

APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: FATORES QUE INFLUENCIAM NO PREÇO DE MOTOCICLETAS USADAS

Gildo Júnior¹
Vinicius Pereira²
Marcos Guerreiro³
Mateus Malicheski⁴
Nicholai Gomes⁵
Ibsem Agrello Dias⁶
Rejane Costa⁷

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 para 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, CSV e Pymongo, para armazenar os dados foi utilizado o Banco de Dados MongoDB. O problema apresentado para nosso grupo tem origem nos dados extraídos no site <https://www.kaggle.com/dan195/ebaymotorcycles> sob o título de *Ebay Motorcycle Prices* tendo como alvo a identificação das variáveis que tem correlação com a preço de motocicletas a venda no *ebay*. O dataset foi obtido do site do ebay O conjunto de dados foi retirado do site do Ebay em 21 de agosto e mostra todas as listagens disponíveis naquele momento. 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 Condição, Modelo, Marca, Milhagem e Números de Revisão do dataset original e a variável Preço como dependente. Nos testes realizados as variáveis que apresentaram correlação significativa foram Milhagem = -0,684 e número de revisão = 0,825.

Palavras-chave: Correlação. Analytics. Preços. Motocicletas. Python.

¹ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: 1gildojunior@gmail.com.

² Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: vini0420@gmail.com.

³ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: marcos.vgm@hotmail.com.

⁴ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: mateusmalicheskidesouza@gmail.com

⁵ Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: nicholais55@yahoo.com.

⁶ Professor Mestre em Administração. Faculdade Cesusc. E-mail: ibsem.dias@gmail.com

⁷ Professora Mestre em Engenharia de Produção. Faculdade Cesusc. E-mail: rejanecestafloripa@gmail.com