**Documentação do Projeto Power BI – Champions League**

**1. Extração dos Dados**

A extração dos dados foi realizada a partir do **Google BigQuery**, onde as consultas foram estruturadas para obter as informações brutas da base de dados.

**2. Storytelling, Definição de Problemas e Objetivos**

Para tornar a análise mais intuitiva e gerar insights valiosos, a estrutura do projeto seguiu uma abordagem de **storytelling**, garantindo que as informações fossem apresentadas de maneira clara e impactante.

**2.1. Definição do Problema**

* Identificar padrões de desempenho entre as equipes.
* Avaliar a relação entre o valor dos elencos e a idade média dos jogadores.
* Analisar a influência do público nos resultados das equipes.
* Verificar como as convocações para seleções impactam o valor do elenco.
* Apesar da base de dados cobrir o período de 2003 até 2023/2024, a maior parte das informações eram "fixas" independentemente da data. Isso afeta a precisão dos dados, mas como o foco do projeto era a aplicação dos conceitos, optou-se por manter a base mesmo com essa limitação.

**2.2. Objetivos do Projeto**

* Criar um painel dinâmico e interativo para facilitar a análise de KPIs esportivos.
* Auxiliar na tomada de decisão com insights visuais bem estruturados.
* Proporcionar uma experiência fluida na análise de equipes, estádios e público.
* Garantir que os dados sejam confiáveis, limpos e processados corretamente.

**3. Processo de ETL e Modelagem**

Para garantir a qualidade e a organização dos dados, o processo de **ETL (Extract, Transform, Load)** seguiu as etapas abaixo:

**3.1. Staging Area**

* Criada uma **Staging Area** no Power Query para armazenar a base de dados original.
* O carregamento dessa tabela foi desativado para manter um **backup** sem impactar o desempenho do modelo.

**3.2. Criação do Data Mart**

* A partir da Staging Area, foi criada uma **tabela Data Mart**, que passou por:
  + **Exclusão de colunas irrelevantes**.
  + **Conversão de tipos de dados** para padronização.
  + **Renomeação de colunas** para facilitar a referência nas tabelas subsequentes.
  + **Otimização para referenciamento**, sem carregamento direto no modelo.

**3.3. Modelagem Star Schema**

* Baseando-se na tabela Data Mart, foram criadas:
  + **Tabela Fato** contendo métricas e medidas quantitativas.
  + **Tabelas Dimensão** (ex.: Dim\_Equipes, Dim\_Estádios) com os atributos descritivos.
* A modelagem seguiu o padrão **Star Schema**, facilitando a performance e a análise dos dados.

**3.4. Web Scraping para Complemento de Dados**

* Como a base original não continha imagens dos **logos dos times** e dos **estádios**, utilizamos **Python** para obter essas informações.
* As bibliotecas utilizadas foram:
  + **Pandas** para manipulação dos dados.
  + **BeautifulSoup** para extração das imagens via Web Scraping.

**3.5. Tratamento e Limpeza dos Dados**

* Os dados foram tratados e limpos utilizando **Power Query**.
* Foram criadas **diversas medidas** para os KPIs, otimizando cálculos dinâmicos no Power BI.

**4. Criação de Dashboards Intuitivos e Interativos**

Foram criados dashboards no Power BI com visualizações organizadas da seguinte forma:

**4.1. Geral**

* **KPIs:**
  + Início dos dados
  + Final dos dados
  + Média do valor do elenco
  + Idade média geral
  + Média do valor dos titulares
* **Gráficos:**
  + Valor total do elenco titular por equipe
  + Idade média por equipe
  + Valor médio dos jogadores titulares
  + Convocações para a seleção principal
  + Relação entre valor do elenco e idade média

**4.2. Equipe**

* **Seleção interativa da equipe** com os seguintes KPIs:
  + Nome da equipe
  + Estádio da equipe
  + Média de público
  + Valor do elenco
  + Idade média
  + Convocados seleção júniores
  + Convocados seleção principal
  + Gols marcados como mandante
  + Gols sofridos como mandante
  + Média de público como mandante por ano
  + Imagens da logo e do estádio da equipe selecionada

**4.3. Público**

* **KPIs:**
  + Maior público
  + Estádio com maior público
  + Menor público
  + Estádio com menor público
* **Gráficos:**
  + Média de público por etapa
  + Média de público por estádio
  + Média de público por temporada
  + Média de público por mês
  + Média de público por horário

**5. Referências Técnicas e Conceitos Utilizados**

* **Ferramentas**: Power BI, Power Query, Google BigQuery, Python, Pandas, BeautifulSoup.
* **Modelagem de Dados**: Star Schema, Tabela Fato, Tabelas Dimensão, Data Mart, Staging Area.
* **Linguagens**: M Language (Power Query), SQL (BigQuery), Python.
* **Técnicas e Conceitos**: ETL, Web Scraping, Validação de Dados, Conversão de Tipos, Renomeação de Colunas, Correção de Erros, Storytelling, KPIs, Dashboards Interativos.

**Desenvolvido por Lucca Ferreira Marques**