



Visualização de Dados 1º Semestre

Projeto 2 - Power BI

Trabalho realizado por:

Daniela Jorge nº 54989
Mafalda Franco nº 55358
Rita Rodrigues nº 54859

Introdução

Neste relatório será feito um estudo sobre o uso da ferramenta Power BI em análise de dados. Através da utilização desta plataforma, é possível criar dashboards e relatórios interativos que permitem uma visualização mais clara e facilitada da informação, auxiliando na tomada de decisões e na identificação de padrões e tendências.

Serão apresentados três dashboards criados no Power BI, cada um com uma finalidade específica. O primeiro dashboard procura relacionar a taxa de abandono escolar, a taxa de desemprego, a taxa de repetição e fatores culturais ao nível NUTS II. É feita uma análise mais detalhada para as regiões do Alentejo e da Área Metropolitana de Lisboa. O segundo dashboard pretende analisar a possível relação entre as escolas localizadas nos municípios da área metropolitana de Lisboa que recebem ajudas sociais e a taxa de repetição escolar dos estudantes que frequentam o ensino secundário nestas escolas. Finalmente, o terceiro dashboard tem como objetivo analisar o sucesso escolar no município de Oeiras.

Ao longo deste relatório, serão apresentados os resultados obtidos através da análise destes dashboards, bem como reflexões e possíveis conclusões sobre os dados apresentados. Esperamos que este estudo possa fornecer informações úteis para a compreensão da realidade da educação na área metropolitana de Lisboa e no município de Oeiras, bem como avaliar a eficácia da ferramenta Power BI na análise de dados.

Power BI

Power BI é uma ferramenta de análise de dados e visualização desenvolvida pela Microsoft. Permite que os utilizadores importem, limpem e transformem dados de diversas fontes, como base de dados do SQL Server, livros do Excel, entre outros. O Power BI oferece várias técnicas de visualização de dados, incluindo gráficos de barras, linhas, colunas, área, bolhas, dispersão, mapas, etc. Além disso, o Power BI também permite criar relatórios personalizados e dashboards para exibir e analisar os dados de uma forma mais intuitiva e fácil de entender. A ferramenta também oferece integração com outras plataformas da Microsoft, como o Office 365, o Dynamics 365 e o Azure, o que permite aos utilizadores trabalhar com os dados de maneira mais eficiente e integrada. A versão Desktop do Power BI permite que os utilizadores criem relatórios e dashboards avançados e que os publiquem na nuvem ou compartilhem com outros utilizadores, enquanto a versão Online é uma versão mais simples, mas ainda assim poderosa, que permite aos utilizadores criar relatórios básicos e dashboards e compartilhá-los com outros utilizadores na nuvem.

No entanto, o Power BI Desktop também apresenta algumas limitações que devem ser consideradas pelos utilizadores. Uma das principais limitações é a quantidade de dados que o software é capaz de processar. O Power BI Desktop tem um limite de 1 GB de dados por relatório, o que pode ser insuficiente para alguns utilizadores que trabalham com conjuntos de dados muito grandes. Além disso, o Power BI Desktop não oferece suporte a todas as fontes de dados disponíveis no mercado, o que pode dificultar a importação de dados de algumas fontes específicas. Outra limitação é a falta de algumas funcionalidades avançadas de análise de dados, como suporte a linguagem de programação R e integração com ferramentas de machine learning.

Ao utilizar o Power BI Desktop, é importante considerar alguns aspectos para garantir que as visualizações de dados sejam eficazes e precisas. Isto inclui a qualidade e a confiabilidade das fontes de dados, a escolha apropriada de técnicas de visualização de dados para representar os dados de maneira clara e compreensível, e a aderência a padrões de design visual para garantir que as visualizações sejam informativas e fáceis de

ler. Além disso, é importante levar em consideração o público-alvo e o objetivo da visualização, com o objetivo de garantir que os relatórios sejam relevantes e úteis.

No geral, o Power BI Desktop é uma ferramenta poderosa e versátil para análise de dados e visualização de informações. No entanto, é importante considerar as suas limitações e avaliar se é a melhor ferramenta para fazer a análise pretendida.

Dados

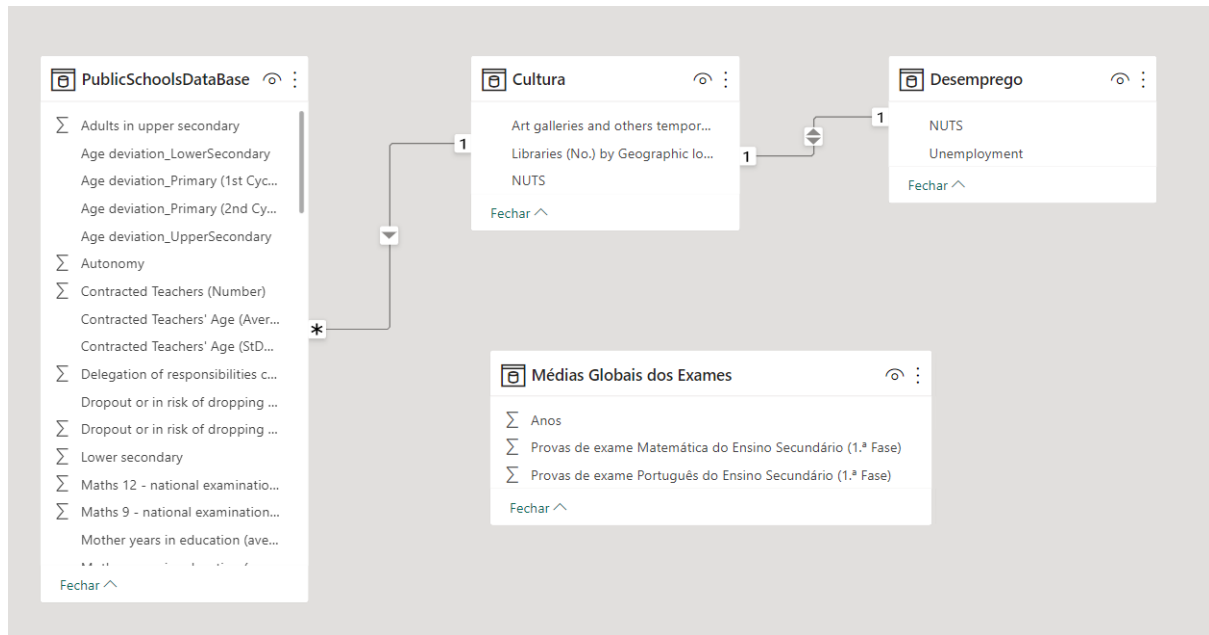
O principal conjunto de dados deste estudo é fornecido pela Direção-Geral de Estatísticas da Educação e Ciência (DGEEC) e diz respeito ao sistema de educação pública português. O conjunto de dados é abrangente e inclui uma ampla gama de variáveis relacionadas com escolas, alunos e professores, entre outros. As variáveis específicas utilizadas no estudo serão descritas com mais detalhes nas seções seguintes deste relatório. O objetivo deste estudo é examinar vários aspectos do sistema de educação pública português e identificar padrões, tendências e insights que possam ser de interesse para educadores e investigadores.

O conjunto de dados de desemprego utilizado neste estudo foi obtido no site do Instituto Nacional de Estatística (INE) e inclui informações sobre as taxas de desemprego em Portugal. Abrange o primeiro trimestre de 2018 e está organizado por regiões NUTS II. A taxa de desemprego é calculada como a percentagem da força de trabalho que está desempregada e ativamente à procura de emprego. É calculada dividindo o número de desempregados pelo tamanho total da força de trabalho (que inclui tanto os empregados quanto os desempregados) e, em seguida, multiplicando o resultado por 100 para apresentá-lo como uma percentagem.

O conjunto de dados de cultura utilizado neste estudo foi também obtido no site do INE e inclui informações sobre o número de galerias de arte e outros espaços de exposição temporária e o número de bibliotecas em Portugal em 2002. O conjunto de dados está organizado por regiões NUTS II. Fornece dados sobre os recursos culturais disponíveis e pode ser usado para explorar padrões e tendências na distribuição desses recursos. O objetivo do uso deste conjunto de dados é examinar a relação entre os recursos culturais e o sistema de educação pública português e compreender como esses recursos podem afetar os resultados educacionais ou outros indicadores de desempenho educacional.

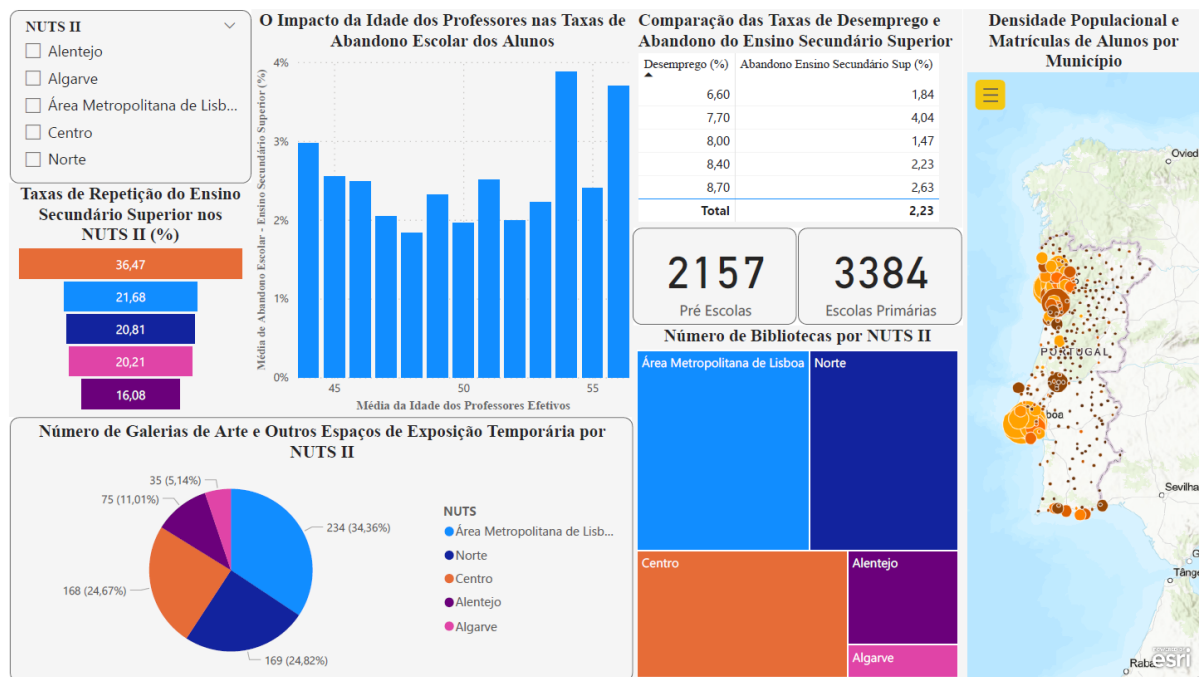
O conjunto de dados sobre as médias globais dos exames nacionais utilizado neste estudo foi obtido no site PORDATA e inclui informações sobre as médias das provas nacionais de 3º ciclo e secundário na 1ª fase desde 2008, apenas são utilizados os dados de 2011 tendo como objetivo relacionar estes valores com as médias obtidas nos mesmos nas diversas escolas portuguesas. Os dados sofreram alterações, uma vez que foram calculados de 0 a 100, o que não coincidia com a escala dos valores do conjunto fornecido inicialmente.

Modelo



Visualizações

Dashboard 1

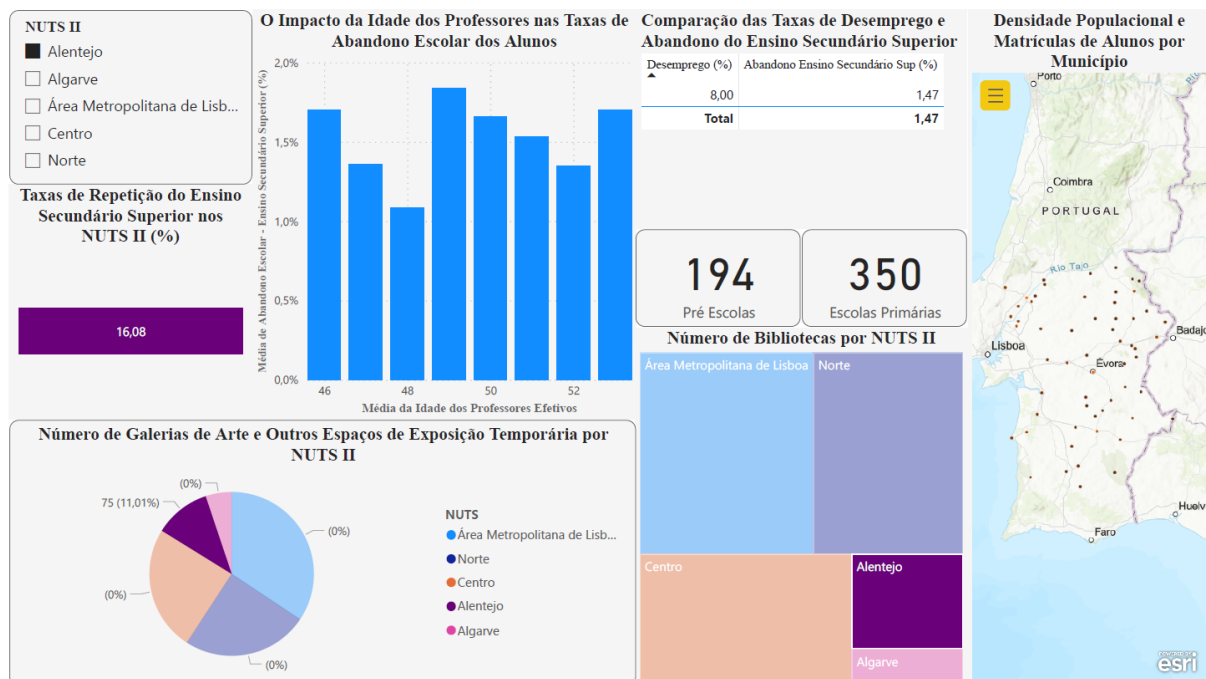


Neste estudo, utilizámos as seguintes visualizações no Power BI para analisar e explorar os dados:

- Segmentação de dados: esta visualização permite dividir os dados em grupos, e comparar os valores entre os grupos.
- Funil: é um tipo de gráfico que permite visualizar a distribuição de um conjunto de dados em diferentes categorias.
- Gráfico de colunas: este gráfico permite comparar os valores de diferentes categorias ao longo de um período de tempo ou em diferentes grupos.

- Tabela: é uma visualização que exibe os dados em forma de tabela, com linhas e colunas. É útil para exibir dados detalhados e para comparar valores em diferentes categorias.
- Mapa ArcGIS: este tipo de mapa permite exibir dados geográficos, como informações sobre regiões, cidades, estados, países, etc. É útil para visualizar padrões e tendências em dados geográficos.
- Cartões: são pequenas visualizações que exibem um único valor ou métrica. São úteis para exibir informações resumidas de maneira concisa e fácil de ler.
- Treemap: é um tipo de gráfico que exibe os dados em forma de árvore, com diferentes níveis de categorias e subcategorias. Cada categoria é representada por um retângulo de tamanho proporcional ao valor da categoria. É útil para visualizar a distribuição de um conjunto de dados em diferentes categorias.
- Gráfico circular: é um tipo de gráfico que exibe os dados em forma de setores de um círculo. Cada setor representa uma categoria ou uma parte do conjunto de dados, e o tamanho de cada setor é proporcional ao valor da categoria. É útil para visualizar a distribuição de um conjunto de dados em diferentes categorias.

Alentejo



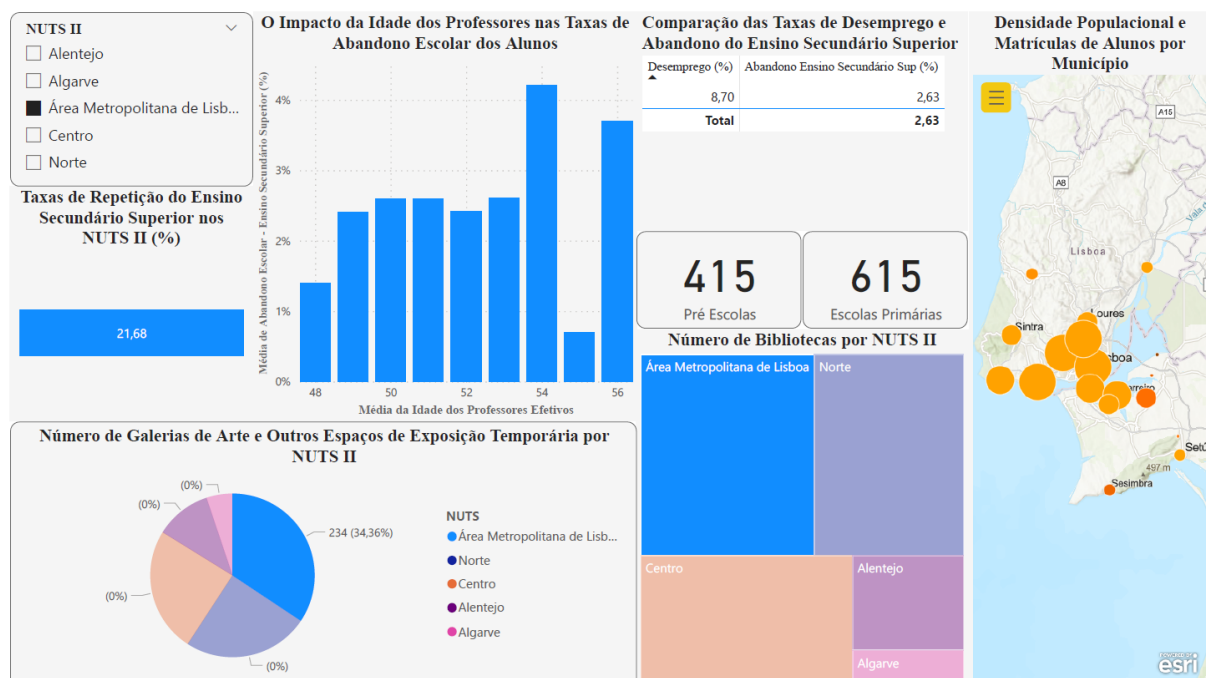
A partir da análise das visualizações fornecidas, podemos observar que a taxa de repetição do ensino secundário superior no Alentejo é de 16,08%, a mais baixa entre as regiões de Portugal. Isso pode ser interpretado como um indicador de bom desempenho dos alunos nesta região no ensino secundário superior. Além disso, o gráfico de colunas aponta para uma idade média dos professores entre 46 e 53 anos, o que pode ser considerado como uma indicação de estabilidade e experiência no mercado de trabalho. Esses fatores podem contribuir para a baixa taxa de abandono escolar do ensino secundário superior no Alentejo, o que sugere uma boa persistência e determinação dos alunos da região em concluir os seus estudos.

No entanto, é importante destacar que a taxa de desemprego no Alentejo é muito elevada, com um valor de 8,00%, em comparação com a baixa taxa de abandono do ensino secundário. Isso pode ser interpretado como uma possível dificuldade dos alunos em

encontrar emprego após concluir os seus estudos, possivelmente devido à falta de oportunidades de emprego na região. Além disso, a análise dos dados culturais (representados pelo gráfico circular e pelo treemap) sugere que o Alentejo é uma das regiões com menor número de galerias de arte e bibliotecas, o que pode ser um fator adicional na escassez de oportunidades de emprego na região. É importante também mencionar que os círculos escuros e pequenos no mapa indicam-nos que o Alentejo apresenta um alto número de matrículas de alunos por município, considerando a densidade populacional da região. Isso pode ser interpretado de duas maneiras: por um lado, pode ser um indicador de uma alta procura por educação na região, com as escolas conseguindo corresponder à procura. Isso é reforçado pelas baixas taxas de repetição e abandono escolar do ensino secundário superior. Por outro lado, pode também sugerir que as escolas estejam sob pressão para corresponder a essa procura, o que pode afetar a qualidade da educação oferecida. Essa possível diminuição da qualidade da educação pode estar relacionada com o fato de alguns indivíduos terem um conjunto limitado de habilidades, o que os torna menos competitivos no mercado de trabalho e, consequentemente, aumenta a taxa de desemprego.

Em resumo, esta análise aponta para um bom desempenho dos alunos no Alentejo no ensino secundário superior, mas também destaca a possibilidade de dificuldades no mercado de trabalho após a conclusão dos estudos. As limitações de recursos culturais e o alto número de matrículas de alunos por município são fatores que podem contribuir para essas dificuldades. No entanto, seria necessário realizar uma análise mais aprofundada para obter uma compreensão mais completa da situação.

Área Metropolitana de Lisboa



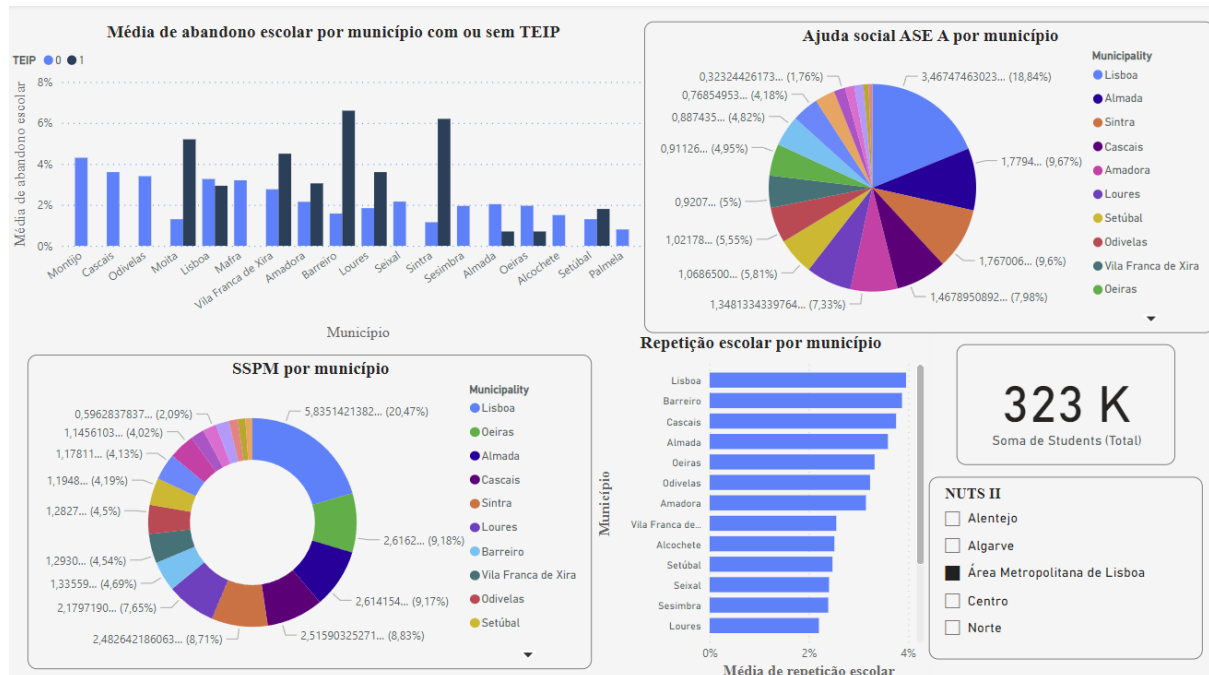
A taxa de abandono escolar na área metropolitana de Lisboa é de 2.63%, o que não está muito longe da média nacional de 2.23%. O gráfico de colunas parece sugerir que existe uma relação entre a média das idades dos professores efetivos e a taxa de abandono escolar dos alunos do ensino secundário superior. As barras que correspondem a médias de idades de 54 e 56 anos apresentam taxas de abandono escolar significativamente mais altas do que as barras correspondentes a outras faixas etárias. No entanto, é importante

considerar que outros fatores podem afetar a taxa de abandono escolar, como o nível socioeconómico dos alunos, o ambiente escolar, o nível de formação dos professores, etc. Apesar da taxa de abandono escolar estar perto da média nacional, a taxa de desemprego na região é de 8.70%, que é a mais elevada de todo o país. Esta situação pode indicar que, apesar de os alunos da área metropolitana de Lisboa terem uma taxa de abandono escolar relativamente baixa, eles podem enfrentar dificuldades no mercado de trabalho após a conclusão dos estudos. É possível que existam fatores específicos da região que contribuam para a alta taxa de desemprego, como a falta de oportunidades de emprego disponíveis ou a existência de uma mão-de-obra altamente qualificada que concorre por um número limitado de vagas.

A área metropolitana de Lisboa é a região que apresenta o maior número de galerias de arte e bibliotecas, de acordo com o gráfico circular e o treemap. Isto pode ser interpretado como uma indicação de um elevado acesso à cultura na região, o que pode ter um impacto positivo na qualidade de vida das pessoas e no desenvolvimento intelectual da população. O mapa aponta para uma grande densidade populacional na área metropolitana de Lisboa, mas um baixo número de matrículas de alunos em escolas públicas devido a apresentar círculos grandes e claros. Isto pode ser explicado pelo fato de existir uma grande população na região, mas talvez um número menor de estudantes. Uma possível explicação para esta diferença pode ser o facto de que a área metropolitana de Lisboa é uma região com um elevado nível de desenvolvimento económico e cultural, o que pode atrair uma população mais velha e com menor interesse em prosseguir os estudos. Outra possibilidade é a existência de muitas escolas privadas na região. De facto, os colégios privados podem atrair uma parcela significativa dos estudantes que residem na área metropolitana de Lisboa, o que pode levar a um menor número de matrículas em escolas públicas. Além disso, é possível que os colégios privados ofereçam uma educação de qualidade superior, o que pode ser atrativo para os pais e responsáveis que desejam garantir uma educação de qualidade para os seus filhos. É importante notar que esta hipótese é apenas uma possibilidade e não está de todo confirmada por estes dados.

Em resumo, esta análise avaliou a situação da área metropolitana de Lisboa em relação ao abandono escolar e ao emprego. A taxa de abandono escolar na região é de 2.63%, o que se encontra próximo da média nacional de 2.23%. A taxa de desemprego na região é de 8.70%, a mais alta do país. A análise destaca ainda a grande densidade populacional na região, mas um baixo número de matrículas de alunos em escolas públicas. Conclui-se que a área metropolitana de Lisboa é uma região com um alto nível de desenvolvimento económico e cultural, mas que enfrenta desafios como uma taxa de desemprego relativamente alta. Seria importante realizar uma análise mais aprofundada para encontrar maneiras de melhorar estes problemas.

Dashboard 2



Com este dashboard pretende-se verificar se existe uma relação entre as escolas dos municípios da área metropolitana de Lisboa que têm ajudas sociais e a taxa de repetição escolar dos alunos que frequentam o ensino secundário nessas escolas.

Para tal, começámos por criar uma segmentação de dados que tem as áreas definidas pelo NUTS II e escolhemos, então, a área Metropolitana de Lisboa para podermos analisar melhor os seus municípios individualmente.

De seguida, criámos os gráficos, figuras e cartões para podermos analisar as variáveis escolhidas em melhor detalhe e mais facilmente:

- **Gráfico de barras vertical.** Neste gráfico, temos a taxa de abandono escolar por município onde é mostrado, através das diferentes barras, se as escolas desses municípios têm o programa TEIP, que é uma iniciativa governamental cujo objectivo é a prevenção e redução do abandono escolar precoce e do absentismo, a redução da indisciplina e a promoção do sucesso educativo de todos os alunos. Na legenda do gráfico, o programa TEIP é representado através de valores binários, as barras com cor azul mais clara têm o valor zero, o que significa que não existe TEIP em certas escolas desse município enquanto que as barras de azul escuro, têm valor 1 e significa que esses municípios têm o programa incorporado nas suas escolas.
- **Gráfico circular.** Neste gráfico, temos representado a taxa de ajuda social ASE A por município.
- **Gráfico em anel.** Neste gráfico, temos a taxa de sucesso da vida escolar dos estudantes por município.
- **Gráfico de barras horizontal.** Neste gráfico, é representado a taxa de repetição escolar por município.
- **Cartão.** Aqui é representado a quantidade de estudantes na área metropolitana de Lisboa em cada município quando seleccionado individualmente.

No gráfico de barras vertical, conseguimos ver que nas escolas de alguns municípios que têm o programa TEIP incorporado a taxa de abandono escolar é maior, isto verifica-se para os municípios do Barreiro, Vila Franca de Xira, Amadora, Moita, Loures, Sintra e Setúbal.

Enquanto que em apenas alguns municípios como Lisboa, Almada e Oeiras as escolas que têm o programa apresentam uma taxa de abandono escolar mais baixa do que as escolas sem o programa.

Uma razão para isto acontecer pode ser o facto de o TEIP se instalar nas escolas em zonas mais desfavorecidas dos municípios e mesmo com a ajuda deste programa a taxa de abandono escolar é maior do que escolas localizadas em zonas melhores mas que não têm este programa.

No gráfico circular, podemos observar facilmente que Lisboa é o município com maior ajuda social ASE A com 18,84%, seguido de Almada com 9,67%, Sintra com 9,6% e Cascais com 7,98%. Com menos ajuda social temos o município de Alcochete e Mafra com menos de 1% de ajuda social.

No gráfico em anel, temos a taxa de sucesso da vida escolar dos alunos dos diferentes municípios e temos em primeiro, Lisboa com uma taxa de 20,47% seguido de Oeiras com 9,18%, Almada com 9,17%, Cascais com 8,83% e Sintra com 8,71%. Com menos de 2% de sucesso, temos os municípios de Alcochete, Mafra e Moita.

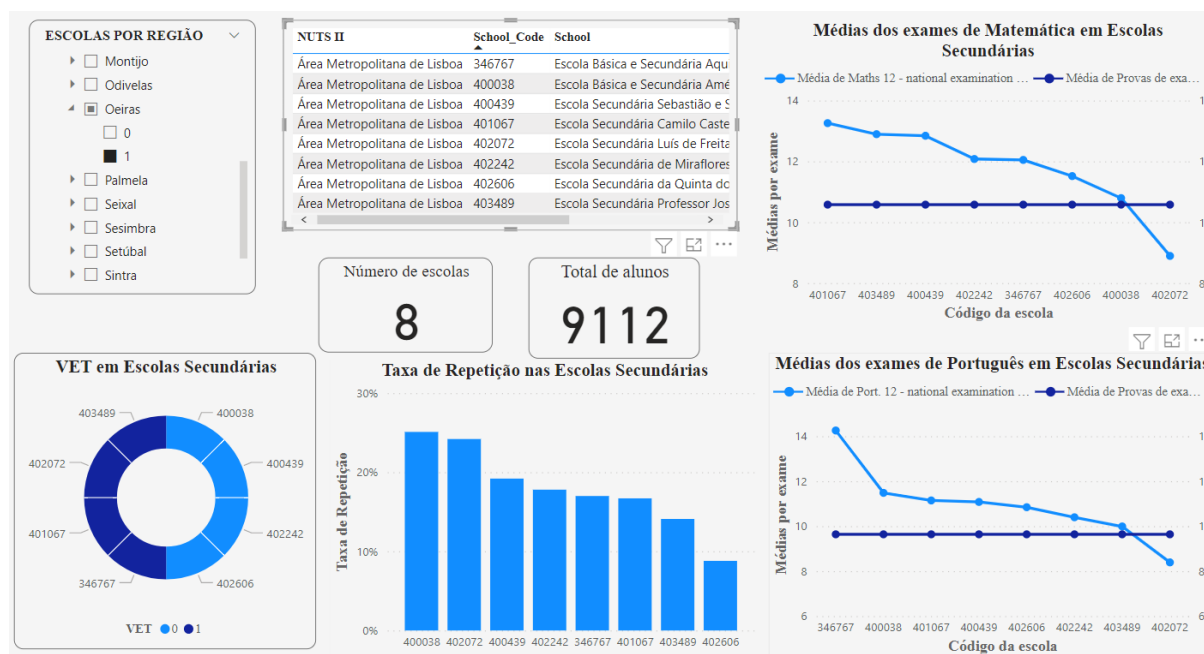
No gráfico de barras horizontal, temos a taxa de repetição escolar por município. Com 3,96% temos o município de Lisboa, seguido do Barreiro, Cascais, Almada e Oeiras. Com menos de 2% temos os municípios de Palmela, Moita e Mafra.

Depois da análise de cada gráfico individualmente, podemos concluir que Lisboa é o município que apesar de ter mais ajudas sociais e programas que ajudam à redução do abandono escolar, tem uma taxa de repetição escolar maior e onde a taxa de abandono escolar é cerca de 3%. Apesar disto, o gráfico em anel contraria um pouco esta análise uma vez que Lisboa é o município com a taxa de sucesso escolar maior.

Uma razão para tal acontecer é o facto de Lisboa ser o município com mais alunos (cerca de 55000 alunos no total), ou seja os apoios sociais podem ser mais altos mas como existem tantos alunos a taxa de repetição também vai ser mais alta comparado com o dos outros municípios.

Por outro lado, temos o município da Moita que tem uma taxa de apoio social ASE e de sucesso escolar mais baixa, o que pode explicar ter a maior taxa de abandono escolar quando analisando o gráfico de barras vertical. A falta de apoios sociais pode ser uma razão pelo qual os estudantes não podem continuar a estudar no ensino secundário.

Dashboard 3



Neste dashboard, possuímos como objetivo analisar o sucesso escolar no município de Oeiras.

Começámos por criar uma segmentação de dados, de forma a podermos selecionar a região e município que desejamos analisar e também uma opção para escolher escolas secundárias ou escolas com outros níveis escolares (representados por valores binários: 0 - Escolas não secundárias e 1 - Escolas Secundárias).

Seguidamente, construímos uma tabela onde as escolas se encontram registadas por região, código identificador e nome, para obtermos uma visão mais simplificada das representações gráficas, dado que em todas elas as escolas estão identificadas, apenas, pelo seu código identificador. A tabela retorna estes dados consoante as escolhas da segmentação de dados.

Criámos dois cartões, de maneira a obtermos informação sobre o total de alunos e escolas num determinado município e caso desejemos saber quantos alunos cada escola possui, basta selecionarmos a escola na tabela e será calculado esse valor.

As análises serão feitas através das representações gráficas:

- **Gráfico em Anel**, onde verificamos se uma escola possui VET, isto é, outros cursos para além dos científico-humanísticos, se sim representa-se a azul escuro (Valor 1), se não a azul claro (Valor 0);
- **Gráfico de Barras**, para verificar a taxa de repetição em cada escola, tendo em conta que em 2011 a taxa de retenção média global foi de 20,8%;
- **Gráficos de Linhas**, para verificar a relação entre notas médias dos exames nacionais de matemática e português no secundário com as notas médias globais nacionais dos mesmos (10,58 a matemática e 9,54 a português).

Nos cartões é possível observar que Oeiras possui um total de 8 escolas e 9112 alunos.

No gráfico de anel, observamos quais as escolas que possuem VET, podendo concluir que metade das escolas em Oeiras possui e a outra metade não.

No gráfico de barras, observamos as taxas de repetição em cada uma das escolas de Oeiras, sendo a Escola Básica e Secundária Amélia Rey Colaço (Linda-a-Velha) a taxa mais alta (25,10%) e a Escola Secundária da Quinta do Marquês a taxa mais baixa (8,80%). Decidindo analisar detalhadamente cada uma destas duas escolas e podemos verificar, através dos gráficos de linhas, que apesar de a Escola Básica e Secundária Amélia Rey Colaço possuir a taxa mais alta de repetição, os seus valores médios dessas mesmas avaliações (11 a matemática e 11 a português) encontram-se acima dos valores médios globais. Para além disto, a taxa encontra-se acima do índice médio global de taxa, significa que será necessário fazer uma intervenção específica a esta escola para perceber o porquê da taxa muito elevada, um motivo provável será a exigência por parte dos docentes. O mesmo acontece com a Escola Secundária Luís de Freitas Branco (Paço de Arcos) que possui uma taxa ligeiramente inferior à mais alta (24,20%) e que tem ambas as médias dos exames nacionais negativas (8 a matemática e 9 a português) e abaixo das médias globais nacionais, o que poderá significar um otimismo por parte dos docentes nos alunos.

Na Escola Secundária da Quinta do Marquês temos a taxa de repetição mais baixa, sendo ambos os valores médios dos exames nacionais na mesma (12 a matemática e 11 a português) acima das médias globais nacionais, o que poderá ser um indicador do bom funcionamento dessa escola.

Para além destas observações, podemos concluir que a Escola Básica e Secundária Amélia Rey Colaço não possui VET que poderá ser um dos motivos para a elevada taxa de repetição, uma vez que os alunos poderão encontrar-se privados de escolhas no que se refere aos cursos secundários disponíveis na mesma.

Conclusão

Através das informações apresentadas nos dashboards e das conclusões que foram retiradas dos mesmos, podemos concluir que, em algumas regiões do país, será necessário fazer uma intervenção na área da educação de forma a que os alunos possam tirar proveito dos seus estudos no presente e, consequentemente, no futuro.