

Bangkok Air Quality Dashboard with Streamlit, Prefect & LakeFS

PROJECT DS1321: Near Real-Time Data Pipeline with Visualization

created at may last commit yesterday Total commits 31/year

🔍 Overview

This project, developed for the **DSI321: Big Data Infrastructure** course, delivers a near real-time air quality monitoring and forecasting platform for Bangkok. Using data from Air4Thai's API, the system provides hourly updates and six-hour forecasts for PM2.5 and AQI levels.

It features:

- Orchestrated workflows via **Prefect**
- Data versioning with **LakeFS**
- Interactive visualization using **Streamlit**
- Forecasting with **ARIMA** models
- Fully containerized setup via **Docker**

🔑 Features

- Live ingestion of AQI & PM2.5 from Bangkok stations
- Forecasting (6h ahead) using ARIMA
- Streamlit dashboard with:
 - Station selector
 - Real-time metrics
 - Heatmap visualization
 - AQI/PM2.5 line charts
- Hourly scheduled flows using Prefect
- Data integrity via LakeFS versioning

🛠 Tools & Technologies

Leverages modern open-source tools:

- **Prefect**: Python-based workflow orchestration and scheduling
- **LakeFS**: Git-like version control system for data lakes
- **Streamlit**: Framework for creating interactive dashboards in Python
- **Docker**: Containerization platform to ensure consistent environments
- **JupyterLab**: Notebook interface for data exploration and testing
- **ARIMA**: Statistical time-series forecasting model for AQI and PM2.5

Tech Stack

Component	Tool/Framework
Orchestration	Prefect
Visualization	Streamlit
Forecasting	ARIMA (statsmodels)
Data Versioning	LakeFS
Containerization	Docker, Docker Compose
Dev Environment	JupyterLab
Data Source	Air4Thai API

Data Schema

Column	Type	Description
<code>timestamp</code>	datetime	Measurement time
<code>stationID</code>	string	Unique station code
<code>nameTH</code>	string	Station name in Thai
<code>areaTH</code>	string	Area in Thai
<code>district</code>	string	Bangkok district name
<code>lat, long</code>	float	Geographic coordinates
<code>AQI.aqi</code>	int	AQI value (0–500)
<code>PM25.value</code>	float	PM2.5 in $\mu\text{g}/\text{m}^3$

Data Quality Checks

- ✓ 1,000+ records collected
- ✓ ≥24 hours of data/station
- ✓ >90% field completeness
- ✓ No object dtypes
- ✓ No duplicated rows

Notebook: [check_data_quality.ipynb](#)

Setup Instructions

1. Clone Repository

```
git clone https://github.com/khwkong/dsi321_2025.git
cd dsi321_2025
```

2. Start Services

```
docker-compose up --build -d
```

3. Access Local Services

- LakeFS: <http://localhost:8001>
- Jupyter: <http://localhost:8888>
- Prefect: <http://localhost:4200>
- Streamlit: <http://localhost:8502>

LakeFS Login:

- Username: `access_key`
- Password: `secret_key`

Create a LakeFS repo:

```
lakectl repo create lakefs://dust-concentration
```

4. Upload Data

```
docker exec -it dsi321-jupyter-1 bash  
python upload.py
```

5. Generate Forecasts

```
python getdata.py  
python forecast.py
```

Or use Prefect UI at <http://localhost:4200>

6. Schedule Flows

```
python deploy.py      # ingestion every hour @ min 25  
python deploy_ml.py   # forecast every hour @ min 27
```

Dashboard Preview

รายงานคุณภาพอากาศในกรุงเทพมหานคร

⌚ ข้อมูลล่าสุดเมื่อ: 18/05/2025 23:47:29

ค้นหาสถานที่หรือเบอร์

สำนักงานเขตคลองเตย (คลองเตย)

สำนักงานเขตคลองเตย (คลองเตย)

AQI 38 – Good

PM2.5 – 19.9 $\mu\text{g}/\text{m}^3$

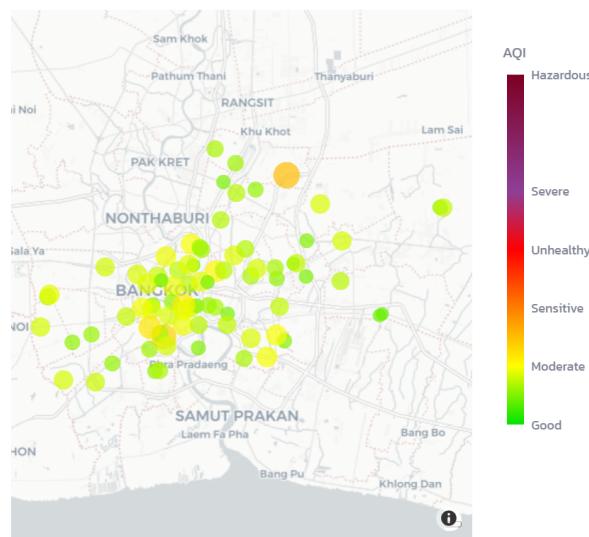
Dashboard

ค่าเฉลี่ยคุณภาพอากาศภายในกรุงเทพฯ

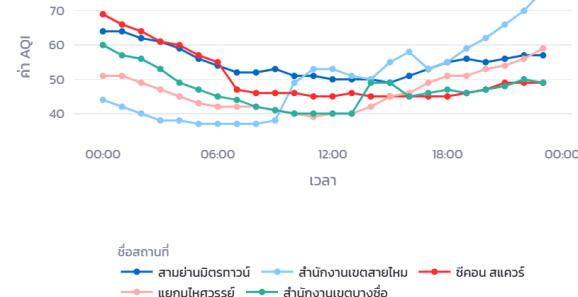
ค่าเฉลี่ย AQI
AQI 37 – Good

ค่าเฉลี่ย PM2.5
19.5 $\mu\text{g}/\text{m}^3$

แผนที่คุณภาพอากาศ



5 สถานที่ในกรุงเทพฯ ที่มีค่า AQI สูงที่สุด (today)



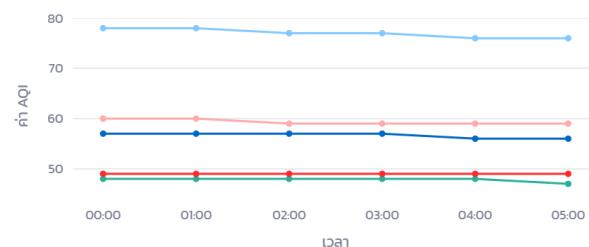
พยากรณ์คุณภาพอากาศล่วงหน้า

เลือกสถานที่

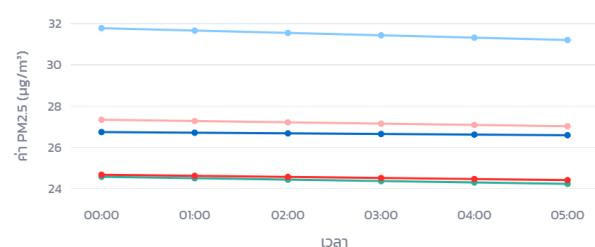
สำนักงานเขตสายไหม × แยกนิ荷วัรดี × สามย่านมีตราชวัน × ชัคอบ สแควร์ × สำนักงานเขตบางซื่อ ×

×

พยากรณ์ AQI



พยากรณ์ PM2.5



สถานที่
● สามย่านมีตราชวัน ● สำนักงานเขตบางซื่อ ● สำนักงานเขตสายไหม
● แยกนิ荷วัรดี ● ชัคอบ สแควร์

สถานที่
● สามย่านมีตราชวัน ● สำนักงานเขตบางซื่อ ● สำนักงานเขตสายไหม
● แยกนิ荷วัรดี ● ชัคอบ สแควร์

ข้อมูลกั้งหมด (ชั่วโมงล่าสุด)

timestamp	nameTH	district	AQI.aqi	PM25.value
18/05/2025 23:00	มหาวิทยาลัยราชภัฏวิ瓜สสเมืองเจ้าพระยา	ธนบุรี	29	16.5
18/05/2025 23:00	รัฐมนตรีทางลงนามเลข 3902	บางขุนเทียน	40	20.9
18/05/2025 23:00	การเคหะแห่งชาติวิภาวดี	ดินแดง	53	25.5
18/05/2025 23:00	โรงพยาบาลกรุงเทพ	มานะนาวา	36	19.1
18/05/2025 23:00	โรงพยาบาลจุฬาลงกรณ์	ปทุมธานี	43	22.2
18/05/2025 23:00	การไฟฟ้าฝ่ายผลิตบี	ธนบุรี	43	22.1
18/05/2025 23:00	สถานีตำรวจนครบาลโชคชัย	วังทองหลาง	44	22.5
18/05/2025 23:00	การเคหะแห่งชาติดินแดง	ดินแดง	45	23.1
18/05/2025 23:00	กรมประชาสัมพันธ์	พญาไท	22	13.4
18/05/2025 23:00	โรงพยาบาลจตุจักร (สิงห์ สิงหเสนี)	วังทองหลาง	34	18.5
18/05/2025 23:00	สำนักงานเขตคลองสาน	มีนบุรี	21	12.4
18/05/2025 23:00	สำนักงานเขตคลองเตย	คลองเตย	43	22.3
18/05/2025 23:00	สำนักงานเขตหนองแขม	หนองแขม	32	17.6

Components

- Real-time readings by station
- Citywide AQI & PM2.5 averages
- AQI heatmap visualization
- Line charts (observed & forecast)
- Interactive table of latest readings

⌚ Forecasting

- Forecast horizon: 6 hours ahead
- ARIMA(1,0,1) per station
- Excludes stations with <24h of data or constant values
- Output saved to:

```
lakefs://dust-concentration/main/forecast/forecast.parquet
```

📁 Repo Structure

```
.
├── data/
│   ├── data.parquet/year=2025/month=5/day=XX/hour=XX/
│   ├── SCHEMA.md
│   └── check_data_quality.ipynb
└── pipeline/
    ├── bangkok_districts.geojson
    └── getdata.py
```

```
├── forecast.py
├── deploy.py
├── deploy_ml.py
└── upload.py
prefect/
├── Dockerfile.jupyter
├── Dockerfile.prefect-worker
├── requirements.txt
└── wait-for-server.sh
visualization/
├── .streamlit/config.toml
└── app.py
.gitignore
README.md
└── docker-compose.yml
└── dashboard_demo.png
```

✉️ Contact

Developer: Panisara Kuiyarat **Email:** kui.panisara@gmail.com