

Análise dos dados de demanda dos cursos de Ciência da Computação e Engenharia da Computação

Lucas Brum, Leonardo Bilhalva, Artur Turatti

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

***Resumo.** Este meta-artigo descreve um sistema a ser desenvolvido sobre os dados dos processos seletivos do Vestibular da UFRGS e quantitativos de alunos de graduação a fim de analisar o aumento recente de graduandos nos cursos de Ciência da Computação e Engenharia da Computação.*

1. Introdução

Durante o semestre 2023/1 na UFRGS foi notado um certo aumento no número de alunos da instituição, o que nos intrigou. Sendo assim, resolvemos estudar esse aumento observando dados dos processos seletivos da UFRGS e de quantitativos de alunos na universidade e criar uma aplicação que demonstre essas análises. Pode ser útil para pesquisas demográficas, censos, estudos relacionados à população no geral, comunidade acadêmica, etc.

Pretendemos utilizar PostgreSQL para armazenagem dos dados e PowerBI para análise. A biblioteca *pandas* do Python pode ser útil também para algumas análises.

Na seção 2, faremos a descrição da base de dados, com auxílio de seus respectivos dicionários de dados. Na seção 3, faremos a descrição do ambiente de desenvolvimento (equipamentos, linguagens de programação e o Sistema de Gerenciamento de Banco de Dados usado para armazenagem dos dados). Na seção 4, descrevemos como será feito o trabalho ao longo do tempo até sua entrega. Na seção 5, serão descritas as funcionalidades propostas (consultas e visualizações que viabilizam a análise). Na seção 6, será descrito o cronograma de como será feito e dividido o trabalho e, por fim, as referências utilizadas até aqui.

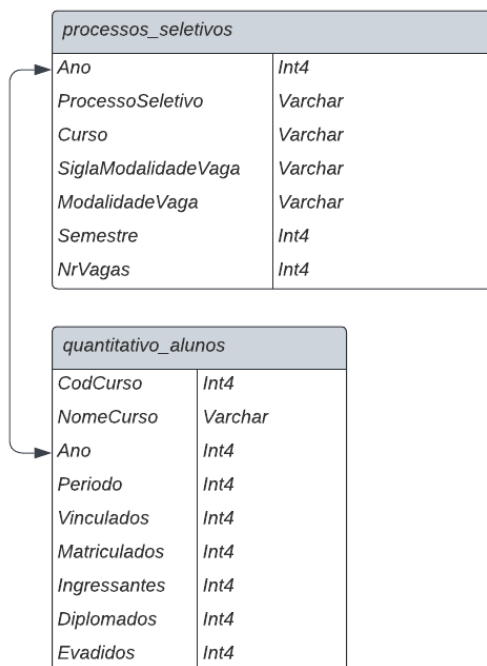
2. Descrição da base de dados

Nesta tarefa, utilizamos duas bases de dados: uma sobre os Processos Seletivos da Graduação, que detém informações numéricas sobre os processos seletivos da graduação da UFRGS e outra sobre o Quantitativo de Alunos de Graduação, que também detém informações demográficas sobre os alunos da universidade. Ambas estão disponíveis na página de dados abertos da UFRGS (<https://dados.ufrgs.br/>).

2.1. Diagrama ER das bases de dados do trabalho

Database ER diagram (crow's foot)

Lucas Brum | July 26, 2023



2.2. Exemplos de dados contidos nas bases

Processos Seletivos da Graduação: no ano de 2018, no Vestibular da UFRGS, o curso de Administração (Diurno), no semestre 1 foram oferecidas 14 vagas para Ampla Concorrência, legendada com a sigla de modalidade de vaga AC/A0.

Quantitativo de alunos de graduação: no ano de 2010, no período 1, o curso de Administração teve 1501 alunos vinculados, 1349 alunos matriculados, 169 alunos ingressantes, 91 alunos diplomados e 24 alunos evadidos.

2.3. Dicionário de dados

Dicionário de dados - Processos Seletivos da graduação

Campo	Tipo	Descrição
Ano	Inteiro	Ano de referência do processo seletivo
ProcessoSeletivo	Texto	Nome do processo seletivo
Curso	Texto	Curso
SiglaModalidadeVaga	Texto	Sigla da modalidade da vaga
ModalidadeVaga	Texto	Modalidade da vaga
Semestre	Texto	Semestre de ingresso
NrVagas	Inteiro	Número de vagas disponibilizadas para o curso, modalidade e semestre do processo seletivo

Dicionário de dados - Quantitativo de alunos da graduação

Campo	Tipo	Descrição
CodCurso	Inteiro	Código do curso
NomeCurso	Texto	Nome do curso
Ano	Inteiro	Ano do período letivo de referência
Periodo	Inteiro	Semestre do período letivo de referência
Vinculados	Inteiro	Número de alunos que possuem vínculo ativo no curso de graduação no período letivo de referência. O aluno pode estar matriculado, em trancamento, em licença, etc.
Matriculados	Inteiro	Número de alunos que estão matriculados em pelo menos uma atividade de ensino do curso de graduação, inclusive em mobilidade acadêmica, no período letivo de referência
Ingressantes	Inteiro	Número de alunos ingressantes em processos

		seletivos (vestibular, transferência interna, ingresso diplomado, etc.) que efetivaram matrícula por curso no período letivo de referência
Diplomados	Inteiro	Número de alunos que, após a conclusão de todos os créditos acadêmicos, apresentaram registro de diplomação por curso no período letivo de referência
Evadidos	Inteiro	Número de alunos que se desligaram por abandono, desistência de vaga, falecimento, transferência interna ou outras formas de saída por curso de graduação no período letivo de referência

3. Ambiente de desenvolvimento

As bases de dados descritas na seção 2 foram obtidas em .csv. Estão armazenados também no PostgreSQL, software que é bem conhecido entre os componentes do grupo. Para algumas visualizações e viabilização das análises, usamos o software PowerBI da Microsoft. Estamos fazendo consultas e demais manipulações usando a linguagem SQL dentro do Postgre

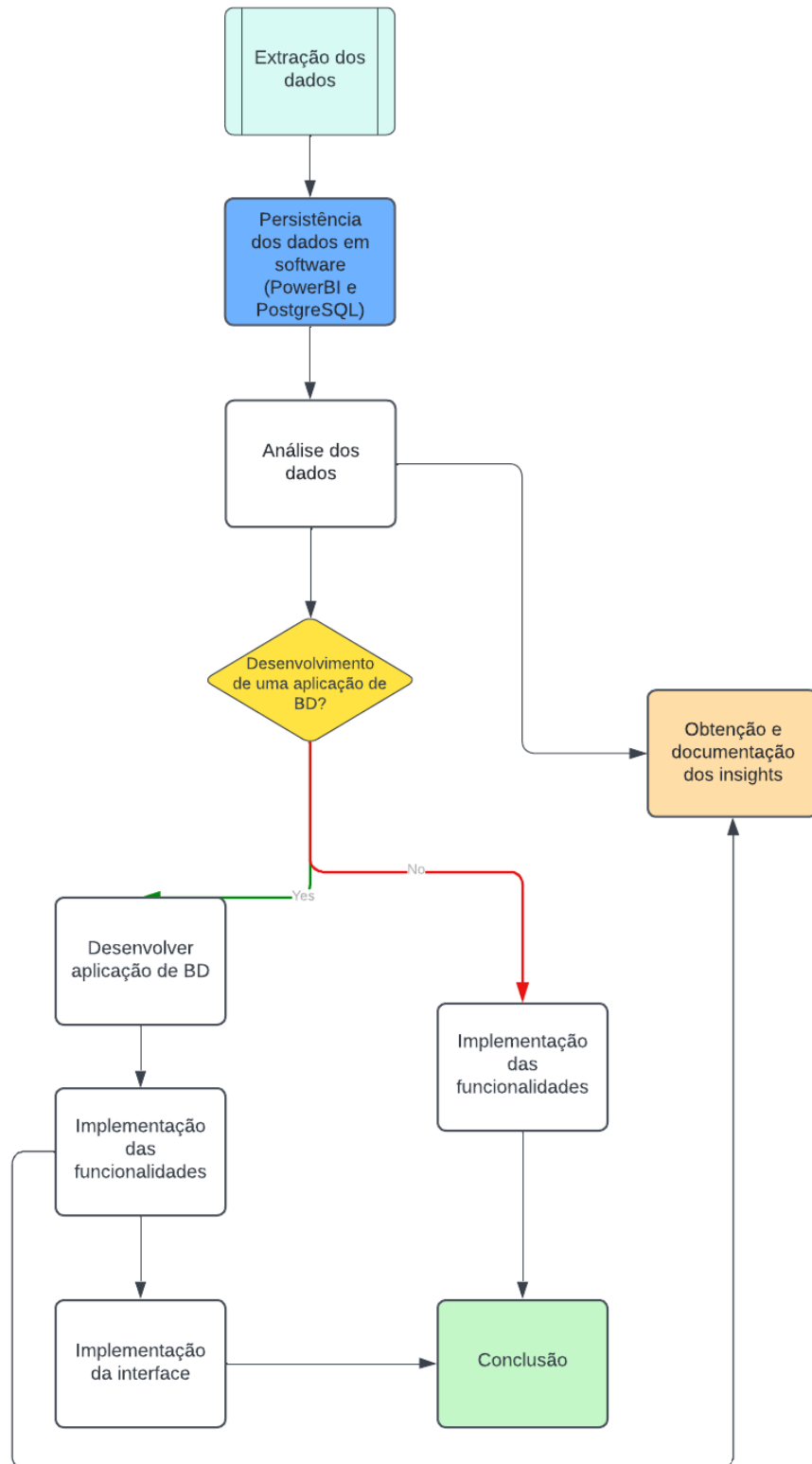
Equipamento 1

- PC MEGAWARE UPD/LX MEGACORP MOD. M7SERIES
- Placa mãe MEGAWARE MW-H61HD-MA
- Processador Intel i7-3770
- Memória 16GB DDR3 SDRAM
- HD Seagate ST1000DM010-2EP102 1000GB
- Sistema operacional Microsoft Windows 10 Home (x64) Build 19045.3208 (22H2)

Equipamento 2

- Placa mãe ASRock Steel Legend B450M mATX DDR4
- Processador AMD Ryzen 5 3600X 6-core 3.8GHz
- Memória 2x8GB DDR4 3000MHz
- Sistema operacional Microsoft Windows 11 Pro 64 bits (10.0, build 22621)
- M.2 NVMe Kingston KC3000 1TB

4. Metodologia



Os dados foram obtidos do portal de Dados Abertos da UFRGS e persistidos no PostgreSQL, alguns insights foram obtidos através do auxílio do uso do Microsoft PowerBI e que inclusive iam contra a possível hipótese de que Ciência e Engenharia da Computação seriam cursos mais requisitados atualmente. Estas novas percepções serão ilustradas em visualizações diversas (gráficos, cartões, interações gráficas, etc). Não foi necessário fazer a limpeza dos dados, consideramos todos os dados importantes para o estudo do caso.

5. Funcionalidades

As funcionalidades serão no formato de cinco consultas, três básicas e duas avançadas. Algumas serão feitas em SQL, outras visualizações. Adicionaremos os insights iniciais que nos levaram a montar as perguntas e o que buscamos com elas.

5.1. Consulta básica

Filtra os conceitos iniciais para a pesquisa:

```
select ano, processoseletivo, curso, nr vagas from processos_seletivos
```

5.2. Consulta sobre evasão dos cursos

Consulta que pesquisa os 5 cursos com maior taxa de evasão entre 5 e 13 anos. Com essa, buscamos entender se o curso de ciência da computação está entre as maiores taxas de evasão. Vieses adicionais: a maior taxa de evasão tem relação com a cursos de exatas?

```
SELECT nomecurso, SUM(evadidos) as total_evadidos  
FROM quantitativo_alunos  
WHERE ano >= 2023 - 5  
GROUP BY nomecurso  
ORDER BY total_evadidos DESC  
LIMIT 5;
```

```
SELECT nomecurso, SUM(evadidos) as total_evadidos  
FROM quantitativo_alunos  
WHERE ano >= 2023 - 13  
GROUP BY nomecurso  
ORDER BY total_evadidos DESC  
LIMIT 5;
```

5.3. Consulta sobre formação e ingresso

Consulta que pesquisa os 5 cursos com maior taxa de evasão/matriculados e ingresso/matriculados. Com essa, buscamos entender se existe uma relação entre os que mais formam com os que mais ingressam alunos. Vieses adicionais: A ciência da computação está nesse grupo? Se sim, tem uma taxa de evasão grande? Para onde vai esse povo?

```
SELECT nomecurso, SUM(diplomados) as total_diplomados  
FROM quantitativo_alunos
```

```

GROUP BY nomecurso
ORDER BY total_diplomados DESC
LIMIT 5;

```

```

SELECT nomecurso, SUM(ingressantes) as total_ingressantes
FROM quantitativo_alunos
GROUP BY nomecurso
ORDER BY total_ingressantes DESC
LIMIT 5;

```

5.4. Consulta de % de top (cursos,período) evadidos

Essa consulta calcula a % de alunos evadidos e imprime em ordem decrescente os mesmos. Podemos tirar a conclusão qual curso tem a maior taxa de evadidos por alunos vinculados, o que mostra uma informação mais certa, pois, cursos que tem mais alunos vinculados, terão mais alunos evadidos proporcionalmente, então é importante puxar a taxa de evadidos/vinculados.

```

SELECT nomecurso, ((evadidos/vinculados)*100)
as porcentagem_evadidos
FROM quantitativo_alunos
ORDER BY porcentagem_evadidos
DESC
LIMIT 5;

```

6. Cronograma

6.1. Datas da disciplina

- 26/07 - Anteprojeto
- 02/08 - Avaliação do Anteprojeto dos Colegas
- 30/08 - Relatório Parcial (metodologia + consultas simples)
- 11/09 - Relatório Final

6.2. Cronograma proposto

Data	Tarefa	Componente do Grupo
02/08	Decidir sobre o desenvolvimento da aplicação	todos
10/08	Protótipo da aplicação	Lucas
17/08	Implementação das consultas e demais análises	Lucas, Leonardo
25/08	Primeira versão da aplicação	Artur
11/09	Versão final da aplicação	todos

7. Considerações finais

Com este trabalho, esperamos aperfeiçoar nossos conhecimentos de banco de dados sob duas principais perspectivas: teórica e prática. No quesito teórico, é natural que precisaremos aperfeiçoar nossos conhecimentos provenientes de experiências prévias, como a cadeira de fundamento de banco de dados, para realizar o todo. No quesito prático, como o nome já diz, entra em voga o momento de relacionar o conteúdo internalizado com as questões críticas ao tema central do projeto.

Para isso, teremos que aperfeiçoar habilidades em linguagem de query (SQL), além de como importar dados da base de dados da UFRGS (que estão em CSV) para dentro das tabelas no PostgreSQL. Também, importante, teremos que achar as queries certas para responder as questões que estávamos nos perguntando, como, por exemplo, qual é o curso que mais tem taxa de evasão. Tudo isso auxiliado pelo fato de ter escolhido de forma correta como organizar as tabelas do BD.

A contribuição social pretendida é tornar acessível o conteúdo e análise de dados sobre os quantitativos de alunos e de processos seletivos. O tema partiu de uma dúvida geral que se move nos corredores e salas de informática: A evasão do curso tem uma taxa alta? Como isso se relaciona com a crescente na adesão, isto é, na procura a ingresso do curso por novos estudantes? De modo geral, todos os integrantes já se depararam com essa dúvida principal sobre a evasão. Porém, ela se mostra vaga por diversas vezes e o trabalho entra justamente nesse ponto para tornar palpável essa análise e possibilitar aprendizados.

Referencias Bibliográficas

Ribeiro, R. (2023). Uma proposta de extração, transformação, carga e visualização para os dados do Censo Escolar, <https://lume.ufrgs.br/handle/10183/259957>

Perrone, S. P. (2023). Uma ferramenta web para a automatização de relatórios da Sociedade Brasileira da Computação sobre dados referentes ao ensino nacional de tecnologia, <https://lume.ufrgs.br/handle/10183/261792>

Site oficial PostgreSQL <https://www.postgresql.org/>

Site oficial PowerBI <https://powerbi.microsoft.com/>

Portal de Dados Abertos da UFRGS: Processos Seletivos da Graduação <https://dados.ufrgs.br/dataset/processos-seletivos-graduacao>

Portal de Dados Abertos da UFRGS: Quantitativo de Alunos da Graduação <https://dados.ufrgs.br/dataset/quantitativo-de-alunos-de-graduacao>

GZH. Com aumento de quase 45% no número de candidatos, UFRGS divulga a densidade do vestibular 2023, <https://gauchazh.clicrbs.com.br/educacao-e-emprego/noticia/2022/12/com-aumento-de-quase-45-no-numero-de-candidatos-ufrgs-divulga-densidade-do-vestibular-2023-clbs0gm5p001x018fhys2dsd9.html>