

BUSINESS INTELLIGENCE

FINAL REPORT

XÂY DỰNG HỆ THỐNG BI & MÔ HÌNH TỐI ƯU HÓA DOANH THU CHO

UBISOFT ENTERTAINMENT SA.

Instructor: MSC Võ Thị Kim Anh

Trần Quang Quân – 520H0569

Nguyễn Phúc Minh Đăng - 521H0497



BỐI CẢNH NGHIÊN CỨU

Ubisoft Entertainment SA

- **Thị trường:** Ngành Game trị giá 300 tỷ USD, cạnh tranh gay gắt.
- **Đối tượng:** Ubisoft - Top 3 nhà phát hành lớn nhưng thị phần doanh thu thấp hơn Activision & EA.
- **Vấn đề (Pain Point):**
 - Ra mắt quá nhiều game nhưng hiệu quả thấp.
 - Phụ thuộc vào các dòng game cũ.
- **Giải pháp:** Cần chuyển đổi từ "Cảm tính" sang "Dữ liệu" (Data-Driven).



MỤC TIÊU DỰ ÁN

- **Mô tả (Descriptive):** Đánh giá vị thế và hiệu suất quá khứ (Power BI).
- **Dự báo (Predictive):** Xây dựng mô hình Machine Learning dự đoán doanh thu (Python).
- **Chỉ dẫn (Prescriptive):** Đề xuất chiến lược tối ưu hóa nguồn lực.



DỮ LIỆU & CÔNG CỤ

Bối cảnh thị trường và thách thức

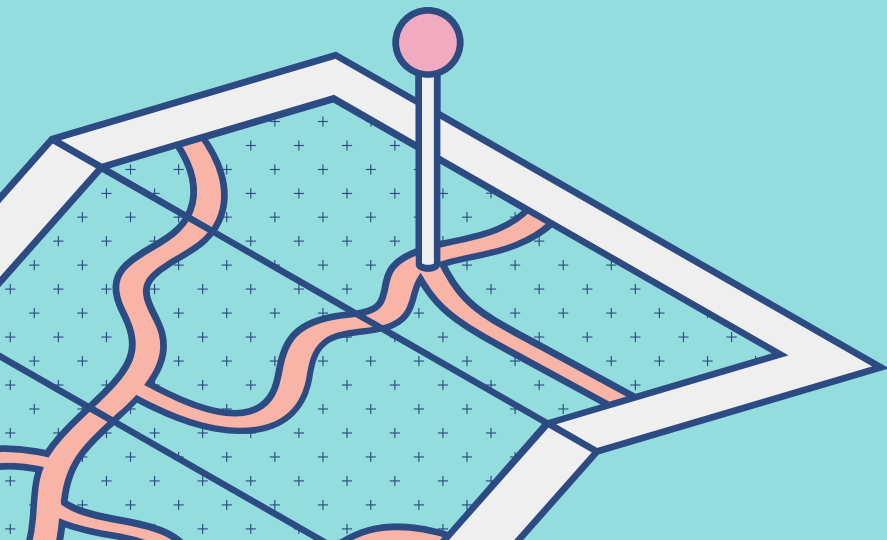
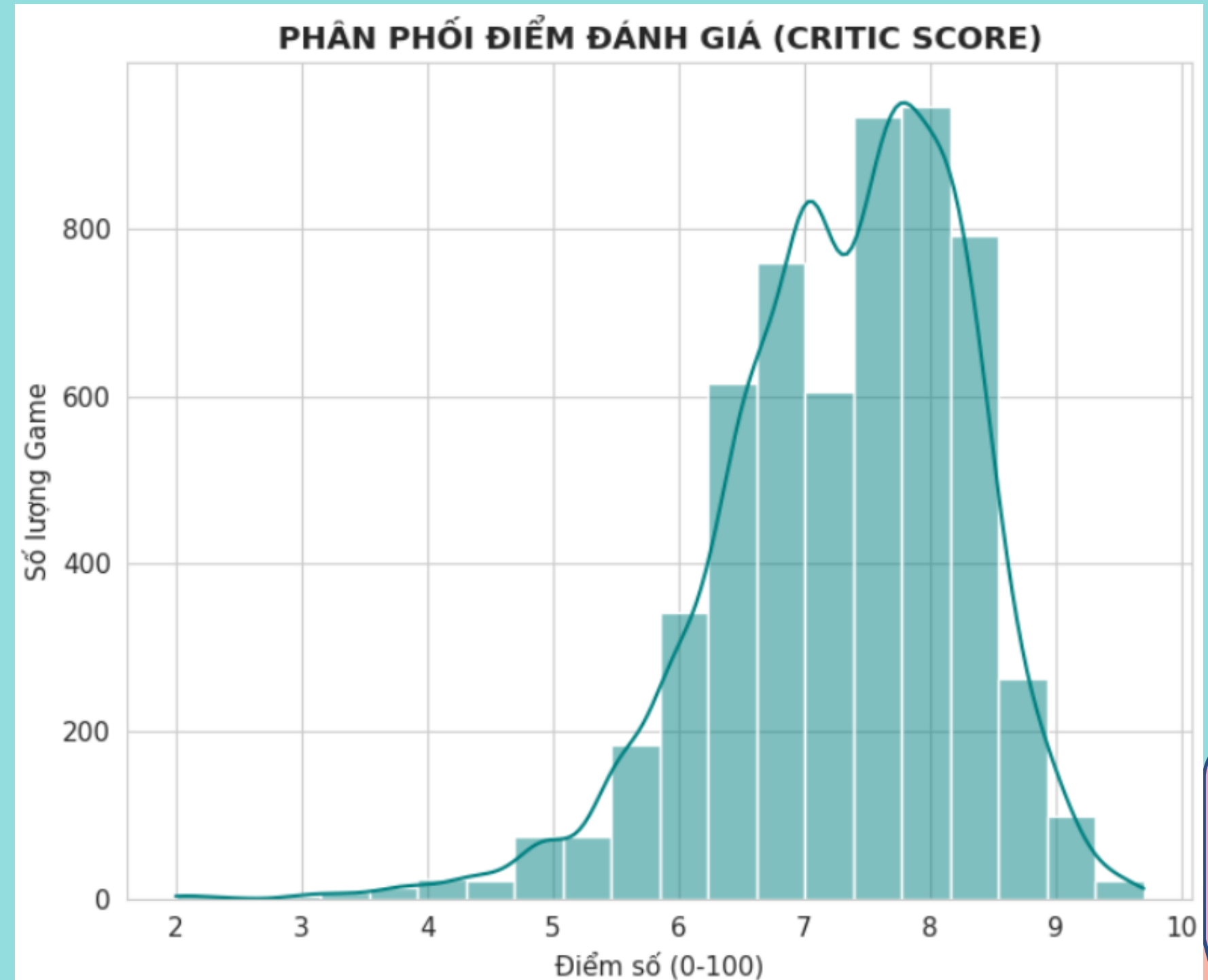
- **Nguồn dữ liệu:** Kaggle Dataset (16,000+ bản ghi, 1980-2020).
- **Công cụ:**
- **Xử lý & Mô hình:** Python (Google Colab), Scikit-learn.
- **Trực quan hóa:** Microsoft Power BI.

3. 5 DÒNG DỮ LIỆU ĐẦU TIÊN:

	title	console	genre	Publisher_Group	developer	critic_score	total_sales	release_date	Year
0	Call of Duty: Black Ops 3	PS4	Shooter	Activision	Treyarch	8.1	15.09	2015-11-06	2015.0
1	Call of Duty: Modern Warfare 3	X360	Shooter	Activision	Infinity Ward	8.7	14.82	2011-11-08	2011.0
2	Call of Duty: Black Ops	X360	Shooter	Activision	Treyarch	8.8	14.74	2010-11-09	2010.0
3	Call of Duty: Black Ops II	X360	Shooter	Activision	Treyarch	8.4	13.86	2012-11-13	2012.0
4	Call of Duty: Black Ops II	PS3	Shooter	Activision	Treyarch	8.0	13.80	2012-11-13	2012.0

XỬ LÝ DỮ LIỆU NÂNG CAO (ADVANCED ETL)

- **Thách thức:** Dữ liệu *Critic_Score* bị thiếu rất nhiều (Missing Values).
- **Giải pháp:** Sử dụng thuật toán **KNN Imputer (K-Nearest Neighbors)**.
- **Cơ chế:**
 - Không điền trung bình (Mean) đơn giản.
 - Tìm các game "láng giềng" tương đồng để điền giá trị.
 - -> Tăng độ chính xác cho mô hình.



XỬ LÝ DỮ LIỆU NÂNG CAO (ADVANCED ETL)

```
# 2. Mã hóa dữ liệu (Chuyển chữ thành số, Máy học chỉ hiểu số)
le = LabelEncoder()
# Tạo các cột mã hóa tạm
df_ml['pub_code'] = le.fit_transform(df_ml['Publisher_Group'].astype(str))
df_ml['genre_code'] = le.fit_transform(df_ml['genre'].astype(str))
df_ml['console_code'] = le.fit_transform(df_ml['console'].astype(str))

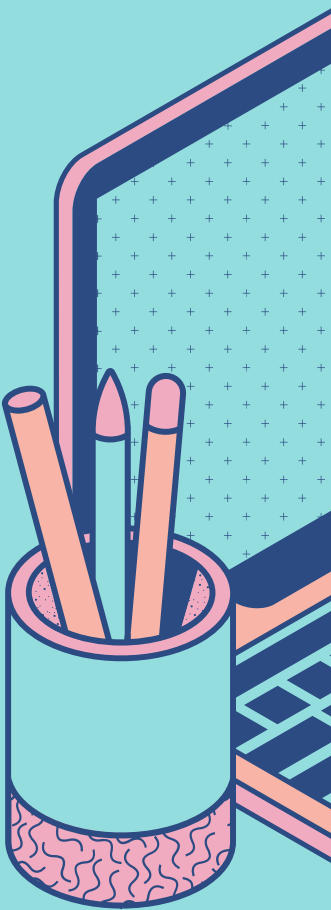
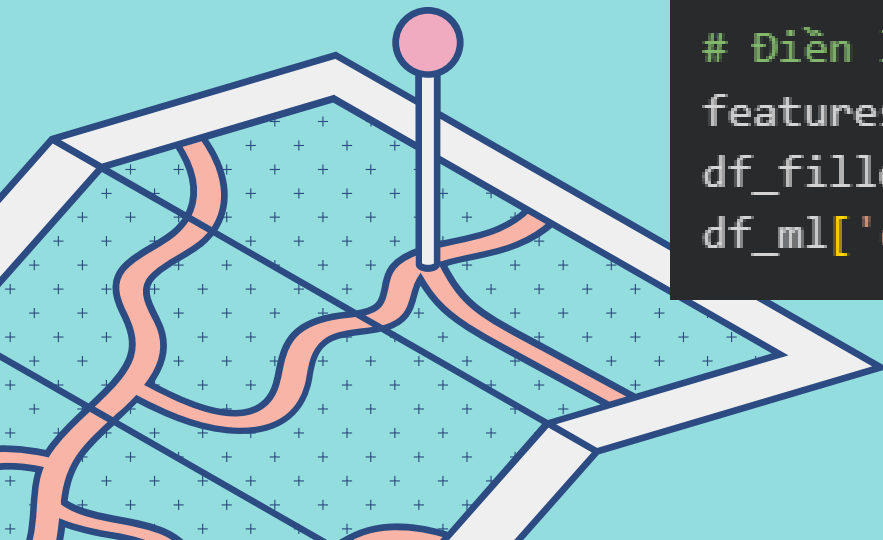
# Năm phát hành, nếu thiếu năm thì điền tạm giá trị trung vị
df_ml['Year_fix'] = df_ml['Year'].fillna(df_ml['Year'].median())

# 3. Chọn các đặc trưng để đưa vào mô hình: Hãng, Thể loại, Hệ máy, Năm -> Để dự đoán: Doanh số
features = ['pub_code', 'genre_code', 'console_code', 'Year_fix', 'total_sales']

# 4. Chạy thuật toán KNN (Tìm 5 hàng xóm gần nhất)
imputer = KNNImputer(n_neighbors=5)
# Hàm này trả về một mảng số đã được điền đầy đủ
df_filled_array = imputer.fit_transform(df_ml[features])

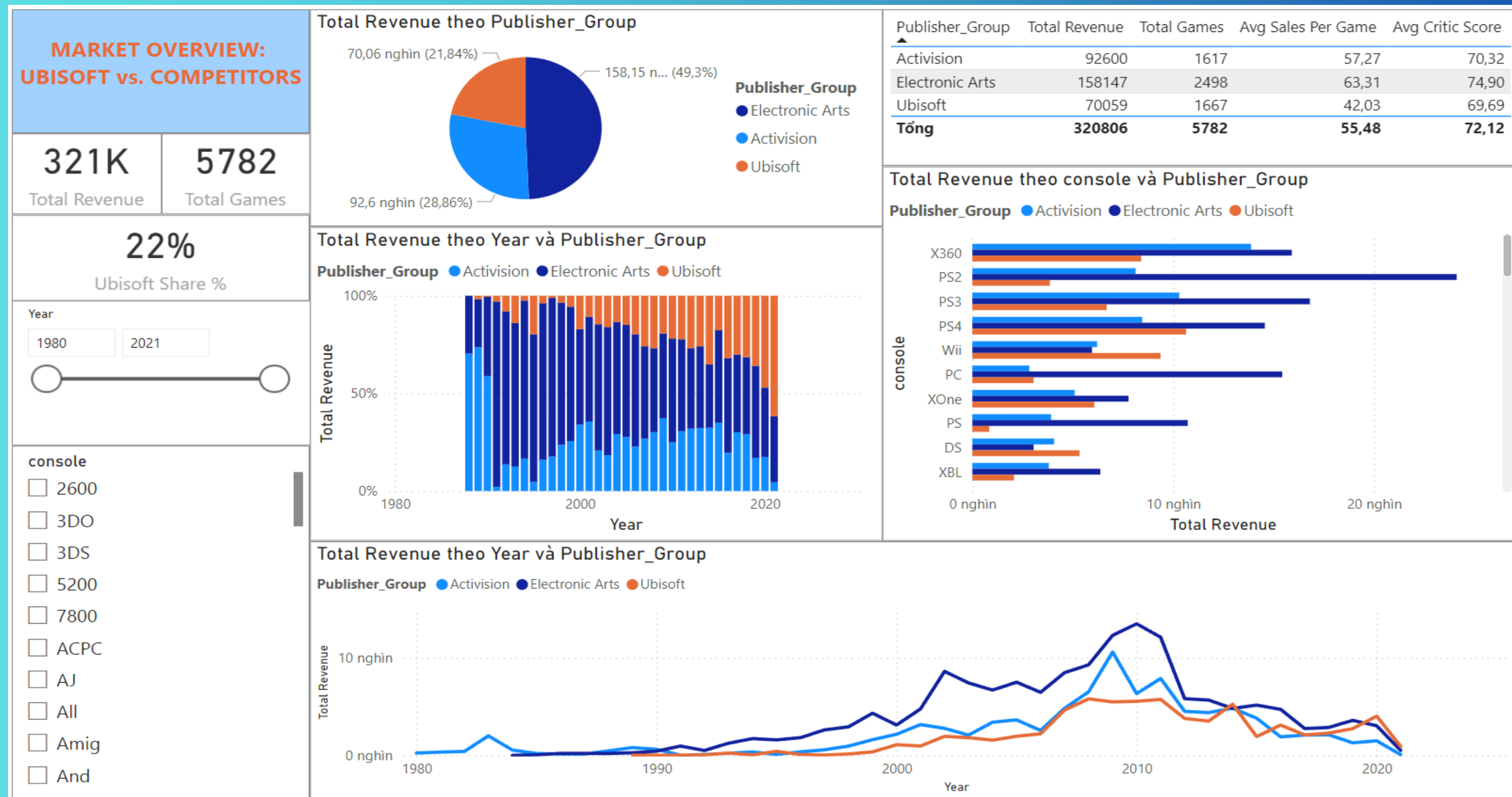
# 5. Gán ngược lại vào DataFrame
# Cột cuối cùng (index 4) trong mảng kết quả chính là total_sales
df_ml['total_sales'] = df_filled_array[:, 4].round(2)

# Điền luôn Critic Score dùng cùng thuật toán
features_score = ['pub_code', 'genre_code', 'console_code', 'Year_fix', 'critic_score']
df_filled_score = imputer.fit_transform(df_ml[features_score])
df_ml['critic_score'] = df_filled_score[:, 4].round(1)
```



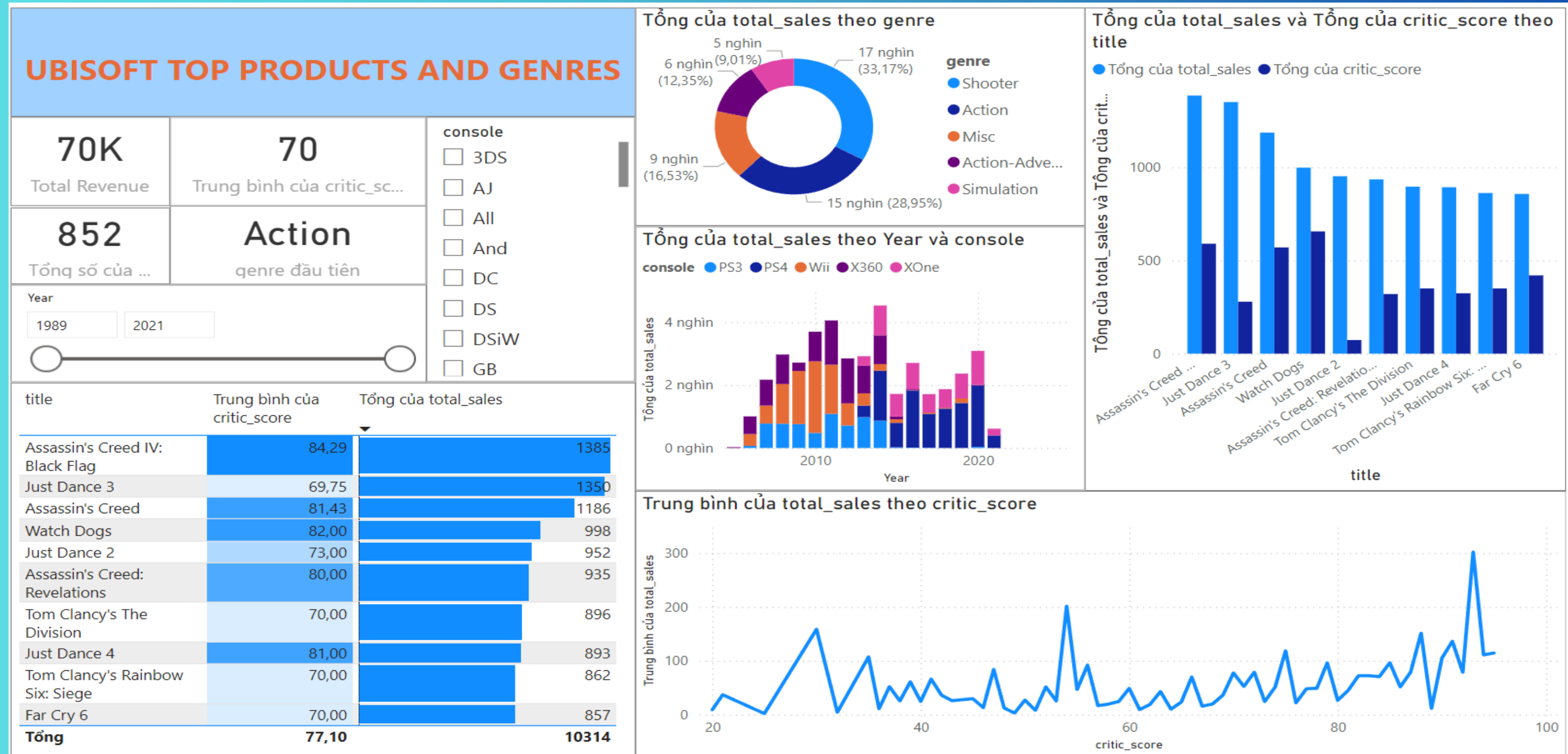
PHÂN TÍCH THỊ TRƯỜNG (POWER BI)

- **Thị phần:** Activision dẫn đầu về doanh thu.
- **Xu hướng:** Doanh thu Ubisoft đạt đỉnh 2010-2012 và đang đi ngang.
- **Nền tảng:** Doanh thu chủ yếu đến từ PlayStation và Xbox.

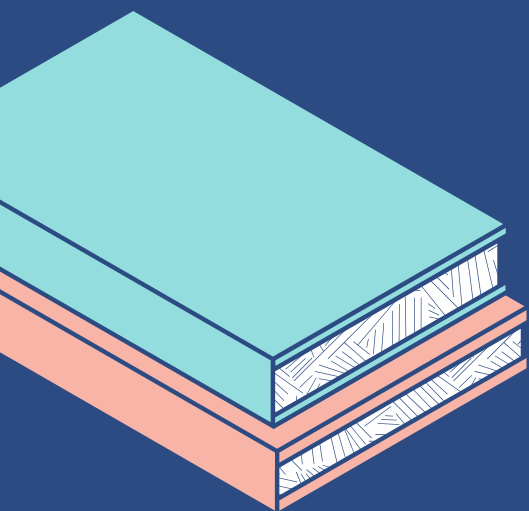


HIỆU SUẤT SẢN PHẨM (POWER BI)

- **Quy luật 80/20:** Doanh thu tập trung vào top 3 Franchise: *Assassin's Creed, Just Dance, Far Cry*.
- **Rủi ro:** Phụ thuộc quá lớn vào thể loại Hành động (Action) & Bắn súng (Shooter).

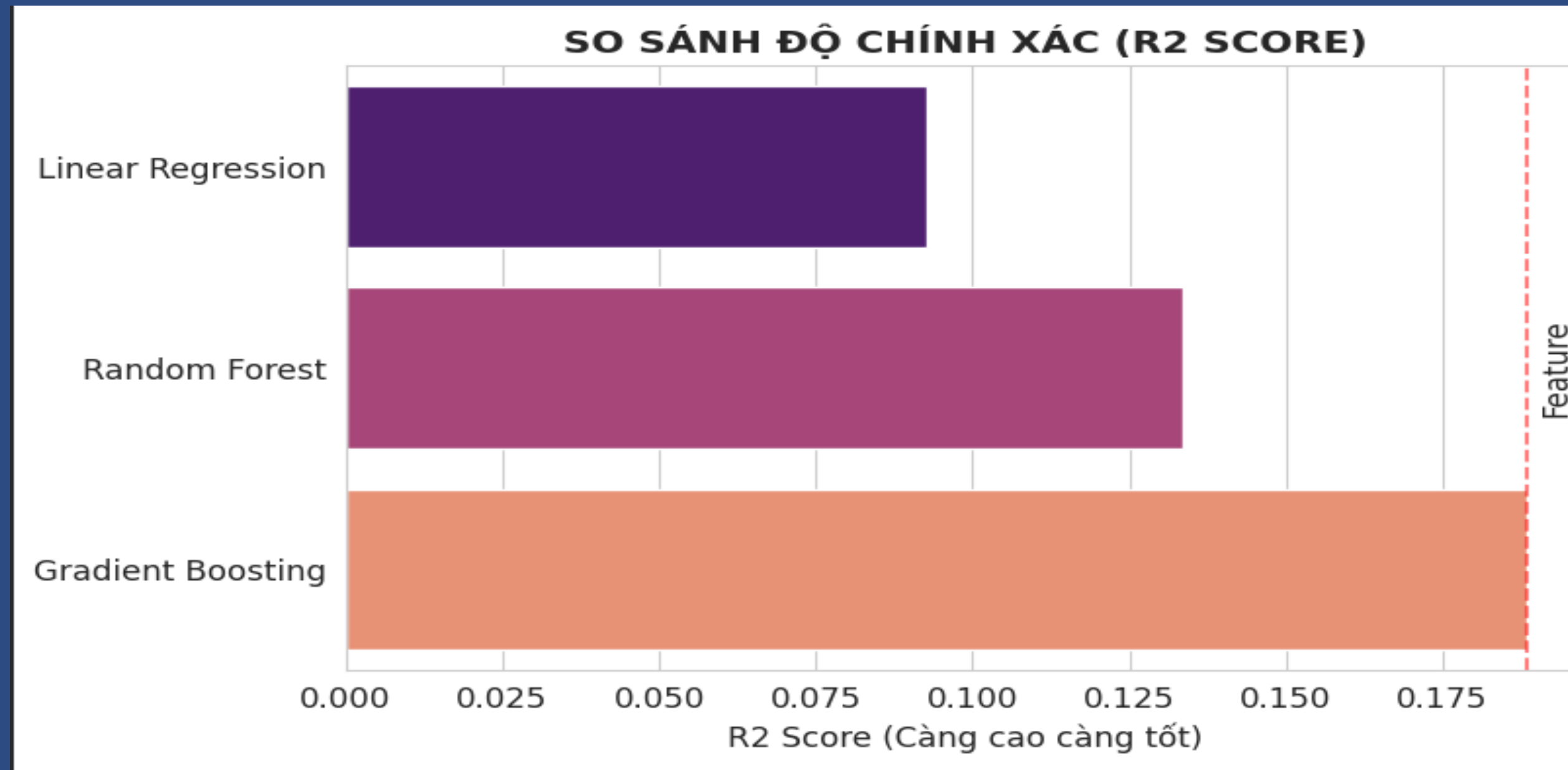


MÔ HÌNH PHÂN TÍCH DỰ ĐOÁN



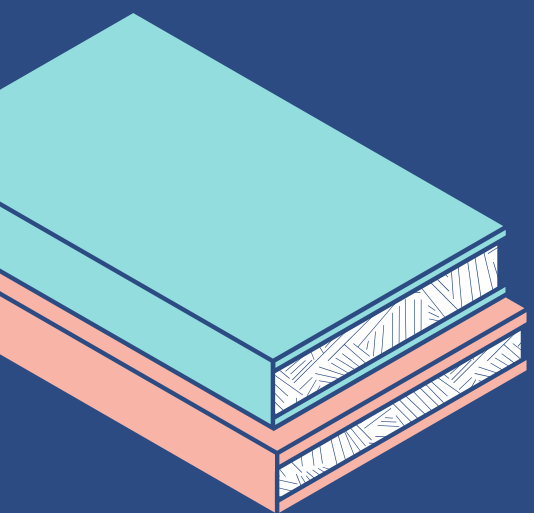
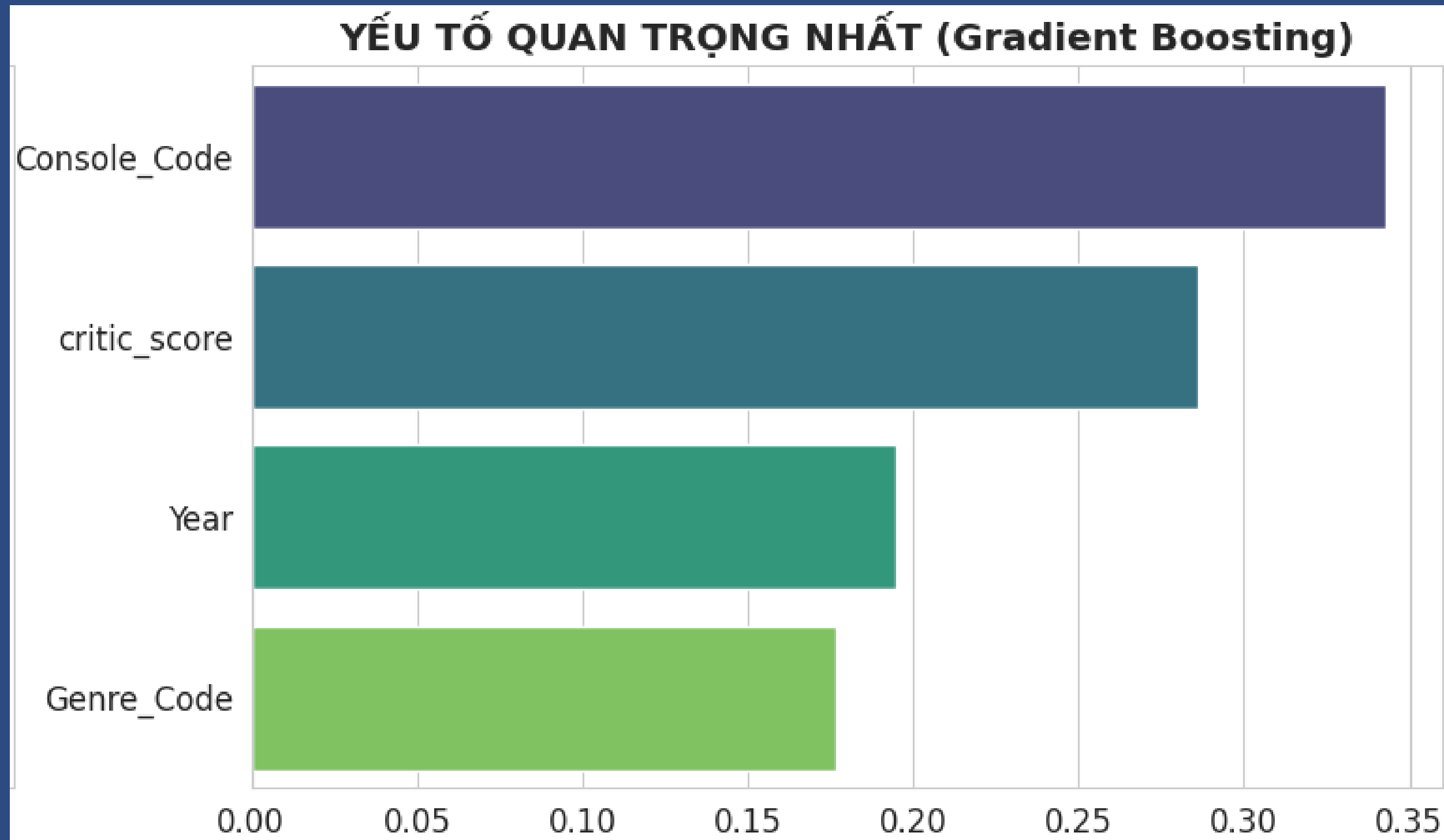
LỰA CHỌN MÔ HÌNH DỰ BÁO

- **Bài toán:** Dự báo doanh thu toàn cầu (Global_Sales).
- **Các thuật toán thử nghiệm:**
 1. Linear Regression (Cơ bản).
 2. Random Forest (Cây quyết định).
 3. Gradient Boosting (Nâng cao).
- **Kết quả:** Gradient Boosting tốt nhất (R2 Score ~ 18%).



CÁC YẾU TỐ ẢNH HƯỞNG (KEY DRIVERS)

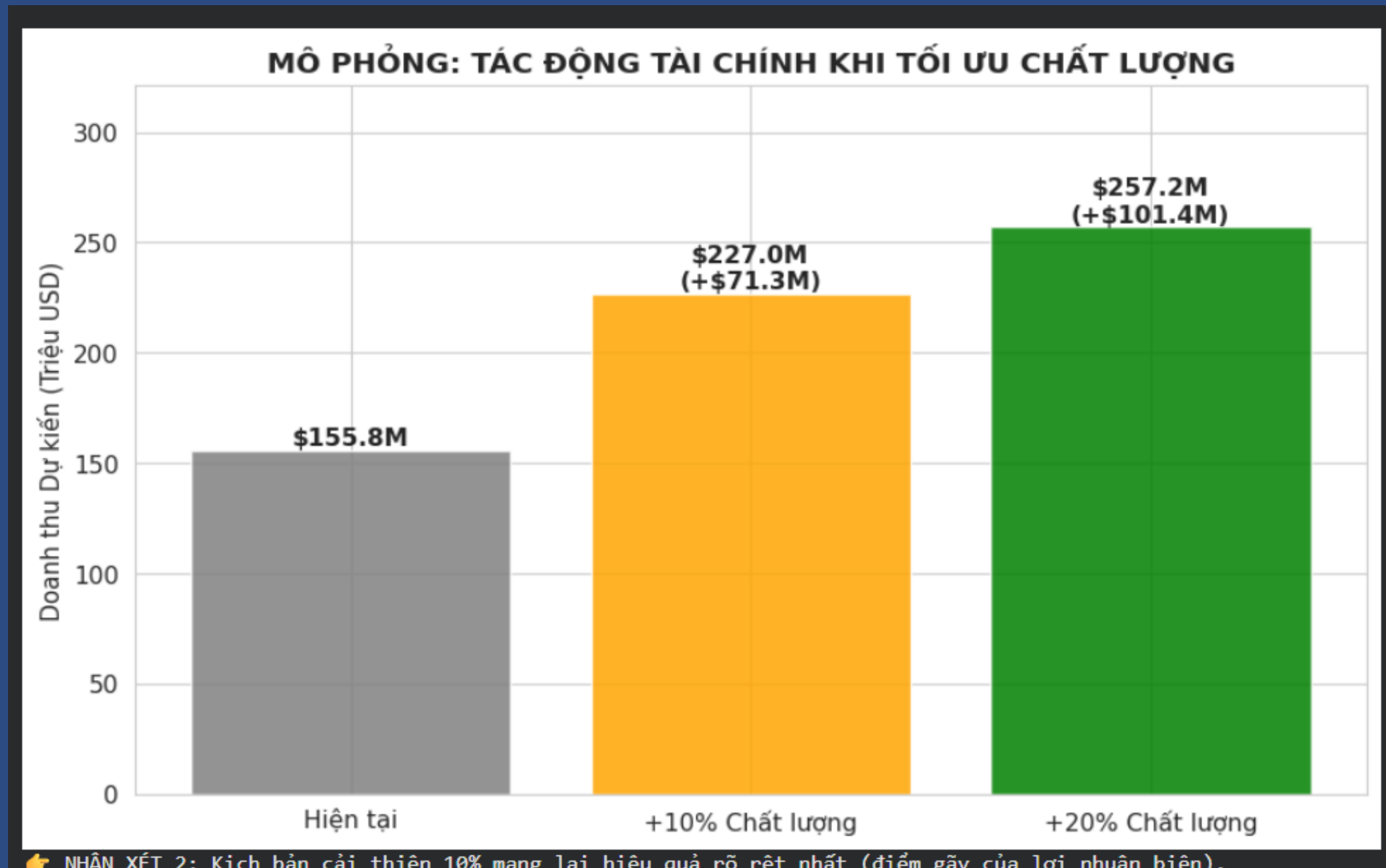
- **Câu hỏi:** Yếu tố nào quyết định game bán chạy?
- **Kết quả từ mô hình:**
 1. **Platform (Hệ máy):** Ảnh hưởng lớn nhất -> Phải chọn đúng sân chơi (Next-Gen).
 2. **Critic Score (Chất lượng):** Ảnh hưởng nhì -> Yếu tố cốt lõi để cạnh tranh.



MÔ PHỎNG CHIẾN LƯỢC (SIMULATION)

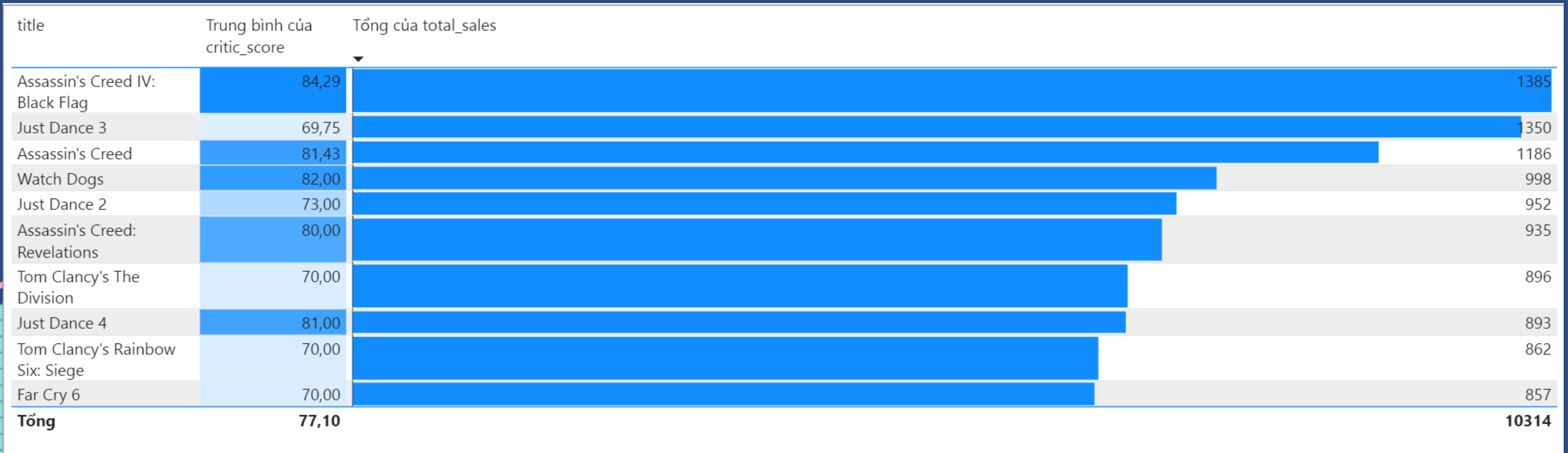
Kịch bản What-if: Nếu tăng chất lượng game (Score) thêm 10%?

- Kết quả dự báo:
- Doanh thu tăng thêm: [Số tiền] Triệu USD.
- ROI (Hoàn vốn) đạt mức cao.
- **Nhận định:** Đầu tư vào chất lượng mang lại lợi nhuận theo cấp số nhân.



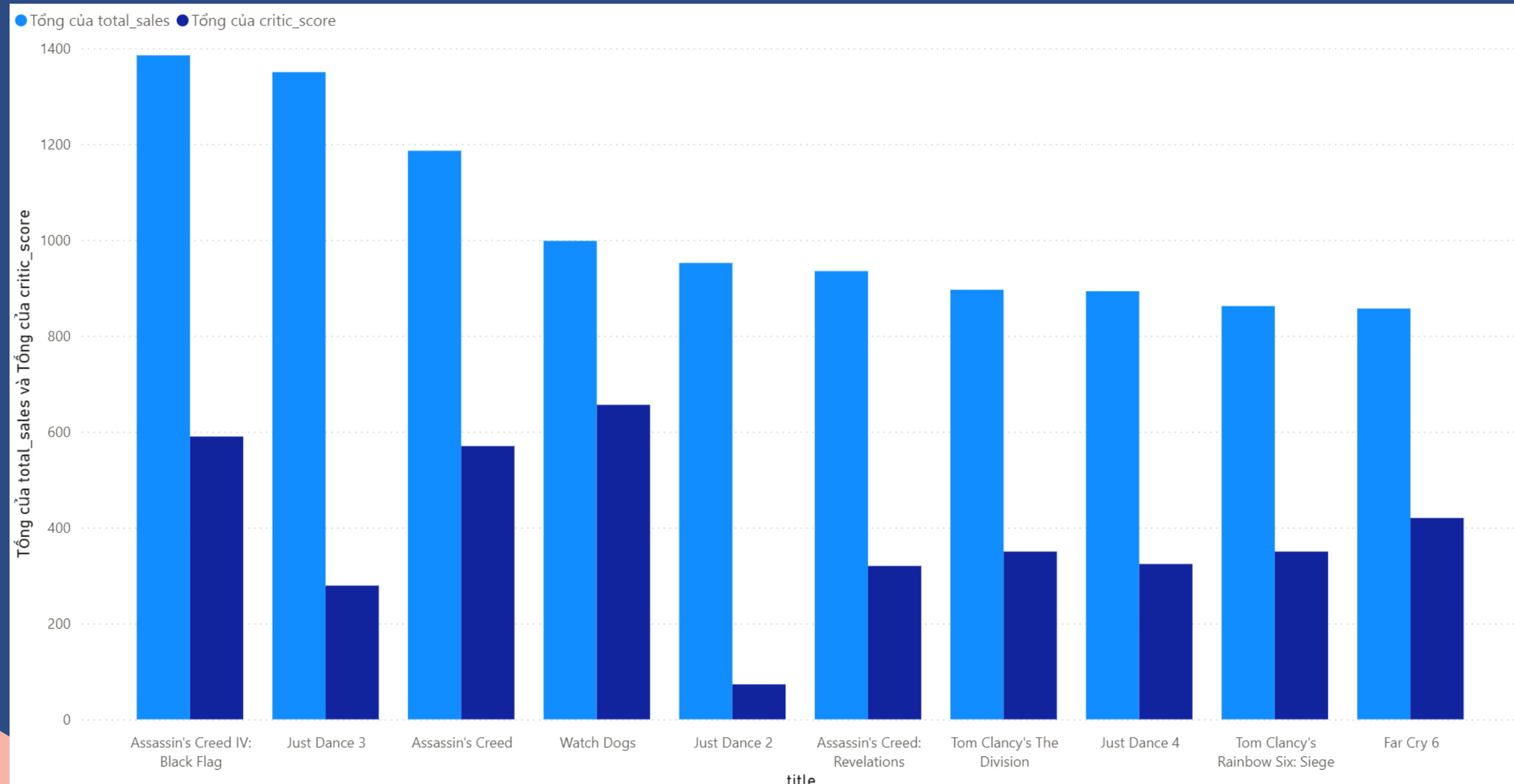
Kết luận

- Mô hình "Số lượng" (Spam nhiều game) đã lỗi thời.
- Cần chuyển dịch sang mô hình "Chất lượng" (Quality-First).
- Dữ liệu khẳng định: Điểm số cao = Doanh thu cao.



Kết luận

- Mô hình "Số lượng" (Spam nhiều game) đã lỗi thời.
- Cần chuyển dịch sang mô hình "Chất lượng" (Quality-First).
- Dữ liệu khẳng định: Điểm số cao = Doanh thu cao.



KHUYẾN NGHỊ QUẢN TRỊ

→ Chiến lược "Kiềng 3 chân":

1. **Nền tảng:** Ưu tiên PS5/Xbox Series X.
2. **Chất lượng:** KPI bắt buộc Critic Score > 80.
3. **R&D:** Tăng ngân sách kiểm thử (QA) lên 20%.



THANK YOU