

[Return to "Data Science Foundations II" in the classroom](#)

Analyzing the NYC Subway Dataset

REVISÃO

REVISÃO DE CÓDIGO

HISTORY

Meets Specifications

Parabéns, você respondeu todas as perguntas corretamente e foi aprovado nesse projeto!

Esse é um projeto com implementações bastante desafiadoras e você se saiu muito bem. 🙌

Continue o bom trabalho e até o próximo projeto. Bons estudos! 😊

Coleta de dados

Quatro arquivos foram coletados e salvos utilizando as bibliotecas do Python responsáveis por coleta de dados na Internet.

Os quatro arquivos se tornaram apenas um, preservando todas as colunas do arquivo de origem.

Os dados foram carregados em um Panda Dataframe. Apenas as linhas onde DESCn possui o valor regular foram selecionadas. As colunas Existsn e Entriesn foram criadas de forma correta. A coluna de hora foi formatada como número inteiro.

Análise dos dados

Foi utilizado Python para calcular a quantidade de dias chuvosos? O valor foi calculado de forma correta?

Foi utilizado Python para calcular a temperatura máxima para dias nublados? O valor foi calculado de forma correta?

Os códigos que exibem as respostas das médias devem exibir as respostas corretas.

Foi utilizado Python para plotar os histogramas. Os histogramas foram plotados de forma a facilitar a identificação da distribuição dos dados. A identificação da distribuição dos dados foi feita de forma correta.

Excelente! Podemos perceber nos histogramas que as duas distribuições são semelhantes e do tipo assimétricas positivas, alterando apenas a quantidade de registros em cada dataset.

MapReduce

A saída do redutor deve estar de acordo com o solicitado no arquivo.

A saída do mapeador deve estar de acordo com o solicitado no arquivo.

O código do redutor e mapeador devem rodar sem erros.

 [BAIXAR PROJETO](#)

RETORNAR