LISTA 2 – Pré-processamento: coletar e analisar dados

1. Disponibilizei alguns arquivos com informações dos jogos do brasileirão, um histórico de cinco anos de jogos. Os arquivos possuem um formato semelhante ao que encontramos em bancos de dados relacionais.
   Abaixo a estrutura de cada um dos arquivos, os campos em itálico representam as chaves que ligam os dados entre os arquivos:
   a. Arquivos **jogos_<ano>**:
      - *id_arena*
      - *id_clube1*
      - *id_clube2*
      - *id_clube_vencedor*
      - *id_estado_vencedor*
      - *id_estado_clube1*
      - *id_estado_clube2*
      - data
      - dia
      - horario
      - rodada
      - qtd_gols_clube1
      - qtd_gols_cube2
   b. Arquivo **estados.csv**:
      - *id_estado*
      - sigla_estado
      - desc_estado
   c. Arquivo **arenas.csv**:
      - *id_arena*
      - desc_arena
   d. Arquivo **clubes.csv**:
      - *id_clube*
      - desc_clube

2. Crie um notebook novo e importe os arquivos, transformando-os em DataFrames. Mantenha os campos chave como índices.

3. O histórico de jogos está dividido por ano de campeonato. Concatene os arquivos mantendo ao final um único DataFrame com o histórico de jogos com todos os anos.

4. Para trabalharmos é importante "desnormalizarmos" os dados, ou seja, incluir todas as informações (jogos, clubes, estados e arenas) em um único DataFrame. Faça essa desnomalização, mantendo ao final apenas um DataFrame com todos os dados.
   Dica: no caso dos clubes e estados, temos mais de uma chave nos jogos, portanto a função *join()* precisará de parâmetros auxiliares para incluir os sufixos nos campos (*lsuffix* e *rsuffix*). Se os nomes dos campos não ficarem da maneira que você quer, renomeie depois com a função *rename()*.

5. Ao chegar nesse ponto já temos o arquivo unificado e pronto para trabalharmos. Agora vamos explorar um pouco os dados. No mesmo notebook, busque as respostas para as perguntas abaixo:
    a. Qual o clube que mais marcou gols em todas as temporadas?
    b. Qual o clube que mais marcou gols na temporada de 2015?
    c. Qual o jogo onde houveram mais gols, em todas as temporadas?
    d. Qual foi a maior goleada de todas as temporadas? Maior goleada nesse contexto é igual a maior diferença de gols entre o que um time marcou em relação ao outro.
    e. Qual foi o estado mais vitorioso (quantidade de partidas) em 2016?