



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
ESCOLA DE INFORMÁTICA APLICADA

Proposta do Trabalho de ED2

Luana Chaves
Romulo Pastore
Yuri Campos
André Luiz Pimentel
Pedro Lucas Barros

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2022

1. Tema escolhido e Base de Dados

O tema escolhido para o trabalho foi: “Análise sobre as vendas em sites de e-commerce no Brasil: A região sudeste apresenta um volume maior e mais rentável de vendas em comparação às outras regiões?”.

Em nossas pesquisas, verificamos que o e-commerce brasileiro tem batido recordes de vendas desde que se iniciou a pandemia no Brasil. Para esclarecer, antes disso, muitas pessoas não tinham segurança suficiente para comprar através de lojas virtuais. Com as mudanças drásticas causadas pela quarentena, a demanda em marketplace aumentou consideravelmente, e o sudeste brasileiro teve um aumento de 17,6% no faturamento do e-commerce já no primeiro semestre de 2021 (dados são da Neotrust em parceria com o Comitê de Métricas da Câmara brasileira da Economia Digital). Tais informações acabaram gerando o questionamento presente no tema.

Para verificarmos a nossa hipótese, utilizaremos o conjunto de dados públicos da Olist Store, que foi obtido através do Kaggle e que pode ser visualizada no repositório do grupo no Github (no caminho ‘databases/crude_data’, através dos arquivos .csv originais). Esse conjunto possui informações de 100 mil pedidos entre 2016 e 2018, feitos em diversos marketplaces no Brasil. Seus recursos permitem visualizar diversas informações sobre as vendas, como: Status do pedido, preço, pagamento, desempenho do frete, vendedor, localização do cliente, atributos do produto e a avaliação do cliente.

Os dados comerciais são reais e, portanto, foram anonimizados e as referências às empresas e parceiros no texto da revisão foram substituídos pelos nomes das grandes casas de Game of Thrones.

2. Formação do grafo e Algoritmo

Cada nó do grafo será um objeto que representará um ‘customer’ ou um ‘seller’ (ambos com campos de “id” e “state”), e as ligações entre os nós serão direcionadas e compostas por uma transação. No objeto ‘customer’, será utilizado um atributo “orders”, que será composto de um array de objetos em que cada objeto é uma transação deste cliente.

Cada transação possuirá valores como data da transação, uma referência para um objeto “seller” dela, uma referência para um objeto que representa o produto sendo transacionado e informações sobre um “review” deixado pelo usuário.

3. Ferramentas Utilizadas

A linguagem que será utilizada para programar o algoritmo do trabalho será Python. Diante disso, a biblioteca utilizada para manipulação e análise dos dados será Pandas. Dentre outras ferramentas que serão utilizadas será o Jupyter Notebook e a visualização gráfica será na linguagem C#, e será feita na plataforma da Unity.

4. Divisão do Grupo

Levando em conta que a divisão dos papéis no projeto seriam de 2 pessoas na documentação, 2 pessoas programando o algoritmo e 1 na visualização; a disposição do grupo foi decidida dessa forma:

- Luana Chaves e Rômulo Pastore na documentação do código
- Pedro Lucas Barros e André Luiz Pimentel na programação
- Yuri Campos na visualização