

Nome completo: Erick de Jesus Lima

Curso: Data Science

Área: TI

Nº do estudo de caso: 85

Lembretes importantes:

- Leia o manual para elaboração de estudo de caso;
- Não é necessário reproduzir o enunciado do estudo de caso;
- Não se preocupe com a ABNT! Seu trabalho pode seguir este template (fonte Arial, tamanho 12 com espaçamento simples);
- O estudo de caso deve ter no mínimo 400 e máximo de 800 palavras contando a partir do título.

A importância da limpeza e do tratamento de dados

No processo de coletas de dados é possível a ocorrência de inconstâncias causadas por motivos variados. Dentre estes, encontram-se o mal levantamento dos dados, o levantamento de parâmetros desnecessários e a omissão de dados sensíveis por parte do dono da informação, tendo em vista a possibilidade de constrangimento gerado com sua exposição, como a orientação sexual ou a religião. Por esse motivo, diversos dados não são preenchidos por alguns dos usuários, resultando na presença de campos em aberto, no seu preenchimento errôneo, dentre outros problemas.

A limpeza e o tratamento de dados são de suma importância para acabar com essas inconstâncias e, para isso, existem várias etapas. Inicialmente é feita a análise de dados que é feita para fim de identificar os erros, o que pode ser feito por meio da utilização de ferramentas de análise qualitativa que analisam os padrões e as restrições para identificar valores inválidos. Após essa etapa, devem ser realizadas a remoção de dados duplicados e de campos vazios, a padronização de datas e horas e, por fim, a correção de digitação e de outros erros.

No caso da ABC Company, os erros capturados na coleta de dados se dão especificamente por tratarem-se de diversas variáveis. Isto ocorreu devido a abrangência de empresas com vários parâmetros que se diferenciam pelo fato de ser um público de vendas diverso e com características diferentes. Com isso diversas empresas divergiram sua visão dos parâmetros obtidos apesar da conformidade existente em seus dados.

Para a resolução do problema específico da ABC Company, o primeiro passo é a utilização da biblioteca Pandas do python, a qual tem o objetivo de realizar a manipulação e a análise dos dados. A instalação e importação dessa biblioteca é feita da seguinte forma:

1. Através do prompt de comando ou terminal, verifique se o pip já está instalado, através do comando ***pip -version***

2. Se o pip estiver instalado, será necessário adicionar as bibliotecas que são indispensáveis para o Pandas. As bibliotecas Numpy que se instala pelo comando: ***pip install numpy*** e a biblioteca matplotlib que é instalada com o comando: ***pip install matplotlib***
3. Por fim será instalado o Pandas, para isso é necessário rodar o comando ***pip install pandas***
4. Feitas todas as etapas anteriores, somente será necessário a importação da biblioteca para o código com o seguinte comando: ***import pandas as pd***
5. Para importar o dataframe, crie uma variável e atribua o seguinte comando ***pd.read_csv("nome_do_arquivo.csv")***

Remoção de dados duplicados com Pandas

Quanto a remoção de dados duplicados, necessita-se realizar um filtro para identificação desses dados, no qual é possível utilizar o método ***duplicated()*** junto com o método ***loc***. Após identificar os dados duplicados, deve-se seguir para a sua exclusão com o método ***drop_duplicates***.

Função Lambda no python

De modo geral, a função lambda serve para criar uma função simples que pode ser utilizada em casos muito específicos. Ela recebe esse nome, pois para declará-la é utilizada a palavra-chave "lambda". A estrutura de uma função lambda é a seguinte: inicia-se a linha com o nome da variável que recebe o atributo lambda junto com os parâmetros da função.

Referências bibliográficas

Apesar das fontes abaixo não terem sido usadas como citações no texto, foram utilizadas no estudo do conteúdo:

https://pandas.pydata.org/docs/user_guide/index.html#user-guide

<https://aws.amazon.com/pt/what-is/data-cleansing/#:~:text=O%20processo%20de%20limpeza%20de,restri%C3%A7%C3%B5es%20para%20identificar%20valores%20inv%C3%A1lidos.>

https://awari.com.br/analise-de-dados-em-pesquisa-qualitativa-descubra-as-estrategias-e-ferramentas-essenciais/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=An%C3%A1lise%20de%20Dados%20Em%20Pesquisa%20Qualitativa:%20Descubra%20as%20Estrat%C3%A9gias%20e%20Ferramentas%20Essenciais#:~:text=1.-,Software%20de%20an%C3%A1lise%20qualitativa%3A,ti%20e%20o%20MAXQDA.

<https://www.fm2s.com.br/blog/cuidados-coleta-de-dados-lado-humano>