

EXPERIMENT-4

```
# eda_inspection_analysis_fixed.py

import pandas as pd

# Load the dataset
df = pd.read_csv(r"C:\Users\REC\Downloads\final_dataset.csv")

print("\n🔍 First 5 Rows:")
print(df.head())

print("\n📄 DataFrame Info:")
print(df.info())

print("\n📊 Column-wise Data Types:")
print(df.dtypes)

print("\n📈 Summary Statistics:")
print(df.describe())

# -----
# 2. Filtering and Subsetting Data
# -----

# Example: Filter rows where AQI > 100
if 'AQI' in df.columns:
    high_aqi = df[df['AQI'] > 100]
    print("\n📄 Rows where AQI > 100:")
    print(high_aqi.head())
```

```
# Example: Filter where PM2.5 > 60 (unhealthy)
```

```
if 'PM2.5' in df.columns:
```

```
    high_pm25 = df[df['PM2.5'] > 60]
```

```
    print("\n 🔥 Rows where PM2.5 > 60:")
```

```
    print(high_pm25.head())
```

```
# Subset only pollutant columns
```

```
pollutants = ['PM2.5', 'PM10', 'NO2', 'SO2', 'CO', 'Ozone', 'AQI']
```

```
existing_pollutants = [col for col in pollutants if col in df.columns]
```

```
print("\n 📌 Subset with pollutant columns:")
```

```
print(df[existing_pollutants].head())
```

```
# -----
```

```
# 3. Descriptive Statistics
```

```
# -----
```

```
# Central Tendency
```

```
print("\n 📊 Measures of Central Tendency:")
```

```
print("\n Mean:\n", df.mean(numeric_only=True))
```

```
print("\n Median:\n", df.median(numeric_only=True))
```

```
print("\n Mode:\n", df.mode(numeric_only=True).iloc[0])
```

```
# Dispersion
```

```
print("\n 📈 Measures of Dispersion:")
```

```
print("\n Range (max - min):\n", df.max(numeric_only=True) - df.min(numeric_only=True))
```

```
print("\n Variance:\n", df.var(numeric_only=True))
```

```
print("\n Standard Deviation:\n", df.std(numeric_only=True))
```


First 5 Rows:

	Date	Month	Year	Holidays_Count	Days	PM2.5	PM10	NO2	SO2	\
0	1	1	2021	0	5	408.80	442.42	160.61	12.95	
1	2	1	2021	0	6	404.04	561.95	52.85	5.18	
2	3	1	2021	1	7	225.07	239.04	170.95	10.93	
3	4	1	2021	0	1	89.55	132.08	153.98	10.42	
4	5	1	2021	0	2	54.06	55.54	122.66	9.70	

	CO	Ozone	AQI
0	2.77	43.19	462
1	2.60	16.43	482
2	1.40	44.29	263
3	1.01	49.19	207
4	0.64	48.88	149

DataFrame Info:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 1461 entries, 0 to 1460

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	Date	1461 non-null	int64
1	Month	1461 non-null	int64
2	Year	1461 non-null	int64
3	Holidays_Count	1461 non-null	int64
4	Days	1461 non-null	int64
5	PM2.5	1461 non-null	float64
6	PM10	1461 non-null	float64
7	NO2	1461 non-null	float64
8	SO2	1461 non-null	float64
9	CO	1461 non-null	float64
10	Ozone	1461 non-null	float64
11	AQI	1461 non-null	int64

dtypes: float64(6), int64(6)

memory usage: 137.1 KB

None

Column-wise Data Types:

```
Date          int64
Month          int64
Year           int64
Holidays_Count int64
Days           int64
PM2.5          float64
PM10           float64
NO2            float64
SO2            float64
CO             float64
Ozone          float64
AQI            int64
dtype: object
```

Summary Statistics:

	Date	Month	Year	Holidays_Count	Days \
count	1461.000000	1461.000000	1461.000000	1461.000000	1461.000000
mean	15.729637	6.522930	2022.501027	0.189596	4.000684
std	8.803105	3.449884	1.118723	0.392116	2.001883
min	1.000000	1.000000	2021.000000	0.000000	1.000000
25%	8.000000	4.000000	2022.000000	0.000000	2.000000
50%	16.000000	7.000000	2023.000000	0.000000	4.000000
75%	23.000000	10.000000	2024.000000	0.000000	6.000000
max	31.000000	12.000000	2024.000000	1.000000	7.000000

	PM2.5	PM10	NO2	SO2	CO \
count	1461.000000	1461.000000	1461.000000	1461.000000	1461.000000
mean	90.774538	218.219261	37.184921	20.104921	1.025832
std	71.650579	129.297734	35.225327	16.543659	0.608305
min	0.050000	9.690000	2.160000	1.210000	0.270000
25%	41.280000	115.110000	17.280000	7.710000	0.610000
50%	72.060000	199.800000	30.490000	15.430000	0.850000
75%	118.500000	297.750000	45.010000	26.620000	1.240000
max	1000.000000	1000.000000	433.980000	113.400000	4.700000

	Ozone	AQI
count	1461.000000	1461.000000
mean	36.338871	202.210815
std	18.951204	107.801076
min	2.700000	19.000000
25%	24.100000	108.000000
50%	32.470000	189.000000
75%	45.730000	284.000000
max	115.870000	500.000000

Rows where AQI > 100:

	Date	Month	Year	Holidays_Count	Days	PM2.5	PM10	NO2	SO2	\
0	1	1	2021	0	5	408.80	442.42	160.61	12.95	
1	2	1	2021	0	6	404.04	561.95	52.85	5.18	
2	3	1	2021	1	7	225.07	239.04	170.95	10.93	
3	4	1	2021	0	1	89.55	132.08	153.98	10.42	
4	5	1	2021	0	2	54.06	55.54	122.66	9.70	

	CO	Ozone	AQI
0	2.77	43.19	462
1	2.60	16.43	482
2	1.40	44.29	263
3	1.01	49.19	207
4	0.64	48.88	149

Rows where PM2.5 > 60:

	Date	Month	Year	Holidays_Count	Days	PM2.5	PM10	NO2	SO2	\
0	1	1	2021	0	5	408.80	442.42	160.61	12.95	
1	2	1	2021	0	6	404.04	561.95	52.85	5.18	
2	3	1	2021	1	7	225.07	239.04	170.95	10.93	
3	4	1	2021	0	1	89.55	132.08	153.98	10.42	
5	6	1	2021	0	3	155.59	180.14	142.71	10.29	

	CO	Ozone	AQI
0	2.77	43.19	462
1	2.60	16.43	482
2	1.40	44.29	263
3	1.01	49.19	207
5	1.18	44.47	252

Subset with pollutant columns:

	PM2.5	PM10	NO2	SO2	CO	Ozone	AQI
0	408.80	442.42	160.61	12.95	2.77	43.19	462
1	404.04	561.95	52.85	5.18	2.60	16.43	482
2	225.07	239.04	170.95	10.93	1.40	44.29	263
3	89.55	132.08	153.98	10.42	1.01	49.19	207
4	54.06	55.54	122.66	9.70	0.64	48.88	149

Measures of Central Tendency:

Mean:

Date	15.729637
Month	6.522930
Year	2022.501027
Holidays_Count	0.189596
Days	4.000684
PM2.5	90.774538
PM10	218.219261
NO2	37.184921
SO2	20.104921
CO	1.025832
Ozone	36.338871
AQI	202.210815

dtype: float64

Median:

Date	16.00
Month	7.00
Year	2023.00
Holidays_Count	0.00
Days	4.00
PM2.5	72.06
PM10	199.80
NO2	30.49
SO2	15.43
CO	0.85
Ozone	32.47
AQI	189.00

dtype: float64


```
Mode:
  Date          1.00
  Month         1.00
  Year         2024.00
  Holidays_Count 0.00
  Days          1.00
  PM2.5         31.18
  PM10         116.78
  NO2           4.82
  SO2           8.01
  CO            0.69
  Ozone         16.85
  AQI           55.00
Name: 0, dtype: float64
```

▴ Measures of Dispersion:

```
Range (max - min):
  Date          30.00
  Month         11.00
  Year           3.00
  Holidays_Count 1.00
  Days           6.00
  PM2.5         999.95
  PM10          990.31
  NO2           431.82
  SO2           112.19
  CO             4.43
  Ozone         113.17
  AQI           481.00
dtype: float64
```

Variance:

Date	77.494662
Month	11.901700
Year	1.251540
Holidays_Count	0.153755
Days	4.007534
PM2.5	5133.805518
PM10	16717.903971
NO2	1240.823645
SO2	273.692644
CO	0.370035
Ozone	359.148138
AQI	11621.071965

dtype: float64

Standard Deviation:

Date	8.803105
Month	3.449884
Year	1.118723
Holidays_Count	0.392116
Days	2.001883
PM2.5	71.650579
PM10	129.297734
NO2	35.225327
SO2	16.543659
CO	0.608305
Ozone	18.951204
AQI	107.801076

dtype: float64