

# BÁO CÁO PHÂN TÍCH DỮ LIỆU EUROPE BIKE STORE SALES

## I. MỤC ĐÍCH CỦA BÀI PHÂN TÍCH

Bài phân tích nhằm đánh giá hiệu quả kinh doanh của chuỗi cửa hàng Europe Bike Store thông qua dữ liệu doanh số bằng Google Colab. Cụ thể, các mục tiêu chính gồm:

- Tìm hiểu xu hướng doanh thu, lợi nhuận theo thời gian, khu vực và danh mục sản phẩm.
- Xác định sản phẩm bán chạy nhất và danh mục có lợi nhuận cao nhất.
- Phát hiện các khu vực hoặc sản phẩm có hiệu suất thấp để đề xuất cải tiến.

## II. MÔ TẢ BỘ DỮ LIỆU

Europe Bike Store Sales là một bộ dữ liệu phổ biến trên Kaggle, thường dùng để phân tích doanh số bán hàng của một chuỗi cửa hàng xe đạp tại nhiều quốc gia châu Âu.

Nguồn dữ liệu: <https://www.kaggle.com/datasets/pepinstaprine/europe-bike-store-sales/data>

Các cột dữ liệu chính:

1. Date: Ngày đặt hàng
2. Day: Ngày
3. Month: Tháng
4. Year: Năm
5. Customer\_Age: Tuổi
6. Age\_Group: Độ tuổi
7. Customer\_Gender: Giới tính
8. Country: Quốc gia
9. State: Khu vực
10. Product Category : Loại sản phẩm
11. SubCategory: danh mục con
12. Product: Tên sản phẩm
13. Order Quantity: Số lượng đặt hàng
14. Unit Cost & Unit Price: Giá vốn và giá bán mỗi đơn vị
15. Cost: Tổng giá vốn ( $\text{Order Quantity} \times \text{Unit Cost}$ )
16. Revenue: Tổng doanh thu ( $\text{Order Quantity} \times \text{Unit Price}$ )
17. Profit: Lợi nhuận ( $\text{Revenue} - \text{Cost}$ )

Dữ liệu có định dạng .xlsx với hàng nghìn bản ghi.

### III. PHÂN TÍCH DỮ LIỆU

#### 1. Đọc và làm sạch dữ liệu

-Sử dụng thư viện Pandas để đọc file excel từ google driver

```
import pandas as pd
df=pd.read_excel('/content/drive/MyDrive/Data Analysis Course/project/Sales.xlsx')
df
```

	Date	Day	Month	Year	Customer_Age	Age_Group	Customer_Gender	Country	State	Product_Category	Sub_Category	Product	Order_Quantity	Unit_Cost	Unit_Price	Profit	Cost	Revenue
0	2013-11-26	26	November	2013	19	Youth (<25)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
1	2015-11-26	26	November	2015	19	Youth (<25)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
2	2014-03-23	23	March	2014	49	Adults (35-64)	M	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	23	45	120	1366	1035	2401
3	2016-03-23	23	March	2016	49	Adults (35-64)	M	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	20	45	120	1188	900	2088
4	2014-05-15	15	May	2014	47	Adults (35-64)	F	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	238	180	418
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
113031	2016-04-12	12	April	2016	41	Adults (35-64)	M	United Kingdom	England	Clothing	Vests	Classic Vest, S	3	24	64	112	72	184
113032	2014-04-02	2	April	2014	18	Youth (<25)	M	Australia	Queensland	Clothing	Vests	Classic Vest, M	22	24	64	655	528	1183
113033	2016-04-02	2	April	2016	18	Youth (<25)	M	Australia	Queensland	Clothing	Vests	Classic Vest, M	22	24	64	655	528	1183
113034	2014-03-04	4	March	2014	37	Adults (35-64)	F	France	Seine (Paris)	Clothing	Vests	Classic Vest, L	24	24	64	684	576	1260
113035	2016-03-04	4	March	2016	37	Adults (35-64)	F	France	Seine (Paris)	Clothing	Vests	Classic Vest, L	23	24	64	655	552	1207

113036 rows x 18 columns

-Sử dụng các hàm để tiền xử lý dữ liệu:

Truy vấn thông tin tổng quan và kiểm tra dữ liệu bị thiếu

```
[44] print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113036 entries, 0 to 113035
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   Date                  113036 non-null  datetime64[ns]
 1   Day                   113036 non-null  int64  
 2   Month                 113036 non-null  object  
 3   Year                  113036 non-null  int64  
 4   Customer_Age          113036 non-null  int64  
 5   Age_Group             113036 non-null  object  
 6   Customer_Gender       113036 non-null  object  
 7   Country               113036 non-null  object  
 8   State                 113036 non-null  object  
 9   Product_Category      113036 non-null  object  
10   Sub_Category          113036 non-null  object  
11   Product               113036 non-null  object  
12   Order_Quantity        113036 non-null  int64  
13   Unit_Cost              113036 non-null  int64  
14   Unit_Price             113036 non-null  int64  
15   Profit                113036 non-null  int64  
16   Cost                  113036 non-null  int64  
17   Revenue               113036 non-null  int64  
dtypes: datetime64[ns](1), int64(9), object(8)
memory usage: 15.5+ MB
None
```

```
print(df.isnull().sum())
```

Date	0
Day	0
Month	0
Year	0
Customer_Age	0
Age_Group	0
Customer_Gender	0
Country	0
State	0
Product_Category	0
Sub_Category	0
Product	0
Order_Quantity	0
Unit_Cost	0
Unit_Price	0
Profit	0
Cost	0
Revenue	0
dtype: int64	

```
print(df.describe())
```

	Date	Day	Year
count	113036	113036.000000	113036.000000
mean	2014-11-23 12:14:55.063519232	15.665753	2014.401739
min	2011-01-01 00:00:00	1.000000	2011.000000
25%	2013-12-22 00:00:00	8.000000	2013.000000
50%	2014-06-27 00:00:00	16.000000	2014.000000
75%	2016-01-09 00:00:00	23.000000	2016.000000
max	2016-07-31 00:00:00	31.000000	2016.000000
std	NaN	8.781567	1.272510

	Customer_Age	Order_Quantity	Unit_Cost	Unit_Price
count	113036.000000	113036.000000	113036.000000	113036.000000
mean	35.919212	11.901660	267.296366	452.938427
min	17.000000	1.000000	1.000000	2.000000
25%	28.000000	2.000000	2.000000	5.000000
50%	35.000000	10.000000	9.000000	24.000000
75%	43.000000	20.000000	42.000000	70.000000
max	87.000000	32.000000	2171.000000	3578.000000
std	11.021936	9.561857	549.835483	922.071219

	Profit	Cost	Revenue
count	113036.000000	113036.000000	113036.000000
mean	285.051665	469.318695	754.370360
min	-30.000000	1.000000	2.000000
25%	29.000000	28.000000	63.000000
50%	101.000000	108.000000	223.000000
75%	358.000000	432.000000	800.000000
max	15896.000000	42978.000000	58074.000000
std	453.887443	884.866118	1309.094674

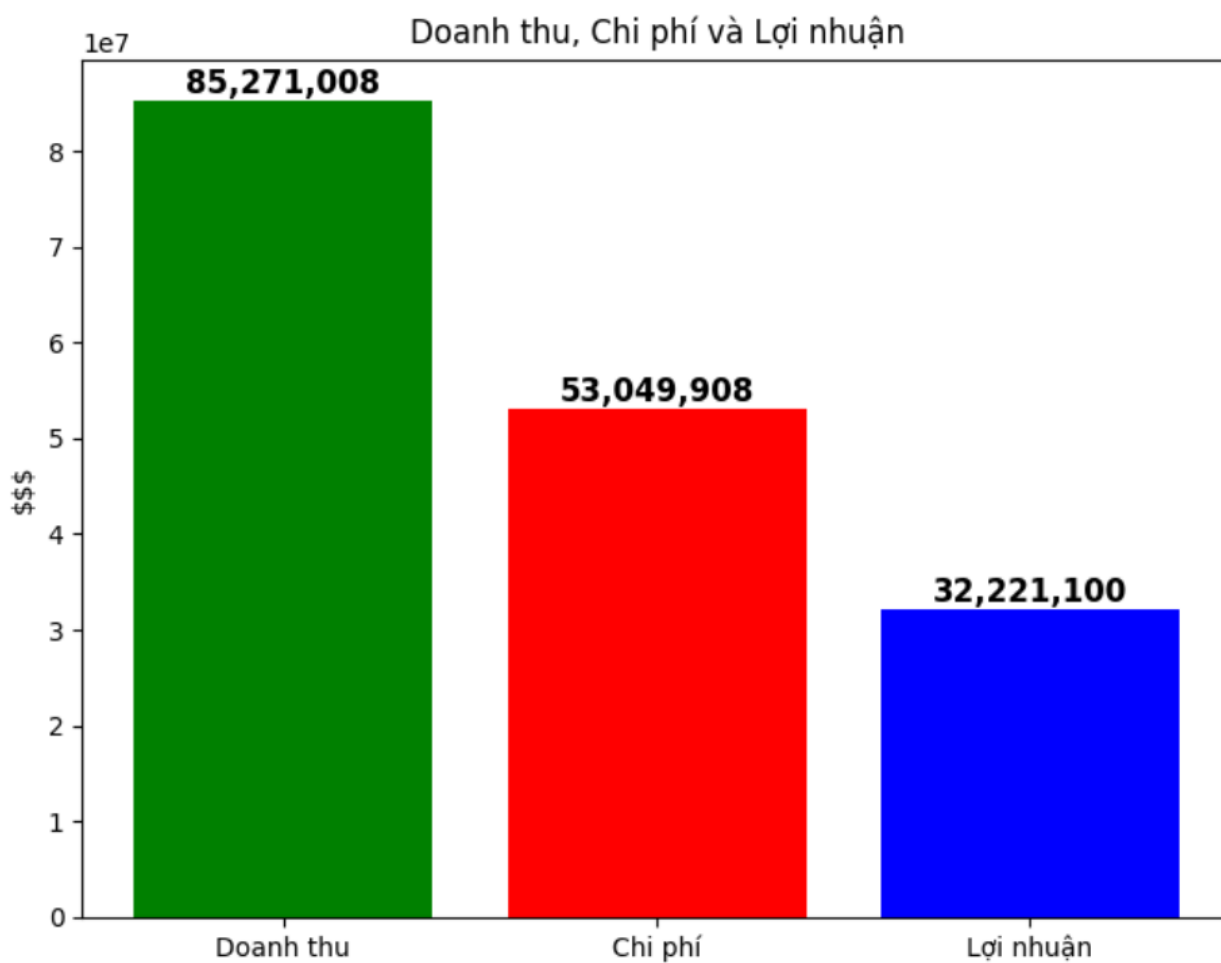
Sử dụng Hàm describe() để hiển thị một bảng thống kê mô tả cho dữ liệu

Tính toán trên dữ liệu các chỉ số quan trọng

Total Unit Cost	30,214,112
Total Unit Price	51,198,348
Total Order Quantity	1,345,316
Total Revenue	85,271,008
Total Cost	53,049,908
Total Profit	32,221,100

## 2. Trực quan hóa và phân tích

### a) Doanh thu, Chi phí và Lợi nhuận:



Doanh thu cao, lợi nhuận là chiếm gần 38%, chứng tỏ doanh nghiệp đang có hiệu suất tốt. Chi phí chiếm 62.2% doanh thu, đây là một mức hợp lý. Tuy nhiên, doanh nghiệp có thể tối ưu chi phí hơn nữa để cải thiện lợi nhuận.

## b) Lợi nhuận theo năm:

Lợi nhuận theo năm



Nhận xét về xu hướng lợi nhuận

- Giai đoạn 2011 - 2013:

Từ 2011 đến 2012, lợi nhuận tăng chậm.

Đến 2013, lợi nhuận tăng đột biến gần gấp đôi so với 2012, cho thấy doanh nghiệp có bước đột phá lớn.

- Giai đoạn 2014 - 2016:

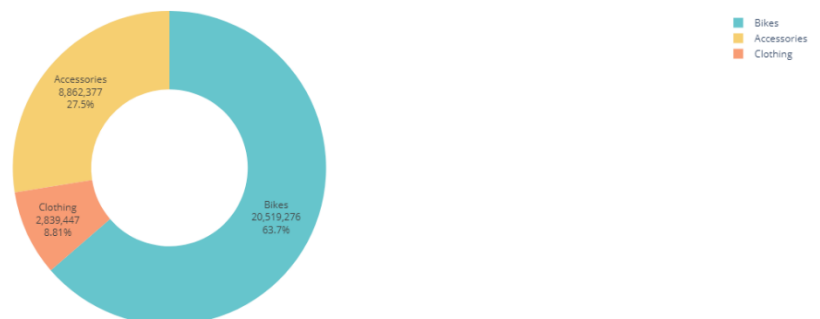
Năm 2014, lợi nhuận giảm nhẹ so với 2013, có thể do sự điều chỉnh của thị trường hoặc yếu tố nội bộ.

Năm 2015, lợi nhuận lại tăng mạnh, đạt mức cao nhất trong giai đoạn (7.53M).

Tuy nhiên, năm 2016 lợi nhuận giảm nhẹ, có thể do thị trường chững lại hoặc doanh nghiệp chưa duy trì được đà tăng trưởng.

## c) Lợi nhuận theo sản phẩm:

Lợi nhuận theo sản phẩm



### 1. Tổng quan về cơ cấu lợi nhuận

- Biểu đồ thể hiện tỷ trọng lợi nhuận theo từng nhóm sản phẩm:
  - Bikes: 20,519,276 (63.7%)
  - Accessories: 8,862,377 (27.5%)
  - Clothing: 2,839,447 (8.81%)

### 2. Nhận xét về lợi nhuận theo sản phẩm

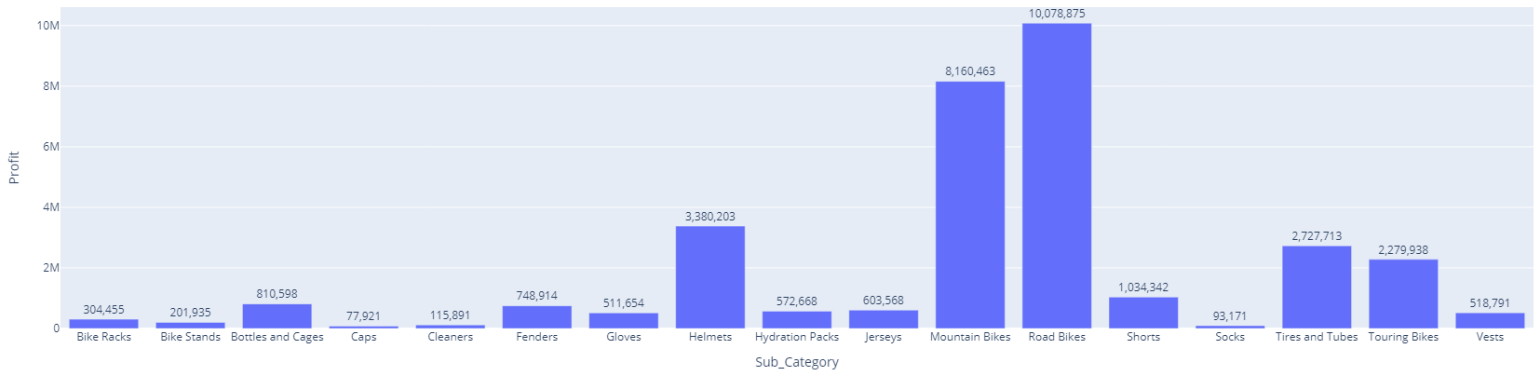
- Bikes (Xe đạp) chiếm tỷ trọng lợi nhuận lớn nhất (63.7%)
  - Đây là dòng sản phẩm mang lại phần lớn lợi nhuận cho doanh nghiệp.
  - Việc duy trì và mở rộng danh mục sản phẩm Bikes có thể tiếp tục gia tăng doanh thu.
  - Có thể tập trung vào phân khúc khách hàng mua Bikes để tối ưu hóa chiến lược marketing.
- Accessories (Phụ kiện) chiếm 27.5% lợi nhuận
  - Đây cũng là một mảng có đóng góp đáng kể.
  - Doanh nghiệp có thể khai thác thêm phụ kiện để tăng doanh thu, chẳng hạn như bán kèm khi khách hàng mua xe đạp.
- Clothing (Quần áo) có tỷ trọng thấp nhất (8.81%)
  - Lợi nhuận từ mảng quần áo thấp hơn đáng kể so với hai nhóm sản phẩm còn lại.
  - Cần đánh giá xem có nên tiếp tục đầu tư mạnh vào Clothing hay chỉ coi đây là dòng sản phẩm bổ trợ.
  - Nếu muốn cải thiện doanh số mảng này, có thể thử các chiến dịch khuyến mãi hoặc giới thiệu sản phẩm thời trang thể thao hấp dẫn hơn.

### 3. Đề xuất chiến lược kinh doanh

- Duy trì và tối ưu hóa mảng Bikes
  - Tiếp tục phát triển sản phẩm và cải thiện dịch vụ hậu mãi để giữ chân khách hàng.
  - Tập trung marketing vào nhóm khách hàng tiềm năng.
- Tận dụng cơ hội tăng trưởng từ Accessories
  - Đẩy mạnh chiến lược bán kèm (cross-selling) với Bikes.
  - Cải thiện danh mục sản phẩm phụ kiện để gia tăng giá trị đơn hàng.
- Cải thiện hiệu suất mảng Clothing
  - Xem xét lại chiến lược giá và sản phẩm để tăng sức cạnh tranh.
  - Nghiên cứu nhu cầu thị trường xem có nên mở rộng hay thu hẹp dòng sản phẩm này.

#### d) Lợi nhuận theo danh mục

Lợi nhuận theo danh mục



#### 1. Tổng quan về lợi nhuận các danh mục

Biểu đồ thể hiện lợi nhuận của từng danh mục sản phẩm. Một số danh mục nổi bật gồm:

- Road Bikes (Xe đạp đường trường): 10,078,875
- Mountain Bikes (Xe đạp địa hình): 8,160,463
- Helmets (Mũ bảo hiểm): 3,380,203
- Tires and Tubes (Lốp và ruột xe): 2,727,713
- Touring Bikes (Xe đạp du lịch): 2,279,938

#### 2. Nhận xét chi tiết theo danh mục

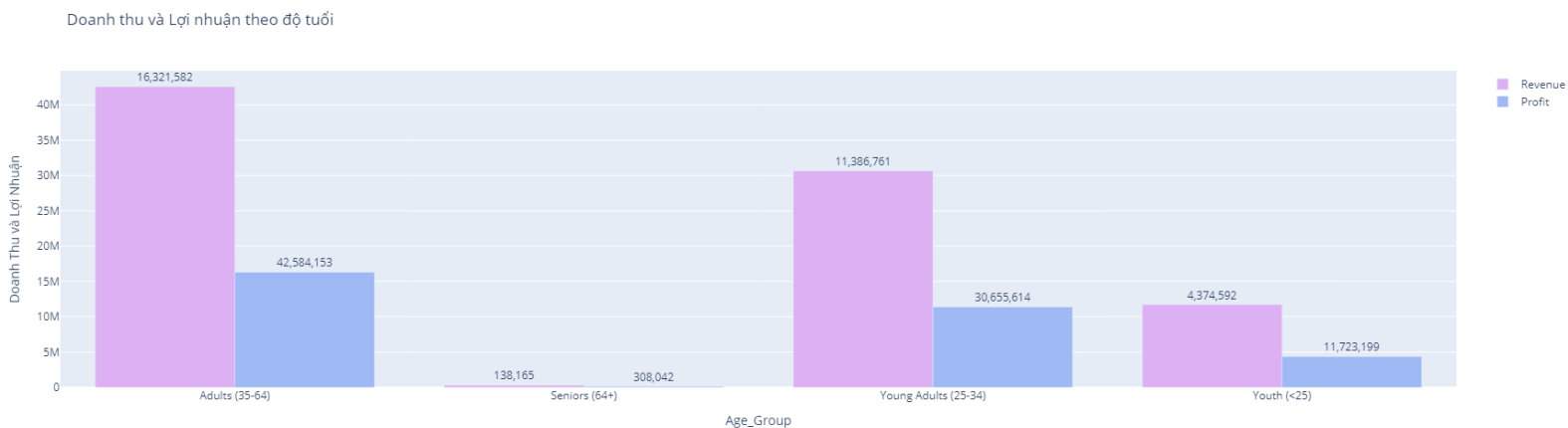
- Nhóm sản phẩm có lợi nhuận cao nhất
  - Road Bikes (10M) và Mountain Bikes (8.1M) đóng góp phần lớn vào tổng lợi nhuận.
  - Các dòng xe đạp này có giá trị cao và có nhu cầu lớn từ khách hàng.
  - Touring Bikes (2.2M) cũng có doanh thu tốt nhưng thấp hơn hai loại trên.
- Danh mục phụ kiện có lợi nhuận đáng kể
  - Helmets (3.38M) là phụ kiện có lợi nhuận cao nhất, cho thấy đây là mặt hàng được khách hàng quan tâm.
  - Tires and Tubes (2.7M) cũng có lợi nhuận tốt, có thể do nhu cầu thay thế thường xuyên.
- Nhóm sản phẩm có lợi nhuận thấp
  - Bike Stands (201K), Caps (77K), Socks (93K) có lợi nhuận khá thấp.
  - Cần xem xét lại chiến lược bán hàng cho các danh mục này, hoặc tập trung hơn vào các sản phẩm có lợi nhuận cao.

#### 3. Đề xuất chiến lược kinh doanh

- Duy trì thế mạnh của Road Bikes & Mountain Bikes

- Đẩy mạnh marketing và ưu đãi cho hai dòng sản phẩm này.
- Cung cấp thêm các gói dịch vụ bảo trì, nâng cấp để gia tăng doanh thu.
- Khai thác tiềm năng từ phụ kiện lợi nhuận cao
- Tiếp tục tập trung vào Helmets và Tires and Tubes.
- Kết hợp chương trình giảm giá hoặc bán kèm với xe đạp.
- Rà soát các danh mục lợi nhuận thấp
- Đánh giá xem có nên tiếp tục kinh doanh các sản phẩm như Bike Stands, Caps, Socks hay không.
- Nếu vẫn duy trì, có thể cải tiến sản phẩm hoặc thay đổi chiến lược tiếp thị.

#### e) Doanh thu và Lợi nhuận theo độ tuổi



#### 1. Tổng quan doanh thu và lợi nhuận theo độ tuổi

Biểu đồ thể hiện doanh thu (màu hồng) và lợi nhuận (màu xanh) của từng nhóm khách hàng theo độ tuổi. Một số điểm đáng chú ý:

- Nhóm Adults (35-64) có doanh thu cao nhất: 42.58M, lợi nhuận 16.32M.
- Nhóm Young Adults (25-34) xếp thứ hai với doanh thu 30.65M, lợi nhuận 11.38M.
- Nhóm Youth (<25) có doanh thu thấp hơn: 11.72M, lợi nhuận 4.37M.
- Nhóm Seniors (64+) có doanh thu và lợi nhuận thấp nhất: 308K doanh thu, 138K lợi nhuận.

#### 2. Nhận xét chi tiết theo độ tuổi

Nhóm khách hàng tiềm năng nhất: Adults (35-64)

- Đây là nhóm khách hàng có đóng góp lớn nhất về doanh thu và lợi nhuận.
- Họ có khả năng chi tiêu cao hơn, có thể quan tâm đến các dòng xe đạp chất lượng cao như Road Bikes, Mountain Bikes.

Young Adults (25-34) cũng là nhóm khách hàng quan trọng

Mặc dù thấp hơn nhóm Adults, nhưng doanh thu 30.65M vẫn rất đáng kể.

Nhóm này có thể quan tâm đến xe đạp thể thao, du lịch, cùng với các phụ kiện như Helmets, Tires and Tubes.

Nhóm Youth (<25) có doanh thu và lợi nhuận thấp hơn

- Doanh thu chỉ 11.72M, lợi nhuận 4.37M, thấp hơn hai nhóm trên.
- Có thể do khả năng chi tiêu hạn chế, tập trung vào sản phẩm giá rẻ hơn hoặc mua ít hơn.
- Nên xem xét các chương trình khuyến mãi hoặc sản phẩm phù hợp hơn với nhóm khách hàng này.

Nhóm Seniors (64+) gần như không đáng kể

- Doanh thu chỉ 308K, lợi nhuận 138K, rất thấp so với các nhóm khác.
- Có thể do nhu cầu sử dụng xe đạp không cao ở nhóm tuổi này.

### 3. Đề xuất chiến lược kinh doanh

Tập trung vào Adults (35-64) và Young Adults (25-34)

- Đẩy mạnh quảng bá các sản phẩm cao cấp, xe đạp chất lượng cao.
- Cung cấp các dịch vụ bảo trì, bảo dưỡng để tăng doanh thu từ nhóm này.

Tìm cách thu hút nhóm Youth (<25)

- Giới thiệu các dòng xe có giá phải chăng hơn.
- Chương trình khuyến mãi, trả góp hoặc giảm giá cho sinh viên có thể giúp tăng doanh thu.

Không cần tập trung nhiều vào Seniors (64+)

- Nếu vẫn muốn khai thác, có thể cung cấp các sản phẩm hỗ trợ như xe đạp trợ lực điện hoặc dịch vụ thuê xe.

## IV. KẾT LUẬN VÀ ĐỀ XUẤT

- **Tập trung sản phẩm chủ lực:** Đẩy mạnh Road Bikes, Mountain Bikes & phụ kiện đi kèm. Giảm bớt danh mục lợi nhuận thấp.
- **Chiến lược giá tối ưu:** Điều chỉnh giá theo nhu cầu, tạo combo sản phẩm để tăng giá trị đơn hàng.
- **Tiếp cận khách hàng tiềm năng:** Tăng cường quảng cáo, ưu đãi nhóm Adults (35-64) & Young Adults (25-34), hỗ trợ trả góp cho Youth (<25).
- **Mở rộng kênh bán hàng:** Tăng cường bán online (website, Shopee, Amazon), mở rộng khu vực tiềm năng.
- **Tối ưu chuỗi cung ứng & dịch vụ:** Giảm chi phí nhập hàng, dự báo nhu cầu chính xác, nâng cấp dịch vụ bảo hành & chăm sóc khách hàng.