THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Introduction to Python

Nikhil Jangamreddy,
HDR Student,
School of Electrical Engineering and Computer Science.

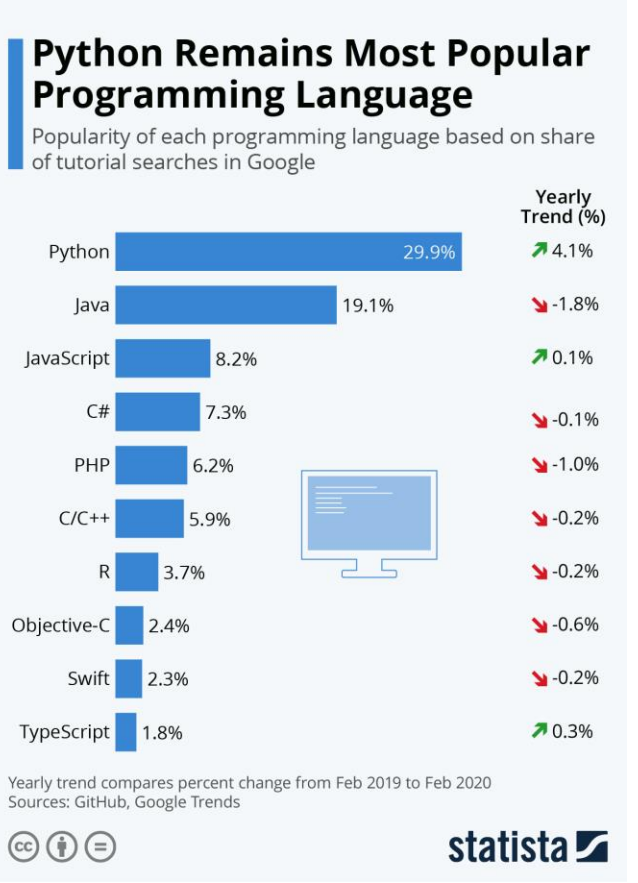
Overview

- Setting up Python for the course.
- Data preparation and Data cleaning.
- Two real world datasets.

Why Python?

Why Python is popular in data science and statistics community?

- Python is a versatile language widely used in data science and statistics due to its simplicity and extensive library support.
- Python Libraries:
  - **NumPy** and **pandas -** manipulate data efficiently, mathematical operations.
  - **Scipy** and **statsmodels** provide methods for advanced statistical tests.
  - **Seaborn** and **Matplotlib** help us visualize data and results in a meaningful way.
  - Data science/Machine learning – Scikit-learn, Tensorflow (Google) Pytorch (Facebook), etc.



Setting up Python for course

- Python 3.x is recommended as Python 2.x is no longer supported.
- To get started, you can install the Anaconda distribution, which is a data science platform that includes Python and a lot of handy tools. For this session, we'll be using Python 3.x as Python 2.x is no longer maintained. (Link)
- How to install libraries in python?
  - Package manager pip, the package manager for Python.
    - Pip install package-name
    - example: pip install numpy
- Online tools: Google colab, Kaggle, Deepnote etc..

Data preparation and cleaning

- https://github.com/jnikhilreddy/AI-for_cyber_resources

Github Codespaces?

- Entirely cloud based and we can directly work on Github code.
- Very helpful for beginners.
- Github copilot.
- Demonstration