

Aplicação de Técnicas de Ciência de Dados

- Engenharia de Software e Ciência da Computação

Stéfani Arnold e Nicolas Hass

Módulo 02/2024

Python3

- Python 3 é o alicerce do projeto, utilizado para implementar e demonstrar as técnicas de ciência de dados.
- Sua versatilidade e uma vasta coleção de bibliotecas permitem abordar todas as etapas da análise, desde a manipulação até a visualização de dados.

Versionamento (Git/GitHub)

- Git e GitHub auxiliam na organização e no versionamento.
- Com ambos, foi possível rastrear mudanças no código e corrigir erros de maneira segura.

Manipulação de Arquivos

- Foram explorados formatos como JSON, conseguidos via pesquisas online - dados da Steam, por ex.
- Sempre respeitando as políticas de uso das fontes.
- Essa abordagem garantiu acesso a dados atualizados e relevantes.

Orientação a Objetos

- Organização e modularidade ao código.
- Classes para representar jogos e suas métricas, encapsulando atributos como título, gênero e tempo médio de jogo.
- Métodos foram implementados para calcular estatísticas ou converter dados entre formatos, promovendo a reutilização e a clareza do código.

Manipulação de banco de dados

- Organizar e recuperar dados de maneira eficiente.
- Dados organizados em json, pré-extraídos de uma base.
- Data scraping no site <https://howlongtobeat.com/>

Análise e visualização de dados

- Análise foi realizada explorando dados coletados sobre a duração de jogos, avaliando correlações com variáveis como gênero, número de vendas e avaliações.
- Ferramentas como Seaborn e Matplotlib foram empregadas para criar gráficos intuitivos, permitindo a visualização de tendências.

Técnicas de Manipulação de dados

- **Inspeção:** Dados foram avaliados para identificar valores ausentes, formatos inconsistentes e possíveis outliers.
- **Limpeza:** Removidos registros duplicados, corrigidos valores inconsistentes e tratados dados ausentes.
- **Transformação:** Dados numéricos foram normalizados, enquanto categorias foram codificadas para facilitar o processamento.
- **Modelagem:** Os dados finais foram organizados em estruturas otimizadas para análise, como graficos, tabelas normalizadas e subconjuntos.

Conclusão

- O presente trabalho buscou cumprir os requisitos da disciplina de Programação para Ciência de Dados.
- Consolidou nosso aprendizado em diversas áreas essenciais.
- Ao aplicar os conteúdos estudados em aula, pretendíamos construir uma análise estruturada que refletisse tanto nosso progresso técnico quanto melhorasse nossa capacidade de resolver problemas do mundo real.
- O projeto proporcionou uma visão e experiência prática das etapas envolvidas em um ciclo completo de ciência de dados: da coleta à apresentação dos resultados.

Obrigado Pela Atenção!

Perguntas?

nicolas.soares@sou.unijui.edu.br

stefani.camargo@sou.unijui.edu.br