

Jornada DATA SCIENCE ÔMEGA

Você Cientista de dados em 2022



Ômega Data Science

Escola online de Data Science

Missão

Impactar carreiras e negócios com uma formação íntegra e fundamentada.

Valores

Fundamentação científica.

Didática.

Prática.

Autonomia do aluno.

Ensino humanizado.



► Ômega Data Science

Escola online de Data Science

Visão

- ▶ Ser o maior centro de excelência na formação de Cientistas de Dados do Brasil.
- ▶ Oferecer um diversificado portfólio de cursos com forte fundamentação estatística, matemática e prática de programação.
- ▶ Incorporar metodologias e ferramentas pedagógicas inovadoras para acelerar o desenvolvimento profissional de nossos alunos.
- ▶ Exercer o ensino humanizado com acompanhamento e integridade.

Professores



Walmes Zeviani
Professor na Ômega

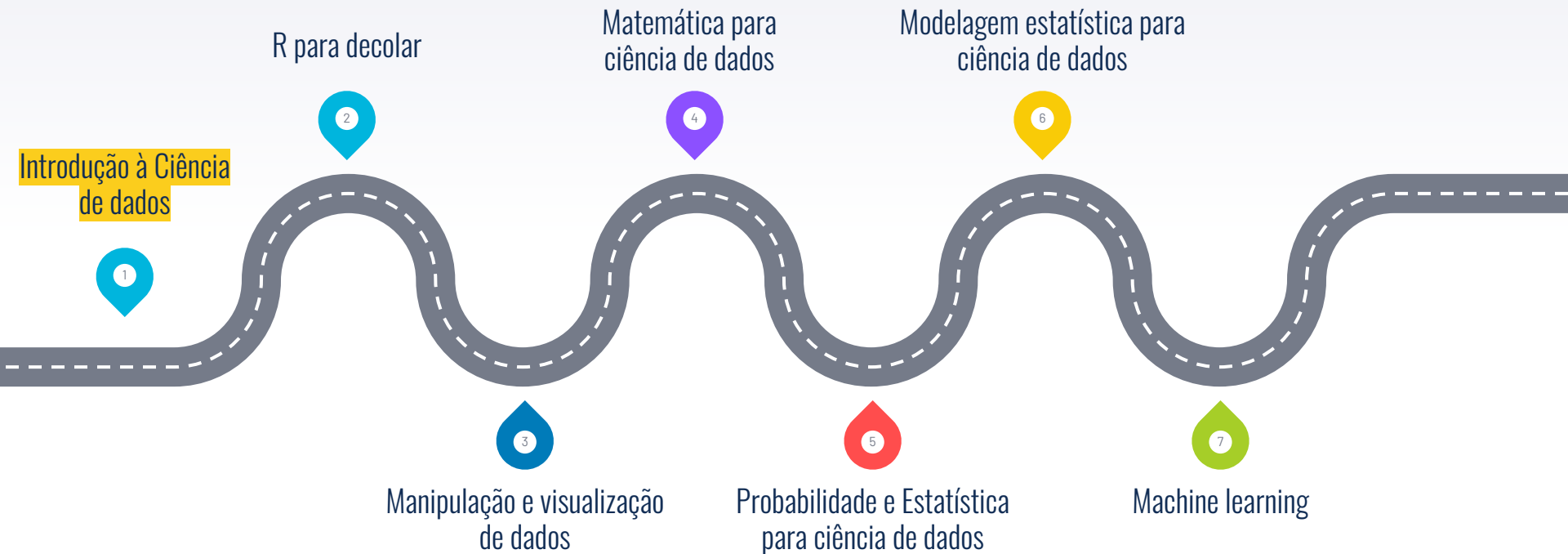


Ômega Data Science
PLATAFORMA DE CURSOS
ONLINE



Wagner Bonat
Professor na Ômega

Jornada em Data Science Ômega



Aproveite a Jornada e construa o seu portfólio

INTRODUÇÃO À CIÊNCIA DE DADOS

Sua jornada começa aqui



► Estrutura do curso

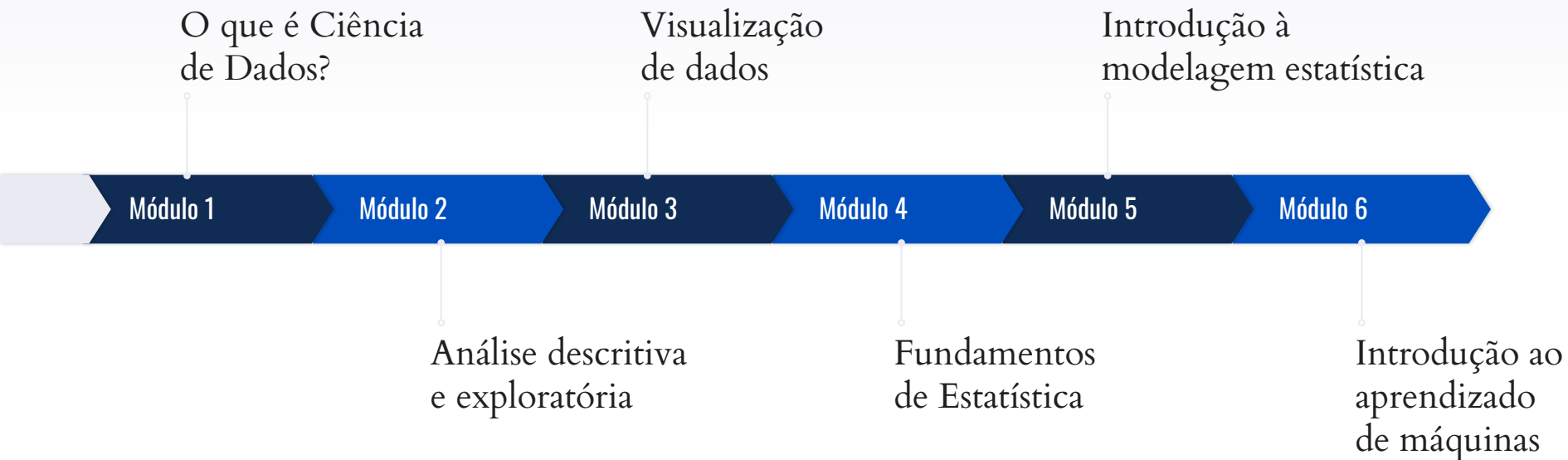
1. O que é ciência de dados?
2. Análise descritiva e exploratória.
3. Visualização de dados.
4. Fundamentos de estatística.
5. Introdução à modelagem estatística.
6. Introdução ao aprendizado de máquina.

Detalhes

- ▶ Carga horária: 18 horas.
- ▶ Nível: Iniciante.
- ▶ Pré-requisito: Não há.
- ▶ Público-alvo: Profissionais de todas as áreas.



Programação do curso



Aspectos operacionais

Vídeos ficam gravados?

O curso todo ficará disponível na plataforma **Sympla Play**. Ver detalhes para se inscrever.

Onde consigo ajuda?

Ao se inscrever em qualquer curso você tem acesso a Comunidade Ômega no **Discord**.

Certificado de participação

Todos os participantes que realizarem ao 75% das tarefas do curso terão seu certificado de participação emitido imediatamente após o curso.

Certificado de performance

Após cada módulo no Sympla Play teremos uma avaliação opcional de performance. Para os que responderem todas em até 30 dias após o término do curso um histórico será gerado e enviado por e-mail.

*Give me six hours to chop down a tree and I will spend
the first four hours sharpening the axe.*

—Abraham Lincoln



MÓDULO 1

O QUE É CIÊNCIA DE DADOS?

Objetivo

Dar uma visão geral da profissão, como atua, onde trabalha, quanto ganha, qual a procura.

Carreiras em dados

Quais os tipos de carreiras que existem e o que elas exigem.

Como estruturar problemas

Como se organiza um projeto de ciência de dados, quais os principais problemas, tarefas e métodos.

Como começar

Qual o caminho mais promissor para iniciar em ciência de dados, principalmente para quem vem de outra área.



MÓDULO 1

O QUE É CIÊNCIA DE DADOS?

1. O que é a Era do Big Data e Analytics?
2. O que é Ciência de Dados?
3. Profissões relacionadas a ciência de dados.
4. Estatísticas sobre a profissão e o mercado de trabalho.
5. Como se tornar um Cientista de Dados?
6. Estruturação das atividades: problemas, tarefas e métodos.
7. Dicas de estudo e portfólio.
8. Projeto I: Escrevendo sobre Ciência de Dados.





t.me/omegadatascience

PARTICIPAR



@omegadatascience

SEGUIR



/OmegaDataScience

INSCREVER



omegadatascience

SEGUIR



@omegadatascienc

SEGUIR



/omegadatascience

SEGUIR

MÓDULO 2

ANÁLISE EXPLORATÓRIA DE DADOS

Objetivo

Motivar a importância da AED, suas principais técnicas e benefícios e como pode ser feita usando o R.

Pré-processamento

Tipos de variáveis e problemas relacionados à qualidade de dados.

Medidas descritivas

Como resumir o comportamento dos dados e compreendê-los.

O CRISP-DM

O exemplo de um ciclo de atividades dentro de um projeto de ciência de dados.



MÓDULO 2

ANÁLISE EXPLORATÓRIA DE DADOS

1. Tipos de variáveis e sua importância.
2. Problemas com a qualidade dos dados.
3. Manipulação de dados.
4. Tarefas de limpeza e preparação.
5. Tarefas de transformação e agregação.
6. Medidas descritivas: posição, dispersão e associação.
7. Prática de ETL com R.
8. Projeto II: Preparando dados para análise.



MÓDULO 3

VISUALIZAÇÃO DE DADOS

Objetivo

Mostrar como a visualização de dados é uma ferramenta importante para a tomada de decisão e desenvolvimento de projetos em ciência de dados.

Princípios de visualização

Conheça quais são as principais decisões que você deve tomar para escolher a melhor forma de visualizar os seus dados.

A função dos gráficos

Como apresentar informação contida nos dados de forma gráfica, qual o melhor gráfico para cada explanação.

Data storytelling

Como fazer narrativas baseadas em métricas e gráficos para comunicar resultados.



MÓDULO 3

VISUALIZAÇÃO DE DADOS

1. A visualização de dados e sua importância.
2. Princípios de visualização de dados.
3. Gráficos e suas funções.
4. Landscape de visualização de dados no R.
5. Gráficos interativos.
6. Dashboards.
7. Prática de relatórios dinâmicos com R.
8. Projeto III: Apresentação e visualização de dados.



MÓDULO 4

FUNDAMENTOS DE ESTATÍSTICA

Objetivo

Dar uma visão geral do pensamento estatístico, seus pilares e principais conceitos.

Problemas de negócios

Entenda como a estatística é usada para resolver problemas e criar novos negócios.

O pensamento analítico

Compreenda como a estatística modela fenômenos aleatórios e permite tomar melhores decisões baseadas em dados.

Inferência estatística

Aprenda como a ciência estatística é construída por meio de um projeto de dados.



MÓDULO 4

FUNDAMENTOS DE ESTATÍSTICA

1. O problema de negócio.
2. O pensamento estatístico.
3. Tomando decisões na presença da incerteza.
4. O desenvolvimento de um produto baseado em dados.
5. Distribuição amostral e a inferência estatística.
6. Separando fontes de variação.
7. Projeto IV: Identificando oportunidades.



MÓDULO 5

INTRODUÇÃO À MODELAGEM ESTATÍSTICA

Objetivo

Motivar as aplicações dos modelos estatísticos, o que são e para que servem.

Componentes

O modelo estatístico e suas partes, como funcionam e como estão conectadas as classes de modelos.

Estimação

Como um modelo é ajustado aos dados e o que isso envolve.

Usando o modelo

De interpretações à predições, como usar modelos para tomar decisões e desenvolver produtos com dados.



MÓDULO 5

INTRODUÇÃO À MODELAGEM ESTATÍSTICA

1. Modelos estatísticos e seus componentes.
2. Objetivos do uso de um modelo.
3. Especificação e ajuste de modelos.
4. Interpretando os resultados e a tomada de decisões.
5. Diferentes tipos de modelos e como escolher o adequado.
6. Projeto V: Especificando e ajustando um modelo.



MÓDULO 6

INTRODUÇÃO AO APRENDIZADO DE MÁQUINAS

Objetivo

Motivar as aplicações de modelos de machine learning e em que difere dos modelos estatísticos.

Tipos de aprendizado

De que forma pode-se aprender com os dados.

Algoritmos

Principais métodos ou algoritmos e os problemas que eles resolvem.

Construir modelos

Como desenvolver um modelo de aprendizado de máquinas.



MÓDULO 6

INTRODUÇÃO AO APRENDIZADO DE MÁQUINAS

1. Tipos de aprendizado de máquina.
2. Etapas de um aprendizado supervisionado.
3. Regressão e classificação.
4. Métricas.
5. Tópicos de aprendizado não supervisionado.
6. Prática com o tidymodels.
7. Projeto VI: Desafiando meu modelo.



Vamos começar!

Telegram: t.me/omegadatascience

Instagram: [@omegadatascience](https://www.instagram.com/omegadatascience)

Twitter: [@omegadatascience](https://twitter.com/omegadatascience)

YouTube: [/OmegaDataScience](https://www.youtube.com/OmegaDataScience)



Ômega Data Science
PLATAFORMA DE CURSOS
ONLINE

omegadatascience.com.br

Era do Big Data e Analytics

O que a Era do Big Data e Analytics significa?

Capacidade de coletar dados em maior **volume, variedade e velocidade** associado a capacidade científica e computacional de analisá-los.

Como podemos nos preparar para tirar proveito dela?

Apropriar de métodos científicos e computacionais que nos permitam **extrair o máximo de informação** de grandes massas de dados.



Dados sempre estiveram entre nós! O que mudou?

- ▶ Censo demográfico brasileiro (1872).
- ▶ Índices financeiros (todo dia na TV).
- ▶ Pesquisas eleitorais.
- ▶ Desenvolvimento de tecnologias.
- ▶ Crescente número de dispositivos inteligentes.
- ▶ Volume, variedade e velocidade que geramos e armazenamos dados mudou!



Estatísticas impressionantes



1

Em 2021, uma pessoa em média gerou 1.7MB de dados a cada segundo.

2

Em 2022, 70% do PIB virá da economia digital.

3

Todo dia 306.4 bilhões de emails são enviados e 500 milhões de Tweets postados.

4

Em janeiro de 2021 a internet tinha 4.66 bilhões de usuários ativos.

5

Para 2030 é esperado 25.44 bilhões de dispositivos IoT.

O que mais está mudando?

- ▶ Dados estruturados (usuais).
- ▶ Textos.
- ▶ Imagens/fotos.
- ▶ Audio.
- ▶ Videos.
- ▶ Redes de relacionamento.
- ▶ Localização.
- ▶ User experience.



Analytics?

- ▶ Data is the new gold!
- ▶ Data is the new oil!

Será?

- ▶ Dados só tem valor se forem analisados!



Aplicações de analytics



Estratégias de marketing



Sistemas educacionais



Detecção de fraudes



Experiência do usuário

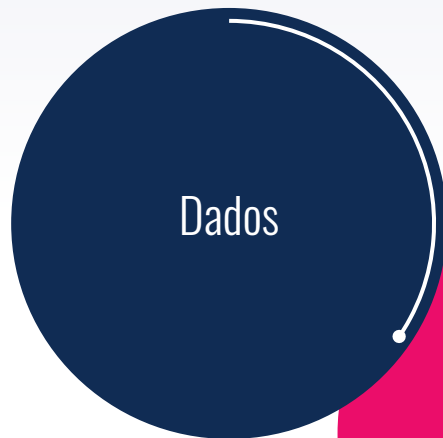


Sistemas de recomendação



People analytics

► Data Science






► O que é Ciência de Dados?

Ciência de dados é o **estudo científico** da criação, validação e transformação de dados para criar **significado/valor**.

Cientistas de dados é o profissional que usa **métodos científicos** para entender e criar **significado/valor** a partir de dados brutos.

50 Best Jobs in America for 2021

Best Jobs: 2021: United States:

Share    

	Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1	Java Developer	\$90,830	4.2/5	10,103	View Jobs
#2	Data Scientist	\$113,736	4.1/5	5,971	View Jobs
#3	Product Manager	\$121,107	3.9/5	14,515	View Jobs
#4	Enterprise Architect	\$131,361	4.0/5	10,069	View Jobs
#5	Devops Engineer	\$110,003	4.0/5	6,904	View Jobs

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)

► O que significa gerar valor?

Regras do jogo

1. Temos duas máquinas M1 e M2.
2. Cada máquina sorteia um número entre 1 e 6.
3. Para ganhar você deve acertar qual o número a máquina irá sortear.

Que número você escolhe para cada máquina?

M1



M2



► O que significa gerar valor?

Regras do jogo

1. Temos duas máquinas M1 e M2.
2. Cada máquina sorteia um número entre 1 e 6.
3. Para ganhar você deve acertar qual o número a máquina irá sortear.

Que número você escolhe para cada máquina?

Vou te ajudar! E agora?

M1



4 2 3 4 5 4 5 4 5 4 1 1 1 1
3 1 2 3 1 6 2 1 1 3 6 4 4 5
3 1 5 6 5 5 2 2 4
6 4 5 5 6 5 3 5 5 2 4 3 6 1
1 6 5 6 1 1 2 3 1 4 4 6 2 2
6 6 4 3 4 5 2 2 3
3 5 3 6 4 4 3 5 4 1 4 4 1 3
3 2 2 6 1 3 6 3 5 1 5 3

M2



5 3 4 4 3 6 5 6 3 2 4 6 3 4
4 3 5 4 5 2 4 6 2 5 2 4 5 3
6 4 5 4 3 6 4 4 3
4 3 1 4 4 5 6 1 2 4 4 4 3
1 4 6 2 5 4 3 4 4 6 5 1 2 2
3 1 6 4 6 3 4 4 3
6 2 6 3 4 5 4 4 3 1 3 3 4 3
6 1 3 4 3 2 2 4 6 4 2 6

► O que significa gerar valor?

Regras do jogo

1. Temos duas máquinas M1 e M2.
2. Cada máquina sorteia um número entre 1 e 6.
3. Para ganhar você deve acertar qual o número a máquina irá sortear.

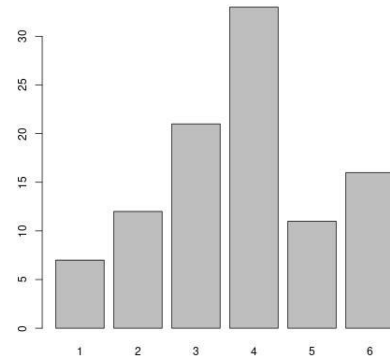
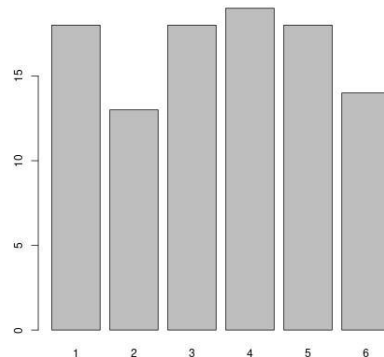
Que número você escolhe para cada máquina?

Agora vou realmente te ajudar!

M1



M2

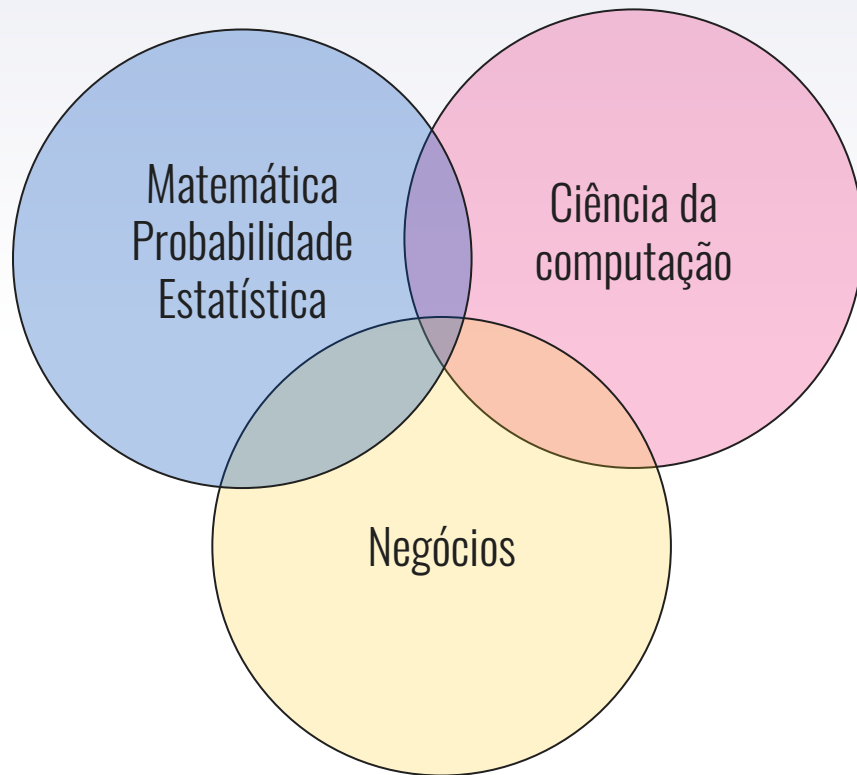


Ciência de dados

Ciência de dados é mais que analisar dados, é criar/aprimorar produtos baseados em dados!

“Person who is better at statistics than any software engineer and better at software engineering than any statistician”

– Josh Wills.



Áreas envolvidas em Ciência de Dados



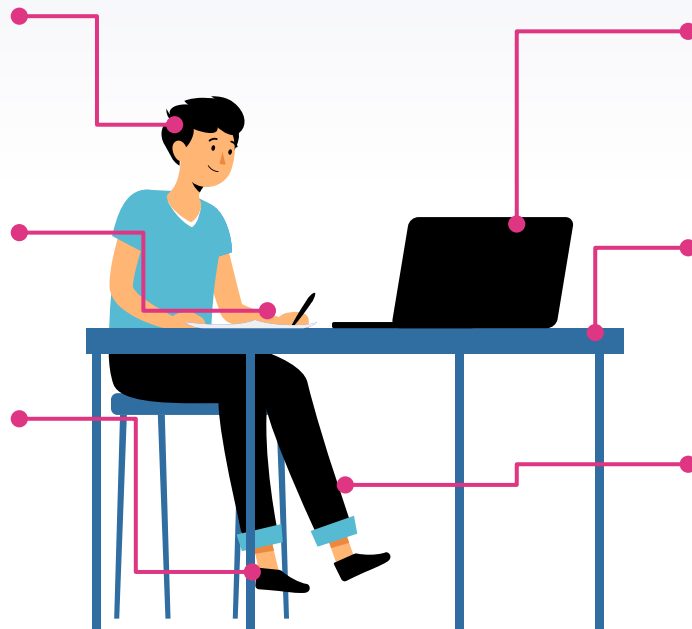
Análise de dados



Big data



Inteligência de negócios



Notebooks e IDEs



Linguagens de programação



Soft skills



Profissões relacionadas

Engenheiro de dados



Constrói e otimiza os sistemas que permitem o cientista e o analista de dados trabalharem.

SQL, Python, R, Cloud.

Analista de dados



Gera valor analisando e comunicando resultados. Ajuda a tomar decisões de negócios.

Comunicação, storytelling, visualização.

Cientista de dados



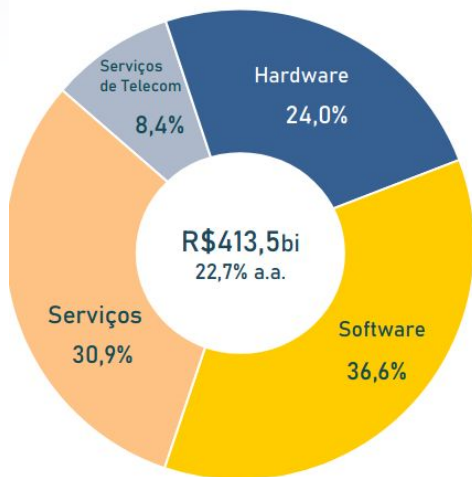
Usa estatística e algoritmos de ML para fazer previsões e responder questões chaves do negócio.

R, Python, SQL, Cloud.

Mercado de trabalho

Perspectivas de investimento de 2021-2024 (R\$ bilhões)

Tecnologias de Transformação Digital



Nuvem

R\$ 181,9 bi | 28% a.a.



Robótica

R\$ 31,4 bi | 6% a.a.



Internet das Coisas

R\$ 74,3 bi | 29% a.a.



Realidade Virtual

R\$ 2,2 bi | 40% a.a.



Big Data & Analytics

R\$ 77,2 bi | 13% a.a.



Redes Sociais

R\$ 14,1 bi | 19% a.a.



Segurança da Informação

R\$ 16,2 bi | 13% a.a.



Impressão 3D

R\$ 0,7 bi | 15% a.a.



Inteligência Artificial

R\$ 11,5 bi | 31% a.a.

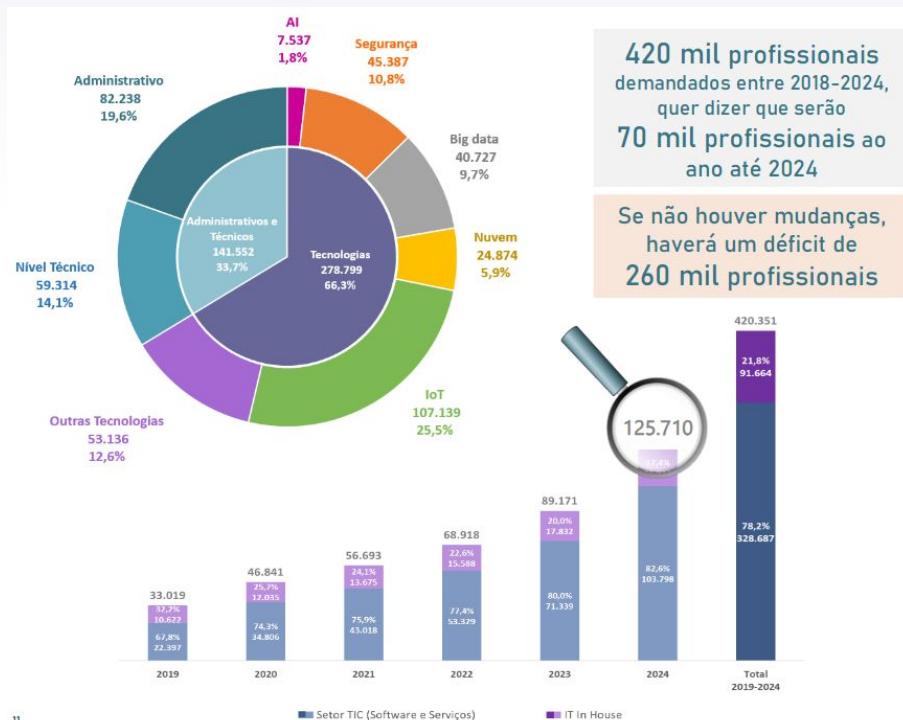


Blockchain

R\$ 3,8 bi | 65% a.a.

Mercado de trabalho

Demanda de profissionais



Mercado de trabalho

Remuneração

Salários de Engenheiro De Dados em Republica Federativa do Brasil

Atualizado em 7 de out. de 2021

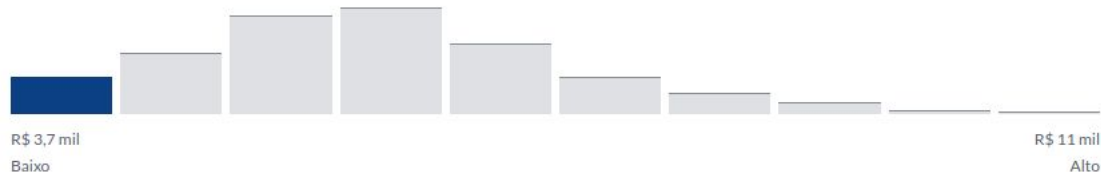
 Confiança **muito alta**

R\$ 7.631 /mês

Salário médio

519 salários

Nenhuma remuneração adicional foi relatada para essa função



Quanto ganha um Engenheiro De Dados?

O salário médio nacional de um Engenheiro De Dados é de R\$ 7.631 por mês em Brasil. Filtre por localização para ver quanto ganha um Engenheiro De Dados na sua região. As estimativas de salários têm como base 519 salários enviados de forma sigilosa ao Glassdoor por pessoas com o cargo de Engenheiro De Dados.

Mercado de trabalho

Remuneração

Salários de Analista De Dados em Republica Federativa do Brasil

Atualizado em 7 de out. de 2021

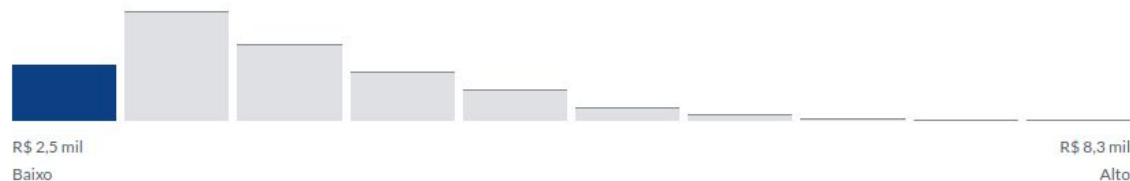
 Confiança **muito alta**

R\$ 4.120 /mês

Salário médio

521 salários

Nenhuma remuneração adicional foi relatada para essa função



Quanto ganha um Analista De Dados?

O salário médio nacional de um Analista De Dados é de R\$ 4.120 por mês em Brasil. Filtre por localização para ver quanto ganha um Analista De Dados na sua região. As estimativas de salários têm como base 521 salários enviados de forma sigilosa ao Glassdoor por pessoas com o cargo de Analista De Dados.

Mercado de trabalho

Remuneração

Salários de Cientista De Dados em Republica Federativa do Brasil

Atualizado em 19 de out. de 2021

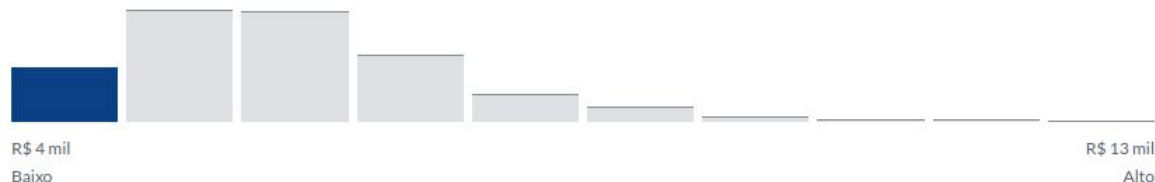
 Confiança **muito alta**

R\$ 7.122 /mês

Salário médio

901 salários

Nenhuma remuneração adicional foi relatada para essa função



Quanto ganha um Cientista De Dados?

O salário médio nacional de um Cientista De Dados é de R\$ 7.122 por mês em Brasil. Filtre por localização para ver quanto ganha um Cientista De Dados na sua região. As estimativas de salários têm como base 901 salários enviados de forma sigilosa ao Glassdoor por pessoas com o cargo de Cientista De Dados.

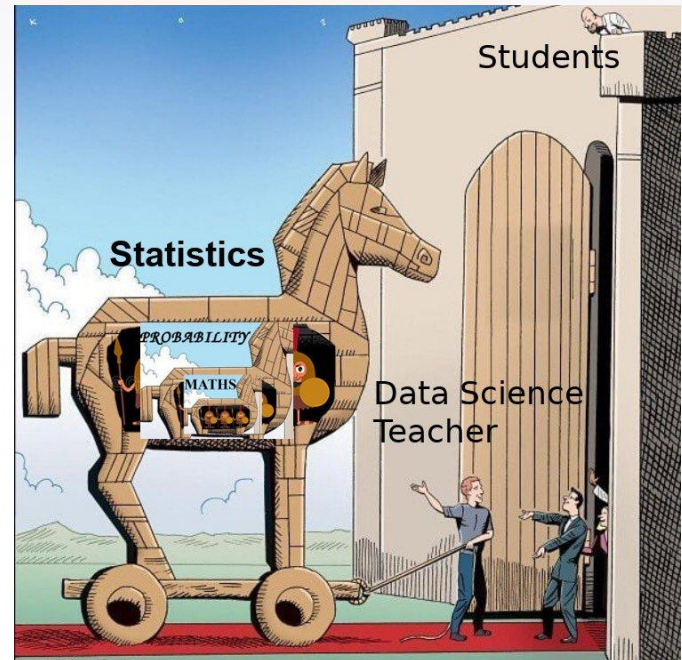
Qual é a formação de um Cientista de Dados?

- ▶ Raros e iniciantes cursos de graduação.
- ▶ Pós-graduações *lato sensu*.
- ▶ Pós-graduações *stricto sensu*.
- ▶ Plataformas de competição (kaggle).
- ▶ Cursos online de aprimoramento.

Desafio!

- ▶ Como formar Cientistas de Dados?
- ▶ Fundamentos

Matemática, Estatística e Programação.



► Técnicas populares

- ▶ Regressão linear múltipla.
- ▶ Modelos lineares generalizados.
- ▶ Árvores de decisão, random forest, ensemble of models.
- ▶ Bagging, boosting, neural networks.
- ▶ Natural language process.
- ▶ Deep learning.
- ▶ Data storytelling.
- ▶ Dashboards e muito mais.

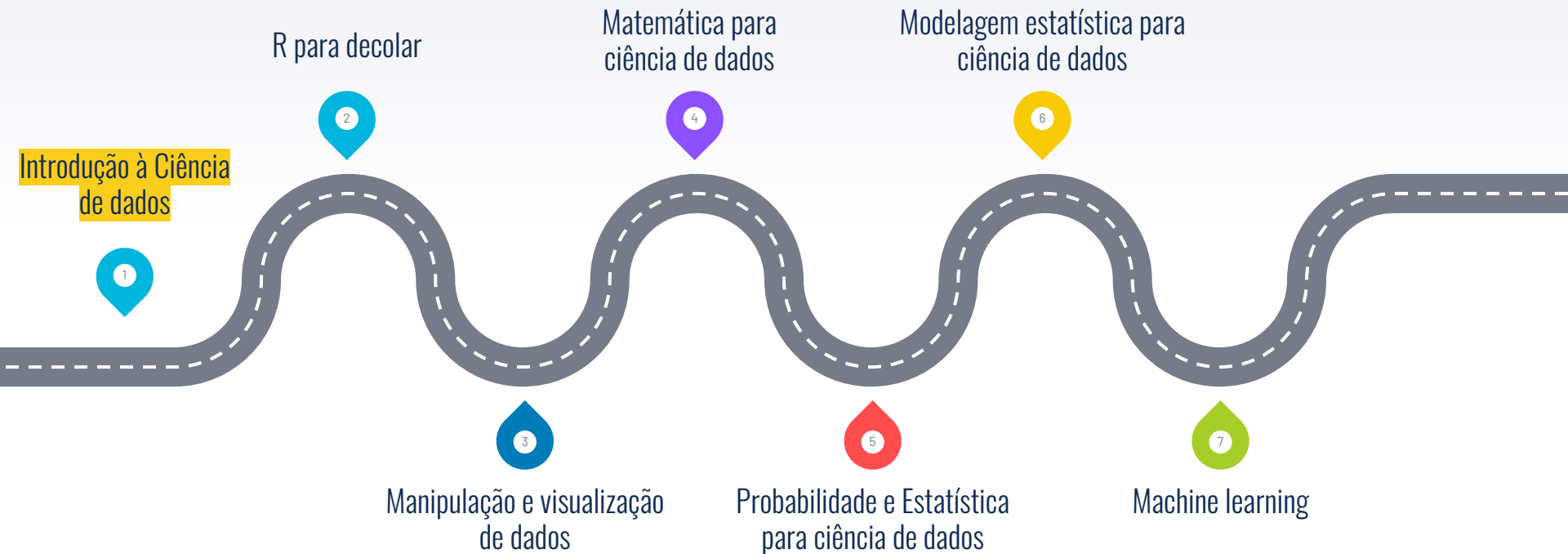


Como que vou aprender tudo isso?

1. Planejamento e disciplina.
2. Foco nos fundamentos e prática.
3. Ferramentas que te ajudem a aprender.
4. Suporte de profissionais experientes.



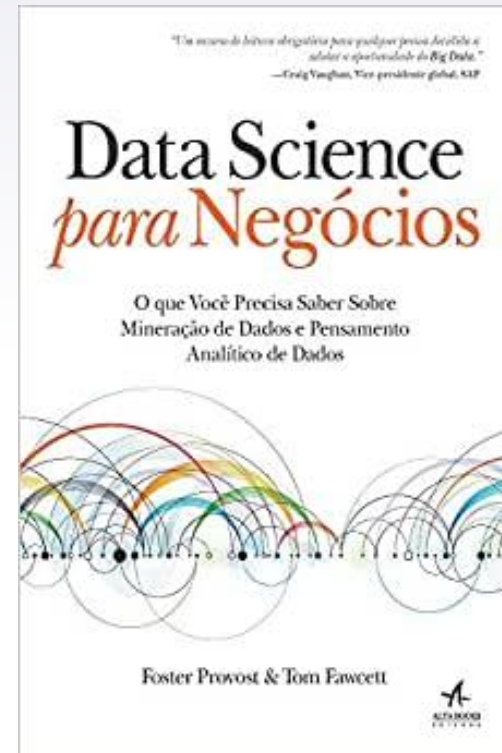
Jornada em Data Science Ômega



Aproveite a Jornada e construa o seu portfólio

Casos clássicos

- ▶ O furacão Frances e o Walmart.
- ▶ Target (varejista USA).
- ▶ Banco Signet (Capital One).



Desafios

Fontes de dados e ferramentas

Seja qual for o negócio em geral tem diversas bases de dados desorganizadas e desconectadas. Uso de diversas ferramentas de forma desordenada.

Falta de otimização

Apesar dos dados existirem em geral é difícil ter a informação certa para subsidiar a tomada de decisão.

Analytics expertise

Muitas empresas não têm pessoal qualificado para transformar dados em insights e ações de negócio.

Predictive insights

Apesar de conseguir analisar dados algumas empresas não conseguem transformá-los em insights de negócio.



► Como se entrega resultados?

Insights acionáveis

Prescrever o que deve ser feito
baseado no conhecimento
adquirido.

Informação

Entender o que aconteceu no
passado.

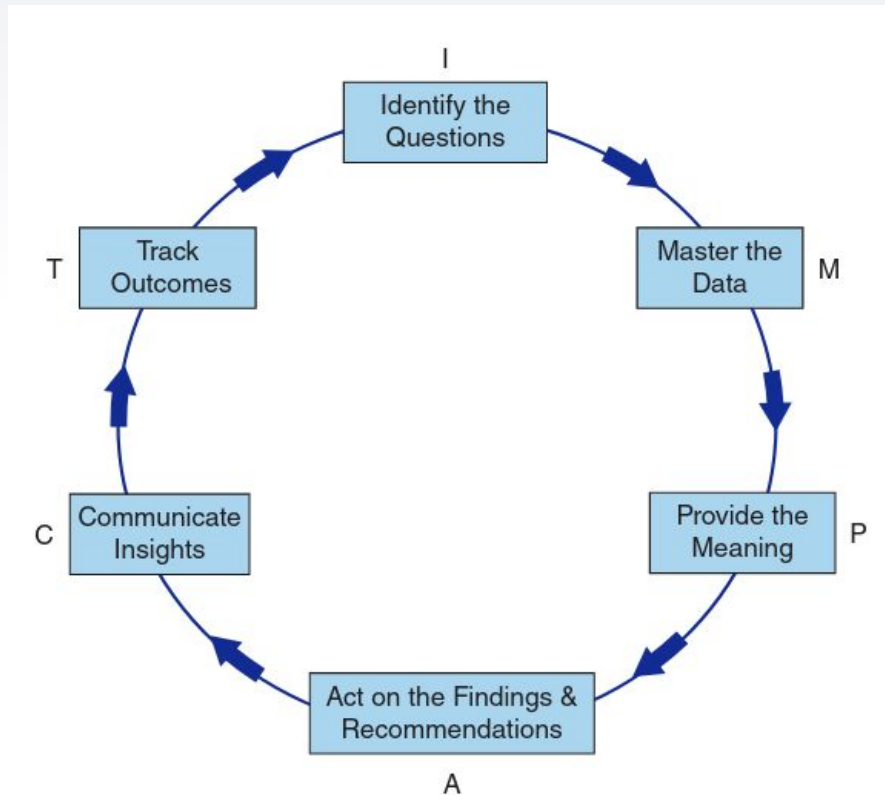
Conhecimento

Entender o que está acontecendo
agora e porque.

Inteligência

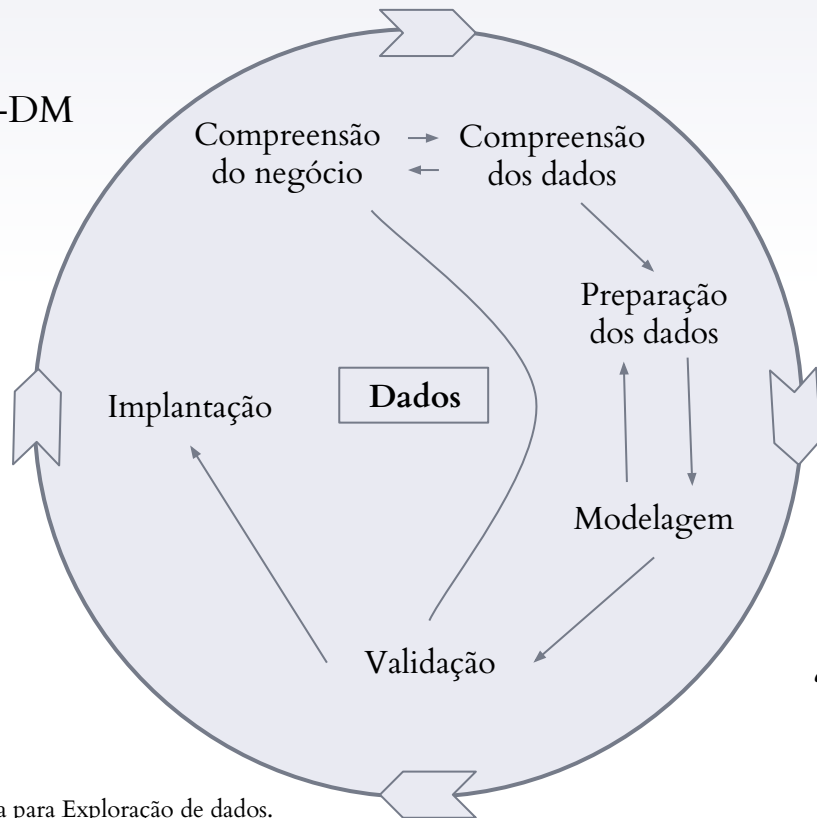
Antecipar o que vai acontecer no
futuro.

IMPACT



▶ Projetos de Ciência de Dados

Metodologia · CRISP-DM



Processo investigativo.
Não apenas construtivo.

“Falhar rápido para ter sucesso mais cedo!”

Tarefas mais populares



Classificação



Regressão



Agrupamento e
coocorrência



Previsão de vínculo



Redução de dados



Modelagem causal

Tipos de análises

Descritiva

O que aconteceu no passado?

Preditiva

O que vai acontecer e por que?

Diagnóstica

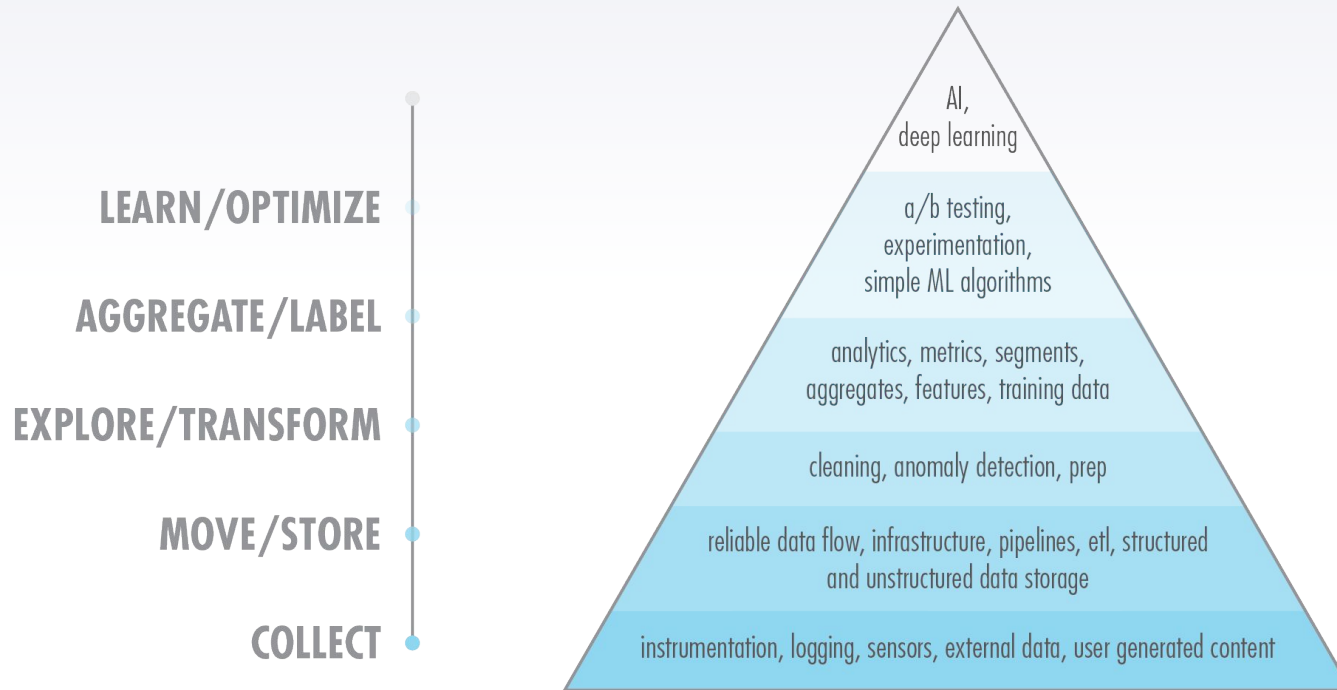
O que está acontecendo agora e por que?

Prescritiva

O que devemos fazer sabendo o que pode acontecer?



THE DATA SCIENCE **HIERARCHY OF NEEDS**



► Como começar

Comece hoje mesmo!

- ▶ Oportunidades em pequenos negócios.
- ▶ Dados transacionais.
- ▶ Cadastros.
- ▶ Métricas do site, loja virtual.

Divulgue-se!

- ▶ Mostre que está aprendendo e engajado.
- ▶ Publique no nosso grupo do discord.
- ▶ Obtenha feedback e melhore.
- ▶ Publique em redes sociais.
- ▶ Compartilhe com amigos e peça feedback.

Emprego

- ▶ Conseguir uma entrevista.
- ▶ Ser aprovado na entrevista.
- ▶ Se manter no emprego.
- ▶ Rede de suporte.

► Oportunidades

Compreenda os seus clientes.

Cadastro de clientes.
Dados transacionais.
Métricas do site.
Formas de aquisição.

Identifique tendências e oportunidades

Dados macroeconômicos.
Análise da concorrência.
Mapeamento e otimização de processos.

Melhore seus produtos e serviços

Review de produtos.
Sites de reclamação.
Mídias sociais.

Começando

- ▶ Faça uma lista de oportunidades para DS na sua empresa atual.
- ▶ Lista os principais tópicos deste módulo e faça um pequeno resumo de cada um.
- ▶ Escolha o que você mais gostou e procure termos relacionados no Google.
- ▶ Leia entre três e cinco matérias sobre o tópico.
- ▶ Escreva uma matéria sobre o tópico.
- ▶ Publique e receba feedback.



R para decolar



Probabilidade e Estatística
para ciência de dados



Modelagem estatística para
ciência de dados



Manipulação e visualização
de dados



Matemática para ciência
de dados



Machine learning