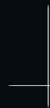


Nové a lepší dataframy v Pythonu



Co je data frame?

Data frame je *tabulková datová struktura*, která se běžně používá v programování a analýze dat. Tato struktura obsahuje řádky a sloupce, kde sloupce představují proměnné a řádky jednotlivá pozorování.

Každý sloupec v *data frame* může obsahovat hodnoty jen jednoho datového typu, ale různé sloupce mohou mít různé datové typy.

Pandas – náš hlavní vzor

- široce používáno
- ostatní knihovny pro zpracování tabulkových dat emulují rozhraní



Columns

Rows

	Name	Score	Attempts	Qualify
0	Anastasia	12.5	1	yes
1	Dima	9.0	3	no
2	Katherine	16.5	2	yes
3	James	NaN	3	no
4	Emily	9.0	2	no

Data

Pandas DataFrame

Náš cíl

- využití čistě pythonských kolekcí
 - velké
 - pomalé
- vhodné pro výuku v úvodním kurzu
 - přátelné (*user friendly*)

Hledání zvířecího maskota



DALL·E

koláž několika roztomilých zvířat

Vítězem je ...



Poznáváte?

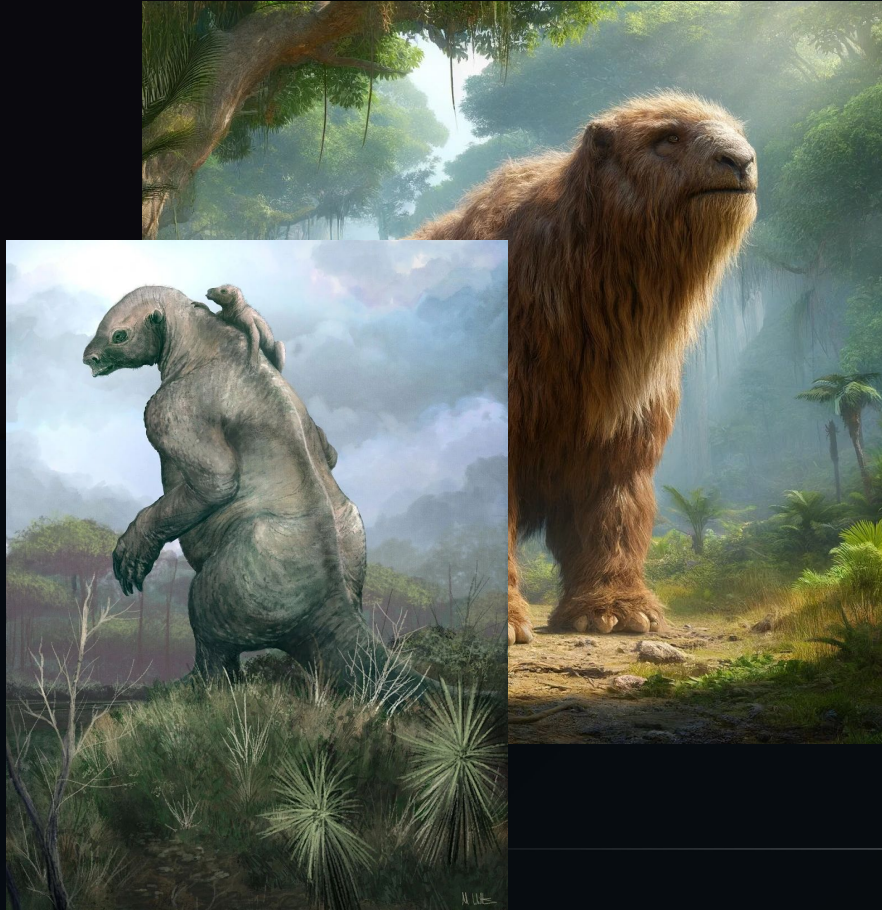
Vítězem je ...



Poznáváte?

- je vymřelé

Vítězem je ...



Poznáváte?

- je vymřelé
- vypadalo úplně jinak

© DALL·E

Vítězem je ...



Poznáváte?

- je vymřelé
- vypadalo úplně jinak
- © DALL·E
- a má nezapamatovatelné jméno

Vítězem je ...



Megatherium
Μεγαθηριον

Kde ho najdete?

github.com/Jiri-Fiser/megatherion

```
from megatherion import DataFrame

df = DataFrame(dict(
    a=Column([None, 3.1415], Type.Float),
    b=Column(["a", 2], Type.String),
    c=Column(range(2), Type.Float)
))
print(df)

df = DataFrame.read_json("data.json")
print(df)
```