

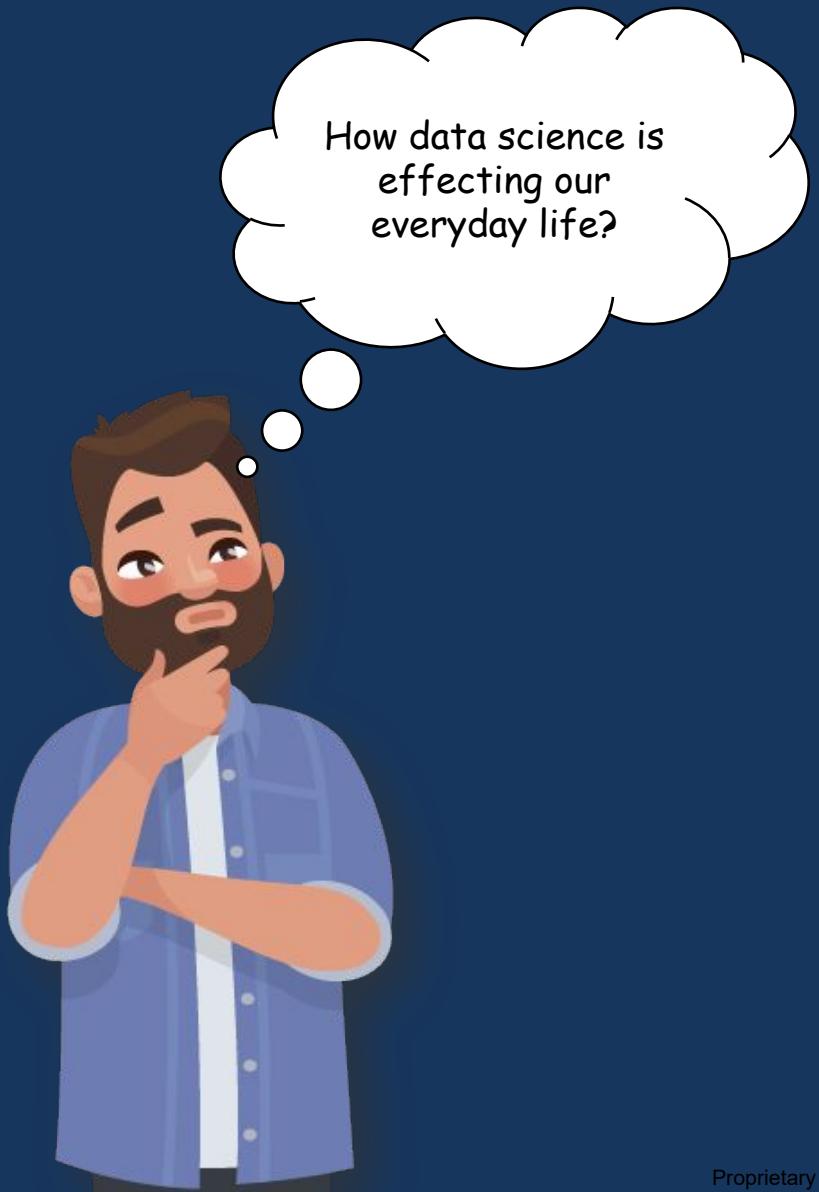
## Credit Card Fraud Analysis Using Data Science

**Sampriti Chatterjee (Great Learning)**

# Agenda

- |   |  |    |  |
|---|--|----|--|
| 1 | Why do we need data science?             | 7  | Basic Syntax for Python                  |
| 2 | What is Data science?                    | 8  | List,Set,Tuple                           |
| 3 | Life cycle of Data science               | 9  | Data manipulation using Numpy and Pandas |
| 4 | Why Python is so popular?                | 10 | What is machine Learning?                |
| 5 | Install python                           | 11 | Supervised Learning: Logistic Regression |
| 6 | Statistical visualization on Python user | 12 | Credit card Fraud analysis using Python  |

# Why do we need Data Science?



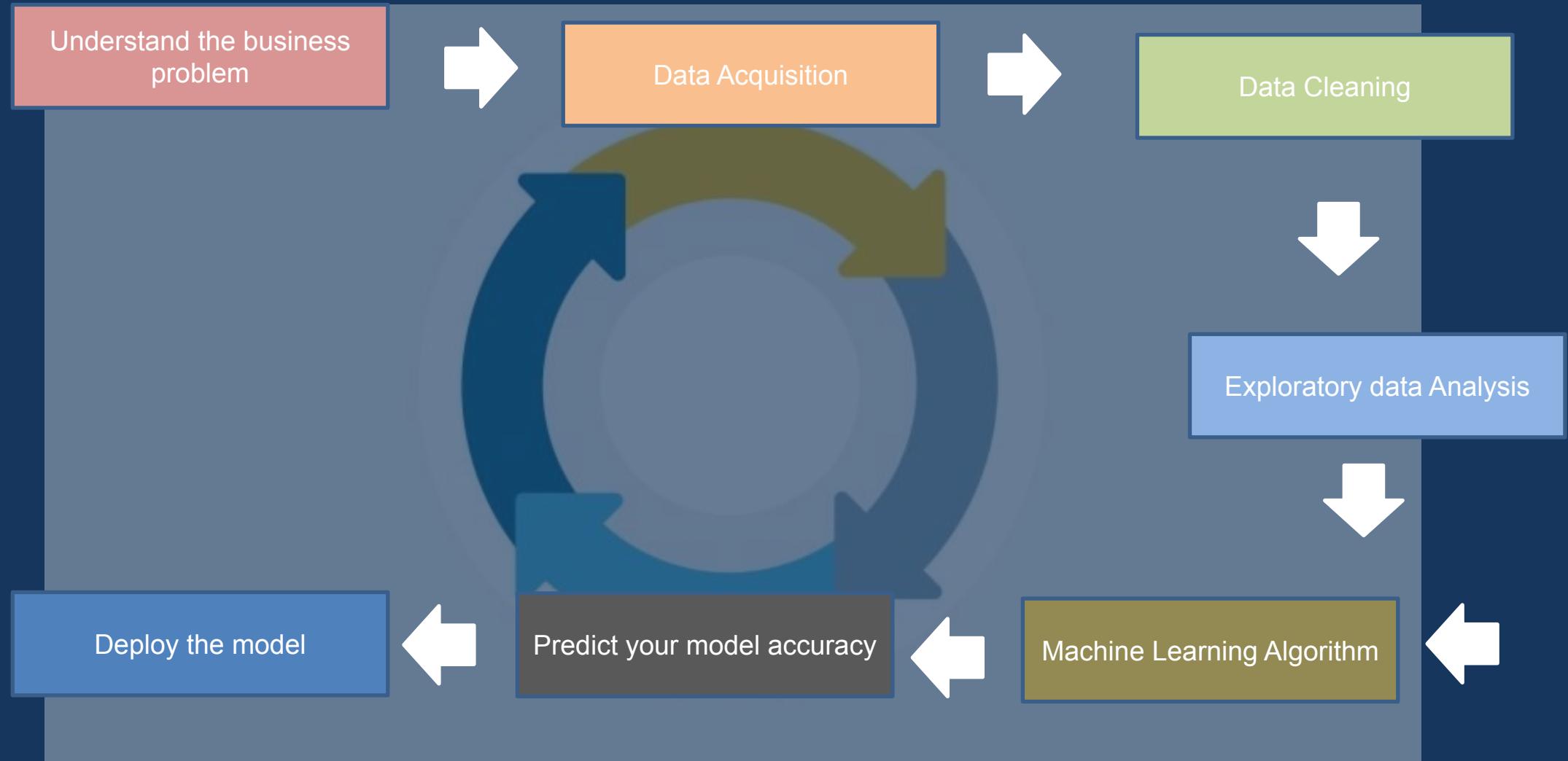
- In the past, we used to have data in a structured format but now as the volume of the data is increasing, so the number of structured data becomes very less, so to handle the massive amount of data we need data science techniques
- Those data can be used to get the proper business insights and the hidden trends from them.
- These insights helps the organization to predict the Future
- Using data science decision making can be faster and effective
- Helps to reduce the production cost
- Build model based on the data to give the ability to the machine to predicts on its own

# What is Data Science?

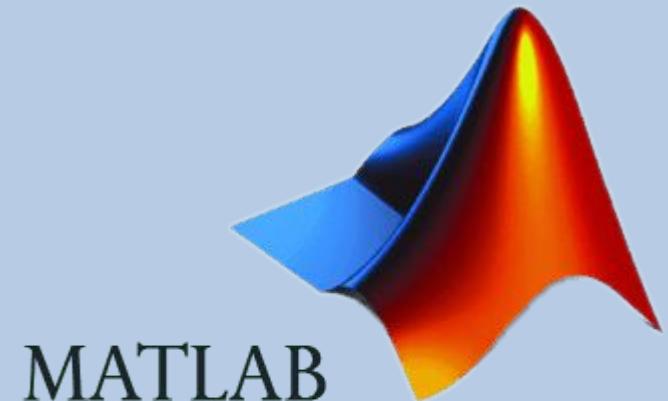
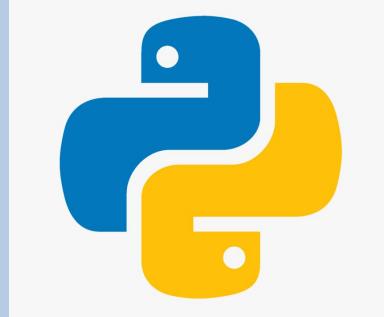


Data science is a process to get some meaningful information from the massive amount of data. In simple terms, read and study the data to get proper intuitive insights. Data Science is a mixture of various tools, algorithms, and machine learning and deep learning concepts to discover hidden patterns from the raw and unstructured data

# Life cycle of Data Science?



# Most Popular Programming Languages For Data Science?

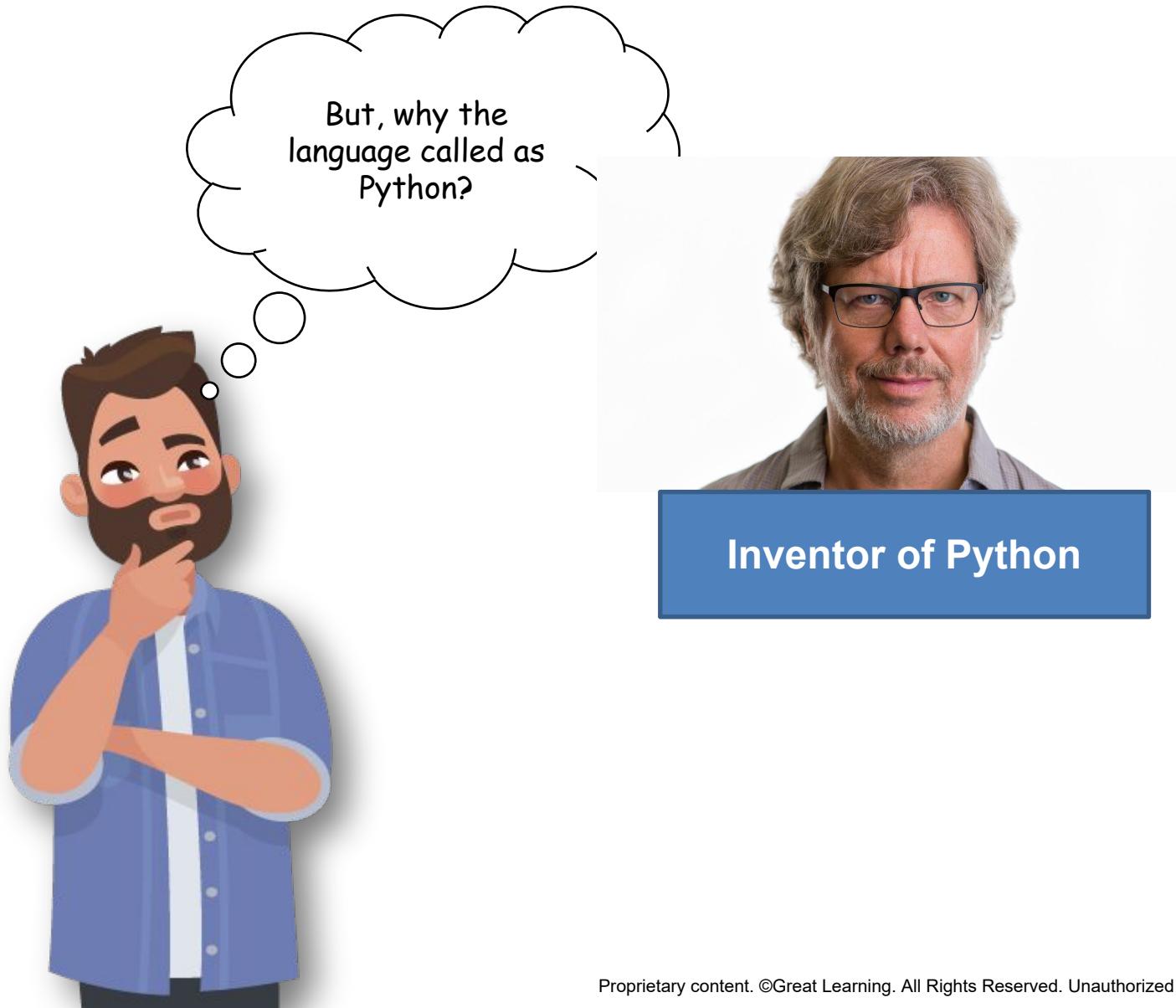


**Python is a popular high level, object oriented and interpreted language**

High level

Interpreted

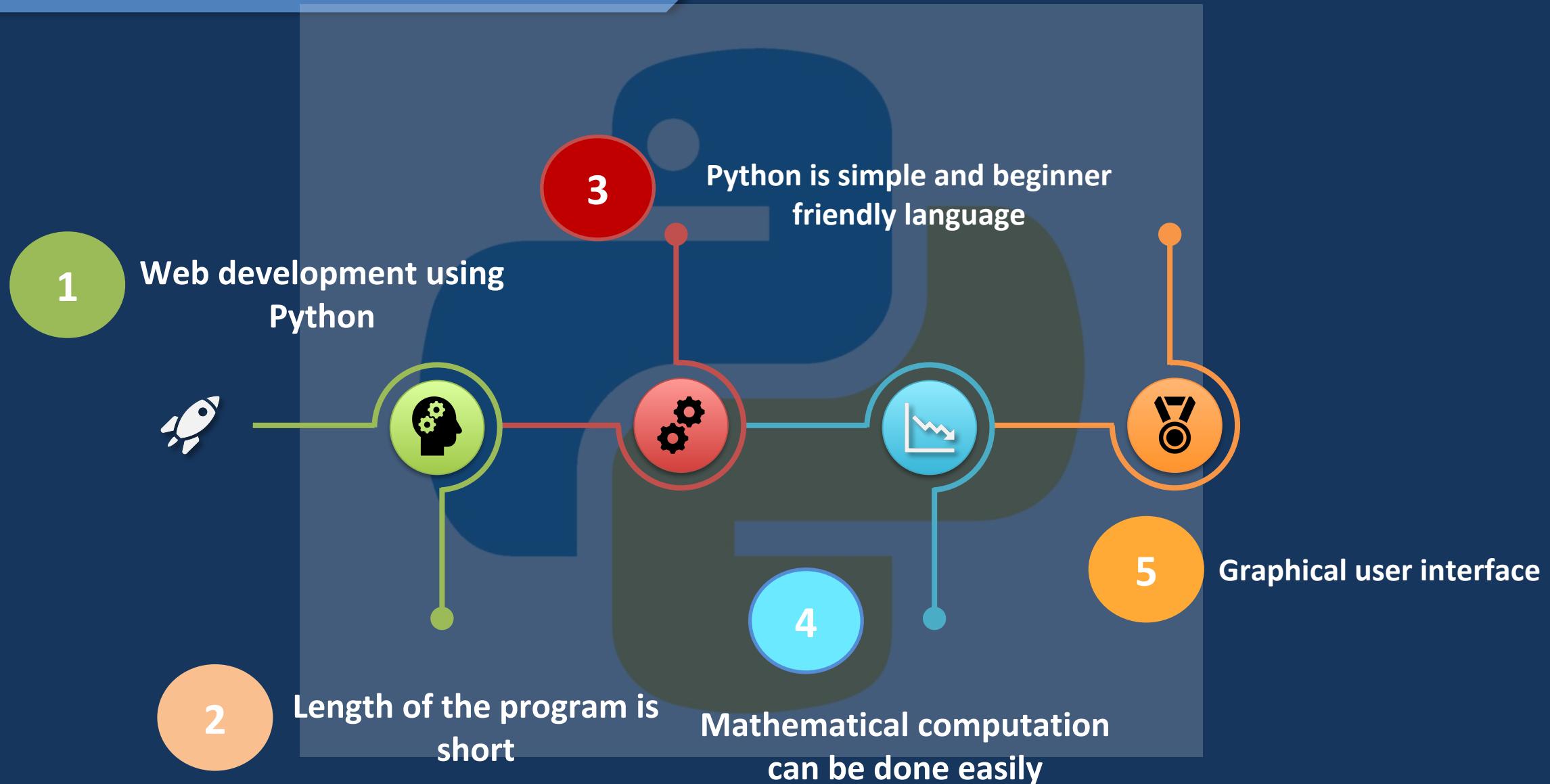
Object oriented



## Important Facts

- Python is invented by Guido van Rossum in 1989
- Rossum used to love watching comedy movies from late seventies
- He needed a short, unique, and slightly mysterious name for his language
- In that time he was watching Monty Python's Flying Circus and from that series he decided to keep his language name python.
- This how Python invented

# Why should you learn Python?



# Why Python is so popular?

1

Largest community for Learners and Collaborators

2

Open source

3

Easy to learn and usable flexibility

4

Huge numbers of Python libraries and Frame work

5

Supports Big Data, Machine Learning and Cloud computing

6

Supports Automation

# Installing Python

This is the site to install Python ->

The screenshot shows the Python.org homepage with a focus on the download section for Mac OS X. The top navigation bar includes links for About, Downloads, Documentation, Community, Success Stories, News, and Events. A prominent yellow button labeled "Download Python 3.8.3" is visible. Below it, text links provide options for Windows, Linux/UNIX, Mac OS X, and Other operating systems. To the right, there's a graphic of two packages descending from the sky on parachutes. The bottom of the page features a footer with a blue bar.

python™

Donate Search GO Socialize

About Downloads Documentation Community Success Stories News Events

**Download the latest version for Mac OS X**

[Download Python 3.8.3](#)

Looking for Python with a different OS? Python for [Windows](#), [Linux/UNIX](#), [Mac OS X](#), [Other](#)

Want to help test development versions of Python? [Prereleases](#), [Docker images](#)

Looking for Python 2.7? See below for specific releases

# Popular IDE for Python: Pycharm

Site to install Python ->

**PyCharm**

Coming in 2020.2 What's New Features Learning Center Buy [Download](#)



Version: 2020.1.2  
Build: 201.7846.77  
3 June 2020

[System requirements](#) [Installation Instructions](#) [Other versions](#)

## Download PyCharm

Windows Mac Linux

**Professional**  
For both Scientific and Web Python development. With HTML, JS, and SQL support.

[Download](#) [Free trial](#)

**Community**  
For pure Python development

[Download](#) [Free, open-source](#)

 Get the Toolbox App to download PyCharm and its future updates with ease

[X] Cookies and IP addresses allow us to deliver and improve our web content and to provide you with a personalized experience. Our website uses cookies and collects your IP address for these purposes.  
=====  
| JetBrains may use cookies and my IP address to collect individual statistics and to provide me with personalized offers and ads subject to the [Privacy Policy](#) and the [Terms of Use](#). JetBrains may use [third-party services](#) for this purpose. I can revoke my consent at any time by visiting the [Opt-Out page](#).  
| [Y]es, I agree [N]o, thanks  
=====  
~ root#

Anaconda installation site->



Individual Edition

## Your data science toolkit

With over 20 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for solo practitioners, it is the toolkit that equips you to work with thousands of open-source packages and libraries.

Download

Google collaboratory link->

The screenshot shows the Google Colaboratory interface. On the left, there's a sidebar titled "Table of contents" with sections like "Getting started", "Data science", "Machine learning", "More Resources", "Machine Learning Examples", and "Section". The main content area has a title "What is Colaboratory?" with a sub-section "Getting started". It explains that Colaboratory allows writing and executing Python in a browser with zero configuration, free access to GPUs, and easy sharing. It also mentions that it's suitable for students, data scientists, and AI researchers. A code cell at the bottom shows a Python script to calculate seconds in a day.

Welcome To Colaboratory

File Edit View Insert Runtime Tools Help

Share

Table of contents

Getting started

Data science

Machine learning

More Resources

Machine Learning Examples

Section

+ Code + Text Copy to Drive

Connect Editing

What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started

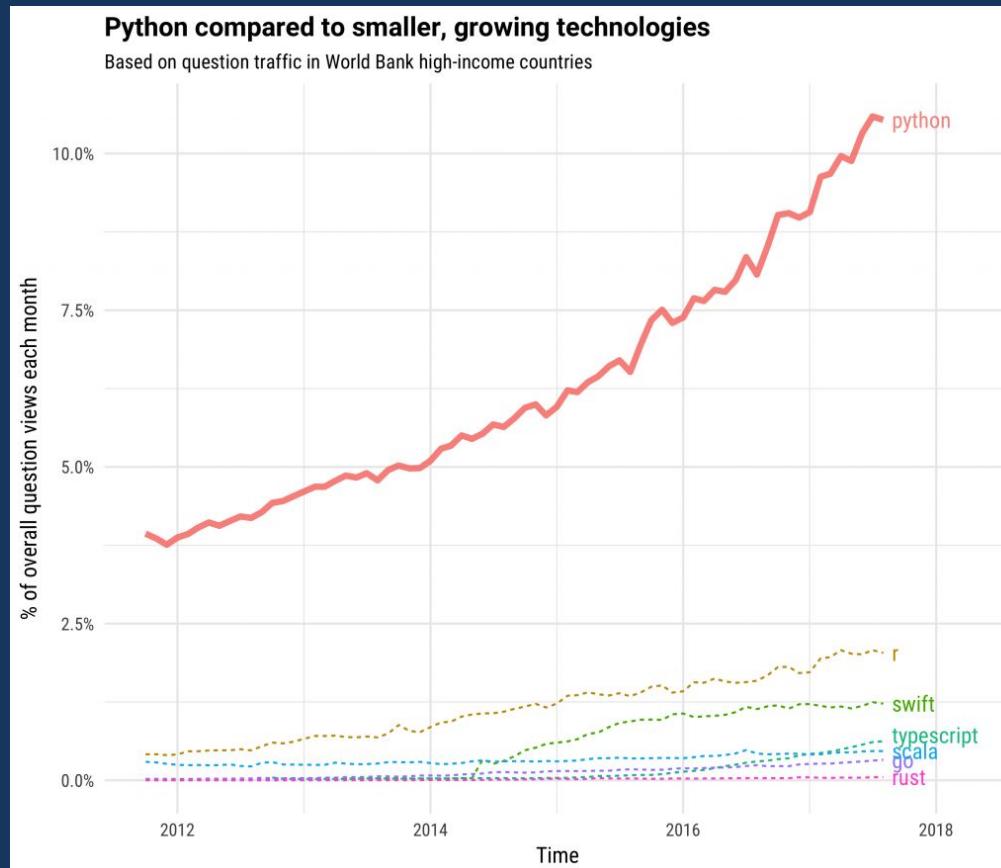
The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

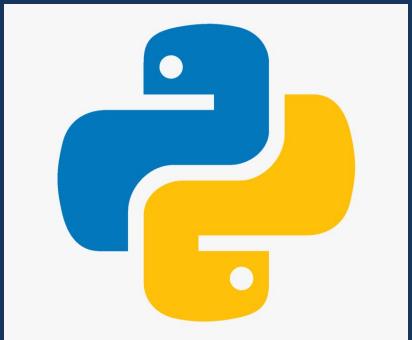
```
[ ] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day
```

86400

# Statistical measurement on Python user



In recent time it is prominent that Python is one of the most popular language because of it's simplicity



Getting started with Python

# How to Store Data?



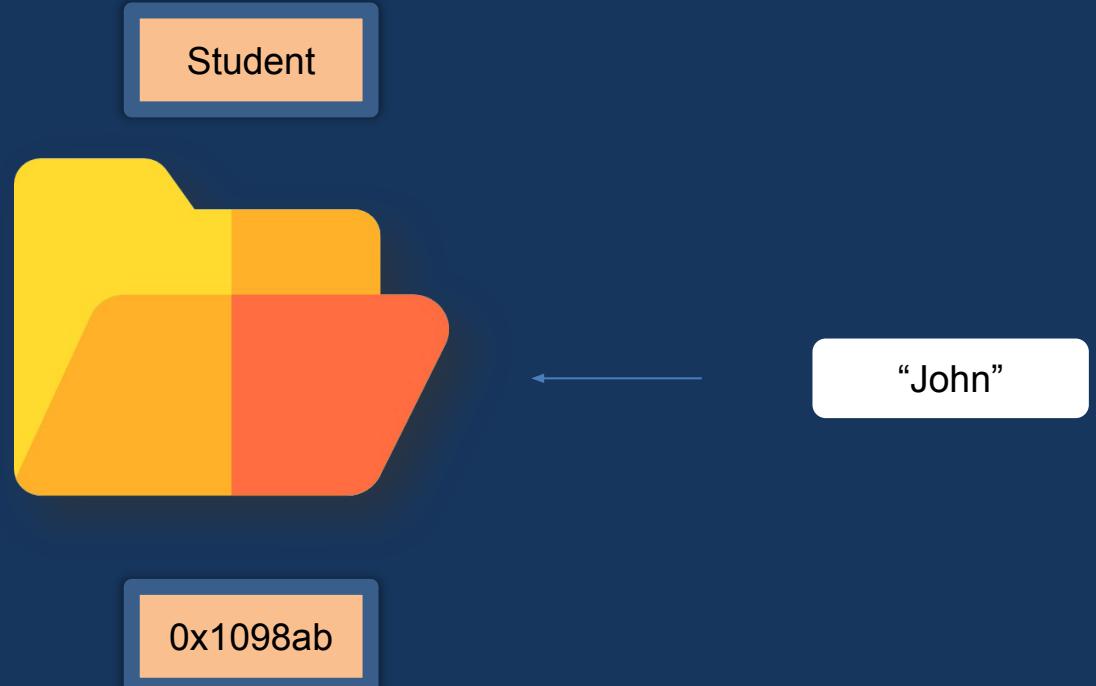
“John”

123

TRUE

# Need of Variables

Data/Values can be stored in temporary storage spaces called variables



## Variable

```
a = 5  
b= 0.4  
c = "Riya"  
print('Value of the variable: ',a)  
print('Value of the variable: ',b)  
print('Value of the variable: ',c)
```

```
Value of the variable: 5  
Value of the variable: 0.4  
Value of the variable: Riya
```

1

Variables are used as containers to store the data values

2

Python has no command to declare a variable

3

A variable is created when you assigned a value to it

# DataTypes in Python

Every variable is associated with a data-type

10, 500

int

3.14, 15.97

float

TRUE, FALSE

Boolean

“Sam”, “Matt”

String

## Datatypes

```
a = 5  
b= 0.4  
c = "Riya"  
print('Type of the variable: ',type(a))  
print('Type of the variable: ',type(b))  
print('Type of the variable: ',type(c))
```

```
Type of the variable: <class 'int'>  
Type of the variable: <class 'float'>  
Type of the variable: <class 'str'>
```

1

Data types are used to classify or categorize of data items

2

Data types represent value which determines what operations can be performed on that data

3

In python data types is assigned when you assign the value to the variable

4

We can check the data type of the variable by calling the type() function

## Basic syntax for Python

### Python string

1

Single line string:

```
data='Hello welcome to Great Learning!'
print(data)
```

```
Hello welcome to Great Learning!
```

2

Multiline string:

```
data = '''Welcome to Great Learning!,
In today's session we will discuss about python project for beginners.'''
print(data)
```

```
Welcome to Great Learning!,
In today's session we will discuss about python project for beginners.
```

3

String as arrays:

```
data='Hello welcome to Great Learning!'
for i in range(len(data)):
    print(data[i])
```

```
H
e
l
l
o
```

4

String slicing:

```
b = "Hello, World!"  
print('String indexing starts from 0: ', b[1:4])
```

```
String indexing starts from 0: ell
```

5

String negative indexing:

```
b = "Hello, World!"  
print('String negative indexing starts from -1: ',b[-6:-1])
```

```
String negative indexing starts from -1: World
```

6

Length of the String:

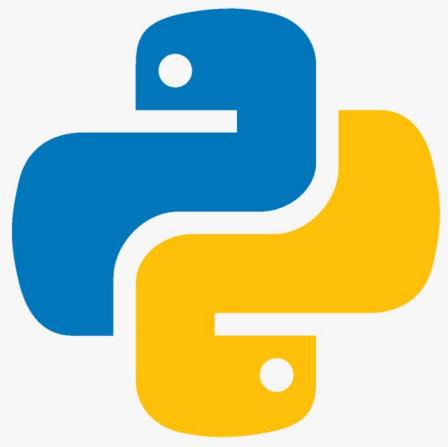
```
print('length of the string: ',len(b))
```

```
length of the string: 13
```

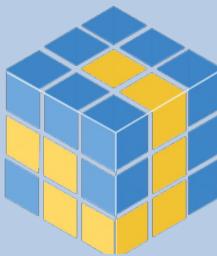


# Python Libraries

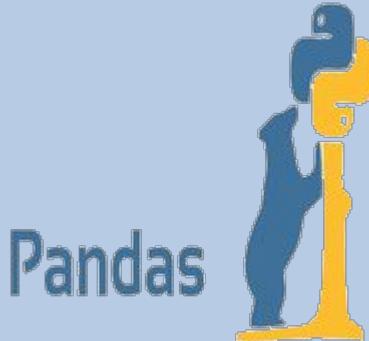
**Data manipulation is a technique which allows to transform, extract, and filter your data efficiently with less time.**



Main two python libraries are used to manipulate the data



**NumPy**



**Pandas**

**Numpy stands for Numerical Python and it is used to perform mathematical and logical operations on arrays**

1 Numpy is a python library

2 Install Numpy: !pip install numpy

3 Import the Library: import numpy as np



# NumPy

**Pandas is a popular data manipulation and analysis library in python which is based on Numpy**

1

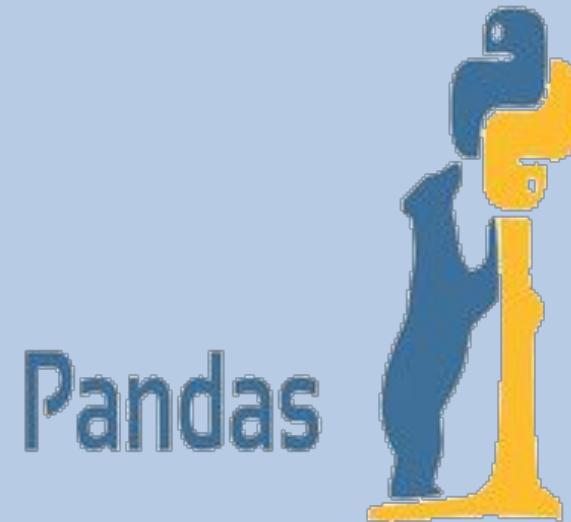
Python is a python library built on top of Numpy

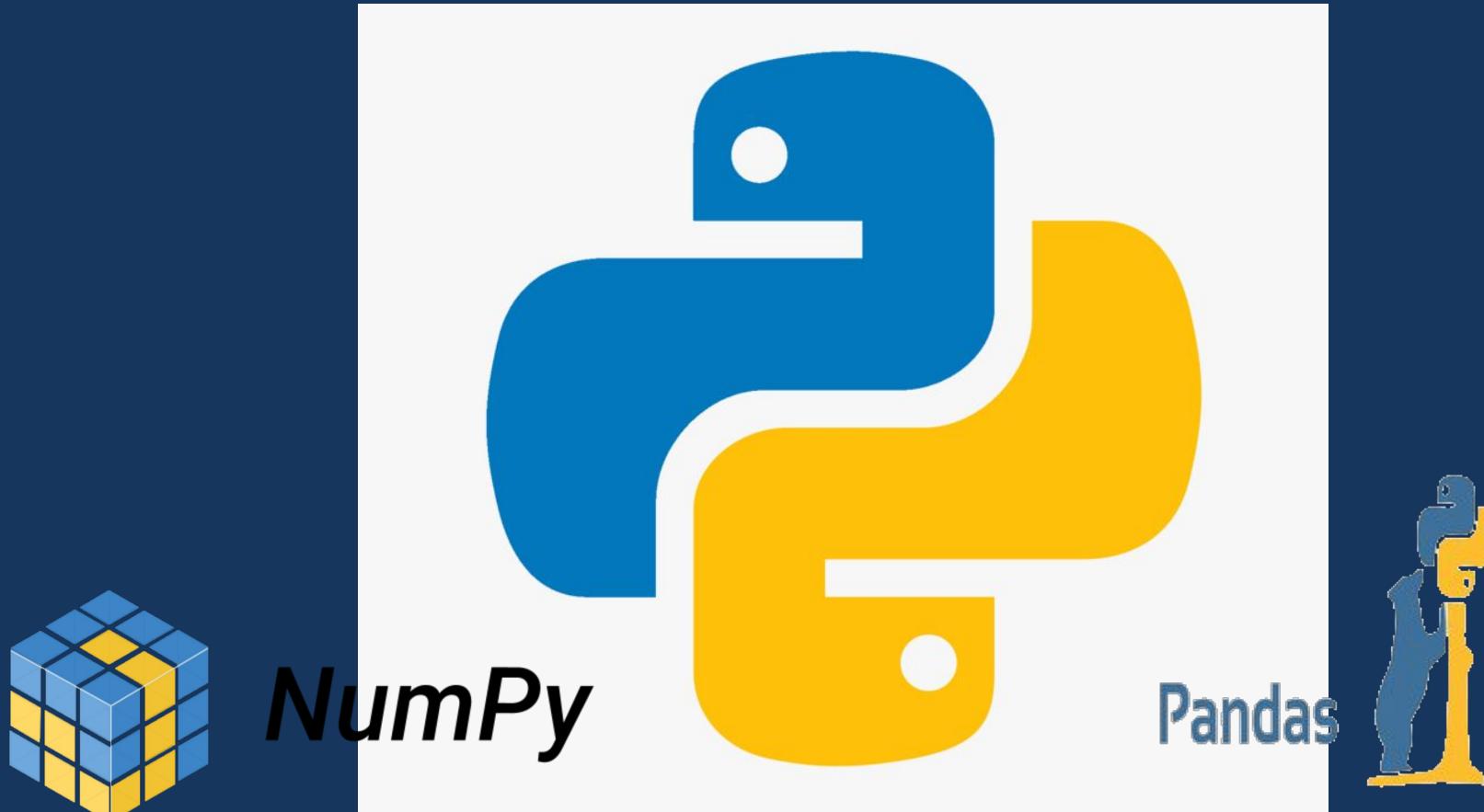
2

Install Numpy: !pip install pandas

3

Import the Library: import pandas as pd





## Demo on Numpy and pandas

## Creating a Dataframe

Panda's data frame is a two-dimensional data structure which is aligned in a tabular fashion with rows and columns



```
In [9]: import pandas as pd  
  
pd.DataFrame({"Name":['Bob', 'Sam', 'Anne'], "Marks": [76, 25, 92]})
```

Out[9]:

	Name	Marks
0	Bob	76
1	Sam	25
2	Anne	92

head()

shape()

describe()

tail()



## Dropping Columns

```
iris.drop('Sepal.Length', axis=1)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa



	Sepal.Width	Petal.Length	Petal.Width	Species
0	3.5	1.4	0.2	setosa
1	3.0	1.4	0.2	setosa
2	3.2	1.3	0.2	setosa
3	3.1	1.5	0.2	setosa
4	3.6	1.4	0.2	setosa

## Dropping Rows

```
iris.drop([1,2,3],axis=0)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa



	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	5.1	3.5	1.4	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa



Machine Learning to build the Model

# What is Machine Learning?

**Machine learning is a sub-set of artificial intelligence (AI) that allows the system to automatically learn and improve from experience without being explicitly programmed**

	Time	V1	V2	V3	V4	V5
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193



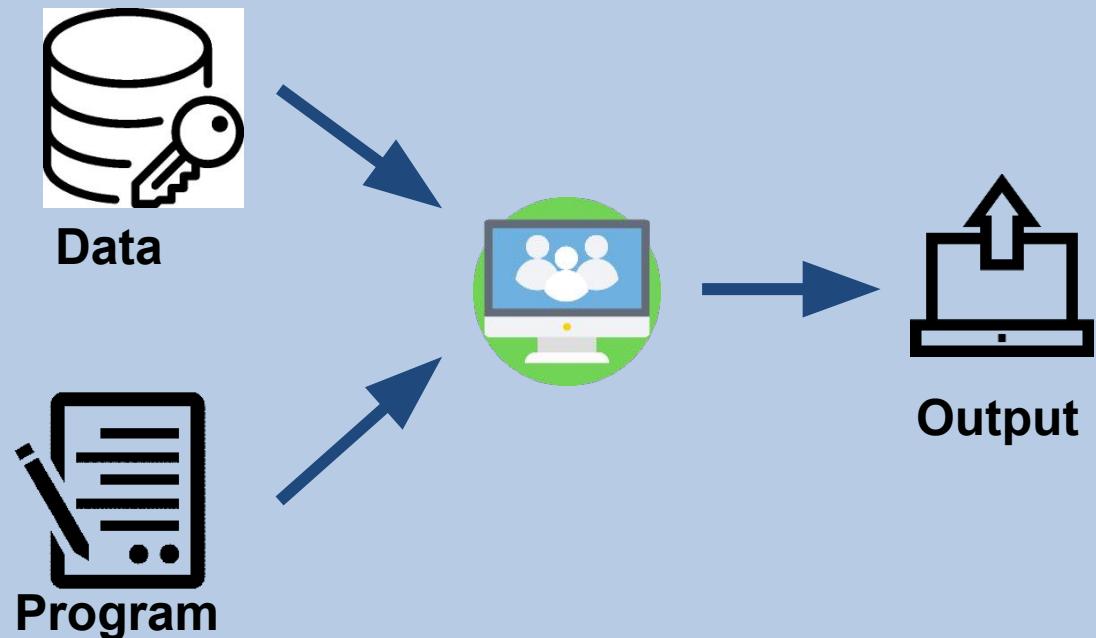
	Time	V1	V2	V3	V4
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828
284805	172788.0	-0.240440	0.530483	0.702510	0.689799
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271

Training Data

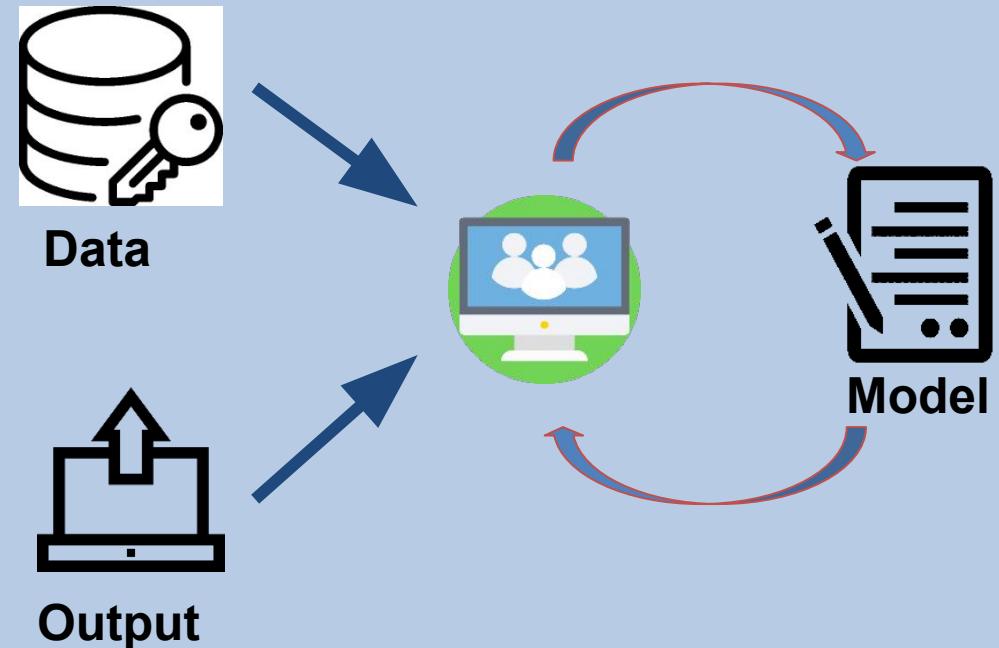
Model Building

Testing Data

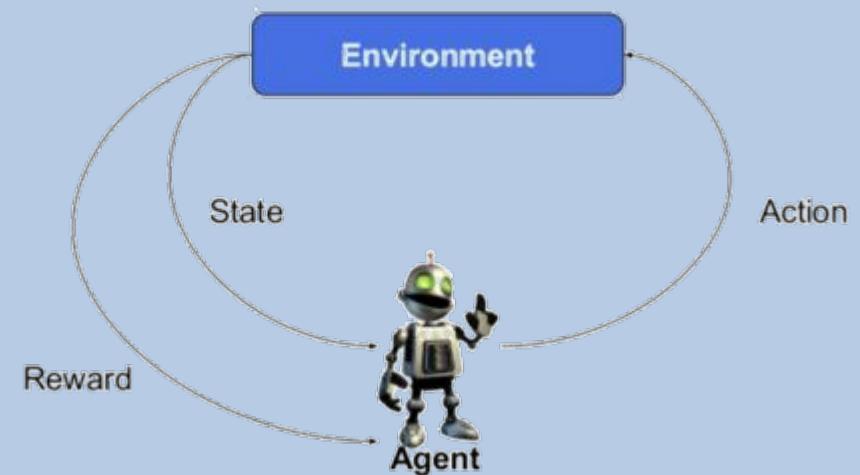
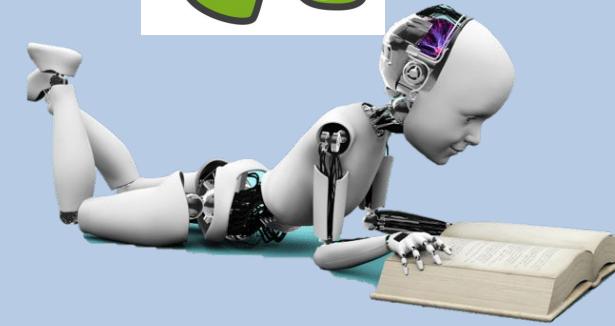
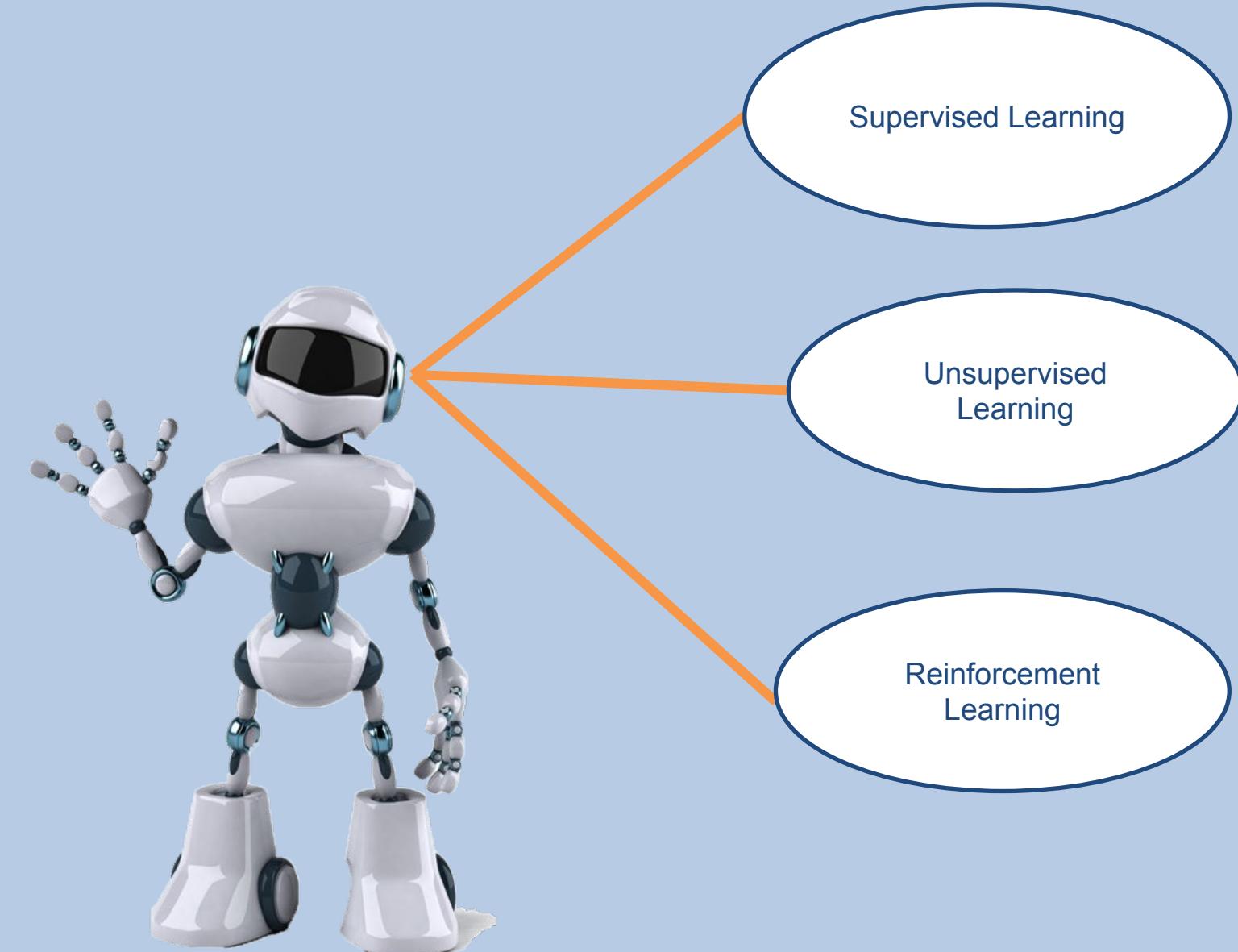
## Traditional Programming



## Machine Learning



# Types Of Machine Learning



# What is Supervised Learning?

Supervised learning works as a supervisor or teacher. Basically, In supervised learning, we teach or train the machine with labeled data (that means data is already tagged with some predefined class). Then we test our model with some unknown new set of data and predict the level for them

- Learning from the labelled data and applying the knowledge to predict the label of the new data(test data), is known as ***Supervised Learning***
- ***Types of Supervised Learning:***
  - **Linear Regression**
  - **Logistic regression**
  - **Decision Tree**
  - **Random Forest**
  - **Naïve Bayes Classifier**



## What is Logistic Regression?

Logistic regression is also a part of supervised learning classification algorithm. It is used to predict the probability of a target variable and the nature of target or dependent variable is discrete, so for the output there will be only two class will be present

- The dependent variable is binary in nature so that can be either 1 (stands for success/yes) or 0 (stands for failure/no).
- Logistic regression is also known as sigmoid function
- *Sigmoid function =  $1 / (1 + e^{-value})$*





# Credit Card Fraud Analysis Using Python

# Thank You