



**TOCANTINS**  
GOVERNO DO ESTADO



Aluno: Iago Batista Antunes Leobas

Disciplina: Mineração de Dados

Data: 15/08/2022

## **Aula 01 - Prática de preparação de base de dados**

Primeiramente, escolha um dataset para o processo de descoberta do conhecimento. Posteriormente, trabalhe com essa base de dados realizando a etapa de transformação dos dados. É necessário que documente todo o processo e o porquê que o dado foi alterado ou retirado do dataset. Bem como, as tipagens do dado.

### **Conjunto Escolhido:**

O conjunto de dados escolhido contém uma lista de jogos com vendas superiores a 100.000 cópias. Foi gerado por uma raspagem de [vgchartz.com](http://vgchartz.com).

Os campos incluem

- Rank - Ranking de vendas gerais
- Name - O nome dos jogos
- Platform - Plataforma do lançamento dos jogos (ou seja, PC, PS4, etc.)
- Year - Ano de lançamento do jogo
- Genre - Gênero do jogo
- Publisher - Editora do jogo
- NA\_Sales - Vendas na América do Norte (em milhões)

- EU\_Sales - Vendas na Europa (em milhões)
- JP\_Sales - Vendas no Japão (em milhões)
- Other\_Sales - Vendas no resto do mundo (em milhões)
- Global\_Sales - Total de vendas mundiais.

O script para extrair os dados está disponível em

<https://github.com/GregorUT/vgchartzScrape> .

## **Hipótese**

A 7ª geração de videogames (PS3, XBOX 360 e Wii) foi bastante marcada pela disputa de vendas entre PS3 vs XBOX 360, minha hipótese surge da premissa que ambas empresas conseguiram suas grandes quantidades de vendas com seus exclusivos de peso, no caso do PS3 com jogos singleplayer voltados a boas histórias e ação, como exemplo The Last of Us e Uncharted, e já o XBOX 360 com jogos de tiro como Halo e Gears of War.

## **Resolução**

Inicialmente foi feita a importação da biblioteca “pandas” para a manipulação dos dados a serem trabalhados.

Após a importação da biblioteca, foi feita a importação da base de dados em formato csv para o meu drive pessoal, e posteriormente, o apontamento de pasta.

Realizado isso, é então chamada uma função para mostrar os 30 primeiros itens da lista.

Posteriormente é feita a impressão de uma matriz com a quantidade de registros, que são 16.598, e a quantidade de colunas, que são 11.

Logo depois é chamada uma função para verificar se há registros nulos na base de dados. Como é apresentado que há, é solicitado então que os registros nulos sejam excluídos.

Após retirar os dados nulos, é imprimido na tela e percebemos que a quantidade de registros já diminuiu.

É então feito um novo agrupamento somente com os dados dos jogos da geração da 7ª geração de videogames, nesse caso o PS3, XBOX 360 e WII, esse grupo é chamado de groupiogo.

Feito esse agrupamento, é então imprimido na tela novamente a quantidade de dados na matriz. É notório que agora a quantidade diminuiu drasticamente, ficando mais interessante de terminar nosso comparativo.

Por fim, é impresso um gráfico de comparação de vendas por gênero de jogos nos videogames da 7ª geração.

Como conclusão, a hipótese se confirma apresentando que o PS3 teve seu maior número de vendas em jogos de ação, e o XBOX 360 com seus jogos de tiro. Vale a pena apontar também que foi obtida a informação que o console da Nintendo, o WII, teve seu sucesso em jogos de esportes, não é atoa que o jogo que encabeça a lista como o mais vendido é o Wii Sports.