> Styling

[ ] ↳ 1 cell hidden

Renewable Energy Optimization

Renewable Energy Optimization
Authored by: Sinan Kilci

Duration: 600 mins
Level: Intermediate
Pre-requisite Skills: Python

The main goal of this project is to use historical environmental data to improve the installation and operation of renewable energy sources like wind turbines and solar panels. The aim is to make energy production more efficient and sustainable by studying environmental factors that affect energy production. The results will help us understand the best places to put renewable energy sources and how to operate them more effectively, using detailed environmental information to boost efficiency and sustainability.

At the end of this use case we will:

- have a clear understanding of how environmental data can be leveraged to optimize the placement and operation of renewable energy sources like wind turbines and solar panels
- access to visualizations that effectively communicate the trends and patterns found in the data, providing an intuitive grasp of complex environmental factors
- explore detailed data analysis and wrangling techniques that refine raw data into actionable insights, highlighting key variables affecting renewable energy efficiency and sustainability
- be exposed to predictive models that forecast energy production potential based on historical environment data, offering strategic insights for future planning and investment.

Introduction

Optimizing Renewable Energy Deployment Using Environmental Sensor Data

In recent years, the global need for renewable energy has been increased as we strive to reduce carbon emissions and reduce the effects of climate change. Renewable energy sources, such as wind turbines and solar panels, play a critical role in this transition. However, the efficiency and effectiveness of these installations are heavily influenced by environmental conditions. Placing a wind turbine in an area with inconsistent wind patterns or a solar panel where sunlight is obstructed can lead to weak performance and increased costs.

This study aims to address these challenges by leveraging detailed environmental data from microclimate and meshed sensor networks. By analyzing factors such as wind speed, solar radiation, temperature, and humidity, we can identify optimal locations and operational strategies for renewable energy installations. The goal is to maximize energy generation, improve cost-efficiency, and reduce environmental impact.

The datasets used in this study are gathered from City of Melbourne Open Data Platform (https://data.melbourne.vic.gov.au) and provide extensive insights into local environmental conditions in Melbourne City. They include:

- Weather Stations Data (ATMOS 41): Historical data collected by weather stations installed in Argyle Square, capturing wind speed, solar radiation, and atmospheric conditions. This information is crucial for assessing site suitability for renewable energy projects.
- Microclimate Sensors Data: Contains climate readings from sensors located within the city, updated every fifteen minutes. This dataset includes ambient air temperature, relative humidity, atmospheric pressure, wind speed and direction, gust wind speed, particulate matter 2.5, particulate matter 10, and noise. It is essential for understanding microclimate variations throughout the day.
- Microclimate Sensor Locations: Provides the historical location and description for each microclimate sensor device installed throughout the city. This data is vital for understanding the spatial distribution of sensors and any relocations that may affect historical data interpretation.
- Microclimate Sensor Readings: Offers environmental readings updated every hour in fifteen-minute increments. It includes data on ambient air temperature, relative humidity, barometric pressure, particulate matter 2.5, particulate matter 10, and average wind speed. This dataset also aligns with EPA Victoria's air quality data, helping correlate microclimate conditions with broader environmental insights.

Importing the required libraries

```
import requests
from io import StringIO
import pandas as pd
import numpy as np
```

```
import numpy as np
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.impute import SimpleImputer
import matplotlib.pyplot as plt
```

Function to request the datasets from their sources (APIv2.1)

```
def request_data(dataset_id, format='csv', delimiter=';', api_key=None):
  base_url = 'https://data.melbourne.vic.gov.au/api/explore/v2.1/catalog/datasets/'
  url = f'{base_url}{dataset_id}/exports/{format}'

  params = {
      'select': '*',
      'limit': -1,
      'lang': 'en',
      'timezone': 'UTC'
  }

  headers = {}
  if api_key:
    headers['Authorization'] = f'Bearer {api_key}'

  try:
    response = requests.get(url, params=params, headers=headers, timeout=10)
    response.raise_for_status()

    if format == 'csv':
      content = response.content.decode('utf-8')
      df = pd.read_csv(StringIO(content), delimiter=delimiter)
      return df
    else:
      raise ValueError(f"Unsupported format: {format}")

  except requests.exceptions.RequestException as e:
    raise Exception(f'API request failed: {e}')
```

Loading datasets

```
microclimate_sensor_readings = request_data('microclimate-sensor-readings', format='csv')
microclimate_sensors_data = request_data('microclimate-sensors-data', format='csv')
meshed_sensor_type_1 = request_data('meshed-sensor-type-1', format='csv')

microclimate_sensor_readings.head()
microclimate_sensors_data.head()
meshed_sensor_type_1.head()
```

	dev_id	time	rtc	battery	solarpanel	command	solar	precip

0	atmos41-32fc	2021-05-14 18:11:23+00:00	10090042.0	4.161	0.024	0.0	0.0
1	atmos41-32fc	2022-05-03 10:25:31+00:00	17564045.0	4.181	0.000	0.0	0.0
2	atmos41-32fc	2022-05-03 21:42:23+00:00	17604657.0	4.143	20.670	0.0	24.0
3	atmos41-32fc	2022-05-04 00:42:43+00:00	17615477.0	4.208	20.694	0.0	68.0
4	atmos41-32fc	2021-05-15 07:55:22+00:00	10139481.0	4.197	0.128	0.0	0.0

Converting time related columns to datetime

```
microclimate_sensor_readings['local_time'] = pd.to_datetime(microclimate_sensor_readings[
microclimate_sensors_data['received_at'] = pd.to_datetime(microclimate_sensors_data['rece
meshed_sensor_type_1['time'] = pd.to_datetime(meshed_sensor_type_1['time'], errors='coerc
```

Check for missing values

```
print("Total rows in wrangled_microclimate_sensor_readings:", len(microclimate_sensor_rea
print("Total rows in wrangled_microclimate_sensors_data:", len(microclimate_sensors_data)
print("Total rows in wrangled_meshed_sensor_type_1:", len(meshed_sensor_type_1))

print("Missing Values in Microclimate Sensor Readings:\n", microclimate_sensor_readings.i
print("Missing Values in Microclimate Sensors Data:\n", microclimate_sensors_data.isnull(
print("Missing Values in Meshed Sensor Type 1:\n", meshed_sensor_type_1.isnull().sum())
```

```
Total rows in wrangled_microclimate_sensor_readings: 56
Total rows in wrangled_microclimate_sensors_data: 52110
Total rows in wrangled_meshed_sensor_type_1: 119925
Missing Values in Microclimate Sensor Readings:
 local_time      0
id              0
site_id         0
sensor_id       0
value           0
type            0
units           0
gatewayhub_id   0
site_status     0
dtype: int64
Missing Values in Microclimate Sensors Data:
 device_id                  0
received_at                0
sensorlocation           803
```

```
latlong                  803
minimumwinddirection    6805
averagewinddirection      20
maximumwinddirection    6805
minimumwindspeed        6805
averagewindspeed          20
gustwindspeed           6805
airtemperature            20
relativehumidity          20
atmosphericpressure       20
pm25                    4725
pm10                    4725
noise                   4725
dtype: int64
Missing Values in Meshed Sensor Type 1:
 dev_id                   0
time                     0
rtc                   2677
battery               2088
solarpanel            2196
command               2129
solar                 2783
precipitation         2784
strikes               2784
windspeed             2199
winddirection         2200
gustspeed             2200
vapourpressure        2784
atmosphericpressure   2200
relativehumidity      2200
airtemp               2783
lat_long              2677
sensor_name           2677
dtype: int64
```

Sorting the data by time

```
microclimate_sensors_data.sort_values('received_at', inplace=True)
```

Extract time-based features

```
microclimate_sensors_data['hour'] = microclimate_sensors_data['received_at'].dt.hour
microclimate_sensors_data['day'] = microclimate_sensors_data['received_at'].dt.day
microclimate_sensors_data['month'] = microclimate_sensors_data['received_at'].dt.month
```

Handling the missing values

```
# Identify existing device_id to sensorlocation and latlong mappings
```

```
location_mapping = microclimate_sensors_data.dropna(subset=['sensorlocation']).set_index('d
latlong_mapping = microclimate_sensors_data.dropna(subset=['latlong']).set_index('device_id

# Fill missing sensorlocation values based on device_id
microclimate_sensors_data['sensorlocation'] = microclimate_sensors_data.apply(
    lambda row: location_mapping.get(row['device_id'], row['sensorlocation']), axis=1
)

# Fill missing latlong values based on device_id
microclimate_sensors_data['latlong'] = microclimate_sensors_data.apply(
    lambda row: latlong_mapping.get(row['device_id'], row['latlong']), axis=1
)

# Check if the missing values in sensorlocation and latlong are filled
print("Missing Values in Sensor Location after filling:\n", microclimate_sensors_data['sens
print("Missing Values in latlong after filling:\n", microclimate_sensors_data['latlong'].is

# Drop rows where all specified wind-related columns are 0
columns_to_check = [
    'minimumwinddirection', 'averagewinddirection', 'maximumwinddirection',
    'minimumwindspeed', 'averagewindspeed', 'gustwindspeed'
]

microclimate_sensors_data = microclimate_sensors_data[
    ~(microclimate_sensors_data[columns_to_check] == 0).all(axis=1)
].copy()

microclimate_sensors_data.dropna(subset=['airtemperature', 'relativehumidity', 'atmospheric
meshed_sensor_type_1.dropna(inplace=True)
```

```
Missing Values in Sensor Location after filling:
 0
Missing Values in latlong after filling:
 0
```

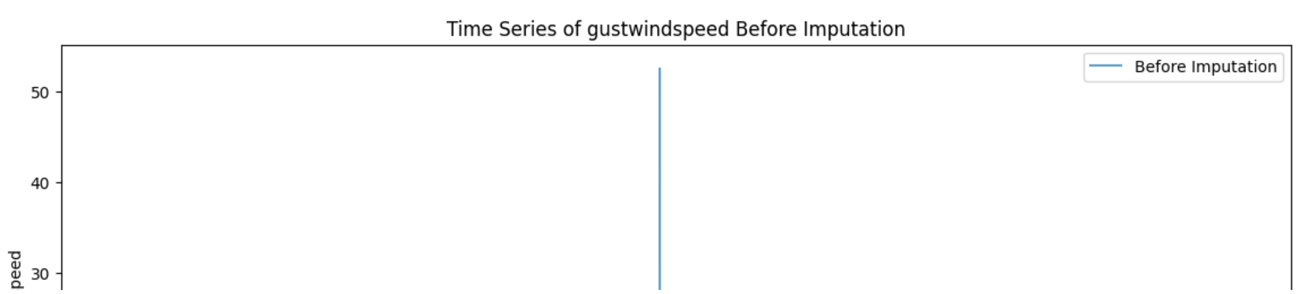
Drop the unnecessary columns

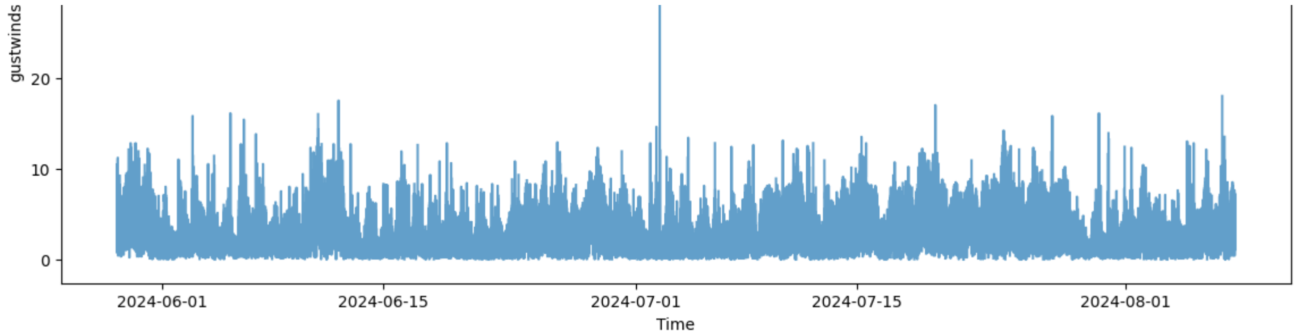
```
microclimate_sensors_data.drop(columns=['pm10', 'pm25', 'noise', 'minimumwinddirection',
```

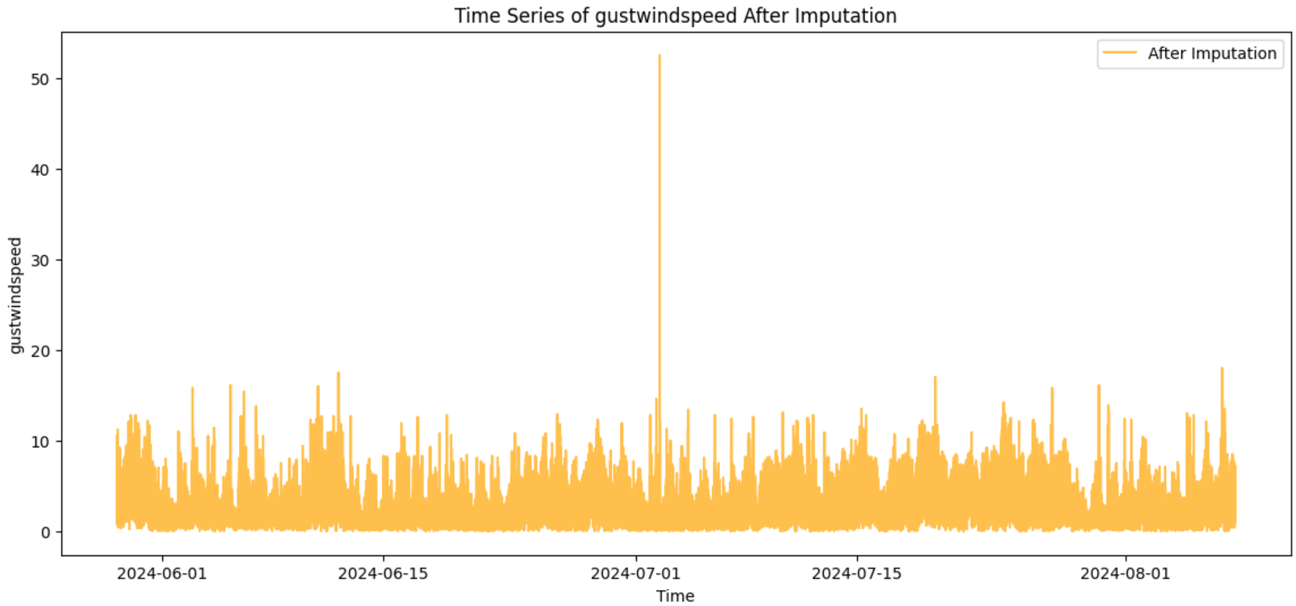
Handling and checking the data in empty fields in Microclimate dataset

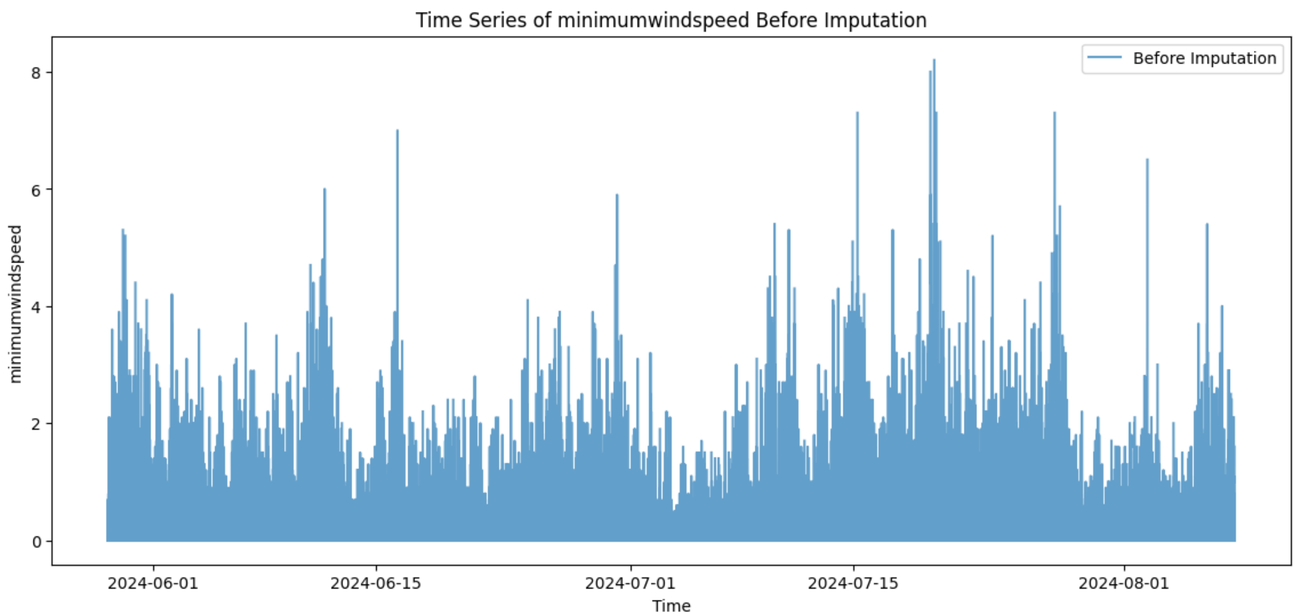
```
def impute_and_plot_knn(target_column, feature_columns, data):
  plt.figure(figsize=(14, 6))
  plt.plot(data['received_at'], data[target_column], label='Before Imputation', alpha=0.7)
  plt.title(f'Time Series of {target_column} Before Imputation')
  plt.xlabel('Time')
  plt.ylabel(target_column)
```

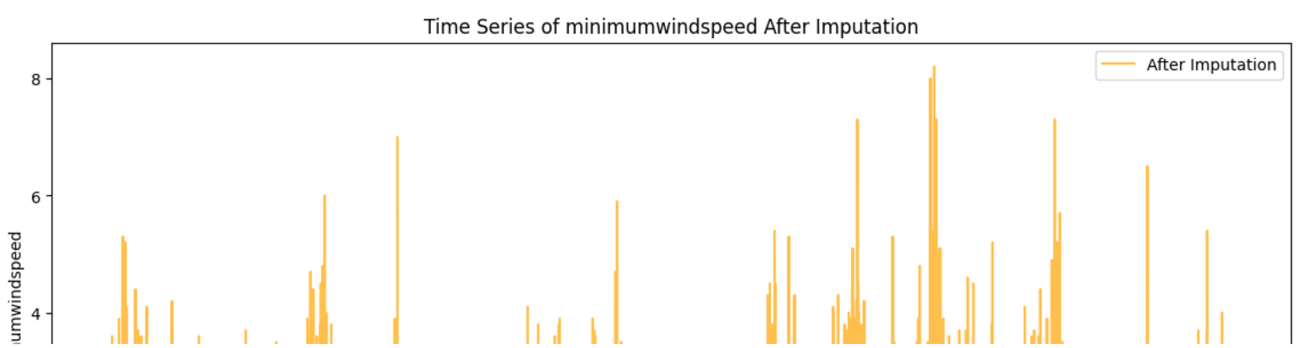
```
    plt.legend()
    plt.show()

    complete_data = data.dropna(subset=[target_column])
    missing_data = data[data[target_column].isnull()]

    X = complete_data[feature_columns]
    y = complete_data[target_column]

    imputer = SimpleImputer(strategy='mean')
    X_imputed = imputer.fit_transform(X)

    X_train, X_test, y_train, y_test = train_test_split(X_imputed, y, test_size=0.2, random_s

    # KNN regression model
    knn_model = KNeighborsRegressor(n_neighbors=5)
    knn_model.fit(X_train, y_train)

    missing_X = missing_data[feature_columns]
    missing_X_imputed = imputer.transform(missing_X)

    predicted_values = knn_model.predict(missing_X_imputed)

    # Ensure non-negative predictions for wind speed
    predicted_values = np.maximum(predicted_values, 0)

    data.loc[missing_data.index, target_column] = predicted_values

    plt.figure(figsize=(14, 6))
    plt.plot(data['received_at'], data[target_column], label='After Imputation', alpha=0.7, c
    plt.title(f'Time Series of {target_column} After Imputation')
    plt.xlabel('Time')
    plt.ylabel(target_column)
    plt.legend()
    plt.show()

# Define target variables and their respective features
targets = ['gustwindspeed', 'minimumwindspeed']
features = ['hour', 'day', 'month', 'averagewindspeed', 'airtemperature', 'relativehumidity

for target in targets:
    impute_and_plot_knn(target, features, microclimate_sensors_data)
```

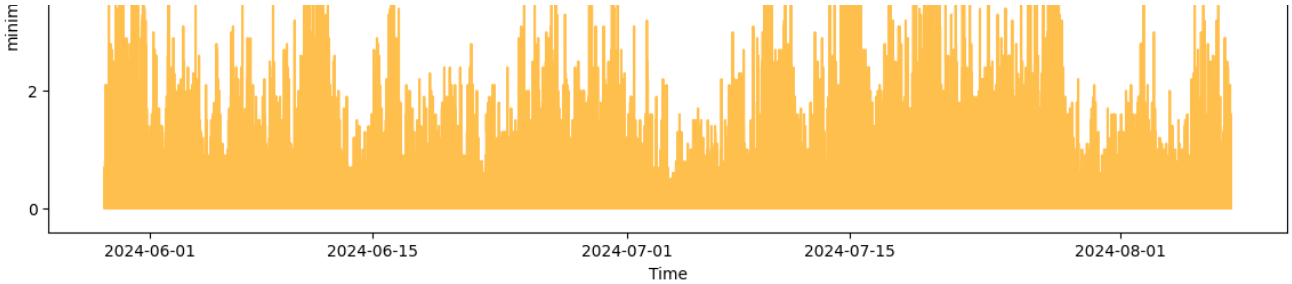



gustwinds
20
10
0
2024-06-01
2024-06-15
2024-07-01
2024-07-15
2024-08-01
Time


Time Series of gustwindspeed After Imputation
After Imputation
gustwindspeed
50
40
30
20
10
0
2024-06-01
2024-06-15
2024-07-01
2024-07-15
2024-08-01
Time


Time Series of minimumwindspeed Before Imputation
Before Imputation
minimumwindspeed
8
6
4
2
0
2024-06-01
2024-06-15
2024-07-01
2024-07-15
2024-08-01
Time

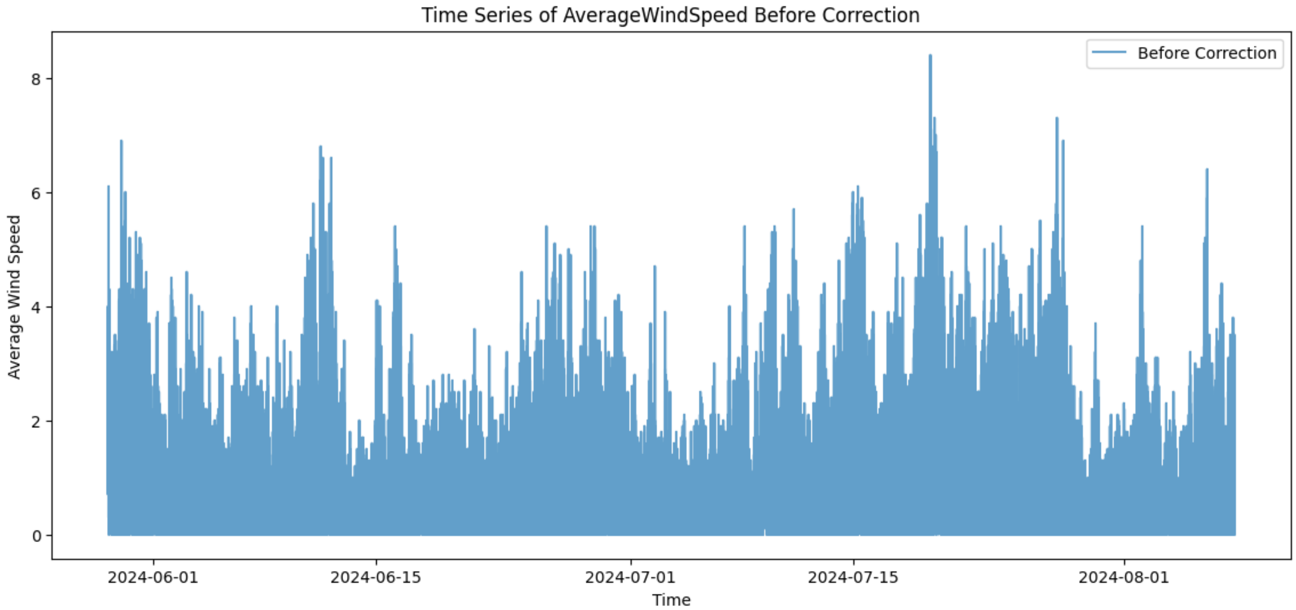

Time Series of minimumwindspeed After Imputation
After Imputation
umwindspeed
8
6
4

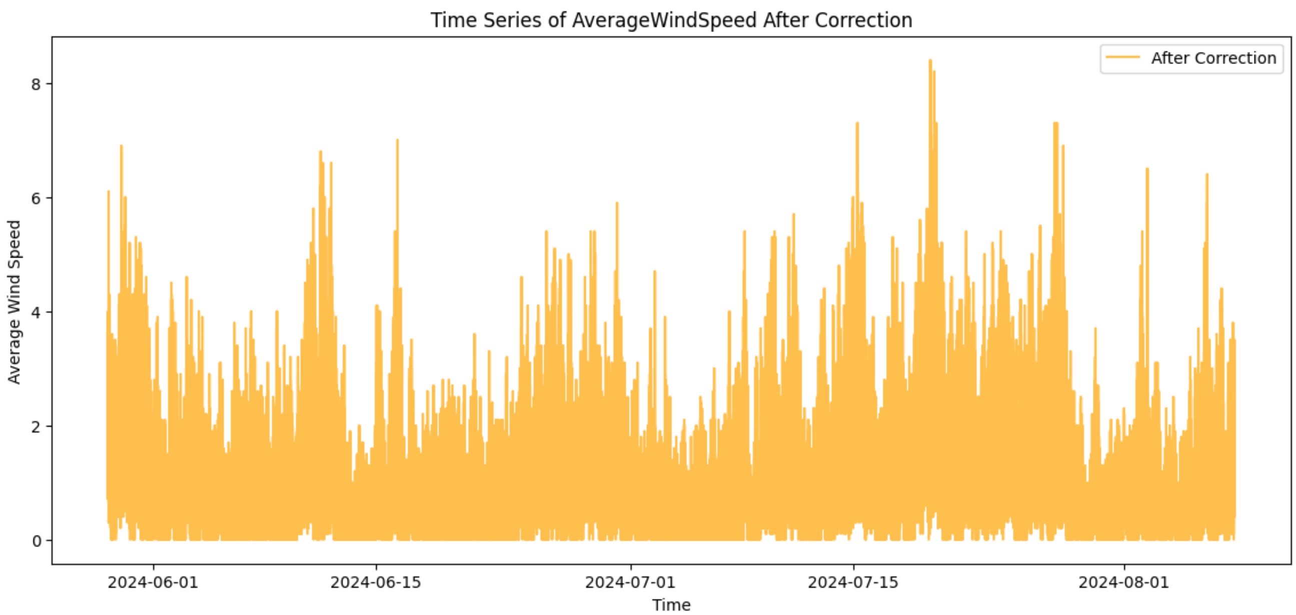

minim
2
0
2024-06-01
2024-06-15
2024-07-01
2024-07-15
2024-08-01
Time

I identified some of the average wind speed values are less then minimum wind speed. I'm using predictive data wrangling method (KNN) to fix that.

```
# Plot before correction
plt.figure(figsize=(14, 6))
plt.plot(microclimate_sensors_data['received_at'], microclimate_sensors_data['averagewind
plt.title('Time Series of AverageWindSpeed Before Correction')
plt.xlabel('Time')
plt.ylabel('Average Wind Speed')
plt.legend()
plt.show()

# Identify rows where AverageWindSpeed is less than MinimumWindSpeed
incorrect_avg_wind = microclimate_sensors_data[microclimate_sensors_data['averagewindspee

complete_data = microclimate_sensors_data.drop(incorrect_avg_wind.index)

# Define features and target
features = ['hour', 'day', 'month', 'minimumwindspeed', 'gustwindspeed', 'airtemperature'
X = complete_data[features]
y = complete_data['averagewindspeed']

imputer = SimpleImputer(strategy='mean')
X_imputed = imputer.fit_transform(X)

# Train the KNN regression model
knn_model = KNeighborsRegressor(n_neighbors=5)
knn_model.fit(X_imputed, y)

missing_X = incorrect_avg_wind[features]
missing_X_imputed = imputer.transform(missing_X)
predicted_values = knn_model.predict(missing_X_imputed)

# Ensure predictions are within bounds
predicted_values = np.clip(predicted_values, incorrect_avg_wind['minimumwindspeed'], inco

microclimate_sensors_data.loc[incorrect_avg_wind.index, 'averagewindspeed'] = predicted_v

# Plot after correction
plt.figure(figsize=(14, 6))
plt.plot(microclimate_sensors_data['received_at'], microclimate_sensors_data['averagewind
```

```
plt.title('Time Series of AverageWindSpeed After Correction')
plt.xlabel('Time')
plt.ylabel('Average Wind Speed')
plt.legend()
plt.show()
```





Check one more time for the missing values

```
print("Missing Values in Microclimate Sensor Readings:\n", microclimate_sensor_readings.i
print("Missing Values in Microclimate Sensors Data:\n", microclimate_sensors_data.isnull(
print("Missing Values in Meshed Sensor Type 1:\n", meshed_sensor_type_1.isnull().sum())
```

```
Missing Values in Microclimate Sensor Readings:
 local_time      0
id              0
site_id         0
sensor_id       0
value           0
type            0
units           0
gatewayhub_id   0
site_status     0
dtype: int64
Missing Values in Microclimate Sensors Data:
 device_id              0
received_at            0
sensorlocation         0
latlong                0
averagewinddirection   0
minimumwindspeed       0
averagewindspeed       0
gustwindspeed          0
airtemperature         0
relativehumidity       0
atmosphericpressure    0
hour                   0
day                    0
month                  0
dtype: int64
Missing Values in Meshed Sensor Type 1:
 dev_id               0
time                 0
rtc                  0
battery              0
```

```
solarpanel            0
command               0
solar                 0
precipitation         0
strikes               0
windspeed             0
winddirection         0
gustspeed             0
vapourpressure        0
atmosphericpressure   0
relativehumidity      0
airtemp               0
lat_long              0
sensor_name           0
dtype: int64
```

Removing outilers

```
def remove_outliers_zscore(df, columns, z_threshold=3.0):
  for column in columns:
    mean_col = df[column].mean()
    std_col = df[column].std()

    z_scores = (df[column] - mean_col) / std_col

    df = df[(z_scores > -z_threshold) & (z_scores < z_threshold)]
  return df

# Columns to check for outliers in microclimate data
columns_to_check_microclimate_sensors_data = [
    'airtemperature', 'relativehumidity', 'atmosphericpressure',
    'minimumwindspeed', 'averagewindspeed', 'gustwindspeed',
    'averagewinddirection'
]

# Columns to check for outliers in meshed sensor data
columns_to_check_meshed_sensor_type_1 = [
    'solar', 'windspeed', 'gustspeed',
    'vapourpressure', 'atmosphericpressure', 'relativehumidity',
    'airtemp'
]

# Remove outliers from the specified columns
wrangled_microclimate_sensors_data = remove_outliers_zscore(microclimate_sensors_data, colu
wrangled_meshed_sensor_type_1 = remove_outliers_zscore(meshed_sensor_type_1, columns_to_che

print("Number of rows after removing outliers from Microclimate data:", len(wrangled_microc
print("Number of rows after removing outliers from Meshed Sensor data:", len(wrangled_meshe
```

```
Number of rows after removing outliers from Microclimate data: 48774
```

```
Number of rows after removing outliers from Meshed Sensor data: 108926
```

NEXT: Data Merging