

Data Science
Academy

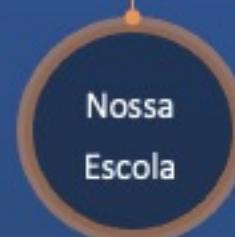


Introdução à Ciência de Dados 3.0



Bem-Vindo(a)!

Data Science Academy



A Data Science Academy é um portal de ensino online especializado em Big Data, Machine Learning, Inteligência Artificial, Blockchain, RPA e tecnologias relacionadas.

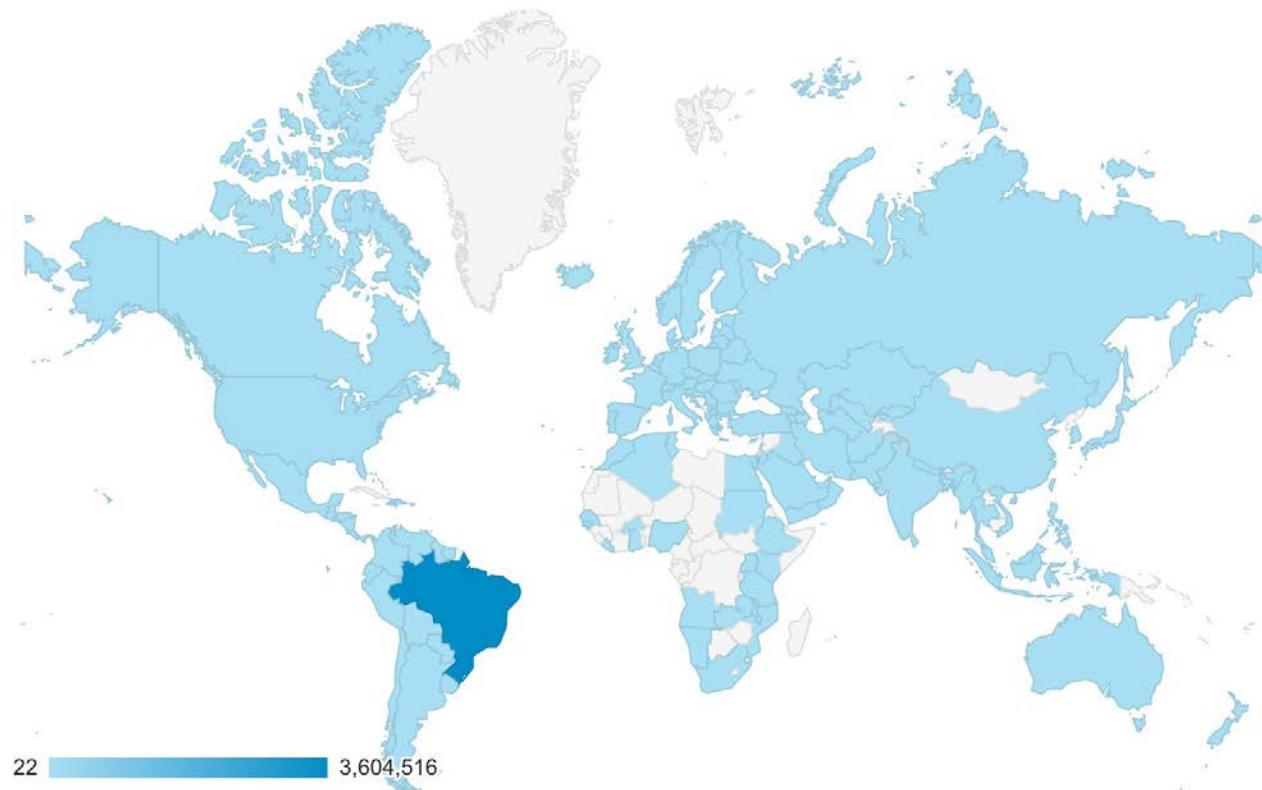
Nosso objetivo é fornecer aos alunos conteúdo de alto nível por meio do uso de computador, tablet ou smartphone, em qualquer lugar, a qualquer hora, 100% online e 100% em português.





Data Science Academy - Localização

No Brasil e no Mundo.





Conteúdo Programático

Conteúdo Programático



Capítulo 1

Introdução



Capítulo 2

O Que é Ciência de Dados?



Capítulo 3

Ciência de Dados e Big Data



Capítulo 4

Ciência de Dados e Estatística



Capítulo 5

Machine Learning e MLOps



Capítulo 6

Aplicações da Ciência de Dados



Capítulo 7

Ciclo de Vida de Projetos de Data Science



Capítulo 8

Soft Skills e Carreiras em Data Science



Capítulo 9

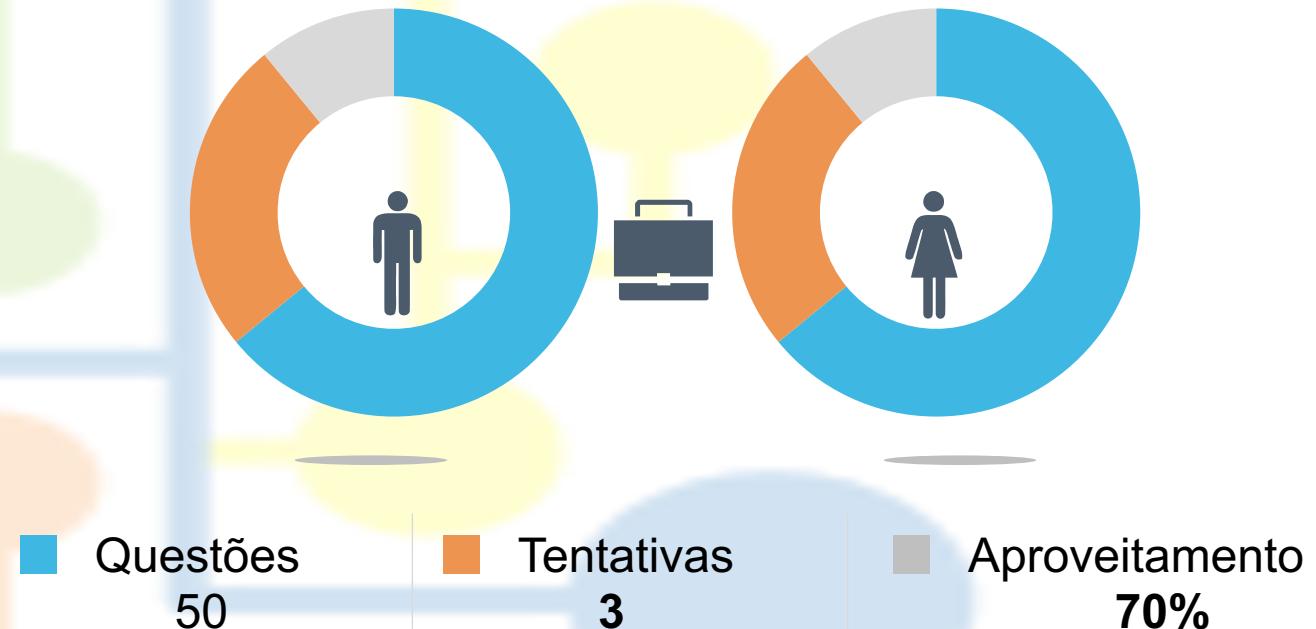
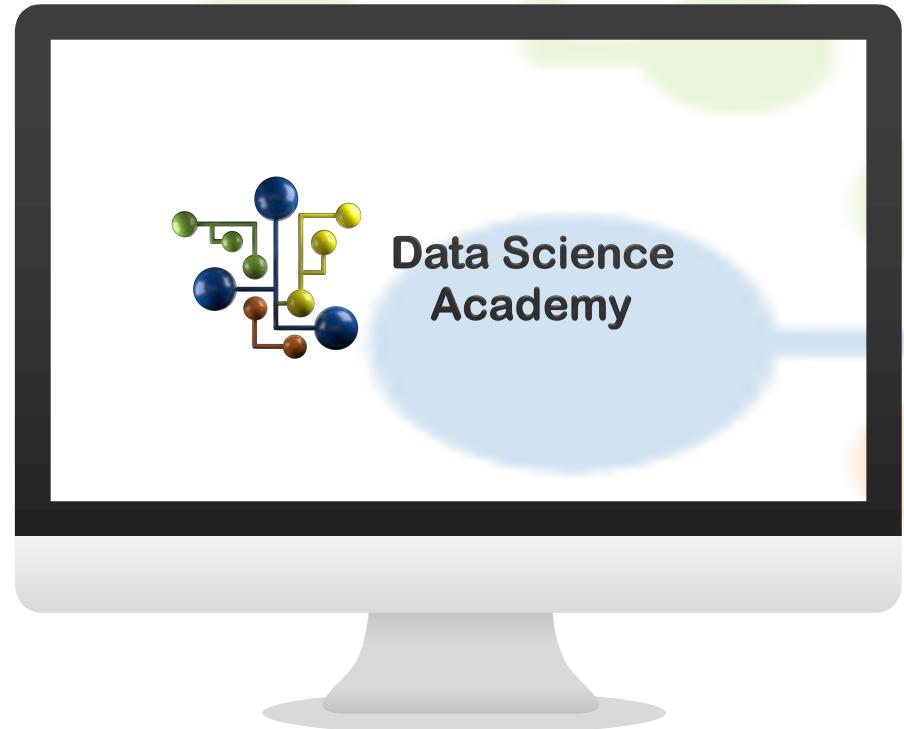
Como Se Tornar um Cientista de Dados



Capítulo 10

Avaliação Final e Certificado de Conclusão

Avaliação Final e Certificado de Conclusão



Qual o PÚblico Alvo Deste Curso?

Este é um curso teórico. Apenas relaxe e aproveite o conhecimento que será trazido a você.

Qualquer pessoa interessada em aprender sobre o universo da Ciência de Dados pode acompanhar este curso!

Este curso é pré-requisito recomendado para todos os demais cursos em nosso portal.



Será que Devo Migrar Minha Carreira Para Ciência de Dados?

- 1- A Ciência de Dados pode ser aplicada a qualquer área de qualquer segmento. Tudo que precisamos é de matéria-prima: dados.
- 2- Compreenda o que é Ciência de Dados e qual a relação com as áreas de negócio!
- 3- Pergunte a si mesmo se você está pronto(a) para estar sempre estudando e se atualizando constantemente.
- 4- Existem diversas carreiras em Ciência de Dados com diferentes perfis. Uma delas pode ser mais aderente ao que você gosta de fazer no dia a dia.
- 5- Migre sua carreira com uma visão de médio/longo prazo!





16 Fatos Importantes Sobre a Ciência de Dados

16 Fatos Importantes Sobre a Ciência de Dados

- O mercado de análise de Big Data deve atingir US\$ 103 bilhões em 2023.
- A baixa qualidade dos dados custa à economia dos EUA até US\$ 3,1 trilhões por ano.
- Em 2020, cada pessoa gerou 1,7 megabytes em apenas um segundo.
- Os usuários da Internet geram cerca de 2,5 quintilhões de bytes de dados por dia.

I ❤️
Fatos

16 Fatos Importantes Sobre a Ciência de Dados

- 95% das empresas citam a necessidade de gerenciar dados não estruturados como um problema para seus negócios.
- 97,2% das organizações estão investindo em Big Data e IA.
- Usando análise de Big Data, a Netflix economiza US\$ 1 bilhão por ano na retenção de clientes.
- Mais de 90% de todos os dados do mundo foram criados nos últimos 2 anos.

I ❤️
Fatos

16 Fatos Importantes Sobre a Ciência de Dados

- O mercado de análises de Big Data em saúde pode valer US\$ 67,82 bilhões em 2025.
- A criação de dados crescerá para mais de 180 zetabytes em 2025.
- As interações de dados aumentaram 5000% entre 2010 e 2020.
- Hoje, uma pessoa levaria aproximadamente 181 milhões de anos para baixar todos os dados da internet.

I ❤️
Fatos

16 Fatos Importantes Sobre a Ciência de Dados

- Em 2020, o mercado de análise de Big Data cresceu 14%.
- O número de dispositivos IoT pode aumentar para 41,6 bilhões em 2025.
- As vagas de empregos para funções em Ciência de Dados alcançaram cerca de 2,7 milhões em 2020 em todo mundo.
- As vagas em Ciência de Dados aumentarão em cerca de 28% até 2026.

I ❤️
Fatos

3



16 Fatos Importantes Sobre a Ciência de Dados

Esses fatos demonstram que precisaremos mais do que nunca de técnicas, ferramentas, procedimentos e processos para analisar os dados, extraíndo valor e ajudando as empresas a aumentar lucro, reduzir custo e sobreviver em um mercado cada vez mais competitivo e orientado a dados (data-driven).

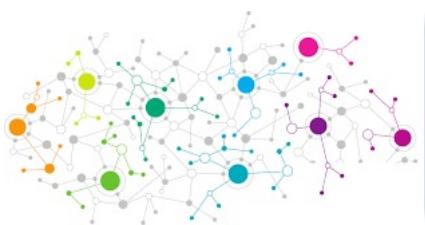


Definição de Ciência de Dados



Ciência de Dados (ou Data Science) é o termo usado para definir a extração de insights de dados que são coletados de várias fontes.

Definição de Ciência de Dados



Utilizando várias técnicas, incluindo modelagem preditiva (Machine Learning), a Ciência de Dados ajuda a analisar e interpretar grandes quantidades de dados.

Definição de Ciência de Dados



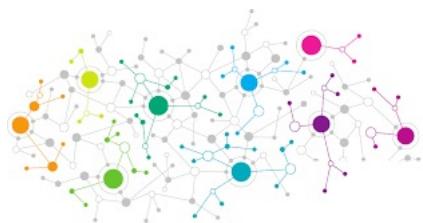
Os profissionais que trabalham com Ciência de Dados são chamados **Cientistas de Dados**, porém muitas outras carreiras estão associadas a Data Science, como veremos mais adiante.

Definição de Ciência de Dados



O termo Ciência de Dados surgiu na EMC (www.emc.com) e vem ganhando popularidade a cada dia.

Definição de Ciência de Dados



Ciência de Dados é o processo para extrair informações valiosas a partir de dados. Como estamos vivendo na era do Big Data, a Ciência de Dados está se tornando um campo cada vez mais importante para explorar e processar grandes volumes de dados, com diferentes formatos, gerados a partir de várias fontes e em diferentes velocidades.

Definição de Ciência de Dados



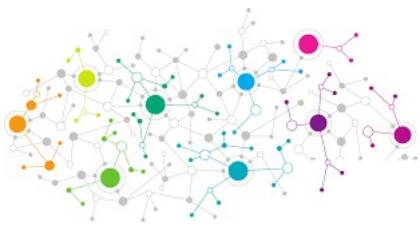
Ciência de Dados é uma grande disciplina em si e consiste em conjuntos de habilidades especializadas, tais como: estatística, matemática, programação, computação e conhecimento de negócios, além de técnicas e teorias, como a análise preditiva, modelagem, mineração de dados e visualização de dados.

Definição de Ciência de Dados



A área de Ciência de Dados não evoluiu da noite para o dia.

Definição de Ciência de Dados



Na verdade, a Ciência de Dados tem estado entre nós há bastante tempo, sob a forma de análise de negócios ou inteligência competitiva, mas somente agora o seu verdadeiro potencial foi percebido e isso se deve em parte ao Big Data.

Definição de Ciência de Dados



O principal objetivo da Ciência de Dados é extrair e interpretar os dados de forma eficaz e apresentá-los em uma linguagem simples para os usuários finais e tomadores de decisão.

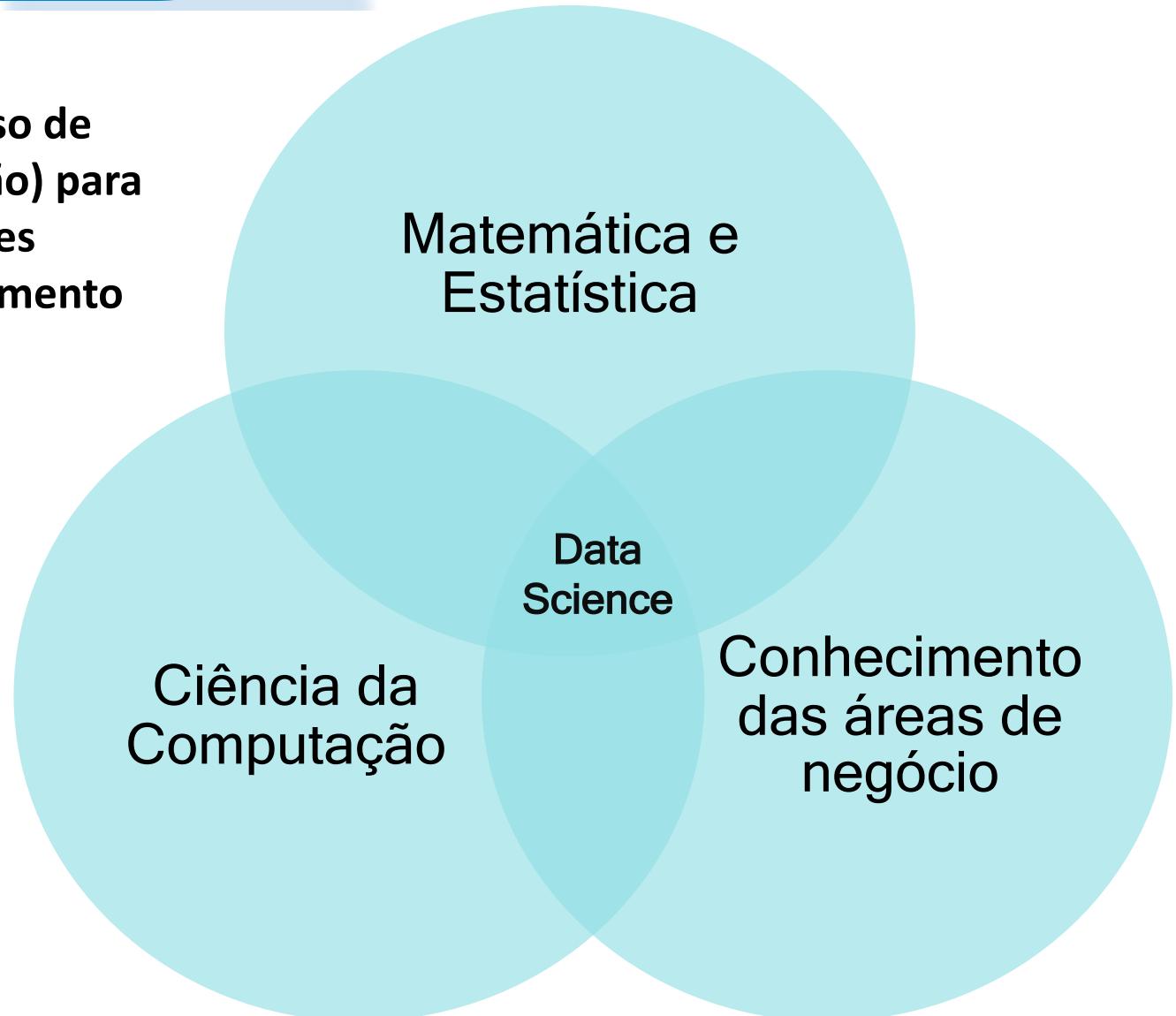
Definição de Ciência de Dados



Assim, a Ciência de Dados visa entregar informações úteis e a capacidade de converter essas informações em produtos "data-driven"!

Áreas de Conhecimento

Em essência, a Ciência de Dados envolve o uso de métodos automatizados (Ciência da Computação) para analisar (Matemática e Estatística) enormes quantidades de dados a fim de extrair conhecimento (áreas de negócio) a partir dos dados.



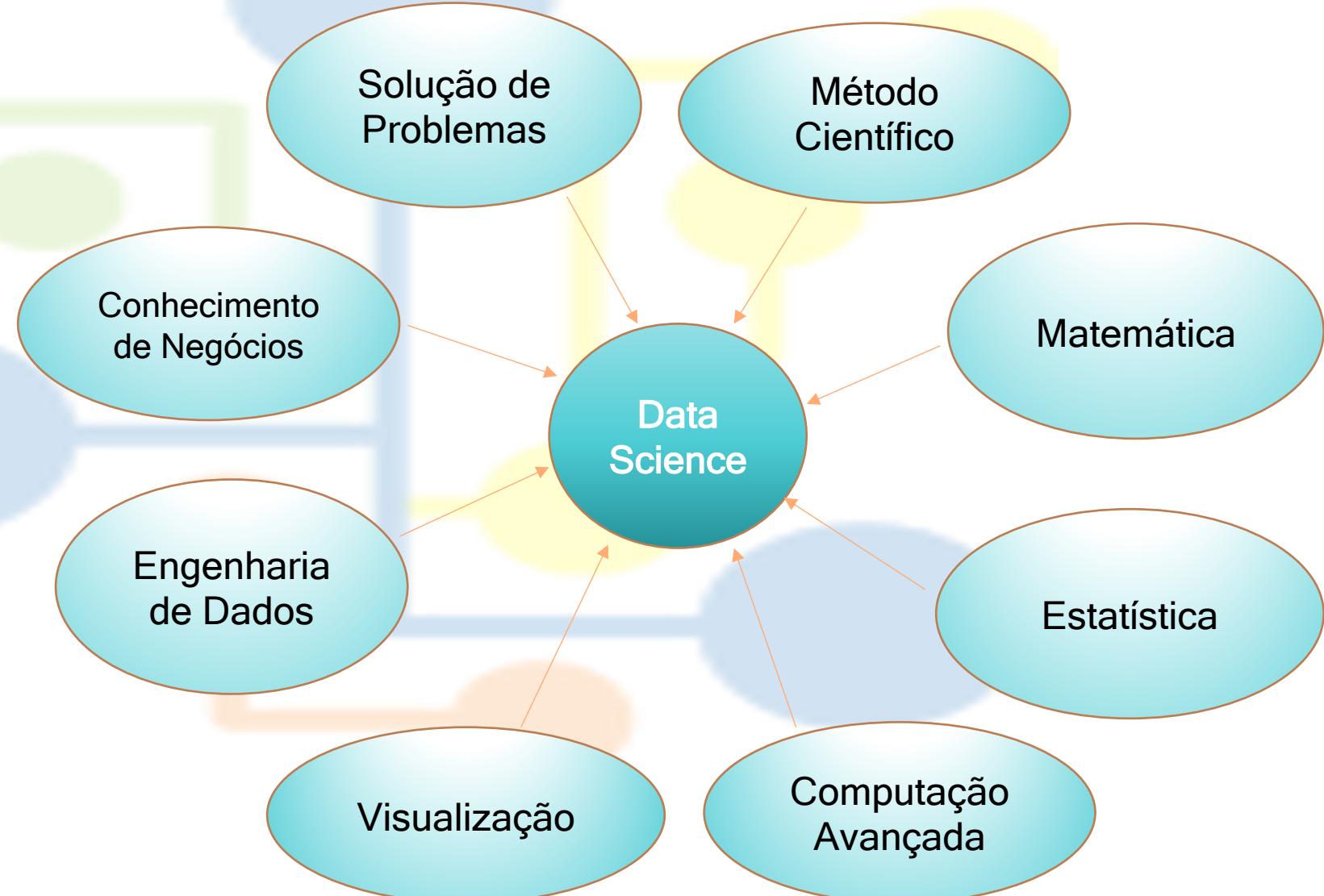
A Importância da Ciência em Ciência de Dados

Perceba que a palavra **chave** em Data Science, não é Data e sim Science.

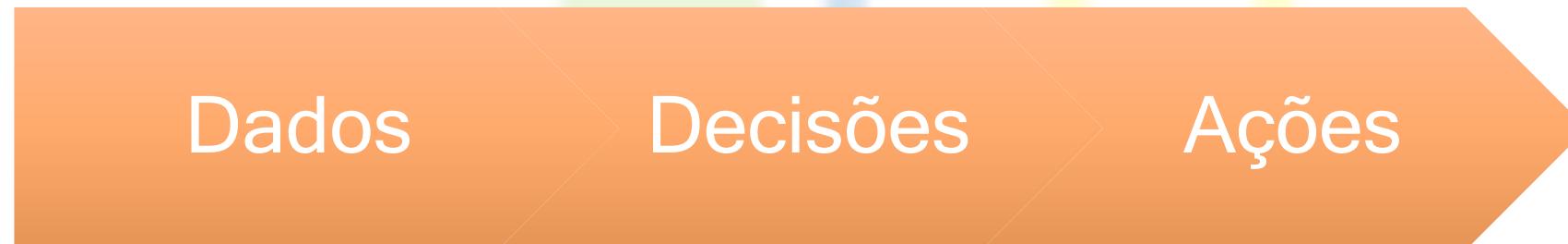


A Importância da Ciência em Ciência de Dados

Ciência de Dados é a exploração e análise de todos os dados disponíveis, sejam eles estruturados ou não, com o objetivo de desenvolver compreensão, extrair conhecimento e formular ações que gerem resultados.



Dados e Tomada de Decisão



Dados e Tomada de Decisão

Exemplo:

Problema identificado: Clientes cancelam seus planos de telefonia com frequência de 2 a 3 meses após assinatura do plano.

- 1- **Dados e Análise** - Dados coletados e analisados, levaram à conclusão que existe uma tendência entre os clientes de operadoras de telefonia em trocar de operadora após o terceiro contato com atendimento ao cliente.
- 2- **Decisão** - Os executivos da empresa, baseados em dados, decidiram rever todos os processos de atendimento ao cliente.
- 3- **Ação** - A empresa contratou uma consultoria para fornecer treinamento especializado de atendimento ao cliente, passou a monitorar as ligações, criou programas de incentivo aos melhores operadores, etc...



Dados e Tomada de Decisão

Dados

Decisões

Ações

O principal objetivo é extrair informações dos dados e transformá-las em conhecimento que possam ser usadas para a tomada de decisões.

Dados e Tomada de Decisão



Dados

Valor

O que deve ser feito?

Acontecerá novamente?

Por que aconteceu?

O que aconteceu?

Decisão

Ação

É Preciso Graduação ou Pós Para Trabalhar com Ciência de Dados?

Vamos começar colocando os pingos nos is.

Existem diversas funções em Ciência de Dados:

- Cientista de Dados
- Engenheiro de Dados
- Analista de Dados
- Engenheiro de Machine Learning
- Engenheiro de IA
- Engenheiro DataOps
- Arquiteto de Dados
- Engenheiro Blockchain*
- Desenvolver RPA*

* Funções mais voltadas pra desenvolvimento embora requeiram conhecimento em dados.

E Preciso Graduação ou Pós Para Trabalhar com Ciência de Dados?

Vamos começar colocando os pingos nos is.

Existem diversas funções em Ciência de Dados:

- Cientista de Dados
- Engenheiro de Dados
- Analista de Dados
- Engenheiro de Machine Learning
- Engenheiro de IA
- Engenheiro DataOps
- Arquiteto de Dados
- Engenheiro Blockchain*
- Desenvolver RPA*

Estas são FUNÇÕES e não PROFISSÕES.

Isso significa que não existe regulamentação e qualquer profissional, com ou sem graduação (ou graduação em qualquer área), pode exercer as funções.

A pergunta ideal seria:

Qual curso de graduação poderia ser mais útil para trabalhar com Ciência de Dados?

* Funções mais voltadas pra desenvolvimento embora requeiram conhecimento em dados.

E Preciso Graduação ou Pós Para Trabalhar com Ciência de Dados?

Um curso de Graduação vai ajudar você com diversos conhecimentos que serão importantes no dia a dia de qualquer profissão.

Um curso de Pós pode fornecer uma visão mais gerencial (MBA) ou uma extensão sobre o que foi aprendido na graduação (Lato Sensu).

Um Mestrado ou Doutorado (Strictu Sensu) vai permitir que você se especialize em um determinado assunto, com foco acadêmico e de pesquisa.

E Preciso Graduação ou Pós Para Trabalhar com Ciência de Dados?

Aqui na DSA não oferecemos cursos de Graduação ou Pós.

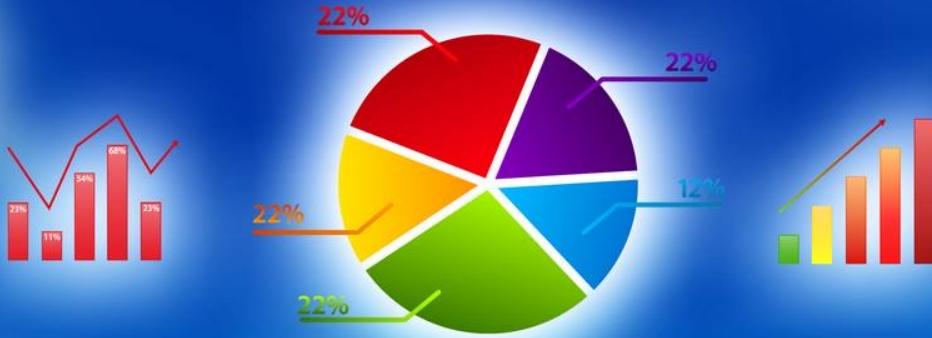
Oferecemos cursos de aperfeiçoamento profissional para quem deseja aprender a trabalhar com Ciência de Dados na prática e com projetos orientados às necessidades do mercado de trabalho.

A carga horária das nossas Formações é superior a cursos de Pós, pois nosso conteúdo é denso, com vasta quantidade de exemplos, exercícios, materiais complementares e projetos práticos.

E Preciso Graduação ou Pós Para Trabalhar com Ciência de Dados?

Exatamente por isso capacitamos profissionais das maiores empresas do Brasil:

<https://www.datascienceacademy.com.br/clientesdsa>



Busque conhecimentos para se tornar um profissional de Analytics.

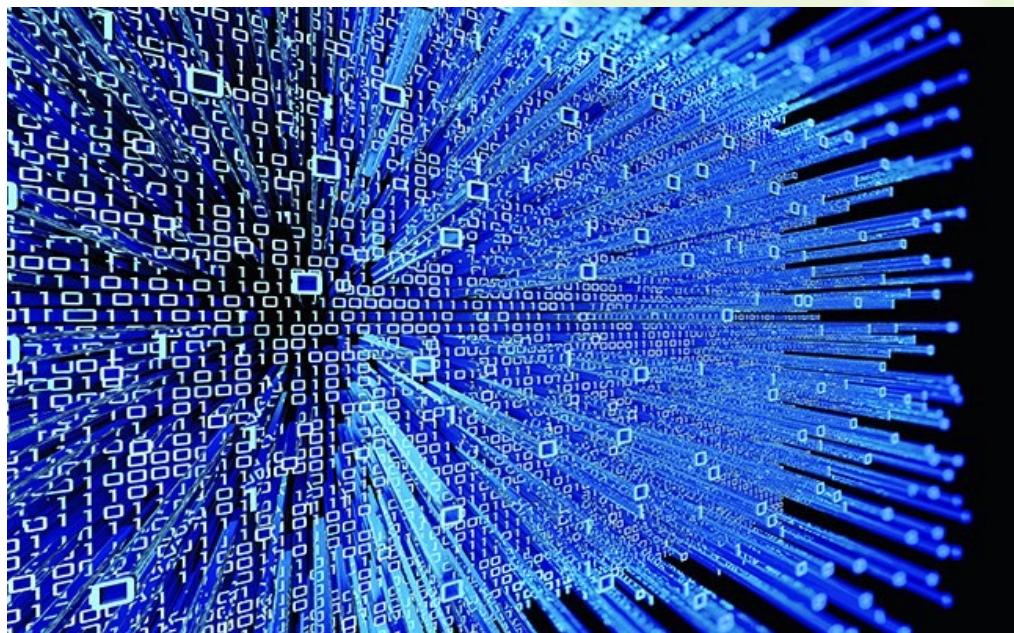


Isso fará toda a diferença em sua carreira!



O Que São Dados?

O Que São Dados?

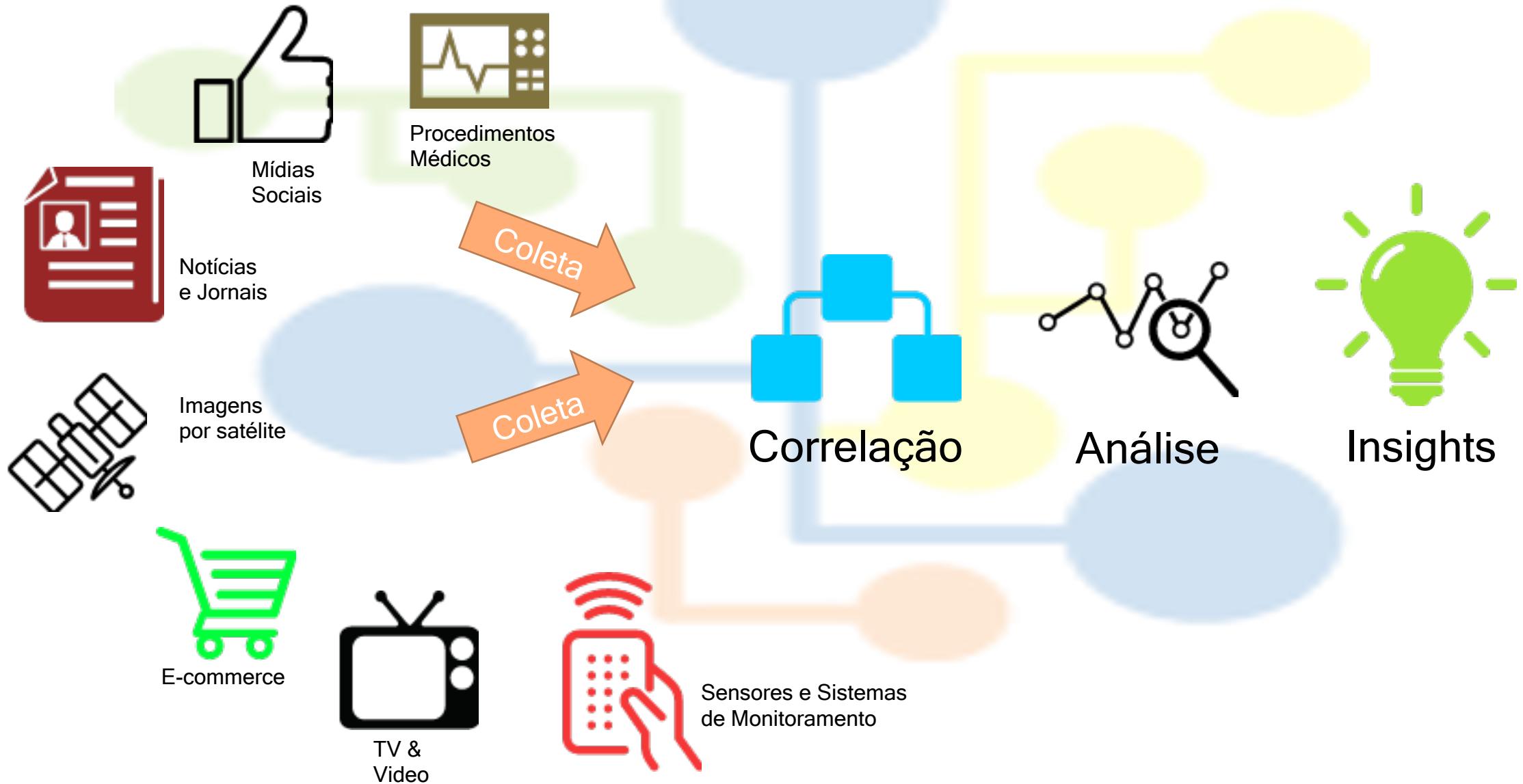


Os dados são fatos individuais, estatísticas ou medidas, que são coletados por meio de observação e/ou medição.

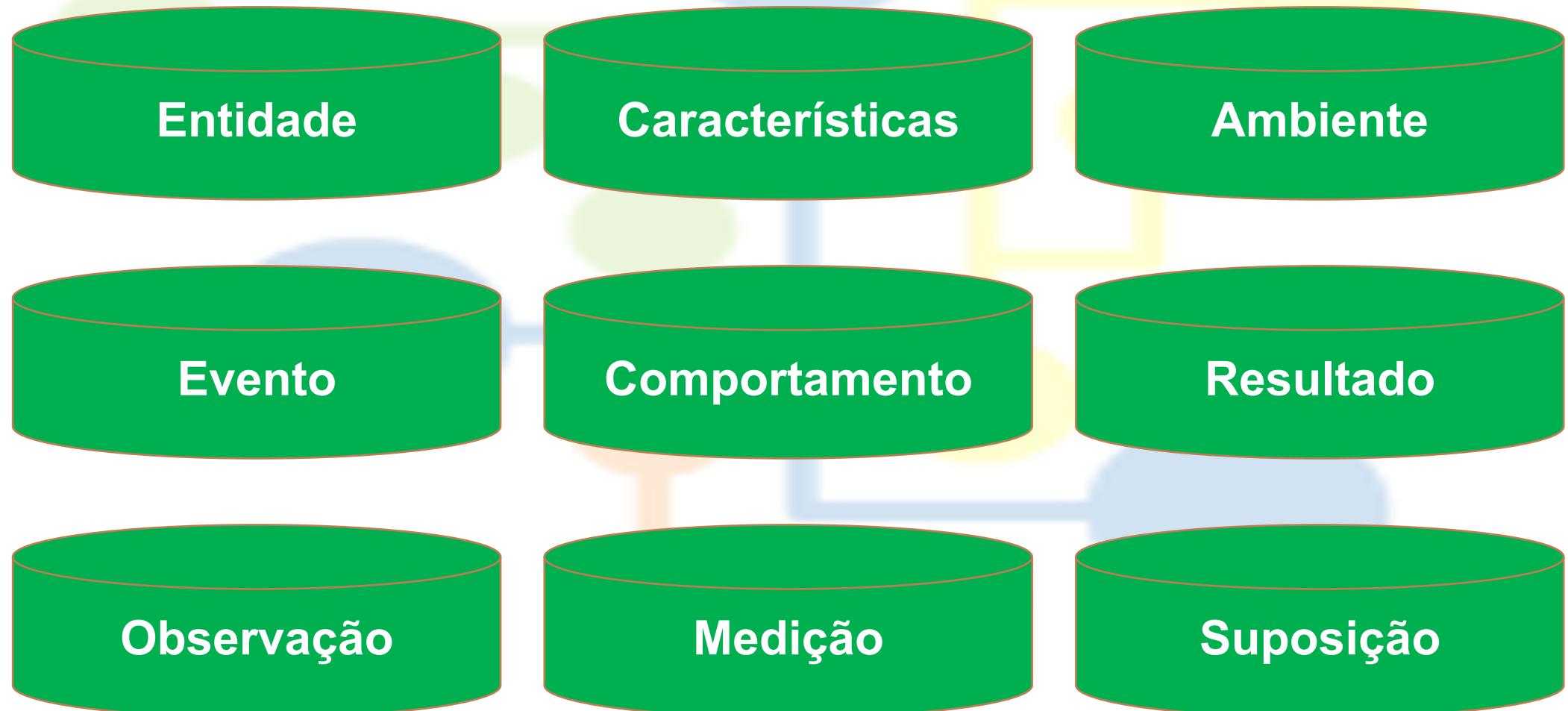
Em um sentido mais técnico, os dados são um conjunto de valores de variáveis qualitativas ou quantitativas sobre eventos, pessoas ou objetos.

Embora os termos "dados" e "informações" sejam frequentemente usados de forma intercambiável, esses termos têm significados distintos.

De Onde Vem os Dados?



O Que os Dados Representam?



Dados e Informação



Dados são coleções de fatos, tais como números, palavras, medições, observações ou mesmo apenas descrições de coisas.

Dados e Informação



Os dados podem ser qualquer caracter, texto, palavras, números, imagens, som ou vídeo e, se não colocados em contexto, significam pouco ou nada para um ser humano.

Dados e Informação



No entanto, a informação é útil e formatada de uma maneira geral, permitindo que seja entendida por um ser humano.

Dados e Informação

Dado

Informação



Dados e Informação

Dado

Informação

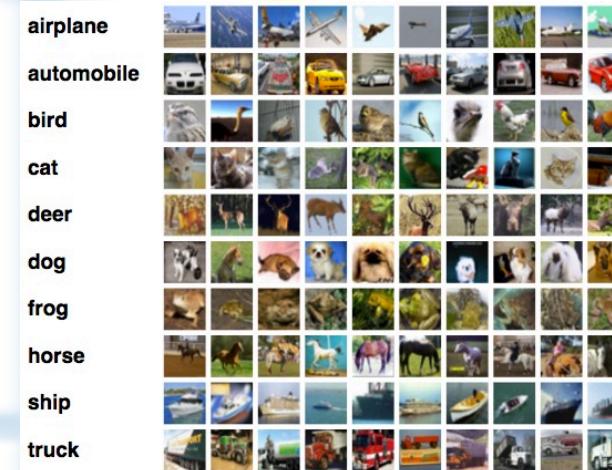


55-21-9999-9999

Telefone Pessoal

O Que São Datasets?

- Conjuntos de dados. ✓
- Coleções de observações. ✓
- Cada observação é tipicamente chamada de registro. ✓
- Cada registro tem um conjunto de atributos que apontam características, comportamentos ou resultados. ✓
- Datasets podem ter 1 registro e 1 atributo e não há limite máximo. ✓



Dataset de Imagens

Branch	Company	ImportCount	ExportCount	WHSCount	Other	TotalCount
HKG	HKG	19	235	0	0	254
ATL	ORD	186	123	1	1	310
BOS	ORD	31	21	0	0	52
CHS	ORD	181	160	0	0	341
CVG	ORD	58	43	0	0	101
DFW	ORD	90	44	0	0	134
DTW	ORD	350	35	1	1	386
EWR	ORD	529	82	0	0	611
IAH	ORD	90	50	5	5	145
JFK	ORD	0	104	0	0	104
LAX	ORD	247	241	12	22	510
MAO	ORD	1	0	0	0	1
MIA	ORD	71	97	9	9	177
MSP	ORD	21	10	0	0	31
ORD	ORD	1477	912	81	81	2470
PDX	ORD	61	26	1	1	88
PIT	ORD	110	60	0	0	170

Dataset de Dados Numéricos e de Texto

E o Big Data?



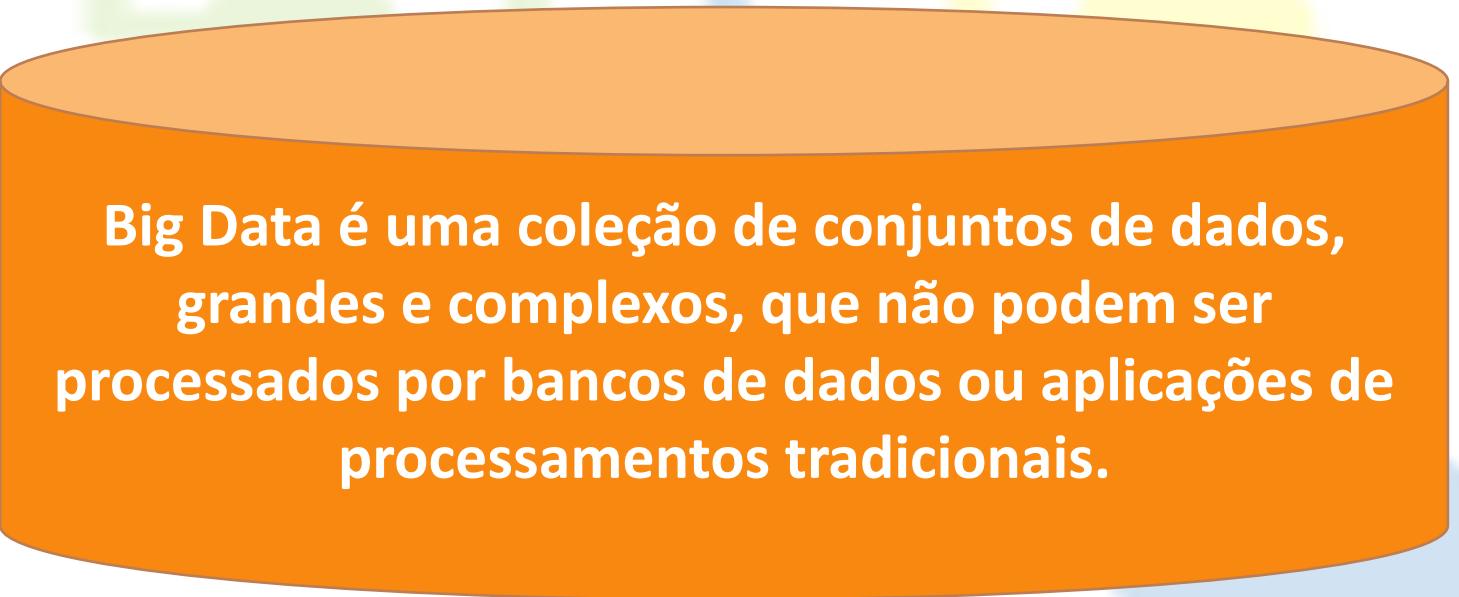
Big Data é mais do que uma palavra da moda no mundo dos negócios.

Todo mundo deixa “rastros” quando cria uma conta em rede social, assiste um vídeo na internet, faz compras online, etc...

São dados como frequência de compra, consumo médio, número de curtidas em posts, pesquisas, etc...

Multiplique isso bilhões de vezes e você começa a perceber o que é o Big Data.

E o Big Data?



Big Data é uma coleção de conjuntos de dados, grandes e complexos, que não podem ser processados por bancos de dados ou aplicações de processamentos tradicionais.

E o Big Data?

Big Data é definido por 4 V's:

- Volume
- Variedade
- Velocidade
- Veracidade



E o Big Data?

O Big Data pode ser composto por dados de diferentes fontes e diferentes formatos.

Fonte dos Dados

Interno



Estruturado



- Resultados de pesquisas
- Registros de vendas
- Medidas de controle de processos
- Bancos de dados de sistemas internos (ERP, CRM)

Não-estruturado



- E-mails, cartas, mensagens de texto
- Legendas de vídeos
- Comentários de clientes
- Mensagens de voz
- Imagens / ilustrações
- Avaliação de funcionários

Externo



- Likes do Facebook, retweets
- Horário de publicação de posts, tweets, updates
- Pontuação em sites de classificação

- Conteúdo publicado em redes sociais
- Comentários em fóruns online
- Imagens
- Vídeos de câmeras de segurança

Ciência de Dados e Big Data



Data
Science

Big Data é o Combustível Para a
Ciência de Dados!

Ciência de Dados e Big Data



Com o volume de dados cada vez maior conseguimos extrair padrões, detectar anomalias, encontrar correlações e fazer previsões de forma cada vez mais precisa.

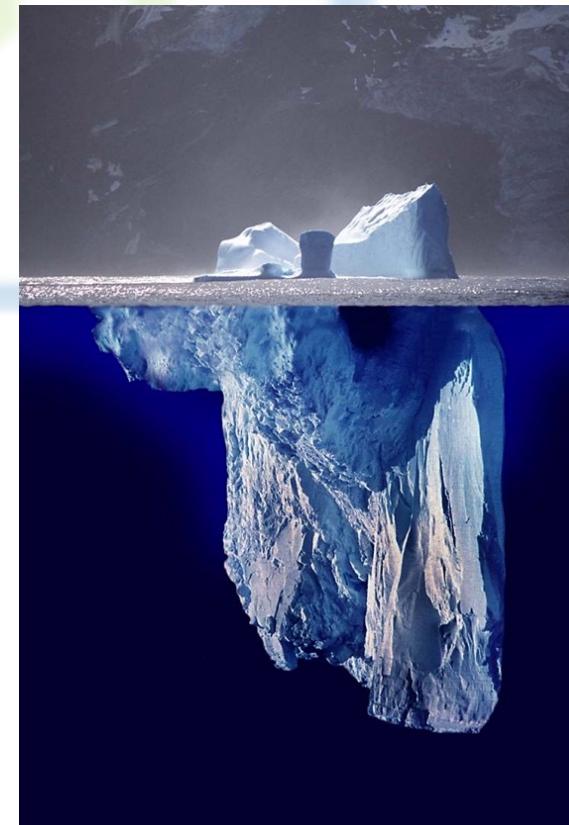
O Big Data foi um dos grandes responsáveis pelo crescimento da Ciência de Dados.

A Oportunidade do Big Data

ERP CRM BI SCM RH

Mídia Social
Análise de Textos
API's
Reconhecimento de Voz
Processamento de Linguagem Natural
Internet das Coisas

Automação
Saas
PaaS



Dados Internos

Estruturados, previsíveis, permanentes, fáceis de obter.

Dados Externos

Não-estruturados, randômicos, voláteis, difíceis de obter.

Ciência de Dados e Big Data

Nós tendemos a superestimar o efeito de uma nova tecnologia no curto prazo e subestimar o efeito no longo prazo.

- Roy Amara

Ciência de Dados e Big Data

Em Deus nós acreditamos. Todos os outros
devem trazer dados.

Willian E. Deming

O Que é e Como Aprender Big Data Analytics?

Big Data Analytics é a aplicação da Ciência de Dados para a análise de Big Data.

Com Big Data Analytics o objetivo é resolver problemas de negócio através da análise de dados com alto volume, variedade e gerados em alta velocidade.



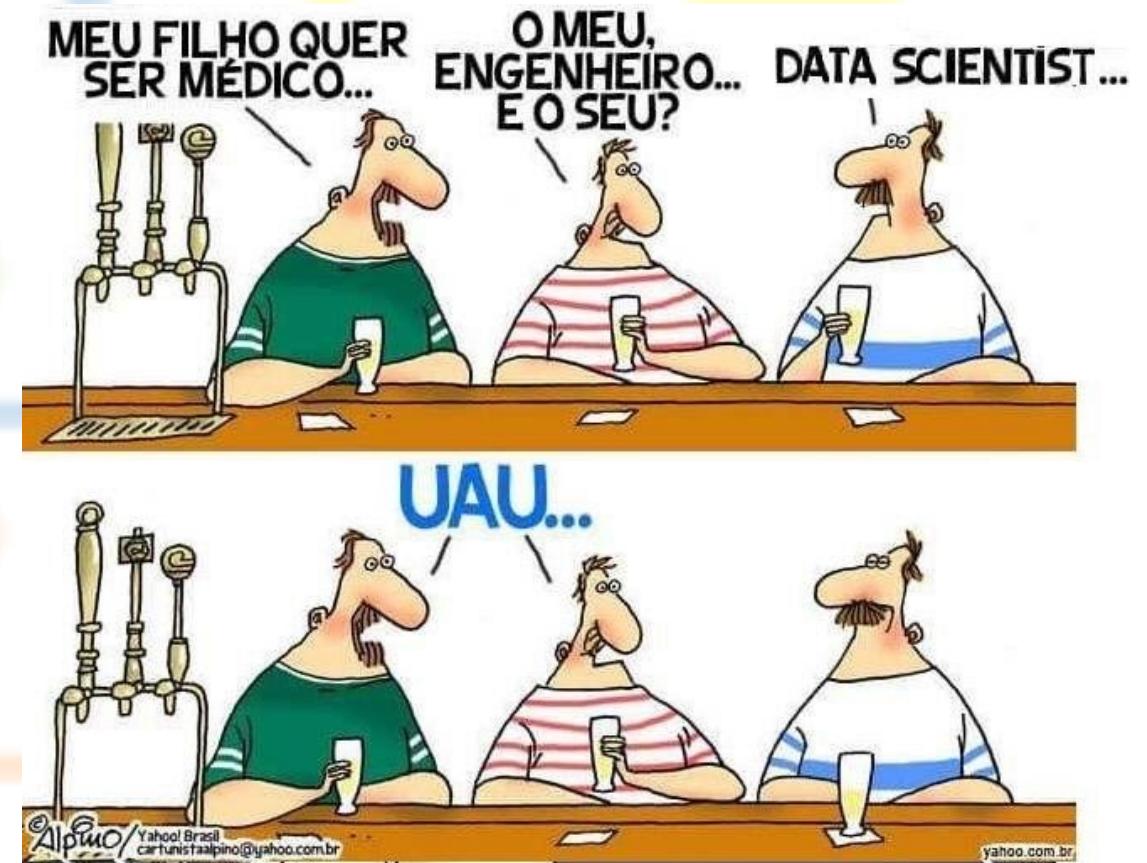
O Que é e Como Aprender Big Data Analytics?

As profissões relacionadas à análise de dados estão entre as mais promissoras na próxima década.



O Que é e Como Aprender Big Data Analytics?

Investir em uma carreira em Ciência de Dados é uma decisão inteligente pois ainda estamos na infância do que pode ser feito com Big Data Analytics.



O Que é e Como Aprender Big Data Analytics?

A Data Science Academy oferece programas de capacitação completos, totalmente práticos e orientados às necessidades do mercado de trabalho, como a Formação Cientista de Dados.

Através de diversos projetos práticos com diferentes ferramentas os alunos aprendem a trabalhar de forma profissional com Big Data Analytics.

Confira o programa completo no link disponível na seção de links úteis ao final do capítulo.

Data Science x Estatística

Ciência de Dados e Estatística
são a mesma coisa?



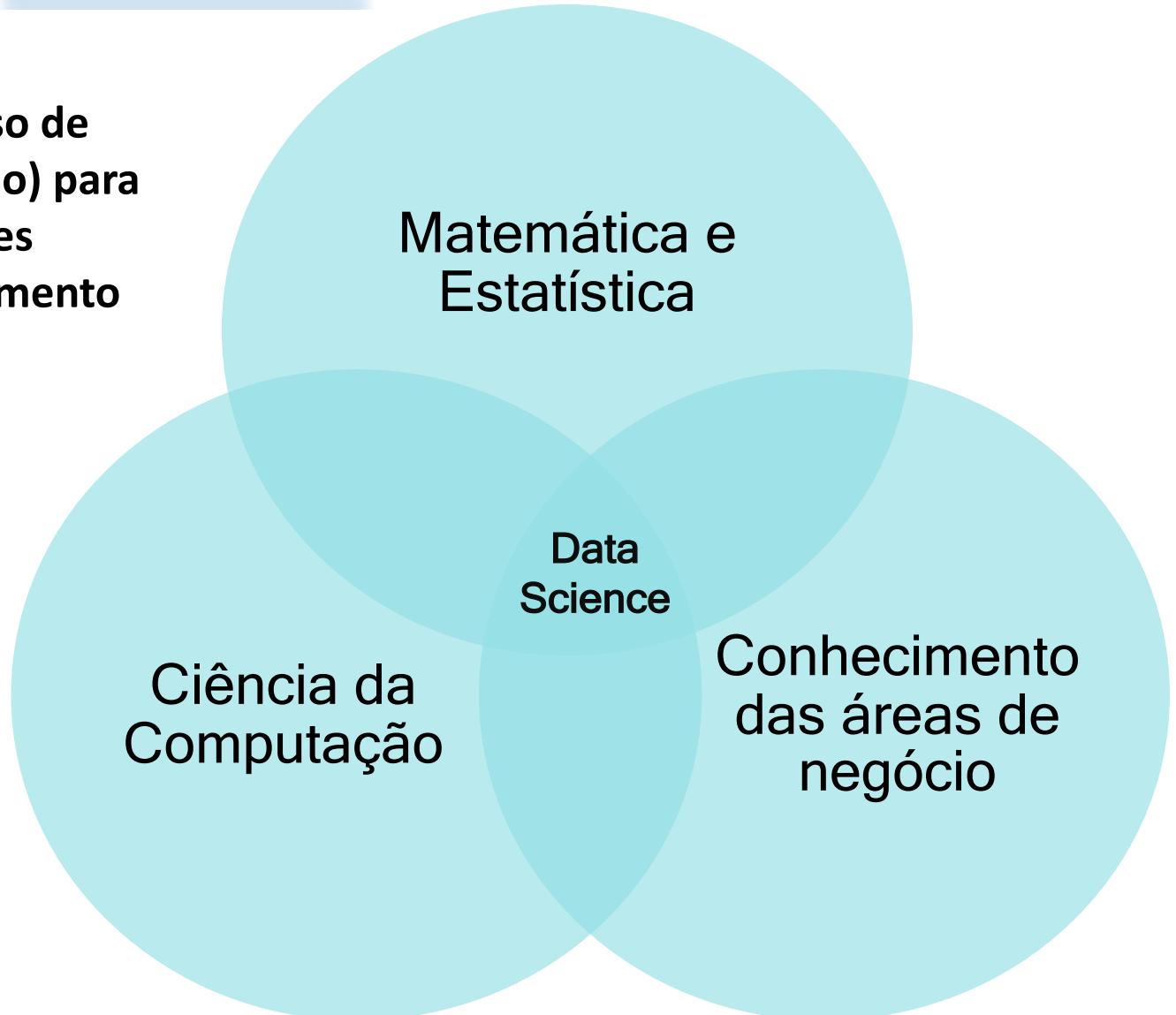
Data Science x Estatística

NÃO



Data Science x Estatística

Em essência, a Ciência de Dados envolve o uso de métodos automatizados (Ciência da Computação) para analisar (Matemática e Estatística) enormes quantidades de dados a fim de extrair conhecimento (áreas de negócio) a partir dos dados.



O Que a American Statistical Association (ASA) Declara Sobre Data Science x Estatística?



“SEM DADOS VOCÊ
É APENAS MAIS
UMA PESSOA COM
UMA OPINIÃO.”

WILLIAM EDWARDS DEMING

O Que a American Statistical Association (ASA) Declara Sobre Data Science x Estatística?

A American Statistical Association (ASA) divulgou recentemente uma declaração sobre o papel da Estatística na Ciência de Dados. O Presidente da ASA, David Morganstein, deu esta declaração no seu comunicado de imprensa:

“Através desta declaração, a ASA e seus membros reconhecem que a ciência de dados abrange mais do que estatística, mas ao mesmo tempo também reconhece que a estatística desempenha um papel fundamental no rápido crescimento deste campo. É nossa esperança que esta declaração possa reforçar a relação da estatística com a ciência de dados e ainda fomentar relacionamentos mútuos de colaboração entre todos os contribuintes na ciência de dados.”

O Que a American Statistical Association (ASA) Declara Sobre Data Science x Estatística?

A declaração evidencia que Estatística é fundamental para a Ciência de Dados, juntamente com gestão de banco de dados, sistemas distribuídos e paralelos, computação, matemática e programação. A sua utilização neste campo emergente capacita pesquisadores para extrair conhecimento e obter melhores resultados de grandes projetos. A declaração também incentiva a colaboração mútua e multifacetada entre estatísticos e cientistas de dados para maximizar o potencial da ciência de dados.

Em resumo, a Estatística desempenha um papel fundamental dentro da Ciência de Dados.

Porém, a Ciência de Dados compreende outras áreas de conhecimento, como já vimos anteriormente.

A Ciência de Dados é uma área multidisciplinar e a Estatística é uma das suas disciplinas.



Definindo Estatística

O Que é Estatística?



Definindo Estatística

A Estatística fornece técnicas e métodos de análise de dados que auxiliam o processo de tomada de decisão nos mais variados problemas onde existe incerteza.



Definindo Estatística

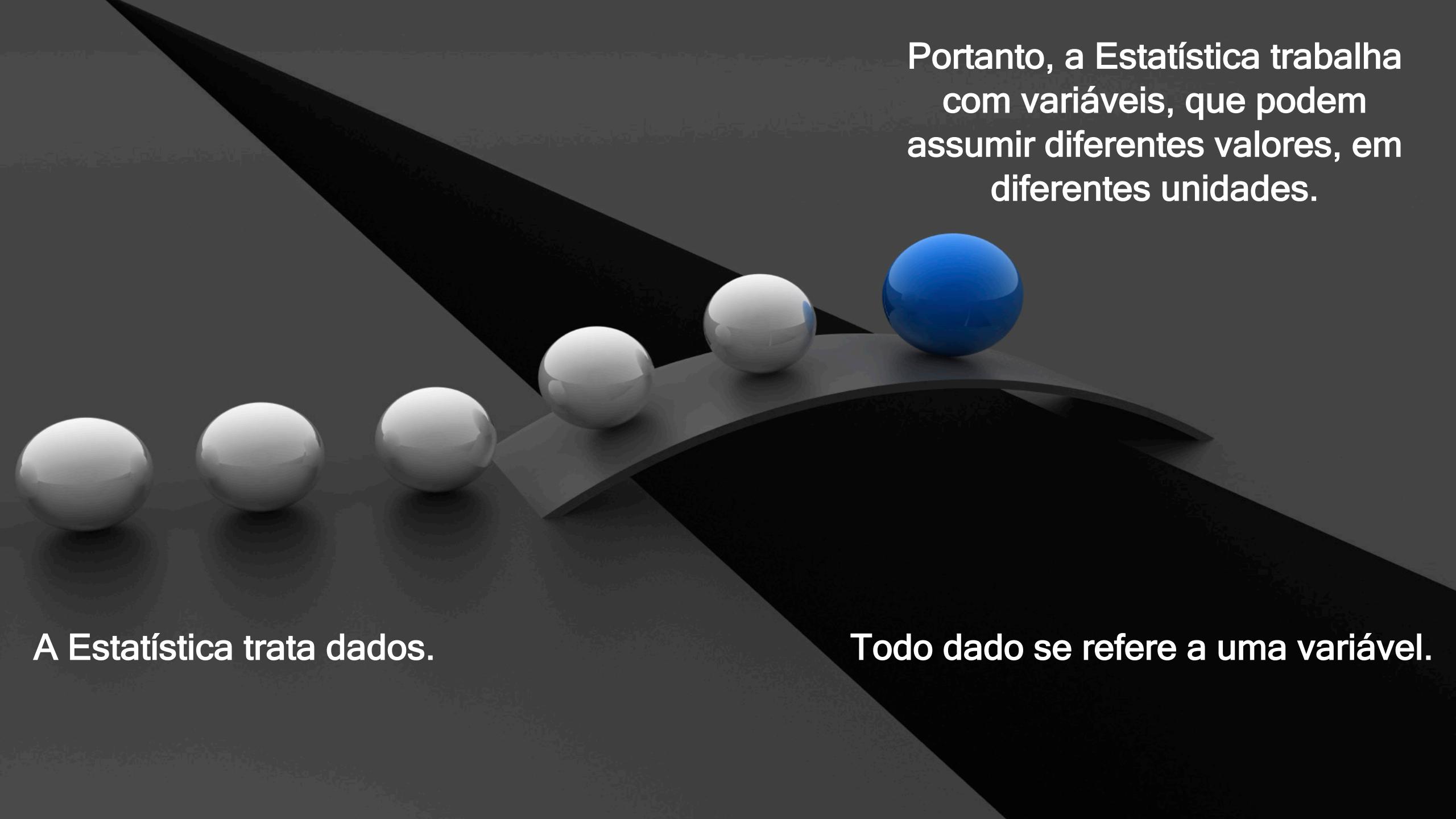
- Bioestatística
- Contabilometria
- Controle de qualidade
- Estatística comercial
- Estatística econômica
- Estatística de engenharia
- Estatística física
- Estatística populacional
- Estatística psicológica
- Estatística social (para todas as ciências sociais)
- Geoestatística
- Pesquisa operacional
- Análise de processo



As 3 Grandes Áreas da Estatística

Estatística é a ciência, parte da **Matemática Aplicada**, que fornece métodos para **coletar, descrever, analisar, apresentar e interpretar** dados, para a utilização dos mesmos na **tomada de decisões**.



The background features a dark gray gradient with a subtle geometric pattern of intersecting diagonal lines. Five glossy spheres are scattered across the surface: one large blue sphere on the right, and four smaller white spheres arranged in a curve on the left.

Portanto, a Estatística trabalha
com variáveis, que podem
assumir diferentes valores, em
diferentes unidades.

A Estatística trata dados.

Todo dado se refere a uma variável.

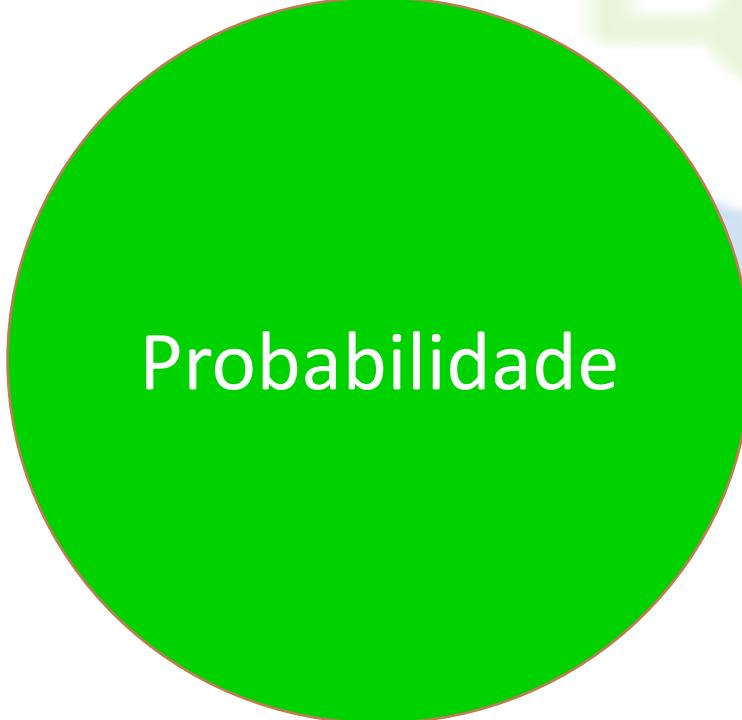
As 3 Grandes Áreas da Estatística

Probabilidade

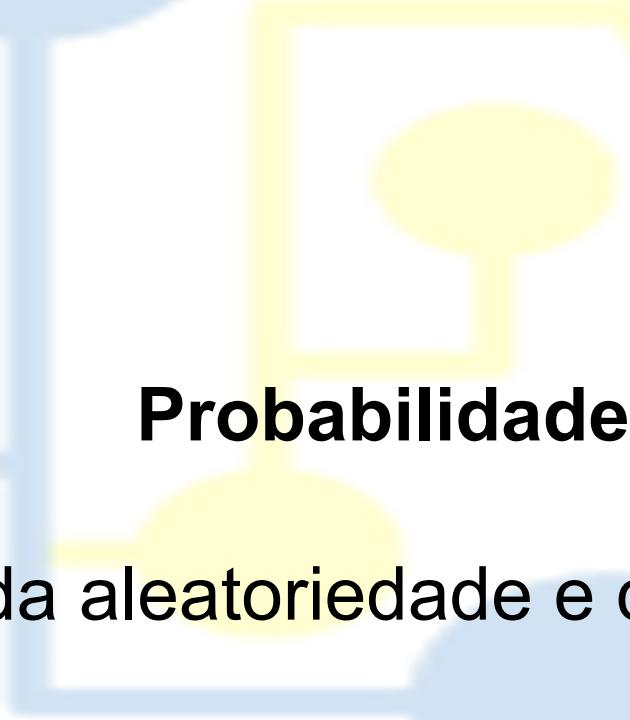
Estatística
Descritiva

Estatística
Inferencial

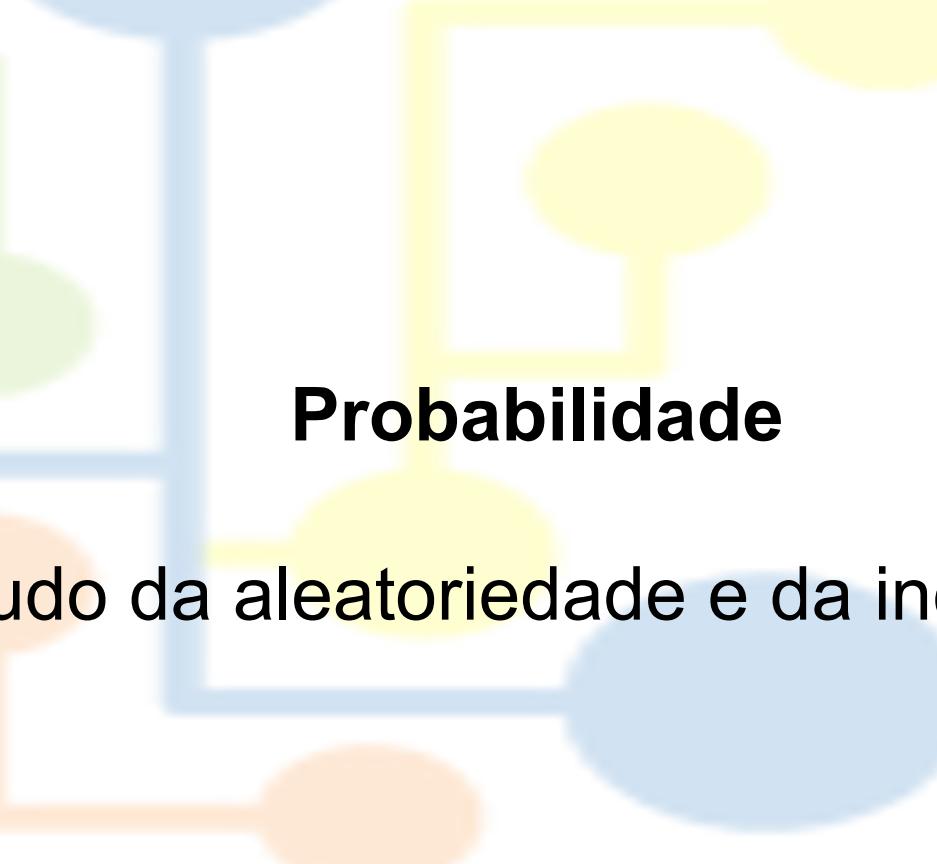
As 3 Grandes Áreas da Estatística



Probabilidade



Probabilidade



Estudo da aleatoriedade e da incerteza.

As 3 Grandes Áreas da Estatística

Estatística Descritiva

Estatística Descritiva

Utiliza métodos para coleta, organização, apresentação, análise e síntese de dados obtidos de uma população ou amostra.

As 3 Grandes Áreas da Estatística



Estatística
Inferencial

Estatística Inferencial

É o processo de estimar informações sobre uma população a partir dos resultados observados em uma amostra.

Qual a Relação Entre Ciência de Dados e Estatística?

A Estatística fornece ferramentas, técnicas e procedimentos que podem ser usados em todas as etapas de projetos de Ciência de Dados.



Qual a Relação Entre Ciência de Dados e Estatística?

Usamos Estatística em diferentes momentos dentro de um projeto de Ciência de Dados. Além disso, alguns algoritmos de Machine Learning são baseados em fundamentos estatísticos.



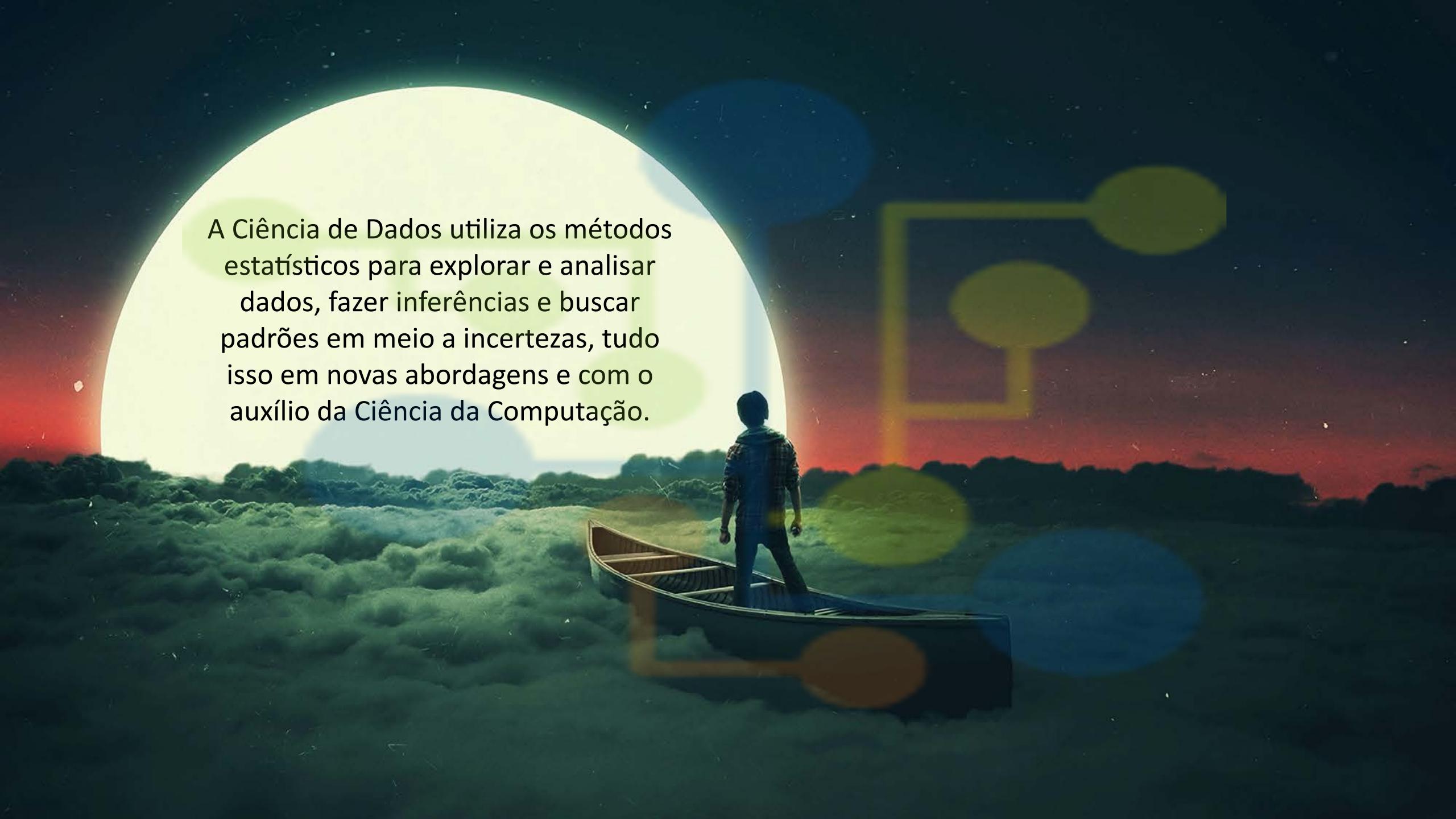
Qual a Relação Entre Ciência de Dados e Estatística?

A Estatística pode ajudar a avaliar se o sucesso de um fundo de investimentos é genuíno ou devido ao acaso, pode ajudar a prever se um determinado assinante vai cancelar sua assinatura este ano ou se uma reivindicação de seguro é fraudulenta.



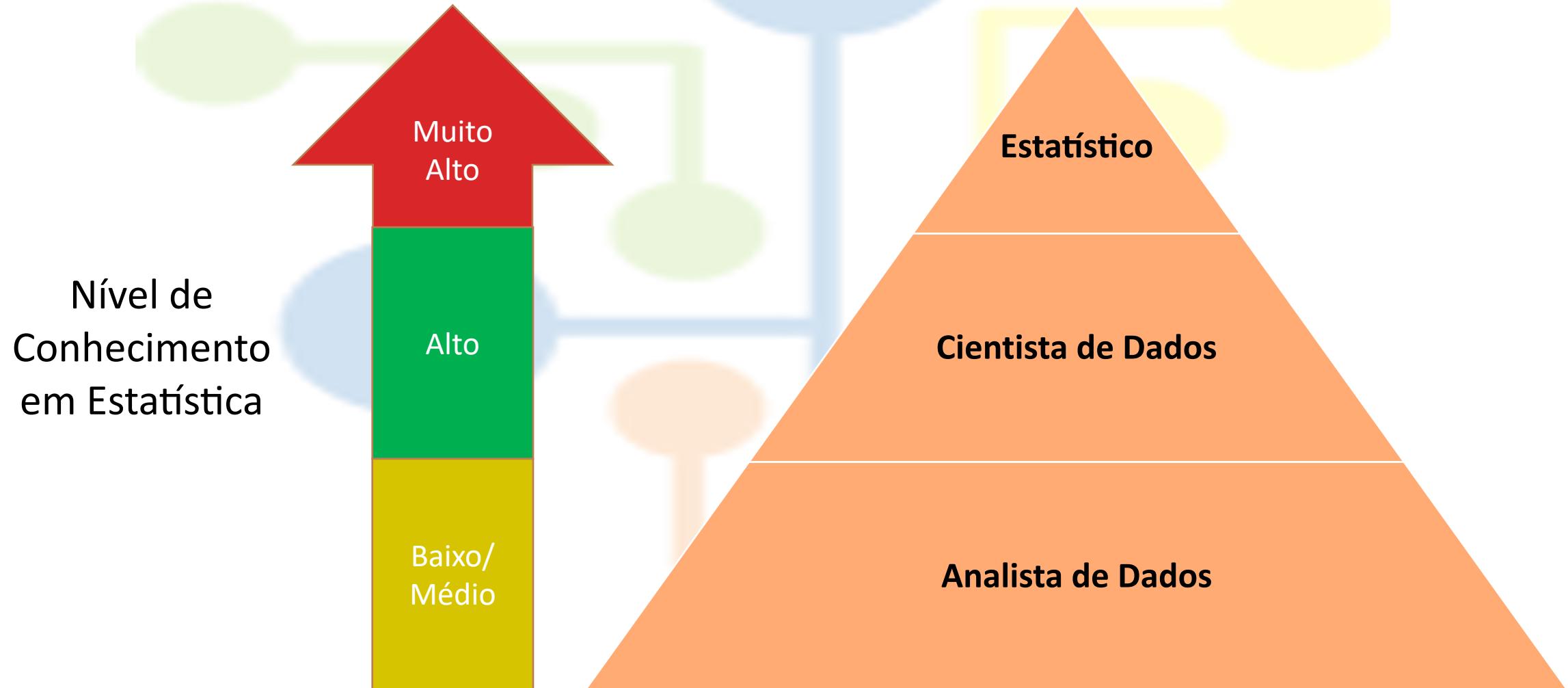
Qual a Relação Entre Ciência de Dados e Estatística?

Vivemos uma era em que a ciência deve prevalecer sobre o empirismo, em que a lógica deve prevalecer sobre o "achismo".



A Ciência de Dados utiliza os métodos estatísticos para explorar e analisar dados, fazer inferências e buscar padrões em meio a incertezas, tudo isso em novas abordagens e com o auxílio da Ciência da Computação.

Como Aprender Estatística Para Ciência de Dados?



Como Aprender Estatística Para Ciência de Dados?

Os Cientistas de Dados possuem 3 características principais:

1. Eles têm um forte conhecimento de Estatística e aprendizado de máquina, pelo menos o suficiente para evitar má interpretação de correlação e causalidade.
2. Eles têm habilidades de computação/programação para usar uma linguagem de programação (como R ou Python) para realizar o trabalho de análise.
3. Eles podem visualizar e resumir seus dados e sua análise de uma maneira que seja significativa para alguém menos familiarizado com os dados, baseado em sua experiência de áreas de negócio.

Como Aprender Estatística Para Ciência de Dados?

Para muitas pessoas, o aprendizado de Estatística pode ser meio obscuro. Isso se deve aos métodos de ensino, muitas vezes artificiais e focados apenas em fórmulas matemáticas, sem qualquer ligação com o mercado de trabalho ou com nossas vidas diárias.



Como Aprender Estatística Para Ciência de Dados?

Entretanto, o rápido crescimento da Ciência de Dados e da análise de dados em geral, colocou a Estatística no centro das atenções, trazendo novas formas de ensinar e aprender Estatística.



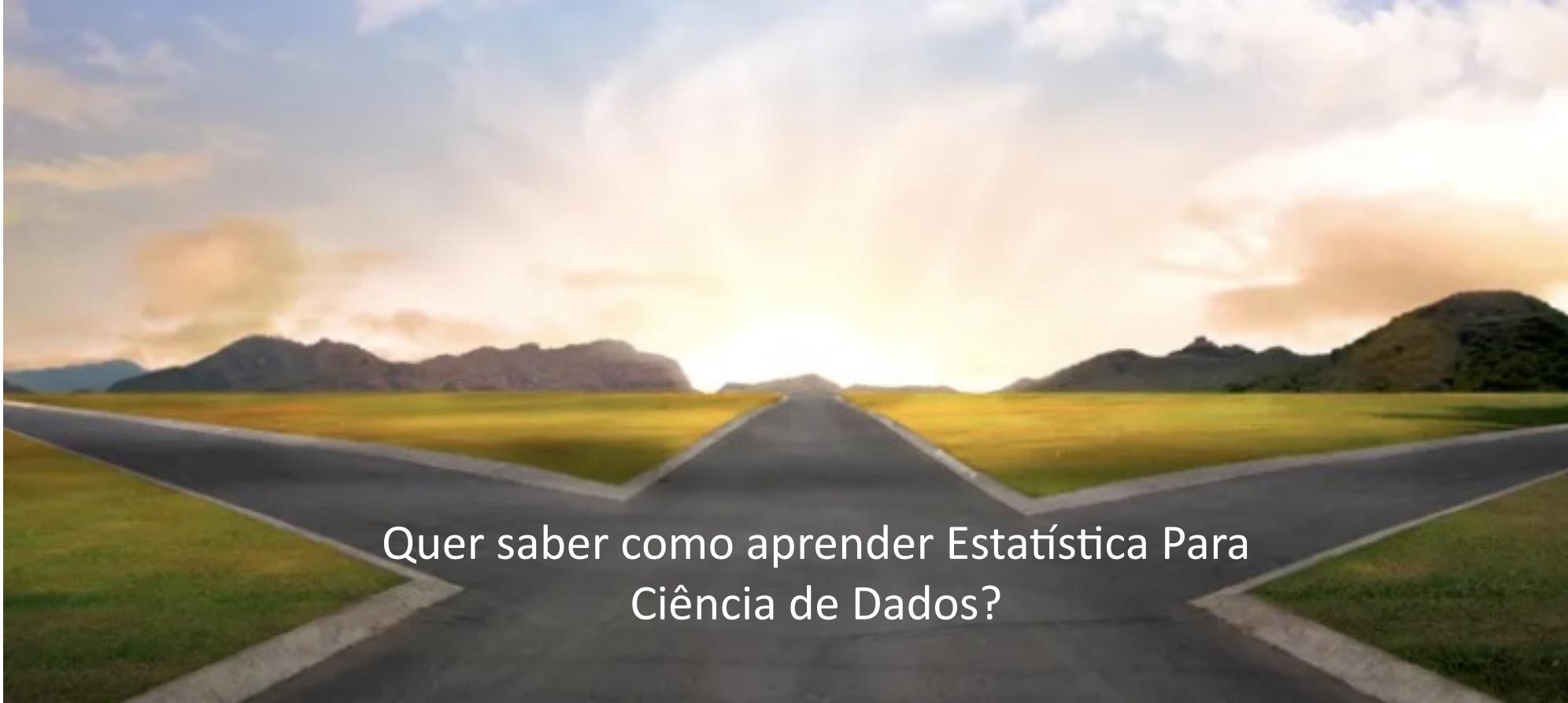
Como Aprender Estatística Para Ciência de Dados?



Aqui na DSA acreditamos que a melhor forma de aprender é fazendo! Por isso oferecemos cursos que aliam teoria e prática na medida certa, com conteúdo atual e voltado às necessidades do mercado de trabalho!

E tudo 100% em português!

Como Aprender Estatística Para Ciência de Dados?



Quer saber como aprender Estatística Para
Ciência de Dados?

Como Aprender Estatística Para Ciência de Dados?

Não espere até que você entenda completamente todos os detalhes sobre um processo, antes de começar a aplicar seu conhecimento.



Como Aprender Estatística Para Ciência de Dados?

A compreensão vem à medida que você aplica o conhecimento.



Como Aprender Estatística Para Ciência de Dados?

Quando aprendemos a falar, nós não esperamos até compreender a estrutura gramatical do idioma, antes de dizer qualquer coisa.



Como Aprender Estatística Para Ciência de Dados?

Quanto mais você aplica o conhecimento, mais você aprende!

Como Aprender Estatística Para Ciência de Dados?

Aprenda a aprender!



Como Aprender Estatística Para Ciência de Dados?

A DSA oferece 2 programas completos para quem deseja aprender Estatística na prática:

- **Formação Cientista de Dados (FCD)**
- **Formação Análise Estatística Para Cientistas de Dados (FAECD)**

A **FCD** coloca o aluno no centro de projetos de Ciência de Dados para aprender como trabalhar com esses projetos do início ao fim, incluindo Estatística.

Como Aprender Estatística Para Ciência de Dados?

A DSA oferece 2 programas completos para quem deseja aprender Estatística na prática:

- **Formação Cientista de Dados (FCD)**
- **Formação Análise Estatística Para Cientistas de Dados (FAECD)**

Na **FAECD** o aluno também trabalha em projetos, mas com foco no aprendizado específico das técnicas, procedimentos e ferramentas estatísticas.

Como Aprender Estatística Para Ciência de Dados?

A DSA oferece 2 programas completos para quem deseja aprender Estatística na prática:

- **Formação Cientista de Dados (FCD)**
- **Formação Análise Estatística Para Cientistas de Dados (FAECD)**

E por qual das duas começar?

Continue acompanhando...

Como Aprender Estatística Para Ciência de Dados?

Imagino que você já tenha se divertido com um Puzzle (quebra-cabeça em inglês). A garotada de hoje talvez não saiba o que é isso, mas quem (assim como eu) é da época pré-mundo-eletrônico já deve ter se divertido com Puzzles. Pois bem, como você monta um quebra-cabeça? Eu vejo duas abordagens:



Como Aprender Estatística Para Ciência de Dados?

1- Abordagem 1 – Você observa atentamente a figura que você quer montar (ou seja, define o objetivo na sua mente) que aqui chamaremos de “Big Picture”, e então começa a ver como as peças se encaixam. Com esta estratégia e a figura na sua mente, as peças individuais quando montadas vão fazendo cada vez mais sentido e mesmo quando você olha para uma peça isolada, seu cérebro tenta associá-la com a Big Picture.

2- Abordagem 2 – Você não observa a figura (ou muitas vezes nem mesmo tem a figura pronta) e então vai investigando as peças individualmente e tentando descobrir as ligações, até que alguma figura faça sentido.

As duas abordagens funcionam e você será capaz de montar o quebra-cabeça. Mas qual das duas é mais eficaz e mais rápida? **Em nossa opinião, a abordagem 1.**

Como Aprender Estatística Para Ciência de Dados?

Concluir a Formação Cientista de Dados dá ao aluno a visão sobre o todo, o Big Picture, sobre o que é e como funciona a Ciência de Dados em todas as suas nuances.

Se eventualmente uma ou outra peça não fizer sentido no meio do caminho, isso é absolutamente normal e esperado durante o processo de aprendizagem e podemos dedicar mais tempo aquela peça depois, sem comprometer o aprendizado do todo.



Como Aprender Estatística Para Ciência de Dados?

Vemos muitos alunos que preferem primeiro aprender tudo sobre Estatística e/ou Matemática (como se isso fosse possível), para só depois começar a aprender como trabalhar com Data Science. Isso não é errado, apenas achamos que leva muito mais tempo.

Se você já comprehende bem o que é Machine Learning e como os modelos funcionam, fica muito mais fácil depois compreender os detalhes, incluindo cada cálculo matemático e/ou estatístico por trás do modelo.



Portanto, acreditamos que concluir a FCD antes da FAECD é mais eficaz.

Aqueles que decidirem começar pela FAECD, também conseguirão montar o quebra-cabeça.

Mas não ache que você tem que aprender tudo sobre um tema, para só depois começar a estudar outro.

O verdadeiro aprendizado ocorre quando você faz conexões, liga os pontos, comprehende os relacionamentos entre os conceitos, principalmente em uma área multidisciplinar como Data Science.



Como Aprender Estatística Para Ciência de Dados?

Há muito a estudar em Ciência de Dados e se o aluno não compreender isso e traçar uma estratégia de estudo, mantendo a constância, o aprendizado nunca sai do ponto zero. Aquela estratégia de “eu estudo quando der”, não funciona! É preciso constância, para que o conhecimento vá sendo construído e meses ou até mesmo anos de estudo para se alcançar um nível razoável de conhecimento.

Comece o quanto antes e se precisar de ajuda pode contar conosco aqui na DSA. Uma das razões do Suporte DSA atender em até 24 horas (incluindo finais de semana e feriados) é exatamente estar lá quando o aluno precisar de ajuda ou tiver alguma dúvida!

Trazemos nos cursos da DSA não apenas o conhecimento técnico, mas estratégias de ensino que tornam o aluno independente, orientado à solução de problemas e que sabe aprender a aprender!

Bons estudos!



Machine Learning e MLOps



Ciência de Dados x Machine Learning

Ciência de Dados x Machine Learning

Todo projeto de Machine Learning é um projeto de Ciência de Dados!

Mas nem todo projeto de Ciência de Dados precisa envolver Machine Learning!



O Que é Machine Learning?

Machine Learning (ou Aprendizado de Máquina) é um sub campo da Inteligência Artificial que permite dar aos computadores a habilidade de aprender sem que sejam explicitamente programados para isso.



O Que é Machine Learning?

Aprendizado é a capacidade de se adaptar, modificar e melhorar seu comportamento e suas respostas, sendo portanto uma das propriedades mais importantes dos seres ditos inteligentes, sejam eles humanos ou não.



O Que é Machine Learning?



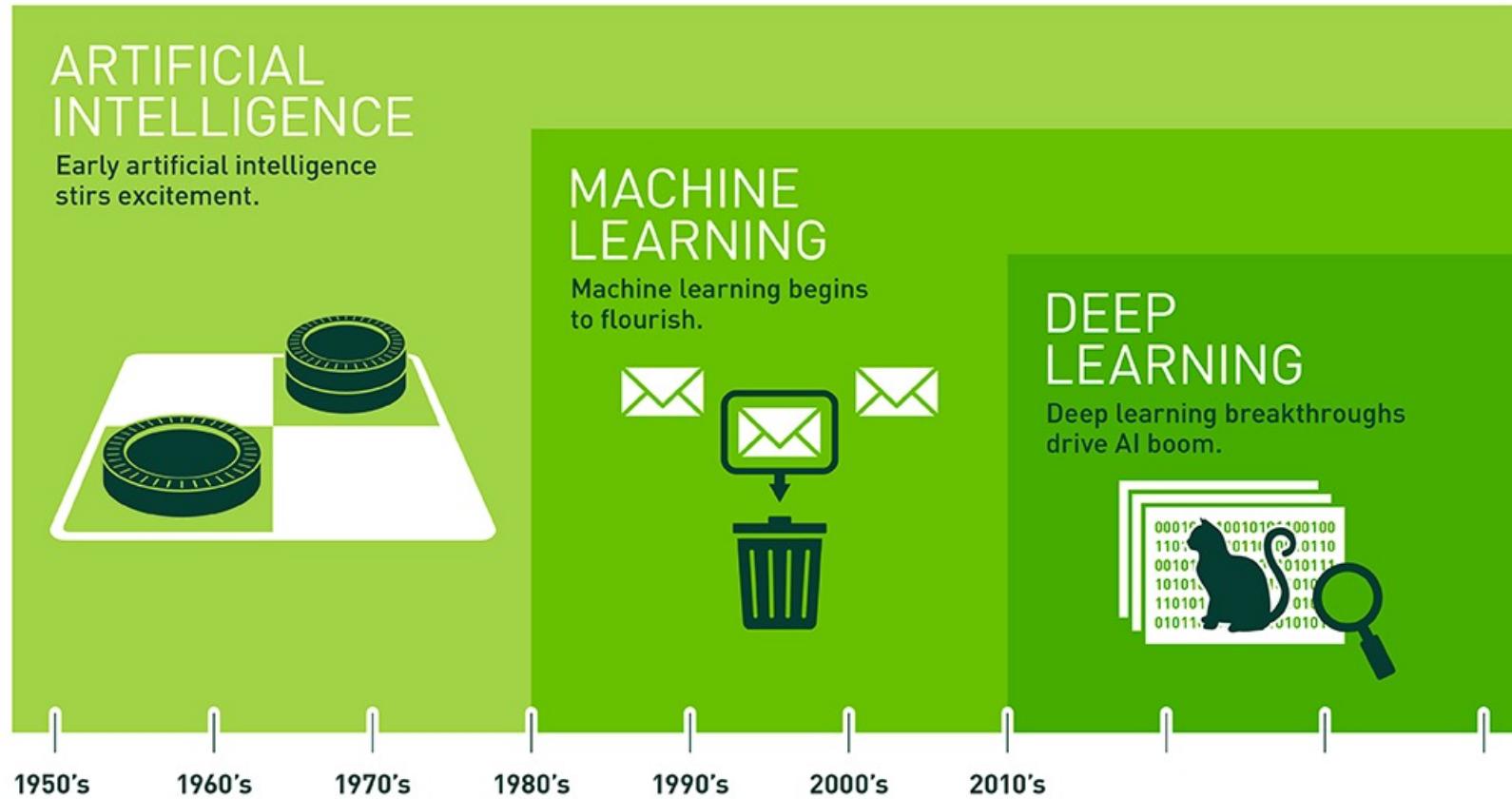
Estamos tentando reproduzir o processo de aprendizado de seres humanos em máquinas, através de algoritmos de Machine Learning.

O Que é Machine Learning?

Machine Learning (ou Aprendizado de Máquina) é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos.

Usando algoritmos que aprendem a partir de dados, o aprendizado de máquina permite que os computadores encontrem padrões ocultos sem que sejam explicitamente programados para procurar algo específico.

O Que é Machine Learning?



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

O Que é Machine Learning?

Machine Learning não é um conceito novo. Mas só agora temos dados em grande quantidade e variedade (Big Data) e capacidade de processamento (especialmente com GPUs – Unidades de Processamento Gráfico), o que nos permite criar modelos de Machine Learning que tenham alta precisão.

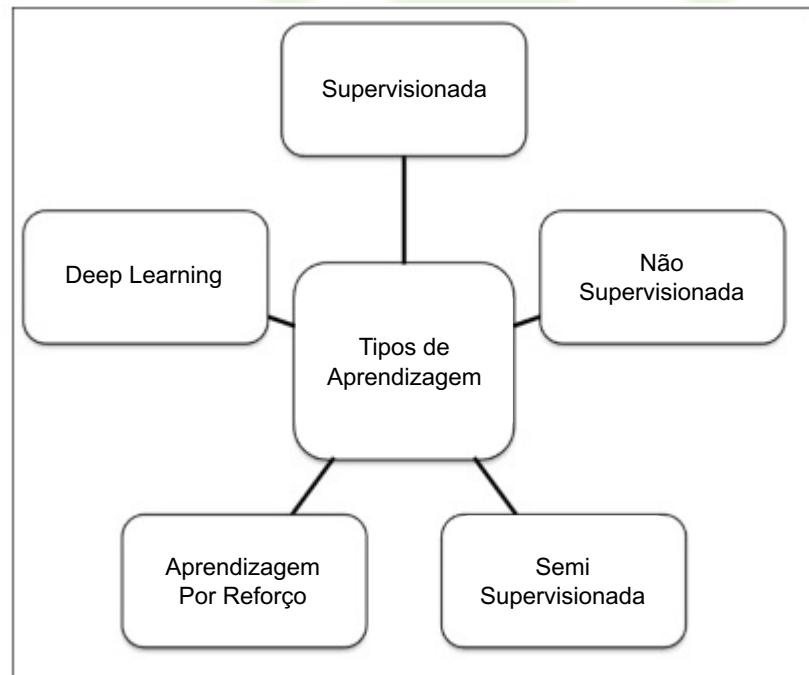


Tipos de Aprendizagem



Estamos tentando reproduzir o processo de aprendizado de seres humanos em máquinas, através de algoritmos de Machine Learning.

Tipos de Aprendizagem

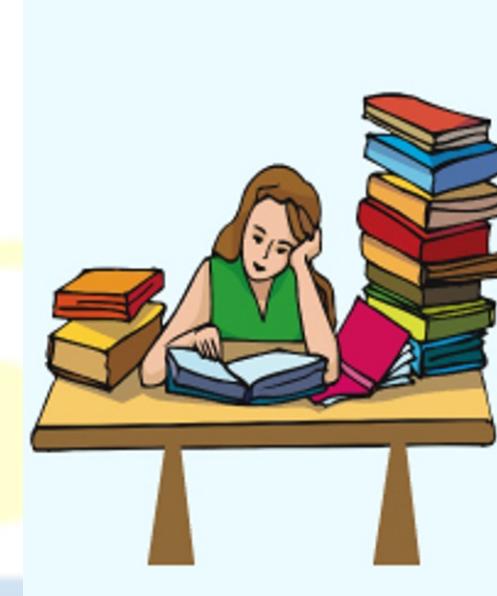


Tipos de Aprendizagem

Aprendizagem Supervisionada



Aprendizado Supervisionado



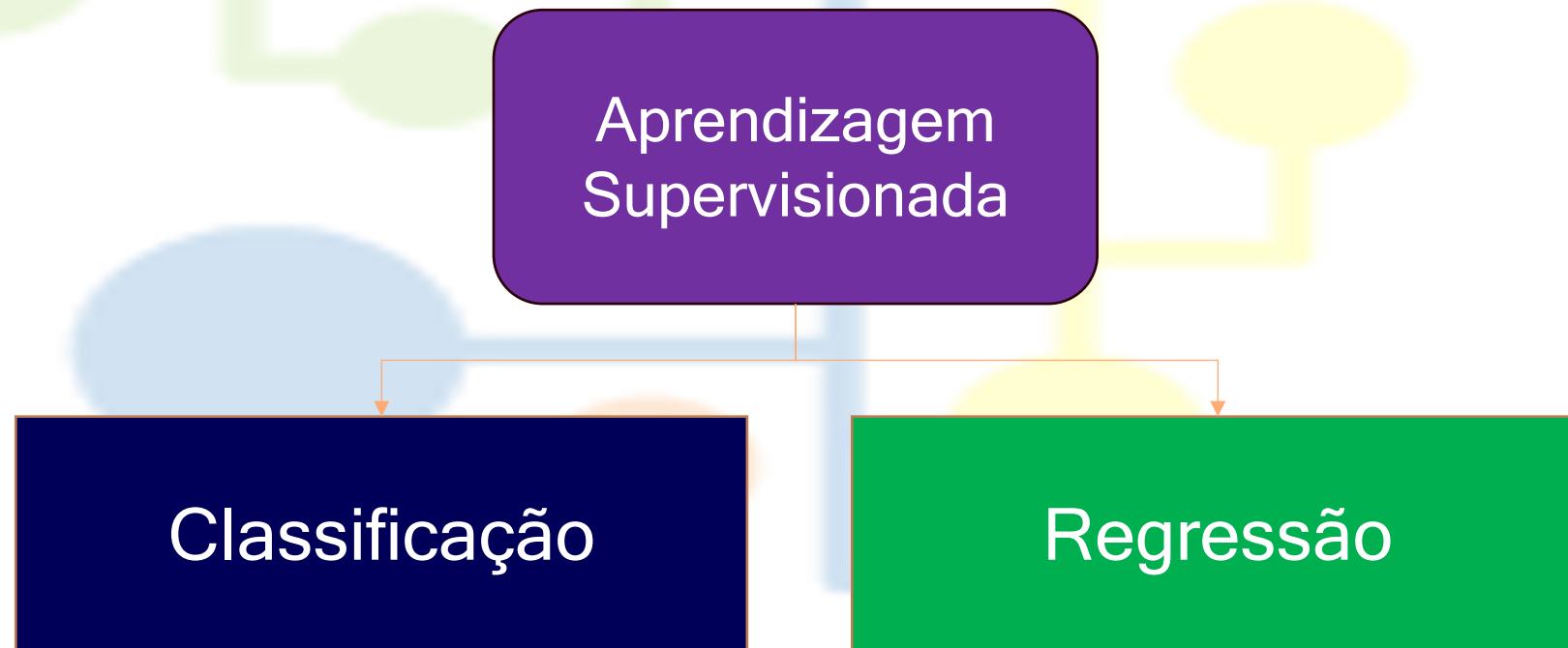
Aprendizado Não Supervisionado

Aprendizagem Supervisionada

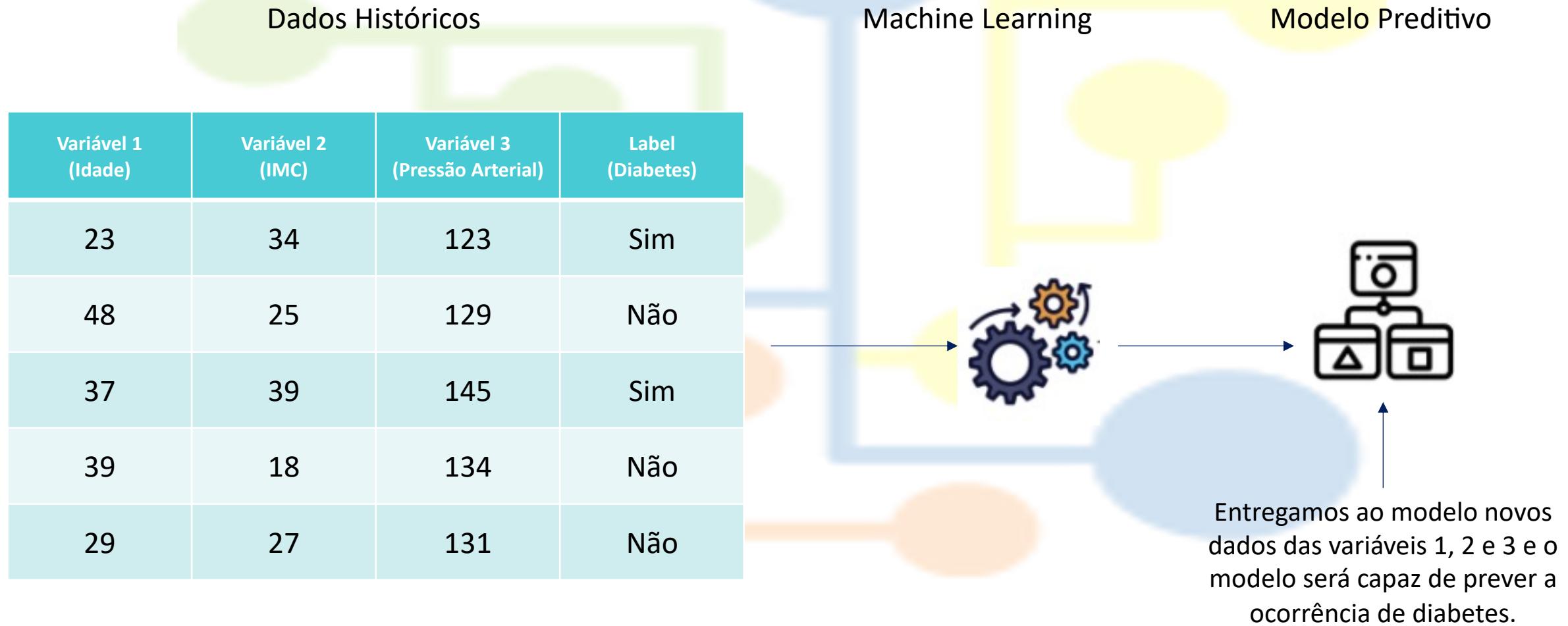


Os algoritmos de aprendizado supervisionado fazem previsões com base em conjuntos de exemplos.

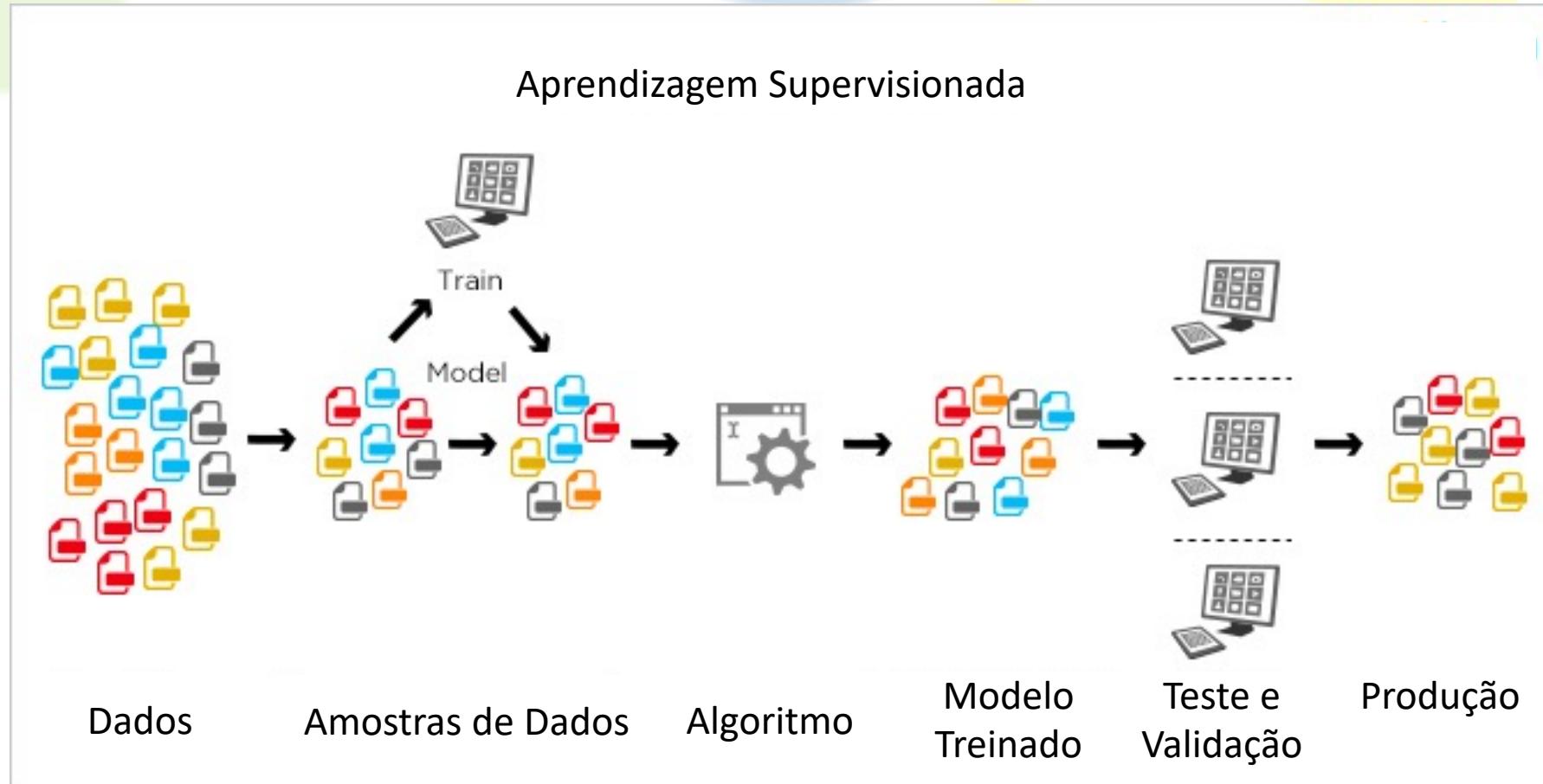
Aprendizagem Supervisionada



Aprendizagem Supervisionada



Aprendizagem Supervisionada



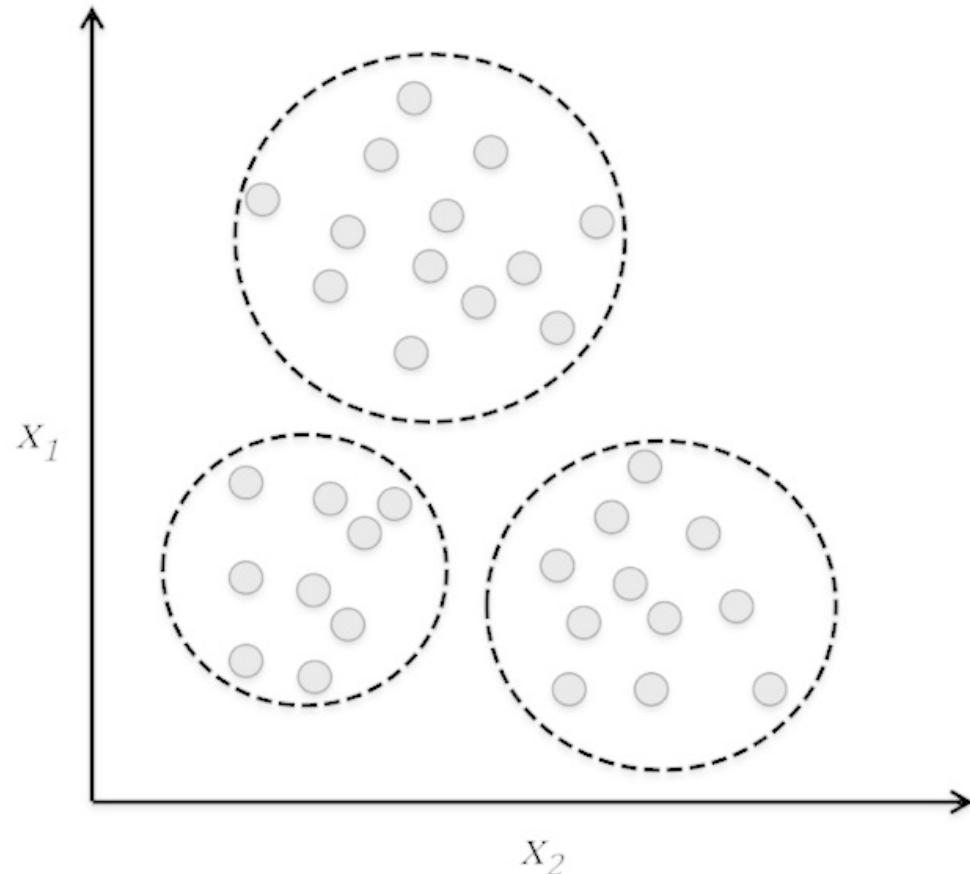


Aprendizagem Não Supervisionada

Aprendizagem Não Supervisionada



Aprendizagem Não Supervisionada



O Aprendizado Não Supervisionado é usado com dados que não possuem rótulos históricos, ou seja, nós não temos variáveis target (as variáveis de saída) para serem estimadas.

O algoritmo não recebe durante o treinamento, os possíveis resultados e ele deve descobrir por si próprio. O objetivo é explorar os dados e encontrar alguma estrutura neles.

Aprendizagem Não Supervisionada

Variável 1 (Idade)	Variável 2 (Número de Compras)	Variável 3 (Valor de Cada Compra)	Segmento
23	12	1200	A
48	3	800	C
37	4	750	B
39	18	400	A
29	9	680	A

Machine Learning

Modelo Preditivo



O modelo é capaz de agrupar os dados por similaridade encontrando grupos (segmentos) de clientes semelhantes.



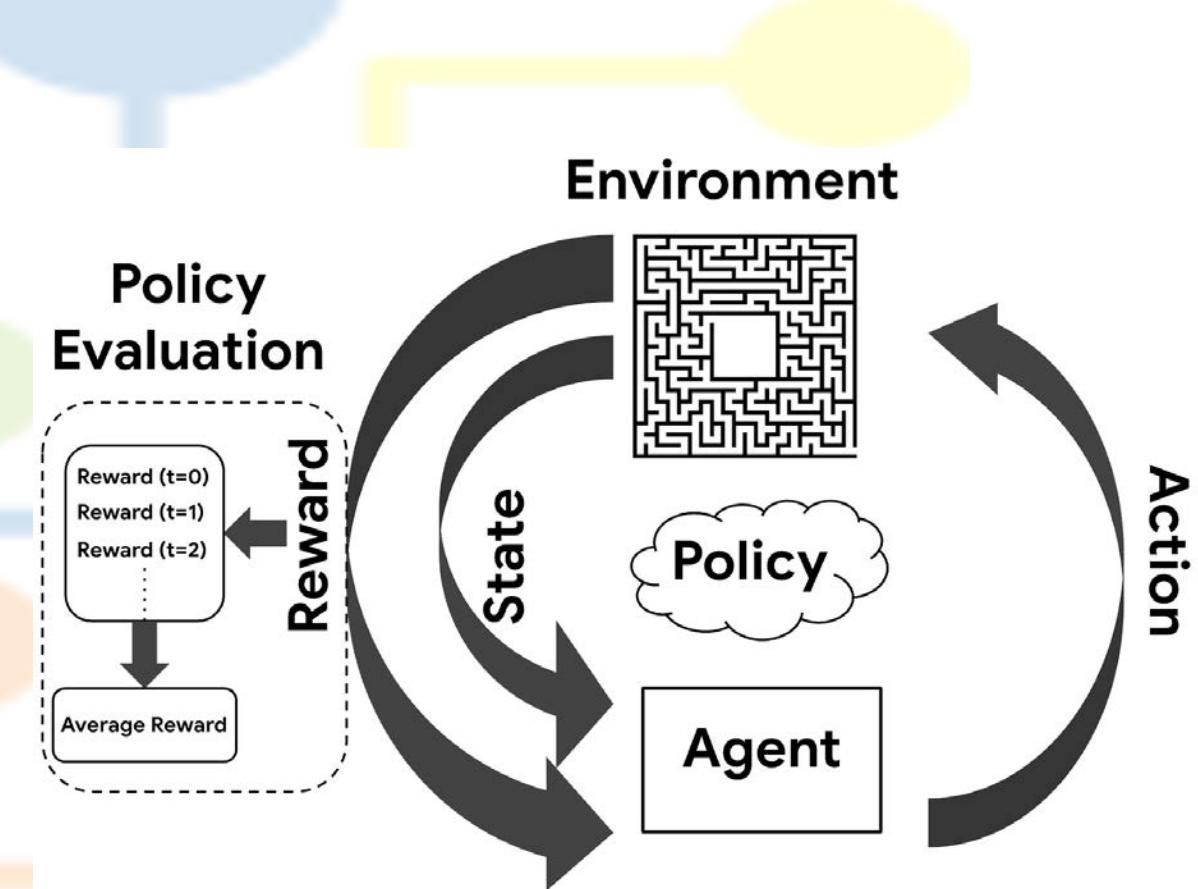
Aprendizagem Por Reforço

Aprendizagem Por Reforço

No aprendizado por reforço um agente é treinado em um ambiente e toma ações com base em políticas. A cada ação que maximiza o objetivo o agente recebe uma recompensa, aprendendo assim a melhor forma de alcançar o objetivo.

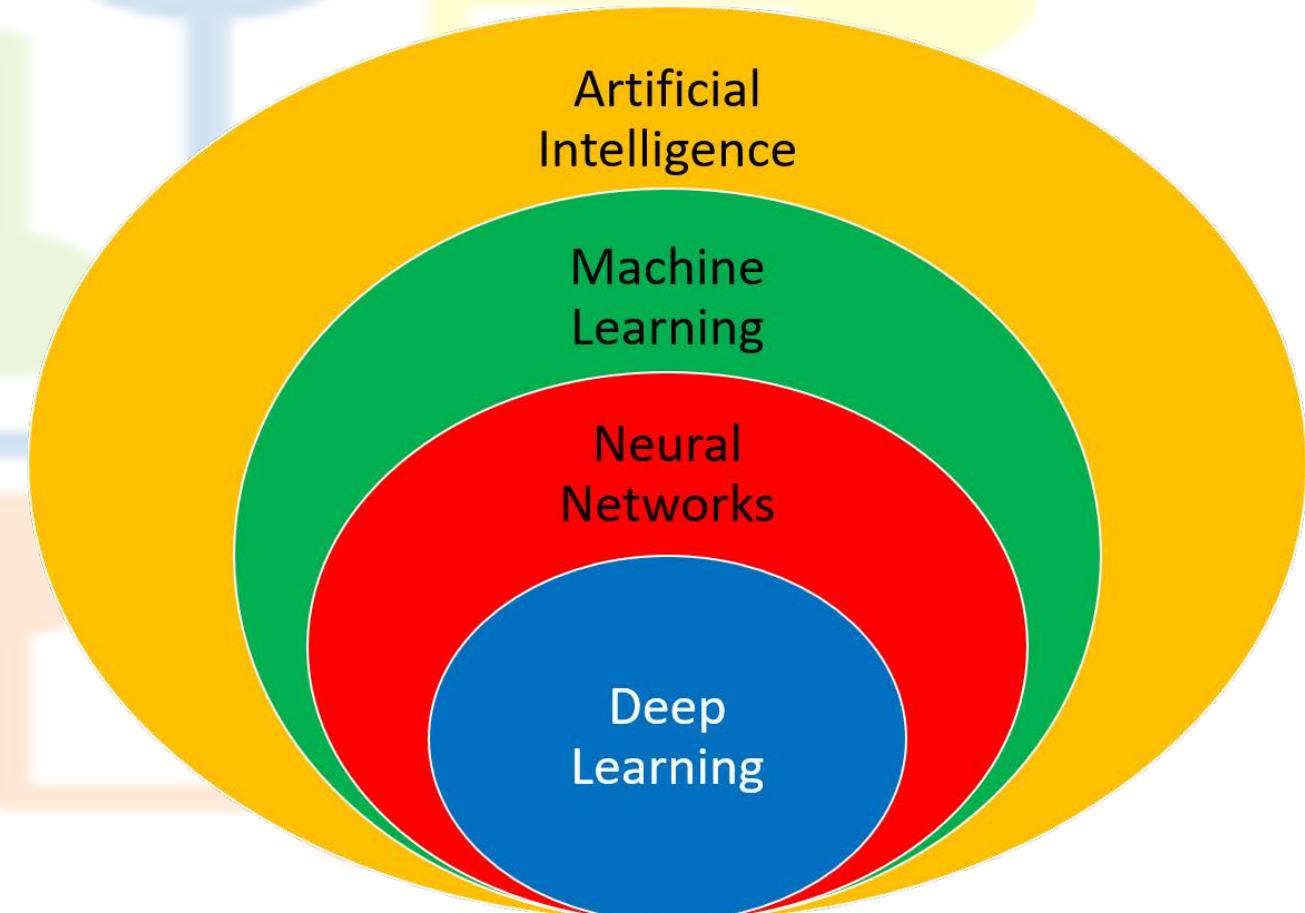
É algo similar ao aprendizado por “tentativa e erro” dos seres humanos.

Essa estratégia de aprendizado é muito usada em robótica e games. É também a principal estratégia de robôs investidores.

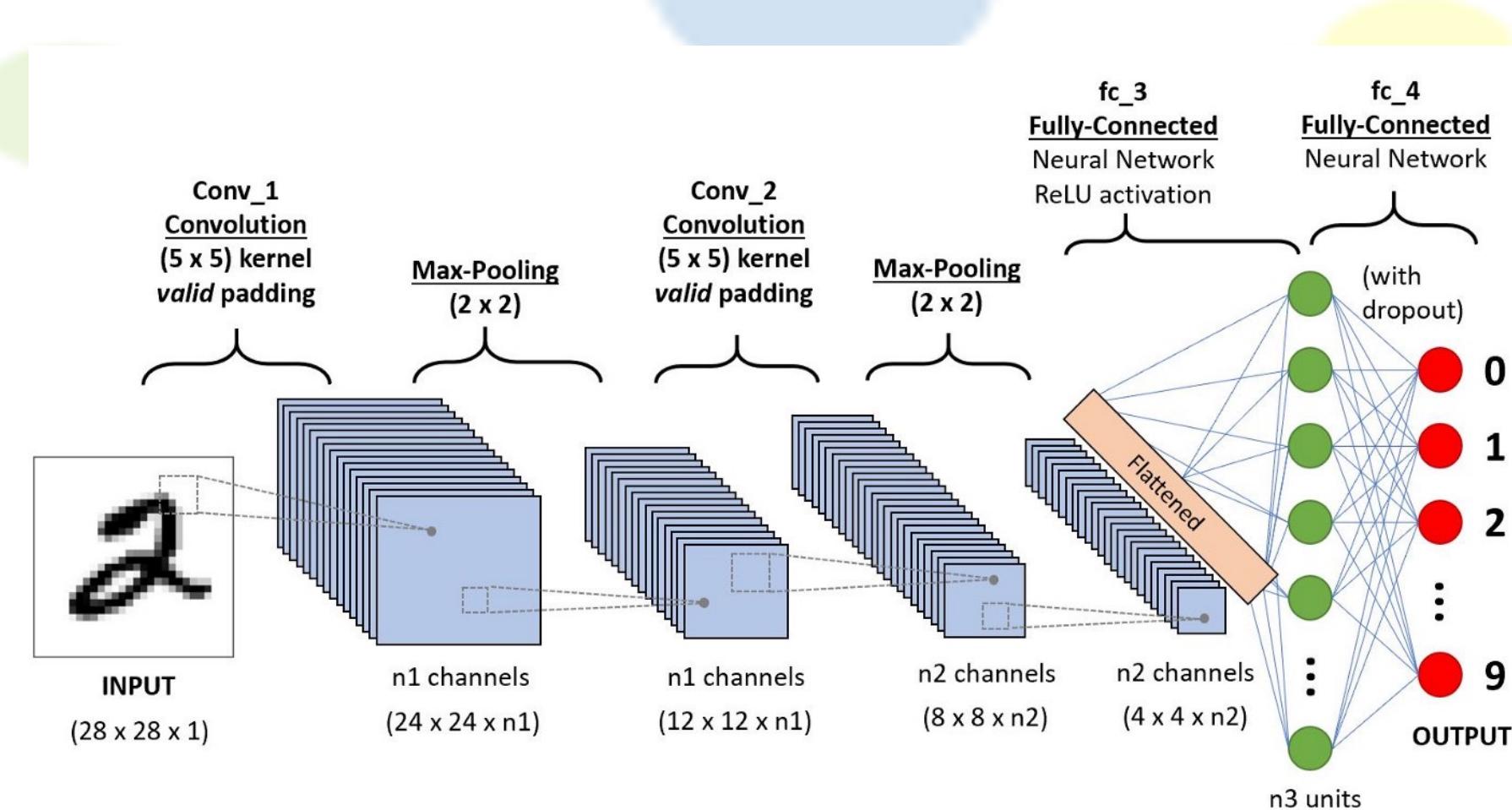


Deep Learning

Deep Learning é uma sub-categoria de Machine Learning.



Deep Learning



Deep Learning

Confira o Deep Learning Book:

<https://www.deeplearningbook.com.br>



Aplicações da Ciência de Dados



Os 5 Pré-Requisitos Para Aplicar a Ciência de Dados

Os 5 Pré-Requisitos Para Aplicar a Ciência de Dados

Estes são os 5 requisitos básicos para que possamos aplicar a Ciência de Dados:

- 1- Problema de Negócio**
- 2- Dados Históricos**
- 3- Padrão Existente nos Dados**
- 4- Capacidade Computacional Para Armazenamento e Processamento de Dados**
- 5- Profissionais Qualificados**



Em Quais Áreas Podemos Aplicar a Ciência de Dados?

Em qualquer área, desde que tenhamos um problema a ser resolvido, dados históricos estejam disponíveis, exista um padrão nos dados, seja possível armazenar e processar os dados e profissionais qualificados possam realizar o trabalho!



Exemplos de Aplicação da Ciência de Dados

Saúde

Exemplos de Aplicação da Ciência de Dados Saúde

- Prever a Ocorrência de Crises Epiléticas
- Manipulação de Dados Para Previsão de Ocorrência de Hepatite
- Análise Exploratória Para Dados de Previsão de Custos Médicos
- Análise de Chamadas Médicas de Emergência com MapReduce
- Estratificação de Risco Usando Registros Eletrônicos de Pacientes Diabéticos



Exemplos de Aplicação da Ciência de Dados Saúde

- Estratificação de Risco Usando Registros Eletrônicos de Eletrocardiogramas (ECG)
- Segmentação de Tumores em Imagens do Cérebro
- Análise de Imagens de Raio-X do Tórax e Detecção de Doenças Pulmonares
- Detecção de Anormalidades em Membros do Corpo Humano com Imagens de Radiografias
- Classificação de DNA Para Prever o Melhor Tratamento Clínico



Exemplos de Aplicação da Ciência de Dados Saúde

Todos esses exemplos citados são Projetos
e Labs da Formação IA Aplicada à
Medicina aqui na DSA.



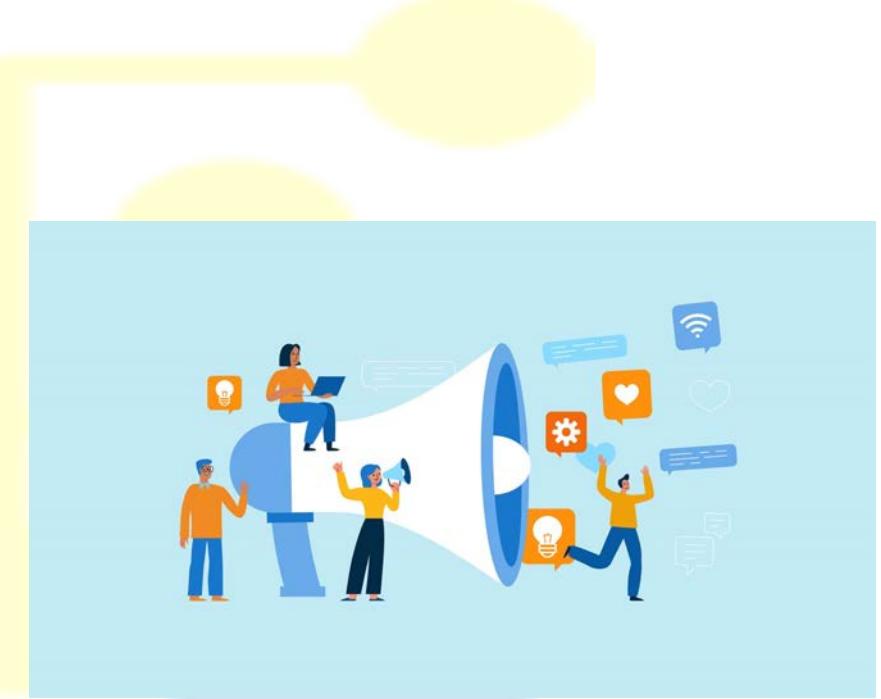


Exemplos de Aplicação da Ciência de Dados

Marketing

Exemplos de Aplicação da Ciência de Dados Marketing

- Páginas com Avaliações de Usuários Aumentam as Vendas de Produtos Online?
- Análise de Indicadores de Performance em Redes de Varejo
- Otimização de Preços e Mix de Produtos
- Data Science Aplicada ao SEO
- Análise da Efetividade de Backlinks nas Campanhas de Marketing Digital



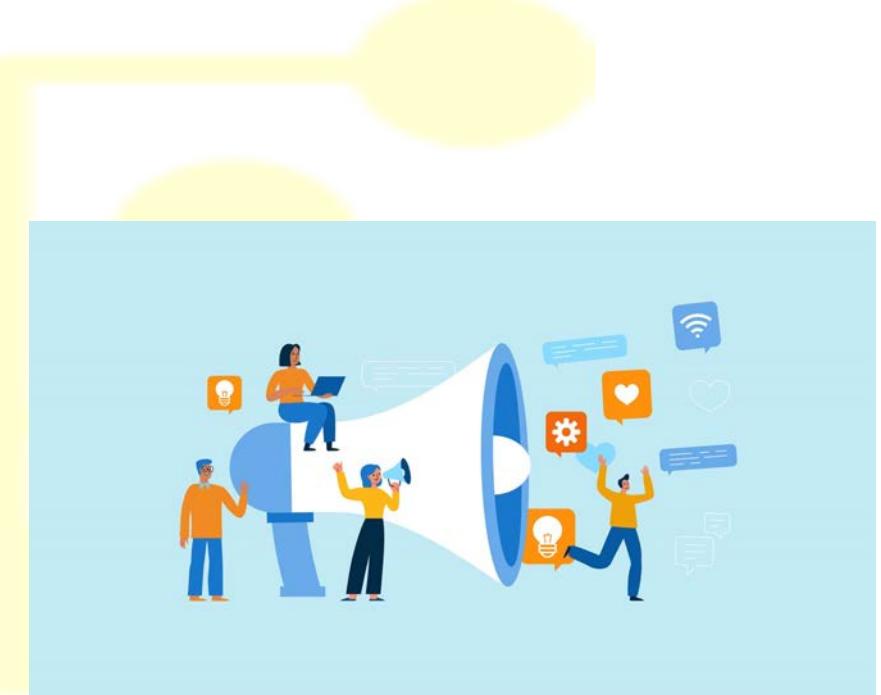
Exemplos de Aplicação da Ciência de Dados Marketing

- Geração Inteligente de Keywords e Previsão de CTR (Click Through Rate) com Machine Learning
- Otimização de Gastos nas Campanhas de Marketing
- Data Science Para Otimização da Taxa de Conversão
- Inferência Causal Para Prever a Efetividade de Campanhas de Marketing Por E-mail
- Análise de Risco Para Monetização de Produto Digital



Exemplos de Aplicação da Ciência de Dados Marketing

Todos esses exemplos citados são Projetos e Labs da Formação Cientista de Dados e Formação Desenvolver Web Para Data Science aqui na DSA.





Exemplos de Aplicação da Ciência de Dados

Finanças

Exemplos de Aplicação da Ciência de Dados Finanças

- Bot Trader - Robô Investidor Para Recomendação de Compra e Venda de Criptomoedas
- Análise de Risco em Operações Financeiras
- Análise de Redes Sociais Como Ferramenta de Detecção de Fraudes Financeiras
- Data Quality Report e Detecção de Fraudes em Transações Imobiliárias
- Text Analytics Para Analisar a Reação do Mercado Sobre as Notícias de Uma Empresa



Exemplos de Aplicação da Ciência de Dados Finanças

- Previsão de Lucro com Ativos Financeiros
- Otimização e Análise de Risco de Portfólio
- Automação da Auditoria Financeira e Contábil
- Detecção de Anomalias em Transações com FinTechs
- Análise de Portfólio de Fundos de Hedge



Exemplos de Aplicação da Ciência de Dados Finanças

Todos esses exemplos citados são
Projetos e Labs da Formação Cientista
de Dados e Formação Engenheiro
Blockchain aqui na DSA.





Exemplos de Aplicação da Ciência de Dados

Vendas

Exemplos de Aplicação da Ciência de Dados Vendas

- Previsão de Demanda e Vendas com Machine Learning
- Análise Preditiva em Vendas B2B
- Engajamento e Análise de Sentimento do Cliente B2B
- Previsão do Resultado das Eleições com Base nos Relatórios Financeiros de Gastos de Campanha
- Previsão do Volume de Entregas de Uma Transportadora ao Longo do Tempo



Exemplos de Aplicação da Ciência de Dados Vendas

- Previsão de Demanda de Pedidos
- Análise de Riscos em Supply Chain
- Previsão de Tendências Macroeconômicas
- Sistema de Recomendação Para Rede de Varejo
- People Analytics - Quais Fatores Mais Causam Atritos no Ambiente de Trabalho de Equipes de Vendas?



Exemplos de Aplicação da Ciência de Dados Vendas

Todos esses exemplos citados são
Projetos e Labs da Formação IA Para
Vendas e Formação Cientista de Dados
aqui na DSA.





Business Intelligence

x

Ciência de Dados

Business Intelligence x Ciência de Dados

Business Intelligence não é a mesma coisa que Data Science (Ciência de Dados).

Business Intelligence x Ciência de Dados

Business Intelligence e Data Science
tem muita coisa em comum e
Cientistas de Dados e Analistas de BI
são como primos.



Business Intelligence x Ciência de Dados

Ambos usam dados com o mesmo objetivo, mas a sua abordagem, tecnologia e função diferem de diversas maneiras.



Business Intelligence x Ciência de Dados

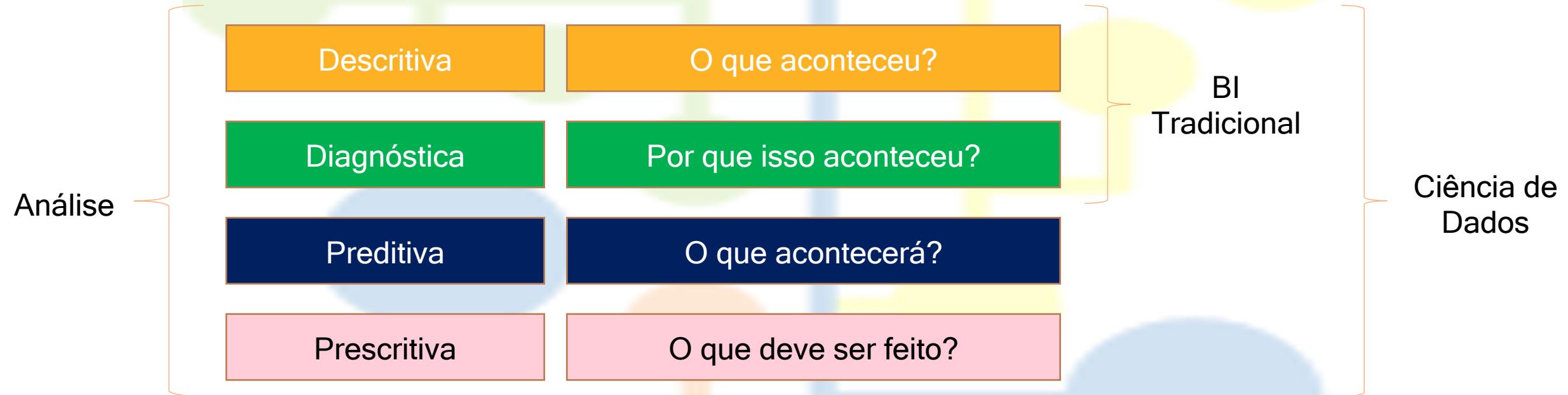
Business
Intelligence

O objetivo do BI é converter dados brutos em insights de negócio, que os líderes empresariais e gestores possam usar para tomar decisões.

Ciência de
Dados

A Ciência de Dados também converte dados brutos em insights de negócio, mas aplica metodologia científica para exploração dos dados, testes de hipótese, modelagem estatística e aprendizado de máquina.

Business Intelligence x Ciência de Dados



O diagrama acima demonstra as diferenças entre BI e Data Science. O tema é abordado em mais detalhes nos cursos da Formação Cientista de Dados aqui na DSA.



Ciclo de Vida de Projetos de Data Science



O Que é um Projeto?

O Que é um Projeto?

A palavra projeto vem da palavra latina projectum, que significa “antes de uma ação”.

Ao longo do tempo, o sentido de “projeto” foi sendo modificado e ampliado, e hoje há vários tipos de projetos, envolvendo desde projetos pessoais até projetos de rodovias, por exemplo.

Fazer um plano antes de colocar algo em prática tornou-se prioridade e hoje há uma disciplina gerencial que se encarrega especialmente disso: a Gestão de Projetos.



O Que é um Projeto?

Um projeto é um esforço único, temporário e progressivo empreendido para criar um produto, serviço ou resultado exclusivo.

Isso significa que um projeto é uma ação específica que tem início e fim determinados (é, portanto, uma ação temporária), e um objetivo claro a ser atingido dentro dos recursos que são destinados a ele (financeiros, materiais e humanos). Geralmente, os projetos são divididos em etapas, as quais vão sendo executadas para gerar entregas.



O Que é um Projeto?

Os projetos não devem ter uma duração indefinida e nem recursos ilimitados — ambos devem ser planejados previamente para estabelecer uma linha de base a ser seguida.

Segundo o Guia PMBOK®, o sucesso de um projeto é medido pela qualidade do produto e do projeto, pela pontualidade, pelo cumprimento do orçamento e pelo grau de satisfação do cliente.

Em Ciência de Dados trabalhamos em diferentes projetos, para resolver diferentes problemas em diferentes áreas de negócio e com diferentes ferramentas.



O Que é um Projeto?

Principais Características de um Projeto:



Projetos de Ciência de Dados

Todo projeto de Data Science deve começar com o objetivo, ou seja, as questões que precisam ser respondidas.



Projetos de Ciência de Dados

Depois de formuladas as questões, busca-se os dados que ajudarão a respondê-las.



Projetos de Ciência de Dados

A maturidade da disciplina de Gestão de Projetos, permite que boas práticas de gerenciamento de projetos, como as definidas pelo PMI (Project Management Institute), possam ser utilizadas em projetos de Data Science.



Projetos de Ciência de Dados

Não há porque reinventar a roda.

Conhecimento e experiência adquiridos em Gestão de Projetos, serão fundamentais para o sucesso de projetos de Data Science.



Projetos de Ciência de Dados



Dica!

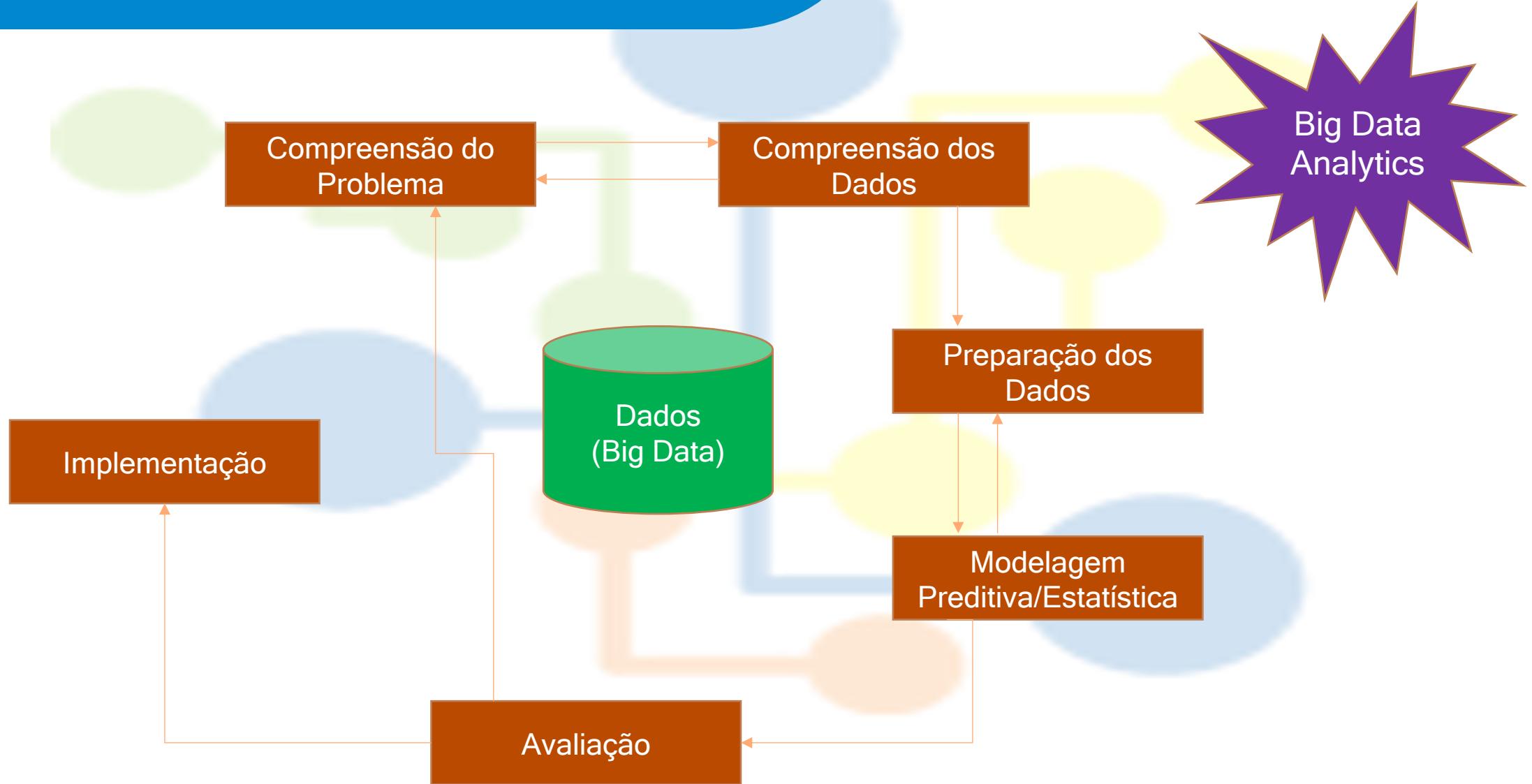
Inicialmente, é vital compreender o problema a ser resolvido.

Avaliar e reavaliar o problema e desenhar uma solução, é um processo iterativo de descoberta.

Isso pode parecer óbvio, mas problemas de negócios raramente vem empacotados de forma clara.

Na prática, o maior desafio em Ciência de Dados, está na identificação do problema.

Projetos de Ciência de Dados





Exemplos de Definição do Problema de Negócio

Exemplos de Definição do Problema de Negócio

Identificar o problema é uma das tarefas principais em um projeto de Data Science!



Atenção!!

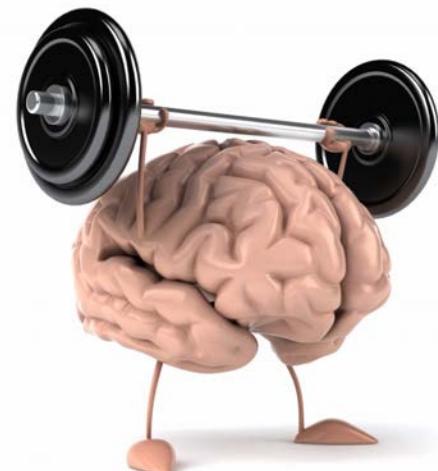
Não subestime esta etapa.

A identificação correta do problema é a chave para o sucesso de projetos de Ciência de Dados.

Exemplos de Definição do Problema de Negócio

Vamos fazer um exercício:

Vamos elaborar algumas perguntas para identificar o problema e a possível solução a ser adotada!



Exemplos de Definição do Problema de Negócio

Quais são os clientes mais rentáveis?

Esta pergunta poderia facilmente ser respondida com uma consulta simples a um banco de dados, usando Linguagem SQL, por exemplo.

Exemplos de Definição do Problema de Negócio

Existe diferença entre os clientes mais rentáveis e a média dos clientes?

Aqui poderíamos usar a Estatística e realizar um teste de hipótese com Linguagem R para confirmar ou rejeitar nossa tese que os clientes mais rentáveis possuem diferença em relação à média dos clientes.

Exemplos de Definição do Problema de Negócio

Algum cliente em particular estará no grupo dos mais rentáveis?
Quanto de faturamento posso esperar vindo desse cliente?

Aqui poderíamos usar análise exploratória com a Linguagem Python a fim de examinar o histórico dos clientes e então usar Machine Learning para criar modelos preditivos e prever se o cliente estará no grupo dos mais rentáveis, quando isso pode ocorrer e ainda prever o faturamento futuro com esse cliente.

Exemplos de Definição do Problema de Negócio

A identificação do problema a ser resolvido, vai impactar todo o projeto de Data Science, não só com relação ao custo, mas também com relação aos recursos que serão adotados.



Exemplos de Definição do Problema de Negócio

E quanto mais complexas as perguntas,
mais a Ciência de Dados pode ajudar
nas respostas!!

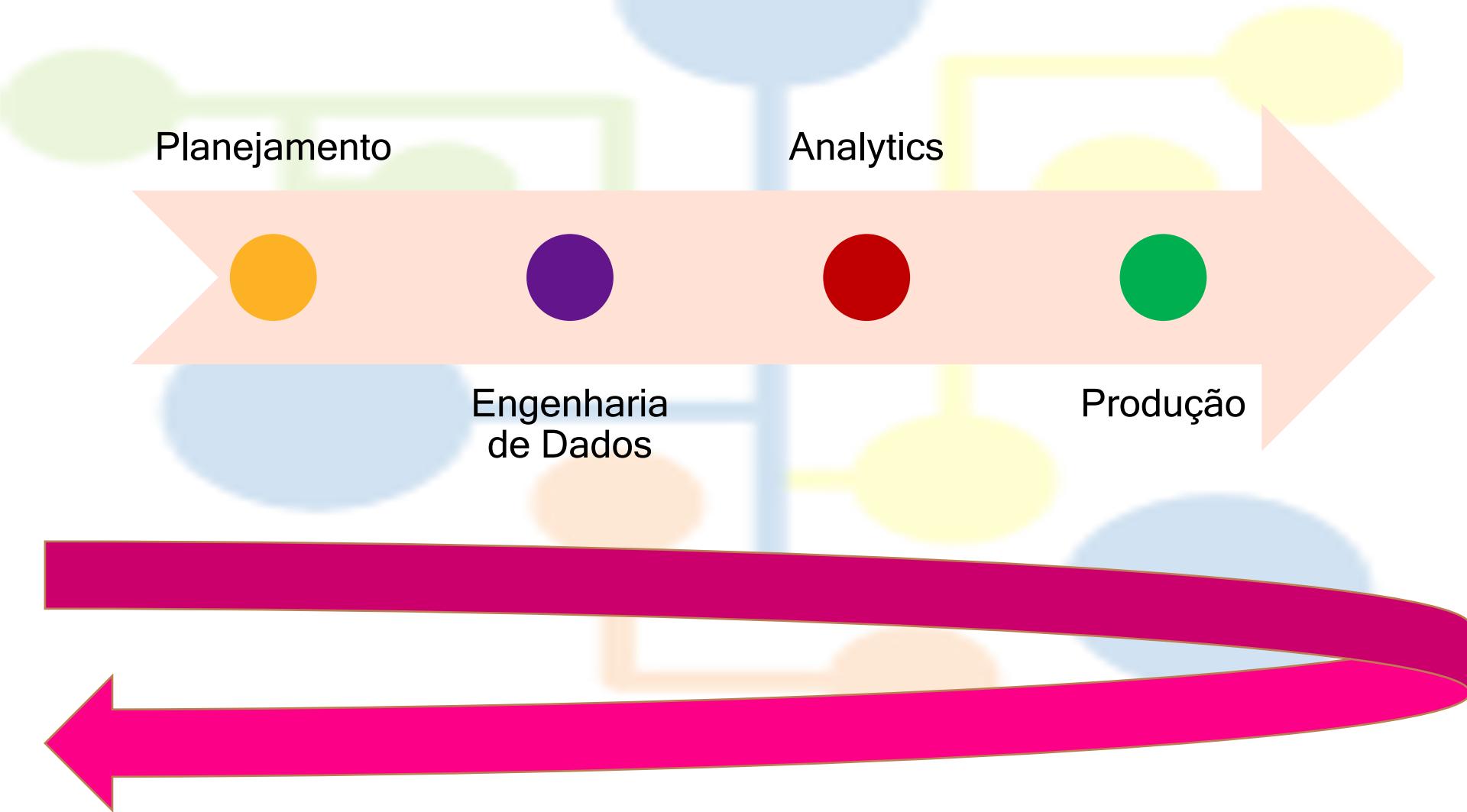


Exemplos de Definição do Problema de Negócio

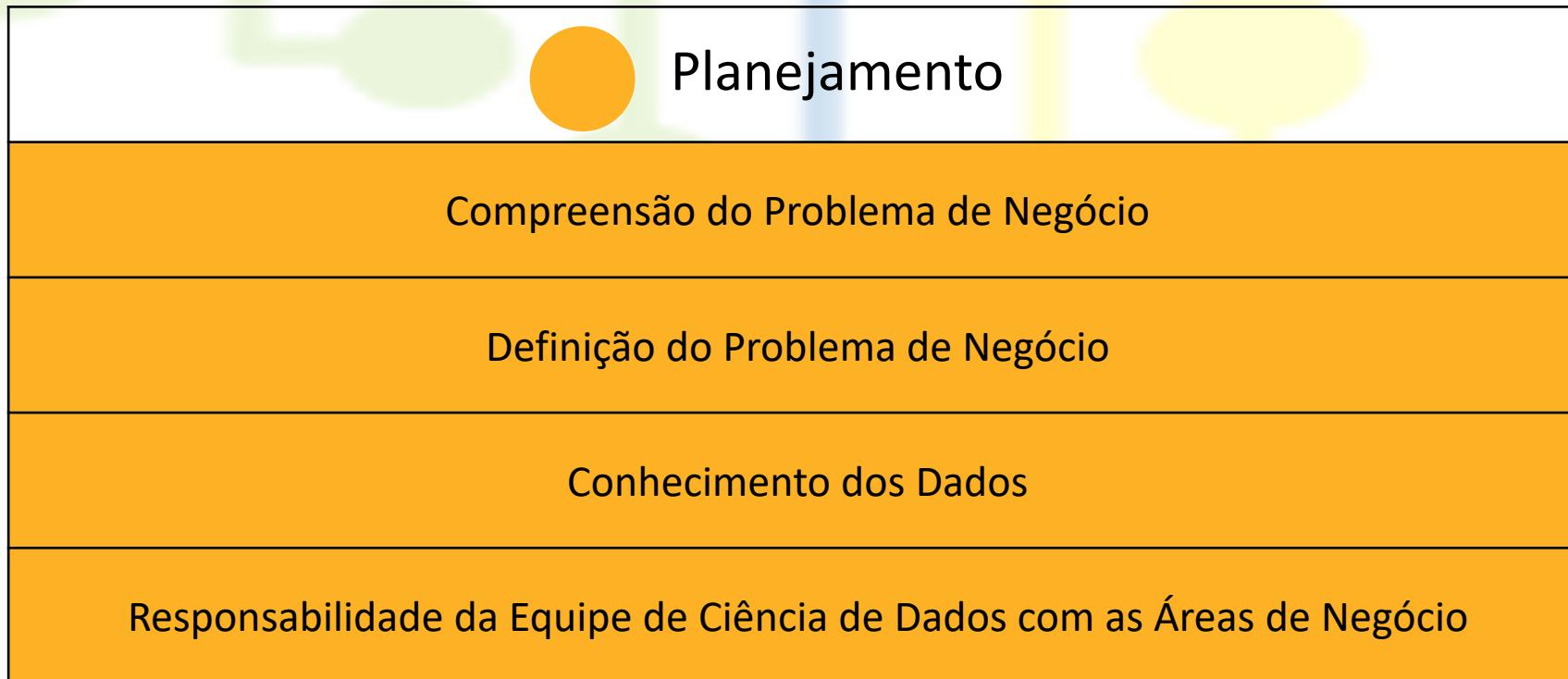


A melhor tecnologia é aquela que
resolve o problema do cliente!

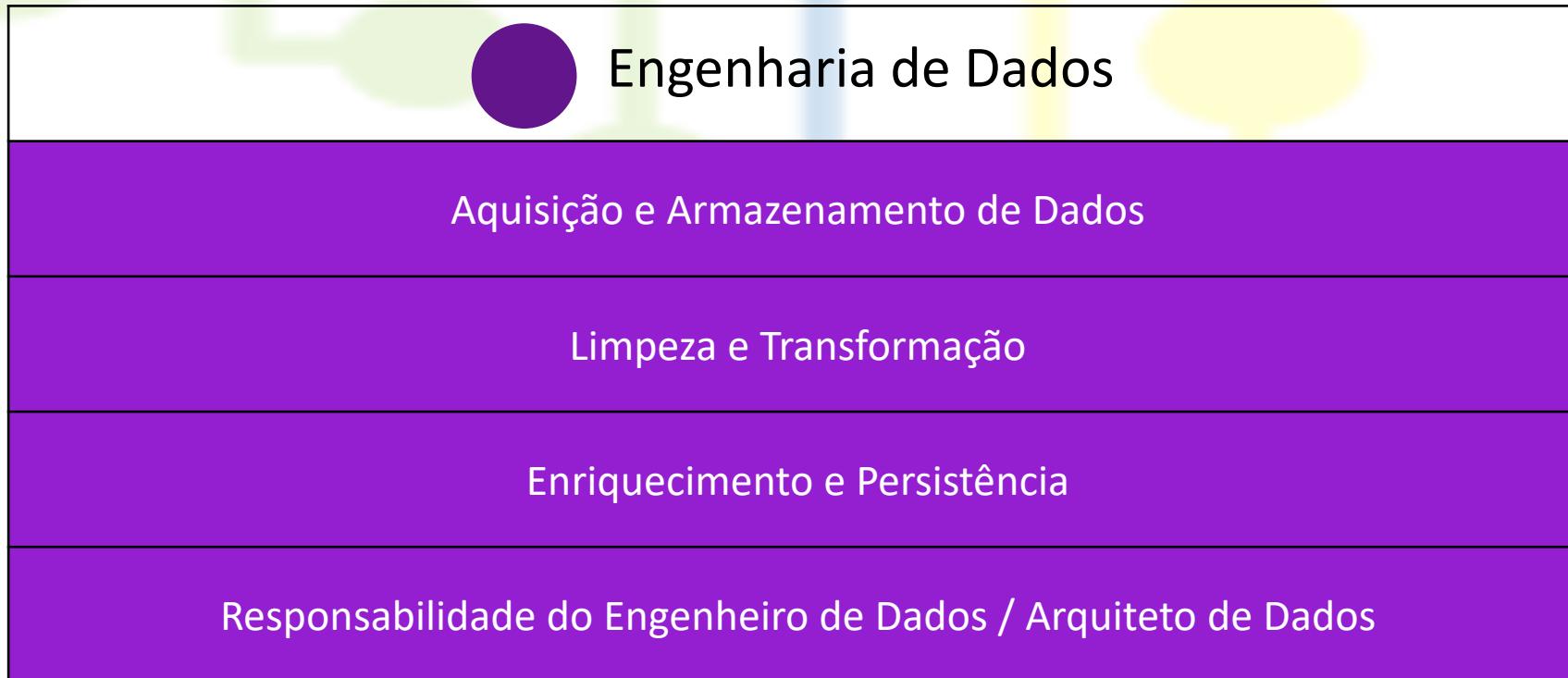
Ciclo de Vida de Projetos de Ciência de Dados



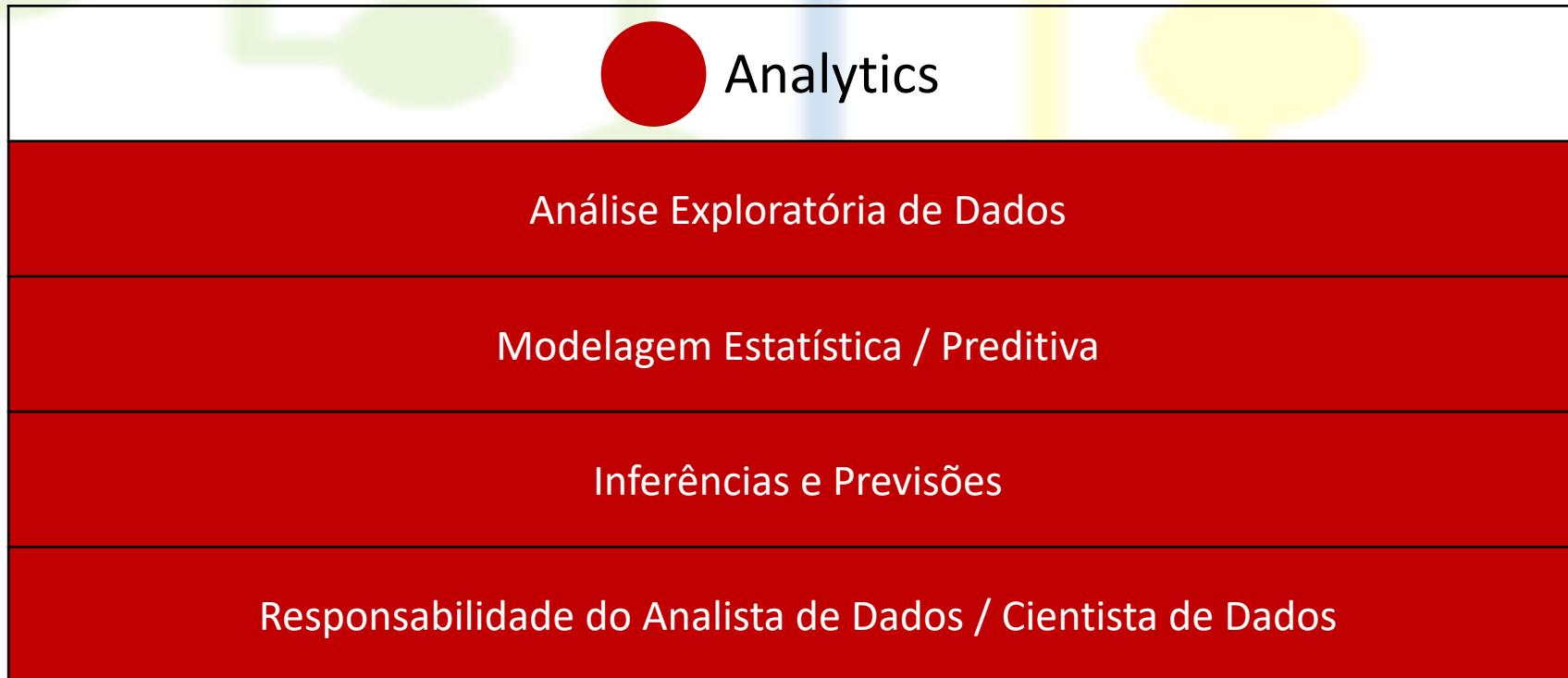
Ciclo de Vida de Projetos de Ciência de Dados



Ciclo de Vida de Projetos de Ciência de Dados



Ciclo de Vida de Projetos de Ciência de Dados



Ciclo de Vida de Projetos de Ciência de Dados

Produção

Construção de Produtos de Dados

Pipelines de Dados e Pipelines de Machine Learning

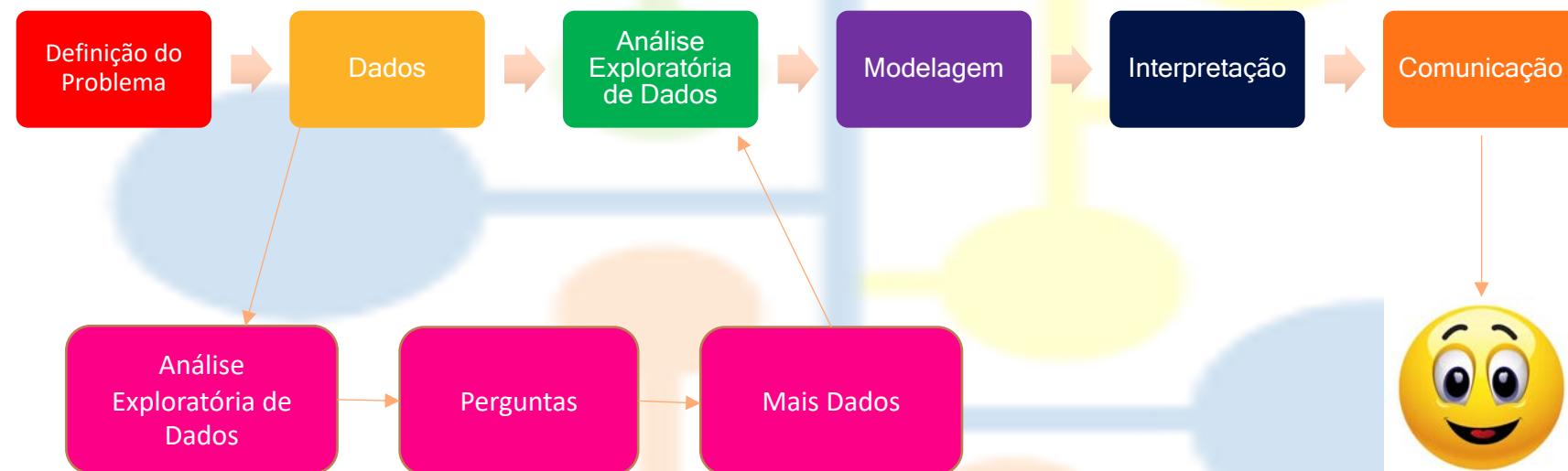
Melhoria Contínua

Responsabilidade do Engenheiro de Machine Learning / Engenheiro Data Ops



Fases de Projetos de Ciência de Dados

Fases de Projetos de Ciência de Dados





Produtos Gerados

Produtos Gerados

A qualidade dos seus outputs será determinada pelas qualidade dos seus inputs.

Bons resultados dependem da formulação de boas questões!

Produtos Gerados

Produtos Gerados em Projetos de Data Science



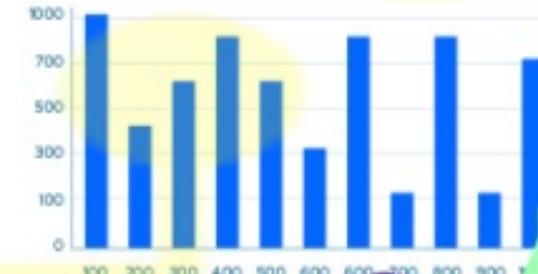
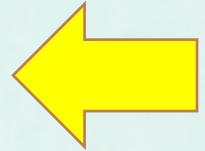


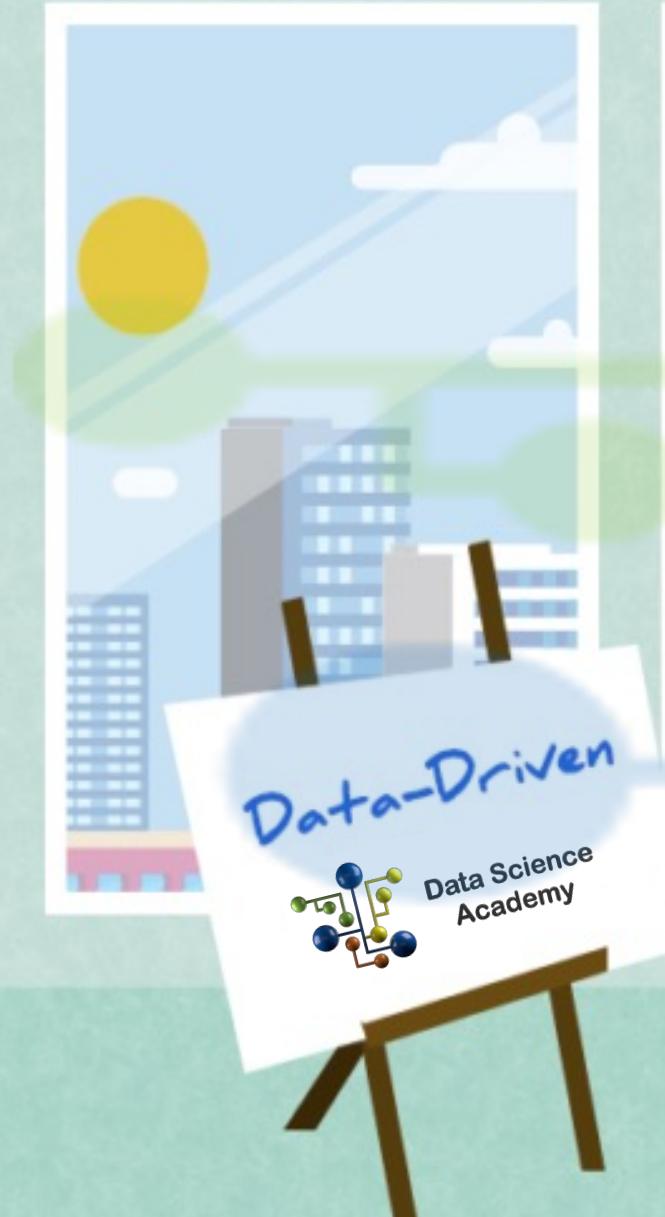
Cultura Orientada a Dados



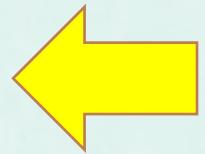


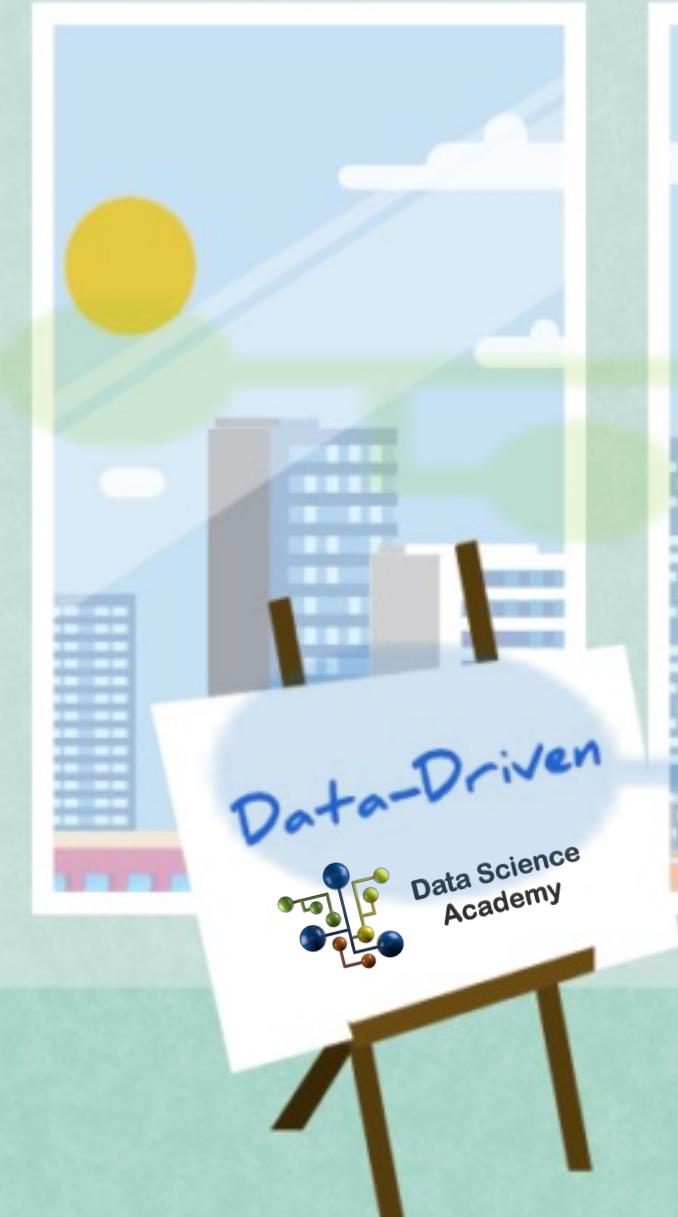
Uma empresa com cultura orientada a dados, reconhece a importância do Big Data e da Ciência de Dados para tomar melhores decisões e oferecer melhores produtos e serviços.



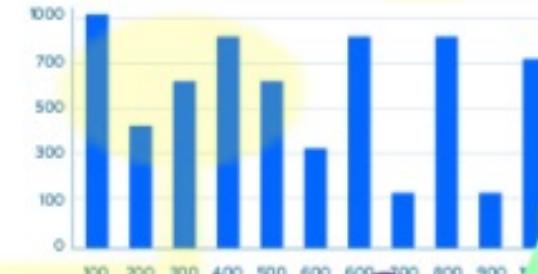
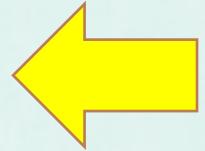


A mudança de paradigma será inevitável.
Aliás, já estamos vivenciando esta
mudança. Terá sucesso, quem melhor se
adaptar a ela.





Os dados já fazem parte do negócio.





Soft Skills e Carreiras em Data Science



O Mercado de Ciência de Dados e Big Data

O Mercado de Ciência de Dados e Big Data

Iniciamos a DSA em Março/2016 quando o mercado brasileiro ainda nem falava sobre Ciência de Dados. Hoje o mercado está super aquecido e já são mais de 500.000 alunos capacitados em nosso portal com uma demanda cada vez maior.

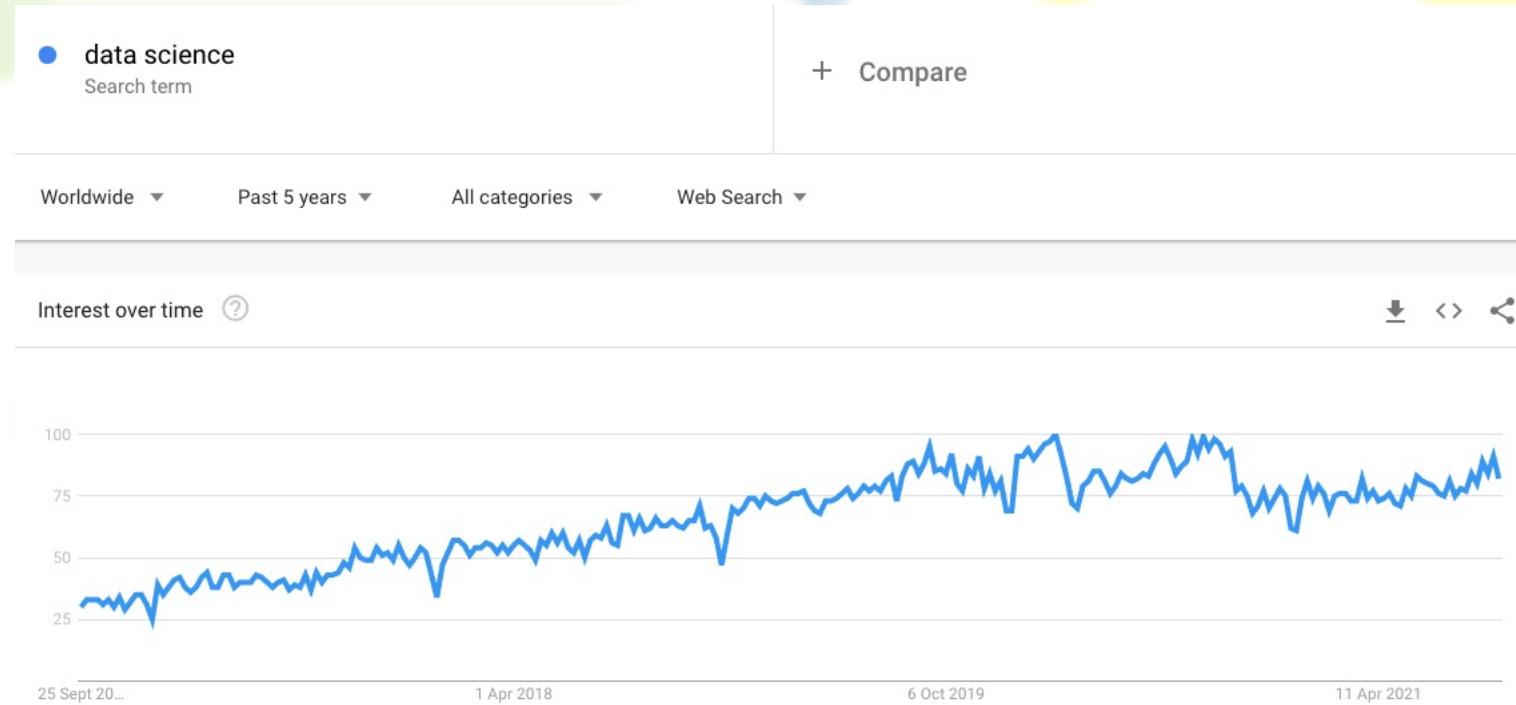


O Mercado de Ciência de Dados e Big Data

Com a explosão da análise de dados e do Big Data, a busca por profissionais capazes de extrair, analisar e gerar insights dos dados, tomou conta de todos os segmentos de negócios uma vez que toda e qualquer área gera dados, a nossa matéria-prima.



O Mercado de Ciência de Dados e Big Data



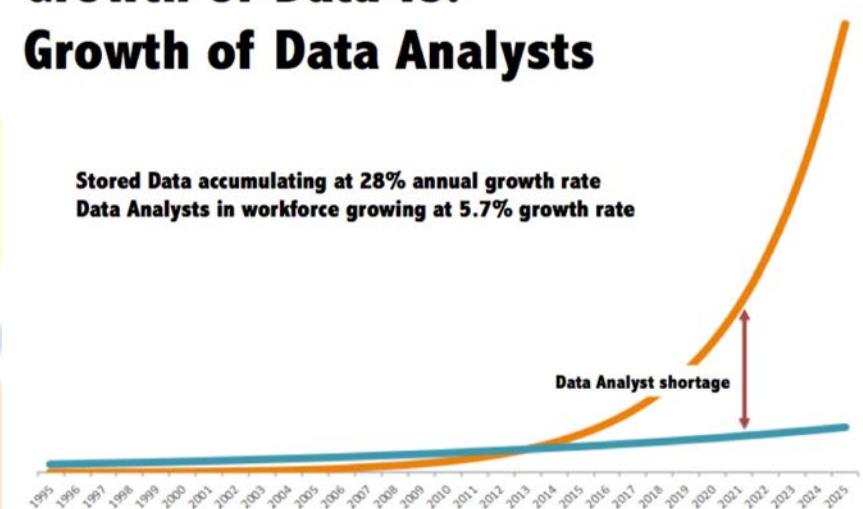
Fonte: <https://trends.google.com>

O Mercado de Ciência de Dados e Big Data

Crescimento do volume de dados x Crescimento da formação de Analistas de Dados

O armazenamento de dados aumenta 28% ao ano, enquanto a formação de profissionais de Análise de Dados aumenta a 5,7% ao ano.

Growth of Data vs. Growth of Data Analysts



<http://www.delphianalytics.net/more-data-than-analysts-the-real-big-data-problem/>

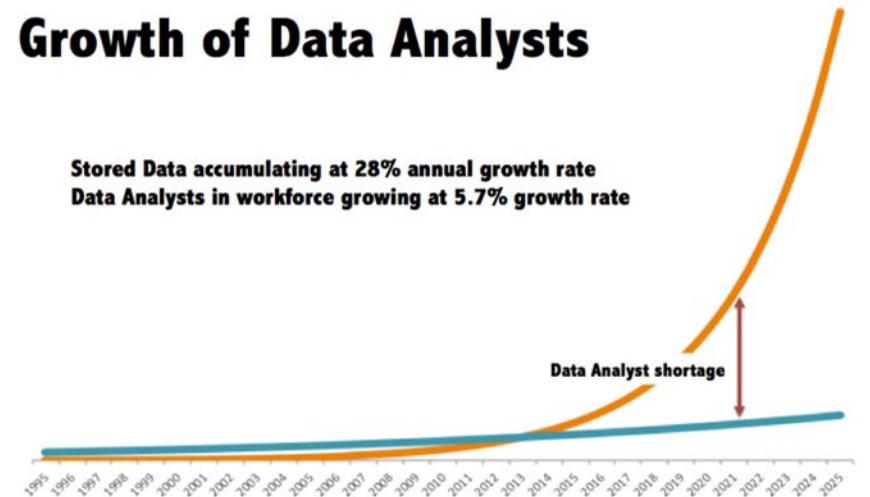
O Mercado de Ciência de Dados e Big Data

Crescimento do volume de dados x Crescimento da formação de Analistas de Dados

Neste ritmo, o que vai acontecer
daqui 5 anos?

Growth of Data vs. Growth of Data Analysts

Stored Data accumulating at 28% annual growth rate
Data Analysts in workforce growing at 5.7% growth rate



<http://www.delphianalytics.net/more-data-than-analysts-the-real-big-data-problem/>

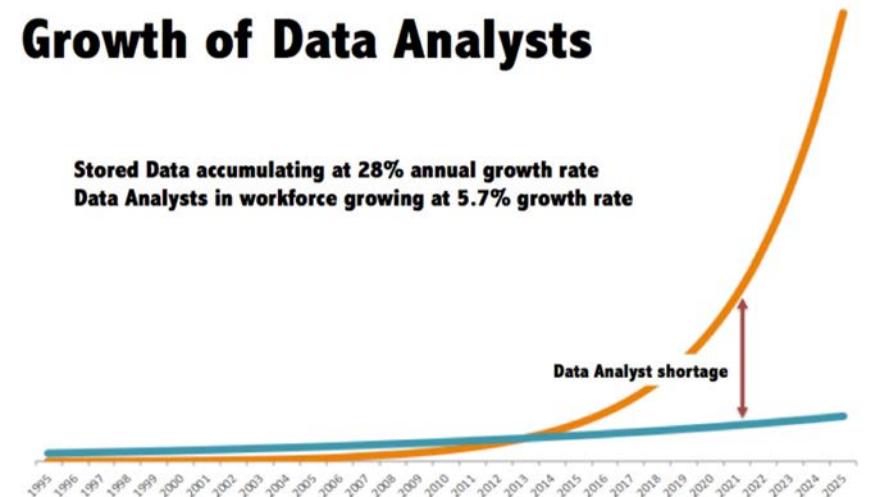
O Mercado de Ciência de Dados e Big Data

Crescimento do volume de dados x Crescimento da formação de Analistas de Dados

O mercado de trabalho não terá profissionais suficientes para suprir a demanda!!!!

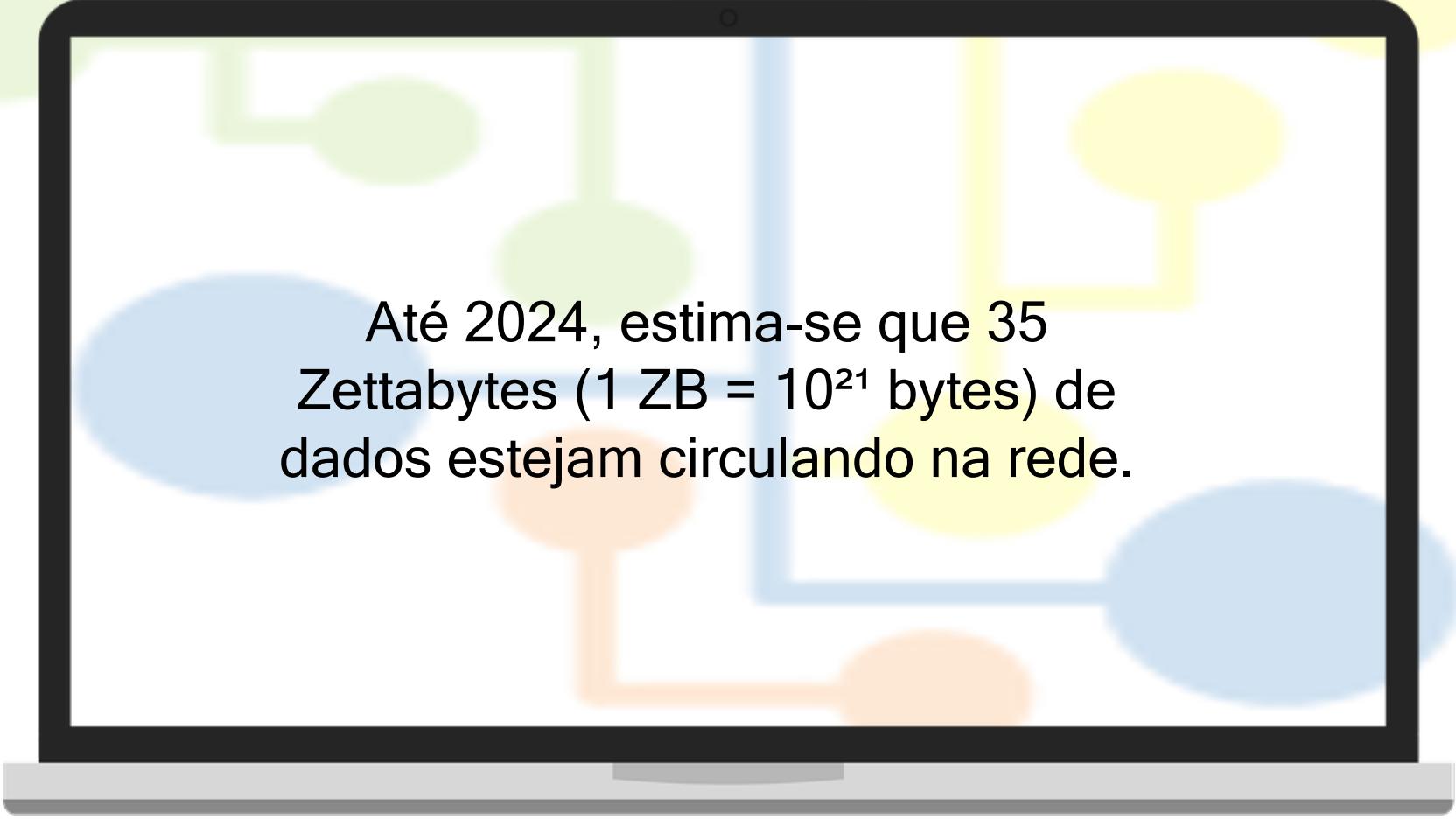
Growth of Data vs. Growth of Data Analysts

Stored Data accumulating at 28% annual growth rate
Data Analysts in workforce growing at 5.7% growth rate



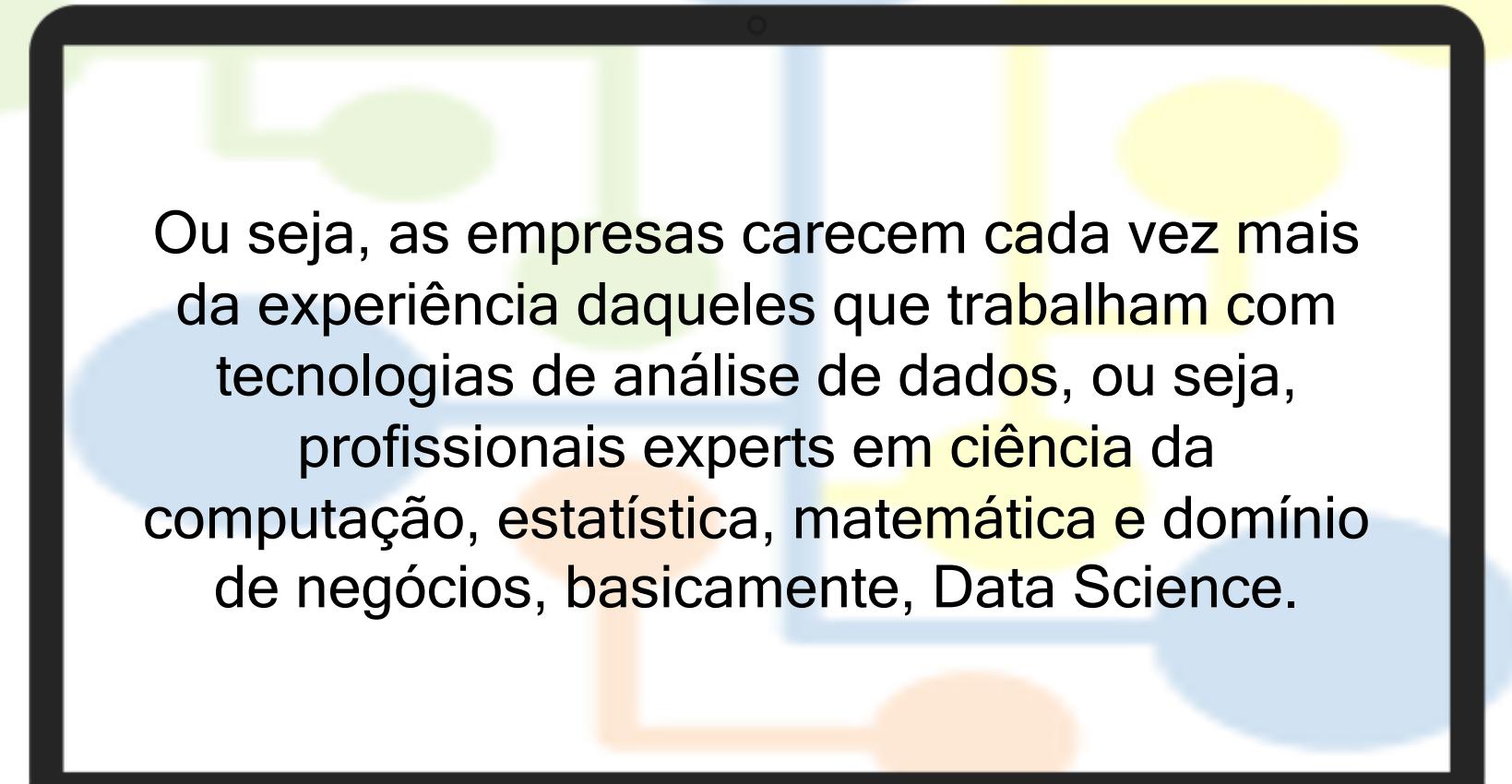
<http://www.delphianalytics.net/more-data-than-analysts-the-real-big-data-problem/>

O Mercado de Ciência de Dados e Big Data



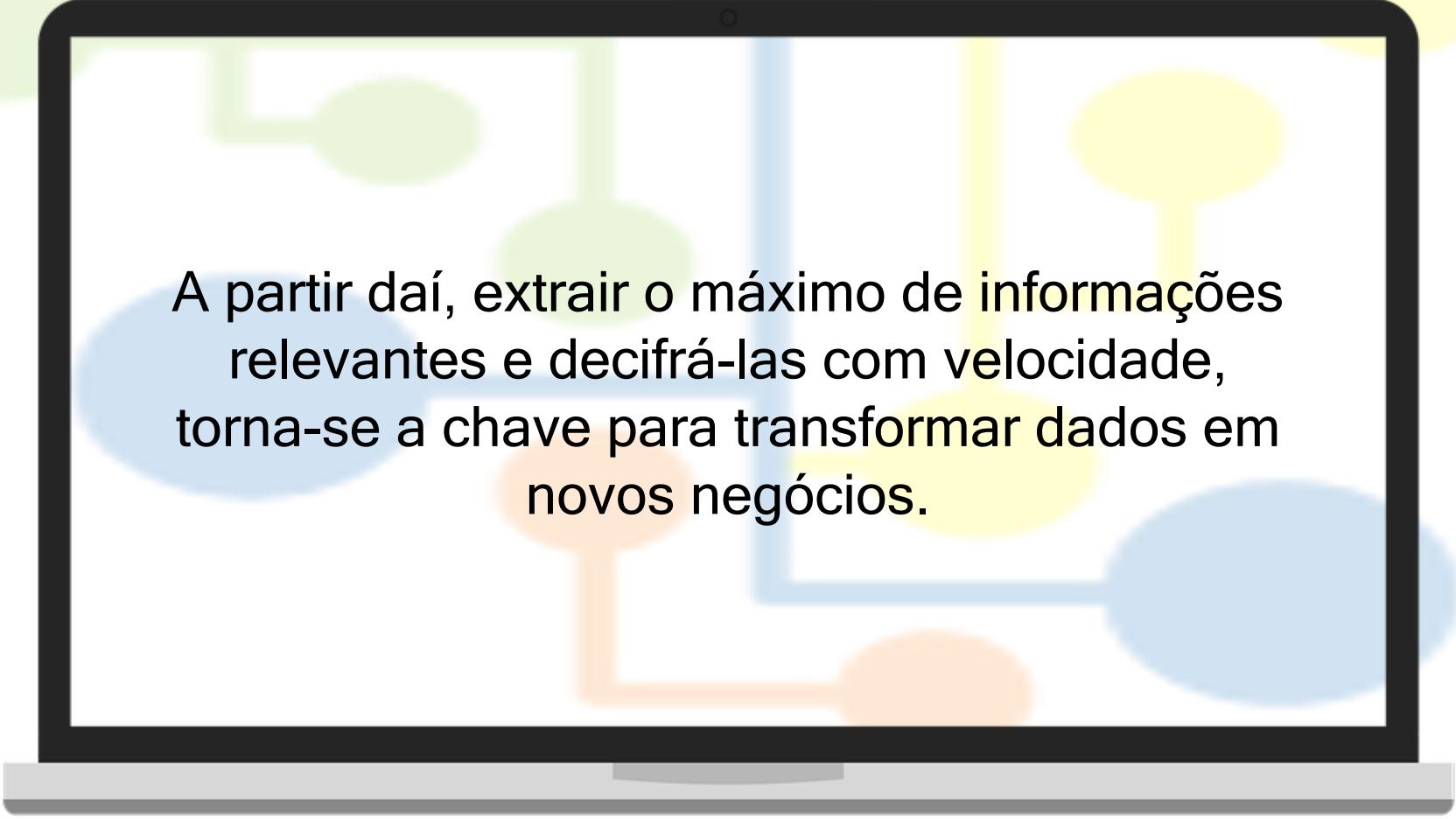
Até 2024, estima-se que 35 Zettabytes ($1 \text{ ZB} = 10^{21}$ bytes) de dados estejam circulando na rede.

O Mercado de Ciência de Dados e Big Data



Ou seja, as empresas carecem cada vez mais da experiência daqueles que trabalham com tecnologias de análise de dados, ou seja, profissionais experts em ciência da computação, estatística, matemática e domínio de negócios, basicamente, Data Science.

O Mercado de Ciência de Dados e Big Data



A partir daí, extrair o máximo de informações relevantes e decifrá-las com velocidade, torna-se a chave para transformar dados em novos negócios.

O Mercado de Ciência de Dados e Big Data

As carreiras ligadas à análise de dados estarão entre as carreiras mais quentes do mercado para os próximos anos.





Carreiras em Data Science

Carreiras em Data Science

Desde o início da DSA participamos de diversos projetos de consultoria em todo o Brasil e tivemos oportunidade de ver na prática como as carreiras em Data Science se desenvolveram ao longo do tempo.



Carreiras em Data Science

Por mais que os conceitos de análise de dados existam há décadas, as tecnologias que surgiram recentemente, permitem fazer coisas que não eram possíveis antes, sem falar no fato que o volume de dados gerado pela humanidade nunca foi tão grande.



Carreiras em Data Science

Na prática, todas as profissões envolvem a coleta e análise de dados.



Carreiras em Data Science

Você vai encontrar muitas carreiras com denominações diferentes. Mais importante do que o título em si é o conhecimento necessário para analisar dados e entregar resultados.



Carreiras em Data Science

Nosso trabalho aqui não é convencer ninguém que uma carreira em Data Science pode render muitos frutos.

Isso já está claro por tudo que se observa no mercado.

Nosso trabalho aqui é ajudar aqueles que já estão convencidos e desejam iniciar ou desenvolver a carreira na área de dados.



Carreiras em Data Science

Vejamos agora as principais carreiras em Data Science!



Analista de Negócios



- Estabelecer os objetivos e o âmbito de sistemas de negócios e de TI.
- Identificar problemas organizacionais e conceber soluções orientadas a dados.
- Realizar análises estatísticas, pesquisas e testes.
- Recomendar mudanças nos processos ou ofertas de produtos para tornar os departamentos internos mais eficientes.
- Conceber novos sistemas ou alterar os existentes.
- Fazer recomendações específicas de TI e apoiar a sua implementação.
- Agir como um elo de ligação entre os gestores e equipes técnicas.
- Coletar requerimentos e necessidades das áreas de negócio.

Curso recomendado na DSA: **Business Analytics**

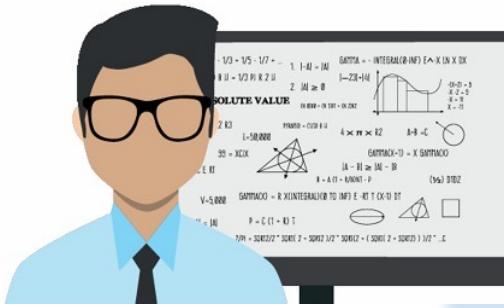
Analista de Dados



- Trabalhar com as equipes de TI, Gestão e / ou Cientistas de Dados para determinar e realizar o processo de análise de dados.
- Coletar dados de fontes primárias e secundárias.
- Efetuar modelagem de dados para análise.
- Realizar limpeza nos dados e descartar informações irrelevantes.
- Analisar e interpretar os resultados utilizando ferramentas estatísticas e técnicas convencionais.
- Identificar tendências, correlações e padrões em conjuntos de dados complexos através de análise exploratória.
- Identificar novas oportunidades para melhoria de processos.
- Concepção, criação e manutenção de bancos de dados relacionais e NoSQL.
- Resolver problemas de código e questões relacionadas a dados.
- Dominar linguagens como R, Python ou SQL e saber construir visualizações eficientes.

Cursos recomendados na DSA: **Formação Analista de Dados**

Cientista de Dados



- Empregar os programas de análise, aprendizado de máquina (Machine Learning) e métodos estatísticos para preparar os dados para uso em modelagem preditiva e prescritiva.
- Explorar e analisar dados de uma variedade de ângulos para determinar fraquezas escondidas, tendências e/ou oportunidades.
- Conceber soluções orientadas a dados para os principais desafios da empresa.
- Criar novos algoritmos para resolver problemas e usar ferramentas para automatizar o trabalho.
- Atuar com o Engenheiro de Dados para extrair grandes volumes de dados de múltiplas fontes internas e externas.
- Comunicar previsões e resultados para a gestão e áreas de negócio através de visualizações de dados eficazes.

Cursos recomendados na DSA: **Formação Cientista de Dados, Formação Análise Estatística Para Cientistas de Dados, Formação Inteligência Artificial e Formação Inteligência Artificial Para Medicina.**

Engenheiro de Dados



- Projetar, construir, instalar, testar e manter sistemas de gerenciamento de dados altamente escaláveis, localmente e/ou na nuvem.
- Pesquisar a aquisição de dados e novos usos para os dados existentes.
- Desenvolver processos de modelagem de dados e mineração.
- Integrar novas tecnologias de gerenciamento de dados com Data Warehouses, Data Lakes e Data Stores.
- Implementar arquiteturas de microsserviços para uso dos dados.
- Criar pipelines de dados e processos ETL (Extração, Transformação e Carga).
- Implementar procedimentos de recuperação de desastres e alta disponibilidade de dados.
- Implementar políticas de segurança no acesso aos dados e compliance com GDPR e LGPD.

Cursos recomendados na DSA: **Formação Engenheiro de Dados, Formação Suporte e Infraestrutura de Big Data e Design e Administração de Cloud Computing AWS**

Engenheiro de Machine Learning



- Projetar sistemas de Machine Learning (ML).
- Efetuar o deploy de modelos de ML através de boas práticas de Engenharia de Software.
- Desenvolver aplicativos de ML para o deploy dos modelos.
- Criar pipelines de ML para o uso dos modelos em produção.
- Executar testes e experimentos de ML.
- Realizar análise estatística e ajuste fino (otimização) dos modelos criados por Cientistas de Dados.
- Treinar e retreinar os modelos de ML conforme necessário.
- Aplicar bibliotecas e frameworks de ML.
- Manter-se atualizado sobre as novas tecnologias.

Cursos recomendados na DSA: **Formação Engenheiro de Machine Learning**

Arquiteto de Dados



- Colaborar com as equipes de TI e Gestão para elaborar uma estratégia de dados que atenda os requisitos da empresa.
- Criar um inventário de dados necessários para implementar a arquitetura.
- Pesquisar novas oportunidades de aquisição de dados.
- Projetar jobs ETL e pipelines de dados.
- Criar um fluxo de dados (pipeline) dentro da empresa.
- Desenvolver modelos de dados.
- Projetar, documentar, construir e implantar arquiteturas e aplicações de banco de dados (por exemplo, grandes bancos de dados relacionais e NoSQL).
- Projetar funcionalidades técnicas (por exemplo, escalabilidade, segurança, desempenho, recuperação de dados, confiabilidade, etc.)
- Implementar medidas para assegurar a precisão dos dados e acessibilidade.
- Criar procedimentos de gestão de metadados e compliance com GDPR e LGPD.

Cursos recomendados na DSA: **Formação Arquiteto de Dados**

Engenheiro DataOps



- Estabelecer plataformas, processos e estruturas de DataOps para melhorar a qualidade dos dados e o ciclo de vida dos dados.
- Implementar integração contínua e entrega de pipelines de dados.
- Criar e implementar estruturas de teste automatizadas para testar pipelines de dados.
- Projetar e implementar frameworks de verificação de qualidade de dados para garantia de qualidade de dados.
- Projetar, implementar e aprimorar sistemas de alerta e monitoramento para a pilha geral de aplicativos de dados.
- Proteção de dados confidenciais da empresa por meio de políticas de segurança apropriadas.
- Projetar e implementar soluções para oferecer suporte a Data Warehouses, Data Lakes e Data Stores.

Cursos recomendados na DSA: **Formação Engenheiro DataOps (checar disponibilidade)** e **Design e Administração de Cloud Computing AWS**

Administrador de Banco de Dados



- Suporte técnico aos bancos de dados existentes.
- Personalização de bancos de dados para necessidades específicas.
- Planejamento e projeto de bancos de dados para necessidades específicas.
- Solução de problemas para atender às necessidades da área de Ciência de Dados.
- Otimização de consultas SQL.
- Supervisão da instalação de novos SGBDs.
- Criar procedimentos de backup, restauração e recuperação de desastres.
- Atuar com bancos de dados relacionais e não relacionais.

Cursos recomendados na DSA: **Design e Implementação de Data Warehouses, Design e Administração de Cloud Computing AWS e SQL Para Data Science**

Estatístico



- Aplicar teorias e métodos estatísticos para resolver problemas práticos de negócios, engenharia, ciências ou outras áreas de conhecimento.
- Analisar e interpretar dados usando ferramentas, algoritmos, modelos estatísticos e software (por exemplo R, SAS, SPSS, Python, etc.)
- Decidir quais dados são necessários para responder a perguntas ou problemas específicos.
- Aplicar testes de hipóteses para validar ou rejeitar teorias e problemas de negócio.
- Realizar modelagem estatística.
- Analisar correlações entre os dados e aplicar testes de causalidade.
- Comunicar as descobertas às partes interessadas e gestores.
- Assessoria organizacional à estratégia de negócios.
- Auxiliar na tomada de decisão.

Cursos recomendados na DSA: **Formação Cientista de Dados e Formação Análise Estatística Para Cientistas de Dados**

Analista de Inteligência de Mercado



O Analista de Inteligência de Mercado é normalmente um Analista de Negócio com foco em análise macroeconômica e análise de competidores.

Outros títulos possíveis:

Analista de Inteligência Estatística
Analista de Dados Econômicos
Analista de Dados de Econometria
Analista de Big Data

Cursos recomendados na DSA: **Formação Analista de Inteligência de Mercado**

E o Empreendedor?



A Ciência de Dados fornece um oceano de oportunidades de empreendedorismo, desde consultoria para montar uma equipe de dados, passando pela construção de Data Lakes, definição de pipelines de dados, deploy de modelos, até entrega de aplicações analíticas completas.

Cursos recomendados na DSA: **Formação Desenvolvedor Web Para Data Science** e **Curso Gratuito de Empreendedorismo Para Data Science (disponível para quem adquire pelo menos uma Formação da DSA)**

Funções Gerenciais em Data Science

Apesar de muitas denominações possíveis, 5 posições chave em liderança de projetos de Data Science são facilmente identificáveis:

- Líder de Equipe de Data Science
- Líder de Projetos de Data Science
- Gerente de Data Science
- Gerente de Dados e Analytics
- Chief Data Officer (CDO)



Soft Skills Mais Importantes Para Carreiras em Data Science

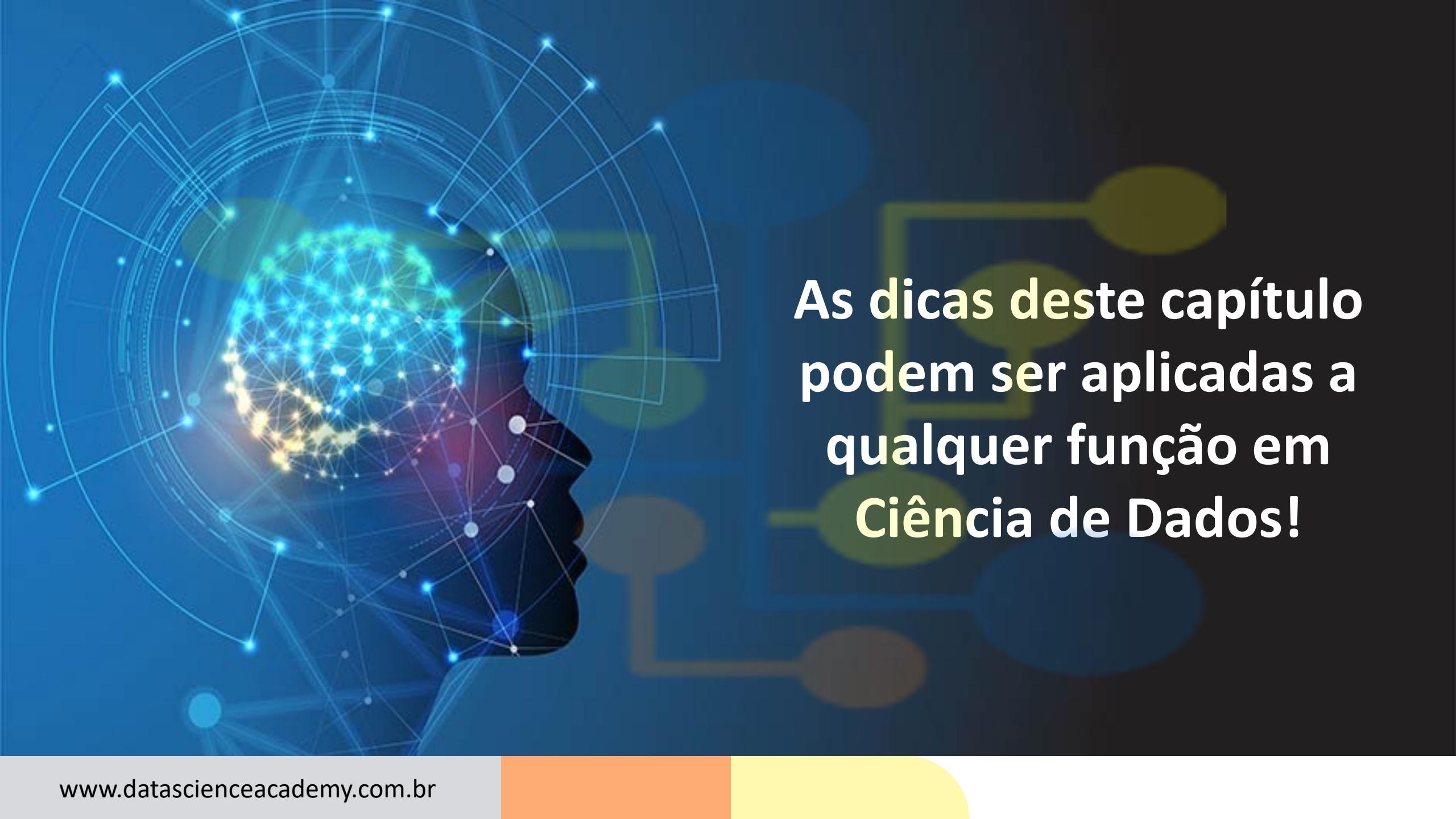


Soft Skills Mais Importantes Para Carreiras em Data Science

Cursos recomendados na DSA: **Curso Gratuito de Soft Skills**
(disponível para quem adquire pelo menos uma Formação da DSA)



Como Se Tornar um Cientista de Dados?



**As dicas deste capítulo
podem ser aplicadas a
qualquer função em
Ciência de Dados!**

Qual a Função Ideal em Ciência de Dados Para o Meu Perfil?

Principais Funções em Data Science

Análise de dados, análise exploratória, visualização e relatórios.

Analista de Dados

Modelagem estatística e modelagem preditiva com Machine Learning e Inteligência Artificial.

Cientista de Dados

Criação e implementação de pipelines de extração e armazenamento de dados.

Engenheiro de Dados

Deploy e implementação de modelos preditivos em sistemas de produção.

Engenheiro de Machine Learning

Design e planejamento de soluções e arquiteturas de dados.

Arquiteto de Dados

Criação, implementação e manutenção da operação de dados.

Engenheiro DataOps

Comece com a Auto-Avaliação

- Programação
- Pensamento Lógico
- Habilidade com Números
- Conhecimento em Banco de Dados



Prepare Seu Computador

- **Hardware**
- **Sistema Operacional**
- **Softwares**



Estatística e Matemática

- Estatística Descritiva (usada principalmente por Analistas de Dados)
- Probabilidade e Estatística Inferencial (usada principalmente por Cientistas de Dados)
- Álgebra Linear e Cálculo (conceitos aplicados em modelagem preditiva e diversos tipos de análises)

A dense collage of mathematical formulas and diagrams, including:

- Algebraic identities: $(a+b)^2 = a^2 + 2ab + b^2$, $c = 2\pi r$, $\sum x_i^2 = 10$.
- Calculus: $\frac{dy}{dx} = \frac{\partial y}{\partial x}$, $\int f(x) dx$, $t = \sqrt{\frac{2}{g}}$.
- Trigonometry: $\sin^2 x \cos^2 x = 1$, $\sin x = \frac{opposite}{hypotenuse}$.
- Geometry: A circle with radius r , area πr^2 , and circumference $2\pi r$.
- Statistics: $\bar{x} = \frac{1}{n} \sum x_i$.
- Probability: $P(A \cap B) = P(A)P(B|A)$.
- Linear Algebra: $A + B = B + A$, $AB = BA$.
- Calculus: $\frac{d}{dx} \int f(x) dx = f(x)$.
- Other: $\alpha A = (1+u^2+v^2)^{-1/2}$, $\alpha = \frac{1}{\sqrt{1+u^2+v^2}}$.

Big Data - Armazenamento e Processamento

- **Curso Gratuito Big Data Fundamentos aqui na DSA**
- **Frameworks de Armazenamento e Processamento de Dados (Hadoop, Spark, Databricks, Cloud Computing)**
- **Bancos de Dados Relacionais e NoSQL**
- **Data Warehouse, Data Lake e Data Store**



Linguagem de Programação e Machine Learning

- Lógica de Programação
- Python, R, Scala, Java, C++
- Machine Learning



Conhecimento de Negócios

- Qual o objetivo da sua análise?
- Trabalhe em projetos sempre como um problema a ser resolvido como objetivo
- Leia, converse, pesquise



Técnicas de Apresentação e Visualização de Dados

- Comunicação
- Cuidados com o idioma
- Técnicas de apresentação
- Saver resumir informação
- Criar gráficos que contem a história dos dados





**Existe o Mito do
Unicórnio?**

Existe o Mito do Unicórnio?

Existe o unicórnio? O profissional que domina tudo isso que discutimos nas aulas anteriores? Consigo realmente trabalhar com Ciência de Dados de forma efetiva?

Sim, existe e você consegue. Não é unicórnio. É profissional como Eu e Você que decidiu estudar, aprender e adquirir o conhecimento.

Muitas pessoas não serão capazes e tentarão convencer os outros que é impossível.

Mas o único limite é o que você impõe a si mesmo.



Existe o Mito do Unicórnio?

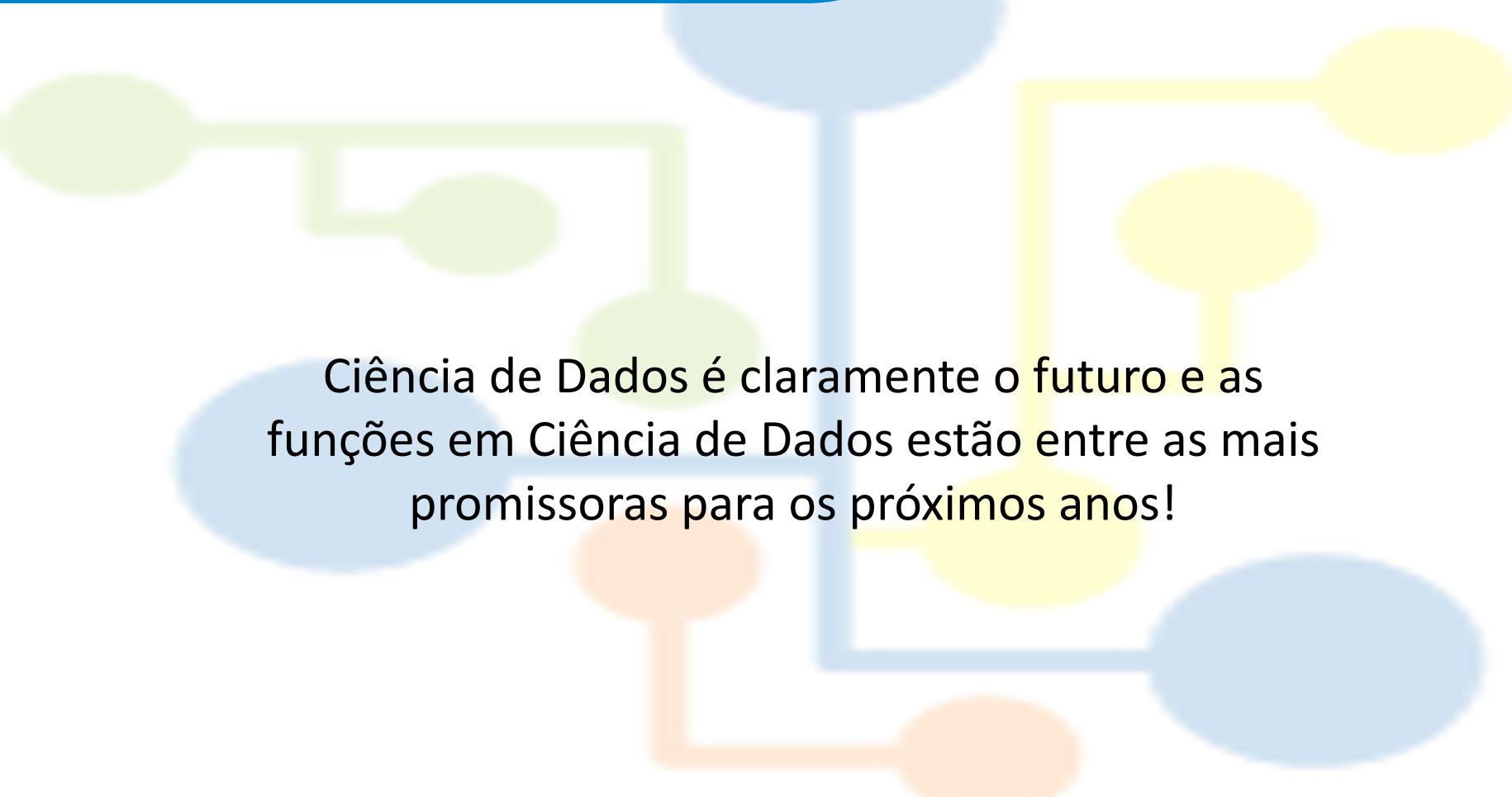
As funções em Ciência de Dados
estão entre aquelas com maior
empregabilidade em todo mundo.





Como se Preparar Para a Profissão do Futuro?

Como se Preparar Para a Profissão do Futuro?



Ciência de Dados é claramente o futuro e as funções em Ciência de Dados estão entre as mais promissoras para os próximos anos!

Como se Preparar Para a Profissão do Futuro?

- 1- Defina seus objetivos profissionais.
- 2- Faça uma auto-avaliação do seu conhecimento atual.
- 3- Certifique-se de ter um computador para suportar seu desenvolvimento profissional.
- 4- Trabalhe em projetos. Muitos projetos. E em diferentes áreas de negócio.
- 5- Desenvolva o raciocínio lógico e orientado a solução de problemas.
- 6- Comunique-se de forma efetiva.
- 7- Programação de computadores vai fazer parte do seu trabalho. Desenvolva essa habilidade.
- 8- Antes de focar em Machine Learning, foque em análise de dados.
- 9- Leia, converse e pesquise.
- 10- Não imponha limites ao seu aprendizado. Você é capaz de aprender o que quiser.





Encerramento



Muito Obrigado!

Tenha Uma Excelente Jornada de Aprendizado.

Data Science Academy