

Realizar análise do arquivo result.csv

O dataset results.csv é um conjunto de dados que inclui resultados de jogos internacionais de futebol, desde o primeiro jogo oficial em 1972 até 2019. Os jogos variam de Copa do Mundo FIFA à FIFA Wild Cup e amistosos. 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.

A atividade por ser realizada instalando o Spark na máquina local, utilizando docker ou o ambiente Databricks Community.

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?