

Avaliação Bimestral de Estatística para Ciência de Dados

Aluno: Allan Wandrey Queiroz
Data: 09/10/2025

1. Introdução e objetivo

Este relatório apresenta a análise descritiva e exploratória do conjunto de dados de vendas de videogames (arquivo `vgsales.csv`), com foco em tratamentos estatísticos aplicados, visualizações e interpretações voltadas para a disciplina de Estatística para Ciência de Dados. O objetivo é demonstrar as etapas de pré-processamento, cálculo de medidas descritivas, análise de correlação entre regiões de venda e identificar tendências relevantes.

2. Tratamento de dados e justificativas

Os passos de limpeza realizados foram: - Normalização de colunas de texto (remoção de espaços extras e padronização de `Genre` e `Platform`). - Imputação de valores ausentes: `Publisher` imputado pela moda: **Electronic Arts**. - `Year` imputado pela mediana: **2007**. Conversão de `Year` para inteiro e exportação do dataset limpo (`vgsales_limpo.csv`). Justificativa: A moda foi escolhida para `Publisher` por se tratar de uma variável categórica; a mediana foi escolhida para `Year` por sua robustez contra outliers.

3. Estatísticas descritivas (colunas de vendas)

Métrica	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16598.000	16598.000	16598.000	16598.000	16598.000
mean	0.265	0.147	0.078	0.048	0.537
median	0.080	0.020	0.000	0.010	0.170
std	0.817	0.505	0.309	0.189	1.555
var	0.667	0.255	0.096	0.036	2.418
min	0.000	0.000	0.000	0.000	0.010
max	41.490	29.020	10.220	10.570	82.740
amplitude	41.490	29.020	10.220	10.570	82.730

4. Matriz de correlação entre regiões de venda

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
NA_Sales	1.000	0.768	0.450	0.635	0.941
EU_Sales	0.768	1.000	0.436	0.726	0.903
JP_Sales	0.450	0.436	1.000	0.290	0.612

Other_Sales	0.635	0.726	0.290	1.000	0.748
Global_Sales	0.941	0.903	0.612	0.748	1.000

5. Visualizações

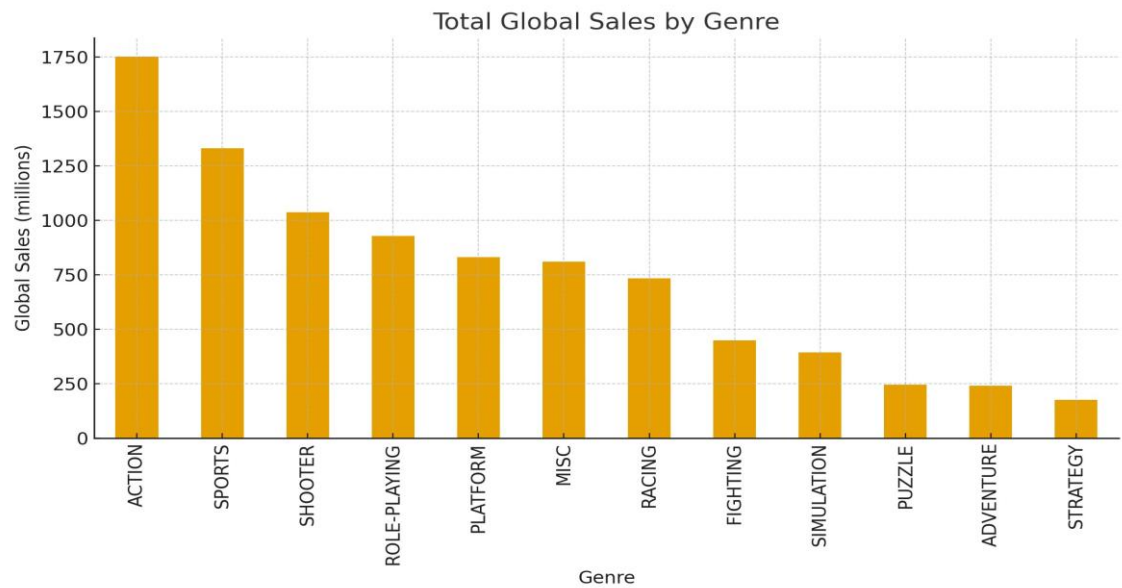


Figura: Vendas globais totais por Gênero

Regional Sales Distribution for Top Genre: ACTION

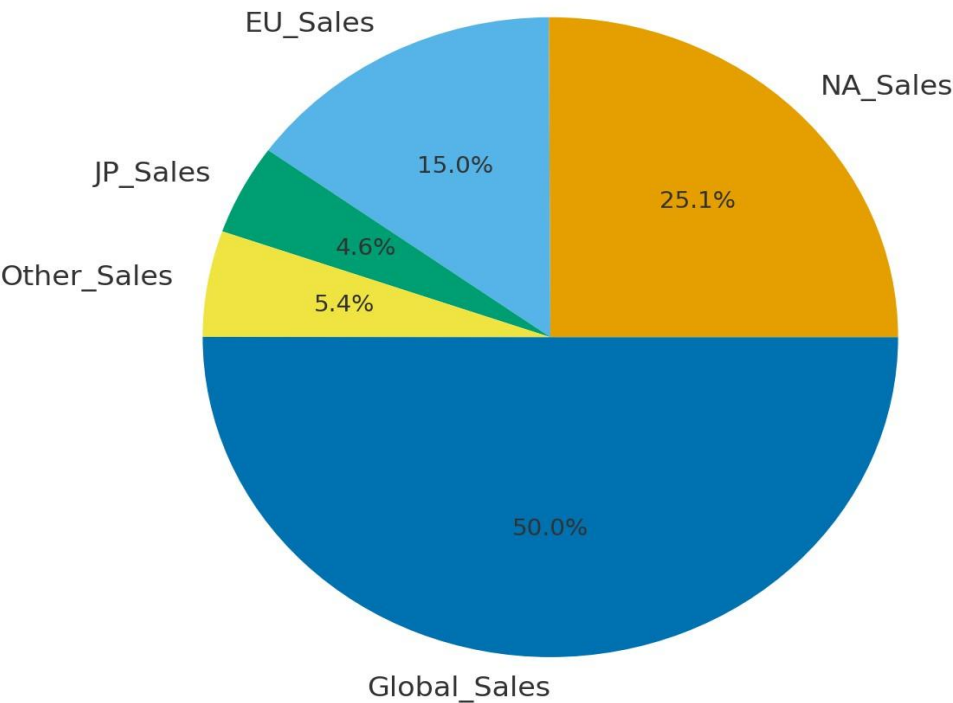


Figura: Distribuição regional de vendas para o gênero principal

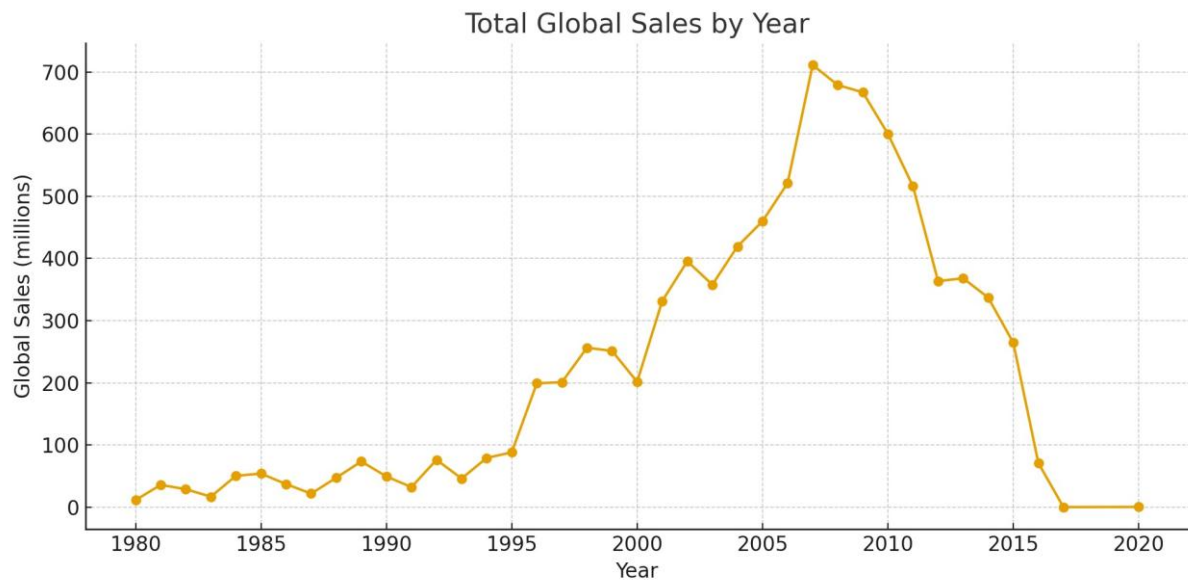


Figura: Evolução das vendas globais por ano

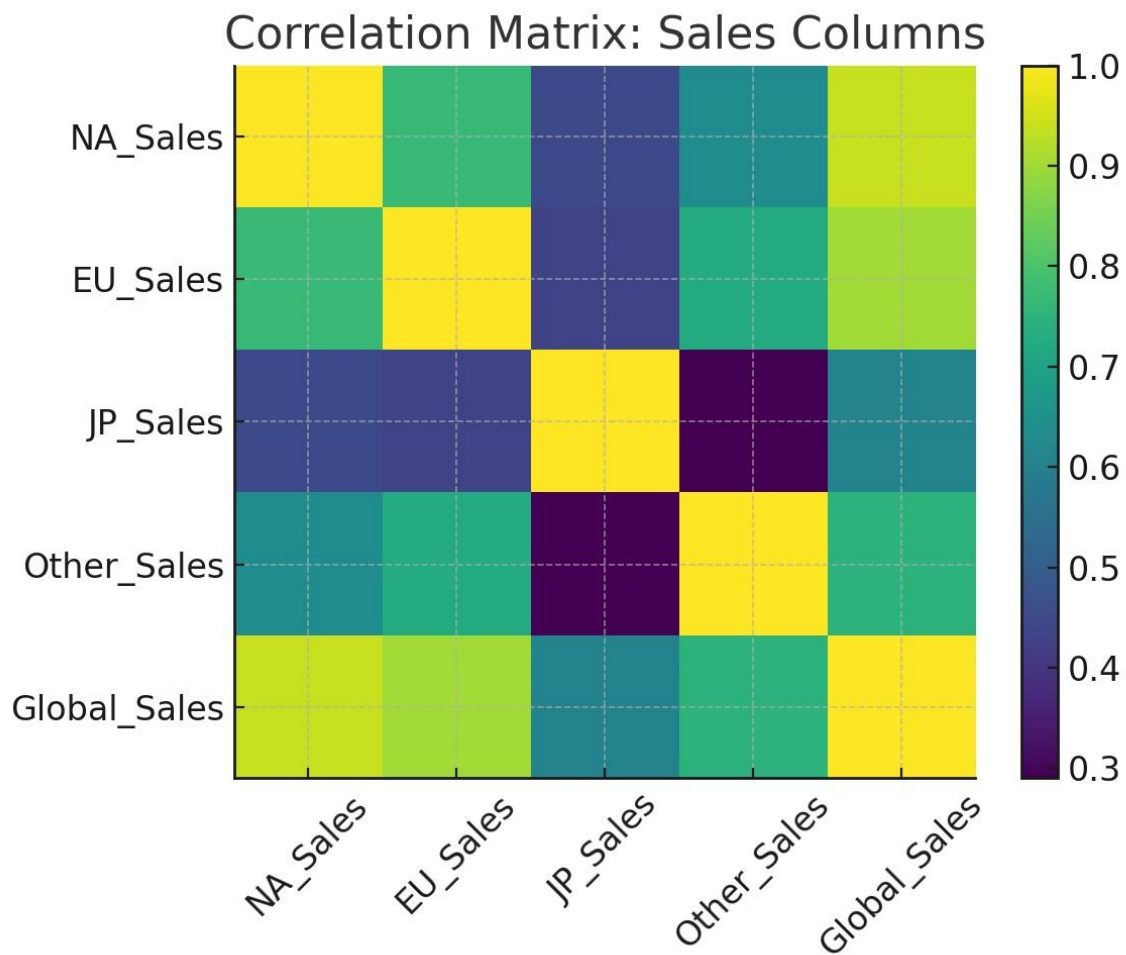


Figura: Heatmap (matriz de correlação entre regiões)

6. Conclusões e observações

- As colunas de vendas apresentam correlações positivas, especialmente entre vendas regionais e Global_Sales, indicando que aumentos regionais tendem a refletir no total global. - A análise por gênero permite identificar os gêneros com maior participação nas vendas globais — recomendamos investigação qualitativa para compreender fatores de sucesso (plataforma, publisher, marketing). - Limitações: o dataset pode conter vieses temporais (lançamentos em anos específicos) e informações ausentes em outras colunas não imputadas automaticamente.