MODULE 01

DATA SCIENCE BOOTCAMP

Introduction to Python and Data Science

Vertical Institute

Reminder

1. Please turn on Zoom camera the whole duration of classes.

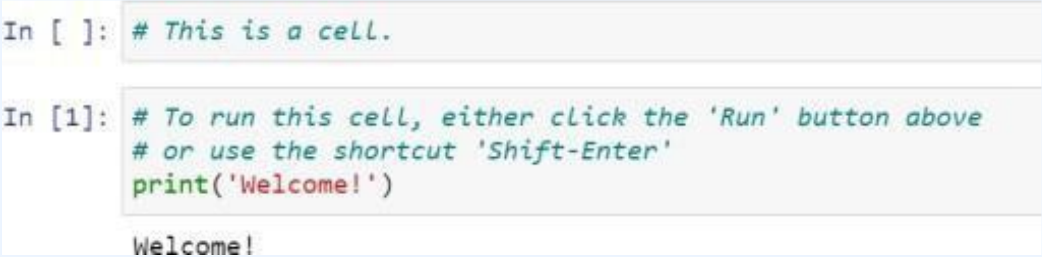
2. At the start of all classes, please rename yourselves to: Name + Last 3 digits and letter of your NRIC. Example: John Tan (123A)

Course Overview

Module 1-2

Starting from ground zero, you will learn fundamental programming concepts by executing basic functions in Python.

You will be able to code comfortably in Python and understand control flow and conditional programming.

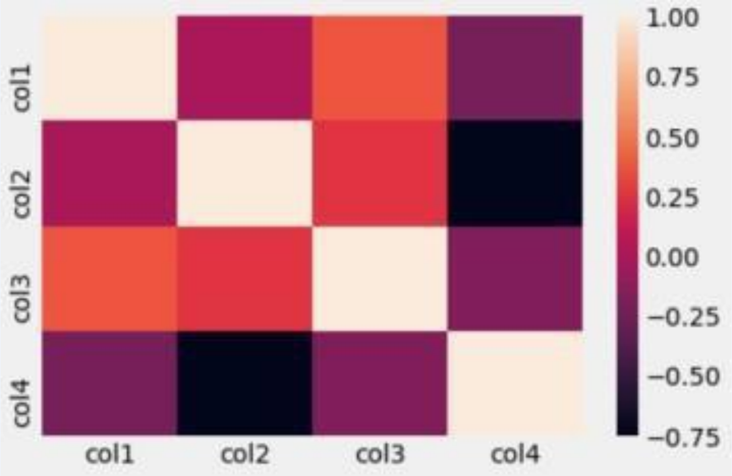


```
In [ ]: # This is a cell.

In [1]: # To run this cell, either click the 'Run' button above
        # or use the shortcut 'Shift-Enter'
        print('Welcome!')

        Welcome!
```

Course Overview

Module 3-4

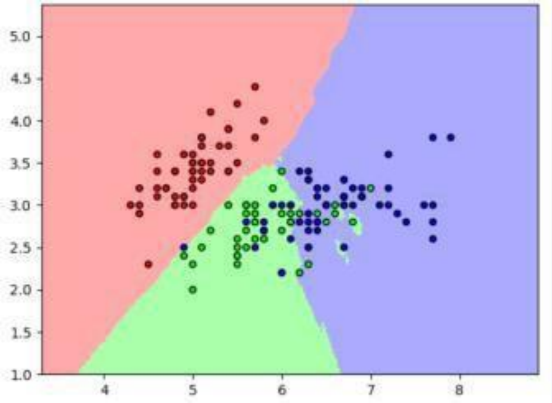
In the next two modules, you will practice exploratory data analysis for cleaning and aggregating data, and understand the basic statistical testing values of your data and more.



Course Overview

Module 5-7

Finally, you will build and refine machine learning models to predict patterns from data sets, tune data parameters for advanced model evaluation.



🧰 Technicalities

1

How is your Wifi connection? Latest version of Zoom is required if you're having trouble with Zoom.

2

Laptop is required. Mobile or tablets screens are too small and the screen will be different from the trainer demo if you use your mobile or tablets.

3

Recording and PDF slides are provided to you in your learners' portal on Vertical Institute's Website.

Things to take note…✍️

To receive the funding support, please take note of the following:

- **Minimum of 75% attendance** (this means that you must attend at least 6 out of 7 lessons.
- **Achieve at least a PASS** for Capstone Project.
- The Capstone Project has to be submitted by the deadline given (1 week from the end of the bootcamp).

In the event that the participant fails to fulfil the above mentioned requirements, the participant is liable for the full amount of the course fee.

What if I missed a lesson?🗓️

If you are missed a lesson, you are allowed to attend make-up class. This will be corresponding class of another intake that same week subjected to slot availability.

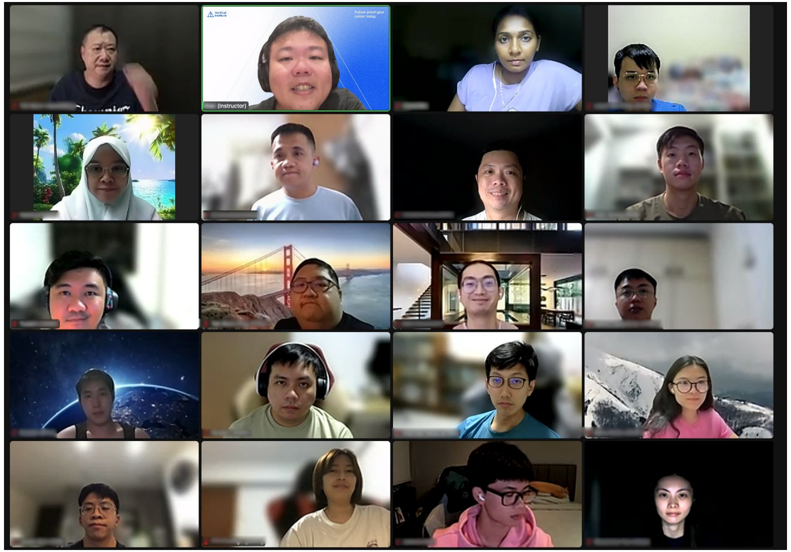
To arrange for a make-up class, you can ***contact the Teaching Assistant or Admin*** and they will be able to arrange for a make-up class nearer to the date.

INTRODUCTION

Hello there! 👋

INTRODUCTION

Instructor Introduction 👋



👋 Now let's meet you!

Now it's your turn to introduce yourself!

- Name
- Occupation/School
- In your spare time, what do you like to do?
- Why do you want to learn Data Science?

Attendance Photo Taking

What is Python and Why Learn it?

Amazon, Google, FB, Netflix – What Do They Have In Common?

FAANG companies love Python and use it for their real-world applications.

What is Python?

- Released in 1991
- Python was created by Guido van
- Rossum
- Flexible
- Easy-to-learn
- Open Source

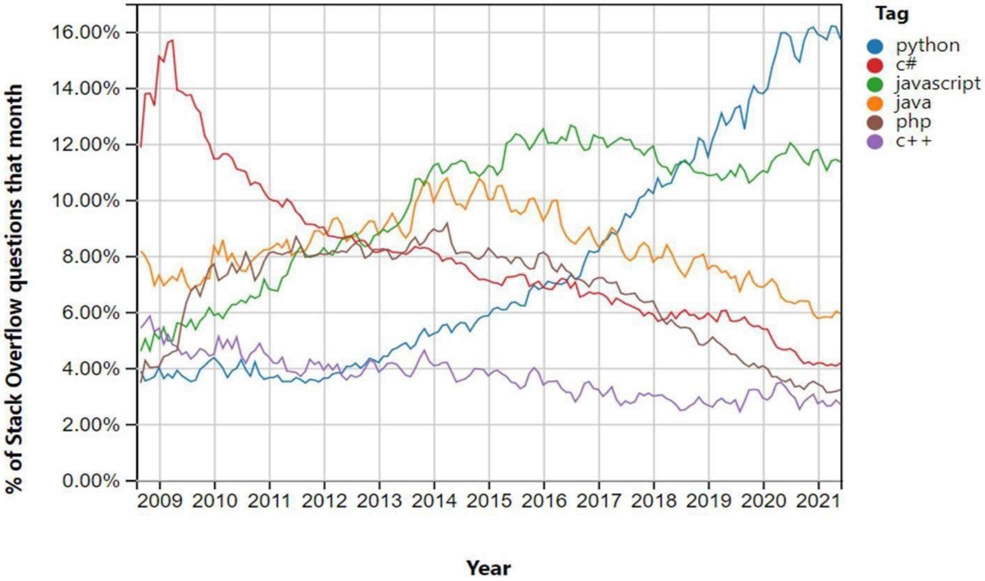
Jupyter (aka. Ipython) Notebook

- Web based application that can
  - Run Python code
  - Contain data
  - Render visualisations
  - Take notes in markdown

www.i-programmer.info/news/216-python/12748-guido-van-rossum-on-python-and-diversity-in-open-source.html

Why Learn Python?

- Named as the most in-demand coding language
- One of the most favorite language used by data scientists
- Being a multi-purpose language, one can build application via python on top of being able to perform data analysis



Jupyter Notebook

What is Jupyter Notebook?

- The Jupyter Notebook is a powerful tool for interactively developing and presenting data science projects.

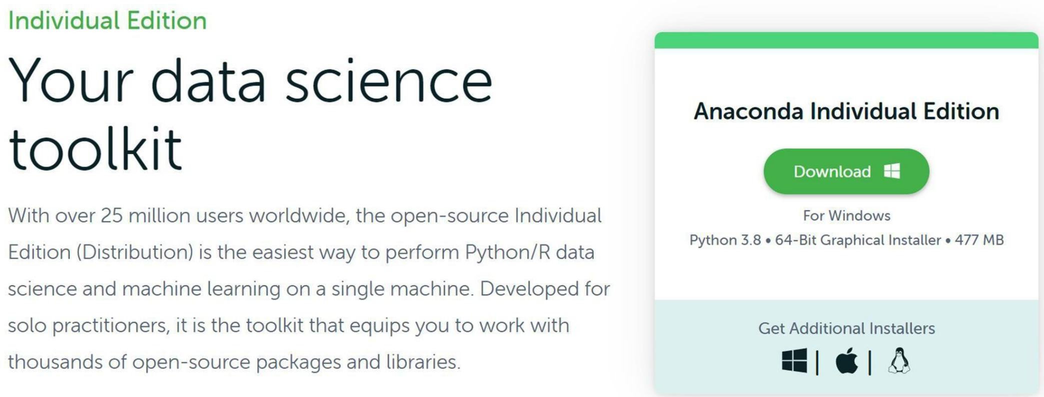
- It's a single document where you can run code, display the output, and also add explanations, formulas, charts, and make your work more transparent, understandable, repeatable, and shareable.

- Free to install and use!

Credit: https://www.dataquest.io/blog/jupyter-notebook-tutorial/

Setting Up Your Jupyter Notebook

Download Anaconda here:
www.anaconda.com/products/individual



Individual Edition

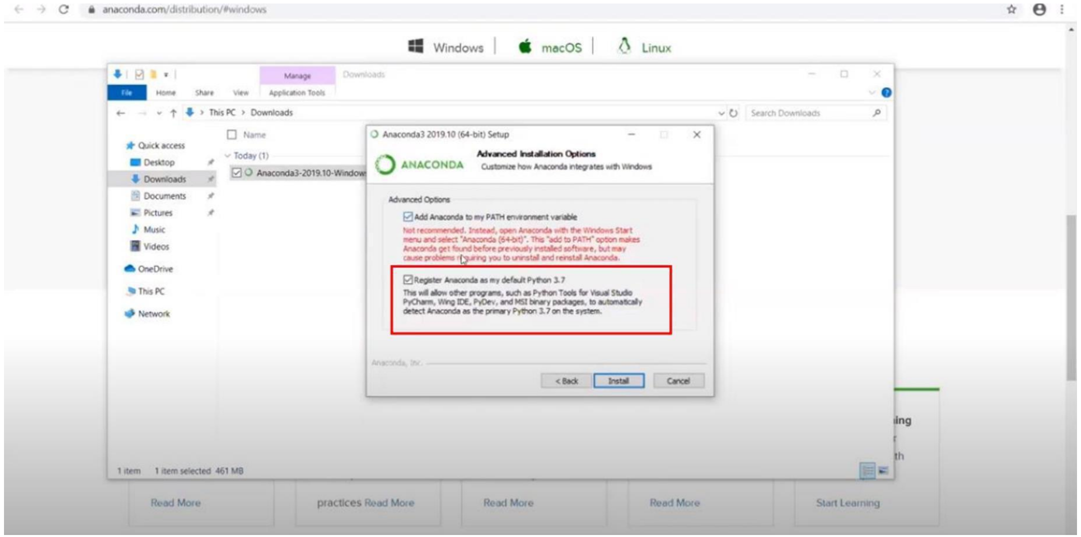
Your data science toolkit

With over 25 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.

Anaconda Individual Edition
Download
For Windows
Python 3.8 • 64-Bit Graphical Installer • 477 MB
Get Additional Installers

Setting Up Your Jupyter Notebook

Launch the installer and follow the recommended settings during the installation.
For Advanced Installation Options, make sure to ‘Register Anaconda as my default Python 3.7



Anaconda3 2019.10 (64-bit) Setup

Advanced Installation Options

Customize how Anaconda integrates with Windows

Advanced Options

☑ Add Anaconda to my PATH environment variable

Not recommended. Instead, open Anaconda with the Windows Start menu and select "Anaconda (64-bit)". This "add to PATH" option makes Anaconda get found before previously installed software, but may cause problems requiring you to uninstall and reinstall Anaconda.

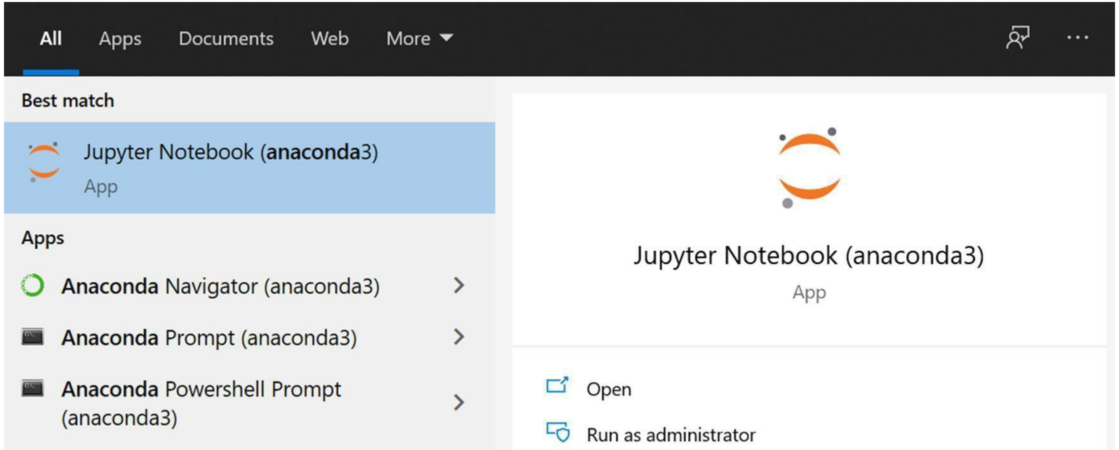
☑ Register Anaconda as my default Python 3.7

This will allow other programs, such as Python Tools for Visual Studio PyCharm, Wing IDE, PyDev, and MSI binary packages, to automatically detect Anaconda as the primary Python 3.7 on the system.

Anaconda, Inc.

Setting Up Your Jupyter Notebook

When the installation is complete, you should be able to access an application called 'Jupyter Notebook'.

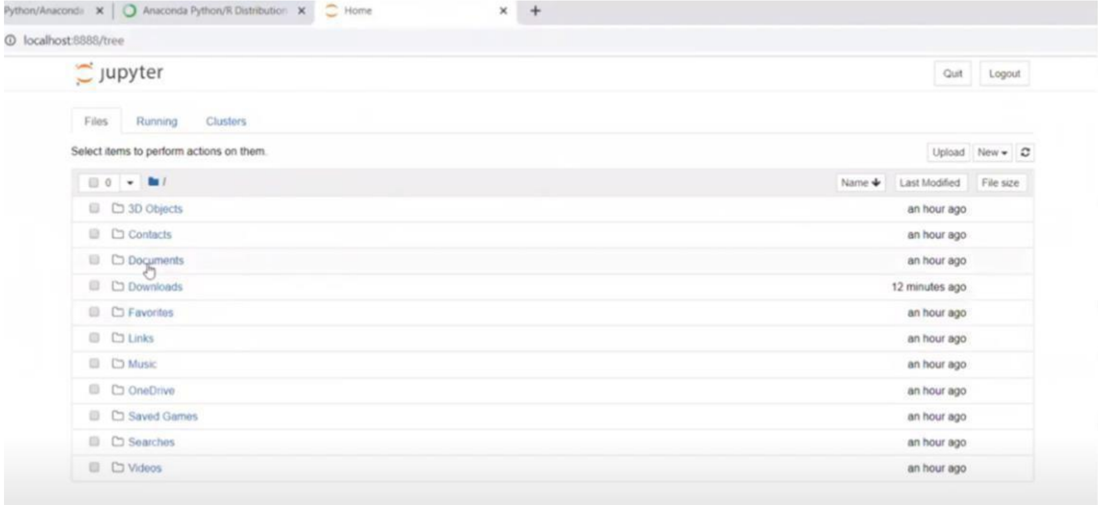


All
Apps
Documents
Web
More
Best match
Jupyter Notebook (anaconda3)
App
Apps
Anaconda Navigator (anaconda3)
Anaconda Prompt (anaconda3)
Anaconda Powershell Prompt (anaconda3)
Jupyter Notebook (anaconda3)
App
Open
Run as administrator

Setting Up Your Jupyter Notebook

- On your web browser, you should be able to see the Jupyter Notebook interface.

- If you're able to arrive at this screen, the installation should have been completed successfully!



Select items to perform actions on them.

Name	Last Modified	File size
/		
3D Objects	an hour ago	
Contacts	an hour ago	
Documents	an hour ago	
Downloads	12 minutes ago	
Favorites	an hour ago	
Links	an hour ago	
Music	an hour ago	
OneDrive	an hour ago	
Saved Games	an hour ago	
Searches	an hour ago	
Videos	an hour ago	

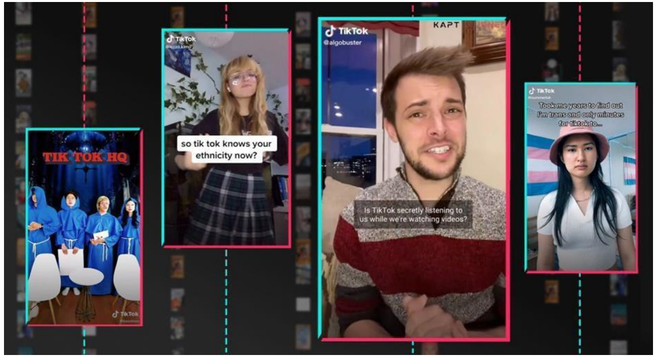
Introduction to Data Science

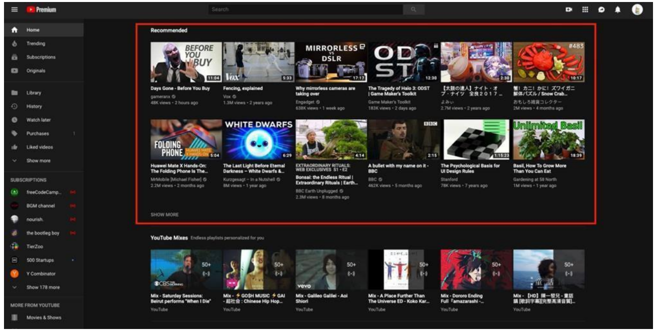
What is Data Science?

Ever wondered how YouTube’s recommendation engine works? Or how TikTok knows exactly what to show you next?

These predictive functionalities are driven by training a computer how to learn using large data sets.

Machine learning is powering innovation in everything from insurance-tech to lending models to fraud detection.

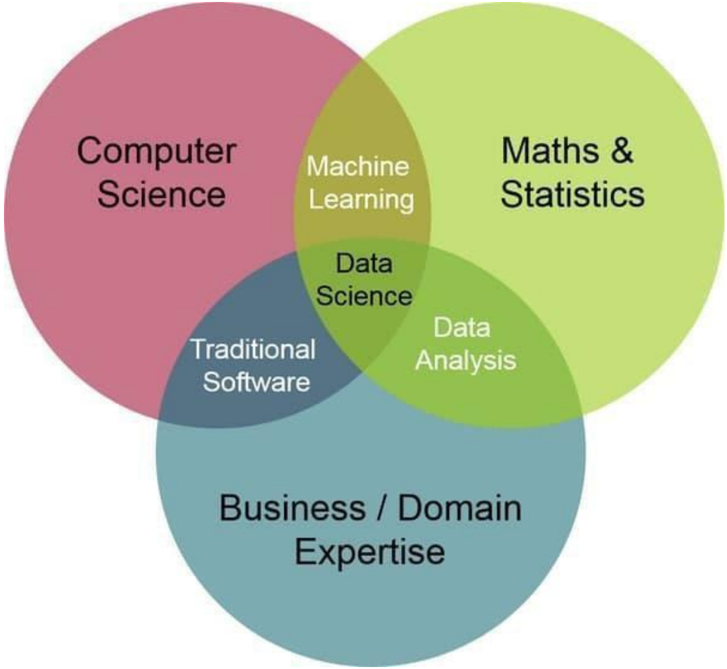

TIK TOK HQ
TikTok
so tik tok knows your ethnicity now?
TikTok
@algobuster
Is TikTok secretly listening to us while we're watching videos?
TikTok
Took me years to find out i'm trans and only minutes for tiktok to...



https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how- tiktok-algorithm-figures-out-your-deepest-desires/6C0C2040-FF25-4827-8528-2BD6612E3796

What is Data Science?

Data science lies at the intersection of business, statistics and computer science.



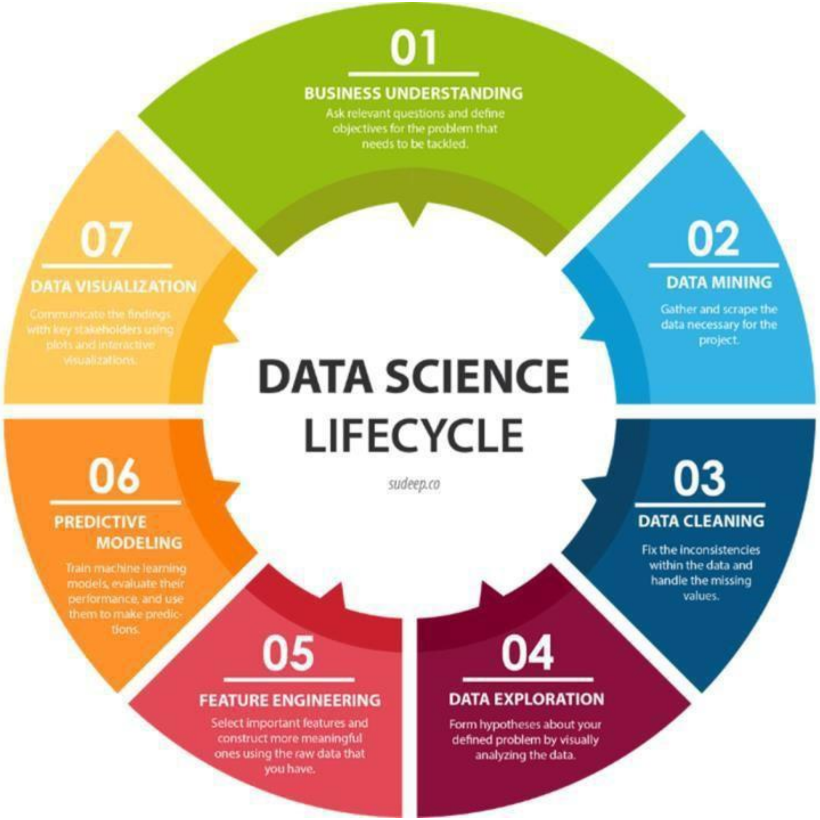
What is Data in Data Science?

Traditional data is data that is structured and stored in databases which analysts can manage from one computer; it is in table format, containing numeric or text value.

Big data, on the other hand, is… bigger than traditional data, and not in the trivial sense. From variety (numbers, text, but also images, audio, mobile data, etc.), to velocity (retrieved and computed in real time), to volume (measured in tera-, peta-, exa-bytes), big data is usually distributed across a network of computers.

https://www.kdnuggets.com/2018/06/what-where-how-data-science.html

Data Science Lifecycle



DATA SCIENCE LIFECYCLE

sudeep.co

01 BUSINESS UNDERSTANDING
Ask relevant questions and define objectives for the problem that needs to be tackled.

02 DATA MINING
Gather and scrape the data necessary for the project.

03 DATA CLEANING
Fix the inconsistencies within the data and handle the missing values.

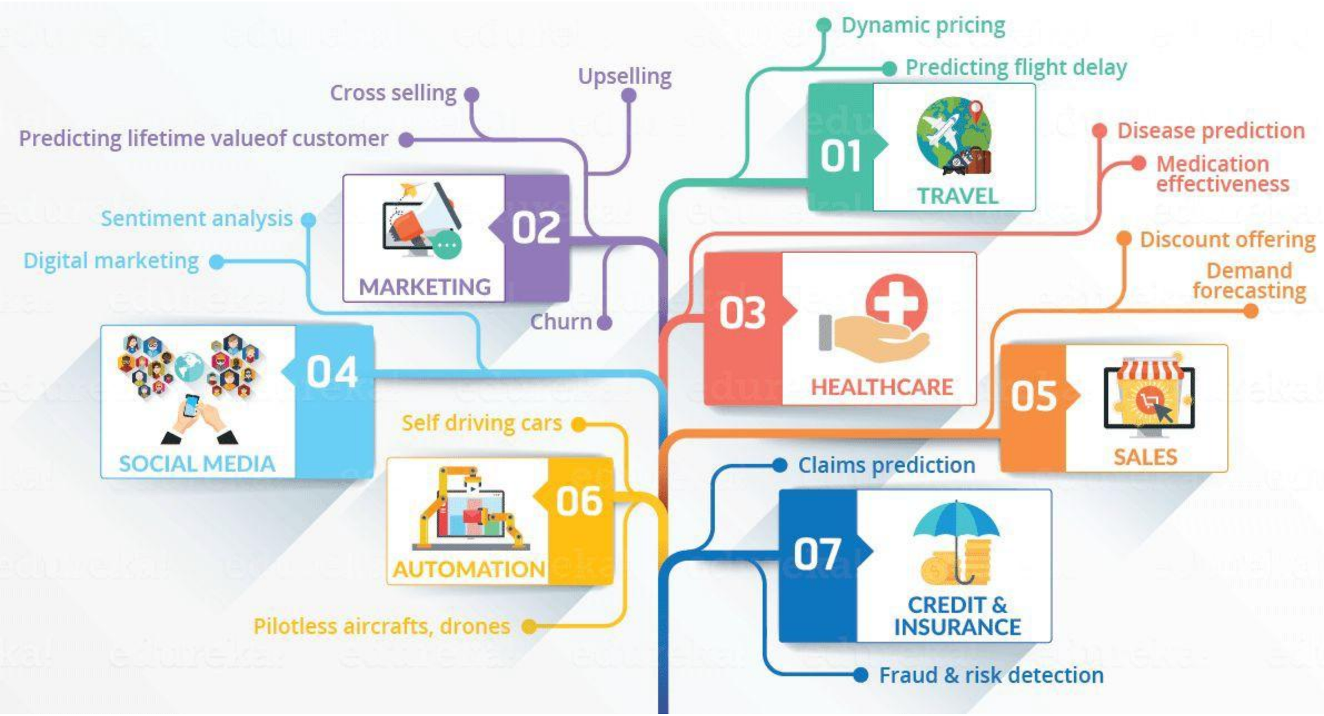
04 DATA EXPLORATION
Form hypotheses about your defined problem by visually analyzing the data.

05 FEATURE ENGINEERING
Select important features and construct more meaningful ones using the raw data that you have.

06 PREDICTIVE MODELING
Train machine learning models, evaluate their performance, and use them to make predictions.

07 DATA VISUALIZATION
Communicate the findings with key stakeholders using plots and interactive visualizations.

Data Science Use Cases



Python Data Types

Python Data Types

Primitives data types are the building blocks for data manipulation and contain pure, simple values of a data. There are 4 primitive variable types.

Primitive Variable Type	Explanation	Examples
Integers	Represents numeric data and more specifically whole numbers from negative infinity to infinity	-1, 2, 50
Float	Short for floating point number, usually used with decimals	-2.1, 2.8, 3.14159
String	Collection of alphabets, words or other characters. Usually enclosed within a pair of single or double quotes	“words”, “1” ,“ ”
Boolean	Takes up the value of True, False. Commonly used for controlling flow of program.	True, False

Python Data Types

Non-primitives are the sophisticated members of data structure family. They don’t just store a value, but rather a collection of values in various formats.

Non-Primitive Variable Type	**Explanation**	**Examples**
Lists	Used to store collection of heterogenous(diverse) items. They are mutable, which means you can change their content without changing their identity	[1,2,3] [‘a’,’b’,’c’] [1,’apple’,3]

Dictionaries	Made of key-value pairs. Key is to identify the item and the value holds as the name suggest, the value of item	x_dict = {‘a’:1,
		‘b’:2 }

Tuples	Tuples are another standard sequence data type however it is immutable, meaning once defined, you cannot delete, add or edit any values inside it	(‘a’,’b’,’c’,’d’,’e’)

Set	Unordered collection of distinct unique objects.	x = set([‘a’,’a’,b’,’c’])
		>>> {‘a’,’b’,’c’}

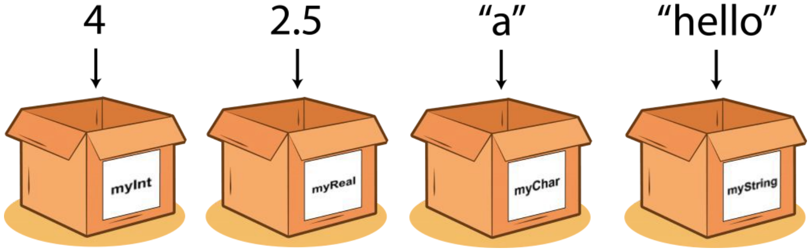
Variables

Variables Assignment

- A variable is a named place in the memory where a programmer can store data and retrieve the data later by using the variable name
- As a programmer, you get to choose the names of the variables

Examples of declaring a variable in python

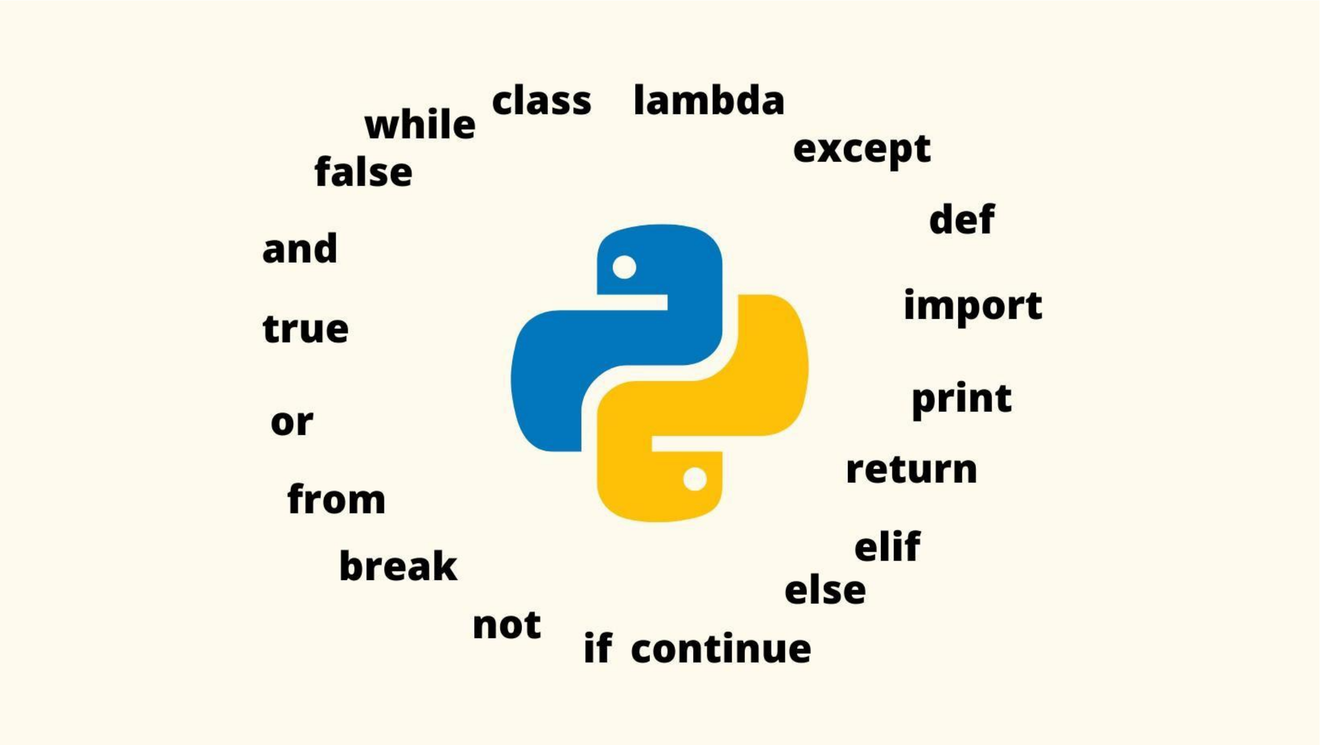
```
a = 1 2 3                 #number
b = ‘Hello’               #string #list
c = [1,2,3]               #dictionary
d = {“1”: “A”, “2”: “B”}
```



Rules of Declaring a Variable

1. Python variable name can contain small case letters (a-z), upper case letters (A-Z), numbers (0-9), and underscore (_).
2. A variable name can’t start with a number.
3. We can’t use reserved **keywords** as a variable name.
4. Python variable can’t contain only digits.
5. A python variable name can start with an underscore or a letter.
6. The variable names are case sensitive.
7. There is no limit on the length of the variable name.

Reserved Keywords



Best Practices

- Variable names should be lowercase.
- A variable's name should be representative of the value(s) it has been assigned.
- If you must use multiple words in your variable name, use an underscore to separate them.

Examples of Invalid Variables

- **9abc:** variable name can’t start with a number.
- **123:** variable name can’t contain only numbers.
- **x-y:** the only special character allowed in the variable name is an underscore.
- **def:** invalid variable name because it’s a reserved keyword.

Built-In functions

Functions are statements that run a specific computation on the input you give it.
Functions are identifiable by the function name followed by round brackets.

- sorted()
- len()
- set()
- list()
- print()
- type()

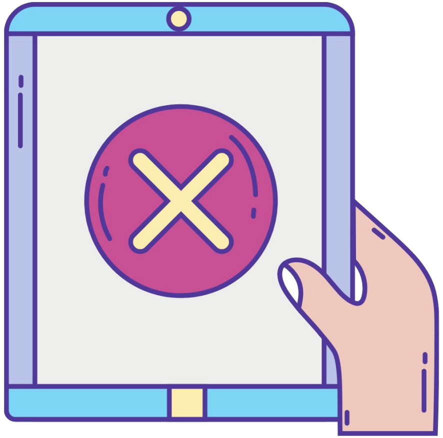
Python Errors

What are Python Errors?

We can make certain mistakes while writing a program that leads to errors when we try to run it.

A python program terminates as soon as it encounters an unhandled error. These errors can be broadly classified into 2 groups:

1. Syntax errors
2. Logical errors (exceptions)



What are Python Errors?

SyntaxError: Code that cannot be interpreted by Python

AttributeError: when you try to call an attribute of an object whose type does not support that method

NameError: using variable that does not exist yet

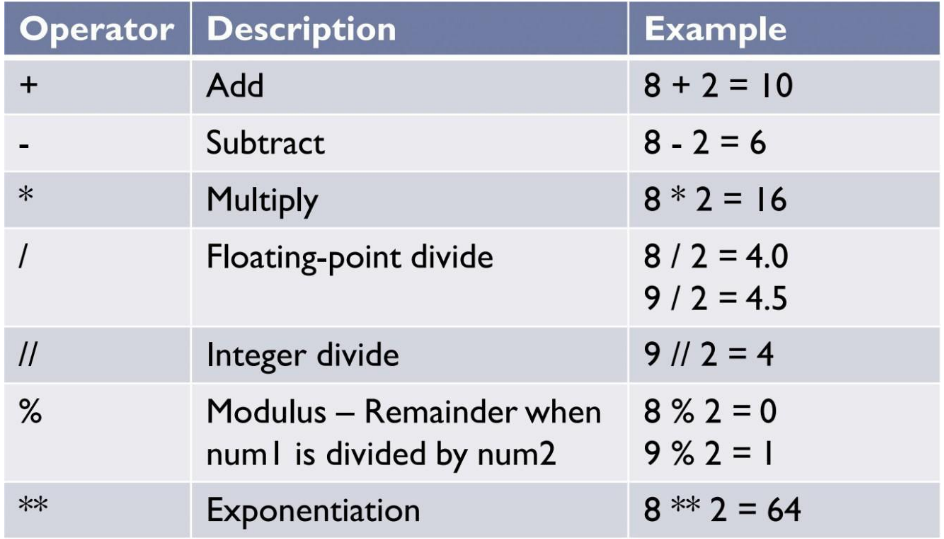
TypeError: Doing operation on an incorrect/unsupported object type

ZeroDivisionError: Due to either a number being divided by zero, or a number being modulo by zero

Python Operators

Arithmetic Operator

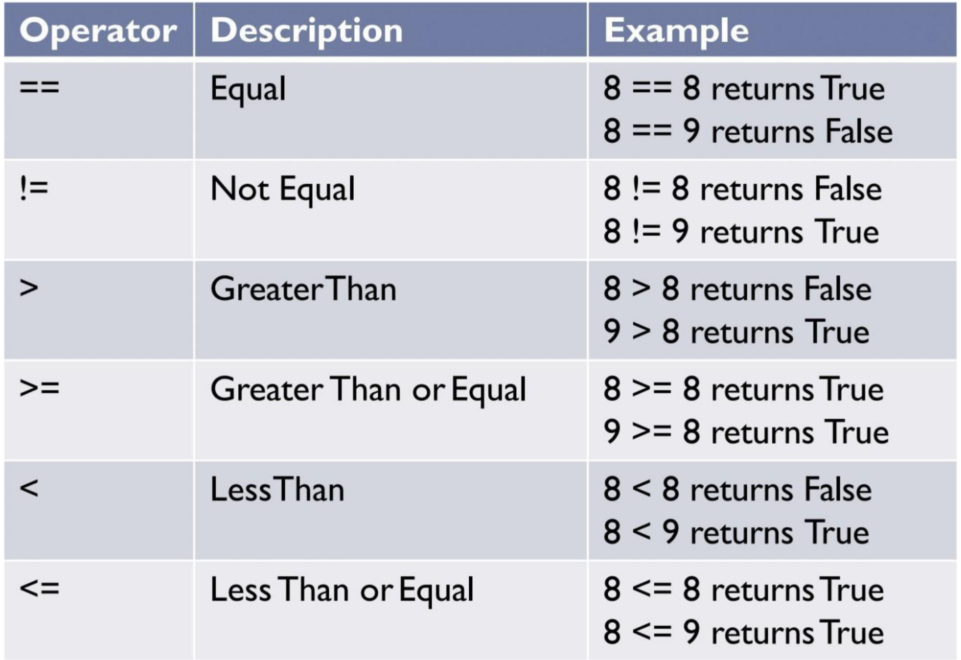
The arithmetic operators perform **addition, subtraction, multiplication, division, exponentiation, and modulus operations**



Operator	Description	Example
+	Add	8 + 2 = 10
-	Subtract	8 - 2 = 6
*	Multiply	8 * 2 = 16
/	Floating-point divide	8 / 2 = 4.0 9 / 2 = 4.5
//	Integer divide	9 // 2 = 4
%	Modulus – Remainder when num1 is divided by num2	8 % 2 = 0 9 % 2 = 1
**	Exponentiation	8 ** 2 = 64

Relational Operator

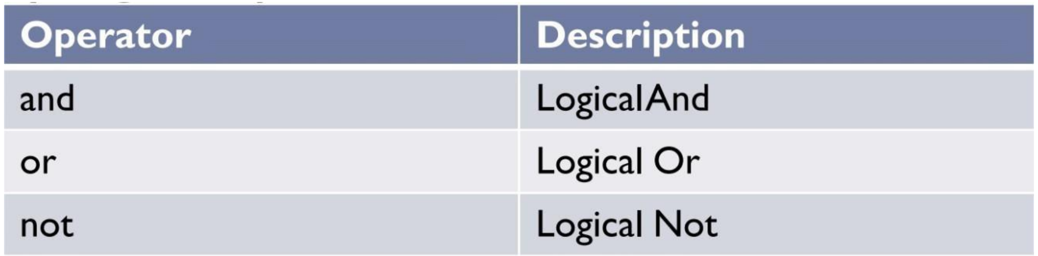
- Operators can also be used to compare objects
- Comparison is required for sorting, sorting helps searching, both of which are fundamental information processing tools
- Takes in 2 operands, returns a boolean (True, False) which is later **used to control program flow, or filter rows** during analytics



Operator	Description	Example
==	Equal	8 == 8 returns True 8 == 9 returns False
!=	Not Equal	8 != 8 returns False 8 != 9 returns True
>	Greater Than	8 > 8 returns False 9 > 8 returns True
>=	Greater Than or Equal	8 >= 8 returns True 9 >= 8 returns True
<	Less Than	8 < 8 returns False 8 < 9 returns True
<=	Less Than or Equal	8 <= 8 returns True 8 <= 9 returns True

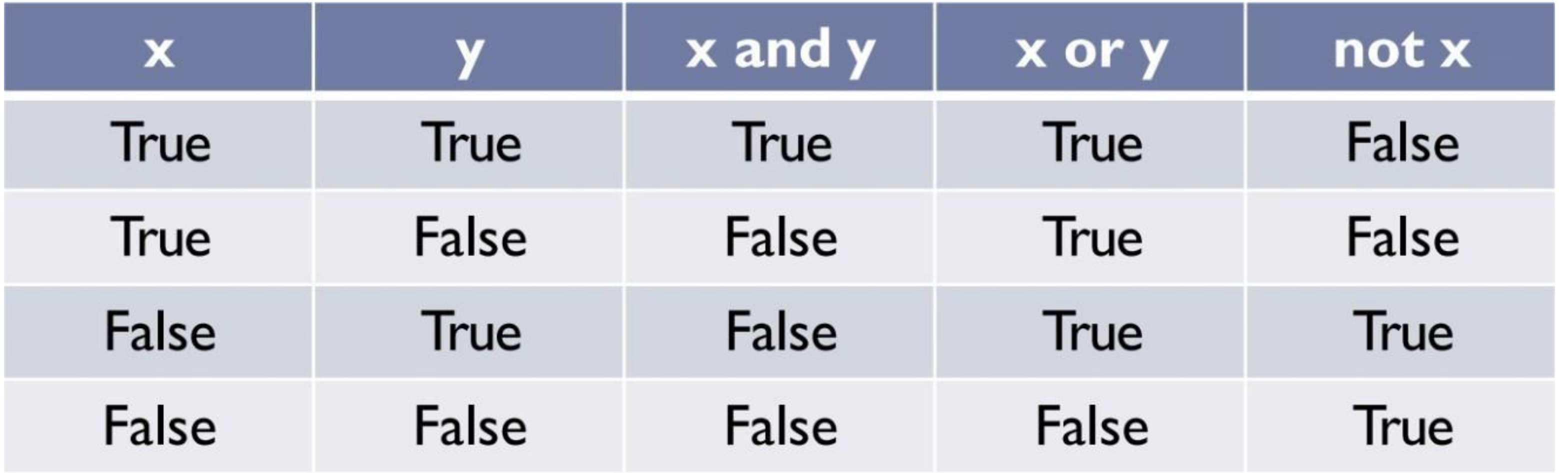
Logical Operator

Each comparison operator creates a single Boolean, logical operators combined booleans to implement logical concepts



Operator	Description
and	LogicalAnd
or	Logical Or
not	Logical Not

Logical Operator Examples



x	y	x and y	x or y	not x
True	True	True	True	False
True	False	False	True	False
False	True	False	True	True
False	False	False	False	True

Operator Precedence and Associativity

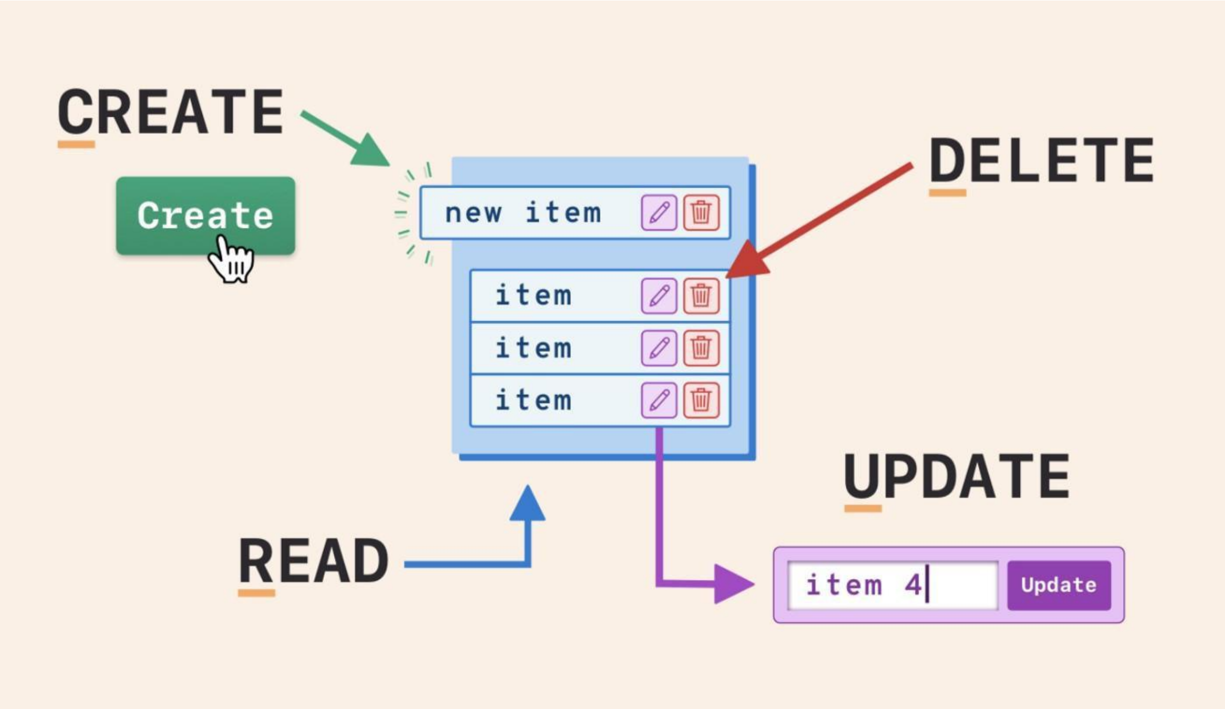
Operator Precedence (high to low):

1. Python operators have different levels of precedence
2. A good practice is to use parentheses to explicitly indicate the desired evaluation precedence

Operator Associativity (the order in which Python evaluates an expression containing multiple operators of the same precedence)

1. Left associativity means that the expression is evaluated from left-to-right (almost all operators)
2. Right associativity means the expression is evaluated from right-to-left

CRUD Framework



CRUD Framework

CRUD is an acronym that comes from the world of computing and refers to the four functions that are considered necessary to implement a persistent storage application.

Create	Allows users to create a new record in the database
Read	Allows users to search and retrieve specific groups in the table and read their values
Update	Allows users to modify existing records that exist in the database
Delete	Allows users to remove records from a database that is no longer needed

CRUD Operations In Finance

A financial institution maintains multiple databases that helps manage to and keep track of existing customers, financial products and spending patterns. Below are some of the common financial tables:

- A **Customer Data Table** includes attributes such as first and last name, personal identification number, contact number, home address, work location, and any other relevant personal details.
- A **Product Table** that includes the company's financial products such as credit cards, loans and trading activities.
- A **Transaction Table** that contains data at the transaction level for each of the customers, including frequency, amount and recency.

Recap time!

What are your favorite takeaways?

Let’s share with each other!

Some things to take note...

Link and resource could be accessed in the Learning Portal.

https://elearn.verticalinstitute.com/users/sign_in

Attendance Photo Taking

MODULE 01

DATA SCIENCE BOOTCAMP

Thank you!

Vertical
Institute