

Formação Cientista de Dados

Módulo 1

www.datascienceacademy.com.br



Data Science Academy

Formação Cientista de Dados



Seja Bem-Vindo

Data Science Academy

www.datascienceacademy.com.br

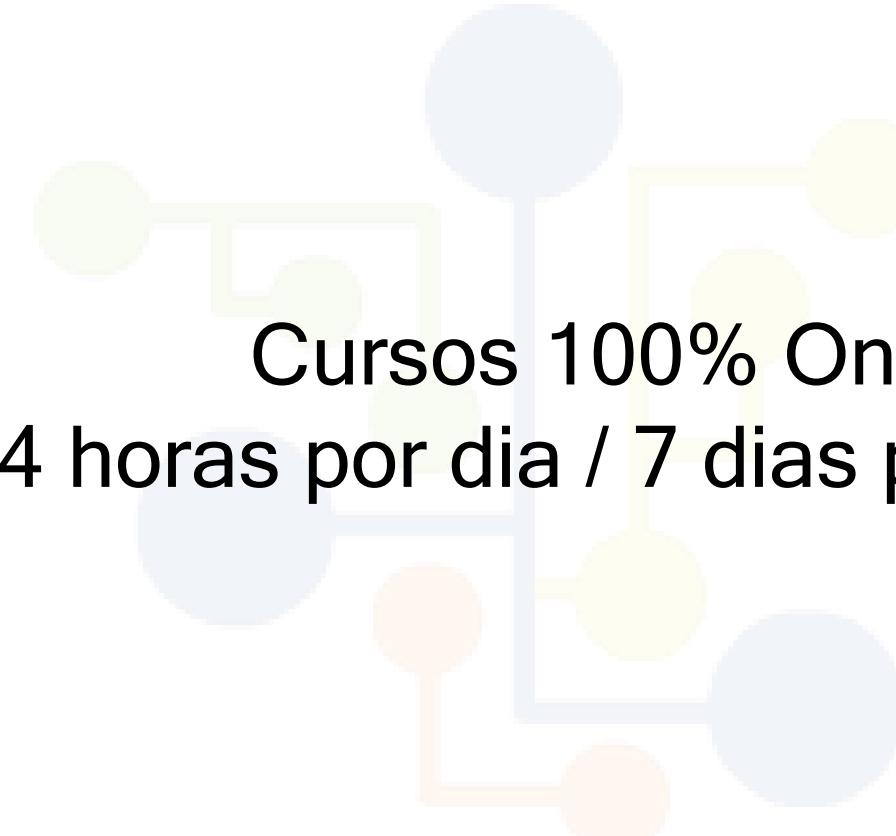
Formação Cientista de Dados

O que é a Formação Cientista de Dados?



Data Science Academy

Formação Cientista de Dados



Cursos 100% Online
24 horas por dia / 7 dias por semana



Data Science Academy

Formação Cientista de Dados

A Formação é composta de 6 cursos



Data Science Academy

Formação Cientista de Dados

Formação composta de 6 cursos

- Big Data Analytics com R e Microsoft Azure Machine Learning
- Big Data Real-Time Analytics com Python e Spark
- Engenharia de Dados com Hadoop
- Machine Learning
- Business Analytics
- Visualização de Dados e Construção de Dashboards



Data Science Academy

Formação Cientista de Dados

Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science Academy

Formação Cientista de Dados

Big Data Analytics com R e Microsoft Azure Machine Learning

Linguagem
R
Capítulos
2, 3 , 4, e 5

Estatística
Capítulo
6

Machine
Learning
Capítulo
7

Azure
Machine
Learning
Capítulos
8, 9, 10 e 11

Data Science Academy



Formação Cientista de Dados

Big Data Real-Time Analytics com Python e Spark



Data Science Academy

Formação Cientista de Dados

Big Data Real-Time Analytics com Python e Spark

Análise de
Dados com
Python
Capítulos
2 e 3

Estatística
Capítulo
4

Machine
Learning
Capítulo
5

Apache
Spark
Capítulos
6, 7, 8 e 9

Data Science Academy

Formação Cientista de Dados

Engenharia de Dados com Hadoop



Data Science Academy

Formação Cientista de Dados

Engenharia de Dados com Hadoop

Cluster
Hadoop

Capítulos
2, 3 e 4

Armazena-
mento de
Dados

Capítulos
5, 6 e 7

Machine
Learning

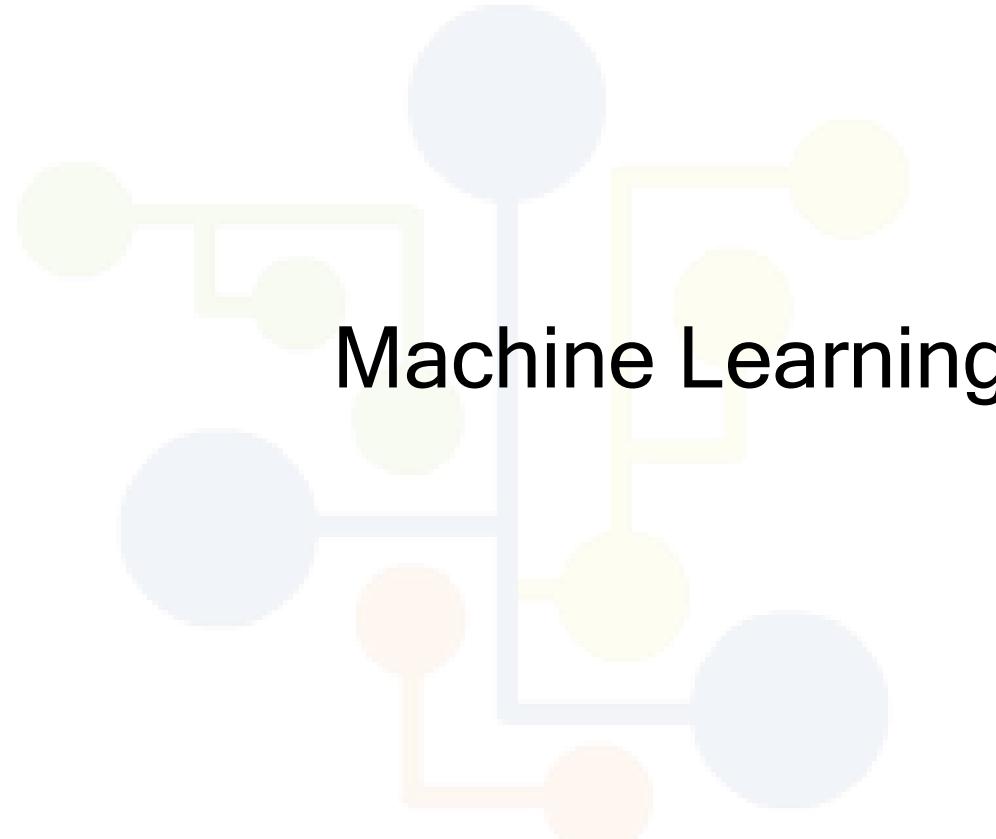
Capítulo
8

Hadoop e
Spark

Capítulo
9



Formação Cientista de Dados



Data Science Academy

Formação Cientista de Dados

Machine Learning



Data Science Academy

Formação Cientista de Dados

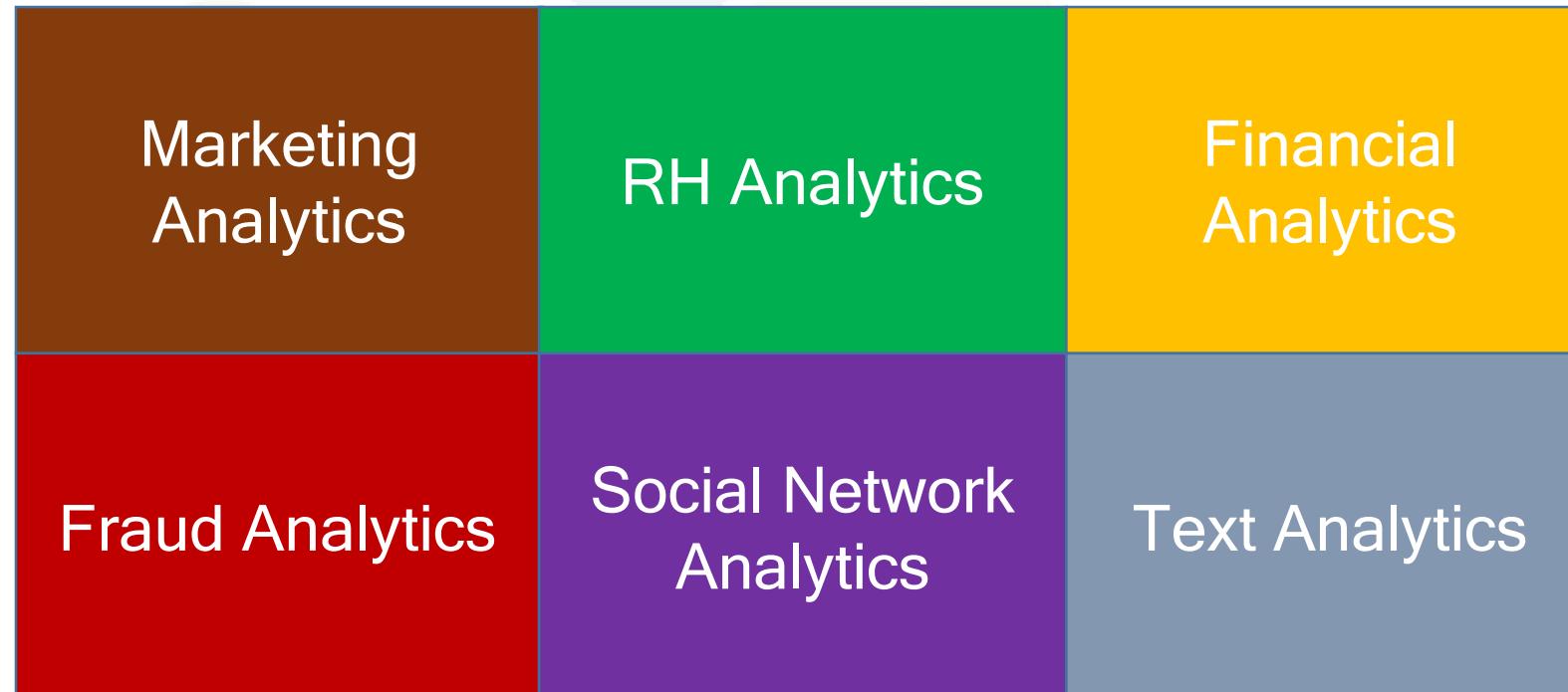
Business Analytics



Data Science Academy

Formação Cientista de Dados

Business Analytics



Data Science Academy

Formação Cientista de Dados

Visualização de Dados e Construção de Dashboards



Data Science Academy

Formação Cientista de Dados

Visualização de Dados e Construção de Dashboards



Data Science Academy

Formação Cientista de Dados

Em todos os cursos



Data Science Academy

Formação Cientista de Dados

Em todos os cursos

- Módulo Introdução (Capítulo 1)
- Projetos (25)
- Projeto Final (1)
- Quizzes e Exercícios
- Avaliação Final (50 questões)
- Fórum Exclusivo



Data Science Academy

Projetos

Projetos



Data Science Academy

Projeto Final

Projeto Final



Data Science Academy

Projeto Final

Projeto Final



Data Science Academy

Certificado

Certificado



Certificado de Conclusão

A Data Science Academy certifica que

Concluiu com êxito o curso de formação on-line:

Com data de finalização em: _____.



www.datascienceacademy.com.br/

Tiago Pereira
CEO



Data Science Academy

Projeto Startup

Projeto Startup



Data Science Academy

Carreira

Preparação para a
Carreira de
Cientista de Dados



Data Science Academy

Recomendação

Pré-requisitos (Recomendação)

- Introdução à Ciência de Dados
- Big Data Fundamentos
- Python Fundamentos para Análise de Dados



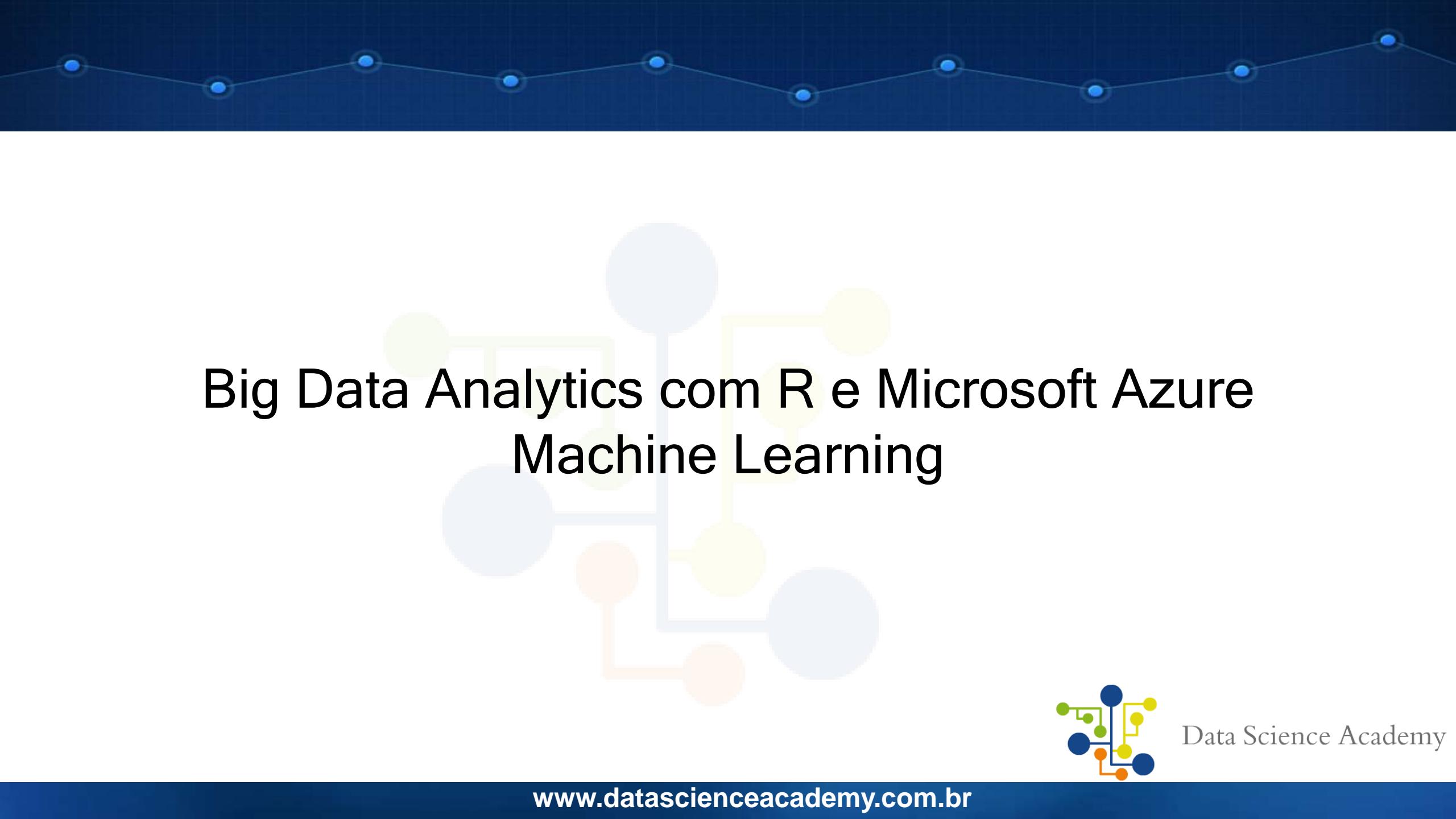
Data Science Academy

Formação Cientista de Dados

Muito Obrigado!



Data Science Academy



Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science Academy

Como este curso está dividido?



Data Science Academy

Como este curso está dividido?

Parte 1
Linguagem R
Capítulos 2, 3, 4, 5

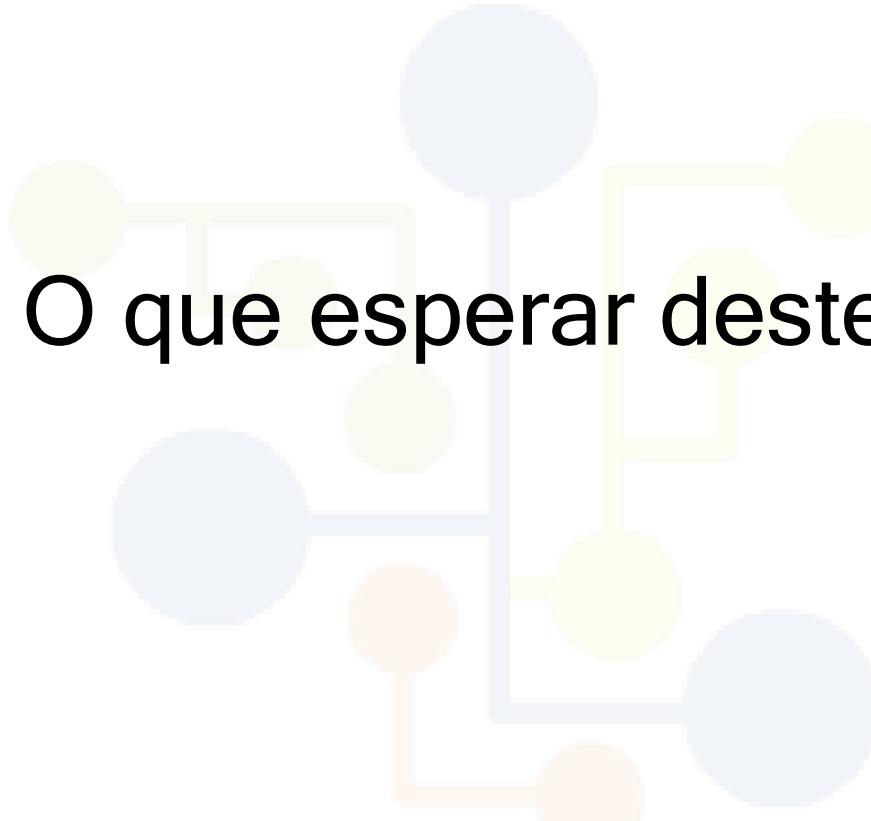
Parte 2
Estatística
Capítulo 6

Parte 3
Machine Learning
Capítulo 7

Parte 4
Azure ML
Capítulos 8, 9, 10



Data Science Academy



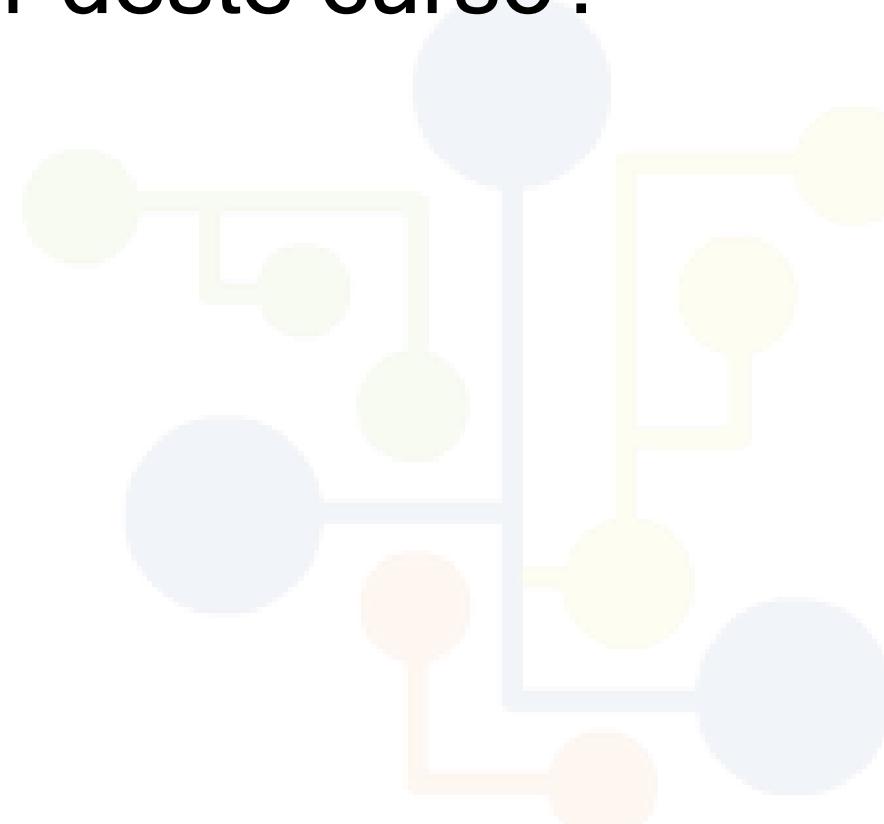
O que esperar deste curso?



Data Science Academy

O que esperar deste curso?

- Praticidade



Data Science Academy

O que esperar deste curso?

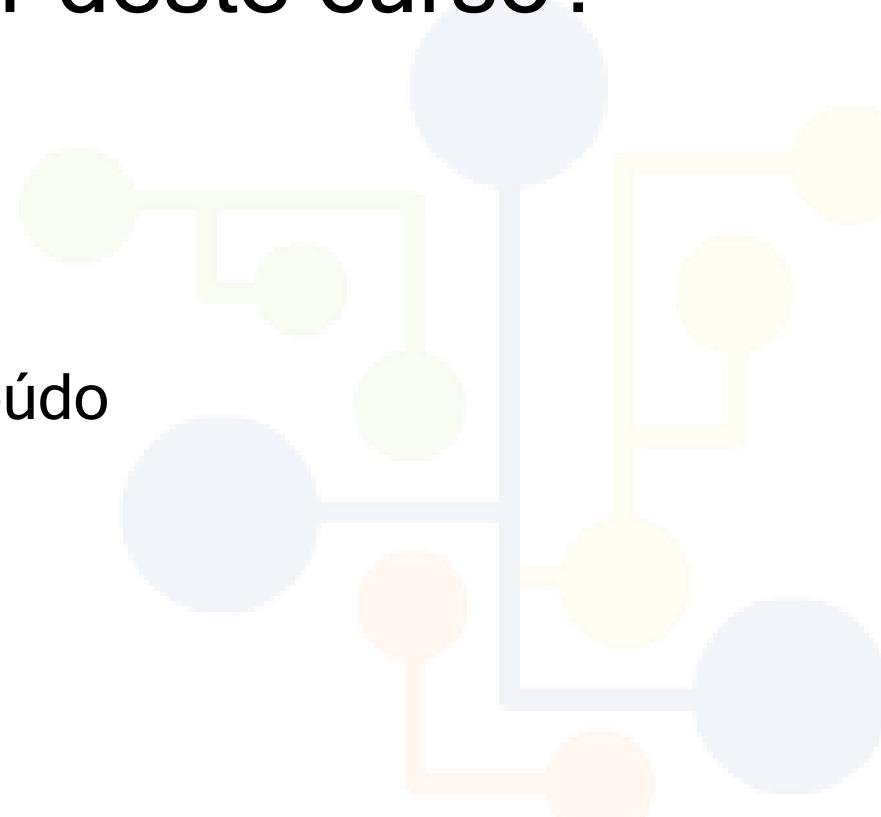
- Praticidade
- Dinamismo



Data Science Academy

O que esperar deste curso?

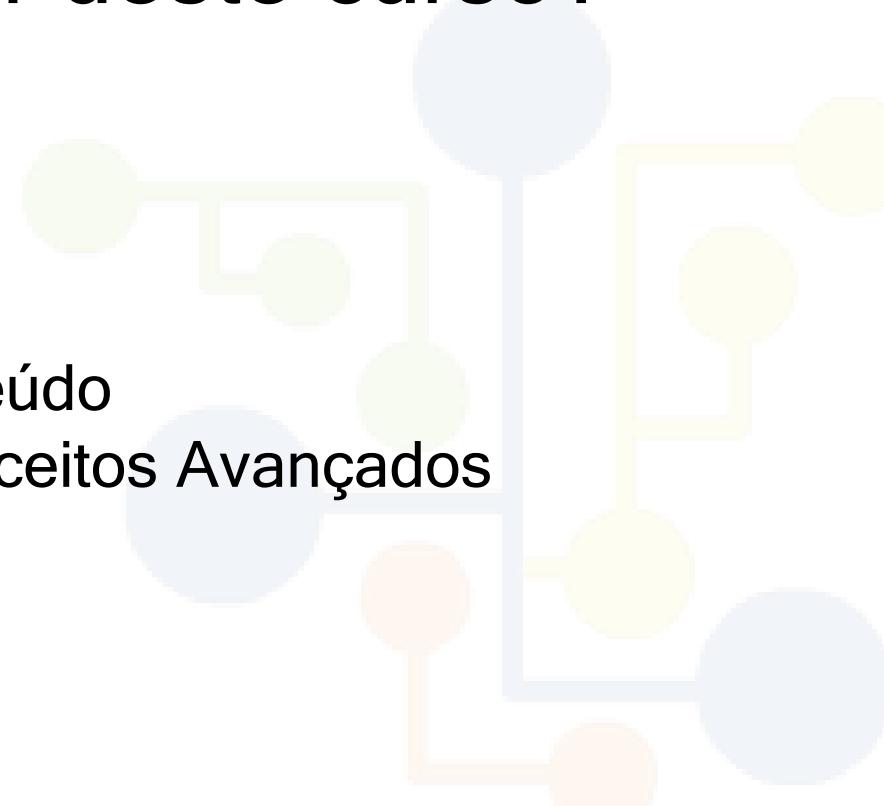
- Praticidade
- Dinamismo
- Muito Conteúdo



Data Science Academy

O que esperar deste curso?

- Praticidade
- Dinamismo
- Muito Conteúdo
- Alguns Conceitos Avançados



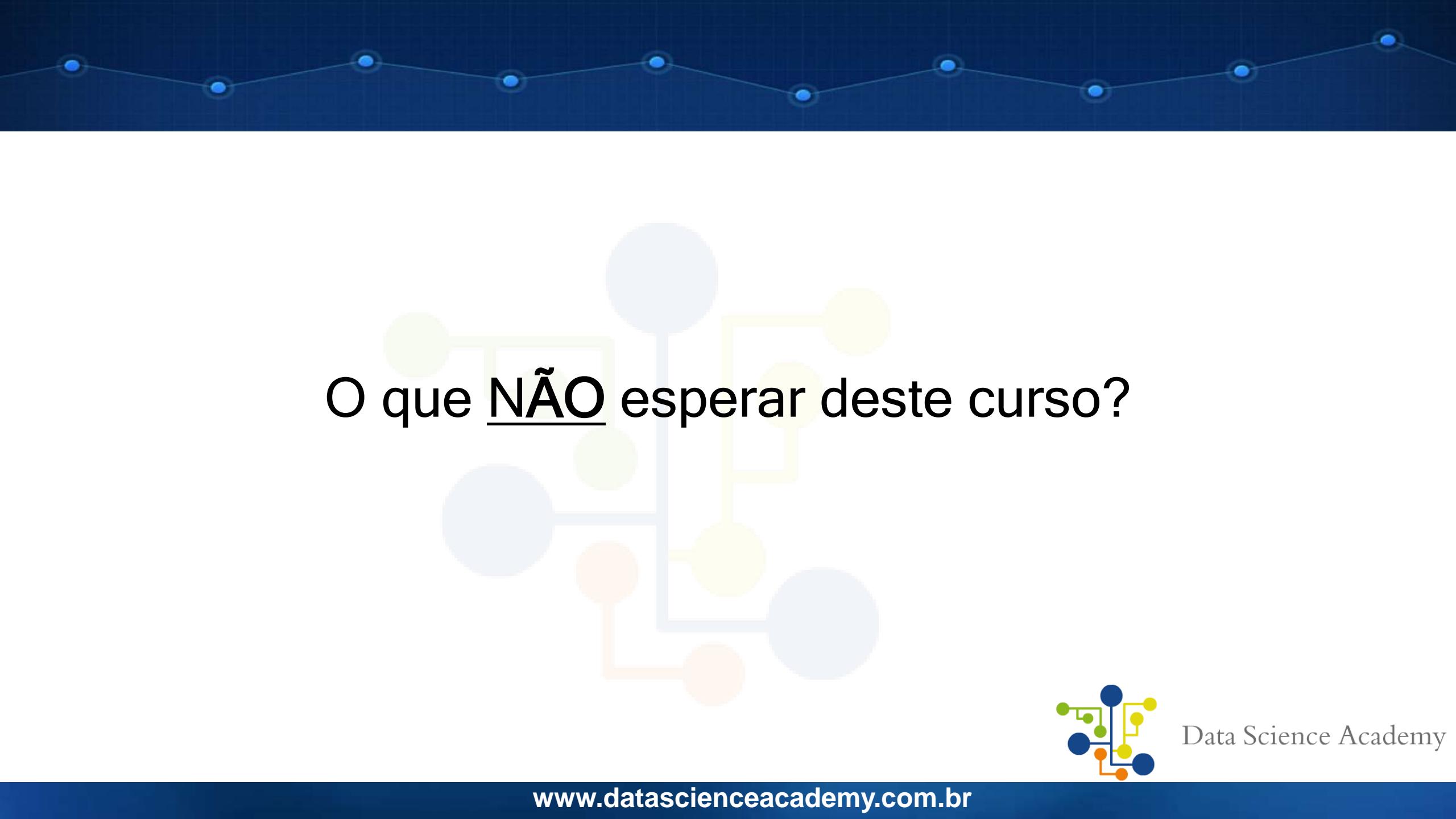
Data Science Academy

O que esperar deste curso?

- Praticidade
- Dinamismo
- Muito Conteúdo
- Alguns Conceitos Avançados
- Objetividade



Data Science Academy



O que NÃO esperar deste curso?



Data Science Academy

O que NÃO esperar deste curso?

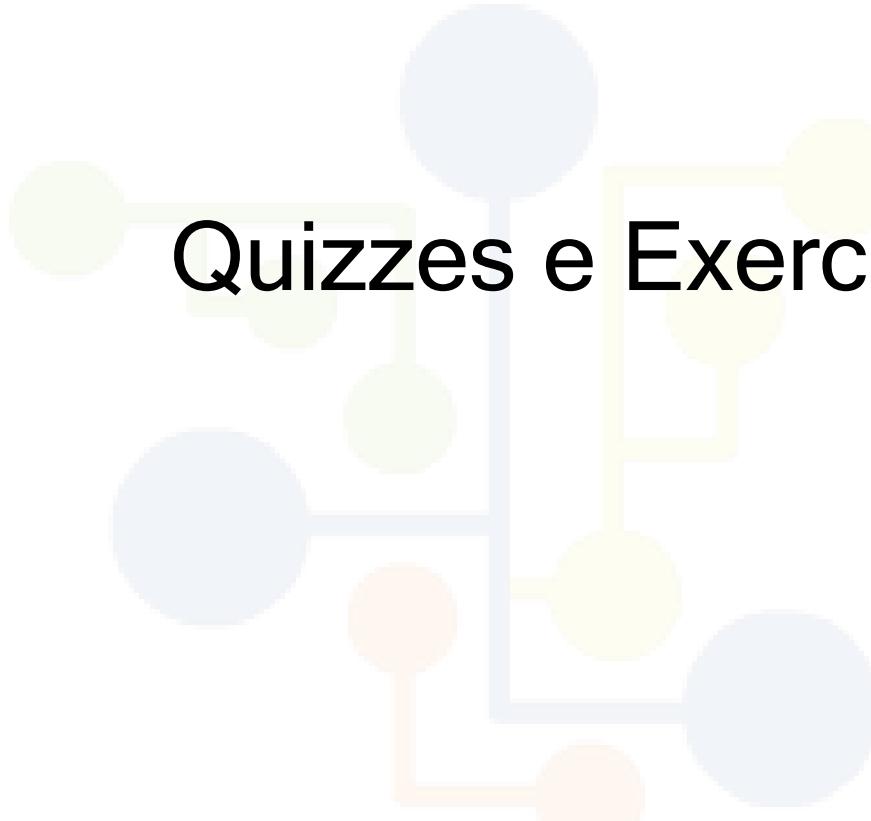
- Este não é um curso só de Estatística
- Comprovar Modelos com Fórmulas Matemáticas
- Não teremos aqui conteúdo de Inteligência Artificial



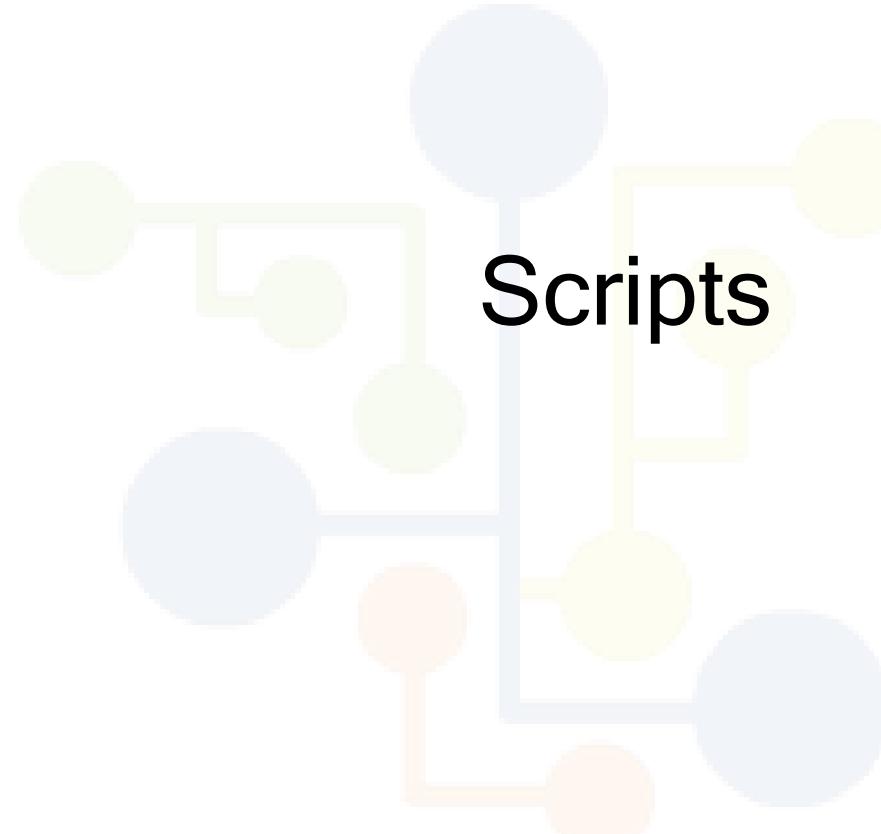
Data Science Academy



Quizzes e Exercícios



Data Science Academy



Data Science Academy

Projetos

Projeto 1 - Análise de Redes Sociais

Projeto 2 - Prevendo a Ocorrência de Câncer

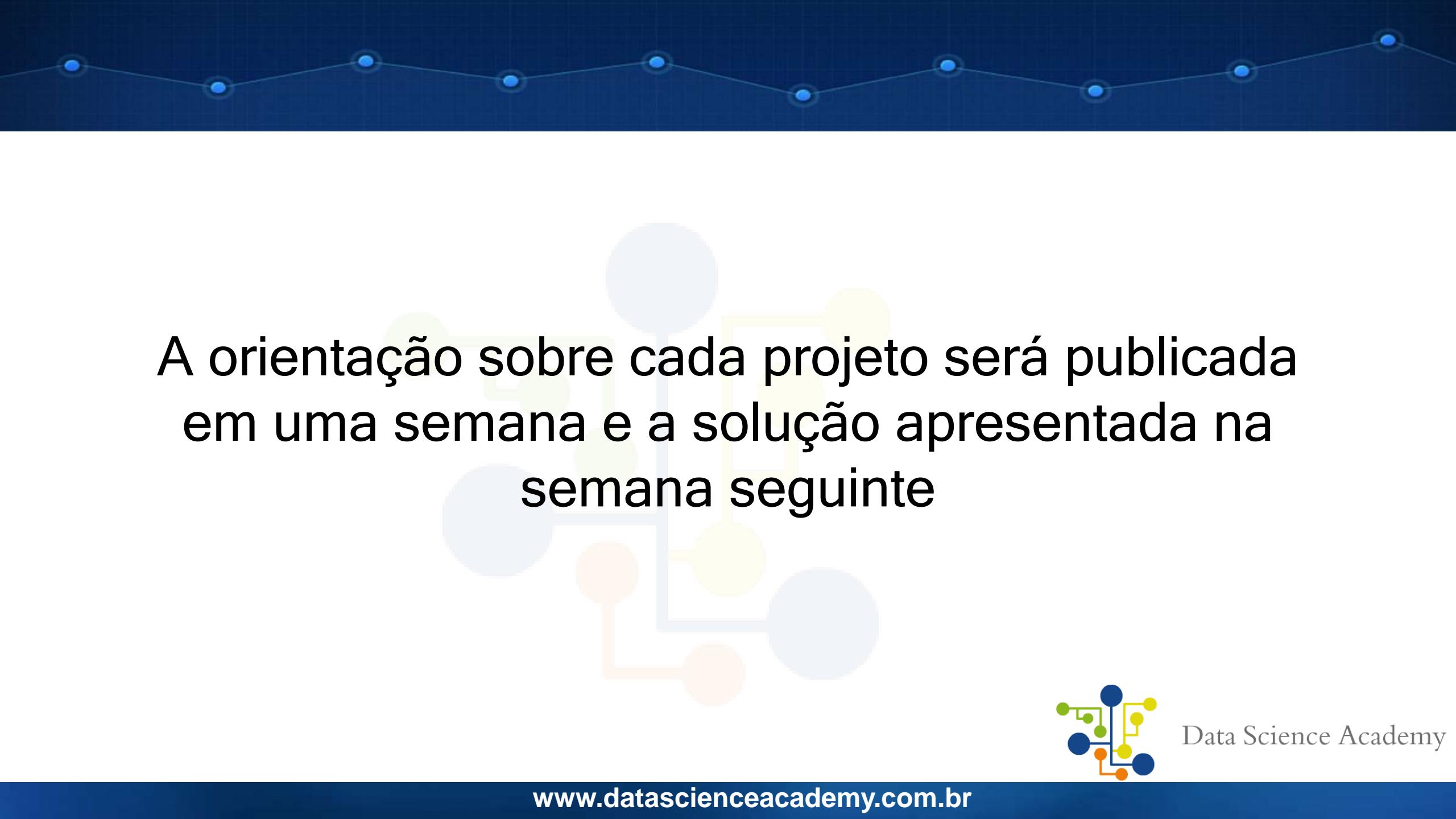
Projeto 3 - Prevendo Despesas Hospitalares

Projeto 4 - Prevendo o Faturamento de uma Rede de Restaurantes

Projeto 5 - Análise de Risco de Crédito



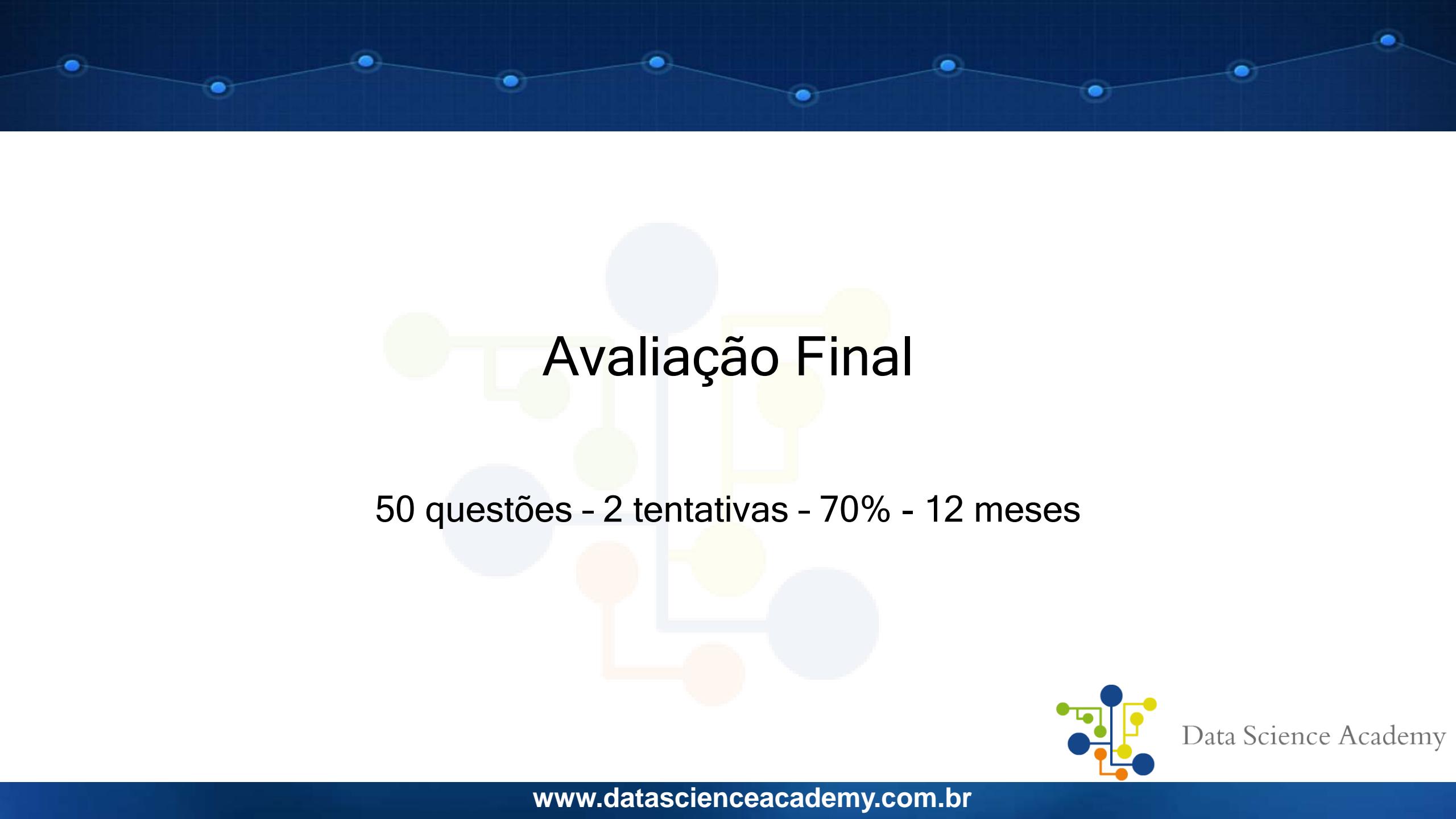
Data Science Academy



A orientação sobre cada projeto será publicada
em uma semana e a solução apresentada na
semana seguinte



Data Science Academy



Avaliação Final

50 questões - 2 tentativas - 70% - 12 meses



Data Science Academy

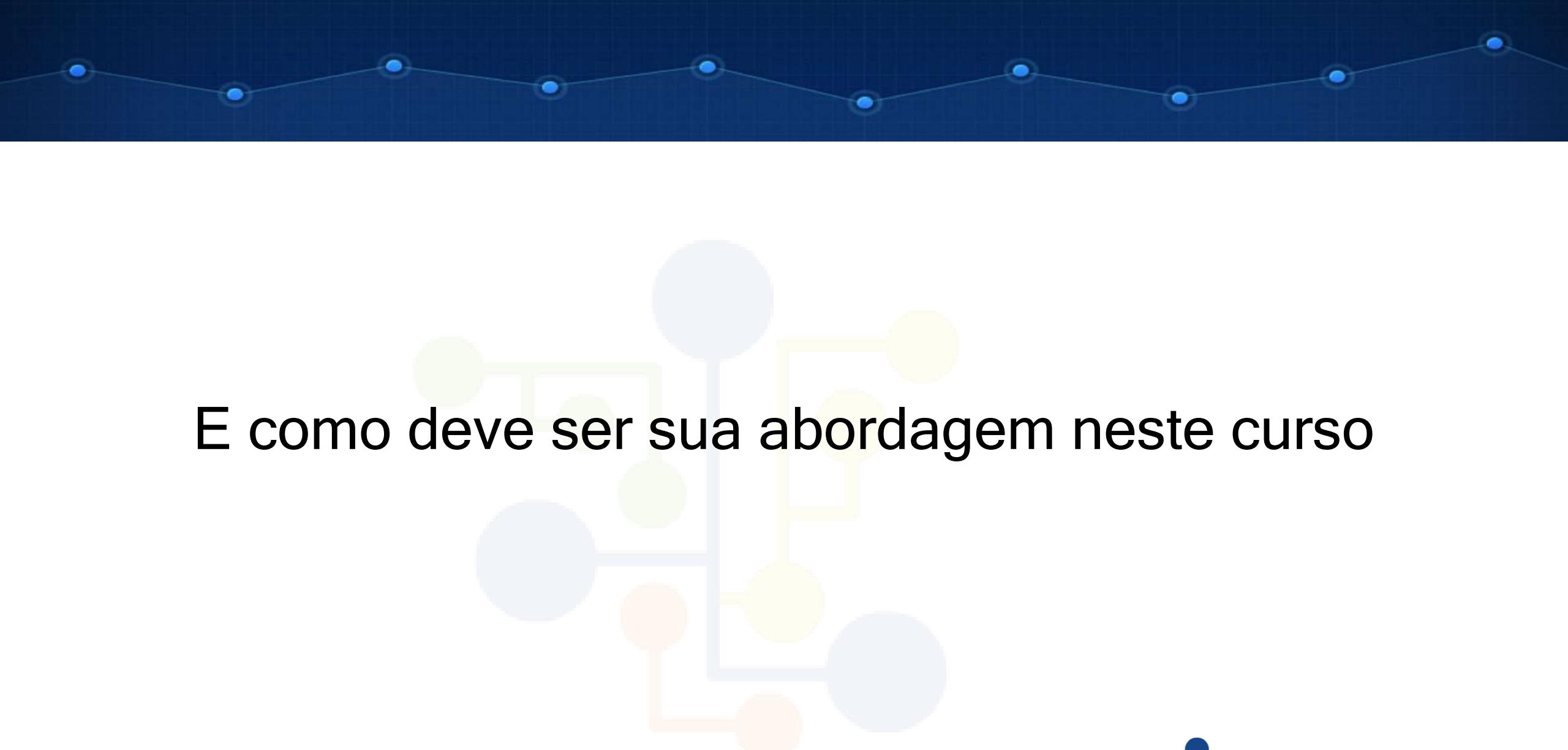


Oracle R Enterprise Microsoft R

Bonus



Data Science Academy



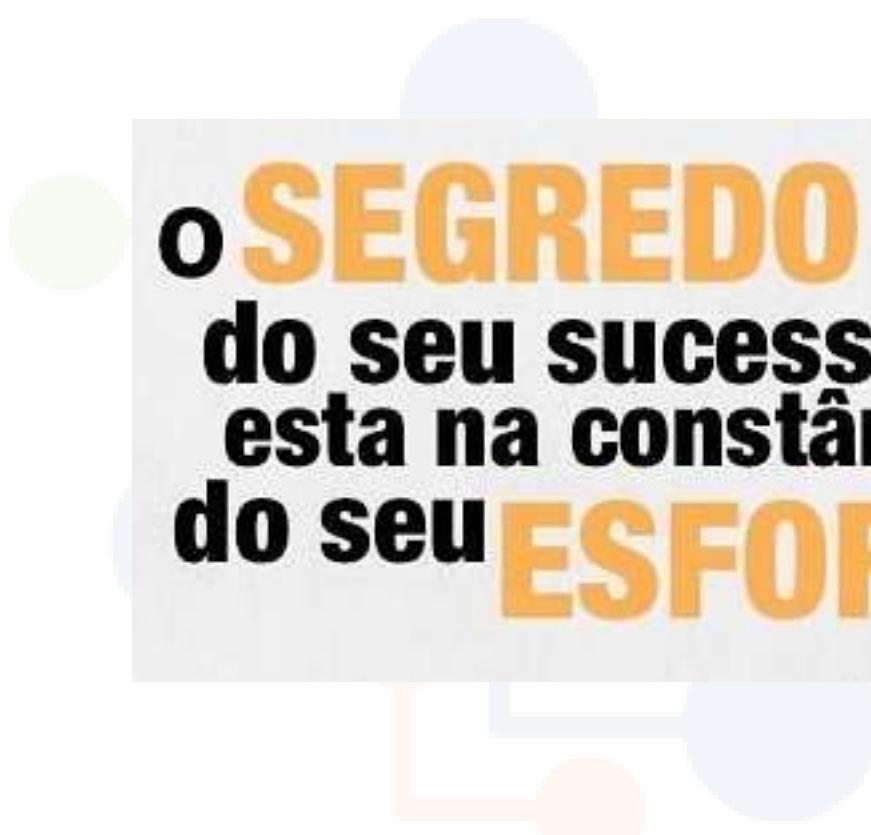
E como deve ser sua abordagem neste curso



Data Science Academy

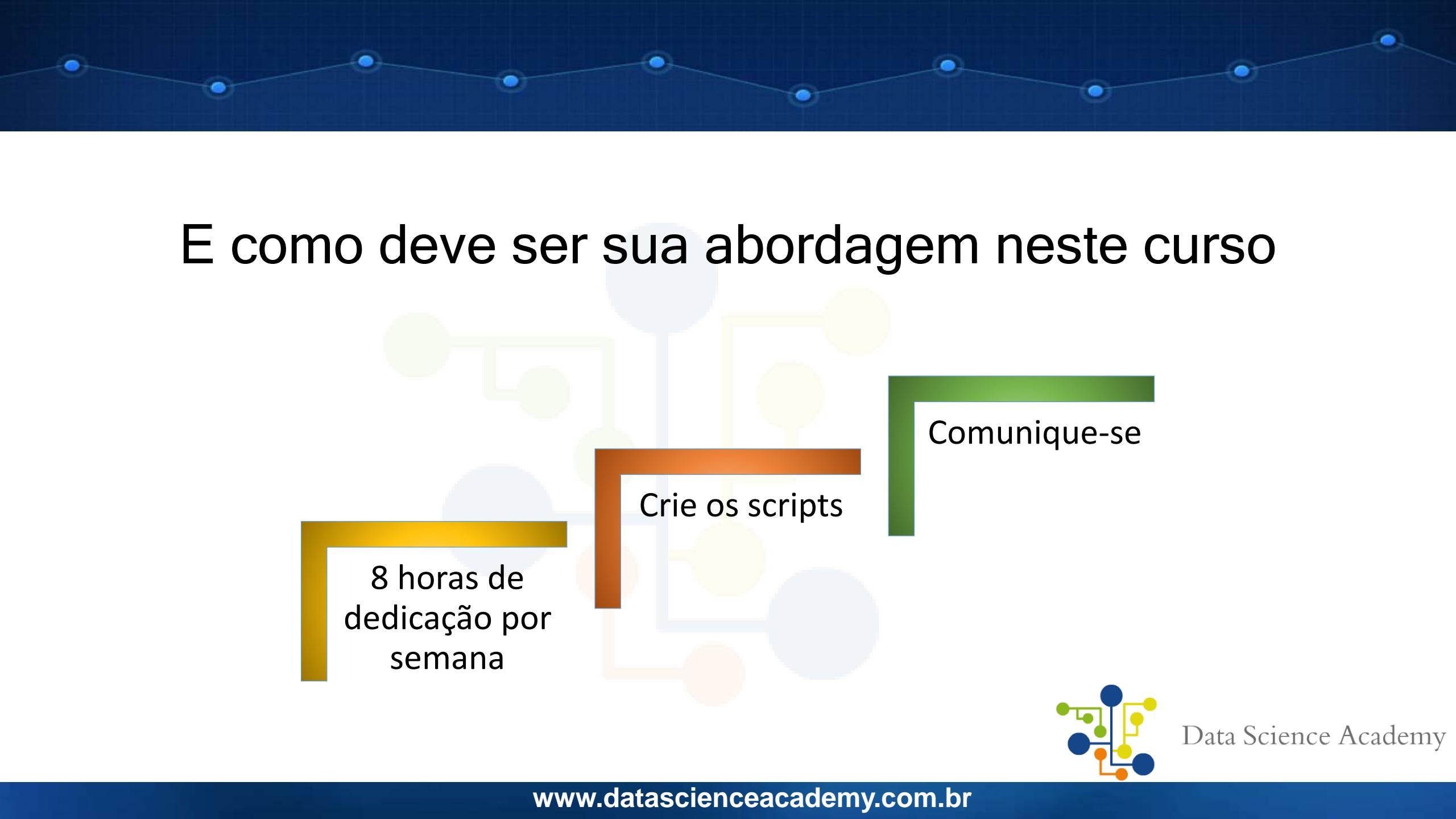


**o SEGREDO
do seu sucesso
esta na constância
do seu ESFORÇO**



Data Science Academy

E como deve ser sua abordagem neste curso



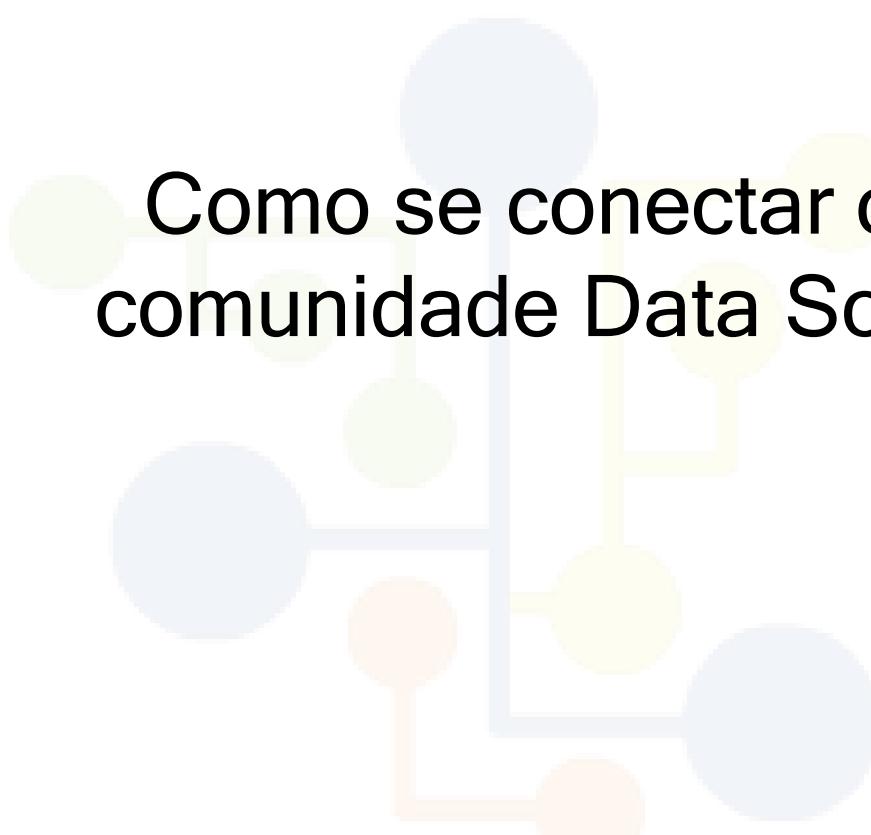
8 horas de
dedicação por
semana

Crie os scripts

Comunique-se



Data Science Academy



Como se conectar com a
comunidade Data Science?



Data Science Academy



www.datasciencecentral.com



Data Science Central

THE ONLINE RESOURCE FOR BIG DATA PRACTITIONERS

ANALYTICS | BIG DATA | HADOOP | DATA PLUMBING | DATAVIZ | JOBS | WEBINARS | DIGEST | NYU STERN MASTERS | SEARCH | CONTACT



Subscribe to DSC Newsletter

MEMBERS



[View All](#)

Data Science Central is the industry's online resource for big data practitioners. From Analytics to Data Integration to Visualization, Data Science Central provides a community experience that includes a robust editorial platform, social interaction, forum-based technical support, the latest in technology, tools and trends and industry job opportunities.

BLOG POSTS

-  Four Recent Salary Surveys for Data Scientists
Posted by Vincent Granville on July 1, 2016 at 11:00am
0 Comments 0 Likes
-  A Plethora of Data Science and Machine Learning Books
Posted by Vincent Granville on June 30, 2016 at 8:49am
0 Comments 0 Likes
-  Complexity frontiers, Artificial Intelligence and Agent-Based Modeling
Posted by Rubens Zimbres on June 26, 2016 at 7:01am
0 Comments 0 Likes

Welcome to
Data Science Central

[Sign Up](#)
[or Sign In](#)

Or sign in with:



Data Science Academy

www.datascienceacademy.com.br



www.kdnuggets.com

KDnuggets™ Data Mining, Analytics, Big Data, and Data Science

Subscribe to **KDnuggets News** | Follow [Twitter](#) [Facebook](#) [LinkedIn](#) | [Contact](#)

search KDnuggets Search

SOFTWARE | NEWS | Top stories | Opinions | Tutorials | JOBS | Academic | Companies | Courses | Datasets | EDUCATION | Certificates | Meetings | Webinars

NEW YORK CITY OCTOBER 23-27, 2016

Empower your business with predictive analytics [LEARN MORE](#)

PAW New York, Oct 23-27: Empower your business with Predictive Analytics. Register today!

Data Mining, Analytics, Big Data, Data Science

[Software](#) (Suites, Text, Visualization)
[Jobs - Industry](#) | [Academic](#)
[Meetings, Conferences](#)
[Companies](#) (Consulting, Products)
[Courses in Big Data, Data Science](#)
[Datasets](#) (APIs/Markets, Gov)
Data Mining Course | Gregory Piatetsky
[Education](#) (online, USA, Europe, cert)
[FAQ](#) | [Polls](#) | [Publications](#) (Books)
[Solutions](#) (Fraud, Data Cleaning)
[Webcasts](#) | [Websites](#) (Blogs, Cartoons, Podcasts)

Most Recent

- Altria: Senior Analyst, Consumer & Marketplace In...
- Upcoming Meetings in Analytics, Big Data, Data Mining,...
- What is Softmax Regression and How is it Related to Lo...
- Three Impactful Machine Learning Topics at ICML 2016
- Text Mining 101: Topic Modeling

→ [KDnuggets Blog Contest: Automated Data Science and Machine Learning](#) - submit by July 15.

→ [KDnuggets 2016 Software Poll results: R, Python Duel As Top Analytics, Data Science software](#)

Latest
[News](#) | [Software](#) | [Tutorials](#)

Subscribe to **KDnuggets News** | Follow [Twitter](#) [@kdnuggets](#) | [Facebook](#) [LinkedIn](#)

Northwestern
PREDICTIVE ANALYTICS
School of Professional Studies

Online MS in Predictive Analytics

Practical guide to Automate Machine Learning [Download Now](#)

Practical Guide to Automate Machine Learning

SEE THE BEAUTY BEHIND BIG DATA REGIS

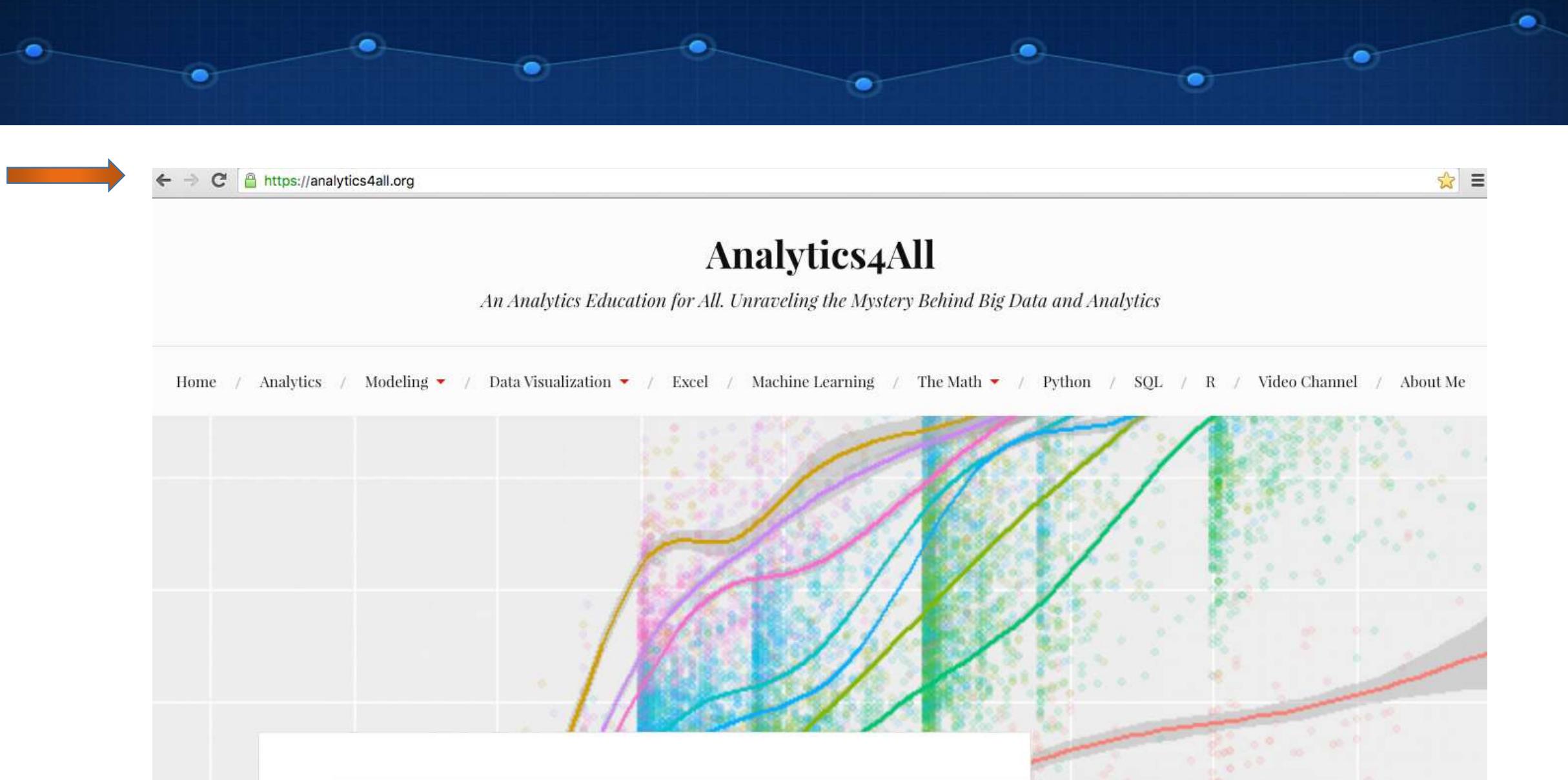


Data Science Academy

The screenshot shows the homepage of www.predictiveanalyticstoday.com. The header features a large blue navigation bar with various sections like 'PREDICTIVE ANALYTICS', 'BUSINESS INTELLIGENCE', 'TEXT ANALYTICS', etc. A prominent orange arrow points from the left towards the top right corner of the page. The main content area includes a banner for 'QDA MINER QUALITATIVE SOFTWARE' and several news cards. One card highlights 'TOTAL MENTIONS' for 'Digimind' with a value of '31,418'. Another card for 'Dataiku DSS' features a logo of interconnected colored circles.



Data Science Academy



Data Science Academy



www.r-bloggers.com

R-bloggers

R news and tutorials contributed by (573) R bloggers

Home | About | RSS | add your blog! | Learn R | R jobs | Contact us

WELCOME!

Follow @rbloggers 33.4K

Here you will find daily news and tutorials about R, contributed by over 573 bloggers.

There are many ways to follow us -

By e-mail:

Your e-mail here

Subscribe

31008 readers BY FEEDBURNER

On Facebook:

R blog... 35 mil curtidas

Curtiu

Você e outros 41 amigos curtiram isso

Combine ArcGIS and R – Clustering Toolbox

July 2, 2016

By Fabio Veronesi

r-bridge-install-master
img
rtools
R Integration.pyt
Install R bindings
Print R Version
R Installation Details
Update R bindings

Last year at the ESRI User Conference in San Diego, there was an announcement of an initiative to bridge ArcGIS and R. This became reality I think early this year with R-Bridge. Basically, ESRI has created an R library that is able to communicate and

ex...

Read more »

Run compiled R packages in AzureML

July 1, 2016

By David Smith

Animated world inequality map

July 1, 2016

By Peter's stats stuff - R

Search & Hit Enter

RECENT POPULAR POSTS

- The useR! 2016 Tutorials
- RMarkdown and Metropolis/Mtheme
- How to write good tests in R

MOST VISITED ARTICLES OF THE WEEK

- How to write the first for loop in R
- Installing R packages
- R tutorials
- Using apply, sapply, lapply in R
- In-depth introduction to machine learning in 15 hours of expert videos
- How to perform a Logistic Regression in R
- Computing and visualizing PCA in R
- Express Intro to dplyr
- The useR! 2016 Tutorials



Data Science Academy



https://rpubs.com

RPubs brought to you by RStudio

Sign in Register

Easy web publishing from R

Write R Markdown documents in RStudio.
Share them here on RPubs. (It's free, and couldn't be simpler!)

Get Started

Recently Published

NOAA Storm Database exploration

Michael Boggs
July 2, 2016

This note shows an exploratory analysis of NOAA storm database data. The goal of this note is to determine what factors influence the number of deaths caused by a single storm. This is done by first loading the data, then performing some basic data cleaning and analysis.

Loading Libraries

Test

R Markdown

Integer Sequence Learning



Data Science Academy



https://try.jupyter.org

jupyter

Hosted by Rackspace

Files Running Clusters

Select items to perform actions on them.

Upload New ▾

Text File	
Folder	
Terminal	
Notebooks	
Apache Toree - Scala	
Bash	
Haskell	
Julia 0.3.2	
Python 2	
Python 3	
R	
Ruby 2.1.5	

Create a new notebook with R

communities
datasets
featured
Welcome Julia - Intro to Gadfly.ipynb
Welcome R - demo.ipynb
Welcome to Haskell.ipynb
Welcome to Python.ipynb
Welcome to Spark with Python.ipynb
Welcome to Spark with Scala.ipynb



Data Science Academy



www.cienciaedados.com

INÍCIO APP SOBRE SLIDE SHARE CONTATO

 Ciência e Dados
Data Science For Professionals

f t w p g+ t

CIÊNCIA E DADOS

DATA SCIENCE, BIG DATA, ANALYTICS, INTERNET OF THINGS



Data Science Academy



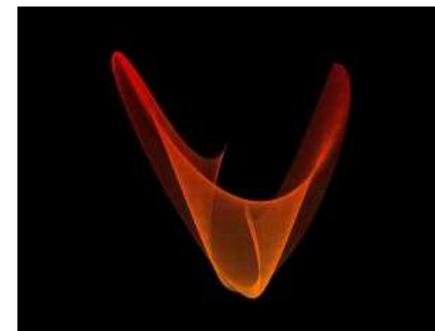
computerworld.com.br/big-data

Big Data

Notícias, tendências, análises, dicas e boas práticas sobre Big Data



Exemplos de aplicações de data mining no mercado brasileiro



Volume, variedade, velocidade, veracidade e valor: Os cinco Vs do Big Data



Quantidade de dados vai mudar estratégias de negócios até 2020

17/06/2016 - 12h40

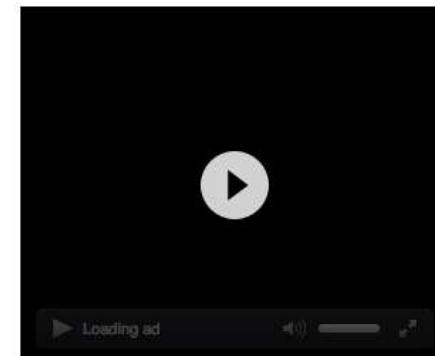
Hu Yoshida, CTO da HDS, afirma que apenas 15% do volume existente de informações será realmente armazenada pelas empresas



Menos de um terço dos projetos de Big Data são lucrativos

16/06/2016 - 14h48

Pesquisa da Capgemini e Informatica aponta que 45% dos projetos envolvendo grandes volumes de dados apenas cobrem os gastos



Mais Lidas



Para ficar de olho: Dez tecnologias emergentes que revolucionarão o mundo



Como cinco minutos por semana podem mudar positivamente sua carreira



Passo a passo para evitar a



Data Science Academy



oestatistico.com.br/2016/05/

HOME SOBRE O BLOG FAQ DO ESTATÍSTICO CONTATO

Arquivos Mensal

MAIO 2016

POR QUE ESTATÍSTICA É TÃO IMPORTANTE?

CURIOSIDADES, ESTATÍSTICA APLICADA, TECNOLOGIA

Estatística! Muita gente tem aversão a esse nome, e sente até arrepios. Associa rapidamente a palavra àqueles cálculos intermináveis, que no final das contas sempre dá 1 ou -1 (às vezes zero); e que gera tremenda dor de cabeça para alunos e profissionais que não são da área, mas precisam cumprir com o aprendizado da disciplina.

Mas a verdade é que seremos cada vez mais dependentes dessa ciência; e sem dúvidas ela vai nortear a sua e a minha vida, praticamente em tudo o que formos fazer. Talvez você já enxergue isso, ou talvez não. Mas ao longo do texto, vamos dar exemplos práticos da influência da estatística no seu dia a dia.

[CONTINUAR LENDO](#)

O Estatístico

Raniere Ramos é Estatístico formado pela Universidade Estadual de São Paulo e acredita que as decisões mais inteligentes são tomadas com base em dados.

PESQUISAR

FACEBOOK



Você curtiu isso



Data Science Academy



unitedstatisticians.com/blog/estatistico-profissao-do-futuro-que-ja-chegou/

TRENDING: Estatístico: a profissão do futuro (que já chegou)

f t g+ SEARCH



Início Encontrar estatísticos



PESQUISAR ...

RECENT POSTS



Saiba como a estatística pode nos ajudar a mantermos vivos!

Jun 29, 2016 | Aplicações da Estatística



Marketing: uma ferramenta dependente dos estudos da estatística

Jun 22, 2016 | Aplicações da Estatística



Como fazer sua carreira de consultor estatístico decolar?

Jun 15, 2016 | Consultoria Estatística, Melhore sua carreira



Data Science Academy



- Users of R Statistical Package
- R Brasil - Programadores
- Big Data Brasil
- Windows Azure - Comunidade Azure
- Estatística Brasil



Data Science Academy



Divirta - se!



Data Science Academy



Data Science Academy



obrigado!



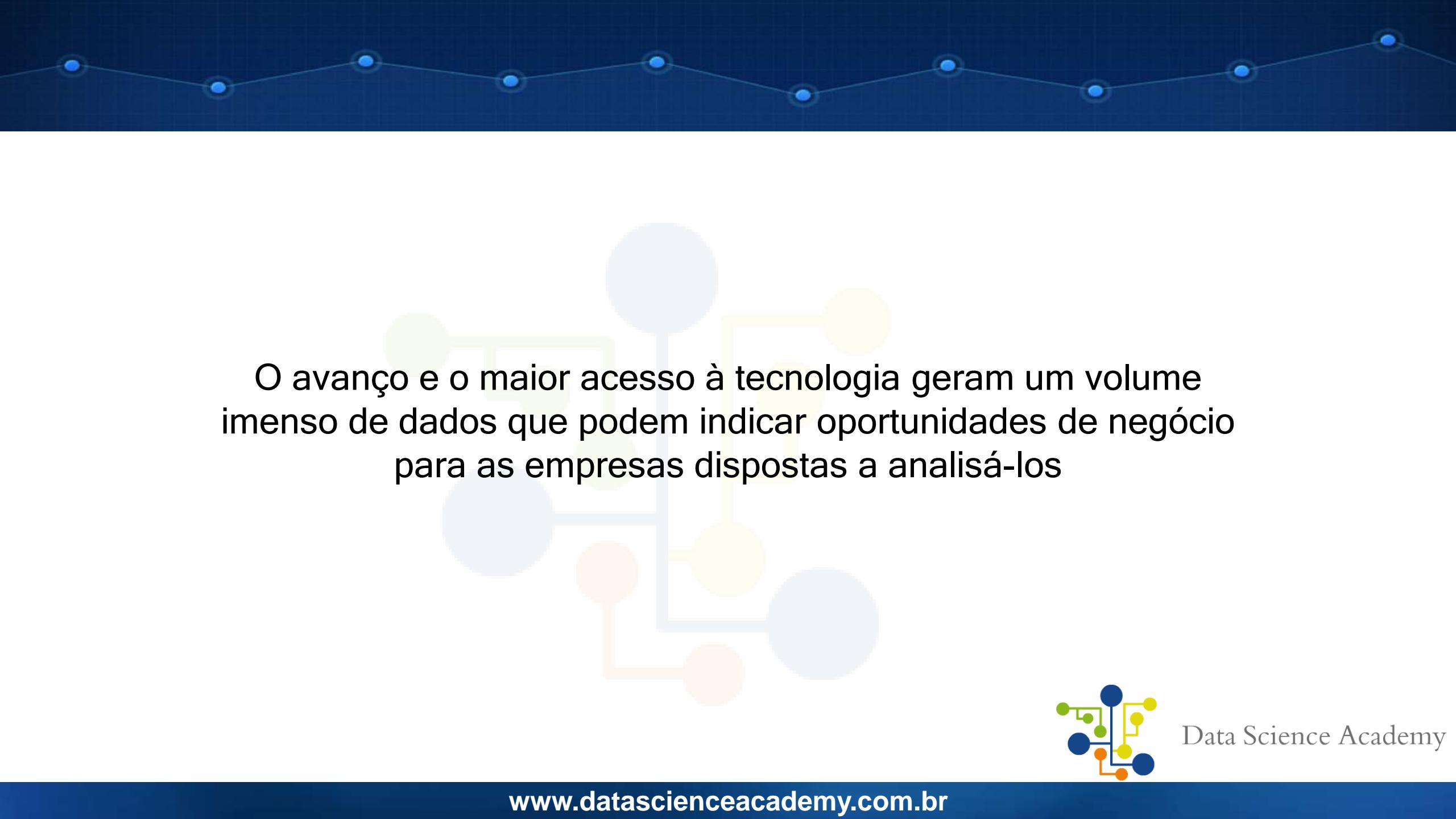
Data Science Academy



A Carreira de Cientista de Dados



Data Science Academy



O avanço e o maior acesso à tecnologia geram um volume imenso de dados que podem indicar oportunidades de negócio para as empresas dispostas a analisá-los



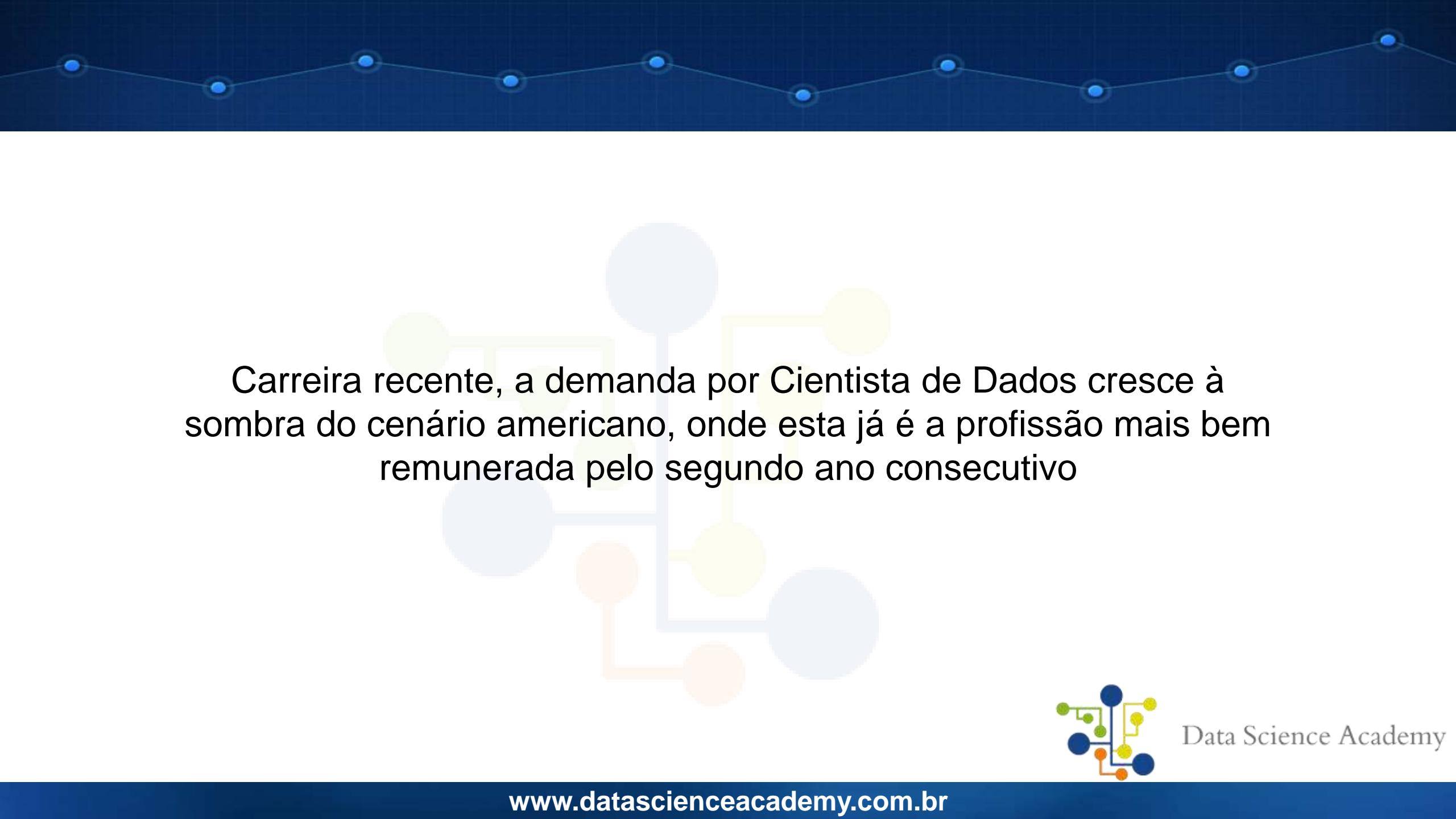
Data Science Academy



A função cabe ao Cientista de Dados, carreira listada pelo Fórum Econômico Mundial como uma das mais relevantes para o mercado até 2025



Data Science Academy



Carreira recente, a demanda por Cientista de Dados cresce à sombra do cenário americano, onde esta já é a profissão mais bem remunerada pelo segundo ano consecutivo

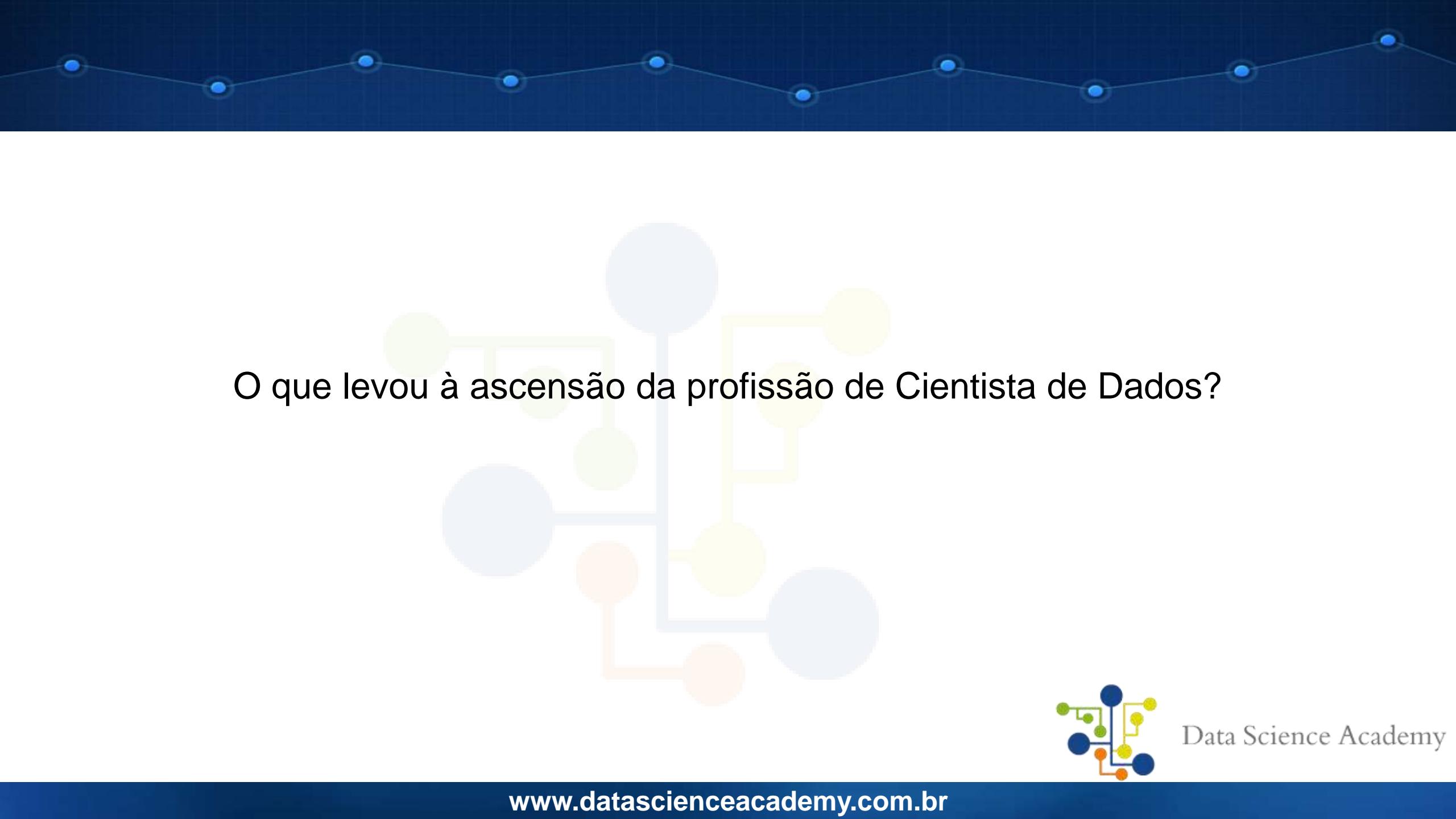


Data Science Academy

“O mercado brasileiro já demanda um bom volume de profissionais, maior até do que a quantidade de pessoas capacitadas para trabalhar na área”, diz Celso Poderoso, coordenador dos cursos de MBA em “Big Data” da Fiap (Faculdade de Informática e Administração Paulista).



Data Science Academy



O que levou à ascensão da profissão de Cientista de Dados?



Data Science Academy



A Carreira

Você já pensou nas vantagens de trabalhar com a análise de dados e ser um dos especialistas mais procurados pelo mercado?



Data Science Academy



Big Data Analytics é fundamental
para a tomada de decisão nas
empresas



Data Science Academy



Uma carreira dedicada à área de Big Data Analytics, pode ser uma das melhores decisões para qualquer profissional com capacidades analíticas



Data Science Academy



A ascensão da análise de
dados estruturados e não
estruturados



Data Science Academy



As previsões de Big Data
estão superando as
previsões de mercado



Data Science Academy



Entre 2015-2020, o Big Data Analytics terá um crescimento de 14,4% ao ano



Data Science Academy



Big Data Analytics é usado em
todo tipo de negócio



Data Science Academy



Já está convencido da
importância da profissão de
Cientista de Dados?



Data Science Academy



Ainda não?



Data Science Academy



Então tenho mais
argumentos!!



Data Science Academy



Existem inúmeras opções de
cargos para os profissionais
de Big Data



Data Science Academy



Big Data Analytics é uma Estratégia
de Negócio



Data Science Academy



E o Cientista de Dados é parte desta
Estratégia



Data Science Academy



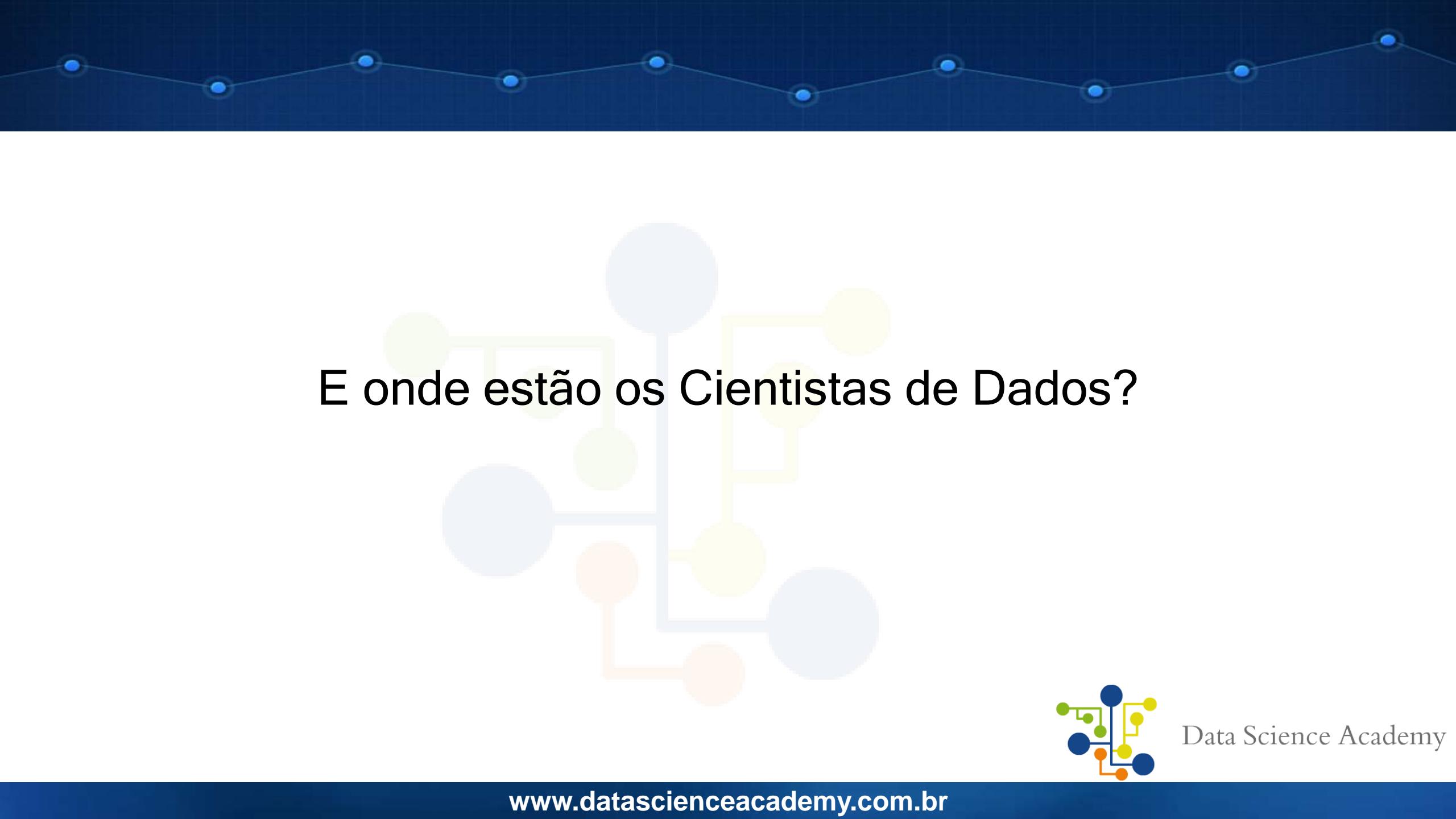
O que é um cientista de dados?

*Conhecimento Acadêmico
(cientistas de dados norte-americanos)



Science Academy





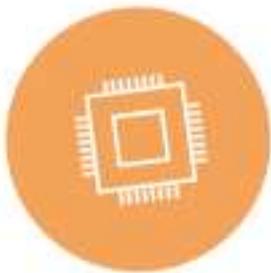
E onde estão os Cientistas de Dados?



Data Science Academy



O número de Cientistas de Dados em todo mundo dobrou nos últimos 4 anos



As áreas de TI e Serviços são as que mais contratam esses profissionais



As principais habilidades são R, Python, Data Mining e Machine Learning

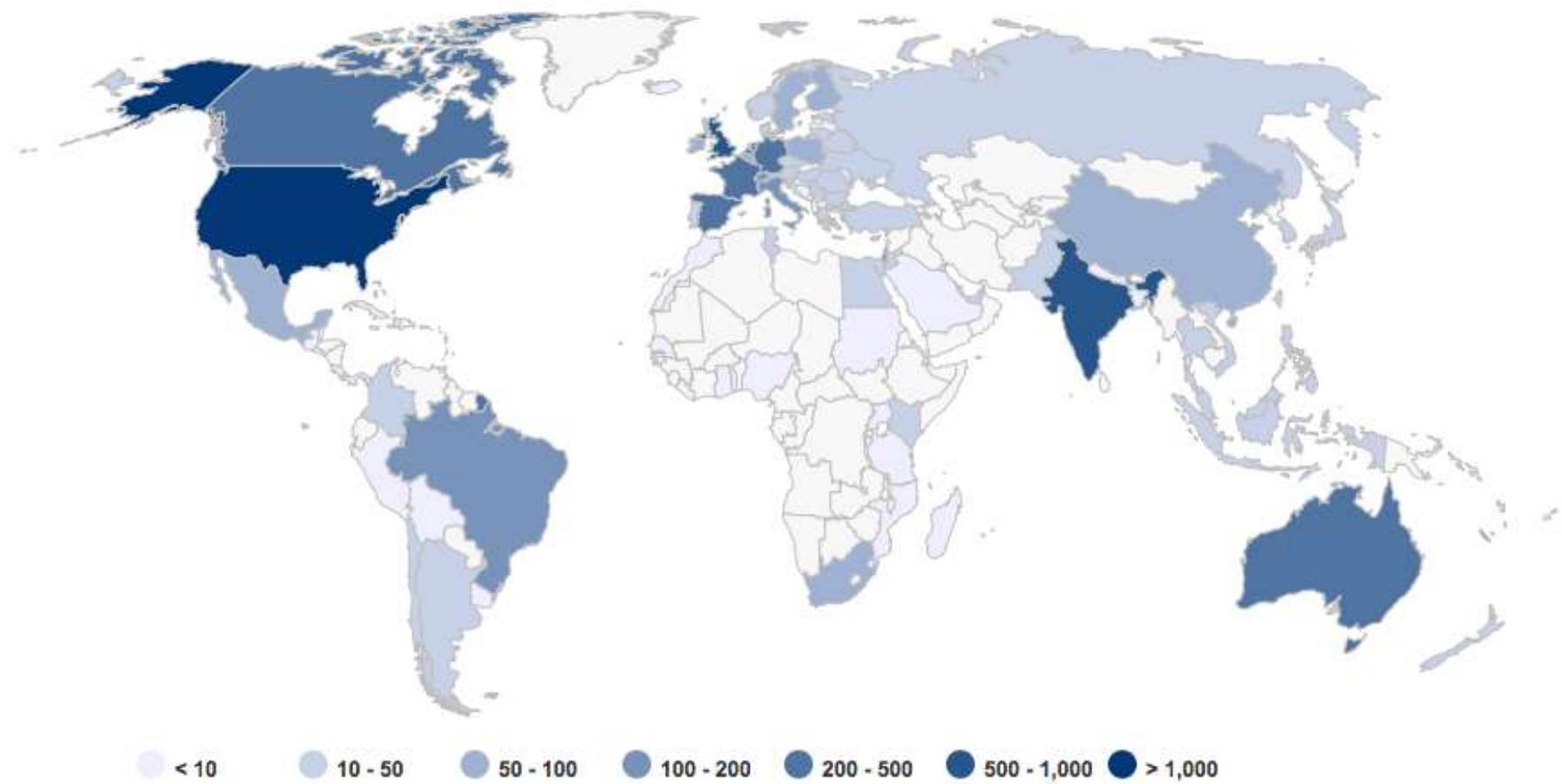


A maioria tem formação em Ciência da Computação, Estatística, Matemática e Física



Data Science Academy

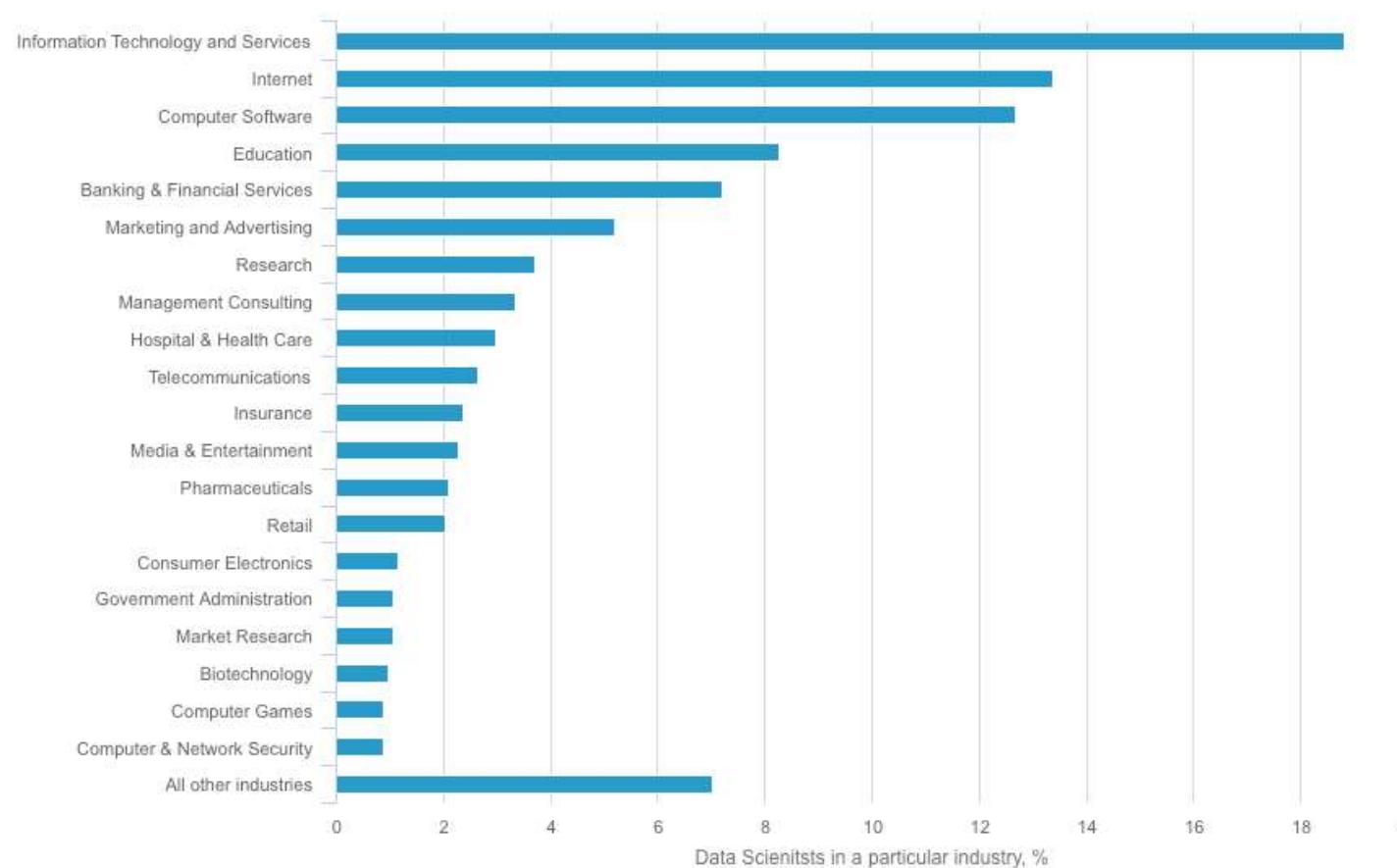
DATA SCIENTISTS WORLDWIDE, LOWER BOUND



RJMETRICS


Science Academy

DATA SCIENTISTS BY INDUSTRY, TOP 20

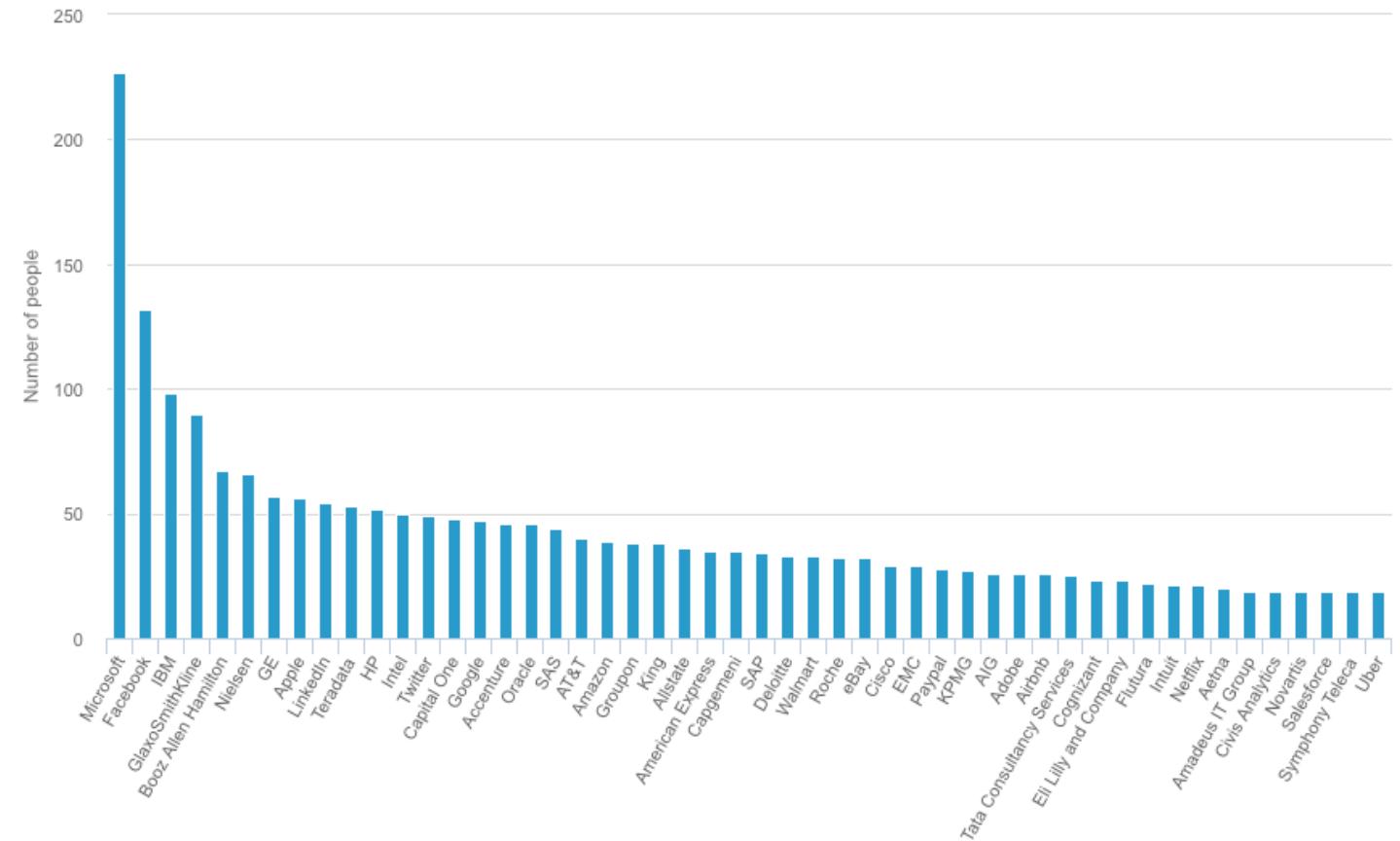


RJMETRICS

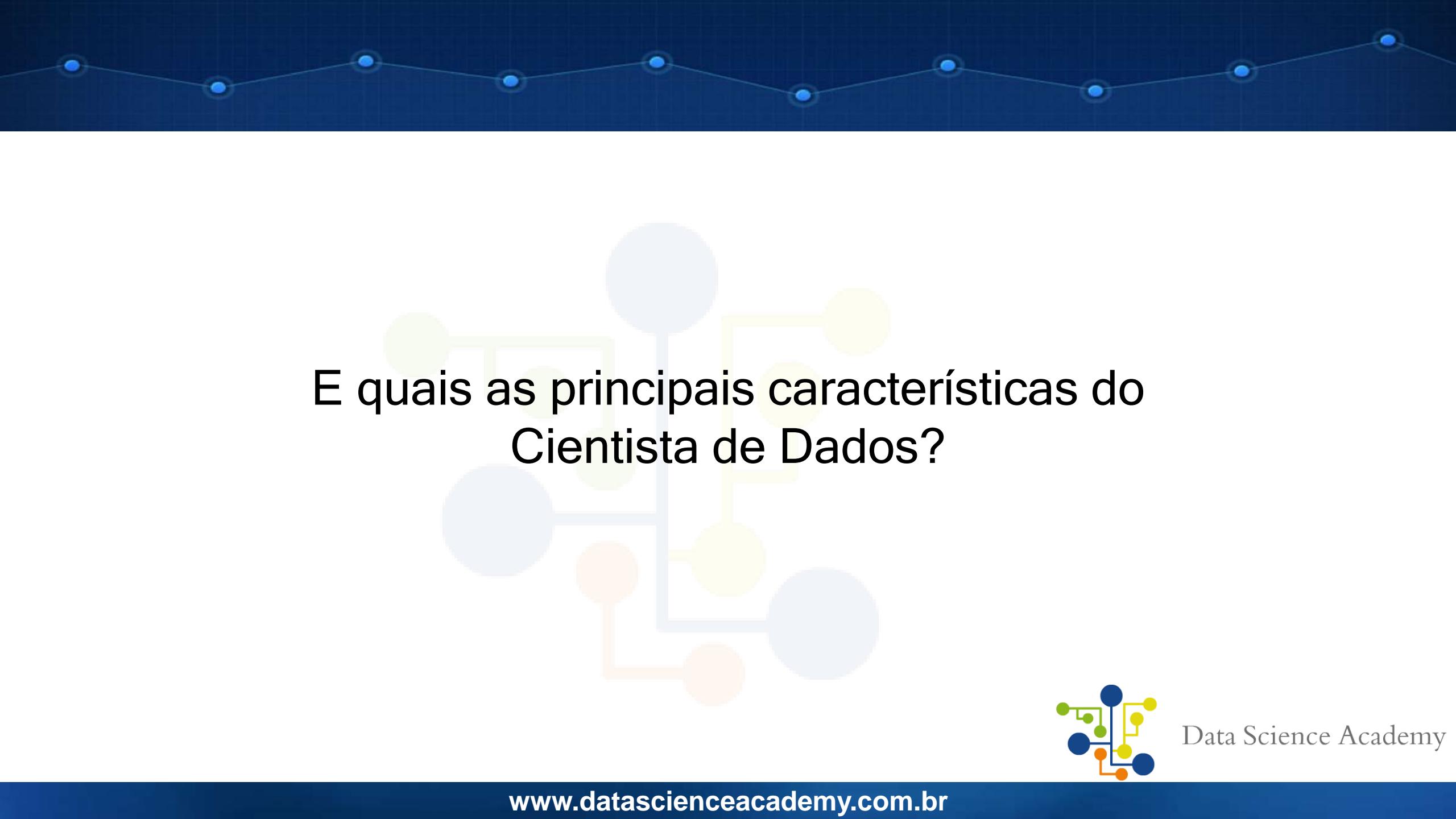


Data Science Academy

DATA SCIENTISTS PER COMPANY, TOP 50



Data Science Academy



E quais as principais características do
Cientista de Dados?



Data Science Academy

E quais as principais características do Cientista de Dados?

- 1- Ferramentas Básicas
- 2- Estatística
- 3- Machine Learning
- 4- Cálculo Multivariado e Álgebra Linear
- 5- Data Munging
- 6- Visualização de Dados e Comunicação
- 7- Pense como um Cientista de Dados



Data Science Academy



obrigado!



Data Science Academy



Por que Cientistas de Dados usam R?



Data Science Academy



Breve História do R



Data Science Academy

Breve História do R

1993	Projeto de pesquisa em Auckland, na Nova Zelândia
1995	R liberado como projeto open-course
1997	Formado o grupo R-Core
2000	Liberada a versão 1.0.0 do R
2003	Criação da R Foundation
2004	Primeira conferência internacional de usuários em Vienna
2015	Formado o R Consortium (com participação da IBM e Microsoft)



Data Science Academy

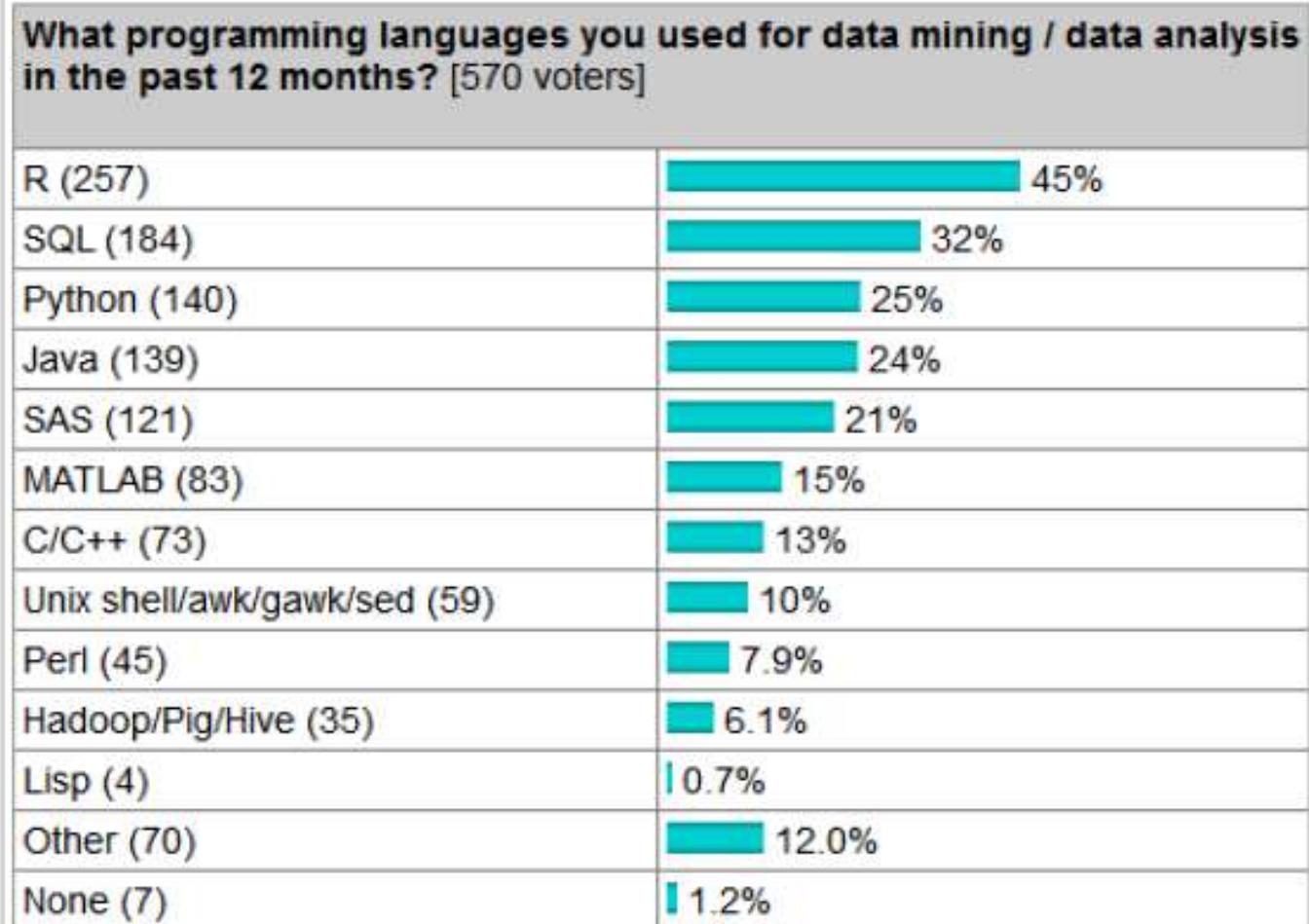
R possui diversas funções para:

- Extração de Dados
- Limpeza de Dados
- Carregamento e Transformação de Dados
- Análise Estatística
- Modelagem Preditiva
- Machine Learning
- Visualização de Dados

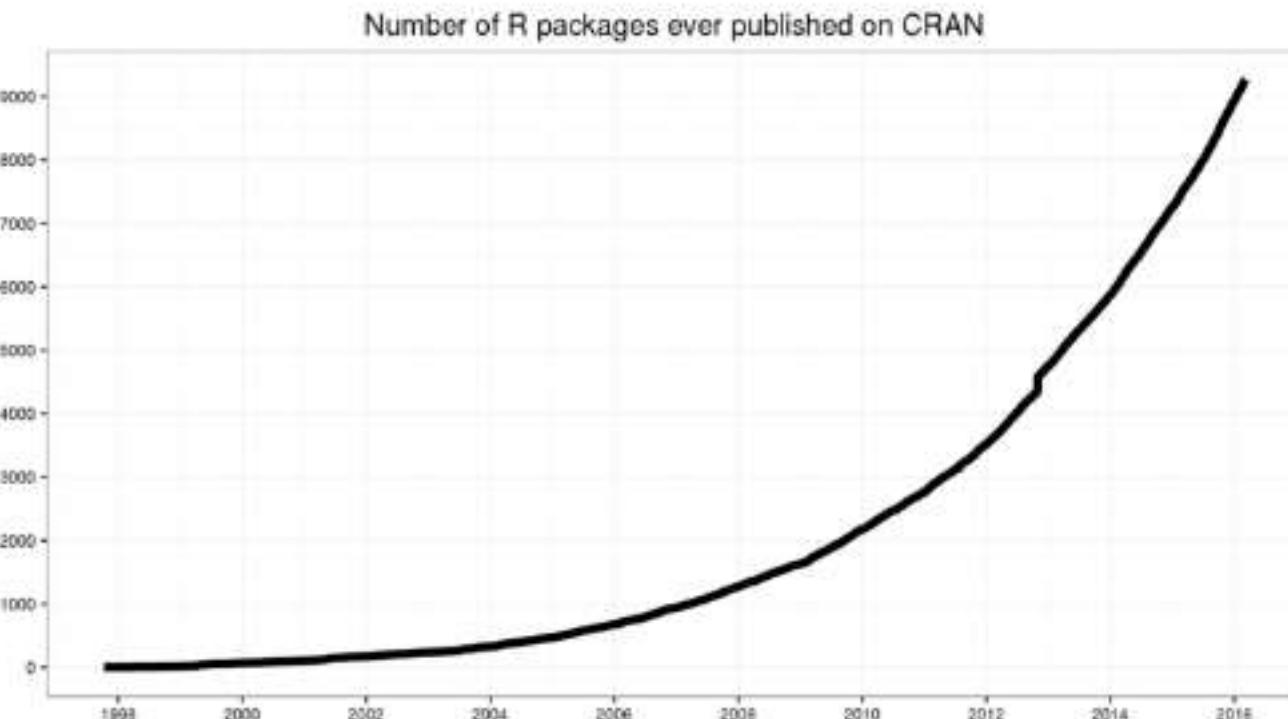


Data Science Academy

Pesquisa realizada no site KD



Data Science Academy



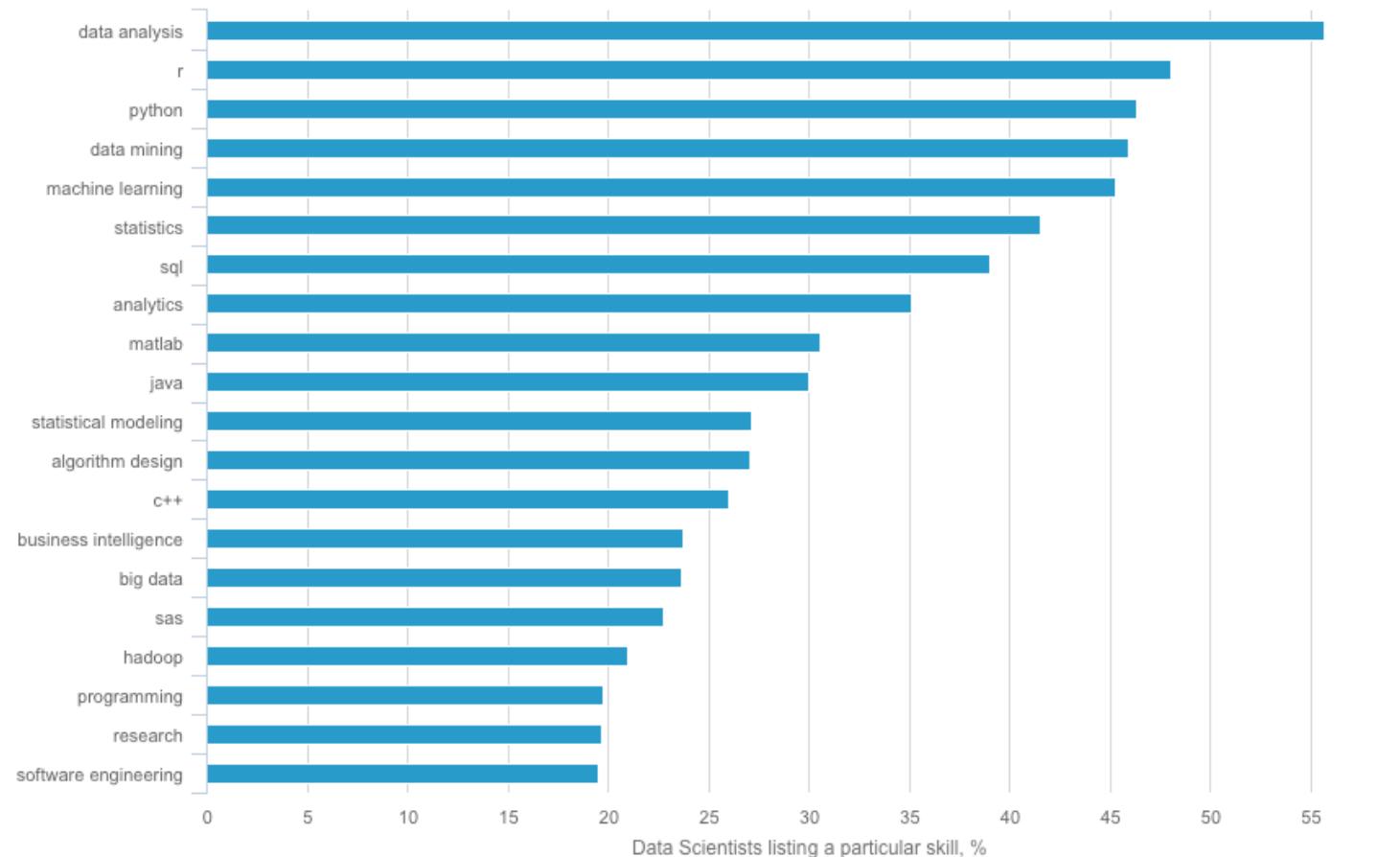
Número de pacotes R publicados por ano no CRAN

Fonte: <http://blog.revolutionanalytics.com/2016/03/16-years-of-r-history.html>



Data Science Academy

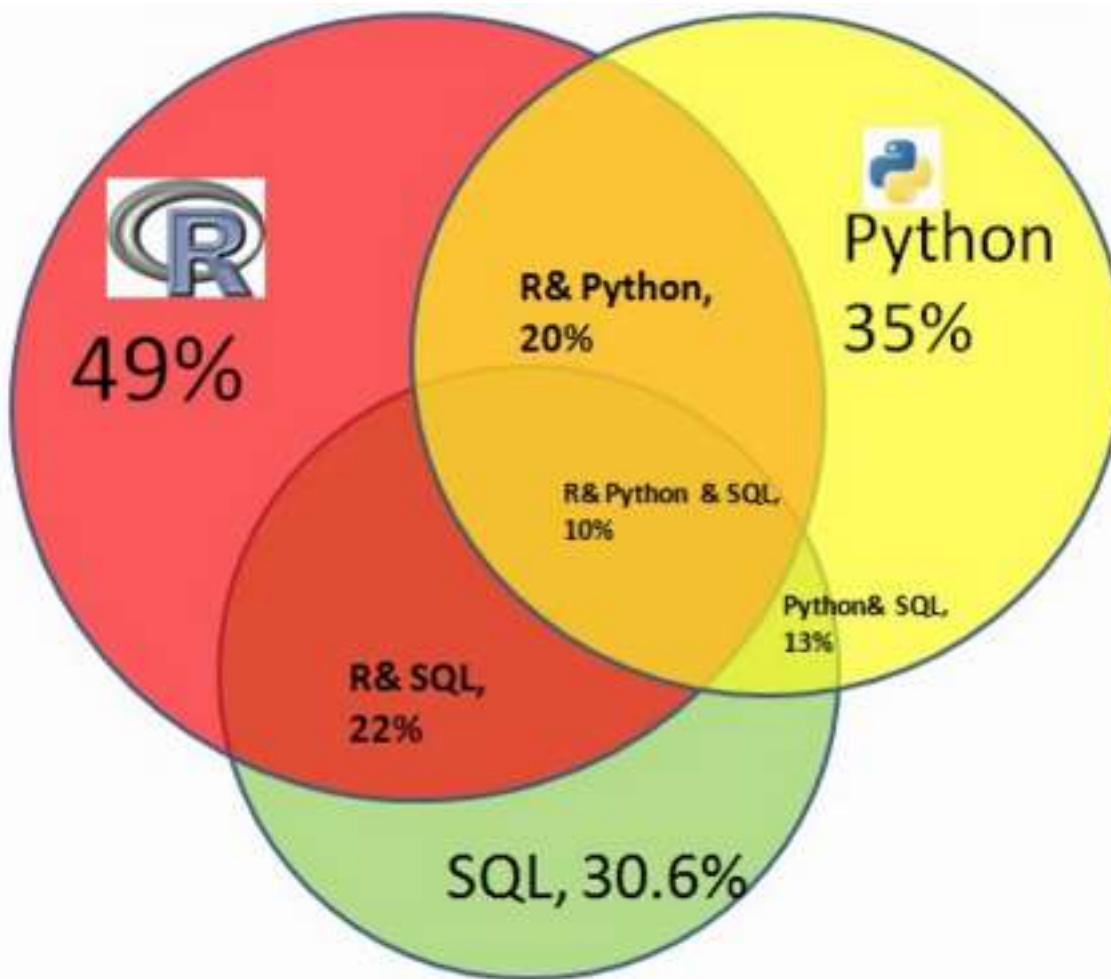
TOP 20 SKILLS OF A DATA SCIENTIST



RJMETRICS



Data Science Academy



Espera-se que o Cientista de Dados domine pelo menos uma destas 3 linguagens



Data Science Academy



IBM
SPSS

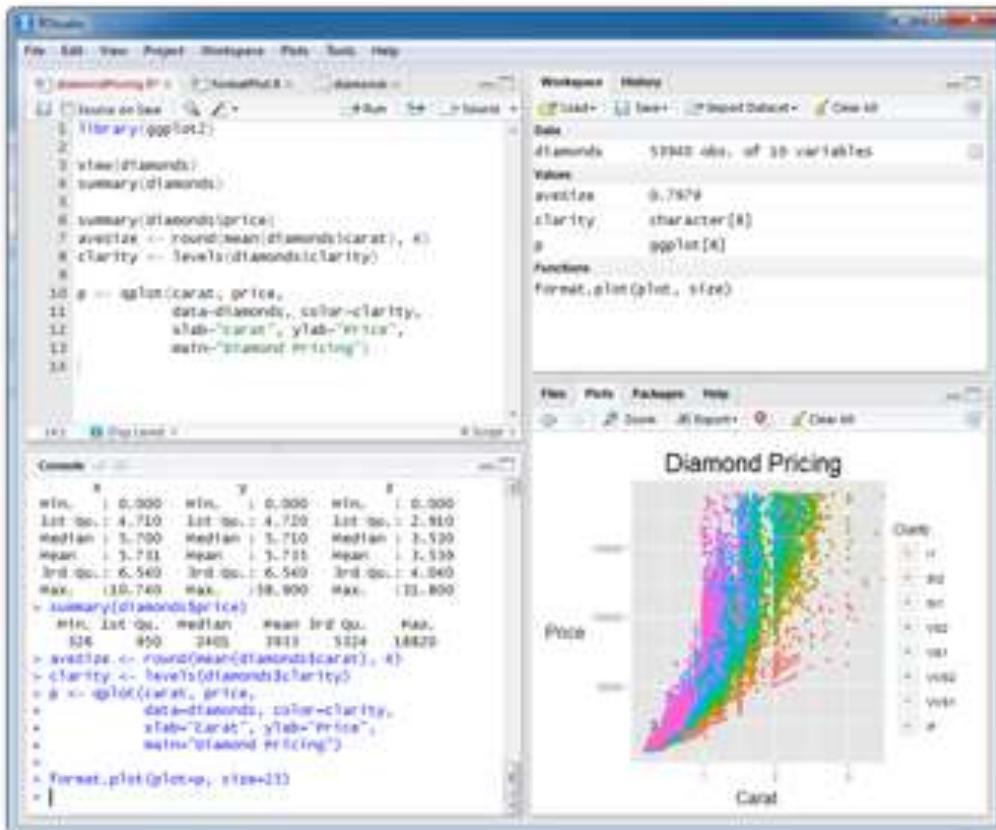
STATA®



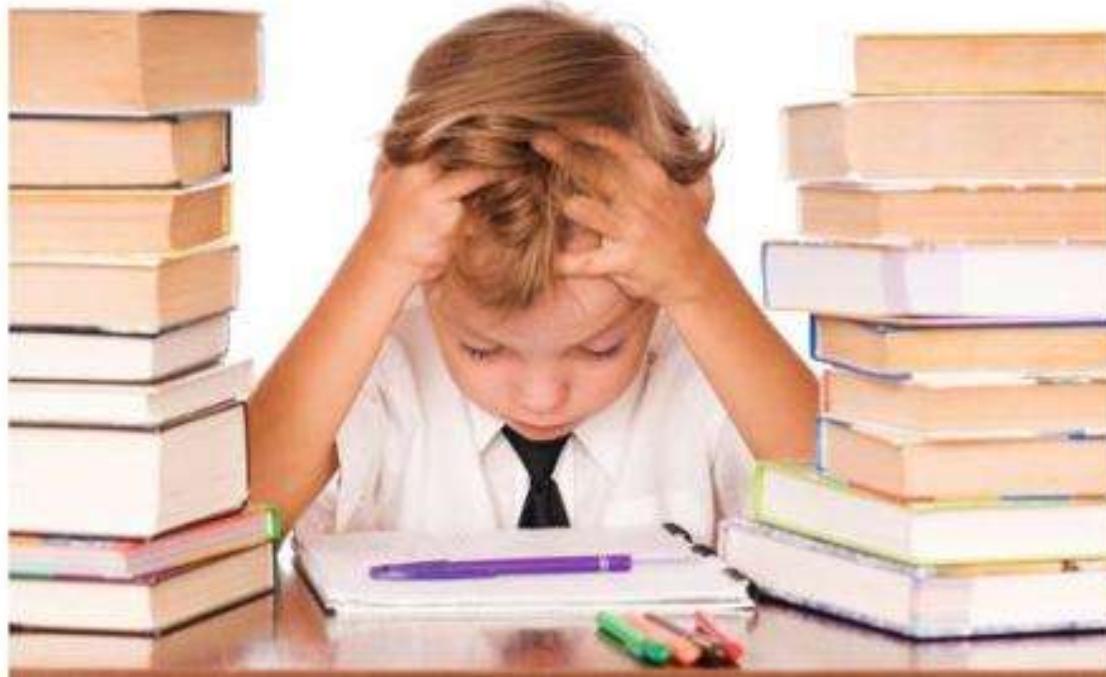
Data Science Academy

Sintaxe

"A" != "a"



Data Science Academy



Dificuldade



Data Science Academy



Vantagens e Desvantagens



Data Science Academy

Vantagens



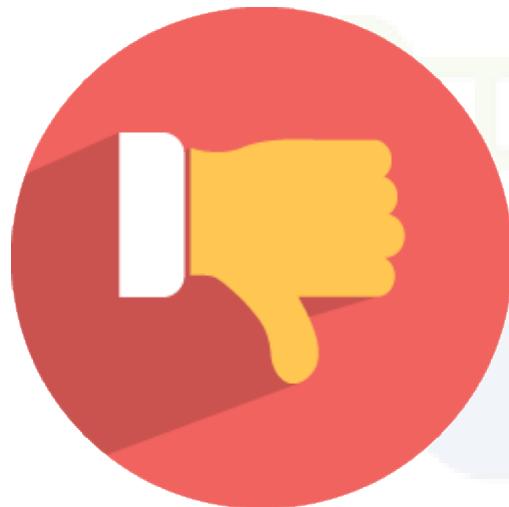
- Grande variedade de pacotes disponíveis
- Flexibilidade e Rapidez
- Machine Learning



Data Science Academy

Desvantagens

- Não há interface gráfica. Tudo é feito por linha de comando
- Limitações no uso de memória, principalmente com datasets muito grandes



Data Science Academy

OPEN SOURCE SOFTWARE



The word cloud includes the following words:

- OPEN: PRODUCT, BUSINESS, PRESS, PDF, BEER, MANY CALLED, HARDWARE, CODE, MOVE, LICENSES, CENTURY, RESEARCH, TERM, COMMUNITY, CULTURAL, PROCESS, EXAMPLE, COMPANIES, TECHNOLOGY, COMPUTER, PROJECTS, MANUFACTURERS, PLATFOM, CREATIVE, INTERNET, PUBLIC, SOCIETY, INTELLECTUAL, POLITICAL, NETWORK, SHARING, SIMILAR, DIGITAL.
- SOURCE: USED, DESIGN, WEB, VOL, LIMITED, DATA, BASED, OTHERS, LIKE, FILM, IDEA, ERIC, PUBLISHING, ANYONE, ALSO, INFORMATION, PHARMACEUTICALS, MESSAGEBOARDS, BLOGS, RAYMOND, INCLUDING.
- SOFTWARE: WORLD, USERS, SCIENCE, VARIOUS, ANOTHER, ACCESS, FREE, WORK, COPYRIGHT, SYSTEM, INNOVATION, USE, ONLINE, EDIT, COMMONS, ISSN, MODEL, PUBLISHING, ANYONE, ALSO, INFORMATION, PHARMACEUTICALS, MESSAGEBOARDS, BLOGS, RAYMOND, INCLUDING.
- OTHER: WORLD, USERS, SCIENCE, VARIOUS, ANOTHER, ACCESS, FREE, WORK, COPYRIGHT, SYSTEM, INNOVATION, USE, ONLINE, EDIT, COMMONS, ISSN, MODEL, PUBLISHING, ANYONE, ALSO, INFORMATION, PHARMACEUTICALS, MESSAGEBOARDS, BLOGS, RAYMOND, INCLUDING.



Data Science Academy



<https://www.r-consortium.org>



Data Science Academy



consortium

<https://www.r-consortium.org>

MEMBERS

R Foundation Member



Platinum



Gold



Silver



Data Science Academy



Oracle R Enterprise

Microsoft R Server



Oracle e Microsoft estão investindo pesado na linguagem R para suas soluções analíticas



Data Science Academy



Versatilidade



Data Science Academy



- **sqldf** - pacote que permite realizar queries SQL em dataframes no R
- **forecast** - modelar séries temporais
- **plyr** - dividir uma estrutura de dados em grupos e aplicar funções a cada grupo
- **stringr** - manipulação de strings
- **Database drivers** - RMongo, RODBC, RMySQL
- **ggplot2** - visualização de dados
- **qcc** - controle de qualidade estatístico
- **randomForest** - pacote para Machine Learning



Data Science Academy



Data Science Academy

www.datasciencacademy.com.br



www.datascienceacademy.com.br



obrigado!



Data Science Academy



Por que R e Microsoft Azure Machine Learning?



Data Science Academy

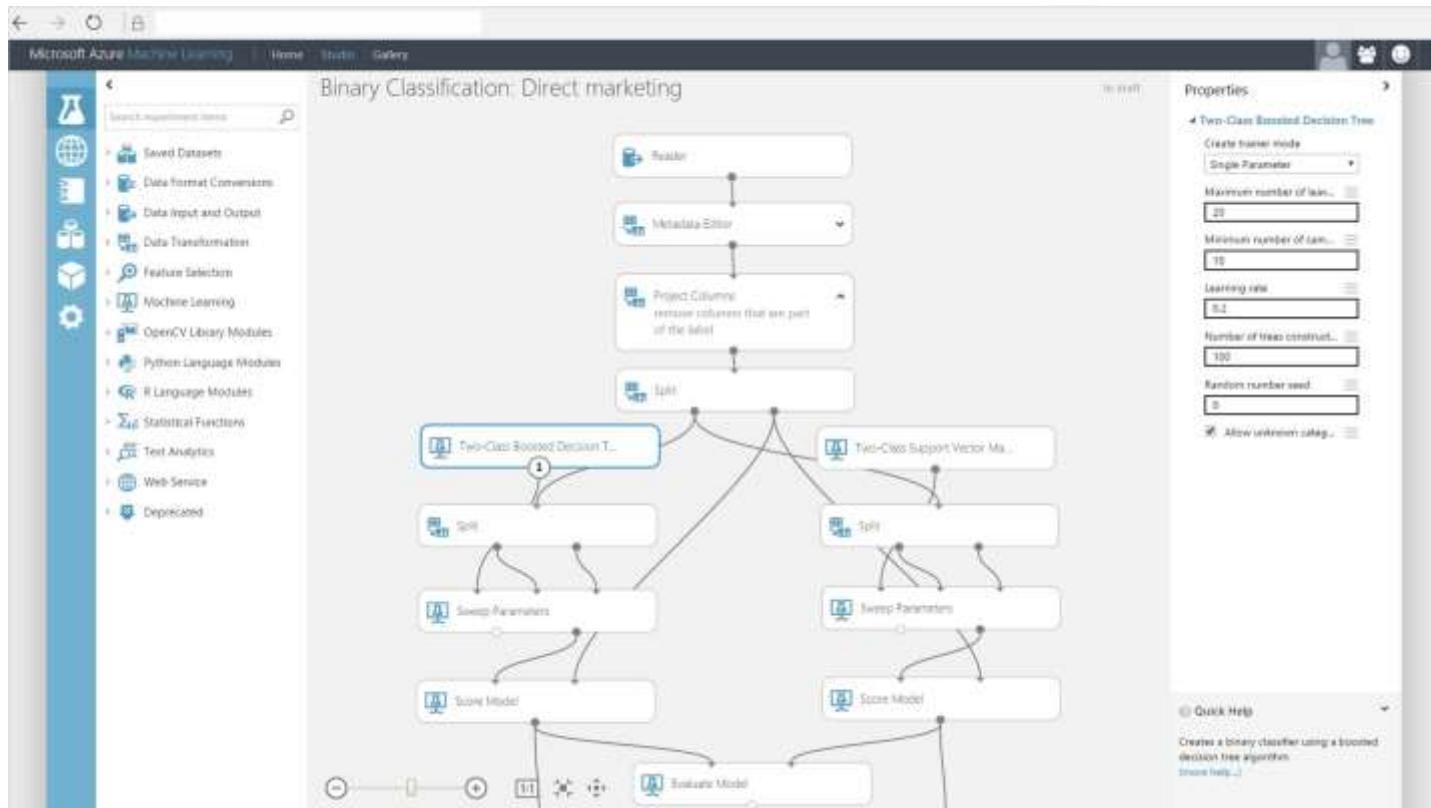


O que é o Azure Machine Learning?

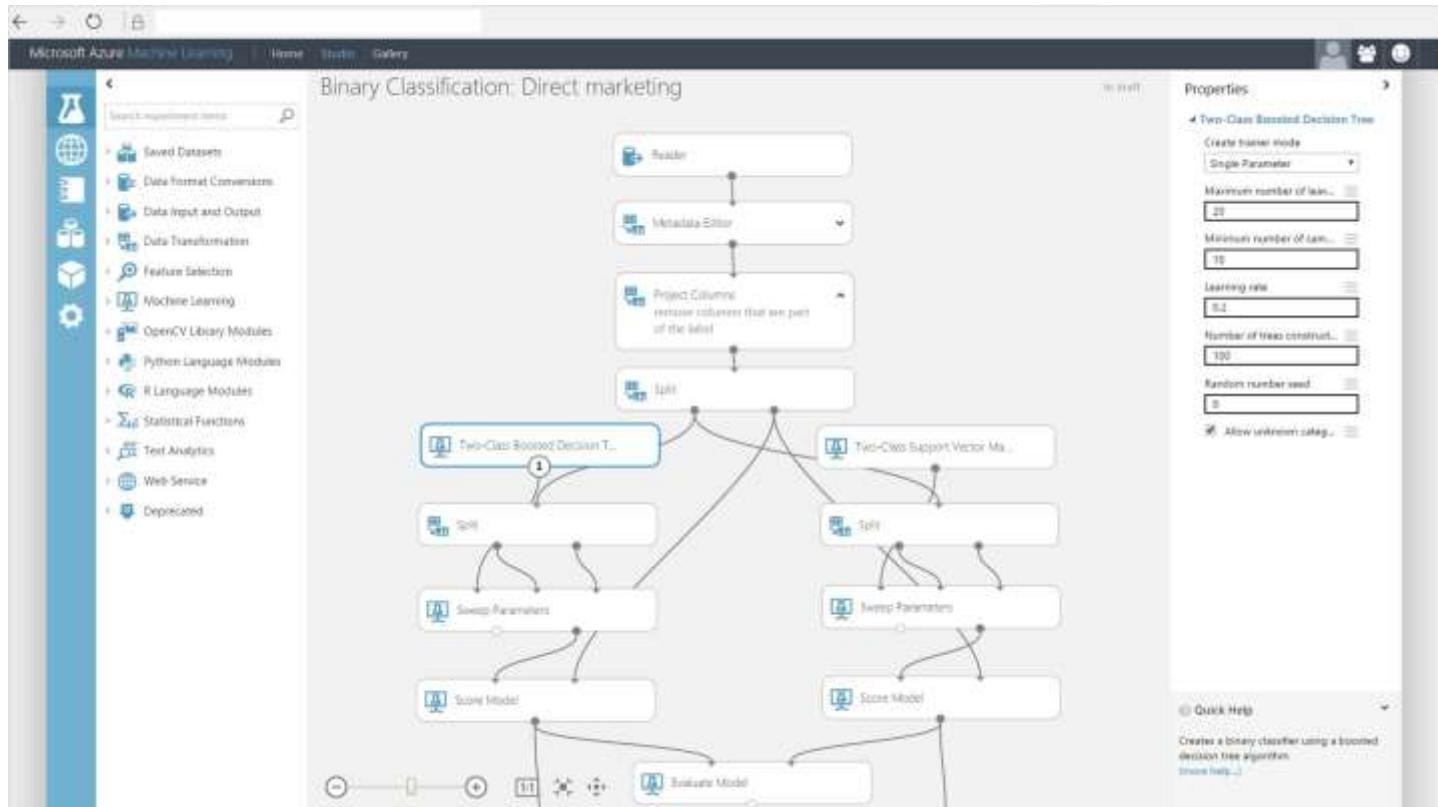


Data Science Academy

O que é o Azure Machine Learning?



Data Science Academy



Tudo na Nuvem



Data Science Academy

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The experiment name is "R word cloud". The experiment flow starts with a "conventions.csv" dataset input, which connects to an "Execute R Script" module. The properties pane on the right displays the R script code:

```
1 #adapted from https://gist.github.com/mollietaylor/3671518
2 library(wordcloud)
3 library(RColorBrewer)
4 library(png)
5
6 conventions <- maml.mapInputPort(1) # class: data.frame
7
8 png("cloud.png")
9 wordcloud(conventions$wordper25k, # words
10   conventions$democrats, # frequencies
11   scale = c(4,1), # size of largest and smallest words
12   colors = brewer.pal(9,"Blues"), # number of colors, palette
13   rot.per = 0) # proportion of words to rotate 90 degrees
14 dev.off()
15
16 wordcloud(conventions$wordper25k,
17   conventions$republicans,
18   scale = c(4,1),
19   colors = brewer.pal(9,"Reds"),
20   rot.per = 0)
```

The experiment status is "In draft" and it was last saved at 11:07:11. The sidebar on the left lists various experiment items: Saved Datasets, Trained Models, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, R Language Modules, Statistical Functions, and Text Analytics.

Machine Learning Studio



Data Science Academy

The screenshot shows the Microsoft Azure Machine Learning studio interface. The experiment name is "R word cloud". The experiment flow starts with a "conventions.csv" dataset input, which feeds into an "Execute R Script" module. The properties pane on the right displays the R script code:

```
1 #adapted from https://gist.github.com/mollietaylor/3671518
2 library(wordcloud)
3 library(RColorBrewer)
4 library(png)
5
6 conventions <- maml.mapInputPort(1) # class: data.frame
7
8 png("cloud.png")
9 wordcloud(conventions$wordper25k, # words
10   conventions$democrats, # frequencies
11   scale = c(4,1), # size of largest and smallest words
12   colors = brewer.pal(9,"Blues"), # number of colors, palette
13   rot.per = 0) # proportion of words to rotate 90 degrees
14 dev.off()
15
16 wordcloud(conventions$wordper25k,
17   conventions$republicans,
18   scale = c(4,1),
19   colors = brewer.pal(9,"Reds"),
20   rot.per = 0)
```

The experiment status is "In draft" and it was last saved at 11:07:11. The left sidebar lists various experiment items like Saved Datasets, Trained Models, and Data Input and Output.

Machine Learning Studio

KNIME



Data Science Academy

Módulos prontos para análises com R, Python e SQL

The screenshot shows the Microsoft Azure Machine Learning studio interface. On the left, a sidebar lists various modules: Home, Saved Datasets, Trained Models, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, R Language Modules, Statistical Functions, and Text Analytics. The main workspace is titled "R word cloud" and is marked as "In draft". It shows a flowchart with a "conventions.csv" dataset input connected to an "Execute R Script" module. The script content is as follows:

```
1 #adapted from https://gist.github.com/mollietaylor/3671518
2 library(wordcloud)
3 library(RColorBrewer)
4 library(png)
5
6 conventions <- maml.mapInputPort(1) # class: data.frame
7
8 png("cloud.png")
9 wordcloud(conventions$wordper25k, # words
10   conventions$democrats, # frequencies
11   scale = c(4,1), # size of largest and smallest words
12   colors = brewer.pal(9,"Blues"), # number of colors, palette
13   rot.per = 0) # proportion of words to rotate 90 degrees
14 dev.off()
15
16 wordcloud(conventions$wordper25k,
17   conventions$republicans,
18   scale = c(4,1),
19   colors = brewer.pal(9,"Reds"),
20   rot.per = 0)
```

The "Properties" pane on the right shows the "Execute R Script" properties, including the R script code, a "Random Seed" of 42, and "Experiment Properties" for start and end times.



Data Science Academy

The screenshot shows the Microsoft Azure Machine Learning studio interface. At the top, there's a navigation bar with links for Why Azure, Products, Documentation, Pricing, Partners, Blog, Resources, and Support. On the right of the navigation bar is a "FREE ACCOUNT" button. Below the navigation bar, the main area has a heading "Simple, scalable, cutting edge" and a sub-heading "A fully managed cloud service that enables you to easily build, deploy, and share predictive analytics solutions." The central part of the screen displays a workflow titled "Binary Classification: Direct marketing". The workflow consists of several nodes connected by arrows: "Reads", "Metadata Editor", "Project Columns", "Split", "Two-Class Boosted Decision Tree", "Two-Class Support Vector Ma.", and "Score". To the left of the workflow is a sidebar with a tree view of available modules under "Machine Learning", including "Saved Datasets", "Data Format Converters", "Data Input and Output", "Data Transformation", "Feature Selection", "OpenCV Library Modules", "Python Language Modules", "R Language Modules", "Statistical Functions", "Text Analytics", "Web Service", and "Deprecated". On the right side of the workflow, there's a "Properties" panel for the "Two-Class Boosted Decision Tree" node, which includes settings for "Data trainer mode" (set to "Single Parameter"), "Maximum number of bins" (20), "Maximum number of splits" (10), "Learning rate" (0.2), "Number of trees constructed" (100), "Random number seed" (0), and a checkbox for "Allow unknown categories".

Diversos Módulos de ML

Azure ML Toolkit



Data Science Academy

Quick Evaluation

Guest Workspace

8-hour trial

No sign-in required.

[Enter](#)

- No hassle instant access
- Stock sample datasets
- ML models built in minutes
- Full range of ML algorithms

Most Popular

Free Workspace

\$0/month

Don't already have a Microsoft account?
Simply [sign up here](#).

[Sign In](#)

- Free access that never expires
- 10 GB storage on us
- R and Python scripts support
- Predictive web services

Enterprise Grade

Standard Workspace

\$9.99/month

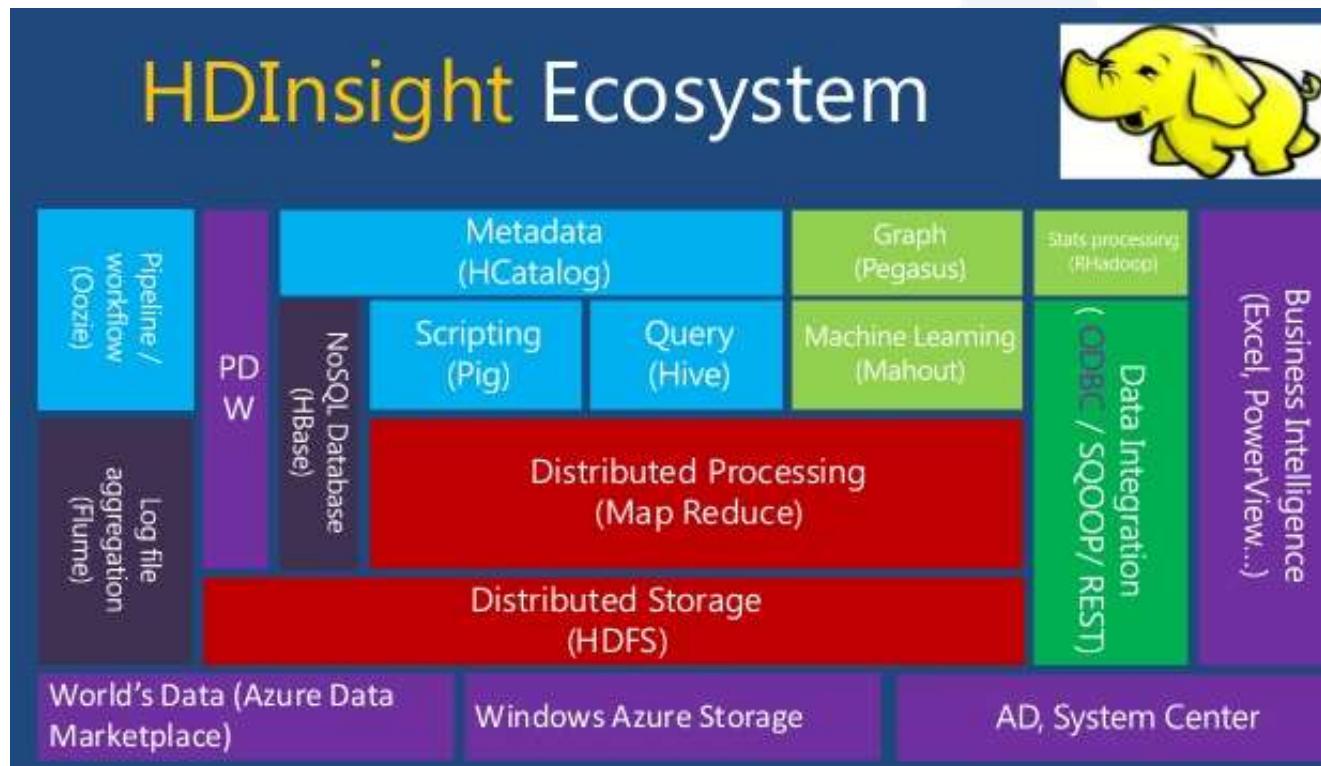
Azure subscription required
Other charges may apply. [Read more](#).

[Create Workspace](#)

- Full SLA Support
- Bring your own Azure storage
- Parallel graph execution
- Elastic Web Service endpoints



Data Science Academy



Azure HDInsight



Data Science Academy

Virtual Machines



Provision Windows and Linux virtual machines in minutes

App Service



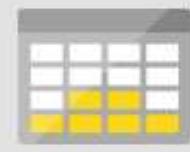
Create web and mobile apps for any platform and any device

SQL Database



Managed relational SQL Database-as-a-service

Storage



Durable, highly available, and massively scalable cloud storage

Cloud Services



Create highly available, infinitely scalable cloud applications and APIs

Document DB



Managed NoSQL document database-as-a-service

Active Directory

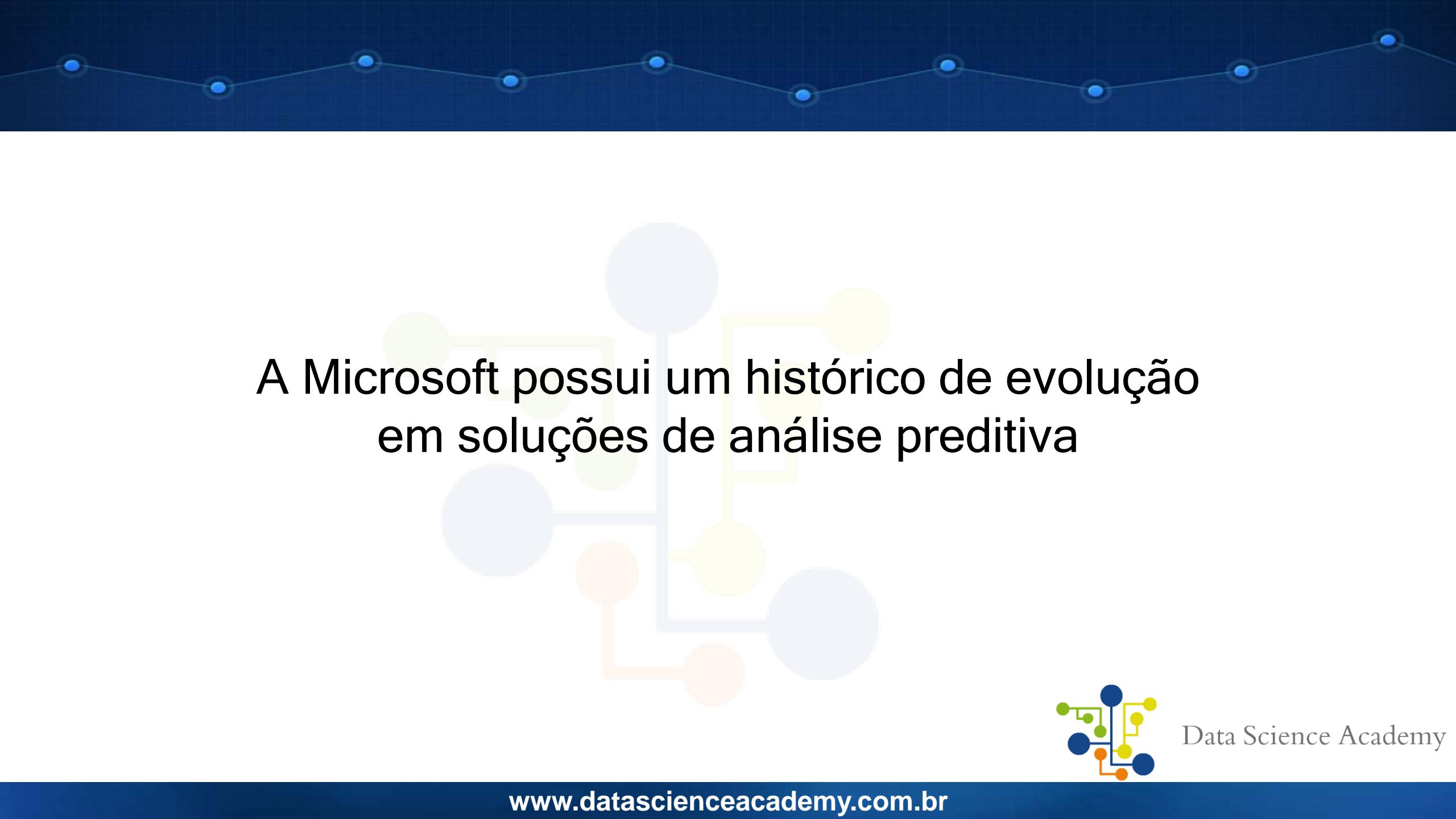


Synchronize on-premises directories and enable single sign-on

Backup



Simple and reliable server backup to the cloud



A Microsoft possui um histórico de evolução
em soluções de análise preditiva

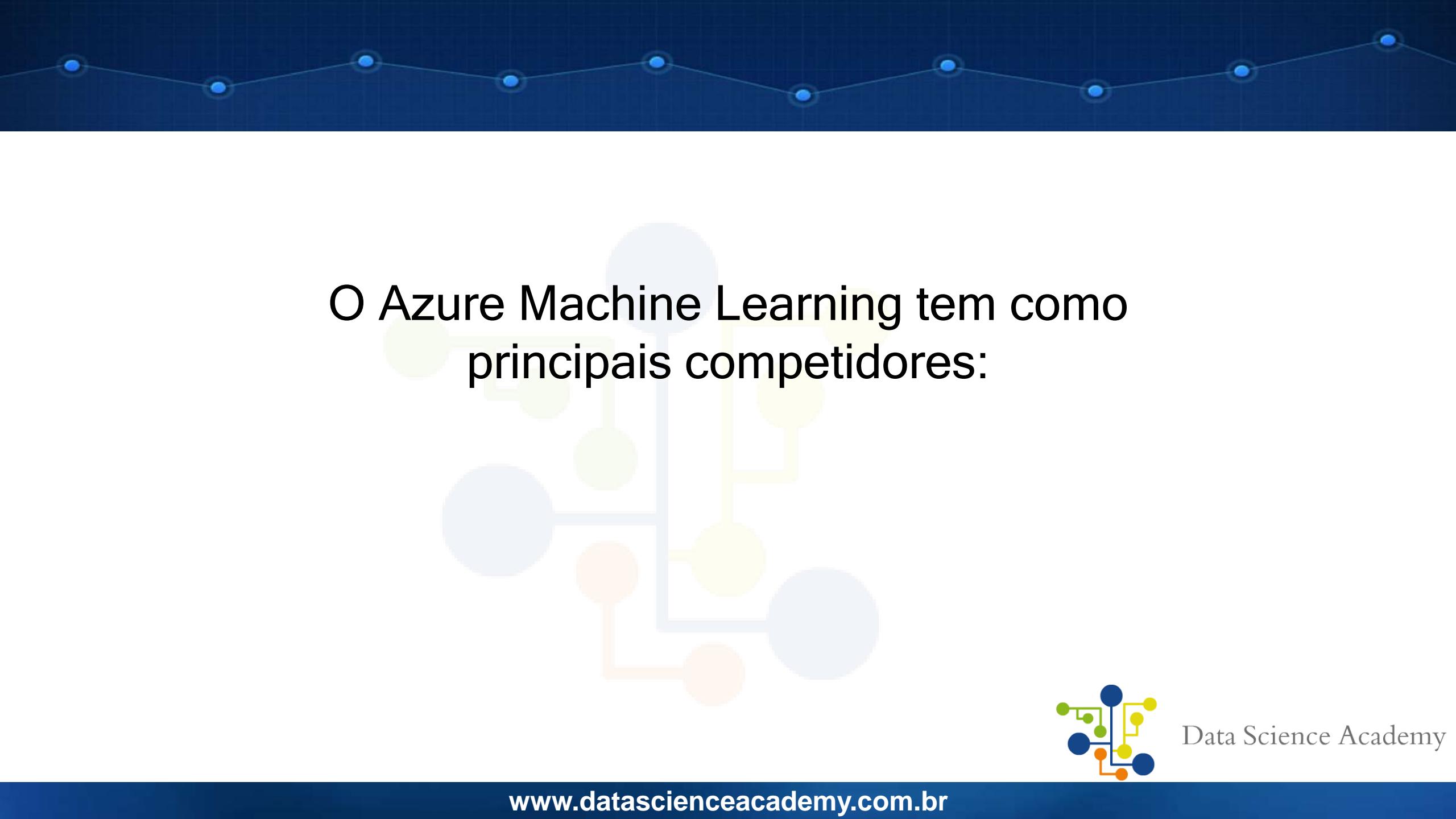


Data Science Academy

Ano	Produto
1999	Filtros de e-mail e Spam no Microsoft Outlook
2004	Começou a incorporar Machine Learning em seus mecanismos de busca
2005	SQL Server 2005 trouxe opções de data mining
2008	O serviço de busca Bing incorporou Machine Learning para previsão de tráfego
2010	Kinect incorporou o reconhecimento de gestos e filtro de ruídos
2014	Primeira versão do Azure Machine Learning
2014	Microsoft lança a plataforma Cortana Intelligence Suite, totalmente baseada em Machine Learning



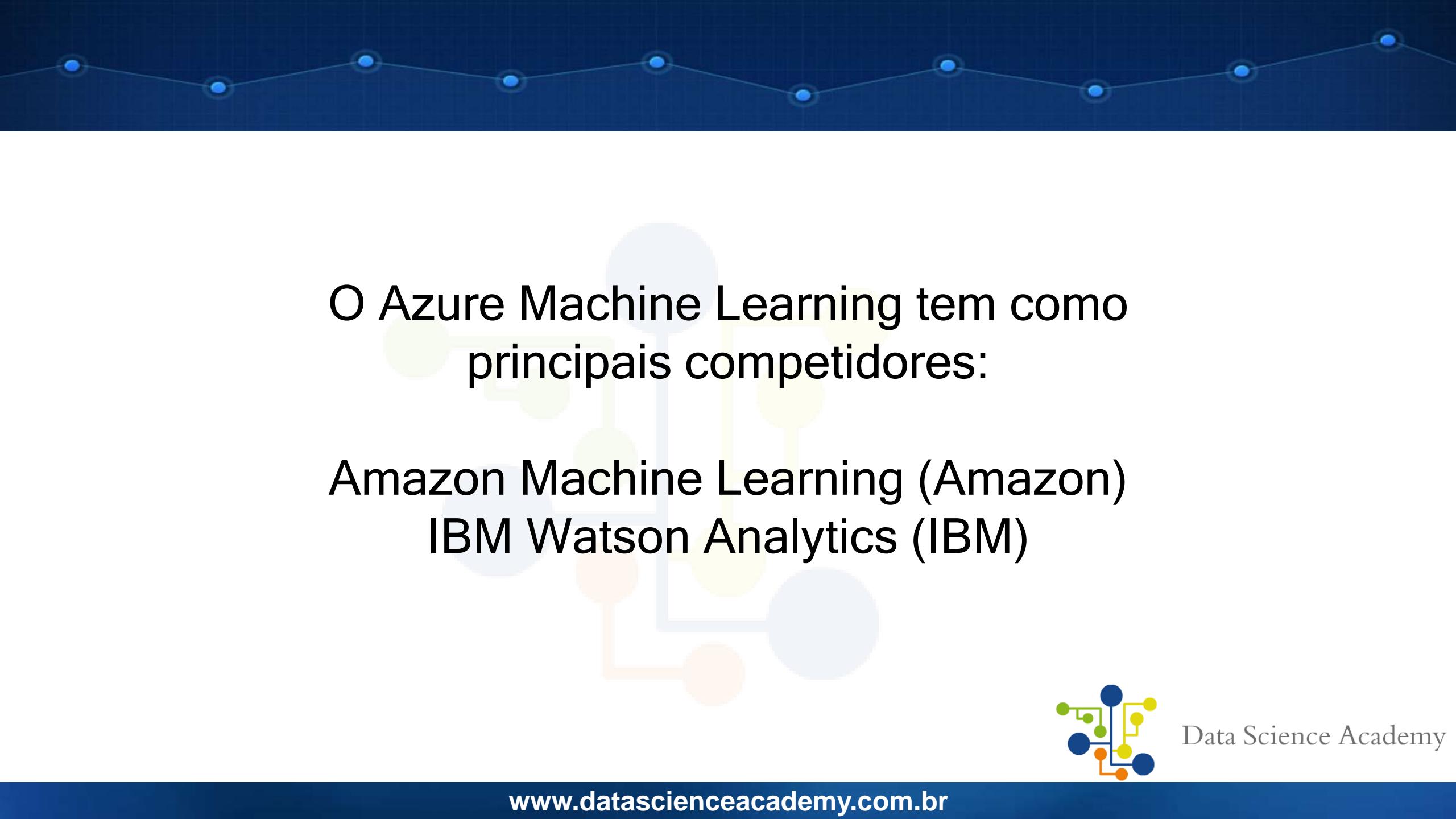
Data Science Academy



O Azure Machine Learning tem como principais competidores:



Data Science Academy



O Azure Machine Learning tem como principais competidores:

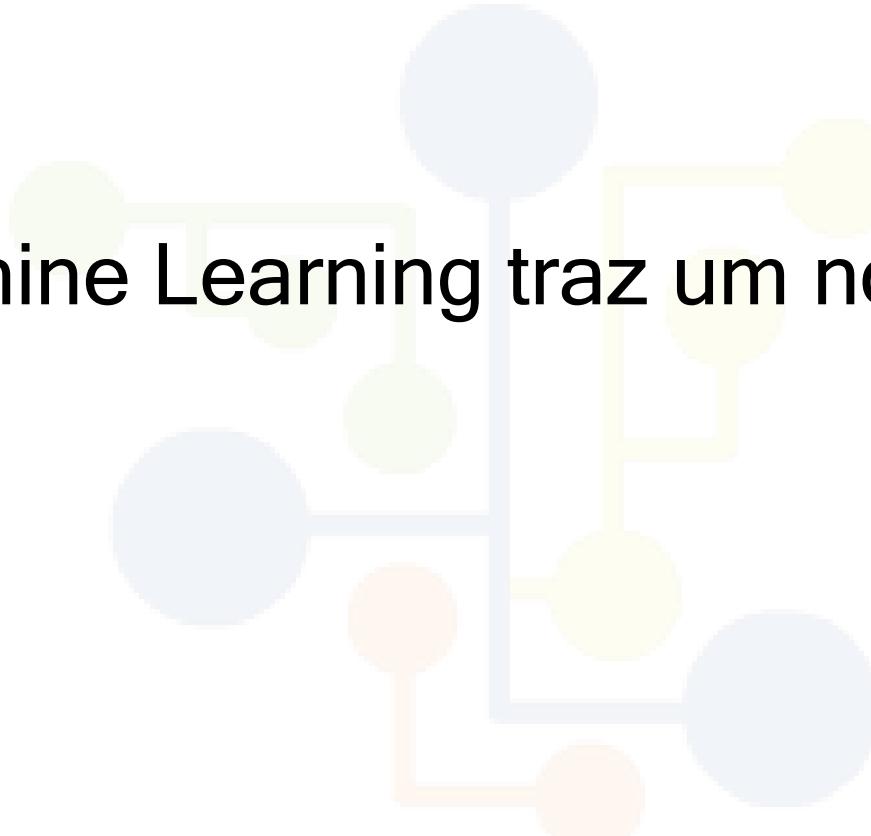
Amazon Machine Learning (Amazon)
IBM Watson Analytics (IBM)



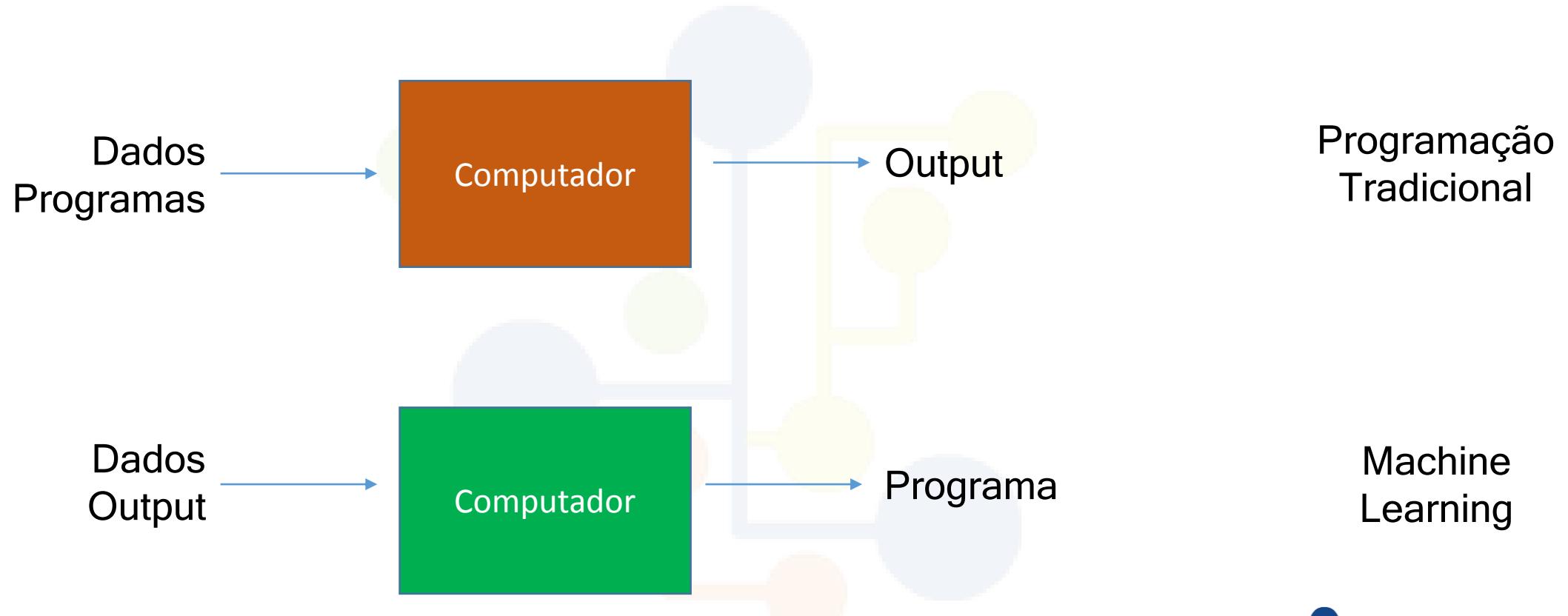
Data Science Academy



Machine Learning traz um novo paradigma



Data Science Academy



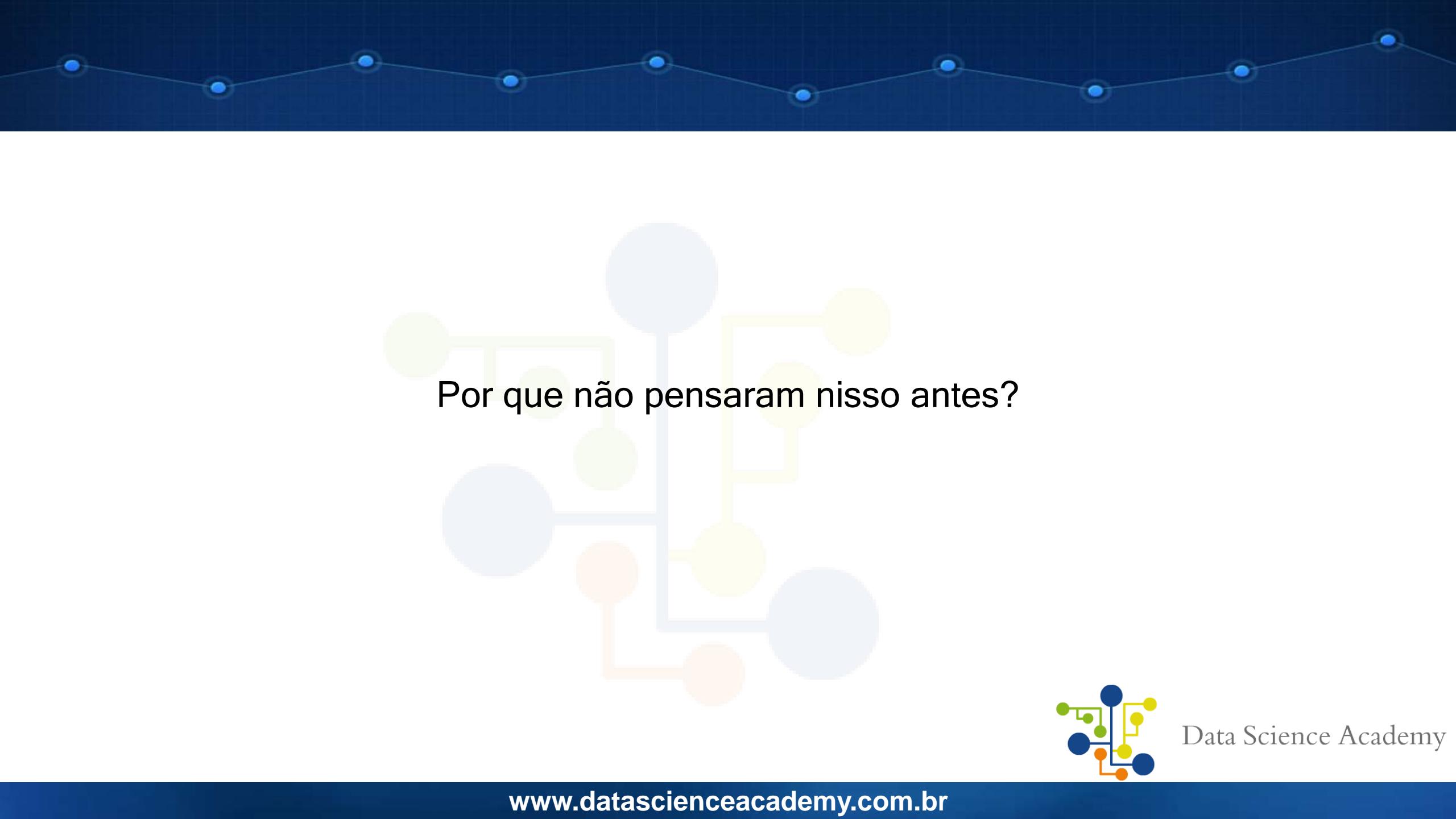
Data Science Academy



A grande vantagem deste novo programa gerado, é que ele pode agora fazer previsões em novos conjuntos de dados, pois ele já aprendeu como fazer



Data Science Academy



Por que não pensaram nisso antes?



Data Science Academy



Tempestade Perfeita

- Crescimento exponencial do volume de dados
- Preço baixo do armazenamento de dados
- Alta capacidade de processamento dos computadores
- E o surgimento do Big Data Analytics



Data Science Academy



Análise Preditiva (Predictive Analytics)



Data Science Academy



Dados são combustível
para a Análise Preditiva



Data Science Academy



E podemos ver a Análise Preditiva em ação todos os dias

- Filtros de Spam dos e-mails
- Programas de concessão de financiamento para casa própria
- Padrões de reconhecimento (voz, imagem)
- Seguro de Vida
- Detecção de Fraudes em Cartões de Crédito
- Controle de Voos
- Engines de Busca na Internet



Data Science Academy



E por que R e Azure Machine Learning?



Data Science Academy

E por que R e Azure Machine Learning?

- Azure Machine Learning é visual, enquanto R é linha de comando (se complementam)
- O R permite estender as funcionalidades do Azure ML
- Esta combinação permite escalabilidade e mais opções de manipulação de dados e criação de modelos
- O Machine Learning Studio traz diversos módulos em R, prontos para uso



Data Science Academy

Machine Learning Workflow



Data Science Academy

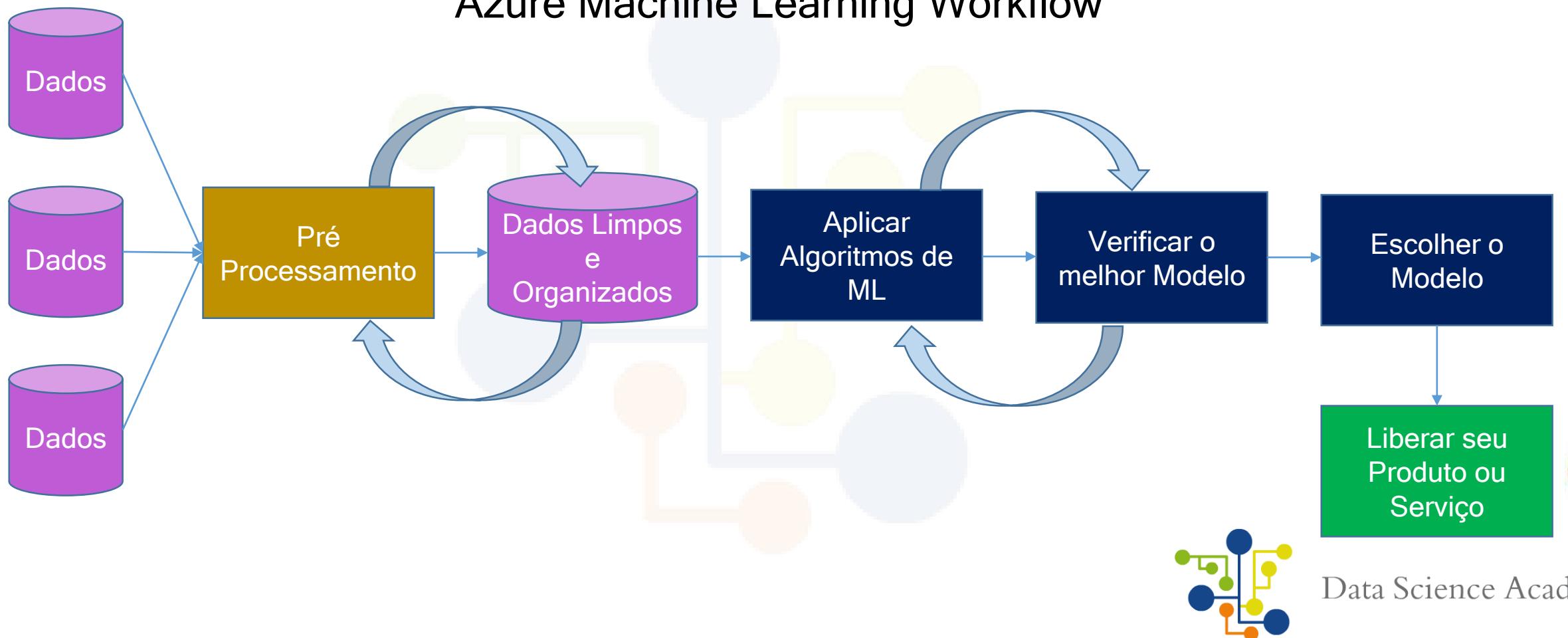
Machine Learning Workflow

Se você não sabe onde quer ir, qualquer caminho serve.
(Alice no País das Maravilhas)



Data Science Academy

Azure Machine Learning Workflow

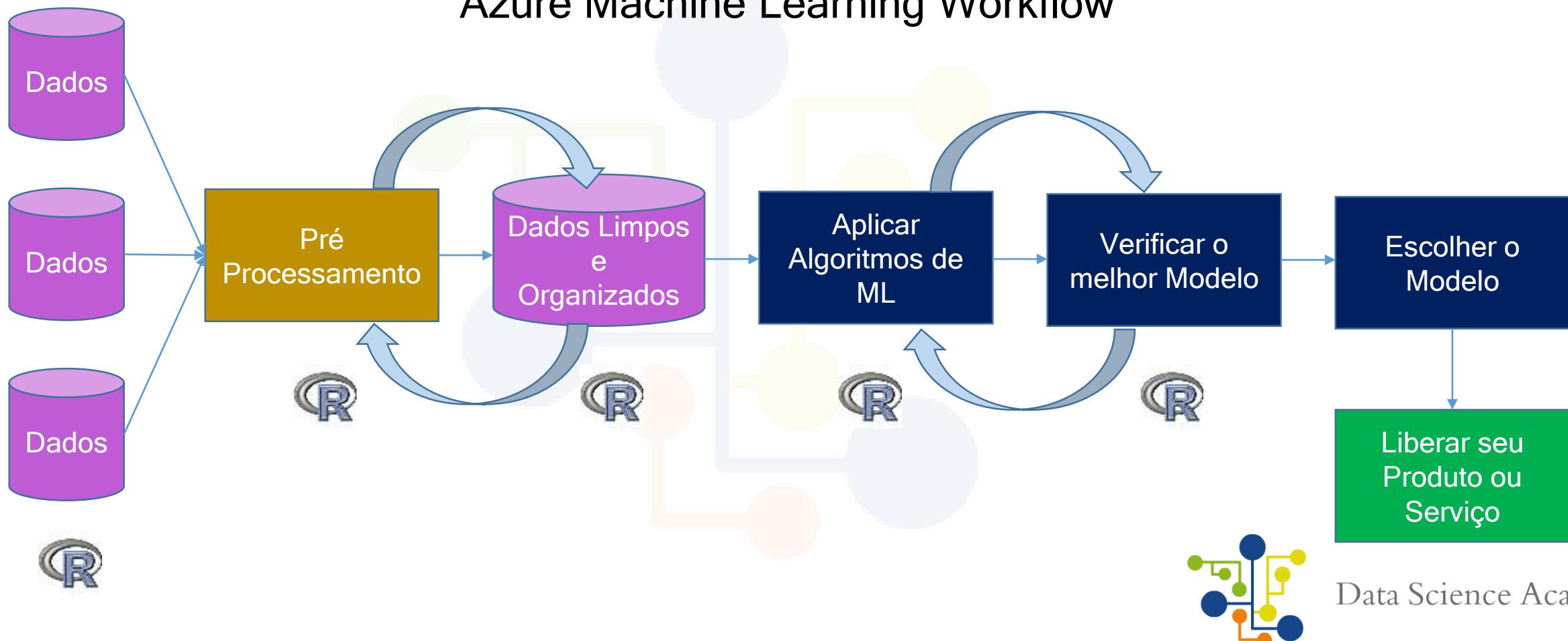


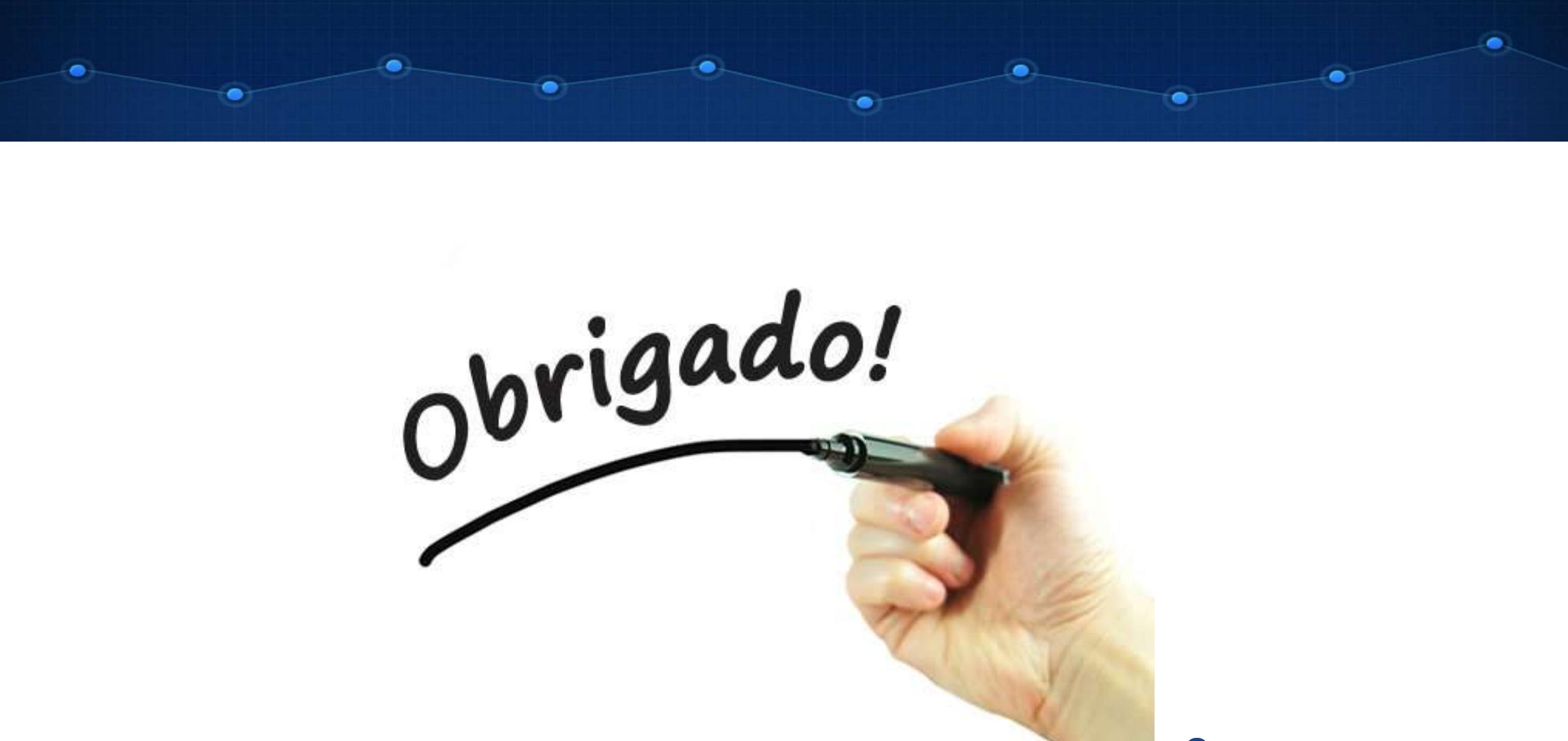
Azure Machine Learning Workflow



Data Science Academy

Azure Machine Learning Workflow





obrigado!



Data Science Academy



Qual a importância do Big Data Analytics?



Data Science Academy



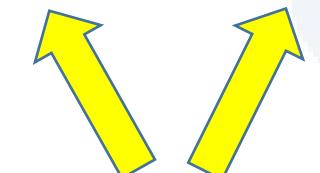
Dados x Informação x Conhecimento x Inteligência



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário	Cargo
Aristóteles	32	R\$ 7.500,00	Analista Financeiro
Sócrates	41	R\$ 9.300,00	Gerente Financeiro



Informação



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

As pessoas trabalham com informações mas a tecnologia armazena dados



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário	Cargo
Aristóteles	32	R\$ 7.500,00	Analista Financeiro
Sócrates	41	R\$ 9.300,00	Gerente Financeiro

Conhecimento



Data Science Academy

Dados x Informação x Conhecimento x Inteligência



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário	Cargo
Aristóteles	32	R\$ 7.500,00	Analista Financeiro
Sócrates	41	R\$ 9.300,00	Gerente Financeiro

Inteligência - capacidade de resolver problemas, usando o conhecimento, através das informações disponíveis

Por exemplo: por que os funcionários do setor financeiro tem queda de performance após 1 ano trabalhando na empresa



Data Science Academy

Dados x Informação x Conhecimento x Inteligência



Big Data Analytics

Extrair conhecimento a partir dos dados



Data Science Academy

Dados x Informação x Conhecimento x Inteligência



Machine Learning

Ensinar algoritmos a usar inteligência, ou seja, usar o conhecimento para resolver problemas



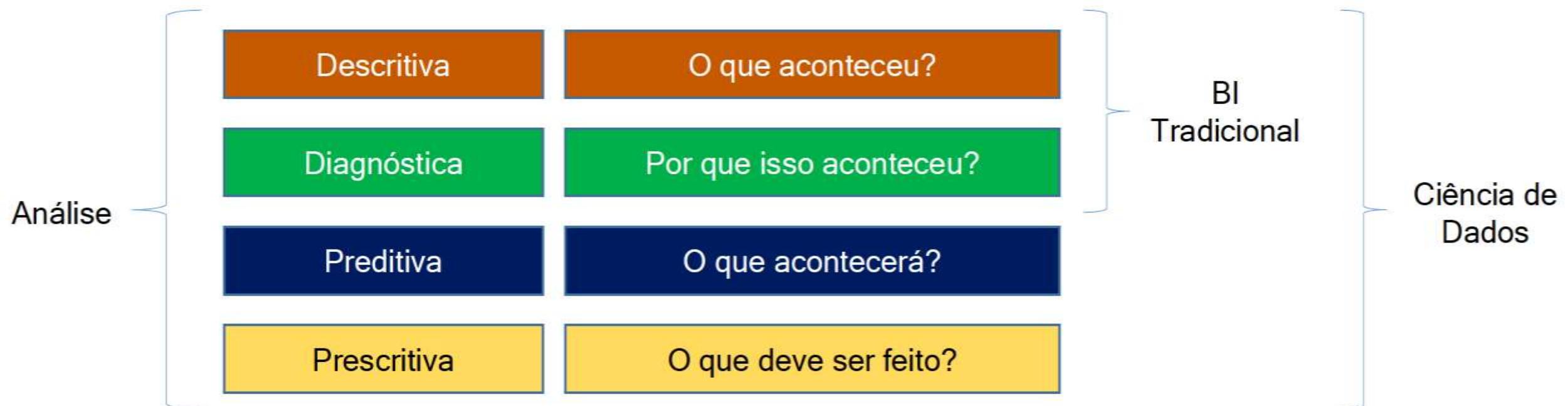
Data Science Academy

Evolução dos Sistemas Analíticos



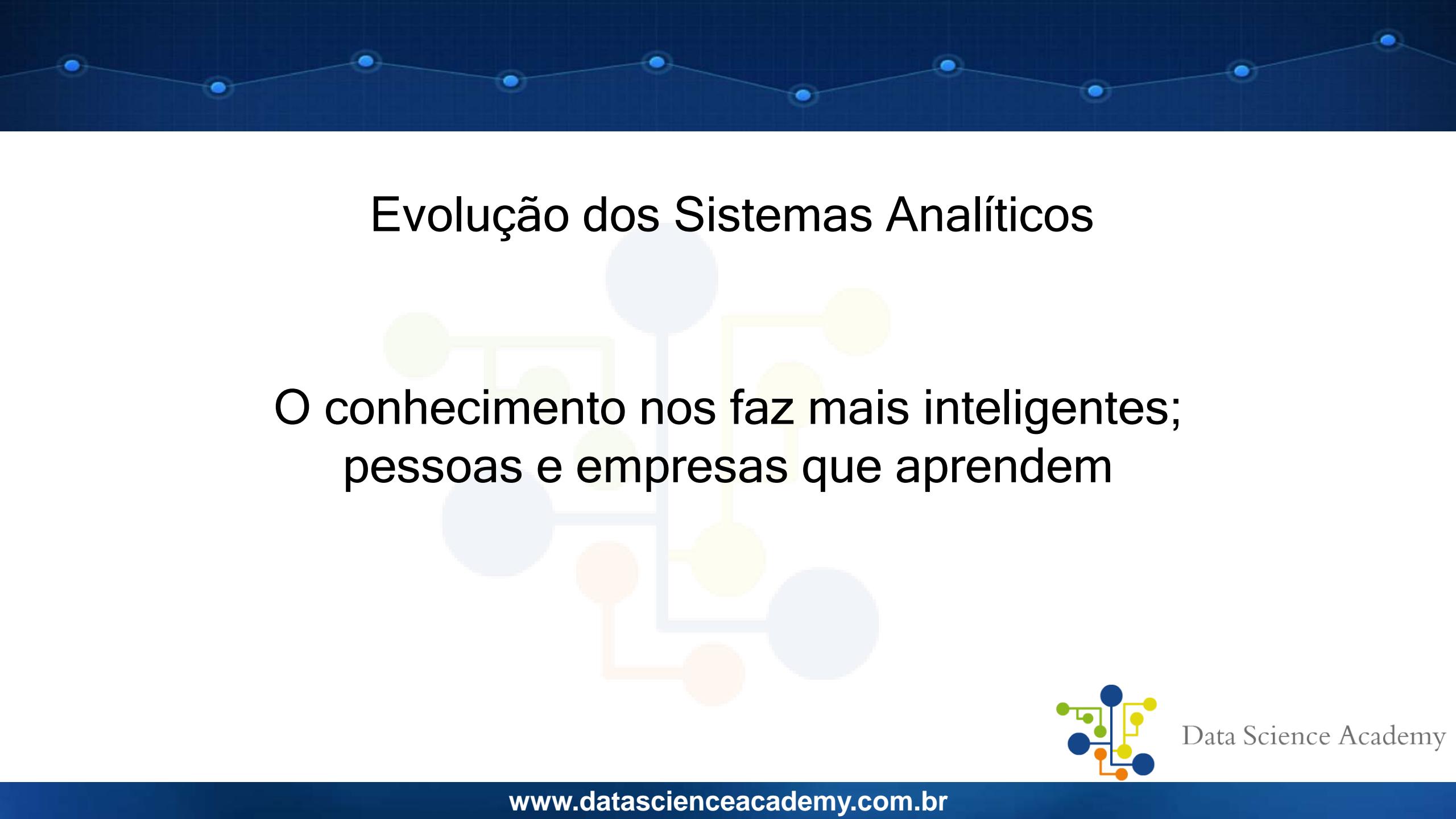
Data Science Academy

Evolução dos Sistemas Analíticos



Data Science Academy

Evolução dos Sistemas Analíticos

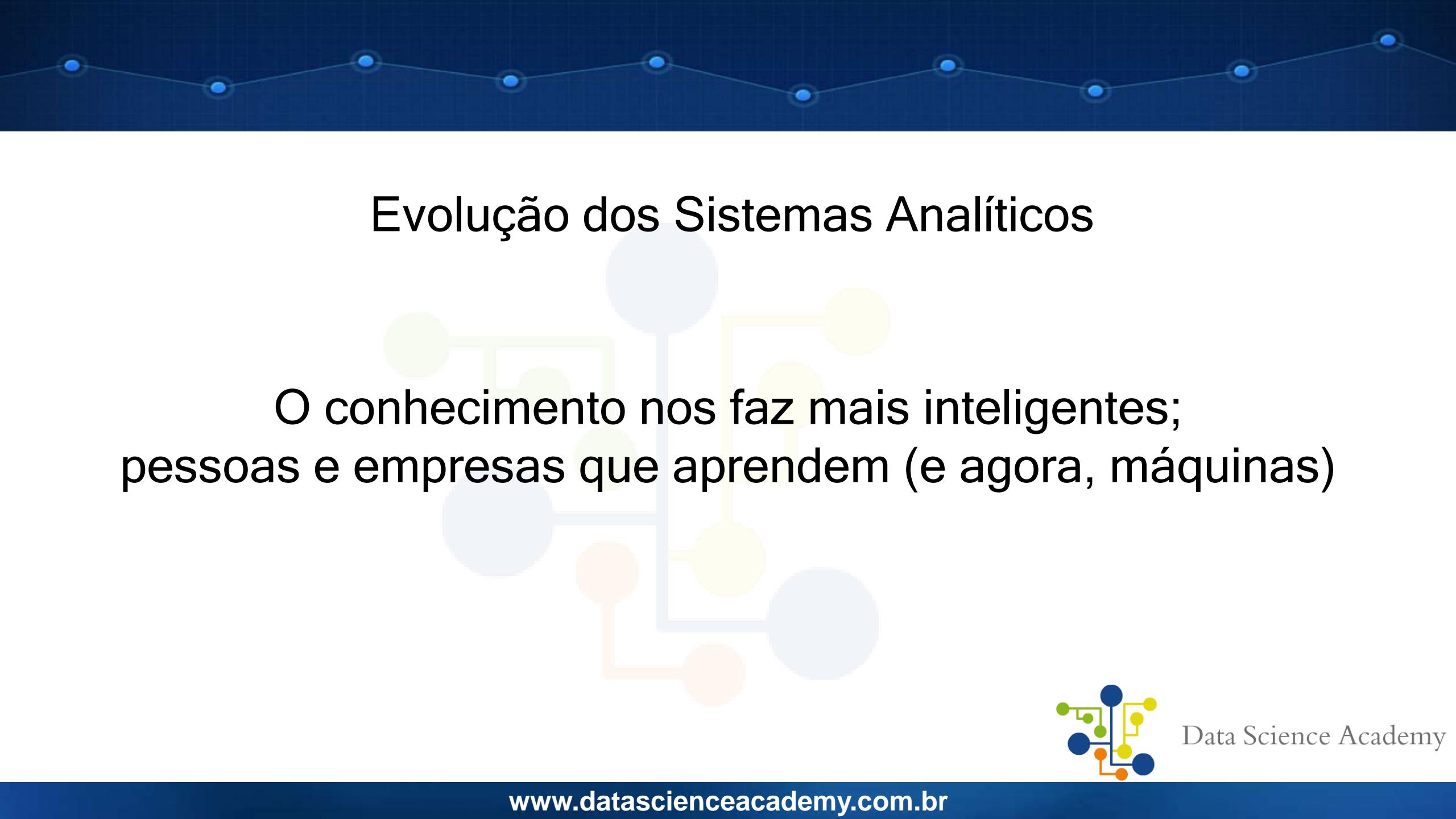


O conhecimento nos faz mais inteligentes;
pessoas e empresas que aprendem



Data Science Academy

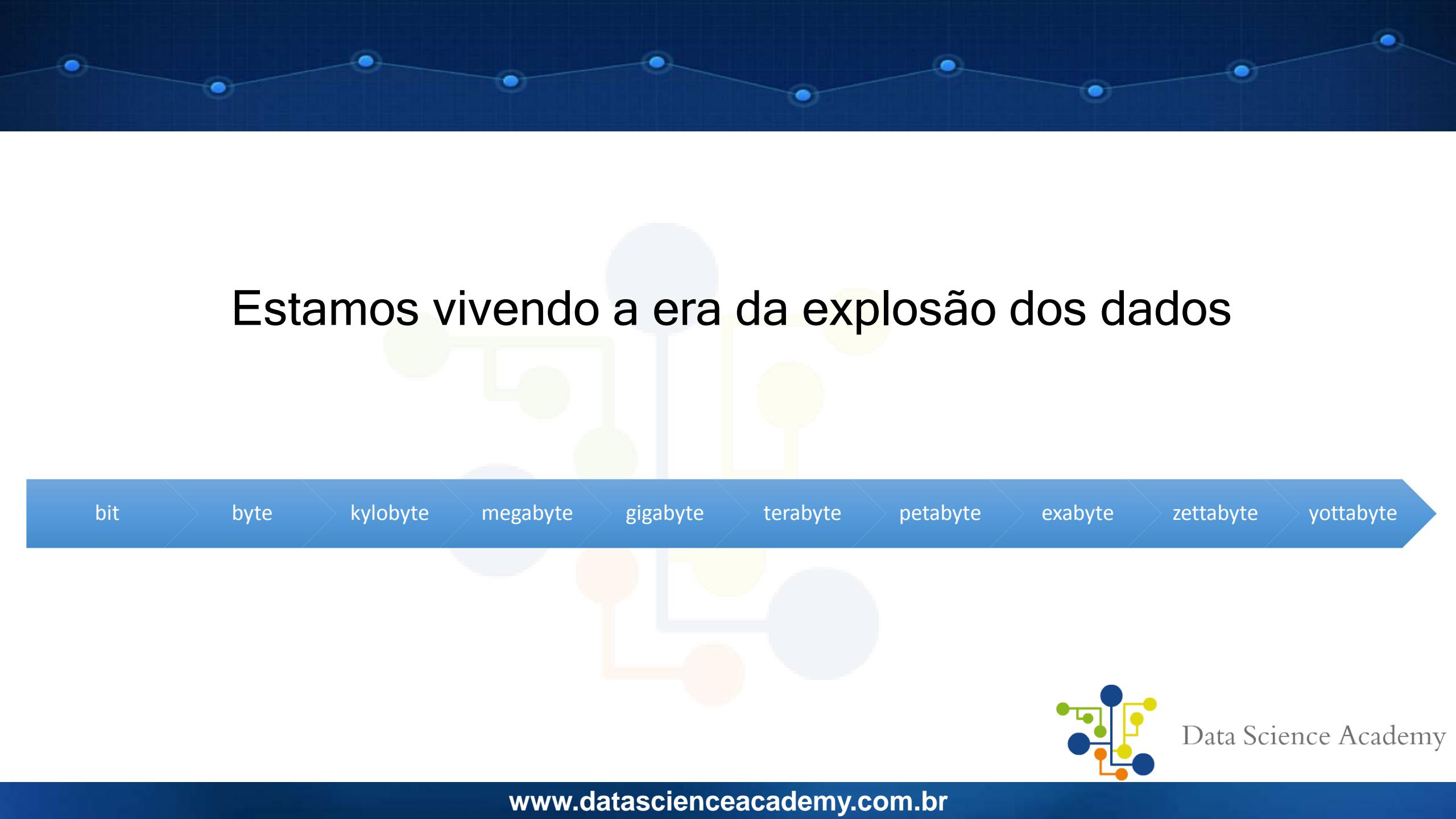
Evolução dos Sistemas Analíticos



O conhecimento nos faz mais inteligentes;
pessoas e empresas que aprendem (e agora, máquinas)



Data Science Academy



Estamos vivendo a era da explosão dos dados

bit

byte

kilobyte

megabyte

gigabyte

terabyte

petabyte

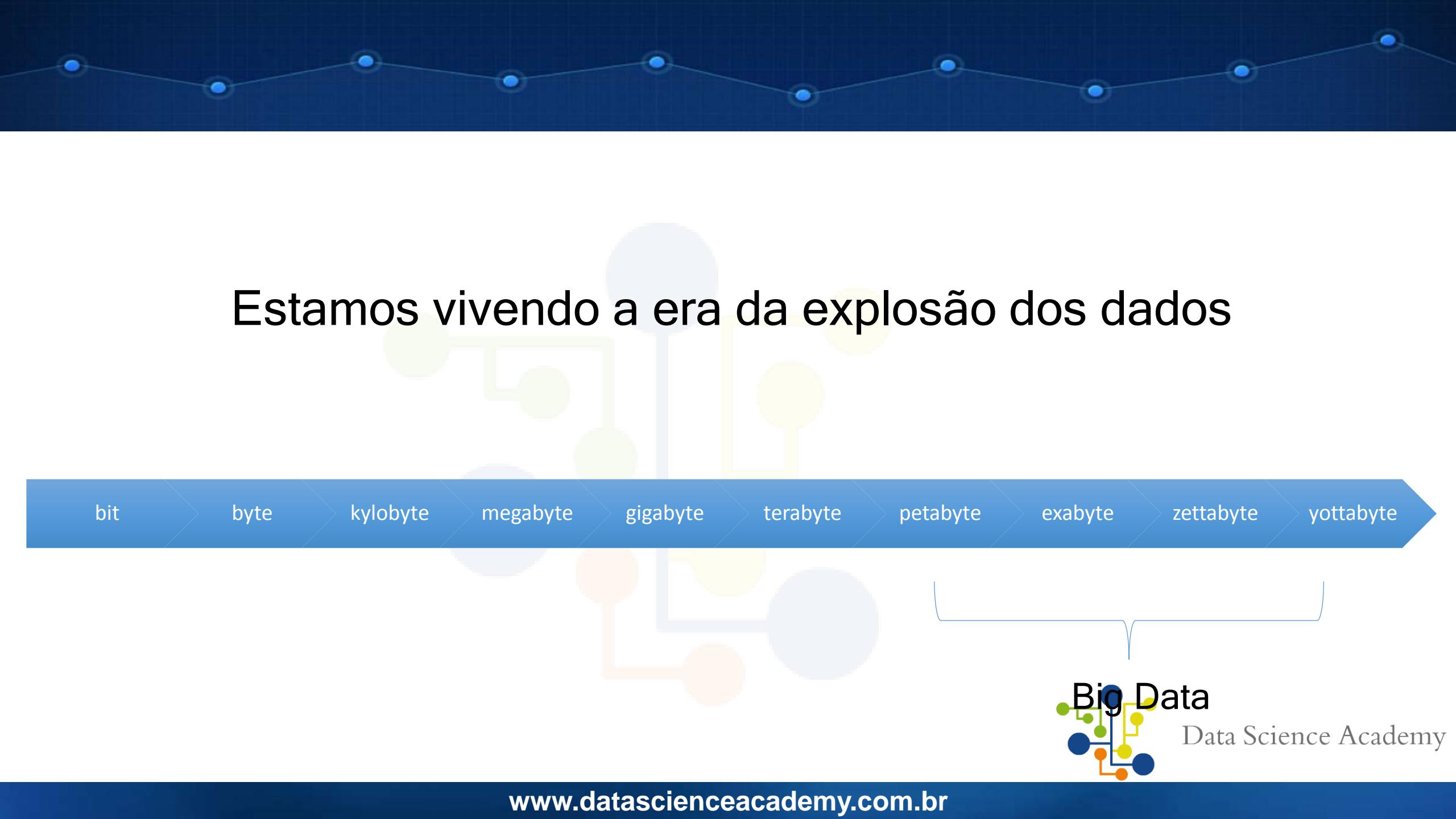
exabyte

zettabyte

yottabyte



Data Science Academy



Estamos vivendo a era da explosão dos dados

bit

byte

kilobyte

megabyte

gigabyte

terabyte

petabyte

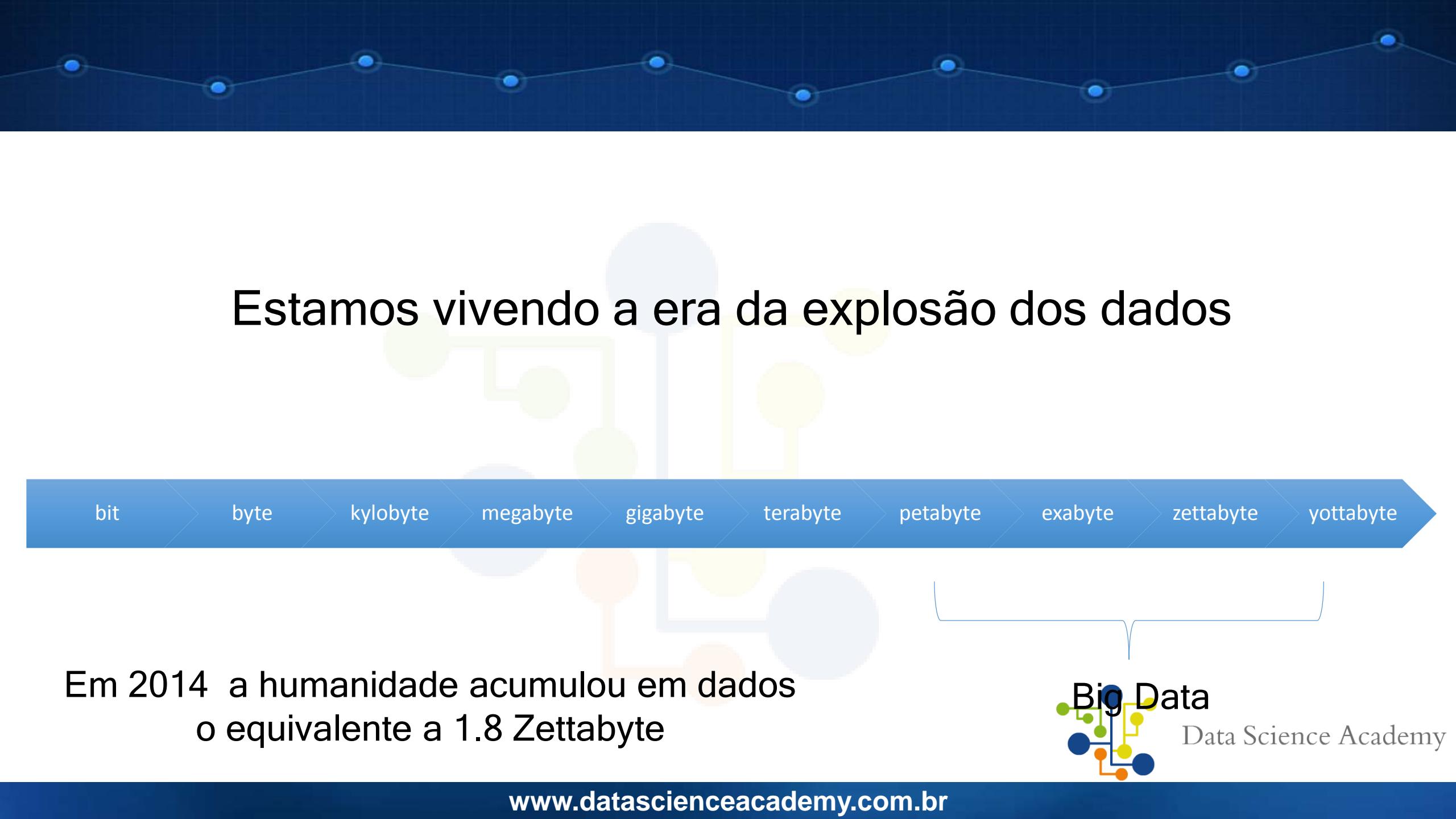
exabyte

zettabyte

yottabyte



Data Science Academy



Estamos vivendo a era da explosão dos dados

bit byte kylobyte megabyte gigabyte terabyte petabyte exabyte zettabyte yottabyte

Em 2014 a humanidade acumulou em dados o equivalente a 1.8 Zettabyte





O que é Big Data Analytics?



O objetivo é simples: melhorar seus processos de trabalho e adquirir insights valiosos acerca das tendências de mercado, comportamento dos consumidores e suas expectativas



Data Science Academy

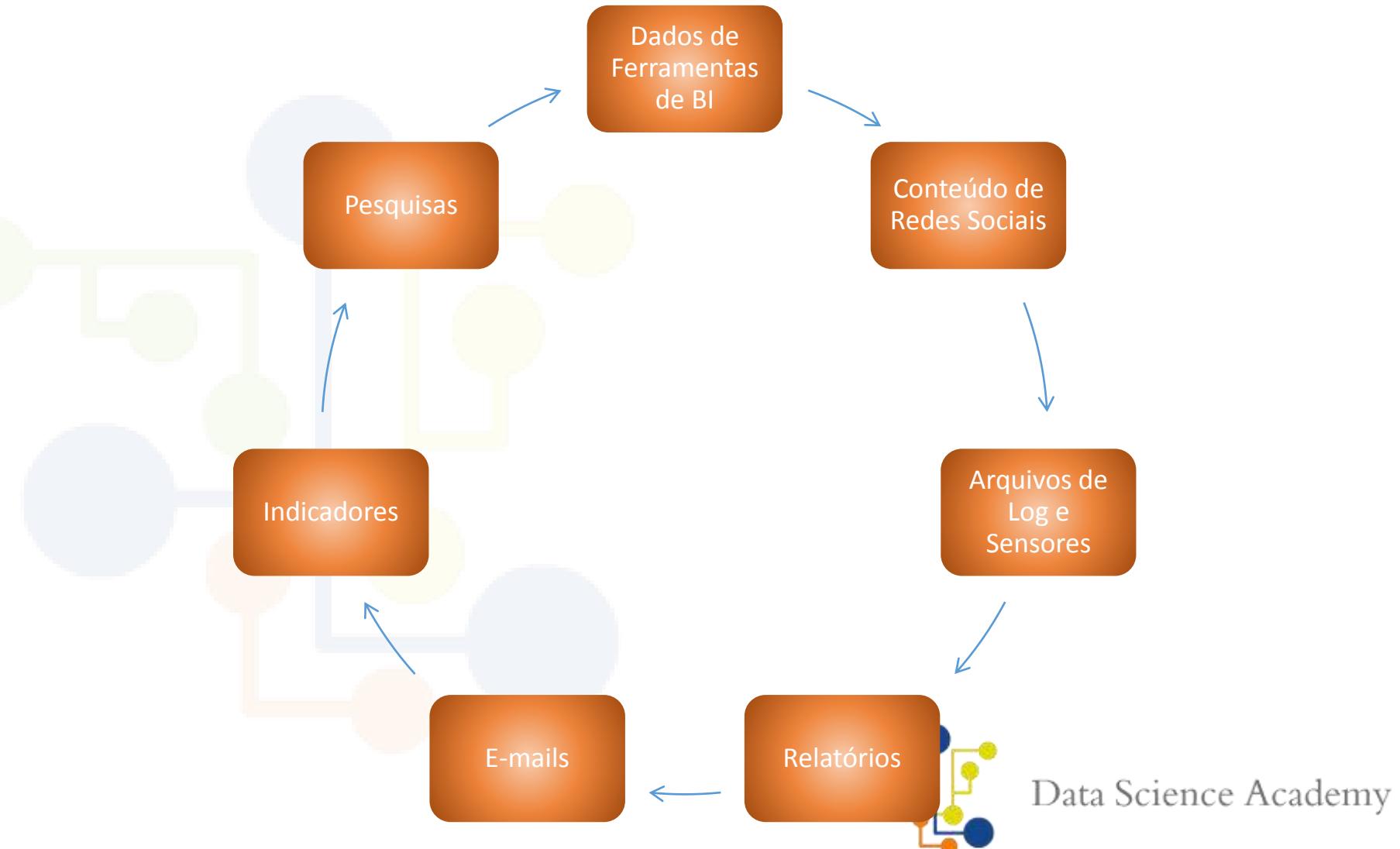
O que é Big Data Analytics?

Big Data Analytics é o trabalho analítico e inteligente em grandes volumes de dados, estruturados ou não-estruturados, que são coletados, armazenados e interpretados por softwares de altíssimo desempenho



Data Science Academy

Matéria-prima

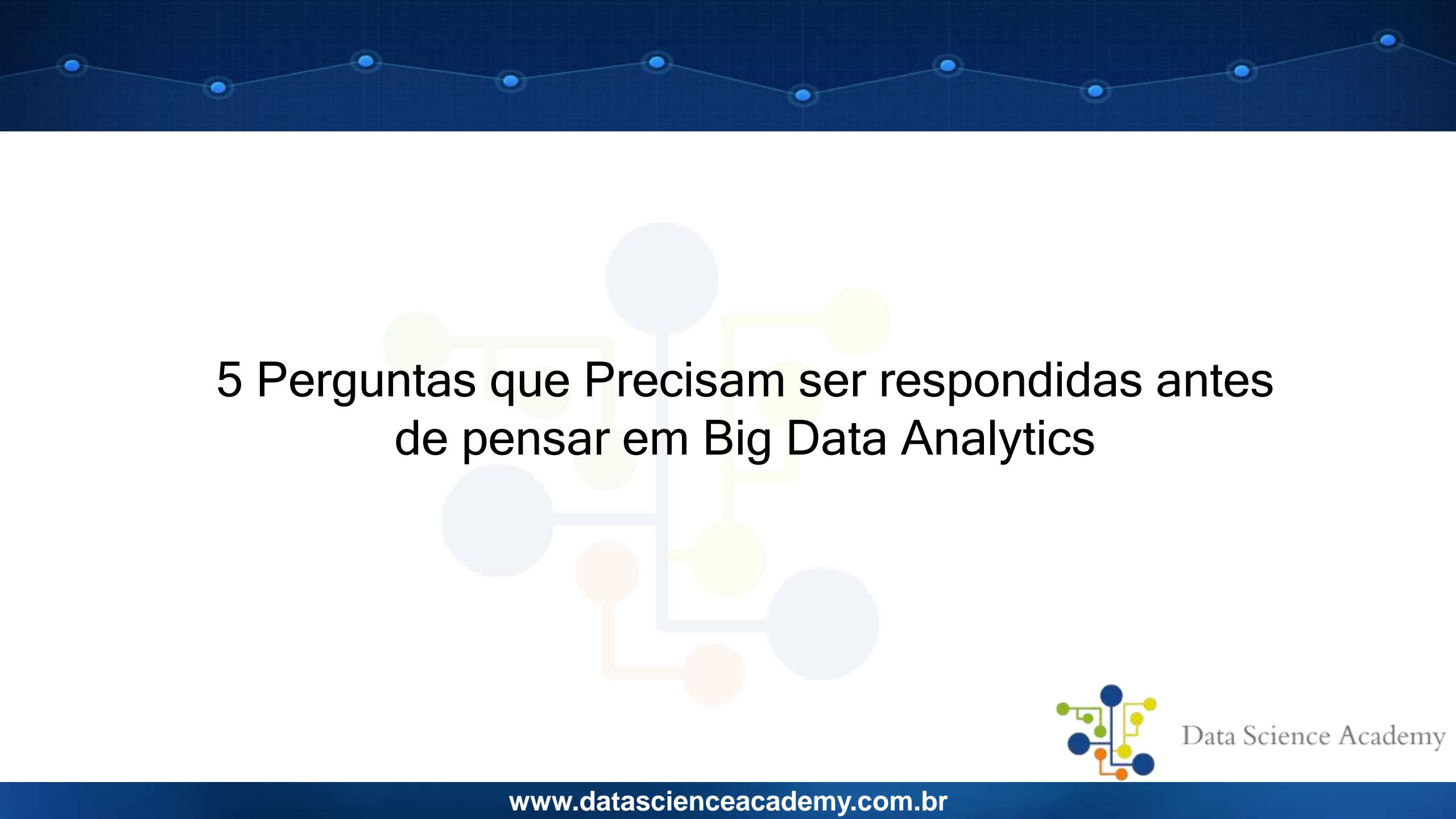


Vantagens e Benefícios

- **Direcionamento das Vendas**
- **Aperfeiçoamento do Processo de Logística**
- Atendimento mais eficiente
- Melhoria na Gestão de Recursos Humanos
- Identificação de Padrões
- Análise da Concorrência



Data Science Academy



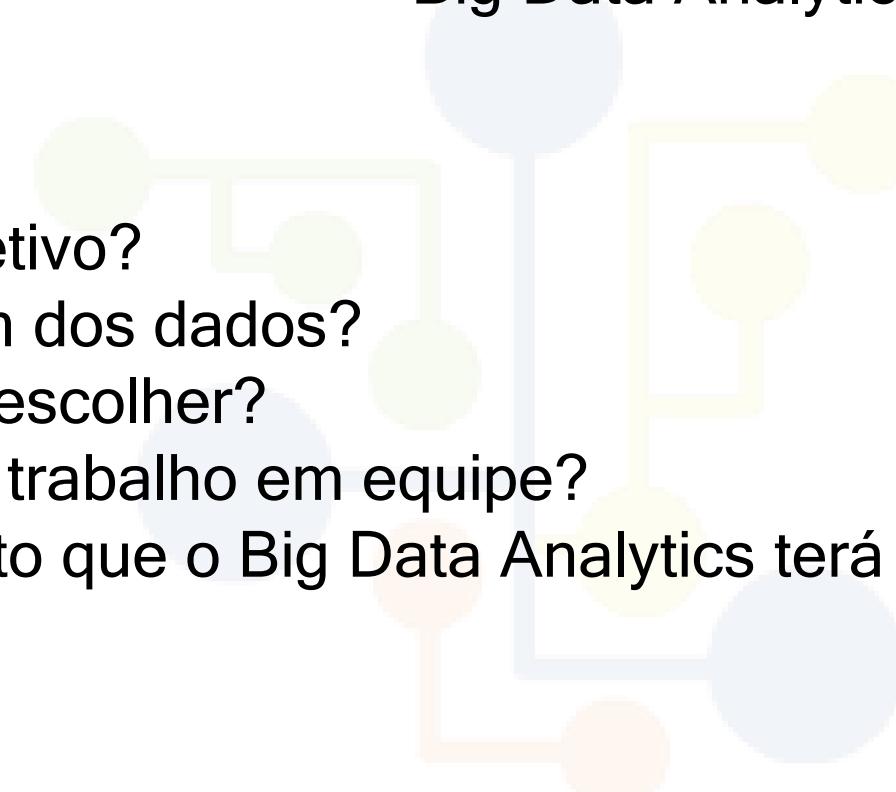
5 Perguntas que Precisam ser respondidas antes de pensar em Big Data Analytics



Data Science Academy



5 Perguntas que Precisam ser respondidas antes de pensar em Big Data Analytics

- 
- 1- Qual seu objetivo?
 - 2- Qual a origem dos dados?
 - 3- Que solução escolher?
 - 4- Este será um trabalho em equipe?
 - 5- Qual o impacto que o Big Data Analytics terá no negócio?



Data Science Academy



Big Data Analytics em Áreas de Negócio



Data Science Academy

Saúde

- Combate a doenças e epidemias
- Dispositivos móveis de monitoramento
- Atendimento médico individualizado
- Wearables e apps



Data Science Academy

A Accenture Healthcare IT Vision 2015, uma pesquisa anual sobre tendências de tecnologia, traz dados interessantes sobre o uso de Big Data na saúde.

The slide features the Accenture logo at the top left and 'Strategy | Consulting | Digital | Technology | Operations' at the top right. The title '2015 Healthcare IT Vision' is in orange. The main heading 'Top 5 eHealth Trends' is prominently displayed in large orange and white text. To the right, a large orange arrow points diagonally upwards with the text 'High performance. Delivered.' Below the main heading is a box containing the text: 'Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.' The background is dark blue with light blue circuit board patterns and glowing blue dots.

Strategy | Consulting | Digital | Technology | Operations

2015 Healthcare IT Vision

Top 5 eHealth Trends

High performance. Delivered.

Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.



Data Science Academy

45% dos executivos da área consultados afirmaram que fazem análise preditiva

The slide features the Accenture logo at the top left and 'Strategy | Consulting | Digital | Technology | Operations' at the top right. A large orange arrow points diagonally upwards from the bottom left towards the center. The title '2015 Healthcare IT Vision' is in orange, followed by 'Top 5 eHealth Trends' in large white text. A callout box contains the text: 'Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.' To the right of the slide is a decorative graphic of interconnected nodes in blue, green, and yellow.

Strategy | Consulting | Digital | Technology | Operations

2015 Healthcare IT Vision

Top 5 eHealth Trends

Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.

High performance. Delivered.



Data Science Academy

59% disseram que já usam algoritmos com o objetivo de fazer softwares operarem com mais inteligência

The slide features the Accenture logo at the top left and a horizontal menu bar with the words "Strategy | Consulting | Digital | Technology | Operations". The main title "2015 Healthcare IT Vision" is in orange. Below it, the section "Top 5 eHealth Trends" is prominently displayed in large orange and white text. To the right, a large orange chevron shape contains the text "High performance. Delivered.". A callout box at the bottom left contains the text: "Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow." The background is dark blue with a faint grid pattern.

Strategy | Consulting | Digital | Technology | Operations

2015 Healthcare IT Vision

Top 5 eHealth Trends

High performance. Delivered.

Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.



Data Science Academy

73% de todos os executivos de saúde relataram ter obtido um ROI positivo depois de investir em tecnologias como dispositivos móveis do tipo wearables

The slide features the Accenture logo at the top left and a horizontal menu bar with the words "Strategy | Consulting | Digital | Technology | Operations". The main title "2015 Healthcare IT Vision" is displayed in orange. Below it, the section "Top 5 eHealth Trends" is prominently shown in large, bold, orange and white text. To the right, a large orange chevron shape contains the slogan "High performance. Delivered." In the bottom left corner of the slide, there is a callout box with a yellow border containing the text: "Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow." The background of the slide has a dark blue theme with abstract circuit board patterns.

Strategy | Consulting | Digital | Technology | Operations

2015 Healthcare IT Vision

Top 5 eHealth Trends

High performance. Delivered.

Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.



Data Science Academy



Construção



Data Science Academy



Logística



Data Science Academy



Logística

Estatística + Matemática + Mídias Sociais + Macroeconomia + Tecnologia + Dados

=

Big Data Analytics



Data Science Academy



CUSTOMER
ANALYSIS
ADVERTISING
MARKET
SALES
PROMOTION
PRICE
INTERNET
TARGET

Marketing

- Análise de mercado mais específica
- Interpretação de Dados mais eficaz
- Mais fontes de dados
- Automação do Marketing
- Experiências mais significativas no PV



Data Science Academy

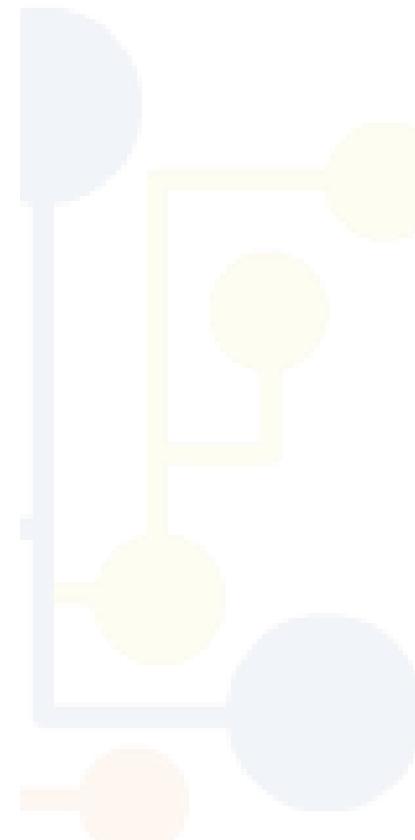


Indústria

- Eficiência no Planejamento (SCM)
- Gestão de Suprimentos (Estoque míni
- Redução de Custos
- Qualidade + dados = + precisão



Data Science Academy



E-commerce



Data Science Academy



Energia



Data Science Academy



A Empresa como um todo

- Mudança de Cultura
- Visão Empresarial *data driven*
- Esforço Top-Down



Data Science Academy

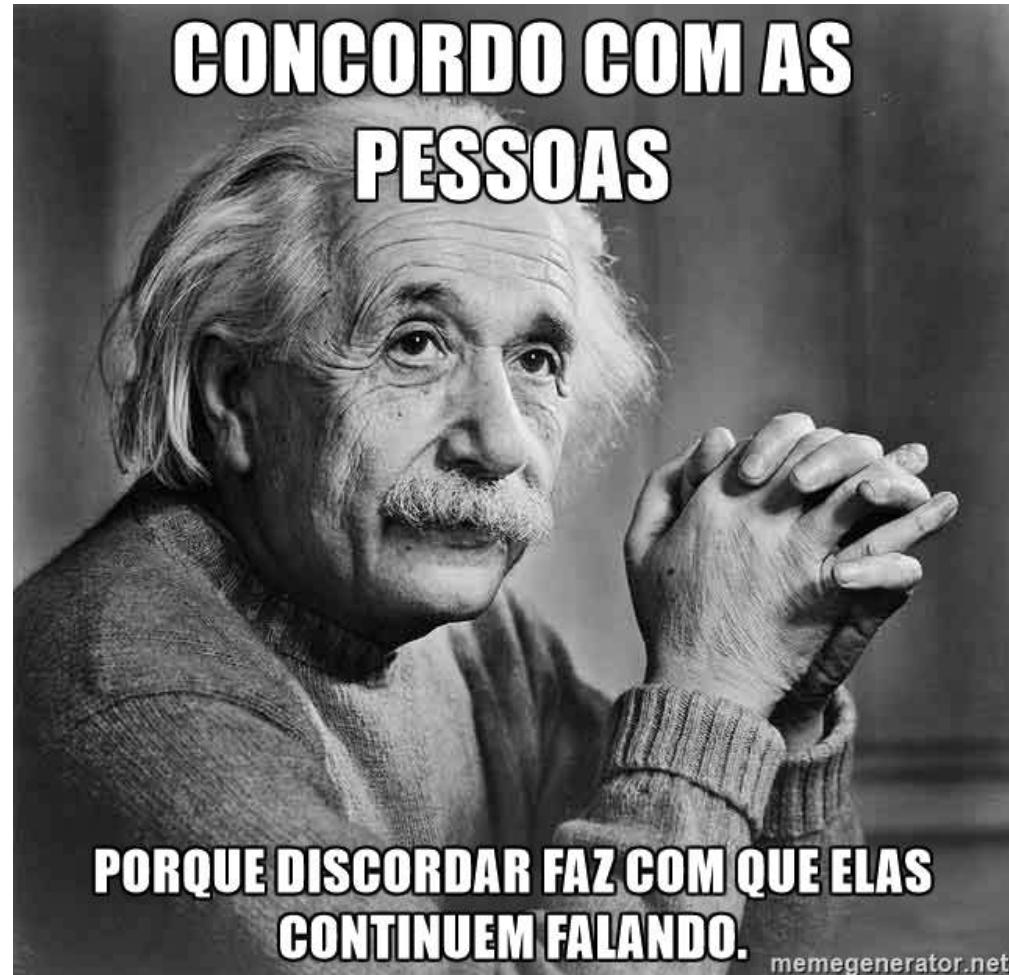


Importante!

- Big Data já não é mais um tema distante
- Big Data fará toda a diferença para seu negócio
- Qualquer empresa pode se beneficiar do Big Data
- Big Data é um trabalho em equipe



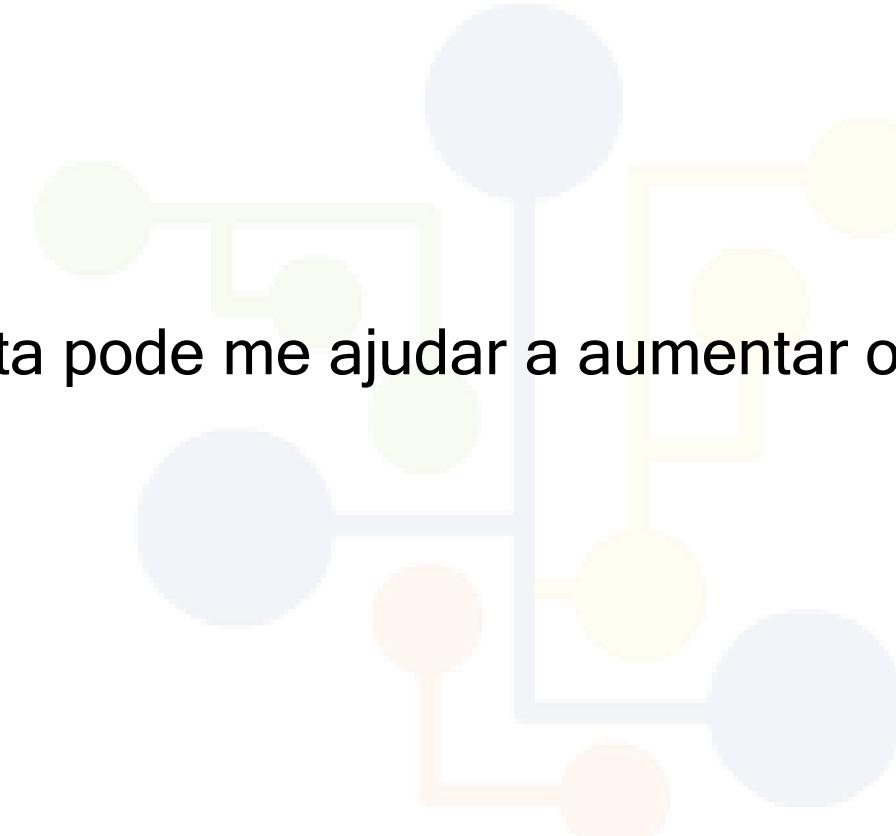
Data Science Academy



Mas por onde eu começo?



Data Science Academy



Como o Big Data pode me ajudar a aumentar o Market Share da empresa?



Data Science Academy

Coletar Dados

Faturamento
Marketing
Clientes
Custos

Efetividade das Campanhas de Marketing

Concorrentes
Redes Sociais



Data Science Academy

Trabalhar Sobre os Dados

Carga

Limpeza

Transformação

Compreensão

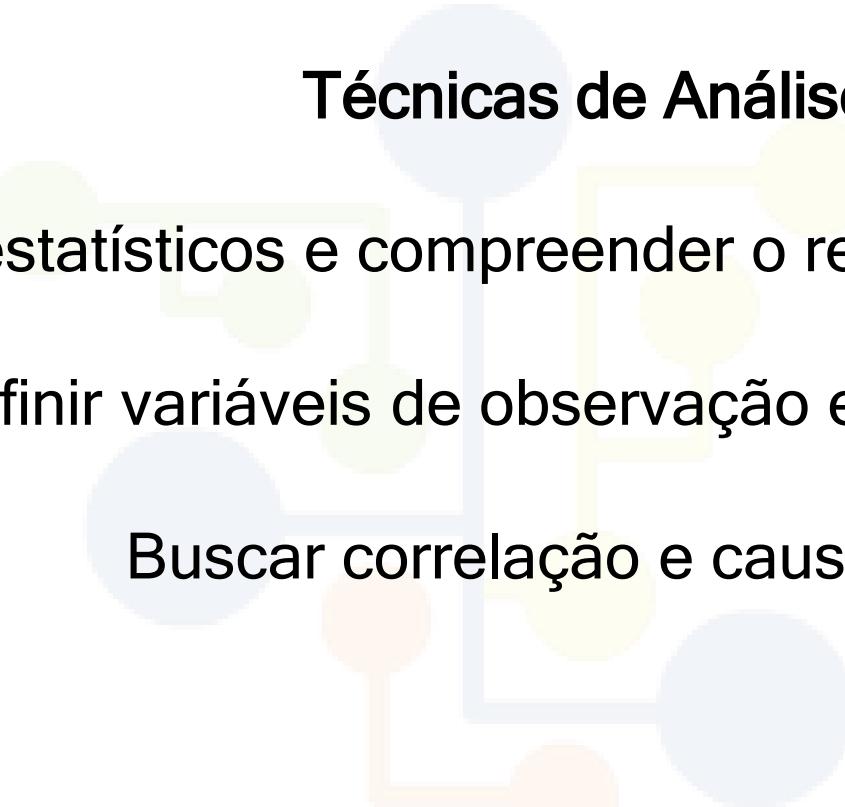


Data Science Academy



Técnicas de Análise

Aplicar modelos estatísticos e compreender o relacionamento entre os dados



Definir variáveis de observação e explanatórias

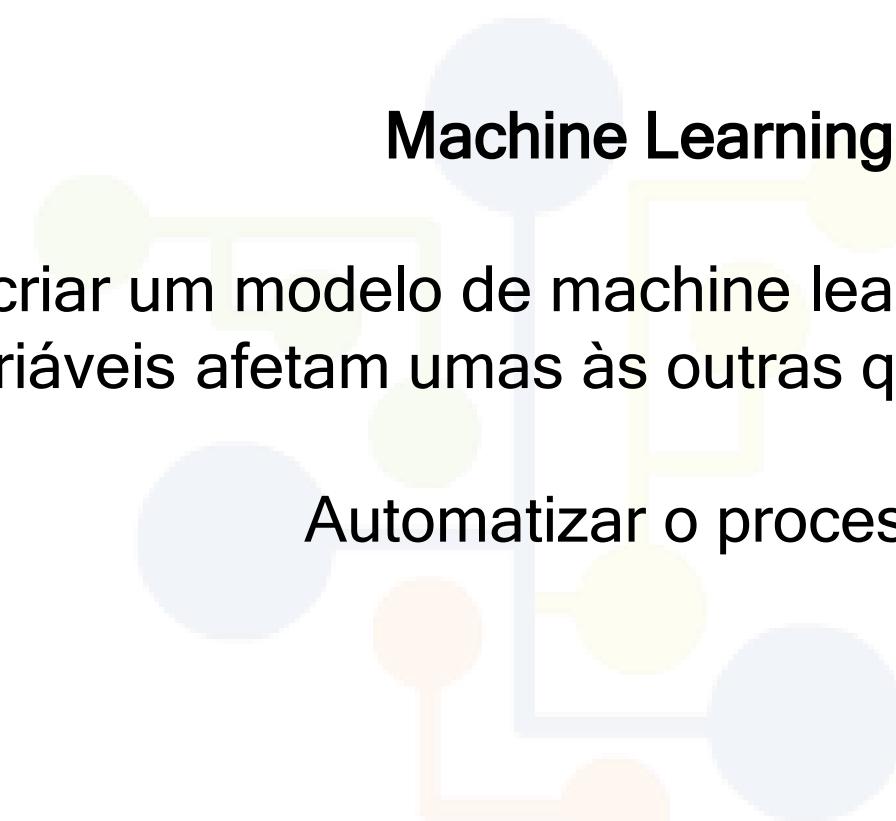
Buscar correlação e causalidade



Data Science Academy



Machine Learning



Juntar tudo e criar um modelo de machine learning, prevendo como estas variáveis afetam umas às outras quando alteradas

Automatizar o processo



Data Science Academy

Apresentar seus Resultados





obrigado!



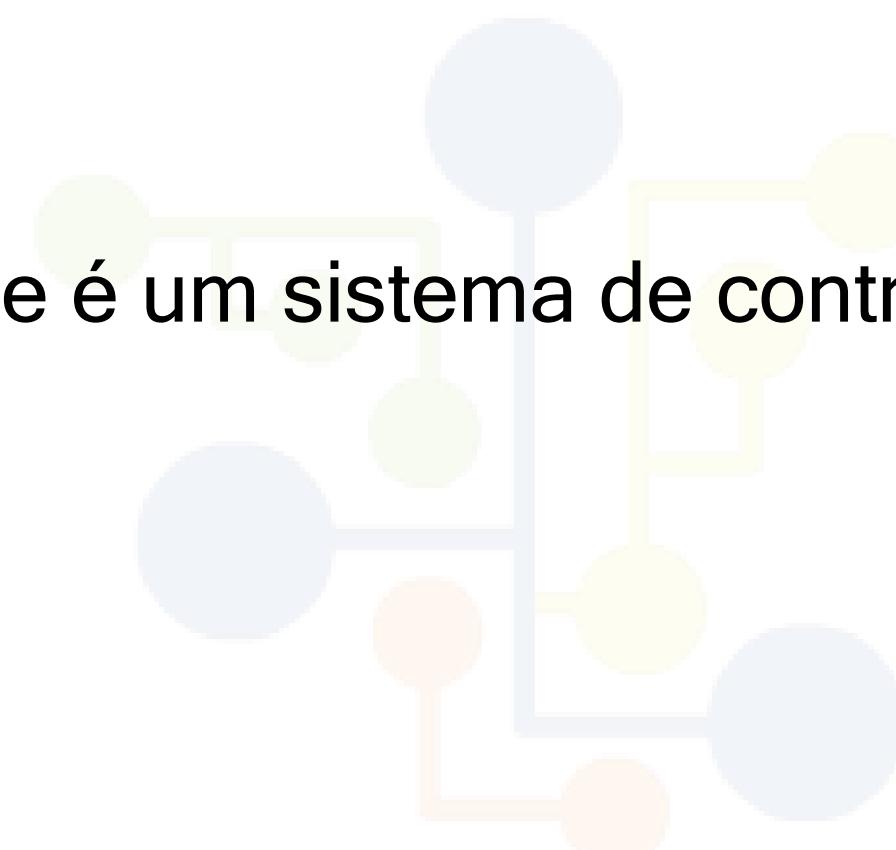
Data Science Academy



Usando o Github



Data Science Academy

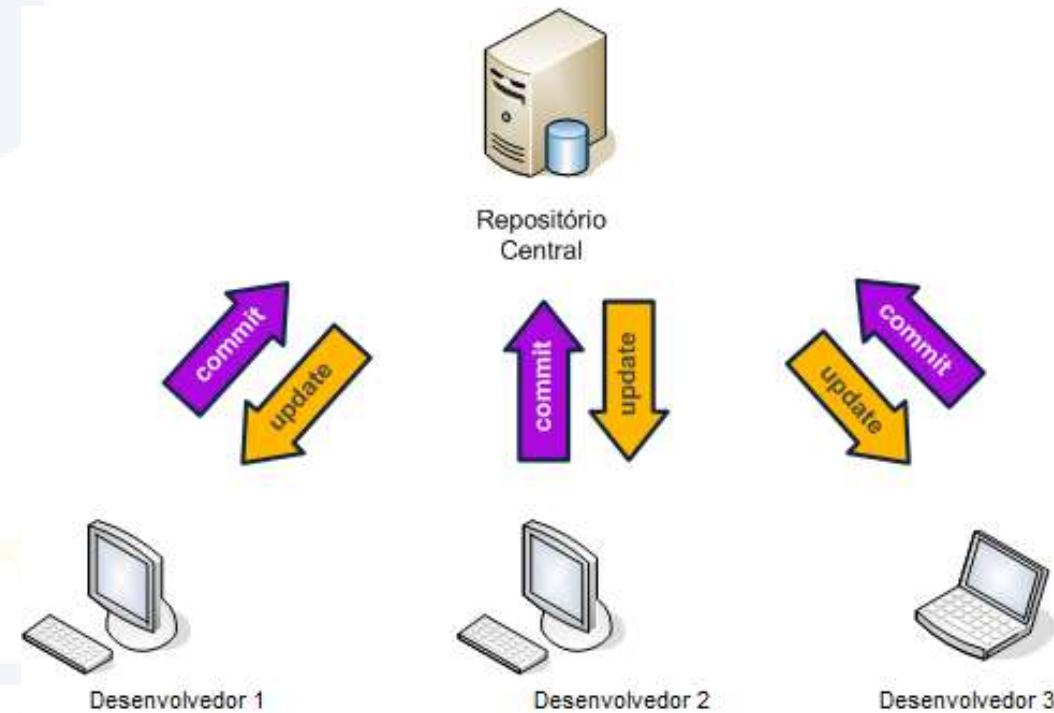


O que é um sistema de controle de versão?



Data Science Academy

Várias pessoas
trabalhando nos
mesmos arquivos



Data Science Academy



Histórico e Conceitos



Data Science Academy

Sistemas de Controle de Versão

- *Concurrent Versions System (CVS)*
- *Subversion (SVN)*
- *Visual SourceSafe (VSS)*
- *Rational ClearCase*
- *Git*



Data Science Academy



Data Science Academy



- Não depender de um servidor central
- Dar ênfase à velocidade
- Integridade dos dados
- Potencializar o trabalho paralelo



Data Science Academy



GitHub



<https://github.com>



Data Science Academy



Search GitHub

Pull requests Issues Gist

Overview Repositories Public activity Edit profile

Popular repositories

 PythonFundamentos Python Fundamentos para Análise de Dados	4 ★
 BigDataAnalyticsHadoop	0 ★
 BigDataAnalyticsSpark	0 ★
 RFundamentos R Fundamentos	0 ★

Customize your pinned repositories

6 contributions in the last year

Contribution settings

Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Less	More										

Summary of pull requests, issues opened, and commits. Learn how we count contributions.

Period: 1 week

Contribution activity

1 commit

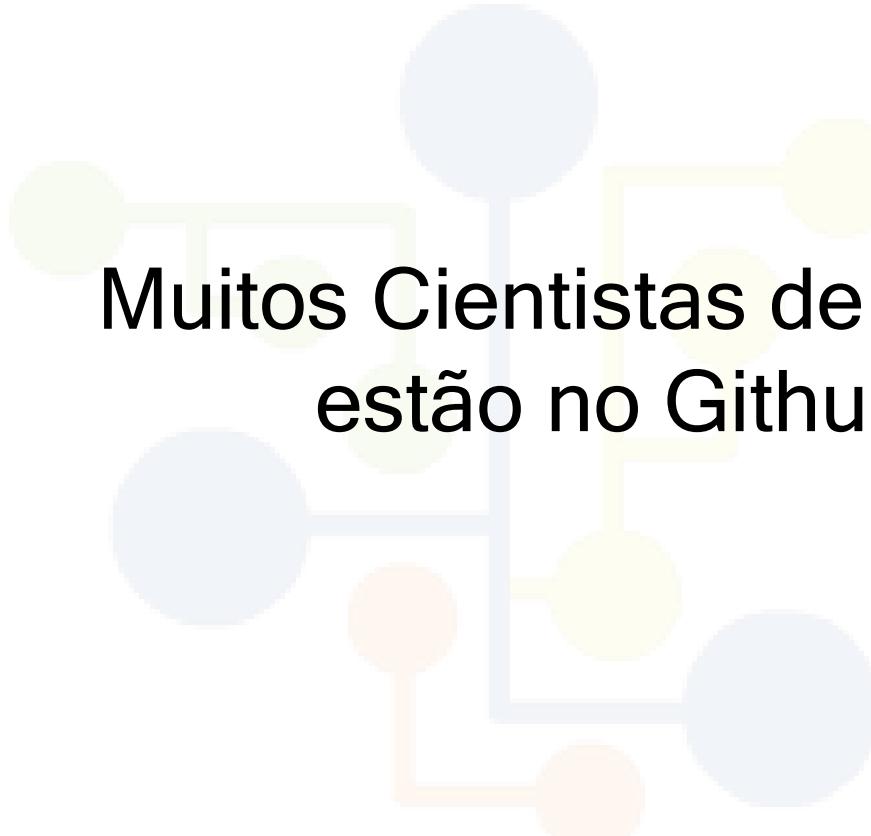
Pushed 1 commit to dsacademybr/PythonFundamentos Jun 26

<https://github.com/dsacademybr>

www.datascienceacademy.com.br



Data Science Academy



Muitos Cientistas de Dados
estão no Github



Data Science Academy



<https://github.com/rasbt>

Personal Open source Business Explore Pricing Blog Support Search GitHub Sign in Sign up

**Sebastian
Raschka**
rasbt

'Data Scientist' and Machine Learning enthusiast with a big passion for Python & open source. Author of 'Python Machine Learning'. Michigan State University.

Michigan State University
 East Lansing, Michigan
 mail@sebastianraschka.com
 http://sebastianraschka.com

Joined on Oct 5, 2013

[Overview](#)

[Repositories](#)

[Public activity](#)

[Follow](#)

Pinned repositories

python-machine-learning-book	3,520 ★
The "Python Machine Learning" book code repository and info resource	
pattern_classification	1,756 ★
A collection of tutorials and examples for solving and understanding machine learning and pattern classification...	
mixtend	484 ★
A library of extension and helper modules for Python's data analysis and machine learning libraries.	
python_reference	1,606 ★
Useful functions, tutorials, and other Python-related things	
biopandas	52 ★
Working with molecular structures in pandas DataFrames	

1,510 contributions in the last year

Summary of pull requests, issues opened, and commits. Learn how we count contributions.

Less More

Data Science Academy

www.datasciencacademy.com.br



<https://github.com/rhiever>

 Personal Open source Business Explore Pricing Blog Support Search GitHub Sign In Sign up



Randy Olson
rhiever

Artificial Intelligence & Machine Learning researcher. Data tinkerer. Community leader for /r/DataIsBeautiful.

 University of Pennsylvania
 Philadelphia, PA
 rso@randalolson.com
 <http://www.randalolson.com>
 Joined on May 8, 2012

1.5k 69 14

Overview Repositories Public activity Follow

Popular repositories

-  [Data-Analysis-and-Machine-Le...](#) 2,457 ★
Repository of teaching materials, code, and data for my data analysis and machine learning projects.
-  [tpot](#) 1,150 ★
A Python tool that automatically creates and optimizes machine learning pipelines using genetic programming.
-  [TwitterFollowBot](#) 562 ★
A Python bot that automates several actions on Twitter, such as following users and favoriting tweets.
-  [datacleaner](#) 301 ★
A Python tool that automatically cleans data sets and readies them for analysis.
-  [reddit-analysis](#) 182 ★
A Python script that parses post titles, self-texts, and comments on reddit and makes word clouds out of the word freq...

812 contributions in the last year



Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

Less More



Data Science Academy



<https://github.com/hmason>

Personal Open source Business Explore Pricing Blog Support Search GitHub Sign in Sign up



Hilary Mason
hmason

Fast Forward Labs
New York, NY
<http://www.hilarymason.com>
Joined on Jul 22, 2008

2k Followers 134 Starred 50 Following

Organizations 

Overview Repositories Public activity Follow

Popular repositories

Repository	Stars
gitmarks	321
ml_class	254
botomatic	233
tc	167
introbot	70

3 contributions in the last year



Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

Less More



Data Science Academy



<https://github.com/mbostock>

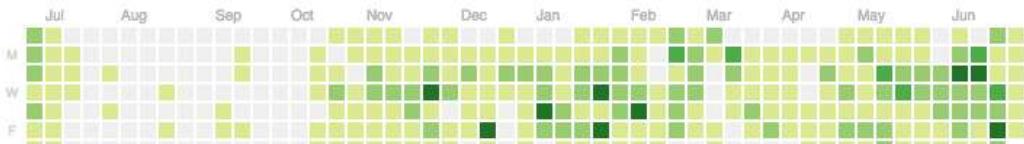
Personal Open source Business Explore Pricing Blog Support Search GitHub Sign in Sign up

 Overview Repositories Public activity Follow

Pinned repositories

-  **d3/d3**
Bring data to life with SVG, Canvas and HTML.  52,082 ★
-  **topojson**
An extension to GeoJSON that encodes topology. 2,366 ★
-  **d3/d3-shape**
Graphical primitives for visualization, such as lines and areas. 1,239 ★
-  **d3/d3-scale**
Encodings that map abstract data to visual representation. 677 ★
-  **d3/d3-geo-projection**
Extended geographic projections for D3. 325 ★

5,546 contributions in the last year



Summary of pull requests, issues opened, and commits. Learn how we count contributions.

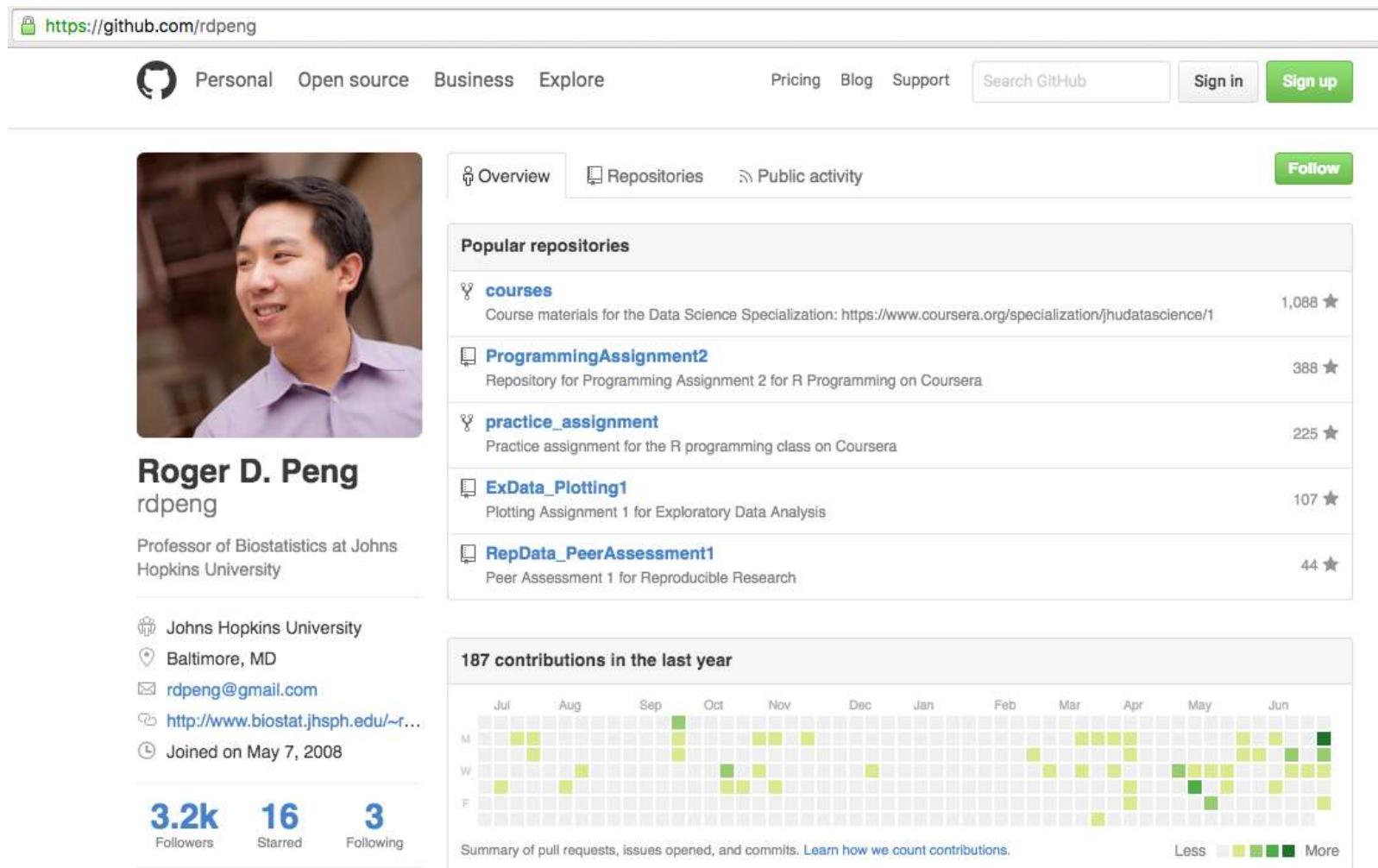
Less More

12.7k Followers **37** Starred **13** Following

Organizations 



Data Science Academy



The image shows a screenshot of Roger D. Peng's GitHub profile page. At the top, there is a navigation bar with links for Personal, Open source, Business, Explore, Pricing, Blog, Support, a search bar labeled "Search GitHub", and buttons for "Sign in" and "Sign up". Below the navigation bar is a large portrait photo of Roger D. Peng, a man with dark hair and a light beard, wearing a purple shirt. To the right of the photo, his name "Roger D. Peng" and GitHub handle "rdpeng" are displayed. Below his name, it says "Professor of Biostatistics at Johns Hopkins University". On the left side of the profile, there is a sidebar with icons for Johns Hopkins University, location in Baltimore, MD, email (rdpeng@gmail.com), and a link to his website (<http://www.biostat.jhsph.edu/~rpeng/>). It also shows the date he joined GitHub, May 7, 2008. At the bottom of the sidebar, there are three summary numbers: 3.2k Followers, 16 Starred, and 3 Following. The main content area of the profile page is titled "Popular repositories" and lists five repositories: "courses" (1,088 stars), "ProgrammingAssignment2" (388 stars), "practice_assignment" (225 stars), "ExData_Plottin1" (107 stars), and "RepData_PeerAssessment1" (44 stars). Below this section is a chart titled "187 contributions in the last year" showing a grid of colored squares representing contributions by month and day. A legend indicates that lighter colors represent "Less" contributions and darker colors represent "More".



Data Science Academy

<https://github.com/amueller>

Personal Open source Business Explore Pricing Blog Support Search GitHub Sign in Sign up

Overview Repositories Public activity Follow

Popular repositories

Repository	Star Count
word_cloud	1,298
introduction_to_ml_with_python	392
scipy_2015_sklearn_tutorial	335
kaggle_insults	102
odscon-sf-2015	76

1,271 contributions in the last year

Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

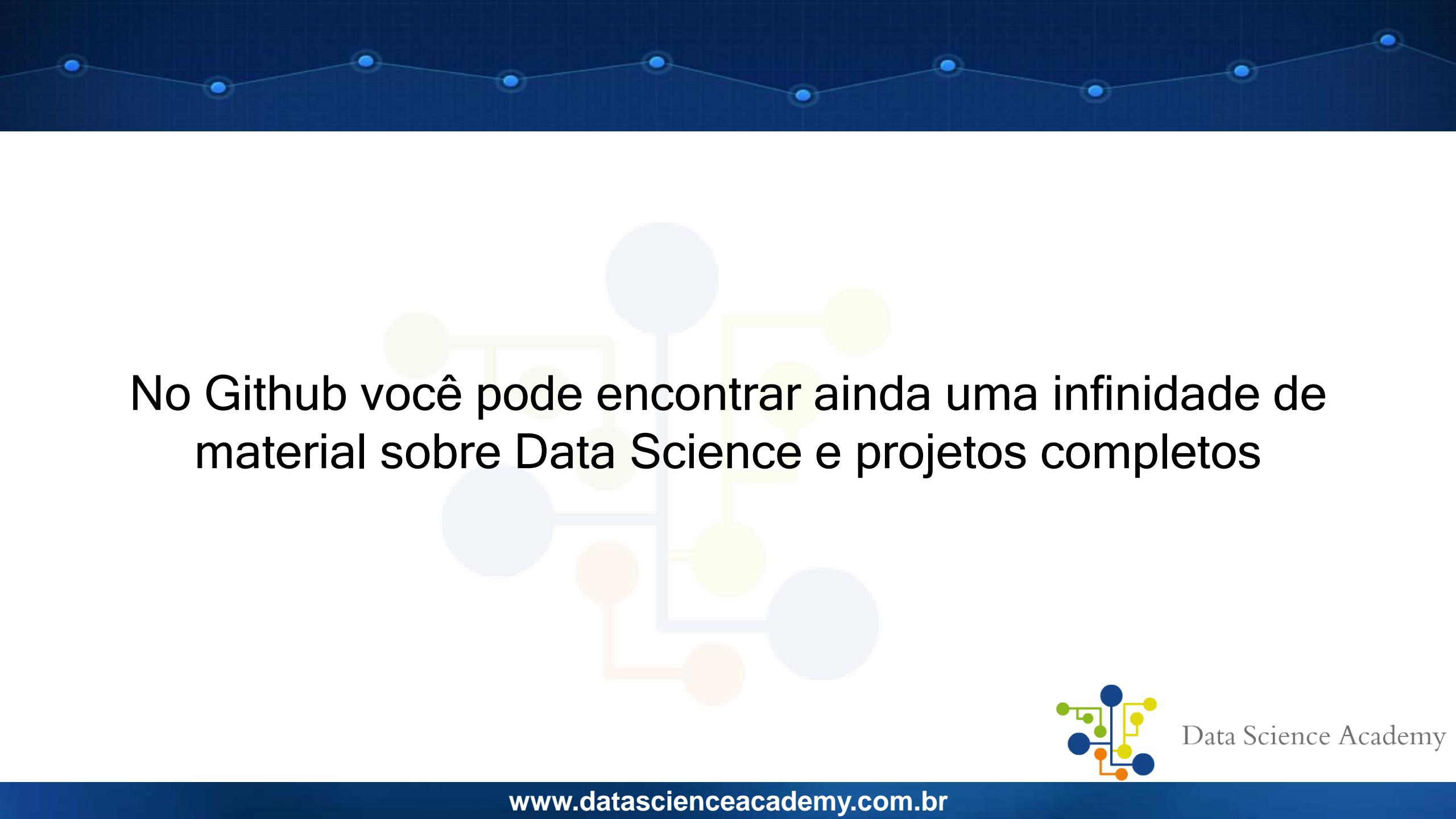
Less More

Organizations

amueller



Data Science Academy



No Github você pode encontrar ainda uma infinidade de material sobre Data Science e projetos completos



Data Science Academy



Principais Conceitos do Git



Data Science Academy

Principais conceitos do Git

Branch

Ramificação do projeto, cada *branch* representa uma versão do seu projeto e podemos seguir uma linha de desenvolvimento a partir de cada *branch*



Data Science Academy

Principais conceitos do Git



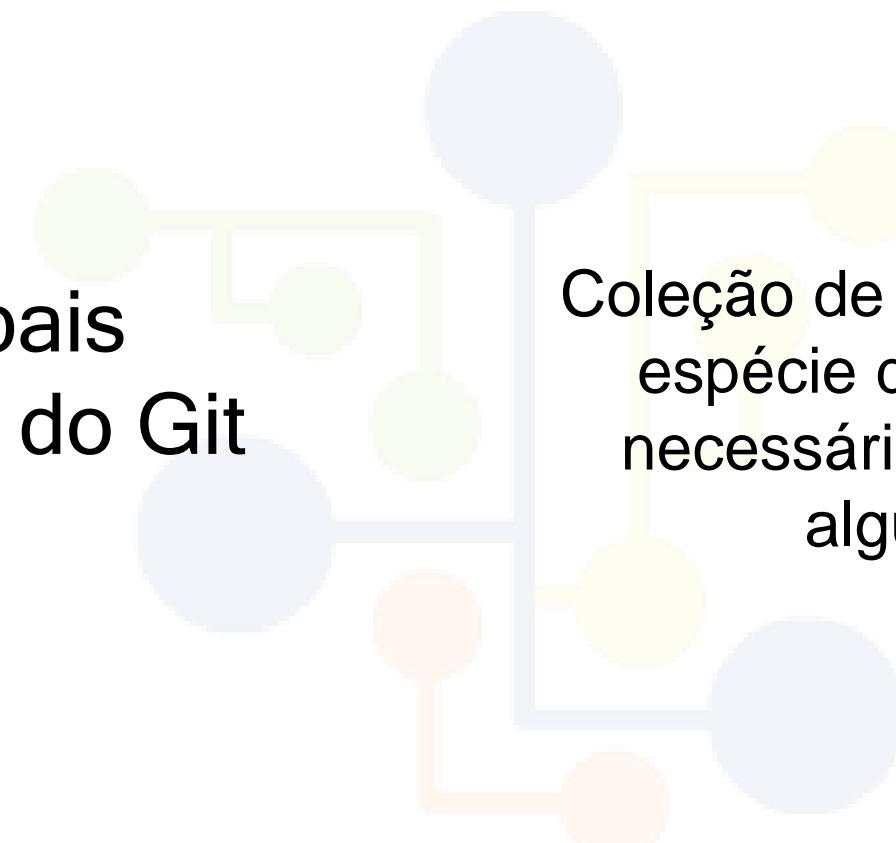
Clone

Cópia local de todos os arquivos de um repositório git



Data Science Academy

Principais conceitos do Git



Commit

Coleção de alterações realizadas, é uma espécie de *checkpoint*, sempre que necessário você pode retroceder até algum *commit* existente



Data Science Academy

Principais conceitos do Git



Fork

Uma bifurcação do projeto, uma cópia do projeto existente para seguir uma nova direção



Data Science Academy

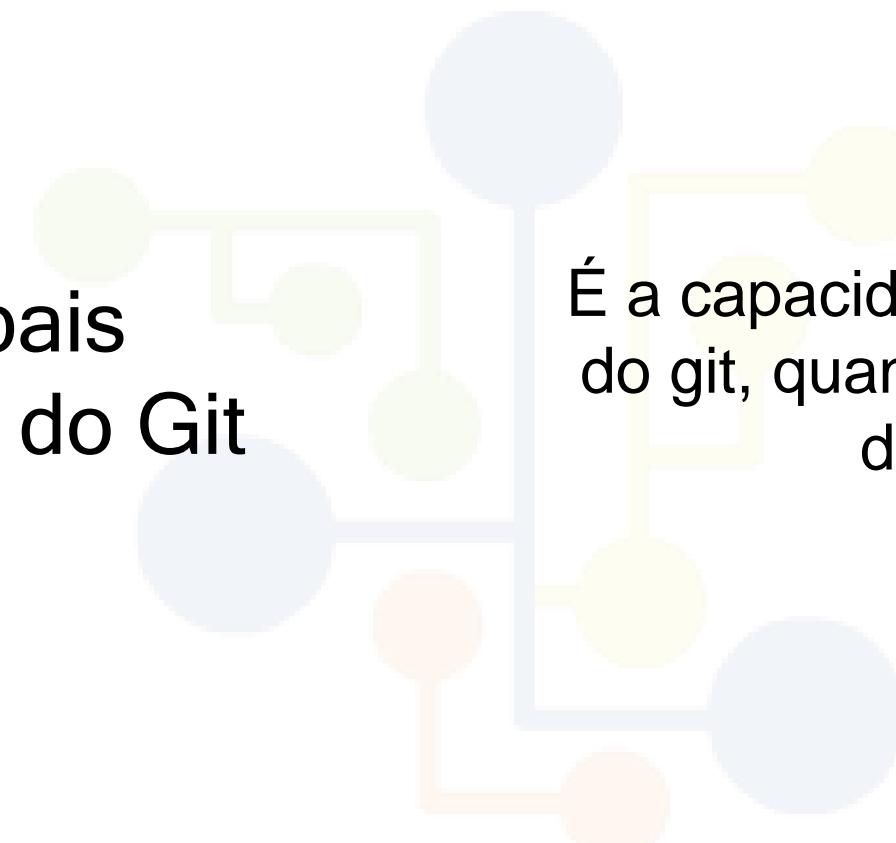
Principais conceitos do Git

Master
Branch padrão de um repositório Git



Data Science Academy

Principais conceitos do Git



Merge

É a capacidade de incorporar alterações do git, quando acontece uma junção de diferentes *branches*



Data Science Academy

Principais conceitos do Git



Pull

Puxa as alterações do repositório remoto



Data Science Academy

Principais conceitos do Git



Push

Empurra as suas alterações para o repositório remoto



Data Science Academy

Principais conceitos do Git

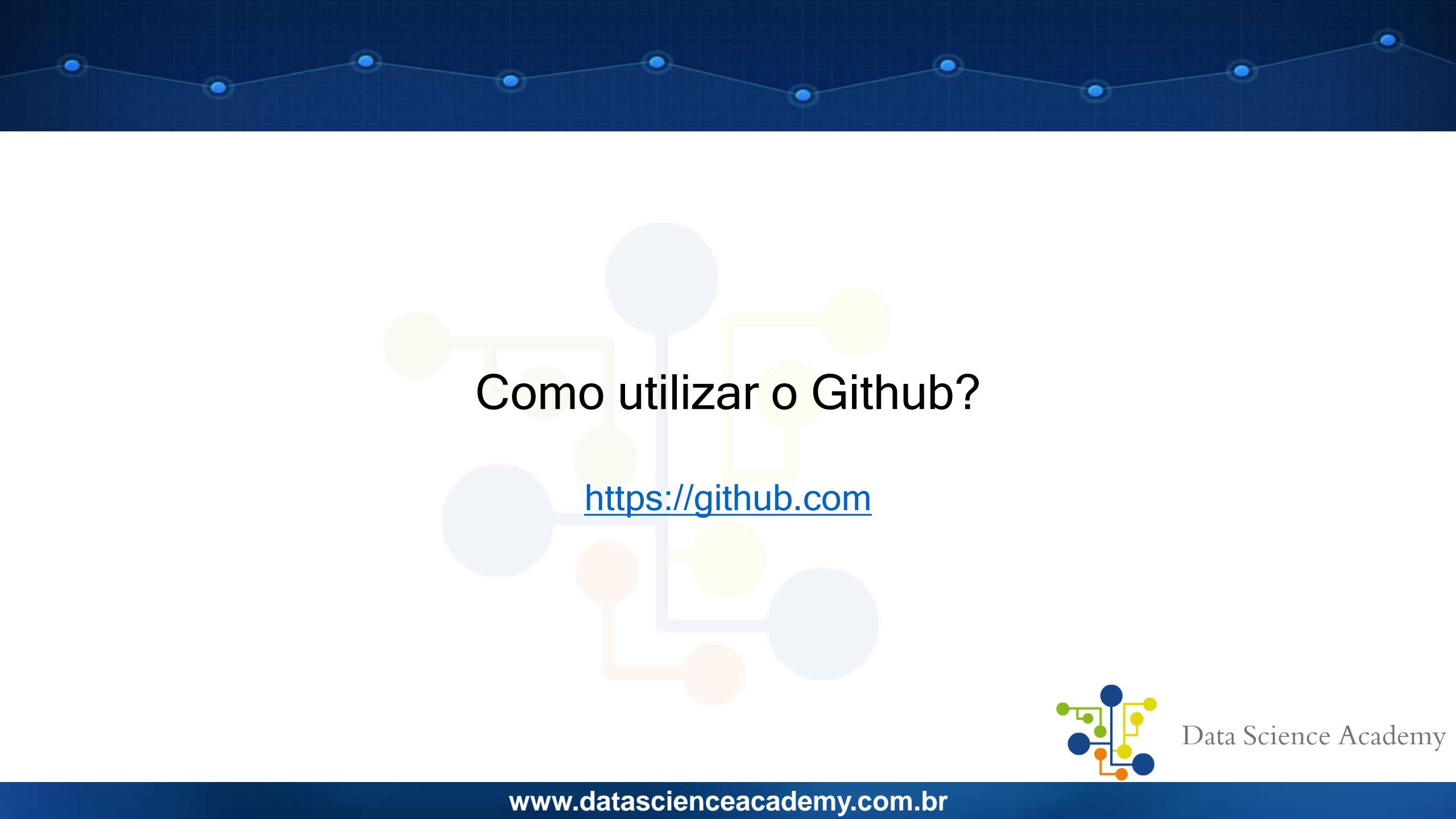


Repositório

Local onde ficam todos os arquivos do projeto, inclusive o histórico e versões



Data Science Academy

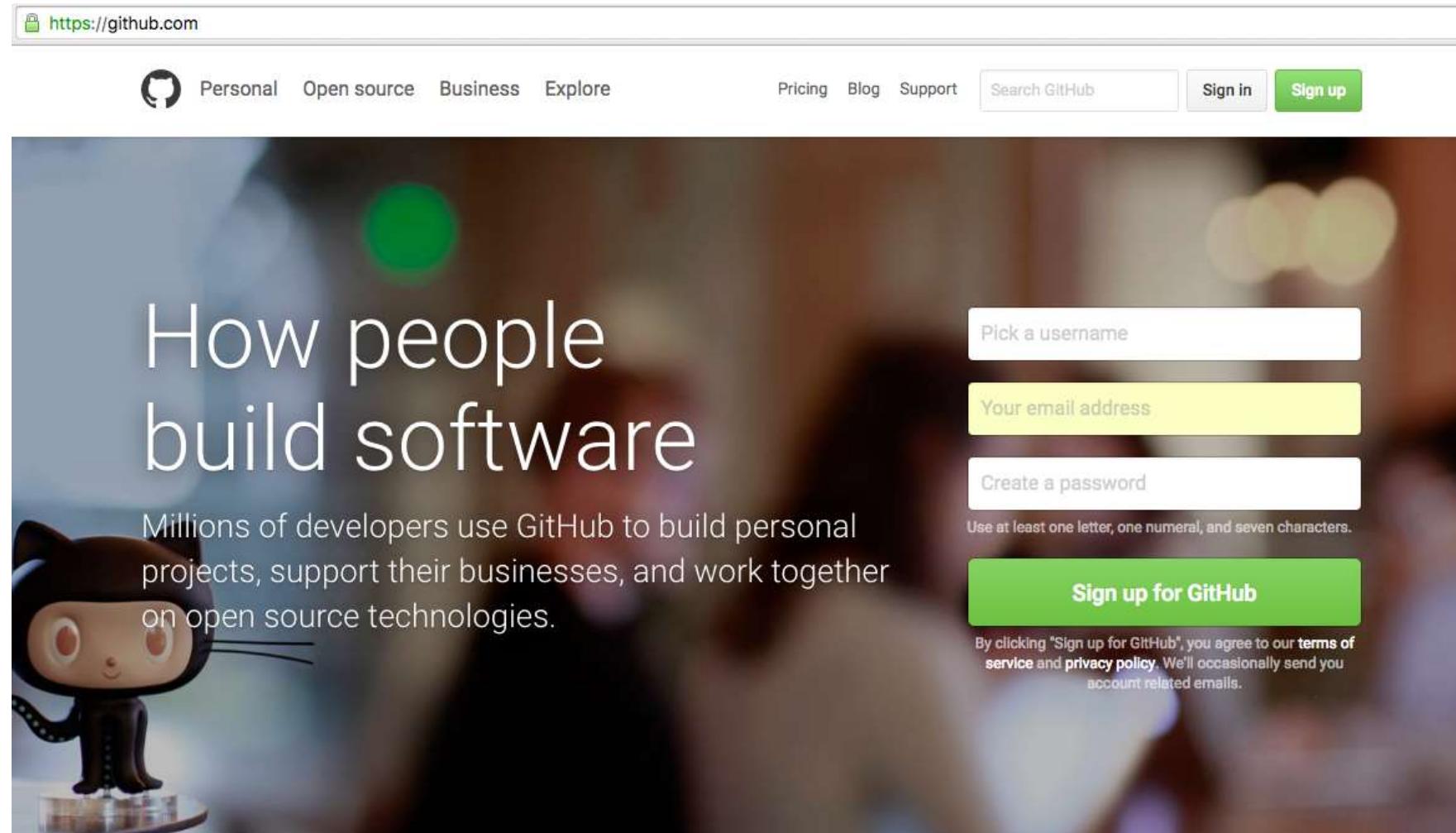


Como utilizar o Github?

<https://github.com>



Data Science Academy



<https://github.com>

www.datascienceacademy.com.br



Data Science Academy

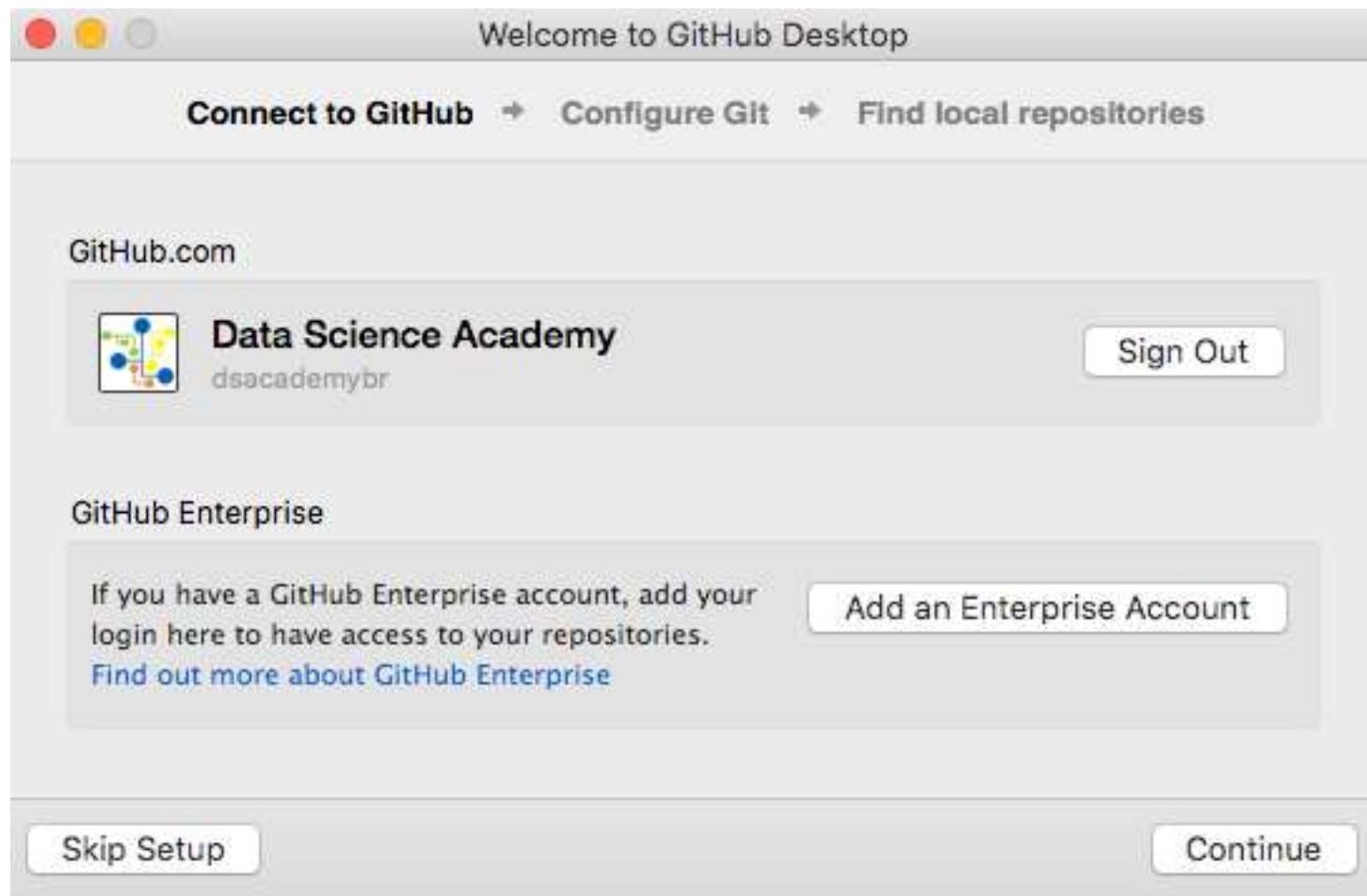


Github Desktop

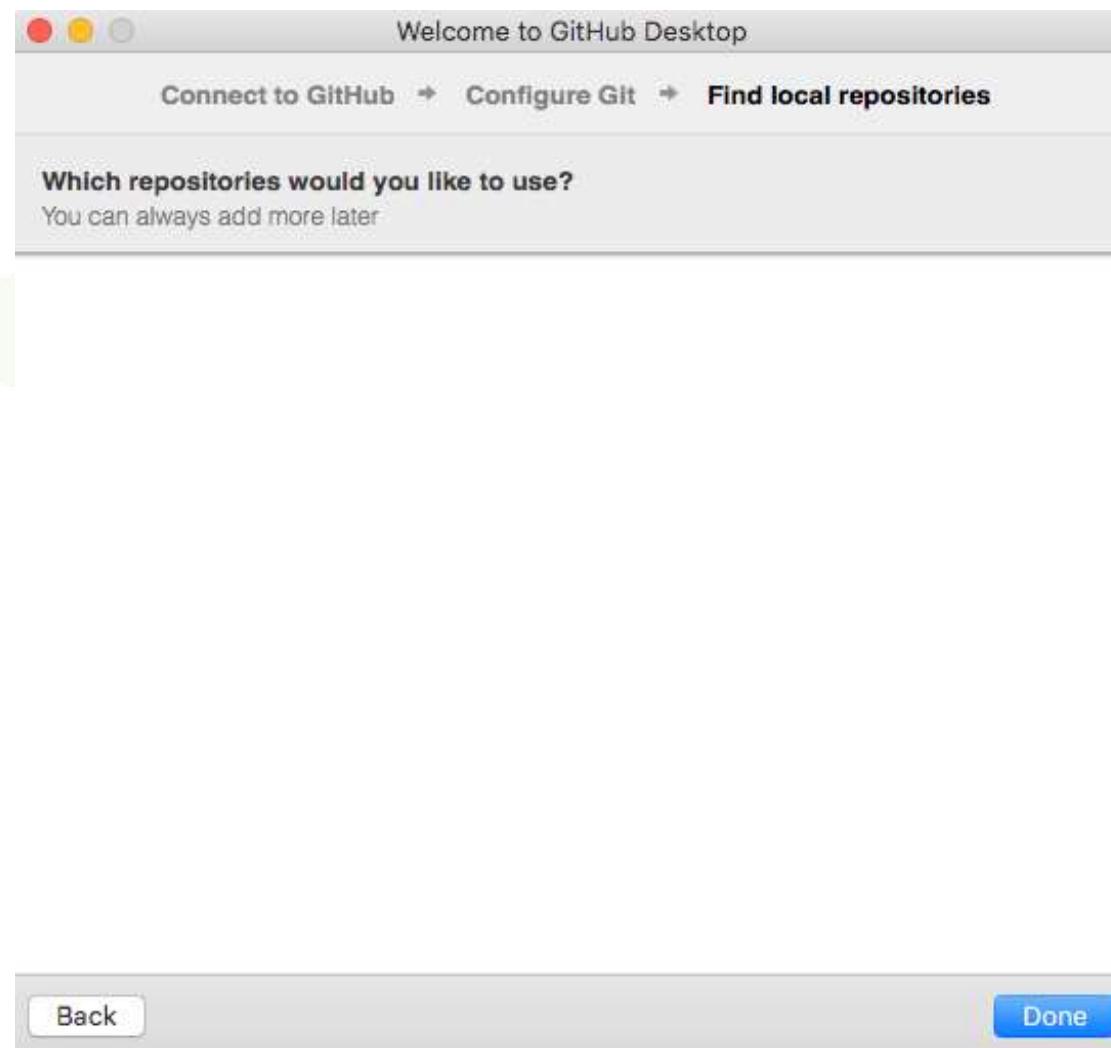
<https://desktop.github.com>



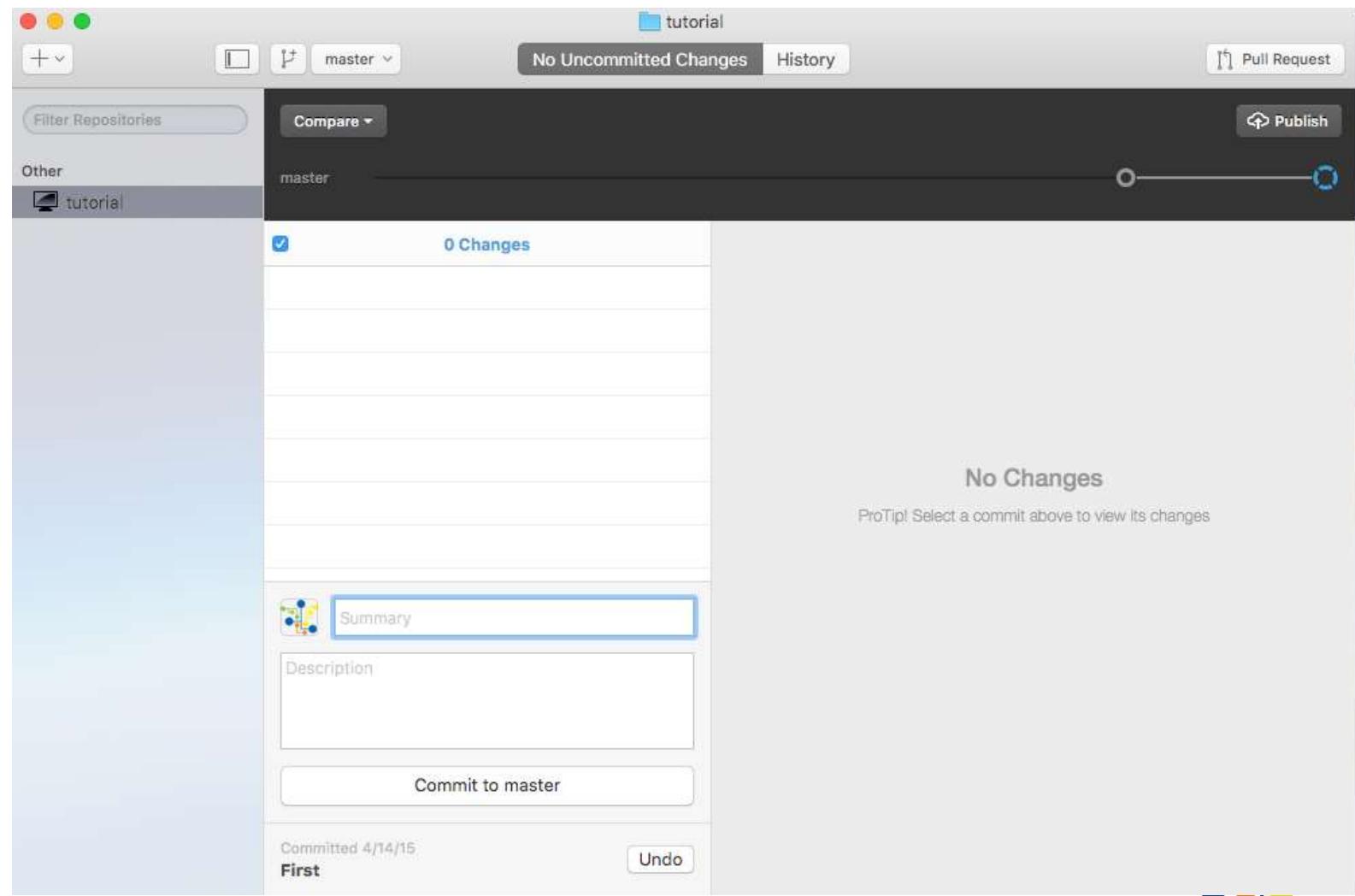
Data Science Academy



Data Science Academy

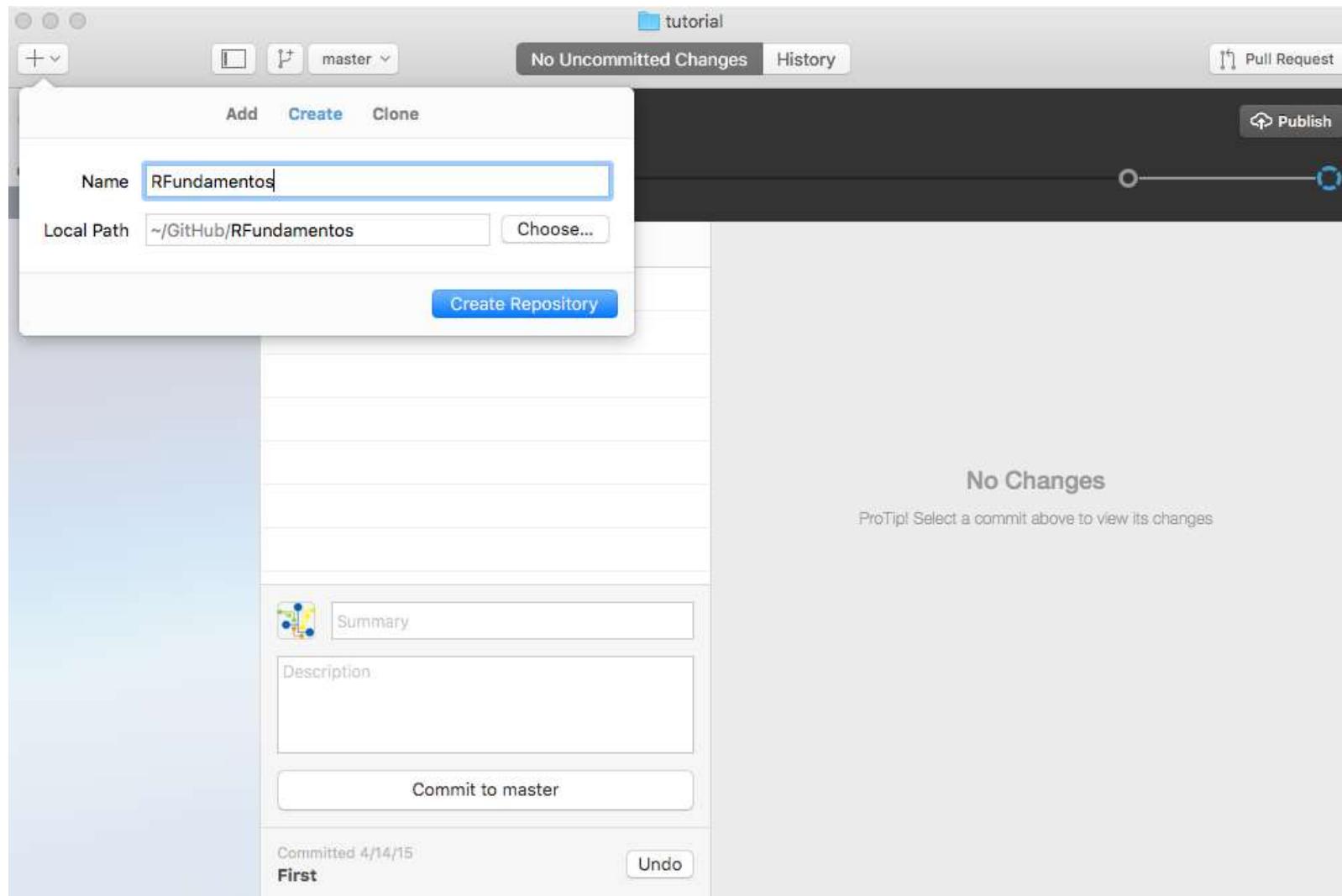


Data Science Academy



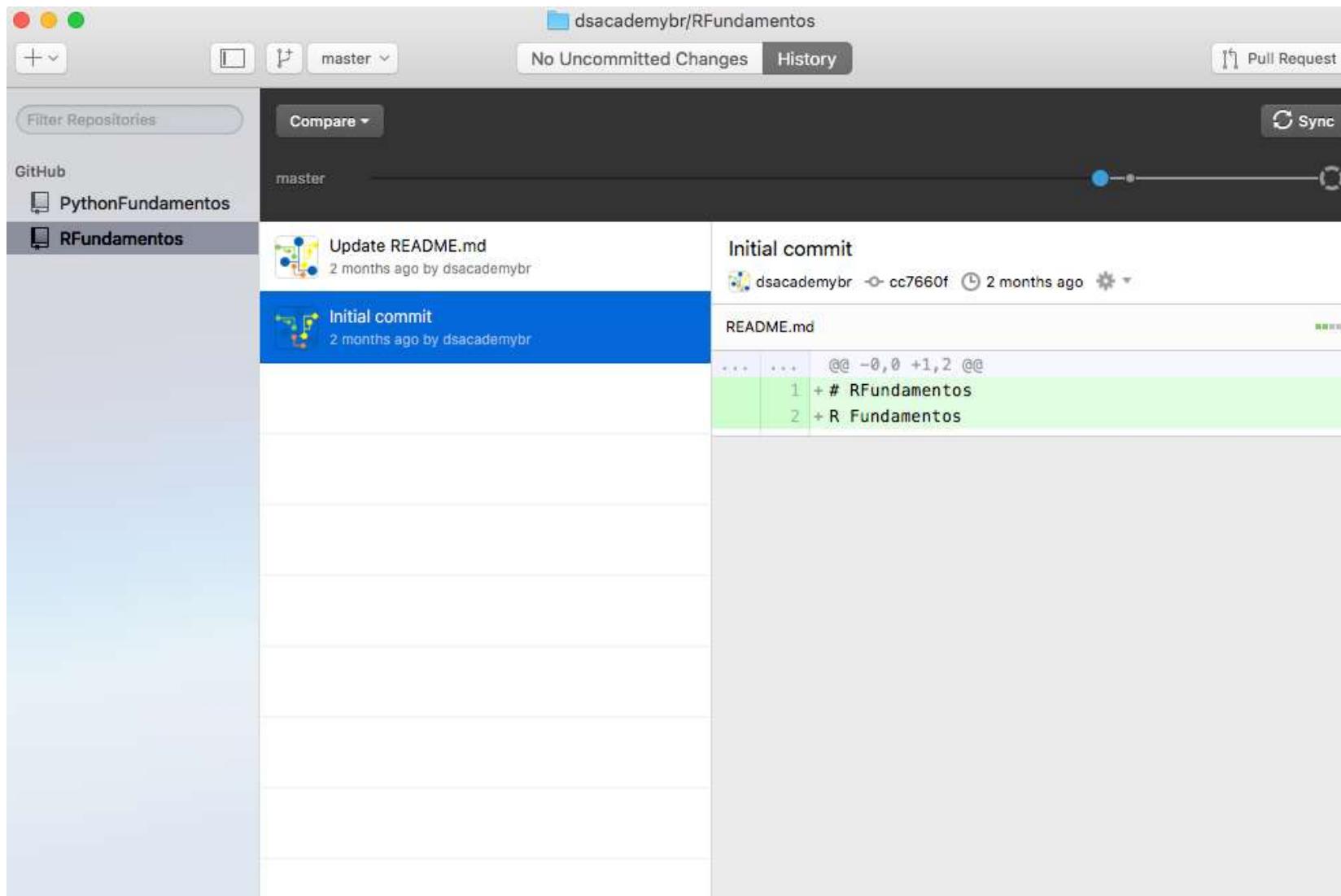
Data Science Academy





Data Science Academy





Data Science Academy



Data Science Academy

dsacademybr

NEW Add a bio

Overview

Repositories

Public activity

Edit profile

Popular repositories

Customize your pinned repositories

PythonFundamentos

Python Fundamentos para Análise de Dados

4 ★

BigDataAnalyticsHadoop

0 ★

BigDataAnalyticsSpark

0 ★

RFundamentos

R Fundamentos

0 ★

6 contributions in the last year

Contribution settings ▾



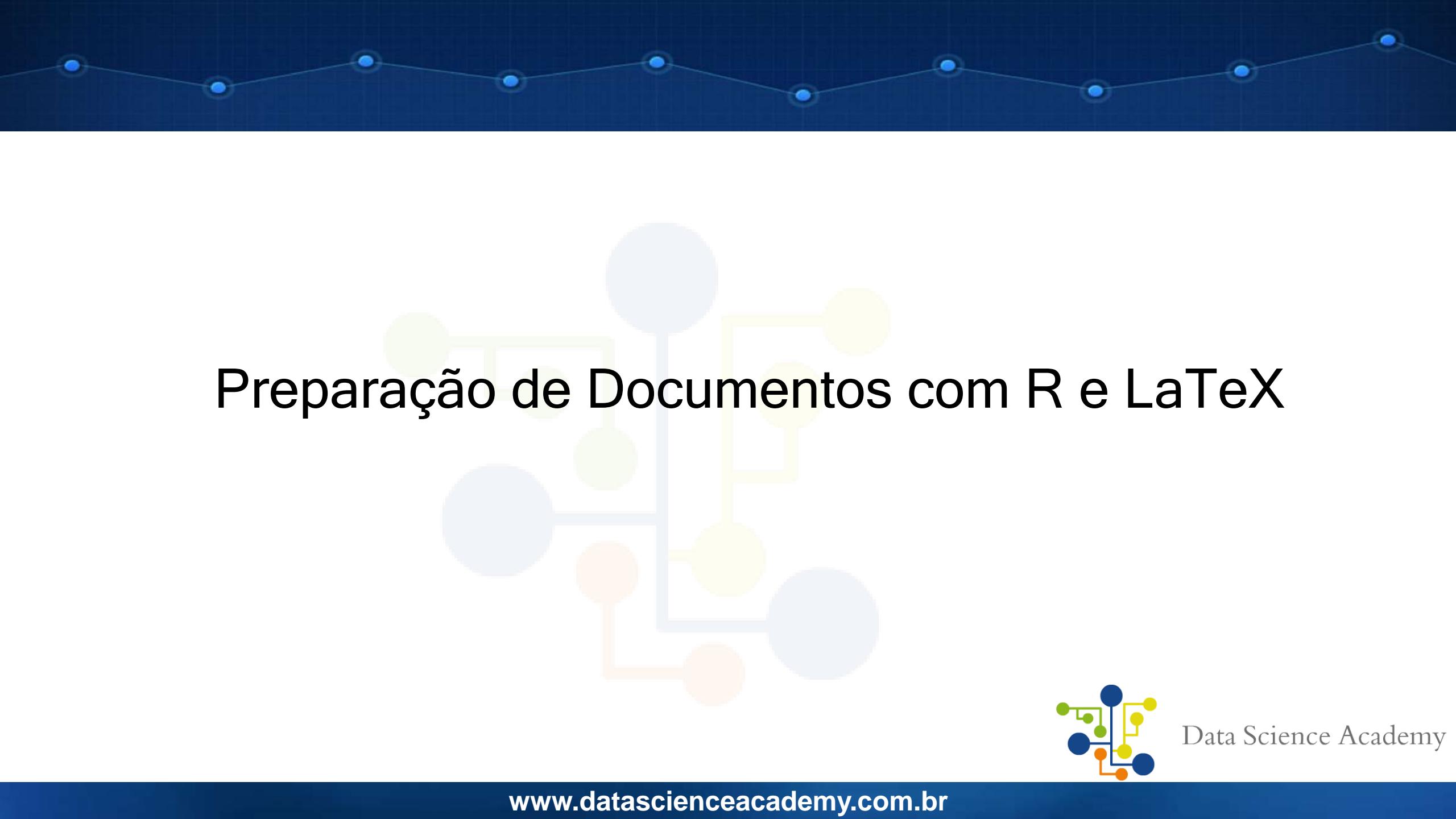
Data Science Academy



obrigado!



Data Science Academy



Preparação de Documentos com R e LaTeX



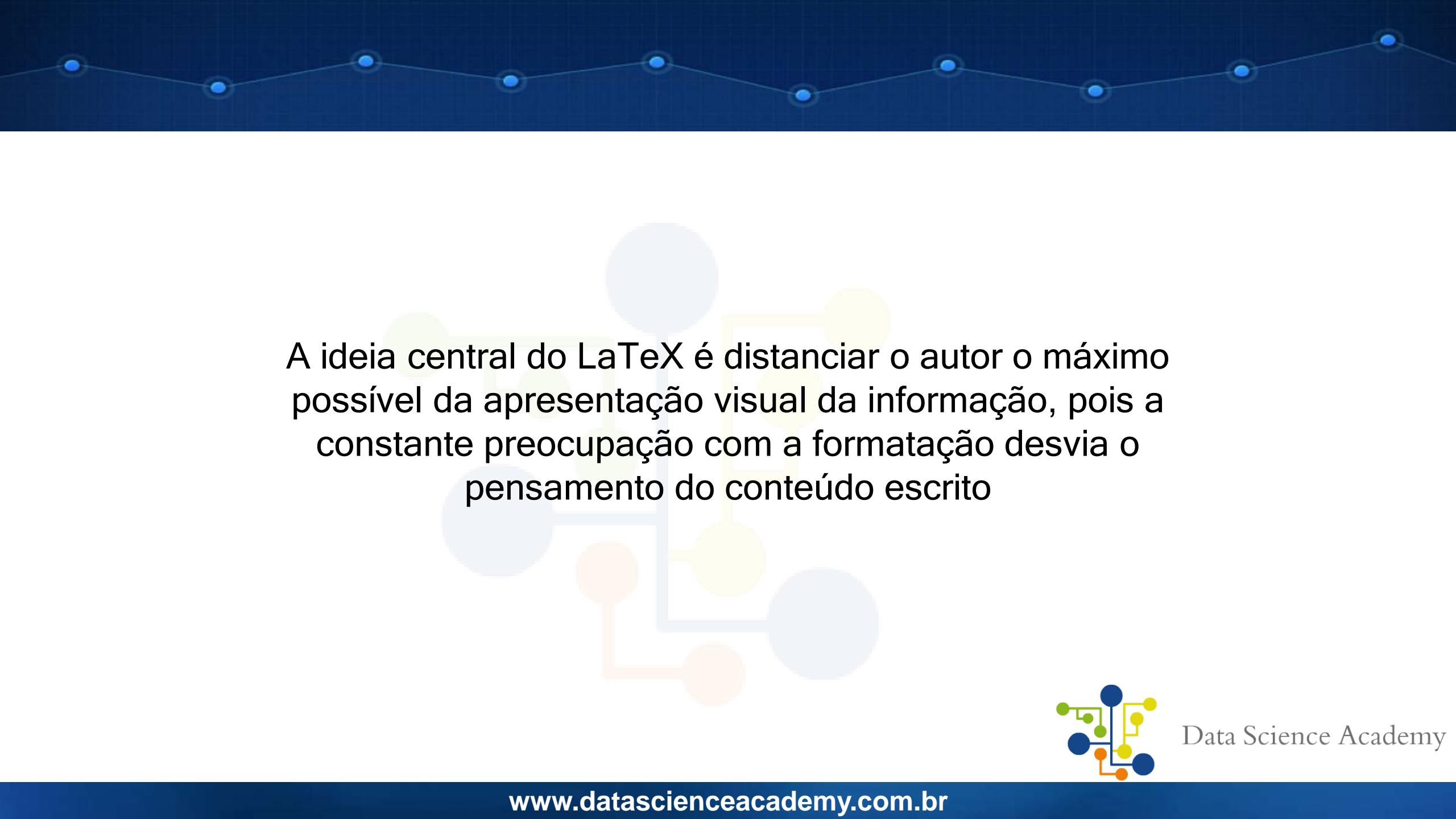
Data Science Academy



O que é LaTeX?



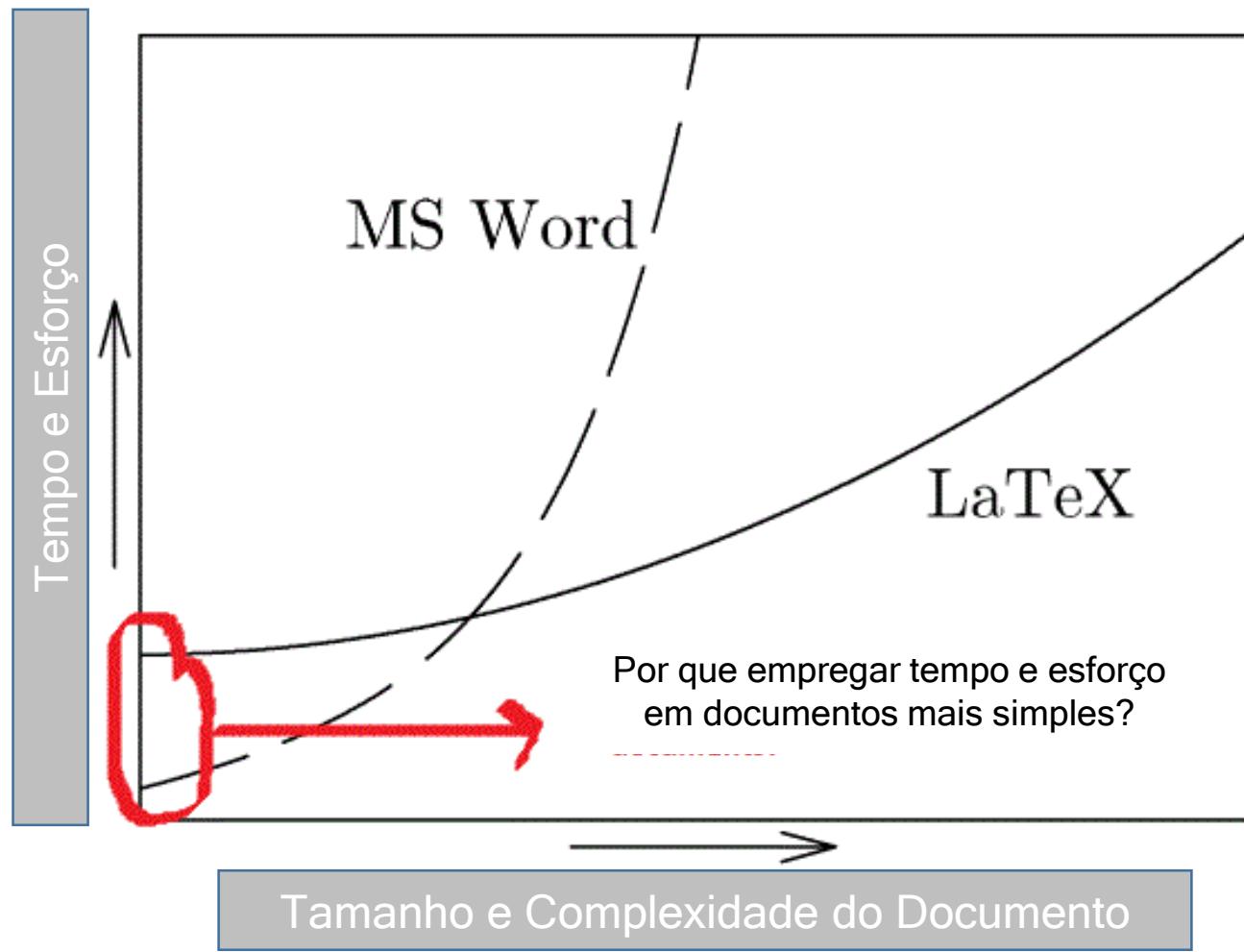
Data Science Academy



A ideia central do LaTeX é distanciar o autor o máximo possível da apresentação visual da informação, pois a constante preocupação com a formatação desvia o pensamento do conteúdo escrito



Data Science Academy



Data Science Academy



Arquivos com extensão .Rnw



Data Science Academy

R e LaTeX

```
1 \documentclass{article}
2 \usepackage[T1]{fontenc}
3
4 \begin{document}
5
6 Here is a code chunk.
7
8 <<foo, fig.height=4>>=
9 1+1
10 letters
11 chartr('xie', 'XIE', c('xie yihui', 'Yihui Xie'))
12 par(mar=c(4, 4, .2, .2)); plot(rnorm(100))
13 @
14
15 You can also write inline expressions, e.g. \$\pi=\Sexpr{pi}s, and \Sexpr{1.598673e8} is a big number.
16
17 \end{document}
```

Arquivos com extensão .Rnw



Data Science Academy

R e LaTeX

Sweave()
knit()



Data Science Academy

R e LaTeX

Sweave()
knit()

Reproducible Research



Data Science Academy

knitr()

```
Usando-Knir.Rtex x | ABC | 
1 \documentclass{article}
2 \usepackage[utf8]{inputenc}
3 \usepackage[english]{babel}
4
5 \begin{document}
6 Você pode digitar seu código R dentro de \LaTeX{} and o código será executado e a saída impressa no documento
7
8 <>>=
9 # Cria sequência de números
10 X = 2:10
11
12 # Mostra medidas estatísticas básicas
13 summary(X)
14
15 @
16 \end{document}
```



Data Science Academy

Sweave()



```
Usando-Knir.Rtex x Usando-Sweave.Rnw x
Format Compile PDF Run
1 \documentclass[a4paper]{article}
2 \usepackage{Sweave}
3
4 \title{Sweave Example 1}
5 \author{Friedrich Leisch}
6
7 \begin{document}
8
9 \maketitle
10
11 In this example we embed parts of the examples from the
12 \texttt{kruskal.test} help page into a \LaTeX{} document:
13
14 - <>>=
15 data(airquality)
16 library(ctest)
17 kruskal.test(Ozone ~ Month, data = airquality)
18 @@
19 which shows that the location parameter of the Ozone
20 distribution varies significantly from month to month. Finally we
21 include a boxplot of the data:
22
23 \begin{center}
24 - <<fig=TRUE,echo=FALSE>>=
25 boxplot(Ozone ~ Month, data = airquality)
26 @@
27 \end{center}
28
29 \end{document}
-----|
```



obrigado!



Data Science Academy

Curta Nossas Páginas nas Redes Sociais

E fique sabendo das novidades em Data Science, Big Data, Internet das Coisas e muito mais...



www.facebook.com/dsacademybr



twitter.com/dsacademybr



www.linkedin.com/company/data-science-academy

www.datascienceacademy.com.br





www.datascienceacademy.com.br

www.datascienceacademy.com.br



Data Science Academy