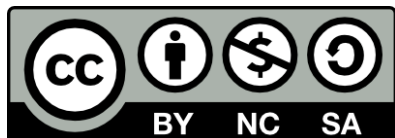


INTI6 – Inteligência Artificial

2023.2

Aula 03

Introdução – Definindo Ciência de Dados



Prof. Everton Silva
everton.silva@ifsp.edu.br



Aula de Hoje

- Uma breve história
- O que é Ciência de Dados?
- Um raio-x dos profissionais de dados do Brasil

Uma breve história

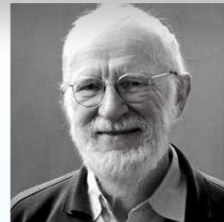
HISTORY OF DATA SCIENCE

1962

*Contribuições em
computação e
estatística.*



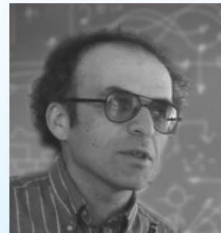
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



Uma breve história

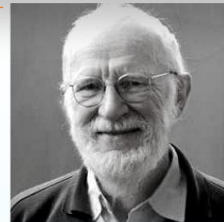
HISTORY OF DATA SCIENCE

1974

*Processamento de dados.
O termo "Data Science" começa a ficar mais claro.*



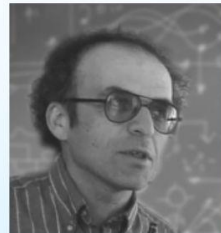
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori

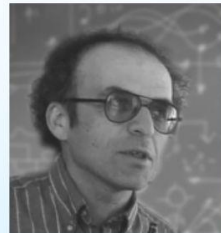


Uma breve história

HISTORY OF DATA SCIENCE

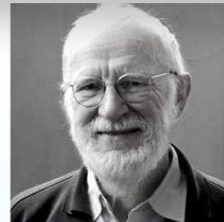


1962, John W. Tukey writes in "The Future of Data Analysis"



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".



1974, Peter Naur publishes the Concise Survey of Computer Methods

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1977, the International Association for Statistical Computing (IASC) was founded

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



1977

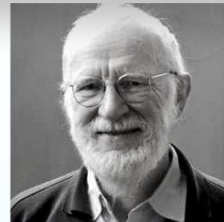
Fomentar o interesse em computação estatística.

Uma breve história

HISTORY OF DATA SCIENCE



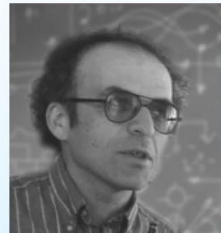
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



1989

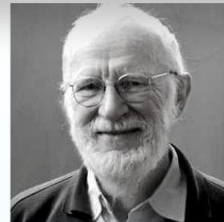
Primeiro workshop de KDD.
KDnuggets

Uma breve história

HISTORY OF DATA SCIENCE



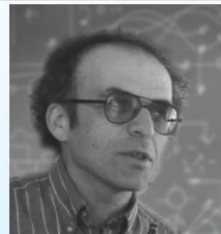
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



1994

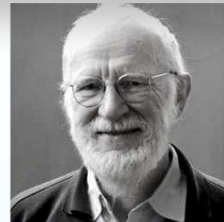
Matéria sobre Marketing de banco de dados.

Uma breve história

HISTORY OF DATA SCIENCE



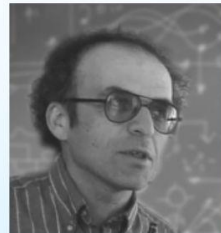
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



1996

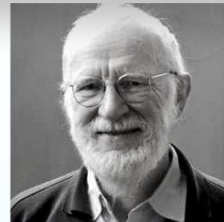
O termo "ciência de dados" foi incluído no título da conferência.

Uma breve história

HISTORY OF DATA SCIENCE



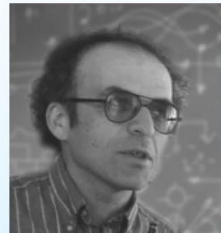
1962, John W. Tukey writes in "The Future of Data Analysis"



1974, Peter Naur publishes the Concise Survey of Computer Methods



1977, the International Association for Statistical Computing (IASC) was founded



1989, Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop

BusinessWeek

1994, BusinessWeek published a cover story on "Database Marketing"



1997, during his inaugural lecture as the H. C. Carver Chair in Statistics at the University of Michigan, Jeff Wu called for statistics to be renamed "data science" and statisticians to be renamed "data scientists".

1996, in the occasion of the conference of International Federation of Classification Societies (IFCS), for the first time, the term "data science" is included in the title of the conference



Infographic by @ingliguori



1997

Estatísticas fossem renomeadas como "ciência de dados".

O que é Ciência de Dados?

Algumas definições

A ciência de dados se preocupa em **analisar dados e extrair conhecimento** útil deles. A construção de modelos preditivos geralmente é a atividade mais importante para um cientista de dados ¹.

A ciência de dados é uma disciplina emergente que se baseia no **conhecimento** da metodologia estatística e da ciência da computação para **criar previsões e percepções** impactantes para uma ampla gama de campos acadêmicos tradicionais ³.

A ciência de dados se preocupa em **analisar Big Data** para **extrair correlações** com estimativas de probabilidade e erro ².

A ciência de dados é a prática de **adquirir, limpar, explorar, visualizar e analisar dados**, utilizando técnicas estatísticas e computacionais, a fim de **extrair conhecimento, insights e valor**, bem como comunicar os resultados para **apoiar a tomada de decisões e resolver problemas em diversos domínios**.

ChatGPT

1. <https://www.kdnuggets.com/tag/data-science>
2. https://www.researchgate.net/publication/285586313_Understanding_Data_Science_An_Emerging_Discipline_for_Data-Intensive_Discovery
3. <https://datascience.harvard.edu>

O que é Ciência de Dados?

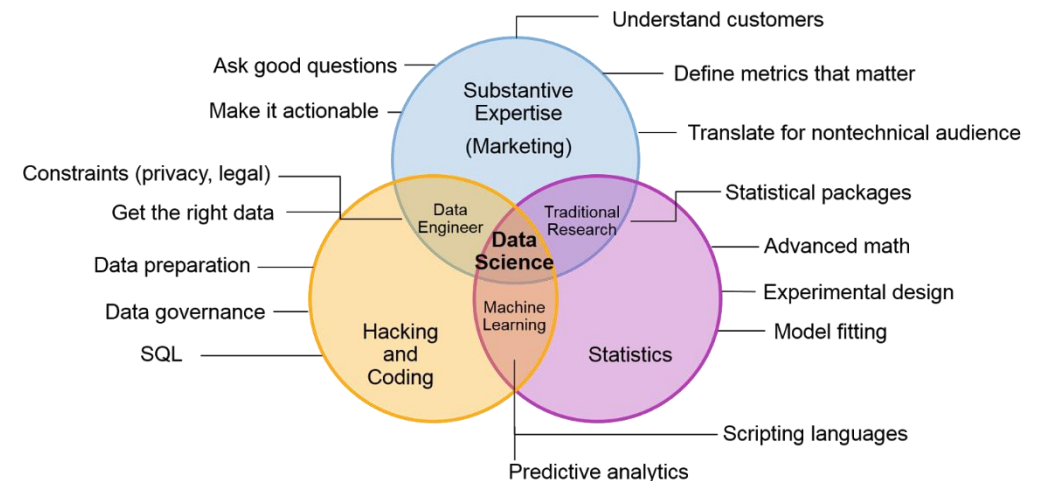
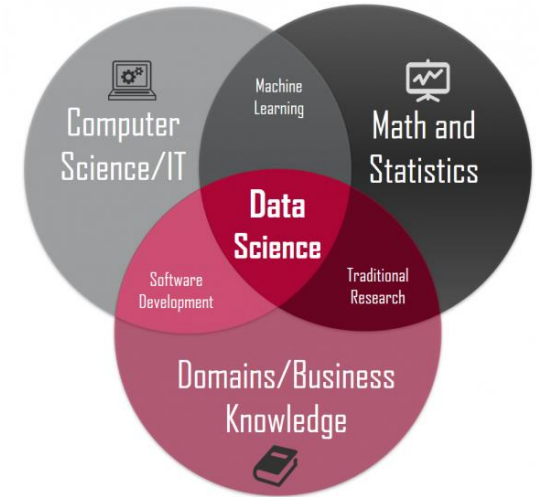
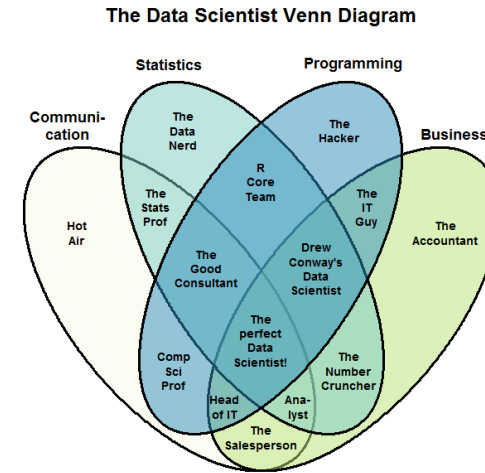
Ciência de dados é um termo que **não tem** uma única e completa definição;

Os métodos, objetivos e aplicações evoluíram com o tempo e avanço da tecnologia;

25 atrás: coleta e limpeza de dados e a aplicação de métodos estatísticos;

Hoje: análise de dados, análise preditiva, mineração de dados, inteligência de negócios, aprendizado de máquina...

Dar sentido aos dados tem uma longa história e é amplamente discutido.



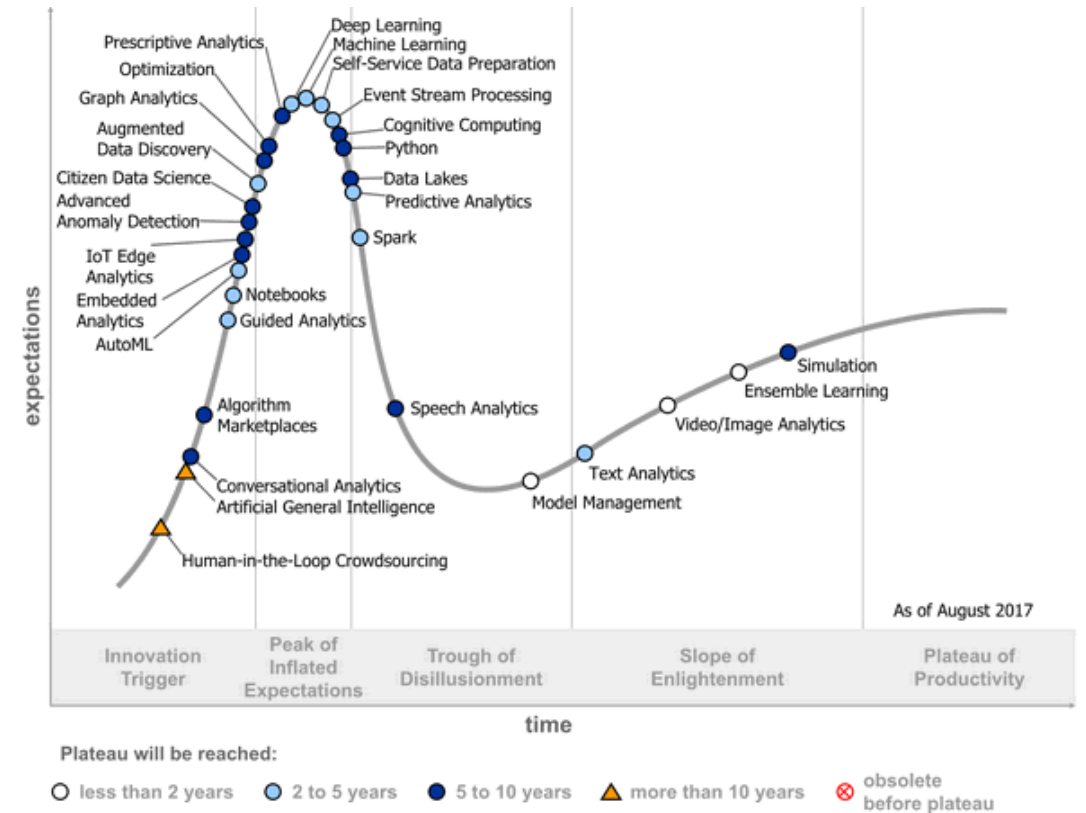
O que é Ciência de Dados?

Os termos “ciência de dados” e “mineração de dados” costumam ser usados alternadamente;

Ciência de dados ganhou notoriedade à medida que vários indivíduos e organizações tentam tirar proveito do “***hype***” em torno dela;

A ciência de dados é um **conjunto de princípios fundamentais** que orientam a extração de conhecimento dos dados;

A mineração de dados é a **extração de conhecimento dos dados**, por meio de tecnologias que incorporam esses princípios;

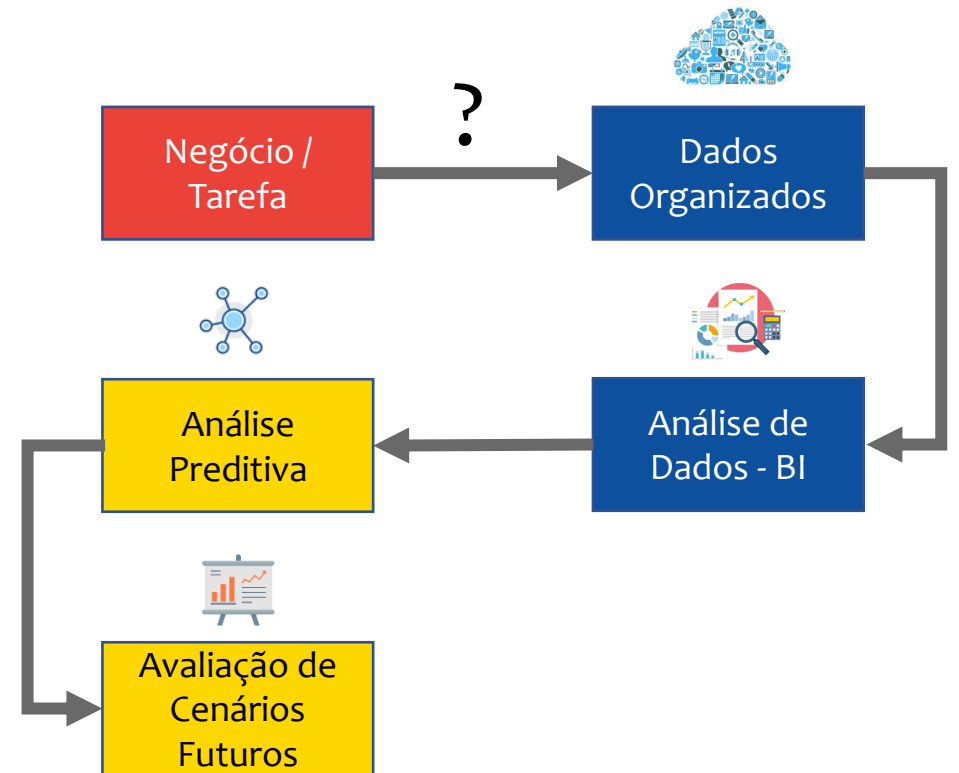


© 2017 Gartner, Inc.

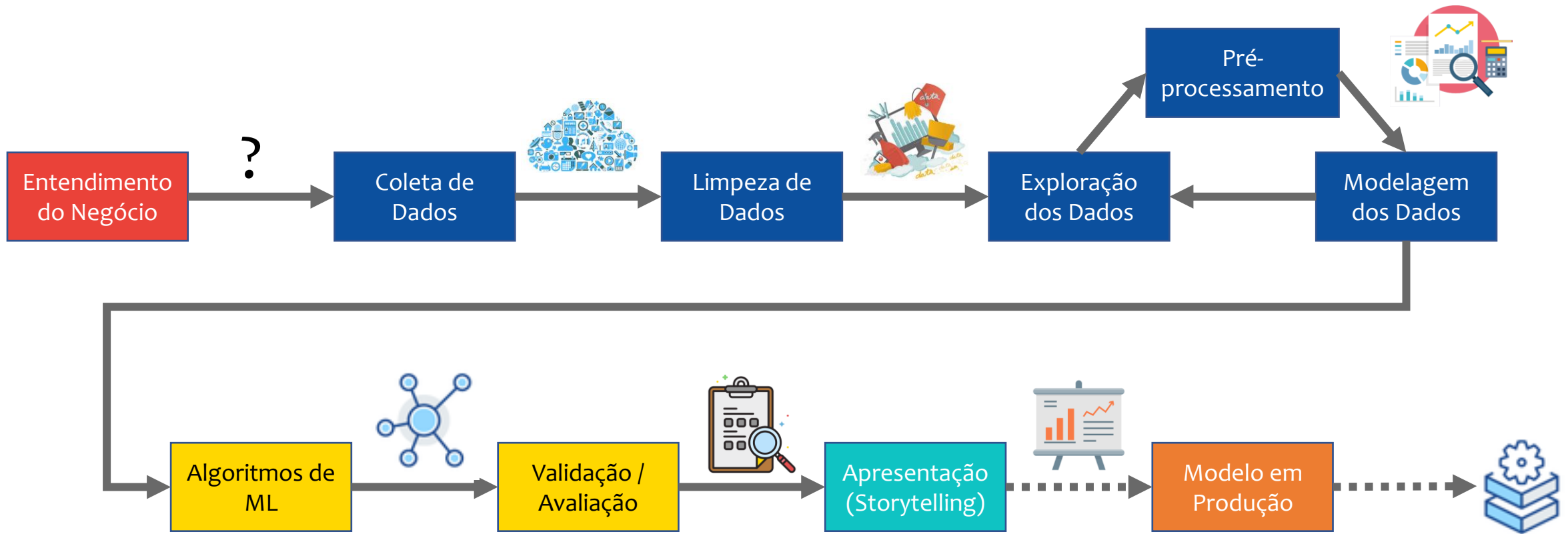
O que é Ciência de Dados?

Como abordar um negócio ou uma tarefa de ciência de dados?

- Cenário 1: você tem os dados adequados e **uma tarefa definida**. Ex.: prever a demanda de um determinado produto para os próximos meses;
- Cenário 2: você recebe uma enorme quantidade de dados e **precisa fazer algo e dizer como** aumentar o lucro para o próximo ano;
- Em ambos os casos devemos ter um conjunto de dados adequado;



Pipeline de Ciência de Dados



Mercado de Trabalho no Brasil

Pesquisa State of Data realizada por Data Hackers e Bain & Company.

Entre 10 de outubro e 28 de novembro de 2022



4.270 respondentes



Diferentes papéis de atuação



Mercado de Trabalho no Brasil

Principais resultados

Alta demanda de profissionais de dados pelas empresas

Aumento da diversidade de cursos de ensino superior

Surgimento de novas posições relacionadas a dados nas empresas

Maior migração de profissionais de outras áreas (**7,7% em 2021 – 13,5% em 2022**)

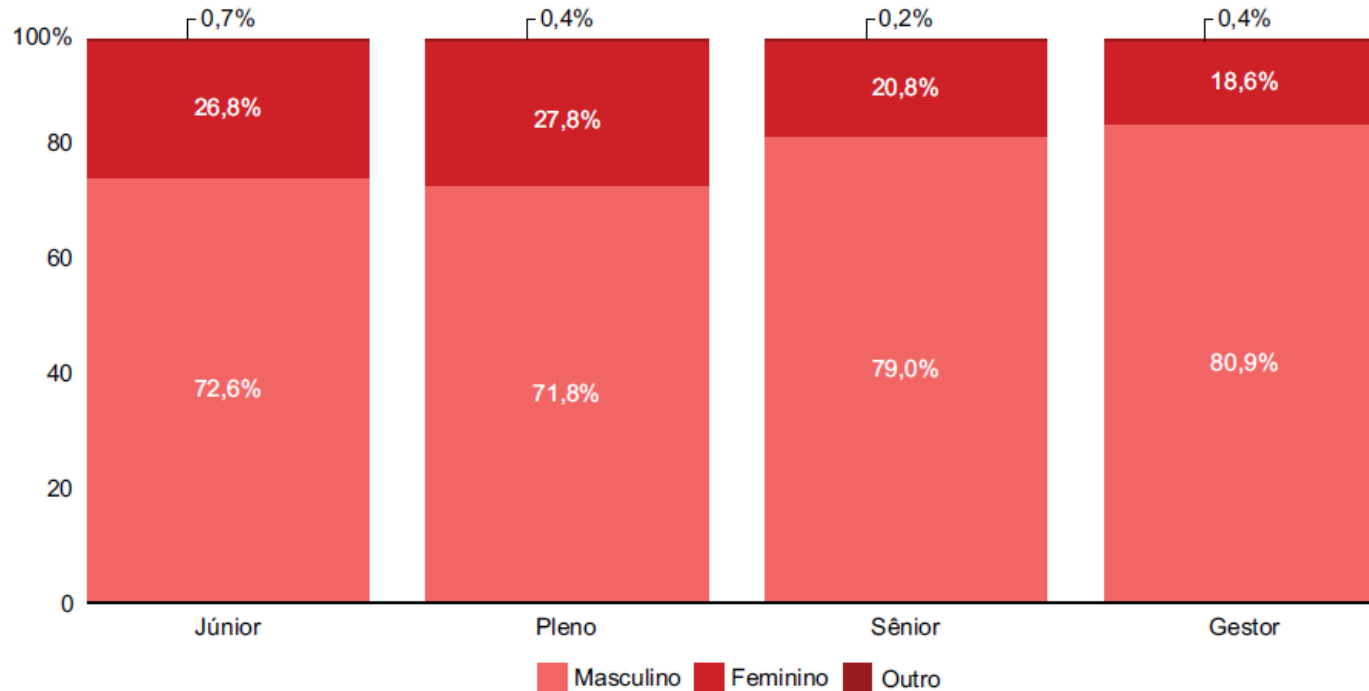
A remuneração aumentou em média cerca de 4% entre 2021 e 2022

Mulheres pardas e pretas ganham em média 23,1% menos que brancos (Gestão)

Mercado de Trabalho no Brasil

Perfil demográfico

Figura 1: Distribuição de gênero por nível de cargo



A participação relativa de pessoas do gênero feminino é maior em **perfis de início de carreira**

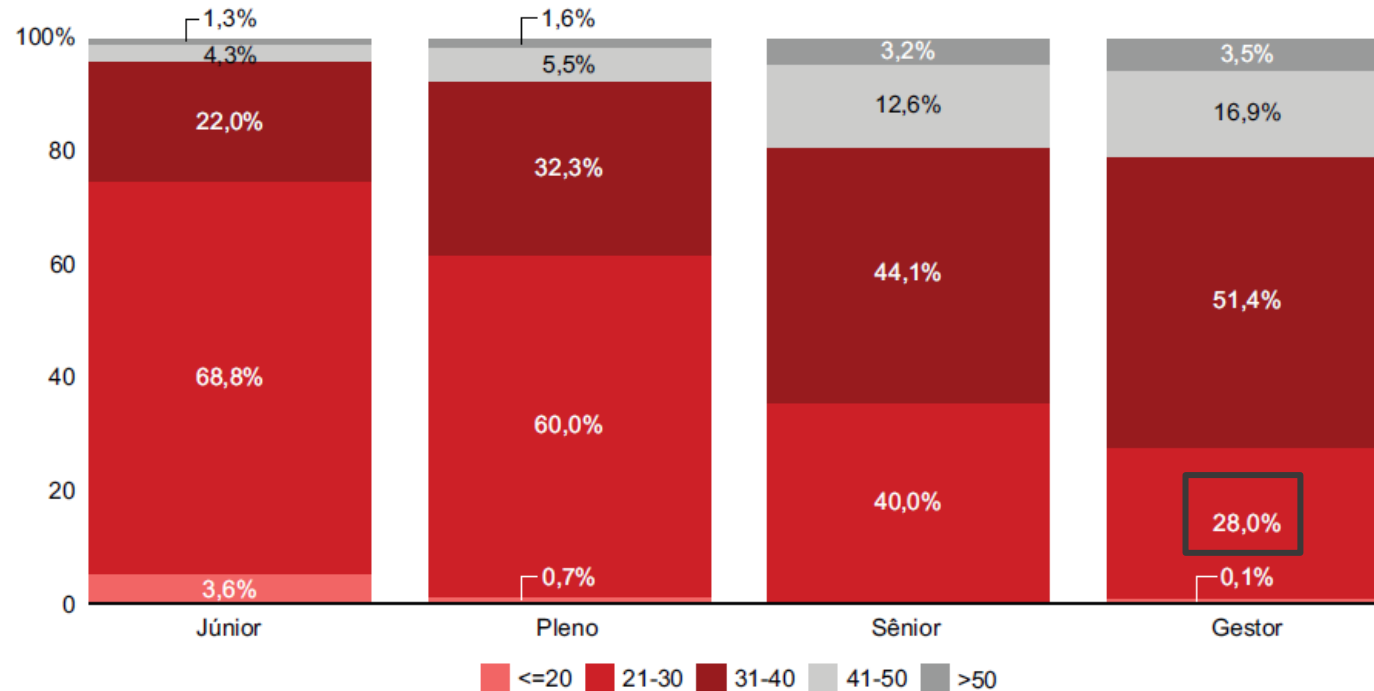
O nível de gestores apresentou relevante aumento em relação à edição anterior da pesquisa, com **18,6%** das respondentes (**versus 13,0%** em 2021)

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Perfil demográfico

Figura 2: Distribuição de idade por nível de cargo



Os níveis de cargo **júnior** e **pleno** possuem maior concentração na faixa dos **21 aos 30 anos**

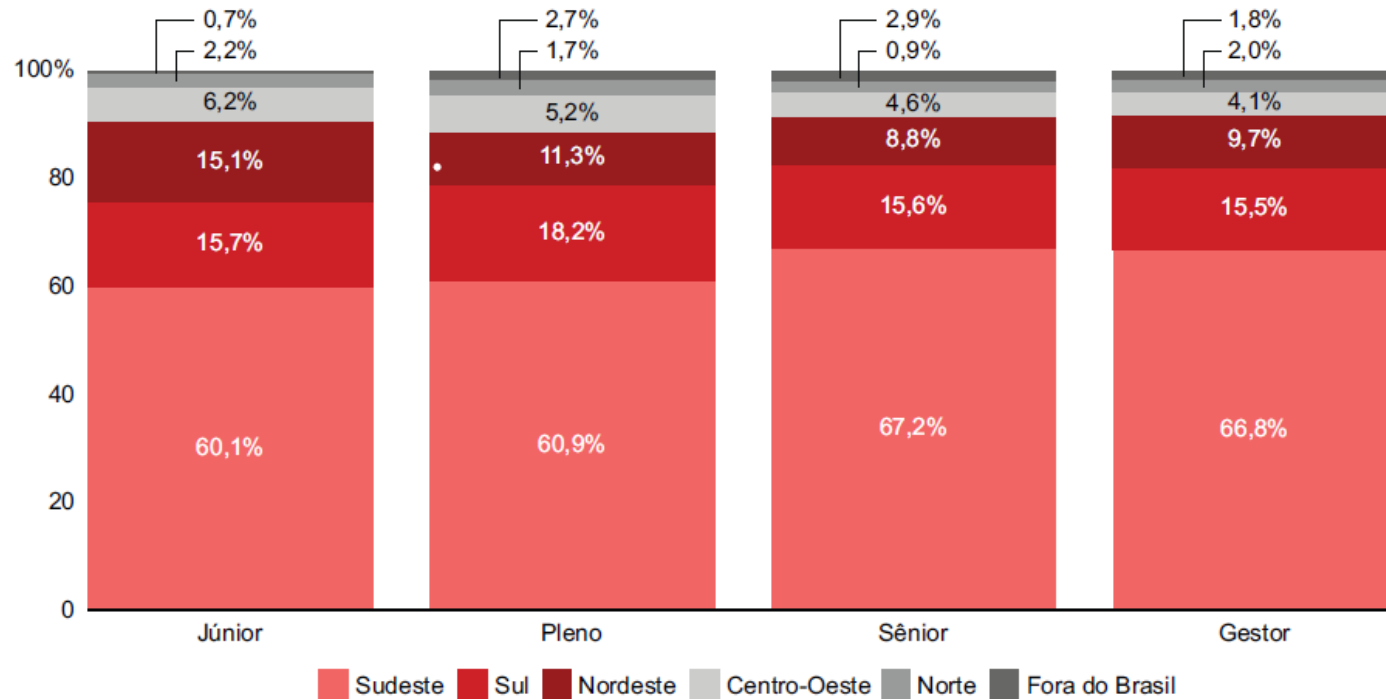
Os níveis de cargo **sênior** e de **gestores** possuem maior concentração na faixa dos **31 aos 40 anos**

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Perfil demográfico

Figura 3: Distribuição de local de residência por nível de cargo



A maior parte dos profissionais de dados no Brasil ainda vive nas regiões **Sudeste e Sul**

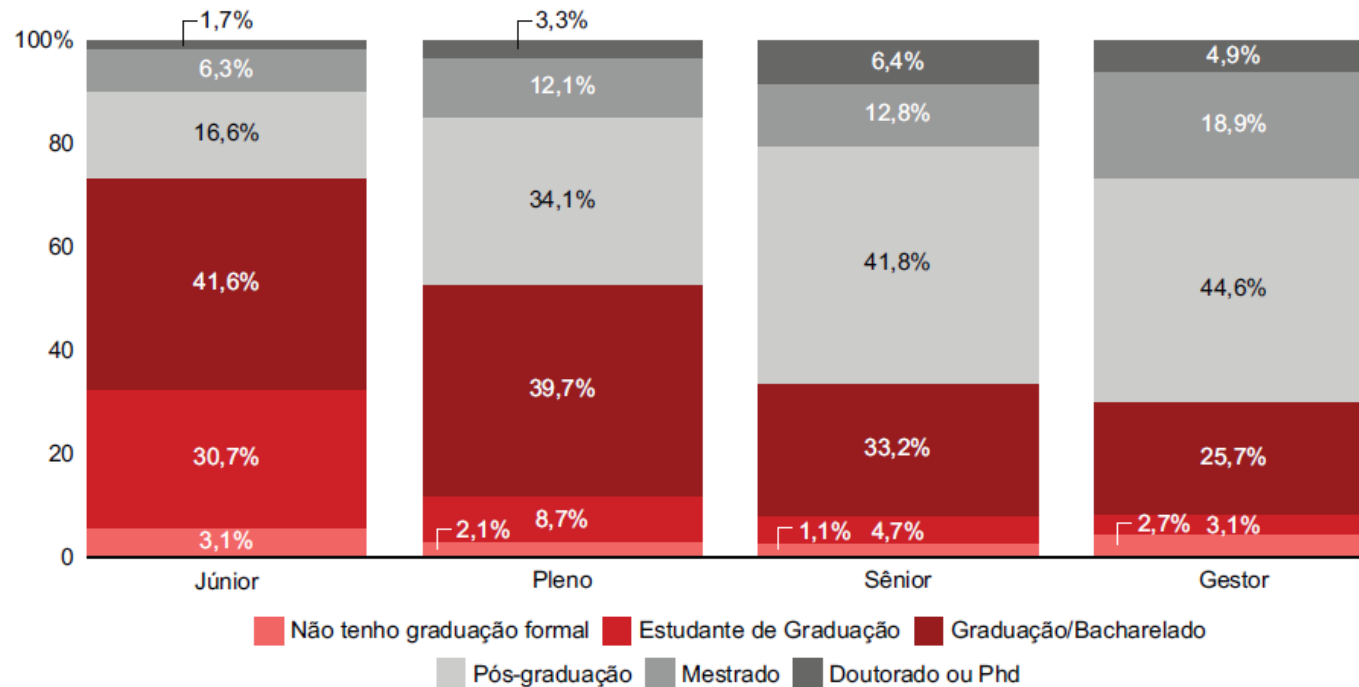
2,0% vive fora do Brasil, representando um leve aumento em relação ao dado de 2021 (1,7%)

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Formação e experiência

Figura 4: Distribuição de nível de instrução formal por nível de cargo



Possuem **alto nível de instrução formal**, com alto índice de pós-graduados no total de respondentes (49%)

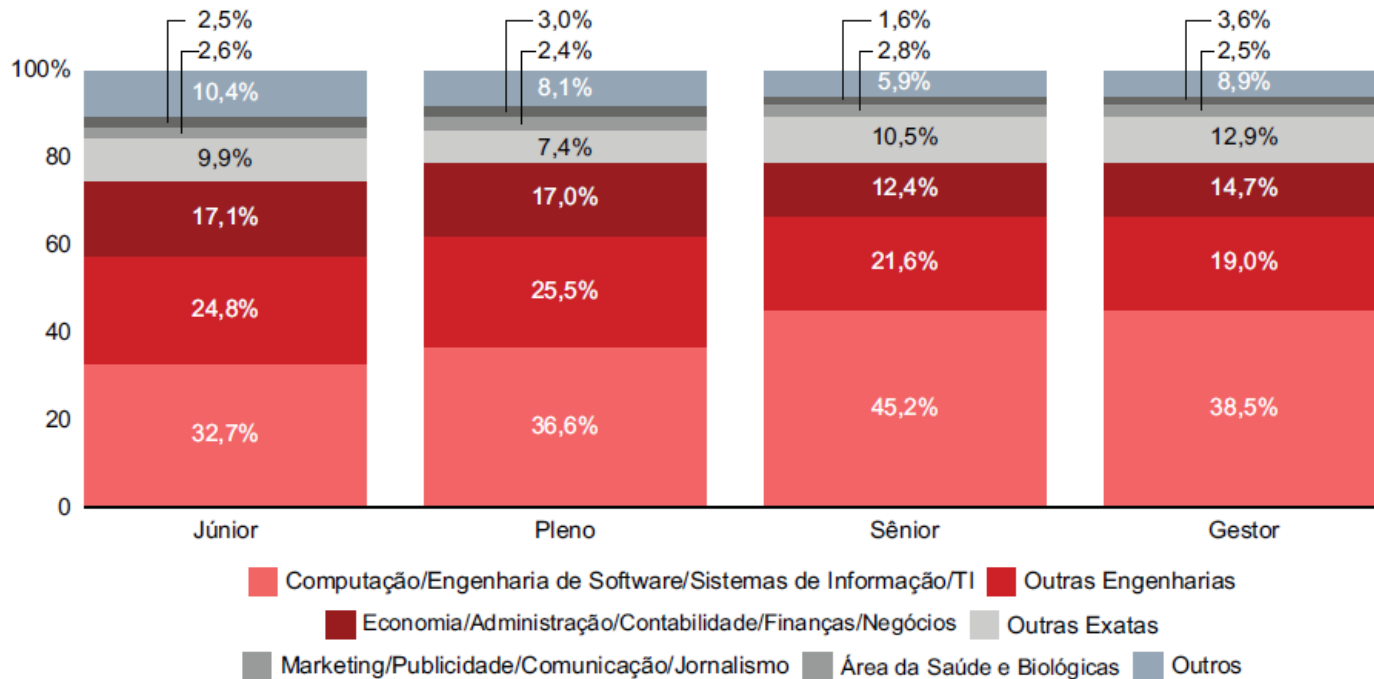
Nos níveis sênior e de gestão, a proporção de respondentes que possuem algum nível de pós-graduação é de **61% e 68%**

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Formação e experiência

Figura 5: Distribuição de área de formação por nível de cargo



Destaque a predominância de áreas relacionadas a TI e engenharia (38,0% e 23,1%)

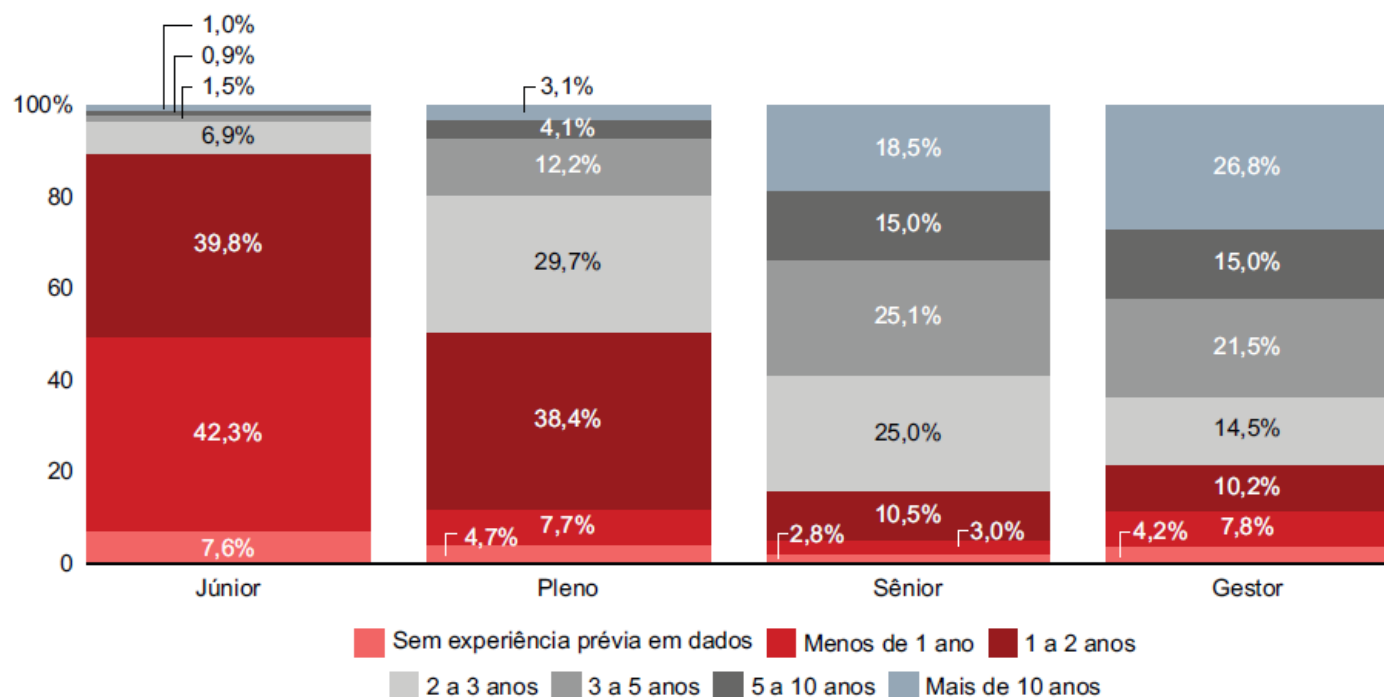
“Outros” compondo 8,3% do total de respondentes, enquanto em 2021 esse número era de 3,3%

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Formação e experiência

Figura 7: Anos de experiência em dados por nível de cargo



Profissionais de nível júnior possuem em sua maioria até 2 anos de experiência (**89,7%**)

Para o nível pleno, a maioria dos profissionais possui entre 1 e 4 anos de experiência (**68,1%**)

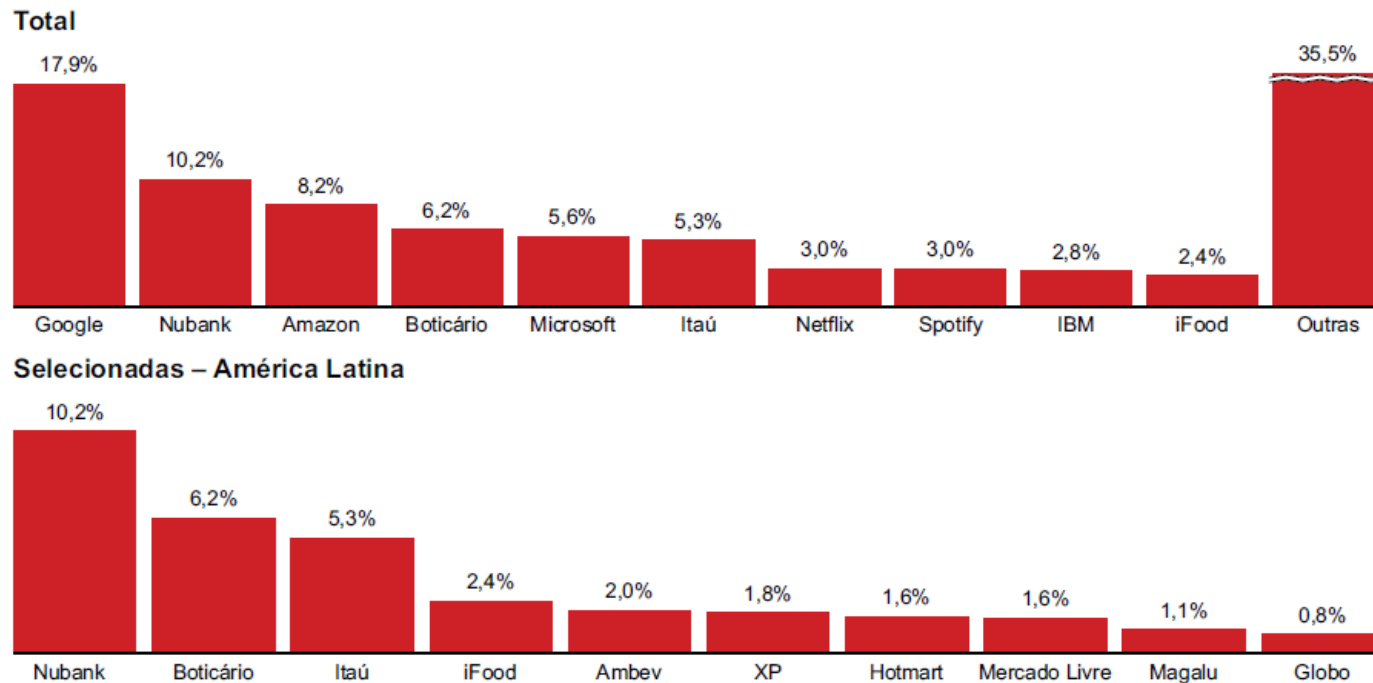
26,8% dos gestores possuem mais de 10 anos de experiência na área de dados

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Formação e experiência

Figura 15: Empresas mais admiradas para atuação na área de dados



A empresa mais admirada por sua atuação em dados é o **Google**, com **17,9%** das menções

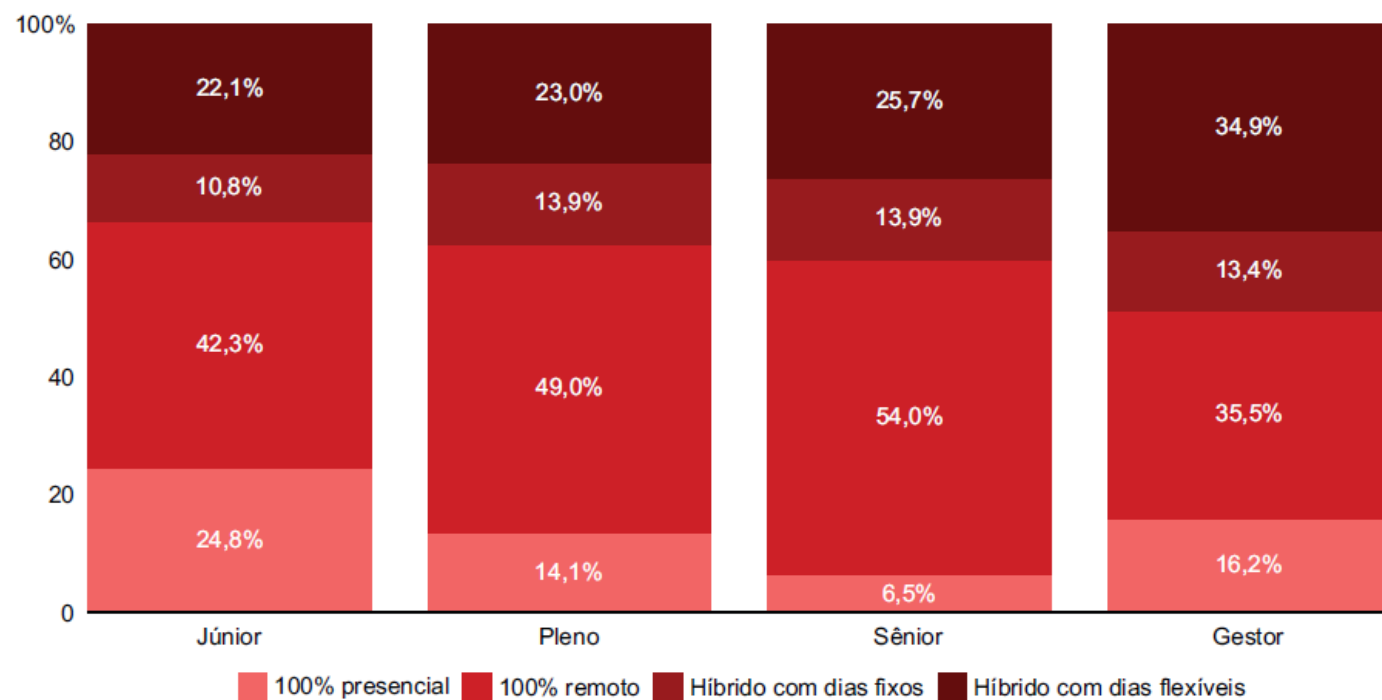
Na América Latina, as empresas mais admiradas são Nubank, Boticário, Itaú, iFood, Ambev, Hotmart, etc...

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Trabalho remoto

Figura 16: Modalidade de trabalho atual por nível de cargo



O nível **júnior** é o que possui maior proporção de profissionais em trabalho **presencial** (24,8%)

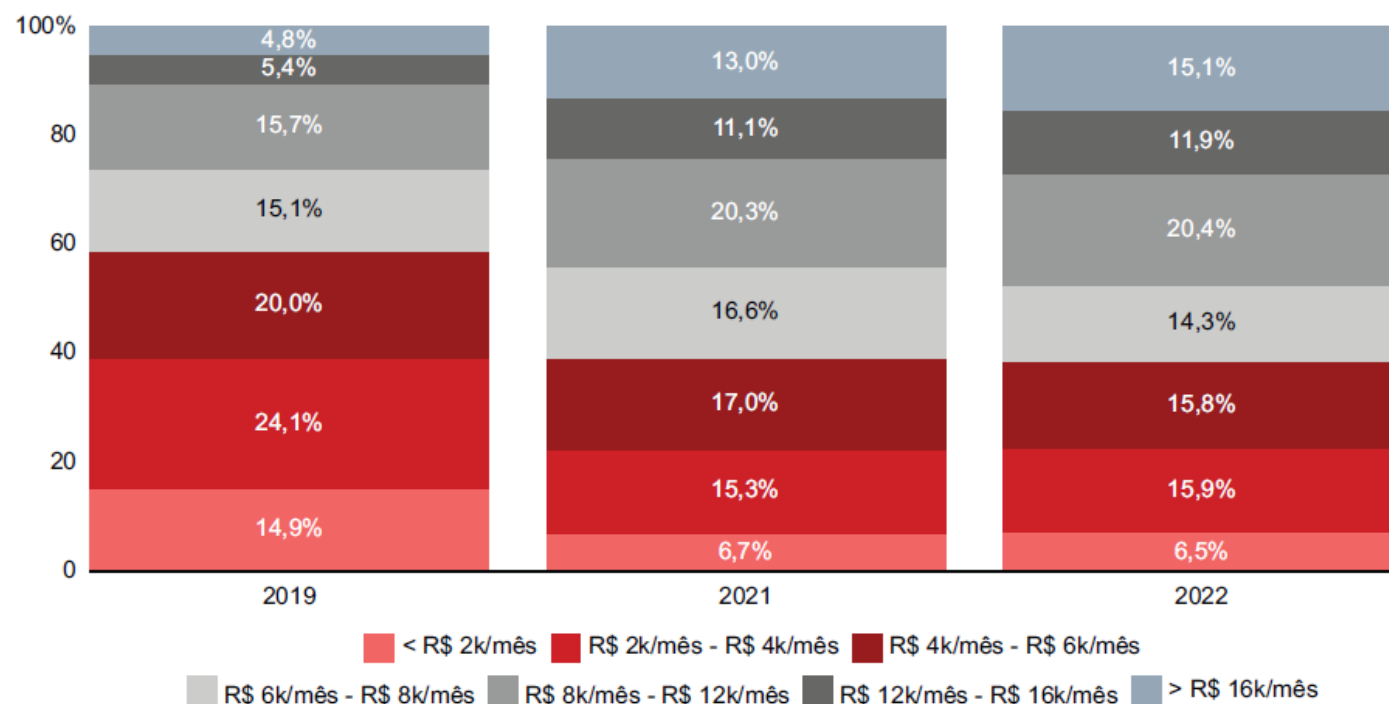
O nível de **gestão** possui a maior proporção de trabalho **híbrido** (48,3%)

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Remuneração

Figura 23: Comparação da remuneração entre 2019 e 2022



Aumento médio entre 2019 e 2021 foi de cerca de 40%

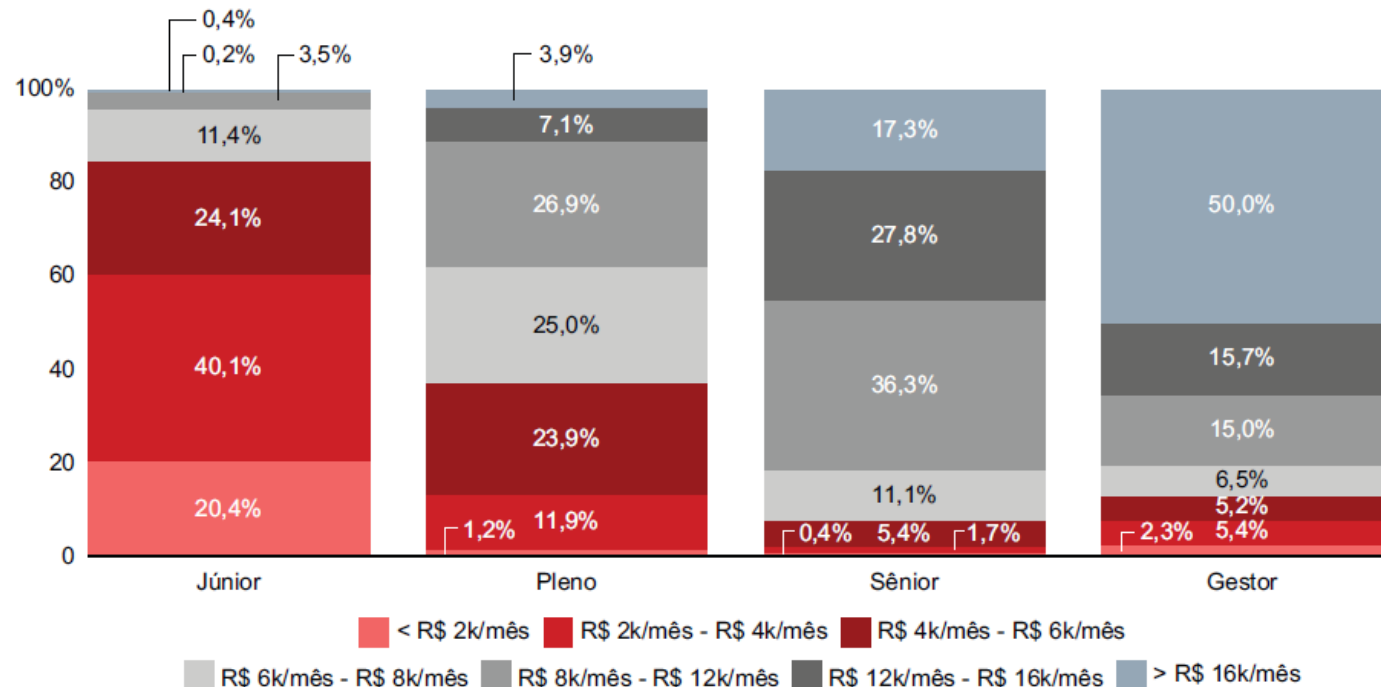
A variação média de 2022 para 2021 foi de cerca de 4%

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Remuneração

Figura 24: Remuneração em 2022 por nível de cargo



A maior concentração de remuneração nos cargos de **nível júnior** se dá nas faixas de até **R\$ 4 mil/mês (60,5%)**

O **nível sênior** possui parcela expressiva com mais de **R\$ 12 mil/mês (45,1%)**

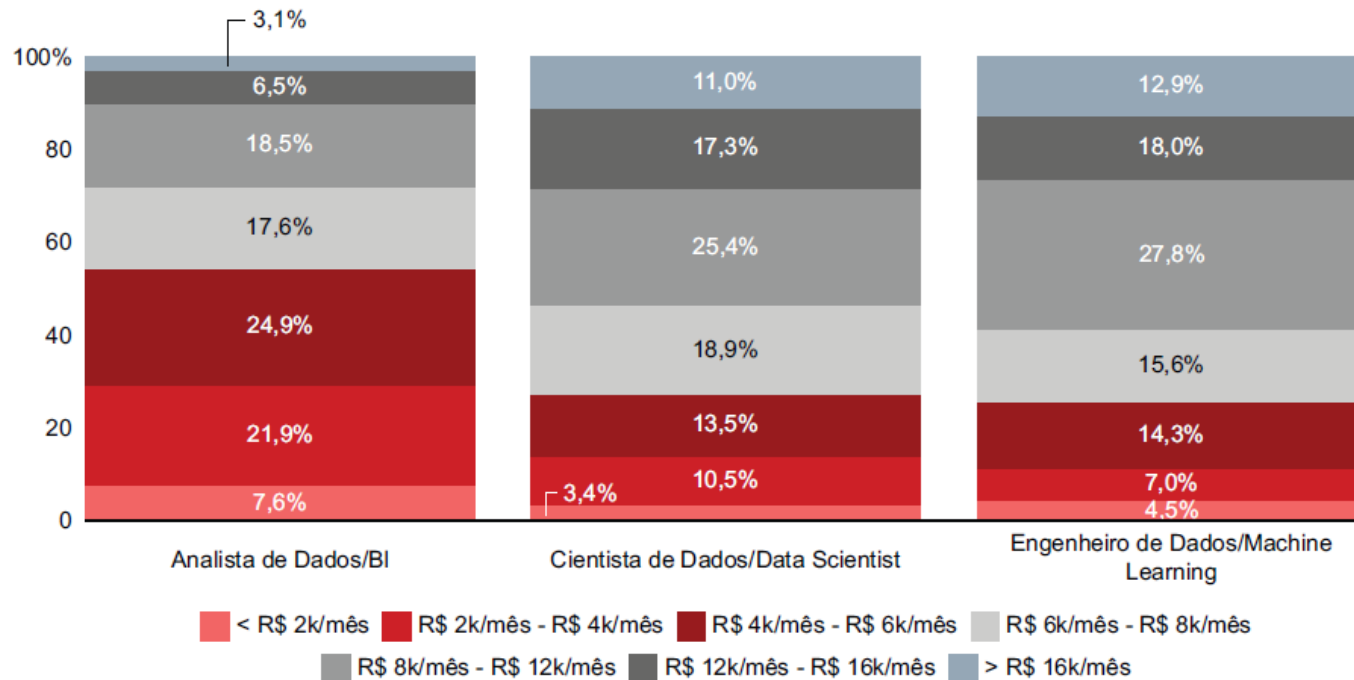
No nível de gestão os profissionais com remuneração acima de **R\$ 12 mil/mês alcançam 65,7%**

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Remuneração

Figura 26: Comparação da remuneração por função



Analistas de dados estão nas faixas mais baixas de remuneração

9,6% dos analistas de dados ganham acima de R\$ 12 mil/mês

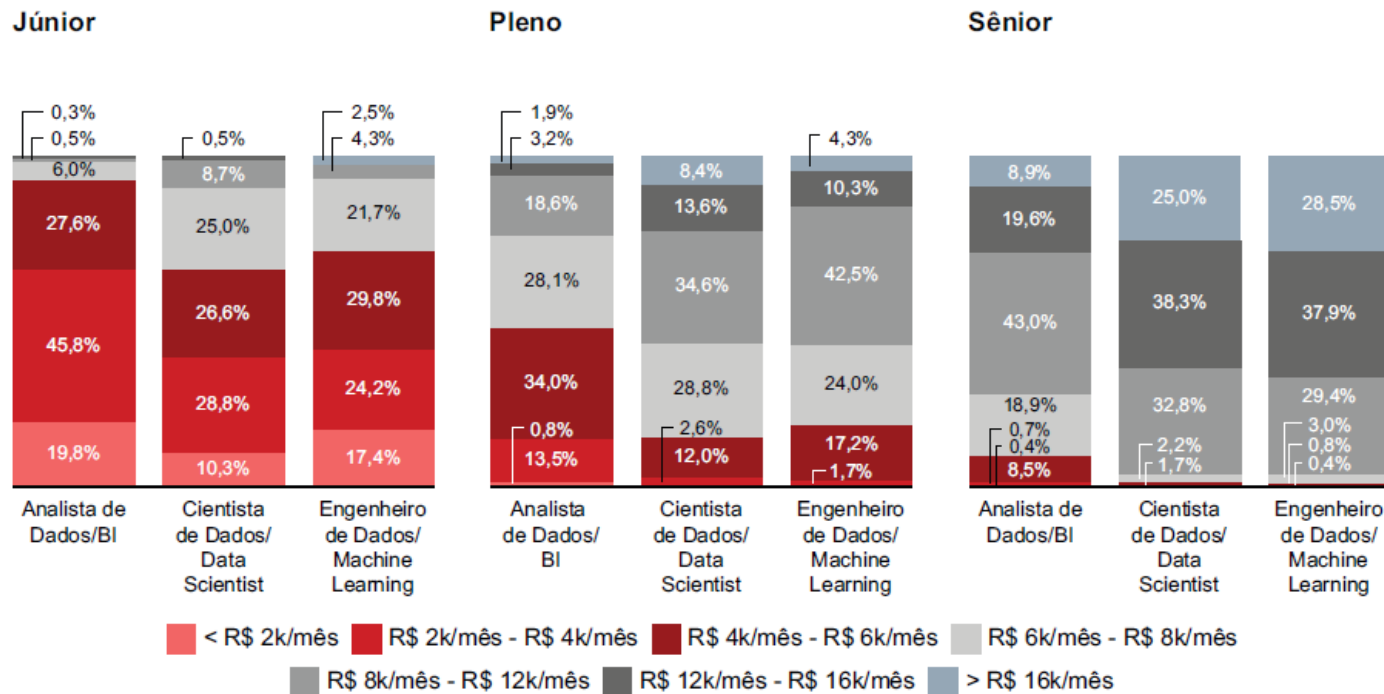
Essa proporção sobe para 28,3% para cientistas de dados e 30,9% para engenheiros de dados

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Remuneração

Figura 27: Comparação da remuneração por função e nível de cargo



A análise por função mostra que a dispersão de remuneração aumenta com o nível de cargo

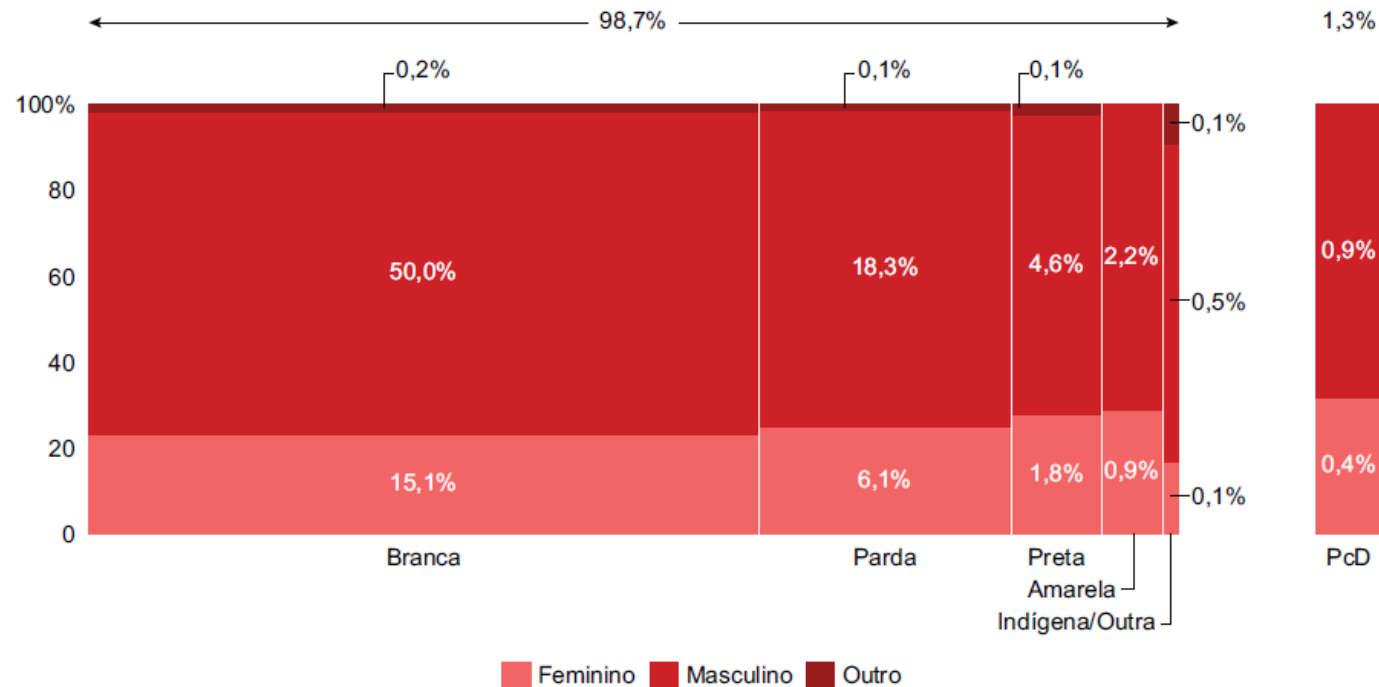
Analistas de dados continuam tendo remuneração inferior

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Diversidade

Figura 28: Distribuição de respondentes por gênero, cor/raça/etnia e PcD da amostra



As **mulheres** ainda são apenas **24,5%** dos profissionais de dados

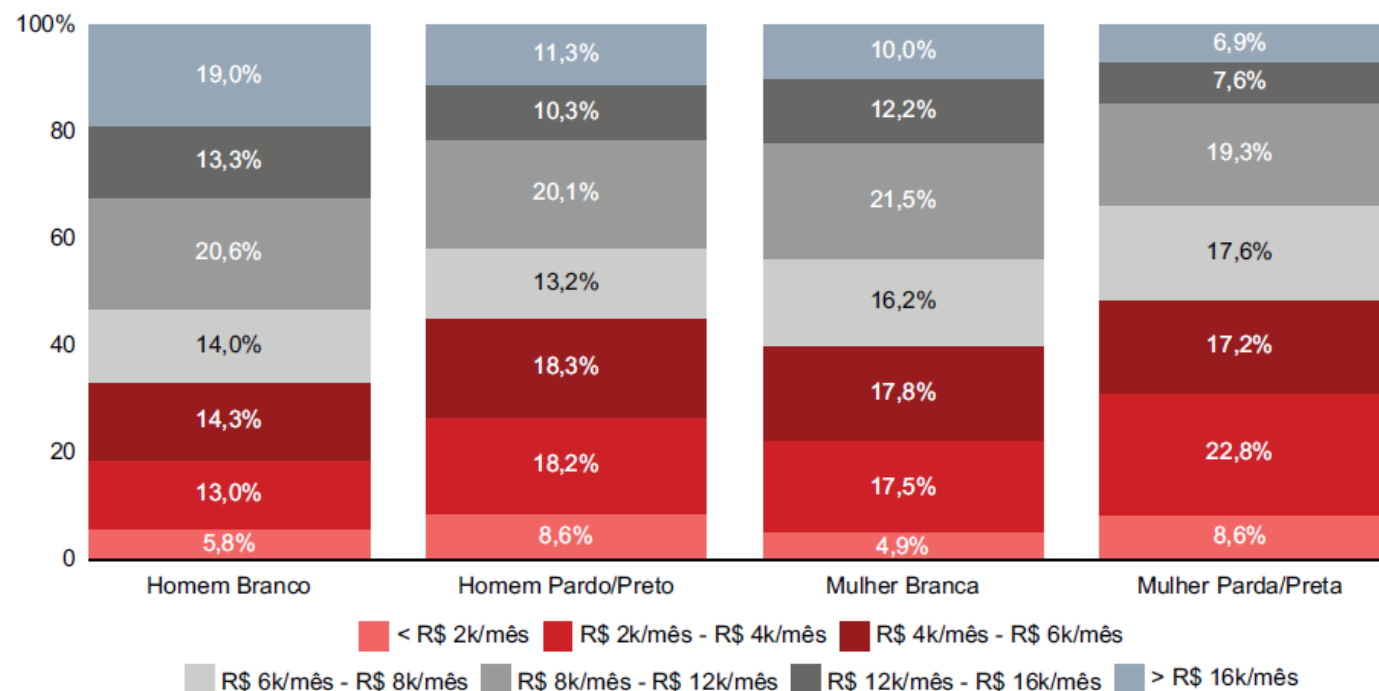
Pardos e pretos, cuja proporção na população brasileira é de 47% e 9,1% respectivamente, representam apenas **24,5% e 6,6%** na amostra

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Mercado de Trabalho no Brasil

Diversidade

Figura 33: Distribuição de remuneração por interseccionalidade de gênero e cor/raça/etnia



19,0% dos homens brancos na faixa de mais de R\$ 16 mil/mês

Homens pardos e pretos esse número é de 11,3% dos respondentes

Mulheres brancas é de 10% e para mulheres pardas e pretas esse número é de apenas 6,9%

Nota: Desconsidera respondentes que se consideram desempregados, apenas estudantes e que preferiram não informar a situação atual de trabalho
Fonte: Pesquisa State of Data Brasil 2022

Bibliografia Básica

- FAWCETT, T.; PROVOST, F. **Data Science para Negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados**. Alta Books Editora, 2018.
- MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big data: A revolution that will transform how we live, work, and think**. Houghton Mifflin Harcourt, 2013.
- GRUS J. **Data science from scratch: first principles with python**. O'Reilly Media, 2019.
- CARVALHO, A. et al. **Inteligência Artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC. 2011.

Outras Referências

- **Data Science History and Overview.** Disponível em: <https://www.kdnuggets.com/2020/11/data-science-history-overview.html>. Acesso: 15 de março de 2021;
- **Defining Data Science: The What, Where and How of Data Science.** Disponível em: <https://365datascience.com/career-advice/career-guides/defining-data-science/>. Acesso 15 de março de 2021;
- **Data Science: A High-Level Overview.** Disponível em: <https://www.linkedin.com/pulse/data-science-high-level-overview-giuliano-liguori-/>. Acesso: 15 de março de 2021;
- **What is Big Data Analytics and How It Helps You Understand Your Customers. .** Disponível em: <https://financesonline.com/what-is-big-data-analytics-and-how-it-helps-you-understand-your-customers/>. Acesso: 22 de março de 2021
- **State of data 2002 - Um raio-x dos profissionais de dados do Brasil.** Disponível em: <https://www.stateofdata.com.br/>. Acesso: 23 de fevereiro de 2023;