

Visualização da informação

Escola de Matemática Aplicada - Fundação Getúlio Vargas

Alunos: *Edilton Brandão & Erick Brito*

Projeto Final: Briefing

Motivação

Em pleno 2022, ano de Copa do Mundo, ano de Hexa!

Não é de hoje que a paixão pelo futebol está no coração dos brasileiros. E o que incendiaria mais essa paixão do que uma Copa do Mundo? Ainda mais com a chance de um hexacampeonato da seleção brasileira.

Tomados pelos sentimentos que esse evento nos proporciona, decidimos fazer nosso projeto final da disciplina sobre **Copa do Mundo**.

Objetivo

A ideia do nosso projeto é criar uma interface via streamlit que contenha resumos das participações de cada seleção nas copas, desde 1930 a 2022 e também algumas visualizações interessantes, permitindo assim, que o usuário tenha um fácil acesso às participações e resultados de uma seleção específica na Copa do Mundo do ano desejado.

Base de Dados

A base de dados que será utilizada no trabalho final é a seguinte: [International football results from 1872 to 2022](#).

Este dataset inclui 44.152 resultados de partidas internacionais de futebol, desde a primeira partida oficial em 1872 até 2022. As partidas variam da Copa do Mundo da FIFA à FIFI Wild Cup e amistosos regulares.

Embora outros torneios estejam presentes, inicialmente nosso interesse é apenas nos resultados da Copa do Mundo da FIFA, por isso ainda vamos fazer uma filtragem na etapa de exploração dos dados, além disso, vamos explicar melhor sobre as colunas presentes no dataset, então aqui vamos dar apenas uma breve olhada nos dados.

Criamos um repositório para o projeto: <https://github.com/Ericks1b/copa-do-mundo-vis>.

Subimos os dados para o repositório e vamos lê-los direto de lá.

```
In [ ]: import pandas as pd
url = "https://raw.githubusercontent.com/Ericks1b/copa-do-mundo-vis/main/data/results.csv"
data = pd.read_csv(url)

In [ ]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44206 entries, 0 to 44205
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   date        44206 non-null   object
 1   home_team    44206 non-null   object
 2   away_team    44206 non-null   object
 3   home_score   44202 non-null   float64
 4   away_score   44202 non-null   float64
 5   tournament   44206 non-null   object
 6   city         44206 non-null   object
 7   country      44206 non-null   object
 8   neutral      44206 non-null   bool
dtypes: bool(1), float64(2), object(6)
memory usage: 2.7+ MB

In [ ]: data.head()

Out [ ]:    date home_team away_team home_score away_score tournament city country neutral
0  1872-11-30  Scotland  England         0.0         0.0  Friendly  Glasgow  Scotland  False
1  1873-03-08   England  Scotland         4.0         2.0  Friendly   London   England  False
2  1874-03-07  Scotland  England         2.0         1.0  Friendly  Glasgow  Scotland  False
3  1875-03-06   England  Scotland         2.0         2.0  Friendly   London   England  False
4  1876-03-04  Scotland  England         3.0         0.0  Friendly  Glasgow  Scotland  False
```

Algumas das principais variáveis que vamos trabalhar são *home_team*, *away_team*, *home_score*, *away_score*, *country* e *date*, pois é a partir delas que vamos obter a informação de, por exemplo, o número de gols de cada país ao longo dos anos.

Tópicos interessantes

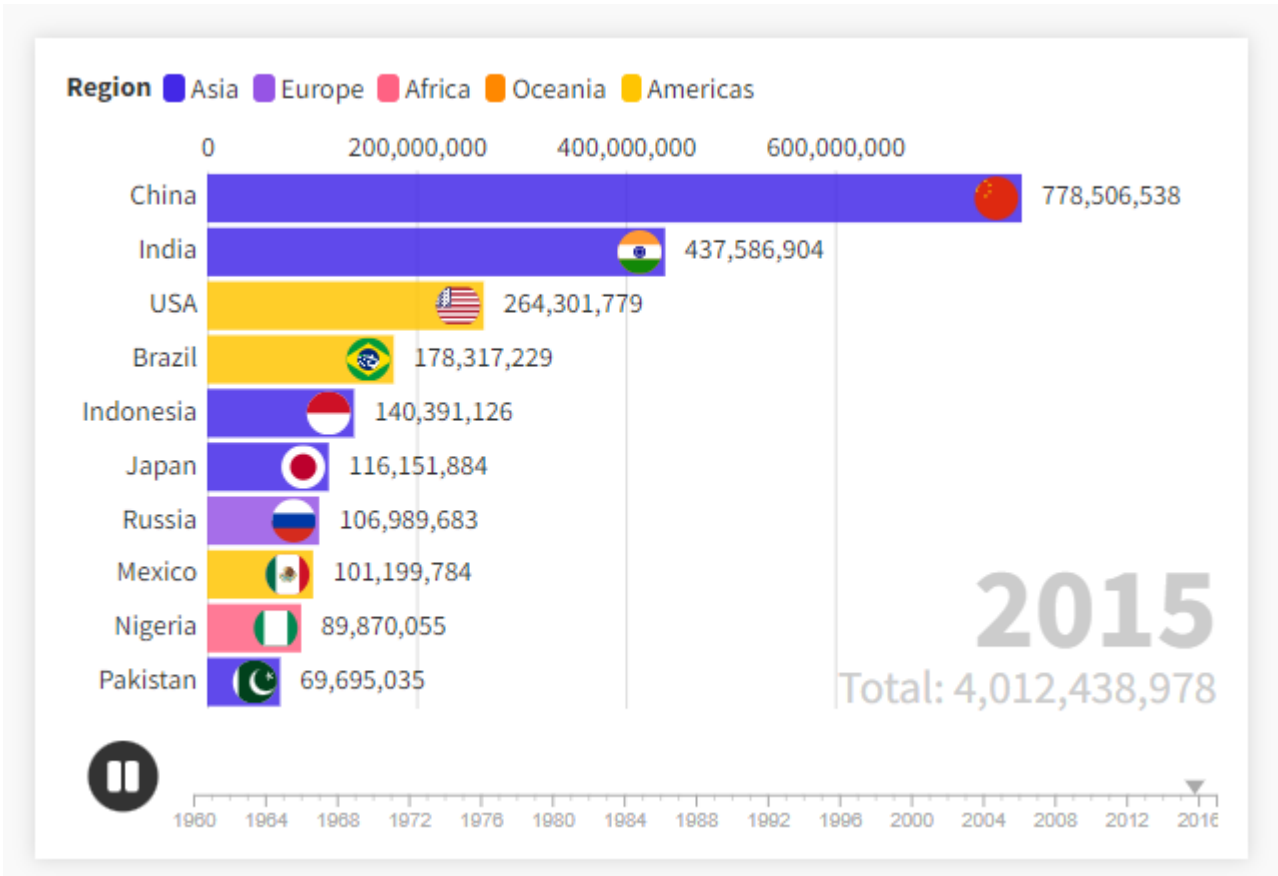
- Hipótese sobre os dados
 - Seleções com mais gols tendem a ter melhores desempenhos na copa (mais participações, mais títulos)
- Quem é o público alvo?
 - Amantes de futebol e pessoas que desejam ter uma visão geral do desempenho das seleções que participaram da Copa do Mundo, ao longo da história
- O que pode ser interessante de encontrar nos dados?
 - Descobrir quais seleções são recordistas de gols
 - Quais as seleções que mais participaram e as que menos participaram de Copa do Mundo
 - Seleções que menos tomaram gols e seus desempenhos

As circunstâncias

- Restrições de prazo
 - 12 de dezembro
- Como a visualização será consumida
 - O objetivo é hospedar a visualização numa aplicação web através da biblioteca Streamlit do Python para que seja possível a interação do usuário com os gráficos.
- Quais são os entregáveis?
 - Relatório final (.pdf)
 - Interface streamlit (web)

Referências de inspiração

Para comparar o saldo de gols acumulativo das seleções ao longo de todas as copas, desejamos criar um *bar chart race*, pois permite uma visualização dinâmica de fácil compreensão e divertida de se visualizar.



<https://app.flourish.studio/@flourish/bar-chart-race#guide>

Desejamos também não só comparar o número de gols, por isso a ideia é que exista um filtro onde o usuário possa escolher o que deseja visualizar (ex: número de faltas, número de títulos, etc..)

Outra referência é um aplicativo interativo para analisar os dados da Bundesliga(campeonato de futebol alemão) da temporada 2013/2014 à temporada 2019/2020 feito via streamlit. Ele resume, de forma interativa, as estatísticas dos jogos que aconteceram nessas temporadas.

<https://tdenzl-bulian-bulian-ifeih.streamlit.app/>