

APLICAÇÃO ANALYTICS COM ESTUDO DE CORRELAÇÃO: FATORES QUE INFLUENCIAM O CONSUMO DE CERVEJA

Artur Ferreira Braz¹
Bruno Pessi Derzi²
Clarissa Costa Martins³
Diéssica Gomes de Andrade⁴
Gabriela da Rocha⁵
Jessica Carolina Piovezani⁶
Patrick Santos⁷
Pedro Rodolfo Kalva Junior⁸
Thiago Cabral⁹
Steve Andrade¹⁰
Luiz Claudio Duarte Dalmolin¹¹

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 Selenium, para armazenar os dados foi utilizado o Banco de Dados Pandas. O problema apresentado para nosso grupo tem origem nos dados extraídos no [site https://www.kaggle.com/dongearge/beer-consumption-sao-paulo](https://www.kaggle.com/dongearge/beer-consumption-sao-paulo) sob o título de *Beer Consumption - Sao Paulo* tendo como alvo a variável de consumo de cerveja e quais variáveis a influenciam. Os dados foram coletados em São Paulo - Brasil, em uma área universitária, onde existem algumas festas com grupos de alunos de 18 a 28 anos de idade. O conjunto de dados utilizado para esta atividade possui 7 atributos, sendo um alvo, com período de um ano. 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 Data, Temperatura média, Temperatura mínima, Temperatura máxima, Precipitação e Final de semana do dataset original e a variável Consumo de cerveja como dependente. Nos testes realizados as variáveis que apresentaram correlação significativa foram Temperatura média = 0,574615 e Temperatura máxima = 0,642672.

Palavras-chave: Correlação. Analytics. Consumo de Cerveja. Python. Pandas.

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

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

³Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: clacm1997@gmail.com.

⁴Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: diessicaandrade0@gmail.com

⁵Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: gabrieladarocha570@gmail.com

⁶Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: jessica.piovezani.8@hotmail.com

⁷Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: trickpa2506@gmail.com

⁸Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: Pedro.kalva@gmail.com

⁹Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: thipcabralz@gmail.com.

¹⁰Graduando em Análise e Desenvolvimento de Sistemas. Faculdade Cesusc. E-mail: steve.cri@hotmail.com.

¹¹ Professor Mestre em Engenharia Elétrica. Faculdade Cesusc. E-mail: lcddalmolin@gmail.com.