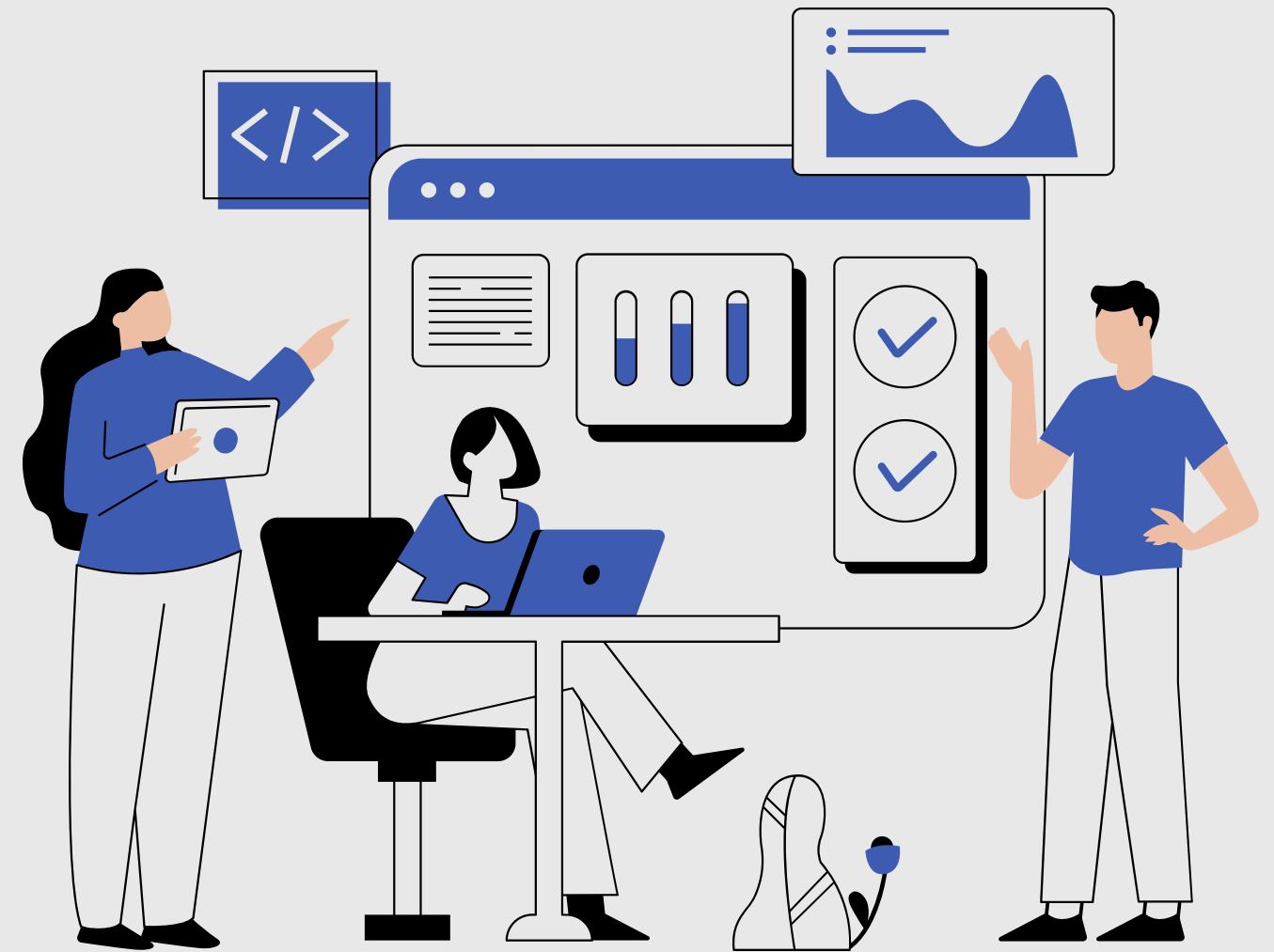


# Mexico Air Pollution (2015 - 2021)

204203 Data analysis project



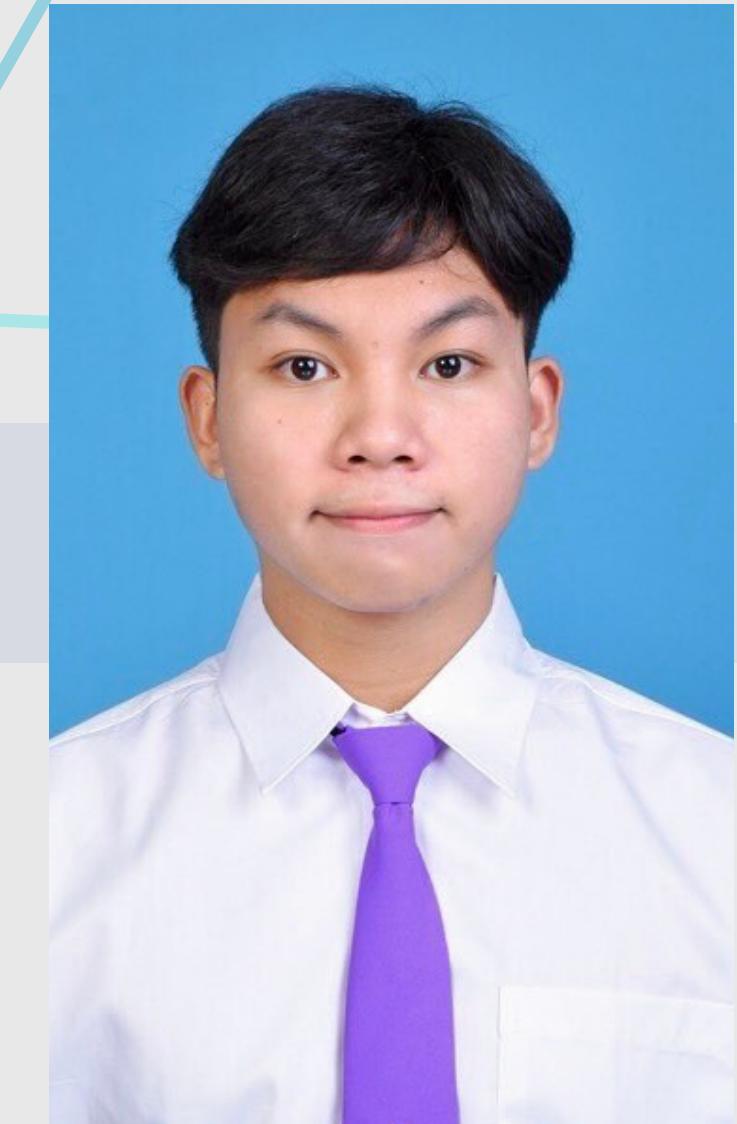
# Member List



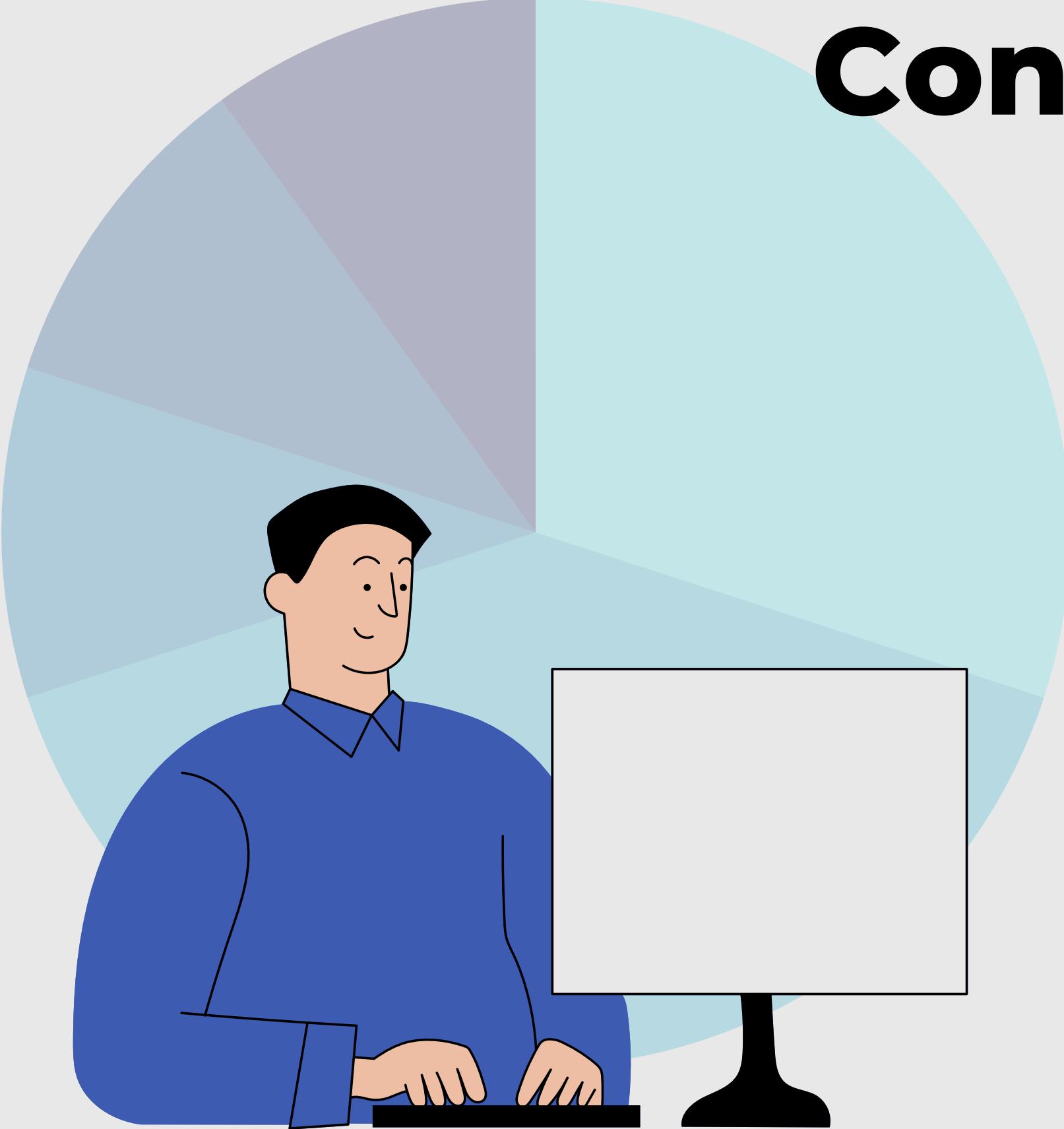
**Name :** ชนากรณ์ ศรีวารมย์  
**ID :** 650510604



**Name :** อภิวิชญ์ ศุภลัษณ์  
**ID :** 650510648



**Name :** จิณณวัตร จันทร์งาม  
**ID :** 650510656



# Content list

- 1. Data Preparation

---

- 2. Statistical Data

---

- 3. Exploration & Visualization

---

- 4. Prediction

---

- 5. Conclusion

---

# Data Preparation

## Data Management - stations\_daily



1. ตรวจสอบความถูกต้องของข้อมูลเบื้องต้น
2. เพิ่มคอลัมน์ month และ year ตามเดือนและปีของข้อมูลแต่ละตัว เพื่อให้ง่ายต่อการค้นหา
3. เปลี่ยนข้อมูลจากคอลัมน์ month และ year ให้เป็นตัวเลข และจัดเรียงข้อมูลตามเดือนและปี
4. ตรวจสอบจำนวนของข้อมูลที่มีในแต่ละปี
5. ตัดปีที่มีข้อมูลของ PM2.5 น้อยกว่าร้อยละ 50
6. ทำการตัดคอลัมน์ที่มีข้อมูลน้อยกว่าร้อยละ 60

# Data Preparation

## Cleansing the outliers - stations\_daily

หลังจากการตัดปีกีมีข้อมูลน้อย และคอลัมน์ที่ข้อมูลมีน้อยออกแล้วนั้น  
พบว่า ยังคงมีข้อมูลที่เป็น outlier และ missing value หลงเหลืออยู่

```
PM2.5      33513
PM10      31938
NOx       49962
O3        27171
CO        33670
HR        40067
NO        54089
NO2      39912
TMP       36498
dtype: int64
```

	PM2.5	PM10	NOx	O3	CO	HR	NO	NO2	TMP
count	1.271940e+05	1.287690e+05	1.107450e+05	133536.000000	127037.000000	1.206400e+05	1.066180e+05	1.207950e+05	124209.000000
mean	2.830052e+33	1.832171e+33	2.906423e+01	0.548284	3.340298	2.316232e+02	3.003726e+01	2.654576e+01	24.307164
std	6.201747e+35	4.357159e+35	9.585072e+03	145.560632	276.000146	5.077423e+04	9.768822e+03	9.177681e+03	633.620598
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
25%	1.214669e+01	2.994232e+01	1.420000e-02	0.018996	0.500000	4.351037e+01	4.577356e-03	8.558333e-03	15.981167
50%	1.866534e+01	4.491667e+01	2.389649e-02	0.025971	0.753667	5.745455e+01	9.150156e-03	1.463492e-02	19.233333
75%	2.714147e+01	6.439130e+01	3.914257e-02	0.034231	1.150000	6.900000e+01	1.791667e-02	2.265548e-02	22.983542
max	1.806160e+38	1.412893e+38	3.189756e+06	53185.473851	84087.353799	1.736131e+07	3.189756e+06	3.189756e+06	202409.533333

จำนวน missing value  
ของแต่ละคอลัมน์

# Data Preparation

## Cleansing the outliers - stations\_daily



1. ทำการกำจัดข้อมูลแต่ละที่น่าจะมีค่าสูงกว่าความเป็นจริง โดยอ้างอิงมาจาก <https://aqicn.org/calculator>
  2. กำจัดข้อมูลที่มีค่าสูงกว่าค่า upper bound และ ข้อมูลที่มีค่าต่ำกว่าค่า lower bound
  3. กำจัด missing value โดยการ fillna ด้วยค่าเฉลี่ยของแต่ละเดือนนั้นของทุกปี
- 2.1 คำนวณค่า first และ third quartile ของข้อมูล
  - 2.2 คำนวณค่า IQR จากสูตร  $Q3 - Q1$
  - 2.3 คำนวณ upper bound และ lower bound เป็นเกณฑ์ที่ใช้ในการตัดข้อมูลที่ โดยใช้สูตร  $Q3 + (1.5 * IQR)$  และ  $Q1 - (1.5 * IQR)$  ตามลำดับ
  - 2.4 กำจัดข้อมูลมีค่าสูงกว่า upper bound หรือต่ำกว่า lower bound

# Data Preparation

## Data Management - stations\_rsinaica

1. ตัดคอลัมน์ station\_id, latitude, longitude, network\_name มาทำเป็น dataframe อันใหม่
2. นำไป merge กับ ข้อมูลรายวันที่ clean แล้ว เพื่อสามารถนำข้อมูลนี้ไปใช้วิเคราะห์ข้อมูลที่เกี่ยวข้องกับพื้นที่ได้

# Statistical Data

Before cleaning :

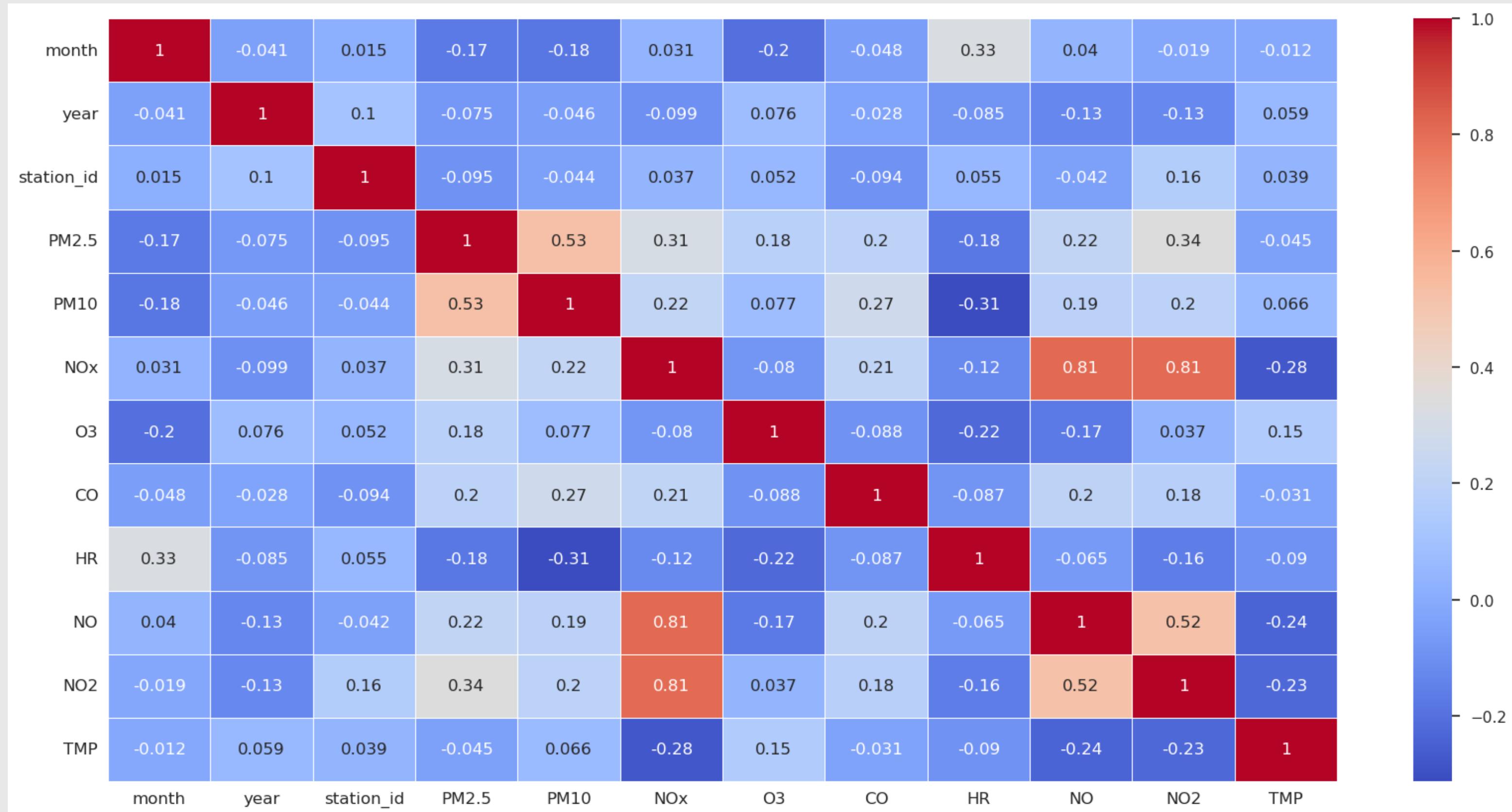
	PM2.5	PM10	NOx	O3	CO	HR	NO	NO2	TMP
count	1.271940e+05	1.287690e+05	1.107450e+05	133536.000000	127037.000000	1.206400e+05	1.066180e+05	1.207950e+05	124209.000000
mean	2.830052e+33	1.832171e+33	2.906423e+01	0.548284	3.340298	2.316232e+02	3.003726e+01	2.654576e+01	24.307164
std	6.201747e+35	4.357159e+35	9.585072e+03	145.560632	276.000146	5.077423e+04	9.768822e+03	9.177681e+03	633.620598
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
25%	1.214669e+01	2.994232e+01	1.420000e-02	0.018996	0.500000	4.351037e+01	4.577356e-03	8.558333e-03	15.981167
50%	1.866534e+01	4.491667e+01	2.389649e-02	0.025971	0.753667	5.745455e+01	9.150156e-03	1.463492e-02	19.233333
75%	2.714147e+01	6.439130e+01	3.914257e-02	0.034231	1.150000	6.900000e+01	1.791667e-02	2.265548e-02	22.983542
max	1.806160e+38	1.412893e+38	3.189756e+06	53185.473851	84087.353799	1.736131e+07	3.189756e+06	3.189756e+06	202409.533333

After cleaning :

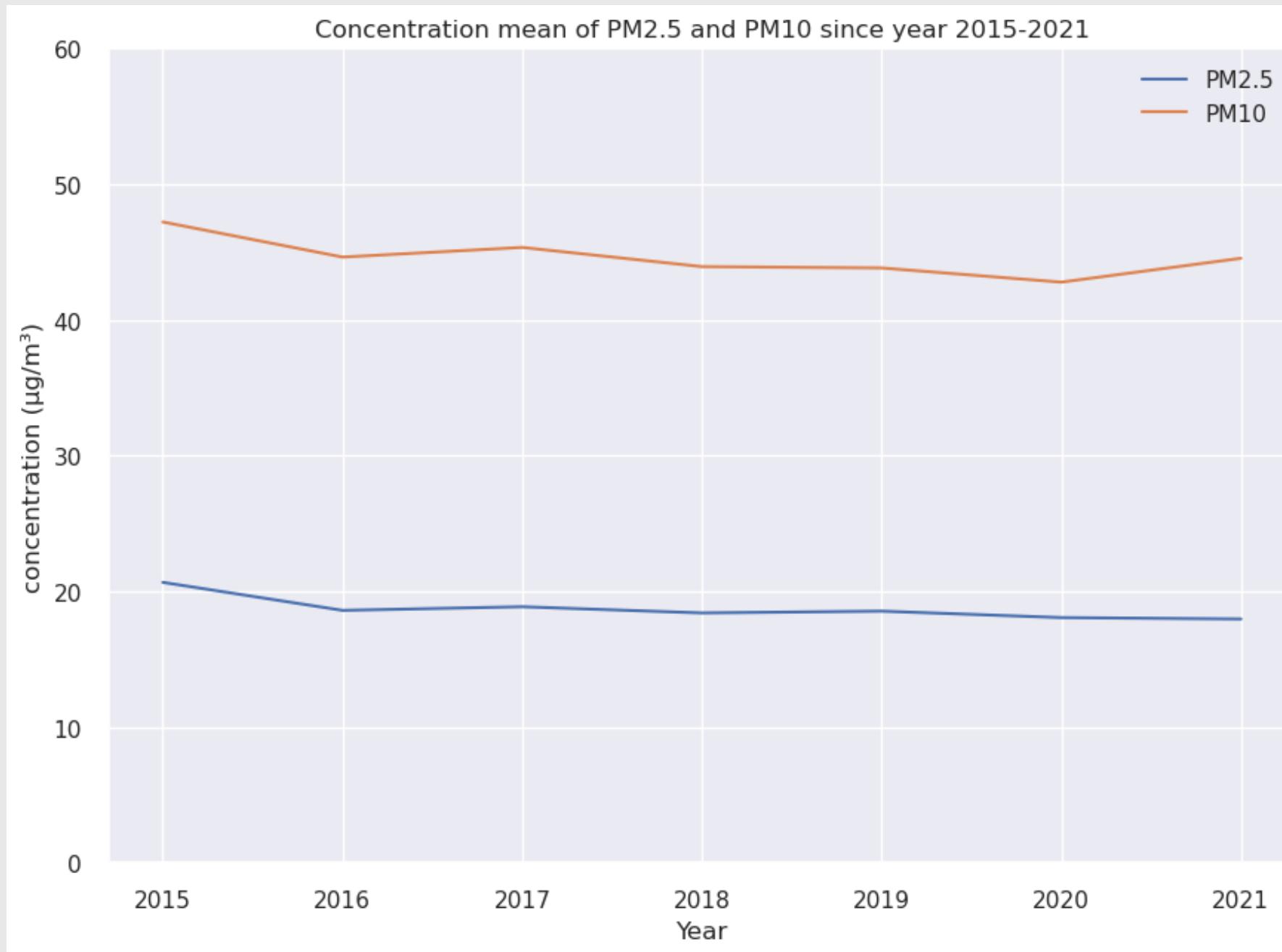
	PM2.5	PM10	NOx	O3	CO	HR	NO	NO2	TMP
count	126341.000000	126341.000000	1.263410e+05	1.263410e+05	1.263410e+05	126341.000000	1.263410e+05	1.263410e+05	126341.000000
mean	18.658039	44.476259	2.585125e-02	2.730267e-02	7.661861e-01	56.273930	1.104895e-02	1.578575e-02	19.475721
std	8.430780	19.217498	1.287838e-02	9.898151e-03	3.600936e-01	15.205113	6.801604e-03	7.886216e-03	4.495546
min	0.000313	0.007538	1.490278e-07	8.109665e-07	5.877500e-39	5.844271	1.298958e-07	8.509144e-08	5.582705
25%	13.312500	32.685951	1.840417e-02	2.145833e-02	5.527917e-01	47.063709	6.626087e-03	1.081905e-02	16.473913
50%	17.583333	42.571429	2.371212e-02	2.660274e-02	7.199639e-01	58.665208	9.789497e-03	1.495367e-02	19.431467
75%	22.776875	54.106826	3.249843e-02	3.363174e-02	9.182917e-01	66.019391	1.414005e-02	1.962632e-02	21.941000
max	48.208333	112.277208	7.688696e-02	5.696278e-02	2.089167e+00	100.000000	3.795833e-02	4.400417e-02	33.312103

# Exploration & Visualization

## Heatmap Correlation



# Trends of PM2.5 and PM10



# แหล่งที่มาของแก๊สต่างๆที่มีผลต่อ PM2.5 และ PM10

## NOx

- เกิดจากการเผาไหม้เชื้อเพลิงในอุณหภูมิสูง
- แหล่งที่มา ยานพาหนะต่างๆ, โรงไฟฟ้า, หม้อไอน้ำที่ใช้ในโรงงานอุตสาหกรรม

## CO

- เกิดจากการเผาไหม้คาร์บอนที่ไม่สมบูรณ์ในเชื้อเพลิงต่างๆ
- เช่น พลาสติก การเผาถ่าน เตาเผา ไฟไหม้ เครื่องยนต์ สารเคมีหรือปฏิกิริยาเคมีบางชนิด

## NO

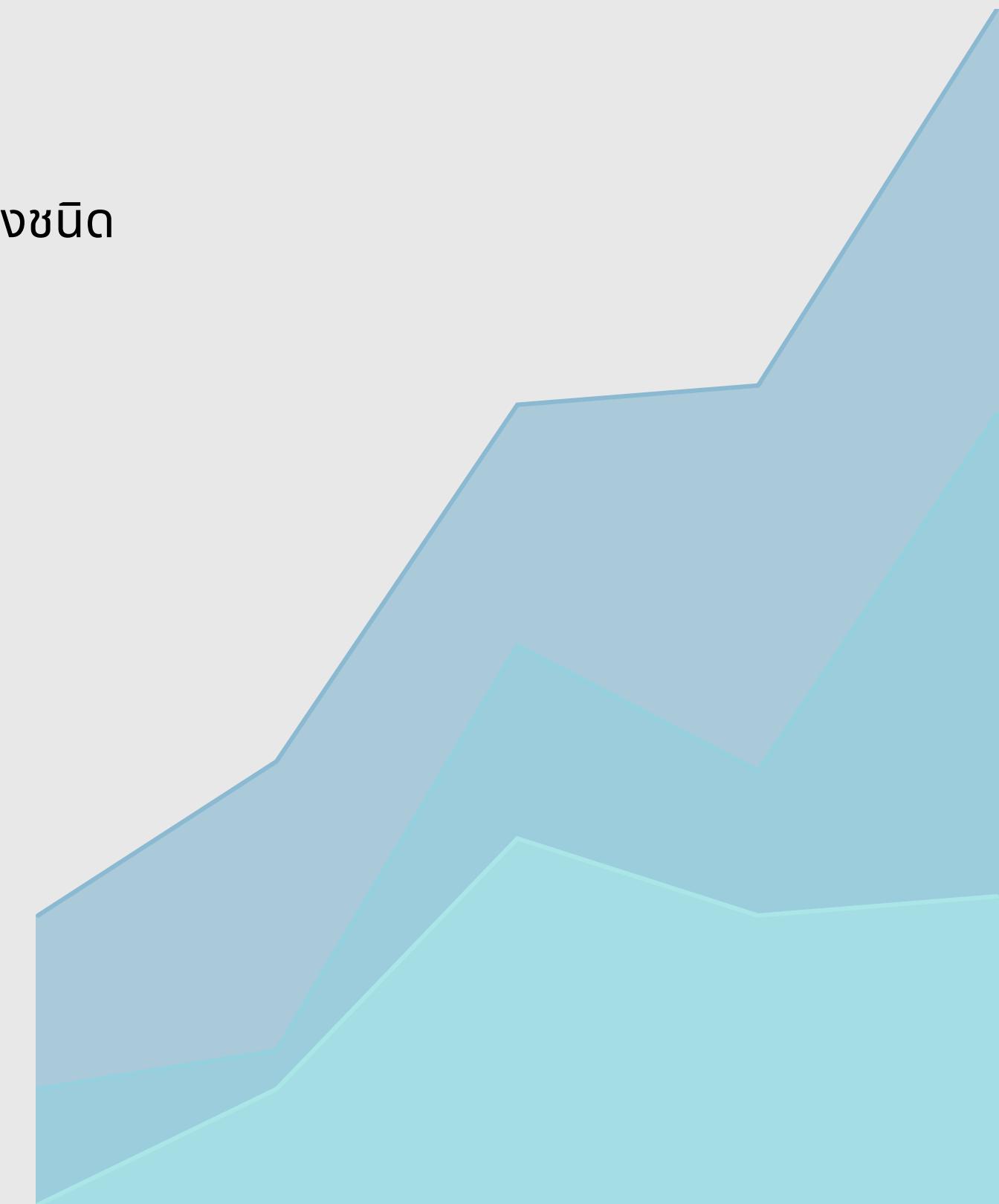
- เกิดขึ้นจากการเผาไหม้สารอินทรีย์ โรงงานอุตสาหกรรมที่ใช้เชื้อเพลิงจากพลาสติก
- เกิดขึ้นตามธรรมชาติระหว่างการเกิดฟ้าผ่า
- พื้นสามารถสังเคราะห์ NO ขึ้นได้

## NO<sub>2</sub>

- สามารถเกิดขึ้นได้เองจากการธรรมชาติ เช่น ฟ้าแลบ ภูเขาไฟระเบิด
- เกิดจากการเผาไหม้เชื้อเพลิงพลาสติก แต่เป็นส่วนน้อย

## O<sub>3</sub>

- เกิดจากการเผาไหม้ในเครื่องยนต์ และกำปฏิกิริยา กับแสงแดด จนเกิดก๊าซโอโซนขึ้น



# จัดหมวดหมู่ เมือง ตามอุตสาหกรรมของเมืองนั้นๆ



## Petroleum

- Salamanca
- Monterrey
- Poza Rica



## Textile

- León
- Toluca
- Aguascalientes
- Tulancingo
- Tlaxcala

## Agricultural

- Tepic
- Aguascalientes
- Oaxaca
- Torreón
- Chihuahua
- Irapuato
- Guerrero



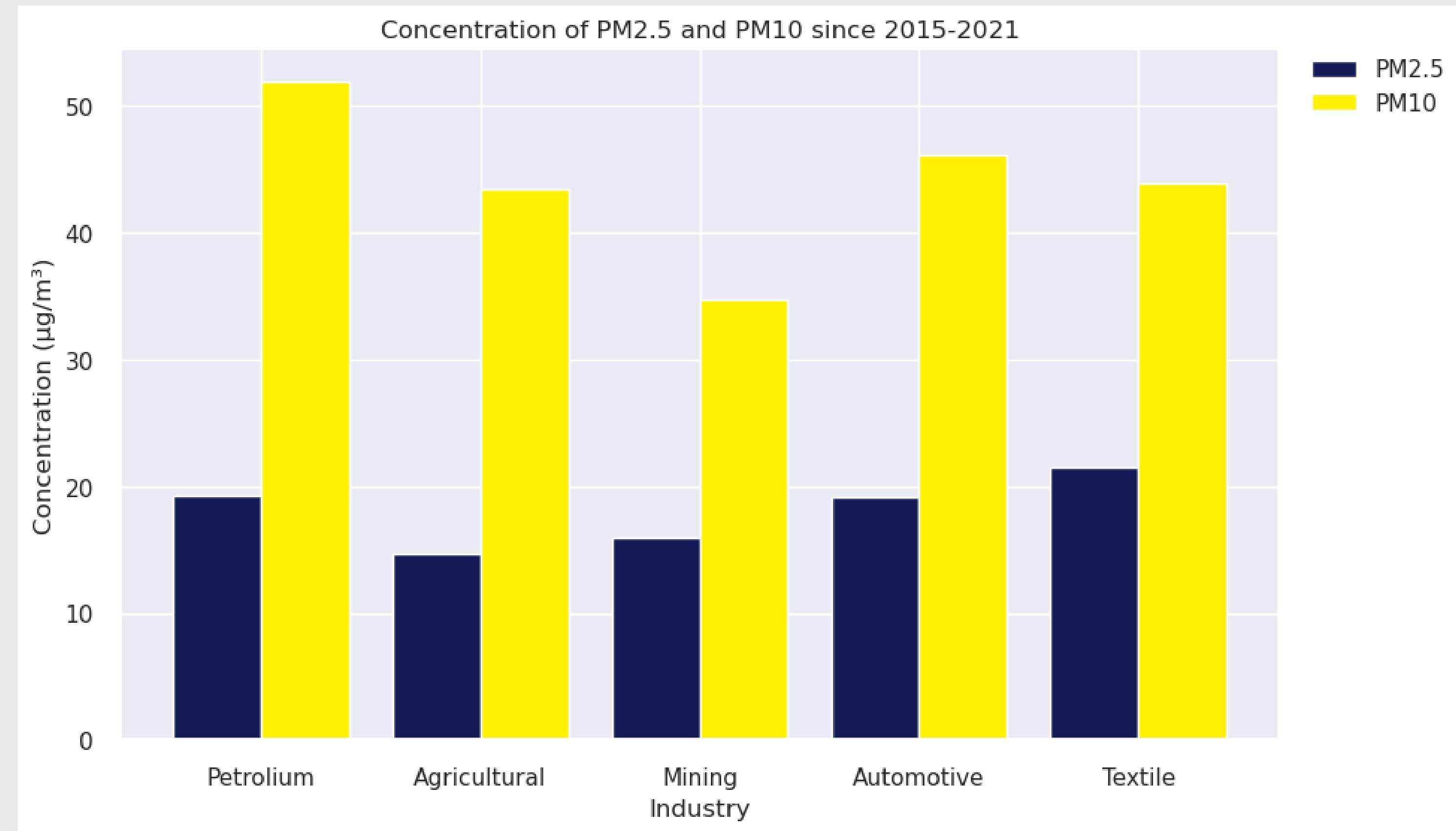
## Mining

- Pachuca
- Aguascalientes
- Chihuahua
- Durango

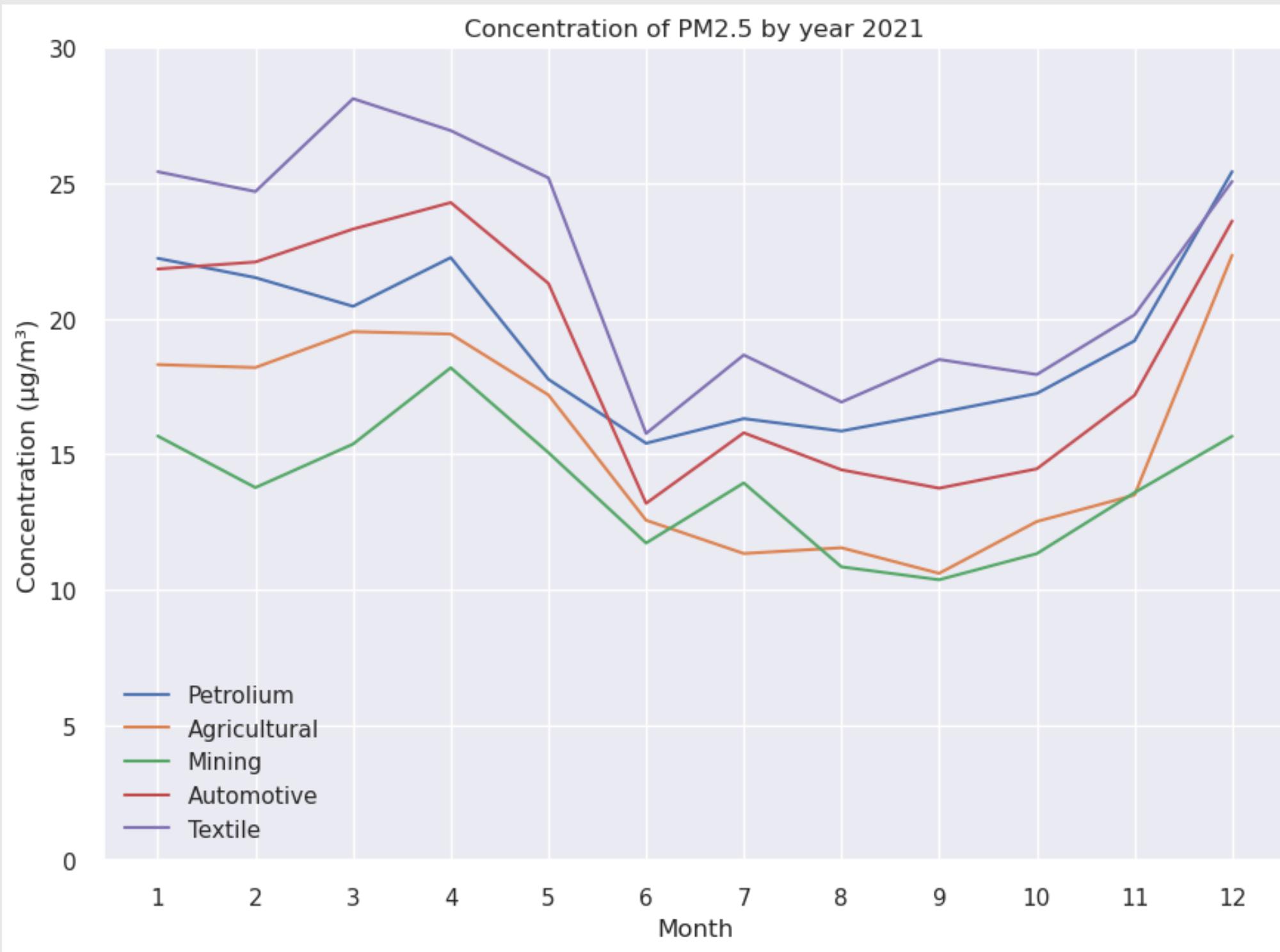
## Automotive and Other Manufacturing

- Toluca
- Puebla
- Celaya
- Tulancingo
- Piedras Negras
- Tepeapulco
- Irapuato
- Saltillo
- Monclova
- Tlaxcala(Electrical Circuits)
- Abasolo
- Guanajuato

# PM2.5 and PM10 emitted by industries in Mexico



# Trends of PM2.5 in each industry in 2021

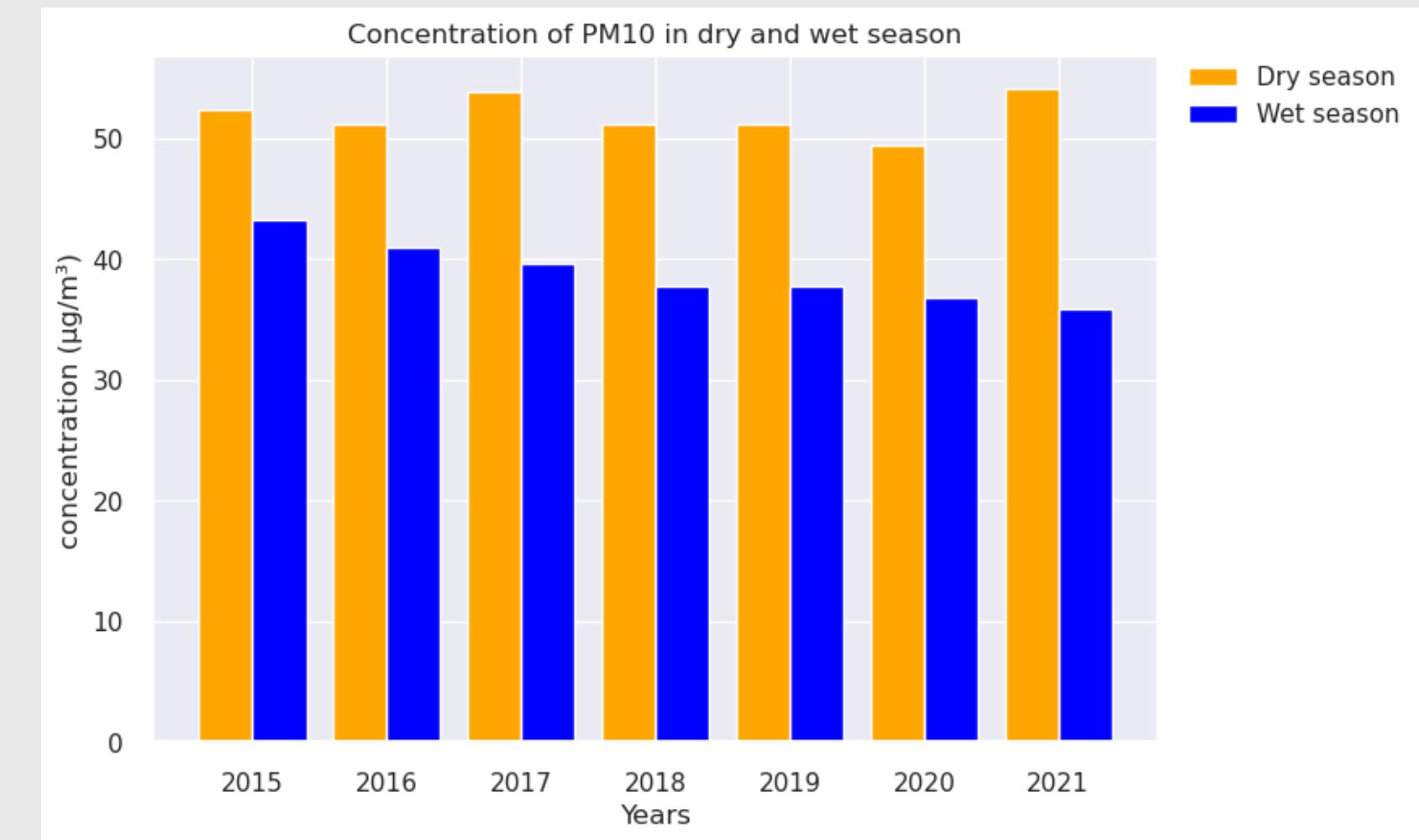
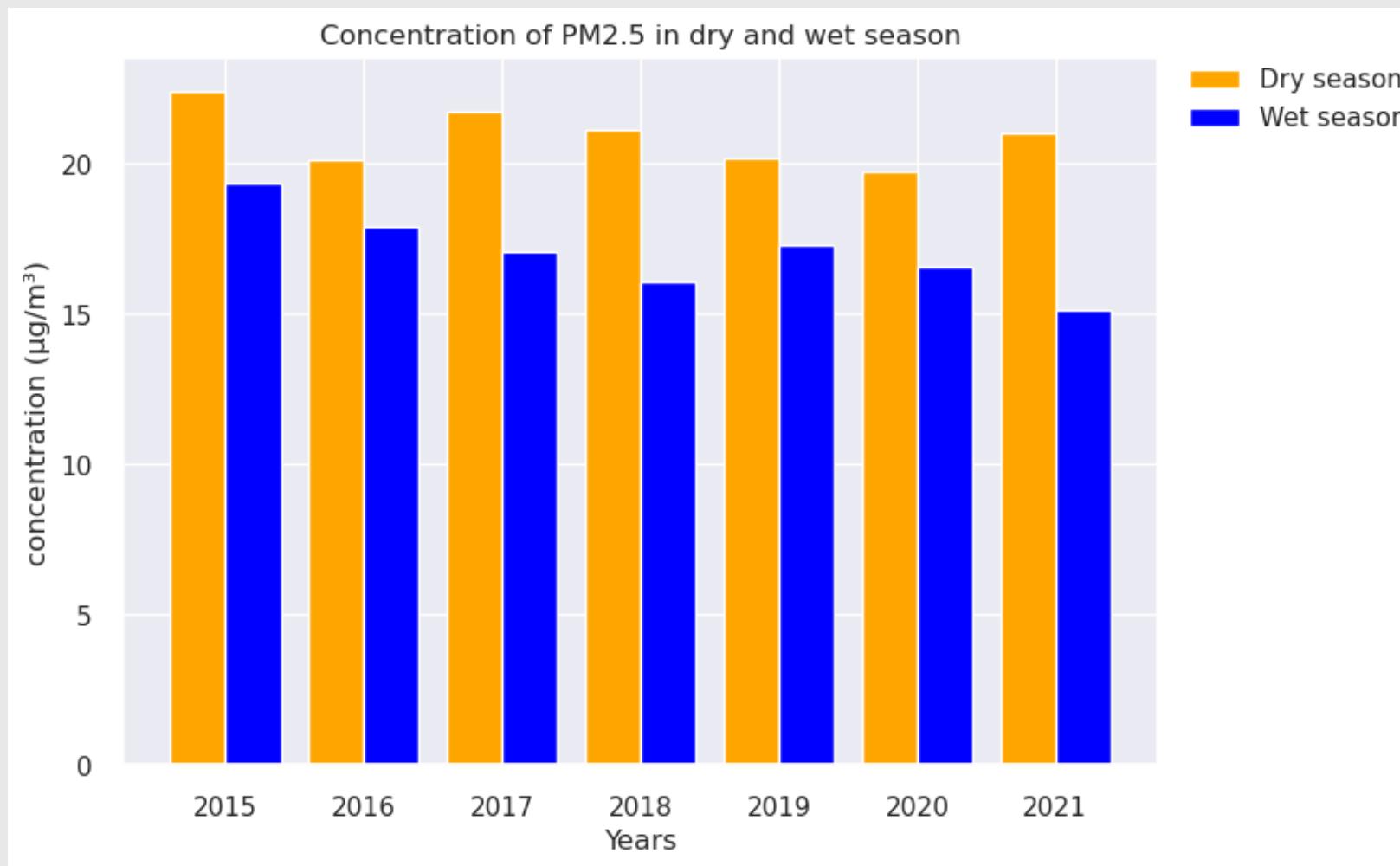


กั้ง 5 อุตสาหกรรม มีช่วงที่ความเข้มข้นของ PM2.5 ขึ้นสูง อยู่ที่เดือน 1-5 จากนั้นเริ่มลด ระดับลงที่เดือนที่ 6 และเริ่มสูงขึ้นอีกครั้งที่เดือน 11

กว่า อุตสาหกรรมก่อผ้า มีค่าความเข้มข้นมากกว่าอุตสาหกรรมอื่นอยู่เสมอ เพราะ ประเทศนี้ ครองส่วนแบ่งตลาดส่งออก (เฉพาะตลาดสหราชอาณาจักร สูงถึง 2.3 พันล้านเหรียญ) จากผลของ FTA-USMCA และอีน ๆ ถึง 12 ข้อตกลง ซึ่งครอบคลุม 46 ประเทศ

และอุตสาหกรรมที่ความเข้มข้นของ PM2.5 ณ จุดสูงสุด รองลงมา คือ อุตสาหกรรมการผลิตยานยนต์และอื่นๆ มีกำลังการผลิตรวมเกือบ 2 ล้านคัน/ปี โดยเฉลี่ย มีมูลค่าตลาดรวม 27,000 ล้านเหรียญสหราชอาณาจักร ข้อมูลล่าสุดในปี 2021 เม็กซิโกมีกำลังการผลิตรวมกว่า 1,685,705 คัน โดยเป็นการผลิตเพื่อการส่งออกถึง 956,530 หรือร้อยละ 56.74 ของยอดการผลิตกั้งหมด

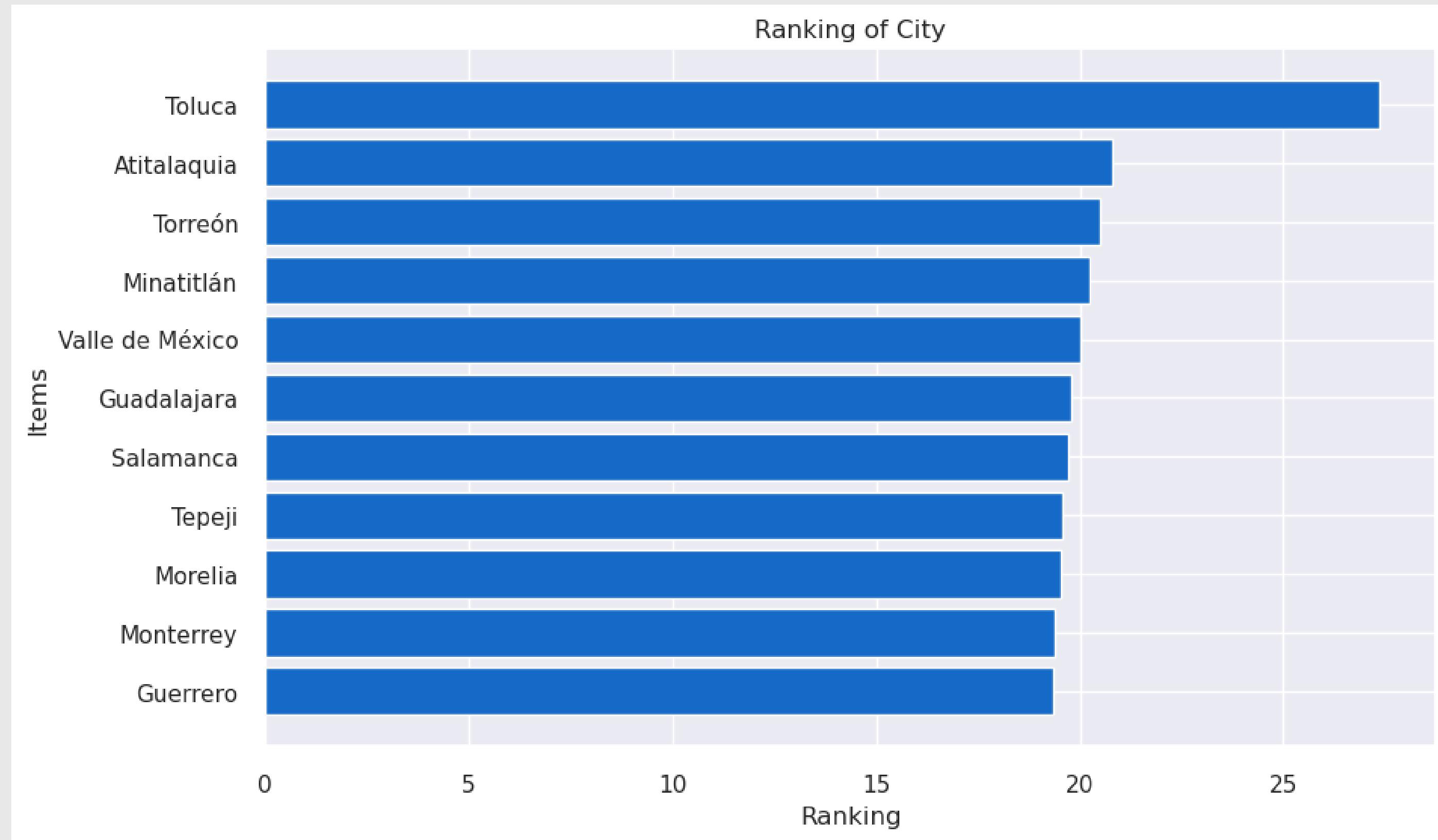
# Dry and Wet season



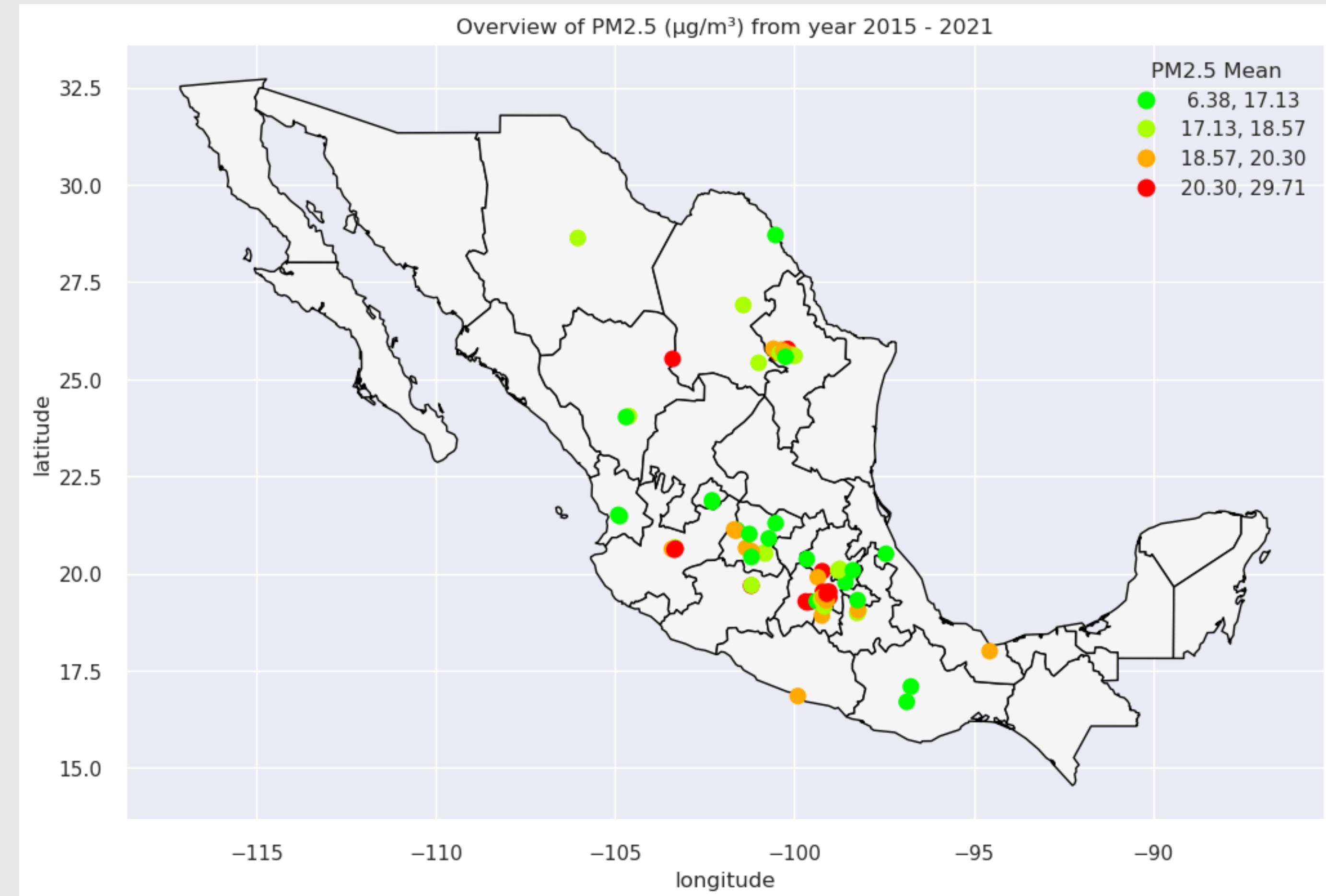
season	TMP	HR
dry_season	17.575776	50.346027
wet_season	21.077721	61.101833

จากแผนภูมิแก่งเราจะทราบว่า dry season ของทุกๆปี จะมีความเข้มข้นของ PM2.5 และ PM10 มากกว่าช่วง wet season และค่าความชื้นสัมพัทธ์ (HR) เวลี่ยของแต่ละช่วงจะแปรผันกับความเข้มข้นของ PM2.5 และ PM10

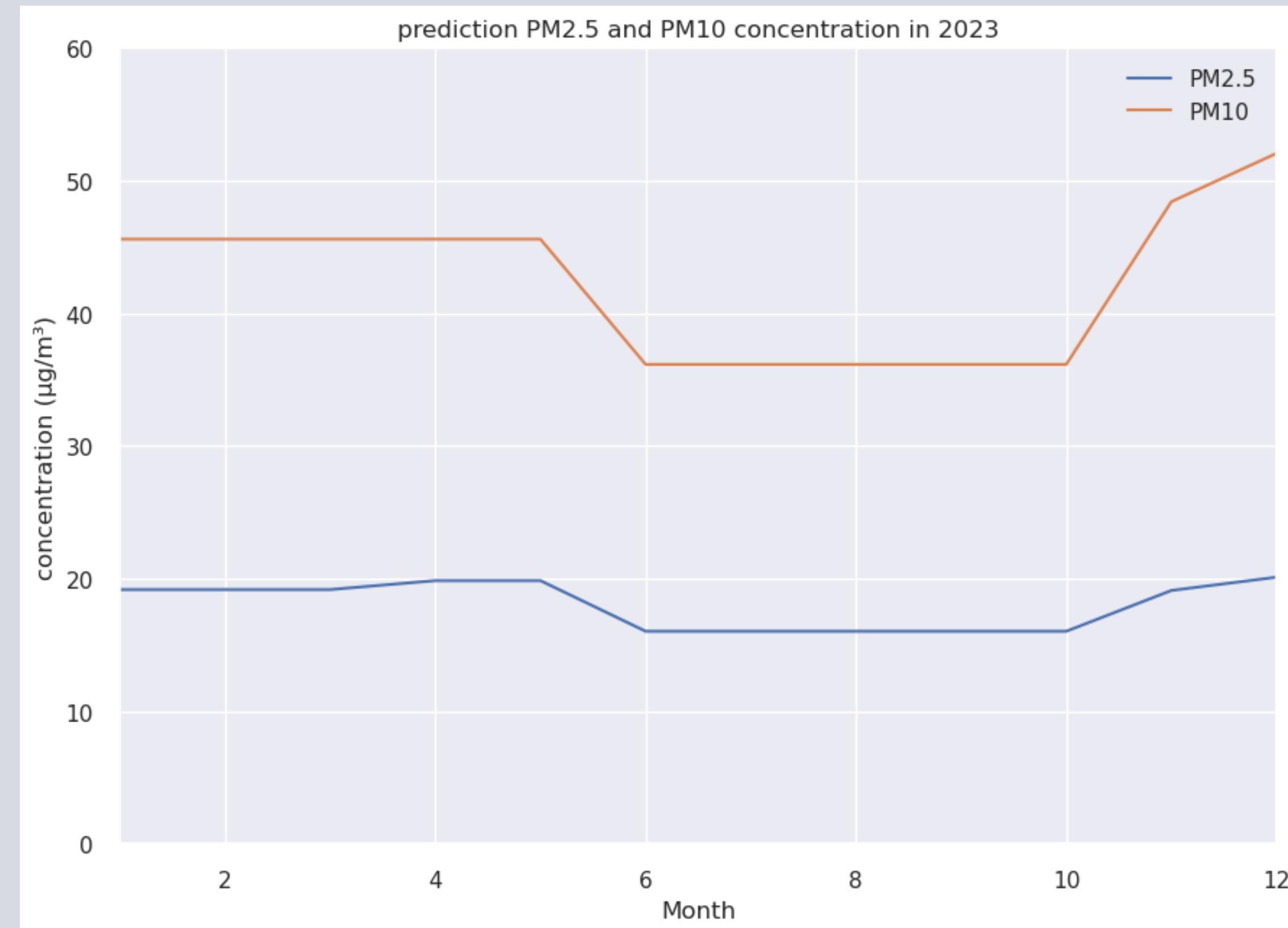
# Mosted polluted city in Mexico since 2015 - 2021



# แผนที่แสดงการรวมค่าเฉลี่ยของ PM2.5 ( $\mu\text{g}/\text{m}^3$ )



# Prediction



# Conclusion

---

เมืองที่มีความเข้มข้นของ PM2.5 มากสุดคือเมืองที่ทำอุตสาหกรรมสิ่งทอ  
ช่วงที่พิคที่สุด คือ ช่วง dry season  
โดยเมืองที่มีค่าเฉลี่ยของ PM2.5 มากที่สุดคือ Toluca

---

**THANK YOU  
FOR YOUR ATTENTION**