

TRƯỜNG ĐẠI HỌC TRÀ VINH
KHOA KỸ THUẬT VÀ CÔNG NGHỆ
BỘ MÔN CÔNG NGHỆ THÔNG TIN



THỰC TẬP ĐỒ ÁN MÔN HỌC
HỌC KỲ II, NĂM HỌC 2024-2025

PHÂN TÍCH BỘ DỮ LIỆU DÒNG SẢN PHẨM HOÀN CHỈNH CỦA NVIDIA

Giảng viên hướng dẫn:

Nguyễn Bảo Ân

Sinh viên thực hiện:

Châu Thanh Thiện – 110122163

Võ Phước Toàn – 110122186

Nguyễn Hoàng Yến – 110122212

Lớp: DA22TTC

Trà Vinh, tháng 02 năm 2025

TRƯỜNG ĐẠI HỌC TRÀ VINH
KHOA KỸ THUẬT VÀ CÔNG NGHỆ
BỘ MÔN CÔNG NGHỆ THÔNG TIN



THỰC TẬP ĐỒ ÁN MÔN HỌC
HỌC KỲ II, NĂM HỌC 2024-2025

PHÂN TÍCH BỘ DỮ LIỆU DÒNG SẢN PHẨM HOÀN CHỈNH CỦA NVIDIA

Giảng viên hướng dẫn:

Nguyễn Bảo Ân

Sinh viên thực hiện:

Châu Thanh Thiện – 110122163

Võ Phước Toàn – 110122186

Nguyễn Hoàng Yến – 110122212

Lớp: DA22TTC

Trà Vinh, tháng 02 năm 2025

[illegible]

(Ký và ghi gõ)

.....

[illegible]

(Ký và ghi rõ họ tên)

LỜI CẢM ƠN

Em xin chân thành cảm ơn các thầy cô trường Đại học Trà Vinh, đặc biệt là các thầy cô khoa Kỹ thuật & Công nghệ đã tạo cơ hội cho em được thực tập và giao lưu. Em có thể tránh được các vấn đề và nhầm lẫn trong môi trường làm việc trong tương lai.

Em xin cảm ơn thầy Nguyễn Bảo Ân. Nhờ sự giúp đỡ và hướng dẫn tận tình của thầy từ đầu đến cuối đồ án, em đã hoàn thành đồ án đúng thời hạn và tích lũy được một vốn kiến thức quý báu.

Mặc dù đã cố gắng hết sức để hoàn thành đề tài này nhưng do hạn chế về thời gian cũng như kiến thức chuyên môn nên trong quá trình nghiên cứu, đánh giá và trình bày đề tài, em còn nhiều thiếu sót. Em rất mong được sự quan tâm, góp ý của thầy bộ môn để đồ án môn học của em được hoàn thành và hoàn thiện hơn. Em xin chân thành cảm ơn.

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN.....	1
1. Giới thiệu	1
2. Phát biểu vấn đề.....	2
3. Giới thiệu về bộ dữ liệu.	3
3.1. Các đặc điểm chính của bộ dữ liệu:	3
3.1.1 Thông tin cơ bản:.....	3
3.1.2 Thông số kỹ thuật:	4
3.1.3 Hiệu suất tính toán:.....	4
CHƯƠNG 2: PHÂN TÍCH DỮ LIỆU.....	5
1. Mô tả dữ liệu.....	5
1.1. Xem trước dữ liệu	5
1.2. Kiểm tra thông tin tổng quan.	5
2. Thống kê mô tả dữ liệu số	5
3. Dữ liệu thiếu và ngoại lệ.....	6
4. Các câu hỏi nghiên cứu.	6
5. Các câu hỏi nghiên cứu.	6
6. Phân tích bộ dữ liệu và trả lời câu hỏi.....	8
CHƯƠNG 3: KẾT LUẬN	25
1. Xu hướng giá và hiệu năng	25
2. Phân khúc thị trường	25
3. Tiến hóa công nghệ	25
4. Mối quan hệ giữa các thông số	25
5. Đặc điểm từng phân khúc	25
DANH MỤC TÀI LIỆU THAM KHẢO.....	27

DANH MỤC HÌNH ẢNH

Hình 1.1 Logo Nvidia.....	1
---------------------------	---

CHƯƠNG 1: TỔNG QUAN

1. Giới thiệu

Nvidia là gì?

Nvidia là một công ty công nghệ đa quốc gia của Mỹ, chuyên thiết kế bộ xử lý đồ họa (GPU - Graphics Processing Unit), hệ thống trên chip (SoC - System on a Chip) và các giải pháp trí tuệ nhân tạo (AI - Artificial Intelligence).



Hình 1.1 Logo Nvidia

Các dòng sản phẩm của Nvidia.

- GPU GeForce – Dành cho gaming và đồ họa chuyên nghiệp.
- GPU Quadro (nay là RTX A-series) – Dành cho thiết kế, kỹ thuật, dựng phim.
- GPU Tesla (nay là A100, H100) – Dành cho AI, deep learning, siêu máy tính.
- GPU Titan – Dành cho workstation cao cấp.
- Nvidia Jetson – Dành cho robot, IoT, AI nhúng.
- Nvidia Shield – Thiết bị giải trí đa phương tiện.

Lý do chọn đề tài:

Trong thời đại công nghệ số phát triển mạnh mẽ, bộ xử lý đồ họa (GPU - Graphics Processing Unit) đã trở thành một trong những thành phần quan trọng nhất trong hệ sinh thái công nghệ. GPU không chỉ đóng vai trò cốt lõi trong xử lý đồ họa và trò chơi điện tử, mà còn được ứng dụng rộng rãi trong các lĩnh vực trí tuệ nhân tạo (AI), học sâu (Deep Learning), điện toán hiệu năng cao (HPC - High-Performance Computing), khoa học dữ liệu và mô phỏng vật lý.

Với nhu cầu ngày càng tăng, các hãng sản xuất GPU như Nvidia, AMD và Intel liên tục tung ra các dòng sản phẩm mới với thông số kỹ thuật mạnh mẽ, nhằm đáp ứng các nhu cầu khác nhau từ người dùng cá nhân đến các trung tâm dữ liệu lớn. Nvidia, một trong những công ty dẫn đầu trong lĩnh vực GPU, đã phát triển nhiều dòng sản phẩm khác nhau, bao gồm:

GPU dành cho người tiêu dùng (Consumer GPUs): Các dòng GeForce RTX và GTX, được thiết kế chủ yếu cho chơi game và các ứng dụng sáng tạo nội dung.

GPU chuyên nghiệp (Professional GPUs): Các dòng Quadro RTX và RTX A-series, hướng đến các chuyên gia đồ họa, kiến trúc sư, và kỹ sư thiết kế.

GPU trung tâm dữ liệu (Datacenter GPUs): Các dòng Tesla, A100, H100, tối ưu hóa cho các ứng dụng AI, phân tích dữ liệu lớn và điện toán đám mây.

Việc hiểu rõ thông số kỹ thuật và hiệu suất của GPU là điều quan trọng đối với các nhà nghiên cứu, doanh nghiệp và người tiêu dùng. Tuy nhiên, với sự phát triển nhanh chóng của công nghệ, việc phân tích dữ liệu GPU trở nên phức tạp hơn, đòi hỏi một cách tiếp cận có hệ thống.

2. Phát biểu vấn đề.

Trên thị trường hiện nay, có rất nhiều dòng GPU với các mức giá, hiệu suất và công nghệ khác nhau. Người dùng, từ game thủ, lập trình viên AI, đến các doanh nghiệp, đều

gặp khó khăn trong việc lựa chọn GPU phù hợp với nhu cầu của họ. Một số câu hỏi quan trọng được đặt ra bao gồm:

Các yếu tố nào ảnh hưởng đến giá GPU? GPU có VRAM cao hơn có thực sự đắt hơn không?

Có mối tương quan nào giữa số lượng CUDA Cores, Tensor Cores và giá GPU không?

GPU trung tâm dữ liệu khác gì so với GPU tiêu dùng về mặt thông số kỹ thuật?

Xu hướng giá GPU thay đổi như thế nào theo thời gian phát hành?

Dữ liệu GPU có thể được sử dụng để dự đoán giá sản phẩm trong tương lai không?

Để trả lời các câu hỏi trên, chúng tôi sẽ sử dụng các phương pháp phân tích dữ liệu, từ thống kê mô tả đến mô hình hồi quy, để khám phá những mối quan hệ quan trọng giữa giá cả và thông số kỹ thuật của GPU.

3. Giới thiệu về bộ dữ liệu.

Bộ dữ liệu được sử dụng trong báo cáo này là dữ liệu về các mẫu GPU của Nvidia, bao gồm nhiều dòng sản phẩm khác nhau từ GeForce, Quadro, đến Tesla và A100. Bộ dữ liệu này cung cấp thông tin chi tiết về giá cả, thông số kỹ thuật, hiệu suất và phân khúc thị trường của từng GPU. Dữ liệu được thu thập từ nhiều nguồn, bao gồm tài liệu chính thức từ Nvidia, trang web của các nhà bán lẻ và các nguồn dữ liệu công khai khác.

3.1. Các đặc điểm chính của bộ dữ liệu:

3.1.1 Thông tin cơ bản:

Model GPU (model) – Tên sản phẩm (ví dụ: RTX 4090, A100, H100).

Giá GPU (price) – Giá thị trường hiện tại hoặc giá gốc khi ra mắt.

Phân khúc thị trường (market) – GPU dành cho người tiêu dùng, chuyên nghiệp hay trung tâm dữ liệu.

Ngày phát hành (release_date) – Ngày chính thức ra mắt GPU.

3.1.2 Thông số kỹ thuật:

Số lõi CUDA (cuda_cores) – Ảnh hưởng đến hiệu suất xử lý đồ họa và tính toán song song.

Số lượng Tensor Cores (tensor_cores) – Tăng tốc độ xử lý cho AI và Machine Learning.

Dung lượng bộ nhớ (vram) – Ảnh hưởng đến khả năng xử lý dữ liệu lớn trong đồ họa và AI.

Xung nhịp (base_clock_mhz, max_boost_clock_mhz) – Ảnh hưởng đến tốc độ xử lý của GPU.

Tiêu thụ điện năng (tdp) – Công suất tiêu thụ điện của GPU, ảnh hưởng đến hiệu suất và nhiệt độ.

Băng thông bộ nhớ (bus_width, bandwidth) – Ảnh hưởng đến tốc độ truyền tải dữ liệu giữa GPU và bộ nhớ.

3.1.3 Hiệu suất tính toán:

Khả năng xử lý số học FP16, FP32, FP64 (fp16, fp32, fp64) – Quyết định hiệu suất trong các tác vụ AI, đồ họa và khoa học dữ liệu.

Bộ nhớ cache (L2 Cache (MB)) – Giúp cải thiện hiệu suất xử lý dữ liệu.

Hỗ trợ NVLink (nv_link) – Công nghệ kết nối GPU để mở rộng bộ nhớ và hiệu suất.

CHƯƠNG 2: PHÂN TÍCH DỮ LIỆU

1. Mô tả dữ liệu.

Tập dữ liệu GPU được thu thập từ nhiều dòng sản phẩm khác nhau, bao gồm các GPU dành cho người tiêu dùng (Consumer), trung tâm dữ liệu (Datacenter), và chuyên nghiệp (Professional). Dữ liệu này chứa thông tin về giá cả, thông số kỹ thuật, hiệu suất tính toán và trạng thái sản phẩm.

1.1. Xem trước dữ liệu

Chúng tôi sử dụng lệnh `df.head()` để xem trước 5 dòng đầu tiên của dữ liệu:

Model	Price	Market	Cuda_cores	Tensor_cores	Vram	Base_clock_mhz	Max_boots_clock_mhz	Tdp
RTX 6000 Ada Generation	7198.00	Professional	18176	568	48	915	2505	300
Tesla L40	8627.99	Datacenter	18176	568	48	735	2600	300
GeForce RTX 4090	1948.99	Consumer	18176	512	24	2230	2520	450

1.2. Kiểm tra thông tin tổng quan.

Lệnh `df.info()` giúp chúng tôi kiểm tra số lượng dòng, cột và kiểu dữ liệu của từng cột.

```
df.info()
```

- Tập dữ liệu có 66 dòng và 28 cột.
- Một số cột chứa giá trị rỗng, chẳng hạn như L2 Cache (MB), Notes, profile, slot.
- Các biến quan trọng như price, vram, cuda_cores, tdp đều ở định dạng số (float64 hoặc int64).

2. Thống kê mô tả dữ liệu số

Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

Lệnh `df.describe()` cung cấp các thống kê mô tả về giá trị lớn nhất, nhỏ nhất, trung bình và độ lệch chuẩn của dữ liệu số.

Thống kê	price	tensor_cores	vram	tdp
Số mẫu	66	66	66	66
Trung bình	2144.62	234.54	20.18	221.85
Độ lệch chuẩn	5468.57	183.21	16.33	87.98
Giá trị nhỏ nhất	18.95	0	4	60
Giá trị lớn nhất	40745	640	80	450

Nhận xét:

- Giá GPU có sự chênh lệch rất lớn: Một số GPU có giá lên đến 40,745 USD, trong khi có những GPU chỉ khoảng 18 USD.
- VRAM dao động từ 4GB đến 80GB, tùy thuộc vào phân khúc GPU.
- TDP (mức tiêu thụ điện năng) trải dài từ 60W đến 450W, phản ánh sự khác biệt lớn giữa GPU tiết kiệm năng lượng và GPU hiệu năng cao.

3. Dữ liệu thiếu và ngoại lệ.

Một số cột trong tập dữ liệu có chứa giá trị rỗng (NaN), đặc biệt là:

- L2 Cache (MB): Chỉ có 27 giá trị hợp lệ.
- Notes, profile, slot: Có nhiều giá trị trống.

Tóm lại, tập dữ liệu GPU chứa đầy đủ thông tin về giá cả, thông số kỹ thuật và hiệu suất, nhưng cũng có một số dữ liệu bị thiếu cần được xử lý trước khi tiến hành phân tích chi tiết.

4. Các câu hỏi nghiên cứu.

Dựa trên tập dữ liệu GPU, chúng tôi đặt ra các câu hỏi nghiên cứu quan trọng nhằm hiểu rõ hơn về mối quan hệ giữa giá GPU và các thông số kỹ thuật:

5. Các câu hỏi nghiên cứu.

Dựa trên tập dữ liệu GPU, chúng tôi đặt ra các câu hỏi nghiên cứu quan trọng nhằm hiểu rõ hơn về mối quan hệ giữa giá GPU và các thông số kỹ thuật:

- 1 Vẽ biểu đồ histogram để hiểu sự phân bố giá của GPU trong tập dữ liệu.
- 2 Dùng histogram để xem mức tiêu thụ điện năng của các GPU phổ biến.
- 3 Kiểm tra xem GPU có nhiều CUDA Cores hơn có giá cao hơn không, vẽ biểu đồ tán xạ (scatter plot).
- 4 Kiểm tra xem dung lượng bộ nhớ có ảnh hưởng đến giá GPU không, vẽ biểu đồ tán xạ (scatter plot).
- 5 So sánh giá GPU giữa các phân khúc thị trường (market), vẽ biểu đồ hộp.
- 6 So sánh mức tiêu thụ điện năng (tdp) giữa các dòng GPU (architecture), vẽ biểu đồ hộp.
- 7 Phân phối số lượng Tensor Cores (tensor_cores) giữa các kiến trúc GPU (architecture), vẽ biểu đồ histogram nhóm.
- 8 Phân phối xung nhịp tối đa (max_boost_clock_mhz) giữa các phân khúc thị trường (market), vẽ biểu đồ histogram nhóm.
- 9 Xu hướng giá GPU theo ngày phát hành (release_date), vẽ biểu đồ đường.
- 10 Xu hướng TDP (tdp) của GPU qua các thế hệ (architecture), vẽ biểu đồ đường.
- 11 Số lượng GPU theo từng kiến trúc (architecture), vẽ biểu đồ cột.
- 12 Số lượng GPU theo phân khúc thị trường (market), vẽ biểu đồ cột.
- 13 Tính toán ma trận tương quan giữa các thông số GPU (vram, tdp, tensor_cores, cuda_cores, price) và vẽ heatmap.
- 14 Tỷ lệ GPU theo phân khúc thị trường (market), vẽ biểu đồ tròn.
- 15 Tỷ lệ GPU hỗ trợ NVLink (nv_link) và không hỗ trợ, Biểu đồ pie chart
- 16 Hồi quy tuyến tính dự đoán giá GPU.

6. Phân tích bộ dữ liệu và trả lời câu hỏi

```
# thư viện
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import pearsonr
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm

# Đọc dữ liệu từ file CSV
df = pd.read_csv("https://raw.githubusercontent.com/VoPhuocToan/DAMH_VPT/refs/heads/main/gpu_cleaned.csv")
```

```
[ ] # Hiển thị dữ liệu
print(df.head())
```

```

model      price  market  ecc_equipped \
0 RTX 6000 Ada Generation 7198.00 Professional Yes
1 Tesla L40 8627.99 Datacenter Yes
2 GeForce RTX 4090 1948.99 Consumer Partial Support
3 RTX 5000 Ada Generation 4499.99 Professional Yes
4 Tesla H100 40745.00 Datacenter Yes

architecture release_date tensor_cores tensor_core_gen cuda_cores \
0 Ada Lovelace 12/3/2022 568 4th Gen (Ada) 18,176
1 Ada Lovelace 8322 568 4th Gen (Ada) 18,176
2 TSMC 4N 4679 512 4th Gen (Ada) 18,176
3 Ada Lovelace 8/9/2023 400 4th Gen (Ada) 12,800
4 Hopper 3/22/2022 640 4th Gen (Hopper) 14,592

base_clock_mhz ... fp64 tdp nv_link bus_interface \
0 915 ... 1,423.00 300 No PCIe 4.0 x16
1 735 ... 1,414.00 300 No PCIe 4.0 x16
2 2,230 ... 1,142.00 450 No PCIe 4.0 x16
3 1,155 ... 1,020.00 250 No PCIe 4.0 x16
4 1,065 ... 25,600.00 350 2-way NVLink PCIe 5.0 x16

L2 Cache (MB) Notes profile slot eol eol_date
0 NaN 4x DisplayPort FHFL Double No No
1 NaN PCIe card (dual-slot) NaN NaN No No
2 72.0 NaN NaN NaN No No
3 NaN 4x DisplayPort FHFL Double No No
4 NaN PCIe card (dual-slot) NaN NaN No No

[5 rows x 28 columns]
```

Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
[ ] # In ra kích thước của tập dữ liệu (số dòng, số cột).
```

```
print("Kích thước của tập dữ liệu:", df.shape) # (số dòng, số cột)
```

↗ Kích thước của tập dữ liệu: (66, 28)

```
▶ # In ra danh sách các cột có trong tập dữ liệu.
```

```
print("Danh sách các cột:")
```

```
for col in df.columns:
```

```
    print(col)
```

↗ Danh sách các cột:

model

price

market

ecc_equipped

architecture

release_date

tensor_cores

tensor_core_gen

cuda_cores

base_clock_mhz

max_boost_clock_mhz

bus_type

bus_width

vram

clock_speed

bandwidth

fp16

fp32

fp64

tdp

nv_link

bus_interface

L2 Cache (MB)

Notes

profile

slot

eol

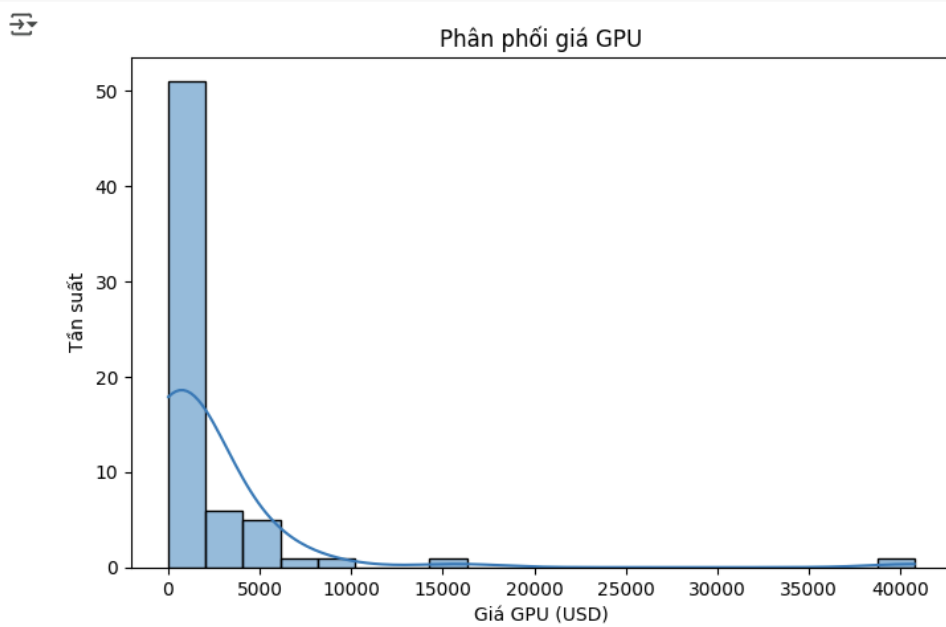
eol_date

Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
# Tính toán giá GPU trung bình theo từng phân khúc thị trường (market).
print("Giá GPU trung bình theo phân khúc thị trường:")
print(df.groupby("market")["price"].mean())
```

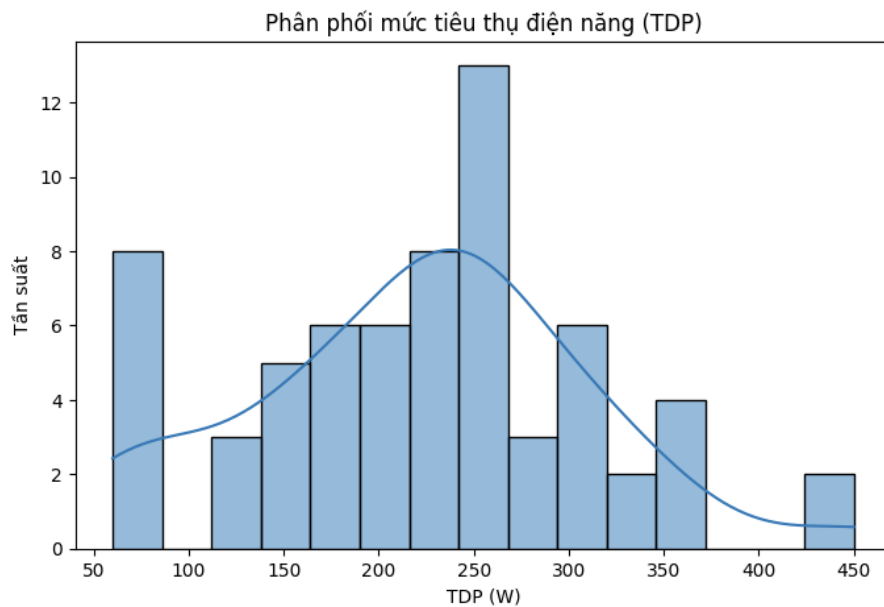
```
Giá GPU trung bình theo phân khúc thị trường:
market
Consumer      478.032593
Datacenter    5368.636111
Legacy         118.627778
Professional  2577.911667
Name: price, dtype: float64
```

```
[ ] #1 Vẽ biểu đồ histogram để hiểu sự phân bố giá của GPU trong tập dữ liệu.
plt.figure(figsize=(8, 5))
sns.histplot(df['price'].dropna(), bins=20, kde=True)
plt.xlabel("Giá GPU (USD)")
plt.ylabel("Tần suất")
plt.title("Phân phối giá GPU")
plt.show()
```



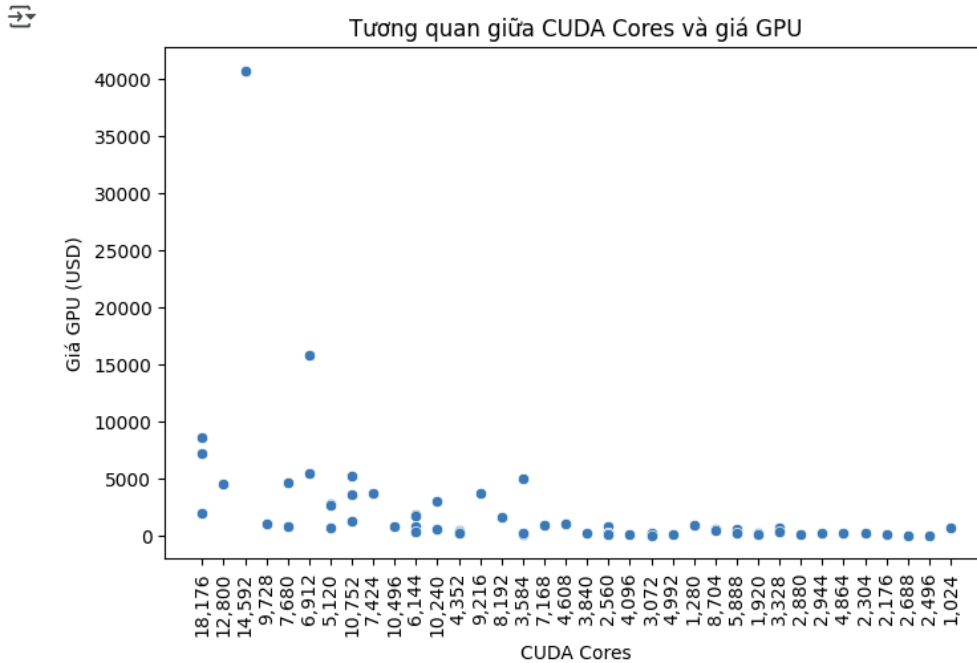
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#2 Dùng histogram để xem mức tiêu thụ điện năng của các GPU phổ biến.  
plt.figure(figsize=(8, 5))  
sns.histplot(df['tdp'].dropna(), bins=15, kde=True)  
plt.xlabel("TDP (W)")  
plt.ylabel("Tần suất")  
plt.title("Phân phối mức tiêu thụ điện năng (TDP)")  
plt.show()
```



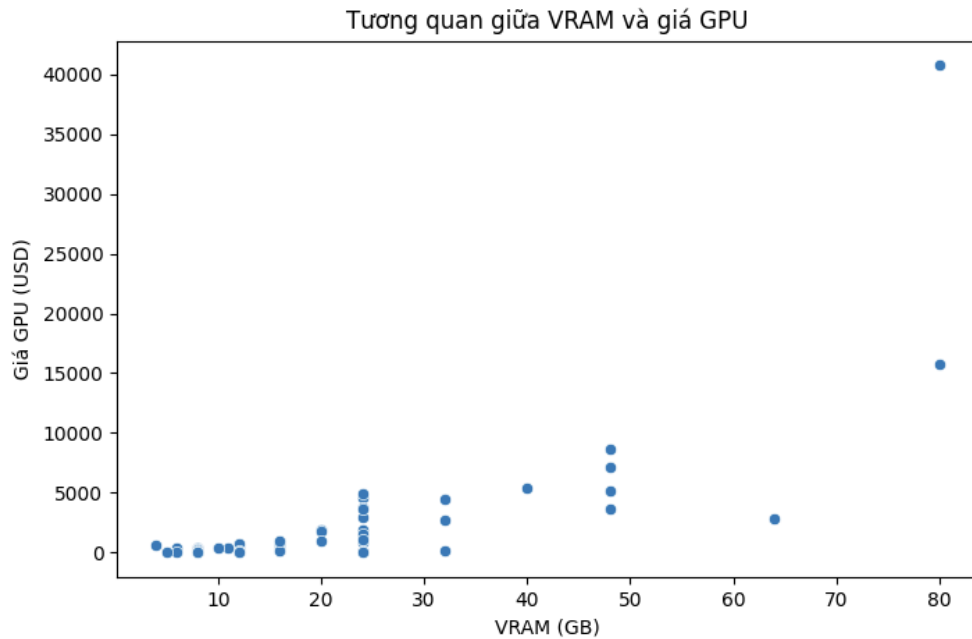
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#3 Kiểm tra xem GPU có nhiều CUDA Cores hơn có giá cao hơn không.  
plt.figure(figsize=(8, 5))  
sns.scatterplot(x=df['cuda_cores'], y=df['price'])  
plt.xlabel("CUDA Cores")  
plt.ylabel("Giá GPU (USD)")  
plt.title("Tương quan giữa CUDA Cores và giá GPU")  
plt.xticks(rotation=90)  
plt.show()
```



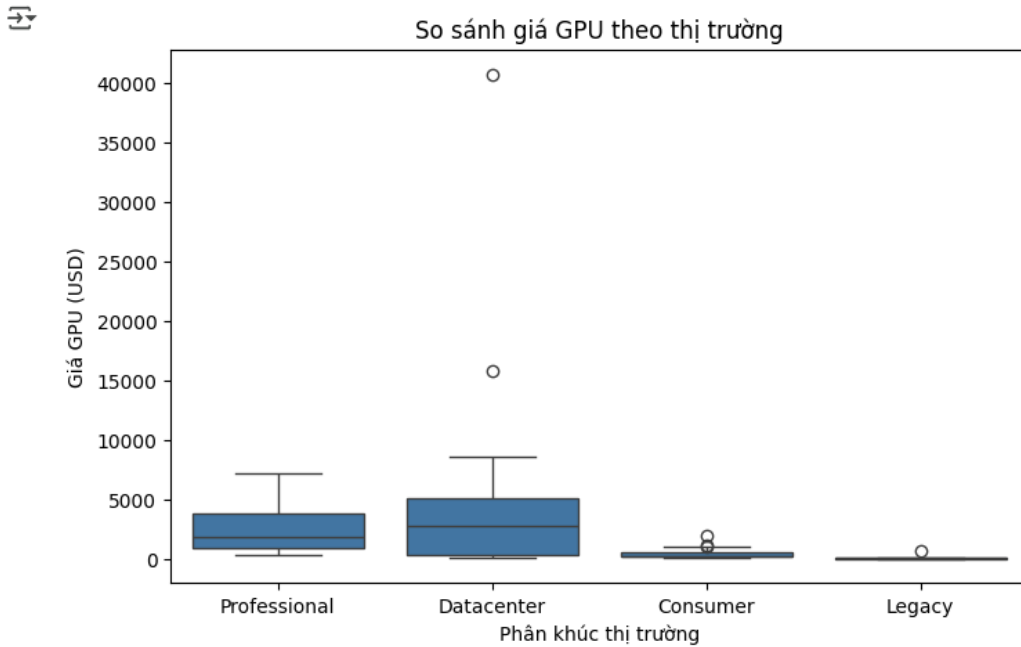
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#4 Kiểm tra xem dung lượng bộ nhớ có ảnh hưởng đến giá GPU không.  
plt.figure(figsize=(8, 5))  
sns.scatterplot(x=df['vram'], y=df['price'])  
plt.xlabel("VRAM (GB)")  
plt.ylabel("Giá GPU (USD)")  
plt.title("Tương quan giữa VRAM và giá GPU")  
plt.show()
```



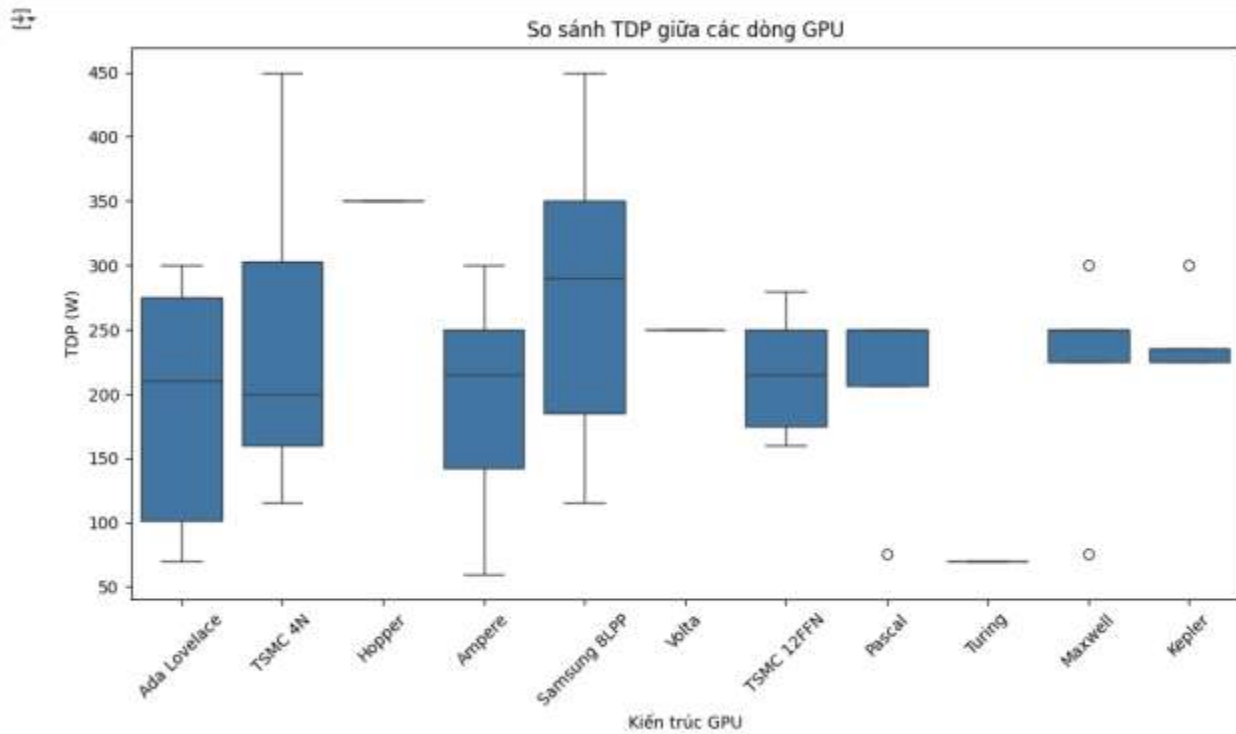
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#5 So sánh giá GPU giữa các phân khúc thị trường (market)
plt.figure(figsize=(8, 5))
sns.boxplot(x=df['market'], y=df['price'])
plt.xlabel("Phân khúc thị trường")
plt.ylabel("Giá GPU (USD)")
plt.title("So sánh giá GPU theo thị trường")
plt.show()
```



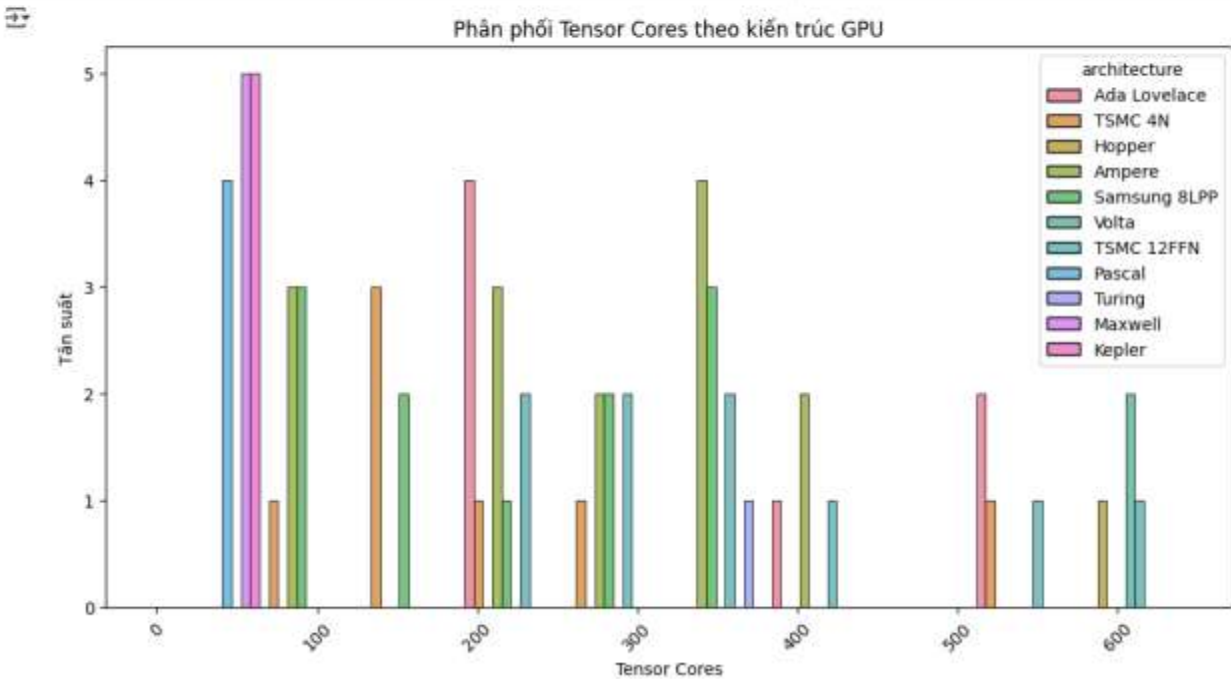
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#6 So sánh mức tiêu thụ điện năng (tđp) giữa các dòng GPU (architecture)
plt.figure(figsize=(12, 6))
sns.boxplot(x=df['architecture'], y=df['tdp'])
plt.xticks(rotation=45)
plt.xlabel("Kiến trúc GPU")
plt.ylabel("TDP (W)")
plt.title("So sánh TDP giữa các dòng GPU")
plt.show()
```



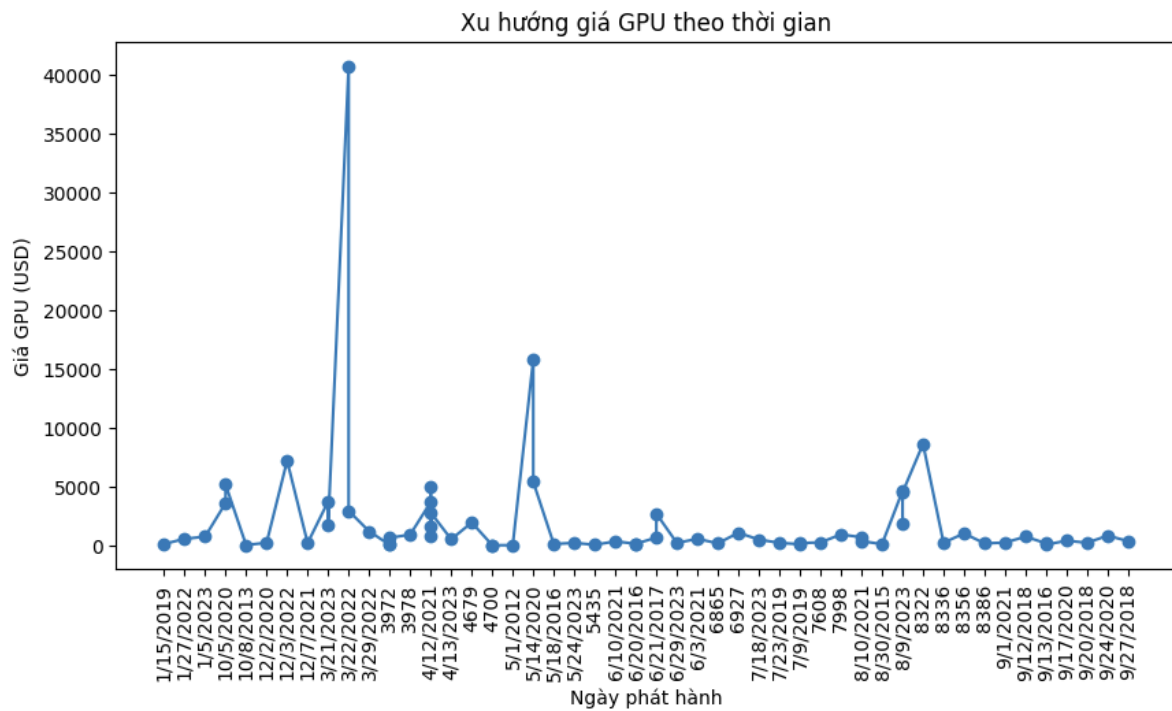
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#7 Phân phối số lượng Tensor Cores (tensor_cores) giữa các kiến trúc GPU (architecture), vẽ biểu đồ histogram nhóm.  
plt.figure(figsize=(12, 6))  
sns.histplot(x=df['tensor_cores'], hue=df['architecture'], multiple="dodge", bins=10)  
plt.xticks(rotation=45)  
plt.xlabel("Tensor Cores")  
plt.ylabel("Tần suất") # Thay đổi nhãn trục y thành "Tần suất"  
plt.title("Phân phối Tensor Cores theo kiến trúc GPU")  
plt.show()
```



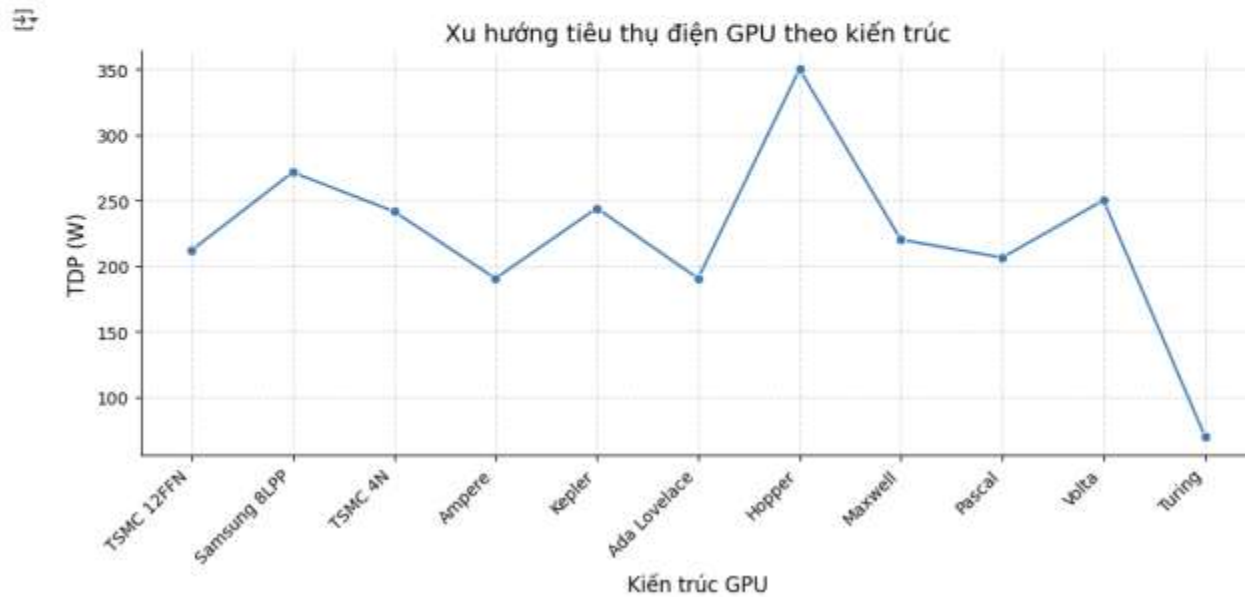
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#9 Xu hướng giá GPU theo ngày phát hành (release_date).
df_sorted = df.sort_values(by="release_date")
plt.figure(figsize=(10, 5))
plt.plot(df_sorted['release_date'], df_sorted['price'], marker='o')
plt.xticks(rotation=90)
plt.xlabel("Ngày phát hành")
plt.ylabel("Giá GPU (USD)")
plt.title("Xu hướng giá GPU theo thời gian")
plt.show()
```

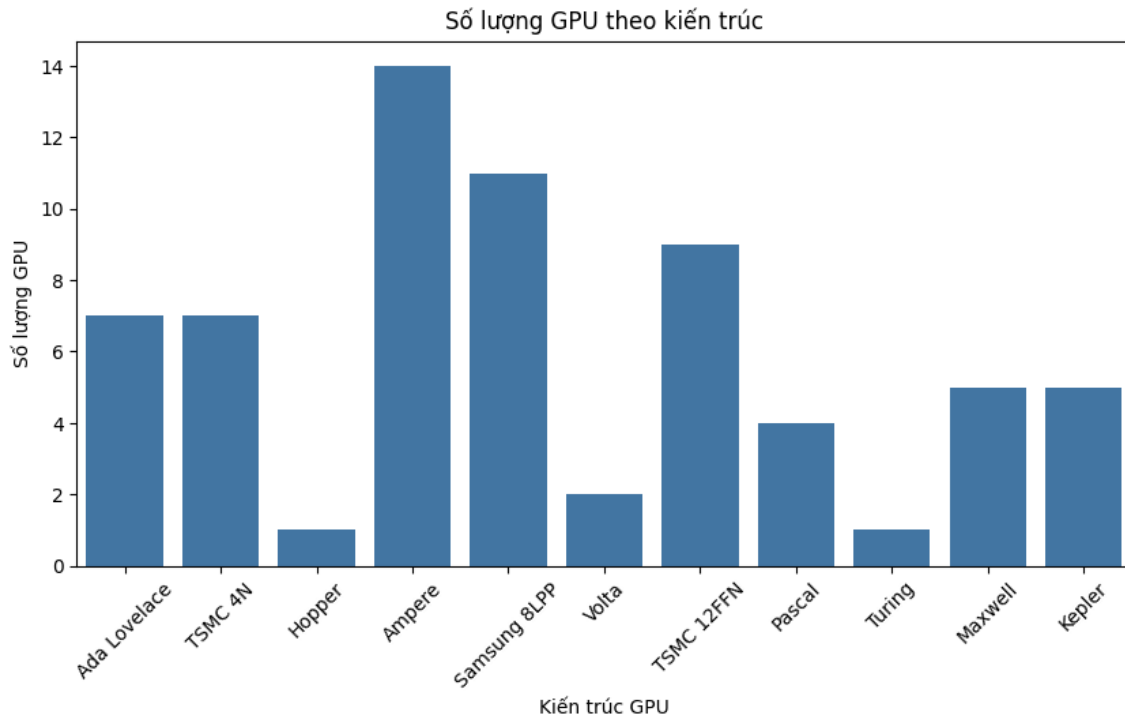


Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#10 Xu hướng TDP (tđp) của GPU qua các thế hệ (architecture).  
plt.figure(figsize=(10, 5))  
sns.lineplot(x=df_sorted['architecture'], y=df_sorted['tdp'], marker='o', errorbar=None)  
plt.xticks(rotation=45, ha='right')  
plt.xlabel("Kiến trúc GPU", fontsize=12)  
plt.ylabel("TDP (W)", fontsize=12)  
plt.title("Xu hướng tiêu thụ điện GPU theo kiến trúc", fontsize=14)  
plt.grid(True, linestyle='--', alpha=0.5)  
sns.despine()  
plt.tight_layout()  
plt.show()
```

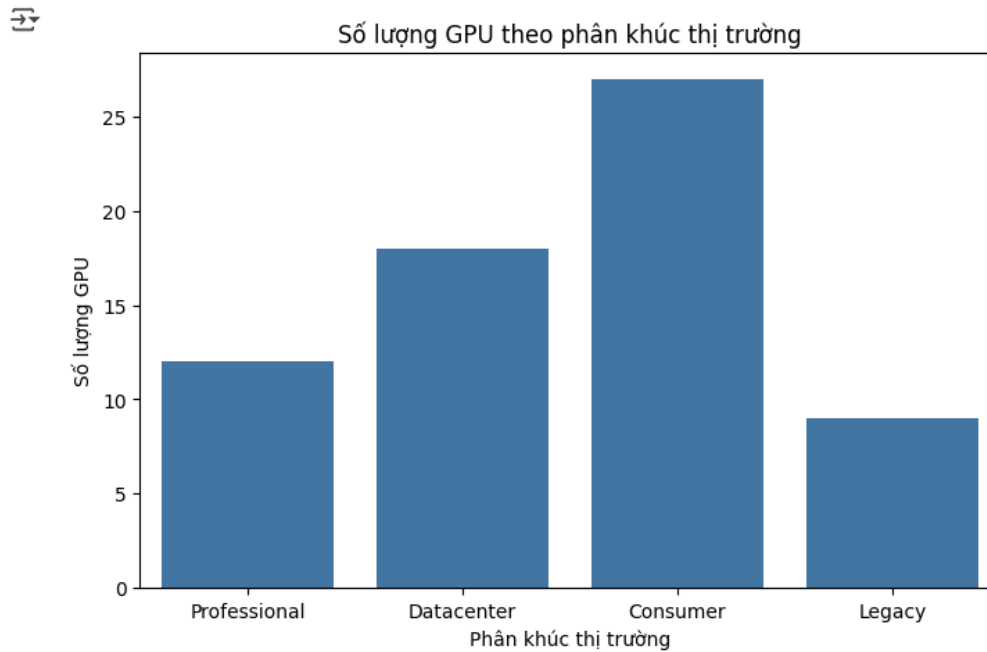


```
#11 Số lượng GPU theo từng kiến trúc (architecture)
plt.figure(figsize=(10, 5))
sns.countplot(x=df['architecture'])
plt.xticks(rotation=45)
plt.xlabel("Kiến trúc GPU")
plt.ylabel("Số lượng GPU")
plt.title("Số lượng GPU theo kiến trúc")
plt.show()
```



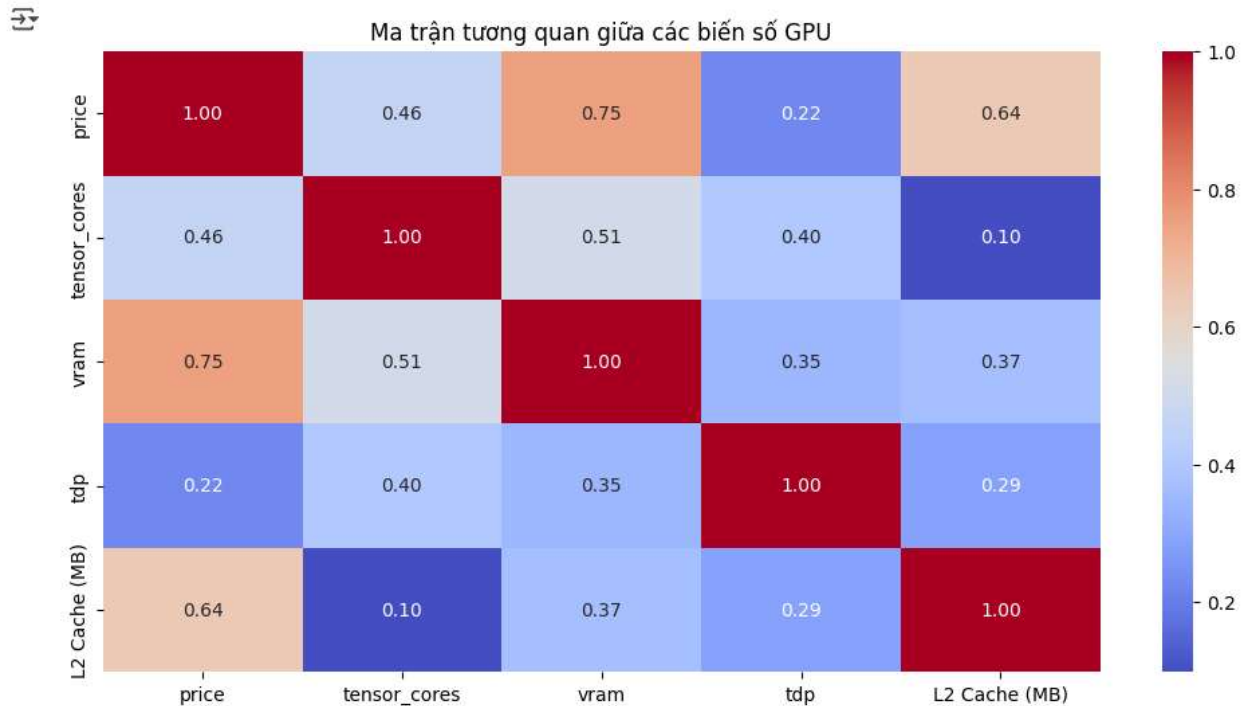
Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

```
#12 Số lượng GPU theo phân khúc thị trường (market), vẽ biểu đồ cột.  
plt.figure(figsize=(8, 5))  
sns.countplot(x=df['market'])  
plt.xlabel("Phân khúc thị trường")  
plt.ylabel("Số lượng GPU")  
plt.title("Số lượng GPU theo phân khúc thị trường")  
plt.show()
```



Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

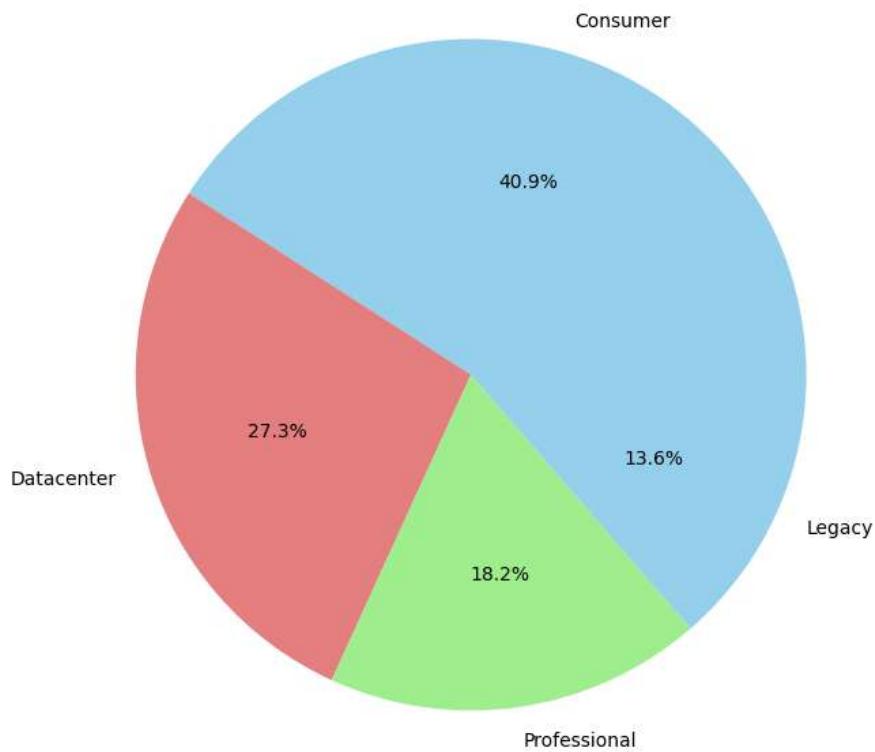
```
#13 Tính toán ma trận tương quan giữa các thông số GPU (vram, tdp, tensor_cores, cuda_cores, price) và vẽ heatmap.  
plt.figure(figsize=(12, 6))  
numerical_df = df.select_dtypes(include=np.number)  
sns.heatmap(numerical_df.corr(), cmap="coolwarm", annot=True, fmt=".2f")  
plt.title("Ma trận tương quan giữa các biến số GPU")  
plt.show()
```



```
#14 Tỷ lệ GPU theo phân khúc thị trường (market)
plt.figure(figsize=(8, 8))
df['market'].value_counts().plot.pie(autopct='%1.1f%%', colors=['skyblue', 'lightcoral', 'lightgreen'])
plt.ylabel("")
plt.title("Tỷ lệ GPU theo phân khúc thị trường")
plt.show()
```



Tỷ lệ GPU theo phân khúc thị trường

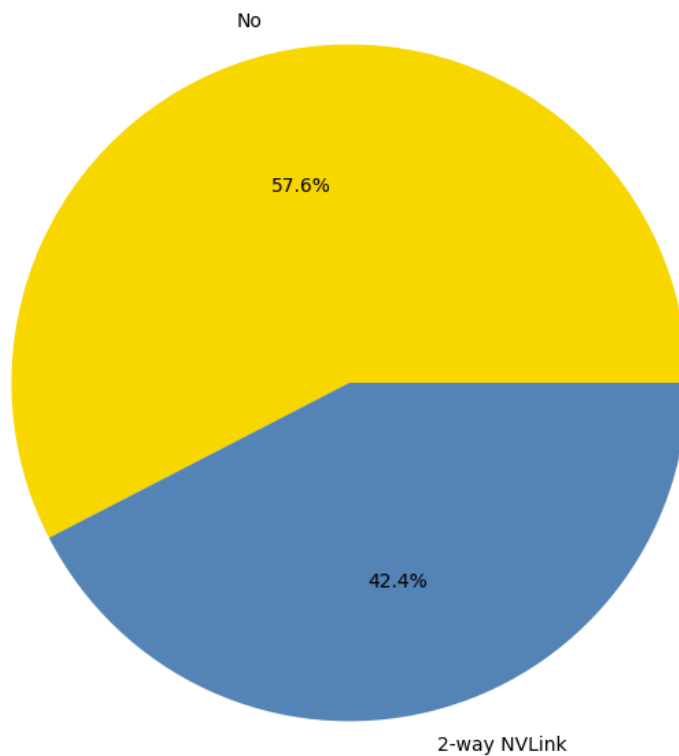


```
#15 Tỷ lệ GPU hỗ trợ NVLink (nv_link) và không hỗ trợ

plt.figure(figsize=(8, 8))
df['nv_link'].value_counts().plot.pie(autopct='%1.1f%%', colors=['gold', 'steelblue'])
plt.ylabel("")
plt.title("Tỷ lệ GPU có NVLink và không có NVLink")
plt.show()
```



Tỷ lệ GPU có NVLink và không có NVLink



```
#16 Hồi quy tuyến tính dự đoán giá GPU.
features = ['vram', 'tensor_cores', 'tdp']
df_clean = df.dropna(subset=features + ['price'])
X = df_clean[features]
y = df_clean['price']

model = LinearRegression()
model.fit(X, y)
print("Hệ số hồi quy:", model.coef_)
print("Intercept:", model.intercept_)
print("R² Score:", model.score(X, y))
```



```
Hệ số hồi quy: [239.10721654  3.9077146  -4.84315716]
Intercept: -2523.0879901180583
R² Score: 0.5812322901661373
```

CHƯƠNG 3: KẾT LUẬN

Từ những câu trả lời trên ta có thể đưa ra 1 số kết luận như sau:

1. Xu hướng giá và hiệu năng

Các GPU dòng Datacenter (Tesla) dẫn đầu về giá, với Tesla H100 đạt mức \$40,000

Có mối tương quan rõ rệt giữa giá và hiệu năng FP32/FP16

GPU thế hệ mới thường có hiệu năng cao hơn và giá thành tương ứng

2. Phân khúc thị trường

Consumer: Giá thành thấp nhất, phù hợp người dùng phổ thông (dòng GeForce RTX)

Professional: Tầm giá trung bình, hướng đến người dùng chuyên nghiệp

Datacenter: Phân khúc cao cấp nhất, chuyên dụng cho trung tâm dữ liệu

Legacy: Các model cũ với giá và hiệu năng thấp

3. Tiến hóa công nghệ

Thế hệ GPU mới (Ada Lovelace) có số lượng CUDA cores và Tensor cores vượt trội

Bảng thông bộ nhớ và tốc độ xử lý tăng qua các thế hệ

Dung lượng VRAM tăng dần, từ 4GB lên đến 80GB

4. Mối quan hệ giữa các thông số

TDP tăng tương ứng với hiệu năng

Số lượng CUDA cores và Tensor cores ảnh hưởng trực tiếp đến hiệu năng

Cards có VRAM lớn thường đi kèm mức giá cao

5. Đặc điểm từng phân khúc

Gaming (Consumer): Cân bằng giữa hiệu năng và giá thành

Phân tích bộ dữ liệu dòng sản phẩm hoàn chỉnh của Nvidia

Professional: Tối ưu về độ ổn định và độ chính xác

Datacenter: Tối ưu cho AI/ML với VRAM lớn và nhiều Tensor cores

Kết luận: NVIDIA có chiến lược phân khúc sản phẩm rõ ràng, với các dòng GPU được thiết kế đặc biệt cho từng đối tượng người dùng khác nhau.

Danh sách thành viên nhóm

MSSV	Họ tên	Lớp	Công việc
110122163	Châu Thanh Thiện	DA22TTC	Chuẩn bị dữ liệu; Chuẩn bị báo cáo
110122186	Võ Phước Toàn	DA22TTC	Chuẩn bị dữ liệu; Chuẩn bị báo cáo
110122212	Nguyễn Hoàng Yến	DA22TTC	Chuẩn bị dữ liệu; Chuẩn bị báo cáo

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] Python Software Foundation, “Python,” python.org. Available: https://www.python.org/?utm_source=chatgpt.com. [Accessed: Feb. 10, 2025].
- [2] V. P. Toan, "GPU Cleaned Dataset," GitHub, Available: https://raw.githubusercontent.com/VoPhuocToan/DAMH_VPT/refs/heads/main/gpu_cleaned.csv. [Accessed: Feb. 10, 2025].
- [3] OpenAI, "ChatGPT," OpenAI, Available: <https://openai.com/chatgpt>. [Accessed: Feb. 10, 2025].
- [4] W3Schools, “Python Tutorial,” W3Schools.com. Available: https://www.w3schools.com/python/?utm_source=chatgpt.com. [Accessed: Feb. 10, 2025].
- [5] A.Nouar, "Nvidia Complete Lineup Dataset," Kaggle, Available: <https://www.kaggle.com/datasets/alexandrenouarcto/nvidia-complete-lineup-dataset>. [Accessed: Feb. 10, 2025].
- [6] Google, “Google Colaboratory Notebook,” Available: <https://colab.research.google.com/drive/193JTN8JbPPoXpOWsmPJqzGjenI27aM6u?pli=1&authuser=2>. [Accessed: Feb. 10, 2025].