**Bộ giáo dục và đào tạo**

**Trường Đại học Ngoại ngữ - Tin học TP.HCM**



**Môn học: Phân tích dữ liệu**

**ĐỀ TÀI : Phân tích doanh số bán hàng của video game**

GVHD: Th.s Vũ Đình Ái

HVTH: 21DH113129 Cao Thảo Vân

21DH113218 Bùi Tuấn Đạt

21DH110660 Nguyễn Hoàng Gia Huy

**Tháng 03 Năm 2024**

**Mục lục**

[Danh mục hình ảnh ii](#_Toc162299594)

[I. Giới thiệu 1](#_Toc162299595)

[II. Nội dung 1](#_Toc162299596)

[III. Phương pháp phân tích 2](#_Toc162299597)

[IV. Kết quả phân tích 2](#_Toc162299598)

[1. Thể loại game nào được làm nhiều nhất 2](#_Toc162299599)

[2. Năm nào có nhiều game được phát hành nhất 3](#_Toc162299600)

[**4.** **Năm nào có doanh thu cao nhất trên toàn thế giới** 4](#_Toc162299601)

[**5.** **Game thuộc thể loại nào được phát hành nhiều nhất trong một năm** 5](#_Toc162299602)

[**6.** **Trò chơi thể loại nào được bán nhiều nhất trong một năm** 6](#_Toc162299603)

[**7.** **Thể loại game nào có giá bán cao nhất toàn cầu** 7](#_Toc162299604)

[**8.** **Nền tảng nào có doanh số cao nhất toàn cầu** 7](#_Toc162299605)

[**9.** **So sánh doanh số theo thể loại của các khu vực** 8](#_Toc162299606)

[**10.** **So sánh doanh số theo nền tảng của các khu vực** 9](#_Toc162299607)

[**11.** **Top 20 Nhà phát hành** 9](#_Toc162299608)

[**12.** **Doanh thu cao nhất toàn cầu theo nhà phát hành** 10](#_Toc162299609)

[**13.** **Thị phần của các nhà phát hành** 11](#_Toc162299610)

[**14.** **Tổng doanh thu theo khu vực** 12](#_Toc162299611)

[V. Kết luận 13](#_Toc162299612)

[Tài liệu tham khảo (IEEE) 14](#_Toc162299613)

# Danh mục hình ảnh

[Hình 1 Quy trình phân tích dữ liệu. 2](#_Toc162299494)

[Hình 2. Những thể loại được làm nhiều nhất 3](#_Toc162299495)

[Hình 3. Những năm phát hành game nhiều nhất. 3](#_Toc162299496)

[Hình 4. Top game phát hành trong 5 năm theo thể loại. 4](#_Toc162299497)

[Hình 5. Những năm có doanh thu cao nhất. 5](#_Toc162299498)

[Hình 6. Thể loại game được phát hành nhiều nhất qua từng năm. 6](#_Toc162299499)

[Hình 7. Thể loại game được bán nhiều nhất qua từng năm. 6](#_Toc162299500)

[Hình 8. Thể loại game có giá bán cao nhất. 7](#_Toc162299501)

[Hình 9. Nền tảng có doanh số cao nhất. 8](#_Toc162299502)

[Hình 10. Biểu đồ heatmap về doanh số theo thể loại của các khu vực. 8](#_Toc162299503)

[Hình 11. Doanh số theo thể loại của các khu vực. 9](#_Toc162299504)

[Hình 12. Doanh số các nền tảng của các khu vực. 9](#_Toc162299505)

[Hình 13. Top 20 nhà phát hành nhiều game nhất. 10](#_Toc162299506)

[Hình 14. Các nhà phát hành có doanh thu cao nhất 10](#_Toc162299507)

[Hình 15. Thị phần của các nhà phát triển theo khu vực. 11](#_Toc162299508)

[Hình 16. Doanh số theo khu vực 12](#_Toc162299509)

[Hình 17. Doanh số theo khu vực 12](#_Toc162299510)

# Giới thiệu

Phần trình bày sẽ giới thiệu về bộ dữ liệu phân tích:

Video Games Sales Analysis And Visualization là bộ dữ liệu được thu thập, xử lý bởi Kaggle, nội dung các phiếu khảo sát là các khía cạnh xung quanh lĩnh vực Machine Learning & Data Science. Đề tài này sẽ tiến hành phân tích các khía cạnh đó để có những cái nhìn sâu sắc hơn về lĩnh vực này đồng thời đưa ra những định hướng học tập và nghiên cứu cho người mới bước chân vào lĩnh vực này.

Để thực hiện đề tài này nhóm đã sử dụng một số công cụ hỗ trợ xử lý dữ liệu như Pandas, Numpy; công cụ trực quan như Matplotlib, Seaborn; đồng thời kết hợp suy luận từ vốn hiểu biết và các khía cạnh dữ liệu liên quan trong bộ dữ liệu để trả lời các câu hỏi đặt ra từ kết quả trực quan hoặc dẫn dắt phân tích đi đúng hướng.

Thông qua thực hiện đề tài này nhóm đã học được rất nhiều điều đặc biệt là tư duy đặt, giải quyết vấn đề thông qua dữ liệu và có thêm một số hiểu biết nhất định về lĩnh vực mình đang theo học.

# Nội dung

Phần trình bày chi tiết các thông tin của của bộ dữ liệu :

* Tên bộ dữ liệu: Video Games Sales Analysis And Visualization.
* Nguồn dữ liệu: Kaggle.
* Cấu trúc gồm: 11 cột là chi tiết cho 14 câu hỏi nhóm tự đặt để đưa ra kết luận tổng quan cho quá trình phân tích dữ liệu bán hàng, 16598 dòng dữ liệu

|  |  |  |
| --- | --- | --- |
| Câu hỏi số (Số cột) | Cột tương ứng trên dataset | Nội dung |
| 0 | Rank | Xếp hạng tổng doanh thu. |
| 1 | Name | Tên trò chơi. |
| 2 | Platform | Nền tảng phát hành trò chơi (ví dụ: PC, PS4, v.v.). |
| 3 | Year | Năm phát hành trò chơi. |
| 4 | Genre | Thể loại trò chơi. |
| 5 | Publisher | Nhà phát hành trò chơi. |
| 6 | NA\_Sales | Doanh số bán hàng ở Bắc Mỹ (tính bằng triệu). |
| 7 | EU\_Sales | Doanh số bán hàng ở Châu Âu (tính bằng triệu). |
| 8 | JP\_Sales | Doanh số bán hàng ở Nhật Bản (tính bằng triệu). |
| 9 | Other\_Sales | Doanh số bán hàng ở phần còn lại của thế giới (tính bằng triệu). |
| 10 | Global\_Sales | Tổng doanh số bán hàng trên toàn thế giới. |

# Phương pháp phân tích

Phần trình bày quy trình phân tích:

Diagram

Description automatically generated

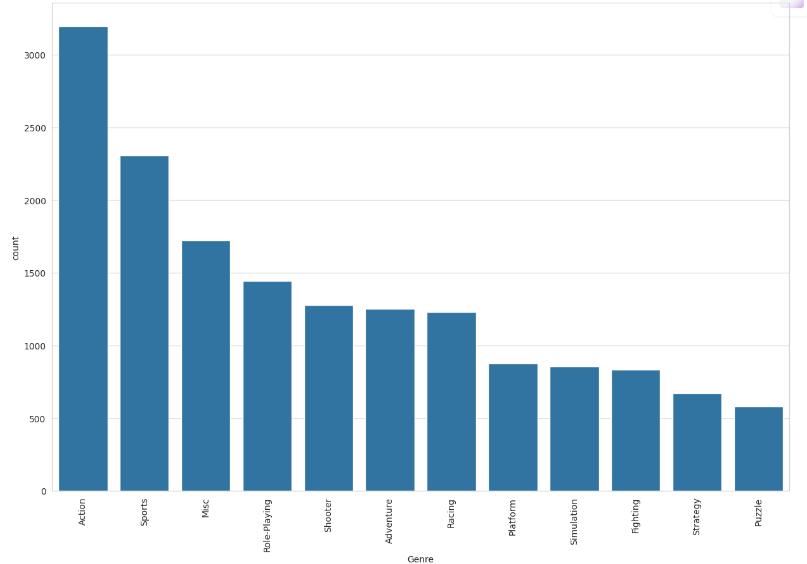
Hình 1 Quy trình phân tích dữ liệu.

Mô tả: Nhóm bắt đầu với việc chọn bộ dữ liệu, kiểm tra các thông tin cơ bản của bộ dữ liệu để đặt ra những vấn đề cơ bản và chung nhất, sau đó tiến hành phân tích thăm dò để giải quyết các vấn đề cơ bản này. Sau khi giải quyết xong nhóm đã hiểu hơn về bộ dữ liệu và phát hiện ra những vấn đề sâu sắc hơn mà bộ dữ liệu mang lại, nhóm tiếp tục tiến hành phân tích thăm dò để giải quyết những vấn đề mới. Cuối cùng nhóm sẽ tổng hợp những kết quả đã phân tích được.

# Kết quả phân tích

Phần trình bày các kết quả phân tích (trực quan dữ liệu, trả lời các câu hỏi liên quan đến bộ dữ liệu đang phân tích

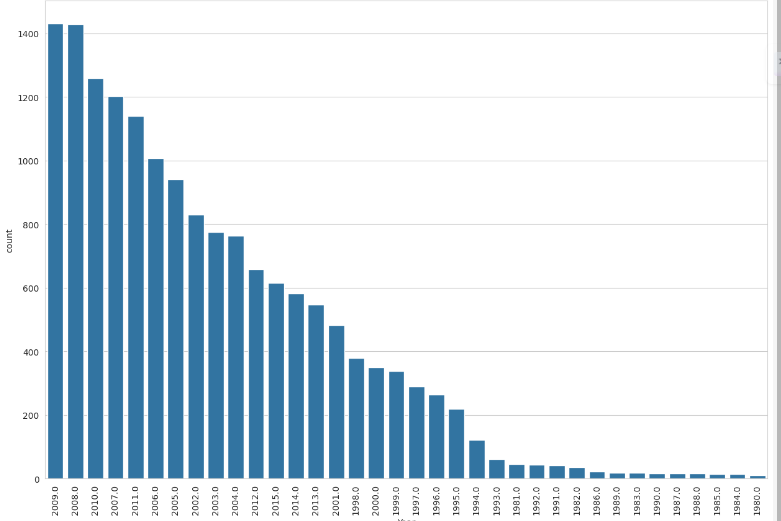
## Thể loại game nào được làm nhiều nhất



Hình 2. Những thể loại được làm nhiều nhất

Những thể loại như Action, Sports, Misc được làm nhiều nhất với số lượng rất lớn, chiếm ưu thế hơn hẳn những cái tên còn lại. Qua đó cũng biết được sở thích của khách hàng.(có 3196 games Action, 2308 games Sports và 1721 games Misc được phát hành)

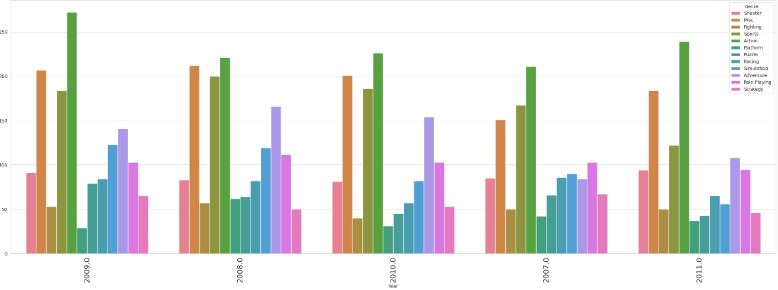
## Năm nào có nhiều game được phát hành nhất



Hình 3. Những năm phát hành game nhiều nhất.

Các năm 2009, 2008, 2010, 2007 và 2011 là những năm có nhiều tựa game được phát hành nhất. Đây cũng có thể là những năm bùng nổ và phát triển của ngành công nghiệp game vì cho ra nhiều tựa game nhất. Cụ thể:

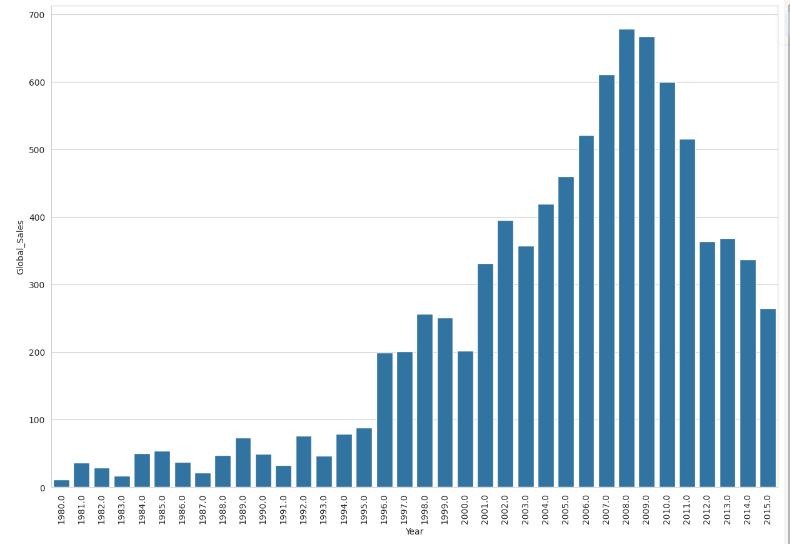
1. 2009.0 -> 1431
2. 2008.0 -> 1428
3. 2010.0 -> 1259
4. 2007.0 -> 1202
5. 2011.0 -> 1139
6. **Top game phát hành trong 5 năm theo thể loại.**



Hình 4. Top game phát hành trong 5 năm theo thể loại.

Như ta có thể thấy theo hình trên thể loại action luôn đứng top đầu trong các thể loại của năm.

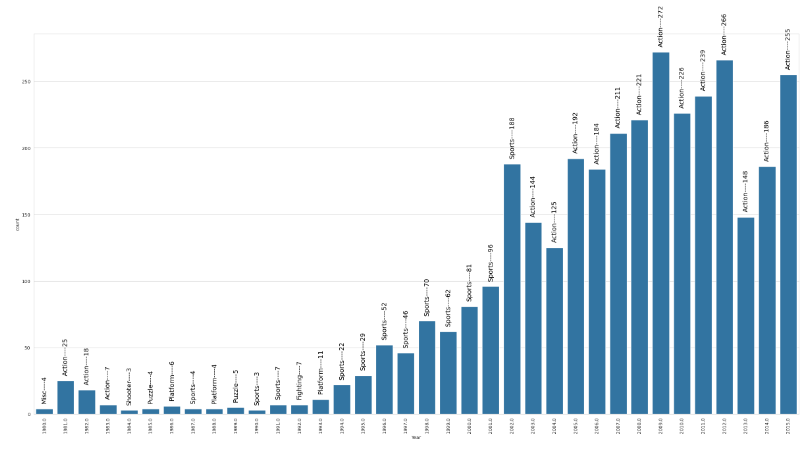
1. **Năm nào có doanh thu cao nhất trên toàn thế giới**



Hình 5. Những năm có doanh thu cao nhất.

Ta có thể thấy tuy 2006 không nằm trong top 5 năm có nhiều game được phát hành nhưng lại nằm trong top 5 năm có doanh thu cao nhất trên toàn thế giới. Cụ thể:

1. 2008 678.90
2. 2009 667.30
3. 2007 611.13
4. 2010 600.45
5. 2006 521.04
6. **Game thuộc thể loại nào được phát hành nhiều nhất trong một năm**

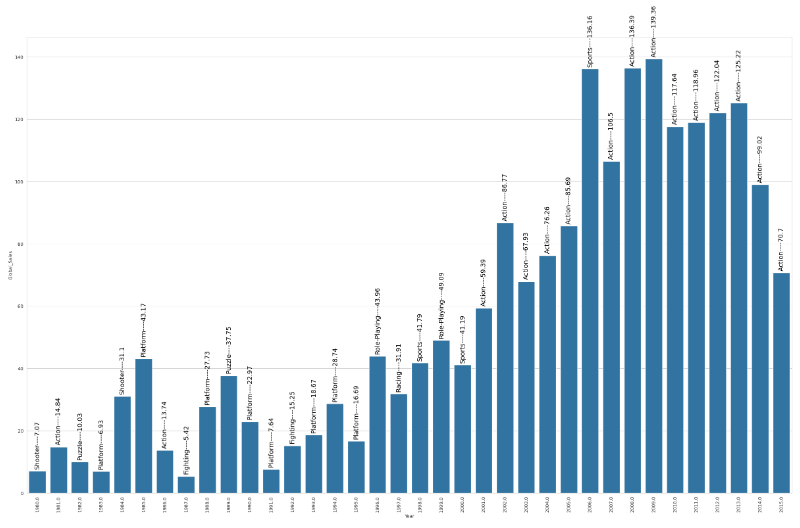


Hình 6. Thể loại game được phát hành nhiều nhất qua từng năm.

Thể loại game Action vẫn chiếm ưu thế và có kết quả tốt nhất là năm 2009 và đổ về những năm sau ta thấy có sự chững lại rồi có sụt giảm. Nhưng vẫn là thể loại được ưa thích nhất. Sau đó là Sport cũng từng là thể loại được rất ưa chuộng. Cụ thể:

2009 Action ---> 272 và 2012 Action ---> 266 games được phát hành.

1. **Trò chơi thể loại nào được bán nhiều nhất trong một năm**

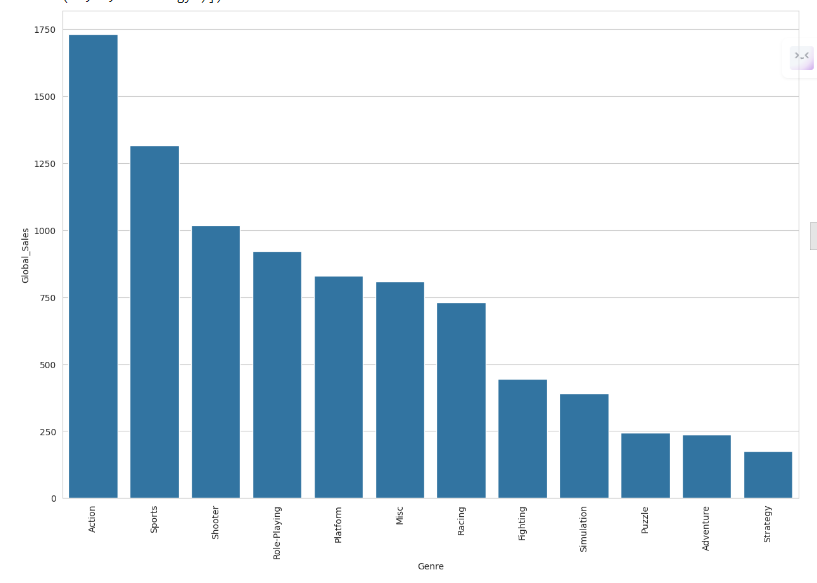


Hình 7. Thể loại game được bán nhiều nhất qua từng năm.

Thể loại game Action không tự nhiên được phát hành nhiều như thế. Ta thấy biểu đồ này thấy được lượt bán nhiều nhất vẫn là Action thể hiện thị yếu mua game của người dùng ảnh hưởng tới lối đi của nhà sản xuất. Cụ thể theo doanh số trên toàn thế giới:

2009 Action ---> 139.36 million và 2008 Action ---> 136.39 miliion

1. **Thể loại game nào có giá bán cao nhất toàn cầu**



Hình 8. Thể loại game có giá bán cao nhất.

Action và Sports vẫn luôn đứng đầu sau đó là Shooter. Vì nguồn cầu lớn về những thể loại trên dẫn đến các nhà sản xuất tập trung phát triển cho ra những sản phẩm tốt nhất để cạnh tranh. Qua đó tăng giá bán sản phẩm. Cụ thể:

Action 🡪1731.26 triệu

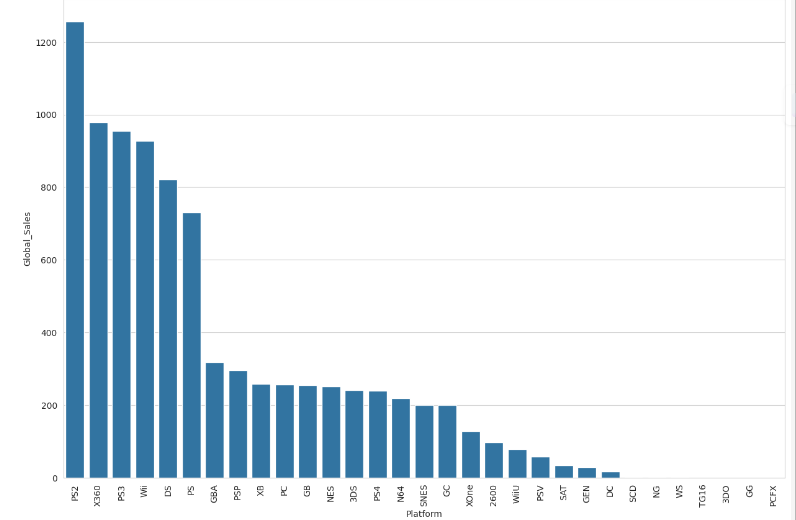
Sports 🡪1316.33 triệu

Shooter 🡪 1019.15 triệu

Role-Playing 🡪 920.57 triệu

Plattform 🡪 829.30 triệu

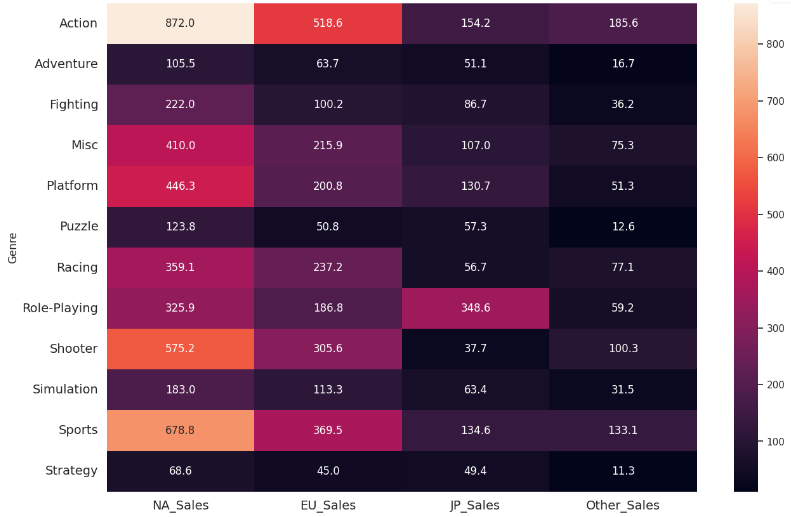
1. **Nền tảng nào có doanh số cao nhất toàn cầu**



Hình 9. Nền tảng có doanh số cao nhất.

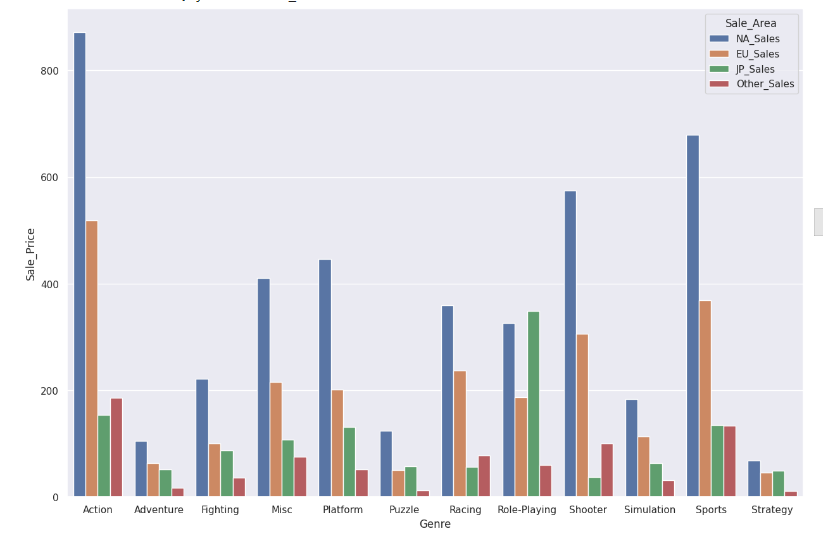
Những nền tảng doanh số cao nhất là PS2, X360, PS3. Và các nền tảng này là những nền tảng trong những năm thị trường game phát triển nhất. Nên doanh số rất cao.

1. **So sánh doanh số theo thể loại của các khu vực**



Hình 10. Biểu đồ heatmap về doanh số theo thể loại của các khu vực.

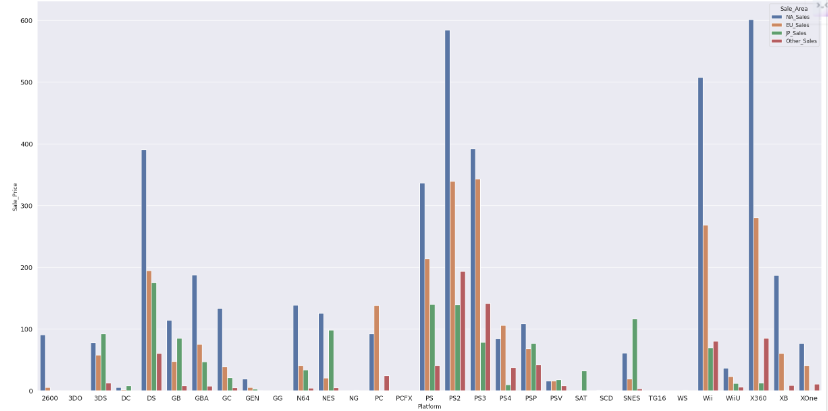
Action, Sport là những thể loại rất được ưa chuộng ở hầu hết các khu vực và có doanh số cao nhất. Thì ở Nhật bản ( JP\_sales ) lại ưa chuộng thể loại Role-Playing hơn và doanh số Role-Playing cũng là cao nhất (348.64 triệu).



Hình 11. Doanh số theo thể loại của các khu vực.

Action, sport và Shooter có doanh thu cao hơn so với những thể loại khác và khu vực Bắc Mỹ (NA\_sales) luôn có doanh số bán hàng cao nhất sau đó là Châu Âu (EU\_sales). Chứng tỏ họ rất yêu thích trò chơi điện tử.

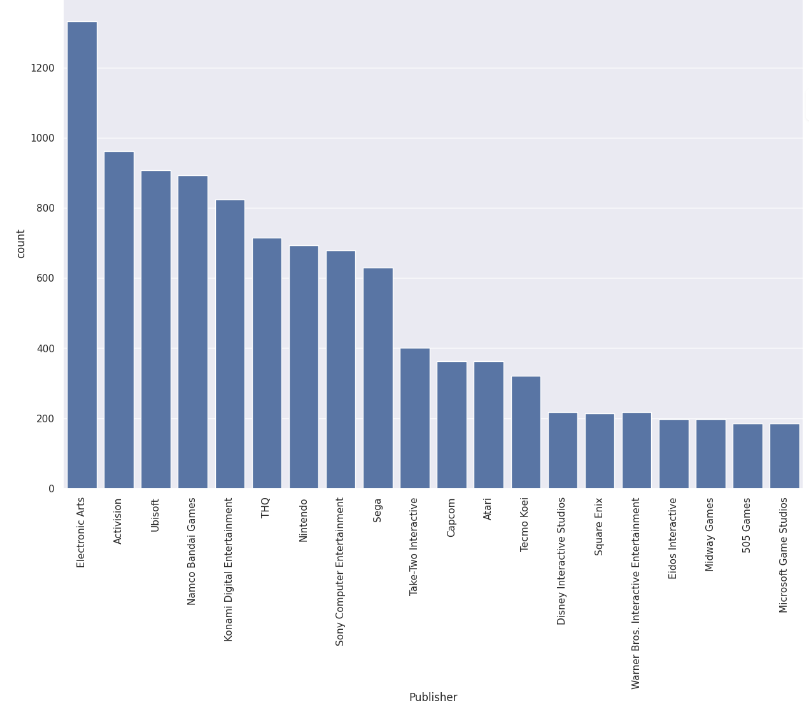
1. **So sánh doanh số theo nền tảng của các khu vực**



Hình 12. Doanh số các nền tảng của các khu vực.

X360, PS2 và Wii đang dẫn đầu thị trường và Bắc Mỹ luôn đứng đầu. Thị trường Bắc Mỹ là thị trường lớn của các nhà sản xuất trò chơi điện tử.

1. **Top 20 Nhà phát hành**



Hình 13. Top 20 nhà phát hành nhiều game nhất.

Ta có EA ( Electronic Arts) là nhà phát hành nhiều game nhất. Sau đó là những cái tên như Activison, Ubisoft.

Electronic Arts 🡪 1319

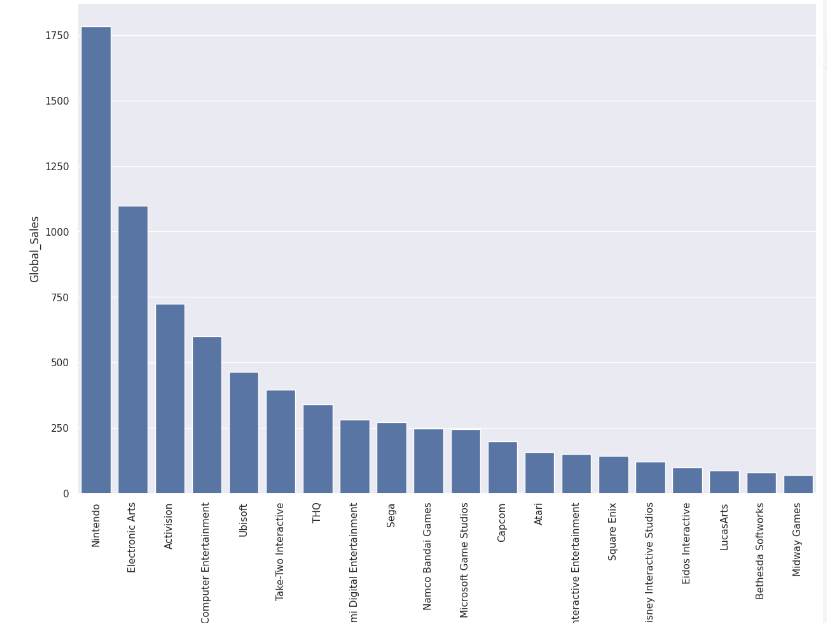
Activision 🡪 952

Ubisoft 🡪 904

Namco Bandai games 🡪 888

Konami Digital Entertainment 🡪 815

1. **Doanh thu cao nhất toàn cầu theo nhà phát hành**



Hình 14. Các nhà phát hành có doanh thu cao nhất

Nintendo kiếm được nhiều tiền hơn so với các nhà phát hành khác, nhưng họ không phát hành nhiều trò chơi nhất. Họ cũng không nằm trong top 5 nhà xuất bản. Chứng tỏ game của Nitendo rất được ưa chuộng và hợp thị yếu người dùng.

Nintendo 🡪 1783.09 triệu

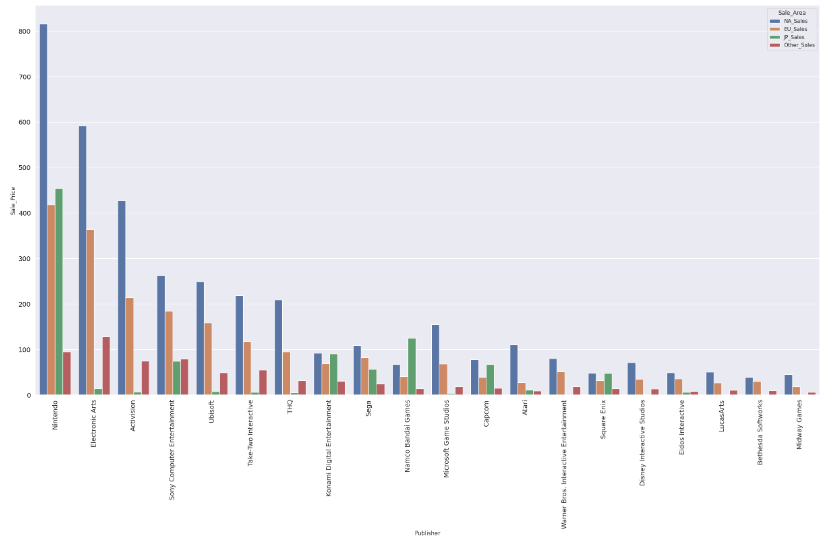
Electronic Arts 🡪 1098.07 triệu

Activision 🡪 723.79 triệu

Sony Computer Entertainment 🡪 600.94 triệu

Ubisoft 🡪 464.67 triệu

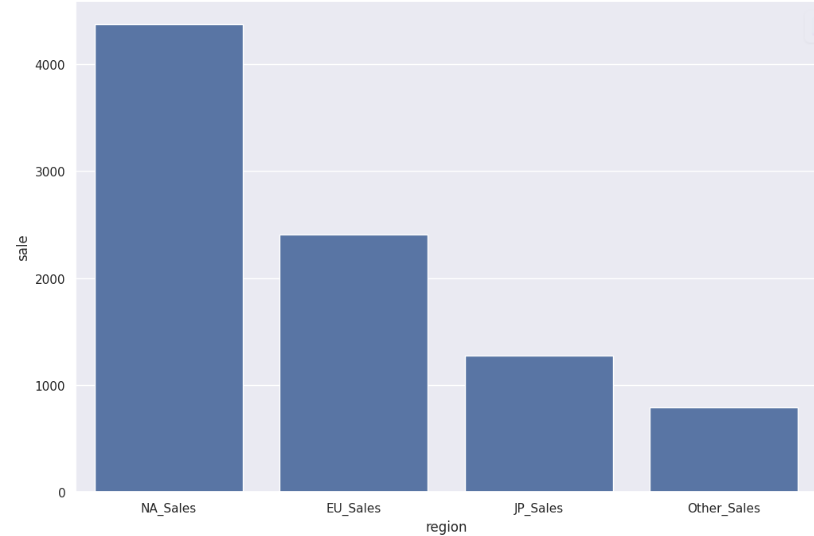
1. **Thị phần của các nhà phát hành**



Hình 15. Thị phần của các nhà phát triển theo khu vực.

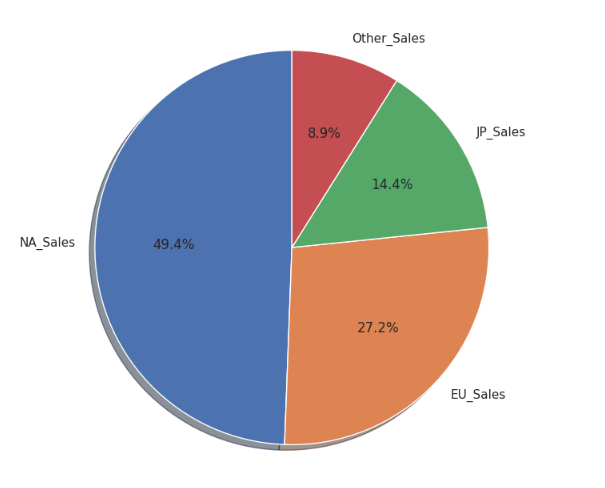
Nintendo có thị phần cao nhất (1783.09) và hơn 0,25 lầm Electronic Arts ở vị trí thứ 2 (1098.07). Sau đó là những Activison (723.79), Sony (600.94), Ubisoft (464.67)… . Nintendo thể hiện được sự lớn mạnh của nhà phát hành này so với phần còn lại.

1. **Tổng doanh thu theo khu vực**



Hình 16. Doanh số theo khu vực

Khu vực Bắc Mỹ là khu vực cực kỳ quang trọng. Chiếm doanh số rất lớn của nền công nghiệp game



Hình 17. Doanh số theo khu vực

Ta thấy được Bắc Mỹ chiếm tới gần 50% doanh thu.

Tổng doanh thu theo từng khu vực:

NA\_Sales 🡪 4370.02

EU\_Sales 🡪 2407.37

JP\_Sales 🡪 1277.27

Other\_Sales 🡪 789.98

# Kết luận

Trong suốt quá trình làm việc, nhóm đã hoàn tất việc EDA trên tất cả các thuộc tính của bộ dữ liệu, tìm hiểu và phân tích sâu hơn nhiều vấn đề được đặt ra sau khi EDA cơ bản bộ dữ liệu. Trên đây là những kết quả được chọn lọc từ những phân tích của nhóm, toàn bộ các phân tích sẽ được trình bày trong source code. Dữ liệu phân tích cung cấp cái nhìn tổng quát về thị trường game theo thể loại, nền tảng, khu vực, và nhà phát hành.

Thông qua quá trình phân tích nhóm đã trả lời được rất nhiều câu hỏi liên quan về lĩnh vực bán hàng cụ thể là video games thông qua bộ dữ liệu nhóm thu thập được, đã dựa trên các biểu đồ và đánh giá theo các tiêu chí bán hàng, thị yếu người dùng. Dựa vào việc phân tích trên chúng em đưa ra những hướng phát triển cho doanh nghiệp như sau:

* Tập trung vào phát triển các tựa game thuộc thể loại Action và Sports, cùng với việc điều chỉnh sản phẩm để phù hợp với nền tảng PS2, X360, và PS3.
* Tận dụng thị trường Bắc Mỹ như một trọng tâm chiến lược doanh nghiệp vì thị trường này chiếm tới gần 50% doanh thu toàn cầu.
* Nghiên cứu thị trường Nhật Bản để tận dụng sở thích ưa chuộng thể loại Role-Playing ở khu vực này.
* Tăng cường hợp tác với các nhà phát hành lớn như Electronic Arts, Activision, và Ubisoft để tận dụng thị phần và doanh số cao.

Tổng quan, nhận diện kỹ lưỡng về sở thích của người chơi và điều chỉnh phản ánh của sản phẩm dựa trên dữ liệu thị trường sẽ giúp doanh nghiệp phát triển mạnh mẽ và cạnh tranh trong ngành công nghiệp game.

# Tài liệu tham khảo

# Quá trình phân tích: <https://www.youtube.com/watch?v=VH2JgqlN2so&t=602s>

1. Quá trình phân tích: <https://www.youtube.com/watch?v=tAo2foiBt4I>
2. Dữ liệu thu thập: <https://www.kaggle.com/code/snanilim/video-games-sales-analysis-and-visualization/input>