

Universidade Tecnológica Federal do Paraná - Campus Dois Vizinhos

Especialização em Ciência de Dados

Resultados da Análise Exploratória de Dados

Nome: Diego Monutti de Souza

Dois Vizinhos - PR

2025

Sumário

1. VISÃO GERAL DA ANÁLISE EXPLORATÓRIA	3
1.1. IDENTIFICAÇÃO	3
1.2. PROBLEMA	3
1.3. OBJETIVO	4
2. MÉTODO (EXECUÇÃO DA AED).....	4
2.1. PERGUNTAS DE PESQUISA	4
2.2. BASE DE DADOS	5
2.3. SELEÇÃO, PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS.....	6
3. RESULTADOS	8

1. VISÃO GERAL DA ANÁLISE EXPLORATÓRIA

1.1. IDENTIFICAÇÃO

- **Título:** *Análise exploratória de vendas globais de jogos eletrônicos*
- **Área:** *Indústria de Jogos / Entretenimento*
- **Autor(a):** Diego Monutti de Souza
- **Data:** 14/08/2025

1.2. PROBLEMA

A indústria global de jogos eletrônicos movimenta bilhões de dólares por ano e representa uma das áreas mais lucrativas do entretenimento. Diante de uma enorme diversidade de títulos, plataformas e públicos regionais, compreender os fatores que influenciam o sucesso comercial de um jogo tornou-se um diferencial competitivo essencial para desenvolvedores, editoras e investidores.

Apesar disso, muitas decisões estratégicas ainda são baseadas em intuições, tendências momentâneas ou recortes parciais do mercado. A falta de uma análise sistemática sobre padrões de vendas, preferências regionais e desempenho por gênero ou plataforma limita a capacidade de antecipar o comportamento do consumidor.

Neste cenário, a análise de uma base consolidada com mais de 16 mil jogos, contendo dados de vendas em várias regiões e plataformas, fornece uma excelente oportunidade para compreender os elementos que determinam o desempenho de um jogo. A motivação principal é transformar esses dados em conhecimento acionável, que contribua com decisões mais estratégicas e com o entendimento aprofundado das tendências do mercado gamer.

1.3. OBJETIVO

O objetivo desta análise é investigar padrões de desempenho comercial no mercado global de jogos eletrônicos, com base em dados de vendas segmentados por gênero, plataforma, região, publicadora e ano de lançamento.

Espera-se, com isso, identificar não apenas os produtos mais vendidos, mas também os fatores que impulsionam esse sucesso em diferentes contextos. A análise também permitirá identificar comportamentos regionais distintos, variações ao longo do tempo e perfis típicos de jogos bem-sucedidos.

Entre os desafios esperados estão o tratamento de dados ausentes, a padronização de variáveis categóricas e a visualização de padrões complexos de forma clara e acessível. O resultado deverá oferecer insights relevantes tanto para aplicações acadêmicas quanto para decisões estratégicas na indústria de games.

2. MÉTODO (EXECUÇÃO DA AED)

2.1. PERGUNTAS DE PESQUISA

Q1 – Quais gêneros de jogos apresentam, em média, os maiores volumes de vendas globais? Essa pergunta busca identificar os tipos de jogos com maior apelo comercial, usando medidas de posição como média e mediana, segmentadas por gênero.

Q2 – Existem diferenças relevantes de preferência por gênero entre as regiões (América do Norte, Europa, Japão)? Com essa investigação, pretende-se verificar se há padrões de consumo regionalizados para determinados estilos de jogos, utilizando agrupamentos e proporções.

Q3 – Qual plataforma (console) concentrou mais vendas ao longo dos anos? E como isso evoluiu ao longo do tempo? Essa pergunta visa identificar as plataformas mais lucrativas e sua variação temporal, por meio de agregações e séries históricas.

Q4 – Quais jogos se destacam como outliers em vendas globais? Objetiva encontrar títulos com desempenho muito acima da média, aplicando análise de outliers com boxplots e medidas interquartis.

2.2. BASE DE DADOS

A base de dados utilizada nesta análise foi obtida da plataforma Kaggle, proveniente do site vgchartz.com via web scraping. O arquivo utilizado é “vgsales new.csv”, com 16.598 registros de jogos que venderam mais de 100 mil cópias. A base inclui variáveis como nome do jogo, plataforma, ano de lançamento, gênero, publicadora e dados de vendas por região (América do Norte, Europa, Japão, outras regiões) e no total global.

Principais variáveis utilizadas:

- Name (nome do jogo)
- Platform (plataforma)
- Genre (gênero)
- Publisher (publicadora)
- Year (ano de lançamento)
- NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales (vendas por região e total global)
- Rank (ranking global de vendas)

A base pode ser acessada neste link:

<https://www.kaggle.com/datasets/willianoliveiragibin/video-game-sales-analyze>

2.3. SELEÇÃO, PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS

O pré-processamento consistiu em diversas etapas que asseguraram a qualidade, consistência e adequação dos dados para as análises exploratórias subsequentes:

- **Remoção de valores ausentes:** identificaram-se 271 registros com valores nulos na coluna Year e 58 na coluna Publisher. Como essas informações são essenciais para as análises temporais e categóricas da base, os registros incompletos foram removidos.
- **Conversão de tipo:** a variável Year, originalmente do tipo float64, foi convertida para int64 com o objetivo de representar corretamente o ano de lançamento dos jogos e permitir ordenações cronológicas adequadas.
- **Verificação de duplicatas:** a base foi inspecionada com o método .duplicated() e constatou-se que não havia registros duplicados.
- **Padronização de nomes de publishers:** diferentes variações nos nomes de publicadoras foram padronizadas para evitar distorções nas análises agregadas. Essa unificação considerou fusões, rebrandings e subdivisões internas. Por exemplo, nomes como "989 Sports", "989 Studios", "SCEA" e "Sony Computer Entertainment" foram todos agrupados sob o nome Sony. Já "Square", "Enix" e "SquareSoft" foram consolidados como Square Enix, refletindo a fusão ocorrida em 2003.

A etapa de transformação teve como objetivo tornar os dados mais expressivos, comparáveis e adequados para análises estatísticas e visuais:

- **Criação da variável Decade:** foi criada uma nova variável categórica a partir do campo Year, representando a década de lançamento de cada jogo (ex.: 1990, 2000, 2010). Essa transformação facilitou a análise de tendências históricas.

- **Cálculo de proporções regionais:** foram adicionadas novas variáveis representando a proporção das vendas de cada região em relação ao total global, como %NA_Sales, %EU_Sales, %JP_Sales e %Other_Sales. Por exemplo, se um jogo vendeu 2 milhões na América do Norte e 4 milhões no total, a coluna %NA_Sales assumiu o valor 0,5.
- **Normalização de variáveis numéricas:** foi aplicada a técnica de normalização Min-Max às colunas de vendas regionais e globais para escalá-las entre 0 e 1. Essa padronização é fundamental para evitar o viés causado por diferentes ordens de grandeza nas análises comparativas e foi essencial para as etapas posteriores de clusterização.
- **Agregações por categoria:** os dados foram agrupados por Genre, Platform, Publisher e Decade, permitindo calcular médias e somatórios de vendas, além de contar a frequência de lançamentos. Isso possibilitou comparações entre categorias ao longo do tempo e entre regiões.
- **Clusterização e redução de dimensionalidade:** com o intuito de investigar padrões não supervisionados entre os jogos, foi aplicada a técnica de KMeans com 3 clusters, considerando as colunas normalizadas de vendas por região. Para facilitar a visualização dos agrupamentos, utilizou-se a técnica de redução de dimensionalidade PCA (Análise de Componentes Principais), que projetou os dados em duas dimensões (pca1 e pca2). O gráfico gerado mostrou separações visuais entre diferentes perfis de vendas. Embora a clusterização não estivesse entre as perguntas de pesquisa, ela foi conduzida como análise exploratória complementar, com o objetivo de observar possíveis perfis distintos de vendas no mercado global de jogos.

Essas etapas garantiram que os dados estivessem limpos, padronizados e prontos para a análise exploratória.

O código completo desta análise, incluindo todos os scripts e notebooks utilizados para o pré-processamento, transformação e geração das visualizações, está disponível no repositório GitHub: <https://github.com/DiegoMonutti/analise-vendas-jogos>

3. RESULTADOS

Q1 – Quais gêneros de jogos apresentam, em média, os maiores volumes de vendas globais?

Essa pergunta busca identificar os tipos de jogos com maior apelo comercial, usando medidas de posição como média e mediana, segmentadas por gênero.

Tabela 1 – Estatísticas de vendas globais por gênero (em milhões de cópias)

Gênero	Média (milhões)	Mediana (milhões)	Desvio Padrão	Qtd. Jogos	Participação Total (%)
Platform	0,95	0,28	2,60	875	9,41
Shooter	0,80	0,23	1,83	1262	11,65
Role-Playing	0,63	0,19	1,72	1470	10,48
Racing	0,59	0,19	1,68	1225	8,25
Sports	0,57	0,22	2,10	2304	14,86
Fighting	0,53	0,21	0,96	836	5,04
Action	0,53	0,19	1,17	3251	19,55
Misc	0,47	0,16	1,33	1686	8,96
Simulation	0,46	0,16	1,21	848	4,43
Puzzle	0,42	0,11	1,58	570	2,75
Strategy	0,26	0,09	0,52	670	1,97
Adventure	0,18	0,06	0,51	1274	2,66

Distribuição de Vendas Globais por Gênero

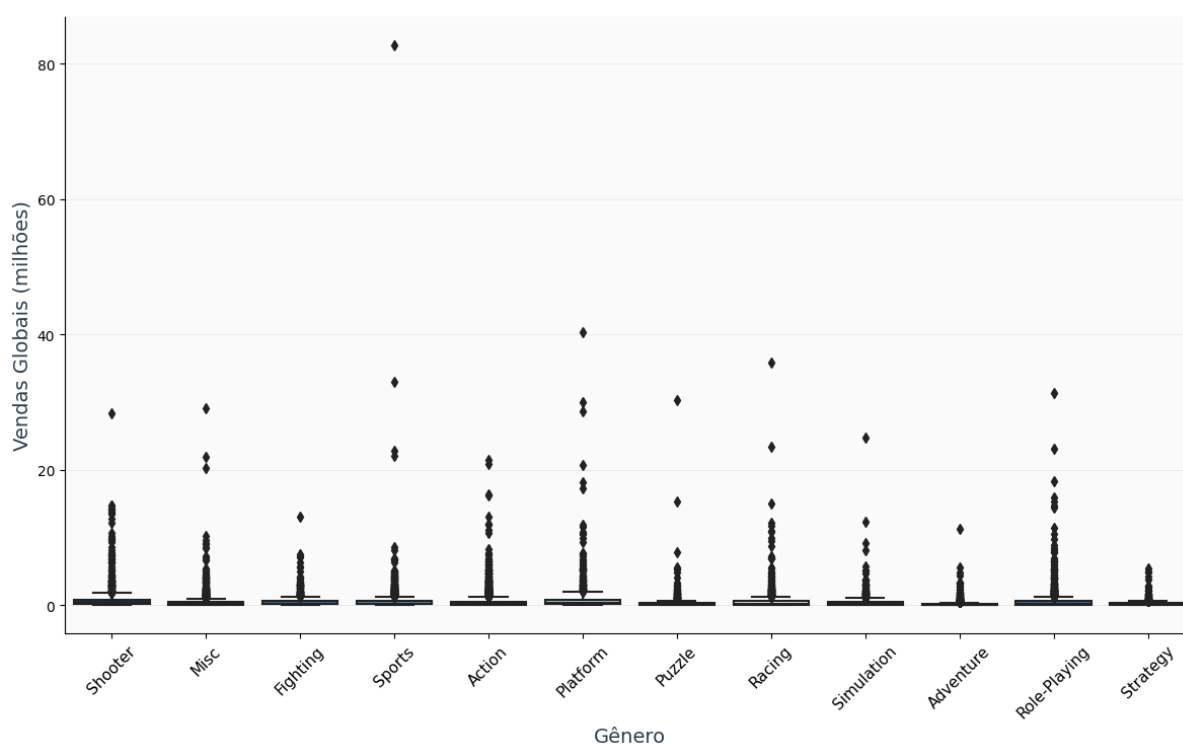


Figura 1- Distribuição de Vendas Globais por Gênero

Média de Vendas Globais por Gênero

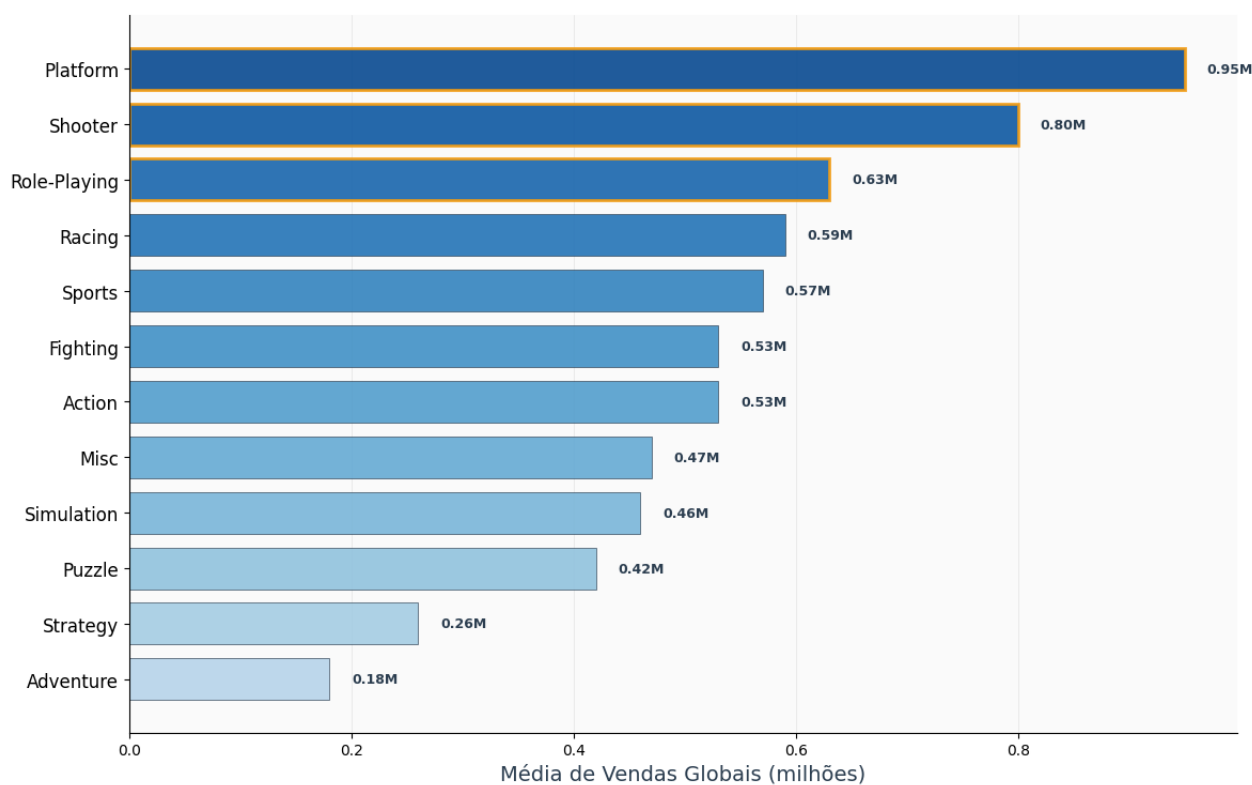


Figura 2 - Média de Vendas Globais por Gênero

Interpretação e Análise dos Resultados:

A análise e visualização dos dados (Figura 2) revela que o gênero **Platform** apresenta o maior volume médio de vendas globais, com **0,95 milhões** de cópias vendidas em média, seguido por **Shooter** (0,80 milhões) e **Role-Playing** (0,63 milhões).

Em contrapartida, os gêneros **Adventure** (0,18 milhões) e **Strategy** (0,26 milhões) apresentaram as menores vendas médias, indicando nichos de mercado mais específicos.

Análise da Variabilidade:

O gênero **Platform** apresenta o maior desvio padrão (2,60), indicando grande variabilidade nas vendas dentro desta categoria. O gênero **Sports** também mostra alta variabilidade (2,10), seguido por **Role-Playing** (1,72) (Figura 1).

Essa alta variabilidade sugere a presença de alguns títulos de performance excepcional que elevam significativamente a média, enquanto muitos jogos do gênero têm vendas mais modestas.

A discrepância entre média e mediana confirma essa interpretação:

- **Platform:** média 0,95 vs mediana 0,28 (diferença de 240%)
- **Shooter:** média 0,80 vs mediana 0,23 (diferença de 248%)
- **Role-Playing:** média 0,63 vs mediana 0,19 (diferença de 232%)

Análise de Mercado:

Participação por volume: O gênero **Action** representa a maior fatia do mercado em quantidade de títulos (19,55%), seguido por **Sports** (14,86%) e **Shooter** (11,65%). Esta distribuição mostra que, embora Action tenha muitos títulos, sua média individual (0,53) é significativamente menor que Platform e Shooter.

Concentração de receita: Apesar de **Platform** representar apenas 9,41% dos títulos, consegue gerar vendas médias 79% superiores ao gênero **Action**, evidenciando maior eficiência comercial por título lançado.

Análise Gráfica:

O boxplot (Figura 1) revela a natureza assimétrica do mercado de jogos, onde a maioria dos títulos concentra-se em vendas baixas (evidenciado pelas caixas comprimidas na base do gráfico), enquanto poucos jogos alcançam vendas excepcionais (outliers visíveis no topo). Esta distribuição justifica perfeitamente a discrepância observada entre médias e medianas, sendo uma característica da indústria de entretenimento digital.

Implicações Estratégicas:

1. **Jogos Platform** oferecem o maior potencial de retorno médio (0,95M), mas com alto risco devido à variabilidade extrema (desvio padrão 2,60).
2. **Shooter e Role-Playing** apresentam equilíbrio interessante entre desempenho comercial e relativa consistência, sendo opções estratégicas mais previsíveis.
3. **Action e Sports** dominam em volume de títulos, mas com retorno médio por título significativamente menor, sugerindo mercados saturados.
4. **Adventure e Strategy** representam nichos especializados com menor potencial comercial médio, mas possivelmente com público fiel.

Conclusão:

A mediana oferece uma visão mais realista do desempenho típico de um jogo dentro de cada gênero, enquanto a média reflete o potencial de títulos excepcionais.

Esta análise sugere que desenvolvedores e editoras devem considerar não apenas o potencial de vendas médio, mas também a variabilidade, competição interna e volume de mercado ao tomar decisões estratégicas sobre desenvolvimento de novos títulos.

A alta variabilidade observada em gêneros como Platform e Sports indica que o sucesso nesses segmentos pode ser mais dependente de fatores como qualidade de

produção, marketing e timing de lançamento do que em gêneros com distribuições mais homogêneas.

Q2 – Existem diferenças relevantes de preferência por gênero entre as regiões (América do Norte, Europa, Japão)?

Com essa investigação, pretende-se verificar se há padrões de consumo regionalizados para determinados estilos de jogos, utilizando agrupamentos e proporções.

Tabela 2 - Proporções Médias de Vendas por Região e Gênero:

<i>Gênero</i>	<i>América do Norte</i>	<i>Europa</i>	<i>Japão</i>	<i>Outras Regiões</i>
<i>Platform</i>	0,6	0,2	0,1	0,1
<i>Puzzle</i>	0,6	0,2	0,2	0,0
<i>Action</i>	0,5	0,3	0,2	0,1
<i>Misc</i>	0,5	0,2	0,3	0,1
<i>Racing</i>	0,5	0,3	0,1	0,1
<i>Shooter</i>	0,5	0,3	0,1	0,1
<i>Simulation</i>	0,5	0,2	0,2	0,1
<i>Sports</i>	0,5	0,2	0,2	0,1
<i>Fighting</i>	0,4	0,2	0,4	0,1
<i>Adventure</i>	0,3	0,2	0,5	0,0
<i>Role-Playing</i>	0,3	0,2	0,5	0,1
<i>Strategy</i>	0,3	0,3	0,3	0,1

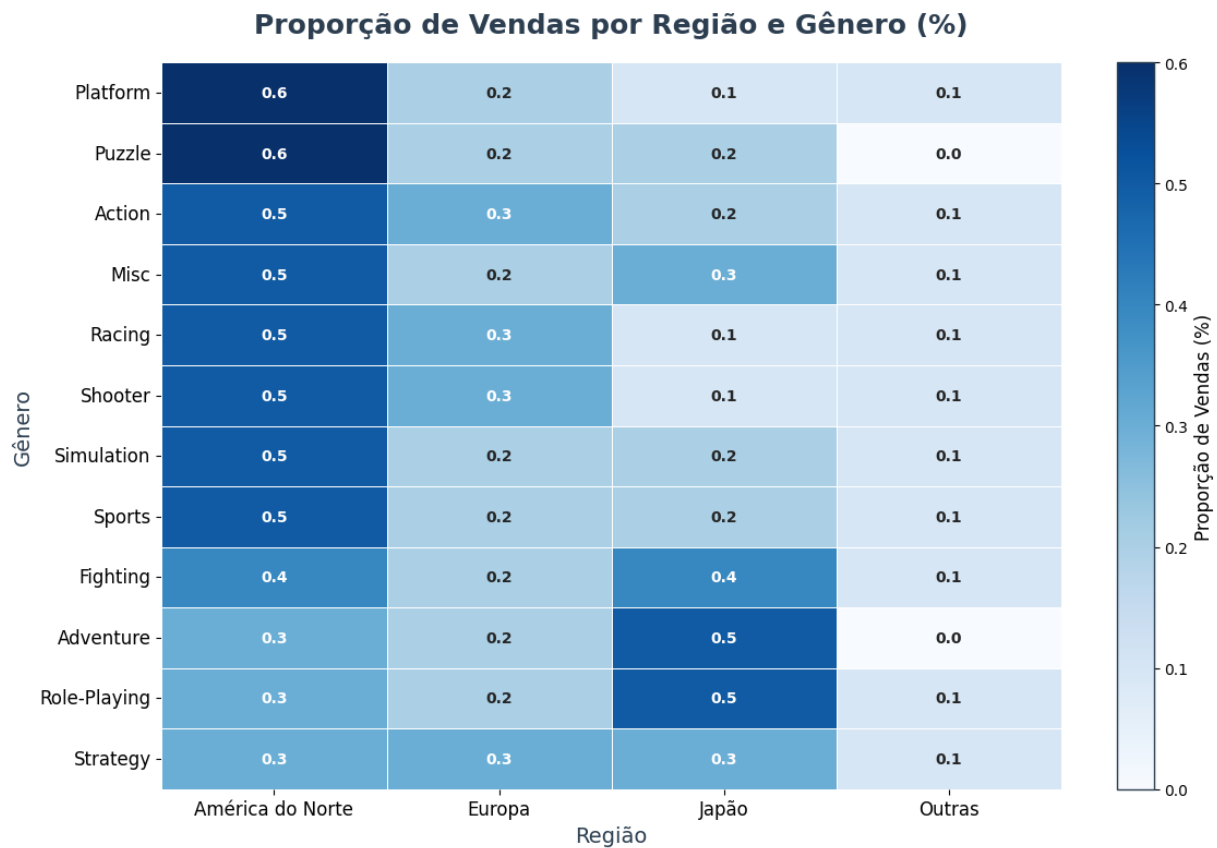


Figura 4 - Proporção de Vendas por Região e Gênero

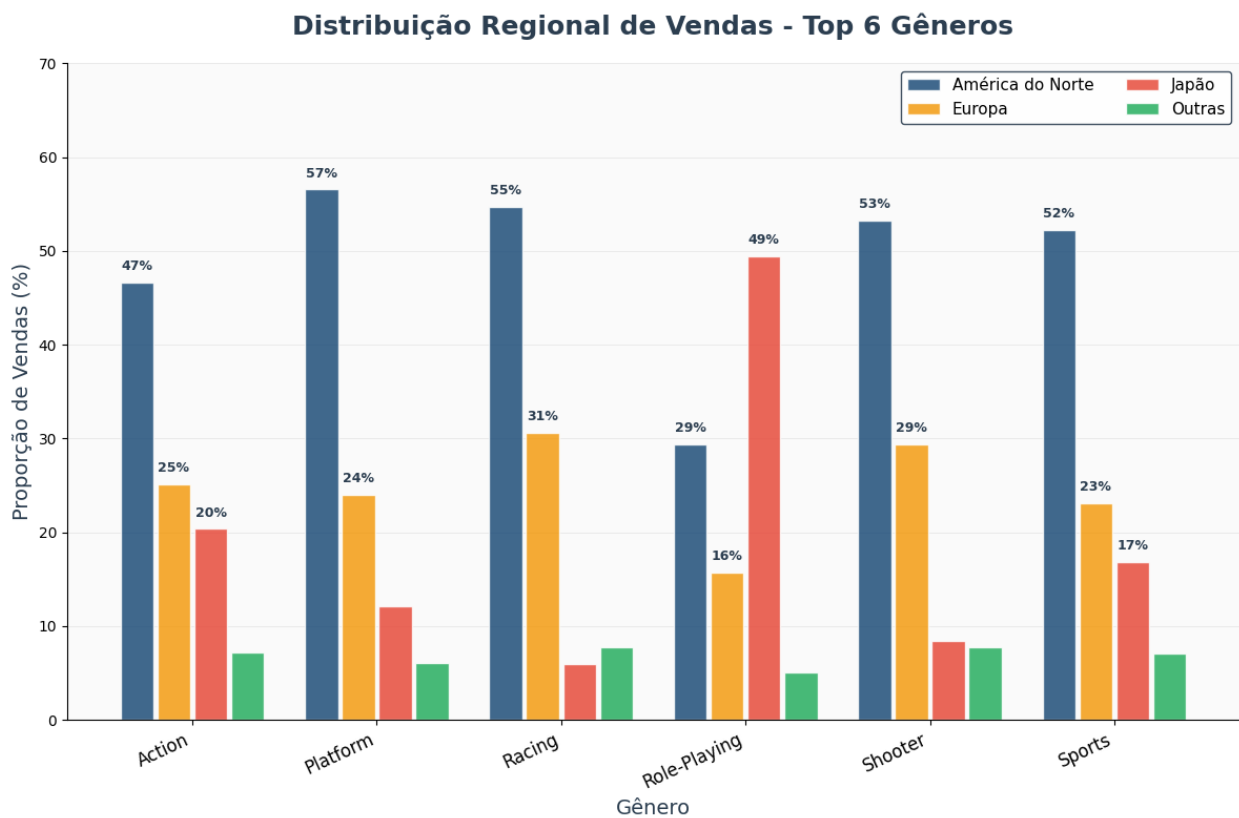


Figura 3 - Distribuição Regional de Vendas

Interpretação e Análise dos Resultados:

A análise revela que a **América do Norte** é o maior mercado consumidor em praticamente todos os gêneros, com destaque absoluto para **Platform (60%)** e **Puzzle (60%)**, superando todas as demais regiões nessas categorias. Esse padrão indica um mercado com forte apelo para experiências acessíveis, rápidas e voltadas ao entretenimento familiar.

O **Japão** apresenta um perfil de consumo diferenciado, concentrando **50% das vendas em Role-Playing** e **50% em Adventure**, gêneros que priorizam narrativa e profundidade. **Fighting** também se destaca no país (40%), refletindo a tradição japonesa de jogos de combate técnico e competitivo.

A **Europa** mantém uma distribuição mais homogênea, com participações consistentes de 20-30% em diversos gêneros como Racing, Action, Shooter e Strategy, sugerindo um mercado diversificado e menos polarizado.

As **Outras Regiões** mostram participação reduzida (0-10%) em praticamente todos os gêneros, o que indica um mercado consumidor ainda em desenvolvimento e potencialmente inexplorado.

Maiores Diferenças Regionais (conforme destacado no gráfico):

- **Platform:** NA (0.6%) vs JP (0.1%) = 0.5pp de diferença
- **Puzzle:** NA (0.6%) vs Other (0.0%) = 0.6pp de diferença
- **Action:** NA (0.5%) vs JP (0.2%) = 0.3pp de diferença
- **Role-Playing:** JP (0.5%) vs Other (0.1%) = 0.4pp de diferença
- **Fighting:** JP (0.4%) vs Other (0.1%) = 0.3pp de diferença

Os gêneros **Platform** (diferença máxima de 50 p.p. entre NA e JP) e **Puzzle** (60 p.p. entre NA e Outras) apresentam as maiores disparidades, evidenciando forte segmentação de público.

Já o gênero **Strategy** apresenta a menor variabilidade (apenas 20 p.p. entre a maior e menor participação), sugerindo apelo mais uniforme globalmente e menor risco geográfico.

Essas variações reforçam a importância de estratégias de marketing segmentadas por região, especialmente para gêneros com forte concentração geográfica.

Análise Gráfica:

O **heatmap (Figura 3)** evidência visualmente as concentrações regionais, com tons mais escuros marcando os gêneros dominantes em cada região (ex.: NA para Platform/Puzzle, JP para RPG/Adventure).

O **gráfico de barras agrupadas (Figura 4)** para os seis gêneros mais populares confirma que não há homogeneização global: cada região apresenta picos distintos, evidenciando preferências culturais específicas.

Padrões Culturais Identificados:

- **América do Norte:** Preferência por jogos de ação direta, puzzle e plataforma, reforçando uma cultura gamer orientada para entretenimento rápido e casual.
- **Japão:** Domínio em RPGs, Adventure e Fighting, refletindo um histórico de valorização de narrativa, progressão e combate técnico competitivo.
- **Europa:** Perfil equilibrado, com presença significativa em múltiplos gêneros, possivelmente devido à diversidade cultural interna.
- **Outras Regiões:** Baixa participação geral, representando potencial de expansão e baixa saturação.

Implicações Estratégicas:

Para desenvolvedores:

1. Jogos Platform e Puzzle devem priorizar localização e campanhas na América do Norte, embora seja necessário considerar alta competição nesse segmento.
2. Títulos RPG e Adventure têm forte apelo no mercado japonês, justificando investimentos focados e localização cultural cuidadosa.
3. A Europa é ideal para portfólios diversificados, absorvendo bem múltiplos gêneros.
4. Jogos híbridos (ex.: Action-RPG) podem ser usados para capturar públicos de mais de uma região simultaneamente.

Para editoras:

1. Estratégias de marketing devem explorar as disparidades regionais identificadas.
2. Strategy apresenta oportunidade para campanhas globais uniformes.
3. Outras Regiões podem ser exploradas como novos mercados com menor concorrência.

Análise da Competitividade

Mercados Saturados:

- América do Norte: alta competição em Platform e Puzzle.
- Japão: concorrência intensa em RPG, Adventure e Fighting.

Oportunidades de Crescimento:

- Outras Regiões: participação de apenas 0-10%, indicando potencial de crescimento.
- Strategy: distribuição equilibrada entre regiões (30% cada), sugerindo menor risco e possibilidade de campanhas globais.

Conclusão:

As diferenças regionais nas preferências de gêneros são significativas e culturalmente enraizadas. América do Norte, Japão e Europa apresentam perfis de consumo distintos, exigindo abordagens personalizadas para maximizar desempenho comercial.

O baixo desempenho em Outras Regiões indica espaço estratégico para crescimento, desde que acompanhado de adaptação cultural e estratégias de localização adequadas.

Além disso, a presença de gêneros com baixa variabilidade, como Strategy, comprova que o mercado global de jogos não é homogêneo, mas oferece nichos de apelo internacional que podem ser explorados de forma uniforme.