**ฝึกทำ Data Visualization ด้วย Matplotlib : ตอน BoxPlot (เบื้องต้น)**

การทำ Data Visualization เป็นส่วนหนึ่งของการทำ EDA หรือ Exploratory Data Analysis ที่จะช่วยให้เราสามารถเห็นภาพรวม การกระจายของข้อมูล ที่เรากำลังจะเข้าไปทำการวิเคราะห์ และ หนึ่งในการทำ Data Visualization อีกอันที่ช่วยทำให้เราเห็นลักษณะการกระจายของข้อมูล รวมถึง Outlier ได้เป็นอย่างดี นั้นก็คือ Boxplots บางครั้งก็เรียกว่า box and whisker plots นั้นเองครับ

มาเริ่มดูกันเลยครับว่าเราจะทำการสร้าง Boxplot ใน python ด้วย Matplotlib ได้อย่างไร

*บทความนี้จะแสดงการเขียน Code บน Jupyter Notebook นะครับ สำหรับใครที่ยังไม่ได้ติดตั้ง สามารถเข้าไปศึกษาวิธีการติดตั้งได้ที่*[*https://jupyter.readthedocs.io/en/latest/install.html*](https://jupyter.readthedocs.io/en/latest/install.html)*ได้เลยครับ*

**เริ่มจากเปิดหน้าต่าง Jupyter Notebook ขึ้นมาก่อน**

**Install a pip package in the current Jupyter kernel**

ก่อนจะเริ่มสร้าง Model เราต้องทำการติดตั้ง Package หรือ Library ที่จำเป็นให้เรียบร้อยก่อน

``` # Install a pip package in the current Jupyter kernel

import sys

!{sys.executable} -m pip install -U numpy scipy scikit-learn matplotlib pandas

จากนั้นก็ Import data ที่เราจะใช้ในการสร้าง boxplot เข้ามาครับ

``` # Import Data

import numpy as np

import pandas as pd

from sklearn.datasets import load\_iris

# save load\_iris() sklearn dataset to iris

# if you'd like to check dataset type use: type(load\_iris())

# if you'd like to view list of attributes use: dir(load\_iris())

iris = load\_iris()

# np.c\_ is the numpy concatenate function

# which is used to concat iris['data'] and iris['target'] arrays

# for pandas column argument: concat iris['feature\_names'] list

# and string list (in this case one string); you can make this anything you'd like..

# the original dataset would probably call this ['Species']

df = pd.DataFrame(data= np.c\_[iris['data'], iris['target']],

columns= iris['feature\_names'] + ['target'])

df.rename(columns={'target': 'species'}, inplace=True)

**แล้วก็มาเริ่มการสร้าง Boxplot กันเลยครับ**