Introduction to Python and Programming

DigitalSkola
Uncover The World Of Digital Skills With Us

Profile

linkedin.com/in/romansyasetyo/

“The best part of programming is the triumph of seeing the machine do something useful ”

- Hilary Mason, Data Scientist at Accel

DigitalSkola
Uncover The World Of Digital Skills With Us

Table of Content

What will We Learn Today?

1. What is Python?
2. Why Python?
3. Python usage in Data Science
4. Python vs R
5. Getting started with Python

Hands on using Google Colab

DigitalSkola
Uncover The World Of Digital Skills With Us

What is Python?

What is Python?

Python is a general-purpose, versatile, and powerful programming language.

Simply said, Python is a programming language just like Java, C++, R, etc.

DigitalSkola
Uncover The World Of Digital Skills With Us

Why Python?

Why Python?

- Python's syntax is simple & intuitive. They're basically English words!
- General programming language, easier to collab with others
- Free and open source language
- Complete libraries and frameworks for data analysis and machine learning
- Many communities and users, so very helpful when you go through Stack Overflow to search for answer
- Scalability

DigitalSkola
Uncover The World Of Digital Skills With Us

Python usage in Data Science

Data Collection & Cleansing

Data Exploration

Data Modelling

Data Visualization

Deploying Model

Data Collection & Cleansing

With Python, we can play with almost all sorts of data that are available in different formats such as:

- Structured Data (e.g. Excel, SQL Table)
- Semi-structured Data (e.g. CSV, JSON)
- Unstructured Data (e.g. text, images)

Data collection usually done by importing files (via upload or url) and scraping from the web.

Data cleansing is to make sure the data is ready to be used for analysis.

Data Exploration

Now that we have clean data, figure out the business question that needs to be answered and then convert that question into a data science question.

For that, we need to explore the data to identify their properties and find the right treatments to provide.

Here, we can uncover initial patterns, characteristics, and points of interest from the data.

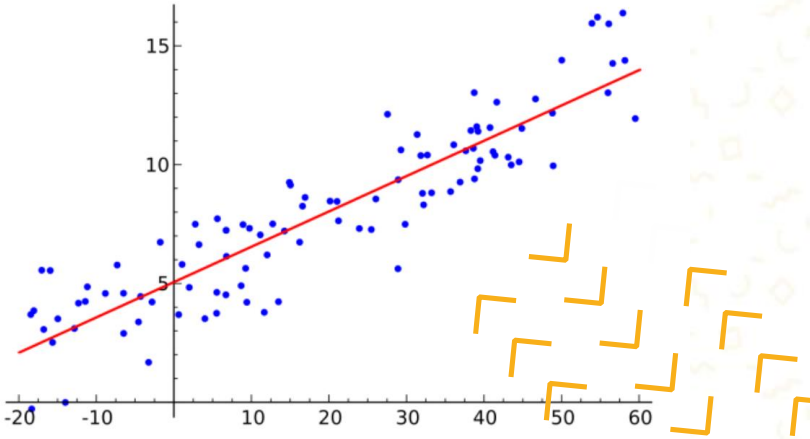
python™

Data Modelling

This is a very crucial phase in the data science process.

Python has many advanced libraries to help us tap the power of machine learning in performing the tasks involved in data modelling.

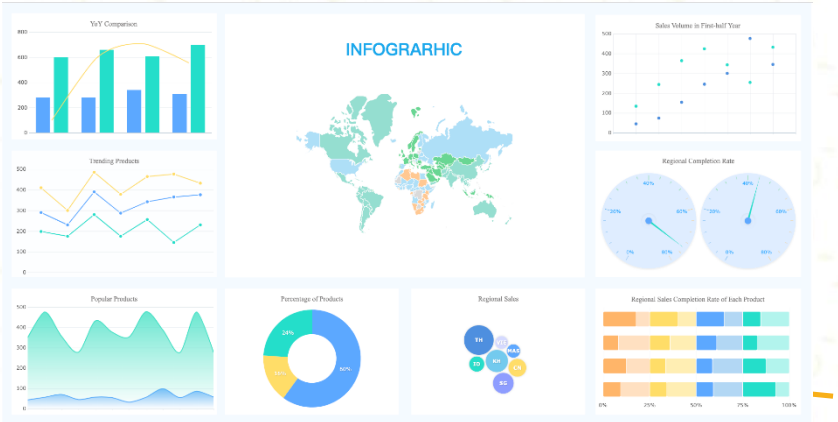
One of the most popular machine learning library in Python is scikit-learn



Data Visualization

In this process, we present the data graphically to show information hidden in the data directly to people.

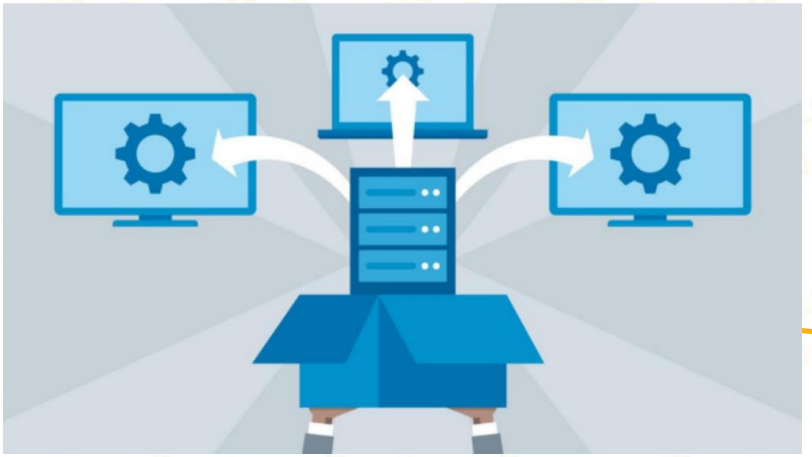
Python has many data visualization packages. **Matplotlib** is the most used library among them for generating basic graphs and charts.



Deploying Model

Deployment of machine learning models or putting models into production means making your models available to the end users or systems.

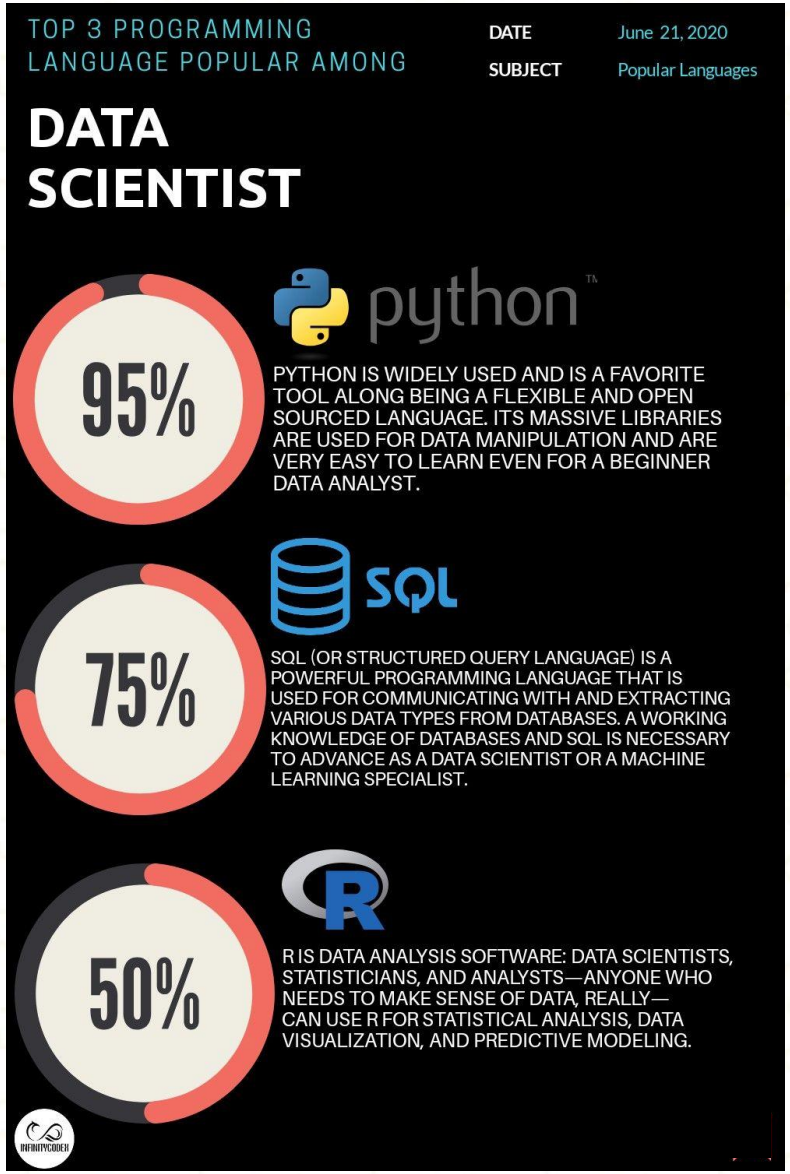
One of the framework that often used for this purpose is **Flask**.

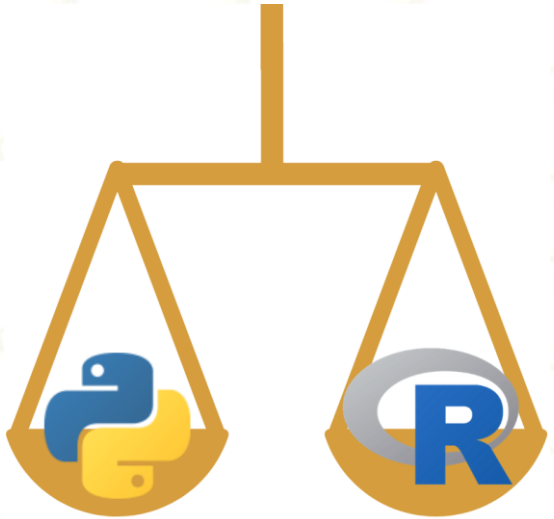


DigitalSkola
Uncover The World Of Digital Skills With Us

Python vs R

Analysis Tool	Similar Superhero	Super Powers in Common
R 	Batman 	• Detective Work • Intelligence • Cunning • Usage of Tools • More Brain than Muscles
Python 	Superman 	• Muscle Power • Super Strength • Elegance • Wide Range • More Muscles than Brain

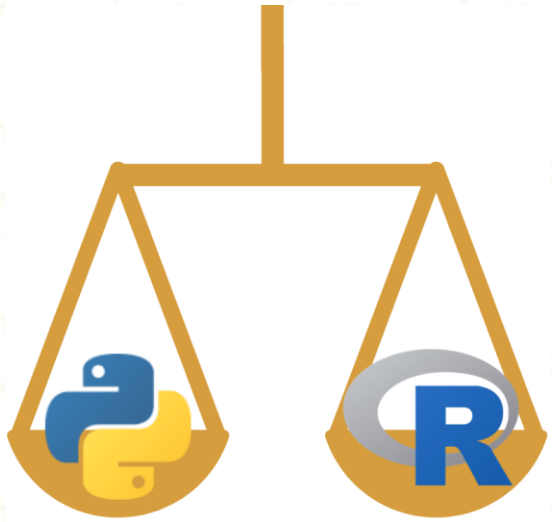




Python vs R

Parameter	Python	R
Objective	Data Science, Web Dev	Data Analysis & Statistical Modeling
Usability	Mostly used on tech side that needs to deploy model	Mostly used on analysis side, academic research, and consulting company
Ease of Learning	Learning curve is relatively linear and smooth	Easier to start, but as it goes advanced more difficult to develop expertise
IDE	Jupyter Notebook, Spyder, PyCharm	RStudio

Python vs R



As we can see, each programming language have their own advantages, and in general both of them can do all of Data Science workflow.

So instead of make them as a rival it will be very worth if we can have an expertise in both of them.

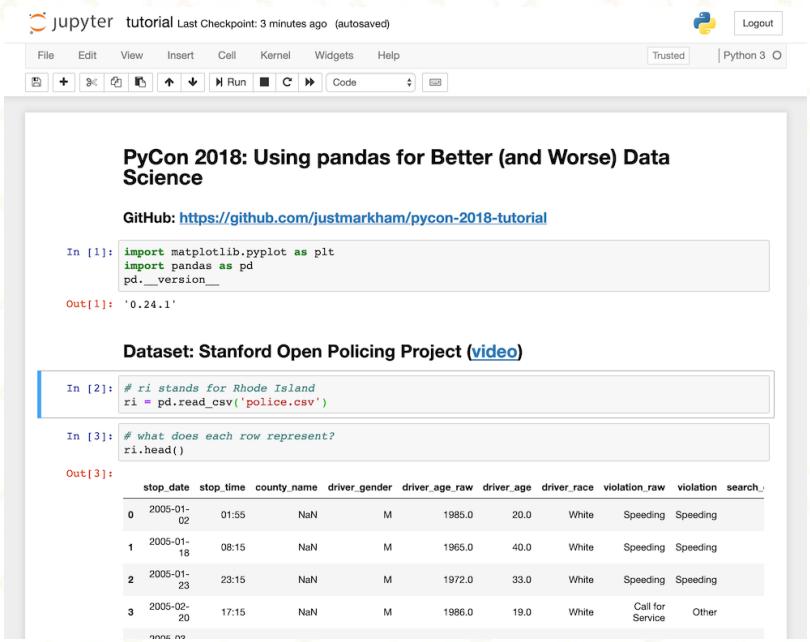
Basically, if you just want to do analysis it doesn't matter which one you use, you just need to deliver the insights.

While if you need to go into production, most company prefers Python.

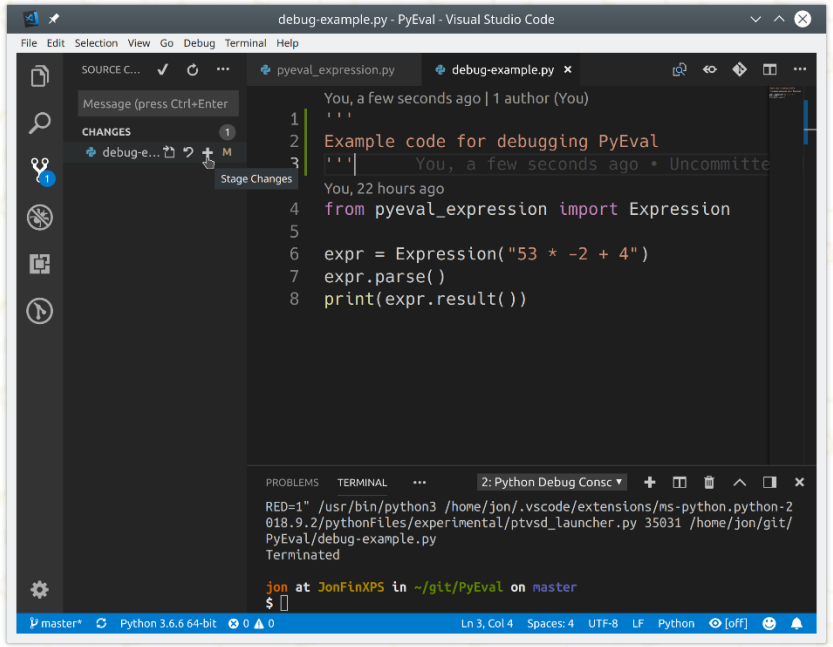
DigitalSkola
Uncover The World Of Digital Skills With Us

Getting Started with Python

Notebook vs IDE



Notebook



IDE

Notebook vs IDE

Notebook	IDE
Easier to explore data	Preferrable for production level
Harder to debug (need to re-run all previous cell)	Easier to debug
Better for documentation	Not good for documentation
File format in .ipynb (can be convert into .py too)	File format in .py

DigitalSkola
Uncover The World Of Digital Skills With Us

Google Colab

https://colab.research.google.com/

(Official Intro : https://www.youtube.com/watch?v=inN8seMm7UI)

About Google Colab

- Basically Notebook environment that placed in cloud and requires no setup to use

- Free cloud service and supports free GPU (GPU is really help especially when running Deep Learning Model)

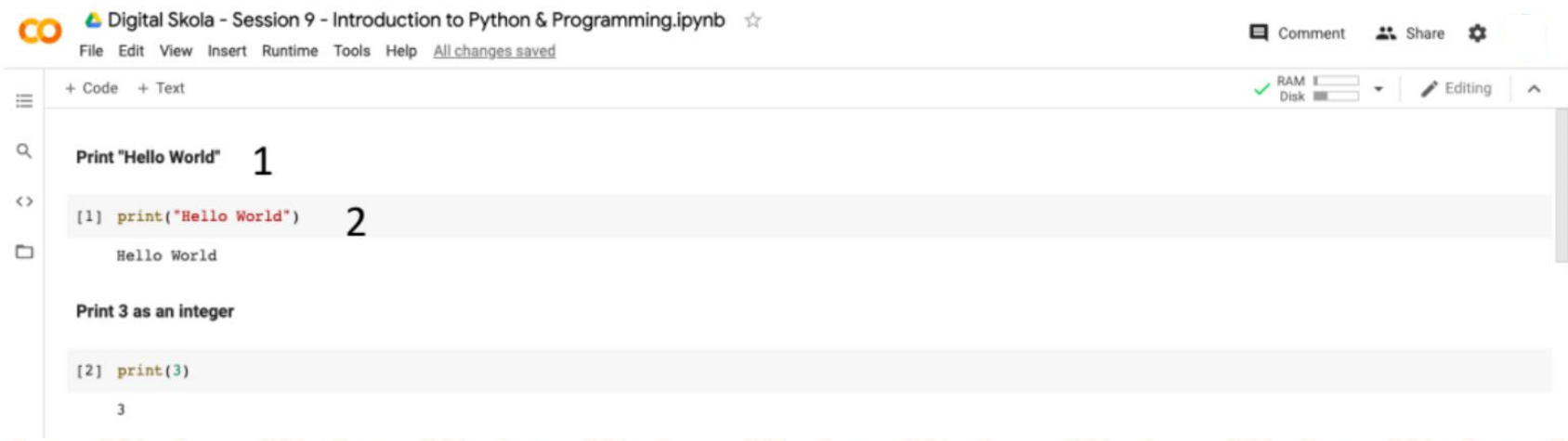
- Can connect to Google Drive and Github

- Can be shared to your teammates (Data Scientist or Data Analyst), so you can work in the same file with your teammates

Google Colab

There are 2 parts in Google Colab :

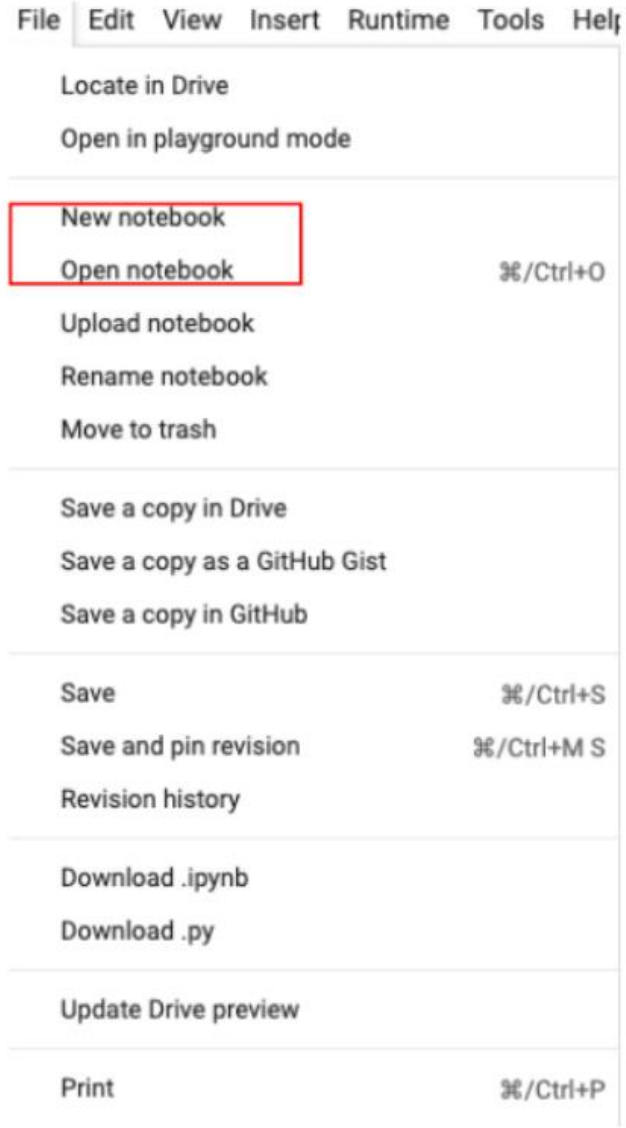
1. Text cell
2. Code cell



DigitalSkola
Uncover The World Of Digital Skills With Us

Google Colab

Create a new or open existing notebook



Google Colab

Basic syntax :

- **!pip freeze** : To check what libraries are already available
- **!pip install <library>** : To install library

Data Types

- Text Type : **str**
- Numeric Types : **int, float, complex**
- Sequence Type : **list, tuple, ranges**
- Mapping Type : **dict**
- Set Types : **set, frozenset**
- Boolean Type : **bool**
- Binary Types : **bytes, bytearray, memoryview**

Our First Code

Type :

print(“Hello World”)

Variable

a place to save some value

Example :

a = 5

That means a holds value 5.

We can update the value of ”a” by type :

a = <new value>

Variable

Variable can also hold the result of another calculation.

Example :

a = 10

b = 5

c = a + b

That means c will holds value of 15

Arithmetic Programming

Assume we have variable a holds 15 and variable b holds 2 (a = 15 , b = 2)

Operation	Description	Example
Addition (+)	Adds values on either side of the operator	a + b = 17
Substraction (-)	Subtract right hand operant from the left hand operant	a – b = 13
Multiplication (*)	Multiplies values on either side of the operator	a * b = 30
Division (/)	Divides left hand operant from the right hand operant	a / b = 7.5
Modulus (%)	Divides left hand operand by right hand operand and returns remainder	a % b = 1
Exponent (**)	Performs exponential (power) calculation on operators	a ** b = 225

DigitalSkola
Uncover The World Of Digital Skills With Us

Google Colab

https://colab.research.google.com/

DigitalSkola

Uncover The World Of Digital Skills With Us

Thank You