# PYKU Frame & Current status

JUNHO LEE (이준호)

# Outline



- 1. Update
  - 1.a. Web crawling :: current status & plan
  - 1.b. Currently having Data
  - 1.c. DNN prediction on SEOUL PM2.5
- 2. PYKU DATA Analysis Frame
  - 2.a. Overview
  - 2.b. Working Environment

# 1.a. Update : Web crawling :: current status & plan



- Current status :
  - Available package:
    - Beautiful soup
    - Selenium
  - Collected Data & Able to be collected
    - Air Quality related data :: SEOUL & BEIJING
    - Weibo, Twitter, Baidu index

- Next step :
  - Let me know, if you need Data, by "WeChat". (1 Month)
  - Using python web crawling package "Scrapy", formalize crawling frame. → (2 weeks)
  - At the mean time,
    - try with "美团", "饿了么", "京东", "淘宝"。。。→ (3~4 days for each)
    - Machine Learning study → (1 Month)
    - Docker & Git Study → (1 Month)

I wiil update, if any useful

# 1.b. Update : Currently having Data



- BEIJING
  - AQI, PM2p5, PM10, SO2, CO, NO2, O3_8h
    - From 2013/12/02 to 2018/06/29 → 4.5 years
    - https://github.com/StudyGroupPKU/Project-pre/blob/master/data_txt/BEIJING_Aqi/Aqi_Beijing_Holi_6.txt

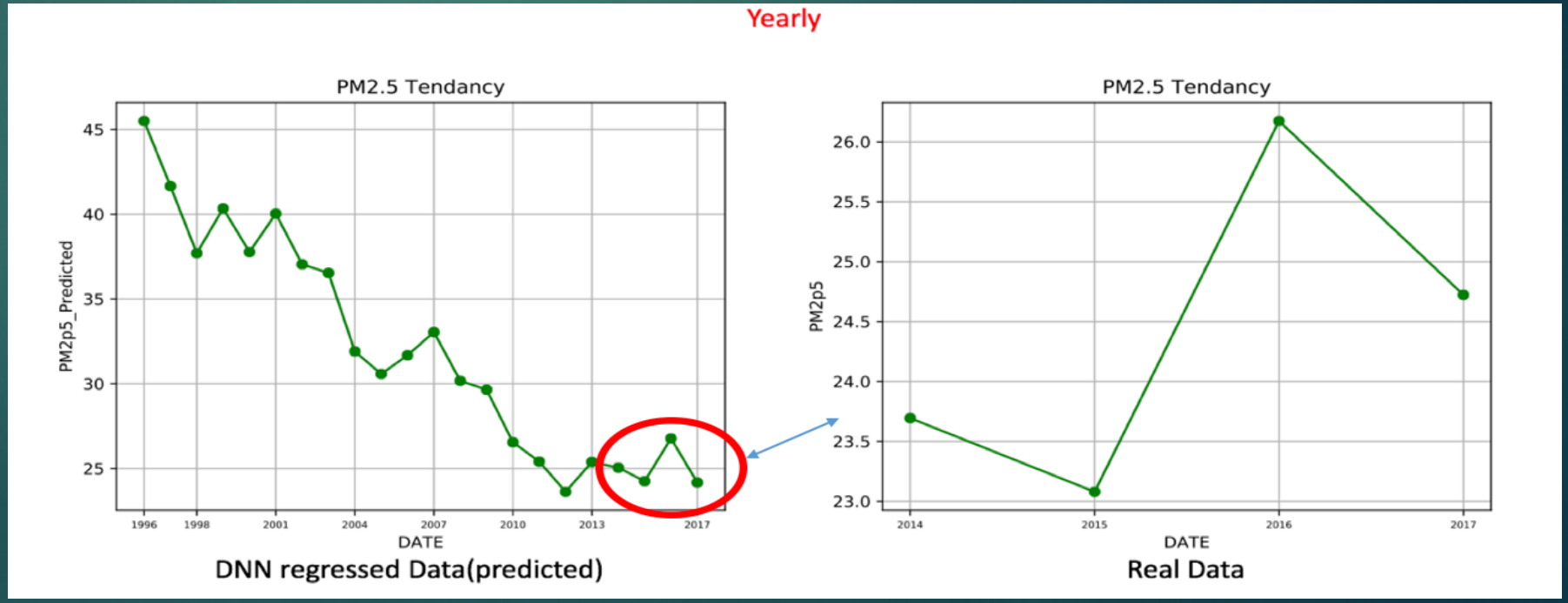
- SEOUL
  - PM2p5, PM10, TEMP_MEAN, H_TEMP, L_TEMP, CLOUD, RAIN, TEMP_Range, SO2, O3, NO2, CO
    - From 2014/01/01 to 2017/12/31 → 4 years
    - https://github.com/StudyGroupPKU/Project-pre/blob/master/data_txt/SEOUL_Aqi/Since14_17Y_Daily_ALL_atmos.txt
  - PM10, TEMP_MEAN, H_TEMP, L_TEMP, CLOUD, RAIN, TEMP_Range, SO2, O3, NO2, CO
    - From 1996/01/01 to 2017/12/31 → 22 years
    - https://github.com/StudyGroupPKU/Project-pre/blob/master/data_txt/SEOUL_Aqi/Since96_17Y_Daily_ALL_atmos.txt

# 1.c. Update : DNN prediction on SEOUL PM2.5



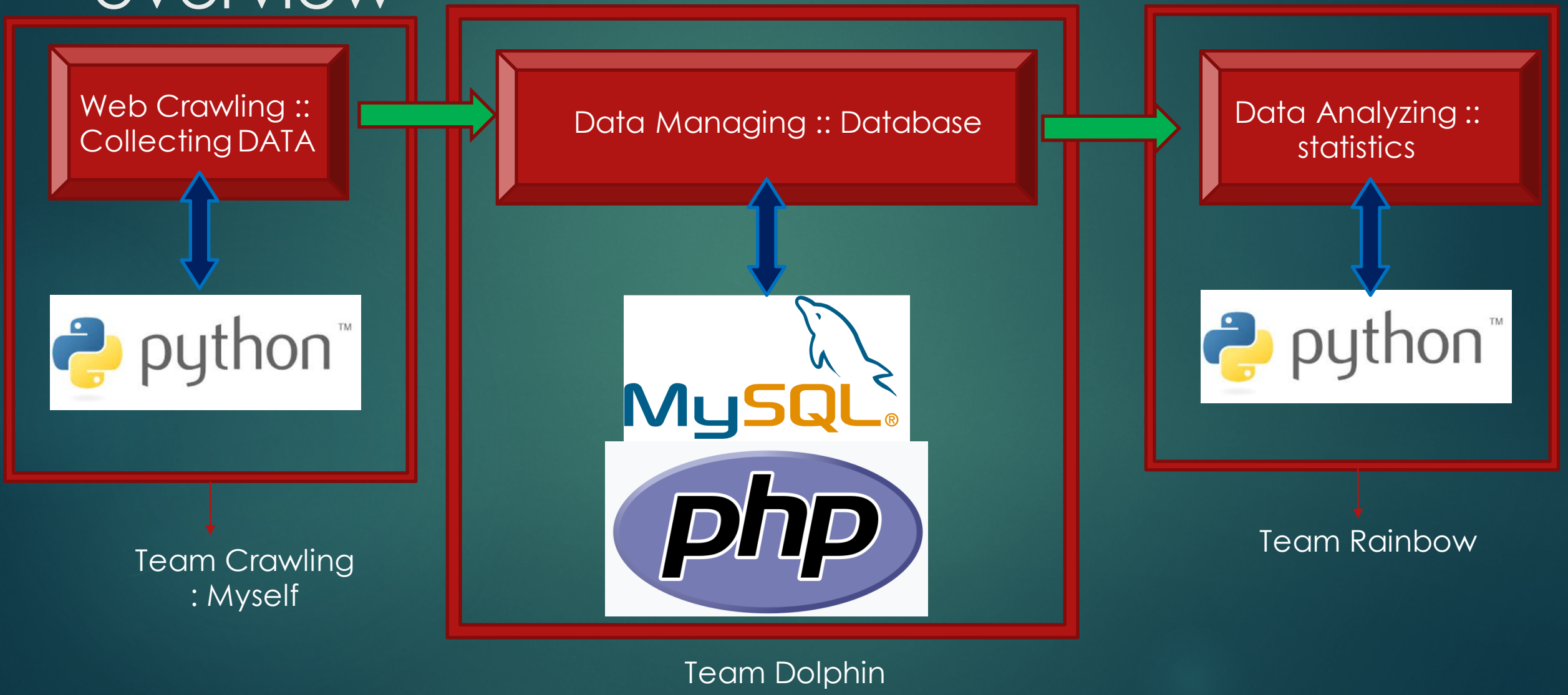
- DNN model construction
  - Using https://github.com/StudyGroupPKU/Project-pre/blob/master/data_txt/SEOUL_Aqi/Since14_17Y_Daily_ALL_atmos.txt Data, PM2.5 predicting model constructed.
  - With https://github.com/StudyGroupPKU/Project-pre/blob/master/data_txt/SEOUL_Aqi/Since96_17Y_Daily_ALL_atmos.txt Data, prediction on PM2.5 performed.

Increase of with PM2.5 highly correlated new DATA means getting better prediction result.



# 2.a. PYKU DATA Analysis Frame : overview





# 2.b. PYKU DATA Analysis Frame : Working Environment



- Tools :
  - Github (GIT). → familiar → easy to share python codes & scripts
  - Docker hub (DOCKER) → kind of know → Providing PYKU working ENV
  - MySQL (Database) → Not known → Saving & providing collected Data → Maybe via web..??
- Docker :: I recommend every works to be performed on Docker ISO, since there would be no compatibility problem with given DOCKER ISO.
- MySQL :: plan??

2.b.

**** We don't save DATA neither on Docker hub nor on GIthub

Github

StudyGroupPKU/Project-pre

.
.
.

python

ROOT
Data Analysis Framework

python
Web crawling

(Team Crawling)

Push Data

Docker PYKU ISO

Docker PYKU container

Working space

Statistical Analysis

ML analysis …

python

git clone
git push

git clone & push

export data

export data

MySQL

php

Storing & sharing & managing collected Data

Team Crawling

Team Dolphin

Team Rainbow