

## APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: VENDAS DE VÍDEO GAMES NO JAPÃO, AMÉRICA DO NORTE, UNIÃO EUROPEIA E OUTROS PAÍSES COM VENDA TOTAL POR GÊNERO

Sabrina Hanich<sup>1</sup>  
João Victor Lunelli<sup>2</sup>  
Ramon Ferreira do Nascimento<sup>3</sup>  
Vinicius Mendes Salum<sup>4</sup>  
Vinicius Azevedo<sup>5</sup>  
Guilherme Arêas Da Silva Perotoni<sup>6</sup>  
Pedro Henrique Malta Côrte<sup>7</sup>  
Gabriel Braun Feistauer<sup>8</sup>  
Gabiella Vitoria De Andrade<sup>9</sup>  
Edinilson da Silva Vida<sup>10</sup>

### RESUMO

Em 8 de maio de 2019 os alunos de todos os semestres de 2019 do curso de Análise e Desenvolvimento de Sistemas participaram do III Aula Aberta. No evento estes foram organizados em equipes com no máximo 10 integrantes, para cada equipe foi entregue um problema de ordem quantitativa. Os professores indicaram um Dataset (conjunto de dados disponibilizados para pesquisa) para cada grupo com dois objetivos a serem alcançados. O primeiro, desenvolver uma rotina automatizada para captura, tratamento e carga dos dados de um site da web um Banco de Dados Não Relacional; o segundo, extrair os dados, armazená-los em uma planilha e realizar um estudo estatístico de correlação para identificar quais variáveis apresentam relação direta com a variável alvo determinada no problema apresentado. Para desenvolver a rotina automatizada utilizamos a linguagem de programação Python, com as bibliotecas Pandas. O problema apresentado para nosso grupo tem origem nos dados extraídos no site <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings> sob o título de *Video Games Sales With Ratings* tendo como alvo a identificação das variáveis que tem correlação entre as vendas de video games no Japão, América do Norte, União Europeia e outros países com a venda total por gênero. Na rotina desenvolvida foi possível identificar o potencial para a automatização dos processos de captura, extração e análise de dados como forma de identificar soluções para problemas que tenham origem dados disponíveis na internet. A partir da implementação desta rotina e do conhecimento das tecnologias aplicadas os integrantes da equipe identificaram diversas outras aplicações. Após execução da rotina que capturou e armazenou os dados foram identificadas e extraídas as variáveis independentes são as vendas no Japão, América do Norte, União Europeia e outros países, do Dataset original e a venda global como dependente. Nos testes realizados as variáveis que apresentaram correlação significativa foram  $A = 0,8767$  e  $B = 0,98788$ .

**Palavras-chave:** Correlação. Vendas. Gênero. Vídeo Games. Python.

<sup>1</sup> Graduanda em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: hanich.sabrina@gmail.com.

<sup>2</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: joao.lunelli8@gmail.com.

<sup>3</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: ramonok@outlook.com.

<sup>4</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: vini-rockmet@hotmail.com

<sup>5</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: viniusazevedooo21@gmail.com

<sup>6</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: gui\_perotoni@hotmail.com

<sup>7</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: phmaltacorte@gmail.com

<sup>8</sup> Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: gabrielfeistauer@gmail.com

<sup>9</sup> Graduanda em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: gdeandrade97.ga@gmail.com

<sup>10</sup> Professor Especialista em Engenharia de Software. Faculdade Cesusc. E-mail: edinilson.vida@faculdadecesusc.edu.br.