

8b11ffbc-90a5-43bb-98d8-9bc190025f8d

March 17, 2025

1 Padrões para possíveis sucesso de vendas

Este projeto visa identificar padrões que determinam o sucesso de um jogo, analisando dados históricos, características do jogo e tendências de mercado. O objetivo é prever possíveis sucessos e planejar campanhas publicitárias mais eficazes, oferecendo insights valiosos para desenvolvedores e profissionais de marketing. Com uma abordagem baseada em dados, buscamos alinhar criatividade e estratégia para maximizar o impacto de lançamentos na indústria de games.

2 Patterns for Potential Sales Success

This project aims to identify patterns that determine the success of a game by analyzing historical data, game characteristics, and market trends. The objective is to predict potential successes and plan more effective marketing campaigns, providing valuable insights for developers and marketing professionals. By adopting a data-driven approach, we seek to align creativity and strategy to maximize the impact of game launches in the industry.

```
[1]: import pandas as pd
import math
from matplotlib import pyplot as plt
import seaborn as sns
import numpy as np
from scipy import stats as st
```

```
[2]: df = pd.read_csv("/datasets/games.csv")
```

2.1 Data Preparation for an Online Video Game Store

```
[3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16715 entries, 0 to 16714
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Name                   16713 non-null  object
1   Platform               16715 non-null  object
2   Year_of_Release       16446 non-null  float64
```

```

3   Genre                16713 non-null  object
4   NA_sales             16715 non-null  float64
5   EU_sales             16715 non-null  float64
6   JP_sales             16715 non-null  float64
7   Other_sales          16715 non-null  float64
8   Critic_Score         8137 non-null   float64
9   User_Score           10014 non-null  object
10  Rating               9949 non-null  object
dtypes: float64(6), object(5)
memory usage: 1.4+ MB

```

```
[4]: df.columns = df.columns.str.lower()
```

```
[5]: df = df.dropna(subset=['year_of_release'])
df['year_of_release'] = df['year_of_release'].astype(int)
```

Convertemos o tipo da coluna `year_of_release` de float para int, já que a coluna contém apenas o ano de lançamento. Essa mudança simplifica a análise, evitando que o programa crie automaticamente informações desnecessárias, como dias e horas, que ocorreriam se usássemos o tipo `datetime`.

We converted the `year_of_release` column from float to int, as the column only contains the release year. This change simplifies the analysis by preventing the program from automatically creating unnecessary information, such as days and hours, which would happen if we used the `datetime` type.

```
[6]: df['user_score'] = df['user_score'].replace("tbd", pd.NA)
df['user_score'] = pd.to_numeric(df['user_score'], errors='coerce')
```

Convertemos a coluna `user_score` de (str) para (float), para que fosse possível fazer cálculos, como médias e medianas. Como alguns valores estavam como ‘TBD’ (ou seja, ainda não tinham sido determinados), decidimos substituí-los por NaN. Isso ajuda a evitar distorções nas análises, já que esses valores não representam uma pontuação real.

We changed the `user_score` column from (str) to (float) so we could perform calculations like averages and medians. Since some values were marked as ‘TBD’ (to be determined), we decided to replace them with NaN. This helps avoid skewing the analysis, as these values don’t represent actual scores.

```
[7]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 16446 entries, 0 to 16714
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                  16444 non-null  object
1   platform              16446 non-null  object
2   year_of_release       16446 non-null  int64
3   genre                  16444 non-null  object

```

```

4   na_sales      16446 non-null  float64
5   eu_sales      16446 non-null  float64
6   jp_sales      16446 non-null  float64
7   other_sales   16446 non-null  float64
8   critic_score   7983 non-null  float64
9   user_score    7463 non-null  float64
10  rating        9768 non-null  object
dtypes: float64(6), int64(1), object(4)
memory usage: 1.5+ MB

```

```

[8]: df['total_sales'] = df['na_sales'] + df['eu_sales'] + df['jp_sales'] +
      ↪df['other_sales']
      df.head(5)

```

```

[8]:
      name platform  year_of_release  genre  na_sales  \
0      Wii Sports      Wii          2006  Sports    41.36
1  Super Mario Bros.      NES          1985  Platform    29.08
2    Mario Kart Wii      Wii          2008   Racing    15.68
3  Wii Sports Resort      Wii          2009   Sports    15.61
4  Pokemon Red/Pokemon Blue      GB          1996  Role-Playing    11.27

      eu_sales  jp_sales  other_sales  critic_score  user_score  rating  \
0      28.96     3.77         8.45          76.0         8.0      E
1       3.58     6.81         0.77           NaN         NaN     NaN
2      12.76     3.79         3.29          82.0         8.3      E
3       10.93     3.28         2.95          80.0         8.0      E
4       8.89    10.22         1.00           NaN         NaN     NaN

      total_sales
0         82.54
1         40.24
2         35.52
3         32.77
4         31.38

```

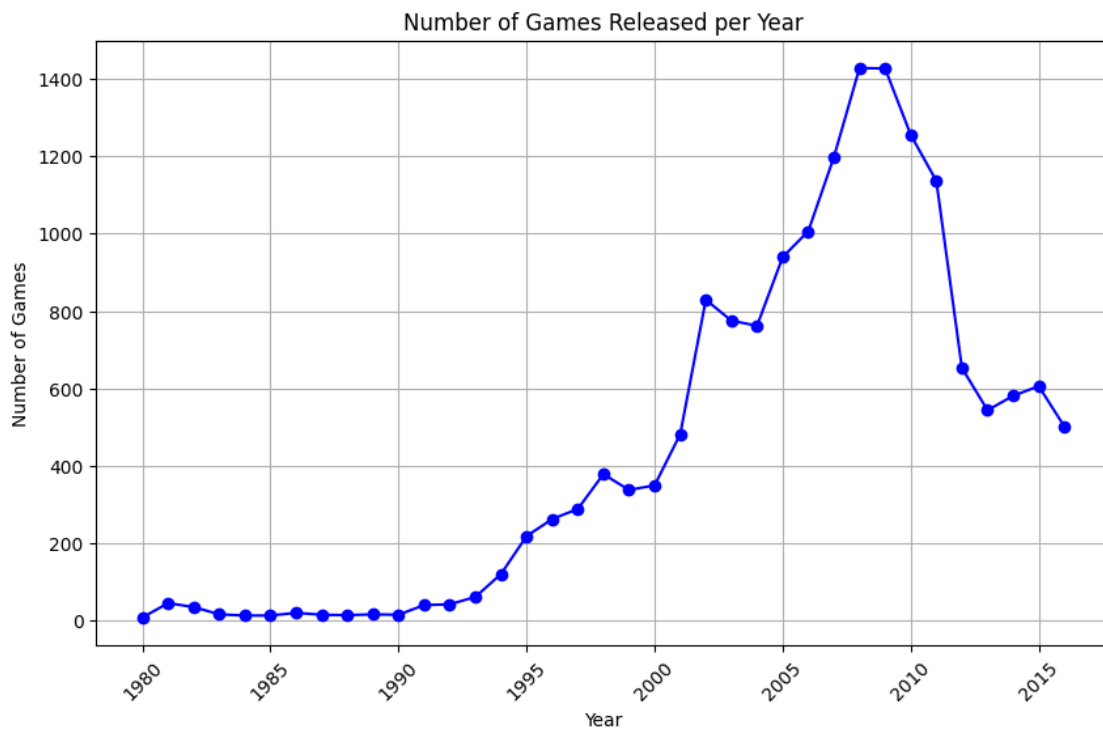
Criamos uma nova coluna nos dados que exibe o total de vendas de cada jogo. Essa coluna foi calculada somando as vendas de todas as regiões (América do Norte, Europa, Japão e outras regiões)

We added a new column to the data showing the total sales for each game. This column was calculated by summing the sales from all regions (North America, Europe, Japan, and other regions) for each individual game.

2.2 Online Video Game Retail Insights

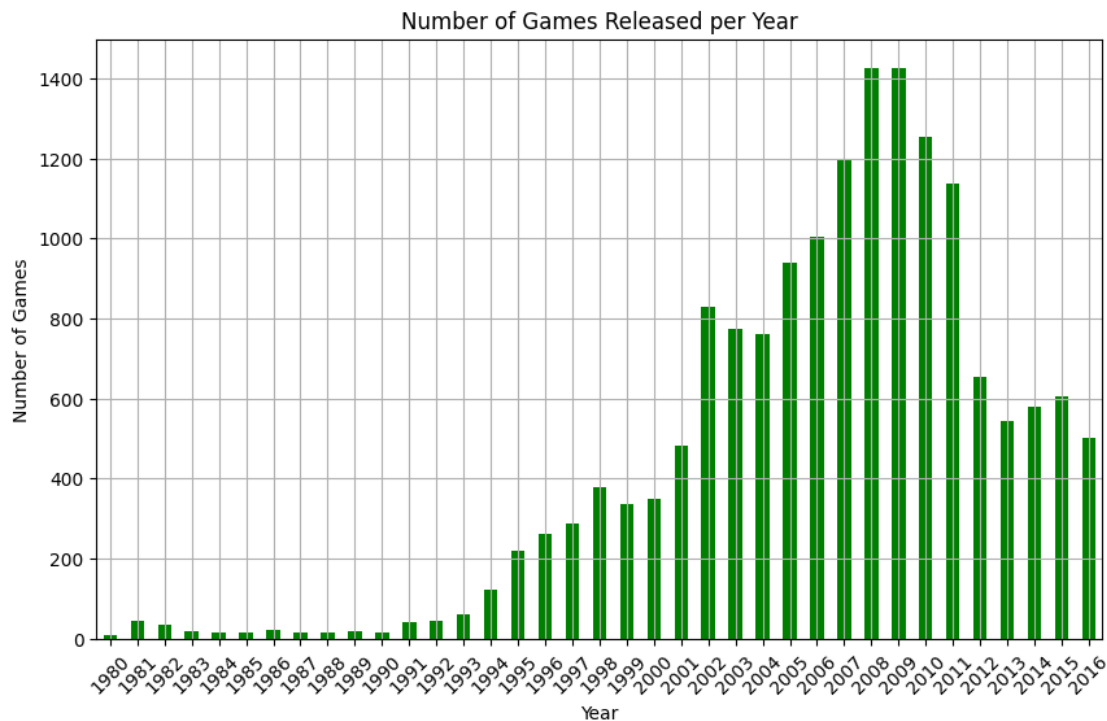
```
[9]: df = df.dropna(subset=['year_of_release'])  
games_per_year = df['year_of_release'].value_counts().sort_index()
```

```
[10]: games_per_year.plot(title = 'Number of Games Released per Year',  
                           xlabel = 'Year',  
                           ylabel = 'Number of Games',  
                           rot = 45,  
                           marker = 'o',  
                           grid = True,  
                           figsize = (10,6),  
                           color = 'blue')  
  
plt.show()
```



```
[34]: games_per_year.plot(kind = 'bar',  
                           title = 'Number of Games Released per Year',  
                           xlabel = 'Year',  
                           ylabel = 'Number of Games',  
                           rot = 45,  
                           grid = True,  
                           figsize = (10,6),  
                           color = 'green')
```

```
plt.show()
```



Podemos observar um grande crescimento no número de lançamentos de jogos ao longo dos anos, com um salto significativo entre 1998 e 2008, quando o número de lançamentos passou de cerca de 400 para mais de 1.400 em um único ano. Esse crescimento pode ser explicado pelos novos consoles, inovações no mundo dos games e avanços tecnológicos. A partir de 2009, o número de lançamentos se estabilizou próximo ao patamar do ano anterior, mas, em seguida, começou a cair, retornando quase ao nível de 1998. Essa queda pode estar relacionada a crises econômicas, mudanças no mercado ou até mesmo à evolução da tecnologia, que pode ter alterado a forma como os jogos são produzidos e distribuídos.

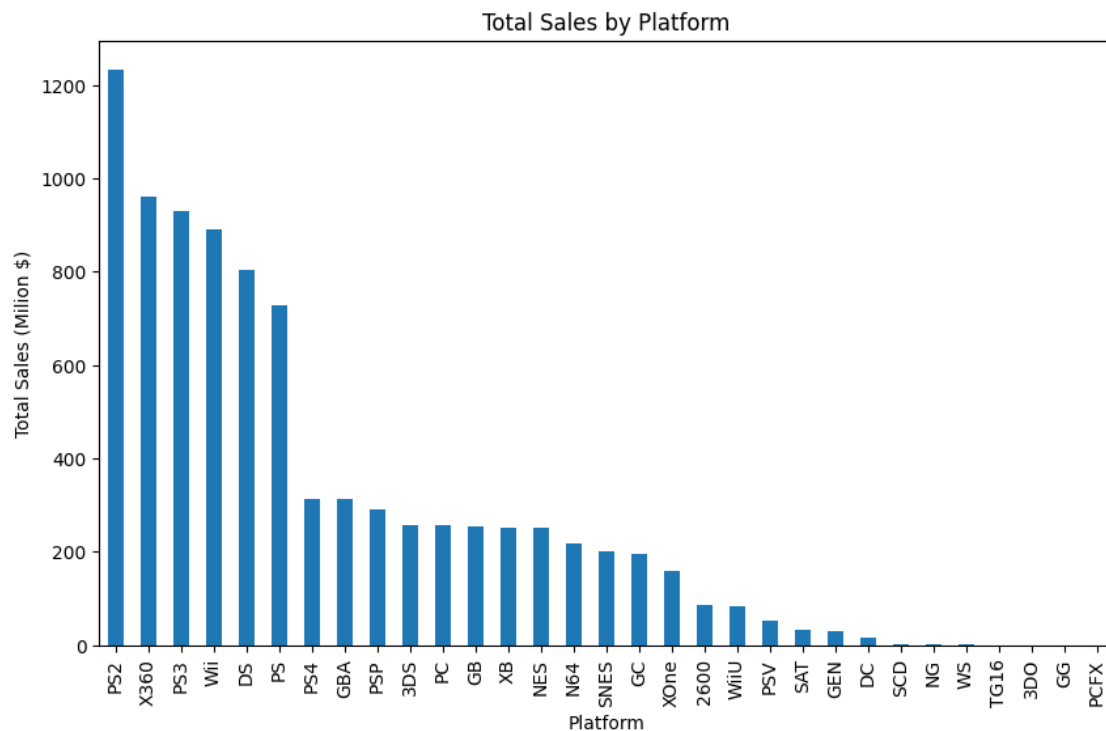
We can observe significant growth in the number of game releases over the years, with a notable surge between 1998 and 2008, when the number of releases jumped from around 400 to over 1,400 in a single year. This growth can be attributed to the launch of new consoles, innovations in the gaming world, and technological advancements. Starting in 2009, the number of releases stabilized near the previous year's level but then began to decline, almost returning to the 1998 figures. This drop may be linked to economic crises, market changes, or even advancements in technology that have altered how games are produced and distributed.

```
[11]: platform_sales = df.groupby('platform')['total_sales'].sum().  
      ↪sort_values(ascending=False)  
platform_sales.head(10)
```

```
[11]: platform
      PS2      1233.56
      X360      961.24
      PS3      931.34
      Wii      891.18
      DS       802.78
      PS       727.58
      PS4      314.14
      GBA      312.88
      PSP      289.53
      3DS      257.81
      Name: total_sales, dtype: float64
```

```
[12]: platform_sales.plot(kind = 'bar',
                           xlabel = 'Platform',
                           ylabel = 'Total Sales (Milion $)',
                           title = 'Total Sales by Platform',
                           figsize = (10,6))

plt.show()
```



```
[13]:
```

```
platform_sales_year = df.groupby(['platform',
    ↳ 'year_of_release'])['total_sales'].sum().reset_index().sort_values(by =
    ↳ 'total_sales', ascending=False)
platform_sales_year.head(10)
```

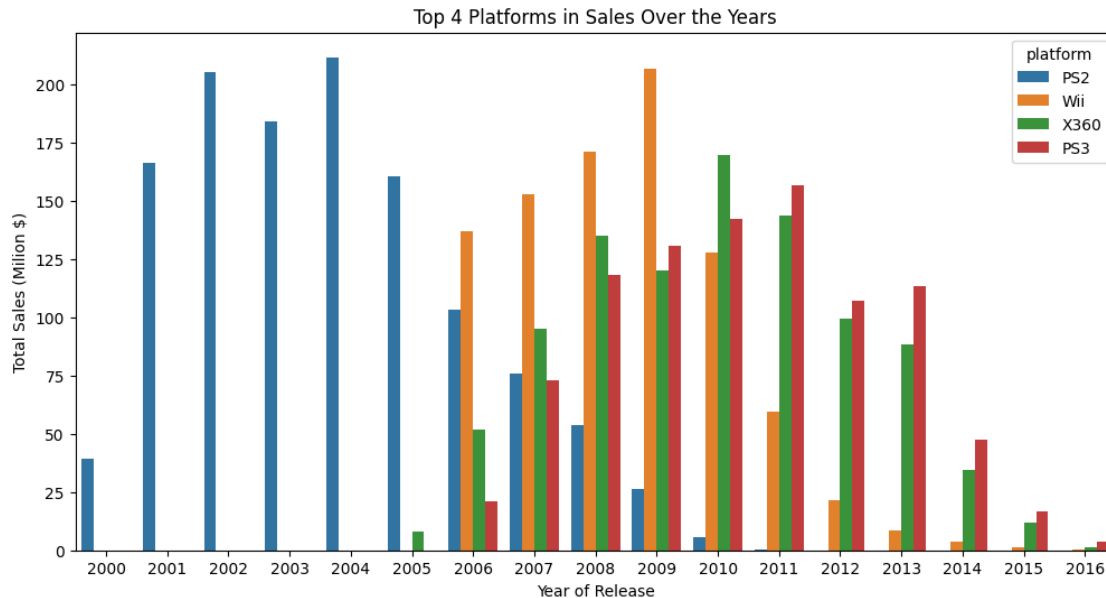
```
[13]:
```

	platform	year_of_release	total_sales
134	PS2	2004	211.81
200	Wii	2009	206.97
132	PS2	2002	205.38
133	PS2	2003	184.31
199	Wii	2008	171.32
218	X360	2010	170.03
124	PS	1998	169.49
131	PS2	2001	166.43
135	PS2	2005	160.66
147	PS3	2011	156.78

```
[14]: platform_sales_total = df.groupby('platform')['total_sales'].sum().
    ↳ sort_values(ascending=False)
top_4_platforms = platform_sales_total.head(4).index
top_4_data = platform_sales_year[platform_sales_year['platform'].
    ↳ isin(top_4_platforms)]

plt.figure(figsize=(12, 6))
sns.barplot(data=top_4_data, x='year_of_release', y='total_sales',
    ↳ hue='platform')
plt.xlabel('Year of Release')
plt.ylabel('Total Sales (Million $)')
plt.title('Top 4 Platforms in Sales Over the Years')

plt.show()
```



O PS2 dominou o mercado por 6 anos, de 2000 a 2005, sendo uma das quatro maiores plataformas em vendas. Após o lançamento do Wii, PS3 e X360, novas plataformas entraram em cena, iniciando uma disputa acirrada pela liderança. O X360 assumiu a primeira posição apenas em 2010, enquanto o Wii teve um desempenho impressionante, liderando as vendas por quatro anos consecutivos (2006-2009) e superando o X360, PS2 e PS3. Durante esse período, o Wii registrou um crescimento constante em suas vendas.

A partir de 2010, o PS3 assumiu a liderança, mas, ao contrário do Wii, suas vendas caíram ano após ano. Vale destacar que nenhuma plataforma conseguiu superar o recorde de vendas do PS2 em 2004, que atingiu 211 milhões. O Wii chegou perto desse valor em 2009, com 206 milhões.

O surgimento de uma nova plataforma depende muito das condições do mercado e dos avanços tecnológicos. Em média, leva cerca de 5 anos para uma nova plataforma surgir e dominar o mercado, enquanto o ciclo de vida de uma plataforma (até seu “desaparecimento”) pode durar até 10 anos, como foi o caso do PS2 e do Wii, que mantiveram vendas significativas durante esse período.

The PS2 dominated the market for 6 years, from 2000 to 2005, standing out as one of the top four platforms in sales. After the release of the Wii, PS3, and X360, new platforms emerged, sparking a fierce competition for leadership. The X360 took the top spot only in 2010, while the Wii delivered an impressive performance, leading sales for four consecutive years (2006-2009) and outperforming the X360, PS2, and PS3. During this period, the Wii saw steady growth in its sales.

Starting in 2010, the PS3 took the lead, but unlike the Wii, its sales declined year after year. It’s worth noting that no platform managed to surpass the PS2’s sales record in 2004, which reached 211 million. Wii came close to this value in 2009 with 206 million.

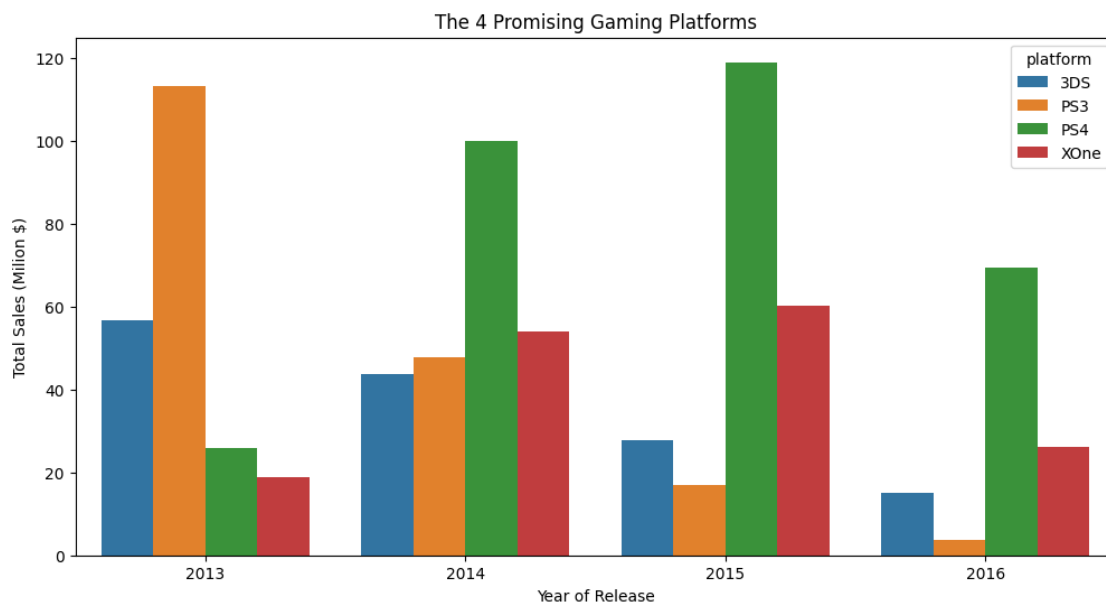
The emergence of a new platform heavily depends on market conditions and technological advancements. On average, it takes about 5 years for a new platform to emerge and dominate the market, while a platform’s lifecycle (until its “disappearance”) can last up to 10 years, as seen with the PS2

and Wii, which maintained significant sales during this period.

```
[15]: future_platforms_sales_year = (df[(df['year_of_release'] >= 2013) &
    ↳ (df['year_of_release'] <= 2016)].groupby(['platform',
    ↳ 'year_of_release'])['total_sales'].sum().reset_index())
top_4_promising_platforms = (future_platforms_sales_year.
    ↳ groupby('platform')['total_sales'].sum().sort_values(ascending=False).
    ↳ head(4).index)
future_4_data =
    ↳ future_platforms_sales_year[future_platforms_sales_year['platform'].
    ↳ isin(top_4_promising_platforms)]

plt.figure(figsize=(12, 6))
sns.barplot(data=future_4_data, x='year_of_release', y='total_sales',
    ↳ hue='platform')
plt.xlabel('Year of Release')
plt.ylabel('Total Sales (Milion $)')
plt.title('The 4 Promising Gaming Platforms')

plt.show()
```



Neste gráfico, estamos analisando as plataformas mais promissoras para os próximos anos, com foco em 2017. Podemos observar que, entre 2014 e 2016, o PS4 liderou as vendas no mercado de games, mantendo uma diferença significativa em relação à segunda plataforma mais promissora, o XOne. Esse desempenho sólido do PS4 indica uma grande chance de continuar sendo uma plataforma dominante nos próximos anos.

Por outro lado, o 3DS e o PS3 vêm registrando quedas sucessivas em suas vendas ao longo dos anos.

Essa tendência sugere que, se nada mudar, essas plataformas correm o risco de serem esquecidas ou substituídas por tecnologias mais recentes nos próximos anos.

In this chart, we are analyzing the most promising platforms for the coming years, with a focus on 2017. We can see that between 2014 and 2016, the PS4 led the sales in the gaming market, maintaining a significant gap over the second most promising platform, the XOne. This strong performance by the PS4 indicates a high likelihood of remaining a dominant platform in the coming years.

On the other hand, the 3DS and PS3 have seen successive declines in their sales over the years. This trend suggests that, if nothing changes, these platforms risk being forgotten or replaced by newer technologies in the near future.

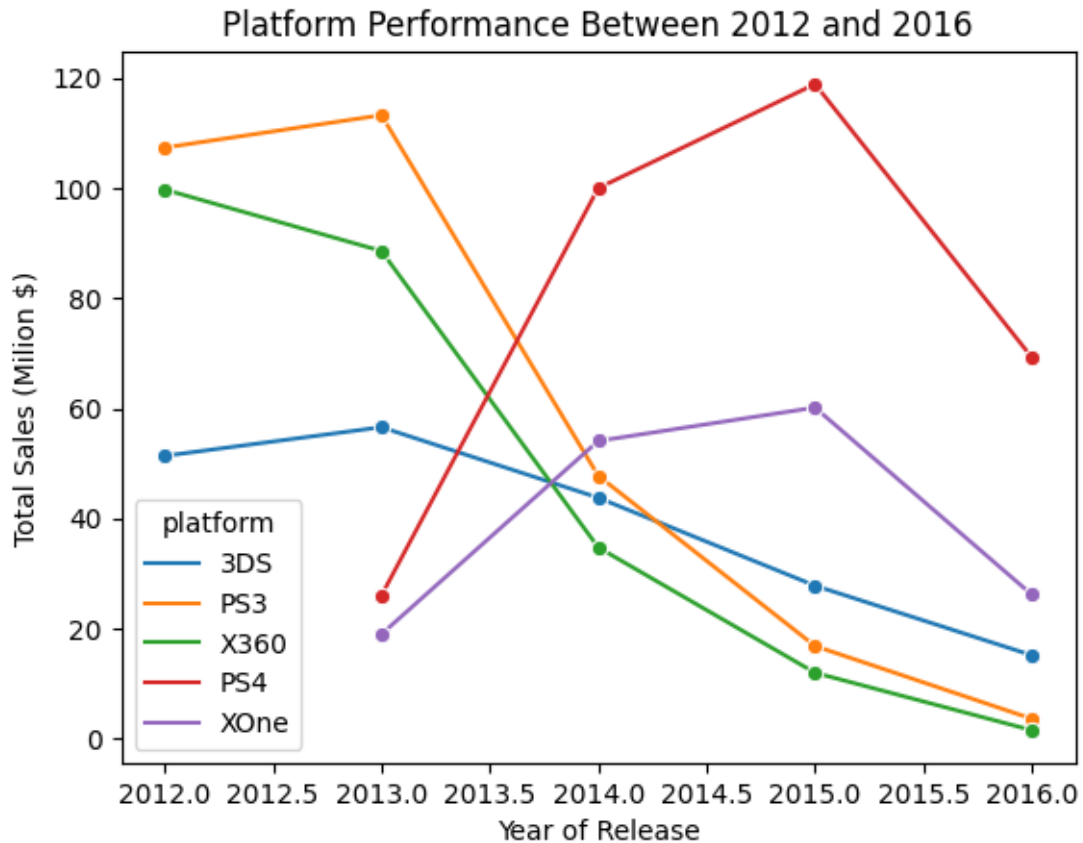
```
[16]: best_platforms_sales = (df[(df['year_of_release'] >= 2012) &
    ↳ (df['year_of_release'] <= 2016)])
best_sales = best_platforms_sales.
    ↳ groupby(['year_of_release', 'platform'])['total_sales'].sum().reset_index()
top_platforms = best_sales.groupby('platform')['total_sales'].sum().
    ↳ sort_values(ascending=False).head(5).index
best_sales_top = best_sales[best_sales['platform'].isin(top_platforms)]
best_sales_top.sort_values(by = ['year_of_release', 'total_sales'], ascending =
    ↳ False)
```

```
[16]:
```

	year_of_release	platform	total_sales
43	2016	PS4	69.25
48	2016	XOne	26.15
40	2016	3DS	15.14
42	2016	PS3	3.60
47	2016	X360	1.52
33	2015	PS4	118.90
39	2015	XOne	60.14
30	2015	3DS	27.78
32	2015	PS3	16.82
38	2015	X360	11.96
23	2014	PS4	100.00
29	2014	XOne	54.07
22	2014	PS3	47.76
20	2014	3DS	43.76
28	2014	X360	34.74
12	2013	PS3	113.25
18	2013	X360	88.58
9	2013	3DS	56.57
13	2013	PS4	25.99
19	2013	XOne	18.96
3	2012	PS3	107.36
8	2012	X360	99.74
0	2012	3DS	51.36

```
[17]: sns.lineplot(data=best_sales_top, x='year_of_release', y='total_sales',
    ↪hue='platform', marker='o')

plt.xlabel('Year of Release')
plt.ylabel('Total Sales (Milion $)')
plt.title('Platform Performance Between 2012 and 2016')
plt.figure(figsize = (12,6))
plt.show()
```



<Figure size 1200x600 with 0 Axes>

Em uma análise do período de 2012 a 2016, observamos que, nos dois primeiros anos (2012 e 2013), o PS3 foi o líder de vendas, apresentando um leve crescimento. No entanto, após o lançamento do PS4, as vendas do PS3 caíram quase 60%, perdendo a liderança para a nova plataforma. Desde então, o PS4 e o XOne, as plataformas mais recentes, dominaram o mercado, mas também registraram quedas nas vendas a partir de 2015.

Entre as 5 plataformas mais vendidas nesse período, nenhuma apresentou crescimento nas vendas. Pelo contrário, todas tiveram quedas significativas, algumas chegando a quase 90%, como o X360, que vendeu 11,96 milhões em 2015 e apenas 1,52 milhões em 2016, correndo o risco de “desaparecer” do mercado.

Se formos falar de plataformas com potencial lucrativo, o PS4 e o XOne, mesmo com quedas nas vendas, ainda se destacam. No entanto, é importante investigar se essas quedas foram causadas por desinteresse do público, crises econômicas ou outros fatores. Uma nova plataforma, bem estudada e elaborada, poderia ter um grande potencial de vendas nos próximos anos, como ocorreu com o lançamento do PS4, que trouxe um crescimento nas vendas de quase 300% em seu primeiro ano.

In an analysis of the period from 2012 to 2016, we observed that in the first two years (2012 and 2013), the PS3 was the sales leader, showing slight growth. However, after the launch of the PS4, PS3 sales dropped by almost 60%, losing the lead to the new platform. Since then, PS4 and XOne, the newer platforms, have dominated the market, but they also experienced sales declines starting in 2015.

Among the top 5 best-selling platforms during this period, none showed sales growth. On the contrary, all of them saw significant declines, with some dropping by almost 90%, such as the X360, which sold 11.96 million in 2015 and only 1.52 million in 2016, risking “disappearing” from the market.

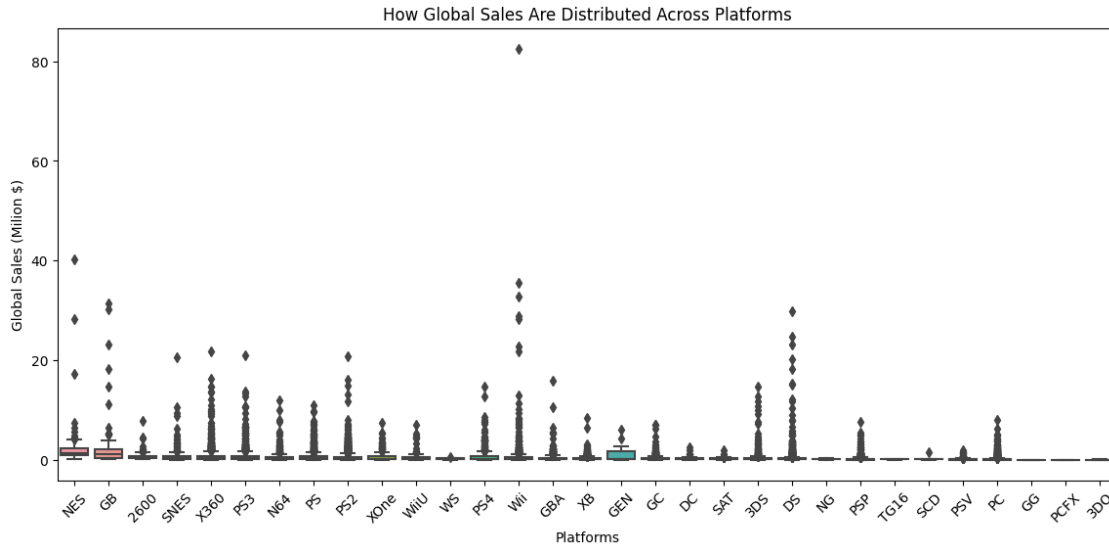
If we talk about platforms with profit potential, the PS4 and XOne, despite their sales declines, still stand out. However, it is important to investigate whether these declines were caused by a lack of consumer interest, economic crises, or other factors. A new platform, well-researched and designed, could have great sales potential in the coming years, as seen with the launch of the PS4, which brought a sales increase of nearly 300% in its first year.

```
[18]: df_filtered = df[df['total_sales'] > 0]
platform_order = df_filtered.groupby('platform')['total_sales'].median().
    ↪sort_values(ascending=False).index

plt.figure(figsize=(14, 6))
sns.boxplot(data=df_filtered, x='platform', y='total_sales',
    ↪order=platform_order)

plt.xlabel('Platforms')
plt.ylabel('Global Sales (Milion $)')
plt.title('How Global Sales Are Distributed Across Platforms')
plt.xticks(rotation=45)

plt.show()
```



As vendas globais dos jogos variam bastante entre as plataformas, como podemos ver no gráfico de caixa. Algumas plataformas têm uma distribuição mais compacta, o que significa que as vendas dos jogos são mais parecidas entre si. Já outras plataformas mostram uma variação maior, com algumas vendas muito altas e outras bem baixas.

Além disso, vemos que há outliers elevados, ou seja, algumas plataformas têm jogos com vendas extremamente altas, mas também com vendas muito baixas. As diferenças nas medianas mostram como as vendas médias variam entre as plataformas, com algumas sendo mais consistentes, enquanto outras têm um desempenho mais instável. Isso reflete a diversidade de sucesso dos jogos em cada plataforma.

Global sales of games vary significantly across platforms, as we can see in the box plot. Some platforms have a more compact distribution, meaning that game sales are more similar to each other. On the other hand, other platforms show greater variation, with some games having very high sales and others very low sales.

Additionally, we see that there are high outliers, meaning that some platforms have games with extremely high sales, but also games with very low sales. The differences in medians show how average sales vary between platforms, with some being more consistent, while others have more unstable performance. This reflects the diversity of success among games on each platform.

```
[19]: platform_means = df_filtered.groupby('platform')['total_sales'].mean().
      ↪sort_values(ascending=False)
      platform_means.head(10)
```

```
[19]: platform
      GB      2.622990
      NES     2.561735
      GEN     1.061034
      SNES     0.836987
```

```
PS4      0.801378
X360     0.780227
2600     0.745517
PS3      0.713670
Wii      0.692986
N64      0.689905
Name: total_sales, dtype: float64
```

Ao analisarmos as vendas médias das plataformas, vemos que o Game Boy (GB) e o Nintendo Entertainment System (NES) se destacam com números muito mais altos, o que indica que esses consoles foram grandes sucessos de vendas. Por outro lado, plataformas mais recentes, como o PlayStation 4 (PS4) e o Xbox One (XOne), apresentam vendas médias mais baixas, embora ainda estejam entre as mais populares.

Em resumo, as plataformas mais antigas continuam a liderar quando se trata de vendas médias, enquanto as mais novas enfrentam uma concorrência maior e têm resultados mais modestos.

When analyzing the average sales of platforms, we see that the Game Boy (GB) and the Nintendo Entertainment System (NES) stand out with significantly higher numbers, indicating that these consoles were major sales successes. On the other hand, more recent platforms, such as the PlayStation 4 (PS4) and the Xbox One (XOne), show lower average sales, even though they remain among the most popular.

In summary, older platforms continue to lead when it comes to average sales, while newer ones face greater competition and achieve more modest results.

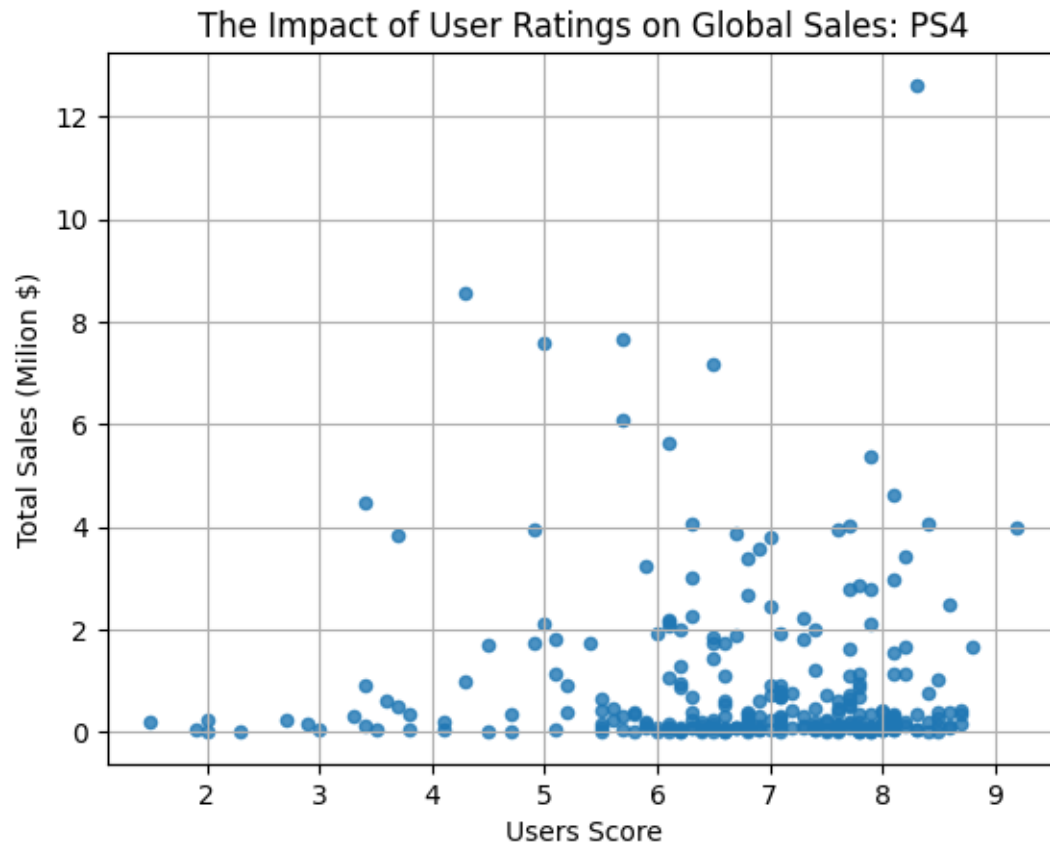
```
[20]: data_ps4 = df[df['platform'] == 'PS4']
      corr_user_sales = data_ps4[['user_score', 'total_sales']].corr().iloc[0, 1]
      corr_critic_sales = data_ps4[['critic_score', 'total_sales']].corr().iloc[0, 1]

      print(corr_user_sales)
      print(corr_critic_sales)
```

```
-0.03195711020455685
0.40656790206178123
```

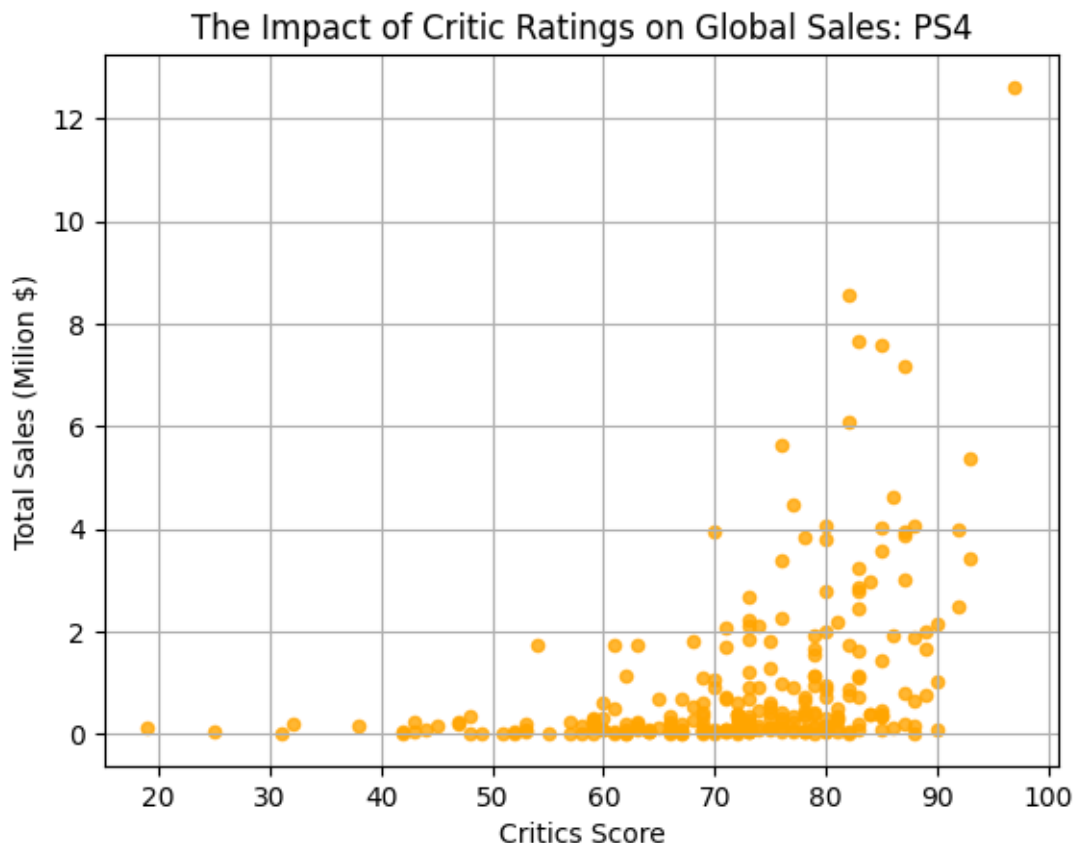
```
[21]: data_ps4.plot(x= 'user_score', y= 'total_sales', kind='scatter', alpha=0.8)
      plt.title('The Impact of User Ratings on Global Sales: PS4')
      plt.xlabel('Users Score')
      plt.ylabel('Total Sales (Milion $)')
      plt.grid(True)

      plt.show()
```



```
[22]: data_ps4.plot(x= 'critic_score', y= 'total_sales', kind='scatter', alpha=0.8,
    ↪color='orange')
plt.title('The Impact of Critic Ratings on Global Sales: PS4')
plt.xlabel('Critics Score')
plt.ylabel('Total Sales (Million $)')
plt.grid(True)

plt.show()
```



Podemos observar que, quando as notas dos críticos são mais altas, as vendas dos jogos também tendem a ser maiores. Isso sugere que, para a plataforma PS4, jogos com avaliações positivas dos críticos geralmente têm um bom desempenho em vendas globais. No entanto, o valor da correlação não é tão alto para afirmar que as vendas elevadas são totalmente influenciadas pelas notas dos críticos. Isso indica que outros fatores, como marketing, popularidade da franquia ou apelo comercial, também desempenham um papel importante no sucesso das vendas.

Por outro lado, a relação entre as notas dos usuários e as vendas não é forte. Isso nos mostra que, para o PS4, as avaliações dos usuários não têm uma influência direta significativa nas vendas dos jogos. Jogos com avaliações altas ou baixas podem ter vendas semelhantes, sugerindo que fatores externos, como campanhas de marketing eficazes ou o apelo de franquias já consolidadas, podem ter um impacto maior no desempenho comercial.

We can observe that when critics' scores are higher, game sales also tend to be higher. This suggests that, for the PS4 platform, games with positive reviews from critics generally perform well in global sales. However, the correlation value is not high enough to claim that high sales are entirely due to critics' scores. This indicates that other factors, such as marketing, franchise popularity, or commercial appeal, also play a significant role in driving sales.

On the other hand, the relationship between user ratings and sales is not strong. This shows that, for the PS4, user reviews do not have a significant direct influence on game sales. Games with high or low user ratings can have similar sales, suggesting that external factors, such as

effective marketing campaigns or the appeal of established franchises, may have a greater impact on commercial performance.

```
[23]: games = df['name'].value_counts()
      games.head(10)
```

```
[23]: Need for Speed: Most Wanted      12
      Ratatouille                      9
      FIFA 14                         9
      LEGO Marvel Super Heroes        9
      FIFA Soccer 13                   8
      Madden NFL 08                   8
      Angry Birds Star Wars           8
      Madden NFL 07                   8
      FIFA 15                         8
      The LEGO Movie Videogame        8
      Name: name, dtype: int64
```

```
[24]: game_fifa15 = df[df['name'] == 'FIFA 15']

      print(game_fifa15[['platform', 'total_sales']])
```

	platform	total_sales
143	PS4	6.08
264	PS3	4.28
494	X360	2.92
754	XOne	2.18
3381	PSV	0.60
3611	Wii	0.56
4279	3DS	0.46
6097	PC	0.29

```
[25]: game_fifa14 = df[df['name'] == 'FIFA 14']

      print(game_fifa14[['platform', 'total_sales']])
```

	platform	total_sales
126	PS3	6.46
271	X360	4.22
469	PS4	3.01
1753	XOne	1.16
4699	PSV	0.41
4809	PC	0.40
4985	Wii	0.38
6871	3DS	0.23
7875	PSP	0.19

No exemplo do jogo FIFA 14, que escolhemos para análise, observamos que a plataforma com mais vendas foi o PS3, com um total de 6,46 milhões de dólares, representando uma diferença de mais

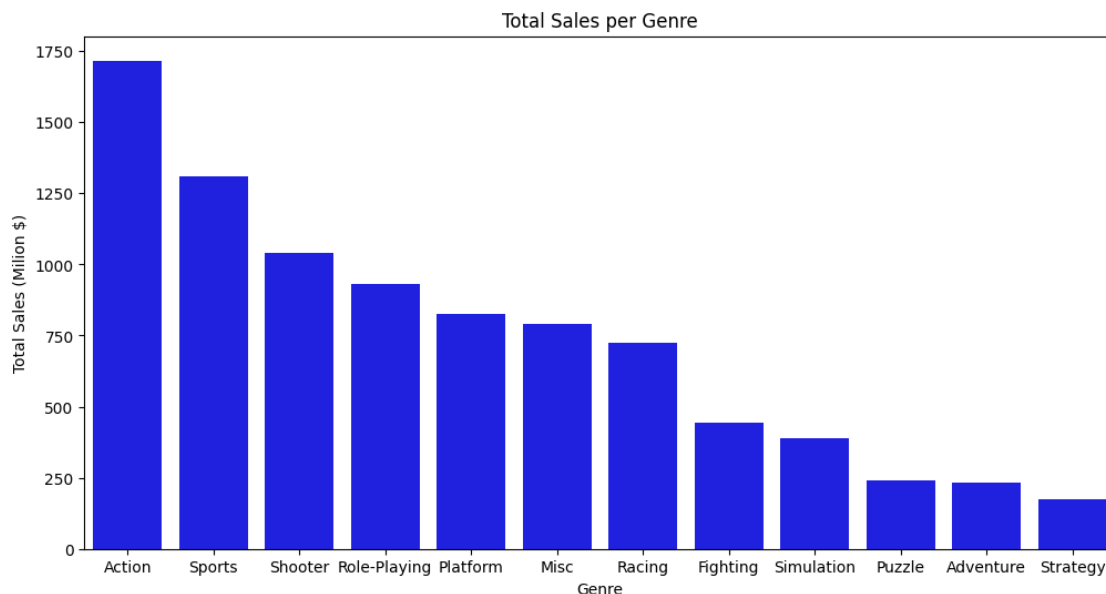
de 50% em relação à segunda plataforma, o X360.

Ao compararmos com o FIFA 15, lançado no ano seguinte, percebemos que a plataforma com maior volume de vendas foi o PS4, o sucessor do PS3, atingindo valores de vendas próximos aos do FIFA 14 no PS3. Isso sugere que jogos que tiveram um bom desempenho comercial em uma plataforma anterior têm grandes chances de repetir o sucesso em sua versão para a nova geração de consoles.

In the example of FIFA 14, which we chose to analyze, we observed that the platform with the highest sales was the PS3, with a total of 6.46 million dollars, representing a difference of more than 50% compared to the second platform, the Xbox 360.

When comparing it to FIFA 15, released the following year, we noticed that the platform with the highest sales volume was the PS4, the successor to the PS3, reaching sales values close to those of FIFA 14 on the PS3. This suggests that games that performed well commercially on a previous platform have a strong chance of repeating their success on the new generation of consoles.

```
[26]: genre_sales = df.groupby('genre')['total_sales'].sum().  
      ↪sort_values(ascending=False)  
genre_counts = df['genre'].value_counts()  
  
plt.figure(figsize=(12,6))  
sns.barplot(x=genre_sales.index, y=genre_sales.values, color = 'blue')  
plt.xlabel('Genre')  
plt.ylabel('Total Sales (Milion $)')  
plt.title('Total Sales per Genre')  
  
plt.show()
```



Os jogos de ação lideram as vendas com um volume significativamente maior que os demais gêneros.

Isso se deve ao seu apelo massivo, com mecânicas envolventes que mantêm o jogador interessado e engajado até a conclusão do jogo.

Em segundo lugar, os jogos de esportes também alcançam um grande número de vendas. Esse gênero se beneficia da popularidade global de esportes como futebol, basquete e outros, atraindo um público fiel. Além disso, a prática de lançar versões atualizadas anualmente, com times e jogadores reais, garante vendas consistentes.

Em terceiro lugar, os jogos de tiros (shooters) conquistam tanto jogadores casuais quanto competitivos, especialmente com a popularidade de modos multiplayer online.

Por outro lado, gêneros como quebra-cabeça (puzzle), aventura e estratégia têm vendas mais baixas. Jogos de quebra-cabeça são mais populares em dispositivos móveis, com vendas menores no mercado tradicional. Jogos de aventura, embora tenham um público fiel, não alcançam o mesmo volume de vendas que gêneros mais massivos. Por fim, os jogos de estratégia atendem a um nicho específico de jogadores, resultando no menor volume de vendas entre os gêneros analisados.

Action games lead in sales with a significantly higher volume than other genres. This is due to their mass appeal, with engaging mechanics that keep players interested and invested until the end of the game.

In second place, sports games also achieve a high number of sales. This genre benefits from the global popularity of sports like soccer, basketball, and others, attracting a loyal audience. Additionally, the practice of releasing updated versions annually, featuring real teams and players, ensures consistent sales.

In third place, shooter games appeal to both casual and competitive players, especially with the popularity of online multiplayer modes.

On the other hand, genres like puzzle, adventure, and strategy have lower sales. Puzzle games are more popular on mobile devices, with fewer sales in the traditional market. Adventure games, while having a dedicated fan base, do not reach the same sales volume as more mainstream genres. Finally, strategy games cater to a specific niche of players, resulting in the lowest sales volume among the analyzed genres.

2.3 User Profile by Region: Platform, Genre, and ESRB Rating Analysis

```
[27]: regions = ['na_sales', 'eu_sales', 'jp_sales', 'other_sales']
top_5_platforms_region = df.groupby('platform')[regions].sum().
    ↪sort_values(by=regions, ascending=False).head(5)
top_5_platforms_region
```

```
[27]:
```

	na_sales	eu_sales	jp_sales	other_sales
platform				
X360	595.74	268.32	12.30	84.88
PS2	572.92	332.63	137.54	190.47
Wii	486.87	258.32	68.28	77.71
PS3	390.13	327.21	79.41	134.59
DS	380.31	188.41	175.02	59.04

Com esses dados, conseguimos perceber que a América do Norte (na_sales) e a Europa (eu_sales)

são, de fato, onde se concentram a maior parte das compras de todas as plataformas, sejam elas Nintendo (Wii, DS), Sony (PS2, PS3) ou Microsoft (X360). O X360 se destaca principalmente na América do Norte, com um total de 595.74 milhões em vendas.

Já no Japão (jp_sales), observamos que plataformas como o DS se destacam com vendas de 175.02 milhões, bem à frente das demais. Isso reflete o fato de que as plataformas portáteis da Nintendo, como o DS, têm uma aceitação muito maior no Japão do que em outros mercados.

Essas análises mostram que os mercados de consoles variam bastante de região para região. Na América do Norte e Europa, os consoles tradicionais de mesa dominam, com destaque para o X360, PS2 e PS3. Já no Japão, a preferência por consoles portáteis, como o DS, é mais forte. Isso revela como cada região tem suas próprias preferências de consumo.

With this data, we can see that North America (na_sales) and Europe (eu_sales) are indeed where the majority of purchases are concentrated across all platforms, whether they are Nintendo (Wii, DS), Sony (PS2, PS3), or Microsoft (X360). The X360 stands out particularly in North America, with total sales of 595.74 million.

In Japan (jp_sales), platforms like the DS stand out with sales of 175.02 million, far ahead of the others. This reflects the fact that Nintendo's portable platforms, such as the DS, have much greater acceptance in Japan compared to other markets.

These analyses show that console markets vary significantly from region to region. In North America and Europe, traditional home consoles dominate, with the X360, PS2, and PS3 leading the way. In Japan, however, the preference for portable consoles, like the DS, is stronger. This highlights how each region has its own unique consumption preferences.

```
[35]: std_platforms_sales = df.groupby('platform')[regions].std().  
      ↪loc[top_5_platforms_region.index]  
std_platforms_sales
```

```
[35]:
```

	na_sales	eu_sales	jp_sales	other_sales
platform				
X360	1.044565	0.480610	0.024821	0.144283
PS2	0.576756	0.360514	0.199480	0.340936
Wii	1.545769	1.088449	0.304530	0.306440
PS3	0.583011	0.593983	0.131381	0.233061
DS	0.537130	0.497225	0.394943	0.128759

Analisando os dados, percebi que o Wii tem o maior desvio padrão nas vendas da América do Norte (1.54) e Europa (1.08). Isso mostra que as vendas do Wii variaram bastante nessas regiões. Alguns jogos venderam muito bem, enquanto outros não foram tão populares.

Já o Xbox 360 tem um desvio padrão alto na América do Norte (1.04), mas bem menor em outras regiões. Isso indica que suas vendas foram bastante instáveis nos EUA e Canadá, mas mais estáveis na Europa e no Japão.

O PS2 tem um desvio padrão mais equilibrado entre as regiões, mas a maior variação aparece no Japão (0.19) e em “Outras Vendas” (0.34). Isso sugere que o desempenho do PS2 nesses mercados foi um pouco mais imprevisível.

Looking at the data, I noticed that the Wii has the highest standard deviation in sales in North

America (1.54) and Europe (1.08). This shows that Wii sales varied a lot in these regions. Some games sold extremely well, while others didn't perform as well.

The Xbox 360 has a high standard deviation in North America (1.04), but much lower in other regions. This indicates that its sales were quite unstable in the US and Canada, but more stable in Europe and Japan.

The PS2 has a more balanced standard deviation across regions, but the highest variation is in Japan (0.19) and "Other Sales" (0.34). This suggests that the PS2's performance in these markets was a bit more unpredictable.

```
[36]: top_5_genre_region = df.groupby('genre')[regions + ['total_sales']].sum().  
      ↪sort_values(by='total_sales', ascending=False).head(5)  
top_5_genre_region
```

```
[36]:
```

	na_sales	eu_sales	jp_sales	other_sales	total_sales
genre					
Action	863.17	510.99	160.14	182.22	1716.52
Sports	671.20	371.33	134.93	132.21	1309.67
Shooter	584.83	314.52	38.68	103.33	1041.36
Role-Playing	330.04	188.24	353.39	59.41	931.08
Platform	444.44	199.39	130.71	51.01	825.55

```
[37]: std_genre_sales = df.groupby('genre')[regions].std().loc[top_5_genre_region.  
      ↪index]  
std_genre_sales
```

```
[37]:
```

	na_sales	eu_sales	jp_sales	other_sales
genre				
Action	0.566716	0.406278	0.165069	0.238090
Sports	1.043081	0.784716	0.225420	0.243224
Shooter	1.209710	0.553901	0.103351	0.200114
Role-Playing	0.678293	0.491155	0.650987	0.125317
Platform	1.504557	0.586754	0.527540	0.165478

Ao analisarmos as vendas por gênero e como elas se comportam em cada região, percebemos padrões interessantes. Tanto na América do Norte (na_sales) quanto na Europa (eu_sales), os jogos de ação dominam o mercado, liderando com uma margem significativa em relação aos outros gêneros.

No entanto, o cenário no Japão (jp_sales) é bastante diferente. Lá, os jogos de Role-Playing (RPG) são os preferidos, com vendas de 353.39 milhões, muito acima de outros gêneros. Esse destaque reflete a forte cultura de RPG no Japão. Além disso, é interessante notar que os jogos de tiro (shooter), que têm um desempenho forte na América do Norte e na Europa, têm vendas muito baixas no Japão, com apenas 38.68 milhões. Isso sugere que o gênero de tiro não ressoa tanto com o público japonês.

Essas análises mostram como as preferências de gênero variam significativamente entre as regiões. Enquanto a América do Norte e a Europa preferem jogos de ação e esportes, o Japão tem uma inclinação clara por RPGs. Essas diferenças culturais e de mercado são essenciais para desenvolvedores e publicadoras ao planejarem lançamentos e estratégias de marketing.

When analyzing sales by genre and how they perform in each region, we notice interesting patterns. Both in North America (na_sales) and Europe (eu_sales), action games dominate the market, leading by a significant margin compared to other genres.

However, the scenario in Japan (jp_sales) is quite different. There, Role-Playing Games (RPGs) are the preferred genre, with sales of 353.39 million, far surpassing other genres. This highlights the strong RPG culture in Japan. Additionally, it is interesting to note that shooter games, which perform well in North America and Europe, have very low sales in Japan, with only 38.68 million. This suggests that the shooter genre does not resonate as much with the Japanese audience.

These analyses show how genre preferences vary significantly between regions. While North America and Europe prefer action and sports games, Japan has a clear inclination toward RPGs. These cultural and market differences are essential for developers and publishers when planning releases and marketing strategies.

```
[29]: top_5_rating_region = df.groupby('rating')[regions + ['total_sales']].sum().
      ↪sort_values(by='total_sales', ascending=False).head(5)
      top_5_rating_region
```

```
[29]:
```

	na_sales	eu_sales	jp_sales	other_sales	total_sales
rating					
E	1274.24	703.87	197.96	231.54	2407.61
T	747.60	421.79	150.70	152.99	1473.08
M	742.89	480.00	63.58	176.06	1462.53
E10+	345.50	183.35	40.20	71.93	640.98
K-A	2.56	0.27	1.46	0.03	4.32

```
[38]: std_rating_sales = df.groupby('rating')[regions].std().loc[top_5_rating_region.
      ↪index]
      std_rating_sales
```

```
[38]:
```

	na_sales	eu_sales	jp_sales	other_sales
rating				
E	1.018775	0.758559	0.305504	0.253824
T	0.456975	0.334333	0.222534	0.119182
M	1.039992	0.653637	0.127521	0.363447
E10+	0.436827	0.269245	0.167071	0.106112
K-A	1.259577	0.075498	0.842931	0.017321

Com os dados fornecidos, podemos perceber que os jogos com a classificação E (Everyone) têm o maior volume de vendas em todas as regiões, indicando que os jogos para todas as idades são extremamente populares, especialmente na América do Norte (na_sales), onde as vendas são muito concentradas.

As classificações T (Teen) e M (Mature) também apresentam grandes volumes de vendas nas regiões da América do Norte e Europa. No entanto, na Europa, os jogos com classificação M superam os T, o que faz sentido, pois a população europeia tende a ser um pouco mais velha que a americana.

Com apenas 4.32 milhões em vendas totais, essa categoria tem uma presença bem pequena e praticamente desapareceu, já que foi substituída pela classificação E (para “Everyone”) em muitas

plataformas.

With the provided data, we can see that games with the E (Everyone) rating have the highest sales volume across all regions, suggesting that games for all ages are extremely popular, especially in North America (na_sales), where the sales are highly concentrated.

The T (Teen) and M (Mature) ratings also show significant sales in North America and Europe. However, in Europe, M rated games surpass T, which makes sense considering the slightly older population in Europe compared to the U.S.

With only 4.32 million in total sales, this category has a very small presence and has essentially disappeared, as it has been replaced by the E (Everyone) rating on many platforms.

2.4 Testes de Hipóteses / Hypothesis Testing

```
[30]: xbox_one_scores = df[df['platform'] == 'XOne']['user_score'].dropna()
pc_scores = df[df['platform'] == 'PC']['user_score'].dropna()

t_stat = st.ttest_ind(xbox_one_scores, pc_scores)
p_value = t_stat.pvalue

alpha = 0.05

print(t_stat)

if p_value < alpha:
    print("Rejeitamos a hipótese nula.")
else:
    print("Não rejeitamos a hipótese nula.")
```

```
Ttest_indResult(statistic=-4.38289526334195, pvalue=1.3037979620149106e-05)
Rejeitamos a hipótese nula.
```

Na hipótese nula (H_0), assumimos que não há diferença significativa entre as médias. Isso significa que as classificações médias dos usuários para as plataformas XOne (XOne) e PC são iguais. Em outras palavras, qualquer diferença observada nas médias é atribuída ao acaso.

Já na hipótese alternativa (H_1), que é a hipótese que queremos testar, sugerimos que há uma diferença significativa entre as médias. Isso implica que as classificações médias dos usuários para XOne e PC são diferentes, indicando que a plataforma pode ter um impacto significativo nas avaliações dos usuários.

Eu escolhi $\alpha = 0.05$, que é um valor comum em testes de hipóteses, indicando que podemos aceitar uma chance de 5% de cometer um erro

In the null hypothesis (H_0), we assume that there is no significant difference between the means. This means that the average user ratings for XOne (XOne) and PC are equal. In other words, any observed difference in the means is attributed to chance.

In the alternative hypothesis (H_1), which is the hypothesis we want to test, we suggest that there is a significant difference between the means. This implies that the average user ratings for XOne and PC are different, indicating that the platform may have a significant impact on user ratings.

I chose $\alpha = 0.05$, which is a common value in hypothesis testing, indicating that we can accept a 5% chance of making a mistake.

```
[31]: action_scores = df[df['genre'] == 'Action']['user_score'].dropna()
sports_scores = df[df['genre'] == 'Sports']['user_score'].dropna()

t_stat = st.ttest_ind(action_scores, sports_scores)
p_value = t_stat.pvalue

alpha = 0.05

print(t_stat)

if p_value < alpha:
    print("Rejeitamos a hipótese nula.")
else:
    print("Não rejeitamos a hipótese nula.")
```

```
Ttest_indResult(statistic=1.8472797735322193, pvalue=0.06480928064420527)
Não rejeitamos a hipótese nula.
```

Na hipótese nula (H_0), partimos do pressuposto de que não existe uma diferença relevante entre as médias. Em outras palavras, as notas médias dadas pelos usuários para os gêneros Action (Ação) e Sports (Esportes) são praticamente iguais.

Por outro lado, na hipótese alternativa (H_1), que é justamente o que queremos investigar, a ideia é que existe, sim, uma diferença significativa entre as médias. Isso quer dizer que as notas médias dos usuários para os gêneros Action e Sports não são as mesmas, sugerindo que o tipo de jogo pode influenciar diretamente as avaliações dos jogadores.

In the null hypothesis (H_0), we start with the assumption that there is no meaningful difference between the means. In other words, the average ratings given by users for the Action and Sports genres are practically the same.

On the other hand, in the alternative hypothesis (H_1), which is what we aim to investigate, the idea is that there is, in fact, a significant difference between the means. This means that the average user ratings for the Action and Sports genres are not the same, suggesting that the type of game may directly influence player ratings.

2.5 Síntese das Conclusões da Análise de Dados / Key Findings from the Data Analysis

Após olhar os dados de vendas de videogames, podemos perceber que algumas plataformas, como PS2, Xbox 360 e PS3, foram as mais populares e dominaram as vendas, principalmente na América do Norte e na Europa. Jogos classificados como “E” (para todas as idades) também se destacaram, o que mostra que os jogos mais acessíveis para todo tipo de público tendem a vender mais.

Também notamos que plataformas mais recentes, como o PS4 e o Xbox One, tiveram um bom desempenho nos primeiros anos após seus lançamentos, com vendas crescendo por até quatro anos. No entanto, após esse período, todas as plataformas começaram a mostrar uma queda nas vendas, o que pode estar relacionado a fatores como uma possível crise econômica ou outros desafios no

mercado. Além disso, os gêneros “Ação” e “Esportes” se destacaram como os mais rentáveis, o que faz sentido, já que esses gêneros são muito populares entre os jogadores.

Além disso, as classificações ESRB têm impacto nas vendas, especialmente na América do Norte, onde os jogos para todas as idades parecem ser os preferidos. No geral, essa análise ajuda a entender melhor como os consumidores se comportam e como o mercado de videogames evolui ao longo do tempo.

After analyzing the video game sales data, we can observe that some platforms, such as the PS2, Xbox 360, and PS3, were the most popular and dominated sales, particularly in North America and Europe. Games rated “E” (for Everyone) also stood out, showing that games accessible to all types of audiences tend to sell more.

We also noticed that more recent platforms, such as the PS4 and Xbox One, performed well in the first few years after their releases, with sales growing for up to four years. However, after this period, all platforms began to show a decline in sales, which could be related to factors such as a possible economic crisis or other market challenges. Additionally, the Action and Sports genres stood out as the most profitable, which makes sense, as these genres are highly popular among players.

Furthermore, ESRB ratings have an impact on sales, especially in North America, where games for all ages seem to be the preferred choice. Overall, this analysis helps us better understand consumer behavior and how the video game market evolves over time.