

**BỘ CÔNG THƯƠNG**  
**TRƯỜNG ĐẠI HỌC KINH TẾ - KỸ**  
**THUẬT CÔNG NGHIỆP**

**KHOA KHOA HỌC ỨNG**  
**DỤNG**

**(Trang bìa)**  
**BÁO CÁO TỔNG KẾT**  
**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN**  
**NĂM HỌC 2024 – 2025**

Tên đề tài:

**NGHIÊN CỨU ỨNG DỤNG CỦA KHOA HỌC DỮ LIỆU**  
**TRONG THƯƠNG MẠI ĐIỆN TỬ**

<b>Giảng viên hướng dẫn:</b>	<b>ThS. Hà Anh Dũng</b>	<b>KHUỖ</b>
<b>Chủ nhiệm đề tài:</b>	<b>Trần Đức Trung</b>	<b>DHKL16A2HN</b>
<b>Thành viên:</b>	<b>Đậu Thị Thảo</b>	<b>DHKL16A2HN</b>
	<b>Đào Xuân Tâm</b>	<b>DHKL16A2HN</b>
	<b>Khuất Thanh</b>	<b>DHKL16A2HN</b>
	<b>Phương</b>	

**HÀ NỘI, 04/2025**

**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC KINH TẾ - KỸ  
THUẬT CÔNG NGHIỆP**

**KHOA KHOA HỌC ỨNG  
DỤNG**

**BÁO CÁO TỔNG KẾT  
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN  
NĂM HỌC 2023 - 2024**

Tên đề tài:

**NGHIÊN CỨU ỨNG DỤNG CỦA KHOA HỌC DỮ LIỆU  
TRONG THƯƠNG MẠI ĐIỆN TỬ**

<b>Giảng viên hướng dẫn:</b>	<b>ThS. Hà Anh Dũng</b>	<b>KHUỖ</b>
<b>Chủ nhiệm đề tài:</b>	<b>Trần Đức Trung</b>	<b>DHKL16A2HN</b>
<b>Thành viên:</b>	<b>Đậu Thị Thảo</b>	<b>DHKL16A2HN</b>
	<b>Đào Xuân Tâm</b>	<b>DHKL16A2HN</b>
	<b>Khuất Thanh Phương</b>	<b>DHKL16A2HN</b>

**Hà Nội, 04/2025**

## MỤC LỤC

MỘT SỐ KÝ HIỆU VÀ CHỮ VIẾT TẮT.....	V
DANH MỤC BẢNG BIỂU HÌNH VẼ.....	VI
LỜI MỞ ĐẦU .....	1
<b>Chương I: QUÁ TRÌNH HÌNH THÀNH VÀ PHÁT TRIỂN CỦA THƯƠNG MẠI ĐIỆN TỬ</b> .....	<b>3</b>
1.1. Khái niệm và sự hình thành ban đầu .....	3
1.2. Giai đoạn Internet bùng nổ (1990–2000).....	3
1.3. Giai đoạn phát triển và chuẩn hóa (2000–2010).....	3
1.4. Giai đoạn bùng nổ và chuyển đổi số (2010–2020) .....	4
1.5. Giai đoạn hậu đại dịch và xu hướng tương lai (2020 – nay) .....	4
<b>Chương 2: ỨNG DỤNG CỦA KHOA HỌC DỮ LIỆU TRONG THƯƠNG MẠI ĐIỆN TỬ</b> .....	<b>6</b>
2.1 Phân tích dữ liệu khách hàng .....	6
2.1.1. Đọc dữ liệu file.....	6
2.1.2. Phân tích doanh số bán hàng.....	6
2.1.3. Phân tích tần suất mua hàng.....	7
<i>Biểu đồ : Phân phối số đơn hàng mỗi khách hàng</i> <b>Error! Bookmark not defined.</b>	
<i>Biểu đồ : Phân phối số ngày giữa các lần mua hàng</i> <b>Error! Bookmark not defined.</b>	
2.1.4. Phân tích vòng đời khách hàng .....	10
2.1.5. Trực quan hóa dữ liệu.....	11
2.1.6. Phân tích khách hàng theo độ tuổi .....	12
2.2. Khoa học dữ liệu và chiến lược Marketing hiệu quả .....	19
2.2.1. Khái niệm cơ bản về Khoa học dữ liệu và chiến lược Marketing hiệu quả	19
2.2.2. Các đặc trưng cơ bản của Khoa học dữ liệu và Marketing .....	19
2.2.3. Tầm quan trọng của Khoa học dữ liệu trong Marketing .....	20
2.2.4. Phân khúc khách hàng trong Marketing .....	21
2.2.5. Đo lường và đánh giá hiệu quả chiến dịch marketing (KPI) .....	22
2.2.6. Tối ưu hóa chiến dịch marketing .....	22
2.2.7. Khoa học dữ liệu giúp dự đoán xu hướng kinh doanh.....	23
2.2.8. Ứng dụng của Khoa học dữ liệu trong Marketing .....	23
2.3. Tối ưu hóa quy trình vận hành và chuỗi cung ứng.....	36
2.3.1. Đọc dữ liệu chuỗi cung ứng .....	36

2.3.2. Phân tích chuỗi cung ứng là gì? .....	37
2.3.3. Dịch vụ chăm sóc khách hàng.....	38
2.3.4. Xử lý mạng lưới chuỗi cung ứng .....	38
2.3.5. Tận dụng tối đa tài sản vào chuỗi cung ứng.....	39
2.3.6. Đo lường và đánh giá hiệu suất làm việc của chuỗi cung ứng .....	41
2.3.7. Tối ưu hóa vận chuyển giao hàng .....	43
2.4. Quản lý tồn kho và giao hàng .....	46
2.4.1. Tối ưu hóa kho hàng.....	46
2.4.2. Quản lý giao hàng .....	52
<b>KẾT LUẬN VÀ KIẾN NGHỊ .....</b>	<b>57</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>58</b>
<b>PHIẾU ĐĂNG KÝ ĐỀ TÀI (<i>Dành cho sinh viên</i>) .....</b>	<b>60</b>
<b>THUYẾT MINH ĐỀ TÀI NGHIÊN CỨU KHOA HỌC .....</b>	<b>61</b>

## **MỘT SỐ KÝ HIỆU VÀ CHỮ VIẾT TẮT**

AI	Trí tuệ nhân tạo
TMĐT	Thương mại điện tử
DN	Doanh nghiệp
NLP	Xử lý ngôn ngữ tự nhiên

## DANH MỤC BẢNG BIỂU HÌNH VẼ

<i>Hình 2.1: Phân phối số đơn hàng mỗi khách hàng .....</i>	<i>9</i>
<i>Hình 2. 2: Phân phối số ngày giữa các lần mua hàng.....</i>	<i>10</i>
<i>Hình 2.3: Biểu đồ phân phối tổng doanh thu từ mỗi khách hàng.....</i>	<i>12</i>
<i>Hình 2.4: Phân phối độ tuổi khách hàng .....</i>	<i>14</i>
<i>Hình 2.5: So sánh giá trị thực tế và giá trị dự đoán .....</i>	<i>18</i>
<i>Hình 2.6: Doanh thu và tốc độ tăng trưởng của thương mại điện tử qua các năm .....</i>	<i>21</i>
<i>Hình 2.7: Thị phần giá trị giao dịch của các sàn thương mại điện tử quý 2/2024 .....</i>	<i>25</i>
<i>Hình 2.8: Hiển thị 5 dòng đầu của dữ liệu. ....</i>	<i>37</i>
<i>Hình 2.9: Biểu đồ mối quan hệ giữa giá sản phẩm và doanh thu .....</i>	<i>38</i>
<i>Hình 2.10: Dùng describe() để thống kê mô tả. ....</i>	<i>40</i>
<i>Hình 2.11: Biểu đồ doanh thu theo từng sản phẩm. ....</i>	<i>41</i>
<i>Hình 2.12: Biểu đồ doanh thu được tạo ra bởi SKU. ....</i>	<i>42</i>
<i>Hình 2.13: Biểu đồ tỉ lệ lỗi trung bình theo từng loại sản phẩm. ....</i>	<i>42</i>
<i>Hình 2.14: Biểu đồ tổng doanh thu theo hãng vận chuyển.....</i>	<i>44</i>
<i>Hình 2.15: Biểu đồ chi phí theo nhà vận chuyển. ....</i>	<i>44</i>
<i>Hình 2.16: Thời gian giao hàng trung bình và chi phí sản xuất trung bình.....</i>	<i>45</i>
<i>Hình 2.17: Biểu đồ chi phí theo phương thức vận chuyển.....</i>	<i>45</i>
<i>Hình 2.18: Biểu đồ thời gian giao hàng trung bình theo nhà vận chuyển. ....</i>	<i>46</i>
<i>Hình 2.19: Bố trí kho hàng 3D.....</i>	<i>50</i>
<i>Hình 2.20: Hiệu quả sử dụng không gian .....</i>	<i>52</i>
<i>Hình 2.21: Dự đoán thời gian giao hàng.....</i>	<i>54</i>
<i>Hình 2.22: Tuyến đường giao hàng .....</i>	<i>55</i>

# LỜI MỞ ĐẦU

## 1. Tổng quan về tình hình nghiên cứu và lý do chọn đề tài

Thương mại điện tử (TMĐT) là lĩnh vực nghiên cứu liên ngành, kết hợp kinh tế, công nghệ, quản trị và hành vi tiêu dùng. Sự phát triển mạnh mẽ của TMĐT trong những thập kỷ gần đây đã thu hút nhiều học giả nghiên cứu từ nhiều góc độ khác nhau.

### 1. Các giai đoạn nghiên cứu chính

**Giai đoạn đầu (trước năm 2000):** Tập trung vào tiềm năng của internet đối với thương mại và các mô hình kinh doanh trực tuyến sơ khai.

**Giai đoạn phát triển (2000 - 2010):** Xuất hiện các nghiên cứu về hạ tầng công nghệ, bảo mật thanh toán, hành vi mua sắm trực tuyến và sự thay đổi trong chuỗi cung ứng.

Giai đoạn bùng nổ (2010 - nay): Nghiên cứu mở rộng sang TMĐT di động, tác động của trí tuệ nhân tạo (AI), dữ liệu lớn (Big Data), blockchain, cá nhân hóa trải nghiệm mua sắm và mô hình kinh doanh đa kênh (Omni-channel).

### 2. Các hướng nghiên cứu chính

Mô hình kinh doanh TMĐT: Nghiên cứu về các loại hình như B2B, B2C, C2C, B2G, xu hướng TMĐT xuyên biên giới.

Hành vi tiêu dùng trực tuyến: Phân tích động cơ mua sắm, trải nghiệm khách hàng, yếu tố ảnh hưởng đến quyết định mua hàng.

Công nghệ và bảo mật trong TMĐT: Nghiên cứu về AI, chatbot, thanh toán điện tử, blockchain, an ninh mạng.

Marketing số và cá nhân hóa: Chiến lược tiếp thị trực tuyến, SEO, quảng cáo trên mạng xã hội, thương mại hội thoại (Conversational Commerce).

Quản lý chuỗi cung ứng điện tử (E-SCM): Tối ưu hóa logistics, giao hàng chặng cuối, kho thông minh.

### 3. Xu hướng nghiên cứu mới

Ứng dụng trí tuệ nhân tạo trong TMĐT

Thương mại điện tử bền vững và thân thiện với môi trường

Ảnh hưởng của TMĐT đến kinh tế và xã hội

Pháp lý và chính sách điều chỉnh TMĐT

Tình hình nghiên cứu về TMĐT đang ngày càng phát triển mạnh mẽ với sự thay đổi nhanh chóng của công nghệ và hành vi tiêu dùng. Việc tiếp tục nghiên cứu và cập nhật các xu hướng mới sẽ giúp doanh nghiệp và nhà nghiên cứu tận dụng hiệu quả những cơ hội mà TMĐT mang lại.

## 2. Mục đích, đối tượng và phạm vi nghiên cứu của đề tài

Mục đích nghiên cứu của đề tài là nghiên cứu áp dụng phần mềm Python để giải quyết các vấn đề về Thương mại điện tử.

**Mục tiêu:** Đề tài nghiên cứu nhằm giải quyết các vấn đề cơ bản:

- Về lý luận: Nghiên cứu lịch sử hình thành và phát triển Thương mại điện tử.
- Nghiên cứu áp dụng phần mềm Python trong các lĩnh vực của Thương mại điện tử: Phân tích dữ liệu khách hàng, Marketing, Tối ưu hóa quy trình vận hành chuỗi cung ứng, Quản lý tồn kho và giao hàng

**Đối tượng nghiên cứu:** Thương mại điện tử.

**Phạm vi nghiên cứu:**

Tất cả các lĩnh vực trong Thương mại điện tử.

### **3. Phương pháp và nhiệm vụ nghiên cứu**

- Phương pháp nghiên cứu tài liệu: Sử dụng các tài liệu do thầy hướng dẫn đưa ra để có cơ sở lý thuyết đơn giản và dễ hiểu nhất.
- Phương pháp vận dụng thực tế: Sử dụng ngôn ngữ lập trình Python để viết các code ứng dụng trong Thương mại điện tử.
- Phương pháp phân tích, tổng hợp, đánh giá.

### **4. Cấu trúc và các kết quả của đề tài**

Ngoài phần lời mở đầu, kết luận và tài liệu tham khảo, đề tài được chia làm 2 chương với cấu trúc và kết quả của từng chương như sau:

- Chương 1: Quá trình hình thành và phát triển của Thương mại điện tử.
- Chương 2: Ứng dụng của Khoa học dữ liệu trong Thương mại điện tử.

### **5. Ý nghĩa các kết quả của đề tài**

Đề tài nghiên cứu viết các code để phân tích dữ liệu khách hàng, chiến lược marketing hiệu quả, tối ưu hóa chuỗi cung ứng, quản lý tồn kho và giao hàng. Các code này giúp cho việc giảng dạy và học tập học phần Thương mại điện tử được thuận tiện hơn. Ngoài ra cũng giúp các công ty kinh doanh online, thương mại điện tử hoạt động thuận tiện hơn. Các phần mềm này có thể chỉnh sửa theo yêu cầu của các công ty kinh doanh online.



# Chương I: QUÁ TRÌNH HÌNH THÀNH VÀ PHÁT TRIỂN CỦA THƯƠNG MẠI ĐIỆN TỬ

## 1.1. *Khái niệm và sự hình thành ban đầu*

Thương mại điện tử (E-commerce) là hình thức mua bán hàng hóa và dịch vụ thông qua các phương tiện điện tử, đặc biệt là mạng Internet. Tuy nhiên, khái niệm này không mới mà đã bắt đầu từ thập niên 1960, gắn liền với sự phát triển của công nghệ truyền dữ liệu điện tử (Electronic Data Interchange – EDI).

Vào những năm 1960, các doanh nghiệp lớn ở Mỹ và châu Âu bắt đầu sử dụng EDI để trao đổi hóa đơn, đơn đặt hàng và các tài liệu kinh doanh một cách tự động hóa giữa các công ty. Hình thức này giúp rút ngắn thời gian xử lý thông tin và giảm lỗi do con người gây ra. Tuy chưa có Internet như ngày nay, nhưng đây chính là tiền đề cho sự ra đời của thương mại điện tử hiện đại.

## 1.2. *Giai đoạn Internet bùng nổ (1990–2000)*

Sự kiện đánh dấu bước ngoặt quan trọng của thương mại điện tử là sự ra đời của trình duyệt web đầu tiên – WorldWideWeb (sau này đổi tên thành Nexus) vào năm 1990 bởi Tim Berners-Lee, cha đẻ của World Wide Web. Đến giữa thập niên 1990, sự phổ biến của trình duyệt Netscape và sự thương mại hóa Internet đã tạo điều kiện cho các doanh nghiệp khai thác nền tảng trực tuyến để kinh doanh.

Trong giai đoạn 1995–2000, hàng loạt công ty thương mại điện tử đầu tiên ra đời. Đáng chú ý nhất là:

- Amazon.com (thành lập năm 1994, hoạt động chính thức năm 1995) – ban đầu là một hiệu sách trực tuyến, sau trở thành "gã khổng lồ" trong lĩnh vực bán lẻ toàn cầu.
- eBay (ra đời năm 1995) – nền tảng đấu giá trực tuyến đầu tiên, cho phép người dùng mua bán hàng hóa trực tiếp với nhau.
- Alibaba (ra đời năm 1999 tại Trung Quốc) – nền tảng thương mại điện tử B2B quy mô lớn, kết nối các doanh nghiệp với nhau.

Giai đoạn này được xem là "thời kỳ sơ khai vàng" của thương mại điện tử, dù còn nhiều hạn chế về kết nối Internet, bảo mật và thói quen tiêu dùng. Tuy nhiên, nó đã đặt nền móng cho sự phát triển bùng nổ sau này.

## 1.3. *Giai đoạn phát triển và chuẩn hóa (2000–2010)*

Đầu những năm 2000, bong bóng dot-com vỡ khiến nhiều công ty công nghệ và thương mại điện tử phá sản. Tuy nhiên, các doanh nghiệp có nền tảng vững chắc như Amazon, eBay vẫn tồn tại và mở rộng mạnh mẽ.

Trong giai đoạn này, thương mại điện tử bước vào thời kỳ phát triển ổn định hơn, với sự chuẩn hóa về hạ tầng kỹ thuật và pháp lý:

- Bảo mật được cải thiện, đặc biệt với sự xuất hiện của giao thức HTTPS, các công nghệ mã hóa và hệ thống thanh toán an toàn hơn như PayPal.

- Cổng thanh toán điện tử như PayPal, Stripe, Alipay phát triển mạnh, tạo điều kiện thuận lợi cho người tiêu dùng và doanh nghiệp.
- Nhiều quốc gia bắt đầu xây dựng khung pháp lý về thương mại điện tử, đảm bảo quyền lợi người tiêu dùng và tạo môi trường cạnh tranh lành mạnh.

Ngoài ra, sự phát triển của logistics và chuỗi cung ứng cũng đóng vai trò quan trọng trong việc thúc đẩy thương mại điện tử.

#### *1.4. Giai đoạn bùng nổ và chuyển đổi số (2010–2020)*

Từ năm 2010 trở đi, thương mại điện tử bước vào giai đoạn tăng trưởng vượt bậc nhờ sự phổ cập của smartphone, mạng 4G, và sự phát triển của mạng xã hội.

- Mobile commerce (m-commerce) – mua sắm qua thiết bị di động – phát triển nhanh chóng.
- Các nền tảng thương mại điện tử đa dạng hơn: app điện thoại, social commerce (bán hàng qua Facebook, Instagram).
- Hàng loạt sàn thương mại điện tử ra đời và phát triển mạnh mẽ ở từng khu vực: Lazada, Shopee, Tiki ở Đông Nam Á; JD.com, Taobao ở Trung Quốc; Etsy ở Mỹ...

Bên cạnh đó, trí tuệ nhân tạo (AI) và dữ liệu lớn (Big Data) bắt đầu được ứng dụng trong cá nhân hóa trải nghiệm khách hàng, quản lý tồn kho, và phân tích hành vi tiêu dùng.

Thương mại điện tử lan rộng sang các lĩnh vực như B2B, C2C, và giao dịch điện tử giữa chính phủ và doanh nghiệp/người dân (G2B, G2C).

#### *1.5. Giai đoạn hậu đại dịch và xu hướng tương lai (2020 – nay)*

Đại dịch COVID-19 từ năm 2020 là chất xúc tác cực mạnh thúc đẩy thương mại điện tử toàn cầu. Trong bối cảnh giãn cách xã hội, người tiêu dùng buộc phải chuyển sang mua sắm trực tuyến.

- Các ngành như thực phẩm, y tế, học tập và giải trí trực tuyến phát triển mạnh.
- Doanh nghiệp nhỏ và vừa cũng nhanh chóng chuyển đổi sang mô hình kinh doanh online.

Sau đại dịch, hành vi tiêu dùng thay đổi: ưu tiên sự tiện lợi, thanh toán không dùng tiền mặt, giao hàng nhanh, và nền tảng đa kênh (omnichannel).

Xu hướng mới trong thương mại điện tử:

- AI và Chatbot hỗ trợ chăm sóc khách hàng.
- AR/VR tăng cường trải nghiệm thử sản phẩm.
- Live commerce (bán hàng livestream) bùng nổ.
- Blockchain và tiền mã hóa được ứng dụng trong thanh toán.
- Green E-commerce – thương mại điện tử thân thiện môi trường ngày càng được quan tâm.

## **KẾT LUẬN**

Lịch sử phát triển của thương mại điện tử là hành trình dài gắn liền với sự tiến bộ của công nghệ, thay đổi hành vi tiêu dùng và sự chuyển dịch mô hình kinh doanh toàn cầu. Từ những trao đổi dữ liệu đơn giản, đến những nền tảng đa kênh phức tạp, thương mại điện tử đã và đang trở thành trụ cột quan trọng của nền kinh tế số hiện đại.

Trong tương lai, thương mại điện tử sẽ không ngừng đổi mới, tích hợp sâu hơn với các công nghệ tiên tiến, đồng thời gắn bó chặt chẽ với các yếu tố bền vững, cá nhân hóa và toàn cầu hóa.

## Chương 2: ỨNG DỤNG CỦA KHOA HỌC DỮ LIỆU TRONG THƯƠNG MẠI ĐIỆN TỬ

### 2.1 Phân tích dữ liệu khách hàng

#### 2.1.1. Đọc dữ liệu file

Dùng thư viện pandas để đọc file, làm sạch dữ liệu và hiển thị các dữ liệu và thông tin của người tiêu dùng:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Đường dẫn đến tệp CSV
file_path = '/data/customer_purchase_history.csv'
df = pd.read_csv(file_path)

# Loại bỏ khoảng trắng trong tên cột nếu có
df.columns = df.columns.str.strip()


# Hiển thị một số dòng dữ liệu đầu tiên
print(df.head())

# Kiểm tra thông tin dữ liệu
print(df.info())

# Kiểm tra các giá trị thiếu
print(df.isnull().sum())

# Xử lý các giá trị thiếu (nếu có)
df = df.dropna()
```

Kết quả chương trình:



	order_id	customer_id	order_date	purchase_amount	quantity
0	1	23	2020-06-29	327.74	9
1	2	74	2023-01-26	265.88	10
2	3	81	2022-10-12	233.18	6
3	4	76	2022-07-31	22.53	5
4	5	92	2021-04-14	162.19	4

#### 2.1.2. Phân tích doanh số bán hàng

Đoạn code sau sử dụng pandas để tìm tổng số khách hàng, đơn hàng và doanh thu:

```

▶ # Tổng số khách hàng
total_customers = df['customer_id'].nunique()
print("Tổng số khách hàng:", total_customers)

# Tổng số đơn hàng
total_orders = df['order_id'].nunique()
print("Tổng số đơn hàng:", total_orders)

# Tổng doanh thu
total_revenue = df['purchase_amount'].sum()
print("Tổng doanh thu:", total_revenue)

# Số lượng sản phẩm mua trung bình mỗi đơn hàng
average_items_per_order = df['quantity'].mean()
print("Số lượng sản phẩm mua trung bình mỗi đơn hàng:", average_items_per_order)

```

Trong đó:

- `df['customer_id'].nunique()`: Đếm số lượng **khách hàng duy nhất** trong cột `customer_id`.
- `nunique()` trả về số lượng giá trị duy nhất (tức là số lượng khách hàng khác nhau).
- `df['order_id'].nunique()`: Đếm số lượng **đơn hàng duy nhất** trong cột `order_id`
- `df['purchase_amount'].sum()`: Tính **tổng doanh thu** bằng cách cộng tất cả các giá trị trong cột `purchase_amount`.

Kết quả chương trình:

```

➡ Tổng số khách hàng: 100
   Tổng số đơn hàng: 1000
   Tổng doanh thu: 255369.43
   Số lượng sản phẩm mua trung bình mỗi đơn hàng: 5.594

```

### 2.1.3. Phân tích tần suất mua hàng

Đoạn code sau phân tích **tần suất mua hàng** bằng Python, sử dụng thư viện Pandas để xử lý dữ liệu và Seaborn để trực quan hóa kết quả. Cụ thể:

```

# Tính số đơn hàng mỗi khách hàng
customer_order_counts = df.groupby('customer_id')['order_id'].nunique()

# Vẽ biểu đồ phân phối số đơn hàng mỗi khách hàng
plt.figure(figsize=(10, 6))
sns.histplot(customer_order_counts, bins=30, kde=True)
plt.title('Phân phối số đơn hàng mỗi khách hàng')
plt.xlabel('Số đơn hàng')
plt.ylabel('Số lượng khách hàng')
plt.show()

# Tính toán số ngày giữa các lần mua hàng của từng khách hàng
df['order_date'] = pd.to_datetime(df['order_date'])
df = df.sort_values(by=['customer_id', 'order_date'])

df['previous_order_date'] = df.groupby('customer_id')['order_date'].shift(1)
df['days_between_orders'] = (df['order_date'] - df['previous_order_date']).dt.days

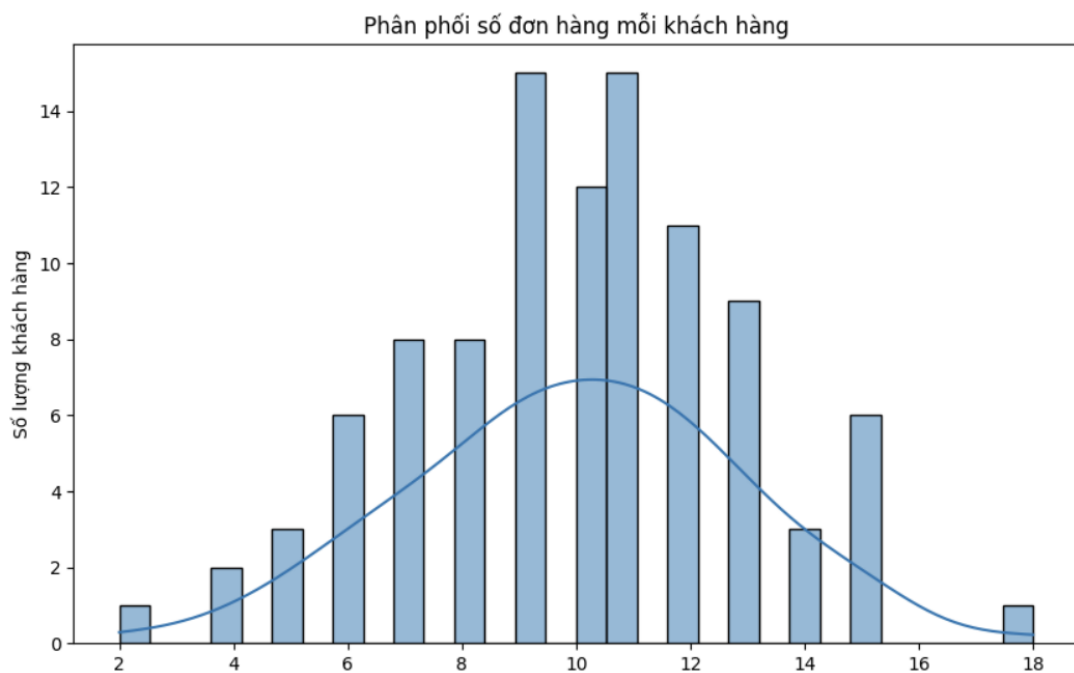
# Vẽ biểu đồ phân phối số ngày giữa các lần mua hàng
plt.figure(figsize=(10, 6))
sns.histplot(df['days_between_orders'].dropna(), bins=30, kde=True)
plt.title('Phân phối số ngày giữa các lần mua hàng')
plt.xlabel('Số ngày')
plt.ylabel('Số lượng')
plt.show()

```

Trong đó:

- **groupby('customer\_id')**: Nhóm dữ liệu theo mã khách hàng (customer\_id).
- **['order\_id'].nunique()**: Đếm số lượng **đơn hàng duy nhất** (order\_id) của từng khách hàng.
- **customer\_order\_counts** sẽ lưu số lượng đơn hàng của từng khách hàng.
- **pd.to\_datetime(df['order\_date'])**: Chuyển đổi cột order\_date về kiểu dữ liệu datetime.
- **sort\_values(by=['customer\_id', 'order\_date'])**:
  - Sắp xếp dữ liệu theo customer\_id trước, rồi theo order\_date tăng dần.
  - Đảm bảo các đơn hàng của từng khách hàng được xếp theo **thứ tự thời gian**.
- **groupby('customer\_id')**: Nhóm theo customer\_id.
- **shift(1)**: Lấy ngày mua hàng **trước đó** của từng khách hàng.
- **previous\_order\_date**: Ngày mua hàng trước đó của cùng một khách hàng.
- **days\_between\_orders**:
  - Tính **khoảng cách giữa các lần mua hàng** bằng cách trừ order\_date cho previous\_order\_date.
  - **.dt.days**: Lấy kết quả dưới dạng **số ngày**.

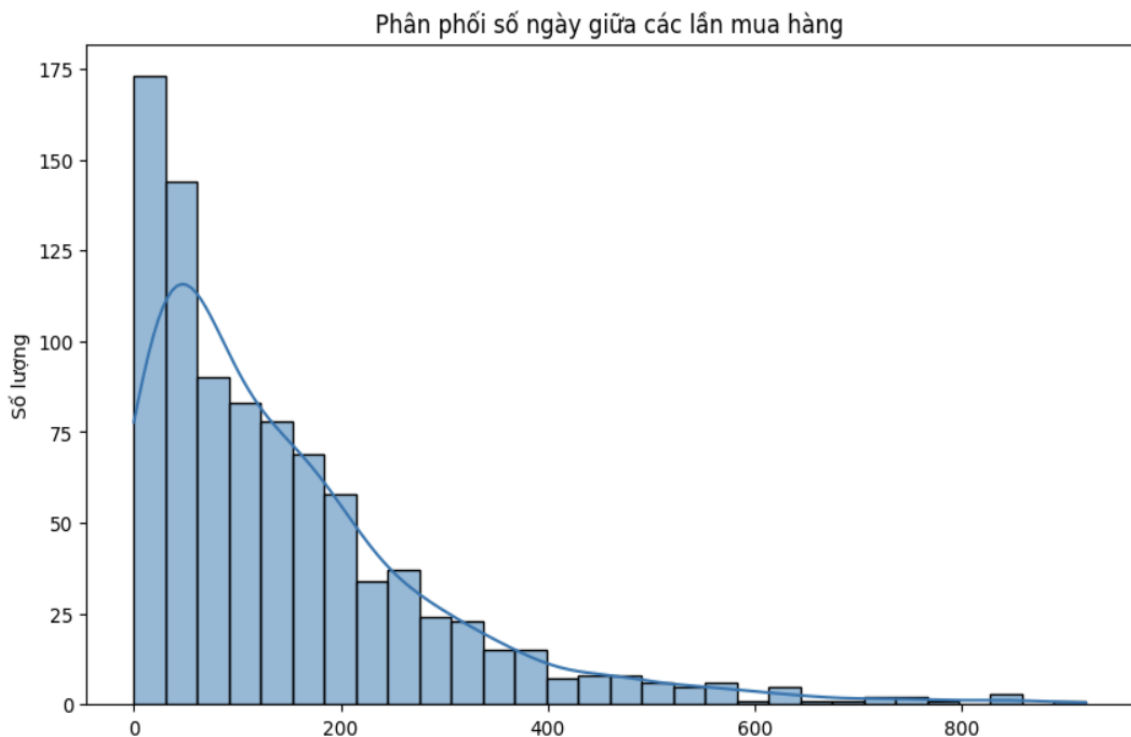
Kết quả chương trình:



*Hình 2.1: Phân phối số đơn hàng mỗi khách hàng*

**Nhận xét:**

- Phân phối có dạng gần giống phân phối chuẩn, hơi lệch phải nhẹ.
- Phần lớn khách hàng đặt từ 8 đến 12 đơn hàng, với đỉnh cao nhất ở khoảng 9–10 đơn.
- Rất ít khách hàng đặt dưới 5 hoặc trên 15 đơn hàng.
- Cho thấy hành vi mua sắm ổn định, với phần lớn khách hàng có tần suất đặt hàng tương đối đều đặn.



**Hình 2. 2:** Phân phối số ngày giữa các lần mua hàng

**Nhận xét:**

- Biểu đồ lệch phải mạnh, phân phối theo dạng phân phối mũ.
- Phần lớn khoảng cách giữa các lần mua hàng là dưới 100 ngày, tập trung nhiều ở khoảng 0–50 ngày.
- Có một số khách hàng có khoảng cách rất dài giữa các lần mua (lên đến trên 800 ngày), nhưng rất hiếm.
- Điều này cho thấy có sự đa dạng trong hành vi mua sắm, một nhóm khách hàng quay lại mua rất thường xuyên, còn một số nhóm khác mua lại sau thời gian dài.

2.1.4. Phân tích vòng đời khách hàng

Đoạn code sau dùng để phân tích giá trị vòng đời khách hàng (**Customer Lifetime Value - CLV**) bằng cách tính tổng doanh thu mỗi khách hàng và giá trị trung bình của CLV. Cụ thể như sau:

```
# Tính tổng doanh thu mỗi khách hàng
customer_revenue = df.groupby('customer_id')['purchase_amount'].sum()

# Tính giá trị vòng đời khách hàng trung bình
average_clv = customer_revenue.mean()
print("Giá trị vòng đời khách hàng trung bình:", average_clv)
```

Trong đó:

- `groupby('customer_id')`:



- Nhóm dữ liệu theo `customer_id`, nghĩa là gộp tất cả các giao dịch của từng khách hàng lại với nhau.

- `['purchase_amount'].sum()`:

- Tính tổng giá trị cột `purchase_amount` cho từng khách hàng.
- **`purchase_amount`** là số tiền khách hàng chi trả trong mỗi đơn hàng.
- Kết quả là tổng doanh thu của từng khách hàng trong khoảng thời gian phân tích.

- **`customer_revenue`** lưu trữ doanh thu của từng khách hàng, với:

- **Chìa khóa (index)** là `customer_id`
- **Giá trị** là tổng doanh thu của khách hàng đó.

-`customer_revenue.mean()`:

- Tính giá trị trung bình của tổng doanh thu trên tất cả khách hàng.
- **Kết quả là giá trị vòng đời khách hàng trung bình (Average CLV).**

=>Điều này cho biết **trung bình mỗi khách hàng chi tiêu bao nhiêu** trong suốt vòng đời của họ (trong khoảng thời gian phân tích).

Kết quả chương trình:

 Giá trị vòng đời khách hàng trung bình: 2553.6943

#### 2.1.5. Trực quan hóa dữ liệu

Đoạn code sau dùng để vẽ biểu đồ phân phối tổng doanh thu của mỗi khách hàng bằng cách sử dụng **Matplotlib** và **Seaborn**. Cụ thể như sau:

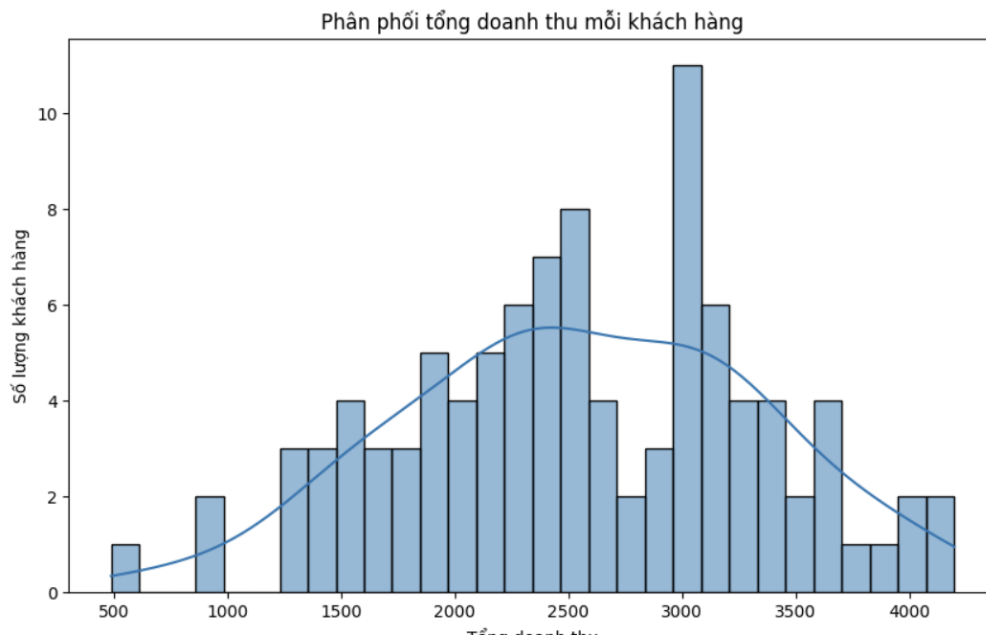
```
# vẽ biểu đồ tổng doanh thu mỗi khách hàng
plt.figure(figsize=(10, 6))
sns.histplot(customer_revenue, bins=30, kde=True)
plt.title('Phân phối tổng doanh thu mỗi khách hàng')
plt.xlabel('Tổng doanh thu')
plt.ylabel('Số lượng khách hàng')
plt.show()
```

Trong đó:

**`sns.histplot(...)`:**

- Vẽ biểu đồ **histogram** để hiển thị phân phối tổng doanh thu của mỗi khách hàng.
- Trục x là **tổng doanh thu** (tổng chi tiêu của từng khách hàng).
- Trục y là **số lượng khách hàng** có cùng mức doanh thu.

Kết quả chương trình:



Hình 2.3: Biểu đồ phân phối tổng doanh thu từ mỗi khách hàng

### Ý nghĩa của biểu đồ:

- Giúp hiểu được phân phối tổng doanh thu của khách hàng:
  - Phần lớn khách hàng chi tiêu bao nhiêu
  - Có nhóm khách hàng chi tiêu cao vượt trội (high-spenders)
  - Có khách hàng nào chi tiêu rất thấp không?
- Từ đó, có thể:
  - Phân khúc khách hàng theo mức chi tiêu.
  - Thiết kế chương trình khuyến mãi cho từng nhóm khách hàng.
  - Phát hiện khách hàng VIP để xây dựng chương trình chăm sóc đặc biệt.

#### 2.1.6. Phân tích khách hàng theo độ tuổi

##### a) Xử lý dữ liệu

```
# Bước 1: Thu thập dữ liệu
# Giả sử dữ liệu khách hàng được lưu trong một file CSV
df = pd.read_csv('/data/customer_data.csv')

# Bước 2: Tiền xử lý dữ liệu
# Kiểm tra dữ liệu thiếu
print(df.isnull().sum())

# Loại bỏ các dòng có dữ liệu thiếu
df = df.dropna()
```

Trong đó:

#### - `df.isnull()`:

- Trả về một DataFrame cùng kích thước với `df`, với giá trị True nếu ô đó bị thiếu (NaN), ngược lại là False.

- **.sum()**:

- Tính tổng số ô bị thiếu (NaN) trên từng cột.
- Kết quả là số lượng giá trị bị thiếu trên mỗi cột trong DataFrame.

- **df.dropna()**:

- Loại bỏ **các dòng** có bất kỳ giá trị nào bị thiếu (NaN).
- Kết quả là DataFrame df chỉ còn lại các dòng đầy đủ dữ liệu.

- **Ghi đè lại df** bằng phiên bản đã loại bỏ dữ liệu thiếu.

Kết quả chương trình:

```
customer_id    0
age            0
gender         0
annual_income  0
spending_score 0
purchase_amount 0
dtype: int64
```

b) *Phân tích khám phá*

```
# Bước 3: Phân tích khám phá
# Tổng quan về dữ liệu
print(df.describe())

# Vẽ biểu đồ phân phối độ tuổi
sns.histplot(df['age'], bins=30, kde=True)
plt.title('Phân phối độ tuổi khách hàng')
plt.xlabel('Độ tuổi')
plt.ylabel('Số lượng')
plt.show()
```

-**df.describe()**:

- Tóm tắt **các thống kê mô tả (Descriptive Statistics)** cho các cột số (numerical columns) trong DataFrame df, bao gồm:
  - **count**: Số lượng giá trị không bị thiếu.
  - **mean**: Giá trị trung bình.
  - **std**: Độ lệch chuẩn (Standard Deviation).
  - **min**: Giá trị nhỏ nhất.
  - **25%, 50%, 75%**: Các **quartile** (phân vị).
  - **max**: Giá trị lớn nhất.

**sns.histplot(...)**:

- Vẽ biểu đồ **histogram** cho cột age để hiển thị **phân phối độ tuổi** của khách hàng.
- Trục x là **độ tuổi**.

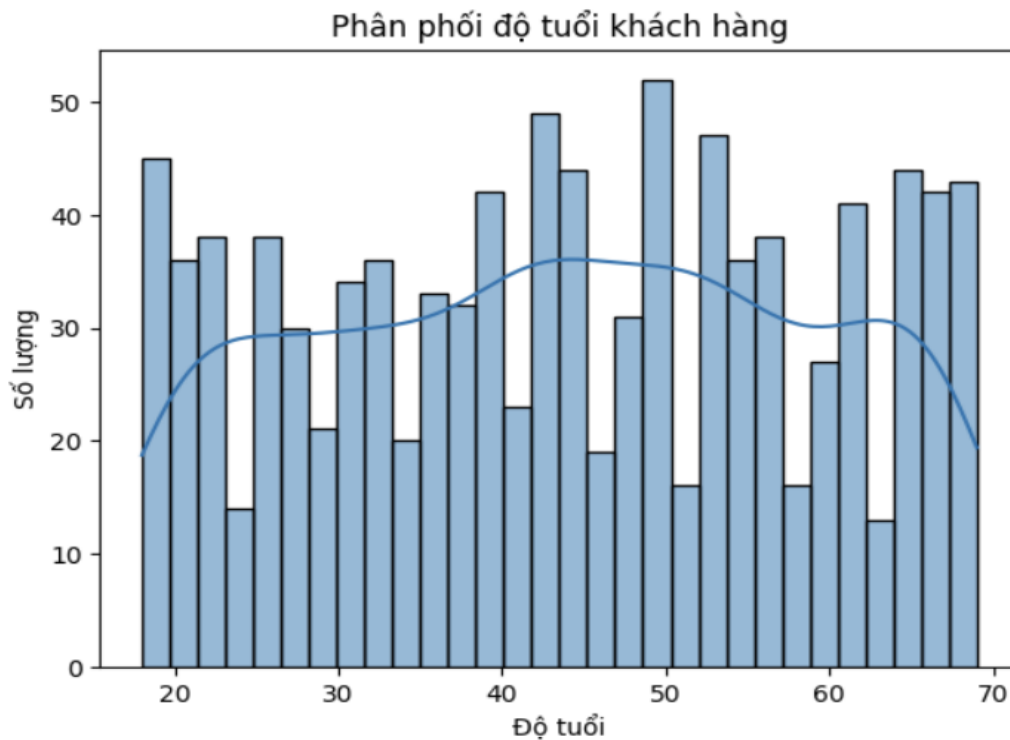
- Trục y là **số lượng khách hàng** trong từng nhóm tuổi.

Kết quả chương trình:

	customer_id	age	annual_income	spending_score	purchase_amount
count	1000.000000	1000.00000	1000.000000	1000.000000	1000.000000
mean	500.500000	43.81900	51203.985000	50.139000	150.438000
std	288.819436	14.99103	14898.526105	28.296897	50.400996
min	1.000000	18.00000	6179.000000	1.000000	-16.000000
25%	250.750000	31.00000	40975.750000	25.750000	115.000000
50%	500.500000	44.00000	51135.500000	51.000000	151.000000
75%	750.250000	56.00000	61131.000000	75.000000	186.000000
max	1000.000000	69.00000	97896.000000	99.000000	284.000000

Kết quả trên giúp ta có cái nhìn tổng quan về:

- Độ tuổi trung bình (mean) và phân phối (std, min, max) của khách hàng.
- Giá trị đơn hàng trung bình và phân phối.
- Phát hiện giá trị bất thường (outliers) nếu có sự chênh lệch lớn giữa mean và median (50%).



Hình 2.4: Phân phối độ tuổi khách hàng

**Ý nghĩa của biểu đồ:**

- Giúp hiểu được phân phối độ tuổi của khách hàng:
  - Nhóm tuổi nào chiếm đa số?
  - Độ tuổi trung bình của khách hàng?
  - Có sự chênh lệch lớn giữa các nhóm tuổi không?
- Thông tin này hữu ích cho:
  - Phân khúc khách hàng theo độ tuổi.

- Thiết kế chiến lược marketing nhắm đến từng nhóm tuổi khác nhau.
- Dự đoán nhu cầu sản phẩm dựa trên hành vi của từng nhóm tuổi.

*c) Phân tích phân khúc khách hàng*

```
# Bước 4: Phân tích phân khúc
# Giả sử chúng ta phân khúc khách hàng theo độ tuổi
df['age_group'] = pd.cut(df['age'], bins=[0, 18, 30, 40, 50, 60, np.inf], labels=['<18', '18-30', '30-40', '40-50', '50-60', '60+'])

# Đếm số lượng khách hàng trong mỗi nhóm tuổi
age_group_counts = df['age_group'].value_counts()
print(age_group_counts)
```

Kết quả của `pd.cut()` là một **cột mới age\_group** trong DataFrame `df`, chứa nhóm tuổi tương ứng cho từng khách hàng.

- `df['age_group'].value_counts()`:

- Đếm **số lượng khách hàng** trong từng nhóm tuổi (`age_group`).
  - Sắp xếp theo thứ tự giảm dần về số lượng.
- `age_group_counts` là một **Series** chứa số lượng khách hàng trong mỗi nhóm tuổi.

Kết quả chương trình:

```
age_group
40-50    218
18-30    214
60+      183
30-40    181
50-60    180
<18       24
Name: count, dtype: int64
Mean Squared Error: 2457.5022893280634
```

Nghĩa là có:

- **300 khách hàng** trong độ tuổi từ **18-30**.
- **80 khách hàng** từ **60 tuổi** trở lên.
- **5 khách hàng** dưới **18 tuổi**.

*d) Ứng dụng thực tế:*

- **Marketing và quảng cáo:**

- Nhắm mục tiêu quảng cáo theo độ tuổi (ví dụ: quảng cáo sản phẩm công nghệ cho nhóm 18-30).

- **Phân tích hành vi mua hàng:**

- Xem nhóm tuổi nào chi tiêu nhiều hơn để tập trung vào họ.

- **Phát triển sản phẩm:**

- Thiết kế sản phẩm phù hợp với nhu cầu và sở thích của từng nhóm tuổi.

*e) Dự đoán mô hình hóa*

```
# Bước 5: Dự đoán và mô hình hóa
# Ví dụ đơn giản với mô hình hồi quy tuyến tính để dự đoán doanh số bán hàng dựa trên độ tuổi
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

# Chọn các đặc trưng và biến mục tiêu
X = df[['age']]
y = df['purchase_amount']

# Chia dữ liệu thành tập huấn luyện và tập kiểm tra
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Huấn luyện mô hình
model = LinearRegression()
model.fit(X_train, y_train)

# Dự đoán trên tập kiểm tra
y_pred = model.predict(X_test)

# Đánh giá mô hình
from sklearn.metrics import mean_squared_error, r2_score
print('Mean Squared Error:', mean_squared_error(y_test, y_pred))
print('R2 Score:', r2_score(y_test, y_pred))
```

#### - Đặc trưng (Feature):

- X: Chọn cột age làm biến đầu vào (input) để dự đoán.

#### □ Biến mục tiêu (Target):

- y: Chọn cột purchase\_amount làm biến đầu ra (output), tức là dự đoán doanh số bán hàng dựa trên độ tuổi.

train\_test\_split(...):

#### - Chia dữ liệu thành hai phần:

- Tập huấn luyện (Training Set): 80% dữ liệu (test\_size=0.2 nghĩa là 20% cho tập kiểm tra).
- Tập kiểm tra (Test Set): 20% dữ liệu để đánh giá mô hình sau khi huấn luyện.

-random\_state=42:

- Đảm bảo kết quả phân chia dữ liệu giống nhau ở mỗi lần chạy (tái lập kết quả).

Kết quả chương trình:

```
Mean Squared Error: 2457.5022893280634
R2 Score: -0.004464565686965738
```

- **MSE = 2457.05**: Sai số trung bình giữa giá trị thực và dự đoán.

- **R<sup>2</sup> = 0.0044**: Mô hình giải thích được **4% sự biến động** của doanh số dựa trên độ tuổi.

#### Ý nghĩa và ứng dụng:

##### • Ý nghĩa:

- Mô hình này giúp dự đoán doanh số dựa trên độ tuổi của khách hàng.

- Nếu  $R^2$  cao, có thể sử dụng tuổi để dự đoán doanh số một cách đáng tin cậy.
- **Ứng dụng thực tế:**
  - Marketing: Tạo chiến lược quảng cáo cho nhóm tuổi có doanh số cao hơn.
  - Phân tích khách hàng: Hiểu rõ mối quan hệ giữa độ tuổi và hành vi mua hàng.
- f) Trực quan hóa dữ liệu

### Mục tiêu:

- So sánh trực quan giữa:
  - Giá trị thực tế (Actual): Doanh số bán hàng thực tế của tập kiểm tra ( $y_{test}$ ).
  - Giá trị dự đoán (Predicted): Doanh số bán hàng được dự đoán bởi mô hình ( $y_{pred}$ ).
- Giúp đánh giá hiệu suất của mô hình và xem xét mức độ chính xác của các dự đoán.

```
# Bước 6: Trực quan hóa dữ liệu
# Vẽ biểu đồ so sánh giá trị thực tế và giá trị dự đoán
plt.scatter(X_test, y_test, color='blue', label='Thực tế')
plt.scatter(X_test, y_pred, color='red', label='Dự đoán')
plt.title('So sánh giá trị thực tế và dự đoán')
plt.xlabel('Độ tuổi')
plt.ylabel('Doanh số bán hàng')
plt.legend()
plt.show()
```

\* **plt.scatter(...):** Vẽ biểu đồ **điểm (Scatter Plot)**, trong đó:

- Trục X: Độ tuổi ( $X_{test}$ ).
- Trục Y: Doanh số bán hàng ( $y_{test}$  và  $y_{pred}$ ).

\* **Hai tập dữ liệu được so sánh:**

#### 1. Điểm màu xanh (Thực tế):

- $X_{test}$  và  $y_{test}$ : Điểm dữ liệu thực tế từ tập kiểm tra.
- `color='blue'`: Màu xanh biểu thị giá trị thực tế.
- `label='Thực tế'`: Nhãn cho chú thích biểu đồ.

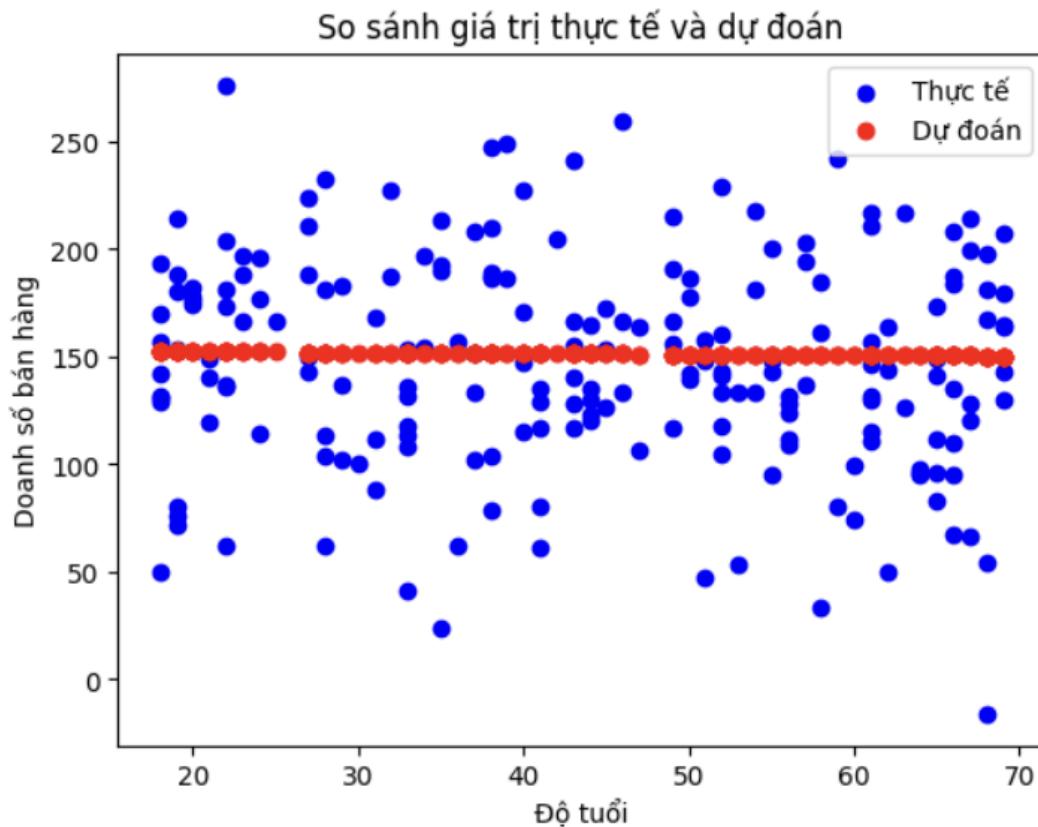
#### 2. Điểm màu đỏ (Dự đoán):

- $X_{test}$  và  $y_{pred}$ : Điểm dữ liệu dự đoán từ mô hình.
- `color='red'`: Màu đỏ biểu thị giá trị dự đoán.
- `label='Dự đoán'`: Nhãn cho chú thích biểu đồ.

### Giải thích:

- Biểu đồ so sánh trực quan giữa giá trị thực tế và giá trị dự đoán.
- Nếu mô hình dự đoán chính xác, các điểm màu đỏ (dự đoán) sẽ gần với các điểm màu xanh (thực tế).

Kết quả chương trình:



Hình 2.5: So sánh giá trị thực tế và giá trị dự đoán

### Ý nghĩa của biểu đồ:

- Biểu đồ này giúp bạn:
  - Đánh giá trực quan độ chính xác của mô hình.
  - Phát hiện lỗi hệ thống nếu mô hình dự đoán không tốt ở một nhóm tuổi cụ thể.
  - Hiểu rõ hơn mối quan hệ giữa độ tuổi và doanh số bán hàng.

### Cách phân tích biểu đồ:

- Nếu các điểm màu đỏ (Dự đoán) gần với các điểm màu xanh (Thực tế), nghĩa là mô hình dự đoán tốt.
- Nếu các điểm cách xa nhau hoặc có mẫu sai lệch rõ rệt, mô hình có thể:
  - Bị lệch (Bias) nếu luôn dự đoán thấp hơn hoặc cao hơn giá trị thực tế.
  - Bị phân tán (Variance) nếu dự đoán thiếu ổn định.



## 2.2. Khoa học dữ liệu và chiến lược Marketing hiệu quả

### 2.2.1. Khái niệm cơ bản về Khoa học dữ liệu và chiến lược Marketing hiệu quả

Khoa học dữ liệu là lĩnh vực nghiên cứu chuyên sâu về dữ liệu. Bằng cách sử dụng kỹ thuật và phương pháp từ thống kê, khoa học máy tính và kinh doanh để phân tích dữ liệu nhằm đưa ra các nhận định có ý nghĩa từ dữ liệu, mang lại lợi ích lớn cho các tổ chức và doanh nghiệp.

Marketing là tập hợp tất cả các hoạt động quảng bá, tiếp thị, thu hút và chăm sóc khách hàng, nhằm đưa sản phẩm hoặc dịch vụ từ nơi sản xuất đến tay người tiêu dùng và duy trì mối quan hệ với khách hàng.

Chiến lược marketing là tất cả những mục tiêu của doanh nghiệp. Được thực hiện hóa bằng: kế hoạch cụ thể, rõ ràng, nhằm mục tiêu tiếp thị sản phẩm đến với người tiêu dùng và phát triển thương hiệu của doanh nghiệp. Chiến lược marketing tập trung thúc đẩy sản phẩm dịch vụ để thu về mức lợi nhuận cao nhất. Có thể nói chiến lược marketing là một bản kế hoạch được đầu tư để đề ra ý tưởng và hành động quảng bá sản phẩm hoặc dịch vụ nhằm tạo ra lợi nhuận.

### 2.2.2. Các đặc trưng cơ bản của Khoa học dữ liệu và Marketing

#### a) Đặc trưng của khoa học dữ liệu

- Khối lượng (Volume): có kích thước, khối lượng lớn.
- Vận tốc (Velocity): tốc độ truyền dữ liệu đến luồng của dữ liệu nhanh chóng.
- Tính đa dạng (Variety): dữ liệu đa dạng về cấu trúc (có cấu trúc, bán cấu trúc, phi cấu trúc).
- Tính xác thực (Veracity): dữ liệu có sẵn đôi khi có sự nhầm lẫn, không chắc chắn về độ chính xác, tạo ra sự nhầm lẫn, thừa hoặc thiếu dữ liệu.
- Giá trị (Value): cần thu thập những dữ liệu hữu ích, loại bỏ những dữ liệu không có lợi cho công ty, doanh nghiệp. Có những dữ liệu không có giá trị sử dụng cần được chuyển đổi thành thứ có giá trị để trích xuất thông tin.

#### b) Đặc trưng của marketing

- Cần thiết (Needs): xuất phát từ nhu cầu cơ bản của con người.
- Mong muốn (Wants): được biểu hiện qua những thứ cụ thể có khả năng thỏa mãn nhu cầu bằng phương thức và nếp sống văn hóa xã hội.
- Nhu cầu (Demands): là những mong muốn có khả năng thanh toán, sở hữu được.
- Sản phẩm (Product): là thứ cần thiết cho mỗi quá trình marketing. Cần có sản phẩm, dịch vụ để trao đổi, mua bán và đáp ứng nhu cầu của khách hàng.
- Lợi ích (Benefit): sự thỏa mãn, đáp ứng nhu cầu của bản thân về sản phẩm, dịch vụ, giá cả, sự tin tưởng vào đơn vị mua và số lượng mua.

- Chi phí (Cost): chi phí thời gian, sức lực, tinh thần, giá cả nguyên vật liệu dùng để sản xuất...

- Sự thỏa mãn của khách hàng (Customers'satisfaction): marketing để thỏa mãn khách hàng mang lại lợi ích thực tế giúp giải quyết vấn đề cho khách hàng.

### 2.2.3. Tầm quan trọng của Khoa học dữ liệu trong Marketing

Khoa học dữ liệu trong marketing không chỉ đơn giản là thu thập và xử lý thông tin mà còn là quá trình hiểu rõ khách hàng của bạn, từ thông tin nhân khẩu học đến sở thích, hành vi, cảm xúc, mong muốn. Từ đó marketer có thể tạo ra thông điệp, chiến lược marketing hiệu quả và nhắm đúng đối tượng mục tiêu.

\* Chương trình trực quan hình ảnh doanh thu và tốc độ tăng trưởng B2C (Business To Consumer) qua các năm

*Bước 1: khai báo thư viện*

```
[ ] import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from matplotlib import cm
```

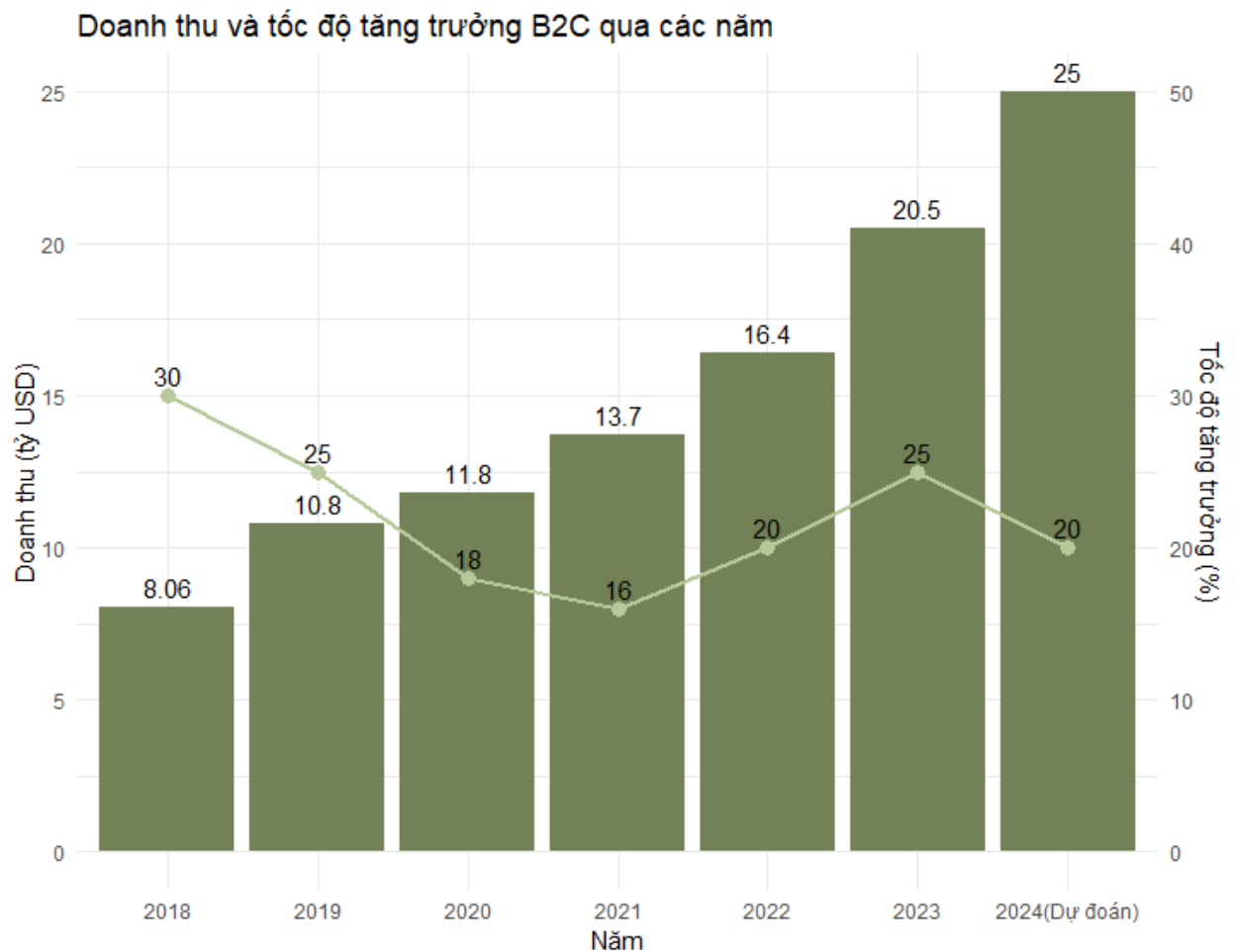
*Bước 2: khai báo dữ liệu*

```
# Dữ liệu1
Nam = ["2018", "2019", "2020", "2021", "2022", "2023", "2024(Dự đoán)"]
Doanhthu = [8.06, 10.8, 11.8, 13.7, 16.4, 20.5, 25]
Tangtruong = [30, 25, 18, 16, 20, 25, 20]
```

*Bước 3: vẽ biểu đồ*

```
# Tạo figure và axis
fig, ax1 = plt.subplots(figsize=(10, 6))
# Vẽ biểu đồ cột
ax1.bar(Nam, Doanhthu, color="#728156", label="Doanh thu (tỷ USD)")
# Gắn nhãn vào các cột
for i, v in enumerate(Doanhthu):
    ax1.text(i, v + 0.2, str(v), ha='center', color='black')
# Tạo trục y thứ hai cho Tốc độ tăng trưởng
ax2 = ax1.twinx()
ax2.plot(Nam, np.array(Tangtruong) / 2, color="#B6C99B", marker='o', label="Tốc độ tăng trưởng (%)", linewidth=2)
ax2.set_ylabel("Tốc độ tăng trưởng (%)", color="black")
# Gắn nhãn vào các điểm của đường
for i, v in enumerate(Tangtruong):
    ax2.text(i, v / 2 + 0.2, str(v), ha='center', color='black')
# Đặt nhãn cho trục y chính
ax1.set_ylabel("Doanh thu (tỷ USD)")
# Tiêu đề và nhãn trục x
plt.title("Doanh thu và tốc độ tăng trưởng B2C qua các năm")
ax1.set_xlabel("Năm")
# Hiện thị biểu đồ
plt.tight_layout()
plt.show()
```

Kết quả:



*Hình 2.6: Doanh thu và tốc độ tăng trưởng của thương mại điện tử qua các năm*

Nhờ có khoa học dữ liệu, doanh thu của ngành thương mại điện tử tăng nhanh với tốc độ vượt bậc. Từ năm 2018-2024 tăng khoảng 16.94 tỷ USD, thị trường đang ngày càng mở rộng mạnh mẽ.

#### 2.2.4. Phân khúc khách hàng trong Marketing

Khoa học dữ liệu giúp dễ dàng phân khúc khách hàng. Việc phân khúc khách hàng giúp doanh nghiệp hiểu rõ hơn về từng nhóm khách hàng cụ thể, từ đó thiết kế ra các chiến dịch tiếp thị phù hợp với từng nhóm giúp tăng hiệu quả tiếp thị và tối ưu hóa chi phí.

Ví dụ về phân khúc khách hàng của Nike

- Độ tuổi: 18 - 44
- Giới tính: nam và nữ
- Lịch sử mua hàng: mua sắm thường xuyên, giá cả đa dạng tùy theo mẫu mã
- Hành vi trên web: tìm kiếm những sản phẩm thể thao và thời trang mới có công nghệ tiên tiến, thiết kế độc đáo để thể hiện bản thân và tạo ra sự khác biệt

- Sở thích và phong cách: đam mê và có niềm yêu thích với thể thao, năng động và có lối sống lành mạnh

Hiểu rõ hành vi và sở thích của khách hàng là yếu tố then chốt giúp marketer tạo ra các chiến dịch marketing hiệu quả. Dựa trên dữ liệu thu thập từ các kênh trực tuyến và ngoại tuyến, các marketer có thể phân tích hành vi mua sắm của khách hàng như tần suất mua hàng, các sản phẩm ưa thích, thời gian mua sắm và các tương tác khác.

Bên cạnh đó, cá nhân hóa trải nghiệm khách hàng cũng là một việc quan trọng trong quá trình là marketing khi có thể tối ưu các hành động tiếp thị quảng cáo phù hợp với tương tác và trải nghiệm của khách hàng từ các dữ liệu cá nhân và hành vi của họ, giúp tăng cường sự hài lòng và lòng trung thành của khách hàng.

#### 2.2.5. Đo lường và đánh giá hiệu quả chiến dịch marketing (KPI)

KPI là số liệu quan trọng để có thể hiểu được tình hình hoạt động và thành công của các doanh nghiệp, là điều cần thiết để đạt được mục tiêu. KPI có thể cung cấp thông tin chi tiết về các khía cạnh khác nhau trong hoạt động kinh doanh (khả năng thu hút khách hàng, tỉ lệ chuyển đổi, doanh thu,...)

Khoa học dữ liệu giúp các marketer có thể dễ dàng đo lường và đánh giá hiệu quả chiến dịch marketing một cách nhanh chóng. Đặc biệt, khi chiến dịch được triển khai trên nhiều kênh với nhiều đối tượng khác nhau thì kỹ năng thu thập, mô hình hóa dữ liệu trong khoa học dữ liệu giúp tiết kiệm thời gian hơn rất nhiều để đưa ra những báo cáo chính xác.

Đối với các nhà quản lý, kỹ năng phân tích dữ liệu sẽ giúp ích rất nhiều trong việc đánh giá nhân sự và phân bổ KPI phù hợp. Với những dữ liệu khách hàng từ quá khứ, dữ liệu về xu hướng thị trường và hiệu suất performance (tiến trình đánh giá khả năng, kết quả của một hành động) trước đó của từng nhân viên để đặt ra được mức KPI khả thi cho cả nhóm và phù hợp với từng người để không bị xa vời với thực tế, thúc đẩy động lực cho từng cá nhân trong nhóm.

#### 2.2.6. Tối ưu hóa chiến dịch marketing

Khoa học dữ liệu giúp tối ưu hóa nội dung thông điệp và quảng cáo. Bằng cách theo dõi và phân tích các phản hồi từ khách hàng, có thể xác định được những yếu tố nào trong nội dung quảng cáo đang thu hút sự chú ý của người tiêu dùng, giúp marketer điều chỉnh nội dung phù hợp và hấp dẫn hơn. Một số công cụ hỗ trợ như Google Analytics và các công cụ SEO khác cũng hỗ trợ chúng ta xác định được những từ khóa và thông điệp nào đang mang lại hiệu quả cao nhất. Từ đó có thể điều chỉnh nội dung để phù hợp với nhu cầu và sở thích của khách hàng.

Để tối ưu hóa chiến dịch marketing thì định vị đúng đối tượng khách hàng là một trong những yếu tố quan trọng nhất. Khoa học dữ liệu cung cấp những thông tin

quý giá như nhân khẩu học, độ tuổi, giới tính, thu nhập, địa lý,...giúp chúng ta nhắm mục tiêu một cách chính xác và hiệu quả hơn.

Ví dụ, Thương hiệu Chanel nhắm đến đối tượng khách hàng mục tiêu là những phụ nữ muốn trở nên sang trọng và sành điệu nên ưu tiên cung cấp rất nhiều dòng sản phẩm thời trang, trang sức, mỹ phẩm cho nữ. Đồng thời kết hợp với 1 số gương mặt đại sứ thương hiệu được phái nữ yêu thích như Jennie, G-dragon,...

Ngoài 2 yếu tố trên, đưa ra quyết định phân bổ ngân sách và nguồn lực cũng là một phần quan trọng của marketing. Bằng cách sử dụng dữ liệu để đưa ra các quyết định này, chúng ta có thể đảm bảo rằng ngân sách của mình được phân bổ một cách hiệu quả, hợp lý để tăng doanh thu, giảm chi phí.

#### 2.2.7. Khoa học dữ liệu giúp dự đoán xu hướng kinh doanh

Khoa học dữ liệu không chỉ giúp marketer hiểu rõ về khách hàng hiện tại mà còn cung cấp những dự báo quan trọng về nhu cầu sản phẩm và xu hướng thị trường trong tương lai. Đây là một công cụ mạnh mẽ để giúp doanh nghiệp đón đầu xu hướng, tối ưu hóa chiến lược marketing và duy trì sự cạnh tranh.

Bằng cách phân tích dữ liệu bán hàng trong quá khứ, doanh nghiệp có thể xác định các mẫu hình và xu hướng theo mùa, từ đó dự báo nhu cầu sản phẩm trong tương lai. Bên cạnh đó, các thuật toán AI và Machine Learning có khả năng phân tích lượng dữ liệu khổng lồ và phức tạp để đưa ra dự đoán chính xác về nhu cầu sản phẩm. Những hệ thống này có thể lọc từ dữ liệu lịch sử và điều chỉnh các mô hình dự báo khi có thêm dữ liệu mới.

Nếu biết cách sử dụng, dữ liệu cũng có thể giúp phân đưa ra các chiến lược có căn cứ, tăng tỉ lệ thành công và giảm tỉ lệ rủi ro. Đồng thời xác định các xu hướng trong hành vi của khách hàng, hiệu suất marketing và các lĩnh vực quan trọng khác. Dựa vào các hiểu biết dựa trên dữ liệu để đưa ra quyết định, chúng ta có thể tin tưởng rằng những hành động của mình có khả năng cao nhất để thúc đẩy tăng trưởng và thành công cho doanh nghiệp.

#### 2.2.8. Ứng dụng của Khoa học dữ liệu trong Marketing

Với sự ra đời của Data Science trong kỷ nguyên kỹ thuật số, người làm marketing có thể am hiểu hơn về thị trường, hiểu sâu hơn về khách hàng mục tiêu.

Sức mạnh của Data Science giúp nhà tiếp thị tận dụng tối đa dữ liệu để cải thiện hiểu biết về thị trường, hiểu rõ hơn về khách hàng và quản lý tương tác của khách hàng qua nhiều nền tảng. Thông qua cách người người dùng tương tác với mọi người để có thể nắm bắt được insight của khách hàng, cung cấp dịch vụ tốt nhất cho người dùng.

Nhờ vào Data Science, các nhà tiếp thị có thể hiểu hơn về quy mô thị trường, khách hàng mục tiêu và cách thị trường vận hành, thay đổi.

##### 2.2.8.1. Ứng dụng trong Digital Marketing

Digital Marketing (marketing điện tử) là tất cả những hoạt động marketing diễn ra trên môi trường công nghệ số, nó tồn tại trong mọi công cụ của marketing mix. Hiểu đơn giản hơn, đây là cách quảng bá sản phẩm hoặc thương hiệu thông qua một hoặc nhiều hình thức truyền thông điện tử.

Trên nền tảng công nghệ số hiện nay, khoa học dữ liệu là bộ phận không thể thiếu để có một chiến lược marketing thành công. Có thể nói dữ liệu chính là mạch máu chính của marketing nói chung và digital marketing nói riêng bởi nhiều ứng dụng hữu ích của nó.

\*Chương trình trực quan hình ảnh thị phần giá trị giao dịch của các sàn TMĐT quý 2/2024

*Bước 1: khai báo thư viện*

```
[ ] import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from matplotlib import cm
```

*Bước 2: khai báo dữ liệu*

```
#Dữ liệu2
data = {
    "Ten": ["Shopee", "TikTok Shop", "Lazada", "Tiki"],
    "Gtri": [71.4, 22.0, 5.9, 0.7],
    "Sluong": [62.38, 19.24, 5.16, 0.58477],
}
df = pd.DataFrame(data)
```

*Bước 3: vẽ biểu đồ*

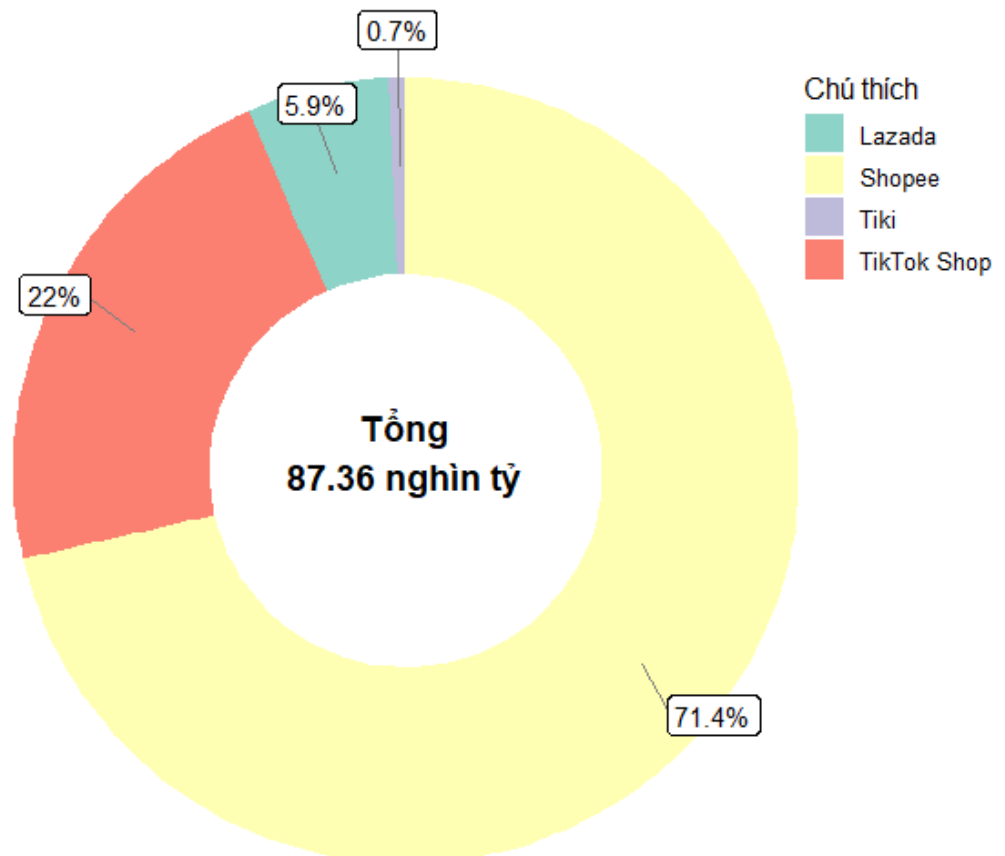
```

# Tính phần trăm
df['phanso'] = df['Gtri'] / df['Gtri'].sum()
# Tổng giá trị
total_value = round(df['Sluong'].sum(), 2)
# Vẽ biểu đồ donut
fig, ax = plt.subplots(figsize=(10, 10))
# Tạo biểu đồ tròn
wedges, texts, autotexts = ax.pie(df['Gtri'], labels=df['Ten'], autopct='', startangle=90,
                                   colors=cm.Set3(np.linspace(0, 1, len(df))), wedgeprops=dict(width=0.3))
ax.legend([f"{name}: {value} nghìn tỷ đồng" for name, value in zip(df['Ten'], df['Sluong'])], loc="center left",
          bbox_to_anchor=(1, 0.5), fontsize=12)
plt.subplots_adjust(right=0.8)
# Thêm nhãn tổng giá trị vào giữa donut
ax.text(0, 0, f"Tổng\n{total_value} nghìn tỷ", ha='center', va='center', fontsize=14, fontweight='bold', color='black')
# Thêm đường nối giữa phần trăm và các phần trong biểu đồ
for i, wedge in enumerate(wedges):
    # Lấy vị trí của wedge
    angle = (wedge.theta2 + wedge.theta1) / 2
    x = np.cos(np.radians(angle))
    y = np.sin(np.radians(angle))
    # Vẽ đường nối từ phần trăm tới wedge
    ax.annotate(f'{df["phanso"][i]*100:.1f}%', xy=(x, y), xytext=(1.2*x, 1.2*y),
               arrowprops=dict(facecolor='black', shrink=0.05, width=1, headwidth=6),
               ha='center', va='center', fontsize=12, color='black')
# Cài đặt các tham số cho biểu đồ
ax.set_title("Thị phần giá trị giao dịch của các sàn TMĐT, Quý 2/2024", fontsize=16, fontweight='bold', ha='center')
# Hiển thị biểu đồ
plt.tight_layout()
plt.show()

```

Kết quả:

### Thị phần giá trị giao dịch của các sàn TMĐT, Quý 2/2024



Hình 2.7: Thị phần giá trị giao dịch của các sàn thương mại điện tử quý 2/2024

2.2.8.2. Dấu chân điện tử (Digital Footprint)

Dấu chân điện tử là tất cả những dữ liệu mà người dùng để lại khi hoạt động trực tuyến, từ việc tìm kiếm trên Google, tương tác trên mạng xã hội cho đến việc thực hiện giao dịch mua sắm trực tuyến. Những thông tin này có thể giúp người làm marketing có thể hiểu được sở thích, đặc điểm người dùng cũng như tính cách hay tâm trạng của họ.

Khoa học dữ liệu có thể phân tích các loại dấu chân điện tử để hiểu rõ hơn về hành vi và sở thích của khách hàng. Ví dụ, hành vi duyệt web, lịch sử mua sắm hoặc tương tác trên mạng xã hội có thể cung cấp thông tin giá trị về những gì khách hàng quan tâm. Bằng cách này, các nhà tiếp thị có thể chia khách hàng thành các nhóm dựa trên thói quen mua sắm, sở thích và hành vi tiêu dùng, từ đó xây dựng các chiến dịch marketing nhắm tới từng nhóm cụ thể.

Bên cạnh đó, các mô hình học máy (machine learning) có thể được áp dụng để dự đoán các xu hướng và hành vi trong tương lai dựa trên dữ liệu lịch sử mà khách hàng đã để lại. Điều này giúp các doanh nghiệp chuẩn bị và đáp ứng nhu cầu một cách kịp thời, có thể chạy các chiến dịch marketing dự đoán một cách tự động. Đồng thời, các dấu chân điện tử cho phép các nhà tiếp thị tối ưu hóa quảng cáo trực tuyến để nhắm đúng đối tượng mục tiêu, giúp doanh nghiệp hiểu rõ ai là người có khả năng mua hàng cao nhất, từ đó điều chỉnh ngân sách và thông điệp quảng cáo phù hợp.

Dấu chân điện tử không chỉ là một nguồn tài nguyên phong phú về hành vi và xu hướng của người tiêu dùng mà còn là một phần quan trọng trong chiến lược digital marketing. Việc áp dụng khoa học dữ liệu để phân tích và sử dụng thông tin từ dấu chân điện tử giúp các doanh nghiệp tạo ra các chiến dịch tiếp thị hiệu quả, cá nhân hóa trải nghiệm khách hàng, tối ưu hóa quảng cáo và cuối cùng là nâng cao hiệu suất kinh doanh.

#### 2.2.8.3. Phân tích ngôn ngữ tự nhiên (Natural Language Processing – NLP)

NLP là một lĩnh vực của khoa học máy tính và trí tuệ nhân tạo, tập trung vào việc giúp máy tính hiểu phân tích và xử lý ngôn ngữ tự nhiên của con người. NLP cho phép các hệ thống máy tính tương tác với con người thông qua ngôn ngữ, nghĩa là các câu, từ và văn bản mà con người sử dụng hàng ngày.

Trong ngành digital marketing, nó có nhiều ứng dụng đáng kể hỗ trợ doanh nghiệp hiểu và tương tác tốt hơn với khách hàng. Thông qua việc phân tích cảm xúc, doanh nghiệp có thể xác định cảm xúc tích cực, tiêu cực hoặc trung tính của khách hàng đối với thương hiệu hoặc sản phẩm, từ đó nhanh chóng phát hiện và xử lý các vấn đề, cải thiện hình ảnh thương hiệu và tăng cường lòng trung thành của khách hàng. Bên cạnh đó, NLP giúp cá nhân hóa nội dung bằng cách tạo ra những thông điệp phù hợp với sở thích và nhu cầu của từng khách hàng, đồng thời tối ưu hóa SEO thông qua phân tích từ khóa và xu hướng tìm kiếm của người tiêu dùng. Một số công cụ NLP có khả năng tạo ra nội dung tự động, rà soát và cải thiện văn phong, ngữ pháp, cấu trúc của nội dung trước khi được công bố, đảm bảo rằng thông điệp truyền



đến khách hàng một cách rõ ràng và hiệu quả. Cuối cùng, NLP cung cấp thông tin để đánh giá và cải tiến các chiến dịch marketing dựa trên dữ liệu lịch sử, hỗ trợ doanh nghiệp phát triển các mô hình và xu hướng từ các chiến dịch trước đó và tối ưu hóa các chiến dịch mới trong tương lai.

Phân tích ngôn ngữ tự động là một công cụ mạnh mẽ trong ngành digital marketing giúp doanh nghiệp tương tác hiệu quả hơn với khách hàng, tối ưu hóa nội dung và nâng cao trải nghiệm người dùng. Bằng cách sử dụng các kỹ thuật NLP, doanh nghiệp có thể thu thập, phân tích dữ liệu một cách linh hoạt, đưa ra những quyết định thông minh hơn và có chiến lược marketing hiệu quả hơn.

#### *2.2.8.4. Hệ thống theo dõi nhãn cầu (eye-tracking)*

Hệ thống theo dõi nhãn cầu là một công nghệ tiên tiến giúp theo dõi và ghi nhận chuyển động của mắt người dùng khi họ tương tác với một trang web, quảng cáo hay sản phẩm. Khi được kết hợp với khoa học dữ liệu, ứng dụng của hệ thống theo dõi nhãn cầu trong ngành digital marketing có thể tạo ra nhiều giá trị thiết thực và thông tin quý giá cho các chiến dịch truyền thông.

Trước tiên, nó giúp tối ưu hóa trải nghiệm người dùng (UX) bằng cách cung cấp dữ liệu về các điểm nhìn, thời gian nhìn vào từng phần của trang và cách người dùng tương tác với nội dung. Qua đó, các nhà phân tích có thể xác định được những yếu tố thu hút sự chú ý để điều chỉnh thiết kế giao diện thân thiện hơn. Thêm vào đó, việc phân tích hành vi người tiêu dùng giúp marketers hiểu rõ hơn về mẫu hành vi tiêu thụ nội dung, từ đó tối ưu hóa các chiến dịch quảng cáo và nội dung trang đích. Hệ thống theo dõi nhãn cầu cũng hỗ trợ cải thiện hiệu suất của quảng cáo bằng cách xác định các yếu tố (màu sắc, hình ảnh, văn bản) hiệu quả trong việc thu hút sự chú ý và tạo nên cảm xúc tích cực từ người tiêu dùng. Đồng thời, hệ thống eye-tracking còn được áp dụng để nghiên cứu phản hồi từ khách hàng về sản phẩm mới, hỗ trợ các nhà sản xuất hiểu rõ hơn về cách người tiêu dùng nhìn nhận những điểm mạnh, yếu trong thiết kế sản phẩm để cải tiến trước khi ra mắt thị trường.

Hệ thống theo dõi nhãn cầu là một công cụ mạnh mẽ khi được tích hợp với khoa học dữ liệu trong digital marketing. Nó không chỉ hỗ trợ trong việc tối ưu hóa trải nghiệm người dùng mà còn cung cấp những insight quý giá về hành vi của người tiêu dùng, từ đó giúp marketers phát triển các chiến lược hiệu quả hơn trong việc thu hút và giữ chân khách hàng, thúc đẩy tỷ lệ chuyển đổi và gia tăng doanh thu cho các doanh nghiệp.

#### *2.2.8.5. Phân chia dữ liệu (clustering)*

Phân chia dữ liệu là một kỹ thuật trong khoa học dữ liệu nhóm các đối tượng tương tự lại với nhau dựa trên các đặc điểm hoặc thuộc tính của chúng. Trong ngành digital marketing, việc phân chia dữ liệu có vai trò cực kỳ quan trọng và ứng dụng đa dạng, từ việc hiểu rõ hành vi khách hàng cho đến tối ưu hóa chiến dịch marketing.

Một trong những ứng dụng quan trọng nhất của việc phân chia dữ liệu trong digital marketing là phân đoạn thị trường. Bằng cách sử dụng các thuật toán phân cụm, doanh nghiệp có thể chia nhỏ đối tượng khách hàng thành các nhóm dựa trên các yếu tố như độ tuổi, giới tính, sở thích, hành vi tiêu dùng, hoặc tình trạng tài chính. Điều này giúp marketers tập trung vào từng phân đoạn với thông điệp và chiến lược truyền thông phù hợp, từ đó nâng cao hiệu quả của chiến dịch marketing.

Bên cạnh đó, phân chia dữ liệu giúp marketers tối ưu hóa ngân sách quảng cáo bằng cách xác định các phân khúc khách hàng có khả năng chuyển đổi cao nhất. Sử dụng có hiệu quả dữ liệu từ các chiến dịch quảng cáo trước đó, các nhà quảng cáo có thể phát hiện được những nhóm khách hàng phản hồi tốt nhất đối với những thông điệp cụ thể để có thể phân bổ ngân sách quảng cáo một cách hiệu quả hơn, hướng tới những phân khúc mang lại ROI (tỷ lệ hoàn vốn đầu tư) tốt nhất. Không chỉ vậy, nó còn hỗ trợ doanh nghiệp trong quy trình phát triển sản phẩm mới phù hợp với nhu cầu thực tế của thị trường.

Phân chia dữ liệu là một công cụ mạnh mẽ trong ngành digital marketing, giúp doanh nghiệp tối ưu hóa hoạt động tiếp thị của mình và nâng cao trải nghiệm khách hàng. Áp dụng kỹ thuật này, marketers có thể phân đoạn thị trường, tối ưu hóa ngân sách quảng cáo, cá nhân hóa trải nghiệm khách hàng và dự đoán hành vi tiêu dùng. Từ đó đưa ra những quyết định thông minh, tối ưu hóa hiệu quả của các chiến dịch marketing, gia tăng doanh thu và cạnh tranh tốt hơn trong thị trường ngày càng khốc liệt.

#### *2.2.8.6. Thu thập dữ liệu chính xác*

Việc thu thập dữ liệu chính xác là một yếu tố then chốt trong khoa học dữ liệu và có tác động sâu rộng đến ngành digital marketing. Trong bối cảnh cạnh tranh gay gắt, doanh nghiệp không chỉ cần thông tin lớn mà còn cần thông tin chính xác để đưa ra các quyết định đúng đắn trong chiến lược marketing.

Thu thập dữ liệu chính xác cho phép marketers có cái nhìn sâu sắc về hành vi và sở thích của khách hàng. Nếu phân tích tốt dữ liệu từ nhiều nguồn như mạng xã hội, trang web, email và các dữ liệu giao dịch, doanh nghiệp có thể xác định được mẫu hành vi tiêu dùng nhằm điều chỉnh chiến lược marketing phù hợp. Hơn nữa, dữ liệu chính xác còn hỗ trợ cá nhân hóa trải nghiệm khách hàng, từ việc cung cấp nội dung và sản phẩm gợi ý dựa trên sở thích cá nhân, cho đến tối ưu hóa chiến dịch quảng cáo bằng cách xác định đối tượng mục tiêu một cách rõ ràng. Ngoài ra, khả năng dự đoán xu hướng và hành vi tương lai nhờ vào các mô hình phân tích dự đoán giúp doanh nghiệp nắm bắt cơ hội và điều chỉnh kịp thời. Việc tổ chức đánh giá hiệu quả của các chiến dịch marketing cũng trở nên dễ dàng hơn với dữ liệu chính xác, từ đó giúp tối ưu hóa quy trình ra quyết định và cải thiện hiệu suất trong chiến dịch quảng cáo cũng như nội dung.

Thu thập dữ liệu chính xác là một yếu tố thiết yếu trong digital marketing, giúp doanh nghiệp hiểu rõ hơn về khách hàng, cá nhân hóa trải nghiệm, tối ưu hóa quảng cáo và dự đoán hành vi tương lai. Khi dữ liệu chính xác được thu thập và phân tích, doanh nghiệp có thể đưa ra các quyết định thông minh hơn, nâng cao hiệu quả các chiến dịch marketing và cải thiện vị thế cạnh tranh trong thị trường.

Dù áp dụng theo cách nào thì Data Science đã trở thành trợ thủ đắc lực cho những người làm Digital Marketing, cung cấp những công cụ và phương pháp mạnh mẽ giúp các doanh nghiệp nắm bắt, phân tích và tận dụng dữ liệu để tối ưu hóa chiến lược tiếp thị. Bằng một loạt ứng dụng trên, doanh nghiệp không chỉ hiểu và phục vụ khách hàng tốt hơn mà còn nâng cao khả năng ra quyết định chính xác và chiến lược dài hạn, đảm bảo sự phát triển và thành công trong một thị trường ngày càng cạnh tranh.

#### 2.2.8.7. Ứng dụng trong Affiliate Marketing

Affiliate Marketing (tiếp thị liên kết) là một mô hình quảng bá sản phẩm, dịch vụ. Trong đó, các nhà phân phối dựa trên năng lực, sự hiểu biết của mình sẽ thực hiện tiếp thị, thu hút khách hàng mua hàng, sử dụng dịch vụ của nhà cung cấp, từ đó nhà phân phối sẽ nhận được hoa hồng từ nhà cung cấp với mỗi đơn hàng thành công hoặc hoàn thành các hành động cụ thể.

Affiliate Marketing là ngành luôn có những sự thay đổi không ngừng trước, yêu cầu các nhà quảng cáo phải luôn đổi mới với thách thức mới. Có nghĩa là chúng ta phải liên tục tìm kiếm các cơ hội và phương pháp mới để tiếp cận nhiều khách hàng hơn cũng như tạo ra nhiều lợi ích hơn. Và khoa học dữ liệu là một công cụ đắc lực để có thể giải quyết được điều đó nhờ những ưu điểm vượt trội của nó.

#### 2.2.8.8. Phân tích để tăng tỷ suất hoàn vốn (ROI)

ROI là một chỉ số tài chính dùng để đo lường hiệu quả của một khoản đầu tư. Nó cho biết mức độ lợi nhuận thu được từ một khoản đầu tư so với chi phí của khoản đầu tư đó, thường được tính toán để so sánh hiệu suất của các khoản đầu tư khác nhau hoặc để đánh giá tính khả thi của một dự án, một chiến dịch marketing, hoặc một hoạt động kinh doanh.

Công thức tính ROI: 
$$ROI = \left( \frac{\text{Lợi nhuận ròng}}{\text{Chi phí đầu tư}} \right) \times 100\%$$

Trong đó:

- Lợi nhuận ròng: tổng doanh thu thu được từ khoản đầu tư trừ đi tổng chi phí đầu tư.
- Chi phí đầu tư: tổng chi phí bỏ ra để thực hiện khoản đầu tư đó.

Việc thu thập và phân tích dữ liệu liên quan đến hành vi người dùng, từ nguồn gốc truy cập đến hành vi mua sắm, cho phép các nhà tiếp thị xác định rõ đối tượng mục tiêu và hiểu rõ hành động của họ trên nền tảng. Sử dụng các công cụ phân tích

như Google Analytics hay các phần mềm chuyên dụng, marketers có thể đo lường hiệu quả của từng kênh tiếp thị, từ đó phát hiện ra các nguồn mang lại tỷ suất hoàn vốn cao nhất. Hơn nữa, dữ liệu cũng giúp xác định các sản phẩm hoặc dịch vụ nào đang thu hút sự quan tâm của khách hàng, cho phép các đối tác liên kết điều chỉnh nội dung và chiến lược quảng bá theo hướng phù hợp hơn. Ngoài ra, khoa học dữ liệu còn hỗ trợ trong việc tối ưu hóa các chiến dịch quảng cáo theo thời gian thực, giúp tự động điều chỉnh ngân sách và nguồn lực cho các kênh mang lại hiệu quả tốt nhất. Đồng thời, phân tích dữ liệu không chỉ giúp tối ưu hóa các chiến dịch ngay lập tức mà còn hỗ trợ xây dựng chiến lược dài hạn dựa trên xu hướng và sở thích của người tiêu dùng.

Lĩnh vực chuyên môn đóng vai trò quan trọng trong bất kỳ hoạt động kinh doanh nào. Các affiliate marketer quảng bá những sản phẩm đúng lĩnh vực chuyên môn của mình thường nghiêm túc hơn và khó để có thể cạnh tranh với họ. Nếu không có khoa học dữ liệu và phân loại thông tin ban đầu theo một phương pháp riêng thì rất khó để đánh bại các đối thủ.

Các affiliate marketer thường sẽ quảng bá trên trang nền tảng của mình để mức chi phí cho mỗi khách hàng tiềm năng sẽ ít hơn. Hơn nữa, các affiliate marketer sẽ rất ít khi thay đổi. Do đó, việc lựa chọn đúng chuyên môn là điều cần thiết để mang lại hiệu quả trong bất kỳ hoạt động kinh doanh nào và tăng ROI. Ví dụ, Võ Hà Linh – tiktoker nổi tiếng trong lĩnh vực chăm sóc sắc đẹp. Cô chủ yếu nhận affiliate marketing các sản phẩm về chăm sóc da, mỹ phẩm đáng sử dụng phù hợp với mọi đối tượng và mọi mức giá hợp lý. Với kiến thức chuyên môn cao về hóa học, làm đẹp,...cô dần lấy được lòng tin của khách hàng, trở thành “chiến thần review” có lượng khách hàng đông đảo bậc nhất hiện nay trên nền tảng Tik Tok.

Trong ngành Affiliate Marketing, khoa học dữ liệu không chỉ giúp nâng cao hiểu biết về khách hàng mà còn tối ưu hóa hiệu suất của các chiến dịch quảng cáo và nội dung, từ đó góp phần quan trọng vào việc tăng tỷ suất hoàn vốn (ROI). Ứng dụng hiệu quả các phân tích dữ liệu sẽ giúp các marketers đưa ra các quyết định thông minh hơn và đạt được kết quả tốt hơn trong các chiến dịch của mình.

#### *2.2.8.9. Khám phá những nguồn lực giúp nâng cấp chiến dịch*

Nhờ khả năng phân tích và khai thác dữ liệu từ nhiều nguồn khác nhau, khoa học dữ liệu đã trở thành một công cụ thiết yếu trong việc nâng cấp chiến dịch Affiliate marketing,

Khoa học dữ liệu cho phép thu thập và phân tích khối lượng lớn dữ liệu từ hành vi người tiêu dùng trên các nền tảng trực tuyến. Thông qua việc phân tích các chỉ số như tỷ lệ nhấp chuột (CTR), thời gian truy cập và mức độ tương tác của người dùng với các liên kết affiliate, marketers có thể xác định xu hướng tiêu dùng và điều chỉnh nội dung quảng cáo cho phù hợp. Điều này không chỉ làm tăng tỷ lệ chuyển đổi mà còn giúp tối ưu hóa trải nghiệm người dùng.

Một trong những yếu tố quan trọng trong Affiliate marketing là việc chọn lựa và quản lý các đối tác liên kết. Khoa học dữ liệu có thể giúp phân tích hiệu suất của từng đối tác dựa trên các chỉ số như doanh thu tạo ra, tỷ lệ chuyển đổi và chi phí. Từ đó, các marketers có thể xác định những đối tác hiệu quả nhất và điều chỉnh ngân sách cho các chiến dịch dựa trên hiệu suất thực tế, đồng thời loại bỏ những liên kết không mang lại giá trị.

Affiliate marketer sử dụng các quảng cáo tốn phí để tiếp cận những khách hàng tiềm năng và phân biệt các yếu tố có thể ảnh hưởng nhiều nhất đến mục tiêu và hiển thị của chiến dịch. Đồng thời cũng sử dụng các cuộc khảo sát thăm dò để theo dõi kết quả của các chiến dịch khác nhau nhằm xây dựng chiến dịch ưu việt nhất.

Việc áp dụng khoa học dữ liệu trong Affiliate marketing không chỉ giúp doanh nghiệp hiểu rõ hơn về khách hàng mà còn tăng cường khả năng tối ưu hóa các chiến dịch quảng cáo. Những nguồn lực từ khoa học dữ liệu không chỉ tạo ra lợi thế cạnh tranh mà còn đảm bảo rằng các chiến dịch tiếp cận đúng đối tượng với thông điệp phù hợp, từ đó gia tăng hiệu quả kinh doanh và sự phát triển bền vững.

#### *2.2.8.10. Tìm kiếm cơ hội tiềm ẩn trong các lĩnh vực chưa được khám phá*

Khoa học dữ liệu đang mở ra nhiều cơ hội tiềm ẩn trong lĩnh vực Affiliate marketing, đặc biệt là trong các lĩnh vực chưa được khám phá hoặc chưa được tận dụng tối đa.

Với việc khai thác dữ liệu từ các nguồn khác nhau như mạng xã hội, diễn đàn trực tuyến và các nền tảng thương mại điện tử có thể mang lại cái nhìn sâu sắc về xu hướng và nhu cầu của người tiêu dùng. Từ những thông tin này, các doanh nghiệp affiliate có thể phát hiện ra những ngách thị trường hoặc sản phẩm mới mà đối thủ chưa chú ý đến, từ đó phát triển các chiến dịch mục tiêu, hấp dẫn hơn.

Một lĩnh vực khác là khám phá các hợp tác mới với các nhà cung cấp hoặc platform chưa được tiếp cận, chẳng hạn như các influencer trên các nền tảng mới nổi như TikTok hay Clubhouse. Việc phân tích hiệu quả của các kênh tiếp thị này có thể giúp các doanh nghiệp tìm ra những nguồn lực tập trung có thể thúc đẩy doanh thu.

Phần lớn các affiliate marketer tập trung vào việc thúc đẩy các đề nghị hấp dẫn cho các cá nhân ở một hoặc nhiều khu vực khác nhau. Affiliate marketing lại đang phát triển nhanh. Theo báo cáo, việc kinh doanh dựa trên web tại châu Á có giá trị tới hơn 1000 tỷ.

Khoa học dữ liệu có thể khám phá các câu khẩu hiệu phù hợp, tập trung vào chúng thông qua các giai đoạn SEO (Search Engine Optimization), PPC (Pay Per Click) và các phương tiện truyền thông trực tuyến để thu hút nhiều khách hàng hơn.

Với việc tìm kiếm dữ liệu, có thể hình dung được việc thu thập dữ liệu về các lĩnh vực khác nhau nơi quảng cáo của bạn đang chạy. Từ đó thực hiện một số thay đổi nhỏ và tập trung vào những lĩnh vực vẫn bị bỏ quên lâu nay.

Ngành Affiliate Marketing có tiềm năng lớn từ ứng dụng khoa học dữ liệu để khám phá và khai thác những cơ hội chưa được nhận diện, từ đó nâng cao hiệu quả tiếp thị và tối ưu hóa lợi nhuận.

#### *2.2.8.11. Chọn sản phẩm phù hợp muốn quảng bá*

Khoa học dữ liệu đóng một vai trò quan trọng trong việc chọn sản phẩm phù hợp để quảng bá trong ngành affiliate marketing, giúp tối ưu hóa hiệu suất và tăng cường tỷ lệ chuyển đổi.

Đầu tiên, phân tích dữ liệu giúp các nhà tiếp thị nắm bắt xu hướng tiêu dùng và sở thích của khách hàng thông qua việc khai thác các nguồn dữ liệu lớn từ mạng xã hội, công cụ tìm kiếm, và các nền tảng thương mại điện tử. Với các công cụ phân tích, marketers có thể xác định những sản phẩm đang được tìm kiếm nhiều nhất, từ đó quyết định sản phẩm nào có tiềm năng đem lại doanh thu cao nhất.

Ngoài ra, việc sử dụng phân tích hành vi người tiêu dùng cho phép các nhà tiếp thị hiểu rõ hơn về các yếu tố ảnh hưởng đến quyết định mua hàng. Các chỉ số như tỷ lệ nhấp chuột (CTR), thời gian tại trang và tỷ lệ thoát sẽ giúp xác định sản phẩm nào thu hút sự chú ý của khách hàng nhất. Người tiếp thị cũng có thể sử dụng kỹ thuật phân tích hồi quy để xác định mối quan hệ giữa các yếu tố cụ thể (như giá cả, mùa vụ hay đặc tính sản phẩm) và quyết định mua hàng của khách hàng, từ đó cải thiện lựa chọn sản phẩm.

Thêm vào đó, việc áp dụng mô hình phân đoạn thị trường thông qua phân tích dữ liệu khách hàng cho phép nhà tiếp thị nhắm đến các nhóm khách hàng cụ thể với sở thích và nhu cầu riêng. Điều này không chỉ giúp chọn sản phẩm phù hợp mà còn điều chỉnh thông điệp quảng cáo sao cho hấp dẫn đối với từng phân khúc. Cuối cùng, phân tích khả năng cạnh tranh cũng hỗ trợ marketers trong việc nhận diện sản phẩm nào được các đối thủ cạnh tranh quảng bá thành công, từ đó tìm ra những sản phẩm có thể chiếm lĩnh thị trường hoặc phát triển các sản phẩm tương tự với lợi thế khác biệt.

Chọn đúng chuyên ngành là cơ sở để đạt được thành công trong affiliate marketing. Chúng ta có thể quảng bá các sản phẩm tùy theo phong cách và sở thích của mình, có thể về thời trang, trang sức, thiết bị điện tử,... Đến thời điểm thích hợp, khi khách hàng bắt gặp bài viết giới thiệu những sản phẩm đúng nhu cầu sẽ quyết định mua.

#### *2.2.8.12. Thiết kế landing page dành riêng cho từng sản phẩm*

Landing page (trang đích) là một trang web độc lập, thường được thiết kế với một mục tiêu cụ thể để thu hút và hướng người truy cập thực hiện một hành động nhất định, chẳng hạn như đăng ký nhận bản tin, tải xuống tài liệu, mua hàng, hoặc điền vào biểu mẫu thông tin. Nó thường được sử dụng trong các chiến dịch marketing trực tuyến với mục đích chuyển đổi lượt truy cập thành khách hàng hoặc khách hàng tiềm năng

Dựa vào dữ liệu về các lời khuyên hay nhận xét từ cộng đồng, marketer có thể lập kế hoạch tạo các trang landing page khác nhau cho mọi mặt hàng và cung cấp cho họ ưu đãi hấp dẫn, điều này có thể nâng cao khả năng thu hút khách hàng. Đồng thời cũng thể hiện sự nghiêm túc trong việc thuyết phục khách hàng mua hàng khi nhìn thấy thông báo hoặc truy cập và trang landing page của một mặt hàng cụ thể.

Việc thu thập và phân tích dữ liệu từ hành vi người dùng, bao gồm lượt truy cập, thời gian trên trang, tỷ lệ thoát và các mẫu hành vi của khách hàng, cho phép các nhà tiếp thị xác định những yếu tố quan trọng nhất mà người dùng quan tâm. Từ đó, họ có thể điều chỉnh nội dung, hình ảnh, và bố cục của landing page để phù hợp với từng đối tượng. Ngoài ra, khoa học dữ liệu cho phép A/B testing (kiểm tra A/B) hiệu quả, nơi các nhà tiếp thị có thể thử nghiệm nhiều phiên bản của một landing page để xem version nào mang lại tỷ lệ chuyển đổi cao nhất. Qua các phân tích từ các thử nghiệm này, họ có thể điều chỉnh tiêu đề, lời kêu gọi hành động (CTA), và hình thức trình bày để nâng cao hiệu suất.

Một ứng dụng khác là việc cá nhân hóa landing page dựa trên thông tin của người dùng, như vị trí địa lý, lịch sử duyệt web, hoặc hành vi tìm kiếm trước đó. Bằng cách này, các landing page có thể cung cấp các khuyến nghị sản phẩm phù hợp hơn, từ đó gia tăng khả năng chuyển đổi. Thêm vào đó, việc dự đoán xu hướng và sở thích của khách hàng thông qua các mô hình phân tích dự đoán giúp tạo ra các nội dung hấp dẫn và liên quan hơn, khuyến khích người dùng thực hiện hành động mong muốn.

Với sự hỗ trợ mạnh mẽ của khoa học dữ liệu, các landing page không chỉ trở nên hiệu quả hơn mà còn tạo ra trải nghiệm cá nhân hóa cho người dùng, từ đó tăng cường tỷ lệ chuyển đổi trong chiến dịch Affiliate marketing.

Những tiến bộ như khoa học dữ liệu và công nghệ AI sẽ khuyến khích và hỗ trợ việc bán hàng nhiều hơn. Việc ứng dụng khoa học dữ liệu một cách đúng đắn và affiliate marketing giúp cho việc phân tích, lên kế hoạch và áp dụng affiliate marketing ngày càng thành công hơn.

#### *2.2.8.13. Ứng dụng trong Social Media Marketing(SMM)*

Social Media là tập hợp các nền tảng, trang web cho phép người dùng tham gia tương tác trực tuyến và chia sẻ nội dung thông tin lên đó như Facebook, Instagram, Youtube, Twitter,...

Social Media Marketing là hình thức tiếp thị sử dụng nền tảng Social Media nhằm mục đích kết nối doanh nghiệp với khách hàng tiềm năng nhằm xây dựng thương hiệu, nâng cao lượng tương tác trên website và nâng cao doanh thu bán hàng.

Ứng dụng khoa học dữ liệu đúng cách giúp SMM tối ưu hóa hiệu suất, đồng thời tạo lợi thế cạnh tranh bằng cách tận dụng dữ liệu để hiểu rõ khách hàng và dự đoán thị trường.

#### *2.2.8.14. Phân tích dữ liệu người dùng*

Phân tích dữ liệu người dùng là một trong những ứng dụng quan trọng của khoa học dữ liệu đối với ngành Social Media Marketing (SMM). Công nghệ và kỹ thuật phân tích dữ liệu hiện đại cho phép các nhà tiếp thị thu thập, xử lý và phân tích khối lượng lớn dữ liệu từ nhiều nguồn khác nhau, từ đó đưa ra các quyết định chiến lược chính xác hơn.

Phân tích dữ liệu người dùng cho phép marketers thu thập thông tin chi tiết về hành vi, sở thích và thói quen của người tiêu dùng trên các nền tảng mạng xã hội như Facebook, Instagram và Twitter. Bằng cách sử dụng các phương pháp phân tích nâng cao, họ có thể xác định đối tượng mục tiêu và phân khúc thị trường một cách chính xác, từ đó phát triển nội dung phù hợp và hấp dẫn hơn. Hơn nữa, dữ liệu từ các bài đăng, tương tác của người dùng và dữ liệu cảm xúc (sentiment analysis) giúp marketers đánh giá mức độ thành công của các chiến dịch quảng cáo và chiến lược nội dung, cho phép họ điều chỉnh kịp thời để nâng cao hiệu quả. Bên cạnh đó, phân tích dữ liệu còn hỗ trợ trong việc dự đoán xu hướng tiêu dùng tương lai, giúp xây dựng các chiến dịch hiệu quả và cá nhân hóa trải nghiệm người dùng.

Thông qua việc sử dụng các công cụ như Google Analytics, Python,...các thương hiệu có thể thu thập và phân tích dữ liệu hành vi của khách hàng trên mạng xã hội như lượt thích, chia sẻ, bình luận,...để nắm rõ sở thích và nhu cầu của khách hàng. Ví dụ, Facebook Ads có thể phân loại người dùng theo nhóm sở thích( thời trang, công nghệ, thể thao,...) và từ đó tạo ra các chiến dịch quảng cáo phù hợp, giúp tăng hiệu quả tiếp cận và chuyển đổi.

Phân tích dữ liệu người dùng trong Social Media Marketing đóng vai trò quan trọng trong việc tối ưu hóa chiến dịch tiếp thị, tạo ra trải nghiệm người dùng cá nhân hóa và dự đoán xu hướng tương lai. Công nghệ và phương pháp phân tích dữ liệu cho phép marketers thu thập thông tin quý giá, từ đó đưa ra các quyết định dựa trên dữ liệu và cải thiện hiệu quả của các chiến dịch marketing một cách đáng kể. Nhờ đó, các doanh nghiệp có thể tối ưu hóa nguồn lực đầu tư và gia tăng lợi nhuận trong môi trường cạnh tranh ngày càng khốc liệt.

#### *2.2.8.15. Dự đoán xu hướng tương lai*

Ứng dụng dự đoán xu hướng của khoa học dữ liệu trong ngành Social Media Marketing (tiếp thị truyền thông xã hội) ngày càng trở nên quan trọng, bởi vì nó giúp các doanh nghiệp hiểu rõ hơn về hành vi người tiêu dùng và tối ưu hóa các chiến dịch marketing.

Dự đoán xu hướng sử dụng dữ liệu lớn và học máy thu thập và phân tích lượng lớn thông tin từ các nền tảng mạng xã hội, từ đó dự đoán những chủ đề hay nội dung nào sẽ trở nên phổ biến trong tương lai. Việc phân tích cảm xúc từ các bài đăng, bình luận và phản hồi của người dùng cho phép marketers hiểu rõ hơn về tâm lý khách hàng và ý kiến của họ đối với sản phẩm hoặc dịch vụ. Ngoài ra, thông qua các mô hình dự đoán, các doanh nghiệp có thể xác định các yếu tố nào ảnh hưởng đến sự



tương tác của người dùng, từ đó tạo ra các nội dung và chiến dịch marketing phù hợp hơn với nhu cầu thực tế của người dùng. Bằng cách này, họ không chỉ nâng cao hiệu suất của chiến dịch mà còn xây dựng được hình ảnh thương hiệu mạnh mẽ, thu hút được sự quan tâm và lòng trung thành từ khách hàng.

Việc dự đoán xu hướng từ khoa học dữ liệu trong Social Media Marketing mang lại một cái nhìn sâu sắc về hành vi người tiêu dùng và thông tin rất quý giá cho các chiến dịch tiếp thị. Từ việc phân tích dữ liệu lớn, dự đoán xu hướng nội dung, phân khúc và định hướng khách hàng đến tối ưu hóa quảng cáo và theo dõi phản hồi thị trường, khoa học dữ liệu đã trở thành một công cụ không thể thiếu cho các marketers trong việc đưa ra quyết định chiến lược, tạo ra nội dung hấp dẫn và duy trì sự tương tác tích cực với khách hàng.

#### *2.2.8.16. Cá nhân hóa nội dung*

Cá nhân hóa nội dung là một trong những ứng dụng mạnh mẽ của khoa học dữ liệu trong lĩnh vực Social Media Marketing. Việc cá nhân hóa nội dung không chỉ giúp thương hiệu tạo sự kết nối sâu sắc hơn với khách hàng mà còn nâng cao hiệu quả của các chiến dịch truyền thông. Đây là quá trình phân tích hành vi và sở thích của từng người dùng để tạo ra các nội dung phù hợp. Điều này không chỉ giúp tăng tỷ lệ nhấp chuột mà còn nâng cao khả năng chuyển đổi khi khách hàng cảm thấy nội dung và sản phẩm được gợi ý thực sự phù hợp với nhu cầu của họ.

Bằng cách phân tích dữ liệu người dùng, bao gồm hành vi tương tác, sở thích, vị trí địa lý và các yếu tố nhân khẩu học, các marketer có thể xây dựng các chiến dịch quảng cáo và nội dung được tối ưu hóa cho từng đối tượng cụ thể. Ví dụ, thuật toán học máy có thể gợi ý sản phẩm hoặc nội dung dựa trên những gì người dùng đã xem hoặc tương tác, từ đó tăng khả năng tiết kiệm thời gian cho người tiêu dùng và nâng cao tỷ lệ chuyển đổi. Thêm vào đó, việc sử dụng các dữ liệu tương tác trên mạng xã hội, như lượt thích, bình luận và chia sẻ, cho phép các nhà tiếp thị theo dõi phản hồi tức thì và điều chỉnh nội dung để thể hiện đúng sở thích và nhu cầu của người dùng. Không chỉ giúp thương hiệu xây dựng mối quan hệ chặt chẽ hơn với khách hàng, cá nhân hóa nội dung còn tạo ra sự khác biệt trong một môi trường cạnh tranh khốc liệt, từ đó gia tăng độ trung thành của khách hàng và tạo ra giá trị lâu dài cho thương hiệu.

Việc cá nhân hóa nội dung thông qua khoa học dữ liệu trong Social Media Marketing là một yếu tố quyết định trong việc nâng cao hiệu quả truyền thông và tạo dựng mối quan hệ bền chặt với khách hàng. Bằng cách sử dụng dữ liệu một cách thông minh, marketers có thể phát triển các chiến dịch phù hợp hơn với nhu cầu và sở thích của từng người tiêu dùng, từ đó tăng cường khả năng cạnh tranh và thúc đẩy doanh thu. Cá nhân hóa nội dung không chỉ đơn thuần là một phương pháp tiếp thị, mà là một chiến lược tổng thể giúp các thương hiệu gần gũi hơn với khách hàng.

#### *2.2.8.17. Dự đoán khủng hoảng truyền thông*

Để có thể xử lý khủng hoảng kịp thời và giảm thiểu thiệt hại, quản lý và duy trì hình ảnh thương hiệu trên mạng xã hội, các doanh nghiệp cần sử dụng phân tích dữ liệu để phát hiện sớm các dấu hiệu của khủng hoảng, chẳng hạn như sự gia tăng bất thường của các bình luận tiêu cực hoặc bài viết lan truyền nhanh chóng.

Bằng cách phân tích dữ liệu lớn từ các nền tảng mạng xã hội, các công cụ phân tích có thể xác định các mô hình hành vi, xu hướng tiêu cực, và tín hiệu cảnh báo sớm về các vấn đề tiềm tàng có thể phát sinh. Thông qua việc theo dõi các chỉ số như độ tương tác, phản hồi của người tiêu dùng, và tình cảm công chúng đối với thương hiệu, doanh nghiệp có thể phát hiện được sự gia tăng của các ý kiến tiêu cực hoặc các cuộc trò chuyện có khả năng phát triển thành khủng hoảng. Các thuật toán học máy cũng có thể giúp xác định mối liên hệ giữa các sự kiện ngoại cảnh và cảm xúc của người tiêu dùng, từ đó cung cấp thông tin chi tiết về nguồn gốc và mức độ nghiêm trọng của các khủng hoảng. Sử dụng những thông tin này, marketer có thể tạo lập các kế hoạch ứng phó nhanh chóng và hiệu quả hơn, giảm thiểu thiệt hại cho thương hiệu, đồng thời duy trì sự tin tưởng và lòng trung thành của khách hàng. Ngoài ra việc kết hợp khoa học dữ liệu trong việc dự đoán khủng hoảng truyền thông, các doanh nghiệp không chỉ có thể bảo vệ hình ảnh của mình mà còn tạo ra cơ hội để củng cố mối quan hệ với khách hàng trong những thời điểm khó khăn.

Việc ứng dụng dự đoán khủng hoảng truyền thông bằng khoa học dữ liệu trong ngành Social Media Marketing không chỉ giúp các doanh nghiệp phát hiện sớm các vấn đề tiềm ẩn mà còn tạo ra khả năng ứng phó nhanh chóng và hiệu quả. Bằng cách kết hợp giám sát thời gian thực, phân tích dữ liệu, và dự đoán xu hướng, các công ty có thể duy trì và bảo vệ hình ảnh thương hiệu, qua đó xây dựng niềm tin với khách hàng và nâng cao sự trung thành của họ.

### 2.3. Tối ưu hóa quy trình vận hành và chuỗi cung ứng

#### 2.3.1. Đọc dữ liệu chuỗi cung ứng

Ví dụ: Phân tích chuỗi cung ứng của 1 công ty thời trang và làm đẹp. Tập dữ liệu này dựa trên chuỗi cung ứng các sản phẩm Trang điểm

- Đọc dữ liệu: Dùng lệnh `read_csv` của thư viện `pandas` để đọc file csv. `Head()` để hiển thị 5 dòng đầu tiên của dữ liệu.

```
1 data = pd.read_csv('supply_chain_data (1).csv')
2 data.head()
```

- Kết quả của chương trình:

	Product type	SKU	Price	Availability	Number of products sold	Revenue generated	Customer demographics	Stock levels	Lead times	Order quantities	...	Location	Lead time	Production volumes	Manufacturing lead time	Manufacturing costs	Inspection results	Defect rates	Trans
0	haircare	SKU0	69.808006		55	802	8661.996792	Non-binary	58	7	96	...	Mumbai	29	215	29	46.279879	Pending	0.226410
1	skincare	SKU1	14.843523		95	736	7460.900065	Female	53	30	37	...	Mumbai	23	517	30	33.616769	Pending	4.854068
2	haircare	SKU2	11.319683		34	8	9577.749626	Unknown	1	10	88	...	Mumbai	12	971	27	30.688019	Pending	4.580593
3	skincare	SKU3	61.163343		68	83	7766.836426	Non-binary	23	13	59	...	Kolkata	24	937	18	35.624741	Fail	4.746649
4	skincare	SKU4	4.805496		26	871	2686.505152	Non-binary	5	3	56	...	Delhi	5	414	3	92.065161	Fail	3.145580
5 rows x 24 columns																			

Hình 2.8: Hiển thị 5 dòng đầu của dữ liệu.

### 2.3.2. Phân tích chuỗi cung ứng là gì?

Phân tích chuỗi cung ứng là quá trình đánh giá mọi giai đoạn của chuỗi cung ứng bắt đầu từ thời điểm doanh nghiệp mua nguyên liệu thô hoặc vật tư từ nhà cung cấp cho đến khi giao sản phẩm cuối cùng cho khách hàng.

Mục đích của phân tích là xác định phần nào của chuỗi cung ứng có thể được cải thiện hoặc rút ngắn để cung cấp sản phẩm nhanh hơn và hiệu quả hơn cho khách hàng.

#### Phân tích chuỗi cung ứng là gì và các loại phân tích khác nhau?

Mỗi phân tích chuỗi cung ứng này có thể tăng hiệu quả chung của hoạt động kinh doanh, từ đó có thể tiết kiệm chi phí đáng kể.

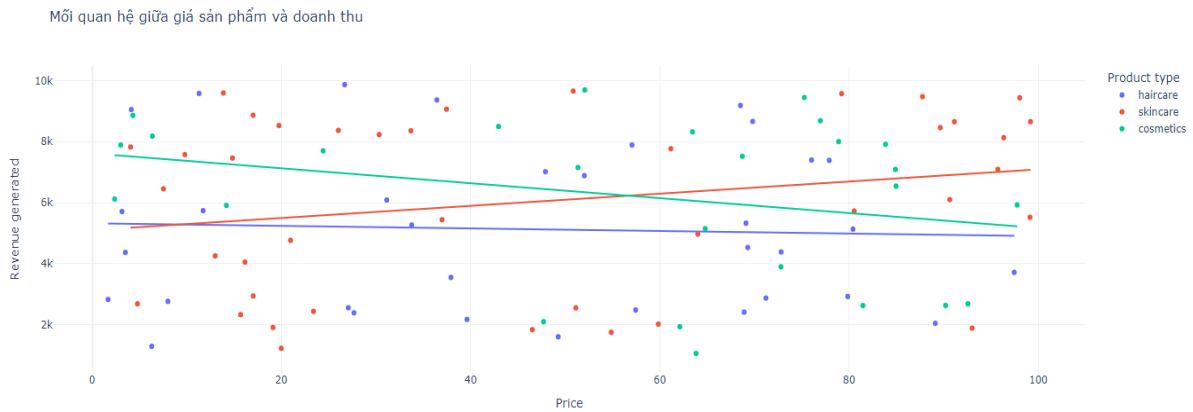
**Phân tích mô tả** tập trung vào việc hiểu những gì đã xảy ra trong quá khứ bằng cách phân tích dữ liệu lịch sử. Phân tích này có thể cung cấp thông tin chi tiết về các số liệu hiệu suất chính, chẳng hạn như mức tồn kho, thời gian giao hàng và hiệu suất giao hàng. Phân tích mô tả có thể giúp xác định các mô hình và xu hướng trong các hoạt động chuỗi cung ứng trước đây, cho phép các tổ chức đưa ra quyết định sáng suốt về các chiến lược trong tương lai.

**Phân tích chẩn đoán** vượt ra ngoài phân tích mô tả bằng cách xác định nguyên nhân gốc rễ của các vấn đề về chuỗi cung ứng. Bằng cách phân tích dữ liệu từ các nguồn khác nhau, chẳng hạn như nhà cung cấp, nhà cung cấp dịch vụ hậu cần và khách hàng, các tổ chức có thể xác định các yếu tố góp phần gây ra sự chậm trễ, gián đoạn hoặc các vấn đề về chất lượng trong chuỗi cung ứng của họ. Điều này có thể giúp họ thực hiện các hành động khắc phục để ngăn ngừa các vấn đề tương tự xảy ra trong tương lai.

Ví dụ: Phân tích Chuỗi cung ứng bằng cách xem xét mối quan hệ giữa giá sản phẩm và doanh thu tạo ra từ sản phẩm:

```
1 fig = px.scatter(data, x = 'Price', y = 'Revenue generated',
2                  title='Mối quan hệ giữa giá sản phẩm và doanh thu',
3                  color = 'Product type', hover_data= 'Number of products sold',
4                  trendline='ols')
5 fig.show()
```

Kết quả:



Hình 2.9: Biểu đồ mối quan hệ giữa giá sản phẩm và doanh thu

### 2.3.3. Dịch vụ chăm sóc khách hàng

Doanh nghiệp cần đảm bảo một số yếu tố liên quan đến dịch vụ khách hàng trong quá trình tối ưu chuỗi cung ứng như:

- Phản hồi nhanh chóng, điều phối thông tin kịp thời
- Hỗ trợ giải quyết khiếu nại nhanh chóng
- Chân thành tiếp nhận và phân tích ý kiến đóng góp của khách hàng

Các công việc trên sẽ đảm bảo rằng doanh nghiệp cung cấp thông tin và giải đáp thắc mắc của khách hàng kịp thời và chính xác, nhằm gia tăng độ tin cậy và hài lòng của khách hàng. Bên cạnh đó, khi thu thập và phân tích phản hồi từ khách hàng doanh nghiệp dễ dàng đánh giá hiệu suất của chuỗi cung ứng, góp phần xác định các khía cạnh có thể cải thiện và thúc đẩy sự phát triển liên tục trong quá trình tối ưu hóa chuỗi cung ứng.

Ví dụ, để xoa dịu một khách hàng bị shop gửi sai sản phẩm, người bán hàng ngay lập tức xin lỗi và sẵn sàng hỗ trợ đổi hàng cho khách, không để họ phải chịu bất kỳ chi phí nào phát sinh để mang lại sự hài lòng cho khách hàng.

### 2.3.4. Xử lý mạng lưới chuỗi cung ứng

Một mạng lưới chuỗi cung ứng tối ưu giúp đảm bảo rằng các hoạt động từ nguồn cung cấp đến khách hàng được thực hiện một cách hiệu quả nhất, với chi phí thấp nhất và thời gian giao hàng ngắn nhất. Dưới đây là cách xây dựng mạng lưới chuỗi cung ứng trong quá trình tối ưu hóa chuỗi cung ứng:

- **Phân tích và đánh giá:** Cần phân tích và đánh giá các yếu tố quan trọng trong chuỗi cung ứng gồm các đối tác, quy trình, hệ thống thông tin và nguồn hàng hóa. Điều này giúp làm rõ vị trí và vai trò của từng thành phần trong chuỗi cung ứng.
- **Xác định mục tiêu và chiến lược:** Mục tiêu có thể bao gồm cải thiện thời gian giao hàng, giảm chi phí vận chuyển, tăng cường khả năng đáp ứng nhu cầu khách hàng, hoặc tối ưu hóa tồn kho.

- **Lựa chọn đối tác:** Đối tác phải có khả năng cung cấp các sản phẩm hoặc dịch vụ chất lượng, đáng tin cậy, và tuân thủ các tiêu chuẩn về chất lượng sản phẩm/dịch vụ, chuẩn mực và đạo đức kinh doanh. Cần xem xét cả khía cạnh về vị trí địa lý để đảm bảo khả năng vận chuyển và phân phối hiệu quả.
- **Thiết kế cấu trúc mạng lưới:** Xây dựng cấu trúc mạng lưới chuỗi cung ứng bao gồm vị trí và chức năng của các cơ sở sản xuất, kho bãi, trung tâm phân phối và các điểm bán hàng. Cần thiết kế một cấu trúc mạng lưới linh hoạt và tối ưu, đáp ứng được yêu cầu về vị trí địa lý, khả năng sản xuất, vận chuyển và phân phối.
- **Theo dõi và đánh giá:** Sau khi triển khai mạng lưới chuỗi cung ứng, cần thực hiện các đánh giá định kỳ để xác định các điểm mạnh và điểm yếu như: chỉ số hiệu suất như thời gian giao hàng, khối lượng tồn kho, độ chính xác dự báo và đáp ứng nhu cầu khách hàng cần được đo lường và theo dõi, từ đó đưa ra các biện pháp cải tiến phù hợp.

#### 2.3.5. Tận dụng tối đa tài sản vào chuỗi cung ứng.

Để tận dụng tối đa tài sản trong quá trình tối ưu hóa chuỗi cung ứng, có một số cách doanh nghiệp có thể áp dụng:

**Chia sẻ tài sản:** Trong khả năng cho phép, hãy chia sẻ tài sản với các đối tác trong chuỗi cung ứng có thể giúp tối ưu hóa việc sử dụng tài sản. Ví dụ, bạn có thể chia sẻ kho lưu trữ, thiết bị hoặc nhân lực với các nhà cung ứng khác để giảm thiểu chi phí và tăng cường hiệu quả.

**Tối ưu hóa quá trình sản xuất:** Áp dụng các phương pháp quản lý chất lượng và quy trình sản xuất hiệu quả để giảm thiểu lãng phí và tăng cường hiệu suất. Sử dụng các công nghệ và phần mềm quản lý sản xuất có thể giúp bạn tối ưu hóa sử dụng tài sản như máy móc, thiết bị và nhân lực.

**Tối ưu hóa vận chuyển và logistics:** Tìm kiếm các lựa chọn vận chuyển và định tuyến hiệu quả, tối thiểu hóa thời gian và chi phí vận chuyển, và sử dụng công nghệ để theo dõi và quản lý hệ thống vận chuyển.

Tận dụng tài sản trong chuỗi cung ứng đòi hỏi một quá trình quản lý thông minh và sử dụng các công nghệ và phương pháp hiệu quả. Bằng cách chia sẻ tài sản, tối ưu hóa sản xuất, vận chuyển và sử dụng công cụ và công nghệ, bạn có thể tận dụng tài sản một cách tối đa để đạt được hiệu quả trong tối ưu hóa chuỗi cung ứng.

**Ví dụ:** Phân tích doanh số theo từng loại sản phẩm:

- Thống kê mô tả: dùng `describe()` để thống kê mô tả bao gồm:

- Count: Tổng các phần tử không bị thiếu
- mean: trung bình của columns
- std: độ lệch chuẩn
- min: giá trị thấp nhất
- max: giá trị lớn nhất

```
1 print(data.describe())
```

Kết quả:

	Price	Availability	Number of products sold	Revenue generated \
count	100.000000	100.000000	100.000000	100.000000
mean	49.462461	48.400000	460.990000	5776.048187
std	31.168193	30.743317	303.780074	2732.841744
min	1.699976	1.000000	8.000000	1061.618523
25%	19.597823	22.750000	184.250000	2812.847151
50%	51.239831	43.500000	392.500000	6006.352023
75%	77.198228	75.000000	704.250000	8253.976921
max	99.171329	100.000000	996.000000	9866.465458

	Stock levels	Lead times	Order quantities	Shipping times \
count	100.000000	100.000000	100.000000	100.000000
mean	47.770000	15.960000	49.220000	5.750000
std	31.369372	8.785801	26.784429	2.724283
min	0.000000	1.000000	1.000000	1.000000
25%	16.750000	8.000000	26.000000	3.750000
50%	47.500000	17.000000	52.000000	6.000000
75%	73.000000	24.000000	71.250000	8.000000
max	100.000000	30.000000	96.000000	10.000000

	Shipping costs	Lead time	Production volumes \
count	100.000000	100.000000	100.000000
mean	5.548149	17.080000	567.840000
std	2.651376	8.846251	263.046861
min	1.013487	1.000000	104.000000
25%	3.540248	10.000000	352.000000
50%	5.320534	18.000000	568.500000
75%	7.601695	25.000000	797.000000
max	9.929816	30.000000	985.000000

	Manufacturing lead time	Manufacturing costs	Defect rates	Costs
count	100.000000	100.000000	100.000000	100.000000
mean	14.770000	47.266693	2.277158	529.245782
std	8.91243	28.982841	1.461366	258.301696
min	1.000000	1.085069	0.018608	103.916248
25%	7.000000	22.983299	1.009650	318.778455
50%	14.000000	45.905622	2.141863	520.430444
75%	23.000000	68.621026	3.563995	763.078231
max	30.000000	99.466109	4.939255	997.413450

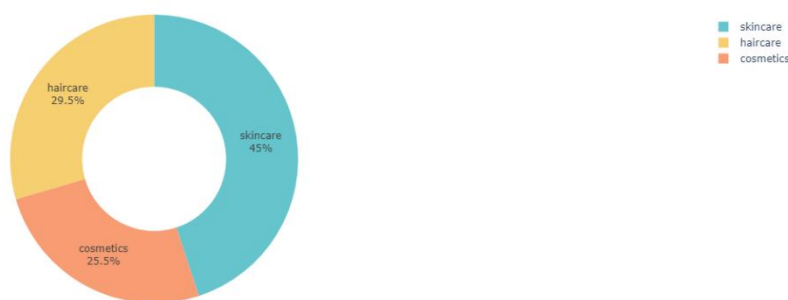
Hình 2.10: Dùng describe() để thống kê mô tả.

- Phân tích doanh số theo từng loại sản phẩm:

```
1 sales_data = data.groupby('Product type')
2   ['Number of products sold'].sum().reset_index()
3 pie_chart = px.pie(sales_data, values = 'Number of products sold',
4                    names = 'Product type',
5                    title = 'Doanh thu số theo từng loại sản phẩm',
6                    hover_data=['Number of products sold'], hole = 0.5,
7                    color_discrete_sequence=px.colors.qualitative.Pastel)
8 pie_chart.update_traces(textposition = 'inside', textinfo = 'percent+label')
9 pie_chart.show()
```

## Kết quả:

Doanh thu số theo từng loại sản phẩm



Hình 2.11: Biểu đồ doanh thu theo từng sản phẩm.

Biểu đồ cho thấy skincare và haircare là các sản phẩm chủ lực, đóng góp lần lượt 48% và 29% tổng doanh thu, trong khi cosmetics chỉ chiếm 23%. Trong tối ưu hóa chuỗi cung ứng, doanh nghiệp nên ưu tiên phân bổ nguồn lực cho skincare và haircare, giảm đầu tư vào cosmetics.

### 2.3.6. Đo lường và đánh giá hiệu suất làm việc của chuỗi cung ứng

Để đảm bảo việc cải thiện và theo dõi hiệu quả của các biện pháp tối ưu hóa, doanh nghiệp cần đo lường và đánh giá hiệu suất làm việc. Dưới đây là một số cách phổ biến để đo lường và đánh giá hiệu suất trong tối ưu hóa chuỗi cung ứng:

**Chỉ số hiệu suất khách quan (KPIs):** Đặt ra và theo dõi các chỉ số hiệu suất khách quan để đánh giá các khía cạnh quan trọng của chuỗi cung ứng. Một số KPIs thường được sử dụng bao gồm: tỷ lệ lỗi đơn hàng, thời gian xử lý đơn hàng, tỷ lệ hoàn trả, tỷ suất sử dụng tài sản, giá trị tồn kho, thời gian giao hàng, và chi phí vận chuyển. Bằng cách so sánh các KPIs với mục tiêu hoặc tiêu chuẩn đã đề ra, doanh nghiệp có thể đánh giá hiệu suất làm việc và xác định các lĩnh vực cần cải thiện.

**Đánh giá đối tác trong chuỗi cung ứng:** Điều này có thể bao gồm việc xem xét độ chính xác và đáng tin cậy của giao hàng, thời gian đáp ứng, chất lượng sản phẩm hoặc dịch vụ, và mức độ tuân thủ các tiêu chuẩn và quy trình chung. Đánh giá hiệu suất đối tác giúp bạn xác định các mối quan hệ yếu và cơ hội để cải tiến và tối ưu hóa chuỗi cung ứng.

**Phân tích chi phí và lợi ích:** Đo lường hiệu suất cũng liên quan đến phân tích chi phí và lợi ích của các biện pháp tối ưu hóa. Xem xét chi phí đầu tư, chi phí vận hành và chi phí tồn kho, so sánh với lợi ích như giảm thiểu lãng phí, tăng cường hiệu suất sản xuất, giảm chi phí vận chuyển hoặc tăng cường sự linh hoạt. Bằng cách đánh giá tổng quan chi phí và lợi ích, bạn có thể đo lường hiệu suất toàn diện và đưa ra quyết định về các biện pháp tối ưu hóa.

**Phản hồi khách hàng:** Phản hồi từ khách hàng là một chỉ số quan trọng để đánh giá hiệu suất chuỗi cung ứng. Thu thập phản hồi từ khách hàng về chất lượng sản phẩm, thời gian giao hàng, dịch vụ khách hàng và các yếu tố liên quan khác.

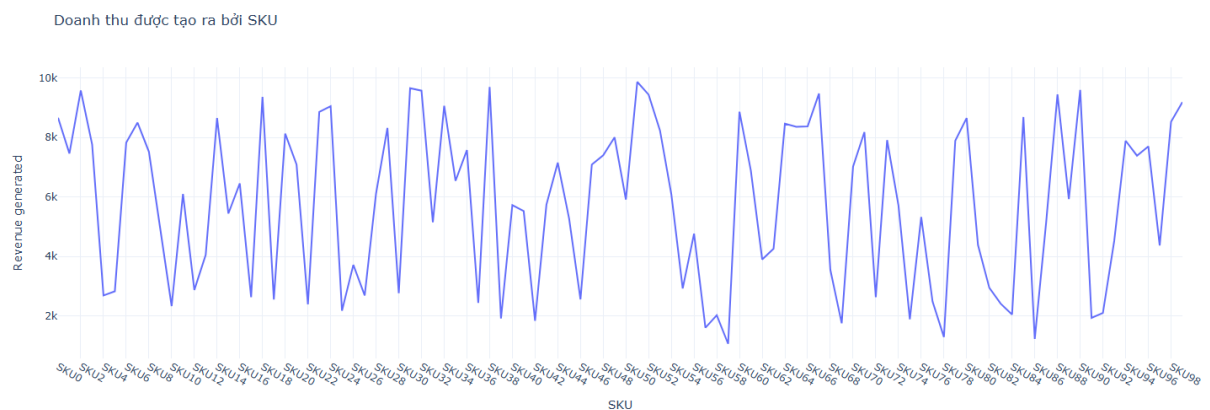


Điều này giúp bạn đánh giá mức độ đáp ứng của chuỗi cung ứng đến nhu cầu và mong muốn của khách hàng và xác định các mục tiêu cải thiện.

**Ví dụ:** Đánh giá doanh thu được tạo ra từ SKU. Có một cột trong tập dữ liệu là SKU. SKU là viết tắt của Stock Keeping Units. Chúng giống như các mã đặc biệt giúp các công ty theo dõi tất cả các mặt hàng khác nhau mà họ đang bán.

```
1 revenue_chart = px.line(data, x='SKU',  
2                           y='Revenue generated',  
3                           title='Doanh thu được tạo ra bởi SKU')  
4 revenue_chart.show()
```

Kết quả:

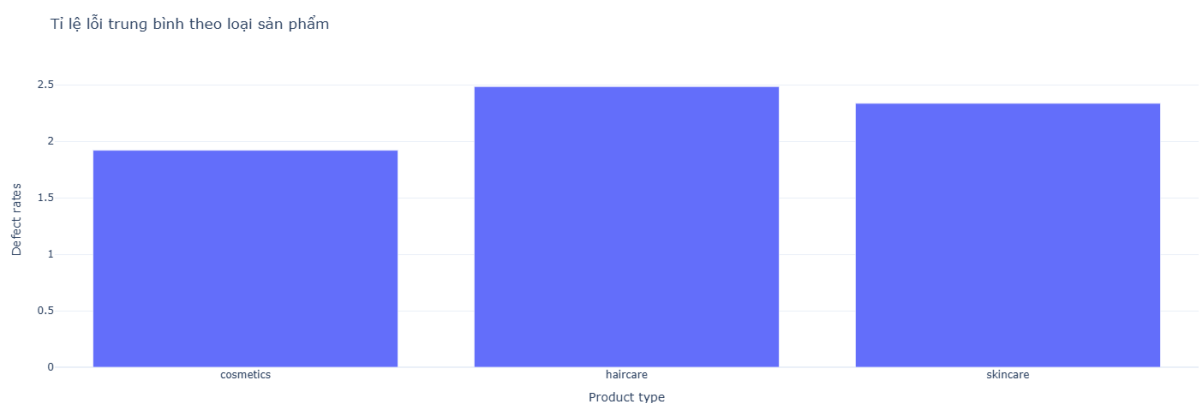


Hình 2.12: Biểu đồ doanh thu được tạo ra bởi SKU.

Tỉ lệ lỗi trung bình theo loại sản phẩm:

```
1 defect_rates_by_product = data.groupby('Product type')['Defect rates'].mean().reset_index()  
2 fig = px.bar(defect_rates_by_product, x='Product type', y='Defect rates',  
3              title='Tỉ lệ lỗi trung bình theo loại sản phẩm')  
4 fig.show()
```

Kết quả:



Hình 2.13: Biểu đồ tỉ lệ lỗi trung bình theo từng loại sản phẩm.

Biểu đồ cho thấy haircare có tỉ lệ lỗi trung bình cao nhất (2.5), tiếp theo là skincare (2.3) và cosmetics (2.0). Điều này cho thấy vấn đề về chất lượng là một thách thức lớn trong chuỗi cung ứng, đặc biệt với haircare, loại sản phẩm đóng góp 29% doanh thu. Doanh nghiệp cần tập trung cải thiện kiểm soát chất lượng, tối ưu



hóa quy trình sản xuất, và quản lý tồn kho hiệu quả hơn để giảm chi phí và tăng doanh thu.

### 2.3.7. Tối ưu hóa vận chuyển giao hàng

#### 2.3.7.1. Sử dụng phần mềm tối ưu hóa lộ trình

Tối ưu hóa tuyến đường giao hàng là một trong những cách hiệu quả nhất để giảm thiểu chi phí logistics. Để đạt được điều này, doanh nghiệp có thể sử dụng phần mềm tối ưu hóa lộ trình nhằm thiết lập tuyến đường vận chuyển tốt nhất cho các đơn hàng dựa trên dữ liệu đầu vào như: lượng đơn hàng, tình hình giao thông, thời gian giao dự tính, vị trí các trạm kiểm soát, khung giờ giao hàng của từng đơn, số lượng và trọng tải xe, v.v.

Thay vì lập kế hoạch một cách thủ công, các phần mềm tối ưu hóa lộ trình tự động sẽ giúp doanh nghiệp phân bổ thời gian và nguồn lực một cách hiệu quả hơn, đồng thời giảm chi phí vận hành và chi phí nhiên liệu.

#### 2.3.7.2. Tối ưu hóa phân bổ tải trọng của phương tiện vận tải

Phân bổ tải trọng là việc sắp xếp hàng hóa để tận dụng tốt nhất khả năng vận chuyển, tối đa hóa giá trị sử dụng của phương tiện. Xe chở hết tải trọng (tận dụng tối đa công suất) là cách tối ưu để giảm chi phí vận chuyển, giảm số lần giao hàng và rủi ro hư hỏng sản phẩm do bị xô lệch trong quá trình di chuyển.

Với khả năng phân bổ tải trọng tốt, doanh nghiệp có thể giảm số lượng phương tiện và nhân lực cần thiết, đồng thời tăng hiệu quả và độ chính xác của hoạt động vận chuyển.

#### 2.3.7.3. Lựa chọn phương tiện vận chuyển phù hợp mục đích

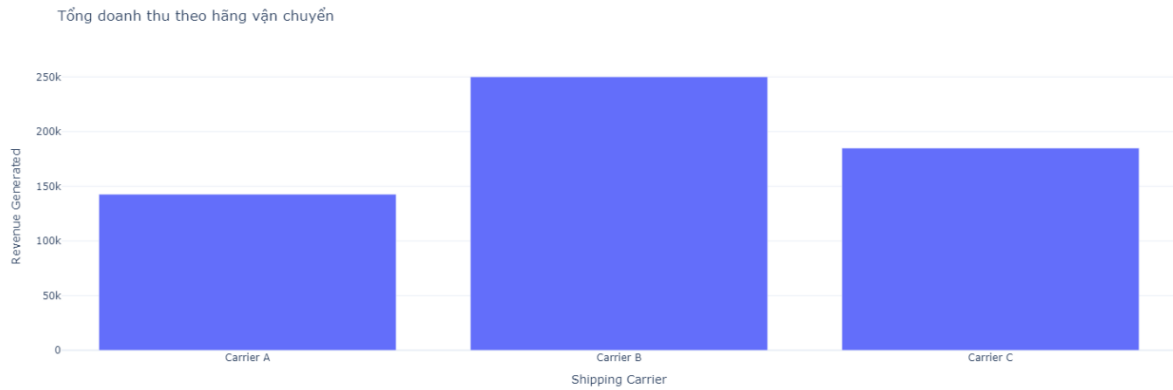
Các công ty cũng cần tính toán các phương án lựa chọn phương tiện phù hợp với các mục đích, điều kiện địa lý và giao thông cụ thể. Chẳng hạn, những hạn chế về quy định giao thông đô thị và tình trạng tắc nghẽn ngày càng gia tăng khiến doanh nghiệp khó có thể sử dụng cùng một phương tiện trên các tuyến đường trong thành phố giống như phương tiện đang sử dụng ở vùng ngoại ô hoặc nông thôn.

Lựa chọn phương tiện vận tải phù hợp và sử dụng tối đa công suất là một giải pháp hữu hiệu để cắt giảm chi phí và thời gian cho quá trình vận chuyển trên đường, từ đó nâng cao hiệu quả giao hàng chặng cuối.

**Ví dụ:** Tổng doanh thu theo hãng vận chuyển:

```
1 total_revenue = data.groupby('Shipping carriers')['Revenue generated'].sum().reset_index()
2 fig = go.Figure()
3 fig.add_trace(go.Bar(x=total_revenue['Shipping carriers'],
4                       y=total_revenue['Revenue generated'])))
5 fig.update_layout(title='Tổng doanh thu theo hãng vận chuyển',
6                   xaxis_title='Shipping Carrier',
7                   yaxis_title='Revenue Generated')
8 fig.show()
```

Kết quả:



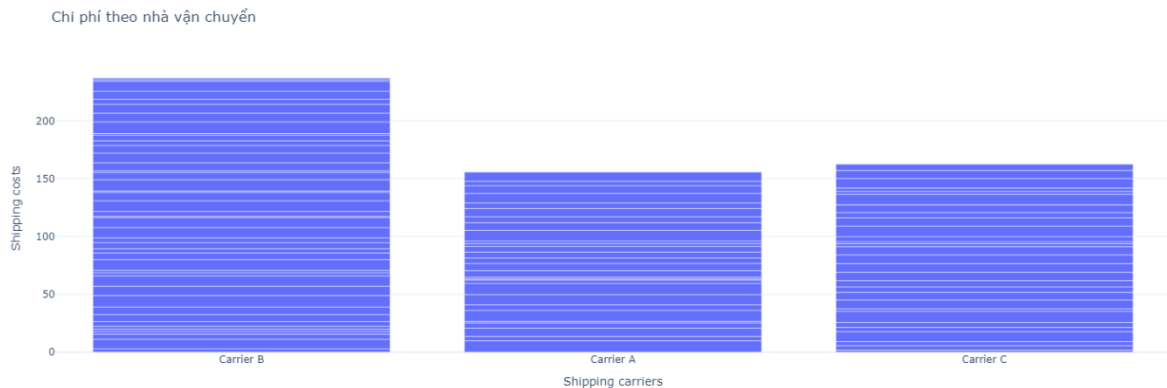
Hình 2.14: Biểu đồ tổng doanh thu theo hãng vận chuyển.

Biểu đồ cho thấy Carrier B đóng góp doanh thu cao nhất (250k), tiếp theo là Carrier C (200k) và Carrier A (120k). Trong tối ưu hóa chuỗi cung ứng, doanh nghiệp nên ưu tiên hợp tác với Carrier B, giảm sử dụng Carrier A, và tận dụng Carrier C như một lựa chọn dự phòng.

Chi phí theo nhà vận chuyển:

```
1 shipping_cost_chart = px.bar(data, x = 'Shipping carriers',
2                               y = 'Shipping costs',
3                               title = 'Chi phí theo nhà vận chuyển')
4 shipping_cost_chart.show()
```

Kết quả:



Hình 2.15: Biểu đồ chi phí theo nhà vận chuyển.

Biểu đồ cho thấy Carrier B có chi phí vận chuyển cao nhất (225k), trong khi Carrier A và Carrier C có chi phí thấp hơn (150k). Tuy nhiên, khi kết hợp với doanh thu (Carrier B: 250k, Carrier C: 200k, Carrier A: 120k), Carrier C là lựa chọn hiệu quả nhất về lợi nhuận, trong khi Carrier A gây lỗ. Trong tối ưu hóa chuỗi cung ứng, doanh nghiệp nên đàm phán giảm chi phí với Carrier B, tăng sử dụng Carrier C, và giảm hoặc ngừng hợp tác với Carrier A.

Công ty đang sử dụng ba hãng vận chuyển để vận chuyển và Hãng vận chuyển B giúp công ty tạo ra nhiều doanh thu hơn. Bây giờ chúng ta hãy xem Thời gian giao hàng trung bình và Chi phí sản xuất trung bình cho tất cả các sản phẩm của công ty:

```

1 avg_lead_time = data.groupby('Product type')['Lead time'].mean().reset_index()
2 avg_manufacturing_costs = data.groupby('Product type')['Manufacturing costs'].mean().reset_index()
3 result = pd.merge(avg_lead_time, avg_manufacturing_costs, on='Product type')
4 result.rename(columns={'Lead time': 'Average Lead Time',
5                        'Manufacturing costs': 'Average Manufacturing Costs'},
6               inplace=True)
7 print(result)

```

Kết quả:

	Product type	Average Lead Time	Average Manufacturing Costs
0	cosmetics	13.538462	43.052740
1	hairecare	18.705882	48.457993
2	skincare	18.000000	48.993157

Hình 2.16: Thời gian giao hàng trung bình và chi phí sản xuất trung bình.

Chi phí theo phương thức vận chuyển:

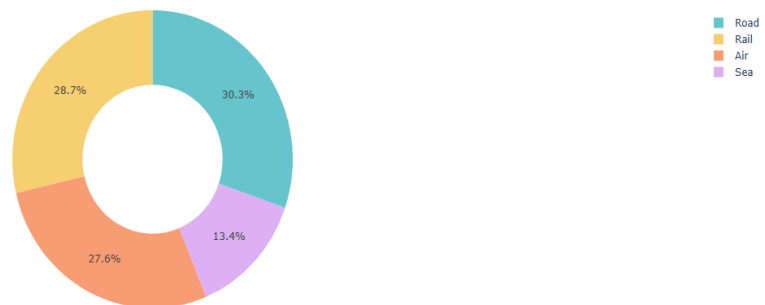
```

1 transportation_chart = px.pie(data,
2                               values='Costs',
3                               names='Transportation modes',
4                               title='Chi phí theo phương thức vận chuyển',
5                               hole=0.5,
6                               color_discrete_sequence=px.colors.qualitative.Pastel)
7 transportation_chart.show()

```

Kết quả:

Chi phí theo phương thức vận chuyển



Hình 2.17: Biểu đồ chi phí theo phương thức vận chuyển.

Vì vậy, công ty chi nhiều hơn cho phương tiện vận tải đường bộ và đường sắt để vận chuyển hàng hóa.

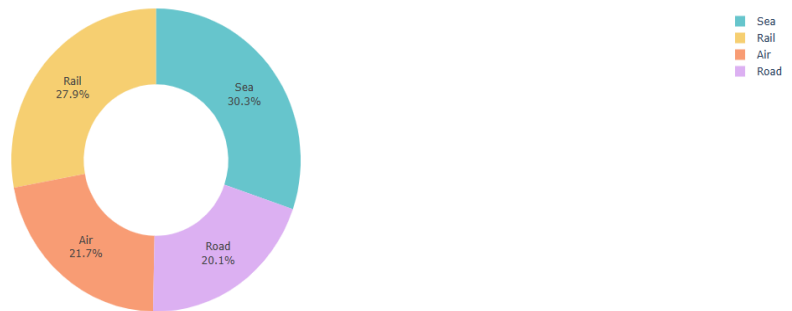
Thời gian giao hàng theo nhà vận chuyển:

```

1 shipping_time = data.groupby('Transportation modes')['Shipping times'].mean().reset_index()
2 shipping_time_chart = px.pie(shipping_time, values = 'Shipping times', names = 'Transportation modes',
3                              title = 'Thời gian giao hàng theo nhà vận chuyển',
4                              hover_data='Shipping times', hole = 0.5,
5                              color_discrete_sequence=px.colors.qualitative.Pastel)
6 shipping_time_chart.update_traces(textposition = 'inside', textinfo = 'percent+label')
7 shipping_time_chart.show()

```

Kết quả của chương trình:



Hình 2.18: Biểu đồ thời gian giao hàng trung bình theo nhà vận chuyển.

Giao hàng theo đường bộ là nhanh nhất mà chi phí cũng rẻ nhất, tuy nhiên với những đơn hàng ở xa sẽ không vận chuyển bằng cách này được. Với những đơn hàng xa thì không nên vận chuyển bằng đường biển vì thời gian vận chuyển sẽ lâu và chi phí cũng cao.

#### 2.3.7.4. Hợp tác thực hiện Logistic thu hồi

Logistics thu hồi (hay Logistics ngược – Reverse Logistics) là quá trình cung cấp hàng hóa mong muốn cho khách hàng, nhận và gửi lại những sản phẩm không mong muốn của họ, và các hoạt động liên quan.

Một kế hoạch giao hàng chặng cuối thành công đòi hỏi doanh nghiệp phải cân đối nhiều yêu cầu về quản lý đơn hàng, quản lý hàng tồn kho, tối ưu hóa tuyến đường, phân phối và vận chuyển. Trong thời đại chuyển đổi số hiện nay, phương án hiệu quả và tất yếu cho các doanh nghiệp là phải tích hợp công nghệ trong điều phối hoạt động giao hàng. Đây là chìa khóa để các công ty tạo sự khác biệt cho thương hiệu và nâng cao hiệu quả kinh doanh của mình.

### 2.4. Quản lý tồn kho và giao hàng

#### 2.4.1. Tối ưu hóa kho hàng

##### 2.4.1.1. Dữ liệu mẫu

Đoạn code sau thông kê các sản phẩm trong kho:

```
# Tạo danh sách sản phẩm với thông tin kích thước và giá
products = pd.DataFrame({
    'product_id': range(1, 11),
    'product_name': ['Bánh mì', 'Sữa hộp', 'Trứng gà', 'Gạo', 'Nước ngọt', 'Kem đánh răng', 'Dầu gội', 'Bột giặt', 'Xà phòng', 'Mì gói'],
    'width': [0.2, 0.1, 0.3, 0.5, 0.2, 0.1, 0.2, 0.3, 0.1, 0.2], # Chiều rộng (m)
    'height': [0.1, 0.2, 0.1, 0.3, 0.3, 0.2, 0.3, 0.4, 0.1, 0.1], # Chiều cao (m)
    'depth': [0.3, 0.2, 0.2, 0.4, 0.2, 0.1, 0.2, 0.3, 0.1, 0.2], # Chiều sâu (m)
    'stock': [50, 100, 80, 30, 120, 60, 40, 20, 70, 90] # Số lượng tồn kho
})
products['volume'] = products['width'] * products['height'] * products['depth']
```

Cấu trúc: Đoạn mã tạo một DataFrame products với các cột: product\_id (mã sản phẩm), name (tên sản phẩm), width (chiều rộng), height (chiều cao), depth (chiều sâu), và stock (số lượng tồn kho).

Dữ liệu:

- product\_id: Từ 1 đến 10.
- name: Danh sách 10 sản phẩm tiêu dùng (Bánh mì, Sữa hộp, v.v.).
- width, height, depth: Các giá trị kích thước (mét), ví dụ: [0.2, 0.1, ...].
- stock: Số lượng tồn kho, ví dụ: [50, 100, ...].

Tính toán: Thêm cột volume (thể tích) bằng công thức: width \* height \* depth.

Mục đích: Quản lý thông tin sản phẩm và tính toán thể tích, có thể phục vụ cho việc lưu trữ hoặc vận chuyển.

```
# Quản lý luồng hàng hóa
# Mô phỏng lộ trình lấy hàng cho một đơn hàng ngẫu nhiên
sample_order = orders[0]['items']
path = [(0, 0, 0)] # Bắt đầu từ cửa kho (0, 0, 0)
total_distance = 0
for item in sample_order:
    pos = product_positions_df[product_positions_df['product_id'] == item][['x', 'y', 'z']].iloc[0]
    path.append((pos['x'], pos['y'], pos['z']))
    # Tính khoảng cách Euclid
    last_pos = path[-2]
    distance = np.sqrt((pos['x'] - last_pos[0])**2 + (pos['y'] - last_pos[1])**2 + (pos['z'] - last_pos[2])**2)
    total_distance += distance
path.append((0, 0, 0)) # Quay lại cửa kho
last_pos = path[-2]
distance = np.sqrt((0 - last_pos[0])**2 + (0 - last_pos[1])**2 + (0 - last_pos[2])**2)
total_distance += distance

print(f"Khoảng cách di chuyển cho đơn hàng mẫu: {total_distance:.2f} mét")

# Hiệu quả sử dụng không gian
# Tính toán không gian sử dụng trên từng kệ, xem xét xếp chồng
shelf_usage = []
for shelf in shelves:
    shelf_id = shelf['shelf_id']
    shelf_products = product_positions_df[product_positions_df['x'] == shelf['x']]
    total_volume = shelf_products['volume'].sum() * shelf_products['stock'].sum()
    volume_used = min(total_volume, shelf_volume) # Không vượt quá dung lượng kệ
    volume_remaining = shelf_volume - volume_used
    usage_percentage = (volume_used / shelf_volume) * 100
    shelf_usage.append({
        'shelf_id': shelf_id,
        'volume_used': volume_used,
        'volume_remaining': volume_remaining,
        'usage_percentage': usage_percentage,
        'products': list(shelf_products['product_name'])
    })
shelf_usage_df = pd.DataFrame(shelf_usage)

print("\nSử dụng không gian trên từng kệ:")
print(shelf_usage_df)
```

Các cột dữ liệu:

- product\_id: Mã sản phẩm, là dãy số từ 1 đến 10, được tạo bằng range(1, 11).
- name: Tên sản phẩm bằng tiếng Việt, gồm: 'Bánh mì', 'Sữa hộp', 'Trứng gà', 'Gạo', 'Nước ngọt', 'Kem đánh răng', 'Đậu phụ', 'Bột giặt', 'Xà phòng', 'Mì gói'. Đây là các mặt hàng tiêu dùng phổ biến.
- width: Chiều rộng (mét), với giá trị: [0.2, 0.1, 0.3, 0.5, 0.2, 0.1, 0.2, 0.3, 0.1, 0.2].
- height: Chiều cao (mét), với giá trị: [0.1, 0.2, 0.1, 0.3, 0.2, 0.3, 0.4, 0.1, 0.1]. Lưu ý: Chỉ có 9 giá trị, thiếu 1 so với các cột khác (10 giá trị), có thể gây lỗi.
- depth: Chiều sâu (mét), với giá trị: [0.3, 0.2, 0.2, 0.4, 0.2, 0.1, 0.2, 0.3, 0.1, 0.2].

- stock: Số lượng tồn kho, với giá trị: [50, 100, 80, 30, 120, 60, 40, 20, 70, 90].

Tính toán cột volume:

- Câu lệnh `products['volume'] = products['width'] * products['height'] * products['depth']` tính thể tích ( $m^3$ ) bằng cách nhân các giá trị của width, height, và depth tương ứng cho từng sản phẩm.

```
# Mô phỏng kho hàng với cấu trúc thực tế
# Kho có 3 kệ, mỗi kệ có 2 tầng, mỗi tầng có 2 vị trí (tổng 12 vị trí)
# Kích thước kệ: chiều rộng 5m, chiều cao 3m, chiều sâu 2m
# Tọa độ kệ: Kệ 1 (x=5), Kệ 2 (x=10), Kệ 3 (x=15)
shelves = [
    {'shelf_id': 1, 'x': 5, 'positions': [{'pos_id': 1, 'y': 1, 'z': 1}, {'pos_id': 2, 'y': 3, 'z': 1}, {'pos_id': 3, 'y': 1, 'z': 3}, {'pos_id': 4, 'y': 3, 'z': 3}],
     'shelf_id': 2, 'x': 10, 'positions': [{'pos_id': 5, 'y': 1, 'z': 1}, {'pos_id': 6, 'y': 3, 'z': 1}, {'pos_id': 7, 'y': 1, 'z': 3}, {'pos_id': 8, 'y': 3, 'z': 3}],
     'shelf_id': 3, 'x': 15, 'positions': [{'pos_id': 9, 'y': 1, 'z': 1}, {'pos_id': 10, 'y': 3, 'z': 1}, {'pos_id': 11, 'y': 1, 'z': 3}, {'pos_id': 12, 'y': 3, 'z': 3}]}
shelf_capacity = {'width': 5, 'height': 3, 'depth': 2}
shelf_volume = shelf_capacity['width'] * shelf_capacity['height'] * shelf_capacity['depth']
```

- Cấu trúc dữ liệu lồng nhau: shelves sử dụng danh sách và từ điển để mô phỏng kho hàng với không gian 3D (x, y, z). Trục x (5, 10, 15) biểu thị vị trí kệ, y và z (1 hoặc 3) biểu thị tầng và độ sâu.
- Tính thể tích: Dung tích mỗi kệ được tính bằng công thức thể tích=chiều rộng×chiều cao×chiều sâu với  $5 \times 3 \times 2 = 30m^3$ . Tổng dung tích kho là  $30 \times 3 = 90m^3$ .

```
# Tối ưu hóa bố trí kho hàng
# Ưu tiên đặt các sản phẩm có tần suất cao và thường mua cùng nhau
top_pair = co_purchase_df.iloc[0]['pair']
product_a, product_b = top_pair
product_positions = {product_a: 1, product_b: 2}

# Đặt các sản phẩm còn lại dựa trên tần suất
remaining_products = products[~products['product_id'].isin([product_a, product_b])]
remaining_products = remaining_products.sort_values('frequency', ascending=False)
remaining_positions = list(range(3, 13))
for i, product in enumerate(remaining_products['product_id']):
    product_positions[product] = remaining_positions[i % len(remaining_positions)]
```

## Lý thuyết:

- **Phân tích tần suất mua chung (Co-Purchase Analysis):** sử dụng dữ liệu từ `co_purchase_df` để xác định cặp sản phẩm thường mua cùng nhau, đặt chúng vào các vị trí gần nhau (giả định là vị trí 1 và 2).
- **Lọc dữ liệu với Pandas:** `~products['product_id'].isin([...])` lọc các sản phẩm không thuộc cặp đã chọn, đảm bảo chúng được phân bổ vào các vị trí còn lại.

**Phân bổ tuần hoàn:** Phép modulo (`i % len(remaining_positions)`) đảm bảo tất cả sản phẩm được gán vị trí trong phạm vi 3 đến 12.

## Kết quả:



Sử dụng không gian trên từng kệ:					
	shelf_id	volume_used	volume_remaining	usage_percentage	\
0	1	6.93	23.07	23.1	
1	2	20.16	9.84	67.2	
2	3	1.50	28.50	5.0	

	products
0	[Trứng gà, Xà phòng, Kem đánh răng, Nước ngọt]
1	[Mì gói, Gạo, Bột giặt, Dầu gội]
2	[Sữa hộp, Bánh mì]

● **Kệ 1:**

- Thể tích đã sử dụng: 6.93 m<sup>3</sup>.
- Thể tích còn lại: 23.07 m<sup>3</sup>.
- Tỷ lệ sử dụng: 23.1%.
- Sản phẩm: Trứng gà, Xà phòng, Kem đánh răng, Nước ngọt.

● **Kệ 2:**

- Thể tích đã sử dụng: 20.16 m<sup>3</sup>.
- Thể tích còn lại: 9.84 m<sup>3</sup>.
- Tỷ lệ sử dụng: 67.2%.
- Sản phẩm: Mì gói, Gạo, Bột giặt, Đậu phụ.

● **Kệ 3:**

- Thể tích đã sử dụng: 1.50 m<sup>3</sup>.
- Thể tích còn lại: 28.50 m<sup>3</sup>.
- Tỷ lệ sử dụng: 5.0%.
- Sản phẩm: Sữa hộp, Bánh mì.

**Ý Nghĩa**

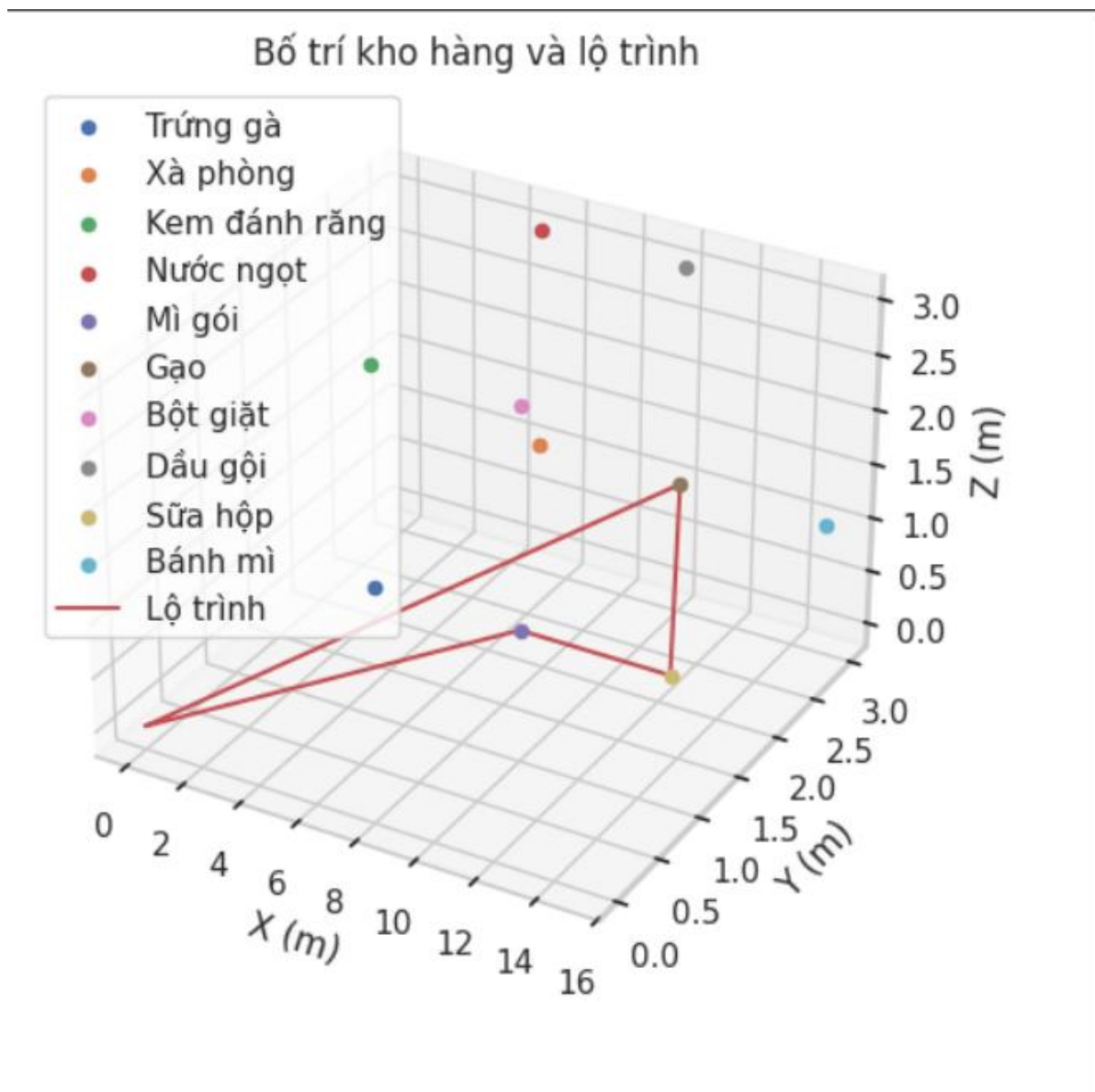
● **Kệ 2 được sử dụng hiệu quả:** Với 67.2% dung tích được sử dụng, Kệ 2 là khu vực hoạt động chính, chứa các sản phẩm thiết yếu có nhu cầu cao. Điều này cho thấy việc phân bổ sản phẩm trên kệ này đã được tối ưu để đáp ứng nhu cầu tiêu thụ.

● **Kệ 1 và Kệ 3 có tiềm năng mở rộng:**

- Kệ 1 (23.1%) còn 23.07 m<sup>3</sup> trống, có thể được sử dụng để nhập thêm các sản phẩm tiêu dùng cá nhân tương tự, duy trì nhóm sản phẩm liên quan.
- Kệ 3 (5.0%) gần như trống (28.50 m<sup>3</sup>), là nơi lý tưởng để nhập các sản phẩm mới, đặc biệt là các mặt hàng có kích thước lớn hoặc số lượng tồn kho cao

*2.4.1.2. Phân tích biểu đồ 3D: Bố trí kho hàng và lộ trình*

*a) Bố trí sản phẩm trong không gian 3D*



*Hình 2.19: Bố trí kho hàng 3D*

Biểu đồ 3D cho thấy các sản phẩm được đặt trên các kệ trong không gian với 3 trục:

- **Bậc X (m)** : Vị trí ngang của kệ, từ 0 đến 16 m, tương ứng với 3 kệ (kệ 1 ở  $x=5$ , kệ 2 ở  $x=10$ , kệ 3 ở  $x=15$ )
- **Cột Y (m)** : Chiều cao tầng trên kệ, từ 0,5 đến 3,0 m (tầng thấp  $y=1$ , tầng cao  $y=3$ ).
- **Pad Z (m)** : Chiều sâu của vị trí, từ 0,5 đến 3,0 m (gần  $z=1$ , xa  $z=3$ ).

Các sản phẩm được biểu hiện bằng các màu khác nhau:

- Trứng gà (màu xanh dương), Xà phòng (màu cam), Kem đánh răng (màu xanh lá), Nước ngọt (màu đỏ), Mì gói (màu tím), Gạo (màu nâu), Bột giặt (màu xám), Đậu phụ (màu hồng), Sữa hộp (màu vàng), Bánh mì (màu xanh lam).

Quan sát layout :

- **Kệ 1 ( $x=5$ )** : Có 4 sản phẩm (Trứng gà, Xà phòng, Kem đánh răng, Nước ngọt), đều nằm ở tầng cao ( $y=3$ ,  $z=3$ ).



- **Kệ 2 (x=10)** : Có 4 sản phẩm (Mì gói, Gạo, Bột rửa, Đậu phụ), nằm ở các tầng khác nhau: Mì gói ( $y=1, z=1$ ), Gạo ( $y=3, z=1$ ), Bột giặt ( $y=1, z=3$ ), Đậu phụ ( $y=3, z=3$ ).

- **Kệ 3 (x=15)** : Có 2 sản phẩm (Sữa hộp, Bánh mì), nằm ở tầng thấp ( $y=1, z=1$ ).

*b) Lộ trình di chuyển*

Dòng đồ trên biểu đồ 3D có thể hiển thị trình chuyển đổi để lấy hàng, kết nối các sản phẩm vị trí:

- Bắt đầu từ vị trí ( $x=5, y=3, z=3$ ) trên Kệ 1 (Trứng gà, Xà phòng, Kem đánh răng, Nước ngọt).

- Di chuyển đến Kệ 2 ( $x=10$ ), lần lượt qua các vị trí: ( $y=1, z=1$ )  $\rightarrow$  ( $y=3, z=1$ )  $\rightarrow$  ( $y=1, z=3$ )  $\rightarrow$  ( $y=3, z=3$ ), tương ứng với Mì gói, Bột, Bột, Đậu phụ.

- Kết thúc tại Kệ 3 ( $x=15, y=1, z=1$ ) để lấy Sữa hộp và Bánh mì.

**Ý nghĩa của lộ trình :**

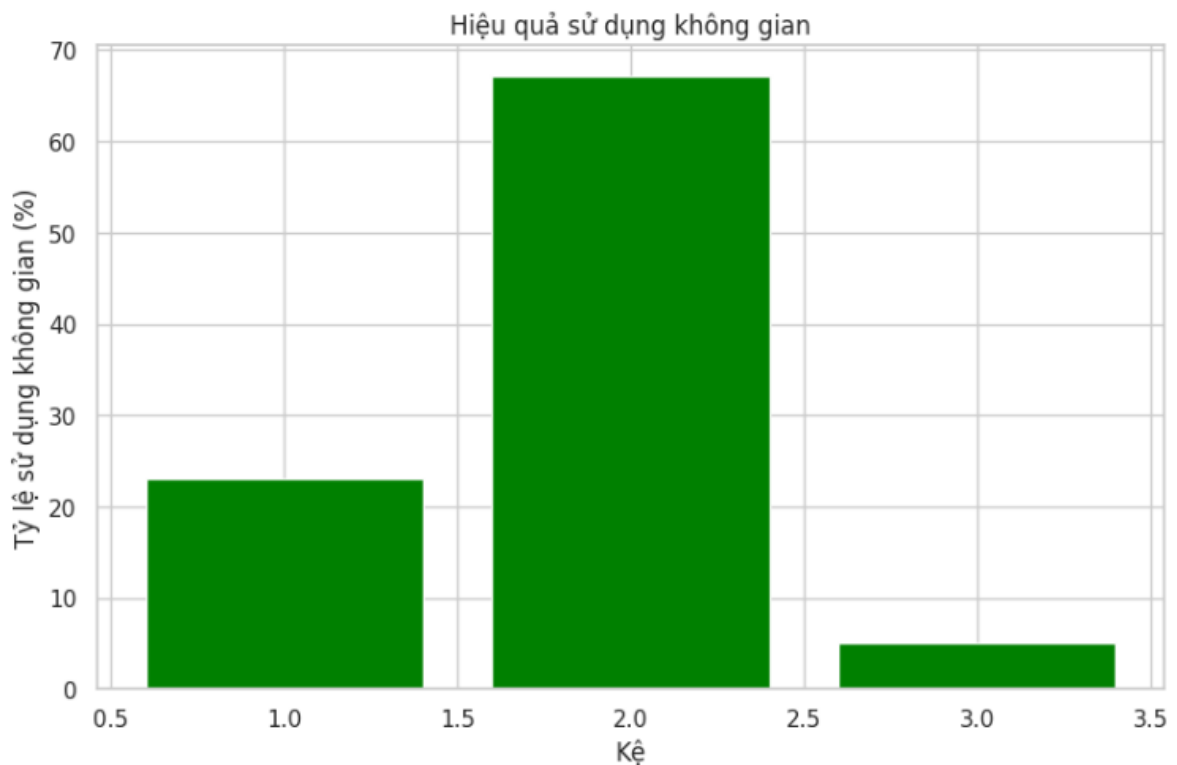
- Lộ trình này được tìm thấy cách nhân viên kho di chuyển để lấy tất cả sản phẩm trong một đơn hàng. Việc đi qua kệ kệ (Kệ 1  $\rightarrow$  Kệ 2  $\rightarrow$  Kệ 3) là hợp lý vì các kệ được đặt đều nhau ( $x=5, 10, 15$ ).

- Trên Kệ 2, lộ trình đi qua các tầng khác nhau ( $y=1$  và  $y=3$ ), cho thấy nhân viên cần chuyển lên xuống để lấy hàng, nhưng các sản phẩm được sắp xếp gần nhau trong cùng một kệ, giúp tiết kiệm thời gian.

- Kệ 1 và Kệ 3 có sản phẩm ở cùng tầng ( $y=3$  cho Kệ 1,  $y=1$  cho Kệ 3), việc lấy hàng trên kệ này khá nhanh.

*2.4.1.3. Phân tích Biểu đồ cột: Hiệu quả sử dụng không gian*

*a) Tỷ lệ sử dụng trên từng kệ*



*Hình 2.20: Hiệu quả sử dụng không gian*

Biểu đồ cột thể hiện tỷ lệ sử dụng không gian (%) trên trục Y, trục x là các kệ (Kệ 1, Kệ 2, Kệ 3):

- Kệ 1 : Tỷ lệ sử dụng 23,1%, tương ứng 6,9 m<sup>3</sup>

⇒ Trứng gà, Xà phòng, Kem đánh răng, Nước ngọt có tiêu thụ trung bình đến thấp. Khách hàng mua các sản phẩm này với tần suất không quá cao, có thể chỉ mua định kỳ hoặc khi cần.

- Kệ 2 : Tỷ lệ sử dụng 67,2%, tương ứng 20,16 m<sup>3</sup>

⇒ Mì gói, Gạo, Bột giặt, Đậu phụ có trình độ tiêu thụ cao. Khách hàng mua các sản phẩm này thường xuyên, có thể hàng ngày hoặc hàng tuần, vì họ là nhu cầu thiết yếu.

- Kệ 3 : Tỷ lệ sử dụng 5,0%, tương ứng 1,5 m<sup>3</sup>.

⇒ Sữa hộp và Bánh mì có khả năng tiêu thụ rất thấp. Khách hàng mua các sản phẩm này không thường xuyên, có thể chỉ mua theo nhu cầu thời gian hoặc không phải mặt hàng chính của kho.

#### 2.4.2. Quản lý giao hàng

```
# Dữ liệu phương tiện giao hàng (tốc độ trung bình km/h, chi phí/km)
VEHICLES = {
    "xe_may": {"speed": 40, "cost": 0.2},
    "o_to": {"speed": 60, "cost": 0.5},
    "xe_tai": {"speed": 50, "cost": 0.7}
}
```

Đây là phần khởi tạo dữ liệu ban đầu, sử dụng một từ điển (dictionary) có tên VEHICLES để lưu trữ thông tin về ba loại phương tiện giao hàng: xe máy ("xe\_may"), ô tô ("o\_to"), và xe tải ("xe\_tai").

Mỗi phương tiện được mô tả bởi hai thuộc tính:

- "speed": Tốc độ trung bình (đơn vị: km/h), lần lượt là 40 km/h cho xe máy, 60 km/h cho ô tô, và 50 km/h cho xe tải.
- "cost": Chi phí vận chuyển trên mỗi km (đơn vị tiền tệ giả định), lần lượt là 0.2, 0.5, và 0.7 cho từng loại phương tiện.

```
# Tạo dữ liệu giả lập để huấn luyện mô hình nhiều
def generate_data(num_samples=100):
    distances = np.random.uniform(5, 100, num_samples)
    traffic_factors = np.random.uniform(0.8, 1.5, num_samples)
    # Tính nhiễu thời gian (phút) do các yếu tố khác (đèn đỏ, thời tiết, v.v.)
    noise = np.random.normal(0, 5, num_samples)
    return pd.DataFrame({'distance': distances, 'traffic_factor': traffic_factors, 'noise': noise})
```

- distances: Khoảng cách giao hàng, phân bố đều từ 5 đến 100 km, đại diện cho các quãng đường thực tế.
- traffic\_factors: Yếu tố giao thông, phân bố đều từ 0.8 (giao thông thuận lợi) đến 1.5 (tắc nghẽn), ảnh hưởng đến thời gian di chuyển.
- noise: Nhiễu thời gian (phút), phân bố chuẩn với trung bình 0 và độ lệch chuẩn 5 phút, mô phỏng các yếu tố bất ngờ như đèn giao thông hoặc sự cố.

### Kết quả:

```
Tuyến đường ngắn nhất từ Hoàn Kiếm đến Cầu Giấy: Hoàn Kiếm -> Hai Bà Trưng -> Cầu Giấy
Khoảng cách: 9.00 km
Hệ số giao thông trung bình: 0.85

Thời gian yêu cầu (Deadline): 40 phút

Thời gian giao hàng dự đoán:
- xe_may: 11.06 phút (Chi phí: $1.80)
- o_to: 7.23 phút (Chi phí: $4.50)
- xe_tai: 16.20 phút (Chi phí: $6.30)

Các phương tiện giao hàng kịp thời:
- xe_may: 11.06 phút
- o_to: 7.23 phút
- xe_tai: 16.20 phút

Phương tiện được chọn: o_to (7.23 phút)
```

Lộ trình: Hoàn Kiếm → Hai Bà Trưng → Cầu Giấy.

Khoảng cách: 9.00 km.

Hệ số giao thông: 0.85.

Thời gian yêu cầu: 48 phút.

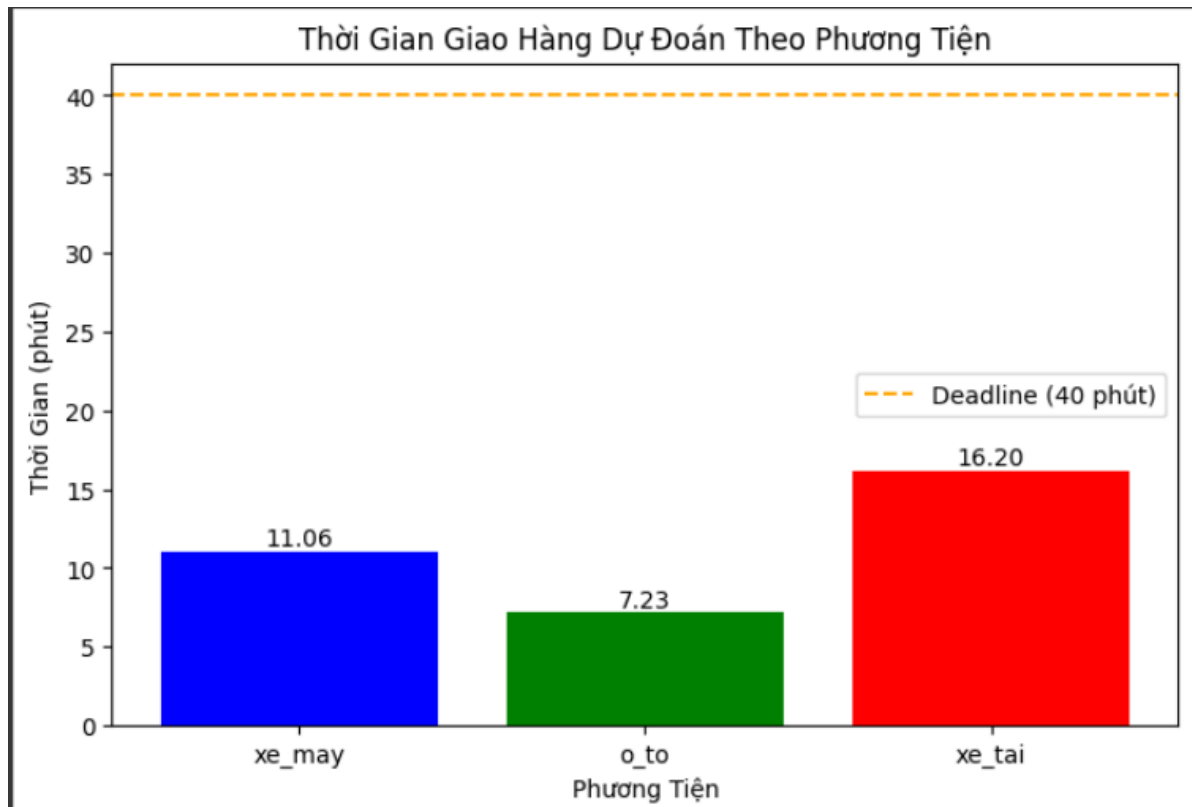
Thời gian và chi phí dự đoán:

- Xe máy: 11.06 phút (\$1.80).
- Ô tô: 7.23 phút (\$4.50).

- Xe tải: 16.20 phút (\$6.30).

Phương tiện được chọn: Ô tô (7.23 phút).

- Ô tô nhanh nhất (7.23 phút), phù hợp với yêu cầu 40 phút.
- Xe tải mất nhiều thời gian nhất (16.20 phút), có thể do tải trọng lớn hoặc tốc độ thấp hơn.
- Hệ số giao thông 0.85 cho thấy điều kiện giao thông ổn định.



Hình 2.21: Dự đoán thời gian giao hàng

- **Trục hoành (x-axis):** Đại diện cho các danh mục là các phương tiện giao hàng (xe máy, ô tô, xe tải).
- **Trục tung (y-axis):** Thể hiện giá trị định lượng, cụ thể là thời gian giao hàng tính bằng phút.
- **Cột dữ liệu:** Chiều cao của mỗi cột tương ứng với giá trị của danh mục mà nó đại diện.
- **Đường tham chiếu (deadline):** Một đường ngang biểu thị ngưỡng giới hạn là 40 phút, giúp đánh giá xem các phương tiện có đáp ứng yêu cầu thời gian hay không.

Trong lĩnh vực logistics, thời gian giao hàng là một yếu tố quan trọng ảnh hưởng đến hiệu quả vận hành và mức độ hài lòng của khách hàng. Việc sử dụng biểu đồ cột kết hợp với đường deadline cho phép dễ dàng xác định phương tiện nào hoạt động hiệu quả nhất và phù hợp với các yêu cầu cụ thể.

**Xe máy (11.06 phút):** Với thời gian giao hàng 11.06 phút, xe máy cho thấy hiệu suất nhanh, thấp hơn đáng kể so với deadline 40 phút. Điều này có thể do khả năng di

chuyển linh hoạt của xe máy trong các khu vực giao thông đô thị đông đúc, khiến nó phù hợp với các đơn hàng yêu cầu giao nhanh.

**Ô tô (7.23 phút):** Ô tô có thời gian giao hàng ngắn nhất, chỉ 7.23 phút, vượt trội so với hai phương tiện còn lại. Kết quả này có thể phản ánh ưu thế của ô tô trong việc tối ưu hóa lộ trình, chẳng hạn như sử dụng đường cao tốc hoặc vận chuyển trong điều kiện giao thông thuận lợi.

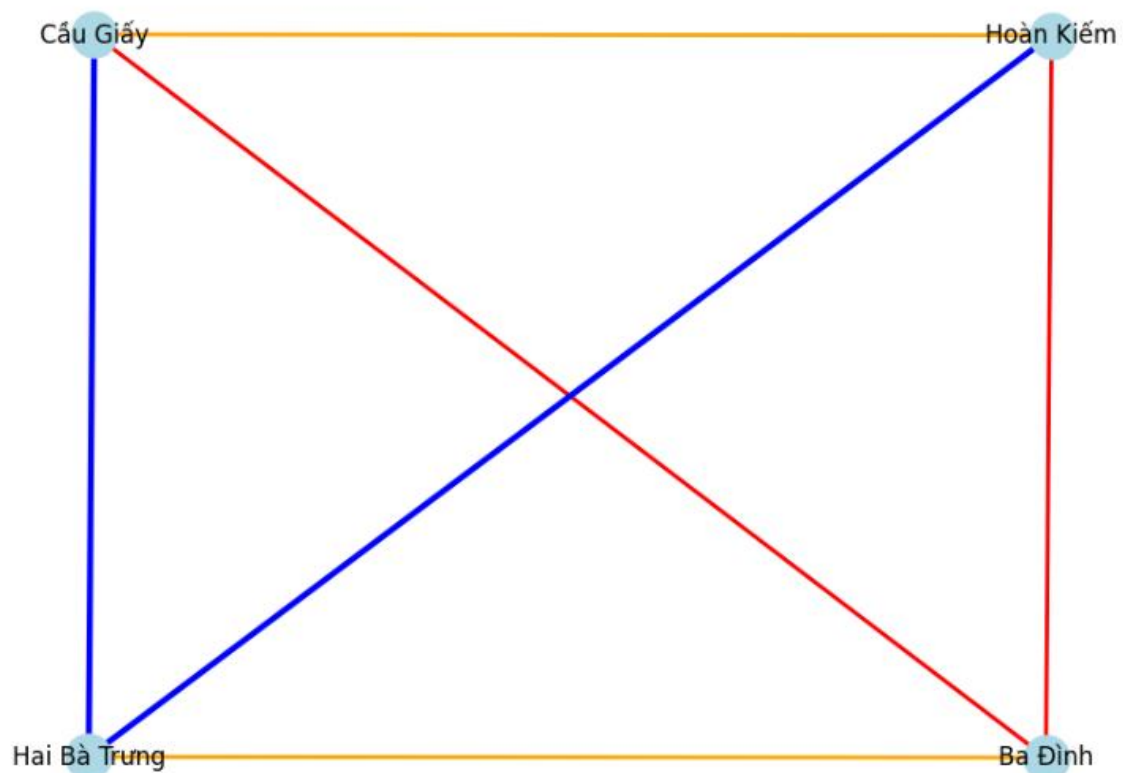
**Xe tải (16.20 phút):** Xe tải ghi nhận thời gian giao hàng 16.20 phút, cao hơn xe máy và ô tô nhưng vẫn dưới ngưỡng deadline. Thời gian dài hơn có thể xuất phát từ việc xe tải thường chở khối lượng hàng hóa lớn, dẫn đến thời gian xử lý và di chuyển tăng lên.

#### Đề xuất:

- Sử dụng ô tô cho các đơn hàng khẩn cấp hoặc lộ trình dài để tận dụng tốc độ vượt trội.
- Áp dụng xe máy cho các khu vực giao thông phức tạp hoặc đơn hàng nhỏ lẻ.
- Tối ưu hóa quy trình xử lý hàng hóa cho xe tải để giảm thời gian giao hàng, đặc biệt khi vận chuyển khối lượng lớn.
- Kết hợp dữ liệu này với các yếu tố khác như chi phí nhiên liệu, khoảng cách giao hàng và điều kiện thời tiết để đưa ra quyết định toàn diện hơn.

---

Tuyến Đường Ngắn Nhất (Màu cạnh: Xanh - Thông thoáng, Cam - Trung bình, Đỏ - Tắc nghẽn)



Hình 2.22: Tuyến đường giao hàng

- **Đỉnh (Nodes)** : Đại diện cho các điểm giao hàng hoặc địa điểm cụ thể, ở đây là Cật Giấy, Hoàn Kiếm, Hai Bà Trưng và Bà Định.

- **Cạnh (Edges)** : Giao diện cho các tuyến đường kết nối giữa các điểm, với màu sắc có thể hiện trạng thái thông tin:

- **Xanh** : Thông thoáng, thời gian di chuyển ngắn nhất.

- **Cam** : Trung bình, thời gian chuyển ở mức chấp nhận được.

- **Đỏ** : Tắc chậm, thời gian di chuyển lâu nhất.

- **Tuyến đường thông thoáng (Xanh)** :

- **Hoàn Kiếm - Hai Bà Trưng** : Đây là tuyến đường chéo màu xanh, cho thấy lộ trình này có giao thông thông thoáng, phù hợp để chuyển nhanh.

- **Cật Giấy - Hai Bà Trưng** : Tuyến đường dọc màu xanh, cũng là một lựa chọn tốt vì không bị tắc.

- **Tuyến đường trung bình (Cam)** :

- **Cật Giấy - Hoàn Kiếm** : Tuyến đường ngang trên cùng, trạng thái giao thông trung bình, có thể sử dụng nhưng không phải tối ưu.

- **Hai Bà Trưng - Bà Định** : Tuyến đường ngang dưới cùng, cũng có trạng thái giao thông trung bình.

- **Tuyến đường tắc (Đỏ)** :

- **Cật Giấy - Bà Định** : Tuyến đường chéo màu đỏ, cho thấy trình này bị tắc nghẽn nghiêm trọng, nên tránh nếu ưu tiên thời gian.

- **Hoàn Kiếm - Bà Định** : Tuyến đường dọc màu đỏ, cũng bị tắc tắc, không phải lựa chọn tốt.

**Ý nghĩa thực tiễn:** Biểu đồ cung cấp thông tin trực quan về tình trạng giao thông giữa các điểm giao hàng, giúp người quản lý logistics lựa chọn tuyến đường tối ưu. Các tuyến đường màu xanh (thông thoáng) nên được ưu tiên để giảm thời gian giao hàng và tăng hiệu quả vận hành. Các tuyến màu đỏ (tắc nghẽn) cần được tránh. Việc xác định các điểm trung chuyển như Hai Bà Trưng có thể giúp tối ưu hóa lộ trình, đặc biệt khi cần đi qua nhiều điểm.

## KẾT LUẬN VÀ KIẾN NGHỊ

Đề tài đã đạt được các kết quả sau:

- Đề tài đã đạt được những phân tích được hành vi mua hàng của khách hàng, đưa ra những phương pháp giúp hiểu biết được hành vi, thói quen của khách hàng giúp cải thiện chất lượng, danh thu đối khách hàng

- Đề tài nghiên cứu thành công tầm quan trọng và ứng dụng của khoa học dữ liệu trong marketing, không chỉ nâng cao tầm nhìn, sự hiểu biết về khách hàng mà còn xây dựng được chiến lược marketing hiệu quả, mang lại những lợi ích rõ rệt về doanh thu và trải nghiệm khách hàng. Đồng thời định hướng cho sự phát triển bền vững của các thương hiệu trong kỷ nguyên số.

- Đề tài đạt được bao gồm giảm chi phí vận hành, cải thiện thời gian giao hàng, tăng hiệu quả sử dụng tài sản, nâng cao chất lượng sản phẩm, và cải thiện sự hài lòng của khách hàng. Các phân tích dữ liệu và chiến lược tối ưu hóa đã giúp doanh nghiệp xác định các lĩnh vực ưu tiên, khắc phục các vấn đề chất lượng, và đưa ra các quyết định chiến lược để tối ưu hóa chuỗi cung ứng.

- Đề tài đã cung cấp giải pháp toàn diện để tối ưu quản lý tồn kho và giao hàng, từ phân tích hành vi khách hàng đến phân bổ không gian kho và lựa chọn phương tiện vận chuyển. Các kết quả này làm cơ sở để nâng cao hiệu quả vận hành và giảm chi phí logistics.

- Hy vọng rằng, với các kết quả đạt được, đề tài này là tài liệu tham khảo, phục vụ cho việc giảng dạy, học tập môn Thương mại điện tử được thuận tiện hơn và có thể được sử dụng trong các công ty kinh doanh online, thương mại điện tử và Logistics.

## TÀI LIỆU THAM KHẢO

- [1] Hà Anh Dũng, Trần Thị Hoàng Yến, “*Nhập môn Khoa học dữ liệu*”, 2024, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.
- [2] Barry Wilkinson, C. Michael Allen, (2004), *Parallel Programming - Techniques and Applications Using Networked Workstations and Parallel Computer*, Prentice Hall, 2nd Edition.
- [3] Joseph JaJa, (1992), *Introduction to Parallel Algorithms*, Addison-Wesley Professional, First edition.
- [4] Laura Igual, Santi Seguí, (2017), *Introduction to Data Science*, Springer International Publishing Switzerland.
- [5] Rachel Schutt, Cathy O’Neil, (2013), *Doing Data Science*, O’Reilly Media, Inc. 405 pages.



# BÁO CÁO TÓM TẮT ĐỀ TÀI NCKH SINH VIÊN

## Đề tài: NGHIÊN CỨU ỨNG DỤNG CỦA KHOA HỌC DỮ LIỆU TRONG THƯƠNG MẠI ĐIỆN TỬ

<i>SVTH</i>	<i>Trần Đức Trung</i>	<i>DHKL16A2HN</i>
<i>GVHD</i>	<i>Hà Anh Dũng</i>	<i>Khoa KHUD</i>

**1. Mục tiêu đề tài:** Đề tài nghiên cứu nhằm giải quyết các vấn đề cơ bản:

- Về lý luận: Nghiên cứu lịch sử hình thành và phát triển Thương mại điện tử.
- Nghiên cứu áp dụng phần mềm Python trong các lĩnh vực của Thương mại điện tử: Phân tích dữ liệu khách hàng, Marketing, Tối ưu hóa quy trình vận hành chuỗi cung ứng, Quản lý tồn kho và giao hàng

### 2. Nội dung nghiên cứu:

**Nội dung 1:** Nghiên cứu ứng dụng của Khoa học dữ liệu để phân tích dữ liệu khách hàng.

**Nội dung 2:** Nghiên cứu ứng dụng của Khoa học dữ liệu trong chiến lược Marketing hiệu quả.

**Nội dung 3:** Nghiên cứu ứng dụng của Khoa học dữ liệu trong tối ưu hóa quy trình vận hành chuỗi cung ứng.

**Nội dung 4:** Nghiên cứu ứng dụng của Khoa học dữ liệu trong quản lý tồn kho và giao hàng.

### 3. Kết luận và kiến nghị:

Đề tài đã tìm ra các code để ứng dụng trong Thương mại điện tử

Ý kiến xác nhận của GVHD

## PHIẾU ĐĂNG KÝ ĐỀ TÀI *(Dành cho sinh viên)*

1. Đơn vị (Cá nhân) chủ trì: Khoa khoa học ứng dụng			2. Cơ quan chủ quản: Trường Đại học Kinh tế Kỹ thuật Công nghiệp				
3. Tên đề tài: “Nghiên cứu ứng dụng của Khoa học dữ liệu trong Thương mại điện tử”							
4. Mã số thuế: 8015032143							
5. Chủ nhiệm đề tài:							
Họ và tên: Trần Đức Trung				Mã số sinh viên: 22174600059			
Khoa: Khoa học ứng dụng				Lớp: ĐHKL16A2HN			
Điện thoại: 0984456463				Email: tdtrung.dhkl16a2hn@sv.uneti.edu.vn			
6. Giảng viên hướng dẫn: Hà Anh Dũng							
Học vị: Thạc sỹ				Chức vụ, đơn vị: Giảng viên, Khoa KHƯD			
Điện thoại:0942899441				Email:hadung@uneti.edu.vn			
7. Tóm tắt nội dung đề tài:							
Nội dung 1: Phân tích dữ liệu khách hàng.							
Nội dung 2: Chiến lược Marketing hiệu quả.							
Nội dung 3: Tối ưu hóa quy trình vận hành chuỗi cung ứng.							
Nội dung 4: Quản lý tồn kho và giao hàng.							
8. Thời gian thực hiện đề tài: 10 tháng (Từ tháng 7 năm 2024 đến tháng 4 năm 2024)							
9. Khối lượng và kinh phí:							
- Khối lượng nghiên cứu khoa học: 65 giờ chuẩn (đối với người hướng dẫn)							
- Chi phí phục vụ đề tài: Theo quy định của nhà trường							
10. Loại hình nghiên cứu:			11. Lĩnh vực khoa học:				
N/C cơ bản	N/C ứng dụng	Triển khai thực nghiệm	Tự nhiên	Kỹ thuật công nghệ	Nông nghiệp	Y học	Xã hội nhân văn
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Ngày tháng năm 2024

## Chủ nhiệm đề tài

## Người hướng dẫn

**Th.S Hà Anh Dũng**  
**Khoa**

**Trần Đức Trung**

## HIỆU TRƯỞNG

Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

**THUYẾT MINH ĐỀ TÀI NGHIÊN CỨU KHOA HỌC**  
(Dành cho sinh viên)

<b>1. Tên đề tài: “Nghiên cứu ứng dụng của Khoa học dữ liệu trong Thương mại điện tử”</b>	<b>2. Mã số:</b>																									
<b>3. Thời gian thực hiện:</b> 10 tháng (Từ tháng 7 năm 2024 đến tháng 4 năm 2025)																										
<b>4. Khối lượng và kinh phí:</b> - Khối lượng nghiên cứu khoa học: 65 giờ chuẩn (đối với người hướng dẫn) - Chi phí phục vụ đề tài: Theo quy định của nhà trường																										
<b>5. Họ tên Chủ nhiệm đề tài:</b> Trần Đức Trung Khoa: Khoa học ứng dụng Lớp: ĐHKL16A2HN Điện thoại: 0984456463 Email: tdtrung.dhkl16a2hn@sv.uneti.edu.vn																										
<b>6. Giảng viên hướng dẫn: Hà Anh Dũng</b> Học hàm, học vị, chuyên môn: Thạc sỹ Chức vụ: Giảng viên Đơn vị công tác: Khoa Khoa học ứng dụng Điện thoại: 0942899441 Email: <a href="mailto:hadung@uneti.edu.vn">hadung@uneti.edu.vn</a>																										
<b>7. Cơ quan chủ quản:</b> Trường Đại học Kinh tế Kỹ thuật - Công nghiệp Địa chỉ: 456 Minh Khai - Hai Bà Trưng - Hà Nội Điện thoại: 04.38621504; Fax: 04.38623938;																										
<b>8. Đơn vị phối hợp chính:</b> Khoa khoa học ứng dụng																										
<b>9. Danh sách những sinh viên tham gia:</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;">TT</th> <th style="width: 25%;">Họ và tên</th> <th style="width: 20%;">Mã sinh viên</th> <th style="width: 15%;">Lớp</th> <th style="width: 35%;">Nội dung nghiên cứu</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">1</td> <td>Trần Đức Trung</td> <td>22174600059</td> <td>ĐHKL16A2HN</td> <td><b>Nội dung 1:</b> Phân tích dữ liệu khách hàng.</td> </tr> <tr> <td style="text-align: center;">2</td> <td>Đậu Thị Thảo</td> <td>22174600003</td> <td>ĐHKL16A2HN</td> <td><b>Nội dung 2:</b> Chiến lược Marketing hiệu quả.</td> </tr> <tr> <td style="text-align: center;">3</td> <td>Khuất Thanh Phương</td> <td>22174600105</td> <td>ĐHKL16A2HN</td> <td><b>Nội dung 3:</b> Tối ưu hóa quy trình vận hành chuỗi cung ứng.</td> </tr> <tr> <td style="text-align: center;">4</td> <td>Đào Xuân Tâm</td> <td>22174600106</td> <td>ĐHKL16A2HN</td> <td><b>Nội dung 4:</b> Quản lý tồn kho và giao hàng.</td> </tr> </tbody> </table>		TT	Họ và tên	Mã sinh viên	Lớp	Nội dung nghiên cứu	1	Trần Đức Trung	22174600059	ĐHKL16A2HN	<b>Nội dung 1:</b> Phân tích dữ liệu khách hàng.	2	Đậu Thị Thảo	22174600003	ĐHKL16A2HN	<b>Nội dung 2:</b> Chiến lược Marketing hiệu quả.	3	Khuất Thanh Phương	22174600105	ĐHKL16A2HN	<b>Nội dung 3:</b> Tối ưu hóa quy trình vận hành chuỗi cung ứng.	4	Đào Xuân Tâm	22174600106	ĐHKL16A2HN	<b>Nội dung 4:</b> Quản lý tồn kho và giao hàng.
TT	Họ và tên	Mã sinh viên	Lớp	Nội dung nghiên cứu																						
1	Trần Đức Trung	22174600059	ĐHKL16A2HN	<b>Nội dung 1:</b> Phân tích dữ liệu khách hàng.																						
2	Đậu Thị Thảo	22174600003	ĐHKL16A2HN	<b>Nội dung 2:</b> Chiến lược Marketing hiệu quả.																						
3	Khuất Thanh Phương	22174600105	ĐHKL16A2HN	<b>Nội dung 3:</b> Tối ưu hóa quy trình vận hành chuỗi cung ứng.																						
4	Đào Xuân Tâm	22174600106	ĐHKL16A2HN	<b>Nội dung 4:</b> Quản lý tồn kho và giao hàng.																						
<b>10. Giới thiệu ý tưởng nghiên cứu:</b>																										

Khoa học dữ liệu là một lĩnh vực liên ngành kết hợp các kiến thức từ toán học, thống kê, khoa học máy tính, và các chuyên ngành khác để phân tích và khai thác thông tin từ dữ liệu. Trong bối cảnh TMĐT, việc ứng dụng KHDL giúp biến dữ liệu thô thành thông tin có giá trị, từ đó hỗ trợ quá trình ra quyết định một cách chính xác và kịp thời. Các doanh nghiệp có thể sử dụng kỹ thuật phân tích dữ liệu để hiểu rõ hơn về hành vi mua sắm, từ đó đưa ra các chiến lược tiếp thị và bán hàng hiệu quả hơn. Điều này không chỉ giúp tăng doanh thu mà còn nâng cao trải nghiệm khách hàng, từ đó tạo ra lợi thế cạnh tranh bền vững. Không chỉ dừng lại ở việc cải thiện hiệu quả hoạt động hiện tại, KHDL còn giúp các doanh nghiệp TMĐT dự đoán và đón đầu các xu hướng tương lai. Phân tích dự báo, dựa trên dữ liệu lịch sử và các yếu tố ảnh hưởng đến thị trường, giúp các doanh nghiệp chuẩn bị và điều chỉnh chiến lược một cách linh hoạt và kịp thời.

Trong học phần *Nhập môn khoa học dữ liệu* đã giúp sinh viên ngành được tiếp cận với các phương pháp và công cụ hiện đại trong phân tích dữ liệu, hiểu được tầm quan trọng của việc khai thác dữ liệu hiệu quả và áp dụng chúng vào các bài toán thực tế trong ngành TMĐT. Cũng chính vì những lý do trên, nhóm chúng em đã lựa chọn đề tài “*Nghiên cứu ứng dụng của Khoa học dữ liệu trong Thương mại điện tử*” làm đề tài nghiên cứu khoa học sinh viên dưới sự hướng dẫn của thầy Hà Anh Dũng.

## **11. Mục tiêu, đối tượng, phạm vi của đề tài:**

### ***Mục tiêu:***

- Đề tài tiến hành phân tích vai trò và tầm quan trọng của KHDL trong ngành thương mại điện tử. Khám phá cách KHDL giúp các doanh nghiệp thương mại điện tử chuyển đổi dữ liệu thô thành thông tin có giá trị. Đánh giá tầm quan trọng của KHDL trong việc nâng cao hiệu quả hoạt động kinh doanh và tạo lợi thế cạnh tranh.
- Khảo sát các ứng dụng cụ thể của KHDL trong các hoạt động khác nhau của thương mại điện tử: tìm hiểu và phân tích hành vi người dùng, gợi ý sản phẩm và dự đoán mua hàng, quản lý kho và chuỗi cung ứng, phân tích dữ liệu khách hàng và tiếp thị.

### ***Đối tượng nghiên cứu:***

- Các doanh nghiệp hoạt động trong lĩnh vực thương mại điện tử từ quy mô nhỏ đến quy mô lớn.
- Những người sử dụng dịch vụ sản phẩm của các nền tảng thương mại điện tử. Họ là những cá nhân có nhu cầu mua sắm trực tuyến và tương tác với các hệ thống thương mại điện tử.
- Các công nghệ và hệ thống hỗ trợ việc thu thập, phân tích dữ liệu trong lĩnh vực thương mại điện tử. Điều này bao gồm các công cụ như Big Data, trí tuệ nhân tạo (AI) và các hệ thống quản lý quan hệ khách hàng (CRM).

***Phạm vi nghiên cứu:*** Tập trung vào các ứng dụng khoa học dữ liệu trong thương mại điện tử từ đầu thế kỷ 21 đến hiện tại, trên quy mô toàn cầu. Bao gồm các khía cạnh như marketing, quản lý khách hàng, quản lý chuỗi cung ứng, quản lý tồn kho và giao hàng.

## **12. Cơ sở lý thuyết và lịch sử nghiên cứu:**

✓ **Cơ sở lý luận về Khoa học dữ liệu: Thông kê toán học, học máy, trí tuệ nhân tạo.**

- Các phần mềm được sử dụng: Python, R, Excel...
- Các thư viện được sử dụng: Numpy, Pandas, Matplotlib...
- Các frameworks được sử dụng: Scikit-learn, TensorFlow, PyTorch...

✓ **Lịch sử nghiên cứu**

Thương mại điện tử xuất hiện lần đầu vào những năm 1970 với sự ra đời của các hệ thống trao đổi dữ liệu điện tử (EDI). Tuy nhiên, sự bùng nổ thực sự của TMĐT chỉ bắt đầu vào cuối những năm 1990 và đầu những năm 2000, khi Internet trở nên phổ biến và các công nghệ thanh toán trực tuyến được phát triển. Các nền tảng TMĐT như Amazon, eBay, và Alibaba đã mở ra một kỷ nguyên mới cho mua sắm trực tuyến, thay đổi cách thức người tiêu dùng mua sắm và doanh nghiệp kinh doanh. Tuy nhiên ban đầu dữ liệu chủ yếu là các thông tin giao dịch đơn giản và phản hồi từ khách hàng. Đến những năm 2000, các công cụ thu thập và phân tích dữ liệu cơ bản đã giúp các doanh nghiệp hiểu hơn về hành vi mua sắm của khách hàng.

Sự phát triển mạnh mẽ của khoa học dữ liệu bắt đầu từ những năm 2010, khi các công cụ và phương pháp phân tích dữ liệu tiên tiến như Machine Learning và Big Data bắt đầu được áp dụng trong thương mại điện tử. Các công ty lớn như Google và Facebook đã tận dụng sức mạnh của dữ liệu để cá nhân hóa trải nghiệm người dùng. Sự phát triển của các phần mềm và công cụ như Hadoop, Spark, cùng các ngôn ngữ lập trình như Python và R đã làm cho việc phân tích dữ liệu trở nên dễ dàng và hiệu quả hơn.

Những năm 2020 chứng kiến sự bùng nổ của dữ liệu lớn và ứng dụng rộng rãi của khoa học dữ liệu trong thương mại điện tử. Khoa học dữ liệu đóng vai trò quan trọng trong việc tối ưu hóa các chiến lược kinh doanh, dự đoán nhu cầu, quản lý chuỗi cung ứng và nâng cao trải nghiệm khách hàng. Các công nghệ mới như trí tuệ nhân tạo (AI) và học máy (Machine Learning) giúp phân tích và dự đoán hành vi của khách hàng một cách chính xác hơn, từ đó cải thiện dịch vụ và tăng cường doanh số.

Hiện nay, khoa học dữ liệu được áp dụng rộng rãi trong nhiều khía cạnh của thương mại điện tử như tiếp thị, bán hàng, dịch vụ khách hàng, và quản lý vận hành. Các công ty thương mại điện tử lớn như Amazon và Alibaba sử dụng khoa học dữ liệu để tối ưu hóa mọi mặt hoạt động của họ. Trong tương lai, sự phát triển liên tục của công nghệ, sự tích hợp sâu hơn giữa AI và khoa học dữ liệu, cùng với sự gia tăng về khối lượng dữ liệu sẽ tiếp tục thúc đẩy sự tiến bộ và thay đổi trong ngành thương mại điện tử.

**13. Phương pháp nghiên cứu :**

- Phương pháp nghiên cứu tài liệu: Sử dụng các tài liệu do thầy hướng dẫn đưa ra để có cơ sở lý thuyết đơn giản và dễ hiểu nhất.
- Phương pháp vận dụng thực tế: Sử dụng ngôn ngữ lập trình Python, R để viết các code ứng dụng trong Thương mại điện tử.
- Phương pháp phân tích, tổng hợp, đánh giá.

<b>14. Nội dung nghiên cứu:</b> <b>Nội dung 1:</b> Phân tích dữ liệu khách hàng. <b>Nội dung 2:</b> Chiến lược Marketing hiệu quả. <b>Nội dung 3:</b> Tối ưu hóa quy trình vận hành chuỗi cung ứng. <b>Nội dung 4:</b> Quản lý tồn kho và giao hàng.					
<b>15. Dạng sản phẩm, kết quả tạo ra: Báo cáo khoa học</b>					
<b>15. Yêu cầu khoa học đối với sản phẩm kết quả tạo ra</b>					
STT	Tên sản phẩm	Số lượng	Mô tả đặc điểm	Yêu cầu khoa học	Địa chỉ áp dụng
1	Báo cáo khoa học	1	Bản mềm và bản cứng	Báo cáo đảm bảo tính logic cao giữa các nội dung. - Thông tin phong phú, đa dạng, cập nhật, có nguồn gốc rõ ràng, chính xác. - Giải pháp có tính khả thi cao, có khả năng vận dụng thực tế.	Khoa KHUĐ và KDTM Trường Đại học Kinh tế Kỹ thuật Công nghiệp
2	Bài tóm tắt kỹ yếu (Mẫu KHCN – 4a)				
<b>16. Tiến độ thực hiện:</b>					
Stt	Nội dung công việc	Kết quả đạt được	Thời gian bắt đầu, kết thúc	Người, đơn vị thực hiện	
1	Xây dựng đề cương	Đề cương chi tiết	7/2024-8/2024	Nhóm sinh viên	
2	Phân tích dữ liệu khách hàng	Các nội dung, code và ví dụ về phân tích hành vi khách hàng, lịch sử mua hàng.	9/2024-1/2025	Trần Đức Trung	
3	Chiến lược Marketing hiệu quả	Các nội dung, code và ví dụ về chiến lược Marketing hàng hóa.	9/2024-1/2025	Đậu Thị Thảo	
4	Tối ưu hóa quy trình vận hành chuỗi cung ứng	Các nội dung, code và ví dụ về Tối ưu hóa quy trình vận hành chuỗi cung ứng.	9/2024-1/2025	Khuất Thanh Phương	
5	Quản lý tồn kho và giao hàng	Các nội dung, code và ví dụ về Quản lý tồn kho và giao hàng.	9/2024-1/2025	Đào Xuân Tâm	

6	Viết báo cáo	Hoàn thành báo cáo	02/2025-04/2025	Nhóm sinh viên
7	Nghiệm thu đề tài	Đề tài nghiệm thu	04/2025	Nhóm sinh viên
<b>17. Dự toán kinh phí thực hiện đề tài</b>				
Mục	Nội dung	Số tiền/ giờ chuẩn	Giải trình chi tiết	Ghi chú
<b>1. Công hướng dẫn</b>				
	Chuyên đề 1:	15	- Phân tích dữ liệu khách hàng.	
	Chuyên đề 2:	15	- Chiến lược Marketing hiệu quả.	
	Chuyên đề 3:	20	- Tối ưu hóa quy trình vận hành chuỗi cung ứng.	
	Chuyên đề 4:	15	- Quản lý tồn kho và giao hàng.	
<b>2. Hàng hóa dịch vụ: Không</b>				
<b>3. Chi khác:</b>				
	Chi hội đồng nghiệm thu		Theo qui chế chi tiêu nội bộ	

Ngày      tháng      năm 2024

**Người hướng dẫn**

**Chủ nhiệm đề tài**

**Hà Anh Dũng**

**Trần Đức Trung**

**Khoa**

Ngày ....tháng ....năm....

**HIỆU TRƯỞNG**

Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

**BIÊN BẢN KIỂM TRA  
TIẾN ĐỘ THỰC HIỆN NHIỆM VỤ NCKH SINH VIÊN NĂM 2025**

Căn cứ Quyết định số ..... /QĐ-ĐHKTKTCN ngày ..... tháng ..... năm 2024 về việc giao nhiệm vụ nghiên cứu khoa học sinh viên năm 2024 – 2025

Căn cứ thông báo số ..... về việc kiểm tra tiến độ thực hiện nhiệm vụ nghiên cứu khoa học sinh viên năm 2025

*Hôm nay, ngày 05 tháng 03 năm 2025*

Tại trường Đại học Kinh tế - Kỹ thuật Công nghiệp tiến hành kiểm tra tiến độ thực hiện đề tài NCKH sinh viên năm 2025

**Thành phần kiểm tra gồm:**

- 1 TS. Trần Thị Hoàng Yến
- 2 TS. Chu Bình Minh
- 3 TS. Lê Xuân Huy
- 4 ThS. Trần Thị Hà

**Nội dung kiểm tra:**

- Các nội dung đã thực hiện đến ngày báo cáo, kết quả đạt được;
- Tình hình giải ngân kinh phí thực hiện nhiệm vụ;
- Những vướng mắc phát sinh (nếu có), hướng xử lý;
- Kế hoạch triển khai các nội dung còn lại của nhiệm vụ để đảm bảo hoàn thành nhiệm vụ theo đúng tiến độ.
- Kiến nghị (nếu có).

**Kết quả kiểm tra như sau:**

Tên đề tài: “Nghiên cứu ứng dụng của Khoa học dữ liệu trong Thương mại điện tử”

Chủ nhiệm đề tài : Trần Đức Trung

Giảng viên hướng dẫn: ThS. Hà Anh Dũng

Các nội dung :

1. Các nội dung đã thực hiện đến ngày báo cáo, kết quả đạt được:
- Phân tích dữ liệu khách hàng;



- Chiến lược Marketing hiệu quả;
- Tối ưu hóa quy trình vận hành chuỗi cung ứng;
- Quản lý tồn kho và giao hàng.

2. Tình hình giải ngân kinh phí thực hiện nhiệm vụ: Không kinh phí.

3. Những vướng mắc phát sinh (nếu có), hướng xử lý:

Không có.

4. Kế hoạch triển khai các nội dung còn lại của nhiệm vụ để đảm bảo hoàn thành nhiệm vụ theo đúng tiến độ:

- Hoàn thiện viết báo cáo.

5. Kiến nghị (nếu có):

Không có.

Biên bản kiểm tra hoàn thành vào hồi 15h30 ngày 05 tháng 03 năm 2025

**Phòng KH-CN**

**Khoa**

**GV hướng dẫn**

**Chủ nhiệm đề tài**

**Hà Anh Dũng**

**Trần Đức Trung**

*(Ký, ghi họ tên)*

*(Ký, ghi họ tên)*

*(Ký, ghi rõ họ tên)*

*(Ký, ghi rõ họ tên)*

*Hà Nội, ngày 05 tháng 03 năm 2025*

**BÁO CÁO TÌNH HÌNH THỰC HIỆN  
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

**I. Thông tin chung:**

1. Tên đề tài: **“Nghiên cứu ứng dụng của Khoa học dữ liệu trong Thương mại điện tử”**

2. Họ, tên chủ nhiệm đề tài và các thành viên tham gia

- Chủ nhiệm đề tài: Trần Đức Trung

Mã số SV: 22174600059

Lớp: ĐHKL16A2HN

Khoa: Khoa học ứng dụng

Điện thoại: 0984456463

Email:

[tdtrung.dhkl16a2hn@sv.uneti.edu.vn](mailto:tdtrung.dhkl16a2hn@sv.uneti.edu.vn)

- Thành viên tham gia:

1) Đậu Thị Thảo

Lớp: ĐHKL16A2HN

Khoa: Khoa học ứng dụng

2) Khuất Thanh Phương

Lớp: ĐHKL16A2HN

Khoa: Khoa học ứng dụng

3) Đào Xuân Tâm

Lớp: ĐHKL16A2HN

Khoa: Khoa học ứng dụng

3. Giảng viên hướng dẫn: Hà Anh Dũng

- Học vị: Thạc sỹ

Chức vụ: Giảng viên

- Điện thoại: 0942899441

Email: [hadung@uneti.edu.vn](mailto:hadung@uneti.edu.vn)

4. Thời gian thực hiện: 10 tháng, từ tháng 08 năm 2024 đến tháng 04 năm 2025

**II. Đánh giá tình hình thực hiện đề tài:**

1. Nội dung nghiên cứu đã thực hiện và sản phẩm đã có

STT	Tiến độ và thời gian	Nội dung giảng viên đã hướng dẫn	Nội dung nghiên cứu đã thực hiện	Sinh viên thực hiện	Sản phẩm	Giảng viên hướng dẫn đánh giá
1	Tháng 8/2024	Xây dựng đề cương.	Đề cương chi tiết.	Xác định mục đích và hướng nghiên cứu.	Đề cương chi tiết.	Hoàn thành tốt
2	Tháng 9-01/2025	Phân tích dữ liệu khách hàng.	Các nội dung, code và ví dụ về phân tích hành vi khách hàng, lịch sử mua hàng.	Trần Đức Trung	Các nội dung, code và ví dụ về phân tích hành vi khách hàng, lịch sử mua hàng.	Hoàn thành tốt
3	Tháng 9-01/2025	Chiến lược Marketing hiệu quả	Các nội dung, code và ví dụ về chiến lược Marketing hàng hóa.	Đậu Thị Thảo	Các nội dung, code và ví dụ về chiến lược Marketing hàng hóa.	Hoàn thành tốt
4	Tháng 9-01/2025	Tối ưu hóa quy trình vận hành chuỗi cung ứng	Các nội dung, code và ví dụ về Tối ưu hóa quy trình vận hành chuỗi cung ứng.	Khuất Thanh Phương	Các nội dung, code và ví dụ về Tối ưu hóa quy trình vận hành chuỗi cung ứng.	Hoàn thành tốt
5	Tháng 9-01/2025	Quản lý tồn kho và giao hàng	Các nội dung, code và ví dụ về Quản lý tồn kho và giao hàng	Đào Xuân Tâm	Các nội dung, code và ví dụ về Quản lý tồn kho và giao hàng.	Hoàn thành tốt

**2. Nội dung chưa hoàn thành (theo đăng kí, nguyên nhân) :** Không có

**3. Kinh phí đã chi cho các nội dung (Nếu có):** Không có

### III. Kế hoạch thực hiện tiếp theo:

STT	Tiến độ và thời gian	Nội dung giảng viên sẽ hướng dẫn	Nội dung thực hiện	Sinh viên thực hiện	Sản phẩm	Ghi chú
6	Tháng 2-4/2025	Hoàn thiện viết báo cáo.	Hoàn thiện viết báo cáo.	Hoàn thiện viết báo cáo.	Báo cáo NCKH sinh viên.	

**IV. Kiến nghị:** Không có

**Khoa**  
(ký, ghi họ tên)

**Giảng viên hướng dẫn**  
(ký, ghi họ tên)

**Chủ nhiệm đề tài**  
(ký, ghi họ tên)

**Hà Anh Dũng**

**Trần Đức Trung**

Số: 854/QĐ - ĐHKTKTCN

Hà Nội, ngày 30 tháng 09 năm 2024

### **QUYẾT ĐỊNH**

Về việc giao nhiệm vụ nghiên cứu khoa học sinh viên năm học 2024 - 2025

#### **HIỆU TRƯỞNG TRƯỜNG ĐẠI HỌC KINH TẾ - KỸ THUẬT CÔNG NGHIỆP**

Căn cứ Quyết định số 7586/QĐ - BCT ngày 25 tháng 8 năm 2014 của Bộ trưởng Bộ Công Thương quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của trường Đại học Kinh tế - Kỹ thuật Công nghiệp;

Căn cứ Nghị quyết số 33/NQ-HĐTĐHKTKTCN ngày 03/10/2023 về việc sửa đổi, bổ sung một số điều Quy chế tổ chức và hoạt động Trường ĐHKTKTCN ban hành kèm theo Nghị quyết số 04/NQ-HĐTĐHKTKTCN ngày 04/5/2022 của Hội đồng trường Trường Đại học Kinh tế - Kỹ thuật Công nghiệp;

Căn cứ Quyết định số 801/QĐ-ĐHKTKTCN ngày 17/9/2024 về việc sửa đổi, bổ sung một số điều của Quy định về hoạt động Khoa học - Công nghệ của Trường Đại học Kinh tế - Kỹ thuật Công nghiệp ban hành kèm theo Quyết định số 722/QĐ-ĐHKTKTCN ngày 24/8/2023;

Căn cứ Quyết định số 838/QĐ-ĐHKTKTCN ngày 25/09/2024 của Hiệu trưởng Trường ĐHKTKTCN về việc ban hành Quy chế Chi tiêu nội bộ của Trường Đại học Kinh tế - Kỹ thuật Công nghiệp;

Theo đề nghị của phòng Khoa học - Công nghệ;

### **QUYẾT ĐỊNH:**

**Điều 1.** Giao nhiệm vụ nghiên cứu khoa học sinh viên năm học 2024-2025 cho các đơn vị, cá nhân thực hiện đề tài (có phụ lục kèm theo);

**Điều 2.** Trường các đơn vị được giao kế hoạch chịu trách nhiệm tổ chức và phối hợp thực hiện các nhiệm vụ khoa học công nghệ theo đúng nội dung, tiến độ và các quy định hiện hành;

**Điều 3.** Các đơn vị có liên quan và các giảng viên, sinh viên có tên trong danh sách chịu trách nhiệm thi hành quyết định này./.

Nơi nhận:

- BGH, HĐT
- Như điều 3;
- Lưu VT, KH-CN

KT. HIỆU TRƯỞNG  
PHÓ HIỆU TRƯỞNG



PGS.TS. Nguyễn Hữu Quang



**DANH SÁCH ĐĂNG KÝ ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN NĂM HỌC 2024-2025**  
(Kèm theo Quyết định số 854/QĐ-DHKTCTCN ngày 30 tháng 09 năm 2024 của Hiệu trưởng)

*Trường Đại học Kinh tế - Kỹ thuật Công nghiệp*

STT		TRƯỜNG ĐẠI HỌC KHOA HỌC CÔNG NGHỆ		Tên đề tài		Sinh viên thực hiện		Lớp		Chức vụ trong đề tài		Giáo viên hướng dẫn		Ghi chú	
KHOA KHOA HỌC ỨNG DỤNG (11)															
1	Nghiên cứu ứng dụng của Khoa học dữ liệu trong Thương mại điện tử	Trần Đức Trung		DHKL16A2HN		Chủ nhiệm DT						ThS. Hà Anh Dũng			
		Đậu Thị Thảo		DHKL16A2HN		Thành viên									
		Khuyết Thanh Phương		DHKL16A2HN		Thành viên									
		Đào Xuân Tâm		DHKL16A2HN		Thành viên									
		Ngô Công Hiếu		DHTM17A1CL		Chủ nhiệm DT						ThS. Nguyễn Thị Hạnh Lê			
2	Tìm khoảng chấp nhận cho nghiệm tối ưu của bài toán quy hoạch tuyến tính tham số	Trần Huy Trường		DHTM17A1CL		Thành viên						ThS. Cao Thị Thanh Xuân			
		Hồ Thị Minh Hằng		DHKL16A2HN		Chủ nhiệm DT									
3	Sử dụng các mô hình học máy để dự đoán khả năng chấp nhận khoản vay của khách hàng	Trần Đức Lương		DHKL16A2HN		Thành viên						CN. Lê Hằng Anh			
		Phạm Thị Ngọc Tú		DHKL16A2HN		Thành viên									
		Nguyễn Ngọc Bắc		DHKL16A2HN		Thành viên									
		Nguyễn Tuấn Hùng		DHKL16A1HN		Chủ nhiệm DT									
4	Xây dựng mô hình CNN sử dụng ngôn ngữ python cho bài toán phân loại ảnh X-quang ngực	Trần Mạnh Hùng		DHKL16A1HN		Thành viên						CN. Lê Thanh Sơn			
		Trương Anh Đức		DHKL16A1HN		Thành viên						CN. Nguyễn Anh Thư			
		Phạm Lê Nhật Minh		DHKL17A1HN		Chủ nhiệm DT									
5	Áp dụng mô hình GLM vào dự đoán khả năng xuất hiện rối loạn giấc ngủ	Phạm Duy Khải		DHKL17A1HN		Thành viên						CN. Trần Bảo Trung			
		Trần Hoàng Bách		DHKL17A1HN		Thành viên									