

Hanna Paola Paranhos

Desenvolvimento de um *dashboard* para análise e visualização dos dados educacionais dos discentes do Instituto de Ciências Exatas e Aplicadas da UFOP

Orientador: Helen De Cassia Sousa da Costa Lima

Coorientador: Janniele Aparecida Soares Araújo

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Agosto de 2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

P223d Paranhos, Hanna Paola.

Desenvolvimento de um dashboard para análise e visualização dos dados educacionais dos discentes do Instituto de Ciências Exatas e Aplicadas da UFOP. [manuscrito] / Hanna Paola Paranhos. - 2021.
49 f.: il.: color., gráf., tab., mapa.

Orientadora: Profa. Dra. Helen de Cassia Sousa da Costa Lima.
Coorientadora: Profa. Dra. Janniele Aparecida Soares Araújo.
Monografia (Bacharelado). Universidade Federal de Ouro Preto.
Instituto de Ciências Exatas e Aplicadas. Graduação em Sistemas de Informação .

1. Análise de dados. 2. Mineração de dados (Computação). 3. Dashboards (Sistemas de informação gerencial). 4. Educação. I. Araújo, Janniele Aparecida Soares. II. Lima, Helen de Cassia Sousa da Costa. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 004.62:378

Bibliotecário(a) Responsável: Sione Galvão Rodrigues - CRB6 / 2526



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS
DEPARTAMENTO DE COMPUTAÇÃO E SISTEMAS



FOLHA DE APROVAÇÃO

Hanna Paola Paranhos

Desenvolvimento de um dashboard para análise e visualização dos dados educacionais dos discentes do Instituto de Ciências Exatas e Aplicadas da UFOP

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação

Aprovada em 30 de agosto de 2021

Membros da banca

Mestra - Helen de Cássia Sousa da Costa Lima - Orientadora - Universidade Federal de Ouro Preto
Doutora - Janniele Aparecida Soares Araújo - Coorientadora - Universidade Federal de Ouro Preto
Doutor - Bruno Rabello Monteiro - Universidade Federal de Ouro Preto
Doutor - George Henrique Godim da Fonseca - Universidade Federal de Ouro Preto

Helen de Cássia Sousa da Costa Lima, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 14/09/2021



Documento assinado eletronicamente por **Helen de Cassia Sousa da Costa Lima, PROFESSOR DE MAGISTERIO SUPERIOR**, em 14/09/2021, às 18:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0220933** e o código CRC **0A9320C7**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.009567/2021-71

SEI nº 0220933

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

Este trabalho é dedicado à memória do meu avô Paulo Paranhos, à memória do meu tio Pedro Paulo Paranhos e a todos que me acompanharam e torceram por mim nesse período.

Agradecimentos

Agradeço a cada Hanna do passado que não desistiu e permaneceu fiel a quem ela é.

Agradeço a minha mãe, Vivian Figueiredo Nogueira por toda luta vivida que me proporcionou a possibilidade de correr atrás das nossas conquistas. Mãe, saiba que todo esforço não foi em vão, hoje sou quem sou, estou onde estou, e chegarei onde quero chegar, por sua causa.

Agradeço ao meu pai, Dalton Morais Paranhos por toda confiança depositada mesmo quando fui "cabeça fraca", todas às vezes que eu estava preocupada, observar o seu jeito de viver, aproveitando cada momento com leveza, é o que equilibrava minha rotina e me deixava aproveitar mais o processo do que esperar para ser feliz apenas no resultado.

Agradeço á minha família, às minhas avós Vivina e Reniria pela torcida, preocupação, orações e apoio no processo e ao meu avô David por todo apoio e carinho. Agradeço aos meus tios e tias, em especial minha tia Lú, por todo amor, e pelo zelo por mim tal qual fosse sua filha. Agradeço aos meus primos e primas que torceram por mim e mesmo que de longe sempre ofereceram seu apoio. Especialmente meus primos, Luca Paranhos e Lucila Otoni que muitas vezes abriram mão da sua rotina para viver um pouco da minha.

Agradeço a minha segunda família da república Erva-Doce e agregados, por toda construção e desconstrução que vivemos. Guilherme, Daniella, Jardel, Victor, Sthael, Douglas, Elielson, Vitória e para os próximos que virão, essa conquista é nossa.

Agradeço a todos meus professores, mestres e em especial minhas orientadoras Helen De Cassia Sousa da Costa Lima e Janniele Aparecida Soares Araújo, obrigada por toda paciência, conversas, orientações e desenvolvimento no tempo que estivemos juntas, ver mulheres na área de tecnologia, executando trabalhos com excelência e com tanta gentileza, é um exemplo da profissional que almejo me tornar no futuro.

Agradeço aos meus amigos e ao meu amor, Rafael Fernandes, por sempre acreditarem em mim, pela companhia nas noites de estudo, nos dias difíceis, e também pelas celebrações em dias bons, todo esforço não foi e não será em vão.

Por fim, agradeço a todos que contribuíram de alguma forma para que a realização desse sonho, fosse possível.

“Science is more than a body of knowledge; it is a way of thinking.”

— Carl Sagan (1934 – 1996),
in: The Demon-Haunted World: Science as a Candle in the Dark.

Resumo

Este trabalho extrai conhecimento dos dados educacionais do campus de uma universidade através do uso de visualizações. O motivo pelo qual está sendo realizada a pesquisa se deve à grande quantidade de dados que, se bem trabalhados, podem gerar bons insumos para a comunidade acadêmica. A motivação deste trabalho consiste em utilizar a análise e visualização dos dados para auxiliar gestores nas tomadas de decisão do campus. Utilizando a metodologia *Knowledge Discovery in Databases* (KDD) e o *software* de *Business Intelligence Power BI* para análise dos dados e criação das visualizações, respectivamente. O resultado atingido foi a criação de 7 dashboards interativos disponibilizados, sendo eles o demográfico, de admissão, de evasão, colação, desempenho aluno, desempenho da matéria e desempenho do departamento.

Palavras-chaves: *KDD. Dashboard. Educação. Análise de Dados. Visualização de dados.*

Abstract

This work proposes to extract knowledge from the educational data of a university campus through the use of visualizations. The reason why a survey is being carried out is due to the large amount of data that, if well worked, can generate good inputs for the academic community. The objective of the work is to use the power of data analysis and visualization for auxiliary managers in decision-making on campus. We use the Knowledge Discovery in Databases (KDD) methodology and the Business Intelligence Power BI software for data analysis and creation of views, respectively. The result achieved was 7 interactive dashboards made available, namely the demographic, admission, dropout, collation, student performance, subject performance and department performance.

Key-words: KDD. Dashboard. Education. Data analysis. Data visualization.

Lista de ilustrações

Figura 1 – Etapas do KDD	15
Figura 2 – Descrição da base de dados	18
Figura 3 – Descrição da base de dados após o pré-processamento	20
Figura 4 – Sumarização dos dados	20
Figura 5 – Resultado da transformação	22
Figura 6 – <i>Dashboard</i> demográfico	23
Figura 7 – <i>Dashboard</i> demográfico: gráfico de coluna a nível de estado	24
Figura 8 – <i>Dashboard</i> demográfico: filtro alunos no sexo feminino	25
Figura 9 – <i>Dashboard</i> de admissão	26
Figura 10 – <i>Dashboard</i> de admissão: gráfico de coluna a nível de semestre	27
Figura 11 – <i>Dashboard</i> de evasão	28
Figura 12 – <i>Dashboard</i> de evasão: filtro alunos evadidos no primeiro período	29
Figura 13 – <i>Dashboard</i> de evasão: filtro do curso de Engenharia de Computação	29
Figura 14 – <i>Dashboard</i> de colação	30
Figura 15 – <i>Dashboard</i> de desempenho aluno	31
Figura 16 – <i>Dashboard</i> de desempenho aluno: filtro de matéria foco em nota e falta	32
Figura 17 – <i>Dashboard</i> de desempenho aluno: filtro de matéria foco em situação aprovado	32
Figura 18 – <i>Dashboard</i> de desempenho matéria	33
Figura 19 – <i>Dashboard</i> de desempenho matéria: detalhes da visualização	34
Figura 20 – <i>Dashboard</i> de desempenho departamento	35
Figura 21 – Como inserir novos dados: Passo 5	41
Figura 22 – Como inserir novos dados: Passo 6	41
Figura 23 – Como inserir novos dados: Passo 7	42
Figura 24 – Como inserir novos dados: Passo 8	42
Figura 25 – Como inserir novos dados: Passo 9	43
Figura 26 – Como inserir novos dados: Passo 10	43
Figura 27 – Como inserir novos dados: Passo 11	44
Figura 28 – Como inserir novos dados: Passo 12	44
Figura 29 – Como inserir novos dados: Passo 13	45
Figura 30 – Como inserir novos dados: Passo 14	45
Figura 31 – Como compartilhar o relatório: Passo 1	46
Figura 32 – Como compartilhar o relatório: Passo 2	47
Figura 33 – Como compartilhar o relatório: Passo 3	47
Figura 34 – Como compartilhar o relatório: Passo 4	48
Figura 35 – Como compartilhar o relatório: Passo 5	48

Figura 36 – Como compartilhar o relatório: Passo 6 49

Sumário

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	<i>Dashboard</i>	14
2.2	Modelagem Multidimensional	14
2.3	Descoberta de conhecimento em banco de dados (KDD)	15
2.4	Plano de desenvolvimento Institucional Ufop 2016-2025	15
2.5	Trabalhos Relacionados	16
3	METODOLOGIA	18
3.1	Seleção	18
3.2	Pré-Processamento	19
3.3	Transformação	21
4	RESULTADOS	23
4.1	<i>Dashboard</i> demográfico	23
4.2	<i>Dashboard</i> de admissão	26
4.3	<i>Dashboard</i> de evasão	28
4.4	<i>Dashboard</i> de colação	30
4.5	<i>Dashboard</i> de desempenho aluno	31
4.6	<i>Dashboard</i> de desempenho matéria	33
4.7	<i>Dashboard</i> de desempenho departamento	35
5	CONSIDERAÇÕES FINAIS	36
5.1	Trabalhos Futuros	37
	REFERÊNCIAS	38
	ANEXOS	39
	ANEXO A – COMO INSERIR NOVOS DADOS	40
	ANEXO B – COMO COMPARTILHAR O RELATÓRIO	46

1 Introdução

A tomada de decisões na educação clássica, é geralmente, alcançada por meio de opiniões e discussões em reuniões, combinadas com um certo número de estatísticas e análises de dados (ZHU, 2018). No entanto, hoje em dia, existem tecnologias orientadas a dados que complementam o método tradicional. Os dados quantitativos, em conjunto com as informações qualitativas vivenciadas por professores e gestores, podem abranger uma melhor eficácia nas decisões tomadas.

Dentro das universidades federais há uma série de dados que não estão sendo utilizados para a extração de conhecimento que auxilie na tomada de decisão. Sendo assim, um dos objetivos deste trabalho é apontar as anomalias que apenas o tratamento e a visualização podem ajudar a identificar. Por exemplo, quais alunos, professores e matérias precisam de mais atenção, se a ementa atende nossos objetivos institucionais, quais decisões estão sendo baseadas em fatos ou apenas em opiniões, quais as causas da evasão de alunos da instituição e por fim, quais informações têm maior utilidade na hora de gerir a instituição.

É possível extrair todos esses conhecimentos com os dados que temos através da utilização do processo *Knowledge Discovery in Databases (KDD)*, um processo iterativo, iterativo e não trivial de descoberta de informações, mediante a identificação de padrões válidos, novos, potencialmente úteis e compreensíveis em dados Aljohani e Davis (2013) para gerar insumos importantes que auxiliem nossos alunos, professores e gestores.

O foco da pesquisa será na parte de visualização e distribuição. Com a visualização é possível obter o entendimento rápido e efetivo do que precisamos. Já com a distribuição é possível descentralizar as informações do campus e aumentar o interesse dos alunos na tomada de decisões importantes, melhorando a publicidade dos dados e fomentando o interesse político dos discentes, que poderão participar em votações para centros acadêmicos, diretorias de institutos e reitoria com base em informações confiáveis. O Instituto de Ciências Exatas e Aplicadas (ICEA) é uma instituição pública, onde todos os papéis são importantes e todos deveriam ter acesso aos fatos ocorridos, dado que existem deveres a serem cumpridos e direitos a serem cobrados.

O uso de *dashboard* em instituições de ensino pode ser uma prática aplicada hoje em dia, auxiliando no monitoramento constante a cada semestre dos resultados alcançados na instituição. A partir do momento que temos a visualização dos dados, é possível definir métricas de desempenho do campus, acompanhá-las facilmente e identificar se as decisões tomadas no passado geraram bons resultados.

Nos institutos educacionais há uma enorme quantidade de dados por aluno, mas

faltam ferramentas que se proponham a reunir todas as fontes e extrair conhecimento a partir disso. A análise de dados desempenha um papel importante na compreensão e disponibilidade da informação. Atualmente existe escassez de conhecimento, uma vez que há a entrada de novos alunos de seis em seis meses, sendo difícil acompanhar o perfil dos alunos de maneira constante e atualizada. As tecnologias de comunicação desempenharam um papel importante na abertura de novas janelas de oportunidades com finalidade de usar o dados disponíveis para descobrir problemas educacionais (ALJOHANI; DAVIS, 2013). Um sistema de visualização em formato de *dashboard* iria simplificar o processo para traçar o retrato da comunidade e auxiliar na tomada de decisão da gestão do campus.

O objetivo geral deste trabalho é utilizar o poder da análise de dados e visualização em *dashboard* para melhorar a tomada de decisão, tendo como intenção descentralizar a informação e compreender o perfil dos discentes do ICEA. A seguir estão listados os objetivos específicos para alcançar o geral:

1. Criação de *Dashboards* interativas;
2. Descentralizar a informação dentro do campos;
3. Auxiliar na tomada de decisão dos gestores da instituição;
4. Conscientizar a organização sobre o perfil dos seus alunos;
5. Incentivar a cultura de tomar decisões orientadas a dados.

Este trabalho está dividido em quatro capítulos, sendo o Capítulo 1 a apresentação geral do tema, objetivos propostos e a identificação do problema, ou seja, a motivação da pesquisa. O Capítulo 2 contempla a fundamentação teórica necessária para compreender os termos, ferramentas e tecnologias abordados no trabalho, juntamente com os trabalhos relacionados com o tema proposto. No Capítulo 4 são descritas todas as etapas desenvolvidas, como sumarização da base de dados, modelagem e visualizações geradas. Finalmente, no Capítulo 5 as considerações finais e propostas para trabalhos futuros.

2 Fundamentação Teórica

Nesta seção é apresentado todo o conhecimento necessário para desenvolver este trabalho, como *dashboard*, *KDD*, *Python* e por fim os trabalhos relacionados.

2.1 *Dashboard*

As visualizações dos dados tem como objetivo identificar informações ocultas, tendências, anomalias ou exceções, e prever acontecimentos.

Few (2006) afirma que um *dashboard* é uma exibição visual das informações mais importantes necessárias para alcançar um ou mais objetivos, consolidados e organizados em uma única tela para que as informações possam ser monitoradas de uma só vez. Através do uso de *dashboard*, o resultado da análise será disponibilizado de forma clara e acessível, de maneira que, quando os discentes e docentes utilizarem a ferramenta, será possível identificar o valor agregado da análise.

O Power BI é uma ferramenta de inteligência de negócios, que oferece uma estrutura para carregar, trabalhar e criar visualizações interativas dos dados em um modelo *self-service*. É possível gerar relatórios, gráficos, personalizar com filtros, e obter informações de forma rápida a partir das visualizações.

Dentro da ferramenta podemos utilizar o Power Query para conectar diversas fontes de dados, realizar a transformação das mesmas e incluí-las nas visualizações. A ferramenta foi escolhida pois ao utilizar o e-mail institucional sua utilização pelo *desktop* é gratuita.

2.2 Modelagem Multidimensional

A modelagem multidimensional é uma técnica de modelar bancos de dados de forma que facilite as consultas em diversos contextos. Esses sistemas homogeneizam e integram dados de organizações em um repositório de dados, a fim de explorar esta representação única e detalhada da organização e extrair conhecimento relevante para a tomada de decisão da organização (ROMERO; ABELLÓ, 2009).

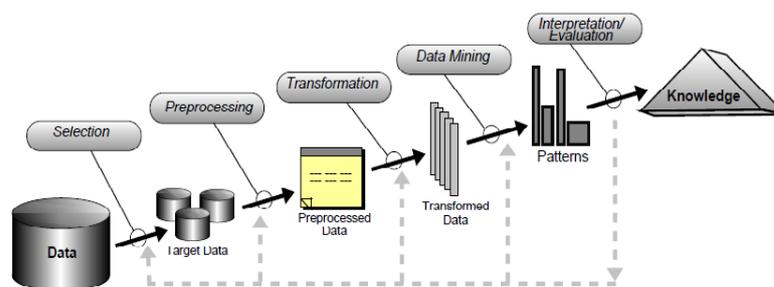
Este tipo de modelagem é obrigatoriamente dividida em tabelas fatos e tabelas dimensões. As tabelas fatos contêm um acontecimento, as chaves de relações das dimensões associadas e métricas de acompanhamento, sendo obrigatoriamente o relacionamento muitos para um com as tabelas dimensões. As tabelas dimensões contêm informações qualitativas com relacionamento um para muitos e permite a visualização em diversas perspectivas, como a intenção de descrever o fato ocorrido. É possível trabalhar com dois

tipos de modelagens multi-dimensionais: o modelo estrela e o modelo floco de neve. A multidimensionalidade fornece uma visualização de dados amigável, fácil de entender e intuitiva para usuários finais não especializados (ROMERO; ABELLÓ, 2009).

2.3 Descoberta de conhecimento em banco de dados (KDD)

A descoberta de conhecimento em banco de dados apresentado na figura 1, é um processo composto por duas metas, sendo a primeira a predição, e a segunda a descrição. A primeira meta tem por finalidade prever o comportamento futuro. Já a descrição tem como foco encontrar padrões com o objetivo de apresentá-los a um usuário de uma forma compreensível para humanos (FAYYAD et al., 1996).

Figura 1 – Etapas do KDD



Fonte : (FAYYAD et al., 1996)

A Figura 1 apresenta todas as etapas do KDD. A etapa de seleção é realizada quando um conjunto de dados é escolhido. Na etapa de pré-processamento é feita a limpeza de dados que não são relevantes ao objetivo. Já na etapa da transformação, os dados são agrupados e carregados para que possa ser feita a análise. A quarta etapa, denominada *Data Mining*, é onde os algoritmos e métodos são escolhidos e aplicados com objetivo de encontrar padrões ou anomalias dentro da base de dados. Na quinta e última etapa, avaliação e interpretação, é comparado o objetivo definido no início do projeto com o resultado de todas as etapas. Caso os resultados atendam os objetivos, podemos partir para a visualização, caso contrário, é possível retornar a qualquer etapa em qualquer momento, e refinar os processos até que o conhecimento seja de fato extraído e seja útil para o usuário de alguma forma.

2.4 Plano de desenvolvimento Institucional Ufop 2016-2025

O Plano de Desenvolvimento Institucional (PDI) tem por objetivo orientar uma gestão participativa, planejada, sustentável e com o compromisso de integrar e aprimorar continuamente as atividades de ensino, pesquisa, extensão e estímulo à inovação da Instituição (UFOP, 2016). O PDI propõe aprimorar doze eixos temáticos a partir de áreas

da Ufop, sendo um deles o "Ensino de Graduação". Para cada eixo existe uma série de objetivos, com metas e indicadores atrelados aos mesmos. Este trabalho se enquadra na proposta do PID de estimular a inovação dentro da universidade, e contribui com dois objetivos citados do Plano de desenvolvimento Institucional, 'Promover a melhoria da qualidade do curso de graduação' e 'Implementar mecanismos de acolhimento, recepção e acompanhamento dos graduandos da UFOP'. No primeiro objetivo, o PDI tem como meta acompanhar indicadores de admissão, como por exemplo, número de ingressantes, notas de vestibular, modalidade de ingresso e desempenho da disciplina. Além de taxas e motivos de evasão, considerando apenas alunos computados no Censo.

Uma das ações atreladas ao segundo objetivo é 'desenvolver ações de acompanhamento dos alunos' e 'intensificar estratégias de alunos com baixo desempenho'. Sendo assim, as visualizações podem contribuir com a identificação e acompanhamento dos pontos de melhoria em diferentes perspectivas, como, tempo, departamento, aluno, matérias e cursos.

2.5 Trabalhos Relacionados

Gonçalves, Silva e Cortes (2018) propõe a utilização do KDD para identificar os problemas de evasão com dados acadêmicos dentro do Instituto Federal do Maranhão (IFMA). Eles propuseram uma abordagem para prever quais alunos são mais propensos a evasão utilizando técnicas de classificação. O autor aplicou os algoritmos de classificação *Naive Bayes*, *J48* e *Support Vector Machine*. Todas as etapas do KDD foram aplicadas e ao analisar os resultados, identificaram que a maior probabilidade de evasão está associado a tempo de curso inferior a 3 semestres ou estudantes com coeficiente de rendimento menor que 5.0 que estão no máximo no 9º semestre. É possível observar que passar pelas etapas do KDD, no contexto educacional, gerou bons insumos para identificar quais alunos precisam de mais atenção, visto que com essas informações gestores podem agir de forma mais eficiente. O trabalho é focado em evasão, mas com a mesma base de dados, é possível gerar conhecimento que auxilie os gestores em outras problemáticas dentro da universidade.

Santos (2017) desenvolveu uma proposta metodológica com intuito de apoiar gestores educacionais partindo de uma análise dos dados armazenados em sistemas acadêmicos. O autor propõe utilizar ferramentas de *Business Intelligence (BI)* para melhorar e tornar o processo de tomada de decisão mais rápido e eficiente, além de reduzir riscos de incertezas, tornando as decisões mais assertivas.

Com intuito de compreender o negócio, identificar as restrições dos problemas e traçar planos para alcançar melhores resultados, o autor mapeou quais cenários necessitavam de maior atenção dos decisores para transformá-los em indicadores dentro de um *dashboard*. Esses cenários mapeados foram referentes às questões de evasão, alunos em grupo de risco e o seu respectivo perfil. O autor passou por todas as etapas do *KDD*, onde

o resultado foi a modelagem e construção de um *data warehouse* que foi utilizado como fonte de dados para as visualizações.

A fim de apurar o impacto que o dashboard gerou para os tomadores de decisão das universidades, foi realizada uma pesquisa onde os resultados foram:

"Para 94,57% dos entrevistados os dados sobre a evasão escolar ficaram mais aparentes (expostos) o que permite que ações de planejamento sejam tomadas visando o melhoramento contínuo. Cerca de 92,75% acreditam que o modelo de aplicação de BI reduziria o índice de evasão escolar"(SANTOS, 2017).

O autor concluí que possuir uma interface amigável e interativa gera a possibilidade de identificar padrões, tendências e anomalias sobre a evasão escolar dentro da instituição. Para Santos (2017), a visualização é, claramente, um dos fatores fundamentais à assimilação e percepção de quem decide.

3 Metodologia

Nesta Capítulo são apresentadas as etapas do *KDD* desenvolvidas neste trabalho. A etapa de seleção é descrita na Seção 3.1, o pré-processamento na Seção 3.2, a transformação na Seção 3.3.

3.1 Seleção

Para a base de dados foram selecionados arquivos com temas distintos e seus respectivos atributos, sendo eles, o arquivo de dados gerais dos alunos, notas, evasão e colação. As informações foram solicitadas pela seção de ensino do campus da universidade. Todas as tabelas disponibilizadas englobam os quatro cursos ofertados atualmente dentro do ICEA, sendo eles: Sistemas de informação, Engenharia da computação, Engenharia elétrica e Engenharia de produção. Os dados cedidos contam com informações desde o primeiro semestre de criação do curso até o semestre período letivo especial do ano de 2020 .

A Figura 2 apresenta a quantidade de arquivos, o tipo da planilha, e os campos inclusos em cada arquivo antes do tratamento dos dados.

Figura 2 – Descrição da base de dados

NOME DOS CAMPOS POR TIPO DE PLANILHA		
QUANTIDADE	NOME	CAMPOS
1	DADOS GERAIS	MATRICULA;SEXO;CIDADE NASCIMENTO;ESTADO NASCIMENTO;PAIS NASCIMENTO;BAIRRO FAMILIAR;CIDADE FAMILIAR;ESTADO FAMILIAR;PAIS FAMILIAR;BAIRRO ACADEMICO;CIDADE ACADEMICO;ESTADO ACADEMICO;PAIS ACADEMICO;REPUBLICA;TIPO REPUBLICA;DATA NASCIMENTO;ANO DE ADMISSAO;SEMESTRE DE ADMISSAO;COD MODO ADMISSAO;DESCRICAO MODO ADMISSAO;PONTUACAO VESTIBULAR;COD CURSO ADMISSAO;COD CURRICULO ADMISSAO;COD HABILITACAO ADMISSAO;COD POLO ADMISSAO;ORIGEM;TURNO;PRIMEIRO HORARIO;ULTIMO HORARIO;COD SITUACAO ALUNO;DESCRICAO SITUACAO ALUNO;CARGA HORARIA CURSO;CARGA HORARIA CURSADA;COD CURSO;CURSO;COD CURRICULO;COD HABILITACAO;COD ENFASE;COD POLO;POLO;CAIXA ARQUIVO;DATA INGRESSO;PARTICIPOU POLITICA AFIRMATIVA;USOU POLITICA AFIRMATIVA;MODALIDADE CONCORRENCIA;MODALIDADE CONCORRENCIA HOMOLOGADA;ALUNO COMPUTADO CENSO;ANO DIPLOMACAO;SEMESTRE DIPLOMACAO;
1	NOTAS	ANO;SEMESTRE;COD DEPARTAMENTO;COD DISCIPLINA;DESCRICAO;COD TURMA;MATRICULA;COR DA PELE;SEXO;COD CURSO;CARATER;MEDIA FINAL;EXAME ESPECIAL;FALTAS;SITUACAO;AULA DADA;GRAVACAO;DESCRICAO MODO ADMISSAO;TIPO ESCOLA
1	COLAÇÃO	ANO;SEMESTRE;MATRICULA;SEXO;TIPO;COD CURSO;CURSO;DATA COLACAO;DATA DIPLOMA;ENDereco FAMILIAR;BAIRRO FAMILIAR;DESCRICAO CIDADE;UF;CEP FAMILIAR;COEFICIENTE;COD CIDADE;DESCRICAO CIDADE;DESCRICAO ACENTUADA CIDADE;PREPOSICAO CIDADE
1	EVASÃO	MATRICULA;COD_CURSO;ANO_EVASAO;SEMESTRE_EVASAO;DT_EVASAO;DESCRICAO;CURSO;ENDereco_FAMILIAR;BAIRRO_FAMILIAR;CIDADE_FAMILIAR;CÓD_UF;CÓD_PAIS;CEP_FAMILIAR;MÓTIVO;FEITO POR REQUERIMENTO;SITUACAO ALUNO ATUAL

Fonte: Elaborado pela autora.

Após a seleção, os arquivos foram carregados para a ferramenta *Power BI* onde foram tratados na próxima etapa de pré-processamento.

3.2 Pré-Processamento

O pré-processamento foi dividido em quatro etapas, sendo a primeira referente à exclusão de tuplas, onde utilizamos dois critérios. O primeiro critério teve como objetivo descartar todos os alunos cujo campo “ALUNO COMPUTADO CENSO” constasse o valor “NÃO”, que refere-se a alunos que não deram prosseguimento na matrícula, ou seja, não chegaram a se tornar discentes. Estas tuplas foram excluídas pois não é de interesse do estudo analisar pessoas não matriculadas. O segundo critério foi desconsiderar tuplas duplicadas, pois elas poderiam gerar ruídos na análise.

A segunda etapa consiste na anonimização dos dados. Foi aplicado uma função do PowerBi para criptografar o número de matrícula de forma que não fosse possível identificar o aluno, a não ser que saiba os cálculos feitos sobre os dados. Esse cálculo só será disponibilizado para os gestores da universidade. Outro processo inserido foi a extração do ano de nascimento a partir da data de nascimento visando proteger a identificação dos alunos. Todas as informações sensíveis e pessoais que pudessem identificar o aluno foram removidas da base de dados.

A terceira etapa é referente à exclusão de colunas (atributos das tabelas), em que foram utilizados os seguintes critérios: (i) conterem alto índice de valores nulos (acima de 60%); (ii) conterem informação que não contribuíssem com os objetivos da visualização, como por exemplo, data de gravação do dado; (iii) conterem informação sem nenhuma diversidade, como por exemplo, cidade acadêmica. Os mesmos critérios também foram utilizados no trabalho de [Gonçalves, Silva e Cortes \(2018\)](#).

Para finalizar foi preciso padronizar colunas para facilitar dados que seriam utilizados na próxima etapa, a de transformação. Foi criada a coluna "CODTEMPO", sendo ela a concatenação da coluna "ANO" e coluna "SEMESTRE". Para a criação da coluna "CODLOCAL" que é a concatenação das colunas "CIDADE FAMILIAR" e "ESTADO FAMILIAR".

Toda a etapa de pré-processamento foi realizada dentro do *Power Query* utilizando a linguagem *M*, vale ressaltar que todos os dados apagados estão salvos na ferramenta, caso seja útil em algum momento revisitar esta etapa. Na Figura 3 é possível observar o resultado das tabelas após o pré-processamento.

Por fim, a Figura 4 apresenta um aspecto quantitativo dos dados, divididas por curso do Icea. A sumarização indica o número de semestres computados, o total de alunos, total de alunos evadidos, total de alunos diplomados, e o número total de notas correspondente à todas as tuplas de notas computadas por matéria e por aluno. Com essa figura é possível observar a visão geral do volume dos dados.

Figura 3 – Descrição da base de dados após o pré-processamento

NOME DOS CAMPOS POR TIPO DE PLANILHA APÓS O PRÉ-PROCESSAMENTO		
QUANTIDADE	NOME	CAMPOS
1	DADOS GERAIS	MATRICULA2;SEXO;CIDADE FAMILIAR;ESTADO FAMILIAR;PAIS FAMILIAR;DATA NASCIMENTO;ANO;SEMESTRE;COD MODO ADMISSAO;DESCRICAO MODO ADMISSAO;PONTUACAO VESTIBULAR;COD CURSO ADMISSAO;COD CURRICULO ADMISSAO;TURNO;COD SITUACAO ALUNO;DESCRICAO SITUACAO ALUNO;CARGA HORARIA CURSO;CARGA HORARIA CURSADA;COD CURSO;CURSO;COD ENFASE;PARTICIPOU POLITICA AFIRMATIVA;USOU POLITICA AFIRMATIVA;MODALIDADE CONCORRENCIA;MODALIDADE CONCORRENCIA HOMOLOGADA;ANO DIPLOMACAO;SEMESTRE DIPLOMACAO; CODTEMPO; CODLOCAL
1	NOTAS	MATRICULA2; ANO;SEMESTRE;COD DEPARTAMENTO;COD DISCIPLINA;DESCRICAO;COD TURMA;SEXO;COD CURSO;CARATER;MEDIA FINAL,EXAME ESPECIAL,FALTAS;SITUACAO;AULA DADA;DESCRICAO MODO ADMISSAO;TIPO ESCOLA
1	COLAÇÃO	MATRICULA2;ANO;SEMESTRE;COD CURSO;COEFICIENTE;
1	EVASÃO	MATRICULA2;COD_CURSO;ANO;SEMESTRE;DESCRICAO;MOTIVO;FEITO POR REQUERIMENTO;

Fonte: Elaborado pela autora.

Figura 4 – Sumarização dos dados

Sumarização dos dados					
CURSO	Número de semestres	# de Alunos	# alunos evadidos	# alunos Diplomados	# de notas
Sistemas de Informação	40	1054	548	238	31557
Engenharia de Computação	31	916	535	89	25446
Engenharia de Produção	43	1229	363	524	50247
Engenharia Elétrica	31	955	457	169	32496

Fonte: Elaborado pela autora.

3.3 Transformação

A transformação foi a etapa de transformar estas mesmas 4 fontes em um banco de dados multidimensional. Essa modelagem foi escolhida pois facilita as consultas e a ferramenta *PowerBi* oferece estrutura para esse tipo de modelo. Não foi utilizado um *Data Warehouse* pois no momento o relatório não tem comportamento de atualizar com alta frequência e nem passar por informações transacionais. No primeiro momento, o conjunto de dados foi analisado para identificar quais são as tabelas fatos e quais são as tabelas dimensões. A tabela fato é aquela que representa um acontecimento, foi identificado 4 fatos, sendo eles, a admissão, a colação, o desempenho, e a evasão.

A tabela fato FT_admissao, é uma réplica da tabela de dados gerais. Nesse processo de transformação desta tabela foram retiradas todos e quaisquer atributos que não tinham relação com o fato de admitir um aluno. As colunas que tinham relação com a admissão, mas que oferecem diferentes perspectivas sobre esse fato foram replicados em outro momento em tabelas dimensões, por exemplo, a perspectiva de aluno e de tempo. Para conseguir uma transformação eficiente com a tabela de dados gerais, foi preciso um pouco mais de cuidado, dado que continham atributos que alimentam as dimensões "DIM_Tempo", "DIM_Aluno", "DIM_Local" e a própria tabela fato de admissão. Então foi preciso atenção para não duplicar atributos nas tabelas e/ou não ultrapassar o limite do tema de cada fato/dimensão.

A tabela fato FT_Desempenho é uma réplica da tabela de nota, exceto das informações referente a alunos e matérias. Permanecendo assim, dados relacionados às notas, faltas, e as chaves primárias das dimensões que contém relação. A tabela fato FT_Colacao passou pelo processo de transformação das informações qualitativas de aluno, mantendo assim apenas os códigos para relacionamento e o coeficiente de rendimento (CR) referente ao aluno graduado. Por fim, a tabela fato FT_Evasão é uma réplica da tabela de evasão pré processada contendo apenas as chaves para relacionamento e o motivo de evasão.

Seguindo para as tabelas de dimensão, que caracterizam, as tabelas fatos, temos 4 dimensões transformadas, a dimensão aluno, local, tempo e matéria. A dimensão DIM_Materia é uma réplica da tabela nota pré processada, com a chave da matéria, o departamento e curso que está relacionada e o nome da mesma. Essa tabela passou por um outro pré processamento de retirada de chaves duplicadas para seguir a metodologia de modelagem multidimensional, sendo a relação da tabela fato com a tabela dimensão de muitos para um. A dimensão matéria oferece perspectiva para a tabela FT_Desempenho.

A tabela dimensão DIM_Local é uma réplica da tabela pré processada de dados gerais apenas com colunas relacionadas à cidade, estado e país do aluno ingressante antes de se matricular na universidade. A tabela dimensão local oferece a perspectiva de local

para o fato FT_admissao. A tabela dimensão DIM_Aluno é uma réplica da tabela dados gerais apenas com colunas que caracterizam esse aluno específico. Foi realizado um pré processamento para retirar as duplicatas visando o relacionamento entre fato e dimensão de muitos para um. A tabela dimensão DIM_Tempo é a agregação das colunas "codtempo", "ano" e "semestre" que se encontram presentes nas quatro tabelas de fonte de dados. Após a agregação, as duplicatas da tabela foram retiradas visando o relacionamento de tabelas fato para com as tabelas dimensão de muitos para um.

As tabelas de dimensão DIM_Aluno e DIM_Tempo oferecem uma perspectiva para todas as quatro tabelas de fato, dado que colação, evasão, admissão e desempenho, sempre estão associadas a um aluno e a um tempo específico.

Os cursos são colunas presentes em todas as tabelas e são consideradas como dimensões degeneradas. As dimensões descaracterizadas, também conhecidas como dimensões degeneradas, são representadas na tabela de fatos como chaves de dimensão, sem que exista a tabela de dimensão (SOARES, 1998).

Após toda a etapa de transformação, o passo seguinte foi gerenciar os relacionamentos das tabelas. Como todas as chaves já estavam criadas, pré-processadas e transformadas nas tabelas corretas, existe uma funcionalidade da ferramenta *Power BI* chamada "Identificar relações". Esta funcionalidade identifica as tabelas que compartilham as mesmas chaves e criam os relacionamentos automaticamente. Após todo o processo de modelagem a próxima etapa é a de visualização. Na Figura 5 podemos observar o resultado da transformação.

Figura 5 – Resultado da transformação



Fonte: Elaborado pela autora.

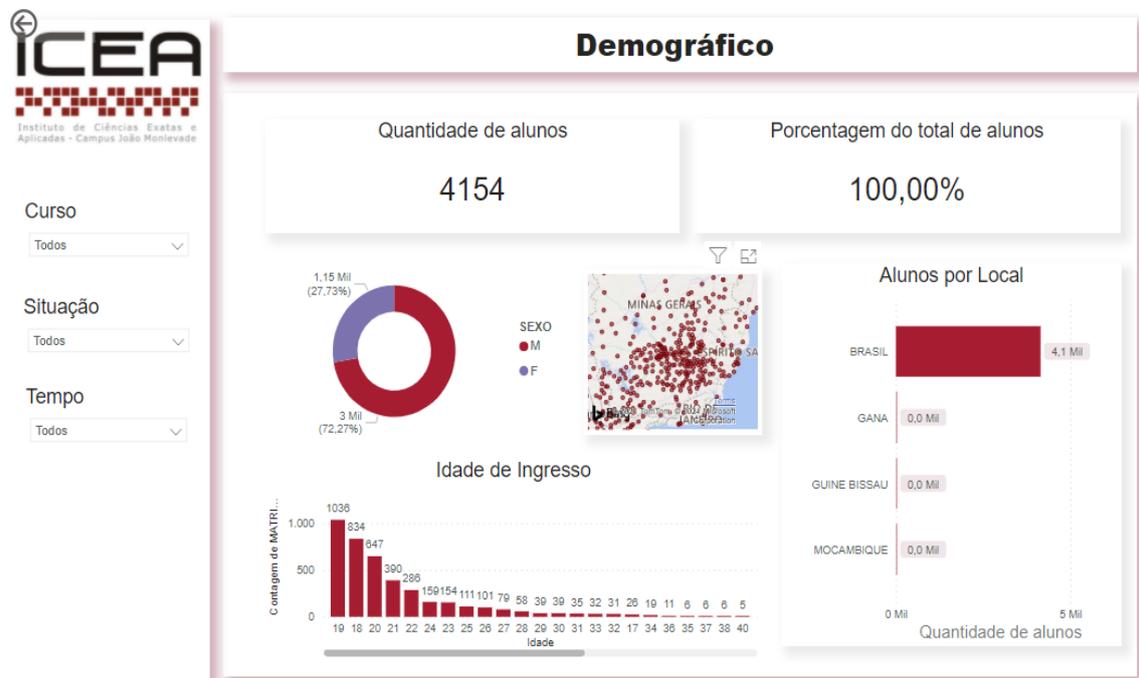
4 Resultados

O relatório é composto por sete *dashboards* e uma página inicial referente ao menu de navegação. Três *dashboards* são referentes aos fatos evasão, admissão e colação respectivamente. Três *dashboards* são referentes ao desempenho, com perspectiva de desempenho de aluno, matéria e departamento. O último *dashboard* aborda informações gerais demográficas dos ingressantes do Icea.

4.1 *Dashboard* demográfico

O objetivo do *dashboard* demográfico na Figura 6, é apresentar o perfil do aluno do Icea, com informações gerais que podem ser filtradas por ano, por semestre, por curso e por situação do aluno, sendo os tipos de situação aprovado, evadido, matriculado, trancado e afastado. Este *dashboard* é composto por seis visualizações, sendo dois cartões, um gráfico de barra, um gráfico de coluna, um gráfico de rosca e um mapa.

Figura 6 – *Dashboard* demográfico



Fonte: Elaborado pela autora.

O mapa apresenta de qual cidade os alunos saíram para estudar no campus Icea. É possível reparar os locais com maior ponto de concentração, que neste caso é Minas Gerais, o que faz sentido dado que o campus fica na cidade de João Monlevade no mesmo estado.

O gráfico de barra apresenta a contagem dos alunos por local de onde vieram, essa visualização apresenta os dados com 3 níveis de hierarquia, a contagem total de alunos por país, por estado e por cidade. A possibilidade de aumentar o nível de detalhamento se chama *drill down*, e a possibilidade de diminuir o nível de detalhamento é chamada de *drill up*. Como mostra a Figura 7

Figura 7 – *Dashboard* demográfico: gráfico de coluna a nível de estado

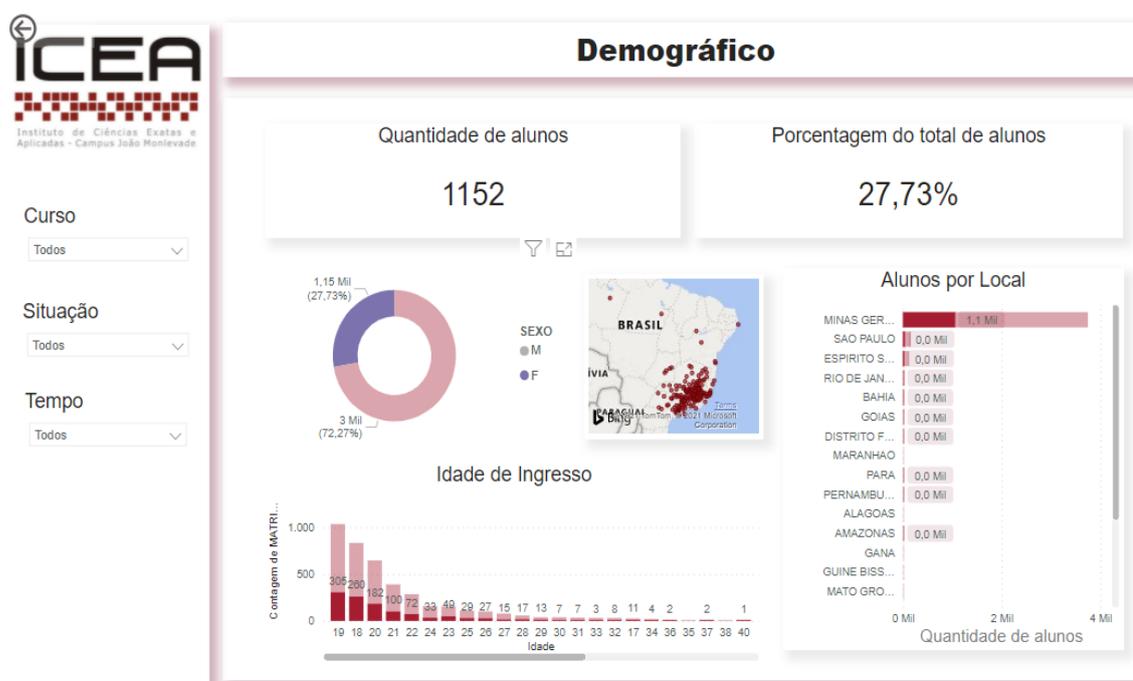


Fonte: Elaborado pela autora.

O gráfico de rosca apresenta a proporção de pessoas de gêneros femininos e masculinos dentro do Icea. A Figura 8 apresenta o comportamento do *dashboard* ao selecionar o dado "Feminino", e como todas as outras visualizações filtram os dados quando alguma informação específica é selecionada.

O gráfico de coluna é feito a partir do cálculo da data do ano de ingresso do aluno menos a data de nascimento do mesmo, concluindo assim a quantidade de pessoas que entraram com cada idade.

Figura 8 – Dashboard demográfico: filtro alunos no sexo feminino

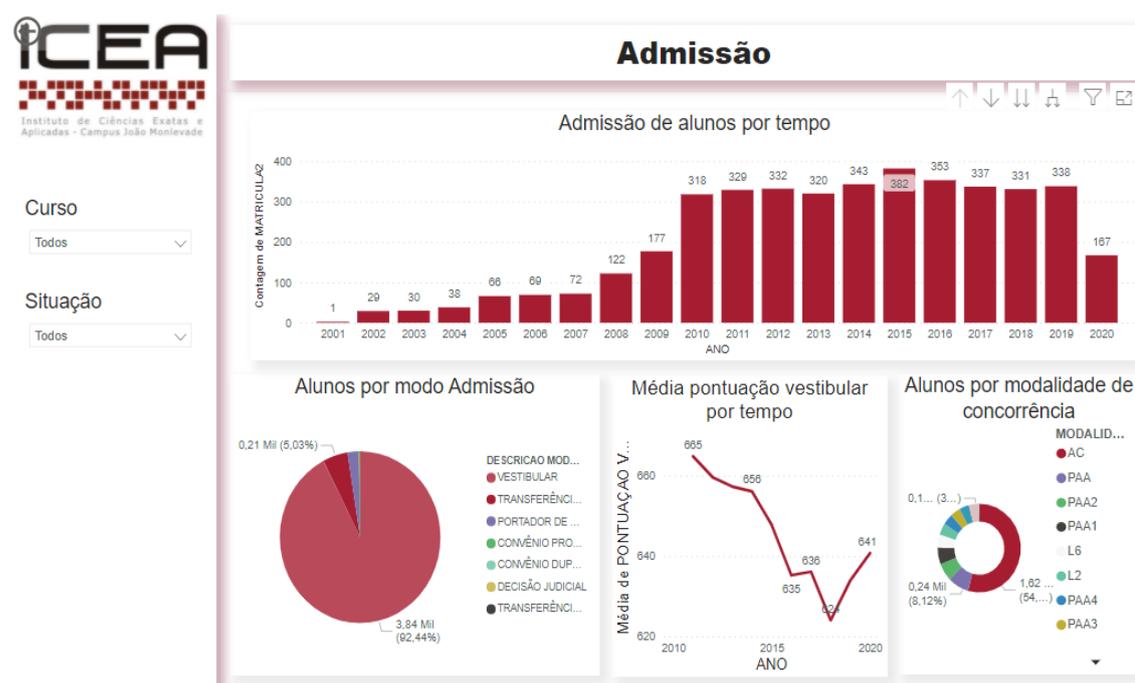


Fonte: Elaborado pela autora.

4.2 Dashboard de admissão

O *dashboard* de admissão tem como objetivo apresentar as informações relacionadas ao ingresso dos alunos do Icea. É composto por dois filtros, sendo eles curso, os tipos de situação atual dos alunos na universidade. Este *dashboard* possui quatro visualizações, sendo elas um gráfico de coluna, um gráfico de pizza, um gráfico de linha e um gráfico de rosca, como visto na figura 9.

Figura 9 – *Dashboard* de admissão



Fonte: Elaborado pela autora.

O gráfico de coluna apresenta a contagem total de alunos admitidos por tempo, com operações de *drill down* e *drill up* oferecendo perspectiva por ano e por semestre, como mostra a Figura 10.

O gráfico de pizza apresenta a proporção da admissão dos alunos por modo de admissão, o gráfico de linha apresenta a média de nota do vestibular dos alunos ingressantes. Este gráfico apresenta dados a partir de 2011, pois foi quando a universidade aderiu ao ENEM como forma de processo seletivo.

O gráfico de rosca apresenta o total e a proporção dos alunos pela modalidade de concorrência, como por exemplo, ampla concorrência e cotas. Os nomes e as especificações das cotas se alteraram de acordo com as edições dos editais, assim como a forma de calcular as notas do vestibular. É importante ressaltar que todas as visualizações respondem às interações com os filtros, modificando os dados visualizados.

Figura 10 – *Dashboard* de admissão: gráfico de coluna a nível de semestre

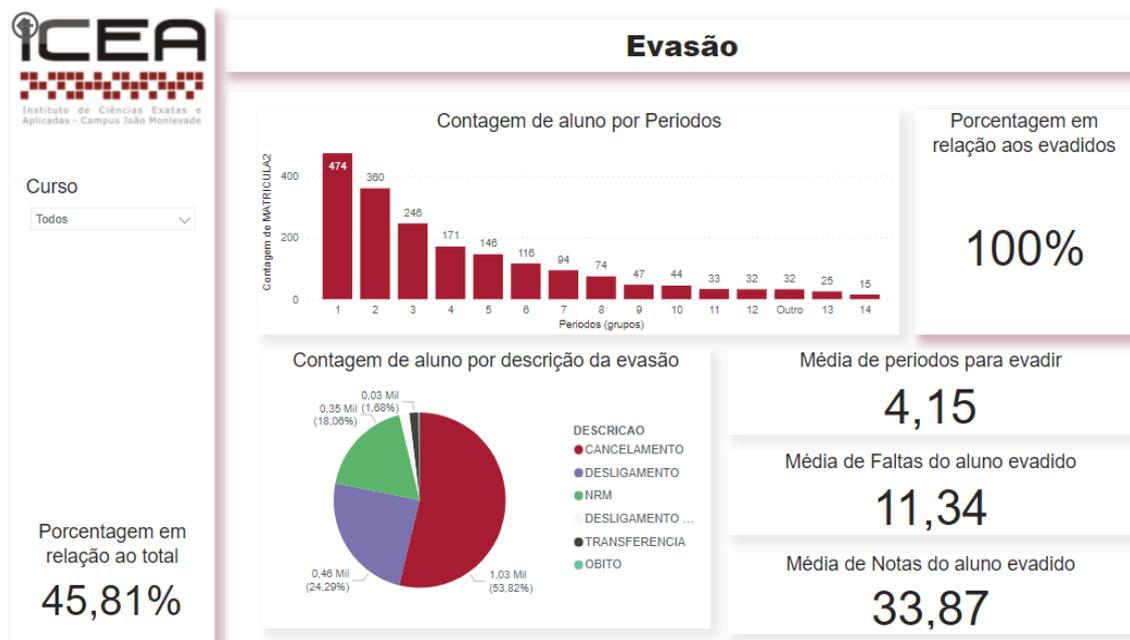


Fonte: Elaborado pela autora.

4.3 Dashboard de evasão

O objetivo do *dashboard* de evasão é apresentar as principais informações atreladas aos alunos evadidos. O mesmo é composto por dois filtros, sendo eles de curso e de departamento, e sete visualizações. A Figura 11 mostra as sete visualizações que compõem o *dashboard*, um gráfico de coluna, um gráfico de pizza e cinco cartões.

Figura 11 – *Dashboard* de evasão



Fonte: Elaborado pela autora.

O gráfico de coluna apresenta a contagem de aluno evadido por período. Ao selecionar algum dos períodos do gráfico, o cartão na lateral direita apresenta a porcentagem referente ao total de alunos do período relacionado. No exemplo da Figura 12, podemos observar que ao selecionar as evasões no primeiro período, o *dashboard* se adapta e apresenta o cartão com a porcentagem referente à proporção de alunos que evadiram no primeiro período.

O cartão à esquerda do *dashboard*, localizado embaixo do filtro de curso apresenta a porcentagem de alunos evadidos perante ao total de alunos ingressos antes no curso. Por exemplo, na Figura 13 podemos observar que 58% dos alunos de Engenharia de computação são considerados evadidos.

Os outros cartões apresentam a média de períodos para evadir, média comum de faltas e a média de notas de alunos evadidos respectivamente. A última visualização é um gráfico de pizza que representa a parcela das evasões em relação ao motivo pelo qual o discente evadiu.

Figura 12 – *Dashboard* de evasão: filtro alunos evadidos no primeiro período



Fonte: Elaborado pela autora.

Figura 13 – *Dashboard* de evasão: filtro do curso de Engenharia de Computação

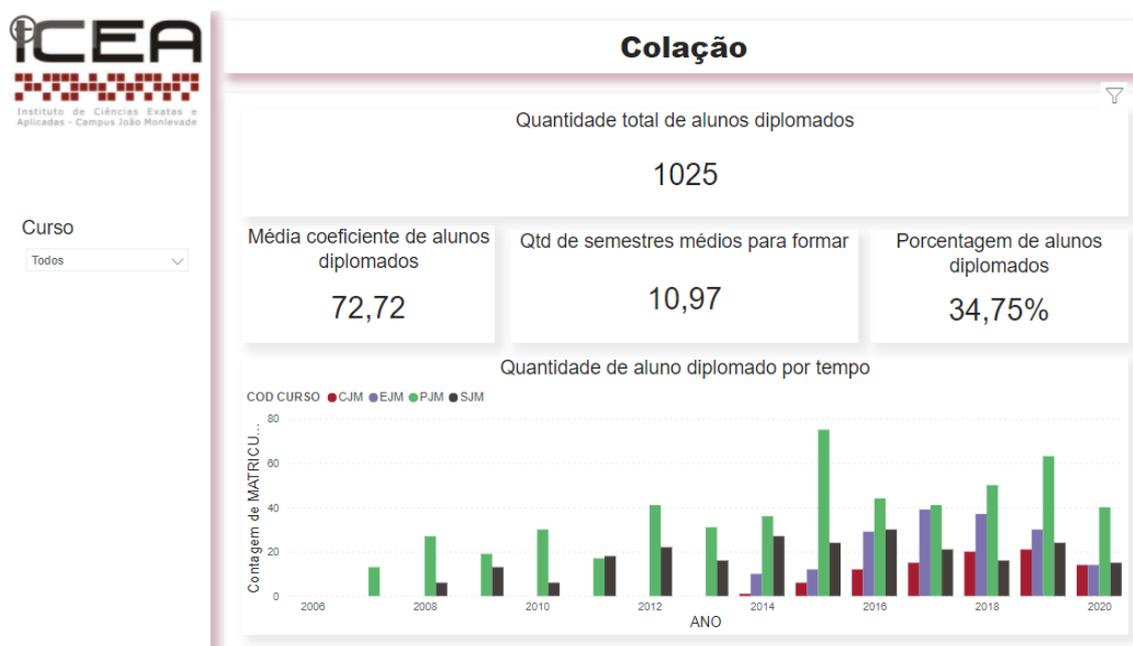


Fonte: Elaborado pela autora.

4.4 Dashboard de colação

O objetivo do *dashboard* de colação é apresentar os dados referentes aos alunos diplomados da universidade no campus ICEA. O *dashboard* apresentado na Figura 14 é composto por 5 visualizações, sendo quatro cartões e um gráfico de coluna.

Figura 14 – *Dashboard* de colação



Fonte: Elaborado pela autora.

O primeiro cartão apresenta a quantidade total de alunos diplomados, na segunda linha o primeiro cartão mais à esquerda é a média de coeficiente dos alunos que formaram, o segundo é a quantidade média de semestres para formar e o terceiro, a porcentagem total de alunos formados, ou seja, total de alunos formados dividido pelo total de alunos ingressantes, desconsiderando os que ainda estão matriculados.

A última visualização é um gráfico de coluna que apresenta o total de alunos formados por tempo, em que cada cor de coluna representa um curso dentro do ICEA.

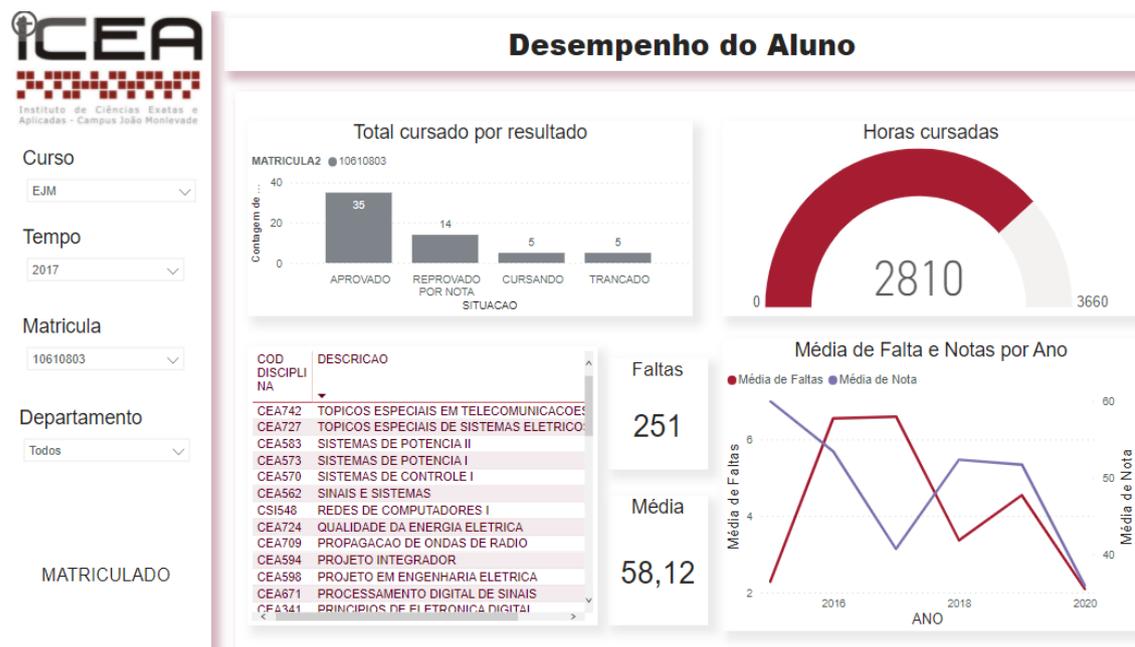
4.5 Dashboard de desempenho aluno

O *dashboard* de desempenho do aluno apresentado na Figura 15 tem como objetivo apresentar o histórico escolar de forma visual. No mesmo existe 4 filtros, de curso, tempo (considerando ano e semestre), de matrícula, (ressaltando que a mesma passou por um processo de anonimização), e por último de departamento. Dentro deste *dashboard* podemos acompanhar sete visualizações, sendo elas um indicador, três cartões, uma tabela, um gráfico de linha e um gráfico de coluna.

O gráfico de coluna chamado "Total cursado por resultado" apresenta a quantidade de matérias feitas por aquele aluno de acordo o resultado que teve na matéria, sendo os tipos de resultados aprovado, reprovado por nota, reprovado por falta, reprovado por falta e nota, cancelado, trancado e cursando.

O indicador de horas cursadas apresenta o total de horas que precisam ser integralizadas no curso, e quantas horas que o aluno em específico já cursou, no exemplo da Figura 15 o aluno em questão cursou 2810 horas de um total de 3660.

Figura 15 – *Dashboard* de desempenho aluno



Fonte: Elaborado pela autora.

A tabela apresenta o código e o nome das matérias cursadas pelo aluno. No gráfico de linha podemos observar as médias de notas e de faltas perante os anos, para conhecer o desempenho do aluno durante o tempo cursado. O cartão "Faltas" apresenta o total de faltas cometidas, assim como o cartão "Média" a média geral daquele aluno, considerando apenas as matérias que as situações são, aprovado, e reprovado, sendo por média, por falta ou ambos, dado que as situações cursando, cancelado e trancado apresentam nota 0 e não devem entrar no cálculo.

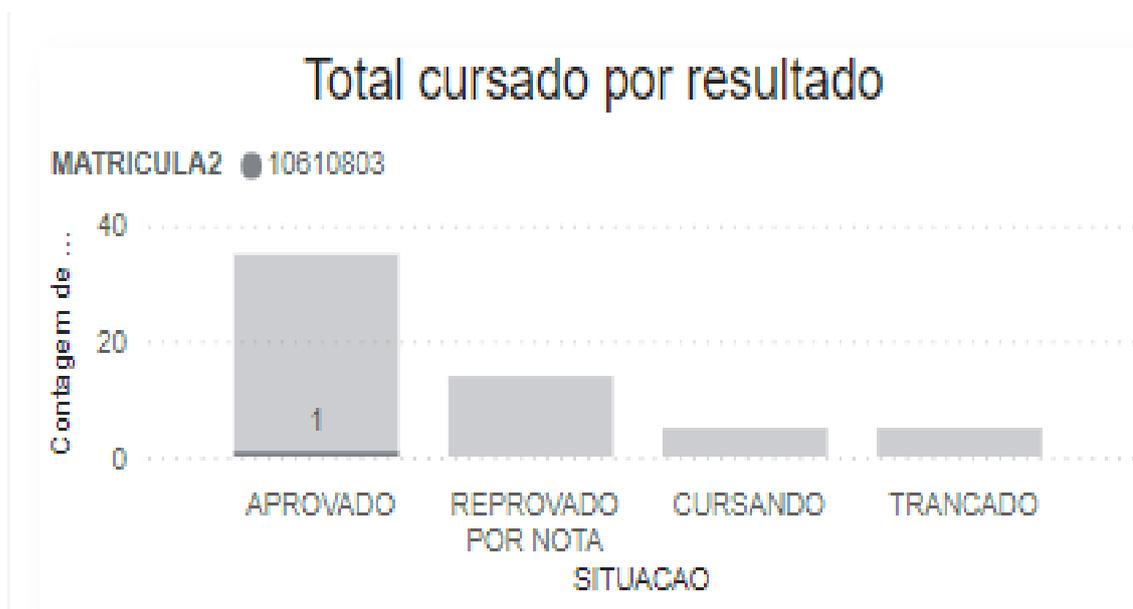
O último cartão apresentado no canto inferior esquerdo é a situação atual deste aluno, se ele está matriculado, evadido, diplomado, trancado ou afastado. Todas as visualizações podem ser filtradas por quaisquer informações dentro do gráfico. Por exemplo, ao selecionar a disciplina CEA594 Projeto Integrador, como visto na Figura 16 e na Figura 17, temos o conhecimento de qual o resultado obtido. Neste caso aprovado, com nota de 80 pontos e o total de 1 falta no ano de 2020.

Figura 16 – Dashboard de desempenho aluno: filtro de matéria foco em nota e falta



Fonte: Elaborado pela autora.

Figura 17 – Dashboard de desempenho aluno: filtro de matéria foco em situação aprovado



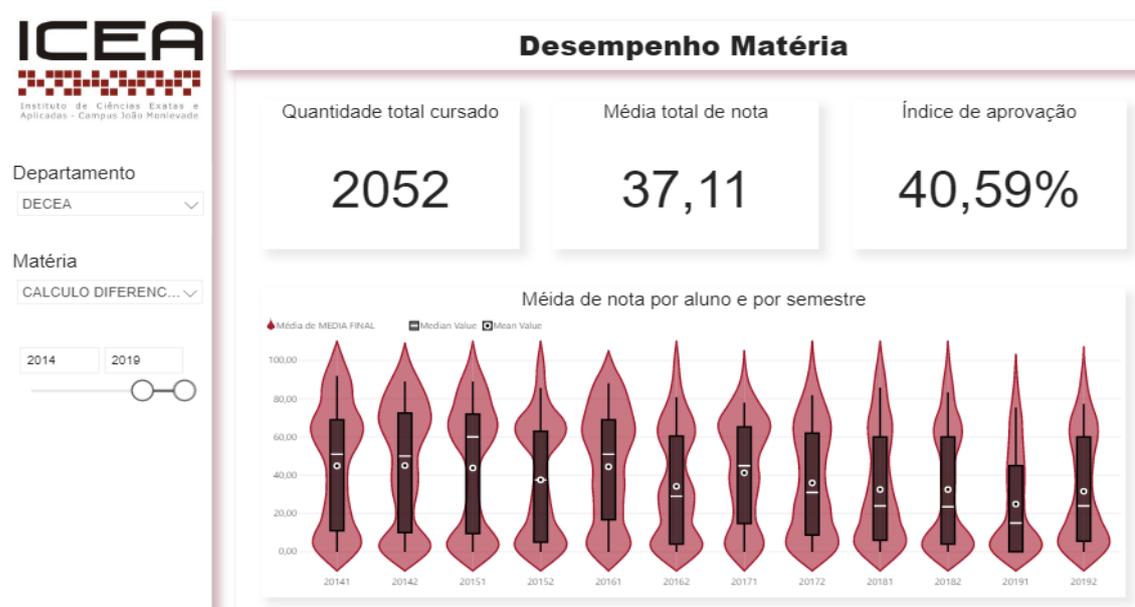
Fonte: Elaborado pela autora.

4.6 Dashboard de desempenho matéria

O dashboard de desempenho da matéria, apresentado na Figura 18, contém três filtros e quatro visualizações. A primeira visualização é denominada "Quantidade total cursada" que significa a quantidade de vezes que aquela matéria e ou conjuntos de matéria foi cursada naquele período de tempo e naquele departamento. A segunda visualização é um cartão que representa a média de todas as notas. A terceira visualização corresponde ao índice de aprovação, ou seja de todas as vezes que essa matéria foi cursada, quantos por cento das vezes o resultado foi uma aprovação. É importante ressaltar que para ser aprovado é necessário que a nota seja maior que 60% e as faltas sejam menores do que 18 naquele semestre.

Como podemos observar no exemplo da Figura 18, a disciplina Cálculo diferencial e integral 1 no departamento Decea, e no período de tempo de 2014 até 2019 foi cursado 2052 vezes, a média de notas é 37,11 e o índice de aprovação referente corresponde a 40,59%.

Figura 18 – Dashboard de desempenho matéria



Fonte: Elaborado pela autora.

A quarta e última visualização é um *violin plot*, é uma combinação de um box plot e um gráfico de densidade de kernel. A visualização de um violin vem acompanhado com as informações de um *blox plot*, ou seja, número de amostra, primeiro, segundo, terceiro quartil, mediana e média.

Os quartis são valores nessa escala que dividem o conjunto de dados em quatro partes, todas elas com o mesmo número de observações. Isso significa que 25% das observações são menores que o primeiro quartil, 50% são menores que o segundo quartil e 75% são menores que o terceiro quartil (FARIAS, 2014).

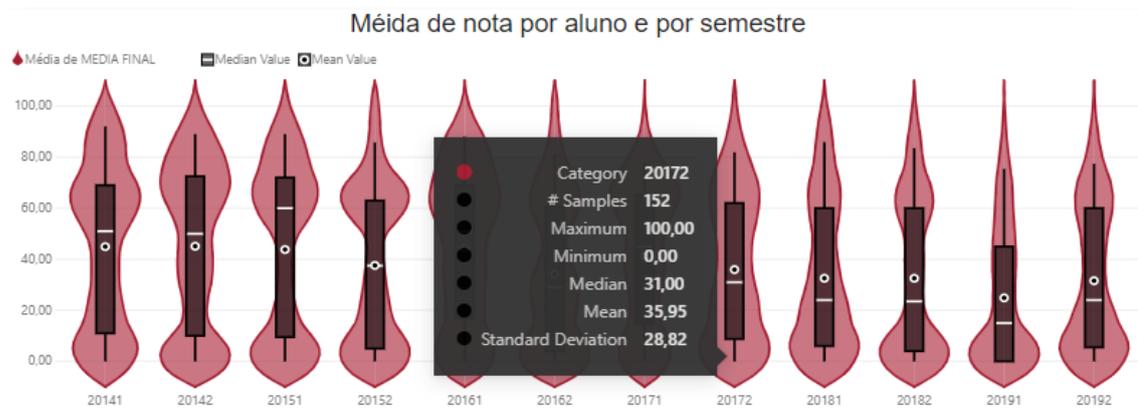
A média é a soma de todos os valores dividido pelo número de amostras, representadas no *violin plot* por um círculo no meio do *box plot*. A mediana é a média dos valores centrais, representado por um traço na visualização. O valor máximo e valor mínimo é representado pelo final e início do traço apresentado no *box plot*.

Um estimador de densidade de kernel é um suavização dos dados do histograma que fornece uma estimativa da densidade de uma variável (HAUGHTON; PHONG, 2004). Um histograma apresenta a distribuição de frequência de uma variável.

O eixo X representa os semestres referentes aos dados e o eixo Y representa a variação da nota que vai de 0 a 100%. Cada *violin plot* apresenta a frequência de um conjunto de dados de dados numéricos, sendo eles a distribuição das notas individuais dos alunos, considerando os alunos que fizeram aquela matéria naquele semestre.

Ao passar o mouse por cima de um *violin*, a visualização apresenta as informações de amostra, desvio padrão, o valor mínimo, valor máximo, média e mediana daquele conjunto de dados. Na Figura 19 podemos observar o comportamento desta interação. Com essa visualização podemos identificar a concentração de desempenho daquela matéria ao longo do tempo e comparar suas variações.

Figura 19 – *Dashboard* de desempenho matéria: detalhes da visualização

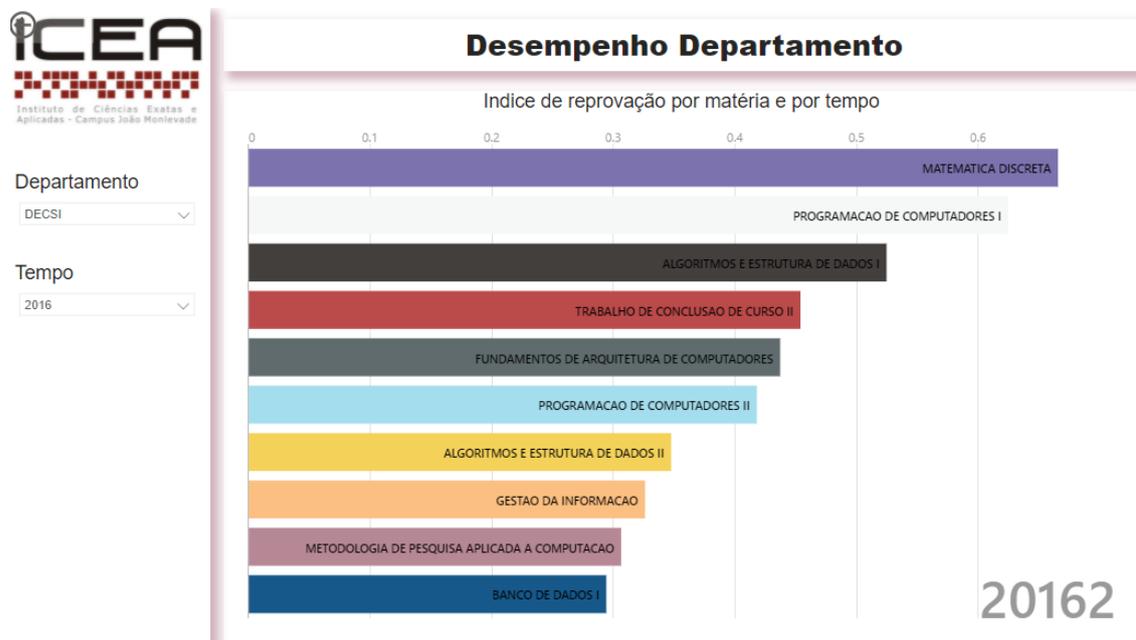


4.7 Dashboard de desempenho departamento

O *dashboard* de desempenho departamento apresentado na Figura 20 contém a visualização denominada *Bar Chart Race*, é um gráfico de barra animado que varia com o tempo. Ele apresenta de forma interativa as 10 matérias com maior índice de reprovação, passando pelos semestres computados na base de dados. Este gráfico é comumente utilizado em redes sociais para representar por exemplo o *valuation* de empresas pelos anos. Também foi utilizado para representar os casos de Covid-19 durante o tempo, comparando diferentes países.

Dentro do *dashboard* podemos deixar ele sem filtros, onde será gerado um ranqueamento de todas as matérias ofertadas no Icea ou podemos filtrar as matérias por departamento. Caso o decisor deseje analisar um período específico ele também pode selecionar um ou mais semestres para visualizar o gráfico estatístico ou a animação dos períodos selecionados respectivamente. Dessa forma, podemos medir o desempenho das matérias, quais se mantiveram no topo, ou seja, com maior índice de reprovação, e conseqüentemente, precisam de mais atenção. No exemplo da Figura 20, é possível observar as 5 matérias com maior índice de reprovação no semestre de 2016.2, considerando apenas o dentro do Decsi.

Figura 20 – Dashboard de desempenho departamento



Fonte: Elaborado pela autora.

5 Conclusão

Neste capítulo são apresentadas as conclusões sobre a pesquisa, seus objetivos, resultados alcançados e trabalhos futuros. O problema identificado está relacionado ao aproveitamento das ferramentas de visualização para extrair informações relevantes dos dados disponíveis.

A versatilidade dos dados permitiu observar diversos aspectos e fatos que estão na nossa rotina. O processo de preparação facilitou a construção da modelagem e visualização dos dados. A flexibilidade e interatividade, entrega o valor de adaptar a informação de acordo com as necessidades de cada gestor. Com a ferramenta de BI foi possível criar um conjunto de *dashboards* interativos que podem auxiliar as decisões de diversos níveis de gestão, como diretoria, departamento e colegiado de curso.

A utilização e propagação desse conhecimento proporcionado tende a trazer vantagens para toda comunidade acadêmica. O trabalho pode ser o meio de acompanhamento de alguns dos indicadores citados no PID. Como por exemplo, no objetivo "Promover a melhoria da qualidade do curso de graduação". A meta deste objetivo é acompanhar os dados de admissão como, número de ingressantes, notas de vestibular, modalidade de ingresso e desempenho da disciplina. Todos esses dados podem ser acompanhados no *dashboard* de admissão, evasão e desempenho matéria respectivamente. No objetivo de "Implementar mecanismos de acolhimento, recepção e acompanhamento dos graduandos, uma das ações propostas é criar um mecanismo de acompanhamento dos alunos, e podemos realizar este acompanhamento pelo *dashboard* desempenho aluno. Além de que, o plano de ação deseja ter esse monitoramento em diferentes perspectivas, como tempo, departamento, aluno, matérias e cursos, e o relatório oferece exatamente esses respectivos pontos de vista.

Uma outra proposta de utilização é os professores mentores de turmas do Icea conseguirem acompanhar esses alunos de forma mais simples, clara e interativa. Será possível analisar a média da turma, o andamento dos alunos, quais os alunos com alto índice de reprovação e em que momento este se encontra no curso. Desta forma será possível uma orientação direcionada e assertiva, pois existe algumas das informações mais importantes disponíveis em um só relatório.

A nível de departamento pode se entender a situação do aluno que tem interesse de abrir vaga em disciplinas, se é uma necessidade ou um desejo. Pelo *dashboard* de desempenho aluno, é possível analisar o contexto do aluno para auxiliar o chefe de departamento a tomar esta decisão. Os decisores do departamento são capazes de acompanhar as disciplinas e tomar decisões que possam mudar o cenário das disciplinas com um histórico de alto índice de reprovação. Além de avaliar se decisões tomadas anteriormente surgiram os

efeitos desejados. Como por exemplo no passado foi decidido separar a turma de Geometria Analítica e Álgebra Linear do curso de Sistemas de Informação com os demais cursos do Icea, com a *dashboard* de matéria é possível reparar se essa decisão reduziu o índice de reprovação.

Foi criado também dois anexos, um com instruções de como manter o relatório atualizado e outro de como compartilhar este relatórios com os demais decisores.

Conclui-se que, ao explorar o potencial dos dados e transformá-los em visualizações pertinentes, os mesmos podem ser utilizados como um sistema de apoio a decisões em contextos educacionais do Icea.

5.1 Trabalhos Futuros

Algumas melhorias podem ser feitas para aprimorar o relatório gerado, como, explorar as visualizações preditivas e prescritivas. Esse tipo de estatística possibilita agir antes mesmo de uma situação se tornar um problema. Pode-se apontar os alunos com risco de evasão, matérias que tendem a subir o índice de reprovação, entre outros contextos. Análises preditivas possibilitam a criação de um plano de ação afim de evitar pioras nos índices do campus.

Outro contexto a ser melhorado é realizar a integração das fontes de dados com os relatórios, facilitando atualizações automáticas e provendo uma maior confiabilidade nos dados, pois a frequência de atualizações será maior. O trabalho também pode ser expandido para outros institutos dentro da Ufop ou até mesmo para outras universidades. Além disso, é possível explorar outras perspectivas como por exemplo um *dashboard* de curso ou criar *dashboards* voltados para programas de pós graduação.

Uma outra melhoria proposta é utilizar dados específicos de cada matéria cursada, para entender o desempenho dos alunos em relação as práticas pedagógicas (trabalhos e provas) expostas durante o semestre, extraindo a base de dados de outras fontes.

Referências

- ALJOHANI, N. R.; DAVIS, H. C. *Learning analytics and formative assessment to provide immediate detailed feedback using a student centered mobile dashboard*. [S.l.], 2013. 262-267 p. Citado 2 vezes nas páginas 12 e 13.
- FARIAS, A. M. L. de. O boxplot. *Universidade Federal Fluminense-Instituto de Matemática*, 2014. Citado na página 33.
- FAYYAD, U. M. et al. Knowledge discovery and data mining: Towards a unifying framework. In: *KDD*. [S.l.: s.n.], 1996. v. 96, p. 82–88. Citado na página 15.
- FEW, S. *Information dashboard design: The effective visual communication of data*. [S.l.]: O'Reilly Media, Inc., 2006. Citado na página 14.
- GONÇALVES, T. C.; SILVA, J. C. da; CORTES, O. A. C. Técnicas de mineração de dados: um estudo de caso da evasão no ensino superior do instituto federal do maranhão. *Revista Brasileira de Computação Aplicada*, v. 10, n. 3, 2018. Citado 2 vezes nas páginas 16 e 19.
- HAUGHTON, D.; PHONG, N. Graphical and numerical descriptive analysis: Exploratory tools applied to vietnamese data. *Journal of Statistics Education*, Taylor & Francis, v. 12, n. 2, 2004. Citado na página 34.
- ROMERO, O.; ABELLÓ, A. A survey of multidimensional modeling methodologies. *International Journal of Data Warehousing and Mining (IJDWM)*, IGI Global, v. 5, n. 2, p. 1–23, 2009. Citado 2 vezes nas páginas 14 e 15.
- SANTOS, J. S. d. Business intelligence: uma proposta metodológica para análise da evasão escolar em instituições federais de ensino. 2017. Citado 2 vezes nas páginas 16 e 17.
- SOARES, V. J. d. A. Modelagem incremental no ambiente de data warehouse. *Rio de Janeiro*, 1998. Citado na página 22.
- UFOP. *PDI UFOP 2016-2025*. 2016. Disponível em: <<https://ufop.br/noticias/institucional/pdi-ufop-2016-2025>>. Citado na página 15.
- ZHU, Y. A data driven educational decision support system. *International Journal of Emerging Technologies in Learning (iJET)*, v. 13, n. 11, p. 4–16, 2018. Citado na página 12.

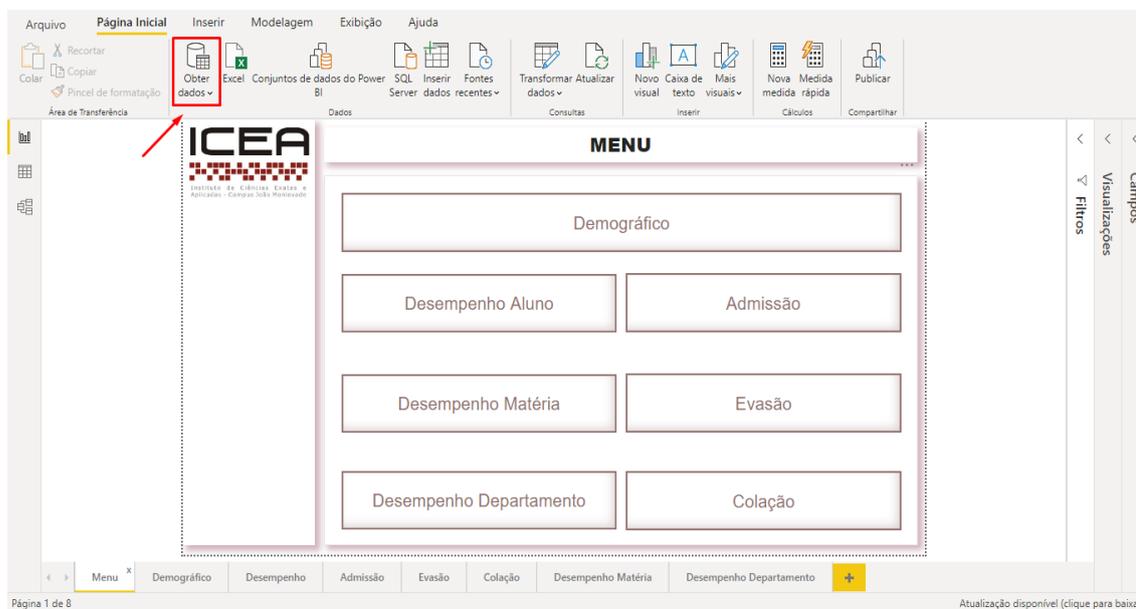
Anexos

ANEXO A – Como inserir novos dados

- a) Entre em contato com a seção de ensino do campus ICEA;
- b) Solicite arquivos com nome “dados_matricula.csv”, “notas_matricula.csv”, “dados_evasao.csv” e “dados_colacao.csv” o formato CSV e com colunas com a respectiva ordem e nome;
 - MATRICULA; SEXO; CIDADE NASCIMENTO; ESTADO NASCIMENTO; PAIS NASCIMENTO; BAIRRO FAMILIAR; CIDADE FAMILIAR; ESTADO FAMILIAR; PAIS FAMILIAR; BAIRRO ACADEMICO; CIDADE ACADEMICO; ESTADO ACADEMICO; PAIS ACADEMICO; REPUBLICA; TIPO REPUBLICA; DATA NASCIMENTO; ANO DE ADMISSAO; SEMESTRE DE ADMISSAO; COD MODO ADMISSAO; DESCRICAO MODO ADMISSAO; PONTUACAO VESTIBULAR; COD CURSO ADMISSAO; COD CURRICULO ADMISSAO; COD HABILITACAO ADMISSAO; COD POLO ADMISSAO; ORIGEM; TURNO; PRIMEIRO HORARIO; ULTIMO HORARIO; COD SITUACAO ALUNO; DESCRICAO SITUACAO ALUNO; CARGA HORARIA CURSO; CARGA HORARIA CURSADA; COD CURSO; CURSO; COD CURRICULO; COD HABILITACAO; COD ENFASE; COD POLO; POLO; CAIXA ARQUIVO; DATA INGRESSO; PARTICIPOU POLITICA AFIRMATIVA; USOU POLITICA AFIRMATIVA; MODALIDADE CONCORRENCIA; MODALIDADE CONCORRENCIA HOMOLOGADA; ALUNO COMPUTADO CENSO; ANO DIPLOMACAO; SEMESTRE DIPLOMACAO;
 - ANO; SEMESTRE; COD DEPARTAMENTO; COD DISCIPLINA; DESCRICAO; COD TURMA; MATRICULA; COR DA PELE; SEXO; COD CURSO; CARATER; MEDIA FINAL; EXAME ESPECIAL; FALTAS; SITUACAO; AULA DADA; GRAVACAO; DESCRICAO MODO ADMISSAO; TIPO ESCOLA;
 - MATRICULA; COD_CURSO; ANO_EVASAO; SEMESTRE_EVASAO; DT_EVASAO; DESCRICAO; CURSO; ENDERECO_FAMILIAR; COD_UF; CIDADE_FAMILIAR; BAIRRO_FAMILIAR; COD_PAIS; CEP_FAMILIAR; MOTIVO; FEITO POR REQUERIMENTO; SITUACAO ALUNO ATUAL;
 - ANO; SEMESTRE; MATRICULA; SEXO; TIPO; COD CURSO; CURSO; DATA COLACAO; DATA DIPLOMA; ENDERECO FAMILIAR; BAIRRO FAMILIAR; DESCRICAO CIDADE; UF; CEP FAMILIAR; COEFICIENTE; COD CIDADE; DESCRICAO CIDADE; DESCRICAO ACENTUADA CIDADE; PREPOSICAO CIDADE.

- c) Mova os arquivos para a mesma pasta que o relatório.;
- d) Abra o relatório;
- e) Clique em obter dados;

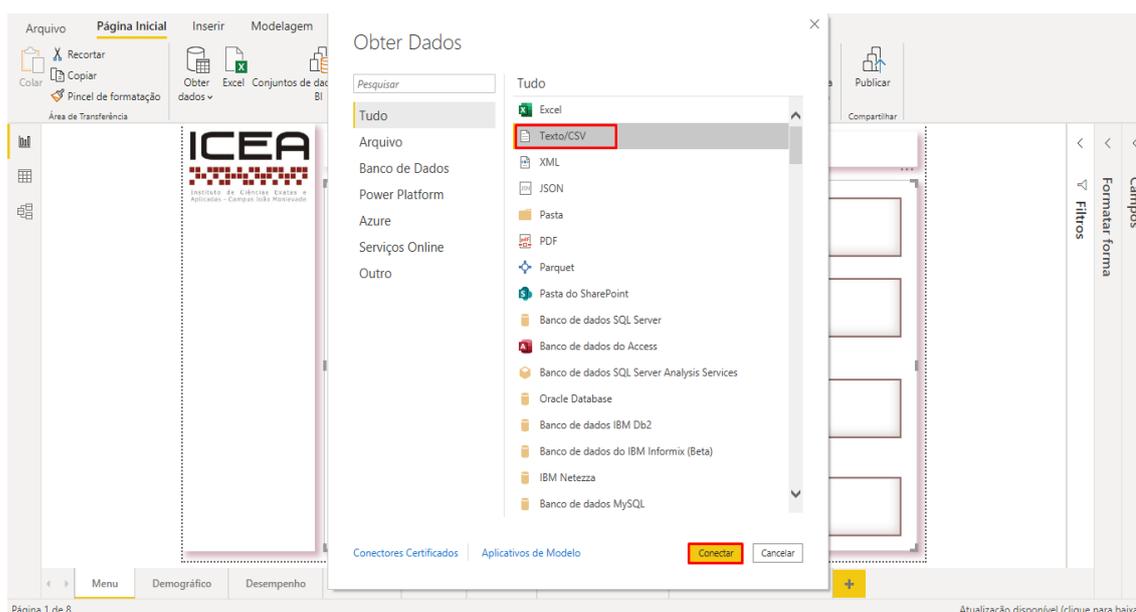
Figura 21 – Como inserir novos dados: Passo 5



Fonte: Elaborado pela autora.

- f) Selecione "texto/csv" e clique em conectar;

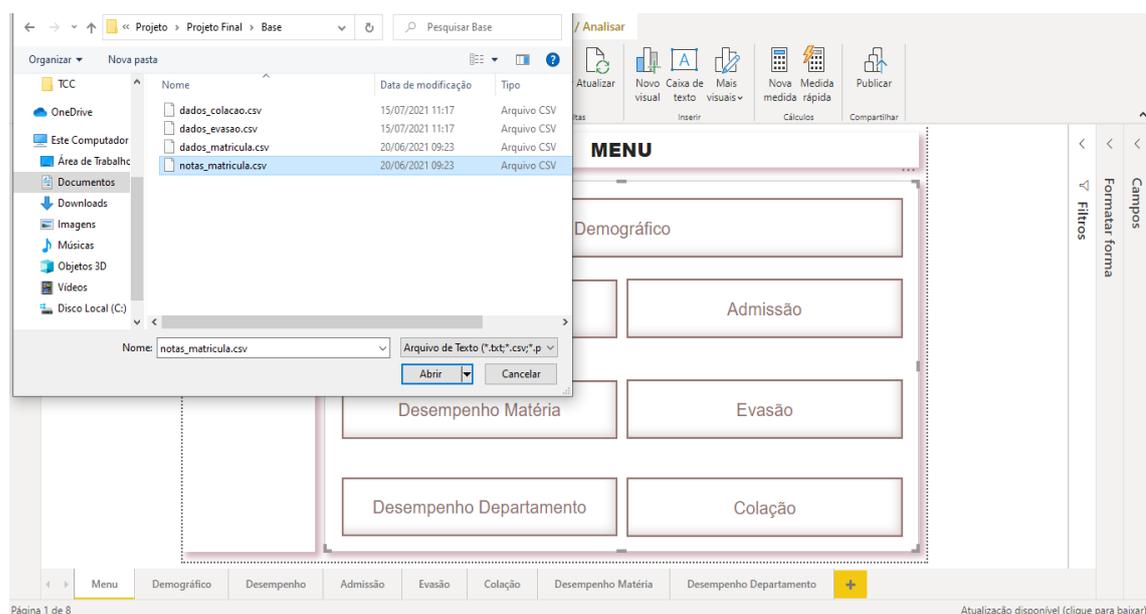
Figura 22 – Como inserir novos dados: Passo 6



Fonte: Elaborado pela autora.

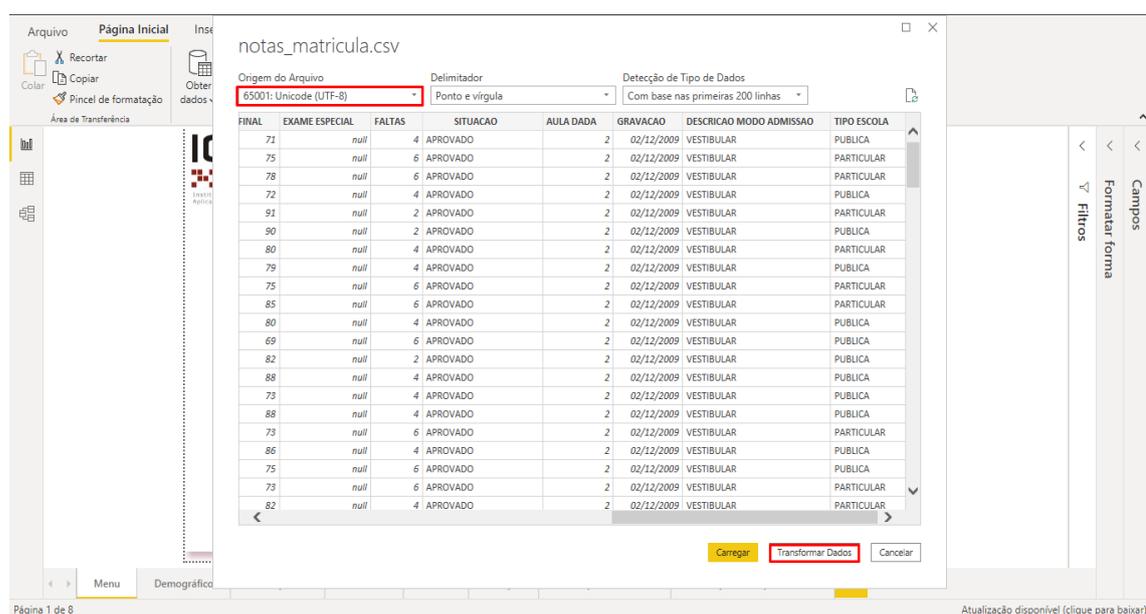
- g) Selecione o arquivo;
- h) Escolha a origem do arquivo como "UTF-8", e clique em "Transformar dados";

Figura 23 – Como inserir novos dados: Passo 7



Fonte: Elaborado pela autora.

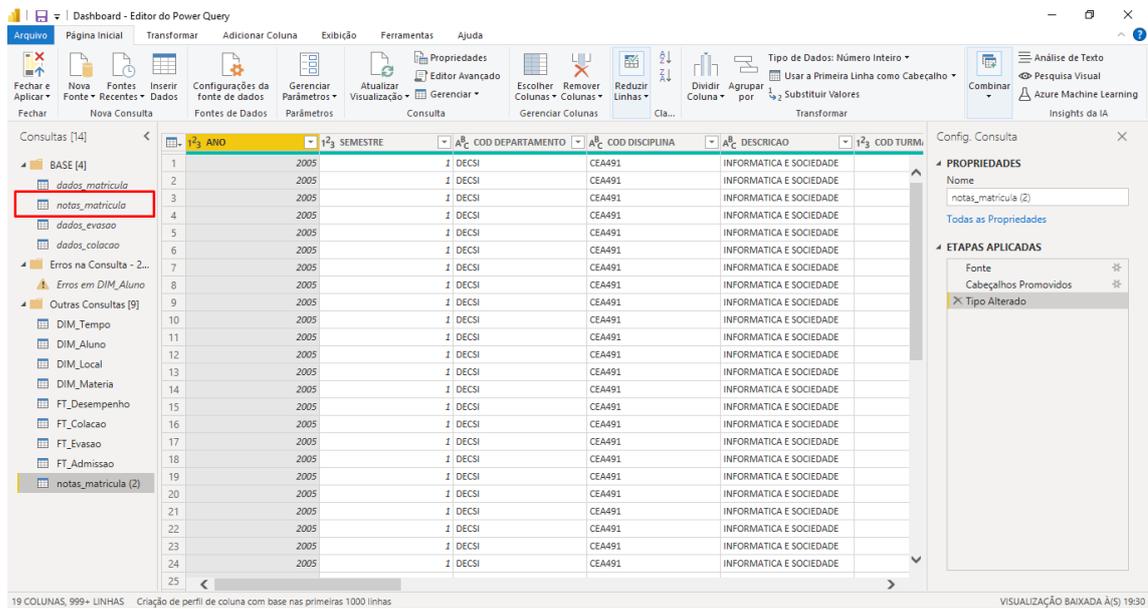
Figura 24 – Como inserir novos dados: Passo 8



Fonte: Elaborado pela autora.

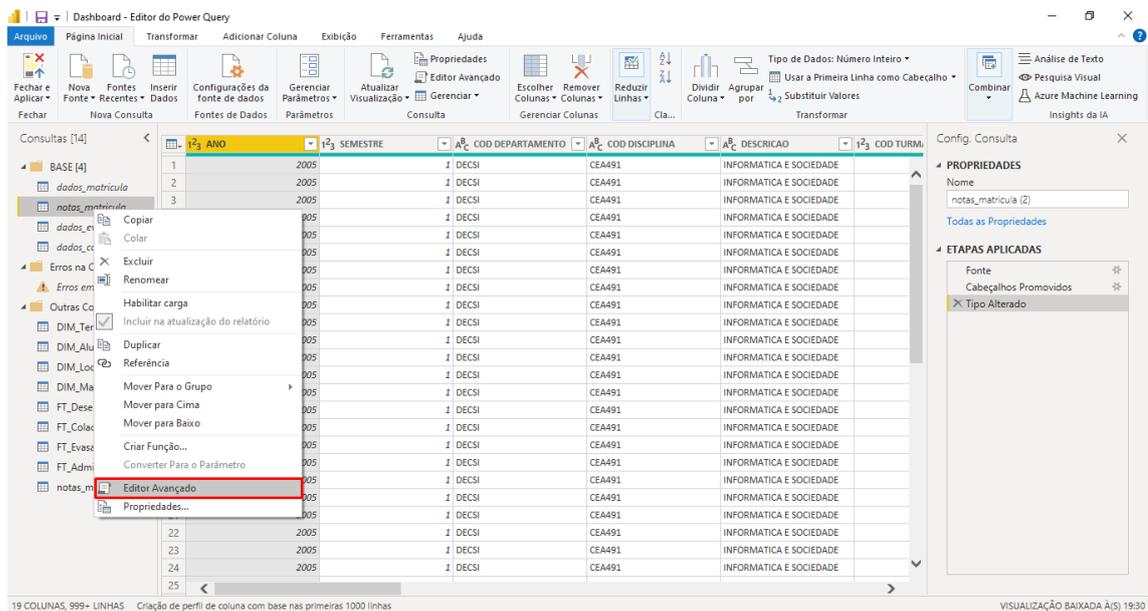
- i) Clique com o botão direito do mouse e selecione o arquivo base do programa;
- j) Clique em editor avançado;
- k) Copie o texto a partir da segunda linha ;
- l) Cole o texto em “Editor avançado” do arquivo que subiu a partir da segunda linha e clique em "Concluído”;
- m) Mova o seu arquivo para arquivo para a pasta “Base”, exclua o arquivo base e altere o nome do novo arquivo, com o nome do arquivo excluído;

Figura 25 – Como inserir novos dados: Passo 9



Fonte: Elaborado pela autora.

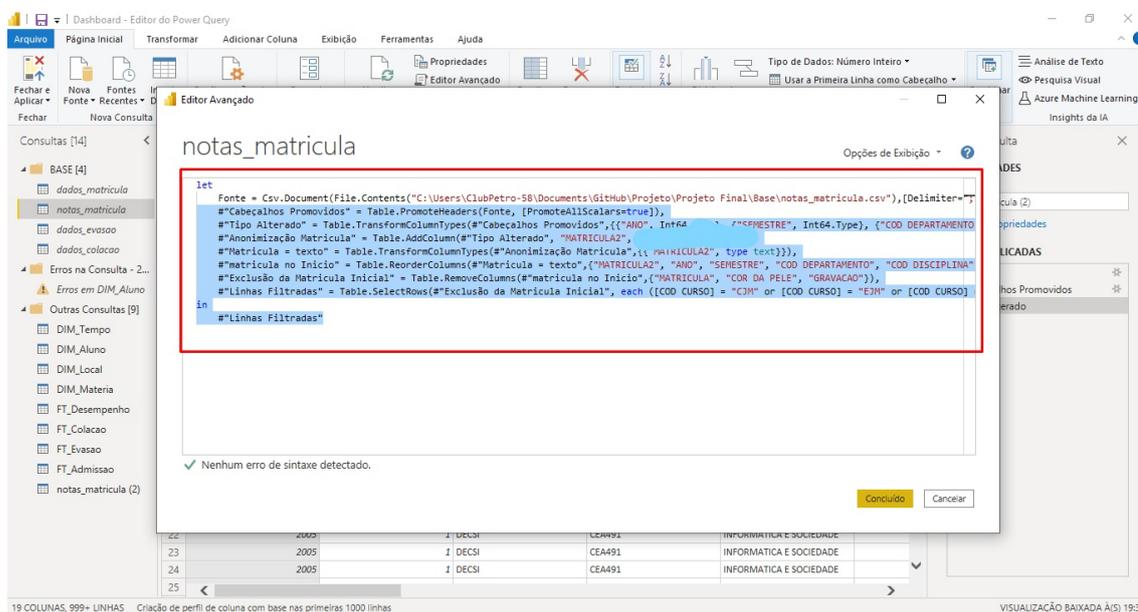
Figura 26 – Como inserir novos dados: Passo 10



Fonte: Elaborado pela autora.

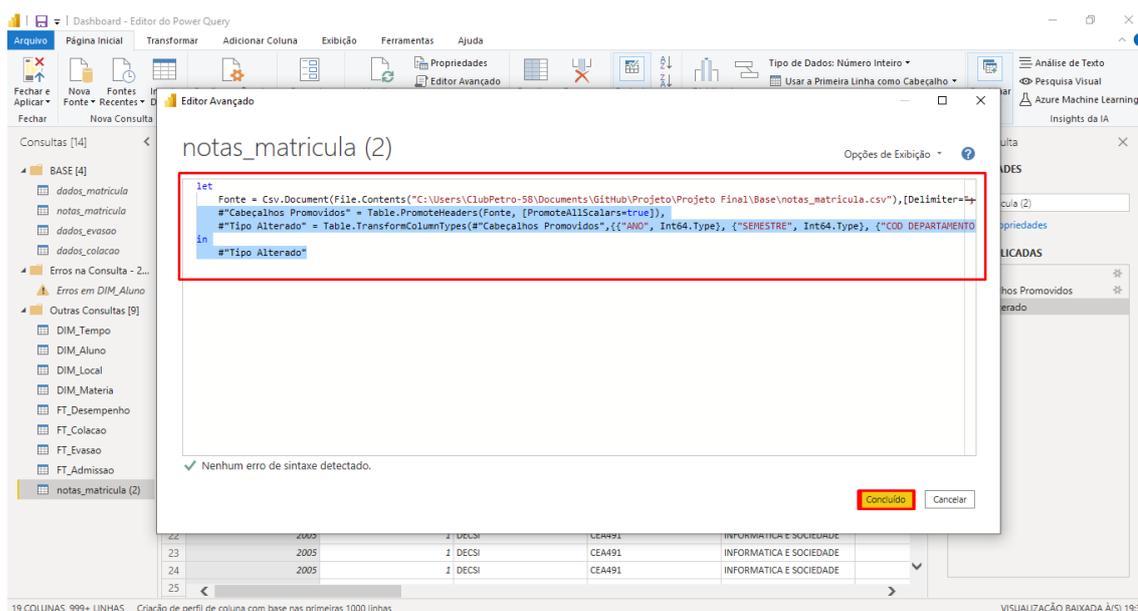
- n) Clique em “Habilitar Carga” para desabilitar a carga;
- o) Repita o processo com os outros 3 arquivos. Para manter os dados atualizados é necessário realizar as substituições das 4 bases selecionadas.

Figura 27 – Como inserir novos dados: Passo 11



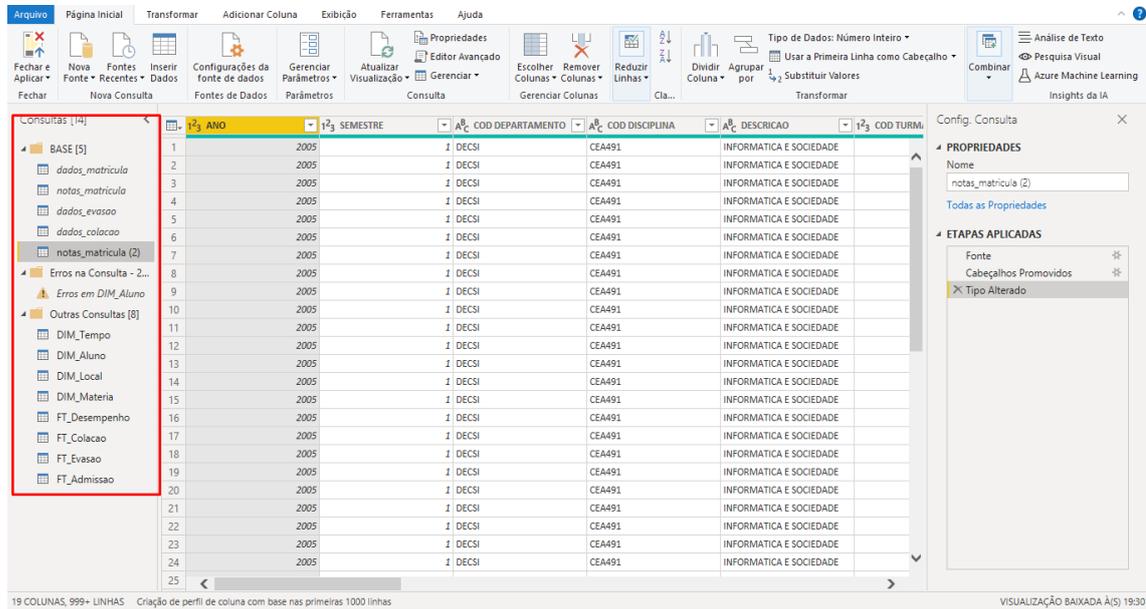
Fonte: Elaborado pela autora.

Figura 28 – Como inserir novos dados: Passo 12



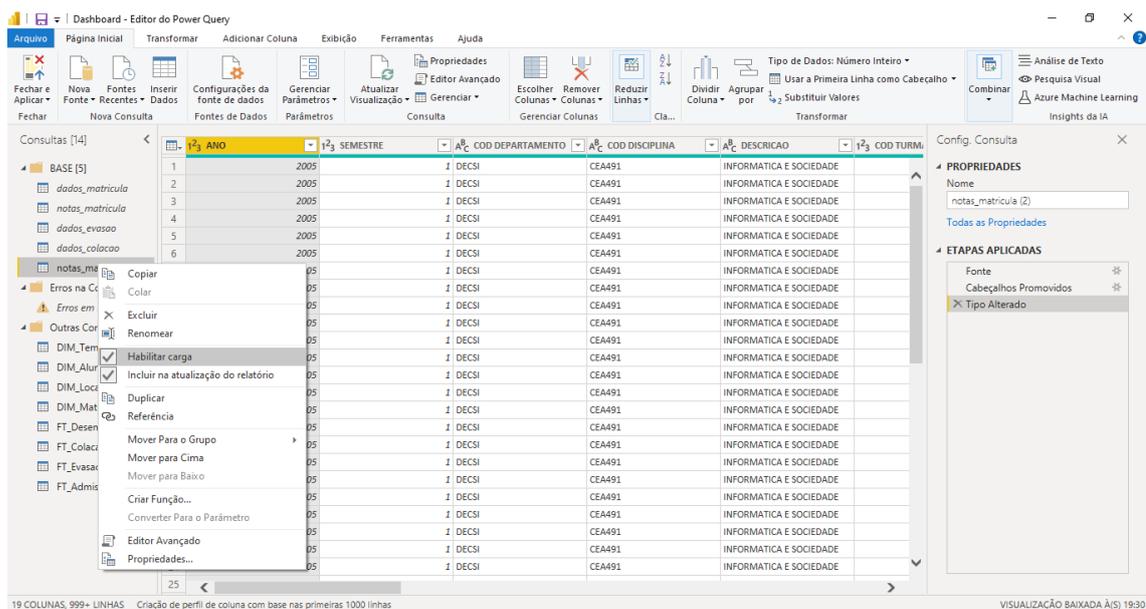
Fonte: Elaborado pela autora.

Figura 29 – Como inserir novos dados: Passo 13



Fonte: Elaborado pela autora.

Figura 30 – Como inserir novos dados: Passo 14

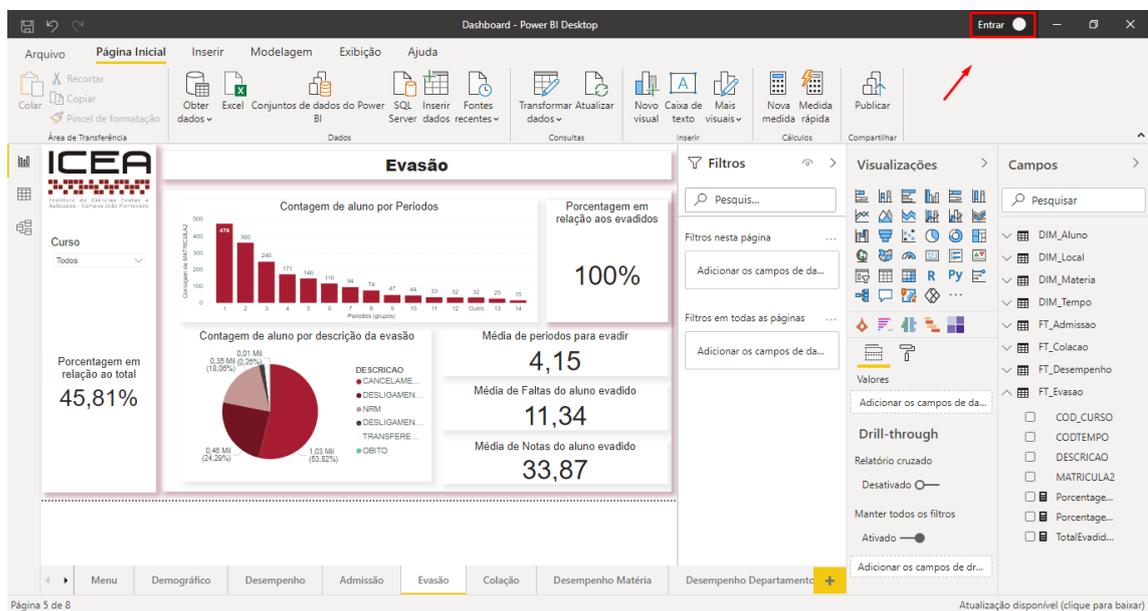


Fonte: Elaborado pela autora.

ANEXO B – Como compartilhar o relatório

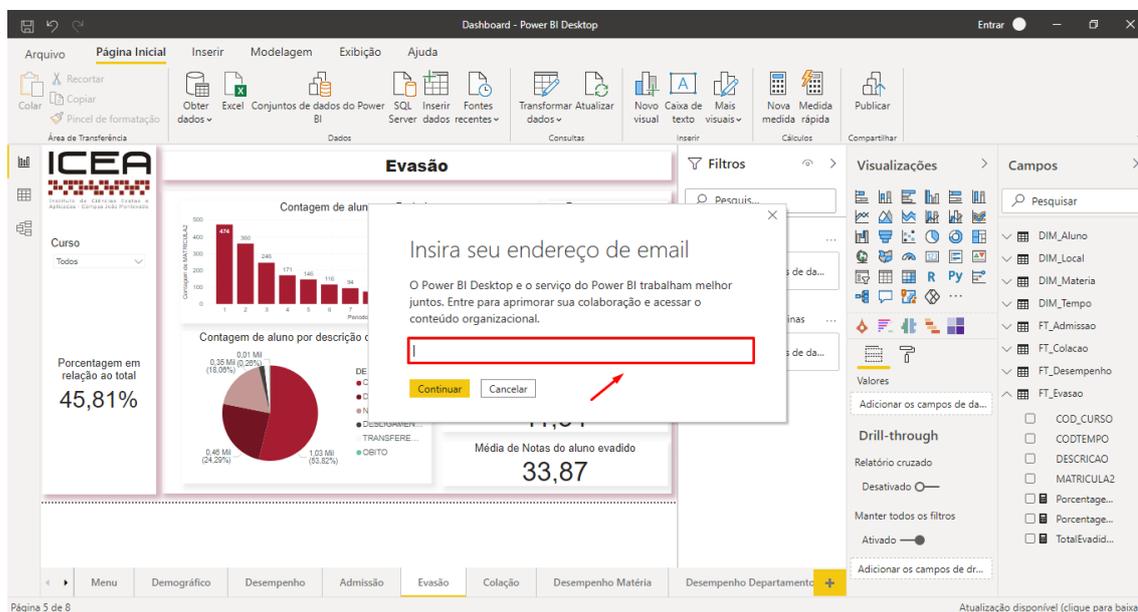
Peça para os decisores as informações de login e senha do *PowerBi online*, clique em “Entrar” no menu superior na lateral direita, clique em “Publicar”, insira essas informações, escolha “Meu *workspace*”, aperte “Selecionar”, e pronto, o arquivo estará disponível para que quem solicitou acesso consiga ver apenas na sua conta pessoal. Segue as imagens do passo a passo.

Figura 31 – Como compartilhar o relatório: Passo 1



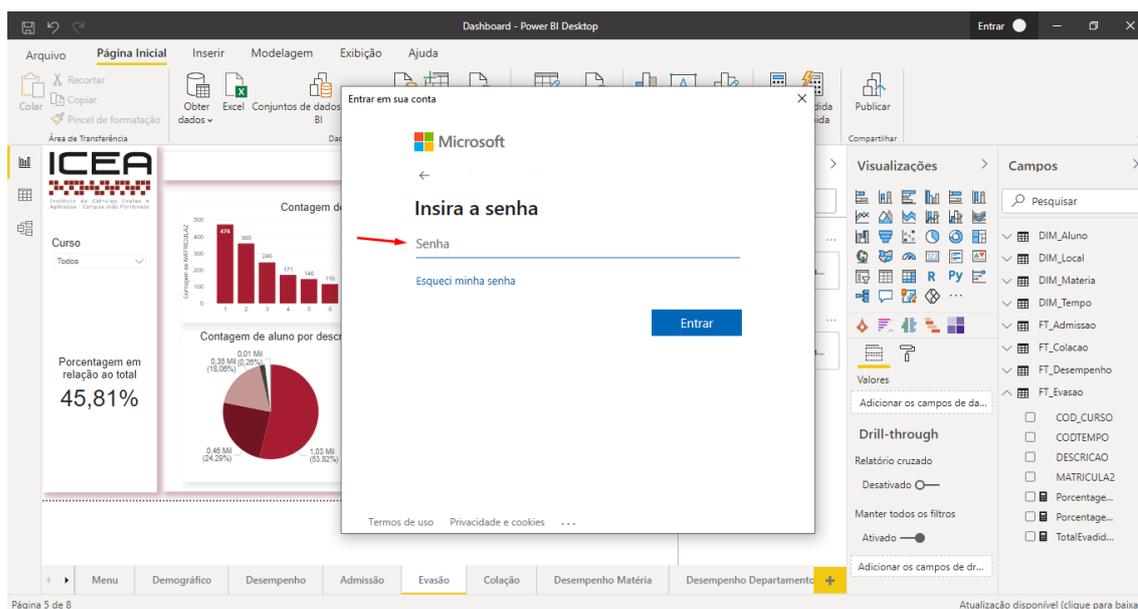
Fonte: Elaborado pela autora.

Figura 32 – Como compartilhar o relatório: Passo 2



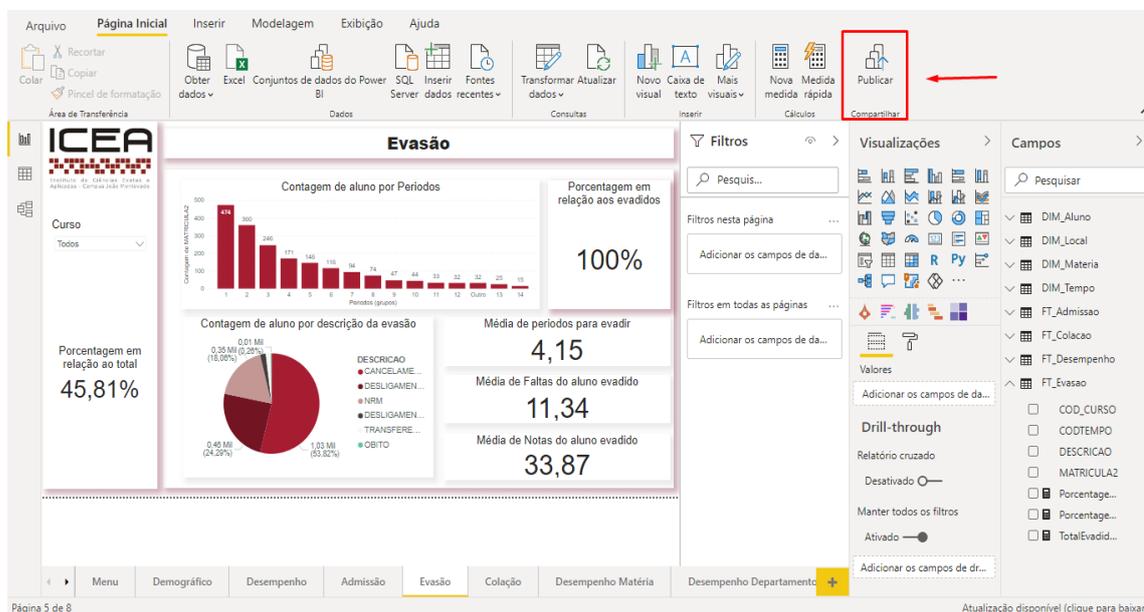
Fonte: Elaborado pela autora.

Figura 33 – Como compartilhar o relatório: Passo 3



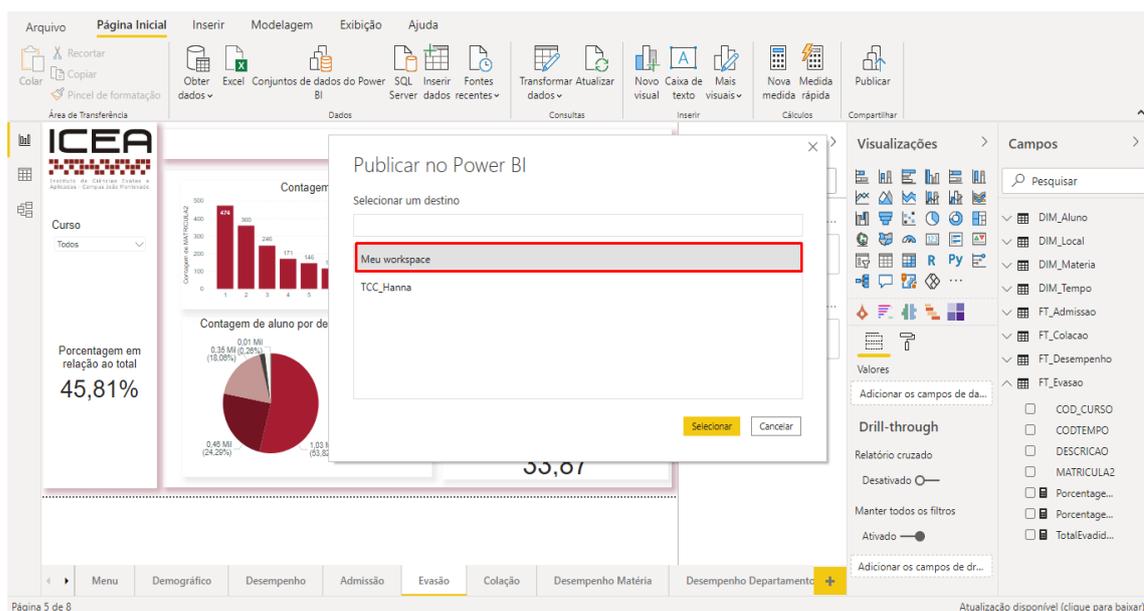
Fonte: Elaborado pela autora.

Figura 34 – Como compartilhar o relatório: Passo 4



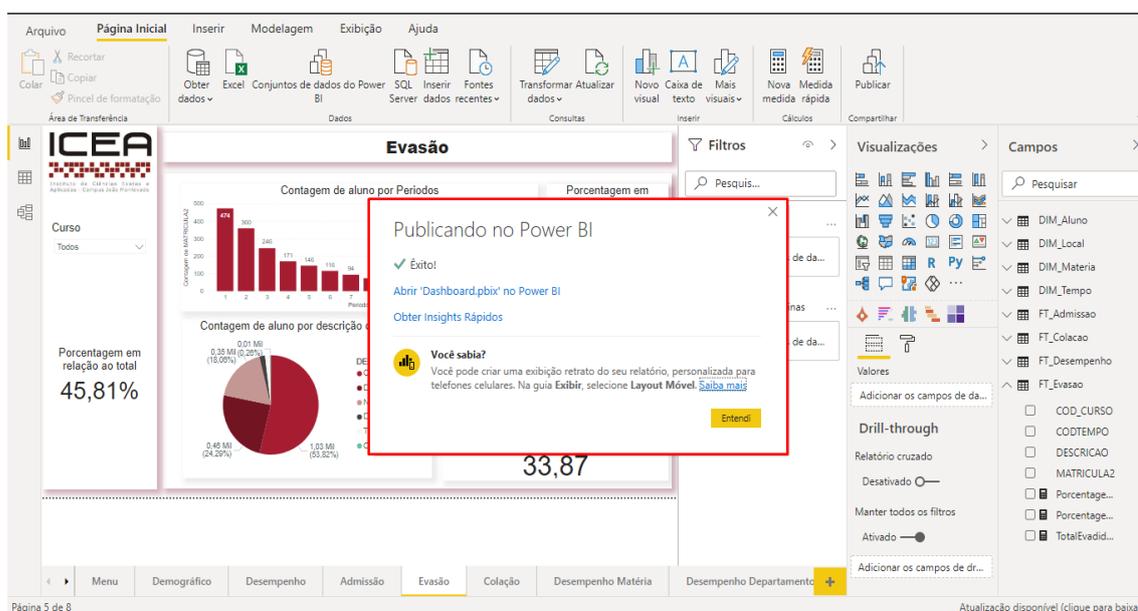
Fonte: Elaborado pela autora.

Figura 35 – Como compartilhar o relatório: Passo 5



Fonte: Elaborado pela autora.

Figura 36 – Como compartilhar o relatório: Passo 6



Fonte: Elaborado pela autora.