20250621_01

June 21, 2025

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

[2]:
```
# Today's goal : merge two dataset cand cleaning.

# Dataset A : Air Quality
air_quality = pd.DataFrame({'city': ['Taipei', 'Taichung', 'Kaohsiung',␣
 ↪'Hsinchu'],
                            'pm25': [35, 42, 55, 28],
                            'ozone': [20, 24, 30, 19]})

# Dataset B: Population
population = pd.DataFrame({'city': ['Taipei', 'Taichung', 'Kaohsiung',␣
 ↪'Hsinchu'],
                           'population': [2600000, 2800000, 2700000, 450000],
                           'density': [9800, 7500, 9400, 5300]})
```

[6]:
```
air_quality.head()
```

[6]:

	city	pm25	ozone
0	Taipei	35	20
1	Taichung	42	24
2	Kaohsiung	55	30
3	Hsinchu	28	19

[7]:
```
population.head()
```

[7]:

	city	population	density
0	Taipei	2600000	9800
1	Taichung	2800000	7500
2	Kaohsiung	2700000	9400
3	Hsinchu	450000	5300

[3]:
```
# Merge datasets
merged = pd.merge(air_quality, population, on = 'city')
```

[8]:
```
merged.head()
```

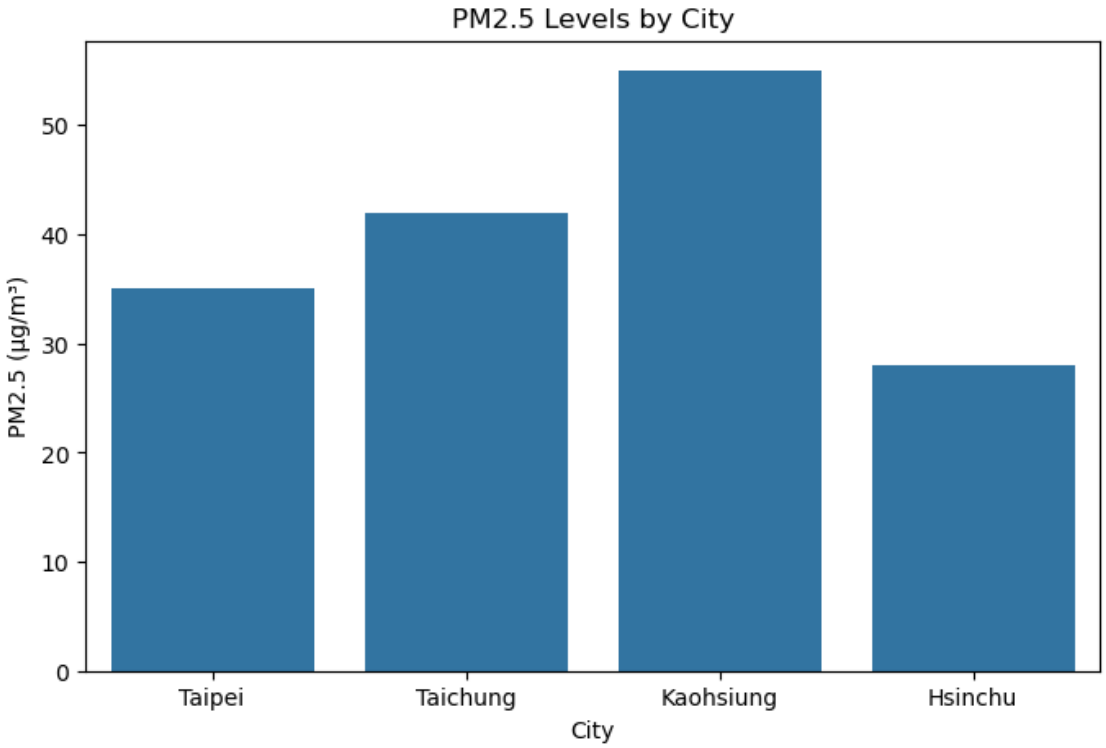
[8]:

	city	pm25	ozone	population	density
0	Taipei	35	20	2600000	9800
1	Taichung	42	24	2800000	7500
2	Kaohsiung	55	30	2700000	9400
3	Hsinchu	28	19	450000	5300

[9]:
```
# Visualization
import seaborn as sns
import matplotlib.pyplot as plt
```

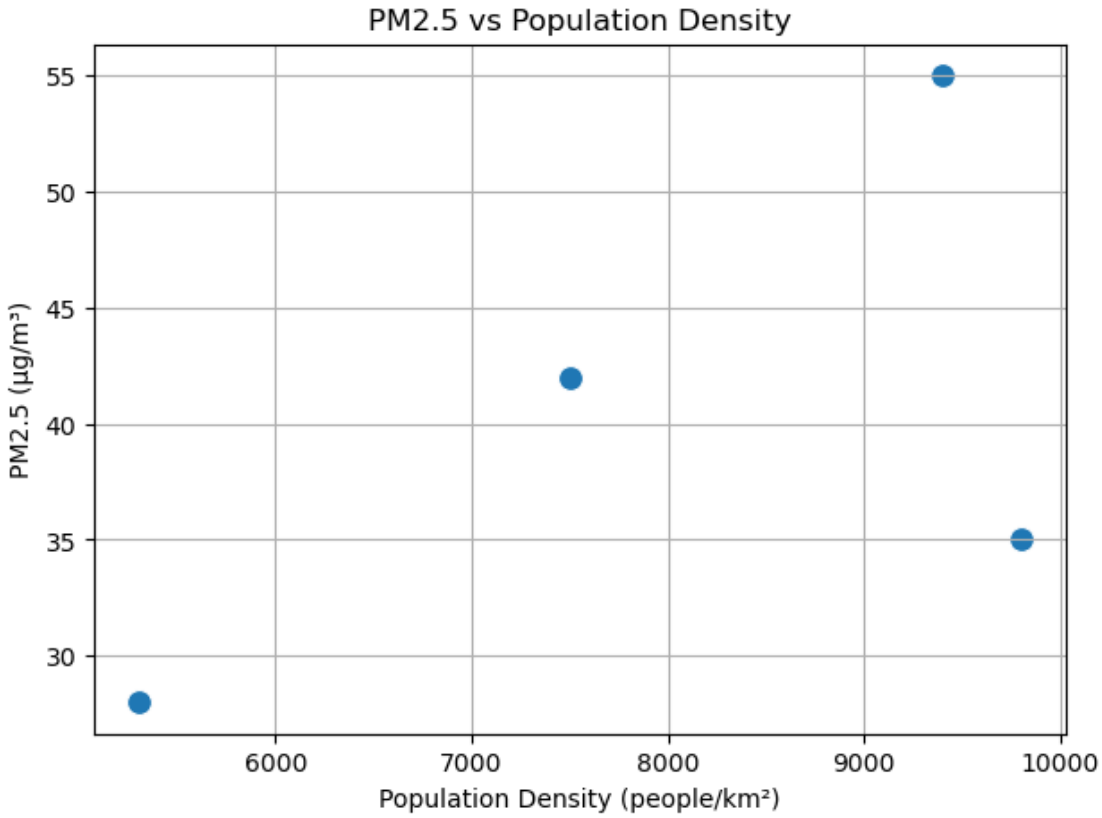
[16]:
```
# PM2.5 by city
plt.figure(figsize = (8, 5))
sns.barplot(data = merged, x = 'city', y = 'pm25')

plt.title("PM2.5 Levels by City")
plt.ylabel("PM2.5 ( g/m³)")
plt.xlabel("City")
plt.show()
```



[18]:
```
# PM2.5 along population density
# s = 100 is to cintrol how big the dot will be
plt.figure(figsize = (7, 5))
sns.scatterplot(data = merged, x = 'density', y = 'pm25', s = 100)
```

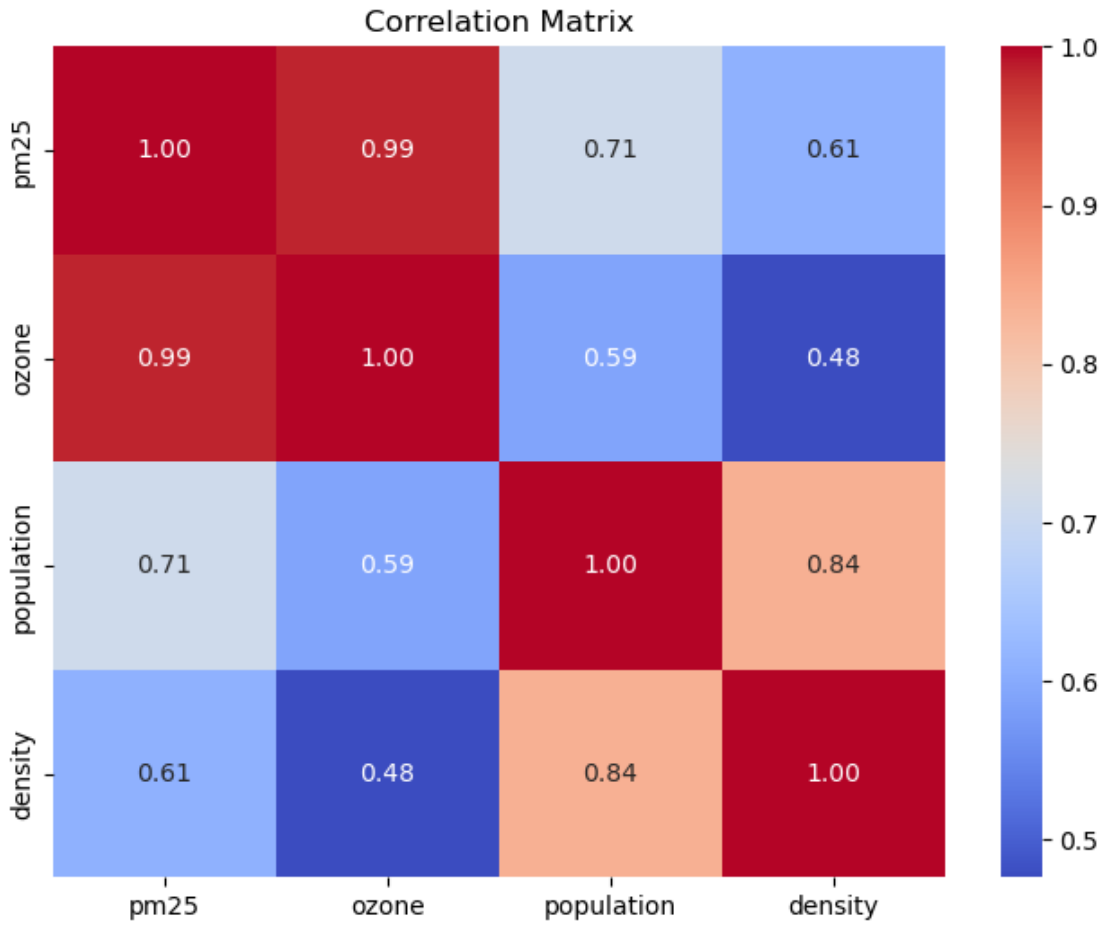
```
plt.title("PM2.5 vs Population Density")
plt.xlabel("Population Density (people/km²)")
plt.ylabel("PM2.5 ( g/m³)")
plt.grid(True)
plt.show()
```



[19]:
```
# Correlation matrix
corr = merged.corr(numeric_only = True)

plt.figure(figsize = (8, 6))
sns.heatmap(corr, annot = True, cmap = 'coolwarm', fmt = ".2f")

plt.title("Correlation Matrix")
plt.show()
```


Correlation Matrix
pm25
ozone
population
density
1.00
0.99
0.71
0.61
0.99
1.00
0.59
0.48
0.71
0.59
1.00
0.84
0.61
0.48
0.84
1.00
pm25
ozone
population
density
1.0
0.9
0.8
0.7
0.6
0.5