

COMPONENTE PRÁTICA

Licenciatura Engenharia Informática

Sistemas de Informação II

David José Nobre Pires | 2019129618

Rui Miguel Sousa Simões | 2018018607

Telmo Eduardo Fonseca Silva | 2019127444

ÍNDICE

Processo de negócio a modular.....	3
Granularidade	3
Dimensões e atributos a incluir no modelo dimensional	4
Factos a ser incluídos no modelo dimensional	5
Modelo dimensional, dados e relação entre tabelas	7
Cálculos e tamanho do modelo concebido	9
Modelo no SQL Server.....	13
Vistas no Visual Studio Data Warehouse.....	14
Dashboards no PowerBI.....	18

Processo de negócio a modular

O processo de negócio a modular para o armazenamento dos dados vai estar associado ao desempenho/sucesso académico dos alunos pois é aquele quando construído vai ter o maior impacto. O nosso objetivo, é que o modelo esteja preparado para responder a todas as perguntas mais importantes e urgentes, como também que os dados sejam imediatamente acessíveis.

Os dados intervenientes nesta situação estão ligados aos alunos, cursos e disciplinas de uma universidade, podemos resumidamente falar sobre eles:

- Em relação aos alunos, vão estar disponíveis dados demográficos e socioeconómicos, incluindo a sua situação financeira;
- No caso dos cursos, pretendemos disponibilizar o tipo de cursos diferentes presentes na instituição, como também a taxa de sucesso de saída, como também a adesão a certo tipo de curso;
- Nas disciplinas, o seu nome e em que semestre de um ano letivo se encontram;

Granularidade

O nível de granularidade com detalhe mais fino disponibilizado pelo modelo será um registo por aluno numa disciplina de um curso durante um período (semestre ou ano letivo).

Através deste registo conseguimos ter acesso ao desempenho académico de cada aluno respetivamente a cada disciplina, podendo depois de uma maneira geral verificar o seu sucesso escolar, ou até mesmo se o aluno tem mais competências numa ou outra área.

Dimensões e atributos a incluir no modelo dimensional

As dimensões/tabelas que vamos inserir no nosso modelo dimensional serão as seguintes dimensões bem como os respetivos atributos:

- Dimensão Aluno, com os atributos:
 - ID do aluno (idaluno, int);
 - Nome (nome, varchar);
 - Idade (idade, int);
 - Sexo (sexo, char);
 - Estado Civil (estadocivil, varchar);
 - Cidade residência (cidaderesidencia, varchar);
 - Renda familiar (rendafamiliar, float);
 - Benefícios financeiros (beneficiosfinanc, float);
 - Status (status, varchar);

- Dimensão Matrículas, com os atributos:
 - ID da matrícula (idmatricula, int);
 - Estado da matrícula (statusmatricula, varchar);
 - Data da matrícula (datamatricula, varchar);
 - Regime da matrícula (tipomatricula, varchar);
 - Período de validade (periodovalidade, int);
 - Foreign key relacionada ao aluno (aluno_idaluno, int NOT NULL);

- Dimensão Curso, com os atributos:
 - ID do curso (idcurso, int);
 - Nome do curso (nomecurso, varchar);
 - Area do curso (areacurso, varchar);
 - Duração do curso (duracaocurso, int);
 - Taxa de sucesso do curso (taxasucesso, float);
 - Taxa de abandono do curso (taxaabandono, float);
 - Popularidade (popularidade, int);

- Coordenador do curso (coordenadorcurso, varchar);
 - Custo do curso (custocurso, int);
 - Foreign key relacionada com a dimensão disciplinas (disciplinas_iddisciplina, int NOT NULL);
- Dimensão Disciplina, com os atributos:
- ID da disciplina (iddisciplina, int);
 - Nome da disciplina (nomediscip, varchar);
 - Número de créditos da disciplina (ects, int);
 - Carga horaria da disciplina (cargahoraria, int);
 - Professor regente da disciplina (profregente, varchar);
 - Tipo da disciplina (tipodiscip, varchar);
 - Dificuldade da disciplina (dificuldade, int);
 - Número de alunos (nralunos, int);
- Dimensão Tempo, com os atributos:
- IDtempo (idtempo, int);
 - Semestre (semestre, int);
 - Ano letivo (anoletivo, int);
 - Dia (dia, int);
 - Mês (mes, int);
 - Ano (ano, int);

Factos a ser incluídos no modelo dimensional

Passando às tabelas de factos, optámos pela abordagem com 2 tabelas. A tabela **DesempenhoAcademico** que permite-nos verificar o desempenho académico dos alunos, através de alguns atributos que temos incluído na tabela como podemos verificar de seguida:

- ID Desempenho Académico (iddesempacad, int);
- Nota final do aluno na disciplina(notafinal, int);

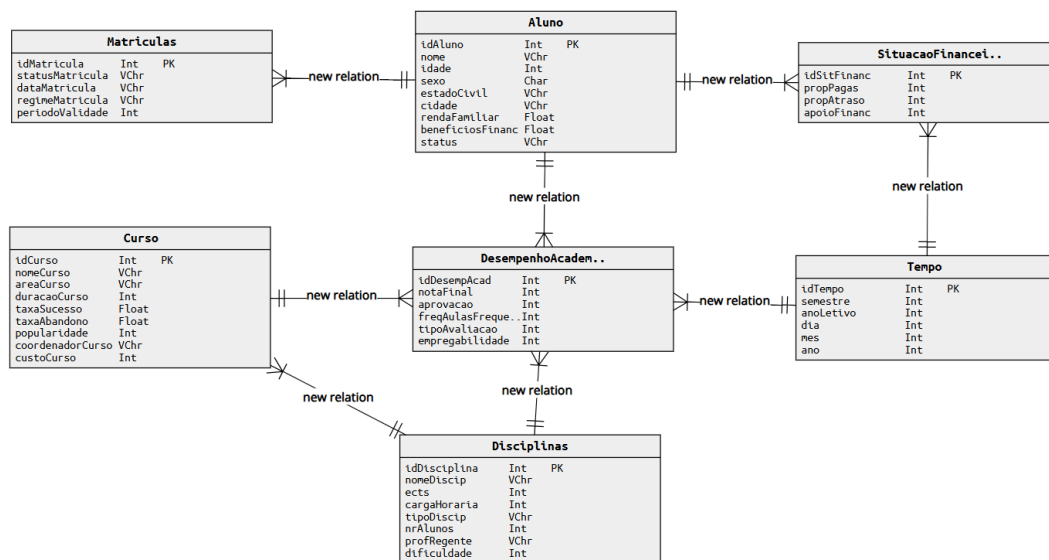
- Aluno aprovado ou não (aprovacao, int);
- Nº aulas frequentadas pelo aluno (freqaulasfrequentadas, int);
- Tipo de avaliação (tipoavaliacao, int);
- Empregabilidade (empregabilidade, int);
- Foreign key relacionada ao tempo (tempo_idtempo, int NOT NULL);
- Foreign key relacionada à disciplina (disciplinas_iddisciplina, int NOT NULL);
- Foreign key relacionada ao curso (curso_idcurso, int NOT NULL);
- Foreign key relacionada ao aluno (aluno_idaluno, int NOT NULL);

A outra tabela de factos é denominada **SituacaoFinanceira**, que permite uma análise da vertente financeira dos alunos e é composta pelos seguintes atributos:

- IDSituacaofinanceira (idsitfinanc, int);
- Propinas pagas (proppagas, int);
- Propinas em atraso (propatraso, int);
- Apoio Financeiro (apoiofinanc, int);
- Foreign key relacionada ao tempo (tempo_idtempo, int NOT NULL);
- Foreign Key relacionada ao aluno (aluno_idaluno, int NOT NULL);

Modelo dimensional, dados e relação entre tabelas

O modelo dimensional obtido é o seguinte modelo em constelação, pois exibe duas tabelas de factos tornando-se mais complexo e distinto de uma modelo em estrela.



Algumas relações entre as diversas tabelas, podemos atentar agora nas ligações entre a tabela de fatos **Desempenho académico** e as outras dimensões:

1. **Desempenho académico** e dimensão **Disciplinas**: cada disciplina tem uma nota diferente correspondente a cada aluno, logo vai estar associado ao sucesso académico de cada um;
2. **Desempenho académico** e dimensão **Cursos**: cada curso diferente na instituição tem diversos alunos, que por sua vez vai ter influência no desempenho do aluno;

3. **Desempenho acadêmico e dimensão Aluno**: neste caso temos a relação mais fácil e simples de entender, pois cada aluno tem um desempenho acadêmico diferente tendo em conta a nota a cada disciplina presente em cada curso em que o mesmo se encontra;
4. **Desempenho acadêmico e dimensão Tempo**: o desempenho acadêmico de cada aluno está ligado a nota que o mesmo obtém em cada disciplina, que por sua vez se encontram em períodos/semestres diferentes, por isso é importante perceber em que espaço temporal a nota foi obtida;
5. **Dimensão Aluno e dimensão Situação financeira**: é importante perceber em que situação financeira cada aluno se encontra, em termos de propinas pagas ou em atraso, como também se foi alvo de algum apoio financeiro;
6. **Dimensão Tempo e dimensão Situação Financeira**: ligada à relação anterior, é também importante saber em que tempo/período o aluno foi alvo de um apoio financeiro, ou em que tempo o aluno teve propinas por pagar (ou não);
7. **Dimensão Aluno e dimensão Matrículas**: uma ligação fulcral para se poder interligar cada matrícula a cada aluno, como também o estado em que a mesma se encontra;
8. **Dimensão Curso e dimensão Disciplinas**: cada curso tem diversas disciplinas. Por sua vez, a tabela curso tem uma Foreign key associada às disciplinas.

Facto DesempenhoAcademico	Armazena medições do desempenho académico, vinculando alunos, disciplinas, períodos e cursos.
Facto SituacaoFinanceira	Armazena medições relativo às condições económicas dos alunos bem como a saúde financeira dos mesmos.
Dimensão Aluno	Detalha as informações pessoais e socioeconómicas dos alunos.
Dimensão Disciplina	Contém informações descritivas sobre cada disciplina.
Dimensão Matrículas	Dispõe informação sobre o tipo da matrícula do aluno.
Dimensão Curso	Oferece informações gerais sobre os cursos disponíveis.
Dimensão Tempo	Representa períodos académicos e suas divisões (anos, semestres).

Cálculos e tamanho do modelo concebido

Dimensão Tempo: 365 dias x 12 meses x 5 anos = 21900 registos para a tabela tempo (usámos apenas 3000 registos);

Dimensão Aluno: 10000 registos, pois as nossas tabelas de facto que correspondem ao desempenho académico e à situação financeira de cada aluno também tem 10000. Para ficar uma para cada um;

Dimensão Matrículas: 10000 registos, pois temos 10000 alunos;

Dimensão Curso: 30 registos, pois como está no enunciado a universidade possui 30 cursos diferentes;

Dimensão Disciplinas: 30 cursos x 30 disciplinas diferentes = 900 registos;

Tabelas de factos DesempenhoAcadémico e SituacaoFinanceira: 10000 registos cada uma, como pedido pelo professor;

Dimensão Disciplinas:

- Iddisciplina(int) = 4 bytes;
- Nomediscip(varchar) = 12 bytes;
- Ects(int) = 4 bytes;
- Cargahoraria(int) = 4 bytes;
- Tipodiscip(varchar) = 12 bytes;
- Nralunos(int) = 4 bytes;
- Profregente(varchar) = 12 bytes;
- Dificuldade(int) = 4 bytes;

Ou seja, $4 + 12 + 4 + 4 + 12 + 4 + 12 + 4 = 56$ bytes, o que dá na realidade $56 \times 900 \text{ registos} = 50400 = 50.4 \text{ KB}$.

Dimensão Curso:

```
CREATE TABLE curso (  
    idcurso          int,  
    nomecurso        varchar(250),  
    areacurso        varchar(250),  
    duracaocurso     int,  
    taxasucesso      float,  
    taxaabandono     float,  
    popularidade     int,  
    coordenadorcurso varchar(250),  
    custocurso       int,  
    disciplinas_iddisciplina int NOT NULL,  
    PRIMARY KEY(idcurso)  
);
```

➤ $4 + 12 + 12 + 4 + 8 + 8 + 4 + 12 + 4 + 4 = 72$ bytes
➤ $72 \text{ bytes} \times 30 \text{ registos} = 2160 = 2.16 \text{ KB}$

Dimensão Aluno:

```
CREATE TABLE aluno (  
    idaluno          int,  
    nome             varchar(250),  
    idade            int,  
    sexo             char(255),  
    estadocivil      varchar(250),  
    cidade           varchar(250),  
    rendafamiliar    float,  
    beneficiosfinanc float,  
    status           varchar(250),  
    PRIMARY KEY(idaluno)  
);
```

➤ $4 + 12 + 4 + 1 + 12 + 12 + 8 + 8 + 12 = 73$ bytes
➤ $73 \times 10000 = 730000 = 730 \text{ KB}$

Dimensão Tempo:

```
CREATE TABLE tempo (  
    idtempo int,  
    semestre int,  
    anoletivo int,  
    dia int,  
    mes int,  
    ano int,  
    PRIMARY KEY(idtempo)  
);
```

➤ $4 + 4 + 4 + 4 + 4 + 4 = 24$ bytes
➤ $24 \times 3000 = 72000 = 72$ KB

Dimensão Matrículas:

```
CREATE TABLE matriculas (  
    idmatricula int,  
    statusmatricula varchar(250),  
    datamatricula varchar(250),  
    regimematricula varchar(250),  
    periodovalidade int,  
    aluno_idaluno int NOT NULL,  
    PRIMARY KEY(idmatricula)  
);
```

➤ $4 + 12 + 12 + 12 + 4 + 4 = 48$ bytes
➤ $48 \times 10000 = 480000 = 480$ KB

Tabela de factos DesempenhoAcademico:

```
CREATE TABLE desempenhoacademico (  
    iddesempacad int,  
    notafinal int,  
    aprovacao int,  
    freqaulasfrequentadas int,  
    tipoavaliacao int,  
    empregabilidade int,  
    tempo_idtempo int NOT NULL,  
    disciplinas_iddisciplina int NOT NULL,  
    curso_idcurso int NOT NULL,  
    aluno_idaluno int NOT NULL,  
    PRIMARY KEY(iddesempacad)  
);
```

➤ $4 \times 10 = 40$ bytes
➤ $40 \times 10000 = 400000 = 400$ KB

Tabela de factos SituacaoFinanceira:

```
CREATE TABLE situacaofinanceira (  
    idsitfinanc      int,  
    proppagas        int,  
    propatraso       int,  
    apoiofinanc      int,  
    tempo_idtempo int NOT NULL,  
    aluno_idaluno int NOT NULL,  
    PRIMARY KEY(idsitfinanc)  
);
```

➤ 4 x 6 = 24 bytes
➤ 24 x 10000 = 240000 = 240 KB

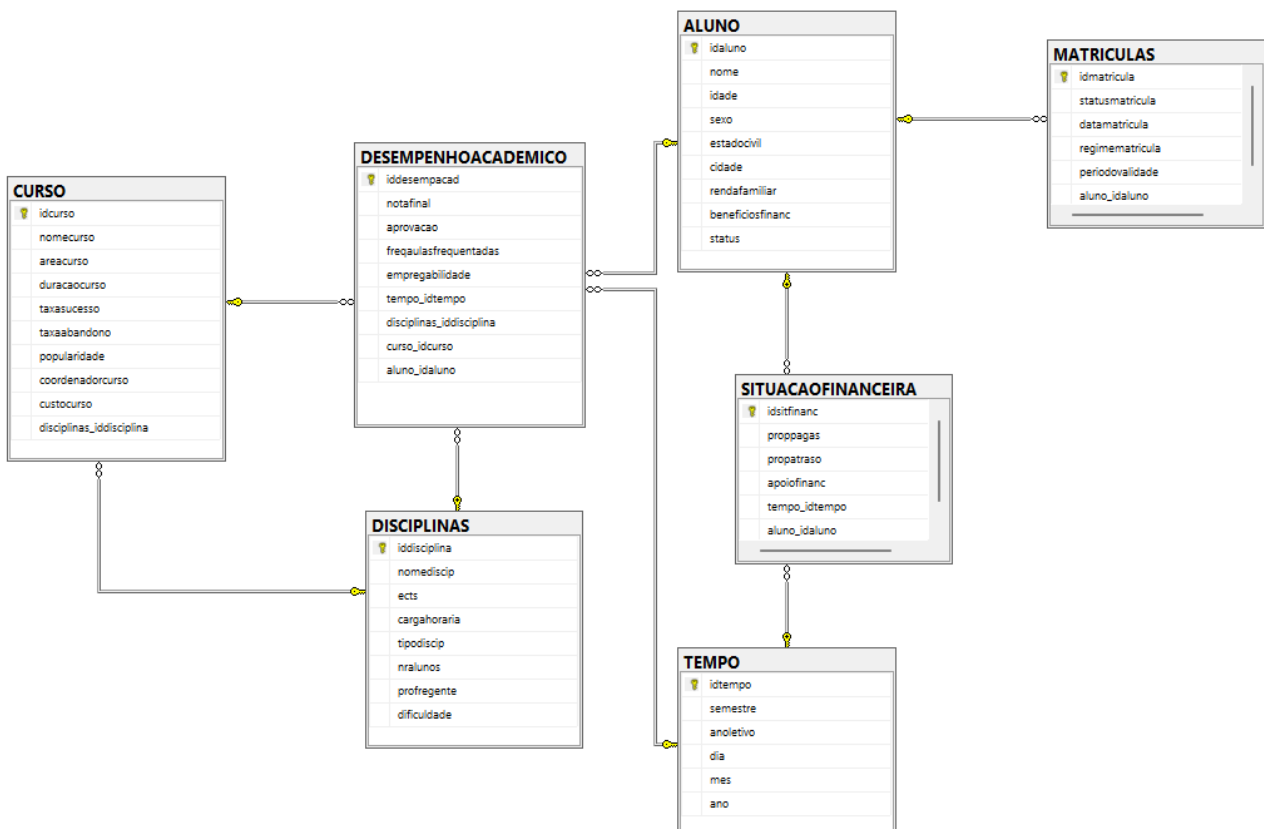
Total de custos:

$50.4 + 2.16 + 730 + 72 + 480 + 400 + 240 = 1974,56 \text{ KB} = 1,97456 \text{ MB}$

$1,97456 - 0,64(\text{valor ocupado pelas tabelas de facto em MB}) = 1,33456$

O valor ocupado pelas dimensões vai ser 1,33456 MB.

Modelo no SQL Server

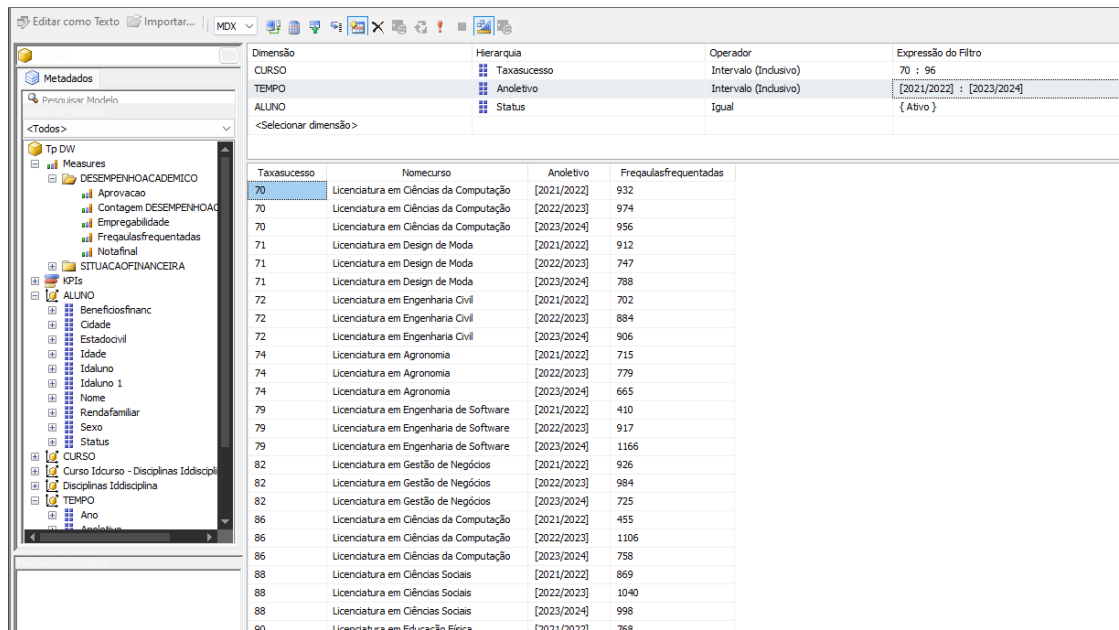


Taxa de sucesso nos primeiros 3 anos (2019 a 2021) de todos os alunos com idade entre os 18 e 26:

Popularidade acima dos 70% para os cursos comparando com a sua empregabilidade:

14

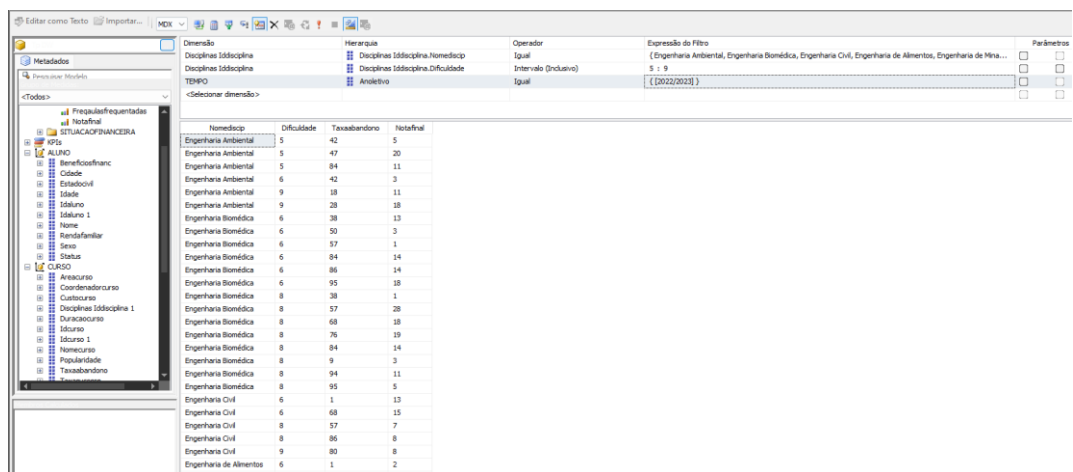
Nome do curso com taxa de sucesso acima de 70%, nos anos letivos entre 2021/2022 a 2023/2024, relacionando com a frequência de aulas frequentadas pelos alunos com status ativo:



Dimensão	Hierarquia	Operador	Expressão do Filtro
CURSO	Taxasuccesso	Intervalo (Inclusivo)	70 : 96
TEMPO	Anoletivo	Intervalo (Inclusivo)	[2021/2022] : [2023/2024]
ALUNO	Status	Igual	{ Ativo }
<Selecionar dimensão>			

Taxasuccesso	Nomecurso	Anoletivo	Frequenciasfrequentadas
70	Licenciatura em Ciências da Computação	[2021/2022]	932
70	Licenciatura em Ciências da Computação	[2022/2023]	974
70	Licenciatura em Ciências da Computação	[2023/2024]	956
71	Licenciatura em Design de Moda	[2021/2022]	912
71	Licenciatura em Design de Moda	[2022/2023]	747
71	Licenciatura em Design de Moda	[2023/2024]	788
72	Licenciatura em Engenharia Civil	[2021/2022]	702
72	Licenciatura em Engenharia Civil	[2022/2023]	884
72	Licenciatura em Engenharia Civil	[2023/2024]	906
74	Licenciatura em Agronomia	[2021/2022]	715
74	Licenciatura em Agronomia	[2022/2023]	779
74	Licenciatura em Agronomia	[2023/2024]	665
79	Licenciatura em Engenharia de Software	[2021/2022]	410
79	Licenciatura em Engenharia de Software	[2022/2023]	917
79	Licenciatura em Engenharia de Software	[2023/2024]	1166
82	Licenciatura em Gestão de Negócios	[2021/2022]	926
82	Licenciatura em Gestão de Negócios	[2022/2023]	984
82	Licenciatura em Gestão de Negócios	[2023/2024]	725
86	Licenciatura em Ciências da Computação	[2021/2022]	455
86	Licenciatura em Ciências da Computação	[2022/2023]	1106
86	Licenciatura em Ciências da Computação	[2023/2024]	758
88	Licenciatura em Ciências Sociais	[2021/2022]	869
88	Licenciatura em Ciências Sociais	[2022/2023]	1040
88	Licenciatura em Ciências Sociais	[2023/2024]	998
91	Licenciatura em Educação Física	[2021/2022]	768

Nota final das disciplinas de Engenharia com dificuldade entre 5 a 9, com a taxa de abandono no ano letivo de 2022/2023:



Dimensão	Hierarquia	Operador	Expressão do Filtro	Parâmetros
Disciplinas Subdisciplina	Disciplinas Subdisciplina Nome Discip	Igual	{ Engenharia Ambiental, Engenharia Biomédica, Engenharia Civil, Engenharia de Alimentos, Engenharia de Mina...	<input type="checkbox"/>
TEMPO	Anoletivo	Intervalo (Inclusivo)	5 : 9	<input type="checkbox"/>
<Selecionar dimensão>				<input type="checkbox"/>

Nome Discip	Dificuldade	Taxa abandono	Nota final
Engenharia Ambiental	5	42	5
Engenharia Ambiental	5	47	20
Engenharia Ambiental	5	84	11
Engenharia Ambiental	6	42	3
Engenharia Ambiental	9	18	11
Engenharia Ambiental	9	28	18
Engenharia Biomédica	6	38	13
Engenharia Biomédica	6	50	3
Engenharia Biomédica	6	17	1
Engenharia Biomédica	6	84	14
Engenharia Biomédica	6	86	14
Engenharia Biomédica	6	95	18
Engenharia Biomédica	8	38	1
Engenharia Biomédica	8	17	28
Engenharia Biomédica	8	68	18
Engenharia Biomédica	8	76	19
Engenharia Biomédica	8	84	14
Engenharia Biomédica	8	9	3
Engenharia Biomédica	8	94	11
Engenharia Biomédica	8	95	5
Engenharia Civil	6	1	13
Engenharia Civil	6	68	15
Engenharia Civil	8	57	7
Engenharia Civil	8	86	8
Engenharia Civil	9	80	8
Engenharia de Alimentos	6	1	2

Renda familiar entre os 1000 e 1500, e verificar se tem propinas em Atraso e/ou usufrui de apoio financeiro, para os alunos que congelaram matrícula ou já desistiram:

Dimensão	Hierarquia	Operador	Expressão do Filtro
ALUNO	Rendafamiliar	Intervalo (Inclusivo)	1000 : 1500
ALUNO	Status	Igual	{ Congelado, Desistente }

Rendafamiliar	Benefícios financ	Status	Propatrasso	Apoio financ
1002	320	Congelado	1	1
1003	145	Desistente	4	1
1010	380	Desistente	0	2
1015	647	Desistente	2	1
1018	612	Congelado	0	1
1024	346	Congelado	1	1
1031	791	Congelado	1	2
1032	1300	Congelado	0	1
1033	231	Congelado	0	0
1035	147	Congelado	1	1
1036	107	Congelado	0	0
1042	633	Desistente	1	0
1049	430	Congelado	0	0
1055	1500	Desistente	1	0
1058	920	Desistente	1	0
1059	291	Congelado	1	2
1064	988	Desistente	1	0
1065	1117	Desistente	1	1
1070	883	Congelado	3	1
1077	283	Desistente	0	0
1079	396	Congelado	0	1
1089	222	Congelado	1	2
1091	1346	Congelado	4	2
1094	633	Desistente	2	0
1095	926	Desistente	0	2
1104	1040	Congelado	2	3

Área do curso com mais popularidade entre os dois sexos comparando a empregabilidade:

Dimensão	Hierarquia	Operador	Expressão do Filtro
CURSO	Popularidade	Intervalo (Inclusivo)	76 :

Popularidade	Sexo	Area curso	Nome curso	Empregabilidade
76	F	marketing	Licenciadu...	7504
76	M	marketing	Licenciadu...	8029
78	F	música	Licenciadu...	9339
78	M	música	Licenciadu...	8283
80	F	direito	Licenciadu...	8874
80	M	direito	Licenciadu...	9108
82	F	marketing	Licenciadu...	9304
82	M	marketing	Licenciadu...	7711

Nome dos alunos que frequentam um certo curso, relacionado com a taxa de sucesso, nota final, a frequência das aulas assistidas e a empregabilidade:
(dados repetidos, uma vez que a gerar os dados com recurso ao “mockaroo” não existe uma coerência nem uma diversidade gigantesca na geração desses mesmos dados, o que provoca esta situação).

Metadados

Pesquisar Modelo

<Todos>

Aprovacao

Contagem DESEMPENHOAC

Empregabilidade

Frequenciasfrequentadas

Nota final

SITUACAOFINANCEIRA

Apoiofinanc

Contagem SITUACAOFINA

Propatrasso

Proppagas

KPIs

ALLUNO

Beneficiosfinanc

Cidade

Estadocivil

Idade

Idaluno

Idaluno 1

Nome

Rendafamiliar

Sexo

Status

CURSO

Areacurso

Coordenadorcurso

Custocurso

Dimensão

Hierarquia

Operador

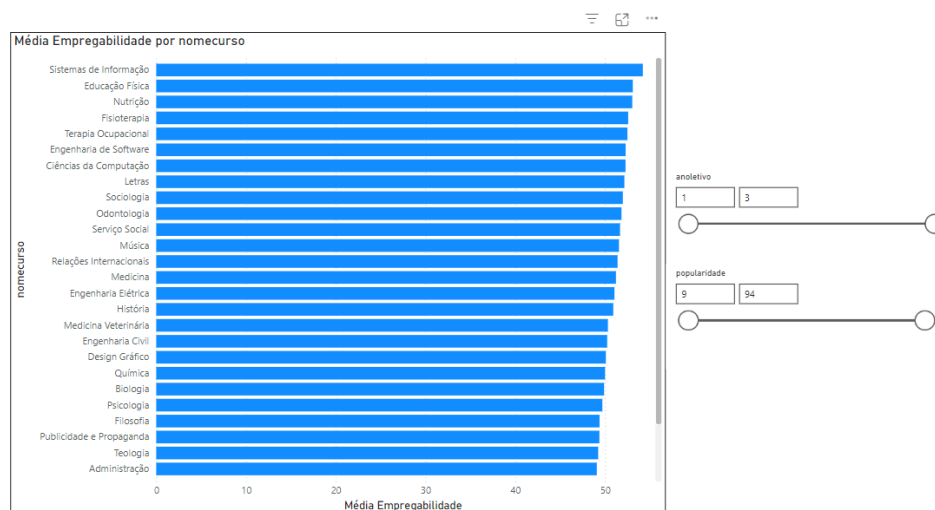
Expressão do Filtro

<Selecionar dimensão>

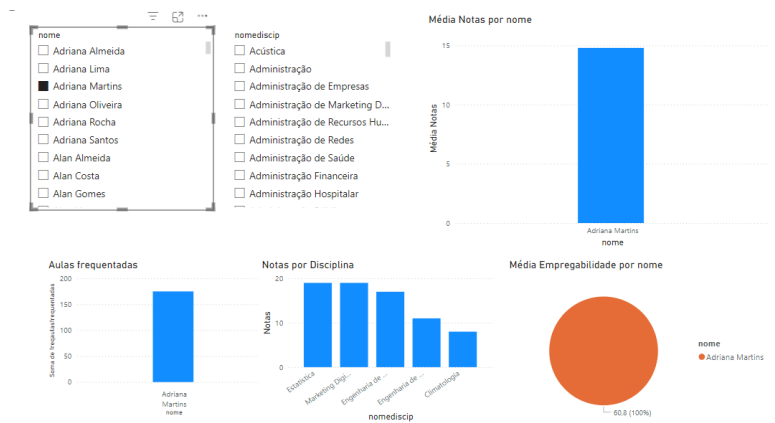
Nome	Nomecurso	Taxasucesso	Nota final	Freqaulasfrequentadas	Empregabilidade
Aarika Elmer	Licenciatura em Educação Física	90	12	17	66
Aarika Elmer	Licenciatura em Engenharia Civil	72	10	99	2
Aarika Elmer	Licenciatura em Marketing	35	19	93	98
Aarika Nordliff	Licenciatura em Engenharia Elétrica	47	10	90	60
Aarika Stachini	Licenciatura em Agronomia	12	15	36	91
Aarika Stachini	Licenciatura em Gestão de Negócios	82	13	42	83
Aaron Lob	Licenciatura em Ciências da Computação	70	13	74	20
Abagael Cornelis	Licenciatura em Agronomia	74	15	39	87
Abigail Cockayne	Licenciatura em Fisioterapia	29	18	42	44
Abba Kerwick	Licenciatura em Administração Pública	92	16	13	52
Abba Kerwick	Licenciatura em Ciências da Computação	70	1	6	44
Abbe Grebner	Licenciatura em Engenharia Elétrica	47	16	22	14
Abbe Grebner	Licenciatura em Gestão de Negócios	82	1	44	84
Abbey Clee	Licenciatura em Gestão de Negócios	7	15	65	15
Abbey Clee	Licenciatura em Relações Internacionais	52	1	99	93
Abbey Mathias	Licenciatura em Fisioterapia	29	6	20	1
Abbey Mogridge	Licenciatura em Relações Internacionais	52	5	85	37
Abbi Hearsey	Licenciatura em Agronomia	12	10	54	59
Abbi Liston	Licenciatura em Design de Moda	71	10	35	16
Abbi Liston	Licenciatura em Engenharia Elétrica	18	16	35	2
Abbi Robatham	Licenciatura em Ciências Sociais	88	6	80	80
Abbie Seaman	Licenciatura em Arquitetura	57	27	3	86
Abbott Newby	Licenciatura em Marketing	35	6	88	43
Abdel Chalfont	Licenciatura em Música	20	10	85	100
Abdel Chattey	Licenciatura em Administração Pública	96	3	90	90
Abdel Chattey	Licenciatura em Engenharia Mecânica	20	7	13	54
Abdel Doss	Licenciatura em Agronomia	74	15	39	87

Dashboards no PowerBI

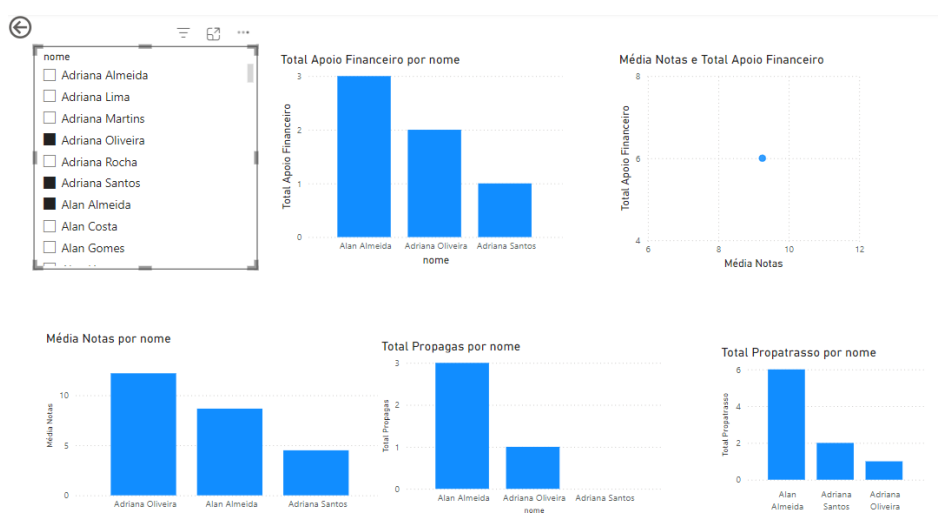
Dashboard interativo que nos permite verificar os cursos com maior empregabilidade presentes nesta instituição. Através de dois sliders distintos, é nos possível filtrar a pesquisa por ano letivo, e popularidade desse mesmo curso.



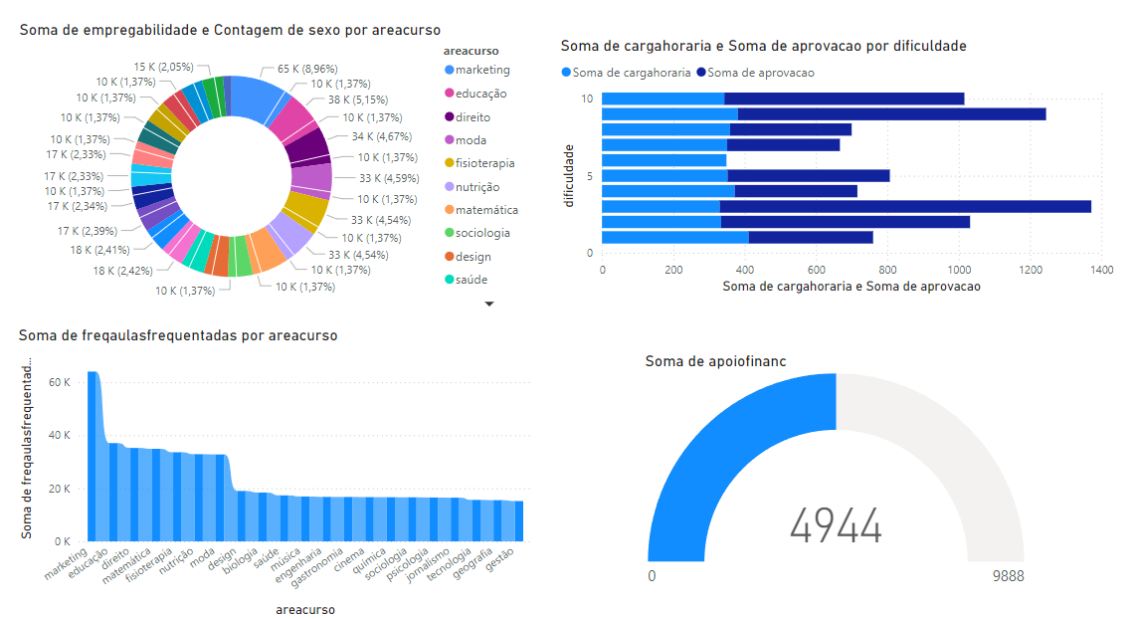
Dashboard interativo, que nos permite verificar o desempenho acadêmico de cada aluno, podendo filtrar a pesquisa pelo nome de cada aluno desta universidade. Primeiro selecionamos o aluno que pretendemos, vemos as disciplinas que o mesmo frequenta, e se necessário ver apenas uma nota em singular de uma cadeira que o aluno tem. Podemos ver a media das notas do aluno, como também o número total de aulas que o mesmo frequentou.



Dashboard interativo, permite-nos relacionar o desempenho académico de cada aluno com a sua situação financeira atual. Podemos verificar o número de propinas pagar ou em atraso, como quantas vezes foi alvo de um apoio financeiro. É também apresentado um gráfico de dispersão que relaciona a sua média com o total de apoios financeiros recebidos.



Dashboard que permite averiguar sobre o diverso tipo de pesquisas relacionadas com o desempenho académico.



Dashboard que recorrendo às cidades dos alunos efetua pesquisas dos mais diversos géneros, como a aprovação por cidades, aprovação por ano letivo, etc.

