

# Mulheres nas Exatas: uma análise das estudantes formadas em TI

**Marina A. Melaré, Daniela Marques**

Tecnólogo em Análise e Desenvolvimento de Sistemas - Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Hortolândia (IFSP) Avenida Thereza Ana Cecon Breda, s/n, Vila São Pedro - Hortolândia – SP – Brasil,

marina.melare@aluno.ifsp.edu.br, marquesdaniela@ifsp.edu.br

**Abstract.** *Several articles and discussions nowadays bring up the difference in female participation in the Information Technology field, both in the executive market and in college education, when compared to male participation. This article presents an analysis of open data provided by the government in the last four years to verify if, even with this topic being highly addressed nowadays, there is still a large variation in the conclusion of courses in this area. In addition, an analysis was also made in the undergraduate area of female interest, as well as the social data provided at the time of enrollment in the courses.*

**Resumo.** *Diversos artigos e discussões atuais ressaltam a diferença da participação feminina na área de Tecnologia da Informação, tanto no mercado de trabalho, como nos cursos superiores, quando se comparada à masculina. Este artigo traz uma análise de dados abertos fornecidos pelo governo brasileiro nos últimos quatro anos para verificar se, mesmo com este tópico sendo bem abordado nos dias de hoje, ainda existe uma discrepância grande na conclusão dos cursos desta área. Além disso, também foi realizada uma análise na área de graduação de interesse feminino, assim como os dados sociais fornecidos no momento de matrícula nos cursos.*

## 1. Introdução

Até pouco tempo atrás, quando uma criança nascia, sua criação era baseada no seu respectivo sexo. Suas brincadeiras, modo de agir, de pensar e, conseqüentemente, suas habilidades eram moldadas pelos estigmas sociais existentes até então.

“Se o gênero pode ser definido como a construção social, cultural e histórica do masculino e do feminino, como fazem pesquisadores tão distantes, [...] também a incompetência técnica das mulheres é o resultado de uma construção social. É em nome da incompetência técnica construída desde a infância na socialização familiar, diferenciando papéis masculinos e femininos, no exercício cotidiano que se dá a técnica para o menino, como por exemplo, consertar o carro, consertar pequenas máquinas, quando o pai vai chamar o menino para fazer estas coisas e o aprendizado da relação de serviço para as meninas: cuidar da boneca, cuidar do irmão menor, ajudar a pôr a mesa. As funções e cargos atribuídos às mulheres são sempre caracterizados por esta relação de serviço” (HIRATA, 2003).

Quando estes estigmas continuam a ditar a forma em que a mulher e o homem são tratados na sociedade, a luta para a quebra dele passa a ser ainda mais difícil, devido a

quantidade de anos em que isso era considerado como verdade absoluta, fazendo a história se manter presa no passado, e o futuro ser mais do mesmo.

“Os homens continuam ditando o que condiz com o “ser homem” e o “ser mulher” ainda no século XXI e o desafio é atrever-se a pensar diferente, a pensar na despolarização dos gêneros, permitindo que tanto homens quanto mulheres possam sentir-se parte de um contexto histórico como sujeitos de sua história, de sua vida, de seu “que fazer” no mundo. Daí decorre a luta. Lutar nessa perspectiva compreende a possibilidade de enfrentamento com o que está imposto, contestando a disseminação de uma concepção de inferioridade feminina e de sua naturalização” (CAMARGO, 2014).

O que pode parecer um pequeno incentivo da parte dos responsáveis pela criança, pode se tornar, em um futuro, uma influência considerável quando o assunto é decisões. Segundo a pesquisadora Ilka Bichara, doutora em Psicologia e professora do Instituto de Psicologia da UFBA (Universidade Federal da Bahia):

“Nas brincadeiras, há negociação, exercício de habilidades sociais complexas, formação de valores e conceitos, como também há emoção. [...] A criança não é um ser passivo que vai apenas reproduzindo aquilo que vê, pelo contrário, ela é um ser ativo, que reinterpreta e recria a realidade em suas brincadeiras e, com isso, cria cultura” (BICHARA, 2012).

As brincadeiras administrativas previamente atribuídas a meninas, podem sim ter impacto na decisão de sua carreira no futuro. Por essa maneira de criação ter sido contrariada há pouco tempo atrás, ainda colhemos os frutos dela nas graduações. No princípio, TI possuía um grande número de mulheres na faculdade, porque até então, os computadores eram usados para tarefas administrativas (SANTOS, 2019). Com o surgimento dos computadores pessoais e sua utilização lúdica para jogos, houve uma inversão de gênero na área da Computação. Até hoje, o maior público na Computação é masculino, o que limita o desenvolvimento de aplicações à uma perspectiva masculino. Em 2014, a Apple lançou pela primeira vez um aplicativo que prometia acompanhar todos os dados envolvendo a saúde de seu usuário, menos dados sobre o ciclo menstrual feminino (DUHAIME-ROSS, 2014). Quanto mais diversidade presente em um time de desenvolvimento, mais a aplicação desenvolvida atenderá a todo tipo de público. No entanto, há iniciativas para aumentar a participação feminina na área de exatas, como discussões levantadas dentro das empresas e programas de incentivo como o *Woman Can Code*<sup>1</sup>. Levando isso em consideração, este artigo pretende verificar se dentro da área da Computação a participação do público feminino mudou por causa dos incentivos governamentais e corporativos, e ainda fazer uma análise conforme a modalidade de ensino. Desta forma, ainda sem dados, uma hipótese de pesquisa foi proposta:

- H1) A conclusão de mulheres nos cursos de TI, nas diferentes modalidades de ensino - bacharel e tecnólogo, é mais baixa quando se comparada a dos homens, mesmo com os incentivos que ocorrem para aumentar a participação feminina.

Conforme os dados foram obtidos e analisados, abordando tanto aspecto social como uma visão mais ampla dos cursos superiores, um leque de possibilidades foi aberto e as seguintes hipóteses surgiram:

---

<sup>1</sup> <https://sharerh.com/women-can-code-oficial/>

- H2) O interesse feminino pela área de TI é menor quando se comparado ao masculino.
- H3) As mulheres possuem um interesse maior em cursos das áreas de Humanas e/ou Biológicas.
- H4) As mulheres possuem uma taxa de graduação maior do que os homens.

Como os dados abertos utilizados datam de 2018 a 2021, as resoluções das hipóteses serão referentes a este período.

Havia um interesse de realizar uma análise dos dados relacionados às bolsas do PROUNI e os dados de matrículas e conclusão dos cursos superiores para verificar se havia relação na obtenção de bolsas e a taxa de conclusão quando separado por gênero. Contudo, essa análise não foi possível pois não era possível relacionar os dados abertos obtidos das bolsas (dados até o ano de 2018) e os dados de matrículas e emissão de diplomas (a partir desse mesmo ano).

Para validar as hipóteses de pesquisa, este artigo realizou uma análise baseada em dados abertos buscando informações sobre a procura e conclusão do público feminino de acordo com a modalidade do curso e aplicou a mineração de dados para esse estudo. Esse artigo está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico, na Seção 3 são descritos trabalhos correlatos, a metodologia é explicada na Seção 4, a Seção 5 mostra o desenvolvimento do trabalho e, por último, a Seção 6 conclui o trabalho.

## 2. Referencial Teórico

As subseções a seguir apresentam definições que serão utilizadas no decorrer deste trabalho. Será descrito o cenário das Mulheres na Computação, seguido de dados abertos e o processo KDD (*Knowledge discovery in databases*).

### 2.1 Mulheres na Computação

Mesmo a primeira linguagem de programação tendo sido inventada por uma mulher (A PRIMEIRA... 2020), ou atualmente termos um rosto feminino liderando o desenvolvimento do programa que tirou a primeira fotografia de um buraco negro pela NASA (GALILEU, 2019), o reflexo da participação feminina na nossa sociedade ainda é, infelizmente, baixo.

De acordo com uma pesquisa realizada pela professora titular do Instituto de Bioquímica Médica Leopoldo de Meis, da Universidade Federal do Rio de Janeiro (UFRJ) Débora Foguel:

“Cientistas mulheres foram premiadas 58 vezes no intervalo entre 1901-2020 em um total de 950 laureados, dentre os quais 27 são organizações. Marie Curie, uma das mulheres dessa minguada lista, acumula duas premiações (em 1903, prêmio de Física e em 1911, prêmio em Química). Dessa forma, o número real de mulheres agraciadas é 57.” (FOGUEL, 2020)

No artigo “*AI Is the Future—But Where Are the Women?*”, Tom Simonite (2018) traz como o desenvolvimento de inteligências artificiais (IAs) tem crescido, mas como a participação feminina dentro deste campo ainda é baixa. As empresas investem no crescimento desse campo. Numa pesquisa realizada pela companhia onde Simonite

trabalha, *WIRED*, eles descobriram que apenas 12% dos pesquisadores de IAs eram mulheres.

“Na empresa *Google*, 21% dos cargos técnicos são preenchidos por mulheres, de acordo com os dados disponibilizados pela companhia em junho. Quando a *WIRED* revisou o tópico relacionado às pesquisas de IA do *Google*, eles listaram 641 pessoas trabalhando em inteligência de máquina, das quais apenas 10% eram mulheres. O *Facebook* relatou no mês passado que 22% dos seus funcionários técnicos são mulheres.” (SIMONITE, 2018, tradução nossa).

Segundo o artigo “Desempenho Acadêmico e o Ingresso no Curso Superior: uma análise das estudantes ingressantes entre 2016 a 2020 do curso Técnico em Informática Integrado ao Ensino Médio”, de 2021, nos últimos 5 anos, houve um aumento considerável de meninas ingressando no curso técnico de informática integrado, oferecido pelo Instituto Federal de São Paulo - Campus Hortolândia. Porém, ao analisar o interesse das alunas durante o curso notou-se que, mesmo com boas notas, o curso não era algo que despertava gosto e/ou curiosidade (MARQUES *et al.*, 2021). Ainda neste artigo, foi levantado o dado de que 88% das estudantes entrevistadas afirmaram que a experiência no curso técnico serviu como experiência para descobrirem se gostavam ou não da área.

Atualmente, centenas de artigos sobre o tema podem ser encontrados. A Sociedade Brasileira de Computação (SBC) teve a iniciativa de criar um *workshop* chamado *Women in Information Technology* (WIT)<sup>2</sup>, onde criou-se um espaço para discussão de assuntos relacionados à gênero no meio da Tecnologia da Informação.

“Organizado na forma de palestras convidadas e painéis, o *workshop* é centrado em debater problemas relacionados à mulher e ao seu acesso à TI, tanto do ponto de vista de mercado de trabalho quanto de inclusão e alfabetização digital. Os temas abordados se concentram na necessidade de educar, recrutar e treinar mulheres, como uma política estratégica para o desenvolvimento e competitividade nacional e regional” (WIT, 2020).

Além do WIT, existe o Fórum Meninas Digitais, que está ativo desde 2011, trazendo um incentivo para alunas do ensino fundamental ao tecnológico para conhecer melhor a área de TI, assim como trazer um incentivo para seguir carreira na área. Mas se tantos trabalhos com o tema da participação feminina na área de Tecnologia da Informação são encontrados, isso significa que o problema ainda existe.

## 2.2 Dados Abertos

Segundo a definição da plataforma de conhecimentos abertos OKFN, embasado também pela plataforma de dados do governo:

“Dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura” (OPEN KNOWLEDGE, 2021, tradução nossa).

---

<sup>2</sup> <https://meninas.sbc.org.br/sobre/women-in-information-technology/>

## 2.2 Processo Knowledge-Discovery in Databases (KDD)

Dentro do campo de mineração de dados, há um processo conhecido como *Knowledge Discovery in Databases* (Descoberta de conhecimento em bancos de dados). Existente desde 1989, este processo representa o processo de coleta de dados e o refinamento dos mesmos de maneira metodológica. Em um nível de detalhes um pouco maior, o processo realiza a extração de dados brutos previamente desconhecidos, vindos de uma base de dados, e os transforma em informações utilizáveis, gerando assim um novo conhecimento sobre um respectivo assunto (TEÓFILO, 2015).

Como representado na Figura 1, o processo KDD possui algumas etapas. Primeiramente, realiza-se uma seleção dos dados relevantes. Após isso, é feito um pré-processamento e formatação para que eles sigam um mesmo padrão. A etapa seguinte é a de mineração de dados, onde os métodos de mineração são aplicados a fim de se extrair padrões previamente desconhecidos, que são então interpretados e analisados, trazendo à tona o novo conhecimento obtido.

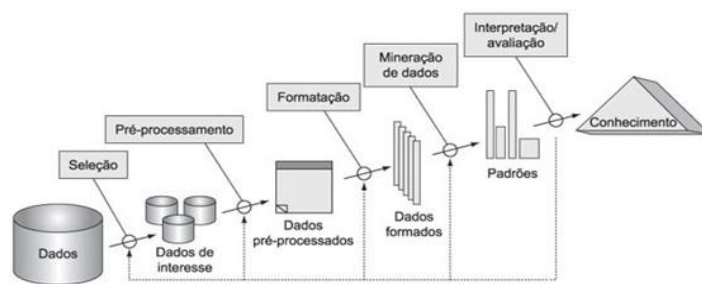


Figura 1. Processo KDD (Fonte: TEÓFILO, 2015)

Já o termo *Data Mining*, traduzido para “Mineração de dados”, é frequentemente confundido, ou então até erroneamente trocado com KDD (KDD... 2021). Porém, como mencionado anteriormente, este é o quarto passo do processo, também conhecido como etapa central, onde algum algoritmo é aplicado para extrair algum padrão de dados. O *Python*, uma linguagem de programação de alto nível e interpretada, é comumente utilizado para essa etapa de mineração. Ele possui uma biblioteca chamada *Pandas*, que é uma ferramenta poderosa e flexível quando se fala de análise de dados. Juntamente com as bibliotecas *matplotlib*, para criação de recursos visuais com os dados, e *NumPy* para a realização de operações matemáticas, os dados podem ser mineirados de uma maneira rápida e precisa (SANTANA, 2017).

## 3. Trabalhos correlatos

Dentre os diversos artigos dentro do tema “mulheres na ciência”, um deles possui uma abordagem semelhante com a que será realizada neste. “Análise da participação das mulheres na ciência: um estudo de caso da área de Ciências Exatas e da Terra no Brasil” (TUESTA; DIGIAMPIETRI; DELGADO; MARTINS, 2018) traz, através da extração de dados da plataforma de currículos Lattes, uma análise comparativa da participação, produção e graduação de mulheres nas área de Ciências Exatas e da Terra no Brasil. Os autores consideraram dados como: sexo, ano de graduação, publicações, ano de início e ano de conclusão de curso em suas pesquisas, feitas com 40.882 pesquisadores.

Outro trabalho, focado na área de Tecnologia da Informação, traz também o comparativo da participação feminina nessa área. “A desigualdade de gênero no campo da Tecnologia da Informação” (CURSINO; MARTINEZ, 2021) extrai dados dos censos de 1980, 2000 e 2010 e também de um serviço de informação ao cidadão (SIC) de uma empresa pública com filiais pelo Brasil e, a partir disso, embasou sua análise quantitativa da porcentagem de mulheres no campo de TI, na qual concluiu-se que o número de mulheres nos cursos da área de TI está diminuindo.

#### **4. Metodologia**

A primeira atividade desta pesquisa foi identificar uma pergunta de pesquisa no contexto de Mulheres na Computação que pudesse ser verificada por meio dos dados abertos disponíveis.

Para essa análise, foi realizado um estudo sobre o conceito e aplicação de dados abertos e *data mining*. Após esse estudo, foi feita também uma pesquisa de dados abertos disponíveis (MEC) para a aplicação desenvolvida consumir, e em paralelo, uma busca sobre o processo KDD. Este processo estabelece relações em grandes volumes de dados, buscando padrões dentro de dados consistentes. Uma vez que os dados foram consolidados, a aplicação desenvolvida em Python os consumiu, gerando os dados necessários para análise e argumentação.

#### **5. Desenvolvimento**

Esta seção descreve o desenvolvimento do trabalho, desde o conceito e aplicação de dados abertos e KDD, até o cenário de implementação e resultados obtidos.

Com base na realidade encontrada em artigos, como os mencionados nos trabalhos correlatos, sobre a diminuição da participação feminina na área de exatas, este artigo tenta corroborar com essas informações. Com isso em mente, uma pergunta de pesquisa e quatro hipóteses foram levantadas. Para comprovar essas informações, foram utilizados dados abertos e o processo de mineração de dados.

##### **5.1. Dados abertos**

O objetivo do uso de dados abertos neste trabalho é conseguirmos extrair deles padrões e conhecimentos que sirvam como base para a discussão levantada. Para isso, foi feita uma procura por dados abertos na web.

Os dados abertos foram extraídos de três fontes diferentes:

1. Sessões de PNP e PROUNI, disponibilizados pelo MEC no endereço <http://dadosabertos.mec.gov.br/>
2. Diplomas de graduação, no endereço [https://dados.gov.br/dataset/diplomas\\_graduacao](https://dados.gov.br/dataset/diplomas_graduacao)
3. Senso de educação superior, no endereço <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>

Todos os dados foram baixados no formato CSV, estando também disponíveis em PDF. Porém, alguns campos com dados iguais possuíam nomes diferentes, como por exemplo “ANO\_CONCLUSAO” e “DATA\_CONCLUSAO”, quando ambos

referenciavam apenas o ano. Uma limitação encontrada foram os dados de bolsa, limitados apenas aos dados do PROUNI.

## 5.2. KDD

### 5.2.1 Seleção

Para a seleção dos dados utilizados para mineração, uma pesquisa foi realizada dentro das plataformas governamentais, buscando dados relacionados à taxa de inscrição em cursos superiores, taxa de graduação, bolsas de estudo, emissão de diplomas, havendo então algum identificador de sexo do estudante, para embasar a pergunta central do artigo. Quando não encontrados nas plataformas do governo, com o auxílio da orientadora do trabalho e das plataformas de pesquisa disponíveis para acesso de todos, estas informações foram encontradas.

A pesquisa nos direcionou para as três bases de dados abertas informadas na Seção 5.1. Após análise do conteúdo, foi realizado o *download* destes arquivos estruturados no formato csv, que são arquivos onde os dados são separados ou delimitados por vírgulas. Na seção referente aos dados do PROUNI, havia a possibilidade de baixar os arquivos no formato XML, que é um tipo de arquivo de linguagem de marcação extensível e é usado para estruturar dados para armazenamento e transporte, porém, devido ao fato de os outros arquivos não possuírem essa opção, e pela escolha da biblioteca *Pandas* para minerar os dados, eles também foram baixados em formato csv.

### 5.2.2 Pré-processamento

Uma vez que os dados foram selecionados e baixados, percebeu-se que algumas das planilhas não possuíam a coluna de sexo do aluno, apenas o nome, o que gerou o esforço de mapear todos para uma coluna separada com o respectivo sexo, baseado no nome disponibilizado nos dados. Uma dificuldade encontrada foi a manipulação das planilhas de microdados de matrícula, pois possuíam um número muito grande de linhas e dados, o que fazia com que a o excel, plataforma utilizada para realizar o pré-processamento, parasse de funcionar.

Após esse processo, para uma melhor organização dos possíveis cenários que poderiam ser extraídos dos dados abertos, criou-se um *board* na plataforma de colaboração e organização de projetos Miro<sup>3</sup>, mapeando todas as colunas de todos os arquivos para o respectivo resultado, onde cada cor de *post-it* representava a planilha que aquela coluna pertencia (Figura 2).



Figura 2. Separação das planilhas por cor

<sup>3</sup>[https://miro.com/app/board/o9J\\_lu1jo88=?invite\\_link\\_id=725996538726](https://miro.com/app/board/o9J_lu1jo88=?invite_link_id=725996538726)

### 5.2.2.1 Relação de mulheres formadas

O primeiro cenário mapeado foi pensado para responder a pergunta de pesquisa: descobrir a relação de mulheres formadas na área de TI, mas também sendo possível analisar a relação dos cursos gerais. Para isso, foram selecionadas as colunas sexo e curso de todas as planilhas referentes à emissão de diplomas (Figura 3).

Relação de mulheres formadas ✓



**Figura 3. Relação de mulheres formadas, utilizando as colunas sexo e nome de curso das planilhas de conclusão**

### 5.2.2.2 Ingresso vs. conclusão

O segundo cenário mapeado tinha como objetivo gerar um comparativo do número de alunos ingressantes nos cursos com o número de alunos concluintes, tanto por cursos gerais, como focado em cursos na área de TI. Para isso, as colunas de sexo e curso das planilhas de matrículas e emissão de diplomas foram utilizadas (Figura 4).



**Figura 4. Matrículas vs. conclusões de alunos nos cursos superiores, utilizando as colunas sexo e nome do curso das planilhas de matrícula e conclusão.**

### 5.2.2.3 Quantidade de alunos que utilizam bolsa

O terceiro cenário buscava levantar a quantidade de alunos, por sexo, que utilizavam bolsa PROUNI e quais os cursos com maiores taxas de bolsistas. Para isso, as colunas tipo\_bolsa, sexo\_beneficiario, modalidade\_ensino\_bolsa e nome\_curso\_bolsa de todas as planilhas do PROUNI foram utilizadas (Figura 5).



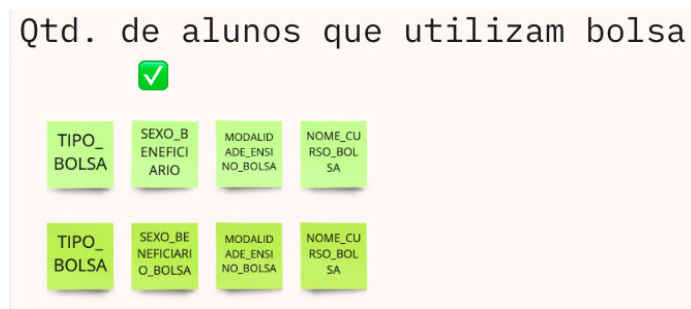


Figura 5. Quantidade de alunos que se beneficiaram de bolsa na graduação

### 5.2.2.4 Informações sociais dos portadores de bolsa

O quarto cenário continuou mapeando dados referente às bolsas PROUNI, porém buscando informações sociais dos portadores de bolsa. As informações selecionadas foram raça, região e se o aluno é ou não deficiente físico. Para isso, as colunas `raca_beneficiario`, `beneficiario_deficiente_fisico` e `regiao_beneficiario` foram utilizadas (Figura 6).

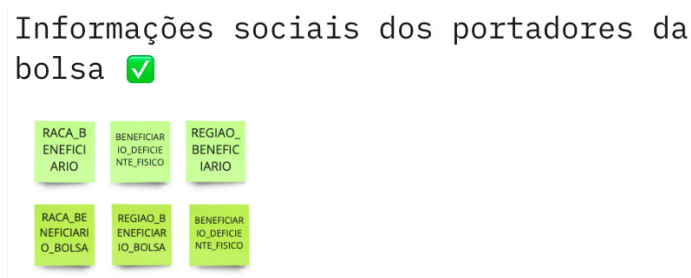


Figura 6. Informações sociais dos beneficiados de bolsa na graduação

### 5.2.2.5 Duração vs. permanência no curso

O quinto cenário buscava realizar uma comparação da média de permanência nos cursos, por sexo do aluno. Para isso, foi calculada a média de permanência geral, utilizando as datas de início e conclusão das planilhas de conclusão que possuíam as duas informações (Figura 7).

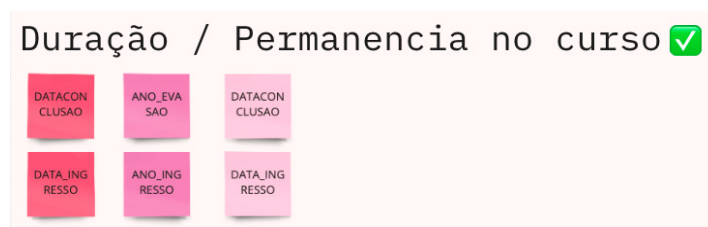
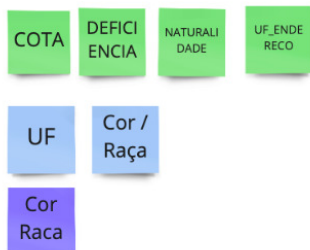


Figura 7. Duração vs. permanência no curso de TI, utilizando as colunas de ano de conclusão e ano de ingresso das planilhas de conclusão.

### 5.2.2.6 Informações sociais matrículas

O sexto cenário, assim como o quarto, realizou um levantamento das informações sociais dos alunos matriculados, numa perspectiva geral, ou seja, portador ou não de bolsa. Para isso, as colunas cota, deficiência, naturalidade, uf e cor e raça foram utilizadas (Figura 8).

## Informações sociais matrículas ✔



**Figura 8. Informações sociais matrículas, utilizando colunas de cota, deficiência, naturalidade de localização das planilhas de matrículas.**

### 5.2.3 Formatação

Para a formatação, utilizou-se *Python* e a biblioteca *pandas*, para formatar e padronizar os nomes das colunas de dados similares. A Figura 9 mostra trecho do código com um exemplo da padronização feita.

```
df_ingressos_sisu = df_ingressos_sisu[['COTA', 'SEXO', 'NATURALIDADE', 'UF_ENDERECO', 'DEFICIENCIA']]
#df_ingressos_sisu.columns = ['SEXO', 'NOME_CURSO', 'DATA_INGRESSO'] #Padronizando colunas
```

**Figura 9. Formatação dos nomes de colunas**

Foi também necessário eliminar todas as linhas duplicadas, em branco e substituir os caracteres especiais contidos neles, para que na etapa de mineração, nenhum problema ou inconsistência ocorresse (Figura 10). Realizou-se a substituição de todos os caracteres especiais existentes, utilizando o método *replace*, assim como a padronização de alguns termos, como transformar todos os estados brasileiros que estavam em maiúsculo em minúsculo.

```
df_dados_prouni['SEXO_BENEFICIARIO_BOLSA'] = df_dados_prouni['SEXO_BENEFICIARIO_BOLSA'].str.replace('Feminino','F')
df_dados_prouni['SEXO_BENEFICIARIO_BOLSA'] = df_dados_prouni['SEXO_BENEFICIARIO_BOLSA'].str.replace('Masculino','M')
df_dados_prouni['NOME_CURSO_BOLSA'] = df_dados_prouni['NOME_CURSO_BOLSA'].str.replace('ÃE','ã')
df_dados_prouni['NOME_CURSO_BOLSA'] = df_dados_prouni['NOME_CURSO_BOLSA'].str.replace('ÃS','ã')
df_dados_prouni['NOME_CURSO_BOLSA'] = df_dados_prouni['NOME_CURSO_BOLSA'].str.replace('Ãi','ã')
df_dados_prouni['NOME_CURSO_BOLSA'] = df_dados_prouni['NOME_CURSO_BOLSA'].str.replace('Ãa','ã')
df_dados_prouni['NOME_CURSO_BOLSA'] = df_dados_prouni['NOME_CURSO_BOLSA'].str.replace('Ã','ã')
df_dados_prouni['RACA_BENEFICIARIO_BOLSA'] = df_dados_prouni['RACA_BENEFICIARIO_BOLSA'].str.replace('A','ã')
df_dados_prouni['RACA_BENEFICIARIO_BOLSA'] = df_dados_prouni['RACA_BENEFICIARIO_BOLSA'].str.replace('ÃE','ã')
df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'] = df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'].str.replace('N','Não deficiente')
df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'] = df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'].str.replace('Não deficiente','N')
df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'] = df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'].str.replace('SIM','Deficiente')
df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'] = df_dados_prouni['BENEFICIARIO_DEFICIENTE_FISICO'].str.replace('S','Deficiente')
df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'] = df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'].str.replace('CENTRO-OESTE','Centro-Oeste')
df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'] = df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'].str.replace('NORDESTE','Nordeste')
df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'] = df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'].str.replace('NORTE','Norte')
df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'] = df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'].str.replace('SUDESTE','Sudeste')
df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'] = df_dados_prouni['REGIAO_BENEFICIARIO_BOLSA'].str.replace('SUL','Sul')
```

**Figura 10. Formatação dos caracteres especiais**

### 5.2.3 Mineração de dados

Para realizar a mineração de dados, a fim de encontrar padrões ou conhecimentos, foram gerados scripts *Python*, disponíveis no Github<sup>4</sup>, que englobaram tanto o processo de leitura, como o de manipulação e exibição dos dados abertos previamente selecionados.

Os scripts se dividiram em 4 arquivos diferentes:

1. **informacoes\_sociais\_matriculas.ipynb**: responsável por gerar conhecimentos relacionados às informações sociais de todas as matrículas disponíveis, como região, deficiência e raça;
2. **ingresso\_vs\_conclusao.ipynb**: responsável por gerar conhecimentos relacionados ao ingresso e conclusão dos alunos em cursos gerais, assim como nos cursos de TI;
3. **alunos\_utilizam\_bolsa\_prouni.ipynb**: responsável por gerar conhecimentos relacionados aos alunos que usam bolsas do PROUNI;
4. **relacao\_formados.ipynb**: responsável por gerar conhecimentos relacionados à formação dos alunos nos cursos superiores.

Em todos os scripts, o primeiro passo a ser realizado foi a leitura de todos os *datasets* por categoria, como previamente definido na Seção 5.2.2. Após realizada a leitura, era necessário padronizar os nomes das colunas, para então juntar todos os dados obtidos em um único *dataset*, para que pudesse ser reutilizado e manipulado com mais facilidade na próxima etapa. Uma vez que a junção foi feita, scripts individuais foram gerados, buscando responder os tópicos mapeados no pré-processamento.

## 6. Resultados

### 6.1 Mulheres em TI

Para responder a pergunta de pesquisa, neste cenário, mapeado na Seção 5.2.2.1, procurou-se por palavras chaves, como “computação”, “sistemas”, “informática”, “informação” e “desenvolvimento”, conforme ilustrado na Figura 11, dentro dos *datasets* de emissão de diplomas ou conclusão do curso. Isso foi feito conforme demonstrado da linha 11 à 15, onde o método *contains()*, que realiza uma procura dentro de uma array por algum valor específico, ficou responsável por encontrar os termos selecionados, utilizando o operador “ou”, representado por “|”, dentro do *dataset*. Também foi realizada a separação de dados pela coluna “SEXO” das planilhas de conclusão, através do método *loc()*, conforme ilustrado da linha 5 à 9.

Para responder a Hipótese 2 (H2), obteve-se o cenário mapeado na Figura 12. Para o curso de Ciência da Computação, dos 205 alunos matriculados, 169 eram homens e 36 mulheres. Nos cursos de Computação, o número permaneceu o mesmo: dos 10 alunos matriculados, 7 eram homens e 3 mulheres. Assim também aconteceu com o curso de Tecnologia em Sistemas de Informação, resultando nos mesmos dados. No curso de Engenharia de Computação, dos 100 alunos matriculados, 87 eram homens e 13 mulheres. No curso de Gestão da Informação, dos 55 alunos matriculados, 40 eram homens e 15 mulheres. No curso de Sistemas de Informação, dos 186 alunos

---

<sup>4</sup> <https://github.com/melarenina/tcc>

matriculados, 147 eram homens e 39 mulheres. Neste caso, a proporção de procura por cursos de TI foi de aproximadamente 81,7% masculina e 18,3% feminina.

```

1 # Grafico: Conclusão área de TI
2
3 concluintes_by_sexo = dados_conclusao.groupby(['SEXO', 'NOME_CURSO']).count().reset_index()
4
5 conclusao_feminina = concluintes_by_sexo.loc[concluintes_by_sexo['SEXO'] == 'F'].
6 groupby(['NOME_CURSO', 'SEXO', 'DATA_CONCLUSAO']).count().reset_index()
7
8 conclusao_masculina = concluintes_by_sexo.loc[concluintes_by_sexo['SEXO'] == 'M'].
9 groupby(['NOME_CURSO', 'SEXO', 'DATA_CONCLUSAO']).count().reset_index()
10
11 conclusao_feminina_ti = conclusao_feminina[conclusao_feminina['NOME_CURSO'].str.
12 contains('SISTEMAS|INFORMÁTICA|INFORMATICA|COMPUTAÇÃO|INFORMACAO|INFORMAÇÃO|DESENVOLVIMENTO')]
13
14 conclusao_masculina_ti = conclusao_masculina[conclusao_masculina['NOME_CURSO'].str.
15 contains('SISTEMAS|INFORMÁTICA|INFORMATICA|COMPUTAÇÃO|INFORMACAO|INFORMAÇÃO|DESENVOLVIMENTO')]
16
17 print(conclusao_feminina_ti)
18 print(conclusao_masculina_ti)

```

Figura 11. Script sobre conclusão dos alunos na área de TI por sexo

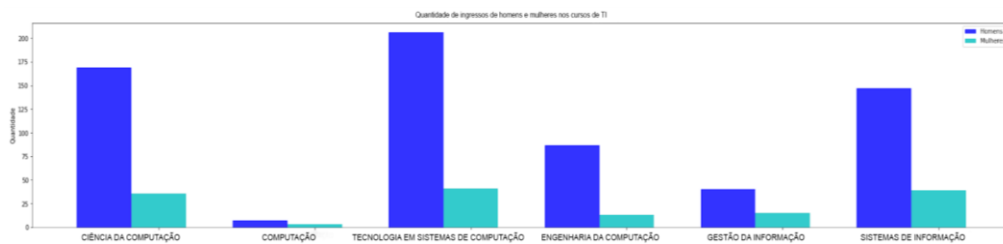


Figura 12. Quantidade de procura de homens e mulheres por curso de TI

Para a Hipótese 1, o objetivo era analisar esses dados pelas modalidades de ensino de Bacharel, Tecnólogo e Engenharia. Porém, nos dados de conclusão dos cursos, não foram encontrados dados relacionados à cursos de Engenharia. Então, ao mapear apenas Bacharel e Tecnólogo, pode-se observar que as mulheres possuem uma preferência muito maior pela modalidade de bacharelado, sendo 1.919 mulheres concluintes em cursos de bacharelado e 107 em cursos tecnólogos (Figura 13).

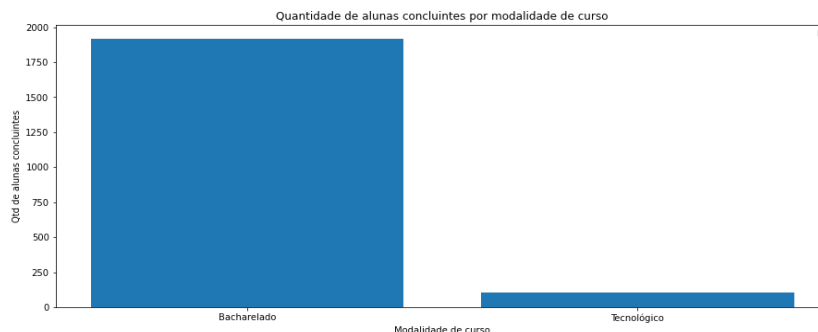
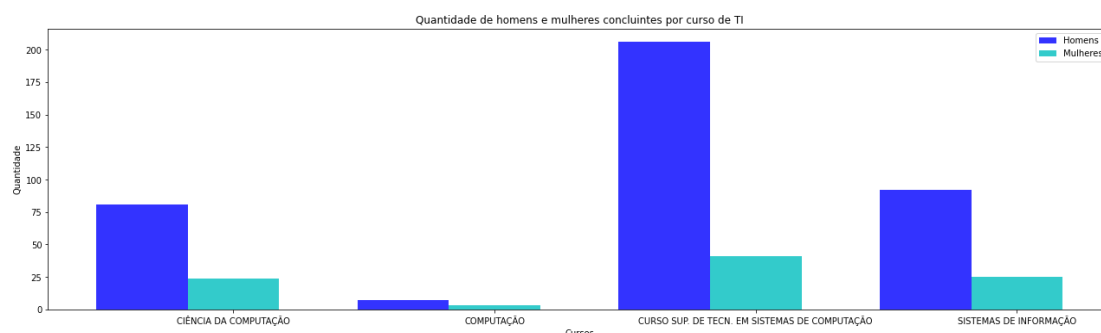


Figura 13. Quantidade de alunas concluintes por modalidade de curso

Em relação aos dados sobre a conclusão de cursos na área de TI, no curso de Ciência da Computação, dos 105 alunos concluintes, 81 são homens e 24 mulheres. Em cursos

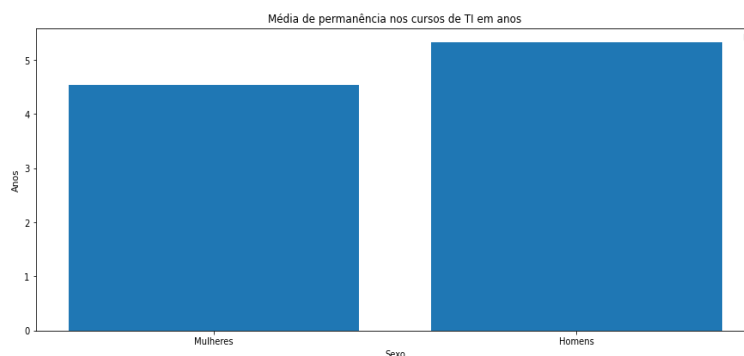
nomeados apenas como “Computação”, dos 10 alunos concluintes, 7 eram homens e 3 mulheres. No curso de Tecnologia em Sistemas de Computação, dos 247 alunos concluintes, 206 eram homens e 42 mulheres. No curso de Sistemas de Informação, haviam 117 alunos concluintes, onde 92 eram homens e 25 mulheres. A Figura 14 apresenta essas informações.



**Figura 14. Quantidade de homens e mulheres concluintes por curso de TI**

Esses dados foram cruciais para confirmar a Hipótese 1 (H1), mostrando que a quantidade de mulheres formadas nos cursos de TI é sim bem menor quando se comparada à de homens. É possível observar que a proporção da conclusão tem uma taxa de aproximadamente 81% dos alunos concluintes sendo homens e 19% mulheres.

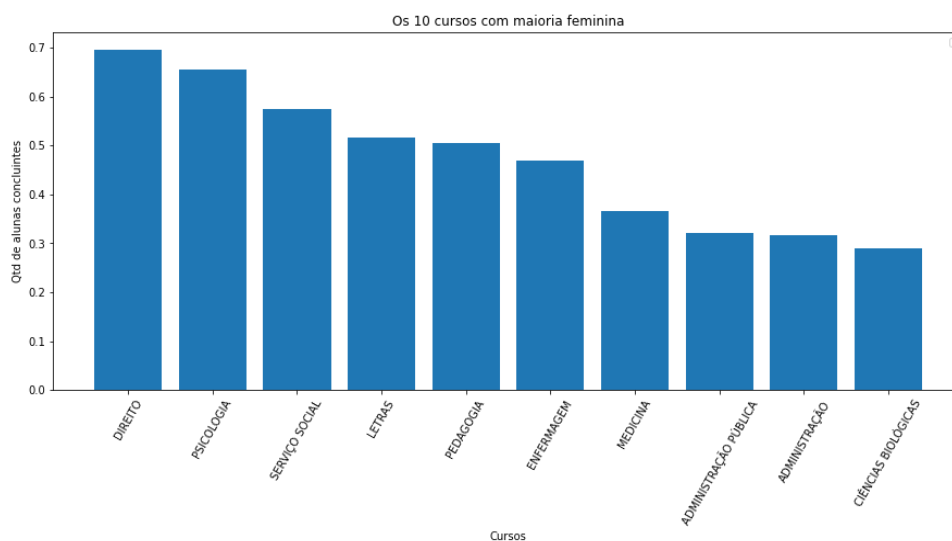
Também foi possível mapear a permanência de homens e mulheres nos cursos da área de TI. A média masculina é de 5,32 anos, enquanto a feminina é de 4,55, considerando apenas os alunos concluintes (Figura 15).



**Figura 15. Média de permanência nos cursos de TI em anos**

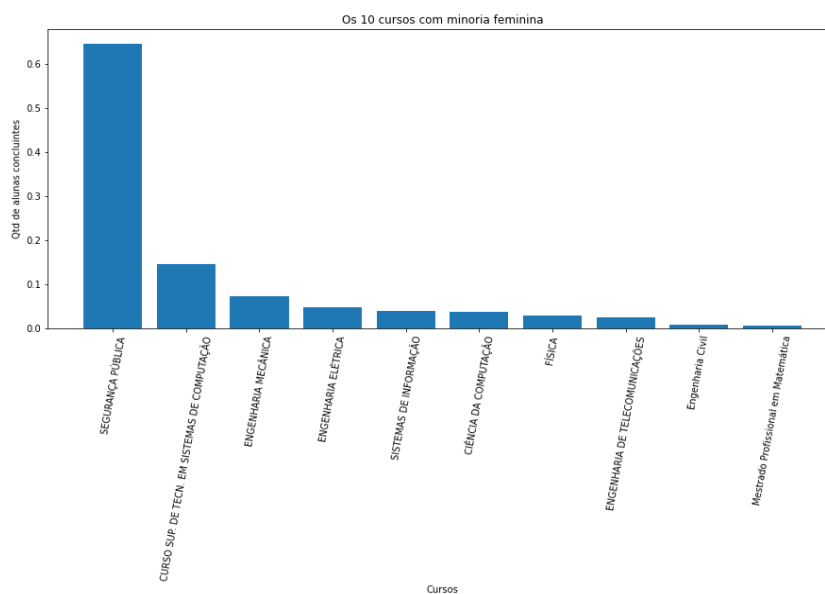
## 6.2 Cursos no geral

Com tantos dados disponíveis, também foi realizada uma análise em cima de todos os cursos disponíveis nos *datasets*. Ao realizar o levantamento dos 10 cursos com mais mulheres, respondendo à Hipótese 3 (H3), é possível observar que são cursos majoritariamente das áreas de humanas e ciências biológicas (Figura 16).



**Figura 16. Os 10 cursos com maioria feminina**

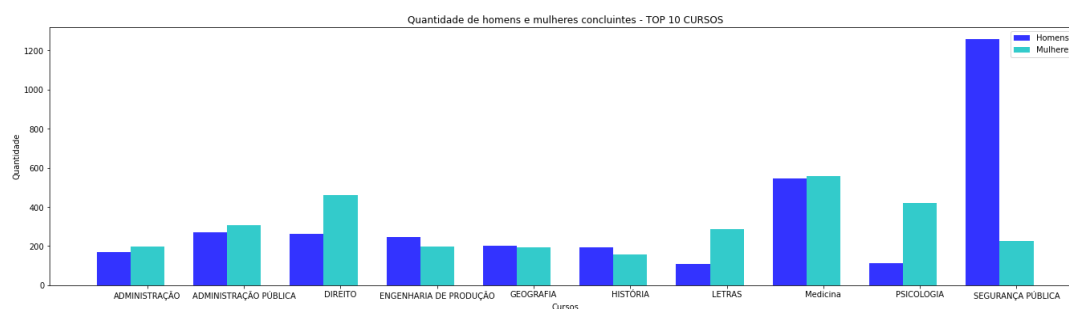
Já os cursos com minoria feminina são predominantemente da área de exatas, onde dos 10 cursos levantados, 3 são relacionados com TI (Figura 17).



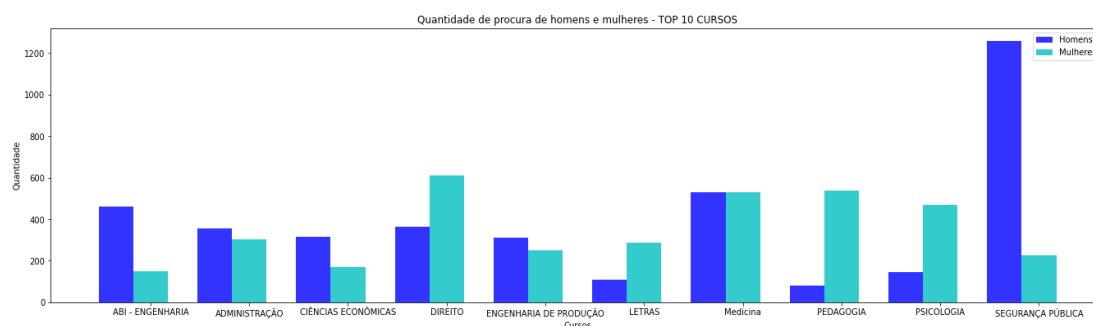
**Figura 17. Os 10 cursos com minoria feminina**

No levantamento dos alunos concluintes no top 10 cursos gerais, pode-se observar que 6 desses cursos possuem mais mulheres e 4 possuem mais homens (Figura 18).

Já no levantamento de alunos matriculados no top 10 cursos gerais, 5 cursos possuem uma taxa maior de homens matriculados, 3 possuem uma taxa maior de mulheres e a Medicina possui a mesma taxa para ambos (Figura 19).

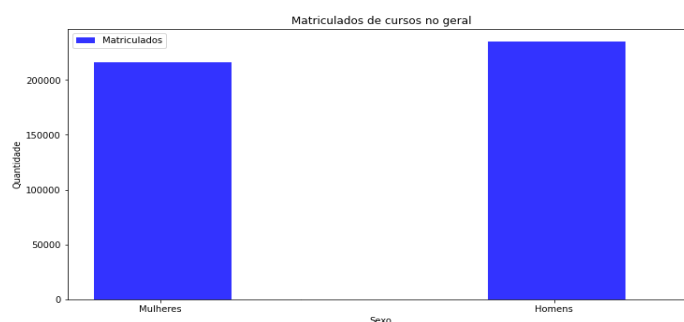


**Figura 18. Quantidade de homens e mulheres concluintes - TOP 10 CURSOS**



**Figura 19. Quantidade de procura de homens e mulheres - TOP 10 CURSOS**

Pode-se observar também, em linhas gerais, que mesmo os homens possuindo uma taxa maior de matrículas em cursos quando comparado a de mulheres, com um total de 237.249 e 218.597 respectivamente, conforme ilustrado na Figura 20, eles ainda assim possuem um número de conclusão menor, sendo 6.874 homens concluintes e 8.873 mulheres, como pode ser observado na Figura 21, comprovando a Hipótese 4 (H4). Ou seja, as mulheres ingressam em menos cursos no geral, mas ainda assim se formam mais do que os homens.

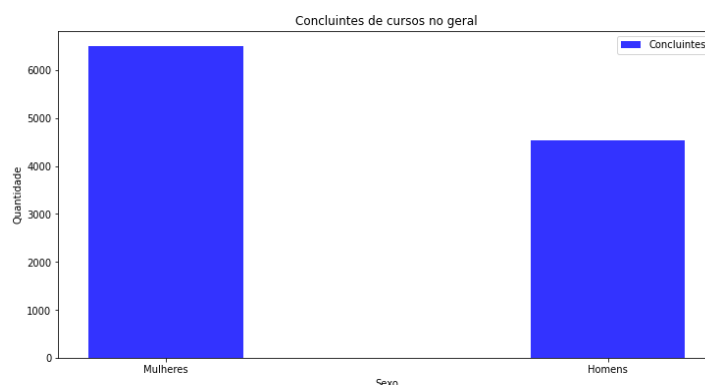


**Figura 20. Quantidade de alunos matriculados em cursos no geral**

### 6.3 Informações sociais

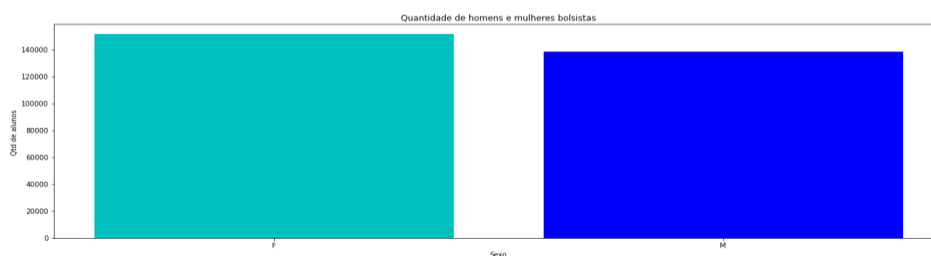
Foi possível também realizar a análise de informações sociais dos estudantes matriculados e portadores de bolsas do PROUNI. A princípio, a intenção era de comparar esses dados de bolsa com os dados de matrículas e conclusão de curso, porém, os dados

disponíveis sobre as bolsas atribuídas datam até 2018, enquanto os dados de matrículas e conclusões estão disponíveis a partir de 2019.



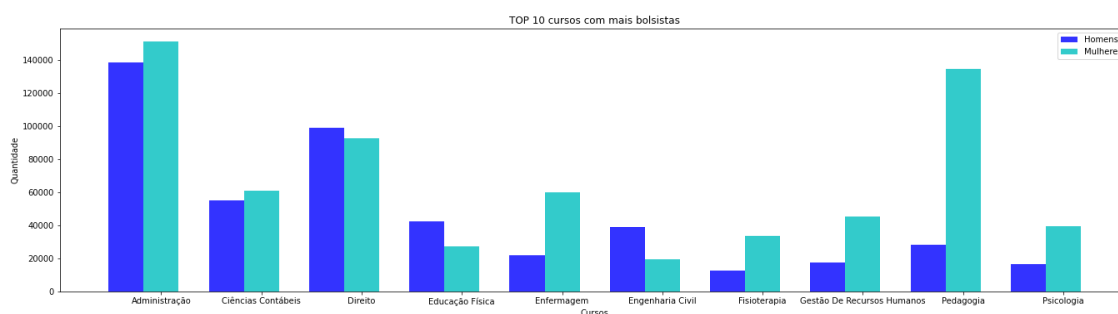
**Figura 21. Quantidade de alunos concluintes em cursos no geral**

Pode-se observar que o número de mulheres bolsistas é maior do que o de homens, sendo 151.428 mulheres e 138.361 homens (Figura 22).



**Figura 22. Quantidade de homens e mulheres bolsistas**

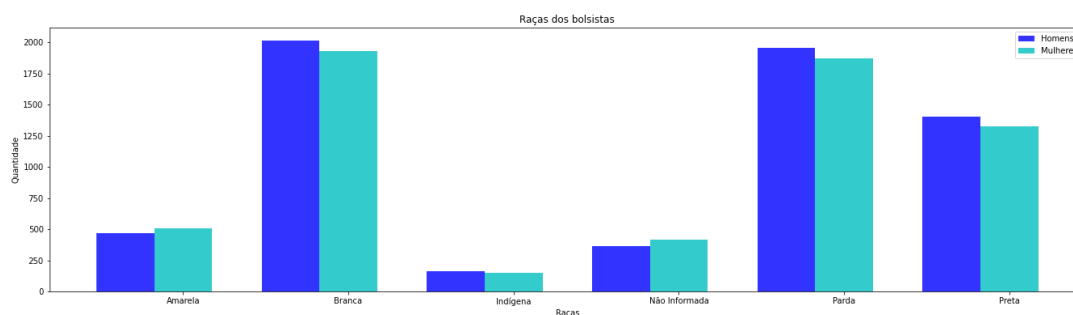
Dentre os 10 cursos com mais bolsistas, as mulheres também continuam com uma taxa maior quando comparada a dos homens, sendo 7 cursos com maioria feminina (Figura 23).



**Figura 23. Cursos superiores brasileiros com as maiores quantidades de bolsistas.**

Também foi possível observar a raça dos alunos portadores de bolsa, sendo as raças branca e parda as com maiores taxas (Figura 24).





**Figura 24. Raças dos bolsistas**

## 7. Conclusão

A participação feminina no mercado de trabalho, quando se trata de áreas de exatas, ainda encontra-se baixa quando comparada à masculina. É possível observar diversas empresas incentivando a participação de mulheres, criando campanhas de inclusão no ambiente executivo e promovendo discussões sobre o assunto.

Este artigo, por meio da análise de dados abertos do governo brasileiro sobre matrículas e conclusões nos cursos superiores, teve o objetivo de descobrir a participação efetiva de mulheres nos cursos de TI, nas diferentes modalidades de ensino - bacharel e tecnólogo, e também analisar quais os cursos e áreas que, atualmente, são de maior interesse para o público feminino.

A partir dos dados obtidos, foi possível observar que, na área de TI, a conclusão de mulheres é realmente menor do que a de homens. A H1 trouxe uma proporção de conclusão masculina nesta área de 81%, enquanto a feminina de apenas 19%.

Já a H2, trazendo uma possível justificativa para os valores obtidos na H1, mostrou que o interesse feminino pela área também não é grande. A proporção de ingresso em cursos na área de Tecnologia da Informação é de 81,7% masculina, enquanto 18,3% é feminina.

Em relação a H3, foi possível observar que o interesse feminino pelas áreas de Humanas e/ou Biológicas é maior quando se comparado às áreas de exatas, visto que, quando filtrado os top 10 cursos com maioria feminina, não havia um curso sequer de exatas. Isso também é embasado pela pesquisa realizada no trabalho correlato “Análise da participação das mulheres na ciência: um estudo de caso da área de Ciências Exatas e da Terra no Brasil” (TUESTA; DIGIAMPIETRI; DELGADO; MARTINS, 2018), o qual teve como um dos resultados de pesquisa um número muito maior de homens em quase todas as subáreas de Ciências Exatas e da Terra.

Considerando a H4, observou-se que mesmo as mulheres ingressando menos que os homens em cursos superiores no geral, elas ainda sim possuem um número maior de conclusão, sendo 8.873 mulheres concluintes e 6.874 homens.

Também foi possível obter dados relacionados às bolsas PROUNI concedidas para os cursos superiores. Estes dados mostraram que o número de mulheres bolsistas, até o ano de 2018, era maior do que o de homens.

Para a realização desta pesquisa, foram aplicados os conhecimentos obtidos nas disciplinas:

- Segurança da informação: utilização apenas de dados abertos, para não violar nenhum aspecto da lei LGPD;
- Metodologia de Pesquisa Científica e Tecnológica: identificar as etapas do processo de pesquisa, como a criação de um cronograma mapeando cada etapa necessária para a conclusão do trabalho, e a elaboração dos relatórios técnico-científicos.
- Algoritmos e Linguagem de Programação: implementação dos scripts *Python* para leitura e geração de dados.
- Gestão de projetos: para organizar o cronograma e informações de pesquisa.
- Engenharia de software: na utilização da ferramenta Github para gerenciamento de configuração.
- Banco de dados e Estatística: de maneira indireta, ajudaram a compreender o mecanismo utilizado pela mineração de dados.

Também obteve-se conhecimento relacionado ao processo KDD e análise de dados, não aplicados anteriormente no curso.

Conclui-se com este trabalho de pesquisa que, mesmo com discussões sobre o assunto, a participação feminina em TI e nas áreas de exatas dos cursos superiores ainda permanece baixa, tanto no quesito procura, como conclusão. Ainda é necessária uma atenção para este tópico, pois ele reflete também no desenvolvimento de tecnologias e soluções utilizadas por toda a sociedade, assim como no bem estar pessoal de mulheres que por muitas vezes possuem o desejo, mas não a oportunidade ou liberdade de atuar nestas áreas.

## 8. Trabalhos Futuros

Por ser um tema de tamanha importância e com diversas áreas de exploração, este trabalho gera dados e questões que podem ser abordadas em trabalhos futuros. Alguns deles sendo:

- Por que as mulheres possuem um interesse maior pelas áreas de Humanas e Ciências Biológicas?
- Porque as mulheres não possuem tanto interesse pela área de TI, visto que no início, eram maioria nessa área? Seria isso talvez por não ter havido um incentivo desde a infância nas gerações passadas? Ou então, por ainda ser uma área predominantemente masculina, não haver tido oportunidades até então para mulheres atuarem efetivamente na área, despertando então o desinteresse por ela?
- Por que as mulheres utilizam mais bolsas PROUNI do que homens? Será o motivo disso, novamente, a falta de incentivo feminino para a realização de cursos superiores? Ou a responsabilidade de cuidar de uma casa e família, junto ao desejo de realizar um curso, mas não tendo o recurso para tal, por não trabalhar?

## 8. Agradecimentos

À Deus, primeiramente, por ter me dado força, perseverança, saúde em meio à uma época tão conturbada e a oportunidade de realizar um curso em uma universidade pública e de qualidade.

À minha família, minha mãe Ilma, meu pai Vlademir, meu irmão Lucas, minha irmã Amanda e meu cunhado Gabriel, por não me deixarem desistir e acreditarem em mim, principalmente quando eu não acreditava.

Aos meus amigos, por sempre me motivarem, acreditarem em mim e me ajudarem a refrescar o pensamento quando precisei. Em especial, ao Renan Lange, pelas horas em ligação me ensinando *Python* e me ajudando com a parte principal do trabalho.

Aos meus colegas de trabalho pelo apoio e ajuda com questões técnicas. Em especial, o Leandro Kimura, meu amigo e mentor, pela ajuda no primeiro rascunho do tema e pelo tanto que me ensinou até aqui.

Ao professor Carlos Santos, por toda a ajuda com o processo KDD, análise de dados e desenvolvimento dos scripts, o meu muito obrigada!

À todos os professores que contribuíram com o conhecimento para desenvolver este trabalho.

À Dani, minha paciente orientadora, que me motivou desde o início, me acalmou, me mostrou que sou capaz e me faz ficar feliz de dividir a profissão com uma mulher com valores como os dela.

E a todos que direta ou indiretamente fizeram parte da minha formação.

## 9. Referências

- BICHARA, Ilka. Brincadeiras na infância influenciam a vida adulta. 2012. Disponível em: <http://www.cienciaecultura.ufba.br/agenciadenoticias/noticias/brincadeiras-na-infancia-influenciam-a-vida-adulta/>. Acesso em: 27 jun. 2021.
- FOGUEL, Débora. AS MULHERES E O NOBEL... O NOBEL E AS MULHERES.... 2020. Disponível em: <http://www.abc.org.br/2020/10/11/as-mulheres-e-o-nobel-o-nobel-e-as-mulheres/>. Acesso em: 27 jun. 2021.
- SIMONITE, Tom. AI Is the Future—But Where Are the Women? 2018. Disponível em: <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>. Acesso em: 27 jun. 2021.
- BRASIL. Censo da Educação Superior. Diretoria de Estatísticas Educacionais. RESUMO TÉCNICO CENSO DA EDUCAÇÃO SUPERIOR 2016. 2016. Disponível em: [http://www.abc.org.br/wp-content/uploads/2019/03/resumo\\_tecnico\\_censo\\_da\\_educacao\\_superior\\_2016.pdf](http://www.abc.org.br/wp-content/uploads/2019/03/resumo_tecnico_censo_da_educacao_superior_2016.pdf). Acesso em: 27 jun. 2021.
- OPEN KNOWLEDGE. What is open? Disponível em: <https://okfn.org/opendata/>. Acesso em: 06 jul. 2021.
- TUESTA, Esteban Fernandez; DIGIAMPIETRI, Luciano Antonio; DELGADO, Karina Valdivia; MARTINS, Nathália Ferraz Alonso. Análise da participação das mulheres na ciência: um estudo de caso da área de Ciências Exatas e da Terra no Brasil (2019). 2018. 62 f. Tese (Doutorado) - Curso de Comunicação e Informação, Universidade de São Paulo, São Paulo, 2018.
- CURSINO, Ariana R.; MARTINEZ, Juliana F. P.. Análise Estatística Descritiva e Regressão da Inserção das Mulheres nos Cursos de TI nos Anos de 2009 a 2018. 2021. 10 f. TCC (Graduação), Faculdade de Tecnologia de São José dos Campos - Professor Jessen Vidal, São José dos Campos, 2021. Disponível em: <https://sol.sbc.org.br/index.php/wit/article/view/15838/15679>. Acesso em: 23 out. 2021.

- MARQUES, Daniela; SILVA, André Constantino da; GUEDES, Gustavo Bartz; SANTOS JUNIOR, Carlos Roberto dos. Desempenho Acadêmico e o Ingresso no Curso Superior: uma análise das estudantes ingressantes entre 2016 a 2020 do curso Técnico em Informática Integrado ao Ensino Médio. 2021. Disponível em: <https://sol.sbc.org.br/index.php/wit/article/view/15841/15682>. Acesso em: 18 jul. 2021.
- WIT 2020 – XV WOMEN IN INFORMATION TECHNOLOGY. 2020. Disponível em: <http://www2.sbc.org.br/csbc2020/wit-2020-xv-women-in-information-technology/>. Acesso em: 25 jul. 2021.
- HIRATA, Helena. Tecnologia, formação profissional e relações de gênero no trabalho. Educação e Tecnologia – Periódico Técnico Científico dos Programas de Pós-Graduação em Tecnologia CEFETs, Curitiba, n. 6, p. 144-156, 2003. Acesso em: 25 jul. 2021.
- CAMARGO, Edson Carpes. Quem disse que não é coisa de menina: provocações acerca das relações de gênero no ensino técnico em agropecuária do IFRS - Campus Bento Gonçalves. Tese (Doutorado em Educação) - Universidade Vale dos Sinos, São Leopoldo, 2014. Acesso em: 25 jul. 2021.
- TEÓFILO, Daniel. KDD – Knowlegde Discovery in Database. 2015. Disponível em: <https://danielteofilo.wordpress.com/2015/02/16/kdd-knowlegde-discovery-in-database/>. Acesso em: 28 set. 2021.
- SANTOS, Carolina Marins. Por que as mulheres “desapareceram” dos cursos de computação?: na década de 1970, cerca de 70% dos alunos do curso de ciências da computação, no ime, eram mulheres; hoje, 15%. Na década de 1970, cerca de 70% dos alunos do curso de Ciências da Computação, no IME, eram mulheres; hoje, 15%. 2019. Disponível em: <https://jornal.usp.br/universidade/por-que-as-mulheres-desapareceram-dos-cursos-de-computacao/>. Acesso em: 08 dez. 2021.
- A primeira programadora da história. 2020. Disponível em: <http://www.ime.unicamp.br/~apmat/ada-lovelace/>. Acesso em: 08 dez. 2021.
- GALILEU, Redação. Katie Bouman: conheça a responsável pela primeira imagem de um buraco negro. 2019. Disponível em: <https://revistagalileu.globo.com/Ciencia/Espaco/noticia/2019/04/katie-bouman-conheca-responsavel-pela-primeira-imagem-de-um-buraco-negro.html>. Acesso em: 08 dez. 2021.
- SANTANA, Rodrigo. Dominando o Pandas: A Biblioteca para Análise de Dados preferida entre os Cientistas de Dados (Parte 1). 2017. Disponível em: <https://minerandodados.com.br/python-para-analise-de-dados/>. Acesso em: 08 dez. 2021.
- KDD and Data Mining. Disponível em: <https://www.datascience-pm.com/kdd-and-data-mining/#:~:text=The%20term%20%E2%80%9Cdata%20mining%E2%80%9D%20is,step%20in%20the%20KDD%20process..> Acesso em: 08 dez. 2021.
- DUHAIME-ROSS, Arielle. Apple promised an expansive health app, so why can't I track menstruation? 2014. Disponível em: <https://www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-app-so-why-cant-i-track>. Acesso em: 13 dez. 2021.

# Documento Digitalizado Público

## Anexo I - Artigo Final

**Assunto:** Anexo I - Artigo Final  
**Assinado por:** Daniela Marques  
**Tipo do Documento:** Projeto  
**Situação:** Finalizado  
**Nível de Acesso:** Público  
**Tipo do Conferência:** Documento Digital

Documento assinado eletronicamente por:

- Daniela Marques, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 13/12/2021 21:46:42.

Este documento foi armazenado no SUAP em 13/12/2021. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifsp.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

**Código Verificador:** 843019

**Código de Autenticação:** ab84830006

