



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO EM

**MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E
EMPRESARIAL**

TRABALHO FINAL DE MESTRADO

**PROJETO ELABORADO PARA A OBTENÇÃO DO GRAU DE
MESTRE**

***ANÁLISE DE DADOS EM BUSINESS INTELLIGENCE DOS
ÚLTIMOS ANOS LETIVOS DO ENSINO SUPERIOR***

MARIANA FERREIRA ANACLETO

ORIENTAÇÃO:

PROFESSOR DOUTOR CARLOS J. COSTA

OUTUBRO – 2022

AGRADECIMENTOS

Nem sempre conseguimos expor em diálogo o que sentimos pelas pessoas. Todavia, as grandes conquistas pessoais sabem melhor junto de pessoas que nos são queridas. É por esse motivo que dedico a primeira página do meu trabalho a todas elas, por me terem apoiado e incentivado a concluir este grande desafio.

Assim sendo, e como não podia deixar de ser, quero agradecer aos meus pais. Johnny e Aninhas, sem vocês nada disto teria sido possível. Obrigada pelos sacrifícios que fizeram para me proporcionarem a melhor educação, por me terem deixado seguir os meus sonhos e por toda a paciência que tiveram para mim em épocas de exames. Sem os milhares de ficheiros que o pai imprimiu e os lanches motivacionais da mãe, finalizar a licenciatura e estar prestes a alcançar o grau de mestre teriam sido muito mais árduos. Agradeço aos meus avós, em especial, ao avô Anacleto e à avó Estela por serem um pilar tão importante na minha vida.

Gostaria ainda de agradecer ao Professor Doutor Carlos J. Costa por ter aceitado entrar nesta aventura comigo e pela sua ajuda e disponibilidade semanalmente. Foi incansável!

E como tudo faz sentido mais tarde ou mais cedo, tenho de agradecer às pessoas incríveis que o ISEG me deu: em gestão, às minhas amigas que sempre acreditaram que ia conseguir e me apoiaram neste percurso, por vezes meio atribulado, e em MQDEE, por ter colocado no meu caminho a Mariana, que foi a minha parceira ao longo desta aventura, a Mada e a Carol por preservarem os níveis de saúde mental ao longo destes meses.

Não quero deixar de mencionar o grande fundador da *Team* Paulo pela sua paciência, o Migas por me continuar a aturar desde o 5º ano, as gémeas que nos últimos dois anos ganharam um lugar muito especial na minha vida, o Chico por ser o amigo que quero levar para a vida e por último, mas não menos importante, o Di pela amizade especial que construímos e bonita caminhada que aceitou dar de mãos dadas comigo.

Acredito que a vida nos contempla, sempre na altura certa, com as pessoas certas. E foi assim que a Helena apareceu na minha vida. A ti, amiga do coração e mana de vida, que depois de tantos desabafos e desesperos ao longo deste processo e dos últimos anos, tinhas sempre as palavras certas no momento certo, sabes que estava destinado sermos amigas.

Apesar dos diferentes papéis de cada um deles na minha vida, sem eles a conclusão desta grande etapa da minha vida não teria sido possível. Do fundo do coração, obrigada!

RESUMO

O século XXI é assinalado pelas constantes mudanças na sociedade nas mais diversas áreas da atividade humana. Uma das principais transformações sentidas está associada ao crescente volume de dados disponibilizados.

Nesta nova sociedade, denominada de Sociedade de Informação, a informação é vista como uma necessidade, mas também como um recurso para alcançar conhecimento.

Porém, o tratamento de grandes quantidades de dados nem sempre é decifrável. Torna-se assim necessário recorrer a mecanismos que permitam processar estes grandes conjuntos de dados de forma que sejam compreendidos e, por conseguinte, forneçam informação útil, nomeadamente para a implementação de estratégias que aumentem a eficácia das organizações no seu meio envolvente.

Circulando cada vez mais informação e de acesso cada vez mais fácil, de dia para dia as organizações nos mais diversos setores atuam num ambiente mais concorrencial e inovador – onde o setor da educação não é exceção. Torna-se assim essencial que as Instituições de Ensino Superior (IES) se munam de novas ferramentas e tecnologias que lhes permitam enfrentar os correntes desafios propostos com a nova realidade em que estão inseridas.

É neste sentido que o presente trabalho tem como principal objetivo, através de uma ferramenta de *Business Intelligence* (BI), caracterizar o Ensino Superior em Portugal nos últimos anos letivos. Assim, recorreu-se ao *Microsoft Power BI* para desenvolver *dashboards* que permitem alcançar os objetivos propostos e contribuem de forma positiva para a tomada de decisões das IES.

Desta forma, a metodologia CRISP-DM seguida permite desenvolver um relatório onde a informação disponibilizada se encontra perceptível em painéis visualmente apelativos, de maneira a contribuir para o melhoramento das decisões dos diferentes Estabelecimentos de Ensino e, por conseguinte, beneficiar de vantagens competitivas perante os restantes e principais rivais.

Palavras-chave: Educação, Instituições de Ensino Superior (IES), CRISP-DM, *Business Intelligence* (BI), *Microsoft Power BI*.

ABSTRACT

The 21st century is marked by the constant changes in society in the most diverse areas of human activity. One of the major transformations is associated with the increasing volume of data available.

In this new society, called the Information Society, information is seen as a necessity and a resource to achieve knowledge.

However, the processing of large amounts of data is not always decipherable. It becomes necessary to use mechanisms that help to process these large data sets in such a way that they are understood and therefore provide useful information for implementing strategies that increase organizations' effectiveness in their environment.

When faced with increased information and increasingly easier access, day-by-day organizations in the most diverse sectors operate in a more competitive and innovative 'environment' – the education sector is no exception. Therefore, it becomes essential that Higher Education Institutions (HEIs) have new tools and technologies that allow them to face the current challenges that come with the new reality in which they are inserted.

In this sense, the present work has as its primary objective a Business Intelligence (BI) tool, characterizing higher education in Portugal in recent school years. Thus, Microsoft Power BI was used to develop dashboards that would achieve the proposed objectives and contribute positively to the decision-making of the HEIs.

In this way, the CRISP-DM methodology followed allows the development of a report, where the information provided is noticeable in visually appealing panels, to contribute to the improvement of the decision of the different Educational Establishments and therefore benefit from competitive advantages over the remaining and main rivals.

Keywords: Education, Higher Education Institutions (HEIs), CRISP-DM, Business Intelligence (BI), Microsoft Power BI.

ÍNDICE

| | |
|---|------------|
| AGRADECIMENTOS..... | II |
| RESUMO | III |
| ABSTRACT | IV |
| ÍNDICE DE FIGURAS..... | VII |
| ÍNDICE DE TABELAS | VII |
| ÍNDICE DAS OPERAÇÕES..... | VII |
| CAPÍTULO 1: INTRODUÇÃO..... | 1 |
| 1.1 Enquadramento e Relevância do Tema | 1 |
| 1.2 Objetivos do Trabalho..... | 1 |
| 1.3 Abordagem Metodológica..... | 3 |
| 1.4 Estrutura do Trabalho..... | 3 |
| CAPÍTULO 2: REVISÃO DA LITERATURA..... | 4 |
| 2.1 A nova e atual Sociedade de Informação | 4 |
| 2.2 <i>Business Intelligence</i> : O Motor de Diferenciação | 5 |
| 2.2.1 Relatórios de Negócio do tipo <i>Dashboards</i> | 6 |
| 2.2.2 Tipos de <i>Dashboards</i> | 7 |
| 2.2.3 Características / Requisitos de um <i>Dashboards</i> | 8 |
| 2.2.4 Arquitetura do Sistema de BI..... | 9 |
| 2.2.5 <i>Data Warehouse</i> (DW) e <i>Data Marts</i> (DM) | 11 |
| 2.2.6 Modelação dos Dados Multidimensionais | 12 |
| 2.2.7 Ferramentas para a Construção de Sistemas de BI..... | 13 |
| 2.2.8 Desafios Enfrentados pelo BI..... | 13 |
| 2.3 A Educação na Sociedade Atual | 14 |
| 2.3.1 O Sistema de Ensino Superior..... | 15 |
| 2.4 O <i>Business Intelligence</i> Associado À Educação | 17 |
| CAPÍTULO 3: METODOLOGIA | 19 |
| CAPÍTULO 4: APLICAÇÃO DA METODOLOGIA – CRISP-DM | 21 |
| 4.1 Fase 1 - <i>Business Understanding</i> | 21 |
| 4.2 Fase 2 - <i>Data Understanding</i> | 22 |
| 4.3 Fase 3 - <i>Data Preparation</i> | 26 |
| 4.4 Fase 4 - <i>Modeling</i> | 30 |
| 4.5 Fase 5 - <i>Evaluation</i> | 34 |
| 4.6 Fase 6 - <i>Deployment</i> | 35 |

| | |
|---|-----------|
| CAPÍTULO 5: RESULTADOS DO PROJETO..... | 36 |
| 5.1 Menu Principal | 36 |
| • <i>Dashboard</i> 1: Página Inicial do Projeto | 36 |
| 5.2 Análise Geral..... | 36 |
| • <i>Dashboard</i> 2: Estabelecimentos – Unidades Orgânicas – Cursos..... | 36 |
| • <i>Dashboard</i> 3: Estabelecimentos – Unidades Orgânicas..... | 38 |
| 5.3 Análise das Entradas | 39 |
| • <i>Dashboard</i> 4: Inscritos | 39 |
| • <i>Dashboard</i> 5: Matriculados..... | 41 |
| 5.3 Análise das Saídas | 42 |
| • <i>Dashboard</i> 6: Diplomados | 42 |
| 5.4 Análise Comparativa..... | 43 |
| • <i>Dashboard</i> 7: Área Curso | 43 |
| CAPÍTULO 6: CONCLUSÃO, LIMITAÇÕES E TRABALHO FUTURO..... | 45 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 47 |
| ANEXOS | 54 |

ÍNDICE DE FIGURAS

| | |
|--|----|
| Figura 1 - Fases do ciclo metodológico CRISP-DM (Fonte: Chapman et al., 2000)..... | 19 |
| Figura 2 - Selecionar e Importar os Dados a Transformar no Microsoft Power BI Desktop (Fonte: Autor)..... | 27 |
| Figura 3 - Criação das relações do Modelo Relacional (Fonte: Autor)..... | 31 |
| Figura 4 - Modelo Relacional (Fonte: Autor) | 32 |
| Figura 5 - Página Inicial do Projeto (Fonte: Autor) | 36 |
| Figura 6 - Análise Geral dos Estabelecimentos – Unidades Orgânicas – Cursos (Fonte: Autor)..... | 37 |
| Figura 7 - Análise Geral dos Estabelecimentos – Unidades Orgânicas (Fonte: Autor) | 38 |
| Figura 8 - Análise das Entradas - Inscritos (Fonte: Autor) | 39 |
| Figura 9 - Análise das Entradas - Matriculados (Fonte: Autor) | 41 |
| Figura 10 - Análise das Saídas - Diplomados (Fonte: Autor) | 42 |
| Figura 11 - Análise Comparativa por Área de Curso (Fonte: Autor)..... | 43 |

ÍNDICE DE TABELAS

| | |
|---|----|
| Tabela I - Nome, Tipo e Descrição das Variáveis Qualitativas retiradas dos diferentes Indicadores | 23 |
| Tabela II - Nome, Tipo e Descrição da Variável Comum a diferentes Indicadores..... | 24 |
| Tabela III - Nome, Tipo e Descrição das Variáveis Quantitativas retiradas dos diferentes Indicadores | 24 |
| Tabela IV - Tabelas de Dimensão e Subdimensão do Modelo Relacional (Chave Primária e Outros Atributos) | 32 |
| Tabela V - Análise SWOT do projeto desenvolvido..... | 35 |

ÍNDICE DAS OPERAÇÕES

| | |
|--|----|
| Operação 1 – Transformação da nova coluna “NumInscritos_PrimeiraOpção” | 29 |
| Operação 2 - Transformação da nova coluna "NumDiplomados_Nota10" | 29 |
| Operação 3 - Medida criada para o cálculo da Nota Média de Ingresso | 30 |
| Operação 4 - Medida criada para o cálculo da Nota Média dos Diplomados | 30 |
| Operação 5 – Transformação da nova coluna da Tabela TD3 - Faculdade..... | 31 |

GLOSSÁRIO DE ABREVIATURAS E SIGLAS

BI – *Business Intelligence*

BIS – *Business Intelligence Systems*

CNA – Concurso Nacional de Acesso

CRISP-DM – *Cross-Industry Standard Process for Data Mining*

DAX - *Data Analysis Expressions*

DGES – Direção-Geral de Ensino Superior

DM – *Data Marts*

DW – *Data Warehouse*

ECTS – *European Credit Transfer and Accumulation System*

EEES – Espaço Europeu de Ensino Superior

ETL – *Extraction, Transformation and Loading*

IES – Instituições de Ensino Superior

IoT – *Internet of Things*

KPI's – *Key Performance Indicator*

MQDEE – Métodos Quantitativos para a Decisão Económica e Empresarial

OA – *Open Access*

OLAP – *Online Analytical Processing.*

ONU – Organização das Nações Unidas

SI – Sistemas de Informação

SQL – *Structured Query Language*

SWOT - *Strengths, Weaknesses, Opportunities, Threats*

TD – Tabelas de Dimensão

TF – Tabela de Factos

TFM – Trabalho Final de Mestrado

CAPÍTULO 1: INTRODUÇÃO

No âmbito do mestrado em Métodos Quantitativos para a Decisão Económica e Empresarial (MQDEE), foi realizado o presente Trabalho Final de Mestrado (TFM) com o intuito de analisar um conjunto de dados relativo às Instituições de Ensino Superior (IES) em Portugal nos últimos oito anos letivos.

1.1 Enquadramento e Relevância do Tema

Com a evolução das novas tecnologias, o mundo moderno ingressou numa nova fase de desenvolvimento tecnológico. Surgem, aliadas a esta fase, diversas transformações na sociedade, como é o caso do crescimento exponencial de dados gerados no quotidiano. Tal leva a que seja cada vez mais relevante recorrer a ferramentas de análise de dados, que forneçam aos utilizadores informação competitiva (Khizbullin et al., 2017; Obskov et al., 2015).

Com a transfiguração da nova sociedade, surge uma maior competitividade no mercado global, inclusive no setor da educação. Dessa forma, torna-se crucial adotar estratégias eficientes face aos seus principais concorrentes e oferecer sistemas de ensino de qualidade (Tsai et al., 2020).

Para tal, e com a maior acessibilidade aos dados nos últimos anos, as técnicas de mineração de dados, também denominadas de *Data Mining*, têm assumido um papel cada vez mais meritório na sociedade (Mandinach & Jimerson, 2016). O interesse em adotar mecanismos que permitam retirar conhecimento a partir de uma grande quantidade de dados têm-se intensificado em diferentes áreas, onde a área educacional não foi exceção (Aldowah et al., 2019).

Assim, e com a consciencialização da importância de se ter dados digitalizados dentro dos sistemas educacionais (Selwyn, 2015), recorrer a ferramentas que auxiliem a tomada de decisão no setor da educação tornou-se numa mais valia, na medida que permite às instituições de ensino destacarem-se das restantes.

1.2 Objetivos do Trabalho

Como forma de responder às novas exigências da sociedade, nomeadamente de traduzir a crescente quantidade de dados disponibilizados em informação útil, o projeto que se perspetiva desenvolver procura trabalhar um conjunto de dados de forma a possibilitar

compreender e dar resposta à seguinte questão: Qual o funcionamento e comportamento do sistema de Ensino Superior em Portugal ao longo dos anos?

Para tal, e através do processamento dos dados e do recurso a técnicas de *Data Science*, pretende-se entender de forma mais simples e intuitiva o panorama nacional deste grau de ensino (Aparicio et al., 2019).

Posto isto, os objetivos gerais propostos a alcançar com a realização deste projeto são:

- Propor uma ferramenta que auxilie os decisores na tomada de decisões no âmbito do Ensino Superior;
- Caracterizar o Ensino Superior em Portugal nos últimos anos letivos.

Este objetivo geral pode ainda ser decomposto nos seguintes objetivos específicos:

- Caracterizar o Ensino Superior em Portugal nos últimos anos letivos quanto à sua Natureza, Tipo de Ensino, Grau e Cursos ministrados nos Estabelecimentos de Ensino;
- Quanto à distribuição geográfica dos Estabelecimentos de Ensino Superior e respetivas Unidades Orgânicas;
- Quanto aos alunos inscritos – análises alusivas à idade, género, nacionalidade, nota de entrada e formas de inscrição;
- Quanto aos alunos diplomados – refletir o número total de diplomados, as suas notas médias e a evolução das suas classificações finais.
- Quanto à evolução existente do Número de Alunos (inscritos e diplomados) VS Notas (entrada e saída) nas diferentes áreas lecionadas.

Com a realização deste trabalho, e através de uma abordagem inovadora e interativa, foi possível, numa perspetiva científica, contribuir com informação relevante para estudos relacionados com a temática. O mesmo tem ainda contribuições práticas na medida em que permite auxiliar os responsáveis dos diferentes Estabelecimentos de Ensino Superior, e respetivas Unidades Orgânicas, na sua tomada de decisões. Este projeto ao permitir que diferentes Instituições de Ensino Superior (IES) tenham conhecimento do ambiente educacional em que se inserem, possibilita tomar decisões estratégicas mais eficientes de maneira a tornar os seus estabelecimentos mais competitivos e, assim, melhorar a posição no *ranking*.

1.3 Abordagem Metodológica

Neste sentido, o projeto proposto visa, com base na abordagem metodológica denominada por CRISP-DM, desenvolver uma solução sustentada por ferramentas de *Business Intelligence* (BI). Complementarmente, para a compreensão da realidade do Ensino Superior e dos conceitos indispensáveis deste tipo de ferramentas, realizou-se de forma preliminar uma fundamentação teórica.

Esta solução de BI irá permitir analisar o conjunto de dados referentes aos últimos oito anos letivos do Ensino Superior em Portugal e terá como objetivo final transformar os mesmos em informação inteligível, apelativa e relevante. Informação essa que será apresentada em relatórios estruturados e interativos, onde a visualização dos dados (*data visualization*) será de fácil interpretação.

1.4 Estrutura do Trabalho

De forma a alcançar o pretendido, o presente TFM será composto por seis capítulos.

No presente capítulo será feita uma breve introdução ao tema estudado. Será enquadrado o tema, estabelecidos os objetivos a alcançar com a realização do projeto e apresentada a estrutura do trabalho.

No subsequente capítulo, através da revisão da literatura, será realizada a fundamentação teórica do tema a abordar. Neste segundo capítulo serão apresentados conceitos e definições, na íntegra e de forma imparcial, relacionados com os alicerces base do trabalho: *Business Intelligence* (BI) e Educação.

No terceiro e quarto capítulo será apresentada a abordagem metodológica a seguir e a sua aplicabilidade no tema escolhido, respetivamente. Depois de exibida a aplicação prática e detalhada de cada fase do processo, no capítulo 5 será apresentada a solução em BI ao tema em estudo, em formato de *dashboards*. Nesta parte do trabalho, será feita ainda uma análise individualizada e pormenorizada dos resultados exibidos.

Posto isto, o sexto e último capítulo do trabalho enuncia as conclusões gerais retiradas do projeto, bem como algumas das limitações que surgiram ao longo da realização do trabalho. Tendo em vista uma perspetiva futura, serão ainda feitas algumas sugestões de novas abordagens ao estudo como forma de o potenciar.

No final do trabalho, encontram-se mencionadas todas as referências bibliográficas utilizadas e foram disponibilizados alguns anexos.

CAPÍTULO 2: REVISÃO DA LITERATURA

Neste capítulo, será feita a fundamentação teórica de alguns conceitos base para o desenvolvimento do presente trabalho, tendo como suporte a apreciação de artigos de carácter científico e técnico-científico, livros e revistas referentes às temáticas abordadas, bem como outros elementos relevantes disponibilizados em formato digital.

2.1 A nova e atual Sociedade de Informação

A nova e atual sociedade depara-se com diversas e profundas mudanças ocorridas simultaneamente nas mais diversas áreas como socioculturais, económicas e tecnológicas (Castells, 2010). Uma das modificações sentidas nesta nova realidade social relaciona-se com a crescente quantidade de bens intangíveis, isto é, de dados disponíveis na sociedade (Vercellis, 2009).

Porém, é necessário transformar este conjunto de dados em informação de qualidade, de maneira a que seja possível, posteriormente, produzir conhecimento relevante para a tomada de decisões (Caldeira & Lopes, 2018). É neste sentido que Boisot e Canals (2004) afirmam que a matéria-prima da informação são os dados e, conseqüentemente, a informação constitui a matéria-prima necessária para alcançar o conhecimento.

Para que isso seja possível, é necessário recorrer a um conjunto de modelos e metodologias de análise que tornem possível transformar a informação num recurso mais estruturado e organizado – o conhecimento (Coutinho & Lisboa, 2011; Lor & Britz, 2007; Vercellis, 2009). Contudo, o acesso a este conjunto de técnicas apenas foi possível com os avanços tecnológicos que permitiram alcançar conhecimento através da informação disponibilizada em tempo real e de forma não estática (Hamidi et al., 2011).

Assim, estas ferramentas ao permitirem transformar informação em conhecimento - recurso mais valioso das organizações do século XXI, segundo Matukhin e Zhitkova (2015) - contribuem significativamente para que as organizações detenham vantagens competitivas bastante consideráveis (Porter & Millar, 1985).

Esta nova sociedade, submersa de informação e de avanços tecnológicos, foi denominada por diversos autores como **Sociedade de Informação** ou Sociedade em Rede segundo Manuel Castells (Castells, 2010; Nataliya, 2015; Voogt & Knezek, 2008).

No entanto, de maneira a acompanhar estas alterações, torna-se crucial considerar o investimento na formação, caso se pretenda contar com indivíduos flexíveis e adaptáveis às diversas mudanças sentidas na sociedade atual (Atak & Erturgut, 2010).

2.2 Business Intelligence: O Motor de Diferenciação

Ter acesso a ferramentas eficazes e eficientes de Sistemas de Informação (SI) – sistemas que se caracterizam pela qualidade e quantidade da sua informação – pode consubstanciar-se na fórmula de alcançar o sucesso numa sociedade em constante mutação (Jourdan et al., 2008). Exemplo destes SI é o caso dos **Sistemas de Business Intelligence (BIS - Business Intelligence Systems)** (Gaardboe & Svarre, 2018).

O conceito de **Business Intelligence (BI)** surge pela primeira vez em 1958 quando Hans Peter Luhn define o termo como “*a capacidade de apreender as inter-relações dos fatos apresentados*” (Luhn, 1958, p. 314) de forma a direcionar a ação para o alcance de um determinado objetivo. Mais tarde, em 1989, Howard Dresner, analista do Grupo Gartner, torna-se patenteador do termo BI, ao defini-lo como um conjunto de conceitos e metodologias com o fim de melhorar a tomada de decisões através da aplicação de sistemas de suporte (Nofal & Yusof, 2013). Desde então, este conceito tem sido alvo de constantes evoluções e redefinido por vários autores ao longo do tempo (Aws et al., 2021).

Para Niu et al. (2021), BI (também conhecido por Inteligência de Negócios) é a “*capacidade de uma empresa alavancar os dados disponíveis de forma significativa*” (p.2). Considera-se que o processo de BI transforma dados brutos em informação útil através de um conjunto de técnicas e conceitos que, implementados através de ferramentas em *softwares* específicos, conseguem fornecer elementos informativos relevantes, de forma compreensível e transparente, para processos de tomada de decisões complexas (Caldeira & Lopes, 2018; Duan & Xu, 2012).

Contudo, se por um lado, Sharda et al. (2015) definem o termo BI utilizando a expressão *umbrela*, em português «guarda-chuva», por consideram que é uma combinação de arquiteturas, ferramentas, base de dados, aplicações e metodologias. Por outro lado, e na perspetiva de Williams e Williams (2008), BI não é apenas uma tecnologia ou uma metodologia, BI é uma nova e poderosa abordagem que, quando bem utilizada, consegue fornecer, de forma significativa, vantagens competitivas perante os demais concorrentes.

Esta ferramenta, ao permitir o acesso a um conjunto de “*informações precisas, atuais e relevantes*” (Khan & Quadri, 2012, p. 65), faz com que estes sistemas de BI se tornem numa mais-valia para as empresas, dado que possibilitam às mesmas ter à sua disposição informações atualizadas, na maioria das vezes, em tempo útil, bem como entender o posicionamento no mercado em que se inserem, fazer comparações com os seus principais concorrentes e ter a perceção de tendências e possíveis mudanças de comportamentos (Teixeira et al., 2014).

Estas diferentes perspetivas fizeram com que se verificasse um crescimento acentuado na pesquisa de processos de BI nas últimas três décadas (Talaoui & Kohtamäki, 2020).

2.2.1 Relatórios de Negócio do tipo *Dashboards*

Tomar decisões precisas e adequadas à altura, com base em informação organizada e pertinente, é umas das maiores preocupações dos decisores. Para tal, recorrer a relatórios de negócio que apresentem a informação referente a um período ou assunto específico, através do auxílio de textos, tabelas e/ou gráficos, torna-se essencial para a tomada de decisões (Eckerson, 2010; Sharda et al., 2015).

É através da análise destes relatórios de negócio que se torna possível identificar relações e padrões de comportamento, bem como verificar possíveis ineficiências na organização (Watson, 2017). Estes indicadores-chave de desempenho, que se encontram fora do intervalo esperado, denominam-se por **KPI's - Key Performance Indicator** (Eckerson, 2010).

Parmenter (2015) define KPI's como indicadores que indicam o que deve ser feito para aumentar o desempenho das atividades ou alcançar um determinado objetivo, numa perspetiva futura. De acordo com Kerzner (2017), estes devem seguir determinadas características **SMART**:

- **Specific** - os *KPI's* devem ser específicos, ou seja, claros e focados nas metas de desempenho;
- **Mensurable** - devem ser mensuráveis/quantitativos, para que se consiga medir e quantificar os objetivos alcançados;
- **Attainable** – devem ser metas alcançáveis;
- **Relevant** - ter em consideração os objetivos e estratégias da empresa é um fator relevante na definição destes indicadores;

- **Time-Based** - delimitar um período em que estes indicadores podem ser verificados.

Os KPI's podem assim ser representados em relatórios de negócios do tipo de **Dashboards** que, por sua vez, segundo a definição de Few e Edge (2007) à *Revista Intelligent Enterprise*, podem ser vistos como uma “*exibição visual das informações mais importantes e necessárias para atingir um ou mais objetivos*” (p.1).

Como se vê, e apesar das várias definições desenvolvidas, todas elas consideram que este tipo de relatórios, após análise de dados complexos, permitem apresentar a informação relevante de uma forma simples num painel que possibilita medir rapidamente o desempenho de uma organização e identificar as áreas que necessitam de atuação imediata de forma a alcançar os objetivos propostos (Bugwandeem & Ungerer, 2019; Kerzner, 2017; Skorka, 2017; Yigitbasioglu & Velcu, 2012). Segundo o estudo realizado por Thomson, a visualização de dados neste tipo de relatórios rápidos e confiáveis tem sido cada vez mais recorrentes (Niu et al., 2021).

2.2.2 Tipos de Dashboards

Porém, de forma que os *dashboards* consigam fornecer a informação certa é necessário que a construção destes tenha em conta que tipo de informação se pretende alcançar. Desta forma, ter em consideração os tipos de *dashboards* é um passo crucial para que se consiga obter os resultados desejados com os mesmos (Sanz, 2018).

Segundo Eckerson (2010) e Few (2006) os *dashboards* podem ser divididos em 3 tipos: **Dashboards Estratégicos**, **Dashboards Analíticos** e **Dashboards Operacionais**.

Os **Dashboards Estratégicos** são frequentemente utilizados por executivos, têm como principal objetivo fornecer informação relativa ao desempenho e progresso organizacional. Baseiam-se em informação atualizada periodicamente, o que permite aos gestores terem uma visão a longo prazo do negócio;

Os **Dashboards Analíticos** dedicam-se à análise de dados e por esse motivo requerem uma quantidade de informação maior que o anterior. Uma particularidade deste tipo de *dashboard* reside no facto de permitir aos gestores decompor a informação (*drill down*) de modo a explorar os dados e identificar quais os motivos que levaram a que certos indicadores de desempenho apresentem determinados resultados. Permitem

identificar problemas e conseqüentemente encontrar soluções de forma a garantir o cumprimento dos objetivos estipulados a curto e longo prazo;

Os **Dashboards Operacionais** têm como principal objetivo supervisionar a atividade quase em tempo real de forma a reduzir a possibilidade de eventuais problemas e prejuízos que possam surgir. Baseiam-se em informações específicas, uma vez que pretendem fornecer informação detalhada de um determinado setor/departamento aos respetivos colaboradores.

2.2.3 Características / Requisitos de um Dashboards

Para além da escolha adequada do tipo de *dashboard* existem outros fatores a ter em consideração para garantir a eficácia dos mesmos (Costa & Aparício, 2019; Eckerson, 2010; Duarte & Costa, 2012; Few, 2006; Janes et al., 2013; Kerzner, 2017; Sanz, 2018), tais como:

- o **layout do dashboard** – deve-se conseguir exibir a informação mais importante e relevante numa só tela;
- a **finalidade do dashboard** – ter em consideração os interesses dos destinatários e os objetivos que pretendem alcançar, pois altera o tipo de informação a utilizar;
- a **quantidade de informação utilizada** – deve-se apresentar apenas a informação estritamente necessária de forma a facilitar o processo de tomada de decisão e evitar a recolha de informação excessiva e de pouca utilidade;
- a «**arte**» do **painel** – recorrer apenas entre 3-5 cores distintas no mesmo painel, e sempre que possível, utilizar cores que tenham em si associadas um significado latente, como é o caso do vermelho a um indicador de alerta ou o verde a um indicativo de sucesso (Costa & Aparício, 2019; Janes et al., 2013);
- a **fonte utilizada** – não devem ser empregues mais de 2-3 tipos de letras diferentes num *dashboard* e quando for necessário destacar uma informação deve-se recorrer a nova fonte da mesma família, utilizar um tamanho de letra maior ou destacar palavras através do **negrito**;
- a **componente gráfica** – recorrer à utilização de quadros, gráficos e/ou tabelas adequados consoante o objetivo a alcançar (comparação, distribuição, composição ou relação), a natureza dos dados (estatísticos ou oscilantes) e/ou o número de variáveis/períodos em análise (Costa & Aparício, 2019). O esquema representado

no Anexo A ajuda a identificar a componente gráfica que melhor representa a informação com base nos fatores anteriormente referidos.

Ao se ter em consideração estas características, Eckerson (2010) acredita que estes painéis conseguem auxiliar, otimizar e acelerar a tomada de decisões, uma vez que fornecem *dashboards* com “*a informação certa, aos utilizadores certos, no momento certo*” (p.8).

Todavia, a visualização de dados através de ferramentas de consulta e relatórios eficientes de apoio às decisões carece de toda uma infraestrutura, isto é, um sistema de procedimentos com diversas fases de processamento, onde grandes quantidades de dados de múltiplas fontes são tratadas de forma a obter elementos informativos relevantes para os processos de tomada de decisão (Krmac, 2011; Laudon & Laudon, 2014). É neste sentido que de seguida serão apresentados os diversos componentes por detrás da arquitetura de um sistema de BI.

2.2.4 Arquitetura do Sistema de BI

Na ótica de Himami et al. (2021), esta arquitetura pode ser decomposta em quatro principais componentes: **Fonte de Dados** (*Data Sources*), **ETL** (*Extraction, Transformation and Loading*), **Repositório de dados** e **Visualização e Análise dos Dados**.

As **fontes de dados** caracterizam-se pela sua heterogeneidade, isto é, podem ser compostas por dados com origem interna ou externa à organização mas também podem contar com dados de diferentes tipos, categorizados por dados estruturados ou não estruturados (Du et al., 2019; Vercellis, 2009).

Eberendu (2016) define dados estruturados ao conjunto de dados apresentados em tabelas e organizados em linhas e colunas, que seguem uma formatação específica (*string*) e que normalmente se encontram armazenados em bases de dados relacionais.

Os dados não estruturados caracterizam-se por serem dados sem uma estrutura particularmente definida. Imagens, textos em formato de *e-mail* e de relatórios, páginas de *Web*, vídeos e dados em forma de áudio são alguns dos exemplos deste tipo de dados, segundo Eberendu (2016). O surgimento destes intensificou-se com a massificação das *IoT - Internet of Things*, também conhecidas como «A *Internet* das Coisas» (Azad et al., 2020).

Transformar grandes quantidades de dados em dados de qualidade e inteligíveis pode ser um desafio a enfrentar pela *data sources* (Maleki & Sabet, 2022), porquanto, coletar diferentes dados exige uma uniformização dos mesmos. Isto é, para que se consiga aceder a todos os dados, independentemente da sua origem, é preciso agrupá-los através de um processo denominado por *ETL* (Rahm, 2016).

Todos os dados, estruturados e não estruturados, quando extraídos das respetivas fontes, sejam internas ou externas, vão ser submetidos a uma etapa de integração e transformação através do processo denominado por *ETL – Extraction, Transformation and Loading* (Simitsis et al., 2005; Talaoui & Kohtamäki, 2020).

Na perspetiva de Souibgui et al. (2019) e de Nwokeji et al. (2018), este procedimento destina-se a realizar as três seguintes funções:

- Primeiramente, a função ***Extraction*** em que se faz uma **extração dos dados** de diversos formatos inseridos nas *data sources*;
- De seguida, a função ***Transformation*** onde se realiza a **transformação dos dados**, nomeadamente a sua limpeza, correção e uniformização, de forma que possam ser convertidos em formatos e estruturas adequadas para que seja possível armazená-los e posteriormente analisá-los;
- E por fim, a função ***Loading*** que se destina a fazer o **carregamento dos dados** que resultaram da etapa anterior, mas agora num novo repositório de dados, com o intuito de os analisar. Repositório esse denominado, tipicamente, por ***Data Warehouse*** e/ou ***Data Marts***.

O sucesso desta etapa, isto é, a obtenção de dados de qualidade neste procedimento, é essencial para a performance dos procedimentos seguintes (Dayal et al., 2009).

Sendo o **repositório de dados** o lugar onde foram carregados os dados limpos, corrigidos e uniformizados, esta componente da infraestrutura de BI visa suportar os dados tratados que se vão analisar (Garani et al., 2019). A este repositório de “*dados agregados, limpos e estruturados*” (Muntean & Surcel, 2013, p. 114) dá-se o nome de ***Data Warehouse (DW)***. Contudo, e para que se consiga fornecer ao utilizador uma estrutura de dados de fácil acesso e compreensão, estes DW são elaborados tendo como base uma técnica de estruturação conhecida por **modelação multidimensional** (Krnetá et al., 2008). Ambos os conceitos vão ser explorados em tópicos seguintes.

Por último, depois de recolhidos os dados heterogéneos (*Data Sources*), do seu devido tratamento através do procedimento *ETL*, e dos mesmos serem carregados no *Data Warehouse*, os dados são modelados em Modelos Multidimensionais como forma de facilitar a sua apresentação e, conseqüentemente, a sua análise. Para tal, é frequente os utilizadores recorrerem a técnicas de **visualização de dados** que permitam a sua **análise** de forma intuitiva e rápida como forma de auxiliarem a sua tomada de decisão. Esta visualização de dados pode ser feita através da criação de **Relatórios do tipo Dashboards** – como vimos anteriormente – ou através de uma técnica denominada por *OLAP – Online Analytical Processing*. Esta última, permite a modelação dos dados armazenados no *Data Warehouse* e que os seus utilizadores detenham a possibilidade de fazer uma análise dos dados mediante diferentes dimensões e conforme as suas necessidades, uma vez que esta técnica, ao utilizar cubos multidimensionais, permite obter uma visualização dos dados em várias perspetivas (Phan & Vogel, 2010).

Os cubos multidimensionais são ferramentas que permitem realizar um vasto tipo de operações que facilitam a análise dos dados. Dentro destas operações, duas das mais importantes e utilizadas são o *Roll-Up* e o *Drill-Down*, que permitem fazer uma agregação dos dados nos cubos de forma, respetivamente, menos ou mais detalhada (Santos & Ramos, 2006).

2.2.5 Data Warehouse (DW) e Data Marts (DM)

Data Warehouse (DW) é um dos principais componentes num sistema de informação dado que possibilita armazenar diversos dados num só local como forma de facilitar o acesso aos seus utilizadores (Caetano & Costa, 2014; Yessad & Labiod, 2016).

Ralph Kimball e *Bill Inmon* são dois dos autores mais conceituados nesta área. Na ótica de *Kimball*, DW define-se como um banco de dados constituído por uma cópia de dados estruturados (Ross & Kimball, 2013). Para *Inmon* (2005), DW é visto como uma “*coleção de dados orientada por assunto, integrada, não volátil e variável no tempo*” (p.29). Apesar das suas perspetivas díspares, ambos consideram que o objetivo destes repositórios é permitir analisar os dados com a expectativa de que se consiga alcançar informação pertinente para a tomada de decisão (Ross & Kimball, 2013).

Não obstante da vantagem intrínseca dos repositórios, ao possibilitarem a armazenagem de grandes volumes de dados limpos, corrigidos e uniformizados, tal vantagem pode dificultar a análise dos dados necessários para a tomada de decisões. Pelo que se torna necessário aceder a esses dados de forma estruturada de maneira a acelerar o processo de

tomada de decisão. Para tal, por vezes torna-se pertinente criar um repositório de proporções menores tendo como base uma DW de maior dimensão (Burstein & Holsapple, 2008).

Estes repositórios de menor dimensão constituídos por um subconjunto de dados de um determinado nível de hierarquia – isto é, dados de uma área ou segmento/departamento específico – denominam-se por **Data Marts (DM)** (Krneta et al., 2016).

Assim, e segundo as diferentes abordagens metodológicas desenvolvidas por *Inmon* e *Kimball*, os DM podem ser respetivamente dependentes ou independentes do DW. Ou seja, se a construção destes repositórios for feita de «cima para baixo», estamos perante a **abordagem Top-Down** desenvolvida por *Bill Inmon* onde, na ótica deste autor, os DM são construídos depois da construção do DW, e por isso, são dependentes destes (Inmon, 2005). (Ver Anexo B) Se pelo contrário, a construção dos DM é feita de «baixo para cima», o respetivo processo metodológico já segue a **abordagem Bottom-Up** desenvolvida por *Ralph Kimball* que defende que a construção de DM é feita por um conjunto de DW, fazendo com que estes sejam independentes dos primeiros (Ross & Kimball, 2013). (Ver Anexo C)

2.2.6 Modelação dos Dados Multidimensionais

Como forma de organizar os dados nos repositórios (*Data Warehouse* e *Data Marts*) recorre-se frequentemente a uma técnica denominada por **Modelação Multidimensional**. Para se compreender a construção destes modelos é necessário apreender 2 conceitos essenciais: a **Tabela de Factos (TF)** e a **Tabela de Dimensão (TD)** (Ross & Kimball, 2013).

A Tabela de Factos (TF) é o centro de todos os modelos desenvolvidos por *Kimball*. É nesta tabela que se encontra a área que se pretende analisar. Esta é composta por um conjunto de atributos, também conhecidos como factos, e por um conjunto de chaves estrangeiras. É através das chaves estrangeiras que é possível relacionar a TF com as várias Tabelas de Dimensão (TD) e, por sua vez, cada TD permite analisar cada uma das perspetivas dos dados em estudo (Santos & Ramos, 2006).

É com base nestes conceitos que *Ralph Kimball* cria um dos processos de estruturação de dados mais conhecido e utilizado: a modelação de dados através de um **Esquema em Estrela (Star Schema)** (Hart & Kuo, 2016). Esta forma de estruturar os dados caracteriza-se pela ligação de várias TD a uma TF central. Esta forma de modelagem em estrela facilita a perceção das relações existentes entre os dados (Ross & Kimball, 2013) (Ver

Anexo D). Uma particularidade deste tipo de estruturação dos dados é que torna possível interligar diferentes Esquemas em Estrela. Esta fusão de esquemas dá origem a um novo modelo, conhecido por **Esquema em Constelações** ou **Esquema em Galáxia**, que se caracteriza por a mesma TD ser partilhada por mais do que uma TF (Santos & Ramos, 2006) (Ver Anexo E).

Este modelo de estruturação de dados esteve na base de um outro, denominado de **Esquema em Floco de Neve** (*Snowflake Schema*), que se caracteriza por ter em conta uma Tabela de Facto, várias Tabelas de Dimensão e várias **Tabela de Subdimensões**. Ou seja, no Esquema em Floco de Neve é possível verificar que, se por um lado, às Tabela de Dimensão estão ligadas Tabelas de Subdimensão, por outro lado, à Tabela de Facto encontram-se ligadas, conseqüentemente, todo o conjunto das Tabelas de Dimensão (Santos & Ramos, 2006) (Ver Anexo F).

2.2.7 Ferramentas para a Construção de Sistemas de BI

Existe uma grande variedade de ferramentas que ajudam a implementar um sistema de BI. Estas ferramentas de análise, cujo objetivo se baseiam no auxílio do processo de decisão, podem ser representadas através do **Quadrante Mágico de Plataformas de Análise e BI** desenvolvido pela Gartner (2022), conforme representado no Anexo G.

Verifica-se que a plataforma *Microsoft* se destaca das demais ao oferecer soluções líderes de gestão e análise de dados. Por esse motivo, o desenvolvimento deste trabalho terá na sua base a ferramenta de visualização de dados conhecida por **Microsoft PowerBI** que, através de relatórios e a visualização de dados de forma interativa, permite, de forma rápida, simples e intuitiva, obter a informação pertinente para a tomada de decisão.

2.2.8 Desafios Enfrentados pelo BI

A maioria das organizações revela que os benefícios esperados com o uso de BIS não vão ao encontro dos resultados na realidade (Aws et al., 2021). Um estudo elaborado por Gartner (2022) detetou que cerca de 70 a 80% dos projetos BIS falham (Ranjbarfard & Hatami, 2020).

Segundo Joshi e Dubbawar (2021), alguns dos desafios que as organizações enfrentam com a implementação de um BI são:

- **Falta de skills** – nem todos os utilizadores adquirem conhecimentos necessários para utilizarem ferramentas de BI;

- **Custo de implementação** – apesar de haver ferramentas gratuitas, algumas técnicas requerem um investimento inicial muito elevado o que pode ser um impedimento para empresas de menor dimensão;
- **Integração dos dados de diferentes fontes** – apesar de ser uma adversidade cada vez menor, ainda existem algumas dificuldades em analisar grande quantidade de dados provenientes de diferentes fontes;
- A **baixa qualidade dos dados** – a elevada quantidade de dados disponíveis leva a que os processos de BI sejam demorados por terem de processar dados indesejados e desnecessários.

2.3 A Educação na Sociedade Atual

O acesso à educação nem sempre foi um direito para todos os Homens. Nos primórdios da história, a educação apenas era concedida a indivíduos que, por mera sorte, nascessem no seio das elites ou seguissem atividades eclesiásticas (Gameiro, 2016; Schippling et al., 2020). Contudo, ao longo dos séculos, o papel da educação na vida em sociedade tem vindo a ser cada vez mais valorizado, compreendido e realçado, sendo considerado um dos principais alicerces da vida do homem em sociedade (Cerdeira & Cabrito, 2018).

Com as mudanças sentidas na sociedade, nomeadamente com o crescimento da industrialização, tornou-se desejável atuar de forma cooperativa. Percebeu-se as vantagens que a educação podia trazer como forma de alcançar mão-de-obra instruída e qualificada (Fomunyan, 2019).

Assim, o século XX ficou caracterizado pela publicação da Declaração Universal dos Direitos do Homem (ONU, 1948). Nesta Declaração, consagrada pelos países membros das Nações Unidas (ONU) em 1948, um dos direitos proclamados é o direito à educação a todos os indivíduos (Artigo 26º).

Ao longo do tempo, e após este direito ter sido consagrado, a sociedade enfrenta um período em que a produção e acesso à informação é cada vez mais veloz (Cruz, 2008). Como forma de alcançar um maior índice de indivíduos que consigam ter pensamento crítico, analisar informação e, conseqüentemente, dar respostas pertinentes e estratégicas, apostar na educação qualificada tornou-se crucial, uma vez que quanto maior o nível de ensino frequentado pelos indivíduos maiores as suas habilitações e competências (Masino & Niño-Zarazúa, 2016).

Assim, a «condição-chave» para o desenvolvimento das sociedades modernas passa por incentivar as populações a frequentar ensinos educacionais cada vez mais elevados – como o *Ensino Superior* (Dziewanowska, 2017).

2.3.1 O Sistema de Ensino Superior

Depois da conclusão do Ensino Secundário, o acesso ao Ensino Superior dita a continuação, ou não, dos estudos numa determinada área específica. Este nível de ensino sofreu uma das reformas educacionais mais impactantes da sua edificação, quando a 1999 em Bolonha, Itália, se deu a **Declaração de Bolonha** (Diogo et al., 2019). Este processo, acordado entre 29 países europeus, inclusive Portugal, teve como principal objetivo promover um Espaço Europeu de Ensino Superior (EEES) que se destacasse pela sua competitividade e qualidade, onde o ensino se baseie num sistema de créditos (ECTS - *European Credit Transfer and Accumulation System*) como forma de promover a coesão europeia e incentivar a mobilidade dos estudantes entre os países, europeus e terceiros, bem como a empregabilidade dos seus diplomados (Davies, 2008; Klemenčič, 2019).

De acordo com esta reforma, os cursos do Ensino Superior estão organizados em três ciclos distintos (Decreto-Lei N.º 74/2006 do Ministério da Ciência, Tecnologia e Ensino Superior, 2006; DGES, 2022b):

- 1º Ciclo destinado à Licenciatura – com duração mínima de 6 semestres onde a formação é orientada para o mercado de trabalho;
- 2º Ciclo destinado ao Mestrado – visto como uma especialização e que tem duração de 3 a 4 semestres; e o
- 3º Ciclo destinado ao Doutoramento – com a durabilidade de 6 a 8 semestres.

Para tal, e independentemente do curso, todos os indivíduos que pretendam submeter a sua candidatura/inscrição de forma a dar entrada no par estabelecimento/ciclo deverão satisfazer um conjunto de condições previamente estabelecidas. Não só apenas poderão candidatar-se ao 2º Ciclo e, consecutivamente, ao 3º Ciclo após a conclusão do Ciclo anterior, como os candidatos têm ainda de satisfazer um conjunto de requisitos fixados para cada par estabelecimento/ciclo de estudos, nomeadamente em termos de pré-requisitos, de classificação mínima em provas de ingresso e de classificação mínima na nota de candidatura (Decreto-Lei N.º 296-A/98 do Ministério da Educação, 1998).

Independentemente do Ciclo, para Almeida e Casanova (2019), é neste nível de ensino que as instituições responsáveis assumem uma responsabilidade significativa na

formação das novas gerações, ao lhes fornecer competências científicas e técnicas, humanas e culturais, assim como competências transversais.

O Ensino Superior em Portugal trata-se de um sistema binário (Simões & Soares, 2010) onde se integra o Ensino Universitário/Politécnico e é ministrado em Estabelecimentos de cariz Público/Privado (DGES, 2022c).

Este sistema educacional conta com Unidades Orgânicas (Faculdades ou Escolas) que se encontram associadas a um único Estabelecimento (Universidades). Estes, podem ainda ser classificados quanto ao seu Tipo de Ensino e Natureza, respetivamente. Pelo que, nos termos dos artigos 3º e 4º da Lei N.º 62/2007 do Regime Jurídico das Instituições de Ensino Superior (2007), quanto ao Tipo de Ensino, as Unidades Orgânicas dividem-se entre ensino Universitário ou Politécnico e os Estabelecimento, quanto à sua Natureza, entre os subsistemas Público ou Privado.

Assim, conforme previsto no artigo 11º da Lei N.º 46/86 de Bases do Sistema Educativo (1986), o sistema conta em formar os indivíduos da comunidade através das suas Unidades Orgânicas (que se compreendem entre Ensino Universitário e Ensino Politécnico). Estas diferenciam-se assim pelo Tipo de Ensino ministrado, isto é, pela orientação e propósito do conhecimento e competências que pretendem desenvolver nos seus formandos (Urbano, 2011). Segundo o n.º 1 do artigo 3º da citada Lei N.º 62/2007 de 10 de setembro, enquanto o **Ensino Universitário** está orientado para oferecer aos indivíduos uma formação técnica sólida e desenvolver as suas capacidades de inovação e de análise crítica, o **Ensino Politécnico** destina-se à formação de cariz profissional, através da transmissão de conhecimentos científicos de carácter teórico e prático (Feitosa et al., 2020; Ferrão & Almeida, 2019).

Como definido no Artigo 4º do Decreto-Lei N.º 74/2006 do Ministério da Ciência, Tecnologia e Ensino Superior (2006), outro aspeto que diferencia ambos é que, enquanto o Ensino Universitário tem competências para formar indivíduos dos três ciclos - isto é, formar licenciados, mestres e doutores - o Ensino Politécnico apenas tem valência para prestar formação a indivíduos do 1º e 2º Ciclo.

Porém, estas Unidades Orgânicas, e independentemente do Tipo de Ensino administrado, são associadas a um Estabelecimento que pode ser caracterizado segundo a sua Natureza. Assim, segundo o estabelecido no Artigo 4º da Lei N.º 62/2007 do Regime Jurídico das Instituições de Ensino Superior (2007), o Ensino Superior pode ter um cariz **Público** –

prestado por instituições que pertencem ao Estado – ou cariz **Privado** – prestado por instituições que pertencem a entidades privadas ou cooperativas.

Com o aumento da aposta em formação qualificada, a oferta dada por parte do subsistema de Ensino Público não consegue dar resposta à crescente procura sentida. Foi como forma de combater, por um lado, essa crescente procura por parte da sociedade e, por outro lado, combater a insatisfação da sociedade mediante a prestação do serviço educacional de cariz Público, que surge o Ensino Superior Privado (Seixas, 2000).

Com a introdução deste novo subsistema de ensino, surge a discussão de saber se o acesso ao ensino deverá ser um direito/bem público, e por isso fornecido de forma igualitária e gratuita através do Estado, ou um privilégio/bem privado que, pelo seu carácter privado se torna um bem transacionável e, assim, apenas os mais favorecidos têm a possibilidade de a ele aceder (Carnoy et al., 2014; McCowan, 2012).

Uma das principais distinções entre estes subsistemas de ensino consiste na forma de acesso ao mesmo. Assim, no Ensino Superior Público, o aluno submete a sua candidatura de forma gratuita onde, conseqüentemente, esta será submetida a um concurso público nacional como forma de dar prioridade de admissão aos alunos que consigam obter notas de admissão mais elevadas, enquanto os alunos que pretendam aceder ao Ensino Superior Privado, necessitam de submeter a sua inscrição mediante o pagamento de uma pré-inscrição (Ferrão & Almeida, 2019).

2.4 O *Business Intelligence* Associado à Educação

Com fim de analisar, de forma fácil e rápida, um conjunto de dados - que se encontram armazenados pelas mais díspares fontes - as organizações estão cada vez mais propensas a recorrer a análises em BI (Drake & Walz, 2018; Wang, 2015).

Esta valorização de sistemas de BI a muito se deve à forma como estas ferramentas projetam a informação em *dashboards*, permitindo tomar decisões mais inteligentes e melhorando significativamente os resultados obtidos (Guster & Brown, 2012). É neste sentido que, as organizações de diferentes áreas, inclusive da educação, se têm sentido cada vez mais atraídas por estes mecanismos de análise (Nidhishree & Badar, 2019).

Assim, ter acesso a instrumentos que permitam analisar o grande volume de dados no contexto do ensino superior, e desses dados extrair conhecimento, poderá levar a vantagens competitivas consideráveis perante os demais estabelecimentos de ensino. Porquanto, os respetivos resultados obtidos, de forma automatizada e sistematizada,

quando utilizados e analisados de forma eficiente, podem constituir uma mais-valia para a IES e conduzir a fatores de diferenciação face aos seus principais concorrentes no setor de ensino (Nobre et al., 2014; Piedade & Santos, 2010).

Nobre et al. (2014) e Nidhishree e Badar (2019) defendem assim que, construir um instrumento no domínio do BI que seja capaz de fornecer informação referente: às candidaturas ao Concurso Nacional de Acesso (CNA) ao ensino superior nos últimos anos; à disposição geográficas das instituições de ensino; e ao número de alunos que concluíram o seu curso por ano, entre outras, poderá contribuir para o melhoramento interno de cada IES bem como para o sucesso externo das mesmas quando comparadas com as suas principais concorrentes.

É nesta ordem de ideias que se torna sugestivo realizar uma análise do Ensino Superior em BI, tendo como base um dos seus principais *stakeholds*: as Instituições de Ensino Superior (IES) (Raimundo & Rosário, 2021).

CAPÍTULO 3: METODOLOGIA

Em 1996, como forma de auxiliar a análise de dados, isto é, o processo de *Data Mining*, um consórcio formado por cinco empresas desenvolveu uma metodologia denominada por *Cross-Industry Standard Process for Data Mining (CRISP-DM)* (Chapman et al., 2000; Costa & Aparicio, 2020; Costa & Aparicio, 2021; Jaggia et al., 2020).

Visto que este modelo se caracteriza pela sua eficiência e se destaca das restantes pela sua vasta aplicabilidade em diversas áreas (Schröer et al., 2021), para a realização do presente projeto será utilizada essa abordagem metodológica.

Uma das mais-valias desta abordagem é a sua composição, isto é, o seu ciclo metodológico que, se caracteriza por um conjunto de seis fases distintas interligadas entre si. A flexibilidade do processo deve-se essencialmente aos seus ciclos e retornos entre as fases. As setas externas representam a natureza cíclica do processo, enquanto as setas internas representam as dependências entre as fases (Chapman et al., 2000).

Assim, de forma que seja possível dar seguimento a uma determinada fase é necessário conhecer os resultados da sua fase antecedente. Esta interdependência entre as fases do processo pode ser assim representada através da direção das setas, uma vez que o fim de uma fase indicará o início da fase seguinte. Este processo pode ser visualmente entendido através da Figura 1:

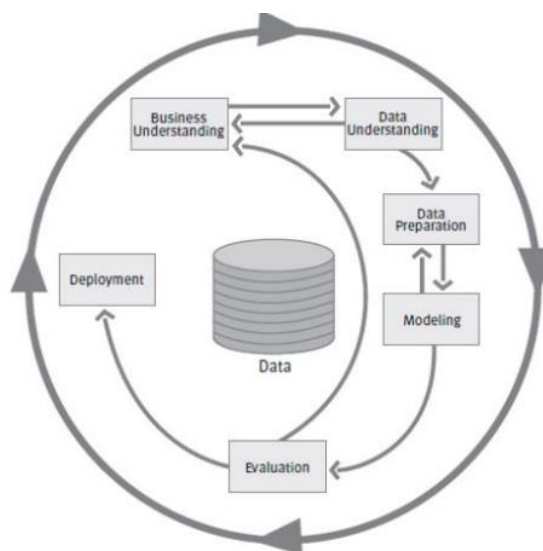


Figura 1 - Fases do ciclo metodológico CRISP-DM (Fonte: Chapman et al., 2000)

Segundo Chapman et al. (2000), e como é retratado na figura anterior, esta metodologia é composta por seis fases em que cada uma delas se descreve, de forma sumária, da seguinte forma (Kruts, 2020; Martins, 2021):

1. **Business Understanding:** Na fase inicial do processo, é fundamental identificar e depreender o problema que precisa de ser solucionado, pelo que é necessário compreender o negócio onde o projeto se insere e definir os objetivos a alcançar com o seu desenvolvimento.
2. **Data Understanding:** Esta fase conta com a seguinte sequência de quatro etapas: recolha e organização dos dados que pretendemos utilizar; descrição desses dados de forma de entender a informação contida nos mesmos; exploração e verificação da qualidade dos dados. Feito isto, é possível descodificar o problema e identificar certos padrões nos dados em análise.
3. **Data Preparation:** Na terceira fase do ciclo metodológico são realizadas as ações de limpeza, construção, integração e formatação dos dados para se obter o conjunto final de dados que vão ser usados no projeto e estes poderem ser usados em ferramentas de *data mining*.
4. **Modeling:** A fase de modelação consiste na seleção e aplicação das técnicas mais adequadas de *data mining* para alcançar os objetivos definidos anteriormente, isto é, na escolha do melhor modelo representativo de conhecimento útil.
5. **Evaluation:** Posteriormente, será feita uma revisão e avaliação de desempenho do modelo por comparação dos resultados obtidos com as referências métricas previamente definidas.
6. **Deployment:** A última fase é à construção efetiva do projeto, em que se organiza e se disponibiliza os resultados aos utilizadores, por exemplo, através da elaboração de um relatório da implementação do processo *data mining* numa Instituição.

CAPÍTULO 4: APLICAÇÃO DA METODOLOGIA – CRISP-DM

Este capítulo compreende a aplicação das seis fases do ciclo metodológico CRISP-DM (Kruts, 2020; Martins, 2021). Na sua implementação foi respeitado o respetivo fluxo dos ciclos e não houve necessidade de retroceder a etapas anteriores, dado que as informações obtidas em cada uma das fases do processo foram tidas como significativas e suficientes para dar continuidade ao ciclo metodológico e alcançar os objetivos propostos.

4.1 Fase 1 - *Business Understanding*

Numa primeira fase da metodologia proposta, e de forma a alcançar os objetivos propostos, é sugestivo que se faça uma compreensão do setor educacional sobre o qual o projeto se vai debruçar: o Ensino Superior.

A educação ao funcionar como uma hierarquia, faz com que o acesso ao Ensino Superior provenha após a conclusão dos dois níveis antecedentes de ensino: o Ensino Básico e o Ensino Secundário.

A educação superior constitui assim o nível mais elevado da hierarquia educacional e é fornecida por Unidades Orgânicas (que lecionam mediante o seu Tipo de Ensino: Universitário ou Politécnico) que, normalmente, se agrupam num Estabelecimento, também denominado por Universidade, de Natureza Privada ou Pública. A frequência neste nível de ensino advém de uma vontade de adquirir saber especializado numa determinada área, o que tem sido cada vez mais procurado pelos indivíduos como forma de adquirirem habilitações para a sua vida futura.

Dentro do Ensino Superior estão inseridos diferentes ciclos – como referido anteriormente na revisão da literatura – onde a conclusão de cada ciclo confere ao estudante um determinado grau académico. Porém, o estudo proposto irá analisar os dados referentes ao 1º Ciclo, ou seja, a Licenciatura bem como, e por existirem cursos que englobam o ciclo objeto do estudo e o 2º Ciclo, os dados referentes a estes cursos denominados de Licenciatura com Mestrado Integrado.

Assim, o projeto desenvolvido pretende disponibilizar informação, pertinente e organizada, como forma de permitir uma análise generalizada do grau académico em análise. Ao considerar a relação entre os três principais objetos em estudo – Estabelecimentos, Unidades Orgânicas e Cursos – nos últimos oito anos letivos, o projeto poderá cooperar na tomada de decisão das IES que assim o entendam.

4.2 Fase 2 - *Data Understanding*

Esta nova fase do processo passa por 4 etapas, como referido anteriormente: a recolha, a descrição, a exploração e a verificação da qualidade dos dados. Caso se faça uma analogia com o processo *ETL*, esta fase diz respeito à extração.

A etapa da recolha dos dados nem sempre é fácil devido às políticas de proteção de dados. Porém, e como forma de ultrapassar este constrangimento, os dados utilizados para a realização do projeto foram recolhidos do *Portal Infocursos* (InfoCursos, 2022) que, é uma base de dados aberta, isto é, uma *Open Access* (OA), que fornece de forma pública, imediata e sem qualquer tipo de limitação, dados em estado bruto aos seus utilizadores (Butler-Henderson et al., 2020). Esta é uma plataforma gerida pela Direção-Geral de Estatísticas da Educação e Ciências com o apoio da DGES, o que permitiu assegurar a qualidade e credibilidade dos dados recolhidos.

De forma a compreender os dados a usar e, assim, facilitar o cumprimento dos objetivos propostos, procedeu-se ainda à descrição e exploração dos dados.

O facto desta plataforma disponibilizar aos seus utilizadores dados relativos aos Cursos Superiores, ministrados em Estabelecimentos de Ensino Superior Portugueses, desde o ano letivo 2013/2014 (através da Base de dados “junho de 2015”) até ao ano letivo 2020/2021 (através da Base de dados “junho de 2022”), e em formato “.xlsx”, permitiu fazer uma agregação dos diferentes dados num só documento Excel.

Para tal, fez-se primordialmente uma recolha de dados relativos a informações gerais onde se considera como variáveis qualitativas todas as informações que caracterizam o acontecimento em análise – a frequência de um curso, num determinado ano, na respetiva Universidade/Faculdade. Assim, as variáveis recolhidas foram:

Tabela I - Nome, Tipo e Descrição das Variáveis Qualitativas retiradas dos diferentes Indicadores

| Indicador | Nome da Variável | Tipo Variáveis Qualitativas | Descrição |
|--|---|-----------------------------|---|
| <i>Ano Letivo</i> | AnoLetivo | Ordinal | Variável que indica o ano letivo da ocorrência analisada. Esta variável está compreendida entre o ano letivo de 2013/2014 ao ano letivo 2020/2021. |
| <i>Estabelecimento</i> Também denominado por Universidade | Estabelecimento <hr/> LocalizaçãoUniversidade <hr/> Website | Nominal | Variável que indica a Universidade da ocorrência analisada. Nota: Cada estabelecimento tem a si associado um <i>idEstabelecimento</i> como forma de simplificar a sua caracterização na Tabela de Factos. Através da DGES, foi possível recolher a informação relativa a estas 2 variáveis que assumem essencialmente carácter informativo das Universidades. |
| <i>Natureza</i> | Natureza | Nominal | Ao indicar se a Universidade é de cariz Público ou Privado, esta variável indica a Natureza de Ensino de determinado Estabelecimento. |
| <i>Unidade Orgânica</i> Também denominado por Faculdade | UnidadeOrganica <hr/> LocalizaçãoFaculdade | Nominal | Variável que indica a Faculdade da ocorrência analisada. Nota: Cada Faculdade tem a si associada uma sigla denominada por <i>idUnidadeOrganica</i> A informação relativa à localização da Unidade Orgânica assume carácter informativo e foi retirada da DGES. |
| <i>Tipo de Ensino</i> | TipoEnsino | Nominal | Variável que indica o tipo de ensino da ocorrência analisada. Podendo assim assumir a característica de Universitário ou Politécnico. |
| <i>Curso</i> | NomeCurso <hr/> Grau <hr/> ÁreaCurso | Nominal | Variável que indica o Curso da ocorrência analisada. Nota: Cada Curso tem a si associado um código, definido por <i>idCurso</i> . O estudo realizado teve como base dados relativos a cursos de 1º Ciclo e de 1º e 2º Ciclo. Estes cursos são classificados pelo seu Grau ao serem denominados por Licenciatura e Mestrado Integrado, respetivamente. Todos os cursos foram agrupados por áreas mediante a classificação feita pela DGES. |

Com base nos dados fornecidos pelo *Portal Infocursos*, foi ainda possível recolher algumas variáveis quantitativas. Os dados desta plataforma estão organizados em diferentes Folhas de Excel, no entanto, a variável apresentada de seguida está disponível em diversas folhas e por esse motivo não será agrupada a nenhum indicador:

Tabela II - Nome, Tipo e Descrição da Variável Comum a diferentes Indicadores

| Nome da Variável | Tipo de Variável | Descrição |
|--|-----------------------|--|
| NumInscritos_NoAnoNoParUniversidadeCurso | Quantitativa Discreta | Variável que indica o número total de inscritos num determinado Curso, no ano em análise, numa determinada Universidade – Faculdade. |

Todas as outras variáveis estão disponíveis em Folhas de Excel específicas mediante os seguintes 6 indicadores em análise: **(1)** Distribuição dos alunos inscritos por sexo; **(2)** Distribuição dos alunos inscritos por nacionalidade; **(3)** Distribuição por idades dos alunos inscritos no Curso; **(4)** Formas de Ingresso no Curso; **(5)** Percentil médio de entrada no Curso; **(6)** Distribuição de classificações finais

Tabela III - Nome, Tipo e Descrição das Variáveis Quantitativas retiradas dos diferentes Indicadores

| Indicador | Nome da Variável | Tipo Variável Quantitativa | Descrição |
|-----------------------------|-----------------------------|----------------------------|--|
| (1) | NumInscritos_Homens | Discreta | Indicador que distribui o total de alunos inscritos por sexo – Feminino e Masculino – no respetivo ano letivo analisado. Este indicador os alunos inscritos em programas de mobilidade internacional. |
| | NumInscritos_Mulheres | | |
| (2) | NumInscritos_Portugueses | Discreta | Indicador que distribui o total de alunos inscritos por nacionalidade – Portuguesa e Estrangeira – no respetivo ano letivo analisado. Este indicador os alunos inscritos em programas de mobilidade internacional. |
| | NumInscritos_Estrangeiros | | |
| (3) | NumInscritos_Idade<=18Anos | Discreta | Indicador que distribui o total de alunos inscritos no Curso nos vários anos letivos por 11 variáveis mediante a sua frequência de idades. |
| | NumInscritos_Idade19Anos | | |
| | NumInscritos_Idade20Anos | | |
| | NumInscritos_Idade21Anos | | |
| | NumInscritos_Idade22Anos | | |
| | NumInscritos_Idade23Anos | | |
| | NumInscritos_Idade24Anos | | |
| NumInscritos_Idade25a27Anos | | | |
| | NumInscritos_Idade28a30Anos | | |

| Indicador | Nome da Variável | Tipo Variável Quantitativa | Descrição |
|-----------|--|----------------------------|--|
| | NumInscritos_Idade31a39Anos | | |
| | NumInscritos_Idade>39Anos | | |
| | NumInscritos_PrimeiroAno_PrimeiraVez | Discreta | Através desta variável é possível conhecer o número total de alunos que se inscreveram no 1º ano, pela primeira vez, nos cursos de Licenciatura 1º Ciclo e em cursos com Mestrado Integrado, num determinado ano letivo através das diferentes formas de ingresso. |
| (4) | %Inscritos_PrimeiraOpcao_Concurso Nacional | Contínua | Através destas 7 variáveis é possível conhecer, em percentagem, os alunos que se inscreveram num determinado curso, num ano, através das diferentes formas de ingresso. Desta forma, caso o Curso participe no Concurso Nacional de Acesso (CNA), os inscritos podem ter ingressado no par Universidade/Curso, num determinado ano, através da sua primeira opção do CNA; através da segunda opção do CNA; através das outras opções do CNA, nomeadamente 3ª, 4ª, 5ª e 6ª opção. Caso, o Curso não participe no CNA, os alunos podem ter ingressado através do Concurso Local ou Concurso Institucional. Ao que diz respeito às mudanças de instituição/Curso; às provas específicas que avaliam a capacidade de um indivíduo maior de 23 anos frequentar o ensino superior; às outras formas previstas na lei, e não indicadas anteriormente, como o caso de concursos e regimes especiais de acesso ao ensino superior, são outras formas de ingresso que podem ser recorridas para aceder a Curso de CNA ou não. |
| | %Inscritos_SegundaOpcao_Concurso Nacional | | |
| | %Inscritos_OutrosRegimeGeral | | |
| | %Inscritos_ConcursoLocal_ou_Concurso Institucional | | |
| | %Inscritos_MudancaCurso | | |
| | %Inscritos_ProvasMajores23Anos | | |
| | %Inscritos_OutrosRegimes | | |
| (5) | NotaIngressoMédia | Contínua | Através desta variável é possível conhecer a nota de ingresso média dos alunos que se matricularam num determinado Curso. |
| | NumMatriculados_CNA | Discreta | Variável que indica o número total de matriculados num determinado Curso através do CNA, no ano em análise, numa determinada Universidade – Faculdade. |
| (6) | NumDiplomados_NoAnoNoPar UniversidadeCurso | Discreta | Esta variável indica o número total de diplomados, no respetivo ano em análise, no par Universidade/Curso. |

| Indicador | Nome da Variável | Tipo Variável Quantitativa | Descrição |
|-----------|--|----------------------------|---|
| | NotaFinal_Media_Diplomados_NoAnoNoParUniversidadeCurso | Contínua | Indicador que permite ter conhecimento da nota final média dos alunos que concluíram o Curso – denominados por diplomados – numa determinada Universidade, no ano letivo em análise. |
| | %ClassifFinalDiplomado_Nota10 | Contínua | Através destas 11 variáveis é possível conhecer a percentagem de alunos diplomados que após concluírem o Curso numa determinada Universidade, no ano letivo em análise, obtiveram classificação final de 10 a 20 valores. |
| | %ClassifFinalDiplomado_Nota11 | | |
| | %ClassifFinalDiplomado_Nota12 | | |
| | %ClassifFinalDiplomado_Nota13 | | |
| | %ClassifFinalDiplomado_Nota14 | | |
| | %ClassifFinalDiplomado_Nota15 | | |
| | %ClassifFinalDiplomado_Nota16 | | |
| | %ClassifFinalDiplomado_Nota17 | | |
| | %ClassifFinalDiplomado_Nota18 | | |
| | %ClassifFinalDiplomado_Nota19 | | |
| | %ClassifFinalDiplomado_Nota20 | | |

Depois da conjugação dos dados numa única folha de Excel obteve-se uma base de dados constituída por 12106 linhas e 53 colunas, onde cada linha do ficheiro corresponde ao total de ocorrências, num ano letivo, num determinado Curso no par Universidade/ Faculdade. Base de dados em Excel essa que, através da sua reorganização, foi possível distribuir os respetivos dados por 7 tabelas distintas, num único *Livro de Excel*, fazendo com que a tabela principal (*Folha do Excel: T. Factos*) ficasse com um total de 45 colunas para as mesmas 12106 ocorrências/linhas.

4.3 Fase 3 - Data Preparation

Nesta fase do processo metodológico – que pode ser analogamente comparável com a etapa de transformação e carregamento de dados do processo ETL – procedeu-se, primeiramente, à importação dos dados recolhidos e distribuídos pelas diferentes tabelas. Para tal, e uma vez que todos os dados constam num *Livro do Excel* proveniente da fase anterior, foi possível importado para o *Microsoft Power BI Desktop* o conjunto de dados que dará suporte ao estudo proposto.

Uma vez inseridos os dados, como visível na seguinte Figura 2, selecionou-se todas as tabelas do ficheiro como forma de proceder aos processos de tratamento e limpeza de dados que, caracterizam esta fase do processo metodológico.

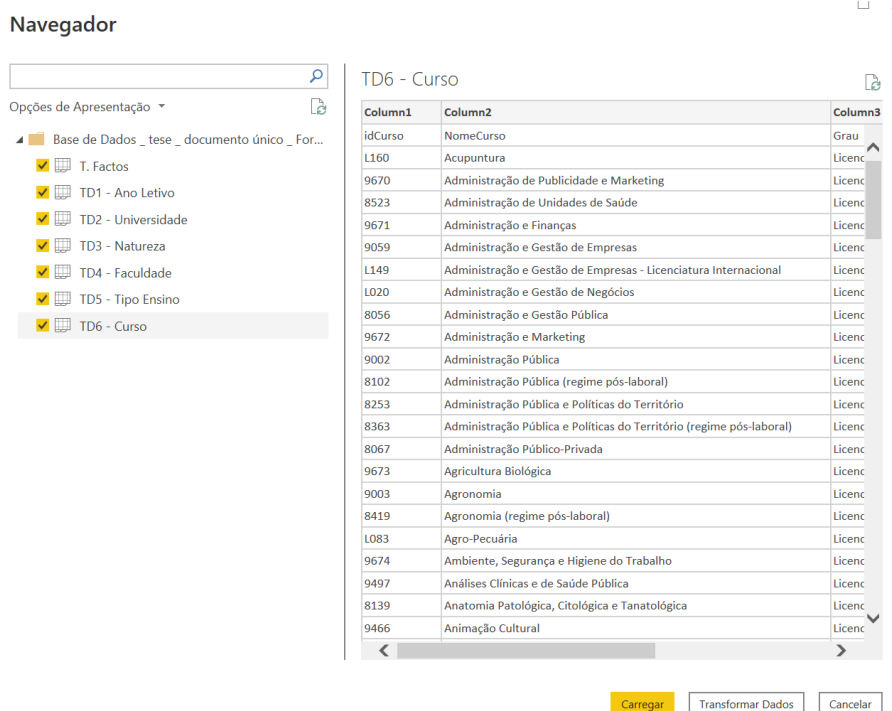


Figura 2 - Selecionar e Importar os Dados a Transformar no Microsoft Power BI Desktop (Fonte: Autor)

Depois de selecionar *Transformar Dados* é aberta uma janela que remete para o Editor do *Power Query*. Este motor de transformação e preparação de dados permitiu realizar as seguintes transformações nos dados importados:

- **Transformação 1:** *Utilizar Primeira Linha como Cabeçalho*

Por o *Power Query* considerar a primeira linha das tabelas de dados como cabeçalho, removeu-se a primeira linha de todas as tabelas cuja primeira linha não correspondesse ao cabeçalho – tal como ocorreu, por exemplo, com a Tabela *TD4 - Curso*.

- **Transformação 2:** *Alterar o Tipo de Dados*

Garantida a estruturação correta das tabelas, passou-se à formatação do *Tipo de Dados* de todas as colunas. Para estas utilizou-se os seguintes tipos de dados: “*Texto*” para variáveis como a “*Natureza*” e “*TipoEnsino*” das Tabelas de Subdimensão; “*Número Decimal*” para variáveis como a “*%Inscritos_PrimeiraOpcao_ConcursoNacional*” da *T. Factos*; e “*Número*

Inteiro” para variáveis como o “*NumInscritos_Homens*” retirada da mesma Tabela.

Depois de definido o *Tipo de Dados* para cada coluna, este é apresentado no lado esquerdo do título da respetiva coluna através do respetivo ícone.

- **Transformação 3: Remover Colunas**

Removeram-se todas as colunas indesejáveis, como colunas em branco e/ou colunas com informação irrelevante para o estudo em questão, nomeadamente colunas da Tabela *TD3 – Faculdade* e da Tabela *TD2 – Universidade*.

- **Transformação 4: Mudar o Nome das colunas**

De forma a tornar os dados mais coerentes no seu todo, alterou-se o nome de determinadas colunas, como por exemplo, nas colunas que contêm a informação da localização das Universidades e das Faculdades, mudou-se o nome da variável “*Morada*” da Tabela *TD2 – Universidade* para “*LocalizaçãoUniversidade*” de forma que esta ficasse em concordância com a variável da Tabela *TD3 – Faculdade* denominada por “*LocalizaçãoFaculdade*”.

- **Transformação 5: Substituir Valores**

Dado a variável “*NotaIngressoMédia*” da Tabela *T. Factos*, por falta de informação, ter o valor zero em algumas ocorrências irá influenciar a análise feita referente à média das notas de ingresso num determinado ano. Assim, procedeu-se à transformação desses valores de zero para *null*.

Realizado este conjunto de transformações, foi selecionada a opção *Fechar e Aplicar* para que os dados tratados fossem carregados e permitisse dar seguimento ao processo metodológico proposto.

Posteriormente, e já fora do *Power Query*, foram ainda criadas *Novas Colunas* na Tabela *T. Factos* para tornar algumas variáveis decimais aplicáveis ao modelo. Assim, uma vez que para todas as ocorrências, a soma das células referentes às variáveis entre “*%Inscritos_PrimeiraOpcao_ConcursoNacional*” e “*%Inscritos_Outros Regimes*”, bem como a soma das células referentes às variáveis entre “*%ClassifFinalDiplomado_Nota10*” e “*%ClassifFinalDiplomado_Nota20*”, tinham sempre um valor total de 100%, transformou-se os dados percentuais das variáveis retiradas dos indicadores “*Formas de Ingresso no Curso*” e “*Distribuição de Classificações Finais*”, respetivamente, em números inteiros, de forma que estas

variáveis indiquem sempre o número total de indivíduos com a mesma característica em cada ocorrência.

Com base na mesma lógica de execução, criou-se assim 18 novas colunas (7 colunas referentes às variáveis das “*Formas de Ingresso no Curso*” e 11 novas colunas alusivas às variáveis da “*Distribuição de Classificações Finais*”), onde através de funções e operadores aritméticos da linguagem *DAX (Data Analysis Expressions)*, se procedeu aos seguintes cálculos:

- **Novas colunas para as variáveis do Indicador “*Formas de Ingresso no Curso*”**
Sendo observável uma relação de proporcionalidade direta entre o “*NumInscritos_NoAnoNoParUniversidadeCurso*” e as variáveis deste indicador, o modelo apresentado na Operação 1, representa um exemplo disto: a transformação da coluna da variável “*%Inscritos_PrimeiraOpcao_ConcursoNacional*” na coluna defina por “*NumInscritos_PrimeiraOpção*”.

```
1 NumInscritos_PrimeiraOpção =
2 DIVIDE(
3 | 'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'
4 | [%Inscritos_PrimeiraOpcao_ConcursoNacional]*'TFactos - Entradas e Saídas num ano _ num curso _ no par
   Universidade_Faculdade'[NumInscritos_NoAnoNoParUniversidadeCurso],100
5 )
```

Operação 1 – Transformação da nova coluna “*NumInscritos_PrimeiraOpção*”

- **Novas colunas para as variáveis do Indicador “*Distribuição de Classificações Finais*”**

Da mesma forma, uma vez que todas as variáveis que constam entre “*%ClassifFinalDiplomado_Nota10*” e “*%ClassifFinalDiplomado_Nota20*” possuem uma relação direta com a variável designada por “*NumDiplomados_NoAnoNoParUniversidadeCurso*”, o exemplo exposto na Operação 2, apresenta o cálculo feito na nova coluna “*NumDiplomados_Nota10*” como forma de tornar a informação da variável denominada por “*%ClassifFinalDiplomado_Nota10*” proveitosa ao estudo.

```
1 NumDiplomados_Nota10 =
2 DIVIDE(
3 | 'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'
4 | [%ClassifFinalDiplomado_Nota10]*'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'
   [NumDiplomados_NoAnoNoParUniversidadeCurso],100
5 )
```

Operação 2 - Transformação da nova coluna “*NumDiplomados_Nota10*”

Por último, uma vez que apenas se tem acesso a dados referentes a cada par Universidade – Curso, surgiu a necessidade de se criar duas medidas na Tabela *T. Factos*, como forma de obter informações generalizadas relativamente às notas médias dos inscritos e dos diplomados. Novas medidas essas que, também através de um conjunto de funções de linguagem DAX, foram conseguidas respetivamente da seguinte forma:

```

1 Nota Média de Ingresso =
2 SUMX(
3     FILTER(
4         'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade',
5         'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'[NotaIngressoMédia]>>0
6     ),
7     [NotaIngressoMédia])/
8 COUNT(
9     'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'[NotaIngressoMédia]
10 )

```

Operação 3 - Medida criada para o cálculo da Nota Média de Ingresso

```

1 Nota Média dos Diplomados =
2 SUMX(
3     FILTER(
4         'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade',
5         'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'
6         [NotaFinal_Media_Diplomados_NoAnoNoParUniversidadeCurso]>>0
7     ),
8     [NotaFinal_Media_Diplomados_NoAnoNoParUniversidadeCurso])/
9 COUNT(
10    'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'
11    [NotaFinal_Media_Diplomados_NoAnoNoParUniversidadeCurso]
12 )

```

Operação 4 - Medida criada para o cálculo da Nota Média dos Diplomados

É ainda de referir que, por se ter constatado que foram disponibilizados novos dados referentes ao último ano letivo em análise, após realizada a primeira versão do conjunto de dados a usar no projeto, decidiu-se adicionar esses novos dados à base de dados final. Desta forma, recorreu-se à opção *Atualizar* do *Microsoft Power BI*, para que os dados carregados, bem como todas os seus elementos visuais a si ligados, fossem atualizados.

4.4 Fase 4 - *Modeling*

De seguida procedeu-se à modelação do conjunto final de dados carregados, isto é, ao cruzamento da informação das diferentes tabelas. Para tal, utilizou-se a opção *Gerir Relações* do *Microsoft Power BI* para criar um modelo relacional que representasse as relações entre as várias tabelas através dos seus atributos em comum.

A vantagem desta ferramenta de gerar automaticamente as relações entre tabelas, permite que o utilizador apenas tenha de verificar se todas as relações existentes foram geradas

automaticamente – podendo criá-las manualmente, caso alguma não tenha sido criada – e se as mesmas são exequíveis.

Exemplificando, e através desta opção criou-se uma relação entre a Tabela *TsubD2 – TipoEnsino* e a Tabela *TD3 – Faculdade*. A cardinalidade entre a Tabela de SubDimensão e a Tabela de Dimensão estabelece uma relação de 1 para muitos (também representada por 1:*) – isto é, cada Tipo de Ensino pode ser lecionado por diversas Faculdades, porém, cada Faculdade apenas leciona mediante um Tipo de Ensino – e a direção de filtro cruzado é definida como única, como apresentado de seguida na Figura 3.

Criar relação

Selecione as tabelas e colunas que estão relacionadas.

TsubD2 - TipoEnsino

TipoEnsino
Universitário
Politécnico

TD3 - Faculdade

| UnidadeOrganica (Faculdade) | LocalizaçãoFaculdade | TipoEnsino |
|---|---|---------------|
| 4002 Academia Nacional Superior de Orquestra | Tv. Galé 36, 1300-472 Lisboa | Politécnico |
| 2710 Atlântica - Escola Universitária de Ciências Empresariais... | Fábrica da Pólvora de, 2730-036 Barcarena | Universitário |
| 1305 Centro de Tecnologias da Saúde | Av. Dom João II Lote 4.69 01, 1990-096 Lisboa | Politécnico |

Cardinalidade: Um para muitos (1:*)

Direção de filtro cruzado: Único

Tornar esta relação ativa

Assumir integridade referencial

Aplicar filtros de segurança em ambas as direções

Figura 3 - Criação das relações do Modelo Relacional (Fonte: Autor)

Com a criação das relações do modelo constatou-se que, com a construção da base de dados no Excel, uma das variáveis (*TipoEnsino*) que devia estar na Tabela *TD3 – Faculdade* se encontrava contida na Tabela de Factos. Desta forma, foi necessário criar uma coluna na TD referida, de forma que esta pudesse ser caracterizada mediante o Tipo de Ensino através da ligação à Tabela *Subdimensão2 – Tipo de ensino* (de forma semelhante como acontecia entre a Tabela *TD2 – Universidade* e a Tabela *Subdimensão 1 – Natureza*). Para tal, recorreu-se a funções de linguagem DAX como pode ser visível de seguida pela Operação 5:

```

1 TipoEnsino =
2 CALCULATE(
3     FIRSTNONBLANK(
4         'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'[TipoEnsino],1
5     ),
6     FILTER(
7         ALL('TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'
8         ),
9         'TFactos - Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade'[idUnidadeOrganica]='TD3 - Faculdade'[idUnidadeOrganica]
10    )
11 )

```

Operação 5 – Transformação da nova coluna da Tabela TD3 - Faculdade

Assim, e visíveis todas as novas relações do modelo, podemos constatar que, de acordo com a modelação multidimensional, o projeto utiliza um Esquema em Floco de Neve – isto é, um Modelo Relacional composto por uma Tabela de Factos (*TFactos – Entradas e Saídas num ano _ num curso _ no par Universidade_Faculdade*), quatro Tabelas de Dimensão (*TD1 – AnoLetivo; TD2 – Universidade; TD3 – Faculdade; TD4 – Curso*) e duas Tabelas de Subdimensão (*TsubD1 – Natureza; TsubD2 – TipoEnsino*) como visível na seguinte Figura 4:

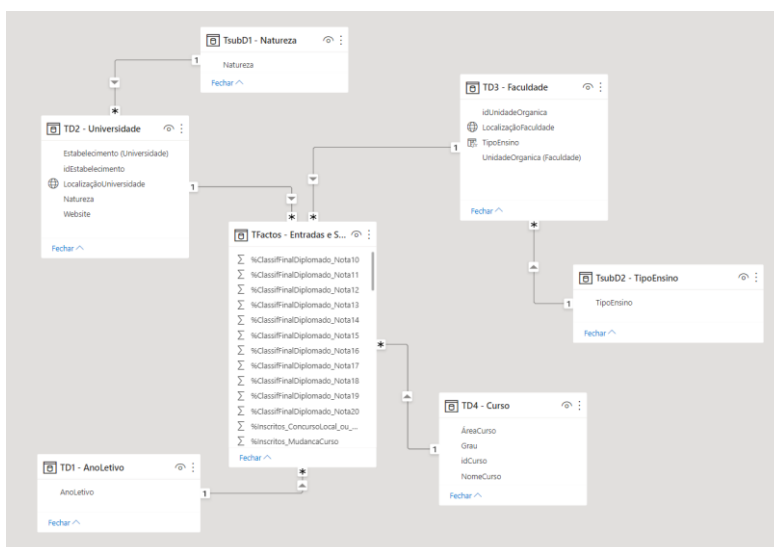


Figura 4 - Modelo Relacional (Fonte: Autor)

De uma forma sucinta, todas as relações consideradas no modelo foram conseguidas através de chaves estrangeiras (atributos comuns entre as tabelas ligadas) que foram representadas na seguinte Tabela IV:

Tabela IV - Tabelas de Dimensão e Subdimensão do Modelo Relacional (Chave Primária e Outros Atributos)

| Tabela de Dimensão (TD) | Chave Primária da TD / Chave Estrangeira da TF | Outros Atributos |
|-------------------------|--|--|
| TD1 – AnoLetivo | <i>AnoLetivo</i> | |
| TD2 – Universidade | <i>idEstabelecimento</i> | Estabelecimento (Universidade) Natureza LocalizaçãoUniversidade Website |
| TD3 – Faculdade | <i>idUnidadeOrganica</i> | UnidadeOrganica (Faculdade) LocalizaçãoFaculdade TipoEnsino |
| TD4 – Curso | <i>idCurso</i> | NomeCurso Grau ÁreaCurso |

| Tabela de Subdimensão (TsubD) | Chave Primária da TsubD / Chave Estrangeira da TD | Outros Atributos |
|-------------------------------|---|------------------|
| TsubD1 – Natureza | <i>Natureza</i> | |
| TsubD2 – TipoEnsino | <i>TipoEnsino</i> | |

Criado o Modelo Relacional, tornou-se possível desenvolver um conjunto de painéis de visualização informativos e interativos, nomeadamente um total de sete *dashboards*, divididos em 5 secções principais: Menu Principal, Análise Geral, Análise das Entradas, Análise das Saídas e Análise Comparativa, que vão ao encontro dos objetivos propostos pelo projeto.

A descrição detalhada de cada um destes *dashboards* e a análise mais minuciosa da informação a retirar de cada um deles será feita no Capítulo 5, porém, de uma forma geral sabe-se que através do primeiro painel (introdutório), pretende-se apresentar o tema em análise e um menu interativo que direciona o utilizador para a análise que pretende (Ver no capítulo seguinte tópico *5.1 Menu Principal*).

Os dois *dashboards* seguintes pretendem fazer uma Análise Geral da constituição e distribuição do Ensino Superior em Portugal (Ver no capítulo seguinte tópico *5.2 Análise Geral*). Assim, no primeiro painel, é possível compreender a interligação das três principais componentes (Estabelecimentos, Unidades Orgânicas e Cursos), retirar informação sobre a Natureza dos Estabelecimentos de ensino e/ou sobre a quantidade de Unidades Orgânicas de cada tipo de ensino existem, bem como perceber a distribuição dos cursos por grau e por área. E, o segundo painel, de forma a complementar a análise, disponibiliza a distribuição dos Estabelecimentos e respetivas Unidades Orgânicas que constituem atualmente o Ensino Superior em Portugal.

Os *dashboards* relativos à Análise das Entradas e à Análise das Saídas apresentam o conjunto de fatores que caracterizam os alunos inscritos e os alunos diplomados, respetivamente, num determinado ano letivo mediante diferentes fatores (Ver no capítulo seguinte tópico *5.3 Análise das Entradas* e *5.4 Análise das Saídas*, respetivamente).

Por último, foi ainda executado um painel que permite realizar uma Análise Comparativa dos dados referentes às entradas e saídas do Ensino Superior nos últimos anos letivos, onde são apresentados o número de alunos (inscritos e diplomados) e as suas notas (entrada e saída), o que permite ao utilizador fazer uma análise evolutiva do comportamento destas variáveis ao longo dos últimos anos, assim como fazer uma análise

meticulosa por Áreas de Curso, caso o deseje (Ver no capítulo seguinte tópico 5.5 *Análise Comparativa*).

4.5 Fase 5 - *Evaluation*

Uma vez que nesta fase de avaliação se deve ter em consideração a usabilidade do projeto (Jooste et al., 2013), procedeu-se a uma apreciação da qualidade e eficácia do modelo desenvolvido, começando-se por verificar se os resultados obtidos permitem alcançar os objetivos propostos.

Para tal, nesta avaliação é crucial ter em consideração: a **experiência do utilizador final**, ou seja, perceber se o projeto satisfaz as necessidades do utilizador e permite auxiliar as suas futuras tomadas de decisões estratégicas; a **acessibilidade das funcionalidades dos dashboards** desenvolvidos no projeto, isto é, ajuizar a performance dos *dashboards* tendo em conta a sua visualização apelativa e interativa, a compreensão da informação apresentada, a sua possibilidade de agrupar diferentes informações num só painel, a diversidade dos elementos visuais e as suas características *user friendly*, entre outros fatores; bem como **identificar possíveis problemas** do projeto. Foi nesse sentido que a avaliação do projeto teve em consideração a opinião de uma especialista na área do BI.

Após a apreciação realizada pela Subdiretora de Unidade Orgânica da Universidade Nova de Lisboa, tomou-se conhecimento de que, apesar do projeto desenvolvido consistir numa boa ferramenta para alcançar os objetivos propostos tem ainda alguns aspetos que, quando considerados e melhorados, podem potenciar futuramente o desempenho do mesmo.

Assim, com base nas observações e conselhos da referida especialista, realizou-se uma análise SWOT referente ao projeto. Este método de análise de projetos, ao ter em consideração uma perspetiva atual e futura, permite identificar, respetivamente, os principais fatores fortes e fracos do projeto – isto é, indicadores de força (*Strengths*) e de fraqueza (*Weaknesses*) – assim com as suas principais oportunidades (*Opportunities*) e ameaças (*Threats*) (Benzaghta et al., 2021; Namugenyi et al., 2019; Phadermrod et al., 2019). Estes indicadores podem ser melhor visualizáveis na seguinte Tabela V:

Tabela V - Análise SWOT do projeto desenvolvido

| | |
|--|---|
| <p style="text-align: center;">FORÇAS</p> <ul style="list-style-type: none"> → Permite uma boa perceção da distribuição geográfica das IES; → Permite construir <i>dashboards</i> com informação mais detalhada; → Permite ter uma boa perceção da evolução do número de alunos e médias por entradas e saídas do Ensino Superior; → Permite perceber a distribuição dos alunos de forma proporcional por área. | <p style="text-align: center;">FRAQUEZAS</p> <ul style="list-style-type: none"> → Considerar dados referentes à mobilidade dos alunos nacionais; → Considerar dados referentes às origens dos estudantes internacionais; → Melhorar a usabilidade do projeto (nomeadamente, a nível de botões). |
| <p style="text-align: center;">OPORTUNIDADES</p> <ul style="list-style-type: none"> → Construção de novos indicadores e elementos visuais que permitam a tomada de decisão; → O projeto desenvolvido, mediante salvas alterações, pode ser adequado a outros setores educacionais. | <p style="text-align: center;">AMEAÇAS</p> <ul style="list-style-type: none"> → Podem surgir novas ferramentas que contruam para melhores resultados; → Limitação da disponibilização do projeto mediante as licenças dos utilizadores. |

4.6 Fase 6 - Deployment

O projeto implementado está pronto a ser utilizado por qualquer Instituição que assim o pretenda. Porém, e caso futuramente alguma Instituição recorra ao mesmo, deverá ser tido em consideração dois aspetos: por um lado, que se trata de um modelo iterativo, pelo que é previsível que seja necessário retroceder a fases anteriores do ciclo metodológico e, por outro lado, dado que o projeto desenvolvido recorreu à ferramenta *Microsoft Power BI Desktop*, deve-se estar ciente de que apenas é possível disponibilizar os *dashboards* realizados e a respetiva visualização interativa entre utilizadores que tenham a licença *Power BI Pro ou Premium* (Microsoft, 2022).

CAPÍTULO 5: RESULTADOS DO PROJETO

Construído o modelo de dados atrás exposto, foi criado um relatório através da ferramenta *Microsoft Power BI*. Este relatório é constituído por 7 *dashboards* divididos em 5 secções principais, que tem por base um conjunto coerente de dados, onde a informação contida nas diferentes páginas se completa como um todo. Assim, neste capítulo serão apresentados e analisados, de forma minuciosa, todos os painéis desenvolvidos.

5.1 Menu Principal

- ***Dashboard 1: Página Inicial do Projeto***



Figura 5 - Página Inicial do Projeto (Fonte: Autor)

A página inicial do projeto destina-se essencialmente a apresentar e introduzir o tema do mesmo. Através de uma ferramenta denominada por *Marcadores*, esta página conta com a interatividade de um conjunto de botões previamente definidos, que funcionam como um Menu. Este tem como finalidade proporcionar ao utilizar uma forma fácil e acessível de aceder diretamente aos *dashboards* relativos à análise selecionada.

Assim, caso o utilizador recorra a um destes botões será sempre levado para o primeiro *dashboard* correspondente à análise escolhida. Uma vez em qualquer painel escolhido, o utilizador tem sempre a possibilidade de regressar Menu principal, através o *Botão Casa*.

5.2 Análise Geral

- ***Dashboard 2: Estabelecimentos – Unidades Orgânicas – Cursos***

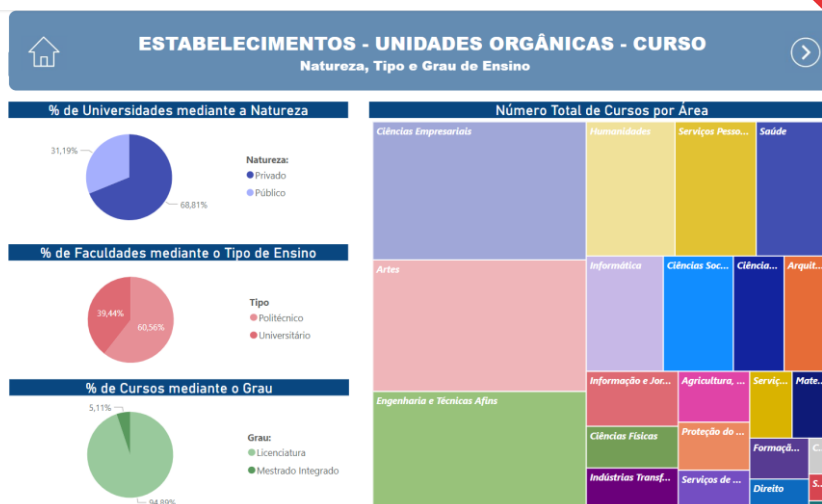


Figura 6 - Análise Geral dos Estabelecimentos – Unidades Orgânicas – Cursos (Fonte: Autor)

A Figura 6 representa o primeiro painel (de dois) desenvolvido para Análise Geral e permite dar a conhecer o panorama geral do Ensino Superior em Portugal.

Através de 3 *Gráficos Circulares* foi possível perceber a «fatia» percentual que representa a Natureza de Ensino das Universidades, o Tipo de Ensino que as Faculdades lecionam bem como o Grau dos vários Cursos em análise, respetivamente. Com a legenda de cada um destes gráficos é possível complementar a informação do valor percentual apurado em cada uma das componentes analisadas através da sua equivalência num número inteiro.

Como forma de melhor transmitir o número de cursos lecionados por área, recorreu-se a um elemento visual denominado por *Treemap*, através do qual é possível agrupar as diferentes áreas em retângulos, cuja dimensão depende do número de cursos lecionados por área. Assim, os retângulos encontram-se dispostos mediante o seu tamanho da parte superior esquerda (áreas onde são lecionados um maior número de cursos) para a parte inferior direita (áreas onde são lecionados um menor número de cursos).

Deste *dashboard* retém-se que o Ensino Superior conta com uma maior percentagem de Universidades de Natureza Privada, ao apresentar um total de 68,81% que representa 75 Universidades Privada. Relativamente às Faculdades, apura-se que a «fatia» maior é dos Politécnicos ao contar com um valor percentual de 60,56%, isto é, 172 Faculdades que lecionam mediante este Tipo de Ensino. Quanto à oferta dos Cursos, é irrefutável que a sua maioria se destina a Cursos apenas de 1º Ciclo (Licenciatura) uma vez que os Cursos de Mestrado Integrado apenas representam 5,11%. Dentro dos Cursos, é de destacar as

áreas das Ciências Empresariais, Artes e Engenharia e Técnicas Afins dado que são as 3 áreas que representam o maior número de Cursos.

Por fim, referir ainda que este *dashboard* conta com um botão no seu canto superior direito que, ao dirigir o utilizador para o próximo *dashboard*, permite dar continuidade à Análise Geral.

- **Dashboard 3: Estabelecimentos – Unidades Orgânicas**



Figura 7 - Análise Geral dos Estabelecimentos – Unidades Orgânicas (Fonte: Autor)

Como forma de aprofundar a Análise Geral, desenvolveu-se o *dashboard* representado na Figura 7. Através deste é possível verificar, por Tipo de Ensino, a disposição geográfica das respetivas Universidades e Faculdades. A cada Estabelecimento e Unidade Orgânica está associado um círculo mediante o Tipo de Ensino lecionado pelas respetivas. Porém, no segundo *Mapa*, a dimensão de cada círculo, representativo de uma Faculdade, varia mediante o número de Cursos que esta leciona.

E assim de fácil perceção que a maior aglomeração de Universidades se encontra em Lisboa e no Porto (através do primeiro *Mapa*). De uma forma semelhante, e dado que uma Universidade conta com várias Faculdades, estas também apesar de se encontrarem mais distribuídas geograficamente, tem uma maior densidade no Centro e Norte de Portugal.

Assim, através do recurso a dois *Mapas*, é possível confirmar a constatação retirada anteriormente de que o Ensino Politécnico é substancialmente mais significativo que o Ensino Universitário.

Caso o utilizador pretenda uma análise mais específica relativamente a uma Universidade ou Faculdade, este painel permite ainda fazer uma *Segmentação de Dados*. Ou seja, basta o utilizador selecionar uma Universidade, para que lhe seja possível determinar geograficamente a sua «sede» e que Faculdades a compõem, bem como as suas disposições geográficas. A operação inversa também é possível, isto é, fazer uma segmentação mediante a seleção de uma Faculdade.

Assim, e como forma de exemplificar (Anexo H), fez-se uma segmentação das Universidades – ao se selecionar a Universidade de Lisboa – como forma de determinar quais as Faculdades que pertencem a esta. Foi assim possível determinar geograficamente a sua «sede» bem como, e através da sua respetiva legenda, algumas características das suas Unidades Orgânicas, como a morada, o Tipo de Ensino que leciona ou até mesmo o seu *website*.

Com esta segmentação tornou-se possível perceber que Faculdades fazem parte da Universidade de Lisboa bem como a sua disposição geográfica. Percebe-se assim que a sua maioria se localiza na Área da Grande Lisboa à exceção do Instituto Superior Técnico (Taguspark) – Unidade Orgânica localizada mais à esquerda no segundo *Mapa*. Dado que a dimensão dos círculos está diretamente associada ao número de Cursos que cada Faculdade leciona, uma vez que o ISEG – Instituto Superior de Economia e Gestão leciona apenas 6 Cursos (como visível na legenda do Anexo H) é das Faculdades com menor dimensão do seu círculo.

5.3 Análise das Entradas

- ***Dashboard 4: Inscritos***

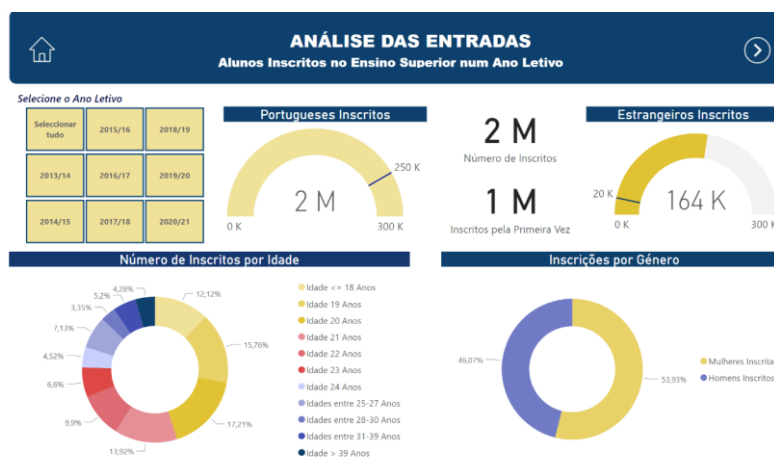


Figura 8 - Análise das Entradas - Inscritos (Fonte: Autor)

Para se fazer a Análise das Entradas realizam-se dois *dashboards* que se complementam. Ambos os painéis contam com elementos de *Segmentação de Dados* previamente sincronizados, que se encontram definidos para realizar inicialmente uma análise de todos os anos (como representado na Figura 8) mas também permitir fazer análises restritas a um Ano Letivo (como feito no Anexo I).

Neste painel, numa primeira instância é possível conhecer o número total de alunos inscritos assim como os que se inscreveram pela primeira vez no Ensino Superior, isto é, que em nenhum dos anos anteriores se tinham candidatado a este nível de ensino. Estas informações são transmitidas através de um elemento visual denominado por *Cartão*. Como forma de complementar a informação do número total de alunos inscritos, recorreu-se ainda a dois *Medidores* que permitem analisar o número de alunos Portugueses e Estrangeiros inscritos neste nível de ensino. Caso estes *Medidores* registassem um total de alunos portugueses/estrangeiros no valor total de 250 mil/20 mil, respetivamente, estipulou-se que seriam um bom indicador para o Sistema de Ensino Português. Adicionalmente, recorreu-se a dois *Gráficos em Anel*, para se fazer uma análise do número de inscritos por idades e por sexo.

Para exemplificar a utilidade deste painel fez-se uma segmentação para o ano de 2020/21 (Anexo I). Desta segmentação verificou-se que dos 256 mil total de alunos inscritos houve 146 mil alunos que se inscreveram no Ensino Superior pela primeira vez. Do total de alunos inscritos, verificou-se que houve 229 mil portugueses a subterem a sua inscrição (não atingido a meta estabelecido dos 250 mil alunos) enquanto 27 mil alunos estrangeiros o fizeram, superando assim a meta dos 20 mil alunos estipulada. Relativamente ao sexo e idade dos alunos, percebe-se que a maior percentagem de inscritos, por um lado, são mulheres (ao apresentar uma percentagem de 56,09%), e por outro, a maioria dos alunos encontra-se na faixa etária dos 19 e 20 anos.

Este painel conta ainda com um botão do canto superior direito que permite ao utilizador passar para o próximo painel desta análise.

• **Dashboard 5: Matriculados**

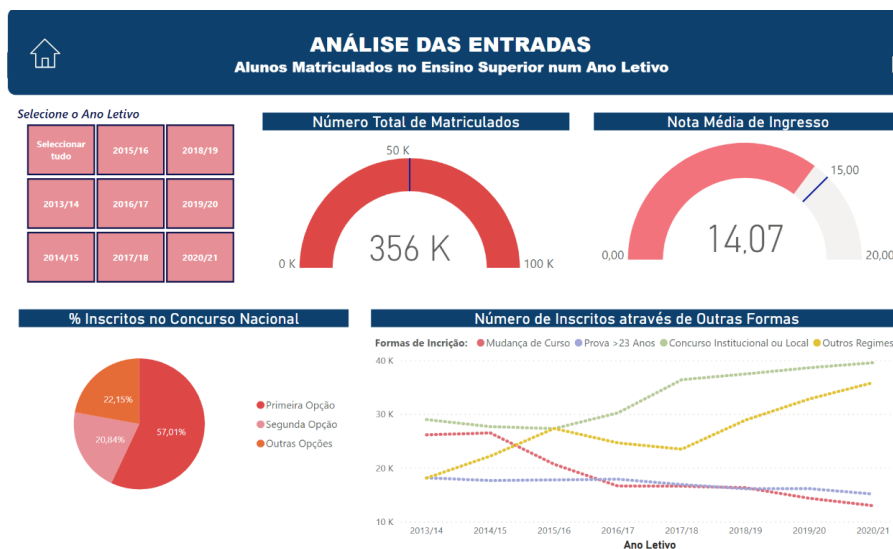


Figura 9 - Análise das Entradas - Matriculados (Fonte: Autor)

No seguimento da análise feita anteriormente, este painel, através de um *Gráfico Circular* e um *Gráfico de Linhas*, permite conhecer, respetivamente, a percentagem e o número de alunos inscritos através de Concurso Nacional ou por Outras Formas.

É possível verificar que a maioria dos alunos quando recorrem ao Concurso Nacional conseguem entrar na sua 1º Opção enquanto que, quando recorrem às Outras Formas de Inscrição, percebe-se, por um lado, que as Mudanças de Curso têm sido cada vez menos recorrentes ao longo dos anos, mas, por outro lado, inscrições por Concurso Institucional/Local e através de Outros Regimes se têm destacado ao longo dos últimos anos letivos.

Todavia, e visto que nem todos os inscritos se chegam a matricular, recorreu-se a dois *Medidores* como forma de observar o número total de matriculados bem com a nota média de ingresso. Estipulou-se que, para estes indicadores representarem uma boa performance, tinham de passar os 50 mil alunos inscritos e a sua média de ingresso devia de superar os 15 valores. Assim, e à semelhança do que se fez anteriormente, tendo em vista os valores apresentados quando feitas outras segmentações, constatou-se que os valores apresentados para o ano letivo de 2020/21 são os que apresentam melhores indicadores: o número total de alunos matriculados no ano atingiu quase a meta estabelecida e a nota média de ingresso ultrapassa a meta estabelecida dos 15 valores (sendo o único ano letivo em estudo que consegue atingir esta meta).

5.3 Análise das Saídas

- **Dashboard 6: Diplomados**

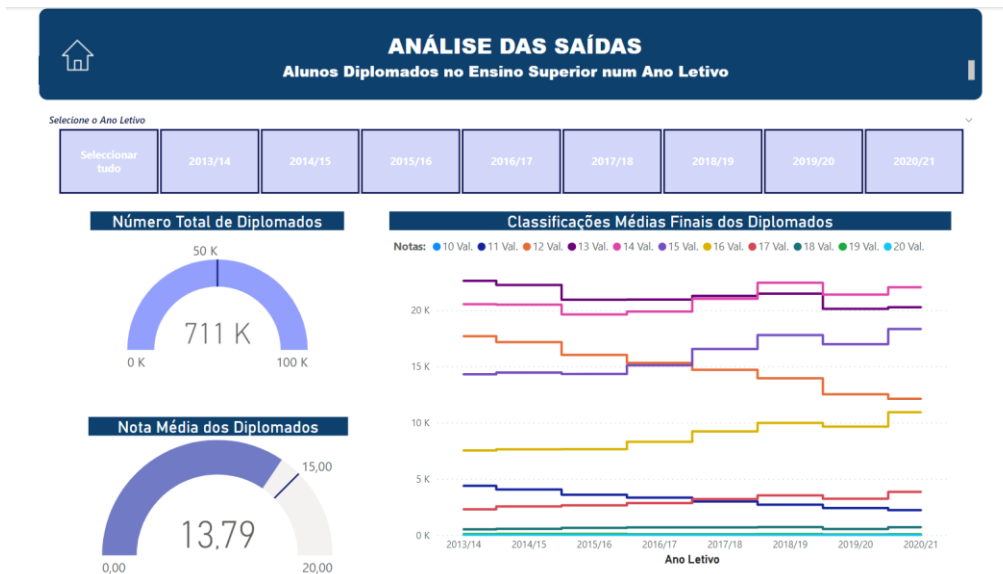


Figura 10 - Análise das Saídas - Diplomados (Fonte: Autor)

Analisadas as entradas, torna-se pertinente realizar a análise referente as saídas dos alunos do Ensino Superior quando estes se tornam diplomados. Para tal, realizou-se o *dashboard* apresentado na Figura 10.

Seguindo a mesma lógica de ideia dos painéis anteriores, os dois primeiros elementos visuais (*Medidores*) permitem analisar o número total de diplomados e a nota média de conclusão. Estipulou-se que ambos os medidores iam ter em consideração os indicadores definidos na análise anterior como forma de se conseguir fazer, posteriormente, uma análise comparativa dos mesmos. Neste sentido, o número total de alunos diplomados e a sua nota média devem assumir, respetivamente, valores de 50 mil alunos e de 15 valores para se considerar que houve um bom desempenho no sistema de ensino num determinado ano letivo.

Neste painel também é possível fazer uma *Segmentação dos Dados* mediante o ano letivo desejado. Assim, e analisados todos os anos de uma forma geral, constatou-se que, apesar do número total de diplomados ultrapassar sempre o limite dos 50 mil alunos, a nota média de conclusão dos alunos não ultrapassa em nenhum dos anos os 15 valores, apresentando assim uma média geral de notas de conclusão no valor de 13,79 valores.

De forma a complementar a informação dos diplomados, torna-se sugestivo perceber qual a tendência e evolução das classificações médias finais dos diplomados. Assim, e através de um *Gráfico de Linhas*, foi possível perceber o «comportamento» que estas classificações têm assumido ao longo dos últimos anos. Observou-se que classificações de 20 valores nunca foram atingidas e que classificações de 10, 18 e 19 valores assumem uma «posição» pouco representativa. As classificações que assumem uma maior tendência, de uma forma geral em todos os anos em análises, são as classificações de 13 e 14 valores. Porém, no ano letivo de 2017/18, as classificações de 14 valores assumem um papel preponderante em relação as de 13 valores.

5.4 Análise Comparativa

- **Dashboard 7: Área Curso**

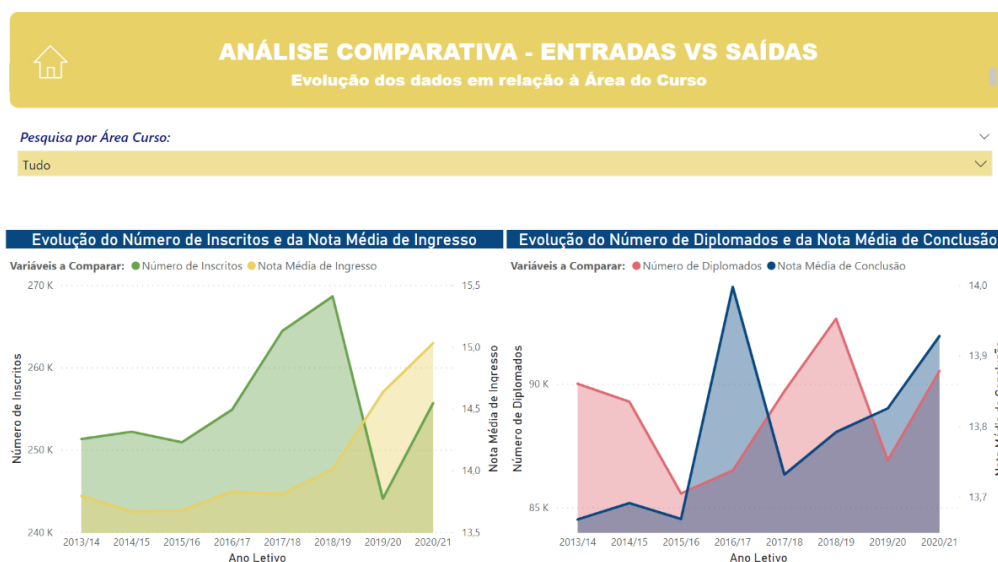


Figura 11 - Análise Comparativa por Área de Curso (Fonte: Autor)

A última análise proposta pelo projeto refere-se à Análise Comparativa. Esta análise permite comparar os principais indicadores das entradas e saídas do Ensino Superior, nomeadamente, o número de alunos (que entra e sai deste nível de ensino) e as notas médias (de entrada e saída) ao longo dos últimos anos letivos.

De forma a fazer a análise desejada recorreu-se a dois *Gráficos de Áreas* que analisassem os indicadores anteriormente referidos tendo em conta as entradas e saídas. Através destes 2 elementos visuais percebe-se que tanto o número de alunos inscritos como a nota média de ingresso, quando comparadas com as respetivas no gráfico das saídas, apresentam sempre valores muito superiores.

Relativamente ao primeiro gráfico, é visível que houve um crescimento significativo ao longo dos anos das notas médias de ingresso, estando estas, ano após ano, a assumir valores mais elevados. Quanto ao número de inscritos, estes destacam-se pelo seu maior pico em 2018/19 consecutivo de um retrocesso no ano seguinte, onde apresenta o valor mais baixo dos anos em análise. Este comportamento foi semelhante quando visto do prisma dos diplomados. Apesar de em 2018/19 se ter verificado o maior número de indivíduos a concluir este nível de ensino, no ano seguinte também se verificou uma grande descida no número total de diplomados.

Relativamente às notas médias de conclusão, estas destacam-se num ano em que o número de alunos diplomados não foi muito significativo, contudo, foi em 2016/17 que tomaram uma posição de «liderança» quando comparadas com outros anos. No entanto, e após se ter verificado uma grande queda deste indicador, desde então que o mesmo tem vindo a crescer nos últimos anos.

Como forma de maximizar a utilidade deste *dashboard* inseriu-se ainda um elemento visual que permitisse fazer uma *Segmentação de Dados*. Assim sendo, este painel permite ainda fazer uma análise mediante a segmentação da Área dos Cursos desejada. Como forma de o exemplificar, fez-se uma segmentação dos Cursos na área das ciências empresariais como pode ser visível no Anexo J.

CAPÍTULO 6: CONCLUSÃO, LIMITAÇÕES E TRABALHO FUTURO

Ao seguir as etapas propostas pela metodologia CRISP-DM e tendo como suporte os dados recolhidos do *Portal Infocursos* e da DGES (DGES, 2022a; InfoCursos, 2022), foi possível alcançar os objetivos proposto com a realização deste projeto: caracterizar o Ensino Superior em Portugal através de uma solução de BI.

Assim, através das funcionalidades da ferramenta *Microsoft Power BI Desktop* e de técnicas de *Data Visualization*, foi possível transformar e estruturar os dados previamente recolhidos, criar um modelo multidimensional assim como, posteriormente, um conjunto painéis, compostos por vários elementos visuais apelativos e de interpretação acessível, que permitissem alcançar o objetivo proposto.

Desta forma, e através da realização de diversos *dashboards* interativos e intuitivos, foi possível contribuir com um conjunto de informação relevante que auxiliasse a tomada de decisão de diferentes IES. Com o acesso a painéis que caracterizam o Ensino Superior em Portugal, permite que estas Instituições consigam entender o ambiente educacional em que se inserem e, conseqüentemente, tenham a capacidade de auxiliar e fundamentar a sua tomada de decisão em informações que as possibilitam ser mais estratégias e eficientes no seu setor.

O projeto foi alvo de uma avaliação por uma especialista na área do BI pelo que, desta apreciação se concluiu que o projeto apesar de poder ser melhorado em alguns aspetos, destaca-se segundo um conjunto de aspetos positivos.

O projeto desenvolvido encontra-se apto a ser utilizado por qualquer Instituição que assim o deseje e caracteriza-se pela sua versatilidade uma vez que, para além de poder ser enquadrado para diferentes IES também pode, mediante salvas alterações, ser aplicável a outros ciclos de ensino caso a base de dados esteja devidamente enquadrada e atualizada.

Todavia, algumas limitações foram sentidas ao longo do desenvolvimento do projeto. O principal entrave sentido esteve relacionado com os dados utilizados. Primeiramente, e quando feita a pesquisa de base de dados relacionadas com o tema, constatou-se que a maioria das bases de dados não tinham acesso livre o que fez com que não fosse possível recolher alguns dados que, numa primeira fase, se tinham considerado como relevantes para executar algumas análises. Exemplo disso, e tendo em conta a temática do estudo, considerou-se que seria interessante perceber a relação entre o número de alunos

matriculados e o número de vagas disponíveis nesse par Curso/Faculdade. No entanto, e dada a indisponibilidade de recolher a informação necessária para a última variável, não foi possível executar a comparação desejada.

Outro dos problemas sentidos com o prosseguimento do trabalho, e apesar de a base de dados ser composta por uma grande quantidade de dados referentes a diversos indicadores, esteve relacionado com a inexistência de informação de alguns campos. A omissão destes dados dificultou o processo, tornando-o mais árduo e trabalhoso, e coloca em causa a qualidade de algumas análises.

Em suma, como sequela destas limitações, o projeto pode ser melhorado futuramente. Para tal, seria interessante melhorar a usabilidade do projeto (como por exemplo, os botões do painel inicial) e apostar na base de dados do projeto através de fusão com outras bases de dados, da inclusão de novas variáveis ou até mesmo realizar uma colaboração com determinada instituição que fornecesse dados de maior qualidade e com relevância para novas abordagens ao estudo. Neste seguimento, seria interessante melhorar a qualidades dos dados referentes aos diplomados (nacionalidade, género e idade dos diplomados) e das Áreas dos Cursos (médias de entrada e saída por Estabelecimentos/Área Curso).

REFERÊNCIAS BIBLIOGRÁFICAS

- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/10.1016/j.tele.2019.01.007>
- Almeida, L. S., & Casanova, J. R. (2019). Desenvolvimento psicossocial e sucesso académico no Ensino Superior. Em *Psicologia da Educação: Temas de Aprofundamento Científico para a Educação XXI* (pp. 101–128). Climepsi Editores.
- Aparicio, S., Aparicio, J. T., & Costa, C. J. (2019). Data Science and AI: Trends Analysis. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), 1-6. IEEE. <https://doi.org/10.23919/CISTI.2019.8760820>
- Atak, M., & Erturgut, R. (2010). Importance of educated human resources in the information age and view of information society organizations on human. *Procedia - Social and Behavioral Sciences*, 2(2), 1452–1456. <https://doi.org/10.1016/j.sbspro.2010.03.217>
- Aws, A. L., Ping, T. A., & Al-Okaily, M. (2021). Towards Business Intelligence Success Measurement in an Organization: A Conceptual Study. *Journal of System and Management Sciences*, 11(2), 155–170. <https://doi.org/10.33168/JSMS.2021.0210>
- Azad, P., Navimipour, N. J., Rahmani, A. M., & Sharifi, A. (2020). The role of structured and unstructured data managing mechanisms in the Internet of things. *Cluster Computing*, 23(2), 1185–1198. <https://doi.org/10.1007/s10586-019-02986-2>
- Benzaghta, M. A., Elwalda, A., Mousa, M., Erkan, I., & Rahman, M. (2021). SWOT analysis applications: An integrative literature review. *Journal of Global Business Insights*, 6(1), 55–73. <https://doi.org/10.5038/2640-6489.6.1.1148>
- Boisot, M., & Canals, A. (2004). Data, information and knowledge: Have we got it right? *Journal of Evolutionary Economics*, 14(1), 43–67. <https://doi.org/10.1007/s00191-003-0181-9>
- Bugwandeen, K., & Ungerer, M. (2019). Exploring the design of performance dashboards in relation to achieving organisational strategic goals. *South African Journal of Industrial Engineering*, 30(2), 161–175. <https://doi.org/10.7166/30-2-2021>
- Burstein, F., & Holsapple, C. W. (2008). *Handbook on decision support systems 2: Variations* (Eds.). Springer Science & Business Media.
- Butler-Henderson, K., Crawford, J., Rudolph, J., Lalani, K., & Sabu, K. M. (2020). COVID-19 in Higher Education Literature Database (CHELD V1): An open access systematic literature review database with coding rules. *Journal of Applied Learning & Teaching*, 3(2), Art. 2. <https://doi.org/10.37074/jalt.2020.3.2.11>
- Caetano, T. V., & Costa, C. J. (2014). Data Warehousing num contexto de Sistemas Integrados. *Atas Da Conferência Da Associação Portuguesa de Sistemas de Informação, Vol. 12*, Art. 0. <http://dx.doi.org/10.18803/capsi.v12.%25p>
- Caldeira, C. P., & Lopes, A. B. P. (2018). Aplicação de Técnicas de Business Intelligence a Base de Dados Prosopográficas. *15th International Conference on Information Systems and Technology Management*, 4967–4979. <https://doi.org/10.5748/9788599693148-15CONTECSI/PS-5983>
- Carnoy, M., Froumin, I., Loyalka, P. K., & Tilak, J. B. G. (2014). The concept of public goods, the state, and higher education finance: A view from the BRICs. *Higher Education*, 68(3), 359–378. <https://doi.org/10.1007/s10734-014-9717-1>
- Castells, M. (2010). *The rise of the network society* (2nd edition). Wiley-Blackwell.

- Cerdeira, L., & Cabrito, B. (2018). Democratização e acessibilidade no ensino superior em Portugal: Mudanças recentes. *Acta Scientiarum. Education*, 40(1), Art. 1. <https://doi.org/10.4025/actascieduc.v40i1.40632>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS inc*, 9(13), 1–73.
- Costa, C. J., & Aparicio, J. T. (2020). POST-DS: A Methodology to Boost Data Science. *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, 1-6. IEEE. <https://doi.org/10.23919/CISTI49556.2020.9140932>
- Costa, C.J., Aparicio, J.T. (2021). A Methodology to Boost Data Science in the Context of COVID-19. In: , et al. *Advances in Parallel & Distributed Processing, and Applications. Transactions on Computational Science and Computational Intelligence*. Springer, Cham. https://doi.org/10.1007/978-3-030-69984-0_7
- Costa, C. J., & Aparício, M. (2019). Supporting the decision on dashboard design charts. *In Proceedings of 254th The IIER International Conference 2019*, 10–15.
- Coutinho, C. P., & Lisbôa, E. S. (2011). Sociedade da informação, do conhecimento e da aprendizagem: Desafios para educação no século XXI. *Revista Educação*, XVIII(1), 5–18.
- Cruz, J. M. D. O. (2008). Processo de ensino-aprendizagem na sociedade da informação. *Educação & Sociedade*, 29(105), 1023–1042. <https://doi.org/10.1590/S0101-73302008000400005>
- Davies, R. (2008). The Bologna process: The quiet revolution in nursing higher education. *Nurse Education Today*, 28(8), 935–942. <https://doi.org/10.1016/j.nedt.2008.05.008>
- Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). Data integration flows for business intelligence. *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, 1–11. <https://doi.org/10.1145/1516360.1516362>
- Decreto-Lei N.º 74/2006 do Ministério da Ciência, Tecnologia e Ensino Superior, § Diário da República - I Série-A, N.º 60 (2006).
- Decreto-Lei N.º 296-A/98 do Ministério da Educação, § Diário da República n.º 222/1998, 1º Suplemento, I-A Série (1998).
- DGES. (2022a). *Acesso ao Ensino Superior 2022—Índices de Cursos (por área de estudos e curso)*. <https://www.dges.gov.pt/guias/indarea.asp>
- DGES. (2022b). *Graus e diplomas do ensino superior*. <https://www.dges.gov.pt/pt/pagina/graus-e-diplomas-do-ensino-superior>
- DGES. (2022c). *Sistema de Ensino Superior Português*. <https://www.dges.gov.pt/pt/pagina/sistema-de-ensino-superior-portugues>
- Diogo, S., Queirós, A., Carvalho, T., Manatos, M., & Soares, D. (2019). 20 Years Of The Bologna Declaration – A Literature Review On The Globalisation Of Higher Education Reforms. *Proceedings of EDULEARN19 Conference*, 5196–5204. <https://doi.org/10.21125/edulearn.2019.1283>
- Drake, B. M., & Walz, A. (2018). Evolving Business Intelligence and Data Analytics in Higher Education. *New Directions for Institutional Research*, 2018(178), 39–52. <https://doi.org/10.1002/ir.20266>
- Du, X., Liu, B., & Zhang, J. (2019). Application of Business Intelligence Based on Big Data in E-commerce Data Analysis. *Journal of Physics: Conference Series*, 1395, 1-6. IOP Publishing. <https://doi.org/10.1088/1742-6596/1395/1/012011>

- Duan, L., & Xu, L. D. (2012). Business Intelligence for Enterprise Systems: A Survey. *IEEE Transactions on Industrial Informatics*, 8(3), 679–687. <https://doi.org/10.1109/TII.2012.2188804>
- Duarte, A & Costa, C. (2012). Information systems: life cycle and success. In Proceedings of the Workshop on Information Systems and Design of Communication (ISDOC '12). Association for Computing Machinery, New York, NY, USA, 25–30. <https://doi.org/10.1145/2311917.2311923>
- Dziewanowska, K. (2017). Value types in higher education – students’ perspective. *Journal of Higher Education Policy and Management*, 39(3), 235–246. <https://doi.org/10.1080/1360080X.2017.1299981>
- Eberendu, A. C. (2016). Unstructured Data: An overview of the data of Big Data. *International Journal of Computer Trends and Technology*, 38(1), 46–50. <https://doi.org/10.14445/22312803/IJCTT-V38P109>
- Eckerson, W. W. (2010). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business* (Second Edition). John Wiley & Sons.
- Feitosa, L. R. C., Marinho-Araujo, C. M., & Almeida, L. D. S. (2020). Psychology In Higher Education In Portugal: Performance In Polytechnic Institutes. *Psicologia Em Estudo*, 25. <https://doi.org/10.4025/psicolestud.v25i0.48061>
- Ferrão, M., & Almeida, L. (2019). Differential effect of university entrance score on first-year students’ academic performance in Portugal. *Assessment & Evaluation in Higher Education*, 44(4), 610–622. <https://doi.org/10.1080/02602938.2018.1525602>
- Few, S. (2006). *Information dashboard design: The effective visual communication of data* (First Edition). O’Reilly Media, Inc.
- Few, S., & Edge, P. (2007). Dashboard confusion revisited. *Perceptual Edge*, 1–6.
- Fomunyam, K. G. (2019). Education and the fourth industrial revolution: Challenges and possibilities for engineering education. *International Journal of Mechanical Engineering and Technology (IJMET)*, 10(8), 271–284.
- Gaardboe, R., & Svarre, T. (2018). Business intelligence success factors: A literature review. *Journal of Information Technology Management*, 29(1), 1–15.
- Gameiro, F. L. (2016). *Elites e educação: Itinerários escolares e percursos profissionais. Alentejo, séculos XIX e XX* [Dissertação, Universidade de Évora].
- Garani, G., Chernov, A., Savvas, I., & Butakova, M. (2019). A Data Warehouse Approach for Business Intelligence. 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 70-75. IEEE. <https://doi.org/10.1109/WETICE.2019.00022>
- Gartner. (2022). *Critical Capabilities for Analytics and Business Intelligence Platforms*. Gartner. <https://www.gartner.com/en/documents/401429>
- Guster, D., & Brown, C. G. (2012). The application of business intelligence to higher education: Technical and managerial perspectives. *Journal of Information Technology Management*, 23(2), 42–62.
- Hamidi, F., Meshkat, M., Rezaee, M., & Jafari, M. (2011). Information technology in education. *Procedia Computer Science*, 3, 369–373. <https://doi.org/10.1016/j.procs.2010.12.062>

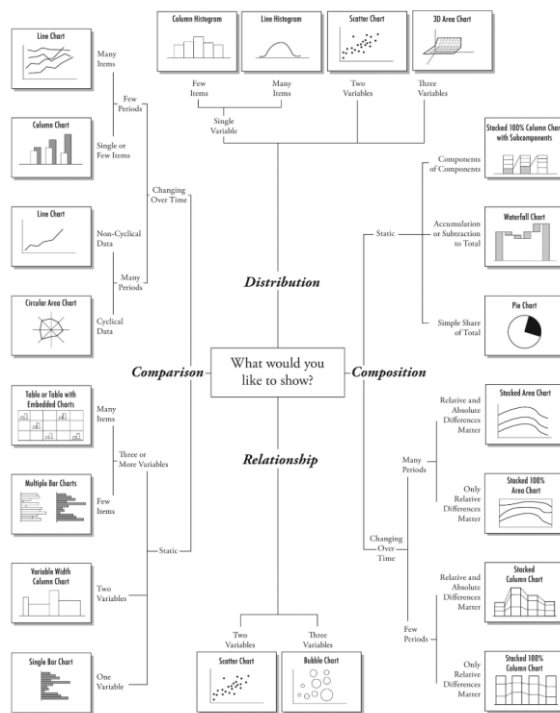
- Hart, R., & Kuo, A. M.-H. (2016). Meeting Health Care Research Needs in a Kimball Integrated Data Warehouse. 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 697-705. IEEE. <https://doi.org/10.1109/DSAA.2016.91>
- Himami, M., Abdullah, A. S., Nurma Yulita, I., & Suryani, M. (2021). Utilization of Data Warehouse in Business Intelligence with Kimball Method at Company XYZ. 2021 International Conference on Artificial Intelligence and Big Data Analytics, 1-6. IEEE. <https://doi.org/10.1109/ICAIBDA53487.2021.9689720>
- InfoCursos. (2022). Dados e Estatísticas de Cursos Superiores. <https://infocursos.medu.pt/>
- Inmon, W. H. (2005). Building the data warehouse (4th edition). Wiley Publishing, Inc.
- Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2020). Applying the CRISP-DM Framework for Teaching Business Analytics. Decision Sciences Journal of Innovative Education, 18(4), 612–634. <https://doi.org/10.1111/dsji.12222>
- Janes, A., Sillitti, A., & Succi, G. (2013). Effective dashboard design. Cutter IT Journal, 26(1), 17–24.
- Jooste, C., van Biljon, J., & Mentz, J. (2013). Usability evaluation guidelines for business intelligence applications. SAICSIT '13: Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, 331–340. <https://doi.org/10.1145/2513456.2513478>
- Joshi, M., & Dubbwar, A. (2021). Review on Business Intelligence, Its Tools and Techniques, and Advantages and Disadvantages. International Journal of Engineering Research, 10(12), 386–391.
- Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). Business Intelligence: An Analysis of the Literature. Information Systems Management, 25(2), 121–131. <https://doi.org/10.1080/10580530801941512>
- Kerzner, H. (2017). Project Management Metrics, KPIs, and Dashboards: A Guide to Measuring and Monitoring Project Performance (Third Edition). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119427599>
- Khan, R. A., & Quadri, S. M. K. (2012). Business intelligence: An integrated approach. Business Intelligence Journal, 5(1), 64–70.
- Khizbullin, F. F., Sologub, T. G., Bulganina, S. V., Lebedeva, T. E., Novikov, V. S., & Prokhorova, V. V. (2017). The direction of transformation of information and communication technology (ict) at the present stage of development into an electronic and information society. Pertanika Journal of Social Sciences and Humanities, 25(S), 45.
- Klemenčič, M. (2019). 20 Years of the Bologna Process in a Global Setting: The external dimension of the Bologna Process revisited. European Journal of Higher Education, 9(1), 2–6. <https://doi.org/10.1080/21568235.2019.1570670>
- Krmac, E. V. (2011). Intelligent Value Chain Networks: Business Intelligence and Other ICT Tools and Technologies in Supply/Demand Chains. Em Supply Chain Management—New Perspectives (Vol. 28, pp. 581–614). IntechOpen. <https://doi.org/10.5772/18850>
- Krneta, D., Jovanovic, V., & Marjanovic, Z. (2016). An Approach to Data Mart Design from a Data Vault. INFOTEH-Jahorina BiH, 15, 473–478.
- Krneta, D., Radosav, D., & Radulovic, B. (2008). Realization business intelligence in commerce using Microsoft Business Intelligence. 2008 6th International Symposium on Intelligent Systems and Informatics, 1-6. IEEE. <https://doi.org/10.1109/SISY.2008.4664943>

- Kruts, K. (2020). Proposta de solução de Business Intelligence no contexto de imigração [MasterThesis, Instituto Superior de Economia e Gestão].
- Laudon, K. C., & Laudon, J. P. (2014). Management information systems: Managing the digital firm (13th Edition). Pearson Education.
- Lei N.º 46/86 de Bases do Sistema Educativo, § Diário da República N.º 237/1986, I Série, II Capítulo - II Secção, III Subsecção (1986).
- Lei N.º 62/2007 do Regime Jurídico das Instituições de Ensino Superior, § Diário da República N.º 174/2007, I Série (2007).
- Lor, P. J., & Britz, J. J. (2007). Is a knowledge society possible without freedom of access to information?. *Journal of Information Science*, 33(4), 387–397. <https://doi.org/10.1177/0165551506075327>
- Luhn, H. P. (1958). A Business Intelligence System. *IBM Journal of Research and Development*, 2(4), 314–319. <https://doi.org/10.1147/rd.24.0314>
- Maleki, R., & Sabet, E. (2022). Business Intelligence Analysis in Small and Medium Enterprises. *International Journal of Innovation in Marketing Elements*, 2(1), 1–11.
- Mandinach, E. B., & Jimerson, J. B. (2016). Teachers learning how to use data: A synthesis of the issues and what is known. *Teaching and Teacher Education*, 60, 452–457. <https://doi.org/10.1016/j.tate.2016.07.009>
- Martins, M. A. J. (2021). Criação de modelo de previsão do Churn na aplicação MBWAY [MasterThesis, Instituto Superior de Economia e Gestão].
- Masino, S., & Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries?. *International Journal of Educational Development*, 48, 53–65. <https://doi.org/10.1016/j.ijedudev.2015.11.012>
- Matukhin, D., & Zhitkova, E. (2015). Implementing Blended Learning Technology in Higher Professional Education. *Procedia - Social and Behavioral Sciences*, 206, 183–188. <https://doi.org/10.1016/j.sbspro.2015.10.051>
- McCowan, T. (2012). Is There A Universal Right To Higher Education?. *British Journal of Educational Studies*, 60(2), 111–128. <https://doi.org/10.1080/00071005.2011.648605>
- Microsoft. (2022). Documentação do Power BI - Power BI. <https://learn.microsoft.com/pt-pt/power-bi/>
- Muntean, M., & Surcel, T. (2013). Agile BI – The Future of BI. *Informatica Economica*, 17(3), 114–124. <https://doi.org/10.12948/issn14531305/17.3.2013.10>
- Namugenyi, C., Nimmagadda, S. L., & Reiners, T. (2019). Design of a SWOT Analysis Model and its Evaluation in Diverse Digital Business Ecosystem Contexts. *Procedia Computer Science*, 159, 1145–1154. <https://doi.org/10.1016/j.procs.2019.09.283>
- Nataliya, P. (2015). Cognitive Management in the Information Society Context. *Procedia - Social and Behavioral Sciences*, 166, 456–459. <https://doi.org/10.1016/j.sbspro.2014.12.553>
- Nidhishree, M. S., & Badar, P. (2019). Study on Web Intelligence in Higher Education: A Review on Business Intelligence Techniques, Frameworks and Usage. *Perspectives in Communication, Embedded-systems and Signal-processing-PiCES*, 3(1), 9–14.
- Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). Organizational business intelligence and decision making using big data analytics. *Information Processing & Management*, 58(6), 102725. <https://doi.org/10.1016/j.ipm.2021.102725>
- Nobre, T., Trigo, A., & Sanches, P. (2014). SBIAES — Business intelligence system for analysis of access to higher education: The case of the Polytechnic Institute of Coimbra. *9th*

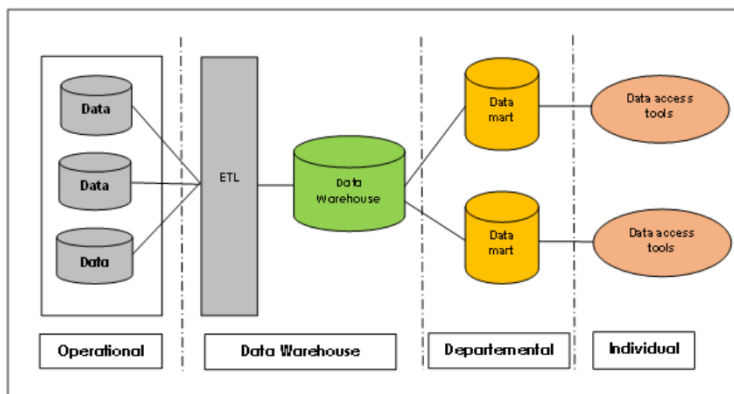
- Iberian Conference on Information Systems and Technologies (CISTI)*, 1-6. IEEE. <https://doi.org/10.1109/CISTI.2014.6876866>
- Nofal, M. I., & Yusof, Z. M. (2013). Integration of Business Intelligence and Enterprise Resource Planning within Organizations. *Procedia Technology*, 11, 658–665. <https://doi.org/10.1016/j.protcy.2013.12.242>
- Nwokeji, J., Aqlan, F., Anugu, A., & Olagunju, A. (2018). Big Data ETL Implementation Approaches: A Systematic Literature Review. In *30th International Conference on Software Engineering and Knowledge Engineering*, 714–715. <https://doi.org/10.18293/SEKE2018-152>
- Obskov, A. V., Pozdeeva, S. I., Matukhin, D. L., & Nizkodubov, G. A. (2015). Educational aspects of interactive foreign language learning in high schools. *Mediterranean Journal of Social Sciences*, 6(4), 319. <https://doi.org/10.5901/mjss.2015.v6n4s3p319>
- ONU. (1948). *Declaração Universal dos Direitos Humanos*. Nações Unidas - ONU Portugal. <https://unric.org/pt/declaracao-universal-dos-direitos-humanos/>
- Parmenter, D. (2015). *Key Performance Indicators—Developing, Implementing, and Using Winning KPIs* (Third Edition). John Wiley & Sons, Inc.
- Phadermrod, B., Crowder, R. M., & Wills, G. B. (2019). Importance-Performance Analysis based SWOT analysis. *International Journal of Information Management*, 44, 194–203. <https://doi.org/10.1016/j.ijinfomgt.2016.03.009>
- Phan, D. D., & Vogel, D. R. (2010). A model of customer relationship management and business intelligence systems for catalogue and online retailers. *Information & Management*, 47(2), 69–77. <https://doi.org/10.1016/j.im.2009.09.001>
- Piedade, M. B., & Santos, M. Y. (2010). Business intelligence in higher education: Enhancing the teaching-learning process with a SRM system. *5th Iberian Conference on Information Systems and Technologies*, 1-5. IEEE.
- Porter, M. E., & Millar, V. E. (1985). How Information Gives You Competitive Advantage. *Harvard Business Review*.
- Rahm, E. (2016). The case for holistic data integration. *East European Conference on Advances in Databases and Information Systems*, 11-27. Springer, Cham.
- Raimundo, R., & Rosário, A. (2021). Blockchain System in the Higher Education. *European Journal of Investigation in Health, Psychology and Education*, 11(1), 276–293. <https://doi.org/10.3390/ejihpe11010021>
- Ranjbarfarid, M., & Hatami, Z. (2020). Critical Success Factors for Implementing Business Intelligence Projects (A BI Implementation Methodology Perspective). *Interdisciplinary Journal of Information, Knowledge, and Management*, 15, 175–202. <https://doi.org/10.28945/4607>
- Ross, M., & Kimball, R. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd Edition). John Wiley & Sons, Inc.
- Santos, M. Y., & Ramos, I. (2006). *Business Intelligence: Tecnologias da informação na gestão de conhecimento*. FCA-Editora de Informática, Lda.
- Sanz, A. C. P. (2018). *Proposta de um dashboard para monitorizar falhas de energia numa rede elétrica inteligente* [MasterThesis, ISCTE - Instituto Universitário de Lisboa].
- Schippling, A., Abrantes, P., & Lopes, J. T. (2020). Educação de elites e a dimensão da internacionalização em Portugal. *Sociologia, Problemas e Práticas*, 94, 119–137.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>

- Seixas, A. M. (2000). O ensino superior privado em Portugal: Políticas e discursos. *Revista Portuguesa de Educação*, 13(2), 53–79.
- Selwyn, N. (2015). Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology*, 40(1), 64–82. <https://doi.org/10.1080/17439884.2014.921628>
- Sharda, R., Delen, D., Turban, E., Aronson, J., & Liang, T.-P. (2015). *Business intelligence and analytics: Systems for decision support* (10th edition). Pearson.
- Simitsis, A., Vassiliadis, P., & Sellis, T. (2005). Optimizing ETL processes in data warehouses. *21st International Conference on Data Engineering (ICDE'05)*, 564-575. IEEE. <https://doi.org/10.1109/ICDE.2005.103>
- Simões, C., & Soares, A. M. (2010). Applying to higher education: Information sources and choice factors. *Studies in Higher Education*, 35(4), 371–389. <https://doi.org/10.1080/03075070903096490>
- Skorka, A. (2017). Successful Dashboard Implementation in Practice: How to Overcome Implementation Barriers and Ensure Long-term Sustainability. *International Journal of Market Research*, 59(2), 239–262. <https://doi.org/10.2501/IJMR-2017-017>
- Souibgui, M., Atigui, F., Zammali, S., Cherfi, S., & Yahia, S. B. (2019). Data quality in ETL process: A preliminary study. *Procedia Computer Science*, 159, 676–687. <https://doi.org/10.1016/j.procs.2019.09.223>
- Talaoui, Y., & Kohtamäki, M. (2020). 35 years of research on business intelligence process: A synthesis of a fragmented literature. *Management Research Review*, 44(5), 677-717. Emerald Publishing Limited. <https://doi.org/10.1108/MRR-07-2020-0386>
- Teixeira, R., Afonso, F., Oliveira, B., Portela, F., & Santos, M. F. (2014). *Business Intelligence to improve the quality of Local Government Services: Case-Study in a Local Government Town Hall*. 1–8.
- Tsai, Y.-S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachler, H., Delgado Kloos, C., & Gašević, D. (2020). Learning analytics in European higher education—Trends and barriers. *Computers & Education*, 155, 103933. <https://doi.org/10.1016/j.compedu.2020.103933>
- Urbano, C. (2011). A (id) entidade do ensino superior politécnico em Portugal: Da Lei de Bases do Sistema Educativo à Declaração de Bolonha. *Sociologia, Problemas e Práticas*, 66, 95–115.
- Vercellis, C. (2009). *Business intelligence: Data mining and optimization for decision making*. John Wiley & Sons.
- Voogt, J., & Knezek, G. (Eds.). (2008). *International handbook of information technology in primary and secondary education* (Vol. 20). Springer Science & Business Media.
- Wang, Y. (2015). Business Intelligence and Analytics Education: Hermeneutic Literature Review and Future Directions in IS Education. *Proceeding of Twenty-First Americas Conference on Information Systems (AMCIS)*, 1–10.
- Watson, H. (2017). Data Visualization, Data Interpreters, and Storytelling. *Business Intelligence Journal*, 22(1), 5–10.
- Williams, S., & Williams, N. (2008). *The Profit Impact of Business Intelligence*. Elsevier Inc.
- Yessad, L., & Labiod, A. (2016). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. *2016 International Conference on System Reliability and Science (ICSRS)*, 95-99. IEEE. <https://doi.org/10.1109/ICSRS.2016.7815845>
- Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1), 41–59. <https://doi.org/10.1016/j.accinf.2011.08.002>

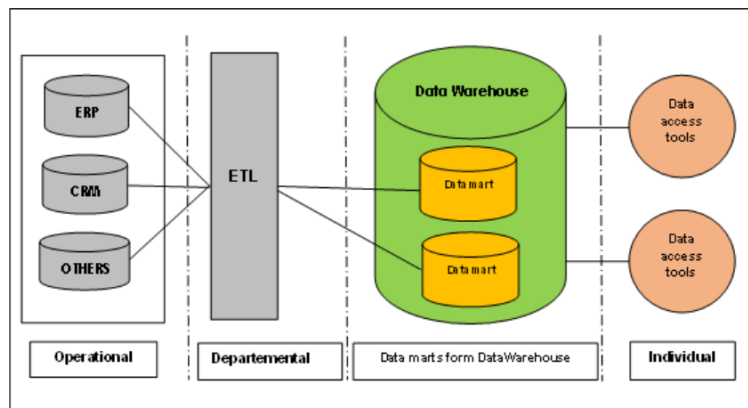
ANEXOS



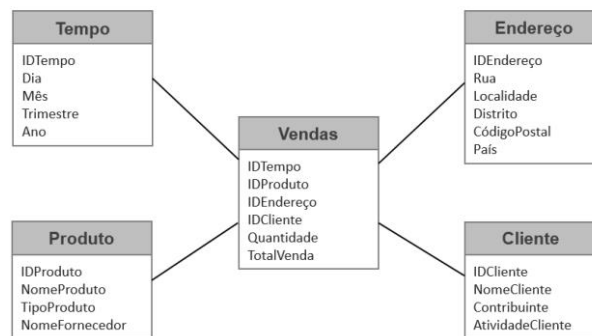
Anexo A - Escolha da Componente Gráfica (Fonte: Abela, 2010)



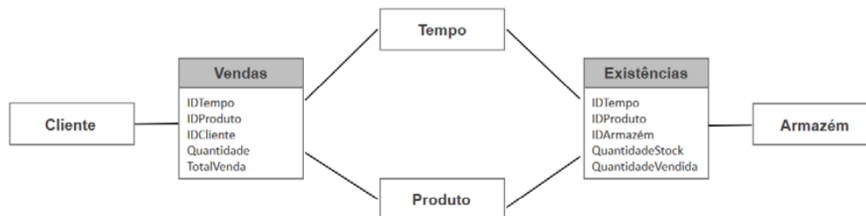
Anexo B - Arquitetura de DW segundo Bill Inmon (Fonte: Yessad & Labiod, 2016)



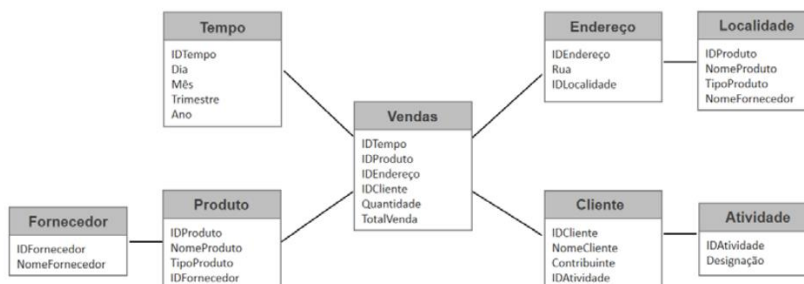
Anexo C - Arquitetura de DW segundo Ralph Kimball (Fonte: Yessad & Labiod, 2016)



Anexos D - Exemplo de um Esquema em Estrela (Fonte: Santos & Ramos, 2006)



Anexos E- Exemplo de um Esquema em Constelação (Fonte: Santos & Ramos, 2006)



Anexos F- Exemplo de Esquema em Floco de Neve (Fonte: Santos & Ramos, 2006)



Anexo G- Quadrante Mágico das Plataformas de Análise e BI (Fonte: Gartner, 2022)

ESTABELECIMENTOS - UNIDADES ORGÂNICAS
Distribuição das Universidades - Faculdades por Tipo de Ensino

Pesquisa por Estabelecimento (Universidade):

Universidade de Lisboa

Pesquisa por Unidade Orgânica (Faculdade):

Tudo

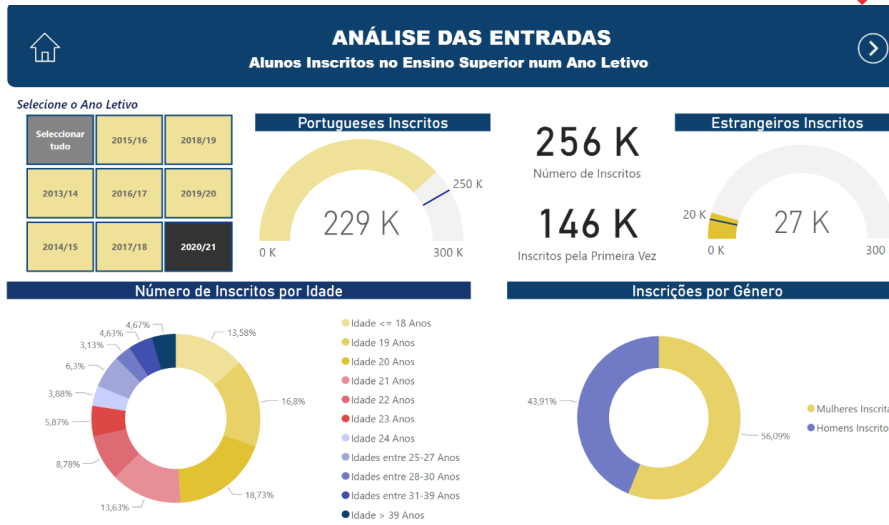
Distribuição das Universidades por Tipo de Ensino

Tipo de Ensino: Universitário

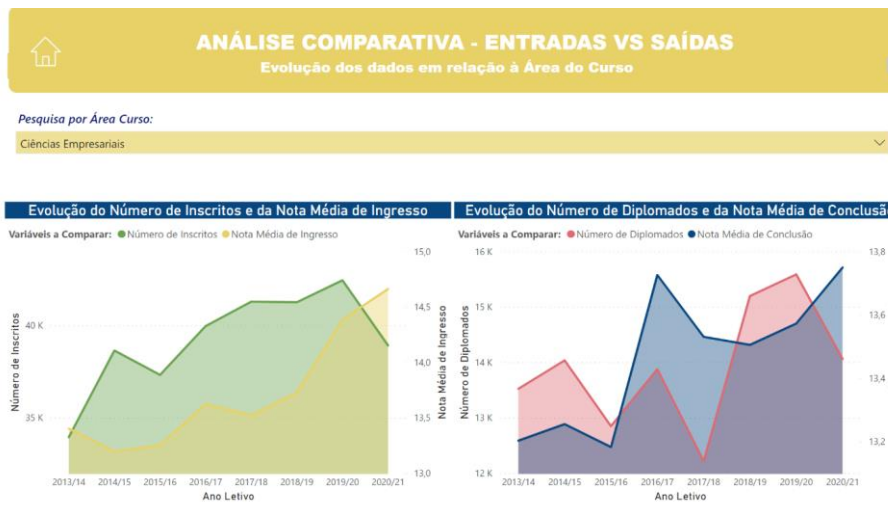
Distribuição das Faculdades por Tipo de Ensino

Tipo de Ensino: Universitário

Anexo H - Aplicação do Filtro "Universidade de Lisboa" na Análise Geral dos Estabelecimentos – Unidades Orgânicas (Fonte: Autor)



Anexo I- Aplicação do Filtro "2020/21" na Análise das Entradas - Inscritos (1) (Fonte: Autor)



Anexo J- Aplicação do Filtro "Ciências Empresariais" na Análise Comparativa por Área de Curso (Fonte: Autor)