

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**TRƯỜNG ĐẠI HỌC ĐẠI NAM**



**BÀI TẬP LỚN**

**TÊN HỌC PHẦN: XÁC SUẤT THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU**

**ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU TÀI CHÍNH CÁ NHÂN**

**Giáo viên hướng dẫn: ThS. Lê Diệu Anh**

**Sinh viên thực hiện:**

STT	Mã sv	Họ và tên	Lớp
1	1771020717	Vương Đức Tuấn	CNTT 17-15
2	1771020650	Nguyễn Quang Thịnh	CNTT 17-15
3	1771020743	Dương Văn Việt	CNTT 17-15

**Hà Nội, năm 2025**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC ĐẠI NAM**



**BÀI TẬP LỚN**

**TÊN HỌC PHẦN: XÁC SUẤT THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU**

**ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU TÀI CHÍNH CÁ NHÂN**

STT	Mã Sinh Viên	Họ và Tên	Ngày Sinh	Điểm	
				Bảng Số	Bảng Chữ
1	1771020717	Vương Đức Tuấn	02/11/2005		
2	1771020650	Nguyễn Quang Thịnh	08/07/2005		
3	1771020743	Dương Văn Việt	02/10/2005		

**CÁN BỘ CHẤM THI**

**Hà Nội, năm 2025**

## LỜI NÓI ĐẦU

Trong cuộc sống hiện đại, quản lý tài chính cá nhân là một yếu tố vô cùng quan trọng đối với mọi hộ gia đình. Việc hiểu rõ về thu nhập và chi tiêu không chỉ giúp gia đình duy trì một cuộc sống ổn định mà còn tạo điều kiện để tiết kiệm, đầu tư và đạt được các mục tiêu tài chính lâu dài. Tuy nhiên, trong thực tế, không phải ai cũng có khả năng quản lý tài chính hiệu quả. Các hộ gia đình thường gặp phải khó khăn trong việc phân bổ ngân sách cho các khoản chi tiêu thiết yếu và không thiết yếu.

Đề tài "Phân tích dữ liệu tài chính cá nhân để tìm hiểu xu hướng chi tiêu của hộ gia đình" nhằm mục đích cung cấp cái nhìn sâu sắc về cách các hộ gia đình chi tiêu hàng tháng, từ đó phát hiện ra những mô hình và xu hướng chi tiêu phổ biến. Qua đó, chúng ta có thể nhận diện được những yếu tố ảnh hưởng lớn đến hành vi chi tiêu, chẳng hạn như thu nhập, số lượng thành viên trong gia đình, ngành nghề và các yếu tố xã hội khác.

Việc phân tích dữ liệu tài chính không chỉ giúp nhận ra các vấn đề trong việc quản lý chi tiêu mà còn đóng vai trò quan trọng trong việc xây dựng một chiến lược tài chính cá nhân hợp lý. Đồng thời, từ những phân tích này, chúng ta cũng có thể đưa ra các khuyến nghị giúp các hộ gia đình tối ưu hóa chi tiêu, cân đối thu nhập và chi phí, hướng đến một cuộc sống tài chính bền vững.

Với dự án này, chúng ta không chỉ hiểu được những yếu tố ảnh hưởng đến chi tiêu mà còn phát hiện ra các thói quen liên quan đến tài chính để thay đổi và cải thiện tình hình tài chính của các hộ gia đình, từ đó góp phần nâng cao chất lượng sống cho mọi người.

# MỤC LỤC

LỜI NÓI ĐẦU .....	3
MỤC LỤC HÌNH ẢNH .....	5
BẢNG CÁC TỪ VIẾT TẮT .....	6
CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI .....	7
1.1. Giới thiệu chung về đề tài lựa chọn .....	7
1.2. Lý do chọn đề tài .....	8
1.3. Mục tiêu .....	9
1.4. Phạm vi nghiên cứu của đề tài .....	9
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT .....	11
2.1. Trình bày các khái niệm thống kê .....	11
2.2. Các phương pháp phân tích dữ liệu tài chính cá nhân .....	13
2.3. Các yếu tố ảnh hưởng đến xu hướng chi tiêu của hộ gia đình .....	15
2.4. Hồi quy tuyến tính .....	16
2.5. Thu thập dữ liệu .....	18
CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU .....	19
3.1. Giới thiệu về dữ liệu .....	19
3.2. Thu thập dữ liệu .....	19
3.3 Chuẩn bị và làm sạch dữ liệu .....	20
3.4 Phân tích mô tả dữ liệu .....	22
3.5 Kiểm tra giả thuyết .....	23
3.6 Mô hình hóa dữ liệu .....	23
3.7 Đánh giá mô hình và kiểm tra độ tin cậy .....	28
3.8 Kết luận .....	28
KẾT LUẬN .....	29
PHÂN CÔNG NHIỆM VỤ .....	32
DANH MỤC TÀI LIỆU THAM KHẢO .....	34

## MỤC LỤC HÌNH ẢNH

<i>Hình 3.1</i> .....	21
<i>Hình 3.2</i> .....	22
<i>Hình 3.3</i> .....	24
<i>Hình 3.4</i> .....	25
<i>Hình 3.5</i> .....	26
<i>Hình 3.6</i> .....	27

## **BẢNG CÁC TỪ VIẾT TẮT**

<b>STT</b>	<b>TỪ VIẾT TẮT</b>	<b>VIẾT ĐẦY ĐỦ</b>
<b>1</b>	<b>XSTK</b>	<b>Xác suất thống kê</b>

# CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

## 1.1. Giới thiệu chung về đề tài lựa chọn

- Trong xã hội hiện đại, quản lý tài chính cá nhân đóng vai trò quan trọng không chỉ đối với cá nhân mà còn đối với các hộ gia đình. Việc hiểu rõ về cách thức chi tiêu và tiết kiệm là một yếu tố quyết định đến sự ổn định tài chính của mỗi gia đình. Tuy nhiên, trong bối cảnh chi phí sinh hoạt ngày càng gia tăng, thu nhập không ổn định, và các yếu tố xã hội luôn biến động, nhiều hộ gia đình gặp khó khăn trong việc phân bổ ngân sách sao cho hợp lý.

- Phân tích dữ liệu tài chính cá nhân là một công cụ mạnh mẽ giúp nhận diện được những xu hướng chi tiêu của các hộ gia đình. Việc thu thập, xử lý và phân tích dữ liệu về thu nhập, chi tiêu, các khoản tiết kiệm có thể cung cấp cái nhìn sâu sắc về cách thức mà các gia đình sử dụng nguồn lực tài chính của mình. Từ đó, các nhà nghiên cứu và các chuyên gia tài chính có thể đưa ra các lời khuyên hữu ích giúp các gia đình tối ưu hóa chi tiêu, cải thiện khả năng tiết kiệm và đạt được các mục tiêu tài chính bền vững.

- Đề tài nhằm mục đích nghiên cứu và phân tích các thói quen chi tiêu của các hộ gia đình, xác định các yếu tố ảnh hưởng đến hành vi tài chính, và tìm kiếm những mô hình chi tiêu có thể áp dụng để cải thiện tình hình tài chính. Dự án này không chỉ giúp người dân có cái nhìn rõ ràng hơn về tình hình tài chính cá nhân mà còn là cơ sở để các chính sách tài chính cá nhân, các chương trình tư vấn tài chính được thiết kế sao cho phù hợp với nhu cầu và đặc điểm của từng nhóm hộ gia đình.

- Phân tích xu hướng chi tiêu của hộ gia đình không chỉ quan trọng đối với các cá nhân mà còn đối với các tổ chức tài chính, ngân hàng, và các cơ quan nghiên cứu kinh tế, giúp họ phát triển các sản phẩm tài chính phù hợp, đồng thời có thể đưa ra các chiến lược giúp hỗ trợ người dân trong việc quản lý tài chính hiệu quả hơn.

- Với sự phát triển mạnh mẽ của công nghệ và các công cụ phân tích dữ liệu, việc thu thập và xử lý thông tin tài chính đã trở nên dễ dàng hơn bao giờ hết. Nhờ đó, chúng ta có thể tiến hành nghiên cứu sâu rộng và chính xác hơn về xu hướng chi tiêu của hộ gia đình, từ đó đóng góp vào sự phát triển bền vững của nền kinh tế cá nhân trong xã hội hiện đại.

## 1.2. Lý do chọn đề tài

- Mỗi gia đình đều phải đối mặt với các quyết định tài chính hàng ngày, từ việc chi tiêu cho các nhu cầu cơ bản như ăn uống, nhà ở đến các khoản chi tiêu cho giáo dục, giải trí, hoặc đầu tư dài hạn. Tuy nhiên, không phải ai cũng có thể quản lý tài chính hiệu quả, và nhiều hộ gia đình gặp phải tình trạng chi tiêu vượt quá thu nhập, dẫn đến những khó khăn tài chính nghiêm trọng.

- Một số lý do lựa chọn đề tài như sau:

+ Tầm quan trọng của việc quản lý tài chính gia đình: Quản lý tài chính là yếu tố quyết định đến sự ổn định và phát triển của mỗi gia đình. Việc chi tiêu hợp lý giúp gia đình duy trì cuộc sống ổn định và có thể tiết kiệm cho các mục tiêu dài hạn. Tuy nhiên, không phải tất cả các hộ gia đình đều biết cách phân bổ ngân sách hợp lý, dẫn đến việc vay mượn, nợ nần và gặp khó khăn tài chính.

+ Hiện trạng chi tiêu chưa hiệu quả của nhiều hộ gia đình: Nhiều gia đình không có kế hoạch tài chính rõ ràng, dẫn đến việc chi tiêu vào những mục không cần thiết hoặc không phù hợp với thu nhập.

+ Sự cần thiết của phân tích dữ liệu tài chính: Với sự phát triển mạnh mẽ của công nghệ và dữ liệu lớn (Big Data), việc thu thập và phân tích dữ liệu tài chính cá nhân không còn là vấn đề khó khăn. Các công cụ phân tích dữ liệu hiện nay giúp nhận diện xu hướng chi tiêu của các hộ gia đình, từ đó đưa ra các khuyến nghị về cách quản lý tài chính hợp lý và hiệu quả hơn.

+ Đáp ứng nhu cầu tăng trưởng của nền kinh tế cá nhân: Trong bối cảnh nền kinh tế ngày càng phát triển, mức độ chi tiêu và tiết kiệm của hộ gia đình đóng vai trò quan trọng trong sự ổn định của nền kinh tế nói chung. Các gia đình có xu hướng tiêu dùng ngày càng lớn hơn, nhưng đồng thời cũng cần phải tiết kiệm và đầu tư thông minh.

+ Tính thực tiễn và gần gũi: Tất cả mọi người đều quan tâm đến tài chính cá nhân: Quản lý tài chính cá nhân là một yếu tố quan trọng trong cuộc sống hàng ngày của mỗi người.



### **1.3. Mục tiêu**

- Với đề tài này chúng ta có thể phân tích được nhiều dữ liệu từ chính dữ liệu đó ta có thể phân tích ra nhiều khía cạnh trong vấn đề chi tiêu và thu nhập của cá nhân cho tới các hộ gia đình.

+ Phân tích thói quen chi tiêu của hộ gia đình: Xác định và phân loại các khoản chi tiêu của hộ gia đình vào các nhóm chi tiêu cụ thể như nhà ở, ăn uống, giải trí, chăm sóc sức khỏe, giáo dục của con cái.

+ Xác định các yếu tố ảnh hưởng đến chi tiêu: Phân tích các yếu tố như thu nhập, số người trong gia đình, độ tuổi của người đứng đầu hộ gia đình, vị trí địa lý (thành thị/nông thôn), và các yếu tố khác có ảnh hưởng đến xu hướng chi tiêu. Đánh giá mức độ ảnh hưởng của từng yếu tố lên các nhóm chi tiêu trong hộ gia đình.

+ Nâng cao nhận thức về tài chính cá nhân: Tăng cường nhận thức của các hộ gia đình về tầm quan trọng của việc theo dõi và quản lý chi tiêu, giúp họ trở nên chủ động hơn trong việc quản lý tài chính cá nhân.

+ Phát hiện các bất hợp lý trong quản lý tài chính: Phân tích dữ liệu giúp phát hiện những bất hợp lý trong việc chi tiêu, chẳng hạn như những cá nhân có thu nhập cao nhưng chi tiêu không hợp lý hoặc những người có thu nhập thấp nhưng lại có khoản chi tiêu quá lớn so với khả năng tài chính của họ.

- Việc phân tích dữ liệu tài chính cá nhân nhằm hiểu rõ về thu nhập và chi tiêu, xác định các yếu tố ảnh hưởng đến tình hình tài chính, và cung cấp các khuyến nghị để tối ưu hóa quản lý tài chính. Mục tiêu chính là giúp các cá nhân và hộ gia đình có thể điều chỉnh hành vi chi tiêu sao cho hợp lý, từ đó cải thiện sức khỏe tài chính cá nhân của họ.

### **1.4. Phạm vi nghiên cứu của đề tài**

- Phạm vi của việc phân tích dữ liệu và nghiên cứu của đề tài có đề cập nhiều tới các phạm vi khác nhau, do đó phạm vi nghiên cứu của đề tài này ở các phạm vi sau:

+ Phạm vi không gian: Nghiên cứu sẽ không giới hạn trong một khu vực cụ thể mà có thể áp dụng cho nhiều vùng, thành phố khác nhau. Tuy nhiên, trong bộ dữ liệu sẽ thu thập dữ liệu từ các cá nhân đều có thu nhập ổn định từ các ngành nghề phổ biến ở đô thị.

+ Phạm vi các yếu tố nghiên cứu: Thu nhập của cá nhân, các khoản chi tiêu của cá nhân đó trong cuộc sống, sự phân bổ chi tiêu và tình trạng tài chính cá nhân.

+ Phạm vi các phương pháp nghiên cứu:

Phân tích tương quan: Nghiên cứu mối quan hệ giữa thu nhập và các khoản chi tiêu, để xác định các yếu tố ảnh hưởng đến việc phân bổ chi tiêu của các cá nhân.

+ Phạm vi thời gian: Phạm vi nghiên cứu chỉ bao gồm các dữ liệu tài chính hiện tại (hoặc gần đây) của các cá nhân, không phân tích về sự thay đổi trong thói quen chi tiêu theo thời gian mà chỉ dừng lại ở mức dữ liệu hiện tại.

## CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

### 2.1. Trình bày về các khái niệm thống kê

- Ngày nay phương pháp thống kê được áp dụng trong tất cả các lĩnh vực có liên quan đến việc ra quyết định, để cho các kết luận chính xác từ một bộ phận so với các dữ liệu và đưa ra quyết định khi đối mặt với kết luận không chắc chắn dựa trên phương pháp thống kê, có một số khái niệm cơ bản rất quan trọng, giúp định hướng và hiểu rõ về các phương pháp phân tích dữ liệu. Sau đây là hai khái niệm cơ bản cần phải hiểu rõ: Tổng thể và mẫu nghiên cứu và các đặc trưng của mẫu và tổng thể.

#### 2.1.1. Tổng thể và mẫu nghiên cứu

- Tổng thể nghiên cứu (Population) và Mẫu nghiên cứu (Sample) là hai khái niệm quan trọng trong bất kỳ nghiên cứu thống kê nào. Chúng liên quan đến phạm vi dữ liệu được thu thập và phân tích.

- Tổng thể nghiên cứu (Population):

+ Khái niệm: Tổng thể nghiên cứu là tập hợp tất cả các đơn vị (cá nhân, sự kiện, đối tượng) mà ta quan tâm nghiên cứu và muốn tìm hiểu về chúng. Tổng thể có thể bao gồm tất cả các phần tử của một nhóm hoặc một hiện tượng cụ thể.

+ Ví dụ: Nếu bạn đang nghiên cứu về xu hướng chi tiêu của các hộ gia đình tại Việt Nam, tổng thể nghiên cứu là tất cả các hộ gia đình ở Việt Nam.

+ Đặc điểm: Tổng thể có thể rất lớn, thậm chí vô hạn, và vì thế đôi khi không thể thu thập dữ liệu từ tất cả các phần tử trong tổng thể.

- Mối quan hệ giữa tổng thể và mẫu:

+ Dữ liệu thu thập từ mẫu sẽ được sử dụng để ước tính các đặc trưng của tổng thể (chẳng hạn như trung bình, phương sai).

+ Một mẫu đại diện sẽ giúp ta suy luận chính xác về tổng thể mà không cần phải thu thập dữ liệu từ tất cả các phần tử trong tổng thể.

### 2.1.2. Các đặc trưng của mẫu và tổng thể

\* Đặc trưng của mẫu:

- Các đặc trưng của mẫu được gọi là thống kê mẫu (sample statistics), và chúng được sử dụng để ước tính các tham số của tổng thể.
- Những đặc trưng này sẽ giúp ta đưa ra các kết luận về tổng thể dựa trên dữ liệu từ mẫu.
- Các đặc trưng của mẫu thường được ký hiệu bằng chữ latin. Ví dụ:
  - + Trung bình mẫu (Sample mean,  $\bar{x}$ ): Giá trị trung bình của các phần tử trong mẫu.
  - + Độ lệch chuẩn mẫu (Sample standard deviation,  $s$ ): Đo lường mức độ phân tán của các giá trị trong mẫu.
  - + Phương sai mẫu (Sample variance,  $s^2$ ): Đo lường sự biến thiên giữa các giá trị trong mẫu.
  - + Tỷ lệ mẫu (Sample proportion,  $p$ ): Tỷ lệ của một đặc điểm trong mẫu.

\* Đặc trưng của tổng thể:

- Thông kê mô tả tổng thể là các chỉ số đo lường cho tổng thể mà không cần phải thu thập thông tin từ từng cá nhân trong tổng thể. Các đặc trưng này giúp mô tả tổng thể về mặt số liệu.
- Các đặc trưng của tổng thể thường được gọi là tham số (parameter). Ví dụ:
  - + Trung bình tổng thể (Population mean,  $\mu$ ): Giá trị trung bình của tất cả các phần tử trong tổng thể.
  - + Độ lệch chuẩn tổng thể (Population standard deviation,  $\sigma$ ): Đo lường mức độ phân tán của các giá trị trong tổng thể.
  - + Phương sai tổng thể (Population variance,  $\sigma^2$ ): Đo lường sự biến thiên giữa các giá trị trong tổng thể.
  - + Tỷ lệ tổng thể (Population proportion,  $p$ ): Tỷ lệ một đặc điểm trong tổng thể.

*\* Sự khác biệt giữa tham số và thống kê:*

- Tham số là các đặc trưng của tổng thể và là giá trị cố định. Tuy nhiên, trong thực tế, chúng ta không thể thu thập dữ liệu từ toàn bộ tổng thể, vì vậy các tham số này thường không thể tính toán trực tiếp.
- Thống kê mẫu là các đặc trưng của mẫu và có thể tính toán được từ dữ liệu mẫu. Thống kê mẫu được sử dụng để ước tính các tham số tổng thể.

*\* Mối quan hệ giữa thống kê mẫu và tham số tổng thể:*

- Độ lệch: Thống kê mẫu có thể khác tham số tổng thể do ảnh hưởng của sự thay đổi ngẫu nhiên trong việc chọn mẫu. Tuy nhiên, với một mẫu đủ lớn và phương pháp chọn mẫu chính xác, thống kê mẫu sẽ gần giống với tham số tổng thể.

## **2.2 Các phương pháp phân tích dữ liệu tài chính cá nhân**

- Để phân tích xu hướng chi tiêu của hộ gia đình, các phương pháp và kỹ thuật phân tích dữ liệu có thể được áp dụng được .

### **2.2.1. Thống kê mô tả (Descriptive Statistics)**

- Các chỉ số thống kê mô tả giúp tóm tắt và trình bày các đặc điểm cơ bản của dữ liệu tài chính. Các chỉ số quan trọng như:
  - + Trung bình (Mean): Giúp hiểu mức độ chi tiêu trung bình của hộ gia đình trong một khoảng thời gian.
  - + Độ lệch chuẩn (Standard Deviation): Giúp đo lường sự biến động hoặc sự phân tán trong chi tiêu của hộ gia đình.
  - + Mốt (Mode): Giúp xác định mức chi tiêu phổ biến nhất trong nhóm hộ gia đình nghiên cứu.
  - + Trung vị (Median): Được sử dụng khi có các giá trị ngoại lệ trong dữ liệu chi tiêu, giúp xác định mức chi tiêu trung bình không bị ảnh hưởng bởi các giá trị cực trị.

- Ví dụ: Giả sử bạn có dữ liệu chi tiêu hàng tháng của 10 hộ gia đình trong một khu vực.

+ Dữ liệu chi tiêu hàng tháng (đơn vị: triệu đồng):

5, 7, 8, 10, 12, 6, 8, 11, 9, 7

+ Bạn có thể tính các chỉ số thống kê mô tả như sau:

Trung bình (Mean):

Trung bình chi tiêu hàng tháng của các hộ gia đình là:

$$\text{Trung bình} = (5+7+8+10+12+6+8+11+9+7) / 10 = 8.3 \text{ triệu đồng}$$

### ***2.2.2. Phân tích hồi quy (Regression Analysis)***

- Là phương pháp thống kê giúp xác định mối quan hệ giữa các yếu tố ảnh hưởng đến chi tiêu của hộ gia đình. Ví dụ, hồi quy có thể giúp xác định mối quan hệ giữa thu nhập và chi tiêu của hộ gia đình, hoặc giữa mức độ tiết kiệm và các yếu tố khác như nợ, thu nhập, và chi tiêu.

+ Hồi quy tuyến tính (Linear regression) có thể được sử dụng để phân tích mối quan hệ giữa chi tiêu và các yếu tố như thu nhập, số người trong gia đình, độ tuổi của người trụ cột, v.v.

+ Hồi quy đa biến (Multiple regression) có thể được sử dụng khi có nhiều yếu tố tác động đồng thời đến chi tiêu của hộ gia đình.

### ***2.2.3. Phân tích tương quan (Correlation Analysis)***

- Phân tích tương quan giúp xác định mức độ liên quan giữa các yếu tố trong dữ liệu tài chính, ví dụ, mối quan hệ giữa chi tiêu và thu nhập, hoặc giữa mức chi tiêu vào các khoản cố định (như nhà cửa, giáo dục) và các khoản chi tiêu biến đổi (như giải trí, du lịch). Hệ số tương quan giúp chúng ta hiểu rõ hơn về các mối liên hệ này.

#### **2.2.4. Phân tích phân phối (Distribution Analysis)**

- Phân tích phân phối giúp hiểu về cách thức phân bố của các giá trị trong dữ liệu chi tiêu. Ví dụ, có thể xác định được tỷ lệ phần trăm hộ gia đình chi tiêu dưới một mức nhất định, hoặc phân tích các mô hình chi tiêu của các nhóm thu nhập khác nhau.

#### **2.3. Các yếu tố ảnh hưởng đến xu hướng chi tiêu của hộ gia đình**

- Khi phân tích xu hướng chi tiêu của hộ gia đình, có một số yếu tố quan trọng ảnh hưởng đến mức chi tiêu của hộ gia đình cần được chú ý:

+ Thu nhập: Thu nhập cao hơn thường dẫn đến chi tiêu cao hơn, tuy nhiên, cách thức chi tiêu của các hộ gia đình có thu nhập khác nhau cũng có thể thay đổi. Hộ gia đình có thu nhập cao có thể dành nhiều cho đầu tư, trong khi các hộ gia đình có thu nhập thấp có thể tập trung nhiều hơn vào các chi phí cơ bản. Ngoài ra trong một hộ gia đình các cá nhân có mức thu nhập khác nhau, do đó thu nhập cao hay thấp sẽ còn dựa vào thu nhập bình quân đầu người trong gia đình.

+ Phí sinh hoạt và cuộc sống: Gia đình nào cũng thế, việc chi tiêu cho các phí sinh hoạt thường ngày như ăn uống, điện, nước và các khoản chi tiêu khác liên quan đến sinh hoạt trong cuộc sống của từng hộ gia đình. Ngoài ra, sẽ còn tồn tại các chi phí phát sinh khác cũng sẽ ảnh hưởng đến mức độ chi tiêu và thu nhập của mỗi cá nhân trong gia đình điều đó sẽ ảnh hưởng đến mức độ chi tiêu của gia đình. Với gia đình có con cái đang ở giai đoạn đi học thì mức độ chi tiêu cho khoản này cũng sẽ ảnh hưởng nhiều tới mức độ chi tiêu của hộ gia đình đó.

+ Đặc điểm nhân khẩu học: Độ tuổi, tình trạng hôn nhân, số lượng con cái, và các yếu tố khác có thể ảnh hưởng lớn đến xu hướng chi tiêu cũng như thu nhập của mỗi cá nhân từ đó ảnh hưởng lớn đến mức độ chi tiêu của hộ gia đình. Ví dụ, hộ gia đình có trẻ em có thể có mức chi tiêu cho giáo dục và chăm sóc sức khỏe cao hơn.

+ Môi trường kinh tế: Tình hình kinh tế nói chung, như tăng trưởng GDP, tỷ lệ thất nghiệp, và sự ổn định của thị trường lao động, cũng có thể ảnh hưởng đến chi tiêu của hộ gia đình.

Khi nền kinh tế phát triển mạnh, hộ gia đình có thể chi tiêu nhiều hơn, ngược lại khi kinh tế suy thoái, chi tiêu có xu hướng giảm.

## 2.4. Hồi quy tuyến tính

- Hồi quy tuyến tính trong **xác suất thống kê** là một phương pháp được sử dụng để mô hình hóa mối quan hệ giữa một biến phụ thuộc và một hoặc nhiều biến độc lập trong ngữ cảnh của xác suất và thống kê. Nó giúp phân tích cách thức mà các yếu tố độc lập tác động lên biến phụ thuộc và dự đoán giá trị của biến này dựa trên các yếu tố đã biết.

- Với bài báo cáo này có thể sử dụng hồi quy tuyến tính để phân tích mối quan hệ giữa thu nhập và chi tiêu của hộ gia đình hoặc giữa các yếu tố khác nhau ảnh hưởng đến chi tiêu.

### 2.4.1. Mô hình hồi quy tuyến tính đơn giản

- Trong mô hình hồi quy tuyến tính đơn giản, mối quan hệ giữa một biến phụ thuộc Y (ví dụ: chi tiêu của hộ gia đình) và một biến độc lập X (ví dụ: thu nhập của hộ gia đình) được mô tả như sau:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

+ Y: Biến phụ thuộc (chi tiêu của hộ gia đình)

+ X: Biến độc lập (thu nhập của hộ gia đình)

+  $\beta_0$  : Hệ số chặn (intercept), đại diện cho giá trị của YYY khi  $X=0$ .

+  $\beta_1$  : Hệ số góc (slope), đại diện cho sự thay đổi trong YYY khi XXX thay đổi một đơn vị.

+  $\epsilon$  Sai số (error), phản ánh sự biến động không thể giải thích được của YYY ngoài các yếu tố trong mô hình.

- Ý nghĩa:

+ Nếu  $\beta_1$  lớn và dương, điều đó có nghĩa là thu nhập có ảnh hưởng tích cực đến chi tiêu: khi thu nhập tăng lên, chi tiêu cũng có xu hướng tăng.



+ Nếu  $\beta_1 = 0$ , điều đó có nghĩa là thu nhập không ảnh hưởng đến chi tiêu.

#### **2.4.2. Hồi quy tuyến tính bội (đa biến)**

- Trong trường hợp phân tích nhiều yếu tố ảnh hưởng đến chi tiêu hơn, có thể sử dụng hồi quy tuyến tính bội (đa biến), trong đó có nhiều biến độc lập. Mô hình hồi quy tuyến tính bội sẽ có dạng:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

+  $X_1, X_2, \dots, X_n$ : Các biến độc lập (ví dụ: thu nhập, độ tuổi, tình trạng hôn nhân, trình độ học vấn, số thành viên trong gia đình, v.v.)

+  $\beta_1, \beta_2, \dots, \beta_n$ : Các hệ số tương ứng với các biến độc lập.

- Ý nghĩa:

+ Mỗi hệ số  $\beta_i$  trong mô hình đại diện cho tác động của biến  $X_i$  đến biến phụ thuộc  $Y$ , trong khi giữ các biến khác cố định. Ví dụ, hệ số  $\beta_1$  sẽ cho bạn biết thay đổi trong thu nhập ( $X_1$ ) sẽ ảnh hưởng đến chi tiêu ( $Y$ ), khi các yếu tố khác (như tuổi, tình trạng hôn nhân) không thay đổi.

#### **2.4.3. Ứng dụng trong đề tài này**

- Phân tích chi tiêu của hộ gia đình: Bạn có thể áp dụng hồi quy tuyến tính để tìm hiểu mối quan hệ giữa thu nhập và các yếu tố khác như mức độ chi tiêu cho các nhu cầu thiết yếu, giải trí, tiết kiệm, đầu tư. Ví dụ, bạn có thể xây dựng mô hình để dự đoán chi tiêu cho thực phẩm, nhà ở, giáo dục dựa trên các yếu tố như thu nhập, tuổi, số lượng người trong gia đình, v.v.

- Dự đoán chi tiêu: Khi có dữ liệu về thu nhập và các yếu tố khác, mô hình hồi quy tuyến tính có thể giúp bạn dự đoán mức chi tiêu của một hộ gia đình cụ thể.

## **2.5. Thu thập dữ liệu**

- Có nhiều nguồn thu thập dữ liệu về từng cá nhân cũng như từng gia đình, tùy vào quy mô dự án. Với dự án này quy mô sẽ nhỏ, do vậy các dữ liệu và thống kê sẽ được thu thập trong một phạm vi nhất định và phù hợp với quy mô của dự án. Dữ liệu được thu thập thông qua quá trình gom nhặt dữ liệu từ các cá nhân trong các gia đình khác nhau từ đó tạo nên một bộ dữ liệu phù hợp với đề tài cũng như quy mô của báo cáo. Các chi phí tiêu dùng của từng cá nhân gia đình cũng như các khoản thu nhập sẽ được thu thập để tạo nên một bộ dữ liệu để đáp ứng nhu cầu của dự án.

## CHƯƠNG 3. PHÂN TÍCH DỮ LIỆU

### 3.1. Giới thiệu về dữ liệu

- Trong nghiên cứu này, chúng tôi sử dụng dữ liệu tài chính cá nhân của các hộ gia đình để tìm hiểu xu hướng chi tiêu. Dữ liệu được thu thập từ các nguồn tài chính cá nhân, bao gồm các báo cáo chi tiêu hàng tháng, dữ liệu thu nhập, khoản tiết kiệm và các khoản chi cho các loại dịch vụ và sản phẩm khác nhau. Mục tiêu là phân tích các yếu tố ảnh hưởng đến chi tiêu của hộ gia đình và tìm ra các xu hướng chi tiêu phổ biến.

- Dữ liệu sơ cấp: Đây là dữ liệu thu thập trực tiếp từ các hộ gia đình thông qua khảo sát, phỏng vấn hoặc bảng câu hỏi trực tuyến, các thông tin trong dữ liệu bao gồm:

+ Thu nhập hàng tháng của hộ gia đình

+ Chi tiêu hàng tháng cho các danh mục như: thực phẩm, điện nước, giải trí, giáo dục, y tế, ...

+ Số tiền tiết kiệm và đầu tư

+ Các khoản vay và nợ tín dụng

- Dữ liệu thứ cấp: Sử dụng dữ liệu từ các báo cáo tài chính cá nhân, các nghiên cứu của tổ chức tài chính và cơ quan chính phủ để bổ sung thêm thông tin hữu ích.

- Công cụ thu thập dữ liệu: Google Forms để thu thập dữ liệu trực tuyến, Excel để xử lý dữ liệu sơ bộ và SQL Database để lưu trữ dữ liệu có cấu trúc.

### 3.2. Thu thập dữ liệu

- Phương pháp thu thập dữ liệu:

+ Dữ liệu được thu thập thông qua các khảo sát từ các hộ gia đình trong khu vực (hoặc từ các nguồn cơ sở dữ liệu tài chính công cộng, nếu có). Các thông tin này được thu thập từ các phương pháp như:

- + Khảo sát trực tuyến hoặc qua điện thoại: Hỏi về chi tiêu hàng tháng, thu nhập, khoản vay và các khoản chi khác.

- + Báo cáo tài chính tự động: Dữ liệu được lấy từ các ứng dụng quản lý tài chính cá nhân hoặc ngân hàng điện tử.

- Thời gian và không gian thu thập dữ liệu:

- + Dữ liệu được thu thập trong khoảng thời gian từ (thời gian bắt đầu) đến (thời gian kết thúc). Dữ liệu đại diện cho các hộ gia đình ở (khu vực cụ thể, ví dụ: thành phố, nông thôn, hoặc toàn quốc).

- Ngoài ra dữ liệu sẽ được tham khảo một phần từ dữ liệu trên internet từ các nguồn uy tín như Github và Kaggle.

### **3.3 Chuẩn bị và làm sạch dữ liệu**

- Xử lý dữ liệu thiếu: Trong quá trình thu thập, một số mẫu có thể thiếu thông tin về một số khoản mục chi tiêu hoặc thu nhập. Để xử lý, sử dụng các phương pháp như:

- + Điền giá trị trung bình cho các khoản chi tiêu bị thiếu.

- + Loại bỏ các mẫu bị thiếu dữ liệu quan trọng (nếu tỷ lệ thiếu quá lớn).

Mã Cá Nhân	Mã Hộ Gia Đình	Thu nhập cá nhân	Nhu cầu thiết yếu	Cuộc sống hiện đại	Giải trí & Trải nghiệm	Dầu tư & Tiết kiệm	Tổng Chi Tiêu Cá Nhân
	GHD_1			17337	7845	22756	57335
C_N2	HGD1_	27539	249600	3317	7146	11130	30892
CN3_		25227	10121	199890	3396	14768	727840
NC_4	HG_D1	48775	13821	6106	1228	1115	
NC_5	HGD1_	44333	22001	18058	5219	8286	53564
CN6_	GHD_2	5078	12225	6799		21817	46678
C_N7		55410	9283	199890	13776	10302	35117
CN8_	HG_D3	46072	23535	19857	4920	4451	52763
CN9_	HG_D3	37276	14387		6742	2972	35420
CN_01	HDG_3	3031	10724	7685	6811	4321	29521
NC_11		52154	21019	12902	4378	2598	40897
C_N12	HGD4_	54756	12864		9525	19889	52484
CN_31	HG_D4	38212	22987	5416	5944	16111	50458
NC_14	HG_D5		5878	13868	8090	19688	47524
C_N15	HG_D6		23448	18811	11576	11331	65166
NC_16	HDG_6		21420	14001	2605	24115	62141
NC_17	HDG_6	28414	19211	8365	9865	21491	58932
C_N18	HG_D7	18263	17179	12249	13090		58379
CN_91	HGD7_	20988	7609	199890	6046	1307	17186
CN_02	HDG_7	599070	4277	3119		20518	34319
CN_12	GHD_8	23382	15308	10442	2796	9079	37625
C_N22	HDG_8	50192	14201	7136		24381	58891
NC_23	HGD9_	19169			8174	960	46897
CN2_4	HG_D9	4545	3270	1278	10829	15882	31259
CN2_5	HGD_01	43220	17886	13231	12145	6971	50233
	HG_D10	39825	7904	199890	5868	19291	37206
NC_27	HDG_11	26525	14664	4818	364	14845	34691
CN2_8	HDG_11	8779			4232	19168	
CN2_9		9845	11948	10809	12760	12205	47722
CN_03	HG_D12	53229	9498	17688	14024	15666	56876
CN_13	HGD1_2	28355	10869		12777	20935	52233
CN3_2	HDG_12	31383	4726		3485	24618	49803

*Hình 3.1: Hình ảnh dữ liệu trước khi lọc*

- Xử lý dữ liệu ngoại lai (outliers): Các dữ liệu ngoại lai có thể xuất hiện, đặc biệt trong các khoản chi tiêu cực đoan hoặc thu nhập rất cao hoặc thấp. Những điểm dữ liệu này có thể gây sai lệch trong mô hình phân tích. Để xử lý:

+ Sử dụng biểu đồ hộp (boxplot) để xác định các điểm ngoại lai.

+ Quyết định loại bỏ hoặc chỉnh sửa các điểm dữ liệu ngoại lai tùy theo mức độ ảnh hưởng của chúng.

- Chuyển đổi dữ liệu: Một số biến trong dữ liệu có thể cần phải chuẩn hóa hoặc mã hóa lại (nếu là dữ liệu phân loại). Ví dụ:

+ Chuyển đổi các khoản chi tiêu thành phần trăm so với thu nhập hàng tháng.

+ Mã hóa các loại chi tiêu (thực phẩm, giải trí, giáo dục, v.v...) thành các giá trị số để dễ dàng phân tích.

A	B	C	D	E	F	G	H
Mã Cá Nhân	Mã Hộ Gia Đình	Thu nhập cá nhân	Nhu cầu thiết yếu	Cuộc sống hiện đại	Giải trí & Trải nghiệm	Đầu tư & Tiết kiệm	Tổng Chi Tiêu Cá Nhân
CN_1	HGD_1	49810	9397	17337	7845	22756	57335
CN_2	HGD_1	27539	9299	3317	7146	11130	30892
CN_3	HGD_1	25227	10121	15975	3396	14768	44260
CN_4	HGD_1	48775	13821	6106	1228	1115	22270
CN_5	HGD_1	44333	22001	18058	5219	8286	53564
CN_6	HGD_2	5078	12225	6799	5837	21817	46678
CN_7	HGD_2	55410	9283	1756	13776	10302	35117
CN_8	HGD_3	46072	23535	19857	4920	4451	52763
CN_9	HGD_3	37276	14387	11319	6742	2972	35420
CN_10	HGD_3	3031	10724	7665	6811	4321	29521
CN_11	HGD_3	52154	21019	12902	4378	2598	40897
CN_12	HGD_4	54756	12864	10206	9525	19889	52484
CN_13	HGD_4	38212	22987	5416	5944	16111	50458
CN_14	HGD_5	26819	5878	13868	8090	19688	47524
CN_15	HGD_6	44150	23448	18811	11576	11331	65166
CN_16	HGD_6	11762	21420	14001	2605	24115	62141
CN_17	HGD_6	28414	19211	8365	9865	21491	58932
CN_18	HGD_7	18263	17179	12249	13090	15861	58379
CN_19	HGD_7	20988	7609	2224	6046	1307	17186
CN_20	HGD_7	40338	4277	3119	6405	20518	34319
CN_21	HGD_8	23382	15308	10442	2796	9079	37625
CN_22	HGD_8	50192	14201	7136	13173	24381	58891
CN_23	HGD_9	19169	24221	13542	8174	960	46897
CN_24	HGD_9	4545	3270	1278	10829	15882	31259
CN_25	HGD_10	43220	17886	13231	12145	6971	50233
CN_26	HGD_10	39825	7904	4143	5868	19291	37206
CN_27	HGD_11	26525	14664	4818	364	14845	34691
CN_28	HGD_11	8779	7563	9344	4232	19168	40307
CN_29	HGD_11	9845	11948	10809	12760	12205	47722
CN_30	HGD_12	53229	9498	17688	14024	15666	56876
CN_31	HGD_12	28355	10869	7652	12777	20935	52233
CN_32	HGD_12	31383	4726	16974	3485	24618	49803
CN_33	HGD_12	8192	24896	3724	11033	10177	49830
CN_34	HGD_13	27709	14478	4333	3110	16873	38794
CN_35	HGD_13	45765	15067	6550	11201	1831	34649
CN_36	HGD_13	43814	4464	3145	2286	23725	33620
CN_37	HGD_14	5575	19360	4628	11003	7614	42605

Hình 3.2: Hình ảnh về dữ liệu đã làm sạch

### 3.4 Phân tích mô tả dữ liệu

- Thống kê mô tả: Chúng tôi tính toán các thống kê mô tả cho các biến tài chính chính, bao gồm:

+ Trung bình và độ lệch chuẩn của thu nhập và chi tiêu trong các danh mục khác nhau.

+ Phương sai và phạm vi cho các khoản chi tiêu để xác định mức độ biến động.

- Biểu đồ và đồ thị: Để hình dung rõ hơn về các xu hướng chi tiêu, chúng tôi sử dụng các công cụ trực quan hóa như:

+ Biểu đồ phân tán (scatter plot): Giúp phân tích mối quan hệ giữa thu nhập và chi tiêu.

+ Histogram: Xem xét phân phối chi tiêu trong các danh mục khác nhau.

+ Biểu đồ hộp (boxplot): Nhận diện các điểm ngoại lai và phân phối chi tiêu.

- Phân tích theo nhóm (segmentation analysis): Phân tích dữ liệu chỉ tiêu của các hộ gia đình theo các nhóm như:

+ Nhóm theo thu nhập: So sánh chỉ tiêu giữa các hộ gia đình có thu nhập cao, trung bình và thấp.

+ Nhóm theo độ tuổi hoặc quy mô hộ gia đình: Tìm hiểu xem chỉ tiêu có thay đổi theo độ tuổi hoặc kích thước của hộ gia đình.

### 3.5 Kiểm tra giả thuyết

- Giới thiệu các giả thuyết nghiên cứu: Với bài báo cáo này, chúng tôi kiểm tra một số giả thuyết liên quan đến xu hướng chỉ tiêu của hộ gia đình, bao gồm:

+ Giả thuyết 1: Mối quan hệ giữa thu nhập và tỷ lệ chi tiêu cho các danh mục như thực phẩm, giải trí.

+ Giả thuyết 2: Mối quan hệ giữa khoản vay và mức độ tiết kiệm của hộ gia đình.

- Kiểm định thống kê: Để kiểm tra các giả thuyết này, chúng tôi sử dụng các phương pháp thống kê như:

+ Kiểm định t (t-test) để so sánh sự khác biệt trong chỉ tiêu giữa các nhóm (thu nhập cao và thu nhập thấp).

+ Phân tích hồi quy tuyến tính để kiểm tra mối quan hệ giữa thu nhập và các khoản chi tiêu, từ đó tìm ra các yếu tố chính ảnh hưởng đến chỉ tiêu của hộ gia đình.

### 3.6 Mô hình hóa dữ liệu

- Phương pháp mô hình hóa:

+ Để hiểu rõ hơn về các yếu tố tác động đến chỉ tiêu, chúng tôi sử dụng mô hình **hồi quy tuyến tính**. Mô hình này giúp đánh giá mối quan hệ giữa các biến độc lập (như thu nhập, tuổi tác, khoản vay) và biến phụ thuộc (chỉ tiêu của hộ gia đình).

+ Mô hình hồi quy tuyến tính có dạng:

$$\text{Chi tiêu} = \beta_0 + \beta_1 \times \text{Thu nhập} + \beta_2 \times \text{Khoản vay} + \beta_3 \times \text{Độ tuổi} + \epsilon$$

+ Trong đó:

Y: là tổng chi tiêu của hộ gia đình.

Các X là các biến giải thích như thu nhập, khoản vay, độ tuổi, v.v...

$\epsilon$  là sai số.

- Kết quả mô hình hóa:

+ Sau khi chạy mô hình hồi quy, chúng tôi thu được các hệ số hồi quy cho từng biến độc lập và các giá trị  $R^2$ , cho biết mức độ phù hợp của mô hình với dữ liệu. Chúng tôi cũng tính toán các giá trị p để kiểm tra sự ý nghĩa thống kê của các yếu tố.

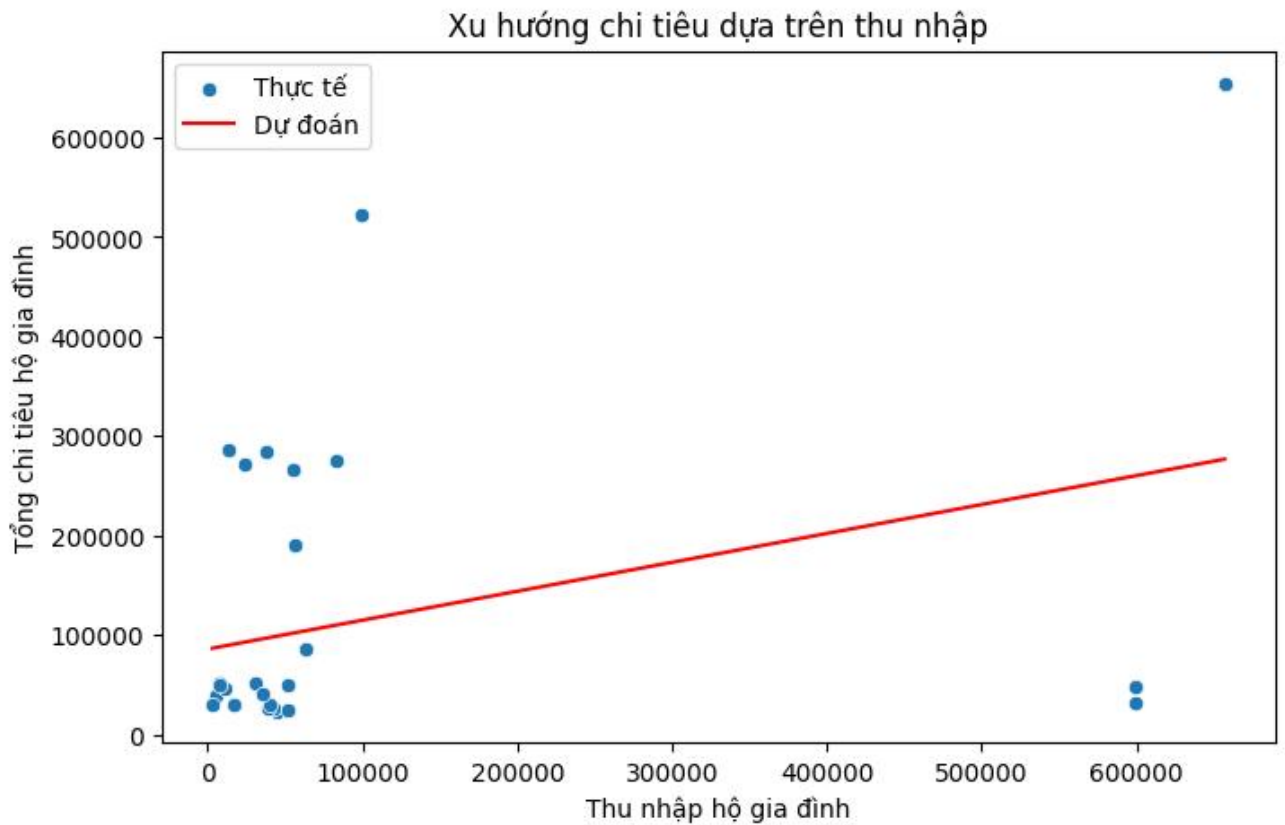
```
# 5. Hồi quy tuyến tính dự đoán tổng chi tiêu dựa trên thu nhập
X = df_household[["Thu_nhap_HGD"]]
y = df_household.iloc[:, 2:-1].sum(axis=1) # Tổng chi tiêu hộ gia đình
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# 6. Đánh giá mô hình
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"MAE: {mae:.2f}, R2 Score: {r2:.2f}")
```

Hình 3.3: Hình ảnh về hồi quy tuyến tính và đánh giá mô hình trong ngôn ngữ lập trình

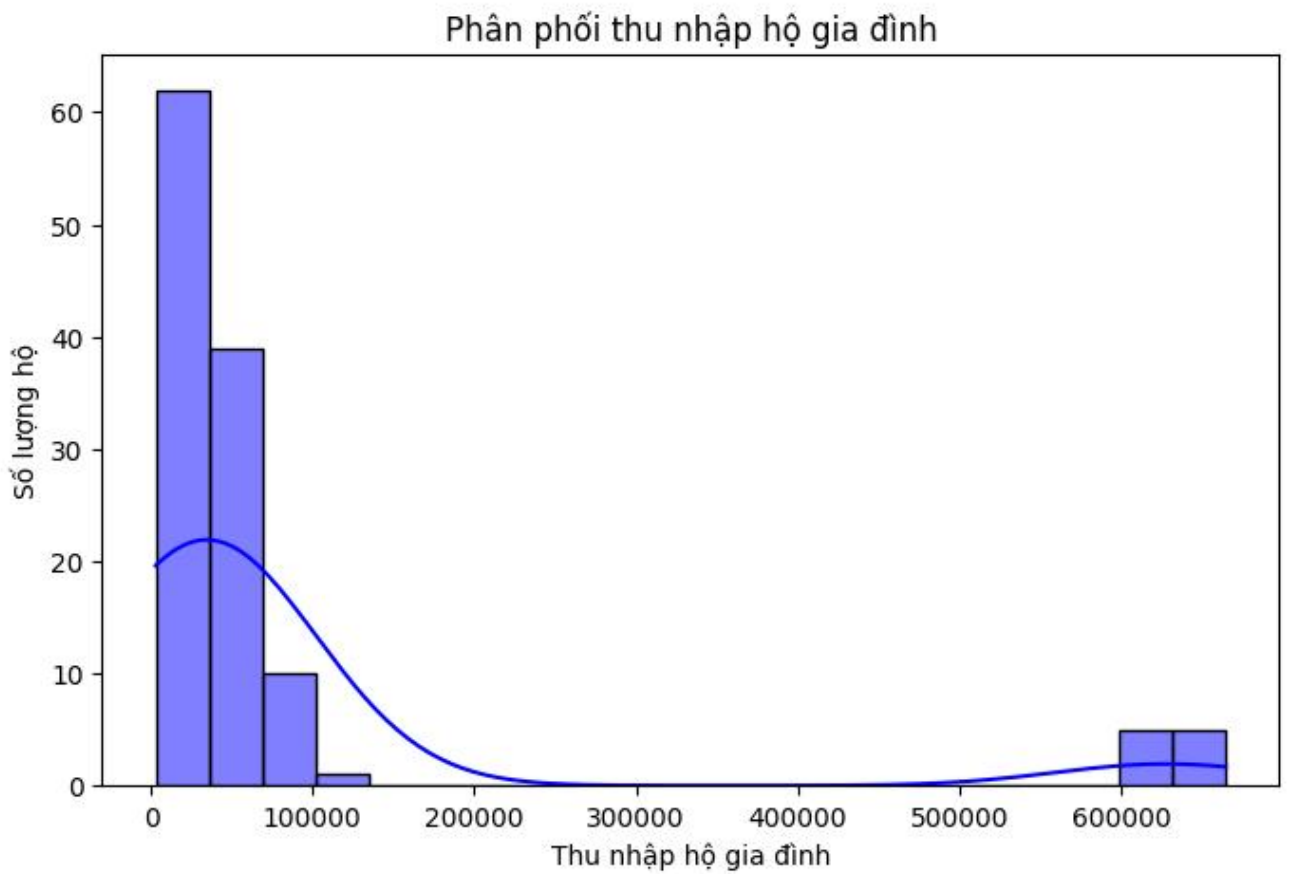


```
# 7. Biểu đồ 1: Xu hướng chi tiêu dựa trên thu nhập
plt.figure(figsize=(8, 5))
sns.scatterplot(x=X_test["Thu_nhap_HGD"], y=y_test, label="Thực tế")
sns.lineplot(x=X_test["Thu_nhap_HGD"], y=y_pred, color='red', label="Dự đoán")
plt.xlabel("Thu nhập hộ gia đình")
plt.ylabel("Tổng chi tiêu hộ gia đình")
plt.title("Xu hướng chi tiêu dựa trên thu nhập")
plt.legend()
plt.show()
```



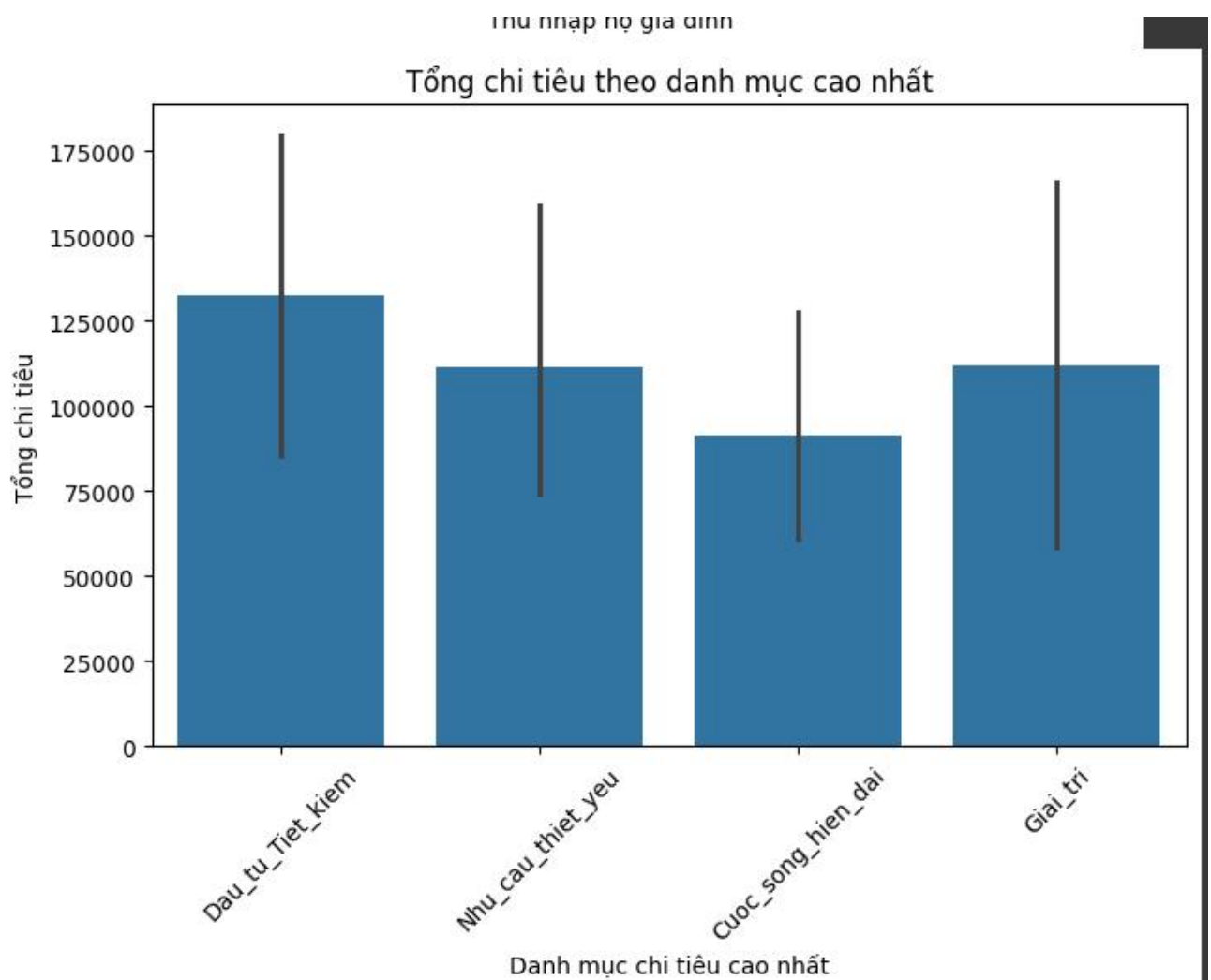
Hình 3.4: Hình ảnh kết quả về quy hồi tuyến tính

```
# 8. Biểu đồ 2: Phân phối thu nhập hộ gia đình
plt.figure(figsize=(8, 5))
sns.histplot(df_household["Thu_nhap_HGD"], bins=20, kde=True, color="blue")
plt.xlabel("Thu nhập hộ gia đình")
plt.ylabel("Số lượng hộ")
plt.title("Phân phối thu nhập hộ gia đình")
plt.show()
```



*Hình 3.5: Hình ảnh kết quả về phân phối thu nhập*

```
# 9. Biểu đồ 3: Tổng chi tiêu theo mức thu nhập
plt.figure(figsize=(8, 5))
sns.barplot(x=df_household["Khoản chi tiêu cao nhất"], y=df_household.iloc[:, 2:-1].sum(axis=1))
plt.xticks(rotation=45)
plt.xlabel("Danh mục chi tiêu cao nhất")
plt.ylabel("Tổng chi tiêu")
plt.title("Tổng chi tiêu theo danh mục cao nhất")
plt.show()
```



Hình 3.6: Hình ảnh kết quả các khoản chi tiêu của gia đình

### 3.7 Đánh giá mô hình và kiểm tra độ tin cậy

- Kiểm tra tính phù hợp của mô hình:

+ Để đảm bảo tính chính xác của mô hình hồi quy, chúng tôi kiểm tra các giả thuyết về phân phối của sai số và sự phù hợp của mô hình. Các sai số phải có phân phối chuẩn và không có sự phụ thuộc.

- Đánh giá độ chính xác của mô hình:

+ Chúng tôi đánh giá mô hình qua các chỉ số như  $R^2$  (đo lường mức độ giải thích của mô hình đối với biến phụ thuộc) và kiểm tra độ lệch chuẩn của sai số dự đoán.

### 3.8 Kết luận

- Qua phân tích dữ liệu tài chính cá nhân của hộ gia đình, chúng tôi nhận thấy rằng thu nhập, khoản vay và các yếu tố nhân khẩu học như độ tuổi có ảnh hưởng rõ rệt đến xu hướng chi tiêu của hộ gia đình. Cụ thể, các hộ gia đình có thu nhập cao có xu hướng chi tiêu nhiều hơn vào các dịch vụ giải trí và đầu tư tài chính, trong khi các hộ gia đình có thu nhập thấp chủ yếu chi tiêu cho các nhu cầu thiết yếu như thực phẩm và hóa đơn dịch vụ. Ngoài ra các cá nhân và các hộ gia đình có nhiều cá nhân có mức thu nhập cao sẽ có những khoản chi tiêu phóng khoáng hơn.

## KẾT LUẬN

### **Kết quả đạt được**

#### ***Kiến thức***

- Qua quá trình nghiên cứu và phân tích dữ liệu tài chính cá nhân của hộ gia đình, tôi đã thu được nhiều kiến thức bổ ích trong lĩnh vực tài chính cá nhân và thống kê, cụ thể:

+ Hiểu rõ hơn về xu hướng chi tiêu của hộ gia đình: Phân tích và nhận diện các yếu tố ảnh hưởng đến việc chi tiêu của hộ gia đình như thu nhập, khoản vay, và các yếu tố nhân khẩu học.

+ Kỹ năng phân tích dữ liệu: Nâng cao khả năng xử lý và phân tích dữ liệu, đặc biệt là việc sử dụng các phương pháp thống kê như hồi quy tuyến tính và kiểm định giả thuyết.

+ Sử dụng phần mềm phân tích dữ liệu: Thành thạo trong việc sử dụng các công cụ phần mềm như Excel, R hoặc Python để phân tích dữ liệu tài chính.

#### ***- Sản phẩm***

+ Báo cáo phân tích chi tiết: Báo cáo này cung cấp các phân tích chi tiết về chi tiêu của hộ gia đình, các yếu tố ảnh hưởng đến chi tiêu, và kết quả của mô hình hồi quy tuyến tính.

+ Mô hình dự đoán chi tiêu: Sản phẩm là một mô hình hồi quy tuyến tính có thể sử dụng để dự đoán xu hướng chi tiêu trong tương lai dựa trên các yếu tố như thu nhập và các khoản vay.

+ Biểu đồ và báo cáo trực quan: Các biểu đồ giúp minh họa rõ ràng các kết quả phân tích như phân phối chi tiêu, sự thay đổi của chi tiêu theo thu nhập, ...

#### ***- Ưu điểm và nhược điểm***

+ *Ưu điểm:*

Phân tích sâu sắc: Dự án đã giúp tôi hiểu rõ hơn về các yếu tố ảnh hưởng đến chi tiêu của hộ gia đình, đặc biệt là mối quan hệ giữa thu nhập và chi tiêu, từ đó cung cấp cái nhìn sâu sắc cho các chính sách tài chính cá nhân.

Kỹ năng phân tích tốt hơn: Tôi đã phát triển được khả năng phân tích số liệu, sử dụng các phương pháp thống kê và mô hình hóa dữ liệu tài chính, điều này rất hữu ích trong công việc nghiên cứu và thực tế.

Khả năng tự học: Qua dự án này, tôi đã học được cách tự giải quyết các vấn đề phát sinh trong phân tích dữ liệu, bao gồm việc xử lý dữ liệu thiếu và kiểm tra giả thuyết.

#### *+ Nhược điểm*

Vì thời gian có hạn, vì vậy không thể tiến hành khảo sát với một mẫu dữ liệu lớn hơn, từ đó sẽ giúp kết quả phân tích có độ tin cậy cao hơn.

Vì kiến thức còn hạn chế chưa thể áp dụng các kỹ thuật phân tích phức tạp hơn như hồi quy phi tuyến tính hoặc mô hình học máy (machine learning) để cải thiện độ chính xác của dự đoán.

Vì điều kiện đầu tư còn hạn hẹp nên việc sử dụng các công cụ phân tích dữ liệu nâng cao (như phần mềm phân tích dữ liệu đắt tiền) còn gặp khó khăn, khiến cho việc áp dụng các phương pháp phân tích hiện đại bị hạn chế.

#### **- Công việc tương lai**

+ Trong tương lai, nếu có thêm thời gian và nguồn lực, chúng tôi sẽ phát triển dự án hơn về nhiều khía cạnh hơn như sau:

Mở rộng nghiên cứu với mẫu dữ liệu lớn hơn: Điều này sẽ giúp làm tăng tính đại diện và độ tin cậy của kết quả nghiên cứu, đồng thời giúp mô hình phân tích chính xác hơn.

Áp dụng các phương pháp phân tích dữ liệu nâng cao: Tôi sẽ tìm hiểu và áp dụng các phương pháp phân tích phức tạp hơn như hồi quy phi tuyến tính, các mô hình học máy, để cải thiện khả năng dự đoán chi tiêu.

Tăng cường phân tích về yếu tố tâm lý: Nghiên cứu thêm về tác động của yếu tố tâm lý (như thói quen chi tiêu, ảnh hưởng xã hội) đến chi tiêu của hộ gia đình.

Xây dựng mô hình dự đoán chi tiêu trong tương lai: Tạo ra một hệ thống dự đoán chi tiêu cho các hộ gia đình trong tương lai, giúp các cá nhân và tổ chức quản lý tài chính hiệu quả hơn.

## PHÂN CÔNG NHIỆM VỤ

STT	MSV	Họ và tên	Nhiệm vụ
1	1771020650	Nguyễn Quang Thịnh	Triển khai dự án báo cáo, thu thập dữ liệu, triển khai dự án trên ngôn ngữ lập trình, kiểm tra và sửa báo cáo.
2	1771020743	Dương Văn Việt	Triển khai dự án báo cáo, kiểm tra và sửa dữ liệu, làm báo cáo, kiểm tra dự án.
3	1771020717	Vương Đức Tuấn	Hỗ trợ trong quá trình làm báo cáo.

**Toàn bộ mã nguồn:** <https://github.com/ThinhNguyen25/PhanTichThuNhapCaNhan.git>



## DANH MỤC TÀI LIỆU THAM KHẢO

- [1]. Bài toán về mô hình tuyến tính hồi quy, <https://viblo.asia/p/bai-toan-hoi-quy-va-mo-hinh-hoi-quy-tuyen-tinh-GrLZDQgOlk0>
- [2]. Hồi quy tuyến tính là gì, <https://aws.amazon.com/vi/whatis/linearregression/#:~:text=H%E1%BB%93i%20quy%20tuy%E1%BA%BFn%20t%C3%ADnh%20l%C3%A0%20m%E1%BB%99t%20k%E1%BB%B9%20thu%E1%BA%ADt,%C4%91%E1%BB%99c%20l%E1%BA%ADp%20nh%C6%B0%20m%E1%BB%99t%20ph%C6%B0%C6%A1ng%20tr%C3%ACnh%20tuy%E1%BA%BFn%20t%C3%ADnh.>
- [3]. Các khái niệm cơ bản trong xác suất thống kê, <https://rdsic.edu.vn/blog/toan/nhung-ung-dung-thuc-te-cua-cong-thuc-xac-suat-thong-ke-vi-cb.html>
- [4]. Các ứng dụng của xác suất thống kê, <https://www.ispacedanang.edu.vn/blog-1/xac-suat-thong-ke-vi-cb.html>
- [5]. Dữ liệu, <https://www.kaggle.com/>
- [6]. Công cụ hỗ trợ và tham khảo, <https://chatgpt.com/>