initial

April 30, 2025

[26]:
```
#Task 1: Import and Clean Data
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

df = pd.read_csv("FloridaBikeRentals.csv", encoding='latin1')

df['Date'] = pd.to_datetime(df['Date'].str.replace('-', '/'), dayfirst=True,␣
 ↪errors='coerce')
df.dropna(subset=['Date'], inplace=True)

for col in ['Temperature(°C)' , 'Wind speed (m/s)' , 'Solar Radiation (MJ/m2)']:
    df[col] = df[col].astype('float32')

for col in ['Seasons', 'Holiday', 'Functioning Day']:
    df[col] = df[col].astype('category')

#Format to JSON
df.to_json("bike_rental_cleaned.json", orient="records", lines=True)
'''
## Data Observations Summary

- No missing or duplicate values were found after investigating the data.
- Date column appears to have inconsistent formating, alternating between '-'␣
 ↪and '/'
'''
```

[5]:
```
#Task 2 Data Processing and Statistical Analysis

import pandas as pd
from sklearn.preprocessing import MinMaxScaler

df = pd.read_csv("FloridaBikeRentals.csv", encoding='latin1')
raw_df = pd.read_csv("FloridaBikeRentals.csv", encoding='latin1')
scaler = MinMaxScaler()
```

```
#Data Transformations

df['Temperature(°C)'] = df['Temperature(°C)'] * 10
df['Visibility_Scaled'] = scaler.fit_transform(df[['Visibility (10m)']])

print(df['Visibility_Scaled'].head(20), "\n\n")


# Basic Statiscal Analysis
print("Before Transformation\n" , raw_df[['Temperature(°C)', 'Humidity(%)',␣
 ↪'Rented Bike Count']].describe())
print("\nAfter Transformation\n" ,df[['Temperature(°C)', 'Humidity(%)', 'Rented␣
 ↪Bike Count']].describe())



'''
## Data Reformating

Changing columns' values to int format for faster processing and categorizing
> df['Holiday']:    'No Holiday' ==> 0 | 'Holiday' ==> 1
> df['Functioning Day']     'No' ==> 0 | 'Yes' ==> 1
> df['Seasons']     'Winter' ==> 0 | 'Spring' ==> 1 | 'Summer' ==> 2 | 'Autumn'␣
 ↪==> 3
'''
df['Date'] = pd.to_datetime(df['Date'].str.replace('-', '/'), dayfirst=True,␣
 ↪errors='coerce')
df.dropna(subset=['Date'], inplace=True)

df['Holiday'] = df['Holiday'].map({'No Holiday': 0, 'Holiday': 1})
df['Functioning Day'] = df['Functioning Day'].map({'No': 0, 'Yes': 1})
df['Seasons'] = df['Seasons'].map({'Winter': 0, 'Spring': 1, 'Summer': 2,␣
 ↪'Autumn': 3})

df.to_csv("bike_rental_processed.csv", index=False)
```

```
0     1.000000
1     1.000000
2     1.000000
3     1.000000
4     1.000000
5     1.000000
6     1.000000
7     1.000000
8     1.000000
9     0.963507
10    0.997973
```

```
11    0.967562
12    1.000000
13    1.000000
14    1.000000
15    1.000000
16    0.388241
17    1.000000
18    1.000000
19    1.000000
Name: Visibility_Scaled, dtype: float64


Before Transformation
        Temperature(°C)  Humidity(%)  Rented Bike Count
count      8760.000000  8760.000000        8760.000000
mean         12.882922    58.226256         704.602055
std          11.944825    20.362413         644.997468
min         -17.800000     0.000000           0.000000
25%           3.500000    42.000000         191.000000
50%          13.700000    57.000000         504.500000
75%          22.500000    74.000000        1065.250000
max          39.400000    98.000000        3556.000000

After Transformation
        Temperature(°C)  Humidity(%)  Rented Bike Count
count      8760.000000  8760.000000        8760.000000
mean        128.829224    58.226256         704.602055
std         119.448252    20.362413         644.997468
min        -178.000000     0.000000           0.000000
25%          35.000000    42.000000         191.000000
50%         137.000000    57.000000         504.500000
75%         225.000000    74.000000        1065.250000
max         394.000000    98.000000        3556.000000
```

[6]:
```
#Task 3 Data Analysis with Pandas

df = pd.read_csv('bike_rental_processed.csv')

#Finding averages to discover trends
seasonal_average_bike_count = df.groupby('Seasons')['Rented Bike Count'].mean()
print(seasonal_average_bike_count, "\n\n0 = Winter\n1 = Spring\n2 = Summer\n3 =␣
 ↪Autumn\n" )

holiday_average_bike_count = df.groupby('Holiday')['Rented Bike Count'].mean()
print(holiday_average_bike_count, '\n\n0 = No Holiday\n1 = Holiday\n')
```

```
functioning_day_average_bike_count = df.groupby('Functioning Day')['Rented Bike␣
 ↪Count'].mean()
print(functioning_day_average_bike_count, '\n\n0 = Bikes are not available\n1 =␣
 ↪Bikes are available\n')


#Creating Distribution Tables
hour_rent_description = df.groupby('Hour')['Rented Bike Count'].describe()
hour_rent_sum = df.groupby('Hour')['Rented Bike Count'].sum()

season_rent_description = df.groupby('Seasons')['Rented Bike Count'].describe()
season_rent_sum = df.groupby('Seasons')['Rented Bike Count'].sum()

temp_rent_description = df.groupby('Temperature(°C)')['Rented Bike Count'].
 ↪describe()
temp_rent_sum = df.groupby('Temperature(°C)')['Rented Bike Count'].sum()

hour_temp_description = df.groupby('Hour')['Temperature(°C)'].describe()
hour_temp_sum = df.groupby('Hour')['Temperature(°C)'].sum()


print('Rented Bikes by Hour Stats\n' , hour_rent_description , hour_rent_sum,␣
 ↪'\n\n')
print('Rented Bikes by Season Stats\n' , season_rent_description ,␣
 ↪season_rent_sum, '\n\n')
print('Rented Bikes by Temperature Stats\n' , temp_rent_description ,␣
 ↪temp_rent_sum, '\n\n')
print('Temperatures by Hour Stats\n' , hour_temp_description , hour_temp_sum,␣
 ↪'\n\n')


#Encoding the dataframe
binary_dataframe = pd.get_dummies(df, columns=['Seasons', 'Holiday',␣
 ↪'Functioning Day'])
binary_dataframe.to_csv("Rental_Bike_Data_Dummy.csv", index=False)
```

```
Seasons
0     225.541204
1     730.031250
2    1034.073370
3     819.597985
Name: Rented Bike Count, dtype: float64

0 = Winter
1 = Spring
2 = Summer
3 = Autumn
```

```
Holiday
0    715.228026
1    499.756944
Name: Rented Bike Count, dtype: float64

0 = No Holiday
1 = Holiday

Functioning Day
0      0.000000
1    729.156999
Name: Rented Bike Count, dtype: float64

0 = Bikes are not available
1 = Bikes are available

Rented Bikes by Hour Stats
       count         mean          std  min    25%     50%     75%     max
Hour
0      365.0   541.460274   364.573274  0.0  196.0   513.0   862.0  1394.0
1      365.0   426.183562   285.528653  0.0  172.0   401.0   669.0  1088.0
2      365.0   301.630137   210.105083  0.0  122.0   265.0   456.0  1254.0
3      365.0   203.331507   143.203525  0.0   79.0   176.0   307.0   644.0
4      365.0   132.591781    90.272058  0.0   51.0   119.0   199.0   421.0
5      365.0   139.082192    95.524974  0.0   57.0   129.0   215.0   383.0
6      365.0   287.564384   222.818717  0.0   97.0   232.0   448.0   807.0
7      365.0   606.005479   482.256817  0.0  211.0   426.0  1065.0  1629.0
8      365.0  1015.701370   761.594619  0.0  401.0   728.0  1785.0  2495.0
9      365.0   645.983562   398.645956  0.0  289.0   680.0   985.0  1401.0
10     365.0   527.821918   323.012687  0.0  212.0   581.0   811.0  1269.0
11     365.0   600.852055   361.955313  0.0  256.0   624.0   934.0  1478.0
12     365.0   699.441096   430.738874  0.0  293.0   709.0  1072.0  1798.0
13     365.0   733.246575   457.724147  0.0  304.0   727.0  1071.0  2000.0
14     365.0   758.824658   488.851830  0.0  329.0   733.0  1099.0  2128.0
15     365.0   829.186301   546.471341  0.0  333.0   785.0  1209.0  2329.0
16     365.0   930.621918   618.018110  0.0  352.0   911.0  1434.0  2479.0
17     365.0  1138.509589   748.947282  0.0  392.0  1184.0  1845.0  2664.0
18     365.0  1502.926027  1029.301642  0.0  531.0  1548.0  2359.0  3556.0
19     365.0  1195.147945   857.420198  0.0  371.0  1224.0  1937.0  2984.0
20     365.0  1068.964384   793.904872  0.0  277.0  1062.0  1726.0  2579.0
21     365.0  1031.449315   753.645522  0.0  271.0  1046.0  1717.0  2505.0
22     365.0   922.797260   660.794396  0.0  273.0   949.0  1501.0  2309.0
23     365.0   671.126027   478.779780  0.0  212.0   656.0  1087.0  1732.0 Hour
0     197633
1     155557
2     110095
3      74216
```

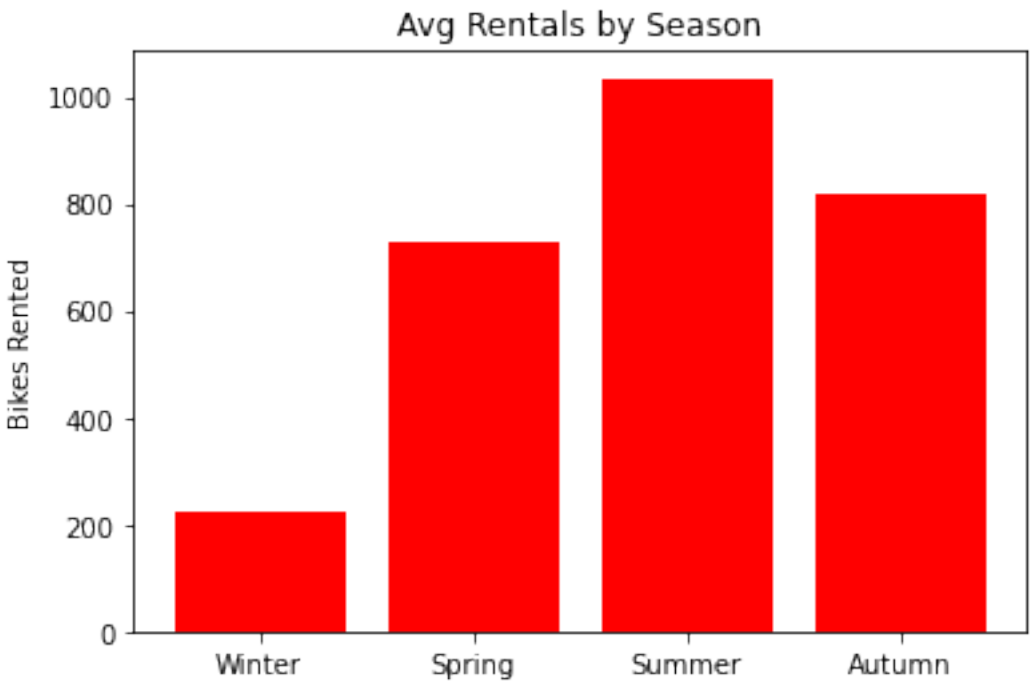
```
4      48396
5      50765
6     104961
7     221192
8     370731
9     235784
10    192655
11    219311
12    255296
13    267635
14    276971
15    302653
16    339677
17    415556
18    548568
19    436229
20    390172
21    376479
22    336821
23    244961
Name: Rented Bike Count, dtype: int64


Rented Bikes by Season Stats
           count         mean         std  min     25%    50%      75%     max
Seasons
0         2160.0   225.541204  150.372236  3.0  110.00  203.0   305.00   937.0
1         2208.0   730.031250  621.509635  0.0  206.00  583.0  1105.25  3251.0
2         2208.0  1034.073370  690.244759  9.0  526.75  905.5  1442.50  3556.0
3         2184.0   819.597985  651.085621  0.0  241.75  763.5  1197.50  3298.0
Seasons
0     487169
1    1611909
2    2283234
3    1790002
Name: Rented Bike Count, dtype: int64


Rented Bikes by Temperature Stats
                   count    mean       std     min      25%     50%      75%  \
Temperature(°C)
-178.0               1.0   322.0       NaN   322.0   322.00   322.0   322.00
-175.0               2.0   145.5  4.949747   142.0   143.75   145.5   147.25
-174.0               1.0    64.0       NaN    64.0    64.00    64.0    64.00
-169.0               1.0    36.0       NaN    36.0    36.00    36.0    36.00
-165.0               1.0    96.0       NaN    96.0    96.00    96.0    96.00
…                  …      …        …      …       …      …       …
 380.0               1.0  1184.0       NaN  1184.0  1184.00  1184.0  1184.00
```

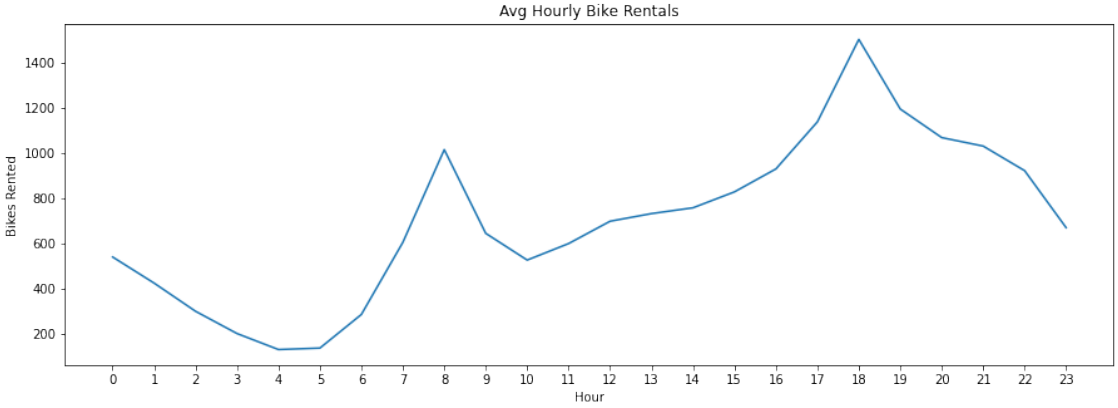
```
 387.0             1.0   475.0       NaN   475.0   475.00   475.0   475.00
 390.0             1.0  1033.0       NaN  1033.0  1033.00  1033.0  1033.00
 393.0             1.0   531.0       NaN   531.0   531.00   531.0   531.00
 394.0             1.0   561.0       NaN   561.0   561.00   561.0   561.00

                    max
Temperature(°C)
-178.0            322.0
-175.0            149.0
-174.0             64.0
-169.0             36.0
-165.0             96.0
…                  …
 380.0            1184.0
 387.0             475.0
 390.0            1033.0
 393.0             531.0
 394.0             561.0

[546 rows x 8 columns] Temperature(°C)
-178.0     322
-175.0     291
-174.0      64
-169.0      36
-165.0      96
          …
 380.0    1184
 387.0     475
 390.0    1033
 393.0     531
 394.0     561
Name: Rented Bike Count, Length: 546, dtype: int64


Temperatures by Hour Stats
       count        mean         std    min   25%    50%    75%    max
Hour
0      365.0  112.863014  113.684317 -159.0  26.0  119.0  207.0  321.0
1      365.0  109.232877  113.449258 -161.0  21.0  115.0  203.0  317.0
2      365.0  105.915068  113.470362 -160.0  20.0  111.0  200.0  315.0
3      365.0  102.936986  113.181192 -160.0  16.0  107.0  196.0  312.0
4      365.0  100.263014  112.908121 -164.0  16.0  103.0  192.0  311.0
5      365.0   97.687671  112.852481 -169.0  11.0   98.0  191.0  307.0
6      365.0   95.605479  113.087528 -174.0  11.0   96.0  189.0  305.0
7      365.0   95.810959  114.992140 -175.0   9.0   99.0  193.0  308.0
8      365.0  101.769863  119.281050 -178.0  11.0  105.0  205.0  318.0
9      365.0  113.758904  122.022525 -175.0  22.0  120.0  215.0  330.0
10     365.0  129.101370  120.987935 -165.0  36.0  144.0  227.0  344.0
```

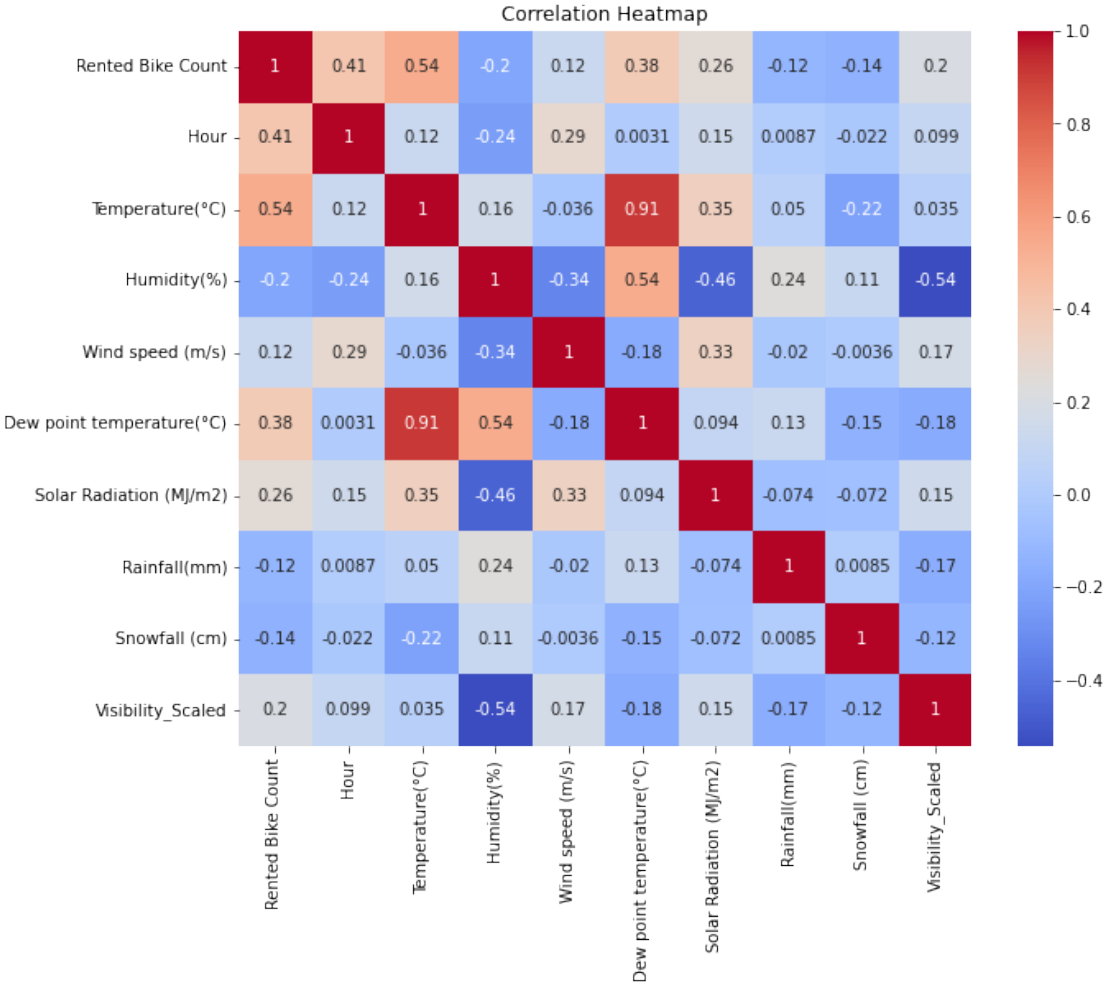
```
11    365.0  143.101370  119.590341 -156.0  48.0  162.0  238.0  358.0
12    365.0  154.621918  118.882293 -144.0  59.0  175.0  249.0  368.0
13    365.0  162.556164  118.778514 -133.0  67.0  183.0  257.0  379.0
14    365.0  168.167123  118.371770 -121.0  73.0  189.0  263.0  387.0
15    365.0  170.361644  118.433250 -114.0  70.0  190.0  264.0  393.0
16    365.0  168.964384  118.781989 -112.0  66.0  188.0  263.0  394.0
17    365.0  162.547945  120.434589 -118.0  60.0  182.0  256.0  390.0
18    365.0  153.008219  121.367766 -130.0  52.0  168.0  250.0  378.0
19    365.0  142.805479  119.007470 -138.0  45.0  153.0  240.0  361.0
20    365.0  133.895890  116.265541 -143.0  37.0  142.0  230.0  345.0
21    365.0  127.452055  114.845266 -147.0  31.0  134.0  221.0  333.0
22    365.0  122.076712  114.195223 -152.0  28.0  128.0  217.0  329.0
23    365.0  117.397260  113.735492 -157.0  28.0  124.0  213.0  324.0 Hour
0     41195.0
1     39870.0
2     38659.0
3     37572.0
4     36596.0
5     35656.0
6     34896.0
7     34971.0
8     37146.0
9     41522.0
10    47122.0
11    52232.0
12    56437.0
13    59333.0
14    61381.0
15    62182.0
16    61672.0
17    59330.0
18    55848.0
19    52124.0
20    48872.0
21    46520.0
22    44558.0
23    42850.0
Name: Temperature(°C), dtype: float64
```

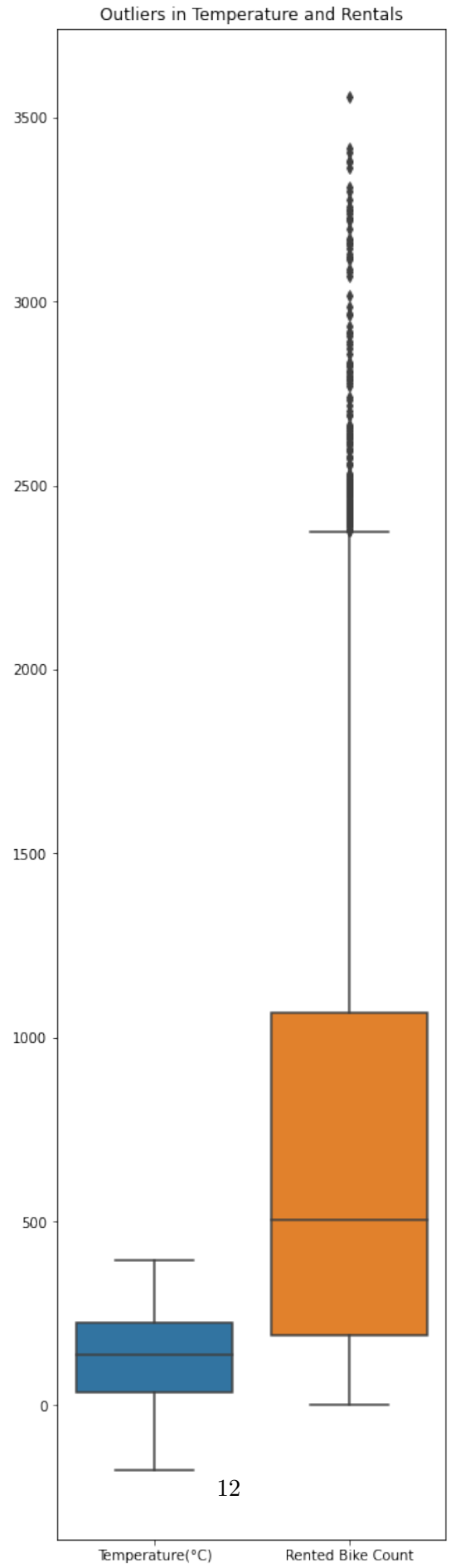
[25]:
```
#Task 4 Data Visualization

import matplotlib.pyplot as plt
import seaborn as sns

# Average Rental By Season Bar Graph
```

```
plt.figure()
rents_by_season = df.groupby('Seasons')['Rented Bike Count'].mean().
 ↪reset_index()
rents_by_season['Seasons'] = rents_by_season['Seasons'].map({0: 'Winter', 1:␣
 ↪'Spring', 2: 'Summer', 3: 'Autumn'})

plt.bar(rents_by_season['Seasons'] , rents_by_season['Rented Bike Count'] ,␣
 ↪color='red', )
plt.title('Avg Rentals by Season')
plt.ylabel('Bikes Rented')
plt.savefig('avg_rentals_by_season.png')

#Average Rental by Hour Line Graph
plt.figure(figsize=(15,5))
plt.xticks([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,␣
 ↪19, 20, 21, 22, 23])
plt.ylabel('Bikes Rented')
rents_by_hour = df.groupby('Hour')['Rented Bike Count'].mean().
 ↪plot(kind='line', title='Avg Hourly Bike Rentals')
plt.savefig('avg_rentals_by_hour.png')

#Correlation Heatmap
numerical_variables = df[['Rented Bike Count' , 'Hour' , 'Temperature(°C)' ,␣
 ↪'Humidity(%)' , 'Wind speed (m/s)' , 'Dew point temperature(°C)' , 'Solar␣
 ↪Radiation (MJ/m2)' , 'Rainfall(mm)' , 'Snowfall (cm)' , 'Visibility_Scaled']]
plt.figure(figsize=(10,8))
sns.heatmap(numerical_variables.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.savefig('correlation_heatmap.png')

#Boxplot for Outliers
plt.figure(figsize=(5, 20))
sns.boxplot(data=df[['Temperature(°C)', 'Rented Bike Count']])
plt.title('Outliers in Temperature and Rentals')
plt.savefig('boxplot_outliers.png')
```


Avg Rentals by Season
Bikes Rented
1000
800
600
400
200
0
Winter
Spring
Summer
Autumn


Avg Hourly Bike Rentals
Bikes Rented
1400
1200
1000
800
600
400
200
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour


Correlation Heatmap
Rented Bike Count
Hour
Temperature(°C)
Humidity(%)
Wind speed (m/s)
Dew point temperature(°C)
Solar Radiation (MJ/m2)
Rainfall(mm)
Snowfall (cm)
Visibility_Scaled
1 0.41 0.54 -0.2 0.12 0.38 0.26 -0.12 -0.14 0.2
0.41 1 0.12 -0.24 0.29 0.0031 0.15 0.0087 -0.022 0.099
0.54 0.12 1 0.16 -0.036 0.91 0.35 0.05 -0.22 0.035
-0.2 -0.24 0.16 1 -0.34 0.54 -0.46 0.24 0.11 -0.54
0.12 0.29 -0.036 -0.34 1 -0.18 0.33 -0.02 -0.0036 0.17
0.38 0.0031 0.91 0.54 -0.18 1 0.094 0.13 -0.15 -0.18
0.26 0.15 0.35 -0.46 0.33 0.094 1 -0.074 -0.072 0.15
-0.12 0.0087 0.05 0.24 -0.02 0.13 -0.074 1 0.0085 -0.17
-0.14 -0.022 -0.22 0.11 -0.0036 -0.15 -0.072 0.0085 1 -0.12
0.2 0.099 0.035 -0.54 0.17 -0.18 0.15 -0.17 -0.12 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Outliers in Temperature and Rentals
3500
3000
2500
2000
1500
1000
500
0
Temperature(°C)
Rented Bike Count

[ ]: