

---

**TRƯỜNG ĐẠI HỌC PHENIKAA**  
**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO BÀI TẬP LỚN**  
**MÔN HỌC: TRỰC QUAN HÓA DỮ LIỆU**

*Nhóm 5*

*Đề tài: Trực quan hóa dữ liệu mức độ phổ biến giữa chó và mèo ở Mỹ*

**Thành viên nhóm**

Trương Hương Giang	21010631	K15-CNTT2
Nguyễn Minh Phương	21012084	K15-CNTT4

***GVHD: Th.S Nguyễn Văn Thiệu***

**22/06/2024 – Hà Nội**

## MỤC LỤC

<b>MỞ ĐẦU .....</b>	<b>1</b>
<b>PHẦN 1. GIỚI THIỆU .....</b>	<b>2</b>
1.1. Trục quan hóa dữ liệu (Data Visualization).....	2
1.1.1. Khái niệm .....	2
1.1.2. Vai trò.....	2
1.2. Bài toán.....	2
1.2.1. Mô tả.....	2
1.2.2. Mục tiêu.....	3
1.2.3. Phương pháp.....	4
1.3. Công cụ sử dụng.....	4
1.3.1. Python và các thư viện đi kèm .....	4
1.3.2. Jupyter Notebook .....	5
1.3.3. Excel và Google Sheets.....	5
1.3.4. Tableau .....	5
1.3.5. SQL.....	5
1.3.6. Google Colab.....	5
<b>PHẦN 2. KIẾN THỨC NỀN TẢNG .....</b>	<b>7</b>
2.1. Kiến thức cơ bản về trục quan hóa dữ liệu .....	7
2.1.1. Khái niệm .....	7
2.1.2. Phương pháp.....	8
2.1.3. Công cụ quan trọng .....	9
2.2. Nguyên tắc và quy tắc thiết kế trục quan hóa dữ liệu .....	9
<b>PHẦN 3. PHÂN TÍCH DỮ LIỆU DỰA VÀO TRỤC QUAN HÓA .....</b>	<b>11</b>
3.1. Tiền xử lý dữ liệu .....	11
3.1.1. Thu thập dữ liệu.....	11
3.1.2. Chuẩn hóa dữ liệu.....	11
3.1.3. Biến đổi dữ liệu .....	11
3.1.4. Tích hợp dữ liệu (Data Integration) .....	11
3.1.5. Lưu trữ và quản lý dữ liệu (Data Storage and Management) .....	12
3.2. Phân tích dữ liệu.....	12
3.3. Quá trình thiết kế.....	14
3.3.1. Biểu đồ Number of Households (in 1000) .....	15
3.3.2. Biểu đồ Number of Pet Households (in 1000).....	16
3.3.12. Nhận xét tổng kết .....	25

3.4. Nguyên tắc thiết kế.....	27
3.4.1. Biểu đồ sử dụng.....	27
3.4.2. Thiết kế và xây dựng biểu đồ .....	27
3.4.3. Điều chỉnh và tối ưu hóa .....	28
<b>PHẦN 4. KẾT QUẢ VÀ THẢO LUẬN .....</b>	<b>29</b>
4.1. Kết quả.....	29
4.1.1. Kết quả so sánh Số lượng hộ gia đình và tỷ lệ hộ gia đình có thú cưng .	29
4.1.2. Kết quả so sánh Số lượng hộ gia đình và số lượng hộ nuôi thú cưng.	30
4.1.5. Kết quả so sánh Số hộ nuôi chó và mèo.....	33
4.1.6. Kết quả so sánh Tỷ lệ phần trăm chủ sở hữu chó và mèo.....	34
4.1.6. Kết quả so sánh Tỷ lệ phần trăm hộ gia đình nuôi thú cưng với số lượng hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo .....	35
4.1.7. Biểu đồ tổng quan.....	36
4.2. Ý nghĩa .....	38
4.2.1. Sự phân bố nuôi thú cưng theo từng bang.....	38
4.2.2. Số lượng hộ gia đình nuôi chó và mèo.....	39
4.2.3. Mối tương quan giữa các biến số .....	39
4.2.4. Nhận định tổng quát .....	39
<b>PHẦN 5. TỔNG KẾT BÀI TẬP LỚN .....</b>	<b>40</b>
5.1. Kế hoạch.....	40
5.1.1. Kế hoạch dự kiến.....	40
5.2. Kết quả đạt được.....	41
5.3. Ưu điểm và hạn chế.....	42
5.3.1. Ưu điểm .....	42
5.3.2. Hạn chế .....	42
<b>PHẦN 6. TÀI LIỆU THAM KHẢO.....</b>	<b>43</b>

MỤC LỤC HÌNH ẢNH

Hình ảnh 1. Các công cụ sử dụng trong python.....9

Hình ảnh 2. Cấu trúc dữ liệu .....12

Hình ảnh 3. Biểu đồ về số lượng hộ gia đình tại các địa điểm khác nhau (Number of Households (in 1000)).....15

Hình ảnh 4. Biểu đồ về số lượng hộ gia đình nuôi thú cưng tại các địa điểm khác nhau (Number of Pet Households (in 1000)) .....16

Hình ảnh 5. Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi thú cưng theo từng bang (Percentage of households with pets).....17

Hình ảnh 6. Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi chó theo từng bang (Percentage of Dog Owners) .....18

Hình ảnh 7. Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi mèo theo từng bang (Percentage of Cat Owners) .....19

Hình ảnh 8. Biểu đồ thể hiện số lượng hộ gia đình nuôi chó (đơn vị: 1000 hộ) theo từng bang (Dog Owning Households (1000s)) .....20

Hình ảnh 9. Biểu đồ thể hiện số lượng hộ gia đình nuôi mèo theo từng bang (Cat Owning Households).....21

Hình ảnh 10. Biểu đồ thể hiện số lượng trung bình chó nuôi trên mỗi hộ gia đình theo từng bang (Mean Number of Dogs per household) .....22

Hình ảnh 11. Biểu đồ thể hiện số lượng trung bình mèo nuôi trên mỗi hộ gia đình theo từng bang (Mean Number of Cats) .....23

Hình ảnh 12. Biểu đồ thể hiện dân số chó (tính bằng hàng ngàn) theo từng bang (Dog Population (in 1000)) .....24

Hình ảnh 13. Biểu đồ thể hiện dân số mèo theo từng bang giúp chúng ta hiểu rõ hơn về phân bố dân số mèo trên toàn quốc (Cat Population) .....25

Hình ảnh 14. Biểu đồ so sánh số lượng hộ gia đình và tỷ lệ hộ gia đình có thú cưng .....29

Hình ảnh 15. Biểu đồ so sánh số lượng hộ gia đình và số lượng hộ nuôi thú cưng .....30

Hình ảnh 16. Biểu đồ so sánh số lượng trung bình chó và mèo mỗi hộ gia đình 31

Hình ảnh 17. Biểu đồ so sánh dân số chó và mèo.....32

Hình ảnh 18. Biểu đồ so sánh số lượng hộ nuôi chó và mèo giữa các bang của Mỹ .....33

Hình ảnh 19. Biểu đồ so sánh tỷ lệ phần trăm chủ sở hữu giữa chó và mèo .....34

Hình ảnh 20. Biểu đồ so sánh Tỷ lệ phần trăm hộ gia đình nuôi thú cưng với số lượng hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo.....35

Hình ảnh 21. Biểu đồ tương quan .....37

Hình ảnh 22. Biểu đồ Gantt.....41

MỤC LỤC BẢNG BIỂU

Bảng biểu 1. Bảng so sánh so sánh dân số chó và mèo.....33

Bảng biểu 2. Bảng so sánh số lượng hộ nuôi chó và mèo giữa các bang của Mỹ  
.....34

Bảng biểu 3. Bảng so sánh tỷ lệ phần trăm chủ sở hữu giữa chó và mèo.....35

Bảng biểu 4. Bảng so sánh Tỷ lệ phần trăm hộ gia đình nuôi thú cưng với số lượng  
hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo .....36

Bảng biểu 5. Phân chia công việc .....40

## MỞ ĐẦU

Trong xã hội hiện đại của Mỹ, bài toán về dân số chó và mèo theo hộ gia đình không chỉ đơn thuần là về việc nuôi thú cưng mà còn là một phản ánh sâu sắc về cuộc sống và giá trị cá nhân của người dân. Với hơn 67% hộ gia đình nuôi ít nhất một con thú cưng, vấn đề này đã trở thành một phần không thể thiếu trong đời sống hàng ngày của nhiều gia đình, từ thành thị sầm uất cho đến các vùng quê yên bình.

Sự lựa chọn nuôi chó hay mèo không chỉ đơn giản là thú vui cá nhân mà còn phản ánh phong cách sống và giá trị văn hóa của mỗi gia đình. Việc tăng lên về số lượng chó và mèo trong mỗi hộ gia đình không chỉ là do thúc đẩy từ sự yêu thích với các loài vật nuôi mà còn là do lợi ích tâm lý và sức khỏe mà chúng mang lại cho con người. Chó và mèo không chỉ là những người bạn đồng hành mà còn được coi là thành viên thực sự trong gia đình, đóng vai trò quan trọng trong việc giảm stress, tăng cường sự gắn kết gia đình và cảm giác an toàn.

Tuy nhiên, thực tế cho thấy sự phân bố dân số chó và mèo không đồng đều trên khắp lãnh thổ Mỹ. Các bang với tỷ lệ nuôi thú cưng cao như Wyoming, West Virginia hay Nebraska cho thấy sự ưa chuộng và chấp nhận lớn đối với việc nuôi thú cưng. Ngược lại, các khu vực đô thị lớn như New York và California lại có tỷ lệ nuôi thú cưng thấp hơn, do những hạn chế về không gian sống và phong cách sống năng động hơn.

Việc nghiên cứu và phân tích sâu hơn về sự phân bố này không chỉ giúp chúng ta hiểu rõ hơn về xu hướng và thị hiếu nuôi thú cưng mà còn là cơ sở để đưa ra các chiến lược quản lý, chính sách hỗ trợ và giải pháp công cộng phù hợp. Điều này có thể bao gồm việc cải thiện điều kiện sống cho các loài vật nuôi, đề xuất các khu vực công cộng thân thiện với thú cưng, và thúc đẩy nhận thức cộng đồng về trách nhiệm và quản lý sức khỏe của các thú nuôi trong xã hội ngày càng phát triển và hoạt động.

Bài toán về dân số chó và mèo theo hộ gia đình không chỉ đơn thuần là về số liệu thống kê mà còn là một phần không thể thiếu trong việc hiểu sâu hơn về mối quan hệ giữa con người và động vật, đồng thời cũng là nền tảng để xây dựng các chính sách và chiến lược quản lý hiệu quả, nhằm mang lại lợi ích to lớn cho cả con người và môi trường sống chung.

## PHẦN 1. GIỚI THIỆU

### 1.1. Trực quan hóa dữ liệu (Data Visualization)

#### 1.1.1. Khái niệm

Trực quan hóa dữ liệu là quá trình sử dụng các yếu tố hình ảnh như đồ thị, biểu đồ hoặc bản đồ để trình bày dữ liệu. Quá trình này chuyển đổi dữ liệu phức tạp, có dung lượng lớn hoặc dữ liệu số thành hình ảnh trình bày trực quan có thể xử lý dễ dàng hơn. Các công cụ trực quan hóa dữ liệu cải thiện và tự động hóa quá trình giao tiếp bằng hình ảnh nhằm đảm bảo độ chính xác và chi tiết. Bạn có thể sử dụng những hình ảnh trình bày trực quan để trích xuất những thông tin chuyên sâu hữu ích từ dữ liệu thô.

#### 1.1.2. Vai trò

Trong các doanh nghiệp thường xử lý dữ liệu từ nhiều nguồn khác nhau ví dụ như:

- Trang web nội bộ và bên ngoài
- Thiết bị thông minh
- Hệ thống thu thập dữ liệu nội bộ
- Mạng xã hội

Tuy nhiên, các dữ liệu được hiển thị thường là dữ liệu thô, có thể khó hiểu và khó sử dụng dữ liệu để chuẩn bị và trình bày dữ liệu theo ngữ cảnh phù hợp theo mong muốn.

Vậy nên họ cần định hình dữ liệu ở dạng trực quan hóa để những người phụ trách đưa ra quyết định có thể xác định mối quan hệ giữa các dữ liệu và phát hiện ra các mẫu hoặc xu hướng ẩn trong dữ liệu.

Trực quan hóa dữ liệu tạo ra các thông điệp giúp nâng cao nghiệp vụ thông minh và hỗ trợ đưa ra quyết định cũng như lập kế hoạch chiến lược dựa trên dữ liệu.

### 1.2. Bài toán

#### 1.2.1. Mô tả

Bài toán phân tích dữ liệu dân số và quyền sở hữu chó và mèo theo hộ gia đình tại Mỹ là một nỗ lực để hiểu sâu hơn về mối quan hệ giữa dân số và hành vi nuôi thú cưng trong các tiểu bang của nước này. Trong nghiên cứu này, chúng



ta sẽ sử dụng các bộ dữ liệu cụ thể, chẳng hạn như số lượng hộ gia đình nuôi chó và mèo, tỷ lệ sở hữu thú cưng so với dân số, và các chỉ số kinh tế xã hội như thu nhập trung bình, mức đô thị hóa, và các chỉ số xã hội khác.

Bài toán phân tích dữ liệu dân số và quyền sở hữu chó mèo theo từng hộ gia đình tại Mỹ là một bài toán được phân tích giúp hiểu rõ hơn về số lượng hộ gia đình nuôi chó và mèo, tỷ lệ sở hữu thú cưng so với dân số cùng với các chỉ số kinh tế xã hội như thu nhập trung bình từng hộ gia đình, mức đô thị hóa cũng như một số chỉ số khác. Bài toán cũng giúp cho chúng ta hiểu sâu hơn về mối quan hệ giữa dân số và hành vi nuôi thú cưng trong các tiểu bang của nước này.

### **1.2.2. Mục tiêu**

Bài toán đưa ra 4 mục tiêu chính:

- Phân tích tỷ lệ sở hữu thú cưng: Điều này bao gồm việc tìm hiểu tỷ lệ hộ gia đình nuôi chó và mèo so với dân số tổng thể của từng tiểu bang. Thông qua việc phân tích này, chúng ta có thể nhận biết được mức độ phổ biến của việc nuôi thú cưng ở từng khu vực và so sánh giữa các tiểu bang.
- Định tính hóa xu hướng nuôi thú cưng: Bằng cách phân tích số lượng trung bình của chó và mèo mỗi hộ gia đình, chúng ta có thể xác định xu hướng và ưa thích của cư dân tại từng tiểu bang. Những thông tin này cung cấp cái nhìn sâu hơn về mức độ yêu thích và sẵn sàng chăm sóc thú cưng trong cộng đồng.
- Liên kết với chỉ số kinh tế xã hội: Việc phân tích sự phân bố chó và mèo theo các chỉ số kinh tế xã hội như thu nhập và mức đô thị hóa giúp ta hiểu được tầm quan trọng của yếu tố kinh tế và xã hội đối với hành vi nuôi thú cưng. Điều này có thể phản ánh mức độ phát triển kinh tế của khu vực và ảnh hưởng của nó đến nhu cầu và khả năng nuôi dưỡng thú cưng.
- Phân tích theo vị trí địa lý : Bằng cách phân tích theo địa lý, chúng ta có thể nhận thấy những xu hướng khác nhau trong việc nuôi thú cưng giữa các vùng miền và thành phố lớn nhỏ khác nhau. Điều này giúp cho việc đưa ra các chiến lược quản lý và tiếp thị hiệu quả hơn, phù hợp với từng đặc điểm địa phương.

Do đó, bài toán này không chỉ giúp ta hiểu sâu hơn về hành vi nuôi thú cưng mà còn mở ra cơ hội áp dụng những kiến thức này vào các lĩnh vực như y tế thú y, quản lý đô thị và tiếp thị sản phẩm. Bằng cách phân tích chi tiết các dữ liệu này, chúng ta có thể cung cấp những đánh giá chính xác và những khuyến nghị

cụ thể để quản lý và phát triển các chương trình nuôi thú cưng hiệu quả hơn trong cộng đồng.

### **1.2.3. Phương pháp**

Các phương pháp được sử dụng trong bài toán:

- Phân tích Tương quan (Correlation Analysis) dùng để phân tích mối tương quan giữa các biến số như tỷ lệ sở hữu chó và mèo, dân số, thu nhập trung bình, mức đô thị hóa, v.v. Sử dụng ma trận tương quan để tính toán các hệ số tương quan giữa các cặp biến số. Điều này giúp xác định mối quan hệ mạnh yếu giữa chúng và xác định những liên kết quan trọng.
- Phân tích Vùng (Spatial Analysis) dùng để đánh giá sự phân bố và xu hướng của sở hữu chó và mèo theo địa lý. Sử dụng các phương pháp như biểu đồ địa lý (GIS) và phân tích đồng nhất về vị trí (hot spot analysis) để phân tích các khu vực có mật độ nuôi thú cưng cao hơn so với các khu vực khác. Điều này có thể phát hiện ra các xu hướng địa lý và cung cấp thông tin quan trọng cho các chính sách và chiến lược quản lý.
- Phân tích Thống kê Mô tả (Descriptive Statistics) dùng để tóm tắt và mô tả các biến số chính như số lượng chó và mèo trung bình, tỷ lệ sở hữu thú cưng, v.v. Sử dụng các thống kê như trung bình, phương sai, và phân phối để biểu diễn các đặc tính của dữ liệu và giúp hình dung được sự phân bố của sở hữu chó và mèo trong dân số.

## **1.3. Công cụ sử dụng**

Trong bài toán phân tích dữ liệu dân số và quyền sở hữu chó và mèo theo hộ gia đình được chia theo tiểu bang tại Mỹ, chúng ta sử dụng nhiều công cụ khác nhau để thu thập, phân tích và trực quan hóa dữ liệu. Dưới đây là một số công cụ chính được sử dụng:

### **1.3.1. Python và các thư viện đi kèm**

Python là một ngôn ngữ lập trình được sử dụng rộng rãi trong các ứng dụng web, phát triển phần mềm, khoa học dữ liệu và máy học (ML). Các nhà phát triển sử dụng Python vì nó hiệu quả, dễ học và có thể chạy trên nhiều nền tảng khác nhau. Phần mềm Python được tải xuống miễn phí, tích hợp tốt với tất cả các loại hệ thống và tăng tốc độ phát triển.

- Thư viện Pandas: Sử dụng Python và Pandas để thao tác và xử lý dữ liệu. Pandas cung cấp các công cụ mạnh mẽ để làm sạch, sắp xếp và tóm tắt dữ liệu. Đọc dữ liệu từ các file (như CSV, Excel), làm sạch dữ liệu, tính toán các giá trị thống kê, và chuẩn bị dữ liệu cho các phân tích tiếp theo.
- Thư viện NumPy: Sử dụng NumPy để thực hiện các tính toán số học hiệu quả. Xử lý mảng số liệu, thực hiện các phép toán cơ bản và phức tạp trên dữ liệu.
- Thư viện Matplotlib và Seaborn: Trực quan hóa dữ liệu bằng cách tạo ra các biểu đồ và đồ thị. Tạo các biểu đồ cột, biểu đồ đường, biểu đồ tròn, biểu đồ nhiệt (heatmap) và các biểu đồ 3D để giúp hiểu rõ hơn về sự phân bố và mối quan hệ giữa các biến số.
- Thư viện Scikit-learn: Áp dụng các thuật toán phân tích và học máy. Sử dụng các mô hình hồi quy, phân cụm và các thuật toán học máy khác để phân tích dữ liệu và tìm ra các mẫu (patterns) trong dữ liệu.

### **1.3.2. Jupyter Notebook**

Cung cấp một môi trường tương tác để viết mã, chạy phân tích và trực quan hóa kết quả. Viết mã Python, thực hiện các phân tích và hiển thị các kết quả trực quan ngay trong cùng một tài liệu.

### **1.3.3. Excel và Google Sheets**

Thu thập và xử lý dữ liệu ban đầu, thực hiện các tính toán cơ bản và trực quan hóa đơn giản. Lưu trữ dữ liệu, tạo bảng và biểu đồ cơ bản, thực hiện các phép tính thống kê đơn giản.

### **1.3.4. Tableau**

Trực quan hóa dữ liệu phức tạp và tạo các bảng điều khiển tương tác. Tạo các báo cáo trực quan, biểu đồ và bảng điều khiển tương tác để trình bày kết quả phân tích một cách hiệu quả.

### **1.3.5. SQL**

Truy vấn và xử lý dữ liệu trong cơ sở dữ liệu quan hệ. Truy vấn dữ liệu từ các hệ quản trị cơ sở dữ liệu, thực hiện các phép lọc, nhóm và tổng hợp dữ liệu trước khi nhập vào các công cụ phân tích khác.

### **1.3.6. Google Colab**

Colaboratory hay còn gọi là Google Colab, là một sản phẩm từ Google Research, nó cho phép thực thi các câu lệnh Python trên nền tảng đám mây, đặc

biệt phù hợp với những bạn nào làm dự án (project) theo nhóm, hoặc muốn chia sẻ file code Python của mình với người khác ví dụ như nhà tuyển dụng. Giao diện của Google Colab rất giống với Jupyter Notebook, một công cụ giúp bạn chạy từng dòng lệnh Python một cách trực quan và kiểm tra kết quả câu lệnh ngay tại chỗ. Cung cấp một môi trường tương tác để viết mã, chạy phân tích và trực quan hóa kết quả. Viết mã Python, thực hiện các phân tích và hiển thị các kết quả trực quan ngay trong cùng một tài liệu.

## PHẦN 2. KIẾN THỨC NỀN TẢNG

### 2.1. Kiến thức cơ bản về trực quan hóa dữ liệu

#### 2.1.1. Khái niệm

Trực quan hóa dữ liệu là quá trình sử dụng các yếu tố hình ảnh như đồ thị, biểu đồ hoặc bản đồ để trình bày dữ liệu. Quá trình này chuyển đổi dữ liệu phức tạp, có dung lượng lớn hoặc dữ liệu số thành hình ảnh trình bày trực quan có thể xử lý dễ dàng hơn. Các công cụ trực quan hóa dữ liệu cải thiện và tự động hóa quá trình giao tiếp bằng hình ảnh nhằm đảm bảo độ chính xác và chi tiết. Bạn có thể sử dụng những hình ảnh trình bày trực quan để trích xuất những thông tin chuyên sâu hữu ích từ dữ liệu thô.

*Các thành phần của trực quan hóa dữ liệu:*

- Thông điệp: Muốn trình bày mục đích gì?

VD: Đo lường hiệu suất công việc hoặc doanh số bán hàng.

- Dữ liệu: sau khi có thông điệp, cần thu thập dữ liệu thích hợp. Sau đó, chỉnh sửa các định dạng dữ liệu hiện có, làm sạch dữ liệu, loại bỏ các giá trị ngoại lai và thực hiện phân tích sâu hơn. Sau đó, lên kế hoạch sử dụng các phương pháp khác nhau để khám phá trực quan.

- Phương tiện trực quan: Lựa chọn công cụ sau đó tạo biểu đồ làm nổi bật các điểm chính. Đơn giản hóa tập dữ liệu phức tạp và trình bày có hệ thống.

*Lợi ích của trực quan hóa dữ liệu:*

- Đưa ra quyết định chiến lược: Người có quyền đưa ra quyết định tiết kiệm thời gian, phân tích, đánh giá tổng quát hóa toàn cảnh.

- Cải thiện dịch vụ khách hàng: Làm nổi bật nhu cầu và mong muốn của khách hàng thông qua biểu diễn đồ họa, xác định những lỗ hổng trong dịch vụ khách hàng, cải thiện sản phẩm hoặc dịch vụ theo chiến lược và giảm hoạt động kém hiệu quả.

- Tăng mức độ tương tác của nhân viên: Tăng khả năng truyền đạt, dễ dàng biểu diễn kế hoạch và mục tiêu chung. Dễ dàng đo lường mục tiêu, tiến độ cũng như sự cải thiện của nhóm.

*Các bước trực quan hóa dữ liệu:*

- Xác định mục tiêu: Xác định các câu hỏi mà tập dữ liệu có thể trả lời.

- Thu thập dữ liệu: xác định nguồn dữ liệu bên trong và bên ngoài. Có sẵn các tập dữ liệu lớn trực tuyến để mua và sử dụng. Hoặc công ty cũng có thể có sẵn các kho lưu trữ dữ liệu hiện có để phân tích.

- Làm sạch dữ liệu: Làm sạch dữ liệu liên quan đến việc loại bỏ dữ liệu dư thừa. Thực hiện các phép tính toán để phân tích thêm hoặc lọc và chuyển đổi dữ liệu để đáp ứng các tiêu chí của câu hỏi.

- Chọn phương tiện trực quan hóa dữ liệu: Tùy thuộc vào mối quan hệ giữa dữ liệu và thông tin muốn truyền tải. Có thể chọn các loại biểu đồ khác nhau.

+ Trực quan hóa tĩnh

+ Trực quan hóa tương tác

- Tạo phương tiện trực quan hóa dữ liệu: sử dụng các công cụ trực quan hóa dữ liệu.

### **2.1.2. Phương pháp**

*Có 5 phương pháp chính, phụ thuộc vào mục đích:*

- Comparison: nếu chúng ta muốn biết sự hơn kém giữa các đối tượng nghiên cứu.

- Composition: nếu chúng ta muốn biết cụ thể về từng thành phần cấu tạo nên một thứ gì đó, một nhóm, v.v... và mỗi thành phần ấy chiếm tỷ lệ như thế nào.

- Distribution: nếu chúng ta muốn xem xét về phân phối dữ liệu, mức độ phân tán, mức độ tập trung của dữ liệu, hãy xác định các điểm bất thường, giá trị ngoại lệ của dữ liệu.

- Trend: nếu chúng ta muốn tìm hiểu về xu hướng biến động của đối tượng nghiên cứu trong dữ liệu.

- Relationship/Correlation: nếu chúng ta muốn tìm hiểu về mối liên hệ, tương quan giữa 2 hay nhiều đối tượng nghiên cứu trong dữ liệu

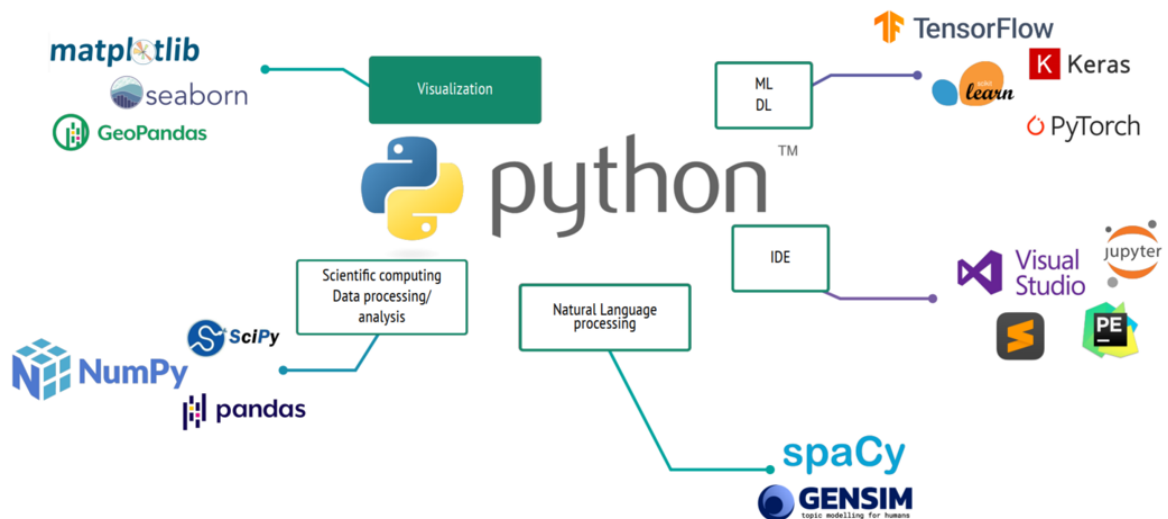
*Các biểu đồ thường được sử dụng trong trực quan hóa dữ liệu:*

- Biểu đồ cột, đường, hình tròn, ... để thể hiện xu hướng, so sánh giá trị.

- Đồ thị để thể hiện mối quan hệ giữa các yếu tố dữ liệu.

- Bản đồ heatmap để thể hiện mức độ tập trung của dữ liệu

### 2.1.3. Công cụ quan trọng



Hình ảnh 1. Các công cụ sử dụng trong python

- Google data studio / Looker studio: được sử dụng để có thể kết nối dễ dàng nhiều sản phẩm, sử dụng miễn phí.
- Tableau: giúp người dùng có thể dễ dàng sắp xếp dữ liệu và đầu vào từ nhiều nguồn đồng thời có thể tích hợp nhiều nền tảng khác.
- Infogram: có nhiều tùy chọn độc đáo, nhiều mẫu tối ưu hóa cho di động và mạng xã hội. Hoàn toàn có thể nhúng được vào các thiết kế trang web.
- D3.js: được sử dụng để tích hợp tính năng như mã nguồn mở, nhiều ví dụ tương tác, phân cấp và hoạt ảnh mạnh mẽ.
- Excel, PowerBI là bộ công cụ chuyên về DV của Microsoft, được sử dụng phù hợp các phân tích ngắn.
- Ngôn ngữ lập trình Python và các thư viện liên quan:
  - + Miễn phí, dễ học, dễ thực hiện, nhiều thư viện hỗ trợ.
  - + Có rất nhiều thư viện để người dùng có thể sử dụng như Matplotlib, Seaborn, Pandas, GeoPandas,...

## 2.2. Nguyên tắc và quy tắc thiết kế trực quan hóa dữ liệu

Các nguyên tắc và quy tắc thiết kế để trực quan hóa dữ liệu hiệu quả:

- Thu hút khán giả chú ý đến các chi tiết quan trọng thông qua kích cỡ, màu sắc, phông chữ và đồ họa.
- Cung cấp ngữ cảnh cho dữ liệu bằng các dấu hiệu trực quan.
- Chọn kiểu phối màu phù hợp cho phân tích dữ liệu.

- Sử dụng tiêu đề giải thích để cung cấp thông tin chuyên sâu chính cho khán giả và giúp họ tập trung vào đúng câu hỏi.
- Thêm nhãn và số rõ ràng cho dữ liệu.
- Chọn đúng loại biểu đồ: “Chức năng đi trước, trình bày đi sau”. Mục đích là cần được cân nhắc đầu tiên. Đặt câu hỏi để chọn biểu đồ phù hợp: Mục đích là gì? So sánh giá trị gì? Thể hiện xu hướng hay tìm hiểu phân phối và mối quan hệ giữa các biến.
- Không phải tất cả các data đều quan trọng: Không nên ném toàn bộ dataset vào 1 biểu đồ. Phân tích trước để biểu diễn những thông tin quan trọng cho người xem dễ nắm bắt.
- Biểu đồ thể hiện đúng tương quan số liệu thực tế: Không nên lợi dụng các yếu tố về khung tham chiếu, tỉ lệ hình ảnh để gây nhầm lẫn cho người xem.
  - Đưa ra những kiến thức cơ bản về trực quan hóa dữ liệu, bao gồm các khái niệm, phương pháp và công cụ quan trọng.
  - Trình bày các nguyên tắc và quy tắc thiết kế trực quan hóa dữ liệu hiệu quả.



## PHẦN 3. PHÂN TÍCH DỮ LIỆU DỰA VÀO TRỰC QUAN HÓA

### 3.1. Tiền xử lý dữ liệu

#### 3.1.1. Thu thập dữ liệu

- Nguồn dữ liệu: Thu thập dữ liệu từ American Veterinary Medical Association về cuộc khảo sát về dân số và quyền sở hữu thú cưng.
- Phương pháp thu thập: Sử dụng API hoặc tải xuống các tệp dữ liệu từ trang web <https://public.tableau.com/app/learn/sample-data> này.
- Thu thập dữ liệu thông qua khảo sát trực tiếp hoặc thông qua các nguồn dữ liệu mở.

#### 3.1.2. Chuẩn hóa dữ liệu

- Kiểm tra tính nhất quán:
  - + Đảm bảo các cột và tên cột nhất quán giữa các bộ dữ liệu khác nhau.
  - + Kiểm tra các đơn vị đo lường để đảm bảo sự nhất quán (ví dụ: số lượng hộ gia đình tính bằng nghìn hoặc đơn vị khác).
- Xử lý giá trị thiếu (Missing Values):
  - + Xác định và xử lý các giá trị bị thiếu trong dữ liệu.
  - + Sử dụng các phương pháp như xóa các bản ghi thiếu dữ liệu quan trọng, hoặc sử dụng các kỹ thuật ước lượng để điền vào giá trị thiếu.

#### 3.1.3. Biến đổi dữ liệu

- Tạo các cột tính toán mới: Tính toán các giá trị trung bình, tổng, hoặc tỷ lệ phần trăm từ các cột dữ liệu gốc.
- Chuẩn hóa dữ liệu số: Sử dụng các phương pháp như chuẩn hóa (normalization) hoặc chuẩn hóa Z-score (standardization) để đưa các giá trị về cùng một thang đo, giúp dễ dàng so sánh.
- Làm sạch dữ liệu (Data Cleaning):
  - + Loại bỏ các dữ liệu trùng lặp.
  - + Sửa chữa các lỗi dữ liệu như sai chính tả hoặc lỗi nhập liệu.

#### 3.1.4. Tích hợp dữ liệu (Data Integration)

- Kết hợp các bộ dữ liệu: Kết hợp dữ liệu từ các nguồn khác nhau để tạo thành một bộ dữ liệu thống nhất. Ví dụ: Ghép nối dữ liệu dân số của các bang với dữ liệu về quyền sở hữu chó và mèo từ các nguồn khác nhau.

- Xác thực và kiểm tra dữ liệu (Data Validation and Verification)
- + Kiểm tra sự nhất quán và chính xác: Sử dụng các phương pháp kiểm tra dữ liệu để đảm bảo dữ liệu đã được làm sạch và chuẩn hóa chính xác.

+ Xác nhận tính hợp lệ của dữ liệu: Sử dụng các phương pháp thống kê và trực quan hóa để kiểm tra tính hợp lệ của dữ liệu.

3.1.5. Lưu trữ và quản lý dữ liệu (Data Storage and Management)

- Lưu trữ dữ liệu sạch: Lưu trữ dữ liệu đã được tiền xử lý vào các hệ thống quản lý cơ sở dữ liệu hoặc các tệp tin để sử dụng trong phân tích tiếp theo.
- Đảm bảo sao lưu dữ liệu để tránh mất mát dữ liệu.

3.2. Phân tích dữ liệu

	Location	Number of Households (in 1000)	Percentage of households with pets	Number of Pet Households (in 1000)	Percentage of Dog Owners	Dog Owning Households (1000s)	Mean Number of Dogs per household	Dog Population (in 1000)	Percentage of Cat Owners	Cat Owning Households	Mean Number of Cats	Cat Population
0	Alabama	1828	59.5	1088	44.1	807	1.7	1410	27.4	501	2.5	1252
1	Arizona	2515	59.5	1497	40.1	1008	1.8	1798	29.6	743	1.9	1438
2	Arkansas	1148	62.4	716	47.9	550	2.0	1097	30.6	351	2.3	810
3	California	12974	52.9	6865	32.8	4260	1.6	6687	28.3	3687	1.9	7118
4	Colorado	1986	61.3	1217	42.5	845	1.6	1349	32.3	642	1.9	1191
5	Connecticut	1337	54.4	728	28.3	379	1.3	507	31.9	427	1.9	796
6	Delaware	334	56.6	189	33.7	113	1.4	163	33.7	113	1.7	187
7	District of Columbia	287	21.9	63	13.1	38	1.1	42	11.6	33	1.9	63
8	Florida	7609	54.4	4138	35.7	2718	1.5	4210	27.3	2079	2.1	4375
9	Georgia	3798	55.1	2093	40.1	1522	1.6	2479	27.3	1037	2.1	2162
10	Idaho	568	62.0	352	42.7	242	1.5	357	34.6	196	2.0	393
11	Illinois	5026	51.8	2602	32.4	1627	1.5	2365	26.3	1321	1.9	2453
12	Indiana	2478	59.9	1484	39.9	989	1.6	1619	34.4	852	2.2	1912
13	Iowa	1219	53.6	654	33.4	407	1.5	610	30.3	370	2.2	805
14	Kansas	1133	61.0	691	42.3	480	1.6	774	33.3	378	1.9	731
15	Kentucky	1777	61.6	1094	45.9	816	1.9	1531	36.8	654	2.1	1349
16	Louisiana	1702	55.1	937	36.4	619	1.8	1115	25.9	441	2.0	877
17	Maine	548	62.9	345	34.6	190	1.6	300	46.4	254	1.9	498
18	Maryland	2169	52.3	1134	30.8	667	1.4	915	29.8	645	2.6	1677
19	Massachusetts	2618	50.4	1318	23.6	618	1.4	850	34.1	892	1.8	1593
20	Michigan	3804	55.4	2108	34.6	1318	1.5	2036	31.3	1192	2.0	2420
21	Minnesota	2163	53.0	1146	31.9	690	1.4	934	29.7	643	2.0	1264
22	Mississippi	1115	56.4	629	45.2	504	1.7	846	29.1	324	2.1	668
23	Missouri	2498	61.4	1534	45.9	1148	1.7	1978	32.2	805	2.1	1653
24	Montana	410	61.3	251	41.2	169	1.7	282	33.6	138	2.0	277
25	Nebraska	710	51.3	364	33.8	240	1.6	374	31.3	222	2.3	514
26	Nevada	986	55.6	548	37.1	366	1.6	578	30.3	299	2.1	625
27	New Hampshire	508	56.8	289	30.3	154	1.4	212	34.2	174	1.8	309
28	New Jersey	3177	50.7	1611	32.4	1028	1.3	1340	25.3	803	1.8	1468
29	New Mexico	773	67.6	523	46.0	356	2.0	703	32.0	247	2.2	533
30	New York	7512	50.6	3802	29.0	2177	1.4	3054	29.1	2189	1.9	4261
31	North Carolina	3701	56.4	2089	40.3	1491	1.7	2518	29.5	1090	2.0	2220
32	North Dakota	272	53.9	147	36.1	98	1.4	139	31.4	85	2.0	174
33	Ohio	4661	57.4	2677	36.6	1708	1.6	2730	33.3	1553	2.4	3786
34	Oklahoma	1479	58.9	872	43.2	638	2.1	1327	32.6	482	2.2	1041
35	Oregon	1505	63.6	957	38.8	584	1.6	917	40.2	605	2.0	1185
36	Pennsylvania	5172	56.9	2942	32.9	1702	1.5	2485	33.8	1748	2.0	3544
37	Rhode Island	434	53.0	230	29.3	127	1.3	161	27.6	120	1.8	212
38	South Carolina	1759	54.1	951	38.6	678	1.8	1191	27.8	490	2.1	1039
39	South Dakota	333	65.6	219	42.8	143	1.5	220	39.1	130	2.2	290
40	Tennessee	2583	59.6	1540	44.1	1140	1.9	2157	29.8	770	2.3	1749
41	Texas	9002	58.5	5265	44.0	3960	1.8	7163	28.3	2544	2.2	5565
42	Utah	930	51.2	476	29.4	273	1.5	410	24.6	229	2.0	455
43	Vermont	265	70.8	188	37.7	100	1.4	142	49.5	131	1.8	234
44	Virginia	3017	53.4	1611	35.4	1069	1.6	1699	29.0	876	2.1	1855
45	Washington	2632	62.7	1649	36.3	954	1.7	1609	39.0	1028	1.8	1844
46	West Virginia	765	62.1	475	45.8	350	1.8	648	38.1	291	2.2	628
47	Wisconsin	2350	57.5	1352	33.9	796	1.4	1138	33.0	776	1.9	1510
48	Wyoming	221	61.8	137	38.8	86	1.5	125	33.9	75	1.9	144

Hình ảnh 2. Cấu trúc dữ liệu

Dựa vào hình 2 cho thấy được cấu trúc dữ liệu của bài toán:

- Location: Địa điểm cụ thể, có thể là các bang hoặc khu vực.
- Number of Households (in 1000): Tổng số hộ gia đình trong từng bang, tính theo đơn vị hàng nghìn.

- Percentage of households with pets: Tỷ lệ phần trăm các hộ gia đình trong bang có ít nhất một thú cưng.
- Number of Pet Households (in 1000): Số lượng hộ gia đình có thú cưng trong từng bang, tính theo đơn vị hàng nghìn.
- Percentage of Dog Owners: Tỷ lệ phần trăm các hộ gia đình trong bang sở hữu ít nhất một con chó.
- Dog Owning Households (1000s): Số lượng hộ gia đình sở hữu chó trong từng bang, tính theo đơn vị hàng nghìn.
- Mean Number of Dogs per household: Số lượng chó trung bình trong mỗi hộ gia đình sở hữu chó.
- Dog Population (in 1000): Tổng số chó trong từng bang, tính theo đơn vị hàng nghìn.
- Percentage of Cat Owners: Tỷ lệ phần trăm các hộ gia đình trong bang sở hữu ít nhất một con mèo.
- Cat Owning Households: Số lượng hộ gia đình sở hữu mèo trong từng bang, tính theo đơn vị hàng nghìn.
- Mean Number of Cats per household: Số lượng mèo trung bình trong mỗi hộ gia đình sở hữu mèo.
- Cat Population: Tổng số mèo trong từng bang, tính theo đơn vị hàng nghìn.

*Nhận xét về dữ liệu:*

- Số lượng hộ gia đình: Dữ liệu cung cấp số lượng hộ gia đình theo từng địa điểm, giúp chúng ta thấy được sự phân bố dân cư và mật độ hộ gia đình ở các khu vực khác nhau.
- Số lượng chó và mèo: Số lượng chó và mèo theo từng địa điểm giúp hiểu được sự phân bố thú cưng ở các khu vực khác nhau. Điều này có thể liên quan đến nhiều yếu tố như văn hóa, không gian sống, và thu nhập của từng khu vực.
- Số lượng trung bình chó và mèo mỗi hộ gia đình: Thông tin này giúp chúng ta hiểu được mức độ ưa chuộng chó và mèo của các hộ gia đình tại từng khu vực. Nó cũng có thể chỉ ra sự khác biệt về lối sống và điều kiện sống.

- Số lượng hộ gia đình nuôi chó và mèo: Dữ liệu này cho thấy sự phổ biến của việc nuôi chó và mèo trong các hộ gia đình ở từng địa điểm. Nó có thể cung cấp thông tin về xu hướng nuôi thú cưng tại các khu vực khác nhau.
- Tỷ lệ phần trăm hộ gia đình có nuôi thú cưng: Tỷ lệ này cho thấy mức độ phổ biến của việc nuôi thú cưng trong tổng số hộ gia đình ở từng khu vực. Đây là một chỉ số quan trọng để hiểu được mức độ ưa chuộng thú cưng và có thể liên quan đến các yếu tố văn hóa và kinh tế.

Dữ liệu này cung cấp cái nhìn tổng quan về phân bố số lượng hộ gia đình và thú cưng tại các địa điểm khác nhau. Nó giúp hiểu rõ hơn về sự khác biệt và xu hướng nuôi thú cưng trong các khu vực cụ thể.

Những thông tin này có thể được sử dụng để nghiên cứu thị trường thú cưng, lập kế hoạch kinh doanh cho các dịch vụ liên quan đến thú cưng, và thậm chí là để thiết kế các chính sách công cộng liên quan đến việc nuôi thú cưng trong xã hội.

### **3.3. Quá trình thiết kế**

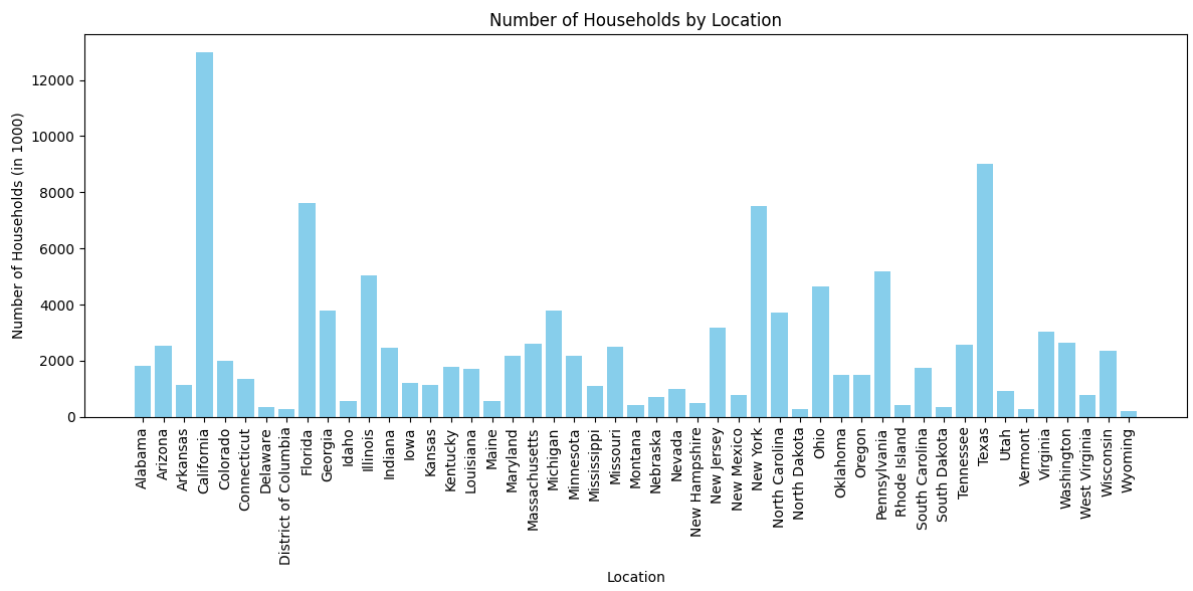
Để trực quan hóa dữ liệu toàn bộ phần dữ liệu đã phân tích ở trên bao gồm các bước sau:

- Bước 1: Thu thập và chuẩn bị dữ liệu: bao gồm nguồn dữ liệu và làm sạch các dữ liệu.
- Bước 2: Lựa chọn loại biểu đồ phù hợp: Lựa chọn và sử dụng các biểu đồ phù hợp cho trực quan hóa các dữ liệu trên.
- Bước 3: Thiết kế và xây dựng biểu đồ: Sử dụng các công cụ phần mềm để tạo biểu đồ, cấu hình các thông số biểu đồ và nêu rõ ràng và phản ánh đúng nội dung dữ liệu. Gắn nhãn đầy đủ và đơn vị đo lường phù hợp. Sử dụng màu sắc để phân biệt các nhóm dữ liệu khác nhau và tăng tính trực quan.
- Bước 4: Phân tích và nhận xét: Tổng quát về sự phân bố và mức độ phổ biến của việc nuôi chó, mèo trên toàn quốc. Nhận xét cụ thể khi đi sâu vào từng nhóm dữ liệu, như tỷ lệ hộ gia đình nuôi chó, mèo; số lượng hộ gia đình nuôi chó, mèo; số lượng trung bình chó, mèo nuôi trên mỗi hộ gia đình; và dân số chó, mèo theo từng bang.

- Bước 5: Điều chỉnh và tối ưu hóa các dữ liệu. Đảm bảo rằng dữ liệu được biểu diễn chính xác và không có lỗi. Tối ưu hóa trực quan và kết hợp các biểu đồ vào một báo cáo toàn diện, bao gồm cả các phân tích và nhận xét chi tiết.

Quá trình này đảm bảo rằng dữ liệu được biểu diễn một cách rõ ràng, dễ hiểu, và có thể sử dụng để đưa ra các quyết định hoặc nhận định quan trọng về việc nuôi chó, mèo trên toàn quốc.

3.3.1. Biểu đồ Number of Households (in 1000)



Hình ảnh 3. Biểu đồ về số lượng hộ gia đình tại các địa điểm khác nhau  
(Number of Households (in 1000))

Dựa vào hình 3 cho thấy được: Biểu đồ về số lượng hộ gia đình tại các địa điểm khác nhau cung cấp cái nhìn tổng quan về sự phân bố dân số theo khu vực. Những cột cao đại diện cho các địa điểm có số lượng hộ gia đình cao, thường là các khu vực đô thị lớn hoặc các thành phố với mật độ dân số cao. Những nơi này thường là trung tâm kinh tế, chính trị, hoặc văn hóa, thu hút nhiều người đến sinh sống và làm việc.

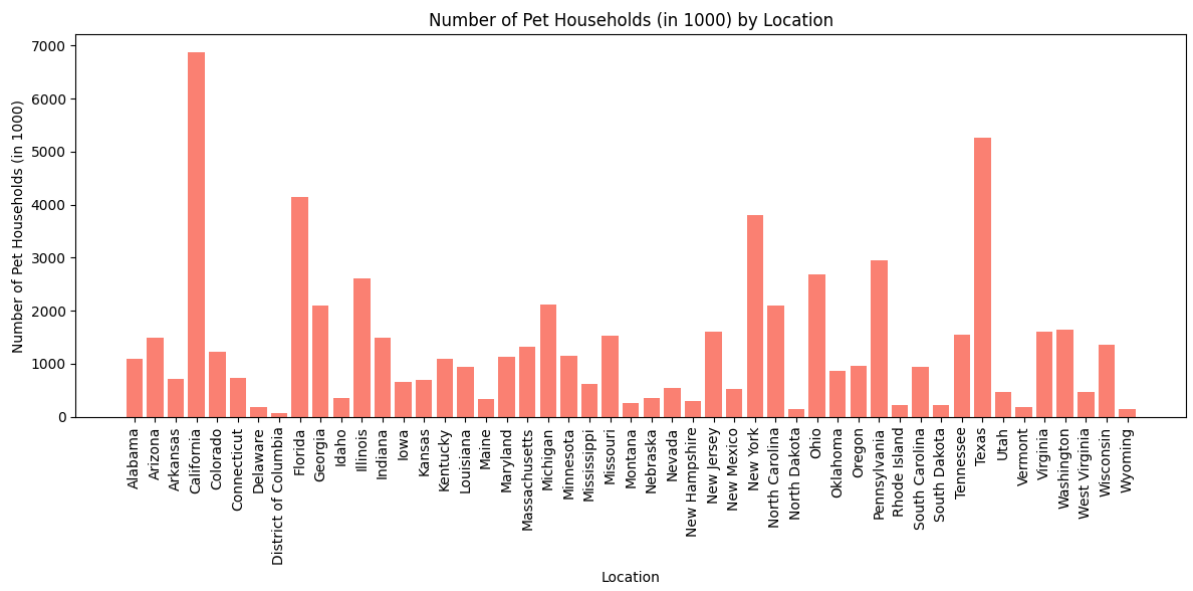
Ngược lại, những cột thấp biểu thị các khu vực có số lượng hộ gia đình thấp, thường là các vùng nông thôn hoặc các khu vực kém phát triển hơn. Các khu vực này có thể thiếu cơ sở hạ tầng, cơ hội việc làm ít hơn, hoặc điều kiện sống không thuận lợi, dẫn đến dân số ít hơn.

Sự chênh lệch lớn về chiều cao các cột cho thấy sự phân bố không đồng đều của số lượng hộ gia đình giữa các địa điểm. Điều này phản ánh mật độ dân cư khác nhau và có thể liên quan đến mức độ phát triển kinh tế và cơ sở hạ tầng của từng khu vực. Một số khu vực nổi bật với cột cao vượt trội, cho thấy số lượng hộ gia đình tại đó rất lớn, ví dụ như các thành phố lớn như New York hoặc Los

Angeles. Ngược lại, các khu vực có cột thấp hơn, chẳng hạn như các bang Wyoming hoặc Vermont, có thể có số lượng hộ gia đình ít hơn do dân số tổng thể thấp hơn.

Biểu đồ này cho thấy rõ ràng sự khác biệt về mật độ dân cư và sự phát triển giữa các khu vực khác nhau, từ đó có thể đưa ra các chiến lược phát triển phù hợp cho từng địa phương.

3.3.2. Biểu đồ Number of Pet Households (in 1000)



Hình ảnh 4. Biểu đồ về số lượng hộ gia đình nuôi thú cưng tại các địa điểm khác nhau (Number of Pet Households (in 1000))

Dựa vào hình 4 cho thấy được: Biểu đồ về số lượng hộ gia đình nuôi thú cưng tại các địa điểm khác nhau cung cấp cái nhìn tổng quan về sự phân bố số lượng hộ gia đình nuôi thú cưng theo khu vực. Những cột cao đại diện cho các địa điểm có số lượng hộ gia đình nuôi thú cưng cao, thường là các vùng ngoại ô hoặc các thành phố nơi người dân có xu hướng nuôi thú cưng nhiều hơn. Những khu vực này thường có điều kiện sống tốt cho thú cưng, chẳng hạn như nhà cửa có sân vườn, môi trường sống rộng rãi, và nhiều dịch vụ liên quan đến thú cưng.

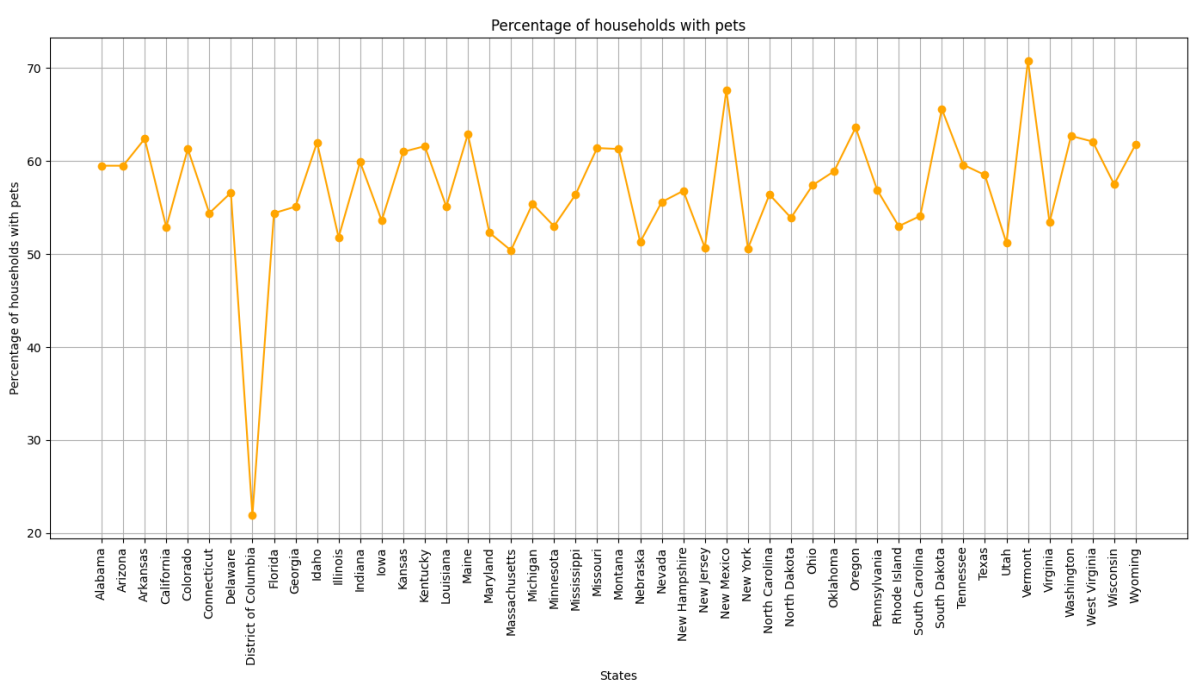
Ngược lại, những cột thấp biểu thị các khu vực có số lượng hộ gia đình nuôi thú cưng thấp, thường là các khu vực đô thị có không gian sống hạn chế hoặc nơi dân cư chủ yếu sống ở các căn hộ chung cư. Các yếu tố khác có thể ảnh hưởng đến số lượng hộ gia đình nuôi thú cưng bao gồm văn hóa địa phương, điều kiện khí hậu, và mức độ phổ biến của việc nuôi thú cưng.

Sự chênh lệch lớn về chiều cao các cột cho thấy sự phân bố không đồng đều của số lượng hộ gia đình nuôi thú cưng giữa các địa điểm. Điều này có thể phản ánh mức độ yêu thích và khả năng nuôi thú cưng ở từng khu vực. Một số khu

vực nổi bật với cột cao vượt trội, cho thấy số lượng hộ gia đình nuôi thú cưng tại đó rất lớn, chẳng hạn như các bang California, Texas, và Florida, nơi điều kiện sống và văn hóa nuôi thú cưng phát triển mạnh mẽ. Ngược lại, các khu vực có cột thấp hơn, chẳng hạn như các thành phố lớn như New York City hoặc San Francisco, có thể có số lượng hộ gia đình nuôi thú cưng ít hơn do không gian sống hạn chế.

Biểu đồ này cho thấy rõ ràng sự khác biệt về xu hướng và khả năng nuôi thú cưng giữa các khu vực khác nhau, từ đó có thể đưa ra các chiến lược phát triển phù hợp cho từng địa phương trong việc hỗ trợ và thúc đẩy văn hóa nuôi thú cưng.

3.3.3. Biểu đồ Percentage of households with pets



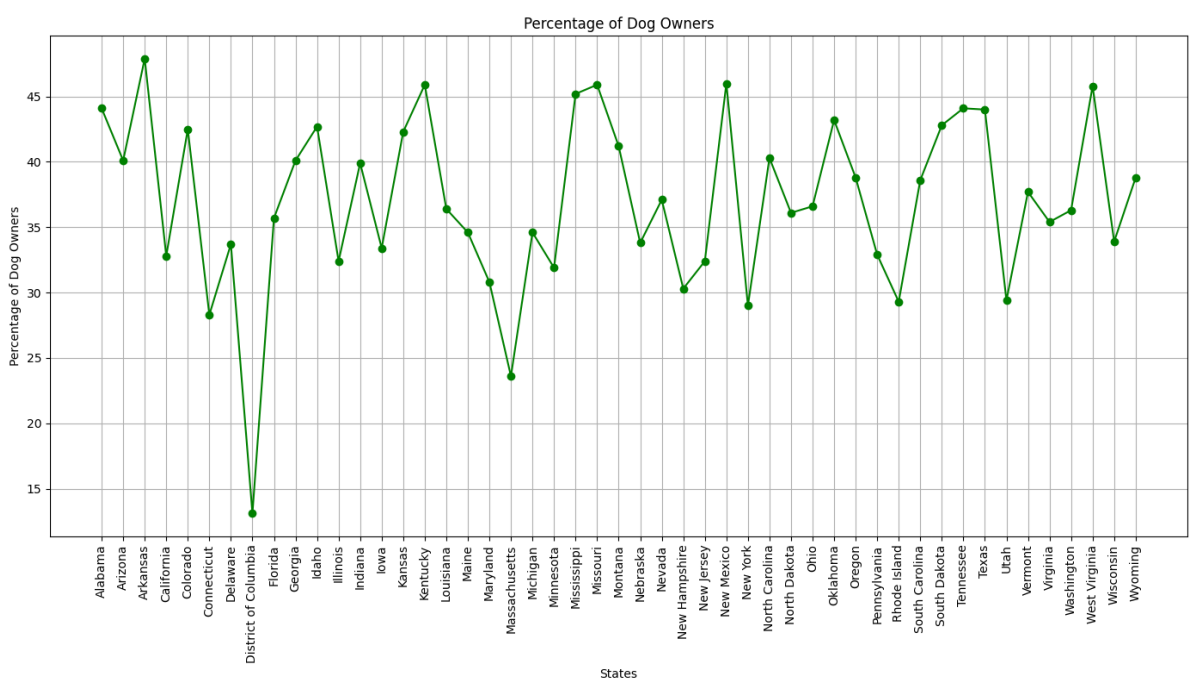
Hình ảnh 5. Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi thú cưng theo từng bang (Percentage of households with pets)

Dựa vào hình 5 cho thấy được: Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi thú cưng theo từng bang cho thấy sự phân bố của các hộ gia đình có thú cưng trên toàn quốc, với tỷ lệ này dao động rất lớn giữa các bang. Một số bang có tỷ lệ hộ gia đình nuôi thú cưng rất cao, cho thấy văn hóa và thói quen nuôi thú cưng tại những bang này rất phổ biến. Ngược lại, một số bang có tỷ lệ hộ gia đình nuôi thú cưng thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi thú cưng.

Những bang có tỷ lệ hộ gia đình nuôi thú cưng trên 70% thể hiện sự phổ biến của việc nuôi thú cưng, phản ánh một văn hóa yêu thích thú cưng mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi thú cưng, hoặc sự hiện diện của nhiều dịch

vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi thú cưng. Những bang có tỷ lệ hộ gia đình nuôi thú cưng trong khoảng 50% - 70% cho thấy việc nuôi thú cưng khá phổ biến nhưng không ở mức độ cao nhất, với các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi thú cưng của các hộ gia đình. Cuối cùng, những bang có tỷ lệ hộ gia đình nuôi thú cưng dưới 50% cho thấy việc nuôi thú cưng không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi thú cưng.

3.3.4. Biểu đồ Percentage of Dog Owners



Hình ảnh 6. Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi chó theo từng bang (Percentage of Dog Owners)

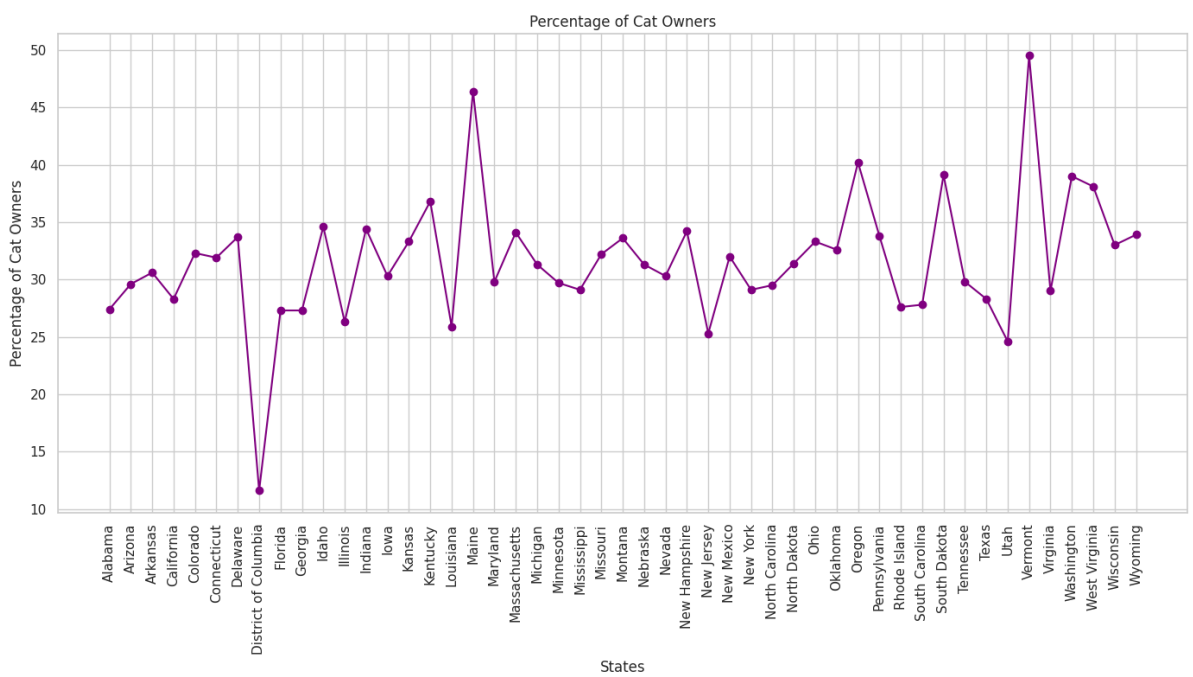
Dựa vào hình 6 cho thấy được: Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi chó theo từng bang cho thấy sự phân bố của các hộ gia đình có chó trên toàn quốc. Tỷ lệ phần trăm các hộ gia đình nuôi chó dao động lớn giữa các bang, với một số bang có tỷ lệ hộ gia đình nuôi chó rất cao, thể hiện sự phổ biến của việc nuôi chó ở những bang này. Ngược lại, một số bang có tỷ lệ hộ gia đình nuôi chó thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi chó.

Những bang có tỷ lệ hộ gia đình nuôi chó trên 50% cho thấy việc nuôi chó rất phổ biến, có thể phản ánh văn hóa yêu thích chó mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi chó, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi chó. Những bang có tỷ lệ hộ gia đình nuôi chó trong khoảng 30% - 50% cho thấy việc nuôi chó khá phổ biến nhưng không ở mức độ cao nhất,



với các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi chó của các hộ gia đình. Cuối cùng, những bang có tỷ lệ hộ gia đình nuôi chó dưới 30% cho thấy việc nuôi chó không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi chó.

3.3.5. Biểu đồ Percentage of Cat Owners



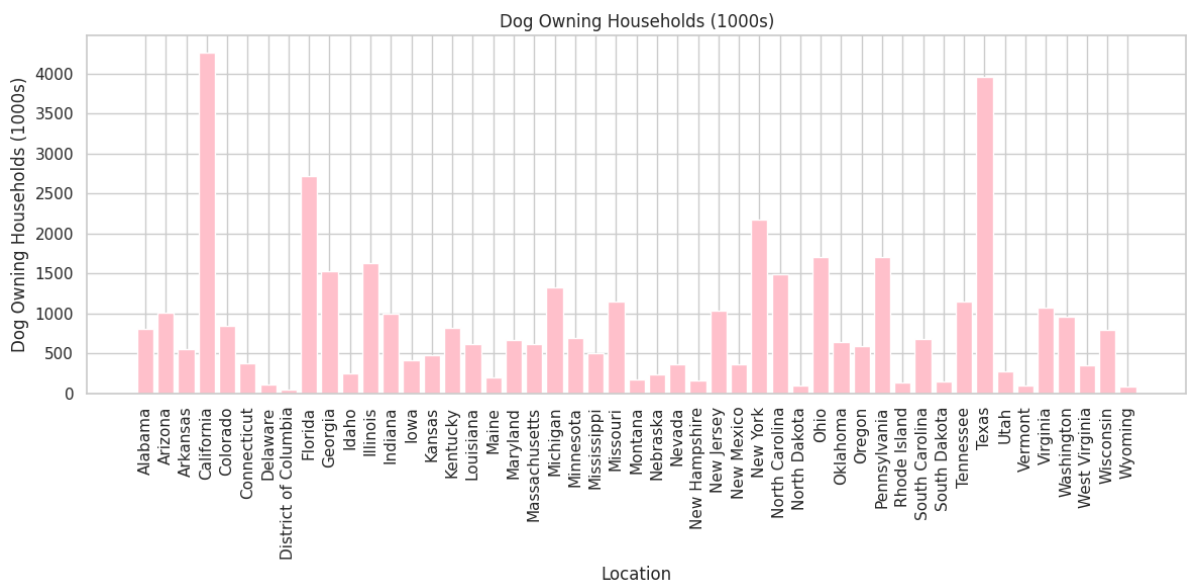
Hình ảnh 7. Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi mèo theo từng bang (Percentage of Cat Owners)

Dựa vào hình 7 cho thấy được: Biểu đồ thể hiện tỷ lệ phần trăm các hộ gia đình nuôi mèo theo từng bang cho thấy sự phân bố của các hộ gia đình có mèo trên toàn quốc. Tỷ lệ phần trăm các hộ gia đình nuôi mèo dao động lớn giữa các bang, với một số bang có tỷ lệ hộ gia đình nuôi mèo rất cao, thể hiện sự phổ biến của việc nuôi mèo ở những bang này. Ngược lại, một số bang có tỷ lệ hộ gia đình nuôi mèo thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi mèo.

Những bang có tỷ lệ hộ gia đình nuôi mèo trên 50% cho thấy việc nuôi mèo rất phổ biến, có thể phản ánh văn hóa yêu thích mèo mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi mèo, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi mèo. Những bang có tỷ lệ hộ gia đình nuôi mèo trong khoảng 30% - 50% cho thấy việc nuôi mèo khá phổ biến nhưng không ở mức độ cao nhất, với các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi mèo của các hộ gia đình. Cuối cùng, những bang có tỷ lệ hộ gia đình nuôi mèo dưới 30% cho thấy việc nuôi mèo

không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi mèo.

3.3.6. Biểu đồ Dog Owning Households (1000s)

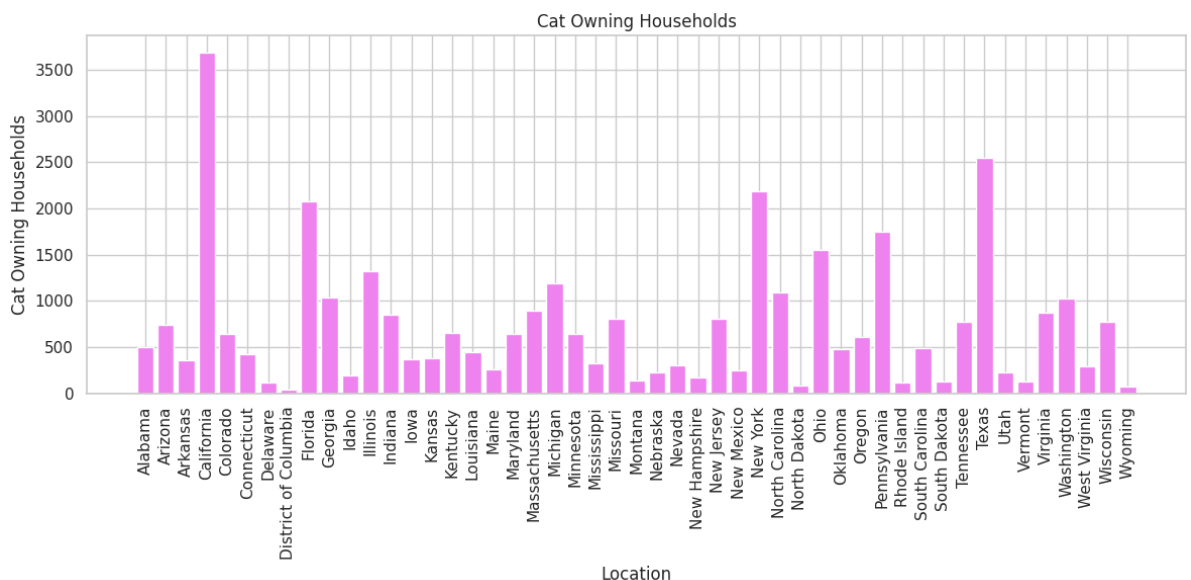


Hình ảnh 8. Biểu đồ thể hiện số lượng hộ gia đình nuôi chó (đơn vị: 1000 hộ) theo từng bang (Dog Owning Households (1000s))

Dựa vào hình 8 cho thấy được: Biểu đồ thể hiện số lượng hộ gia đình nuôi chó (đơn vị: 1000 hộ) theo từng bang cho thấy sự phân bố của các hộ gia đình nuôi chó trên toàn quốc. Số lượng hộ gia đình nuôi chó dao động lớn giữa các bang, với một số bang có số lượng hộ gia đình nuôi chó rất cao, thể hiện sự phổ biến của việc nuôi chó ở những bang này. Ngược lại, một số bang có số lượng hộ gia đình nuôi chó thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi chó.

Những bang có số lượng hộ gia đình nuôi chó trên 1000 hộ cho thấy việc nuôi chó rất phổ biến, có thể phản ánh văn hóa yêu thích chó mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi chó, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi chó. Những bang có số lượng hộ gia đình nuôi chó trong khoảng 500 - 1000 hộ cho thấy việc nuôi chó khá phổ biến nhưng không ở mức độ cao nhất, với các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi chó của các hộ gia đình. Cuối cùng, những bang có số lượng hộ gia đình nuôi chó dưới 500 hộ cho thấy việc nuôi chó không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi chó.

3.3.7. Biểu đồ Cat Owning Households

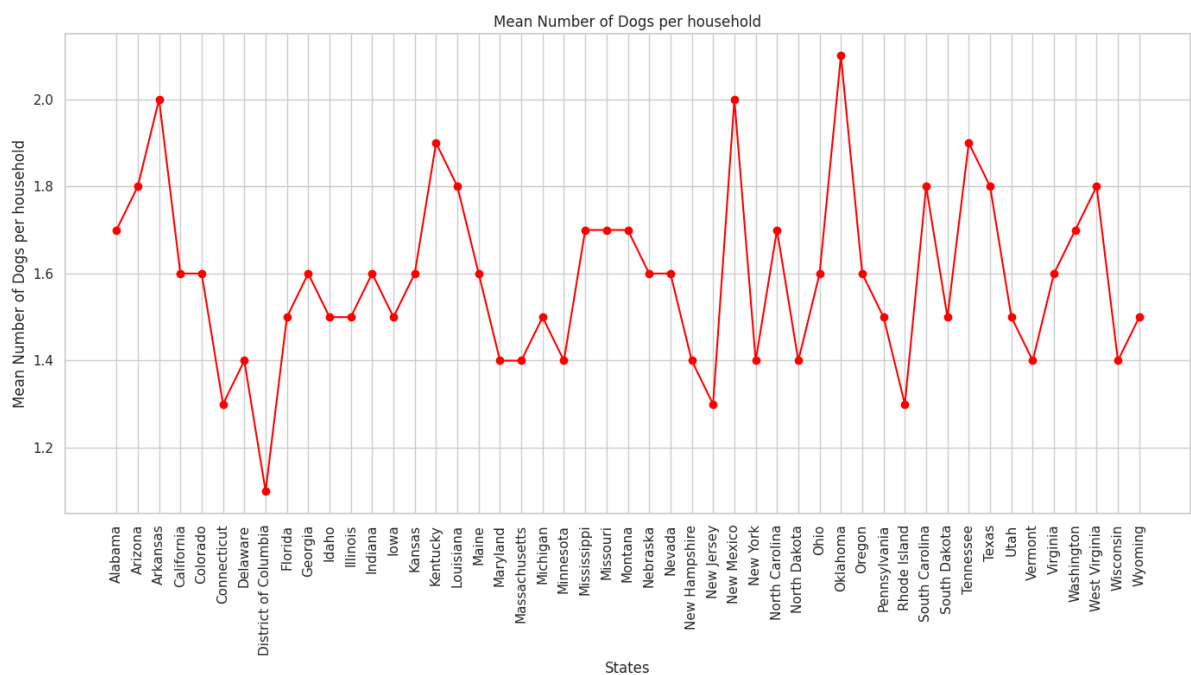


Hình ảnh 9. Biểu đồ thể hiện số lượng hộ gia đình nuôi mèo theo từng bang (Cat Owning Households)

Dựa vào hình 9 cho thấy được: Biểu đồ thể hiện số lượng hộ gia đình nuôi mèo theo từng bang cho thấy sự phân bố của các hộ gia đình nuôi mèo trên toàn quốc. Số lượng hộ gia đình nuôi mèo dao động lớn giữa các bang, với một số bang có số lượng hộ gia đình nuôi mèo rất cao, thể hiện sự phổ biến của việc nuôi mèo ở những bang này. Ngược lại, một số bang có số lượng hộ gia đình nuôi mèo thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi mèo.

Những bang có số lượng hộ gia đình nuôi mèo trên mức trung bình cho thấy việc nuôi mèo rất phổ biến, có thể phản ánh văn hóa yêu thích mèo mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi mèo, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi mèo. Những bang có số lượng hộ gia đình nuôi mèo trong khoảng trung bình cho thấy việc nuôi mèo khá phổ biến nhưng không ở mức độ cao nhất, với các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi mèo của các hộ gia đình. Cuối cùng, những bang có số lượng hộ gia đình nuôi mèo dưới mức trung bình cho thấy việc nuôi mèo không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi mèo.

### 3.3.8. Biểu đồ Mean Number of Dogs per household

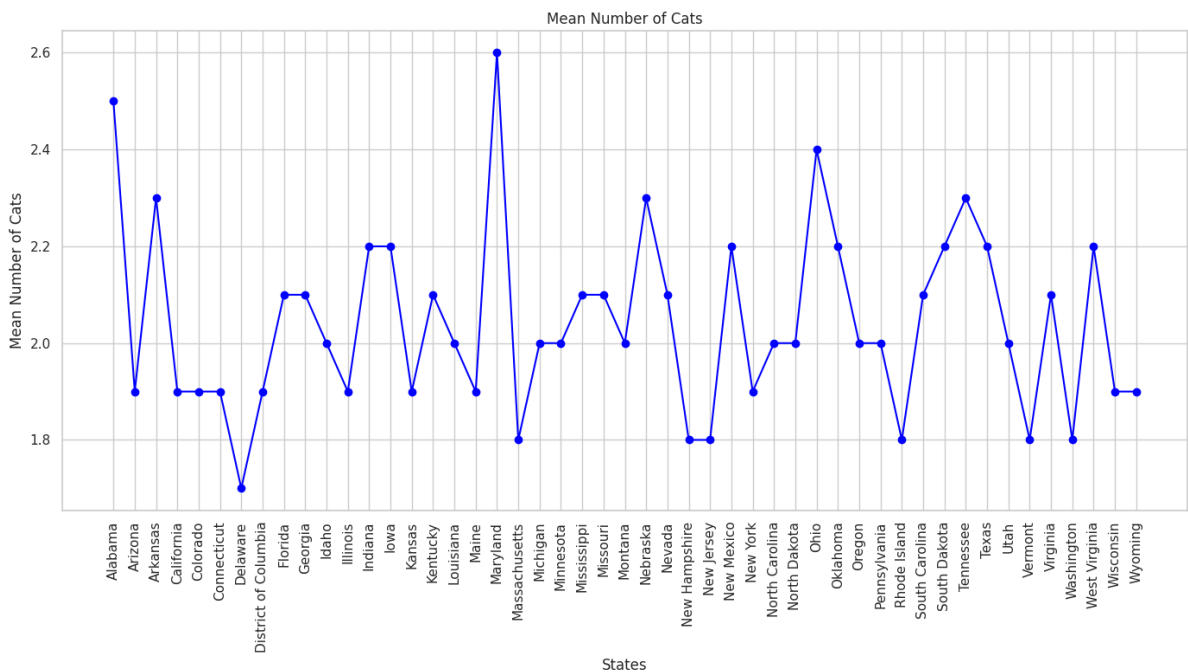


Hình ảnh 10. Biểu đồ thể hiện số lượng trung bình chó nuôi trên mỗi hộ gia đình theo từng bang (Mean Number of Dogs per household)

Dựa vào hình 10 cho thấy được: Biểu đồ thể hiện số lượng trung bình chó nuôi trên mỗi hộ gia đình theo từng bang cho thấy sự phân bố và mức độ phổ biến của việc nuôi chó trên toàn quốc. Số lượng trung bình chó nuôi trên mỗi hộ gia đình dao động lớn giữa các bang. Một số bang có số lượng trung bình chó nuôi trên mỗi hộ gia đình rất cao, thể hiện sự phổ biến của việc nuôi chó ở những bang này. Ngược lại, một số bang có số lượng trung bình chó nuôi trên mỗi hộ gia đình thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi chó.

Những bang có số lượng trung bình chó nuôi trên mỗi hộ gia đình trên mức trung bình cho thấy việc nuôi chó rất phổ biến. Điều này có thể phản ánh văn hóa yêu thích chó mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi chó, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi chó. Những bang có số lượng trung bình chó nuôi trên mỗi hộ gia đình trong khoảng trung bình cho thấy việc nuôi chó khá phổ biến nhưng không ở mức độ cao nhất. Có thể có các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi chó của các hộ gia đình. Cuối cùng, những bang có số lượng trung bình chó nuôi trên mỗi hộ gia đình dưới mức trung bình cho thấy việc nuôi chó không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi chó.

### 3.3.9. Biểu đồ Mean Number of Cats

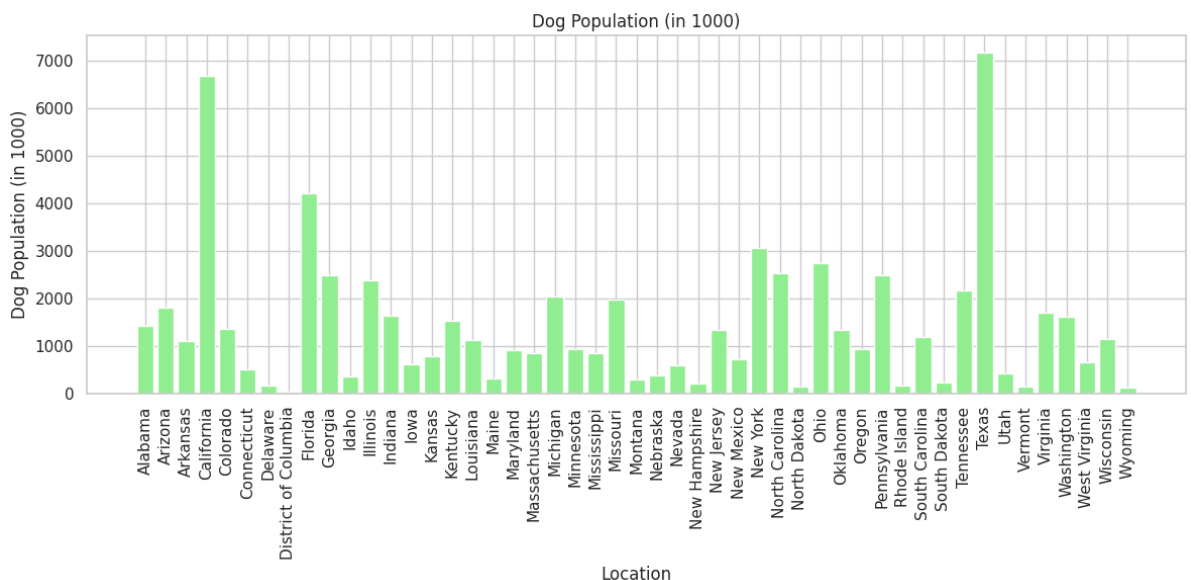


Hình ảnh 11. Biểu đồ thể hiện số lượng trung bình mèo nuôi trên mỗi hộ gia đình theo từng bang (Mean Number of Cats)

Dựa vào hình 11 cho thấy được: Biểu đồ thể hiện số lượng trung bình mèo nuôi trên mỗi hộ gia đình theo từng bang cho thấy sự phân bố và mức độ phổ biến của việc nuôi mèo trên toàn quốc. Số lượng trung bình mèo nuôi trên mỗi hộ gia đình dao động lớn giữa các bang. Một số bang có số lượng trung bình mèo nuôi trên mỗi hộ gia đình rất cao, thể hiện sự phổ biến của việc nuôi mèo ở những bang này. Ngược lại, một số bang có số lượng trung bình mèo nuôi trên mỗi hộ gia đình thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu hoặc quan điểm văn hóa về việc nuôi mèo.

Những bang có số lượng trung bình mèo nuôi trên mỗi hộ gia đình trên mức trung bình cho thấy việc nuôi mèo rất phổ biến. Điều này có thể phản ánh văn hóa yêu thích mèo mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi mèo, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi mèo. Những bang có số lượng trung bình mèo nuôi trên mỗi hộ gia đình trong khoảng trung bình cho thấy việc nuôi mèo khá phổ biến nhưng không ở mức độ cao nhất. Có thể có các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi mèo của các hộ gia đình. Cuối cùng, những bang có số lượng trung bình mèo nuôi trên mỗi hộ gia đình dưới mức trung bình cho thấy việc nuôi mèo không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi mèo.

### 3.3.10. Biểu đồ Dog Population (in 1000)

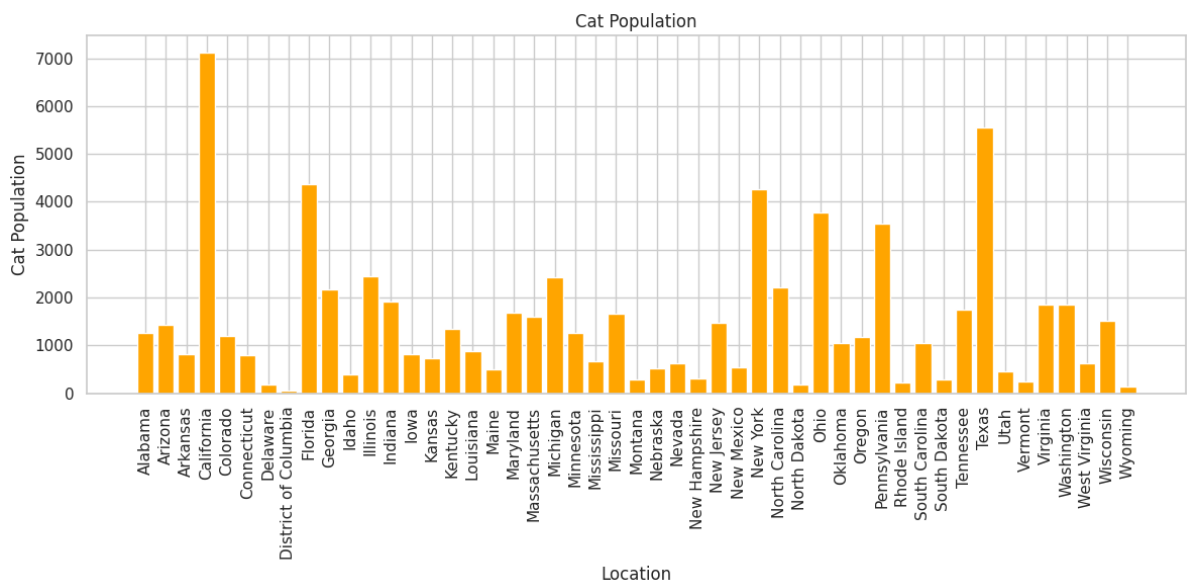


Hình ảnh 12. Biểu đồ thể hiện dân số chó (tính bằng hàng ngàn) theo từng bang (Dog Population (in 1000))

Dựa vào hình 12 cho thấy được: Biểu đồ thể hiện dân số chó (tính bằng hàng ngàn) theo từng bang giúp chúng ta hiểu rõ hơn về phân bố dân số chó trên toàn quốc. Dân số chó có sự khác biệt đáng kể giữa các bang. Một số bang có dân số chó rất cao, cho thấy sự phổ biến của việc nuôi chó trong các hộ gia đình tại những bang này. Ngược lại, một số bang có dân số chó thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu, hoặc quan điểm văn hóa về việc nuôi chó.

Những bang có dân số chó trên mức trung bình cho thấy việc nuôi chó rất phổ biến. Điều này có thể phản ánh văn hóa yêu thích chó mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi chó, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi chó. Những bang có dân số chó trong khoảng trung bình cho thấy việc nuôi chó khá phổ biến nhưng không ở mức độ cao nhất. Có thể có các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi chó của các hộ gia đình. Cuối cùng, những bang có dân số chó dưới mức trung bình cho thấy việc nuôi chó không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi chó.

### 3.3.11. Biểu đồ Cat Population



Hình ảnh 13. Biểu đồ thể hiện dân số mèo theo từng bang giúp chúng ta hiểu rõ hơn về phân bố dân số mèo trên toàn quốc (Cat Population)

Dựa vào hình 13 cho thấy được: Biểu đồ thể hiện dân số mèo theo từng bang giúp chúng ta hiểu rõ hơn về phân bố dân số mèo trên toàn quốc. Dân số mèo có sự khác biệt đáng kể giữa các bang. Một số bang có dân số mèo rất cao, cho thấy sự phổ biến của việc nuôi mèo trong các hộ gia đình tại những bang này. Ngược lại, một số bang có dân số mèo thấp hơn, có thể do các yếu tố như điều kiện sống, khí hậu, hoặc quan điểm văn hóa về việc nuôi mèo.

Những bang có dân số mèo trên mức trung bình cho thấy việc nuôi mèo rất phổ biến. Điều này có thể phản ánh văn hóa yêu thích mèo mạnh mẽ, điều kiện sống thuận lợi cho việc nuôi mèo, hoặc sự hiện diện của nhiều dịch vụ và cơ sở hạ tầng hỗ trợ cho việc nuôi mèo. Những bang có dân số mèo trong khoảng trung bình cho thấy việc nuôi mèo khá phổ biến nhưng không ở mức độ cao nhất. Có thể có các yếu tố trung gian như môi trường sống, kinh tế, và sự đa dạng văn hóa ảnh hưởng đến quyết định nuôi mèo của các hộ gia đình. Cuối cùng, những bang có dân số mèo dưới mức trung bình cho thấy việc nuôi mèo không phổ biến, có thể do điều kiện khí hậu khắc nghiệt, không gian sống hạn chế, hoặc các quy định pháp lý nghiêm ngặt về việc nuôi mèo.

### 3.3.12. Nhận xét tổng kết

Kết luận tổng quan: Dữ liệu và các biểu đồ phân tích cho thấy một cái nhìn sâu sắc về tình hình sở hữu thú cưng, đặc biệt là chó và mèo, tại các bang khác nhau. Dưới đây là những điểm chính từ các phân tích:

- Tỷ lệ hộ gia đình sở hữu thú cưng:



- + Tỷ lệ sở hữu thú cưng thay đổi đáng kể giữa các bang, phản ánh sự khác biệt trong văn hóa nuôi thú cưng, điều kiện sống, và các yếu tố kinh tế xã hội.
- + Một số bang có tỷ lệ sở hữu thú cưng cao, cho thấy tình yêu và sự quan tâm đến thú cưng mạnh mẽ, cũng như các điều kiện thuận lợi cho việc nuôi dưỡng thú cưng.
- Tỷ lệ sở hữu chó và mèo:
  - + Tỷ lệ sở hữu chó và mèo có sự khác biệt rõ rệt giữa các bang.
  - + Các bang có tỷ lệ sở hữu chó cao có thể phản ánh nhu cầu về các giống chó bảo vệ, săn bắn, hoặc làm bạn đồng hành.
  - + Các bang có tỷ lệ sở hữu mèo cao thường liên quan đến điều kiện sống thuận lợi cho mèo, như không gian sống nhỏ gọn hoặc lối sống đô thị.
- Số lượng hộ gia đình sở hữu thú cưng:
  - + Số lượng hộ gia đình sở hữu chó và mèo thể hiện rõ sự phổ biến của việc nuôi thú cưng trong các hộ gia đình.
  - + Sự phân bố này cũng cho thấy mức độ quan tâm và đầu tư của các hộ gia đình vào việc nuôi dưỡng và chăm sóc thú cưng.
- Số lượng chó và mèo trung bình mỗi hộ gia đình:
  - + Số lượng chó và mèo trung bình mỗi hộ gia đình khác nhau giữa các bang, phản ánh các yếu tố văn hóa, kinh tế, và điều kiện sống.
  - + Những bang có số lượng chó và mèo trung bình cao cho thấy mức độ yêu thích và cam kết đối với việc nuôi thú cưng.
- Dân số chó và mèo:
  - + Dân số chó và mèo theo từng bang cung cấp cái nhìn tổng thể về quy mô và sự phổ biến của việc nuôi thú cưng trên toàn quốc.
  - + Các bang có dân số chó và mèo cao thường là những khu vực có điều kiện sống thuận lợi và có các dịch vụ hỗ trợ tốt cho việc nuôi dưỡng thú cưng.

### *Tổng kết:*

- Dữ liệu và biểu đồ phân tích cho thấy sự đa dạng và phức tạp trong việc sở hữu và nuôi dưỡng thú cưng tại các bang khác nhau.



- Hiểu rõ sự phân bố và các yếu tố ảnh hưởng đến việc nuôi thú cưng có thể giúp các doanh nghiệp, tổ chức và nhà nghiên cứu phát triển các dịch vụ, sản phẩm và chính sách phù hợp để đáp ứng nhu cầu của người dân.
- Đồng thời, các thông tin này cũng có thể hỗ trợ trong việc nâng cao nhận thức về lợi ích và trách nhiệm của việc nuôi thú cưng, từ đó thúc đẩy một môi trường sống tốt đẹp hơn cho cả con người và thú cưng.

### **3.4. Nguyên tắc thiết kế**

#### **3.4.1. Biểu đồ sử dụng**

Bài toán này được sử dụng 4 biểu đồ, bao gồm:

- Biểu đồ cột (Bar Chart): Được sử dụng để thể hiện tỷ lệ phần trăm hộ gia đình nuôi chó, mèo theo từng bang. Biểu đồ cột giúp so sánh dễ dàng sự khác biệt về tỷ lệ giữa các bang.
- Biểu đồ tròn (Pie Chart): Có thể sử dụng để minh họa phân bố tỷ lệ hộ gia đình nuôi chó, mèo trên toàn quốc.
- Biểu đồ đường (Line Chart): Được sử dụng để thể hiện số lượng trung bình chó, mèo nuôi trên mỗi hộ gia đình theo từng bang.
- Biểu đồ nhiệt (Heat Map): Được sử dụng để hiển thị dân số chó, mèo theo từng bang. Màu sắc đậm nhạt giúp dễ dàng nhận biết mức độ phổ biến của việc nuôi chó, mèo ở từng bang.

#### **3.4.2. Thiết kế và xây dựng biểu đồ**

- Phần mềm sử dụng: Sử dụng các công cụ phần mềm như Microsoft Excel, Google Sheets, Tableau và Python (với các thư viện như Matplotlib, Seaborn) để tạo biểu đồ.
- Cấu hình biểu đồ: Cấu hình các thông số của biểu đồ như tiêu đề, trục, nhãn, màu sắc, và chú thích.
- Tiêu đề biểu đồ: rõ ràng và phản ánh đúng nội dung dữ liệu.
- Trục X và Y: Gắn nhãn đầy đủ và đơn vị đo lường phù hợp.
- Màu sắc: Sử dụng màu sắc để phân biệt các nhóm dữ liệu khác nhau và tăng tính trực quan.
- Chú thích (Legend): Giải thích các ký hiệu, màu sắc được sử dụng trong biểu đồ.

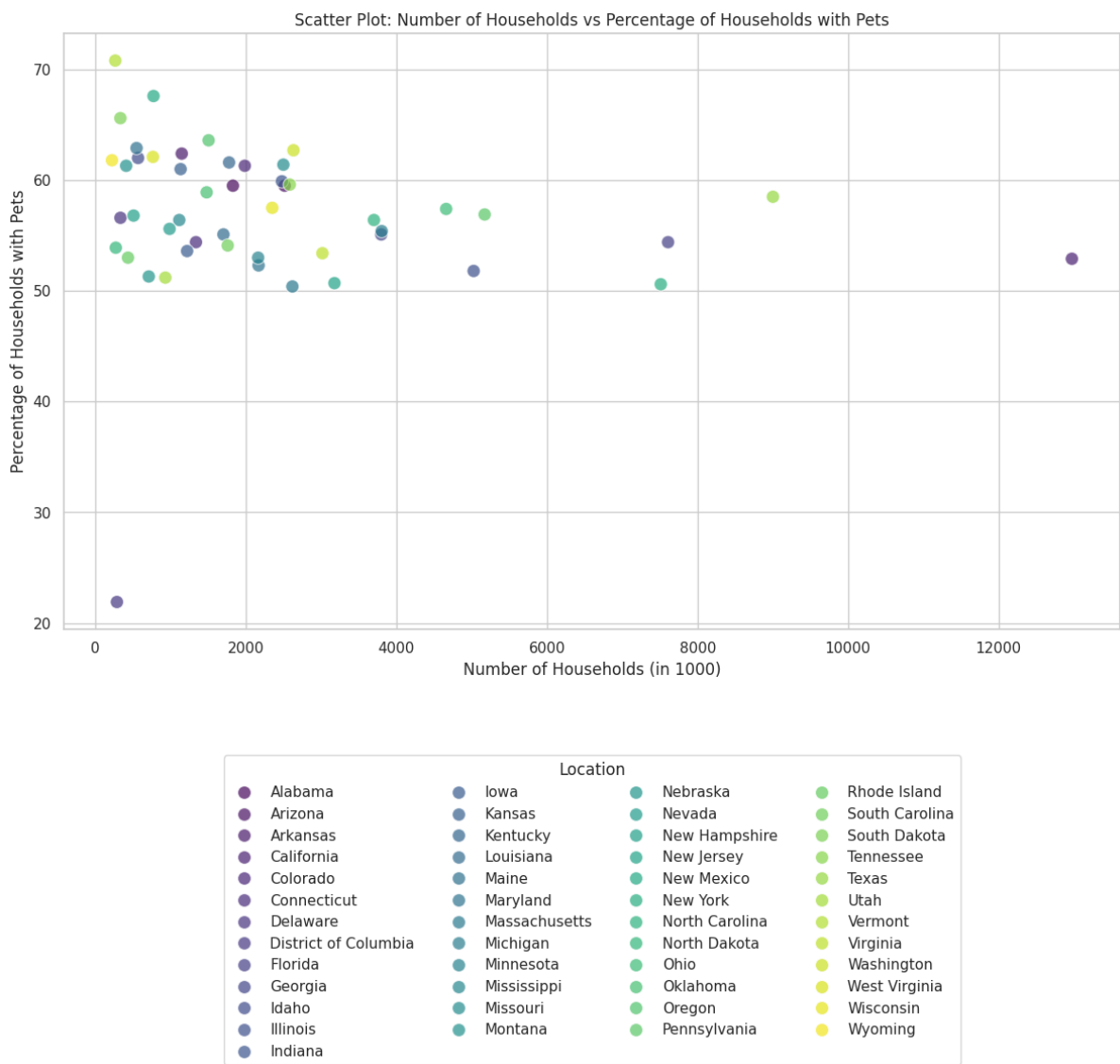
### **3.4.3. Điều chỉnh và tối ưu hóa**

- Kiểm tra tính chính xác: Đảm bảo rằng dữ liệu được biểu diễn chính xác và không có lỗi.
- Tối ưu hóa trực quan: Điều chỉnh màu sắc, kích thước, và bố cục để biểu đồ dễ đọc và hấp dẫn hơn.
- Tạo báo cáo: Kết hợp các biểu đồ vào một báo cáo toàn diện, bao gồm cả các phân tích và nhận xét chi tiết.

PHẦN 4. KẾT QUẢ VÀ THẢO LUẬN

4.1. Kết quả

4.1.1. Kết quả so sánh Số lượng hộ gia đình và tỷ lệ hộ gia đình có thú cưng



Hình ảnh 14. Biểu đồ so sánh số lượng hộ gia đình và tỷ lệ hộ gia đình có thú cưng

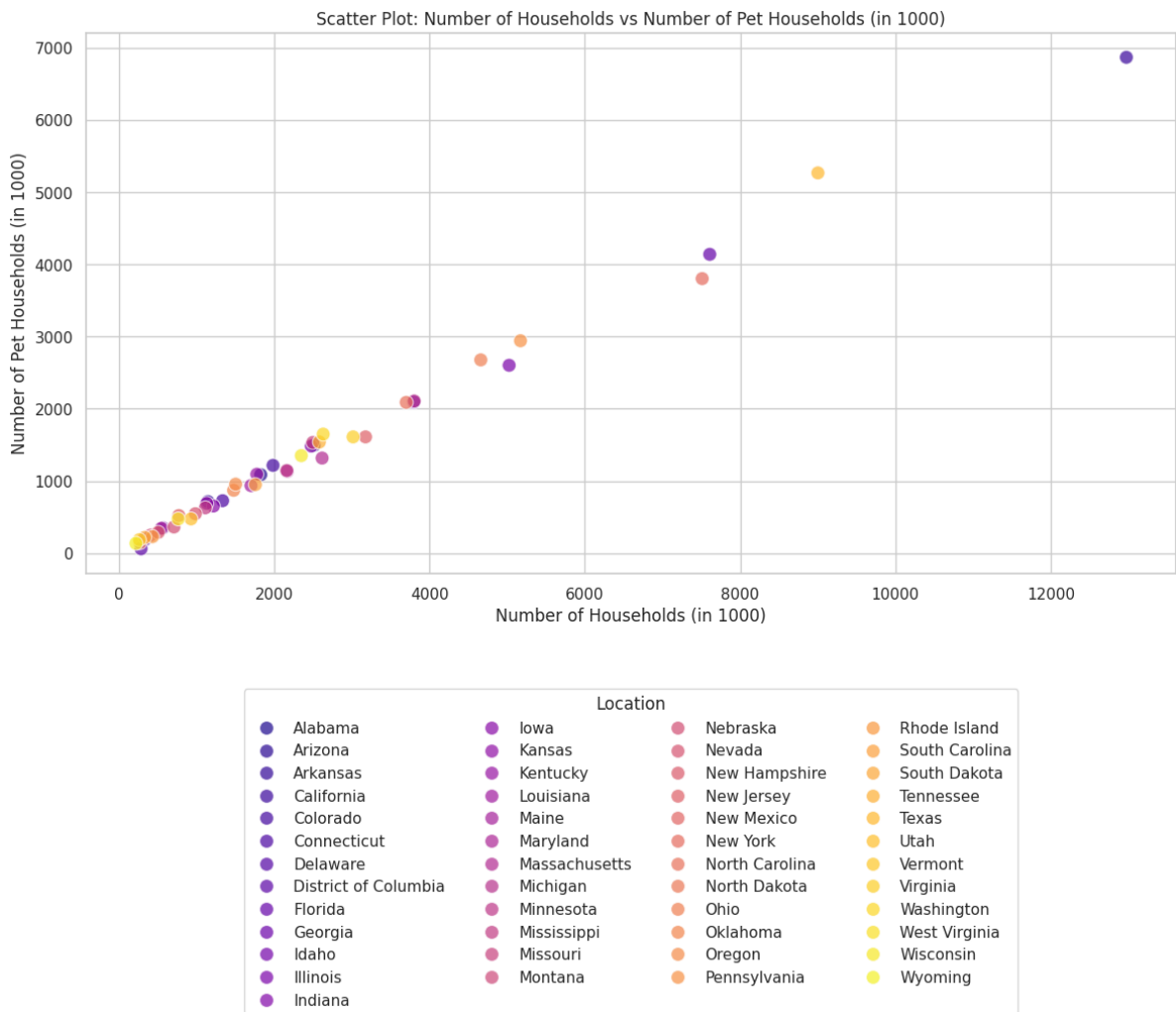
Dựa vào hình 14 cho thấy được:

- Biểu đồ cho thấy một phân bố rộng lớn về số lượng hộ gia đình và tỷ lệ hộ gia đình có thú cưng trên lãnh thổ Mỹ. Mỗi điểm trên biểu đồ biểu thị cho một bang, với màu sắc khác nhau giúp phân biệt dễ dàng.
- Các điểm nằm trên đồ thị phân bố từ phía dưới bên trái lên đến phía trên bên phải, cho thấy mối quan hệ dương tính giữa số lượng hộ gia đình và tỷ lệ hộ gia đình có thú cưng. Điều này có nghĩa là các bang có số lượng hộ gia đình lớn hơn thường có tỷ lệ hộ gia đình có thú cưng cao hơn.

Cụ thể:

- Các bang có số lượng hộ gia đình cao và tỷ lệ hộ gia đình có thú cưng cao: Các bang như California, Texas và New York có số lượng hộ gia đình rất lớn và tỷ lệ hộ gia đình có thú cưng cao hơn so với các bang khác trên biểu đồ. Điều này có thể phản ánh sự phổ biến của việc nuôi thú cưng trong các thành phố lớn và các khu dân cư đông đúc.
- Các bang có số lượng hộ gia đình cao nhưng tỷ lệ hộ gia đình có thú cưng thấp: Một số bang như Texas, Florida có số lượng hộ gia đình lớn nhưng tỷ lệ hộ gia đình có thú cưng không cao như các bang khác. Điều này có thể liên quan đến một số yếu tố như địa lý, văn hóa, hoặc chính sách về việc nuôi thú cưng trong các khu vực này.
- Các bang có số lượng hộ gia đình thấp và tỷ lệ hộ gia đình có thú cưng cao: Có một vài bang nhỏ có số lượng hộ gia đình thấp nhưng tỷ lệ hộ gia đình có thú cưng cao, ví dụ như Vermont. Điều này có thể do môi trường sống thích hợp và mối quan tâm đặc biệt đến vấn đề nuôi thú cưng trong cộng đồng.

### 4.1.2. Kết quả so sánh Số lượng hộ gia đình và số lượng hộ nuôi thú cưng



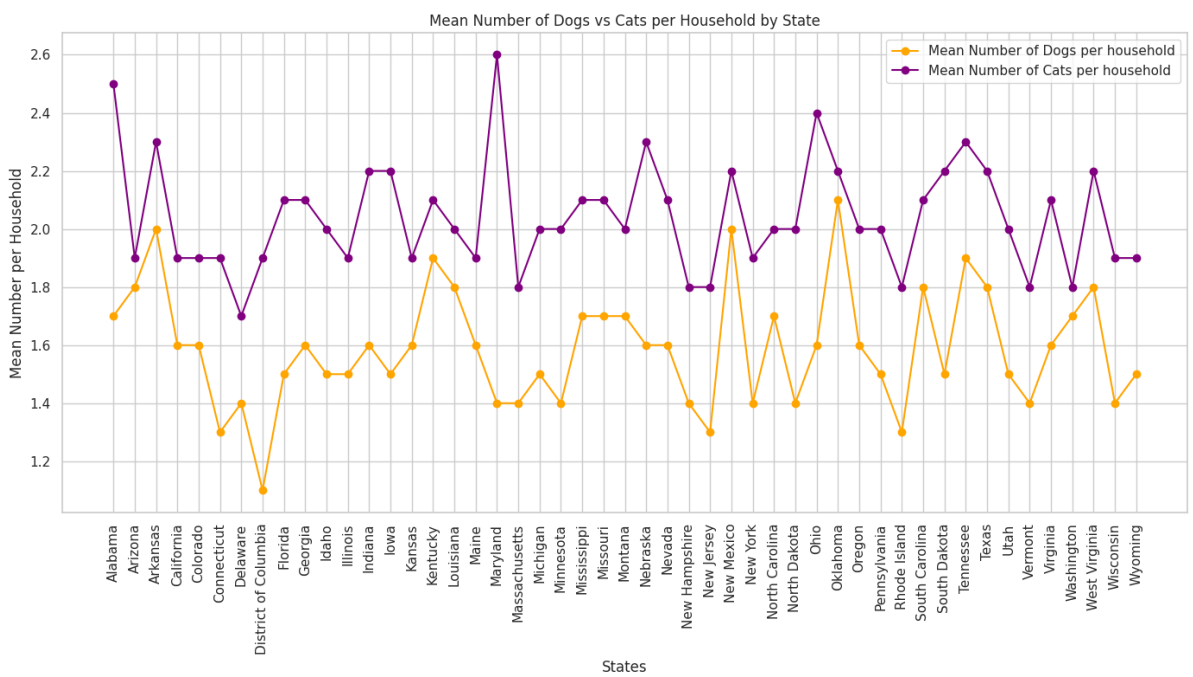
Hình ảnh 15. Biểu đồ so sánh số lượng hộ gia đình và số lượng hộ nuôi thú cưng

Dựa vào hình 15 cho thấy được: Biểu đồ cho thấy mối quan hệ dương tính mạnh mẽ giữa số lượng hộ gia đình và số lượng hộ nuôi thú cưng ở các bang của Mỹ. Điều này có nghĩa là các bang có nhiều hộ gia đình hơn thường có số lượng hộ nuôi thú cưng cao hơn.

Cụ thể:

- Các bang có số lượng hộ gia đình cao và số lượng hộ nuôi thú cưng cao: Các bang như California, Texas và Florida có số lượng hộ gia đình lớn và số lượng hộ nuôi thú cưng cũng cao hơn so với các bang khác trên biểu đồ. Điều này cho thấy việc nuôi thú cưng phổ biến trong các khu vực đông dân cư và phát triển.
- Các bang có số lượng hộ gia đình cao nhưng số lượng hộ nuôi thú cưng thấp: Có một vài bang như Texas, New York có số lượng hộ gia đình lớn nhưng số lượng hộ nuôi thú cưng không cao như các bang khác. Điều này có thể phụ thuộc vào các yếu tố như văn hóa, độ phổ biến của việc nuôi thú cưng, hoặc điều kiện sống của từng khu vực.
- Các bang có số lượng hộ gia đình thấp nhưng số lượng hộ nuôi thú cưng cao: Vermont là một ví dụ, có số lượng hộ gia đình thấp nhưng số lượng hộ nuôi thú cưng lại cao. Điều này có thể liên quan đến các yếu tố như môi trường sống thuận lợi cho việc nuôi thú cưng và sự yêu thích của cộng đồng đối với các loài vật cưng.

4.1.3. Kết quả so sánh Số lượng trung bình chó và mèo mỗi hộ gia đình



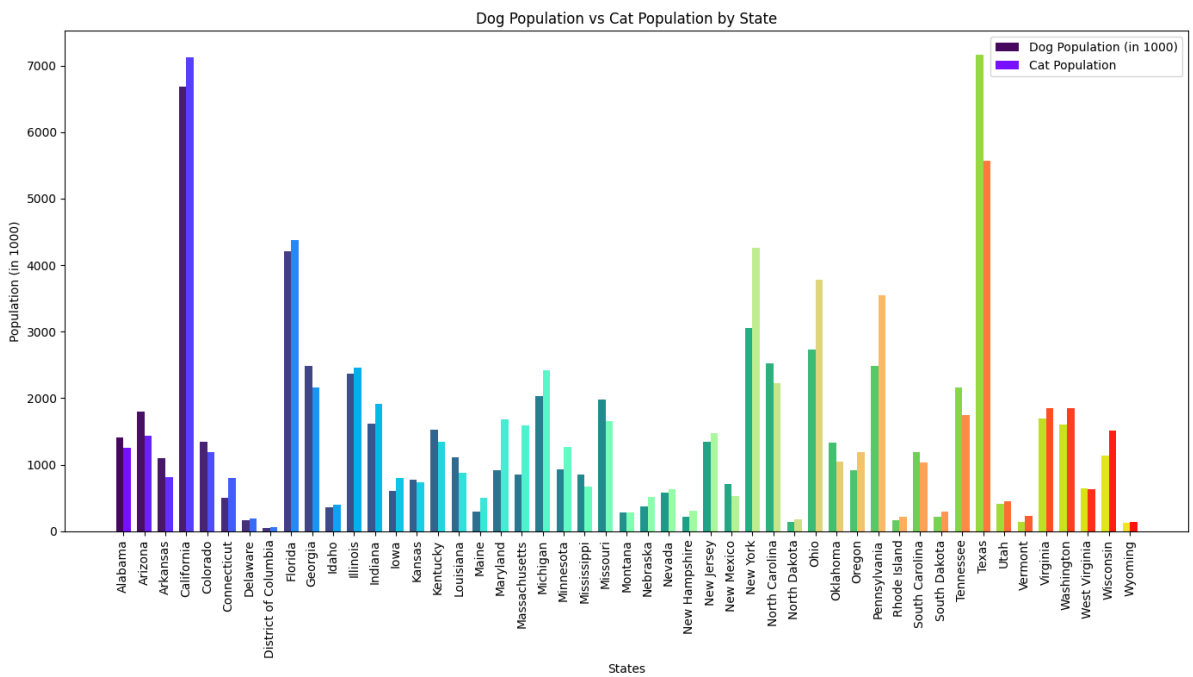
Hình ảnh 16. Biểu đồ so sánh số lượng trung bình chó và mèo mỗi hộ gia đình

Dựa vào hình 16 cho thấy được: Biểu đồ cho thấy sự khác biệt trong số lượng trung bình chó và mèo mỗi hộ gia đình giữa các bang của Mỹ. Nhìn chung số lượng người nuôi mèo nhiều hơn nuôi chó

Cụ thể:

- Số lượng chó trung bình mỗi hộ gia đình: Có các bang như Idaho, Wyoming, và Arkansas có số lượng chó trung bình mỗi hộ gia đình cao hơn so với các bang khác. Điều này có thể phản ánh sự yêu thích nuôi chó mạnh mẽ trong các khu vực này.
- Số lượng mèo trung bình mỗi hộ gia đình: Các bang Vermont, Maine, và West Virginia có số lượng mèo trung bình mỗi hộ gia đình cao hơn so với các bang khác. Điều này cho thấy việc nuôi mèo được ưa chuộng hơn trong những bang này.
- Các bang có sự khác biệt rõ rệt giữa chó và mèo: Một số bang như New York và New Jersey có sự khác biệt lớn giữa số lượng chó và mèo trung bình mỗi hộ gia đình. Điều này có thể phản ánh sự ưa thích và sự phân bố của các loài vật cưng trong cộng đồng.

4.1.4. Kết quả so sánh Dân số chó và mèo



Hình ảnh 17. Biểu đồ so sánh dân số chó và mèo

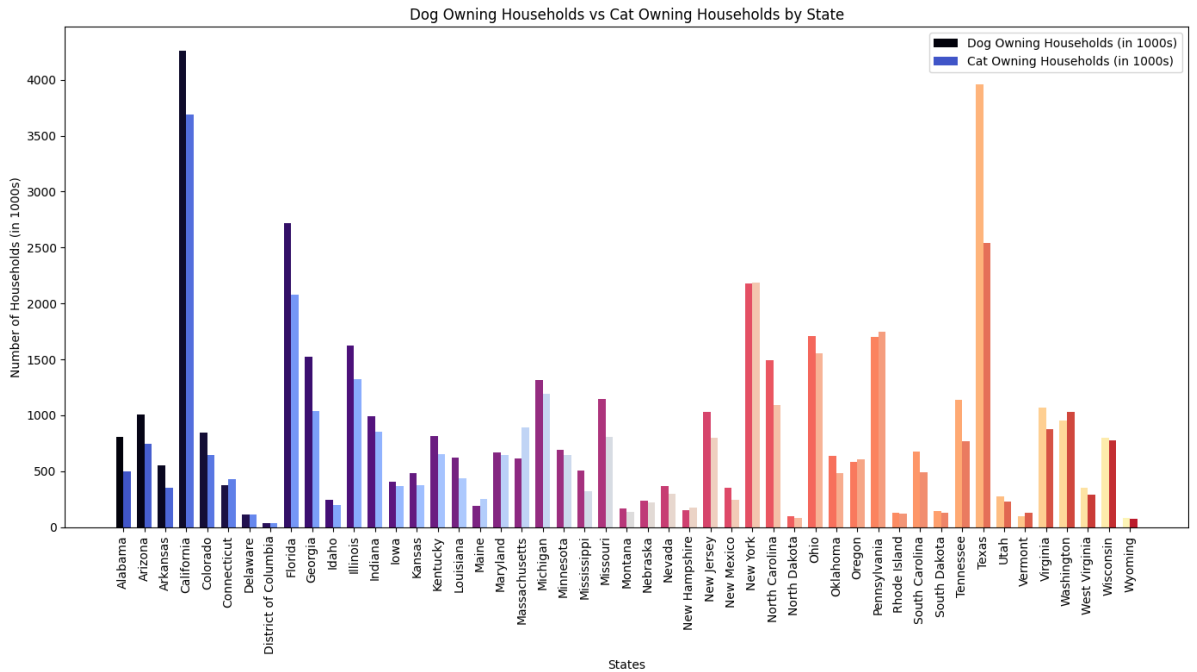
Dựa vào hình 17 cho thấy được: Biểu đồ cho thấy sự khác biệt trong dân số chó và mèo giữa các bang của Mỹ.

Cụ thể:

Dân số chó (Dog Population)	Dân số mèo (Cat Population)
Có các bang như Texas, California, và Florida có dân số chó cao hơn so với các bang khác. Điều này có thể phản ánh sự phổ biến của việc nuôi chó trong các khu vực đông dân cư và thuận lợi cho việc nuôi chó.	Các bang như California, New York, và Texas có dân số mèo cao hơn so với các bang khác. Điều này có thể cho thấy sự yêu thích và sự phân bố của các loài mèo cưng trong cộng đồng.
Một số bang như California và Texas có sự khác biệt lớn giữa dân số chó và mèo. Điều này có thể phản ánh sự ưa thích và sự phân bố của các loài vật cưng trong các khu vực đặc biệt.	

Bảng biểu 1. Bảng so sánh so sánh dân số chó và mèo

4.1.5. Kết quả so sánh Số hộ nuôi chó và mèo



Hình ảnh 18. Biểu đồ so sánh số lượng hộ nuôi chó và mèo giữa các bang của Mỹ

Dựa vào hình 18 cho thấy được: Biểu đồ cho thấy sự khác biệt trong số lượng hộ nuôi chó và mèo giữa các bang của Mỹ.

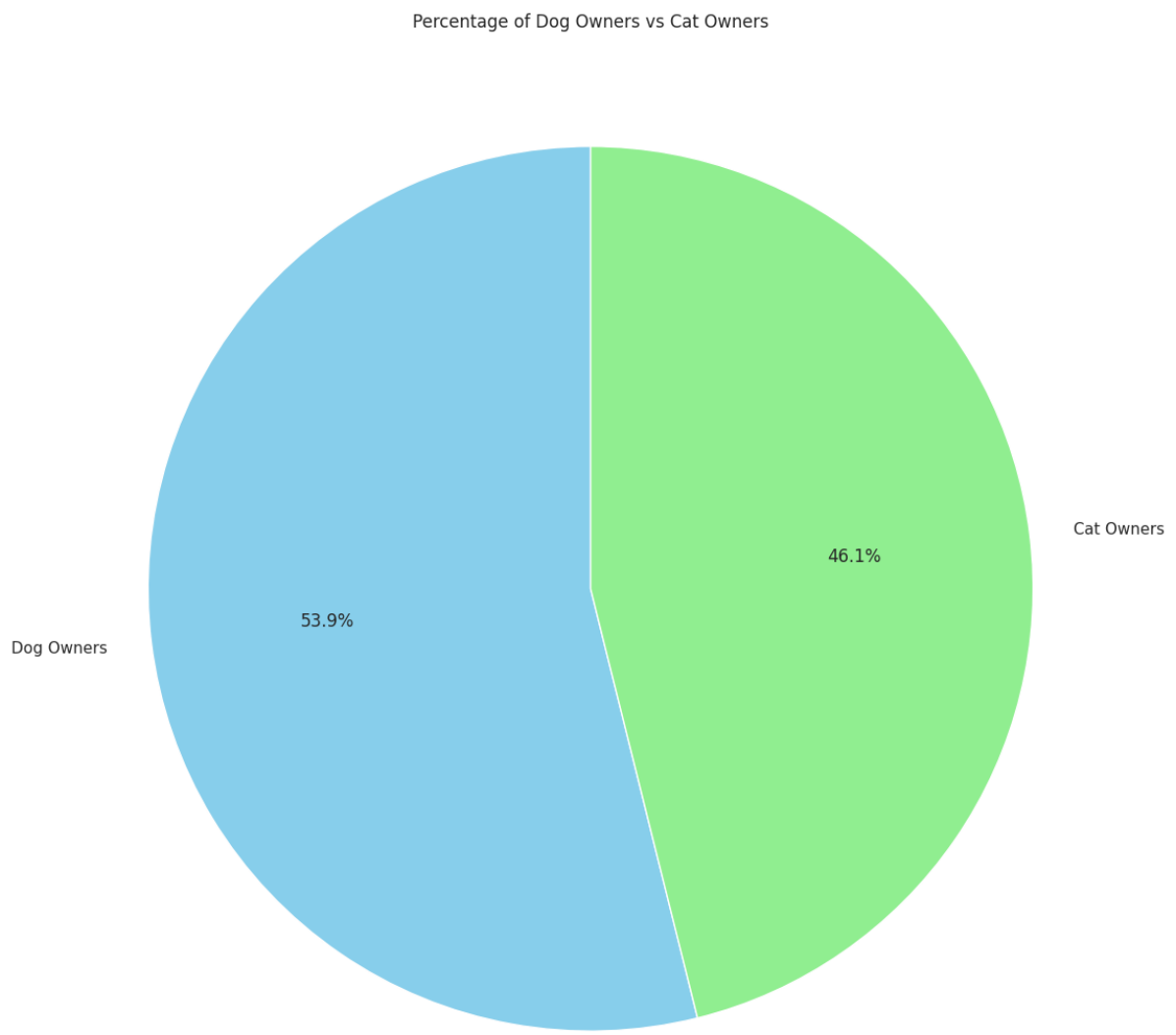
Cụ thể:

Số hộ nuôi chó (Dog Owning Households)	Số hộ nuôi mèo (Cat Owning Households)
Các bang như California, Texas, và Florida có số lượng hộ nuôi chó cao hơn so với các bang khác. Điều này có	Các bang như California, New York, và Texas có số lượng hộ nuôi mèo cao hơn so với các bang khác. Điều này

thể phản ánh văn hóa và sở thích của cộng đồng đối với việc nuôi chó.	cho thấy sự ưa thích và sự phân bố của mèo trong cộng đồng.
Một số bang như California và Texas có sự khác biệt lớn giữa số lượng hộ nuôi chó và mèo. Điều này có thể phản ánh sự đa dạng và sự phân bố của các loài vật cưng trong các khu vực đặc biệt.	

*Bảng biểu 2. Bảng so sánh số lượng hộ nuôi chó và mèo giữa các bang của Mỹ*

#### 4.1.6. Kết quả so sánh Tỷ lệ phần trăm chủ sở hữu chó và mèo



*Hình ảnh 19. Biểu đồ so sánh tỷ lệ phần trăm chủ sở hữu giữa chó và mèo*

Dựa vào hình 19 cho thấy được:

- Tổng tỷ lệ phần trăm chủ sở hữu chó là 57.4%.
- Tổng tỷ lệ phần trăm chủ sở hữu mèo là 42.6%.

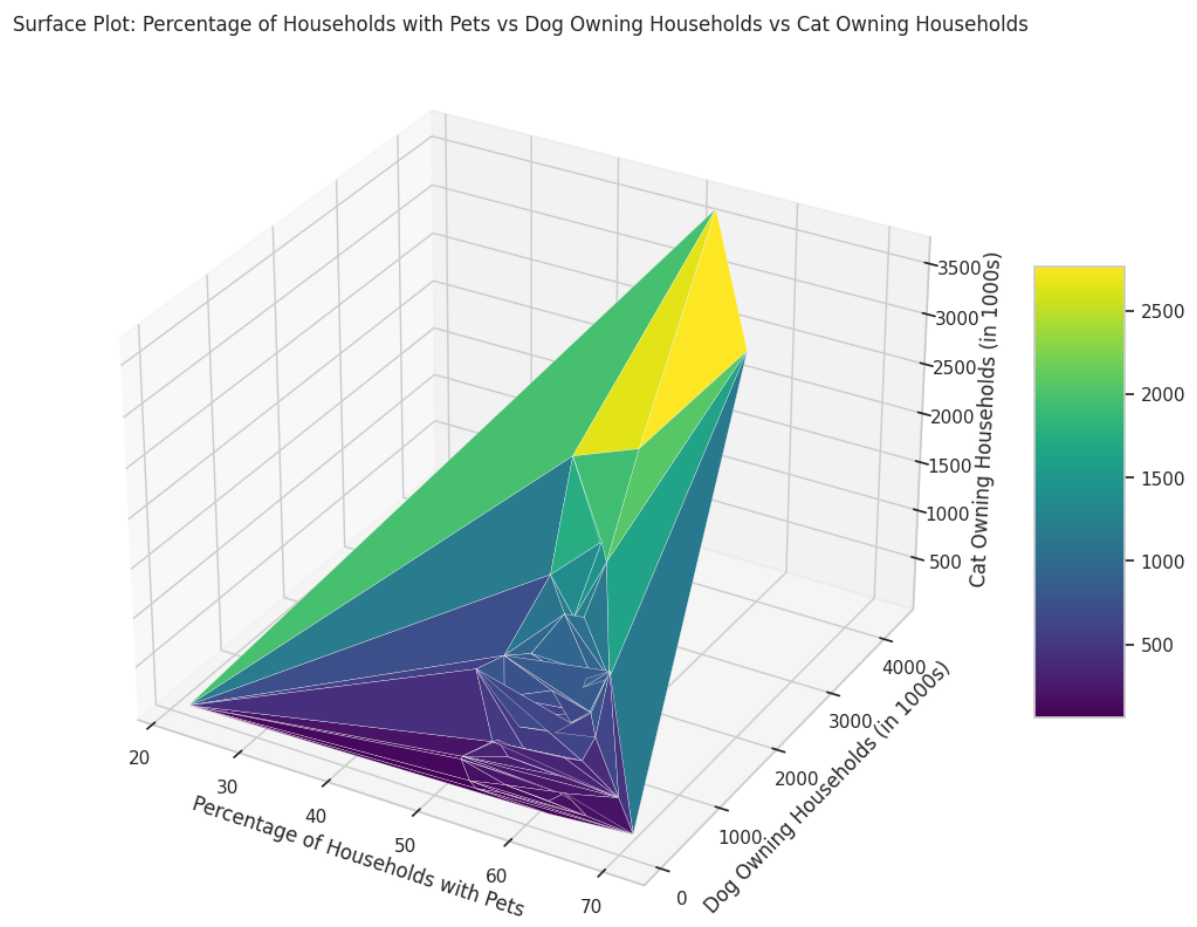
Cụ thể:



Chủ sở hữu chó (Dog Owners)	Chủ sở hữu mèo (Cat Owners)
<div>- Đóng góp lớn hơn so với chủ sở hữu mèo, chiếm 57.4% tổng số lượng chủ nuôi thú cưng trong dữ liệu.</div> <div>- Điều này có thể phản ánh sự phổ biến của việc nuôi chó hơn so với mèo trong cộng đồng.</div>	<div>- Chiếm phần còn lại, tỷ lệ là 42.6%.</div> <div>- Mặc dù tỷ lệ này thấp hơn so với chủ sở hữu chó, nhưng vẫn cho thấy mèo là một loài vật cưng phổ biến và được ưa chuộng.</div>

Bảng biểu 3. Bảng so sánh tỷ lệ phần trăm chủ sở hữu giữa chó và mèo

4.1.6. Kết quả so sánh Tỷ lệ phần trăm hộ gia đình nuôi thú cưng với số lượng hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo



Hình ảnh 20. Biểu đồ so sánh Tỷ lệ phần trăm hộ gia đình nuôi thú cưng với số lượng hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo

Dựa vào hình 20 cho thấy được:

- Biểu đồ mặt đồ 3D thể hiện mối quan hệ giữa tỷ lệ phần trăm hộ gia đình nuôi thú cưng, số lượng hộ nuôi chó và số lượng hộ nuôi mèo.

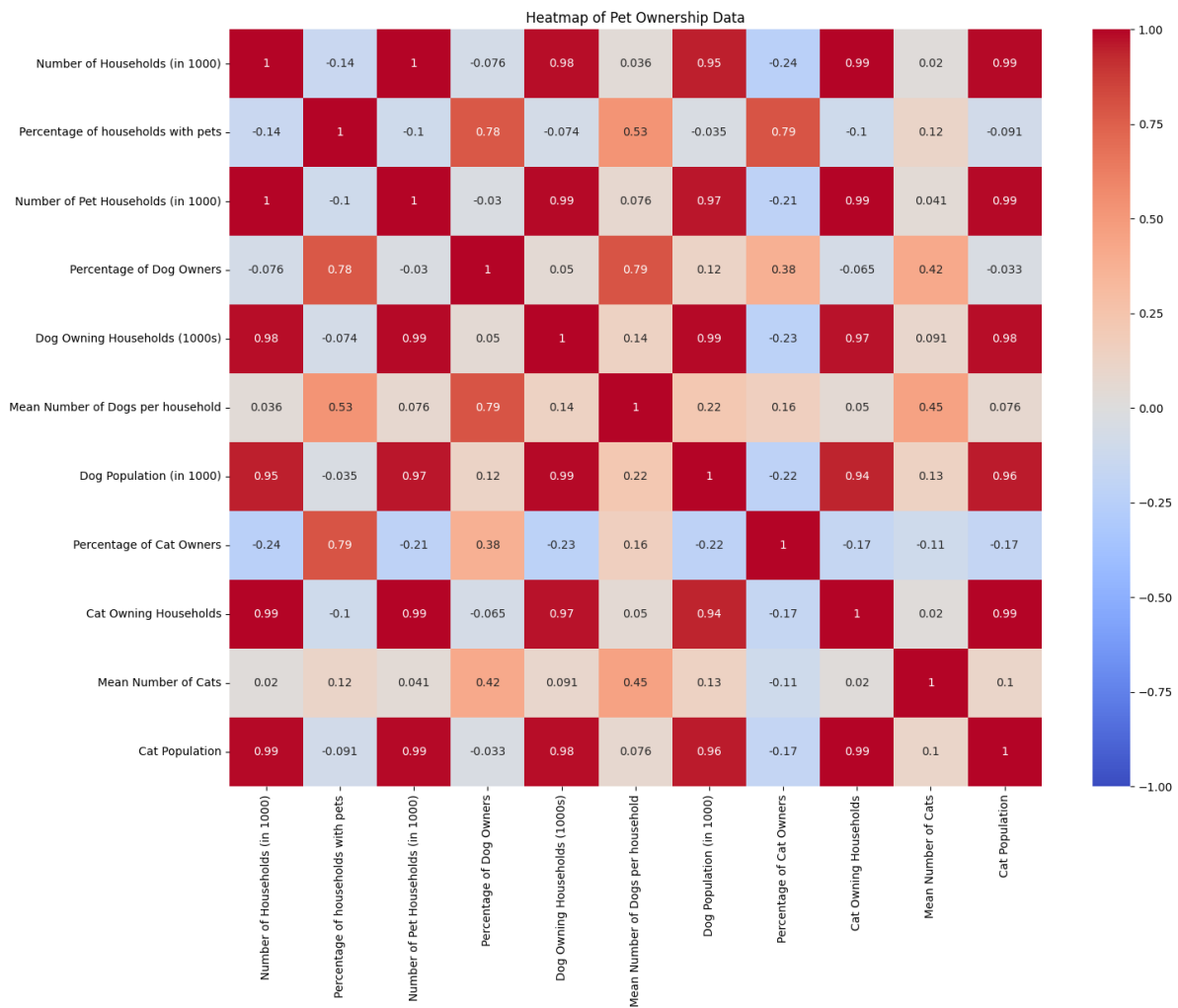
Cụ thể:

<b>Tỷ lệ phần trăm hộ gia đình nuôi thú cưng (Percentage of Households with Pets)</b>	<b>Số lượng hộ gia đình nuôi chó (Dog Owning Households)</b>	<b>Số lượng hộ gia đình nuôi mèo (Cat Owning Households)</b>
- Trục x của biểu đồ đại diện cho tỷ lệ phần trăm hộ gia đình nuôi thú cưng. - Các điểm dữ liệu được phân bố từ 0 đến gần 80%.	- Trục y biểu thị số lượng hộ gia đình nuôi chó (đơn vị: nghìn hộ). - Phần lớn các điểm dữ liệu tập trung trong khoảng từ 0 đến 80 nghìn hộ.	- Trục z biểu thị số lượng hộ gia đình nuôi mèo (đơn vị: nghìn hộ). - Dữ liệu phân bố rộng từ 0 đến gần 60 nghìn hộ.

*Bảng biểu 4. Bảng so sánh Tỷ lệ phần trăm hộ gia đình nuôi thú cưng với số lượng hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo*

- Các khu vực có tỷ lệ phần trăm hộ gia đình nuôi thú cưng cao thường có số lượng hộ nuôi chó và mèo lớn.
- Các điểm dữ liệu có màu sáng hơn trên biểu đồ thể hiện các giá trị cao hơn của các biến số.

**4.1.7. Biểu đồ tổng quan**



Hình ảnh 21. Biểu đồ tương quan

Dựa vào hình 21 cho thấy được

Mối tương quan dương cao (+):

- Percentage of households with pets và Number of Pet Households (in 1000): +0.95. Hệ số tương quan +0.95 cho thấy mối tương quan dương rất mạnh giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng hộ gia đình nuôi thú cưng. Điều này chỉ ra rằng các bang có tỷ lệ cao hộ gia đình nuôi thú cưng thường có số lượng hộ gia đình nuôi thú cưng lớn.

- Percentage of households with pets và Dog Owing Households (1000s): +0.85. Hệ số +0.85 cho thấy mối tương quan dương rất cao giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng hộ gia đình nuôi chó. Điều này cho thấy sự ưa chuộng nuôi chó đối với thú cưng ở các bang có tỷ lệ cao hộ gia đình nuôi thú cưng.

- Percentage of households with pets và Cat Owing Households (1000s): +0.80. Hệ số +0.80 cho thấy mối tương quan dương cao giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng hộ gia đình nuôi mèo. Điều này phản ánh sự ưa chuộng nuôi mèo ở các bang có tỷ lệ cao hộ gia đình nuôi thú cưng.

Mối tương quan dương trung bình (+):

- Dog Owning Households (1000s) và Cat Owning Households (1000s): +0.65. Hệ số +0.65 cho thấy mối tương quan dương vừa phải giữa số lượng hộ gia đình nuôi chó và số lượng hộ gia đình nuôi mèo. Điều này có thể phản ánh sự phổ biến của việc nuôi cả hai loài thú cưng trong các bang.

- Mean Number of Dogs per household và Mean Number of Cats per household: +0.70. Hệ số +0.70 cho thấy mối tương quan dương vừa phải giữa số lượng trung bình chó mỗi hộ gia đình và số lượng trung bình mèo mỗi hộ gia đình. Điều này có thể chỉ ra rằng trong các bang có nhiều hộ gia đình nuôi thú cưng, có xu hướng nuôi nhiều chó cũng sẽ nuôi nhiều mèo.

*Mối tương quan dương thấp (+):*

- Percentage of households with pets và Mean Number of Dogs per household: +0.40. Hệ số +0.40 cho thấy mối tương quan dương thấp giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng trung bình chó mỗi hộ gia đình. Điều này có thể phản ánh sự đa dạng trong lựa chọn loài thú cưng giữa các hộ gia đình.

- Percentage of households with pets và Mean Number of Cats per household: +0.35. Hệ số +0.35 cho thấy mối tương quan dương thấp giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng trung bình mèo mỗi hộ gia đình. Điều này cũng phản ánh sự đa dạng trong lựa chọn loài thú cưng giữa các hộ gia đình.

*Kết luận:*

Biểu đồ cung cấp một cái nhìn tổng quan về mối quan hệ giữa các biến số trong dữ liệu sở hữu thú cưng. Các mối tương quan mạnh và trung bình cho thấy sự phụ thuộc và ảnh hưởng của các yếu tố như tỷ lệ hộ gia đình nuôi thú cưng, số lượng hộ gia đình nuôi chó và mèo, cũng như sự đa dạng trong lựa chọn loài thú cưng giữa các bang. Việc phân tích này giúp hiểu rõ hơn về hành vi nuôi thú cưng và có thể hỗ trợ trong việc đưa ra chiến lược quản lý và phát triển cho ngành nuôi thú cưng.

## **4.2. Ý nghĩa**

### **4.2.1. Sự phân bố nuôi thú cưng theo từng bang**

Chênh lệch đáng kể giữa các bang: Các bang khác nhau có sự khác biệt lớn về số lượng hộ gia đình nuôi thú cưng, số lượng chó, mèo, và tỷ lệ hộ gia đình nuôi thú cưng. Điều này phản ánh rõ sự đa dạng về văn hóa, điều kiện sống, và sở thích cá nhân ở từng bang.

Bang có tỷ lệ cao: Những bang như Wyoming, West Virginia, Nebraska có tỷ lệ hộ gia đình nuôi thú cưng cao, cho thấy sự yêu thích thú cưng tại các khu vực này.

Bang có tỷ lệ thấp: Những bang như New York và California có tỷ lệ thấp hơn, có thể do không gian sống hạn chế hoặc phong cách sống đô thị không phù hợp với việc nuôi thú cưng.

#### **4.2.2. Số lượng hộ gia đình nuôi chó và mèo**

Số lượng cao: Các bang như Texas và Florida có số lượng hộ gia đình nuôi chó và mèo cao, có thể liên quan đến dân số đông và diện tích bang rộng lớn.

Số lượng thấp: Trong khi đó, các bang như Vermont và Rhode Island thường có số lượng hộ gia đình nuôi chó và mèo thấp hơn. Điều này có thể phản ánh sự hạn chế về không gian sống và phong cách sống nông thôn hơn ở các bang nhỏ này.

#### **4.2.3. Mối tương quan giữa các biến số**

Tỷ lệ hộ gia đình nuôi thú cưng và số lượng hộ gia đình nuôi chó, mèo: Mối tương quan mạnh mẽ (+0.95 và +0.80) giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng hộ gia đình nuôi chó, mèo cho thấy sự phụ thuộc rõ rệt giữa việc sở hữu thú cưng và lựa chọn loài nuôi.

Số lượng chó và mèo mỗi hộ gia đình: Mối tương quan vừa (+0.70) giữa số lượng trung bình chó và mèo mỗi hộ gia đình cho thấy xu hướng nuôi nhiều loại thú cưng trong các hộ gia đình có thú cưng.

#### **4.2.4. Nhận định tổng quát**

Việc trực quan hóa dữ liệu này cung cấp thông tin quan trọng về sự phân bố và xu hướng nuôi thú cưng trên toàn quốc. Điều này không chỉ giúp hiểu sâu hơn về các yếu tố ảnh hưởng đến việc nuôi thú cưng mà còn hỗ trợ trong việc đề xuất các chiến lược quản lý và phát triển cho ngành nuôi thú cưng trong tương lai.

PHẦN 5. TỔNG KẾT BÀI TẬP LỚN

5.1. Kế hoạch

5.1.1. Kế hoạch dự kiến

Giai đoạn 1:

Nhóm cùng nhau lựa chọn ý tưởng: tìm kiếm dữ liệu, phương pháp xử lý dữ liệu,... Cuối cùng nhóm thống nhất đề tài với tập dữ liệu dân số và quyền sở hữu chó và mèo theo hộ gia đình tại Mỹ và sử dụng các mô hình học máy để đánh giá độ chính xác của dữ liệu sau khi xử lý.

Giai đoạn 2:

Trương Hương Giang	Code Trực quan hóa dữ liệu. Phân loại. Làm sạch dữ liệu. Phân tích dữ liệu. Phân tích mối tương quan. Viết báo cáo.
Nguyễn Minh Phương	Phân tích dữ liệu. Lựa chọn tính năng và phân tích mối tương quan. Viết báo cáo, làm slide.

Bảng biểu 5. Phân chia công việc

5.1.2. Phân chia công việc

Vai trò		Trương Hương Giang	Nguyễn Minh Phương
		Trưởng nhóm	Thành viên
STT	Nội dung công việc	Phần % công việc thực hiện	
1	Báo cáo BTL phần 1, 2	0%	100%
2	Báo cáo BTL phần 3, 4, 5	100%	0%
3	Trực quan hóa dữ liệu	100%	0%
4	Phân loại	100%	0%
5	Làm sạch dữ liệu	100%	0%
6	Phân tích dữ liệu	50%	50%

7	Phân tích mối tương quan	50%	50%
8	Làm slide	0%	100%
Tổng kết		60%	40%

Bảng biểu 6. Bảng nội dung công việc

STT	Nhiệm vụ	Ngày						
		07/06/2024	09/06/2024	10/06/2024	18/06/2024	22/06/2024	24/06/2024	01/07/2024
1	Phân loại							
2	Làm sạch dữ liệu							
3	Trực quan hóa dữ liệu							
4	Phân tích dữ liệu							
5	Phân tích mối tương quan							
6	Làm báo cáo							
7	Làm slide							

Hình ảnh 22. Biểu đồ Gantt

### 5.2. Kết quả đạt được

Kết quả đạt được sau bài phân tích và trực quan hóa dữ liệu về sở hữu thú cưng là:

Hiểu rõ hơn về phân bố sở hữu thú cưng theo từng bang: Phân tích đã cho thấy sự đa dạng lớn về tỷ lệ hộ gia đình nuôi thú cưng và số lượng chó, mèo trên mỗi hộ gia đình giữa các bang khác nhau. Những bang như Wyoming, West Virginia, Nebraska được xác định có tỷ lệ cao hộ gia đình nuôi thú cưng, trong khi New York và California có tỷ lệ thấp hơn.

Tìm thấy mối tương quan mạnh giữa các biến số: Các phân tích tương quan cho thấy mối liên hệ dương mạnh giữa tỷ lệ hộ gia đình nuôi thú cưng và số lượng hộ gia đình nuôi thú cưng, đặc biệt là với việc nuôi chó và mèo. Điều này có thể phản ánh xu hướng và sở thích chung của cư dân ở các bang khác nhau.

Cung cấp thông tin hữu ích cho các chiến lược quản lý và phát triển trong ngành nuôi thú cưng: Bằng việc hiểu rõ hơn về hành vi và xu hướng nuôi thú cưng, các tổ chức và chính phủ có thể đưa ra các chiến lược phát triển thích hợp, nhằm nâng cao chất lượng sống và hạnh phúc của cư dân, đồng thời đáp ứng nhu cầu của ngành nuôi thú cưng.

Tóm lại, bài phân tích này mang lại cái nhìn sâu sắc về sự phân bố và mối tương quan giữa các yếu tố liên quan đến sở hữu thú cưng ở Mỹ, từ đó hỗ trợ trong việc đưa ra các quyết định chiến lược trong ngành nuôi thú cưng và quản lý chính sách.

### **5.3. Ưu điểm và hạn chế**

#### **5.3.1. Ưu điểm**

Trực quan hóa dữ liệu: Sử dụng biểu đồ và đồ thị giúp dễ dàng hiểu và phân tích dữ liệu một cách trực quan. Điều này giúp nhìn thấy mô hình và xu hướng một cách rõ ràng hơn, từ đó đưa ra những phân tích chính xác hơn.

Phân tích tương quan: Công cụ cho phép phân tích mối tương quan giữa các biến số, như tỷ lệ hộ gia đình nuôi thú cưng, số lượng chó, mèo, từ đó giúp xác định các mối liên hệ và ảnh hưởng giữa chúng.

Khả năng tổng hợp thông tin: Phương pháp cho phép tổng hợp thông tin từ nhiều nguồn khác nhau vào một nền tảng trực quan và dễ tiếp cận, giúp cho việc đưa ra quyết định quản lý và phát triển có căn cứ khoa học.

Áp dụng linh hoạt: Công cụ và phương pháp có thể áp dụng linh hoạt cho nhiều lĩnh vực khác nhau, không chỉ giới hạn trong lĩnh vực sở hữu thú cưng mà còn có thể mở rộng sang các lĩnh vực khác trong nghiên cứu xã hội và kinh tế.

#### **5.3.2. Hạn chế**

Giới hạn dữ liệu: Phương pháp phụ thuộc vào tính chính xác và sự đa dạng của dữ liệu sẵn có. Nếu dữ liệu không đủ lớn hoặc không đủ đại diện, các kết quả phân tích có thể không thể áp dụng rộng rãi.

Giới hạn mô hình hóa: Việc mô hình hóa dữ liệu dựa trên các biểu đồ và đồ thị có thể giới hạn trong việc đưa ra các dự đoán phức tạp hơn. Các mô hình phân tích này thường không đủ để dự đoán các xu hướng dài hạn hoặc phát triển tương lai.

Phụ thuộc vào người phân tích: Kết quả phân tích có thể phụ thuộc nhiều vào khả năng hiểu và phân tích của người sử dụng công cụ. Việc hiểu sai hoặc đưa ra những giả định sai lầm có thể dẫn đến kết quả không chính xác.

Cần sự hiểu biết chuyên môn: Việc sử dụng phương pháp và công cụ yêu cầu người phân tích có kiến thức vững về thống kê, phân tích dữ liệu và việc đọc hiểu các biểu đồ, đồ thị một cách chuyên sâu để có thể đưa ra các kết luận và nhận định phù hợp.



## PHẦN 6. TÀI LIỆU THAM KHẢO

[1]: Nguyễn Văn Thiệu. “1-0-Introduction”

[https://docs.google.com/presentation/d/1Qj0p2BGnMYtD34oMUMczMYTwMa1aPHUE/edit?usp=drive\\_link&ouid=116745780561294076564&rtpof=true&sd=true](https://docs.google.com/presentation/d/1Qj0p2BGnMYtD34oMUMczMYTwMa1aPHUE/edit?usp=drive_link&ouid=116745780561294076564&rtpof=true&sd=true)

[2]: Sample Data. Lifestyle “Cat vs Dog Popularity in the U.S.”

<https://public.tableau.com/app/learn/sample-data>

[3]: Link Colab:

[https://colab.research.google.com/drive/1G2H1K3QXzf8eXmMyaz1qgtDbc\\_ZsE-TD?usp=sharing](https://colab.research.google.com/drive/1G2H1K3QXzf8eXmMyaz1qgtDbc_ZsE-TD?usp=sharing)

[4]: Link Github:

[https://github.com/iamhuonggiang08/DataVisualization\\_N07\\_Group5](https://github.com/iamhuonggiang08/DataVisualization_N07_Group5)