

Visualização de Informação | Visualização de dados disponíveis num dataset

Lucas Barros | 83895, João Abílio Rodrigues | 84732

Mestrado em Engenharia Informática, Universidade de Aveiro, 2019

Resumo

Este relatório tem como propósito a apresentação de uma solução para permitir visualizar dados informativos sobre o mundo do futebol. Ao longo do mesmo será explicado todo o raciocínio das ferramentas de visualização usadas e serão apresentados alguns resultados e feedback.

Motivação e Objetivos

O d3 é uma ferramenta muito poderosa que pode ser usada para criar visualizações gráficas para uma melhor interpretação e visualização de dados.

No âmbito da Unidade Curricular de Visualização de Informação, foi proposto aos alunos a escolha de um dataset de forma a realizar uma plataforma de visualização de dados e permitir a utilizadores aceder a informação útil. Nesta plataforma temos como objetivos acompanhar a evolução de jogadores, ver estatísticas de equipas e permitir ao utilizador comparar equipas dentro de ligas.

Utilizadores e questões

De modo a compreender as necessidades desta plataforma, criamos personas, pois são estas que nos fazem perceber o que o nosso sistema necessita. Ao longo deste processo registámos perguntas a que este trabalho teria de responder para ser concluído com sucesso.

Caraterização dos utilizadores e o seu contexto

Francisco Almeida, de 20 anos, é um estudante universitário que gosta de estar a par de todos os pormenores do futebol. Tendo pouco tempo para se atualizar devido aos estudos na Universidade, dá jeito a ele ter um sítio onde possa ver de forma rápida e atualizada esses dados.

Sérgio Trindade, de 46 anos, é um adepto de futebol que faz apostas no Placard e gostaria de estar a par das estatísticas das equipas e jogadores para realizar apostas mais seguras.

Perguntas para responder

Q1: Qual o melhor marcador da Bundesliga?

Q2A: Ver a capacidade de passe curto de um jogador (ex. Bruno Fernandes).

Q2B: Ver um overview das capacidades de um jogador.

Q3: Ver estatísticas de uma equipa (ex. Sporting Clube de Portugal)

Q4: Ver a equipa de uma liga (ex. Bundesliga) que tem o melhor score.

Dataset

O nosso trabalho foi baseado num dataset que apresenta várias estatísticas sobre o mundo do futebol, desde jogadores a ligas.

O dataset tem o nome de “European Soccer Database” e foi adquirido do website Kaggle [1].

O formato do dataset é *SQLite*. De forma a podermos utilizar os dados, criamos um script em *Python3* para processar a informação para vários ficheiros com formato *CSV (Comma Separated Values)*, sendo repartidas por várias colunas.

Solução

No desenvolvimento da solução criamos um protótipo de baixa fidelidade para começar a pensar no design e visualizações a utilizar, com a ajuda do feedback dos utilizadores. De seguida procedemos à implementação do site, em que tivemos em conta o feedback obtido dos utilizadores e as questões a responder.

Protótipo de baixa fidelidade e feedback

O protótipo de baixa fidelidade foi um protótipo em papel. Tínhamos 3 tipos de visualizações: gráfico circular, gráfico de barras e gráfico de ranking. O feedback obtido sobre o nosso protótipo revelou que seria inútil utilizar o gráfico circular e que seria benéfico nas nossas visualizações utilizar um *starplot*.

Protótipo funcional

Neste processo passamos ao desenvolvimento do site. As próximas figuras representam as visualizações que respondem às perguntas feitas.

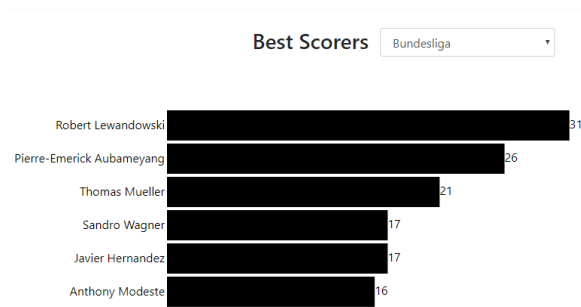


Figura 1: Visualização para responder à questão Q1.

Bruno Fernandes - 75

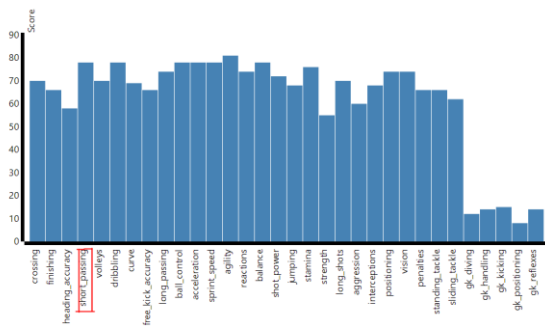


Figura 2: Visualização para responder à questão Q2A.

Esta visualização, na primeira implementação era uma tabela, no entanto, foi-nos sugerido durante a apresentação que devido ao volume dos dados e às suas características em relação aos limites dos seus possíveis valores, que mudássemos para um gráfico de barras.

Bruno Fernandes - 75

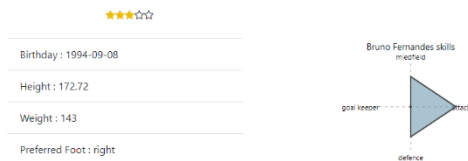


Figura 3: Visualização para responder à questão Q2B

Sporting CP

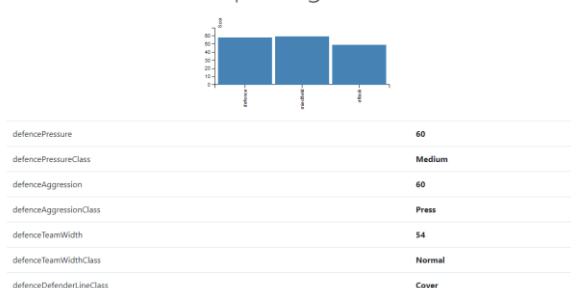


Figura 4: Visualização para responder à questão Q3.

Bundesliga

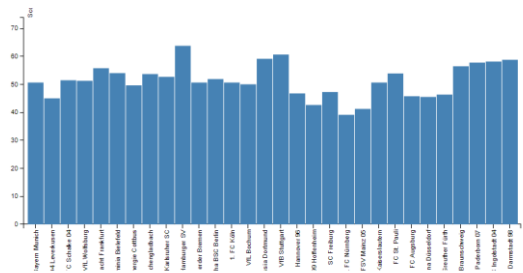


Figura 5: Visualização para responder à questão Q4.

Desafios na implementação

Os dois desafios que tivemos na implementação foram aprender a reproduzir visualizações com a livreria javascript d3[2], principalmente o *starplot*, devido à pouca documentação relativa a esta visualização e, como obter os dados em SQLite para CSV. Para tal, foi utilizado a biblioteca de Python3 “sqlite”. Para produzir o starplot, foi utilizado um script auxiliar que se encontra na pasta “scripts”.

Avaliação e mudanças no protótipo

A transição entre o protótipo de baixa fidelidade e o protótipo funcional, como se viu anteriormente, sofreu algumas alterações, devido ao feedback recebido, assim como algumas das questões apresentadas não serem totalmente respondidas no protótipo em papel que depois no site foram solucionadas. Assim acabamos por apresentar as seguintes visualizações: starplot, gráfico de barras e gráfico de ranking.

Conclusão e Trabalho Futuro

Com a realização deste trabalho aprendemos a manipular dados e a criar visualizações em d3.

Como trabalho futuro, pensámos em desenvolver um sistema de autocomplete na barra de procura, no entanto tal não foi possível pois não encontrámos documentação d3 que mostrasse como o fazer, no entanto com um pouco mais de tempo, podia ter sido possível proceder à implementação. Também pensamos em mostrar mais estatísticas que não apenas os melhores marcadores na página inicial, por exemplo, os jogadores com mais assistências, mais cartões (amarelos e vermelhos), porém tal não foi possível também devido à falta de tempo

References

- [1] European Soccer Database <https://www.kaggle.com/hugomathien/soccer>
- [2] Data-Driven Documents, <https://d3js.org/>