PYTHON DATA ESSENTIALS

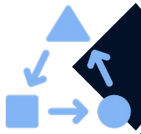
- INTRODUCTIONS
- LEARNING OBJECTIVES
- PYTHON OVERVIEW

AGENDA

DATA SCIENCE & PYTHON
KEY LEARNING OBJECTIVES

Be able to describe the key elements of the Python data science eco-system

Gain some hands-on experience of data wrangling with Python



Be able to load data into Python and perform basic operations

Be able to create static data visualizations in Python

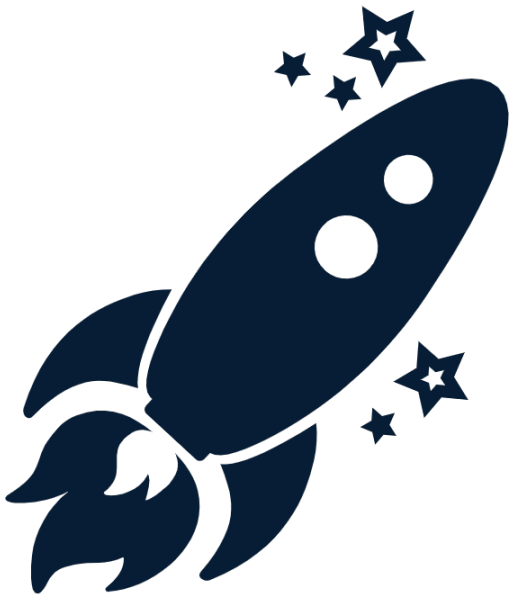
Neueda

- WHAT IS PYTHON?
- WHAT CAN YOU DO WITH IT?
- WHY IS IT SO POPULAR?

PYTHON – THE TOP 3 QUESTIONS

WHAT IS PYTHON?

The world's fastest growing programming language



- Software Engineers
- Data Analysts
- Accountants
- Mathematicians
- Scientists
- + Kids

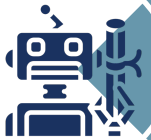
WHAT CAN YOU DO WITH IT?

Data Analysis

Analyze big datasets much more effectively than with Excel

Automation

Automate repetitive tasks such as copying files, sending emails or generating reports



Artificial Intelligence and Machine Learning

Combine these top two applications – Data Analysis & Automation – alongside the huge range of free utilities and you have the perfect AI platform

Building Apps and Web Sites

Traditional programming tasks

Neueda

WHY IS IT SO POPULAR?

Beginner Friendly

uses considerably less code than other languages while producing similar outcomes, “reads like English”

Batteries Included

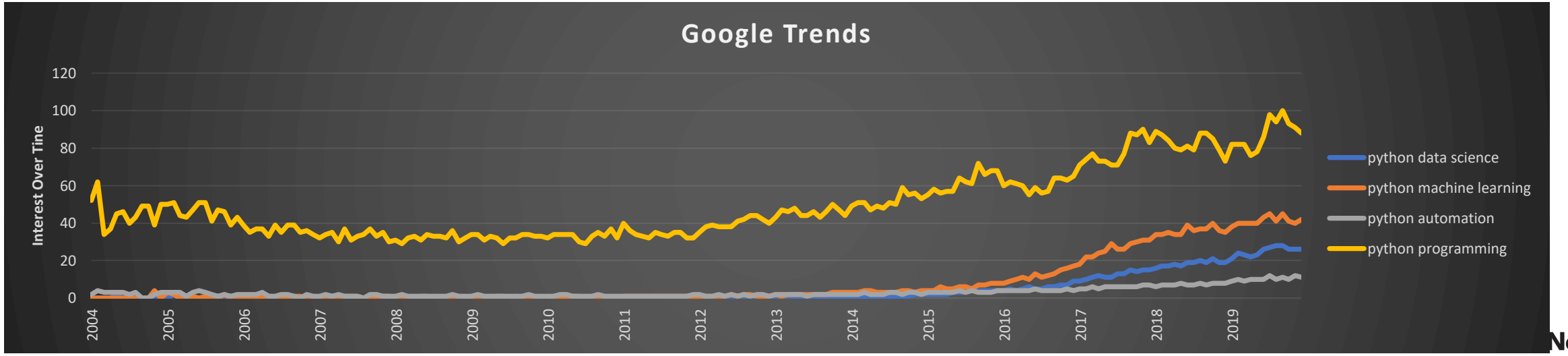
ships with many standard libraries right out of the box

Libraries

a wealth of open source libraries built on Python

Versatility

mobile & web, ML, NLP data science, gaming, task automation, finance



Interpreted, high level programming language

First released in 1991 by Guido van Rossum with goals of

- small core language
- a large standard library
- easily extensible interpreter

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Readability counts.

The Zen of Python

HISTORY & ADOPTION BY BUSINESS COMMUNITY

1989
- Core Datatypes & Collections
- Classes, functions, exceptions

1994 – V 1.x
- Functional Programming (map,reduce,filter)
- Keyword arguments

2000 – V 2.x
- Comprehensions
- Generators
- Context Manager

2008 – V 3.x
- Not compatible with V 2.x
- Reduce feature duplication by removing old ways of doing things

YouTube	Google	Instagram	Reddit	Reddit	Dropbox	Spotify
Every time you watch a video, you're executing Python code	An official language at Google used – in things such as system administration tools and many Google App Engines	It's application servers are written in Python using Django as the web framework	Rewritten in Python in 2005 to gain greater development flexibility	Python module, Luigi, is used to power the Radio and Discover features, as well as the recommendations	Both the Dropbox server (running on the cloud) and desktop client software were primarily written in Python.	LiveNode, one of the internal systems that manages the display of content on the webpage, is partly written in Python

Neueda

PYTHON FOR DATA SCIENCE – KEY COMPONENTS
1. PYTHON

- A programming language i.e. the actual language syntax
- An executable program – it takes as it's input Python code and does things based on what it finds in that code
  - This is referred to as the "Python Interpreter"
- The "Python Interpreter" is commonly installed for you e.g. it's running "inside" Jupyter or Dataflame

PYTHON FOR DATA SCIENCE – KEY COMPONENTS
2. JUPYTER

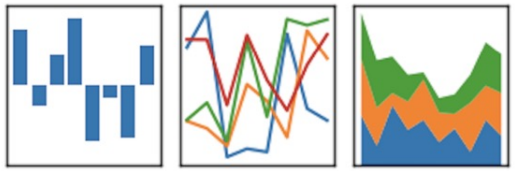
- A "web application" that lets you write python code in a web browser
- The Python code is sent to a "backend" to be executed, the result is displayed in the web browser
- The Citi enhanced version of Jupyter is called Dataflame

PYTHON FOR DATA SCIENCE – KEY COMPONENTS
3. PANDAS

- A "Python Library" i.e. it adds extra features to the core Python language
- Designed specifically for fast manipulation of tables of data
- The core element in Pandas is called a "DataFrame"
- A DataFrame represents a single table of data

pandas

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



PYTHON FOR DATA SCIENCE – KEY COMPONENTS
4. ANACONDA

- A data science platform designed to simplify installing data science tools
- Commonly used to simplify the installation of Python, Jupyter, Pandas
- Anaconda includes Python, Jupyter and Pandas in it's default installation

PYTHON FOR DATA SCIENCE – KEY COMPONENTS
5. PACKAGE INSTALLERS - PIP / CONDA

- Pip is the default Python package management system "Pip Installs Packages"
  - The default source of packages is pypi.org

- Conda is the package management system that comes with Anaconda
- Pip and Conda work in a very similar way, the main differences are
  - They get their packages from different sources
  - Conda is used to install more than just Python packages

LET'S GET STARTED USING PYTHON!

QUESTIONS