Projeto: Análise de Dados de Desempenho de Futebol com soccerdata e Power BI

Objetivo geral

Coletar, organizar e analisar dados completos de desempenho de times e jogadores das principais ligas de futebol (principalmente europeias) dos últimos anos, utilizando a biblioteca soccerdata. Os dados serão preparados para análises e visualizações no Power BI, incluindo dados históricos, estatísticas por partidas e temporadas, para dar suporte a análises avançadas.

Sprints

Sprint 1: Preparação do ambiente e levantamento dos dados disponíveis

Objetivo:

* Preparar o ambiente Python com a biblioteca soccerdata atualizada.
* Entender quais métodos e tipos de dados estão disponíveis na versão atual da biblioteca.
* Testar coleta de dados básicos de times e jogadores para temporadas recentes.

Atividades:

* Instalar e atualizar soccerdata corretamente (sem erros de git ou versão).
* Criar script para listar métodos públicos da classe FBref.
* Criar script para testar quais métodos retornam dados (DataFrames) para as ligas e temporadas desejadas.
* Mapear as colunas e tipos de dados que cada método retorna.

Entregáveis:

* Documentação rápida dos métodos e seus outputs (quais dados podem ser coletados).
* Script base para coleta inicial.

Sprint 2: Coleta e armazenamento dos dados históricos por temporada

Objetivo:

* Coletar dados por temporada das principais ligas europeias (2010 a 2026).
* Salvar os dados localmente organizados por ano e time em arquivos CSV.

Atividades:

* Criar script que itere por temporadas e ligas, usando o método read\_team\_match\_stats para coletar estatísticas por jogo.
* Para cada time na temporada, salvar CSV contendo dados detalhados.
* Tratar exceções e garantir logs de erros para temporadas sem dados.

Entregáveis:

* Pasta organizada com CSVs de dados por ano e time.
* Relatório de erros/ausências de dados.

Sprint 3: Ampliação da coleta para estatísticas individuais e análises avançadas

Objetivo:

* Coletar dados detalhados de jogadores (por temporada e por partida).
* Integrar dados de diferentes métodos (read\_player\_season\_stats, read\_player\_match\_stats).
* Analisar as colunas disponíveis e formatar os dados para análises comparativas.

Atividades:

* Criar script para coleta e armazenamento dos dados individuais.
* Padronizar nomes de times e jogadores para facilitar cruzamento.
* Documentar o dicionário de colunas mais relevantes para análise.

Entregáveis:

* Dataset organizado de jogadores com suas estatísticas históricas.
* Scripts documentados para coleta e limpeza.

Sprint 4: Preparação e integração dos dados no Power BI

Objetivo:

* Criar fluxos de ETL para importar dados dos CSVs para o Power BI.
* Construir relatórios iniciais de análise de desempenho por time e jogador.
* Desenvolver dashboards interativos com filtros por temporada, liga e estatísticas.

Atividades:

* Mapear os arquivos CSV para tabelas relacionais no Power BI.
* Construir modelos de dados para relacionar times, jogadores, temporadas e métricas.
* Criar visualizações e KPIs essenciais (gols, assistências, posse, etc).

Entregáveis:

* Power BI Desktop com relatórios e dashboards publicados.
* Guia rápido para atualização dos dados e relatórios.

Sprint 5: Sazonalidade, automatização e monitoramento

Objetivo:

* Desenvolver um catálogo sazonal para organizar os dados conforme campanhas, torneios e eventos.
* Automatizar a coleta e atualização dos dados periodicamente.
* Monitorar a qualidade e integridade dos dados.

Atividades:

* Definir lógica para catalogar temporadas e torneios.
* Criar scripts para agendamento e execução automática (ex: via cron ou Windows Task Scheduler).
* Implementar logs e alertas para falhas ou dados incompletos.

Entregáveis:

* Sistema automatizado para coleta e atualização dos dados.
* Relatórios de controle de qualidade.

Observações finais

* Cada sprint pode durar de 1 a 2 semanas, dependendo da sua disponibilidade.
* Podemos detalhar e criar scripts específicos para cada sprint, se quiser.
* Sempre podemos revisar e adaptar o plano conforme novas necessidades ou descobertas.