

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**  
**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Đề tài: Phân tích sản phẩm thời trang trên sàn thương mại TIKI**

**Môn học: Trực quan hóa dữ liệu**

**Lớp: 21\_21**

**Giảng viên hướng dẫn:**

**TS. Bùi Tiến Lên**

**Người thực hiện:**

**Võ Minh Khuê – 21120486**

**Nguyễn Ngọc Gia Minh - 21120501**

**Nguyễn Gia Phúc - 21120529**

**Nguyễn Đình Nhật Trí - 21120576**

**Nguyễn Thủy Uyên - 21120590**

## Mục lục

I.	TỔNG QUAN VỀ NHÓM:.....	3
II.	GIỚI THIỆU BỘ DỮ LIỆU:.....	3
III.	DASHBOARD: .....	4
IV.	ĐÁNH GIÁ: .....	9
1.	Kết hợp nguồn dữ liệu đáng tin cậy: .....	9
2.	Phù hợp với mục đích:.....	10
3.	Rõ ràng và dễ hiểu:.....	10
4.	4 & 5. Sự liên kết, tương tác và điều hướng: .....	11
5.	Thiết kế hấp dẫn: .....	12
6.	Mối quan hệ và chiến lược cho các mặt hàng: .....	12
7.	Tích hợp AI: .....	14

## I. TỔNG QUAN VỀ NHÓM:

MSSV	Họ tên	Phân công
21120529	Nguyễn Gia Phúc	Tích hợp AI, Dashboard
21120486	Võ Minh Khuê	Viết báo cáo, thiết kế slides
21120501	Nguyễn Ngọc Gia Minh	Dashboard, viết báo cáo
21120576	Trần Đình Nhật Trí	Dashboard, viết báo cáo
21120590	Nguyễn Thủy Uyên	Dashboard, viết báo cáo, slides

## II. GIỚI THIỆU BỘ DỮ LIỆU:

Bộ dữ liệu được thu thập từ nền tảng thương mại điện tử TIKI, một trong những trang mua sắm trực tuyến hàng đầu tại Việt Nam.

Do số lượng lớn các mặt hàng trên TIKI và vì mục đích giáo dục, nên chỉ thu thập các mặt hàng phổ biến (dựa trên xếp hạng sản phẩm của TIKI). Điều này không chỉ giảm kích thước dữ liệu mà còn loại bỏ các ID sản phẩm 'rỗng' không mong muốn.

Giấy phép: Được crawl từ public API từ TIKI. Sử dụng cho mục đích giáo dục

Link dataset: <https://www.kaggle.com/datasets/michaelminhpham/vietnamese-tiki-e-commerce-dataset>

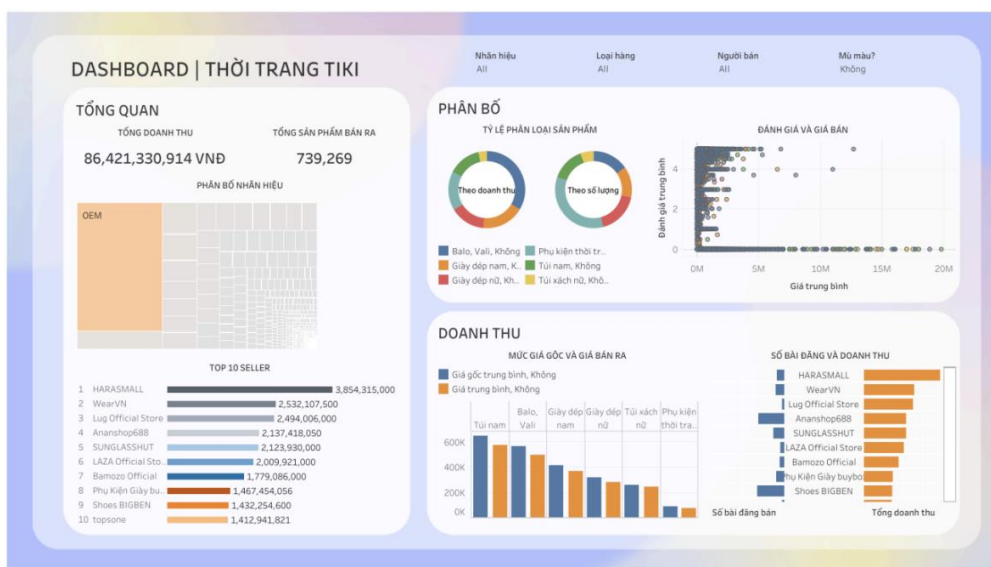
Ý nghĩa các thuộc tính:

- **name:** Tiêu đề sản phẩm.
- **description:** Mô tả sản phẩm.
- **original\_price:** Giá gốc trước khi có giảm giá.
- **price:** Giá hiện tại sau khi có thể áp dụng giảm giá.
- **fulfillment\_type:** Loại phương thức giao hàng.
- **brand:** Nhà sản xuất hoặc thương hiệu đăng ký.
- **review\_counts:** Số lượng đánh giá.
- **rating\_average:** Điểm đánh giá trung bình của sản phẩm.
- **pay\_later:** Cho phép chương trình mua hàng ngay và trả sau.
- **current\_seller:** Tên người bán hiện tại.
- **date\_created:** số ngày từ ngày cập nhật cuối cùng và ngày tạo
- **number\_of\_images:** số lượng hình ảnh sản phẩm đã tải lên
- **vnd\_cashback:** số tiền hoàn lại theo mệnh giá VND

- **has\_video**: Đúng nếu sản phẩm có video clip, nếu không thì Sai
- **quantity\_sold**: tổng số lượng đã bán trong lịch sử

### III. DASHBOARD:

Tổng quan:



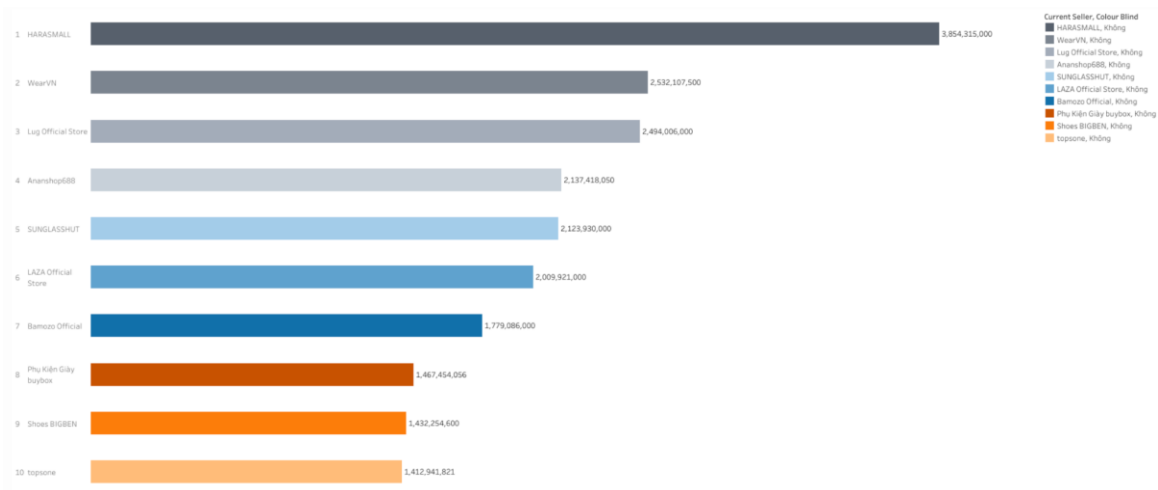
#### A. Phân tích biểu đồ:

##### 1. Phân bố nhãn hiệu:

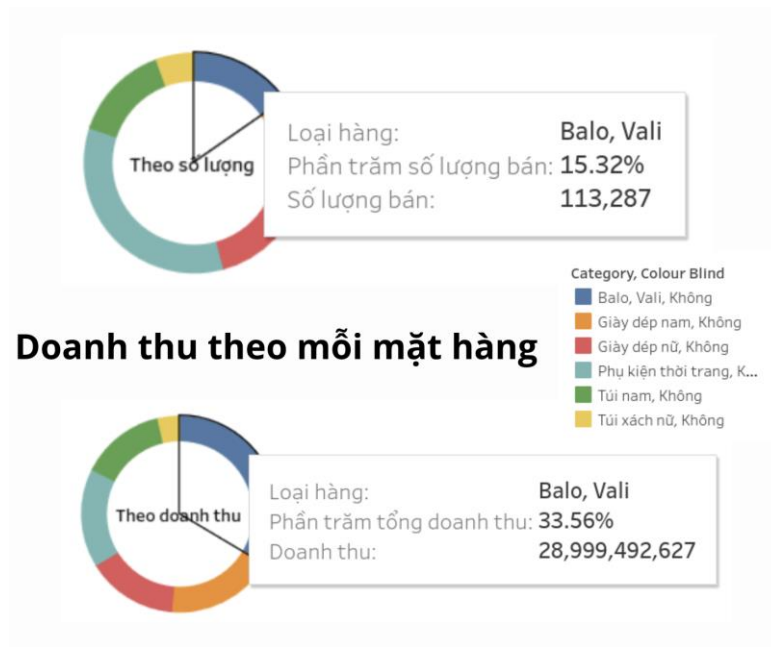


- **Nhận xét:**

- OEM có thị phần lớn nhất trên thị trường về mặt doanh số, cho thấy các sản phẩm OEM được bán nhiều nhất trong số tất cả các thương hiệu được đại diện.
  - Có nhiều thương hiệu nhỏ hơn với mức độ thị phần khác nhau. Nhiều thương hiệu khác chiếm phần nhỏ hơn trên thị trường, cho thấy bối cảnh cạnh tranh cao.
- Lí do chọn Tree map:
    - Dữ liệu có tính chất phân cấp.
    - Dễ so sánh doanh thu giữa các nhãn hiệu.
    - Thiết kế đẹp mắt.
2. Top 10 nhà bán lẻ:



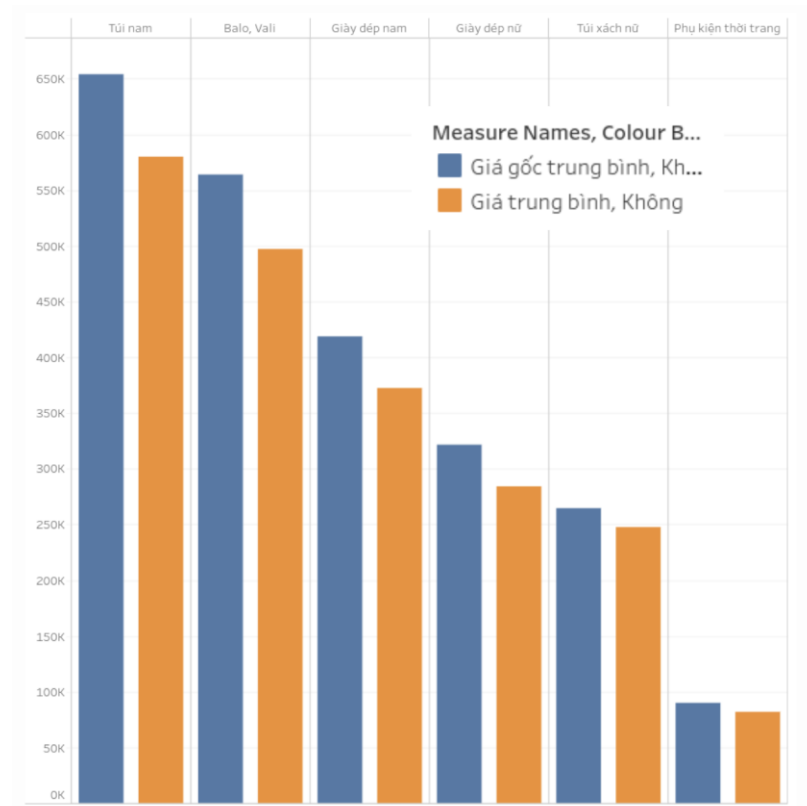
- Nhận xét:
    - HARAS.MALL dẫn đầu về doanh thu, cho thấy sự hiện diện mạnh mẽ trên thị trường
    - Những nhà bán lẻ tuy lọt vào top 10 nhưng lại có doanh thu thấp hơn nhiều so với những người bán hàng đầu, nhưng cho thấy tiềm năng tăng trưởng.
  - Lí do chọn Bar chart:
    - Dễ dàng quan sát và so sánh doanh thu giữa các nhà bán lẻ.
3. Tổng sản phẩm bán được theo loại mặt hàng



- Nhận xét:
  - Phân loại sản phẩm theo doanh thu và số sản phẩm bán ra có sự chênh lệch.
  - Loại hàng Balo, Vali chiếm tỷ trọng cao trong doanh thu, về sản phẩm bán ra thì Phụ kiện thời trang chiếm ưu thế.
- Lí do chọn Pie chart:
  - So sánh tỷ lệ sản phẩm, doanh thu giữa các loại hàng.

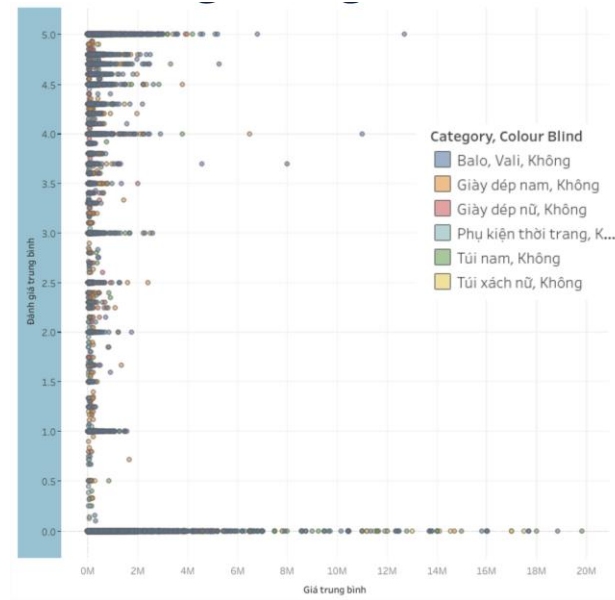
-> Proximity

#### 4. Giá gốc và giá bán ra:

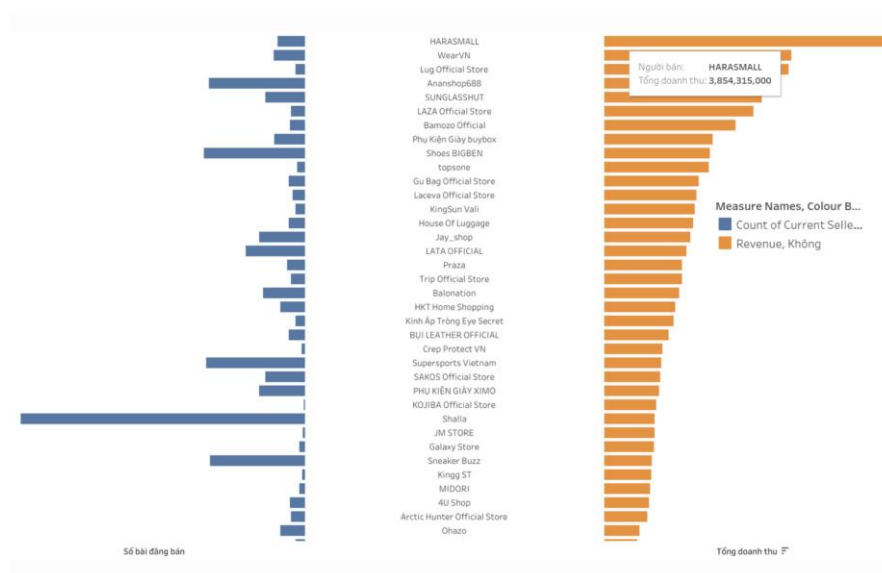


- Nhận xét:
  - Trên tất cả các danh mục, giá bán trung bình luôn thấp hơn giá ban đầu trung bình, cho thấy mức chiết khấu hoặc giảm giá. Túi xách, Balo/ Vali nam có mức giảm giá cao nhất.
  - Các danh mục có giá cao hơn như Túi xách nam và Ba lô/Vali vẫn giảm giá đáng kể, có thể cho thấy chiến lược xả hàng hoặc thu hút người mua.
  - Mức giảm giá đáng kể, đặc biệt là đối với các mặt hàng có giá cao hơn, đây có thể là một chiến thuật để tăng doanh số bán hàng hoặc thâm nhập thị trường.
- Lí do chọn Bar chart:
  - So sánh mức độ chênh lệch giữa giá gốc và giá bán ra thực tế.

## 5. Đánh giá và giá bán:



- Nhận xét:
    - Nhìn chung ở mức giá nào cũng có đánh giá từ 1-5 sao, tuy nhiên sản phẩm càng có giá trị thì đánh giá cao xuất hiện nhiều.
    - Có một số sản phẩm có xếp hạng 0, có thể là sản phẩm chưa được xếp hạng.
    - Balo, Vali và Giày dép nam là các phân loại được ưa chuộng.
  - Lí do chọn Scatter chart:
    - Giúp quan sát mối quan hệ và tìm ra những pattern.
6. Doanh thu và số bài đăng





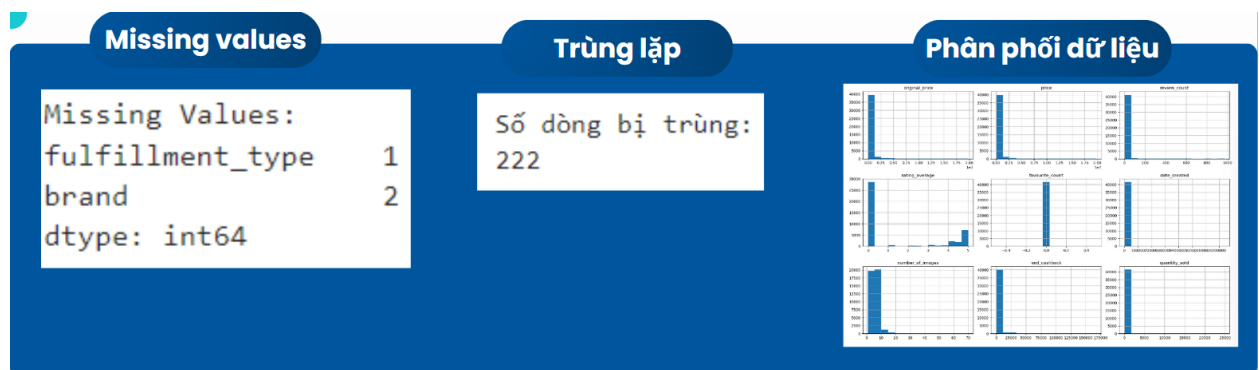
- Nhận xét:
  - Shalla nổi bật là người dẫn đầu về doanh thu dù không có số lượng bài đăng cao nhất.
  - Có sự chênh lệch giữa số lượng bài đăng và doanh thu được tạo ra, cho thấy rằng các yếu tố khác ngoài số lượng bài đăng, chẳng hạn như chất lượng sản phẩm, danh tiếng thương hiệu hoặc hiệu quả tiếp thị, có thể đóng một vai trò quan trọng trong việc tạo doanh thu.
- Lí do chọn Butterfly chart:
  - So sánh hiệu quả hoạt động của các nhà bán lẻ.
  - Thiết kế đẹp mắt.

#### IV. ĐÁNH GIÁ:

##### 1. Kết hợp nguồn dữ liệu đáng tin cậy:

Bộ dữ liệu được crawl từ public API được cung cấp từ TIKI.

Dữ liệu trước tiên xử lý:



Sau khi tiền xử lý:

```
print(df_cleaned.head(5))
```

	name	original_price	price
0	tàng đeo cài áo	10000	10000
1	Bộ Móc May Áo Ngực Loại 2 Móc Cài	7000	7000
2	Miếng Nối Rộng Tăng Size Áo Ngực 4 Móc Cài-Móc...	8000	8000
3	GOTI - Ghim Cài Áo Hình Chiếu Lá Trắng Đinh Ng...	15000	13500
4	Vi nút bấm áo dài 36 nút hiệu POP	13600	13600

	fulfillment_type	brand	review_count	rating_average	favourite_count
0	dropship	OEM	2	5.0	0
1	dropship	OEM	1	4.0	0
2	dropship	OEM	1	5.0	0
3	dropship	GOTI	22	4.8	0
4	dropship	OEM	8	5.0	0

	pay_later	current_seller	date_created
0	False	Siêu thị vàng mã Thiên Hương Phúc	539
1	False	Comliti Shop	602
2	False	Comliti Shop	603
3	False	GOTI Official Store	729
4	False	Shop Hảy Cát	636

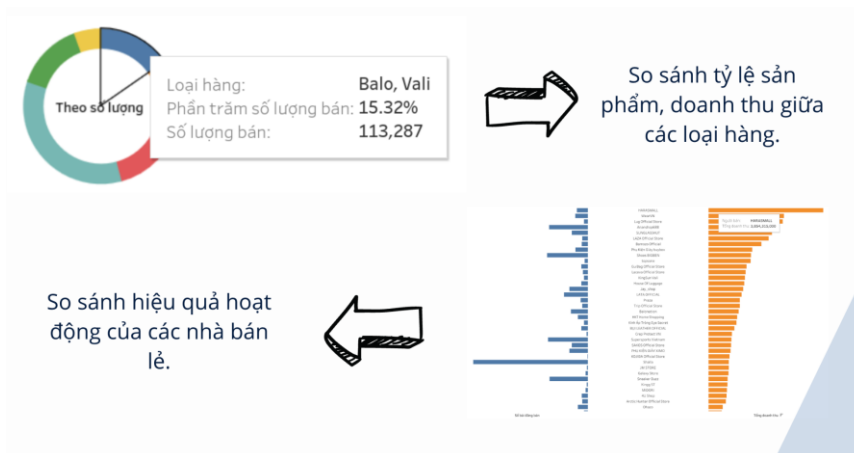
  

	number_of_images	vnd_cashback	has_video	category
0	3	0	False	Phụ kiện thời trang
1	6	0	True	Phụ kiện thời trang
2	7	0	True	Phụ kiện thời trang
3	4	361	False	Phụ kiện thời trang
4	3	0	False	Phụ kiện thời trang

	quantity_sold
0	106
1	5
2	17
3	66
4	63

## 2. Phù hợp với mục đích:



## 3. Rõ ràng và dễ hiểu:

Bố cục của các biểu đồ được sắp xếp trên dashboard tạo sự hài hòa và dễ đọc.

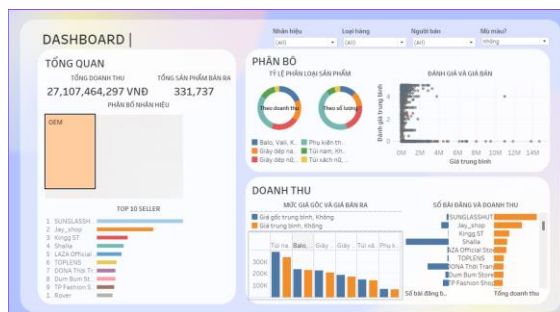
Tuy nhiên, vẫn còn một số chỗ bị khuyết dữ liệu do kích thước chữ lớn và nhiều trường dữ liệu.



#### 4. 4 & 5. Sự liên kết, tương tác và điều hướng:

Sử dụng tính năng Action của dashboard để tăng sự liên kết và tích hợp giữa các biểu đồ.

Sự thay đổi giữa các biểu đồ khi chọn các nhãn hàng

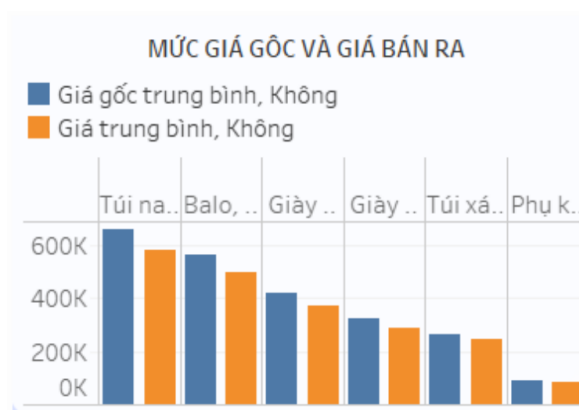
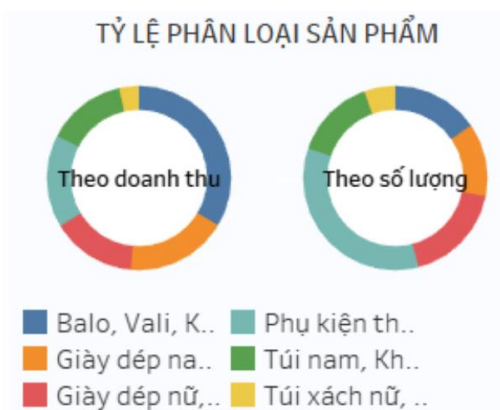


OEM



SAKOS

## 5. Thiết kế hấp dẫn:



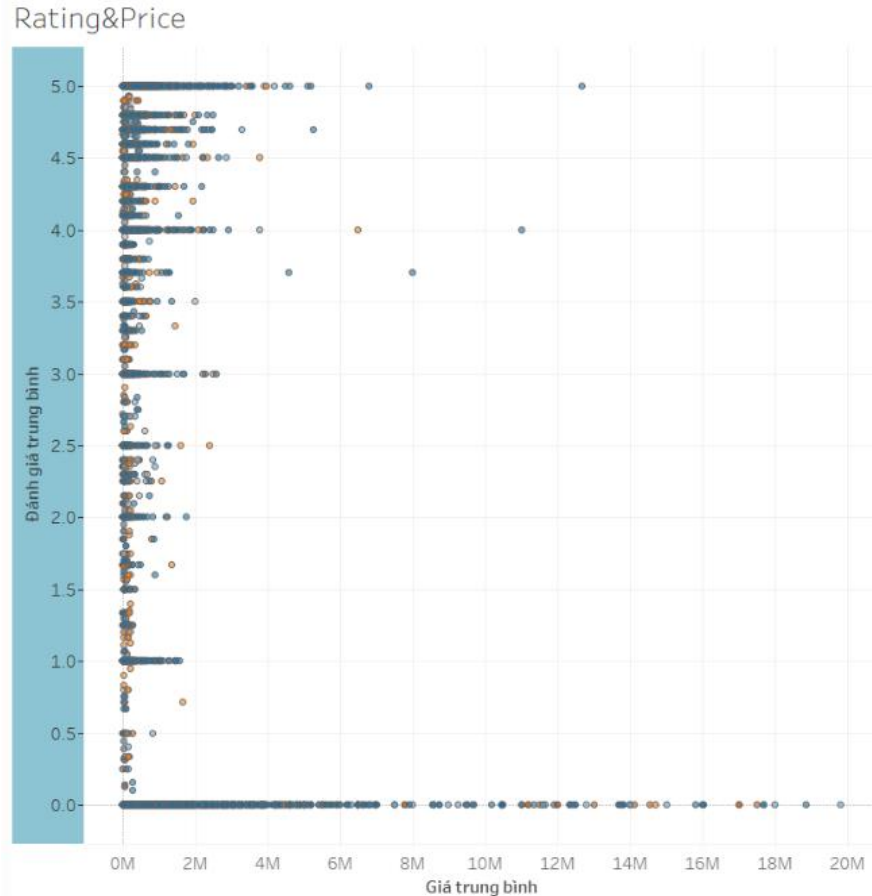
Màu sắc đa dạng

Sử dụng cặp màu tương phản

Ngoài ra dashboard có chế độ dành cho những người mù màu



## 6. Mối quan hệ và chiến lược cho các mặt hàng:



Biểu đồ scatter thể hiện mối quan hệ giữa giá bán và đánh giá của khách hàng

Từ biểu đồ scatter bên có thể rút ra một số chiến lược như:

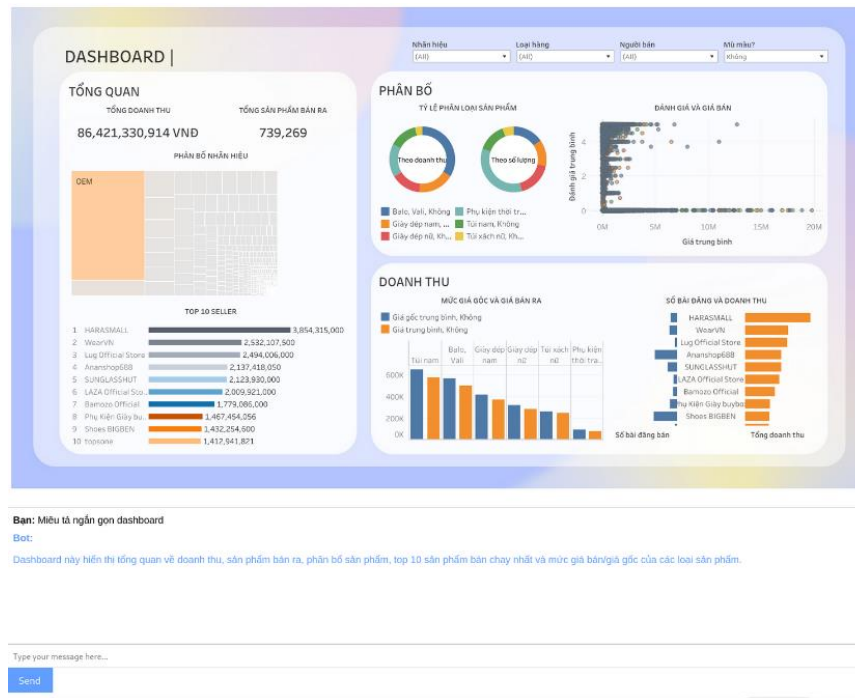
Để đạt ratings cao từ khách hàng:

- Tập trung vào chất lượng sản phẩm thay vì giá để duy trì sự hài lòng của khách hàng.
- Phân tích các đặc điểm của các sản phẩm được đánh giá cao để nhân rộng thành công này trên các danh mục khác.

Chiến lược định giá:

- Xem xét định giá cao hơn cho các sản phẩm chất lượng cao đã nhận được đánh giá tốt liên tục.
- Đảm bảo giá cả cạnh tranh cho các sản phẩm, đặc biệt là trong các khoảng giá thấp, để nâng cao đánh giá.
- Đầu tư vào túi xách nữ và giày dép nam, vì các danh mục này cho thấy sự cân bằng tiềm năng của đánh giá tốt ở nhiều mức giá khác nhau.

## 7. Tích hợp AI:



### Tính năng:

- Dùng Flask làm backend.
- Sử dụng Google Gemini Flash 1.5.
- Có thể hiển thị nội dung markdown từ LLM.
- Hỗ trợ phân tích dashboard.