

Dados Portugal: Visualização de Dados Estatísticos

Diogo Carvalho - 92969 - dfac@ua.pt

Rafael Baptista - 93367 - rafaelbaptista@ua.pt

Visualização de Informação, Mestrado em Engenharia Informática, Universidade de Aveiro, 2022

Resumo

Neste relatório está apresentado todo o trabalho realizado para o desenvolvimento da nossa plataforma Web, que possui o objetivo de permitir explorar dados sobre diferentes áreas da nossa sociedade, mais especificamente, a área da população, saúde e educação em Portugal com dados referentes aos últimos anos.

1. Motivação e Objetivos

Decidimos selecionar este tema uma vez que consideramos que o conjunto de dados que iremos tratar ao longo deste projeto contém informação relevante relativa a aspetos importantes da nossa sociedade, como atributos (idade, género...) da população portuguesa e estatísticas relativas ao setor da educação e saúde.

Com estas informações pretendemos aprofundar os nossos conhecimentos e dos nossos colegas relativos à caracterização da população portuguesa e ao desenvolvimento de importantes setores que constituem pilares fundamentais da nossa sociedade e que são muitas vezes referenciados e alvo de discussão, quer seja a um nível mais informal ou até mesmo a nível político.

Posto isto, temos como principal objetivo colmatar a falta de uma plataforma online que permita a visualização deste tipo de dados de uma forma detalhada e eficiente, permitindo retirar conclusões e suportar decisões com base na análise das visualizações de dados desenvolvidas.

2. Utilizadores e Questões

Numa primeira fase do projeto, efetuou-se um estudo e planeamento da plataforma. Neste planeamento, foi efetuada a seleção do tipo de questões que podem ser respondidas pela plataforma, bem como os principais utilizadores alvo da aplicação.

2.1 Caracterização dos Utilizadores

Neste estudo, que serviu posteriormente como suporte para a implementação da nossa plataforma, verificámos que o grupo de potenciais utilizadores da nossa plataforma é composto por qualquer indivíduo que pretenda obter informações relativas à caracterização da população portuguesa, seja com propósito profissional ou apenas para satisfação de curiosidade pessoal, existindo uma maior tendência de utilização por parte de pessoas ligadas aos setores da saúde e da educação, visto que estes setores possuem também uma área de análise na nossa plataforma. Este conjunto de pessoas pode incluir, por exemplo, indivíduos interessados em dados estatísticos sobre os temas referidos, que procuram compreender a evolução destes dados ao longo do tempo ou simplesmente analisar os dados mais recentes sobre cada área.

Para além destes utilizadores, a nossa plataforma pode também ser utilizada para fins académicos, onde pode ser efetuada a recolha de dados, seja no seu estado bruto ou através da aquisição da imagem das visualizações criadas no desenvolvimento da plataforma.

2.2 Enumeração Questões

Nesta mesma análise ao conjunto de utilizadores, também definimos algumas das possíveis questões relevantes que um utilizador conseguiria obter uma resposta com a utilização da nossa aplicação. Segue-se a lista de questões definidas durante essa análise:

1. Qual era a esperança média de vida da população portuguesa em 1980?
2. Qual é a densidade populacional no distrito de Faro?
3. Qual era o número de internados nos Hospitais em 2020?
4. Quantos médicos na especialidade de cirurgia geral existiam em Portugal no ano 2010?

5. Qual é o número de mulheres que estudaram no ensino superior em 2000?
6. Compare a média das notas dos exames de secundário do ano de 2020 e 2021. Em qual dos anos se obteve uma maior média para a disciplina de Matemática A?

3. Conjunto de Dados

O conjunto de dados utilizado para o desenvolvimento da plataforma contém informação relativa às áreas populacional, educação e saúde.

Todos os dados utilizados foram recolhidos da mesma fonte. A obtenção dos mesmos foi efetuada através da plataforma online da PORDATA. Todos os dados contêm um carácter nacional, à exceção da distribuição da população portuguesa por distritos, que possui um carácter regional.

Mais concretamente, relativamente à área da população obtivemos os seguintes dados:

- Distribuição da população portuguesa por faixa etária ao longo dos últimos anos.
- Evolução da esperança média de vida ao longo dos últimos anos.
- Distribuição da população portuguesa por género ao longo dos últimos anos.
- O valor da densidade populacional por distrito em 2020.

Relativamente à área da saúde:

- O número total de camas em diferentes tipos de estabelecimentos de saúde ao longo dos últimos anos.
- O número de consultas, internamentos e urgências por estabelecimento de saúde ao longo dos últimos anos.
- O número de médicos por especialidade ao longo dos últimos anos.

Relativamente à área da educação:

- O número de pessoas por nível de ensino e por género ao longo dos últimos anos.
- A média das notas dos exames de secundário em cada disciplina ao longo dos últimos anos.
- Taxa de abandono precoce de educação por género dos jovens de 18 aos 25 anos ao longo dos últimos anos.

4. Solução de Visualização

Começámos por construir um protótipo de baixa fidelidade realizado em papel por forma a conseguirmos realizar testes de usabilidade a diversos utilizadores durante as aulas práticas da cadeira de Visualização de Informação.

Para a execução dos testes de usabilidade com o protótipo de baixa fidelidade desenvolvemos, para além do próprio protótipo, um guia para o observador, isto é, pessoa que esta a conduzir o teste e outro para o participante, ou seja, pessoa que está a desempenhar o teste.

Os guiões tinham como principal objetivo não só guiar os participantes ao longo de todo o teste, reduzindo assim a necessidade de intervenções por parte do observador, permitindo que os resultados dos testes fossem mais precisos e próximos da realidade, bem como também permitir a que tanto o observador como o participante preenchessem um pequeno questionário para cada tarefa relativo à dificuldade da tarefa, à capacidade ou não de conclusão da mesma, à necessidade de intervenção do observador e onde era possível escrever observação caso o participante desejasse.

Com base no feedback destes testes, adaptámos as mudanças ao nosso protótipo e começámos a desenvolver a nossa aplicação.

4.1 Protótipo de Baixa Fidelidade

Com o objetivo de realizar testes de usabilidade que permitissem validar a nossa solução de implementação, bem como com o propósito de utilizar como suporte e guia durante a implementação da nossa plataforma, desenvolvemos o protótipo de baixa fidelidade em papel que se segue:

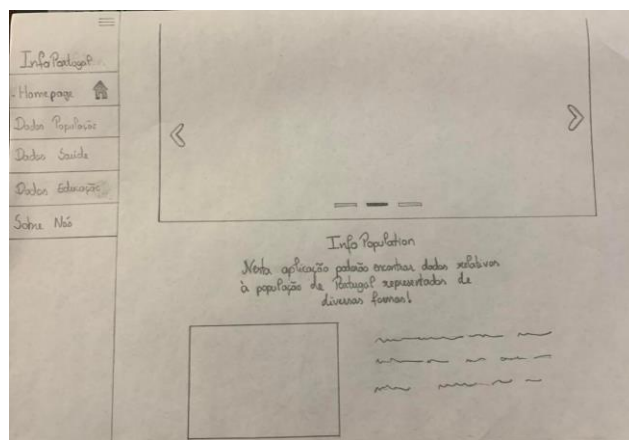


Figure 1: Página inicial

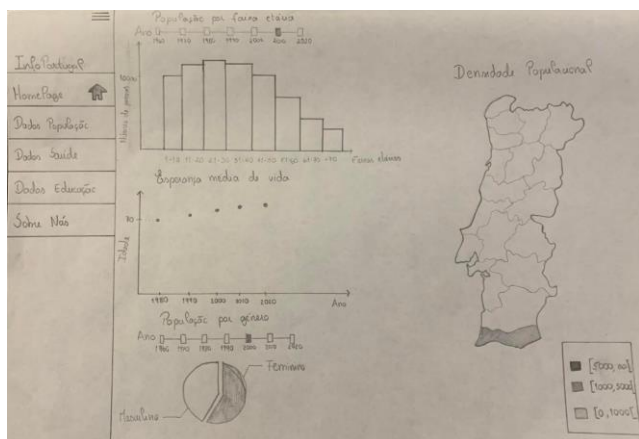


Figure 2: Página população

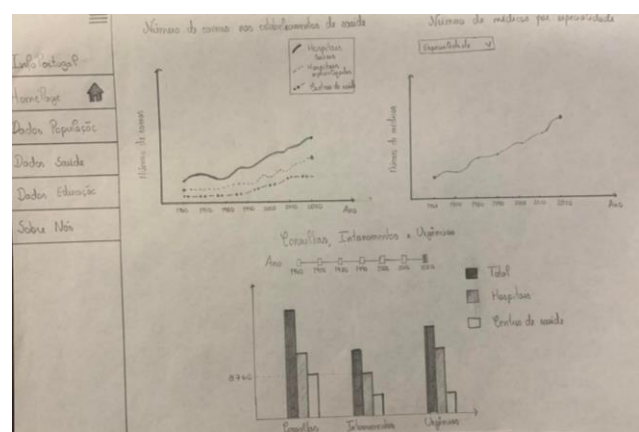


Figure 3: Página saúde

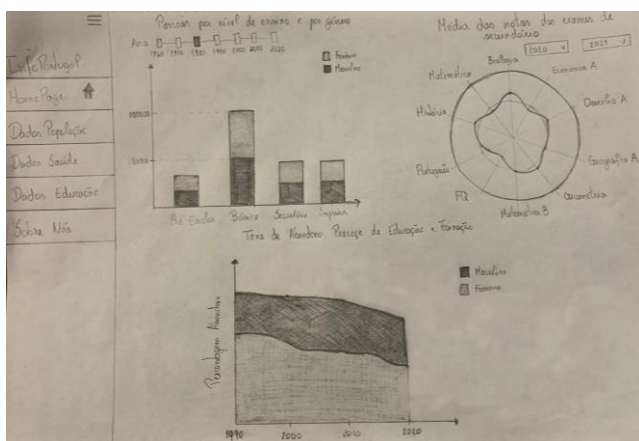


Figure 4: Página educação

No desenvolvimento deste protótipo, decidimos distribuir as diferentes áreas dos dados (população, saúde e educação) em páginas distintas, de modo que o utilizador consiga encontrar os dados que pretende de forma mais rápida e eficiente.

Iniciando pela página que contém os dados relativos à população, decidimos representar os dados com as seguintes visualizações:

- Histograma para o número de população pelas diferentes faixas etárias, com opção de filtragem por ano.
- Gráfico de pontos para a esperança média de vida ao longo dos anos.
- Pie chart para o número de população por género, com opção de filtragem por ano.
- Mapa de Portugal com escala de cores para representar a densidade populacional de cada distrito do país.

Relativamente aos dados referentes à área da saúde decidimos utilizar as seguintes visualizações:

- Gráfico de linhas para o número de camas dos diferentes estabelecimentos de saúde por ano.
- Gráfico de colunas triplas para o número de consultas, internamentos e urgências nos diferentes estabelecimentos de saúde, com opção de filtragem por ano.
- Gráfico de linhas para o número de médicos por especialidade ao longo dos anos, com opção de filtragem pela especialidade.

Por fim, para os dados referentes à área da educação decidimos representar através das seguintes visualizações:

- Gráfico de barras empilhadas para o número de pessoas por nível de ensino e por género, com opção de filtragem por ano.
- Gráfico de radar para representar a média das notas dos exames de secundário ao longo dos anos, com opção de filtragem e comparação entre anos.
- Gráfico de área para representar a taxa de abandono precoce de educação e formação por género ao longo dos anos.

Como referido acima, após a construção do protótipo de baixa fidelidade, foram realizados testes de usabilidade.

Através dos resultados obtidos com esses testes e do feedback emitido pelos utilizadores decidimos efetuar algumas mudanças relativas ao protótipo:

- No gráfico de médicos por especialidade ao longo dos anos, foi referido que seria útil uma comparação mais direta entre os valores do número de médicos em cada especialidade. Posto isto, invés de colocarmos um gráfico de

uma só linha com um input de alteração da especialidade, optamos por colocar todas as especialidades no gráfico e dar a opção ao utilizar de através de um conjunto de switches on/off, retirar e colocar as linhas que desejar, para comparar as diferentes especialidades entre si.

- Em todos os gráficos utilizados que fazem uso de barras para representar dados, era complicado verificar qual o valor ao certo que uma determinada barra possui, e por vezes, quando as alturas das barras eram idênticas e a sua distância elevada, poderia ser difícil compará-las. Posto isto, optamos por criar uma tooltip em que, sempre que o utilizar passar o rato do computador por cima de uma determinada barra, surge uma informação mais detalhada da mesma com o valor concreto que a mesma representa.

Com base no protótipo inicial e nas mudanças referidas acima, começamos o desenvolvimento do protótipo funcional.

4.2 Protótipo Funcional

Tendo como base o protótipo desenvolvido anteriormente, a aplicação possui uma página inicial e possui um menu lateral com links para as páginas relativas a cada área já referida.

Começando com a área da população, face às visualizações definidas anteriormente optámos por desenvolver tudo conforme o protótipo, à exceção do pie chart, que, por motivos estéticos, acabámos por desenvolver um donut chart.

Analisando cada gráfico de forma mais específica:

Na população por faixa etária o utilizador consegue filtrar os dados por ano, e ao colocar o cursor por cima de cada barra, consegue obter o valor exato que aquela barra representa, bem como é mostrado uma linha horizontal a tracejado por forma a facilitar a comparação de várias faixas etárias.

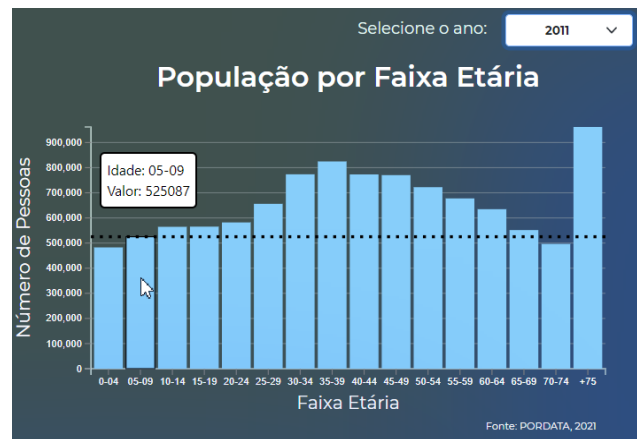


Figure 5: Histograma. População por Faixa Etária.

Na visualização da esperança média de vida, o utilizador consegue também interagir com o gráfico. Quando coloca o cursor em cada ponto do gráfico, obtém o valor exato daquele ponto e também é mostrada a linha horizontal, de forma idêntica à visualização anterior

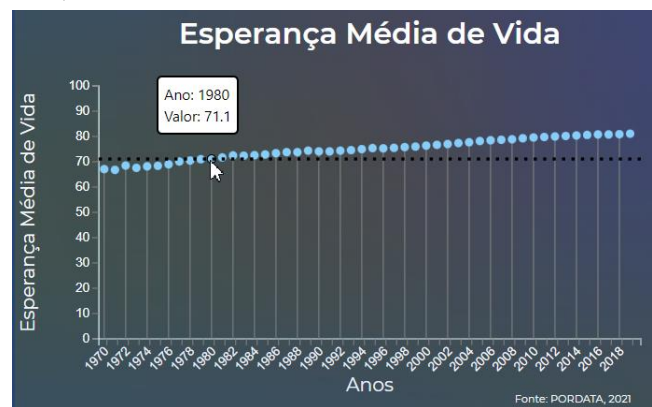


Figure 6: Gráfico de Pontos. Esperança Média de Vida. Resposta à Questão 1.

Para representar a distribuição da população por género, optámos por desenvolver um gráfico de Donut.

Deste modo, o utilizador tem a possibilidade de escolher o ano que pretende obter os dados através de um filtro e tem também a possibilidade de obter o valor exato quando coloca o cursor por cima de cada secção do gráfico.

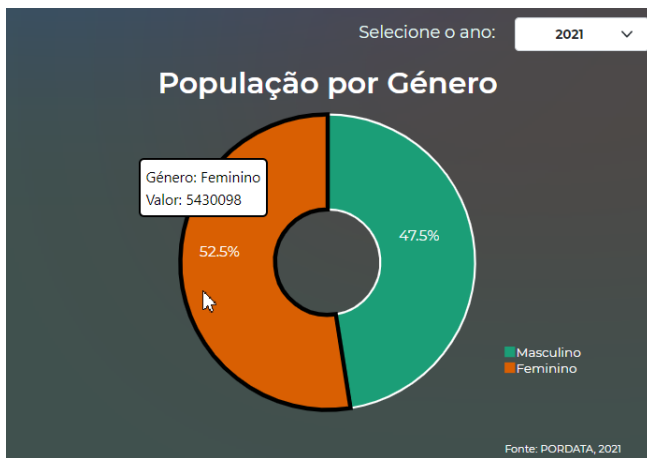


Figure 7: Gráfico Donut. População por Género.

Para representar a densidade populacional, construímos um mapa de Portugal com os diferentes distritos, incluindo as regiões autónomas da madeira e açores, onde o utilizador consegue ter uma noção dos dados apenas a visualizar o mapa através das cores atribuídas a cada distrito.

Para além disso, também é possível obter os valores exatos de densidade para cada distrito quando é colocado o cursor por cima do distrito desejado.

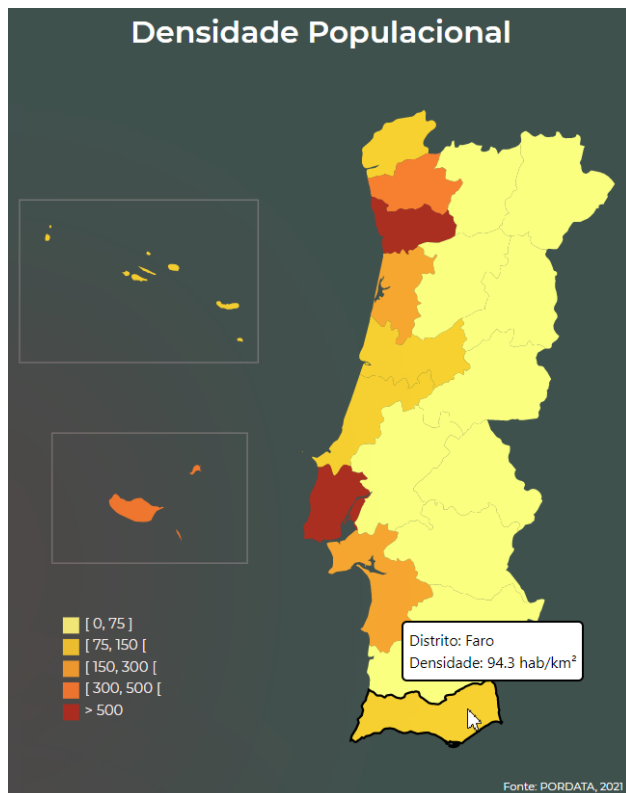


Figure 8: Mapa de Portugal. Densidade Populacional. Resposta à Questão 2.

Prosseguindo para a área da saúde, começando com o número de camas por estabelecimento de saúde, trata-se de um gráfico de linhas sendo que cada tipo de estabelecimento tem uma cor de linha associada. As cores foram seleccionadas de forma a se conseguirem distinguir facilmente.

Para além disso, o utilizador tem a possibilidade de colocar o cursor no ponto que pretender para obter o valor exato do mesmo.

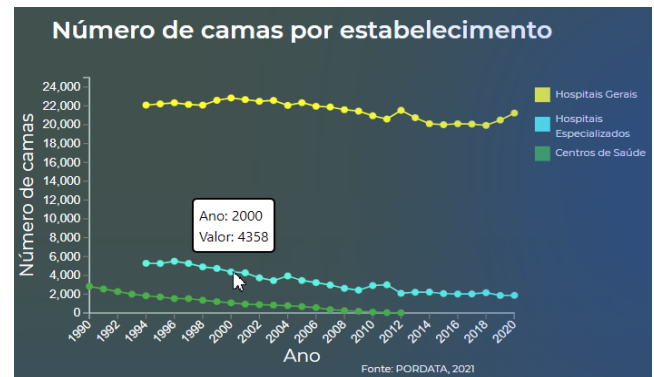


Figure 9: Gráfico de linhas. Número de camas por estabelecimento.

Relativamente à representação do número consultas, internamentos e urgências por tipo de estabelecimento de saúde ao longo dos últimos anos, a opção foi a utilização de um gráfico de colunas triplas. Visto existir a falta de dados relativamente a alguns anos para alguns estabelecimentos, foi adicionada uma representação para os dados não disponíveis como demonstrado na figura através de um padrão cinzento.

O utilizador tem a capacidade de filtrar o ano que pretende visualizar os dados, bem como a capacidade de obter o valor concreto das barras ao colocar o cursor nas mesmas, apresentando ainda uma linha horizontal pontuada para uma mais fácil comparação entre barras.



Figure 10: Gráfico de colunas triplas. Número de Consultas, Internamentos e Urgências. Resposta à Questão 3.

Relativamente à representação do número de médicos por especialidade, trata-se de um gráfico de linhas sendo que cada especialidade tem uma cor associada.

Como referido anteriormente, foi desenvolvida uma forma que permite ao utilizador ter a possibilidade de filtrar quais as especialidades que pretende visualizar no gráfico, podendo adicionar ou retirar à sua vontade.

Para além disso tem também a possibilidade de saber os valores exatos dos dados ao colocar o cursor no ponto que pretender.

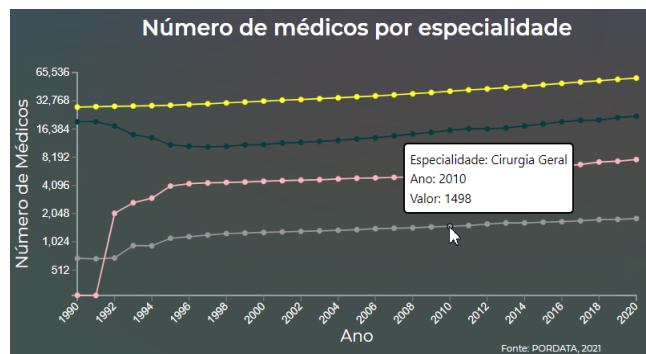


Figure 11: Gráfico de linhas. Número de médicos por especialidade. Resposta à Questão 4.

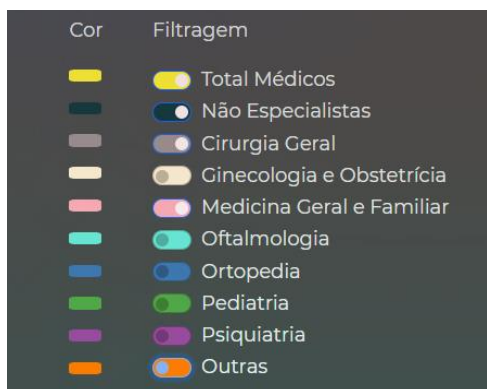


Figure 12: Filtragem do Gráfico da Visualização do Número de médicos por especialidade.

Por fim, na área da educação, para o número de alunos por nível de ensino e por género, implementámos um gráfico de colunas empilhadas em que cada género tem uma cor diferente e o utilizador tem também a possibilidade de escolher o ano que pretende ver os dados. Este gráfico de colunas empilhadas permite ao utilizador conseguir obter uma perceção mais rápida do valor total de alunos (masculinos e femininos) por tipo de ensino num determinado ano.

Para além disso, o utilizador consegue saber os valores exatos bem como o valor total ao colocar o cursor por cima de cada coluna.



Figure 13: Gráfico de colunas empilhadas. Número de Alunos por nível de ensino e género. Resposta à Questão 5.

Avançando para a representação da média das notas dos exames do ensino secundário, estes dados estão representados sob a forma de um gráfico de radar ou gráfico aranha. Esta representação permite ao utilizador sobrepor os dados de dois anos distintos, permitindo ao mesmo comparar os resultados dos diferentes anos.

Posto isto, o utilizador é então capaz de escolher dois anos de forma a poder compará-los, sendo que cada um irá ter associada uma cor diferente. Para além disso, o utilizador ao colocar o cursor nos pontos, consegue obter os valores exatos das notas.



Figure 14: Gráfico de Radar. Média das notas dos exames de secundário. Resposta à Questão 6.

Por último, a taxa de abandono precoce de educação e formação por género de adolescentes dos 18 aos 25 anos, está representada num gráfico de área. Esta representação permite ao utilizador ter uma noção mais concreta da

evolução dos dados ao longo dos anos, bem como perceber qual das áreas (que neste caso se associam às ‘classes’ masculino e feminino) tem um papel maior para os valores da taxa que estamos a observar.

Em adição, é colocada uma linha com o total, ou seja, com a evolução da taxa ao longo dos anos sem discriminação de género. O utilizador tem a capacidade de adicionar ou remover esta linha através de um switch on/off.

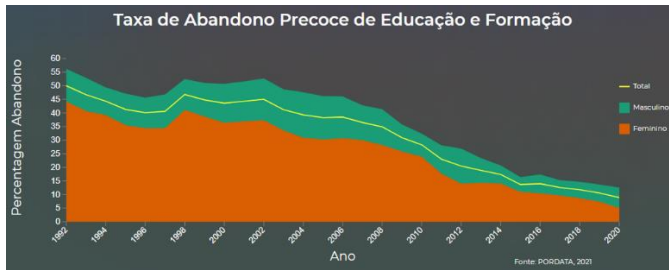


Figure 15: Gráfico de área. Taxa de abandono precoce de educação e formação.

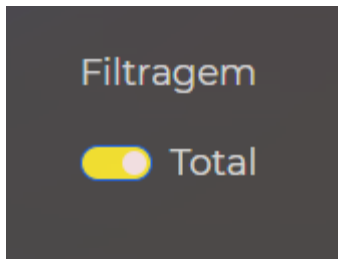


Figure 16: Filtragem do Gráfico de Visualização da Taxa de abandono precoce de educação e formação

4.3 Desafios de Implementação

Para desenvolvermos a nossa plataforma utilizamos as seguintes tecnologias:

- HTML;
- CSS;
- Javascript
- Biblioteca D3.js (para a criação dos gráficos de visualização).

Toda a componente de design e estilo da plataforma foi desenvolvida de raiz. Todos os aspetos funcionais da aplicação foram também criados por nós.

Relativamente às representações gráficas, as mesmas foram desenvolvidas na sua grande maioria por nós, utilizando por base implementações já existentes na internet, sobretudo no site oficial da biblioteca d3.js que possui exemplos de código base para os diferentes tipos de gráficos.

Relativamente aos desafios que encontrámos, o maior desafio foi o desenvolvimento das próprias visualizações em si, bem como o trabalho associado para as tornar naquilo que idealizámos e que acreditamos que permita dar resposta às necessidades dos nossos utilizadores. Associado a este ponto, o desenvolvimento do código utilizando a biblioteca D3.js acabou por se revelar um desafio por si só, uma vez que era uma ferramenta nova para nós, onde necessitamos de aprender e explorar primeiro para compreender a utilizar a mesma.

Por fim, visto que os dados que obtivemos da PORDATA e utilizámos na nossa plataforma possuíam um formato não definido por nós, precisámos também de converter os dados num formato que fosse útil para nós e que pudéssemos utilizar para representar, necessitando assim de algum pré processamento visto que os dados estavam em Excel.

4.4 Avaliação e alterações no protótipo

Relativamente à avaliação do nosso protótipo funcional, para efetuar a mesma, procedemos à realização de um conjunto de testes de usabilidade com a nossa nova plataforma desenvolvida.

Estes testes foram realizados para um conjunto de seis pessoas, todas elas pertencentes ao nosso agregado familiar. Durante o teste, foi pedido a cada um dos participantes que respondesse às mesmas questões apresentadas na secção 2.2. De forma idêntica aos testes realizados para o protótipo de baixa fidelidade, foi registada a dificuldade de execução, necessidade de ajuda e capacidade de conclusão da tarefa por parte do participante.

Relativamente aos resultados obtidos, todos os participantes conseguiram realizar as tarefas sem necessidade de ajuda. No entanto, na secção de observações, dois dos participantes referiram que, para responder à pergunta número 1 “Qual era a esperança média de vida da população portuguesa em 1980?” foi difícil decidir se os dados relativamente à mesma se encontravam na área referente à população ou na área referente à saúde. Não foram então efetuadas quaisquer alterações ao protótipo, visto que esta pequena observação levantada na verdade não constitui um problema porque, caso o utilizador se engane e aceda à área da saúde, rapidamente irá tentar encontrar os dados na secção de população.

Por fim, foi realizada uma avaliação heurística da solução final pelos dois autores do trabalho, seguindo as dez heurísticas descritas por Jakob Nielsen. Nesta fase, cada um

dos avaliadores efetuou de forma independente uma avaliação com 5 níveis para cada heurística e no fim houve uma discussão dos resultados obtidos.

Na discussão final, a única heurística que foi avaliada com um valor igual ou superior a 1 por qualquer um dos dois avaliadores, e que por isso poderia não estar a ser cumprida, foi a heurística relativa à Ajuda e Documentação, visto que a nossa plataforma não possui qualquer documentação. No entanto, consideramos que a nossa plataforma é bastante simples e intuitiva, e qualquer dúvida na interação com a mesma pode ser esclarecida através da área de contacto presente na plataforma.

5. Conclusão e Trabalho Futuro

Concluimos assim este projeto com uma avaliação satisfatória do nosso trabalho, onde conseguimos alcançar os objetivos a que nos propusemos, desenvolvendo uma plataforma capaz de dar resposta às questões levantadas e que pode servir de apoio a qualquer utilizador que pretenda obter dados e/ou analisar os mesmos.

Para além disso, estamos também bastante satisfeitos com o desenvolvimento das nossas capacidades relativas à biblioteca d3.js. Conseguimos desenvolver diversas representações gráficas distintas, todas elas capazes de dar resposta às necessidades dos nossos utilizadores, enquanto representam os dados apelativa e interactivamente.

Como trabalho futuro, desafiamos qualquer pessoa que tenha interesse em desenvolver uma plataforma mais robusta semelhante a esta a primeiramente efetuar um maior número de testes de usabilidade do nosso protótipo funcional, visto que os nossos testes foram diminutos e todos eles compostos por pessoas do nosso agregado familiar, o que pode ter acrescentado algum tipo de tendência aos nossos resultados. Adicionalmente, podem também ser acrescentadas outras visualizações de dados às áreas já existentes por forma a completar o conteúdo da plataforma, e caso seja útil, acrescentar ainda dados de novas áreas como a área do desporto, emprego e económica.

6. Referências

- [1] D3.js, Data Driven Documents, <https://d3js.org/>
- [2] PorData, Milhares de Estatísticas sobre Municípios, Portugal e Europa, <https://www.pordata.pt/Home>
- [3] PorData, Portugal, <https://www.pordata.pt/Portugal>
- [4] PorData, Municípios, <https://www.pordata.pt/Municipios>