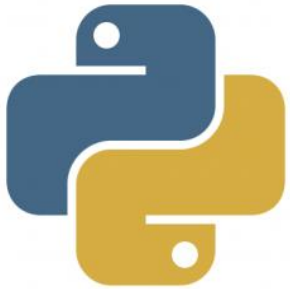
Introduction to



python

for Data Analysis

Course Outline

- Introduction to Python for Data Analysis
- Introduction to Python Data Types
- Introduction to Pandas (Series & Dataframe)
- Python Input/Output (xlsx, csv, sql, html)
- Data Exploration with Python
- Data Cleaning with Python
- Data Wrangling with Python
- Data Visualization with Matplotlib/Seaborn

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

python

Why Python For Data Analysis?

- Built for Speed
- Very simple and easy to learn
- Powerful libraries for everything
- Free and open- source
- It offers more ready-made tools for technologies such as data science, machine learning, artificial intelligence and robotics than other languages.

python

What is Anaconda?

- Anaconda is a free and open source distribution of R and Python languages. Focused on providing the tools necessary for Data science.
- In it, you will find the Anaconda Navigator, which is the graphical alternative to the command-line interface. This makes it easy for users to launch applications and manage packages and environments without using the command-line commands.

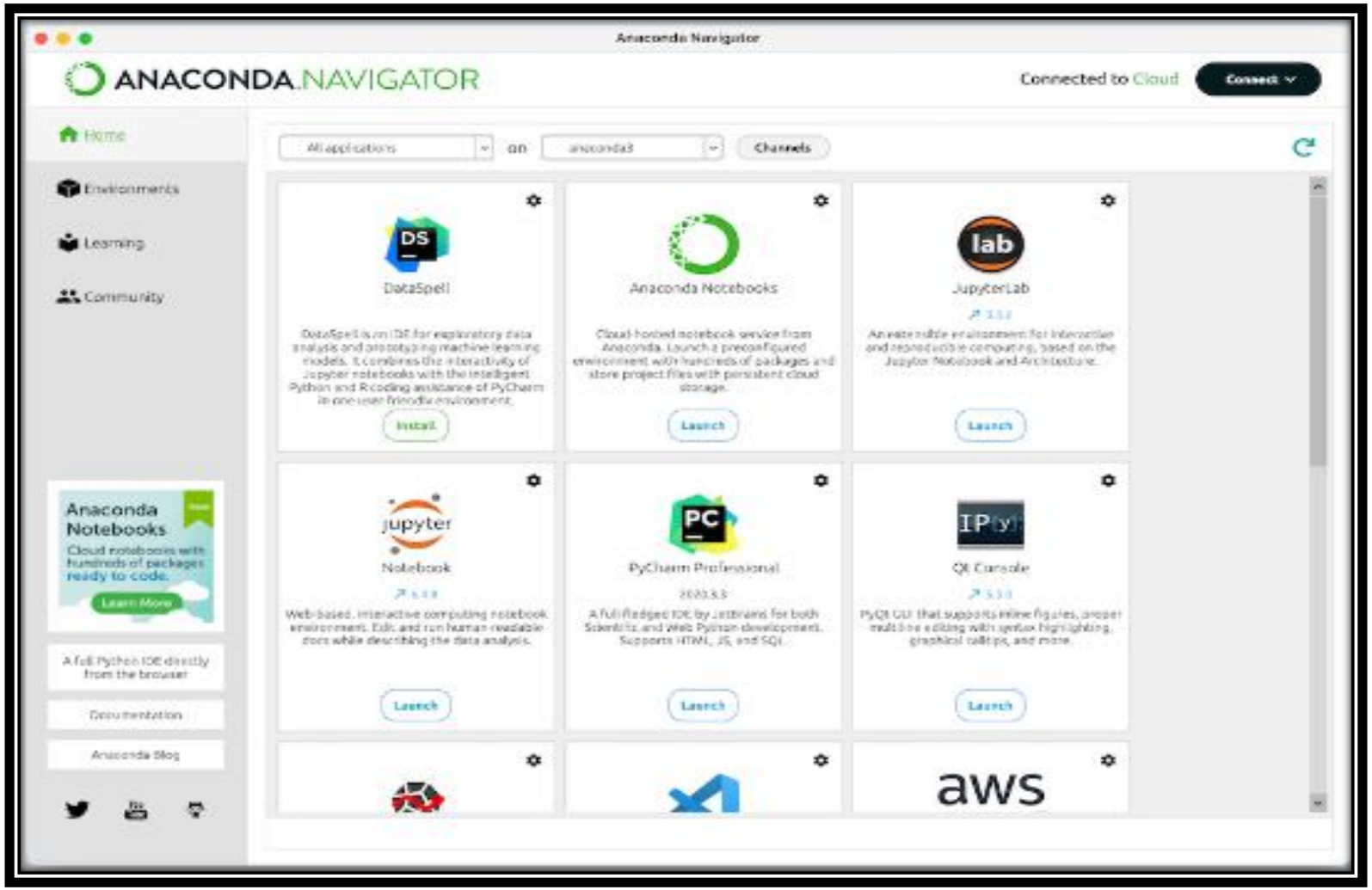
Anaconda Navigator

python

jupyter Notebook

Anaconda houses multiple other software's including

JUPYTER NOTEBOOK

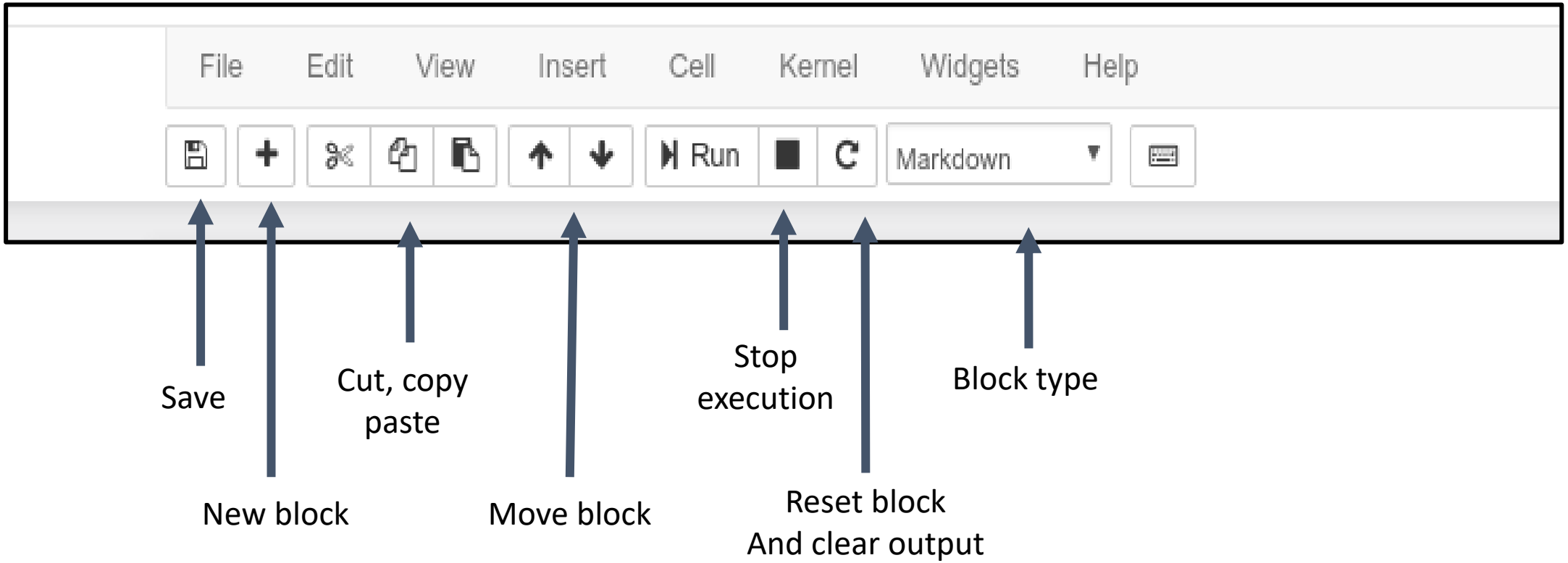


Jupyter Notebook

- Originally called Ipython, the name Jupyter comes from the fact that it supports writing code in three popular languages: Julia and Python and R which are popular for data science.
- Jupyter Notebook is an open source, web based interface application which accommodates live code and visualization all in one place.
- Supports more than 40 programming languages.

python

Exploring the Interface



jupyter Notebook

Working Modes

- **THE COMMAND MODE (The Blue Cell):** The command mode essentially means you're in command of everything i.e. you can add a cell, delete a cell.
- **THE EDIT MODE (The Green Cell):** The edit mode means you're in the cell. Essentially means you can edit a cell. When you want to get out the edit mode, you press the ESC button.

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Lets have a comment!

- Comments are useful when your code needs further explanation either for your future self and anybody else.
- Useful when you want to remove the code from execution but not permanently
- Comments in Python are done with #

NOTE: The shortcut of (Ctrl + /) can also be very handy

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Assignment Statement

Assignment statement simply means taking something and assigning it to something else(or an object).

For example

```
x = 5
name = “Ben”
age = 49
```

Python is an OOP (Object Oriented Programming) Language

which means that every thing can be kept inside an object

HiiT ...Your Premium Training Partner World Class IT Services since 1996

Print Statement

We can print any statement on console in python using the **PRINT** keyword

```
print()

print("This is my first program")

print('My name is', name)

print('My number is' + name)

print('My number is {}'. format(name))
```


HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Variables

VARIABLE		REASON
var1	Valid	Starts with letter
var1?	Invalid	Only _ is allowed
var1_	Valid	Starts with letter and contains _
1var	Invalid	Starts with number
_ var	Valid	Can start with _
_1var	Valid	Can start with _
var_1	Valid	_ and numbers are allowed
. var	Invalid	Can only start with _ or A-Z or a-z

- Variables are defined to store values such as characters (text), integers (numbers), Boolean values (True or False) etc.
- Variable name must start with only letter or underscore (not number)
- Variable name should contain only number, letter or underscore

Python Datatypes

Data Type	Example
Integer	5, 10, 22, 36
Float	5.2, 3.6, 5.3
Complex	3 + 4j, 2j
Boolean	True, False
String	“Hello”, “23.5”
List	[1,2,3, “Hello”, True]
Tuple	(1,2,3, “Hello”, True)
Set	{1,2,3, “Hello”, True}
Dictionary	{“Name” : “Sam”, “Age” : 15}

HiiT ...Your Premium Training Partner | World Class IT Services since 1996

Common Python Libraries

- **Pandas:** The cornerstone of our data analysis with python
- **Matplotlib:** The foundational library for visualization.
- **Numpy:** The numeric library for all calculation in Python.
- **Seaborn:** A statistical visualization tool built on top of matplotlib
- **Statsmodels:** A library with many advanced statistical functions
- **Scipy:** Advanced scientific computing
- **Scikit-learn:** The most popular machine learning library

HiiT ...Your Premium Training Partner World Class IT Services since 1996