

PYTHON DVT TEACHING MODULE

INTRO DATA ANALYTIC

What is Data Analysis

> A process of inspecting, cleansing, transforming and modeling data with the goal of discovering useful information, informing conclusion and supporting decision-making.

[Definition by Wikipedia.](#)

Auto-managed closed tools



Programming Languages



Auto-managed closed tools

👎 Closed Source 🙅

👎 Expensive 💸

👎 Limited 😞

👍 Easy to learn 🙄

Programming Languages

👍 Open Source 🙌

👍 Free (or very cheap) 💰

👎 Extremely Powerful 💪

👎 Steep learning curve 🙄

Data Extraction

Data Cleaning

Data Wrangling

Analysis

Action

- SQL
- Scrapping
- File Formats
 - CSV
 - JSON
 - XML
- Consulting APIs
- Buying Data
- Distributed Databases

- Missing values and empty data
- Data imputation
- Incorrect types
- Incorrect or invalid values
- Outliers and non relevant data
- Statistical sanitization

- Hierarchical Data
- Handling categorical data
- Reshaping and transforming structures
- Indexing data for quick access
- Merging, combining and joining data

- Exploration
- Building statistical models
- Visualization and representations
- Correlation vs Causation analysis
- Hypothesis testing
- Statistical analysis
- Reporting

- Building Machine Learning Models
- Feature Engineering
- Moving ML into production
- Building ETL pipelines
- Live dashboard and reporting
- Decision making and real-life tests

PYTHON ECOSYSTEM:

The libraries we use...

- [pandas](#): The cornerstone of our Data Analysis job with Python
- [matplotlib](#): The foundational library for visualizations. Other libraries we'll use will be built on top of matplotlib.
- [numpy](#): The numeric library that serves as the foundation of all calculations in Python.
- [seaborn](#): A statistical visualization tool built on top of matplotlib.
- [statsmodels](#): A library with many advanced statistical functions.
- [scipy](#): Advanced scientific computing, including functions for optimization, linear algebra, image processing and much more.
- [scikit-learn](#): The most popular machine learning library for Python (not deep learning)

INSTALLATION PYTHON AND JUPYTER NOTEBOOK

MAC

- install python

```
brew install python3
```

<https://programwithus.com/learn/python/install-python3-mac>

 Step by step guide to install Python3 on a Mac | Free Python Resource • programwithus.com

WINDOW

- install python

<https://realpython.com/installing-python/#how-to-install-python-on-windows>

 Python 3 Installation & Setup Guide – Real Python • realpython.com

INSTALLING LIBRARY AND JUPYTER LAB

- install jupyter lab

```
pip install jupyterlab
```

- install library numpy

```
pip install numpy
```

- install library matplotlib

```
pip install matplotlib
```

- install library pandas

```
pip install pandas
```

- install library bokeh

```
pip install bokeh
```

- start jupyter lab notebook

```
jupyter lab
```

USING NOTEBOOK JUPYTER LAB

- IMPORT NOTES AND NECESSARY FILES

<https://drive.google.com/drive/folders/1X7otewwPpmL-GdN9IziljoURiWcVQ90o?usp=sharing>

TUTORIAL SESSION

PART 1 : Understanding CELL

PART 2 : Working with code

PART 3 : Interacting with data

PART 4 : Exporting to excel

PART 5 : Cleaning Data

PART 6 : Reading Data with other resources

PART 6 : Forecasting Data **

Exercise