PS2

November 12, 2025

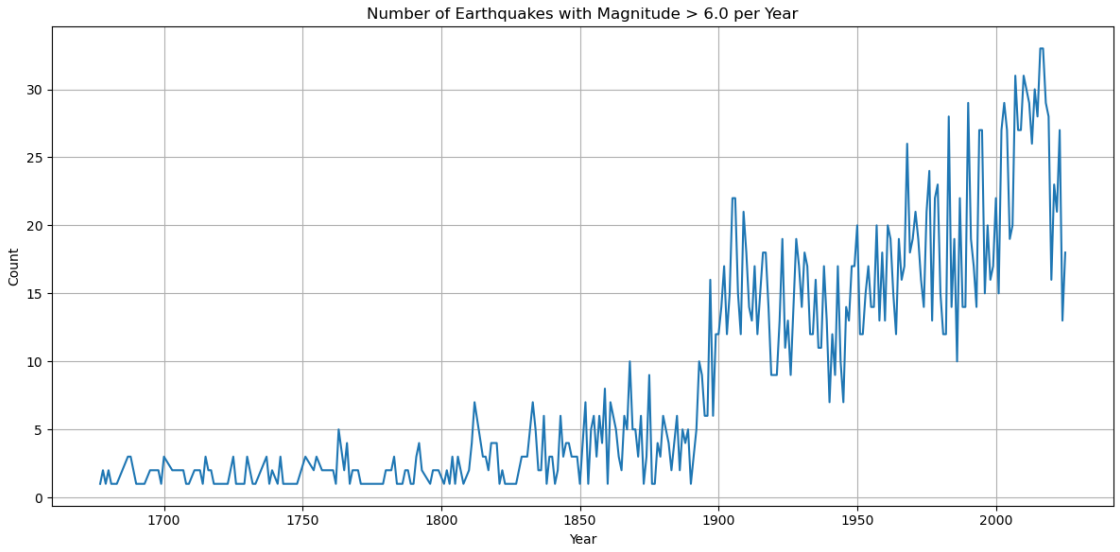
[4]:
```
#1. Significant earthquakes since 2150 B.C.
import pandas as pd

#   TSV
file_path = r"D:\code\earthquakes-2025-11-05_21-17-35_+0800.tsv"
Sig_Eqs = pd.read_csv(file_path, sep='\t', encoding='utf-8',␣
 ↪on_bad_lines='skip', low_memory=False)
```

[5]:
```
#  Deaths
Sig_Eqs['Deaths'] = pd.to_numeric(Sig_Eqs['Deaths'], errors='coerce').fillna(0)

#
deaths_by_country = Sig_Eqs.groupby('Country')['Deaths'].sum().
 ↪sort_values(ascending=False)

#
print("Top 10 countries by total earthquake deaths:")
print(deaths_by_country.head(10))
```

```
Top 10 countries by total earthquake deaths:
Country
CHINA         2139210.0
TURKEY        1199742.0
IRAN          1014453.0
ITALY          498219.0
SYRIA          419226.0
HAITI          323484.0
AZERBAIJAN     319251.0
JAPAN          242445.0
ARMENIA        191890.0
PAKISTAN       145083.0
Name: Deaths, dtype: float64
```

[6]:
```
Sig_Eqs['Mag'] = pd.to_numeric(Sig_Eqs['Mag'], errors='coerce')

#
Sig_Eqs['Year'] = pd.to_numeric(Sig_Eqs['Year'], errors='coerce')
Sig_Eqs['Mo'] = pd.to_numeric(Sig_Eqs['Mo'], errors='coerce')
```

```
Sig_Eqs['Dy'] = pd.to_numeric(Sig_Eqs['Dy'], errors='coerce')
Sig_Eqs['Date'] = pd.to_datetime(dict(year=Sig_Eqs['Year'],␣
  ↪month=Sig_Eqs['Mo'], day=Sig_Eqs['Dy']), errors='coerce')

#    > 6.0
high_mag_eqs = Sig_Eqs[Sig_Eqs['Mag'] > 6.0].copy()

#
high_mag_eqs['Year'] = high_mag_eqs['Date'].dt.year
eqs_per_year = high_mag_eqs.groupby('Year').size()

#
import matplotlib.pyplot as plt
plt.figure(figsize=(12, 6))
eqs_per_year.plot()
plt.title("Number of Earthquakes with Magnitude > 6.0 per Year")
plt.xlabel("Year")
plt.ylabel("Count")
plt.grid(True)
plt.tight_layout()
plt.show()
```



[7]:
```
def CountEq_LargestEq(country):
    df_country = Sig_Eqs[Sig_Eqs['Country'] == country].copy()
    total_eqs = len(df_country)

    #   Mag   Date
    df_country['Mag'] = pd.to_numeric(df_country['Mag'], errors='coerce')
```

```
    df_country['Date'] = pd.to_datetime(df_country['Date'], errors='coerce')

    #
    if df_country['Mag'].isna().all():
        largest_eq_date = None
    else:
        max_mag_idx = df_country['Mag'].idxmax()
        largest_eq_date = df_country.loc[max_mag_idx, 'Date']

    return total_eqs, largest_eq_date

#
results = []
for country in Sig_Eqs['Country'].dropna().unique():
    total, date = CountEq_LargestEq(country)
    results.append((country, total, date))

#
results_sorted = sorted(results, key=lambda x: x[1], reverse=True)

#
print("Earthquake count and largest magnitude date by country:")
for country, total, date in results_sorted:
    print(f"{country}: {total} earthquakes, largest on {date}")
```

```
Earthquake count and largest magnitude date by country:
CHINA: 623 earthquakes, largest on NaT
JAPAN: 424 earthquakes, largest on 2011-03-11 00:00:00
INDONESIA: 421 earthquakes, largest on 2004-12-26 00:00:00
IRAN: 388 earthquakes, largest on NaT
TURKEY: 358 earthquakes, largest on 1939-12-26 00:00:00
ITALY: 333 earthquakes, largest on 1915-01-13 00:00:00
GREECE: 289 earthquakes, largest on NaT
USA: 280 earthquakes, largest on 1964-03-28 00:00:00
PHILIPPINES: 230 earthquakes, largest on 1897-09-21 00:00:00
MEXICO: 214 earthquakes, largest on 1787-03-28 00:00:00
CHILE: 200 earthquakes, largest on 1960-05-22 00:00:00
PERU: 194 earthquakes, largest on 1716-02-06 00:00:00
RUSSIA: 158 earthquakes, largest on 1952-11-04 00:00:00
PAPUA NEW GUINEA: 107 earthquakes, largest on 1919-05-06 00:00:00
INDIA: 102 earthquakes, largest on 1950-08-15 00:00:00
TAIWAN: 101 earthquakes, largest on 1920-06-05 00:00:00
COLOMBIA: 82 earthquakes, largest on 1826-06-18 00:00:00
NEW ZEALAND: 72 earthquakes, largest on NaT
ECUADOR: 69 earthquakes, largest on 1906-01-31 00:00:00
AFGHANISTAN: 68 earthquakes, largest on 1909-07-07 00:00:00
VENEZUELA: 65 earthquakes, largest on NaT
VANUATU: 64 earthquakes, largest on 1913-10-14 00:00:00
```

SOLOMON ISLANDS: 63 earthquakes, largest on 1977-04-21 00:00:00
ALGERIA: 57 earthquakes, largest on 1980-10-10 00:00:00
ALBANIA: 56 earthquakes, largest on 1893-06-14 00:00:00
PAKISTAN: 53 earthquakes, largest on 1945-11-27 00:00:00
CROATIA: 53 earthquakes, largest on NaT
GUATEMALA: 47 earthquakes, largest on 1942-08-06 00:00:00
FRANCE: 43 earthquakes, largest on 1817-03-11 00:00:00
MYANMAR (BURMA): 43 earthquakes, largest on 1912-05-23 00:00:00
EL SALVADOR: 42 earthquakes, largest on 1776-05-30 00:00:00
NICARAGUA: 39 earthquakes, largest on 1898-04-29 00:00:00
ARGENTINA: 39 earthquakes, largest on 1894-10-27 00:00:00
USA TERRITORY: 38 earthquakes, largest on 1902-09-22 00:00:00
COSTA RICA: 37 earthquakes, largest on 1950-10-05 00:00:00
SPAIN: 35 earthquakes, largest on NaT
SYRIA: 33 earthquakes, largest on NaT
SWITZERLAND: 31 earthquakes, largest on NaT
PORTUGAL: 28 earthquakes, largest on NaT
AZORES (PORTUGAL): 28 earthquakes, largest on 1968-02-28 00:00:00
NEW CALEDONIA: 28 earthquakes, largest on 1875-03-28 00:00:00
TAJIKISTAN: 28 earthquakes, largest on 1907-10-21 00:00:00
AUSTRALIA: 26 earthquakes, largest on 1989-05-23 00:00:00
TONGA: 25 earthquakes, largest on 1919-04-30 00:00:00
ISRAEL: 24 earthquakes, largest on NaT
IRAQ: 24 earthquakes, largest on 1864-12-02 00:00:00
PANAMA: 23 earthquakes, largest on 1882-09-07 00:00:00
KERMADEC ISLANDS (NEW ZEALAND): 23 earthquakes, largest on 1986-10-20 00:00:00
SOUTH KOREA: 22 earthquakes, largest on NaT
SLOVENIA: 22 earthquakes, largest on NaT
CANADA: 22 earthquakes, largest on 1949-08-22 00:00:00
NEPAL: 21 earthquakes, largest on NaT
MOROCCO: 21 earthquakes, largest on 2023-09-08 00:00:00
HAITI: 21 earthquakes, largest on 1842-05-07 00:00:00
FIJI: 20 earthquakes, largest on 1919-01-01 00:00:00
DOMINICAN REPUBLIC: 19 earthquakes, largest on 1946-08-04 00:00:00
BOLIVIA: 19 earthquakes, largest on 1994-06-09 00:00:00
JAMAICA: 19 earthquakes, largest on 1899-06-14 00:00:00
BULGARIA: 18 earthquakes, largest on 1904-04-04 00:00:00
AZERBAIJAN: 17 earthquakes, largest on NaT
ICELAND: 17 earthquakes, largest on 1912-05-06 00:00:00
BANGLADESH: 17 earthquakes, largest on 1918-07-08 00:00:00
EGYPT: 16 earthquakes, largest on 1995-11-22 00:00:00
KYRGYZSTAN: 15 earthquakes, largest on 1911-01-03 00:00:00
GEORGIA: 15 earthquakes, largest on 1905-10-21 00:00:00
SERBIA: 15 earthquakes, largest on 1922-03-24 00:00:00
ROMANIA: 15 earthquakes, largest on 1977-03-04 00:00:00
HONDURAS: 15 earthquakes, largest on 2025-02-08 00:00:00
CUBA: 15 earthquakes, largest on 2020-01-28 00:00:00
LEBANON: 14 earthquakes, largest on 1759-11-25 00:00:00

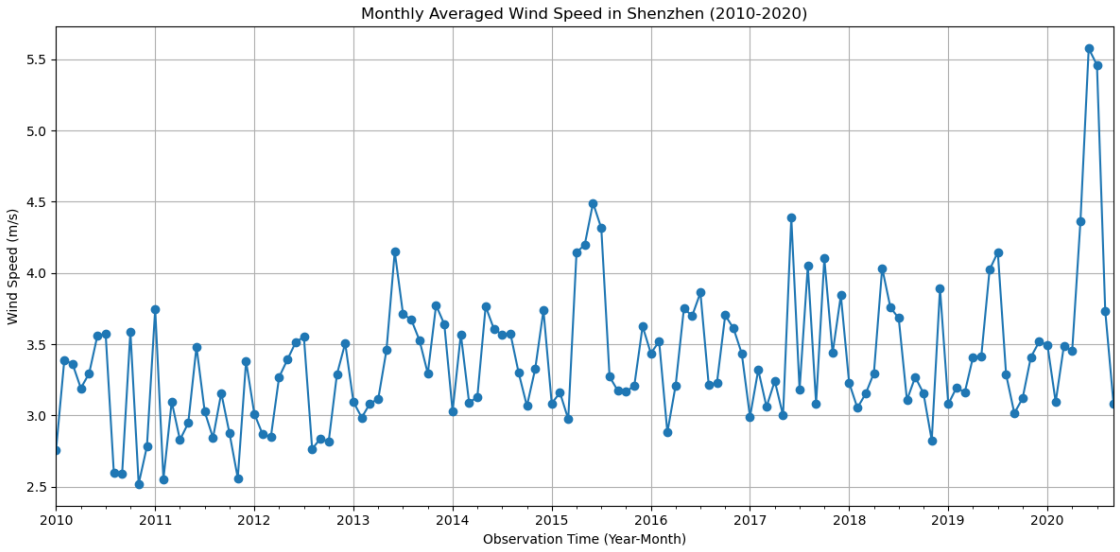
UK: 14 earthquakes, largest on NaT
UZBEKISTAN: 14 earthquakes, largest on 1976-04-08 00:00:00
SOUTH AFRICA: 14 earthquakes, largest on 1942-11-10 00:00:00
MACEDONIA: 13 earthquakes, largest on 1979-05-24 00:00:00
ARMENIA: 13 earthquakes, largest on 1988-12-07 00:00:00
BRAZIL: 13 earthquakes, largest on 1963-11-09 00:00:00
TURKMENISTAN: 12 earthquakes, largest on 1895-07-08 00:00:00
UKRAINE: 12 earthquakes, largest on NaT
BOSNIA-HERZEGOVINA: 11 earthquakes, largest on 1969-10-27 00:00:00
YEMEN: 10 earthquakes, largest on 1982-12-13 00:00:00
MONTENEGRO: 10 earthquakes, largest on 1979-04-15 00:00:00
ETHIOPIA: 10 earthquakes, largest on 1906-08-25 00:00:00
MARTINIQUE: 10 earthquakes, largest on 1906-12-03 00:00:00
TUNISIA: 9 earthquakes, largest on 1957-02-20 00:00:00
KAZAKHSTAN: 9 earthquakes, largest on 1889-07-11 00:00:00
GERMANY: 9 earthquakes, largest on 1978-09-03 00:00:00
GUADELOUPE: 9 earthquakes, largest on 1843-02-08 00:00:00
CYPRUS: 8 earthquakes, largest on 1953-09-10 00:00:00
AUSTRIA: 8 earthquakes, largest on NaT
TRINIDAD AND TOBAGO: 8 earthquakes, largest on 1888-01-10 00:00:00
TANZANIA: 8 earthquakes, largest on 1910-12-13 00:00:00
SAMOA: 8 earthquakes, largest on 1917-06-26 00:00:00
POLAND: 8 earthquakes, largest on 2004-09-21 00:00:00
ATLANTIC OCEAN: 7 earthquakes, largest on 1941-11-25 00:00:00
CONGO: 7 earthquakes, largest on 1992-09-11 00:00:00
SOUTH GEORGIA AND THE SOUTH SANDWICH ISLANDS: 7 earthquakes, largest on
1929-06-27 00:00:00
NORTH KOREA: 6 earthquakes, largest on NaT
ERITREA: 6 earthquakes, largest on 1875-11-02 00:00:00
MONGOLIA: 6 earthquakes, largest on 1905-07-09 00:00:00
VIETNAM: 6 earthquakes, largest on 1935-11-01 00:00:00
ANTARCTICA: 6 earthquakes, largest on 1998-03-25 00:00:00
JORDAN: 5 earthquakes, largest on NaT
GHANA: 5 earthquakes, largest on 1862-07-10 00:00:00
HUNGARY: 5 earthquakes, largest on 1834-10-15 00:00:00
RWANDA: 5 earthquakes, largest on 2015-08-07 00:00:00
BHUTAN: 5 earthquakes, largest on 2009-09-21 00:00:00
THAILAND: 4 earthquakes, largest on 2014-05-05 00:00:00
SOUTH SUDAN: 4 earthquakes, largest on 1990-05-20 00:00:00
MICRONESIA, FED. STATES OF: 4 earthquakes, largest on 1911-08-16 00:00:00
UGANDA: 4 earthquakes, largest on 1912-07-09 00:00:00
MALAWI: 4 earthquakes, largest on 1989-03-10 00:00:00
SLOVAKIA: 3 earthquakes, largest on 2004-01-10 00:00:00
ANTIGUA AND BARBUDA: 3 earthquakes, largest on 1690-04-16 00:00:00
KENYA: 3 earthquakes, largest on 1928-01-06 00:00:00
INDIAN OCEAN: 3 earthquakes, largest on 1928-03-09 00:00:00
MALAYSIA: 3 earthquakes, largest on 1976-07-26 00:00:00
NETHERLANDS: 3 earthquakes, largest on 1992-04-13 00:00:00

```
SAUDI ARABIA: 3 earthquakes, largest on 2009-05-19 00:00:00
MOZAMBIQUE: 3 earthquakes, largest on 2006-02-22 00:00:00
LIBYA: 2 earthquakes, largest on 1963-02-21 00:00:00
OMAN: 2 earthquakes, largest on NaT
URUGUAY: 2 earthquakes, largest on None
FRENCH GUIANA: 2 earthquakes, largest on 1885-08-04 00:00:00
SAINT LUCIA: 2 earthquakes, largest on None
TOGO: 2 earthquakes, largest on NaT
CANARY ISLANDS: 2 earthquakes, largest on None
UK TERRITORY: 2 earthquakes, largest on 1983-11-30 00:00:00
COTE D'IVOIRE: 2 earthquakes, largest on 1879-02-11 00:00:00
SOLOMON SEA: 2 earthquakes, largest on 1895-03-06 00:00:00
CAMEROON: 2 earthquakes, largest on 1945-09-12 00:00:00
PACIFIC OCEAN: 2 earthquakes, largest on 1932-11-02 00:00:00
LAOS: 2 earthquakes, largest on 2007-05-16 00:00:00
IRELAND: 1 earthquakes, largest on None
SIERRA LEONE: 1 earthquakes, largest on 1795-05-20 00:00:00
NORWAY: 1 earthquakes, largest on 1819-08-31 00:00:00
GRENADA: 1 earthquakes, largest on None
BARBADOS: 1 earthquakes, largest on None
SAINT VINCENT AND THE GRENADINES: 1 earthquakes, largest on None
FRENCH POLYNESIA: 1 earthquakes, largest on 1848-07-12 00:00:00
BRITISH VIRGIN ISLANDS: 1 earthquakes, largest on None
SRI LANKA: 1 earthquakes, largest on None
TASMAN SEA: 1 earthquakes, largest on 1892-01-26 00:00:00
MONTSERRAT: 1 earthquakes, largest on None
KIRIBATI: 1 earthquakes, largest on 1905-06-30 00:00:00
PALAU: 1 earthquakes, largest on 1914-10-23 00:00:00
CENTRAL AFRICAN REPUBLIC: 1 earthquakes, largest on 1921-09-16 00:00:00
GABON: 1 earthquakes, largest on 1974-09-23 00:00:00
BELGIUM: 1 earthquakes, largest on 1983-11-08 00:00:00
GUINEA: 1 earthquakes, largest on 1983-12-22 00:00:00
DJIBOUTI: 1 earthquakes, largest on 1989-08-20 00:00:00
BERING SEA: 1 earthquakes, largest on 1991-02-21 00:00:00
WALLIS AND FUTUNA (FRENCH TERRITORY): 1 earthquakes, largest on 1993-03-12
00:00:00
SUDAN: 1 earthquakes, largest on 1993-08-01 00:00:00
BURUNDI: 1 earthquakes, largest on 2004-02-24 00:00:00
CZECH REPUBLIC: 1 earthquakes, largest on 2008-11-22 00:00:00
MADAGASCAR: 1 earthquakes, largest on 2017-01-11 00:00:00
ZAMBIA: 1 earthquakes, largest on 2017-02-24 00:00:00
COMOROS: 1 earthquakes, largest on 2018-05-15 00:00:00
```

[29]:
```python
#2. Wind speed in Shenzhen from 2010 to 2020
import pandas as pd
import matplotlib.pyplot as plt
```

```
# 1.
file_path = r"D:\code\2281305.csv"

# NOAA ISD          CSV      POS 61-70
#      "WND"       /
#       df.head()
df = pd.read_csv(file_path, low_memory=False)

# 2.
# NOAA ISD    "DATE"   "YEARMODAHRMN"      YYYYMMDDHHMM
#     "DATE"
if "DATE" in df.columns:
    df["DATE"] = pd.to_datetime(df["DATE"], errors="coerce")
elif "YEARMODAHRMN" in df.columns:
    df["DATE"] = pd.to_datetime(df["YEARMODAHRMN"], format="%Y%m%d%H%M",␣
 ↪errors="coerce")

# 3.
# POS 65-70       "WND"     "ddd,dd,dd,dd"
#     "WND"      /
def parse_wind(wnd_str):
    try:
        # NOAA WND  : "ddd,dd,dd,dd" →
        parts = str(wnd_str).split(",")
        if len(parts) >= 4 and parts[3] != "9999":
            return float(parts[3]) / 10.0  #     m/s *10
    except:
        return None
    return None

if "WND" in df.columns:
    df["WIND_SPEED"] = df["WND"].apply(parse_wind)
else:
    #        "wind_speed"
    df["WIND_SPEED"] = pd.to_numeric(df["wind_speed"], errors="coerce")

# 4.
df = df.dropna(subset=["DATE", "WIND_SPEED"])
df["YEAR_MONTH"] = df["DATE"].dt.to_period("M")
monthly_avg = df.groupby("YEAR_MONTH")["WIND_SPEED"].mean()

# 5.
plt.figure(figsize=(12,6))
monthly_avg.plot(marker="o", linestyle="-")
plt.title("Monthly Averaged Wind Speed in Shenzhen (2010-2020)")
plt.xlabel("Observation Time (Year-Month)")
plt.ylabel("Wind Speed (m/s)")
```

```
plt.grid(True)
plt.tight_layout()
plt.show()

# 6.
from scipy.stats import linregress

x = monthly_avg.index.to_timestamp()
x_num = (x - x.min()).days  #
slope, intercept, r_value, p_value, std_err = linregress(x_num, monthly_avg.
 ↪values)

print(f"  : {slope:.4f} m/s per day, p ={p_value:.4f}")
if p_value < 0.05:
    if slope > 0:
        print("       ")
    else:
        print("       ")
else:
    print("      ")
```




```
  : 0.0002 m/s per day, p =0.0000
```

[7]:
```
#3. Explore a data set

import pandas as pd

file_path = r"D:\code\      .csv"
```

```
#    GBK
df = pd.read_csv(file_path, encoding='gbk', on_bad_lines='skip')

#
print(df.head())
print(df.columns)



#          AQI   NaN
df.columns = df.columns.str.strip()
df_clean = df.dropna()
df_clean = df_clean[df_clean['AQI '] >= 0]  #   AQI
```

```
                  AQI     AQI   PM2.5  PM10  So2  No2    Co  O3
0     2022/3/1          25        9      6    17    2   10  0.25  68
1     2022/3/2          43       52     16    44    2   25  0.40  49
2     2022/3/3          57       98     24    68    4   26  0.56  53
3     2022/3/4          72      158     27    97    2   24  0.47  54
4     2022/3/5          32       25      7    31    2   10  0.23  66
Index([' ', ' ', '  ', 'AQI  ', ' AQI ', 'PM2.5', 'PM10', 'So2', 'No2',
       'Co', 'O3'],
      dtype='object')
```

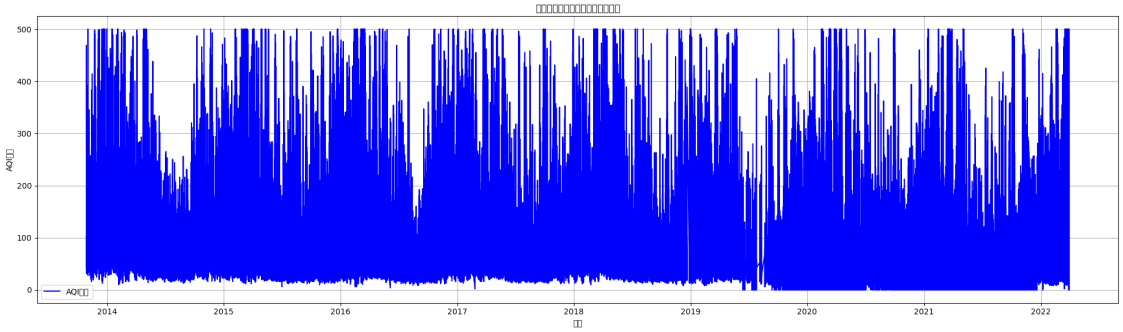
[11]:
```
import matplotlib.pyplot as plt

#       ' '    datetime
df_clean[' '] = pd.to_datetime(df_clean[' '], errors='coerce')
df_clean = df_clean.dropna(subset=[' '])  #

#
df_clean = df_clean.sort_values(' ')

#   AQI
plt.figure(figsize=(12, 6))
plt.plot(df_clean[' '], df_clean['AQI '], label='AQI ', color='blue')
plt.xlabel(' ')
plt.ylabel('AQI ')
plt.title('        ')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```

```
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 26085 (\N{CJK UNIFIED IDEOGRAPH-65E5}) missing from font(s)
DejaVu Sans.
```

```
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 26399 (\N{CJK UNIFIED IDEOGRAPH-671F}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 25351 (\N{CJK UNIFIED IDEOGRAPH-6307}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 25968 (\N{CJK UNIFIED IDEOGRAPH-6570}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 20840 (\N{CJK UNIFIED IDEOGRAPH-5168}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 22269 (\N{CJK UNIFIED IDEOGRAPH-56FD}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 20027 (\N{CJK UNIFIED IDEOGRAPH-4E3B}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 35201 (\N{CJK UNIFIED IDEOGRAPH-8981}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 22478 (\N{CJK UNIFIED IDEOGRAPH-57CE}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 24066 (\N{CJK UNIFIED IDEOGRAPH-5E02}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 31354 (\N{CJK UNIFIED IDEOGRAPH-7A7A}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 27668 (\N{CJK UNIFIED IDEOGRAPH-6C14}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 36136 (\N{CJK UNIFIED IDEOGRAPH-8D28}) missing from font(s)
DejaVu Sans.
```

```
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 37327 (\N{CJK UNIFIED IDEOGRAPH-91CF}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 26102 (\N{CJK UNIFIED IDEOGRAPH-65F6}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 38388 (\N{CJK UNIFIED IDEOGRAPH-95F4}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 24207 (\N{CJK UNIFIED IDEOGRAPH-5E8F}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 21015 (\N{CJK UNIFIED IDEOGRAPH-5217}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
C:\Users\45527\AppData\Local\Temp\ipykernel_293256\1673751191.py:18:
UserWarning: Glyph 22270 (\N{CJK UNIFIED IDEOGRAPH-56FE}) missing from font(s)
DejaVu Sans.
  plt.tight_layout()
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
25351 (\N{CJK UNIFIED IDEOGRAPH-6307}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
25968 (\N{CJK UNIFIED IDEOGRAPH-6570}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
20840 (\N{CJK UNIFIED IDEOGRAPH-5168}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
22269 (\N{CJK UNIFIED IDEOGRAPH-56FD}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
20027 (\N{CJK UNIFIED IDEOGRAPH-4E3B}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
35201 (\N{CJK UNIFIED IDEOGRAPH-8981}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
22478 (\N{CJK UNIFIED IDEOGRAPH-57CE}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
24066 (\N{CJK UNIFIED IDEOGRAPH-5E02}) missing from font(s) DejaVu Sans.
```

```
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
31354 (\N{CJK UNIFIED IDEOGRAPH-7A7A}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
27668 (\N{CJK UNIFIED IDEOGRAPH-6C14}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
36136 (\N{CJK UNIFIED IDEOGRAPH-8D28}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
37327 (\N{CJK UNIFIED IDEOGRAPH-91CF}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
26102 (\N{CJK UNIFIED IDEOGRAPH-65F6}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
38388 (\N{CJK UNIFIED IDEOGRAPH-95F4}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
24207 (\N{CJK UNIFIED IDEOGRAPH-5E8F}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
21015 (\N{CJK UNIFIED IDEOGRAPH-5217}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
22270 (\N{CJK UNIFIED IDEOGRAPH-56FE}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
26085 (\N{CJK UNIFIED IDEOGRAPH-65E5}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
D:\anaconda\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning: Glyph
26399 (\N{CJK UNIFIED IDEOGRAPH-671F}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



[14]:
```
#
print("AQI      ")
print(df_clean['AQI '].describe())

#
print("  ", df_clean['AQI '].max())
print("  ", df_clean['AQI '].min())
print("  ", df_clean['AQI '].median())
print("  ", df_clean['AQI '].std())
print("    AQI  > 500 ", (df_clean['AQI '] > 500).sum())
```

```
AQI
count    974449.000000
mean         70.868673
std          46.885637
min           0.000000
25%          42.000000
50%          60.000000
75%          84.000000
max         500.000000
Name: AQI , dtype: float64
   500
   0
   60.0
   46.88563718814081
    AQI  > 500  0
```

[ ]: