Day 4 Notes - Numpy, Pandas

# Cool-Demo video (start of the day)

[Computer learning to play breakout](https://www.youtube.com/watch?v=V1eYniJ0Rnk)

# Cool-Demo exercise: Transformers (end of the day)

Online demo (writing with transformer) : <https://transformer.huggingface.co/>

Code repo : <https://github.com/elephantscale/cool-demos>

Notebook : <https://github.com/elephantscale/cool-demos/blob/main/transformers/GPT2_text_generation.ipynb>

Read about GPT2

<https://openai.com/blog/better-language-models/>

Available models : https://huggingface.co/transformers/pretrained\_models.html

# Misc Links

**Jupyter magic functions**

<https://ipython.readthedocs.io/en/stable/interactive/magics.html>

**Regular expression testing**

<https://regex101.com/>

<https://pythex.org/>

[Regex for US phone numbers](https://stackoverflow.com/questions/123559/how-to-validate-phone-numbers-using-regex)

**Pandas on Spark - Koalas**

<https://koalas.readthedocs.io/en/latest/>

<https://databricks.com/blog/2020/06/24/introducing-koalas-1-0.html>

**Modern, Efficient Data formats**

<https://arrow.apache.org/>

# Extra Lab

Analyzing flight data with Pandas

Start with the repo : https://github.com/elephantscale/python-data-science-workshop

https://github.com/elephantscale/python-data-science-workshop/blob/main/pandas/2-pandas-exploration.ipynb

# Numpy

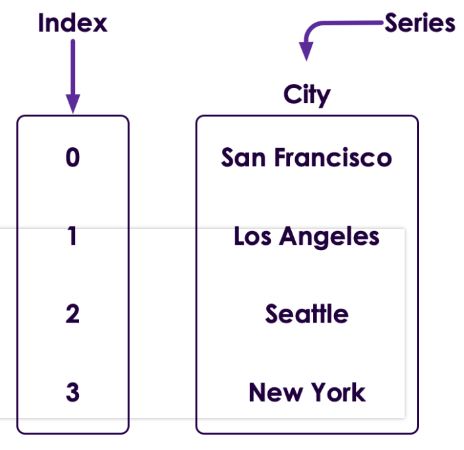
<https://towardsdatascience.com/why-numpy-is-so-fundamental-78ae2807300>

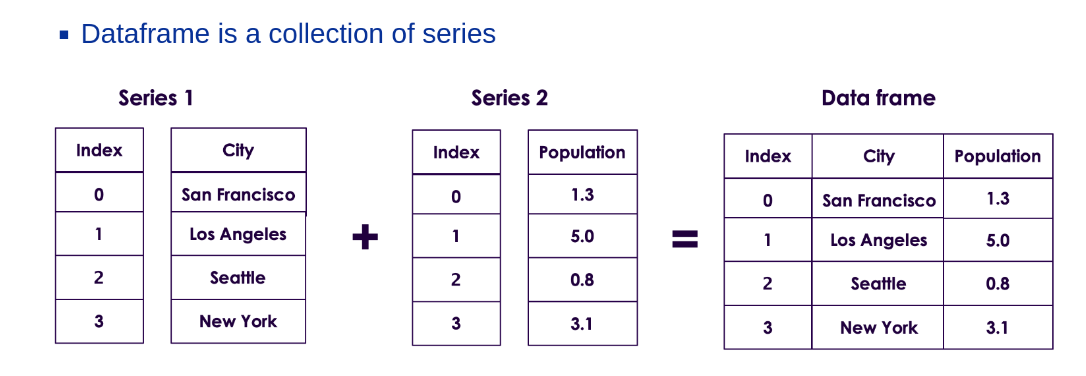
# Caching

<https://cython.readthedocs.io/en/latest/src/userguide/numpy_tutorial.html>

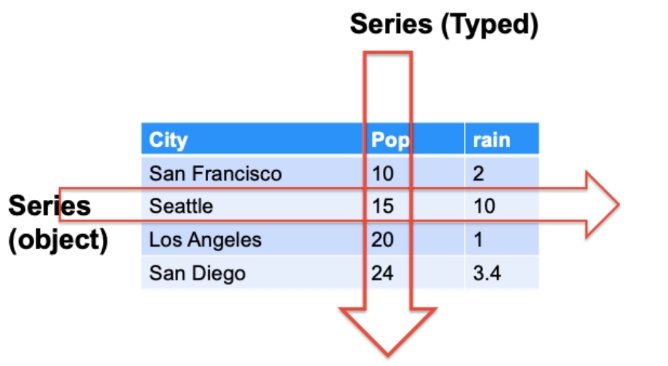
# Pandas

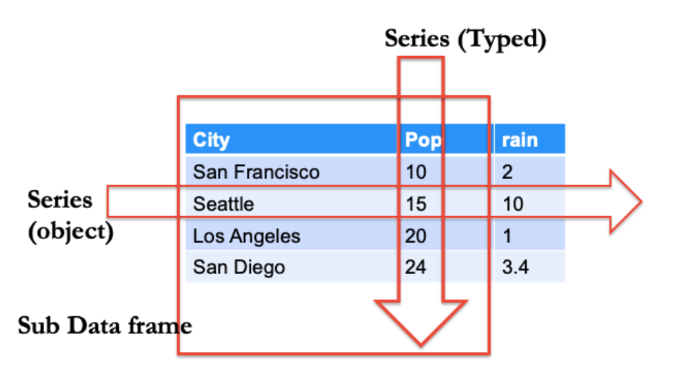
## Series

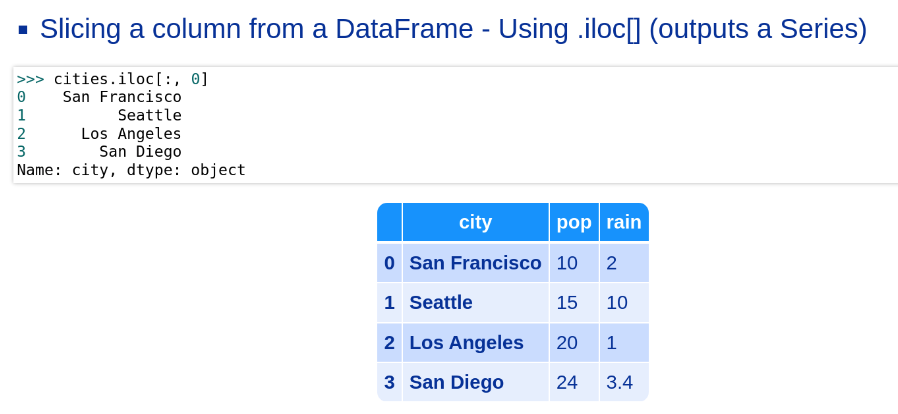


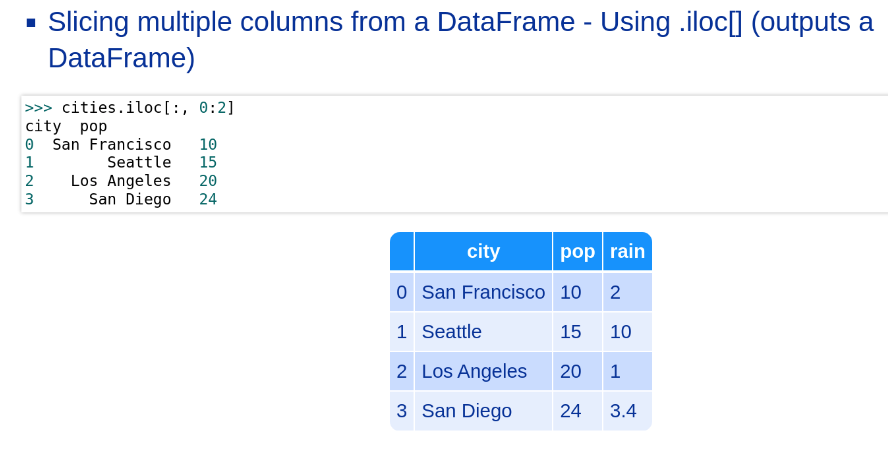


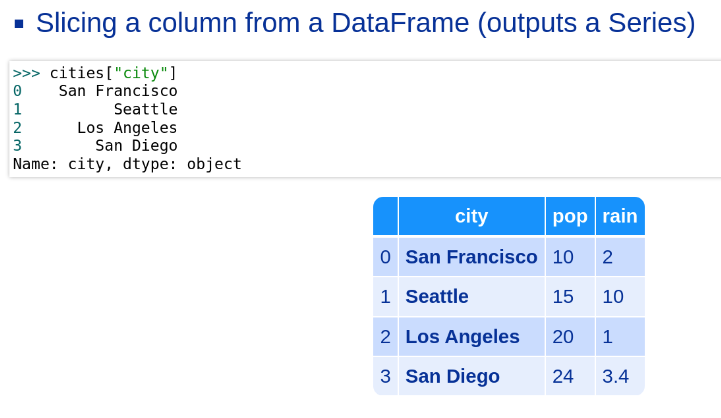
## DataFrame Slicing

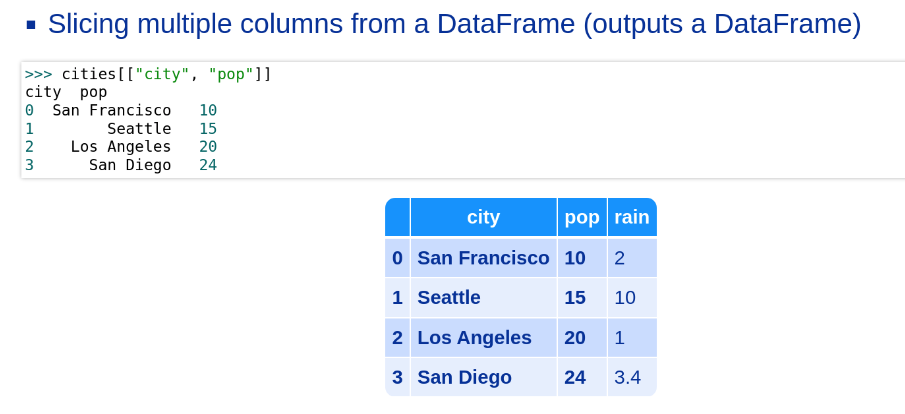


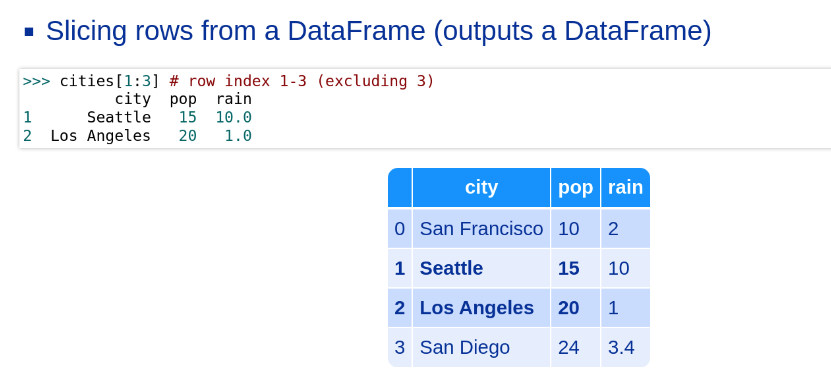












## Understanding Pandas Axis

[Pandas Axis Explained](https://railsware.com/blog/python-for-machine-learning-pandas-axis-explained/) - great article explaining axis with examples!

