



INSTITUTO  
UNIVERSITÁRIO  
DE LISBOA

---

Análise de Dados para a deteção de Estratégias de Sucesso na Educação

João Tomás Ferreira Guerra

Mestrado em Informática e Gestão

Orientador:

Doutor Luís Miguel Martins Nunes, Professor Associado  
ISCTE-IUL

Co-Orientadora:

Doutora Elsa Alexandra Cabral da Rocha Cardoso, Professora Auxiliar  
ISCTE-IUL

Outubro, 2023





TECNOLOGIAS  
E ARQUITETURA

---

Escola de Tecnologias e Arquitetura

Análise de Dados para a deteção de Estratégias de Sucesso na Educação

João Tomás Ferreira Guerra

Mestrado em Informática e Gestão

Orientador:

Doutor Luís Miguel Martins Nunes, Professor Associado  
ISCTE-IUL

Co-Orientadora:

Doutora Elsa Alexandra Cabral da Rocha Cardoso, Professora Auxiliar  
ISCTE-IUL

Outubro, 2023



## **Agradecimentos**

Gostaria de expressar os meus agradecimentos a todos os que me apoiaram a realizar este trabalho de investigação. Este projeto representa um grande passo na minha jornada académica e não teria sido possível sem o apoio de todos os envolvidos.

Em primeiro lugar à minha família, a quem devo tudo, agradeço por todo o apoio incondicional, amor e compreensão durante este longo percurso. Valorizo o facto de permanecerem sempre do meu lado e acima de tudo encorajarem-me a enfrentar os desafios e a saber lidar com eles para que me torne mais forte no futuro.

Agradeço profundamente aos meus orientadores, professor Luís e professora Elsa, por toda a dedicação e disponibilidade para me ajudar em qualquer altura, especialmente nos momentos mais complicados deste projeto. Pelo incentivo dado de modo a alcançar os meus objetivos académicos e por todo o acompanhamento durante este longo período.

Um reconhecimento à FCT - Fundação para a Ciência e Tecnologia, I.P. pelo parcial financiamento deste trabalho no âmbito dos projetos UIDB/04466/2020 e UIDP/04466/2020.

Agradeço também à DGEEC por permitir e facilitar o acesso aos dados utilizados nesta investigação, fornecendo desde sempre um ambiente colaborativo e confortável para o bom desenvolvimento da minha pesquisa. Um especial reconhecimento à Dra. Susana Fernandes, que desempenhou um papel crucial na disponibilização dos dados da DGEEC. O seu apoio consistente em ajudar-me a compreender os dados, bem como a partilha de novos conjuntos de dados, sempre que necessário, foram elementos-chave para o sucesso deste trabalho.

Por fim, um obrigado aos meus colegas de curso. Enquanto atravessávamos juntos esta etapa das nossas vidas, sempre se mostraram solidários, garantindo que ninguém ficasse para trás.



## Resumo

A educação é um dos fatores determinantes para o desenvolvimento de um país. O nível secundário, o último da educação obrigatória em Portugal, é fundamental para a preparação para a universidade ou para o mundo do trabalho. Para avaliar se este nível está a cumprir o seu propósito é essencial analisar vários fatores que podem ser relevantes como indicadores do seu sucesso. Um indicador de sucesso educacional consegue descrever qualitativa ou quantitativamente o sucesso de uma escola, pelo que a presença de constantes diferenças e variações nas medidas de sucesso demonstradas pelas mesmas ao longo dos anos, tem levado à necessidade de perceber e analisar quais desses indicadores influenciam a performance educacional de uma instituição. A presente dissertação, baseia-se na metodologia CRISP-DM, e apresenta uma análise multinível de dados educacionais de escolas do ensino secundário português com o objetivo de analisar as características e tendências dos indicadores aconselhados na literatura. Posteriormente, são analisadas as relações dos indicadores com as medidas de sucesso na educação, no sentido de descobrir qual a influência de cada um no desempenho escolar no ensino secundário.

**Palavras-Chave:** Ensino secundário, sucesso escolar, indicadores de sucesso, estratégias de sucesso, dados educacionais.



## **Abstract**

Education is one of the key factors in a country's development. Secondary education, the last level of mandatory education in Portugal, is fundamental for preparing for university or the world of work. In order to assess whether this level is fulfilling its purpose, it is essential to analyse several factors that may be relevant as indicators of its success. An educational success indicator can describe qualitatively or quantitatively the success of a school, and the presence of constant differences and performance variables demonstrated by them over the years has led to the need to understand and analyse which of these indicators influence an institution's educational performance. This dissertation is based on the CRISP-DM methodology and presents a multilevel analysis of educational data from Portuguese secondary schools with the aim of analysing the characteristics and trends of the indicators recommended in the literature. Afterwards, the relationship between the indicators and the measures of success in education is analysed to discover the influence of each one on school performance in secondary education.

**Keywords:** Secondary education, school success, success indicators, success strategies, educational data.



# Índice

1. Introdução .....	1
1.1. Motivação .....	1
1.2. Descrição do problema .....	2
1.3. Contexto Teórico.....	2
1.4. Objetivos de Investigação.....	3
1.5. Perguntas de investigação.....	3
1.6. Abordagem metodológica.....	4
1.7. Estrutura da dissertação .....	5
2. Revisão de Literatura.....	7
2.1. Revisão Sistemática de literatura PRISMA.....	7
2.2. Identificação de palavras-chave e expressões de pesquisa .....	7
2.3. Seleção de estudos.....	8
2.4. Resultados de revisão de literatura.....	9
2.4.1 Análise de revisão .....	9
2.4.2 Indicadores de sucesso no ensino secundário .....	10
2.4.3 Identificação das fontes de dados .....	13
3. Compreensão e preparação de dados.....	15
3.1 Seleção dos indicadores para investigação .....	15
3.2 Pedido e extração de dados.....	17
3.3 Descrição de dados disponibilizados.....	18
3.4 Preparação de dados .....	19
3.5 Análise descritiva dos dados.....	22
4. Análise exploratória.....	27
4.1 Indicadores calculados .....	27
4.2 Análise de medidas de sucesso .....	28
4.3 Análise de Indicadores por escola.....	34
4.3 Análise Indicadores por curso.....	48
4.4 Análise Indicadores por turma .....	51
4.5 Resumo dos resultados .....	55
5. Conclusão.....	57
5.1 Limitações de pesquisa e trabalho futuro .....	59
Referências Bibliográficas .....	62
Apêndice A – Estudos revistos na Revisão de Literatura.....	63
Apêndice B – Tabelas de dados disponibilizadas .....	65

Apêndice C – Medidas e Indicadores Calculados .....	69
Apêndice D – Gráficos Indicador Tamanho da Escola .....	79
Apêndice E – Gráficos Indicador Tamanho da Equipa Docente .....	81
Apêndice F – Gráficos Indicador Média de Alunos por Turma .....	8
Apêndice G – Gráficos Indicador Número de cursos disponibilizados .....	83
Apêndice H – Gráficos Indicador Matrículas Anuladas.....	84
Apêndice I – Gráficos Indicador Número de Alunos Transferidos .....	85
Apêndice J – Gráficos Indicador Fatores Socioeconómicos dos alunos .....	86
Apêndice K – Gráficos Indicador Nível médio de qualificação dos professores .....	87
Apêndice L – Gráficos Indicador Média de anos de experiências dos professores .....	88

## Lista de Tabelas

Tabela 1. Descrição de palavras-chave e expressões de pesquisa.....	7
Tabela 2. Critérios de inclusão e de exclusão. ....	8
Tabela 3. Seleção de filtros.....	8
Tabela 4. Estudos com maior importância e os seus contributos.....	10
Tabela 5. Indicadores de sucesso investigados por cada estudo. ....	11
Tabela 6. Fontes de dados de cada estudo selecionado. ....	14
Tabela 7. Medidas de Sucesso selecionadas para estudo. ....	15
Tabela 8. Indicadores de sucesso sugeridos e selecionados para estudo.....	16
Tabela 9. Campos excluídos de cada tabela disponibilizada. ....	19
Tabela 10. Número de turmas por Curso e Ano de Ensino. ....	23
Tabela 11. Medidas de sucesso calculadas. ....	27
Tabela 12. Indicadores de sucesso calculados ....	28
Tabela 13. Correlação do Indicador Tipo de escola com todas as medidas de sucesso .....	35
Tabela 14. Correlação do Indicador Tamanho da escola com todas as medidas de sucesso. ...	37
Tabela 15. Correlação do Indicador Tamanho da equipa docente com todas as medidas de sucesso. ....	38
Tabela 16. Correlação do Indicador Rácio tamanho da escola face ao tamanho da equipa docente com todas as medidas de sucesso. ....	39
Tabela 17. Correlação do Indicador Média do número de alunos com todas as medidas de sucesso. ....	40
Tabela 18. Correlação do Indicador Número de Cursos disponibilizados com todas as medidas de sucesso.....	42
Tabela 19. Correlação do Indicador Número de matrículas anuladas por escola com todas as medidas de sucesso.....	43
Tabela 20. Correlação do Indicador Número alunos transferidos por escola com todas as medidas de sucesso.....	43
Tabela 21. Correlação do Indicador Média dos Fatores Socioeconómicos dos estudantes por escola com todas as medidas de sucesso. ....	44
Tabela 22. Correlação do Indicador nível de qualificação média dos professores por escola com todas as medidas de sucesso. ....	46
Tabela 23. Correlação do Indicador Média de anos de experiência dos professores com todas as medidas de sucesso.....	47
Tabela 24. Correlação do Indicador Tamanho do Curso com as medidas de sucesso em análise ao nível de detalhe por curso.....	50
Tabela 25. Correlação do Indicador Número de matrículas anuladas com as medidas de sucesso em análise ao nível de detalhe por curso.....	51
Tabela 26. Correlação do Indicador Número de alunos transferidos com as medidas de sucesso em análise ao nível de detalhe por curso.....	51
Tabela 27. Correlação do Tamanho da turma com as medidas de sucesso em análise ao nível de detalhe por curso.....	53
Tabela 28. Correlação do Número de matrículas anuladas com as medidas de sucesso em análise ao nível de detalhe por curso.....	54
Tabela 29. . Correlação do Número de alunos transferidos com as medidas de sucesso em análise ao nível de detalhe por curso.....	54

## Lista de Figuras

Figura 1. Metodologia CRISP-DM Adaptado de CRISP-DM Methodology (Chapman et al., 2000).	4
Figura 2. Fases de seleção de estudos. Adaptado de PRISMA <i>methodology flowchart</i> (Page et al., 2021).	9
Figura 3. Número de estudos selecionados por ano de publicação.	9
Figura 4. Processo de preparação de dados.	18
Figura 5. Tabelas disponibilizadas pela DGEEC.	18
Figura 6. Número de escolas por Natureza e Ano Letivo.	22
Figura 7. Número de Alunos por Ano de Ensino e Ano Letivo.	23
Figura 8. Número de Professores por Ano Letivo.	24
Figura 9. Número de Exames por Ano de Ensino e Ano Letivo.	24
Figura 10. Número de alunos por Escalão ASE e Ano Letivo.	25
Figura 11. Médias de Exames de 1ª e 2ª fase por Ano Letivo (escala 0- 200).	29
Figura 12. Médias de Exames de 1ª e 2ª Fase por Curso e Ano Letivo (escala 0- 200).	29
Figura 13. Média Final por Ano Letivo (escala 0-200).	30
Figura 14. Média Final por Curso e Ano Letivo (escala 0-200).	31
Figura 15. Percentagem de alunos presentes em exames de segunda fase.	31
Figura 16. Percentagem de alunos presentes em segunda fase por Curso e Ano Letivo.	32
Figura 17. Percentagem de alunos por Ano Letivo.	32
Figura 18. Percentagem de alunos reprovados por Curso e Ano Letivo.	33
Figura 19. Número de escolas por categoria de alunos desistentes e por Ano Letivo.	34
Figura 20. <i>Heat map</i> de correlações entre as medidas de sucesso ao nível de detalhe escolar.	34
Figura 21. Médias de Exames de 1ª Fase pelas Médias de Exames de 2ª Fase por escola (escala 0-200).	35
Figura 22. Tipo de escola pelas Médias Finais (escala 0-200).	36
Figura 23. Tipo de escola pelas Médias de Exames de 1ª Fase (escala 0-200).	36
Figura 24. Tipo de escola pelas Médias de Exames de 2ª Fase (escala 0-200).	36
Figura 25. Tipo de escola pela Percentagem de Alunos em 2ª Fase.	37
Figura 26. Tamanho médio das escolas pela percentagem média de alunos reprovados ao longo dos 6 anos letivos.	37
Figura 27. Tamanho médio da equipa docente pela média de alunos reprovados (%), ao longo dos 6 anos letivos.	38
Figura 28. Tamanho médio da equipa docente pela Média de Exames de 2ª Fase, ao longo dos 6 anos letivos (escala 0-200).	38
Figura 29. Rácio do tamanho da escola face ao tamanho da equipa docente por percentagem de alunos reprovados, ao longo dos 6 anos letivos.	40
Figura 30. Rácio do tamanho da escola face ao tamanho da equipa docente por percentagem de alunos reprovados, ao longo dos 6 anos letivos.	40
Figura 31. Média de Alunos por Turma pelas Médias de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).	41
Figura 32. Média de Alunos por Turma pelas Médias de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).	41
Figura 33. Média de Alunos por Turma pelas Médias Finais por escola, ao longo dos 6 anos letivos (escala 0-200).	41
Figura 34. Média de Alunos por Turma pela percentagem de alunos reprovados, ao longo dos 6 anos letivos.	41

Figura 35. Número de cursos disponibilizados pela percentagem de alunos reprovados por escola. ....	42
Figura 36. Média de Escalões ASE pelas Médias de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).....	45
Figura 37. Média de Escalões ASE pelas Médias de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).....	45
Figura 38. Média de Escalões ASE pelas Médias Finais por escola, ao longo dos 6 anos letivos (escala 0-200).....	45
Figura 39. Média de Escalões ASE pela percentagem de alunos reprovados por escola, ao longo dos 6 anos letivos.....	45
Figura 40. Qualificação Média dos Professores pela percentagem de Alunos Reprovados por escola, ao longo dos 6 anos letivos.....	46
Figura 41. Qualificação Média dos Professores pela Média de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200). ....	46
Figura 42. Qualificação Média dos Professores pela Média de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos. ....	46
Figura 43. Anos de Experiência Média dos Professores pelas Médias de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).....	47
Figura 44. Anos de Experiência Média dos Professores pelas Médias de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).....	47
Figura 45. Anos de Experiência Média dos Professores pelas Médias Finais por escola, ao longo dos 6 anos letivos (escala 0-200).....	48
Figura 46. Anos de Experiência Média dos Professores pela percentagem de Alunos reprovados por escola, ao longo dos 6 anos letivos .....	48
Figura 47. <i>Heat map</i> de correlações entre as medidas de sucesso ao nível de detalhe por curso.....	48
Figura 48. Média de Exames de 1ª Fase pelas Médias Finais do curso de Artes Visuais, ao longo dos 6 anos letivos (escala 0-200).....	49
Figura 49. Média de Exames de 1ª Fase pelas Médias Finais do curso de Ciências e Tecnologias, ao longo dos 6 anos letivos (escala 0-200).....	49
Figura 50. Média de Exames de 1ª Fase pelas Médias Finais do curso de Ciências Socioeconómicas, ao longo dos 6 anos letivos (escala 0-200). ....	49
Figura 51. Média de Exames de 1ª Fase pelas Médias Finais do curso de Línguas e Humanidades, ao longo dos 6 anos letivos (escala 0-200). ....	49
Figura 52. Tamanho dos cursos de Artes Visuais por escola pela percentagem de alunos reprovados, ao longo dos 6 anos letivos. ....	50
Figura 53. Tamanho dos cursos de Ciências e Tecnologias por escola pela percentagem de alunos reprovados, ao longo dos 6 anos letivos. ....	50
Figura 54. Tamanho dos cursos de Ciências Socioeconómicas por escola pela percentagem de alunos, ao longo dos 6 anos letivos. ....	50
Figura 55. Tamanho dos cursos de Línguas e Humanidades por escola pela percentagem de alunos reprovados, ao longo dos 6 anos letivos. ....	50
Figura 56. <i>Heat map</i> de correlações entre as medidas de sucesso ao nível de detalhe por turma.....	51
Figura 57. Média de Exames de 1ª Fase pelas Médias Finais das turmas de 10º ano, ao longo dos 6 anos letivos (escala 0-200).....	52
Figura 58. Média de Exames de 1ª Fase pelas Médias Finais das turmas de 11º ano, ao longo dos 6 anos letivos (escala 0-200).....	52
Figura 59. Média de Exames de 1ª Fase pelas Médias Finais das turmas de 12º ano, ao longo dos 6 anos letivos (escala 0-200).....	52

Figura 60. Tamanho da turmas pelas Médias de Exames de 1ª Fase das turmas de 10º ano, ao longo dos 6 anos letivos (escala 0-200).....	53
Figura 61. Tamanho da turmas pelas Médias de Exames de 1ª Fase das turmas de 11º ano, ao longo dos 6 anos letivos (escala 0-200).....	53
Figura 62. Tamanho das turmas pelas Médias de Exames de 1ª Fase das turmas de 12º ano, ao longo dos 6 anos letivos (escala 0-200).....	53
Figura 63. Tamanho da turmas pela Percentagem de alunos reprovados das turmas de 10º ano, ao longo dos 6 anos letivos. ....	54
Figura 64. Tamanho da turmas pela Percentagem de alunos reprovados das turmas de 11º ano, ao longo dos 6 anos letivos. ....	54
Figura 65. Tamanho da turmas pela Percentagem de alunos reprovados das turmas de 12º ano, ao longo dos 6 anos letivos.. ....	54

## Lista de Siglas

CRISP-DM	<i>CRoss Industry Standard Process for Data Mining</i>
OCDE	Organização para a Cooperação e Desenvolvimento Económico
ODS	Objetivos de Desenvolvimento Sustentável
PRISMA	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
RSL	Revisão Sistemática de Literatura
PISA	<i>Programme for International Student Assessment</i>
DGEEC	Direção-Geral de Estatísticas da Educação Ciência
IA>AP	Inteligência Artificial para a Administração Pública
CFD	Classificação Final à Disciplina
ASE	Ação Social Escolar



# CAPÍTULO

## 1. Introdução

O rápido desenvolvimento e variedade das tecnologias de informação nos dias de hoje impactou um aumento significativo na quantidade e complexidade dos dados disponíveis, proporcionando o importante desafio de saber como fazer a sua gestão e análise de forma correta, para que seja possível efetuar uma tomada de decisão dos processos pretendidos (Şen et al., 2012). As instituições educativas têm uma grande presença neste desafio, pertencentes a uma indústria que lida com um exponencial crescimento do volume de dados, capazes de fazer assumir diferentes objetivos e posições no que toca à sua análise sectorial (Aruna & Priya, 2021). A análise de dados educativos pode desempenhar um papel muito importante quando se trata de investigar resultados, comportamentos e padrões dos estudantes ou professores académicos (Govindasamy & Velmurugan, 2019).

De acordo com a Organização para a Cooperação e Desenvolvimento Económico (OCDE), o nível educacional secundário em Portugal é frequentemente visto como uma qualificação mínima para ingressar com sucesso no mercado de trabalho ou no ensino superior, no entanto 17% dos jovens adultos não chegaram a atingir esse nível de educação, um valor 3% mais elevado que a média verificada pela OCDE (Markus, 2022). Números como estes tornam necessária a discussão sobre sucesso escolar no setor secundário em Portugal. Trata-se de um tópico multicausal, complexo de analisar, que apresenta múltiplos indicadores indireta e diretamente relacionados. A identificação desses indicadores é uma tarefa exigente devido à sua diversidade (Martínez Abad & Chaparro Caso López, 2017), o que tem vindo a implicar um aumento significativo no número de investigações sobre o tópico, que apresentam uma grande importância para a indústria, com o intuito de evoluir o setor educativo e também para elevar o progresso dos estudantes (Aruna & Priya, 2021).

Indicadores de sucesso no setor secundário não provêm apenas das competências dos estudantes, estando presentes múltiplos outros fatores externos que garantem a necessidade e importância da investigação de indicadores que justificam um desempenho educacional positivo (Leiber, 2019).

### 1.1. Motivação

A deteção de estratégias de sucesso no setor secundário da indústria educacional tem sido um tópico discutido e investigado não só para obter melhores formandos nas mais variadas áreas oferecidas, mas também para tornar possível o desenvolvimento da educação que apresenta um moderado avanço tendo como comparação outras indústrias. Desde o início do primeiro momento de aprendizagem ou avaliação escolar que os alunos e as suas instituições são expostos aos impactos dos indicadores de desempenho educacionais, numa altura em que os mesmos indivíduos estão focados em alcançar bons resultados e ingressar em boas universidades ou empregos. Estes indicadores são dependentes de uma grande variedade de dados com altos níveis de complexidade, que podem influenciar todo o sucesso de uma escola, exigindo uma análise multivariada profunda com diferentes abordagens.

Atualmente, devido à disponibilidade pública para algumas das bases de dados escolares, existem alternativas para analisar essa eficácia educativa e perceber as possíveis razões para o sucesso ou insucesso de uma escola. Compreender e melhorar o sucesso escolar ajuda não só a promover a igualdade das oportunidades educacionais como também a reduzir disparidades e assegurar que a educação seja verdadeiramente inclusiva e de alta qualidade para todos, alinhando-se assim com os Objetivos de Desenvolvimento Sustentável (ODS), nomeadamente o ODS 4, "Educação de Qualidade".

## **1.2. Descrição do problema**

Segundo Silva, L (2020), até os anos 60 do século XX, era aceite o conceito de que as escolas não tinham nenhuma influência no sucesso educacional dos alunos. Acreditava-se que o desempenho dos alunos era apenas determinado pelo seu ambiente pessoal e pelas suas habilidades. O que consta atualmente é que todos os comportamentos dos estudantes são diferentes e que múltiplas escolas têm, portanto, as suas políticas e métodos de ensino, onde estão presentes indicadores escolares que fazem variar para além do desempenho dos seus alunos, os resultados e as suas qualidades enquanto instituição, ao longo dos anos. O sucesso escolar em Portugal é uma área que apresenta uma necessidade de investigação com o intuito de perceber o que torna uma escola eficaz e o que tem vindo a afetar o sucesso da mesma. Com especial enfoque no ensino secundário, onde se preparam os alunos nos últimos três anos de ensino obrigatório, é importante perceber que indicadores estão associados à variação do desempenho e através da análise dos mesmos tentar detetar estratégias de sucesso.

## **1.3. Contexto Teórico**

O conceito de sucesso esteve desde sempre associado à indústria da educação, no entanto, conseguir definir o termo é uma tarefa exigente à qual muitos investigadores têm vindo a tentar cumprir devido à sua complexidade e dependência para inúmeros fatores (Gonzaga & Miguel, 2019). Apesar da adversidade, segundo Perrenoud P. (2003), o sucesso escolar relaciona dois tópicos fundamentais: o desempenho de um aluno e o desempenho de uma escola ou do seu sistema escolar. Num primeiro ponto, compreende-se sucesso de um aluno quando, num modo geral, num ambiente de avaliação das aprendizagens adquiridas, o desempenho e o envolvimento nas tarefas educativas levam a que os resultados obtidos sejam atingidos ou a superarem os que anteriormente foram planeados para o mesmo (Costa, 2013). No que toca ao desempenho de uma escola, uma boa instituição apresenta sucesso quando atingem os seus objetivos, asseguram condições de trabalho e garantem também o êxito dos seus estudantes.

No entanto, o sucesso pode tomar várias perspetivas dependendo do nível de escolaridade em questão. Um estudante do ensino secundário apresenta diferentes objetivos escolares dos estudantes do ensino primário, cujo foco principal é na aquisição dos conhecimentos e competências básicas, ou de um estudante do ensino superior, cujo propósito é especializar-se numa área profissional específica.

O sucesso, portanto, deve ser assumido e investigado de acordo com os fatores distintos que o nível da educação impõe para os indivíduos que nele atuam.

No ensino secundário, onde os estudantes começam a planear o seu futuro e descobrir as suas vocações profissionais, o empenho e o sucesso educativo ganham importância, validando ainda mais a atenção que os determinantes de performance exteriores às competências dos alunos requerem (Costa, 2013).

### **Avaliação do ensino secundário em Portugal**

Um estudante que frequenta o ensino secundário em Portugal apresenta-se nos últimos três dos 12 anos do ensino básico e obrigatório português, que dão acesso ao ensino superior em função dos seus resultados durante esses anos. O processo de avaliação de um estudante do ensino secundário em Portugal depende de dois fatores quantitativos:

**Avaliação escolar interna:** A avaliação escolar interna é calculada através da soma de todos os resultados obtidos nas disciplinas durante os três anos do ensino secundário, dividida pelo número total de disciplinas estudadas.

**Avaliação escolar externa:** A avaliação escolar externa do ensino secundário é caracterizada por um único exame final por disciplina, no final de cada ano de ensino. Estes exames eram anteriormente obrigatórios para as disciplinas que se centram na área de competências inscritas no início do primeiro ano, tendo um peso de 30% nas notas finais das mesmas, mas que de momento (outubro de 2023) são opcionais, deixando de ter qualquer peso para o resultado final à disciplina.

## **1.4. Objetivos de Investigação**

Como principal objetivo desta investigação pretende-se calcular o máximo de indicadores e medidas de sucesso possíveis e detetar as relações e o peso de influência de cada indicador no sucesso educacional do setor secundário em Portugal. Procura-se colecionar dados de várias escolas secundárias ao longo do seu tempo de ensino, no sentido de calcular o máximo de indicadores e medidas de sucesso possíveis. Os indicadores e medidas de sucesso a usar neste projeto de investigação são selecionados e revistos através da revisão de literatura, onde se pretende integrar os mais estudados nos últimos anos. Posteriormente, tendo como base os indicadores que foram possíveis de calcular, pretende-se analisar as características e as tendências de cada um, comparando cada indicador de modo a tentar detetar as relações e as suas influências no sucesso das escolas do setor secundário em Portugal.

## **1.5. Perguntas de investigação**

Este estudo irá tentar responder às seguintes perguntas de investigação:

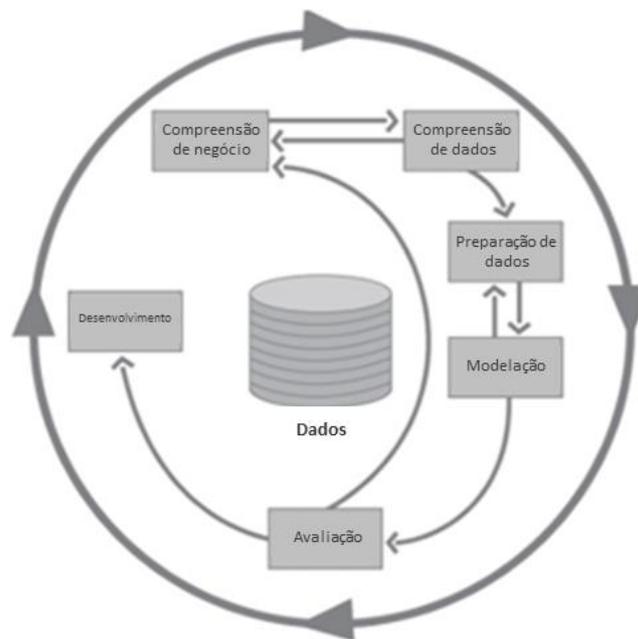
PI1: De acordo com os indicadores sugeridos na revisão de literatura, e com base nos dados educativos disponibilizados, que indicadores de sucesso escolar é possível calcular?

PI2: Quais as características e tendências das medidas de sucesso calculadas ao longo dos anos letivos?

PI3: Que indicadores educacionais estabelecem relação com as medidas de sucesso e em que medida influenciam o desempenho escolar no ensino secundário?

## 1.6. Abordagem metodológica

Na presente investigação, a abordagem metodológica utilizada é designada por “*CRoss Industry Standard Process for Data Mining*” (CRISP-DM). Esta metodologia, como é possível verificar através da Figura 1, está dividida em seis principais etapas: compreensão de negócios, compreensão de dados, preparação de dados, modelação, avaliação e desenvolvimento. Cada uma das fases mencionadas é composta por várias tarefas que podem estar ou não relacionadas entre si e encontram-se interligadas com uma estrutura sequencial e iterativa. No entanto, este tipo de metodologia tem flexibilidade para que a execução de algumas tarefas possa ser realizada numa ordem distinta.



**Figura 1.** Metodologia CRISP-DM

Adaptado de CRISP-DM Methodology (Chapman et al., 2000).

Adaptando a metodologia à presente dissertação, a etapa de compreensão de negócios corresponde aos primeiros dois capítulos de introdução e de revisão de literatura, onde é elaborada uma análise e compreensão da investigação e do problema associado à mesma. Conta com a presença ainda dos objetivos a cumprir, os seus requisitos e as perguntas de investigação a responder num último ponto.

Uma vez obtidos os dados para estudo, as etapas de compreensão e preparação dos mesmos fazem correspondência ao terceiro capítulo designado pelo nome das tarefas em questão, isto é, compreensão e preparação dos dados. Neste segmento pretende-se obter um foco principal nos dados a utilizar para garantir a sua fiabilidade e qualidade, proceder à sua organização, limpeza e formatação para poder adotar diferentes abordagens numa fase seguinte.

As fases de modelação e de avaliação são incorporada no capítulo quatro, onde são feitas análises e aplicados modelos para completar o objetivo delineado e resumir os resultados obtidos através de cada análise realizada.

A etapa de desenvolvimento está incorporada no capítulo de conclusão deste estudo, onde são mencionados os resultados obtidos e mencionadas as respostas às perguntas de investigação.

## **1.7. Estrutura da dissertação**

O presente estudo encontra-se estruturado em cinco capítulos que descrevem as diferentes etapas realizadas até à conclusão do mesmo.

Num primeiro capítulo é feita uma introdução ao tema principal de investigação, onde são descritos os principais objetivos e a abordagem metodológica utilizada.

O capítulo dois é dedicado à revisão de literatura, onde se pretende analisar estudos e obter um enquadramento para a presente investigação.

O terceiro capítulo destina-se à recolha, compreensão e tratamento de dados a utilizar onde são descritos os dados recolhidos, referenciadas as fontes dos mesmos e especificados os significados de cada um;

Num quarto capítulo é efetuada a modelação e análise relacional dos dados e indicadores recolhidos com recurso a várias técnicas, fazendo também um resumo de todos os resultados obtidos.

O quinto e último capítulo apresenta os resultados de análise obtidos e as conclusões da presente pesquisa, respondendo a cada uma das perguntas de investigação. São ainda feitas referências às limitações verificadas no estudo, bem como possíveis trabalhos futuros a desenvolver.

No fim do documento é possível verificar ainda todas as referências servidas como base de conhecimento para a investigação em causa, bem como os apêndices.



## CAPÍTULO

### 2. Revisão de Literatura

O capítulo de revisão de literatura visa identificar várias perspetivas sobre o tema especificado nas seções anteriores e pretende-se rever quais os indicadores de desempenho mais estudados, que apresentam influência no sucesso académico das escolas que prestam o ensino secundário. Como metodologia, segue uma Revisão Sistemática de Literatura (RSL) de acordo com as regras baseadas em “*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*” (PRISMA) (Page et al., 2021), com o objetivo de encontrar e analisar todas as informações e indicadores necessários para tomar uma posição apropriada e fornecer as respostas às perguntas de investigação anteriormente mencionadas. Através deste tipo de metodologia, é possível verificar uma estratégia explícita de investigação baseada em literatura científica.

#### 2.1. Revisão Sistemática de literatura PRISMA

O PRISMA é um método que facilita a recolha de revisões bibliográficas a utilizar como base para revisões sistemáticas por outros investigadores. Este método também tem alguns benefícios: permite a replicação de métodos de revisão e fornece uma estrutura de revisão com procedimentos específicos e pré-definidos (Page et al., 2021).

#### 2.2. Identificação de palavras-chave e expressões de pesquisa

Nesta secção, sendo a primeira fase da metodologia PRISMA, são especificadas as palavras-chave utilizadas para a pesquisa de artigos científicos (Tabela 1). Através das palavras-chave e com o apoio dos conectores "AND" e "OR" foi possível criar duas expressões de pesquisa para selecionar os artigos relevantes que levariam por sua vez à recolha de indicadores de sucesso (Tabela 1).

Tanto as palavras-chave como as expressões de pesquisa foram selecionadas de modo a obter estudos relativos à análise de sucesso no ensino secundário e os seus indicadores, especificando a palavra “PISA” (*Programme for International Student Assessment*) que se trata de um programa desenvolvido pela OCDE de modo a avaliar as competências de leitura, matemática e ciências de alunos de 80 países diferentes, que se encontram no fim da sua escolaridade obrigatória.

**Tabela 1.** Descrição de palavras-chave e expressões de pesquisa.

<b>Palavras-Chave</b>	<i>Secondary schools, academic performance, success factors, multilevel analysis, educational data, PISA.</i>
<b>Expressão de pesquisa 1</b>	<i>("high school" OR "secondary school" OR "secondary education") AND ("academic success" OR "academic performance" OR "academic achievement") AND ("factors" OR "determinants" OR "effects") AND "multilevel analysis" AND "educational data"</i>
<b>Expressão de pesquisa 2</b>	<i>("school success" OR "school achievement") AND "student performance" AND "PISA" AND "multilevel analysis" AND ("factors" OR "determinants" OR "effects")</i>

### 2.3. Seleção de estudos

Numa segunda etapa da metodologia PRISMA, todo o processo de seleção de estudos é efetuado. A base de dados utilizada para a recolha de artigos científicos é a Scopus que através do seu processo de filtragem, foi possível definir os artigos mais relevantes e apropriados para o tópico desta investigação. Posteriormente, foram estabelecidos os critérios de inclusão e de exclusão (Tabela 2) cujos seus principais objetivos são assegurar que os conhecimentos adquiridos tenham um conteúdo mais recente e que sejam também redigidos numa linguagem acessível.

**Tabela 2.** Critérios de inclusão e de exclusão.

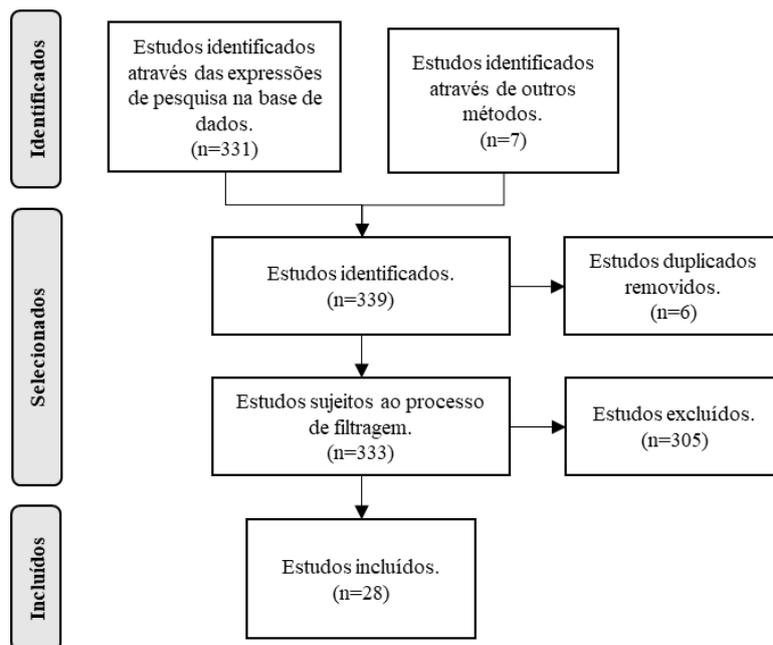
<b>Critérios de inclusão</b>	<b>Critérios de exclusão</b>
Artigo escrito em inglês ou português.	Artigo escritos num outro idioma.
Data de publicação entre 2010-2022.	Data de publicação mais antiga que 2010.

No que toca à filtragem dos estudos a analisar, foram criados dois tipos de filtros para restringir os artigos devolvidos pela base de dados Scopus (Tabela 3). Primeiramente, foram feitas pesquisas diferentes para verificar a presença de cada uma das expressões de pesquisa em qualquer parte de um artigo. Apenas os artigos que continham a expressão de pesquisa indicada foram devolvidos. Em segundo lugar, foi aplicado o primeiro filtro para restringir o número de estudos, onde apenas foram indicados os artigos que preenchiam os critérios de inclusão. Por fim, a secção onde se presencia o resumo de cada artigo foi lida com o intuito de verificar a sua relevância para o tópico pretendido desta investigação.

**Tabela 3.** Seleção de filtros.

<b>Filtro</b>	<b>Descrição</b>
Sem filtros	Verificação da presença da expressão de pesquisa em qualquer parte do artigo.
Filtro #1	Aplicação dos critérios de inclusão e de exclusão.
Filtro #2	Leitura do resumo ( <i>abstract</i> ) de cada artigo de modo a garantir relevância para a investigação.

A recolha dos artigos foi garantida através da ferramenta Mendeley Reference Manager para extrair, organizar e fornecer informações sobre os estudos, tais como o nome do autor, tipo de estudo, ano de publicação, fontes, referências, entre outros. Esta ferramenta também ajudou no processo de referenciamento de todos os estudos utilizados como base de conhecimento desta investigação. Através da Figura 2, é possível verificar as fases de seleção dos estudos científicos da referida base de dados onde foram identificados e removidos estudos duplicados, aplicados todos os filtros e elegidos por fim para a possibilidade de recolha de informação. Para complementar a revisão bibliográfica, a partir de conselhos dos professores orientadores, foi possível adquirir outros artigos relevantes para a investigação, identificados na Figura 2 como “outros métodos”.



**Figura 2.** Fases de seleção de estudos.  
Adaptado de PRISMA *methodology flowchart* (Page et al., 2021).

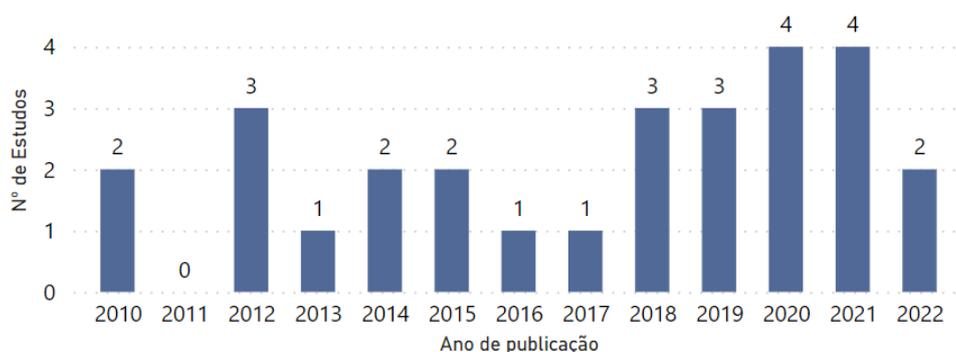
## 2.4. Resultados de revisão de literatura

Numa última etapa da metodologia PRISMA, procede-se à apresentação dos resultados obtidos através da revisão dos artigos científicos selecionados. Primeiramente uma análise aos artigos é facultada e após tal, são enumerados os indicadores de sucesso mais estudados tendo como base toda a informação obtida durante o processo de revisão de literatura.

### 2.4.1 Análise de revisão

Uma vez colecionados todos os estudos servidos como base de conhecimento para a presente investigação, um total de 28 artigos científicos apresentam-se identificados e listados no Apêndice A, designado por “Estudos revistos na Revisão de Literatura”.

Através do ano de publicação de cada um dos artigos científicos e uma vez todos registados, é possível indicar que os anos 2020 e 2021 foram os que apresentaram mais estudos elaborados (4) ao contrário do ano de 2011 que não obteve qualquer registo, de todos os incluídos (Figura 3).



**Figura 3.** Número de estudos selecionados por ano de publicação.

De forma a distinguir os estudos que apresentam maior relevância, através da Tabela 4 é possível notar algumas informações sobre os mesmos, no que toca aos seus contributos e resultados a considerar durante a realização deste projeto de investigação, bem como os rankings enquanto artigos científicos, disponibilizados através da plataforma Scimago.

**Tabela 4.** Estudos com maior importância e os seus contributos.

<b>ID</b>	<b>Referência</b>	<b>Ranking</b>	<b>Contributos</b>
1	(Creemers & Kyriakides, 2010)	-	Proporciona uma compreensão do significado de indicadores educacionais e como devem ser investigados.
2	(Marks, 2010)	Q1	Análise dos efeitos escolares no desempenho dos estudantes que acedem ao ensino superior. Referencia a importância de análise dos fatores nos alunos que apresentam piores resultados e não em fatores particulares escolares no seu geral.
15	(Ma et al., 2018)	Q1	Análise relacional entre fatores na área de ciências do ensino secundário. Conclui na sua investigação que a variação do sucesso académico provém aproximadamente de 77% de fatores dos alunos, 8% de fatores das salas de aula, 3% dos professores e 12% de fatores da escola
19	(Martínez-Abad et al., 2020)	Q1	Análise dos principais fatores associados à eficácia escolar no ensino secundário. Seleção das melhores e piores escolas com base nos dados obtidos através do estudo da OCDE: PISA 2015 para identificação dos fatores que mais se relacionam com sucesso e insucesso escolar.
20	(Maheshwari et al., 2020)	Q4	Análise das variáveis que mais afetam o abandono do ensino e formação de um modelo de previsão que permite medir a possibilidade do mesmo acontecimento.

#### **2.4.2 Indicadores de sucesso no ensino secundário**

Analisar indicadores que apresentam peso na variação do sucesso de uma escola do ensino secundário é um desafio que capta o interesse de vários investigadores pela sua complexidade (Şen et al., 2012). Podem ser definidos como dados estatísticos que descrevem e caracterizam variáveis educativas e que permitem a avaliação ou a monitorização de escolas, alunos, professores ou outros aspetos escolares. Tendem a dar informação acerca da qualidade e do sucesso alcançado, podendo por isso ser agrupados em várias dimensões (Creemers & Kyriakides, 2010). Indicadores de performance escolares variam também ao longo dos diferentes países e regiões devido às culturas e oportunidades vividas em cada um deles, pelo que necessitam de ser analisados detalhadamente para que ocorram modificações positivas no âmbito pretendido (Alam et al., 2021).

De acordo com os estudos integrados na revisão bibliográfica efetuada e sendo o principal objetivo agrupar indicadores de performance para sustentar o presente projeto de investigação em escolas secundárias portuguesas, destaca-se a Tabela 5, onde é possível verificar quais os indicadores mais abordados para cada medição de sucesso estudada.

**Tabela 5.** Indicadores de sucesso investigados por cada estudo.

<b>Medidas de sucesso investigadas</b>	<b>Indicadores investigados para cada medição</b>		<b>Estudos</b>
Média de avaliação externa do ensino secundário	Indicadores escolares	Tamanho da escola	9, 11, 12, 13, 15
		Rácio número de alunos/número de professores	9, 11, 13, 15
		Recursos escolares disponibilizados	9, 12, 15
		Tipo de localidade da escola (rural/urbana)	11, 12, 15
		Fatores socioeconômicos da escola	12, 15
	Indicadores de professores	Nível médio de qualificação dos professores	9, 15
		Média de anos de experiência como professor	15
	Indicadores de estudantes	Nível de educação parental	11, 12, 13
Fatores socioeconômicos dos estudantes		15	
Média final do ensino secundário	Indicadores escolares	Recursos escolares disponibilizados	2, 18, 24
		Fatores socioeconômicos da escola	2, 24, 28
		Tipo de localidade da escola (rural/urbana)	2, 24
		Rácio número de alunos / número de professores	18, 24
		Tamanho da escola	18, 24
		Tipo de escola (privada/pública)	4
	Indicadores de estudantes	Fatores socioeconômicos dos estudantes	2, 18, 24, 28
		Nível de educação parental	28
Resultados PISA	Indicadores escolares	Fatores socioeconômicos da escola	2, 3, 8, 10, 16, 19, 27
		Tamanho da escola;	3, 7, 8, 16, 19, 25, 27
		Recursos escolares disponibilizados	2, 3, 5, 16, 25
		Tipo de localidade da escola (rural/urbana)	7, 8, 10, 27
		Rácio número de alunos/número de professores	7, 10, 16, 19
		Média do número de horas de aulas por semana	10, 22
		Tipo de escola (privada/pública)	7, 8, 10
		Número de alunos que mudaram de escola no secundário	16, 19
	Média do número de estudantes que abandonam no ensino secundário	20	
	Indicadores de professores	Nível de qualificação dos professores	19, 27
Indicadores de estudantes		Nível de educação parental	3, 5, 7, 22
	Taxa de abandono escolar	Indicadores escolares	Média do número de estudantes que abandonam no ensino secundário
Rácio número de alunos/número de professores			20

Ao proceder à análise e investigação de sucesso na educação é necessário distinguir como se mede e avalia esse mesmo sucesso. A cada tipo de indicador estudado na revisão de literatura está associado uma medição de desempenho. Segundo Lemos (2013), verificam-se dois tipos de medidas: medidas de aproveitamento e medidas de participação, sendo que cada uma toma e assume princípios distintos. As primeiras medidas dão ênfase ao aproveitamento obtido, medido pelas classificações periódicas, exames, taxas de aprovação e médias finais internas e/ou externas. Por sua vez, as medidas de participação referem-se às taxas de desistência escolar e abandono precoce do ensino, que analisam o número de estudantes que desistem da sua educação face aos que completam todo o seu ensino obrigatório.

Através da revisão efetuada, para a deteção de sucesso numa escola do ensino secundário as medidas de aproveitamento mais estudadas foram a média final do ensino secundário, a média de avaliação externa do ensino secundário e ainda os resultados PISA de cada escola em análise.

Como medidas de participação, o estudo de Maheshwari et al. (2020) é o único que as aplica, investigando o peso de alguns indicadores escolares na medição da taxa de abandono, de forma a avaliar o sucesso das mesmas.

Todas as medidas mencionadas para além de terem a capacidade de distinguir sucesso de insucesso têm o peso de vários indicadores associados, que justificam, ou não, a sua variação:

### **Indicadores dos estudantes**

Quando são referidos indicadores relativos ao aluno, são evidenciados dados que descrevem o mesmo e que podem ser julgados ao ponto de interferirem no desempenho académico do indivíduo. Grande número de estudos relacionam o desempenho escolar com as características de um aluno (Pangeni, 2014), sendo fatores como o seu nível socioeconómico e o nível de educação parental os mais estudados, de todos os artigos analisados para esta investigação.

Uma grande percentagem dos estudos revistos analisam ainda o fator motivacional e de bem-estar de um estudante através da realização de questionários e entrevistas, dada a indisponibilidade de armazenamento desse tipo de valores em bases de dados educativas.

### **Indicadores escolares**

O número de investigações que estudam características escolares como fatores influenciadores de sucesso no ensino secundário têm vindo a apresentar resultados variáveis (Wei et al., 2012). De acordo com Marks (2010), existem três tipos de fatores que uma escola pode apresentar: os seus fatores composicionais, estruturais e ainda as práticas ou processos escolares. Fatores composicionais de uma escola descrevem a mesma através do seu contexto académico, o tipo de escola que se trata (pública/privada), o tipo de ensino que presta, o número de professores face ao número de alunos, entre outros.

Quanto aos fatores estruturais, em vários países, os recursos disponibilizados por uma escola estão associados ao aumento do nível de performance dos seus alunos (Chowa et al., 2015), o seu tamanho, o número de salas de aula que disponibiliza ou até mesmo a posse de uma biblioteca no interior do seu estabelecimento são considerados fatores a ter em conta quando se investiga sucesso no ensino.

Já as suas práticas e processos enquanto escola, são valorizadas as ações que promovem um clima apropriado ao ensino, proporcionam atividades extracurriculares, uma melhor relação entre aluno e professor e partilham valores de sensibilização aos seus alunos (Liu et al., 2015).

### **Indicadores dos professores**

Os próprios professores são também vistos como influentes no sucesso escolar apresentando um grande peso no número de estudos que os tomam como indicadores de sucesso (Ma et al., 2018). Sendo os indivíduos que mais estabelecem contacto com os alunos, são também eles que prestam o maior exercício escolar, a aprendizagem. Alguns investigadores têm atingido resultados e detetado evidências significativas quando analisam a influência de certas características dos professores para o sucesso académico, como as suas qualificações e o número de anos de experiência (Chowa et al., 2015).

#### **2.4.3 Identificação das fontes de dados**

Dado o intervalo temporal entre cada estudo e as diferentes regiões onde os indicadores são investigados, é necessário indicar e perceber de onde vêm todos os dados para a deteção de sucesso escolar em cada um dos casos (Tabela 6). Através desta informação é possível também verificar as fontes de dados usadas em cada um dos trabalhos.

**Tabela 6.** Fontes de dados de cada estudo selecionado.

<b>ID</b>	<b>Referência</b>	<b>Fonte de dados</b>
1	(Creemers & Kyriakides, 2010)	<ul style="list-style-type: none"> <li>• Não utilizou dados provenientes de fontes externas.</li> </ul>
2	(Marks, 2010)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2003.</li> </ul>
3	(Sun et al., 2012)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2006.</li> </ul>
4	(Şen et al., 2012)	<ul style="list-style-type: none"> <li>• Direção Geral de Tecnologias da Educação do Ministério da Educação Nacional na Turquia.</li> </ul>
5	(Wei et al., 2012)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2003.</li> </ul>
6	(Lemos, 2013)	<ul style="list-style-type: none"> <li>• Gabinete de Estatística e Planeamento da Educação (GEPE);</li> <li>• PORDATA.</li> </ul>
7	(Pangeni, 2014)	<ul style="list-style-type: none"> <li>• Questionários.</li> </ul>
8	(Shera, 2014)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2009 e 2012;</li> </ul>
9	(Chowa et al., 2015)	<ul style="list-style-type: none"> <li>• Projeto experimental <i>YouthSave</i> (recolha de dados escolares durante 5 anos).</li> </ul>
10	(Liu et al., 2015)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2003.</li> </ul>
11	(Börkan & Bakış, 2016)	<ul style="list-style-type: none"> <li>• e-School (Sistema de informação de gestão educacional na Turquia)</li> </ul>
12	(Takashiro, 2017)	<ul style="list-style-type: none"> <li>• <i>Trends in International Mathematics and Science Study (TIMSS)</i></li> </ul>
13	(Granvik Saminathen et al., 2018)	<ul style="list-style-type: none"> <li>• Questionários.</li> </ul>
14	(Cristina & Caldeira, 2018)	<ul style="list-style-type: none"> <li>• Entrevistas.</li> </ul>
15	(Ma et al., 2018)	<ul style="list-style-type: none"> <li>• <i>Program for Regional Assessment and Promotion of Basic Education quality (PRABEQ)</i></li> </ul>
16	(Martínez-Abad, 2019)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2015.</li> </ul>
17	(Gonzaga & Miguel, 2019)	<ul style="list-style-type: none"> <li>• Direção-Geral de Estatísticas da Educação Ciência (DGEEC);</li> <li>• PORDATA.</li> </ul>
18	(Karaman & Atar, 2019)	<ul style="list-style-type: none"> <li>• Questionários.</li> </ul>
19	(Martínez-Abad et al., 2020)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2015.</li> </ul>
20	(Maheshwari et al., 2020)	<ul style="list-style-type: none"> <li>• Portal do Governo de livre acesso da Índia</li> </ul>
21	(Silva, 2020)	<ul style="list-style-type: none"> <li>• Direção-Geral de Estatísticas de Educação e Ciência (DGEEC)</li> </ul>
22	(Câmara Leme et al., 2020)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2009 e 2012;</li> <li>• Programa Internacional para a Avaliação das Competências dos Adultos (PIAAC) 2012</li> </ul>
23	(Larson et al., 2021)	<ul style="list-style-type: none"> <li>• <i>Maryland Safe and Supportive Schools (MDS3)</i></li> </ul>
24	(Alam et al., 2021)	<ul style="list-style-type: none"> <li>• Coleção de dados de várias escolas (<a href="http://dx.doi.org/10.17632/637d4s7vjh.1">http://dx.doi.org/10.17632/637d4s7vjh.1</a>)</li> </ul>
25	(Altun & Kalkan, 2021)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2015.</li> </ul>
26	(Khan & Ghosh, 2021)	<ul style="list-style-type: none"> <li>• Não utilizou dados provenientes de fontes externas.</li> </ul>
27	(Langi & Jeon, 2022)	<ul style="list-style-type: none"> <li>• OCDE: Estudo PISA 2019.</li> </ul>
28	(Huberts et al., 2022)	<ul style="list-style-type: none"> <li>• Escola secundária;</li> <li>• Governo holandês.</li> </ul>

## CAPÍTULO

### 3. Compreensão e preparação de dados

Este capítulo visa enquadrar os dados envolvidos neste projeto, mostrando quais os indicadores selecionados e pretendidos para análise, as suas origens, apresentando também as suas características para um melhor entendimento.

#### 3.1 Seleção dos indicadores para investigação

Uma vez analisados os resultados obtidos na revisão de literatura da presente investigação, é possível selecionar os indicadores e medidas de sucesso para análise, destacando também quais os dados pretendidos e necessários para que se consiga proceder ao cálculo dos mesmos. A fim de promover uma análise mais minuciosa dos indicadores de sucesso educacional, procedeu-se ainda à desagregação dos mesmos em três níveis de detalhe distintos: indicadores de nível escolar, indicadores ao nível de curso e indicadores ao nível da turma.

As medidas de sucesso em termos de aproveitamento selecionadas para esta investigação estão representadas na Tabela 7.

**Tabela 7.** Medidas de Sucesso selecionadas para estudo.

Nível de detalhe	ID	Medida de Sucesso
Medidas de sucesso a calcular para cada um dos níveis de detalhe	1	Média de Exames de 1ª Fase
	2	Média de Exames de 2ª Fase
	3	Médias Finais
	4	Resultados PISA
	5	Percentagem de alunos em exames 2ª Fase
	6	Percentagem de alunos reprovados
	7	Nº de Alunos desistentes

Relativamente aos indicadores que podem ter influência e peso nas medidas de sucesso, a Tabela 8 apresenta todos os indicadores respetivos para cada nível de detalhe.

**Tabela 8.** Indicadores de sucesso sugeridos e selecionados para estudo.

<b>Nível de detalhe</b>	<b>ID</b>	<b>Indicador</b>	<b>Dados</b>
Indicadores por escola	1	Tipo de escola	Natureza da escola (Pública/Privada)
	2	Tamanho da escola	Nº de alunos
	3	Tamanho da equipa docente	Nº de professores
	4	Rácio tamanho da escola face ao tamanho da equipa docente	Nº de alunos
			Nº de professores
	5	Nº médio de alunos por turma	Nº de alunos
			Nº de turmas
	6	Recursos escolares disponibilizados	Nº de cursos
			Nº de salas de aula disponíveis em cada escola
			Nº de computadores disponíveis por escola
			Nº de bibliotecas disponíveis por escola
	7	Média do nº de horas de aulas por semana	Nº de lugares nos refeitórios por escola
			Nº de horas de aulas por semana
	8	Tipo de localidade da escola	Tipo de localidade da escola (Rural/Urba)
	9	Fatores socioeconómicos da escola	Nível socioeconómico da escola
10	Nº de matrículas anuladas	Nº de matrículas anuladas	
11	Nº de alunos transferidos	Nº de alunos transferidos	
12	Fatores socioeconómicos dos estudantes	Contagem de alunos por cada grau de escalão ASE	
13	Nível de educação parental dos estudantes	Nº de parentes por cada grau de qualificação.	
14	Nível médio de qualificação dos professores	Nº de professores por cada grau de qualificação	
15	Média de anos de experiência dos professores	Nº de anos de experiência de cada professor	
Indicadores por curso	16	Tamanho dos cursos	Nº de alunos por curso
	17	Tamanho da equipa docente	Nº de professores
	18	Nº de matrículas anuladas	Nº de matrículas anuladas
	19	Nº de alunos transferidos	Nº de alunos transferidos
	20	Nível de educação parental dos estudantes	Nº de parentes por cada grau de qualificação.
	21	Nível médio de qualificação dos professores	Nº de professores por cada grau de qualificação
	22	Média de anos de experiência dos professores	Nº de anos de experiência de cada professor
Indicadores por turma	23	Tamanho das turmas	Nº de alunos por turma
	24	Tamanho da equipa docente	Nº de professores
	25	Nº de matrículas anuladas	Nº de matrículas anuladas
			Nº de alunos transferidos
	27	Nível de educação parental dos estudantes	Nº de parentes por cada grau de qualificação.
	28	Nível médio de qualificação dos professores	Nº de professores por cada grau de qualificação
	29	Média de anos de experiência dos professores	Nº de anos de experiência de cada professor

## 3.2 Pedido e extração de dados

A obtenção dos dados para este estudo, em parceria com o projeto Inteligência Artificial para a Administração Pública (IA>AP), envolveu também uma colaboração essencial com a Direção-Geral de Estatísticas da Educação e Ciência (DGEEC), visando adquirir informações detalhadas e representativas que permitissem a análise objetiva dos fatores influenciadores do desempenho escolar. A DGEEC é um “serviço central da administração direta do Estado português”, cuja finalidade passa por analisar e divulgar dados estatísticos relativos à educação e ciência, fornecendo também informações precisas e fiáveis sobre o sistema educativo português, de modo a apoiar a formulação de políticas e melhorar os resultados educativos.<sup>1</sup>

A obtenção dos dados envolveu um processo de solicitação formal, onde foram necessários os seguintes passos: uma preparação da solicitação, tendo sido crucial delinear com precisão os objetivos do estudo e identificar os indicadores específicos a serem analisados, permitindo a formalização de um pedido oficial à DGEEC. Em consonância com as diretrizes de privacidade de dados, foi necessária a assinatura de uma declaração de confidencialidade, assegurando a conformidade com os regulamentos de proteção de dados. Todo o processo de trabalho e extração de dados, segundo o formalizado na declaração, teve de ser efetuado nos próprios estabelecimentos da empresa, respeitando assim as normas éticas e agindo com todas as precauções necessárias para proteger a privacidade das pessoas e instituições cujos dados foram incluídos.

A parceria estabelecida com a DGEEC revelou-se valiosa em diversos aspetos, tendo sido demonstrado um compromisso sólido em apoiar e contribuir para o êxito do projeto. Contudo, é necessário destacar que o processo de extração de dados enfrentou desafios substanciais, que merecem uma análise cuidadosa.

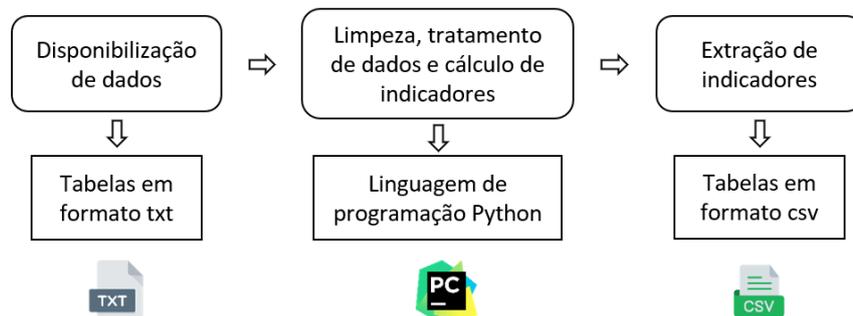
Um dos principais obstáculos foi a complexidade decorrente das rígidas políticas de privacidade de dados. A necessidade de proteger informações sensíveis é de extrema importância, pelo que as restrições associadas a essas políticas tornaram a extração de dados uma tarefa demorada e complexa, o que por vezes exigiu uma rigorosa aprovação prévia para aceder a informações específicas.

Ademais, as limitações técnicas das máquinas disponíveis nos estabelecimentos da DGEEC também se revelaram um fator significativo. As máquinas disponíveis não estavam idealmente equipadas para a tarefa de extração de dados em grande escala, resultando em atrasos e interrupções frequentes no processo. A aquisição e instalação dos programas necessários para a extração de dados também provou ser um desafio, pois a compatibilidade desses programas com o ambiente existente gerou questões complexas. Adicionalmente, notou-se que os dados disponibilizados pela DGEEC nem sempre estavam completamente alinhados com as necessidades de pesquisa, ocorrendo a necessidade de fazer vários ajustes ao pedido inicial. Tudo isto gerou inúmeras revisões e ajustes ao projeto, resultando num processo de aquisição de dados bastante mais longo do que o previsto inicialmente. Em resumo, apesar da cooperação positiva com a DGEEC e o seu apoio dedicado, o processo de extração de dados enfrentou vários desafios, incluindo restrições de privacidade, limitações técnicas, questões de software e inconsistências nos dados fornecidos.

---

<sup>1</sup> <https://www.dgeec.mec.pt/np4/dgeec/>

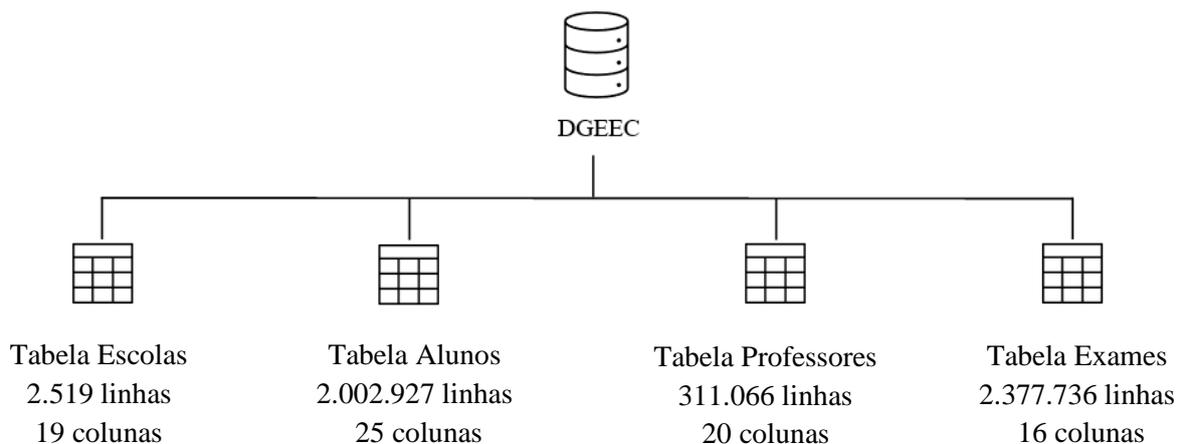
Todos os conjuntos de dados apresentavam-se estruturados em formato de ficheiro de texto, pelo que, para iniciar o processo de preparação de dados, foi necessário importá-los para o ambiente de desenvolvimento integrado PyCharm e, posteriormente, utilizar a linguagem de programação Python através também da versatilidade da biblioteca Pandas. Estas escolhas técnicas foram feitas devido às vantagens que o Python e o Pandas oferecem no tratamento e manipulação eficiente de conjuntos de dados de grande volume, no sentido de limpar as tabelas, a fim de prepará-las para o cálculo de indicadores. Após a conclusão das tarefas de limpeza e de cálculo, também todo o processo de extração de dados foi efetuado através da linguagem de programação anteriormente designada, permitindo extrair os indicadores finais de maneira mais eficaz, para ficheiros no formato CSV (Figura 4).



**Figura 4.** Processo de preparação de dados.

### 3.3 Descrição de dados disponibilizados

Esta secção tem como finalidade apresentar uma visão e descrição dos dados fornecidos que constituem a base para o cálculo e análise de indicadores de sucesso no contexto das escolas do ensino secundário. Os registos de dados que foram disponibilizados para posterior extração e cálculo de indicadores foram divididos em quatro tabelas distintas, contendo um grande volume de informações, tal como demonstrado na Figura 5.



**Figura 5.** Tabelas disponibilizadas pela DGEEC.

Relativamente a estas tabelas de dados disponibilizadas, todos os seus campos e respetivos significados, estão devidamente documentadas e disponíveis no Apêndice B deste documento, intitulado como “Tabelas de dados disponibilizadas”. Este apêndice serve como uma referência detalhada para garantir a compreensão completa e precisa dos dados disponibilizados para este estudo.

### 3.4 Preparação de dados

O processo de preparação em grandes conjuntos de dados é uma etapa crucial em qualquer estudo que envolva a extração de informações, com o objetivo de assegurar uma maior qualidade nos resultados. Dentro desta etapa é descrito o processo de limpeza e tratamento de dados, onde foram encontrados alguns desafios, podendo ser destacadas algumas complexidades inerentes.

#### Exclusão de campos não desejados

No sentido de simplificar as estruturas dos dados, reduzindo a complexidade dos mesmos e permitir que apenas informações essenciais fossem retidas para o cálculo dos indicadores, foram identificados e eliminados todos os campos que não eram relevantes para a análise (Tabela 9).

**Tabela 9.** Campos excluídos de cada tabela disponibilizada.

Tabela Escolas	Tabela Alunos	Tabela Professores	Tabela Exames
CNUTSI	ESTABELECIMENTO	GENERO	INTERNO
NUTSI	CGENERO	CATEGORIA	PARAMELHORIA
CCONCELHO	CIDADE	IDADE	PARAMELHORIA
CONCELHO	MÃE_HABILITAÇÃO	HR_COMP_LETIVA	USOUTOLERANCIA
CNATUREZA2	PAI_HABILITAÇÃO	HR_D4	CURSO_DESCR
NATUREZA2	EE_HABILITAÇÃO	HR_COMP_N_LETIVA	-
CTIPOLOGIA	MÃE_NACIONALIDADE	DIRETOR_ESCOLA	-
TIPOLOGIA	PAI_NACIONALIDADE	CNACIONALIDADE	-
CNUTSII	MÃE_PROFISSÃO	NACIONALIDADE	-
NUTSII	PAI_PROFISSÃO	COMP_FORMACAO	-
CNUTSIII	MÃE_SIT_EMPREGO	ATIVIDADE	-
NUTSIII	PAI_EMPREGO	-	-
CAGRUPAMENTO	EE_PARENTESCO	-	-
AGRUPAMENTO	EE_NACIONALIDADE	-	-
-	EE_PROFISSÃO	-	-
-	EE_SIT_EMPREGO	-	-

À exceção do grande número de campos que não se enquadravam com o objetivo final deste estudo, surgem outros, que apesar de terem importância para a investigação e permitirem calcular alguns dos indicadores sugeridos, apresentavam alguns desafios: os campos que prestavam informações geográficas das escolas (“CONCELHO” e “NUTS I”, “NUTS II” e “NUTS III”) poderiam ser utilizados no sentido de substituir o indicador sugerido pelos estudos revistos, “Tipo de Localidade da escola”, que tinha como propósito distinguir escolas localizadas em zonas rurais ou urbanas.

No entanto, não se procedeu à utilização dos mesmos devido ao elevado nível de detalhe, complexidade e tempo que estes encargariam, deixando como sugestão para trabalhos e projetos de investigação futuros.

Para além destes, destaca-se também os campos “MÃE\_HABILITAÇÃO” e “PAI\_HABILITAÇÃO” que seriam úteis para o cálculo do indicador sugerido “Nível de educação parental dos estudantes”. Estes campos foram excluídos e não considerados para análise devido ao elevado número de valores ausentes nas tabelas disponibilizadas. Esta escolha de não proceder ao cálculo do indicador em questão através destes campos, veio garantir a qualidade e a credibilidade dos dados selecionados para a presente investigação.

### **Valores ausentes, duplicados e/ou incorretos**

Foram identificados e tratados registos com valores ausentes em campos importantes para a análise final. Este tipo de situação verificou-se na tabela “Exames”, mais especificamente na coluna CFD (Classificação Final à Disciplina) onde bastantes registos se encontravam a NULL. De forma a manter a qualidade dos dados para o cálculo da medida de sucesso que estava dependente destes (Médias Finais), procedeu-se à exclusão dos registos que lhes estavam associados.

A deteção e remoção de registos duplicados foram também realizadas para garantir que cada entrada no conjunto de dados fosse única e não introduzisse valores indevidos nas análises. A coluna “TURMA” da tabela “Alunos” foi o único exemplo do descrito, uma vez que ao invés de apresentar valores únicos para cada turma de uma escola, apresentava o mesmo valor para turmas do mesmo ano letivo, mas de escolas, cursos ou anos de ensino diferentes. Foi criada uma chave identificadora de cada turma através da chave primária composta pelos campos “ANOLETIVO”, “CESCOLA”, “ANO\_ENSINO”, “CURSO” e “TURMA”.

### **Exclusão de registos não desejados**

Nesta etapa do processo de tratamento de dados, para que estes fossem filtrados tendo em conta o âmbito deste projeto de estudo, foram ainda excluídos alguns registos presentes nas tabelas disponibilizadas pela DGEEC. Na tabela “Professores”, na coluna “CICLO\_DOCENCIA”, de modo que fossem apenas contemplados dados de professores que lecionavam no ensino secundário foram filtrados todos os registos com o valor “Secundário”.

Foram também tomadas medidas para refinar o conjunto de dados na tabela “Exames”. O foco estava na coluna “EXAME”, que diz respeito à disciplina à qual o aluno realizou exame, estando presentes inicialmente um total de 52 disciplinas distintas. O objetivo, portanto, passou por filtrar e reter apenas as disciplinas de Inglês, Português, Geometria Descritiva A, Filosofia, Matemática A, Matemática B, Biologia e Geologia, Física e Química A, Francês, Geografia A, Matemática Aplicada as Ciências Sociais., Desenho A, História A, História B, Economia A, Alemão e Espanhol para análise posterior. Para alcançar esta seleção específica, as linhas que não continham as referidas disciplinas foram removidas da tabela.

Este procedimento de eliminação resultou numa redução significativa de dados, passando de 52 disciplinas em análise para apenas 17. Desta forma, o conjunto de dados foi concentrado nas disciplinas de interesse, tornando-o mais relevante e alinhado com os objetivos da análise futura.

## **Inconsistência nas tipologias de dados**

Durante a análise das tabelas, surgiram inconsistências nas tipologias dos dados, particularmente em relação à representação dos anos letivos. Enquanto a tabela "Exames" adotava um formato de dados inteiros (por exemplo, "2014") para indicar os anos letivos, as demais tabelas utilizavam um formato de intervalo em varchar ("2015/2016").

Esta disparidade nos formatos de dados entre as tabelas complicou a integração e a comparação das informações, pelo que foi necessário alinhar as tipologias de dados e permitir a correlação entre as tabelas. Foi implementado um processo de modificação da tipologia de dados na coluna "ANOLETIVO" da tabela "Exames". Toda esta ação foi executada com o objetivo de assegurar a uniformidade na representação dos anos letivos em todas as tabelas, simplificando, assim, a integração e a análise conjunta dos dados.

## **Inconsistência nos valores temporais e sazonais dos dados**

Observou-se uma inconsistência nos valores temporais e sazonais dos dados presentes nas quatro tabelas disponibilizadas. A tabela "Exames" continha informações que abrangiam o período do ano letivo de 2015/2016 a 2020/2021, enquanto as tabelas "Escolas", "Alunos" e "Professores" tinham dados que iam de 2013/2014 a 2020/2021. Para garantir a coesão dos dados e a consistência nas análises posteriores, optou-se por focar apenas nos registros que eram comuns a todas as tabelas, isto é, foram selecionados para análise e tratamento somente os anos letivos que estavam presentes em todas as tabelas. Esta abordagem assegurou que os dados utilizados estivessem harmonizados em termos temporais, possibilitando análises mais robustas e coerentes.

## **Conversão de dados categóricos em numéricos (Codificação)**

O procedimento de codificação foi aplicado a duas colunas de diferentes tabelas com a finalidade de facilitar análises futuras e permitir que algoritmos matemáticos pudessem ser aplicados de forma eficiente. A primeira coluna que sofreu codificação foi a "ESCALAO\_ASE" da tabela "Alunos", que originalmente continha quatro categorias, sendo que cada uma delas foi mapeada para um valor numérico, como segue:

“Não Beneficia”: 1; “Escalão C”: 2; “Escalão B”: 3; “Escalão A”: 4.

Igualmente, o mesmo processo foi executado na coluna "HABILITACAO" da tabela "Professores" conforme o seguinte: “Outra”: 1; “Bacharelato”: 2; “Licenciatura”: 3; “Mestrado/Doutoramento”: 4.

As codificações descritas permitiram que as categorias fossem representadas de maneira numérica, tornando os dados mais adequados para análises estatísticas e facilitando o cálculo de métricas, quando necessário.

Este tipo de abordagem foi executado de forma precisa no sentido de não só preservar a integridade das informações como também garantir que a representação numérica refletisse fielmente o significado das categorias originais.

## Categorização e segmentação de dados

Nesta seção, é abordado o processo de categorização e segmentação de dados aplicado ao conjunto de dados referentes à coluna “ACONTECIMENTO” da tabela “Alunos”. O objetivo principal desta etapa foi facilitar o cálculo de indicadores relevantes para a análise, conseguindo respeitar ainda, rigorosamente, a privacidade dos dados pessoais dos alunos. Tal como solicitado pela DGEEC, esta etapa serviu também para proteger a privacidade dos dados dos alunos e evitar a divulgação de informações sensíveis, uma vez que um dos requisitos impostos pela DGEEC para a extração de informação foi a adequação da mesma no sentido de não divulgar informações que dizem respeito à quantificação num número menor ou igual a 3. Como tal, as seguintes categorizações foram implementadas:

Número de Desistentes para qualquer nível de detalhe:

Categoria “0”: Nenhum aluno registado como desistente;

Categoria “ $\leq 3$ ”: Indica que o número de desistentes é igual ou inferior a 3;

Categoria “ $> 3$ ”: Indica que o número de desistentes é superior a 3.

Número de Matrículas Anuladas para qualquer nível de detalhe:

Categoria “0”: Nenhuma matrícula anulada registada.

Categoria “ $\leq 3$ ”: Indica que o número de matrículas anuladas é igual ou inferior a 3;

Categoria “ $> 3$ ”: Indica que o número de matrículas anuladas é superior a 3.

Número de Alunos Transferidos para qualquer nível de detalhe:

Categoria “0”: Nenhum aluno registado como transferido.

Categoria “ $\leq 3$ ”: Indica que o número de alunos transferidos é igual ou inferior a 3;

Categoria “ $> 3$ ”: Indica que o número de alunos transferidos é superior a 3.

## 3.5 Análise descritiva dos dados

Após todo o processo de limpeza e tratamento de dados, foi possível realizar uma análise descritiva relativa a todos os dados utilizados para cálculo e estudo de indicadores de sucesso escolar. O conjunto de dados utilizado apresenta um retrato abrangente do cenário educacional, englobando durante um período de seis anos letivos (2015/2016 a 2020/2021) tanto escolas públicas como privadas. No total, os dados abrangem 673 escolas secundárias, das quais 525 (78.01%) são de natureza pública, enquanto as restantes 148 (21.99%) são classificadas como escolas privadas. No decorrer dos anos em análise, é possível indicar que o número de escolas se manteve constante face à sua natureza verificando-se um maior no número de escolas públicas face às privadas (Figura 6).

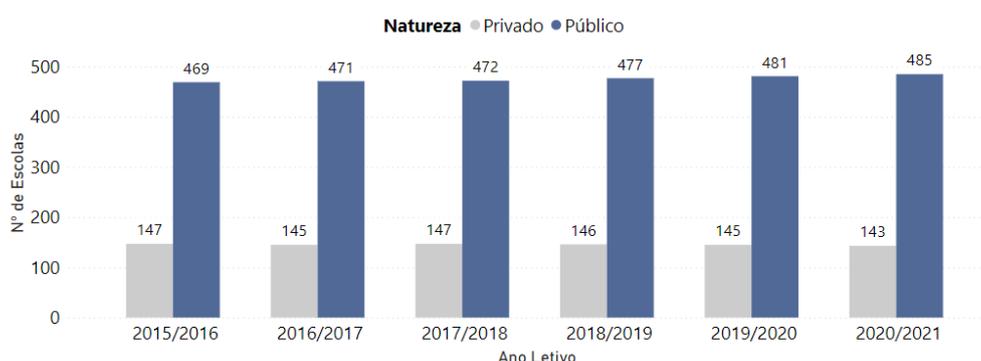
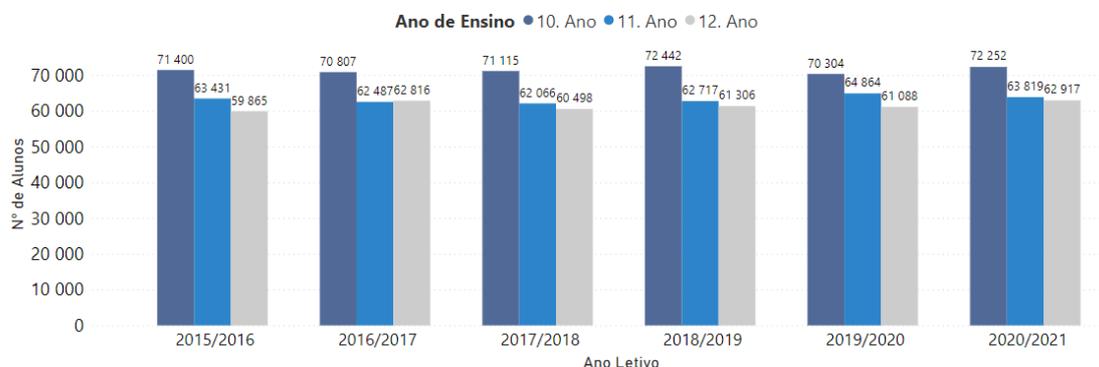


Figura 6. Número de escolas por Natureza e Ano Letivo.

No que diz respeito aos alunos, o conjunto de dados contempla informações detalhadas sobre um total de 575 568 alunos, matriculados nos três anos de ensino (10º, 11º e 12º ano) em escolas secundárias, ao longo dos seis anos considerados. Também estes números se mantiveram constantes nos anos em análise (Figura 7).



**Figura 7.** Número de Alunos por Ano de Ensino e Ano Letivo.

Já em termos de turmas, todos os alunos referenciados anteriormente encontram-se distribuídos por 70.927 turmas no total, estando estas desagregadas por um amplo espectro de quatro cursos (Artes Visuais, Ciências e Tecnologias, Ciências Socioeconómicas e Línguas e Humanidades). Através da análise descritiva da tabela 10, podemos observar uma distribuição detalhada do número de turmas por curso e ano de ensino. É possível indicar que o curso "Ciências e Tecnologias" tem o maior número de turmas em todos os anos, com um total de 29 633 turmas (41.78%). Por outro lado, o curso "Artes Visuais" tem a menor quantidade de alunos, com um total de 7 848 (11.06%).

O ano de ensino que mais se destaca é o 10º ano, que apresenta o maior número de estudantes (25 062) face aos outros anos de ensino em análise.

**Tabela 10.** Número de turmas por Curso e Ano de Ensino.

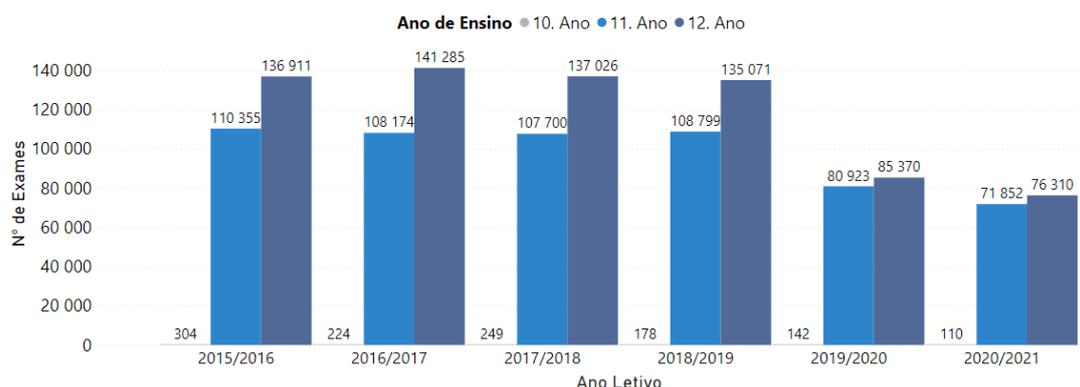
Curso	10º ano	11º ano	12º ano	Total
Artes Visuais	2 762	2 550	2 536	<b>7 848</b>
Ciências e Tecnologias	10 358	9 513	9 762	<b>29 633</b>
Ciências Socioeconómicas	4 552	4 162	4 357	<b>13 071</b>
Línguas e Humanidades	7 390	6 564	6 421	<b>20 375</b>
<b>Total</b>	<b>25 062</b>	<b>22 789</b>	<b>23 076</b>	<b>70 927</b>

Os dados envolvem também informações relativas a todos os professores que exerceram as suas funções na educação do ensino secundário. A Figura 8 que expõe o número de professores ao longo de vários anos letivos, facilita a análise e a tendência geral de aumento no número de professores durante o período considerado. De 2015/2016 a 2020/2021, há um aumento constante no número de professores, começando com 31.821 professores e aumentando gradualmente, atingindo os 33.988 professores. No entanto, a variação anual é relativamente pequena, com a maior diferença a ser de 816 professores entre o ano letivo de 2016/2017 e 2017/2018, e a menor diferença de apenas 51 professores entre 2017/2018 e 2018/2019. Estas flutuações provam que número de professores se manteve relativamente estável, com variações que não ultrapassam os 2,6% em qualquer ano letivo.



**Figura 8.** Número de Professores por Ano Letivo.

É possível também analisar de forma geral os dados relacionados com os exames de 1ª fase realizados ao longo dos anos letivos e diferentes anos de ensino. Tanto os resultados dos exames, como os resultados finais são apresentados numa escala de 0 a 200. A Figura 9 apresenta o número de exames realizados em cada ano de ensino durante os anos letivos de 2015/2016 a 2020/2021. Os dados revelam flutuações significativas no número de exames de 1ª Fase realizados de um ano letivo para outro que serão abordadas no capítulo seguinte deste trabalho de dissertação. O ano de ensino cujos alunos mais realizam exames foi desde sempre o 12º ano seguido do 11º ano, com algumas oscilações, mas mantendo um padrão geral de variação.



**Figura 9.** Número de Exames por Ano de Ensino e Ano Letivo.

Analisando as condições socioeconómicas dos alunos das escolas secundárias, em Portugal os alunos são classificados nos escalões ASE (Ação Social Escolar) com base nas suas condições socioeconómicas e necessidades de financiamento para o acesso à educação. Cada escalão representa um nível diferente de apoio, e sua distribuição é baseada no salário familiar e noutros fatores socioeconómicos.

Com base na coluna disponibilizada pela DGEEC “ESCALAO\_ASE”, que foi também codificada de modo a obter médias ao nível de detalhe escolar, é possível descrever o significado de cada escalão:

**Não Beneficia (Escalão 0):** Os alunos deste escalão não têm direito a nenhum tipo de apoio social escolar. O salário familiar destes alunos é considerado suficiente para pagar todas os gastos relacionados com a educação.

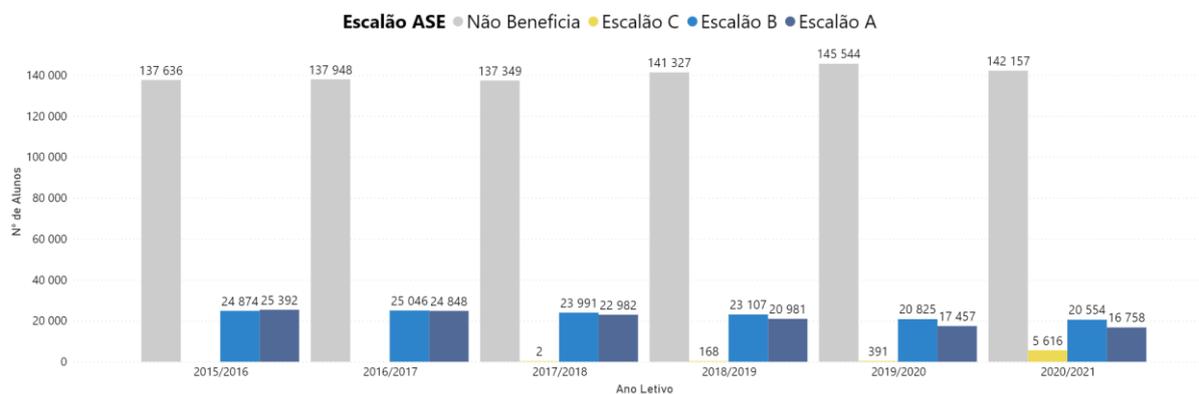
**Escalão C (Escalão 1):** Os alunos elegíveis ao escalão C, enquadram-se dentro do rendimento familiar máximo para obter benefícios sociais escolares.

Este tipo de estudantes recebe benefícios, como subsídios nas refeições escolares e material escolar, mas que geralmente são menos amplos do que os dos alunos dos escalões A e B.

Escalão B (Escalão 2): Os alunos deste escalão têm rendimentos familiares um pouco mais altos do que os alunos do Escalão A, mas também estão em condições socioeconómicas que justificam a necessidade de apoio social escolar. Os benefícios contemplados neste escalão são o acesso mais barato a refeições escolares, subsídios para transporte e apoio financeiro para materiais escolares.

Escalão A (Escalão 3): Já os alunos do escalão A têm um rendimento mais baixo e por isso têm direito ao maior apoio social escolar disponível. O apoio designado contempla subsídios para o transporte escolar, refeições escolares gratuitas, subsídios para material escolar, contando ainda com outros tipos de apoio financeiro.

A Figura 10, apresenta o número de estudantes para cada escalão ao longo de todos os anos letivos. A maioria dos alunos enquadra-se no escalão "Não Beneficia," enquanto o número de alunos enquadrados nos escalões A e B se tem mantido idêntico e constante durante todo o intervalo de tempo em análise. Quando se procede à análise do escalão com menos benefícios, o “Escalão C”, desde o ano letivo de 2018/2019 que se tem verificado um aumento significativo no número de alunos a ele associado, tendo sido atingido um pico de 5 616 alunos em 2020/2021.



**Figura 10.** Número de alunos por Escalão ASE e Ano Letivo.



## CAPÍTULO

### 4. Análise exploratória

Este capítulo concentra-se na especificação e análise dos indicadores calculados, abordando questões-chave, tendências e *insights* relevantes que emergem das mesmas análises. Ao longo deste capítulo são também exploradas criticamente as medidas de sucesso que foram derivadas dos dados, explorando como estas se relacionam com os indicadores e como contribuem para diferentes percepções daquilo que é o sucesso educativo. A escolha e a relação entre os indicadores e as medidas de sucesso a analisar baseiam-se nos estudos descritos da revisão de literatura deste projeto e visa responder às perguntas de investigação deste estudo, abrindo caminhos para novos trabalhos de pesquisa, cruciais para o sistema educacional em foco. Para realizar a análise estatística e visualização dos dados deste estudo, optou-se por uma abordagem metodológica que envolveu o uso de ferramentas específicas. Para o cálculo e análise entre os indicadores, recorreu-se à linguagem de programação Python e a bibliotecas adequadas como Pandas, NumPy, SciPy, Seaborn e Matplotlib. De modo a fornecer uma análise visual recorreu-se ainda ao software Power BI, com o intuito de criar gráficos.

#### 4.1 Indicadores calculados

Com base nos indicadores pré-selecionados e alinhando os mesmos com todos os dados disponibilizados pela DGEEC, esta seção visa especificar todos os indicadores possíveis de calcular, assegurando assim a continuidade e coerência na descrição dos dados em estudo e respondendo à primeira pergunta de investigação deste projeto de estudo. Através da Tabela 11 e Tabela 12 é possível verificar quais as medidas de sucesso, bem como todos os indicadores associados a cada nível de detalhe pré-estabelecido (escolar, por curso e por turma).

Tabela 11. Medidas de sucesso calculadas.

Nível de detalhe	ID	Medidas de Sucesso
Todos os níveis de detalhe	1	Média de Exames de 1ª fase
	2	Média de Exames de 2ª fase
	3	Médias Finais
	5	Percentagem de alunos em exames de 2ª Fase
	6	Percentagem de alunos reprovados
	7	Nº de alunos desistentes

**Tabela 12.** Indicadores de sucesso calculados

Nível de detalhe	ID	Indicador	Dados
Indicadores por escola	1	Tipo de escola	Natureza da escola (Público/Privado)
	2	Tamanho da escola	Nº de alunos
	3	Tamanho da equipa docente	Nº de professores
	4	Rácio tamanho da escola face ao tamanho da equipa docente	Nº de alunos
			Nº de professores
	5	Média do nº de alunos por turma	Nº de alunos
			Nº de turmas
	6	Nº de cursos disponibilizados	Nº de cursos
	10	Nº de matrículas anuladas	Nº de matrículas anuladas
	11	Nº de alunos transferidos	Nº de alunos transferidos
	12	Fatores socioeconómicos dos estudantes	Nº de alunos por cada grau de escalão ASE
14	Nível médio de qualificação dos professores	Nº de professores por cada grau de qualificação	
15	Média de anos de experiência dos professores	Nº de anos de experiência de cada professor	
Indicadores por curso	16	Tamanho do curso (média do nº de alunos por curso)	Nº de alunos
			Nº de turmas
	18	Nº de matrículas anuladas	Nº de matrículas anuladas
19	Nº de alunos transferidos	Nº de alunos transferidos	
Indicadores por turma	23	Tamanho das turmas	Nº de alunos
			Nº de turmas
	25	Nº de matrículas anuladas	Nº de matrículas anuladas
26	Nº de alunos transferidos	Nº de alunos transferidos	

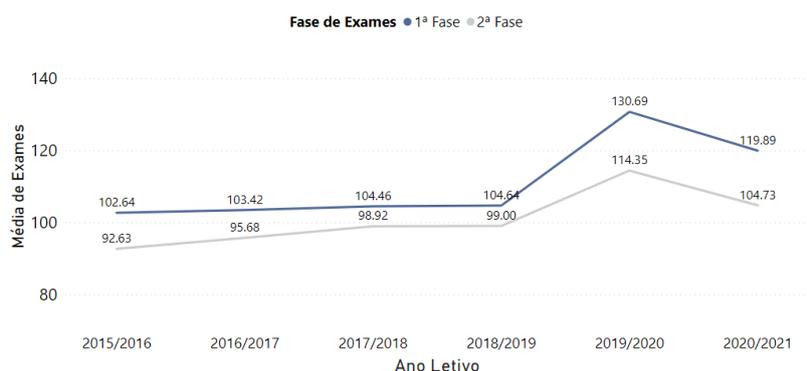
É importante ainda indicar que todas as informações detalhadas relativas a cada uma das medidas e indicadores estão devidamente documentadas no Apêndice C deste trabalho, intitulado por “Medidas e Indicadores de Sucesso Calculados”, onde para cada um dos mesmos, são fornecidos os seguintes detalhes: nome do indicador, número de campos utilizados e número de observações analisadas. Para além destes, são dados detalhes também quanto aos seus campos: nome do campo, tipologia do campo, número de valores únicos, valor mínimo, valor máximo, média e desvio padrão.

## 4.2 Análise de medidas de sucesso

As medidas de sucesso calculadas para este projeto de investigação fornecem um quadro objetivo para avaliar o alcance e a eficácia dos estudantes do ensino secundário. Nesta seção, são analisadas em detalhe as medidas e como estas têm sofrido alterações ao longo do tempo.

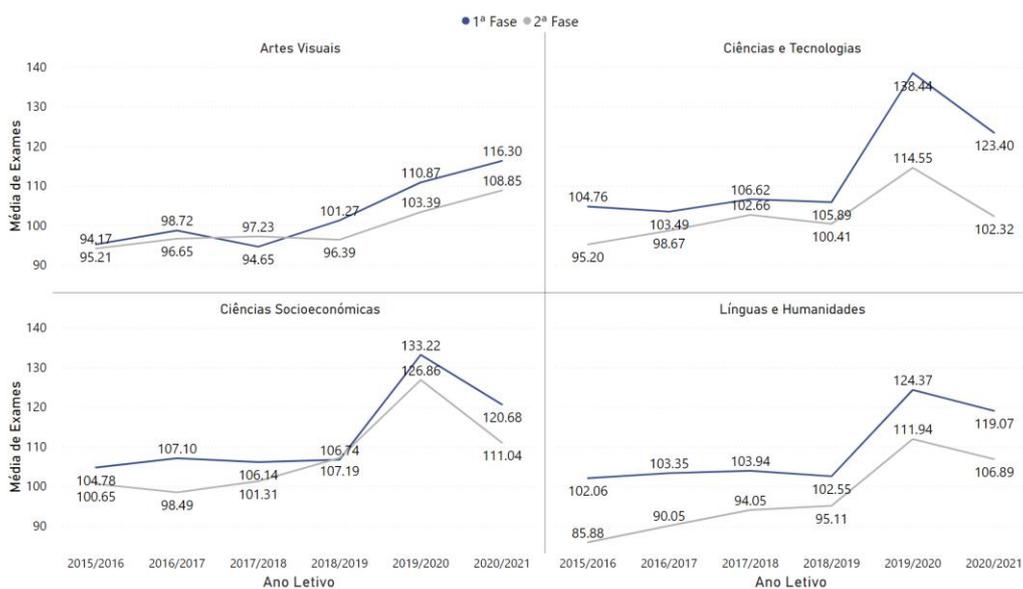
## Médias de Exames de 1ª e 2ª Fase

Tradicionalmente, o sucesso dos alunos tem sido associado ao desempenho nas suas avaliações, e uma medida importante desse desempenho são as médias externas obtidas. A Figura 11 traduz uma visão geral das tendências e variações nas médias das notas dos exames de primeira e segunda fase ao longo dos anos letivos. No geral, as médias apresentam flutuações, mas não seguem uma tendência clara de aumento ou diminuição. O ano letivo de 2019/2020 destaca-se face aos restantes, através das médias notavelmente mais altas, fazendo contraste com anos letivos de 2015/2016 a 2018/2019 que mostram médias relativamente mais baixas.



**Figura 11.** Médias de Exames de 1ª e 2ª fase por Ano Letivo (escala 0- 200).

Quando a análise é feita num nível de detalhe superior, por curso, verifica-se em todos eles a mesma subida de média no ano letivo 2019/2020 (Figura 12). Quando o foco é comparar resultados entre estes, é possível destacar que o melhor curso em primeira fase é "Ciências e Tecnologias" com uma média global de 113.55 e que o curso que apresenta piores médias é "Artes Visuais" com uma média global de 102.30. Em exames de segunda fase o melhor curso é o de "Ciências Socioeconómicas" com 106.68 de média global, sendo "Línguas e Humanidades" o pior, com uma média de 95.98.



**Figura 12.** Médias de Exames de 1ª e 2ª Fase por Curso e Ano Letivo (escala 0- 200).

O aumento significativo nas médias dos exames de ambas as fases durante o ano letivo de 2019/2020 pode ser atribuído, em parte, às mudanças significativas nas leis educacionais que foram implementadas como resultado da pandemia de COVID-19. Durante esse período único, as regulamentações educacionais sofreram alterações para acomodar as circunstâncias únicas do sistema educacional.

Uma das mudanças mais notáveis foi o fim da obrigação dos alunos de realizar exames. Nesse novo cenário, todos os exames tornaram-se opcionais, sendo apenas necessários fazer exame os alunos que optassem por utilizá-los como provas de ingresso ao ensino superior. Toda esta flexibilidade nas avaliações apresentou influência na abordagem dos alunos em relação aos exames, criando um ambiente acadêmico diferente do habitual. As condições excepcionais causadas pela pandemia tiveram impactos significativos no desempenho acadêmico e na dinâmica dos exames escolares.

### Médias Finais

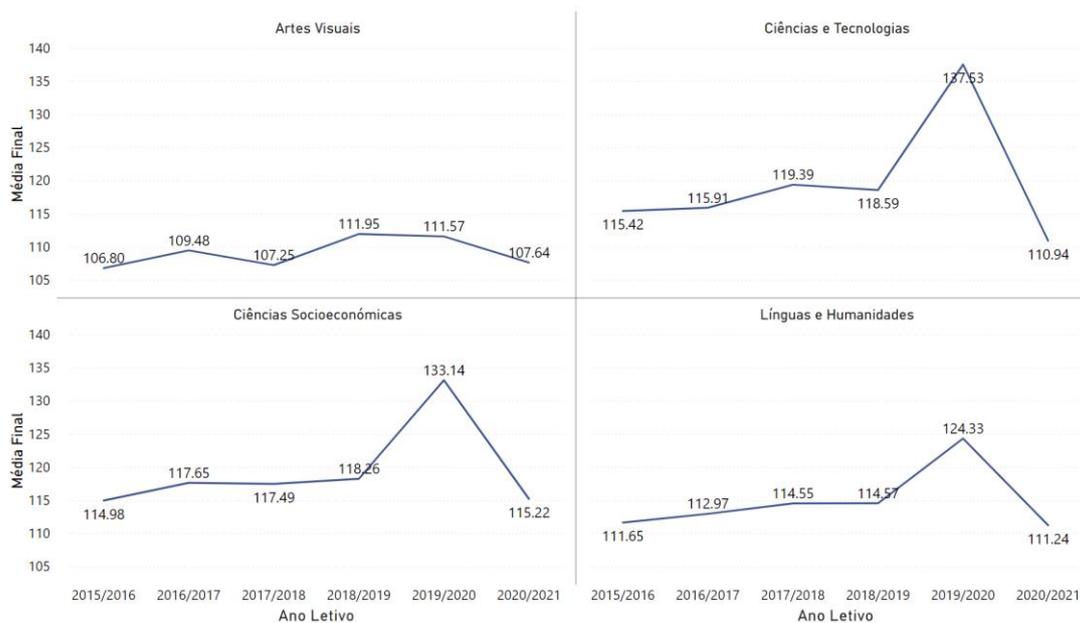
A medida relativa à média final escolar indica o sucesso que os alunos têm no final da sua avaliação tanto interna como externa. A análise de dados temporal a esta medida de sucesso pode ser feita com base na Figura 15. Nos anos letivos 2015/2016 e 2016/2017, as médias finais escolares mantiveram-se em 117.3 valores, sugerindo uma certa estabilidade nesse período. No entanto, a partir do ano letivo 2017/2018, observa-se um aumento na média, atingindo os 119.24 valores. Essa tendência ascendente continuou nos anos seguintes, com a média da nota final alcançando 130.61 em 2019/2020.

Tal aumento pode também ser justificado face às mesmas razões descritas na análise das medidas de sucesso acima mencionadas. No entanto, no ano letivo 2020/2021, existe uma queda nas médias finais, que diminuiu para 110.18, o record mínimo atingido de todos os anos em análise.



**Figura 13.** Média Final por Ano Letivo (escala 0-200).

Esta análise, quando feita ao nível de detalhe dos cursos do ensino secundário, pode-se verificar através da Figura 16 que, uma vez mais, o curso de "Ciências e Tecnologias" se destaca com uma média final de todos os anos letivos mais alta, obtendo uma média de aproximadamente 119.92, ao contrário dos alunos de "Artes Visuais", com a média mais baixa de todos os cursos, de aproximadamente 109.15. Em todos estes, verifica-se uma subida de rendimento no ano letivo de 2019/2020 e uma descida repentina em 2020/2021.



**Figura 14.** Média Final por Curso e Ano Letivo (escala 0-200).

### Percentagem de alunos em exames de 2ª Fase

A segunda fase dos exames nacionais é uma oportunidade adicional para que os estudantes demonstrem seu conhecimento, pelo que o número de alunos que participam nessa mesma fase pode também vir a medir parte do sucesso alcançado por uma escola. Fazendo uma análise à Figura 13, é revelada uma tendência no comportamento dos alunos em relação à segunda fase de exames ao longo dos anos letivos.

No ano letivo 2015/2016, observou-se uma percentagem significativa de 45%, no entanto, nos anos subsequentes, esses valores sofreram flutuações notáveis. No ano letivo 2016/2017, houve um aumento no número de alunos, tendo sido atingido o pico máximo de 48%. A partir desse ano letivo o número de presenças dos estudantes tem vindo a diminuir, atingindo recordes no ano letivo 2019/2020 devido às mesmas razões que fizeram elevar os resultados obtidos neste mesmo tipo de avaliações, isto é, como as políticas de obrigatoriedade de realização em certos exames foi alterada, a disposição dos alunos para participar em segundas fases de exames mudou.



**Figura 15.** Percentagem de alunos presentes em exames de segunda fase.

Tendo como base a Figura 14 que indica o número de alunos que repetiram exames por curso em cada ano letivo, destaca-se o curso de Ciências e Tecnologias, que emerge como o que consistentemente atrai o maior número de alunos que precisam de ir a segunda fase. A percentagem média de alunos que repetiram exames para este curso foi de aproximadamente 44.51%, significativamente superior ao curso de Ciências Socioeconómicas com uma percentagem média de 37.25%.

O curso de Artes Visuais aparece como terceiro neste ranking, onde cerca de 25.87% dos alunos repetem exames. Por fim, o curso de Línguas e Humanidades apresenta consistentemente a percentagem média mais baixa de alunos que repetem o exame, com uma percentagem de 15.25%.

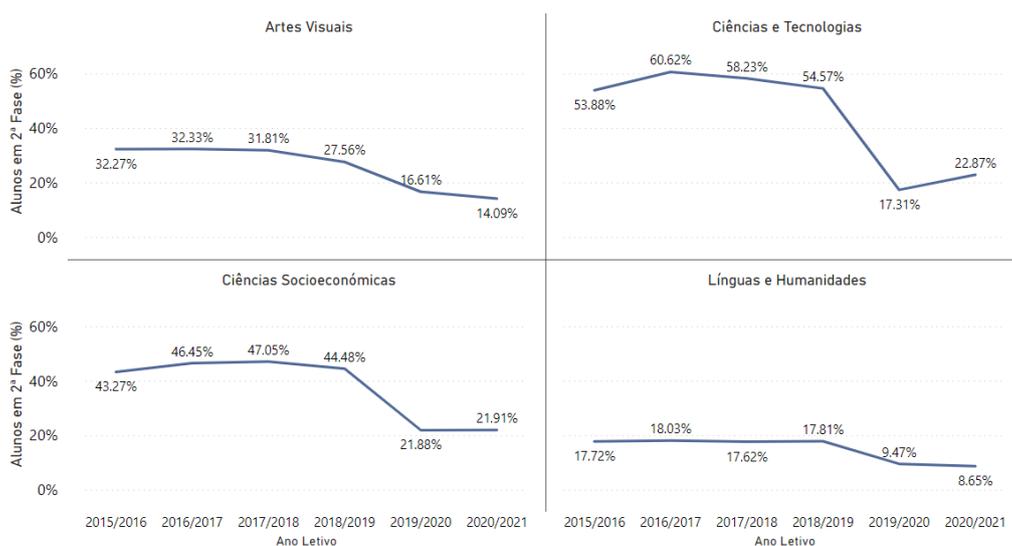


Figura 16. Percentagem de alunos presentes em segunda fase por Curso e Ano Letivo.

### Percentagem de alunos reprovados

A métrica de foco nesta secção traduz o número de alunos reprovados face ao número total de alunos matriculados em cada ano, que transformada em percentagem, permite avaliar a sucesso dos alunos ao longo do tempo. A Figura 17 revela uma série de informações, sendo possível indicar que o ano letivo 2016/2017 apresenta-se como o pior relativamente à taxa de reprovação (13.93%). No entanto, a partir do ano letivo 2017/2018, houve uma tendência de redução nas reprovações, com 12.87% em 2017/2018 e 12.29% no ano seguinte. O declínio mais significativo no número de reprovações ocorreu no ano letivo 2019/2020, com apenas 9.07% dos alunos reprovados.

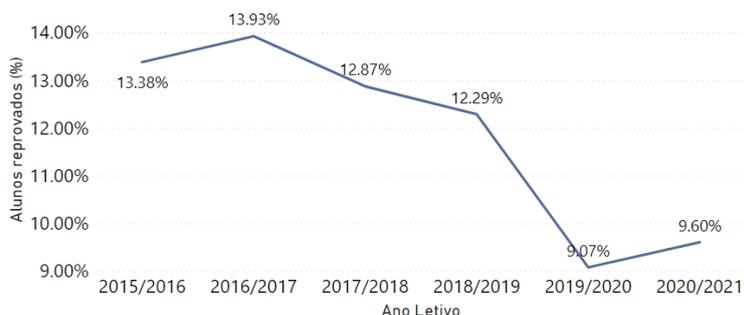
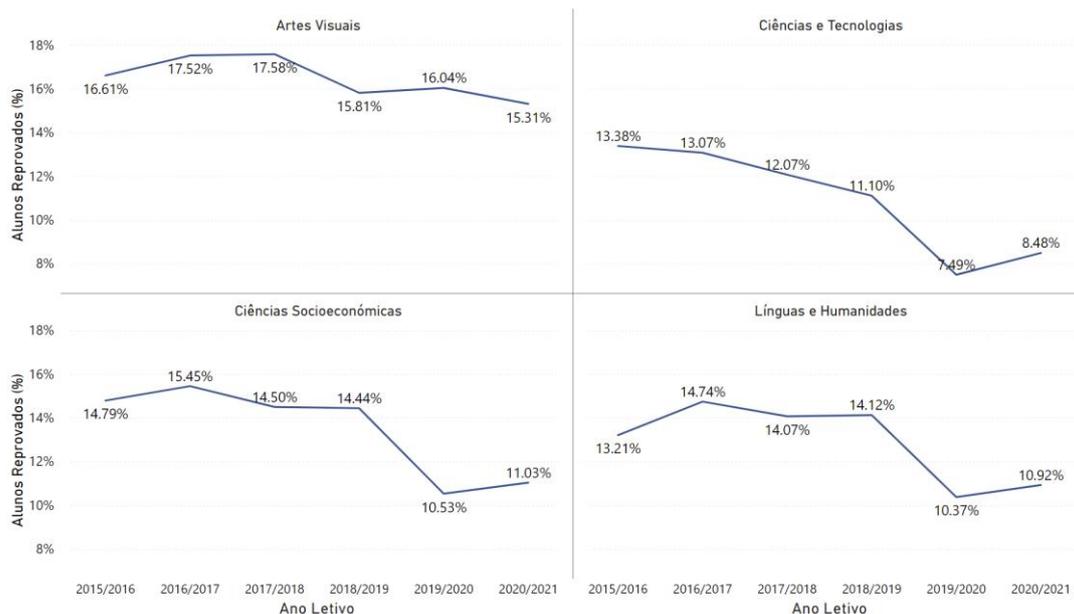


Figura 17. Percentagem de alunos por Ano Letivo.

Ao analisar as taxas médias de alunos reprovados ao longo dos anos e cursos, identificam-se algumas tendências importantes (Figura 18). O curso de "Artes Visuais", com uma taxa média de aproximadamente 16.55% de alunos reprovados é relativamente alta em comparação com os outros cursos e sugere que os alunos desse curso enfrentam desafios adicionais naquilo que são as suas avaliações. Por outro lado, o curso de "Ciências e Tecnologias", apresenta uma percentagem média de aproximadamente 11.27% de alunos reprovados, o mais baixo dos quatro, indicando um desempenho acadêmico geralmente mais sólido. A percentagem mais elevada durante intervalo de tempo de análise foi atingida pelo curso de "Artes Visuais" (17.58%) em 2018/2019, apesar de no ano seguinte ter sofrido uma melhoria de 1.77%.



**Figura 18.** Percentagem de alunos reprovados por Curso e Ano Letivo.

### Número de alunos desistentes

Relativamente à última medida de sucesso para análise, a Figura 19 demonstra a evolução do número de alunos desistentes nas escolas ao longo dos seis anos letivos. As desistências, tal como detalhado no capítulo de preparação de dados, foram categorizadas em três grupos "0" (nenhum aluno desistente), " $\leq 3$ " (até 3 alunos desistentes) e " $> 3$ " (mais de 3 alunos desistentes).

Como observação inicial é importante referir que em todos os anos letivos, foi sempre maior o número de escolas sem qualquer aluno desistente do que o número de escolas com pelo menos 1 aluno desistente. Como tendência notória salienta-se que durante o ano letivo de 2015/2016, havia um total de 336 escolas que não registaram nenhuma desistência, 216 escolas com até 3 desistentes e 52 escolas com mais de 3 desistentes e que esse padrão se veio a repetir até o ano de 2018/2019, com algumas variações. O ano letivo de 2019/2020 tratou-se de um ano positivo, mas invulgar, com um aumento no número de escolas sem desistentes (428 escolas), enquanto o número de escolas com mais de 3 desistentes diminuiu significativamente (16 escolas).

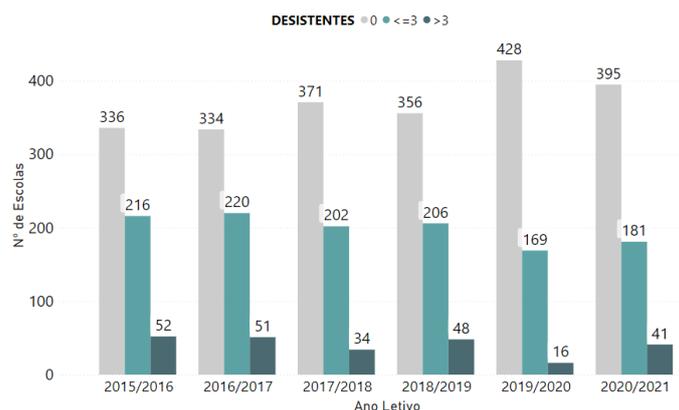


Figura 19. Número de escolas por categoria de alunos desistentes e por Ano Letivo.

### 4.3 Análise de Indicadores por escola

Após a análise das medidas de sucesso, cabe agora avaliar de forma abrangente todos os indicadores previamente calculados, estabelecendo relações entre estes e as medidas de sucesso anteriormente mencionadas. No sentido de identificar de forma precisa e fundamentada quais desses indicadores exercem influência significativa sobre o sucesso educacional, foram calculadas as correlações entre cada um dos indicadores e medidas de sucesso estudados, de acordo com o coeficiente de correlação linear de Pearson. Esta análise adotou ainda uma estratégia de formação de *clusters*, sendo este um método e um tipo de abordagem útil para identificar padrões e relações complexas entre variáveis. A formação dos *clusters*, tem como objetivo agrupar as escolas com base em características semelhantes, sendo que, neste caso, foram agrupadas escolas com base nas medidas de sucesso que mais se correlacionavam. Através da Figura 20, que fornece um *heat map* com as correlações entre todas as medidas de sucesso, é possível verificar que as medidas "Média de Exames de 1ª Fase" e "Média de Exames de 2ª Fase" foram as escolhidas pois demonstram a correlação mais significativa, uma correlação positiva de 0.77.

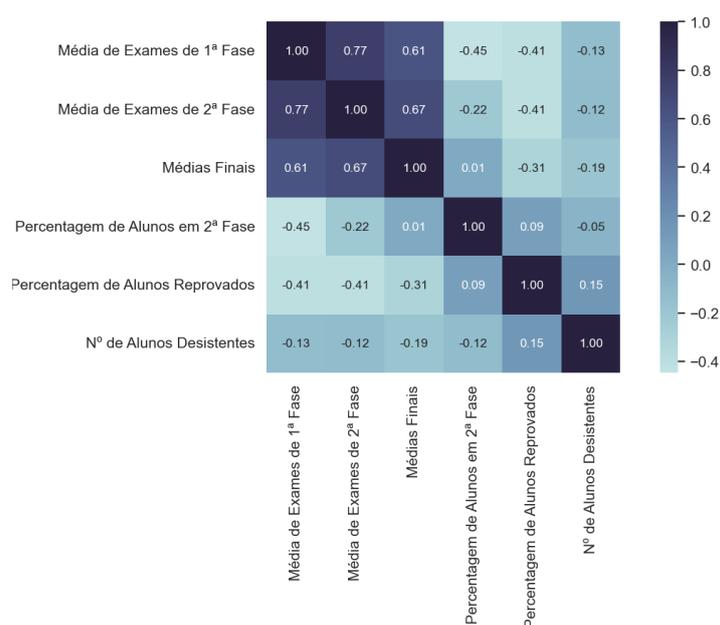
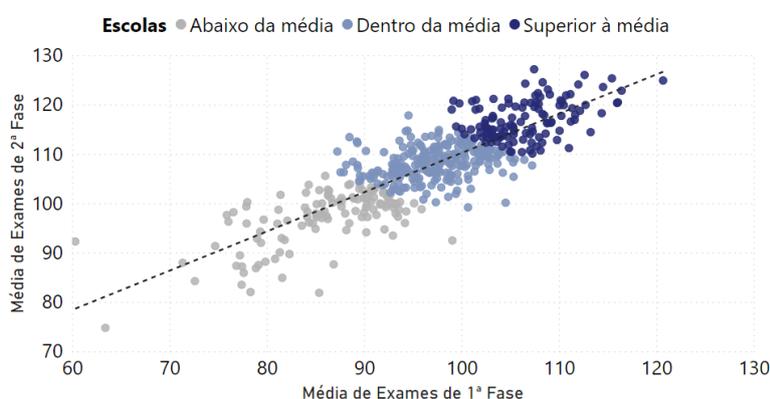


Figura 20. Heat map de correlações entre as medidas de sucesso ao nível de detalhe escolar.

Utilizou-se ainda técnicas de análise de *cluster*, como k-means, para agrupar as escolas em três categorias distintas: *cluster* "abaixo da média" que representa as escolas com um desempenho médio inferior nas duas medidas de sucesso ao longo dos anos (n=125); *cluster* "dentro da média" que representa as escolas com um desempenho médio nas duas medidas de sucesso ao longo dos anos (n=230) e *cluster* "superior à média" que se refere às escolas que demonstram um desempenho superior nas duas medidas de sucesso ao longo dos anos (n=123).

A formação desses *clusters*, tal como demonstrado na Figura 21, vem tornar possível uma análise mais precisa e detalhada das influências dos indicadores educacionais, revelando como diferentes escolas, agrupadas com base no seu desempenho, podem se relacionar com os indicadores de forma distinta.



**Figura 21.** Médias de Exames de 1ª Fase pelas Médias de Exames de 2ª Fase por escola (escala 0-200).

## Tipo de escola

Analisando o primeiro indicador, a Tabela 13 fornece as correlações entre as medidas de sucesso escolar em estudo e o indicador Tipo de escola, que faz a distinção entre a natureza das escolas, que podem ser públicas ou privadas. Esta análise visa compreender como o tipo de escola pode influenciar o desempenho dos alunos em diversos aspectos.

**Tabela 13.** Correlação do Indicador Tipo de escola com todas as medidas de sucesso

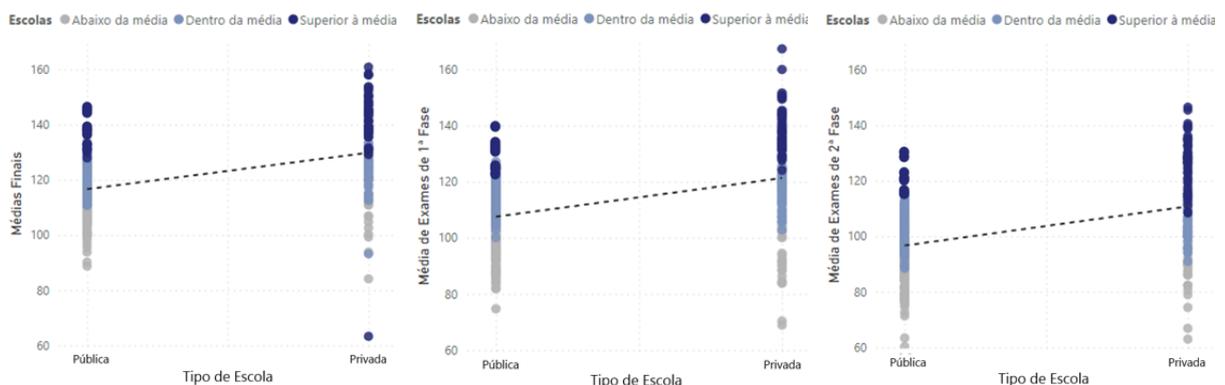
Variável	Correlação
Média de Exames de 1ª fase	0.4543
Média de Exames de 2ª fase	0.4324
Média Final	0.5166
Porcentagem de alunos reprovados	-0.0392
Porcentagem de alunos em exames de 2ª fase	0.3764
Nº de alunos desistentes	-0.123

Na análise das correlações entre o desempenho escolar e o tipo de escola, foram identificadas relações significativas. As escolas privadas apresentam não só melhores resultados na primeira fase de exames, com uma correlação positiva de 0,4543 como também na segunda fase, com uma correlação positiva de 0,4324. Além disso, as instituições privadas têm médias finais mais altas (correlação positiva de 0.5166).

No que toca ao número de alunos que vão a exames de segunda fase, apesar destas demonstrarem melhores resultados, são também as escolas privadas que mais tendem a ter alunos a repetir os seus exames, mostrando uma correlação moderada de 0.3764. No sentido de verificar visualmente estas relações, servem as Figuras 22 a 25, que destacam as diferenças entre cada tipo de escola no desempenho académico dos alunos.

A análise das correlações e dos gráficos fornecem evidências que indicam diferenças marcantes entre as escolas públicas e privadas em termos de sucesso escolar, no entanto, é fundamental ressaltar que tais observações refletem associações, e não causalidade direta.

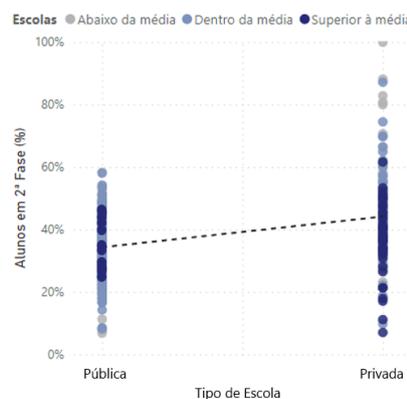
Relativamente aos *clusters*, é notável que a maior parte das escolas “superiores à média” são do tipo 1, isto é, de natureza privada, verificando-se também melhores médias finais. Já as escolas “dentro da média” e “abaixo da média” estão repartidas pelos dois tipos de escola.



**Figura 22.** Tipo de escola pelas Médias Finais (escala 0:200).

**Figura 23.** Tipo de escola pelas Médias de Exames de 1ª Fase (escala 0:200).

**Figura 24.** Tipo de escola pelas Médias de Exames de 2ª Fase (escala 0:200).



**Figura 25.** Tipo de escola pela Percentagem de Alunos em 2ª Fase.

## Tamanho da escola

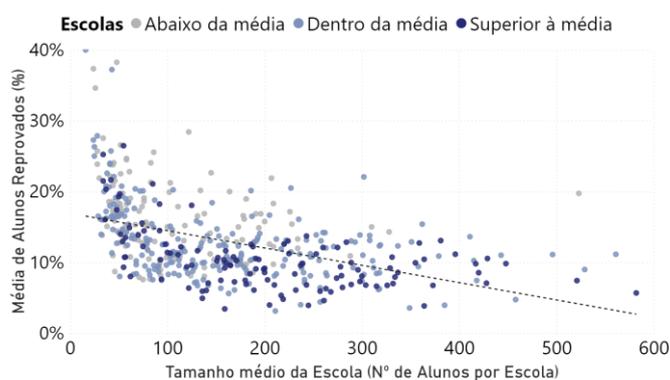
Em seguida procedeu-se à análise das relações entre o tamanho das escolas e as diversas medidas de sucesso escolar (Tabela 14), com o objetivo de compreender como o número de alunos de uma instituição educacional pode influenciar o desempenho dos próprios. Primeiramente, a correlação com maior destaque trata-se de uma correlação moderada e positiva entre o tamanho da escola e o número de alunos desistentes, com um coeficiente de correlação de 0.3139, sugerindo que, em média, escolas maiores tendem a ter um maior número de alunos que desistem da sua educação.

Já em termos de percentagens de alunos que reprovam, observa-se uma correlação negativa, embora não muito forte (-0.2662), sugerindo que um número elevado de alunos de uma escola pode ter uma influência negativa, mas pouco pronunciada, nas taxas de reprovação dos seus alunos, isto é, tendem a reprovar menos.

**Tabela 14.** Correlação do Indicador Tamanho da escola com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	0.0416
Média de Exames de 2ª fase	0.0874
Média Final	0.0334
Percentagem de alunos reprovados	-0.2662
Percentagem de alunos em exames de 2ª fase	-0.1059
Nº de alunos desistentes	0.3139

Através do gráfico de dispersão representado na Figura 26, é possível verificar a correlação negativa destacada, através da sua linha de tendência decrescente. Pode também ser feita uma análise aos *clusters*, onde a grande maioria das escolas “superiores à média” se encontram situadas abaixo da linha de tendência, sugerindo que apresentam, em média, percentagens de alunos reprovados menores que as restantes. Quanto ao número de alunos por escola, este indicador não permite diferenciar escolas “superiores à média” das restantes, uma vez que independentemente do *cluster* em questão, as escolas apresentam-se dispersas ao longo da linha de tendência.



**Figura 26.** Tamanho médio das escolas pela percentagem média de alunos reprovados ao longo dos 6 anos letivos.

Esta análise, pode também ser realizada para cada ano letivo em estudo (ver Apêndice D) no sentido de verificar e detetar se as diferenças no indicador em questão impulsionam variações na medida de sucesso. O único ano onde as variações são notórias é, uma vez mais, em 2019/2020 onde as percentagens de reprovações diminuem significativamente, mas o tamanho médio das escolas mantém-se estável.

No que toca ao número de alunos desistentes, também presente no Apêndice C deste documento, é possível visualizar o número de escolas pelo número médio de alunos por escola associado às categorias do número de alunos desistentes.

Um maior tamanho médio da escola está associado à categoria que representa um número de alunos desistentes superior. Nesse mesmo gráfico pode-se ainda visualizar os valores para os diferentes *clusters* criados, onde o destaque vai para as escolas com os resultados dos exames “superior à média”, que se apresentam em menor número quando se analisa a categoria de mais de três alunos desistentes.

### Tamanho da equipa docente

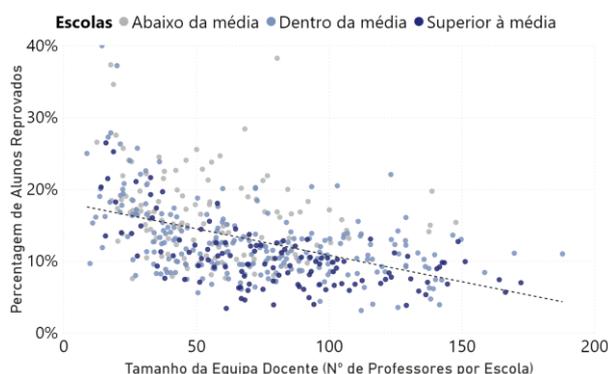
Relativamente ao tamanho da equipa docente, que é calculado através do número de professores do ensino secundário em cada escola, a correlação que mais se destaca é a percentagem de alunos reprovados. O valor negativo (-0.2478) presente na Tabela 15 indica que a diminuição do número de reprovações está associada a um maior número de professores, isto é, quanto maior o tamanho da equipa docente, em número, menor o número de alunos reprovados face aos matriculados.

A correlação entre a média de exames de segunda fase também demonstra uma correlação mais impactante que as restantes, desta vez positiva (0.2000), sugerindo que o número de professores está associado ao aumento das médias dos exames que são repetidos. Estas correlações destacam a importância da gestão dos professores por parte das escolas, que precisa de ser adaptada às necessidades dos alunos.

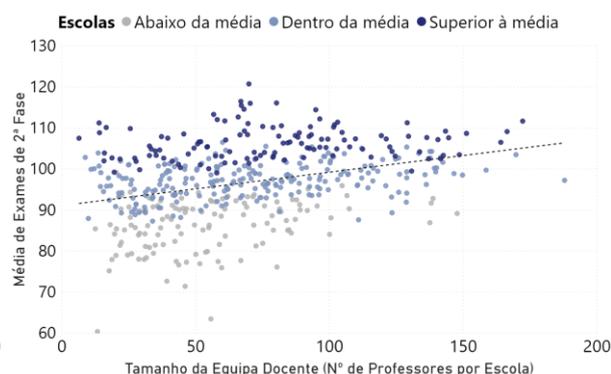
**Tabela 15.** Correlação do Indicador Tamanho da equipa docente com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	0.1444
Média de Exames de 2ª fase	0.2000
Média Final	0.1587
Percentagem de alunos reprovados	-0.2478
Percentagem de alunos em exames de 2ª fase	-0.1152
Nº de alunos desistentes	0.1006

Relativamente aos *clusters* tanto a Figura 28 como a Figura 29 indicam que não existe diferença notória entre o número de professores de cada um. Apesar da percentagem de alunos reprovados por escola diminuir e a média de exames de segunda fase aumentar ao longo da linha de tendência de cada gráfico, o tamanho médio da equipa docente mantém-se disperso em cada um dos três *clusters*.



**Figura 27.** Tamanho médio da equipa docente pela média de alunos reprovados (%), ao longo dos 6 anos letivos.



**Figura 28.** Tamanho médio da equipa docente pela Média de Exames de 2ª Fase, ao longo dos 6 anos letivos (escala 0-200).

Quando a análise é feita para cada ano letivo, os gráficos fornecidos no Apêndice E, não fornecem flutuações significativas neste indicador que fizessem alterar o sucesso da percentagem de alunos reprovados, mantendo-se constante em todos os *clusters* criados.

### **Rácio tamanho da escola face ao tamanho da equipa docente**

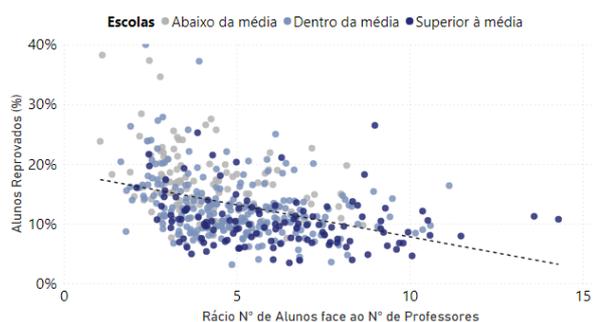
O rácio entre o tamanho da escola face ao tamanho da equipa docente quantifica o número de alunos para cada professor, isto é, quanto maior for o rácio, menor o número de professores para cada estudante. Quando se relaciona as medidas de sucesso com este indicador, não existem correlações fortes e significativas (Tabela 16), no entanto, a percentagem de alunos reprovados e a média de exames de primeira fase apresentam as maiores correlações. A primeira, tratando-se de uma correlação negativa (-0.1619) sugere que, em média, quanto maior o número de alunos face aos seus professores, a percentagem de alunos reprovados tende a diminuir. Este valor vem contrariar a correlação verificada no ponto anterior, onde se conclui que quanto maior o número de professores, menor a percentagem de alunos que reprovam. Uma possível interpretação para esta incoerência de valores poderá ser que para atingir bons resultados nesta medida de sucesso, não basta apenas ter um elevado número de professores a lecionar numa escola sendo também importante ter em consideração o número de alunos que nela estudam.

Quanto à correlação com a média de exames de primeira fase (0.1296), o valor positivo, embora não muito forte, indica que quanto maior for o rácio, maiores serão também as médias obtidas nesse tipo de avaliação.

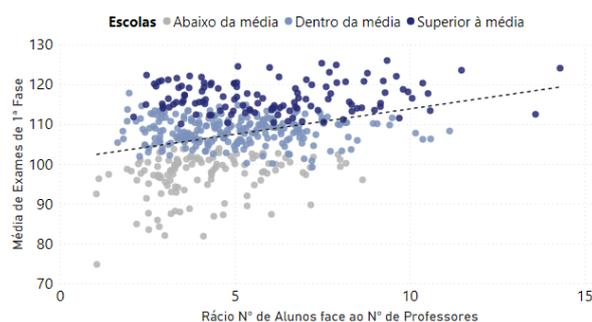
**Tabela 16.** Correlação do Indicador Rácio tamanho da escola face ao tamanho da equipa docente com todas as medidas de sucesso.

<b>Variável</b>	<b>Correlação</b>
Média de Exames de 1ª fase	0.1296
Média de Exames de 2ª fase	0.1037
Média Final	0.0658
Percentagem de alunos reprovados	-0.1619
Percentagem de alunos em exames de 2ª fase	-0.0468
Nº de alunos desistentes	0.1213

Os gráficos de dispersão representados nas Figuras 30 e 31 vêm justificar os valores destacados na tabela anterior, fornecendo uma linha de tendência decrescente no caso da medida de sucesso de percentagem de alunos reprovados, e uma linha crescente no caso da média de exames de primeira fase. Os *clusters* neles incorporados indicam que o rácio entre o número de alunos face ao número de professores não é suficiente para distinguir escolas “superior à média” de escolas “abaixo da média” uma vez que estas se encontram dispersas pelo gráfico em questão, sem nenhum tipo de agrupamento ou tendência.



**Figura 29.** Rácio do tamanho da escola face ao tamanho da equipa docente por percentagem de alunos reprovados, ao longo dos 6 anos letivos.



**Figura 30.** Rácio do tamanho da escola face ao tamanho da equipa docente pela Média de Exames de 1ª Fase, ao longo dos 6 anos letivos (escala 0-200).

### Média do número de alunos por turma

Quanto ao indicador que representa o número médio de alunos por turma em cada escola, segundo as correlações calculadas (Tabela 17), tanto as medidas de sucesso que dizem respeito às médias de exames como as médias finais do ensino secundário, fornecem correlações negativas de  $-0.1700$ ,  $-0.1265$  e  $-0.1308$ , respetivamente. Estes valores sugerem que, em média, à medida que o tamanho médio das turmas aumenta, as notas dos alunos tendem a diminuir, diminuindo assim as médias dessas avaliações escolares. A tendência, portanto, é as escolas secundárias, com turmas mais pequenas, apresentarem melhores resultados nesse tipo de avaliações.

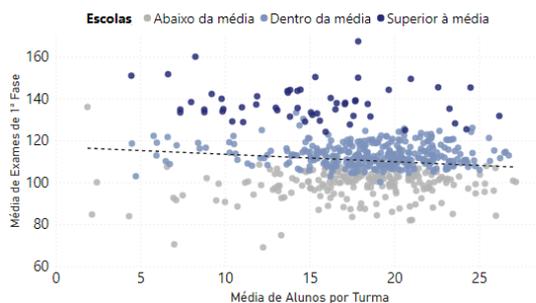
No que diz respeito à percentagem de alunos que reprovam, existe uma correlação positiva significativa entre esse indicador e o tamanho médio das turmas, provando que, em média, as escolas com turmas mais numerosas tendem a ter mais alunos reprovados.

**Tabela 17.** Correlação do Indicador Média do número de alunos com todas as medidas de sucesso.

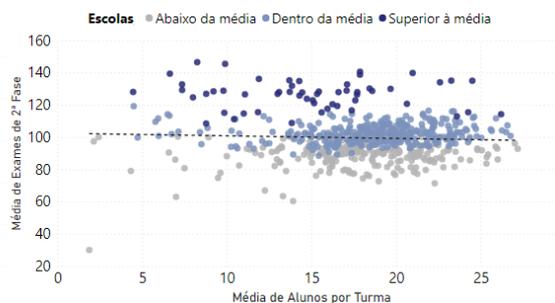
Variável	Correlação
Média de Exames de 1ª fase	$-0.1700$
Média de Exames de 2ª fase	$-0.1265$
Média Final	$-0.1308$
Percentagem de alunos reprovados	$0.1404$
Percentagem de alunos em exames de 2ª fase	$0.0241$
Nº de alunos desistentes	$0.0475$

Estes valores podem ser traduzidos para um formato visual (Figuras 32 a 35), através dos gráficos de dispersão que fornecem as linhas de tendência para cada indicador discutido.

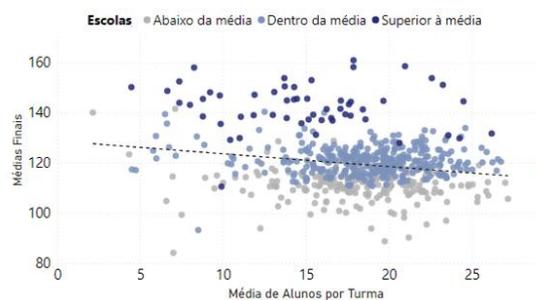
Todos os *clusters* apresentam valores médios de alunos por turma dispersos, verificando-se somente uma certa tendência de que um grande número de escolas opta por agrupar os seus alunos em turmas de 15 a 25 estudantes.



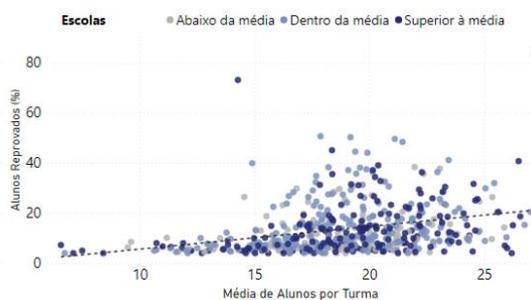
**Figura 31.** Média de Alunos por Turma pelas Médias de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 32.** Média de Alunos por Turma pelas Médias de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 33.** Média de Alunos por Turma pelas Médias Finais por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 34.** Média de Alunos por Turma pela percentagem de alunos reprovados, ao longo dos 6 anos letivos.

A relação entre a média de alunos por turma e as médias de exames de primeira fase, em cada ano letivo, encontra-se representada no Apêndice F deste documento no sentido de conseguir também detetar mudanças repentinas deste indicador, que possam influenciar a medida de sucesso em causa. De acordo com as análises realizadas, não ocorre nenhuma flutuação no decorrer dos anos, pelo que, com exceção do ano de 2019/2020, as médias de exames escolares em primeira fase mantêm-se relativamente constantes.

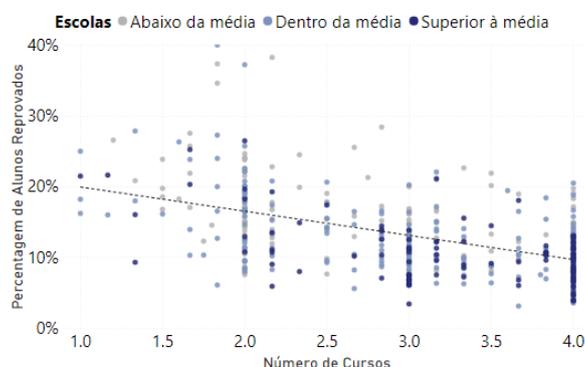
### Número de cursos disponibilizados

Os resultados da análise de correlações não mostram muitos valores notáveis entre as medidas de sucesso e número de cursos disponíveis em cada escola. Através da Tabela 18, verifica-se que a correlação mais significativa é a correlação negativa moderada (-0.2154), com a percentagem de alunos reprovados, que indica que a redução no número de cursos disponíveis impactua positivamente nesta medida, isto é, quanto maior o número de cursos disponibilizados, menor a percentagem de alunos reprovados. Relativamente à percentagem de alunos que estiveram presentes em exames nacionais de segunda fase, a correlação negativa (-1.372) mostra que o aumento do número de cursos disponíveis está associado à diminuição dessa medida de sucesso. Em suma, aumentar a variedade de cursos disponíveis, parece ser uma estratégia de sucesso, embora não muito forte, no sentido de reduzir a percentagem de reprovações. No entanto é importante salientar que estas relações não apresentam causalidade direta pelo que outros fatores devem também ser considerados.

**Tabela 18.** Correlação do Indicador Número de Cursos disponibilizados com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	0.0399
Média de Exames de 2ª fase	0.0622
Média Final	0.0724
Percentagem de alunos reprovados	-0.2154
Percentagem de alunos em exames de 2ª fase	-0.1372
Nº de alunos desistentes	-0.0982

Os gráfico de dispersão que justifica a informação relativa aos indicador com maior correlação acima mencionado (Figura 36) permite também verificar que as escolas agrupadas com resultados dos exames “superior à média” tendem a disponibilizar um maior número de cursos. Esta tendência verifica-se também quando é feita uma análise ano a ano (ver Apêndice G).



**Figura 35.** Número de cursos disponibilizados pela percentagem de alunos reprovados por escola, ao longo dos 6 anos letivos.

### Número de matrículas anuladas

No que toca ao número de matrículas anuladas por escola, este indicador apenas apresenta correlações significativas com a medida de sucesso relativa ao número de alunos desistentes (Tabela 19). Anular uma matrícula difere de desistir da educação, na medida em que o aluno pode anular uma matrícula em qualquer disciplina que pretenda e ainda prosseguir e concluir os seus estudos, no entanto, é apenas avaliado nesse ano pelos seus resultados externos, ou seja, cem por cento da nota da disciplina anulada estará dependente dos resultado obtido no exame da mesma. A correlação verificada entre este indicador e a medida de sucesso especificada (0.2074) indica que quanto maior o número de matrículas anuladas numa escola, maior será a tendência para que os seus alunos desistam da própria educação.

Este tipo de análise permite dar destaque e importância aos fatores não apenas associados aos resultados finais dos alunos, sendo também necessário entender a principal razão destes valores, através de uma abordagem holística, no sentido de promover o sucesso académico e reduzir a taxa de desistência educacional.

**Tabela 19.** Correlação do Indicador Número de matrículas anuladas por escola com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	-0.0792
Média de Exames de 2ª fase	-0.0193
Média Final	-0.0316
Percentagem de alunos reprovados	0.0219
Percentagem de alunos em exames de 2ª fase	0.0583
Nº de alunos desistentes	0.2074

No sentido de visualizar esta correlação e os *clusters* criados, sendo a medida de sucesso “Nº de alunos desistentes” e o indicador “Nº de matrículas anuladas” variáveis categóricas, optou-se por uma abordagem de quantificar o número de escolas por cada categoria criada de matrículas anuladas e por cada categoria criada de alunos desistentes (ver Apêndice H).

Os gráficos de linhas representados fornecem, portanto, a relação entre as variáveis descritas, sendo possível indicar que escolas “superiores à média”, contrariamente às escolas “abaixo da média”, se contabilizam em menor número em todas as categorias de número de alunos desistentes, quando comparadas por este indicador. Isto é, escolas com melhores superiores à média tendem a ser menos quando avaliadas pelo número de alunos desistentes, sabendo a categoria de número de alunos com matrículas anuladas

### Número de alunos transferidos

Relacionando o número de alunos transferidos com as medidas de sucesso em estudo, a única correlação que se destaca na Tabela 20 é entre o indicador em questão e o número de alunos desistentes, com uma correlação positiva, embora não muito forte de 0.1082. Tal como na análise mencionada no ponto anterior, esta correlação significa que quanto maior o número de alunos transferidos verificado numa escola, maior a tendência de o número de alunos desistentes crescer.

**Tabela 20.** Correlação do Indicador Número alunos transferidos por escola com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	-0.0336
Média de Exames de 2ª fase	-0.0103
Média Final	-0.0126
Percentagem de alunos reprovados	0.0613
Percentagem de alunos em exames de 2ª fase	-0.0232
Nº de alunos desistentes	0.1082

A correlação discutida pode ainda ser visualizada através dos gráficos presentes no Apêndice I. Os gráficos demonstram o número de escolas para cada categoria de número de alunos transferidos e de número de alunos desistentes. Quanto maior o valor do indicador, maior será o impacto no número de desistentes de uma escola. Relativamente aos *clusters*, é possível aferir que são menos as escolas que apresentam médias de exames “superior à média” quando relacionadas com o número de alunos desistentes e de transferidos.

## Fatores socioeconómicos dos estudantes

Os fatores socioeconómicos dos estudantes de uma escola do ensino secundário, medidos através do escalão ASE dos alunos, enquadram-se como um dos indicadores de sucesso investigados e sugeridos na etapa de revisão de literatura deste estudo. Quando relacionado com as medidas de sucesso, surgem correlações às quais se pode tirar algumas conclusões (Tabela 21): relativamente à correlação com as médias obtidas, tanto dos exames nacionais de ambas as fases como das médias finais, ao longo dos anos letivos, verifica-se que estão inversamente relacionadas (-0.3802, -0.4166 e -0.25). Estes resultados indicam que os alunos com melhores condições socioeconómicas (com um escalão ASE menor) tendem a alcançar resultados mais altos nas avaliações em questão.

No que toca à percentagem de alunos reprovados de uma escola, a correlação positiva de 0.2133 indica que, à medida que os fatores socioeconómicos dos estudantes melhoram, a percentagem de alunos reprovados diminui. Isso sugere que escolas que têm estudantes com condições socioeconómicas mais favoráveis têm uma taxa média de reprovação menor.

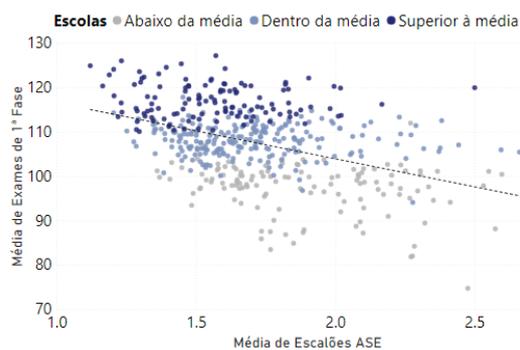
Como resumo, os resultados da análise destacam uma influência substancial dos fatores socioeconómicos dos estudantes de uma escola no seu desempenho académico, onde escolas que têm alunos associados a escalões ASE mais baixos tendem a obter melhores médias externas e finais diminuindo a probabilidade de reprovar. Estas descobertas revelam a importância das políticas e das intervenções que pretendem reduzir as disparidades socioeconómicas na educação do ensino secundário, proporcionando oportunidades mais equitativas para todos os alunos.

**Tabela 21.** Correlação do Indicador Média dos Fatores Socioeconómicos dos estudantes por escola com todas as medidas de sucesso.

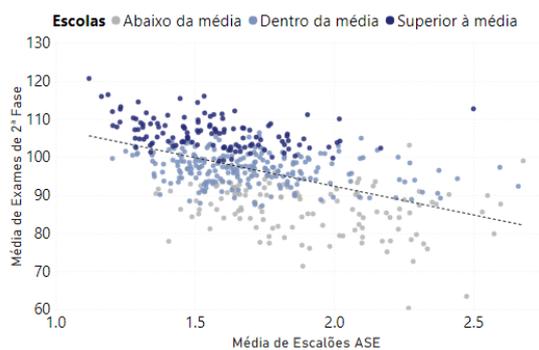
Variável	Correlação
Média de Exames de 1ª fase	-0.3802
Média de Exames de 2ª fase	-0.4166
Média Final	-0.25
Percentagem de alunos reprovados	0.2133
Percentagem de alunos em exames de 2ª fase	0.0703
Nº de alunos desistentes	-0.082

Os gráficos de dispersão representados nas Figuras 37, 38, 39 e 40 avaliam as relações entre as medidas de sucesso destacadas e o indicador em questão, alinhando também as mesmas com os *clusters* de escolas criados.

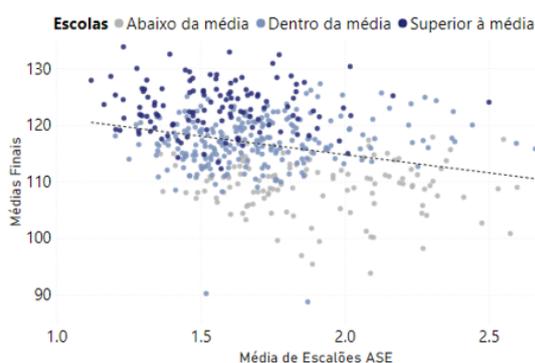
Algumas das escolas agrupadas como “Abaixo da média” e “Dentro da média” revelam uma análise interessante ao se localizarem em zonas dos gráficos mais à direita, onde os escalões ASE são superiores, quando comparadas com as escolas “Superior à média”.



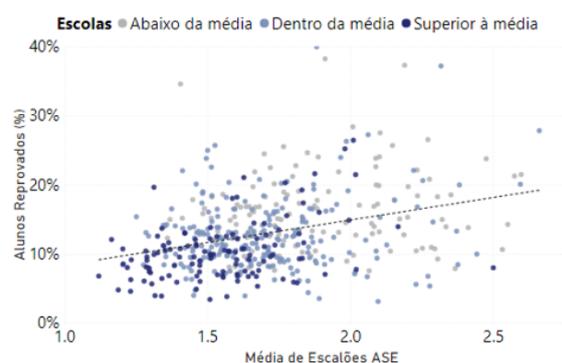
**Figura 36.** Média de Escalões ASE pelas Médias de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 37.** Média de Escalões ASE pelas Médias de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 38.** Média de Escalões ASE pelas Médias Finais por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 39.** Média de Escalões ASE pela percentagem de alunos reprovados por escola, ao longo dos 6 anos letivos.

Disponibilizando uma análise por ano letivo (ver Apêndice J), o resultado é idêntico, uma vez que escolas incluídas no *cluster* com melhores médias de exames apresentam, ano após ano, melhor nível de escalões ASE ao contrário das escolas “abaixo da média”, que educam alunos, em média, com piores condições socioeconómicas.

### Nível médio de qualificação dos professores

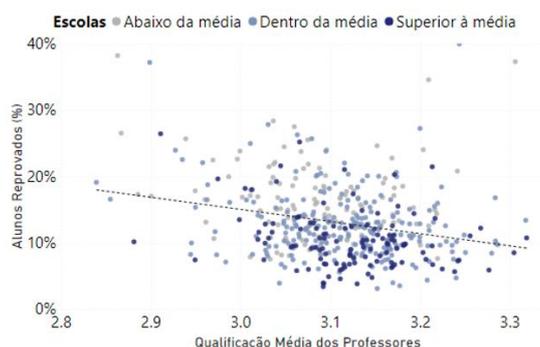
Um professor para lecionar no ensino secundário precisa naturalmente de apresentar um certo nível de qualificação no sentido de conseguir prestar o melhor ensino aos seus alunos. Apesar das escolas em Portugal sentirem cada vez mais a falta de professores, a Tabela 22 vem demonstrar que a qualificação dos mesmos necessita também de estar de acordo com o ensino prestado. Atualmente, segundo novas atualizações, para combater a ausência de docentes em algumas escolas, o Ministério da Educação tem vindo a baixar os requisitos mínimos de qualificação para poder exercer a função de docente. Estas implementações podem indiretamente vir a diminuir aquilo que é o sucesso académico dos alunos.

A correlação negativa (-0.2193) verificada entre este indicador e a percentagem de alunos reprovados demonstra que, em média, quanto melhor for a qualificação dos professores, menor será essa taxa de reprovação. Quanto aos resultados médios verificados nos exames nacionais, as correlações de 0.1589 de primeira fase e de 0.1496 de segunda fase, partilham o mesmo tipo de importância quanto ao nível de qualificação dos professores, uma vez que tendencialmente, escolas com docentes melhor qualificados tendem a ter melhores médias nesse tipo de avaliações.

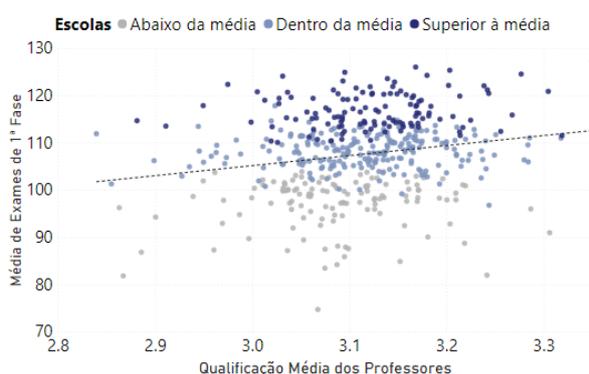
**Tabela 22.** Correlação do Indicador nível de qualificação média dos professores por escola com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	0.1589
Média de Exames de 2ª fase	0.1496
Média Final	0.0989
Porcentagem de alunos reprovados	-0.2193
Porcentagem de alunos em exames de 2ª fase	-0.0775
Nº de alunos desistentes	-0.0186

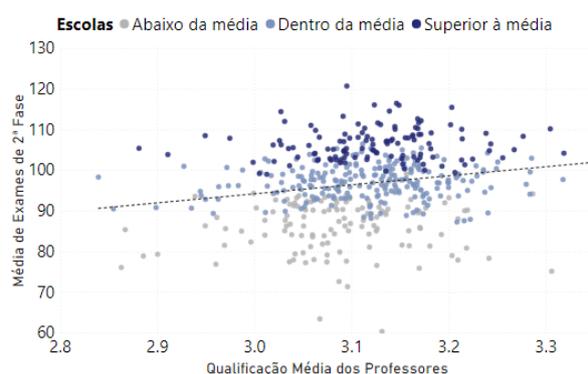
O gráfico de dispersão representado na Figura 41, no sentido de justificar a primeira correlação mencionada, demonstra uma linha de tendência decrescente, ao contrário das Figuras 42 e 43 que fornecem gráficos para as duas últimas correlações destacadas, cujas curvas de tendência são crescentes. Ao proceder à análise dos *clusters* formados, a média de anos de experiência não tende a apresentar influência na distinção entre escolas “superiores à média” das restantes. Os registos dos três *clusters* em questão estão dispersamente localizados nos gráficos sendo apenas possível indicar que as escolas tendem a ter professores cuja sua qualificação se situa entre os 3.0 e os 3.2, isto é, com o grau de formação de licenciatura.



**Figura 40.** Qualificação Média dos Professores pela percentagem de Alunos Reprovados por escola, ao longo dos 6 anos letivos.



**Figura 41.** Qualificação Média dos Professores pela Média de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 42.** Qualificação Média dos Professores pela Média de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos, (escala 0-200).

Ao longo dos anos letivos, analisando os gráficos presentes no Apêndice K, não se apuram variações nas qualificações médias dos professores das escolas de qualquer *cluster*, não sendo, por isso, um indicador capaz de distinguir escolas “superior à média” de escolas “abaixo da média”, quando a análise é feita unicamente por este.

## Média de anos de experiência dos professores

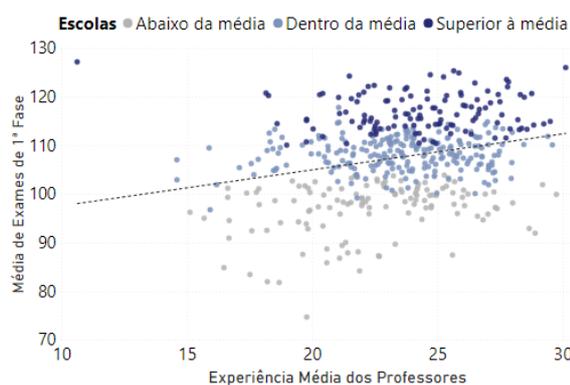
Segundo a Tabela 23, as correlações verificadas indicam também que o número médio de anos de experiência dos professores de uma escola tem impacto em algumas das medidas de sucesso estudadas.

As correlações com mais destaque surgem entre o indicador em estudo e as medidas de sucesso relativas às médias obtidas nos exames de primeira e segunda fase, valores como 0.2254 e 0.2477, respetivamente. Esta análise refere que quanto maior for a experiência dos professores, melhor serão, em média, os resultados obtidos nessas medidas.

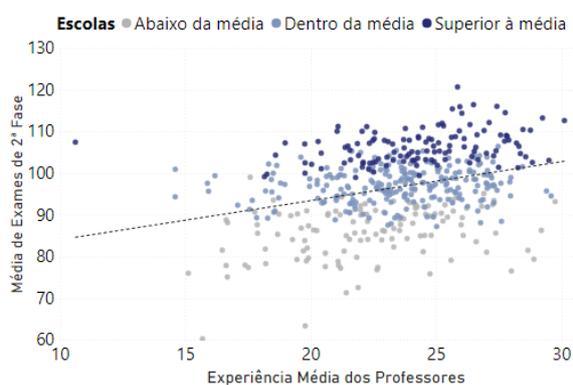
**Tabela 23.** Correlação do Indicador Média de anos de experiência dos professores com todas as medidas de sucesso.

Variável	Correlação
Média de Exames de 1ª fase	0.2254
Média de Exames de 2ª fase	0.2477
Média Final	0.1141
Percentagem de alunos reprovados	-0.1052
Percentagem de alunos em exames de 2ª fase	-0.1037
Nº de alunos desistentes	0.0215

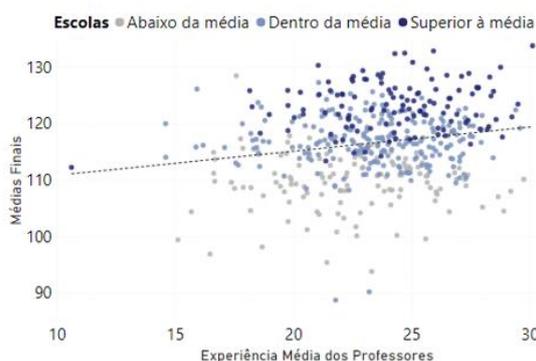
Os *clusters* criados no início desta seção, representados nas Figuras 44, 45 e 46 traduzem as correlações destacadas em cima através das linhas de tendência crescentes, e decrescentes no caso da Figura 47. É possível indicar que as escolas que prestam o ensino secundário tendem a reter professores cujo número médio de anos de experiência se situa entre os 20 e os 30 anos e que não existe uma diferença notória deste indicador quando se compara escolas “superiores à média” de escolas “abaixo da média”, estando estes valores dispersos ao longo da linha de tendência.



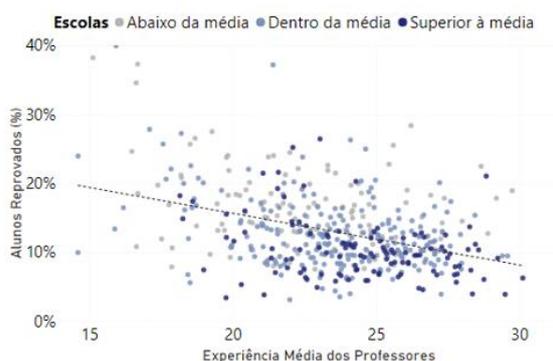
**Figura 43.** Anos de Experiência Média dos Professores pelas Médias de Exames de 1ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 44.** Anos de Experiência Média dos Professores pelas Médias de Exames de 2ª Fase por escola, ao longo dos 6 anos letivos (escala 0-200).



**Figura 45.** Anos de Experiência Média dos Professores pelas Médias Finais por escola, ao longo dos 6 anos letivos (escala 0-200).



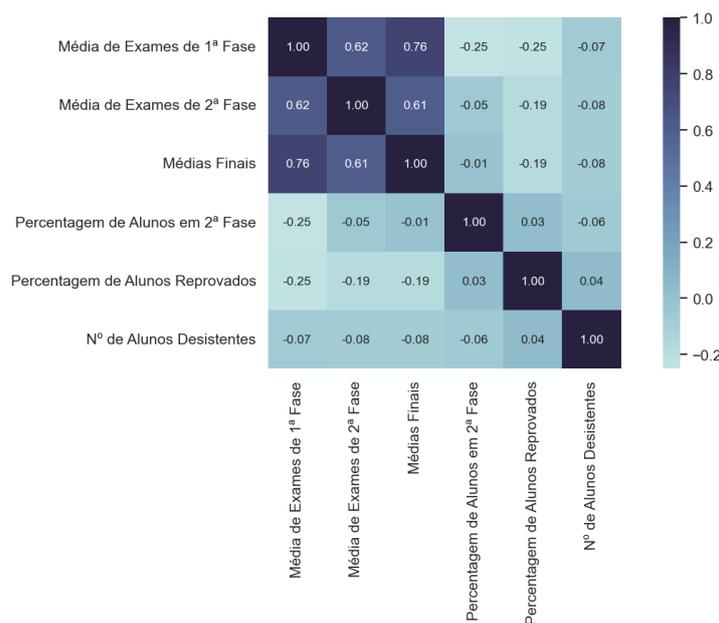
**Figura 46.** Anos de Experiência Média dos Professores pela percentagem de Alunos Reprovados por escola, ao longo dos 6 anos letivos.

Os gráficos de dispersão que relacionam a medida de sucesso com maior correlação com este indicador (média de exames de primeira fase) para cada ano letivo permitem ainda detetar que não existem flutuações na experiência média dos professores de cada *cluster*, no decorrer do tempo (ver Apêndice L). Em todos os seis anos o indicador mantém-se estável sem discrepância entre cada *cluster*.

### 4.3 Análise Indicadores por curso

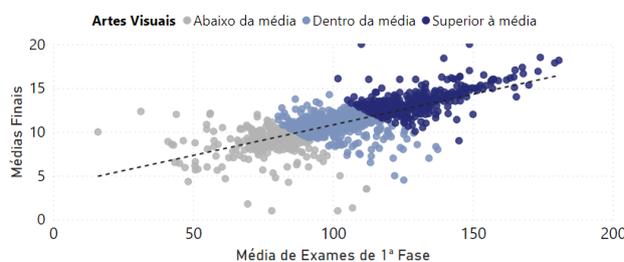
O foco central desta secção passa agora por analisar e detalhar as relações entre todos os indicadores de sucesso calculados para o nível de detalhe por curso, relacionando-os, uma vez mais, com todas as medidas de sucesso. Para a criação de *clusters*, com o apoio visual de um *heat map* (Figura 48), utilizou-se a mesma abordagem que no nível de detalhe escolar, no sentido de conseguir agrupar os cursos das várias escolas com mais semelhanças, escolhendo a correlação com mais significância entre as correlações de todas as medidas de sucesso.

A maior correlação verificada, entre as Médias de Exames de 1ª Fase e Médias Finais, foi, portanto, a escolhida para o agrupamento dos vários cursos das escolas secundárias em estudo, com uma correlação positiva de 0.76.

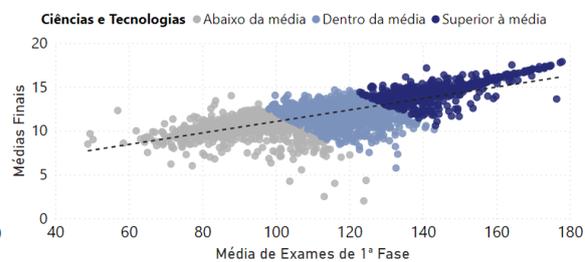


**Figura 47.** Heat map de correlações entre as medidas de sucesso ao nível de detalhe por curso.

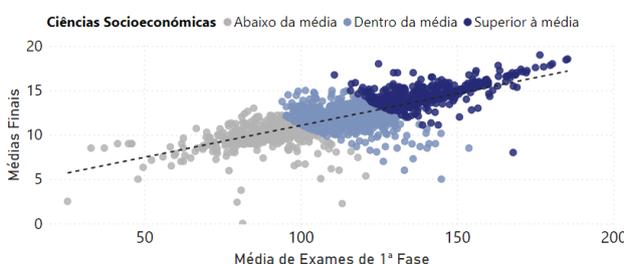
Os *clusters* foram ainda separados por curso, no sentido de auxiliar a visualização dos dados e conseguir obter resultados mais filtrados e detalhados. Seguindo o mesmo procedimento da secção anterior, adotou-se uma estratégia de agrupar os diferentes cursos das diferentes escolas por três *clusters* (“Abaixo da média”, “Dentro da média” e “Acima da média”) quando analisados pelas duas variáveis selecionadas em cima, como se pode verificar através dos gráficos de dispersão das Figuras 49 a 52.



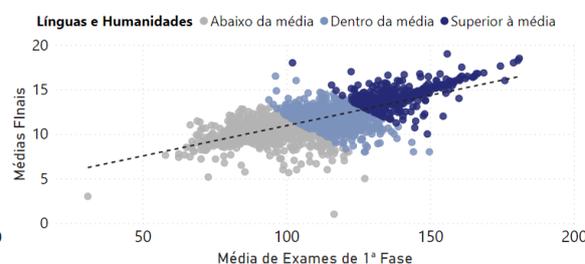
**Figura 48** – Média de Exames de 1ª Fase pelas Médias Finais do curso de Artes Visuais, ao longo dos 6 anos letivos.



**Figura 49** – Média de Exames de 1ª Fase pelas Médias Finais do curso de Ciências e Tecnologias, ao longo dos 6 anos letivos.



**Figura 50.** – Média de Exames de 1ª Fase pelas Médias Finais do curso de Ciências Socioeconómicas, ao longo dos 6 anos letivos.



**Figura 51.** – Média de Exames de 1ª Fase pelas Médias Finais do curso de Línguas e Humanidades, ao longo dos 6 anos letivos.

Os *clusters* relativos aos cursos de Artes Visuais, 350 dos mesmos são classificados como “superior à média”, 706 como “dentro da média” e 301 como “abaixo da média” quando são analisados tendo em conta as médias de exames de 1º Fase e as médias finais. Nos cursos de Ciências e Tecnologias, 696 enquadram-se como “superior à média”, 1554 como “dentro da média” e 1195 como “abaixo da média”. No que toca aos *clusters* dos cursos de ciências socioeconómicas, 505 são “superiores à média”, 1112 estão “dentro da média” e 559 “abaixo da média”. Por fim, nos cursos de Línguas e Humanidades, 467 foram agrupados como apresentarem resultados “superior à média”, 1303 apresentam-se “dentro da média” e 1238 “abaixo da média”.

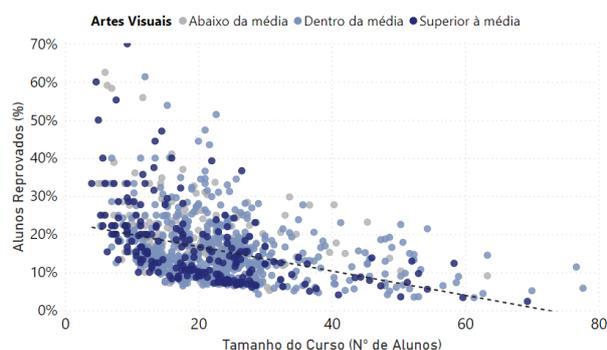
### Tamanho do Curso

O tamanho do curso, sendo o primeiro indicador de sucesso a analisar dentro do nível de detalhe por curso, é calculado através da soma do número de alunos para cada curso em cada escola. Este indicador apenas apresenta uma correlação relevante quando se relaciona com a medida de sucesso de percentagem de alunos reprovados por curso, com um valor de correlação de -0.2335 (Tabela 24). Embora fraco, este valor sugere que quanto maior o número de alunos de um curso, menor será a percentagem de alunos reprovados.

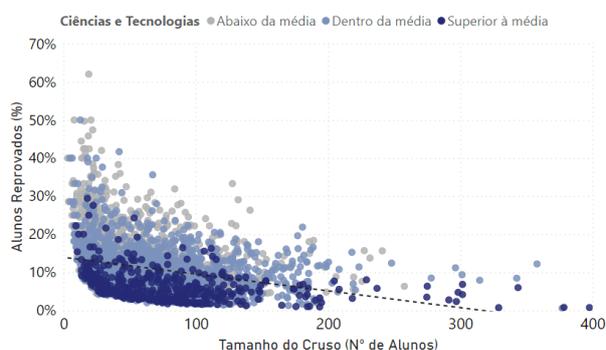
**Tabela 24.** Correlação do Indicador Tamanho do Curso com as medidas de sucesso em análise ao nível de detalhe por curso.

Variável	Correlação
Média de Exames de 1ª fase	0.0844
Média de Exames de 2ª fase	0.0333
Média Final	0.0572
Percentagem de alunos reprovados	-0.2335
Percentagem de alunos em exames de 2ª fase	-0.1332
Nº de alunos desistentes	0.0907

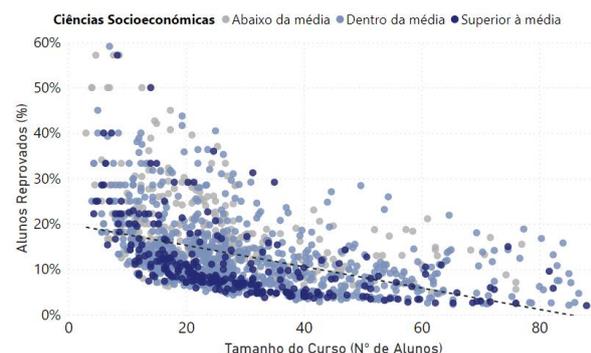
Os gráficos de dispersão criados, quando impostos e analisados perante a relação do indicador em questão com a medida de sucesso destacada, sugerem que ambos os cursos contribuem e apresentam o mesmo nível de coeficiente de correlação, ao apresentarem linhas de tendência decrescentes (Figuras 53 a 56). Quanto aos *clusters* formados anteriormente, é possível indicar que os cursos “superiores à média” quando avaliados pelas médias de exame de 1ª fase e médias finais, tendem a apresentar um menor número de alunos reprovados, isto é, abaixo da linha de tendência.



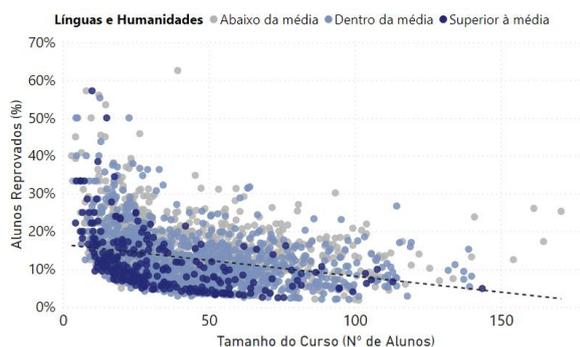
**Figura 52.** Tamanho dos cursos de Artes Visuais por escola pela percentagem de alunos reprovados, ao longo dos 6 anos letivos.



**Figura 53.** Tamanho dos cursos de Ciências e Tecnologias por escola pela percentagem de alunos reprovados, ao longo dos 6 anos letivos.



**Figura 54.** Tamanho dos cursos de Ciências Socioeconómicas por escola pela percentagem de alunos, ao longo dos 6 anos letivos.



**Figura 55.** Tamanho dos cursos de Línguas e Humanidades por escola pela percentagem de alunos reprovados, ao longo dos 6 anos letivos.

### Número de matrículas anuladas e Número de alunos transferidos

Os indicadores que fornecem o número de matrículas anuladas e o número de alunos transferidos por curso em cada escola, segundo a Tabela 25 e Tabela 26, respetivamente, não apresentam qualquer correlação significativa com nenhuma das medidas de sucesso analisadas ao nível de detalhe por curso.

Estes valores sugerem que quando se pretende prever ou justificar o sucesso académico de um curso de uma certa escola, os indicadores referenciados, segundo os dados disponibilizados, não aparentam ser suficientes para o fazer ou não tendem a ter qualquer efeito de influência no mesmo.

**Tabela 25.** Correlação do Indicador Número de matrículas anuladas com as medidas de sucesso em análise ao nível de detalhe por curso.

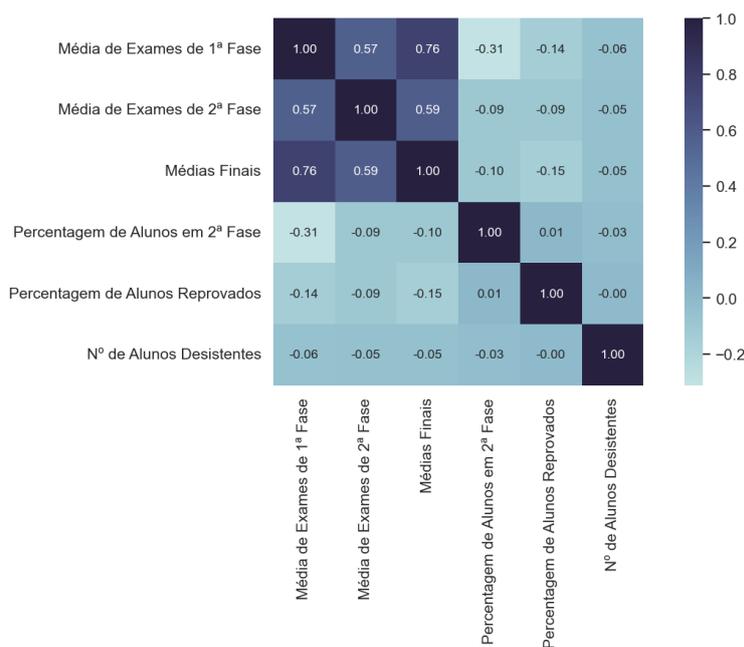
Variável	Correlação
Média de Exames de 1ª fase	-0.0443
Média de Exames de 2ª fase	-0.037
Média Final	-0.0148
Percentagem de alunos reprovados	0.0026
Percentagem de alunos em exames de 2ª fase	0.0263
Nº de alunos desistentes	0.0197

**Tabela 26.** Correlação do Indicador Número de alunos transferidos com as medidas de sucesso em análise ao nível de detalhe por curso.

Variável	Correlação
Média de Exames de 1ª fase	-0.0183
Média de Exames de 2ª fase	-0.0344
Média Final	-0.0082
Percentagem de alunos reprovados	-0.0081
Percentagem de alunos em exames de 2ª fase	-0.0112
Nº de alunos desistentes	0.0384

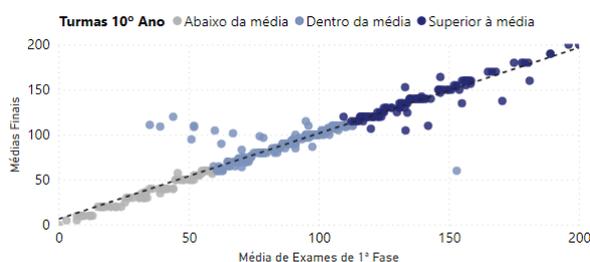
#### 4.4 Análise Indicadores por turma

Esta secção conduz a uma última análise detalhada das relações entre os indicadores de desempenho calculados e as medidas de sucesso, desta vez para o nível de detalhe por turma. Para a criação dos *clusters*, com o auxílio de uma representação visual em forma de *heat map* (Figura 57), adotou-se a mesma abordagem referida nas duas últimas secções, no sentido de agrupar turmas com características semelhantes. Neste processo, foi selecionada a correlação mais significativa entre as diversas medidas de sucesso. A correlação que apresenta mais destaque é a que relaciona as médias dos exames da 1ª fase e as médias finais, com uma correlação positiva de 0,76. Essa correlação foi, portanto, a escolhida para agrupar as diversas turmas em estudo nas escolas secundárias.

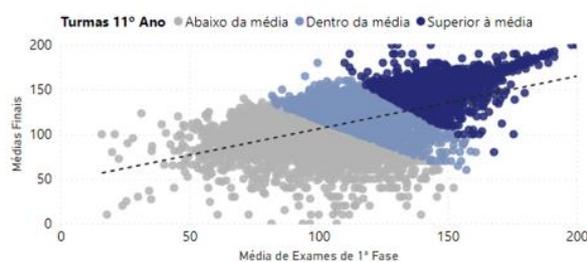


**Figura 56 – Heat map** de correlações entre as medidas de sucesso ao nível de detalhe por turma.

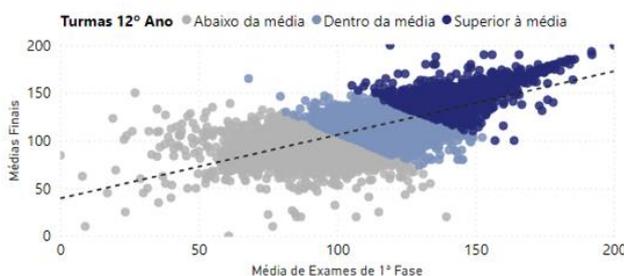
Para a formação dos *clusters*, fez-se ainda uma distinção das turmas pelos três anos de ensino que a educação secundária em Portugal inclui, de forma a ter um nível de detalhe superior. Para cada um dos anos em questão foram agrupadas as turmas com base nas medidas de sucesso especificadas (Figura 58, 59 e 60), tendo sido obtidas para o 10º ano 84 turmas “abaixo da média”, 111 turmas “dentro da média” e 180 turmas “acima da média”. Os valores para os anos seguintes são bastante superiores uma vez que são estes que tendem a fazer mais exames: no 11º ano, 5091 turmas estão “abaixo da média”, 7690 das turmas “dentro da média” e 3338 turmas têm médias combinadas por essas medidas de sucesso “superiores à média”; para o 12º ano, 5555 turmas estão “abaixo da média”, 6951 “dentro da média” e as restantes 3175 são “superiores à média”.



**Figura 57** - Média de Exames de 1ª Fase pelas Médias Finais das turmas de 10º ano, ao longo dos 6 anos letivos.



**Figura 58**. Média de Exames de 1ª Fase pelas Médias Finais das turmas de 11º ano, ao longo dos 6 anos letivos.



**Figura 59**. Média de Exames de 1ª Fase pelas Médias Finais das turmas de 12º ano, ao longo dos 6 anos letivos.

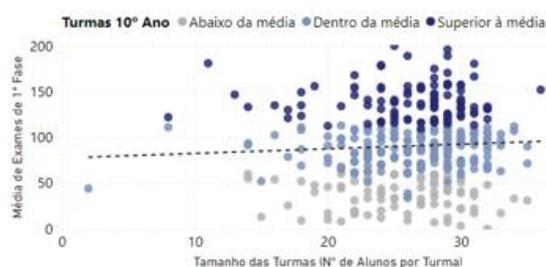
## Tamanho da turma

Analisando o primeiro indicador e calculando os coeficientes de correlação para todas as medidas de sucesso (Tabela 27), o destaque vai para a percentagem de alunos reprovados, com uma correlação positiva de 0.4441, sugerindo que quanto maior o número de alunos numa turma, maior será, em média, a percentagem de alunos reprovados. Por outro lado, em conformidade com a análise feita ao indicador de nível de detalhe escolar “média de alunos por turma” na secção 4.2, as médias obtidas nos três tipos de avaliação (1ª fase, 2ª fase e final) apresentam também correlações significativas, embora mais fracas e negativas, de -0.118, -0.1207 e -0.1001, respetivamente, indicando que quanto maior for o tamanho das turmas, pior serão as médias obtidas nessas medidas de sucesso.

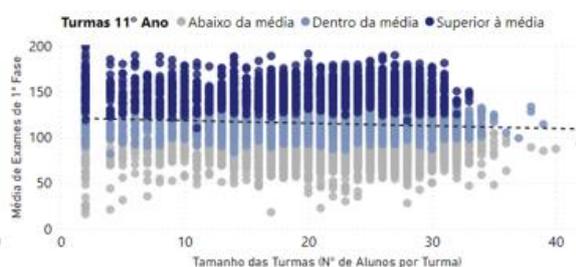
**Tabela 27.** Correlação do Tamanho da turma com as medidas de sucesso em análise ao nível de detalhe por curso.

Variável	Correlação
Média de Exames de 1ª fase	-0.1180
Média de Exames de 2ª fase	-0.1207
Média Final	-0.1001
Percentagem de alunos reprovados	0.4441
Percentagem de alunos em exames de 2ª fase	0.0310
Nº de alunos desistentes	0.0298

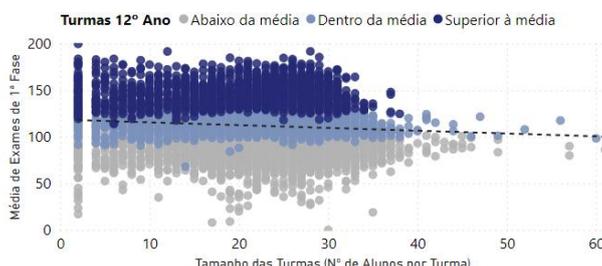
Quanto aos *clusters* formados, através dos gráficos de dispersão detalhados nas Figuras 61, 62 e 63, que demonstram a relação entre cada medida de sucesso destacada e o indicador “Tamanho da turma”, é possível retirar algumas informações úteis. As turmas de 10º ano, sendo as representadas em menor número nesta análise, são também as que contrariam o sentido da correlação demonstrada entre o tamanho da turma e a média de exames de 1ª fase, fornecendo uma linha de tendência crescente. No que toca aos *clusters*, nas turmas do mesmo ano de ensino, tanto as “superiores à média” como as restantes apresentam um número de alunos idêntico, onde a maioria tem 20 a 30 alunos por turma. Pelo contrário, as turmas tanto de 11º ano como de 12º ano, que apresentam linhas de tendência decrescentes e de acordo com a correlação sugerida, tendem a ter números de alunos por turma mais dispersos.



**Figura 60.** Tamanho das turmas pelas Médias de Exames de 1ª Fase das turmas de 10º ano, ao longo dos 6 anos letivos.



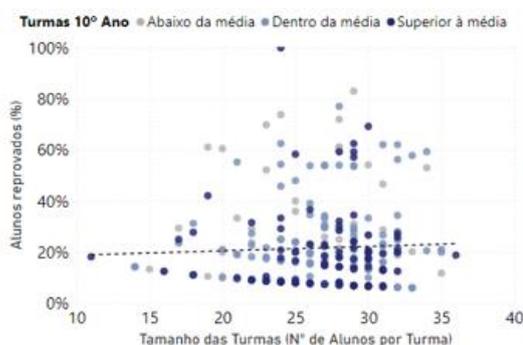
**Figura 61.** Tamanho das turmas pelas Médias de Exames de 1ª Fase das turmas de 11º ano, ao longo dos 6 anos letivos.



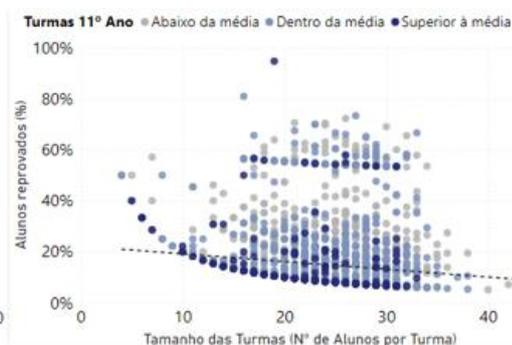
**Figura 62.** Tamanho das turmas pelas Médias de Exames de 1ª Fase das turmas de 12º ano, ao longo dos 6 anos letivos.

Fazendo desta vez referência à correlação mais significativa, entre o tamanho das turmas e a percentagem de alunos reprovados, as Figuras 64, 65 e 66 traduzem a relação entre o indicador e a medida de sucesso em questão. Através das mesmas, é possível indicar que a situação se repete, onde as turmas que estão associadas ao 10º ano de ensino contrariam o sentido do coeficiente de correlação discutido, com uma linha de tendência crescente.

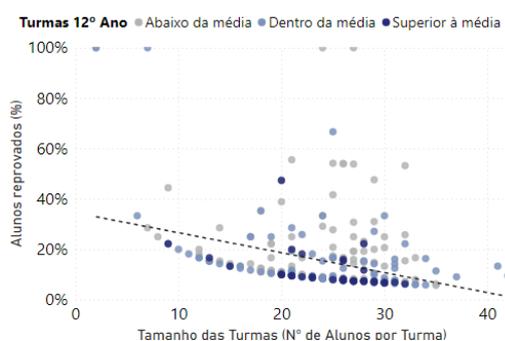
Já os *clusters* das turmas formados, indicam que quando o ano de ensino é 11º ou 12º, as “superiores à média” apresentam uma menor percentagem de alunos reprovados e que tende a anular-se à medida que o tamanho das mesmas diminui.



**Figura 63.** Tamanho da turmas pela Percentagem de alunos reprovados das turmas de 10º ano, ao longo dos 6 anos letivos.



**Figura 64.** Tamanho da turmas pela Percentagem de alunos reprovados das turmas de 11º ano, ao longo dos 6 anos letivos.



**Figura 65.** Tamanho da turmas pela Percentagem de alunos reprovados das turmas de 12º ano, ao longo dos 6 anos letivos.

### Número de matrículas anuladas e Número de alunos transferidos

Relativamente aos últimos dois indicadores para análise, tal como sucedido no nível de detalhe por curso, o número de matrículas anuladas e o número de alunos transferidos não fornecem correlações relevantes quando relacionadas com as medidas de sucesso por cada turma (Tabela 28 e 29).

Variável	Correlação
Média de Exames de 1ª fase	-0.0341
Média de Exames de 2ª fase	-0.012
Média Final	-0.0298
Percentagem de alunos reprovados	0.0013
Percentagem de alunos em exames de 2ª fase	-0.0056
Nº de alunos desistentes	-0.0243

**Tabela 28.** Correlação do Número de matrículas anuladas com as medidas de sucesso em análise ao nível de detalhe por curso.

Variável	Correlação
Média de Exames de 1ª fase	-0.0334
Média de Exames de 2ª fase	-0.0197
Média Final	-0.0348
Percentagem de alunos reprovados	-0.0002
Percentagem de alunos em exames de 2ª fase	0.0299
Nº de alunos desistentes	-0.0136

**Tabela 29.** Correlação do Número alunos transferidos com as medidas de sucesso em análise ao nível de detalhe por curso.

## 4.5 Resumo dos resultados

A secção que se segue serve para retratar todas as análises e resultados obtidos nos capítulos anteriores, respondendo também às restantes perguntas de investigação.

A seleção dos dados para o cálculo dos indicadores mostrou-se uma mais-valia no sentido de facilitar todo o processo de pedido e aquisição dos mesmos, no entanto, nem todos os indicadores e medidas de sucesso foram possíveis de calcular. Os dados disponibilizados não cobriam todos os 29 indicadores e 21 medidas de sucesso sugeridos na revisão de literatura da presente investigação. Considerando tal indisponibilidade nas bases de dados da DGEEC e tendo em conta a inconsistência de outros dados nas tabelas disponibilizadas, foram possíveis de calcular 17 indicadores e 18 medidas de sucesso, de acordo com os três níveis de detalhe estudados (escola, curso e turma).

Quando analisadas todas as medidas de sucesso, é possível obter uma visão valiosa do desempenho das escolas e dos seus estudantes de diversas formas, ao longo dos anos letivos. Tanto as medidas de sucesso de aproveitamento como as de participação demonstraram variações e tendências relevantes, especialmente no ano onde o impacto do COVID-19 se sentiu em todo o país (2019/2020), uma vez que foram implementadas novas leis nesse ano, que deram aos alunos a opção de realizar exames apenas às disciplinas que necessitavam para iniciar os seus percursos no ensino superior. As medidas que traduziam as médias finais e de exames atingiram o pico máximo dos anos em investigação, tendo sido um ano de aparente sucesso tanto ao nível da escola como ao nível do curso, onde as médias aumentaram cerca de 16% em primeira fase, 7.65% em segunda fase e 5.5% nos resultados finais. A percentagem de alunos reprovados bem como a de alunos que repetiam exames desceu nesse mesmo ano, com uma diminuição de 3.22% e de 28.91%, respetivamente. Quanto ao número de alunos desistentes, o número de escolas com mais do que três alunos desistentes diminuiu de 48 para apenas 16 escolas. O curso que mais apresentou sucesso nos exames, nas médias finais e na percentagem de reprovações foi o de Ciências e Tecnologias, tendo sido apenas pior que Línguas e Humanidades na percentagem de alunos em segunda fase. O curso que mais apresentou alunos desistentes foi o de Línguas e Humanidades.

Ao relacionar todos os indicadores calculados com todas as medidas de sucesso, surgem algumas correlações e pontos de influência relevantes para a análise e deteção de estratégias de sucesso na educação secundária. Ao nível de detalhe escolar, a medida de sucesso média de exames de primeira fase apresentou correlações significativas com os indicadores tipo de escola, média de alunos por turma, fatores socioeconómicos dos estudantes, nível médio de qualificação e média de anos de experiência dos professores. Quanto às médias de exames de segunda fase, os coeficientes de correlação em destaque foram os mesmos que nos de primeira fase tendo sido apresentado também um valor relativamente moderado com o tamanho da equipa docente. As médias finais correlacionaram-se com os mesmos indicadores de sucesso que as médias de primeira fase não tendo apenas relação significativa com o indicador nível de qualificação médio dos professores. Quanto à percentagem de alunos reprovados, os indicadores relativos ao tamanho da escola, tamanho da equipa docente, número de cursos disponibilizados, fatores socioeconómicos dos estudantes e o nível médio de qualificação dos professores apresentaram correlações mais significativas tendo como comparação os restantes indicadores.

A medida de sucesso que avalia a percentagem de alunos em segunda fase demonstrou apenas correlação significativa com o indicador tipo de escola. Por fim quanto à medida de sucesso de participação, número de alunos desistentes, os indicadores que demonstraram peso de influência no sucesso desta medida foram o tamanho da escola, o número de alunos transferidos e o número de matrículas anuladas.

Na análise e relação ao nível de detalhe por curso e por turma, o único indicador em destaque é o tamanho dos mesmos, isto é, tanto o tamanho dos cursos como o tamanho das turmas mostraram correlações significativas: o primeiro relativamente à percentagem de alunos reprovados e o segundo quanto à média de exames de primeira fase.

Quando é feita uma análise aos *clusters* criados para cada tipo de nível de detalhe, é possível descrever e mostrar tendências relativas a cada tipo de agrupamento. Escolas “superiores à média” apesar de não apresentarem um número de alunos e de docentes fixo e constante, são maioritariamente de natureza privada, disponibilizando um maior número de cursos (3 e 4 cursos). Os seus alunos terminam as suas disciplinas através das suas avaliações internas e externas (menor número de matrículas anuladas), tendem a solicitar menos pedidos de transferências e apresentam escalões ASE mais baixos (melhores fatores socioeconómicos). Pelo contrário, escolas “abaixo da média”, que têm piores resultados nas medidas de sucesso de ambas as fases dos exames nacionais, têm também percentagens mais altas de alunos reprovados e médias finais mais baixas. Este tipo de escolas, também não demonstra tendências relativamente ao seu tamanho, tanto ao nível de alunos como de professores, no entanto, ao longo dos anos, são estas as escolas que educam alunos com maiores necessidades económicas, uma vez que atingem os maiores registos médios deste indicador.

Já no nível de detalhe por curso, os que se inserem no *cluster* “superior à média” demonstram, em média, um menor número de alunos, vindo a obter melhores resultados médios em primeira fase, melhores resultados finais e uma percentagem mais reduzida de alunos reprovados.

As turmas, por sua vez, quando “superiores à média” no que toca às médias de primeira fase e médias finais, tendem a obter também uma menor percentagem de alunos reprovados.

## CAPÍTULO

### 5. Conclusão

A presente pesquisa explorou a relação entre uma série de indicadores de sucesso e medidas de desempenho académico sugeridos na revisão de literatura, num contexto educacional do ensino secundário em Portugal. Para o cálculo desses mesmos indicadores foi efetuado um pedido formal de colaboração à DGEEC, em parceria com o projeto do Iscte IA>AP. Depois de ultrapassados todos os desafios e burocracias referentes ao pedido e aquisição dos dados, foi disponibilizado um diverso conjunto de informações úteis. Com base nessas informações, é possível fornecer uma resposta a cada uma das perguntas de investigação pré-estabelecidas:

**PI1:** De acordo com os indicadores sugeridos na revisão de literatura, e com base nos dados educativos disponibilizados, que indicadores de sucesso escolar é possível calcular?

Apesar de nem todas as medidas e indicadores selecionados terem sido calculados, devido à falta de dados ou inconsistência dos mesmos, as tabelas fornecidas permitiram atingir os principais objetivos deste projeto de investigação. Foram estudados 17 indicadores e 18 medidas de sucesso de acordo com os três níveis de detalhe (escola, curso e turma), detalhados individualmente no Apêndice C deste documento. Foram calculados indicadores relativos ao tipo de escola, o seu tamanho tanto ao nível de alunos como de docentes, o rácio entre esses mesmos tamanhos, o número médio de alunos por turma, o número de cursos disponibilizados, os fatores socioeconómicos dos alunos e o número de matrículas anuladas e de alunos transferidos. Quanto aos professores, calculou-se os seus níveis médios de qualificação e a média de anos de experiência dos mesmos. Foram calculados também para o nível de detalhe por curso e por turma o tamanho de cada um, no sentido de distinguir o número de alunos presentes tanto em cada curso como em cada turma.

As medidas de sucesso foram todas calculadas para os três níveis de detalhe: médias dos exames de ambas as fases e médias finais, percentagem de alunos que repetiram exames, percentagem de alunos que reprovaram e ainda o número de alunos desistentes.

**PI2:** Quais as características e tendências das medidas de sucesso calculadas ao longo dos anos letivos?

As análises efetuadas a cada medida de sucesso revelaram tendências e detalhes interessantes para a deteção de estratégias de sucesso na educação do ensino secundário. Tanto as medidas de sucesso de aproveitamento como as de participação elevaram a importância e influência das leis implementadas no ensino, no desempenho académico dos estudantes. O ano letivo de 2019/2020 revelou uma melhoria significativa nas medidas de sucesso, uma vez que foi dada a possibilidade aos alunos de realizarem exames apenas às disciplinas necessárias para entrar no ensino superior. Os melhores registos de sucesso das medidas em análise foram verificados no ano letivo em questão, onde as médias aumentaram cerca de 16% em primeira fase, 7.65% em segunda fase e 5.5% nos resultados finais. O número de alunos reprovados e o número de alunos presentes em exames de segunda fase diminuíram em 3.22% e 28.91%, respetivamente. O mesmo se fez sentir no número de alunos desistentes, pois o número de escolas com mais do que três alunos desistentes decresceu de 48 para 16 escolas.

Na análise realizada ao nível de detalhe por curso, Ciências e Tecnologias apresentou o maior sucesso nos exames, nas médias finais e na percentagem de reprovações, tendo sido apenas pior que Línguas e Humanidades na percentagem de alunos que repetiram exames. O curso onde os alunos mais desistiam da sua educação verificou-se no curso de Línguas e Humanidades.

**PI3:** Que indicadores educacionais estabelecem relação com as medidas de sucesso e em que medida influenciam o desempenho escolar no ensino secundário?

Os resultados das análises efetuadas revelam também *insights* sobre quais indicadores exercem maior influência nas medidas de sucesso. Os indicadores tipo de escola, média de alunos por turma, fatores socioeconómicos dos alunos, nível médio de qualificação e a média de anos de experiência dos professores mostraram correlações significativas com a média de exames de primeira fase no nível de detalhe escolar. Estes indicadores fizeram-se também notar com valores de correlações significativos com a medida de sucesso de média de exames de segunda fase, acrescentando ainda o indicador tamanho médio da equipa docente. As médias finais correlacionaram-se com os mesmos indicadores de sucesso que as médias de primeira fase não tendo apenas relação significativa com o indicador nível de qualificação médio dos professores. Os indicadores relativos ao tamanho da escola, tamanho da equipa docente, número de cursos disponibilizados, fatores socioeconómicos dos estudantes e o nível médio de qualificação dos professores apresentaram correlações mais significativas tendo como comparação os restantes indicadores, com a medida de sucesso percentagem de alunos reprovados. Já a percentagem de alunos em segunda fase demonstrou apenas correlação significativa com o indicador tipo de escola. Por fim, os indicadores que demonstraram peso de influência no sucesso da medida número de alunos desistentes foram o tamanho da escola, o número de alunos transferidos e o número de matrículas anuladas.

Na análise e relação ao nível de detalhe por curso e por turma, o único indicador em destaque é o tamanho dos mesmos, isto é, o tamanho dos cursos mostrou peso de influência na percentagem de alunos reprovados e o tamanho das turmas apresentou correlações significativas com a média de exames de primeira fase.

Foram ainda criados *clusters* de acordo com cada nível de detalhe, no sentido de distinguir escolas, cursos e turmas “superiores à média” das restantes. Estas análises permitiram fazer uma descrição geral de cada escola com base nos indicadores estudados: escolas “superiores à média”, contrariamente às escolas “abaixo da média”, são maioritariamente privadas e oferecem mais cursos, apesar de não terem um número constante de alunos e professores. Os seus alunos anulam menos matrículas, realizam menos pedidos de transferência e apresentam melhores fatores socioeconómicos. Os cursos que mais se destacaram foram os “superiores à média”, ao apresentarem um menor número de alunos, vindo a obter melhores resultados médios em primeira fase, melhores resultados finais e uma percentagem mais reduzida de alunos reprovados. Por fim, as turmas que são “superiores à média” nas médias de primeira fase e finais tendem a ter também uma taxa menor de reprovação.

Compreender o significado e o nível de influência de cada indicador nas variações de cada medida de sucesso, disponibilizou orientações sobre como essa compreensão pode ser aplicada para detetar estratégias de sucesso e melhorar o desempenho das escolas e dos seus alunos.

## **5.1 Limitações de pesquisa e trabalho futuro**

Para garantir uma avaliação precisa dos resultados, é importante destacar as principais limitações do presente trabalho de pesquisa. Uma das limitações fundamentais esteve relacionada com a complexidade de aquisição dos dados, onde o processo do pedido e extração dos mesmos, embora essencial para a pesquisa, tornou-se bastante moroso e burocrático devido às políticas de privacidade exigidas. A necessidade de anonimização e confidencialidade dos dados, bem como a inexperiência do investigador em lidar com dados reais, foram dois motivos tomados como limitação à profundidade da pesquisa.

Olhando para o futuro, surgem oportunidades que podem ser consideradas com base nas lições aprendidas e nas limitações enfrentadas. Uma abordagem promissora a ter em conta em trabalhos futuros poderá ser a diversificação das fontes de dados, isto é, explorar diferentes fontes pode ajudar a superar as limitações dos dados inconsistentes e oferecer uma visão mais abrangente do sucesso escolar no ensino secundário. Através de novas fontes, estudos futuros poderão calcular todos os indicadores e medidas não analisadas neste projeto de investigação, podendo também considerar ainda a análise de novos indicadores de sucesso escolar, permitindo uma análise mais holística e precisa.

Em suma, apesar dos desafios encontrados durante a obtenção e análise de dados neste estudo, as limitações identificadas fornecem uma base sólida para que ocorra uma melhoria em pesquisas futuras, que podem contribuir para avanços notáveis no campo da educação e do sucesso escolar em Portugal.

## Referências Bibliográficas

- Alam, T. M., Mushtaq, M., Shaukat, K., Hameed, I. A., Sarwar, M. U., & Luo, S. (2021). A novel method for performance measurement of public educational institutions using machine learning models. *Applied Sciences (Switzerland)*, *11*(19). <https://doi.org/10.3390/app11199296>
- Altun, A., & Kalkan, Ö. K. (2021). Cross-national study on students and school factors affecting science literacy. *Educational Studies*, *47*(4), 403–421. <https://doi.org/10.1080/03055698.2019.1702511>
- Aruna, P., & Priya, N. (2021). Analysis of Machine Learning techniques for Predicting Student Success in an Educational Institution. *ACM International Conference Proceeding Series*, 168–173. <https://doi.org/10.1145/3494885.3494916>
- Börkan, B., & Bakış, O. (2016). Determinants of academic achievement of middle schoolers in Turkey. *Kuram ve Uygulamada Eğitim Bilimleri*, *16*(6), 2193–2217. <https://doi.org/10.12738/estp.2016.6.0227>
- Câmara Leme, A., Escardíbul, J. O., Nunes, L. C., Balcão Reis, A., & Seabra, C. (2020). The effect of a specialized versus a general upper-secondary school curriculum on students' performance and inequality. A difference-in-differences cross-country analysis. *Applied Economics*, *52*(39), 4317–4331. <https://doi.org/10.1080/00036846.2020.1734183>
- Chowa, G. A. N., Masa, R. D., Ramos, Y., & Ansong, D. (2015). How do student and school characteristics influence youth academic achievement in Ghana? A hierarchical linear modeling of Ghana YouthSave baseline data. *International Journal of Educational Development*, *45*, 129–140. <https://doi.org/10.1016/j.ijedudev.2015.09.009>
- Creemers, B. P. M., & Kyriakides, L. (2010). *Validity of Educational Indicators*. 10.1016/B978-0-08-044894-7.00274-8
- Cristina, S., & Caldeira, R. (2018). *Políticas Educativas, Organizações e Sucesso Escolar Estudo de caso numa Escola Secundária/3º ciclo*.
- Gonzaga, H., & Miguel, P. (2019). *Provision of Academic Data for Research: A Step for Academic Success*. <https://doi.org/10.3233/EFI-190349>
- Govindasamy, K., & Velmurugan, T. (2019). Preprocessing and feature extraction process in predicting students performance using clustering technique. *International Journal of Recent Technology and Engineering*, *8*(1), 2407–2413.
- Granvik Saminathen, M., Brolin Låftman, S., Almquist, Y. B., & Modin, B. (2018). Effective schools, school segregation, and the link with school achievement. *School Effectiveness and School Improvement*, *29*(3), 464–484. <https://doi.org/10.1080/09243453.2018.1470988>
- Huberts, L. C. E., Schoonhoven, M., & Does, R. J. M. M. (2022). Multilevel process monitoring: A case study to predict student success or failure. *Journal of Quality Technology*, *54*(2), 127–143. <https://doi.org/10.1080/00224065.2020.1828008>
- Jorge Barbosa Costa, C. (2013). *Sucesso escolar no ensino secundário*. <https://estudogeral.sib.uc.pt/bitstream/10316/24218/3/Sucesso%20escolar%20no%20ensino%20secundario.pdf>

- Karaman, P., & Atar, B. (2019). The effects of student and school level characteristics on academic achievement of middle school students in Turkey. *Journal of Measurement and Evaluation in Education and Psychology*, 10(4), 391–405. <https://doi.org/10.21031/EPOD.564819>
- Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, 26(1), 205–240. <https://doi.org/10.1007/s10639-020-10230-3>
- Langi, M., & Jeon, M. (2022). Identifying and Supporting Academically Low-Performing Schools in a Developing Country: An Application of a Specialized Multilevel IRT Model to PISA-D Assessment Data. *Psychometrika*. <https://doi.org/10.1007/s11336-022-09893-3>
- Larson, K. E., Pas, E. T., Bottiani, J. H., Kush, J. M., & Bradshaw, C. P. (2021). A Multidimensional and Multilevel Examination of Student Engagement and Secondary School Teachers' Use of Classroom Management Practices. *Journal of Positive Behavior Interventions*, 23(3), 149–162. <https://doi.org/10.1177/1098300720929352>
- Leiber, T. (2019). A general theory of learning and teaching and a related comprehensive set of performance indicators for higher education institutions. *Quality in Higher Education*, 25(1), 76–97. <https://doi.org/10.1080/13538322.2019.1594030>
- Lemos, V. (2013). Políticas públicas de educação: Equidade e sucesso escolar. *Sociologia, Problemas e Praticas*, 73, 151–159. <https://doi.org/10.7458/SPP2013732812>
- Liu, H., van Damme, J., Gielen, S., & van den Noortgate, W. (2015). School processes mediate school compositional effects: Model specification and estimation. *British Educational Research Journal*, 41(3), 423–447. <https://doi.org/10.1002/berj.3147>
- Ma, X., Wu, X., Yuan, J., & Luo, X. (2018). Students, Classrooms, Teachers, and Schools: Competing Effects on Science Achievement. *Teachers College Record*, 120, 48.
- Maheshwari, E., Roy, C., Pandey, M., & Rautray, S. S. (2020). Prediction of Factors Associated with the Dropout Rates of Primary to High School Students in India Using Data Mining Tools. *Advances in Intelligent Systems and Computing*, 1013, 242–251. [https://doi.org/10.1007/978-981-32-9186-7\\_26](https://doi.org/10.1007/978-981-32-9186-7_26)
- Marks, G. N. (2010). What aspects of schooling are important? School effects on tertiary entrance performance. *School Effectiveness and School Improvement*, 21(3), 267–287. <https://doi.org/10.1080/09243451003694364>
- Markus, S. (2022). *Education at a Glance 2022 OECD Indicators PUBE*. [https://www.oecd.org/education/education-at-a-glance/EAG2022\\_X3-A.pdf](https://www.oecd.org/education/education-at-a-glance/EAG2022_X3-A.pdf)
- Martínez Abad, F., & Chaparro Caso López, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, 28(1), 39–55. <https://doi.org/10.1080/09243453.2016.1235591>
- Martínez-Abad, F. (2019). Identification of Factors Associated With School Effectiveness With Data Mining Techniques: Testing a New Approach. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02583>
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M. J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66. <https://doi.org/10.1016/j.stueduc.2020.100875>

- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. In *The BMJ* (Vol. 372). BMJ Publishing Group. <https://doi.org/10.1136/bmj.n160>
- Pangeni, K. P. (2014). Factors determining educational quality: Student mathematics achievement in Nepal. *International Journal of Educational Development*, 34(1), 30–41. <https://doi.org/10.1016/j.ijedudev.2013.03.001>
- Perrenoud, P. (2003). *Sucesso Na Escola: Só O Currículo, Nada Mais Que O Currículo! Student Evaluation-Standardized Tests* (Vol. 119).
- Pete Chapman, Clinton, J., Randy Kerber, Thomas Khabaza, & Thomas Reinartz. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.
- Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39(10), 9468–9476. <https://doi.org/10.1016/j.eswa.2012.02.112>
- Shera, P. (2014). School effects, gender and socioeconomic differences in reading performance: A multilevel analysis. *International Education Studies*, 7(11), 28–39. <https://doi.org/10.5539/ies.v7n11p28>
- Silva, L. M. (2020). *Sucesso Escolar Dos Alunos No Ensino Secundário Dos Cursos Científico-Humanísticos*. <http://hdl.handle.net/10400.14/33740>
- Sun, L., Bradley, K. D., & Akers, K. (2012). A Multilevel Modelling Approach to Investigating Factors Impacting Science Achievement for Secondary School Students: PISA Hong Kong Sample. *International Journal of Science Education*, 34(14), 2107–2125. <https://doi.org/10.1080/09500693.2012.708063>
- Takashiro, N. (2017). A multilevel analysis of Japanese middle school student and school socioeconomic status influence on mathematics achievement. *Educational Assessment, Evaluation and Accountability*, 29(3), 247–267. <https://doi.org/10.1007/s11092-016-9255-8>
- Wei, Y., Clifton, R. A., & Roberts, L. W. (2012). School Resources and the Academic Achievement of Canadian Students. In *Alberta Journal of Educational Research* (Vol. 57, Issue 4).

## Apêndice A – Estudos revistos na Revisão de Literatura

<b>ID</b>	<b>Referência</b>	<b>Título</b>	<b>Tipo</b>
1	(Creemers & Kyriakides, 2010)	Validity of Educational Indicators	Journal Article: Research and Evaluation of Educational Effectiveness
2	(Marks, 2010)	What aspects of schooling are important? School effects on tertiary entrance performance	Journal Article: School Effectiveness and School Improvement
3	(Sun et al., 2012)	A Multilevel Modelling Approach to Investigating Factors Impacting Science Achievement for Secondary School Students: PISA Hong Kong Sample	Journal Article: International Journal of Science Education
4	(Şen et al., 2012)	Predicting and analyzing secondary education placement-test scores: A data mining approach	Journal Article: Expert Systems with Applications
5	(Wei et al., 2012)	School Resources and the Academic Achievement of Canadian Students	Report: Alberta Journal of Educational Research
6	(Lemos, 2013)	Políticas públicas de educação: Equidade e sucesso escolar	Journal Article: Association for Computing Machinery
7	(Pangeni, 2014)	Factors determining educational quality: Student mathematics achievement in Nepal	Journal Article: International Journal of Educational Development
8	(Shera, 2014)	School effects, gender and socioeconomic differences in reading performance: A multilevel analysis	Journal Article: International Education Studies
9	(Chowa et al., 2015)	How do student and school characteristics influence youth academic achievement in Ghana? A hierarchical linear modeling of Ghana YouthSave baseline data	Journal Article: International Journal of Educational Development
10	(Liu et al., 2015)	School processes mediate school compositional effects: Model specification and estimation	Journal Article: British Educational Research Journal
11	(Börkan & Bakış, 2016)	Determinants of academic achievement of middle schoolers in Turkey	Journal Article: Kuram ve Uygulamada Eğitim Bilimleri
12	(Takashiro, 2017)	A multilevel analysis of Japanese middle school student and school socioeconomic status influence on mathematics achievement	Journal Article: Educational Assessment, Evaluation and Accountability
13	(Granvik Saminathen et al., 2018)	Effective schools, school segregation, and the link with school achievement	Journal Article: School Effectiveness and School Improvement
14	(Cristina & Caldeira, 2018)	Políticas Educativas, Organizações e Sucesso Escolar Estudo de caso numa Escola Secundária/3º ciclo	Report

15	(Ma et al., 2018)	Students, Classrooms, Teachers, and Schools: Competing Effects on Science Achievement	Journal Article: Teachers College Record
16	(Martínez-Abad, 2019)	Identification of Factors Associated With School Effectiveness With Data Mining Techniques: Testing a New Approach	Journal Article: Frontiers in Psychology
17	(Gonzaga & Miguel, 2019)	Provision of Academic Data for Research: A Step for Academic Success	Report
18	(Karaman & Atar, 2019)	The effects of student and school level characteristics on academic achievement of middle school students in Turkey	Conference Proceedings: Journal of Measurement and Evaluation in Education and Psychology
19	(Martínez-Abad et al., 2020)	Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment	Journal Article: Studies in Educational Evaluation
20	(Maheshwari et al., 2020)	Prediction of Factors Associated with the Dropout Rates of Primary to High School Students in India Using Data Mining Tools	Conference Proceedings: Advances in Intelligent Systems and Computing
21	(Silva, 2020)	Sucesso Escolar dos Alunos no Ensino Secundário dos Cursos Científico-Humanísticos	Report
22	(Câmara Leme et al., 2020)	The effect of a specialized versus a general upper-secondary school curriculum on students' performance and inequality. A difference-in-differences cross-country analysis	Journal Article: Applied Economics
23	(Larson et al., 2021)	A Multidimensional and Multilevel Examination of Student Engagement and Secondary School Teachers' Use of Classroom Management Practices	Journal Article: Journal of Positive Behavior Interventions
24	(Alam et al., 2021)	A novel method for performance measurement of public educational institutions using machine learning models	Journal Article: Applied sciences
25	(Altun & Kalkan, 2021)	Cross-national study on students and school factors affecting science literacy	Journal Article: Educational Studies
26	(Khan & Ghosh, 2021)	Student performance analysis and prediction in classroom learning: A review of educational data mining studies	Journal Article: Education and Information Technologies
27	(Langi & Jeon, 2022)	Identifying and Supporting Academically Low-Performing Schools in a Developing Country: An Application of a Specialized Multilevel IRT Model to PISA-D Assessment Data	Journal Article: Psychometrika
28	(Huberts et al., 2022)	Multilevel process monitoring: A case study to predict student success or failure	Journal Article: Journal of Quality Technology

## Apêndice B – Tabelas de dados disponibilizadas

**Tabela Escolas:**

<b>Campo</b>	<b>Significado</b>
ANO_LETIVO	Ano letivo a que os dados se referem.
CECOLA	Código identificador de cada escola.
ESCOLA	Nome da escola.
CNUTSI	Código de Nomenclatura de Unidades Territoriais para fins estatísticos, de nível 1, constituída por 3 unidades, a que pertence a escola.
NUTSI	Nomenclatura de Unidades Territoriais para fins estatísticos, de nível 1, constituída por 3 unidades, a que pertence a escola.
CCONCELHO	Código de concelho da escola.
CONCELHO	Concelho da escola.
CNATUREZA	Código da natureza da escola (0 ou 1).
NATUREZA	Natureza da escola (Publica ou Privada).
CNATUREZA2	Código de Natureza 2 da escola.
NATUREZA2	Natureza 2 da escola.
CTIPOLOGIA	Código do tipo de entidade.
TIPOLOGIA	Tipologia da escola.
CNUTSII	Código de Nomenclatura de Unidades Territoriais para fins estatísticos, de nível 2, constituída por 7 unidades, a que pertence a escola.
NUTSII	Nomenclatura de Unidades Territoriais para fins estatísticos, de nível 2, constituída por 7 unidades, a que pertence a escola.
CNUTSIII	Código de Nomenclatura de Unidades Territoriais para fins estatísticos, de nível 3, constituída por 25 unidades, a que pertence a escola.
NUTSIII	Nomenclatura de Unidades Territoriais para fins estatísticos, de nível 3, constituída por 25 unidades, a que pertence a escola.
CAGRUPAMENTO	Código de Agrupamento da escola.
AGRUPAMENTO	Agrupamento da escola.

**Tabela Alunos:**

<b>Campos</b>	<b>Significado</b>
FKENTIDADE_ALUNO	Código identificador de cada aluno.
ANOLETIVO	Ano letivo a que os dados se referem.
ESTABELECIMENTO	Tipo de estabelecimento da descola.
ANO_ENSINO	Ano de ensino.
CESCOLA	Código identificador de cada escola.
CURSO	Curso do aluno.
TURMA	Turma do aluno.
SITUACAO	Situação do aluno.
ACONTECIMENTO	Descritivo do acontecimento de cada aluno (Desistiu/Excluído/Retido por faltas; Transitou/Concluiu; Não Transitou/não concluiu; Transferido).
CGENERO	Código de género do aluno.
CIDADE	Cidade do aluno.
ESCALAO_ASE	Escalão de Apoio Social Escolar (ASE) do aluno, que inclui os escalões "Não Beneficia", "Escalão A", "Escalão B" e "Escalão C".
MÃE_HABILITAÇÃO	Nível de habilitação da mãe do aluno.
PAI_HABILITAÇÃO	Nível de habilitação do pai do aluno.
EE_HABILITAÇÃO	Nível de habilitação do Encarregado de Educação do aluno.
MÃE_NACIONALIDADE	Nacionalidade da mãe do aluno.
PAI_NACIONALIDADE	Nacionalidade do pai do aluno.
MÃE_PROFISSÃO	Profissão da mãe do aluno.
PAI_PROFISSÃO	Profissão do pai do aluno.
MÃE_SIT_OEMPREGO	Situação de emprego da mãe do aluno.
PAI_SIT_EMPREGO	Situação de emprego do pai do aluno.
EE_PARENTESCO	Grau de parentesco do Encarregado de educação do aluno.
EE_NACIONALIDADE	Nacionalidade do Encarregado de educação do aluno.
EE_PROFISSÃO	Profissão do Encarregado de educação do aluno.
EE_SIT_EMPREGO	Situação de emprego do Encarregado de educação do aluno.

**Tabela Professores:**

<b>Campos</b>	<b>Significado</b>
FKENTIDADEprofessor	Código identificador de cada professor.
ANOLETIVO	Ano letivo a que os dados se referem.
CESCOLA	Código da escola onde o professor leciona.
GENERO	Gênero do professor.
HABILITACAO_ACADEMICA	Nível de habilitação acadêmica do professor.
CICLO_DOCENCIA	Ciclo de docência à qual o professor presta ensino.
CATEGORIA	Categoria do professor.
IDADE	Idade do professor.
HR_COMP_LETIVA	Número de horas semanais letivas do professor.
HR_D4	Nº de horas semanais do docente afetas às atividades totais do professor.
HR_COMP_N_LETIVA	Número de horas da componente não letiva do professor.
DIRETOR_ESCOLA	Indicador se o professor é ou não diretor da escola.
APOSENTACAO	Número de anos de serviço que contam para a aposentação do professor.
ANTES_PROFISSIONALIZACAO	Nº de anos de serviço efetuado antes da profissionalização do professor.
APOS_PROFISSIONALIZACAO	Nº de anos de serviço efetuado depois da profissionalização.
COD_NACIONALIDADE	Código da nacionalidade do professor.
NACIONALIDADE	Nacionalidade do professor.
COMPONENTE_FORMACAO	Componente de formação do professor.
GRUPO_DOCENCIA	Grupo de docência do professor.
ATIVIDADE	Descrição do tipo de atividade do professor (Educação Pré-Escolar; 1º Ciclo do Ensino Básico; 2º e 3º Ciclos Ensino Básico; Ensino Secundário).

**Tabela Exames:**

<b>Campos</b>	<b>Significado</b>
ANO	Ano letivo a que os dados se referem.
ESCOLA_DESCR	Nome da escola.
ESCOLA_CODDGEEC	Código identificador de cada escola.
FKENTIDADE	Código identificador de cada aluno.
ESCOLA_PUBPRIV	Natureza da escola (PUB/PRIV).
FASE	Identificador da fase de cada exame.
EXAME	Código identificador de cada tipo de exame (código da disciplina).
EXAME_DESCR	Descritivo de cada tipo de exame (nome das disciplinas).
INTERNO	Código identificador do aluno à disciplina do exame (Interno/Externo) .
PARAMELHORIA	Código identificador se o exame tinha a finalidade de melhoria de nota.
PARAINGRESSO	Código identificador se o exame serviria de ingresso ao ensino superior.
USOUTOLERANCIA	Código identificador de que o aluno utilizou tolerância em cada exame.
CURSO_DESCR	Curso do aluno.
SITFREQ_DESCR	Descritivo da situação de exame do aluno (Admitido; Reprovou Frequência).
CLASS_EXAM	Classificação de exame do aluno.
CFD	Classificação final à disciplina do aluno.

## Apêndice C – Medidas e Indicadores Calculados

Medida de Sucesso						
Média de Exames de 1ª Fase (Escolar)						
Nº de Campos: 7			Nº de observações: 59 387			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_EXAME	float	21 306	0	200	110.74	32.29
N_ALUNOS	int	338	1	504	21.91	33.51

Medida de Sucesso						
Média de Exames de 1ª Fase (Curso)						
Nº de Campos: 9			Nº de observações: 83 600			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_EXAME	float	22 377	0	200	111.08	34.21
N_ALUNOS	int	250	1	415	15.66	33.51

Medida de Sucesso						
Média de Exames de 1ª Fase (Turma)						
Nº de Campos: 10			Nº de observações: 136 557			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	varchar	8 408				
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_EXAME	float	20 070	0	200	110.82	33.83
N_ALUNOS	int	43	1	58	9.66	8.36

Medida de Sucesso						
Média de Exames de 2ª Fase (Escolar)						
Nº de Campos: 6			Nº de observações: 44 006			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_EXAME_2F	float	9 671	0	200	100.18	31.85

Medida de Sucesso						
Média de Exames de 2ª Fase (Curso)						
Nº de Campos: 8			Nº de observações: 92 809			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_EXAME_2F	float	9 374	-1	200	100.54	32.79

Medida de Sucesso						
Média de Exames de 2ª Fase (Turma)						
Nº de Campos: 9			Nº de observações: 44 006			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	611	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	int	8 055	-	-	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_EXAME_2F	float	7 263	-1	200	100.43	32.80

Medida de Sucesso						
Médias Finais (Escolar)						
Nº de Campos: 7			Nº de observações: 40 527			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_FINAL	float	10 618	0.0	200.0	118.2	2.97
N_ALUNOS	int	346	1	504	24.07	34.81

Medida de Sucesso						
Médias Finais (Curso)						
Nº de Campos: 9			Nº de observações: 40 527			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_FINAL	float	10 257	0.0	200.0	117.4	32.9
N_ALUNOS	int	257	1	416	15.66	23.06

Medida de Sucesso						
Médias Finais (Turma)						
Nº de Campos: 10			Nº de observações: 137 597			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	int	8 409	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
MEDIA_FINAL	float	5 774	0.0	200.0	117.54	32.58
N_ALUNOS	int	46	1	58	9.57	8.52

Medida de Sucesso						
Percentagem de Alunos em 2ª Fase (Escolar)						
Nº de Campos: 7				Nº de observações: 40 511		
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	611	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
N_ALUNOS_2F	int	146	1	298	9.91	14.59
PERCENTAGEM	float	2 893	0.0	1.00	0.361	0.183

Medida de Sucesso						
Percentagem de Alunos presentes em 2ª Fase (Curso)						
Nº de Campos: 9				Nº de observações: 57 215		
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	611	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
CEXAME	int	17	122	835	-	-
EXAME	varchar	17	-	-	-	-
N_ALUNOS_2F	int	131	1	298	7.32	10.80
PERCENTAGEM	float	4 574	0.0	1.0	0.29	0.22

Medida de Sucesso						
Percentagem de Alunos em 2ª Fase (Turma)						
Nº de Campos: 8				Nº de observações: 26 716		
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	604	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	int	7 320	-	-	-	-
N_ALUNOS_2F	int	41	1	53	11.28	7.20
PERCENTAGEM	float	93	0.0	1.0	0.48	0.27

Medida de Sucesso						
Percentagem de Alunos reprovados (Escolar)						
Nº de Campos: 6			Nº de observações: 6 552			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	632	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
N_REPROVADOS	int	104	0	162	17.68	15.37
N_TOTAL	int	464	7	635	148.03	94.91
PERCENTAGEM	float	2 861	0	1	0.131	0.078

Medida de Sucesso						
Percentagem de Alunos reprovados (Curso)						
Nº de Campos: 8			Nº de observações: 18 185			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	632	101615	1824324	-	-
CCURSO	int	17	1	17	-	-
CURSO	varchar	17	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
N_REPROVADOS	int	60	0	88	5.75	6.13
N_TOTAL	int	279	1	538	49.58	39.30
PERCENTAGEM	float	1 703	0	1	0.134	0.106

Medida de Sucesso						
Percentagem de Alunos reprovados (Turma)						
Nº de Campos: 9			Nº de observações: 24 663			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	622	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	int	5 860	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
N_REPROVADOS	int	21	2	27	3.12	2.13
N_TOTAL	varchar	48	0	50	21.83	7.98
PERCENTAGEM	float	215	0.00	1.00	0.21	0.25

Medida de Sucesso						
Nº de Alunos desistentes (Escolar)						
Nº de Campos: 4			Nº de observações: 14 276			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	647	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CDESISTENCIA	varchar	3	0	>3	-	-

Medida de Sucesso						
Nº de Alunos desistentes (Curso)						
Nº de Campos: 6			Nº de observações: 38 283			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	640	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CDESISTENCIA	varchar	3	0	>3	-	-

Medida de Sucesso						
Nº de Alunos desistentes (Turma)						
Nº de Campos: 7			Nº de observações: 3 890			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	433	101615	1824324	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	int	1 296	-	-	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CDESISTENCIA	varchar	3	0	>3	-	-

Indicador de Sucesso						
Tipo de Escola (Escolar)						
Nº de Campos: 2			Nº de observações: 673			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
CESCOLA	int	673	101615	1824324	-	-
NATUREZA	varchar	2	-	-	-	-

Indicador de Sucesso						
Tamanho da Escola (Escolar)						
Nº de Campos: 5			Nº de observações: 17 739			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	673	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
NATUREZA	varchar	2	-	-	-	-
N_ALUNOS	int	338	4	504	181.537	90.05

Indicador de Sucesso						
Tamanho da Equipe Docente (Escolar)						
Nº de Campos: 3			Nº de observações: 4 337			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	658	101615	1824324	-	-
N_PROFESSORES	int	185	2	223	67.32	37.09

Indicador de Sucesso						
Rácio tamanho da escola face ao tamanho da equipa docente (Escolar)						
Nº de Campos: 5			Nº de observações: 4 242			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	558	101615	1824324	-	-
N_ALUNOS	int	988	2	1672	362.87	274.04
N_PROFESSORES	int	180	2	200	68.15	36.15
RACIO	float	3 406	0.11	54	5.21	4.24

Indicador de Sucesso						
Média de Alunos por Turma (Escolar)						
Nº de Campos: 5			Nº de observações: 38 283			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CURSO	varchar	4	-	-	-	-
MEDIA_ALUNOS	float	59	1	95	19.16	7.45

Indicador de Sucesso						
Número de cursos disponibilizados (Escolar)						
Nº de Campos: 5			Nº de observações: 38 283			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	668	101615	1824324	-	-
N_CURSOS	varchar	4	0	4	2.88	0.98

Indicador de Sucesso						
Número de matrículas anuladas (Escolar)						
Nº de Campos: 4			Nº de observações: 14 226			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	568	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CMATRÍCULAS_ANULADAS	varchar	3	0	>3	-	-

Indicador de Sucesso						
Número de alunos transferidos (Escolar)						
Nº de Campos: 4			Nº de observações: 14 276			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	576	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CTRANSFERIDOS	varchar	3	0	>3	-	-

Indicador de Sucesso						
Fatores Socioeconômicos dos alunos (Escolar)						
Nº de Campos: 6			Nº de observações: 15 873			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	646	101615	1824324	-	-
ESCALAO_ASE	varchar	4	-	-	-	-
CESCALAO	int	4	1	4	-	-
N_ALUNOS_ASE	int	862	2	1739	111.68	168.11
MEDIA_ASE	float	2765	1.0	4.0	1.71	0.33

Indicador de Sucesso						
Qualificação dos professores (Escolar)						
Nº de Campos: 6			Nº de observações: 12 944			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	492	101615	1824324	-	-
HABILITACAO	varchar	4	-	-	-	-
CHABILITACAO	int	4	1	4	2.84	1.04
N_PROFESORES	int	150	0	172	22.7	28.92
MEDIA_HABILITACAO	float	1295	2.00	4.00	3.10	0.10

Indicador de Sucesso						
Média de anos de experiência dos professores (Escolar)						
Nº de Campos: 4			Nº de observações: 4337			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	492	101615	1824324	-	-
MEDIA_ANOS_EXPERINCIA	float	3730	0	33.03	22.75	3.59
N_PROFESSORES	int	183	0	223	68.2	37.98

Indicador de Sucesso						
Tamanho do Curso (Curso)						
Nº de Campos: 5			Nº de observações: 10 609			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	612	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CURSO	varchar	4	-	-	-	-
N_ALUNOS	int	552	2	1193	110.88	106.87

Indicador de Sucesso						
Número de matrículas anuladas (Curso)						
Nº de Campos: 6			Nº de observações: 29 366			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	8	2015/2016	2020/2021	-	-
CESCOLA	int	640	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
CMATRÍCULAS_ANULADAS	varchar	3	0	>3	-	-

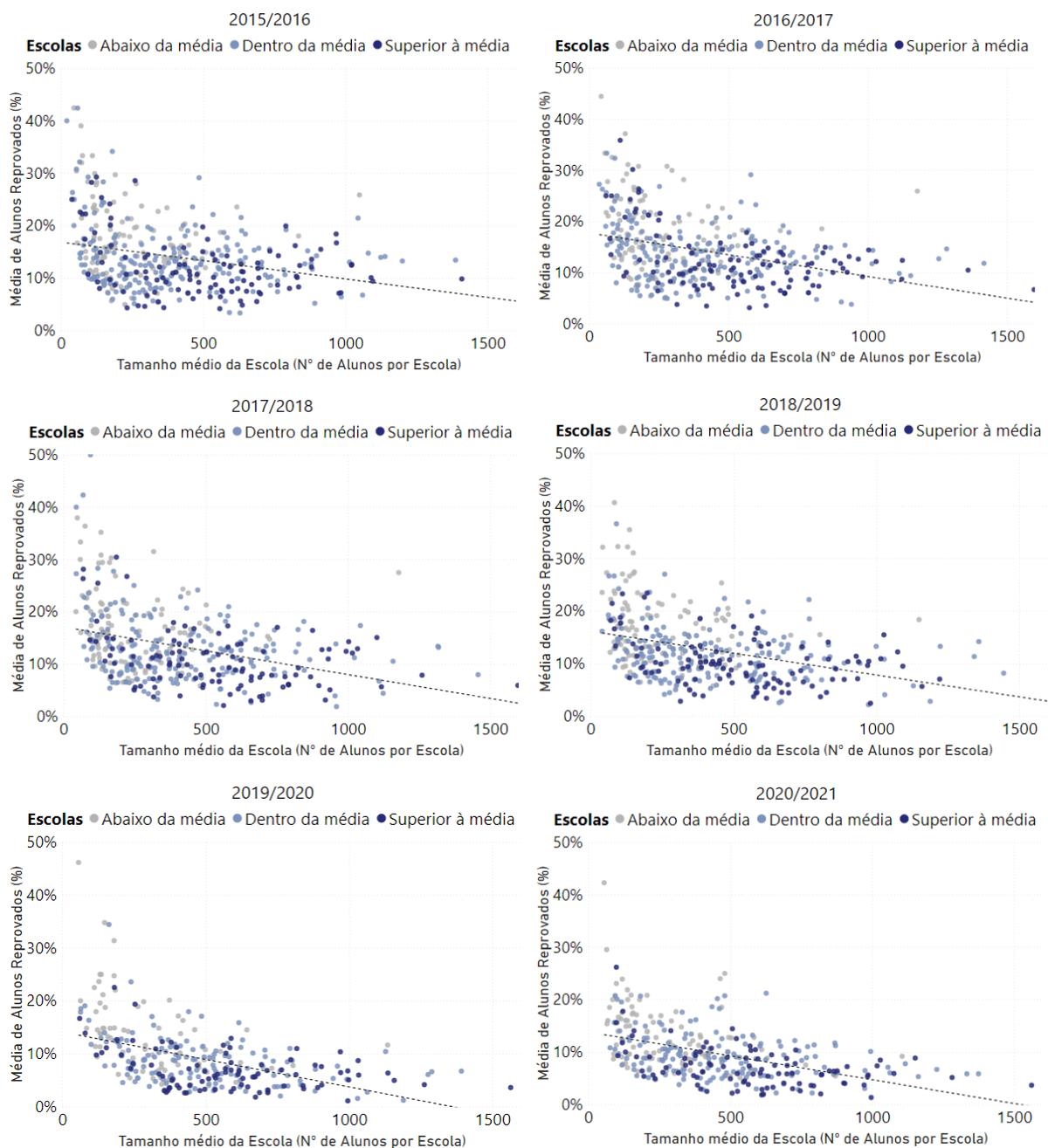
Indicador de Sucesso						
Nº de Alunos transferidos (Curso)						
Nº de Campos: 6			Nº de observações: 29 366			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	8	2015/2016	2020/2021	-	-
CESCOLA	int	640	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
N_DESISTENCIAS	varchar	3	0	>3	-	-

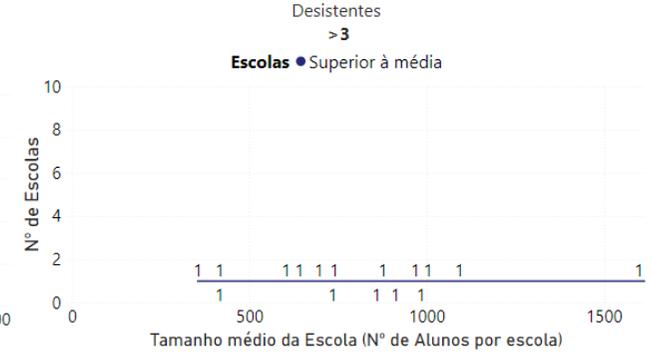
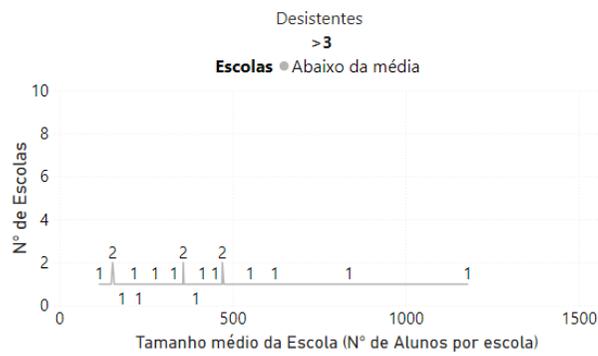
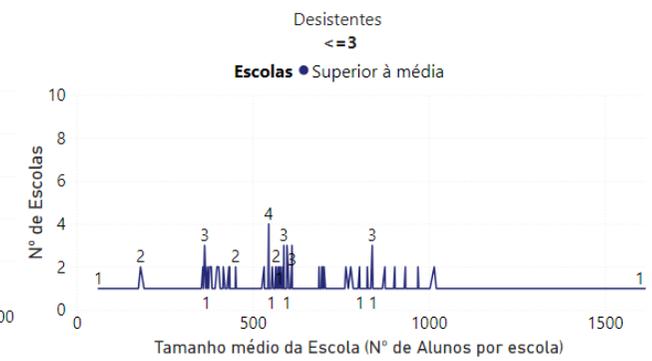
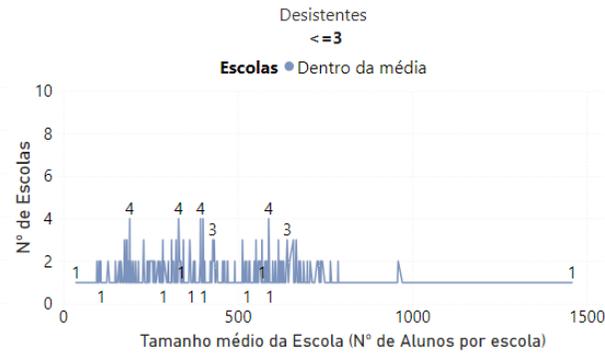
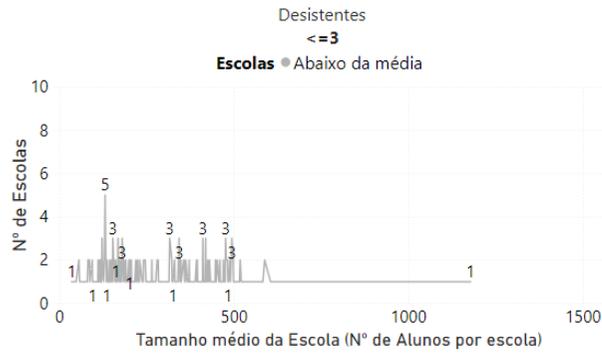
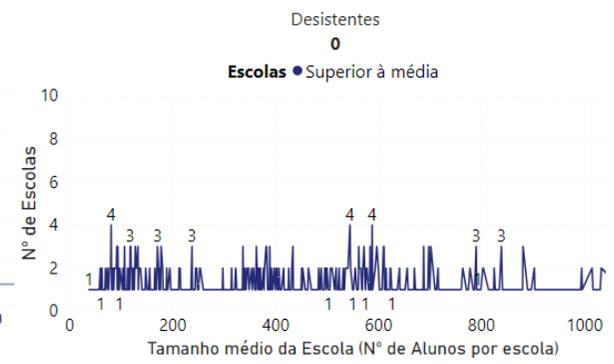
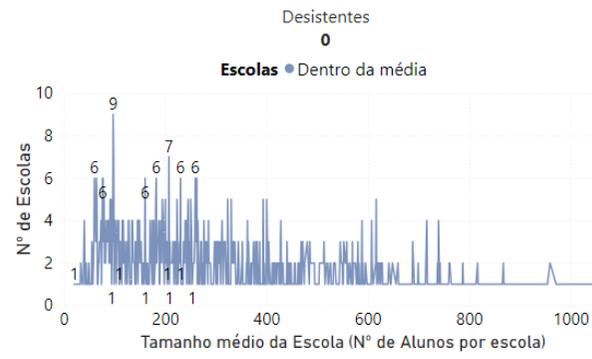
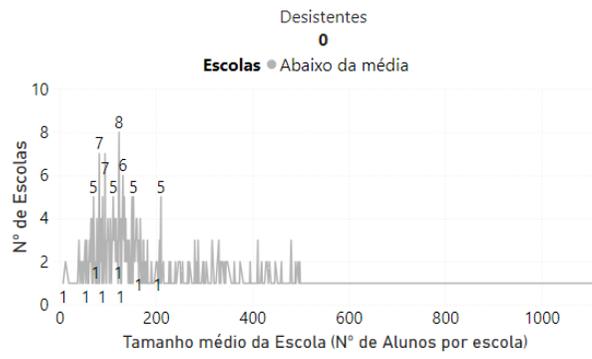
Indicador de Sucesso						
Tamanho das turmas (Turma)						
Nº de Campos: 7			Nº de observações: 51 922			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	653	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	varchar	13 825	-	-	-	-
N_ALUNOS	int	338	2	144	23.44	7.17

Indicador de Sucesso						
Número de matrículas anuladas (Turma)						
Nº de Campos: 7			Nº de observações: 6 429			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	545	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	varchar	1987	-	-	-	-
CMATRÍCULAS_ANULADAS	varchar	3	0	>3	-	-

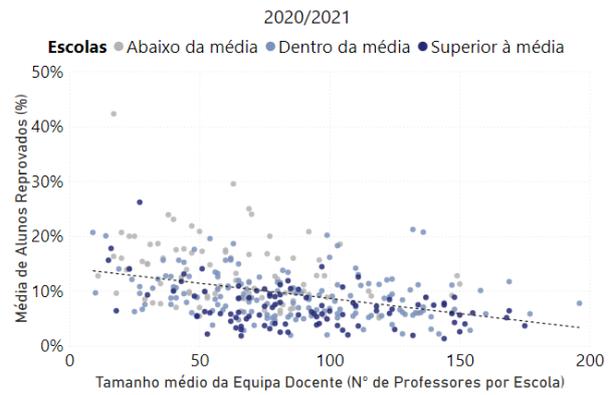
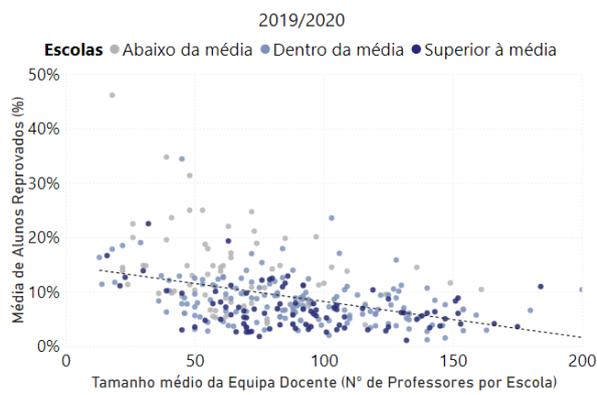
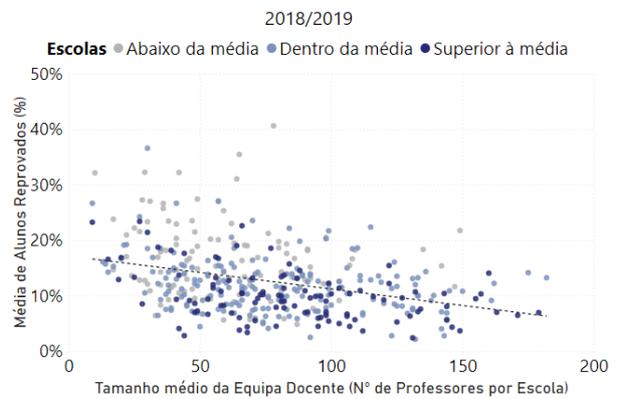
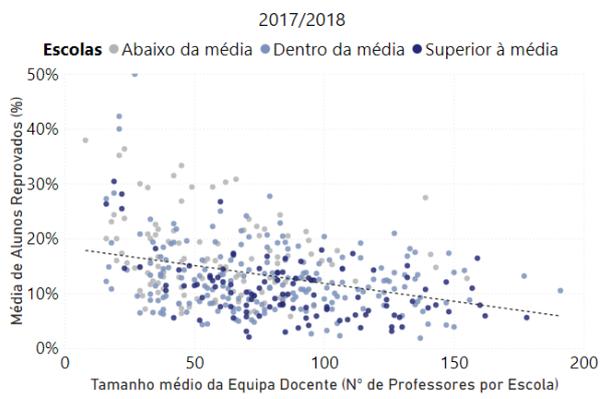
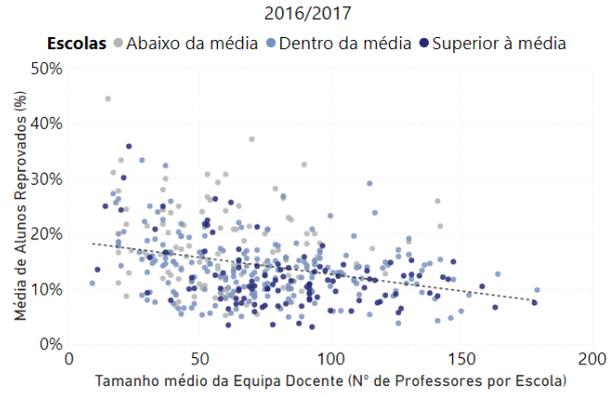
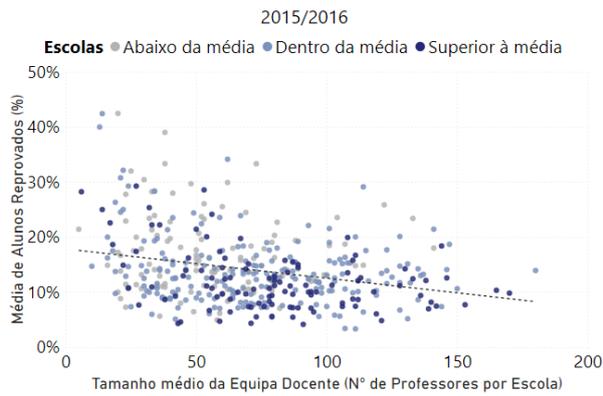
Indicador de Sucesso						
Nº de alunos transferidos (Turma)						
Nº de Campos: 7			Nº de observações: 6 348			
Campo	Tipo	# Valores únicos	Mínimo	Máximo	Média	Desvio padrão
ANOLETIVO	varchar	6	2015/2016	2020/2021	-	-
CESCOLA	int	569	101615	1824324	-	-
ANO_ENSINO	varchar	3	10. ano	12. ano	-	-
CCURSO	int	4	1	4	-	-
CURSO	varchar	4	-	-	-	-
TURMA	varchar	1803	-	-	-	-
CTRANSFERIDOS	varchar	3	0	>3	-	-

## Apêndice D – Gráficos Indicador Tamanho da Escola

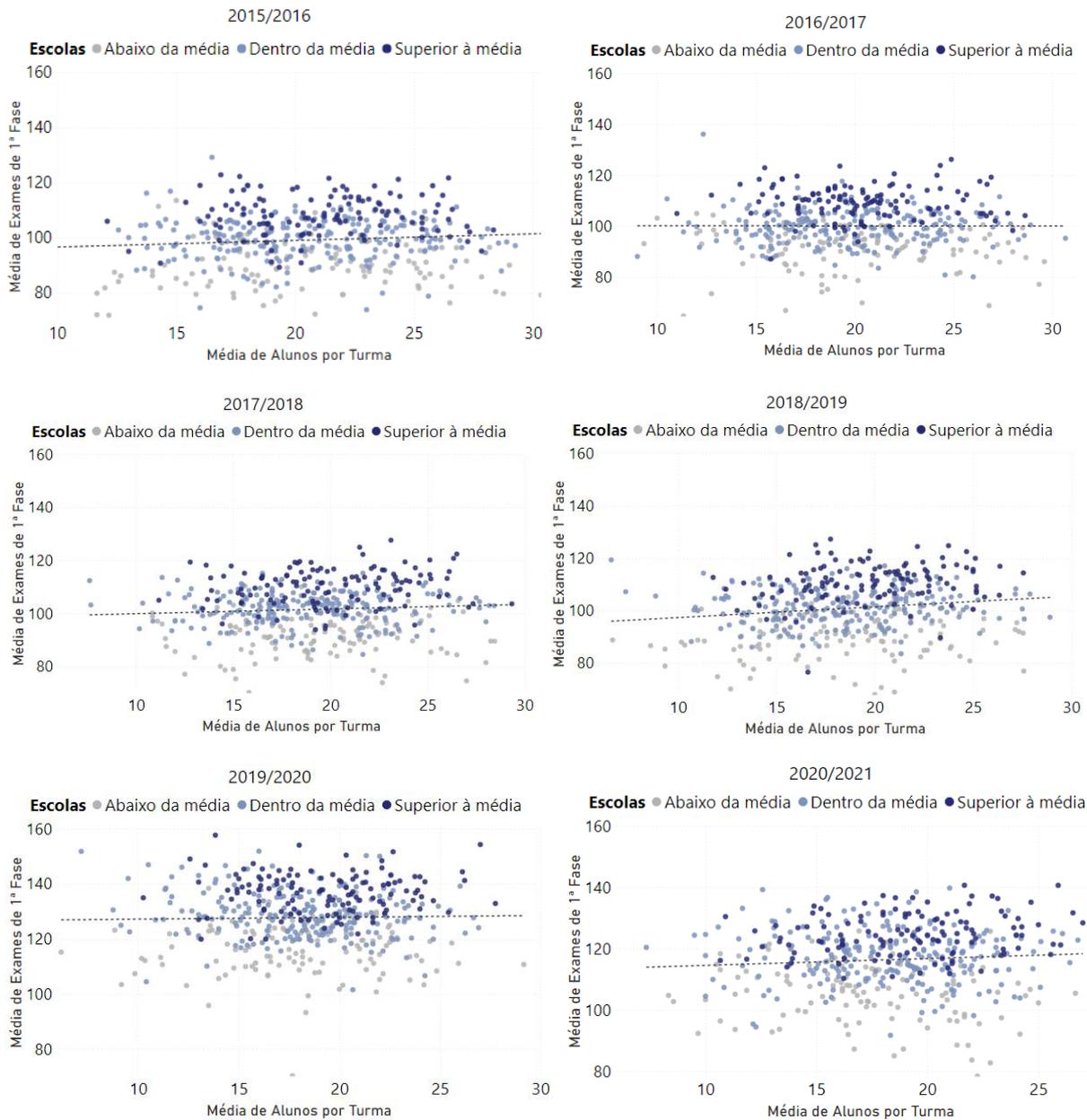




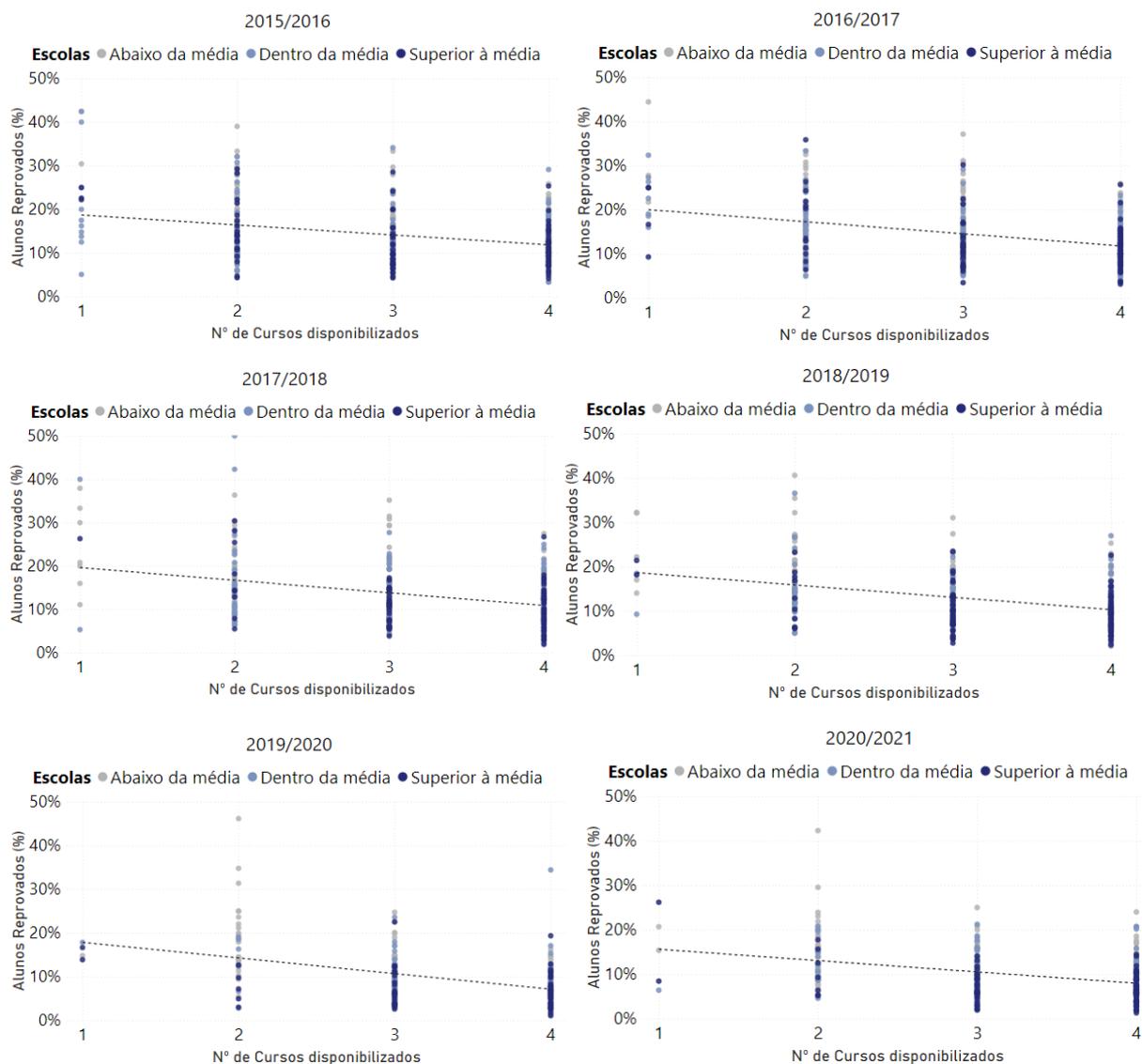
# Apêndice E – Gráficos Indicador Tamanho da Equipe Docente



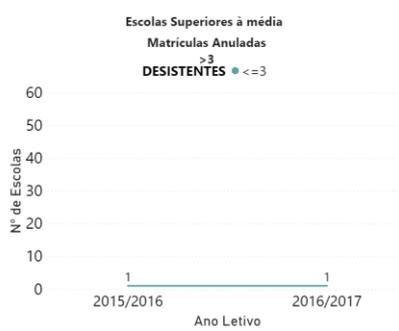
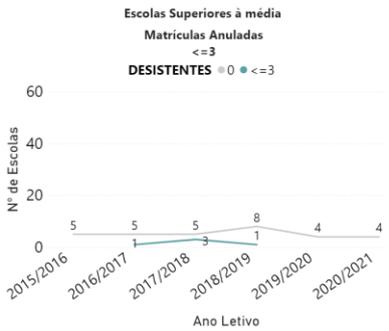
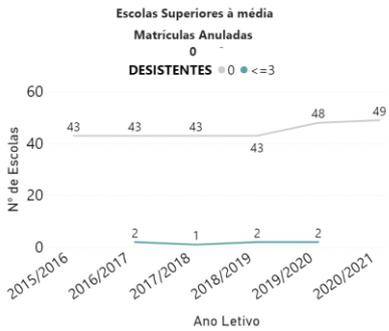
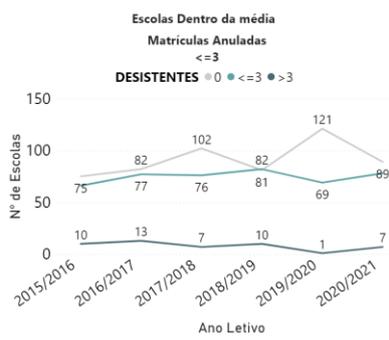
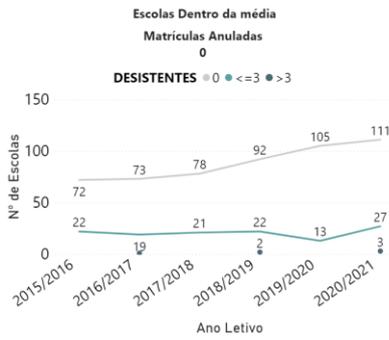
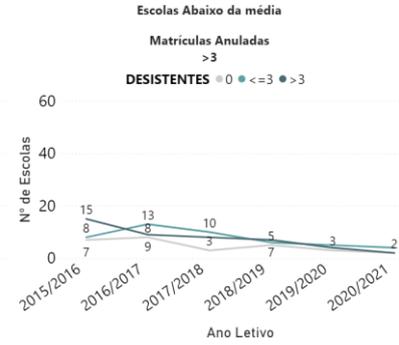
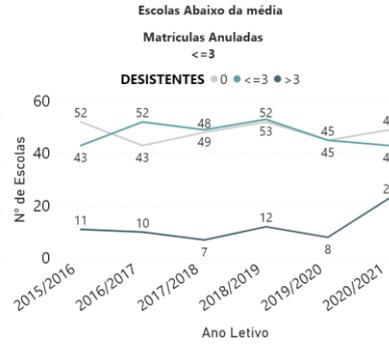
# Apêndice F – Gráficos Indicador Média de Alunos por Turma



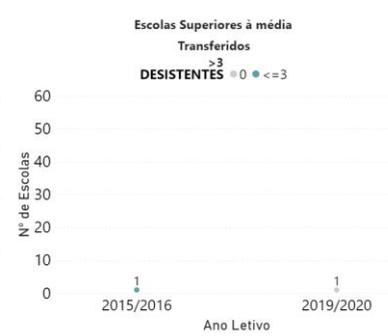
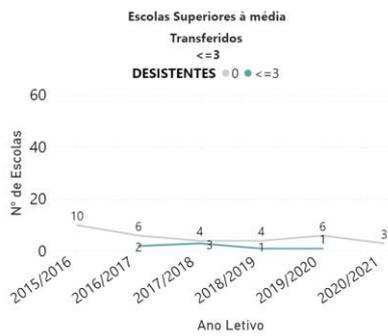
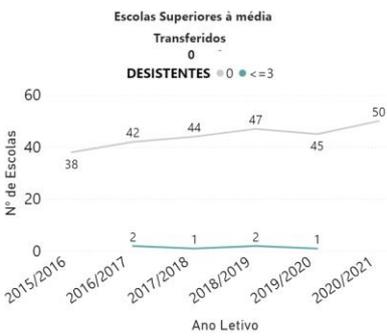
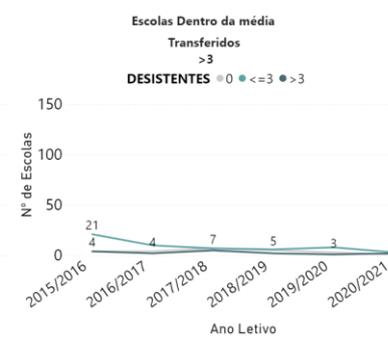
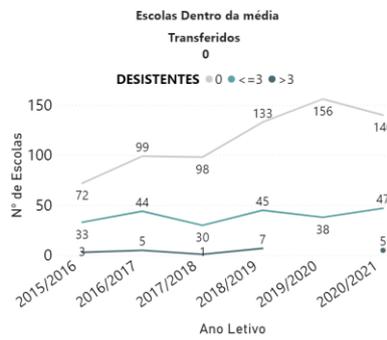
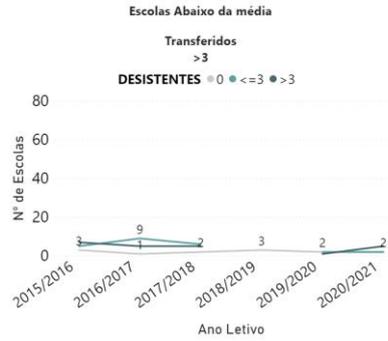
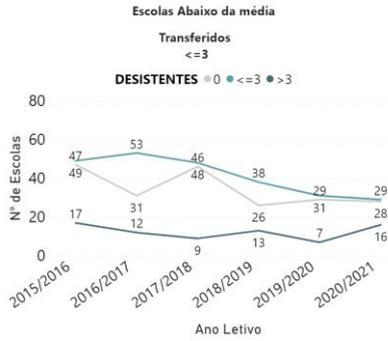
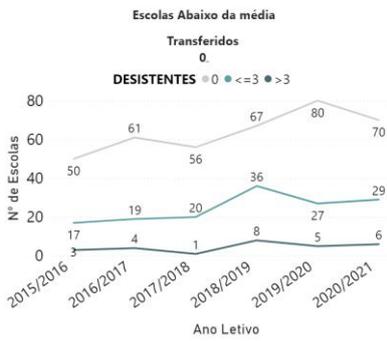
## Apêndice G – Gráficos Indicador Número de cursos disponibilizados



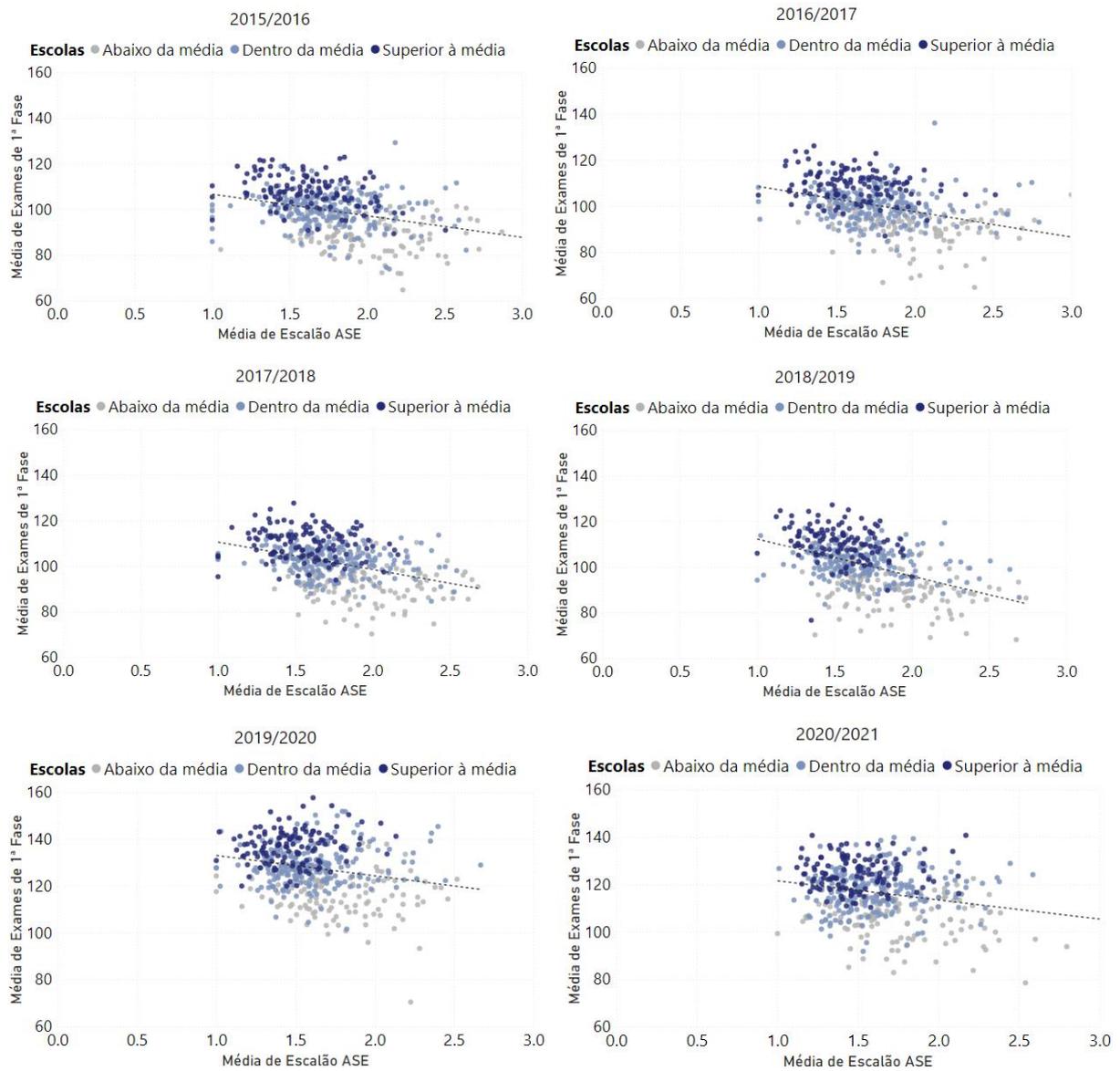
# Apêndice H – Gráficos Indicador Matrículas Anuladas



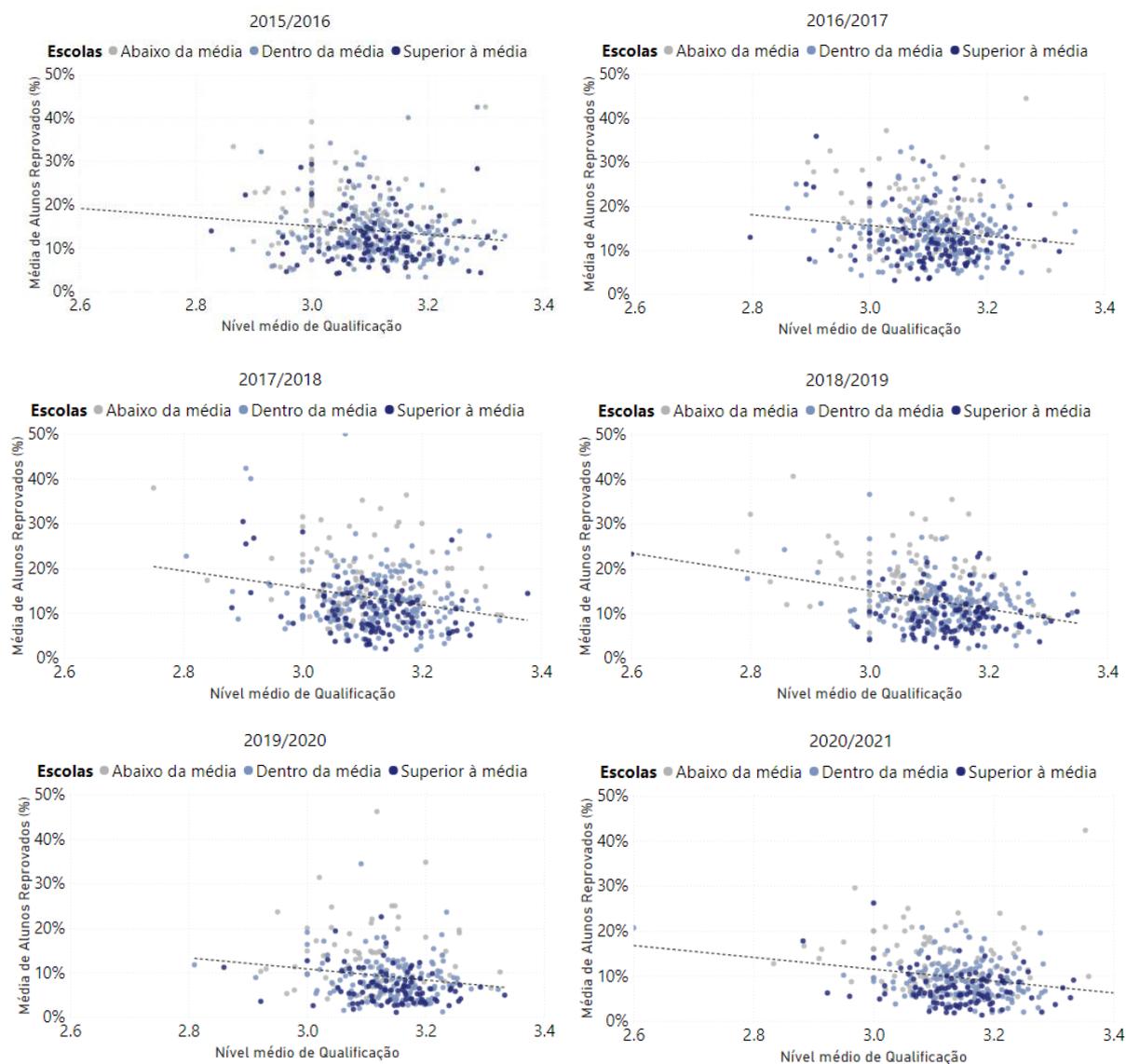
# Apêndice I – Gráficos Indicador Número de Alunos Transferidos



# Apêndice J – Gráficos Indicador Fatores Socioeconômicos dos alunos



## Apêndice K – Gráficos Indicador Nível médio de qualificação dos professores



# Apêndice L – Gráficos Indicador Média de anos de experiências dos professores

