

GUIA PRÁTICO PARA A ANÁLISE DE DADOS

Censo da Educação Superior e ENADE



**SOU
CIÊNCIA**



Centro de Estudos Sociedade, Universidade e Ciência

**GUIA PRÁTICO PARA
A ANÁLISE DE DADOS:
Censo da educação superior
e ENADE**

São Paulo, 2023



O SoU_Ciência é um Centro de Estudos e Think Tank sobre Sociedade, Universidade e Ciência. É um Grupo de Pesquisa multidisciplinar cadastrado no CNPq, sediado na Unifesp e composto por uma equipe de pesquisadores de todos os campi e de outras universidades, com histórico de pesquisa, inovação e gestão em Educação Superior e Ciência, Tecnologia e Inovação (CT&I).

Coordenação Geral SoU_Ciência:

Soraya Soubhi Smaili (Unifesp)

Coordenação de Pesquisa:

Maria Angélica Pedra Minhoto (Unifesp) e
Pedro Arantes (Unifesp)



O Grupo de Dados do SoU_Ciência atua na coleta, tratamento, integração, documentação, proteção, preservação e disponibilização de dados para consulta dos pesquisadores associados ao Centro e à comunidade externa. O objetivo geral é contribuir para tornar o Centro de Estudos uma referência nacional na consolidação e divulgação de dados sobre a Educação Superior, Ciência e Tecnologia no Brasil.

O presente documento foi elaborado pelo Grupo de Dados do SoU_Ciência. São autores do documento os seguintes pesquisadores:

Coordenação

Dra. Thaís Cavalcante Martins

Apoio Técnico

Me. Antonio Guerreiro Silva Serrano

Ma. Ecila Alves de Oliveira Migliori

Colaboração

Dra. Daniela Leal Musa

Grupo de Dados

Thaís Cavalcante Martins

Alexsandro Cardoso Carvalho

Ecila Alves de Oliveira Migliori

Antonio Guerreiro Silva Serrano

Daniela Leal Musa

Meyrele Torres Nascimento

Fabiana Yumi Sato Kayanoki

Ricardo Ireno dos Santos

Apoio

Fundação Tide Setúbal

Instituto Serrapilheira

Emendas Parlamentares



Nosso conteúdo está disponível sob a licença Creative Commons Atribuição 4.0 Internacional, e pode ser compartilhado e reutilizado para trabalhos derivados, desde que citada a fonte.

SoU_Ciência

Coordenação Geral SoU_Ciência: Soraya Soubhi Smaili (Unifesp)

Coordenação de Pesquisa: Maria Angélica Pedra Minhoto (Unifesp) e Pedro Arantes (Unifesp)

Pesquisadores associados: Carlos Eduardo Bielchowsky (UFRJ), Daniela Leal Musa (Unifesp), Débora Foguel (UFRJ), Décio Semensato (Unifesp), Flávia Calé (USP), Gabriela de Brelàz (Unifesp), Hironobu Sano (UFRN), Ioshiaqui Shimbo (UFSCar), Iraneide Soares da Silva (UESPI), Joaze Bernardino Costa (UnB), Leandro Innocentini Lopes de Faria (UFSCar), Lidiane Cristina da Silva (Unifesp), Marcos Xavier (Unifesp), Maurício Moura (George Washington University), Natalino Neves da Silva (UFMG), Nelson Cardoso Amaral (UFG), Rogério Schlegel (Unifesp), Roniberto Morato do Amaral (UFSCar), Solange Ledi Gonçalves (Unifesp), Veneziano de Castro Araújo (Unifesp), Wilma de Nazaré Baía Coelho (UFPA).

Bolsistas de pós-doc: Maria Nilza da Silva, Sheina Koffler, Thaís Cavalcante Martins, Vanessa Sígolo, Vivian Peron.

Pesquisadores de pós-doc: André Luiz Vieira Dias, Danilo Braun Santos, Jade Percassi, Thiago Borges de Aguiar.

Pesquisadores colaboradores: Claudia Guedes Araújo Silva, Danilo Ferreira Malta, Evilasio da Silva Salvador, Fabio Goveia, Fabio Malini, Fernando Sfair Kinker, Maurício Ernica, Romualdo Luiz Portela de Oliveira, Mirhiane Mendes de Abreu, Tamyres Cristine Xavier Barbosa.

Bolsista de apoio técnico: Alexsandro Cardoso Carvalho, Antonio Guerreiro Silva Serrano, Ecila Alves de Oliveira Migliori, Fabiana Yumi Sato Kayanoki, Giovanna Nardini, Marina Mendes da Costa, Meyrele Torres Nascimento, Ricardo Ireno dos Santos.

Bolsistas de iniciação científica: Allan de Sousa Araújo, Ananery Lopes Ribeiro, Denisson Guimarães do Carmo, Giovanna Novais Campos, Ligia Andreussi Chrisostomo, Luisa Brito Rodrigues, Marina Merisia Barbosa de Souza, Samuel Montezuma Lemos, Vinicius Macena, Wesley Penariol Silva.

APRESENTAÇÃO

Este Guia não é um curso de análise de dados sobre o Censo da Educação Superior (CES) e do Exame Nacional de Desempenho dos Estudantes (Enade). Ele é uma referência prática sobre ferramentas computacionais que podem servir como facilitadoras para a utilização dos dados disponibilizados pelo Centro de Estudos SoU_Ciência. O que significa que o seu foco está na recomendação de ferramentas, software e na linguagem de programação.

Embora este Guia esteja alinhado com a tendência crescente do uso das Ciências de Dados em pesquisas acadêmicas, sua ênfase está no compartilhamento de informações que podem beneficiar os usuários comuns, isto é, o público interessado nos conjuntos de dados disponibilizados pelo Centro.

Como os conjuntos de dados possuem tamanhos diversos é possível que surjam problemas relacionados às limitações de memória, de armazenamento e de processamento em computadores convencionais. Aqui, o público poderá encontrar alternativas para lidar com estas questões.

O Guia oferece sugestões para o armazenamento dos dados, informações sobre a configuração mínima e ideal para a leitura dos arquivos e sugestões de software e linguagens de programação para analisar os dados.

Apesar de seu caráter técnico, o conteúdo deste Guia provavelmente será mais útil para os iniciantes. Ainda que também sirva para que os usuários com nível intermediário aprendam algo novo sobre armazenamento de dados, documentação e programação.

Um objetivo paralelo deste Guia é oferecer uma introdução à Análise Exploratória de Dados (AED) utilizando as linguagens de programação *Python* e *R* através do *Google Colab*. Neste espaço, o público poderá acessar algumas dicas sobre como analisar dados a partir de exemplos do Censo da Educação Superior e do Enade. Ainda assim, a ênfase está no código e na adaptação de exemplos, e não na descoberta de tendências que são típicas da pesquisa científica.

O material dos notebooks não cobre as amplas potencialidades de análise de dados nestes ambientes de programação, mas abrange fundamentos básicos para quem deseja trabalhar com estas linguagens.

Da mesma forma, estas não são as únicas ferramentas ou linguagens que podem ser utilizadas para analisar os dados disponibilizados. Inúmeras pesquisas científicas, que têm como fonte o CES e Enade, são desenvolvidas utilizando outras ferramentas. Neste caso, não existe melhor ou pior, somente uma questão de habilidade pessoal e custo-benefício relacionado às ferramentas.

Esperamos que a divulgação deste Guia contribua para fornecer incentivos para a utilização dos microdados educacionais, oferecendo informações úteis para quem deseja analisar grandes conjuntos de dados, como é o caso do Censo da Educação Superior e do Enade.

Boa leitura!

Os autores

Agradecemos todas as críticas e sugestões recebidas de pesquisadores e usuários que estão colaborando para a melhoria na qualidade da prestação das informações.

Todos os links apresentados neste Guia Prático estão ativos na data de sua publicação, em junho de 2023. Encontrou algum link quebrado? Nos ajude a contribuir com a aprendizagem de estudantes, pesquisadores, gestores e com a ciência aberta.

Avise-nos sobre erros em souciencia@unifesp.br

Dados Internacionais de Catalogação na Publicação (CIP)

Guia prático para a análise de dados: censo da educação superior e ENADE / Antônio Guerreiro Silva Serrano, Daniela Leal Musa, Ecila Alves de Oliveira Migliori; Thaís Cavalcante Martins (coordenadora). -- São Paulo: SoU_Ciência, 2023.
PDF.

ISBN 978-65-87312-82-8

O SoU_Ciência é um Centro de Estudos e Think Tank sobre Sociedade, Universidade e Ciência.

1. Educação superior – Brasil. 2. Avaliação. 3. Universidades e faculdades – Brasil – Exames. I. Serrano, Antônio Guerreiro Silva. II. Migliori, Ecila Alves de Oliveira. III. Musa, Daniela Leal. IV. Martins, Thaís Cavalcante, *coord.* II. Título.

CDD 378.81

Elaborado por Creuza Andréa Trindade dos Santos – CRB 2/1352

SUMÁRIO

1. Requisitos Básicos	9
1.1 Configurações do Computador	9
1.2 Sistema Operacional	9
1.4 Ferramentas de Análise de Dados	11
1.4.1 Softwares	11
1.4.2 Linguagens de programação	12
2. Análise Exploratória de Dados	14
2.1 Download dos Dados	14
2.2 Softwares	14
2.3 Linguagens de Programação	15
REFERÊNCIAS	16

1. Requisitos Básicos

O trabalho com grandes volumes de dados demandam alguns conhecimentos mínimos de hardware, software e, dependendo do tamanho e complexidade do conjunto de dados, linguagens de programação. Este guia fornece um panorama geral sobre as ferramentas computacionais indicadas para facilitar o trabalho com os bancos de dados disponibilizados pelo SoU_Ciência.

1.1 Configurações do Computador

Os conjuntos de dados do SoU_Ciência contêm arquivos no formato CSV de diversos tamanhos, da ordem de MB (*megabyte*) até GB (*gigabyte*), sendo que alguns deles são arquivos com milhões de linhas. Ao realizar a análise de dados em um computador convencional (*laptop* ou *desktop*), podem surgir problemas relacionados, principalmente, à limitações de armazenamento e memória RAM da máquina diante de arquivos muito grandes.

A **Tabela 1** indica configurações de hardware para ter um desempenho satisfatório. Isso é, um cenário em que o usuário possa conduzir as análises de dados e chegar aos seus resultados em um tempo razoável. Para arquivos com tamanho da ordem de MB, a configuração mínima atenderá plenamente as demandas de processamento de dados, já para arquivos da ordem de GB, recomenda-se as configurações intermediária (desempenho aceitável) e ideal (alto desempenho).

Tabela 1 - Configuração de hardware para análise do banco de dados.

	MÍNIMA	INTERMEDIÁRIA	IDEAL
RAM	8 GB	16 GB	32 GB
SSD	256 GB	480 GB	1 TB
HD	512 GB	512 GB	1 TB
Processador	Intel Core i5 AMD Ryzen 5	Intel Core i7 AMD Ryzen 7	Intel Core i7 AMD Ryzen 7

Fonte: Elaboração própria (Equipe de Dados do SoU_Ciência)

Consulte o arquivo **Leia-me do banco de dados do SoU_Ciência** para saber mais sobre o tamanho de cada arquivo.

1.2 Sistema Operacional

Os sistemas operacionais (*SOs*) desempenham um papel importante em trabalhos que envolvem dados, fornecendo uma plataforma estável e eficiente para a execução de análises e modelagem estatística. Análises de dados podem ser executadas nos três *SOs* mais conhecidos: *Windows*, *Mac* e *Linux*.

O Windows é um *SO* desenvolvido pela *Microsoft*, sendo o mais utilizado em todo o mundo. Ele oferece uma interface gráfica amigável e ampla compatibilidade com uma grande variedade de softwares. O *Windows* é, geralmente, a primeira opção para os interessados em trabalhar com

dados, sendo especialmente popular entre os iniciantes, pois muitas ferramentas de análise de dados têm suporte nativo para o SO. O *Microsoft Excel*, por exemplo, é uma ferramenta comumente utilizada para análise de dados no Windows, juntamente com softwares como o *Power BI* e o *SQL Server*.

O *Mac* é um SO desenvolvido pela *Apple* e é amplamente adotado por profissionais do design web e cientistas de dados. O *Mac* é conhecido por sua estabilidade e alto desempenho, o que é crucial para executar análises de dados de elevada complexidade. A linguagem de programação *Python*, amplamente utilizada na análise de dados, é nativamente suportada no *Mac*, o que facilita o desenvolvimento de código e a utilização de bibliotecas populares como o *pandas*, *numpy* e *scikit-learn*.

O Linux é um SO de código aberto amplamente utilizado em ambientes de Ciência de Dados. A grande vantagem do *Linux* é sua flexibilidade e capacidade de personalização, com excelente usabilidade, sendo compatível com a maioria das rotinas de trabalho ligadas à Ciência de Dados. O *Linux* é especialmente popular entre os interessados em trabalhar com dados com conhecimento de nível intermediário e entre os profissionais da área.

A escolha do SO depende das habilidades pessoais e dos interesses do projeto. Independentemente da escolha, todos os SOs oferecem recursos efetivos para realizar análises de dados de qualquer nível de complexidade.

1.3 Plataforma em Nuvem

A computação em nuvem permite o acesso de serviços de computação remotamente, bastando apenas estar conectado com a internet. Os dados não ficam armazenados em um computador pessoal ou um servidor local, mas, sim, numa rede global de servidores espalhados por todo o mundo para armazenar dados, executar aplicações e fornecer serviços aos usuários. Existem diversos serviços em nuvem voltados para quem está interessado em trabalhar com volumosos conjuntos de dados, sendo os principais: *AWS*, *Azure* e *Google Cloud*. Alguns serviços oferecem planos gratuitos, enquanto outros oferecem planos pagos.

Na **Tabela 2**, encontra-se uma recomendação de configurações mínima e ideal de serviços em nuvem para que seja possível a leitura dos arquivos CSV dos bancos de dados do SoU_Ciência e execução de scripts de programação de análise exploratória de dados.

Tabela 2 - Configuração de serviço em nuvem para análise do banco de dados.

	MÍNIMA	IDEAL
RAM	30 GB	> 30 GB
Armazenamento	5 GB	≥ 15 GB

Fonte: Elaboração própria (Equipe de Dados SoU_Ciência)

Embora o armazenamento em nuvem seja recomendado, ele não é obrigatório. Os bancos de dados disponibilizados pelo SoU_Ciência também podem ser armazenados localmente.

1.4 Ferramentas de Análise de Dados

Em um mundo cada vez mais orientado a dados (*data driven*), quantidades gigantescas de dados são produzidas a cada segundo, gerando um interesse crescente de pesquisa em todas as áreas do conhecimento, fornecendo informações relevantes para a comprovação de hipóteses e identificação de padrões de comportamentos.

Atualmente, tem se tornado cada vez mais comum utilizar ferramentas projetadas para organizar, categorizar e inferir conclusões estatísticas de diversos conjuntos de dados.

Os softwares são ferramentas que apresentam uma interface gráfica projetada para tornar a usabilidade mais fácil e intuitiva, já as linguagens de programação exigem do usuário uma maior capacidade de abstração e um conhecimento mais aprofundado de informática. Neste caso, não existe melhor ou pior, somente uma questão de habilidade pessoal e custo-benefício do ferramental.

1.4.1 Softwares

Existem muitas opções de softwares no mercado para utilização em projetos de pesquisa, com versões gratuitas e/ou pagas. Na Tabela 3, indica-se os softwares mais comuns, apontando se há restrição de tamanho de arquivo suportado, bem como se a utilização exige o pagamento de licenças.

Tabela 3 - Tamanho máximo do arquivo que pode ser carregado na versão mais recente e tipo de licenciamento do software ou aplicação web.

SOFTWARE	TAMANHO	GRATUITO
<i>Planilhas Google</i>	≤ 20 MB	Não*
<i>Microsoft Excel</i>	-	Não*
<i>LibreOffice Calc</i>	-	Sim
<i>IBM SPSS Statistics</i>	-	Não*
<i>STATA</i>		Não
<i>Statistica</i>		Não
<i>Power BI</i>		Não**
<i>Tableau</i>		Não*
<i>Pentaho</i>		Sim
<i>Metabase</i>	-	Não**

“-” Não há um tamanho máximo de arquivo especificado. Logo, o desempenho do programa depende principalmente do sistema operacional e configuração de hardware do computador, e também do plano contratado.

“*” Ferramenta paga, mas possui um plano gratuito ou de teste para estudantes e/ou professores de instituições de ensino credenciadas.

“**” Ferramenta paga, mas possui um plano gratuito (*freemium*).

Fonte: Elaboração própria (Equipe de Dados do SoU_Ciência)

Consulte o arquivo [Leia-me do banco de dados do SoU_Ciência](#) para saber mais sobre o tamanho de cada arquivo.

Caso o software escolhido não tenha sido mencionado na **Tabela 3**, consulte a documentação técnica do mesmo para obter as informações necessárias.

1.4.2 Linguagens de programação

Com as linguagens de programação, ocorre o mesmo fenômeno dos softwares, existem muitas opções e, de tempos em tempos, surge uma nova, tida como revolucionária. Este guia não pretende examinar as funcionalidades de cada linguagem, mas apenas indicar as soluções mais populares atualmente.

Independente do tamanho do arquivo CSV do banco de dados do SoU_Ciência, qualquer linguagem de programação será capaz de executar uma análise do conjunto de dados. O desempenho das linguagens de programação está ligada diretamente à capacidade de processamento e armazenamento do computador ou plataforma em nuvem, para isso, certifique-se sempre de ter a configuração mais adequada para execução dos scripts de programação. Leia o capítulo 1. Requisitos Básicos deste manual.

A Tabela 4 apresenta as principais linguagens de programação e algumas informações básicas para nortear a escolha da linguagem mais apropriada, segundo as características de cada projeto de pesquisa.

Tabela 4- Linguagens de programação, tipo de licenciamento e acessibilidade do código-fonte.

LINGUAGEM	GRATUITO	CÓDIGO ABERTO
<i>Python</i>	Sim	Sim
<i>R</i>	Sim	Sim
<i>SQL</i>	Sim	-
<i>Julia</i>	Sim	Sim
<i>Scala</i>	Sim	Sim
<i>SAS*</i>	Não	Não
<i>Matlab</i>	Não	Não

“-” Depende do Sistema Gerenciador de Banco de Dados (SGBD). Por exemplo, o *MySQL* e o *PostgreSQL* são SGBDs que utilizam SQL e são softwares de código aberto, enquanto o *Microsoft SQL Server* é um produto proprietário que também utiliza SQL.

“*” SAS é um software comercial que possui a sua própria linguagem de programação, chamada SAS Programming Language.

Fonte: Elaboração própria (Equipe de Dados do SoU_Ciência)

Python é uma linguagem de multipropósito, isto é, ela é utilizada não só para a análise de dados, mas também para o desenvolvimento de outros tipos de aplicação. Ela tem alcançado um grande crescimento de popularidade nos últimos anos, tornando-se a mais famosa e popular entre os profissionais que trabalham com dados. Por ser tão popular, existe uma imensa quantidade e variedade de conteúdos técnicos disponíveis pela internet, além de uma comunidade muito atuante na criação constante de novas bibliotecas.

R é uma linguagem que foi desenvolvida por matemáticos e estatísticos, criada com o intuito de lidar com um volume massivo de dados. Sua utilização é muito comum no meio acadêmico brasileiro e internacional, mas também fora dele, contando com uma comunidade gigantesca e muito ativa. Existem bibliotecas em *R* para resolver problemas bem específicos no tratamento e análise de dados, mas também para lidar com problemas de pesquisa. Sendo que, é comum que o usuário, seja ele um cientista de dados ou um pesquisador, utilize *Python* e *R* simultaneamente em projetos.

SQL, cuja sigla significa Structured Query Language, é uma linguagem de programação para manipular bancos de dados relacionais (tabelas com linhas e colunas). O intuito do *SQL* é analisar e/ou executar tarefas em tabelas, principalmente através dos comandos básicos inserir (*insert*), pesquisar (*search*), atualizar (*update*) e excluir (*delete*). Não só isso, mas principalmente, escrever comandos de consultas (*queries*) que lidam com diferentes parâmetros para obter informações específicas dentro de um banco de dados.

As três linguagens citadas são populares nos trabalhos que envolvem grandes volumes de dados. As demais, se encontram restritas a determinados segmentos- não possuindo uma grande comunidade de usuários. Diante disso, a disponibilidade e o custo para a resolução de problemas, típicos da análise de dados, são fatores que pesam bastante na escolha destas linguagens e devem ser levados em consideração no desenvolvimento de projetos.

2. Análise Exploratória de Dados

O objetivo da Análise Exploratória de Dados (**AED**) é gerar uma análise prévia do conjunto de dados antes da aplicação de qualquer técnica estatística. Assim, é possível entender melhor os dados e as relações existentes entre as variáveis analisadas.

As etapas da AED variam dependendo do autor, mas, em geral, elas abrangem os seguintes estágios:

- I. Definir claramente o objetivo da análise** - Qual(is) a(s) informação(ões) que eu desejo extrair desse conjunto de dados? Qual(is) hipótese(s) eu desejo validar?
- II. Escolha da fonte dos dados** - arquivos no formato CSV ou JSON, site da internet, texto, etc.
- III. Metodologia de aquisição dos dados** - leitura de arquivos no formato CSV ou JSON, consumo de API, Web Scraping, etc.
- IV. Formatação dos dados** - converter para um formato legível pela ferramenta de análise de dados, por exemplo, data no formato aaaa-mm-dd.
- V. Limpeza dos dados** - eliminar outliers, dados ausentes, inconsistentes, etc.
- VI. Seleção dos dados** - aplicação de filtros para utilizar somente os dados de interesse.
- VII. Agregação/Desagregação dos dados** - união/separação de dataframes gerados a partir de múltiplos arquivos no formato CSV, por exemplo.
- VIII. Intercorrelação de variáveis** - identificação do grau de interdependência entre as variáveis através de um mapa de calor (heatmaps), por exemplo.
- IX. Visualização dos dados** - extrair informações de um conjunto de dados graficamente através de um histograma, um diagrama de caixas, gráfico de dispersão, etc.

2.1 Download dos Dados

Os bancos de dados do SoU_Ciência estão disponíveis no repositório institucional da Universidade Federal de São Paulo e estão em formato CSV delimitados por ponto e vírgula (;). Os arquivos encontram-se compactados (em formato .zip) e devem ser descompactados no seu disco rígido.

2.2 Softwares

A equipe de Dados do SoU_Ciência elenca aqui uma série de tutoriais ou documentação técnica para consulta sobre os softwares indicados neste guia (seção 1.4.1 Softwares) onde podem ser encontradas informações de como executar as etapas de uma AED.

- [Google Planilhas](#)
- [MS Excel](#)
- [LibreOffice Calc](#)
- [IBM SPSS Statistics](#)
- [STATA](#)
- [Statística](#)

- [Power BI](#)
- [Tableau](#)
- [Pentaho](#)
- [Metabase](#)

2.3 Linguagens de Programação

A equipe de Dados do SoU_Ciência disponibiliza abaixo um passo a passo de como executar uma AED utilizando as linguagens de programação *Python* e *R* através do *Google Colab* (editor de código em nuvem da *Google*). O *Colab* permite a escrita de textos em formato *LaTeX* juntamente com código de programação, sendo assim, é possível, em um único ambiente, (denominado *notebook*), fazer cálculos, executar códigos, plotar gráficos, e compartilhar na forma de um link, como a seguir:

- [Notebook Python](#)
- [Notebook R](#)

2.4 Recomendações de desempenho

Nem sempre executar uma AED é uma tarefa simples, principalmente quando o computador de trabalho não possui uma configuração robusta que comporte um arquivo CSV com milhões de linhas em tempo hábil. Neste caso, a equipe de Dados do SoU_Ciência compartilha aqui dois notebooks com algumas dicas simples de como analisar os dados sem consumir tanta memória e processamento do computador.

- [Notebook Python](#)
- [Notebook R](#)

REFERÊNCIAS

INTEL. Intel Core Processor Family. Disponível em: <<https://www.intel.com/content/www/us/en/products/details/processors/core.html>>. Acesso em: 16 maio 2023.

AMD. Ryzen Processors. Disponível em: <<https://www.amd.com/en/processors/ryzen>>. Acesso em: 16 maio 2023.

IBM. O que é ciência de dados?. Disponível em: <<https://www.ibm.com/br-pt/topics/data-science>>. Acesso em: 16 maio 2023.

Linux Kernel Organization. Documentação do Linux. Disponível em: <<https://www.kernel.org/doc/html/latest/>>. Acesso em: 18 maio 2023.

Microsoft. Documentação oficial do Windows. Disponível em: <<https://learn.microsoft.com/pt-br/windows/>>. Acesso em: 18 maio 2023.

Apple Inc. Documentação oficial da Apple e do macOS. Disponível em: <<https://developer.apple.com/documentation/>>. Acesso em: 18 maio 2023.

RICE UNIVERSITY. Programming Languages for Data Science. Disponível em: <<https://csweb.rice.edu/academics/graduate-programs/online-mds/blog/programming-languages-for-data-science>>. Acesso em: 16 maio 2023.

DATOS.GOB.ES. Cuando utilizar cada lenguaje de programación en ciencia de datos. Disponível em: <<https://datos.gob.es/es/blog/cuando-utilizar-cada-lenguaje-de-programacion-en-ciencia-de-datos>>. Acesso em: 16 maio 2023.

DAT Linux. Disponível em: <<https://datlinux.com/>>. Acesso em: 15 maio 2023.

Amazon Web Services. Disponível em: <<https://aws.amazon.com/pt/>>. Acesso em: 15 maio 2023.

Microsoft. Azure. Disponível em: <<https://azure.microsoft.com/pt-br>>. Acesso em: 15 maio 2023.

Google Cloud. Disponível em: <<https://cloud.google.com/?hl=pt-br>>. Acesso em: 15 maio 2023.

Forbes. How The World Became Data-Driven And What's Next? Por Kevin Casey. Publicado em 20 de maio de 2020. Disponível em: <<https://www.forbes.com/sites/googlecloud/2020/05/20/how-the-world-became-data-driven-and-whats-next/?sh=285adf0757fc>>. Acesso em: 15 de maio de 2023.

MIT Connection Science. Data-Driven Society. Disponível em: <https://connection.mit.edu/sites/default/files/publication-pdfs/data%20driven%20society%20sci%20amer_0.pdf>. Acesso em: 15 maio 2023.

BOOTCAMP CVN COLUMBIA. Python basics guide. Disponível em: <<https://bootcamp.cvn.columbia.edu/blog/python-basics-guide/>>. Acesso em: 15 maio 2023.

GOOGLE. Planilhas Google. Disponível em: <<https://support.google.com/docs/answer/37603?hl=pt-BR>>. Acesso em: 15 maio 2023.

WORKSPACETIPS.IO. Google Docs Size Limitations. Disponível em: <<https://workspacetips.io/tips/docs/google-docs-size-limitations/#:~:text=Uploaded%20spreadsheet%20files%20that%20are,detailed%20spreadsheet%20size%20limits%20tip>>. Acesso em: 15 maio 2023.

IBM. Planning Analytics 2.0.0 Known issues: Microsoft Excel limits. IBM, 2021. Disponível em: <<https://www.ibm.com/docs/en/planning-analytics/2.0.0?topic=issues-microsoft-excel-limits>>. Acesso em: 15 maio 2023.

The Document Foundation. LibreOffice 7.2 Calc Guide. Disponível em: <<https://nextcloud.documentfoundation.org/s/nwyEYBmQTS9B6pk>>. Acesso em: 15 maio 2023.

STATA CORPORATION. Stata: Release 18. Statistical Software. College Station, TX: StataCorp LLC, 2023. Disponível em: <<https://www.stata.com/bookstore/users-guide/>>. Acesso em: 15 maio 2023.

TIBCO SOFTWARE INC. TIBCO Statistica Quick Reference Guide Version 14.0.1. [S.l.], 2021. Disponível em: <https://docs.tibco.com/pub/stat/14.0.1/doc/pdf/TIB_stat_14.0.1_quick_ref.pdf?id=1>. Acesso em: 15 maio 2023.

MICROSOFT. Power BI. [S.l.], 2023. Disponível em: <<https://learn.microsoft.com/pt-br/power-bi/>>. Acesso em: 15 maio 2023.

TABLEAU SOFTWARE LLC. Tableau. [S.l.], 2023. Disponível em: <<https://www.tableau.com/pt-br>>. Acesso em: 15 maio 2023.

HITACHI VANTARA LLC. Pentaho Documentation Version 9.1. [S.l.], 2021. Disponível em: <<https://help.hitachivantara.com/Documentation/Pentaho/9.1>>. Acesso em: 15 maio 2023.

METABASE INC. Metabase Documentation Version latest. [S.l.], 2023. Disponível em: <<https://www.metabase.com/docs/latest/>>. Acesso em: 15 maio 2023.

Python Software Foundation. Python Documentation. Disponível em: <<https://www.python.org/doc/versions/>>. Acesso em: 15 maio 2023.

The R Project for Statistical Computing. Other documentation. Disponível em: <<https://www.r-project.org/other-docs.html>>. Acesso em: 15 maio 2023.

AMAZON WEB SERVICES. O que é SQL?. Disponível em: <<https://aws.amazon.com/pt/what-is/sql/>>. Acesso em: 16 maio 2023.

Oracle Corporation. MySQL Documentation. Disponível em: <<https://dev.mysql.com/doc/>>. Acesso em: 15 maio 2023.

The PostgreSQL Global Development Group. PostgreSQL Documentation. Disponível em: <<https://www.postgresql.org/docs/>>. Acesso em: 15 maio 2023.

Microsoft Learn. Documentação do SQL Server. Disponível em: <<https://learn.microsoft.com/pt-br/sql/?view=sql-server-ver16>>. Acesso em: 15 maio 2023.

The Julia Language. Julia Documentation v1. Disponível em: <<https://docs.julialang.org/en/v1/>>. Acesso em: 15 maio 2023.

The Scala Programming Language. Scala Documentation. Disponível em: <<https://docs.scala-lang.org/>>. Acesso em: 15 maio 2023.

SAS Institute Inc. SAS Documentation. Disponível em: <<https://support.sas.com/en/documentation.html>>. Acesso em: 15 maio 2023.

MathWorks. MATLAB Documentation. Disponível em: <<https://www.mathworks.com/help/matlab/>>. Acesso em: 15 maio 2023.

SchemaSpy. Disponível em: <<https://schemaspy.org/>>. Acesso em: 15 maio 2023.

SCHEMASPY. Schemaspy documentation. [S.l.]: Read the Docs, 2023. Disponível em: <<https://schemaspy.readthedocs.io/en/latest/>>. Acesso em: 16 maio 2023.

SCHEMASPY. Repositório do Schemaspy no GitHub. Disponível em: <<https://github.com/schemaspy/schemaspy>>. Acesso em: 16 maio 2023.

COLAB RESEARCH. Introdução à Análise Exploratória de Dados utilizando a biblioteca Pandas. Disponível em: <<https://colab.research.google.com/drive/1ZK1zqBeBeFJ315yizvAt5UQyc7TNI1yy?usp=sharing>>. Acesso em: 16 maio 2023.

COLAB RESEARCH. Google Colab Tutorial. Disponível em: <<https://colab.research.google.com/drive/16pBJQePbqkz3QFV54L4NikOn1kwpuRrj>>. Acesso em: 16 maio 2023.

ZUUR, Alain F.; IENO, Elena N.; MEESTERS, Erik HWG.; outros. A Beginner's Guide to R. 2009. Springer.

DALGAARD, Peter. Introductory Statistics with R. Springer, 2008.

GUTTAG, John V. Introduction to Computation and Programming Using Python. Mit Press, 2013.

IGUAL, Laura; SEGUÍ, Santi. Introduction to Data Science. Springer, 2017.

LEE, Kent D. Python Programming Fundamentals. Springer, 2011.