

A low-angle, close-up shot of a child's legs as they walk on a paved path. The child is wearing tan cargo pants and grey sneakers with blue accents. The background is a soft-focus landscape with trees and a bright, low sun creating a warm, golden glow and lens flare. The overall mood is peaceful and hopeful, symbolizing the beginning of a journey.

Lesson #01

First Steps on Data Science

July 2019



Ivanovitch Silva



ivan@imd.ufrn.br



<https://github.com/ivanovitchm>



https://twitter.com/ivanovitch_slv



AGENDA FOR LONG-TERM LEARNING

Data Science Foundation Data Pipeline

- Store
- Collect
- Cleaning
- Analyse
- Visualize



- 
- A close-up photograph of a spiral-bound notebook. A wooden pencil lies diagonally across the left side of the frame. The notebook's pages show a calendar for the month of May, with dates 27 (L), 28 (S), 29 (M), and 30 (T) visible. The spiral binding is on the right side. The background is slightly blurred, focusing attention on the notebook and pencil.
- 1. How to become a Data Scientist**
 - 2. Platforms for Data Science**
 - 3. Python Crash Course**



Introduction





Group Survey

Who am I?
My master/phd research is about ...



How to Become a **Data Scientist**



DATA Engineer

A Data-Driven Program

DATA Scientist

Develops, constructs, tests, and maintains architectures. Such as databases and large-scale processing systems.

Cleans, massages and organizes (big) data. Performs descriptive statistics and analysis to develop insights, build models and solve a business need.



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

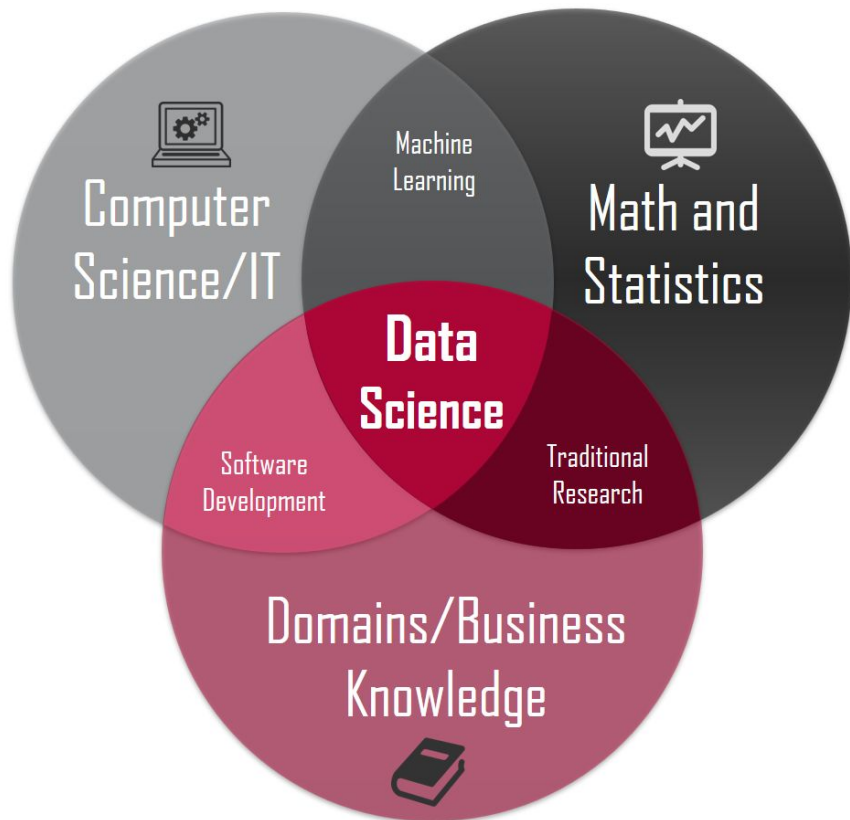


PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



which programming
language to learn first (DS)?





visualize

Seaborn
Folium

matplotlib

Leaflet



Bokeh



plotly

analyze



Keras

PYTORCH



NetworkX



PyMC3

PyGSP

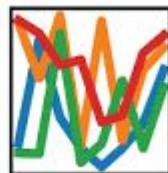
collect, pre-processing, prepare



NumPy

pandas

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



data comes from everywhere



Data Files
(XML, CSV, Excel, JSON, ...)



Database
(MySQL, Oracle, ...)



API



Sites



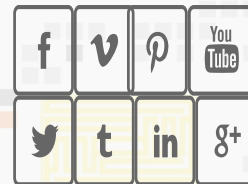
Text and reports



Maps



Image and videos



Social Media



PESQUISAR DADOS

Ex.: cursos



Etiquetas mais comuns

requisição

requisições

materiais

GRUPOS



Comunicados e
Documentos



Contratos e
Convênios



Despesas e
Orçamento



Ensino



Extensão



Institucional



Materiais



Patrimônio



Pesquisa



Pessoas

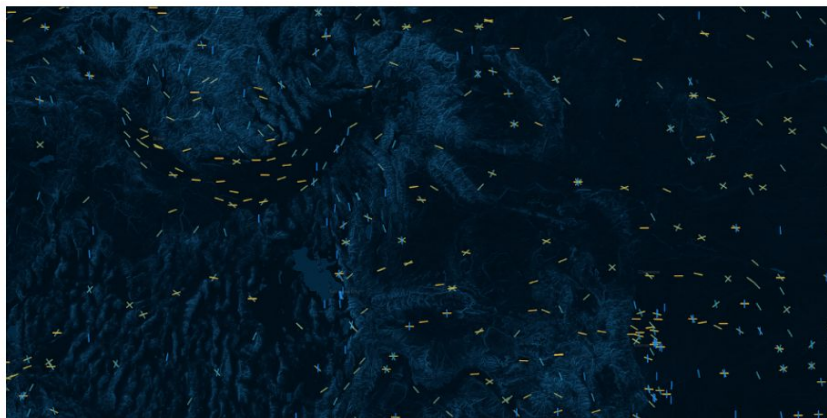


Processos



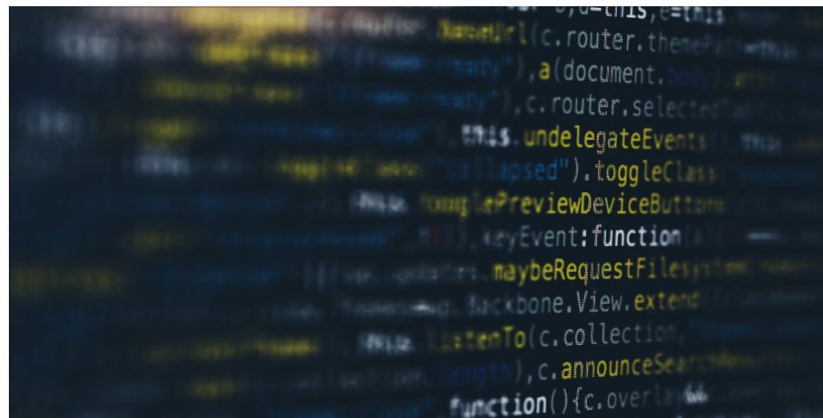
Towards Data Science

Sharing concepts, ideas, and codes

[DATA SCIENCE](#)[MACHINE LEARNING](#)[PROGRAMMING](#)[VISUALIZATION](#)[AI](#)[JOURNALISM](#)[EVENTS](#)[SUBMIT](#)[Follow](#)

Making of: Trails of Wind

How we created a map of the global architecture of airport runways, which turned out to be a wind map.



Elements of Functional Programming in Python

Learn how to use the lambda, map, filter, and reduce functions in Python to transform data structures.

13 Essential Newsletters for Data Scientists: Remastered

You're going to want this data science and AI-focused content



Conor Dewey in Towards Data Science

Follow

Mar 17 · 6 min read





Data science Platforms

<https://www.kdnuggets.com/2019/02/gartner-2019-mq-data-science-machine-learning-changes.html>

<https://medium.com/etteam/10-of-the-best-platforms-for-data-science-and-machine-learning-36a61ec1a676>

Simple Jupyter demo

This cell has text formatted using the markdown language, which gets rendered like regular html.
The next cell has some code:

```
In [57]: import random
         for i in range(3):
           print random.random()
         x = 10

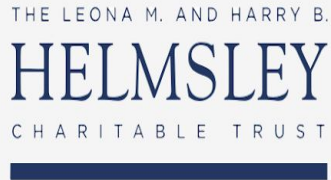
0.10564822904
0.153941700348
0.518503128416
```

Here is another text cell, with some *formatting*.



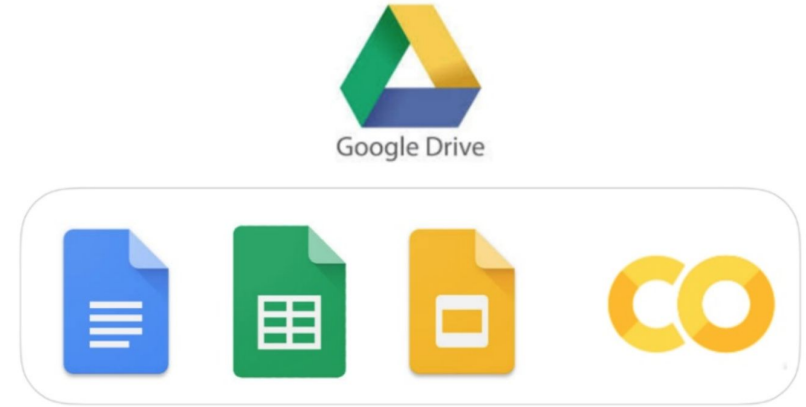
Sponsors

Project Jupyter receives direct funding from the following sources:



Google Colaboratory

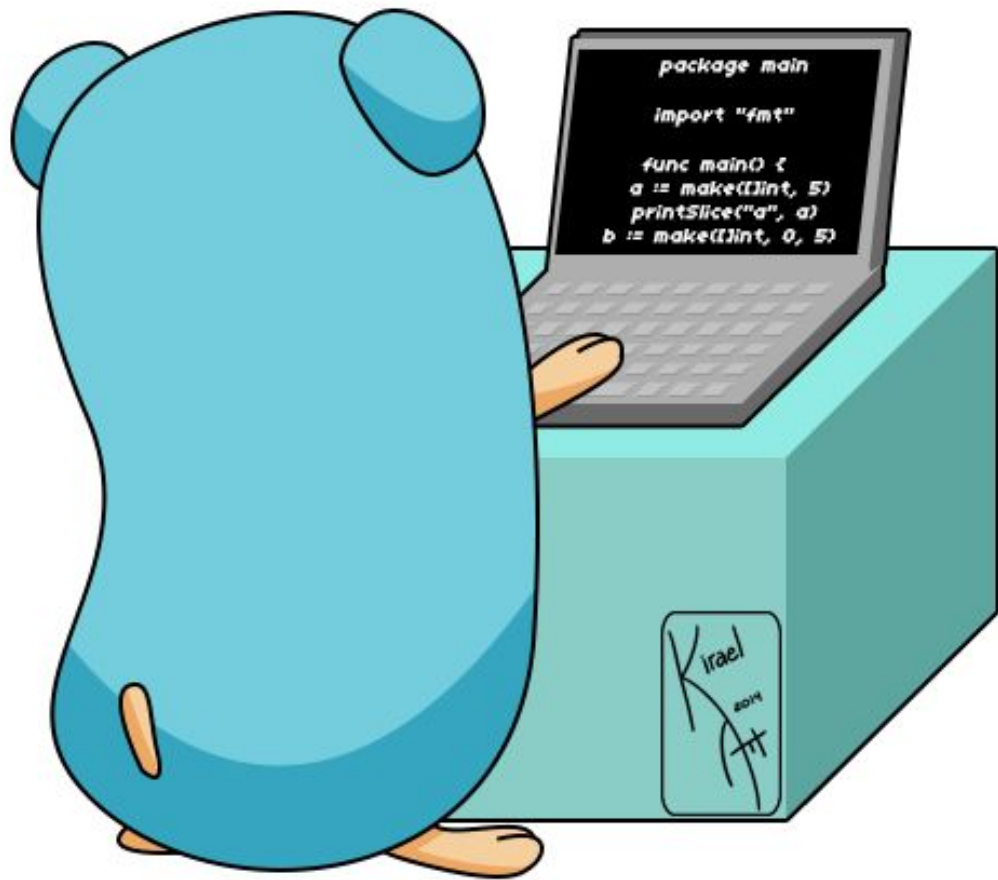
<https://colab.research.google.com/>

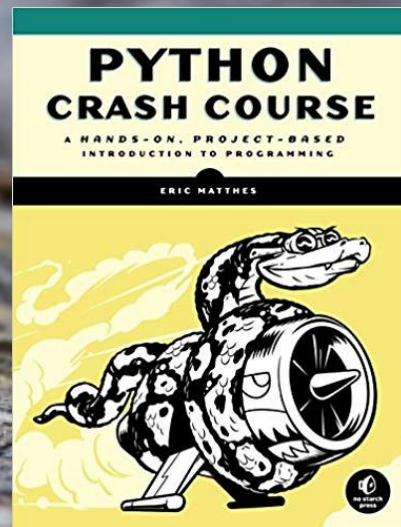


Colaboratory is a Google research project created to help disseminate **machine learning education and research**. It's a Jupyter notebook environment that **requires no setup** to use and **runs entirely in the cloud**.

Colaboratory notebooks are stored in Google Drive and can be shared just as you would with Google Docs or Sheets. Colaboratory is **free to use**.

Hands on





Lists of Lists

**Hello
World**

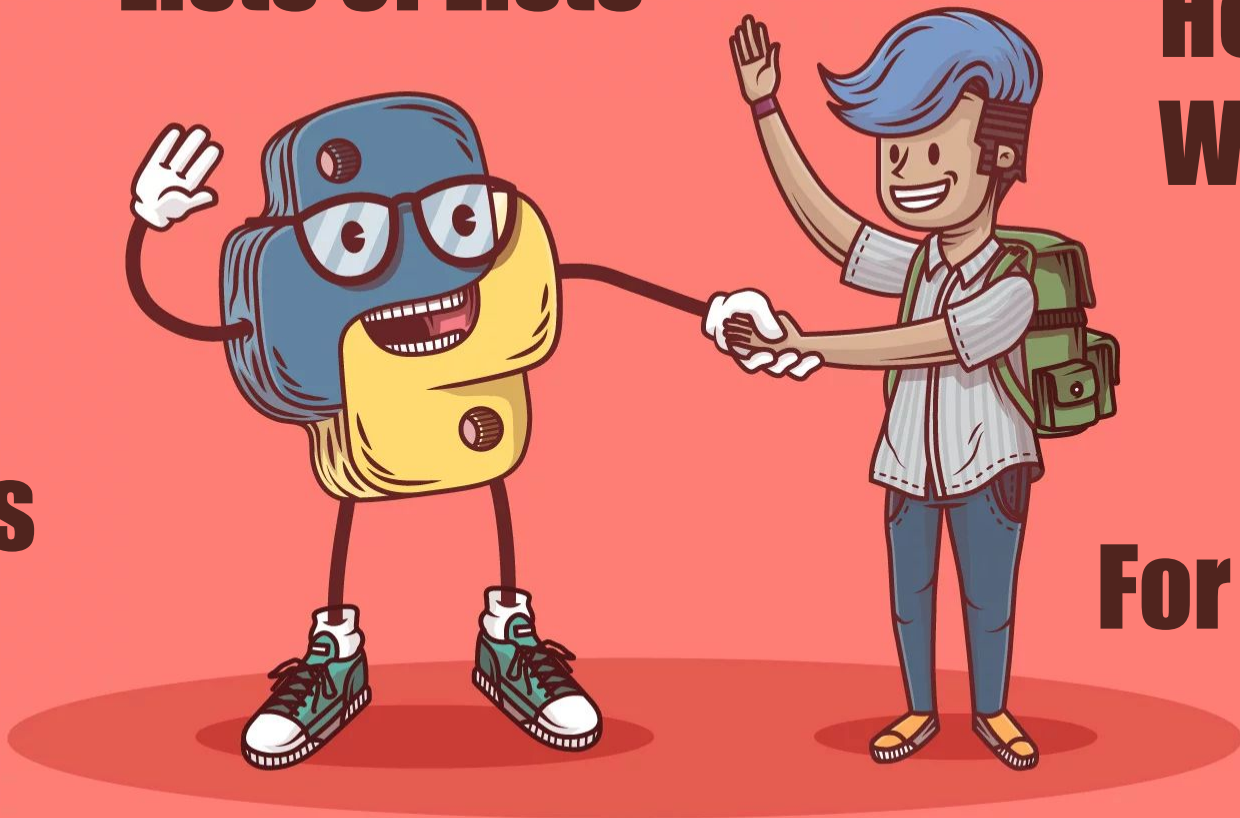
Lists

Files

For Loops

Conditional Statements

Real Python



Dataset


Released Under GPL 2

^

337

Mobile App Store (7200 apps)

Analytics for Mobile Apps



Ramanathan • updated 8 months ago (Version 7)

Data

Overview

Kernels (69)

Discussion (8)

Activity

Download (6 MB)

New Kernel

⋮

Data (6 MB)

⛶

Data Sources

AppleStore.csv

7198 x 17

appleStore_descriptio...

7197 x 4

About this file

Apple Store

Columns

~

user_rating

user_rating_ver

A ver

A cont_rating

A prime_genre

sup_devices.num

iPadSc_urls.num

lang.num

| | track_name | price | currency | rating_count_tot | user_rating |
|---|-------------------------|--------------|-----------------|-------------------------|--------------------|
| 0 | Facebook | 0.0 | USD | 2974676 | 3.5 |
| 1 | Instagram | 0.0 | USD | 2161558 | 4.5 |
| 2 | Clash of Clans | 0.0 | USD | 2130805 | 4.5 |
| 3 | Temple Run | 0.0 | USD | 1724546 | 4.5 |
| 4 | Pandora - Music & Radio | 0.0 | USD | 1126879 | 4.0 |



AppleStore.csv

[Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)[ivanovitchm](#) / [datascience4demography](#)[Watch](#) ▾

0

[★ Star](#)

0

[Fork](#)

0

[Code](#)[Issues](#) 0[Pull requests](#) 0[Projects](#) 0[Wiki](#)[Security](#)[Insights](#)[Settings](#)

No description, website, or topics provided.

[Edit](#)[Manage topics](#)[3 commits](#)[1 branch](#)[0 releases](#)[1 contributor](#)Branch: [master](#) ▾[New pull request](#)[Create new file](#)[Upload files](#)[Find File](#)[Clone or download](#) ▾[ivanovitchm](#) Update README.md

Latest commit 8ad6efa 8 minutes ago

[README.md](#)

Update README.md

8 minutes ago



README.md



Data Science for Demography

- Kickoff: **Quartas Demográficas** talk
- Lesson #01
 - Platforms for Data Science
 - Google Colab
 - Python crash course (hello world, loops, list, list of lists, files)

Lesson #01 Python crash course.ipynb

