

การได้มาของข้อมูล (Data Collection)

การได้มาของข้อมูล (Data Collection) หมายถึง กระบวนการรวบรวมข้อมูลที่เกี่ยวข้องและเป็นประโยชน์เพื่อนำไปใช้ในการวิจัย การวิเคราะห์ หรือการตัดสินใจในด้านต่าง ๆ โดยการเก็บรวบรวมข้อมูลนั้นสามารถทำได้หลายรูปแบบ ขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการศึกษา

ประเภทของการเก็บรวบรวมข้อมูล

1. ข้อมูลปฐมภูมิ (Primary Data) ข้อมูลที่เก็บรวบรวมขึ้นใหม่โดยตรงจากแหล่งข้อมูล ตัวอย่างวิธีการ:

- การสัมภาษณ์
- การสำรวจ/แบบสอบถาม
- การทดลอง
- การสังเกตการณ์

2. ข้อมูลทุติยภูมิ (Secondary Data) ข้อมูลที่มีการจัดเก็บไว้แล้วจากแหล่งข้อมูลอื่น ตัวอย่างแหล่งข้อมูล:

- รายงานวิจัย
- ฐานข้อมูลออนไลน์
- เอกสารสถิติ
- หนังสือและบทความ

สำหรับ Mini Project นี้ คณะผู้จัดทำเลือกใช้ข้อมูลจากกรมควบคุมมลพิษ อ้างอิงจากลิงก์เว็บ <http://air4thai.pcd.go.th/webV3/#/Home> โดยใช้ข้อมูลคุณภาพอากาศ ในวันที่ 28 มกราคม - 3 กุมภาพันธ์ 2568 อ้างอิงได้จากลิงก์เว็บ <http://air4thai.pcd.go.th/webV3/#/History>

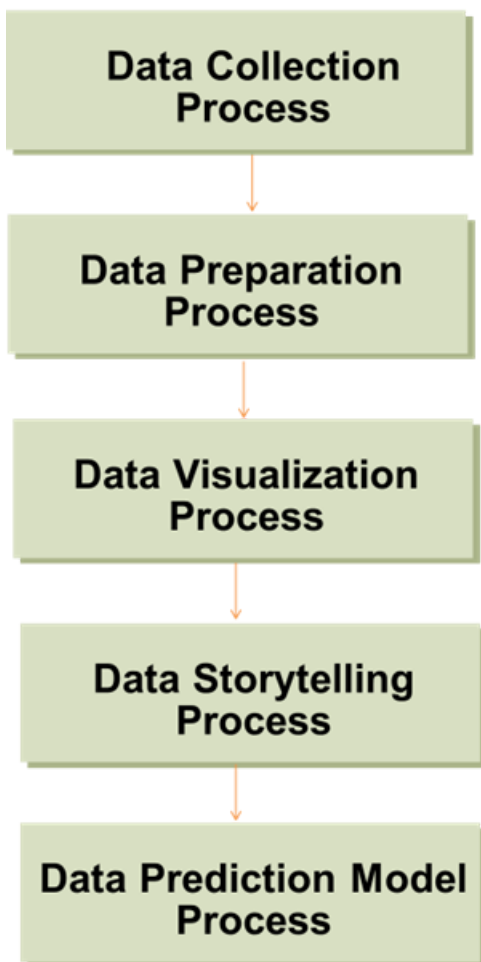
ค่าฝุ่น PM2.5 เป็นหนึ่งในปัจจัยสำคัญที่ส่งผลกระทบต่อคุณภาพอากาศและสุขภาพของประชาชน การศึกษาข้อมูลค่าฝุ่น PM2.5 ในช่วงเวลาสั้น ๆ สามารถให้ข้อมูลเชิงลึกเกี่ยวกับแนวโน้มและปัจจัยที่ส่งผลกระทบต่อระดับมลพิษในระยะสั้น ดังนั้น คณะผู้จัดทำจึงได้นำข้อมูลสถิติค่าฝุ่น PM2.5 ในช่วง 7 วันที่ผ่านมา มาวิเคราะห์และนำเสนอข้อมูลเพื่อศึกษาแนวโน้มและความสัมพันธ์ของค่าฝุ่นกับปัจจัยแวดล้อม เช่น อุณหภูมิ ความชื้น และปริมาณฝุ่น PM10 จุดมุ่งหมายของการศึกษานี้คือเพื่อให้เข้าใจถึงความผันผวนของค่าฝุ่น PM2.5 ในช่วงสั้น ๆ และเป็นแนวทางสำหรับการเฝ้าระวังคุณภาพอากาศในอนาคต

วัตถุประสงค์

1. ส่วนกระบวนการจัดเตรียมข้อมูล
2. ส่วนการนำเสนอข้อมูล
3. ส่วนการพยากรณ์ข้อมูล

- เพื่อนำเสนอข้อมูลค่าฝุ่น PM2.5 ในรูปแบบแผนภูมิต่าง ๆ
- เพื่ออภิปรายผลลัพธ์แผนภูมินำเสนอข้อมูลที่ได้ในเชิงลึก โดยมุ่งเน้นที่แนวโน้มและปัจจัยที่ส่งผลต่อค่าฝุ่น PM2.5
- เพื่อศึกษาโครงสร้างข้อมูลค่าฝุ่น PM2.5 ในช่วง 7 วันล่าสุด
- เพื่อจัดเตรียมข้อมูลค่าฝุ่น PM2.5 และปัจจัยแวดล้อมให้พร้อมต่อการนำเสนอและวิเคราะห์แนวโน้ม
- เพื่อสร้างแบบจำลองพยากรณ์ข้อมูลแนวโน้มค่าฝุ่น PM2.5
- เพื่อหาประสิทธิภาพของแบบจำลองพยากรณ์ที่สร้างขึ้น

วิธีการดำเนินงาน



- ดำเนินการเก็บข้อมูล

การเก็บรวบรวมในครั้งนี้เป็นการเก็บข้อมูลจากข้อมูลทุติยภูมิ (Secondary Data) จากลิงก์เว็บ <http://air4thai.pcd.go.th/webV3/#/History> เขียนคำสั่งในการเก็บรวบรวมข้อมูลดังนี้

```
In [1]: import pandas as pd

url = "https://raw.githubusercontent.com/inkgrix/miniproject-i-DataViz/refs/heads/main/pm2-5.csv"
data = pd.read_csv(url)
data
```

Out[1]:

| | date_time | location | station | pm2_5 | pm10 | temperature |
|----|-----------|---------------|--|-------|-------|-------------|
| 0 | 1/28/2025 | Bangkok | สำนักงานเขตดินแดง | 29.0 | 77.0 | 28.0 |
| 1 | 1/28/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 29.9 | 82.0 | 28.0 |
| 2 | 1/28/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 32.5 | NaN | 29.0 |
| 3 | 1/28/2025 | Nonthaburi | กรมควบคุมโรค | 26.1 | 55.0 | 29.0 |
| 4 | 1/28/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 31.1 | 65.0 | 29.0 |
| 5 | 1/28/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 21.7 | NaN | 28.0 |
| 6 | 1/28/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 45.2 | NaN | 30.0 |
| 7 | 1/28/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 11.3 | 54.0 | NaN |
| 8 | 1/29/2025 | Bangkok | สำนักงานเขตดินแดง | 28.9 | 61.0 | 28.0 |
| 9 | 1/29/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 29.9 | 75.0 | 28.0 |
| 10 | 1/29/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 30.3 | NaN | 28.0 |
| 11 | 1/29/2025 | Nonthaburi | กรมควบคุมโรค | 24.4 | 46.0 | 30.0 |
| 12 | 1/29/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 27.9 | 56.0 | 29.0 |
| 13 | 1/29/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 16.7 | NaN | 28.0 |
| 14 | 1/29/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 33.1 | NaN | 31.0 |
| 15 | 1/29/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 22.5 | 60.0 | NaN |
| 16 | 1/30/2025 | Bangkok | สำนักงานเขตดินแดง | 33.8 | 77.0 | 30.0 |
| 17 | 1/30/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 37.9 | 96.0 | 30.0 |
| 18 | 1/30/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 35.1 | NaN | 30.0 |
| 19 | 1/30/2025 | Nonthaburi | กรมควบคุมโรค | 30.8 | 51.0 | 32.0 |
| 20 | 1/30/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 25.2 | 62.0 | 30.0 |
| 21 | 1/30/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 42.2 | NaN | 30.0 |
| 22 | 1/30/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 43.1 | NaN | 33.0 |
| 23 | 1/30/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 29.6 | 66.0 | NaN |
| 24 | 1/31/2025 | Bangkok | สำนักงานเขตดินแดง | 51.5 | 104.0 | 31.0 |
| 25 | 1/31/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 54.9 | 160.0 | 30.0 |
| 26 | 1/31/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 13.3 | NaN | 32.0 |
| 27 | 1/31/2025 | Nonthaburi | กรมควบคุมโรค | 47.2 | 85.0 | 33.0 |
| 28 | 1/31/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 46.8 | 87.0 | 33.0 |
| 29 | 1/31/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 43.3 | NaN | 30.0 |
| 30 | 1/31/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 63.2 | NaN | 32.0 |
| 31 | 1/31/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 48.2 | 65.0 | NaN |
| 32 | 2/1/2025 | Bangkok | สำนักงานเขตดินแดง | 59.4 | 111.0 | 33.0 |
| 33 | 2/1/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 58.9 | 185.0 | 31.0 |
| 34 | 2/1/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 51.4 | NaN | 33.0 |
| 35 | 2/1/2025 | Nonthaburi | กรมควบคุมโรค | 45.6 | 78.0 | 34.0 |
| 36 | 2/1/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 50.0 | 97.0 | 35.0 |
| 37 | 2/1/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 39.4 | NaN | 32.0 |
| 38 | 2/1/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 79.4 | NaN | 33.0 |
| 39 | 2/1/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 62.9 | 83.0 | NaN |
| 40 | 2/2/2025 | Bangkok | สำนักงานเขตดินแดง | 40.1 | 71.0 | 34.0 |
| 41 | 2/2/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 45.6 | 92.0 | 32.0 |
| 42 | 2/2/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 49.4 | NaN | 34.0 |
| 43 | 2/2/2025 | Nonthaburi | กรมควบคุมโรค | 40.6 | 82.0 | 36.0 |

| | date_time | location | station | pm2_5 | pm10 | temperature |
|----|-----------|---------------|--|-------|------|-------------|
| 44 | 2/2/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 42.0 | 72.0 | 36.0 |
| 45 | 2/2/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 31.8 | NaN | 32.0 |
| 46 | 2/2/2025 | Samut Sakhon | แขวงทางหลวงสมุทรสาคร | 55.3 | NaN | 34.0 |
| 47 | 2/2/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 40.1 | 54.0 | NaN |
| 48 | 2/3/2025 | Bangkok | สำนักงานเขตดินแดง | 42.1 | 77.0 | NaN |
| 49 | 2/3/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 47.4 | 84.0 | 32.0 |
| 50 | 2/3/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 44.5 | NaN | 34.0 |
| 51 | 2/3/2025 | Nonthaburi | กรมควบคุมโรค | 43.0 | 82.0 | 36.0 |
| 52 | 2/3/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 46.3 | 93.0 | 35.0 |
| 53 | 2/3/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 28.6 | NaN | 32.0 |
| 54 | 2/3/2025 | Samut Sakhon | แขวงทางหลวงสมุทรสาคร | 52.7 | NaN | 34.0 |
| 55 | 2/3/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 17.7 | 51.0 | NaN |

- ดำเนินการสำรวจข้อมูล

```
In [2]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 56 entries, 0 to 55
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   date_time    56 non-null     object
1   location     56 non-null     object
2   station      56 non-null     object
3   pm2_5        56 non-null     float64
4   pm10         35 non-null     float64
5   temperature  48 non-null     float64
dtypes: float64(3), object(3)
memory usage: 2.8+ KB
```

คำอธิบายสารสนเทศข้อมูล

ชุดข้อมูลค่าฝุ่น PM2.5 ในช่วง 7 วันล่าสุดมีรายละเอียดดังนี้

- 1. มีจำนวนรายการข้อมูล (รายการ) จำนวน **56** แถว
- 2. มีจำนวนมิติข้อมูลหรือคอลัมน์ จำนวน **6** คอลัมน์ โดยคอลัมน์ทั้งหมดประกอบด้วยข้อมูลประเภท ตัวเลข (float64) และ ตัวอักษร (object) ซึ่งพบว่าบางคอลัมน์มีค่าว่าง ตัวอย่างเช่น คอลัมน์ 'pm10' มีข้อมูลเพียง 35 รายการ และคอลัมน์ 'temperature' มีข้อมูลเพียง 48 รายการ

```
In [3]: data.describe()
```

| | pm2_5 | pm10 | temperature |
|-------|-----------|------------|-------------|
| count | 56.000000 | 35.000000 | 48.000000 |
| mean | 39.300000 | 79.885714 | 31.479167 |
| std | 13.487058 | 28.312096 | 2.396714 |
| min | 11.300000 | 46.000000 | 28.000000 |
| 25% | 29.825000 | 61.500000 | 30.000000 |
| 50% | 40.350000 | 77.000000 | 32.000000 |
| 75% | 47.250000 | 86.000000 | 33.000000 |
| max | 79.400000 | 185.000000 | 36.000000 |

คำอธิบาย

จากการสำรวจค่าทางสถิติของข้อมูลด้วยคำสั่ง describe() พบว่ามีมิติข้อมูล 'pm2_5' มีค่ากลางเท่ากับ 39.30 ค่าสูงสุด 79.40 ค่าต่ำสุด 11.30

การจัดเตรียมข้อมูล (Data Preparation)

In [4]: `data`

Out[4]:

| | date_time | location | station | pm2_5 | pm10 | temperature |
|----|-----------|---------------|--|-------|-------|-------------|
| 0 | 1/28/2025 | Bangkok | สำนักงานเขตดินแดง | 29.0 | 77.0 | 28.0 |
| 1 | 1/28/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 29.9 | 82.0 | 28.0 |
| 2 | 1/28/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 32.5 | NaN | 29.0 |
| 3 | 1/28/2025 | Nonthaburi | กรมควบคุมโรค | 26.1 | 55.0 | 29.0 |
| 4 | 1/28/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 31.1 | 65.0 | 29.0 |
| 5 | 1/28/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 21.7 | NaN | 28.0 |
| 6 | 1/28/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 45.2 | NaN | 30.0 |
| 7 | 1/28/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 11.3 | 54.0 | NaN |
| 8 | 1/29/2025 | Bangkok | สำนักงานเขตดินแดง | 28.9 | 61.0 | 28.0 |
| 9 | 1/29/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 29.9 | 75.0 | 28.0 |
| 10 | 1/29/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 30.3 | NaN | 28.0 |
| 11 | 1/29/2025 | Nonthaburi | กรมควบคุมโรค | 24.4 | 46.0 | 30.0 |
| 12 | 1/29/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 27.9 | 56.0 | 29.0 |
| 13 | 1/29/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 16.7 | NaN | 28.0 |
| 14 | 1/29/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 33.1 | NaN | 31.0 |
| 15 | 1/29/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 22.5 | 60.0 | NaN |
| 16 | 1/30/2025 | Bangkok | สำนักงานเขตดินแดง | 33.8 | 77.0 | 30.0 |
| 17 | 1/30/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 37.9 | 96.0 | 30.0 |
| 18 | 1/30/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 35.1 | NaN | 30.0 |
| 19 | 1/30/2025 | Nonthaburi | กรมควบคุมโรค | 30.8 | 51.0 | 32.0 |
| 20 | 1/30/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 25.2 | 62.0 | 30.0 |
| 21 | 1/30/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 42.2 | NaN | 30.0 |
| 22 | 1/30/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 43.1 | NaN | 33.0 |
| 23 | 1/30/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 29.6 | 66.0 | NaN |
| 24 | 1/31/2025 | Bangkok | สำนักงานเขตดินแดง | 51.5 | 104.0 | 31.0 |
| 25 | 1/31/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 54.9 | 160.0 | 30.0 |
| 26 | 1/31/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 13.3 | NaN | 32.0 |
| 27 | 1/31/2025 | Nonthaburi | กรมควบคุมโรค | 47.2 | 85.0 | 33.0 |
| 28 | 1/31/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 46.8 | 87.0 | 33.0 |
| 29 | 1/31/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 43.3 | NaN | 30.0 |
| 30 | 1/31/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 63.2 | NaN | 32.0 |
| 31 | 1/31/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 48.2 | 65.0 | NaN |
| 32 | 2/1/2025 | Bangkok | สำนักงานเขตดินแดง | 59.4 | 111.0 | 33.0 |
| 33 | 2/1/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 58.9 | 185.0 | 31.0 |
| 34 | 2/1/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 51.4 | NaN | 33.0 |
| 35 | 2/1/2025 | Nonthaburi | กรมควบคุมโรค | 45.6 | 78.0 | 34.0 |
| 36 | 2/1/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 50.0 | 97.0 | 35.0 |
| 37 | 2/1/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 39.4 | NaN | 32.0 |
| 38 | 2/1/2025 | Samut Sakhon | แขวงการทางสมุทรสาคร | 79.4 | NaN | 33.0 |
| 39 | 2/1/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 62.9 | 83.0 | NaN |
| 40 | 2/2/2025 | Bangkok | สำนักงานเขตดินแดง | 40.1 | 71.0 | 34.0 |
| 41 | 2/2/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 45.6 | 92.0 | 32.0 |
| 42 | 2/2/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 49.4 | NaN | 34.0 |
| 43 | 2/2/2025 | Nonthaburi | กรมควบคุมโรค | 40.6 | 82.0 | 36.0 |

| | date_time | location | station | pm2_5 | pm10 | temperature |
|----|-----------|---------------|--|-------|------|-------------|
| 44 | 2/2/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 42.0 | 72.0 | 36.0 |
| 45 | 2/2/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 31.8 | NaN | 32.0 |
| 46 | 2/2/2025 | Samut Sakhon | แขวงทางหลวงสมุทรสาคร | 55.3 | NaN | 34.0 |
| 47 | 2/2/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 40.1 | 54.0 | NaN |
| 48 | 2/3/2025 | Bangkok | สำนักงานเขตดินแดง | 42.1 | 77.0 | NaN |
| 49 | 2/3/2025 | Bangkok | กรมอุตุนิยมวิทยาบางนา | 47.4 | 84.0 | 32.0 |
| 50 | 2/3/2025 | Bangkok | มหาวิทยาลัยเกษตรศาสตร์ | 44.5 | NaN | 34.0 |
| 51 | 2/3/2025 | Nonthaburi | กรมควบคุมโรค | 43.0 | 82.0 | 36.0 |
| 52 | 2/3/2025 | Pathumthani | มหาวิทยาลัยกรุงเทพ | 46.3 | 93.0 | 35.0 |
| 53 | 2/3/2025 | Samut Prakan | บ้านพักกรมอุตสาหกรรมพื้นฐานและการเหมืองแร่ | 28.6 | NaN | 32.0 |
| 54 | 2/3/2025 | Samut Sakhon | แขวงทางหลวงสมุทรสาคร | 52.7 | NaN | 34.0 |
| 55 | 2/3/2025 | Nakhon Pathom | อ่างเก็บน้ำประปา | 17.7 | 51.0 | NaN |

นำเสนอข้อมูลด้วยภาพ

1. การจัดเตรียม

- จัดเตรียม Font และข้อมูลก่อนการนำเสนอ

ในการนำเสนอข้อมูลในครั้งนี้มีภาษาไทยเป็นหลัก จึงเรียกใช้งานรูปแบบ Font ชื่อ Sriracha-Regular.ttf จึงได้ดำเนินการ upload และกำหนดเป็นรูปแบบหลักในการนำเสนอ

- Sriracha-Regular.ttf

```
In [5]: import matplotlib as mpl
import matplotlib.pyplot as plt

mpl.font_manager.fontManager.addfont('./font/Sriracha-Regular.ttf')
mpl.rc('font', family='Sriracha')
mpl.rcParams.update({'font.size': 15})
```

- เรียกใช้งาน Library ที่เกี่ยวข้อง

```
In [6]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

- สํารวจข้อมูลเบื้องต้น

```
In [7]: g = sns.PairGrid(data,
                        hue = "location",
                        palette = "YlOrRd")

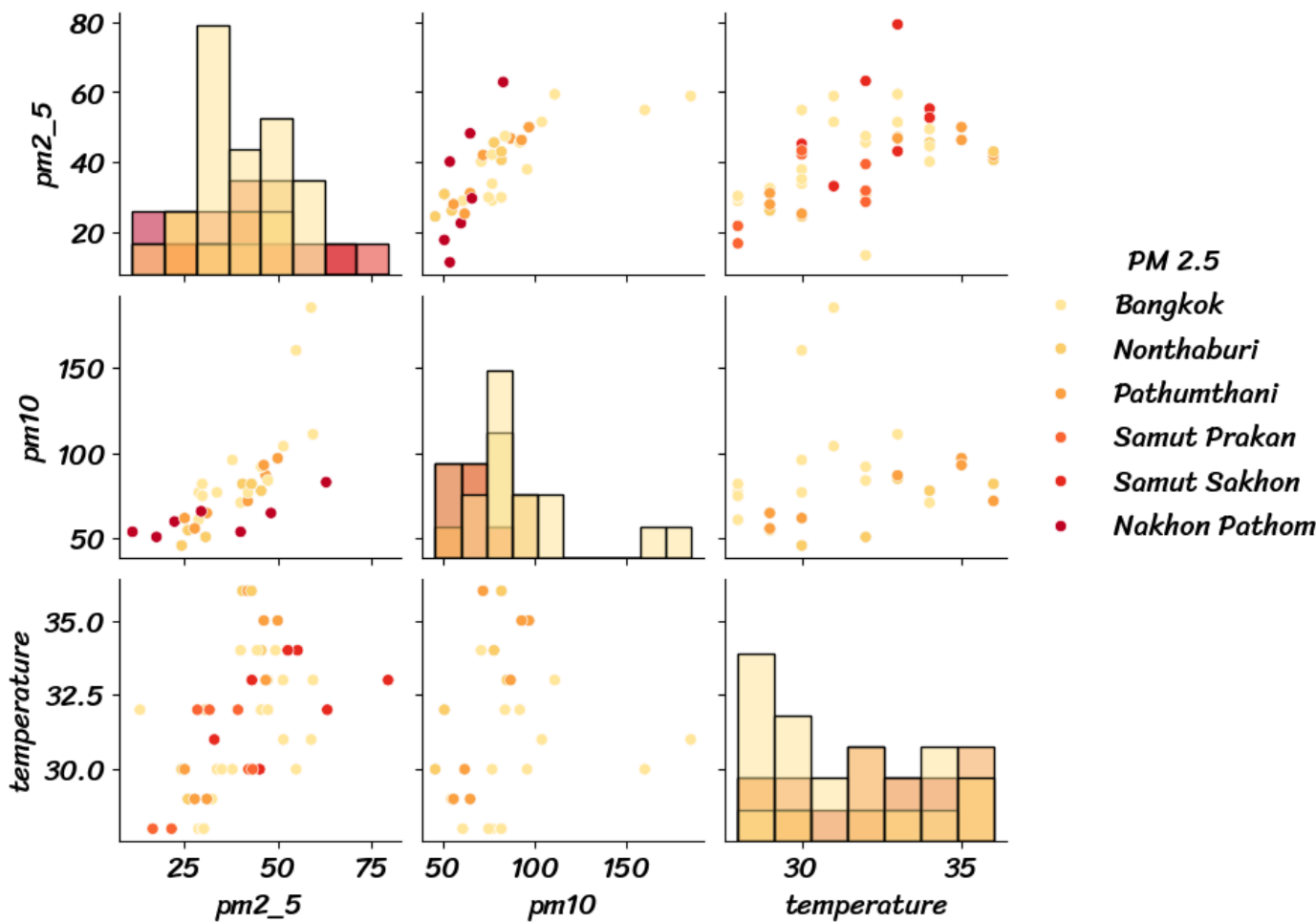
g.map_diag(sns.histplot)
g.map_offdiag(sns.scatterplot)

g.fig.suptitle("แผนภูมิสำรวจรูปแบบความสัมพันธ์ของข้อมูลเบื้องต้น\n (Exploratory Data Analysis)",
               y = 1.02,
               fontsize = 16)

g.add_legend()
g.legend.set_title("PM 2.5")

g.tight_layout()
plt.show()
```

แผนภูมิสำรวจรูปแบบความสัมพันธ์ของข้อมูลเบื้องต้น (Exploratory Data Analysis)



คำอธิบาย การวิเคราะห์ข้อมูลเบื้องต้น (Exploratory Data Analysis) โดยใช้แผนภูมิ Pair Plot เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรทั้งหมดในชุดข้อมูล ผ่านกราฟกระจาย (scatter plot) และฮิสโตแกรม (histogram) สำหรับแต่ละตัวแปร

รายละเอียดของแผนภูมิ

แกนแนวดิ่งและแนวนอน:

แต่ละแกนแสดงตัวแปร เช่น pm2_5, pm10, temperature และ location

กราฟกระจาย (Scatter Plot)

แสดงความสัมพันธ์ระหว่างคู่ของตัวแปร ถ้ามีการเรียงตัวในทิศทางที่ชัดเจนแสดงถึงความสัมพันธ์ที่ดี หากกระจายแบบสุ่มแสดงถึงความสัมพันธ์ที่น้อยหรือไม่มีเลย

ฮิสโตแกรม (Histogram)

แสดงการกระจายของแต่ละตัวแปรในแถวทแยงมุมหลัก โดยการกระจายแสดงถึงลักษณะของข้อมูล

การกระจายข้อมูล

แสดงการกระจายที่อาจมีความเอนเอียง (skewed distribution) สำหรับบางตัวแปร เช่น pm10 และ temperature ที่กระจุกตัวอยู่ในช่วงค่าต่ำ

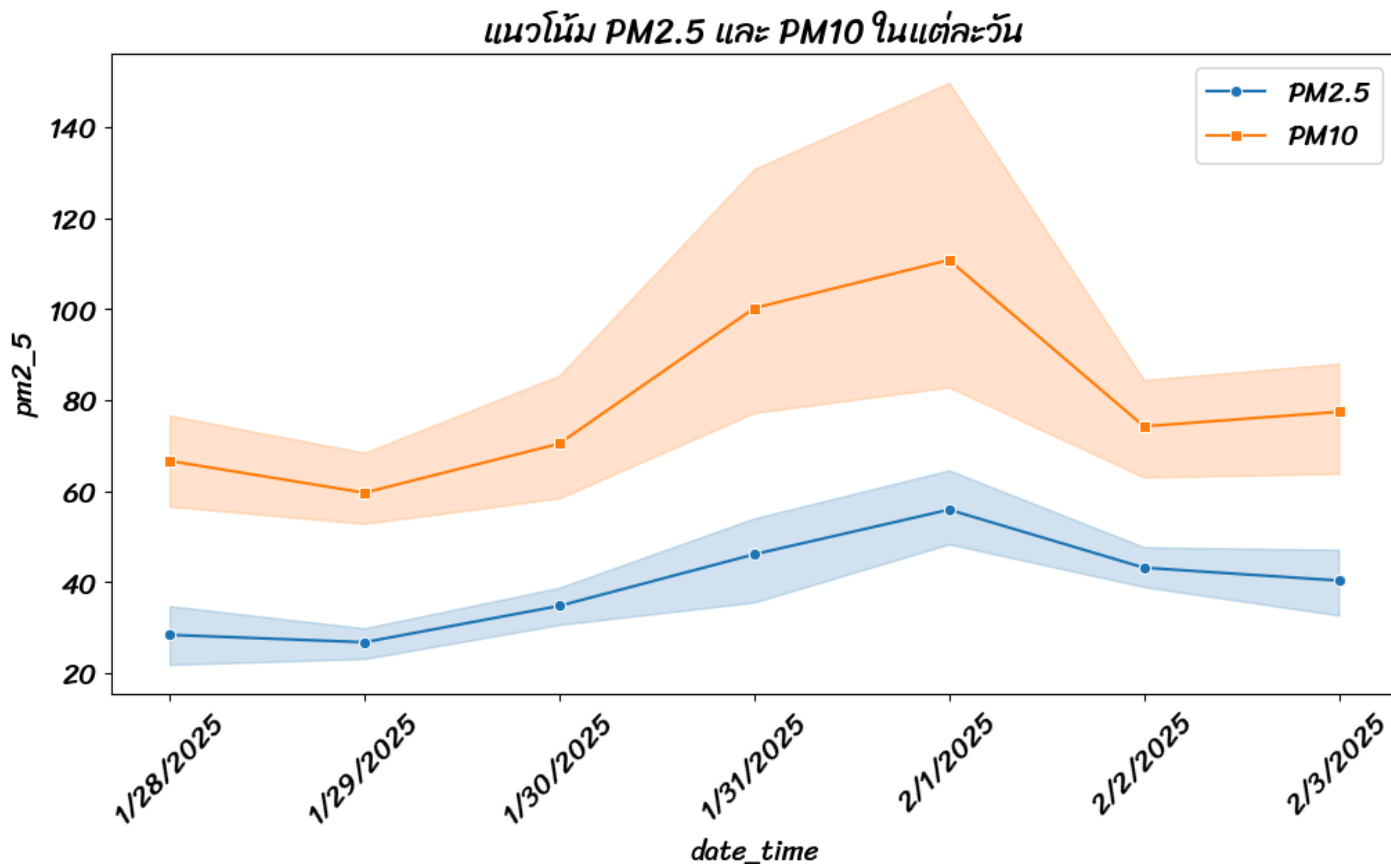
ความสัมพันธ์เชิงเส้น

ตัวแปร pm2_5, pm10 และ temperature แสดงถึงความสัมพันธ์ที่มีแนวโน้มเป็นเชิงบวก โดยจุดกระจายเรียงตัวเป็นเส้นทแยงมุม

สรุปแล้ว แผนภูมิ Pair Plot ใช้เพื่อแสดงภาพรวมความสัมพันธ์และการกระจายข้อมูลของแต่ละตัวแปร ซึ่งเป็นเครื่องมือสำคัญในการวิเคราะห์ข้อมูลเบื้องต้นก่อนการทำการวิเคราะห์เชิงลึกหรือการสร้างโมเดลต่อไป

2.5 การพิจารณาแนวโน้ม PM2.5 และ PM10 ในแต่ละวัน


```
In [8]: plt.figure(figsize=(12,6))
sns.lineplot(data=data, x="date_time", y="pm2_5", label="PM2.5", marker="o")
sns.lineplot(data=data, x="date_time", y="pm10", label="PM10", marker="s")
plt.xticks(rotation=45)
plt.title("แนวโน้ม PM2.5 และ PM10 ในแต่ละวัน")
plt.legend()
plt.show()
```



คำอธิบาย

การนำเสนอข้อมูลในรูปแบบแผนภูมิเส้น (line plot) โดยใช้คำสั่ง `sns.lineplot()` เพื่อแสดงให้เห็นการเปลี่ยนแปลงของค่าฝุ่น PM2.5 และ PM10 ในแต่ละวัน และทำให้สามารถเปรียบเทียบแนวโน้มของทั้งสองตัวแปรได้อย่างชัดเจน

```
In [9]: import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from mpl_toolkits.mplot3d import Axes3D

colors = data["pm2_5"]

fig = plt.figure(figsize=(11, 8))
ax = fig.add_subplot(111, projection='3d')

sc = ax.scatter(data["pm2_5"], data["pm10"], data["temperature"],
                c=colors, cmap="viridis", alpha=0.7, edgecolors='k', s=60) # เพิ่มขนาดจุด

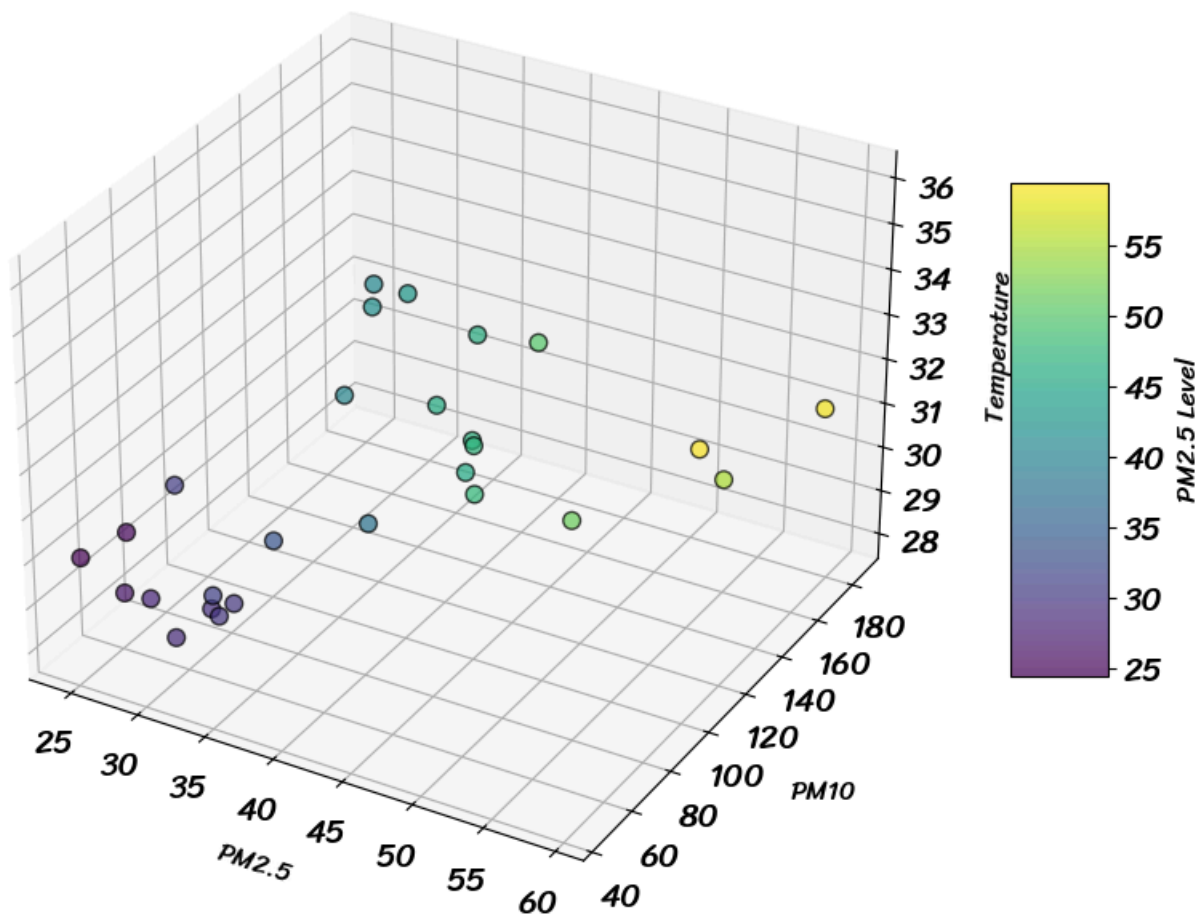
ax.set_xlabel("PM2.5", fontsize=12, labelpad=10)
ax.set_ylabel("PM10", fontsize=12, labelpad=15)
ax.set_zlabel("Temperature", fontsize=12, labelpad=10)
ax.set_title("3D Scatter Plot ของ PM2.5, PM10 และอุณหภูมิ", fontsize=15)

# ขยับกราฟไปทางซ้าย
pos = ax.get_position()
ax.set_position([pos.x0 - 0.05, pos.y0, pos.width, pos.height])

cbar = fig.colorbar(sc, ax=ax, shrink=0.5, aspect=5)
cbar.set_label("PM2.5 Level", fontsize=12)

plt.show()
```

3D Scatter Plot ของ PM2.5, PM10 และอุณหภูมิ



คำอธิบาย

ในภาพ 3D Scatter Plot นี้ แสดงข้อมูลเกี่ยวกับความสัมพันธ์ระหว่างค่าฝุ่น PM2.5, PM10 และอุณหภูมิ โดยแสดงใน 3 แกนดังนี้:

แกน X (PM2.5): แสดงค่าฝุ่น PM2.5 ในแต่ละช่วงเวลา โดยค่าฝุ่นจะเพิ่มจากซ้ายไปขวาบนแกนนี้ แกน Y (PM10): แสดงค่าฝุ่น PM10 โดยค่าของมันจะแสดงการเปลี่ยนแปลงในลักษณะคล้ายกับ PM2.5 บนแกน Y แกน Z (Temperature): แสดงอุณหภูมิในแต่ละช่วงเวลา โดยค่าอุณหภูมิจะถูกแสดงบนแกน Z

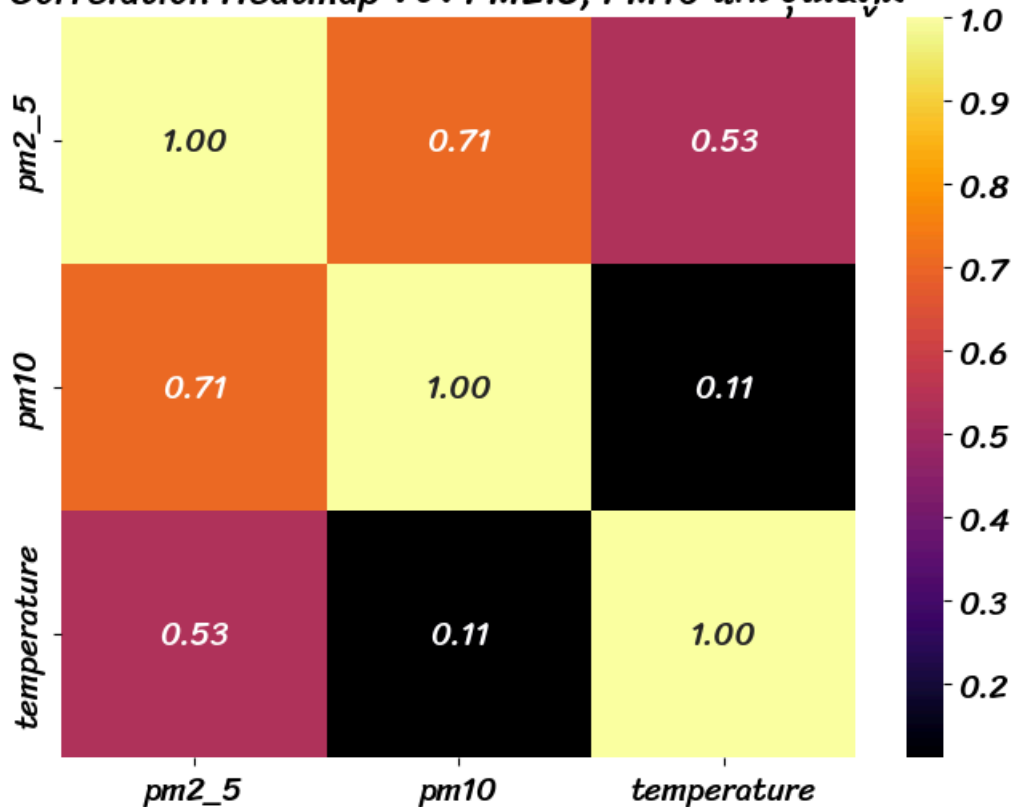
สีของจุด (Colormap): จุดในกราฟจะถูกแสดงด้วยสีที่แตกต่างกัน โดยใช้สีจาก cmap="viridis", สีวงที่ปลายแกน X หมายถึงค่าต่ำ และสีเหลืองที่ปลายแกน Z หมายถึงค่ามาก เพื่อแสดงระดับความเข้มข้นของค่าฝุ่น PM2.5

การแสดงผลในกราฟ:

กราฟนี้ช่วยให้สามารถสังเกตเห็นการกระจายตัวและแนวโน้มของค่าฝุ่น PM2.5 และ PM10 พร้อมกับอุณหภูมิในแต่ละวัน โดยสามารถเห็นว่ามีความสัมพันธ์กันอย่างไรในมิติสามมิติ สีของจุดในกราฟแสดงถึงระดับความเข้มข้นของฝุ่น PM2.5 ทำให้สามารถสังเกตเห็นลักษณะความสัมพันธ์ระหว่างข้อมูลได้อย่างชัดเจน

```
In [10]: plt.figure(figsize=(8,6))
sns.heatmap(data[["pm2_5", "pm10", "temperature"]].corr(), annot=True, cmap="inferno", fmt=".2f")
plt.title("Correlation Heatmap ของ PM2.5, PM10 และอุณหภูมิ")
plt.show()
```

Correlation Heatmap ของ PM2.5, PM10 และอุณหภูมิ



คำอธิบาย

ในกราฟนี้แสดง Correlation Heatmap ของค่าฝุ่น PM2.5, PM10 และอุณหภูมิ ซึ่งช่วยในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

กราฟนี้แสดงให้เห็นว่า PM2.5 และ PM10 มีความสัมพันธ์กันในเชิงบวก (ค่าสูงสุดใกล้ 1) ซึ่งหมายความว่าเมื่อค่า PM2.5 เพิ่มขึ้น ค่า PM10 ก็มีแนวโน้มที่จะเพิ่มขึ้นเช่นกัน\

อุณหภูมิ (Temperature) มีความสัมพันธ์ที่ต่ำกว่ากับทั้ง PM2.5 และ PM10 ซึ่งแสดงว่าความสัมพันธ์ระหว่างอุณหภูมิและค่าฝุ่นนั้นไม่สูงมาก

สรุป

กราฟนี้ช่วยในการเข้าใจความสัมพันธ์ระหว่างตัวแปรในชุดข้อมูลและแสดงความสัมพันธ์ระหว่างค่าฝุ่น PM2.5, PM10 และอุณหภูมิในรูปแบบที่สามารถเข้าใจได้ง่ายผ่านสีและตัวเลขในกราฟ