

Impacto da COVID-19 nos Estudantes Universitários no Brasil

Trabalho final da disciplina de Ciência de Dados

Roque Maitino Neto

Ana Cláudia Lima

Marcelo Bressanin

Nilton Ferst

Gabrielly Richene

2023-01-30

1. Considerações iniciais

Este trabalho destina-se à realização de Análise Exploratória de Dados em dados relacionados ao impacto da pandemia de Covid 19 em alunos do Ensino Superior. Tais dados foram obtidos pela realização de levantamento conduzido pelo professor João Pedro Albino entre alunos do ensino superior. Os dados obtidos foram armazenados em planilha Excel e, conforme mencionado, formam base para este trabalho.

2. Análise Exploratória de Dados

2.1 Introdução e conceituação de Análise Exploratória de Dados

Análise exploratória de dados (AED) é um processo de investigação e descrição dos dados, com o objetivo de entender sua estrutura e detectar padrões e tendências. A AED pode ser efetivada de nas modalidades monovariada, bivariada e multivariada. A utilização de gráficos facilita a compreensão dos dados, conforme destacam Escovedo e Koshiyama, o conceito de Data Science (ou Ciência de Dados), [...] refere-se à coleta de dados de várias fontes para fins de análise, com o objetivo de apoiar a tomada de decisões, utilizando geralmente grandes quantidades de dados de forma sistematizada”. (ESCOVEDO e KOSHIYAMA, 2020, p. 3).

A análise exploratória de dados pode ser feita de diversas formas, dentre as quais destacamos a análise monovariada, bivariada e multivariada. Os resultados obtidos com estes diferentes tipos de análises permitem a composição de diferentes tipos de gráficos, tendo em vista facilitar a compreensão dos dados analisados, conforme destacam Escovedo e Koshiyama:

É muito importante compreender bem os seus dados a fim de obter os melhores resultados possíveis nos algoritmos de Machine Learning, aplicados posteriormente. Para aprender mais sobre seus dados de forma mais fácil e intuitiva, é possível usar a visualização de dados, ou seja, a criação de gráficos a partir dos dados brutos. Após entendê-los, você poderá limpar, transformar e apresentar melhor os dados que você possui, e posteriormente, aplicar os algoritmos de Machine Learning mais adequados para eles. Através dos gráficos criados, será possível identificar valores discrepantes, faltantes ou inválidos, além de sugerir a necessidade da aplicação de alguma transformação de dados. Examinar visualmente os relacionamentos entre os atributos pode ajudar a identificar atributos redundantes, por exemplo. (ESCOVEDO e KOSHIYAMA, 2020, p. 88).

Nesse sentido, e no que se refere a estratégias para a visualização de quantidades massivas de informações, vale lembrar com Marquesone que,

[...] Nós, humanos, temos em nosso cérebro uma incrível capacidade para compreender padrões por meio do sentido visual. Esse potencial é tanto que, de toda a informação sensorial que temos, 80% é destinada a esse sentido [...] Por termos esta capacidade, a visualização de dados permite que nós tenhamos uma

compreensão aperfeiçoada do que nos é apresentado, por meio de uma perspectiva totalmente diferente. (...). (MARQUESONE, 2017, p.190).

Assim, nos parece pertinente afirmar que a utilização de técnicas de análise exploratória de dados seja não apenas um recurso importante para o manuseio e a interpretação de grandes conjuntos de informações, a partir de problemas levantados a priori, como também para sua apresentação de forma a facilitar a avaliação das conclusões obtidas.

2.2 Introdução à Análise Univariada de Dados

Análise univariada: é a análise de uma única variável. Essa análise é usada para entender a distribuição e as características de uma única variável, como sua média, desvio padrão, mediana, quartis, entre outros. Uma análise exploratória monovariada dos dados coletados nesta pesquisa do Impacto da Covid-19 na vida dos estudantes das universidades poderia começar comparando a situação dos estudantes entrevistados que tinham acesso à internet com aqueles que não tinham.

2.3 Introdução à Análise Bivariada de Dados

Trata-se da análise de duas variáveis. Essa análise é usada para entender as relações entre as variáveis, como por exemplo, se existe uma relação linear entre elas, ou se uma variável é independente da outra. Um exemplo de análise bivariada seria a comparação dos estudantes que tinham ou não acesso à internet com a capacidade de prosseguir nos estudos.

2.4 Introdução à Análise Multivariada de Dados

Análise multivariada é a análise de mais de duas variáveis. Essa análise é usada para entender as relações entre várias variáveis e detectar padrões e tendências complexas. As técnicas multivariadas incluem análise de componentes principais, análise de cluster, análise discriminante, entre outros. A comparação entre os estudantes que foram vacinados e aqueles que não foram e diversas outras variáveis, a exemplo do nível de ansiedade, a disposição para atividades, a capacidade de prosseguir estudos, a capacidade de socialização, o bem-estar psicológico, a qualidade de vida e o desempenho escolar, é um exemplo de análise multivariada. É possível que os estudantes vacinados tenham menos ansiedade e sejam mais propensos a prosseguir com seus estudos, socializar e ter melhor bem-estar psicológico e qualidade de vida.

Escovedo e Koshiyama explicam que “não existe uma receita pronta para o trabalho de análise exploratória de dados, uma vez que ele dependerá dos dados a serem analisados e dos insights que cada etapa de análise provocará.” (ESCOVEDO e KOSHIYAMA, 2020, p. 99) Os três tipos de análise exploratória de dados podem ser realizados utilizando diferentes ferramentas e técnicas estatísticas, como gráficos, tabelas, estatísticas descritivas, entre outros. O objetivo principal é identificar as características do conjunto de dados e entender como as variáveis podem ser relacionadas entre si.

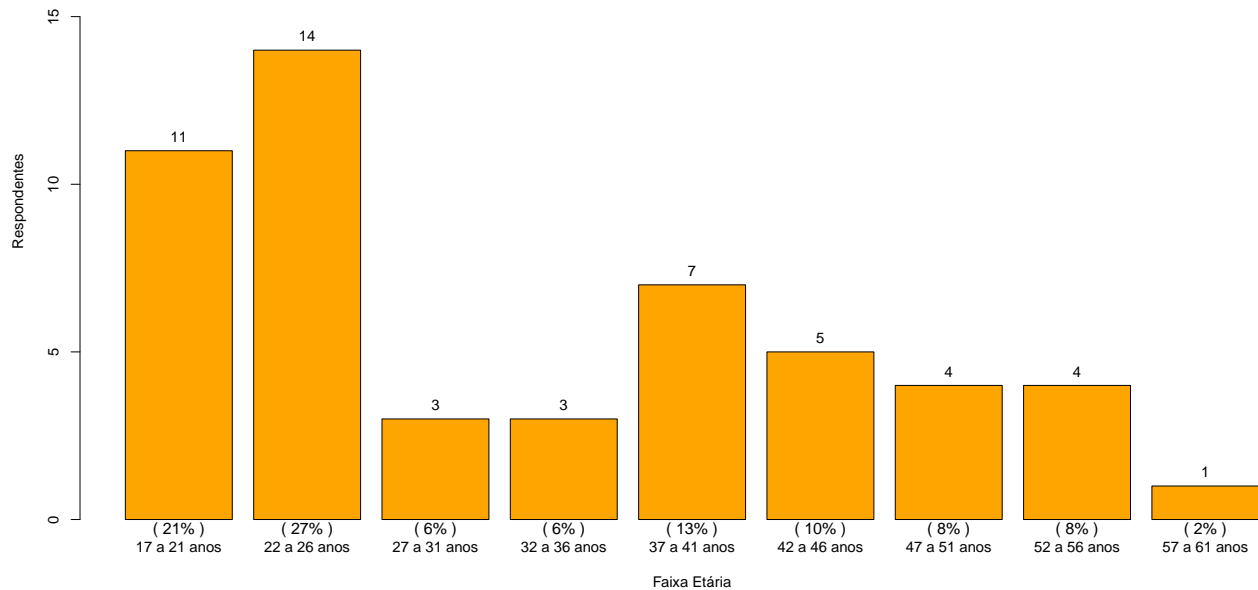
3. Análises elaboradas a partir dos dados disponibilizados

Feitas as considerações teóricas sobre a Análise Exploratória de Dados, o trabalho passa a abordar as relações existentes entre os dados relacionados ao impacto da pandemia de Covid 19 causado em certa comunidade acadêmica. Nesta seção do artigo serão mostrados e discutidos, portanto, os gráficos resultantes da AED realizada.

A amostra utilizada neste artigo é composta por **52** observações que foram coletadas durante os anos de 2020 a 2023.

Como pode ser observado no Gráfico 1, os dados da amostra correspondem à faixa etária de estudantes entre 17 a 61 anos.

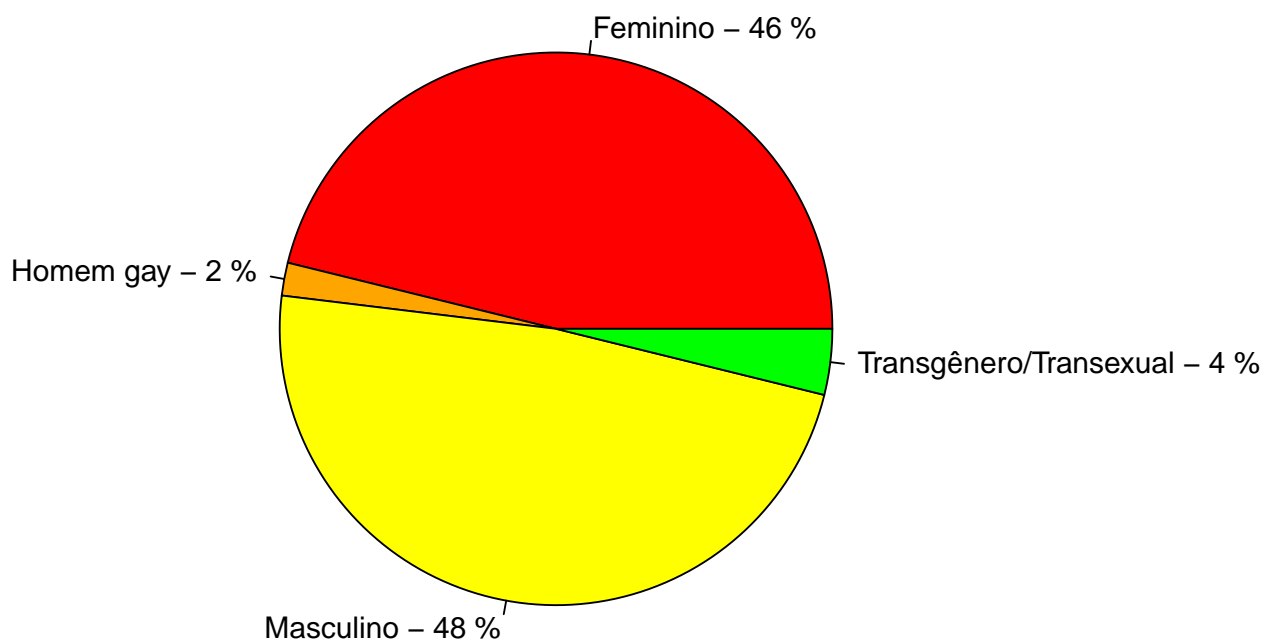
Gráfico 1: Faixa etária dos respondentes



A maior quantidade de estudantes da amostra coletada está na faixa de 22 a 26 anos, com 14 participantes, como mostra o Gráfico 1.

Como forma de complementar a análise dos respondentes, o Gráfico 2 mostra que os gêneros femininos e masculinos apresentaram uma certa paridade, com respondentes femininos representando 46 % da amostra e os masculinos 48 % da amostra.

Gráfico 2: Quantidade de respondentes por gênero

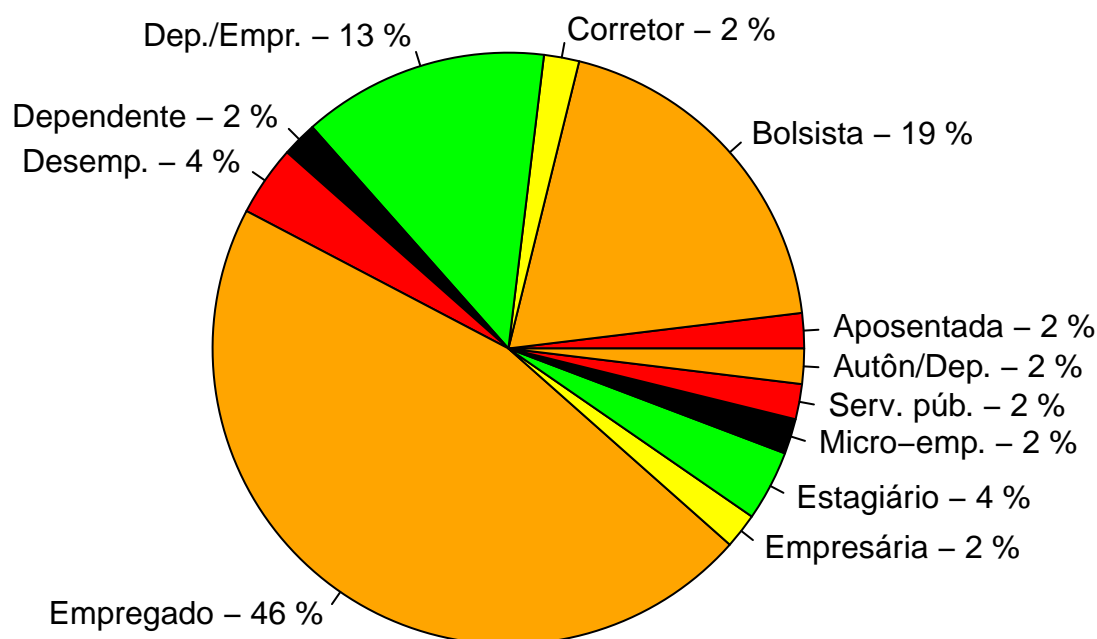


Ainda cabe pontuar que o Gráfico 2 mostra a distribuição de respondentes por gênero.

Para plotar o gráfico utilizou-se uma tabela de frequência de gênero presente na variável *casos_genero* para criar uma representação do tipo “pizza”, usando a biblioteca gráfica *pie()*. As cores do gráfico são definidas como “red”, “orange”, “yellow”, “green”, e os rótulos do gráfico são criados como a combinação do nome do gênero e a porcentagem correspondente de respondentes. Também é calculada a porcentagem de respondentes de cada gênero.

Já o gráfico mostrado no Gráfico 3 representa o status de emprego do grupo de alunos entrevistados. Os dados utilizados para a construção do gráfico são uma tabela com a quantidade de respondentes em cada categoria funcional, que inclui “Aposentado”, “Bolsa”, “Corretor de Imóveis”, “Dependente com vínculo empregatício”, “Dependente”, “Desempregado”, “Empregado”, “Empreendedor”, “Estagiário”, “Microempreendedor”, “Funcionário Público”, “Tatuador Autônomo com ajuda dos pais”.

Gráfico 3: Quantidade de respondentes por situação empregatícia



O Gráfico 3, do tipo pizza, mostra a porcentagem de entrevistados em cada categoria de emprego, com cada segmento representando uma categoria diferente e os percentuais de entrevistados em cada categoria.

Os dados constantes no gráfico demonstram que o maior grupo de respondentes é de empregados, representando 46 % do total de entrevistados, enquanto o segundo maior grupo é o de bolsistas com 19 % do total de respondentes.

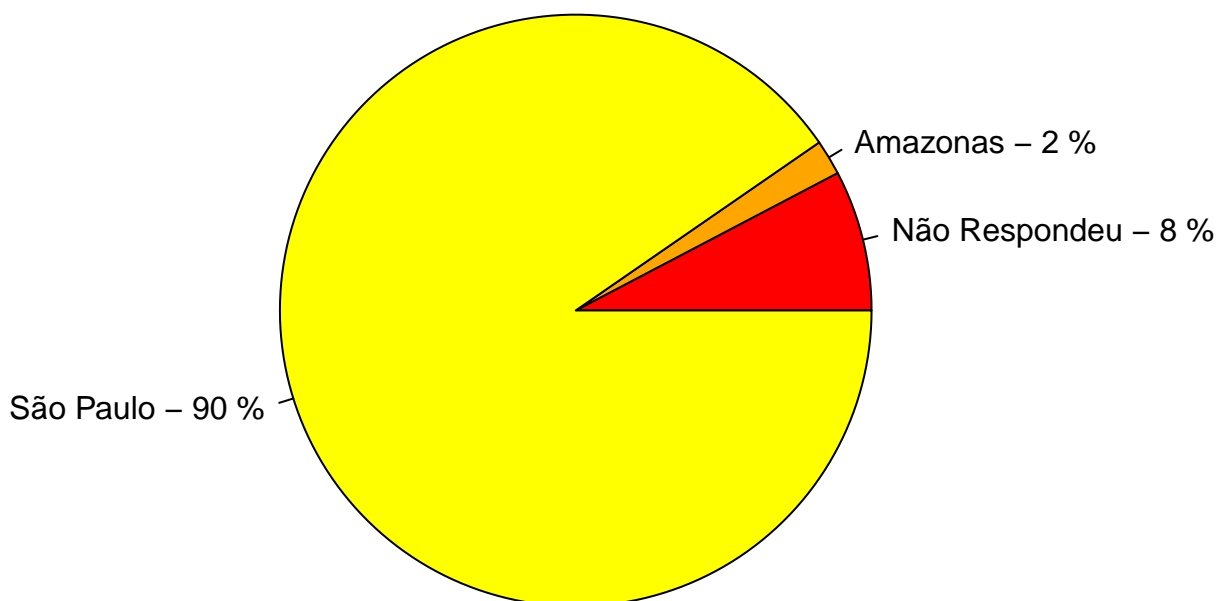
Os menores grupos são os de “Agente Imobiliário”, “Aposentado”, “Autônomo/Dependente dos pais”, “Micro-empresário” e “Corretor” com apenas 2 % dos inquiridos.

No geral, o gráfico fornece uma representação clara e concisa da situação profissional dos entrevistados da pesquisa e permite uma comparação fácil das diferentes categorias.

A próxima figura (Gráfico 4) é uma representação da distribuição dos respondentes da pesquisa por estado de residência.

Os dados usados para criar o gráfico são uma tabela do número de respondentes em cada estado, que inclui “Amazonas”, “São Paulo” e “Sem resposta”. O gráfico é do tipo pizza que mostra a porcentagem de entrevistados em cada estado, com cada segmento representando um estado diferente e indicando a porcentagem de entrevistados naquele estado.

Gráfico 4: Quantidade de respondentes por estado



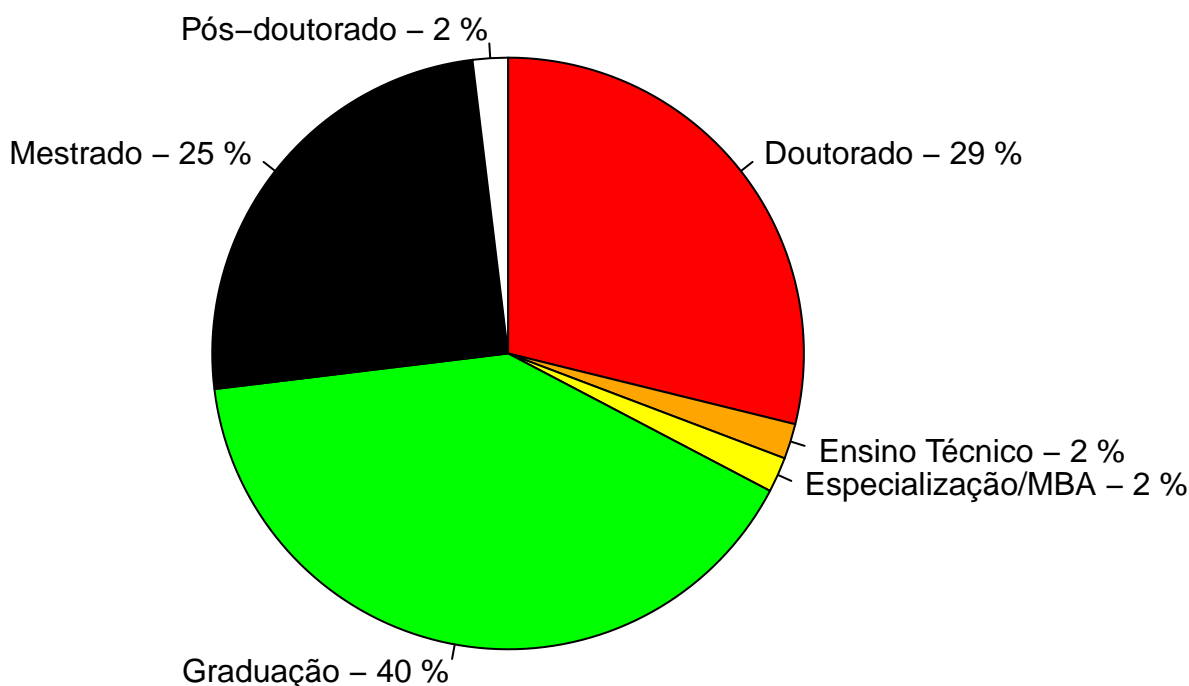
Os dados mostram que o maior grupo de respondentes são os paulistas, respondendo por 90 % do total, enquanto o segundo maior grupo é o “Sem Resposta” com 8 % do participantes. O menor grupo é representado pelo estado do “Amazonas” com apenas 2 % dos entrevistados.

No geral, o Gráfico 4 fornece uma representação clara da distribuição dos entrevistados por estado, mas seria ainda mais informativo se houvesse mais estados representados.

É possível que o tamanho da amostra não foi suficientemente grande para generalizar as informações para todo o país, portanto é importante levar este fato em consideração quando da interpretação dos dados.

O próximo gráfico (Gráfico 5), também no formato pizza, mostra a distribuição dos níveis de ensino dos respondentes.

Gráfico 5: Quantidade de respondentes por nível de ensino



A partir desse gráfico, podemos observar que a maioria dos respondentes (40 %) frequentam curso de graduação, seguido por estudantes de mestrado (25 %) e doutorado (29 %). A menor porcentagem de respondentes pertencem ao ensino técnico (2 %) ou especialização/MBA (2 %) e pós-doutorado (2 %).

4. Discussão dos resultados

É importante notar que essa análise é baseada em uma amostra de dados fornecida, portanto, não pode ser generalizada para a população geral.

A partir dos dados coletados na pesquisa, é possível observar que a pandemia de Covid-19 teve impacto significativo na comunidade acadêmica estudada.

No que diz respeito à *gênero*, a amostra mostrou que houve uma maior representatividade de mulheres (46 %) do que de homens (2 %).

Quanto ao grau de instrução, a maioria dos respondentes cursa ensino superior (40 %). Já em relação a empregabilidade, a maioria dos respondentes (46 %) está empregada ou trabalhando de forma remota.

E por fim, a maioria dos respondentes (27%) pertence à faixa etária entre 22 a 26 anos.

É importante mencionar que esses dados são baseados em uma amostra específica e não refletem necessariamente a realidade de toda a comunidade acadêmica. Além disso, é possível que existam outras variáveis relevantes que não foram incluídas nesta análise exploratória.

O primeiro gráfico apresentado, intitulado “Gráfico 1: Faixa etária dos respondentes”, mostra a distribuição de idade dos universitários que participaram da pesquisa, divididos em faixas etárias. O gráfico mostra que a maioria dos respondentes tem entre 22 e 36 anos.

O segundo gráfico, intitulado “Gráfico 2: Número de respondentes por gênero”, mostra a distribuição de gênero dos universitários participantes da pesquisa. O gráfico mostra que a maioria dos respondentes é do gênero feminino (46 %).

Podemos notar ainda, no Gráfico 3, que a maioria dos entrevistados estão em situação empregatícia regular (46 %), o que pode indicar uma estabilidade econômica no contexto da pandemia.

Já os Gráficos 4 e 5 demonstram a predominância de paulistas entre os entrevistados (2 %), em sua maioria com formação universitária e pós-graduação (mestrado e doutorado), indicando que o público avaliado possui concentração no estado de São Paulo e que demonstra índices educacionais relevantes, quando comparados à estatística do IBGE, que, em 2010, cerca de 11% da população brasileira possuía formação universitária. A partir dos dados coletados é possível observar como a pandemia de Covid-19 afetou a comunidade acadêmica, tendo em vista as diferenças entre gênero, idade, grau de instrução, situação financeira e emprego. Os dados demonstram que 30,77% dos respondentes disseram que seu desempenho escolar diminuiu durante a pandemia, enquanto 26,92% disseram que tiveram evolução no desempenho escolar, ao passo que 38,46% afirmaram que não houve alteração no desempenho escolar e 3,85% não avaliaram ou não souberam se houve alguma alteração.

Ressalta-se que o baixo índice de alteração no desempenho escolar deve ser observado considerando-se o elevado grau de instrução dos respondentes, sendo que a maioria (40%) frequentam cursos de graduação, seguido por mestrado (25%) e doutorado (29%). A menor porcentagem de respondentes vem do ensino técnico (2%) ou especialização/MBA (2%) e pós-doutorado (2%).

A pesquisa demonstrou ainda que a pandemia afetou muito a capacidade de socialização, considerando-se que 38,46% dos respondentes disseram que essa capacidade ficou pior do que antes, 19,23% destacaram que ficou “muito pior que antes”, 36,54% afirmaram que não houve alteração e apenas 5,77% disseram que ficou melhor que antes.

O grande impacto da pandemia entre os respondentes foi no bem-estar psicológico, uma vez que 50% respondeu que ficou pior do que antes, 17,31% disseram que ficou “muito pior que antes”, 23,08% afirmaram que não houve alteração, 3,85% declararam que ficou melhor que antes, 3,85% que ficou “muito melhor do que antes” e 1,92% não avaliaram ou disseram que não sabiam se houve alteração.

5. Considerações finais

Outras diversas correlações poderiam ser feitas com a análise multivariada dos dados coletados. No entanto, a análise exploratória feita neste estudo já demonstra que a pandemia do Covid-19 trouxe grande impacto na saúde mental dos respondentes, dificultando a capacidade de socialização e piorando o bem-estar psicológico. O grande insight que a análise desses dados nos traz é sobre o surgimento da necessidade de criação de uma política pública para melhorar a saúde mental dos estudantes universitários em razão dos problemas decorrentes da pandemia do Covid-19.

6. Referências

ESCOVEDO, Tatiana; KOSHIYAMA, Adriano. Introdução a Data Science: Algoritmos de Machine Learning e métodos de análise. São Paulo: Casa do Código, 2020, 272 p.
MARQUESONE, Rosangela. Big Data: Técnicas e Tecnologias para extração de valor dos dados. São Paulo: Casa do Código, 2017, 258 p.
XIE, Yihui; DERVIEUX, Christophe; RIEDERER, Emily. R Markdown Cookbook. Chapman & Hall/CRC, 07 nov. 2022. Disponível em: <https://bookdown.org/yihui/rmarkdown-cookbook/multi-column.html>. Acesso em: 30 jan. 2023.