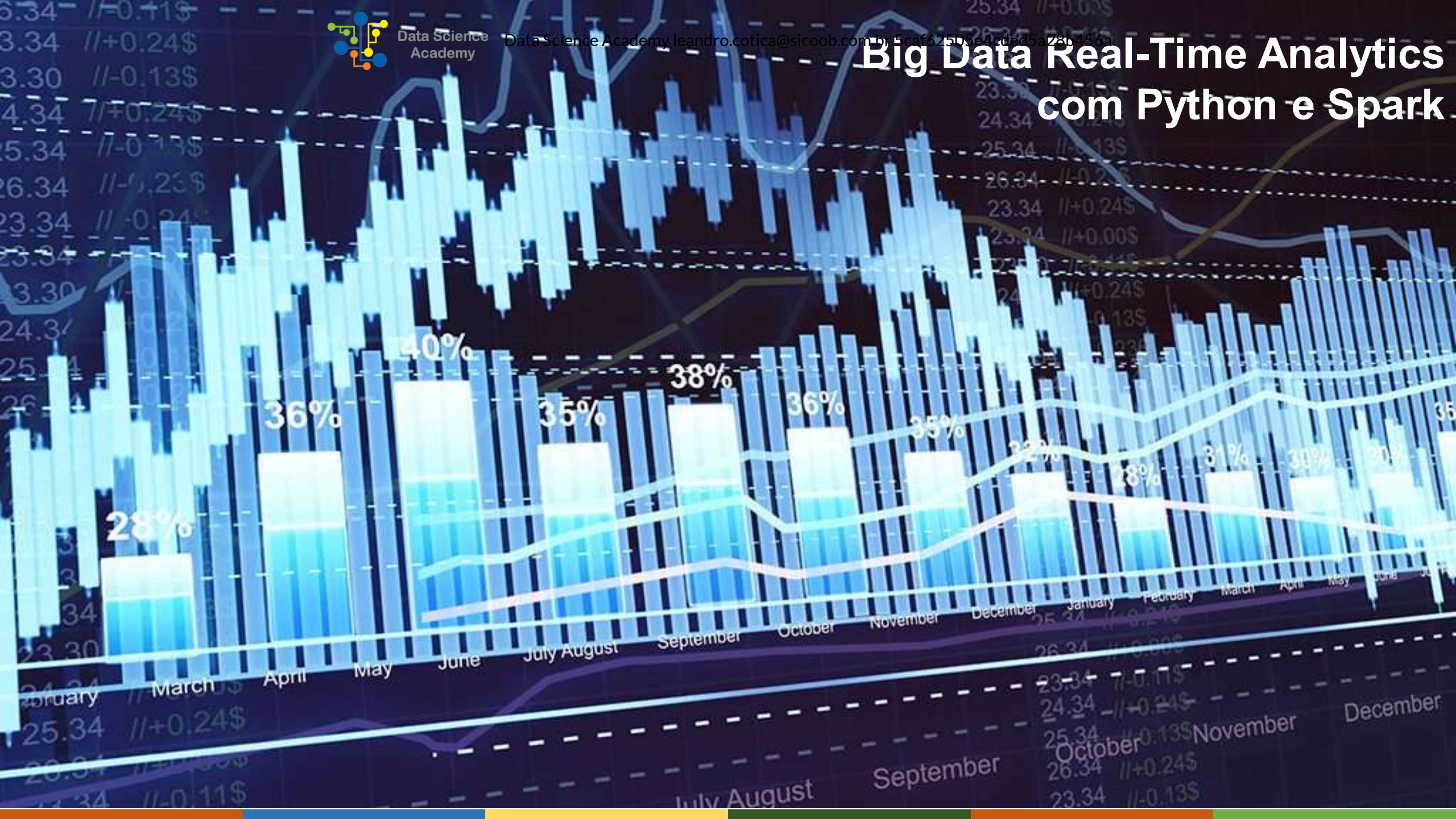


Big Data Real-Time Analytics com Python e Spark





Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Big Data Real-Time Analytics com Python e Spark

Seja muito bem-vindo(a)!



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python

Manipulação de Dados com Python





Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python

Python Fundamentos para Análise de Dados

Inscreva-se gratuitamente!

54 horas





Introdução a Análise de Dados com Python

O que estudaremos neste capítulo?

- Módulos Python para Análise de Dados
- Estruturas de Dados
- NumPy
- Pandas
- Pré-Processamento



Introdução a Análise de Dados com Python

Preciso realmente aprender uma linguagem de programação para trabalhar como Cientista de Dados?



Introdução a Análise de Dados com Python

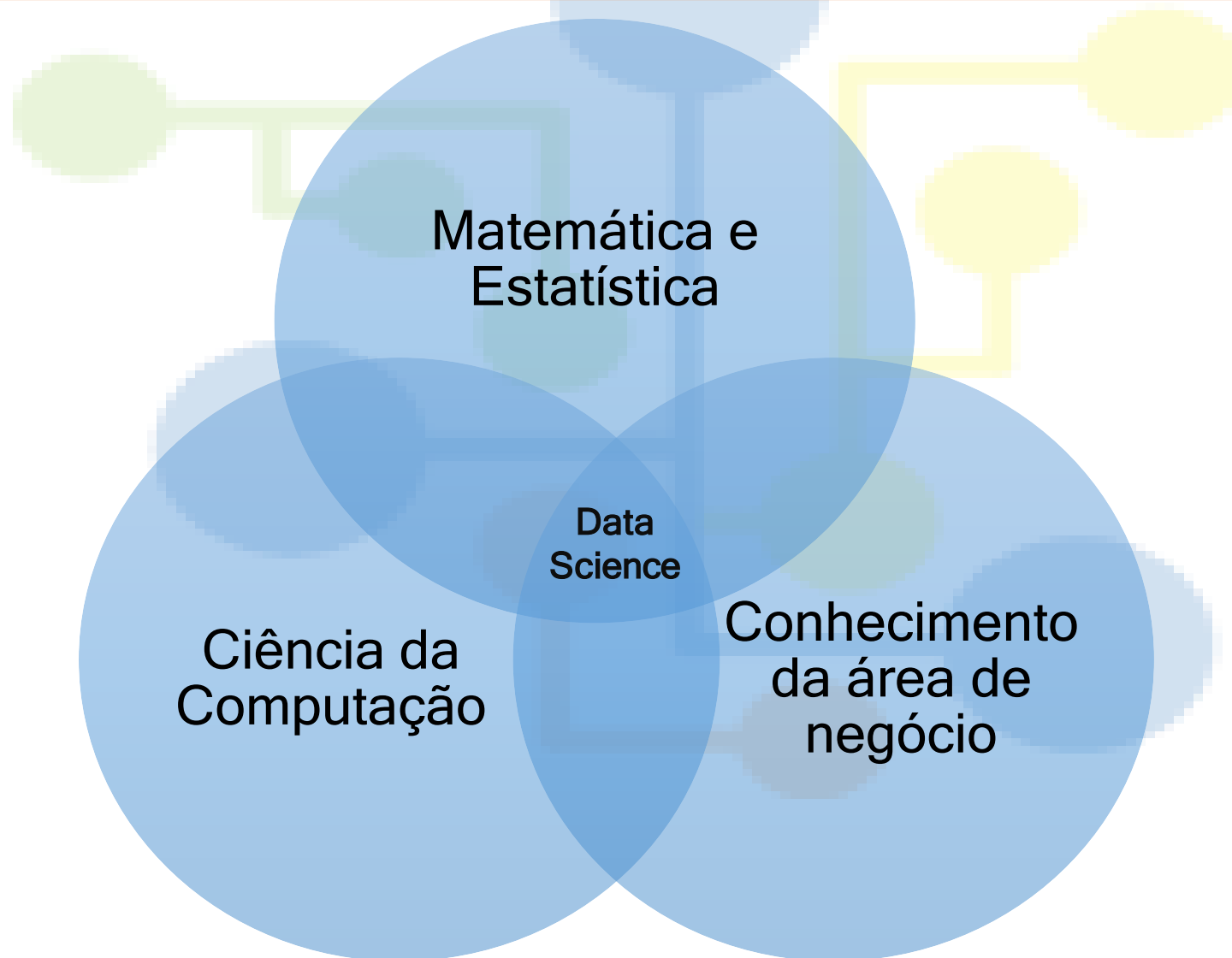
Você precisa aprender a usar uma ferramenta analítica!



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python





Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python





Introdução a Análise de Dados com Python

Eu tenho que escolher uma das duas?





Introdução a Análise de Dados com Python

Precisamos realmente escolher?

Por que não usar as duas?



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python





Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python



Mas é muito trabalho
e muito material!!!



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python



Sim, é isso mesmo!



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python



Quem disse que
seria fácil?



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python



Quer ou não se tornar
um Cientista de Dados?



Introdução a Análise de Dados com Python

Mas eu poderia utilizar soluções
analíticas proprietárias?



Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Introdução a Análise de Dados com Python

Formação Análise Estatística Para Cientistas de Dados

Sua Base para Carreira de Cientista de Dados Começa Aqui.

Matricule-se já!



Introdução a Análise de Dados com Python

Mas então por que estou aprendendo linguagens de programação como ferramenta analítica?



Introdução a Análise de Dados com Python

- R e Python são linguagens gratuitas e podem ser usadas livremente, reduzindo custos com licenciamento de software, por exemplo.
- Embora R e Python requeiram mais trabalho manual para criar uma solução de análise, elas oferecem muito mais flexibilidade.
- Todo o conhecimento adquirido com R e Python, pode ser facilmente aplicado em outras soluções analíticas.
- Soluções Microsoft, Oracle e IBM, dão suporte a R e Python, como forma de estender as funcionalidades de suas soluções.
- As duas linguagens possuem uma grande e ativa comunidade e muita documentação disponível.



Introdução a Análise de Dados com Python

A pergunta que deveria ser feita é:

Há alguma razão para não aprender R ou Python para
Data Science?

NÃO!





Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Big Data Real-Time Analytics com Python e Spark

Por Que Linguagem Python Para Análise de Dados?





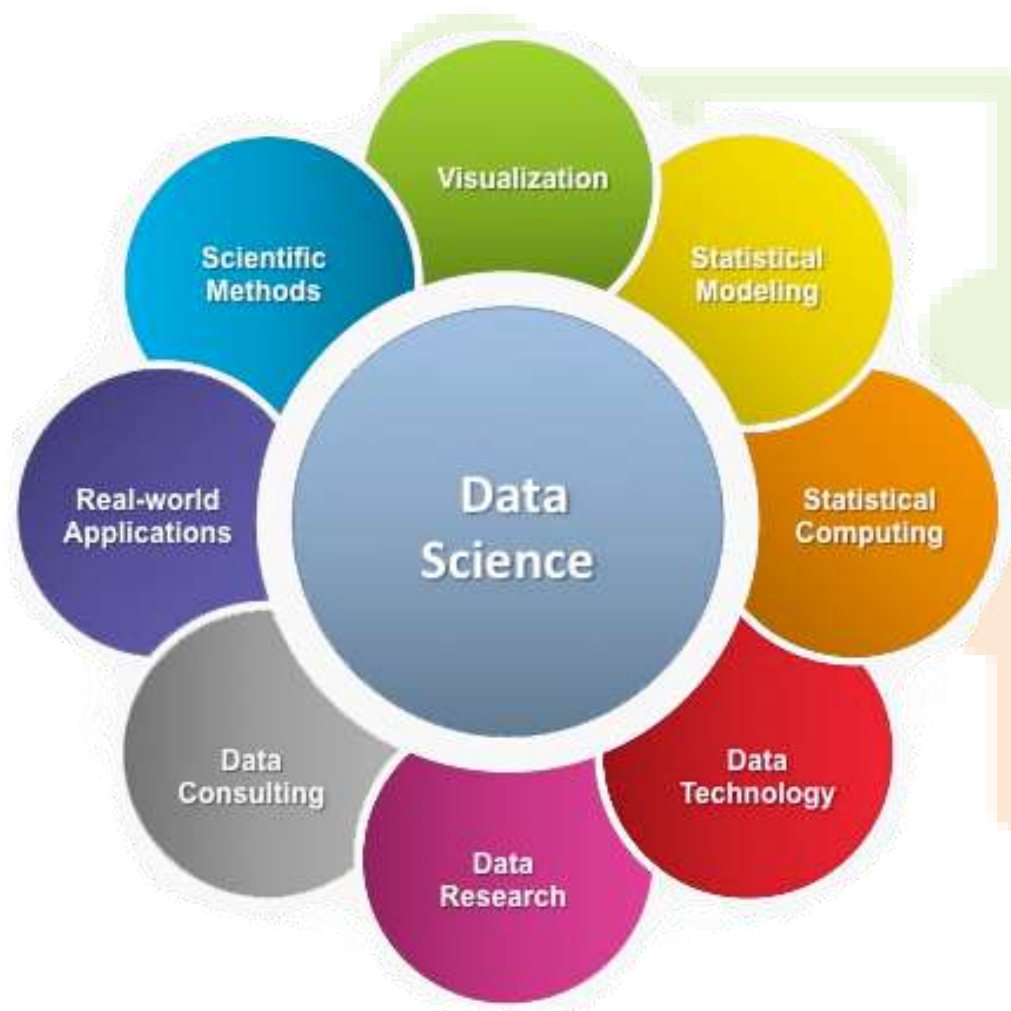
Por Que Linguagem Python Para Análise de Dados?

Cada vez mais dados para analisar
(Big Data)

Cada vez menos tempo
(Precisamos de Soluções de Análise em Tempo Real)



Por Que Linguagem Python Para Análise de Dados?



A boa notícia é que a linguagem Python pode nos ajudar em todas as etapas do processo de análise de dados!



Por Que Linguagem Python Para Análise de Dados?



- Fácil de aprender
- Muito popular
- Uso geral
- Linguagem interpretada
- Comunidade



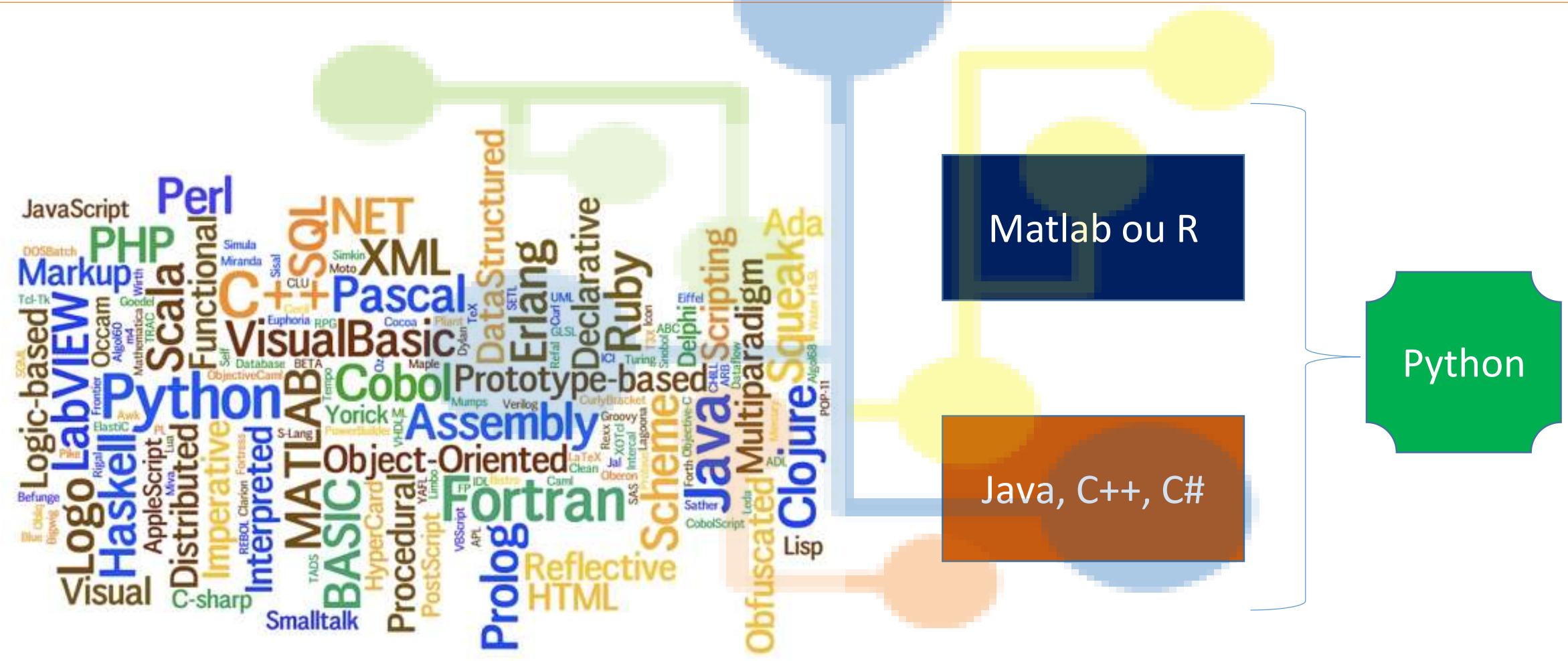
Por Que Linguagem Python Para Análise de Dados?



- R
- SAS
- Matlab
- Stata
- C e C++
- Fortran
- Cython (<http://cython.org>)



Por Que Linguagem Python Para Análise de Dados?





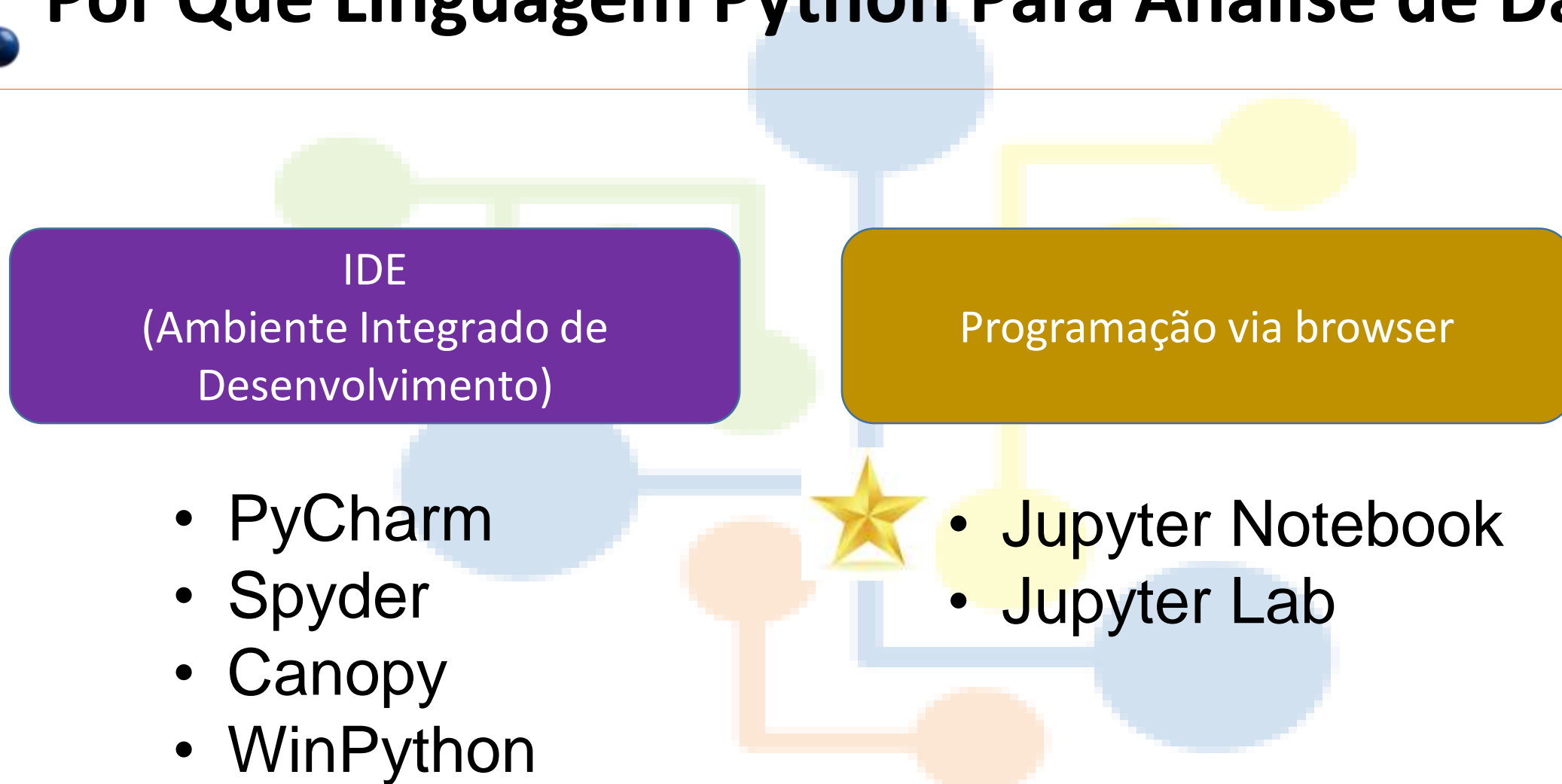
Por Que Linguagem Python Para Análise de Dados?

Quando não usar
Python?





Por Que Linguagem Python Para Análise de Dados?





Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Big Data Real-Time Analytics com Python e Spark

Principais Módulos Python Para Análise de Dados





Principais Módulos Python Para Análise de Dados



Python, quanto linguagem de programação, possui seus comandos e funções built-in.

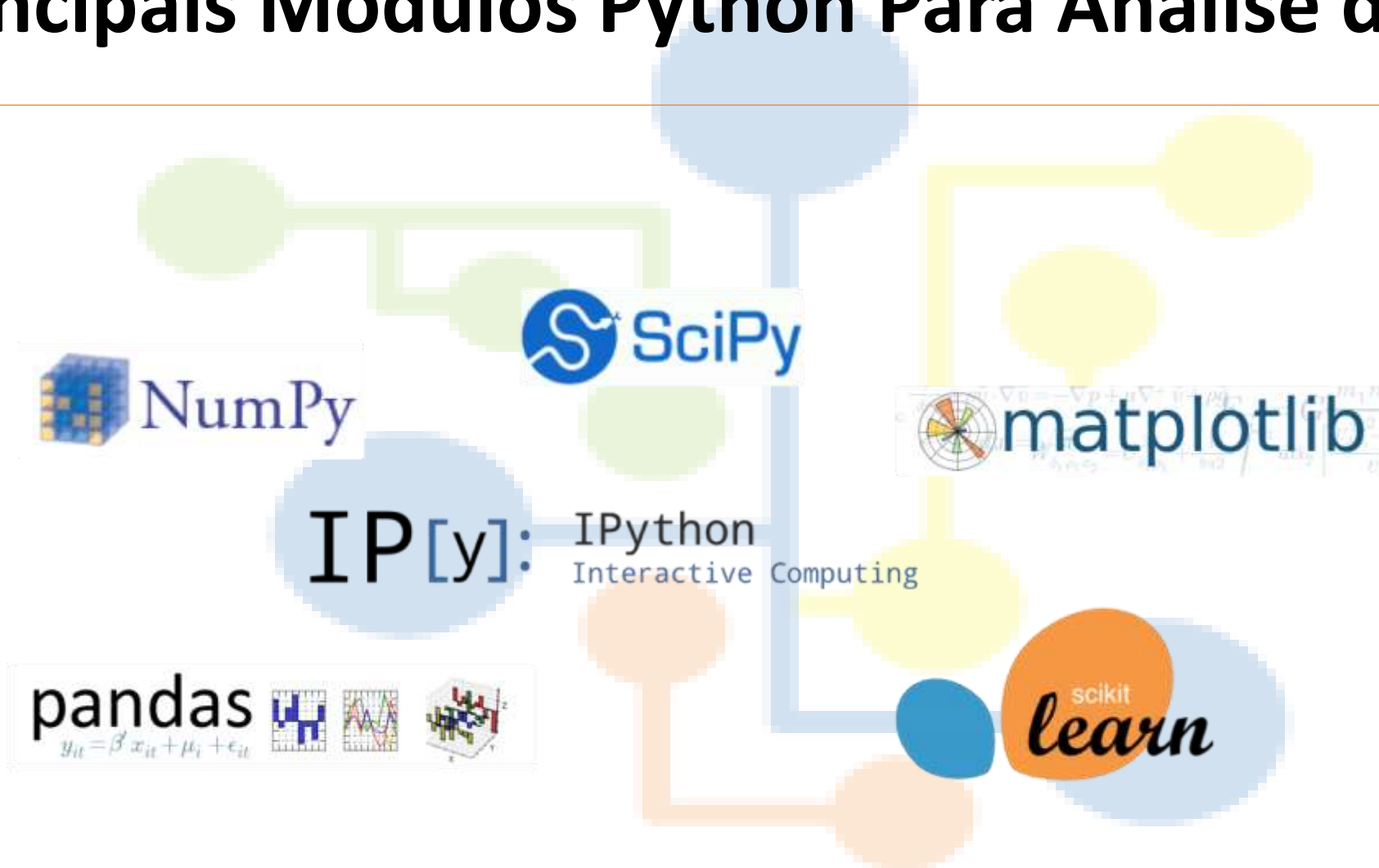
É o que chamamos de linguagem pura.



Data Science
Academy

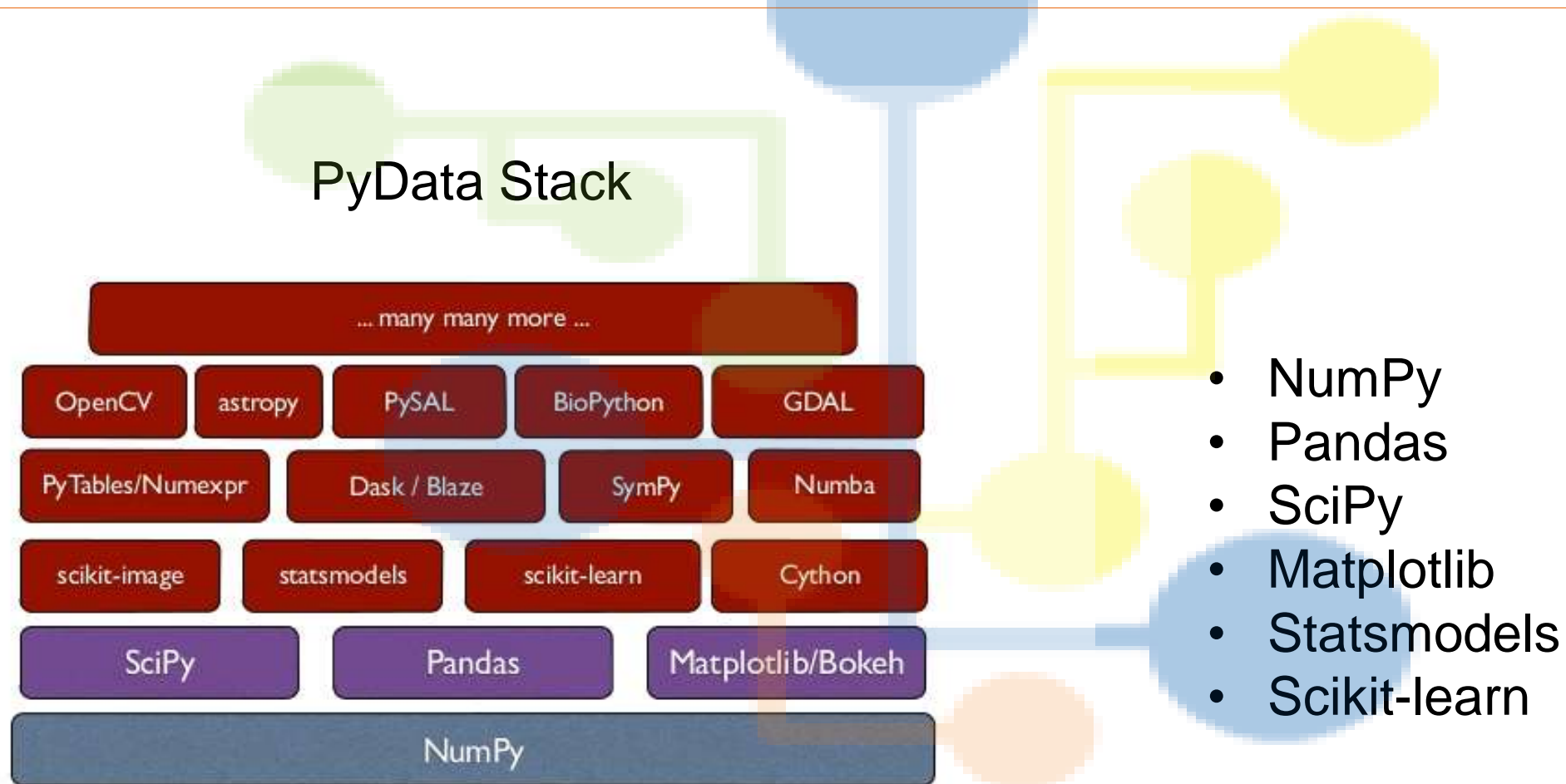
Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

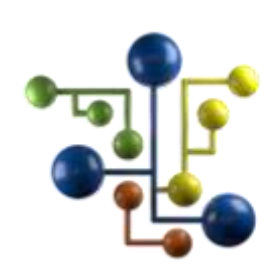
Principais Módulos Python Para Análise de Dados





Principais Módulos Python Para Análise de Dados

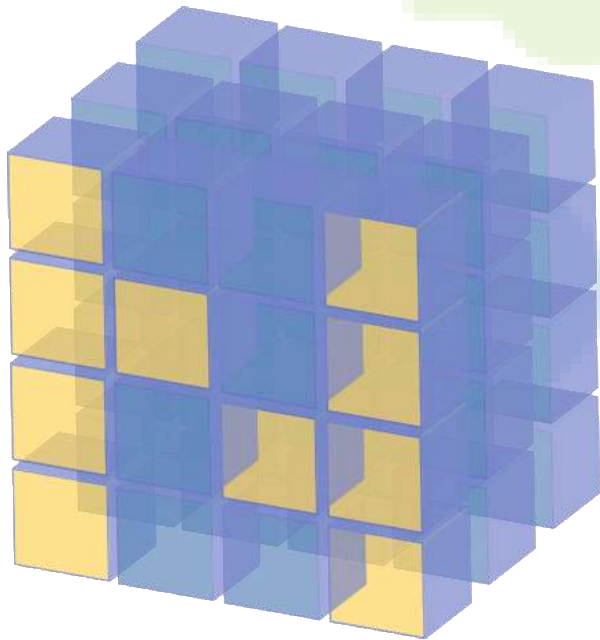




Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Principais Módulos Python Para Análise de Dados



NumPy



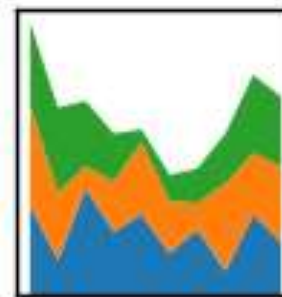
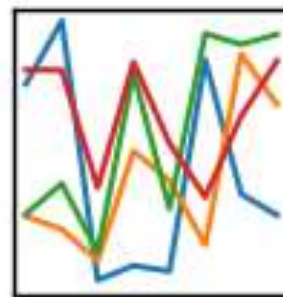
Data Science
Academy

Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Principais Módulos Python Para Análise de Dados

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

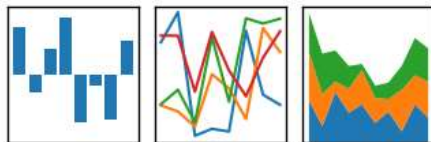




Principais Módulos Python Para Análise de Dados

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



- Pode processar dados em diferentes formatos, como dados de séries temporais, matrizes, dados estruturados ou não estruturados.
- Facilita muito o trabalho de carga e importação de dados em arquivos csv ou de bancos de dados.
- Fornece funções para as mais variadas etapas de pré-processamento, como: subsetting, slicing, filtros, merge, agrupamentos, ordenação, reshape.
- Permite facilmente tratar dados missing.
- Pode ser usado para converter dados, bem como para aplicar modelagem estatística.
- É totalmente integrado com outros pacotes Python, como SciPy, NumPy e Scikit-Learn.



Data Science
Academy

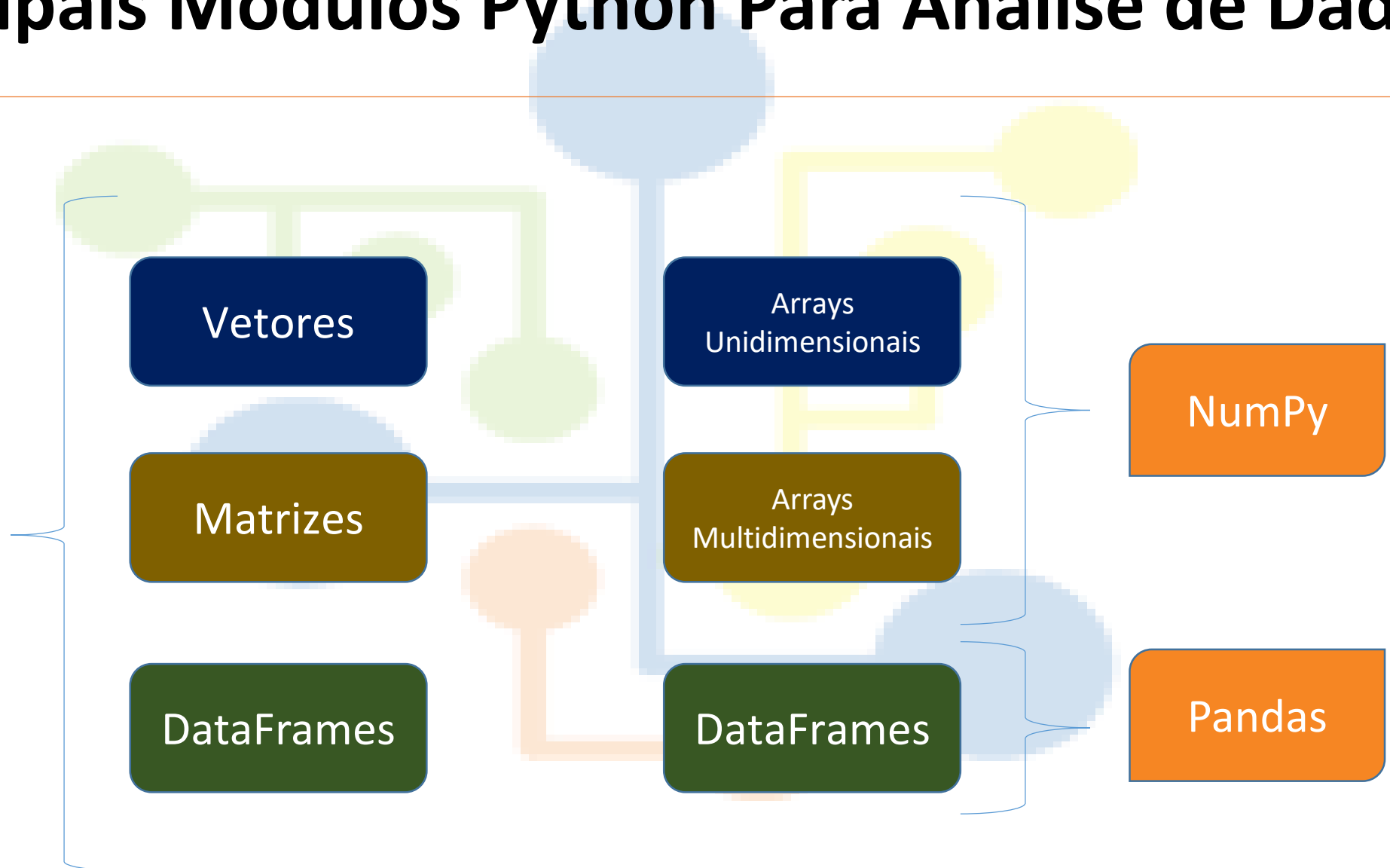
Data Science Academy leandro.cotica@sicoob.com.br 5caf62505e4cded5a28b456a

Principais Módulos Python Para Análise de Dados





Principais Módulos Python Para Análise de Dados





Data Science
Academy

Data Science Academy leandro.cotica@data-science-academy.com.br 5caf62505e4cded5a28b456a



Tenha uma Excelente Jornada de Aprendizagem.

Muito Obrigado por Participar!

Equipe Data Science Academy