

Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science Academy

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3



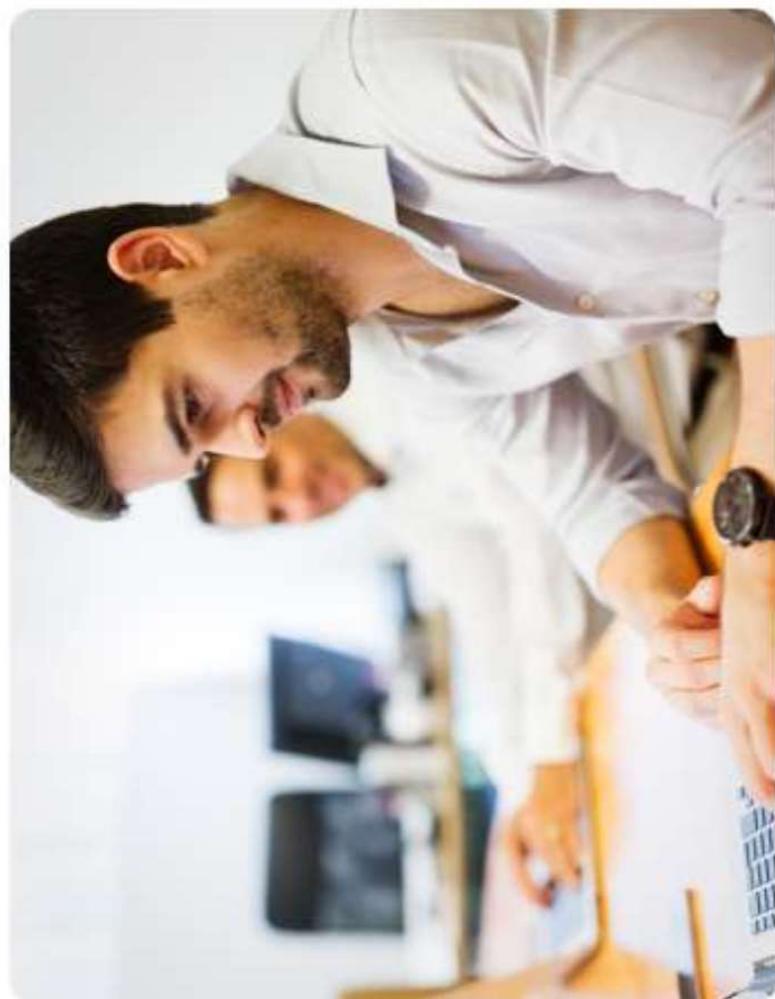
Seja Bem-Vindo



Data Science Academy

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Data Science
Academy



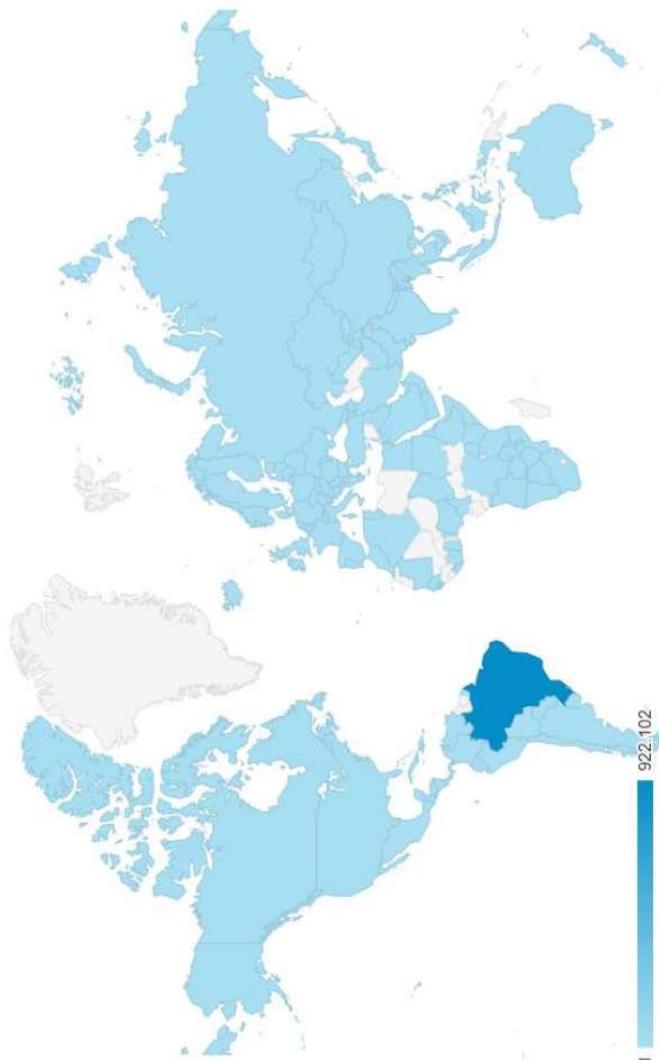
Data Science Academy

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



No Brasil e no Mundo





Seja Bem-Vindo



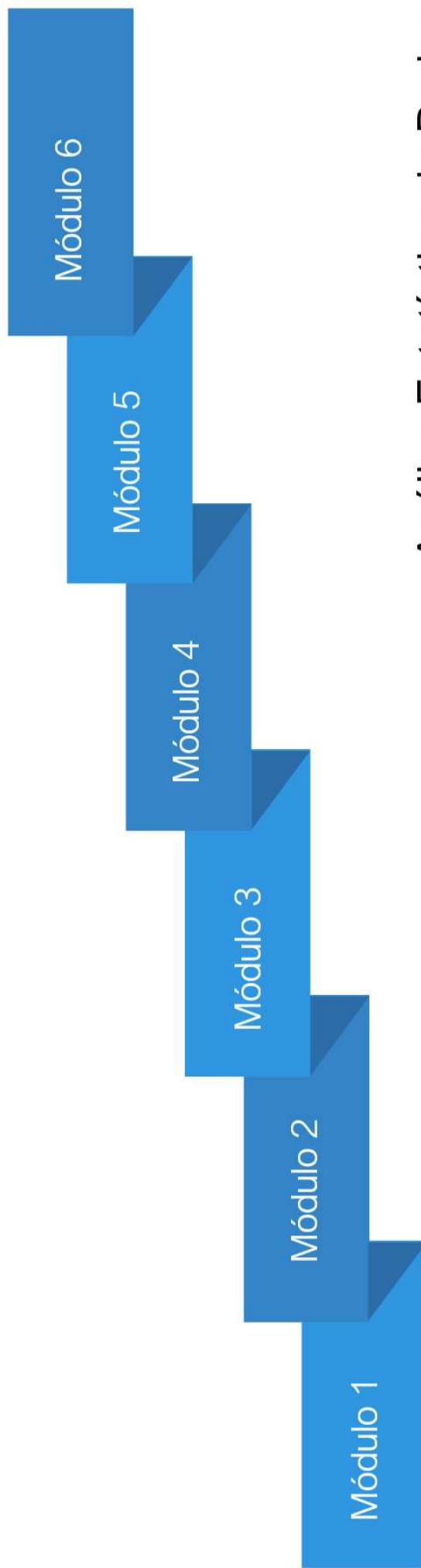
**Este é o Primeiro Curso da
Formação Cientista de Dados**

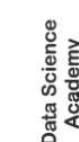


Big Data Analytics com R e Microsoft Azure Machine Learning

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Estrutura do Curso



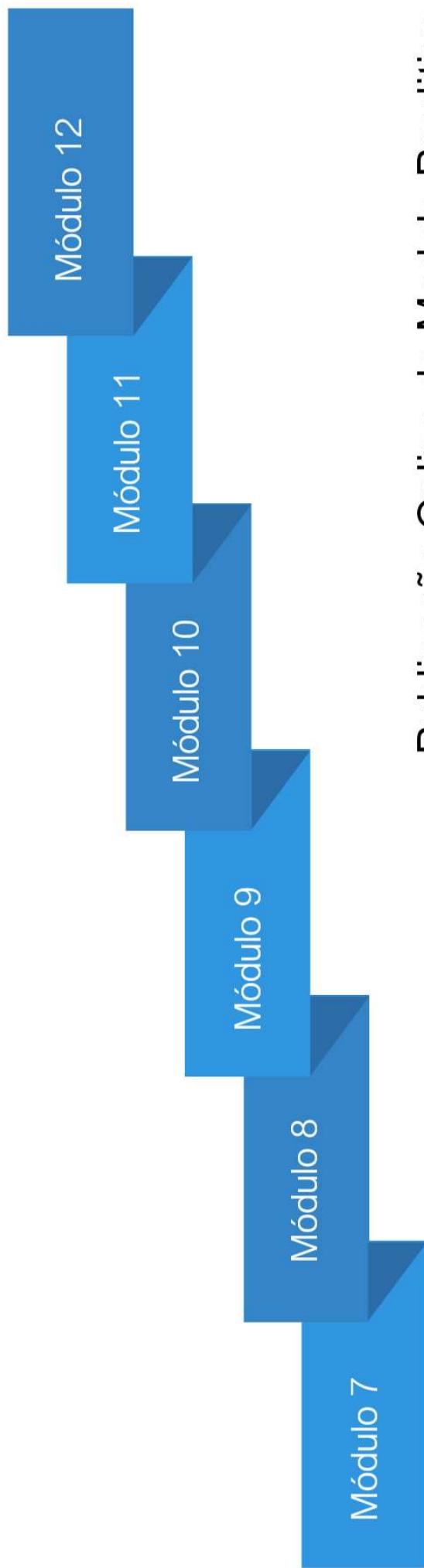


Data Science
Academy

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Big Data Analytics com R e Microsoft Azure Machine Learning

Estrutura do Curso



Publicação Online do Modelo Preditivo



Big Data Analytics com R e Microsoft Azure Machine Learning

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

**KEEP
CALM
QUE ISTO
AINDA NÃO
ACABOU**



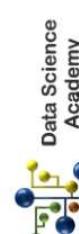
Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Big Data Analytics com R e Microsoft Azure Machine Learning

Mini-Projetos para
estudo dirigido

Projetos com
feedback

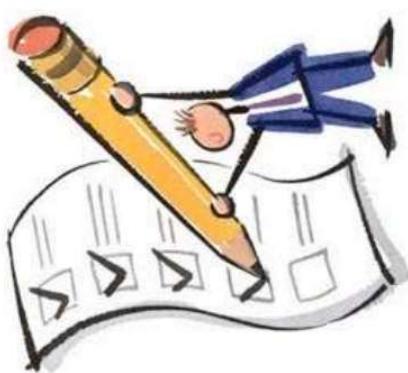
Bônus



Big Data Analytics com R e Microsoft Azure Machine Learning

Avaliação Final

- 50 questões
- 120 minutos de duração
- 3 tentativas





Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Big Data Analytics com R e Microsoft Azure Machine Learning

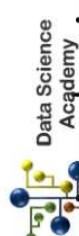


Certificado de Conclusão



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

O que esperar deste curso?



Big Data Analytics com R e Microsoft Azure Machine Learning

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

O que esperar deste curso?

- Teoria e Prática na Medida Certa
- Dinamismo
- Muito Conteúdo
- Alguns Conceitos Avançados
- Objetividade
- Exercícios



O que NÃO esperar deste curso?

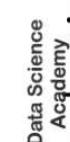


O que NÃO esperar deste curso?

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



E como deve ser sua abordagem neste curso?



E como deve ser sua abordagem neste curso?

4 a 6 horas de dedicação por semana

Prepare seu computador e reproduza todas as aulas

Comunique-se



Bibliografia, Referências e Links Úteis



Big Data Analytics com R e Microsoft Azure Machine Learning

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3





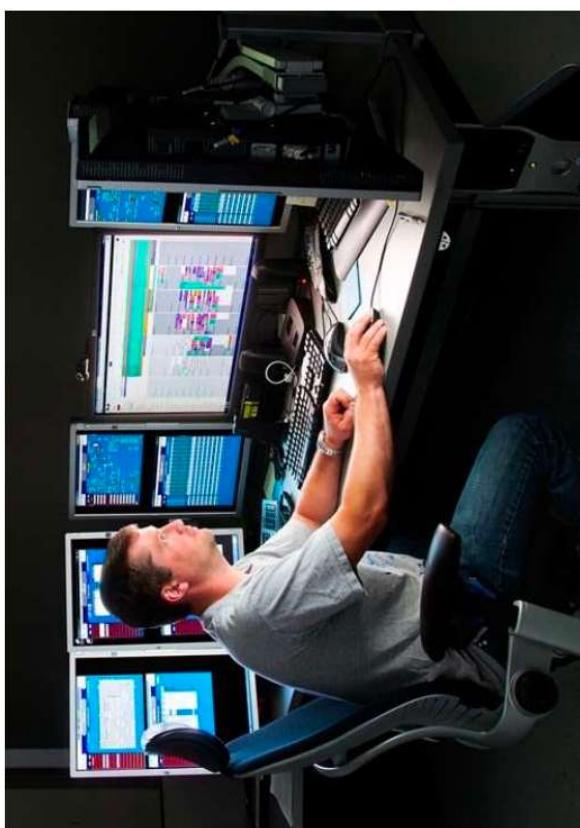
:Divirta-se



A Carreira de Cientista de Dados

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

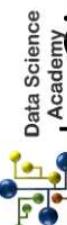
Pelo quarto ano seguido a profissão de Cientista de Dados é a que teve maior demanda em todo mundo.





A Carreira de Cientista de Dados

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3



Estes são os 8 passos para começar sua carreira como Cientista de Dados:

- 1- Faça uma auto-avaliação
- 2- Prepare seu computador
- 3- Estatística e Matemática
- 4- Big Data
- 5- Linguagem de Programação e Machine Learning
- 6- Conhecimento de Negócios
- 7- Técnicas de Apresentação e Visualização de Dados
- 8- Pratique!!



A Carreira de Cientista de Dados

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Erros que devem ser evitados ao longo da caminhada:

Achar que o aprendizado
é fácil e rápido

Começar por problemas
muito complexos

Focar apenas em
programação



Por que Cientistas de Dados usam R?

Por que Cientistas de Dados usam R?





Por que Cientistas de Dados usam 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)

Breve História da Linguagem R



Por que Cientistas de Dados usam R?





Por que Cientistas de Dados usam R?



Versatilidade



Por que Cientistas de Dados usam R?

Linguagem R pode ser usada:

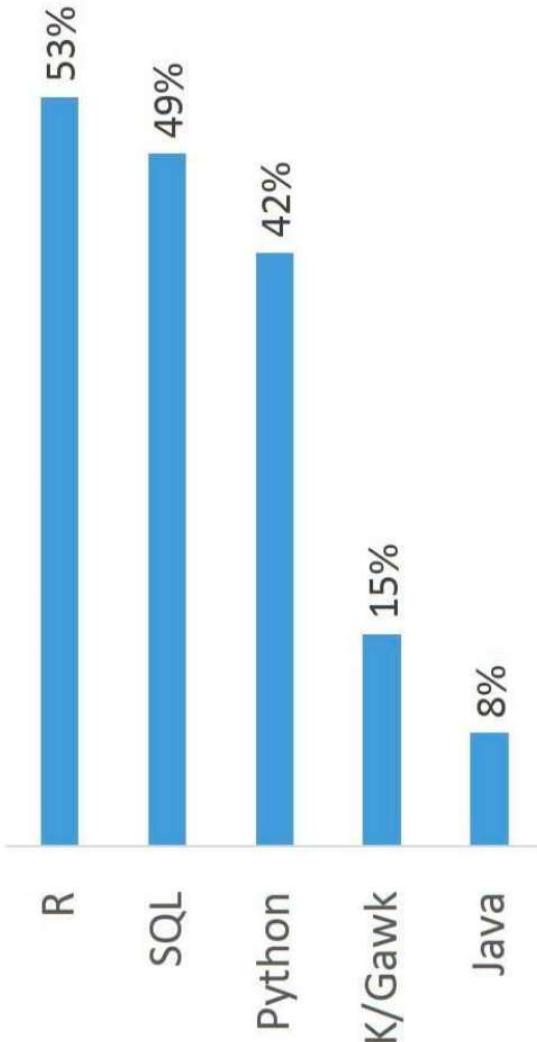
- Extração de Dados
- Limpeza de Dados
- Carregamento e Transformação de Dados
- Análise Estatística
- Modelagem Preditiva
- Machine Learning
- Visualização de Dados





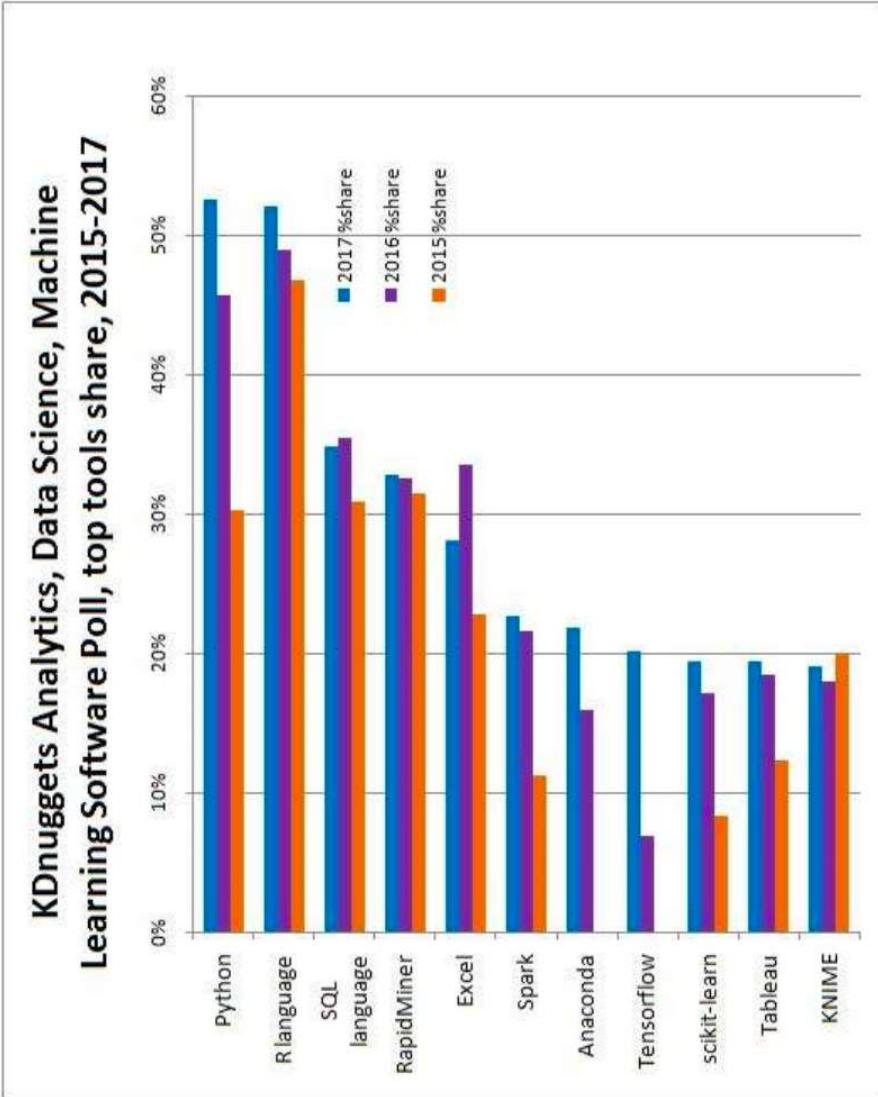
Por que Cientistas de Dados usam R?

Top 5 of programming languages mostly used for data science's activities





Por que Cientistas de Dados usam R?





Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Vantagens e Desvantagens





Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Vantagens

- Grande variedade de pacotes disponíveis
- Flexibilidade e Rapidez
- Análise Estatística





Desvantagens

- Não há interface gráfica (tudo é feito por linha de comando)
- Limitações no uso de memória principalmente com datasets muito grandes



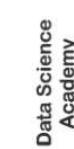


Por que R e Microsoft Azure Machine Learning?

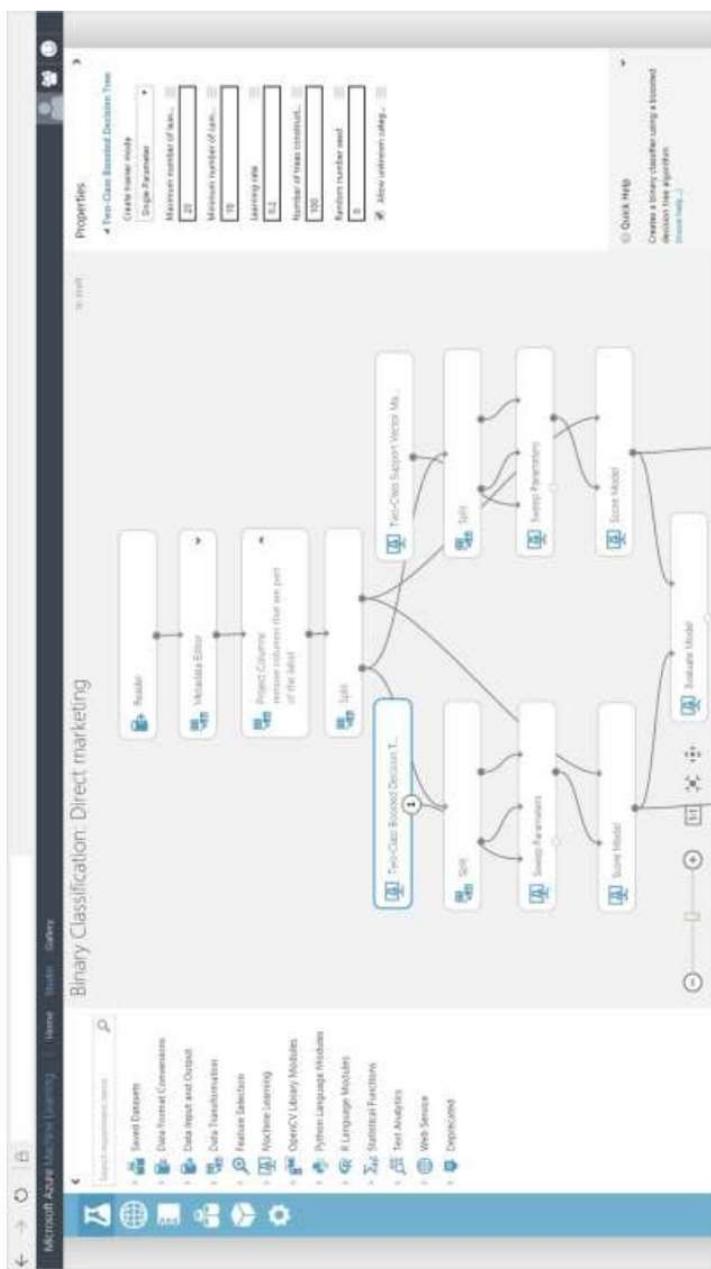
Por que R e Microsoft Azure Machine Learning?



Por que Re Microsoft Azure Machine Learning?



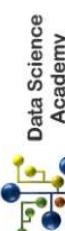
Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3



O que é o Azure Machine Learning?



Por que Re Microsoft Azure Machine Learning?



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

The screenshot shows the Microsoft Azure Machine Learning studio interface. On the left, there's a navigation bar with options like '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 area displays a workflow titled 'R word cloud'. This workflow consists of several steps: 'convention.csv' (a dataset), 'Execute R Script' (with code to load 'wordcloud'), 'convention\$words', 'conventions\$documents', 'conventions\$republicans', and 'conventions\$republicans'. A connection from 'convention\$words' leads to 'Execute R Script' (with code to generate a word cloud). Another connection from 'conventions\$documents' leads to 'conventions\$republicans'. The 'Execute R Script' step has a preview window showing a word cloud visualization. Below the workflow, there's an 'Experiment Properties' section with fields for 'START TIME', 'END TIME', and 'Execute R Script'. The 'Execute R Script' section also includes a note: 'Execute the given R script using R 3.1.0 [more...]'.

```

1 #adapted from https://gitlab.com/mrdlctaylor/2671578
2 library(wordcloud)
3 library(RColorBrewer)
4 library(png)
5 conventions <- mail.mapInputPort(1) # class: data.frame
6
7 png("cloud.png")
8 wordCloud(conventions$wordper25k, # words
9 conventions$documents, # frequencies
10 conventions$republicans, # size of largest and smallest words
11 scale = c(4,1), # size of largest and smallest words
12 colors = brewer.pal(9, "Blues"), # number of colors, palette
13 rot.per = 0) # proportion of words to rotate 90 degrees
14 dev.off()
15 wordCloud(conventions$wordper25k,
16 conventions$republicans,
17 scale = c(4,1),
18 colors = brewer.pal(9, "Reds"),
19 rot.per = 0)
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

```

Machine Learning Studio



Por que Re Microsoft Azure Machine Learning?

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

R word cloud

Draft saved at 11:07:11

Properties

Execute R Script

R Script

```
1 #adapted from https://github.com/mellicode/R/2015.08
2 library(wordCloud)
3 library(RColorBrewer)
4 library(png)
5 conventions <- maml.mapInputPort(1) # class: data.frame
6
7 plot("Cloud.Rnw")
8 wordCloud(conventions$words[1:25k], # words
9   conventions$descripts, # frequencies
10  scale = c(4,1), # size of largest and smallest words
11  colors = brewer.pal(10, "Blues"), # number of colors, palette
12  rot.per = 0) # proportion of words to rotate 90 degrees
13 dev.off()
14
15 wordCloud(conventions$words[25k],
16   conventions$publicans,
17   scale = c(4,1),
18   colors = brewer.pal(9, "Reds"),
19   rot.per = 0)
20
21 dev.off()
```

Random Seed A2

Experiment Properties

START TIME END TIME

Execute R Script

Execute the given R script using R 3.1.0
(more...)

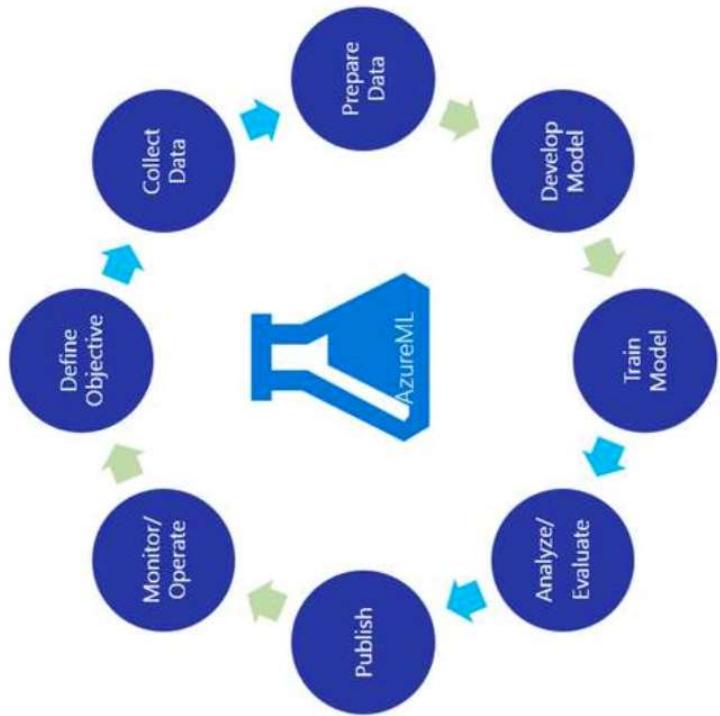
NEW NEW HISTORY SAVE DISCARD CHANGES CANCEL ABORT

Módulos prontos para análises com R, Python e SQL



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Azure Machine Learning Workflow



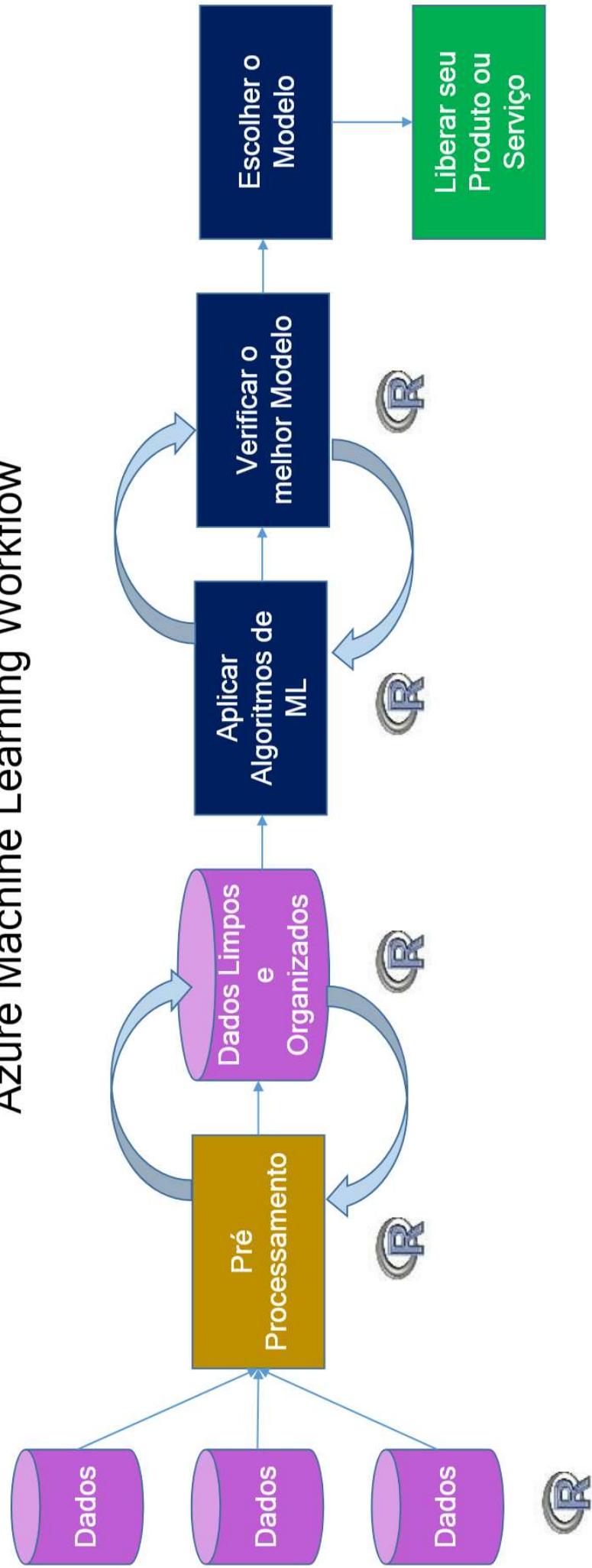
Azure Machine Learning Workflow



Azure Machine Learning Workflow

Data Science Academy Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

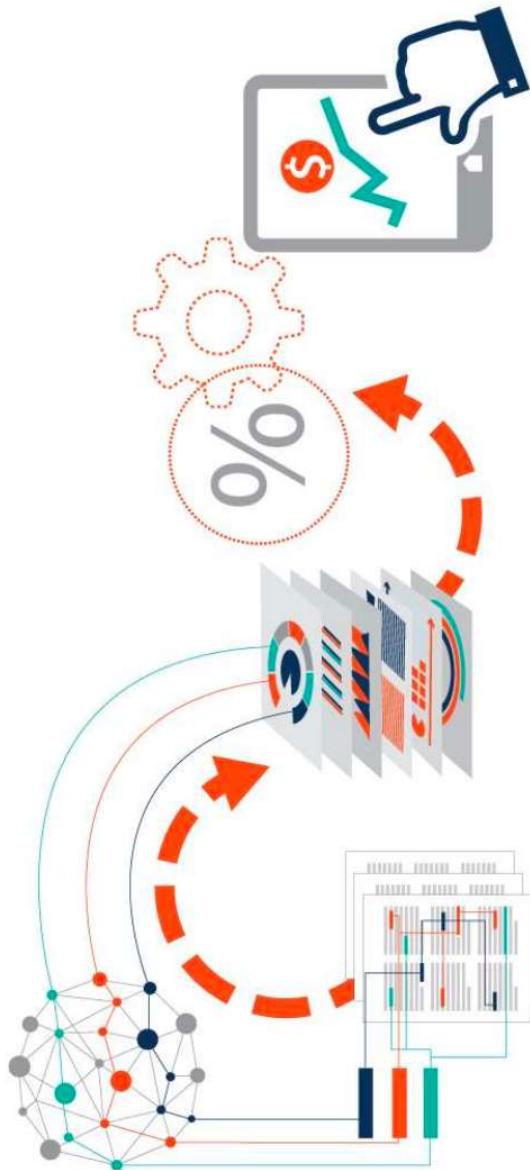
Azure Machine Learning Workflow





Dados x Informação x Conhecimento x Inteligência

Dados x Informação x Conhecimento x Inteligência





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



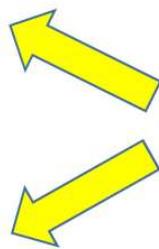
Dado



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



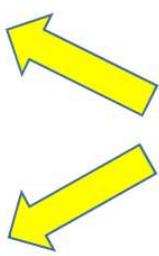
Dados x Informação x Conhecimento x Inteligência

As pessoas trabalham com informações mas a tecnologia armazena dados



Dados x Informação x Conhecimento x Inteligência

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



Conhecimento



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Dados x Informação x Conhecimento x Inteligência



Data Science
Academy





Dados x Informação x Conhecimento x Inteligência

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

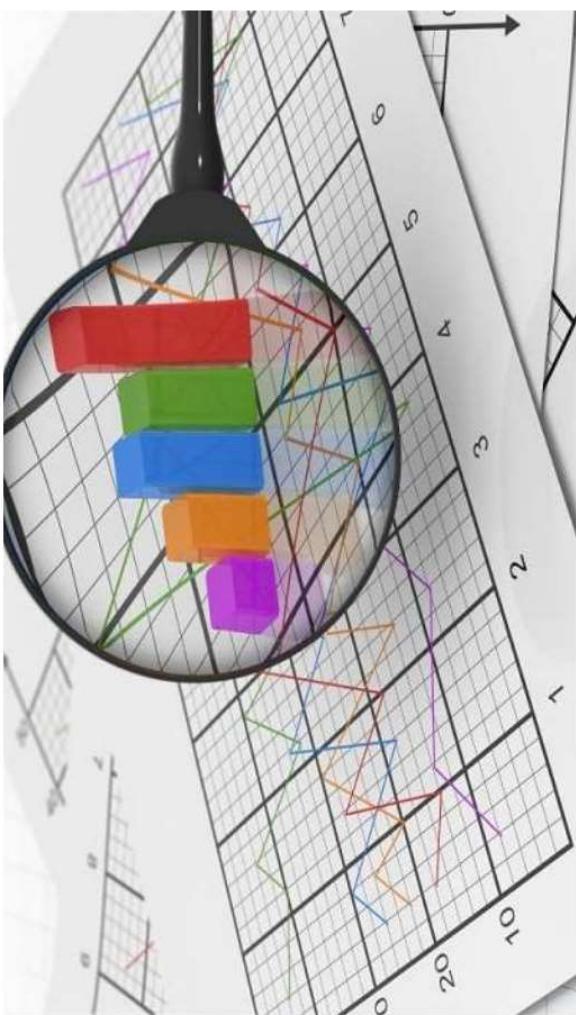
Inteligência - capacidade de resolver problemas, usando o conhecimento, através das informações disponíveis

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



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Dados x Informação x Conhecimento x Inteligência



Big Data Analytics

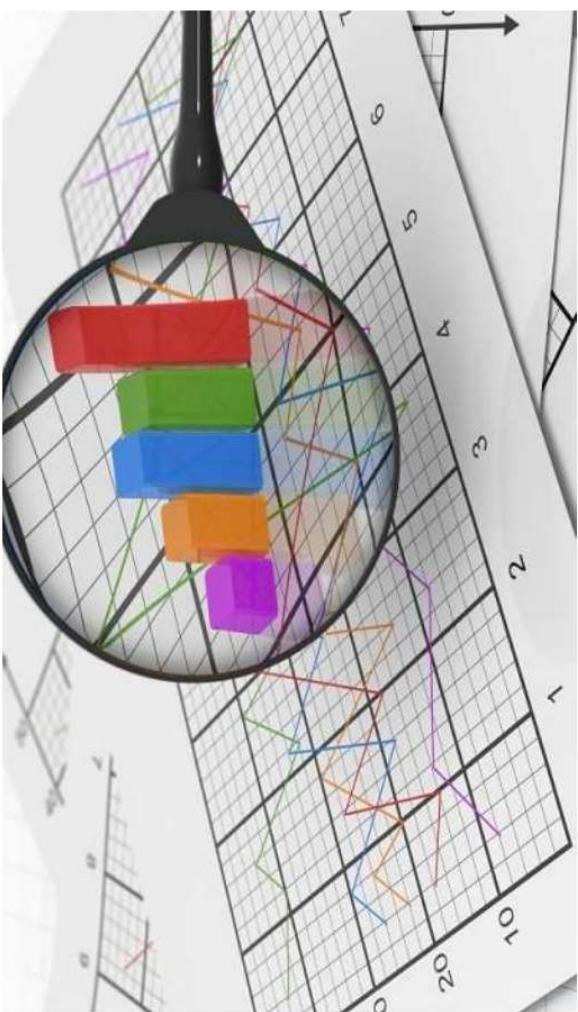
Extrair conhecimento a partir dos dados (muitos dados)



Dados x Informação x Conhecimento x Inteligência

Machine Learning

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

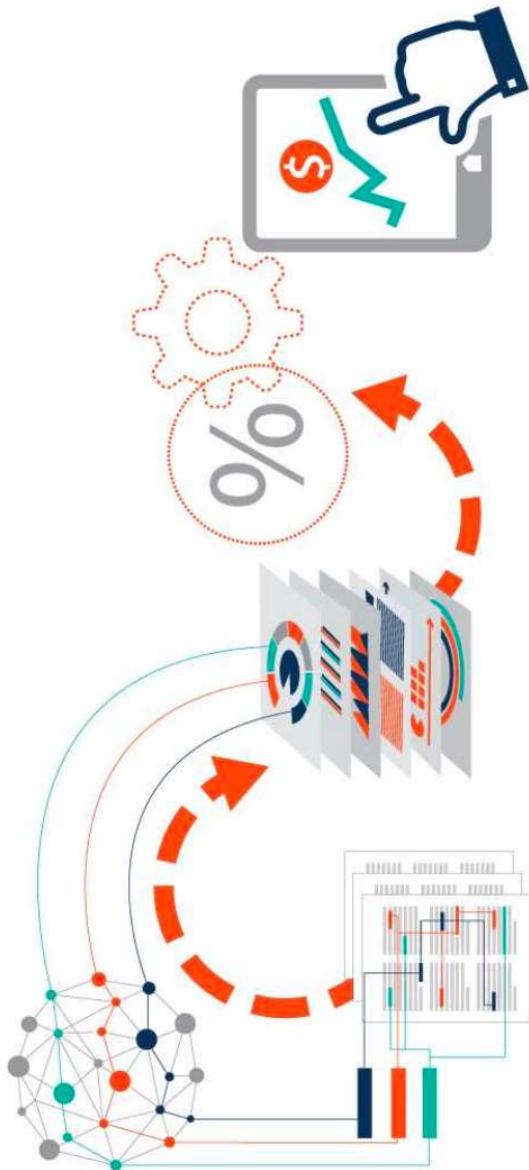




A Importância do Big Data Analytics

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

A Importância do Big Data Analytics





A Importância do Big Data Analytics

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

Data Science
Academy

Tempoade Perfeita

- Crescimento exponencial do volume de dados
- Preço baixo do armazenamento de dados
- Alta capacidade de processamento dos computadores





A Importância do Big Data Analytics



Dados são combustível para a Análise Preditiva



A Importância do Big Data Analytics

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



A Importância do Big Data Analytics

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Data Science
Academy

Tudo isso é alimentado por dados. Muitos dados. Big Data.



A Importância do Big Data Analytics

Tudo isso é alimentado por dados. Muitos dados. Big Data.

Quando aplicamos Data Science para análise de Big Data temos o que é chamado de **Big Data Analytics** e sua importância é cada vez maior no mundo atual.



Data Science e a Evolução dos Sistemas Analíticos

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a6/2005e4cd1ac8b45a3

Data Science e a Evolução dos Sistemas Analíticos



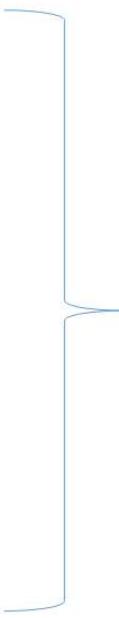


Data Science e a Evolução dos Sistemas Analíticos

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a6/2005e4cd1acb8b45a3

Estamos vivendo a era da explosão dos dados

bit byte kilobyte megabyte gigabyte terabyte petabyte exabyte zettabyte yottabyte



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

Big Data



Data Science
Academy

Data Science e a Evolução dos Sistemas Analíticos

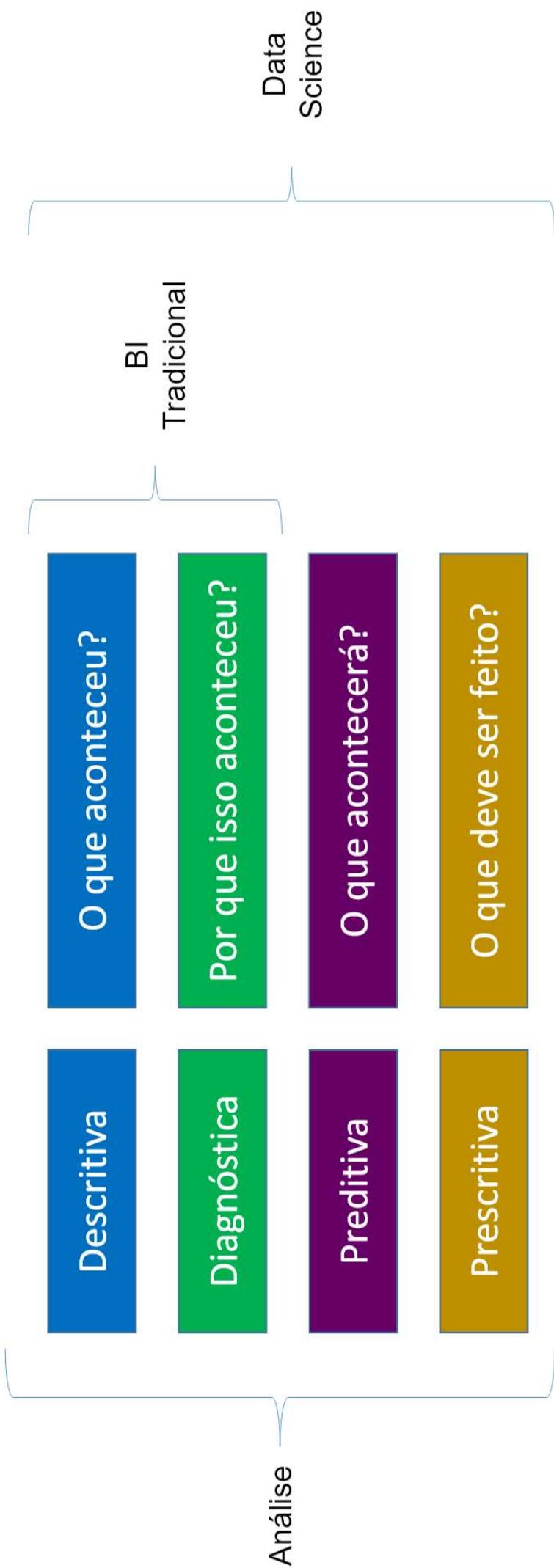
Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cde1acb8b45a3





Data Science e a Evolução dos Sistemas Analíticos

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a6/2005e4cd1acb8b45a3

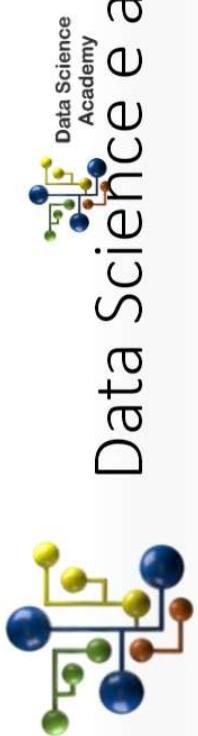




Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Data Science e a Evolução dos Sistemas Analíticos

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



Data Science e a Evolução dos Sistemas Analíticos

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

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



Data Science Academy Data Science Academy phelipe.utsempreboni@outlook.com 5c8a6/2005e4cded1acb8b45a3

Data Science e Inteligência Artificial - Há Diferença?

Data Science e Inteligência Artificial - Há Diferença?





Data Science e Inteligência Artificial - Há Diferença?

Data Science e Inteligência Artificial - Há Diferença?





Data Science e Inteligência Artificial - Há Diferença?





Data Science Academy Data Science Academy phelipe.utsempreboni@outlook.com 5c8a6/2005e4cded1acb8b45a3

Data Science e Inteligência Artificial - Há Diferença?

Data Science é uma área multidisciplinar que envolve Estatística, Matemática, Machine Learning, Programação, Banco de Dados, Conhecimento de Negócios, Técnicas de Apresentação e Visualização de Dados.



Data Science e Inteligência Artificial - Há Diferença?

Data Science é uma área multidisciplinary que envolve Estatística, Matemática, Machine Learning, Programação, Banco de Dados, Conhecimento de Negócios, Técnicas de Apresentação e Visualização de Dados.

Aplicamos Data Science para analisar dados, construir modelos preditivos e resolver problemas de negócio. Ou seja, aplicamos Data Science quando estamos trabalhando com Machine Learning e Inteligência Artificial.



A Importância do Portfólio Para o Cientista de Dados

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

A Importância do Portfólio Para o Cientista de Dados





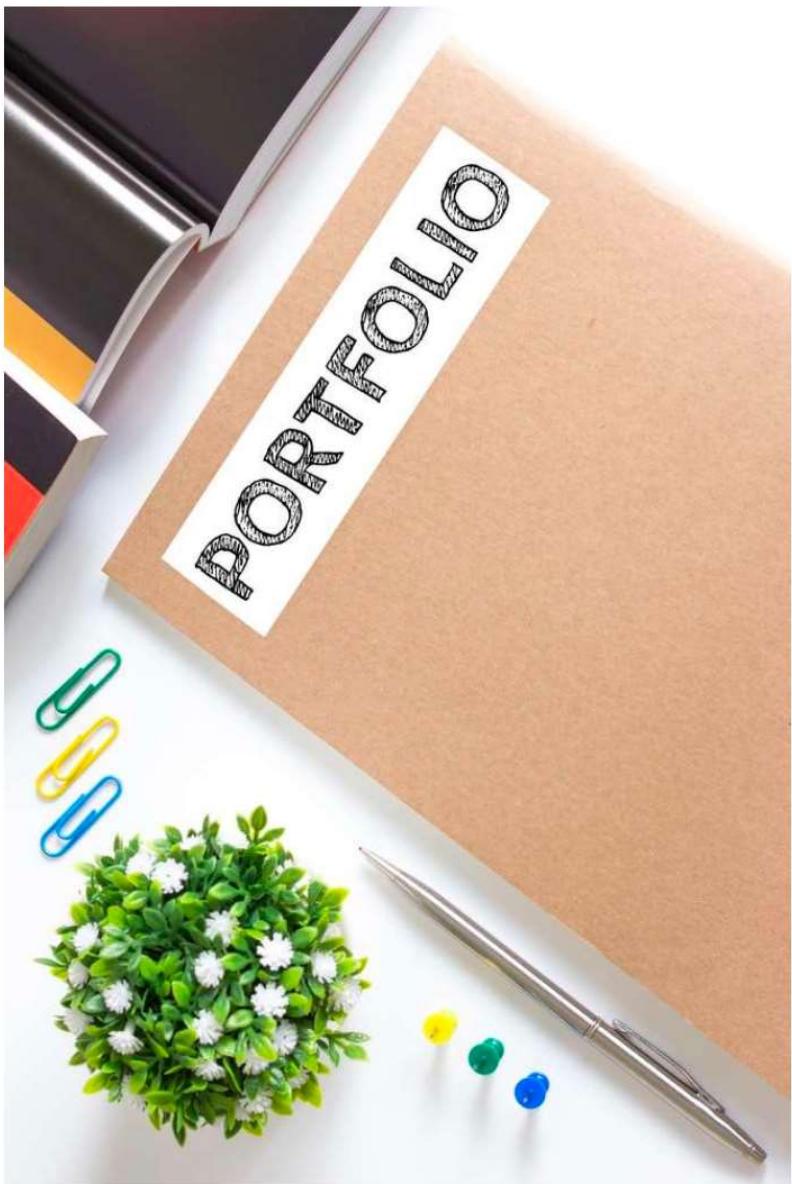
Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

No curso **Preparação Para Carreira de Cientista de Dados** o aluno aprende como iniciar um portfólio usando o Github.



A Importância do Portfólio Para o Cientista de Dados

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3





A Importância do Portfólio Para o Cientista de Dados

Fundamental para quem está iniciando em Data Science	Visibilidade do seu conhecimento	Demonstração das suas habilidades	Oportunidade de praticar e resolver problemas	Interação com outros profissionais	Aprendizado contínuo
--	----------------------------------	-----------------------------------	---	------------------------------------	----------------------



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cded1acb8b45a3

O Que Há de Novo na FCD 2.0?



2.0

O Que Há de Novo na FCD 2.0?



O Que Há de Novo na FCD 2.0?

Data Science Academy Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Formação Cientista de Dados 1.0

Programa pioneiro no Brasil criado em 2016

Grande quantidade de conteúdo em mais de 340 horas

Cursos Práticos e voltados ao Mercado de trabalho

Milhares de alunos capacitados

- + 1200 vídeos
- + 200 manuais em pdf
- + 320 scripts



O Que Há de Novo na FCD 2.0?

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Formação Cientista de Dados 2.0

Conteúdo
100%
atualizado e
revisado

Novos
conteúdos,
scripts,
projetos e
exercícios

Nova
abordagem
de ensino

Mais
dinamismo e
interatividade



O Que Há de Novo na FCD 2.0?

Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Formação Cientista de Dados 2.0

Esta atualização não tem custo para os alunos já matriculados na Formação ou nos cursos individuais da Formação.



O Que Há de Novo na FCD 2.0?

Data Science Academy Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

Formação Cientista de Dados 2.0

1. Big Data Analytics com R e Microsoft Azure Machine Learning 2.0
2. Big Data Real-Time Analytics com Python e Spark 2.0
3. Engenharia de Dados com Hadoop e Spark 2.0
4. Machine Learning com R e Python 2.0
5. Business Analytics 2.0
6. Visualização de Dados e Design de Dashboards 2.0
7. Preparação Para Carreira de Cientista de Dados 2.0
 - 7.1 – Curso de Sistema Operacional Linux 2.0
 - 7.2 – E-Gov Analytics
 - 7.3 – Empreendedorismo em Data Science



O Que Há de Novo na FCD 2.0?

Data Science Academy Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cdde1acb8b45a3

Como será a atualização para a versão 2.0:

- 1- Todos os alunos matriculados nos cursos da FCD terão a nova versão automaticamente.
Nada precisa ser feito.
- 2- Os novos capítulos já atualizados serão liberados todos os dias entre 00:00 e 01:00 da manhã pelo horário de Brasília. Faça um refresh no browser se estiver acessando a DSA neste horário.
- 3- A cada atualização, o capítulo antigo será inteiramente removido e o capítulo novo publicado.
- 4- Seguiremos a ordem dos cursos da FCD e toda atualização vai ocorrer entre Janeiro e Março/2019.



O Que Há de Novo na FCD 2.0?

Ainda não comecei a Formação, o que fazer?

Aproveite e comece agora com a versão 2.0

Estou no meio do curso, o que devo fazer?

Caso você precise muito do Certificado, conclua o curso e depois revise com a versão 2.0. Ou então, volte ao início e aproveite o novo conteúdo.

Já conclui o curso, o que devo fazer?

Você pode refazer o curso e aproveitar o novo conteúdo ou apenas revisar as aulas onde teve mais dúvidas. Quem já obteve o Certificado, não precisa refazer a avaliação final do curso.



Data Science Academy phelipe.utsempreboni@outlook.com 5c8a62005e4cd1acb8b45a3

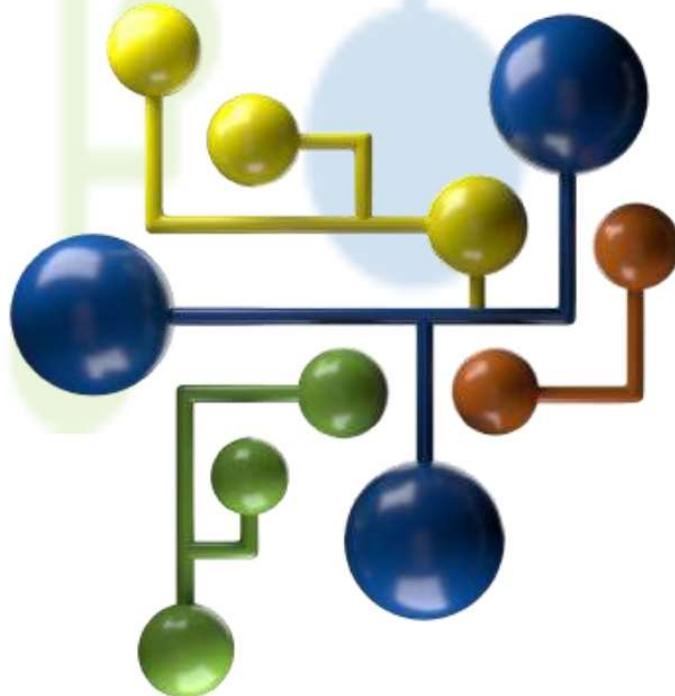
O Que Há de Novo na FCD 2.0?





Data Science Academy phelipe.utsempreboni@outlook.com 5c8a6/2005e4cded1acb8b45a3

Muito Obrigado por Participar!



Tenha uma Excelente Jornada de Aprendizagem.

Equipe Data Science Academy