

## Project 2 (Open Project) Manual

Topic: Choice 1) PM2.5

Group: ใจผมสลายชะมุง

Member:

6231340921	Pongsapak	Pulthasthan
6231343821	Peeranath	Theerawatanachai
6231367921	Sahutsawat	Sivakosit
6230054221	Kharit	Tasanakowit

### Directory structure

- Folder มีทั้งหมด 5 Folder ได้แก่
  1. **cleaned\_data**: เก็บ Data ตั้งต้นที่ผ่านการ clean แล้ว ซึ่งสามารถนำไปวิเคราะห์ หรือทำ Visualize ต่อได้
  2. **data**: เก็บ Data ตั้งต้นที่ได้ทำการ download มาตั้งแต่ต้น
  3. **outsources\_data**: เก็บ Data ที่ทำการ download มาจากภายนอก เช่น ตารางแปลงชื่อจังหวัด
  4. **processed\_data**: เก็บ Data ที่ผ่านการ process แล้ว
  5. **spark**: เก็บ Library ของ spark
- File มีทั้งหมด 10 Files ได้แก่
  1. **Bangkok\_and\_Metropolitan\_Clustering\_2021.ipynb**: ไฟล์ที่ทำการทำ clustering stations ต่าง ๆ ที่วัดค่า PM2.5 ในเขตกรุงเทพมหานครและปริมณฑล โดยใช้ข้อมูล PM2.5 ในปี 2021 จาก SFA และ PCD (Before)
  2. **Cleaning\_PCD\_Data\_After\_for\_Visualization.ipynb**: ไฟล์ที่ใช้ในการเตรียม PCD Data After สำหรับ Visualization
  3. **Cleaning\_PCD\_Data\_After.ipynb**: ไฟล์ที่ใช้ในการเตรียม PCD Data After
  4. **Cleaning\_PCD\_Data\_Before.ipynb**: ไฟล์ที่ทำการ merge และ clean ข้อมูลจาก station ต่าง ๆ ในข้อมูล PCD Data Before โดยใช้ Spark
  5. **Creating\_Dictionary\_both\_SFA\_PCD.ipynb**: ไฟล์ที่ใช้สร้าง dictionary ของทั้ง SFA, PCD before เพื่อให้รู้ว่าแต่ละข้อมูลเป็นของจังหวัดไหน(SFA\_merge คือ SFA\_DICTIONARY join SFA\_DATA)
  6. **EDA\_PCD\_Data.ipynb**: ไฟล์ซึ่งนำข้อมูล PCD ทั้งหมดมาทำการวิเคราะห์ โดยมีการวิเคราะห์ค่าเฉลี่ยของ PM2.5 ในแต่ละเดือนของจังหวัดต่าง ๆ รวมถึงมีการทำ clustering แต่ละจังหวัดในประเทศไทย
  7. **Mapping\_PCD\_Data\_Before\_Station.ipynb**: ไฟล์ซึ่งทำการรวบรวมข้อมูลสถานีต่าง ๆ ที่ทำการวัดค่าของข้อมูล PCD (Before)
  8. **Mapping\_Thailand\_Province\_With\_Region\_and\_Coordinate.ipynb**: ไฟล์ซึ่งรวบรวมข้อมูลจังหวัดต่าง ๆ ในประเทศไทยเข้าด้วยกัน โดยมีข้อมูลชื่อจังหวัด (TH/EN) ภูมิภาค และการทำการ scrape latitude, longitude จาก Google Map
  9. **paired\_t\_test.ipynb**: ไฟล์ที่ใช้ทำ Statistical analysis ของ PCD Data
  10. **Visualization\_with\_PowerBI.pbix**: ไฟล์ที่ใช้ทำ Visualization SFA Data ด้วย Power BI

## หมายเหตุ

- หากต้องการ reproduce code ในบางไฟล์ต้องมีการเปลี่ยน basic\_path (cd ตอนต้น) ก่อนถึงจะสามารถทำงานได้
- หากต้องการ reproduce code ไฟล์ 4. Cleaning\_PCD\_Data\_Before.ipynb จำเป็นต้อง Setup library Spark ก่อนถึงจะสามารถทำงานได้
- หากต้องการดูไฟล์ PowerBI สามารถดาวน์โหลดไฟล์ 10. Visualization\_with\_PowerBI.pbix แล้วนำไปใช้งานได้เลย