

José Manuel Guimarães Pereira

Análise do desempenho de estudantes numa Instituição de Ensino Superior - fatores condicionantes e estratégias: uma aproximação baseada em Business Intelligence

Relatório de Projeto de Mestrado apresentado à Universidade Portucalense Infante D. Henrique para obtenção do grau de Mestre em Informática, Especialização em Engenharia de Software, realizado sob a orientação da Doutora Filomena Castro Lopes e Doutora Isabel Seruca, Professoras Associadas do Departamento de Inovação Ciência e Tecnologia da Universidade Portucalense.

Departamento de Ciência e Tecnologia, Universidade Portucalense
Fevereiro de 2022

AGRADECIMENTOS

Às minhas orientadoras, Professora Doutora Filomena Lopes e Professora Doutora Isabel Seruca, pela sua orientação, conselhos prestados e acima de tudo pela sua dedicação e apoio, fundamental para a elaboração deste projeto.

À minha família, pais, irmã, avós e à minha prima Ana Bela Guimarães.

A todos, que participaram direta ou indiretamente na realização deste projeto e que contribuíram para o sucesso do trabalho final, o meu sincero obrigado.

RESUMO

Análise do desempenho de estudantes numa Instituição de Ensino Superior - fatores condicionantes e estratégias: uma aproximação baseada em Business Intelligence

Um sistema de Business Intelligence (BI) inclui ferramentas e técnicas que, , permitem analisar os dados gerados por uma organização de forma a extrair informação relevante que suporte a tomada de decisão. As Instituições de Ensino Superior (IES) poderão beneficiar da utilização de sistemas de BI em várias vertentes, nomeadamente na análise do desempenho académico dos alunos.

Este trabalho tem como objetivo endereçar a utilização de sistemas de BI no Ensino Superior (ES) que atuem no âmbito da monitorização e análise do desempenho académico e demonstrar as várias etapas da respetiva implementação e desafios associados, bem como as mais valias que estes aportam. Apresenta-se uma prova de conceito da aplicação do BI numa IES Portuguesa privada, que analisa as características dos alunos matriculados em vários cursos e determina a influência de diversos fatores no seu sucesso académico. Esses fatores incluem dados relacionados com a formação académica antes de ingressar no ES e dados sociodemográficos que serão cruzados com informações sobre o desempenho dos alunos ao longo e na conclusão do curso na IES.

A ideia subjacente será suportar o processo de garantia da qualidade do ensino, sugerindo melhorias tanto para o processo de ensino-aprendizagem como para a comunicação da oferta. Com o conhecimento obtido, a IES poderá desenhar com mais facilidade uma estratégia diferenciadora para os alunos inscritos, tendo em conta os respetivos perfis e necessidades evidenciadas, melhorando a comunicação da oferta formativa e criando oportunidades a nível curricular e extracurricular, bem como iniciativas de projetos adaptados aos vários segmentos de alunos.

Palavras-chave: Ensino Superior; Análise desempenho estudantes; Business Intelligence

ABSTRACT

Analysis of student's performance in a Higher Education Institution - conditioning factors and strategies: an approach based on Business Intelligence

A Business Intelligence (BI) system combines tools and techniques that, supported by technology, allow to use the data generated by an organization in order to extract relevant information that supports decision making. Higher Education Institutions (HEI) may benefit from the use of BI systems in several aspects, in addition to administrative management, namely in the analysis of students' academic performance.

This work aims to address the use of BI systems in Higher Education (HE) that are focused on the monitoring and analysis of academic performance and demonstrate the various stages of their implementation and associated challenges, as well as their added value. A proof of concept is presented for the application of BI in a private Portuguese HEI, which analyzes the characteristics of students enrolled in various study programmes and determines the influence of different factors on the academic success of students. These factors include data related to the academic training of students before joining the HEI and sociodemographic data that will be cross-referenced with information on the performance of students throughout and at the conclusion of the study programme at HEI.

The underlying idea will be to support the education quality assurance process, suggesting improvements for both the training process and offer. With the knowledge obtained, the HEI will be able to more easily design a differentiating strategy for the enrolled students, taking into account their respective profiles and evidenced needs, improving the communication of the training offer and creating opportunities at curricular and extracurricular level, as well as providing project initiatives adapted to the various segments of students.

Keywords: Higher Education; Student's performance analysis; Business Intelligence.

ÍNDICE

1	Introdução	1
1.1	Objetivos e contribuição do trabalho	2
1.2	Metodologia adotada	3
1.3	Estrutura do trabalho	3
2	Revisão da Literatura.....	5
2.1	Conceito de Business Intelligence	5
2.2	Arquitetura e componentes de um sistema de Business Intelligence	6
2.2.1	Processo ETL.....	7
2.2.2	Data Warehouse.....	8
2.2.3	Análise OLAP	14
2.2.4	Data Mining	15
2.2.5	Monitorização e visualização dos KPIs	18
2.3	Ferramentas de Software de Business Intelligence	18
2.4	Business Intelligence no Ensino Superior	21
3	Caso de estudo	24
3.1	Descrição.....	24
3.2	Dados a analisar.....	25
3.3	Correlações e KPIs.....	26
3.4	Desenho do DW	29
3.4.1	Dimensões.....	30

3.4.2	Factos.....	34
3.5	Processo ETL.....	34
3.6	Configuração da plataforma Qlik Sense.....	40
3.7	Criação dos gráficos na plataforma Qlik Sense.....	44
3.8	Análise OLAP	48
3.8.1	Correlação entre a nota final do curso e a escola frequentada no ensino secundário.....	49
3.8.2	Segmentos de alunos por regime de ingresso e nota final de curso	50
3.8.3	Segmentos de alunos por situação profissional dos pais e correlação com a nota final de curso.....	51
3.8.4	Segmentos de alunos por nível de escolaridade dos pais e correlação com a nota final de curso.....	53
3.8.5	Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso.....	55
3.8.6	Segmentos de alunos por origem geográfica e nota de ingresso ...	56
3.8.7	Segmentos de alunos caracterizados pela sua performance académica	57
3.8.8	Correlação entre a nota final de curso e a nota de ingresso	58
3.8.9	Segmentos de alunos por género e nota de ingresso, relacionados com a nota final de curso.....	59
3.8.10	Eficiência no ensino - percentagem de graduados (por ano letivo)	60
3.9	Análise dos KPI's.....	61
3.9.1	Número de alunos inscritos por género (por ano letivo)	61

3.9.2	Número de alunos por idade.....	62
3.9.3	Número de alunos por origem geográfica em Portugal	63
3.9.4	Número de alunos por grau de escolaridade da mãe.....	64
3.9.5	Número de alunos por grau de escolaridade do pai.....	65
3.9.6	Número de alunos por situação profissional da mãe.....	66
3.9.7	Número de alunos por situação profissional do pai.....	66
3.9.8	Número de alunos deslocados da sua residência permanente	67
3.9.9	Número de alunos inscritos por ano curricular (por ano letivo).....	67
3.9.10	Nota média por UC por ano letivo	68
3.9.11	Percentagem de alunos por intervalo de nota, por UC e por ano letivo	69
3.9.12	Número médio de inscrições até ao final do curso por aluno	70
3.9.13	Percentagem de aprovações por UC e por ano letivo	71
3.9.14	Percentagem de graduados em N anos (por ano letivo)	72
3.9.15	Percentagem de graduados em N+1 anos (por ano letivo)	72
3.9.16	Percentagem de graduados em N+2 anos (por ano letivo)	73
3.9.17	Percentagem de graduados em mais de N+2 anos (por ano letivo)	73
3.9.18	Nota média de conclusão por curso e por ano letivo.....	74
3.9.19	Dashboards	75
4	Conclusão	76
4.1	Reflexão sobre o trabalho realizado.....	76

4.2	Obstáculos enfrentados	77
4.3	Constrangimentos e limitações ao trabalho realizado	78
4.4	Sugestões para Futuras Intervenções	79
5	Referências bibliográficas.....	81
6	Anexos	83
6.1	Correlação entre a nota final do curso e a escola frequentada no ensino secundário (cf. Figura 35)	83
6.2	Segmentos de alunos por regime de ingresso e nota final de curso (cf. Figura 36)	84
6.3	Segmentos de alunos por situação profissional dos pais e correlação com a nota final de curso (cf. Figura 37, Figura 38).....	85
6.4	Segmentos de alunos por nível de escolaridade dos pais e correlação com a nota final de curso (cf. Figura 39, Figura 40).....	87
6.5	Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso (cf. Figura 41).....	89
6.6	Segmento de alunos por género e nota de ingresso, relacionados com a nota final de curso (cf. Figura 45).....	90
6.7	Número de alunos por idade (cf. Figura 48).....	91
6.8	Número médio de inscrições até ao final do curso por aluno (cf. Figura 58)	92
6.9	Dashboards (cf. Figura 65)	93

LISTA DE ACRÓNIMOS

Acrónimo	Significado
BI	Business Intelligence
CE	Chave Estrangeira
CP	Chave Primária
DCBD	Descoberta de conhecimento em bases de dados
DM	Data Mining
DW	Data Warehouse
ES	Ensino Superior
ETL	Extract, Transform, Load
HEI	Higher Education Institution
IES	Instituição de Ensino Superior
KPI	Key Performance Indicator
OLAP	Online Analytical Processing
RGPD	Regulamento Geral sobre a Proteção de Dados
SIUPT	Sistema de Informação da Universidade Portucalense
UC	Unidade Curricular
UPT	Universidade Portucalense

ÍNDICE DE FIGURAS

Figura 1 - Arquitetura de um sistema de BI (Ahmad et al., 2020)	6
Figura 2 - O processo ETL (Fatima, 2020)	7
Figura 3 - Esquema em estrela adaptado (Myers, 2019).....	10
Figura 4 - Esquema em floco de neve (Snowflake)	12
Figura 5 - Esquema em constelação (Last night Study, n.d.).....	13
Figura 6 - Análise OLAP (Myers, 2019).....	14
Figura 7 - Fases do processo de DCBD (Vaz, 2016).....	16
Figura 8 - Magic Quadrant for Analytics and Business Intelligence Platforms (Analytics and Business Intelligence Platforms, 2020).....	19
Figura 9 - Esquema em estrela do DW do sistema	29
Figura 10 - Limpeza/Uniformização de dados	35
Figura 11 - Carregamento de dados para a tabela Dim_Country	35
Figura 12 - Carregamento de dados para a tabela Dim_AdmissionType.....	35
Figura 13 - Carregamento de dados para a tabela Dim_Course	36
Figura 14 - Carregamento de dados para a tabela Dim_School	36
Figura 15 - Carregamento de dados para a tabela Dim_Study	36
Figura 16 - Carregamento de dados para a tabela Fact_Grades.....	37
Figura 17 - Carregamento de dados para a tabela Dim_Student.....	38
Figura 18 - Carregamento de dados para a tabela Dim_SchoolYear.....	39
Figura 19 - Datawarehouse na implementação do projeto em Qlik Sense.....	39
Figura 20 - Modelo dados na implementação do projeto em Qlik Sense	40

Figura 21 - Criação de nova dimensão.....	40
Figura 22 - Dimensão para traduzir a flag Type da Dimensão School	41
Figura 23 - Dimensão para criar intervalos para a nota ingresso.....	41
Figura 24 - Dimensão para criar intervalos para a nota final de curso	42
Figura 25 - Dimensão para criar intervalos para nota de UC	42
Figura 26 - Dimensão para criar categorias de desempenho académico baseadas no total de anos e nota final	43
Figura 27 - Dimensão para criar categorias de desempenho académico baseadas no total de anos.....	43
Figura 28 - Dimensão para criar intervalos de idades.....	44
Figura 29 - Primeira etapa na criação de um gráfico	44
Figura 30 - Escolher dimensão previamente criada (EntryMark Interval).....	45
Figura 31 - Inserir expressões relativas às medidas.....	46
Figura 32 - Gráfico defeito.....	46
Figura 33 - Alterar orientação do gráfico	47
Figura 34 - Alterar localização da legenda do gráfico	47
Figura 35 - Correlação entre a nota final do curso e a escola frequentada no ensino secundário	49
Figura 36 - Segmentos de alunos por regime de ingresso e nota final de curso.	50
Figura 37 - Segmentos de alunos por situação profissional do pai e correlação com a nota final de curso	51
Figura 38 - Segmentos de alunos por situação profissional da mãe e correlação com a nota final de curso	52

Figura 39 - Segmentos de alunos por nível de escolaridade da mãe e correlação com a nota final de curso	53
Figura 40 - Segmentos de alunos por nível de escolaridade do pai e correlação com a nota final de curso	54
Figura 41 - Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso	55
Figura 42 - Segmentos de alunos por origem geográfica e nota de ingresso	56
Figura 43 - Segmentos alunos caracterizados pela sua performance académica	57
Figura 44 - Correlação entre a nota final de curso e a nota de ingresso.....	58
Figura 45 - Segmentos de alunos por género e nota de ingresso, relacionados com a nota final de curso	59
Figura 46 - Eficiência no ensino - percentagem de graduados (ano letivo 2016/2017).....	60
Figura 47 - Número de alunos inscritos por género (ano letivo 2017/2018).....	61
Figura 48 - Número de alunos por idade	62
Figura 49 - Número de alunos por origem geográfica em Portugal.....	63
Figura 50 - Número de alunos por grau de escolaridade da mãe	64
Figura 51 - Número de alunos por grau de escolaridade do pai	65
Figura 52 - Número de alunos por situação profissional da mãe	66
Figura 53 - Número de alunos por situação profissional do pai	66
Figura 54 - Número de alunos deslocados da sua residência permanente.....	67
Figura 55 - Número de alunos inscritos por ano curricular (por ano letivo)	67
Figura 56 - Nota média por UC por ano letivo	68

Figura 57 - Percentagem de alunos por intervalo de nota, por UC e por ano letivo (Ciência Política – 2016/2017)	69
Figura 58 - Número médio de inscrições até ao final do curso por aluno.....	70
Figura 59 - Percentagem de aprovações por UC e por ano letivo (Estatística – 2013/2014).....	71
Figura 60 - Percentagem de graduados em N anos (2017/2018)	72
Figura 61 - Percentagem de graduados em N+1 anos (2017/2018)	72
Figura 62 - Percentagem de graduados em N+2 anos (ano letivo 2017/2018) ...	73
Figura 63 - Percentagem de graduados em mais de N+2 anos (ano letivo 2017/2018).....	73
Figura 64 - Nota média de conclusão por curso e ano letivo	74
Figura 65 - Dashboard de análise sociodemográfica e de desempenho académico	75

ÍNDICE DE TABELAS

Tabela 1 - Diferenças entre base de dados operacionais e DW	9
Tabela 2 - Diferenças entre tabela de factos e tabela de dimensão	11
Tabela 3 - Descrição das operações possíveis num cubo OLAP	15
Tabela 4 - Tarefas e técnicas de DM.....	17
Tabela 5 - Comparação de ferramentas de BI (Sukumar, 2021)	21
Tabela 6 - Lista dos KPIs identificados.....	27
Tabela 7 - Lista dos KPIs identificados (continuação)	28
Tabela 8 - Tabela Dim_AdmissionType do DW	30
Tabela 9 - Tabela Dim_SchoolYear do DW	30
Tabela 10 - Tabela Dim_Course do DW.....	31
Tabela 11 - Tabela Dim_School do DW	31
Tabela 12 - Tabela Dim_Student do DW	33
Tabela 13 - Tabela Dim_Study do DW	33
Tabela 14 - Tabela Fact_Grades do DW	34

1 INTRODUÇÃO

A tendência de massificação verificada nos últimos anos no ES em Portugal tem proporcionado o acesso a grupos de estudantes muito diferentes socioculturalmente uns dos outros, com maior presença de estudantes não tradicionais, ou seja, trabalhadores-estudantes, estudantes internacionais, estudantes de cursos de especialização profissional e tecnológica e de regimes especiais.

Associada à diversidade de alunos que acedem ao ES, surge a questão das suas competências e motivações académicas, uma vez que o sistema de *numerus clausus*¹ instituído favorece o aumento da percentagem de alunos não colocados nos pares “curso-instituição” de primeira opção. Por outro lado, a implantação de sistemas de garantia de qualidade nas IES leva a que estas instituições se preocupem cada vez mais com questões relacionadas com a melhoria da eficiência do ensino-aprendizagem.

As IES precisam, portanto, de ferramentas tecnológicas que lhes permitam compreender melhor quem são os seus candidatos e quais os fatores que influenciam a taxa de sucesso académico dos alunos matriculados nos seus cursos. Com esta informação disponível, poderão atuar de forma mais precisa no acompanhamento específico e mais personalizado de grupos de alunos com características de percurso educativo/resultados académicos identificados, bem como otimizar os seus recursos na oferta formativa e conseqüentemente na captação de alunos.

Os sistemas BI, como sistemas idealmente focados na análise e monitorização do desempenho das organizações, poderão servir esse propósito, ou seja, ser utilizados no setor do ES, não só na área da gestão administrativa como também na análise do desempenho académico, com o objetivo de suportar melhores tomadas de decisão que contribuam para a qualidade do ensino e a melhoria do processo ensino-aprendizagem.

Este projeto pretende endereçar a utilização de sistemas de BI no ES que atuem no âmbito da monitorização e análise académica e demonstrar as várias etapas da respetiva implementação e desafios associados, bem como as mais valias que estes aportam.

¹ Limite máximo estabelecido de indivíduos ou entidades que podem ser admitidos num organismo ou instituição (in Dicionário Priberam da Língua Portuguesa [consultado em 30-01-2022])

1.1 Objetivos e contribuição do trabalho

O trabalho apresentado neste documento resulta de um Projeto realizado numa IES privada, a Universidade Portucalense (UPT).

Pretende-se analisar, utilizando uma prova de conceito de um sistema de BI, as características dos alunos matriculados nos vários cursos da UPT, e determinar a influência de diversos fatores no seu sucesso escolar. Esses fatores incluem dados relacionados com a formação académica dos alunos antes de ingressar no ES e dados sociodemográficos que serão cruzados com informações sobre o desempenho dos alunos ao longo e na conclusão do curso na IES.

A contribuição deste trabalho pode ser sintetizada em dois aspetos principais:

Por um lado, demonstra-se a utilidade da aplicação de um sistema de BI para monitorização do desempenho académico dos alunos de uma IES e para um melhor conhecimento dos fatores externos e sociodemográficos com impacto nesse desempenho, permitindo a definição de iniciativas que contribuam para a melhoria da qualidade e da eficiência do ensino-aprendizagem.

Assim, o estudo dos indicadores e a segmentação dos alunos realizados no âmbito deste projeto poderão servir de base, numa futura implementação com dados reais (obtidos com o devido consentimento nos termos da Lei da Proteção de Dados Pessoais), para um melhor conhecimento dos perfis dos alunos da IES alvo e para a compreensão dos fatores que influenciam o desempenho académico dos alunos.

Com o conhecimento obtido, a IES poderá desenhar com mais facilidade uma estratégia diferenciadora, tendo em conta os respetivos perfis e necessidades evidenciadas, melhorando a comunicação da oferta formativa e criando oportunidades a nível curricular e extracurricular.

Por outro lado, o modelo de dados definido para o *Data Warehouse* (DW), os indicadores de eficiência formativa e de caracterização dos alunos e as correlações identificadas poderão servir de base a outros projetos e iniciativas de análise do desempenho dos alunos no ES. A utilização de uma ferramenta de software de BI adequada para o efeito será também explorada neste trabalho para averiguar o nível de conhecimentos técnicos necessário, a usabilidade da ferramenta e as facilidades de visualização gráfica disponibilizadas.

1.2 Metodologia adotada

Para desenvolver o projeto reportado neste documento recorreu-se a uma metodologia baseada numa prova de conceito. Esta metodologia está associada frequentemente à decisão da implementação de soluções tecnológicas e permite avaliar a viabilidade do projeto a ser desenvolvido e se este resulta numa mais-valia perceptível para o negócio, justificando o investimento a ser efetuado. A prova de conceito resulta na documentação conceptual e implementação que visa testar se uma determinada aplicação tecnológica satisfaz os objetivos predefinidos antes de ser tomada a decisão final da implementação e feito o esforço do investimento.

No âmbito deste projeto, não sendo possível utilizar dados reais dos alunos da IES analisada por questões relacionadas com a implementação do Regulamento Geral da Proteção de Dados (RGPD), a utilização desta metodologia permitirá a demonstração do objetivo principal do trabalho.

1.3 Estrutura do trabalho

O presente documento está dividido em 4 capítulos, estruturados no sentido de apresentar o estudo e trabalho realizados durante o projeto.

Assim, no primeiro capítulo é feito um enquadramento geral do tema a abordar e é contextualizado o estudo a desenvolver. São identificados os objetivos, principais contribuições e metodologia a seguir para a concretização do trabalho.

O segundo capítulo apresenta uma revisão bibliográfica sobre os conceitos, arquitetura e componentes de um sistema de BI, bem como uma visão geral das várias vertentes da aplicação do BI no ES e uma análise comparativa dos principais softwares de BI considerados líderes no mercado.

No terceiro capítulo é apresentada a prova de conceito da aplicação do sistema de BI ao caso da UPT, com a documentação das várias etapas envolvidas e resultados obtidos. É efetuada uma descrição dos dados envolvidos, a identificação da análise e correlações necessárias, os principais Indicadores Chave de Desempenho (KPIs), o modelo do DW considerado e o processo ETL. São ainda apresentados os principais resultados da análise OLAP e os *dashboards* desenvolvidos.

No último capítulo são revisitados os objetivos do projeto, é efetuada uma análise à prova de conceito desenvolvida e são feitas considerações sobre os resultados obtidos e limitações associadas. Um balanço do trabalho realizado é apresentado, finalizando com sugestões para futuras intervenções.

2 REVISÃO DA LITERATURA

2.1 Conceito de Business Intelligence

O conceito de *Business Intelligence* (BI) foi evoluindo ao longo do tempo até à sua utilização atual. O termo “BI” foi utilizado pela primeira vez em 1865 por Richard Millar Devens, num livro intitulado *Cyclopaedia of Commercial and Business Anecdotes*, para descrever como Henry Furnese lucrou ao recolher informação de modo a poder agir antes dos seus concorrentes (Foote, 2017).

Mais tarde, em 1958, Hans Peter Luhn’s distinguiu BI como uma solução tecnológica caracterizando-a como um “sistema com capacidade de aprender as inter-relações dos factos apresentados de forma a orientar a tomada de decisão em direção a um objetivo desejado (Luhn, 1958). Por outras palavras, a capacidade de analisar e identificar relações entre os dados, utilizando depois o resultado desta análise para orientar a tomada de decisão para atingir um objetivo (Lester, 2018).

Em 1989, Gartner Howard Dersner (Apraxine & Stylianou, 2017) definiu BI como sendo um “guarda chuva” que abrange as aplicações, infraestruturas, ferramentas e melhores práticas que permitam o acesso e análise da informação, de modo a facilitar as tomadas de decisão.

Em 2013 foi publicado o livro *Oracle Data Warehousing*, escrito por *Robert Stackowiak, Joseph Rayman e Rick Greenwald* (Stackowiak et al., 2013) em que o processo de BI é definido como a recolha de grandes quantidades de dados, transformação, análise e apresentação da informação num conjunto de relatórios de alto nível condensados e dirigidos com informação essencial à tomada de decisão.

Atualmente, os sistemas de BI utilizam os dados disponíveis nas organizações e dados externos, transformando-os em informação relevante para suporte à tomada de decisão. Para esse fim, são utilizadas um conjunto de ferramentas que permitem, entre outras funcionalidades, a criação de relatórios, que serão utilizados pela gestão de topo e cada vez mais por outros colaboradores das empresas, para auxiliar na tomada de decisão (Santos & Ramos, 2017).

Na elaboração deste trabalho será considerada a definição de BI disponibilizada em (Ramos, 2016): “*BI consiste num conjunto de estratégias, tecnologias e*

metodologias cujo objetivo é ajudar na transformação dos dados disponíveis nas organizações em informação de qualidade e a informação em conhecimento permitindo assim às organizações tomar decisões de forma mais acertada e no tempo certo”.

2.2 Arquitetura e componentes de um sistema de Business Intelligence

Atualmente os gestores responsáveis pela tomada de decisão sabem que a informação atempada e precisa é crucial para melhorar o desempenho das organizações. Por esta razão, a utilização dos sistemas de BI tem aumentado significativamente nas organizações (Santos & Ramos, 2017). Como discutido na secção anterior, um sistema de BI combina dados com ferramentas analíticas, de forma a disponibilizar informação relevante para a tomada de decisão.

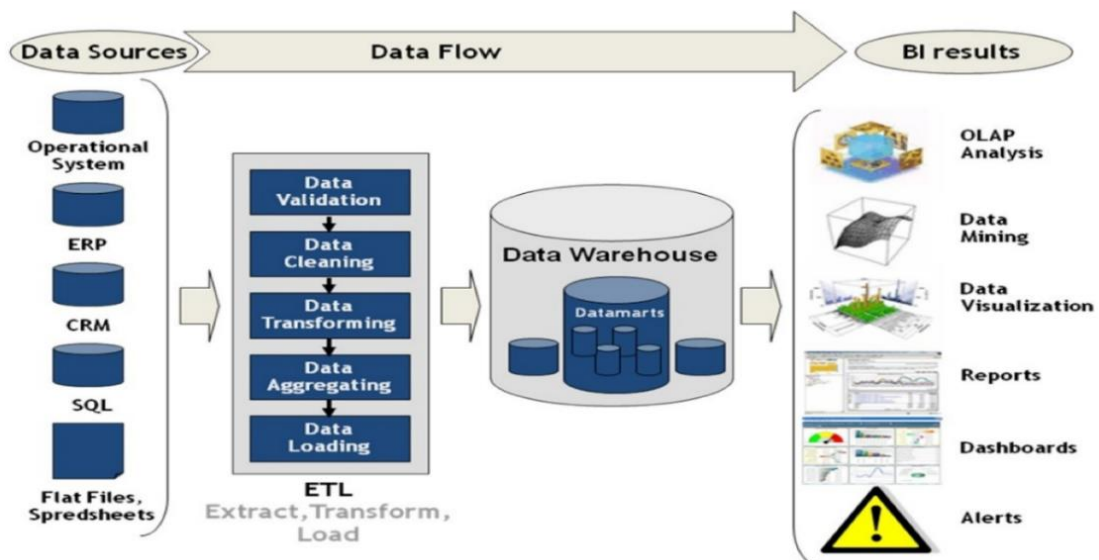


Figura 1 - Arquitetura de um sistema de BI (Ahmad et al., 2020)

A Figura 1 apresenta a arquitetura de um sistema típico de BI. As componentes principais desta arquitetura serão descritas nas secções 2.2.1 a 2.2.5.

2.2.1 Processo ETL

O processo designado por *Extract, Transform, Load* (ETL) é o processo que permite que os dados sejam extraídos de bases de dados ou sistemas operacionais, e transformados e carregados numa nova base de dados com capacidade para armazenar grandes volumes de dados designada por *Data Warehouse* (DW) (Dedić & Stanier, 2016). Genericamente, o resultado do processo de ETL é a consolidação de dados oriundos de diversas fontes – designadas sistemas organizacionais – num DW ou *data mart*.

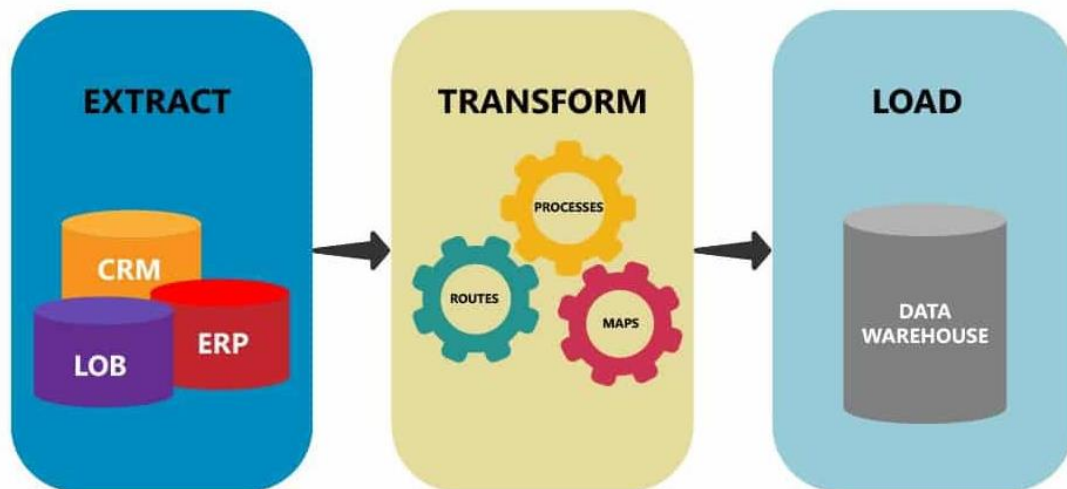


Figura 2 - O processo ETL (Fatima, 2020)

Assim, o processo ETL inclui 3 etapas, representadas graficamente na Figura 2 e descritas no texto que se segue:

- Extrair (*Extract*);
- Transformar (*Transform*);
- Carregar (*Load*).

A extração dos dados permite a recolha dos dados de vários sistemas e bases de dados operacionais ou de outros ficheiros.

A transformação dos dados consiste na conversão dos dados extraídos para o formato pretendido. O nível de transformação dos dados dependerá do formato inicial

dos mesmos; quanto mais este for semelhante ao formato final, menos operações de transformação serão necessárias. A transformação pode envolver as seguintes etapas:

- Limpeza;
- Filtragem;
- Normalização dos dados;
- Validação do fluxo dos dados;
- Validação dos limites dos dados;
- Transposição de linhas e colunas;
- Junção;
- Separação;
- Ordenação.

O carregamento dos dados é a última etapa do processo e consiste na inserção dos dados extraídos e transformados no DW. Este carregamento pode ser dividido em três fases:

- Carregamento inicial;
- Atualização incremental;
- Atualização completa.

2.2.2 Data Warehouse

O conceito de *Data Warehouse* (DW) remonta aos finais dos anos 80, quando os investigadores da IBM Barry Devlin e Paul Murphy desenvolveram o “*Business DW*”. Este conceito representa um modelo de armazenamento para os dados provenientes dos sistemas operacionais. Assim, dados que anteriormente se encontravam espalhados por vários repositórios agora passam a ser agregados num único repositório (o DW), permitindo aos utilizadores, com uma única ferramenta, consultar informação de maneira mais eficiente (Shah, 2014).

A análise dos dados num DW é efetuada sob a forma de consultas predefinidas, *ad-hoc* ou pela elaboração de relatórios que normalmente permitem a monitorização das alterações ocorridas nos principais indicadores ao longo do tempo.

Segundo (Inmon, 1996), um DW é uma base de dados mantida de uma forma autónoma e orientada para a análise dos dados, ao invés das bases de dados operacionais que são focadas no processamento de dados. Toda a informação é

mapeada temporalmente e deve permitir o armazenamento de dados de vários anos (dados históricos), pelo que os requisitos de espaço de armazenamento são exigentes (bases de dados com capacidade de armazenamento na ordem dos *terabytes* e *petabytes*). Os dados são armazenados segundo estruturas otimizadas para a análise dos dados e que potenciem a rapidez da análise. Os utilizadores finais podem fazer consultas complexas ao DW, sendo permitidos apenas acessos de leitura, libertando assim as bases de dados operacionais para a recolha, armazenamento e manipulação da informação do dia-a-dia.

Na Tabela 1 são apresentadas as principais diferenças entre as bases de dados operacionais e um DW.

Base de dados operacionais	DW
Objetivos operacionais	Registo histórico
Acessos de leitura/escrita	Acessos só de leitura
Acesso por transações predefinidas	Acesso por questões <i>ad-hoc</i> e relatórios periódicos
Acesso a poucos registos de cada vez	Acesso a muitos registos de cada vez
Dados atualizados em tempo real	Carregamentos periódicos de mais dados
Estrutura otimizada para atualizações	Estrutura otimizada para análise de dados

Tabela 1 - Diferenças entre base de dados operacionais e DW

Modelação dos dados num DW

As bases de dados tradicionais recorrem a técnicas de modelação, tais como modelos relacionais e normalização, com o foco na eficácia e consistência dos dados. No entanto, no universo do BI este tipo de estrutura não é o mais adequado, dado que frequentemente a estrutura das bases de dados torna-se muito complexa para os utilizadores finais poderem executar consultas. Por norma, uma base de dados operacional é constituída por várias tabelas ligadas por uma complexa teia de ligações.

Na criação de um DW a técnica de modelação mais utilizada é a modelação multidimensional. A abordagem foi inicialmente definida por Kimball em 1996, embora

este autor alegue não ter sido o inventor mas meramente tendo documentado o que já era praticado há muitos anos (Moody & Kortink, 2003).

Esta técnica tem como principais objetivos a criação de um DW de fácil compreensão e utilização, bem como a otimização do desempenho nas consultas. A modelação dimensional poderá ser efetuada de acordo com um dos seguintes três esquemas:

- Esquema em estrela;
- Esquema em floco de neve;
- Esquema em constelação.

Esquema em estrela

O esquema em estrela é o esquema mais utilizado para modelar dados de forma multidimensional. Ao contrário de um esquema relacional de uma base de dados, que é desenhado para ser eficiente no armazenamento dos dados e evitar redundância dos mesmos, o esquema em estrela é desenhado de forma a facilitar as consultas (Santos & Ramos, 2017).

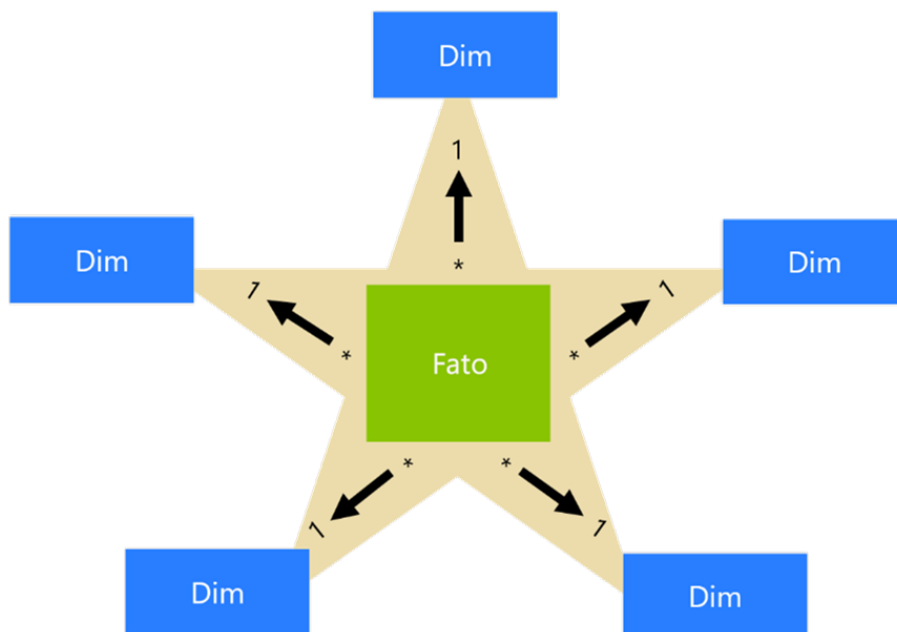


Figura 3 - Esquema em estrela adaptado (Myers, 2019)

O esquema em estrela foi desenvolvido por Kimball em 1997 e tem como principais características a desnormalização das tabelas. No esquema em estrela a informação é organizada em tabelas, cada uma correspondente a um determinado assunto. Para alcançar este objetivo a informação é dividida em dois tipos de tabelas: uma tabela de factos, que contém os dados a analisar (informação quantitativa), e várias tabelas dimensão, que incluem informação qualitativa complementar aos factos, conforme ilustrado na Figura 3 (Garani & Helmer, 2012).

A relação entre a tabela de factos e as tabelas dimensão é uma relação de (n:1), uma vez que um registo de uma tabela dimensão pode estar associado a vários registos da tabela de factos (Santos & Ramos, 2017).

Por norma, os registos contidos na tabela de factos são dados numéricos que podem ser analisados (aspectos quantitativos ou medidas da análise), mas nesta tabela também se encontram as chaves estrangeiras correspondentes às tabelas dimensão. As tabelas dimensão correspondem aos aspectos qualitativos da análise e normalmente permitem responder às perguntas “a quem?”, “quando?”, “onde?”, “por quem?”, “como?”, entre outras. Um esquema em estrela tem por norma uma tabela de factos e várias dimensões (Garani & Helmer, 2012).

Tabela de Factos	Tabela de dimensão
Constituída por um conjunto de atributos numéricos (dados a analisar) e por um conjunto de atributos (chaves estrangeiras) que se relacionam com as diferentes dimensões.	Existem tantas dimensões quantas as vertentes qualitativas pelas quais se pretende analisar os factos
Normalizada	Desnormalizada, apresentam na sua maioria uma grande quantidade de atributos
Contém grande parte dos registos do DW	Contém poucos registos comparadas com a tabela de factos

Tabela 2 - Diferenças entre tabela de factos e tabela de dimensão

A terminologia do modelo multidimensional provém do facto de o modelo em estrela poder ser visualizado como um cubo, em que cada dimensão do modelo pode caracterizar uma face do cubo. As tabelas dimensão são, por norma, desnormalizadas

podendo conter uma ou mais hierarquias. Segundo (Santos & Ramos, 2017), as características que melhor representam o modelo em estrela podem ser observadas na Tabela 2.

Por ser uma estrutura simples, o esquema em estrela é de fácil compreensão visual, facilitando a sua discussão, assim como as análises dos utilizadores.

Esquema em floco de neve

Um esquema em floco de neve (*Snowflake*) é muito similar ao esquema em estrela, com a particularidade de as dimensões estarem parcialmente ou completamente normalizadas. O esquema deixa de ter uma estrutura regular, uma vez que uma dimensão pode ter sub-dimensões (Santos & Ramos, 2017).

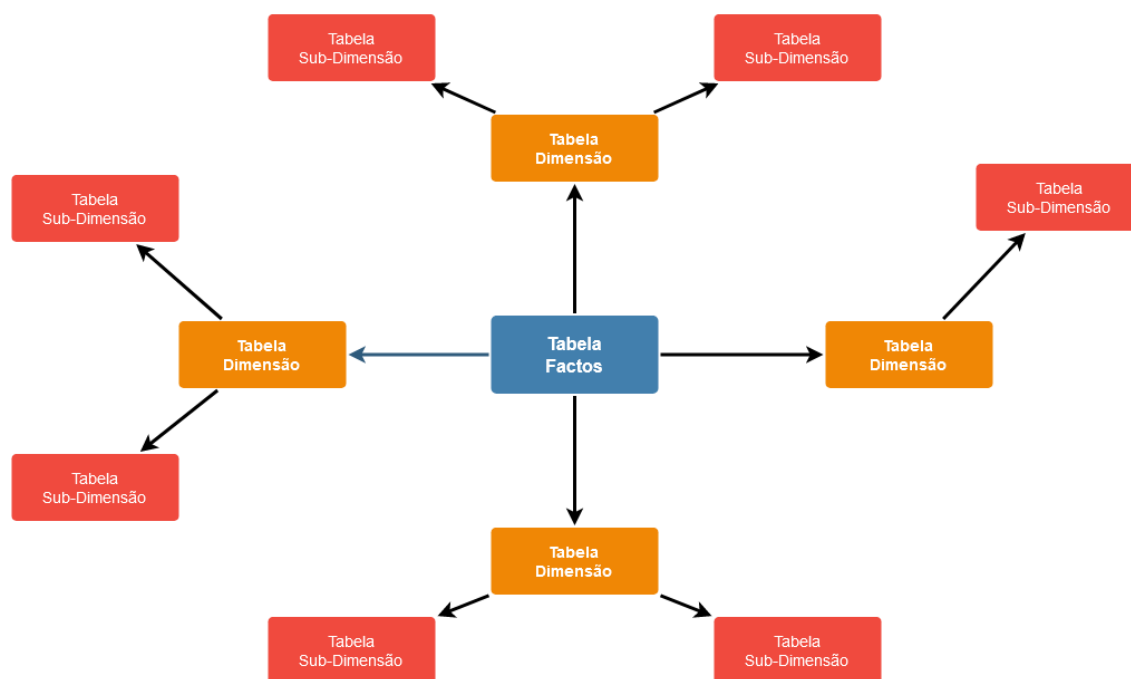


Figura 4 - Esquema em floco de neve (*Snowflake*)

O esquema em floco de neve foi desenvolvido com o objetivo de reduzir a desnormalização das tabelas dimensão e pode ser obtido normalizando as tabelas dimensão até um certo nível, criando sub-tabelas, conforme ilustrado na Figura 4. Neste esquema, a adição de nova informação torna-se mais simples e a duplicação

(redundância) da informação é reduzida. No entanto, o aumento do número de tabelas tem como consequência o aumento do número de *joins* necessários para as análises a efetuar, pois vão existir mais ligações extra entre dimensões e sub-dimensões, tornando-as mais lentas (Garani & Helmer, 2012).

A maior desvantagem deste esquema está associada à dificuldade de interpretação por parte do utilizador, e à perda de desempenho devido às normalizações (Santos & Ramos, 2017).

Esquema em constelação

O esquema em constelação é um esquema em que múltiplas tabelas de factos estão interligadas pela partilha de tabelas dimensão.

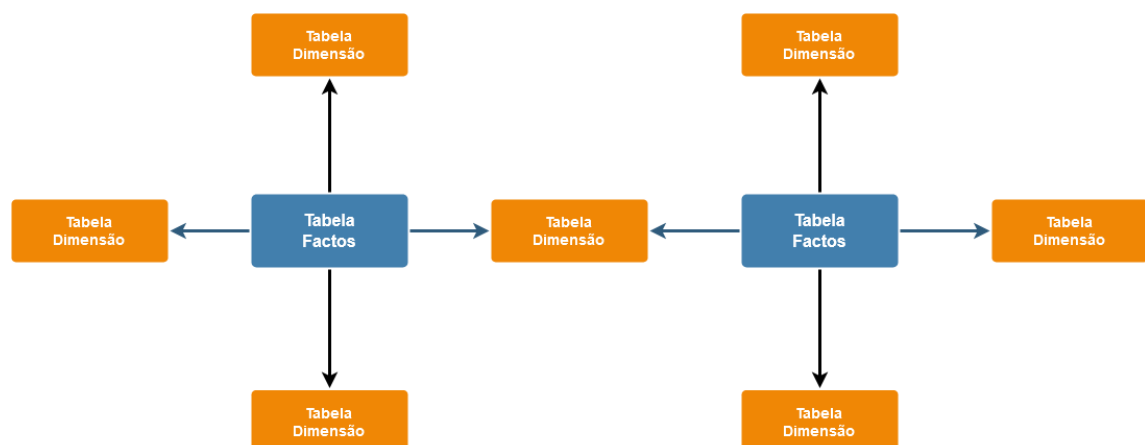


Figura 5 - Esquema em constelação (Last night Study, n.d.)

Conforme ilustrado na Figura 5, este esquema pode ser visto como vários esquema em estrela, com a particularidade de entre eles existirem dimensões comuns (partilhadas) (Santos & Ramos, 2017).

Uma das grandes desvantagens deste esquema é a complexidade de interpretação por parte do utilizador e a perda de desempenho devido à normalização. Em contrapartida, uma das vantagens é a flexibilidade na obtenção de informação, pois tem uma maior riqueza e flexibilidade na implementação (S, 2018).

2.2.3 Análise OLAP

Um DW pode ser explorado de várias formas, sendo que a mais comum recorre à tecnologia *OnLine Analytical Processing* (OLAP). Esta tecnologia permite a criação de cubos, como o apresentado na Figura 6, para analisar a informação sob várias perspetivas ou dimensões (Myers, 2019).

Assim, de acordo com a Figura 6, as vendas podem ser analisadas de acordo com as dimensões *Product* (Produto), *Geography* (Local) e *Time* (Tempo). As dimensões estão, por sua vez, organizadas em diferentes níveis de hierarquia que permitem diferentes níveis de granularidade na análise efetuada. Normalmente, a análise das vendas na dimensão “Tempo” poderá ser feita ao nível do ano, trimestre (*quarter*), mês e dia. A análise das vendas para a dimensão “Local” poderá ser feita ao nível do continente, país e cidade, etc.

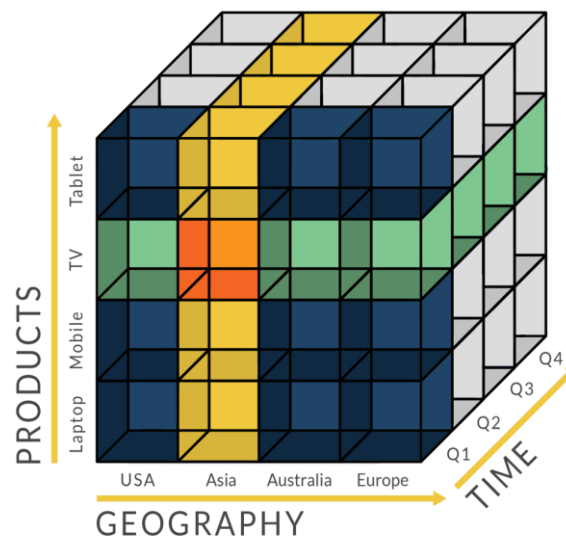


Figura 6 - Análise OLAP (Myers, 2019)

Os cubos permitem analisar os factos disponíveis, na tabela de factos, pelas diferentes dimensões. Para a análise poder dar resposta a todas as questões do negócio, a modelação dos dados deve ser efetuada de forma a garantir que toda a informação necessária seja considerada (Santos & Ramos, 2017).

A tecnologia OLAP é utilizada em muitas aplicações de gestão de negócios, planeamento, orçamento, previsões atmosféricas e financeiras, modelos, simulações e relatórios de DW. O OLAP permite aos utilizadores executarem análises *ad-hoc* de

dados, sobrepondo várias dimensões e fornecendo assim uma visão mais alargada e o entendimento necessário para uma melhor tomada de decisão (OLAP.com, n.d.).

De acordo com (Santos & Ramos, 2017) existem quatro operações principais para manipulação de dados nos cubos, conforme apresentado na Tabela 3.

Operação OLAP	Descrição
<i>Drill-down</i>	Esta operação permite navegar pelos dados partindo do todo até ao detalhe. Um exemplo é explorar um facto (vendas), por ano, por trimestre e finalmente por mês, utilizando a hierarquia ano->trimestre->mês.
<i>Roll-up</i>	Esta operação é o oposto do <i>drill-down</i> . Permite navegar pelos dados utilizando uma dada hierarquia (mês->trimestre->ano)
<i>Slice and Dice</i>	Esta operação permite navegar pelos dados restringindo os mesmos. O corte (<i>Slice</i>) permite selecionar uma porção dos dados restringindo, para isso, a análise a uma dimensão do cubo. A redução (<i>Dice</i>) permite efetuar a análise cruzando duas ou mais dimensões.
<i>Pivot (rotate)</i>	Esta operação permite apresentar uma representação alternativa dos dados, alterando o posicionamento das dimensões nos eixos da análise gráfica.

Tabela 3 - Descrição das operações possíveis num cubo OLAP

2.2.4 Data Mining

A mineração de dados, vulgarmente designada por *Data Mining* (DM), é um processo que integra uma das fases do processo de descoberta de conhecimento em bases de dados (DCBD). Consiste na procura de relacionamentos, padrões e modelos implícitos nos dados, que não se conseguem descobrir a “olho nu”, isto é, sem recorrer a algoritmos específicos de mineração de dados. O DM é frequentemente utilizado como uma das técnicas de análise de dados consideradas num projeto de implementação de BI.

No processo de DCBD estão incluídas as fases de seleção, tratamento, pré-processamento, DM e interpretação de resultados, e o seu objetivo é a obtenção de

conhecimento (Piedade, 2011). O processo DCBD, além de iterativo é também iterativo uma vez que, sempre que seja necessária uma tomada de decisão, é requerida a presença do utilizador. A Figura 7 ilustra as diferentes fases do processo (Santos & Ramos, 2017), que são também descritas no texto que se segue.

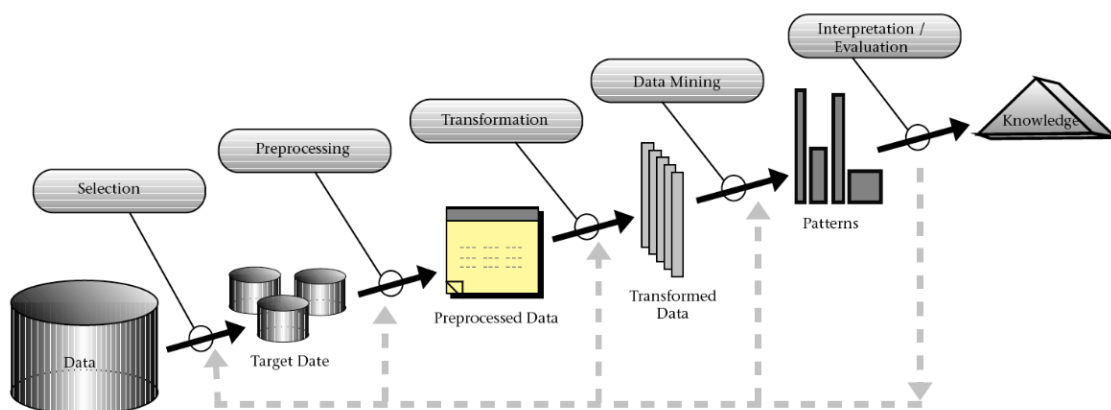


Figura 7 - Fases do processo de DCBD (Vaz, 2016)

Seleção

A fase de seleção consiste na seleção dos dados armazenados nos diversos repositórios, DW, entre outros. Os dados selecionados são os dados que vão ser necessários aos algoritmos de DM. Nesta seleção é também efetuada uma limpeza dos dados eliminando atributos que não têm interesse para o processo de descoberta de conhecimento (Piedade, 2011).

Pré-processamento dos dados

Nesta fase é realizada uma redução do espaço de pesquisa diminuindo o número de linhas ou de colunas a analisar. Para isso é feita uma transformação ao nível dos atributos, convertendo atributos contínuos em atributos discretos. Um exemplo dessa transformação consiste em substituir as idades por faixas etárias (Santos & Ramos, 2017).

Tratamento dos dados

Nesta fase do processo procede-se à limpeza e verificação dos dados, exceto se os mesmos tiverem vindo de um DW. Entre os procedimentos habituais nesta fase, destaca-se a identificação de registos duplicados, valores inconsistentes ou valores que não fazem sentido no contexto em que estão a ser utilizados (Santos & Ramos, 2017).

Data Mining

Esta fase consiste na pesquisa onde os dados provenientes da fase de pré-processamento vão ser analisados usando técnicas de DM. Para atingir os objetivos propostos pode ser necessário utilizar mais do que uma técnica. Esta fase é iniciada com a definição dos objetivos a atingir. Em função do tipo de resultado pretendido é definida uma tarefa (ou método de DM) a executar e é também identificada a técnica de DM a utilizar, que se revela mais adequada aos dados, de modo a obter padrões. A Tabela 4 apresenta as tarefas ou métodos principais utilizados no DM e técnicas respetivas.

Interpretação de resultados

Esta fase consiste na análise, verificação e validação dos resultados obtidos através dos algoritmos previamente selecionados. A validação é feita através da aplicação dos padrões encontrados a novos conjuntos de dados (Santos & Ramos, 2017).

Tarefa	Descrição	Técnica
Classificação	Enquadrar os dados em classes predefinidas, identificando a classe a que cada elemento pertence	Árvore de decisão Redes neuronais Algoritmos genéticos
Regressão	Prever o valor de uma variável contínua (variável dependente) em função de várias variáveis independentes	Regressão linear Regressão não linear Regressão logística
Segmentação (clustering)	Identificar um conjunto de segmentos (clusters) que dividem os dados analisados. Os segmentos surgem de agrupamentos que são detetados nos dados e que obedecem a métricas de similaridade	Redes neuronais Vizinhos mais próximos
Sumariação	Descrever um determinado conjunto de dados, fornecendo descrições resumidas do mesmo	Árvore de decisão Algoritmos genéticos Vizinhos mais próximos
Modelação de dependências	Identificar um modelo que descreva dependências significativas entre variáveis. Incluem-se como casos particulares e descrição de associações e de sequências nos dados.	Regras de associação

Tabela 4 - Tarefas e técnicas de DM

2.2.5 Monitorização e visualização dos KPIs

Um sistema de BI tem por base a definição, análise e monitorização de um conjunto de Indicadores de Desempenho ou *Key Performance Indicators* (KPIs). Estes indicadores constituem as medidas quantitativas ou métricas que permitem avaliar o desempenho ou grau de concretização de objetivos de negócio a atingir num período de tempo.

Um KPI é utilizado para definir um marco num determinado assunto, frequentemente entendido como um fator crítico de sucesso que se pretende alcançar, servindo para avaliar o seu progresso de uma forma objetiva (Qlik, 2021b)

Num sistema de BI, os KPIs são definidos no início do projeto em conjunto com os *stakeholders* do sistema. A implementação do sistema de BI permite monitorizar esses indicadores, que são geralmente disponibilizados aos utilizadores em formatos gráficos nomeadamente *dashboards* e *scorecards*,

2.3 Ferramentas de Software de Business Intelligence

Existem no mercado vários software de BI. Estes têm em comum o objetivo de recolher grandes volumes de dados provenientes de uma ou várias fontes internas e/ou externas, organizando-os e transformando-os em informação útil ou relevante que irá depois ser disponibilizada ao utilizador.

De acordo com (Richardson et al., 2020), atualmente existem três software líderes no mercado de software de BI: Microsoft Power BI, Tableau, Qlik Sense.

Estas soluções pertencem ao “Quadrante Mágico” das plataformas de BI publicado pela consultora Gartner (cf. Figura 8) e serão resumidamente descritas no texto que se segue, destacando-se os pontos fortes e fracos apresentados por cada uma.



Figura 8 - Magic Quadrant for Analytics and Business Intelligence Platforms (Analytics and Business Intelligence Platforms, 2020)

Qlik

O Qlik aposta em ferramentas de *Machine Learning* e Inteligência Artificial, recorrendo para isso ao *Qlik Cognitive Engine* que, em conjunto com o *Qlik Associative Engine*, oferecem sugestões de análise. O Qlik oferece flexibilidade nas suas soluções fornecendo uma solução SaaS, *on-premises* e uma solução mista (Qlik, 2021a).

Vantagens

- Facilidade na implementação com soluções SaaS, *on-premises* e mistas;
- *Qlik's Data Literacy* é um programa que ajuda utilizadores, com diferentes graus académicos, a utilizar, compreender e migrar da antiga versão para as novas (QlikView para Qlik Sense).

Desvantagens

- Licenciamento varia de versão para versão, o que aumenta a complexidade na escolha da versão;
- Poucos utilizadores comparativamente com outros líderes de mercado.

Microsoft Power BI

O Microsoft Power BI oferece um grande conjunto de funcionalidades, nomeadamente, carregamento dos dados e respetiva preparação, descoberta de relações entre os dados de forma gráfica, dashboards interativos e análise avançada dos dados. Pode ser utilizado localmente (*on-premises*) ou como SaaS (*cloud azure*). A versão SaaS tem atualizações semanais. Das funcionalidades disponibilizadas em 2020 destacam-se ferramentas avançadas de análise de dados recorrendo a tecnologia de inteligência artificial e deteção de anomalias (Microsoft, 2020).

Vantagens

- Integração com Office 365 e Microsoft Teams.

Desvantagens

- A Microsoft não dá a possibilidade de escolha do serviço cloud, visto que o Power BI só corre no serviço Azure da Microsoft.

Tableau

O Tableau distingue-se dos seus concorrentes por oferecer uma experiência visual para exploração dos dados bastante agradável, permitindo aos utilizadores o acesso à informação com poucos cliques. Em 2020, reforçou e melhorou a área de preparação e gestão dos dados. Em relação à preparação dos dados lançou novas funcionalidades para a modelação dos mesmos, tornando mais fácil a análise entre tabelas em diferentes níveis (Richardson et al., 2020).

Vantagens

- Mantém uma experiência agradável e simples apesar das várias atualizações ao longo do tempo;
- Oferece uma ferramenta simples que permite integração com o Salesforce;

Desvantagens

- Não é uma ferramenta desenvolvida de raiz para a cloud; embora existam ferramentas online tais como Tableau Online e Tableau CRM, a maioria das instalações é feita on-premises.

A Tabela 5 apresenta uma síntese comparativa das três ferramentas anteriormente descritas.

	Microsoft Power BI	Tableau	Qlik Sense
Integração	Pode-se ligar com várias fontes de dados, mas está perfeitamente otimizado para se conectar com Microsoft Azure	Pode-se ligar com várias fontes de dados	Pode-se ligar com várias fontes de dados
ETL integrado	Não	Não	Sim
Modelação de dados	Simples	Simples	Avançada
Combinação de dados	Permite, mas em menor escala, comparada com Tableau e Qlik Sense	Permite	Permite uma maior combinação de dados comparada com outros softwares
Análise de dados	Sim, fornece análise e casos de uso através da Azure ML Web Services	Sim, fornece modelos de inteligência artificial através de extensões	Sim
Visualização dos dados	Sim, permite manipulação dos dados com acesso a gráficos simples	Sim, simplicidade na obtenção de gráficos através de uma interface bastante intuitiva, o que possibilita que utilizadores com diferentes capacidades consigam obter gráficos.	Sim, possui um forte <i>backend</i> que permite a manipulação dos dados, o que possibilita a obtenção de vários gráficos diferentes.

Tabela 5 - Comparação de ferramentas de BI (Sukumar, 2021)

2.4 Business Intelligence no Ensino Superior

Atualmente as Instituições de Ensino Superior (IES) atuam num meio altamente competitivo, em que fatores internos e externos podem ter um impacto na performance da instituição. Nesse contexto, uma apropriada monitorização do desempenho das IES em várias vertentes é essencial, tanto mais que existem normalmente grandes volumes de dados envolvidos, gerados pelas atividades do dia-a-dia destas organizações.

As soluções de BI no ES são normalmente aplicadas nas áreas da gestão administrativa, análise académica (garantia da qualidade do ensino e melhoria do desempenho académico), marketing e gestão.

A análise dos dados relacionada com a caracterização dos estudantes, admissões, retenções e desempenho pode ser referenciada como análise académica (Apraxine & Stylianou, 2017). Alguns exemplos de intervenções neste âmbito que poderão beneficiar da implementação de soluções de BI são:

- Admissão de estudantes;
- Área de tutoria especial;
- Melhorar níveis de desistência;
- Avaliação do corpo docente;
- Melhoria dos currículos;
- Ajudar os alunos na preparação de planos de estudo;
- Determinar quais os cursos mais acessíveis para ensino pelo e-learning.

No campo da garantia da qualidade do ensino e da melhoria do desempenho académico, com base na análise dos dados relevantes, as instituições podem ter uma visão mais abrangente dos seus estudantes, analisando as suas preferências académicas e desempenho, e tomar iniciativas de forma a melhorar os seus resultados e taxas de retenção.

Assim, as instituições podem desenvolver programas direcionados a cada grupo de estudantes, contribuindo para um melhor desempenho individual e, conseqüentemente, para o aumento do *rating* da instituição. Adicionalmente, podem ser utilizadas ferramentas de DM, de modo a dividir os estudantes em grupos com base no seu desempenho e, assim, ajustar o processo de ensino-aprendizagem a cada grupo (Baepler & Murdoch, 2010).

Genericamente, a utilização de ferramentas de BI permite aos responsáveis pelo ensino analisarem a informação académica dos alunos. Já a monitorização permite aos professores tomarem ações que venham a prevenir o risco de os alunos falharem e até mesmo levar a um aumento da performance no geral. Através de ferramentas de BI, os professores podem identificar e avaliar muitos indicadores sociodemográficos e de performance, tais como número de alunos, média de idades, performance académica, taxas de retenção e ajustar as atividades letivas e metodologias de ensino-aprendizagem com base nos dados obtidos (Apraxine & Stylianou, 2017).

No campo do marketing e gestão, a análise da informação académica pode fornecer à gestão das IES uma visão sobre os cursos que os alunos preferem, a procura do mercado, as unidades curriculares (UC) mais escolhidas (quando existe essa opção). Com essa informação as IES poderão ajustar a oferta, criando ou descontinuando cursos de uma forma mais sustentada (Franklin Jr., 2015).

A informação extraída pode também contribuir para um aumento da vantagem competitiva da instituição. Informação sobre a origem demográfica dos alunos, idade, género e médias podem ajudar o departamento de marketing a ajustar as políticas de modo a reduzir custos ou expandir as campanhas e outros *targets*. Ter uma noção sobre os cursos mais populares entre os estudantes e os requisitos do mercado, permite à instituição fazer um melhor planeamento, o que proporciona uma vantagem competitiva adicional (Baepler & Murdoch, 2010).

Uma IES possui em geral um grande número de alunos, o que gera uma grande quantidade de informação, desde o início do processo (admissão) até ao final do processo (graduação). Estes dados são críticos no processo de tomada de decisão. Através de ferramentas de BI, podem estar centralizados e serem disponibilizados a todos os departamentos sob a forma de relatórios previamente construídos, apenas com a informação necessária a cada um, melhorando todo o processo interno (Apraxine & Stylianou, 2017).

Como qualquer outra organização, uma IES tem um conjunto de objetivos estratégicos que pretende alcançar. A utilização de uma solução de BI pode ajudar a alcançar os objetivos traçados, produzindo informação de qualidade e suportando melhor o processo de tomada de decisão.

Atualmente as IES utilizam diferentes abordagens ao BI, mas independentemente da abordagem, os objetivos da análise dos dados são na generalidade os mesmos: transformar o enorme volume de dados que possuem em informação útil. Embora existam atualmente no mercado vários software que permitem grande parte dos resultados esperados, estes são muitas vezes pouco flexíveis com relatórios standards e não têm funcionalidades necessárias para os diferentes processos existentes numa universidade. Para contornar esse problema, as IES optam frequentemente por desenvolver ou até mesmo adquirir um software de BI mais robusto, com mais funcionalidades e mais flexível (Fred Friedrich, 2013).

3 CASO DE ESTUDO

3.1 Descrição

Neste capítulo será apresentado o caso de estudo da UPT, tendo sido desenvolvida uma prova de conceito para um sistema de BI, que permita analisar a influência de vários fatores no sucesso académico dos alunos. Esses fatores incluem dados relacionadas com o background académico dos alunos anterior ao ingresso no ES e dados sociodemográficos que serão cruzados com a performance dos alunos durante e na conclusão do curso. Este projeto contribuirá para a melhoria e personalização do processo ensino-aprendizagem na UPT.

Utilizando um sistema de BI, uma IES poderá melhor conhecer os alunos no que diz respeito ao seu background académico, comportamento e preferências, informação demográfica e resultados académicos obtidos (incluindo taxas de retenção e graduação), assim como as correlações entre os diferentes aspetos. Esta informação poderá contribuir para uma melhoria no processo de tomada de decisão, bem como, para a maximização de benefícios, capitalizando oportunidades e ganho de vantagem competitiva (Apraxine & Stylianou, 2017).

A UPT é uma IES privada criada em junho de 1986 e localizada no Porto. Está organizada em seis departamentos:

- Arquitetura e Multimédia Gallaecia;
- Ciência e Tecnologia;
- Direito;
- Economia e Gestão;
- Psicologia e Educação;
- Turismo, Património e Cultura.

Nos vários departamentos são disponibilizados cursos de 1º, 2º e 3º ciclo, assim como cursos não conferentes de grau. Atualmente, a universidade tem cerca de 3300 alunos. Sendo uma IES privada, a sua principal estrutura de gestão está dividida num Conselho de Administração e na Reitoria. A UPT assumiu o princípio de que a qualidade e a garantia da qualidade são os principais vetores para o seu funcionamento e desenvolvimento.

3.2 Dados a analisar

Para o estudo a desenvolver foi gerado aleatoriamente pela UPT um conjunto de 730 registos fictícios baseados em dados originais de alunos matriculados nos cursos de 1º ciclo nos anos letivos de 2011/2012 a 2018/2019. Foram consideradas as seguintes variáveis para análise:

- Escola do ensino secundário frequentada;
- Regime e nota de ingresso;
- Sexo;
- Idade;
- Região de proveniência;
- Nível educacional e situação profissional dos pais.

Também por questões relacionadas com o RGPD, os dados foram anonimizados, tendo sido retirados todos os campos que permitissem a identificação individual dos alunos e respetivos pais, e correlacionados com dados relativos ao desempenho dos alunos ao longo do curso, nomeadamente:

- Inscrições por UC e nota obtida;
- Número de anos para conclusão do curso;
- Classificação final.

O Qlik Sense foi o software de BI escolhido para o desenvolvimento deste projeto, devido às facilidades de visualização gráfica disponibilizadas e à possibilidade de integração com outros projetos da UPT já implementados com este software.

Foram realizadas reuniões com o Diretor de um dos departamentos da UPT para recolher informações adicionais sobre os objetivos de negócio da instituição, bem como ajudar na identificação de questões de negócio que deveriam ser abordadas neste estudo. Foram também realizadas reuniões com representantes do Centro de Sistemas de Informação da UPT para identificação dos dados do sistema operacional académico (SIUPT) a serem gerados e a incluir no estudo a realizar.

3.3 Correlações e KPIs

Com base numa análise preliminar dos dados fornecidos pela UPT e nas entrevistas realizadas com os representantes da Instituição foram identificados os seguintes aspetos de negócio a serem abordados pelo estudo:

- a) Qual a correlação entre a nota de ingresso do aluno, o desempenho nas suas UC e a nota final de curso?
- b) Qual a correlação entre a nota final de curso e a escola frequentada pelo aluno no secundário?
- c) Segmentos de alunos por regime de ingresso e nota final de curso.
- d) Segmentos de alunos por situação profissional dos pais.
- e) Segmentos de alunos por nível de escolaridade dos pais.
- f) Segmentos de alunos por género e nota de ingresso.
- g) Segmentos de alunos por origem geográfica (região de proveniência) e nota de ingresso.
- h) Segmentos de alunos com performance académica correspondente a:
 - a. Excelente (curso concluído em n anos e nota final ≥ 16).
 - b. Bom (curso concluído em n anos e nota final ≥ 13 e < 16).
 - c. Razoável (curso concluído em n anos e nota final < 13).
 - d. Suficiente (curso concluído em n+1 anos e nota final ≥ 12).
 - e. Insuficiente (curso concluído em n+1 anos e nota final ≥ 10 e < 12).
 - f. Outros (curso concluído em n+2 anos ou mais).
- i) Correlação entre a nota final de curso e a nota de ingresso.
- j) Segmentos de alunos por género e nota de ingresso, relacionados com a nota final.
- k) Eficiência no ensino – Percentagem de graduados (por ano letivo).

Um conjunto de KPIs foi também definido como relevante para o estudo a desenvolver. Estes indicadores de desempenho são apresentados na Tabela 6, organizados de acordo com os objetivos relacionados a alcançar e que contribuem para o propósito geral deste projeto, sendo também incluída a respetiva métrica de cálculo.

Objetivo	KPI	Métrica de calculo
A - Conhecer as características dos alunos	1 - Número de alunos inscritos por género	Nº de alunos inscritos, por sexo
	2 - Número de alunos por idade	Nº de alunos inscritos, por idade
	3 - Número de alunos por origem geográfica em Portugal	Nº de alunos inscritos, agrupados por distrito de origem
	4 - Número de alunos por origem sociodemográfica dos pais – escolaridade	Nº de alunos inscritos por escolaridade dos pais
	5 - Número de alunos por origem socioeconómica dos pais – situação profissional	Nº de alunos inscritos por situação profissional dos pais
	6 - Número de alunos deslocados da sua residência permanente	Nº de alunos deslocados da residência permanente
	7 - Número de alunos com necessidades especiais	Nº de alunos com necessidades especiais
	8 - Número de alunos inscritos por ano curricular	Nº de alunos por ano curricular
	9 - Número de alunos que cancelaram a inscrição	Nº de alunos que cancelaram a inscrição
	10 - Número de alunos por motivo de cancelamento da inscrição	Nº de alunos por motivo de cancelamento
	11 - Taxa de abandono por ano curricular	Nº total de estudantes que cancelaram a inscrição no ano curricular / Nº total de estudantes do ano curricular
	12 - Número de alunos que suspenderam a inscrição	((Nº total de estudantes no ano N-1 – nº de Diplomados no ano N-1 - (Nº total de estudantes no ano – Nº de reingressos)) -nº estudantes que anularam
	13 - Número de alunos por motivo de suspensão da inscrição	Nº de alunos inscritos por motivo de inscrição
	14 - % de alunos com estatuto trabalhador-estudante	Nº de estudantes com ETE / Nº total de estudantes

Tabela 6 - Lista dos KPIs identificados

Objetivo	KPI	Métrica de calculo
B - Melhorar o desempenho do aluno	1 - Número médio de matrículas por ano curricular	Somatório do nº de inscrições no ano n/Total de inscrições no ano n
	2 - Número médio de inscrições até ao final do curso	Somatório do nº de inscrições de cada estudante diplomado no curso no ano N/Total de estudantes graduados no ano n
	3 - % de aprovações por UC (relativa ao total de alunos inscritos)	Nº de alunos com nota $\geq 9,5$ /Nº de alunos inscritos na UC
	4 - % de aprovações por UC (relativa ao total de alunos que completaram a avaliação)	Nº de alunos com nota $\geq 9,5$ /Nº de alunos que se submeteram a avaliação
	5 - % de alunos inscritos no primeiro ano, pela primeira vez, com aprovação em pelo menos 75% dos ECTS em que se inscreveram	Nº de estudantes para os quais o somatório do nº de ECTS das UCs em que teve aprovação/60 é $\geq 0,75$
	6 - Nota média por UC	Somatório das notas dos alunos na UC / Total de estudantes submetidos a avaliação
	7 - % de alunos por intervalo de nota, por UC	Nº de alunos submetidos a avaliação na UC/Total de alunos submetidos a avaliação na UC agrupados por intervalo de nota
C - Aumentar a eficiência do ensino	1 - % de graduados em N anos	Nº de diplomados em N anos / Nº de diplomados no total
	2 - % de graduados em N+1 anos	Nº de diplomados em N+1 anos / Nº de diplomados no total
	3 - % de graduados em N+2 anos	Nº de diplomados em N+2 anos / Nº de diplomados no total
	4 - % de graduados em >N+2 anos	Nº de diplomados em >N+2 anos / Nº de diplomados no total
	5 - Total de graduados	Somatório de alunos aprovados em todas as UCs do curso
	6 - Nota média de conclusão do curso	Somatório da nota de conclusão de todos os alunos do curso / Nº de alunos do curso
	7 - % de graduados em cada gama da escala Europeia de ECTS	Ver DL 42/2005

Tabela 7 - Lista dos KPIs identificados (continuação)

Todos os KPIs apresentados na Tabela 6 e Tabela 7 possuem uma periodicidade anual de cálculo. Devido à falta de dados facultados necessários para o cálculo, não foi possível calcular os seguintes KPIs: grupo A (7, 9, 10, 11, 12, 13, 14), grupo B (4, 5), e grupo C (7).

3.4 Desenho do DW

No desenho do DW foram consideradas as correlações e KPIs identificadas na Secção 3.3, pois estas têm implicações nos dados que têm de ser armazenados no DW. Foi considerado um modelo multidimensional para o DW baseado no esquema em estrela. Para este esquema de modelação dos dados são necessários dois tipos de tabelas: a tabela de factos que inclui os factos a serem analisados e as tabelas dimensão que são as tabelas pelas quais os factos irão ser analisados.

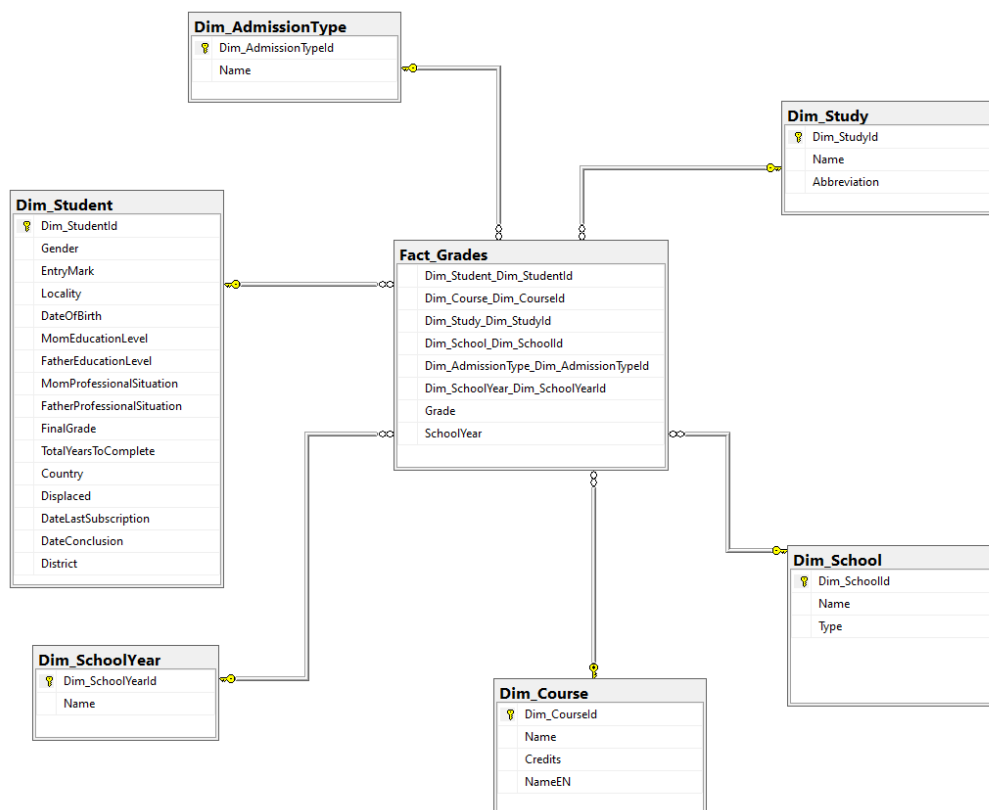


Figura 9 - Esquema em estrela do DW do sistema

O desenho final do DW é apresentado na Figura 9. Assim, a tabela de factos contém as notas dos alunos, que constituem os aspetos quantitativos da análise, enquanto as dimensões constituem os aspetos qualitativos da análise, isto é, os dados

sociodemográficos dos alunos, tipo de regime de ingresso, ano letivo, curso e UC em que o aluno está inscrito e escola secundária que o aluno frequentou antes do ingresso no ES.

3.4.1 Dimensões

O DW apresentado na Figura 9 é composto por seis tabelas dimensão que serão descritas em mais detalhe nesta secção.

Dim_AdmissionType (Tipo de Regime de Ingresso)

Dimensão que contém a informação referente aos tipos de regime de ingresso (cf. Tabela 8).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_AdmissionTypeId	Bigint	X		ID da dimensão	
Name	Nvarchar			Tipo de ingresso	Concurso institucional Alunos externos Concurso Especial Maiores de 23 Mudança de curso Transferência Etc.

Tabela 8 - Tabela Dim_AdmissionType do DW

Dim_SchoolYear (Ano letivo)

Dimensão que contém a informação referente aos anos letivos (cf. Tabela 9).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_SchoolYearId	Bigint	X		ID da dimensão	
Name	Nvarchar			Ano letivo	2011/2012 2012/2013 2013/2014 Etc.

Tabela 9 - Tabela Dim_SchoolYear do DW

Dim_Course (Unidade Curricular)

Dimensão que contém a informação referente às UC (cf. Tabela 10).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_CourseId	Bigint	X		ID da dimensão	
Name	Nvarchar			Designação da UC em PT	Álgebra Linear Algoritmia e Programação Alterabilidade de Materiais Análise de Relatórios e Contas Etc.
Credits	Int			Número de créditos da UC	[0,100]
NameEN	Nvarchar			Designação da UC em EN	

Tabela 10 - Tabela Dim_Course do DW

Dim_School (Escola Secundária)

Dimensão que contém a informação referente às escolas secundárias (cf. Tabela 11).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_SchoolId	Bigint	X		ID da dimensão	
Name	Nvarchar			Nome da escola secundária	Colégio D. Duarte Colégio de Gaia E.S. Maia Etc.
Type	Int			Flag que indica se a escola é privada ou não	0 – Pública 1 - Privada

Tabela 11 - Tabela Dim_School do DW

Dim_Student (Aluno)

Dimensão que contém a informação referente aos alunos (cf. Tabela 12).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_StudentId	Bigint	X		ID da dimensão	
Gender	Nvarchar			Género do aluno	M F
EntryMark	Int			Nota de ingresso do aluno	[95-200]
Locality	Nvarchar			Localidade de origem do aluno	Portela Póvoa de Varzim Valbom Varziela Etc.
DateOfBirth	Date			Data de nascimento do aluno	AAAA-MM-DD
MomEducationLevel	Nvarchar			Habilitações literárias da mãe	Não sabe ler nem escrever Ensino Superior – Licenciatura Ensino Superior – Bacharelato Etc.
FatherEducationLevel	Nvarchar			Habilitações literárias do pai	Não sabe ler nem escrever Ensino Superior – Licenciatura Ensino Superior – Bacharelato Etc.
MomProfessionalSituation	Nvarchar			Situação profissional da mãe	Doméstica/o Reformado/a Desempregado/a Trabalha por conta de outrem Etc.
FatherProfessionalSituation	Nvarchar			Situação profissional do pai	Doméstica/o Reformado/a Desempregado/a Trabalha por conta de outrem Etc.
FinalGrade	Int			Nota final de conclusão do curso do aluno	[10-200]

Campo	Tipo	CP	CE	Descrição	Gama de valores
TotalYearToComplete	int			Total de anos para conclusão do curso	[3-indefinido]
Country	nvarchar			País de naturalidade	Portugal Espanha Guernesey Etc.
Displaced	int			Flag que indica se aluno é deslocado da residência	0 - Não deslocado 1 - Deslocado
DateLastSubscription	date			Data da última inscrição	AAAA-MM-DD
DateConclusion	date			Data de conclusão do curso	AAAA-MM-DD
District	nvarchar			Distrito	Porto Braga Vila Real Aveiro Etc.

Tabela 12 - Tabela Dim_Student do DW

Dim_Study (Curso)

Dimensão que contém a informação referente aos cursos (cf. Tabela 13).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_StudyId	Bigint	X		ID da dimensão	
Name	Nvarchar			Nome do curso	Finanças Gestão Tecnologias e Sistemas de Informação Etc.
Abbreviation	Int			Abreviatura do nome do curso	

Tabela 13 - Tabela Dim_Study do DW

3.4.2 Factos

O DW apresentado na Figura 9 contém uma tabela de factos que será descrita em mais detalhe nesta secção.

Fact_Grades (Notas)

Esta tabela de factos contém um conjunto de chaves estrangeiras que permite a interligação às diferentes tabelas dimensão; contém ainda um campo com os factos a analisar (cf. Tabela 14).

Campo	Tipo	CP	CE	Descrição	Gama de valores
Dim_Student_Dim_StudentId	Bigint	X	X	Id da dimensão Dim_Student	
Dim_Course_Dim_CourseId	Bigint	X	X	Id da dimensão Dim_Course	
Dim_Study_Dim_StudyId	Bigint	X	X	Id da dimensão Dim_Study	
Dim_School_Dim_SchoolId	Bigint	X	X	Id da dimensão Dim_School	
Dim_AdmissionType_ Dim_AdmissionTypeId	Bigint	X	X	Id da dimensão Dim_AdmissionType	
Grade	int			Nota final da disciplina	[95-200]

Tabela 14 - Tabela Fact_Grades do DW

3.5 Processo ETL

Os dados foram exportados do sistema de informação da Universidade Portucalense (SIUPT) para o formato Excel. Com base nos registos originais, foi gerado aleatoriamente um conjunto de dados fictícios para os campos relativos aos dados sociodemográficos dos alunos e de desempenho académico; para além disso, os registos foram anonimizados, tendo sido retirados os campos que permitissem a identificação individual dos alunos, assim como a identificação dos respetivos pais.

Foi efetuado um pré-processamento de correção dos dados, uma vez que os campos vazios (“null”) provenientes da base de dados do sistema operacional foram colocados no formato Excel como *string*, provocando um erro. Seguidamente, o Excel

foi importado a partir de uma base de dados SQL, tendo sido utilizados um conjunto de scripts para limpeza, transformação e padronização de formatos.

A título de exemplo, o script apresentado na Figura 10 efetua uma limpeza/uniformização dos dados, percorrendo todos os registos presentes na tabela candidatura (CA_SITP_ID_PAI, CA_SITP_ID_MAE, CA_ESC_ID_PAI, CA_ESC_ID_MAE, CA_PA_ID_NATURAL) e inscrição (I_CLASSIFICACAO_FINAL) da base de dados “upt” corrigindo os valores NULL considerados como *string* e colocando-os corretamente como NULL.

```
UPDATE upt.dbo.candidatura SET CA_SITP_ID_PAI=null WHERE CA_SITP_ID_PAI='NULL';
UPDATE upt.dbo.candidatura SET CA_SITP_ID_MAE=null WHERE CA_SITP_ID_MAE='NULL';
UPDATE upt.dbo.candidatura SET CA_ESC_ID_PAI=null WHERE CA_ESC_ID_PAI='NULL';
UPDATE upt.dbo.candidatura SET CA_ESC_ID_MAE=null WHERE CA_ESC_ID_MAE='NULL';
UPDATE upt.dbo.candidatura SET CA_PA_ID_NATURAL=null WHERE
CA_PA_ID_NATURAL='NULL';
UPDATE upt.dbo.inscricao SET I_CLASSIFICACAO_FINAL=null WHERE
I_CLASSIFICACAO_FINAL='NULL'
UPDATE upt.dbo.inscricao SET I_DATA_CONCLUSAO=null WHERE I_DATA_CONCLUSAO='NULL'
UPDATE upt.dbo.inscricao SET I_ANO=null WHERE I_ANO='NULL'
```

Figura 10 - Limpeza/Uniformização de dados

Terminada a fase de preparação dos dados, foram utilizados um conjunto de scripts para migração da informação para o DW em SQL (Figura 11 a Figura 18).

Assim, o script apresentado na Figura 11 extrai os países existentes na tabela “pais” da base de dados “upt” e carrega-os para a tabela “Dim_Country” da base de dados “datawarehouse”.

```
INSERT INTO Dim_Country(Dim_CountryId, Name, Abbreviation, Description,
Nationality)
SELECT PA_ID, PA_NOME, PA_ABREVIATURA, PA_DESC, PA_NACIONALIDADE FROM
upt.dbo.pais
```

Figura 11 - Carregamento de dados para a tabela Dim_Country

O script apresentado na Figura 12 extrai os tipos de admissão existentes na tabela “tipo_ingresso” da base de dados “upt” e carrega-os para a tabela “Dim_AdmissionType” da base de dados “datawarehouse”.

```
INSERT INTO datawarehouse.dbo.Dim_AdmissionType(Dim_AdmissionTypeId, Name)
SELECT TI_ID, TI_NOME FROM upt.dbo.tipo_ingresso
```

Figura 12 - Carregamento de dados para a tabela Dim_AdmissionType

O script apresentado na Figura 13 extrai as UC existentes na tabela “disciplina” da base de dados “upt” e carrega-as para a tabela “Dim_Course” da base de dados “datawarehouse”.

```
INSERT INTO datawarehouse.dbo.Dim_Course(Dim_CourseId, Name, Credits, NameEN)
SELECT D_ID, D_NOME, D_ECTS, D_NOME_EN FROM upt.dbo.disciplina
```

Figura 13 - Carregamento de dados para a tabela Dim_Course

O script apresentado na Figura 14 extrai as escolas existentes na tabela “escola” da base de dados “upt” e percorre todas as candidaturas para extrair também as “outras escolas” e carrega-as para a tabela “Dim_School” da base de dados “datawarehouse”.

```
SET IDENTITY_INSERT datawarehouse.dbo.Dim_School ON
INSERT INTO datawarehouse.dbo.Dim_School(Dim_SchoolId, Name, Type)
SELECT ES_ID, ES_NOME, ES_PRIVADA FROM upt.dbo.escola
SET IDENTITY_INSERT datawarehouse.dbo.Dim_School OFF

INSERT INTO datawarehouse.dbo.Dim_School(Name)
SELECT CAH_OUTRA_ESCOLA FROM upt.dbo.candidatura_habilitacao
where CAH_OUTRA_ESCOLA <> 'null'
and CAH_OUTRA_ESCOLA NOT IN (SELECT ES_NOME FROM upt.dbo.escola) GROUP BY
CAH_OUTRA_ESCOLA
```

Figura 14 - Carregamento de dados para a tabela Dim_School

O script apresentado na Figura 15 extrai os cursos existentes na tabela “curso” da base de dados “upt” e carrega-os para a tabela “Dim_Study” da base de dados “datawarehouse”.

```
INSERT INTO datawarehouse.dbo.Dim_Study(Dim_StudyId, Name, Abbreviation)
SELECT C_ID, C_NOME, C_ABREVIATURA FROM upt.dbo.curso
```

Figura 15 - Carregamento de dados para a tabela Dim_Study

O script apresentado na Figura 16 percorre as tabelas “inscrição”, “inscricao_disciplina”, “uc”, “curso”, “tipo_curso”, “disciplina”, “candidatura”, “candidatura_habilitacao” e “edicao” da base de dados “upt” para obter as inscrições nas UC e cursos e respetivas classificações e carrega-as para a tabela “Fact_Grades” da base de dados “datawarehouse”.

```

INSERT INTO datawarehouse.dbo.Fact_Grades(
Dim_Student_Dim_StudentId,
Dim_Course_Dim_CourseId,
Dim_Study_Dim_StudyId,
Dim_School_Dim_SchoolId,
Dim_AdmissionType_Dim_AdmissionTypeId,
Dim_SchoolYear_Dim_SchoolYearId,
Grade,
SchoolYear
)
OUTPUT inserted.Dim_Student_Dim_StudentId
SELECT
I_A_ID,
D_ID,
C_ID,
case when CAH_ES_ID is null then (select TOP(1) Dim_SchoolId FROM
datawarehouse.dbo.Dim_School WHERE Name=CAH_OUTRA_ESCOLA) else CAH_ES_ID end,
CA_TI_ID,
(SELECT Dim_SchoolYear.Dim_SchoolYearId FROM Dim_SchoolYear WHERE
Dim_SchoolYear.Name=E_ANOLECTI),
ID_NOTA_FINAL,
I_ANO
FROM upt.dbo.inscricao
INNER JOIN upt.dbo.inscricao_disciplina ON I_ID=ID_I_ID
INNER JOIN upt.dbo.uc ON ID_UC_ID=UC_ID
INNER JOIN upt.dbo.curso ON UC_C_ID=C_ID
INNER JOIN upt.dbo.tipo_curso ON C_TC_ID=TC_ID
INNER JOIN upt.dbo.disciplina ON UC_D_ID=D_ID
INNER JOIN upt.dbo.candidatura ON I_CA_ID=CA_ID
INNER JOIN upt.dbo.candidatura_habilitacao ON CA_ID=CAH_CA_ID
INNER JOIN upt.dbo.edicao ON I_E_ID=E_ID

```

Figura 16 - Carregamento de dados para a tabela Fact_Grades

O script apresentado na Figura 17 percorre as tabelas “inscrição”, “candidatura”, “candidatura_habilitacao”, “pais”, “escolaridade”, “situação_profissional”, “edição”, “curso” e “tipo_curso” da base de dados “upt” para extrair a lista de alunos e carrega-os para a tabela “Dim_Student” da base de dados “datawarehouse”. Visto que não há maneira de identificar os alunos devido às regras de RGPD a maneira encontrada foi utilizar o id do aluno presente no campo “I_A_ID” na tabela “inscrição” na base de dados “upt”.

```

INSERT INTO datawarehouse.dbo.Dim_Student(Dim_StudentId,
Gender,
EntryMark,
Locality,
DateOfBirth,
MomEducationLevel,
FatherEducationLevel,
MomProfessionalSituation,
FatherProfessionalSituation,
FinalGrade,
DateConclusion,
DateLastSubscription,
TotalYearsToComplete,
Country,
Displaced,
District)
SELECT
I_A_ID,
CA_SEXO,
ISNULL(CAST(CAH_CLASSIFICACAO AS nvarchar(50)), CAST(CA_M23_CLASSIFICACAO AS
nvarchar(50))) AS 'CAH_CLASSIFICACAO',
LOCALIDADE,
CA_DATA_NASC,
escolaridadeMae.ESC_NOME as 'escolaridade_mae',
escolaridadePai.ESC_NOME as 'escolaridade_pai',
situacaoMae.SITP_NOME as 'situacao_mae',
situacaoPai.SITP_NOME as 'situacao_pai',
I_CLASSIFICACAO_FINAL,
(SELECT convert(date, (SELECT TOP 1 value FROM STRING_SPLIT(I_DATA_CONCLUSAO, '
')),5)) as 'dataConclusao',
I_DATA_CRIADA,
I_INSCRICOES_MESMO_CURSO,
PA_NOME,
I_DESLOCADO,
(SELECT top 1 nome_distrito FROM [upt].[dbo].codigos_postais INNER JOIN
[upt].[dbo].distrito ON
[upt].[dbo].codigos_postais.cod_distrito=[upt].[dbo].distrito.cod_distrito WHERE
nome_localidade like [LOCALIDADE])
FROM [upt].[dbo].[inscricao]
INNER JOIN upt.dbo.candidatura ON CA_ID=I_CA_ID
INNER JOIN upt.dbo.candidatura_habilitacao ON CA_ID=CAH_CA_ID
LEFT JOIN upt.dbo.pais ON PA_ID=CA_PA_ID_NATURAL
LEFT JOIN upt.dbo.escolaridade as escolaridadeMae ON
CA_ESC_ID_MAE=escolaridadeMae.ESC_ID
LEFT JOIN upt.dbo.escolaridade as escolaridadePai ON
CA_ESC_ID_PAI=escolaridadePai.ESC_ID
LEFT JOIN upt.dbo.situacao_profissional as situacaoMae ON
CA_SITP_ID_MAE=situacaoMae.SITP_ID
LEFT JOIN upt.dbo.situacao_profissional as situacaoPai ON
CA_SITP_ID_PAI=situacaoPai.SITP_ID
INNER JOIN upt.dbo.edicao ON I_E_ID=E_ID
INNER JOIN upt.dbo.curso ON E_C_ID=C_ID
INNER JOIN upt.dbo.tipo_curso ON C_TC_ID=TC_ID,
(SELECT max(I_ID) as idUtilizador,max(I_DATA_CRIADA) as transaction_date FROM
[upt].[dbo].[inscricao] INNER JOIN upt.dbo.candidatura ON CA_ID=I_CA_ID group by
CA_A_ID) max_user
where [upt].[dbo].[inscricao].I_ID=max_user.idUtilizador and
I_DATA_CRIADA=max_user.transaction_date
order by I_A_ID

```

Figura 17 - Carregamento de dados para a tabela Dim_Student

O script apresentado na Figura 18 extrai os anos letivos existentes na tabela “edicao” da base de dados “upt” e carrega-os para a tabela “Dim_SchoolYear” da base de dados “datawarehouse”.

```
INSERT INTO datawarehouse.dbo.Dim_SchoolYear(Name)
SELECT distinct E_ANOLECTI from upt.dbo.edicao
```

Figura 18 - Carregamento de dados para a tabela Dim_SchoolYear

O DW em SQL Server resultante do processo ETL foi depois exportado para o projeto de implementação em Qlik Sense, correspondendo às representações de modelos de dados ilustradas nas Figura 19 e Figura 20.

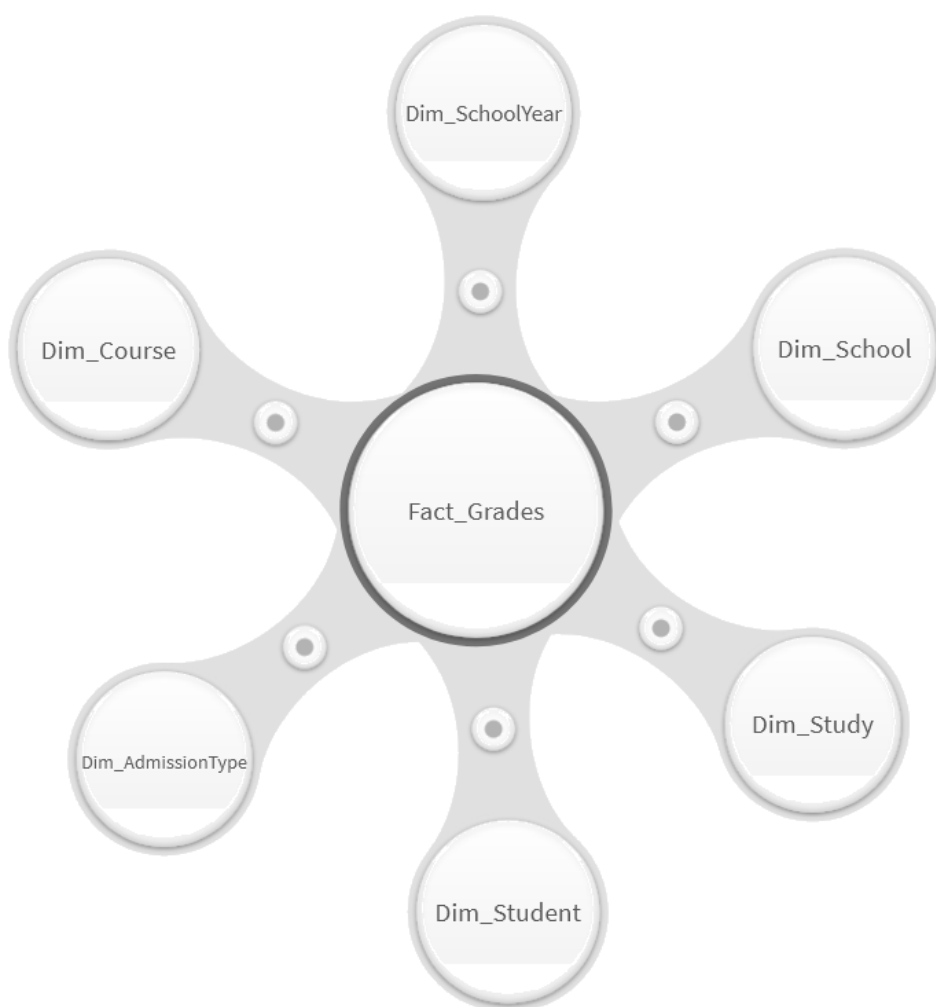


Figura 19 - Datawarehouse na implementação do projeto em Qlik Sense



Figura 20 - Modelo dados na implementação do projeto em Qlik Sense

3.6 Configuração da plataforma Qlik Sense

Antes de proceder à criação dos gráficos apresentados na Secção 3.8 (Figura 41, Figura 42, Figura 43, Figura 44, Figura 45, Figura 46, Figura 57, Figura 59), foi necessário criar algumas dimensões específicas (dimensões calculadas) no Qlik Sense. A criação destas dimensões foi feita através do editor de expressões, requerendo alguns conhecimentos específicos de lógica de programação (cf. Figura 21).

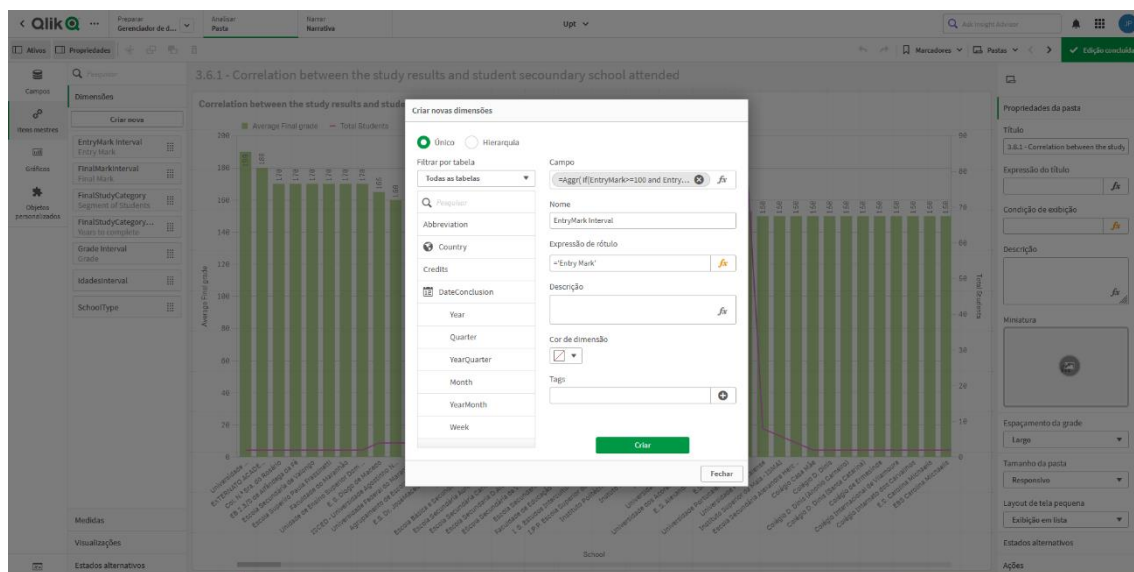


Figura 21 - Criação de nova dimensão

SchoolType

A Figura 22 apresenta a sequência de comandos utilizada para criar a dimensão SchoolType, necessária para implementação da conversão num formato textual da *flag* booleana “Type” da dimensão “School”.

```
=Aggr(if(Type=1, 'Private',  
        if(Type=0, 'Public', 'Outros')), Type)
```

Figura 22 - Dimensão para traduzir a flag Type da Dimensão School

EntryMarkInterval

A Figura 23 apresenta a sintaxe para a criação dos intervalos de notas de ingresso relativos ao campo (EntryMark), que armazena a nota de ingresso do aluno. Este campo foi dividido em 5 grupos de notas:

- [10-12[;
- [12-14[;
- [14-16[;
- [16-18[;
- [18-20].

```
=Aggr(if(EntryMark>=100 and EntryMark<120, '[10-12[',  
        if(EntryMark>=120 and EntryMark<140, '[12-14[',  
        if(EntryMark>=140 and EntryMark<160, '[14-16[',  
        if(EntryMark>=160 and EntryMark<180, '[16-18[',  
        if(EntryMark>=180 and EntryMark<=200, '[18-20]')  
))))), EntryMark)
```

Figura 23 - Dimensão para criar intervalos para a nota ingresso

FinalGradeInterval

A Figura 24 apresenta a sintaxe para a criação de intervalos de notas finais relativos ao campo (FinalGrade), que armazena a nota final de curso do aluno. Este campo foi dividido em 5 grupos de notas:

- [10-12[;
- [12-14[;
- [14-16[;
- [16-18[;
- [18-20].

```
=Aggr(if(FinalGrade>=100 and FinalGrade<120, '[10-12[',
      if(FinalGrade>=120 and FinalGrade<140, '[12-14[',
        if(FinalGrade>=140 and FinalGrade<160, '[14-16[',
          if(FinalGrade>=160 and FinalGrade<180, '[16-18[',
            if(FinalGrade>=180 and FinalGrade<=200, '[18-20]')
          )))
      ),FinalGrade)
```

Figura 24 - Dimensão para criar intervalos para a nota final de curso

Grade

A Figura 25 apresenta a sintaxe para a criação de intervalos de notas relativos ao campo (Grade), que armazena a nota final de uma UC. Este campo foi dividido em 6 grupos de notas:

- [0-10[;
- [10-12[;
- [12-14[;
- [14-16[;
- [16-18[;
- [18-20].

```
=Aggr(if(Grade>=0 and Grade <100, '[0-10[',
      if(Grade>=100 and Grade<120, '[10-12[',
        if(Grade>=120 and Grade<140, '[12-14[',
          if(Grade>=140 and Grade<160, '[14-16[',
            if(Grade>=160 and Grade<180, '[16-18[',
              if(Grade>=180 and Grade<=200, '[18-20]')
            ))))
      ),Grade)
```

Figura 25 - Dimensão para criar intervalos para nota de UC

FinalStudyCategoryYearAndGrade

A Figura 26 apresenta a sintaxe para a criação das categorias de desempenho acadêmico. Foram consideradas 5 categorias, de acordo com o total de anos necessário para completar o curso e a nota final do mesmo:

- Excellent;
- Good;
- Reasonable;
- Fair;
- Poor.

```
=if(TotalYearsToComplete=3 and FinalGrade>=160, 'Excellent',
  if(TotalYearsToComplete=3 and FinalGrade>=130, 'Good',
    if(TotalYearsToComplete=3 and FinalGrade<140, 'Reasonable',
      if(TotalYearsToComplete=4 and FinalGrade>=120, 'Fair',
        if(TotalYearsToComplete>=4 and FinalGrade>10,
          'Poor', 'Others'))))))
```

Figura 26 - Dimensão para criar categorias de desempenho acadêmico baseadas no total de anos e nota final

FinalStudyCategoryYear

A Figura 27 apresenta a sintaxe para a criação dos grupos relativos ao total de anos para completar o curso. Foram considerados 4 grupos:

- 3 Years;
- 4 Years;
- 5 Years;
- Bigger than 5 Years.

```
=if(TotalYearsToComplete=3, '3 Years',
  if(TotalYearsToComplete=4, '4 Years',
    if(TotalYearsToComplete=5, '5 Years',
      if(TotalYearsToComplete>5, 'Bigger than 5 Years', 'Others'))))
```

Figura 27 - Dimensão para criar categorias de desempenho acadêmico baseadas no total de anos

AgesInterval

A Figura 28 apresenta a sintaxe para a criação dos grupos de faixas etárias utilizados. Foram criados seis grupos:

- [18-20[;
- [20-22[;
- [22-24[;
- [24-26[;
- [26-28[;
- >28.

```

=if(Year(DateLastSubscription)-Year(DateOfBirth)>=18 and Year(DateLastSubscription)-
Year(DateOfBirth)<20, '[18,20[',
if(Year(DateLastSubscription)-Year(DateOfBirth)>=20 and Year(DateLastSubscription)-
Year(DateOfBirth)<22, '[20,22[',
  if(Year(DateLastSubscription)-Year(DateOfBirth)>=22 and Year(DateLastSubscription)-
Year(DateOfBirth)<24, '[22,24[',
    if(Year(DateLastSubscription)-Year(DateOfBirth)>=24 and Year(DateLastSubscription)-
Year(DateOfBirth)<26, '[24,26[',
      if(Year(DateLastSubscription)-Year(DateOfBirth)>=26 and
Year(DateLastSubscription)-Year(DateOfBirth)<28, '[26,28[',
        if(Year(DateLastSubscription)-Year(DateOfBirth)>=28, '>28', 'Others')))))))

```

Figura 28 - Dimensão para criar intervalos de idades

3.7 Criação dos gráficos na plataforma Qlik Sense

Depois das dimensões necessárias à criação dos gráficos estarem criadas, a criação dos gráficos é relativamente simples e o processo é maioritariamente *drag-and-drop*.

Seguidamente será ilustrada a criação de um dos gráficos elaborados neste trabalho e apresentado na Secção 3.8, Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso (cf. Figura 41).

Num primeiro passo, procede-se à escolha do tipo de gráfico a utilizar a partir da lista disponibilizada pelo Qlik Sense, arrastando-se o mesmo para a área de trabalho (cf. Figura 29). Neste exemplo é utilizado um gráfico de barras.

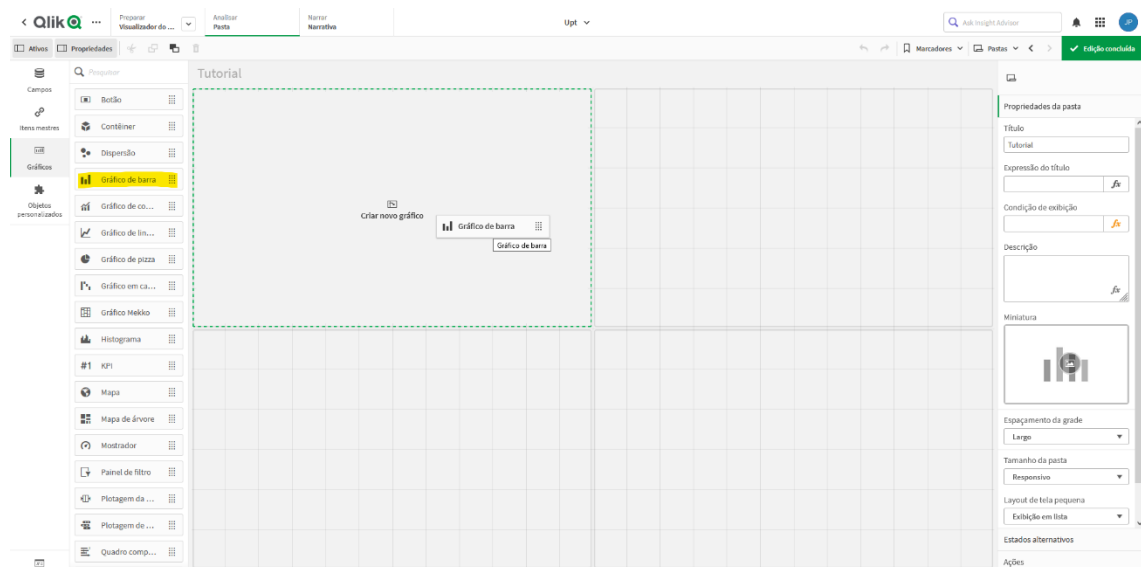


Figura 29 - Primeira etapa na criação de um gráfico

O próximo passo será definir que dimensões e medidas serão analisadas no gráfico. Neste exemplo será utilizada uma dimensão previamente criada “EntryMark Interval”, e como medidas serão definidas as seguintes quatro expressões:

Male Average Final Mark

=avg(if(Gender='M',FinalGrade))/10

Female Average Final Mark

=avg(if(Gender='F',FinalGrade))/10

% of Male Students

=count(if(Gender='M',[Dim_StudentId-Dim_Student_Dim_StudentId])) /
count(TOTAL [Dim_StudentId-Dim_Student_Dim_StudentId])

% of Female Students

=count(if(Gender='F',[Dim_StudentId-Dim_Student_Dim_StudentId])) /
count(TOTAL [Dim_StudentId-Dim_Student_Dim_StudentId])

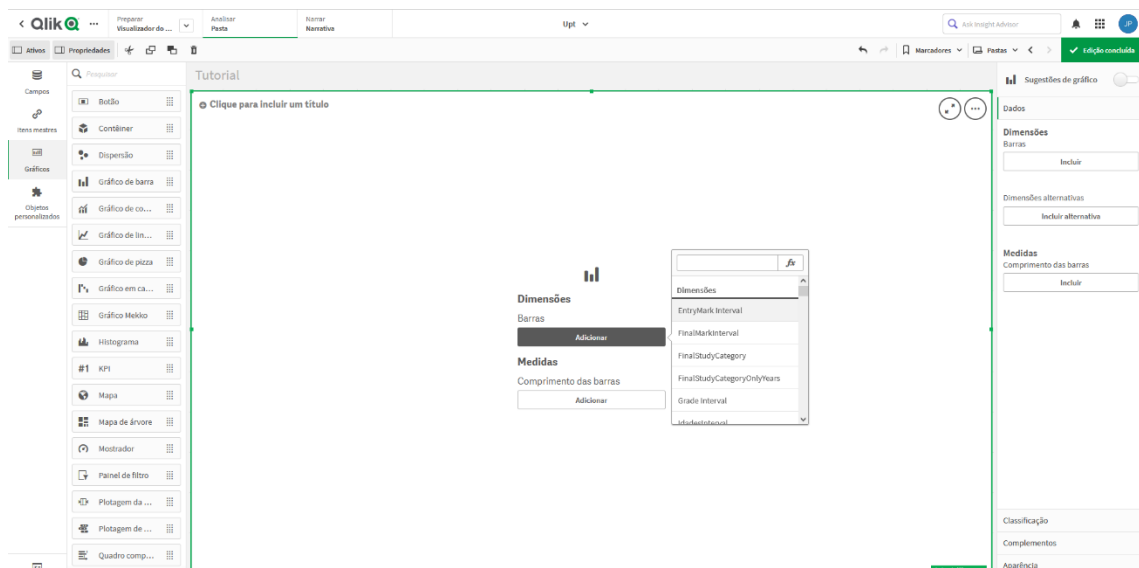


Figura 30 - Escolher dimensão previamente criada (EntryMark Interval)

Para adicionar a dimensão, será necessário escolher a dimensão previamente criada “EntryMark Interval” (cf. Figura 30).

Relativamente às medidas, é necessário criá-las uma a uma e inserir as respetivas expressões (cf. Figura 31).

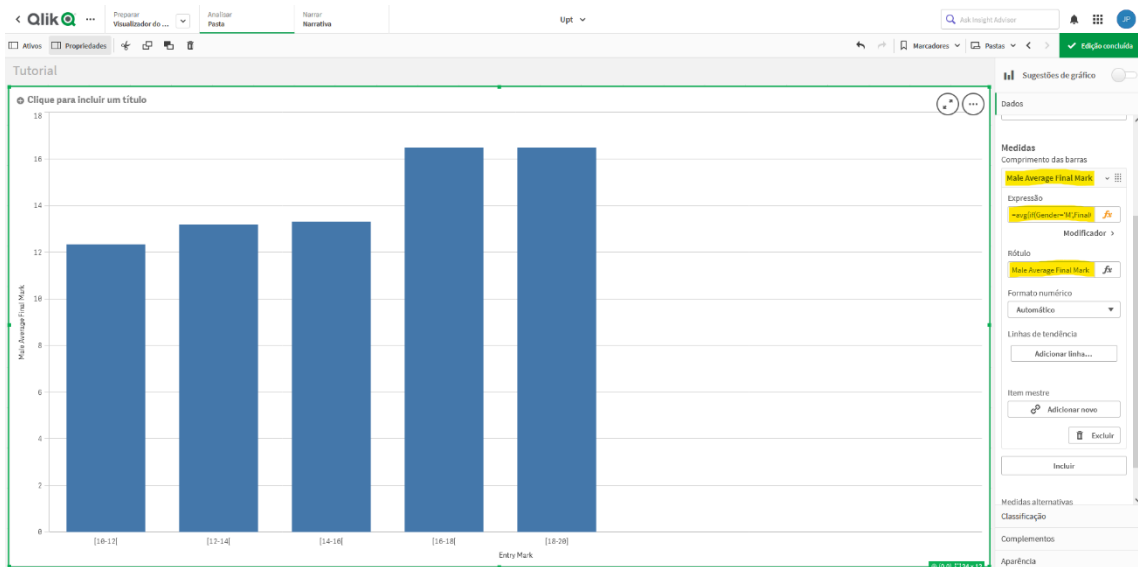


Figura 31 - Inserir expressões relativas às medidas

A Figura 32 apresenta o gráfico de barras verticais resultante.

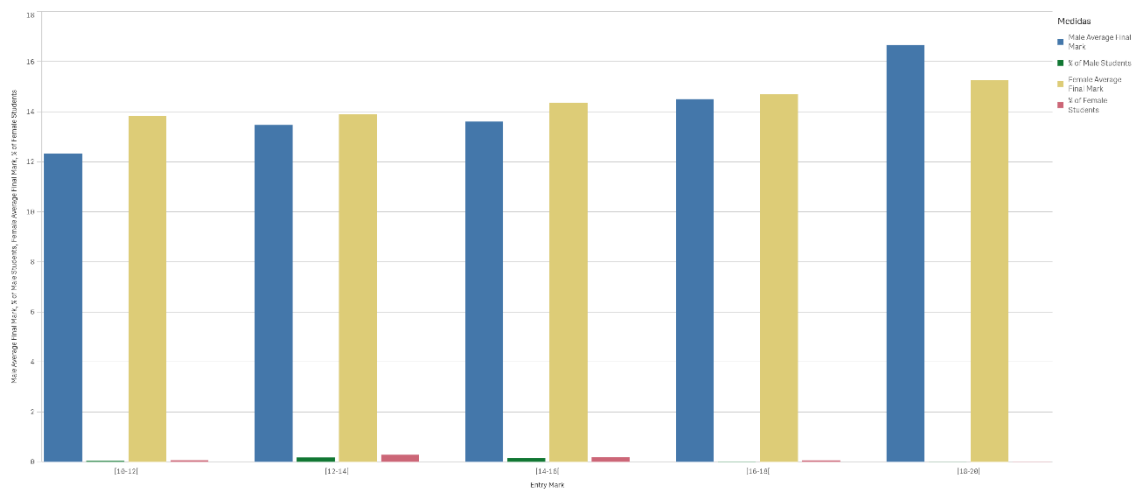


Figura 32 - Gráfico defeito

De forma a obter a aparência do gráfico correspondente à ilustrada na Figura 41, será necessário efetuar algumas alterações de layout, nomeadamente alteração da disposição das barras de vertical para horizontal e alteração do local de posicionamento da legenda. Estas alterações são realizadas na aba “Aparência -> Geral” para alterar a orientação das barras (cf. Figura 33) e na aba “Aparência -> Cores e Legenda” para alterar a localização da legenda (cf. Figura 34).

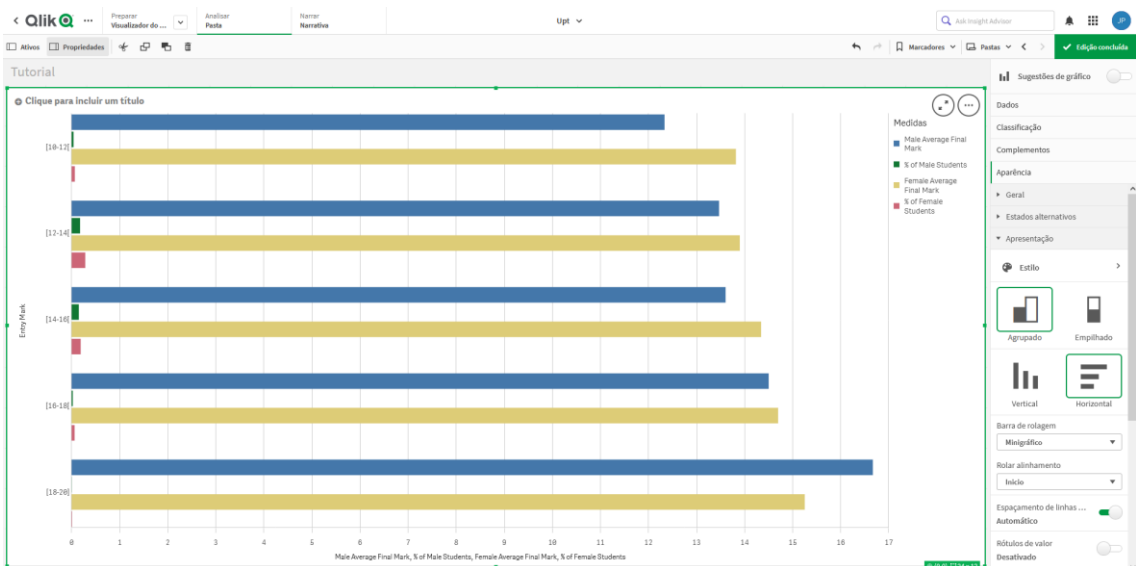


Figura 33 - Alterar orientação do gráfico

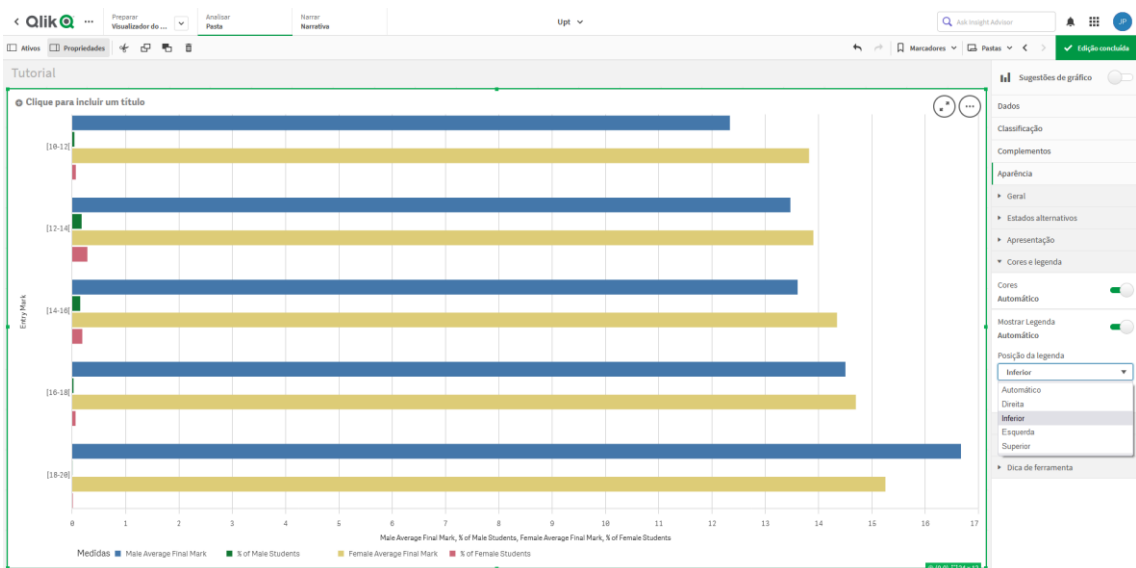


Figura 34 - Alterar localização da legenda do gráfico

3.8 Análise OLAP

Nesta secção apresentam-se os resultados da análise multidimensional efetuada, baseada no DW definido na Secção 3.4 e, em particular, das correlações previamente identificadas na Secção 3.3.

Por uma questão de melhoria de leitura, versões ampliadas de algumas figuras (Figura 35, Figura 36, Figura 37, Figura 38, Figura 39, Figura 40, Figura 41, Figura 45, Figura 48, Figura 58 e Figura 65) são apresentadas em anexo.

3.8.1 Correlação entre a nota final do curso e a escola frequentada no ensino secundário

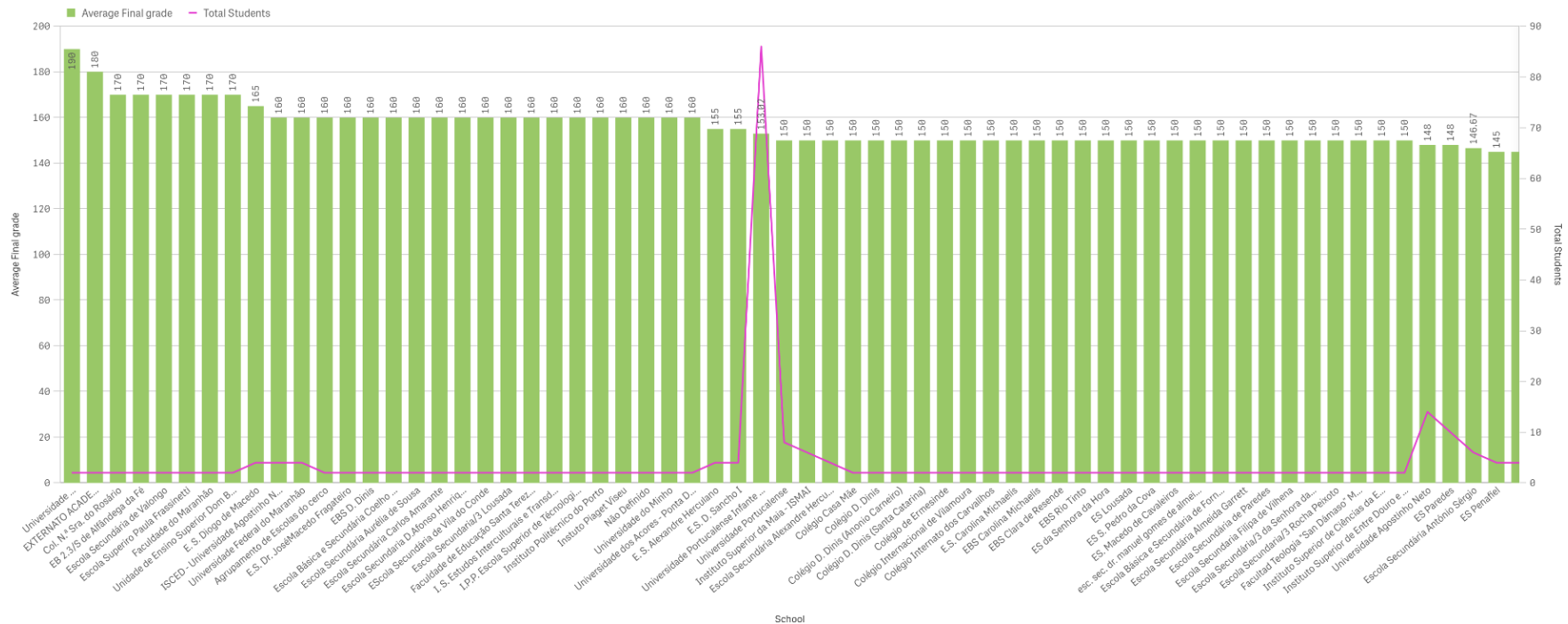


Figura 35 - Correlação entre a nota final do curso e a escola frequentada no ensino secundário

A Figura 35 apresenta uma representação parcial dos resultados obtidos, uma vez que existe um grande número de escolas envolvidas. Não foi possível identificar uma correlação significativa entre as duas variáveis: nota final de curso do aluno e escola de origem.

3.8.2 Segmentos de alunos por regime de ingresso e nota final de curso

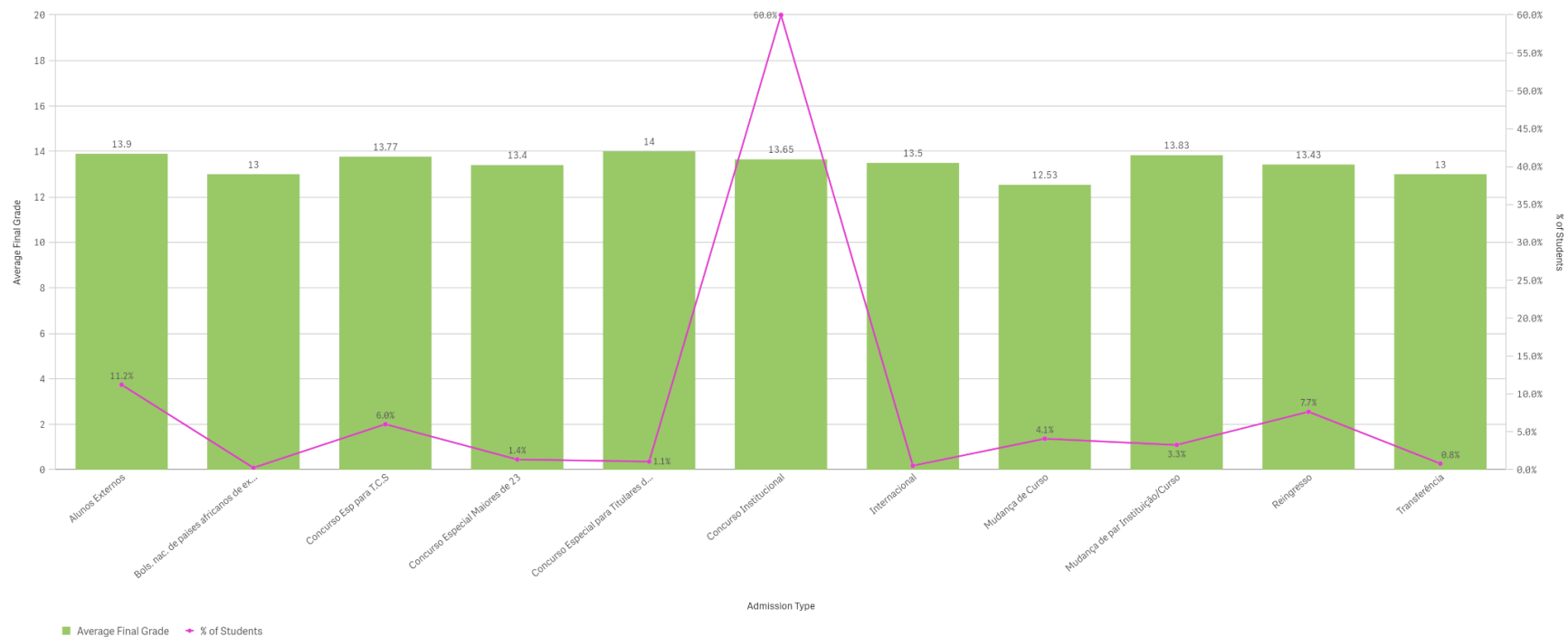


Figura 36 - Segmentos de alunos por regime de ingresso e nota final de curso

Pela análise da Figura 36 verifica-se que, embora se identifiquem vários regimes de ingresso, o mais comum é o regime “Concurso Institucional” com 60% de alunos. No entanto, o regime a que corresponde uma melhor nota final de curso (em termos médios) é o regime “Concurso Especial para Titulares de um Diploma de Especialização Tecnológica” com nota média de conclusão de curso de 14 valores.

3.8.3 Segmentos de alunos por situação profissional dos pais e correlação com a nota final de curso

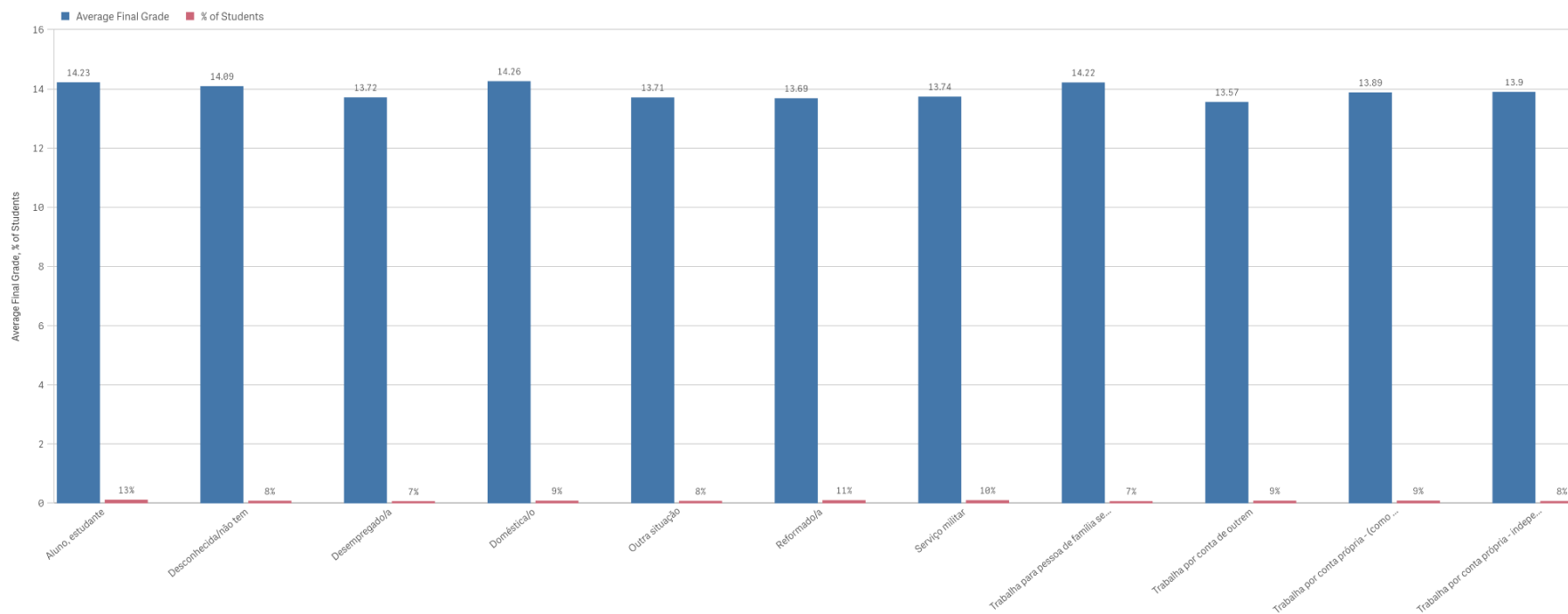


Figura 37 - Segmentos de alunos por situação profissional do pai e correlação com a nota final de curso

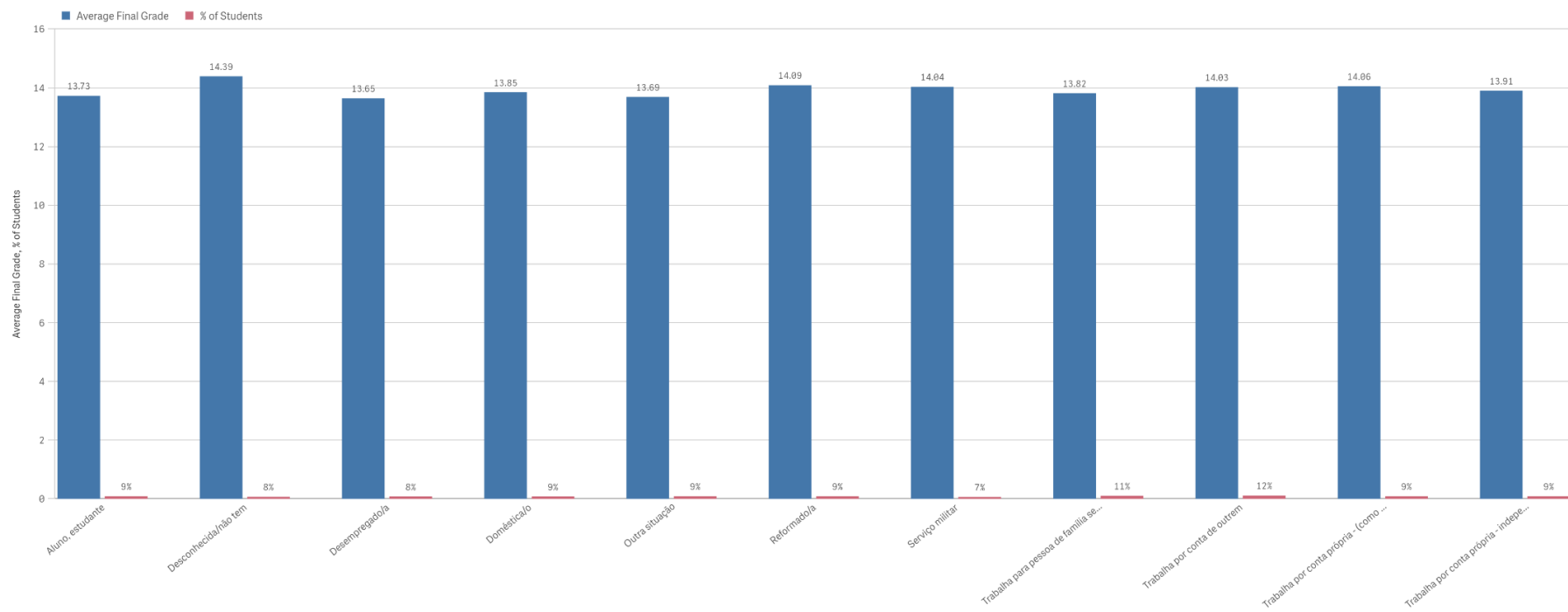


Figura 38 - Segmentos de alunos por situação profissional da mãe e correlação com a nota final de curso

Pela análise dos gráficos apresentados na Figura 37 e Figura 38, a situação profissional quer do pai, quer da mãe do aluno não tem impacto significativo na nota final (em termos médios) do curso do aluno.

3.8.4 Segmentos de alunos por nível de escolaridade dos pais e correlação com a nota final de curso

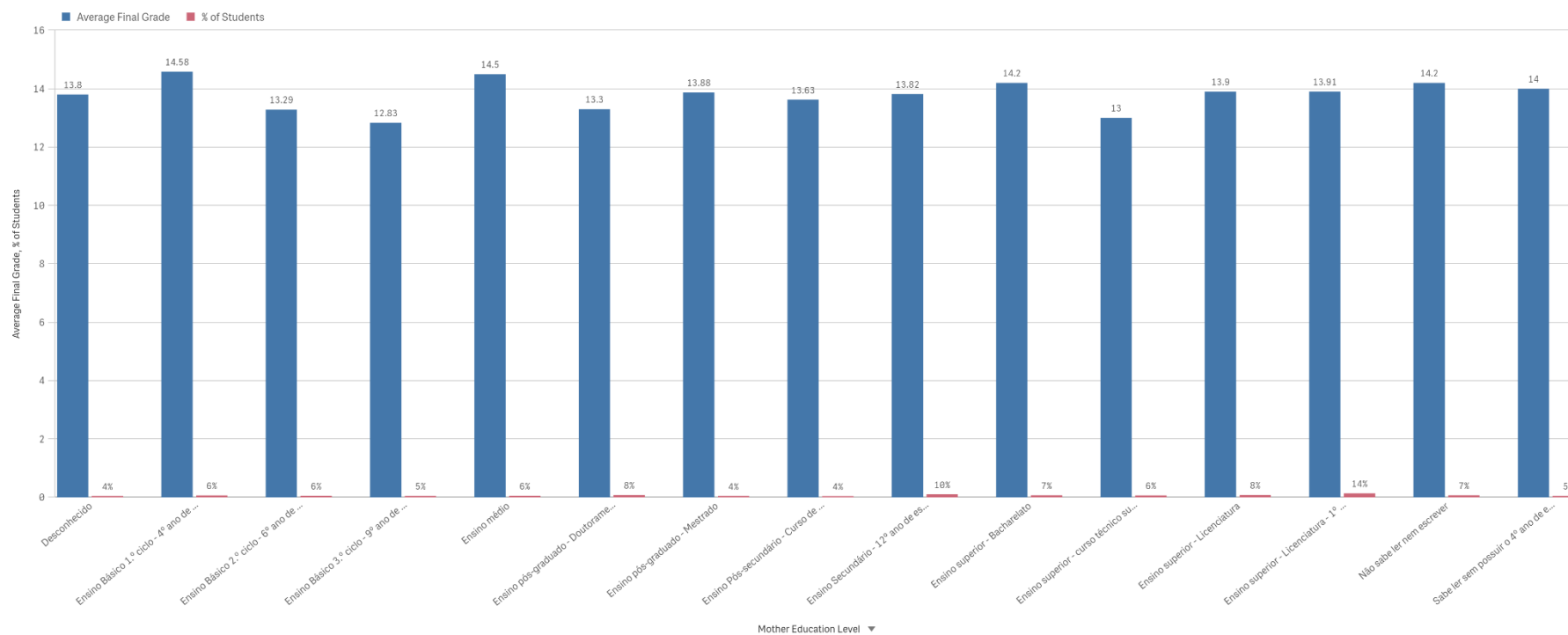


Figura 39 - Segmentos de alunos por nível de escolaridade da mãe e correlação com a nota final de curso

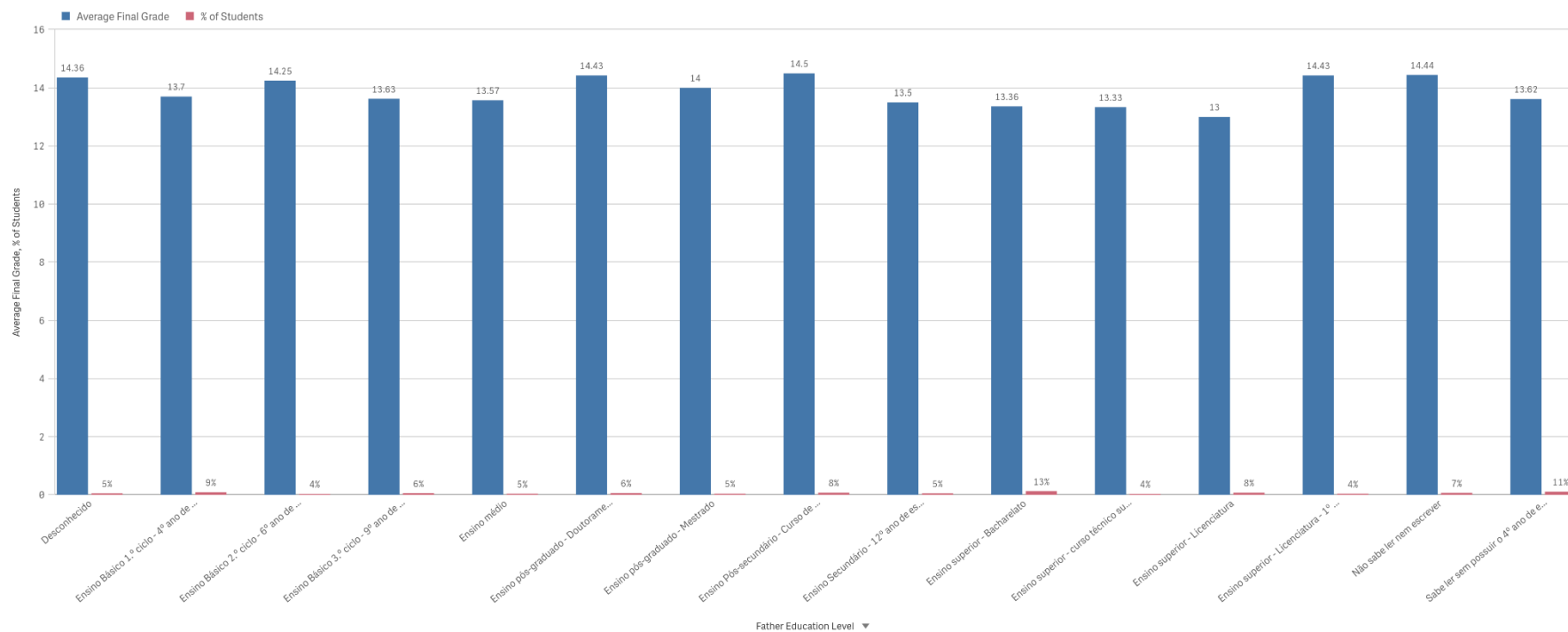


Figura 40 - Segmentos de alunos por nível de escolaridade do pai e correlação com a nota final de curso

Pela análise dos gráficos apresentados na Figura 39 e Figura 40, a escolaridade quer do pai, quer da mãe do aluno não tem impacto significativo na nota final (em termos médios) do curso do aluno.

3.8.5 Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso

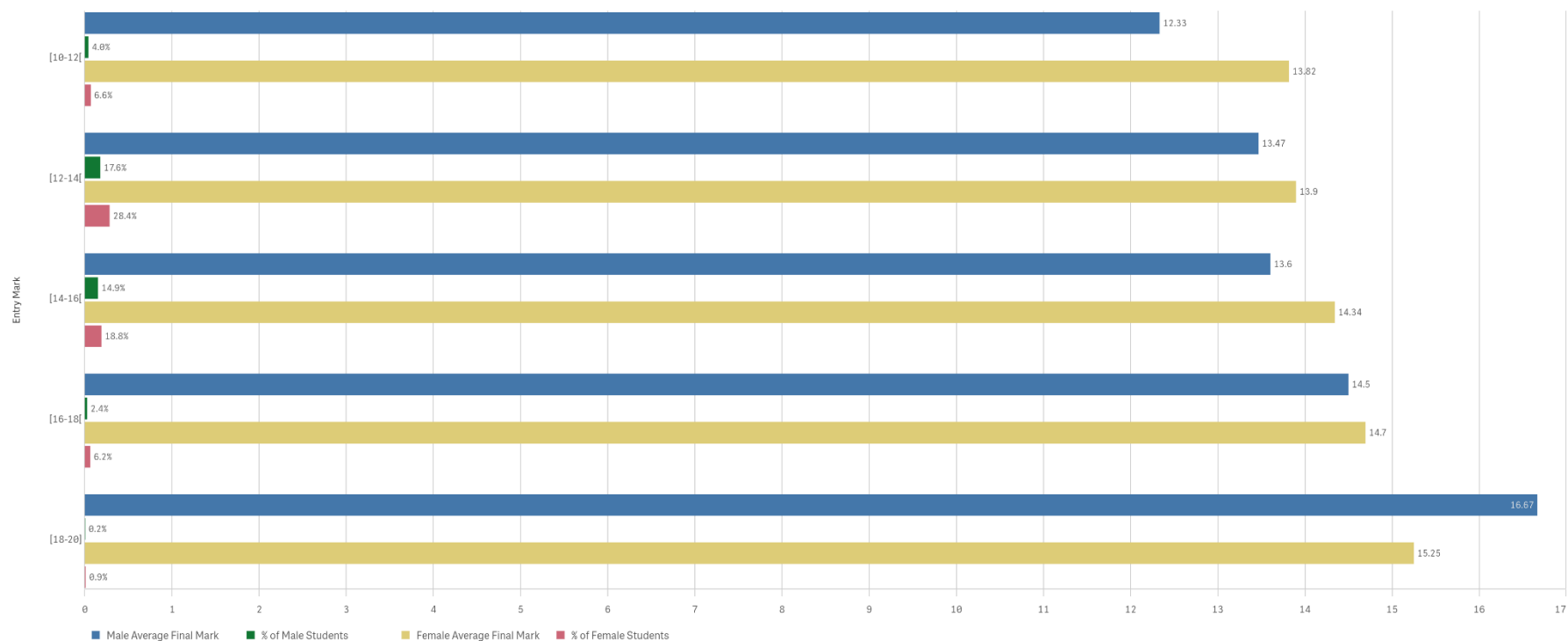


Figura 41 - Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso

Da análise da Figura 41 verifica-se que a percentagem mais significativa de alunos é a correspondente à gama de notas de ingresso no intervalo [12-14[com uma distribuição de 28,4% para os alunos do sexo feminino e 17,6% dos alunos do sexo masculino, seguida do segmento de alunos com notas de ingresso no intervalo [14-16[, com uma distribuição de 18,8% para os alunos do sexo feminino e 14,9% para os alunos do sexo masculino.

3.8.6 Segmentos de alunos por origem geográfica e nota de ingresso

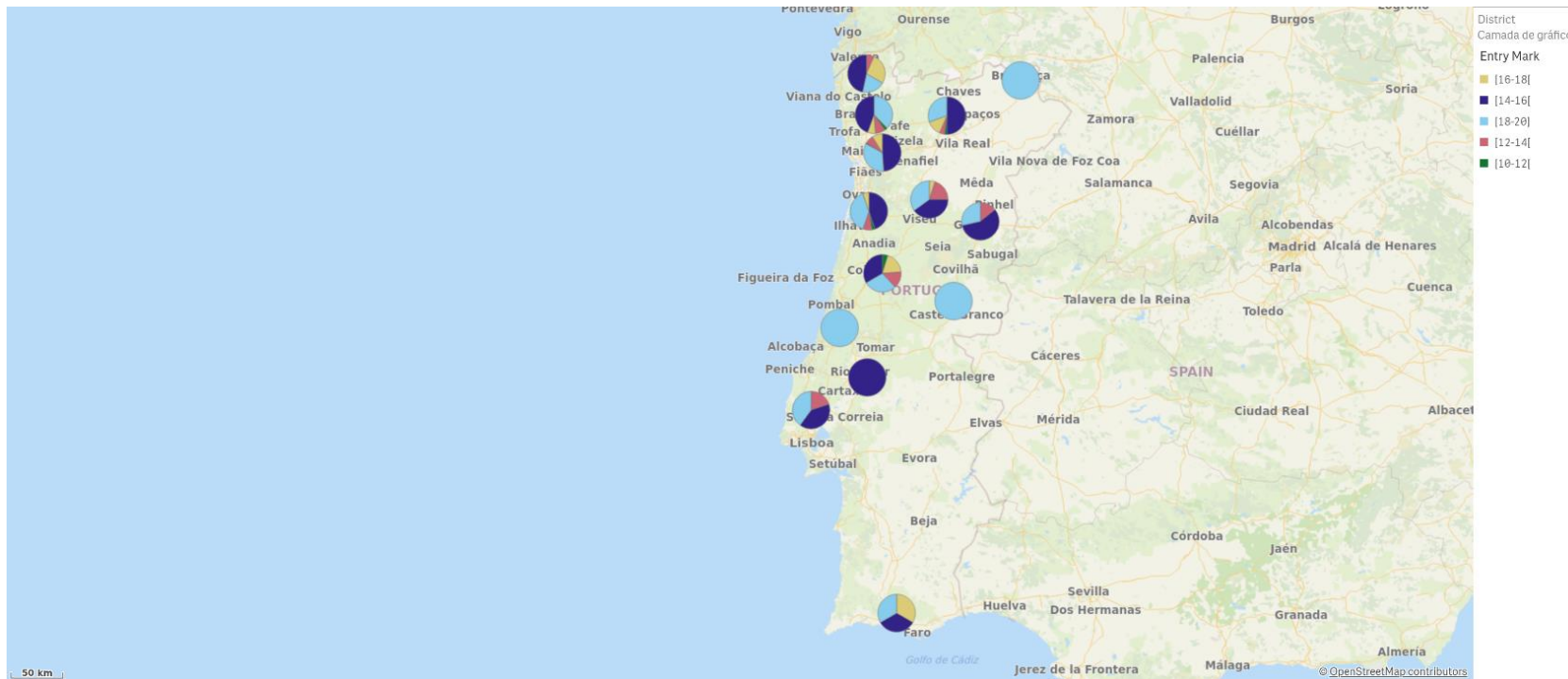
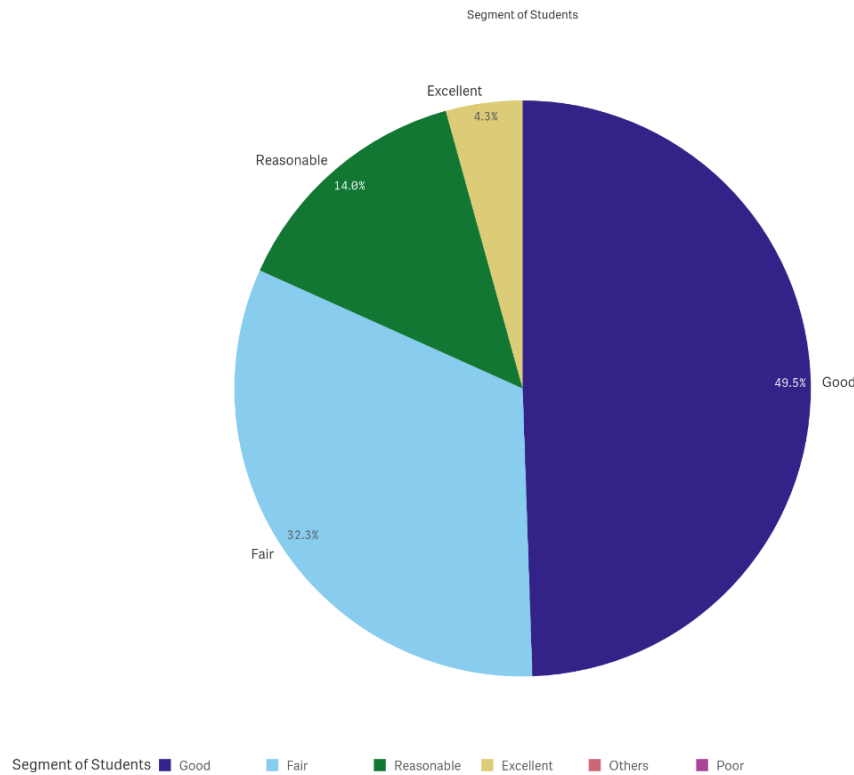


Figura 42 - Segmentos de alunos por origem geográfica e nota de ingresso

A Figura 42 apresenta a proveniência geográfica dos alunos, ilustrada de acordo com a densidade de alunos registada em cada localidade. A maioria dos alunos provém do Norte de Portugal, mais especificamente dos concelhos da área metropolitana do Porto, e de cidades do litoral do país. Verifica-se que as melhores notas de ingresso, em termos relativos, provêm de alunos de Bragança, Castelo Branco e Pombal.

3.8.7 Segmentos de alunos caracterizados pela sua performance académica



Os segmentos dos alunos foram criados seguindo os seguintes critérios:

- Excellent – Curso concluído em n anos e nota final ≥ 16 ;
- Good – Curso concluído em n anos e nota final ≥ 13 e < 16 ;
- Reasonable – Curso concluído em n anos e nota final < 13 ;
- Fair – Curso concluído em $n+1$ anos e nota final ≥ 12 ;
- Poor – Curso concluído em $n+1$ anos e nota final ≥ 10 e < 12 ;
- Other – Curso concluído em $n+2$ anos ou superior.

Figura 43 - Segmentos alunos caracterizados pela sua performance académica

Considerando a amostra de alunos fornecida, 49,5% dos alunos situam-se no segmento “Good” em termos de desempenho académico, seguindo-se 32,3% no seguimento “Fair” (cf. Figura 43).

3.8.8 Correlação entre a nota final de curso e a nota de ingresso

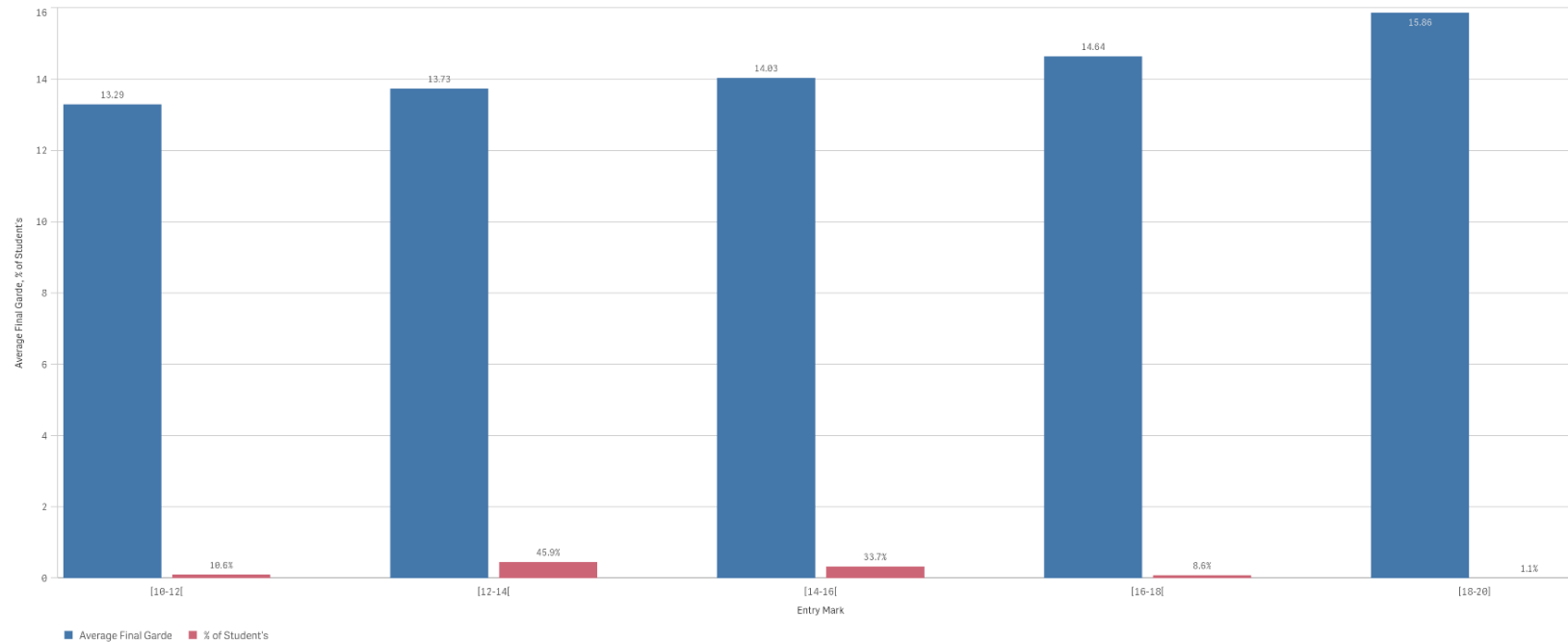


Figura 44 - Correlação entre a nota final de curso e a nota de ingresso

Da análise do gráfico da Figura 44, verifica-se que existe uma correlação positiva entre a nota de ingresso e a nota final do aluno, isto é, alunos com notas de ingresso mais altas (ex: gama [18-20]) têm também em média, notas finais mais altas de conclusão do curso. Esta correlação verifica-se para todos os segmentos de notas de ingresso. No entanto, a percentagem de alunos correspondente ao segmento de notas de ingresso que originam melhores notas de conclusão do curso é de apenas 1,1% (segmento de notas de ingresso [18-20]).

3.8.9 Segmentos de alunos por género e nota de ingresso, relacionados com a nota final de curso

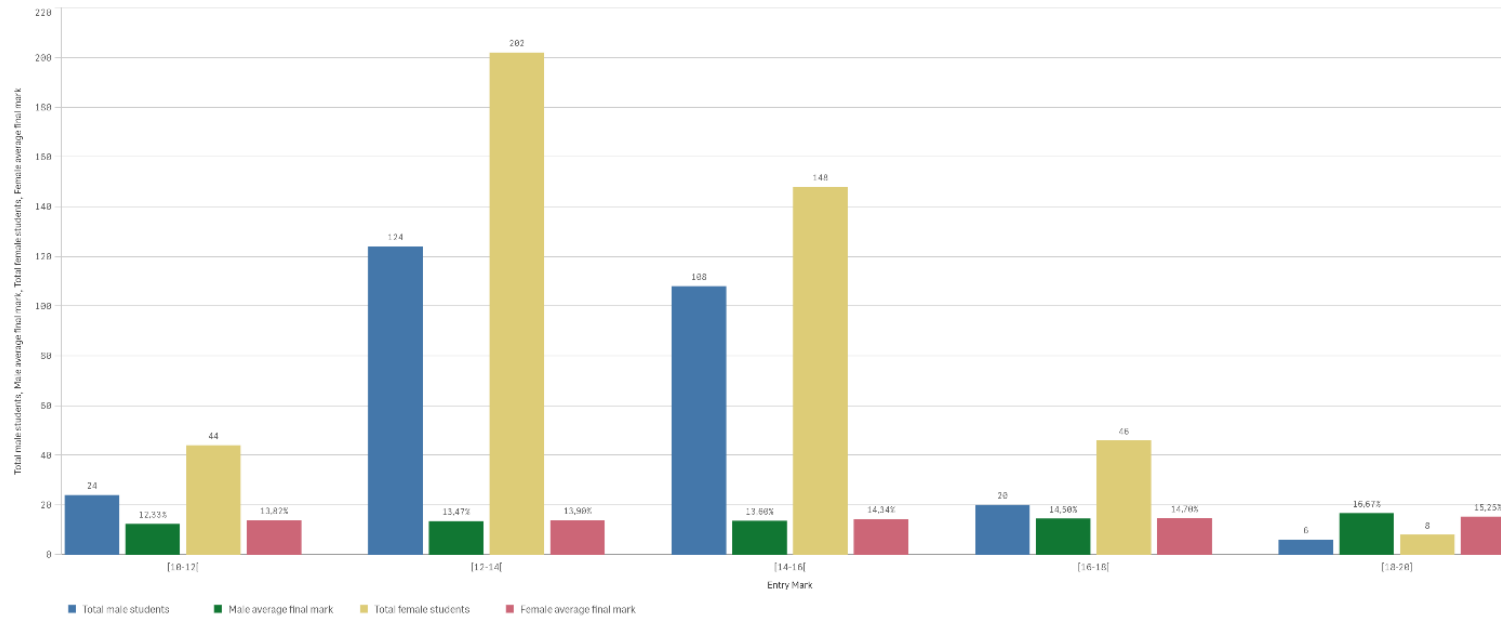


Figura 45 - Segmentos de alunos por género e nota de ingresso, relacionados com a nota final de curso

Os intervalos de notas de ingresso mais significativos em termos de número de alunos são os correspondentes aos grupos [12-14] e [14-16], conforme ilustrado na Figura 45. Porém, é no intervalo de notas ingresso [18-20] que existe uma melhor correlação com os resultados da nota final de curso, sendo que os alunos deste intervalo apresentam melhor desempenho em termos médios (nota média de conclusão de curso de 15,25% para os alunos do sexo feminino e 16,67% para alunos do sexo masculino).

3.8.10 Eficiência no ensino - percentagem de graduados (por ano letivo)

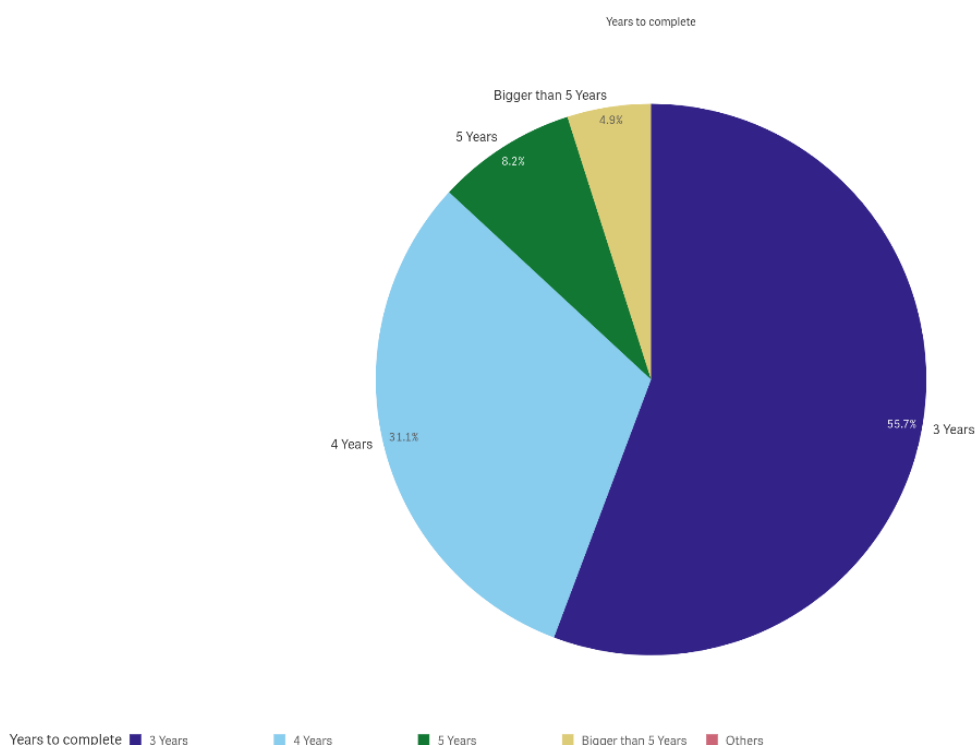


Figura 46 - Eficiência no ensino - percentagem de graduados (ano letivo 2016/2017)

O gráfico da Figura 46 apresenta, para o ano letivo 2016/2017 (previamente selecionado), as percentagens dos alunos que se licenciaram em n anos, $n+1$, $n+2$ e em mais de $n+2$ anos. 55,7% dos alunos concluem o curso em n (3) anos, o que corresponde à expectativa de duração dos cursos, enquanto 31,1% dos alunos concluem o curso em $n+1$ (4) anos.

3.9 Análise dos KPI's

Nesta secção apresentam-se os resultados da análise multidimensional efetuada baseada no DW definido na Secção 3.4 e, em particular, dos KPI's previamente identificados na Secção 3.3.

3.9.1 Número de alunos inscritos por género (por ano letivo)

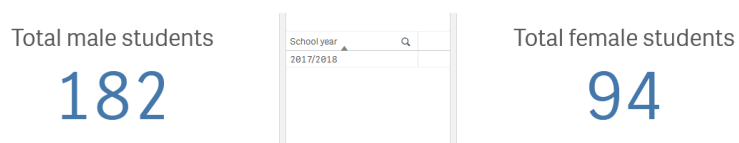


Figura 47 - Número de alunos inscritos por género (ano letivo 2017/2018)

A Figura 47 ilustra a distribuição dos alunos inscritos por género, após a seleção de um ano letivo (2017/2018).

3.9.2 Número de alunos por idade

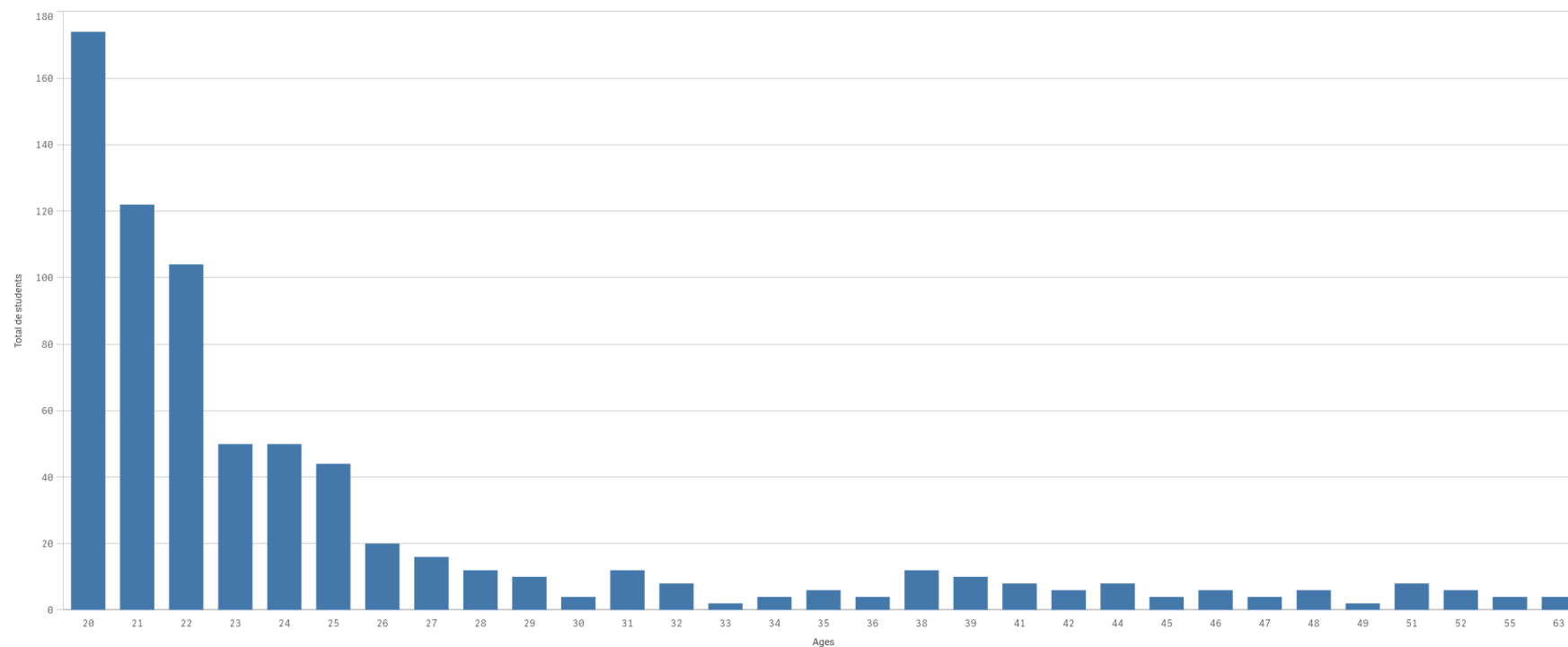


Figura 48 - Número de alunos por idade

O gráfico da Figura 48 apresenta o total de alunos por idade. Verifica-se que a maioria dos alunos têm idades compreendidas entre 21 e 23 anos.

3.9.3 Número de alunos por origem geográfica em Portugal

District	Q	Total students	% of students
Totais		730	100.0%
Porto		270	37.0%
Vila Real		92	12.6%
Braga		90	12.3%
Aveiro		76	10.4%
Coimbra		42	5.8%
-		42	5.8%
Viseu		40	5.5%
Viana do Castelo		30	4.1%
Guarda		14	1.9%
Lisboa		10	1.4%
Faro		6	0.8%
Ilha de São Miguel		4	0.5%
Leiria		4	0.5%
Santarém		4	0.5%
Bragança		2	0.3%
Castelo Branco		2	0.3%
Ilha das Flores		2	0.3%

Figura 49 - Número de alunos por origem geográfica em Portugal

A tabela da Figura 49 apresenta o total de alunos em valor e percentagem por distrito em Portugal. Verifica-se que 37% dos alunos são provenientes do distrito do Porto; 5,8% dos alunos não declararam o distrito ou este não pertence a Portugal.

3.9.4 Número de alunos por grau de escolaridade da mãe

Escolaridade Mãe	Q	Total de alunos
Totais		730
Ensino superior - Licenciatura - 1º Ciclo		60
Ensino superior - Licenciatura		58
Ensino médio		54
Ensino Secundário - 12º ano de escolaridade ou equivalente		52
Ensino pós-graduado - Doutoramento		50
Ensino superior - Bacharelato		50
Não sabe ler nem escrever		50
Desconhecido		48
Ensino Básico 1.º ciclo - 4º ano de escolaridade (antiga 4ª classe)		48
Ensino pós-graduado - Mestrado		46
Ensino Pós-secundário - Curso de Especialização Tecnológica		46
Ensino superior - curso técnico superior profissional ou equivalente		44
Sabe ler sem possuir o 4º ano de escolaridade (antiga 4ª classe)		44
Ensino Básico 3.º ciclo - 9º ano de escol. (antigo 5º ao liceal ou ensino téc.)		40
Ensino Básico 2.º ciclo - 6º ano de escol. (antigo 2º ao liceal ou ciclo prep.)		36
-		4

Figura 50 - Número de alunos por grau de escolaridade da mãe

A Figura 50 apresenta o número de alunos por grau de escolaridade da mãe. De acordo com os dados origem, existem 60 alunos cuja escolaridade da Mãe corresponde ao grau de Licenciatura.

3.9.5 Número de alunos por grau de escolaridade do pai

Escolaridade Pai	Q	Total de alunos
Totais		730
Ensino pós-graduado - Doutoramento		64
Ensino médio		56
Ensino Básico 2.º ciclo - 6º ano de escol. (antigo 2º ao liceal ou ciclo prep.)		54
Ensino superior - Bacharelato		52
Ensino superior - Licenciatura - 1º Ciclo		50
Sabe ler sem possuir o 4º ano de escolaridade (antiga 4ª classe)		50
Ensino Básico 1.º ciclo - 4º ano de escolaridade (antiga 4ª classe)		48
Ensino Pós-secundário - Curso de Especialização Tecnológica		48
Ensino superior - curso técnico superior profissional ou equivalente		48
Não sabe ler nem escrever		48
Ensino superior - Licenciatura		46
Desconhecido		44
Ensino Básico 3.º ciclo - 9º ano de escol. (antigo 5º ao liceal ou ensino téc.)		42
Ensino Secundário - 12º ano de escolaridade ou equivalente		42
Ensino pós-graduado - Mestrado		34
-		4

Figura 51 - Número de alunos por grau de escolaridade do pai

A Figura 51 apresenta o número de alunos por grau de escolaridade do pai. De acordo com os dados origem, existem 64 alunos cuja escolaridade do Pai corresponde ao grau de Doutoramento.

3.9.6 Número de alunos por situação profissional da mãe

Situação profissional mãe	Q	Total de alunos
Totais		730
Trabalha por conta de outrem		80
Trabalha para pessoa de família sem receber remuneração		76
Desempregado/a		68
Trabalha por conta própria - (como empregador)		68
Aluno, estudante		66
Doméstica/o		66
Reformado/a		66
Trabalha por conta própria - independente (sem empregados)		64
Desconhecida/não tem		58
Outra situação		58
Serviço militar		56
-		4

Figura 52 - Número de alunos por situação profissional da mãe

A tabela da Figura 52 apresenta o número de alunos por situação profissional da mãe. Observa-se que 80 alunos têm a mãe na situação “Trabalho por conta de outrem”, enquanto 68 alunos têm a mãe em situação de desemprego.

3.9.7 Número de alunos por situação profissional do pai

Situação profissional pai	Q	Total de alunos
Totais		730
Aluno, estudante		96
Reformado/a		72
Serviço militar		72
Trabalha por conta própria - (como empregador)		70
Doméstica/o		68
Desconhecida/não tem		64
Trabalha por conta própria - independente (sem empregados)		62
Trabalha por conta de outrem		60
Outra situação		56
Trabalha para pessoa de família sem receber remuneração		54
Desempregado/a		52
-		4

Figura 53 - Número de alunos por situação profissional do pai

A tabela da Figura 53 apresenta o número de alunos por situação profissional do pai. Observa-se que 96 alunos têm o pai na situação “Alunos, estudante”.

3.9.8 Número de alunos deslocados da sua residência permanente

Number of students displaced from their permanent residence

104

Figura 54 - Número de alunos deslocados da sua residência permanente

A Figura 54 indica o total de alunos inscritos deslocados da sua residência permanente.

3.9.9 Número de alunos inscritos por ano curricular (por ano letivo)

Ano letivo	Q	Ano curricular	Q	Total matrículas
Totais				-
2011/2012		1		4
2011/2012		2		2
2011/2012		3		4
2011/2012		4		2
2012/2013		1		10
2012/2013		2		20
2012/2013		3		4
2013/2014		1		56
2013/2014		2		4
2014/2015		1		110
2014/2015		2		140
2014/2015		3		116
2014/2015		4		18
2015/2016		1		250
2015/2016		2		188
2015/2016		3		206
2015/2016		4		4
2016/2017		1		26
2016/2017		2		102
2016/2017		3		126
2016/2017		4		44
2017/2018		1		14
2017/2018		2		56
2017/2018		3		86
2017/2018		4		114
2018/2019		1		2
2018/2019		2		30
2018/2019		3		14
2018/2019		4		12

Figura 55 - Número de alunos inscritos por ano curricular (por ano letivo)

A Figura 55 apresenta uma lista com o número de inscrições por ano letivo e ano curricular. Observa-se que o segundo ano curricular do ano letivo 2015/2016 foi o que registou um maior número de inscrições.

3.9.10 Nota média por UC por ano letivo

Unidade curricular	2017/2018								
	Administração e Gestão da Educação	Ciências da Educação	Direito	Direito Especialização em Ciências Jurídico-Políticas	Economia	Gestão	Psicologia Clínica e da Saúde	Solicitadoria	Turismo
Alojamento e Hospitalidade	-	-	-	-	-	-	-	-	199
Análise Financeira e de Projetos	-	-	-	-	183	199.33333	-	-	-
Análise Microeconómica de Mercados	-	-	-	-	152.33333	177	-	-	-
Avaliação em Contexto Educativo	-	-	-	-	-	-	-	-	-
Ciência Política	-	-	179.4	190.2	-	-	-	-	-
Comércio Internacional	-	-	-	-	151	-	-	-	-
Comunicação e Gestão de Conflitos	176	-	-	-	-	-	-	-	-
Contabilidade Analítica	-	-	-	-	162.66667	-	-	-	-
Contabilidade de Gestão	-	-	-	-	153	200	-	-	-
Contabilidade Financeira	-	-	-	-	186.5	151	-	-	-
Contabilidade Geral	-	-	-	-	146.33333	129.5	-	-	-
Contencioso Administrativo e Tributário	-	-	178.87097	-	-	-	-	-	-
Contencioso da União Europeia	-	-	200	-	-	-	-	-	-
Contratos Cíveis	-	-	-	186	-	-	-	-	-
Contratos em especial	-	-	174.07895	-	-	-	-	-	-
Contratos Internacionais	-	-	189.14286	-	-	-	-	-	-
Controlo de Gestão	-	-	-	-	-	164	-	-	-
Desenvolvimento e Operacionalização de Projetos Turísticos	-	-	-	-	-	-	-	-	193.3
Desenvolvimento Económico	-	-	-	-	169.25	-	-	-	-
Direito Administrativo I	-	-	176.6	-	-	-	-	-	-
Direito Administrativo II	-	-	161.55556	-	-	-	-	-	-
Direito Comercial	-	-	156.83333	-	-	-	-	-	-
Direito Constitucional	-	-	182	-	-	-	-	150	-
Direito da Banca e Seguros	-	-	180.62162	-	-	-	-	-	-
Direito da Família	-	-	170.33333	-	-	-	-	-	-
Direito da Insolvência	-	-	189.57143	-	-	-	-	181.16667	-
Direito da União Europeia	-	-	191.33333	-	-	-	-	200	-
Direito das Obrigações I	-	-	181.5	-	-	-	-	-	-

Figura 56 - Nota média por UC por ano letivo

A tabela da Figura 56 apresenta a nota média por UC para um ano letivo selecionado (2017/2018), indicando o curso correspondente. Nesta tabela é possível observar diferentes médias finais para a mesma UC, mas em cursos diferentes. Por exemplo, a UC “Contabilidade Geral” tem uma média final de 186,5 no curso de “Economia” e 151 no curso de “Gestão”.

3.9.11 Percentagem de alunos por intervalo de nota, por UC e por ano letivo

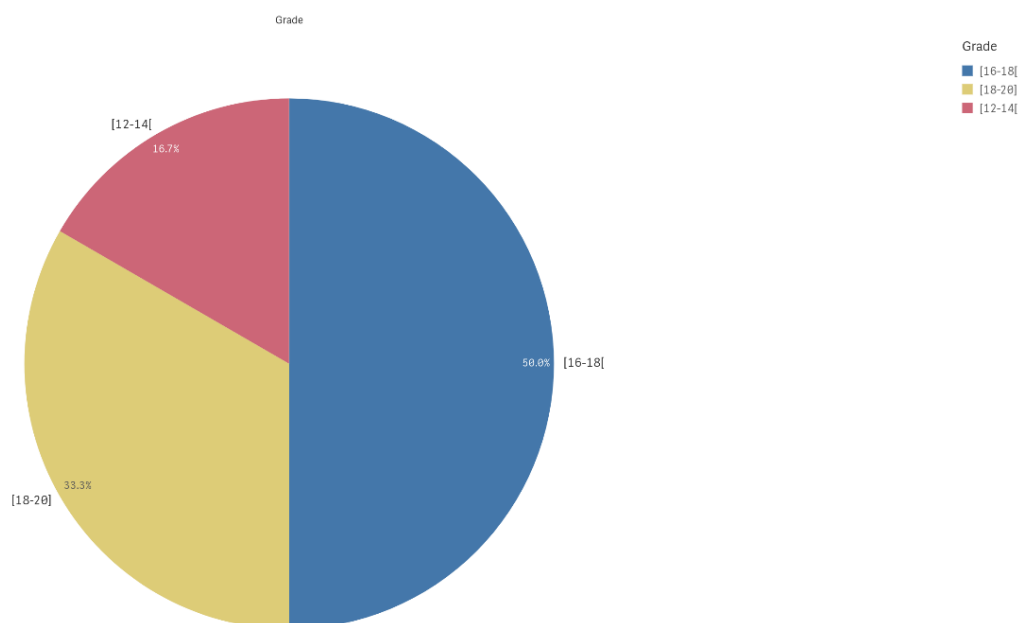


Figura 57 - Percentagem de alunos por intervalo de nota, por UC e por ano letivo (Ciência Política – 2016/2017)

Da análise do gráfico da Figura 57 verifica-se que 50% dos alunos na UC “Ciência Política” (previamente selecionada) e no ano letivo 2016/2017 (também previamente selecionado) obtêm uma nota final entre [16-18], seguindo-se os intervalos [18-20[com uma percentagem de 33,3% e [12-14[com percentagem de 16,7%.

3.9.12 Número médio de inscrições até ao final do curso por aluno

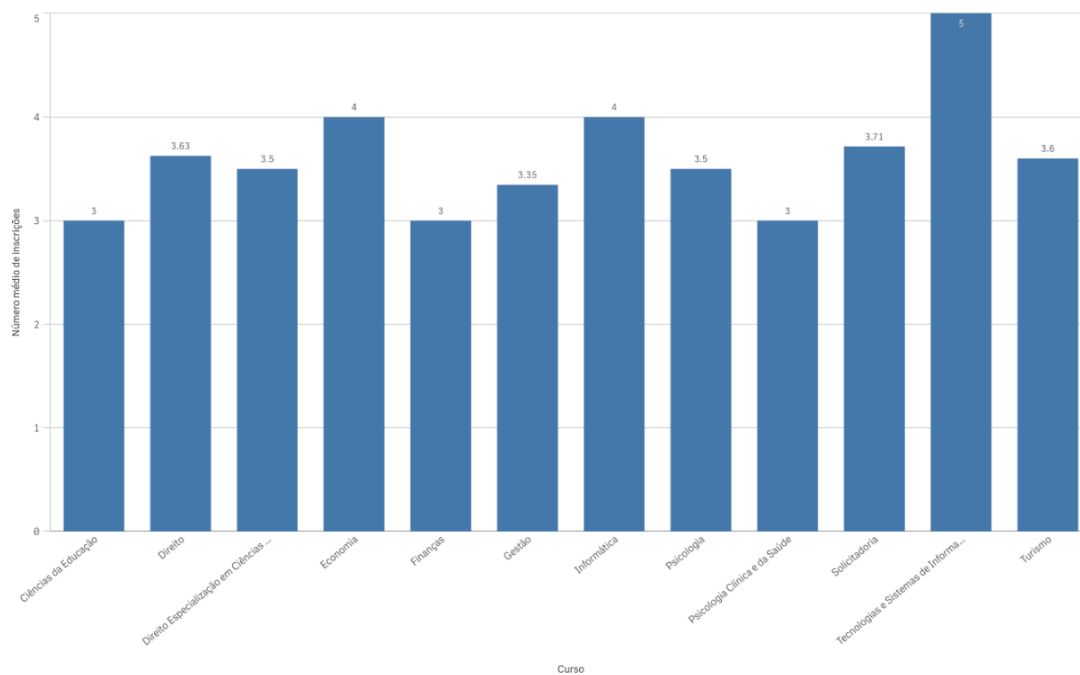


Figura 58 - Número médio de inscrições até ao final do curso por aluno

No gráfico da Figura 58 observa-se que o curso com maior número de inscrições por aluno até ao final do curso é “Tecnologias e Sistemas de Informação” com 5 inscrições.

3.9.13 Percentagem de aprovações por UC e por ano letivo

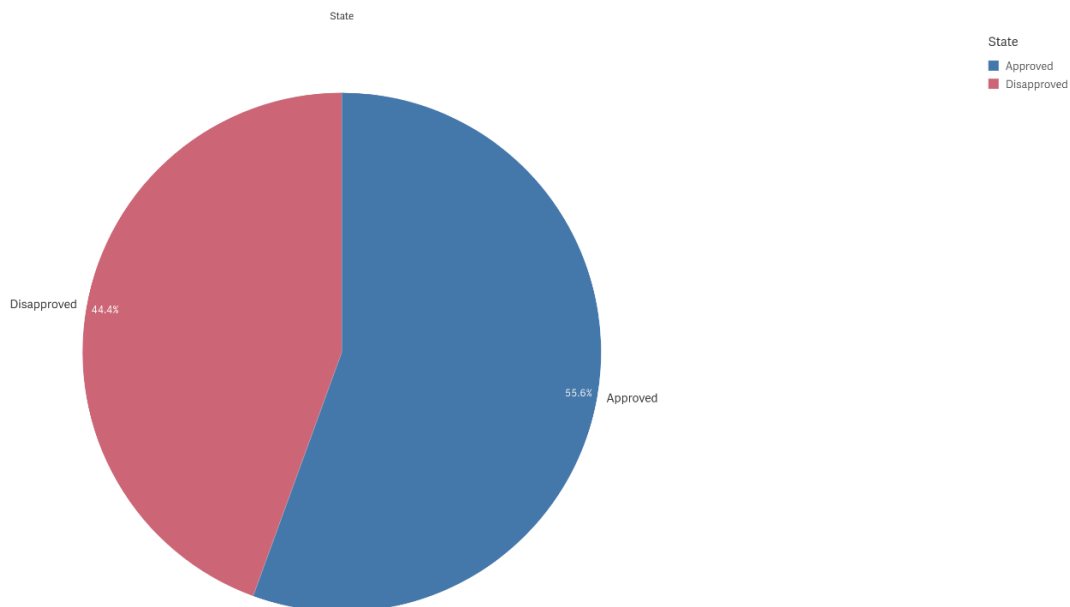


Figura 59 - Percentagem de aprovações por UC e por ano letivo (Estatística – 2013/2014)

A Figura 59 apresenta a percentagem de aprovações para a UC “Estatística” (previamente seleccionada) no ano letivo “2013/2014” (também previamente seleccionado). Verifica-se que 44,4% dos alunos não obtiveram aprovação.

3.9.14 Percentagem de graduados em N anos (por ano letivo)

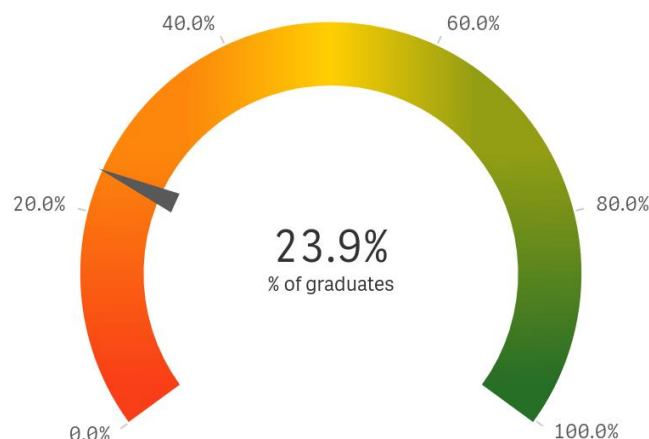


Figura 60 - Percentagem de graduados em N anos (2017/2018)

Pela análise do gráfico da Figura 60, verifica-se que a percentagem de alunos graduados em n anos para o ano letivo de 2017/2018 (previamente seleccionado) é de 23,9%. O valor apresentado por este indicador encontra-se na zona “laranja-vermelho” em relação às metas alvo previamente definidas.

3.9.15 Percentagem de graduados em N+1 anos (por ano letivo)

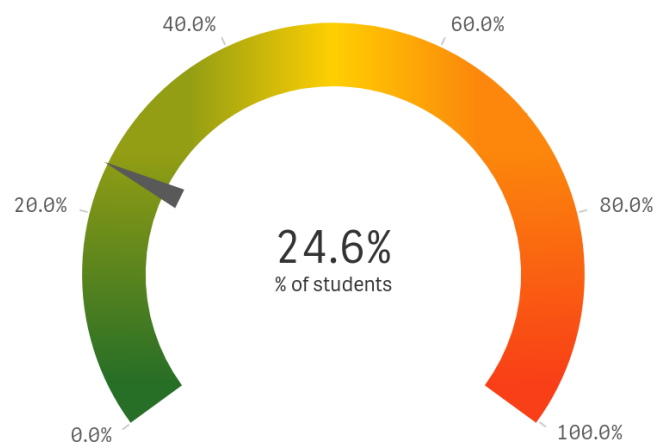


Figura 61 - Percentagem de graduados em N+1 anos (2017/2018)

Pela análise do gráfico da Figura 61, verifica-se que a percentagem de alunos graduados em n+1 anos para o ano letivo de 2017/2018 (previamente seleccionado) é de 24,6%. O valor apresentado por este indicador encontra-se na zona “verde amarelado” em relação às metas alvo previamente definidas.

3.9.16 Percentagem de graduados em N+2 anos (por ano letivo)

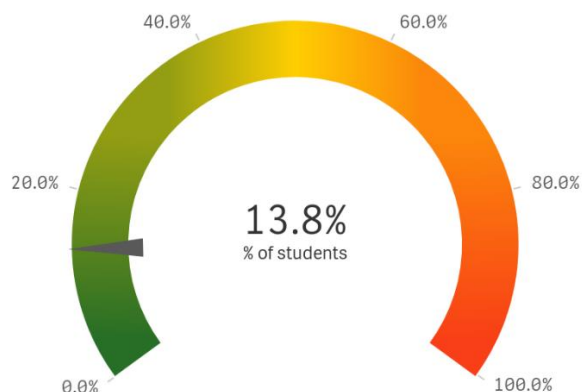


Figura 62 - Percentagem de graduados em N+2 anos (ano letivo 2017/2018)

Pela análise do gráfico da Figura 62, verifica-se que a percentagem de alunos graduados em n+2 anos para o ano letivo de 2017/2018 (previamente seleccionado) é de 13,8%. O valor apresentado por este indicador encontra-se na zona “verde” em relação às metas alvo previamente definidas.

3.9.17 Percentagem de graduados em mais de N+2 anos (por ano letivo)

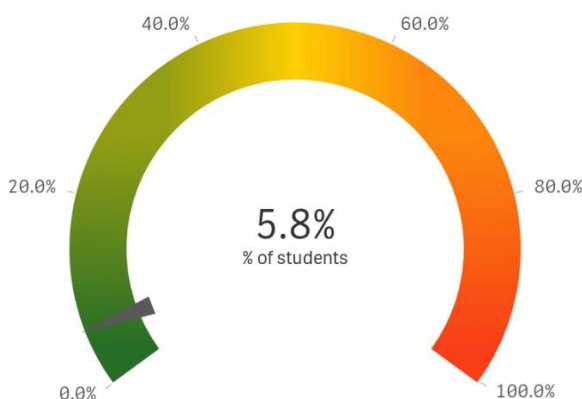


Figura 63 - Percentagem de graduados em mais de N+2 anos (ano letivo 2017/2018)

Pela análise do gráfico da Figura 63, verifica-se que a percentagem de alunos graduados em n+2 anos para o ano letivo de 2017/2018 (previamente seleccionado) é de 5,8%. O valor apresentado por este indicador encontra-se na zona “verde” em relação às metas alvo previamente definidas.

3.9.18 Nota média de conclusão por curso e por ano letivo

Study	School Year							
	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019
Administração e Gestão da Educação	-	-	160	-	156.66667	160	150	-
Ciência Jurídica Forense	-	-	-	-	-	155	-	-
Ciências da Educação	-	-	-	165	166.66667	-	166	-
Ciências da Educação - Área de Especialização em Educação e Intervenção Sociocomunitária	-	-	-	-	-	170	-	180
Direito	130	126	131.33333	134.08163	145	135.08772	135.39683	137.14286
Direito Especialização em Ciências Jurídico-Empresariais	-	150	150	160	150	150	-	160
Direito Especialização em Ciências Jurídico-Políticas	-	-	-	-	-	-	164	164
Direito Especialização em Ciências Jurídico-Processuais	-	-	-	-	-	140	-	-
Economia	-	120	130	127.14286	134.5	133.07692	122.5	120
Educação Social	-	140	140	140	148	145	-	-
Finanças	-	-	-	150	150	140	-	140
Gestão	130	120	128	129.66667	138.97436	135	126.66667	136
Gestão da Hospitalidade	-	-	-	-	141.25	140	-	-
Informática	-	-	-	-	147.5	137.5	-	150
Património Artístico Conservação e Restauro	-	-	-	-	-	150	-	-
Património Artístico, Conservação e Restauro	160	-	-	-	-	-	-	-
Psicologia	110	120	120	127	142.08333	132.5	-	-
Psicologia Clínica e da Saúde	-	-	-	158	153.33333	153.125	155.71429	158
Solicitadoria	120	120	122.5	120	125	122.30769	122.30769	-
Tecnologias e Sistemas de Informação	-	130	-	-	-	-	-	-
Turismo	-	120	120	125	144.5	130	139.61538	130

Figura 64 - Nota média de conclusão por curso e ano letivo

A tabela da Figura 64 apresenta as notas médias de conclusão por curso e por ano letivo. A título exemplificativo, a nota média do curso de Direito no ano letivo 2015/2016 foi de 150 (15 valores).

3.9.19 Dashboards

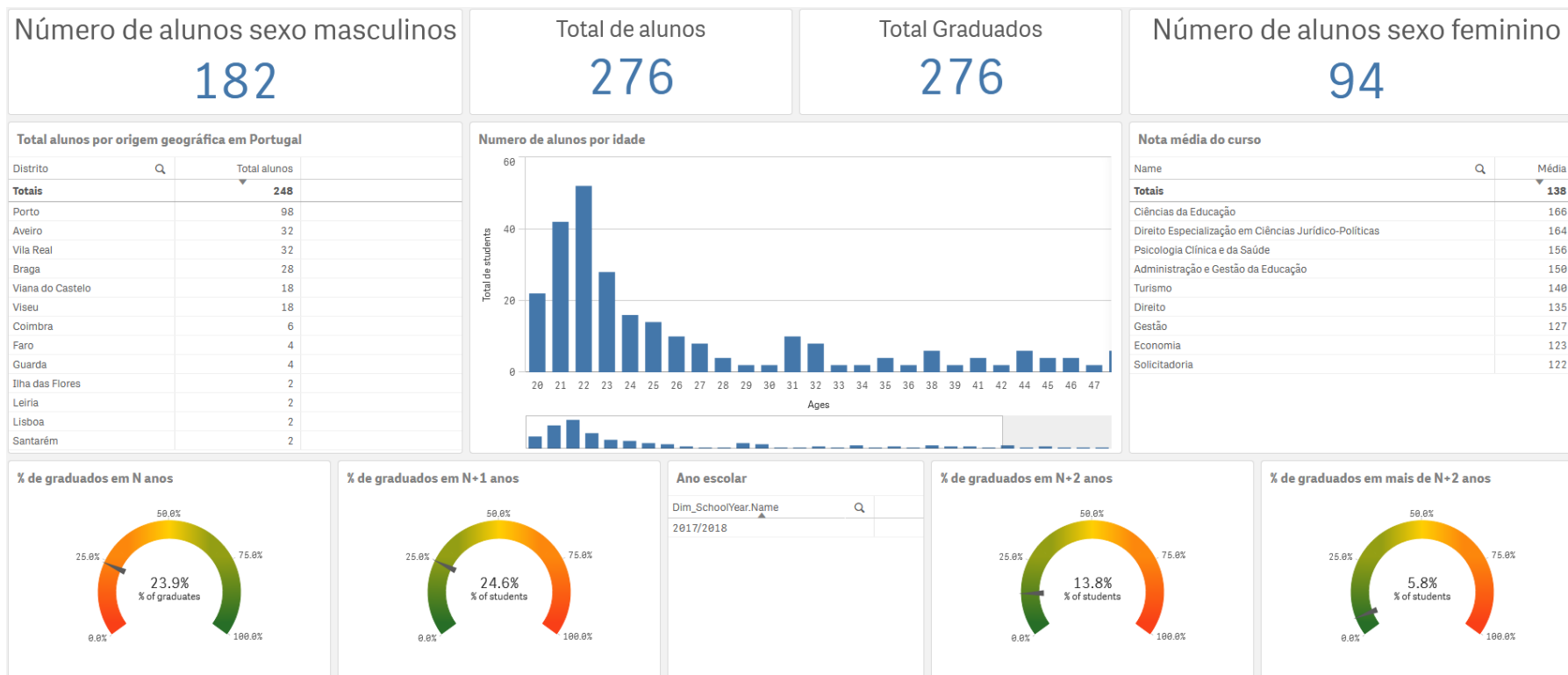


Figura 65 - Dashboard de análise sociodemográfica e de desempenho académico

A Figura 65 apresenta um exemplo de um *dashboard* com alguns dos indicadores relevantes sociodemográficos e de eficiência de ensino descritos nas secções anteriores relativos a um ano letivo selecionado (2017/2018).

4 CONCLUSÃO

4.1 Reflexão sobre o trabalho realizado

Este projeto teve como objetivo genérico endereçar a utilização de um sistema de BI no ES que atue no âmbito da monitorização e análise do desempenho académico e demonstrar as várias etapas da respetiva implementação e desafios associados, bem como as mais valias que estes aportam.

O projeto foi realizado numa IES privada, a UPT. Pretendeu-se analisar, utilizando uma prova de conceito de um sistema de BI, as características dos alunos matriculados nos vários cursos da UPT, e determinar a influência de diversos fatores no sucesso escolar dos alunos. Esses fatores incluíram dados relacionados com a formação académica dos alunos antes de ingressar no ES e dados sociodemográficos que foram cruzados com informações sobre o desempenho dos alunos ao longo e na conclusão do curso na IES.

Os resultados obtidos neste projeto mostram que os sistemas de BI poderão ter um papel importante no suporte do processo de garantia da qualidade do ensino de uma EIS.

Analisando a primeira contribuição definida para o projeto, apesar de os dados de origem serem fictícios, devido a constrangimentos resultantes da entrada em vigor do RGPD posterior ao início do projeto, o que dificulta a interpretação com significado útil dos dados analisados, poder-se-á afirmar que esta prova de conceito permitiu a demonstração das mais-valias que um sistema de BI poderá trazer ao negócio, neste caso particular, à oferta formativa e processo de ensino-aprendizagem de uma IES.

Com base nos resultados obtidos, e com o conhecimento que se demonstrou ser possível obter dos perfis dos alunos e dos fatores que condicionam o respetivo desempenho académico, poderão ser exploradas oportunidades a nível curricular e extracurricular da oferta formativa, bem como iniciativas de prestações adaptadas e projetos de tutoria aos vários segmentos de alunos, e assim justificando um possível investimento a ser efetuado.

Assim, o modelo do DW proposto permitiu dar resposta a todas as necessidades de informação do projeto e o sistema de BI desenvolvido permitiu correlacionar os dados

em análise e facultando informação útil sobre os alunos, aos agentes de educação, de forma a se poderem implementar novas estratégias de ensino-aprendizagem mais ajustados ao público-alvo.

Relativamente à segunda contribuição definida para o projeto, o modelo de dados definido para o DW, os indicadores de eficiência formativa e de caracterização dos alunos e as correlações identificadas poderão servir de base a outros projetos e iniciativas de análise do desempenho dos alunos no ES.

A utilização da ferramenta de software de BI adotada neste trabalho, o Qlik Sense, mostrou-se adequada para atingir os objetivos do negócio. As facilidades de visualização gráfica disponibilizadas pela ferramenta revelaram-se satisfatórias, considerando a galeria de gráficos disponível e a possibilidade de criação de *dashboards*. No entanto, a criação de gráficos requiere a utilização de comandos específicos, que não são intuitivos para um utilizador sem conhecimentos técnicos. O processo ETL e a criação do DW puderam ser efetuados de uma forma sistematizada, embora requerendo conhecimentos técnicos, especialmente na etapa da transformação e limpeza dos dados.

4.2 Obstáculos enfrentados

A implementação desta prova de conceito alertou também para vários aspetos condicionantes, que obrigatoriamente terão de ser superados, a nível de implementação, para garantir que o investimento valerá a pena.

A disponibilidade dos dados a analisar, nos sistemas operacionais do negócio, é crucial. Como se pôde constatar, alguns indicadores (cf. Secção 3.3) não puderam ser calculados por falta de dados disponibilizados. Para além disso, por questões de segurança, não foi permitida pelo Centro de Sistemas de Informação da UPT a importação direta dos dados do sistema operacional para o sistema de BI desenvolvido.

O processo ETL foi exigente e os conhecimentos tecnológicos revelaram-se indispensáveis para se poder implementar esta fase com sucesso.

Relativamente ao nível de envolvimento e conhecimento necessários para se conduzir com sucesso um projeto de BI numa EIS, verificou-se que é crucial:

- 1) Envolver os utilizadores finais e os agentes de tomada de decisão para a importância e mais-valia da implementação de um projeto deste tipo para a IES;
- 2) Envolver igualmente os especialistas da área da Informática/Sistemas de Informação da organização para essa relevância, de forma a que prestem o apoio técnico necessário e se contornem os habituais problemas de relutância na prestação de informação.

Por um lado, a EIS tem de garantir o apoio técnico através de especialistas da área de Informática, por forma a possibilitar o processo automático de ETL, a partir do sistema operacional. Por outro, a participação de colaboradores da área do negócio, neste caso particular, diretores de departamento, coordenadores de curso e docentes, na equipa de implementação, é importante para se poderem discutir as correlações dos dados relevantes e a granularidade dos dados a considerar, assim como a definição dos KPIs, respetivas metas a atingir e periodicidade de cálculo.

4.3 Constrangimentos e limitações ao trabalho realizado

No decurso do projeto existiram algumas limitações ao trabalho realizado que dificultaram a concretização dos objetivos inicialmente previstos. Uma das dificuldades encontradas foi a entrada em vigor do RGPD, após o início do projeto. Esta situação não permitiu que os dados reais dos alunos fossem facultados pela IES e resultou em sucessivos atrasos no desenvolvimento do projeto pela necessidade de auscultação de vários órgãos da IES e pelo impasse gerado. Não foi, assim, possível sugerir estratégias concretas de ensino-aprendizagem e de comunicação da oferta como inicialmente previsto.

Problemas pessoais, a nível de saúde do autor deste trabalho, que o impossibilitaram de usar o computador contribuíram para o atraso no desenvolvimento do projeto. Paralelamente a situação pandémica vigente, com o aumento repentino de casos em janeiro de 2021 e confinamento obrigatório decretado em Portugal, dificultou os contactos do autor com o Centro de Sistemas de Informação da IES. Esta situação não permitiu um ritmo de trabalho sistemático, contribuindo para atrasos no processo e causando alguma desmotivação no autor.

4.4 Sugestões para Futuras Intervenções

Os sistemas de BI constituem uma ferramenta importante para a melhoria do desempenho das organizações, nomeadamente as IES. Para garantir o sucesso do investimento nestes sistemas, é necessário que as IES se consciencializem da importância atual das Tecnologias de Informação na sociedade e nas organizações em geral, designada por muitos pela era da Transformação Digital, e assim, dotar as suas instituições de práticas e competências necessárias para o sucesso da implementação de projetos de BI, e em geral, de qualquer outro tipo de intervenção de Tecnologias e Sistemas de Informação. As atividades de intervenção de Tecnologias e Sistemas de Informação não envolvem só aspetos tecnológicos, mas também organizacionais e de negócio.

Relativamente ao caso concreto do sistema de BI proposto neste projeto, antevê-se como necessário para trabalho futuro:

- 1) Obter consentimento explícito dos alunos no ato da matrícula para utilização dos seus dados para análise de perfis de comportamento de negócio e, assim, possibilitar a utilização de dados reais;
- 2) Obter uma melhor sensibilização e comprometimento dos especialistas da área de Informática/Sistemas de Informação da organização para o interesse deste tipo de projeto para a persecução dos objetivos de negócio da IES;
- 3) Permitir uma ligação direta do sistema de BI ao sistema operacional para o processo ETL e atualização periódica dos dados do DW;
- 4) Implementação de técnicas de *data mining* que permitam a melhoria do conhecimento dos perfis de alunos da IES e a previsão de resultados académicos;
- 5) Sensibilização dos docentes da IES da importância do conhecimento gerado pelo projeto para a reformulação necessária das suas práticas de ensino-aprendizagem;
- 6) Articulação com o departamento de Marketing da IES para definição das iniciativas a tomar para melhoria da relação aluno-IES e comunicação adequada da oferta formativa.

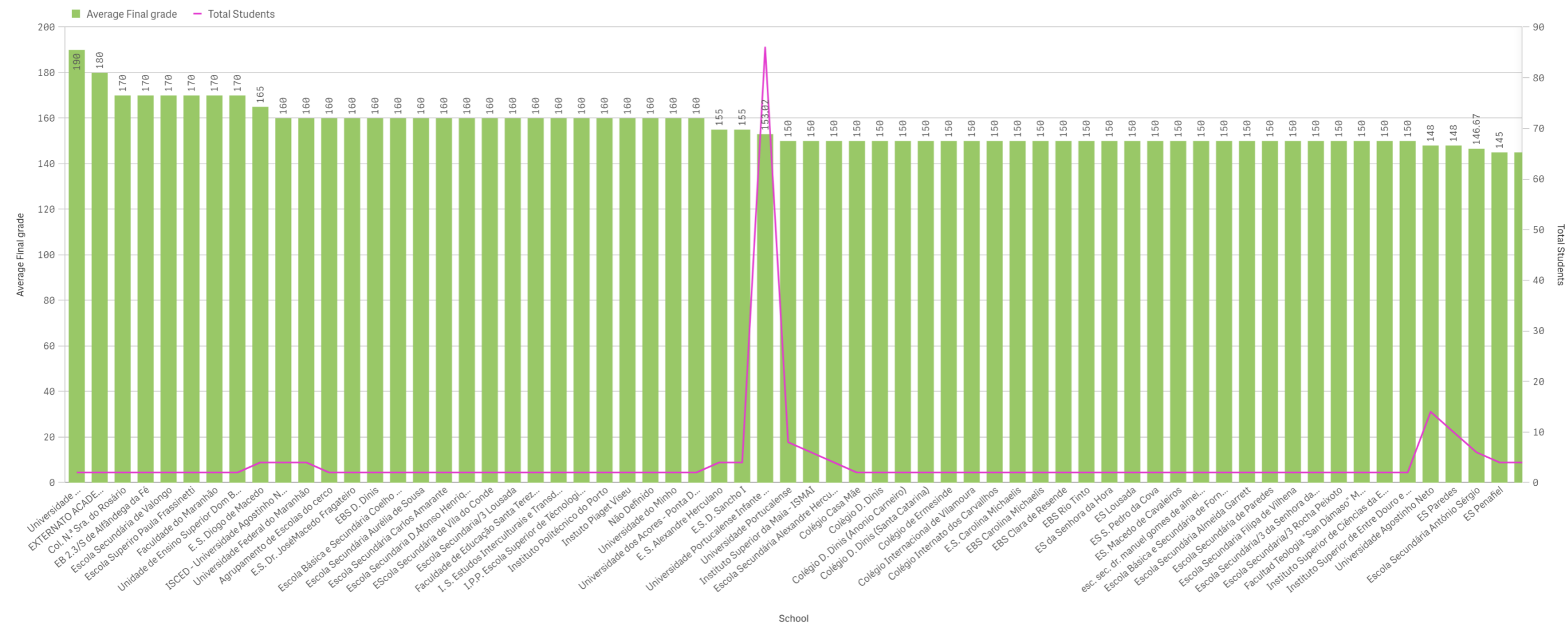
5 REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmad, S., Miskon, S., Alabdan, R., & Tlili, I. (2020). Exploration of influential determinants for the adoption of business intelligence system in the textile and apparel industry. *Sustainability (Switzerland)*, 12(18), 1–21. <https://doi.org/10.3390/su12187674>
- Analytics and Business Intelligence Platforms.* (2020). <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>
- Apraxine, D., & Stylianou, E. (2017). Business Intelligence in a Higher Educational Institution. *Proceedings of 2017 IEEE Global Engineering Education Conference (Educon2017)*, April, 1735–1746.
- Baepler, P., & Murdoch, C. (2010). Academic Analytics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*, 4(2). <https://doi.org/10.20429/ijstotl.2010.040217>
- Dedić, N., & Stanier, C. (2016). An evaluation of the challenges of multilingualism in data warehouse development. *ICEIS 2016 - Proceedings of the 18th International Conference on Enterprise Information Systems*, 1(May), 196–206. <https://doi.org/10.5220/0005858401960206>
- Fatima, N. (2020). *ETL vs. ELT: Qual é a diferença?* <https://www.astera.com/pt/tipo/blog/etl-vs-elt-qual-é-a-diferença/>
- Foote, K. D. (2017). *A Brief History of Business Intelligence.* <https://www.dataversity.net/brief-history-business>
- Franklin Jr., C. (2015). *Using Business Intelligence To Change Student Lives.* Informationweek. <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=113299331&lang=pt-br&site=ehost-live>
- Fred Friedrich. (2013). *Good BI Governance Is Just Good Business.* Educause Review. <https://er.educause.edu/articles/2013/12/good-bi-governance-is-just-good-business>
- Garani, G., & Helmer, S. (2012). *International Journal of Data Warehousing and Mining.* October. <https://doi.org/10.4018/jdwm.2012100102>
- Inmon, W. H. (1996). *Building the Data Warehouse (2nd Ed.).* John Wiley & Sons, Inc.
- Last night Study. (n.d.). *Fact constellation Schemas for Multidimensional Modal.* Data Warehouse. Retrieved June 10, 2021, from <http://www.lastnightstudy.com/Show?id=52/Fact-constellation-Schemas-for-Multidimensional-Modal>
- Lester, J. (2018). Learning Analytics in Higher Education. *Learning Analytics in Higher Education*, 178. <https://doi.org/10.4324/9780203731864>
- Luhn, H. P. (1958). A Business Intelligence System. *IBM Journal of Research and*

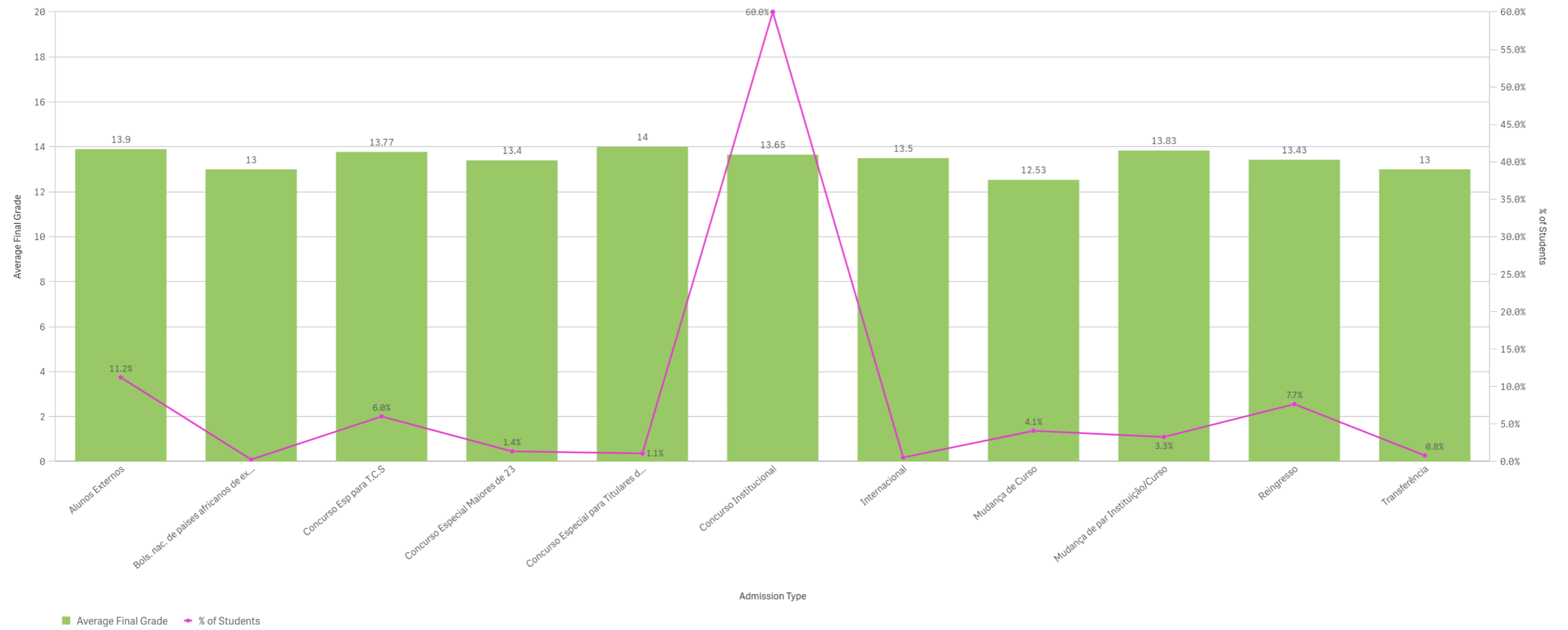
- Development*, 2(4), 314–319. <https://doi.org/10.1147/rd.24.0314>
- Microsoft. (2020). *Porquê o Power BI*. <https://powerbi.microsoft.com/pt-pt/why-power-bi/>
- Moody, D. L., & Kortink, M. a R. (2003). From ER Models to Dimensional Models: Bridging the Gap between OLTP and OLAP Design, Part I. *Business Intelligence Journal*, January 2003, 7–24.
- Myers, P. (2019). *Entenda o esquema em estrela e a importância para o Power BI*. <https://docs.microsoft.com/pt-br/power-bi/guidance/star-schema>
- OLAP.com. (n.d.). *What is the definition of OLAP*. Retrieved January 6, 2021, from <https://olap.com/olap-definition/>
- Piedade, M. (2011). *Business Intelligence no suporte ao conceito e à prática de Student Relationship Management em instituições de Ensino Superior*. 286. http://repositorium.sdum.uminho.pt/bitstream/1822/20461/1/TeseDout_MariaBeatrizPiedade_23Abr2012.pdf
- Qlik. (2021a). *Welcome to Qlik Help*. <https://help.qlik.com/>
- Qlik. (2021b). *What is a KPI*. <https://www.qlik.com/us/kpi>
- Ramos, S. (2016). *Business Intelligence (BI) & Analytics*. SolidQ.
- Richardson, J., Sallam, R., Schlegel, K., Kronz, A., & Sun, J. (2020). *Magic Quadrant for Analytics and Business Intelligence Platforms*. Gartner. https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb&ocid=mkto_eml_EM597235A1LA1
- S, V. (2018). *Data Warehouse Fact Constellation Schema and Design*. <https://dwgeek.com/data-warehouse-fact-constellation-schema-design.html/>
- Santos, M., & Ramos, I. (2017). *Business Intelligence - Da Informação ao Conhecimento* (FCA (Ed.); 3ª).
- Shah, F. (2014). *The History of BI: The 1980's and 90's*. Dataconomy.Com. <https://dataconomy.com/2014/07/the-history-of-bi-the-1980s-and-90s/>
- Stackowiak, R., Rayman, J., & Greenwalk, R. (2013). Oracle Data Warehousing and Business Intelligence Solutions. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).
- Sukumar, N. (2021). *Comparison of the Best Business Intelligence Plataforms in 2021*. <https://www.ideas2it.com/blogs/comparison-of-the-best-business-intelligence-bi-tools-of-2021/>
- Vaz, A. L. (2016). *Data mining*. <https://arthurlambletvaz.medium.com/data-mining-e-machine-learning-289aa966d4f7>

6 ANEXOS

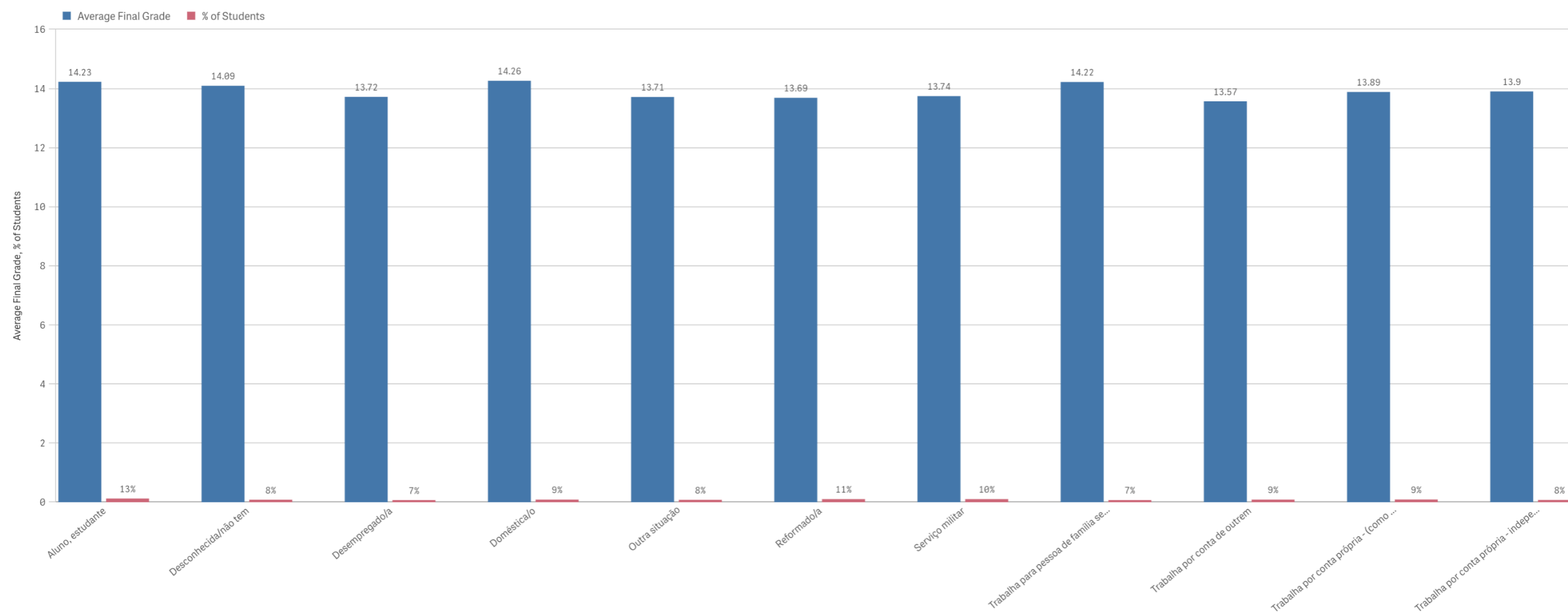
6.1 Correlação entre a nota final do curso e a escola frequentada no ensino secundário (cf. Figura 35)

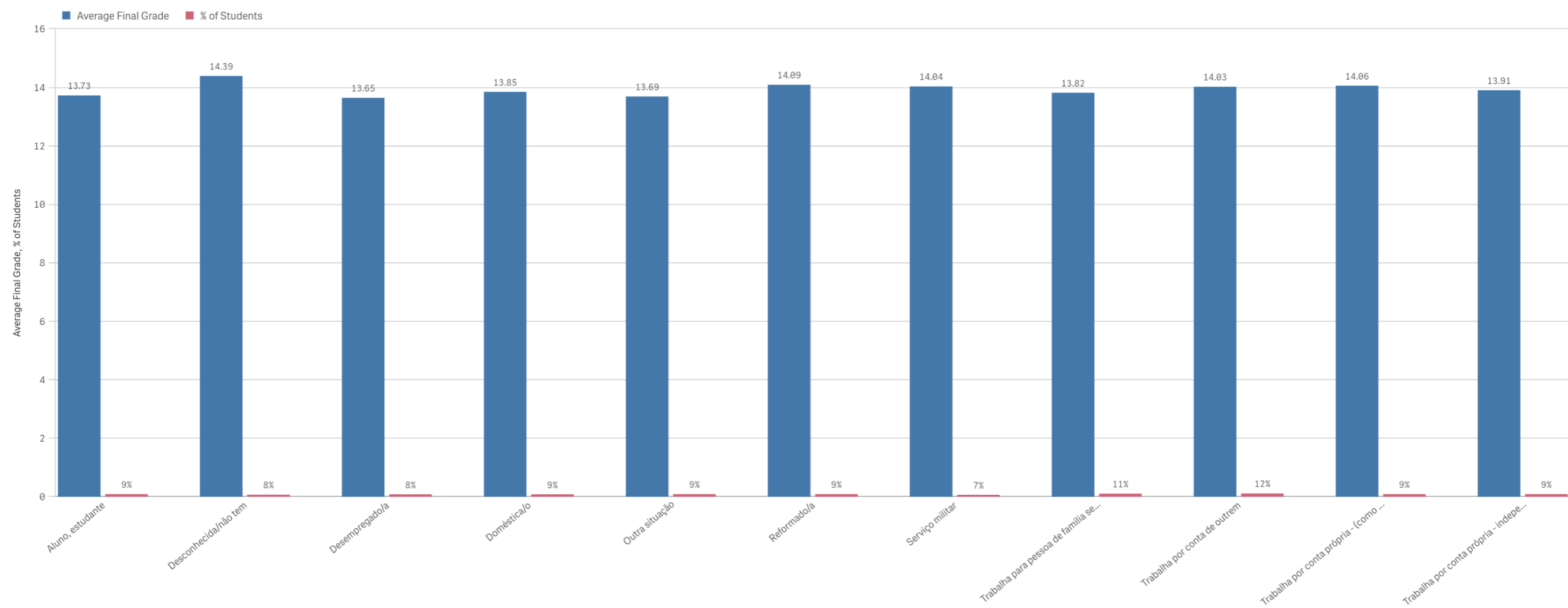


6.2 Segmentos de alunos por regime de ingresso e nota final de curso (cf. Figura 36)

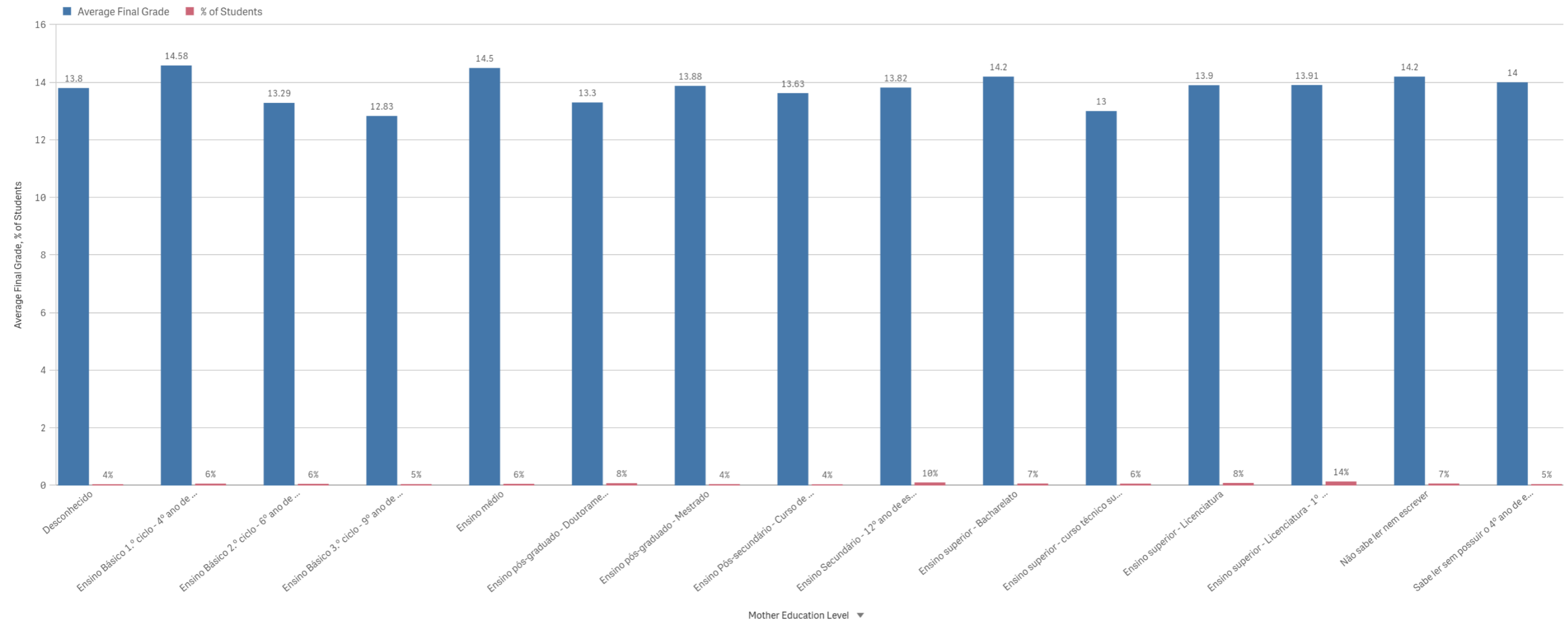


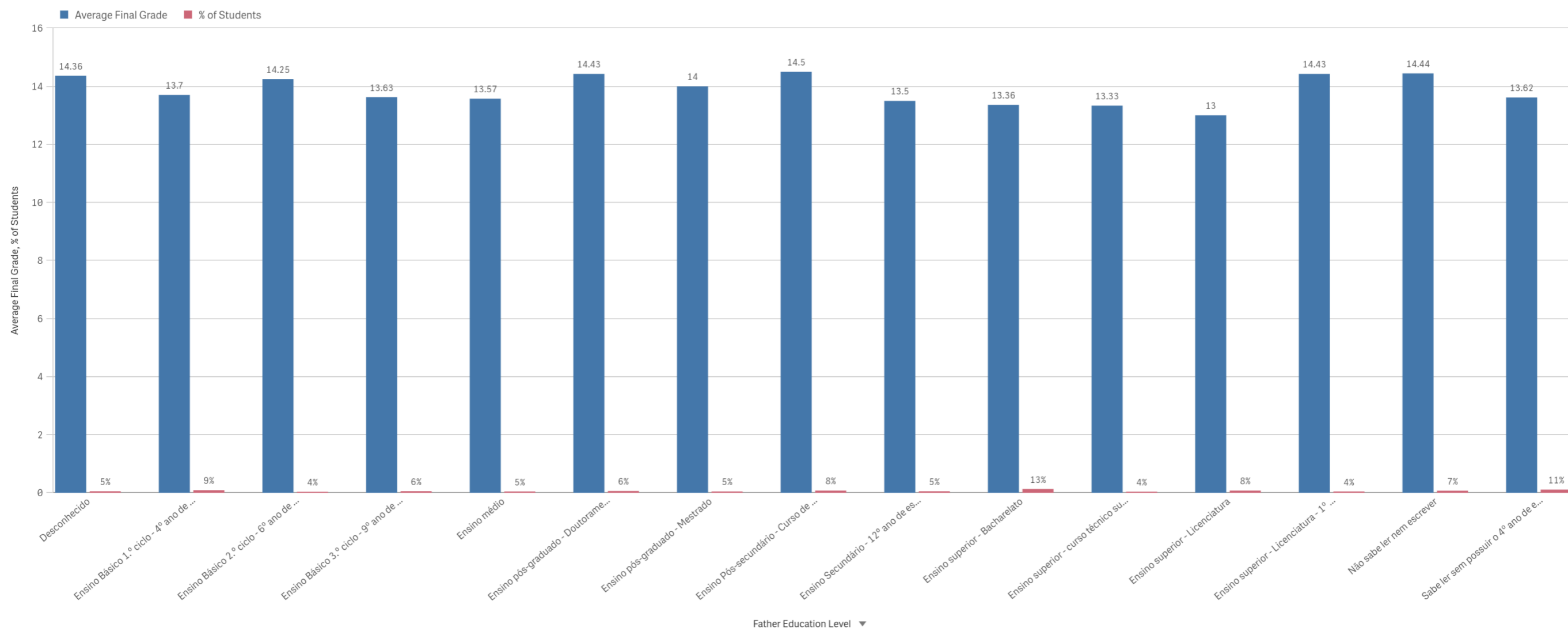
6.3 Segmentos de alunos por situação profissional dos pais e correlação com a nota final de curso (cf. Figura 37, Figura 38)



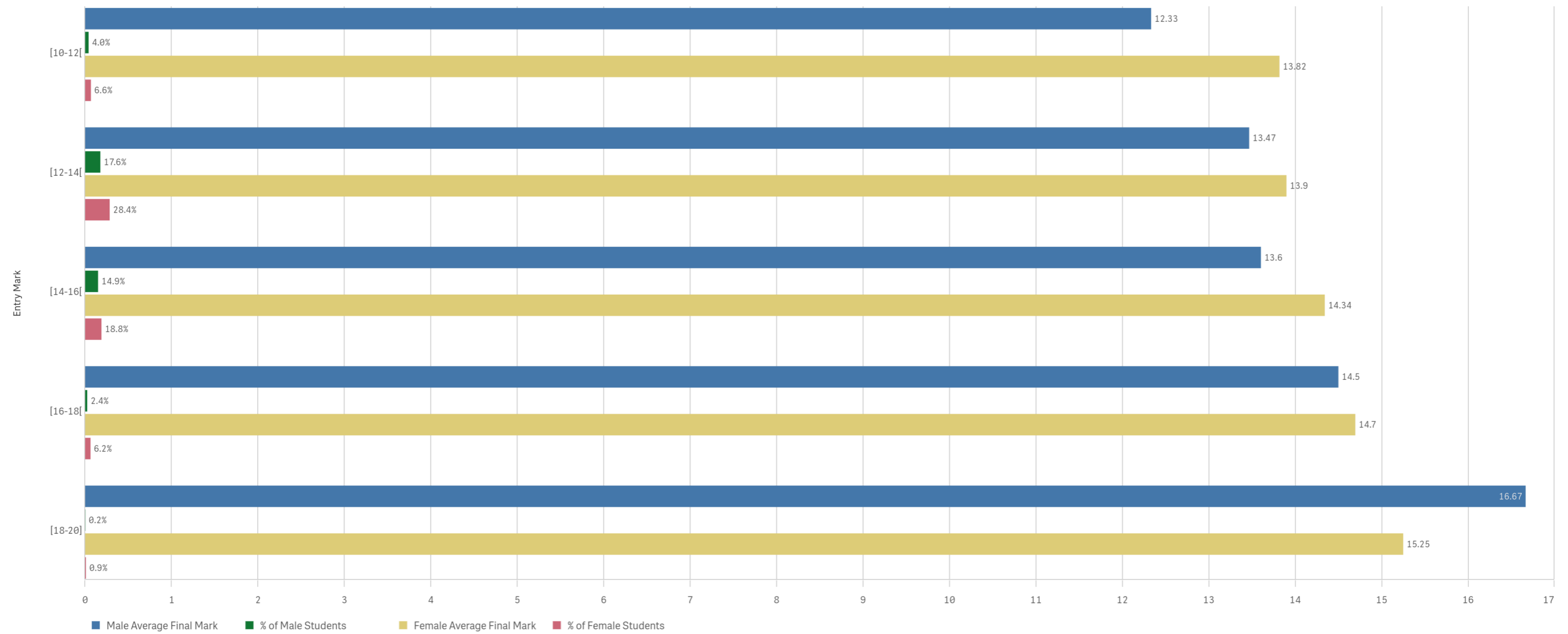


6.4 Segmentos de alunos por nível de escolaridade dos pais e correlação com a nota final de curso (cf. Figura 39, Figura 40)

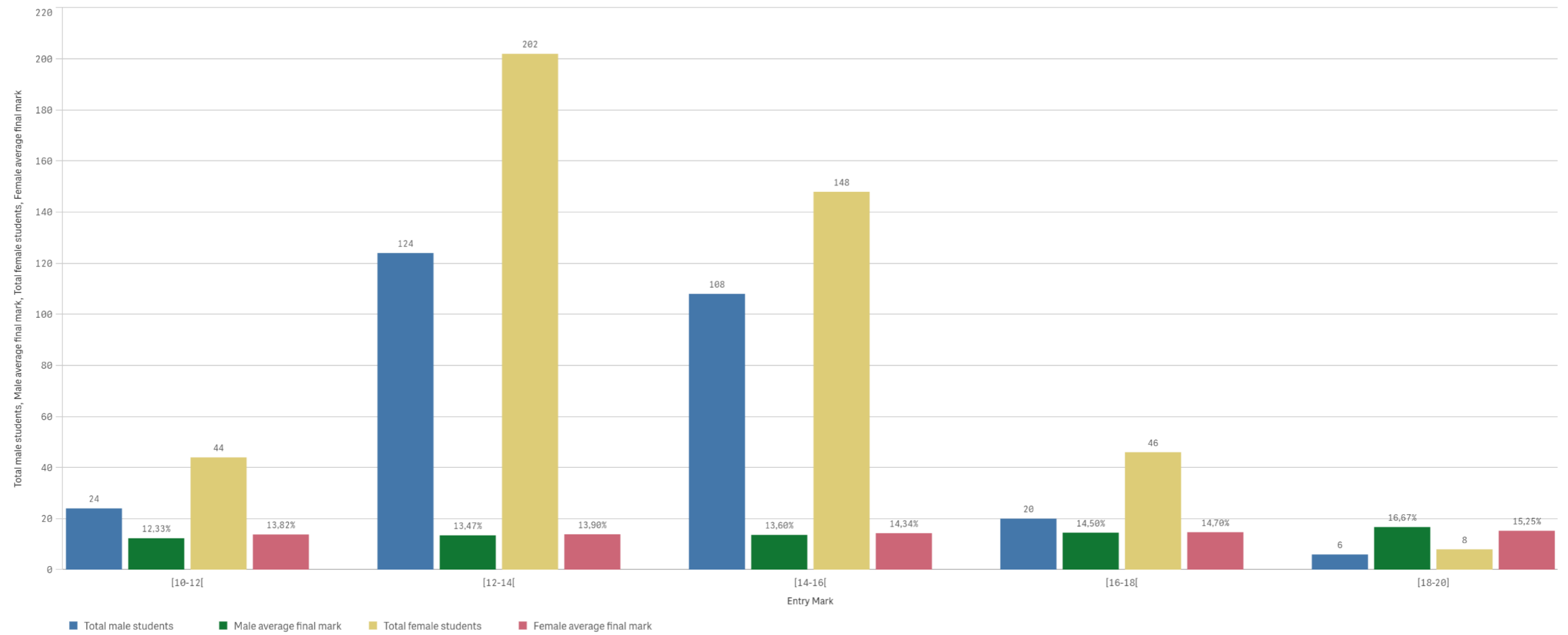




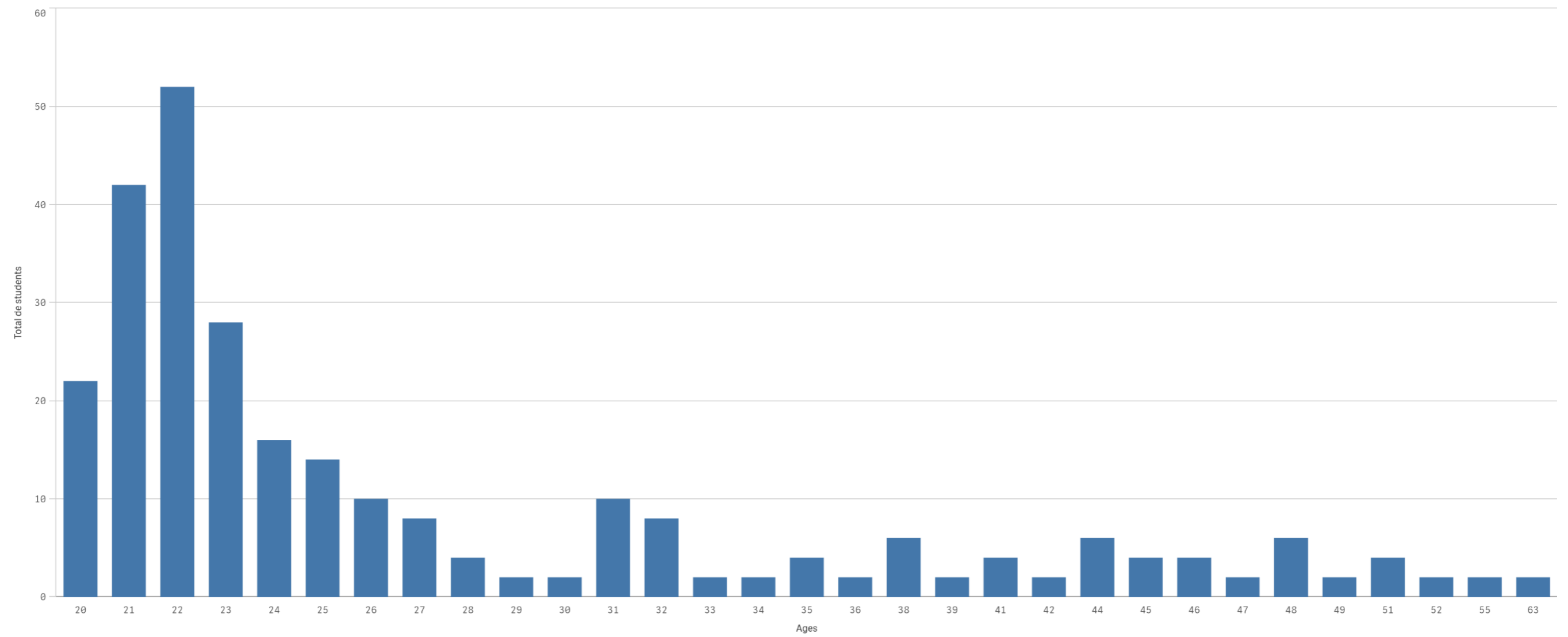
6.5 Segmentos de alunos por género e nota de ingresso e correlação com a nota final de curso (cf. Figura 41)



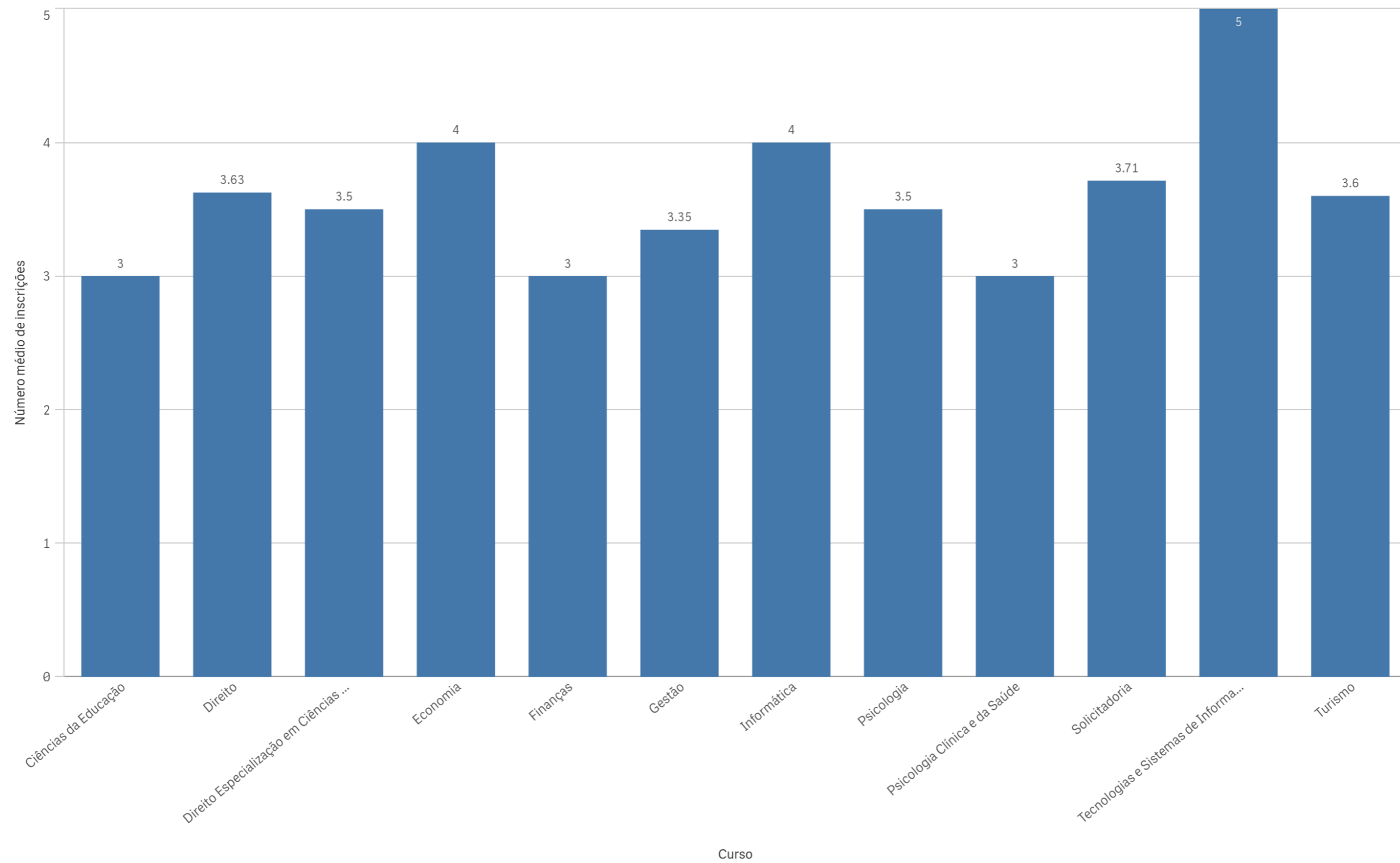
6.6 Segmento de alunos por género e nota de ingresso, relacionados com a nota final de curso (cf. Figura 45)



6.7 Número de alunos por idade (cf. Figura 48)



6.8 Número médio de inscrições até ao final do curso por aluno (cf. Figura 58)



6.9 Dashboards (cf. Figura 65)

