

Nome completo: Jose Carlos Merlos Junior

Curso: Ciência de Dados

Área: Tecnologia da Informação

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.

Limpeza e Padronização de Dados

Os esforços para realizar análise de dados para a tomada de decisões estratégicas são um passo fundamental para o sucesso em qualquer segmento. No caso hipotético apresentado no âmbito de vendas, trata-se de apenas um exemplo onde podemos analisar e observar o quão importante é o trabalho para a identificação de tendências e padrões, melhorias de processos, previsões, antecipações e tomadas de decisões.

No caso apresentado, algumas medidas iniciais de tratamento de dados foram efetuadas, restando alguns pontos que foram apresentados para tratamento dos dados.

Para começar, vamos utilizar a biblioteca Pandas para importar dados de um arquivo .CSV e adicioná-lo ao dataframe:

```
!pip install pandas # Instalando a biblioteca do Pandas  
import pandas as pd # Importando as bibliotecas do Pandas  
df = pd.read_csv('arquivo.csv') # importando o dataframe para a variável df  
com as informações do arquivo CSV pela função do pandas read_csv
```

Posteriormente vamos remover os dados duplicados:

```
df = df.drop_duplicates() -- removendo os dados duplicados do meu  
dataframe que foi carregado anteriormente em memória com a função do  
pandas drop_duplicates
```

Poderíamos também utilizar a função lambda para aprimorar o tratamento dos meus dados que carregamos no dataframe (df) a partir de um arquivo csv.

Muito utilizada para criar funções anônimas de forma concisa de forma simples e rápida, como por exemplo podemos aplicar diversas transformações a colunas de um dataframe.

Como exemplo poderíamos realizar o tratamento de valores negativos nos dados da empresa com a função apresentada:

No exemplo abaixo estamos transformando a coluna preco e vendas do nosso dataframe **df** para converter os valores negativos em positivos através da função **abs(x)** de forma a transformar o valor em absoluto, e que aplica a operacaopara cada valor das colunas x (utilizando o método **apply**)

```
df['preco'] = df['preco'].apply(lambda x: abs(x))  
df['vendas'] = df['vendas'].apply(lambda x: abs(x))
```

Como mais um exemplo prático, caso a ABC Company decida atribuir descontos aos seus produtos, poderíamos adicionar uma coluna atribuindo uma porcentagem de desconto aos produtos da ABC Company:

```
df['Desconto'] = df['Preco'].apply(lambda x: x * 0.9) # Nesse exemplo foi utilizado o método apply na coluna Preco do nosso dataframe df. Dessa forma, a função lambda recebe cada valor x da coluna Preco e calcula o valor para atribuir o desconto a cada dado da coluna. No caso atribuímos 10% de desconto.
```

Discussão

Através desse caminho, a equipe da ABC Company continuará tendo êxito na transformação de dados em informações através do uso das técnicas e ferramentas apropriadas para tomadas de decisões.

Com a função lambda não é diferente, trata-se de uma ferramenta poderosa para realizar manipulação de dados como explicado anteriormente.

Conclusão

É nítido o quão fundamental é realizarmos estudos das informações que são geradas pela empresa e, para que tenhamos essas informações confiáveis para estudos e insights, faz-se necessária a aplicação de estudos e técnicas como introduzido no presente trabalho.

Para que a ABC Company continue tendo sucesso na extração de informações e insights valiosos dos dados, é necessário um trabalho contínuo e não único. Propostas de melhoria incluem a realização de automatização dos processos de limpeza e tratamento, além de um sistema de monitoramento da qualidade dos dados para que as análises permaneçam confiáveis.

