
ald
eia
Sejam
bem-vindos!
Utilize a nossa redes de wi-fi:
#ALDEIA
utilizando a senha
b8xygf

A Aldeia é muito mais que espaço

Somos um movimento de desenvolvimento de realizadores.

Temos tudo que realizadores precisam para fazer uma ideia dar certo.

http://aldeia.cc

Cursos

Confrarias

Coworking

Offices

Networking

Eventos

Acelerações

Não passe perrengue

Tem água e café à vontade, e um doce e um salgado para você pegar na hora que quiser.

Temos banheiros nos dois andares da **Cândido**:

- Primeiro andar: atrás da recepção
- Segundo andar: ao lado da escada

E atrás da recepção na unidade **Estação**.

Se algo não estiver certo, fale com a nossa equipe

Faça parte da nossa Tribo

Receba os **materiais do curso** e seu **certificado** de participação por meio da nossa comunidade virtual.

Acesse https://aldeia.cc/chamado e faça sua solicitação para fazer parte da plataforma, utilizando o e-mail da compra do curso para se identificar.

Tire uma foto deste QR code e vá direto para a página da Tribo


Curso de Data Science
Charles Adriano dos Santos
Rafael Roberto Dias
ald
eia

Agenda

1 – Agenda

2 – Welcome to R – Parte I

3 – Homework - ETL

4 – Namorando Dados (SQL)

5 – R – Parte II

6 – Welcome to Python

Manhã

Horário	Assunto
09:30	Welcome to R – Parte I
11:30	Homework - ETL
12:30	Almoço

Tarde

Horário	Assunto
13:30	Namorando Dados (SQL)
15:00	R – Parte II: Qualidade dos Dados e Variáveis Relevantes
17:00	Welcome to Python: Básico, Numpy, Pandas e Banco

Nos Episódios Anteriores...

Profissão Data Science

Estatística & Ciência da Computação

Desafio Agro XP

- ETL
- Modelagem de Dados
- Banco de Dados
- Queries SQL

Welcome to R – Parte I

1 – Agenda

2 – Welcome to R – Parte I

3 – Homework - ETL

4 – Namorando Dados (SQL)

5 – R – Parte II

6 – Welcome to Python

Quais são os principais softwares Estatísticos?

- **MiniTab - Software Matemático e Estatístico**
- **SAS - Statistical Analysis System**
- **SPSS - Statistical Package for the Social Sciences**
- **S-PLUS - Versão paga do R**
- **Python - Linguagem Interpretada**
- **R - (Ross e Robert)**

Detalhes Software R

- **Linguagem Alto Nível -** Longe do código de máquina e mais próximo à linguagem humana
- **Interpretada -** O programa resultante não é executado diretamente pelo sistema operacional ou processador
- **Script -** Programas escritos para um sistema de tempo que automatiza a execução de tarefas
- **Orientada a objetos -** Abstração, Encapsulamento, Herança e Polimorfismo

Detalhes Software R

O R disponibiliza uma ampla variedade de

- Técnicas estatísticas
- Gráficos
- Modelos Lineares
- Modelos não Lineares
- Testes estatísticos clássicos
- Análises de Séries Temporais
- Classificação
- Agrupamento
- Machine Learning
- Artificial Inteligence

Detalhes Software R

- O R é utilizado através de um Interpretador de comandos
- Ao escrever 4 + 4 na linha de comando, obtém-se o seguinte resultado:

```
> 4 + 4
[1] 8
> 
```

- A linguagem R suporta matrizes aritméticas, escalares, vetores, matrizes, quadros de dados (similares a tabelas numa base de dados relacional) e listas

Detalhes Software RStudio

- RStudio é um software livre de ambiente de desenvolvimento, e que possui uma interface gráfica amigável

- O R Studio é uma interface para o R, com diversas utilidades diferentes que a tornam uma ferramenta mais simples em comparação ao R

- Ele possui duas versões: RStudio Desktop, que roda localmente em desktop e RStudio Server, que permite acessá-lo usando um navegador web enquanto ele roda em um servidor GNU/Linux remoto

Aprendendo Linguagem R no RStudio

Bora Praticar?

Homework - ETL

1 – Agenda

2 – Welcome to R – Parte I

3 – Homework - ETL

4 – Namorando Dados (SQL)

5 – R – Parte II

6 – Welcome to Python

O Trabalho do Cientista de Dados > Desafio Curso

1. Definição do problema e levantamento de perguntas a serem respondidas ☑
2. Planejamento do processo de Data Science ☑
3. Coleta de dados☑
4. Processamento e limpeza dos dados
5. Armazenamento dos dados ☑
6. Análise de dados
7. Construção e validação de algoritmos e modelos
8. Data Visualization
9. Disseminação da informação
10. Colocar modelo em produção

Homework - ETL

1 – Agenda

2 – Welcome to R – Parte I

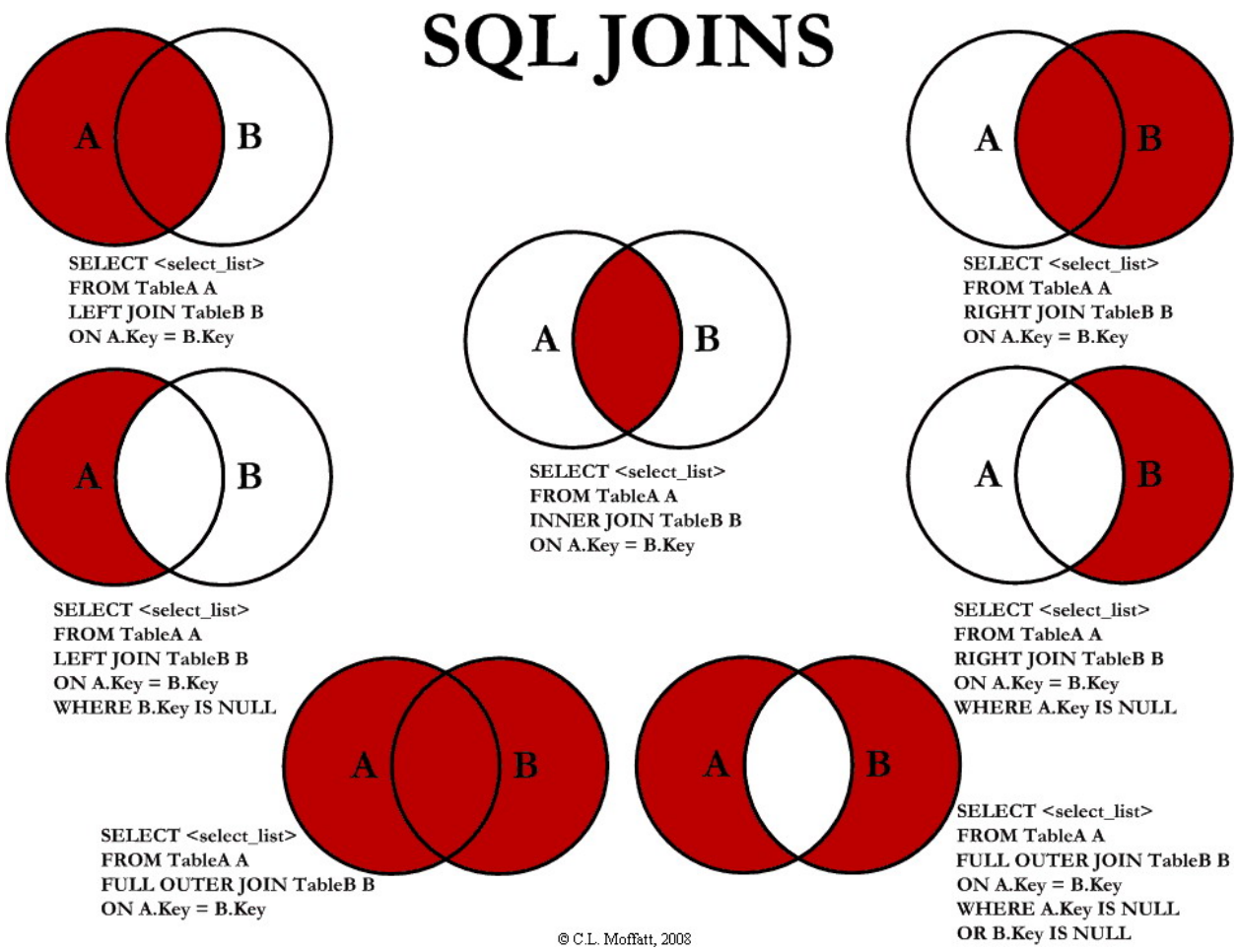
3 – Homework - ETL

4 – Namorando Dados (SQL)

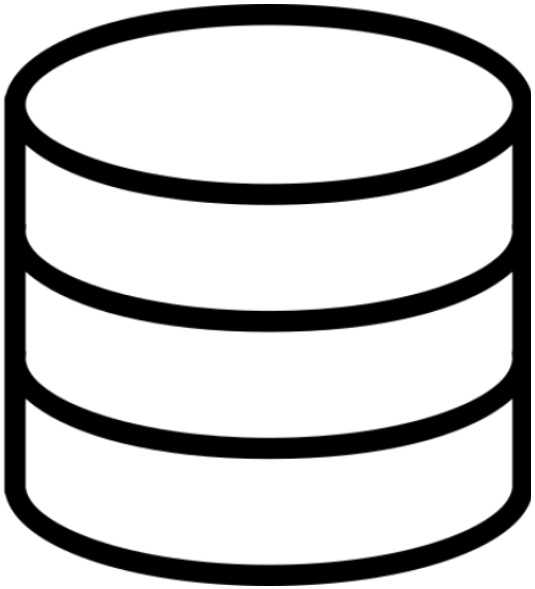
5 – R – Parte II

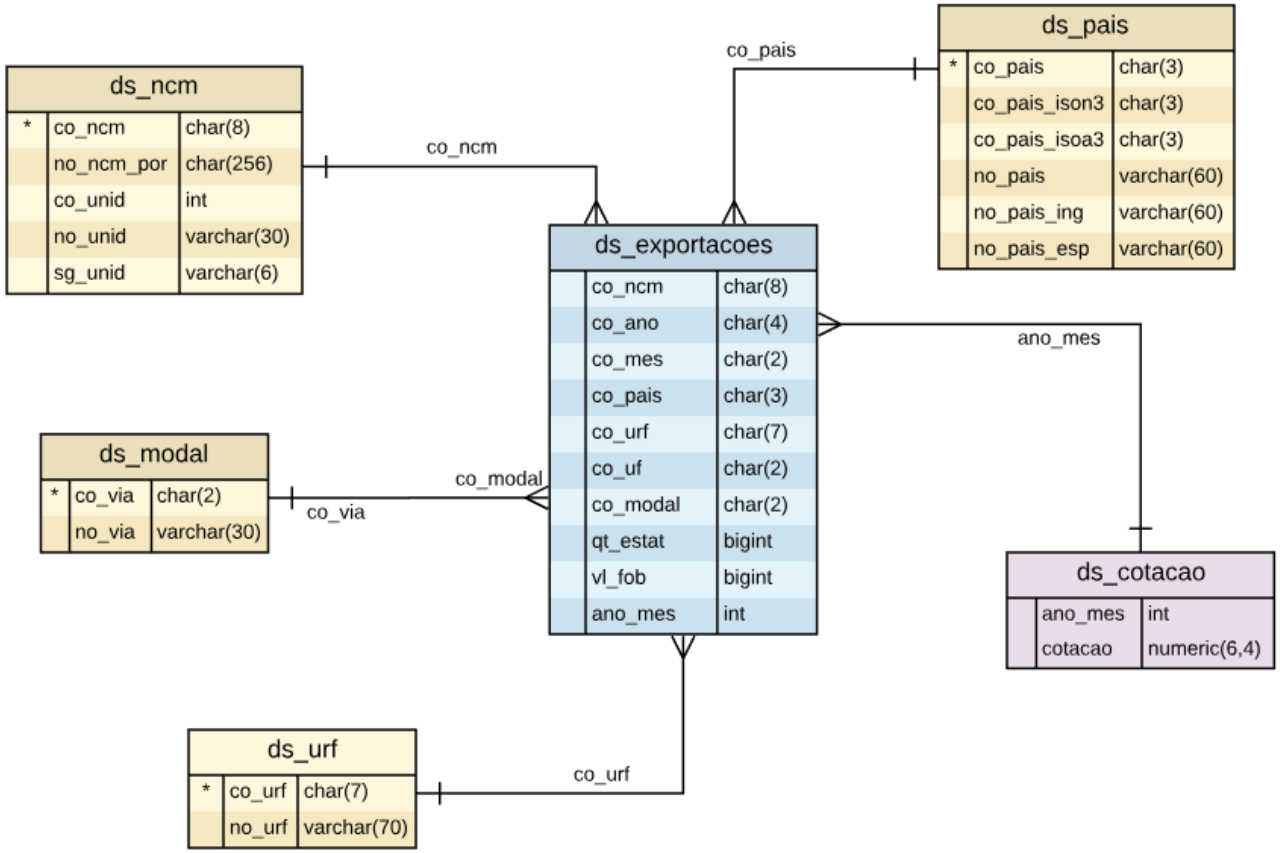
6 – Welcome to Python

Namorando os Dados (Queries SQL)



Desafio – Modelo de Dados





Namorando os Dados (Queries SQL)

Mãos à obra Pessoal!!!

R - Parte II: Qualidade dos Dados e Variáveis Relevantes

1 – Agenda

2 – Welcome to R – Parte I

3 – Homework - ETL

4 – Namorando Dados (SQL)

5 – R – Parte II

6 – Welcome to Python

O Trabalho do Cientista de Dados > Desafio Curso

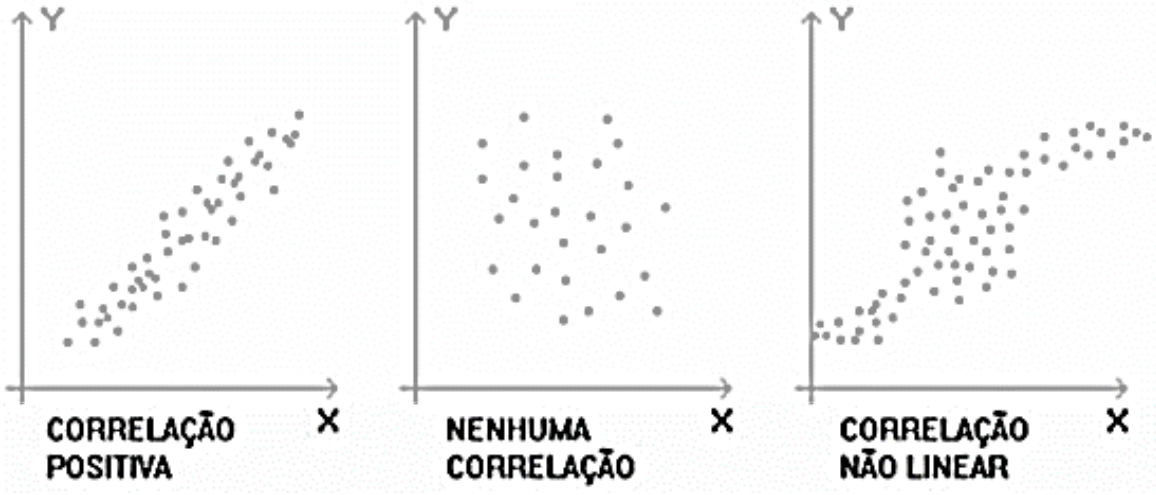
1. Definição do problema e levantamento de perguntas a serem respondidas ☑
2. Planejamento do processo de Data Science ☑
3. Coleta de dados☑
4. Processamento e limpeza dos dados
5. Armazenamento dos dados ☑
6. Análise de dados
7. Construção e validação de algoritmos e modelos
8. Data Visualization
9. Disseminação da informação
10. Colocar modelo em produção

Analisando a Qualidade dos Dados

- Objetivo nesta etapa do estudo é verificar a qualidade dos dados para entender quais tem potencial de fazer parte do estudo
- Foco maior em verificar se existem dados faltantes ou nulos que podem interferir no estudo
- Também aqui começa o entendimento de como cada variável ajuda a explicar o evento em estudo
- Aqui começam as **descobertas** do Cientista de Dados

Variáveis Relevantes

- Objetivo nesta etapa do estudo é verificar a como as variáveis se relacionam entre si
  - **Foco maior aqui é entender a correlação entre as variáveis**
- O modelo ou a metodologia que será utilizada para responder as perguntas do estudo dependem dos achados desta etapa



Welcome to Python: Básico, Numpy, Pandas e Banco

1 – Agenda

2 – Welcome to R – Parte I

3 – Homework - ETL

4 – Namorando Dados (SQL)

5 – R – Parte II

6 – Welcome to Python

Python

Mar 2019	Mar 2018	Change	Programming Language	Ratings	Change
1	1		Java	14.880%	-0.06%
2	2		C	13.305%	+0.55%
3	4	↑	Python	8.262%	+2.39%
4	3	↓	C++	8.126%	+1.67%
5	6	↑	Visual Basic .NET	6.429%	+2.34%
6	5	↓	C#	3.267%	-1.80%
7	8	↑	JavaScript	2.426%	-1.49%
8	7	↓	PHP	2.420%	-1.59%
9	10	↑	SQL	1.926%	-0.76%
10	14	⇈	Objective-C	1.681%	-0.09%
11	18	⇈	MATLAB	1.469%	+0.06%
12	16	⇈	Assembly language	1.413%	-0.29%
13	11	↓	Perl	1.302%	-0.93%
14	20	⇈	R	1.278%	+0.15%
15	9	⇊	Ruby	1.202%	-1.54%
16	60	⇈	Groovy	1.178%	+1.04%
17	12	⇊	Swift	1.158%	-0.99%
18	17	↓	Go	1.016%	-0.43%
19	13	⇊	Delphi/Object Pascal	1.012%	-0.78%
20	15	⇊	Visual Basic	0.954%	-0.79%

Fonte: https://www.tiobe.com/tiobe-index/

Python – Me Dê Motivos

Linguagem em forte ascenção (3ª linguagem mais amada Stack Overflow)

Curva de Aprendizado Baixa

Free (Licença GLP)

Estável (1ª versão 1991)

Multiplataforma (Windows, Linux, MacOS e etc.)

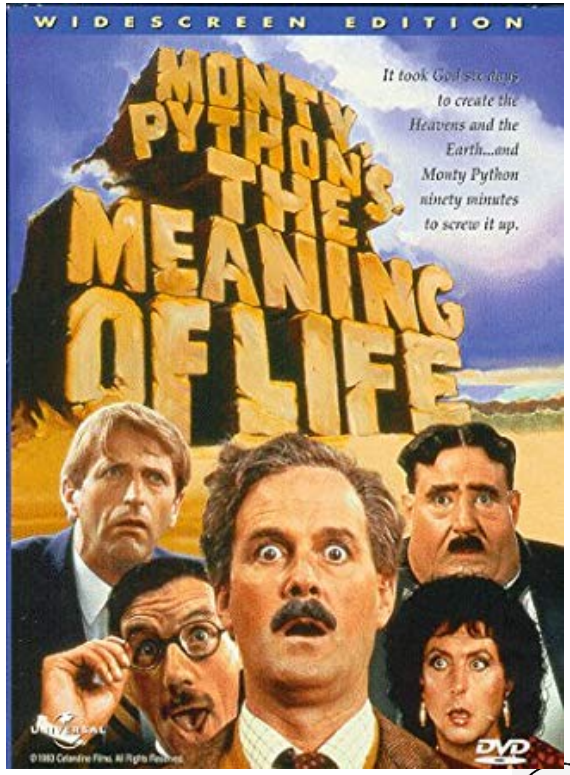
Comunidade

Data Science → Ótimos pacotes

Python – História

Pai do Python →

Guido van Rossum



A inspiração do nome →

Python – História

Versão 2 (2.7) **x Versão 3 (3.5)**

3/4 Paradigmas de Programação:

- **Programação Imperativa** → Ações/Comandos de um programa
- **Programação Orientada o Objeto** → Abstração, Encapsulamento, Herança e Polimorfismo
- **Programação Funcional** → Soluções como problemas de funções

Interpretada

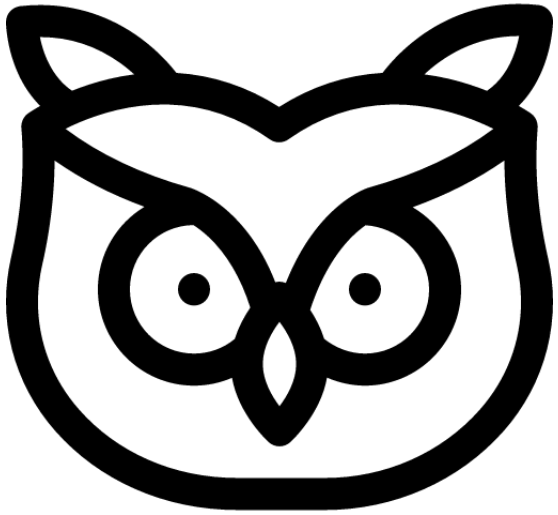
Python – Hands-on

Python – Versão 2 x Versão 3

Python 2.X	Python 3.X
There's ASCII str type and unicodetype, but no separate type to handle bytes of data	All strings (str) are Unicode strings; two byte classes are introduced: bytes and bytearray
Two types of integers: C-based integers (int) and Python long integer (long)	All integers are long but referred to by the int type
Return type of division is int if operands are integers: 5 / 4 gives 1; 4 / 2 gives 2	Return type of division is float even if operands or result are integers: 5 / 4 gives 1.25; 4 / 2 gives 2.0
round(16.5) returns a float of value 16.0	round(16.5) returns an int of value 16
Unorderable types can be compared	Comparison of unorderable types raises a TypeError
print is a statement: print "Hello World!"	print() is a built-in function: print("Hello World!")
range() returns a list of numbers while xrange() returns an object for lazy evaluation	range() returns an object for lazy evaluation similar to Python 2 xrange(); and range()method __contains__ speeds up lookups
Functions/methods map(), filter(), zip(), dict.items(), dict.keys(), dict.values() return lists	These function/methods return objects for lazy evaluation
raw_input() returns input as strand input() evaluates the input as a Python expression	input() will return a string similar to Python 2 raw_input()
Raising exceptions: raise IOError("file error") or raise IOError, "file error"	Raising exceptions: raise IOError("file error")
Handling exceptions: except NameError, err: or except (TypeError, NameError), err:	Handling exceptions: except NameError as err or except (TypeError, NameError) as err
On generators, a method or function call: g.next() or next(g)	On generators, only a function call: next(g)
Loop variables in a comprehension leak to global namespace	Loop variables are limited in scope to the comprehension

Fonte: https://devopedia.org/python-2-vs-3

Quero Saber Mais...

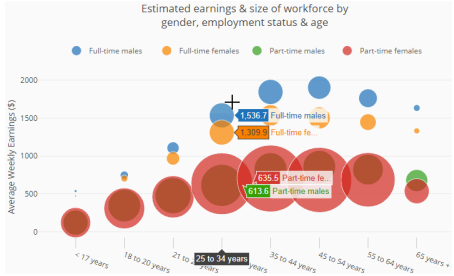


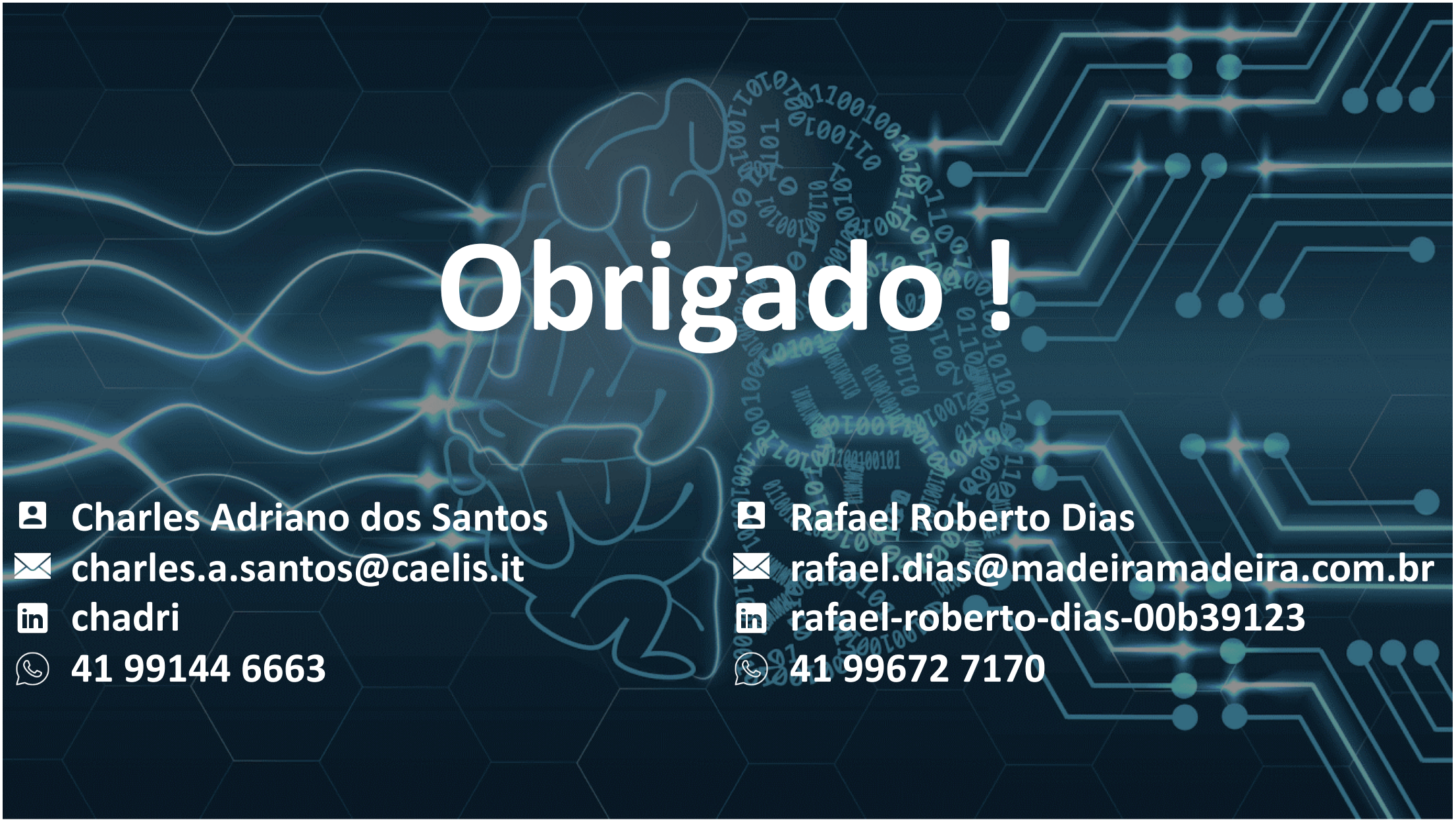
Os 35 Melhores Cursos de Python gratuitos disponíveis pra você

Towards Data Science

Sharing concepts, ideas, and codes

Data Manipulation for Machine Learning with Pandas





Charles Adriano dos Santos
charles.a.santos@caelis.it
chadri
41 99144 6663

Rafael Roberto Dias
rafael.dias@madeiramadeira.com.br
rafael-roberto-dias-00b39123
41 99672 7170