**TRƯỜNG ĐẠI HỌC PHENIKAA**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A logo for a university

Description automatically generated**

**BÁO CÁO BÀI TẬP LỚN**

**MÔN HỌC: TRỰC QUAN HÓA DỮ LIỆU**

***Nhóm 9   
Đề tài: Trực quan hóa dữ liệu Star Wars Character Details***

**Thành viên nhóm**

|  |  |  |
| --- | --- | --- |
| **Ngô Thị Linh** | **21011613** | **K15-CNTT3** |

***GVHD: Th.S Nguyễn Văn Thiệu***

**04/07//2024 – Hà Nội**

**Mục lục**

[MỞ ĐẦU 4](#_Toc171172920)

[PHẦN 1. GIỚI THIỆU 5](#_Toc171172921)

[**1.1. Trực quan hóa dữ liệu (Data Visualization) 5**](#_Toc171172922)

[1.1.1. Khái niệm 5](#_Toc171172923)

[1.2.3. Phương pháp 7](#_Toc171172924)

[**1.3. Công cụ sử dụng 7**](#_Toc171172925)

[1.3.1. Python và các thư viện đi kèm 7](#_Toc171172926)

[1.3.2. Jupyter Notebook 8](#_Toc171172927)

[1.3.3. Excel và Google Sheets 8](#_Toc171172928)

[1.3.4. Tableau 9](#_Toc171172929)

[1.3.5. SQL 9](#_Toc171172930)

[1.3.6. Google Colab 9](#_Toc171172931)

[PHẦN 2. KIẾN THỨC NỀN TẢNG 10](#_Toc171172932)

[**2.1. Kiến thức cơ bản về trực quan hóa dữ liệu 10**](#_Toc171172933)

[2.1.1. Khái niệm 10](#_Toc171172934)

[2.1.2. Phương pháp 11](#_Toc171172935)

[2.1.3. Công cụ quan trọng 12](#_Toc171172936)

[**2.2. Nguyên tắc và quy tắc thiết kế trực quan hóa dữ liệu 12**](#_Toc171172937)

[PHẦN 3. PHÂN TÍCH DỮ LIỆU DỰA VÀO TRỰC QUAN HÓA 14](#_Toc171172938)

[**3.1. Tiền xử lý dữ liệu 14**](#_Toc171172939)

[3.1.1. Thu thập dữ liệu 14](#_Toc171172940)

[3.1.2. Chuẩn hóa dữ liệu 14](#_Toc171172941)

[3.1.3. Biến đổi dữ liệu 14](#_Toc171172942)

[3.1.4. Tích hợp dữ liệu (Data Integration) 14](#_Toc171172943)

[3.1.5. Lưu trữ và quản lý dữ liệu (Data Storage and Management) 15](#_Toc171172944)

[**3.2. Phân tích dữ liệu 15**](#_Toc171172945)

[**3.3. Quá trình thiết kế 17**](#_Toc171172946)

[3.1.1. Birth\_year 18](#_Toc171172947)

[3.1.2. Height 19](#_Toc171172948)

[3.1.3. Mass 19](#_Toc171172949)

[3.1.4. Gender 20](#_Toc171172950)

[3.1.5. Skin\_color 21](#_Toc171172951)

[3.1.6. Hair\_color 22](#_Toc171172952)

[3.1.7. Eye\_color 23](#_Toc171172953)

[3.1.8. Homeworld 24](#_Toc171172954)

[3.3.9. Nhận xét tổng kết 25](#_Toc171172955)

[**3.4. Nguyên tắc thiết kế 26**](#_Toc171172956)

[3.4.1. Biểu đồ sử dụng 26](#_Toc171172957)

[3.4.2. Thiết kế và xây dựng biểu đồ 26](#_Toc171172958)

[3.4.3. Điều chỉnh và tối ưu hóa 26](#_Toc171172959)

[PHẦN 4. KẾT QUẢ VÀ THẢO LUẬN 27](#_Toc171172960)

[**4.1. Kết quả 27**](#_Toc171172961)

[4.1.1. Kết quả so sánh chiều cao và cân nặng của nhân vật(10 đối tượng) 27](#_Toc171172962)

[4.1.2. Biểu đồ phân phối chiều cao và màu mắt 28](#_Toc171172963)

[4.1.3. Biểu đồ phân phối khối lượng và hành tinh quê hương 28](#_Toc171172964)

[4.1.4. Biểu đồ phân phối năm sinh và hành tinh quê hương 29](#_Toc171172965)

[4.1.5. Biểu đồ phân phối năm sinh và khối lượng 30](#_Toc171172966)

[4.1.6. Biểu đồ phân phối giới tính và chiều cao 31](#_Toc171172967)

[4.1.7. Biểu đồ phân phối giới tính và khối lượng 33](#_Toc171172968)

[**4.2. Ý nghĩa 33**](#_Toc171172969)

[4.2.1. Những Yếu Tố Đặc Điểm Của Nhân Vật 34](#_Toc171172970)

[4.2.2. Dự Đoán Và Triển Vọng 34](#_Toc171172971)

[4.2.3. Hợp Tác Và Cộng Đồng 34](#_Toc171172972)

[PHẦN 5. TỔNG KẾT BÀI TẬP LỚN 35](#_Toc171172973)

[**5.1. Kế hoạch 35**](#_Toc171172974)

[5.1.1. Kế hoạch dự kiến 35](#_Toc171172975)

[5.1.2. Giai đoạn làm việc 35](#_Toc171172976)

[**5.2. Kết quả đạt được 35**](#_Toc171172977)

[**5.3. Ưu điểm và hạn chế 36**](#_Toc171172978)

[5.3.1. Ưu điểm 36](#_Toc171172979)

[5.3.2. Hạn chế 37](#_Toc171172980)

[PHẦN 6. TÀI LIỆU THAM KHẢO 39](#_Toc171172981)

**Danh sách hình ảnh**

[*Hình ảnh 1. Các công cụ sử dụng trong python* 12](#_Toc171173670)

[*Hình ảnh 2. Cấu trúc dữ liệu* 16](#_Toc171173671)

[*Hình ảnh 3.Biểu đồ về năm sinh của các nhân vật* 18](#_Toc171173672)

[*Hình ảnh 4.Biểu đồ về chiều cao của các nhân vật* 19](#_Toc171173673)

[*Hình ảnh 5.Biểu đồ về khối lượng của các nhân vật* 20](#_Toc171173674)

[*Hình ảnh 6.Biểu đồ về giới tính của các nhân vật* 21](#_Toc171173675)

[*Hình ảnh 7.Biểu đồ về màu da của các nhân vật* 22](#_Toc171173676)

[*Hình ảnh 8.Biểu đồ về màu tóc của các nhân vật* 22](#_Toc171173677)

[*Hình ảnh 9.Biểu đồ về màu mắt của các nhân vật* 23](#_Toc171173678)

[*Hình ảnh 10.Biểu đồ về màu tóc của các nhân vật* 24](#_Toc171173679)

[*Hình ảnh 11. Biểu đồ so sánh chiều cao và cân nặng của 10 đối tượng* 27](#_Toc171173680)

[*Hình ảnh 12. Biểu đồ phân phối chiều cao và màu mắt* 28](#_Toc171173681)

[*Hình ảnh 13. Biểu đồ phân phối khối lượng và hành tinh quê hương* 29](#_Toc171173682)

[*Hình ảnh 14. Biểu đồ phân phối năm sinh và hành tinh quê hương* 30](#_Toc171173683)

[*Hình ảnh 15. Biểu đồ phân phối năm sinh và khối lượng* 31](#_Toc171173684)

[*Hình ảnh 16. Biểu đồ phân phối giới tính và chiều cao* 32](#_Toc171173685)

[*Hình ảnh 17. Biểu đồ phân phối giới tính và khối lượng* 33](#_Toc171173686)

[*Hình ảnh 18 . Biểu đồ Gantt* 35](#_Toc171173687)

# MỞ ĐẦU

Trong vũ trụ rộng lớn của Star Wars, không chỉ là những cuộc phiêu lưu đầy hành động và những trận chiến quyết liệt, mà còn là một thế giới phức tạp, nơi mà nhân vật từng được khắc họa với độ chi tiết tinh xảo, phản ánh sâu sắc về con người và những giá trị tinh thần cao quý.

Mỗi nhân vật trong vũ trụ này không chỉ đơn giản là một hình mẫu anh hùng, mà là biểu tượng của sự đấu tranh cho sự tự do và công lý, một tín ngưỡng vững chắc về những lý tưởng cao cả hơn bản thân. Từ những chiến binh dũng cảm như Luke Skywalker và Rey, đến những nhân vật đầy sức mạnh và sự dẻo dai như Darth Vader và Kylo Ren, mỗi cá nhân đều mang trong mình một câu chuyện phức tạp, một hành trình tìm kiếm bản thân và những lựa chọn quan trọng giữa sự ám ảnh và lòng từ bi.

Việc nghiên cứu và phân tích sâu sắc về từng nhân vật trong Star Wars không chỉ giúp ta khám phá ra sự phát triển và sâu sắc hơn về tính cách của họ, mà còn là cơ hội để tiếp cận với những thông điệp sâu sắc về can đảm, tình bạn và sự hy sinh. Những câu chuyện này cũng là đề tài để xây dựng các phương pháp và công cụ phân tích hiệu quả, giúp hỗ trợ quyết định và truyền tải thông tin một cách sinh động và rõ ràng.

Bài toán này không chỉ dừng lại ở việc khai phá dữ liệu mà còn là việc tái hiện lại những câu chuyện sâu sắc và mang đến những góc nhìn mới về một trong những vũ trụ khoa học viễn tưởng đáng kính trong lịch sử điện ảnh. Đó là một cơ hội để chúng ta cảm nhận và khám phá những giá trị về lòng can đảm, sự đoàn kết và nghị lực vượt qua mọi thử thách, từ đó mở ra những suy ngẫm sâu sắc về bản chất của con người và vũ trụ mênh mông này.

# 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 về các nhân vật trong vũ trụ Star Wars là để hiểu sâu hơn về mối quan hệ giữa các thuộc tính nhân vật và các giá trị văn hóa, tinh thần mà họ đại diện. Trong nghiên cứu này, tôi sẽ sử dụng các bộ dữ liệu cụ thể, chẳng hạn như chiều cao, màu mắt, và các chỉ số khác để phân tích và trực quan hóa thông tin.

Bài toán này giúp tôi hiểu rõ hơn về các thuộc tính của nhân vật trong Star Wars, nhận biết các mẫu và xu hướng giữa các thuộc tính như chiều cao và màu mắt, cùng với các chỉ số khác. Nghiên cứu cũng giúp làm sáng tỏ mối liên hệ giữa các nhân vật và các giá trị văn hóa, tinh thần mà họ thể hiện, từ đó cung cấp cái nhìn sâu sắc hơn về vũ trụ Star Wars.

**1.2.2. Mục tiêu**

Bài toán đưa ra 4 mục tiêu chính:

- Phân tích các thuộc tính nhân vật: Điều này bao gồm việc tìm hiểu các thuộc tính như chiều cao, màu mắt của các nhân vật trong vũ trụ Star Wars. Thông qua việc phân tích này, chúng ta có thể nhận biết được các đặc điểm nổi bật và xu hướng chung của các nhân vật trong từng bộ phim hoặc từng phần của series.

- Định tính hóa xu hướng và tính cách nhân vật: Bằng cách phân tích các thuộc tính như chiều cao, màu mắt, giới tính và vai trò trong cốt truyện, chúng ta có thể xác định xu hướng và đặc điểm chung của các nhân vật. Những thông tin này cung cấp cái nhìn sâu hơn về sự phát triển và tính cách của nhân vật trong cộng đồng Star Wars.

- Liên kết với giá trị văn hóa và xã hội: Việc phân tích các thuộc tính nhân vật theo bối cảnh văn hóa và xã hội giúp ta hiểu được tầm quan trọng của các yếu tố văn hóa và xã hội đối với sự phát triển và hành vi của các nhân vật. Điều này có thể phản ánh các giá trị văn hóa mà bộ phim muốn truyền tải và ảnh hưởng của chúng đến khán giả.

- Phân tích theo vị trí và vai trò trong cốt truyện: Bằng cách phân tích các nhân vật theo vị trí địa lý và vai trò trong cốt truyện, chúng ta có thể nhận thấy những xu hướng khác nhau trong việc phát triển nhân vật giữa các phần của series và các câu chuyện riêng lẻ. Điều này giúp cho việc đưa ra các chiến lược sáng tạo và tiếp thị hiệu quả hơn, phù hợp với từng đặc điểm của từng phần phim.

Do đó, bài toán này không chỉ giúp ta hiểu sâu hơn về các nhân vật trong Star Wars mà còn mở ra cơ hội áp dụng những kiến thức này vào các lĩnh vực như phân tích văn hóa, truyền thông và tiếp thị sản phẩm liên quan đến Star Wars. 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 chiến lược liên quan đến thương hiệu Star Wars hiệu quả hơn trong cộng đồng.

### 1.2.3. Phương pháp

Các phương pháp được sử dụng trong bài toán phân tích dữ liệu các nhân vật trong vũ trụ Star Wars bao gồm:

- Phân tích Tương quan (Correlation Analysis): Được sử dụng để phân tích mối tương quan giữa các thuộc tính của nhân vật như chiều cao, màu mắt, giới tính, và vai trò trong cốt truyện. 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 nhận biết những liên kết quan trọng giữa các thuộc tính nhân vật.

- Phân tích Vùng (Spatial Analysis): Được dùng để đánh giá sự phân bố và xu hướng của các nhân vật theo bối cảnh địa lý và các hành tinh trong vũ trụ Star Wars. Sử dụng các phương pháp như bản đồ hành tinh (GIS) và phân tích điểm nóng (hot spot analysis) để xác định các khu vực có mật độ nhân vật đặc biệt cao. Điều này có thể phát hiện ra các xu hướng địa lý và cung cấp thông tin quan trọng cho việc phân tích và phát triển cốt truyện.

- Phân tích Thống kê Mô tả (Descriptive Statistics): Được dùng để tóm tắt và mô tả các thuộc tính chính của nhân vật như số lượng nhân vật theo từng loại, chiều cao trung bình, phân bố màu mắt, 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 các thuộc tính nhân vật trong vũ trụ Star Wars.

## 1.3. Công cụ sử dụng

Trong bài toán phân tích dữ liệu về các nhân vật trong vũ trụ Star Wars, 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 json: Sử dụng để thao tác với dữ liệu JSON (JavaScript Object Notation). JSON là định dạng dữ liệu phổ biến để trao đổi thông tin giữa máy chủ và ứng dụng web. Chuyển đổi dữ liệu từ định dạng JSON sang các đối tượng Python và ngược lại. Đọc và ghi dữ liệu JSON từ các tệp.

- Thư viện requests: Sử dụng để gửi các yêu cầu HTTP. Đây là một thư viện đơn giản và dễ sử dụng để lấy dữ liệu từ các API web. Gửi các yêu cầu HTTP như GET, POST, PUT, DELETE. Xử lý phản hồi từ máy chủ và quản lý các tham số URL, tiêu đề HTTP và các dữ liệu khác liên quan đến yêu cầu.

- 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 plotly.graph\_objects và plotly.express: Sử dụng để tạo các biểu đồ tương tác và phức tạp. Plotly cung cấp các công cụ để tạo và tùy chỉnh các loại biểu đồ như biểu đồ đường, biểu đồ thanh, biểu đồ hình tròn, và nhiều loại biểu đồ khác. Plotly Express đơn giản hóa quy trình tạo biểu đồ với ít mã hơn và cung cấp các hàm tích hợp để dễ dàng tạo các biểu đồ đẹp mắt và trực quan.

- Thư viện matplotlib.pyplot: Sử dụng để tạo các biểu đồ tĩnh trong Python. Matplotlib cung cấp các công cụ để tạo ra các biểu đồ 2D và 3D chi tiết. Tạo các biểu đồ đường, biểu đồ thanh, biểu đồ hình tròn, biểu đồ phân tán, và nhiều loại biểu đồ khác. Tùy chỉnh các yếu tố của biểu đồ như màu sắc, nhãn, tiêu đề, và phong cách. Lưu các biểu đồ dưới nhiều định dạng hình ảnh khác nhau như PNG, SVG, PDF.

### 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

Tableau là phần mềm hỗ trợ phân tích (Data Analyst) và trực quan hóa dữ liệu (Data Visualization). Tableau giúp tổng hợp các dữ liệu này từ các dãy số thành những hình ảnh, biểu đồ trực quan, xây dựng các dashboard và các phân tích (Self-services).

### 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 sẽ 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*

Trong bài toán phân tích dữ liệu liên quan đến các nhân vật trong Star Wars, chúng ta sử dụng nhiều công cụ và thư viện khác nhau để thu thập, xử lý và trực quan hóa dữ liệu. Dưới đây là các công cụ chính được sử dụng:

- Google Data Studio / Looker Studio: Được sử dụng để kết nối dễ dàng với nhiều sản phẩm và dịch vụ khác nhau. Đây là một công cụ miễn phí của Google, cho phép người dùng tạo các báo cáo và bảng điều khiển (dashboard) tùy chỉnh, trực quan, giúp theo dõi và phân tích dữ liệu một cách hiệu quả.

- Tableau: Giúp người dùng dễ dàng sắp xếp dữ liệu và nhập dữ liệu từ nhiều nguồn khác nhau. Tableau cung cấp khả năng tích hợp với nhiều nền tảng khác, tạo ra các biểu đồ và bảng điều khiển tương tác, hỗ trợ phân tích dữ liệu phức tạp một cách trực quan và dễ dàng.

- Infogram: Có nhiều tùy chọn độc đáo, với nhiều mẫu tối ưu hóa cho di động và mạng xã hội. Infogram cho phép người dùng nhúng các biểu đồ và bản đồ vào thiết kế trang web, cung cấp các công cụ trực quan để tạo ra các biểu đồ và infographic chuyên nghiệp.

- D3.js: Được sử dụng để tạo các biểu đồ tương tác và hoạt ảnh mạnh mẽ. Đây là một thư viện JavaScript mã nguồn mở, cung cấp nhiều ví dụ tương tác và hỗ trợ phân cấp dữ liệu, cho phép người dùng tạo ra các biểu đồ và hình ảnh động phức tạp.

- Excel và Power BI: Là bộ công cụ chuyên về trực quan hóa dữ liệu (DV) của Microsoft. Excel phù hợp cho các phân tích dữ liệu ngắn hạn và đơn giản, trong khi Power BI cung cấp khả năng tạo các báo cáo và bảng điều khiển phức tạp hơn, kết nối dữ liệu từ nhiều nguồn khác nhau.

- Ngôn ngữ lập trình Python và các thư viện liên quan: Python là ngôn ngữ lập trình miễn phí, dễ học và có nhiều thư viện hỗ trợ mạnh mẽ cho việc phân tích và trực quan hóa dữ liệu. Các thư viện phổ biến bao gồm:

* Matplotlib: Sử dụng để tạo các biểu đồ tĩnh 2D và 3D chi tiết.
* Seaborn: Xây dựng trên Matplotlib, cung cấp các biểu đồ thống kê đẹp mắt.
* Pandas: Cung cấp các công cụ mạnh mẽ để thao tác và xử lý dữ liệu.
* GeoPandas: Mở rộng Pandas để làm việc với dữ liệu không gian địa lý….

## 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: Sử dụng kích cỡ, màu sắc, phông chữ và đồ họa để thu hút sự chú ý của khán giả vào các chi tiết quan trọng trong dữ liệu. Những yếu tố này giúp làm nổi bật thông tin cần truyền tải và giúp khán giả dễ dàng nhận diện các điểm mấu chốt.

- Cung cấp ngữ cảnh cho dữ liệu bằng các dấu hiệu trực quan: Ngữ cảnh là yếu tố quan trọng để khán giả hiểu rõ dữ liệu. Sử dụng các dấu hiệu trực quan như đường lưới (gridlines), chú thích (annotations) và màu sắc để cung cấp bối cảnh cho dữ liệu, giúp khán giả dễ dàng theo dõi và hiểu rõ các xu hướng và mối quan hệ.

- Chọn kiểu phối màu phù hợp cho phân tích dữ liệu: Màu sắc đóng vai trò quan trọng trong trực quan hóa dữ liệu. Chọn bảng màu phù hợp giúp tăng tính thẩm mỹ và cải thiện khả năng đọc hiểu của biểu đồ. Hãy chắc chắn rằng màu sắc được sử dụng có độ tương phản đủ để phân biệt các phần tử khác nhau trong biểu đồ.

- 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. Tiêu đề nên ngắn gọn nhưng đủ ý nghĩa để khán giả nắm bắt được nội dung chính của biểu đồ.

- Thêm nhãn và số rõ ràng cho dữ liệu: Nhãn và số liệu rõ ràng giúp khán giả dễ dàng hiểu và phân tích dữ liệu được trình bày. Sử dụng các phông chữ dễ đọc và đặt nhãn một cách hợp lý để tránh gây rối mắt.

- 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 tập tin people.json trong dự án SWAPI (Star Wars API). Dữ liệu này bao gồm thông tin chi tiết về các nhân vật trong vũ trụ Star Wars.

- Phương pháp thu thập: Sử dụng API hoặc tải xuống các tệp dữ liệu từ trang web <https://github.com/phalt/swapi/blob/master/resources/fixtures/people.json> 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 chiều cao của nhân vật từ các nguồn khác nhau thành bảng chiều cao hoàn chỉnh.

- Xác thực và kiểm tra dữ liệu (Data Validation and Verification)

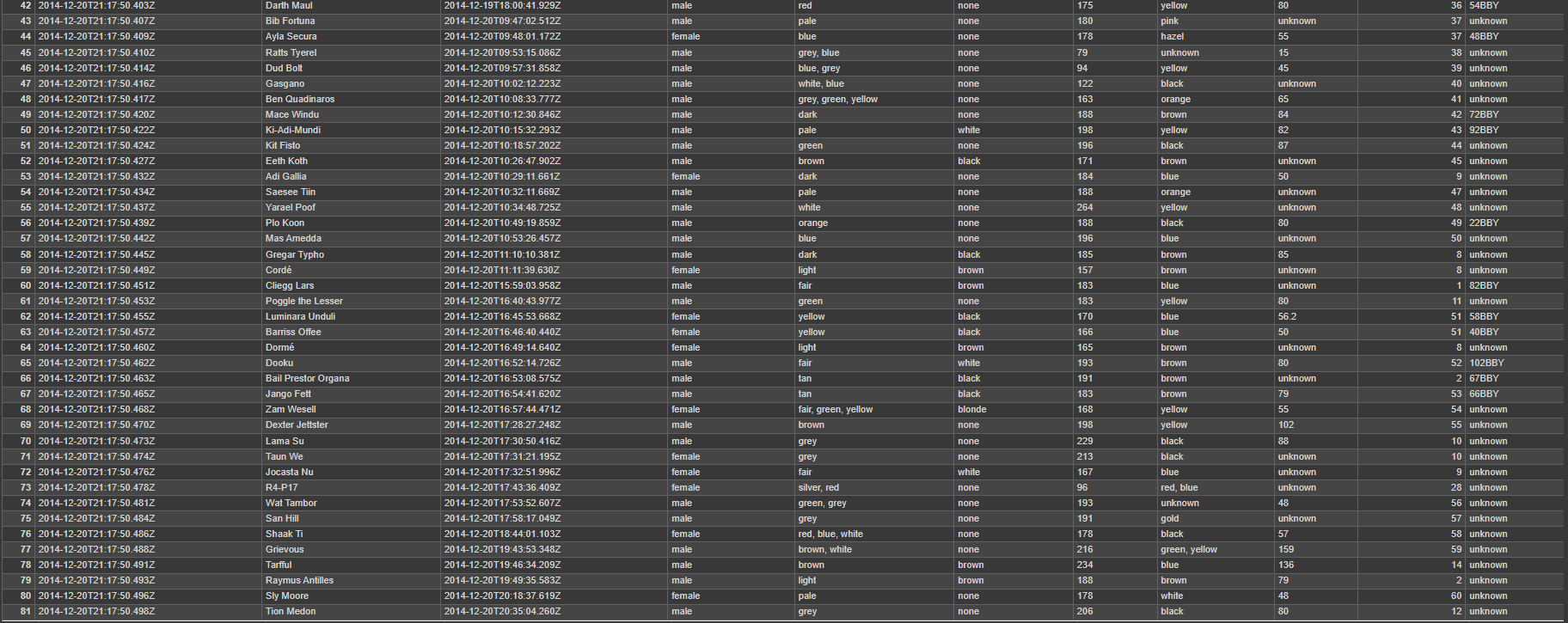
* Kiểm tra sự nhất quán và chính xác: Sử dụng các phương pháp kiểm tra dữ liệu để đảm bảo dữ liệu đã được làm sạch và chuẩn hóa chính xác.
* Xác nhận tính hợp lệ của dữ liệu: Sử dụng các phương pháp thống kê và trực quan hóa để kiểm tra tính hợp lệ của dữ liệu.

### 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



*Hình ảnh 2. Cấu trúc dữ liệu*

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

* edited: Thời gian chỉnh sửa
* name: Tên
* created: Thời gian tạo
* gender: Giới tính
* skin\_color: Màu da
* hair\_color: Màu tóc
* height: Chiều cao
* eye\_color: Màu mắt
* mass: Khối lượng
* homeworld: Hành tinh quê hương
* birth\_year: Năm sinh

*Nhận xét về dữ liệu:*

- Thời gian chỉnh sửa cho biết lần cuối cùng dữ liệu về nhân vật được cập nhật, giúp đảm bảo tính chính xác và mới nhất của thông tin.

- Tên của các nhân vật giúp chúng ta xác định rõ ràng từng cá nhân trong vũ trụ Star Wars, từ đó dễ dàng phân tích các thuộc tính khác liên quan đến họ.

- Thời gian tạo ghi lại thời điểm mà thông tin về nhân vật được đưa vào cơ sở dữ liệu, cung cấp bối cảnh về quá trình cập nhật dữ liệu.

- Giới tính của các nhân vật giúp hiểu rõ hơn về sự đa dạng và phân bố giới tính trong vũ trụ Star Wars.

- Màu da của các nhân vật giúp nghiên cứu sự đa dạng về chủng tộc và loài trong Star Wars, cũng như thể hiện các đặc trưng văn hóa và sinh học khác nhau.

- Màu tóc cung cấp thông tin về ngoại hình của các nhân vật, giúp phân tích các đặc điểm nhân dạng và phong cách cá nhân.

- Chiều cao của các nhân vật giúp nghiên cứu về sự đa dạng hình thể và so sánh giữa các loài khác nhau trong vũ trụ Star Wars.

- Màu mắt cung cấp thêm một khía cạnh về ngoại hình và có thể được sử dụng để phân loại các nhân vật theo những đặc điểm nhận diện cụ thể.

- Khối lượng của các nhân vật giúp nghiên cứu về đặc điểm thể chất và so sánh giữa các nhân vật khác nhau.

- Hành tinh quê hương cho biết nguồn gốc địa lý của các nhân vật, giúp phân tích sự phân bố và nguồn gốc của các loài trong vũ trụ Star Wars.

- Năm sinh của các nhân vật cung cấp thông tin về độ tuổi và thế hệ, giúp hiểu rõ hơn về lịch sử và dòng thời gian của các sự kiện trong Star Wars.

Dữ liệu này cung cấp cái nhìn tổng quan về các nhân vật trong Star Wars, giúp hiểu rõ hơn về sự đa dạng và đặc điểm của các nhân vật. Những thông tin này có thể được sử dụng để nghiên cứu về các yếu tố văn hóa, sinh học và xã hội trong vũ trụ Star Wars, cũng như lập kế hoạch cho các nghiên cứu và sản phẩm liên quan.

## 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 việc thu thập dữ liệu từ dự án Star Wars và thực hiện các bước làm sạch và chuẩn bị dữ liệu để sử dụng cho phân tích và trực quan hóa.

- Bước 2: Lựa chọn loại biểu đồ phù hợp: Chọn và sử dụng các loại biểu đồ như biểu đồ cột, biểu đồ tròn, biểu đồ đường, … để phù hợp với từng loại dữ liệu và mục đích phân tích.

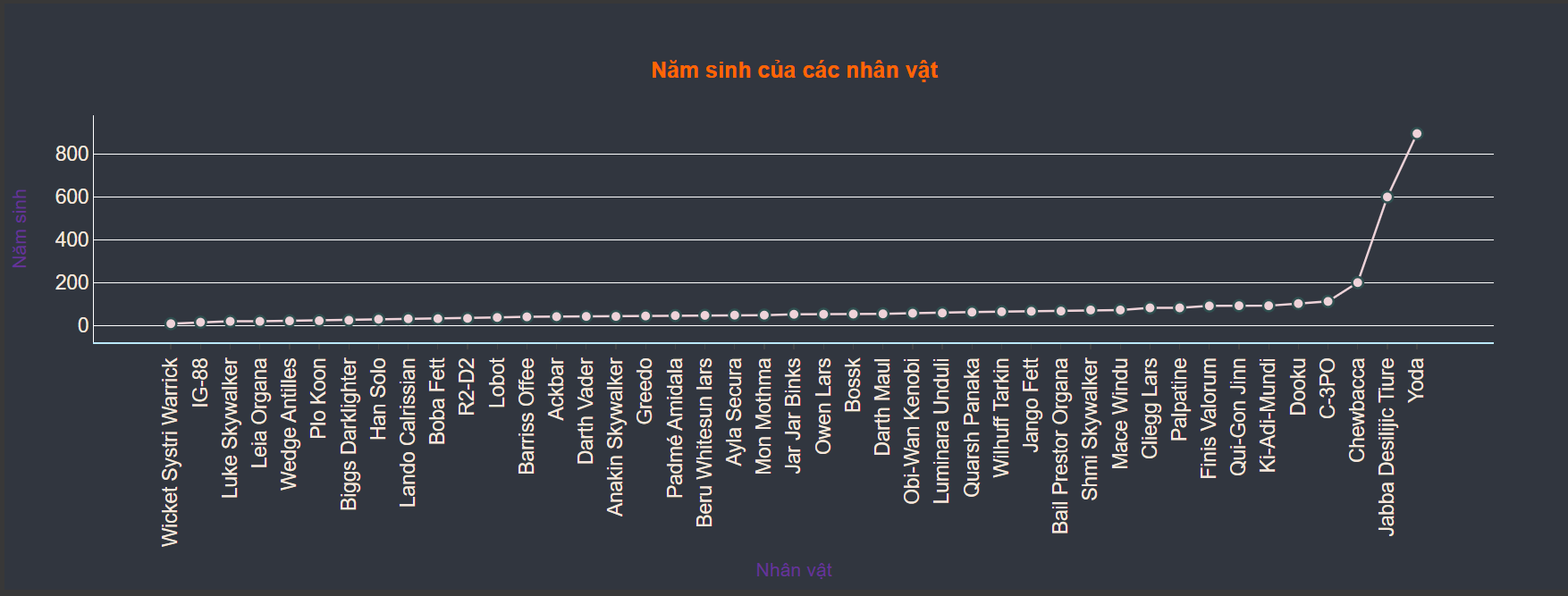
- Bước 3: Thiết kế và xây dựng biểu đồ: Sử dụng các công cụ như Plotly, Matplotlib để tạo và cấu hình các biểu đồ, đảm bảo rằng chúng phản ánh chính xác nội dung dữ liệu. Gắn nhãn đầy đủ và sử dụng màu sắc để làm nổi bật các nhóm dữ liệu khác nhau.

- Bước 4: Phân tích và nhận xét: Phân tích các yếu tố liên quan đến các đặc điểm của nhân vật trong vũ trụ Star Wars giúp hiểu rõ hơn về các xu hướng và đặc điểm của từng nhân vật ,đóng góp vào việc phát triển và mở rộng vũ trụ này. Việc áp dụng các phân tích sẽ hỗ trợ trong việc tối ưu hóa và phát triển vũ trụ Star Wars một cách bền vững và sáng tạo hơn.

- Bước 5: Điều chỉnh và tối ưu hóa 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 về vũ trụ Star Wars Character Details.

Quá trình này giúp đảm bảo rằng dữ liệu được biểu diễn một cách rõ ràng, dễ hiểu và có thể sử dụng để hỗ trợ quyết định hoặc nhận định quan trọng liên quan đến việc phát triển vũ trụ Star Wars Character Details.

### 3.1.1. Birth\_year

*Hình ảnh 3.Biểu đồ về năm sinh của các nhân vật*

*Dựa vào hình 3 cho thấy được*: Các nhân vật có năm sinh gần trận chiến Yavin được hiển thị rõ ràng trên biểu đồ với các điểm trên đường biểu diễn. Điều này giúp chúng ta dễ dàng xác định các nhân vật sinh ra trước và sau sự kiện quan trọng này. Những nhân vật này có thể đóng vai trò quan trọng trong các sự kiện gần trận chiến Yavin và có liên quan đến các sự kiện chính trong dòng thời gian Star Wars.

Các nhân vật có năm sinh xa trận chiến Yavin được hiển thị ở hai phía của biểu đồ, với các giá trị năm sinh dương (trước trận chiến Yavin) và âm (sau trận chiến Yavin). Những nhân vật này có thể thuộc các giai đoạn khác nhau trong lịch sử Star Wars và có vai trò trong các sự kiện không liên quan trực tiếp đến trận chiến Yavin.

Biểu đồ cho thấy sự phân bố năm sinh không đồng đều của các nhân vật trong Star Wars, phản ánh sự đa dạng và phong phú của các giai đoạn lịch sử trong vũ trụ này. Một số nhân vật có năm sinh nổi bật gần các sự kiện chính như trận chiến Yavin, cho thấy họ có vai trò quan trọng trong các sự kiện này. Ngược lại, một số nhân vật có năm sinh cách xa các sự kiện chính, có thể đóng vai trò trong các câu chuyện phụ hoặc các giai đoạn ít được chú ý hơn trong dòng thời gian Star Wars.

Biểu đồ cung cấp cái nhìn tổng quan về sự phân bố năm sinh của các nhân vật trong vũ trụ Star Wars, giúp chúng ta hiểu rõ hơn về mối quan hệ giữa các nhân vật và các sự kiện lịch sử trong dòng thời gian này.

### 3.1.2. Height

****

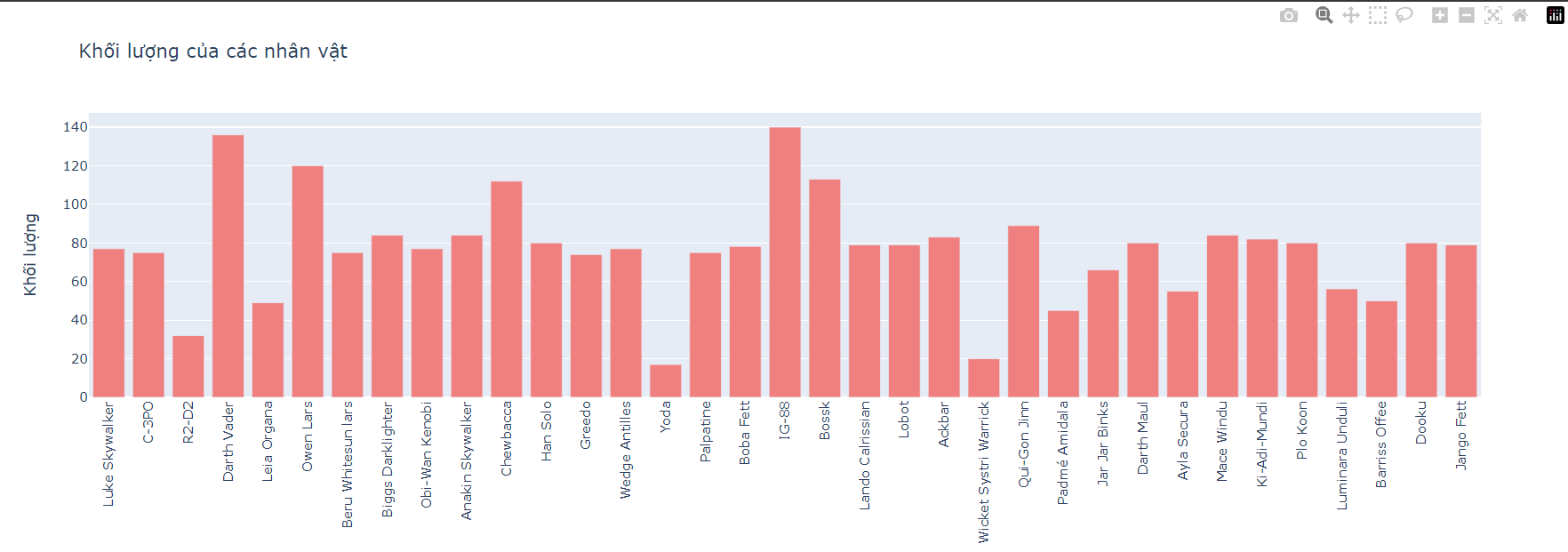
*Hình ảnh 4.Biểu đồ về chiều cao của các nhân vật*

*Dựa vào hình 4 cho thấy được*: Các nhân vật có chiều cao phổ biến được thể hiện rõ ràng trên biểu đồ với các cột cao hơn. Điều này giúp chúng ta dễ dàng xác định chiều cao phổ biến nhất trong các nhân vật. Những nhân vật này có thể đại diện cho những nhân vật phổ biến hoặc những nhân vật chính trong vũ trụ Star Wars. Các nhân vật có chiều cao không phổ biến được thể hiện ở hai phía của biểu đồ, với các giá trị chiều cao xa so với giá trị phổ biến nhất. Những nhân vật này có thể thuộc các chủng tộc hoặc loài khác nhau, hoặc có những đặc điểm ngoại hình đặc biệt trong dòng thời gian Star Wars.

Biểu đồ cho thấy sự phân bố chiều cao không đồng đều của các nhân vật trong Star Wars, phản ánh sự đa dạng và phong phú của các nhân vật trong vũ trụ này. Một số nhân vật có chiều cao phổ biến, cho thấy họ có thể đóng vai trò quan trọng hoặc được nhiều người biết đến trong các sự kiện chính của Star Wars. Ngược lại, một số nhân vật có chiều cao không phổ biến có thể đóng vai trò trong các câu chuyện phụ hoặc có những đặc điểm ngoại hình đặc biệt, làm cho họ khác biệt trong vũ trụ Star Wars.

Biểu đồ cung cấp cái nhìn tổng quan về sự phân bố chiều cao của các nhân vật trong vũ trụ Star Wars, giúp chúng ta hiểu rõ hơn về mối quan hệ giữa các nhân vật và các đặc điểm ngoại hình của họ.

### 3.1.3. Mass

****

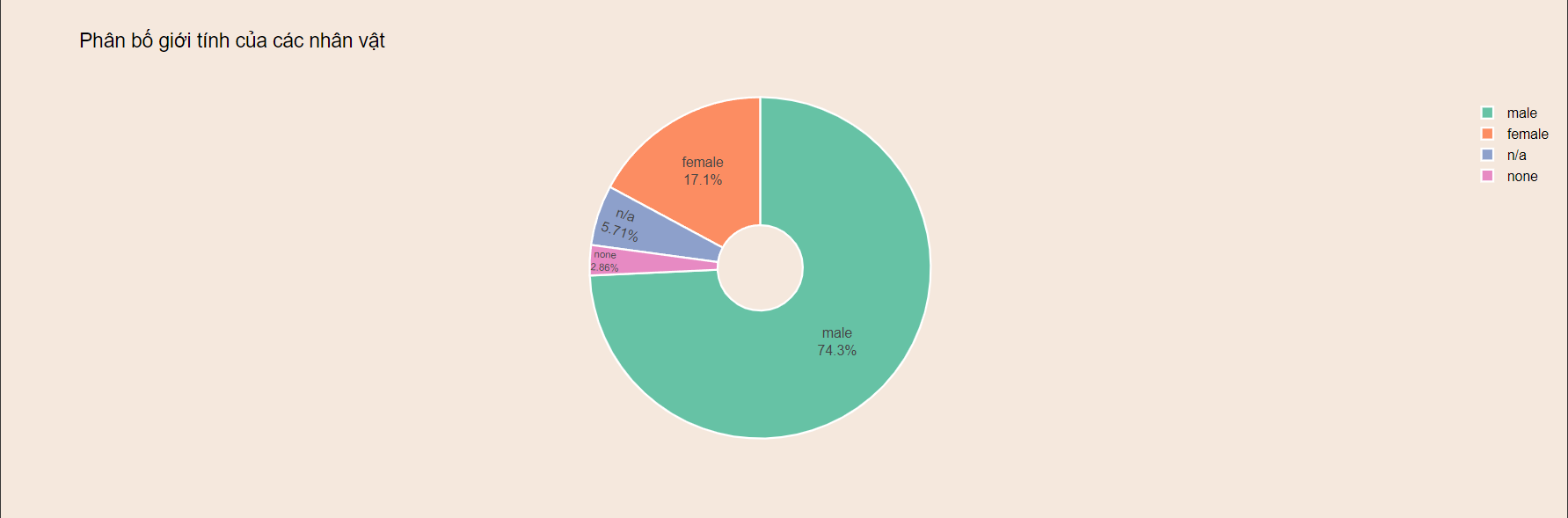
*Hình ảnh 5.Biểu đồ về khối lượng của các nhân vật*

*Dựa vào hình 5 cho thấy được*: Các nhân vật có khối lượng lớn nổi bật với các cột dài hơn. Đây có thể là những nhân vật có kích thước lớn hoặc có vai trò đặc biệt trong các cuộc chiến. Những nhân vật này có thể đại diện cho các loài hoặc chủng tộc có đặc điểm ngoại hình lớn hơn hoặc mạnh mẽ hơn trong vũ trụ Star Wars. Các nhân vật có khối lượng nhỏ được thể hiện qua các cột ngắn hơn. Đây có thể là các nhân vật nhỏ bé hơn hoặc có vai trò ít nổi bật hơn. Những người này ở khu vực có thể thiếu cơ sở hạ tầng, ít cơ hội việc làm, hoặc điều kiện sống không thuận lợi, dẫn đến dân số ít hơn.

Sự khác biệt rõ rệt về chiều cao của các cột cho thấy sự phân bố không đồng đều của khối lượng giữa các nhân vật. Điều này phản ánh sự đa dạng về đặc điểm ngoại hình và sinh học của các nhân vật trong vũ trụ Star Wars. Một số nhân vật có cột cao vượt trội, cho thấy khối lượng của họ rất lớn. Đây có thể là các nhân vật như Chewbacca, Jabba the Hutt, hoặc các sinh vật có kích thước lớn khác. Ngược lại, các nhân vật có cột thấp hơn có thể là các nhân vật như Yoda, các droid nhỏ, hoặc các sinh vật có kích thước nhỏ hơn. Ví dụ, các nhân vật như R2-D2 hoặc BB-8 có thể có khối lượng thấp hơn do kích thước nhỏ gọn.

Biểu đồ cung cấp cái nhìn tổng quan về phân bố khối lượng của các nhân vật trong vũ trụ Star Wars, giúp hiểu rõ hơn về sự đa dạng và phong phú của các nhân vật.

### 3.1.4. Gender

****

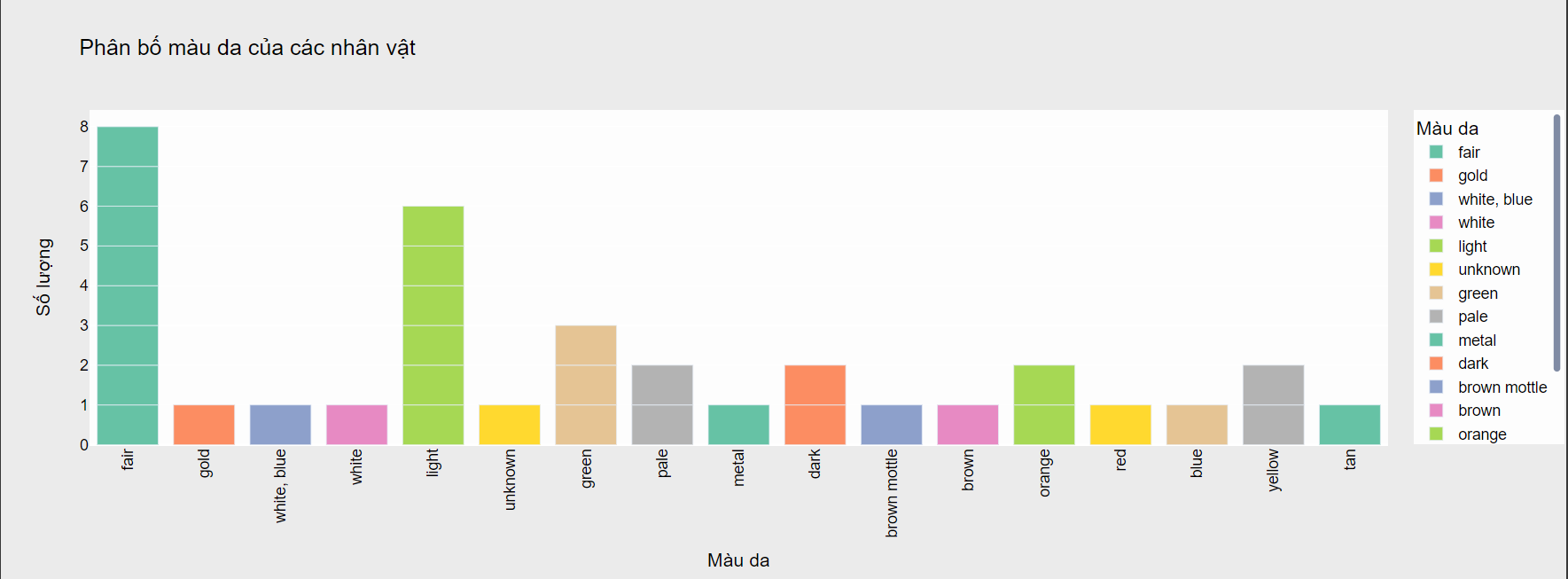
*Hình ảnh 6.Biểu đồ về giới tính của các nhân vật*

*Dựa vào hình 6 cho thấy được*: Các giới tính chiếm tỷ lệ lớn sẽ có các phần lớn hơn trong biểu đồ tròn. Điều này cho thấy sự phổ biến của giới tính đó trong danh sách các nhân vật. Ví dụ, nếu phần lớn các nhân vật là nam, thì phần màu đại diện cho giới tính nam sẽ chiếm diện tích lớn hơn. Các giới tính chiếm tỷ lệ nhỏ sẽ có các phần nhỏ hơn trong biểu đồ tròn. Điều này cho thấy sự ít phổ biến của giới tính đó trong danh sách các nhân vật. Ví dụ, nếu số lượng nhân vật không rõ giới tính hoặc giới tính khác (ngoài nam và nữ) là ít, thì phần màu đại diện cho các giới tính này sẽ chiếm diện tích nhỏ hơn.

Sự khác biệt về kích thước các phần trong biểu đồ tròn cho thấy sự phân bố không đồng đều của các giới tính trong danh sách các nhân vật. Điều này có thể phản ánh xu hướng tạo nhân vật của các nhà biên kịch và đạo diễn. Một số giới tính có phần lớn hơn, cho thấy số lượng nhân vật thuộc giới tính đó rất nhiều. Đây có thể là nam hoặc nữ, tùy thuộc vào cách xây dựng nhân vật trong phim. Ngược lại, các giới tính có phần nhỏ hơn cho thấy số lượng nhân vật thuộc giới tính đó ít hơn. Ví dụ, các nhân vật có giới tính không rõ ràng hoặc giới tính khác có thể chiếm tỷ lệ nhỏ trong tổng số nhân vật.

Biểu đồ cung cấp cái nhìn tổng quan về phân bố giới tính của các nhân vật trong vũ trụ Star Wars, giúp hiểu rõ hơn về sự đa dạng và phong phú của các nhân vật từ góc độ giới tính.

### 3.1.5. Skin\_color

****

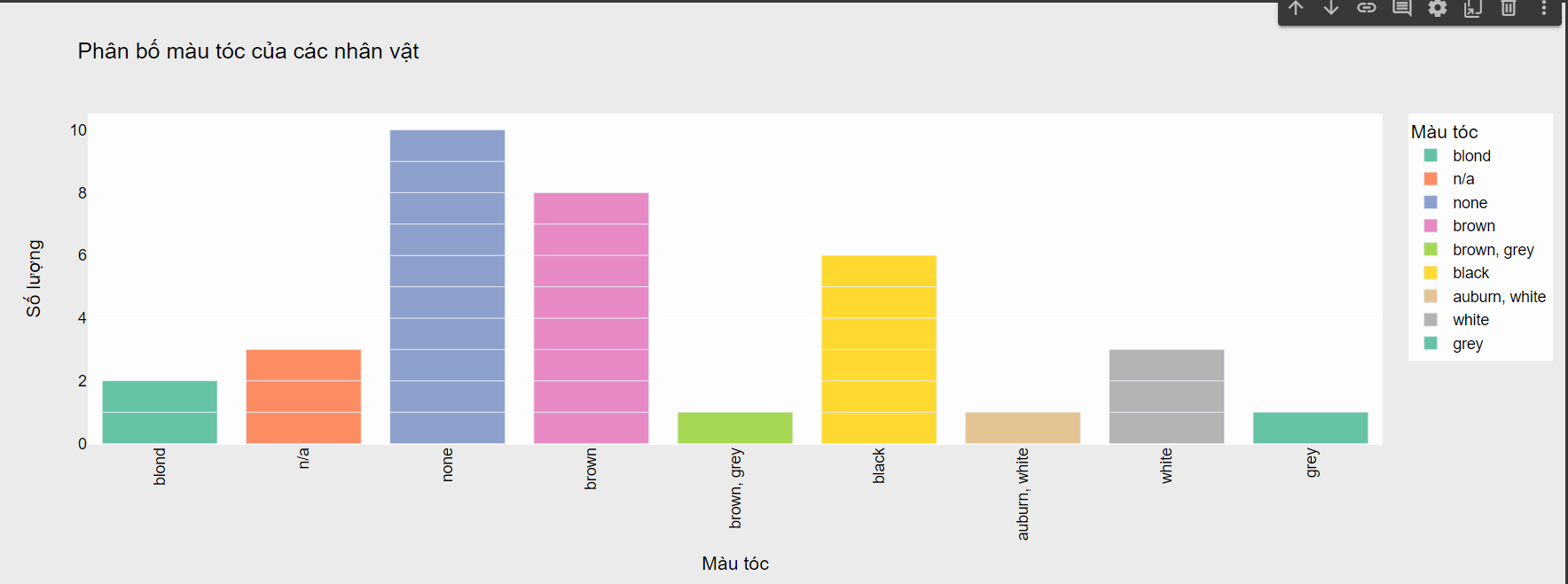
*Hình ảnh 7.Biểu đồ về màu da của các nhân vật*

*Dựa vào hình 7 cho thấy được*: Các màu da phổ biến sẽ có cột cao hơn và nổi bật hơn trong biểu đồ. Điều này cho thấy sự đa dạng và phổ biến của các màu da này trong danh sách các nhân vật. Ví dụ, nếu màu da như "màu da trắng" hoặc "màu da xanh lá cây" chiếm tỷ lệ lớn, thì các cột tương ứng sẽ cao hơn và có màu sắc rõ ràng hơn. Các màu da ít phổ biến sẽ có cột thấp hơn và có thể màu sắc không rõ ràng trong biểu đồ. Điều này cho thấy sự hiếm có và ít được sử dụng của các màu da này trong danh sách các nhân vật. Ví dụ, các màu da như "màu da cam" hoặc "màu da đen" có thể có cột thấp hơn do số lượng nhân vật thấp.

Biểu đồ thể hiện sự đa dạng màu da của các nhân vật trong vũ trụ Star Wars, từ những màu da phổ biến đến những màu da ít thường thấy. Các màu da chiếm tỷ lệ lớn sẽ có cột cao và rõ ràng hơn, dễ thu hút sự chú ý của người xem. Những màu da ít phổ biến có thể không được biểu thị rõ ràng trong biểu đồ, nhưng vẫn quan trọng để hiểu sự đa dạng trong thiết kế các nhân vật.

Biểu đồ cung cấp cái nhìn tổng quan về sự đa dạng màu da của các nhân vật trong vũ trụ Star Wars, từ các màu da phổ biến đến các màu da ít thường thấy.

### 3.1.6. Hair\_color



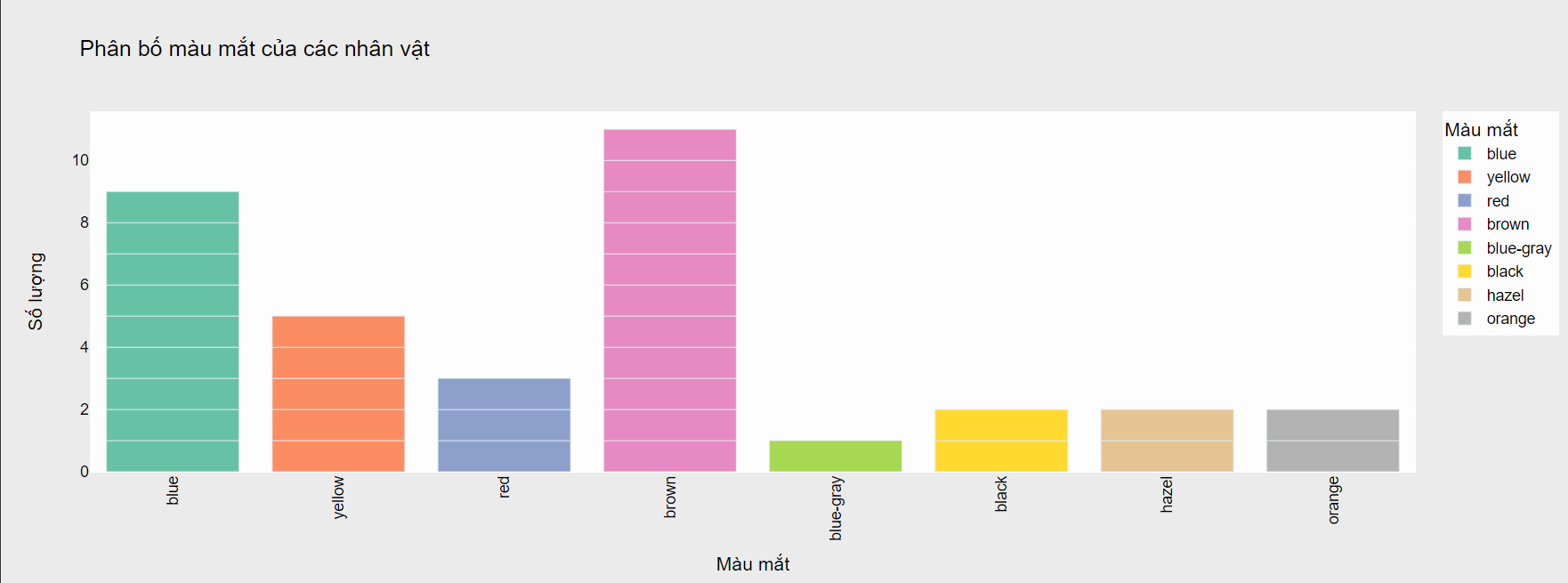
*Hình ảnh 8.Biểu đồ về màu tóc của các nhân vật*

*Dựa vào hình 8 cho thấy được*: Các màu tóc phổ biến sẽ có cột cao hơn và nổi bật hơn trong biểu đồ. Điều này cho thấy sự đa dạng và phổ biến của các màu tóc này trong danh sách các nhân vật. Ví dụ, nếu màu tóc như "nâu" hoặc "đen" chiếm tỷ lệ lớn, thì các cột tương ứng sẽ cao hơn và có màu sắc rõ ràng hơn. Các màu tóc ít phổ biến sẽ có cột thấp hơn và có thể màu sắc không rõ ràng trong biểu đồ. Điều này cho thấy sự hiếm có và ít được sử dụng của các màu tóc này trong danh sách các nhân vật. Ví dụ, các màu tóc như "bạc" hoặc "đỏ" có thể có cột thấp hơn do số lượng nhân vật thấp.

Biểu đồ thể hiện sự đa dạng màu tóc của các nhân vật trong vũ trụ Star Wars, từ những màu tóc phổ biến đến những màu tóc ít thường thấy. Các màu tóc chiếm tỷ lệ lớn sẽ có cột cao và rõ ràng hơn, dễ thu hút sự chú ý của người xem. Những màu tóc ít phổ biến có thể không được biểu thị rõ ràng trong biểu đồ, nhưng vẫn quan trọng để hiểu sự đa dạng trong thiết kế các nhân vật.

Biểu đồ cung cấp cái nhìn tổng quan về sự đa dạng màu tóc của các nhân vật trong vũ trụ Star Wars, từ các màu tóc phổ biến đến các màu tóc ít thường thấy.

### 3.1.7. Eye\_color

****

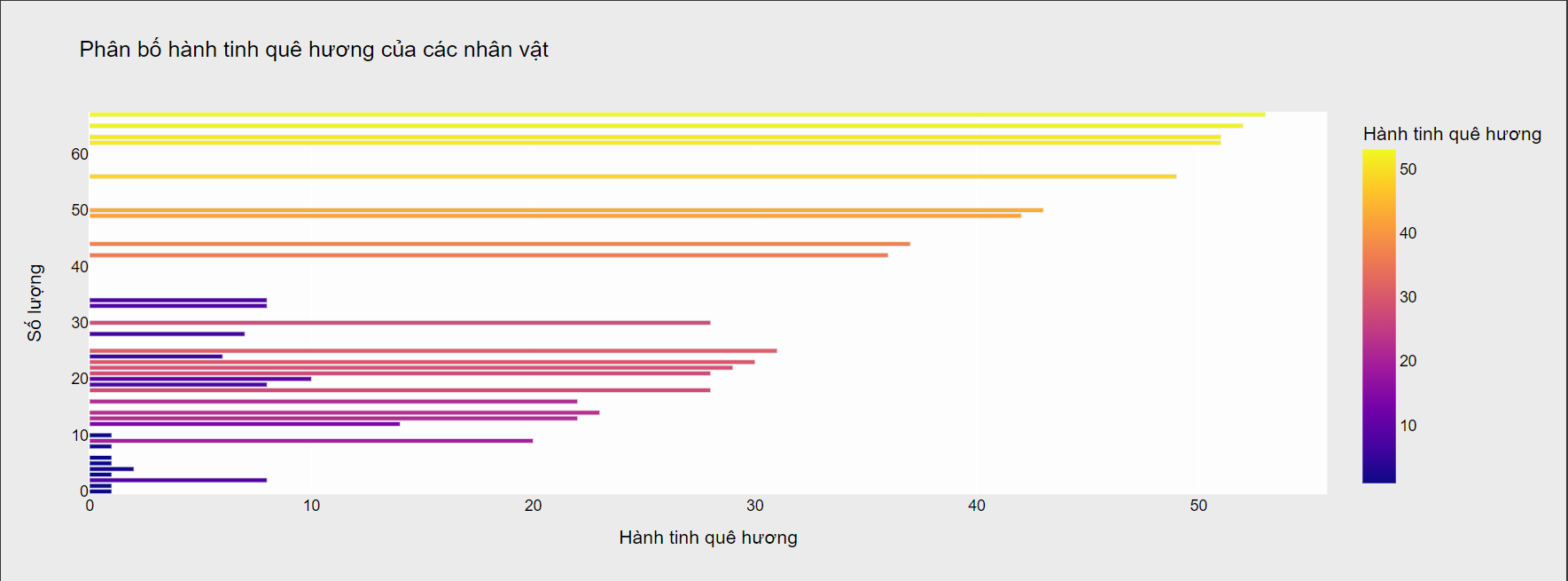
*Hình ảnh 9.Biểu đồ về màu mắt của các nhân vật*

*Dựa vào hình 9 cho thấy được*: Các màu mắt phổ biến sẽ có cột cao hơn và nổi bật hơn trong biểu đồ. Điều này cho thấy sự đa dạng và phổ biến của các màu mắt này trong danh sách các nhân vật. Ví dụ, nếu màu mắt như "nâu" hoặc "xanh" chiếm tỷ lệ lớn, thì các cột tương ứng sẽ cao hơn và có màu sắc rõ ràng hơn. Các màu mắt ít phổ biến sẽ có cột thấp hơn và có thể màu sắc không rõ ràng trong biểu đồ. Điều này cho thấy sự hiếm có và ít được sử dụng của các màu mắt này trong danh sách các nhân vật. Ví dụ, các màu mắt như "đỏ" hoặc "vàng" có thể có cột thấp hơn do số lượng nhân vật thấp.

Biểu đồ thể hiện sự đa dạng màu mắt của các nhân vật trong vũ trụ Star Wars, từ những màu mắt phổ biến đến những màu mắt ít thường thấy. Các màu mắt chiếm tỷ lệ lớn sẽ có cột cao và rõ ràng hơn, dễ thu hút sự chú ý của người xem. Những màu mắt ít phổ biến có thể không được biểu thị rõ ràng trong biểu đồ, nhưng vẫn quan trọng để hiểu sự đa dạng trong thiết kế các nhân vật.

Biểu đồ cung cấp cái nhìn tổng quan về sự đa dạng màu mắt của các nhân vật trong vũ trụ Star Wars, từ các màu mắt phổ biến đến các màu mắt ít thường thấy.

### 3.1.8. Homeworld



*Hình ảnh 10.Biểu đồ về màu tóc của các nhân vật*

*Dựa vào hình 10 cho thấy được*: Các hành tinh có số lượng nhân vật đáng kể sẽ có cột cao và nổi bật hơn trong biểu đồ. Điều này cho thấy sự phổ biến và sự xuất hiện nhiều của nhân vật từ những hành tinh này. Ví dụ, các hành tinh như "Tatooine", "Naboo", hay "Coruscant" có thể có số lượng nhân vật nhiều hơn so với các hành tinh khác. Các hành tinh ít có nhân vật đặc trưng sẽ có cột thấp hơn và ít nổi bật trong biểu đồ. Điều này có thể phản ánh sự hiếm có và ít được sử dụng của những hành tinh này trong vũ trụ Star Wars. Ví dụ, các hành tinh như "Dagobah" hoặc "Kamino" có thể có số lượng nhân vật ít hơn do các nhân vật từ đó ít xuất hiện trong các tác phẩm của Star Wars.

Biểu đồ thể hiện sự đa dạng các hành tinh quê hương của các nhân vật trong vũ trụ Star Wars, từ những hành tinh phổ biến đến những hành tinh ít được biết. Các hành tinh có nhiều nhân vật thường có cột cao và rõ ràng hơn, dễ thu hút sự chú ý của người xem. Những hành tinh ít có nhân vật thường có cột thấp hơn và có màu sắc không rõ ràng hơn, nhưng vẫn quan trọng để hiểu sự đa dạng và chi tiết trong vũ trụ Star Wars.

Biểu đồ cung cấp cái nhìn tổng quan về sự đa dạng các hành tinh quê hương của các nhân vật trong vũ trụ Star Wars.

### 3.3.9. Nhận xét tổng kết

***Kết luận tổng quan***: Dữ liệu và các biểu đồ phân tích về vũ trụ Star Wars cung cấp một cái nhìn sâu sắc về các đặc điểm và đặc tính của các nhân vật trong loạt phim. Dưới đây là những điểm chính từ các phân tích:

* Tần suất xuất hiện của nhân vật:
* Tần suất xuất hiện của các nhân vật trong vũ trụ Star Wars biểu thị sự phổ biến và sự quan trọng của từng nhân vật trong câu chuyện.
* Các nhân vật nổi bật thường có tần suất xuất hiện cao hơn, phản ánh vai trò quan trọng và mức độ hấp dẫn đối với khán giả.
* Phân bố theo chiều cao và cân nặng của nhân vật:
* Chiều cao và khối lượng của các nhân vật được phân bố rộng rãi và phản ánh sự đa dạng trong các loài và chủng tộc trong vũ trụ Star Wars.
* Các biểu đồ phân tích chi tiết giúp hiểu rõ hơn về các đặc điểm vật lý của từng nhân vật và sự khác biệt giữa chúng.
* Phân bố theo giới tính và đặc điểm sinh học:
* Phân bố giới tính, màu da, màu tóc, và màu mắt của các nhân vật cho thấy sự đa dạng trong các đặc điểm nhân cách và bản sắc cá nhân của từng nhân vật.
* Những đặc điểm này không chỉ làm phong phú thêm cho câu chuyện mà còn làm nổi bật sự đa dạng văn hóa và bản sắc trong vũ trụ Star Wars.
* Hành tinh quê hương của nhân vật:
* Phân bố hành tinh quê hương của các nhân vật cho thấy sự đa dạng và phong phú của các địa điểm và môi trường sống trong vũ trụ Star Wars.
* Các hành tinh nổi bật thường có số lượng nhân vật đáng kể, phản ánh vai trò quan trọng của chúng trong câu chuyện và cuộc sống của các nhân vật.

*Tổng kết:*

Dữ liệu và biểu đồ phân tích từ vũ trụ Star Wars giúp hiểu rõ hơn về các nhân vật và các yếu tố ảnh hưởng đến cuộc sống của họ. Những thông tin này không chỉ hỗ trợ trong việc nghiên cứu và phân tích về vũ trụ này mà còn giúp phát triển các sản phẩm, dịch vụ và chính sách phù hợp. Đồng thời, chúng cũng nâng cao nhận thức về sự đa dạng và giá trị của việc khám phá và khai thác các khía cạnh của vũ trụ Star Wars.

## 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 chiều cao, màu da, màu tóc, màu mắt, cân nặng, hành tinh quê hương. Biểu đồ cột giúp so sánh dễ dàng sự khác biệt về tỷ lệ giữa các đối tượng trong bảng.

- Biểu đồ tròn (Pie Chart): Có thể sử dụng để minh họa phân bố tỷ lệ giới tính.

- Biểu đồ đường (Line Chart): Được sử dụng để thể hiện năm sinh của nhân vật.

### 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.

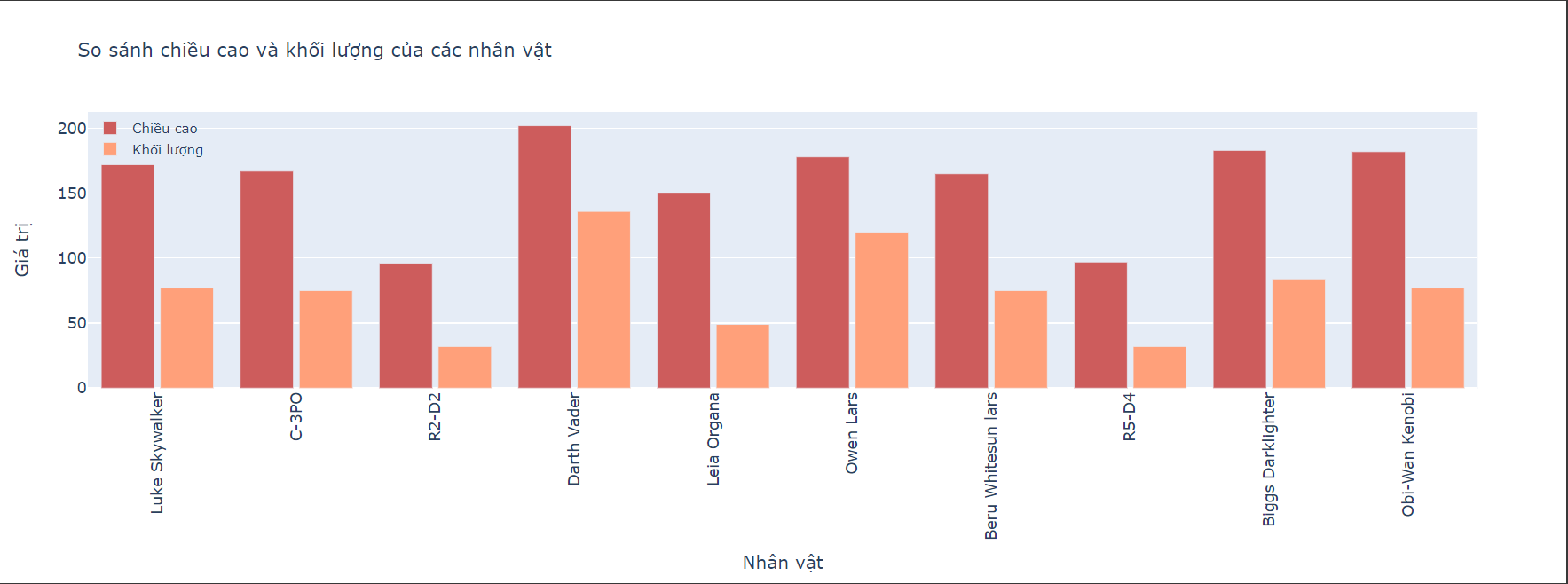
- Tối ưu hóa trực quan: Điều chỉnh màu sắc, kích thước, và bố cục để biểu đồ dễ đọc và hấp dẫn hơn.

- 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 chiều cao và cân nặng của nhân vật(10 đối tượng)



*Hình ảnh 11. Biểu đồ so sánh chiều cao và cân nặng của 10 đối tượng*

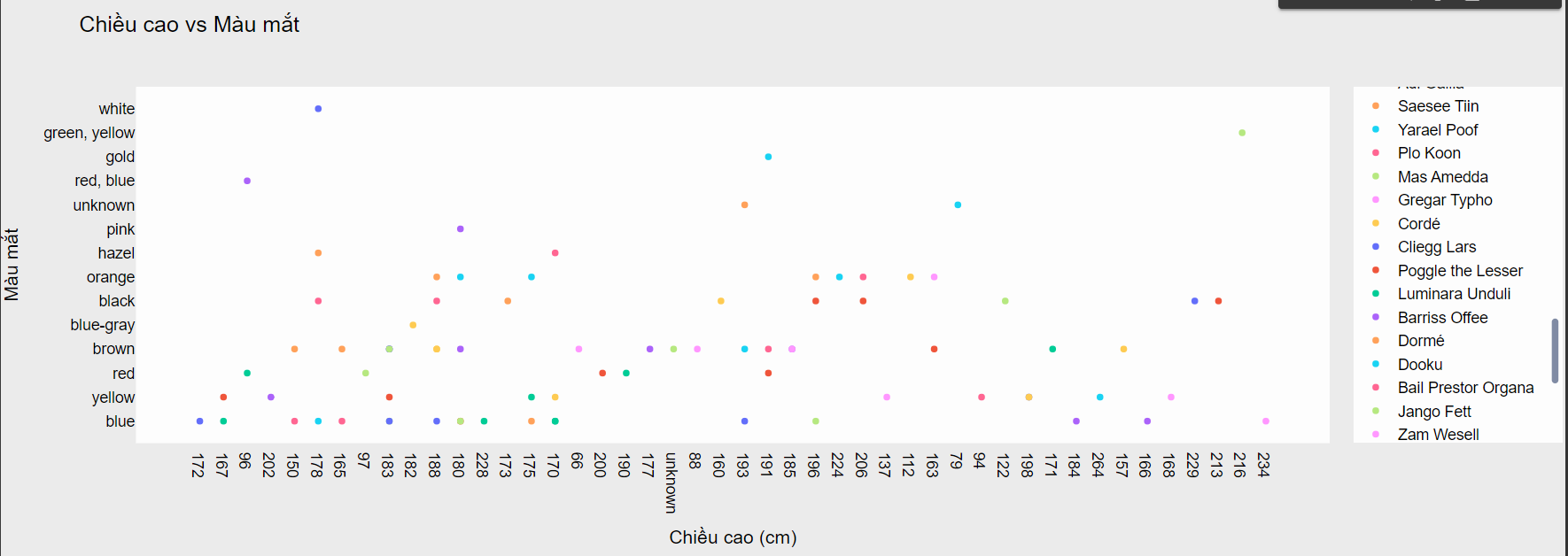
*Dựa vào hình 14 cho thấy được:* Biểu đồ cho thấy sự so sánh giữa chiều cao (tính bằng cm) và khối lượng (tính bằng kg) của các nhân vật trên cùng một đồ thị. Mỗi cột biểu thị cho một nhân vật, với màu sắc khác nhau giúp phân biệt dễ dàng (10 đối tượng ). Các cột có chiều cao và khối lượng cao hơn thường được thể hiện bằng các cột dài hơn, cho thấy sự khác biệt về chiều cao và khối lượng giữa các nhân vật.

*Cụ thể:*

* Các nhân vật có chiều cao và khối lượng lớn: Các nhân vật như Darth Vader và Biggs Darklighter có chiều cao và khối lượng rất lớn, được biểu thị bởi các cột dài. Điều này phản ánh sự mạnh mẽ và thể trạng lớn của họ trong vũ trụ Star Wars.
* Các nhân vật có chiều cao cao nhưng khối lượng lớn: Có một số nhân vật như R5-D4 có chiều cao cao nhưng khối lượng thấp, được thể hiện bởi cột chiều cao dài và cột khối lượng ngắn. Điều này cho thấy những nhân vật này có hình dáng và cân nặng khác biệt so với các nhân vật khác.
* Các nhân vật có chiều cao thấp và khối lượng nhỏ: Một số nhân vật nhỏ như R2-D2 có cả chiều cao và khối lượng thấp, được biểu thị bởi các cột ngắn. Điều này thể hiện sự nhỏ bé và dễ dàng di chuyển của họ trong các tình huống khác nhau.

Biểu đồ này cung cấp thông tin hữu ích về đặc tính vật lý của các nhân vật trong vũ trụ Star Wars, giúp hiểu rõ hơn về sự đa dạng và đặc điểm riêng biệt của từng nhân vật.

**4.1.2. Biểu đồ phân phối chiều cao và màu mắt**



*Hình ảnh 12. Biểu đồ phân phối chiều cao và màu mắt*

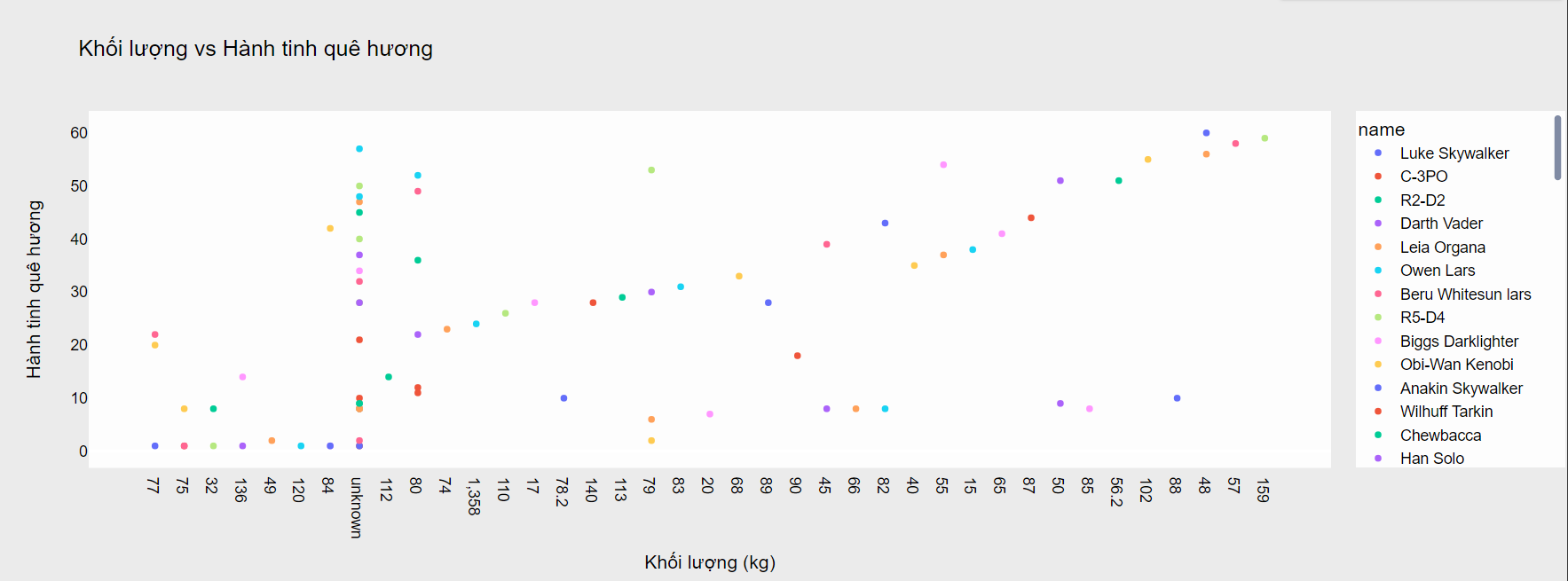
*Dựa vào hình 12 cho thấy được:* Biểu đồ scatter này biểu thị sự phân bố của các nhân vật trong vũ trụ Star Wars dựa trên chiều cao và màu mắt của họ. Mỗi điểm trên biểu đồ biểu thị cho một nhân vật cụ thể, được mã hóa màu sắc để dễ dàng phân biệt. Màu sắc của từng điểm biểu thị cho tên của nhân vật tương ứng, giúp xác định và nhận diện mỗi nhân vật một cách nhanh chóng trên biểu đồ.

*Cụ thể:*

* Mối quan hệ giữa khối lượng và hành tinh quê hương:
* Biểu đồ cho thấy không có mối quan hệ rõ ràng giữa chiều cao và màu mắt của các nhân vật. Điều này cho thấy rằng khối lượng của mỗi nhân vật không phụ thuộc vào hành tinh mà họ đến từ.
* Sự đa dạng về khối lượng và hành tinh quê hương:
* Các nhân vật có chiều cao và màu mắt đa dạng, từ màu xanh da trời, xanh lá cây,.. .
* Điều này phản ánh sự phong phú và đa dạng trong vũ trụ Star Wars, nơi mà các nhân vật có thể đến từ nhiều chủng tộc khác nhau.

Biểu đồ scatter này không chỉ minh họa sự phân bố của chiều cao và màu mắt trong vũ trụ Star Wars mà còn giúp nhìn nhận sự đa dạng và phong phú của các nhân vật trong loạt phim này. Việc hiểu rõ sự phân phối này có thể giúp người hâm mộ và các nhà nghiên cứu tạo ra những câu chuyện phong phú hơn và khám phá thêm về thế giới Star Wars.

**4.1.3. Biểu đồ phân phối khối lượng và hành tinh quê hương**

****

*Hình ảnh 13. Biểu đồ phân phối khối lượng và hành tinh quê hương*

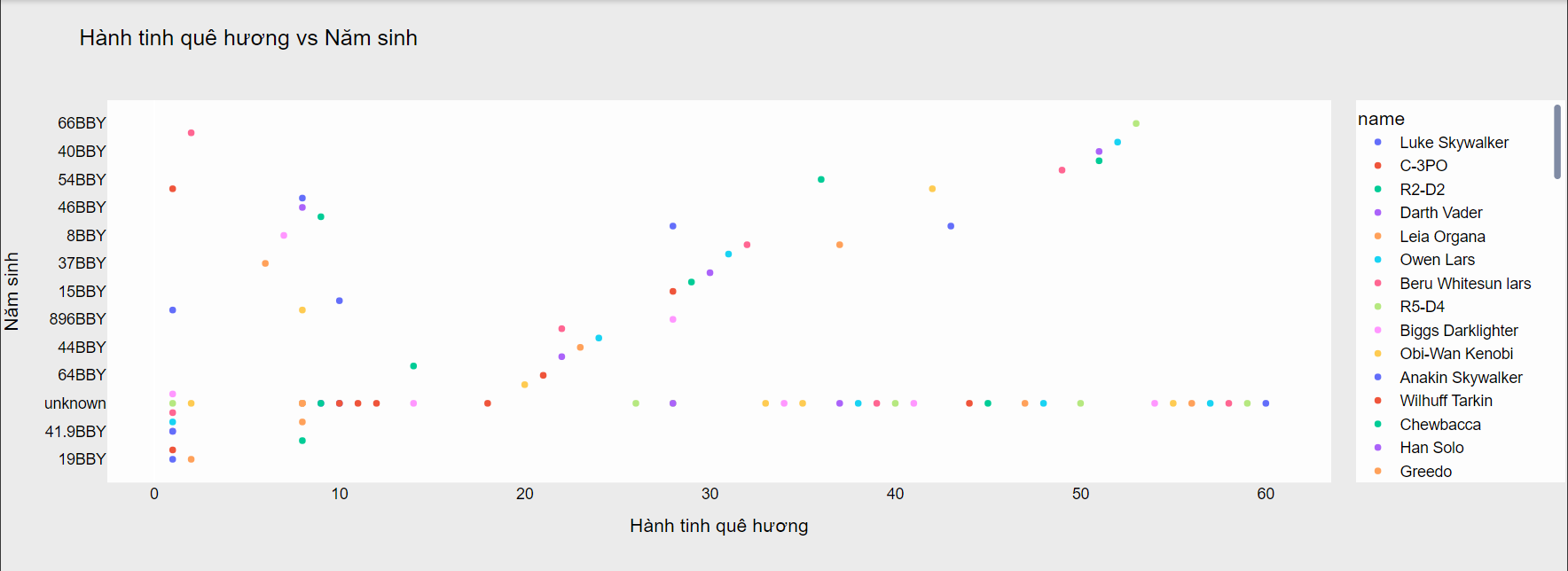
*Dựa vào hình 13 cho thấy được:* Biểu đồ scatter này biểu thị sự phân bố của các nhân vật trong vũ trụ Star Wars dựa trên khối lượng và hành tinh quê hương của họ. Mỗi điểm trên biểu đồ biểu thị cho một nhân vật cụ thể, được mã hóa màu sắc để dễ dàng phân biệt. Màu sắc của từng điểm biểu thị cho tên của nhân vật tương ứng, giúp xác định và nhận diện mỗi nhân vật một cách nhanh chóng trên biểu đồ.

*Cụ thể:*

* Mối quan hệ giữa khối lượng và hành tinh quê hương:
* Biểu đồ cho thấy không có mối quan hệ rõ ràng giữa khối lượng và hành tinh quê hương của các nhân vật. Điều này cho thấy rằng khối lượng của mỗi nhân vật không phụ thuộc vào hành tinh mà họ đến từ.
* Sự đa dạng về khối lượng và hành tinh quê hương:
* Các nhân vật có khối lượng và hành tinh quê hương đa dạng, từ các hành tinh lớn như Earth và Mars đến các hành tinh nhỏ như Pluto và Moon.
* Điều này phản ánh sự phong phú và đa dạng trong vũ trụ Star Wars, nơi mà các nhân vật có thể đến từ nhiều hành tinh khác nhau.

Biểu đồ scatter này không chỉ minh họa sự phân bố của khối lượng và hành tinh quê hương trong vũ trụ Star Wars mà còn giúp nhìn nhận sự đa dạng và phong phú của các nhân vật trong loạt phim này. Việc hiểu rõ sự phân phối này có thể giúp người hâm mộ và các nhà nghiên cứu tạo ra những câu chuyện phong phú hơn và khám phá thêm về thế giới Star Wars.

**4.1.4. Biểu đồ phân phối năm sinh và hành tinh quê hương**

****

*Hình ảnh 14. Biểu đồ phân phối năm sinh và hành tinh quê hương*

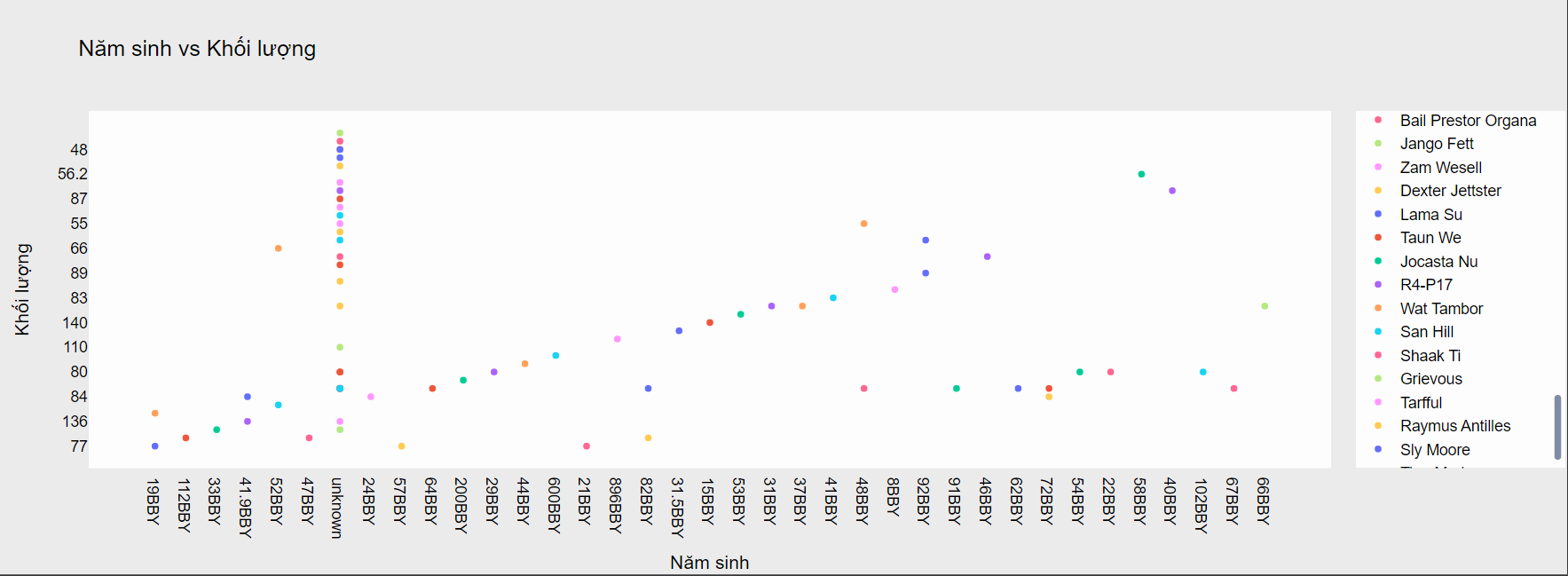
*Dựa vào hình 14 cho thấy được:* Biểu đồ scatter này minh họa sự phân bố của các nhân vật trong vũ trụ Star Wars dựa trên hành tinh quê hương và năm sinh của họ. Mỗi điểm trên biểu đồ biểu thị cho một nhân vật cụ thể, với màu sắc khác nhau giúp phân biệt từng nhân vật.

*Cụ thể:*

* Mối quan hệ giữa hành tinh quê hương và năm sinh:
* Biểu đồ cho thấy sự phân bố của các nhân vật trên các hành tinh khác nhau và năm sinh của họ.
* Không có mối quan hệ rõ ràng giữa hành tinh quê hương và năm sinh của các nhân vật. Điều này cho thấy rằng năm sinh của mỗi nhân vật không phụ thuộc vào hành tinh mà họ đến từ.
* Sự đa dạng về hành tinh quê hương và năm sinh:
* Các nhân vật được sinh ra trên nhiều hành tinh khác nhau, từ những hành tinh như Earth và Mars đến những hành tinh xa xôi trong vũ trụ.
* Năm sinh của các nhân vật cũng đa dạng, phản ánh sự đa dạng và phong phú trong vũ trụ Star Wars.

Biểu đồ scatter này không chỉ giúp hiểu rõ hơn về sự phân bố của các nhân vật dựa trên hành tinh quê hương và năm sinh mà còn thể hiện sự đa dạng và phong phú của vũ trụ Star Wars. Việc nghiên cứu và hiểu rõ sự phân bố này có thể giúp người hâm mộ khám phá và khám phá thêm về các chi tiết thú vị trong thế giới Star Wars.

**4.1.5. Biểu đồ phân phối năm sinh và khối lượng**

****

*Hình ảnh 15. Biểu đồ phân phối năm sinh và khối lượng*

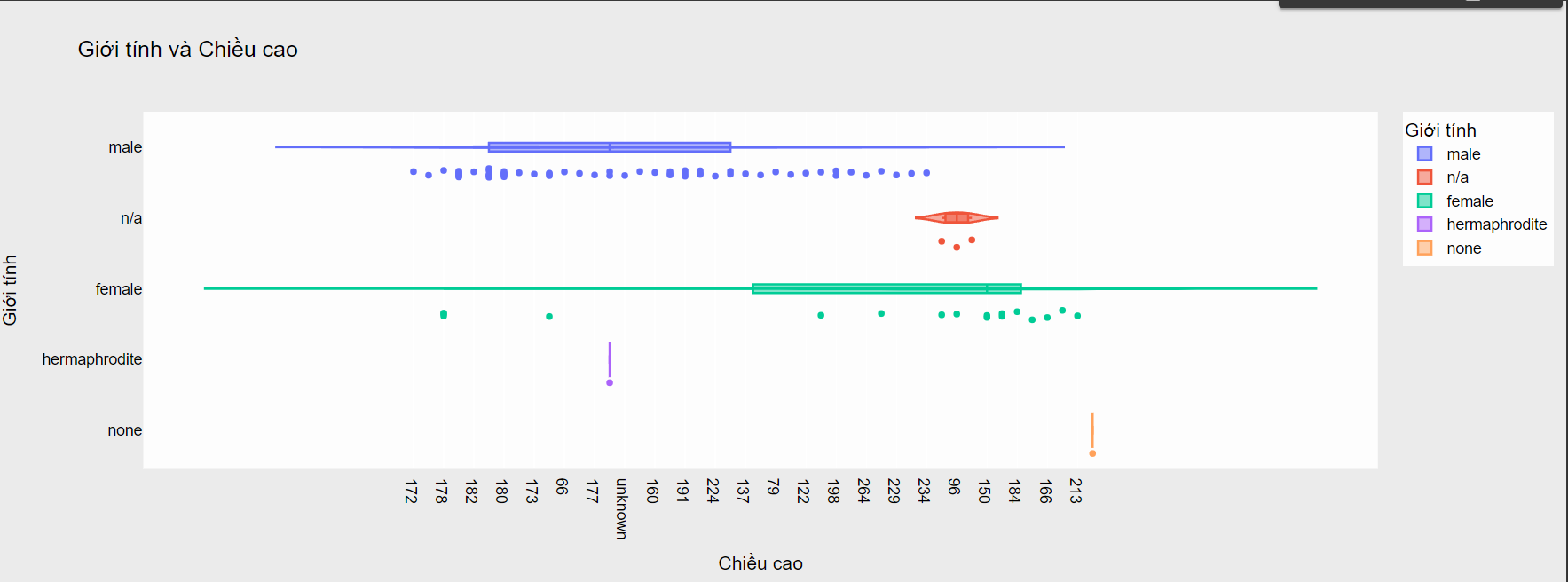
*Dựa vào hình 15 cho thấy được:* Biểu đồ scatter này biểu thị sự phân bố của các nhân vật trong vũ trụ Star Wars dựa trên năm sinh và khối lượng của họ. Mỗi điểm trên biểu đồ biểu thị cho một nhân vật cụ thể, được mã hóa màu sắc để dễ dàng phân biệt.

*Cụ thể:*

* Mối quan hệ giữa năm sinh và khối lượng:
* Biểu đồ cho thấy không có mối quan hệ rõ ràng giữa năm sinh và khối lượng của các nhân vật. Điều này cho thấy rằng khối lượng của mỗi nhân vật không phụ thuộc vào năm sinh của họ.
* Sự đa dạng về năm sinh và khối lượng:
* Các nhân vật có năm sinh và khối lượng đa dạng, từ các giá trị năm sinh và khối lượng khác nhau.
* Điều này phản ánh sự phong phú và đa dạng trong các tính cách và khía cạnh về đặc điểm sinh lý của các nhân vật trong vũ trụ Star Wars.

Biểu đồ scatter này cung cấp cái nhìn tổng quan về sự đa dạng và phân bố của các nhân vật trong loạt phim Star Wars dựa trên năm sinh và khối lượng. Việc phân tích này có thể giúp nhìn nhận sâu hơn về các tính cách và sự khác biệt giữa các nhân vật, từ đó làm giàu thêm hiểu biết về vũ trụ này.

**4.1.6. Biểu đồ phân phối giới tính và chiều cao**

****

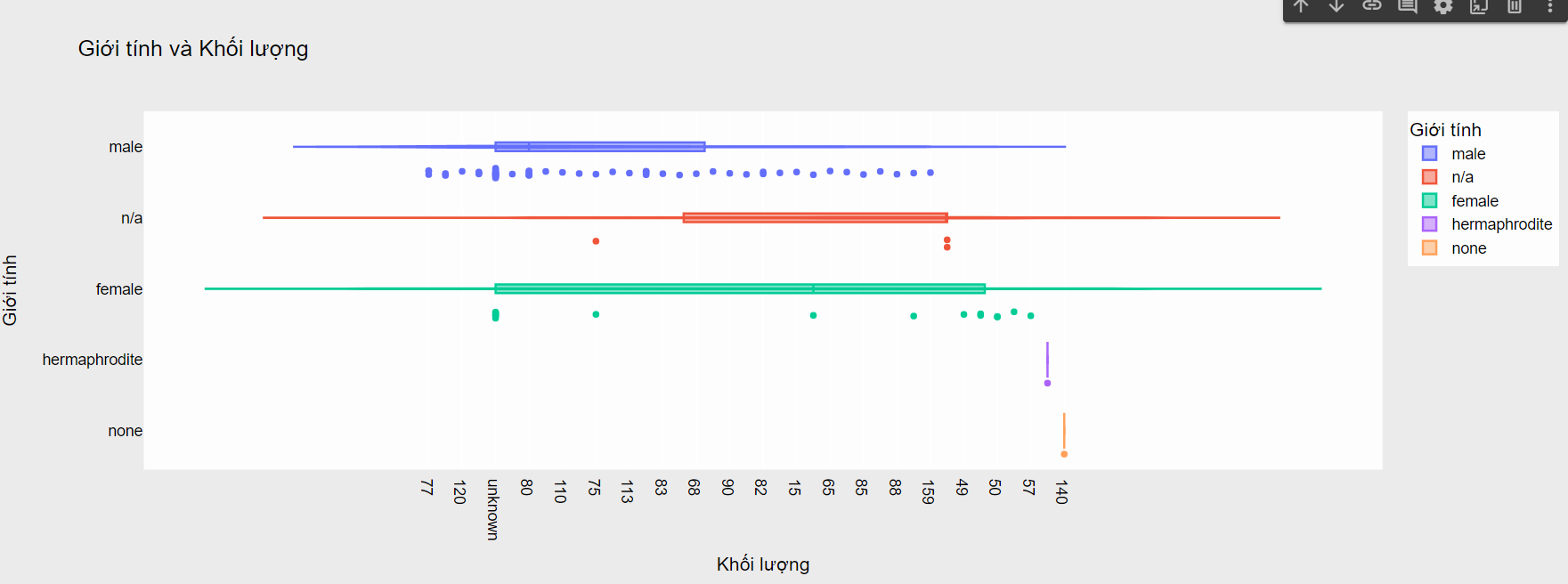
*Hình ảnh 16. Biểu đồ phân phối giới tính và chiều cao*

*Dựa vào hình 16 cho thấy được:* Biểu đồ violin plot này thể hiện phân phối chiều cao của nhân vật trong vũ trụ Star Wars theo từng giới tính. Mỗi violin plot biểu thị phân phối dữ liệu của một giới tính cụ thể trên trục ngang, với độ dày của violin plot thể hiện mật độ phân bố dữ liệu.

*Cụ thể:*

* Phân phối chiều cao theo giới tính:
* Đối với nam giới, phân phối chiều cao tập trung chủ yếu từ khoảng 170cm đến 180cm, với một vài điểm ngoại lệ ở các giá trị cao hơn.
* Đối với nữ giới, phân phối chiều cao có xu hướng tập trung từ khoảng 160cm đến 170cm, với một vài trường hợp ngoại lệ ở các giá trị thấp hơn và cao hơn.
* Đặc điểm phân phối:
* Cả nam giới và nữ giới đều có phân phối chiều cao khá rộng, cho thấy sự đa dạng về chiều cao giữa các nhân vật trong vũ trụ Star Wars.
* Việc sử dụng violin plot giúp hiển thị chi tiết phân bố dữ liệu và so sánh sự khác biệt giữa các nhóm giới tính.

Biểu đồ violin plot là công cụ hữu ích để thể hiện phân phối dữ liệu chiều cao theo giới tính trong vũ trụ Star Wars. Phân tích từ biểu đồ này có thể giúp chúng ta hiểu sâu hơn về đặc điểm về chiều cao của các nhân vật và sự đa dạng giới tính trong loạt phim này.

**4.1.7. Biểu đồ phân phối giới tính và khối lượng**

*Hình ảnh 17. Biểu đồ phân phối giới tính và khối lượng*

*Dựa vào hình 17 cho thấy được:* Biểu đồ violin plot này biểu thị phân phối khối lượng của nhân vật trong vũ trụ Star Wars theo từng giới tính. Mỗi violin plot trên biểu đồ biểu thị phân phối dữ liệu của một giới tính cụ thể trên trục ngang, với độ dày của violin plot thể hiện mật độ phân bố dữ liệu.

*Cụ thể:*

* Phân phối khối lượng theo giới tính:
* Đối với giới tính Nam (Male), phân phối khối lượng có xu hướng tập trung từ khoảng 70kg đến 90kg, với một vài trường hợp ngoại lệ ở các giá trị thấp hơn và cao hơn.
* Đối với giới tính Nữ (Female), phân phối khối lượng tương đối đồng đều, phân bố chủ yếu từ khoảng 50kg đến 70kg.
* Đặc điểm phân phối:
* Mỗi giới tính có đặc trưng phân phối riêng, cho thấy sự đa dạng về khối lượng trong từng nhóm giới tính khác nhau.
* Việc sử dụng violin plot giúp hiển thị chi tiết phân bố dữ liệu và so sánh sự khác biệt giữa các nhóm giới tính.

Biểu đồ violin plot là công cụ hữu ích để thể hiện phân phối dữ liệu khối lượng theo từng giới tính của các nhân vật trong vũ trụ Star Wars. Phân tích từ biểu đồ này có thể giúp chúng ta hiểu sâu hơn về đặc điểm khối lượng và sự đa dạng của các nhân vật theo giới tính trong loạt phim này.

## 4.2. Ý nghĩa

Tổng kết lại, Phân tích các yếu tố liên quan đến các đặc điểm của nhân vật trong vũ trụ Star Wars không chỉ giúp hiểu rõ hơn về các xu hướng và đặc điểm của từng nhân vật mà còn đóng góp vào việc phát triển và mở rộng vũ trụ này. Việc áp dụng các phân tích này sẽ hỗ trợ trong việc tối ưu hóa và phát triển vũ trụ Star Wars một cách bền vững và sáng tạo hơn. Dưới đây là những điểm chính từ các phân tích:

**4.2.1. Những Yếu Tố Đặc Điểm Của Nhân Vật**

Phân phối và đặc điểm: Biểu đồ Scatter Plot, Histogram Chart, Violin Plot, Line Chart cho thấy sự phân bố đặc điểm như chiều cao, cân nặng, màu mắt và năm sinh giữa các nhân vật. Mặc dù có sự biến động, nhưng chúng đều cho thấy xu hướng và sự đa dạng trong các đặc điểm của từng nhân vật.

So sánh trực quan: Các biểu đồ Bar Plot và Pie Chart thể hiện rõ ràng sự khác biệt và phân bố giữa các nhân vật theo từng đặc điểm, từ màu mắt đến hành tinh quê hương, giúp nhìn thấy sự đa dạng và phổ biến trong loạt phim.

**4.2.2. Dự Đoán Và Triển Vọng**

Chiến lược và phát triển: Dựa trên các phân tích về các đặc điểm của nhân vật, có thể phát triển chiến lược và các sản phẩm, dịch vụ mới để phù hợp và mở rộng thế giới Star Wars.

Nghiên cứu và phát triển: Tiếp tục nghiên cứu và phát triển các khía cạnh về các nhân vật và câu chuyện trong vũ trụ, từ đó mở ra những cơ hội mới trong việc phát triển và mở rộng thế giới Star Wars.

**4.2.3. Hợp Tác Và Cộng Đồng**

Xây dựng cộng đồng: Xây dựng các mối quan hệ hợp tác với cộng đồng fan hâm mộ và nhà nghiên cứu để phát triển và mở rộng vũ trụ Star Wars, từ sản phẩm đến trải nghiệm người dùng.

# PHẦN 5. TỔNG KẾT BÀI TẬP LỚN

## 5.1. Kế hoạch

### 5.1.1. Kế hoạch dự kiến

*Giai đoạn 1:*

Lựa chọn ý tưởng: tìm kiếm dữ liệu, phương pháp xử lý dữ liệu,… Cuối cùng nhóm thống nhất đề tài **Star Wars Character Details** với tập dữ liệu và sử dụng các mô hình học máy để đánh giá độ chính xác của dữ liệu sau khi xử lý.

*Giai đoạn 2:*

Code Trực quan hóa dữ liệu. Phân loại. Làm sạch dữ liệu. Phân tích dữ liệu. Phân tích mối tương quan.Lựa chọn tính năng. Viết báo cáo, làm slide.

**5.1.2. Giai đoạn làm việc**



*Hình ảnh 18 . Biểu đồ Gantt*

## 5.2. Kết quả đạt được

Kết quả đạt được sau bài phân tích và trực quan hóa dữ liệu về Star Wars Character Details là:

***Hiểu rõ hơn về phân bố các đặc điểm của nhân vật Star Wars***

Phân tích dữ liệu các nhân vật Star Wars cung cấp cái nhìn tổng quan về sự đa dạng và phong phú của các nhân vật. Các đặc điểm như chiều cao, màu mắt, giới tính được phân tích kỹ lưỡng để hiểu rõ hơn về cách các nhân vật này được thiết kế và phát triển.

Ví dụ, biểu đồ phân bố chiều cao cho thấy sự chênh lệch lớn giữa các nhân vật. Trong khi một số nhân vật như Chewbacca có chiều cao vượt trội, những nhân vật khác như Yoda lại có chiều cao rất khiêm tốn. Tương tự, màu mắt cũng rất đa dạng, từ những màu sắc phổ biến như nâu và xanh đến những màu sắc đặc biệt như đỏ và tím. Phân tích này giúp người xem nhận ra rằng sự đa dạng về đặc điểm nhân vật là một yếu tố quan trọng tạo nên sự hấp dẫn của vũ trụ Star Wars.

***Tìm thấy mối tương quan giữa các đặc điểm nhân vật***

Phân tích tương quan giữa các đặc điểm nhân vật giúp tìm ra những mối liên hệ quan trọng. Chẳng hạn, có một mối tương quan mạnh mẽ giữa chiều cao và cân nặng của các nhân vật, điều này là hợp lý khi xem xét về mặt sinh học. Ngoài ra, còn có những mối liên hệ thú vị giữa màu mắt và chủng tộc của nhân vật, cho thấy cách mà các nhà làm phim sử dụng màu mắt để phân biệt các chủng tộc khác nhau trong vũ trụ Star Wars.

Những phát hiện này không chỉ giúp người xem hiểu rõ hơn về các nhân vật mà còn cung cấp cơ sở cho các nghiên cứu và phát triển nhân vật trong tương lai. Các nhà làm phim có thể sử dụng những thông tin này để tạo ra các nhân vật mới với đặc điểm độc đáo và hấp dẫn hơn.

***Cung cấp thông tin hữu ích cho việc phát triển các nhân vật trong tương lai***

Hiểu rõ hơn về các đặc điểm và mối quan hệ giữa các nhân vật không chỉ giúp các nhà làm phim phát triển những nhân vật mới mà còn giúp họ cải thiện chất lượng nội dung phim. Bằng cách sử dụng các phân tích này, họ có thể tạo ra các nhân vật phù hợp với mong đợi của người hâm mộ, đồng thời đưa ra các cốt truyện và tình tiết phong phú và hấp dẫn hơn.

Những thông tin này cũng có thể được sử dụng trong việc quảng bá và tiếp thị phim. Biết được những đặc điểm và xu hướng nào thu hút khán giả sẽ giúp các nhà sản xuất tối ưu hóa chiến lược tiếp thị, từ đó tăng cường hiệu quả quảng bá và doanh thu của phim.

Tóm lại, bài phân tích này mang lại cái nhìn sâu sắc về các đặc điểm và mối tương quan giữa các nhân vật trong Star Wars, từ đó hỗ trợ trong việc phát triển nhân vật và quản lý nội dung phim. Nó cung cấp những thông tin quan trọng giúp các nhà làm phim và các nhà nghiên cứu hiểu rõ hơn về cách các nhân vật được thiết kế và phát triển, đồng thời đưa ra những gợi ý hữu ích cho việc tạo ra các nhân vật mới và cốt truyện phong phú hơn trong tương lai.

## 5.3. Ưu điểm và hạn chế

### 5.3.1. Ưu điểm

***Trực quan hóa dữ liệu:***

Dễ dàng hiểu và phân tích: Sử dụng biểu đồ và đồ thị giúp người xem dễ dàng nhận ra các mẫu và xu hướng trong dữ liệu. Điều này rất hữu ích cho việc phân tích và đưa ra các kết luận chính xác. Ví dụ, biểu đồ phân bố chiều cao và cân nặng cho phép dễ dàng nhận ra những nhân vật có ngoại hình đặc biệt nổi bật.

Hỗ trợ trình bày: Các biểu đồ và đồ thị giúp truyền đạt thông tin một cách hiệu quả, hỗ trợ trong việc trình bày kết quả phân tích cho các bên liên quan hoặc khán giả.

***Phân tích tương quan:***

Xác định mối liên hệ: Công cụ phân tích tương quan giúp xác định mối liên hệ giữa các biến số, chẳng hạn như giữa chiều cao và cân nặng. Điều này giúp hiểu rõ hơn về cấu trúc và đặc điểm của dữ liệu.

Phát hiện xu hướng: Phân tích tương quan giúp phát hiện các xu hướng và mẫu hình trong dữ liệu, từ đó hỗ trợ trong việc dự đoán và lập kế hoạch.

***Khả năng tổng hợp thông tin:***

Tích hợp từ nhiều nguồn: Phương pháp cho phép tổng hợp thông tin từ nhiều nguồn khác nhau vào một nền tảng duy nhất. Điều này giúp có được cái nhìn toàn diện và đa chiều về dữ liệu.

Hỗ trợ quyết định: Việc tổng hợp thông tin giúp đưa ra các quyết định quản lý và phát triển có căn cứ khoa học và đáng tin cậy.

***Áp dụng linh hoạt***:

Đa lĩnh vực: Công cụ và phương pháp phân tích có thể được áp dụng trong nhiều lĩnh vực khác nhau, từ nghiên cứu xã hội và kinh tế đến phát triển sản phẩm và dịch vụ.

Điều chỉnh dễ dàng: Phương pháp phân tích linh hoạt có thể được điều chỉnh để phù hợp với các nhu cầu và yêu cầu cụ thể của từng lĩnh vực.

### 5.3.2. Hạn chế

***Giới hạn dữ liệu:***

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

Khả năng thu thập dữ liệu: Việc thu thập dữ liệu đầy đủ và chính xác có thể gặp nhiều khó khăn, đặc biệt là khi dữ liệu phân tán và không nhất quán.

***Giới hạn mô hình hóa:***

Khả năng dự đoán hạn chế: Việc mô hình hóa dữ liệu dựa trên các biểu đồ và đồ thị có thể hạn chế 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.

Độ chính xác mô hình: Mô hình hóa dữ liệu có thể không đủ chính xác để áp dụng trong mọi trường hợp, đặc biệt là khi đối diện với các biến đổi và yếu tố bất ngờ.

***Phụ thuộc vào người phân tích:***

Kỹ năng và kinh nghiệm: 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.

Sai sót con người: Dù công cụ phân tích mạnh mẽ, nhưng sai sót trong quá trình nhập liệu và xử lý dữ liệu do con người vẫn có thể xảy ra.

***Cần sự hiểu biết chuyên môn:***

Yêu cầu kiến thức chuyên sâu: Việc sử dụng phương pháp và công cụ phân tích yêu cầu người phân tích có kiến thức vững về thống kê, phân tích dữ liệu và đọc hiểu các biểu đồ, đồ thị một cách chuyên sâu.

Đào tạo và hướng dẫn: Cần có sự đào tạo và hướng dẫn cho những người mới bắt đầu hoặc không có nền tảng vững về phân tích dữ liệu để họ có thể sử dụng hiệu quả các công cụ và phương pháp này.

# 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>

[2]: Link Colab:

<https://colab.research.google.com/drive/1ssMzZR40sCmhw0ePY0JlZb7aVy1x4SZz?usp=sharing>

[3]: Data: ”Star Wars Character Details”

<https://github.com/Ngolinh141/DataVisualization_N07_Group9/blob/main/people.json>

[4]: Link Github:

<https://github.com/Ngolinh141/DataVisualization_N07_Group9>