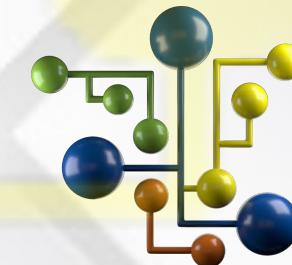


Microsoft

POWER BI

Para Data Science, versão 2.0



Data Science
Academy

Sobre Nós

Data Science Academy

A Data Science Academy (DSA) é um portal de ensino online especializado em Big Data, Data Science, Machine Learning, Inteligência Artificial, Arquitetura de Dados, Blockchain, RPA (Automação Robótica de Processos), Desenvolvimento de Chatbots 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.



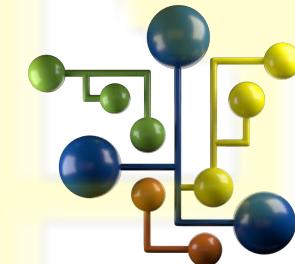
Conteúdo Programático (20 Módulos)

Parte 1
Análise de Dados com
Power BI

Parte 2
Gráficos, Mapas e
Dashboards

Parte 3
Estatística, R, Python e
Machine Learning

Parte 4
SQL Analytics,
Automação, Agentes
Virtuais, Power Apps



**Data Science
Academy**

Conteúdo Programático

Parte 1
Análise de Dados com
Power BI

01

Introdução

02

Primeiros Passos com Power BI Desktop

03

Business Intelligence no Power BI

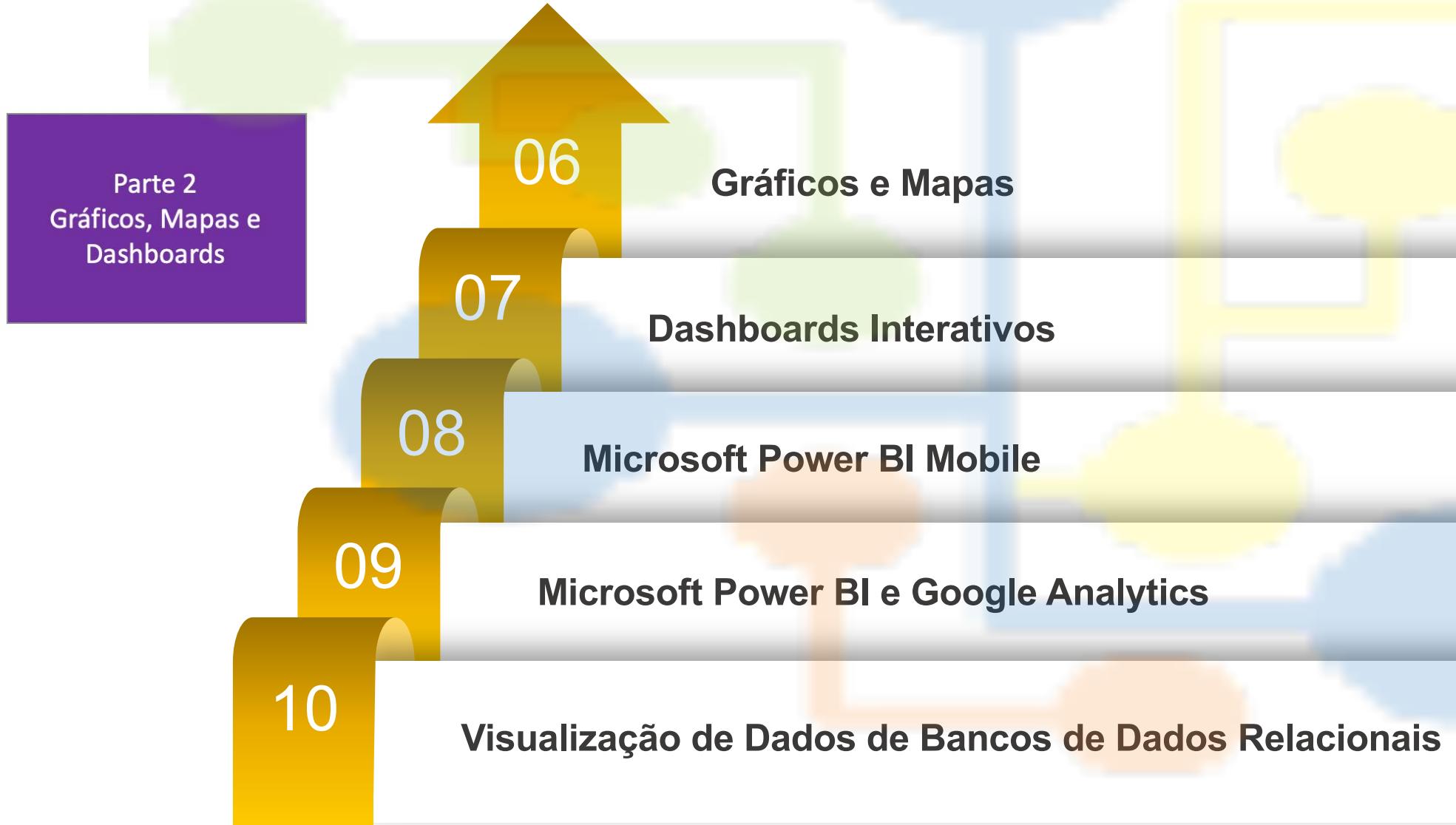
04

Modelagem e Relacionamento

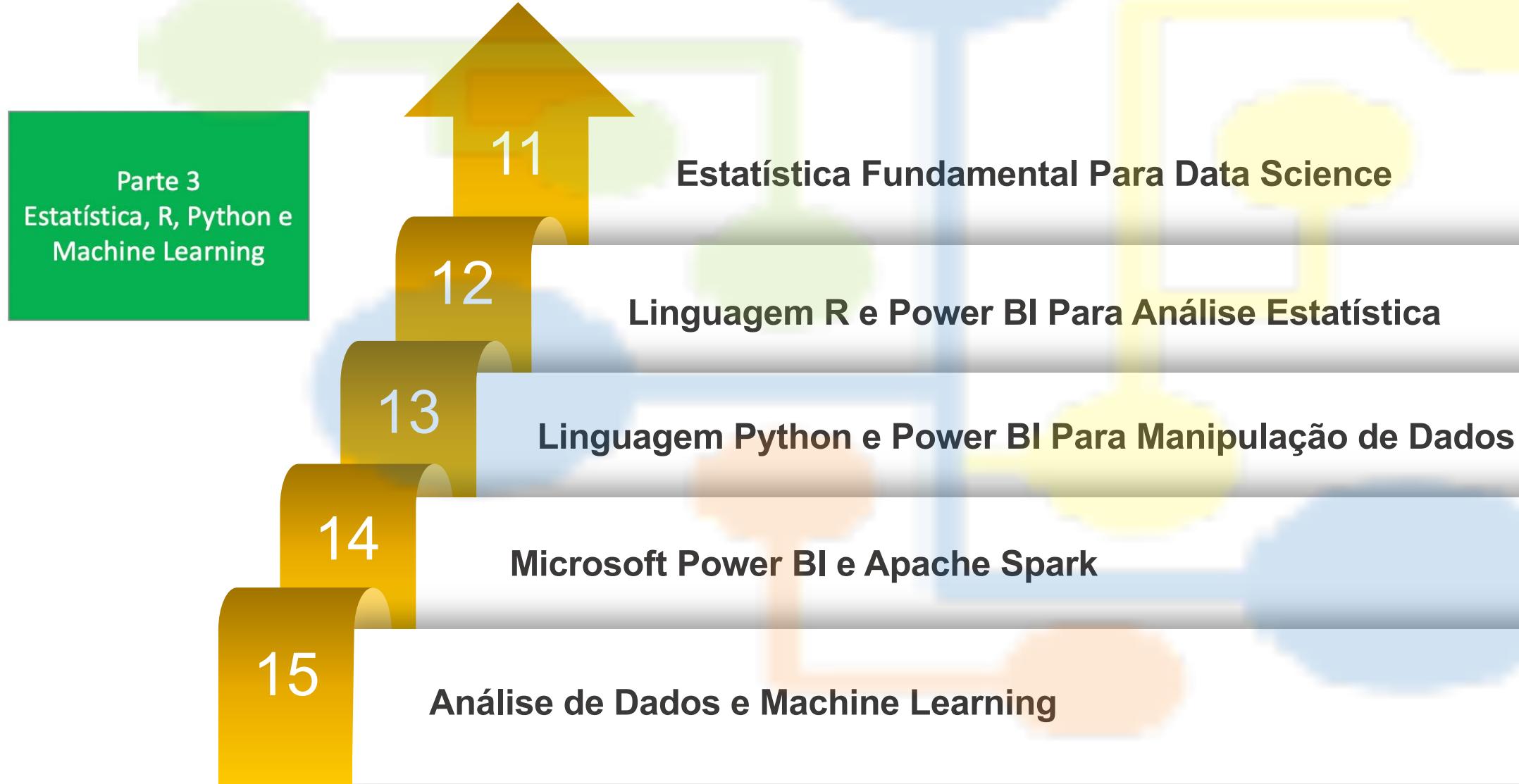
05

Séries Temporais, Agregação e Filtros

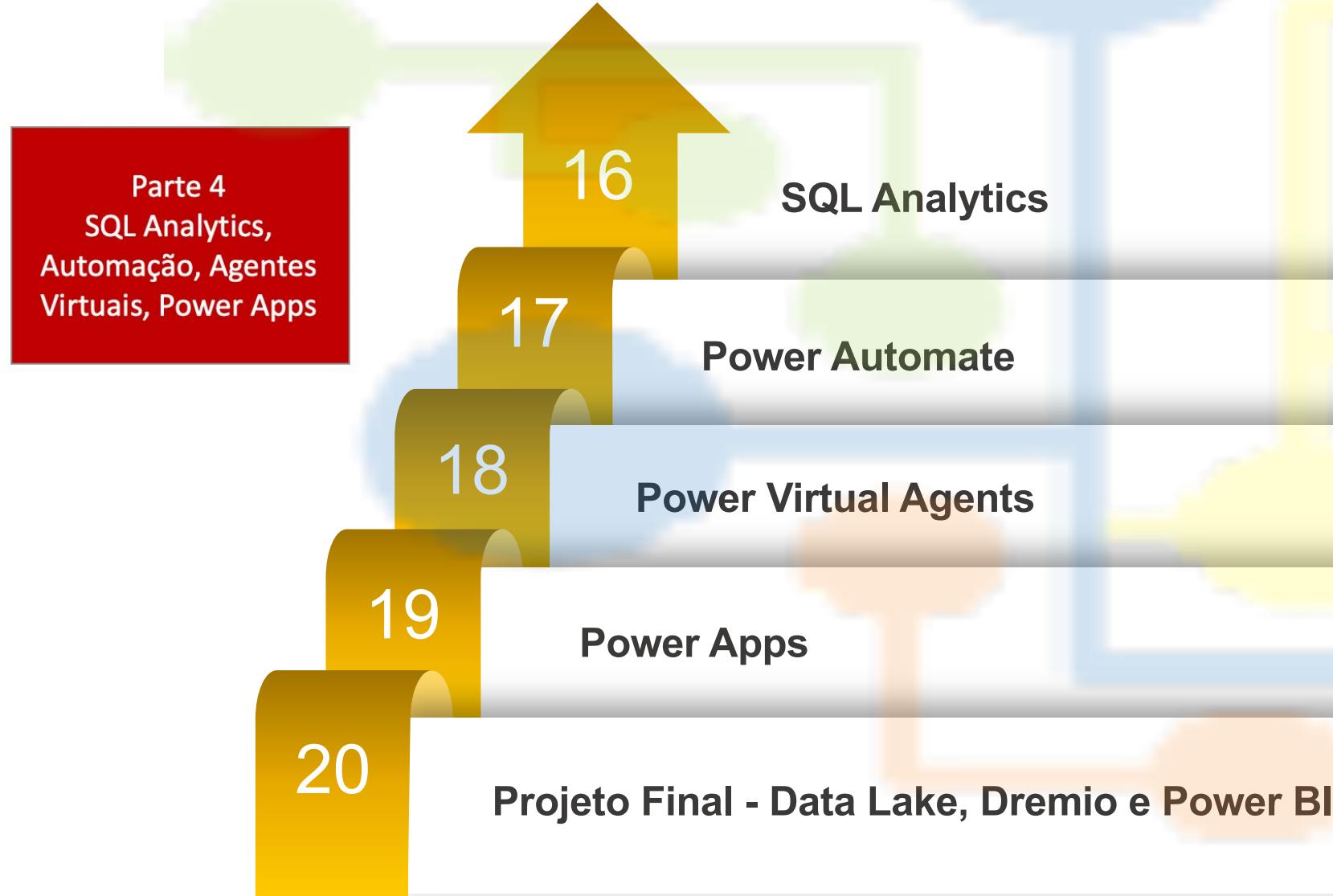
Conteúdo Programático



Conteúdo Programático



Conteúdo Programático



Tem certeza que esse curso é gratuito?



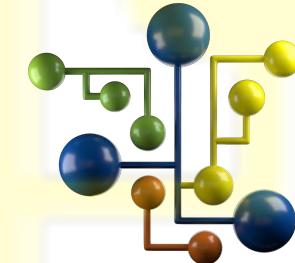
O Que Esperar Deste Curso?

Curso Denso,
com Muito
Conteúdo
(Padrão DSA)

Orientado às
Necessidades
do Mercado
de Trabalho

Objetivo e
Prático

Este é um
Curso de
Introdução



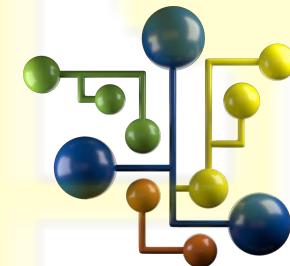
**Data Science
Academy**

Requerimentos de Hardware



Recomendação:

Processador Core i5 ou i7 ou Equivalente
4 GB de Memória RAM (no mínimo), sendo
8 GB de Memória RAM o ideal



Data Science
Academy

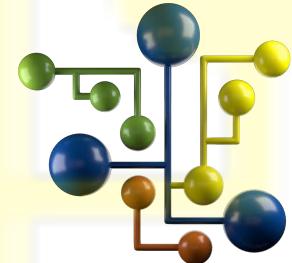
Requerimentos de Software



Power BI Desktop (Requer SO Windows)

Windows 10

Outros Softwares Serão Instalados Durante o Curso.
Certifique-se de Ter Privilégios de Administrador na
Sua Máquina.



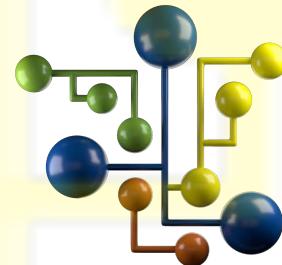
Data Science
Academy

Avaliação Final e Certificado de Conclusão



Para obter o certificado de conclusão, você deverá fazer a avaliação final e obter aproveitamento igual ou superior a 70%.

São 3 tentativas e se não for aprovado deverá aguardar 30 dias para solicitar uma nova tentativa!



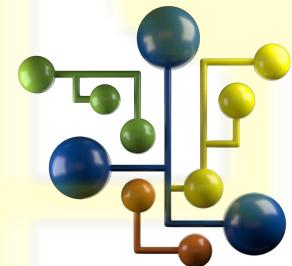
**Data Science
Academy**

Avaliação Final e Certificado de Conclusão



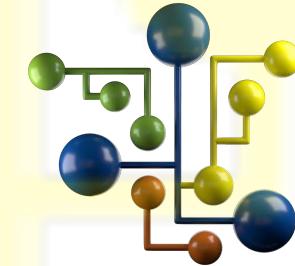
A avaliação final terá 50 questões e você deverá responder em até 120 minutos.

TODOS os itens de aprendizagem do curso deverão ser acessados para que avaliação final seja liberada!



**Data Science
Academy**

Dados x Informação x
Conhecimento x Inteligência



Data Science
Academy

Dados x Informação x Conhecimento x Inteligência

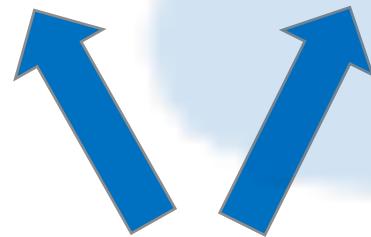
Funcionário	Idade	Salário Mensal	Cargo
Sócrates	32	R\$ 18.000,00	Cientista de Dados
Aristóteles	55	R\$ 17.500,00	Engenheiro de Dados



Dado

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário Mensal	Cargo
Sócrates	32	R\$ 18.000,00	Cientista de Dados
Aristóteles	55	R\$ 17.500,00	Engenheiro de Dados



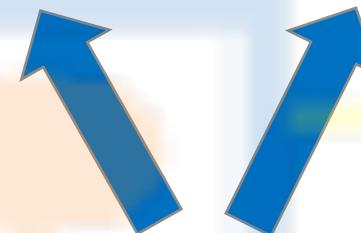
Informação

Dados x Informação x Conhecimento x Inteligência

As pessoas trabalham com informações mas a tecnologia armazena dados!

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário Mensal	Cargo
Sócrates	32	R\$ 18.000,00	Cientista de Dados
Aristóteles	55	R\$ 17.500,00	Engenheiro de Dados



Conhecimento

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário Mensal	Cargo
Sócrates	32	R\$ 18.000,00	Cientista de Dados
Aristóteles	55	R\$ 17.500,00	Engenheiro de Dados

Já o conceito de Inteligência (alguns chamam Sabedoria) está acima de conhecimento.

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário Mensal	Cargo
Sócrates	32	R\$ 18.000,00	Cientista de Dados
Aristóteles	55	R\$ 17.500,00	Engenheiro de Dados

Inteligência é a capacidade de resolver problemas, usando o conhecimento, através das informações disponíveis.

Dados x Informação x Conhecimento x Inteligência

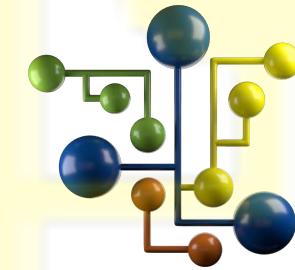
Funcionário	Idade	Salário Mensal	Cargo
Sócrates	32	R\$ 18.000,00	Cientista de Dados
Aristóteles	55	R\$ 17.500,00	Engenheiro de Dados

A diferença está na forma como cada um utiliza o conhecimento que tem e as conexões que faz em seu cérebro. Isto é inteligência, ou seja, saber resolver problemas utilizando o conhecimento que possui. E isto se aplica também a poder resolver problemas novos usando adaptações, analogias, etc.

Dados x Informação x Conhecimento x Inteligência

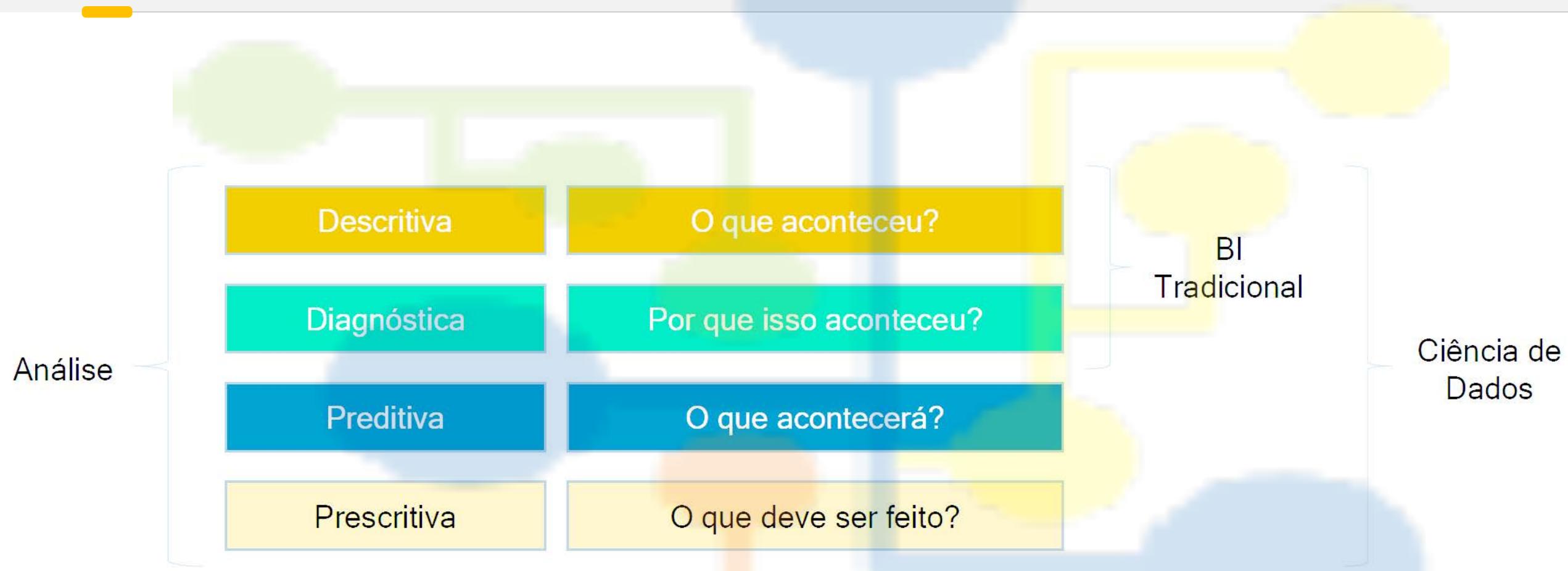
Nosso objetivo é começar com os dados, transformá-los em informações e conhecimento e permitir que tomadores de decisão usem sua inteligência para resolver problemas, a partir do conhecimento adquirido.

Business Intelligence
x
Data Science

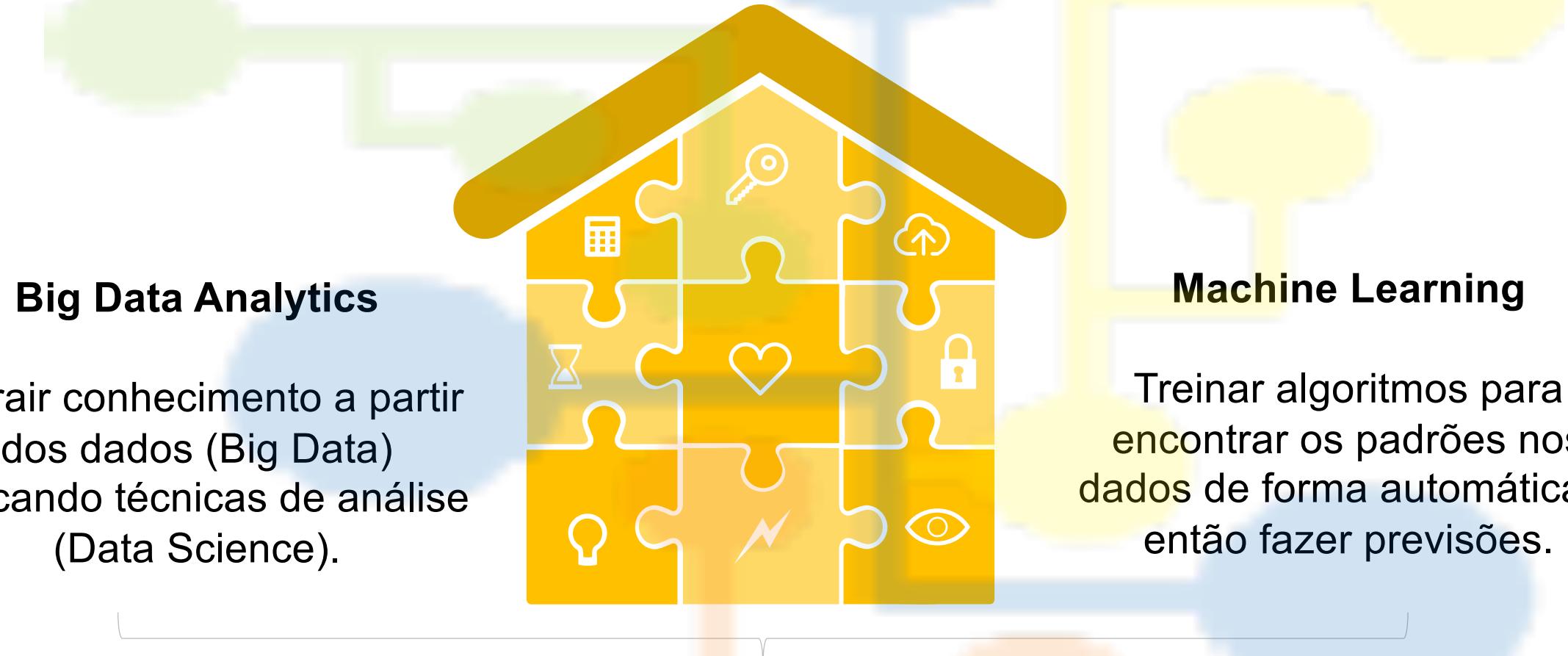


Data Science
Academy

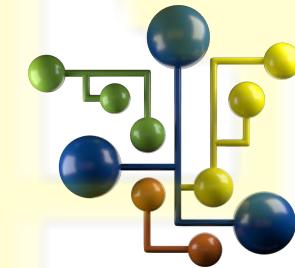
Business Intelligence x Data Science



Big Data Analytics x Machine Learning



Primeiros Passos com o Power BI Desktop



Data Science
Academy

Primeiros Passos com Power BI Desktop

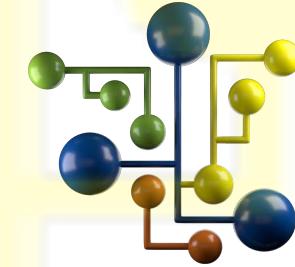
- Fontes de Dados do Power BI
- Carregar Arquivos CSV e Excel
- Power BI Desktop Workflow
- Explorar os Dados
- Construir Visualizações
- Formatar Legenda, Título e Labels

Primeiros Passos com Power BI Desktop

Estudo de Caso
Construindo um Dashboard de Vendas

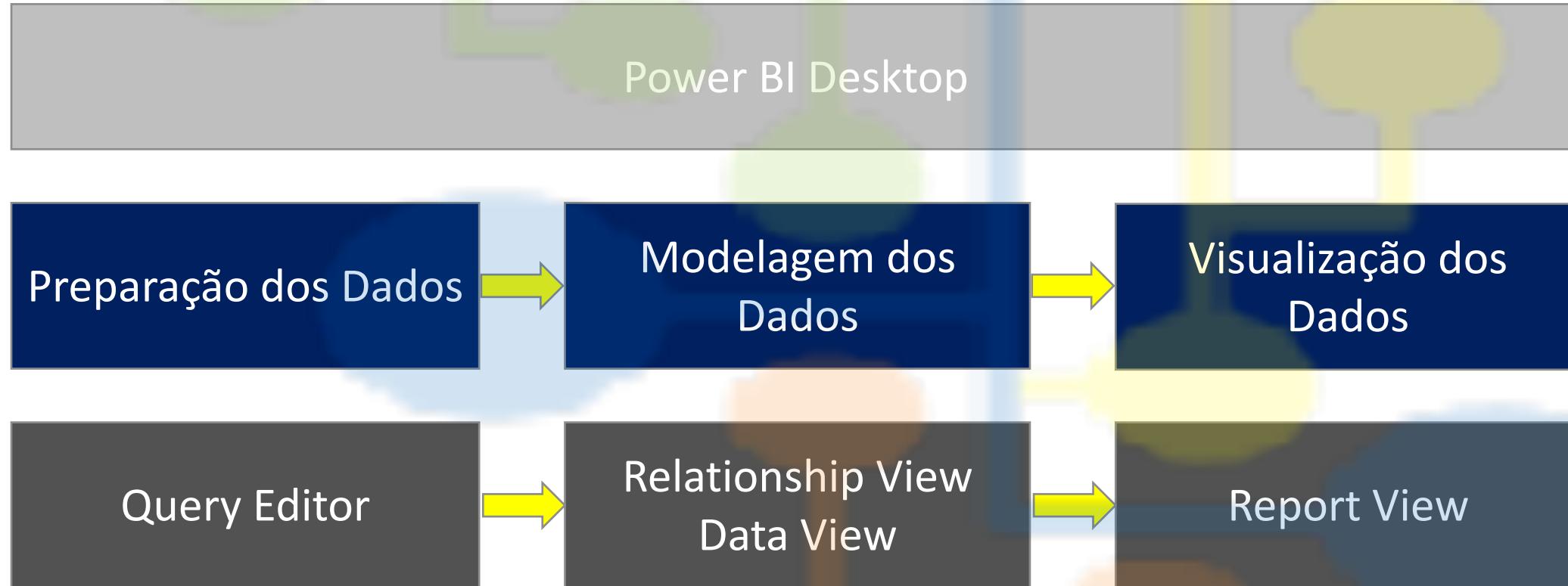


Power BI Desktop Workflow



Data Science
Academy

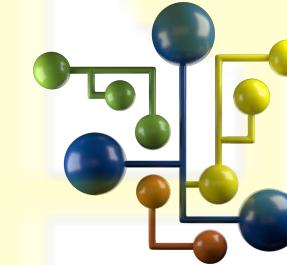
Power BI Desktop Workflow



Business Intelligence no Power BI



Você é Pago Para
Resolver Problemas



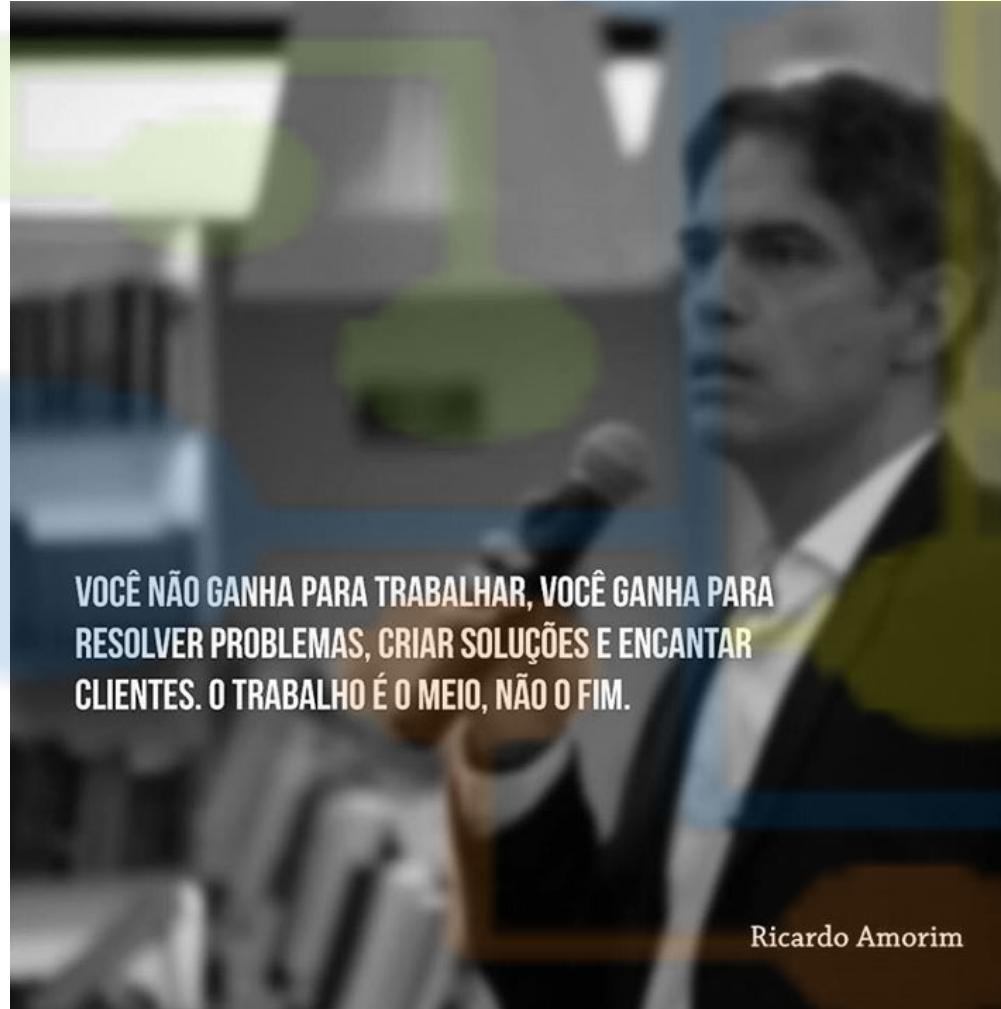
Data Science
Academy

Você é Pago Para Resolver Problemas

Para que você usa o Power BI?

- Para criar gráficos! 
- Para criar relatórios! 
- Para modelar os dados sem a necessidade de programação! 
- Para integrar com outros produtos Microsoft! 
- Para resolver problemas de negócio, entregar soluções e ajudar a empresa ou cliente a alcançar seus objetivos e estratégias. 

Você é Pago Para Resolver Problemas



Você é Pago Para Resolver Problemas



Você é Pago Para Resolver Problemas

1. Quando se deparar com mensagens de erro, analise!
2. Quando se deparar com mensagens de erro, pesquise!
3. Não sabe? Aprenda!
4. Desenvolva habilidades com foco na solução de problemas.
5. Observe as perguntas que uma pessoa faz! Elas dizem muito mais do que as respostas!



Você é Pago Para Resolver Problemas

O mundo tem 2 grupos de pessoas:

As que criam problemas.

As que resolvem problemas.

Tente se manter sempre no segundo grupo, onde a concorrência é muito menor!



Introdução a Business Intelligence no Power BI



Business
Intelligence

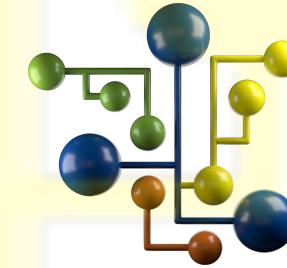
Introdução a Business Intelligence no Power BI

BI é um termo criado pelo Gartner Group definido como "um termo genérico que inclui aplicações, infraestrutura, ferramentas e melhores práticas que permite o acesso e análise de informações para melhorar e otimizar decisões e desempenho".

Introdução a Business Intelligence no Power BI

BI é, portanto, um conceito!

Modelo de Organização dos Dados



Data Science
Academy

Modelo de Organização dos Dados

Para o Estudo de Caso 2:

Posso levar os dados para o Power BI da forma que estão?

Quais são os requerimentos de organização dos dados?

Como mantendo os dados em um formato de atualização sob demanda?

Como visualizar os dados por diferentes visões e ângulos?

Modelo de Organização dos Dados

Precisamos carregar os dados de uma forma que facilite a extração de relatórios e a entrega dos resultados aos tomadores de decisão!

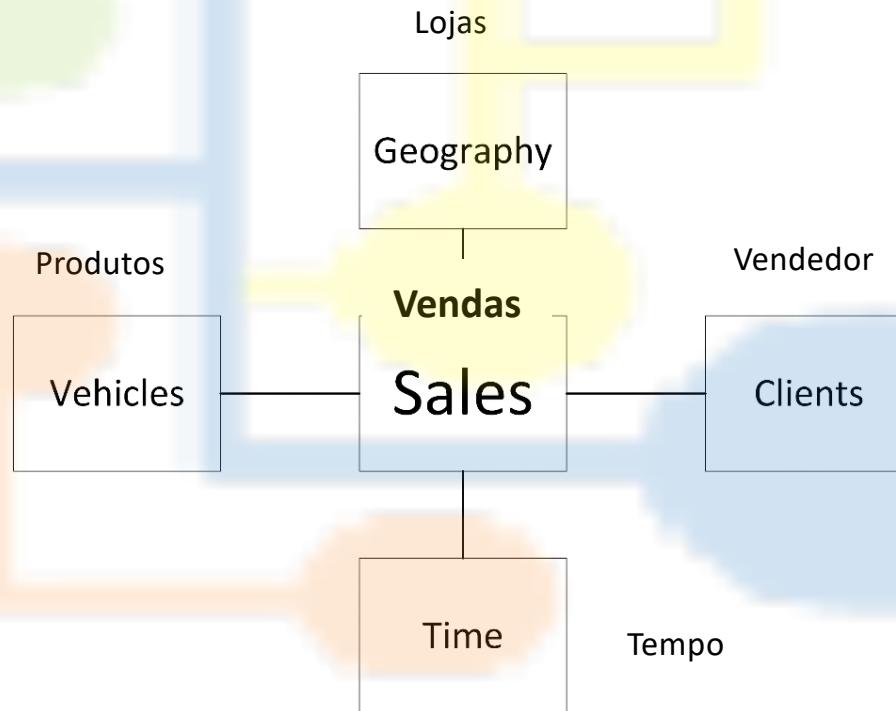
Precisamos de um modelo de organização dos dados!

Modelo de Organização dos Dados

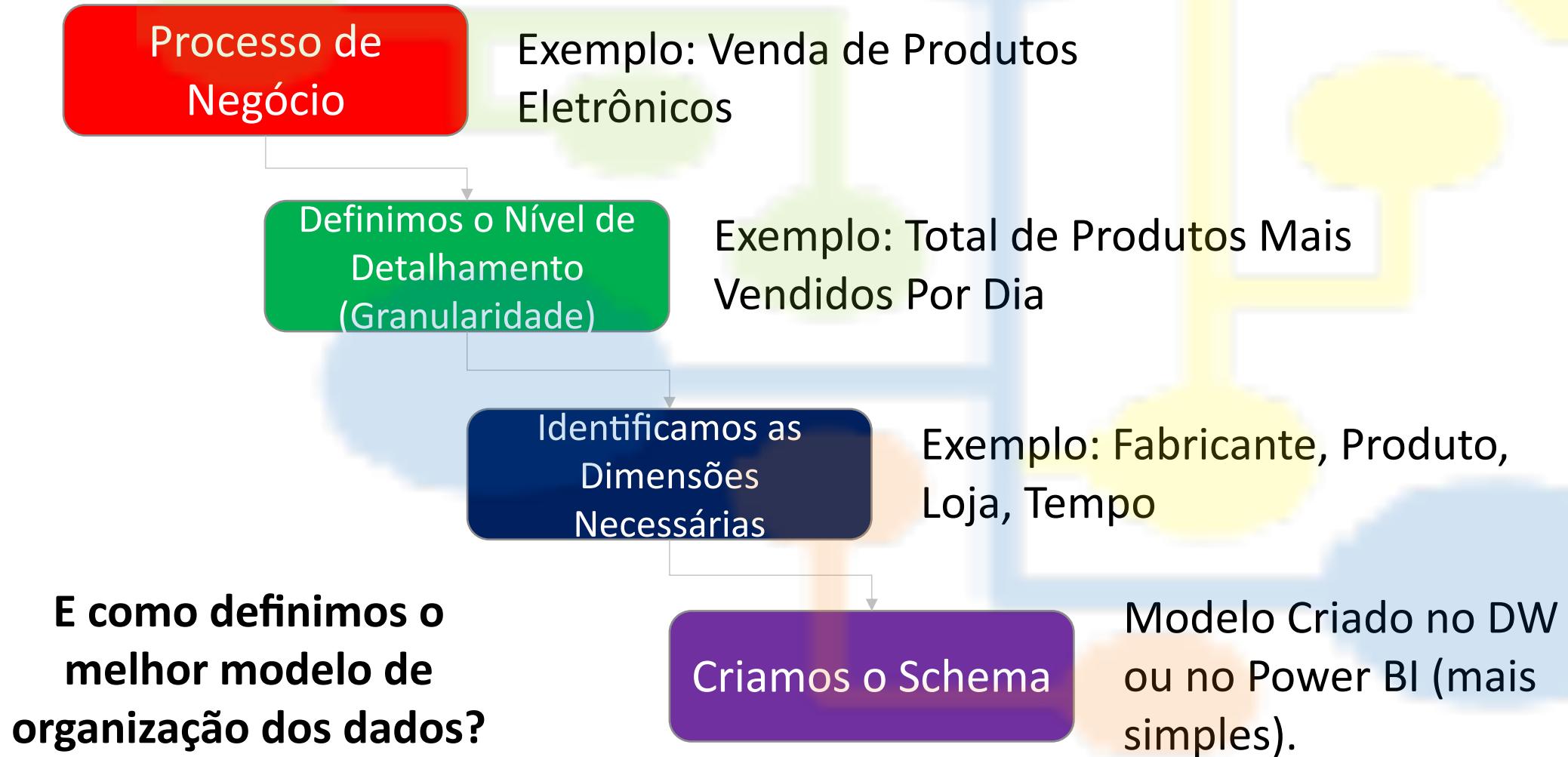
Os dados estão com uma representação tabular, o que pode ser útil para gráficos básicos!

A	B	C	D	E	F	G	H	I	J	K	L
ID-Produto	Produto	Categoria	Segmento	Fabricante	Loja	Cidade	Estado	Vendedor	ID-Vendedor	Data Venda	ValorVenda
2	SKU-0000001	LG K10 TV Power	Celulares	Corporativo	LG	SP8B21	São Paulo	Ana Teixeira	1009	04-10-2012	R\$ 679,00
3	SKU-0000002	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Josias Silva	1006	01-01-2012	R\$ 832,00
4	SKU-0000003	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Josias Silva	1006	02-02-2012	R\$ 790,00
5	SKU-0000004	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Mateus Gonçalves	1003	03-03-2012	R\$ 765,32
6	SKU-0000005	Lavadora 11 Kg	Eletrônicos	Doméstico	Electrolux	SP8B21	São Paulo	Artur Moreira	1004	04-04-2012	R\$ 459,89
7	SKU-0000006	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Rodrigo Fagundes	1005	04-05-2012	R\$ 409,99
8	SKU-0000007	Geladeira Duplex	Eletrônicos	Doméstico	Electrolux	SP8B21	São Paulo	Josias Silva	1006	04-06-2012	R\$ 1,000,91
9	SKU-0000008	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	A9990	Belo Horizonte	Minas Gerais	1007	04-07-2012	R\$ 1,229,00
10	SKU-0000009	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Mateus Gonçalves	1003	04-08-2012	R\$ 1,300,00
11	SKU-0000010	Geladeira Duplex	Eletrônicos	Doméstico	Electrolux	SP8B21	São Paulo	André Ferreira	1002	04-09-2012	R\$ 1,290,00
12	SKU-0000011	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Ana Teixeira	1009	04-09-2012	R\$ 1,287,00
13	SKU-0000012	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	A9990	Belo Horizonte	Minas Gerais	1005	04-11-2012	R\$ 1,651,00
14	SKU-0000013	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Josias Silva	1008	04-11-2012	R\$ 1,100,00
15	SKU-0000014	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Mateus Gonçalves	1003	04-12-2012	R\$ 1,190,00
16	SKU-0000015	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Artur Moreira	1004	04-12-2012	R\$ 1,190,98
17	SKU-0000016	Lavadora 11 Kg	Eletrônicos	Doméstico	Electrolux	A9990	Belo Horizonte	Minas Gerais	1005	02-01-2013	R\$ 877,00
18	SKU-0000017	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Josias Silva	1008	02-02-2013	R\$ 982,00
19	SKU-0000018	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Fernando Zambrini	1007	02-03-2013	R\$ 872,00
20	SKU-0000019	Lavadora 11 Kg	Eletrônicos	Doméstico	Electrolux	SP8B21	São Paulo	Mateus Gonçalves	1003	02-04-2013	R\$ 799,00
21	SKU-0000020	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	André Ferreira	1002	02-05-2013	R\$ 899,00
22	SKU-0000021	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Ana Teixeira	1009	02-06-2013	R\$ 799,00
23	SKU-0000022	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Josias Silva	1008	02-07-2013	R\$ 987,00
24	SKU-0000023	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Josias Silva	1005	02-08-2013	R\$ 699,90
25	SKU-0000024	Lavadora 11 Kg	Eletrônicos	Doméstico	Electrolux	SP8B21	São Paulo	Mateus Gonçalves	1003	02-09-2013	R\$ 789,23
26	SKU-0000025	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Artur Moreira	1008	02-10-2013	R\$ 789,34
27	SKU-0000026	Lavadora 11 Kg	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Rodrigo Fagundes	1005	02-11-2013	R\$ 764,20
28	SKU-0000027	Lavadora 11 Kg	Eletrônicos	Doméstico	Electrolux	SP8B21	São Paulo	Josias Silva	1008	02-12-2013	R\$ 1,245,90
29	SKU-0000028	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Fernando Zambrini	1007	02-01-2013	R\$ 1,345,87
30	SKU-0000029	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Mateus Gonçalves	1003	02-02-2013	R\$ 1,234,12
31	SKU-0000030	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	André Ferreira	1002	02-03-2013	R\$ 1,245,90
32	SKU-0000031	Geladeira Duplex	Eletrônicos	Doméstico	Brastemp	SP8B21	São Paulo	Maria Fernandes	1001	02-04-2013	R\$ 1,345,87

Mas o que precisamos é de um modelo que facilite a visualização dos dados por diferentes ângulos e sob demanda.



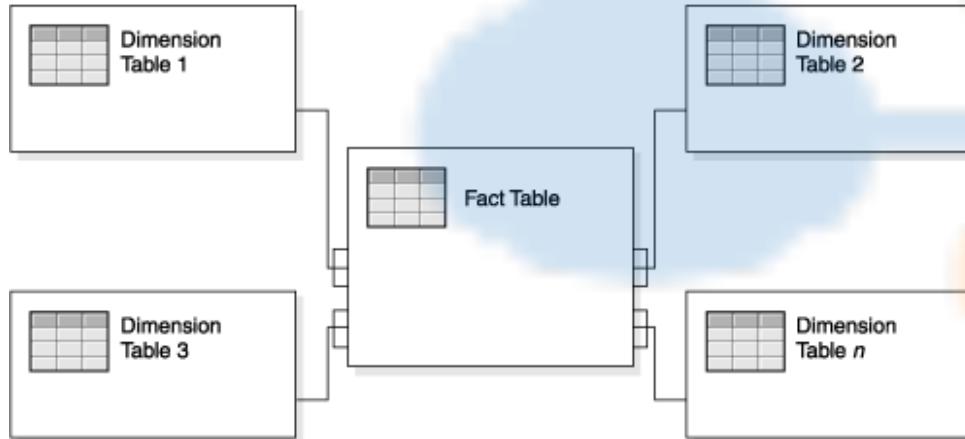
Modelo de Organização dos Dados



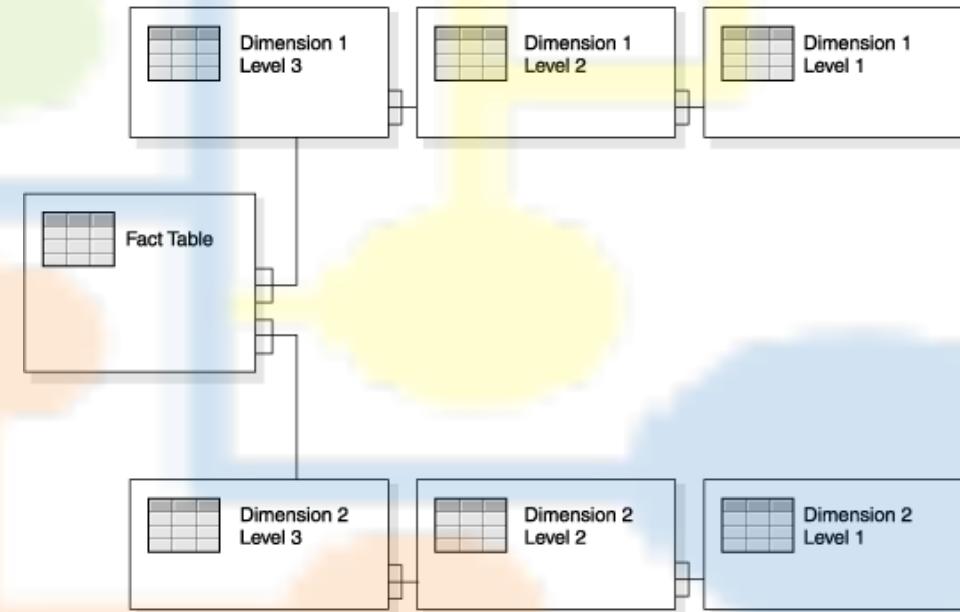
Modelo de Organização dos Dados

Modelos de Organização de Dados Mais Usados em BI

Star Schema



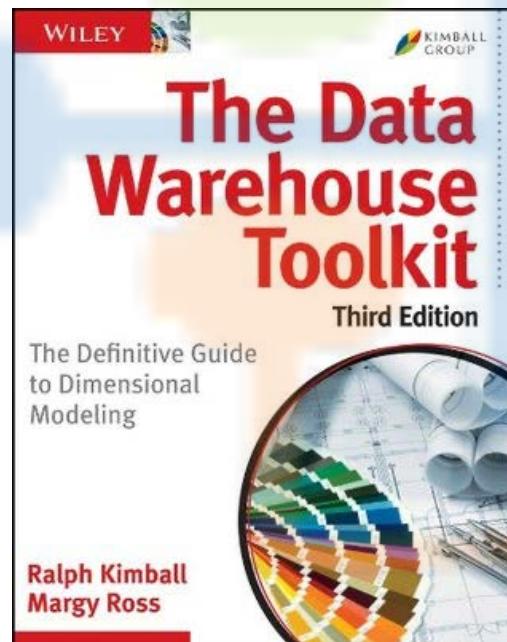
Snowflake Schema



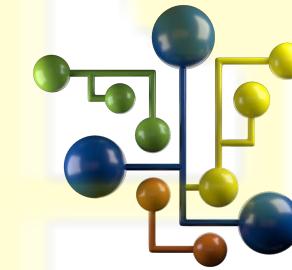
Modelo de Organização dos Dados

Se quiser estudar esse assunto em mais detalhes, ele é abordado no curso **Design e Implementação de Data Warehouses**, da Formação Engenheiro de Dados, aqui na DSA.

Também recomendamos o livro:

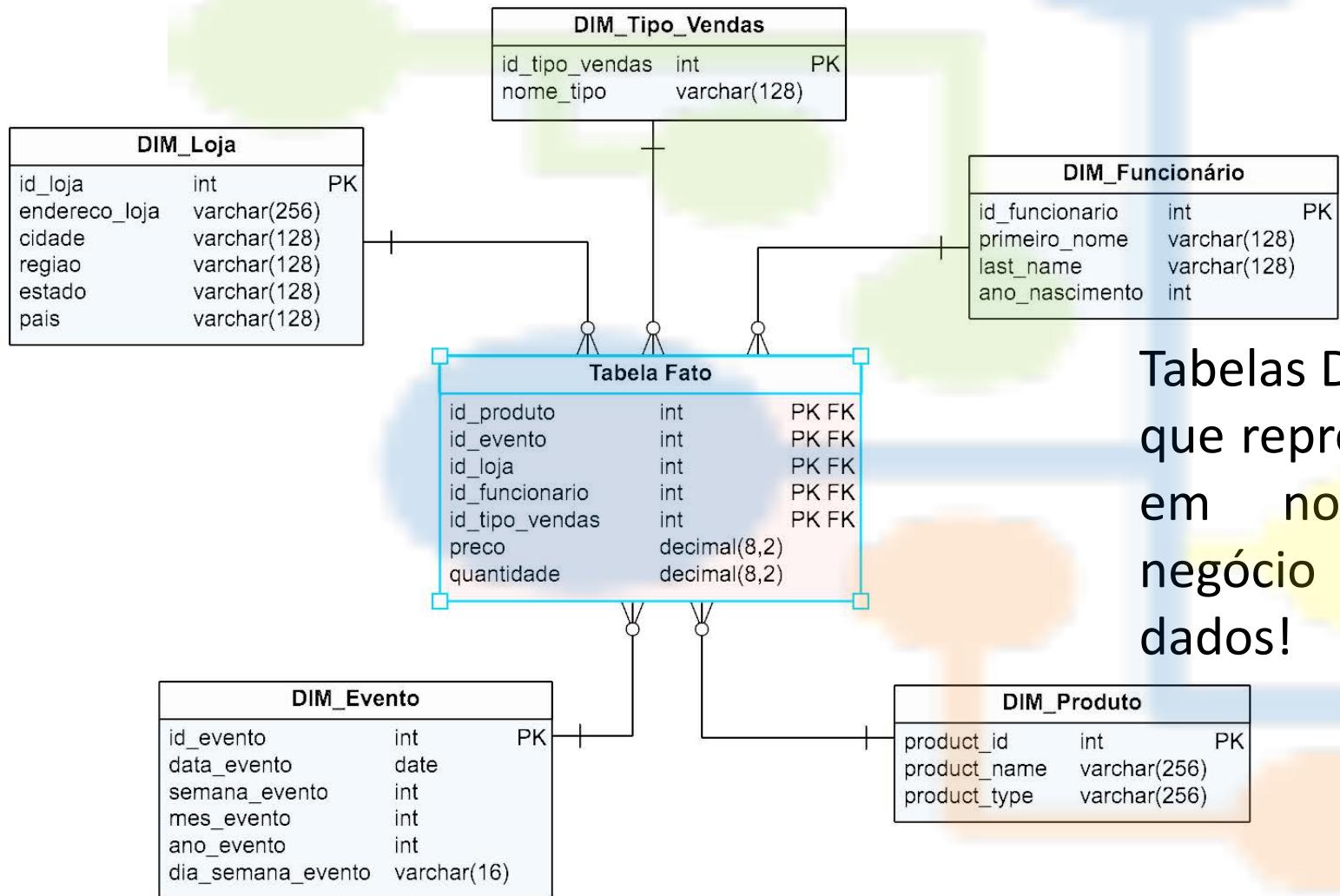


Modelo Star Schema



Data Science
Academy

Modelo Star Schema



Tabelas DIM são as **dimensões** que representam as entidades em nosso problema de negócio e que classificam os dados!

Modelo Star Schema

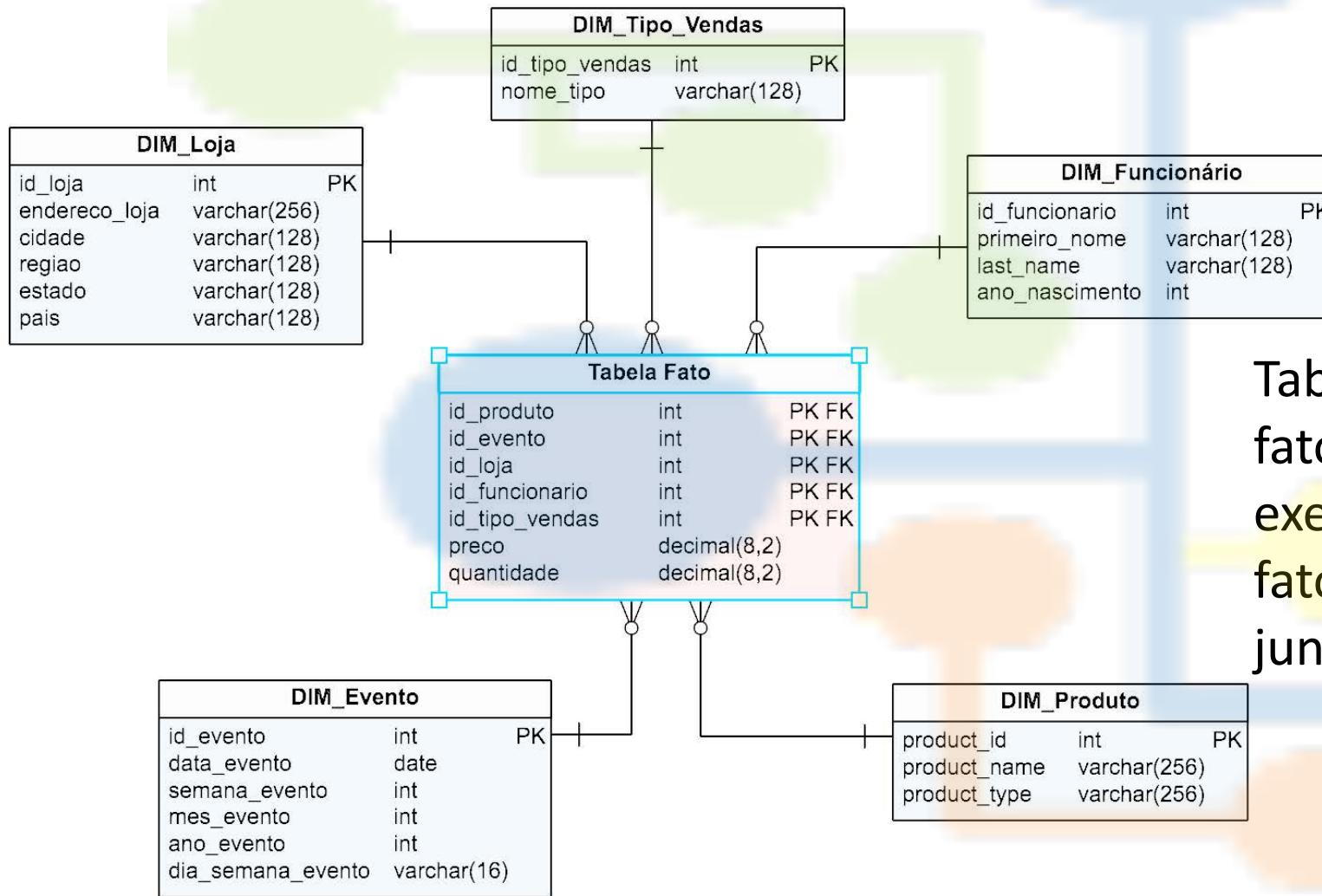
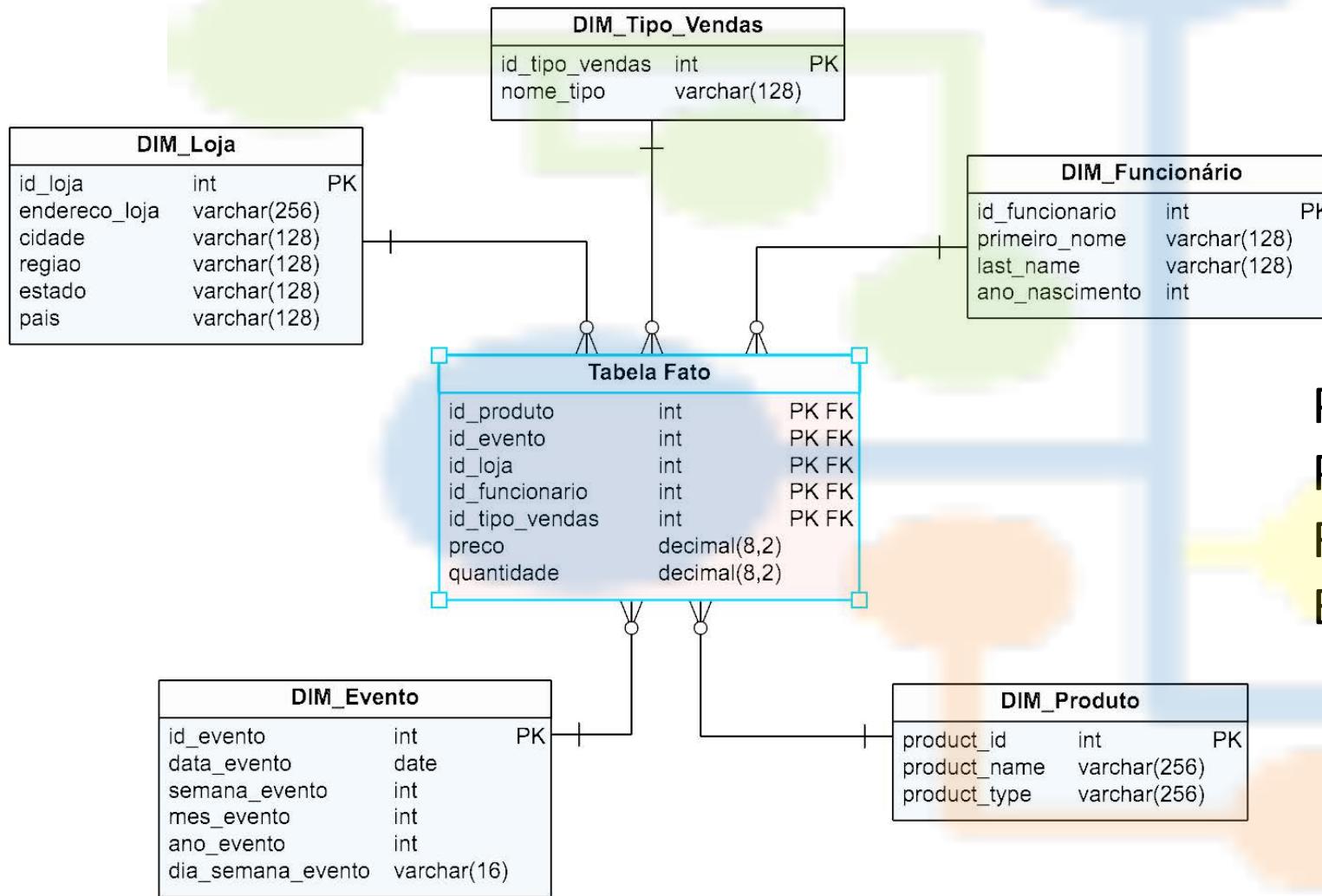


Tabela FATO representa um fato, um detalhe, por exemplo, uma venda. Cada fato acontece a partir da junção das dimensões.

Modelo Star Schema

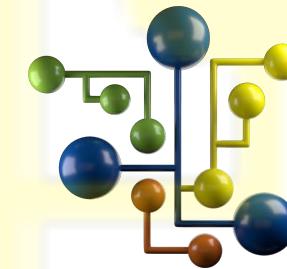


PK = Primary Key (Chave Primária)
FK = Foreign Key (Chave Estrangeira)

Modelagem, Relacionamento e DAX

Modelagem, Relacionamento
e Introdução ao DAX

As 4 Fases do Aprendizado



Data Science
Academy

As 4 Fases do Aprendizado

Como foi seu aprendizado até aqui no curso? Teve dúvidas, dificuldades?

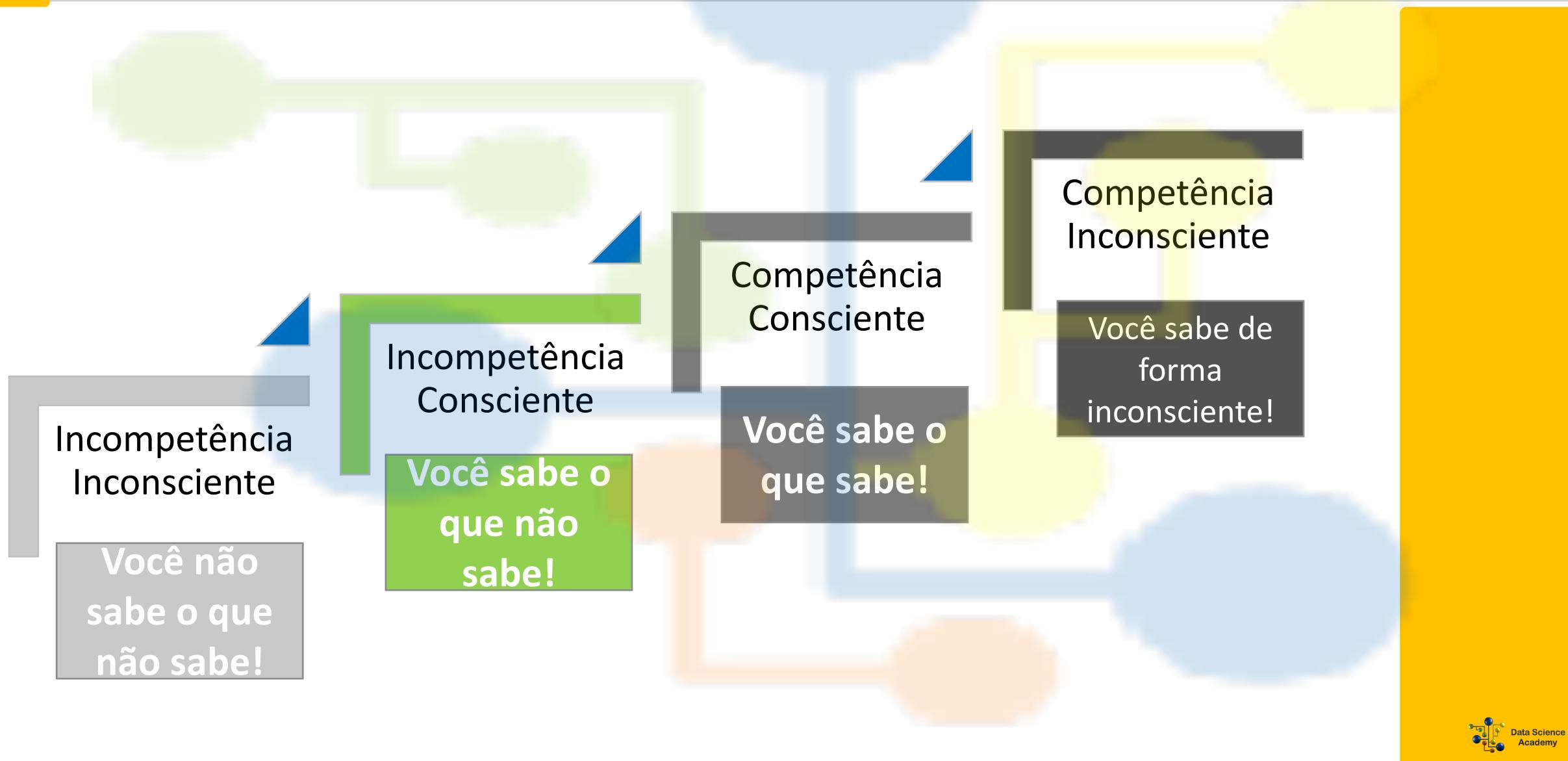
Ótimo, isso faz parte do processo de aprendizado.

Como foi seu aprendizado até aqui no curso? Não teve dúvidas ou dificuldades?

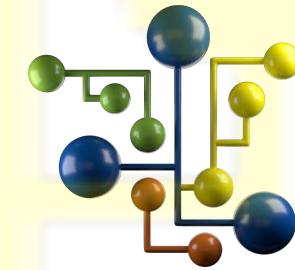
Ótimo, então o conceito já está claro para você.

Cada pessoa tem seu tempo de aprendizado, sua forma e velocidade!
Descubra o que funciona melhor para você! Experimente!

As 4 Fases do Aprendizado

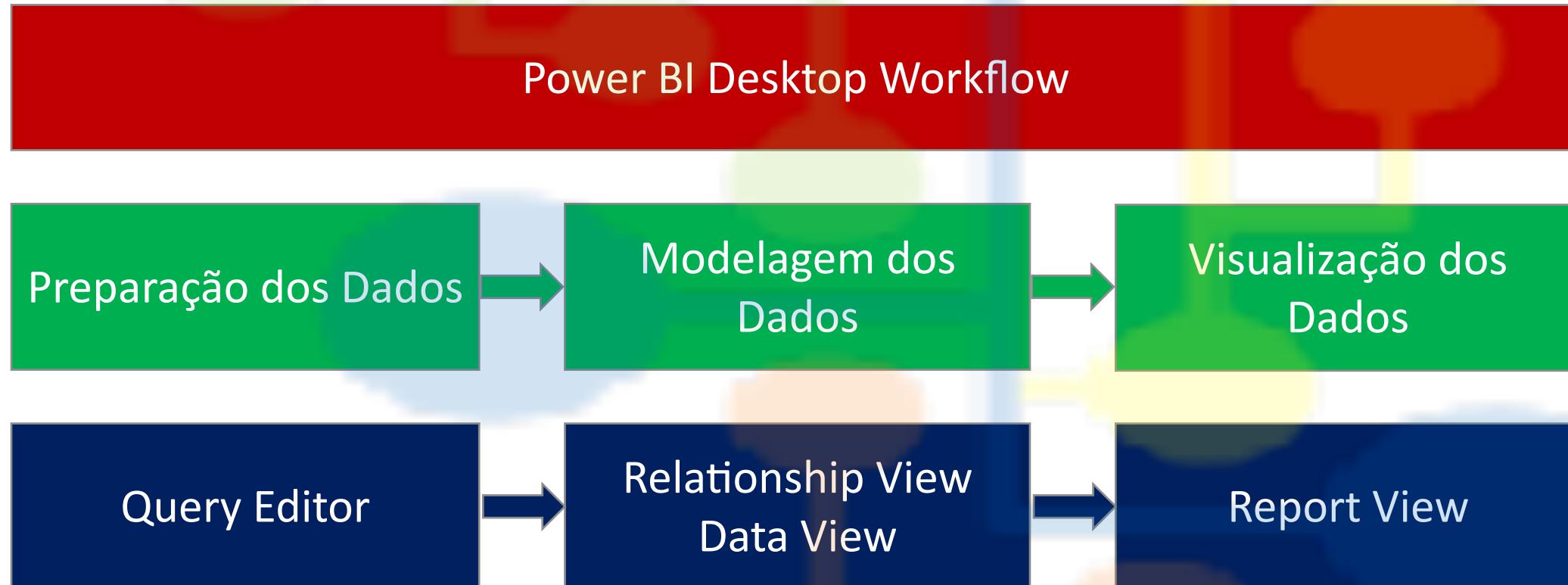


Power BI Desktop Workflow

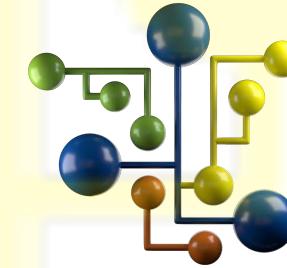


Data Science
Academy

Power BI Desktop Workflow



Relacionamentos e Cardinalidade



Data Science
Academy

Relacionamentos e Cardinalidade

Um para Muitos ($1:*$) e Muitos para Um ($*:1$)

Um para Um ($1:1$)

Muitos para Muitos ($*:*$)

Relacionamentos e Cardinalidade

Um para Muitos (1:*) e Muitos para Um (*:1)

Tabela PRODUTOS

ID-Produto	Produto	Categoria	Segmento	Fabricante
1000	LG K10 TV Power	Celulares	Corporativo	LG
1001	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp
1002	Notebook Dell 8 GB	Eletrônicos	Corporativo	Dell
1003	Desktop HP 16 GB	Eletrônicos	Doméstico	HP
1004	Impressora Deskjet	Eletrônicos	Doméstico	HP
1005	Samsung Galaxy 8	Celulares	Corporativo	Samsung

Tabela VENDAS

ID-Venda	Data	ID-Produto
VE001	28/06/2020	1000
VE002	28/06/2020	1001
VE003	29/06/2020	1000
VE004	30/06/2020	1002
VE005	01/07/2020	1005
VE006	01/07/2020	1005

Relacionamentos e Cardinalidade

Um para Muitos (1:*) e Muitos para Um (*:1)

Tabela PRODUTOS

ID-Produto	Produto	Categoria	Segmento	Fabricante
1000	LG K10 TV Power	Celulares	Corporativo	LG
1001	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp
1002	Notebook Dell 8 GB	Eletrônicos	Corporativo	Dell
1003	Desktop HP 16 GB	Eletrônicos	Doméstico	HP
1004	Impressora Deskjet	Eletrônicos	Doméstico	HP
1005	Samsung Galaxy 8	Celulares	Corporativo	Samsung

Cada produto é único!

Tabela VENDAS

ID-Venda	Data	ID-Produto
VE001	28/06/2020	1000
VE002	28/06/2020	1001
VE003	29/06/2020	1000
VE004	30/06/2020	1002
VE005	01/07/2020	1005
VE006	01/07/2020	1005

Cada produto pode ser vendido diversas vezes!

Relacionamentos e Cardinalidade

Um para Muitos (1:*) e Muitos para Um (*:1)

Tabela PRODUTOS

ID-Produto	Produto	Categoria	Segmento	Fabricante
1000	LG K10 TV Power	Celulares	Corporativo	LG
1001	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp
1002	Notebook Dell 8 GB	Eletrônicos	Corporativo	Dell
1003	Desktop HP 16 GB	Eletrônicos	Doméstico	HP
1004	Impressora Deskjet	Eletrônicos	Doméstico	HP
1005	Samsung Galaxy 8	Celulares	Corporativo	Samsung

ID-Produto é PK
(Primary Key ou Chave Primária)

Tabela VENDAS

ID-Venda	Data	ID-Produto
VE001	28/06/2020	1000
VE002	28/06/2020	1001
VE003	29/06/2020	1000
VE004	30/06/2020	1002
VE005	01/07/2020	1005
VE006	01/07/2020	1005

ID-Produto é FK
(Foreign Key ou Chave Estrangeira)

Relacionamentos e Cardinalidade

Um para Um (1:1)

Tabela VENDEDOR

ID-Registro	Vendedor	ID-Loja	Cidade Loja
1000	Ana Teixeira	SP8821	São Paulo
1001	Josias Silva	RJ9832	Rio de Janeiro
1002	Rodrigo Fagundes	RJ9832	Rio de Janeiro
1003	Fernando Zambrini	A9990	Belo Horizonte
1004	Artur Moreira	SP8821	São Paulo
1005	Mateus Gonçalves	SP8821	São Paulo

Tabela REGISTRO

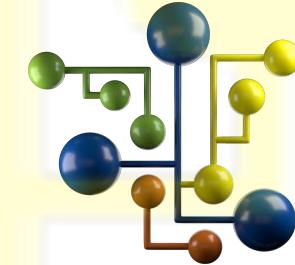
ID-Registro	Data Emissão Registro	Validade
1000	28/06/2021	2023
1001	29/06/2021	2023
1002	29/06/2022	2024
1003	29/06/2023	2025
1004	30/06/2020	2024
1005	01/07/2019	2022

Relacionamentos e Cardinalidade

Muitos para Muitos (*:*)

(Cross Filter Direction no Power BI)

Séries Temporais, Agregação e Filtros



Data Science
Academy

Séries Temporais, Agregação e Filtros

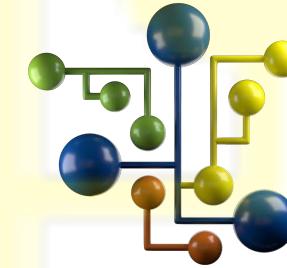
Estudo de Caso 3
(Limpeza e Transformação de Dados)

Séries Temporais

Agregação

Filtros

Gráficos e Mapas



Data Science
Academy

Introdução

Parte 1
Análise de Dados com
Power BI

Parte 2
Gráficos, Mapas e
Dashboards

Parte 3
Estatística, R, Python e
Machine Learning

Parte 4
SQL Analytics,
Automação, Agentes
Virtuais, Power Apps

Introdução

Parte 1
Análise de Dados com
Power BI



Introdução

Primeiros Passos com Power BI Desktop

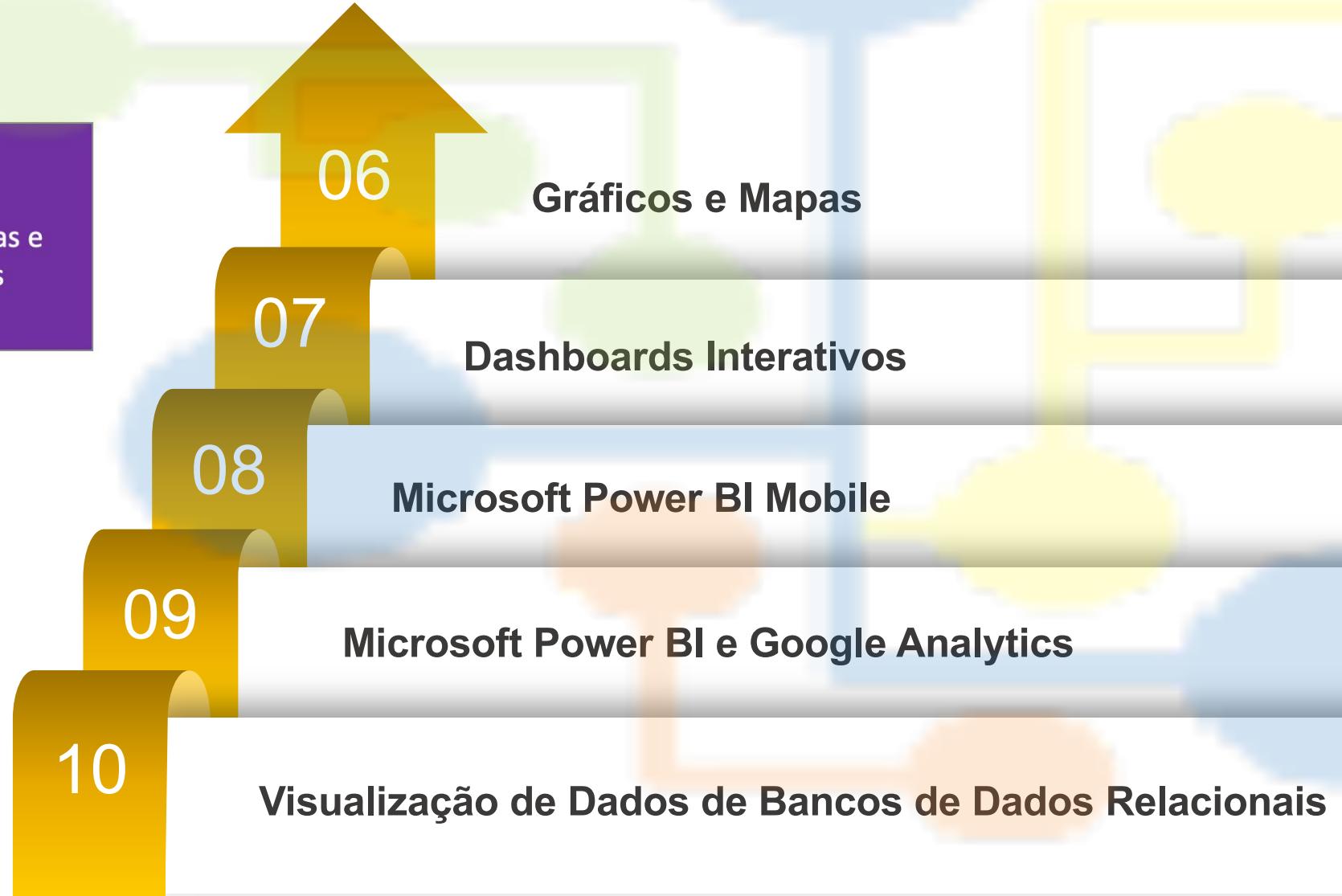
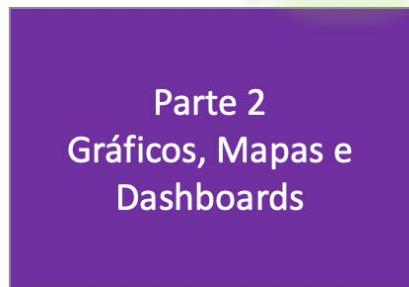
Business Intelligence no Power BI

Modelagem e Relacionamento

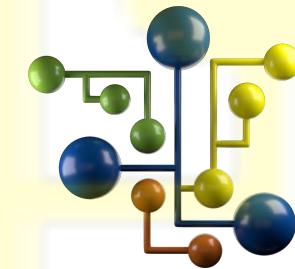
05

Séries Temporais, Agregação e Filtros

Introdução



Qual a Melhor Estratégia de Aprendizado?



Data Science
Academy

Qual a Melhor Estratégia de Aprendizado?



Qual a Melhor Estratégia de Aprendizado?



Estratégia A – 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.



Estratégia B – 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.

Qual a Melhor Estratégia de Aprendizado?

As duas estratégias funcionam e você vai montar o quebra-cabeça.

Mas qual das duas é mais efetiva e leva a um aprendizado mais rápido?

Qual a Melhor Estratégia de Aprendizado?



Estratégia A

Mais eficiente e mais rápida!



Estratégia B

Muito mais lenta, pode causar
frustração e dificulta o aprendizado!

Qual a Melhor Estratégia de Aprendizado?

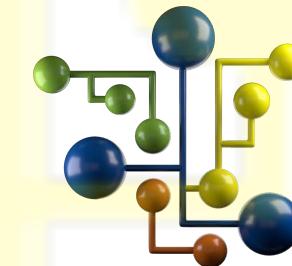
Com Ciência de Dados a ideia é a mesma!

Primeiro compreenda o processo de Ciência de Dados e depois entre no detalhamento, se necessário.

Por isso não recomendamos começar o aprendizado por Estatística ou Matemática, mas sim pela solução de problemas com Data Science.

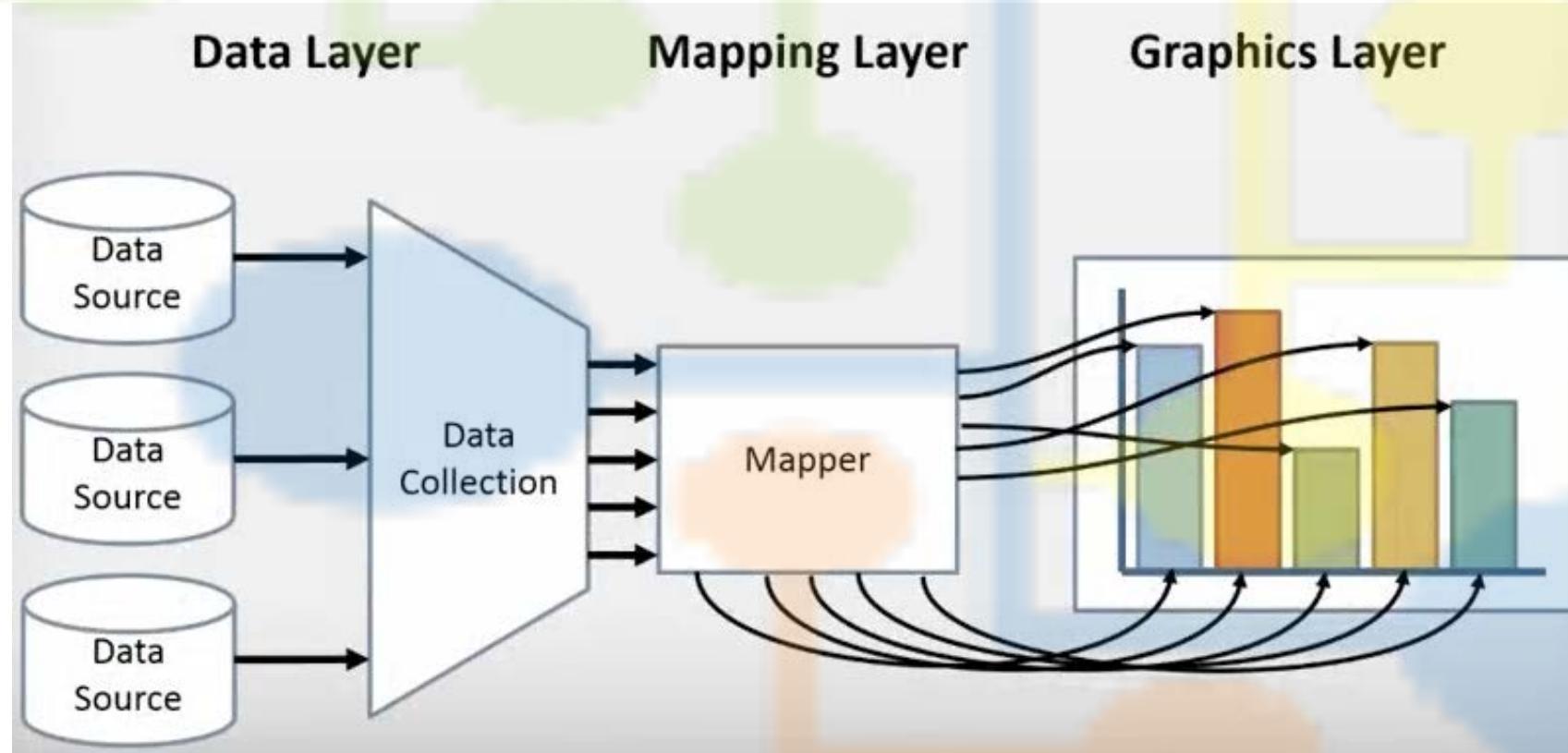
Recomendamos fazer primeiro a **Formação Cientista de Dados** e depois a **Formação Análise Estatística Para Cientistas de Dados**, aqui na DSA!

Data Visualization Framework

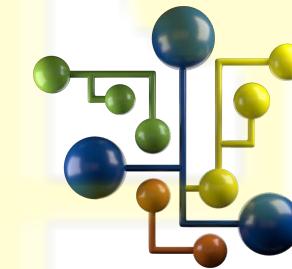


Data Science
Academy

Data Visualization Framework



Visualização de Dados Numéricos e Não Numéricos



Data Science
Academy

Visualização de Dados Numéricos e Não Numéricos

A

Código	Nome	Cargo
1001	Pelé	Cientista de Dados
1002	Zico	Engenheiro de Machine Learning
1003	Romário	Cientista de Dados

B

Índice	Número de Quartos	Preço Casa (R\$)
1	5	250.000
2	3	180.000
3	6	300.000

Visualização de Dados Numéricos e Não Numéricos

A

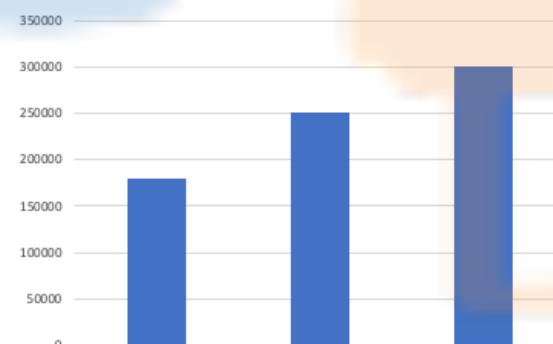
Gráfico da Tabela A



Número de pessoas
por cargo

B

Gráfico da Tabela B



Número de casas por
valor da casa

Visualização de Dados Numéricos e Não Numéricos

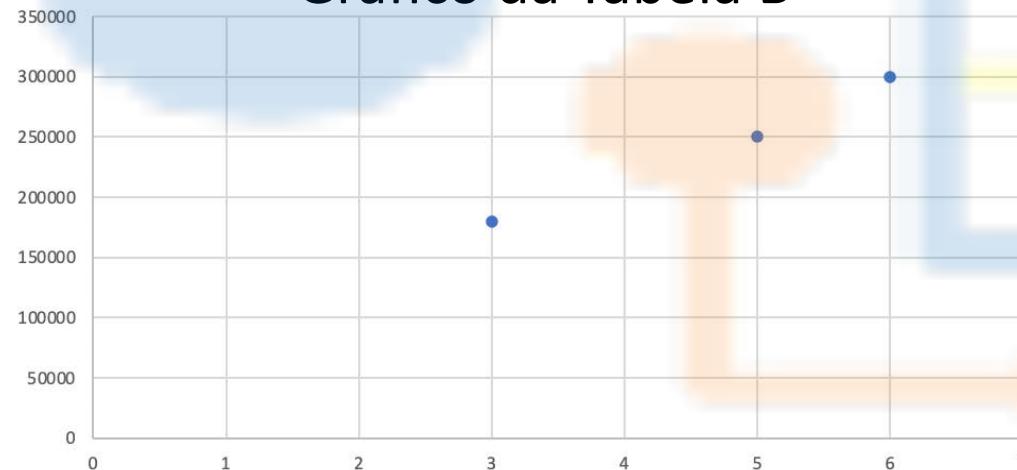
A

Gráfico da Tabela A



B

Gráfico da Tabela B



Número de pessoas
por cargo

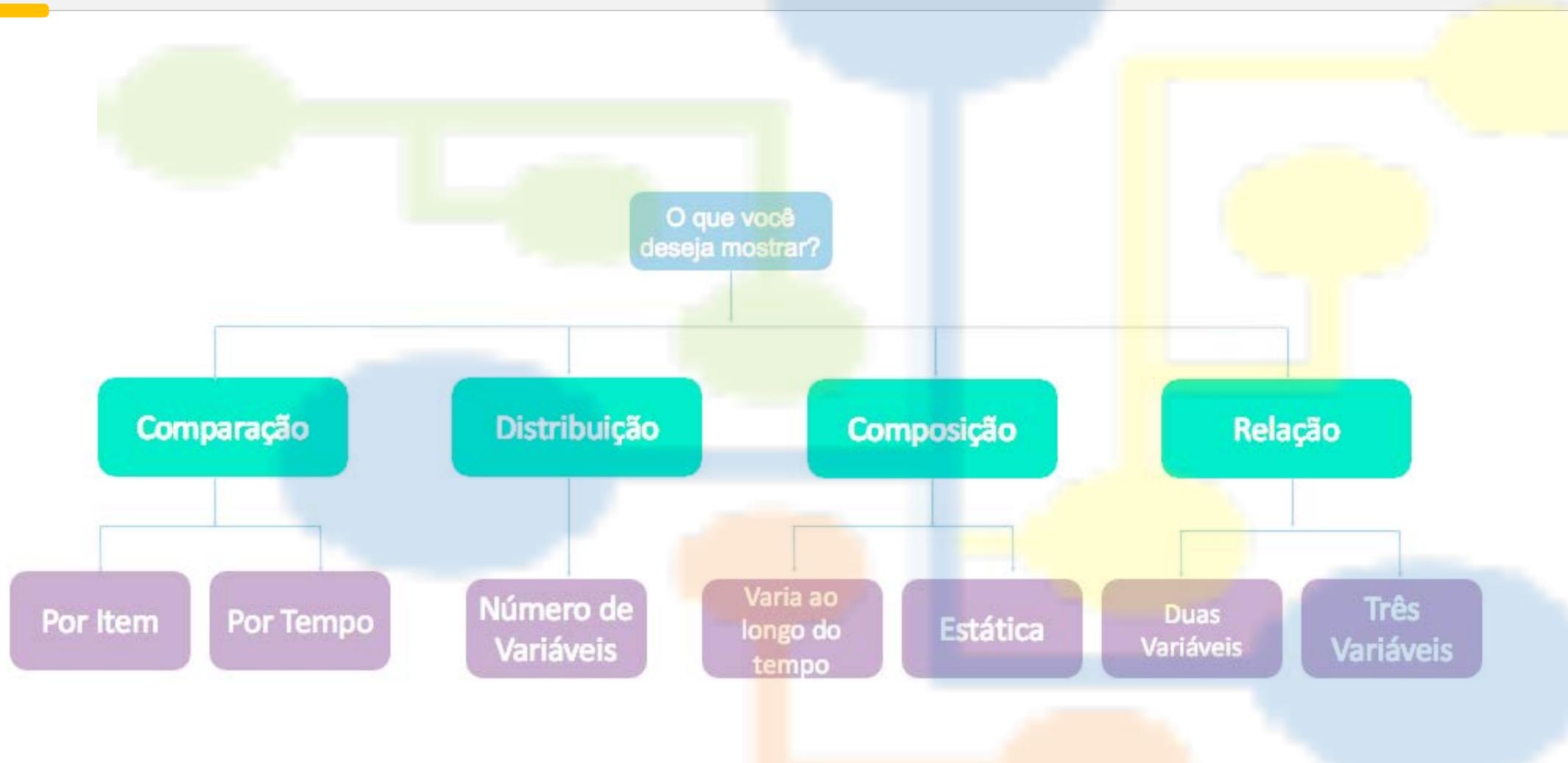
Relação entre número de
quartos e valor da casa

Visualização de Dados Numéricos e Não Numéricos

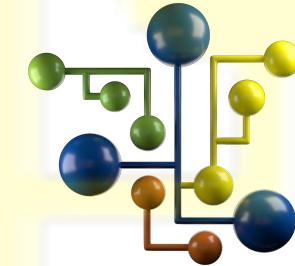


Você não achou que criar gráficos seria a parte mais fácil do processo, não é mesmo?

Visualização de Dados Numéricos e Não Numéricos



O Que é um Dashboard?



Data Science
Academy

O Que é um Dashboard?



O Que é um Dashboard?

Um Dashboard é uma exibição visual das informações mais importantes necessárias para alcançar um ou mais objetivos, consolidado e organizado em uma única tela para que as informações possam ser monitoradas e/ou analisadas ao mesmo tempo.

O Que é um Dashboard?

Um Dashboard é uma ferramenta de gestão e visualização de informações que é usado para monitorar indicadores de desempenho (KPI's), métricas e outros pontos de dados relevantes para o negócio, departamento ou projeto.

O Que é um Dashboard?

Com o uso de visualizações de dados (gráficos, mapas, tabelas e elementos visuais), o Dashboard simplifica o complexo processo de análise de dados e provê ao usuário uma visão clara do que já ocorreu, da situação atual ou eventuais previsões.

O Que é um Dashboard?

Dashboard Como Ferramenta de Gerenciamento de Informações



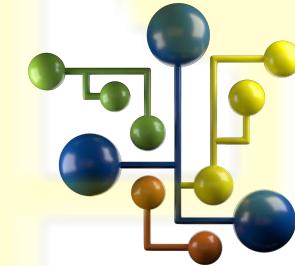
O Que é um Dashboard?



Um Dashboard bem projetado é uma notável ferramenta de gerenciamento de informações.

Mini-Projeto 1

Análise de Dados de E-Commerce
com Power BI e Google Analytics



Data Science
Academy

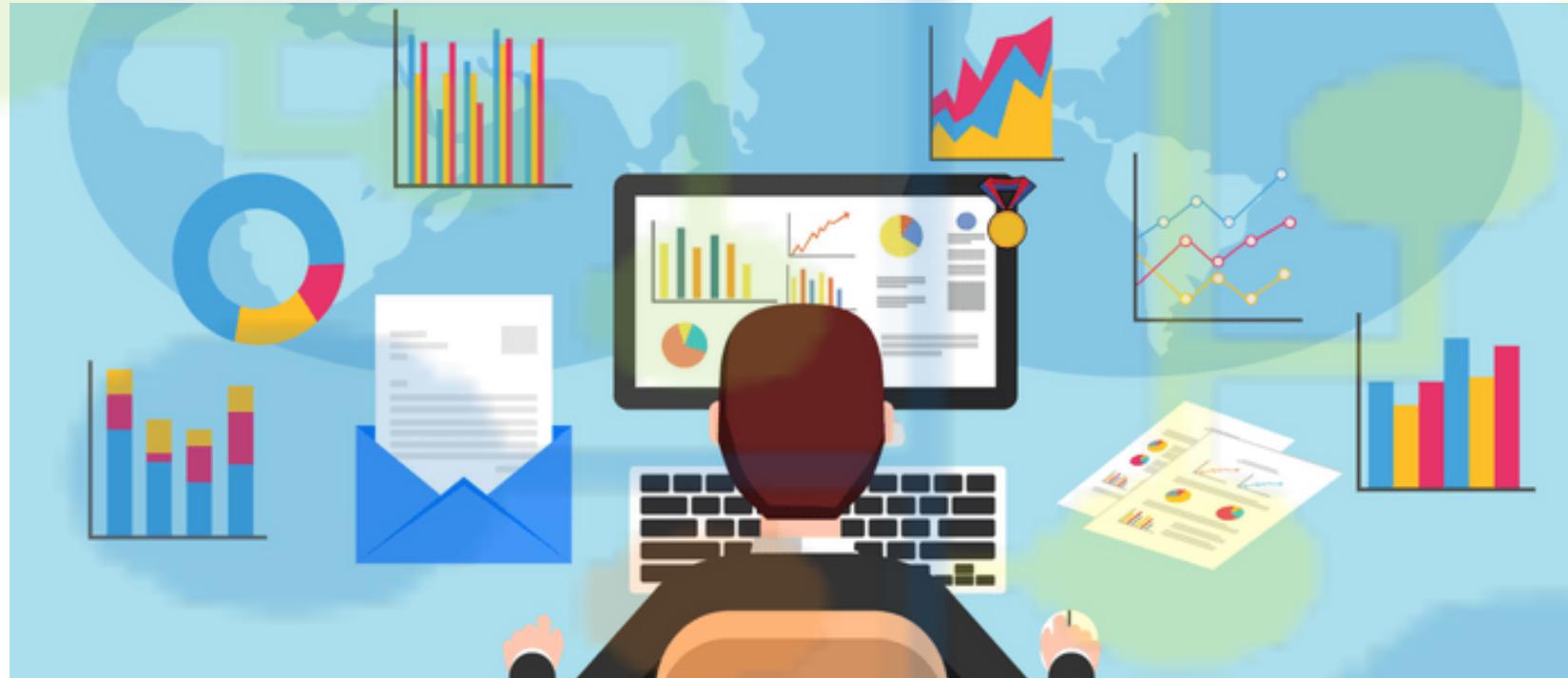
Mini-Projeto 1

Definição do Problema

Uma empresa comercializa produtos online. A empresa configurou o Google Analytics para extrair dados sobre o perfil dos visitantes do portal de e-commerce e agora o Gestor da área de Marketing gostaria de ter as respostas às perguntas abaixo:

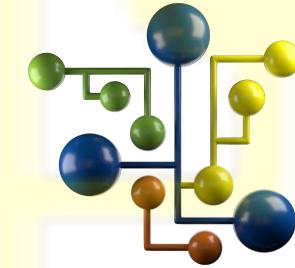
1. Como os clientes mais acessam nosso portal, por busca orgânica ou paga?
2. Quanto tempo em média um visitante permanece em nosso portal por dia do mês?
3. Qual a principal fonte de acesso ao nosso portal?
4. Qual o sistema operacional mais usado para acessar nosso portal?
5. Qual o dispositivo mais usado para acesso ao nosso portal?
6. Qual o total de faturamento por dia?

Mini-Projeto 1



Seu trabalho como analista é analisar os dados e entregar as respostas ao Gestor de Marketing!

O Que São Bancos de Dados Relacionais?



Data Science
Academy

O Que São Bancos de Dados Relacionais?

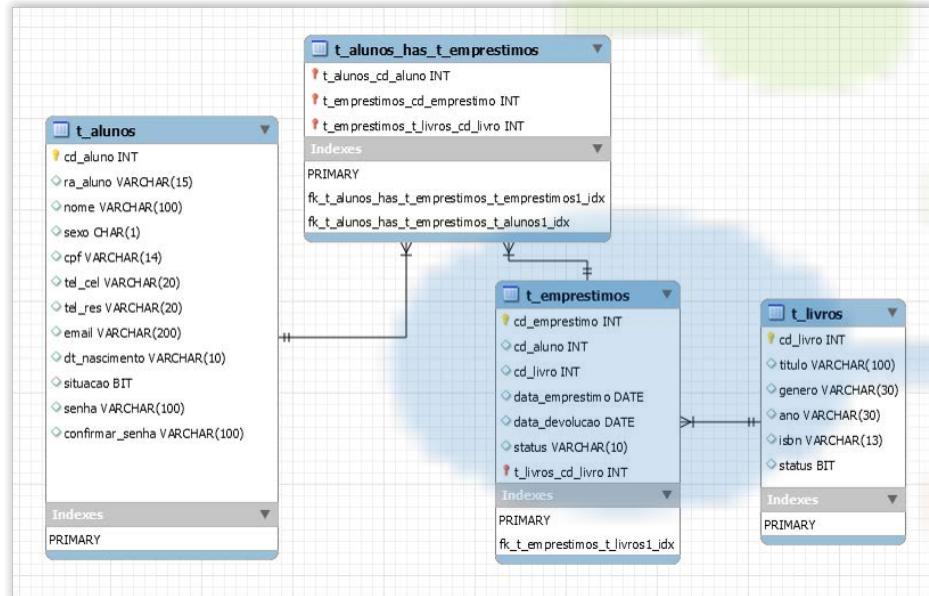


Um banco de dados relacional é um tipo de banco de dados que armazena e fornece acesso a pontos de dados relacionados entre si.

Para garantir que os dados sejam sempre precisos e acessíveis, os bancos de dados relacionais seguem determinadas regras de integridade.

Por exemplo, uma regra de integridade pode especificar que linhas duplicadas não são permitidas em uma tabela para eliminar o potencial de informações errôneas que entram no banco de dados.

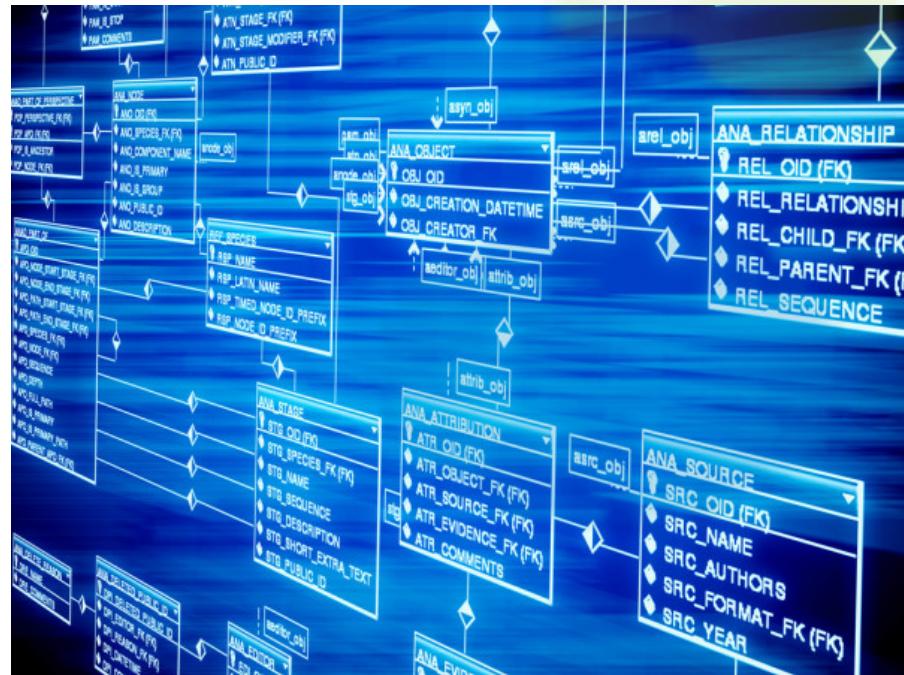
O Que São Bancos de Dados Relacionais?



Bancos de dados relacionais são baseados no modelo relacional, uma maneira intuitiva e direta de representar dados em tabelas.

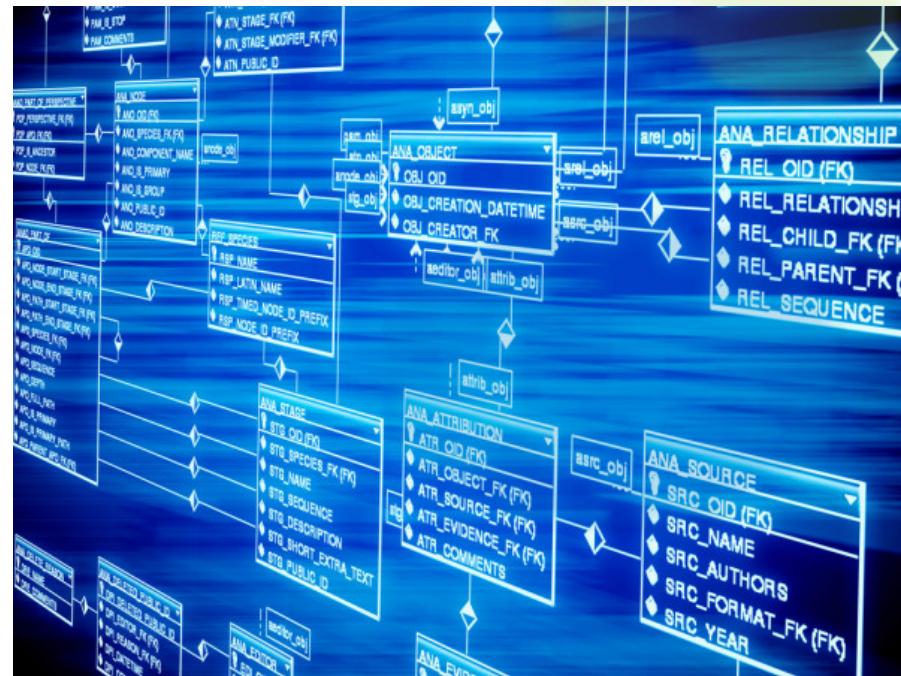
Em um banco de dados relacional, cada linha na tabela é um registro com uma ID exclusiva chamada chave. As colunas da tabela contêm atributos dos dados e cada registro geralmente tem um valor para cada atributo, facilitando o estabelecimento das relações entre os pontos de dados.

O Que São Bancos de Dados Relacionais?



Modelo Lógico x Modelo Físico

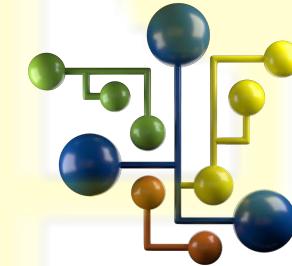
O Que São Bancos de Dados Relacionais?



O modelo relacional significa que as estruturas de dados lógicas: tabelas de dados, views e índices são separadas das estruturas de armazenamento físico.

Essa separação significa que os Administradores de Banco de Dados (DBAs – Database Administrators) podem gerenciar o armazenamento de dados físicos sem afetar o acesso a esses dados como uma estrutura lógica. Por exemplo, a renomeação de um arquivo de banco de dados não renomeia as tabelas armazenadas nele.

Sistemas Gerenciadores de Bancos de Dados



Data Science
Academy

Sistemas Gerenciadores de Bancos de Dados



Esses são SGBDs (Sistemas Gerenciadores de Bancos de Dados).

Cada SGBD pode gerenciar um ou mais bancos de dados.

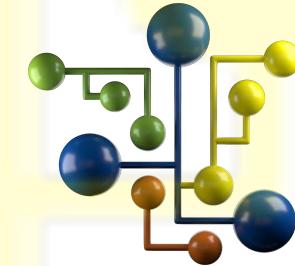
Bancos de Dados Relacionais

x

Bancos de Dados Não-Relacionais

x

Data Stores



Data Science
Academy

Bancos de Dados Relacionais x Bancos de Dados Não-Relacionais x Data Stores

Modelagem de Bancos de Dados Relacionais, Não Relacionais e Data Stores

Modelagem de Dados, SQL Analytics, Integração, Governança e Segurança. A base do trabalho de um Arquiteto de Dados.
Este é o segundo curso da Formação Arquiteto de Dados.

44 horas | Acesso Exclusivo Turmas Fechadas



Bancos de Dados Relacionais x Bancos de Dados Não-Relacionais x Data Stores

SGBDs Para Bancos de Dados Relacionais

Oracle
SQL Server
PostgreSQL
MySQL

SGBDs Para Bancos de Dados Não-Relacionais

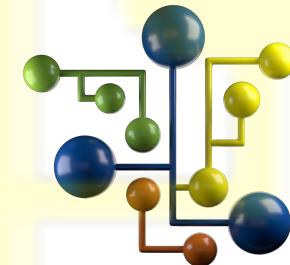
MongoDB
Apache Cassandra
Apache Hbase
Amazon DynamoDB

Data Stores

Apache Hadoop
Amazon S3
Cassandra File System
GlusterFS
Lustre

Estudo de Caso 4

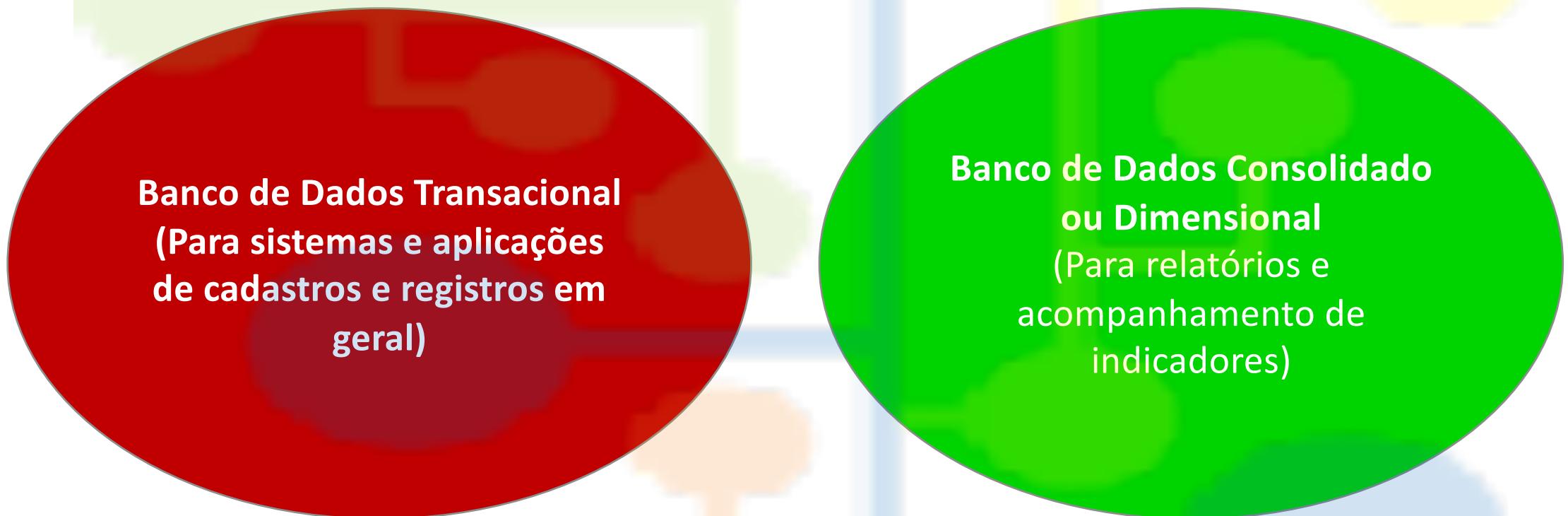
Construindo um Data Warehouse de Vendas



Data Science
Academy

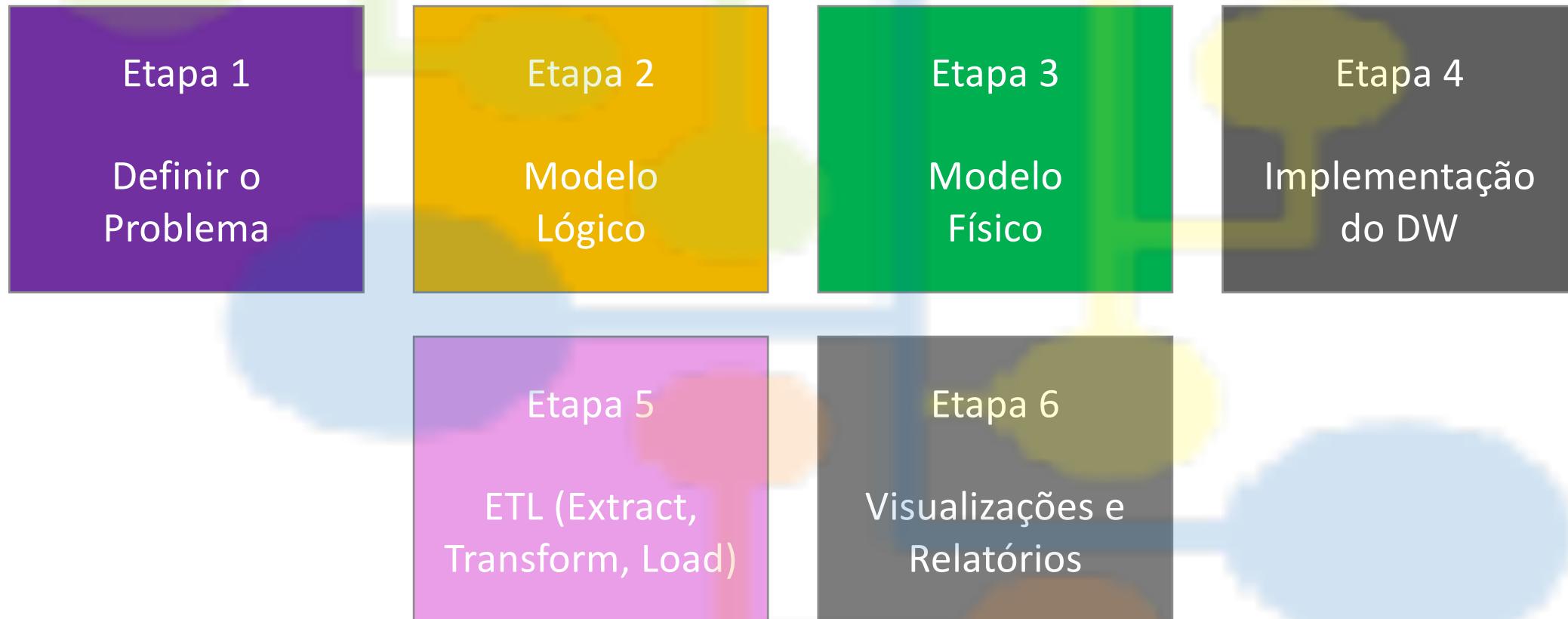
Estudo de Caso 4

Construindo um Data Warehouse de Vendas

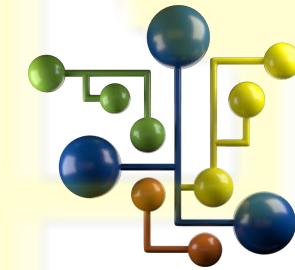


Estudo de Caso 4

Construindo um Data Warehouse de Vendas

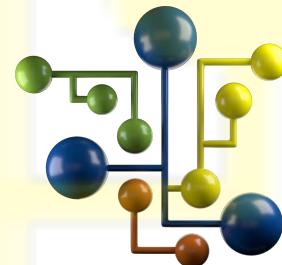
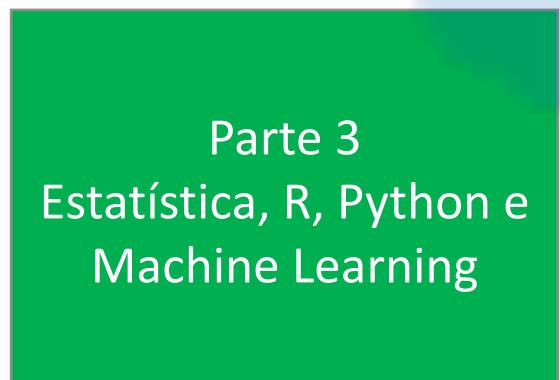


Estatística Fundamental Para Data Science



Data Science
Academy

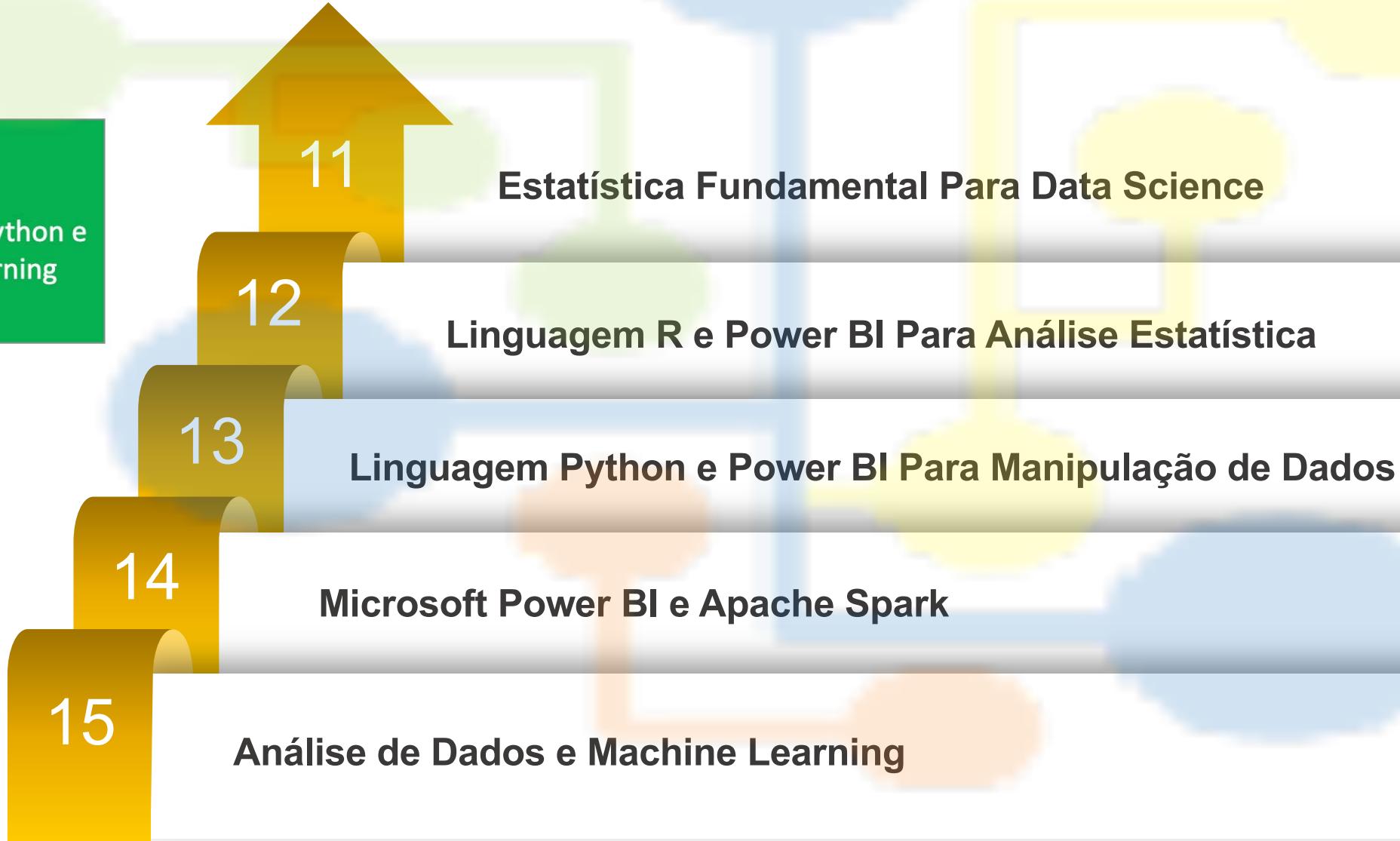
Conteúdo Programático (20 Módulos)



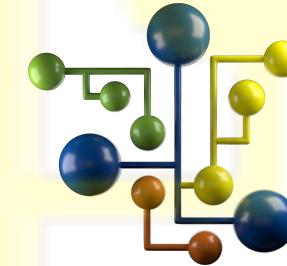
**Data Science
Academy**

Introdução

Parte 3
Estatística, R, Python e
Machine Learning



Estatística
x
Data Science



Data Science
Academy

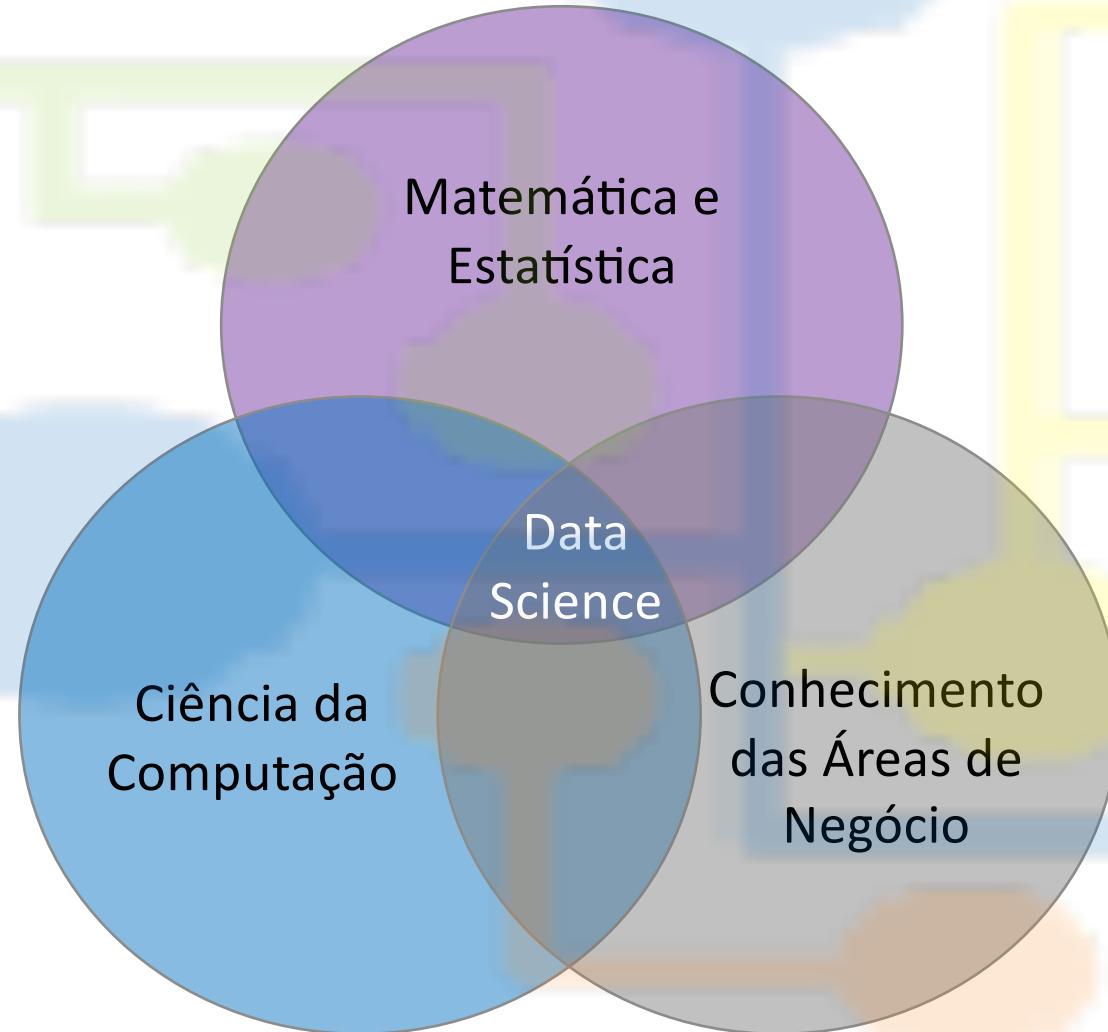
Estatística x Data Science

Estatística e Data Science são a mesma coisa?

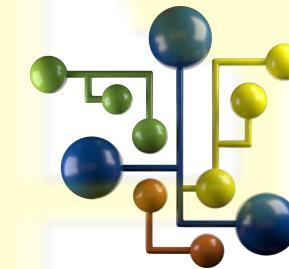
Estatística x Data Science

NÃO

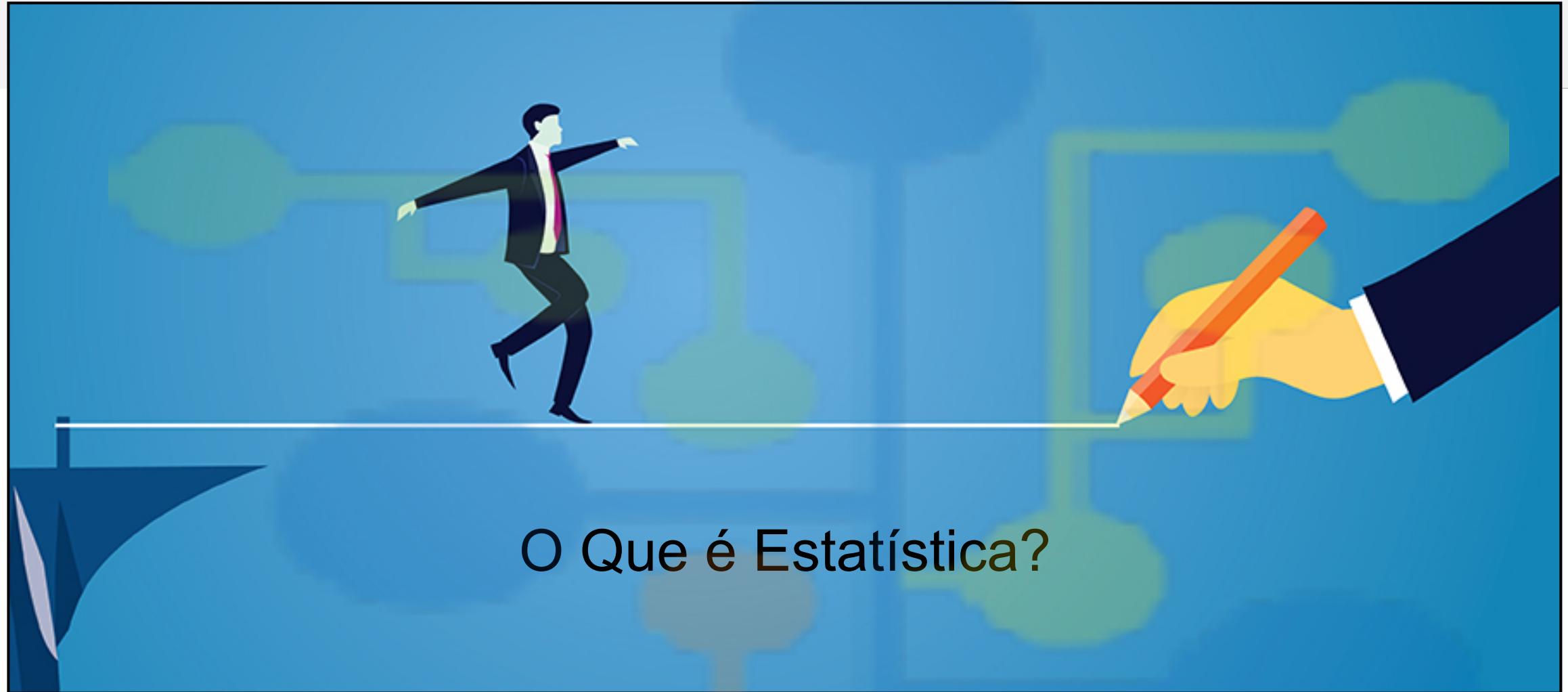
Estatística x Data Science



Definindo Estatística



Data Science
Academy



O Que é Estatística?

A Estatística pode ser pensada como a ciência de aprendizagem a partir de dados.

O Que é Estatística?

No mundo moderno, a alta competitividade na busca de tecnologias e de mercados tem provocado uma constante corrida pela informação.

Definindo Estatística

O Que é Estatística?

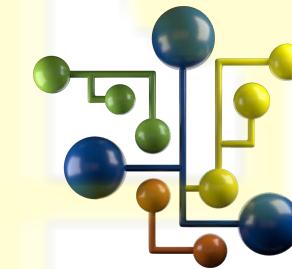
A Estatística fornece técnicas e métodos de análise de dados que auxiliam o processo de tomada de decisão nos mais variados problemas onde existe incerteza.

Áreas Onde Aplicamos Estatística

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

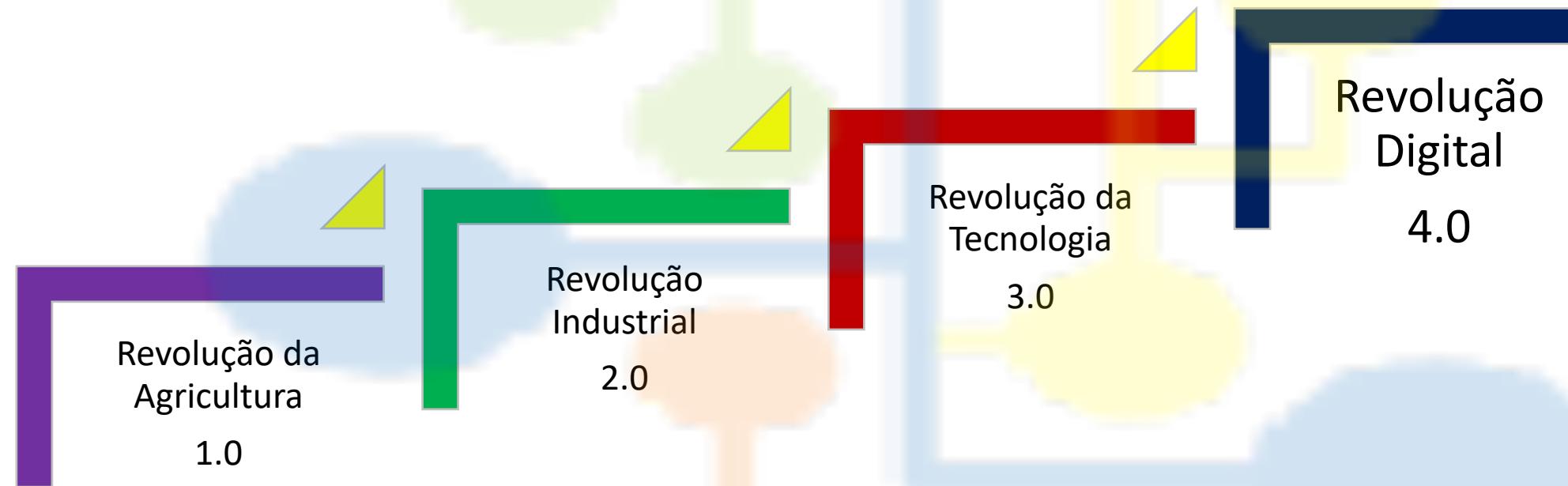


Estatística e a Quarta Revolução Industrial



Data Science
Academy

Estatística e a Quarta Revolução Industrial



Estatística e a Quarta Revolução Industrial



E qual a matéria-prima da quarta revolução industrial?

Dados



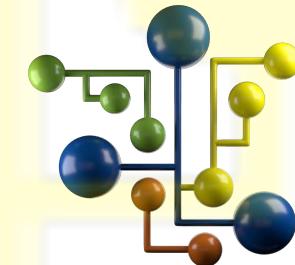
Estatística e a Quarta Revolução Industrial

E o que precisamos para obter informação, conhecimento e inteligência a partir dos dados?

Técnicas de análise de dados, como as que são fornecidas pela Estatística!



As Principais Áreas da Estatística



Data Science
Academy

As Principais Áreas da Estatística

As Principais Áreas da
Estatística



Probabilidade

Estudo da aleatoriedade e da incerteza.



Estatística Descritiva

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

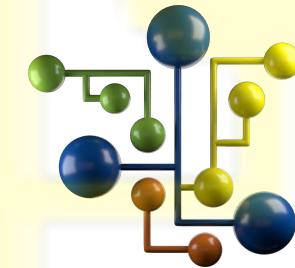


Estatística Inferencial

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



E o Que Devo Aprender Primeiro?



Data Science
Academy

E o Que Devo Aprender Primeiro?

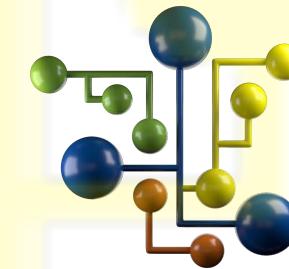


Formação Cientista
de Dados



Formação Análise
Estatística Para
Cientistas de Dados

População e Amostra



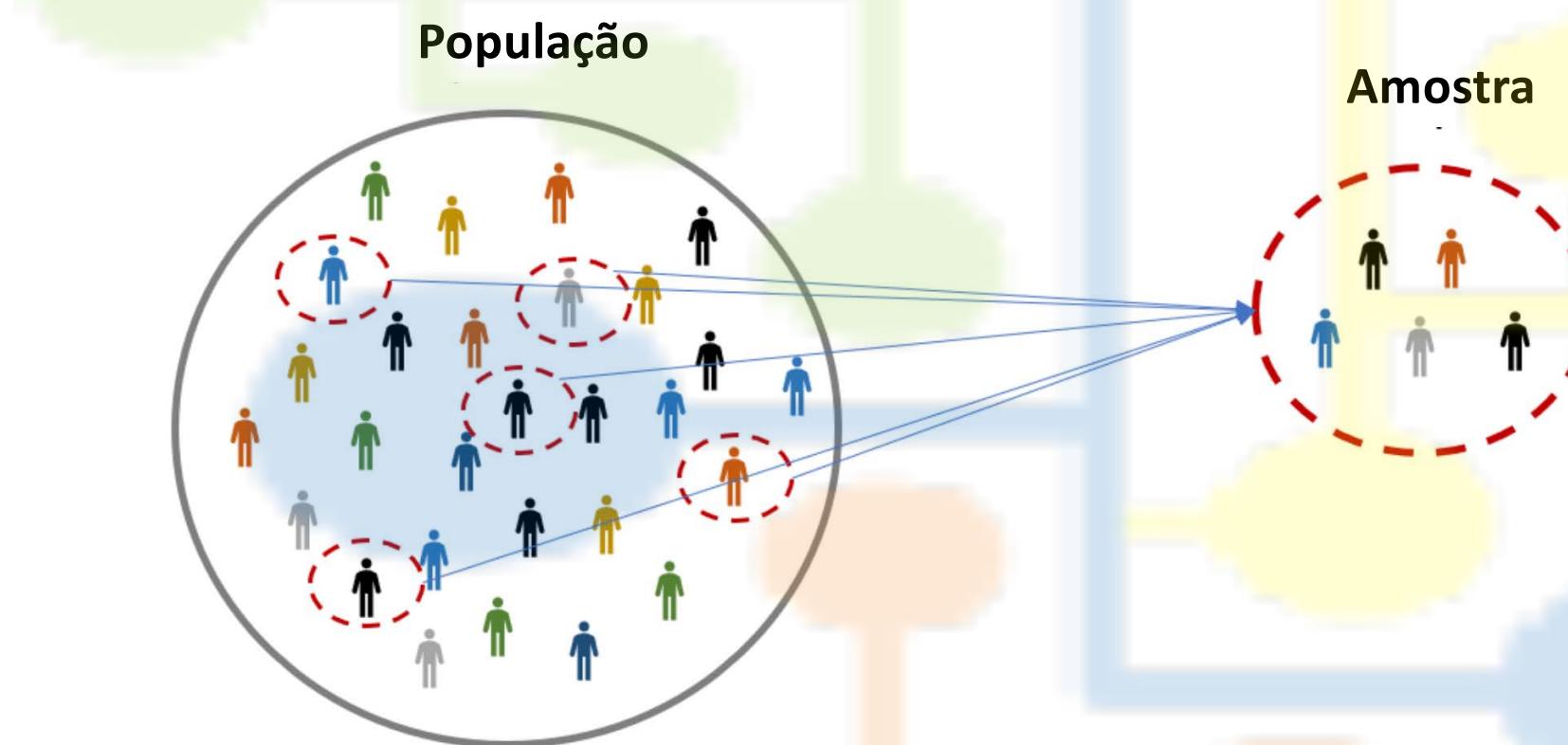
Data Science
Academy

População e Amostra

Sempre que você se deparar com um novo conjunto de dados, uma das primeiras perguntas que devem ser feitas é:

Qual é minha população e qual é minha amostra?

População e Amostra



População e Amostra

Exemplo: Pesquisa Eleitoral

Os institutos de pesquisa, examinam uma amostra e a partir disso, deduzem informações sobre toda a população.



População e Amostra

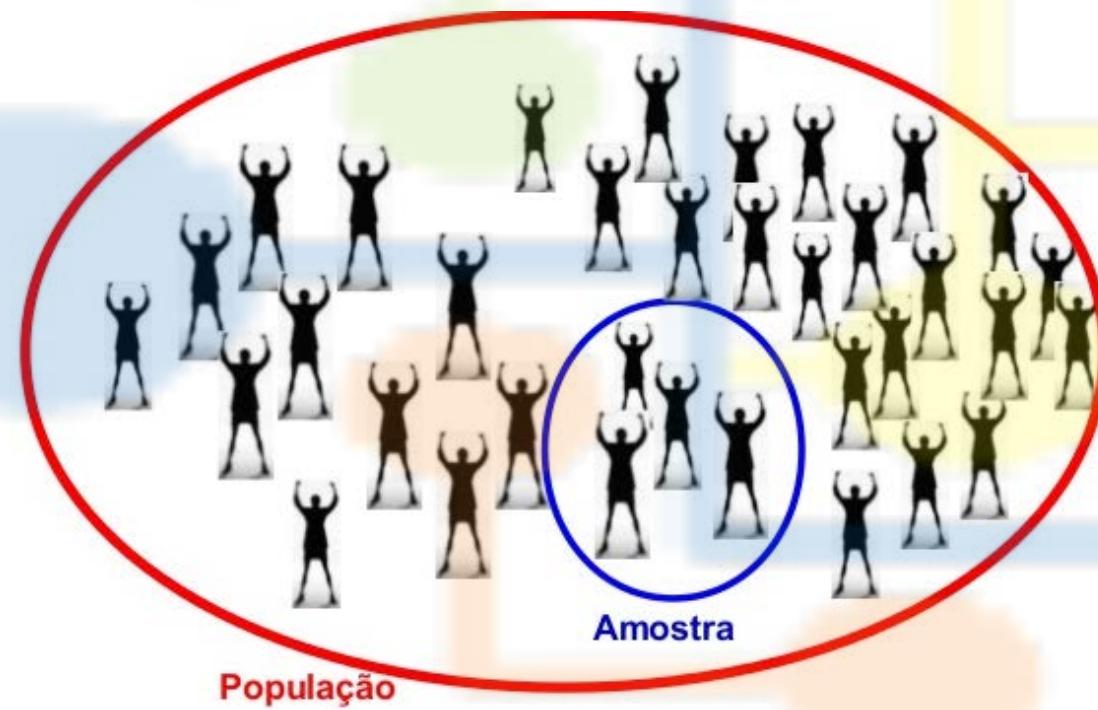
É Sopa!





População e Amostra

Como garantir que a **amostra** representa fielmente a **população**?



População e Amostra

É Sopa novamente!



População e Amostra

Randomização



População e Amostra

Ok, entendi.

Mas como eu faço com indivíduos ou com os meus dados??

População e Amostra

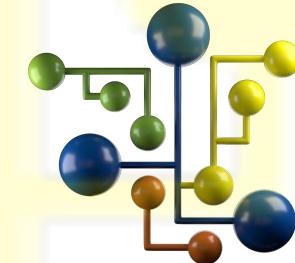
Simplesmente, você coleta sua **amostra** de forma randomizada, sem escolher exatamente quem fará parte da amostra.

População e Amostra

A Estatística oferece diversas técnicas e ferramentas para o processo de amostragem



Técnicas de Amostragem

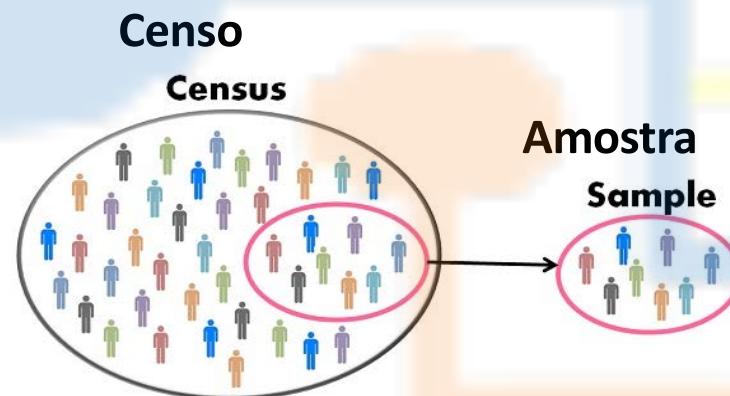


Data Science
Academy

Técnicas de Amostragem

Amostragem: usa a coleta, organização, apresentação e análise dos dados como meio de estudar os parâmetros de uma população.

Censo: é a técnica que seleciona e avalia **todos** os elementos da população quando se realiza uma pesquisa.



Técnicas de Amostragem

Por que não medir uma **população inteira**, ao invés de medir apenas uma **amostra**?



Técnicas de Amostragem

Dependendo das circunstâncias, medir uma população inteira seria **caro demais** ou até mesmo inviável.



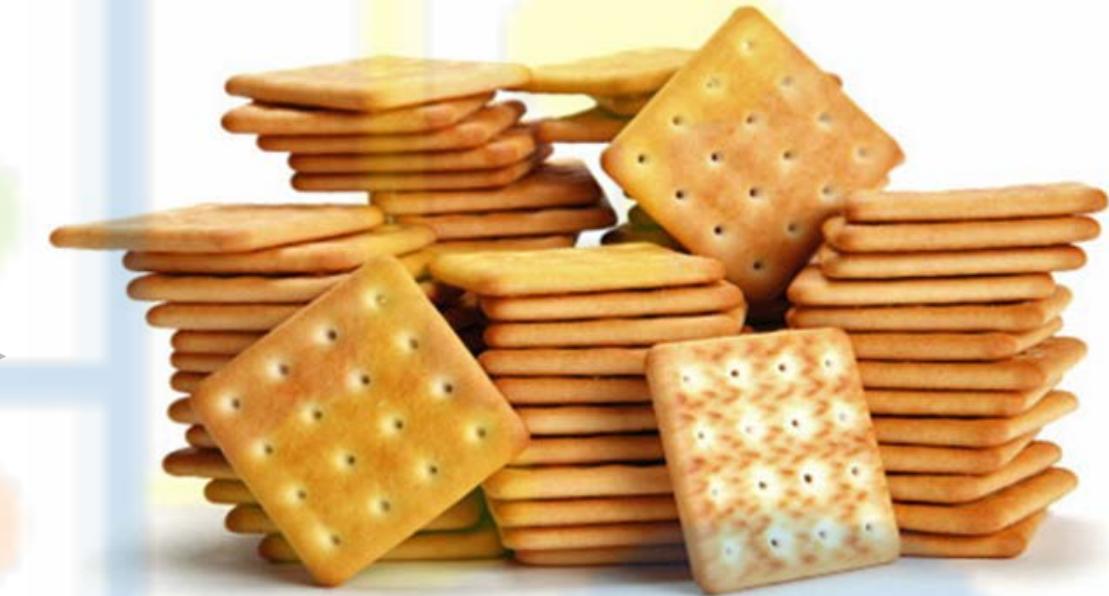
Técnicas de Amostragem



Técnicas de Amostragem



Inferência



Técnicas de Amostragem

Parabéns!!

Você acabou de ter a definição de
Estatística Inferencial.

Técnicas de Amostragem

Lembra da sopa?



Técnicas de Amostragem



Os dados devem ser tratados com muito cuidado, de modo que as conclusões refletem a realidade.

Técnicas de Amostragem

Ao escolher uma amostra...



Técnicas de Amostragem

E como deve ser a amostragem?

Probabilística

X

Não Probabilística

Técnicas de Amostragem

Amostragem Probabilística

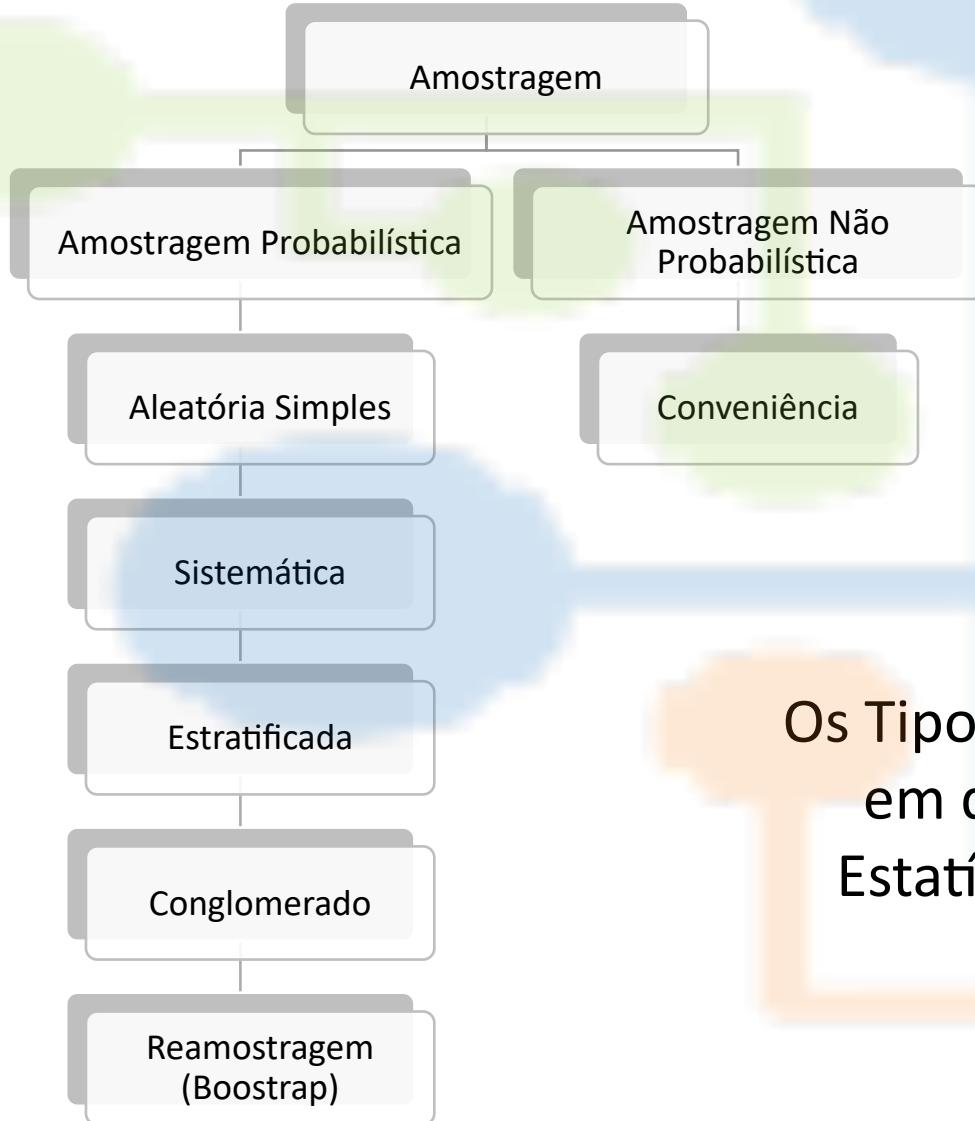
- Amostragem Simples ao Acaso
- Amostragem Sistemática
- Amostragem Por Conglomerado
- Amostragem Estratificada
- Reamostragem (Bootstrap)

Técnicas de Amostragem

Amostragem Não Probabilística

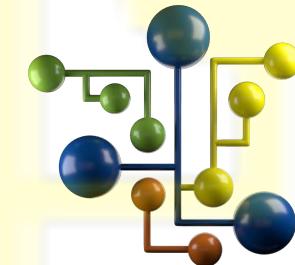
- Amostragem a Esmo
- Amostragem Intencional
- Amostragem Por Voluntários

Técnicas de Amostragem



Os Tipos de Amostragem são estudados
em detalhes na Formação Análise
Estatística Para Cientistas de Dados.

Parâmetro
X
Estatística



Data Science
Academy

Parâmetro x Estatística

Parâmetro – característica sobre a população.

Valores calculados usando dados da população são chamados de **parâmetros**.

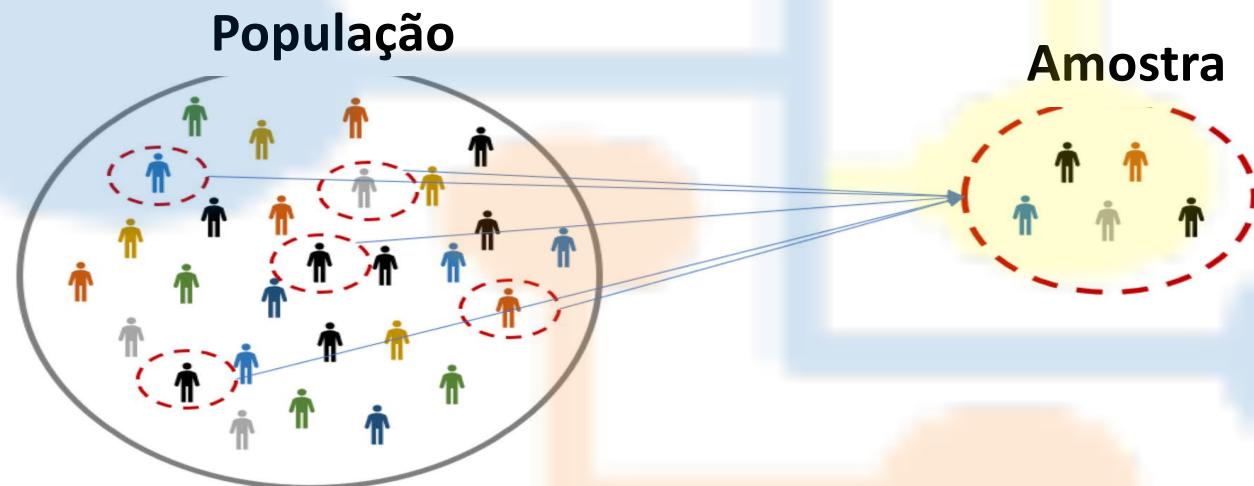
Parâmetro x Estatística

Estatística – característica sobre a amostra.

Valores calculados usando dados da amostra são chamados de **estatísticas**.

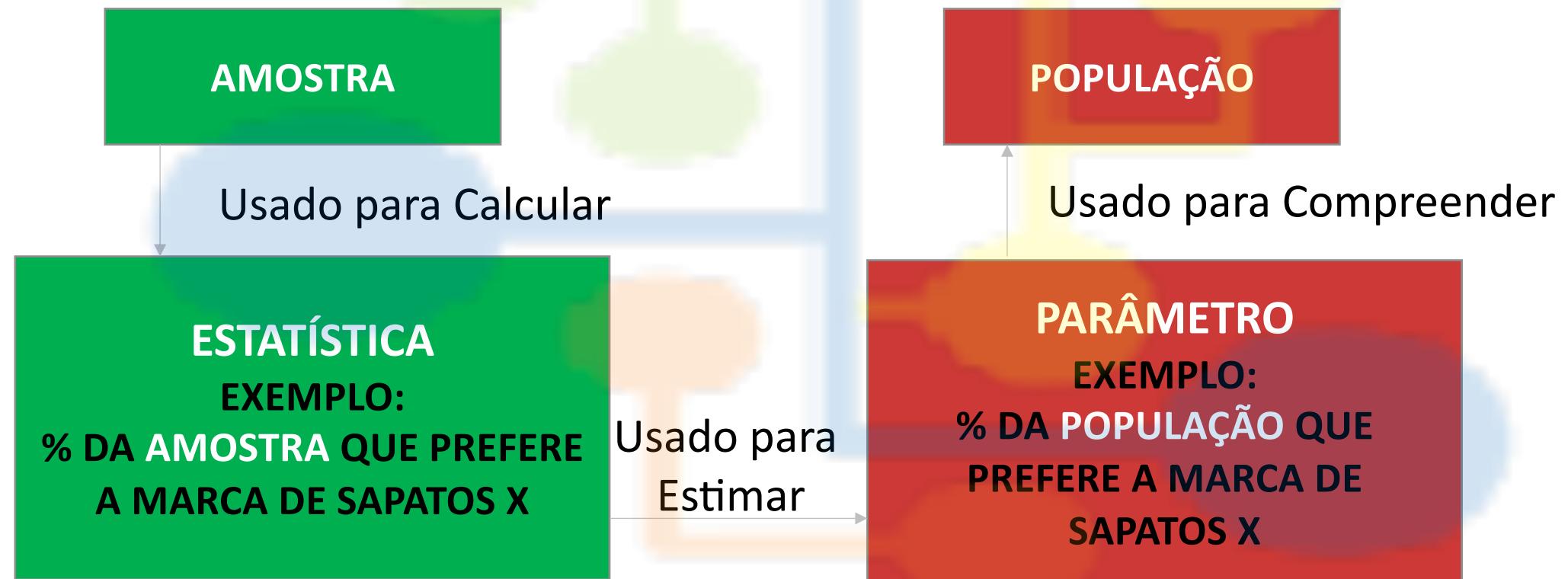
Parâmetro x Estatística

Estatística Inferencial realiza deduções e conclusões sobre a população, baseadas nos resultados obtidos da análise da amostra.

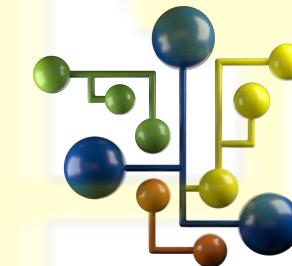


Parâmetro x Estatística

A **Estatística Inferencial** realiza deduções e conclusões sobre a população, baseadas nos resultados obtidos na análise da amostra.



Dados Primários
X
Dados Secundários



Data Science
Academy

Dados Primários x Dados Secundários

Vamos relembrar um conceito fundamental!



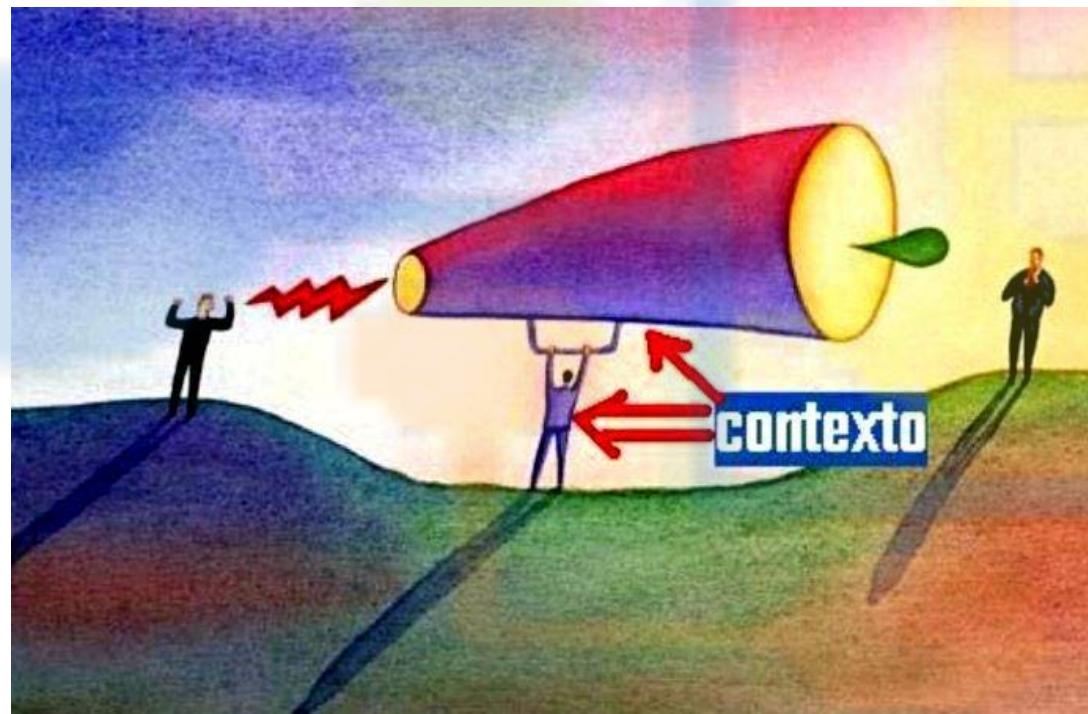
Dados Primários x Dados Secundários

- **Dados** – valores coletados através de observação ou medição.
- **Informação** – dados que são transformados em fatos relevantes e usados para um propósito específico.



Dados Primários x Dados Secundários

Dados não fazem sentido, se não forem colocados em um contexto!



Dados Primários x Dados Secundários

Os dados podem ser obtidos através de duas fontes principais:

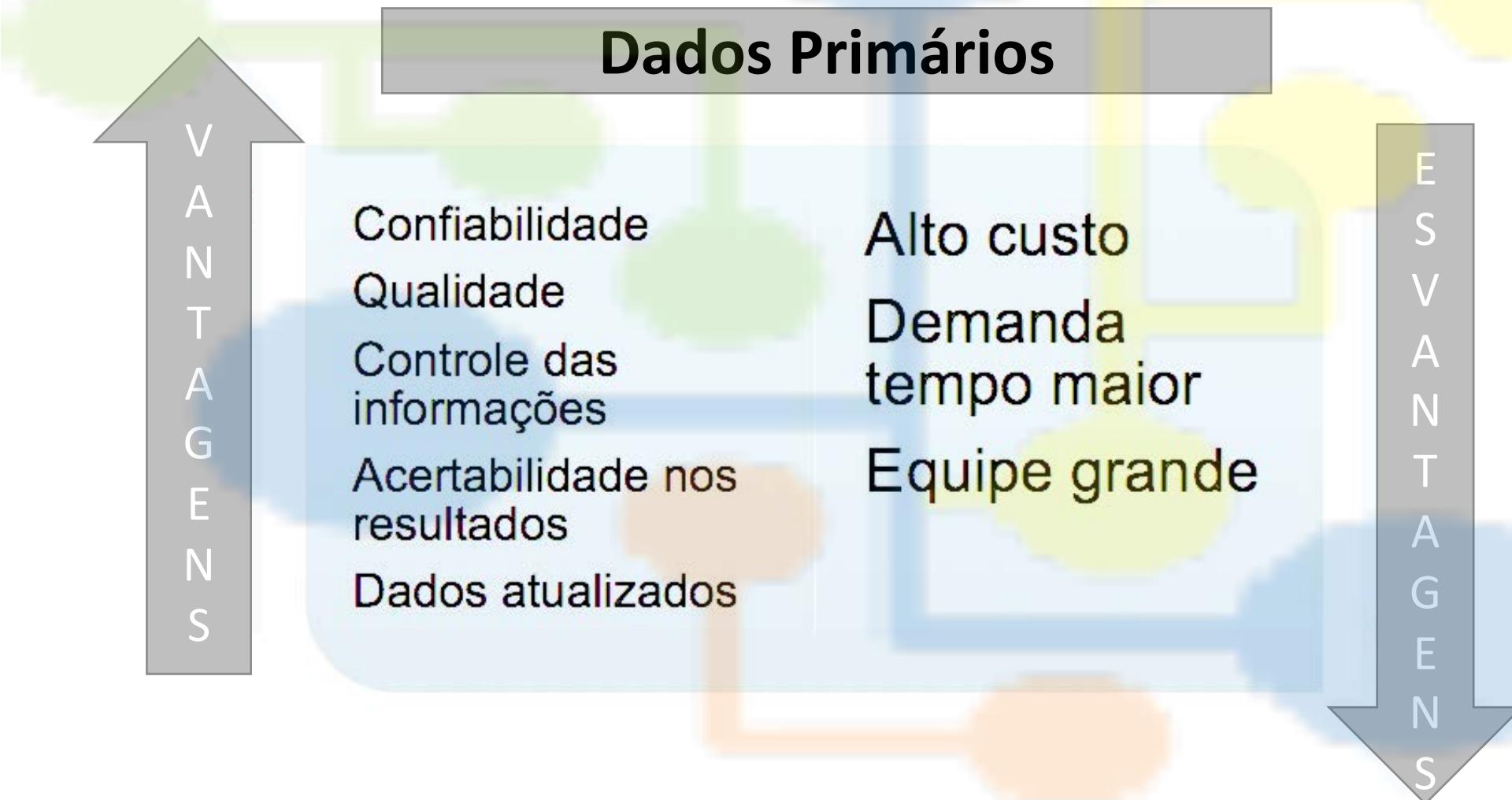
Dados Primários

- Coletados por quem faz a análise
- Confiáveis
- Possuem maior controle

Dados Secundários

- Coletados por terceiros
- Não Confiáveis
- Não possuem muito controle

Dados Primários x Dados Secundários



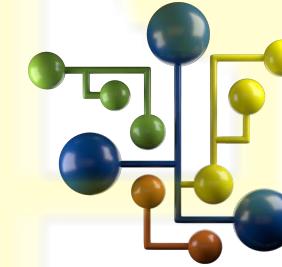
Dados Primários x Dados Secundários



Dados Primários x Dados Secundários

Informação → Conhecimento

Observações
X
Variáveis



Data Science
Academy

Observações x Variáveis

Observação

Uma observação é uma ocorrência de um item de dados específico que é gravada sobre uma unidade de dados.

Outras Nomenclaturas: Registro, Ocorrência do Evento, Instância ou Linha em uma tabela.

Observações x Variáveis

Variável

Variável é a característica de interesse que é medida em cada elemento da amostra ou população. Como o nome sugere, seus valores variam de elemento para elemento.

As variáveis podem ter valores numéricos ou não numéricos.

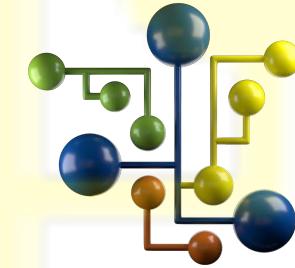
Outras Nomenclaturas: Atributo, Característica ou Coluna em uma tabela.

Observações x Variáveis

The diagram illustrates the relationship between observations and variables. A box labeled "Observações" (Observations) has a bracket pointing to a table containing five rows of data. Another box labeled "Variáveis" (Variables) has a bracket pointing to the column headers of the table.

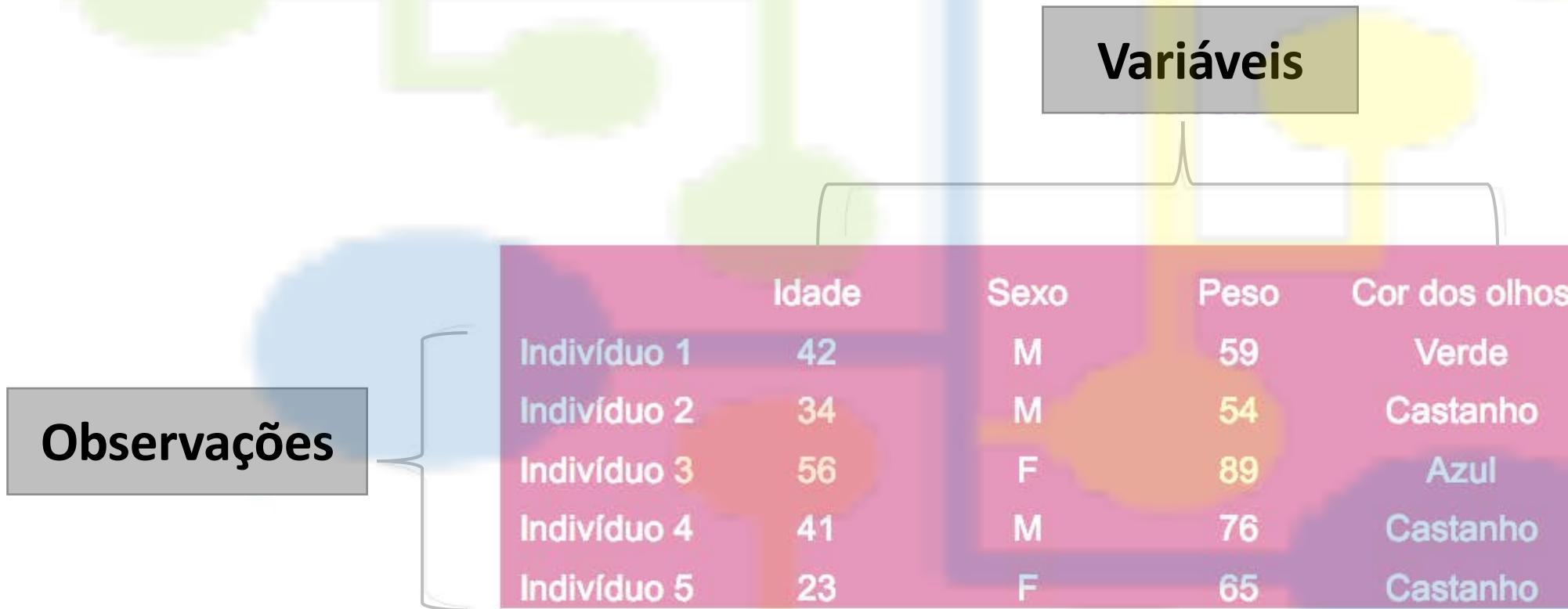
	Idade	Sexo	Peso	Cor dos olhos
Indivíduo 1	42	M	59	Verde
Indivíduo 2	34	M	54	Castanho
Indivíduo 3	56	F	89	Azul
Indivíduo 4	41	M	76	Castanho
Indivíduo 5	23	F	65	Castanho

Tipos de Variáveis



Data Science
Academy

Tipos de Variáveis



Tipos de Variáveis

Como classificar os tipos de dados (tipos de variáveis)?

Tipos de Variáveis

Os dados podem conter variáveis:

Qualitativas – utilizam termos **descritivos** para descrever algo de interesse. Ex: cor dos olhos, estado civil, religião, sexo, grau de escolaridade, classe social, tipo sanguíneo, cor da pele, etc...

Tipos de Variáveis

Os dados podem conter variáveis:

Quantitativas – representadas por valores numéricos que podem ser **contados** ou **medidos**. Ex: número de crianças em uma sala de aula, peso do corpo humano, idade, número de filhos, etc...

Tipos de Variáveis

Dentro desta classificação, podemos ter variáveis:



Tipos de Variáveis

Um dado classificado como "**idade**" pode ser **quantitativo**.
Ex.: 11, 15, 18, 25, 42 anos.

Entretanto, se esse dado for informado por "**faixa etária**" ele é **qualitativo** (ordinal).
Ex: 0 – 5 anos, 6 – 12 anos, 13 – 18 anos, 19 – 28 anos

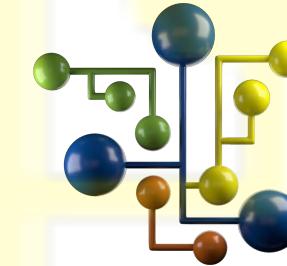


Tipos de Variáveis



É muito importante classificar os dados de forma correta, pois isso permitirá a você escolher o melhor teste estatístico a ser utilizado na análise dos dados.

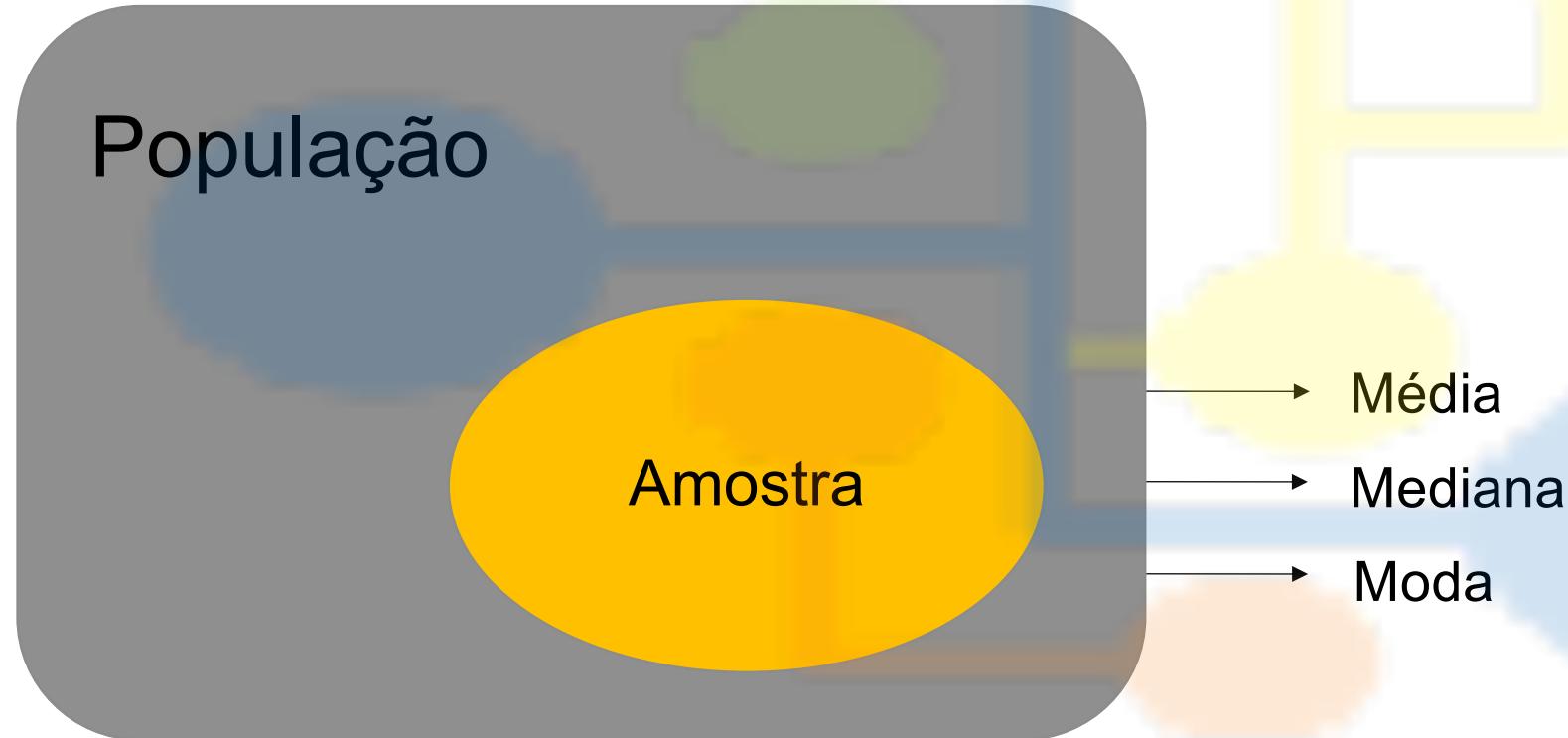
Medidas de Posição



Data Science
Academy

Medidas de Posição

As medidas de posição são valores que representam a tendência de concentração dos dados observados. As mais importantes são as medidas de tendência central. As três medidas de tendência central mais utilizadas são: média aritmética, moda e mediana.



Medidas de Posição

Média



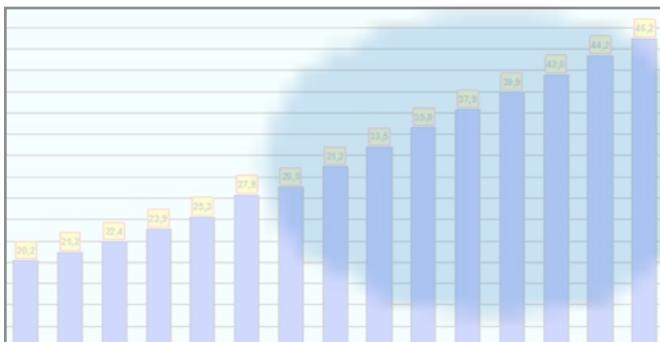
Sem dúvida, **médias** são as formas mais simples de identificar tendências em um conjunto de dados.

A média de uma variável é a soma de todos os registros de uma variável dividido pelo número de registros.

As médias são usadas em diversos testes estatísticos.

Medidas de Posição

Mediana



Definimos a **Mediana** de um conjunto de dados como o valor que divide um conjunto de dados (ordenados) em duas partes com a mesma quantidade de dados.

Se o número de elementos n na amostra for ímpar, a Mediana será: $(n + 1) / 2$

Se o número de elementos n na amostra for par, a Mediana será: $(n / 2) + 1$

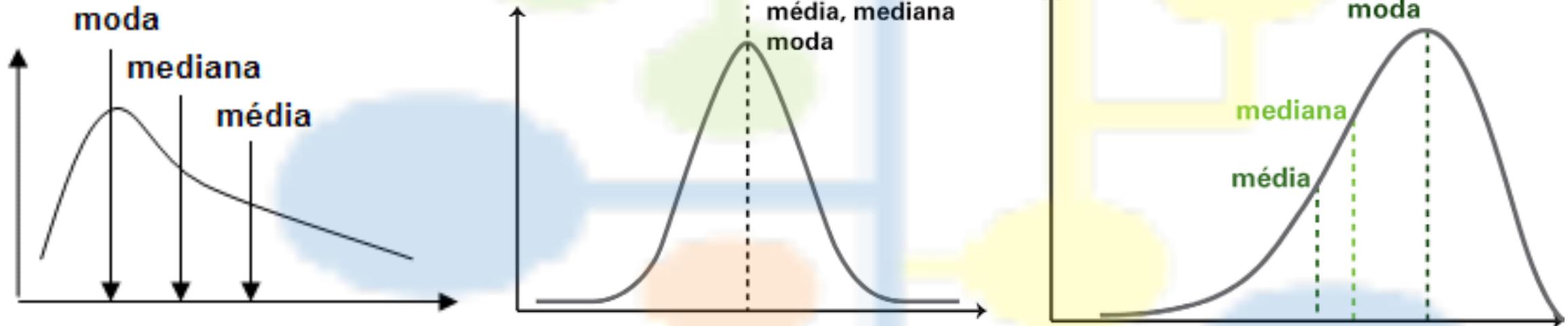
Medidas de Posição

Moda

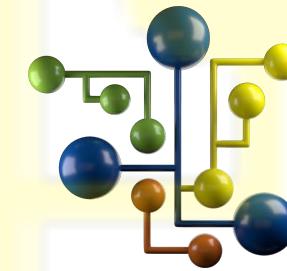


É o valor de maior frequência na amostra, ou seja, o que aparece mais frequentemente no conjunto de dados.

Medidas de Posição



Medidas de Dispersão



Data Science
Academy

Medidas de Dispersão

Uma maneira de descrever um conjunto de dados é através de **medidas de dispersão**. Elas descrevem a amplitude dos dados, ou seja, quanto espalhados os dados estão dentro de um conjunto.

As duas principais medidas de dispersão:
Variância e Desvio Padrão.

Medidas de Dispersão

A **variância** mede a amplitude (variabilidade) dos dados em relação à média.

Medidas de Dispersão

O **desvio padrão** é usado para medir a variabilidade entre os números em um conjunto de dados. Assim como o nome sugere, o desvio padrão é um padrão de desvio (distância) da média.

Em termos bem simples, o **desvio padrão** é a **distância média, da média**.

Medidas de Dispersão

**Exemplo de Uso das Medidas de Dispersão
Analizando as Notas de Julio e Fernanda**



Medidas de Dispersão

Julio – cursa 6 disciplinas na faculdade de Estatística e obteve as seguintes notas e média no exame final:

Disciplinas	Notas
Disciplina 1	100
Disciplina 2	100
Disciplina 3	100
Disciplina 4	50
Disciplina 5	50
Disciplina 6	50

Média Final = 75

Medidas de Dispersão

Fernanda – também cursa 6 disciplinas na faculdade de Estatística e obteve as seguintes notas e média no exame final:

Disciplinas	Notas
Disciplina 1	75
Disciplina 2	74
Disciplina 3	76
Disciplina 4	77
Disciplina 5	75
Disciplina 6	74

Média Final = 75

Medidas de Dispersão

Julio e Fernanda tiveram a mesma média final, mas será que o aprendizado foi similar?

Considerando que a nota de aprovação de cada disciplina seja 70, o que podemos concluir?

Como diferenciar as duas distribuições da variável nota?



Medidas de Dispersão

Julio e Fernanda tiveram a mesma média final, mas será que o aprendizado foi similar?

Considerando que a nota de aprovação de cada disciplina seja 70, o que podemos concluir?

Como diferenciar as duas distribuições da variável nota?

**Calculando a Variância e
o Desvio Padrão**

Medidas de Dispersão

Resultados do Julio:	
Média da Amostra =	75
Variância da Amostra =	750
Desvio Padrão da Amostra =	27.39

A análise das notas do Julio revelam:

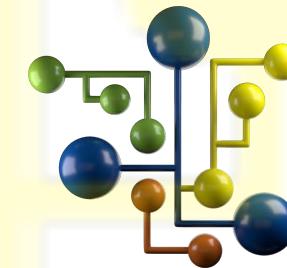
- Comportamentos diferentes de estudo.
- Maior variabilidade nas notas.
- O desvio padrão foi alto, entre 27 Pontos para mais ou para menos.
- Terá de cursar 3 disciplinas novamente para receber o diploma de graduação.

Resultados da Fernanda:	
Média da Amostra =	75
Variância da Amostra =	1.37
Desvio Padrão da Amostra =	1.17

A análise das notas da Fernanda revelam:

- Comportamentos similares de estudo.
- Manteve a variabilidade baixa com notas mais uniformes.
- O desvio padrão foi baixo, entre 1,17 Pontos para mais ou para menos.
- Receberá o diploma de graduação.

Coeficiente de Variação



Data Science
Academy

Coeficiente de Variação

O **coeficiente de variação** (CV) mede o desvio padrão em termos de percentual da média.

- Um CV alto indica alta variabilidade dos dados, ou seja, menos consistência dos dados.
- Um CV menor indica mais consistência dentro do conjunto de dados.

Coeficiente de Variação

Quando comparamos a consistência entre 2 conjuntos de dados em relação a suas médias, o ideal é utilizar o **coeficiente de variação**.

Coeficiente de Variação

Como calculamos o Coeficiente de Variação (CV):

$$CV = \frac{s}{x} \times 100$$

Onde: **S** = Desvio Padrão
X = Média

Coeficiente de Variação



Valor Médio da
Ação na Bolsa
de Valores \$
55.62
Desvio Padrão
\$ 5.10



Valor Médio da
Ação na Bolsa
de Valores \$
24.86
Desvio Padrão
\$ 3.60

Coeficiente de Variação

Os Seguintes Dados Foram Coletados:



$$\text{Nike} \rightarrow CV = S / X (100) = \$5.10 / \$55.62 (100) = 9.2\%$$



$$\text{Adidas} \rightarrow CV = S / X (100) = \$3.60 / \$24.86 (100) = 14.5\%$$

Coeficiente de Variação



$$\text{Nike} \rightarrow CV = S / X (100) = \$5.10 / \$55.62 (100) = 9.2\%$$



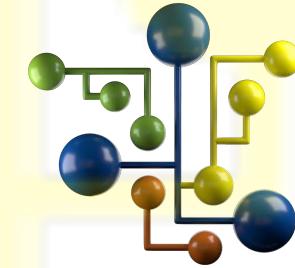
$$\text{Adidas} \rightarrow CV = S / X (100) = \$3.60 / \$24.86 (100) = 14.5\%$$

Coeficiente de Variação

Um investidor se sentiria mais seguro em adquirir ações da **Nike**, pois o preço das ações teria uma variação menor, podendo assim evitar perdas e permitindo ao investidor ter um investimento mais seguro.



Medidas de Posição Relativa



Data Science
Academy

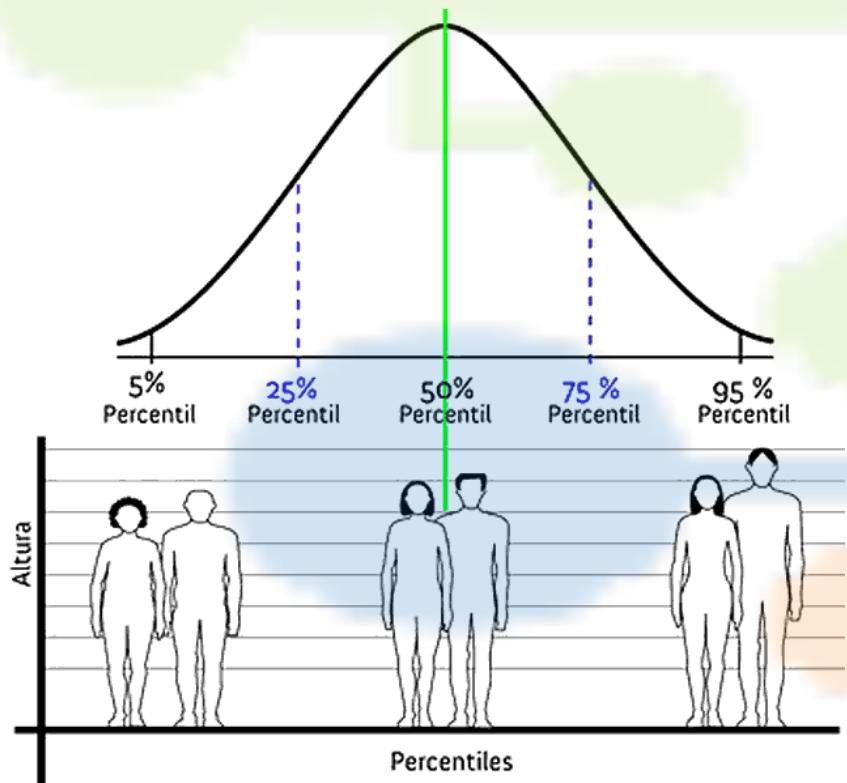
Medidas de Posição Relativa

Os dados podem ser medidos em termos de posição relativa, que compara a posição de um valor, em relação a outro valor dentro do conjunto de dados.

Medidas de Posição Relativa

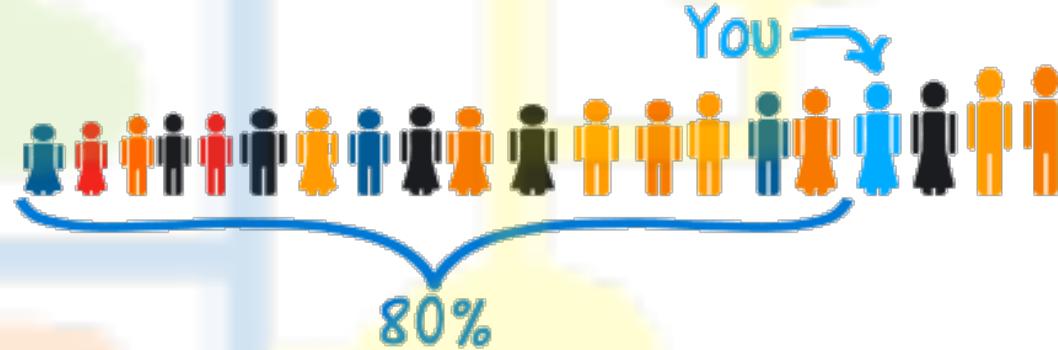
Percentil e quartil são as medidas mais comuns de posição relativa.

Medidas de Posição Relativa



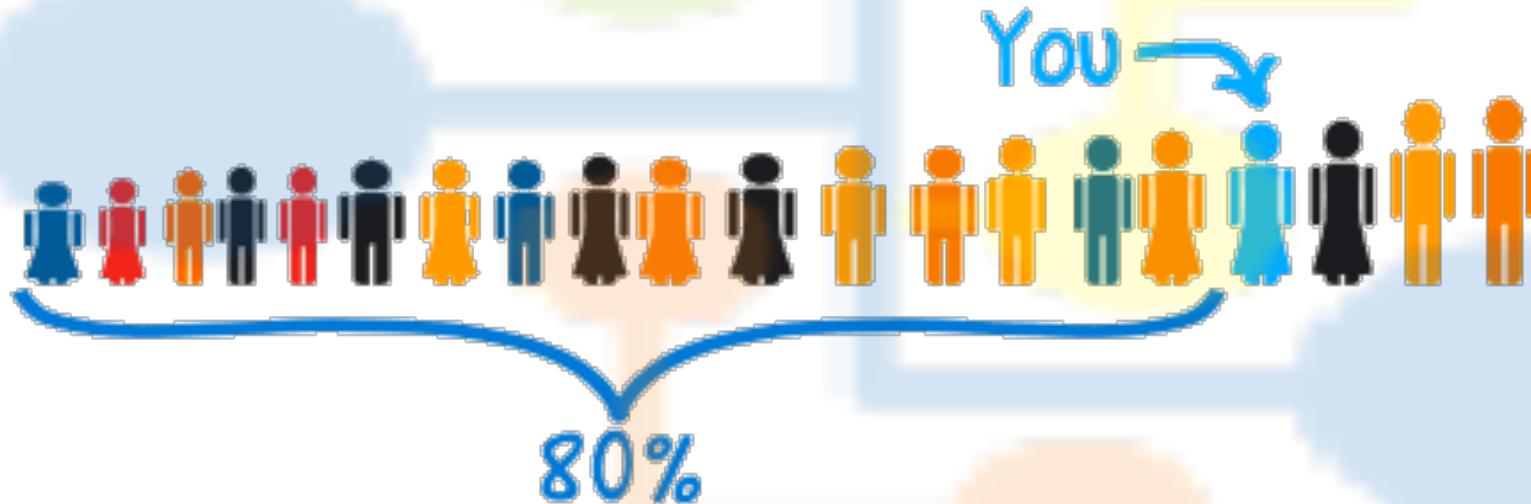
Medidas de Posição Relativa

Percentil



Medidas de Posição Relativa

A maneira mais fácil de informar a posição relativa é por meio do uso do **percentil**.



Medidas de Posição Relativa

Percentil e Porcentagem são a mesma coisa?

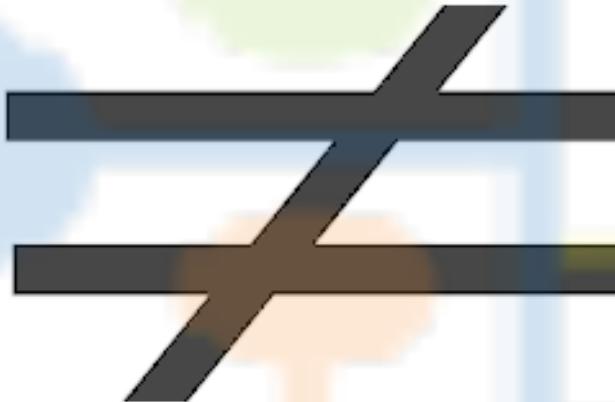


Medidas de Posição Relativa

Percentil e Porcentagem **não** são a mesma coisa.

Percentil

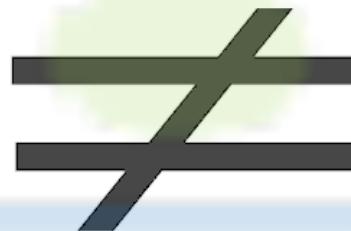
Porcentagem



Medidas de Posição Relativa

Percentil e Porcentagem **não** são a mesma coisa.

Percentil



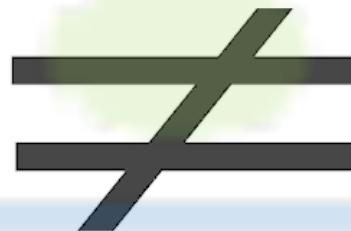
Porcentagem

Porcentagem (%): Proporção calculada em relação a uma grandeza de cem unidades. A porcentagem pode ser encontrada multiplicando o valor numérico por 100.

Medidas de Posição Relativa

Percentil e Porcentagem não são a mesma coisa.

Percentil

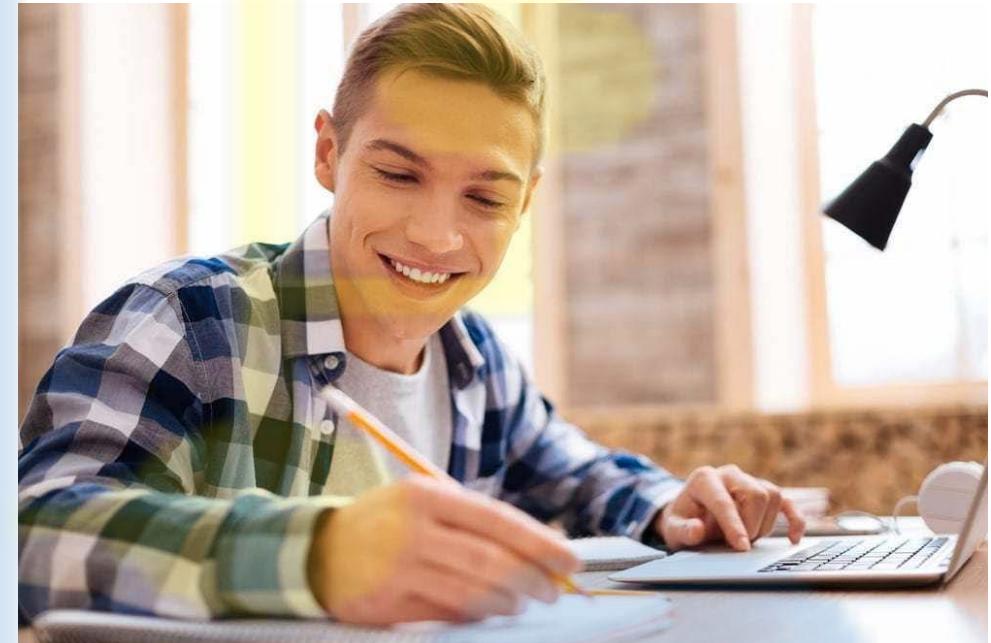


Porcentagem

Percentil: É o ponto da distribuição dos resultados ordenados da amostra (por ordem crescente dos dados) em 100 partes de igual amplitude. Por exemplo, um resultado no percentil 90 significa que 90% dos resultados se situam nesse ponto ou abaixo dele.

Medidas de Posição Relativa

Suponha que um aluno tenha conseguido nota **36** em um exame de admissão em uma universidade, cujo valor máximo era **45**.



Medidas de Posição Relativa

Supondo que além de informar a você que o aluno conseguiu nota 36, eu dissesse que ele ficou em

97º percentil

Isso significa que o aluno foi melhor que **97%** dos outros alunos que prestaram o mesmo exame!



Medidas de Posição Relativa

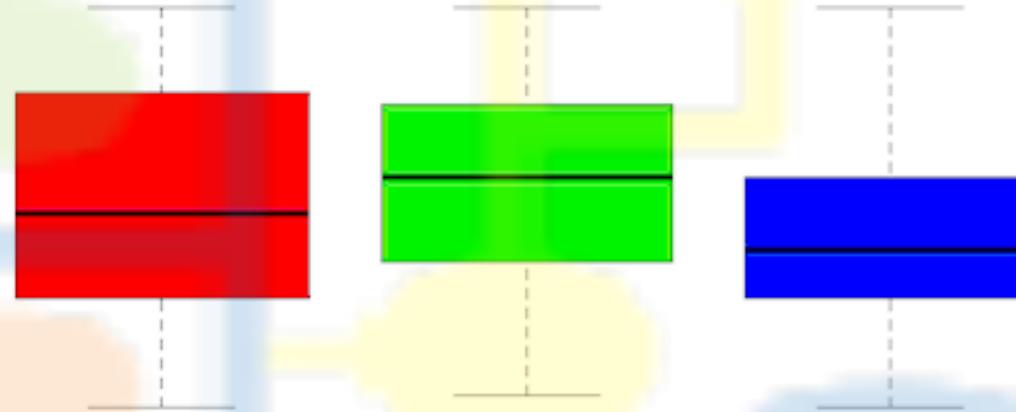
Perceba que se dividirmos **36/45**, o aluno teve um aproveitamento de **80%**.

Esta informação **NÃO** é a mesma coisa que o **percentil**.



Medidas de Posição Relativa

Quartil



Medidas de Posição Relativa

Quartil é simplesmente um percentil específico de interesse.

Medidas de Posição Relativa

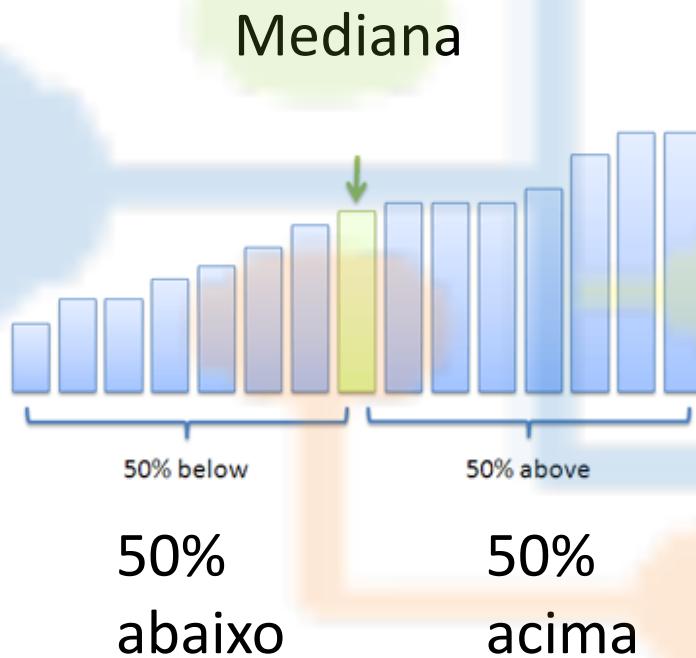
Quartis são valores que dividem uma tabela de dados em quatro partes iguais:

- O **primeiro quartil** é o valor que constitui **25% percentil**.
- O **segundo quartil** é o valor que constitui **50% percentil**.
- O **terceiro quartil** é o valor que constitui **75% percentil**.
- O **quarto quartil** é o valor que constitui **100% percentil**.

Medidas de Posição Relativa

Perceba que o **segundo quartil** é a **mediana**, ou seja,

50º percentil



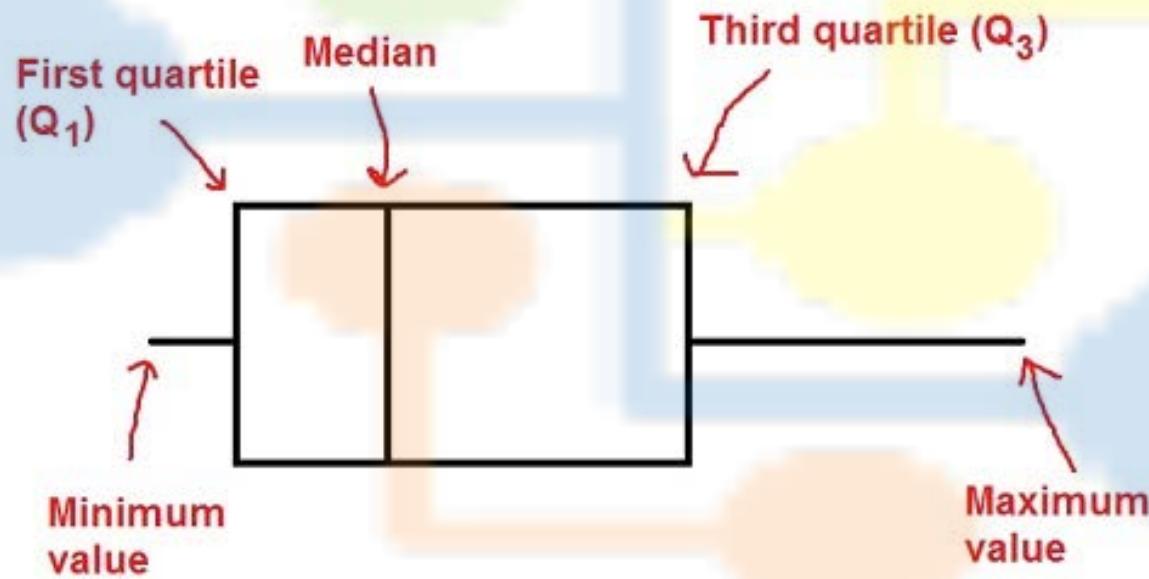
Medidas de Posição Relativa

Temos ainda os intervalos interquartis:

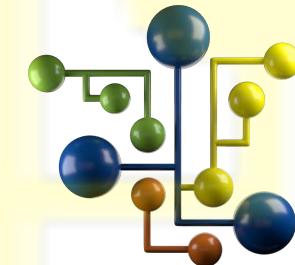
- Intervalo Interquartil $\rightarrow Q_3 - Q_1$
- Intervalo Semi-interquartil $\rightarrow (Q_3 - Q_1) / 2$
- Quartil Médio $\rightarrow (Q_3 + Q_1) / 2$

Medidas de Posição Relativa

Os intervalos interquartis são fundamentais para saber interpretar um boxplot, uma importante ferramenta de análise exploratória dos dados:



Métodos Estatísticos Para Análise de Dados



Data Science
Academy

Métodos Estatísticos Para Análise de Dados



Métodos Gráficos ou
Tabulares

$$CV = \frac{s}{x} \times 100$$

Métodos
Numéricos

Métodos Estatísticos Para Análise de Dados

- Tabela de Frequência
- Tabela de Contingência
- Gráficos de Linhas
- Gráficos de Barras
- Gráfico de Pareto
- Histogramas
- Gráficos de Caixa (boxplots)
- Diagramas de dispersão
- Gráfico Temporal
- Ogiva (frequência cumulativa)
- Ramo e folhas
- Gráficos de Pontos
- Gráfico de Quartis

Métodos Gráficos ou Tabulares



Métodos Estatísticos Para Análise de Dados

Métodos Numéricos

$$CV = \frac{s}{x} \times 100$$

- Média
- Mediana
- Moda
- Quartis
- Desvio Padrão
- Variância
- Intervalo Interquartil
- Coeficiente de Variação
- Coeficiente de Assimetria
- Curtose
- Coeficiente de Correlação Linear
- Covariância
- Coeficientes de Associação

Métodos Estatísticos Para Análise de Dados

A técnica utilizada vai depender do tipo de variável



Métodos Estatísticos Para Análise de Dados

Nos manuais em pdf ao final deste capítulo você vai encontrar exemplos de como explorar alguns desses métodos estatísticos para análise de dados.

Pratique-os no Power BI.

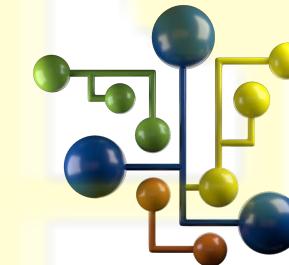
Linguagem Python

x

Linguagem R

x

DAX



Data Science
Academy

Linguagem Python x Linguagem R x DAX

Análise de
Dados com
Power BI



Análise de
Dados com
Power BI e
DAX



Análise de
Dados com
Power BI, DAX
e R



Análise de
Dados com
Power BI, DAX,
R e Python



Linguagem Python x Linguagem R x DAX

O Power BI oferece funcionalidades básicas de manipulação de dados e criação de gráficos.

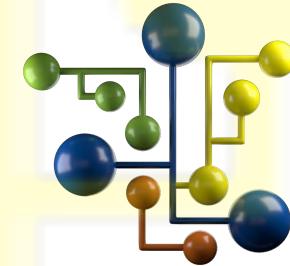
Ao usar o DAX, conseguimos customizar o processo de análise de dados e criar regras de negócio avançadas.

Com Linguagem R trazemos para o Power BI uma poderosa plataforma de análise estatística, além de gráficos avançados.

Com a Linguagem Python temos manipulação de dados, modelos de Machine Learning e IA e gráficos avançados.

Mini-Projeto 3

Você Está Preparado Para Isso?
Tem Certeza?



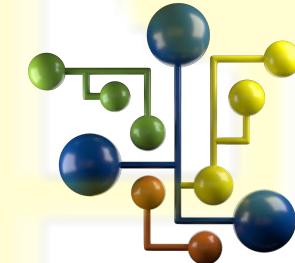
Data Science
Academy

Mini-Projeto 3 - Você Está Preparado Para Isso? Tem Certeza?



Cena do Filme Matrix - 1999

O Que é o Apache Spark?



Data Science
Academy

O Que é o Apache Spark?



O Que é o Apache Spark?



Um cluster de computadores permite usar de forma conjunta a capacidade computacional de diversas máquinas aumentando o poder de armazenamento e processamento.

Mas aí temos um problema!

Como gerenciar o armazenamento e processamento distribuído através de diversos computadores?

O Que é o Apache Spark?



Para gerenciar o armazenamento distribuído podemos usar o **Apache Hadoop HDFS** (Hadoop Distributed File System).

Para gerenciar o processamento distribuído podemos usar o **Apache Spark**.

O Que é o Apache Spark?

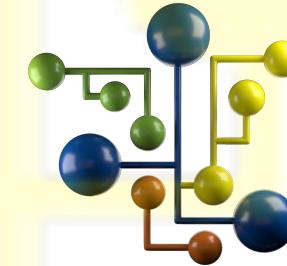
O Apache Spark é um framework para processamento distribuído que possui diversos módulos para diferentes tarefas em Ciência de Dados.

Se precisamos processar grandes volumes de dados, de forma distribuída através de um cluster de computadores e com alta velocidade, o Apache Spark tem se mostrado uma das melhores opções.

O Apache Spark é open-source.

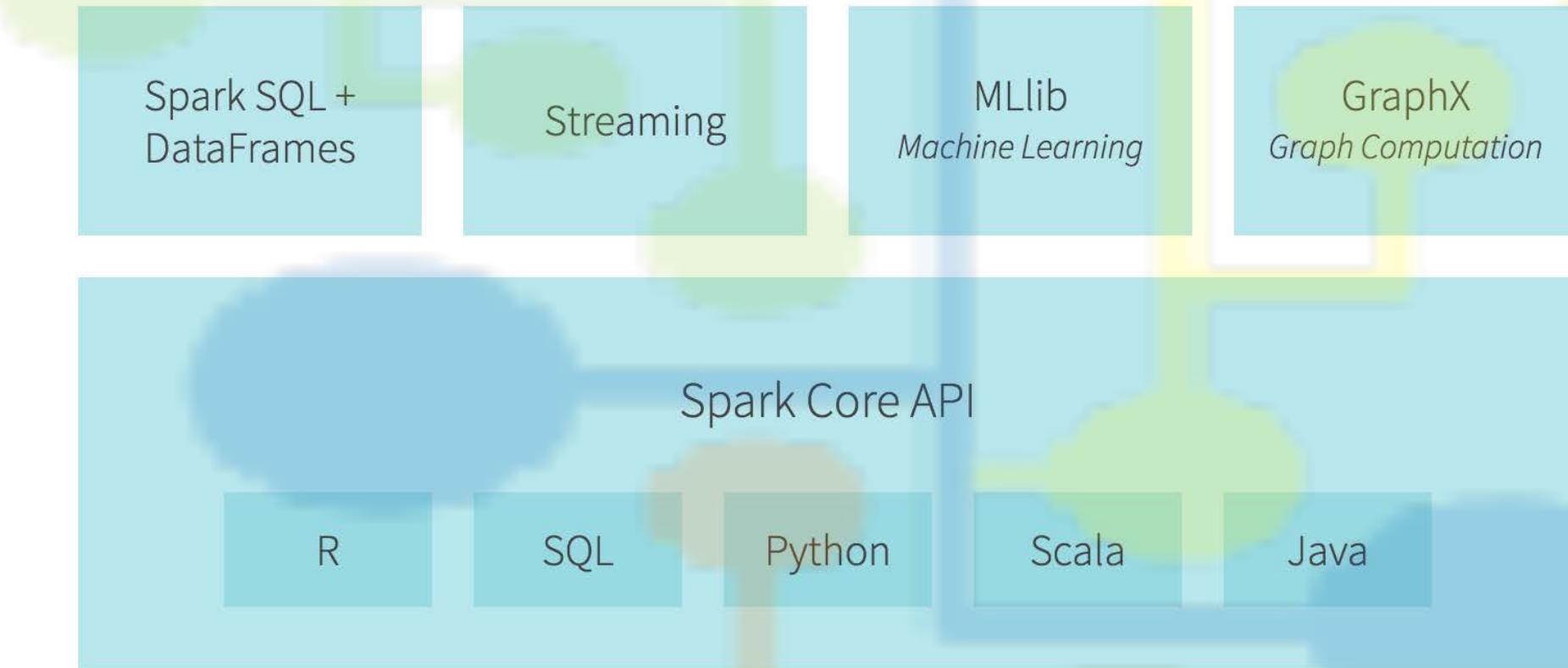


Apache Spark Framework

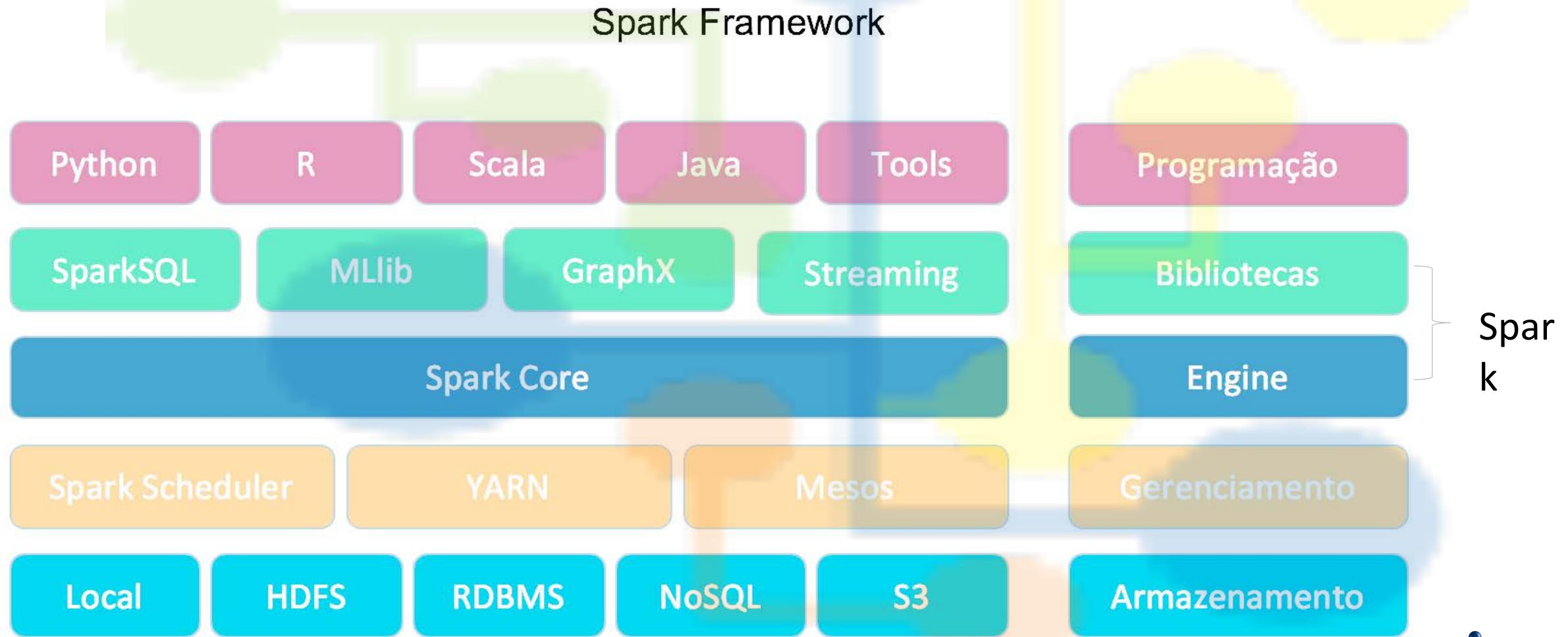


Data Science
Academy

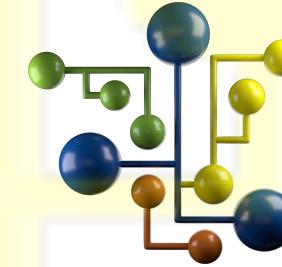
Apache Spark Framework



Apache Spark Framework



Afinal, Por Que Precisamos de Machine Learning?

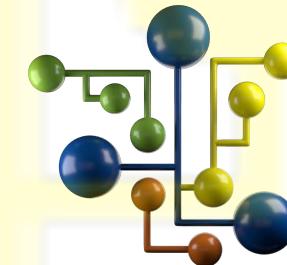


Data Science
Academy

Afinal, Por Que Precisamos de Machine Learning?

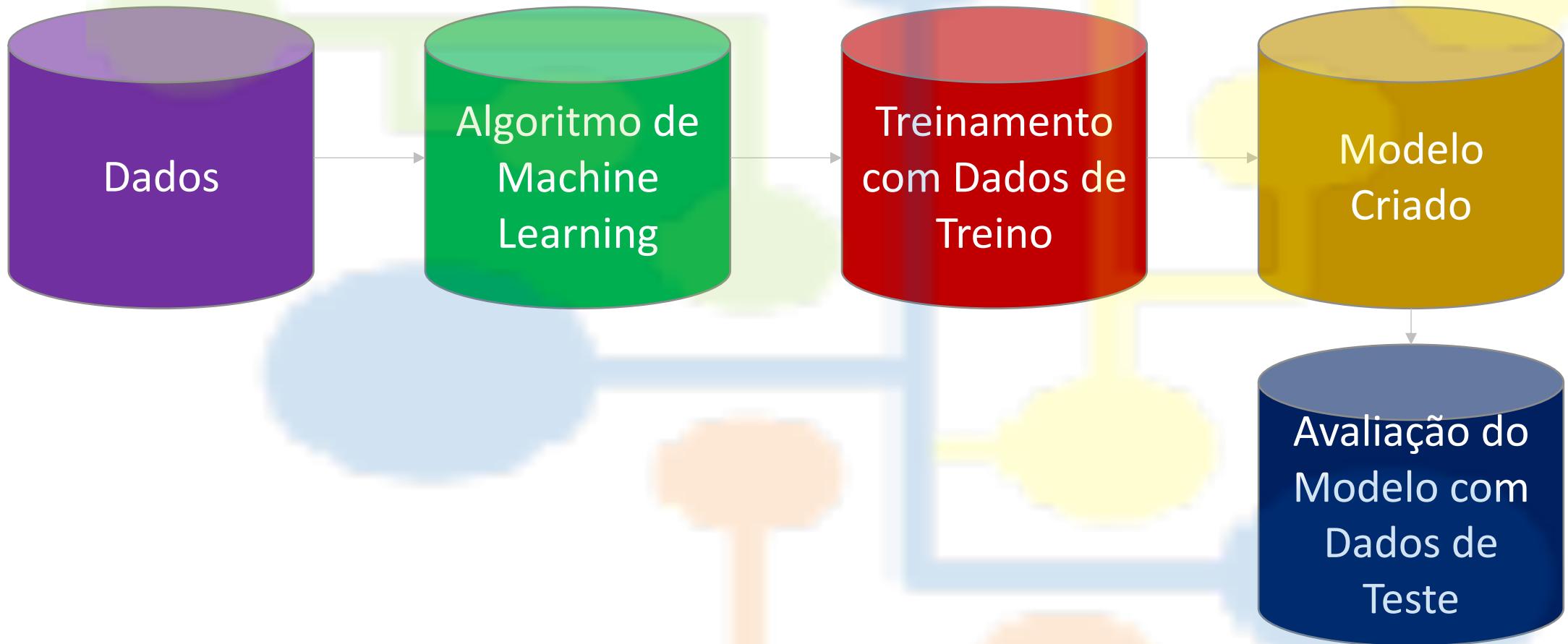
Variáveis Preditoras (Recursos, Atributos, Características, Variáveis Independentes)					Variável Alvo (Target, Variável Dependente)
Idade	Gênero	Saldo no Cartão	Educação	Valor a Pagar	Pagou a Fatura?
34	F	2.000	Superior	350	Não
27	M	3.500	Superior	400	Sim
41	M	5.600	Superior	1.500	Não
39	F	1.200	Ensino Médio	120	Sim

Tipos de Aprendizado em Machine Learning



Data Science
Academy

Tipos de Aprendizado em Machine Learning



Tipos de Aprendizado em Machine Learning

Aprendizado
Supervisionado
Regressão

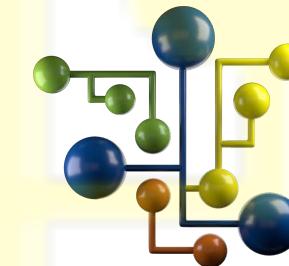
Aprendizado
Supervisionado
Classificação

Aprendizado
Não
Supervisionado
Clusterização

Aprendizado
Por Reforço

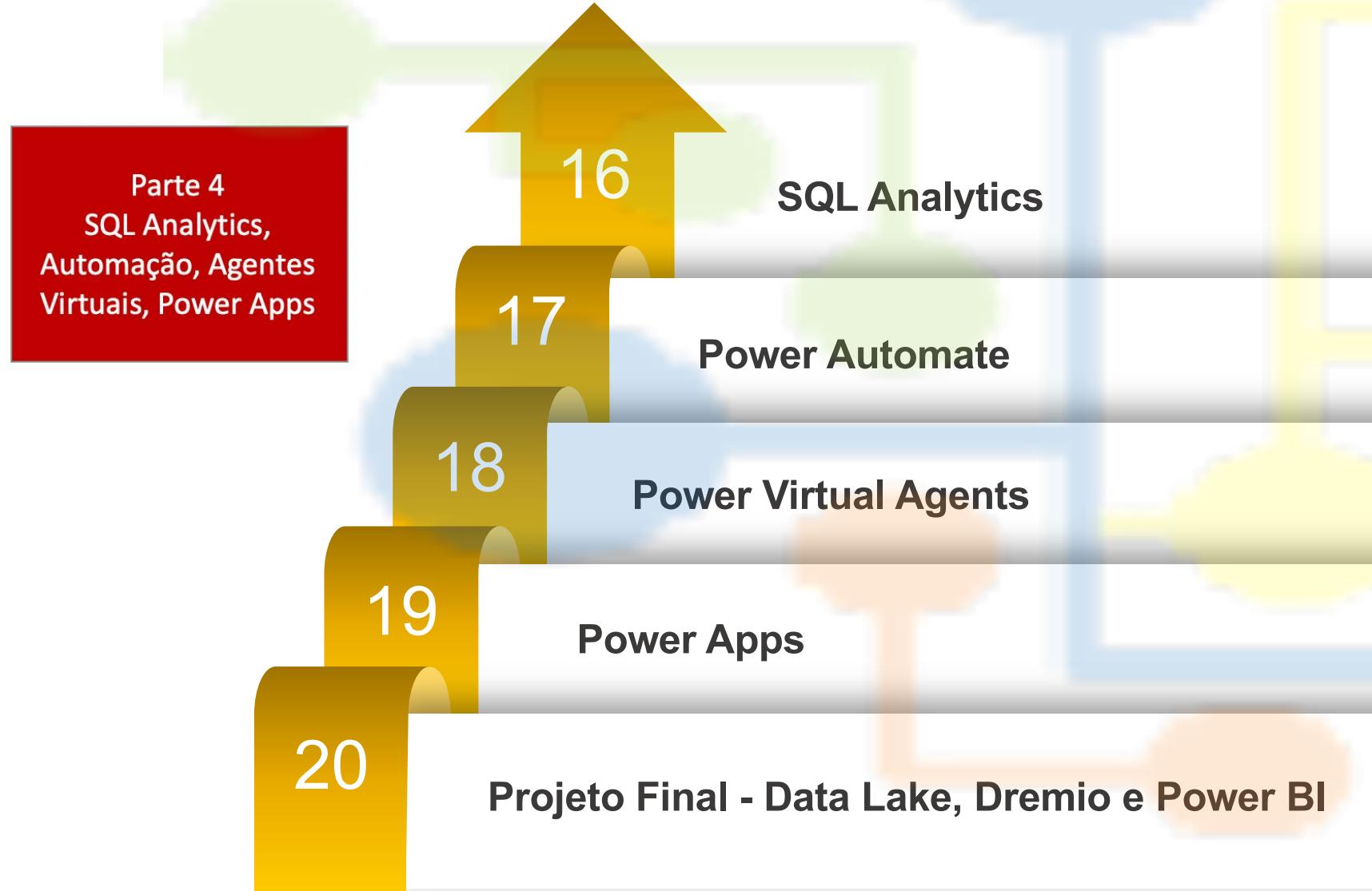


SQL Analytics

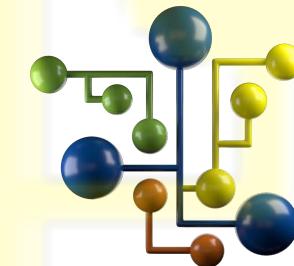


Data Science
Academy

Introdução



O Que é SQL Analytics?



Data Science
Academy

O Que é SQL Analytics?



The background of the slide features a white surface with several semi-transparent, rounded rectangular shapes in various colors: light green, light blue, light yellow, and light orange. Superimposed on these shapes are faint icons related to data and technology, such as a gear, a bar chart, a database symbol, and a network connection.

SQL

Structured Query Language,

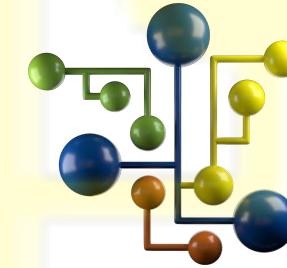
ou

Linguagem de Consulta Estruturada

O Que é SQL Analytics?

SQL Analytics é a utilização de Linguagem SQL para manipulação, transformação e análise de dados.

Introdução à Linguagem SQL



Data Science
Academy

Introdução à Linguagem SQL

Linguagem SQL

Instruções DQL
(Data Query Language)

Instruções DML
(Data Manipulation Language)

Instruções DDL
(Data Definition Language)

Instruções DCL
(Data Control Language)

Instruções DTL
(Data Transaction Language)

Select

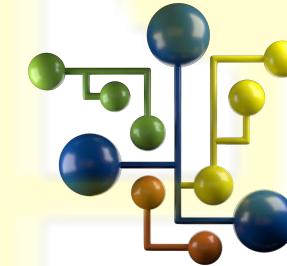
Insert
Update
Delete

Create
Alter
Drop

Grant
Revoke

Commit
Rollback

Microsoft Power Platform



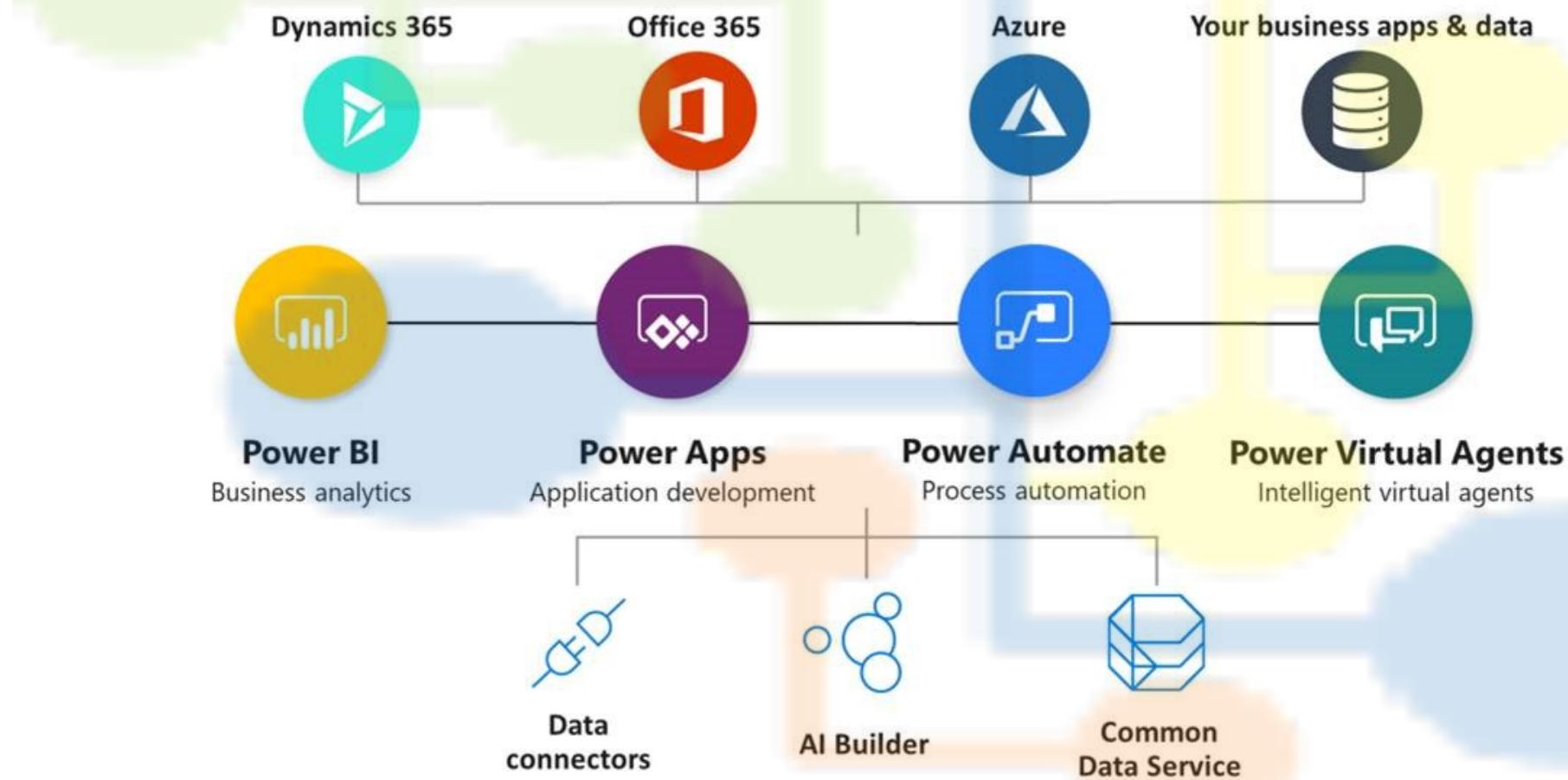
Data Science
Academy

Microsoft Power Platform

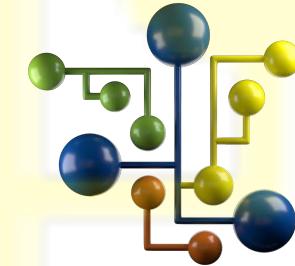


Low Code

Microsoft Power Platform

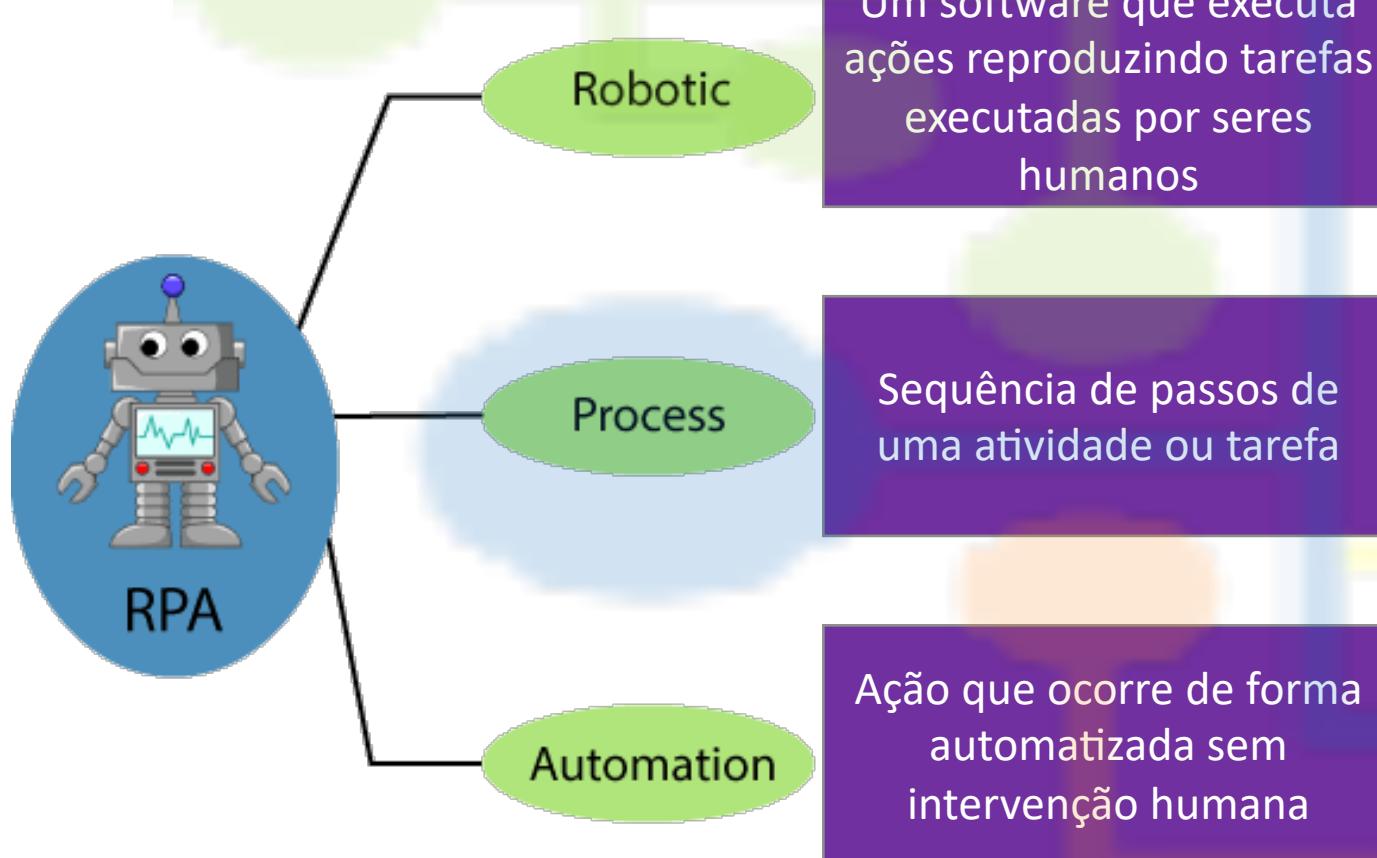


O Que é Automação Robótica de Processos?



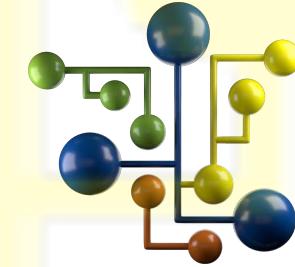
Data Science
Academy

O Que é Automação Robótica de Processos?



Automação Robótica de Processos ou Robotic Process Automation (RPA) é um conjunto de ações cujo objetivo é automatizar processos de negócio.

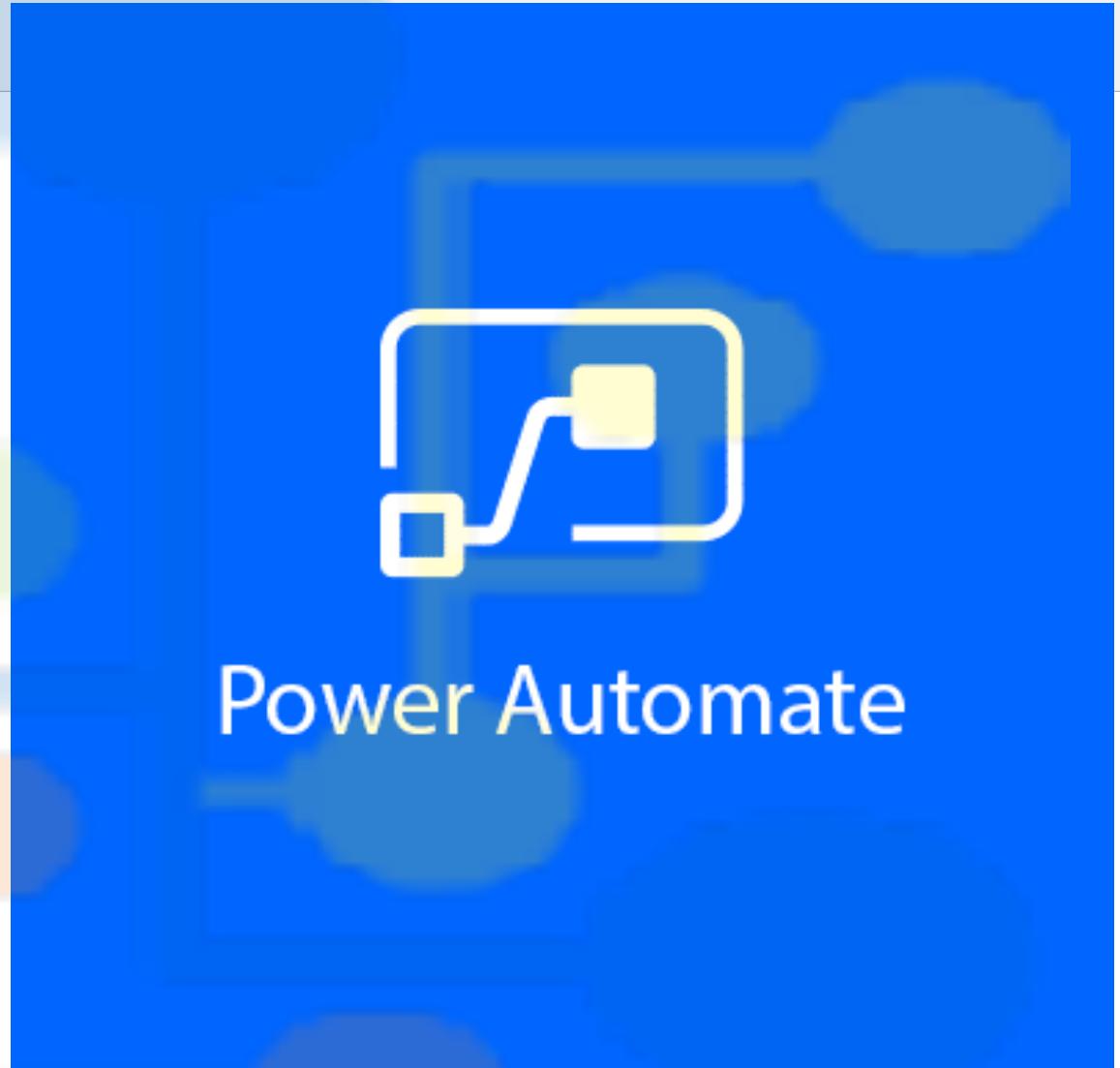
O Que é o Microsoft Power Automate?



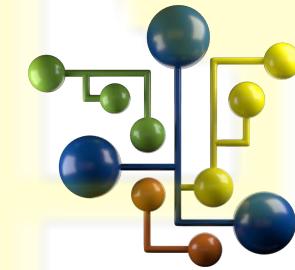
Data Science
Academy

O Que é o Microsoft Power Automate?

- O Power Automate é uma plataforma de desenvolvimento de “baixo código” (low code) para a construção de fluxos de trabalho automatizados.
- O desenvolvimento é conduzido por uma interface de usuário gráfica simples e intuitiva que vem com centenas de conectores pré-construídos e milhares de modelos (templates) prontos para uso.
- Esses conectores, ações e modelos permitem que o desenvolvedor cidadão (Citizen Developer) se conecte facilmente a aplicativos e serviços populares durante o processo de desenvolvimento do fluxo de trabalho automatizado.

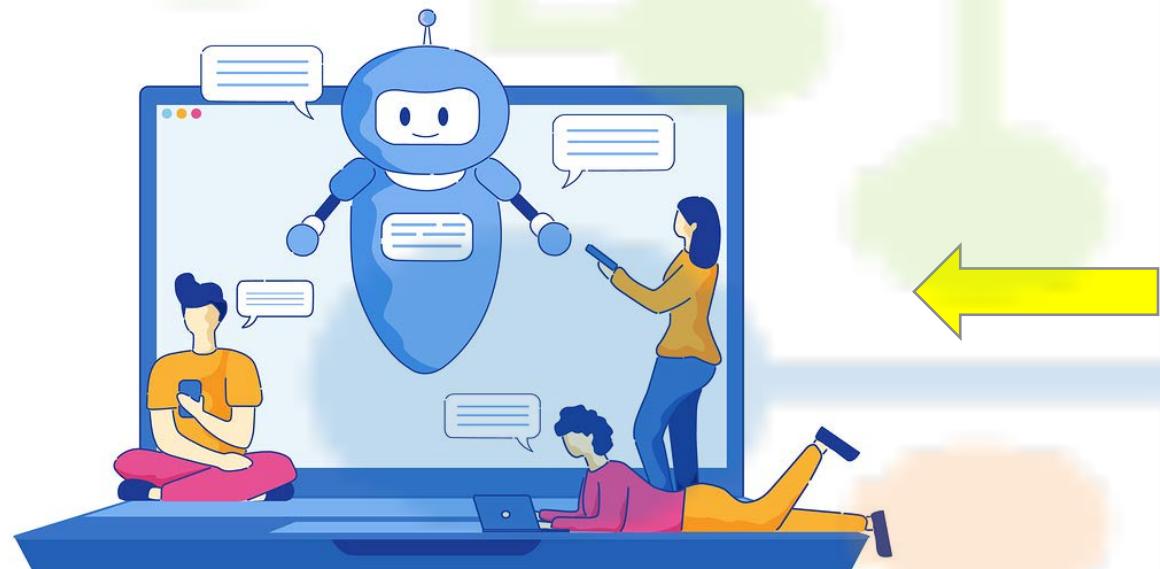


O Que é um Chatbot?



Data Science
Academy

O Que é um Chatbot?



Isso não é um chatbot!

O Que é um Chatbot?

```
31     def __init__(self, path=None, debug=False):
32         self.file = None
33         self.fingerprints = set()
34         self.logdups = True
35         self.debug = debug
36         self.logger = logging.getLogger(__name__)
37         if path:
38             self.file = open(os.path.join(path, 'seen.txt'), 'a')
39             self.file.seek(0)
40             self.fingerprints.update(self.file.read().split())
41
42     @classmethod
43     def from_settings(cls, settings):
44         debug = settings.getboolean('SUPERVISOR_DEBUG')
45         return cls(job_dir(settings), debug)
46
47     def request_seen(self, request):
48         fp = self.request_fingerprint(request)
49         if fp in self.fingerprints:
50             return True
51         self.fingerprints.add(fp)
52         if self.file:
53             self.file.write(fp + os.linesep)
54
55     def request_fingerprint(self, request):
56         return request_fingerprint(request)
```



Isso é um chatbot!

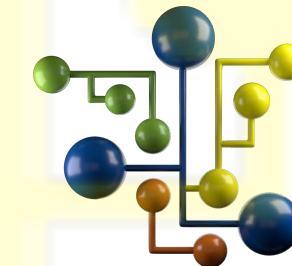
O Que é um Chatbot?

De acordo com Dicionário Oxford, um chatbot é:

“Um programa de computador projetado para simular conversas com usuários humanos, especialmente pela Internet.”

É um assistente que se comunica conosco por meio de mensagens de texto, um agente virtual que se integra a web sites, aplicativos ou mensageiros instantâneos e ajuda os empresários a se aproximarem dos clientes. O chatbot é um sistema automatizado de comunicação com os usuários.

Business Applications

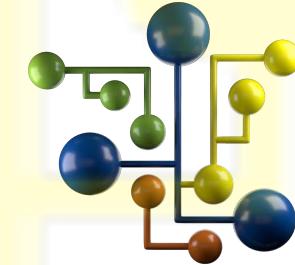


Data Science
Academy

Business Applications



Low Code é Realmente Vantajoso Para Uma Empresa?



Data Science
Academy

Low Code é Realmente Vantajoso Para Uma Empresa?



Vantagens do Low Code:

- 1- Velocidade
- 2- Agilidade
- 3- Desenvolvimento “Multiexperience”
- 4- Inovação
- 5- Redução do Envolvimento da TI

Low Code é Realmente Vantajoso Para Uma Empresa?



Desvantagens do Low Code:

- 1- Dificuldade em atender requerimentos de negócio
- 2- Não pode ser usado em aplicações mais complexas
- 3- Limitado em termos de segurança e flexibilidade
- 4- Menor controle sobre execução das apps
- 5- Apps simples e menos sofisticadas

Low Code é Realmente Vantajoso Para Uma Empresa?

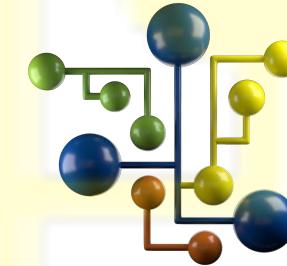


Low Code é Realmente Vantajoso Para Uma Empresa?

Para pequenas e (algumas) médias empresas o uso de soluções baseadas em low code pode sim trazer benefícios, principalmente em termos de velocidade e redução de custos.

Para (algumas) médias e grandes empresas o uso de soluções low code ainda não seria capaz de atender os requerimentos e complexidade dos processos de negócio, além das questões de compliance e segurança.

Encerramento



Data Science
Academy



Encerramento

Promessas feitas no início do curso de Power BI:

Promessa	Status
Curso prático com exemplos claros e objetivos	
Material de qualidade e excelência	✓
Grande quantidade de conteúdo relevante	✓
Diferentes áreas de estudo	✓
BI e Data Science	✓
Integração do Power BI com Linguagens R e Python	✓
Introdução à Microsoft Power Platform	✓
Certificado de conclusão e curso totalmente gratuito	✓

Encerramento

Próximos Passos:

- Quer trabalhar com Análise de Dados, Ciência de Dados e/ou IA?
 - Formação Cientista de Dados
 - Formação Análise Estatística Para Cientistas de Dados
 - Formação Analista de Inteligência de Mercado
 - Formação Desenvolver Microsoft Para Data Science e IA
 - Formação Desenvolver Web Para Data Science
 - Formação Inteligência Artificial
 - Formação Inteligência Artificial Para Medicina

Encerramento

Próximos Passos:

- Quer trabalhar com Engenharia ou Arquitetura de Dados?
 - Formação Engenheiro de Dados
 - Formação Arquiteto de Dados
 - Formação Suporte e Infraestrutura de Big Data

Encerramento

Próximos Passos:

- Quer trabalhar com Automação Robótica de Processos?
 - Formação Desenvolvedor RPA

Encerramento

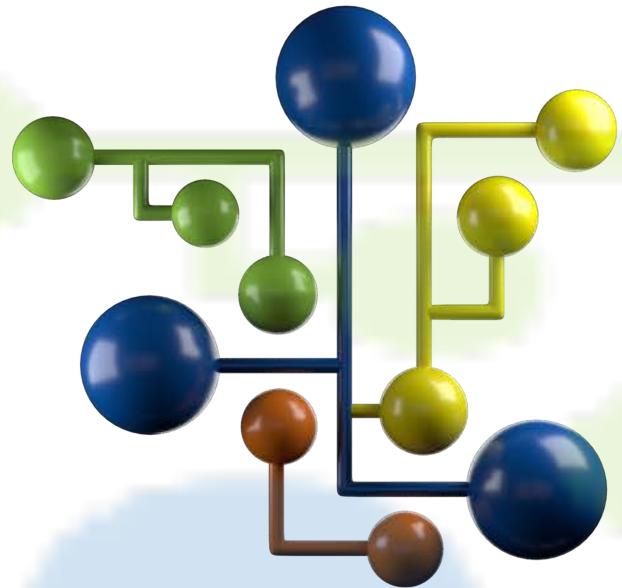
Próximos Passos:

- Quer trabalhar com Blockchain?
 - Formação Engenheiro Blockchain

Encerramento

Para visualizar o catálogo de cursos acesse:

www.datascienceacademy.com.br



Data Science Academy

Muito Obrigado!

Continue Trilhando Uma Excelente
Jornada de Aprendizagem.