

UNIVERSIDADE DE SÃO PAULO – USP
ESCOLA DE ENGENHARIA DE SÃO CARLOS – EESC

Trabalho de Estatística Descritiva: Impacto da Análise de Críticos e de Usuários na Venda de Jogos Eletrônicos

Alunos:

Miguel Rodrigues Fonseca - 14682196

Luís Filipe Silva Forti - 14592348

Professor:

Vicente Garibay Cancho

São Carlos - SP
2025

Sumário

1	Introdução	1
2	Metodologia Empregada	2
	2.1 Variáveis do conjunto que serão utilizadas	2
	2.2 Ferramentas	2
	2.3 Coeficientes de correlação	3
	2.4 Gráficos	3
3	Resultados Obtidos	4
	3.1 Médias e coeficientes de correlação gerais	4
	3.2 Região NA	5
	3.3 Região EU	6
	3.4 Região JP	7
	3.5 Restante do mundo	8
	3.6 Análise	9
4	Conclusão	9

1 Introdução

Durante todo o contexto deste trabalho iremos fazer a análise, descrição e tentar resumir da melhor forma possível nosso problema sugerido com um banco de dados presente na plataforma [Kaggle](#), uma das fontes de dados indicada pelo professor.

O [conjunto de dados que iremos utilizar](#) trata a respeito, principalmente, da venda de diversos jogos eletrônicos desde 1980 até o ano de 2016. O conjunto de dados possui também diversas informações qualitativas e quantitativas, das quais algumas serão alvo de análises neste trabalho, como: notas de críticos e usuários, gênero dos jogos (Ação, Aventura, RPG, etc.), empresa que publicou o jogo, empresa que desenvolveu o jogo, plataforma do jogo (PC, Xbox360, Playstation 4, etc.), quantidade de jogadores, e, é claro, as vendas tanto gerais quanto específicas de parte das maiores regiões consumidoras de jogos eletrônicos (Japão, América do Norte e Europa).

Ao nos depararmos com esse tema, que consideramos relevante por se tratar de um mercado diretamente envolvido com o que aprendemos no nosso curso na área de programação, pensamos no seguinte problema que pode ser respondido ao analisar os dados que possuímos:

O quão relevante as análises dos críticos e do público em geral são para as vendas de jogos em cada região do mundo?

Para o qual formulamos a seguinte hipótese:

As avaliações dos críticos tender a ter menor relevância quando comparadas as avaliações dos usuários. Assim, jogos com boas avaliações dos jogadores tendem a vender mais.

Ao respondermos este problema esperamos que seja possível ter um melhor entendimento de como funciona o mercado de jogos eletrônicos e suas relações internas ao identificar se acontece algum impacto de vendas com a opinião geral do público e/ou da crítica especializada em alguma das três maiores regiões do mercado internacional de videogames.

Esse tipo de análise é relevante pois é comum que veículos de mídia recebam antecipadamente jogos para que eles possam ser avaliados antes do seu lançamento oficial. Então, ao realizar nossas análises, poderemos fazer inferências sobre o impacto real desse costume no público consumidor de videogame e se eles são mais influenciados por essa prática ou pelas suas próprias avaliações na hora de comprar um jogo novo.

2 Metodologia Empregada

Nesta seção estão os métodos de análise que foram utilizados para chegar em nossas conclusões, quais dados do conjunto foram utilizados, ferramentas utilizadas e códigos escritos.

2.1 Variáveis do conjunto que serão utilizadas

Ao todo utilizaremos seis informações do conjunto de dados para cada jogo: vendas na América do Norte, vendas na Europa, vendas no Japão, vendas no restante do mundo, nota dos críticos e nota dos usuários. A tabela com todas as informações do conjunto de dados pode ser vista abaixo:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	82.0	73.0	8.3	709.0	Nintendo	E
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	86.0	73.0	8	192.0	Nintendo	E
7	Wii Play	Wii	2006.0	Misc	Nintendo	13.56	9.18	2.93	2.84	28.52	58.0	41.0	6.6	129.0	Nintendo	E
8	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14.44	6.94	4.70	2.24	28.32	87.0	80.0	8.4	594.0	Nintendo	E
11	Mario Kart DS	DS	2005.0	Racing	Nintendo	9.71	7.47	4.13	1.90	23.21	91.0	64.0	8.6	464.0	Nintendo	E
...
16667	E.T. The Extra-Terrestrial	GBA	2001.0	Action	NewKidCo	0.01	0.00	0.00	0.00	0.01	46.0	4.0	2.4	21.0	Fluid Studios	E
16677	Mortal Kombat: Deadly Alliance	GBA	2002.0	Fighting	Midway Games	0.01	0.00	0.00	0.00	0.01	81.0	12.0	8.8	9.0	Criterion Games	M
16696	Metal Gear Solid V: Ground Zeroes	PC	2014.0	Action	Kojima Digital Entertainment	0.00	0.01	0.00	0.00	0.01	80.0	20.0	7.6	412.0	Kojima Productions	M
16700	Breach	PC	2011.0	Shooter	Destiny	0.01	0.00	0.00	0.00	0.01	61.0	12.0	5.8	43.0	Atomic Games	T
16706	STORM: Frontline Nation	PC	2011.0	Strategy	Unknown	0.00	0.01	0.00	0.00	0.01	60.0	12.0	7.2	13.0	SimBin	E10+

Figura 1 – Tabela do CSV mostrando todas as variáveis

Já aqui podemos ver como ficou nosso conjunto após edição para conter apenas as variáveis quantitativas que serão analisadas.

	User_Score	Critic_Score	NA_Sales	EU_Sales	JP_Sales	Other_Sales
0	0.500000	36.000000	0.130000	0.000000	0.000000	0.010000
1	0.500000	39.000000	0.080000	0.020000	0.000000	0.000000
2	0.600000	72.000000	0.090000	0.000000	0.000000	0.010000
3	0.700000	32.000000	0.040000	0.000000	0.000000	0.000000
4	0.900000	28.000000	0.090000	0.000000	0.000000	0.010000
5	1.000000	13.000000	0.020000	0.010000	0.000000	0.010000
6	1.000000	42.000000	0.080000	0.010000	0.000000	0.010000
7	1.200000	35.000000	0.090000	0.060000	0.000000	0.010000
8	1.200000	43.000000	0.070000	0.020000	0.000000	0.010000

Figura 2 – Tabela do CSV mostrando as variáveis utilizadas

2.2 Ferramentas

Para realização do trabalho usamos a linguagem python com as bibliotecas pandas, numpy, matplotlib e scipy. A biblioteca pandas foi necessária para ler e separar os dados relevantes do arquivo CSV, já a matplotlib nos auxiliou na criação dos gráficos e boxplots que serão mostrados posteriormente, por fim a numpy e scipy foram usadas para diversas coisas, dentre elas o cálculo dos coeficientes de correlação de Pearson e de Kendall.

2.3 Coeficientes de correlação

Os coeficiente de Pearson e Kendall foram calculados com as relações entre vendas nas três regiões e no restante do mundo com as notas recebidas (uma vez para a notas de críticos e uma vez para as notas de usuários). Para isso usamos as funções *pearsonr* e *kendalltau* da biblioteca *scipy* após separar as colunas necessárias do CSV.

2.4 Gráficos

Os gráficos de barras e boxplots seguiram uma linha igual a dos coeficientes, nós realizamos as análises com as mesmas variáveis em busca de alguma relação entre ela com auxílio da biblioteca *matplotlib* de uma forma como abaixo:

```
#Plota o gráfico de médias
plt.figure(figsize=(20, 10))
#Valores levemente deslocados em x para visualização no gráfico
plt.bar(x-2, mediaVendasNA, label="América do Norte")
plt.bar(x-1, mediaVendasEU, label="Europa")
plt.bar(x, mediaVendasJP, label="Japão")
plt.bar(x+1, mediaVendasOutros, label="Restante do mundo")
plt.xlabel("Nota dos críticos")
plt.ylabel("Milhões de vendas média")
plt.legend()
plt.show()
```

Figura 3 – Plot das médias

```
#Boxplot vendas NA
fig = plt.figure(figsize = (20, 10))
plt.boxplot(vendasNA, positions = x, widths=np.ones(x.size)*0.3, patch_artist=True,
            boxprops=dict(facecolor="tab:blue", color="black"),
            capprops=dict(color="black"),
            whiskerprops=dict(color="black"),
            flierprops=dict(color="tab:blue", markeredgecolor="tab:blue"),
            medianprops=dict(color="black"),
            )

plt.xlabel("Nota dos usuários")
plt.ylabel("Milhões de vendas")
plt.title("América do Norte")
plt.show()
```

Figura 4 – Plot de um boxplot

Vale ressaltar também que realizamos uma análise com mapa de calor, onde os eixos X e Y são, respectivamente, as notas dadas pelos usuário se críticos e a intensidade é o número médio de vendas:

```
#Heatmap vendas NA
data_pivoted = pontGeral.pivot(index="Critic_Score", columns="User_Score", values="NA_Sales")
ax = sns.heatmap(data_pivoted)
plt.title("Média de vendas América do Norte")
plt.xlabel("Nota dos usuários")
plt.ylabel("Nota dos críticos")
plt.show()
```

Figura 5 – Plot de um mapa de calor

3 Resultados Obtidos

3.1 Médias e coeficientes de correlação gerais

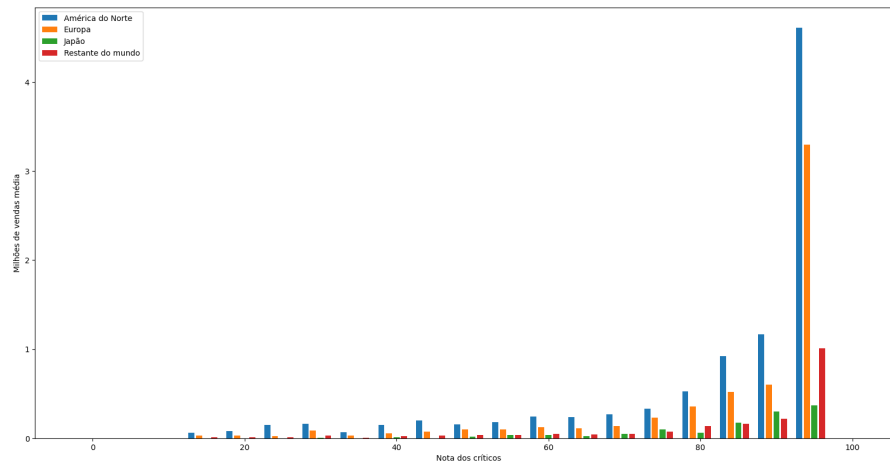


Figura 6 – Média de vendas com relação a nota dos críticos

Região	Coefficiente de Pearson	Valor de p (Pearson)	Coefficiente de Kendall	Valor de p (Kendall)
América do Norte	0.269524	7.33088e-114	0.219327	9.08928e-156
Europa	0.245257	5.28097e-94	0.247378	1.5257e-192
Japão	0.149347	2.52654e-35	0.173338	3.17344e-77
Restante do mundo	0.225858	4.59384e-79	0.226661	6.21837e-154

Figura 7 – Tabela dos coeficientes obtidos para as notas de críticos

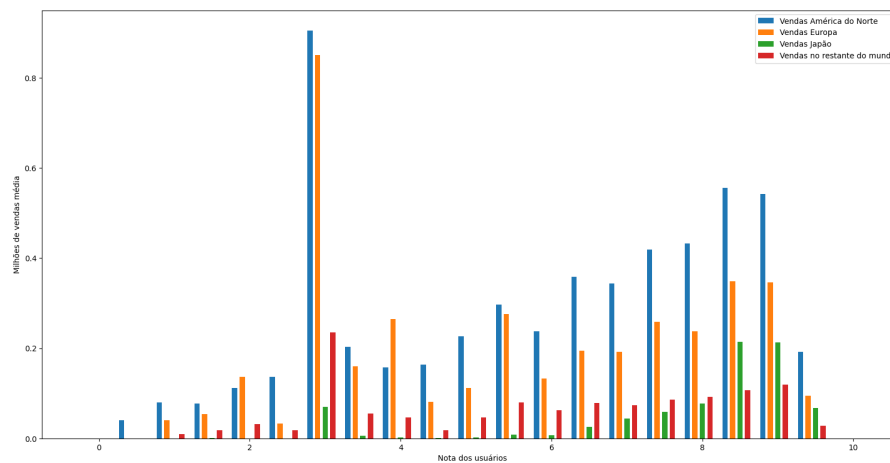


Figura 8 – Média de vendas com relação a nota dos usuários

Região	Coefficiente de Pearson	Valor de p (Pearson)	Coefficiente de Kendall	Valor de p (Kendall)
América do Norte	0.0941167	6.72869e-15	0.087911	1.83295e-26
Europa	0.0597539	7.85615e-07	0.0869413	2.72752e-25
Japão	0.129782	5.14859e-27	0.140418	3.40542e-51
Restante do mundo	0.0582722	1.463e-06	0.052201	1.28183e-09

Figura 9 – Tabela dos coeficientes obtidos para as notas de usuários

3.2 Região NA

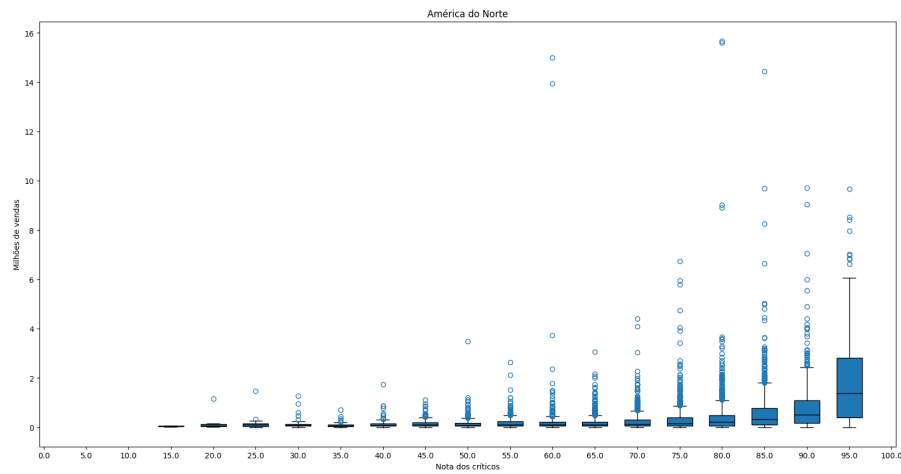


Figura 10 – Boxplot de vendas com relação a nota dos críticos: NA

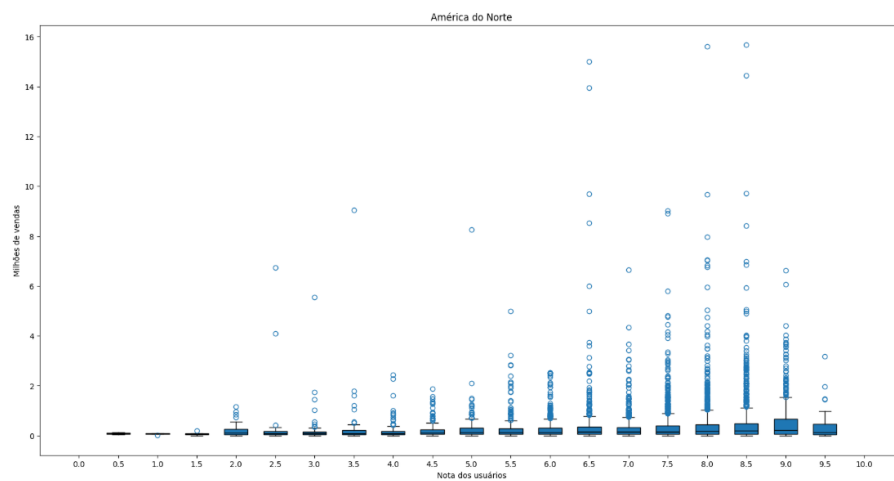


Figura 11 – Boxplot de vendas com relação a nota dos usuários: NA

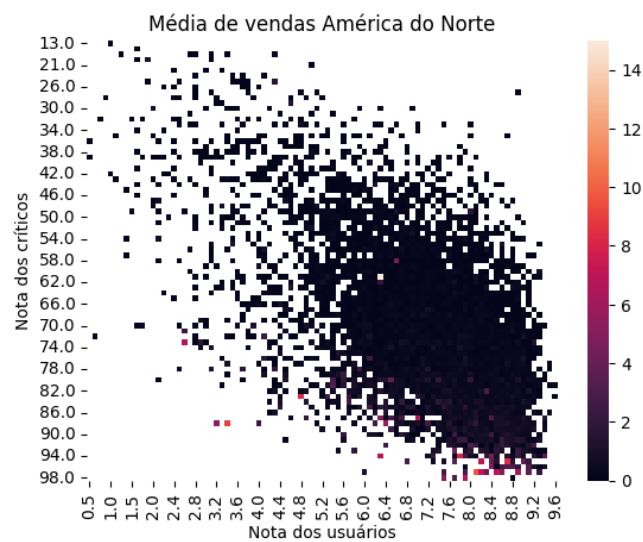


Figura 12 – Mapa de calor NA

3.3 Região EU

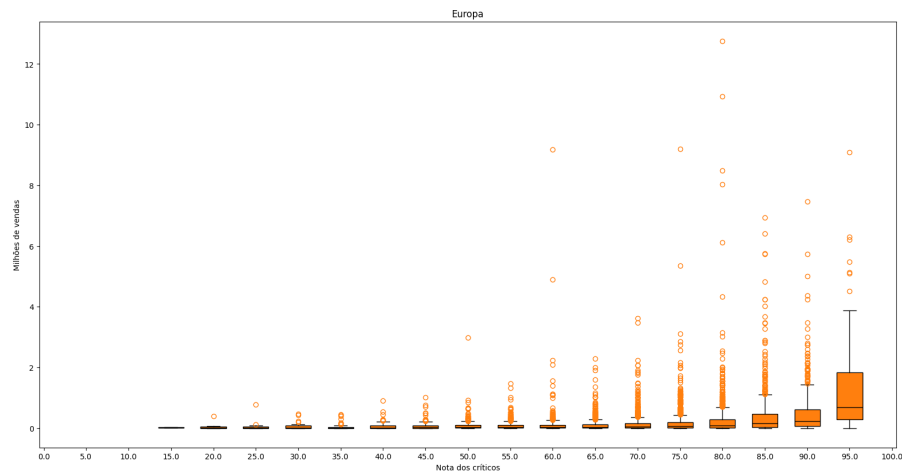


Figura 13 – Boxplot de vendas com relação a nota dos críticos: EU

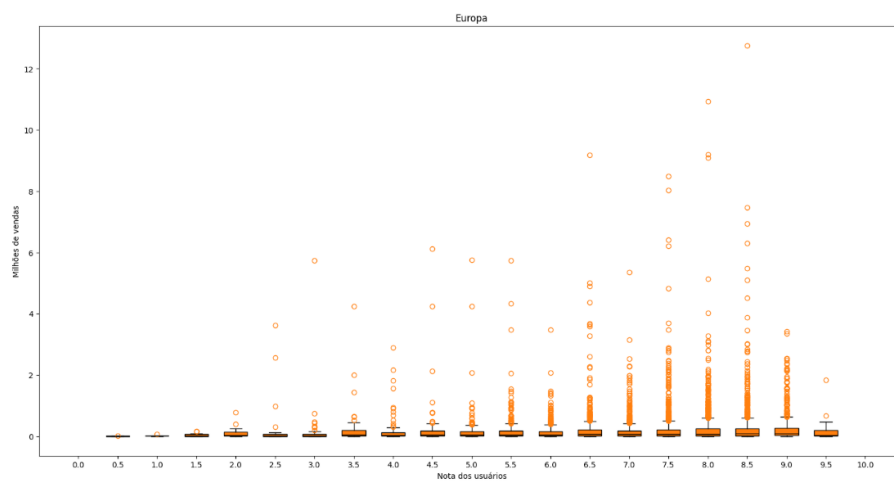


Figura 14 – Boxplot de vendas com relação a nota dos usuários: EU

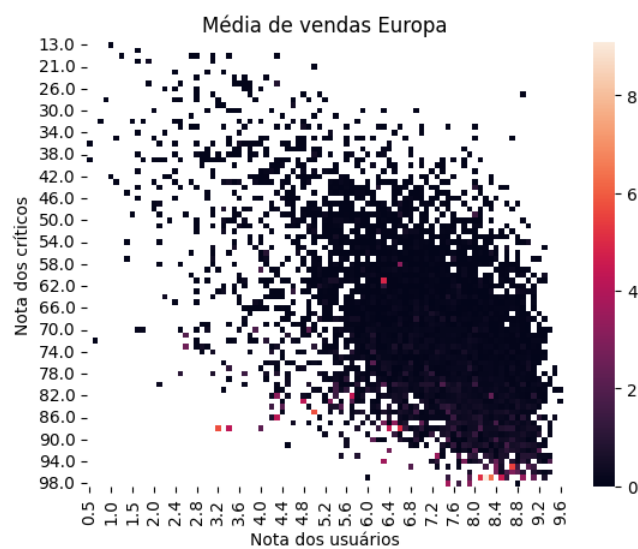


Figura 15 – Mapa de calor EU

3.4 Região JP

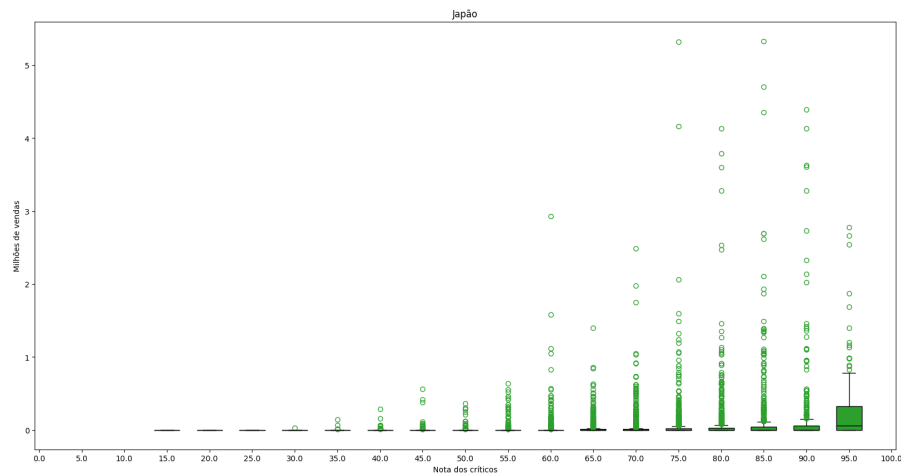


Figura 16 – Boxplot de vendas com relação a nota dos críticos: JP

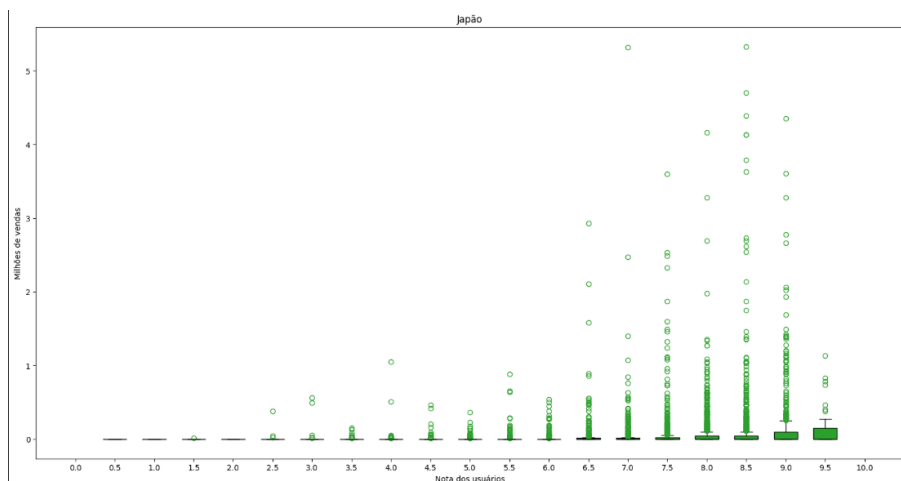


Figura 17 – Boxplot de vendas com relação a nota dos usuários: JP

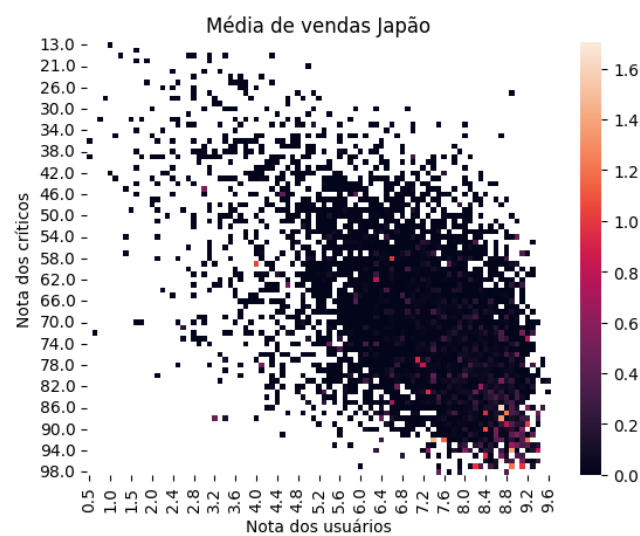


Figura 18 – Mapa de calor JP

3.5 Restante do mundo

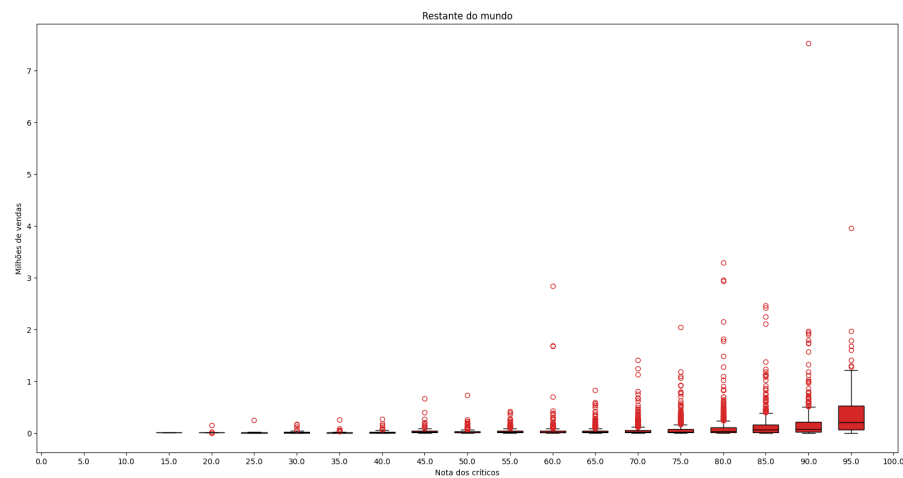


Figura 19 – Boxplot de vendas com relação a nota dos críticos: restante do mundo

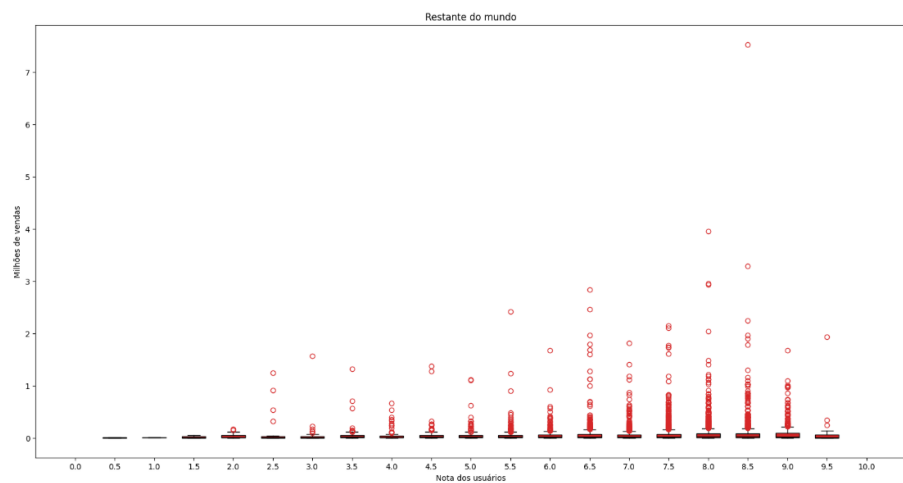


Figura 20 – Boxplot de vendas com relação a nota dos usuários: restante do mundo

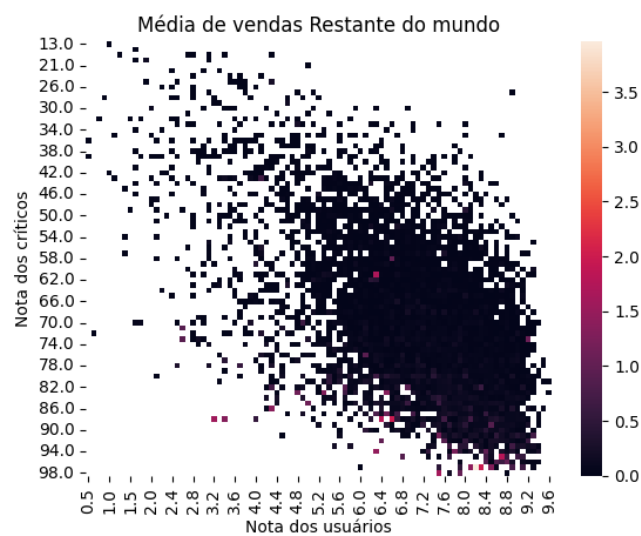


Figura 21 – Mapa de calor para o restante do mundo

3.6 Análise

Pelos resultados obtidos dos coeficientes de Pearson e de Kendall (Imagens 7 e 9), percebem-se algumas características interessantes. Primeiramente é chamativo os baixíssimos valores de p para todas as análises, mostrando que, com certeza, há uma dependência entre o número de vendas de um jogo e suas avaliações.

Segundamente fica visível uma grande disparidade entre os coeficientes da América do Norte, Europa e do restante do mundo nas análises das avaliações dos críticos e dos usuários. Nas análises dos críticos os valores são consideravelmente maiores que nas análises dos usuários, cerca de 3 vezes maiores, no mínimo. Isto prova que a análise dos críticos é muito mais relevante para o número de vendas nessas regiões do que as análises do usuários.

Por fim, o Japão mostrou um comportamento diferente dos demais. Seus coeficientes são aproximadamente os mesmos para ambas as análises, além de serem razoavelmente menores que os dos demais. Isto demonstra que, no Japão, a origem da análise reflete pouco no número de vendas, importando apenas o valor, embora este represente pouca importância pros resultados.

Todas essas análises se apresentaram nos gráficos obtidos, onde os valores obtidos pelas análises dos críticos mostraram uma progressão muito menos caótica do que aquela das análises dos usuários, exceto pelo Japão, que demonstrou uma progressão mais estável em ambos os casos.

Os boxplots, por sua vez, demonstraram que, mesmo que a média siga estas características, a grande maioria dos jogos ainda vendem relativamente pouco, independente da nota. Isto fica perceptível na mediana dos gráficos, que está sempre nos valores mais baixos. Portanto, embora jogos com maiores notas vendam mais, a probabilidade de ter vendas baixas é consideravelmente constante para todas as notas.

Por fim, os gráficos de calor novamente provam a importância das análises dos críticos, pois a concentração de valores altos está toda na região de avaliação alta dos críticos, independente das avaliações dos usuários. Exceto no Japão, que possui valores maiores também nos valores mais medianos das análises dos críticos, confirmando novamente sua excepcionalidade em relação aos demais.

4 Conclusão

A partir das análises realizadas, foi possível observar que existe uma relação entre as avaliações (notas) atribuídas aos jogos eletrônicos e seu desempenho em vendas nas diferentes regiões do mundo.

Em particular, os resultados obtidos mostraram que as notas dos críticos possuem uma correlação mais forte e consistente com as vendas, especialmente nas regiões da América do Norte e Europa. Esse resultado contradiz nossa hipótese inicial do trabalho, que propunha que as análises dos usuários têm maior impacto comercial do que as avaliações dos críticos.

Além disso foi possível perceber como avaliações no geral não tem um impacto tão grande no mercado Japonês, como evidenciado pelos seus coeficientes de Pearson e Kendall. Assim descobrimos uma outra perspectiva que não foi pensada no nosso problema.

Os dois coeficientes se complementam na análise: enquanto Pearson identifica relações lineares, Kendall permite detectar padrões mais sutis de associação. A comparação entre ambos revelou que a influência das notas dos críticos é mais uniforme e previsível, enquanto as notas dos usuários tendem a apresentar maior variabilidade e impacto mais irregular sobre as vendas, especialmente em regiões como o Japão.

Além dos coeficientes, o uso de gráficos de média, boxplots e mapas de calor contribuiu significativamente para a visualização e interpretação dos dados. Isso se deve ao fato de que, após calcular os coeficientes, foi possível observar as relações apontadas pelos cálculos com esse outro ponto de vista gráfico, que também é importante ao expressar resultados estatísticos.

Dessa forma, conclui-se que as críticas especializadas têm papel relevante e mais consistente na decisão de compra dos consumidores, sobretudo nos mercados ocidentais. Isso reforça a importância do marketing pré-lançamento focado em impressões da mídia especializada. Para futuros estudos, seria interessante considerar outras variáveis (como o gênero do jogo ou o estúdio desenvolvedor, que são variáveis qualitativas) para ampliar ainda mais a compreensão sobre os fatores que afetam o sucesso comercial de um jogo.