

Realizar análise do arquivo results.csv

O dataset é um conjunto de dados que inclui resultados de jogos internacionais de futebol.

Os jogos são estritamente internacionais masculinos e os dados não incluem Jogos

Olímpicos ou jogos em que pelo menos uma das equipes foi a seleção B do país, sub-23 ou uma seleção da liga.

O dataset foi retirado do Kaggle.

Metadados:

date - data da partida

home_team - time mandante

away_team - time visitante

home_score - placar do time mandante

away_score - placar do time visitante

tournament - torneio

city - cidade onde a partida foi realizada

country - país onde a partida foi realizada

neutral - desconsiderar

Importar o arquivo no Spark e responder as perguntas abaixo utilizando python ou SQL.

Para cada questão, informar a resposta e o código utilizado para chegar ao resultado.

- 1 - Quantos registros existem na base?
- 2 - Quantas equipes únicas mandantes existem na base?
- 3 - Quantas vezes as equipes mandantes saíram vitoriosas?
- 4 - Quantas vezes as equipes visitantes saíram vitoriosas?
- 5 - Quantas partidas resultaram em empate?
- 6 - Quantas partidas foram realizadas em cada país?
- 7 - Qual país teve mais partidas?
- 8 - Qual a partida com maior número de gols?
- 9 - Qual a maior goleada?
- 10 - Quantos jogos ocorreram no Brasil?