

Nome completo: Paulo Emílio Nery Assis Oliveira

Curso: Data Science

Área: Tecnologia

Nº do estudo de caso: 85

Limpeza e Tratamento de Dados com Python

A qualidade dos dados é fundamental para garantir a precisão das análises e decisões empresariais. No contexto da ABC Company, que coleta e analisa dados de vendas de várias lojas em todo o país, a equipe de análise de dados enfrentava desafios com dados incompletos, inconsistentes ou incorretos. Para solucionar esses problemas, eles optaram por realizar a limpeza e o tratamento dos dados utilizando a linguagem de programação Python, com foco na biblioteca pandas.

Instalação e Importação da Biblioteca Pandas e Leitura do DataFrame:

- Para iniciar o processo, é necessário instalar a biblioteca pandas, o comando a ser utilizado é:
``pip install pandas``.
- Em seguida para utilizar a biblioteca pandas, é necessário importa-la, para isso utilizamos o código:
``import pandas as pd``.
- Para importar o dataframe e carregar os dados a partir de um arquivo CSV chamado "dados.csv", o código seria:
``df = pd.read_csv("dados.csv")``.

Remoção de Dados Duplicados:

A equipe esqueceu-se de remover dados duplicados durante o processo de limpeza e tratamento, os quais poderiam comprometer a análise. Para resolver essa questão no Pandas, pode-se utilizar o método “drop_duplicates()” no DataFrame. O código para remover duplicados seria:

```
`df = df.drop_duplicates()`
```

Função Lambda em Python:

A função lambda é uma função anônima em Python, frequentemente utilizada para operações simples em um curto período de tempo. No contexto do caso apresentado, a equipe utilizou a função lambda para corrigir preços negativos e valores de vendas. O código abaixo demonstra como foi aplicada a função lambda para calcular o valor absoluto das colunas 'price' e 'sales':

```
`df['price'] = df['price'].apply(lambda x: abs(x))`  
`df['sales'] = df['sales'].apply(lambda x: abs(x))`
```

Discussão e Solução:

O estudo de caso apresentado destaca a importância da limpeza e tratamento de dados na área de Data Science, especialmente no contexto empresarial. Ao utilizar ferramentas como o Python e bibliotecas como o “pandas”, é possível automatizar e otimizar esse processo, garantindo a qualidade dos dados e, consequentemente, a confiabilidade das análises realizadas.

Para evitar problemas semelhantes no futuro, a equipe da ABC Company deve adotar práticas de limpeza de dados mais abrangentes, incluindo a remoção de duplicatas e a verificação de consistência em todas as etapas do processo. Além disso, é fundamental investir em treinamento e capacitação da equipe para garantir o domínio das ferramentas e técnicas necessárias para lidar eficientemente com os dados.

Conclusão:

A utilização de técnicas de limpeza e tratamento de dados, aliada ao uso adequado de bibliotecas como o Pandas em Python, é essencial para garantir a qualidade e confiabilidade das análises de dados. Isso contribui significativamente para a tomada de decisões estratégicas nas organizações, fornecendo insights precisos e confiáveis para impulsionar o sucesso empresarial.

Este estudo de caso ilustra como a aplicação de conceitos de limpeza e tratamento de dados pode melhorar significativamente a qualidade dos dados e, por consequência, as análises resultantes. Continuar investindo em aprimoramento e atualização das técnicas e ferramentas utilizadas é essencial para se manter competitivo no cenário empresarial em constante evolução.