Data Science Bootcamp - WoMakersCode

**Métodos e resultados de previsão de vendas em e-commerce utilizando um conjunto de dados público da Olist Store**

Desafio 4 - Kaggle #1 - Brazilian E-Commerce Public Dataset by Olist

Grupo 1 - Atena:

MURTINHO, A. B.[[1]](#footnote-0); LEITE, M. F.[[2]](#footnote-1); MONACHESI, P. P[[3]](#footnote-2).; PINHEIRO, T[[4]](#footnote-3).

RESUMO: xxxxxxxxxxxxxxxxxxx.

Palavras-chave: *Análise de Dados*; *Tratamento Estatístico*; *Data Science*; *Comércio Eletrônico*.

**Introdução**

O presente trabalho visa documentar a análise de dados realizada por estudantes do Grupo 1, denominado Atena, para a apresentação do *Hack Day* promovido pela *WoMakersCode* como trabalho final decorrente do *bootcamp* (curso intensivo) de *Data Science* (CIência de Dados) ministrado entre 18 de outubro e 17 de dezembro de 2019 em laboratórios da Universidade São Judas Tadeu - campus da Avenida Angélica, próximo à estação Paulista de metrô.

A análise visa o tratamento e a inferência estatística de um *dataset* (conjunto de dados) disponibilizado pela *Olist Store* (uma empresa de tecnologia SaaS - *Software as a Service* - Olist, 2019), e que configura um dos desafios com dados publicados pela plataforma Kaggle (2019).

Esse dataset é composto por um conjunto de oito tabelas com diferentes informações relacionadas a ordens de venda geradas por diferentes comércios eletrônicos (dentre eles as Lojas Americanas, Submarino, entre outras), incluindo data de compra, tipo de produto, prazo de entrega, análise do consumidor, origem das solicitações, dentre outras variáveis.

Dentro disso, questiona-se qual a possível relação entre essas oito tabelas, com o objetivo principal de gerar uma previsão de vendas relacionadas a esse conjunto de dados e, como objetivos secundários, trabalhar com as diversas ferramentas e técnicas que foram expostas durante as aulas do bootcamp.

**Metodologia**

Para dar início à análise, inicialmente os dados foram coletados em formato “csv.” (dados tabulados e divididos por vírgula, conforme disponibilizados pela plataforma Kaggle) e tratados por meio do programa Microsoft Power BI, que conta com uma interface e ferramentas voltadas para *business intelligence* (inteligência analítica - Microsoft, 2019).

Em sequência, os dados foram exportados para um Jupyter Notebook (aplicação que permite criar e compartilhar documentos com linhas executáveis de códigos, visualizações e textos, facilitando a modelagem estatística e a visualização de dados, dentre outras funcionalidades - Project Jupyter, 2019). A linguagem de programação adotada foi Python - versão 3 - e diversas bibliotecas correlacionadas, como Pandas (2019), NumPy (2019) e Seaborn (WASKOM, 2018).

Para a análise estatística, foram usados……..

**Desenvolvimento**

x

**Discussão**

x

**Considerações finais**

x

**Referências**

KAGGLE. ***Brazilian E-Commerce Public Dataset by Olist***. Kaggle, Datasets, 2019. Disponível em: <<https://www.kaggle.com/olistbr/brazilian-ecommerce/#olist_products_dataset.csv>>. Acesso em: 6 dez. 2019.

MCKINNEY, W.***Python for Data Analysis****: Data Wrangling with Pandas, NumPy, and IPython*. California: O'Reilly, 2018.

MICROSOFT. ***Turn data into opportunity.*** Microsoft, Power BI, 2019. Disponível em: <<https://powerbi.microsoft.com/pt-br/>>. Acesso em: 6 dez. 2019.

NUMPY. ***The NumPy community***. NumPy developers, Community, 2019. Disponível em: <<https://numpy.org/community.html>>. Acesso em: 6 dez. 2019.

OLIST. **Dúvidas Frequentes**. Olist, FAQ. Curitiba, 2019. Disponível em: <<https://olist.com/faq/>>. Acesso em: 6 dez. 2019.

PANDAS. ***The pandas project***. Pandas, About, nov. 2019. Disponível em: <<https://pandas.pydata.org/about.html>>. Acesso em: 6 dez. 2019.

PROJECT JUPYTER. ***About us***. Jupyter, About. U.S., 14 nov. 2019. Disponível em: <<https://jupyter.org/about>>. Acesso em: 6 dez. 2019.

WASKOM, M. ***Seaborn:*** *statistical data visualization*. Seaborn, 2012-2018. Disponível em: <<https://seaborn.pydata.org/>>. Acesso em: 6 dez. 2019.

1. Amanda de Britto Murtinho, e-mail: [ab.murtinho@hotmail.com](mailto:ab.murtinho@hotmail.com), LinkedIn: <https://www.linkedin.com/in/amandareznor/>. [↑](#footnote-ref-0)
2. Michele Francisca Leite, e-mail: [michele.fl18@gmail.com](mailto:michele.fl18@gmail.com), LinkedIn: <https://www.linkedin.com/in/michele-leite-91360062/>. [↑](#footnote-ref-1)
3. Paola Picolo Monachesi, e-mail: [paolapicolomonachesi@gmail.com](mailto:paolapicolomonachesi@gmail.com), LinkedIn: <https://www.linkedin.com/in/paola-monachesi/>. [↑](#footnote-ref-2)
4. Taís Pinheiro, e-mail: [tais141@gmail.com](mailto:tais141@gmail.com), LinkedIn: <https://www.linkedin.com/in/taispin/>. [↑](#footnote-ref-3)