INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Análise Descritiva de Dados para Caraterização de Curso Universitário

Bruno Alexandre Baião de Oliveira

Mestrado em Gestão de Sistemas de Informação

Orientadora:
Doutora Ana Maria Carvalho de Almeida, Professora Associadacom Agregação,
Iscte – Instituto Universitário de Lisboa

Orientadora:
Doutora Ana Rita Henrique Peixoto, Professora Auxiliar,
Iscte – Instituto Universitário de Lisboa

Dezembro, 2025

[Esta página foi deixada intencionalmente em branco]

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciência e Tecnologias de Informação

Análise Descritiva de Dados para Caraterização de Curso Universitário

Bruno Alexandre Baião de Oliveira

Mestrado em Gestão de Sistemas de Informação

Orientadora:
Doutora Ana Maria Carvalho de Almeida, Professora Associada com Agregação,
Iscte – Instituto Universitário de Lisboa

Orientadora:
Doutora Ana Rita Henrique Peixoto, Professora Auxiliar,
Iscte – Instituto Universitário de Lisboa

Dezembro, 2025

[Esta página foi deixada intencionalmente em branco]

"O começo é, por definição, um ato de coragem."
Hannah Arendt

[Esta página foi deixada intencionalmente em branco]

Agradecimentos

A conclusão do presente trabalho, assim como, todo o meu sucesso no ensino superior, é dedicado ao maior pilar e amor da minha vida: a minha mulher. Um profundo agradecimento, porque seria impossível conseguir alcançar tudo o que me propus sem o seu amor, carinho, compreensão e apoio incondicional durante todo este percurso. É graças a ela que tenho força para conseguir fazer mil e uma coisas ao mesmo tempo, pelo fruto do seu suporte incansável. A sorte dá muito trabalho. Tudo isto, é por nós, para que o futuro que sonhamos juntos se torne realidade.

Agradeço à minha família, em especial aos meus irmãos e irmã, que são a luz da minha vida. Cada passo que dou, penso em ser uma referência para eles, que lhes mostra que é possível acreditar, persistir e nunca se dar por vencido. Espero que este meu esforço seja um exemplo da importância de lutar, todos os dias, para sermos um pouco melhores do que fomos ontem.

Agradeço ao ISCTE - Instituto Universitário de Lisboa, e aos docentes que me acompanharam neste percurso, pelo acompanhemento, ensinamentos e pela disponibilização dos dados e informações necessários para o desenvolvimento da investigação, bem como pela confiança depositada no meu trabalho.

E por fim, a quem me acompanhou mais de perto, nesta jornada, a Prof. Ana de Almeida e a Prof. Ana Rita Peixoto, que me ajudaram a fechar mais um belo capítulo da minha vida, um enorme obrigado.

[Esta página foi deixada intencionalmente em branco]

Resumo

O presente trabalho incide sobre o desenvolvimento de uma ferramenta de apoio à decisão com base em *dashboards* interativos, aplicados à análise do percurso académico dos estudantes no ensino superior, com o objetivo de ser um instrumento valioso para apoiar decisões estratégicas nas áreas de gestão académica e admissões. A investigação decorreu num contexto real, com dados das Licenciaturas em Informática e Gestão de Empresas e em Ciência de Dados.

Através da metodologia CRISP-DM, foi possível estruturar o projeto em várias fases, desde a compreensão do problema e dos requisitos analíticos definidos pelos docentes, até à preparação rigorosa dos dados e desenvolvimento do *dashboard*. A preparação dos dados incluiu etapas de limpeza, transformação e normalização, bem como a criação de identificadores únicos para cada estudante, possibilitando análises longitudinais e transversais. Adicionalmente, foram desenvolvidas colunas calculadas com *Data Analysis Expressions (DAX)* que permitiram criar métricas personalizadas adaptadas à lógica da análise pretendida.

O *dashboard* final, desenvolvido em Power BI, incorpora um conjunto de visualizações, que permitem explorar indicadores como notas de ingresso, cursos de origem, distribuição geográfica, médias anuais, percentagem de créditos concluídos, e desempenho dos trabalhadores-estudantes. Cada visualização foi desenhada de forma a responder a questões analíticas colocadas pelos docentes, com foco na clareza, segmentação e comparação entre grupos.

Os resultados obtidos permitiram identificar padrões no desempenho dos estudantes, a influência da nota e do curso de origem, o impacto do estatuto de trabalhador-estudante, e a existência de variações ao longo dos anos.

Palavras Chave: *Power BI*; *CRISP-DM*; visualização de dados; desempenho académico; ensino superior; sistema de apoio à decisão

[Esta página foi deixada intencionalmente em branco]

Abstract

This dissertation focuses on the development of a decision support tool based on interactive dashboards applied to the analysis of students' academic pathways in higher education with the aim of being a valuable instrument to support strategic decisions in the areas of academic management and admissions. The research was carried out in a real institutional context, using data from the Bachelor's degrees in Informatics and Business Management and Data Science.

By adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, the project was structured into distinct yet interconnected phases, ranging from the understanding of the problem and analytical requirements defined by faculty members to the preparation of the data and the development of the dashboard. The data preparation process included steps of cleaning, transformation, and normalization, as well as the creation of unique identifiers for each student. Additionally, calculated columns were developed using DAX to create customized metrics tailored to the intended analytical logic.

The final dashboard, developed in Power BI, integrates a comprehensive set of visualizations that allow the exploration of key indicators such as admission grades, previous academic background, geographic distribution, annual averages, percentage of credits completed, and the performance of student-workers. Each visualization was designed to directly address analytical questions raised by the faculty, with a focus on clarity, segmentation, and group comparison.

The results obtained enabled the identification of relevant patterns in students' performance, namely the influence of entry grades and prior academic background, the impact of student-worker status, and variations across academic years.

KEYWORDS: *Power BI*; *CRISP-DM*; *data visualization*; *academic performance*; *higher education*; *decision support system*

[Esta página foi deixada intencionalmente em branco]

Conteúdo

Agradecimentos iii

Resumo v

Abstract vii

Lista de Figuras xi

Lista de Acrónimos xiii

Capítulo 1. Introdução 1
1.1. Contextualização 2
1.2. Questão de Investigação e Objetivos 3
1.3. Principais Contribuições 4
1.4. Estrutura do Documento 5

Capítulo 2. Revisão Sistemática da Literatura 7
2.1. Metodologia da Revisão 7
2.2. Resultados da Revisão Sistemática da Literatura 9
2.2.1. Técnicas de Análise Exploratória e Mineração de Dados 9
2.2.2. Fatores na Caraterização de Estudantes 10
2.2.3. Ferramentas de Visualização e Sistemas de Suporte à Decisão 11
2.2.4. Previsão do Desempenho e Retenção de Estudantes 11
2.2.5. Gestão de Cursos e Otimização dos Programas Educativos 12
2.2.6. Síntese da Revisão da Literatura 12

Capítulo 3. Desenho e Desenvolvimento 15
3.1. Abordagem Geral 15
3.2. Levantamento dos Requisitos dos Docentes 16
3.3. Compreensão dos Dados 18
3.4. Preparação, Tratamento e Modelação dos Dados 19
3.5. Desenvolvimento do *Dashboard* Interativo 22

Capítulo 4. Resultados e Discussão 25
4.1. Visualizações e Resultados Obtidos no *Dashboard* 25
4.1.1. Licenciatura de Informática e Gestão de Empresas 25
4.1.2. Funcionalidades Interativas e Filtros Dinâmicos 38
4.1.3. Utilidade para a Tomada de Decisão 39

4.2. Discussão dos Resultados 41
4.2.1. Relação entre Perfil de Ingresso e Desempenho Académico 41
4.2.2. Impacto do Estatuto de Trabalhador-Estudante 41
4.2.3. Distribuição Geográfica e Diversidade de Perfis 42
4.2.4. Comparação com Estudos de Visualização e Apoio à Decisão e Implicações Estratégicas 43

Capítulo 5. Conclusão 45
5.1. Síntese do Trabalho Desenvolvido 45
5.2. Principais Conclusões 47
5.3. Principais Contribuições do Trabalho 48
5.4. Limitações do Estudo 49
5.5. Trabalho Futuro 50

Bibliografia 53

Apêndice A. Dicionário de Dados 57

Apêndice B. *Dashboard Power BI* - LIGE 65

Apêndice C. *Dashboard Power BI* - LCD 67

Lista de Figuras

Figure 1.1 Os 4 tipos de análise de dados [2] 2

Figure 1.2 Pilares da análise de dados [2] 3

Figure 2.1 Contagem de Artigos ao Longo dos Anos (2011-2024) 8

Figure 3.1 Metodologia e Pipeline da Solução Desenvolvida 17

Figure 4.1 Distribuição dos alunos da LIGE por curso anterior no momento do ingresso 26

Figure 4.2 Nota de ingresso média por curso anterior dos alunos da LIGE 27

Figure 4.3 Distribuição geográfica dos alunos da LIGE por distrito de residência 28

Figure 4.4 Evolução da nota de ingresso média dos alunos da LIGE por ano letivo 29

Figure 4.5 Distribuição dos alunos da LIGE por desempenho e ano letivo 30

Figure 4.6 Relação entre a Média Final do Ano Curricular e a Média de Ingresso por Ano Letivo 32

Figure 4.7 Distribuição dos alunos da Licenciatura em Ciência de Dados (LCD) por curso anterior no momento do ingresso 33

Figure 4.8 Nota de ingresso média por curso anterior dos alunos da LCD 34

Figure 4.9 Distribuição geográfica dos alunos da LCD por distrito de residência 35

Figure 4.10 Evolução da nota de ingresso média dos alunos da LCD por ano letivo 35

Figure 4.11 Distribuição dos alunos da LCD por desempenho e ano letivo 36

Figure 4.12 Relação entre a Média Final do Ano Curricular e a Média de Ingresso por Ano Letivo 38

Figure 4.13 Menu Inicial (HOME) 39

Figure 4.14 Filtro de trabalhador-estudante 39

Figure 4.15 Filtro do Ano letivo 39

Figure B.1 *Dashboard Power BI* - LIGE - Análise de Ingresso 65

Figure B.2 *Dashboard Power BI* - LIGE - Análise de Progresso 65

Figure C.1 *Dashboard Power BI* - LCD - Análise de Ingresso 67

Figure C.2 *Dashboard Power BI* - LCD - Análise de Progresso 67

[Esta página foi deixada intencionalmente em branco]

Lista de Acrónimos

BIRCH: Balanced Iterative Reducing and Clustering using Hierarchies

CRISP-DM: Cross-Industry Standard Process for Data Mining

DAX: Data Analysis Expressions

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

DL: Deep Learning

EDM: Educational Data Mining

ISCTE: ISCTE - Instituto Universitário de Lisboa

LCD: Licenciatura em Ciência de Dados

LIGE: Licenciatura em Informática e Gestão de Empresas

ML: Machine Learning

PCA: Principal Component Analysis

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyse

RGPD: Regulamento Geral sobre a Proteção de Dados

SMOTE: Synthetic Minority Over-sampling Technique

[Esta página foi deixada intencionalmente em branco]

CAPíTULO 1

Introdução

O processo de admissão, no ensino superior, tem apresentado grandes desafios, dada a complexidade e competição intensa entre os candidatos. Como resposta, as universidades têm investido em estratégias para melhorar a admissão de alunos com maior potencial, através de ferramentas tecnológicas e analíticas que apoiam estratégicamente na tomada de decisão, baseada em evidências [1].

A *Data Analytics* tem emergido como uma abordagem essencial, para ajudar as instituições a recolher, processar e interpretar dados sobre os candidatos. Tecnologias como o *Machine Learning (ML)* e *Deep Learning (DL)*, têm vindo a oferecer benefícios na previsão do desempenho académico e profissional, a promoção de uma maior diversidade entre os candidatos admitidos e o fortalecimento dos padrões académicos [2].

O uso de exames padronizados, ou testes de áreas específicas, podem ajudar a medir as competências e as aptidões dos candidatos. Contudo, surgem questões como a justiça, confiabilidade e eficácia em prever o sucesso académico dos estudantes. O que ressalta a importância de considerar outros fatores, tais como, as caraterísticas demográficas e os aspetos socioeconómicos que influenciam as oportunidades e a performance dos alunos [3].

A aplicação de ferramentas como Educational Data Mining (EDM), têm um papel bastante importante. São responsáveis pela extração de informações úteis e padrões de grandes bases de dados. As técnicas preditivas têm demonstrado capacidade para antecipar resultados académicos e identificar os melhores candidatos. Isto é um ponto de destaque, quando as universidades enfrentam a árdua tarefa de equilibrar os critérios de inclusão social, diversificação dos admitidos e manutenção dos altos padrões de qualidade [4].

Embora a análise de dados ofereça soluções inovadoras, também ela está repleta de desafios, tais como, a integridade e qualidade dos dados, como também a necessidade de especialização técnica para processá-los. Além disso, as interpretações inadequadas dos dados podem levar a más decisões, o que reforça a necessidade de combinar métodos quantitativos com avaliações qualitativas e expertise humana [4].

Posto isto, a análise de dados em admissões vai além de uma abordagem técnica. O uso da Data Analytics reflete uma mudança de paradigma que procura alinhar critérios académicos, diversidade e sucesso dos estudantes, promovendo as práticas de admissão mais eficazes, inclusivas e baseadas em evidências. Esta abordagem representa um avanço significativo no uso estratégico de dados para a melhoria contínua do ensino superior.

1.1. Contextualização

Data analytics, é a disciplina que estuda os dados, não tratados, para conseguir retirar conclusões e tomar decisões. O objetivo é dar resposta a questões e poder tomar decisões com estas respostas, tal como, é o caso dos processos de admissão ao ensino superior, onde os dados ajudam a identificar padrões e prever resultados futuros com maior precisão. No contexto dos processos de admissão, a aplicação de técnicas de análise de dados permite que as instituições compreendam melhor os perfis dos candidatos, podendo prever o possível desempenho e ajustar estratégias que garantam a qualidade académica. Existem grandes desafios nos processos de admissão, tais como, a competição elevada, que exige métodos mais eficazes para identificar o potencial do candidato, as desigualdades socioeconómicas, que podem influenciar significativamente o desempenho académico e as oportunidades de acesso ao ensino superior e, os dados insuficientes, que afetam a precisão das análises e previsões [2], [5], [6].

Existem quatro categorias fundamentais, cada uma com aplicações importantes para o tema:

- Análise descritiva - Ajuda a compreender o desempenho passado dos candidatos, ao identificar padrões gerais, como médias de desempenho.
- Análise diagnóstica - Explora relaçóes de causa e efeito, o que ajuda a entender a razão de certos candidatos terem sucesso ou enfrentarem maiores dificuldades.
- Análise preditiva - Usa técnicas avançadas como ML, para projetar o desempenho futuro dos candidatos, o que permite selecionar de forma mais estratégica.
- Análise prescritiva - sugere ações específicas, como ajustes nos critérios de admissão, de forma a garantir maior inclusão e equidade sem comprometer padrões académicos.

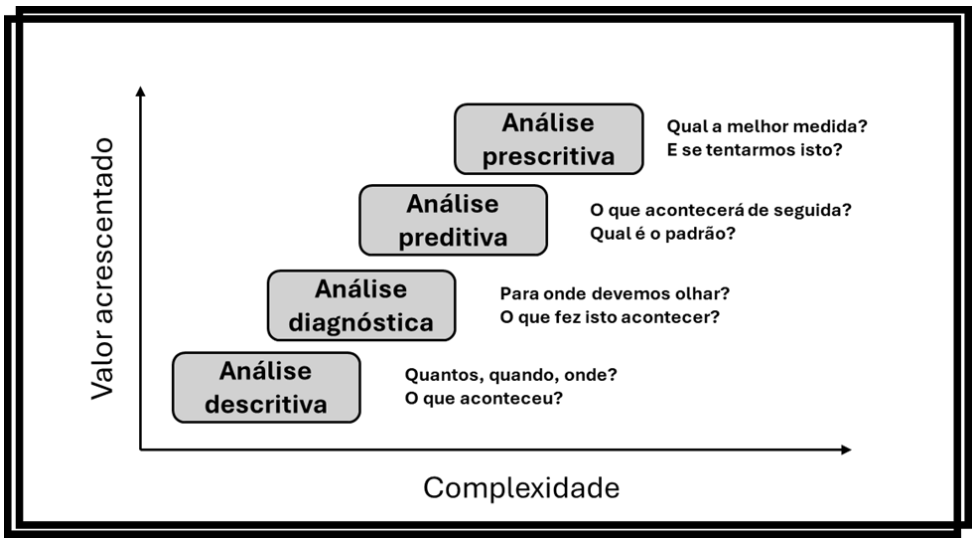


FIGURA 1.1. Os 4 tipos de análise de dados [2]

A metodologia Cross-Industry Standard Process for Data Mining (CRISP-DM) é reconhecida pela aplicabilidade em projetos de mineração de dados. Esta abordagem

abrange desde a definição do problema até à visualização dos resultados. Ferramentas como dashboards interativos, têm desempenhado um papel crucial na comunicação de *insights*, por sintetizar um grande volume de dados em visualizações intuitivas e acessíveis, promovendo uma melhoria na compreensão dos padrões e tendências, o que facilita no processo de tomada de decisão [5], [6]. Para que estas análises sejam eficazes, é necessário que se sigam as cinco etapas estruturais:

- Definição do problema/objetivos;
- Recolha e armazenamento dos dados;
- Limpeza e processamento dos dados;
- Análise dos dados;
- Visualização e comunicação dos dados.

Este processo organizado é a base para decisões mais inclusivas e alinhadas aos objetivos institucionais.

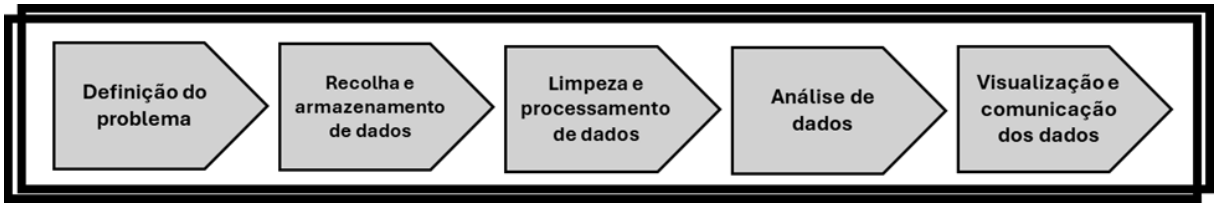


FIGURA 1.2. Pilares da análise de dados [2]

No contexto, da presente investigação, a análise de dados pode ser aplicada para:

- Identificar os candidatos mais alinhados com os programas de licenciaturas oferecidas, pela instituição, minimizando a probabilidade de insucesso académico;
- Compreender padrões na procura de licenciaturas, sugerindo potenciais ajustes no portefólio de cursos;
- Criar relatórios visuais interativos, como ferramenta de suporte à tomada de decisão e promover estratégias baseadas em evidências.

1.2. Questão de Investigação e Objetivos

Este estudo procura explorar o papel da análise de dados no contexto dos processos de admissão aos programas de licenciaturas do ISCTE - Instituto Universitário de Lisboa (ISCTE), com o objetivo de melhorar a tomada de decisão. A questão de investigação é: De que forma a análise de dados pode apoiar as direções de curso do ISCTE na compreensão dos perfis de estudantes que ingressam nas licenciaturas e na melhoria da tomada de decisão estratégica?

O objetivo é desenvolver um modelo de análise, baseado nos dados dos alunos do ISCTE, que permita identificar candidatos mais alinhados com os programas existentes e compreender tendências e padrões na procura por parte dos estudantes. Para isto, é necessário analisar fatores, que influenciam o sucesso académico, como o contexto demográfico (distrito de residência) e o percurso académico (curso anterior, média da licenciatura,

áreas de estudo). É também importante, avaliar a relação entre a procura dos cursos e as caraterísticas dos candidatos, com o objetivo de entender se há áreas de maior interesse que não estão atualmente a ser contempladas pelos programas existentes e, com isto, ser possível propor recomendações estratégicas que possam ajudar nos critérios de admissão, criação de programas introdutórios para candidatos de áreas menos alinhadas e melhorar o desempenho académico. Por fim, construir um *dashboard* que apresente os resultados desta análise que explora as caraterísitcas e padrões dos candidatos, o alinhamentos entre os perfis e desempenho esperado e que possa ser uma ferramenta para suporte à decisão. Este objetivo advém da necessidade das direções de curso em aceder a uma visão integrada e interpretável dos dados dos estudantes. Atualmente, a informação encontra-se dispersa e existe uma falta de ferramentas de visualização que permitam interpretar os dados de forma clara e comparável, dificultando a compreensão dos perfis de entrada, a identificação de fatores associados ao sucesso académico e a deteção de perfis de risco. Esta limitação compromete a capacidade de análise e de tomada de decisão, o que torna essencial o desenvolvimento de uma solução analítica que consolide e apresente os dados de uma forma acessível.

1.3. Principais Contribuições

O presente trabalho, adota uma abordagem estruturada, onde explora múltiplas variáveis que influenciam o desempenho dos estudantes, o que permite uma análise mais completa do perfil dos candidatos e a aplicação da metodologia CRISP-DM como referência para todas as etapas do ciclo de análise de dados.

No âmbito tecnológico, é proposta a construção de uma ferramenta prática para suporte à decisão, onde, através de um *dashboard* interativo em *Power BI*, sejam apresentados visualmente os dados analisados, permitindo explorar de uma forma mais intuitiva os padrões e tendências dos perfis dos candidatos. O *dashboard* proporciona uma maior facilidade de acesso aos dados e uso da ferramenta, por ser amigável para utilizadores não especializados.

Este estudo contribui para uma gestão mais eficiente e baseada em dados nos processos de admissão, o que permitirá ao ISCTE melhorar o desenho dos programas das licenciaturas e identificar os candidatos com maior potencial de sucesso. Visando assim baixar a taxa de insucesso, ajustar os cursos às necessidades dos candidatos, incluindo a criação de programas introdutórios e dar suporte à decisão. O que é possível através do *dashboard* desenvolvido.

Os resultados desta investigação, beneficiam tanto a instituição com os candidatos. Através da aplicação prática da análise de dados no contexto da admissão em licenciaturas, o trabalho desenvolvido promove processos mais justos, baseados em factos, alinhados com os objetivos instituicionais de excelência académica e inovação.

1.4. Estrutura do Documento

Este documento está estruturado em quatro capítulos principais, organizados de forma a refletir a lógica sequencial do desenvolvimento do projeto.

Capítulo 1 — Introdução: Apresenta o enquadramento do tema, os objetivos do trabalho, a motivação por detrás da investigação e a estrutura da dissertação.

Capítulo 2 — Revisão Sistemática da Literatura: Explora os principais conceitos teóricos e práticos relacionados com a visualização de dados, o contexto do ensino superior, o uso de dashboards e a metodologia CRISP-DM, que sustenta o desenvolvimento do projeto.

Capítulo 3 — Desenho e Desenvolvimento: Descreve detalhadamente o processo de recolha, limpeza, tratamento e transformação dos dados, bem como a construção de variáveis calculadas e a estruturação dos datasets utilizados nas análises.Apresenta o processo de desenvolvimento do dashboard no Power BI, com especial destaque para as visualizações implementadas, as funcionalidades interativas e a justificação técnica das escolhas visuais realizadas.

Capítulo 5 — Conclusão: Resume os principais resultados obtidos, reflete sobre as limitações do projeto e propõe direções para trabalhos futuros, nomeadamente no que diz respeito à expansão da ferramenta e ao aprofundamento das análises.

[Esta página foi deixada intencionalmente em branco]

CAPíTULO 2

Revisão Sistemática da Literatura

A análise exploratória de dados tem tido destaque como ferramenta para a caraterização de cursos do ensino superior. Desta forma, é possível identificar padrões nos dados dos estudantes, compreender as relações entre variáveis e apoiar decisões estratégicas, permitindo desenhar e melhorar os programas académicos. Estudos recentes, destacam que a análise exploratória contribui mais do que apenas identificar caraterísticas de um curso, também oferece uma nova visão que permite fazer ajustes a certos critérios de admissão e estratégias de ensino [7], [1].

O presente capítulo é uma etapa de relevo para consolidar o conhecimento atual sobre as técnicas de análise exploratória e mineração dos dados institucionais que têm sido aplicados para caraterizar os cursos académicos e entender os fatores que influenciam o desempenho dos estudantes. É apresentada uma revisão estruturada, baseada em critérios rigorosos de seleção, análise e sintese das publicações para esta dissertação.

2.1. Metodologia da Revisão

A metodologia adotada segue as diretrizes *Preferred Reporting Items for Systematic Reviews and Meta-Analyse (PRISMA)*, de forma a garantir rigor e transparência. Os passos seguidos foram:

(1) Definição do objetivo de revisão: Exploração de como a análise de dados contribui para a melhoria dos processos de admissão no ensino superior.
(2) Base de dados consultada: Scopus
(3) Palavras-chave e termos procurados:
    - data AND science AND admissions AND mining AND masters [3 documento encontrado - 0 incluído];
    - data AND analytics AND admissions AND mining AND masters [1 documento encontrado - 1 incluído];
    - predicting AND student AND admission AND success AND data AND mining [19 documentos encontrados - 8 incluídos]
    - mining AND data AND student AND admission AND performance AND dashboard OR visualization [5 documentos encontrados - 1 incluído]
    - (preditct* OR analytic*) AND (higher AND education OR masters OR universities) AND course AND management AND (academic AND sucess OR performance) AND profile [5 documentos encontrados - 2 incluídos]
    - (higher AND education OR masters OR universities) AND ("course management"OR "curriculum design"OR "program optimization") AND ("data

analytics"OR "dashboard"OR "performance tracking") [13 documentos encontrados - 3 incluídos]

- ("universities"OR "higher education") AND ("course management"OR "program optimization") AND ("decision support"OR "education *insights*") AND ("dashboard"OR "performance tracking") [22 documentos encontrados - 5 incluídos]

(4) Critérios de Inclusão/Exclusão:

- Como se pode observar na Figura 2.1, foram incluídos artigos publicados entre 2011 e 2024.
- Foram selecionados 24 artigos que investigam a aplicação de técnicas de análise exploratória de dados para identificar padrões em bases de dados académicas. Além disso, os estudos escolhidos exploraram o uso de ferramentas de visualização, como dashboards interativos, para comunicar os resultados de forma clara e acessível.
- Foram excluídos os estudos que se focavam exclusivamente em técnicas preditivas, sem conexão direta com a caracterização de cursos ou a melhoria dos programas académicos. O objetivo foi garantir que os artigos selecionados abordassem aspectos analíticos que pudessem contribuir diretamente para a compreensão e melhoramento dos processos educativos.

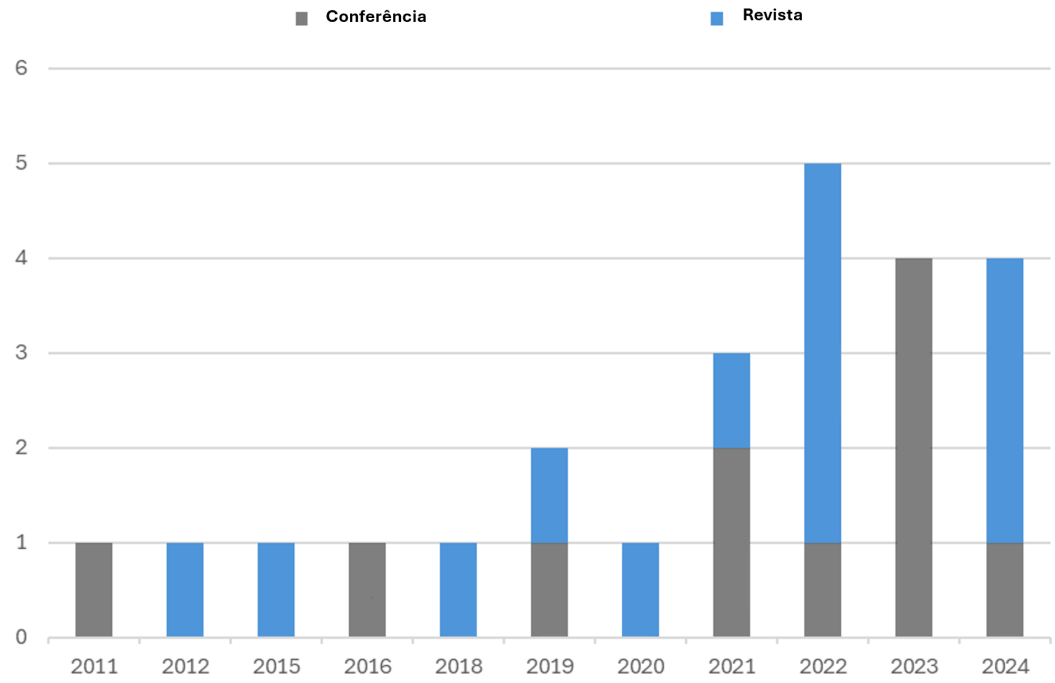


FIGURA 2.1. Contagem de Artigos ao Longo dos Anos (2011-2024)

2.2. Resultados da Revisão Sistemática da Literatura

Neste subcapítulo serão abordadas as técnicas de análise exploratória e mineração de dados, os fatores que caraterizam os estudantes, as ferramentas de visualização e os sistemas de suporte à decisão, a previsão do desempenho e retenção dos estudantes, a gestão dos cursos e a otimização dos programas educativos e, por fim, a discussão dos resultados.

2.2.1. Técnicas de Análise Exploratória e Mineração de Dados

A literatura destaca a utilização de técnicas de análise exploratória de dados e mineração de dados educativos (EDM), como uma abordagem eficaz na caraterização de cursos académicos e previsão do sucesso académico. A análise de relações entre variáveis académicas, como notas em disciplinas introdutórias e o desempenho geral, é pertinente para ajustar o desenho curricular e prever áreas de maior dificuldade nos cursos [8]. O uso de técnicas de *clustering*, como *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* e *Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)*, e de análise de correlações, também é amplamente destacado na literatura, como ferramentas fundamentais, para agrupar os estudantes com caraterísticas idênticas e compreender fatores que possam afetar o desempenho, identificando grupos de alto risco de abandono ou baixa participação, o que ajuda a implementar estratégias focalizadas[9], [10].

Modelos preditivos, como o *Grandient Boosting*, *Support Vector Machines* e *Random Forest*, são abordados e explorados, para prever o desempenho com base em dados demográficos e históricos, alcançando precisões de até 94%. Em outros estudos, é demonstrado como as técnicas de regressão linear e árvores de decisão podem ser aplicadas para explorar fatores preditivos em diferentes contextos académicos [11]–[14]. Um dos artigos detalha a aplicação de modelos de classificação para prever o sucesso em admissões de mestrado, destacando a integração de análises exploratórias para melhorar as tomadas de decisão. O estudo conduzido pela Universidade Politécnica de Timisoara, na Roménia, utilizou dados históricos de admissões e modelos preditivos, como o Gradient Boosting Classifier, que alcançou uma precisão de 87,83%. Estes modelos foram usados para analisar as preferências dos candidatos, as notas de ingresso e dados demográficos, onde foram identificados padrões que previram quais dos estudantes tinham uma maior probabilidade de ter sucesso em cursos específicos [15]. Adicionalmente, a aplicação de análises temporais demonstrou-se eficaz para monitorizar o progresso dos estudantes ao longo de um semestre. Ao dividir os períodos em intervalos, do primeiro ao quarto trimestre, foi possível identificar padrões e ajustar estratégias . Este tipo de abordagem reforça a utilidade da análise exploratória e da mineração de dados educativos na criação de estratégias de ensino adaptadas [9].

Para além das abordagens preditivas, a literatura destaca a utilização de ferramentas de visualização e sistemas de apoio à decisão. Os *Dashboards* dinâmicos têm sido amplamente utilizados para monitorizar o progresso académico, identificar as dificuldades

em disciplinas específicas e analisar os padrões de procura por cursos. A visualização de dados tem demonstrado ser uma ferramenta de grande impacto para a interpretação de grandes volumes de informação, permitindo que os gestores de curso e docentes tenham uma visão mais abrangente e detalhada do percurso académico dos estudantes e das estratégias mais eficazes para melhorar o desempenho global [16].

O pré-processamento dos dados tem um papel central na implementação destas técnicas. Os métodos de normalização e padronização garantem que os algoritmos operem com variáveis homogéneas, o que reduz enviesamentos das análises. As estratégias de redução de dimensionalidade, como o *Principal Component Analysis (PCA)*, são frequentemente utilizadas para otimizar os modelos ao eliminar redundâncias nos conjuntos de dados. Além disso, a literatura destaca a importância de técnicas como o *Synthetic Minority Over-sampling Technique (SMOTE)* para equilibrar as bases de dados desajustadas, o que torna as previsões mais fiáveis e representativas [17].

Apesar destas técnicas oferecerem melhorias substanciais na gestão e na previsão do sucesso académico, ainda existem bastantes desafios. A qualidade dos dados continua a ser um fator crítico para a fiabilidade das análises, sendo assim, fundamental que as instituições adotem boas práticas na recolha e tratamento dos dados. Além disto, há preocupações éticas relacionadas com a privacidade dos estudantes e a transparência no uso de modelos preditivos. O desenvolvimento de regulamentação clara e a implementação de protocolos institucionais são passos essenciais para garantir que a análise de dados académicos seja utilizada de uma forma ética e benéfica para todos os intervenientes no processo educativo [18].

A combinação de técnicas quantitativas e qualitativas tem sido apontada como uma das estratégias mais eficazes para uma compreensão abrangente do desempenho académico. A integração de feedbacks dos estudantes com análises preditivas permite desenvolver abordagens pedagógicas mais eficazes e adaptadas. A tendência emergente da integração da inteligência artifical na análise de dados académicos representa um avanço significativo e acaba por oferecer possibilidades mais personalizadas em termos de aprendizagem e gestão do ensino superior [9].

2.2.2. Fatores na Caraterização de Estudantes

Os estudos analisados identificaram fatores importantes, na caraterização dos estudantes, tais como, variáveis académicas, demográficas e comportamentais. Destacou-se que o desempenho inicial do estudante, como notas em disciplinas básicas ou introdutórias, ligadas à Matemática ou ciências, é um dos fatores preditivos mais importantes do sucesso nos programas académicos [8], [11], [17], [19], [20].Outros autores expandiram essa análise ao incluir fatores demográficos, como a idade, rendimento familiar e escolaridade dos pais, que influenciam diretamente as taxas de retenção e conclusão [1], [12]. Por outro lado, outros fatores, como hábitos de estudo e a prática de atividades extracurriculares, foram apontados como elementos cruciais para a compreensão do perfil dos estudantes e as suas necessidades específicas [10], [18], [21].

Apesar dos avanços tecnológicos, muitas instituições enfrentam dificuldades em alinhar os seus programas às necessidades do mercado de trabalho, particularmente em áreas tecnológicas. A introdução de ferrramentas tecnológicas como a utilização de dashboards e outros sistemas de apoio à decisão têm vindo a combater estas dificuldades de uma forma mais dinâmica e reativa [17], [18].

2.2.3. Ferramentas de Visualização e Sistemas de Suporte à Decisão

A literatura, demonstra que ferramentas de visualização interativa, como o *PerformanceVis*, permitem identificar padrões através de grandes volumes de dados dos estudantes, o que ajuda a compreender variáveis como taxas de retenção e desempenho em disciplinas específicas [7]. O uso de sistemas analíticos, como ferramenta de suporte à decisão tem sido amplamente discutidos na literatura. A visualização de dados, particularmente, através de dashbaords, emergiu como uma abordagem para melhorar a tomada de decisão, no ensino superior. Estudos apontam que estas ferramentas facilitam a interpretação dos dados, até dos mais complexos, o que permite que os diretores de curso compreendam rapidamente os padrões de desempenho académico e as taxas de retenção. Pode, por exemplo usar-se um *dashboard* para acompanhar o desempenho dos estudantes em tempo real, o que facilita e possibilita identificar áreas de maior dificuldade [16], [22]. A exploração dos dashboards interativos, demonstra o impacto na análise de dados e a sua utilidade é enfatizada na identificação padrões.[7].Projetos como o QualiChain, que têm como principal objetivo permitir que as instituições educativas verifiquem e validem dados educativos de uma forma segura e eficiente, através da exploração de tecnologias, como blockchain, que reforça o potencial enorme para garantir maior confiabilidade e transparência na verificação de dados académicos, ao mesmo tempo que dá suporte à decisão para otimizar o design curricular e os processos administrativos[16]–[18].

Sistemas híbridos, que combinam a análise exploratória de dados com modelos preditivos, são eficazes para caraterizar os cursos e identificar áreas de melhoria, pela sua capacidade de integrar múltiplas fontes de dados e de fornecer *insights* detalhados. [19].

Destaca-se que o uso de dashboards melhora significativamente a acessibilidade das análises, permitindo que os diretores de cursos ou professores, interpretem os dados e tomem decisões informadas , como também alinhar as estratégias em tempo real consoante o comportamento e necessidade dos estudantes e das instituições [7], [19], [22]. Estas técnicas de visualização ajudam a identificar lacunas no programa académico e a ajustar estratégias de ensino [21].

2.2.4. Previsão do Desempenho e Retenção de Estudantes

A previsão do desempenho e a redução das taxas de abandono académico são áreas com um grande foco na aplicação da análise de dados na educação. Os modelos baseados em *ML* como o *Random Forest* e o *Logistic Regression*, demonstraram ser eficazes para

identifcar estudantes em risco de abandono, já as técnicas não supervisionadas, como clustering, revelaram padrões comportamentais que podem ser usados para desenvolver intervenções personalizadas [9], [14] Estes modelos são de classificação binária, onde o objetivo é categorizar os estudantes em dois grupos, "em risco"ou "não em risco"ou "aprovado"ou "reprovado", como demontrado em estudos recentes [14].

Uma das abordagens, envolve a análise de *logs* em plataformas educativos, como o Moodle. Estes dados fornecem informações sobre o comportamento dos estudantes, como a frequência de login, o tempo gasto em materiais dos programas e participações no fórum. Os estudos demonstram que estas informações podem ser usadas para prever resultados académicos com alta precisão, o que permite que os professores adaptem as suas estratégias [9], [10].

2.2.5. Gestão de Cursos e Otimização dos Programas Educativos

A gestão de cursos e a otimização dos programas educativos são áreas críticas para garantir que as instituições de ensino superior atendam às expectativas dos estudantes e do mercado de trabalho. A literatura destaca que, é de grande interesse adotar abordagens inovadoras baseadas em dados, integrando tecnologias emergentes e estratégias para uma tomada de decisão informada [17], [18].

A relação entre educação e mercado de trabalho é um dos pilares da gestão educacional moderna. Muitas instituições ainda operam de maneira isolada, desenvolvendo programas com base em critérios estritamente académicos, sem considerar as exigências dos empregadores e da indústria em constante evolução. O uso de análises baseadas em dados ajuda a identificar competências exigidas pelo mercado, promovendo a inclusão destas competências nos programas educativos [17], [18].

A revisão sistemática também destaca iniciativas voltadas para a criação de programas personalizados e flexíveis. Isso inclui a oferta de cursos modulares ou programas adaptáveis que permitem aos estudantes adquirir competências específicas de acordo com as suas aspirações profissionais. A personalização é complementada por ferramentas analíticas que monitorizam as escolhas e o progresso dos estudantes, sugerindo ajustes quando necessário [10], [23].

É vital implementar modelos de controlo de qualidade no ensino superior, de forma a garantir a relevância e eficácia dos programas académicos. Os estudos sugerem que o uso de benchmarks internacionais e a integração de sistemas tecnológicos podem ajudar as instituições a alinhar os objetivos académicos à procura do mercado de trabalho [17], [18].

2.2.6. Síntese da Revisão da Literatura

Os artigos revistos, evidenciam que a análise exploratória de dados oferece uma base robusta para a caraterização de cursos. No entanto, desafios como a qualidade e integridade dos dados, podem vir a ser um barreira. Dois dos artigos, sugerem que investir na infraestrutura técnológica é essencial para retirar os maiores benefícios da análise

exploratória de dados [1], [7]. Além disso, a necessidade de integrar fatores qualitativos nos modelos analíticos é expressivo, porque embora as variáveis quantitativas ofereçam *insights* valiosos, fatores como a motivação e as competências interpessoais ainda estão por ser explorados.

Embora os modelos quantitativos possam oferecer *insights* valiosos, há uma lacuna na incorporação de fatores qualitativos, como a motivação, as competências interpessoais e o contexto cultural. Estes elementos, muitas vezes intangíveis, desempenham um papel crucial no desempenho académico, mas ainda estão pouco representados nos modelos analíticos. A integração destes fatores pode enriquecer a análise e oferecer uma visão mais holística do perfil e das necessidades dos estudantes [10], [17].

A revisão também aponta para várias oportunidades na utilização da análise de dados no ensino superior. Tecnologias emergentes, como a inteligência artificial e *ML*, oferecem um grande potencial para personalizar a aprendizagem e melhorar a experiência dos estudantes. A inclusão de análises temporais e preditivas, aliadas a estratégias baseadas em dados, pode transformar a gestão académica, permitindo que as instituições adaptem os programas de forma dinâmica e proativa [10], [18].

Por fim, esta revisão sistemática confirma que a análise exploratória de dados é uma abordagem poderosa e indispensável para caraterizar os cursos do ensino superior. As ferramentas analíticas, como dashboards, ampliam o impacto desta análise e torna os resultados mais acessíveis e intuitivos.

[Esta página foi deixada intencionalmente em branco]

CAPíTULO 3

Desenho e Desenvolvimento

Este capítulo descreve a metodologia utilizada para desenvolver o estudo, que combina uma análise exploratória de dados com a construção de uma ferramenta interativa de suporte à decisão. Foi adotada a metodologia CRISP-DM, que se revelou a mais adequada para organizar as diferentes fases do projeto, desde a recolha e preparação dos dados, até à análise e visualização dos resultados [5].

3.1. Abordagem Geral

A metodologia CRISP-DM é composta por seis fases principais que, embora apresentadas de forma sequencial, são iterativas e interdependentes. No contexto deste estudo, as etapas foram cuidadosamente adaptadas às necessidades da investigação e à realidade dos dados académicos fornecidos.

A primeira fase, denominada Compreensão do Negócio, teve como ponto de partida o levantamento dos requisitos junto dos principais *stakeholders*, neste caso, os docentes e diretores das licenciaturas em Licenciatura em Informática e Gestão de Empresas (LIGE) e Licenciatura em Ciência de Dados (LCD), do ISCTE. Esta fase permitiu compreender em profundidade as necessidades informacionais da direção dos cursos, as dificuldades sentidas na monitorização do desempenho dos estudantes, como também os principais indicadores desejados para apoio à tomada de decisão. A recolha destes requisitos foi pertinente para garantir que todas as fases seguintes estivessem alinhadas com objetivos concretos e relevantes do ponto de vista institucional.

Na fase de Compreensão dos Dados, procedeu-se à análise inicial das bases de dados fornecidas pela instituição, contendo informação detalhada sobre os estudantes inscritos nas licenciaturas referidas. Esta etapa permitiu compreender a estrutura geral dos dados, os tipos de variáveis disponíveis (como dados demográficos, percurso académico anterior, desempenho em cada ano letivo, entre outros), como também detetar inconsistências, dados em falta, formatos heterogéneos e potenciais outliers. A análise exploratória inicial foi mandatória para antecipar os desafios de tratamento que se seguiriam, ao mesmo tempo que permitiu validar a viabilidade das perguntas analíticas pretendidas.

A fase de Preparação dos Dados foi implementada em *Python* [1], utilizando bibliotecas como *Pandas* [2] e *NumPy* [3] para a limpeza, transformação e uniformização das

[1]Python Software Foundation. Python Language Reference. https://www.python.org
[2]McKinney, W. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 2010.
[3]Harris, C. R. et al. Array programming with NumPy. Nature, 585, 357–362, 2020.

variáveis [9]. Esta etapa incluiu o tratamento de valores nulos, a normalização de categorias (como áreas de estudo, estatutos académicos e cursos de origem), a conversão de formatos numéricos e textuais e a consolidação dos dados em estruturas finais coerentes e bem definidas. O objetivo foi assegurar a qualidade e integridade da informação, preparando-a para ser utilizada no ambiente de visualização do *Power BI*.

De seguida, na fase de Modelação, foram criadas colunas calculadas no *Power BI*, recorrendo à linguagem *Data Analysis Expressions (DAX)*, de forma a enriquecer os dados e permitir segmentações analíticas mais robustas. Métricas como a percentagem de créditos concluídos, a classificação por grupos de desempenho e a ordenação lógica das faixas foram desenvolvidas com o objetivo de suportar visualizações mais informativas e facilitar comparações entre diferentes perfis de estudantes.

A fase de Visualização e Comunicação consistiu na construção do *dashboard* interativo em *Power BI*, organizado em quatro páginas principais (duas para cada curso analisado) e uma página inicial com o objetivo de se tornar uma página de navegação para as páginas principais do *dashboard*. Este *dashboard* foi desenhado para oferecer uma navegação intuitiva e clara, através de gráficos, mapas e filtros interativos, permitindo aos docentes explorar os dados de forma dinâmica, validar hipóteses e identificar padrões relevantes de desempenho e evolução dos alunos ao longo dos anos.

Por fim, a fase de Avaliação centrou-se na verificação da qualidade final do produto analítico. Foram conduzidos testes de validação dos filtros, métricas e visualizações, com o objetivo de garantir que todas as perguntas analíticas previamente levantadas pudessem ser respondidas através do *dashboard*. Esta etapa incluiu também sessões informais de validação com os docentes envolvidos, que confirmaram a utilidade das análises produzidas e a pertinência das funcionalidades implementadas.

A adoção da metodologia CRISP-DM, como é possível ver na Figura 3.1, permitiu garantir um processo estruturado e iterativo, assegurando o rigor técnico do tratamento de dados e a utilidade prática dos resultados produzidos. Cada fase foi concebida não como um fim em si mesma, mas como parte de um ciclo contínuo de aprendizagem e melhoria, em resposta às necessidades reais dos utilizadores finais da solução.

3.2. Levantamento dos Requisitos dos Docentes

A definição de requisitos foi um dos pilares fundamentais para o desenvolvimento deste projeto. A docente Ana Maria de Almeida, orientadora deste trabalho e ex-Diretora da LCD, foi a responsável pela recolha de requisitos entre direções de curso. O principal objetivo foi identificar as necessidades práticas sentidas pelas direções de curso no acompanhamento dos estudantes, assim como os indicadores considerados relevantes para caracterizar os perfis dos alunos, o seu desempenho e a sua progressão ao longo do tempo.

A recolha de requisitos permitiu sistematizar um conjunto de objetivos analíticos que orientaram toda a construção do *dashboard* interativo, como também o processo de preparação, análise e visualização dos dados.

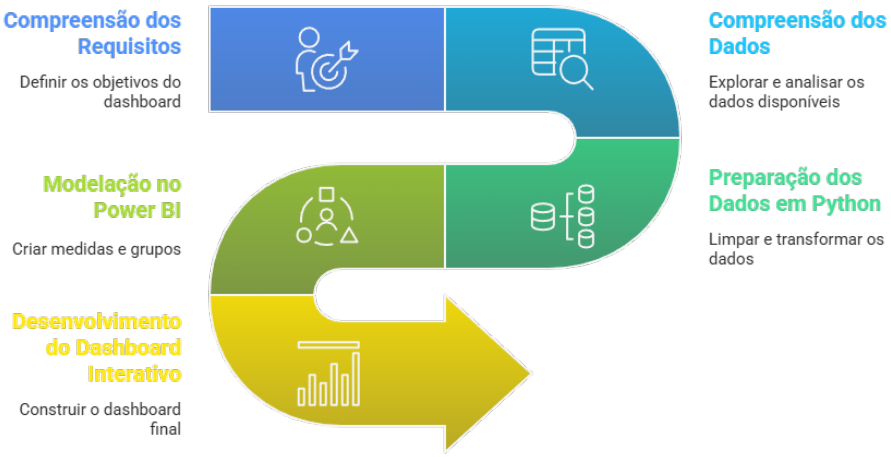


FIGURA 3.1. Metodologia e Pipeline da Solução Desenvolvida

Seria necessário fazer uma caracterização dos novos alunos. Pretendia-se compreender de forma aprofundada o perfil de entrada dos estudantes admitidos no primeiro ano curricular. Para isso, foi indicado como prioritário analisar a distribuição das notas de ingresso e as áreas de estudo de proveniência no ensino secundário. A componente geográfica também se revelou importante, sendo proposta a criação de uma visualização da distribuição da residência dos alunos, com segmentação por distrito. Outro aspeto considerado importante foi a possibilidade de relacionar o desempenho académico inicial (nota de ingresso) com a área de estudo.

Comparar os diferentes anos letivos, com o intuito de analisar a evolução dos indicadores ao longo do tempo. Esta análise permitiria às direções de curso identificar alterações estruturais nos perfis dos candidatos e perceber se existiam variações significativas nos padrões de entrada.

Um dos principais objetivos da análise foi avaliar o desempenho académico dos estudantes, com especial interesse em segmentá-los com base na percentagem de créditos concluídos, de forma a identificar diferentes níveis de aproveitamento e padrões de sucesso. Foram definidos cinco grupos: estudantes que concluíram 100% dos créditos a que se inscreveram, entre 60% e 80%, entre 30% e 50%, menos de 30%, e aqueles que não concluíram nenhum crédito. A segmentação dos estudantes pelos intervalos de percentagem de créditos concluídos segue as orientações das docentes responsáveis, que recomendaram o uso destes *thresholds* por refletirem níveis distintos e de fácil interpretação do progresso académico, relevantes para a análise institucional. Para além disso, foi considerado essencial analisar a distribuição das médias finais, a comparação entre a média do ano anterior e a média do ano atual (ou de ingresso, no caso dos estudantes do primeiro ano) por ano curricular. Por fim, foi identificada a necessidade de analisar especificamente o

desempenho dos estudantes com estatuto de trabalhador-estudante, comparando os seus resultados com os restantes colegas.

Este levantamento detalhado de requisitos foi determinante para a seleção das variáveis a integrar no modelo analítico, para o desenho das métricas a calcular e para a definição das visualizações incluídas no *dashboard* final. Ao alinhar as decisões técnicas com as necessidades dos utilizadores finais, foi possível garantir que o produto desenvolvido teria aplicabilidade prática e utilidade real na gestão e acompanhamento dos estudantes.

Deste modo, o modelo analítico e o dashboard foram concebidos para responder às seguintes questões:

- Qual o curso de proveniência dos alunos antes do ingresso no curso?
- Existem diferenças nas notas de ingresso consoante o curso anterior dos estudantes?
- De onde são, geograficamente, os alunos que ingressam no curso?
- Como tem evoluído a nota de ingresso dos estudantes ao longo dos anos?
- Como tem evoluído o desempenho académico dos estudantes ao longo dos anos?
- Existe relação entre a média de ingresso e a média final obtida pelos alunos em cada ano curricular?

3.3. Compreensão dos Dados

De forma a garantir a conformidade com os princípios legais e éticos de proteção de dados, todos os dados utilizados neste projeto foram disponibilizados pelo ISCTE de forma completamente anonimizada. Esta medida visou assegurar que nenhuma informação pessoal ou sensível pudesse ser associada a qualquer estudante individual, protegendo assim a sua identidade e a sua privacidade, em conformidade com o Regulamento Geral sobre a Proteção de Dados (RGPD).

A anonimização dos dados consistiu na remoção ou substituição de todos os identificadores que pudessem permitir a identificação dos indivíduos. Isto incluiu nomes próprios, números de estudante, endereços de email institucionais e quaisquer outros elementos que, isoladamente ou em combinação com outros, pudessem identificar uma pessoa. O processo de anonimização foi efetuado pela equipa responsável pela extração dos dados na instituição, antes da entrega da base para análise, garantindo que os dados chegavam ao investigador já protegidos de acordo com as boas práticas. Foram disponibilizadas duas bases de dados distintas, pelo ISCTE, com informações sobre os estudantes inscritos na Licenciatura em Informática e Gestão de Empresas entre os anos de 2014 e 2024 e na Licenciatura em Ciência de Dados entre os anos de 2019 e 2024. Para cada um dos cursos, foram fornecidos ficheiros no formato .xlsx, tendo informação detalhada sobre os estudantes inscritos. Apesar da anonimização, a base de dados manteve-se suficientemente rica do ponto de vista analítico. Foram preservadas variáveis relevantes como:

- **Dados demográficos**:
  - Idade do estudante;

  - Sexo;
  - Nacionalidade;
  - Região de residência.
- **Percurso académico prévio**:
  - Curso de origem e área de estudo no secundário;
  - Exames nacionais realizados e respetivas classificações;
  - Média de ingresso no ensino superior.
- **Dados de inscrição e desempenho**:
  - Ano de ingresso e regime de inscrição;
  - Número de créditos inscritos e concluídos;
  - Média obtida no ano letivo corrente e anterior.
- **Outras variáveis relevantes**:
  - Estatuto de trabalhador-estudante;
  - Situação académica (ativo, desistente, concluído);
  - Possíveis equivalências atribuídas.

Todos os campos foram descritos detalhadamente no dicionário de dados (ver Apêndice A), que serviu como base para compreender o significado, o formato e as possíveis categorias de cada variável. A compreensão prévia dos dados foi essencial para identificar inconsistências, variáveis redundantes e outliers antes do início da fase de tratamento.

Para garantir a consistência entre diferentes cursos e anos letivos, os datasets foram integrados em quatro dataframes finais, preservando um identificador único por estudante. Esta estrutura permitiu realizar análises transversais, comparando padrões entre anos e cursos. O dicionário de dados encontra-se no Apêndice A e serve como referência para o significado e o formato de cada variável utilizada.

3.4. Preparação, Tratamento e Modelação dos Dados

A preparação dos dados constituiu uma das fases mais críticas deste projeto, uma vez que a fiabilidade das análises depende diretamente da qualidade da informação utilizada. Esta etapa foi desenvolvida integralmente em *Python*, utilizando bibliotecas especializadas como *Pandas*, *NumPy* e *Matplotlib*, que permitiram manipular eficientemente grandes volumes de dados e realizar operações complexas de limpeza e transformação.

O primeiro passo consistiu na remoção de dados ausentes e inconsistentes. Foram identificadas linhas com valores nulos em variáveis consideradas essenciais, como médias académicas ou créditos inscritos, que foram tratadas com estratégias diferenciadas consoante a natureza e relevância da variável em causa.

No caso de colunas consideradas sem utilidade, por apresentarem apenas um valor único ou por estarem completamente vazias, optou-se pela sua eliminação. Adicionalmente, foi calculada a percentagem de dados em falta por coluna, sendo eliminadas todas aquelas que apresentavam mais de 80% de valores nulos, por se considerar que não contribuiriam significativamente para a análise.

Os dados numéricos em falta foram imputados utilizando a mediana ou a média, consoante a variável em questão, procurando sempre a abordagem que causasse menos distorção nas distribuições e que minimizasse o risco de enviesamento nos resultados. Já os dados categóricos em falta foram imputados com valores neutros ou descritivos considerados como "nulos informativos", de modo a não introduzir ruído analítico. Exemplos deste processo incluem:

- `'Estatutos'`: imputado com `'Sem estatuto'`;
- `'Último Estado'`: imputado com `'Matriculado'`;
- `'Ano Letivo do Último Estado'`: imputado com `'Sem inscrição'`;
- `'Ano Lectivo: Inscrito em UCs Obrigatórias Atrasadas'`: imputado com `'Desconhecido'`.

Seguiu-se a fase de uniformização de categorias, essencial para garantir a consistência semântica ao longo de todo o conjunto de dados. Variáveis como os nomes dos cursos de origem, as áreas de estudo ou os estatutos académicos surgiam com diferentes grafias, abreviaturas e estruturas de preenchimento entre anos e entre os dois cursos analisados. Através de mapeamentos personalizados e processos sistemáticos de normalização, foi possível consolidar estas categorias, preservando a riqueza semântica da informação.

De forma a ultrapassar as limitações impostas pela anonimização dos dados, revelou-se essencial criar um identificador único para cada estudante. A inexistência de um identificador direto impedia a realização de análises individuais e longitudinais. Assim, foi criada uma chave composta, com base na combinação das seguintes variáveis: `distrito de nascimento`, `concelho de nascimento`, `freguesia de nascimento`, `distrito`, `concelho`, `freguesia`, `localidade`, `código postal` e `área do código postal`. Esta abordagem permitiu representar cada estudante de forma inequívoca dentro da amostra analisada.

Como resultado, foi possível identificar 1573 estudantes únicos na LIGE e 868 estudantes na LCD, assegurando as condições necessárias para análises personalizadas e comparações interanuais fiáveis.

Para permitir uma análise mais pormenorizada do grupo de trabalhadores-estudantes e já com o objetivo de incluir um filtro dedicado no *dashboard* interativo — foi também criada uma nova coluna binária, que indica se um estudante possui ou não o estatuto de trabalhador-estudante. Esta coluna foi derivada diretamente da informação presente na variável `'Estatutos'`, sendo codificada da seguinte forma:

- Se o valor em `'Estatutos'` corresponder a `'Trabalhador-Estudante'`: novo valor = `'Sim'`;
- Em todos os restantes casos: novo valor = `'Não'`.

Esta variável permitiu não só segmentar os estudantes de forma clara, como também viabilizou a análise comparativa do seu desempenho em relação aos restantes colegas, através de filtros e visualizações específicas no *Power BI*.

Após todas estas fases de limpeza, enriquecimento e transformação, os dados foram integrados em quatro estruturas principais:

- `df_todos_final` e `dc_df_todos_final`: estruturas que agregam os registos históricos de cada estudante, contemplando múltiplas observações ao longo dos anos letivos;
- `dadosunicos_tratados_final` e `dc_dadosunicos_tratados_final`: versões finais consolidadas, onde cada estudante surge apenas uma vez, com os dados mais relevantes do seu percurso académico.

Esta separação foi estratégica: permitiu realizar tanto análises longitudinais (com múltiplas observações por aluno ao longo do tempo) como transversais (uma linha por estudante), assegurando que cada registo no modelo representava inequivocamente um aluno distinto, através da utilização e validação dos identificadores únicos criados.

A integração posterior com o *Power BI* exigiu não apenas a importação destes datasets tratados, mas também o enriquecimento dos mesmos através da criação de colunas adicionais, utilizando a linguagem *DAX*. Esta modelação foi realizada já no ambiente do *Power BI*, com o objetivo de suportar análises mais dinâmicas e segmentações interativas, respondendo a requisitos concretos dos docentes.

Três colunas foram especialmente relevantes neste processo. A primeira, denominada `Perc_Creditos`, calcula a percentagem exata de créditos concluídos por cada estudante com base na fórmula:

$$\texttt{Perc_Creditos} = \frac{\text{Créditos Concluídos}}{\text{Créditos Inscritos}} \times 100$$

Esta métrica tornou-se a base de muitas das visualizações analíticas, por permitir a comparação proporcional entre estudantes com diferentes cargas horárias e níveis de progressão.

A segunda coluna criada, `Faixa_Desempenho`, classifica os estudantes em grupos de aproveitamento com base na percentagem de créditos concluídos. Esta segmentação foi definida em conjunto com os docentes, resultando em cinco grupos: estudantes que concluíram 100% dos créditos, entre 60% e 80%, entre 30% e 50%, menos de 30%, e aqueles que não concluíram qualquer crédito. Esta classificação foi essencial para visualizar padrões de desempenho e comparar perfis entre diferentes grupos de estudantes.

Por fim, foi criada a coluna `Faixa_Ordem`, responsável por estabelecer uma ordenação lógica e personalizada dos grupos definidos anteriormente. Esta coluna é indispensável no contexto do *Power BI*, uma vez que evita que as categorias sejam organizadas de forma alfabética nos gráficos — o que prejudicaria a interpretação visual dos dados e dificultaria a leitura das tendências.

A implementação destas colunas diretamente no *Power BI* trouxe dois benefícios concretos. Em primeiro lugar, permitiu maior eficiência computacional, evitando sobrecarga durante o pré-processamento em *Python*. Em segundo, aumentou a interatividade das visualizações, já que os cálculos passaram a ser feitos dinamicamente, permitindo filtros e análises em tempo real por parte dos utilizadores.

A escolha das ferramentas ao longo do projeto foi orientada por critérios de flexibilidade, robustez e integração. *Python* foi utilizado para todas as operações de pré-processamento, exploração inicial e tratamento intensivo dos dados. O *Excel* desempenhou um papel complementar na validação do dicionário de dados e inspeção preliminar dos ficheiros recebidos. Finalmente, o *Power BI* foi selecionado como ambiente principal de visualização e comunicação dos resultados, não só pela sua capacidade de ligação direta a fontes de dados tratadas, mas também pela sua versatilidade na construção de dashboards interativos com filtros dinâmicos, segmentações e cálculos personalizados.

Esta sequência do tratamento intensivo em *Python*, passando pela modelação em *DAX*, até à visualização final no *Power BI* constituiu um pipeline analítico coeso e funcional, garantindo que os dados fossem não só corretos e bem estruturados, como também úteis e facilmente interpretáveis por quem deles depende para a tomada de decisões pedagógicas e estratégicas.

3.5. Desenvolvimento do *Dashboard* Interativo

A fase de desenvolvimento do *dashboard* interativo constituiu um momento-chave deste projeto, uma vez que materializou os objetivos definidos na fase de levantamento de requisitos e tornou os dados acessíveis de forma clara, explorável e visualmente intuitiva. O *Power BI* foi a ferramenta escolhida para esta tarefa, pelas suas capacidades avançadas de visualização, interatividade e ligação dinâmica a fontes de dados tratadas previamente.

O objetivo central do *dashboard* foi permitir que os docentes e responsáveis de curso pudessem consultar, analisar e comparar indicadores relevantes sobre os estudantes das licenciaturas em Informática e Gestão de Empresas e em Ciência de Dados, ao longo de diferentes anos letivos. A ferramenta foi desenvolvida de forma a responder diretamente aos requisitos previamente definidos, descritos no ponto 3.2, em conjunto com os docentes, estando estruturada para fornecer respostas claras a questões analíticas concretas.

Em primeiro lugar, foi priorizada a análise dos novos alunos, permitindo visualizar a distribuição das notas de ingresso, as áreas de estudo no ensino secundário, facilitando a compreensão do perfil de entrada dos estudantes. Em segundo lugar, foi implementada uma visualização geográfica da origem dos alunos, através de mapas interativos que apresentam a localização de residência por distrito, contribuindo para uma leitura espacial imediata da distribuição dos estudantes.

A avaliação do desempenho académico foi também uma prioridade, incluindo indicadores como a percentagem de créditos concluídos, as médias anuais e a sua evolução ao longo do percurso académico. Esta análise é essencial para identificar padrões de sucesso, retenção ou dificuldades ao longo do tempo. A comparação temporal entre diferentes anos, permite detetar tendências estruturais no perfil e desempenho dos estudantes.

Por fim, foi dada especial atenção ao grupo dos trabalhadores-estudantes, permitindo a sua segmentação e comparação com os restantes alunos ao nível do aproveitamento académico, nomeadamente através da média e da percentagem de créditos concluídos. Esta

funcionalidade responde diretamente ao interesse manifestado pelos docentes em compreender as especificidades e desafios enfrentados por este grupo em particular. Todas as visualizações estão interligadas, permitindo uma navegação fluida e uma exploração intuitiva dos dados por parte dos utilizadores finais, como se pode observar nos apêndices B.1, B.2, C.1, C.2.

[Esta página foi deixada intencionalmente em branco]

CAPíTULO 4

Resultados e Discussão

Este capítulo apresenta os resultados obtidos a partir do desenvolvimento do dashboard e da análise realizada sobre os dados dos estudantes. São descritas as principais visualizações construídas, bem como os padrões identificados relativamente ao perfil de ingresso, desempenho académico e a evolução ao longo dos anos letivos. Realiza-se, adicionalmente, a discussão destes resultados, relacionando-os com a literatura e com os objetivos definidos para o estudo, de forma, a avaliar o contributo da análise de dados para o apoio à decisão das direções de curso.

4.1. Visualizações e Resultados Obtidos no *Dashboard*

A estrutura do *dashboard* foi organizada de forma a proporcionar uma navegação clara e lógica, separando as análises por curso. Foram criadas quatro páginas principais, duas para cada licenciatura em análise: LIGE e a LCD. Cada uma destas páginas apresenta um conjunto coeso de visualizações sobre o perfil e desempenho dos estudantes, mantendo uma disposição visual consistente para facilitar a interpretação. A escolha dos gráficos foi cuidadosamente feita para garantir que cada tipo de visualização respondesse de forma eficaz às perguntas analíticas levantadas, permitindo observar padrões e realizar comparações relevantes entre diferentes grupos de estudantes, anos letivos e perfis. Através das visualizações interativas, dos filtros dinâmicos e das métricas calculadas com *DAX*, foi possível identificar padrões relevantes, validar hipóteses formuladas pelos docentes e extrair conclusões sobre os fatores que influenciam o sucesso académico e os perfis de ingresso.

4.1.1. Licenciatura de Informática e Gestão de Empresas

A presente secção detalha as visualizações incluídas no *dashboard* interativo desenvolvido em *Power BI*, apresentando a fundamentação por detrás da escolha de cada tipo de gráfico e a ligação aos requisitos analíticos definidos pelos docentes.

A Figura 4.1 responde diretamente à questão colocada pelos docentes: "Qual o curso de proveniência dos alunos antes do ingresso no curso?". Ao apresentar a quantidade de alunos por curso de proveniência, é possível compreender qual o background académico predominante dos estudantes que ingressam na LIGE, sendo este um indicador relevante para avaliar o alinhamento entre a formação prévia e as exigências do plano curricular da licenciatura.

Foi utilizado um gráfico de barras verticais simples, dada a natureza categórica da variável "Curso Anterior" e o objetivo de visualizar rapidamente o volume de alunos por

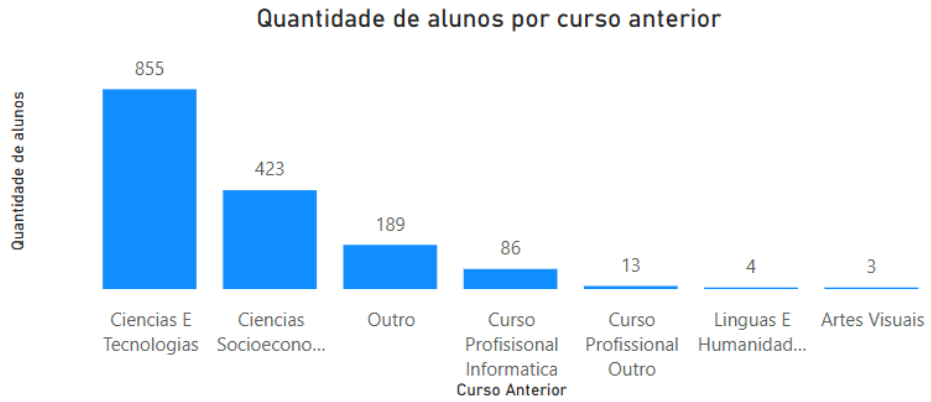


FIGURA 4.1. Distribuição dos alunos da LIGE por curso anterior no momento do ingresso

categoria. Este tipo de gráfico é particularmente eficaz para representar comparações diretas entre grupos e facilitar a leitura de frequências absolutas.

O gráfico, da Figura 4.1, evidencia a predominância clara dos estudantes provenientes dos cursos de Ciências e Tecnologias, que representam a maior fatia com um total de 855 alunos. Esta concentração confirma que a maioria dos candidatos a cursos de natureza tecnológica ou científica provém desta via de ensino secundário. Em segundo lugar surgem os alunos provenientes de Ciências Socioeconómicas, num total de 423 estudantes, revelando também uma participação significativa neste percurso académico.

As categorias de Outros cursos (189 alunos) e Cursos Profissionais de Informática (86 alunos) aparecem com valores bastante mais reduzidos, o que sugere que, embora sejam percursos possíveis de acesso, o seu contributo para o total de inscritos é consideravelmente menor. Ainda mais residuais são os números de alunos vindos de Cursos Profissionais de outras áreas (13), de Línguas e Humanidades (4) e de Artes Visuais (3), que praticamente não têm expressão no universo analisado.

Esta distribuição mostra que o perfil típico de ingresso é fortemente marcado pelas áreas científico-tecnológicas, seguidas das socioeconómicas, enquanto as restantes áreas desempenham apenas um papel marginal. Esta informação é relevante, uma vez que permite compreender melhor a adequação dos percursos anteriores ao perfil de exigência do curso e também orientar estratégias de divulgação ou apoio dirigidas a alunos provenientes de vias de ensino menos representadas.

A Figura 4.2 dá resposta à questão: "Existem diferenças nas notas de ingresso consoante o curso anterior dos estudantes?". Esta visualização permite observar o desempenho médio no ingresso para cada curso de origem, possibilitando uma análise comparativa direta.

A escolha de um gráfico de barras verticais mantém-se adequada neste contexto, pois continua-se a trabalhar com uma variável categórica (curso anterior), agora cruzada com uma variável contínua (nota média de ingresso). Este tipo de gráfico é eficaz para

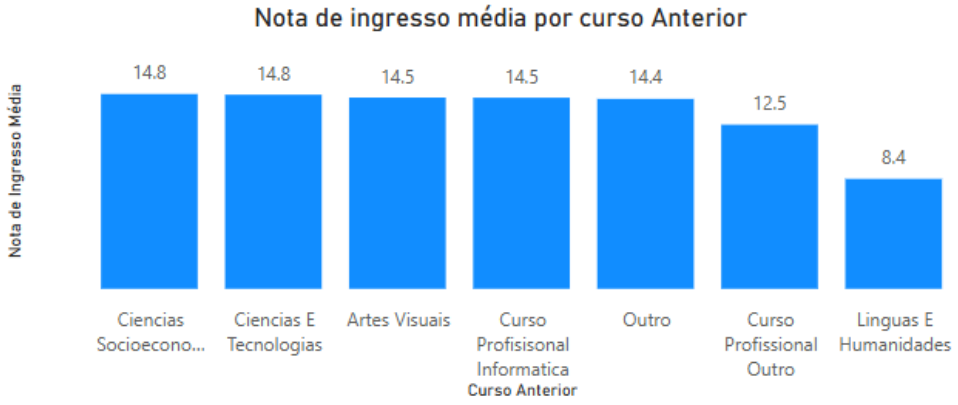


FIGURA 4.2. Nota de ingresso média por curso anterior dos alunos da LIGE

evidenciar contrastes e tendências entre categorias, facilitando a identificação de padrões de desempenho conforme a formação prévia.

Esta análise é relevante para os docentes porque permite relacionar o background académico com o sucesso esperado à entrada, oferecendo uma primeira leitura sobre a adequação ou eventual desfasamento entre a preparação pré-universitária dos alunos e as exigências iniciais do curso.

A análise do gráfico, da Figura 4.2, revela que as notas de ingresso médias apresentam variações relevantes em função do curso de origem. Os estudantes provenientes de Ciências Socioeconómicas e de Ciências e Tecnologias registam as notas mais elevadas, ambas situadas nos 148 valores, o que mostra que estas duas vias de ensino não apenas concentram o maior número de alunos, como também garantem desempenhos consistentes no momento da candidatura. De seguida, encontram-se os alunos de Artes Visuais e de Cursos Profissionais de Informática, ambos com uma média de 14.5 valores, praticamente ao mesmo nível dos percursos científico-tecnológicos e socioeconómicos. Esta proximidade sugere que, apesar de terem menor representatividade em termos de quantidade de estudantes, os alunos que ingressam a partir destas áreas chegam ao ensino superior com níveis de preparação semelhantes.

Os estudantes classificados como Outros apresentam uma média ligeiramente inferior (14.4 valores), enquanto aqueles que provêm de Cursos Profissionais de Outras Áreas evidenciam uma diferença mais acentuada, com uma média de apenas 12.5 valores. Por fim, os candidatos de Línguas e Humanidades destacam-se por apresentar a nota de ingresso média mais baixa (84 valores), o que sugere uma clara desvantagem competitiva face às restantes origens escolares.

Por fim, o gráfico evidencia uma relação entre a via de ensino anterior e o desempenho no ingresso, reforçando o papel estruturante das áreas de Ciências e Tecnologias e de Ciências Socioeconómicas não só em termos de quantidade de alunos, mas também de qualidade académica medida pelas notas de entrada.

A Figura 4.3 responde à questão colocada pelos docentes: "De onde são, geograficamente, os alunos que ingressam no curso?".



FIGURA 4.3. Distribuição geográfica dos alunos da LIGE por distrito de residência

Esta visualização apresenta a distribuição dos estudantes da LIGE por distrito de residência no momento do ingresso, utilizando um mapa interativo com pontos georreferenciados. Esta abordagem permite identificar, de forma imediata e intuitiva, quais as regiões do país com maior ou menor representatividade.

A utilização de um mapa interativo foi considerada a forma mais eficaz de representar esta informação, já que permite simultaneamente observar a dispersão geográfica e o peso relativo de cada distrito. Cada ponto representa um distrito e o seu tamanho é proporcional ao número de alunos aí residentes. A visualização aproveita a capacidade do *Power BI* para ligar automaticamente os dados de localização aos mapas da Microsoft Azure, oferecendo uma experiência de leitura clara, direta e espacialmente contextualizada. Este tipo de análise geográfica é particularmente útil para apoiar decisões relacionadas com estratégias de divulgação regional do curso, identificação de zonas de captação e possíveis áreas com menor penetração, interpretação de possíveis padrões de migração académica.

O mapa, da Figura 4.3, permite observar a distribuição geográfica dos estudantes de acordo com o distrito de proveniência. Nota-se uma forte concentração de alunos nos distritos da faixa litoral, em particular na Área Metropolitana de Lisboa, que se destaca como o principal polo de origem. Esta concentração é expectável, tendo em conta a densidade populacional da região e a proximidade face à instituição de ensino.

Para além de Lisboa, registam-se também valores expressivos nos distritos do Centro do país, como Setúbal, Santarém, Leiria e Coimbra, reforçando a ideia de que a maioria dos estudantes provém de zonas urbanas e próximas da costa. Em contraste, os distritos do interior apresentam menor representatividade, o que pode estar associado tanto à densidade populacional mais baixa como a fatores de acessibilidade e mobilidade.

É ainda visível a presença de estudantes oriundos das Regiões Autónomas dos Açores e da Madeira, embora em números bastante mais reduzidos, o que demonstra que, apesar da distância geográfica, existem candidatos que optam por este curso.

O gráfico evidencia uma clara assimetria regional, marcada por uma forte concentração de estudantes no litoral e uma participação mais residual das regiões insulares e

do interior do país. Esta informação pode ser particularmente útil para delinear estratégias de divulgação do curso em áreas menos representadas, contribuindo para uma maior diversificação territorial dos estudantes.

A Figura 4.4 procura responder à pergunta analítica: "Como tem evoluído a nota de ingresso dos estudantes ao longo dos anos?".



FIGURA 4.4. Evolução da nota de ingresso média dos alunos da LIGE por ano letivo

Esta visualização apresenta a média das notas de ingresso dos alunos da LIGE, segmentada por ano letivo. O objetivo principal deste gráfico é permitir a análise longitudinal da qualidade média dos ingressantes, tal como solicitado pelos docentes, permitindo observar flutuações e tendências ao longo do tempo.

A escolha de um gráfico de linhas com área preenchida prende-se com a necessidade de representar uma variável contínua (nota de ingresso média), realçar visualmente a variação ano após ano e destacar os pontos de inflexão, picos ou quedas abruptas, que podem motivar análises adicionais ou levar à identificação de fenómenos externos (por exemplo, alterações nos critérios de acesso ao ensino superior, impacto da pandemia, etc.).

Através da Figura 4.4, conclui-se que a evolução das notas de ingresso médias ao longo dos anos letivos revela uma tendência global de crescimento, embora marcada por oscilações em determinados períodos. Em 2014/2015, a média situava-se em torno dos 13.8 valores, registando um aumento progressivo até 2017/2018, quando atingiu aproximadamente 14.9 valores. No ano seguinte observou-se uma ligeira quebra, mas a tendência voltou a ser positiva, culminando em 2020/2021, ano em que se registou o valor mais elevado, com cerca de 15.9 valores. A partir desse ponto verifica-se alguma instabilidade: em 2021/2022 há uma descida acentuada para valores próximos dos 15.1, seguida de uma nova recuperação em 2022/2023,com a média a voltar a subir para cerca de 15.6 valores. Contudo, em 2023/2024 regista-se nova queda, voltando a níveis semelhantes aos de 2018/2019, em torno dos 14.6 valores. Este comportamento sugere que, apesar da trajetória de crescimento observada durante grande parte da década, existem fatores que

influenciam as notas de ingresso, podendo ser devido à variação do número de vagas ou à procura pelo curso. Em termos gerais, a análise permite concluir que o perfil de ingresso tem vindo a consolidar-se em patamares mais elevados do que os registados no início do período estudado, demonstrando uma maior competitividade e exigência no acesso.

A Figura 4.5 responde diretamente à questão levantada pelos docentes: "Como tem evoluído o desempenho académico dos estudantes ao longo dos anos?". Neste gráfico é apresentada a distribuição anual dos estudantes da LIGE por níveis de desempenho académico, definidas de acordo com a percentagem de créditos concluídos em relação aos créditos em que o estudante se inscreveu. Esta visualização fornece uma perspetiva abrangente sobre o aproveitamento académico dos estudantes ao longo do tempo, permitindo identificar padrões de sucesso, retenção ou insucesso, conforme solicitado nas sessões de levantamento de requisitos. O gráfico empilhado, da Figura 4.5, permite observar a evolução do rendimento académico dos estudantes ao longo dos anos letivos, agrupando-os por categorias de conclusão: 100%, 60–80%, 30–50%, <30% e 0% concluído.

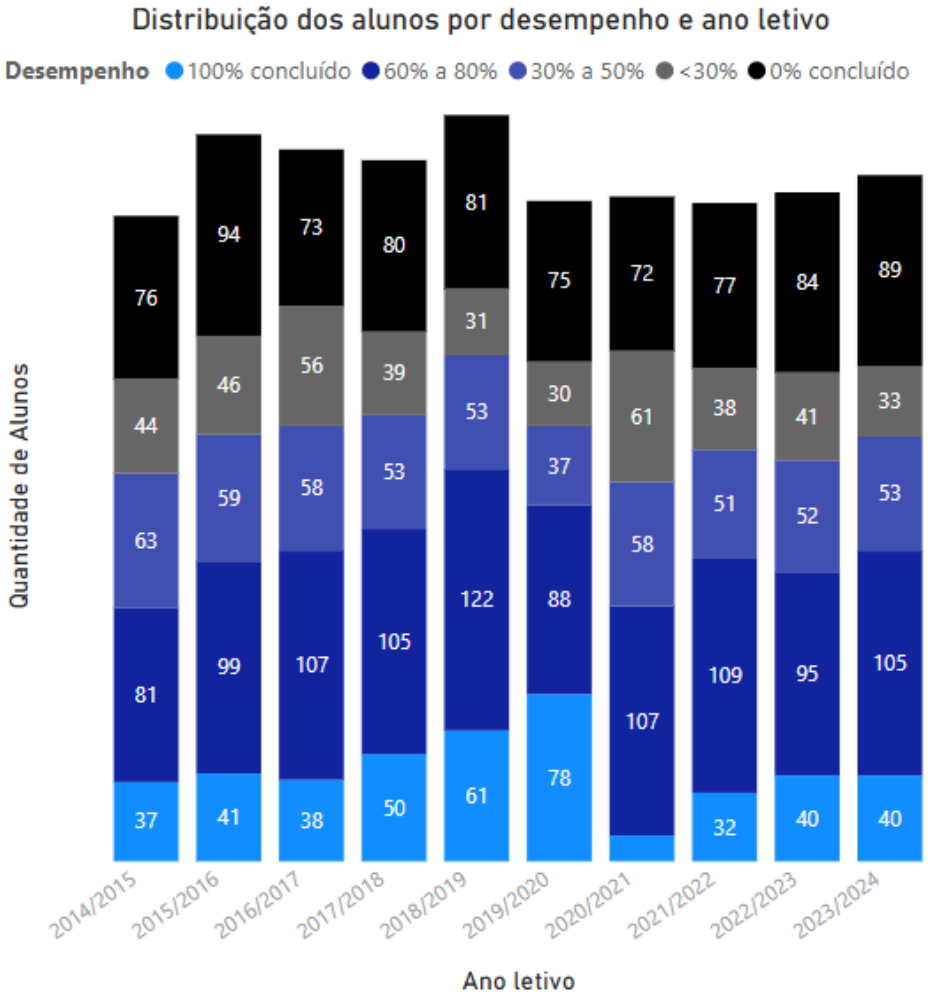


FIGURA 4.5. Distribuição dos alunos da LIGE por desempenho e ano letivo

De uma forma geral, nota-se que apenas uma pequena fração dos alunos consegue concluir integralmente (100%) o plano de estudos em cada ano, variando entre cerca de

30 e 60 estudantes. Este grupo apresenta alguma oscilação ao longo do tempo, com destaque para 2019/2020, ano em que se registou o maior número de conclusões totais (61 alunos). A faixa de 60 a 80% de conclusão representa uma parte consistente e significativa dos estudantes, mantendo-se sempre como uma das categorias mais numerosas, com valores que chegam a ultrapassar os 120 alunos em 2018/2019. Este padrão sugere que muitos estudantes progridem de forma sólida, ainda que não completem integralmente todas as unidades curriculares num único ano. Já os grupos de desempenho mais baixos, em especial os alunos com 0% de conclusão, mantêm-se sempre presentes em números relativamente elevados. Em alguns anos, como 2015/2016 e 2016/2017, este grupo atinge valores superiores a 90 estudantes, o que evidencia que uma parte significativa dos inscritos não apresenta aproveitamento académico. Embora este número oscile ao longo da série, continua a representar uma preocupação persistente em praticamente todos os anos letivos. Por fim, os grupos intermédios (30–50% e <30% de conclusão) oscilam em função das coortes, mas confirmam que uma proporção não negligenciável de alunos progride apenas parcialmente, acumulando unidades curriculares em atraso. Em síntese, o gráfico mostra uma distribuição heterogénea do desempenho académico: enquanto existe uma base sólida de estudantes que avança de forma consistente (60–80% de conclusão), coexistem valores preocupantes de insucesso, com destaque para os alunos que não concluem qualquer unidade curricular. Estes resultados sugerem a necessidade de medidas adicionais de acompanhamento, sobretudo direcionadas aos estudantes com menores taxas de progressão, de modo a reduzir o peso da reprovação e promover trajetórias académicas mais lineares.

A Figura 4.6 responde à questão analítica: "Existe relação entre a média de ingresso e a média final obtida pelos alunos em cada ano curricular?". O gráfico apresenta, para cada ano letivo, a comparação entre a média final do ano curricular (representada pelas barras azuis) e a média de ingresso (representada pela linha preta).

Esta visualização permite observar simultaneamente a evolução do desempenho académico efetivo ao longo dos anos curriculares e a forma como este se relaciona com a preparação prévia dos estudantes à entrada do curso.

A comparação entre os dois indicadores sugere que a média de ingresso, embora importante, não é determinante para o desempenho final: anos em que a nota de ingresso é mais elevada não correspondem necessariamente a médias finais superiores. Este resultado reforça a ideia de que fatores como a adaptação ao ensino superior, a motivação individual e o apoio pedagógico desempenham um papel central na consolidação do desempenho académico.

A utilização de um gráfico combinado, com barras para as médias finais e linha para a média de ingresso, é adequada para representar variáveis distintas mas relacionadas, permitindo acompanhar tendências paralelas e realizar comparações diretas entre o perfil de entrada e os resultados académicos obtidos.

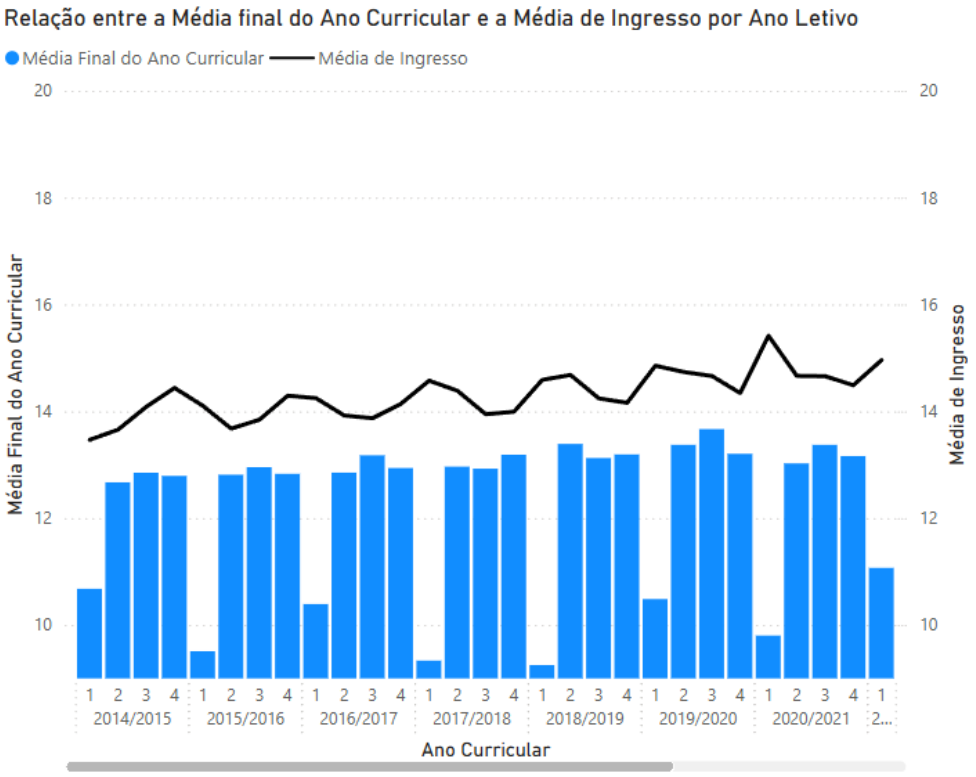


FIGURA 4.6. Relação entre a Média Final do Ano Curricular e a Média de Ingresso por Ano Letivo

O gráfico, da Figura 4.6, apresenta a comparação entre a média final obtida pelos estudantes em cada ano curricular e a média de ingresso dos mesmos, ao longo dos di- ferentes anos letivos. As barras representam as médias finais do 1.º, 2.º, 3.º e 4.º anos curriculares, enquanto a linha preta traduz a evolução da média de ingresso. No todo, observa-se que as médias finais se mantêm relativamente consistentes, situando-se na maior parte dos casos entre os 12 e os 14 valores. Esta estabilidade sugere que, apesar das diferenças no perfil de ingresso ao longo dos anos, a evolução do desempenho académico dos estudantes tende a convergir para um mesmo patamar, independentemente do ano curricular em que se encontram. A média de ingresso, por sua vez, apresenta valores mais elevados, entre 14 e 16 valores, com oscilações mais acentuadas ao longo do período. Contudo, não se verifica uma relação direta entre os dois indicadores: anos em que a média de ingresso é mais elevada não correspondem necessariamente a médias finais superiores. Este padrão confirma que, embora a nota de ingresso constitua um bom indicador da preparação inicial dos alunos, não é determinante para o sucesso académico subsequente. Em particular, os estudantes do 1.º ano curricular continuam a registar valores médios ligeiramente mais baixos do que os dos anos seguintes, o que reflete as dificuldades típicas de adaptação ao ensino superior. Já os alunos do 2.º, 3.º e 4.º anos apresentam desempenhos bastante próximos, com médias consistentes, evidenciando maior estabilidade após a fase inicial de integração. Em síntese, o gráfico demonstra que a nota de ingresso, apesar de relevante no processo de seleção, não explica por completo o desempenho académico ao longo do curso. O ensino superior funciona como um nivelador, atenuando as

diferenças iniciais e promovendo uma relativa homogeneidade de resultados entre os anos curriculares.

4.1.1.1. *Licenciatura em Ciência de Dados*

Tal como no caso da LIGE, foram produzidas visualizações específicas para a LCD, com o objetivo de responder às mesmas questões analíticas colocadas pelos docentes durante o levantamento de requisitos.

As visualizações seguintes respondem, assim, às mesmas perguntas, mas aplicadas ao contexto dos estudantes da LCD. Seguem-se as figuras com as respetivas descrições e interpretações.

Na Figura 4.7, a distribuição dos alunos por curso anterior demonstra que a maioria dos estudantes ingressa a partir da área de Ciências e Tecnologias, que representa 393 alunos. Esta predominância reflete a ligação natural entre a formação científica adquirida no secundário e as exigências do curso superior, traduzindo uma forte correspondência entre perfil de ingresso e percurso académico escolhido.

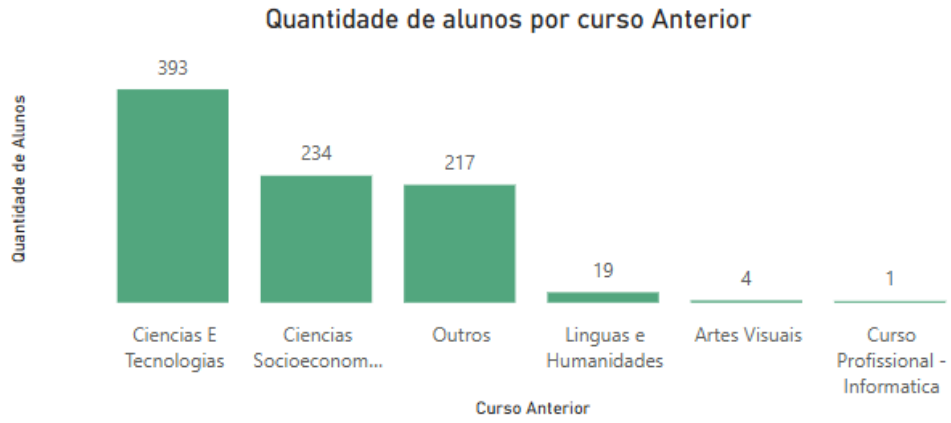


FIGURA 4.7. Distribuição dos alunos da LCD por curso anterior no momento do ingresso

Em segundo lugar surgem os alunos de Ciências Socioeconómicas, com um total de 234 estudantes, demonstrando também uma representatividade significativa. A categoria Outros, com 217 alunos, ocupa a terceira posição, sugerindo uma diversidade de percursos escolares que, embora menos estruturados dentro das vias tradicionais, também contribuem de forma relevante para o ingresso. Já os estudantes provenientes de Línguas e Humanidades são bastante menos expressivos, contabilizando apenas 19 casos, seguidos de Artes Visuais, com 4 alunos, e de Cursos Profissionais de Informática, com apenas 1 aluno. Estas últimas categorias praticamente não têm impacto na composição global do corpo discente, revelando que a ligação entre estes percursos e o curso superior em análise é muito ténue.

Em síntese, o gráfico confirma que a grande maioria dos estudantes provém das áreas de Ciências e Tecnologias e de Ciências Socioeconómicas, enquanto as restantes vias de ensino têm uma contribuição residual. Este padrão reforça a ideia de que o perfil de

ingresso está fortemente concentrado em percursos ligados à ciência e à economia, o que pode influenciar tanto o desempenho académico futuro como as estratégias de captação de novos estudantes.

O gráfico, da Figura 4.8, evidencia diferenças significativas entre as notas de ingresso médias dos estudantes consoante o curso de origem. Os melhores resultados pertencem aos alunos provenientes de Ciências e Tecnologias, que apresentam uma média de 158 valores, muito próxima da registada pelos alunos de Artes Visuais, com 157 valores. Este dado é particularmente interessante, pois, apesar do número reduzido de alunos oriundos de Artes Visuais, os que ingressam demonstram um desempenho de entrada elevado, superando inclusive os de algumas áreas tradicionalmente mais representadas.

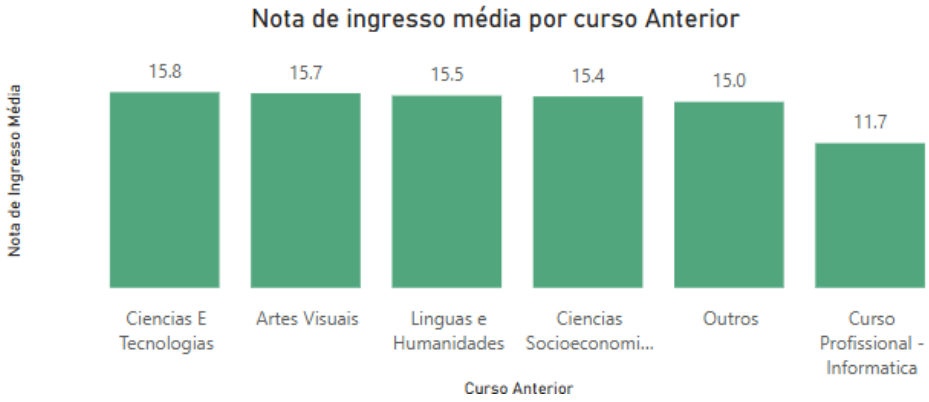


FIGURA 4.8. Nota de ingresso média por curso anterior dos alunos da LCD

Em seguida surgem os estudantes de Línguas e Humanidades, com uma média de 15.5 valores, e os de Ciências Socioeconómicas, com 15.4 valores, ambos também a revelar desempenhos sólidos no momento de ingresso. A categoria Outros apresenta uma média ligeiramente inferior, de 15 valores, mantendo-se ainda assim relativamente próxima das restantes vias principais. No extremo oposto, encontram-se os alunos de Cursos Profissionais de Informática, cuja média de ingresso se situa nos 11.7 valores, uma diferença expressiva face às restantes categorias. Este resultado sugere uma menor competitividade ou preparação académica à entrada destes estudantes, o que pode refletir as especificidades do percurso profissionalizante face às vias científico-humanísticas. De forma geral, os dados confirmam que as áreas de Ciências e Tecnologias e de Artes Visuais concentram os estudantes mais bem posicionados em termos de nota de ingresso, enquanto os cursos profissionais, em particular os ligados à informática, apresentam um desempenho consideravelmente mais baixo.

O mapa, da Figura 4.9, mostra a proveniência geográfica dos estudantes e confirma um padrão já identificado na LIGE: a maioria dos alunos provém da faixa litoral, com especial destaque para a Área Metropolitana de Lisboa, que surge como o principal polo de origem. A concentração de pontos nesta região evidencia a forte procura por parte dos distritos de Lisboa e de Setúbal, o que é coerente com a maior densidade populacional e a proximidade face à instituição.

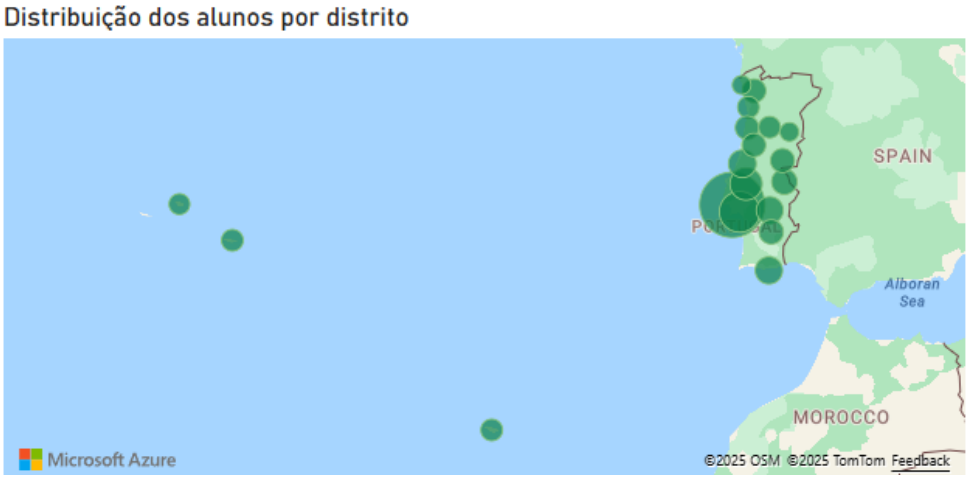


FIGURA 4.9. Distribuição geográfica dos alunos da LCD por distrito de residência

Verifica-se também a presença de estudantes de distritos do centro do país, como Santarém, Leiria e Coimbra, embora em menor número relativamente a Lisboa. Já o interior apresenta uma participação bastante reduzida, sinalizando assimetrias regionais que refletem tanto a distribuição populacional como questões de acessibilidade e mobilidade. As Regiões Autónomas dos Açores e da Madeira estão igualmente representadas, ainda que com um número residual de estudantes. Estes dados, contudo, mostram que mesmo regiões geograficamente mais distantes contribuem para a diversidade da população estudantil. No conjunto, o gráfico revela uma forte concentração no litoral e uma participação residual do interior e das ilhas, reforçando a necessidade de refletir sobre estratégias que promovam uma maior captação de alunos provenientes de zonas menos representadas.

Na Figura 4.10, entre 2019/2020 e 2023/2024, verifica-se uma evolução marcada por variações significativas nas notas médias de ingresso. O ano letivo de 2019/2020 apresenta o valor mais baixo do período, próximo dos 151 valores, mas logo em 2020/2021 ocorre uma subida acentuada, atingindo cerca de 157 valores.

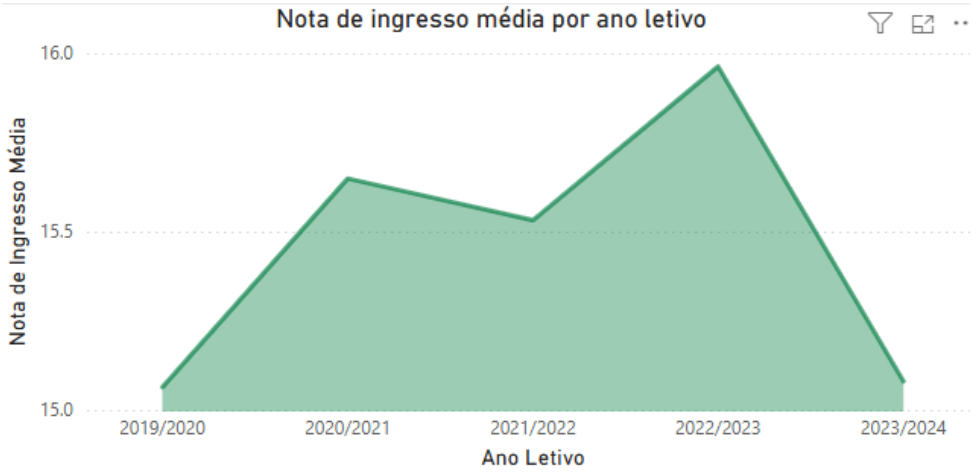


FIGURA 4.10. Evolução da nota de ingresso média dos alunos da LCD por ano letivo

No ano seguinte, 2021/2022, regista-se uma ligeira descida, situando a média em torno dos 155 valores, mas em 2022/2023 observa-se o pico mais elevado de toda a série, com a nota média a aproximar-se dos 160 valores. Esta subida traduz um momento de forte competitividade no acesso, possivelmente influenciado por uma maior procura do curso ou por critérios de ingresso mais exigentes.

Por fim, em 2023/2024 regista-se uma quebra significativa, com a média a voltar a níveis próximos dos observados em 2019/2020, cerca de 151 valores. Esta oscilação demonstra que, apesar de existir uma tendência de valorização das notas de ingresso ao longo dos anos, fatores conjunturais continuam a exercer um impacto direto no perfil dos candidatos admitidos.

O gráfico revela um cenário de instabilidade recente, com uma trajetória ascendente até 2022/2023, seguida de uma descida abrupta em 2023/2024, o que pode estar associado à diminuição da procura ou a alterações no processo de candidatura e seleção.

O gráfico empilhado, da Figura 4.11, mostra a evolução do desempenho académico entre 2019/2020 e 2023/2024, considerando as categorias de conclusão do percurso académico (100%, 60–80%, 30–50%, <30% e 0%).

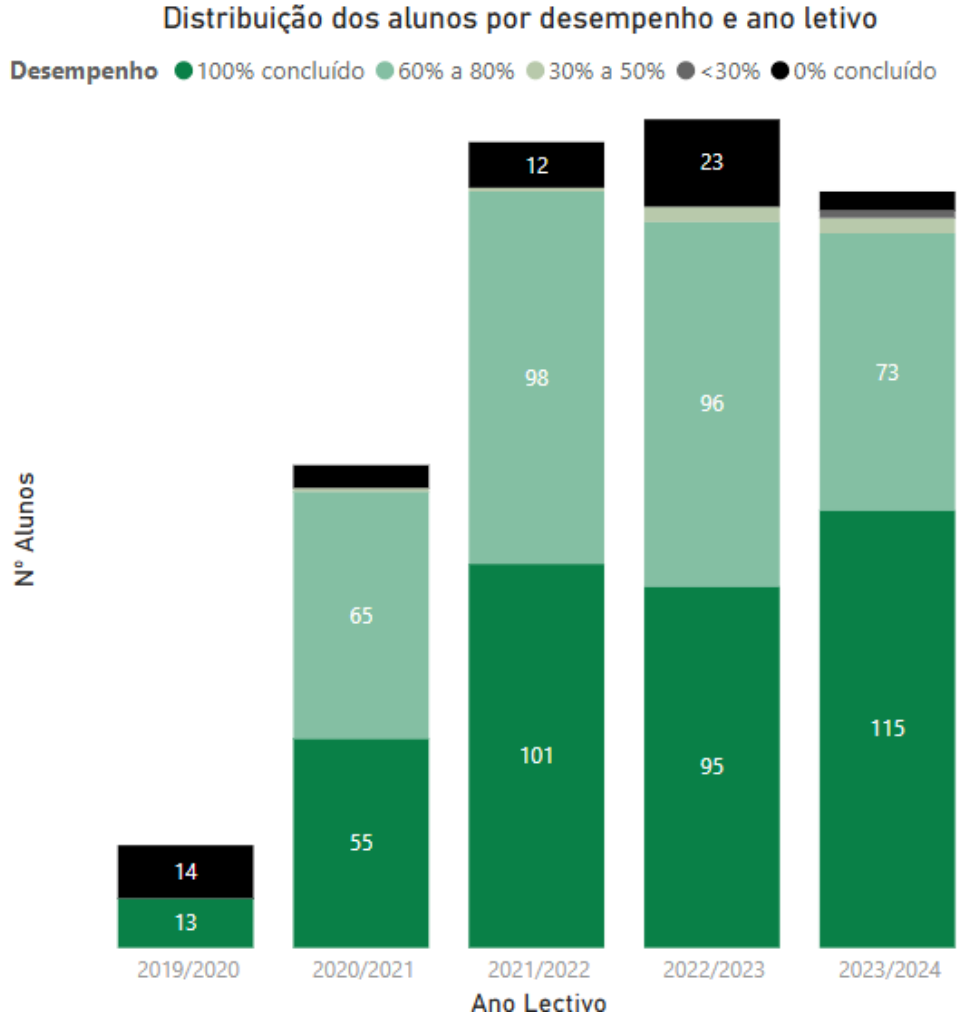


FIGURA 4.11. Distribuição dos alunos da LCD por desempenho e ano letivo

Em 2019/2020, o número de estudantes era bastante reduzido, repartindo-se maioritariamente entre os que não concluíram nenhuma unidade curricular (14 alunos) e os que atingiram 100% de conclusão (13 alunos), revelando uma polarização entre sucesso total e insucesso absoluto.

Já em 2020/2021, verifica-se um aumento expressivo do número total de estudantes, com destaque para as categorias intermédias. A maior parte concluiu entre 60% e 80% do percurso (65 alunos), enquanto um número significativo conseguiu concluir integralmente (55 alunos). O grupo sem qualquer aproveitamento foi bastante pequeno, sinalizando uma melhoria global no desempenho.

Nos anos seguintes, 2021/2022 e 2022/2023, o número de alunos manteve-se elevado, e observa-se uma estabilização nas categorias dominantes: entre 95 e 101 alunos concluíram integralmente (100%), enquanto entre 96 e 98 alunos concluíram entre 60% e 80%. Contudo, surgem novamente casos de insucesso absoluto, com 12 alunos em 2021/2022 e 23 em 2022/2023, confirmando que a taxa de reprovação, embora minoritária, continua a existir.

Por fim, em 2023/2024, os resultados apontam para um aumento do número de estudantes que concluem 100% do percurso (115 alunos), mantendo-se ainda um contingente expressivo nas categorias de 60–80% de conclusão (73 alunos). Os grupos de insucesso total ou de baixas taxas de progressão são, neste ano, residuais, o que indica um cenário de desempenho globalmente mais positivo.

De forma geral, o gráfico demonstra uma evolução muito favorável ao longo do tempo, com crescimento contínuo do número de estudantes que atingem a conclusão total e diminuição dos casos de insucesso absoluto. Esta tendência sugere não apenas uma melhor preparação dos alunos à entrada, como também possíveis melhorias no acompanhamento académico e nas condições de progressão ao longo do curso.

O gráfico, da Figura 4.12, relaciona a média final do ano curricular com a média de ingresso por ano letivo e ano curricular. Oferece uma perspetiva sobre a articulação entre o desempenho prévio dos estudantes e os resultados efetivamente obtidos ao longo do percurso académico. As barras verdes evidenciam uma tendência de relativa estabilidade das médias finais, situando-se, na maioria dos casos, entre os 13 e os 15 valores, independentemente do ano letivo ou do ano curricular considerado. Esta consistência sugere que o desempenho dos estudantes no ensino superior tende a convergir para um patamar comum, ainda que o ponto de partida em termos de nota de ingresso apresente alguma variabilidade.

A linha preta, que representa a média de ingresso, demonstra oscilações mais significativas, com valores a variar entre cerca de 150 e 160 valores ao longo do período em análise. No entanto, não se observa uma relação direta entre os dois indicadores: anos letivos em que a média de ingresso é mais elevada não correspondem necessariamente a médias finais superiores. Tal constatação permite inferir que, embora a nota de ingresso

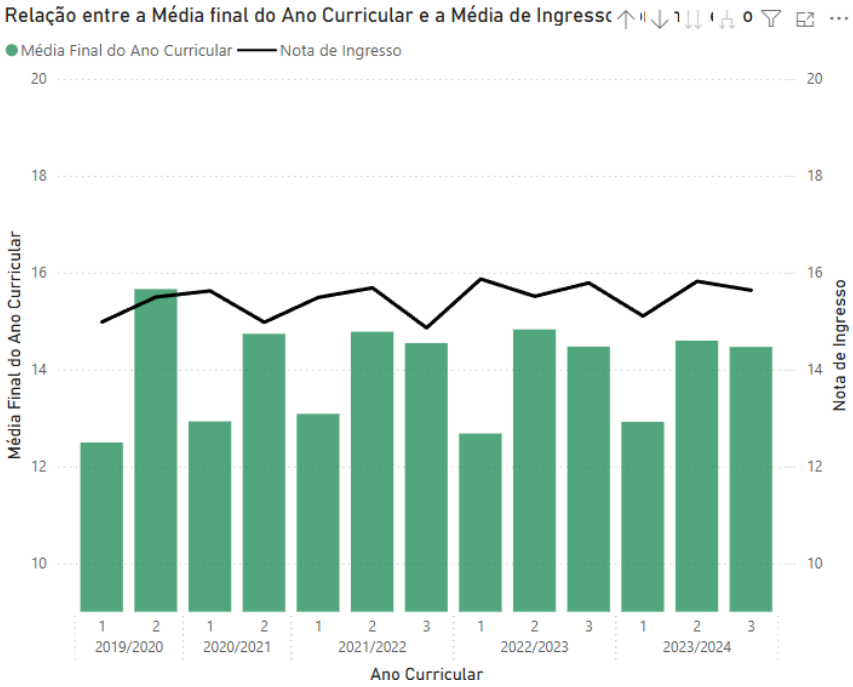


FIGURA 4.12. Relação entre a Média Final do Ano Curricular e a Média de Ingresso por Ano Letivo

constitua um critério importante de seleção e reflita, em certa medida, a preparação inicial dos estudantes, ela não se revela determinante para o desempenho académico subsequente.

Esta leitura aponta para a relevância de fatores adicionais que influenciam o percurso académico, tais como a adaptação às exigências do ensino superior, a qualidade do acompanhamento pedagógico, a motivação individual ou as condições de suporte institucional. Em síntese, o gráfico evidencia que o ensino superior funciona como um nivelador do desempenho académico, atenuando as diferenças observadas no momento do ingresso e conduzindo a resultados finais relativamente homogéneos entre os diferentes anos curriculares.

4.1.2. Funcionalidades Interativas e Filtros Dinâmicos

Um dos elementos mais relevantes do *dashboard* é a possibilidade de realizar análises segmentadas de forma dinâmica. Para isso, foram implementados diversos filtros interativos, como a seleção por ano letivo, estatuto de trabalhador-estudante, número total de alunos na análise, menu *HOME* que dispõe dos botões de navegação entre dashboards, botão *HOME* que direciona a navegação para o menu inicial e setas de navegação. Estes filtros permitem que o utilizador explore os dados de forma personalizada, adaptando as visualizações aos seus interesses específicos.

A Figura 4.13 mostra o menu inicial do dahsboard, *home*, com os botões que permitem aceder às análises de ingresso e de progresso de cada curso.

A segmentação por estatuto de trabalhador-estudante, por exemplo, permite isolar este grupo e compará-lo com os restantes estudantes. A Figura 4.14 mostra o botão



FIGURA 4.13. Menu Inicial (HOME)

que permite realizar o filtro descrito. Foi considerado prioritário pelos docentes, dada a necessidade de compreender melhor as especificidades deste perfil académico.

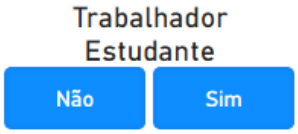


FIGURA 4.14. Filtro de trabalhador-estudante

O filtro do Ano Letivo constitui uma funcionalidade essencial para a análise detalhada da informação, na medida em que permite restringir a análise a um período temporal específico ou, em alternativa, selecionar vários anos em simultâneo. A Figura 4.15 mostra os vários botões que permitem realizar os filtros descritos. Com isto, o utilizador pode optar por realizar uma análise pormenorizada de um único ano letivo, centrando a atenção nos indicadores correspondentes a este período, ou proceder a uma comparação entre diferentes anos, identificando padrões, tendências e variações ao longo do tempo. Esta flexibilidade garante uma leitura mais rica e contextualizada dos dados, adaptando a visualização às necessidades concretas da análise em curso.



FIGURA 4.15. Filtro do Ano letivo

4.1.3. Utilidade para a Tomada de Decisão

O *dashboard* final representa uma ferramenta de suporte importante para os docentes e para as direções de curso. Através da sua utilização, torna-se possível identificar padrões de sucesso académico que estão associados a áreas de estudo, notas de ingresso e

características demográficas, o que permite compreender melhor os fatores que influenciam o percurso dos estudantes. Para além disso, possibilita monitorizar o desempenho de grupos específicos, como é o caso dos trabalhadores-estudantes ou dos grupos de estudantes de determinado ano letivo, facilitando uma análise direcionada a realidades particulares. O *dashboard* permite ainda avaliar a adequação de determinados perfis de ingresso ao contexto curricular dos cursos, contribuindo para perceber se existem desalinhamentos ou dificuldades específicas.

A informação disponibilizada serve também de base para fundamentar decisões relacionadas com estratégias de acompanhamento, definição de critérios de ingresso ou implementação de medidas de apoio que respondam de forma mais eficaz às necessidades dos estudantes. Neste sentido, o *dashboard* cumpre plenamente o seu papel como instrumento de apoio à decisão, combinando rigor analítico com acessibilidade visual e interatividade, tornando-se um recurso essencial para a gestão e o planeamento académico.

A análise realizada através do *dashboard* permitiu retirar um conjunto de conclusões úteis para a gestão pedagógica e estratégica das licenciaturas em estudo. Os dados mostram, de forma clara, que os estudantes oriundos de cursos secundários mais próximos da área das TIC e da Gestão apresentam melhores resultados, quer no ingresso, quer no percurso académico. Esta evidência reforça a importância da correspondência entre o perfil de entrada e as exigências curriculares do curso, confirmando que a preparação prévia desempenha um papel decisivo na progressão dos alunos.

Outro aspeto relevante prende-se com a distribuição geográfica. Verifica-se uma forte concentração de estudantes nos distritos do litoral, sobretudo na Área Metropolitana de Lisboa, enquanto os distritos do interior e as Regiões Autónomas estão pouco representados. Esta realidade deve ser considerada na definição de estratégias de divulgação, no sentido de diversificar a origem dos candidatos e alargar a base de recrutamento.

No que diz respeito ao desempenho académico, destaca-se o 1.º ano curricular como o momento de maior fragilidade. As médias mais baixas e a maior variabilidade dos resultados nesta fase mostram que a transição para o ensino superior continua a ser um desafio para muitos alunos. Este dado aponta para a necessidade de medidas específicas de integração e apoio pedagógico logo no início do percurso, capazes de reduzir as taxas de insucesso e de abandono.

A comparação entre as médias de ingresso e as médias finais mostra ainda que não existe uma relação direta entre estes dois indicadores. Anos em que a média de ingresso é mais elevada não correspondem necessariamente a melhores resultados finais. Isto demonstra que, para além da seleção inicial, fatores como a adaptação ao ensino superior, a motivação individual e o acompanhamento prestado pela instituição têm um impacto decisivo no sucesso académico.

Em síntese, o *dashboard* assume-se como um instrumento de apoio à decisão que vai para além da simples visualização de dados. Ao proporcionar uma visão integrada e atualizada sobre o percurso dos estudantes, oferece bases sólidas para decisões estratégicas

relacionadas com o acompanhamento, a gestão pedagógica e a definição de políticas de captação e apoio.

4.2. Discussão dos Resultados

Os resultados obtidos a partir da análise exploratória e do *dashboard* desenvolvido confirmam, em grande medida, os padrões identificados na literatura sobre desempenho académico, perfil dos estudantes e fatores determinantes do sucesso no ensino superior.

4.2.1. Relação entre Perfil de Ingresso e Desempenho Académico

Os resultados obtidos confirmam que o perfil de ingresso dos estudantes constitui um fator relevante para compreender o seu desempenho ao longo do percurso académico. A análise evidenciou que os estudantes provenientes das áreas de Ciências e Tecnologias e Ciências Socioeconómicas representam a maioria dos ingressos e apresentam, em média, notas de entrada mais elevadas, refletindo uma preparação de base mais sólida. Em contrapartida, os alunos oriundos de cursos profissionais, sobretudo fora da área da informática, chegam ao ensino superior com notas de ingresso mais baixas, o que se traduz numa maior vulnerabilidade académica, em particular no primeiro ano curricular.

Apesar destas diferenças iniciais, as médias finais observadas revelam uma tendência para a convergência do desempenho ao longo dos anos. Ou seja, ainda que as notas de ingresso indiquem o nível de preparação inicial, o contexto de ensino superior funciona como um nivelador, permitindo que estudantes com trajetórias distintas atinjam desempenhos relativamente próximos à medida que avançam no curso. Esta constatação está em linha com diversos estudos da literatura [4], [10], que reconhecem as notas de ingresso como preditores significativos do sucesso académico, mas não exclusivos.

O primeiro ano curricular surge, contudo, como o período mais crítico, com médias mais baixas e maior variabilidade entre estudantes, refletindo as dificuldades associadas à adaptação ao ensino superior. Este padrão, também identificado por [9], [11], reforça a necessidade de políticas institucionais orientadas para um acompanhamento mais próximo neste momento de transição, mitigando o impacto das fragilidades de preparação prévia.

Em síntese, o perfil de ingresso influencia de forma clara o desempenho inicial, mas tende a perder peso ao longo do percurso, uma vez que fatores como a adaptação, o apoio pedagógico, a motivação e as condições institucionais assumem um papel crescente na explicação do sucesso académico. Esta evidência sublinha a importância de utilizar a informação de ingresso não apenas como critério de seleção, mas sobretudo como instrumento de diagnóstico para orientar estratégias de apoio diferenciadas e promover a equidade no ensino superior.

4.2.2. Impacto do Estatuto de Trabalhador-Estudante

A análise do impacto do estatuto de trabalhador-estudante revela diferenças no percurso académico quando comparado com os estudantes a tempo inteiro. Os dashboards permitem observar que, LIGE como na LCD, os trabalhadores-estudantes representam

uma proporção significativa da população estudantil, o que reforça a importância de compreender como este estatuto influencia os resultados.

Nos gráficos de desempenho por ano letivo, verifica-se que os estudantes sem estatuto apresentam maior concentração nos intervalos de conclusão entre 60% e 100% das unidades curriculares, evidenciando trajetórias mais consistentes e lineares ao longo do tempo. Em contrapartida, entre os trabalhadores-estudantes é mais visível a presença nas categorias de menor conclusão ($< 30\%$ ou 0%), sinalizando maiores dificuldades em manter um ritmo regular de aprovação.

A relação entre a nota de ingresso e a média final do ano curricular também ilustra esta diferença. Para os estudantes sem estatuto, as médias tendem a alinhar-se com as notas de ingresso, mostrando uma continuidade do desempenho. Já entre os trabalhadores-estudantes, nota-se um desfasamento maior, sobretudo nos primeiros anos, em que a exigência de conciliar trabalho com estudo se traduz num impacto mais evidente na progressão académica.

Estes resultados confirmam aquilo que a literatura tem vindo a evidenciar: embora o estatuto de trabalhador-estudante permita maior flexibilidade administrativa e académica, ele não elimina os constrangimentos associados à dupla condição (profissional e académica), refletindo-se em trajetórias mais heterogéneas e, por vezes, mais longas. No entanto, é igualmente visível que, em alguns casos, a experiência profissional contribui para consolidar competências que permitem aos estudantes estabilizar e recuperar o desempenho em fases mais avançadas do curso.

Assim, o impacto do estatuto de trabalhador-estudante deve ser entendido não apenas como um fator de risco para o insucesso e a desistência, mas também como um indicador para a necessidade de políticas de acompanhamento diferenciadas. Medidas como a flexibilização na avaliação contínua, o reforço do ensino híbrido e a promoção de tutorias direcionadas podem mitigar as dificuldades identificadas e potenciar trajetórias académicas mais próximas das dos colegas sem estatuto.

4.2.3. Distribuição Geográfica e Diversidade de Perfis

A análise da distribuição geográfica dos estudantes evidencia uma forte concentração de origem nos distritos do litoral, em particular na Área Metropolitana de Lisboa, que surge como o principal polo de proveniência. Esta tendência é explicada não apenas pela densidade populacional da região, mas também pela proximidade geográfica à instituição, que constitui um fator determinante no processo de escolha do curso [10]. Outros distritos da região centro, como Setúbal, Santarém, Leiria e Coimbra, também se encontram representados, ainda que em menor escala.

Em contraste, os distritos do interior apresentam valores significativamente mais baixos, refletindo assimetrias territoriais que se relacionam tanto com a menor densidade populacional como com questões de acessibilidade e mobilidade. A presença de estudantes provenientes das Regiões Autónomas dos Açores e da Madeira, embora residual, confirma a abrangência nacional do curso e acrescenta diversidade ao perfil dos ingressos [15].

Para além da dimensão geográfica, os dados apontam para uma diversidade considerável de percursos de formação anteriores. A maioria dos estudantes ingressa a partir das áreas de Ciências e Tecnologias ou Ciências Socioeconómicas, o que corresponde às vias de ensino secundário com maior expressão nacional e maior alinhamento com os requisitos curriculares do curso [1]. Contudo, observa-se também a presença, ainda que em menor número, de estudantes oriundos de percursos como Línguas e Humanidades, Artes Visuais e Cursos Profissionais. Esta heterogeneidade contribui para a pluralidade de perfis dentro da população estudantil, mas também reforça a necessidade de estratégias de acompanhamento diferenciadas, de forma a apoiar grupos que chegam com preparação menos próxima da área científica dominante [8], [21].

Assim, a leitura conjunta da distribuição geográfica e da diversidade de perfis de ingresso permite concluir que, embora o curso consiga captar estudantes de diferentes regiões e formações, persiste uma forte concentração no litoral e uma prevalência clara das vias científico-tecnológicas e socioeconómicas. Este cenário sublinha a importância de políticas de divulgação e de captação mais ativas em territórios do interior e nas regiões autónomas, como também a implementação de mecanismos de integração académica que garantam equidade de oportunidades de sucesso entre perfis de ingresso distintos.

4.2.4. Comparação com Estudos de Visualização e Apoio à Decisão e Implicações Estratégicas

O *dashboard* interativo desenvolvido neste projeto cumpre efetivamente o objetivo de suporte à decisão, ao permitir a exploração visual de métricas académicas e a identificação de padrões notórios. Tal como evidenciado em estudos de visualização aplicada ao contexto educativo [7] e em investigações sobre sistemas de *dashboards* como instrumentos de apoio à decisão [6], a utilização de ferramentas interativas facilita a interpretação de fenómenos complexos e democratiza o acesso à informação por parte de diferentes atores institucionais, como docentes, coordenadores de curso ou gabinetes de apoio ao estudante.

A presença de filtros dinâmicos — como o estatuto académico, o curso de origem ou o ano de ingresso — garante flexibilidade e adaptabilidade da ferramenta, o que reforça o seu valor enquanto instrumento de gestão pedagógica. Esta capacidade de ajustar a análise em função das necessidades específicas aproxima-se do conceito de visualização centrada no utilizador, destacada na literatura recente sobre learning analytics e educational data mining [9], [10].

A análise realizada não só validou tendências previamente identificadas na literatura como também trouxe novos insights para a gestão académica e institucional. Entre as implicações mais expressivas, destacam-se três eixos fundamentais: (i) a relevância do curso de origem como variável preditora do desempenho académico no ensino superior, em linha com investigações sobre fatores de ingresso e sucesso estudantil [8], [21]; (ii) a necessidade de medidas específicas de apoio a trabalhadores-estudantes, cuja trajetória académica tende a ser mais heterogénea, exigindo políticas de acompanhamento diferenciadas para promover a equidade no acesso e no sucesso; (iii) a utilidade de dashboards

interativos como ferramentas de acompanhamento contínuo do progresso académico e de monitorização de indicadores-chave, permitindo às instituições não apenas compreender o passado e o presente dos estudantes, mas também antecipar cenários futuros e agir preventivamente.

Deste modo, esta dissertação contribui para preencher lacunas identificadas na revisão da literatura, ao propor uma abordagem prática e orientada à realidade institucional do ensino superior. Em síntese, os resultados obtidos alinham-se com os principais achados da investigação científica, mas acrescentam perspetivas adicionais orientadas à personalização das estratégias pedagógicas e à otimização da gestão curricular. O uso de visualizações interativas reforça-se, assim, como uma estratégia que não apenas apoia a leitura retrospetiva dos dados, mas também fundamenta decisões mais eficazes e informadas, sustentando a construção de um modelo de ensino superior mais adaptado, inclusivo e eficiente.

Em síntese, os resultados obtidos alinham-se com os principais achados da literatura, mas também oferecem perspetivas adicionais para a personalização de estratégias pedagógicas e de gestão. O uso de visualizações interativas permite, não só compreender o passado e o presente, mas também construir cenários preditivos para apoiar decisões mais informadas no futuro.

CAPíTULO 5

Conclusão

5.1. Síntese do Trabalho Desenvolvido

Este trabalho analisou, de forma sistemática, a relação entre o perfil de ingresso e o desempenho académico no ensino superior, com foco nos cursos de LIGE e LCD do ISCTE. Para tal, foi consolidada uma base de dados institucional que integra variáveis demográficas, notas de ingresso, cursos/áreas de origem, médias por ano curricular, taxas de conclusão de créditos e estatutos académicos, incluindo o estatuto de trabalhador-estudante. Sobre esta base, foi concebido e implementado um *dashboard* interativo em Power BI, pensado como instrumento de apoio à decisão para docentes e órgãos de gestão.

A condução do projeto seguiu a lógica da metodologia CRISP-DM, garantindo rastreabilidade entre objetivos e resultados [5]. Na fase de compreensão do problema, foram definidos, com os utilizadores finais, os principais *use cases*: caracterizar notas de ingresso e sua distribuição; mapear proveniências geográficas e percursos de origem; estudar a relação entre ingresso, médias e créditos concluídos; e avaliar o impacto do estatuto de trabalhador-estudante na progressão. Na compreensão e preparação de dados, procedeu-se à normalização de escalas (e.g., transformação de notas de ingresso para a escala 0–20 quando necessário), ao tratamento sistemático de omissões (preenchimento por mediana/média em variáveis numéricas e categorias informativas nas variáveis nominais) e à harmonização semântica de categorias de curso de origem (agregação em *Ciências e Tecnologias*, *Ciências Socioeconómicas*, *Línguas e Humanidades*, *Artes Visuais*, *Curso Profissional – Informática* e *Outros*), assegurando comparabilidade analítica entre anos e coortes de ingresso.

A fase de modelação, por opção metodológica, privilegiou a análise exploratória e a construção de indicadores derivados (médias por ano curricular, rácios de conclusão, cortes por estatuto e por origem) em detrimento de modelos preditivos avançados, reservando estes últimos para trabalho futuro. A avaliação focou-se na utilidade prática para os utilizadores e na consistência dos resultados face à literatura em *learning analytics* e *educational data mining*, com especial atenção a padrões recorrentes reportados em contextos comparáveis [10].

O *dashboard* final disponibiliza múltiplas vistas coordenadas e filtros dinâmicos (ano letivo, ano curricular, estatuto, curso de origem, localização geográfica), permitindo *drill-down* do agregado institucional para grupos específicos e, destes, para métricas detalhadas. As escolhas de visualização seguiram princípios de legibilidade e análise temporal/relacional: séries para evolução anual de médias e notas; barras empilhadas para proporções de conclusão por grupo; mapas e tabelas para distribuição territorial; *small*

multiples para comparar anos curriculares. Esta abordagem alinha-se com a evidência de que visualizações interativas facilitam a leitura de fenómenos complexos e democratizam o acesso à informação para diferentes perfis de decisores [6], [7].

Do ponto de vista substantivo, o trabalho consolidou resultados centrais: (i) o curso/área de origem é informativo sobre o desempenho inicial, mas o efeito tende a atenuar-se ao longo do percurso; (ii) o 1.º ano curricular concentra maior variabilidade e risco, confirmando a necessidade de medidas de integração; (iii) o estatuto de trabalhador-estudante está associado a trajetórias mais heterogéneas, mas com evidência de disciplina e recuperação em fases posteriores; (iv) a origem geográfica mostra concentração em distritos do litoral, com implicações para estratégias de recrutamento e apoio. Para além de confirmar tendências referidas na literatura, o contributo distintivo deste trabalho reside na operacionalização institucional: transformar dados dispersos em um sistema de monitorização contínua, utilizável por docentes e direções, que liga diretamente indicadores a decisões pedagógicas e de gestão [7], [10].

De uma forma complementar, a revisão sistemática da literatura evidenciou a crescente utilização de técnicas de EDM e *Learning Analytics* no ensino superior para apoiar processos de decisão e otimizar a gestão académica. Vários estudos analisados [3], [4], [8], [11] confirmam que algoritmos preditivos e modelos baseados em dados podem identificar precocemente fatores de risco, apoiar a seleção de candidatos mais bem preparados e melhorar o acompanhamento de estudantes ao longo do seu percurso. A literatura também destaca que *dashboards* interativos e sistemas de visualização [6], [7], [22] ajudam a melhorar o acesso democrático à informação e também facilitam a leitura de fenómenos complexos. Por outro lado, a maioria das investigações salienta que, embora as notas de ingresso e os dados demográficos sejam preditores relevantes do desempenho inicial, o seu impacto tende a diluir-se ao longo do percurso académico, dando lugar a variáveis comportamentais e contextuais [9], [19], [21]. De forma transversal, os trabalhos revistos apontam para a necessidade de integrar sistemas institucionais dispersos e adotar metodologias maduras, como a CRISP-DM [5], para garantir rastreabilidade, reprodutibilidade e explicabilidade dos modelos. Esta evidência fundamenta a opção metodológica desta dissertação, centrada na preparação rigorosa dos dados, no desenho de indicadores derivados e na criação de um *dashboard* interativo como passo fundamental para o apoio à tomada de decisão.

Em síntese, a dissertação entrega através de conjunto de dados preparados, um *dashboard* interativo orientado a perguntas concretas de gestão académica e uma base analítica que comprova a mais-valia de integrar EDA e visualização para leitura rápida de padrões e identificação de grupos-alvo. Este resultado cria as condições para a próxima etapa: incorporar componentes preditivas e mecanismos de alerta precoce, mantendo a experiência de utilização e a explicabilidade como critérios centrais.

5.2. Principais Conclusões

A exploração dos dados académicos e a interação com o *dashboard* permitiram formular um conjunto de conclusões robustas, com relevância científica e institucional.

Em primeiro lugar, confirmou-se que os perfis de ingresso funcionam como preditores do desempenho académico. Estudantes com notas de ingresso mais elevadas e oriundos de cursos científico-tecnológicos ou socioeconómicos tendem a apresentar melhores resultados no início da licenciatura, confirmando padrões já identificados na literatura [15], [21]. No entanto, o efeito não é determinista: foram observados casos de estudantes com notas mais baixas que alcançaram médias finais superiores, sinalizando que fatores não captados na base de dados — como motivação, disciplina individual ou apoio institucional — também desempenham um papel significativo.

Em segundo lugar, destacou-se o impacto do estatuto de trabalhador-estudante. Contrariando uma visão comum de maior fragilidade, estes estudantes revelaram, em média, taxas de conclusão de créditos iguais ou superiores às dos colegas a tempo inteiro, particularmente a partir do segundo ano. Este resultado sugere que a experiência profissional pode contribuir para a aquisição de competências de organização e resiliência, ainda que à custa de trajetórias menos lineares [10]. A constatação reforça a importância de políticas diferenciadas de apoio, que tirem partido deste perfil sem negligenciar os desafios adicionais da conciliação trabalho-estudo.

Em terceiro lugar, a análise dos cursos de origem revelou diferenças estruturais entre as duas licenciaturas estudadas. A LIGE apresenta uma maior concentração de estudantes oriundos de áreas diretamente ligadas à informática e à gestão, enquanto a LCD capta uma diversidade maior de proveniências, incluindo percursos menos tradicionais. Este resultado evidencia que a Ciência de Dados, enquanto área emergente, tem um apelo transversal, mas exige mecanismos de nivelamento que mitiguem assimetrias na preparação de base.

Em quarto lugar, a distribuição geográfica mostrou um padrão de concentração acentuada na Área Metropolitana de Lisboa e nos distritos de Setúbal e Santarém, com menor representatividade das regiões do interior. Este fenómeno confirma as assimetrias territoriais no acesso ao ensino superior já reportadas em estudos nacionais e internacionais [9]. Tal evidencia a necessidade de estratégias de captação mais inclusivas e de medidas de apoio específicas para estudantes deslocados.

Por último, a análise temporal permitiu identificar estabilidade relativa nas médias de ingresso e de desempenho ao longo dos últimos cinco anos. Esta consistência sugere que, apesar de variações conjunturais, existem padrões estruturais persistentes que caracterizam estas licenciaturas. O dado é importante porque indica que os fenómenos observados não são meramente circunstanciais, mas refletem tendências institucionais consolidadas.

Em síntese, as conclusões deste trabalho apontam para três eixos fundamentais: (i) o perfil de ingresso é preditivo mas não determinista, devendo ser complementado por dados

qualitativos e contextuais; (ii) o estatuto de trabalhador-estudante constitui simultaneamente um desafio e uma oportunidade, exigindo políticas ajustadas; (iii) a concentração geográfica e a diversidade de percursos de origem obrigam a repensar estratégias de integração e apoio académico. Estes resultados, para além de alinharem com a literatura em *learning analytics* e *educational data mining*, oferecem contributos práticos diretos à gestão académica do ISCTE.

5.3. Principais Contribuições do Trabalho

O presente estudo oferece contributos de grande interesse tanto no plano prático-institucional como no plano científico, reforçando a importância da análise exploratória de dados e da visualização interativa no ensino superior.

No plano prático, a principal contribuição reside no desenvolvimento de um *dashboard* interativo que disponibiliza uma visão integrada sobre múltiplos indicadores de percurso académico. Através de filtros dinâmicos e visualizações intuitivas, a ferramenta permite que docentes, coordenadores de curso e órgãos de gestão acedam de forma rápida e flexível a informação crítica para a tomada de decisão. Ao democratizar o acesso aos dados, o *dashboard* torna-se um instrumento de suporte contínuo à monitorização do desempenho e à identificação de padrões emergentes, alinhando-se com a tendência internacional de utilização de ferramentas de apoio à decisão em contexto educacional [6], [7].

No plano estratégico, o trabalho fornece uma base empírica para a definição de políticas institucionais mais direcionadas. A confirmação da relevância do curso de origem e das notas de ingresso como preditores de desempenho académico permite repensar mecanismos de orientação e nivelamento inicial, com vista a reduzir desigualdades de preparação. Do mesmo modo, os resultados relativos ao estatuto de trabalhador-estudante oferecem evidência concreta para fundamentar políticas diferenciadas de apoio, desde modalidades flexíveis de avaliação a estratégias de acompanhamento personalizado [10].

Do ponto de vista científico, esta dissertação contribui para a literatura em *educational data mining* e *learning analytics*, demonstrando como metodologias exploratórias podem ser operacionalizadas em contextos institucionais reais. A integração de dados administrativos com técnicas de análise e visualização não só confirma achados de estudos prévios sobre predição de desempenho e abandono [15], [21], como acrescenta valor pela vertente aplicada: a transformação dos dados em informação acionável para gestores e docentes. Além disso, o estudo reforça a relevância da visualização interativa como meio de comunicação de insights complexos a públicos não técnicos, ampliando a aplicabilidade das abordagens de ciência de dados ao contexto académico.

Em síntese, o trabalho deixa três contributos principais: (i) a conceção de uma ferramenta prática de suporte à decisão para a gestão académica; (ii) a produção de evidência que sustenta a definição de políticas institucionais mais inclusivas e equitativas; e (iii) o reforço da literatura científica sobre análise exploratória e visualização de dados no ensino

superior. Ao conjugar estas dimensões, a dissertação posiciona-se como um contributo útil tanto para a comunidade académica como para a prática de gestão educacional.

5.4. Limitações do Estudo

Apesar das contribuições alcançadas, importa reconhecer um conjunto de limitações que condicionam a generalização e o alcance dos resultados.

A primeira limitação prende-se com a natureza e o âmbito da base de dados utilizada. Os dados analisados dizem respeito exclusivamente ao ISCTE e a dois cursos de licenciatura em particular LIGE e LCD). Esta delimitação garante consistência e comparabilidade interna, mas restringe a validade externa dos resultados. Fenómenos como padrões de ingresso, impacto do estatuto de trabalhador-estudante ou distribuição geográfica podem manifestar-se de forma distinta em outras instituições ou áreas disciplinares, o que limita a possibilidade de extrapolação direta [10].

A segunda limitação decorre da ausência de variáveis qualitativas ou contextuais. O estudo centrou-se em dimensões administrativas — notas de ingresso, cursos de origem, médias, créditos concluídos, estatutos — sem contemplar fatores como motivação, hábitos de estudo, envolvimento em atividades extracurriculares, condições socioeconómicas ou apoio familiar. Estes fatores são amplamente reconhecidos pela literatura como determinantes no desempenho académico [9], [21], e a sua ausência pode explicar casos em que estudantes contrariaram as tendências esperadas (e.g., notas de ingresso baixas mas médias finais elevadas).

A terceira limitação refere-se à qualidade e consistência dos dados textuais registados nos sistemas institucionais. Em várias variáveis, como a identificação do distrito ou do curso de origem, observaram-se entradas incorretas ou pouco padronizadas, incluindo situações em que eram introduzidos nomes de países em campos destinados a distritos ou múltiplas variantes ortográficas para a mesma categoria. Este tipo de ruído obrigou a um esforço adicional de normalização e pode ter introduzido enviesamentos na análise. A ausência de mecanismos de validação automática — como listas pré-definidas de categorias ou restrições de entrada — compromete a uniformidade da informação recolhida e aumenta a probabilidade de erro. Esta constatação reforça a importância de implementar sistemas de recolha de dados mais robustos, com regras de validação integradas, que assegurem consistência semântica e minimizem o trabalho de limpeza subsequente.

A quarta limitação decorre da abordagem metodológica. A dissertação privilegiou a análise exploratória e descritiva, com recurso a estatísticas simples e visualizações interativas, mas não integrou modelos preditivos mais sofisticados, como regressões multivariadas ou algoritmos de *ML*. Embora esta opção tenha sido deliberada visando a usabilidade e a transparência da ferramenta, limita a capacidade de identificar relações complexas entre variáveis e de antecipar trajetórias de risco de forma automatizada.

Por último, importa referir que o *dashboard* foi concebido como protótipo académico, com validação essencialmente técnica e conceptual. Não foram conduzidos testes formais de usabilidade junto de diferentes perfis de utilizadores, nem foi avaliada a sua integração

efetiva nos processos de gestão institucional. A ausência desta validação prática constitui uma limitação para aferir o impacto real da ferramenta no quotidiano académico.

Resumindo, as limitações identificadas não invalidam os resultados obtidos, mas devem ser entendidas como pontos de atenção e, simultaneamente, como oportunidades para investigação e desenvolvimento futuros.

5.5. Trabalho Futuro

As limitações identificadas abrem espaço para várias linhas de investigação e desenvolvimento que podem reforçar e expandir os contributos deste estudo.

Em primeiro lugar, a nível de dados, será pertinente integrar novas variáveis de natureza qualitativa e contextual, tais como fatores socioeconómicos, níveis de motivação, envolvimento em atividades extracurriculares e indicadores de bem-estar estudantil. A incorporação destas dimensões poderá enriquecer substancialmente as análises, permitindo compreender melhor os fatores não administrativos que influenciam o sucesso académico.

Em segundo lugar, importa melhorar a qualidade e a consistência dos dados recolhidos pelos sistemas institucionais. A introdução de mecanismos de validação automática, como menus suspensos, listas pré-definidas de cursos e distritos ou restrições de entrada em campos textuais, pode reduzir significativamente o ruído e garantir maior uniformidade semântica. Este aperfeiçoamento não só simplificaria a fase de limpeza e preparação, como também aumentaria a fiabilidade das análises subsequentes.

Em terceiro lugar, sugere-se a integração de metodologias preditivas mais sofisticadas, incluindo algoritmos de *ML* supervisionados e não supervisionados, capazes de identificar padrões complexos e de antecipar, em tempo real, estudantes em risco de insucesso ou abandono. A combinação entre análises exploratórias e preditivas ampliaria o valor estratégico do *dashboard*, transformando-o numa ferramenta não apenas descritiva, mas também preventiva.

Outra linha de evolução passa pela expansão do âmbito empírico. A aplicação do mesmo modelo de análise a outros cursos ou instituições permitiria realizar comparações interdisciplinares e interinstitucionais, aumentando a robustez dos resultados e permitindo identificar tendências gerais do ensino superior português. Adicionalmente, estudos longitudinais que acompanhem os estudantes para além do ciclo de estudos inicial, incluindo percursos de diplomados no mercado de trabalho, enriqueceriam a compreensão do impacto a longo prazo das variáveis de ingresso.

Finalmente, será fundamental proceder a uma validação prática e participada do *dashboard*. Testes de usabilidade junto de docentes, coordenadores e responsáveis institucionais permitiriam aferir a eficácia da ferramenta no apoio à tomada de decisão, como também recolher sugestões para melhorias futuras. A implementação piloto em contextos reais de gestão académica poderá transformar o protótipo desenvolvido numa solução efetivamente integrada no ecossistema institucional.

Em síntese, os trabalhos futuros deverão articular-se em três eixos centrais: enriquecimento das variáveis analisadas, sofisticação das metodologias aplicadas e validação

prática do *dashboard*. O avanço nestas direções poderá consolidar o papel da análise de dados e da visualização interativa como instrumentos estratégicos para a promoção do sucesso académico e para uma gestão mais informada e equitativa do ensino superior.

[Esta página foi deixada intencionalmente em branco]

Bibliografia

[1] A. I. Al-Alawi, M. A. A. Alfateh e A. M. Alrayes, «Educational Data Mining Utilization to Support the Admission Process in Higher Education Institutions: A Systematic Literature Review,» em *2023 International Conference On Cyber Management And Engineering (CyMaEn)*, 2023, pp. 332–339. DOI: 10.1109/CyMaEn57228.2023.10051077.

[2] K. Thongkaew, S. Vasupongayya, J. Phongsooknirun e T. Loywirat, «Data Analytics for Admission Process: Bachelor of Engineering Program,» em *Advances in Transdisciplinary Engineering*. IOS Press BV, nov. de 2023, vol. 41, pp. 671–680, ISBN: 9781643684406. DOI: 10.3233/ATDE230663.

[3] C. Hourigan, «From Data Entry to Evidence-Based Decision-Making: How Data and Analysis Can Drive Improvements in Admissions Policies.,» 2011. URL: https://api.semanticscholar.org/CorpusID:156237221.

[4] H. A. Mengash, «Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems,» *IEEE Access*, vol. 8, pp. 55 462–55 470, 2020. DOI: 10.1109/ACCESS.2020.2981905.

[5] J. A. Solano, D. J. L. Cuesta, S. F. U. Ibáñez e J. R. Coronado-Hernández, «Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test,» em *Procedia Computer Science*, vol. 198, Elsevier B.V., 2021, pp. 512–517. DOI: 10.1016/j.procs.2021.12.278.

[6] A. Zagorecki, J. Ristvej, L. K. Comfort e T. Lovecek, «Executive Dashboard Systems for Emergency Management,» *Communications - Scientific letters of the University of Zilina*, vol. 14, pp. 82–89, 2 jun. de 2012, ISSN: 13354205. DOI: 10.26552/com.C.2012.2.82-89. URL: http://komunikacie.uniza.sk/doi/10.26552/com.C.2012.2.82-89.html.

[7] H. Deng, X. Wang, Z. Guo, A. Decker, X. Duan, C. Wang, G. A. Ambrose e K. Abbott, «PerformanceVis: Visual analytics of student performance data from an introductory chemistry course,» *Visual Informatics*, vol. 3, pp. 166–176, 4 2019. DOI: 10.1016/j.visinf.2019.10.004.

[8] P. Vuttipittayamongkol, «Predicting factors of academic performance,» em *2016 2nd Asian Conference on Defence Technology, ACDT 2016*, 2016, pp. 161–166, ISBN: 9781509022571. DOI: 10.1109/ACDT.2016.7437662.

[9] J. Pecuchova e M. Drlik, «Enhancing the Early Student Dropout Prediction Model Through Clustering Analysis of Students' Digital Traces,» *IEEE Access*, vol. 12, pp. 159 336–159 367, 2024. DOI: 10.1109/ACCESS.2024.3486762.

[10] K. Ahmad, W. Iqbal, A. El-Hassan, J. Qadir, D. Benhaddou, M. Ayyash e A. Al-Fuqaha, «Data-Driven Artificial Intelligence in Education: A Comprehensive Review,» *IEEE Transactions on Learning Technologies*, vol. 17, pp. 12–31, 2024. DOI: 10.1109/TLT.2023.3314610.

[11] S. Alwarthan, N. Aslam e I. Khan, «Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review,» *Applied Computational Intelligence and Soft Computing*, vol. 2022, 2022. DOI: 10.1155/2022/8924028.

[12] J. Kanchana, G. Amarasinghe, V. Nanayakkara e A. S. Perera, «A Data Mining Approach for Early Prediction Of Academic Performance of Students,» em *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, 2021, pp. 01–08. DOI: 10.1109/TALE52509.2021.9678558.

[13] S. Ray e M. Saeed, *Applications of educational data mining and learning analytics tools in handling big data in higher education*. 2018, pp. 135–160, ISBN: 9783319764726. DOI: 10.1007/978-3-319-76472-6_7.

[14] F. Qiu, G. Zhang, X. Sheng, L. Jiang, L. Zhu, Q. Xiang, B. Jiang e P.-K. Chen, «Predicting students' performance in e-learning using learning process and behaviour data,» *Scientific Reports*, vol. 12, 1 2022. DOI: 10.1038/s41598-021-03867-8.

[15] M. Ursan e M. Bucos, «Predictive Analytics Models for Student Admission to Master Programs in Romania,» em *2022 15th International Symposium on Electronics and Telecommunications, ISETC 2022 - Conference Proceedings*, 2022, ISBN: 9781665451505. DOI: 10.1109/ISETC56213.2022.10010056.

[16] A. Rabelo, M. Rodrigues, C. Nobre, S. Isotani e L. Zárate, «Educational data mining and learning analytics: a review of educational management in e-learning,» *Information Discovery and Delivery*, vol. 52, pp. 149–163, 2 2024. DOI: 10.1108/IDD-10-2022-0099.

[17] I. Kolodii, I. Piatnytska, A. Shkodyn, K. Herasymiuk e N. Ponomarenko, «Implementation of Quality Control of the Educational Process in Higher Education Institutions Based on Models,» *Journal of Curriculum and Teaching*, vol. 11, pp. 40–48, 9 2022. DOI: 10.5430/JCT.V11N9P40.

[18] C. Kontzinos, O. Markaki, P. Kokkinakos, V. Karakolis, S. Skalidakis e J. Psarras, «University process optimisation through smart curriculum design and blockchain-based student accreditation,» em *18th International Conference on WWW/Internet 2019*, 2019, pp. 93–100, ISBN: 9789898533944. DOI: 10.33965/icwi2019_2019131012.

[19] J. Ballaho, «Predicting Student's Success in Programming Courses: A Decision Support System for Admission in Computer Science and Information Technology Programs,» em *6th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2024*, 2024, pp. 60–64, ISBN: 9798350389692. DOI: 10.1109/IICAIET62352.2024.10729909.

[20] C. F. de Oliveira, S. R. Sobral, M. J. Ferreira e F. Moreira, «How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education: A Systematic Literature Review,» *Big Data and Cognitive Computing*, vol. 5, n.º 4, 2021, ISSN: 2504-2289. DOI: 10.3390/bdcc5040064. URL: https://www.mdpi.com/2504-2289/5/4/64.

[21] R. Kavitha, C. R. Krupa e J. Menezes, «Determining the Factors Influencing the Academic Accomplishment of Students and Predicting their Success Using Machine Learning Techniques,» em *2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2023*, 2023, ISBN: 9798350306811. DOI: 10.1109/ICAECA56562.2023.10200741.

[22] N. Tuah, A. Yoag, D. Nizam e C. Chin, «A Dashboard-based System to Manage and Monitor the Progression of Undergraduate IT Degree Final Year Projects,» *Pertanika Journal of Science and Technology*, vol. 30, pp. 235–256, 1 2022. DOI: 10.47836/pjst.30.1.13.

[23] S. Schiller, M. Goul, L. Iyer, R. Sharda, D. Schrader e D. Asamoah, «Build your dream (Not just big) analytics program,» *Communications of the Association for Information Systems*, vol. 37, pp. 811–826, 1 2015. DOI: 10.17705/1cais.03740.

[Esta página foi deixada intencionalmente em branco]

APÊNDICE A

Dicionário de Dados

Group/field	**Description**	**Data Type**
1ª Vez?	Indica se a inscrição realizada foi a primeira do estudante	Boolean
Acordo	Acordo de ingresso do estudante no curso	String
Algum Estado de Matrícula Inativo?	Indica se a matrícula já teve estado inativo (ex.: Interrupção de Estudos)	String
Ano Curricular (Ano Anterior)	Ano curricular do ano letivo anterior	Integer
Ano Curricular (Início do Ano Lectivo)	Ano curricular no início do ano letivo	Integer
Ano de Início	Ano em que o curso começou a ser frequentado pelo estudante	String
Ano e Matrícula	Ano letivo referente aos dados	String
Ano Lectivo: Créditos Aprovados Obrigatórios	Nº de créditos aprovados em UCs obrigatórias do ano curricular	Integer
Ano Lectivo: Créditos Inscritos Obrigatórios	Nº de créditos inscritos em UCs obrigatórias do ano curricular	Integer
Ano Lectivo: Data de Conclusão	Data do último momento de avaliação das UCs inscritas	Date
Ano Lectivo: Inscrito em UCs Obrigatórias Atrasadas	Indica se o estudante está inscrito em UCs obrigatórias em atraso	Boolean

Campo	Descrição	Data Type
Ano Lectivo: Inscrito em UCs Obrigatórias em Avanço	Indica se o estudante está inscrito em UCs obrigatórias em avanço	Boolean
Ano Lectivo: Média Ponderada	Média ponderada do ano letivo (por créditos)	Decimal
Ano Lectivo: Média Simples	Média simples do ano letivo	Decimal
Anos de Conclusão de Outras Matrículas?	Anos letivos em que o estudante concluiu outras matrículas	String
Apuramento Final: Ano de Conclusão	Ano letivo em que o estudante concluiu o curso	String
Apuramento Final: Concluído?	Indica se o estudante tem condições para apuramento final	Boolean
Apuramento Final: Data de Conclusão	Data da última avaliação no curso	Datetime
Apuramento Final: Nota Descritiva	Nota qualitativa (ex.: doutoramento)	String
Apuramento Final: Nota Final	Nota final atribuída após conclusão (arredondada)	Integer
Apuramento Final: Nota Original	Nota final calculada	Decimal
Apuramento Final: Processado?	Indica se os dados finais foram processados	Boolean
Apuramento Parcial: Ano de Conclusão	Ano letivo da última avaliação parcial	String

Campo	Descrição	Data Type
Apuramento Parcial: Concluído?	Indica se o estudante tem condições para apuramento parcial	Boolean
Apuramento Parcial: Créditos	Total de ECTS contabilizados para o apuramento parcial	Integer
Apuramento Parcial: Data de Conclusão	Data da última avaliação parcial	Datetime
Apuramento Parcial: Nota Descritiva	Nota qualitativa	String
Apuramento Parcial: Nota Final	Média final para o apuramento parcial	Integer
Apuramento Parcial: Nota Original	Média calculada para o apuramento parcial	Decimal
Apuramento Parcial: Processado?	Indica se o apuramento parcial foi efetuado	Boolean
Área de Investigação	Área de investigação do curso	String
Área do Código Postal	Área correspondente ao código postal	String
Código do Curso (Subsequente)	Código do curso de ciclos subsequentes. Curso de mestrado do qual o estudante finalista de Licenciatura está a frequentar UC.	String
Código Oficial do Curso	Código oficial do curso nas entidades reguladoras	Integer
Código Postal	Código postal	String
Concelho	Concelho de residência	String
Concelho de Nascimento	Concelho onde nasceu	String

Campo	**Descrição**	**Data Type**
Créditos (Início do Ano Lectivo)	Créditos (ECTS) que o estudante tem no início do ano letivo	Integer
Créditos Inscritos	Nº total de créditos inscritos	Integer
Créditos Inscritos (Extra-Curricular)	Nº de créditos inscritos em UCs extra-curriculares	Integer
Créditos Inscritos (Isoladas)	Nº de créditos inscritos em UCs isoladas	Integer
Créditos Inscritos (Subsequentes)	Nº de créditos inscritos em UCs subsequentes	Integer
Data de Início	Data em que o curso começou a ser frequentado pelo estudante	Date
Data de Inscrição	Data da última inscrição em que o estudante se inscreveu no curso	Datetime
Data de Nascimento	Data de nascimento	Datetime
Data Última Atualização de Inscrição	Data da última modificação nos dados da inscrição	Datetime
Data Último Estado	Data em que o último estado iniciou	Datetime
Distrito	Distrito de residência	String
Distrito de Nascimento	Distrito onde nasceu	String
Estatutos	Lista de estatutos do estudante (ex.: trabalhador estudante)	String
Fase de Entrada	Para estudantes de Licenciatura ou Mestrado integrado que ingressaram via CNA. Indica a fase de ingresso do estudante no curso	Integer
Freguesia	Freguesia de residência	String
Freguesia de Nascimento	Freguesia onde nasceu	String

Campo	**Descrição**	**Data Type**
Género	Género do estudante	String
Habilitação Anterior - Curso	Curso na instituição anterior	String
Habilitação Anterior - Instituição	Instituição anterior do estudante	String
Habilitação Anterior - Nível	Nível da graduação na instituição anterior	String
Habilitação Anterior - País	País da instituição anterior	String
Identificação do Curso	Código interno do curso na instituição	String
Ingresso	Forma de ingresso do estudante no curso	String
Inscrições Anuladas?	Indica se o estudante tem inscrições anuladas	Boolean
Inscrições em Melhoria?	Indica se o estudante tem inscrições em melhorias de nota	Boolean
Localidade	Localidade onde reside	String
Média (Início do Ano Lectivo)	Média no início do ano letivo	Decimal
Média Atual	Média ao momento da recolha dos dados	Decimal
Nacionalidade	Nacionalidade do estudante	String
Nº de Anos de Inscrição (inclui atual)	Nº de inscrições em anos letivos diferentes, incluindo o atual	Integer
Nº de Anos de Inscrição c/ precedentes (inclui atual)	Nº de inscrições em anos letivos diferentes, incluindo cursos precedentes	Integer
Nº de Anos de Inscrição p/ Prescrição (Bonificado)	Nº de anos de inscrição considerados para efeitos de prescrição, incluindo bonificações	Integer
Nº de Inscrições	Nº total de inscrições em UCs no ano letivo	Integer

Campo	**Descrição**	**Data Type**
Nº Inscrições (Extra-Curricular)	Nº de inscrições em UCs extra-curriculares	Integer
Nº Inscrições (Isoladas)	Nº de inscrições em UCs isoladas	Integer
Nº Inscrições (Subsequentes)	Nº de inscrições em UCs subsequentes	Integer
Nº Registos de Mobilidade: Incoming	Nº de registos em que o estudante ingressou em regime de mobilidade em Portugal	Integer
Nº Registos de Mobilidade: Outgoing	Nº de registos em que o estudante ingressou em regime de mobilidade no estrangeiro	Integer
Nº Total de Matrículas	Nº total de matrículas do estudante	Integer
Nome Completo do Curso	Nome completo do curso inscrito	String
Nome de Graduação do Curso	Nome do curso atribuído ao grau académico	String
Nota de Ingresso	Nota de ingresso do estudante no curso	Integer
Opção de Colocação	Opção escolhida pelo estudante no momento da candidatura ao curso	Integer
País de Nascimento	País onde nasceu	String
País de Residência	País onde reside	String
Plano Curricular	Plano de estudos atual que o estudante está a frequentar	String
Plano Curricular (Subsequente)	Plano de estudos do curso de ciclos subsequentes. Curso de mestrado do qual o estudante finalista de Licenciatura está a frequentar UC.	String
Planos Curriculares	Número de planos de estudos que o estudante teve associado no curso	Integer

Campo	**Descrição**	**Data Type**
Pode Prescrever?	Indica se o estudante pode ser excluído do curso por ultrapassar tempo máximo	Boolean
Previsão de Ano Curricular (Seguinte)	Previsão do ano curricular no ano letivo seguinte, considerando aprovações/creditações	Integer
Proveniência - Curso	Curso frequentado pelo estudante antes da entrada	String
Proveniência - Instituição	Instituição de ensino de onde o estudante veio	String
Proveniência - Nível	Nível de ensino anteriormente frequentado	String
Proveniência - País	País da instituição de ensino anterior	String
Ramo Principal	Se o curso prever ramos, indica o ramo escolhido	String
Ramo Secundário	Não aplicável no ISCTE	String
Regime	Regime inscrito (ex.: Tempo Integral/Parcial)	String
Reingressou (inclui precedentes)?	Indica se o estudante já reingressou anteriormente, considerando cursos precedentes	Boolean
Reingressou em Anos Anteriores?	Indica se o estudante reingressou no curso em anos letivos anteriores	Boolean
Reingressou no Ano Lectivo?	Indica se o estudante reingressou no curso no ano letivo atual	Boolean
Segunda Nacionalidade	Segunda nacionalidade (caso exista)	String
Tem Creditações?	Indica se o estudante tem creditações para o curso	Boolean
Tipo de Curso	Tipologia do Curso (Licenciatura, pós-graduação, mestrado, doutoramento)	String
Última Inscrição	Ano letivo da última inscrição	String

Campo	**Descrição**	**Data Type**
Último Estado	Último estado da matrícula	String
Unidade do Curso	Unidade onde o curso é lecionado dentro da instituição (ex.: Escola)	String

Este apêndice apresenta o dicionário de dados utilizado no presente trabalho, descrevendo detalhadamente cada uma das variáveis incluídas na base de dados institucional.

APÊNDICE B

Dashboard Power BI - LIGE

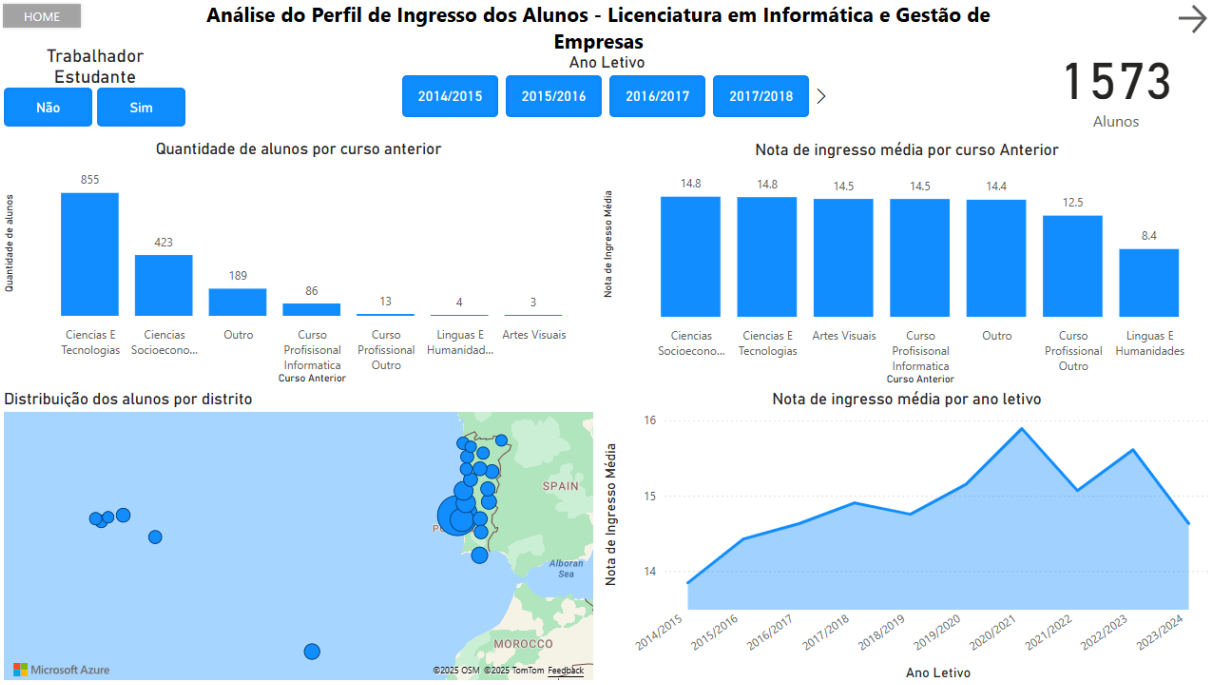


FIGURA B.1. *Dashboard Power BI* - LIGE - Análise de Ingresso

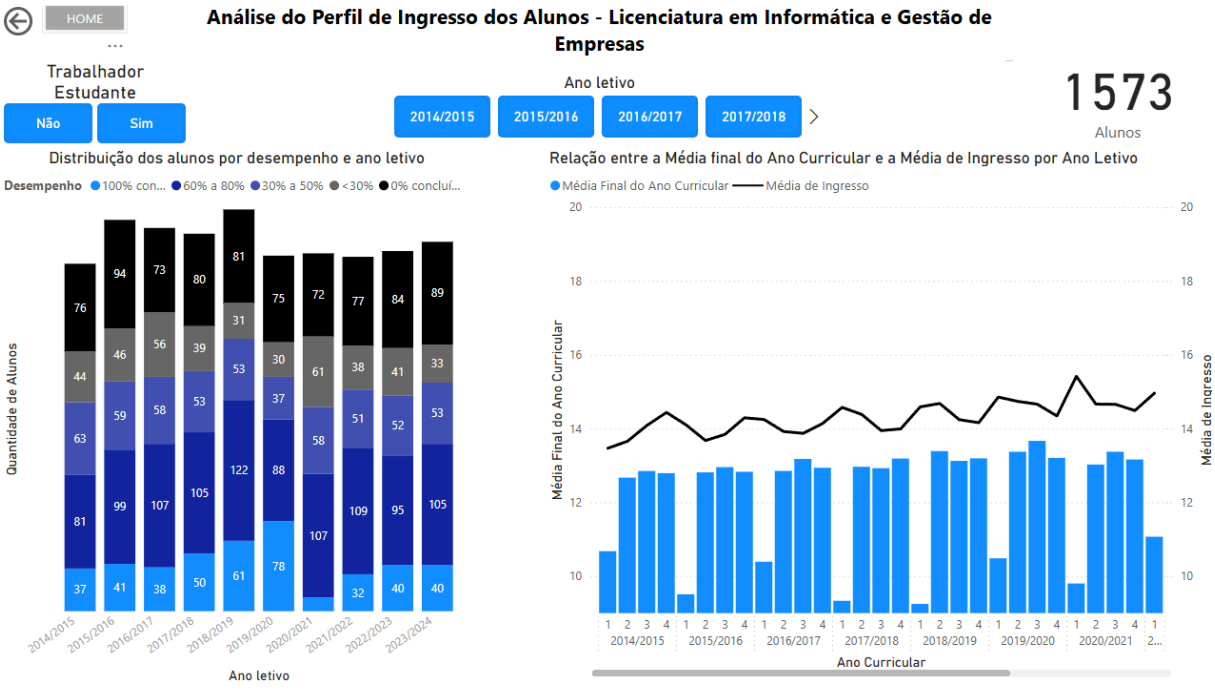


FIGURA B.2. *Dashboard Power BI* - LIGE - Análise de Progresso

Este apêndice inclui as imagens dos *dashboards* interativos, da Licenciatura de Informática e Gestão de Empresas, desenvolvidos em *Power BI*. Cada *dashboard* foi criado para responder a questões analíticas, permitindo explorar indicadores académicos de uma forma segmentada, comparativa e visualmente intuitiva.

APÊNDICE C

Dashboard Power BI - LCD

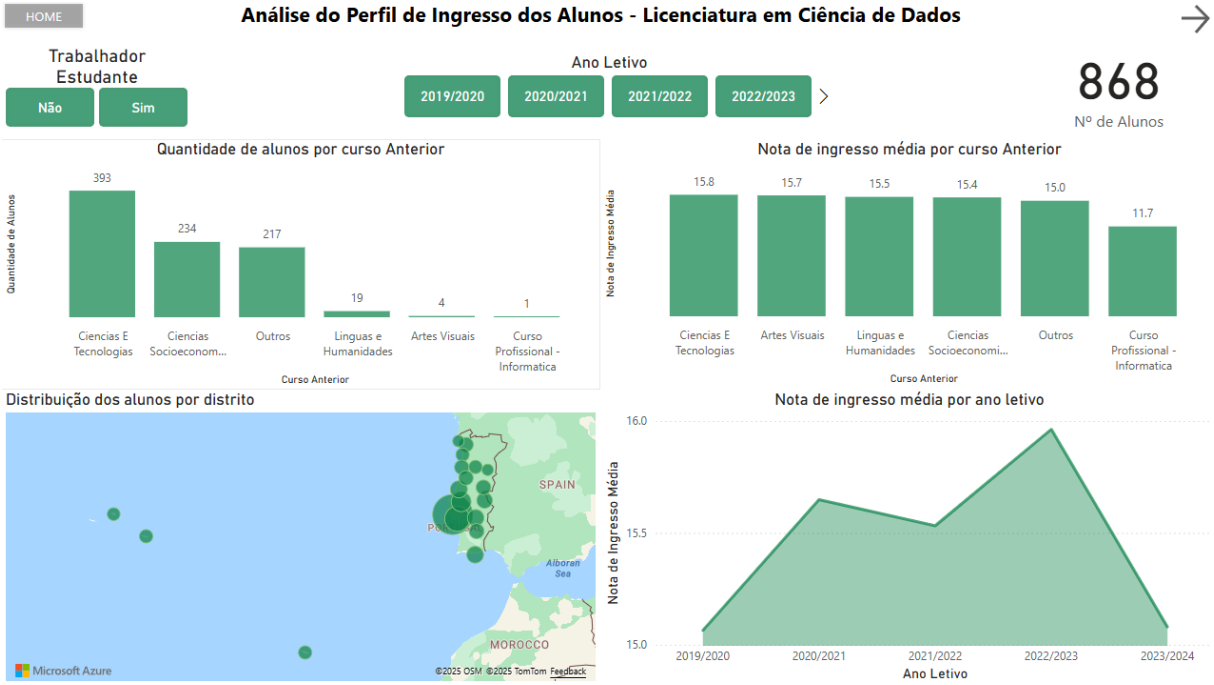


FIGURA C.1. *Dashboard Power BI* - LCD - Análise de Ingresso

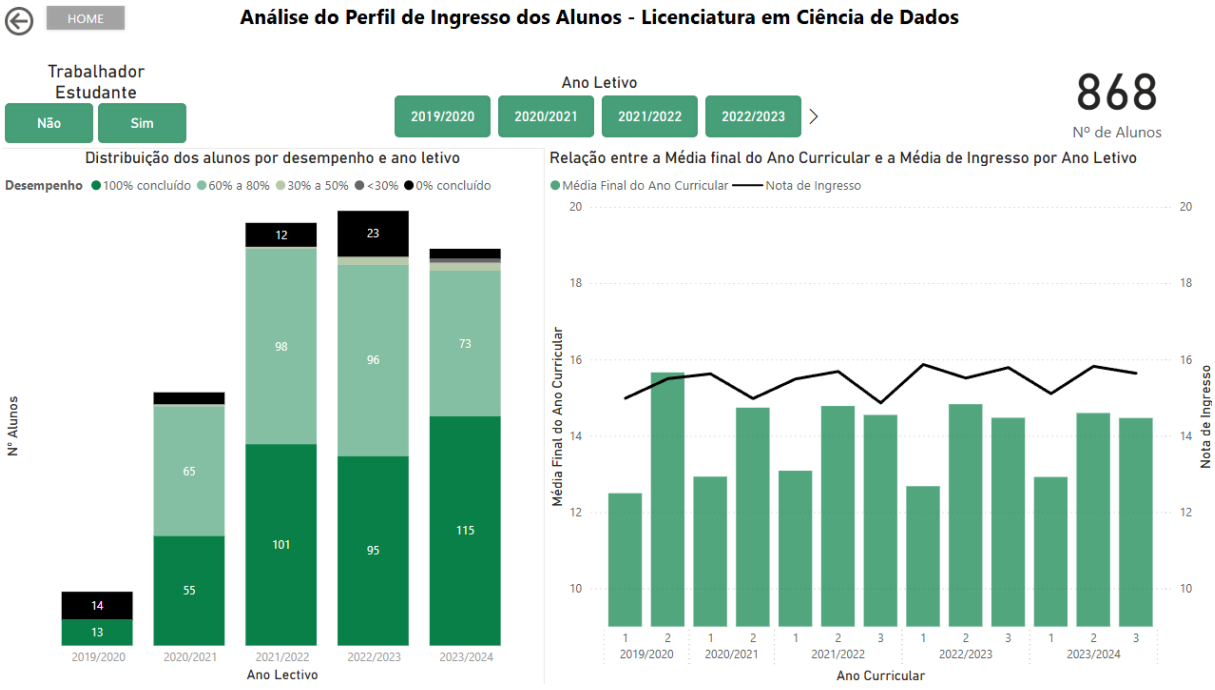


FIGURA C.2. *Dashboard Power BI* - LCD - Análise de Progresso

Este apêndice inclui as imagens dos *dashboards* interativos, da Licenciatura de Ciência de Dados, desenvolvidos em *Power BI*. Cada *dashboard* foi criado para responder a questões analíticas, permitindo explorar indicadores académicos de uma forma segmentada, comparativa e visualmente intuitiva.