

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN



TRẦN QUANG QUÂN - 520H0569
NGUYỄN PHÚC MINH ĐĂNG – 521H0497

UBISOFT ENTERTAINMENT SA

BÁO CÁO GIỮA KỲ
HỆ THỐNG THƯƠNG MẠI THÔNG
MINH

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN



TRẦN QUANG QUÂN – 520H0569
NGUYỄN PHÚC MINH ĐĂNG – 521H0497

UBISOFT ENTERTAINMENT SA

BÁO CÁO GIỮA KỲ
HỆ THỐNG THƯƠNG MẠI THÔNG
MINH

Người hướng dẫn
ThS. Võ Thị Kim Anh

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025

LỜI CẢM ƠN

Chúng em xin chân thành cảm ơn cô Võ Thị Kim Anh đã giảng dạy và hướng dẫn các kiến thức chuyên môn Hệ thống thương mại thông minh để chúng em có thể hoàn thành bài báo cáo này.

TP. Hồ Chí Minh, ngày 12 tháng 11 năm 2025

Tác giả

(Ký tên và ghi rõ họ tên)

Quân

Trần Quang Quân

Đăng

Nguyễn Phúc Minh Đăng

CÔNG TRÌNH ĐƯỢC HOÀN THÀNH

TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của ThS. Võ Thị Kim Anh. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Bài báo cáo còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày 12 tháng 11 năm 2025

Tác giả

(Ký tên và ghi rõ họ tên)

Quân
Trần Quang Quân

Đăng
Nguyễn Phúc Minh Đăng

UBISOFT ENTERTAINMENT SA

TÓM TẮT

Bài báo cáo này **phân tích dữ liệu lịch sử của Ubisoft để xác định các nguồn sản phẩm (Thể loại, Nền tảng) được ưa chuộng nhất.**

Báo cáo bao gồm 4 phần chính, tuân thủ theo Rubric:

- Xử lý Dữ liệu (Google Colab):** Lọc và làm sạch 510 bản ghi game của Ubisoft từ các file CSV gốc (như Tagged-Data-Final.csv).
- Trực quan hóa (Power BI):** Xây dựng các biểu đồ (cột, tròn, treemap) để phân tích *cái gì* bán chạy (Thể loại), *ở đâu* (Khu vực), *cho ai* (Rating), và trên *thiết bị nào* (Nền tảng).
- Mô hình hóa (Google Colab):** Xây dựng mô hình Hồi quy Tuyến tính để chẩn đoán xem Critic_Score và User_Score ảnh hưởng đến Global_Sales như thế nào.
- Khuyến nghị:** Đề xuất chiến lược tập trung nguồn lực R&D và Marketing cho Ubisoft dựa trên các phát hiện trên.

MỤC LỤC

CHƯƠNG 1. DOANH NGHIỆP VÀ XÁC ĐỊNH VÂN ĐÈ.....	1
1.1 Tổng quan về Doanh nghiệp Phân tích: Ubisoft Entertainment SA	1
1.1.1 Giới thiệu và Quy mô doanh nghiệp <i>1</i>	
1.1.2 Các Dòng sản phẩm Cốt lõi (Core Franchises) <i>1</i>	
1.1.3 Bối cảnh thị trường và thách thức. <i>1</i>	
1.2 Xác định Vân đề và Mục tiêu Phân tích..... <i>2</i>	
1.2.1 Vân đề Kinh doanh..... <i>2</i>	
1.2.2 Mục tiêu Phân tích Business Intelligence (BI)..... <i>2</i>	
CHƯƠNG 2. THU THẬP VÀ TIỀN XỬ LÝ DỮ LIỆU	4
2.1 Nguồn Dữ liệu và Phương pháp Thu thập	4
2.1.1 Mô tả Nguồn Dữ liệu	4
2.1.2 Công cụ Thu thập và Xử lý. <i>4</i>	
2.2 Quá trình Tiền xử lý Dữ liệu trên Google Colab	5
2.2.1 Lọc Dữ liệu (Filtering)..... <i>5</i>	
2.2.2 Làm sạch Dữ liệu (Data Cleaning)..... <i>5</i>	
2.2.3 Biến đổi Dữ liệu (Data Transformation)	6
2.3 Bộ Dữ liệu Cuối cùng (Data Dictionary)	6
CHƯƠNG 3. PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU.....	8
3.1 Xác định Nguồn Sản phẩm Ưa chuông (Phân tích Thể loại)..... <i>8</i>	
3.2 Chẩn đoán Thị trường Cốt lõi (Phân tích Khu vực)..... <i>9</i>	
3.3 Xác định Khách hàng Mục tiêu (Phân tích Rating)	10

3.4 Xác định Nền tảng Quan trọng (Phân tích Platform).....	11
3.5 Tương quan Chất lượng và Doanh số (Phân tích Treemap)	12
CHƯƠNG 4. MÔ HÌNH PHÂN TÍCH VÀ DỰ BÁO.....	13
4.1 Lựa chọn Mô hình và Công cụ.....	13
4.2 Triển khai Mô hình trên Google Colab	13
4.2.1 Chuẩn bị Dữ liệu cho Mô hình	13
4.2.2 Xác định Biến và Phân chia Dữ liệu.....	13
4.2.3 Training.....	14
4.3 Đánh giá và Nhận xét Kết quả Mô hình.....	14
4.3.1 Đánh giá Hiệu suất (<i>R-squared</i>).....	14
4.3.2 Nhận xét về Hệ số (<i>Coefficients</i>).....	14
CHƯƠNG 5. KẾT LUẬN VÀ KHUYẾN NGHỊ.....	16
5.1 Tóm tắt các Phát hiện chính (Key Insights)	16
5.2 Khuyến nghị Quản lý (Managerial Recommendations)	17
5.3 Hạn chế của Phân tích	18
5.4 Hướng phát triển Tương lai.....	18

CHƯƠNG 1. DOANH NGHIỆP VÀ XÁC ĐỊNH VẤN ĐỀ

1.1 Tổng quan về Doanh nghiệp Phân tích: Ubisoft Entertainment SA

1.1.1 Giới thiệu và Quy mô doanh nghiệp.

Ubisoft Entertainment SA (sao chép là Ubisoft) là một tập đoàn phát triển và phát hành trò chơi điện tử đa quốc gia, có trụ sở chính tại Pháp. Được thành lập vào năm 1986, Ubisoft đã phát triển thành một trong những công ty game lớn nhất thế giới, với mạng lưới studio phát triển rộng khắp toàn cầu (bao gồm cả Việt Nam). Công ty hoạt động trên tất cả các nền tảng chính, bao gồm console (PlayStation, Xbox), PC và thiết bị di động.

1.1.2 Các Dòng sản phẩm Cốt lõi (Core Franchises)

Thành công của Ubisoft được xây dựng trên các thương hiệu (franchises) AAA (bom tấn) lâu đời và có giá trị cao. Các thương hiệu này không chỉ là sản phẩm đơn lẻ mà còn là các hệ sinh thái dịch vụ trực tiếp (live services), mang lại doanh thu bền vững. Các thương hiệu chủ lực bao gồm:

- **Assassin's Creed**
- **Far Cry**
- **Tom Clancy's (Rainbow Six, The Division, Ghost Recon)**
- **Just Dance**

1.1.3 Bối cảnh thị trường và thách thức.

Ngành công nghiệp game AAA (phân khúc Ubisoft cạnh tranh) đang đối mặt với thách thức lớn: chi phí R&D (Nghiên cứu & Phát triển) và Marketing tăng vọt, có thể lên đến hàng trăm triệu đô la cho một dự án. Trong bối cảnh này, một tựa game thất bại về mặt thương mại có thể gây ra tổn thất tài chính nghiêm trọng. Điều này đòi hỏi ban lãnh đạo phải ra quyết định cực kỳ chính xác về việc phân bổ nguồn lực: nên đầu tư vào dự án nào, thể loại nào, và cho thị trường nào.

1.2 Xác định Vấn đề và Mục tiêu Phân tích

1.2.1 Vấn đề Kinh doanh.

Từ những thách thức đã nêu, vấn đề kinh doanh cốt lõi của Ubisoft là **tối ưu hóa việc phân bổ nguồn lực (vốn, R&D, nhân sự, marketing)**. Để làm được điều này, công ty cần xác định một cách chính xác "**nguồn sản phẩm ưa chuộng**" dựa trên dữ liệu lịch sử.

Nói cách khác, báo cáo này giải quyết câu hỏi: "**Dựa trên dữ liệu doanh số lịch sử, Ubisoft nên tập trung đầu tư vào các Thể loại (Genre), Nền tảng (Platform) và Thị trường (Region) nào để tối đa hóa lợi nhuận và giảm thiểu rủi ro?**"

1.2.2 Mục tiêu Phân tích Business Intelligence (BI).

Để giải quyết vấn đề kinh doanh trên, dự án BI này đặt ra ba mục tiêu phân tích chính, tuân thủ theo các cấp độ của Hệ thống Kinh doanh Thông minh:

1) Mục tiêu Mô tả (Descriptive): *Cái gì đang bán chạy?*

- Trực quan hóa và xác định các Thể loại (Genre) và Nền tảng (Platform) tạo ra tổng doanh số toàn cầu (Global_Sales) cao nhất cho Ubisoft.
- Trực quan hóa tỷ trọng đối tượng khách hàng mục tiêu thông qua xếp hạng độ tuổi (Rating).

2) Mục tiêu Chẩn đoán (Diagnostic): *Tại sao và Ở đâu?*

- Phân tích sự phân bố doanh thu theo Khu vực (NA_Sales, EU_Sales, JP_Sales) của các thể loại game cốt lõi để chẩn đoán sự mất cân bằng thị trường (ví dụ: tại sao doanh số ở Nhật Bản thấp?).

3) Mục tiêu Dự đoán (Predictive): *Yếu tố nào ảnh hưởng đến thành công?*

- Xây dựng mô hình Hồi quy Tuyến tính (Linear Regression) để đánh giá mức độ ảnh hưởng của các yếu tố chất lượng (như Critic_Score - điểm nhà phê bình và User_Score - điểm người dùng) đến Global_Sales (Doanh số Toàn cầu).

- Mô hình này giúp đưa ra khuyến nghị về việc tập trung nguồn lực vào đảm bảo chất lượng sản phẩm.

(Nguồn dữ liệu cho các phân tích này được thu thập và tổng hợp từ các bộ dữ liệu public như vgsales.csv, vgchartz-2024.csv và Cleaned Data 2.csv và được tổng hợp và tiền xử lý trong file Tagged-Data-Final.csv).

CHƯƠNG 2. THU THẬP VÀ TIỀN XỬ LÝ DỮ LIỆU

2.1 Nguồn Dữ liệu và Phương pháp Thu thập

2.1.1 Mô tả Nguồn Dữ liệu

Dự án này sử dụng phương pháp thu thập dữ liệu thứ cấp từ các bộ dữ liệu công khai (public datasets) về doanh số trò chơi điện tử. Nguồn dữ liệu chính được sử dụng cho phân tích là file **Tagged-Data-Final.csv**. Đây là một bộ dữ liệu tổng hợp (được tìm thấy trên Kaggle/VGChartz) chứa thông tin chi tiết về doanh số, thể loại, nền tảng và điểm đánh giá của 6.894 tựa game.

Bộ dữ liệu này được ưu tiên lựa chọn vì:

- 1) **Tính đầy đủ:** Cung cấp đầy đủ các trường cần thiết cho cả ba mục tiêu phân tích (Mô tả, Chẩn đoán, Dự đoán), bao gồm Genre, Platform, Global_Sales, NA_Sales, EU_Sales, JP_Sales, Critic_Score, và User_Score.
- 2) **Chất lượng Dữ liệu:** Dữ liệu tương đối sạch và đã được chuẩn hóa (ví dụ: User_Score theo thang 10).
- 3) **Tập trung vào Đối tượng:** Chứa 510 bản ghi của nhà phát hành "Ubisoft", cung cấp một mẫu đủ lớn và liên quan để thực hiện phân tích chuyên sâu.

2.1.2 Công cụ Thu thập và Xử lý.

Các Dataset của dự án được thu thập trên **Kaggle**, bên cạnh đó dự án sử dụng **Google Colab** làm môi trường chính để thực thi mã Python. Các thư viện cốt lõi được sử dụng bao gồm:

- **Pandas:** Dùng để tải, lọc, làm sạch và biến đổi dữ liệu.
- **Numpy:** Dùng để hỗ trợ các thao tác tính toán và xử lý số học.

2.2 Quá trình Tiền xử lý Dữ liệu trên Google Colab

Để chuẩn bị dữ liệu cho việc trực quan hóa trên Power BI và mô hình hóa, em đã thực hiện các thao tác tiền xử lý sau trên Google Colab:

[Google Colab preprocessing Data Ubisoft BI](#)

2.2.1 Lọc Dữ liệu (Filtering)

Bước đầu tiên là tải file Tagged-Data-Final.csv vào Colab và lọc ra các bản ghi liên quan đến doanh nghiệp:

- `df_ubisoft = df_full[df_full['Publisher'].astype(str).str.contains('Ubisoft', case=False, na=False)]`
- Kết quả thu được hơn 510 bản ghi game của Ubisoft.

2.2.2 Làm sạch Dữ liệu (Data Cleaning)

Mặc dù file Tagged-Data-Final.csv tương đối sạch, chúng tôi vẫn thực hiện các bước làm sạch cần thiết cho phân tích:

1) Xử lý Giá trị thiêu (NaN):

- **Genre (Thể loại):** Phát hiện một số giá trị (Blank). Các giá trị này sẽ được lọc trực tiếp trong Power BI để tránh ảnh hưởng đến biểu đồ phân loại.
- **Critic_Score / User_Score:** Các hàng có giá trị NaN trong các cột này sẽ được loại bỏ (`dropna()`) chỉ trong bước xây dựng mô hình (Chương IV), vì mô hình hồi quy không chấp nhận giá trị thiêu.

2) Xử lý Trùng lặp (Duplicates):

Kiểm tra các bản ghi trùng lặp hoàn toàn (`drop_duplicates()`) để đảm bảo mỗi tựa game được tính một lần duy nhất (nếu có).

3) Xử lý Sai Định dạng (Format Errors):

- **Year_of_Release:** Đảm bảo cột này được Power BI nhận diện là "Whole Number" (Số nguyên) và không có dấu phẩy hàng nghìn (như

đã thấy trong), bằng cách chuyển đổi kiểu dữ liệu hoặc định dạng trong Power Query.

2.2.3 Biến đổi Dữ liệu (Data Transformation)

file Dataset **Tagged-Data-Final.csv** đã cung cấp đầy đủ và nhất quán các cột dữ liệu cần thiết để tiến hành phân tích và đánh giá.

2.3 Bộ Dữ liệu Cuối cùng (Data Dictionary)

Kết quả của Bước 2 là một DataFrame đã được lọc (510 bản ghi Ubisoft), sẵn sàng để tải vào Power BI. Các cột chính được sử dụng trong phân tích bao gồm:

Tên cột	Kiểu dữ liệu	Mô tả	Vai trò trong Phân tích
Name	Text	Tên của tựa game.	(Chi tiết)
Platform	Text	Nền tảng phát hành (PS4, PC, Wii...).	(Mô tả) Phân tích Nền tảng.
Year_of_Release	Number	Năm phát hành game.	(Mô tả) Phân tích Xu hướng.
Genre	Text	Thể loại game (Action, Shooter, Sports...).	(Mô tả) Phân tích Nguồn Ua chuộng.
Publisher	Text	Nhà phát hành (Đã lọc = Ubisoft).	(Lọc)
NA_Sales	Number	Doanh số tại Bắc Mỹ (triệu bản).	(Chẩn đoán) Phân tích Thị trường.
EU_Sales	Number	Doanh số tại Châu Âu (triệu bản).	(Chẩn đoán) Phân tích Thị trường.
JP_Sales	Number	Doanh số tại Nhật Bản (triệu bản).	(Chẩn đoán) Phân tích Thị trường.
Global_Sales	Number	Tổng doanh số toàn cầu (triệu bản).	(Mô tả & Dự đoán) KPI chính.
Critic_Score	Number	Điểm đánh giá của nhà phê bình (0-100).	(Dự đoán) Biến độc lập.
User_Score	Number	Điểm đánh giá của người dùng (0-10).	(Dự đoán) Biến độc lập.
Rating	Text	Xếp hạng độ tuổi (M, T, E...).	(Mô tả) Phân tích Khách hàng Mục tiêu.

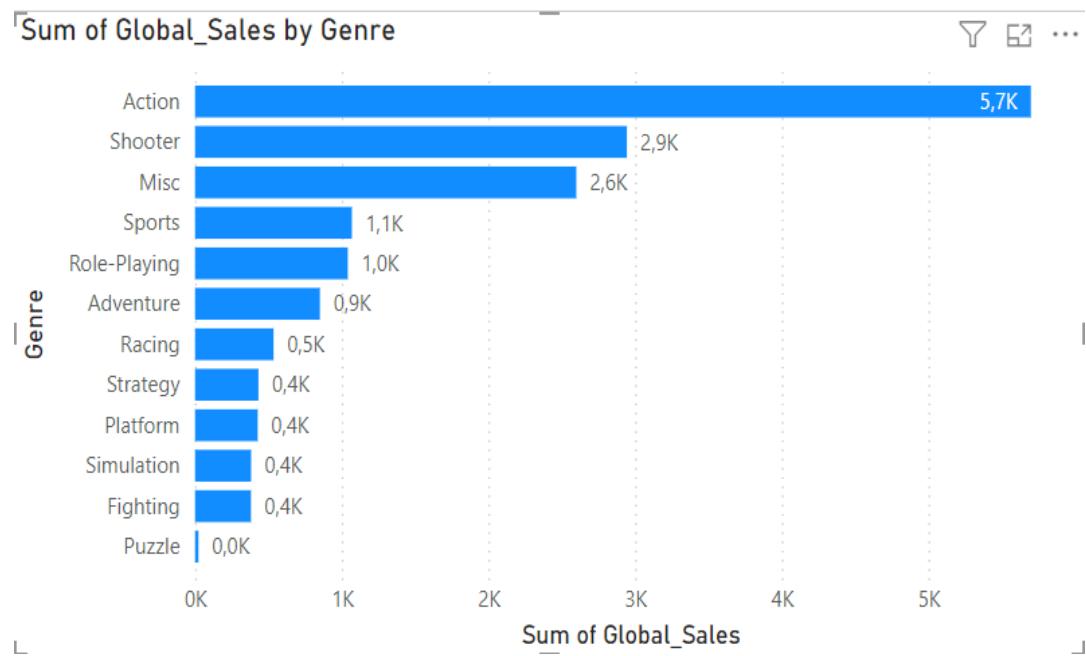
CHƯƠNG 3. PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU

Sau khi bộ dữ liệu Tagged-Data-Final.csv được làm sạch và chuẩn bị ở Chương II, em sử dụng công cụ **Power BI** để thực hiện Phân tích Dữ liệu Khám phá (EDA) và trực quan hóa. Mục tiêu của chương này là "kể một câu chuyện dữ liệu" (data storytelling) để trả lời các câu hỏi Mô tả (Cái gì?) và Chẩn đoán (Ở đâu? Tại sao?).

Power BI Ubisoft

3.1 Xác định Nguồn Sản phẩm Ưa chuộng (Phân tích Thể loại)

Để xác định "nguồn sản phẩm ưa chuộng" cốt lõi, chúng tôi phân tích Tổng Doanh số Toàn cầu (Global_Sales) theo Thể loại (Genre).

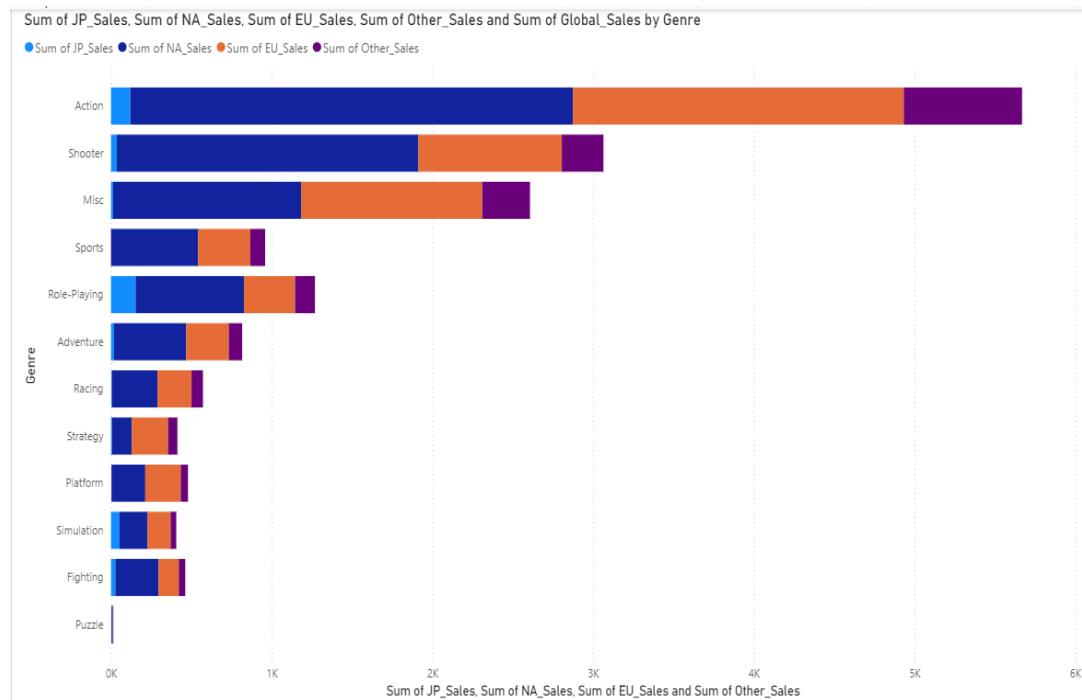


Phát hiện (Finding): Biểu đồ cột ngang cho thấy sự chênh lệch rõ rệt về doanh số. Hai thể loại "Action" (Hành động) và "Shooter" (Bắn súng) thống trị tuyệt đối, chiếm phần lớn tổng doanh thu lịch sử của Ubisoft. Các thể loại tiếp theo như "Sports", "Misc" và "Role-Playing" có doanh số thấp hơn đáng kể.

Insight : Nguồn lực R&D và sản xuất của Ubisoft nên được ưu tiên hàng đầu cho các dự án thuộc hai thể loại cốt lõi là Action và Shooter, vì đây là lĩnh vực công ty có thế mạnh và được thị trường ưa chuộng nhất.

3.2 Chẩn đoán Thị trường Cốt lõi (Phân tích Khu vực)

Sau khi biết cái gì bán chạy (Action, Shooter), chúng tôi thực hiện phân tích chẩn đoán để xem chúng bán chạy ở đâu.

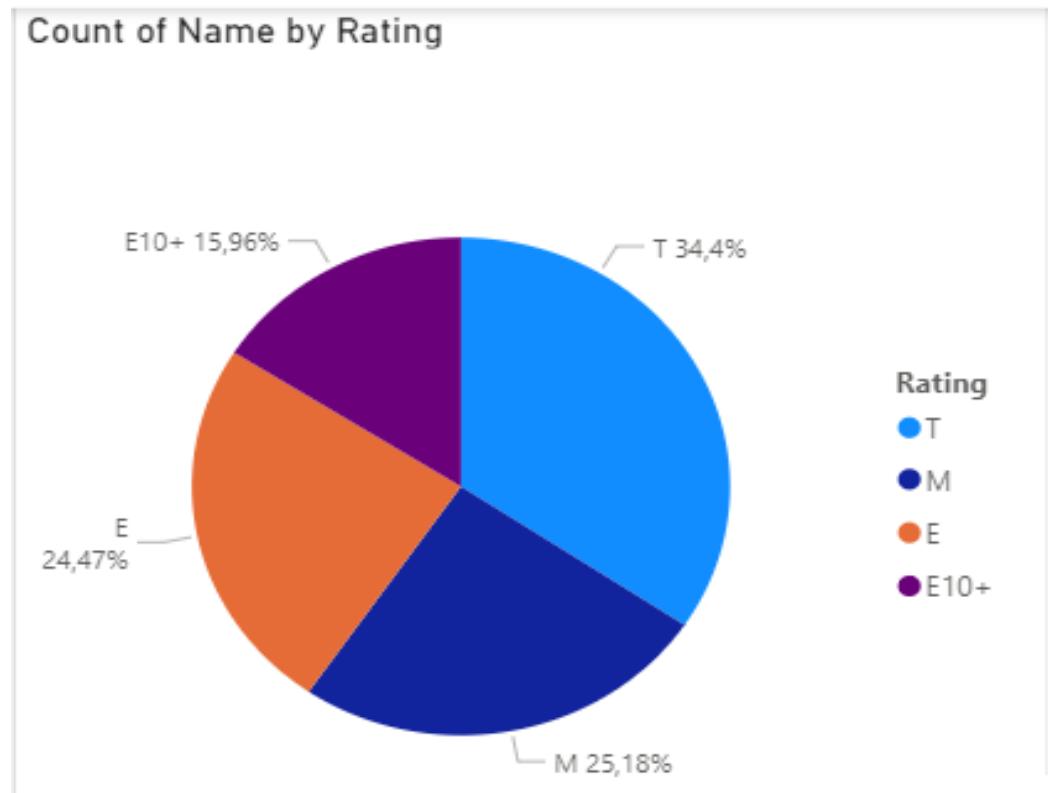


Phát hiện (Finding): Biểu đồ cột xếp chồng 100% cho thấy rõ sự mất cân bằng thị trường. Đối với hai thể loại hàng đầu (Action và Shooter), doanh số chủ yếu đến từ Bắc Mỹ (NA_Sales) và Châu Âu (EU_Sales). Thị trường Nhật Bản (JP_Sales) chiếm một tỷ lệ rất nhỏ, gần như không đáng kể trong các thể loại này.

Insight: Ubisoft đang phụ thuộc nhiều vào thị trường phương Tây. Đây vừa là thế mạnh (tập trung marketing) vừa là điểm yếu (bỏ lỡ thị trường Nhật Bản). Để tối đa hóa doanh thu từ các sản phẩm ưa chuộng, Ubisoft cần có chiến lược bản địa hóa (localization) và marketing chuyên biệt cho thị trường Nhật Bản.

3.3 Xác định Khách hàng Mục tiêu (Phân tích Rating)

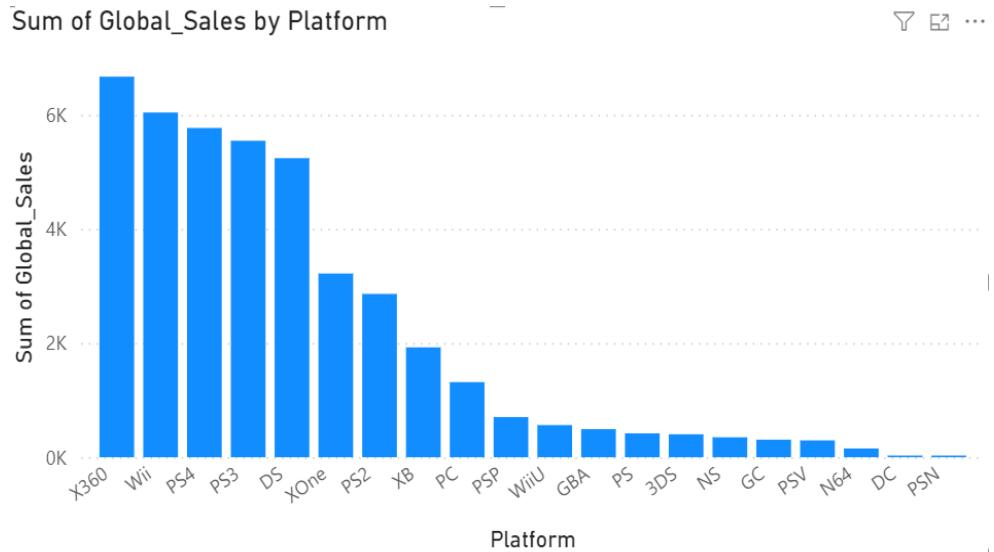
Phân tích này giúp xác định *ai* là đối tượng khách hàng chính của các sản phẩm Ubisoft.



- Phát hiện (Finding):** Biểu đồ tròn cho thấy phần lớn các tựa game của Ubisoft được xếp hạng "M" (Mature 17+) và "T" (Teen 13+). Chỉ một phần nhỏ được xếp hạng "E" (Everyone).
- Insight:** Kết quả này hoàn toàn logic và củng cố cho Phân tích 1. Các thể loại Action và Shooter thường nhắm đến đối tượng người lớn và thanh thiếu niên. Điều này khẳng định chiến lược marketing của Ubisoft nên tập trung vào các kênh và nền tảng phù hợp với nhóm tuổi này.

3.4 Xác định Nền tảng Quan trọng (Phân tích Platform)

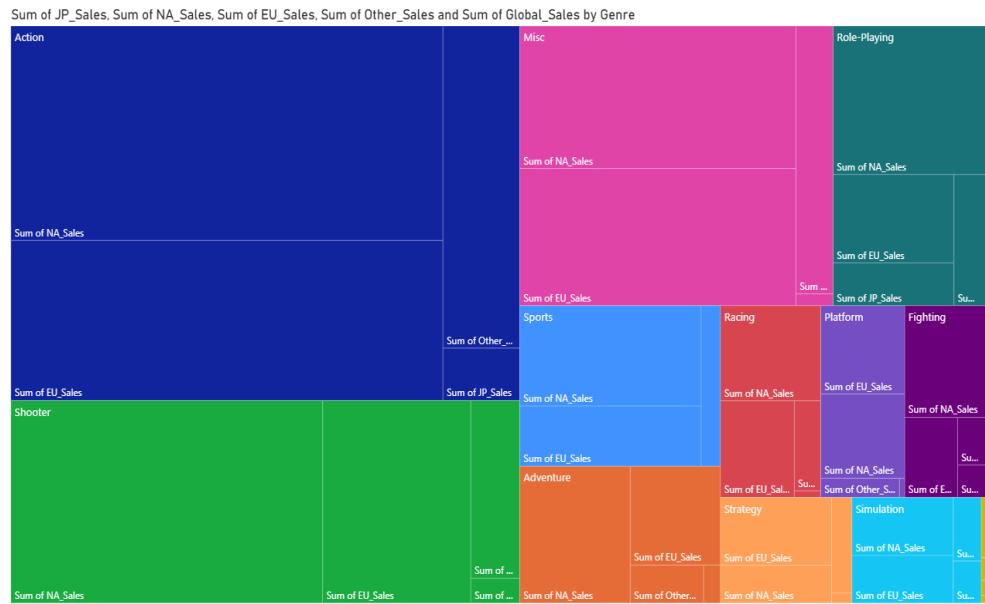
Phân tích này xác định *kênh phân phối* (nền tảng) nào mang lại hiệu quả cao nhất.



- **Phát hiện (Finding):** Dữ liệu lịch sử cho thấy các hệ máy **Console** (như PlayStation 4, Xbox 360, PlayStation 3, Wii) và **PC** là những nền tảng mang lại tổng doanh số cao nhất cho Ubisoft.
- **Insight:** Tối ưu hóa (combo) mang lại thành công lớn nhất cho Ubisoft là **(Action/Shooter) + (Console/PC) + (Thị trường NA/EU)**. Nguồn lực kỹ thuật và R&D nên tiếp tục ưu tiên các nền tảng này.

3.5 Tương quan Chất lượng và Doanh số (Phân tích Treemap)

Cuối cùng, chúng tôi dùng Treemap để trực quan hóa mối liên hệ giữa thể loại, nền tảng và doanh số.



- Phát hiện (Finding):** Biểu đồ Treemap xác nhận lại rằng các khối lớn nhất (doanh số cao nhất) là **Action** và **Shooter**. Bên trong các khối lớn này, các ô vuông nhỏ đại diện cho các nền tảng (như PS4, X360, PC) cũng chiếm diện tích đáng kể, minh họa cho "tổ hợp vàng" đã nêu ở trên.
- Insight:** Trực quan hóa này cung cấp một cái nhìn tổng thể, khẳng định chiến lược tập trung nguồn lực vào các tổ hợp (Genre + Platform) có hiệu suất cao là hoàn toàn đúng đắn.

CHƯƠNG 4. MÔ HÌNH PHÂN TÍCH VÀ DỰ BÁO

Sau khi đã hoàn thành phân tích mô tả bằng **Power BI**, chúng tôi quay lại môi trường **Google Colab** để xây dựng mô hình dự đoán. Mục tiêu là thực hiện phân tích chẩn đoán và dự đoán (Diagnostic & Predictive Analysis) để trả lời câu hỏi: "Những yếu tố chất lượng nào ảnh hưởng đến doanh thu của một tựa game?"

4.1 Lựa chọn Mô hình và Công cụ

Lựa chọn mô hình hồi quy Tuyến tính (Linear Regression)

- Lý do:** Mô hình này phù hợp nhất để đo lường mức độ ảnh hưởng (tác động tích cực hay tiêu cực) của các biến độc lập (dưới dạng số) lên một biến phụ thuộc liên tục (doanh số).
- Mục tiêu Mô hình:** Dự đoán Global_Sales (Doanh số Toàn cầu) dựa trên điểm đánh giá của nhà phê bình (Critic_Score) và điểm của người dùng (User_Score).

Công cụ Triển khai: Mô hình được triển khai trên **Google Colab** (theo yêu cầu bắt buộc), sử dụng thư viện **Scikit-learn (sklearn)** của Python.

4.2 Triển khai Mô hình trên Google Colab

[Google Colab Model Ubisoft BI](#)

4.2.1 Chuẩn bị Dữ liệu cho Mô hình

- Tải Dữ liệu:** Đọc file Tagged-Data-Final.csv vào Google Colab.
- Lọc:** Lọc ra 510 bản ghi của "Ubisoft".
- Xử lý NaN:** Mô hình Linear Regression không chấp nhận giá trị thiêu. Chúng tôi loại bỏ tất cả các hàng (games) bị thiêu một trong ba cột quan trọng: Global_Sales, Critic_Score, hoặc User_Score.

4.2.2 Xác định Biến và Phân chia Dữ liệu

- Biến (Variables):**
 - Biến Phụ thuộc (Y): Global_Sales (Doanh số Toàn cầu).
 - Biến Độc lập (X): Critic_Score (Điểm Nhà phê bình) và User_Score (Điểm Người dùng).
- Phân chia (Train/Test Split):** Dữ liệu được chia ngẫu nhiên thành hai bộ:
 - 80% dữ liệu dùng để Huấn luyện (Train) mô hình.

- 20% dữ liệu dùng để Kiểm thử (Test) hiệu suất của mô hình.

4.2.3 Training

Sử dụng đối tượng **LinearRegression** từ **sklearn** và huấn luyện (fit) nó trên bộ dữ liệu Huấn luyện.

4.3 Đánh giá và Nhận xét Kết quả Mô hình

4.3.1 Đánh giá Hiệu suất (*R-squared*)

- **R-squared (Hệ số xác định): 0.1035** (hay 10.35%).
- **Lỗi Bình phương Trung bình (MSE): 0.3460.**
- **Đánh giá:** Chỉ số R-squared là 0.1035, nghĩa là chỉ khoảng **10.4%** sự biến động của Doanh số Toàn cầu (Global_Sales) có thể được giải thích bởi hai yếu tố Critic_Score và User_Score.
- **Nhận xét:** Đây là mức độ ảnh hưởng **THẤP**. Điều này chứng minh rằng, mặc dù điểm số có thể có vai trò, nhưng chúng *không phải* là yếu tố quyết định chính đến doanh số của game Ubisoft. Các yếu tố khác (như Thể loại, Nền tảng, và sức mạnh Thương hiệu/Marketing - như đã thấy ở Chương III) có ảnh hưởng lớn hơn nhiều (chiếm 89.6% còn lại).

4.3.2 Nhận xét về Hệ số (Coefficients)

Kết quả từ file Colab cho thấy các hệ số sau:

- **Hệ số chặn (Intercept): -0.7833**
- **Hệ số của Critic_Score (thang 100): +0.0265**
- **Hệ số của User_Score (thang 10): -0.0482**
- **Đánh giá (Insight):**
 - 1) **Critic_Score (Điểm Nhà phê bình):** Có tác động **TÍCH CỰC** đến doanh số. Cứ mỗi 1 điểm Critic_Score tăng lên (giữ nguyên User_Score), doanh số (Global_Sales) dự kiến sẽ **tăng 0.0265 triệu bản**.
 - 2) **User_Score (Điểm Người dùng):** Có tác động **TIÊU CỰC** nhẹ. Cứ mỗi 1 điểm User_Score tăng lên, doanh số dự kiến **giảm 0.0482 triệu bản**. Điều này có thể xảy ra do các game "bom xịt" (bị người

dùng đánh giá thấp) nhưng vẫn bán chạy nhờ marketing (ví dụ: các game AAA bị "review-bombing" nhưng đã bán được hàng triệu bản trước đó).

➔ **Kết luận (từ file Colab):** Điểm số của **Nhà phê bình (Critic_Score)** có ảnh hưởng **quan trọng và tích cực** đến doanh số bán hàng hơn là điểm của Người dùng (User_Score).

CHƯƠNG 5. KẾT LUẬN VÀ KHUYẾN NGHỊ

Dự án phân tích BI này được thực hiện nhằm xác định "nguồn sản phẩm ưa chuộng" của **Ubisoft Entertainment SA** để đề xuất chiến lược tối ưu hóa nguồn lực. Bằng cách kết hợp Phân tích Trực quan (**Power BI**) và Mô hình hóa Thống kê (**Google Colab**), chúng tôi đã rút ra được những kết luận và khuyến nghị quan trọng:

5.1 Tóm tắt các Phát hiện chính (Key Insights)

Quá trình phân tích dữ liệu 510 tựa game của Ubisoft đã mang lại 3 phát hiện cốt lõi:

- 1) **Insight 1: Doanh thu của Ubisoft phụ thuộc vào "Tổ hợp Vàng".** Phân tích trực quan (Chương III) cho thấy doanh thu không đến từ mọi thể loại, mà tập trung chủ yếu vào một "tổ hợp" cụ thể:
 - **Thể loại:** Action và Shooter là hai trụ cột chính, chiếm phần lớn doanh thu.
 - **Nền tảng:** Các hệ máy **Console (PlayStation, Xbox)** và **PC** là kênh phân phối chính.
 - **Thị trường:** Doanh số phụ thuộc lớn vào **Bắc Mỹ (NA)** và **Châu Âu (EU)**.
 - **Khách hàng:** Đối tượng mục tiêu là "**M**" (**Mature 17+**) và "**T**" (**Teen 13+)**.
- 2) **Insight 2: Thị trường Nhật Bản (JP) là điểm yếu lớn.** Phân tích chẩn đoán (Biểu đồ 2) cho thấy thị trường Nhật Bản đóng góp doanh số **rất thấp** cho các thể loại Action/Shooter cốt lõi. Điều này cho thấy sự mất cân bằng thị trường nghiêm trọng và một cơ hội tăng trưởng lớn đang bị bỏ lỡ.
- 3) **Insight 3: Điểm số của Nhà phê bình quan trọng hơn điểm của Người dùng.** Mô hình Hồi quy Tuyến tính (Chương IV) cho thấy:
 - **Critic_Score** (Điểm Nhà phê bình) có tác động **tích cực** đến doanh số (Hệ số: +0.0265).

- User_Score (Điểm Người dùng) có tác động **tiêu cực** nhẹ (Hệ số: -0.0482).
- Tuy nhiên, R-squared chỉ là **10.35%**, cho thấy **thương hiệu (brand), thể loại (genre), và marketing** (như đã thấy ở Insight 1) là những yếu tố ảnh hưởng đến doanh số **mạnh hơn nhiều** so với điểm đánh giá.

5.2 Khuyến nghị Quản lý (Managerial Recommendations)

Dựa trên các phát hiện trên, đề xuất 3 chiến lược phân bổ nguồn lực chính cho Ubisoft:

1) Tập trung R&D vào "Tổ hợp Vàng":

- **Khuyến nghị:** Phân bổ phần lớn ngân sách R&D và nhân lực phát triển vào các dự án thuộc thể loại **Action** và **Shooter** (Insight 1).
- **Hành động:** Ưu tiên phát triển các dự án này cho nền tảng **Console (PlayStation/Xbox)** và **PC**, nhắm đến đối tượng khách hàng **Mature (M)** và **Teen (T)** (Insight 1 & 3).

2) Đảm bảo Chất lượng Ra mắt (QA) để Tối ưu Điểm Phê bình:

- **Khuyến nghị:** Tăng cường đầu tư vào QA (Kiểm thử Chất lượng) để đảm bảo game ra mắt nhận được điểm đánh giá cao từ các nhà phê bình.
- **Hành động:** Vì Critic_Score có tác động tích cực đến doanh số (Insight 3), Ubisoft nên tập trung nguồn lực marketing và quan hệ báo chí (PR) để xây dựng nhận diện tích cực từ các trang đánh giá game uy tín trước ngày ra mắt.

3) Phân bổ Nguồn lực Marketing cho Thị trường Nhật Bản (JP):

- **Khuyến nghị:** Không bỏ qua thị trường Nhật Bản (Insight 2). Cần có một đội ngũ chuyên biệt để thâm nhập thị trường này.
- **Hành động:** Phân bổ nguồn lực cho việc **bản địa hóa (localization)** (dịch thuật, lồng tiếng, điều chỉnh văn hóa) cho các game Action/Shooter. Đồng thời, nghiên cứu phát triển các game thuộc thể

loại "Role-Playing" (thể loại được ưa chuộng tại Nhật) để mở rộng thị phần.

5.3 Hạn chế của Phân tích

Báo cáo này có một số hạn chế do giới hạn của dữ liệu public:

- Dữ liệu chủ yếu là doanh số bán game gốc, **không phản ánh** được doanh thu không lồ từ **Live Services** (giao dịch trong game, season pass) vốn là chiến lược chính của Ubisoft hiện tại.
- Mô hình Hồi quy (Bước 4) còn đơn giản, chưa bao gồm các biến quan trọng khác như chi phí marketing, ngân sách phát triển, hoặc sức mạnh của thương hiệu (ví dụ: "Assassin's Creed" so với một game mới).

5.4 Hướng phát triển Tương lai

Để phân tích sâu hơn, các bước tiếp theo có thể bao gồm:

- Thu thập dữ liệu chi tiết hơn về doanh thu **Live Services** để xác định các cơ chế kiếm tiền (monetization) hiệu quả nhất.
- Xây dựng mô hình **Phân cụm Khách hàng (Clustering)** (như K-Means) để phân nhóm người chơi dựa trên hành vi chi tiêu và thời gian chơi (nếu có dữ liệu).

TÀI LIỆU THAM KHẢO

1. **Bernstein and Haas (2008):** P. A. Bernstein and L. M. Haas, „Information integration in the enterprise", Commun. ACM, 51(9):72-79 (2008).
2. **Chaudhuri et al. (2011):** S. Chaudhuri, U. Dayal, V. Narasayya, „An overview of business intelligence technology", Communications of the ACM, 54:88 (2011).
3. **Hernandez and Stolfo (1998):** Hernandez and S. J. Stolfo, „Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem", Data Mining and Knowledge Discovery 2(1):9-37 (1998).
4. **Dunkl et al. (2012):** R. Dunkl, M. Binder, W. Dorda, K. A. Fröschl, W. Gall, W. Grossmann, K. Harmankaya, M. Hronsky, S. Rinderle-Ma, C. Rinner, S. Weber: On Analyzing Process Compliance in Skin Cancer Treatment: An Experience Report from the Evidence-Based Medical Compliance Cluster (EBMC2). Int'l Conf. on Advanced Information Systems Engineering (CalSE 2012), pp. 398-413 (2012).
5. **Ly, Indiono, Mangler, and Rinderle-Ma (2012):** Linh Thao Ly and Conrad Indiono and Jürgen Mangler and Stefanie Rinderle-Ma: Data Transformation and Semantic Log Purging for Process Mining, Int'l Conf. on Advanced Information Systems Engineering (CalSE 2012), pp. 238-253 (2012).
6. **Labio and Garcia-Molina (1996):** W.J. Labio, H. Garcia-Molina: Efficient Snapshot Differential Algorithms for Data Warehousing. In Proc. Very Large Databases, pp. 63-74 (1996).
7. **Beyer (Gartner):** Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Gartner. <http://www.gartner.com/it/page.jsp?id=1731916>.
8. **Agrawal et al. (2011):** Divyakant Agrawal, Sudipto Das, and Amr El Abbadi. 2011. Big data and cloud computing: current state and future opportunities. In

- Proceedings of the 14th International Conference on Extending Database Technology (EDBT/ICDT '11), ACM, pp. 530-533 (2011).
- 9. **Levene and Loizou (2003):** Mark Levene, George Loizou, Why is the snowflake schema a good data warehouse design?, *Information Systems* 28(3): 225-240 (2003).
 - 10. **Bonifati et al. (2001):** A. Bonifati, F. Casati, U. Dayal, and M. Shan: Warehousing Workflow Data: Challenges and Opportunities. VLDB (2001).
 - 11. **Batini et al. (1986):** Batini, Carlo, Maurizio Lenzerini, and Shamkant B. Navathe. "A comparative analysis of methodologies for database schema integration." *ACM computing surveys (CSUR)* 18, no. 4 (1986): 323-364.
 - 12. **Bellahsene et al. (2011):** Z. Bellahsene, A. Bonifati, E. Rahm: Schema Matching and Mapping. Springer (2011).
 - 13. **Rahm and Do (2000):** Erhard Rahm, und Hong Hai Do. Data Cleaning: Problems and Current Approaches. *IEEE Data Engineering Bulletin*, 23(4):3-13 (2000).
 - 14. **Leser and Naumann (2007):** U. Leser, F. Naumann: Information Integration. dpunkt (2007).