

Analytics e Inteligência Artificial

Aula 18

**Projeto Disciplina Parte 4 -
Clusterização**





BUSINESS SCHOOL

Graduação, pós-graduação, MBA, Pós-MBA, Mestrado Profissional, Curso In Company e EAD



CONSULTING

Consultoria personalizada que oferece soluções baseadas em seu problema de negócio



RESEARCH

Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino



Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais **escolas de negócio do mundo**, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 **projetos de consultorias** em organizações públicas e privadas.



Único curso de graduação em administração a receber as notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única Business School brasileira a figurar no ranking LATAM



Signatária do Pacto Global da ONU



Membro fundador da ANAMBA - Associação Nacional MBAs



Credenciada pela AMBA - Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB - Association to Advance Collegiate Schools of Business



Filiada a EFMD - European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação

O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de **Big Data, Analytics** e **Inteligência Artificial**.



Profª Drª Alessandra
Montini

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil
Os diretores foram professores de grandes especialistas do mercado
+10 anos de atuação
+1000 alunos formados

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria
- Larga experiência de mercado na resolução de *cases*
- Participação em Congressos Nacionais e Internacionais
- Professor assistente que acompanha o aluno durante todo o curso

Estrutura

- 100% das aulas realizadas em laboratórios
- Computadores para uso individual durante as aulas
- 5 laboratórios de alta qualidade (investimento +R\$2MM)
- 2 Unidades próximas a estação de metrô (com estacionamento)



Corpo Diretivo

COORDENADORES DO LABDATA | ATUAÇÃO ACADÊMICA E PROFISSIONAL

4



Profª Dra.
**Alessandra
Montini**

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Têm muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em estatística aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Membro do Conselho Curador da FIA, Coordenadora de Grupos de Pesquisa no CNPQ, Parecerista da FAPESP e Colunista de grandes Portais de Tecnologia.

 [linkedin.com/in/alessandramontini/](https://www.linkedin.com/in/alessandramontini/)



Prof. Dr.
Adolpho Walter Canton

Diretor do LABDATA-FIA. Consultor em Projetos de *Analytics*, *Big Data* e Inteligência Artificial. Professor FEA – USP. PhD em Estatística Aplicada pela *University of North Carolina at Chapel Hill*, Estados Unidos.

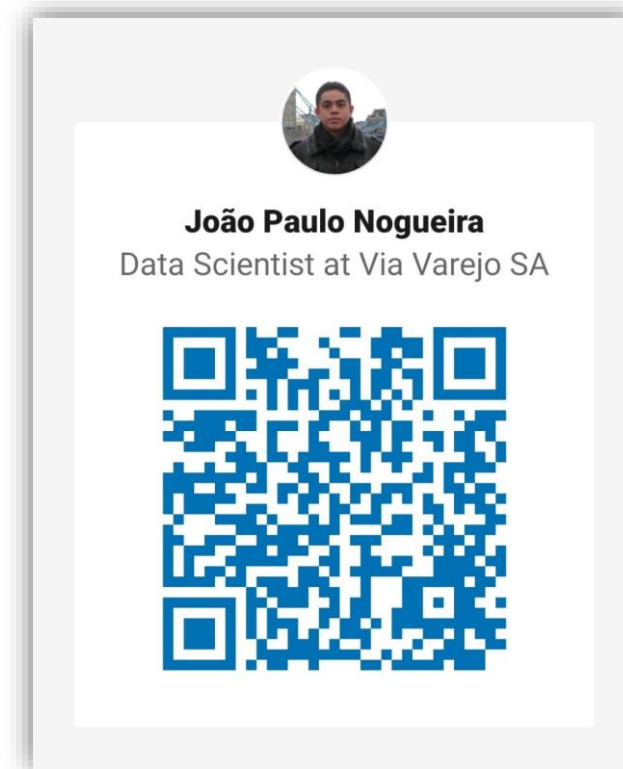


Currículo – ProfJoão Nogueira

FORMAÇÃO ACADÊMICA | EXPERIÊNCIA PROFISSIONAL

5

- **(2019-Presente)** – Professor nos cursos de Extensão, Pós e MBA em Big Data e Data Mining na Fundação Instituto de Administração (FIA) – www.fia.com.br
- **(2018-Presente)** – Cientista de Dados na Via Varejo – <https://viavarejo.com.br>
- **(2016-Presente)** – Doutorando em Física Computacional e Estatística pelo Departamento de Física na Universidade Federal do Ceará – <https://fisica.ufc.br>
- **(2014-2016)** – Mestre em Física da Matéria Condensada pelo Departamento de Física na Universidade Federal do Ceará - <https://fisica.ufc.br>
- **(2012-2013)** – Estudante Intercambista na Universidade de Coimbra – Portugal – <https://www.uc.pt>
- **(2010-2014)** – Bacharel em Física pela Universidade Federal do Ceará – <http://www.ufc.br>
- **Contatos:**
 - E-mail: joaonogueira@fisica.ufc.br



Conteúdo Programático da Disciplina Projeto de Inteligência Artificial



Data	Horário	Tema
09/03/2021	19:00	Aula 1 - Introdução ao Ambiente de Desenvolvimento
11/03/2021	19:00	Aula 2 - Revisão de Python
16/03/2021	19:00	Aula 3 - Manipulação de Dados
18/03/2021	19:00	Aula 4 - Análise Exploratória de Dados
23/03/2021	19:00	Aula 5 - Projeto da disciplina - Parte 1 - Análise Exploratória de Dados
25/03/2021	19:00	Aula 6 - Introdução, Motivação e Framework de Machine Learning
06/04/2021	19:00	Aula 7 - Analytical Base Table
08/04/2021	19:00	Aula 8 - Aprendizagem Supervisionada - Classificação
13/04/2021	19:00	Aula 9 - Aprendizagem Supervisionada - Classificação
15/04/2021	19:00	Aula 10 - Aprendizagem Supervisionada - Classificação
20/04/2021	19:00	Aula 11 - Projeto da disciplina - Parte 2 - Machine Learning - Classificação
22/04/2021	19:00	Aula 12 - Projeto da disciplina - Parte 2 - Machine Learning - Classificação
27/04/2021	19:00	Aula 13 - Aprendizagem Supervisionada - Regressão
29/04/2021	19:00	Aula 14 - Aprendizagem Supervisionada - Regressão
04/05/2021	19:00	Aula 15 - Projeto da disciplina - Parte 3 - Machine Learning - Regressão
06/05/2021	19:00	Aula 16 - Aprendizagem Não-Supervisionada
11/05/2021	19:00	Aula 17 - Aprendizagem Não-Supervisionada
13/05/2021	19:00	Aula 18 - Projeto da disciplina - Parte 4 - Machine Learning - Clusterização
18/05/2021	19:00	Aula 19 - AutoML
20/05/2021	19:00	Aula 20 - Demonstração de Deploy de Machine Learning

Conteúdo da Aula

- 1. Introdução
- 2. Aprendizagem Não Supervisionada
- 3. Exercícios
- Entregas e Prazos



Material das aulas

- Iremos utilizar o Google Colab para desenvolver os códigos durante as aulas.
- Acesse <https://bit.ly/tutorial-colab-projeto> para realizar o tutorial de utilização do Google Colab.



1. Introdução



1. Introdução

BASES DE DADOS

- Essa etapa do projeto da disciplina iremos aplicar os conceitos sobre Aprendizagem Não Supervisionada para segmentar os clientes da base do iFood.



2. Aprendizagem Não Supervisionada



2. Aprendizagem Supervisionada - Regressão

BASES DE DADOS



ORDERS

- **ORDERS** - Informações sobre os pedidos realizados.

Icon made by [Srip](https://www.flaticon.com) from www.flaticon.com
@2020 LABDATA FIA. Copyright all rights reserved.



2. Aprendizagem Supervisionada - Regressão

PROBLEMA

- O departamento de Marketing quer **entender o comportamento dos clientes** do iFood, para isso fez uma solicitação para o time de Ciência de Dados para criar modelos de Machine Learning para segmentar os clientes.



2. Aprendizagem Supervisionada - Regressão

PROBLEMA

- O objetivo desses modelos é entender o comportamento dos clientes realizando duas segmentações:
 - Segmentação por tipo de comida
 - Segmentação RFV (Para esse exercício será utilizado somente a frequência e o valor).



2. Aprendizagem Supervisionada - Regressão

O QUE IREMOS APRENDER?

- **Criar a ABT** de segmentação por tipo de comida
- Executar o método do cotovelo para encontrar o número ideal de clusters
- Rodar o algoritmo KMeans com o **número ideal de clusters** encontrado
- Analisar os resultados através de um mapa de calor
- Caso necessário, utilizar uma **Surrougate Tree** para encontrar as variáveis mais importantes
- Criar a ABT de segmentação por RFV (usando apenas a Frequência e o Valor)
- Repetir os passos anterior para a **segmentação por RFV**
- Combinar o resultado das duas segmentações em uma só tabela contendo 3 colunas:
`customer_id`, `cluster_categorias`, `cluster_fv`
- Escrever os principais **insights** retirados da análise



3. Exercícios



4. Exercícios

BASES DE DADOS

- Utilize o notebook abaixo para realizar os exercícios.



Abra o arquivo "**aula18-projeto-parte4-aprendizagem-nao-supervisionada.ipynb**"



4. Entregas e Prazos



5. Entregas e Prazos

BASES DE DADOS

- A entrega deverá ser submetida no portal do aluno, com o upload de apenas um notebook no formato `.ipynb`, com o código executado
- Recomendo também enviar uma cópia do notebook para o email do monitor: denner.lopes86@faculdefia.edu.br
- No campo de **Assunto** deve ser exatamente:
 - Entrega – TURMA 2 BIG DATA EAD – Projeto Parte 2 - Classificação **(obrigatório)**
- No **Corpo** do e-mail deve ser adicionado o seu **nome completo**.
- **Prazo: 07/02 até às 23:59**
- O projeto final terá um total de **100 pontos**.
- A entrega dos projetos acontecerá em 4 partes e cada parte vale **25 pontos**

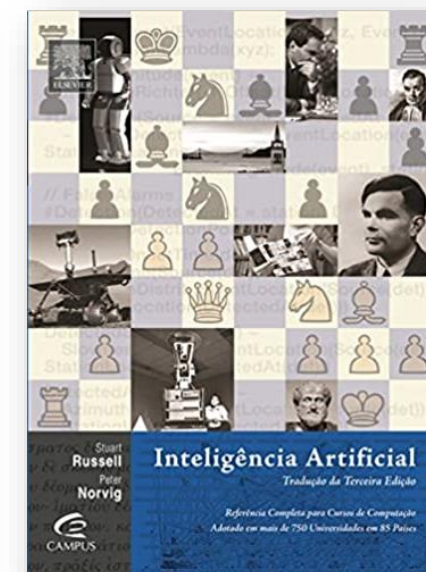
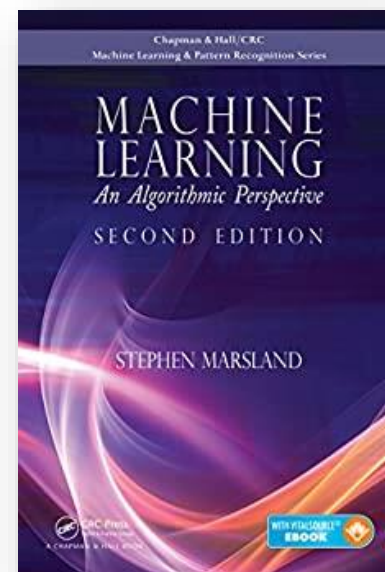
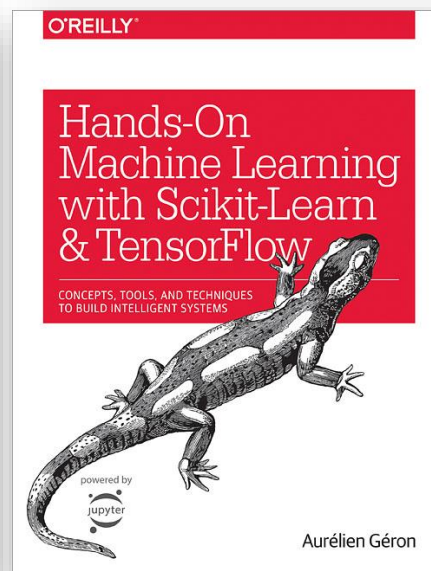
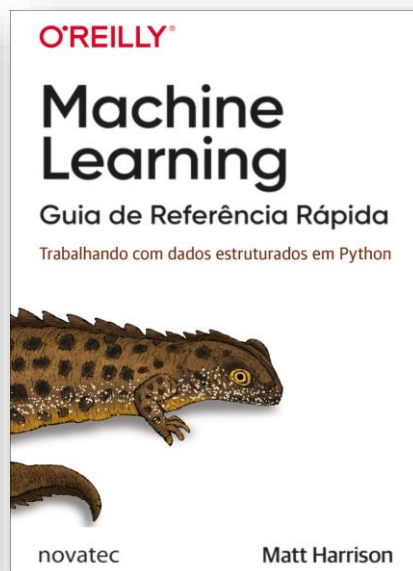
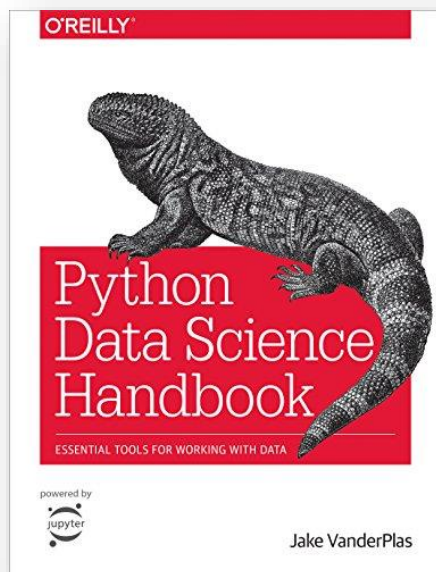


Referências Bibliográficas



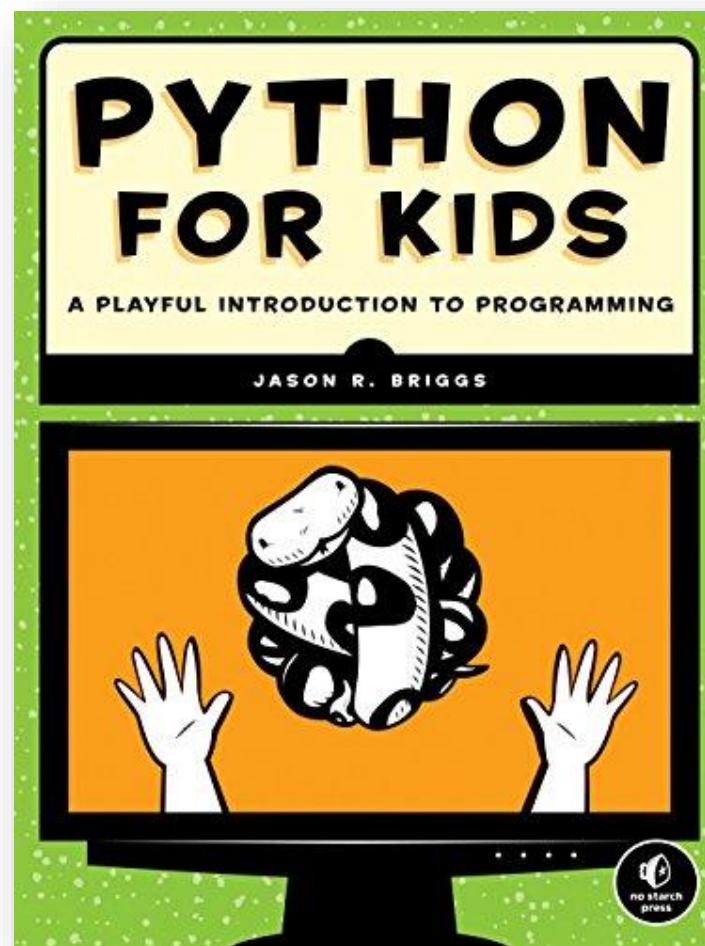
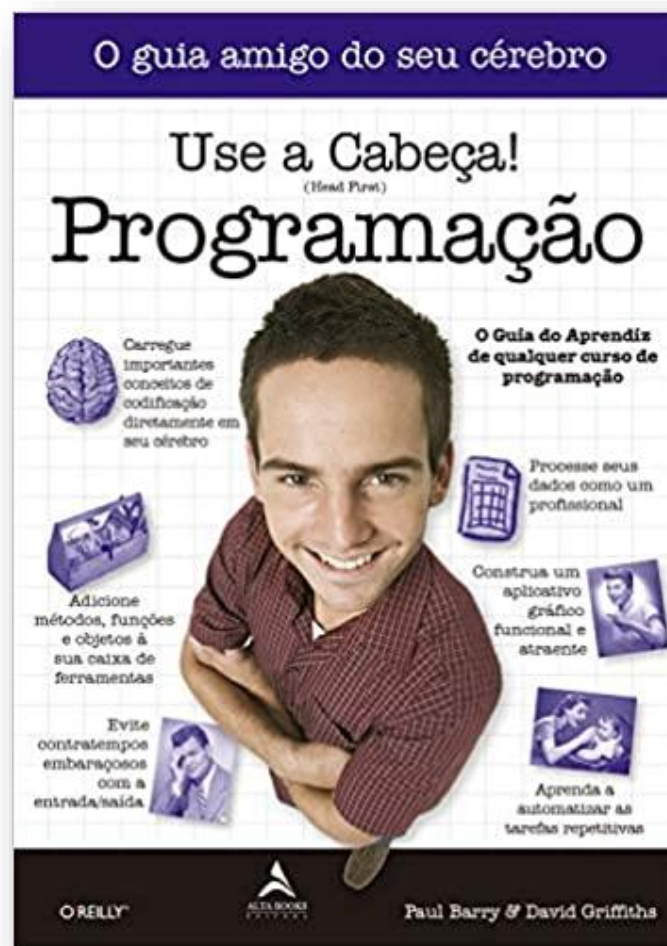
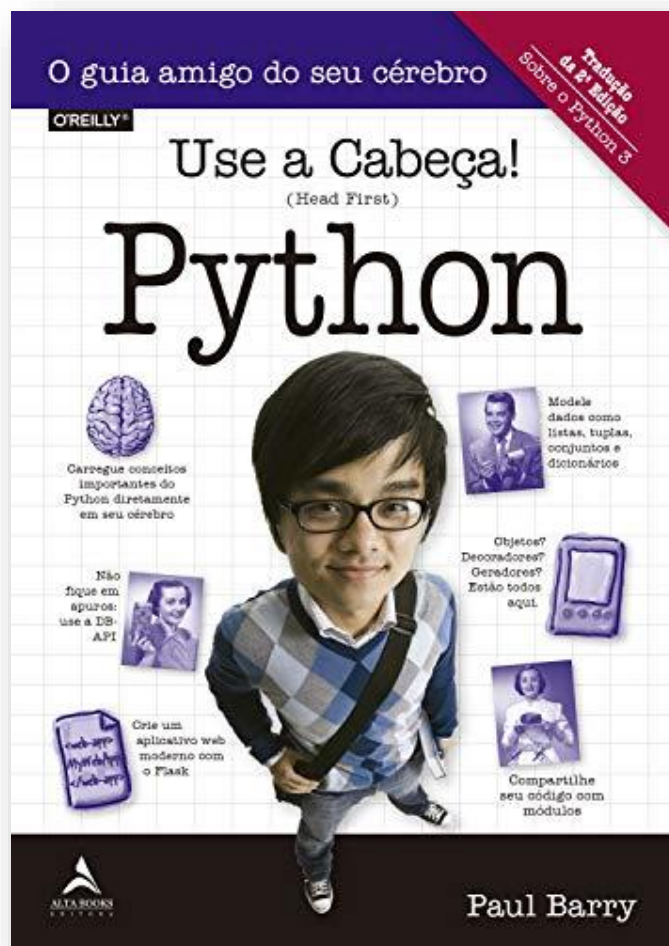
Referências Bibliográficas

LIVROS



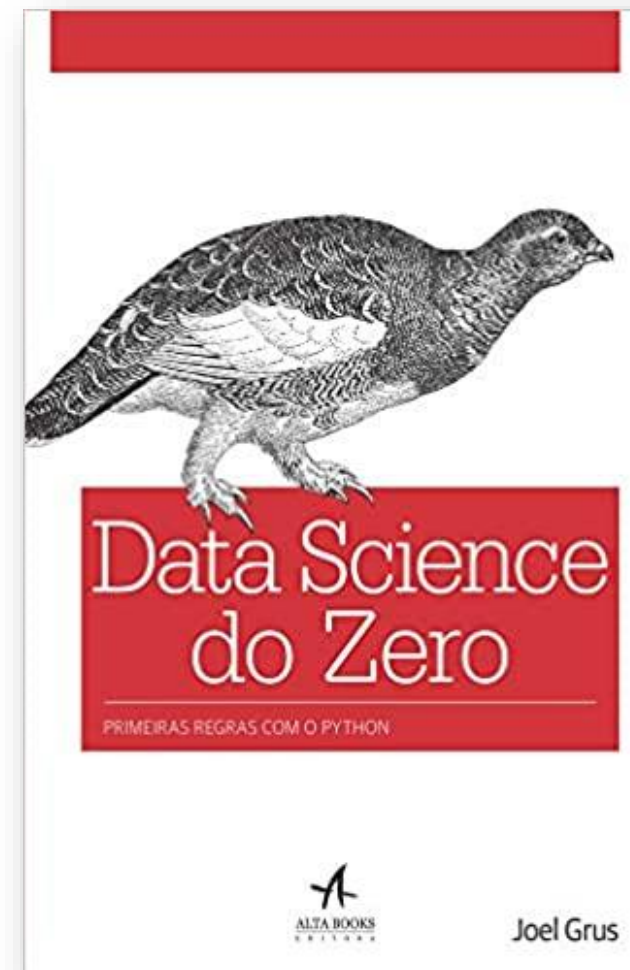
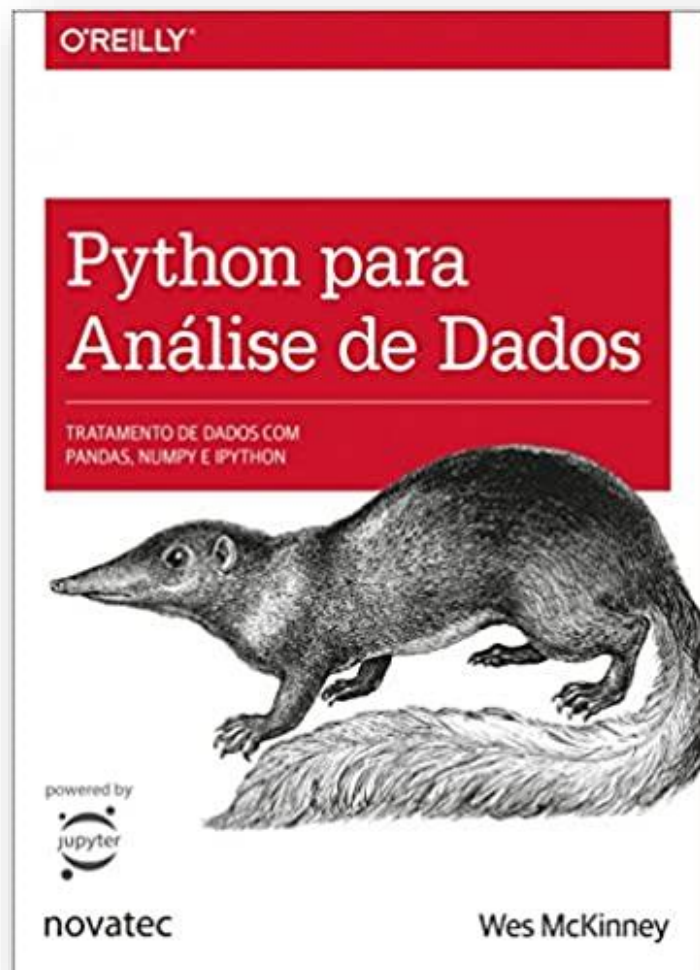
Referências Bibliográficas

LIVROS



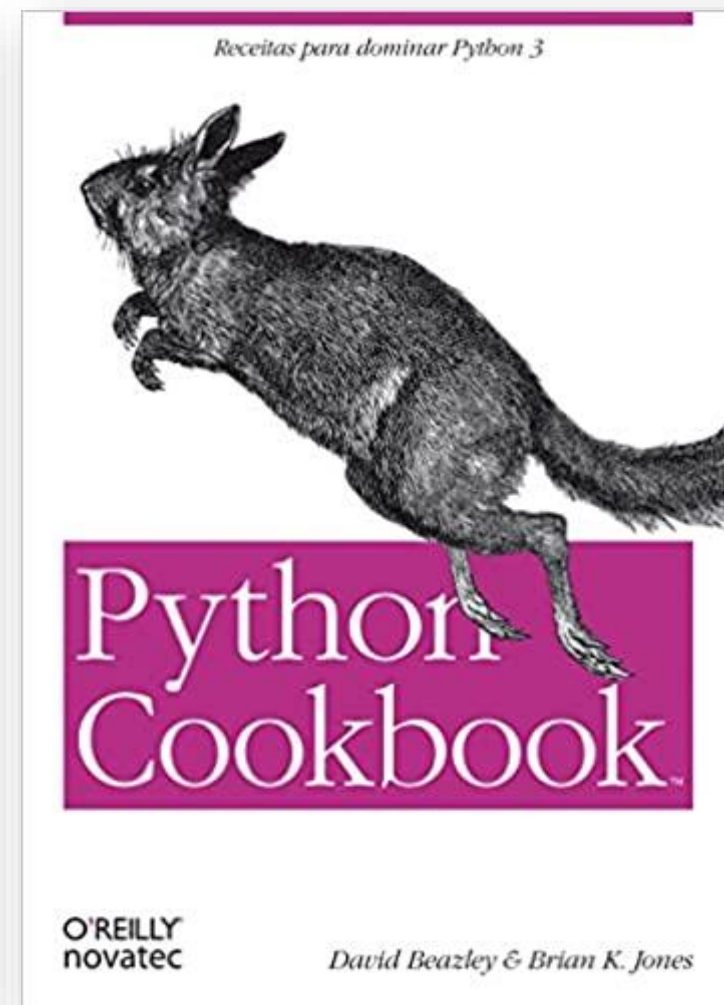
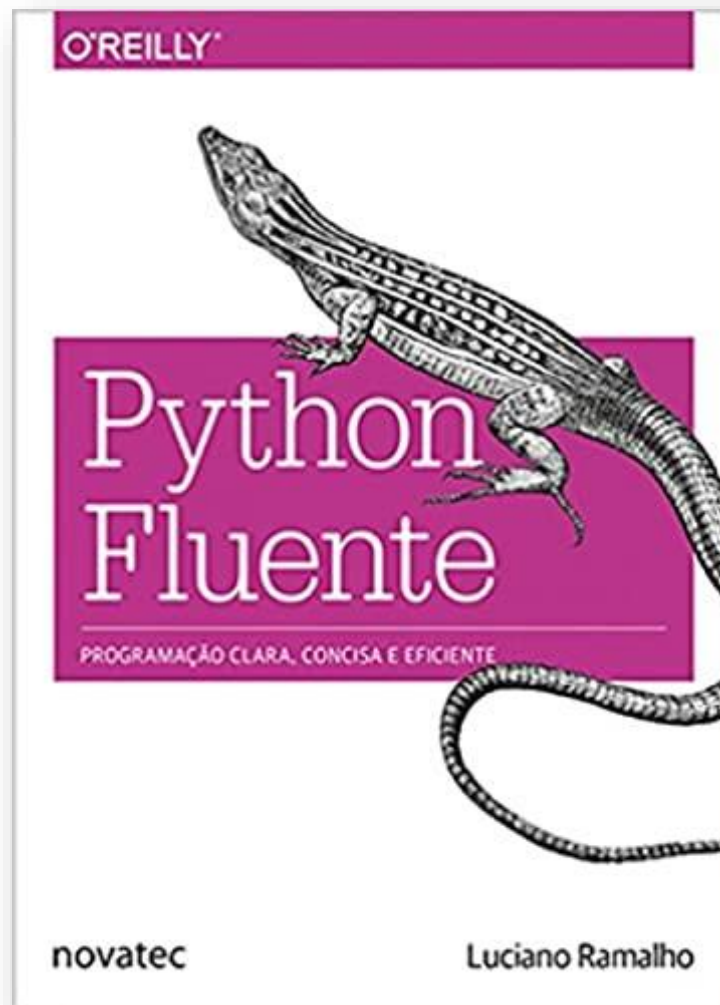
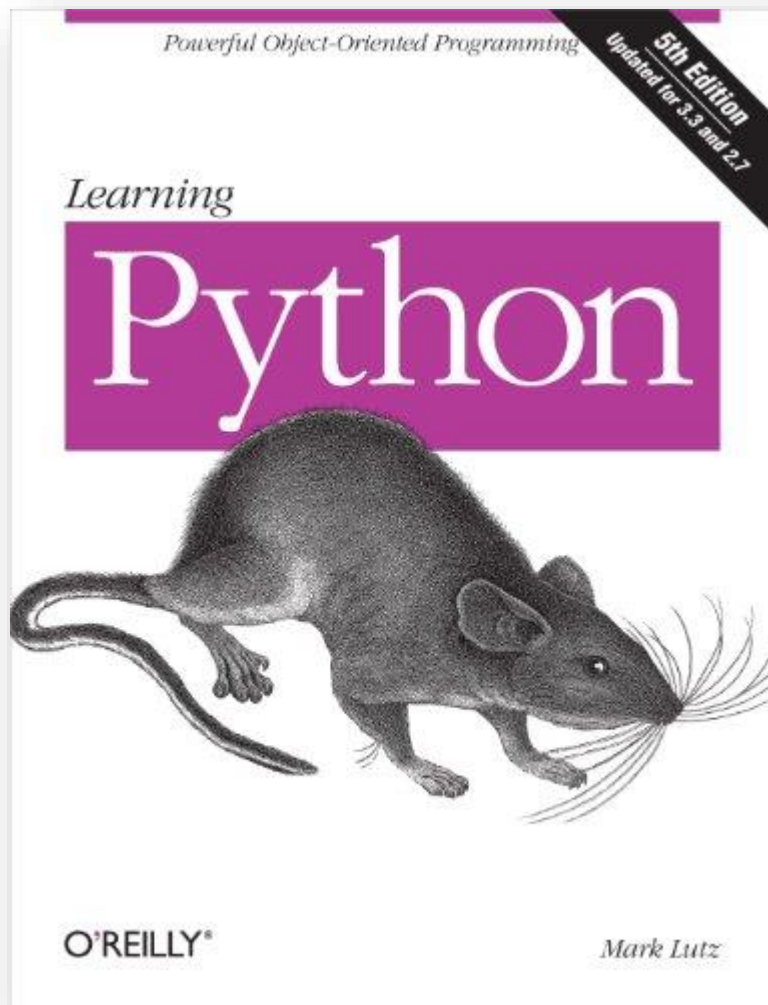
Referências Bibliográficas

LIVROS



Referências Bibliográficas

LIVROS



Referências Bibliográficas

LINKS, ÍCONES, IMAGENS

- As referências de links utilizados podem ser visualizados em <http://urls.dinomagri.com/refs>
- Tutoriais disponíveis no site oficial do Pandas - <http://pandas.pydata.org/pandas-docs/stable/>
- Livro de receitas disponíveis no site oficial do Pandas - <http://pandas.pydata.org/pandas-docs/stable/cookbook.html>
- As imagens foram Icon made by [Srip](#), [Pixel perfect](#), [Eucalyp](#) e [Prettycons](#) from www.flaticon.com

