

Mineração de Wumpus

JULIANA PINHO MARCHI

177291

LETICIA SOUSA DE OLIVEIRA

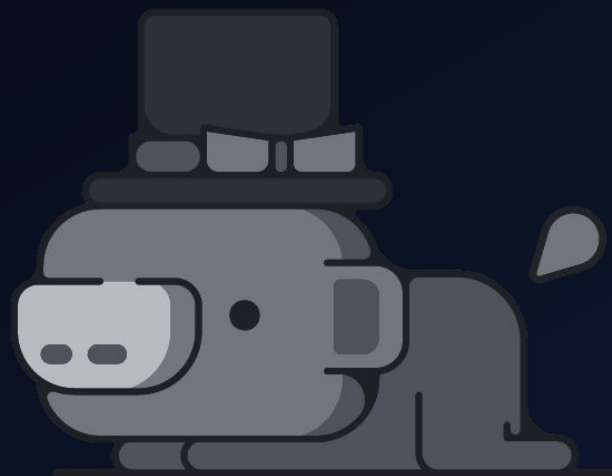
201506

LUCIANO SOUZA GOMES DO NASCIMENTO

202215

THIAGO HENRIQUE VIOTTO

187583



DESCRIÇÃO DO PROBLEMA

- Uma empresa de venda de jogos quer personalizar a página do site de acordo com a localização do usuário, levando em consideração os jogos mais vendidos na presente localidade.



- Neste trabalho, iremos analisar as plataformas, as editoras e os gêneros mais vendidos em cada localização e em escala global em um aspecto temporal.

ATRIBUTOS DA BASE

- A base de dados utilizada no desenvolvimento do trabalho é uma base transacional
 - Platform: plataforma;
 - Year: ano de lançamento;
 - Genre: gênero;
 - Publisher: editora;
 - NA_Sales: vendas na América do Norte;
 - EU_Sales: vendas na Europa;
 - JP_Sales: vendas no Japão;
 - Other_Sales: vendas nas demais localidades;
 - Global_Sales: vendas globais (soma de todas as outras).

Platform,Year,Genre,Publisher,NA_Sales,EU_Sales,JP_Sales,Other_Sales,Global_Sales

PRÉ-PROCESSAMENTO

- A base era ranqueada e possuía mais de 16,500 jogos → Diminuída para os primeiros 4 mil registros.
- *Year* = “N/A” → alguns removidos, alguns imputados em pesquisa manual; Também utilizando nos valores ausentes do *Platform* e *Publisher*.
- Categorização dos atributos numéricos *Sales* e *Year* (na associação, valores numéricos não são indicados) .
- A base de dados original possuía poucos ruídos, à pedido foram integrados 10% de dados incompletos.

CATEGORIAS DE VENDA

(em milhões)	Baixíssimo	Baixo	Médio	Médio-Alto	Alto	Altíssimo
NA_Sales	menor que 0,5	0,5 a 1	1 a 2,5	2,5 a 5	5 a 15	maior que 15
EU_Sales	menor que 0,5	0,5 a 1	1 a 2,5	2,5 a 5	5 a 15	maior que 15
JP_Sales	menor que 0,5	0,5 a 1	1 a 2,5	2,5 a 5	5 a 15	maior que 15
Other_Sales	menor que 0,5	0,5 a 1	1 a 2,5	2,5 a 5	5 a 15	maior que 15
Global_Sale	menor que 0,5	0,5 a 1	1 a 5	5 a 10	10 a 20	maior que 20

CATEGORIAS DE ANO

	Muito Antigo	Antigo	Médio	Novo
Ano	Antes de 2000	2000 a 2007	2007 a 2013	2013 a 2016

TAREFA UTILIZADA: ASSOCIAÇÃO

- A associação possibilitará uma análise profunda sobre cada atributo da base de dados, podendo prever determinadas regras a fim de promover as vendas de jogos.
- O algoritmo utilizado será o APRIORI, visto que ele permite encontrar itens frequentes numa transação e é um dos mais populares e com melhor eficiência.
- As regras de associação podem fornecer indícios de possíveis relações existentes entre os atributos envolvidos nas vendas de jogos. Tais relações podem ajudar a compreender e prever quais as áreas para vendas de determinados jogos, além das análises das mudanças de costumes e gostos em relação a determinados games ao passar dos anos.



EXPERIMENTOS

1. VENDAS E GÊNEROS

Neste experimento foi levado em consideração apenas os gêneros e todas as vendas. Os registros da base foram diminuídos com o intuito de encontrar mais facilmente quais as plataformas mais vendidas em determinados locais.

Experimento 04:

Níveis mínimos de suporte: 0.03%

Níveis mínimos de confiança: 91%

- {EU_Sales = Altissima } -> {Genre = Sports}
Suporte 0.03% Confiança: 100% Lift: 6.07
- {Genre = Action} -> {JP_Sales = Baixissimo}
Suporte:15% Confiança:93% Lift: 1.08
- {Genre = Sports} -> {JP_Sales = Baixissimo}
Suporte 15% Confiança: 91% Lift: 1.06
- {Genre = Shooter} -> {JP_Sales = Baixissimo}
Suporte 9% Confiança: 98% Lift: 1.13

2. VENDAS E EDITORAS

Neste experimento foi levado em consideração apenas as editoras e todas as vendas. Os testes foram feitos separadamente, testando as editoras com as vendas de determinada região, assim sucessivamente. Os registros da base foram diminuídos com o intuito de encontrar mais facilmente quais os gêneros preferidos em determinados anos.

Experimento 06:

Total de registros: 999

Níveis mínimos de suporte: 0,1%

Níveis mínimos de confiança: 10%

- {Global_Sales = Altissima} -> {Publisher = Nintendo}

Suporte: 100% Confiança: 85% Lift: 3.7

- {Other_Sales = Medio alto} -> {Publisher = Nintendo}

Suporte: 50% Confiança: 63% Lift: 2.72

- {NA_Sales = Alta} -> {Publisher=Nintendo}

Suporte: 5% Confiança: 50% Lift: 2.18

- {EU_Sales = Alta} -> {Publisher=Nintendo}

Suporte: 0,8% Confiança: 54% Lift :2.35

- {JP_Sales = Alta} -> {Publisher = Nintendo}

Suporte: 0,8% Confiança: 100% Lift: 4.36



EXPERIMENTOS COM RUÍDOS

3. VENDAS E GÊNEROS

Neste experimento foi levado em consideração apenas os gêneros e todas as vendas.

Experimento 11:

Níveis mínimos de suporte: 0.03%

Níveis mínimos de confiança: 49%

- {Genre = Action} --> {Global_Sales = Medio}
Suporte: 7% Confiança: 50% Lift: 1.07
- {Genre = Action} --> {JP_Sales = Baixissimo}
Suporte: 15% Confiança: 89% Lift: 1.15
- {Genre = Sports} --> {JP_Sales = Baixissimo}
Suporte: 13% Confiança 88% Lift: 1.13

CONCLUSÃO

- Plataforma mais vendida: PS2 -> Xbox 360 -> PS3;
- Gênero: favorito da Europa é o Sports; Shooter, Action e Sports não são populares no Japão;
- PS3 foi preferido por outros locais do mundo;
- Mudança nas preferências das plataformas;
- Gêneros Action, Sports, Role-Playing, Misc e Shooter foram preteridos nos anos médios. Os gêneros Fighting e Simulation foram mais frequentes no período antigo, enquanto que Adventure e Strategy no período muito antigo. Os gêneros Action e Shooter são mais frequentes no período novo.

- A Nintendo predominou como a editora mais vendida em todo o mundo por vários anos, principalmente pelo período antigo. Já nos anos mais recentes a Nintendo perdeu para outras editoras como a Capcom, Electronic Arts e Take-Two Interactive.
- Nos resultados com ruídos, analisamos que alguns experimentos as regras foram muito inconsistentes e ineficientes.