Python Fundamentals.

Voor Data Scientists en Data Engineers

wortell

Onderwerpen

- Welkom, introductie, kennismaking
- Hoogover uitleg: wat is Python? Wat is Pandas?
- Jupyter Notebook
- Inlezen
- Inspectie en verkennen van data
- Selectie en filteren
- Data wrangling
- Joins
- Visualisatie
- Eigen functies
- Installatie en opzet van Jupyter, Python en Pandas

wortell

Introductie.

Waar sta je nu?

- Waar werk je?
- Welke software?
- Gebruik je al Python?

Wat is je leerdoel?

wortell

Waarom Python + Pandas?

Eating the World!

- Leesbare code
- Snel te schrijven
- Efficiënte libraries

Data-analyse, ML, wrangling

- Maar ook handig om snel een API aan te spreken
- .. of een website te scrapen

Spark, Databricks, ...

- Koalas / PySpark

Leuk om iets nieuws te leren

wortell

Waarom GEEN Python / Pandas?

Higher-level language

- Dus ook langzamer
- … ”premature optimization is the root of all evil” (Donald Knuth)

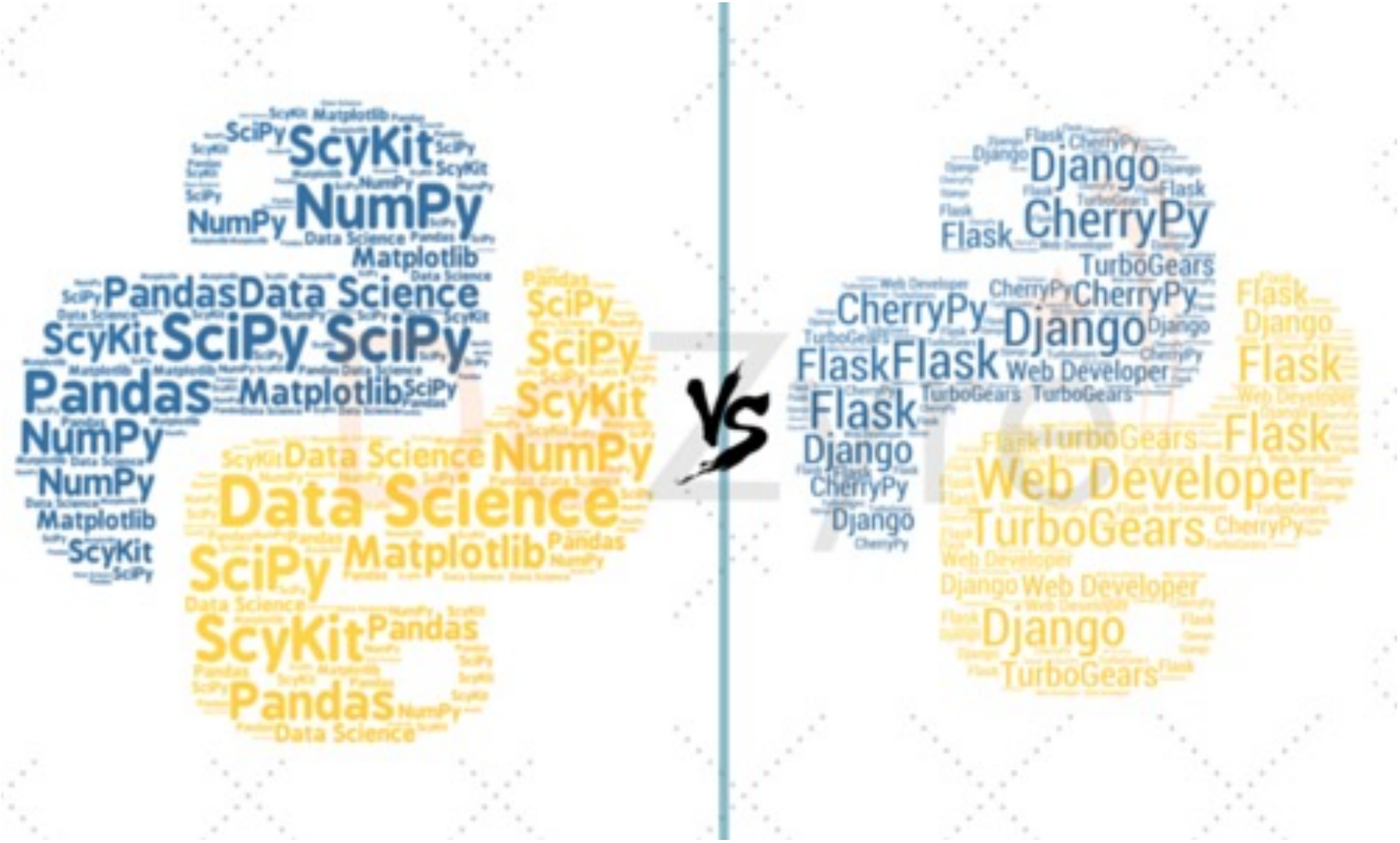
Pandas: alles in-memory

- Voor echt grote datasets gedistribueerde oplossingen (Spark / Dask)

Niet geschikt voor mobile

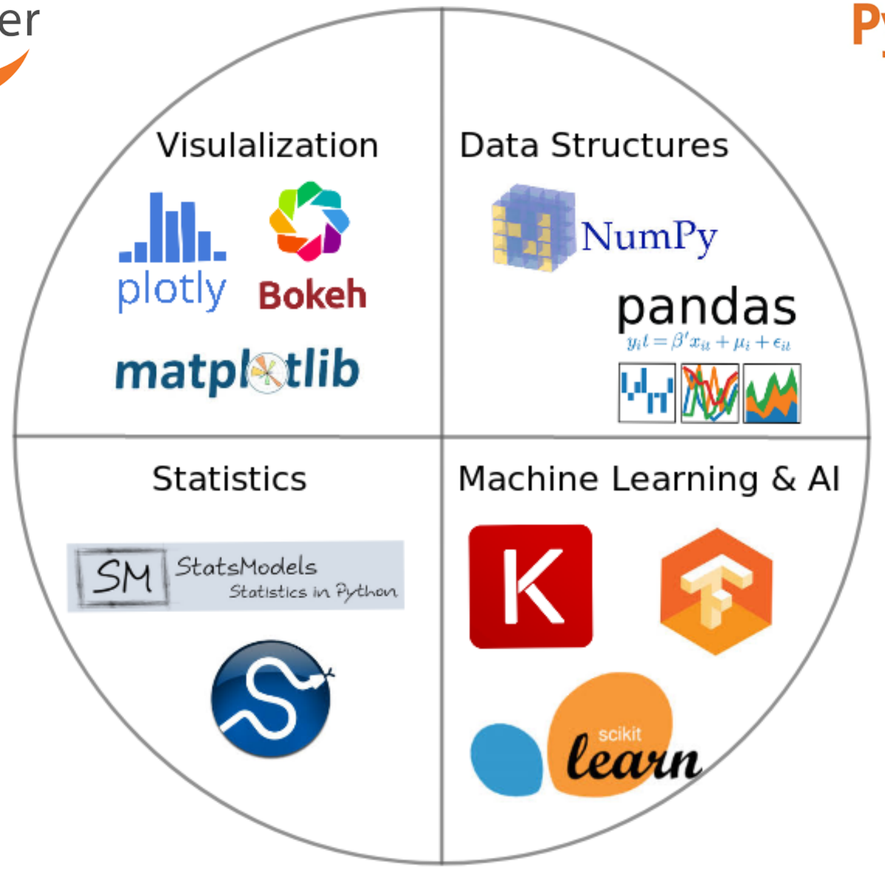
wortell

Wegen naar Python



wortell

Het Python Data-ecosysteem



jupyter

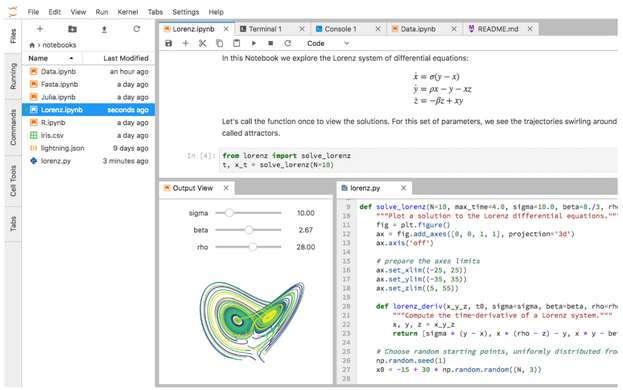
PySpark

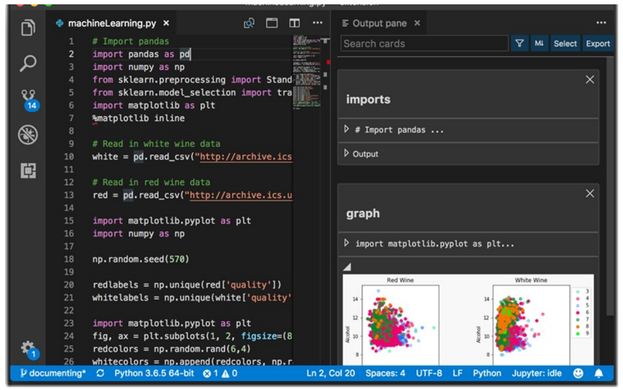
Apache Airflow

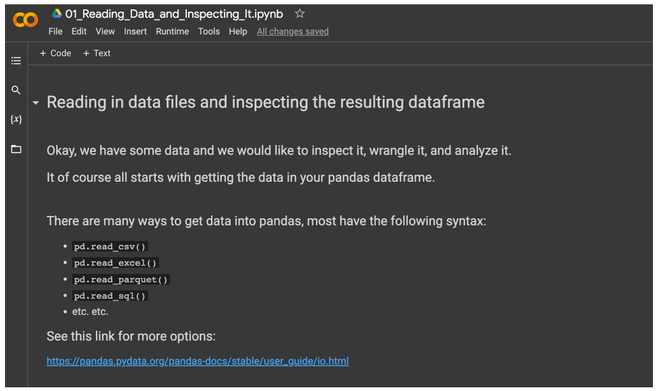
wortell

Jupyter

- Intro Jupyter
- Jupyter Notebook, Jupyter Lab, VS Code, Google Colab, ...

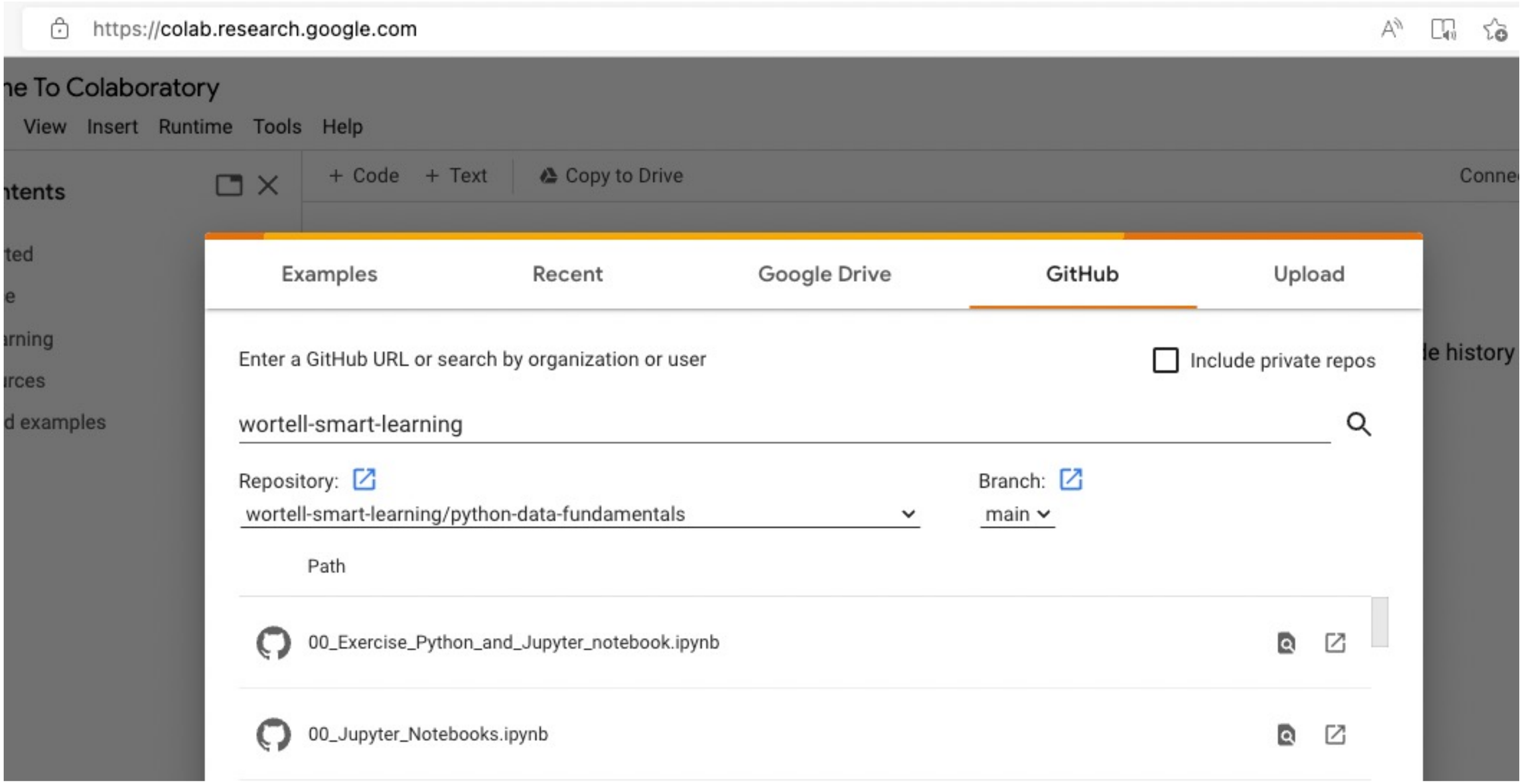






wortell

Eerste stappen met Jupyter in Colab



Datatypes in Python

Name	Type	Description
Integers	int	Whole numbers, such as: **3 300 200**
Floating point	float	Numbers with a decimal point: **2.3 4.6 100.0**
Strings	str	Ordered sequence of characters: **"hello" 'Sammy' "2000" "楽しい"**
Lists	list	Ordered sequence of objects: **[10,"hello",200.3]**
Dictionaries	dict	Unordered Key:Value pairs: **{"mykey" : "value" , "name" : "Frankie"}**
Tuples	tup	Ordered immutable sequence of objects: **(10,"hello",200.3)**
Sets	set	Unordered collection of unique objects: **{"a","b"}**
Booleans	bool	Logical value indicating **True** or **False**

wortell

Jupyter: tips & tricks

- Shift + Enter to run code
- Tab completion
- Nieuwe cell: Escape gevolgd door a (above) of b (below) of dd (delete)
- Shift Tab to see arguments and information about methods, functions or classes
- Magic commands, such as ls
- ? or ?? to get extra help and info

wortell

Pandas intro

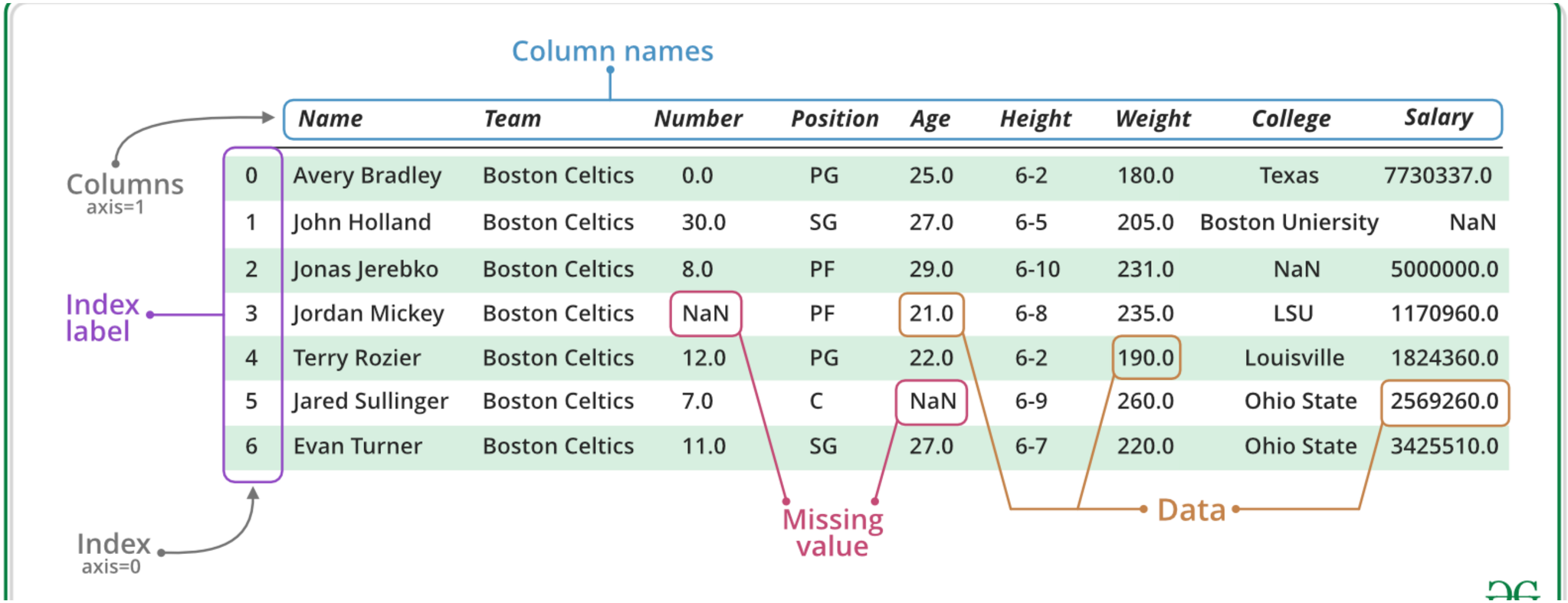
- Data inlezen pd.read_csv()
- Data inspectie df.info() df.head() df.describe()
- Data selectie df[df.column == ‘value’] df.loc[df.column == ‘value’, :]
- Data wrangling df[‘column’].fillna() df.drop_duplicates()
- Data joinen df.merge(df2, how=‘inner’, on=‘column_name’)
- Data visualisatie px.scatter(df, x, y)

pandas

pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

wortell

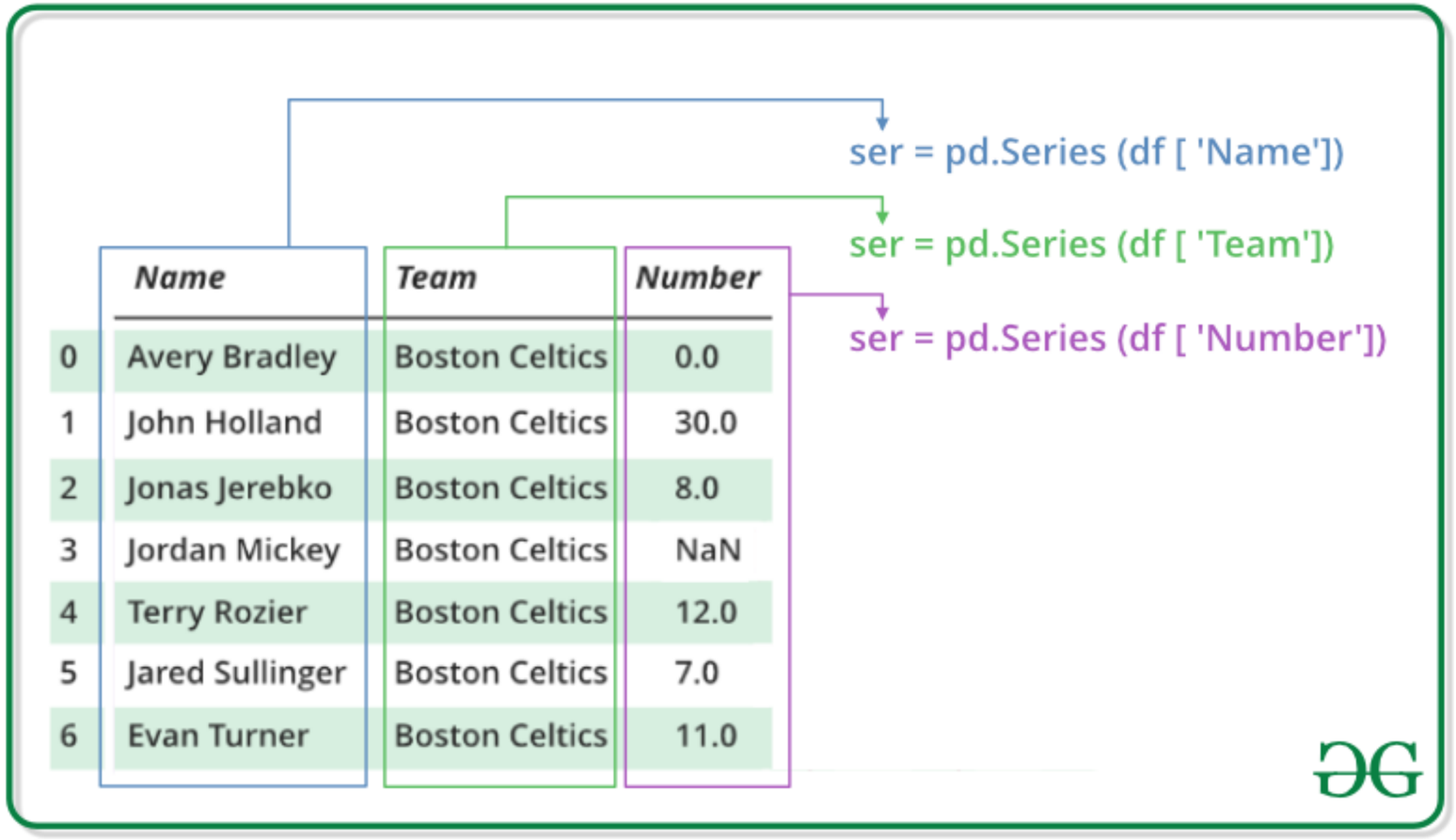
Pandas: the DataFrame (df)



	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston Uniersity	NaN
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0	6-8	235.0	LSU	1170960.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0	6-2	190.0	Louisville	1824360.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN	6-9	260.0	Ohio State	2569260.0
6	Evan Turner	Boston Celtics	11.0	SG	27.0	6-7	220.0	Ohio State	3425510.0

wortell

Pandas: the Series



	Name	Team	Number
0	Avery Bradley	Boston Celtics	0.0
1	John Holland	Boston Celtics	30.0
2	Jonas Jerebko	Boston Celtics	8.0
3	Jordan Mickey	Boston Celtics	NaN
4	Terry Rozier	Boston Celtics	12.0
5	Jared Sullinger	Boston Celtics	7.0
6	Evan Turner	Boston Celtics	11.0

wortell

Next slides contains elements you can copy and paste with the ‘format’ tool.

wortell

Photo Side.

Big photo covering the side

A photo doesn’t always have to be circular or rounded. It can also contain 50% or less/more on the side as a separator.

Photo Circular.

How do you do this?

To get the circular photo, make sure that you have a square photo and simply copy the formatting of the sample to your required photo. If the photo isn't squared yet, read below on how to do that.

How to crop on 1:1 ratio

- Paste your photo
  - Ideal situation would be a 1:1/square photo
  - If not, you can crop it with the instructions below
- Go to Picture Format > Crop > Square 1:1
- Copy and paste the formatting from the sample on the right to the new one.

wortell

Drop Shadow Element.

Drop Shadow

Voor vlakken achter een tekst, of andere elementen die kadering nodig hebben. (Niet verplicht)

Just a sample

- Bullet 1
- Bullet 2
- Bullet 3

wortell

- Just a big bullet
- Another one
  - Sub bullet
  - Another sub bullet

wortell