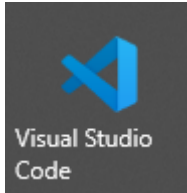


## โจทย์ข้อที่ 1

1. source code ในโปรเจค section\_1 นี้ run ด้วย VSCODE



2. package ที่ใช้ใน source code
  - pip install pandas openpyxl
  - pip install selenium
  - pip install beautifulsoup4
3. selenium web driver ในโปรเจคนี้ run ด้วย **Chrome**



4. วิธี Run file ใน vscode : ไปที่ แถบ menu

Terminal >> new\_terminal

Run :

Python main\_1.py

Python main\_2.py

```
keyboardInterrupt
PS E:\python\PSU\section_1> python main_1.py
```

5. Demo :

โจทย์ข้อที่ 1 ส่วนที่ 1 : <https://youtu.be/pB8TdShS1Qg>

โจทย์ข้อที่ 1 ส่วนที่ 2 : [https://www.youtube.com/watch?v=M-\\_fqU1xldo](https://www.youtube.com/watch?v=M-_fqU1xldo)

## โจทย์ข้อที่ 2

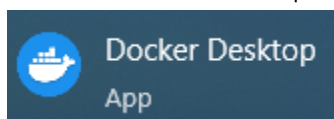
1. source code ใน โปรเจค section\_2 นี้ run บน apache airflow
2. ขอข้ามใน โจทย์ข้อที่ 2 ส่วนที่ 1 เนื่องจาก ใช้เวลาศึกษาไม่ทัน

### โจทย์ข้อที่ 2 การสร้าง Pipeline, ฐานข้อมูล, และการจัดการคุณภาพข้อมูล (คะแนนเต็ม 30 คะแนน)



- (15 คะแนน) การจัดการคุณภาพข้อมูลด้วย OOP (Data Manipulation):
  - เขียนสคริปต์ในรูปแบบ Object-Oriented Programming (OOP) เพื่อจัดการคุณภาพข้อมูลตามมาตรฐานหลักเกณฑ์การประเมินคุณภาพข้อมูลสำหรับหน่วยงานภาครัฐ <https://standard.dga.or.th/standard/dg-std/5785/>
  - สร้าง Class (เช่น DataProcessor) ที่มี Method สำหรับ:
    - Standardization: จัดการข้อมูลให้อยู่ในรูปแบบมาตรฐาน (เช่น รูปแบบชื่อผู้แต่ง, ชื่อวารสาร)
    - Handling Missing Values: จัดการค่าว่างในคอลัมน์ที่สำคัญ
    - Deduplication: สร้างตรรกะเพื่อระบุและรวมข้อมูลที่ซ้ำซ้อนกันจากแหล่งต่างๆ
    - เป็นต้น

3. วิธีติดตั้ง apache\_airflow บน window 10 และ setup environment ที่ใช้ในโจทย์ข้อสอบ
  - 3.1 ติดตั้ง environment ของ ubuntu บน window 10
  - 3.2 ติดตั้ง docker desktop
  - 3.3 เปิด command line ของ ubuntu ที่ได้ติดตั้งไว้



### 3.4 ติดตั้ง apache\_airflow ด้วย docker บน window (ไม่มี Docker ใน ubuntu ให้ไปหาติดตั้งมาก่อน)

#### 1. Docker-based Installation : Install Docker and Docker Compose.

```
## mkdir apache_airflow
## cd apache_airflow
## sudo apt-get update
## sudo apt-get install ca-certificates curl gnupg lsb-release
## curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --
dearmor -o /usr/share/keyrings/docker-archive-keyring.gpg
## echo "deb [arch=$(dpkg --print-architecture) signed-
by=/usr/share/keyrings/docker-archive-keyring.gpg]
https://download.docker.com/linux/ubuntu $(lsb_release -cs) stable" | sudo tee
/etc/apt/sources.list.d/docker.list > /dev/null
## sudo apt-get update
## sudo apt-get install docker-ce docker-ce-cli containerd.io docker-compose-
plugin
```

-----

Download Airflow Docker Compose file.

```
## curl -LfO 'https://airflow.apache.org/docs/apache-airflow/stable/docker-
compose.yaml'
```

-----

Create necessary directories and set environment variable:

```
## mkdir -p ./dags ./logs ./plugins
## echo -e "AIRFLOW_UID=$(id -u)" > .env
```

-----

Initialize the database.

```
##docker compose up airflow-init
```

-----

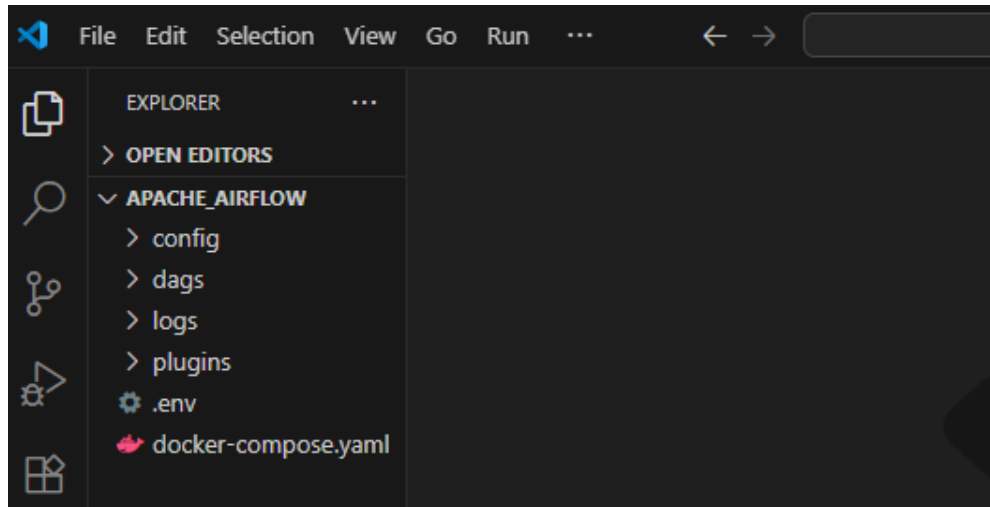
Run Airflow.

```
##docker compose up -d
```

#### 2.Open in vs code

#code .

**3.เมื่อเปิด vscode จะได้โครงสร้าง project ดังรูป**



3.5 ถ้าติดตั้งถูกต้อง run บน browser ด้วย

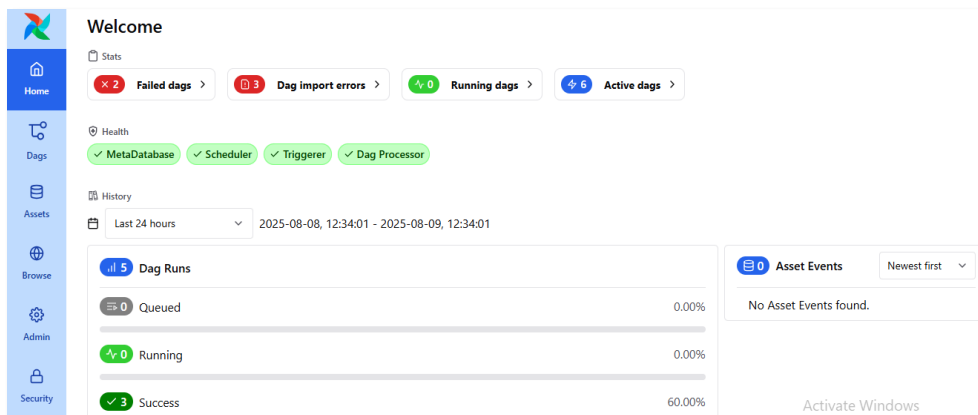
<http://localhost:8080>

Login

Username : airflow

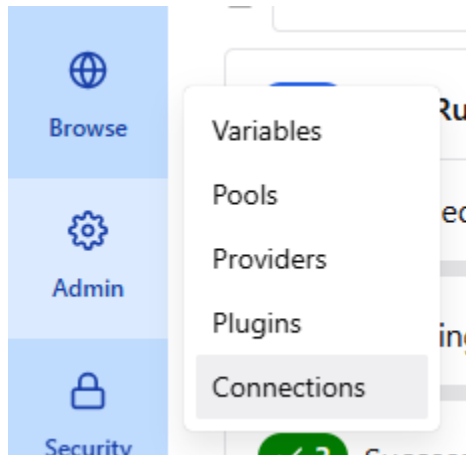
Password : airflow

จะเข้าหน้าเว็บต่อไปนี้



#### 4. Set up environment ของ postgresql บน apache\_airflow

##### 4.1 ไปที่ admin >> connection



##### 4.2 ไปที่ add connection



##### 4.3 ตั้งค่า postgresql connection

A screenshot of the 'Add Connection' form in the Apache Airflow web interface. The form has a title 'Add Connection' and a close button (X). It contains the following fields:

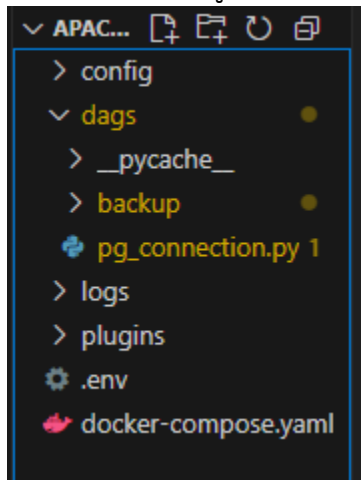
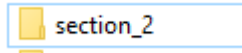
- 'Connection ID \*': A text input field.
- 'Connection Type \*': A dropdown menu with 'Postgres' selected. Below it, a small text note says: 'Connection type missing? Make sure you have installed the corresponding Airflow Providers Package.'
- 'Standard Fields': A section with a collapse/expand arrow (upward arrow) containing three text input fields:
  - 'Description': A text input field.
  - 'Host': A text input field.
  - 'Login': A text input field.

##### 4.4 ถ้าตั้งค่าถูกต้องจะได้

Connection Id	Connection Type	Description	Host	Port	
pg-1f2e1a3e	postgres	pg-1f2e1a3e	pg-1f2e1a3e-punvarat-7a2d.baivencloud.com	22067	<a href="#">✎</a> <a href="#">🗑</a>

5. จาก โจทย์ข้อที่ 2 ส่วนที่ 2 เพิ่มไฟล์ชื่อ pg\_connection.py ใน dags

Source code อยู่ใน Folder



Demo :

โจทย์ข้อที่ 2 ส่วนที่ 2 : [https://www.youtube.com/watch?v=VQQPB-\\_saKM](https://www.youtube.com/watch?v=VQQPB-_saKM)

### โจทย์ข้อที่ 3

#### ผลลัพธ์ที่ได้จากการ run sql

1. แสดงจำนวนผลงานวิจัยทั้งหมดในแต่ละปี โดยเรียงจากปีล่าสุดไปหาเก่าสุด

Sql :

```
SELECT * FROM public.research order by years desc;
```

Output :

	id [PK] integer	title text	author text	years integer	data_resource character varying (255)	university text
5	5	พฤติกรรมการใช้บริการ และสภ...	สุชาดา กองสวัสดิ์	2025	TCI	สงขลานครินทร์
6	44	แนวทางการพัฒนาคุณภาพพระศ...	สุชาดา กองสวัสดิ์	2025	TCI	ราชภัฏสวนดุสิต
7	43	ปัจจัยที่ส่งผลต่อผลผลิตภาพในกา...	สุชาดา กองสวัสดิ์	2025	TCI	สงขลานครินทร์
8	59	การพัฒนาระบบยื่นขอกำหนด...	เครือมาส แก้วทอน	2025	OpenAlex	ธรรมศาสตร์
9	9	ปัจจัยที่ส่งผลต่อผลผลิตภาพในกา...	สุชาดา กองสวัสดิ์	2025	TCI	สงขลานครินทร์
10	10	แนวทางการพัฒนาคุณภาพพระศ...	สุชาดา กองสวัสดิ์	2025	TCI	ราชภัฏสวนดุสิต
11	1	ปัจจัยที่มีอิทธิพลต่อการมีผลงา...	สุชาดา กองสวัสดิ์	2024	TCI	สงขลานคริ...
12	2	การศึกษาประสิทธิภาพของการ...	ณัฏฐา ศิริรักษ์	2024	TCI	สงขลานครินทร์
13	8	การพัฒนาระบบบริหารจัดการคว...	อัญชลี พัฒนพันธ์ชัย	2024	TCI	สงขลานครินทร์
14	14	ปัจจัยที่มีผลต่ออัลเวสระยะยั้งช...	วิชญพงษ์ โพธิ์พิรุฬห์	2024	Scopus	วลัยลักษณ์
15	26	ปัจจัยที่มีผลต่ออัลเวสระยะยั้งช...	เครือมาส แก้วทอน	2024	OpenAlex	วลัยลักษณ์
16	25	ปัจจัยที่มีอิทธิพลต่อการมีผลงา...	สุชาดา กองสวัสดิ์	2024	TCI	สงขลานคริ...

2. ค้นหารายชื่อผู้แต่ง 10 อันดับแรกที่มีจำนวนผลงานตีพิมพ์มากที่สุด

Sql :

```
SELECT author,count(*) as n FROM public.research  
group by author order by n desc limit 10;
```

Output :

	author text	n bigint
1	เครื่องมือส แก้วทอง	14
2	สุชาติ กองสวัสดิ์	10
3	ธนพล กิ่งภูเขา	10
4	ศิริพร เทพอรุณ	8
5	ณัฏฐา ศิริรักษ์	6
6	วิษณุ โรจน์สุวรรณ	4
7	บรรณกร แซ่ลิ้ม	4
8	กมลวรรณ แสงทอง	4
9	ปริญญ อาทิลวาเลีย	2
10	นุฬามิตรอฟลี แวหะมะ	2



### 3. แสดงรายชื่อผลงานวิจัยที่มาจากแหล่งข้อมูล 'Scopus' และตีพิมพ์หลังปี 2022

Sql :

```
SELECT * FROM public.research where data_resource = 'Scopus'
and years < 2022
```

Output :

	id [PK] integer	title text	author text	years integer	data_resource character varying (255)	university text
1	11	ปัจจัยที่มีอิทธิพลต่อการมีผลงานตีพิมพ์ในวารสารทางวิชาการระดับ...	สุชาติ กองสวัสดิ์	2021	Scopus	ธรรมศาสตร์
2	17	ความรู้ ทัศนคติ และพฤติกรรมการออกกำลังกายของนักศึกษา มห...	วิษณุ ใจจรูญสุวรรณ	2021	Scopus	วลัยลักษณ์
3	18	การพัฒนาระบบบริหารจัดการความร่วมมือด้านต่างประเทศและเค...	กมลวรรณ แสง...	2020	Scopus	สงขลานครินทร์
4	19	ปัจจัยที่ส่งผลต่อผลผลิตภาพในการทำวิจัยของอาจารย์ในสายสังคม...	ธนพล กิ่งภูเขา	2020	Scopus	สงขลานครินทร์
5	21	ปัจจัยที่ส่งผลต่อผลผลิตภาพในการทำวิจัยของอาจารย์ในสายสังคม...	ธนพล กิ่งภูเขา	2021	Scopus	ธรรมศาสตร์
6	22	การพัฒนาระบบบริหารจัดการความร่วมมือด้านต่างประเทศและเค...	ธนพล กิ่งภูเขา	2020	Scopus	สงขลานครินทร์
7	45	ปัจจัยที่มีอิทธิพลต่อการมีผลงานตีพิมพ์ในวารสารทางวิชาการระดับ...	สุชาติ กองสวัสดิ์	2021	Scopus	ธรรมศาสตร์
8	51	ความรู้ ทัศนคติ และพฤติกรรมการออกกำลังกายของนักศึกษา มห...	วิษณุ ใจจรูญสุวรรณ	2021	Scopus	วลัยลักษณ์
9	52	การพัฒนาระบบบริหารจัดการความร่วมมือด้านต่างประเทศและเค...	กมลวรรณ แสง...	2020	Scopus	สงขลานครินทร์
10	53	ปัจจัยที่ส่งผลต่อผลผลิตภาพในการทำวิจัยของอาจารย์ในสายสังคม...	ธนพล กิ่งภูเขา	2020	Scopus	สงขลานครินทร์
11	55	ปัจจัยที่ส่งผลต่อผลผลิตภาพในการทำวิจัยของอาจารย์ในสายสังคม...	ธนพล กิ่งภูเขา	2021	Scopus	ธรรมศาสตร์
12	56	การพัฒนาระบบบริหารจัดการความร่วมมือด้านต่างประเทศและเค...	ธนพล กิ่งภูเขา	2020	Scopus	สงขลานครินทร์

4. ค้นหารายชื่อวารสารที่มีผลงานของบุคลากร ม.อ. ตีพิมพ์มากที่สุด 5 อันดับแรก

Sql :

```
SELECT author,count(*) as n FROM public.research  
where university like '%สงขลานครินทร์%' group by  
author order by n desc limit 5;
```

Output :

	author text	n bigint
1	เครื่องมือส แก้วท...	6
2	กมลวรรณ แสง...	4
3	สุชาดา กองสวัสดิ์	4
4	ธนพล กิ่งภูเขา	4
5	ณัฏฐา ศิริรักษ์	4