

# Meu Processo de Criação do Script de Análise de Dados com Manus

**Data:** 04 de Junho de 2025

## 1. Minha Solicitação Inicial

Precisei de um script Python para análise de dados e um modelo básico de Machine Learning. Pedi ao Manus para criar um código modular usando Pandas, NumPy, Matplotlib, Seaborn e Scikit-learn, que fosse fácil de usar em um Jupyter Notebook. O fluxo deveria incluir carregamento de dados (de um `dados_exemplo.csv`), Análise Exploratória (EDA) com gráficos, pré-processamento (valores ausentes, codificação categórica) e um modelo simples de Regressão Linear com avaliação.

## 2. Primeira Versão Recebida

Manus entregou rapidamente um script `analise_ml.py` que continha todas as funções que pedi, bem comentadas e com a estrutura para notebook (`# %%`). Além disso, gerou um `dados_exemplo.csv` para que eu pudesse testar o script imediatamente.

## 3. Pedido de Organização da Estrutura

Percebi que seria melhor organizar o projeto em uma estrutura de pastas mais padrão para ciência de dados. Enviei uma imagem de exemplo da estrutura que eu queria, com pastas como `data/`, `src/`, `notebooks/`, `reports/`, e arquivos como `.gitignore`, `README.md` e `requirements.txt`.

Manus prontamente reorganizou tudo, moveu os arquivos para os locais corretos (`dados_exemplo.csv` para `data/raw/`, `analise_ml.py` para `src/`) e criou os arquivos auxiliares. Recebi um arquivo zip (`analise_de_dados_projeto.zip`) com tudo organizado.

## 4. Problema na Execução do Notebook

Ao tentar rodar o script `analise_ml.py` célula por célula no meu ambiente de notebook (VS Code), notei que nada acontecia quando eu executava as células que chamavam as funções (como a que carregava os dados). Parecia que as funções não estavam sendo definidas ou as chamadas não estavam sendo executadas corretamente nesse modo interativo.

Reportei isso ao Manus.

## 5. Correção para Execução Célula a Célula

Manus identificou que o problema era o bloco `if __name__ == '__main__':` no final do script, que impedia a execução das chamadas de função quando rodado célula a célula. Ele refatorou o script, removendo esse bloco e colocando as chamadas de função em células `# %% [code]` separadas no final do arquivo.

Recebi uma nova versão do projeto zipado (`analise_de_dados_projeto_v2.zip`) com essa correção.

## 6. Dúvida sobre Erros `NameError`

Mesmo com a correção anterior, ao executar as células de chamada, comecei a receber erros como `NameError: name 'carregar_dados' is not defined` e `NameError: name 'df' is not defined`. Enviei uma captura de tela desses erros para o Manus.

## 7. Esclarecimento e Ajuste Final

Manus explicou que esses erros `NameError` acontecem em notebooks se eu não executar as células que definem as funções (as primeiras células do script) antes de tentar usar essas funções ou as variáveis criadas por elas (as últimas células). Ele também adicionou um comentário bem claro no script, logo antes das células de execução, me lembrando de rodar todas as definições primeiro.

## 8. Resultado Final

Agora tenho um projeto de análise de dados completo e bem organizado:

- Script Python ( `src/analise_ml.py` ) modular e pronto para ser executado célula a célula no notebook.
- Dataset de exemplo ( `data/raw/dados_exemplo.csv` ).
- Estrutura de pastas padrão.
- Arquivos auxiliares ( `.gitignore` , `README.md` , `requirements.txt` ).
- Instruções claras no script sobre a ordem de execução das células para evitar erros.

O processo foi iterativo e o Manus foi ajustando o trabalho com base nas minhas necessidades e nos problemas que encontrei.