

# Análise de dados do Quantitativo de alunos da UFRGS

Lucas Matheus Dias Brum

Bacharel em Ciência da

Computação

Porto Alegre, Brasil

lmdbrum@inf.ufrgs.br

## ABSTRACT

Este artigo tem como objetivo descrever uma experiência científica e também fazer uma análise crítica sobre o programa Microsoft Power BI, referência em análise de dados usando como base de dados quantitativo de alunos da UFRGS.

## Palavras-chave do autor

análise de dados; quantitativo de alunos; gráficos; visualização de informações;

## INTRODUÇÃO

A fim de verificar o porquê de um certo aumento na população acadêmica no instituto de informática da Universidade Federal do Rio Grande do Sul, foi feita uma pesquisa nos dados abertos da UFRGS para verificar se houve aumento recente na busca pelos cursos de Ciência da Computação e Engenharia da Computação na universidade. Porém, não foi apresentado nenhum aumento nem mesmo com análises SQL ou com auxílio dos gráficos. Com esse ponto de partida foram aplicadas várias análises e escolhido o software Microsoft Power Bi para ambientar as análises e, por fim, fazer uma análise crítica sobre a plataforma, que será discutida aqui.

## EXTRAÇÃO DOS DADOS

A UFRGS mantém uma plataforma onde se obtém dados abertos acerca de assuntos pertinentes à Universidade, o Dados UFRGS. Ali, num primeiro momento, foram obtidos os dados sobre Processos Seletivos feitos pela UFRGS e também os dados sobre Quantitativo de alunos formalmente relacionados à Universidade a fim de estabelecer uma relação entre oferecimento de vagas e ocupação das mesmas.

As tabelas incluem informações pertinentes para análise tais como:

Processos Seletivos: Ano - ano de referência do processo seletivo; ProcessoSeletivo - nome do processo seletivo; Curso - nome do curso; SiglaModalidadeVaga - sigla da modalidade da vaga; ModalidadeVaga - modalidade da vaga; Semestre - semestre de ingresso; NrVagas - número de vagas disponibilizadas para o curso, modalidade e semestre do processo seletivo.

Quantitativo de alunos: CodCurso - código do curso; NomeCurso - nome do curso; Ano - ano do período letivo de referência; Período - semestre do período letivo de referência; Vinculados - 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 - 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 - Número de alunos ingressantes em processos seletivos (vestibular, transferência interna, ingresso diplomado, etc.) que efetivaram matrícula por curso no período letivo de referência; Diplomados - Número de alunos que, após a conclusão de todos os créditos acadêmicos, apresentaram registro de diplomação por curso no período letivo de referência; Evadidos - 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.

Os dados obtidos possuem atributos categóricos (Nome do curso), ordenados - quantitativos - sequenciais (Ingressantes, Evadidos, Diplomados, Vinculados, Matriculados, Período,

## PRIMEIROS PROCEDIMENTOS

Depois de extraídos os dados, foi feita uma experiência com a linguagem Python utilizando a biblioteca Numpy para uma primeira análise. Ao mesmo passo, foi usado o software PowerBI da Microsoft para análise também, verificando se os resultados convergiam e comprovando sua eficácia.

No primeiro momento, seria feito uma comparação entre as duas formas de análise, porém, com o passar do tempo, foi observada uma maior facilidade em elaborar gráficos e visuais de forma mais rápida e entendível, portanto, apenas o PowerBI foi usado até o final do trabalho para auxílio das visualizações.

Decidida a plataforma que seria usada para as análises, foram feitos os primeiros visuais. No PowerBI, existe um recurso simples chamado Cartão, que resume um *insight* que se sobressaia sobre os dados, que neste caso pode ser o curso mais evadido, curso mais ingressado, etc.

Partin foram usados vários Cartões para nos dar as respostas mais diretamente - as informações que mais se sobressaem dentre outras - de nossa hipótese inicial, de que os cursos de Ciência e Engenharia da Computação seriam os mais buscados na UFRGS atualmente, que infelizmente foram invalidados, já que na verdade os cursos mais ingressados

são Ciências jurídicas e sociais, administração, letras e etc. como será mostrado mais adiante.

Posteriormente, a plataforma foi explorada com os visuais que poderiam ser feitos através da mesma. De gráficos mais simples e conhecidos até os diferenciados. Gráficos de barra duplo, gráfico de linha, gráfico estilo pizza/rosca - estes foram experimentados porém não faziam sentido para o dashboard final pois prejudicava a visualização dos mesmos nas figuras já que são muitos dados sobre nomes dos cursos, mas parecem interessantes para dados de uma tabela considerada inicialmente, a de processos seletivos onde os dados foram descritos e discriminados através de modalidades de vaga e forma de ingresso.

**VISUALIZAÇÕES EMPREGADAS**

**Cartões**

Com os cartões, é possível resumir informações que predominam nos dados, destacando os dados que mais ou menos se propagam. Nesse caso, foi relevante atentar qual curso mais ingressado, mais evadido e bem como totalizar os dados referentes aos matriculados na UFRGS até aquele momento, ingressantes, entre outros.

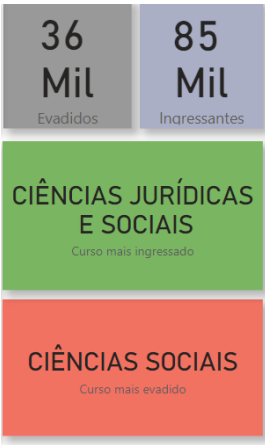


Figura 1. Elementos da visualização Cartão do PowerBI.

**Barra dupla**

Com o gráfico de barras duplas, é possível e efetiva uma comparação direta entre dois aspectos de interesse.- melhor ainda se forem antagônicos. Aqui usa-se a soma de ingressantes e evadidos para se ter um panorama direto dessa discrepância entre os que entram e saem da Universidade.

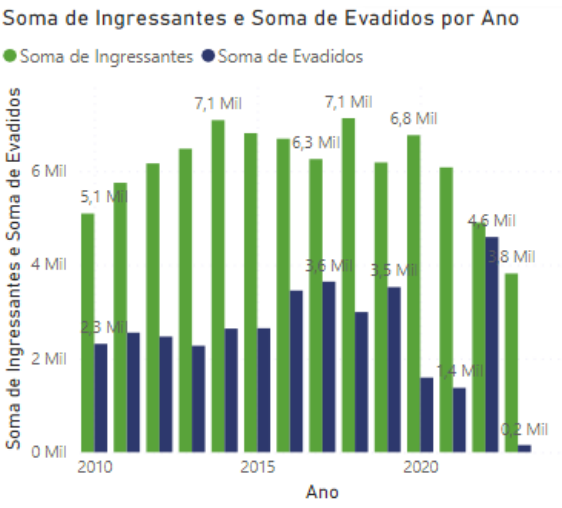


Figura 2. Gráficos de barra dupla.

**Linha**

Um visual mais simples, que observa o comportamento linear de determinado aspecto de interesse, que se comporta de maneira síncrona aos valores (pontos) dos dados de amostra.

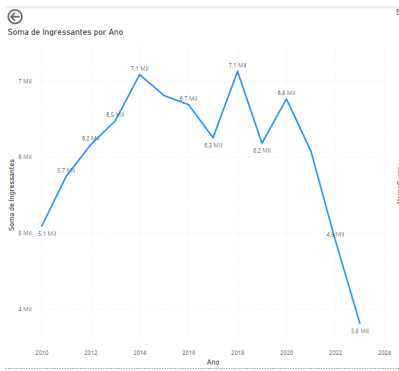


Figura 3. Gráfico de linha simples.

**Pizza**

Dentro de um total de entradas distintas, delimita-se as parcelas baseadas em categorias relevantes. Observamos aqui o que mais parecia ser oportuno em um gráfico do estilo: as porcentagem de forma de ingresso bem como o processo seletivo em questão.

Com viés de totalidade dos dados em questão, os gráficos de pizza e de rosca encarregam-se de particionar dados em questão, dada uma categoria .

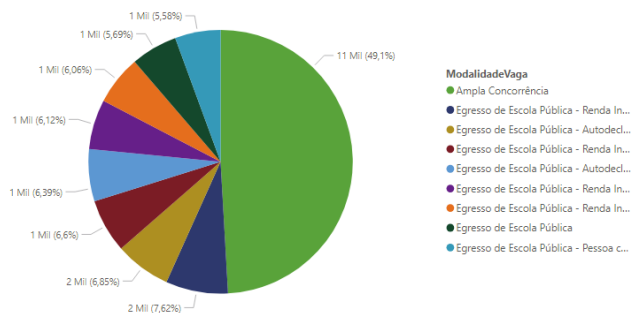


Figura 3. Gráfico de pizza.

### Treemap

Geralmente bonitos visualmente, os treemaps mostram os dados de forma hierárquica fortemente aliada aos valores conforme uma dada classificação. No nosso caso, utiliza-se a descendência do maior para o menor, o que resultou num treemap pouco variado, muito possivelmente pela forma como o PowerBI dispõe os dados, o que causa confusões, pois aparentemente o software não distingue - pelo menos a um nível visivelmente perceptível - as várias categorias dos valores.

### Pizza/Rosca

Quando utilizamos visuais que parecem ser importantes para um conjunto, pensamos que seria viável e até mesmo indispensável se criar um dashboard que sintetize essa análise, montando gráficos que trabalhem em conjuntos em torno de eixos como Ano, Curso e Período, que sustentam e coordenam consistentemente a dinamicidade dos dados.

### Segmentação de dados

A segmentação de dados, na opinião do autor, o elemento mais efetivo e administrativo do PowerBI, serve para selecionar os parâmetros que montam e resumem o visual. Nesta análise foram escolhidos Ano, Período e NomeCurso.

### SOBRE O DASHBOARD FINAL

Compõe-se o dashboard com um pequeno resumo de todas as visualizações estudadas, gráficos de barra duplo, segmentação de dados, cartões e treemap. A segmentação de dados De forma síncrona e legível, funcionam quase todos os visuais, exceto os de curso mais evadido e mais ingressado, que congelam no mesmo nome do curso, já que as escolhas monopolizam o curso em destaque.

### POWERBI

Microsoft PowerBI é uma plataforma unificada e escalável para Business Intelligence (BI) ou Inteligência de Negócios em português, referência em análise e visualização de dados. É possível utilizar a plataforma criando uma conta gratuita no Microsoft Fabric, plataforma que possui um pacote de serviços de análise. No primeiro momento, pode-se requerer uma avaliação gratuita do PowerBI, que pode ser renovada futuramente depois de trinta dias. Possui

versões web, Desktop e até mesmo mobile. Para esta análise, foi utilizada a versão Desktop no Windows 10.

Para a instalação Desktop, basta acessar o Microsoft Store, aplicativo que vem com a instalação do Windows 10 e procurar na barra de pesquisa PowerBI e clicar em 'adquirir'. Não há versão do PowerBI para Linux ou MacOS, porém, há a versão Mobile para Android e iOS.



### Interface

A interface do PowerBI no geral é positiva, organizada dentre tantas possibilidades, porém um pouco poluída na parte em que se escolhem as visualizações a serem dispostas nas páginas, já que existem muitas visualizações e muitas formas de se dispor os dados visualmente. A plataforma é realmente muito rica em possibilidades de se fazer um dashboard simples ou até um mais avançado e conta com muitos comandos para isso, o que efetiva uma análise satisfatória, principalmente no quesito dos negócios no jargão das empresas, seja qual tamanho elas tenham.

### CONSIDERAÇÕES FINAIS

A análise desses dados surgiu da percepção de que, nos últimos semestres, especialmente no último (2023/1, conforme datado neste artigo), o Instituto de Informática da UFRGS e o Restaurante Universitário, em particular o RU 06, localizado no Campus do Vale da Universidade, entre os bairros Agronomia e a entrada de Viamão, estavam enfrentando superlotação. No entanto, os dados coletados não corroboram a hipótese de que os cursos oferecidos pelo Instituto de Informática eram os mais procurados naquele momento.

Esses dados, entretanto, serviram como uma oportunidade para aplicar conceitos de Visualização de Informações, explorando diversas formas de representação visual para tornar os dados mais intuitivos e compreensíveis. A experimentação de diferentes abordagens visuais e a exploração de novas maneiras de visualizar dados tornam-se interessantes não apenas do ponto de vista visual e gráfico, mas também na busca por abordagens mais interativas, tornando as informações menos estáticas e mais tangíveis.

Porém, também é necessário pontuar que a ausência de autosave na plataforma PowerBI pode confundir os usuários e também a não-maleabilidade para dispor o dashboard em tela cheia para melhor exibição - pelo menos na versão Desktop são pontos bastante problemáticos - principalmente para usuários desatentos. Também devido à alta gama de

comandos, pode ser que alguns usuários se percam, seja executando os comandos do Power Query, com o intuito de se usar os preceitos de Data Analysis Functions (DAX), bastante valorizados pela plataforma, bem como os de ETL (Extract, Transform & Load).

Em suma, a visualização de informações desempenha um papel vital na comunicação eficaz de dados complexos. Gráficos, barras, gráficos de pizza e outras representações

visuais não apenas simplificam a compreensão, mas também destacam tendências e insights cruciais. Em um mundo cada vez mais orientado por dados, dominar a arte da visualização é essencial para a tomada de decisões informadas e a disseminação eficaz do conhecimento. Portanto, a utilização estratégica dessas ferramentas visuais é um componente essencial para a pesquisa, comunicação e compreensão de informações em diversas áreas do conhecimento.

## REFERÊNCIAS

1. Tamara Munzner. 2015. *Visualization Analysis & Design*
2. Dados de Processos Seletivos da UFRGS, encontrados em <https://dados.ufrgs.br/dataset/processos-seletivos-ggraduacao>
3. Dados de Quantitativo de Alunos da UFRGS encontrados em, <https://dados.ufrgs.br/dataset/quantitativo-de-alunos-de-graduacao>
4. GZH. 2022. *Com aumento de quase 45% no número de candidatos, UFRGS divulga densidade do vestibular 2023*. Disponível em: <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>
5. Pedro Perrone. 2023. *Uma ferramenta web para a automatização de relatórios da Sociedade Brasileira de Computação sobre dados referentes ao ensino nacional de tecnologia*. Disponível em: <https://lume.ufrgs.br/handle/10183/261792>
6. Rafael Ribeiro. 2023. *Uma proposta de extração, transformação, carga, e visualização para os dados do Censo Escolar*. Disponível em <https://lume.ufrgs.br/handle/10183/259957>