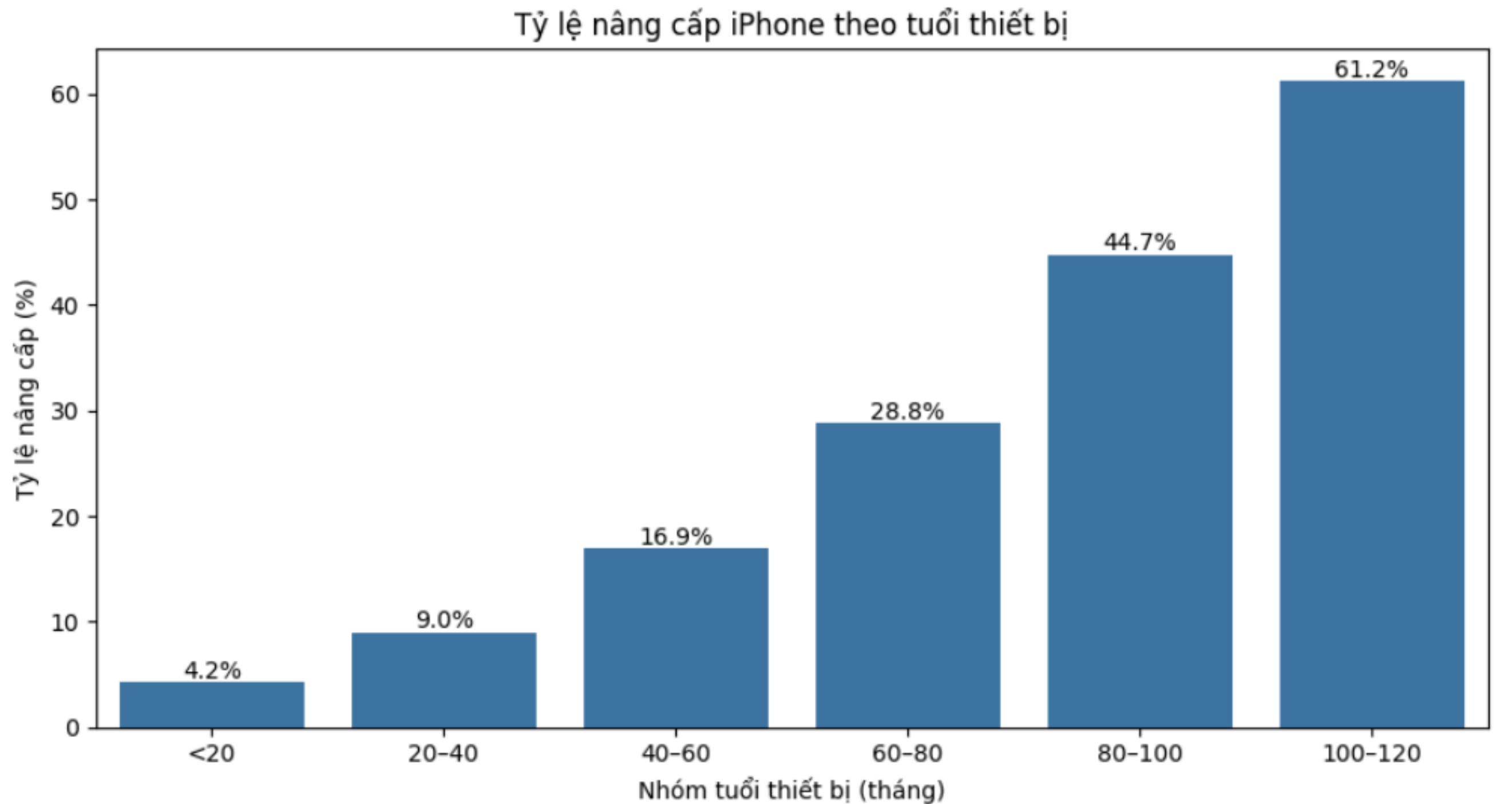
Mối tương quan giữa thu nhập và tỷ lệ nâng cấp theo cấp độ thành phố



Nhóm Metropolitan có:

- Thu nhập trung TB cao nhất
- Tỷ lệ nâng cấp điện thoại cao nhất

- Thu nhập trung bình giảm dần theo Urban và Rural
- Có sự chênh lệch giữa tỷ lệ nâng cấp điện thoại của Urban và Rural. Tuy nhiên, không mạnh như Metropolitan

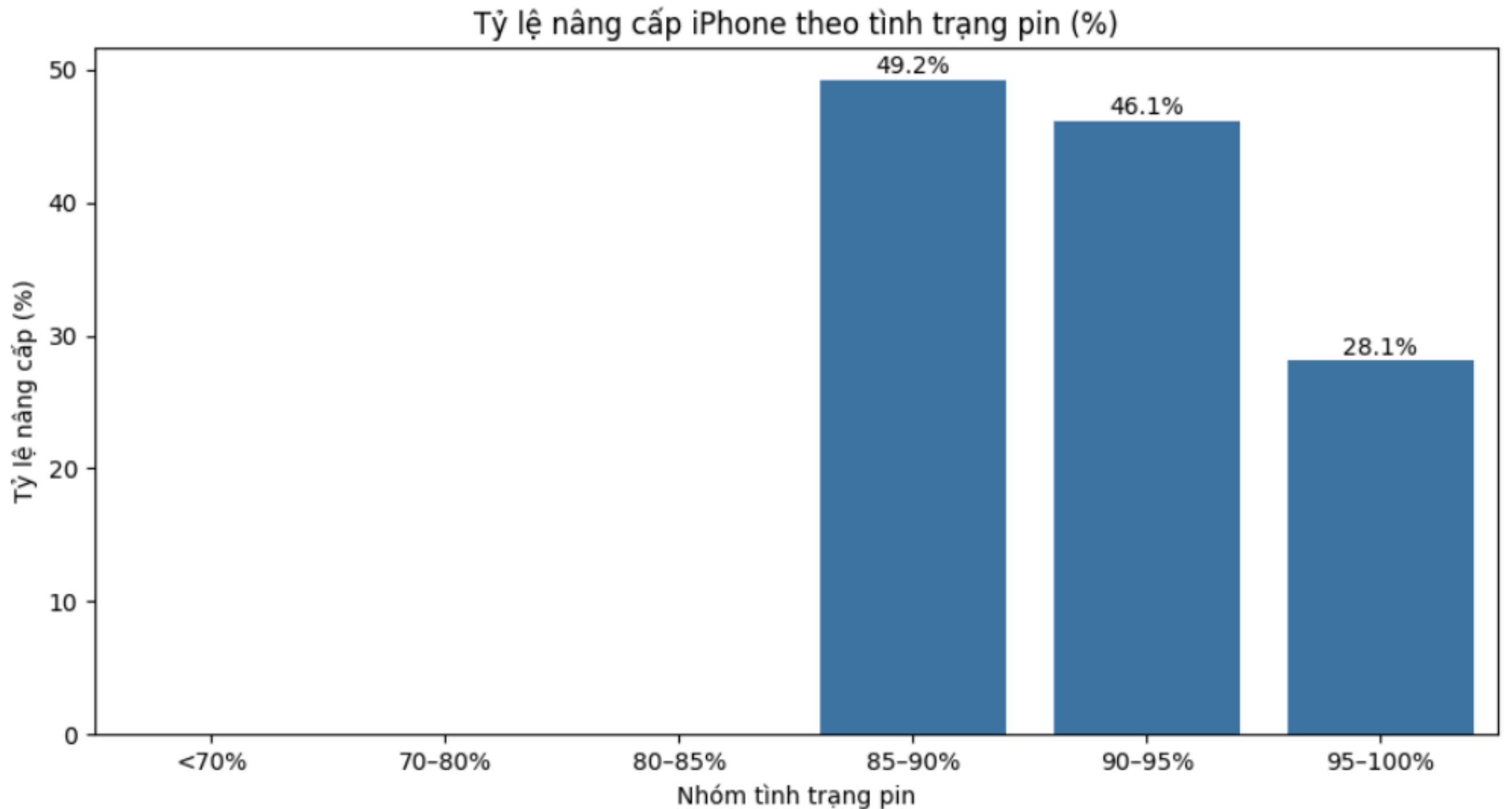


- Tỷ lệ nâng cấp tỷ lệ thuận với tuổi của thiết

- Nhóm <20 và 20-40 có tỷ lệ rất thấp
- Tuy nhiên tăng vọt ở nhóm 80-100 và đặc biệt là nhóm 100-120, đạt ngưỡng 61.2%

Tuổi của thiết bị ảnh hưởng mạnh mẽ đến quyết định nâng cấp điện thoại

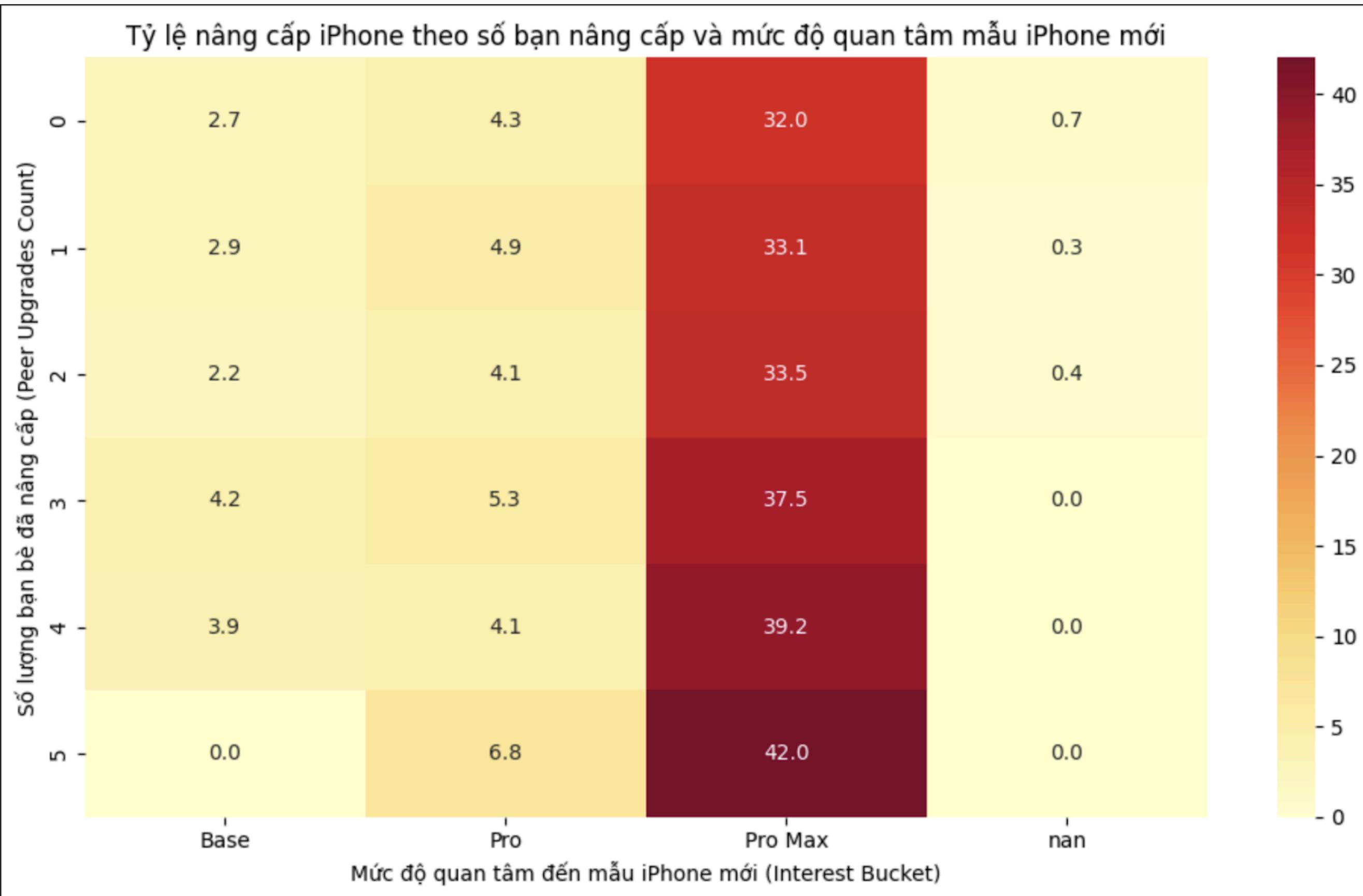
TÂM LÝ, HÀNH VI CỦA NGƯỜI TIÊU DÙNG



- Tỷ lệ nâng cấp ở nhóm >95% tương đối thấp

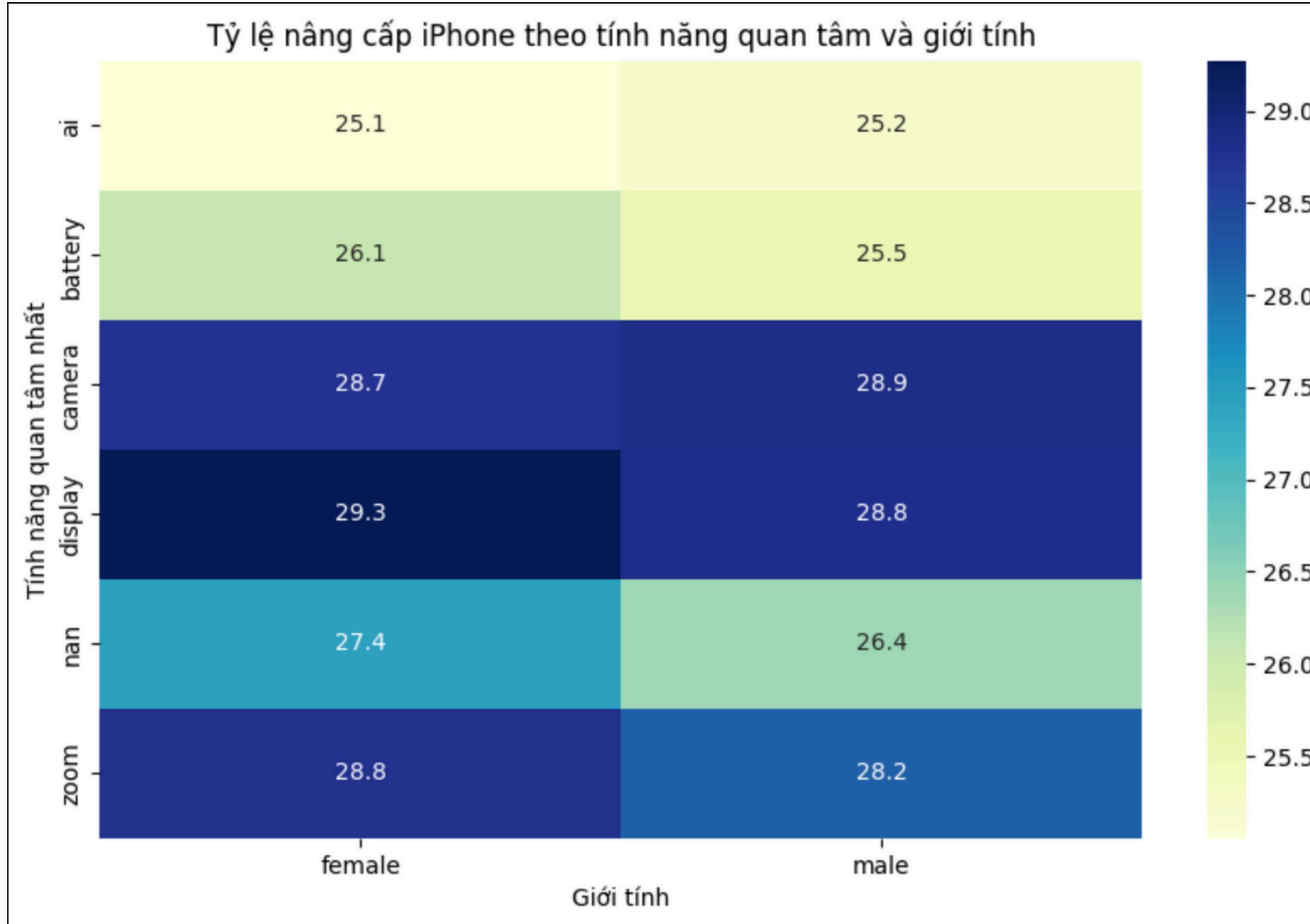
- Tăng vọt ở nhóm 90-95% và cao nhất ở nhóm 85-90%

- Hầu như không có người dùng pin dưới 80%
- Sự nhạy cảm của người tiêu dùng, nhận thấy hiệu năng giảm là thay pin, không đợi <80%
- Tuy nhiên có thể xuất phát từ thiên lệch quan sát (observation bias)



- **Người tiêu dùng quan tâm phân khúc cao cấp (Pro Max) có tỷ lệ nâng cấp cao nhất và dễ bị ảnh hưởng bởi bạn bè xung quanh**

- **Phân khúc Base và Pro ổn định hơn, không bị ảnh hưởng nhiều bởi bạn bè xung quanh**

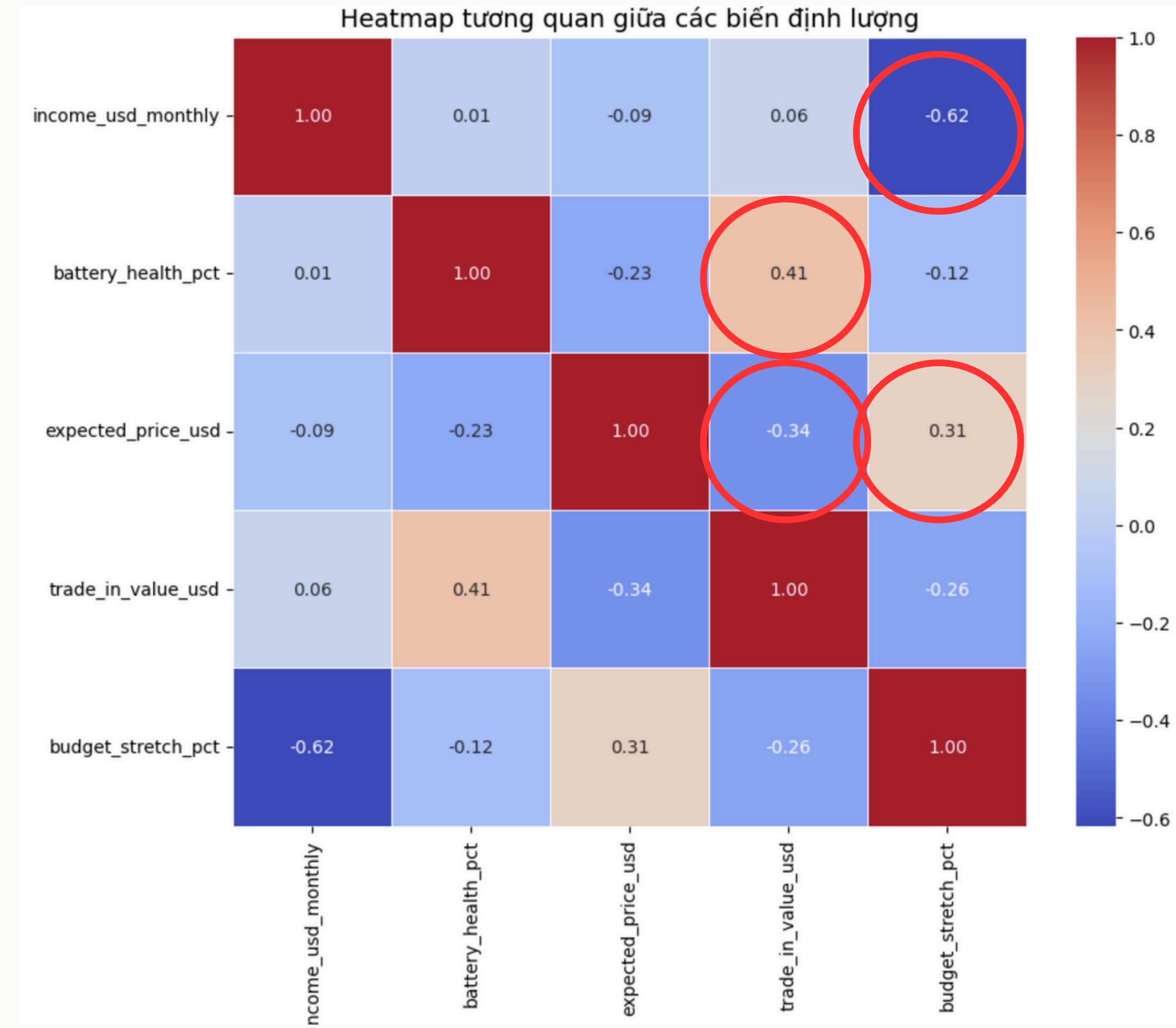


- Xu hướng nâng cấp dựa trên trải nghiệm thị giác và hình ảnh (Display và Camera)

Giới tính không phải là yếu tố quyết định trong hành vi của người tiêu dùng về việc nâng cấp điện thoại.

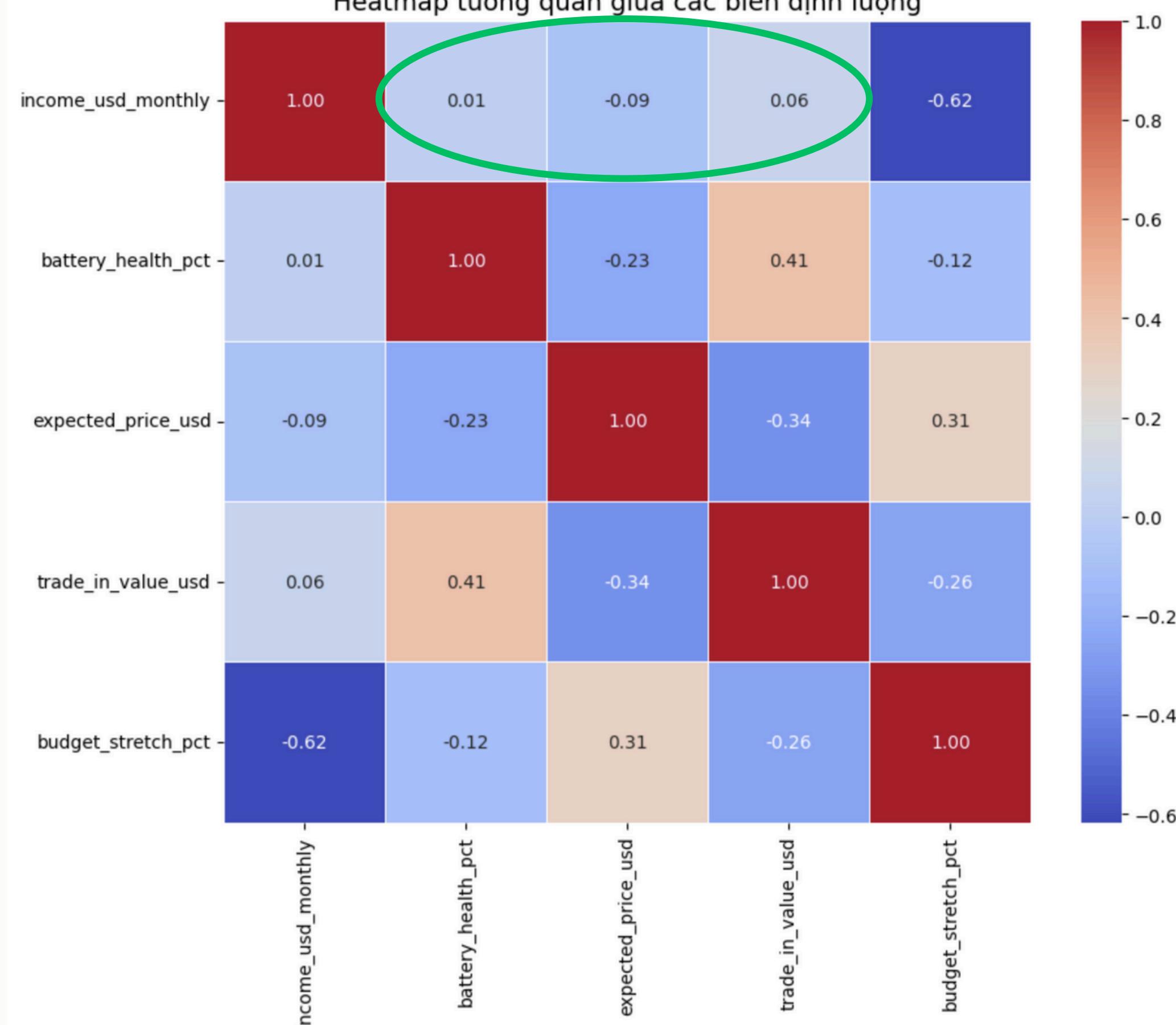
III

PHÂN TÍCH TƯƠNG QUAN



PHÂN TÍCH TƯƠNG QUAN

Heatmap tương quan giữa các biến định lượng

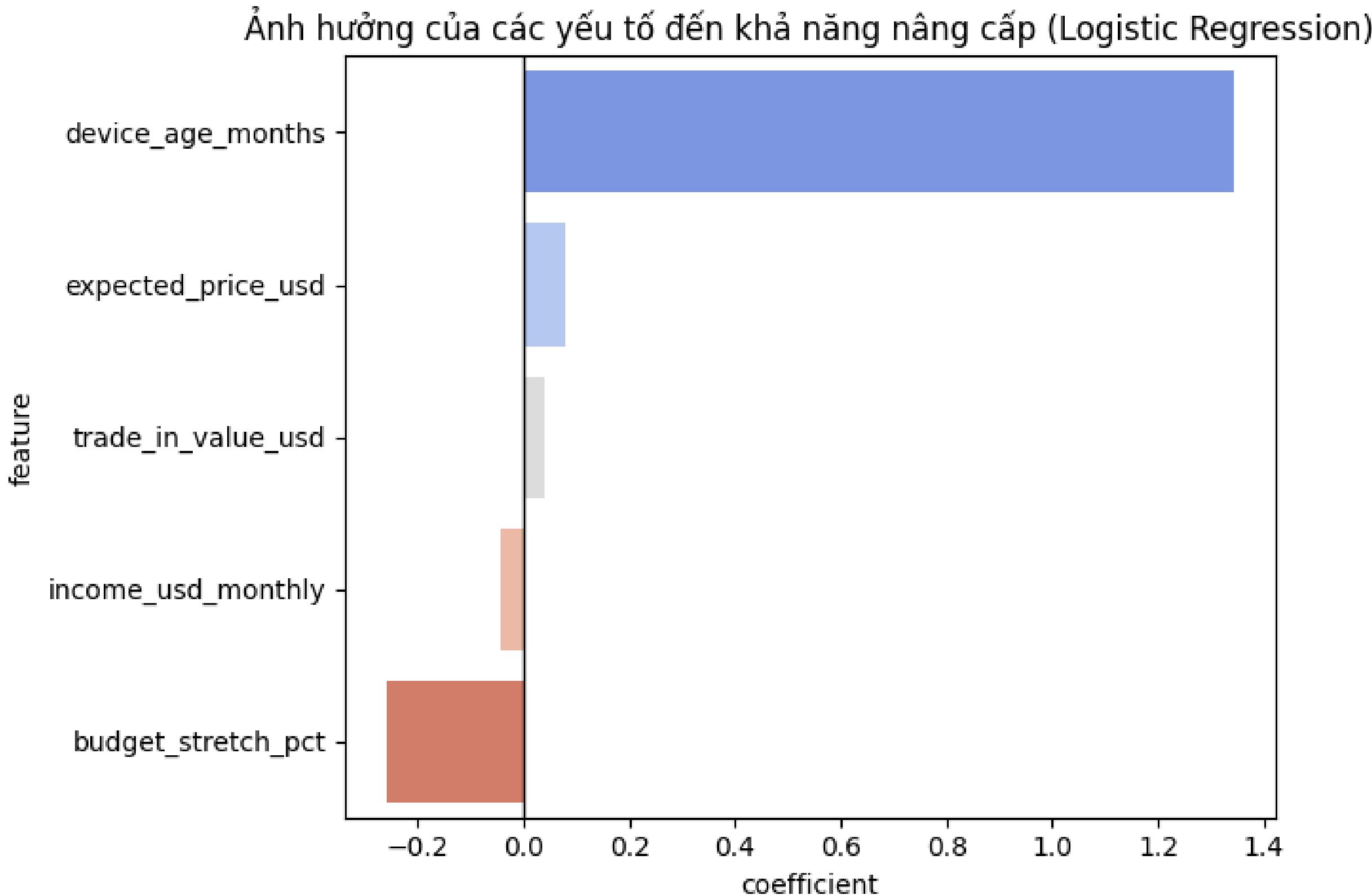


ỨNG DỤNG LOGISTIC MODEL

income_usd_monthly,
expected_price_usd,
device_age_months,
trade_in_value_usd và
budget_stretch_pct.

==== Classification report ===					
		precision	recall	f1-score	support
	0	0.82	0.88	0.85	25457
	1	0.62	0.49	0.55	9959
	accuracy			0.77	35416
	macro avg	0.72	0.69	0.70	35416
	weighted avg	0.76	0.77	0.76	35416
==== Confusion matrix ===					
	[[22475 2982]				
	[5098 4861]]				
==== Hệ số và Odds Ratio ===					
	feature	coefficient	odds_ratio		
1	device_age_months	1.342156	3.827		
3	expected_price_usd	0.078357	1.082		
2	trade_in_value_usd	0.040939	1.042		
0	income_usd_monthly	-0.043720	0.957		
4	budget_stretch_pct	-0.258450	0.772		

ỨNG DỤNG LOGISTIC MODEL



IV

THỬ NGHIỆM PHÂN LỚP

TIỀN XỬ LÝ

Bỏ các cột không cần thiết:

```
dfM = dfM.drop(['device_age_group', 'battery_group', 'launch', 'respondent_id', 'age_group'], axis=1)  
dfM.head()
```

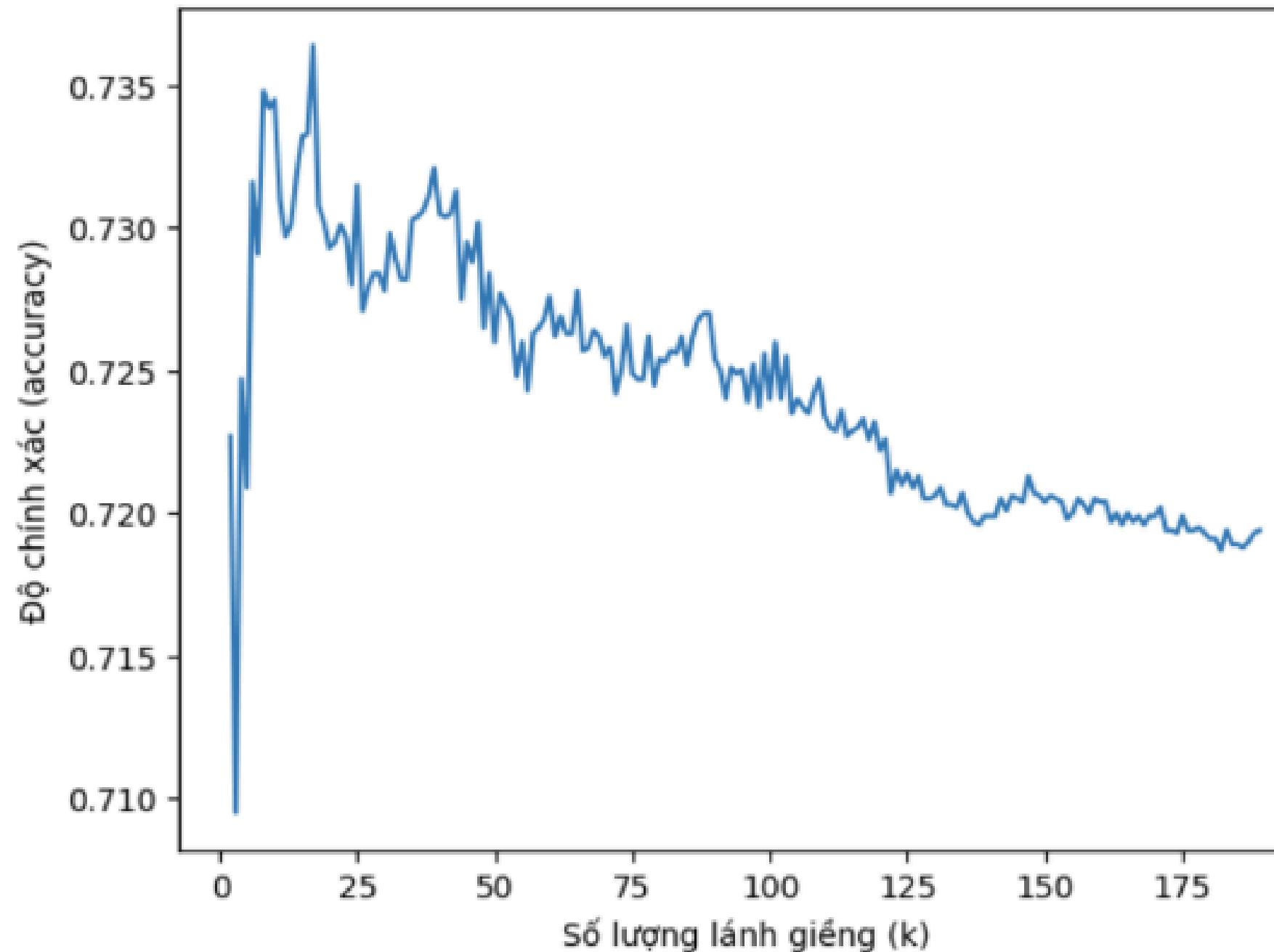
One-Hot Encoding các biến định tính:

```
categorical_cols = [  
    ... 'gender', 'country', 'city_tier', 'current_brand',  
    ... 'is_storage_full', 'ecosystem_count', 'ad_recall', 'interest_new_model_bucket', 'most_interested_function'  
]  
  
df_encoded = pd.get_dummies(dfM, columns=categorical_cols, drop_first=True)  
df_encoded['upgrade_within_90days'] = df_encoded['upgrade_within_90days'].astype(int)  
df_encoded.head()
```

✓ 0.1s

TIỀN XỬ LÝ

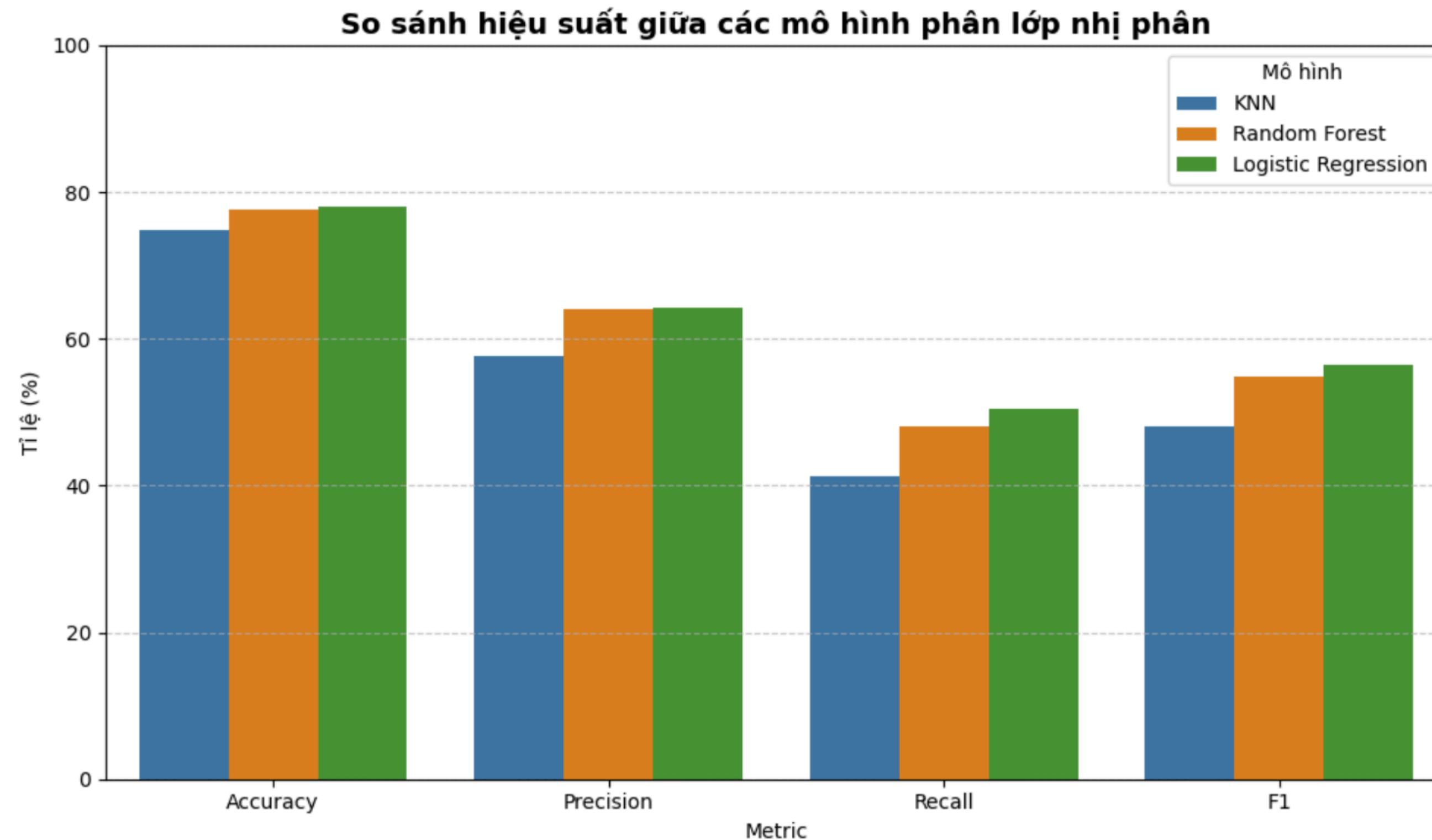
Thử nghiệm qua các giá trị K để tìm giá trị phù hợp nhất



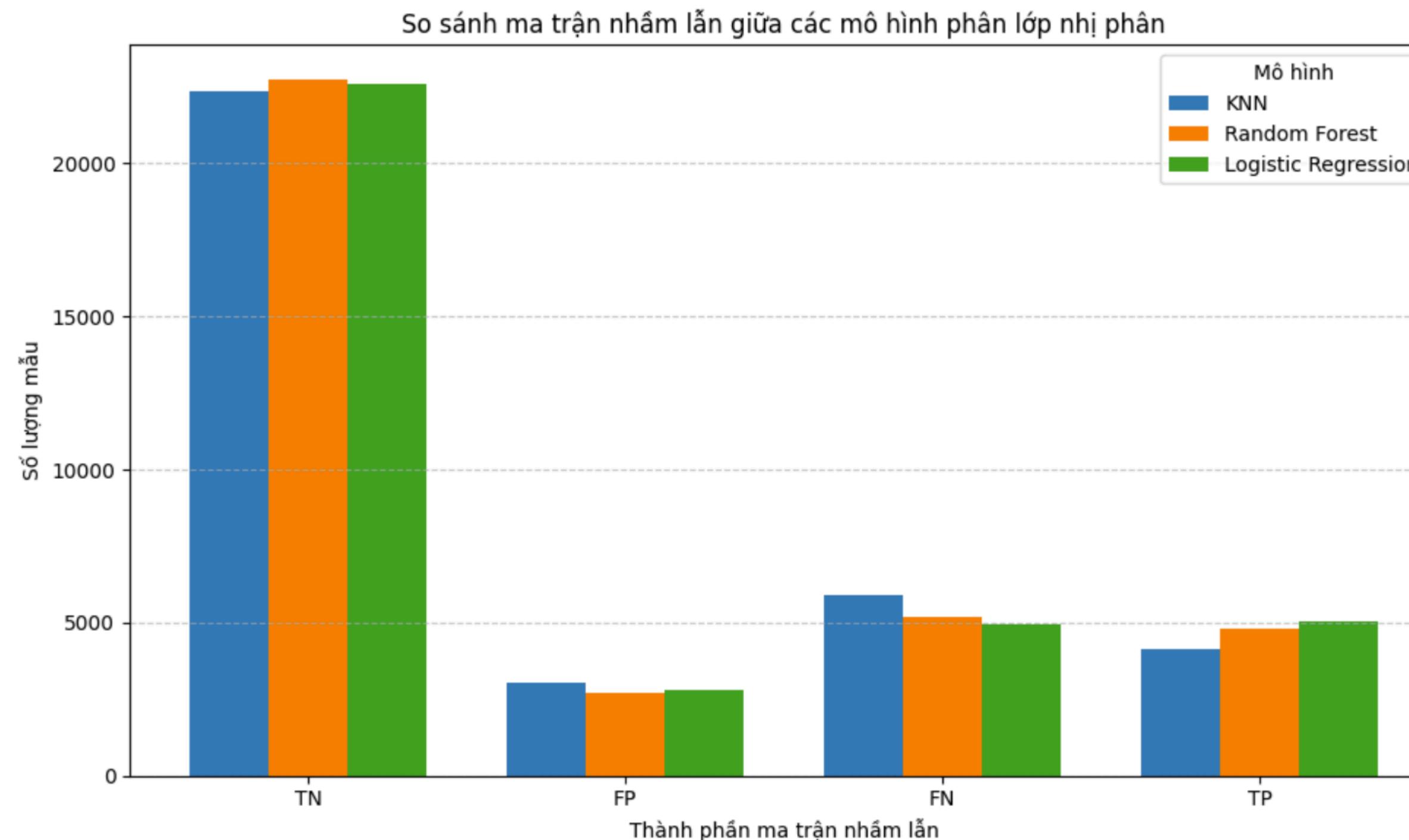
```
k = list_k[np.argmax(list_acc)]
print(f'Giá trị k = {k:d} tạo mô hình k-NN có độ chính xác cao nhất.')
✓ 0.0s
```

Giá trị k = 17 tạo mô hình k-NN có độ chính xác cao nhất.

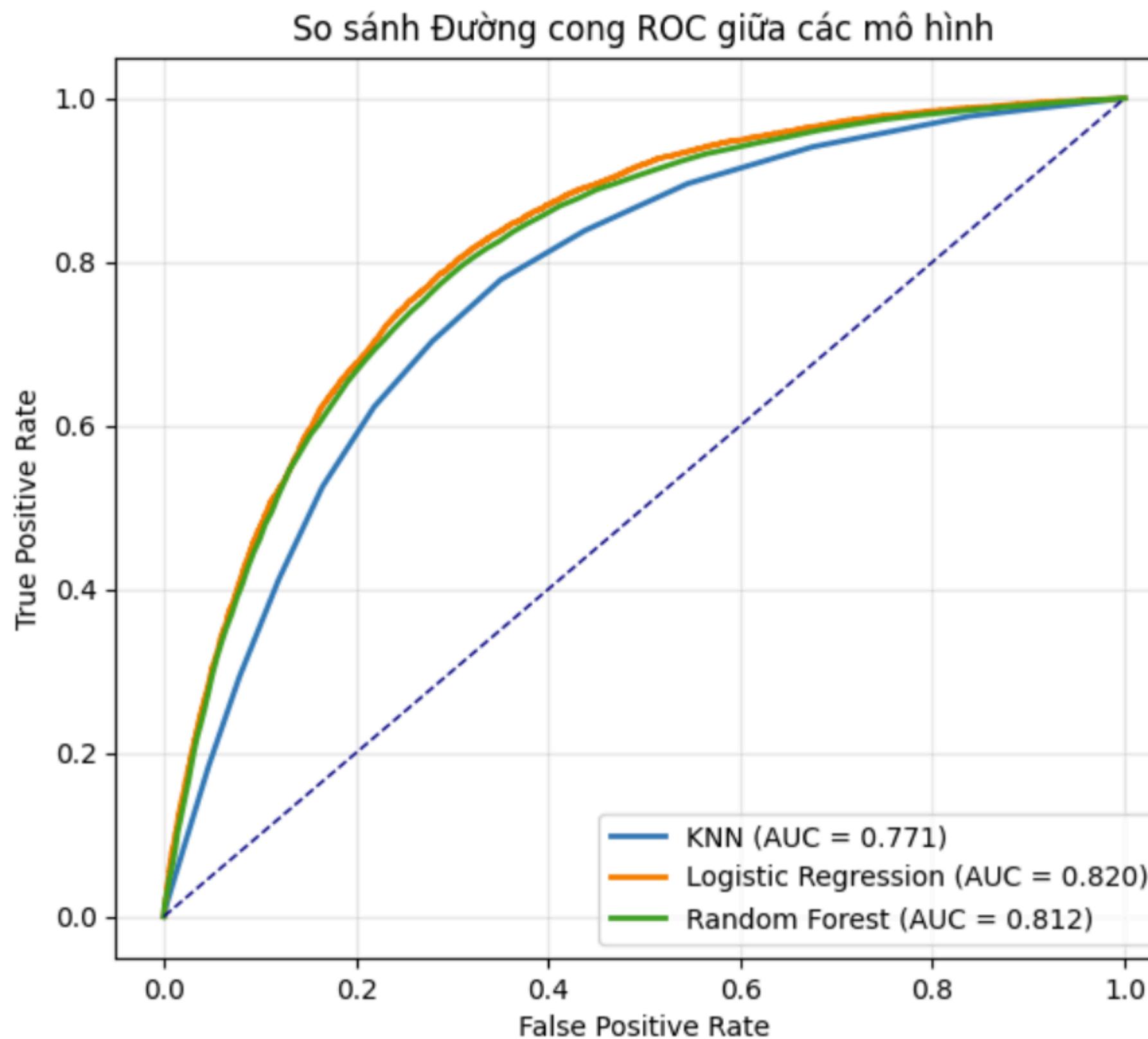
SO SÁNH CÁC MODEL



SO SÁNH CÁC MODEL



SO SÁNH CÁC MODEL



V

KẾT LUẬN

- Nhóm tuổi 26–35, sống ở đô thị lớn, có thiết bị cũ và pin yếu → xác suất nâng cấp cao nhất.
- Người dùng Apple có tỷ lệ nâng cấp cao và gắn bó hệ sinh thái hơn Android.
- Thị trường phát triển ổn định, còn Việt Nam & Ấn Độ nhạy cảm về giá nhưng chi tiêu mạnh khi đã quyết định.
- Yếu tố ảnh hưởng mạnh nhất: Tuổi thiết bị, sức khỏe pin, giá thu cũ, mức sẵn sàng chi.

ỨNG DỤNG

- Hỗ trợ doanh nghiệp tối ưu marketing, trade-in
- Cung cấp thêm kỹ năng tiền xử lý dữ liệu - EDA – trực quan – mô hình hóa dữ liệu.