















### **UNTAR untuk INDONESIA**

# Analisis Data

## Lakukan langkah-langkag berikut:

- Masuk ke <a href="https://drive.google.com/drive/my-drive">https://drive.google.com/drive/my-drive</a>
- Buat Folder BMKG.
- Unggah file Anda ke folder BMKG.





```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette samples, silhouette score
import matplotlib.cm as cm
import matplotlib.pyplot as plt
from scipy.spatial import distance
from sklearn import preprocessing
import time
```





```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
path = '_/content/drive/MyDrive/BMKG/Data BMKG Ternate.xlsx'
df = pd.read_excel(path,sheet_name='Sheet1')
```

df.head()

	Date	Tn	Tx	Tavg	RH_avg	RR	SS	ff_x	ddd_x	ff_avg	ddd_car
0	2010-01-01	25.0	32.6	27.4	86.0	8888.0	7.4	6.0	315.0	2.0	NE
1	2010-01-02	24.0	32.8	27.5	84.0	2.0	8.0	5.0	315.0	3.0	NW
2	2010-01-03	24.0	32.0	27.0	85.0	0.0	6.5	7.0	360.0	4.0	NW
3	2010-01-04	23.0	32.8	28.0	80.0	8888.0	6.7	7.0	360.0	4.0	NW
4	2010-01-05	21.0	30.8	26.7	86.0	5.0	2.0	8.0	315.0	6.0	NW





df.tail()

	Date	Tn	Tx	Tavg	RH_avg	RR	SS	ff_x	ddd_x	ff_avg	ddd_car
4955	2023-07-27	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4956	2023-07-28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4957	2023-07-29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4958	2023-07-30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4959	2023-07-31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

df = df.iloc[0:4838,:]

df.tail()

	Date	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
4833	2023-03-27	25.4	31.0	27.4	86.0	26.6	0.6	7.0	300.0	3.0	W
4834	2023-03-28	24.6	30.4	27.1	88.0	51.8	2.7	7.0	320.0	3.0	NW
4835	2023-03-29	24.0	28.8	25.6	92.0	26.8	3.0	5.0	300.0	2.0	W
4836	2023-03-30	24.4	28.2	26.7	90.0	6.9	2.0	4.0	150.0	2.0	SW
4837	2023-03-31	24.8	31.0	27.3	87.0	15.9	8.0	6.0	260.0	3.0	W













```
df.columns
Index(['Date', 'Tn', 'Tx', 'Tavg', 'RH_avg', 'RR', 'ss', 'ff_x', 'ddd_x',
       'ff_avg', 'ddd_car'],
      dtype='object')
df.rename(columns = {'Tn': 'Minimum Temperature', 'Tx': 'Maximum Temperature', 'Tavg': 'Average Temperature', 'RH_avg': 'Humidity',
                      'RR': 'Rainfall' ,'ss': 'Sunshine', 'ff_x': 'Maximum Wind Speed', 'ddd_x': 'Wind Direction' ,'ff_avg': 'Average Wind Speed',
                      'ddd car':'Wind Mode'}, inplace = True)
```















#### df.head()

	Date	Minimum Temperature	Maximum Temperature	Average Temperature	Humidity	Rainfall	Sunshine	Maximum Wind Speed	Wind Direction	Average Wind Speed	Wind Mode
0	2010-01-01	25.0	32.6	27.4	86.0	8888.0	7.4	6.0	315.0	2.0	NE
1	2010-01-02	24.0	32.8	27.5	84.0	2.0	8.0	5.0	315.0	3.0	NW
2	2010-01-03	24.0	32.0	27.0	85.0	0.0	6.5	7.0	360.0	4.0	NW
3	2010-01-04	23.0	32.8	28.0	80.0	8888.0	6.7	7.0	360.0	4.0	NW
4	2010-01-05	21.0	30.8	26.7	86.0	5.0	2.0	8.0	315.0	6.0	NW















```
df['Minimum Temperature'] = df['Minimum Temperature'].replace([8888], None)
df['Maximum Temperature'] = df['Maximum Temperature'].replace([8888], None)
df['Average Temperature'] = df['Average Temperature'].replace([8888], None)
df['Humidity'] = df['Humidity'].replace([8888], None)
df['Rainfall'] = df['Rainfall'].replace([8888], None)
df['Sunshine'] = df['Sunshine'].replace([8888], None)
df['Maximum Wind Speed'] = df['Maximum Wind Speed'].replace([8888], None)
df['Wind Direction'] = df['Wind Direction'].replace([8888], None)
df['Average Wind Speed'] = df['Average Wind Speed'].replace([8888], None)
```





```
df['Minimum Temperature'][df['Minimum Temperature'] <= 14] = None
df['Maximum Temperature'][df['Maximum Temperature'] <= 14] = None
df['Average Temperature'][df['Average Temperature'] <= 14] = None</pre>
```

```
df['Minimum Temperature'][df['Minimum Temperature'] >= 45] = None
df['Maximum Temperature'][df['Maximum Temperature'] >= 45] = None
df['Average Temperature'][df['Average Temperature'] >= 45] = None
```





## df['Sunshine'][df['Sunshine'] >= 13] = None

```
df['Maximum Wind Speed'][df['Maximum Wind Speed'] >= 30] = None
df['Average Wind Speed'][df['Average Wind Speed'] >= 30] = None
```





#### df.describe()

	Minimum Temperature	Maximum Temperature	Average Temperature	Humidity	Sunshine	Maximum Wind Speed	Wind Direction	Average Wind Speed
count	4393.000000	4374.000000	4404.000000	4404.000000	4403.00000	4540.000000	4541.000000	4541.000000
mean	24.681516	31.232190	27.236149	83.513624	5.76968	5.549119	252.718344	2.338912
std	0.951753	1.160027	0.817979	4.967077	3.02550	1.897462	106.480358	0.883951
min	18.000000	23.200000	23.900000	53.000000	0.00000	2.000000	0.000000	0.000000
25%	24.000000	30.600000	26.700000	81.000000	3.50000	4.000000	150.000000	2.000000
50%	25.000000	31.400000	27.300000	84.000000	6.20000	5.000000	315.000000	2.000000
75%	25.000000	32.000000	27.800000	87.000000	8.00000	6.000000	340.000000	3.000000
max	28.000000	36.000000	31.800000	96.000000	12.20000	27.000000	360.000000	9.000000













### df.isna().sum()/len(df)\*100

Date	0.000000
Minimum Temperature	9.198016
Maximum Temperature	9.590740
Average Temperature	8.970649
Humidity	8.970649
Rainfall	21.723853
Sunshine	8.991319
Maximum Wind Speed	6.159570
Wind Direction	6.138900
Average Wind Speed	6.138900
Wind Mode	6.118231
dtype: float64	





```
df = df.ffill()
df = df.bfill()
```

df.isna().sum()/len(df)\*100

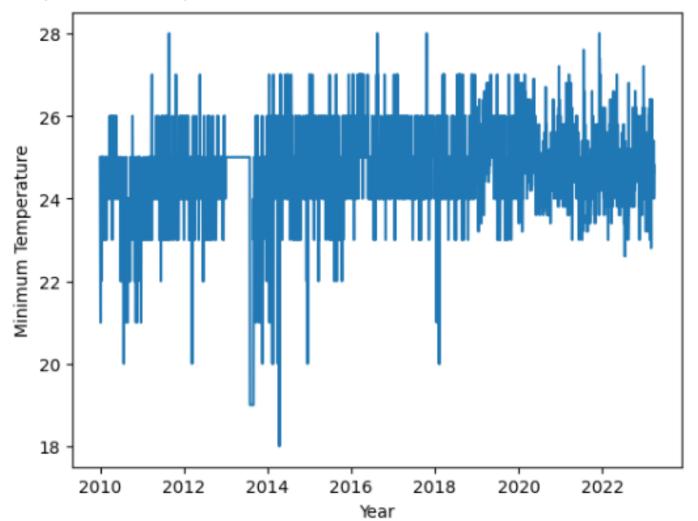
Date	0.0
Minimum Temperature	0.0
Maximum Temperature	0.0
Average Temperature	0.0
Humidity	0.0
Rainfall	0.0
Sunshine	0.0
Maximum Wind Speed	0.0
Wind Direction	0.0
Average Wind Speed	0.0
Wind Mode	0.0
dtype: float64	





```
plt.plot(df['Date'],df['Minimum Temperature'])
plt.ylabel('Minimum Temperature')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')











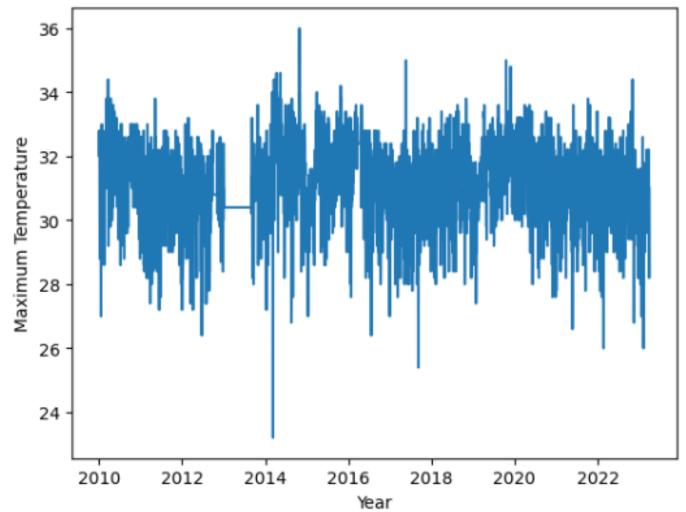






```
plt.plot(df['Date'],df['Maximum Temperature'])
plt.ylabel('Maximum Temperature')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')











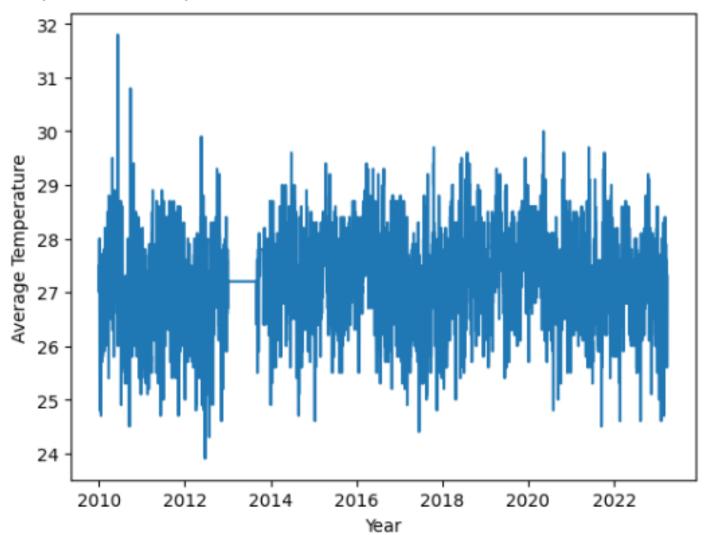






```
plt.plot(df['Date'],df['Average Temperature'])
plt.ylabel('Average Temperature')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')











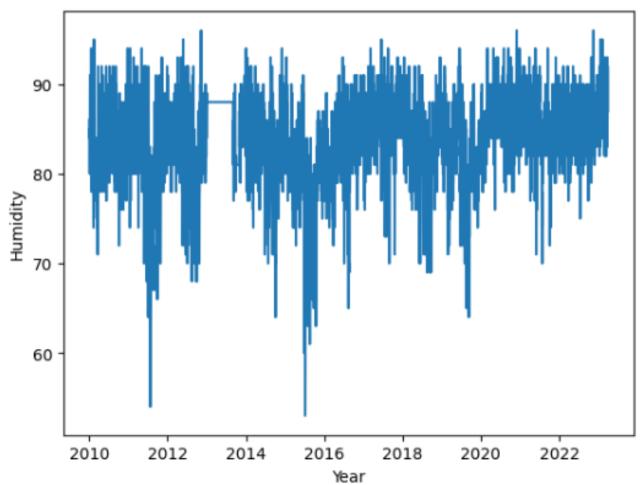






```
plt.plot(df['Date'],df['Humidity'])
plt.ylabel('Humidity')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')











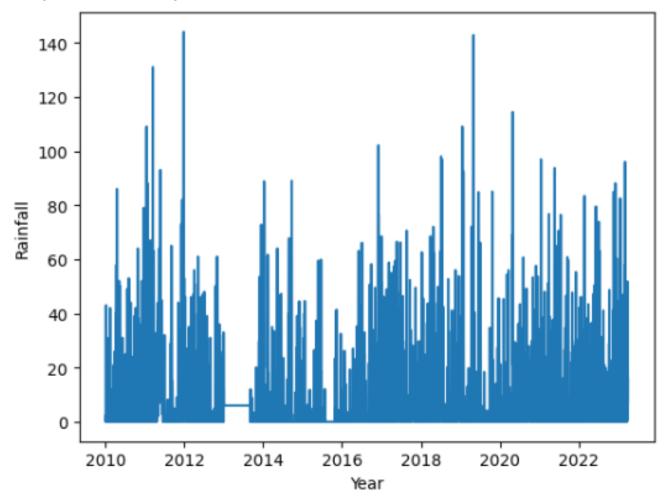






```
plt.plot(df['Date'],df['Rainfall'])
plt.ylabel('Rainfall')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')















```
plt.plot(df['Date'],df['Maximum Wind Speed'])
plt.ylabel('Maximum Wind Speed')
plt.xlabel('Year')
Text(0.5, 0, 'Year')
    25
    20 -
 Maximum Wind Speed
    15
    10
     5 ·
```

2016

Year

2014

2012

2010











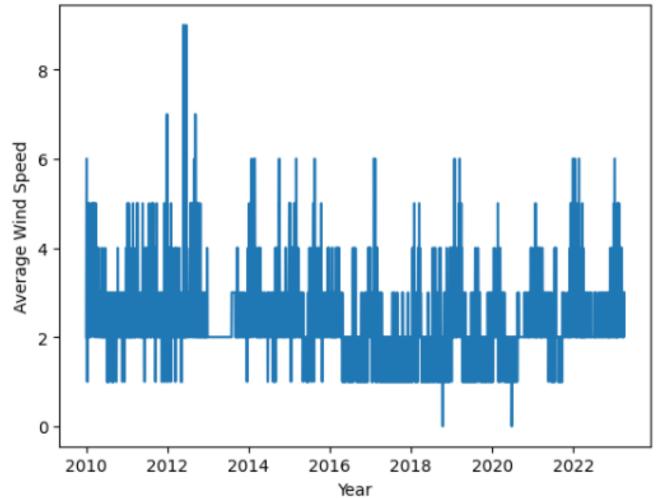
2020

2022

2018

```
plt.plot(df['Date'],df['Average Wind Speed'])
plt.ylabel('Average Wind Speed')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')





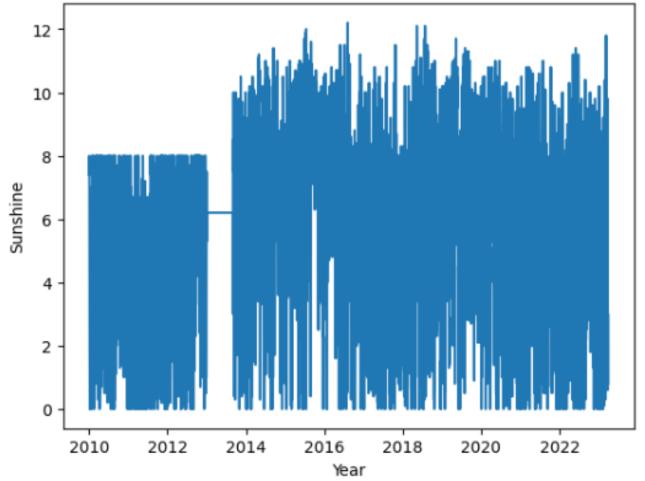






```
plt.plot(df['Date'],df['Sunshine'])
plt.ylabel('Sunshine')
plt.xlabel('Year')
```

Text(0.5, 0, 'Year')

















```
df['Month'] = df['Date'].dt.month
df['Year'] = df['Date'].dt.year
df['Day'] = df['Date'].dt.day_name()
```





#### df.tail()

Date	Minimum Temperature	Maximum Temperature	Average Temperature	Humidity	Rainfall	Sunshine	Maximum Wind Speed	Wind Direction	Average Wind Speed	Wind Mode	Month	Year	Day
<b>4833</b> 2023-03-27	25.4	31.0	27.4	86.0	26.6	0.6	7.0	300.0	3.0	W	3	2023	Monday
<b>4834</b> 2023-03-28	24.6	30.4	27.1	88.0	51.8	2.7	7.0	320.0	3.0	NW	3	2023	Tuesday
<b>4835</b> 2023-03-29	24.0	28.8	25.6	92.0	26.8	3.0	5.0	300.0	2.0	W	3	2023	Wednesday
<b>4836</b> 2023-03-30	24.4	28.2	26.7	90.0	6.9	2.0	4.0	150.0	2.0	SW	3	2023	Thursday
<b>4837</b> 2023-03-31	24.8	31.0	27.3	87.0	15.9	8.0	6.0	260.0	3.0	W	3	2023	Friday









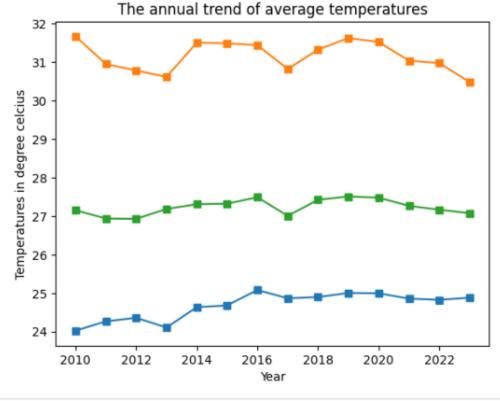








```
var = ['Minimum Temperature','Maximum Temperature','Average Temperature']
dfy1 = df.groupby(['Year']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Temperatures in degree celcius')
plt.title('The annual trend of average temperatures')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.25),ncol=3)
```







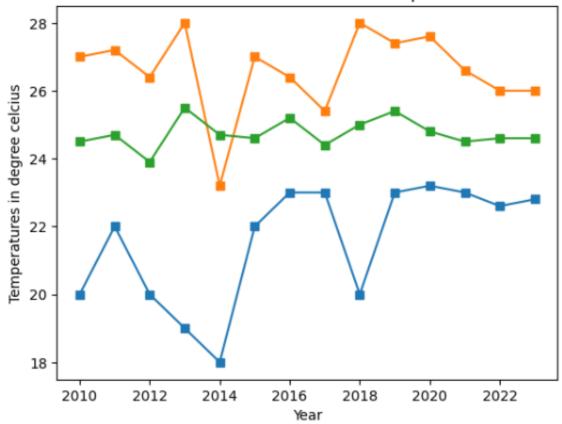




```
var = ['Minimum Temperature','Maximum Temperature','Average Temperature']
dfy1 = df.groupby(['Year']).min()[var]
dfy1.plot(marker = 's')
plt.ylabel('Temperatures in degree celcius')
plt.title('The annual trend of minimum temperatures')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.25),ncol=3)
```

<matplotlib.legend.Legend at 0x7f6ce2649990>

#### The annual trend of minimum temperatures





**IDONESIA** 







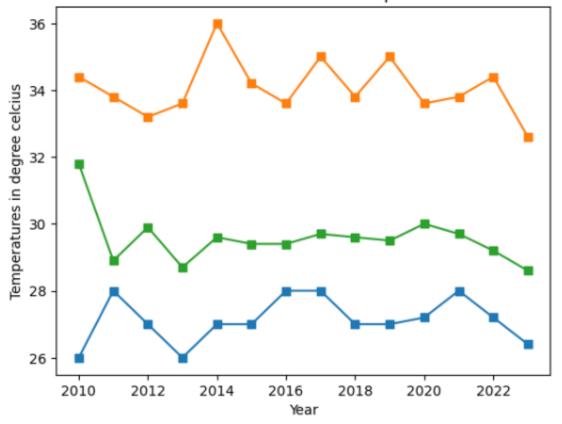


```
var = ['Minimum Temperature','Maximum Temperature','Average Temperature']
dfy1 = df.groupby(['Year']).max()[var]
dfy1.plot(marker = 's')
plt.ylabel('Temperatures in degree celcius')
plt.title('The annual of maximum temperatures')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.25),ncol=3)
```

<matplotlib.legend.Legend at 0x7f6ce22f3eb0>

Minimum Temperature

#### The annual of maximum temperatures













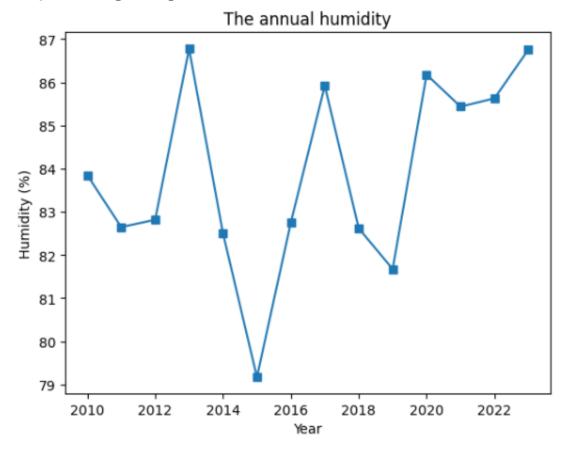


Maximum Temperature

Average Temperature

```
var = ['Humidity']
dfy1 = df.groupby(['Year']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Humidity (%)')
plt.title('The annual humidity')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.3),ncol=3)
```

<ipython-input-33-0965333cc0cc>:2: FutureWarning: The default value of numeric
 dfy1 = df.groupby(['Year']).mean()[var]
<matplotlib.legend.Legend at 0x7f6ce2442c20>













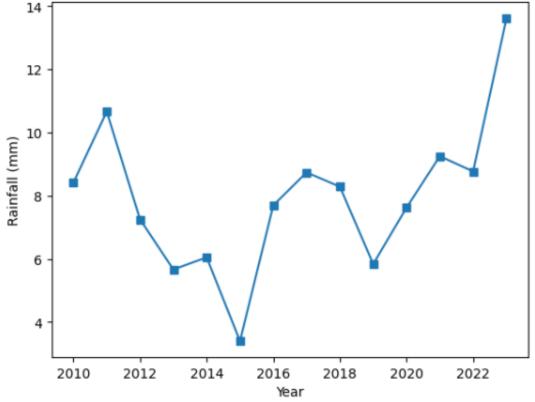




```
var = ['Rainfall']
dfy1 = df.groupby(['Year']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Rainfall (mm)')
plt.title('The annual rainfall')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.3),ncol=3)
```

<ipython-input-34-4b7110445749>:2: FutureWarning: The default value of num
dfy1 = df.groupby(['Year']).mean()[var]
<matplotlib.legend.Legend at 0x7f6ce26a71c0>

#### The annual rainfall











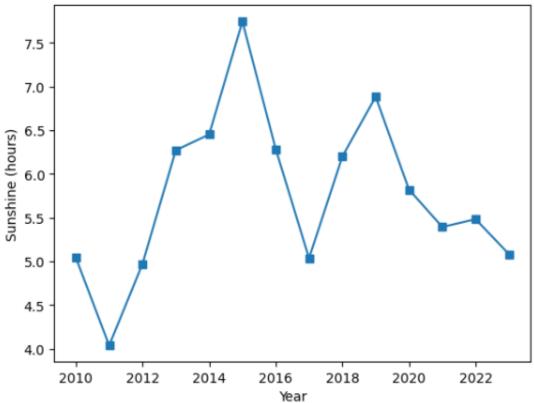




```
var = ['Sunshine']
dfy1 = df.groupby(['Year']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Sunshine (hours)')
plt.title('The annual sunshine duration')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.3),ncol=3)
```

<ipython-input-35-ed987b4ebaaf>:2: FutureWarning: The default value of num
 dfy1 = df.groupby(['Year']).mean()[var]
<matplotlib.legend.Legend at 0x7f6ce245fdc0>

#### The annual sunshine duration













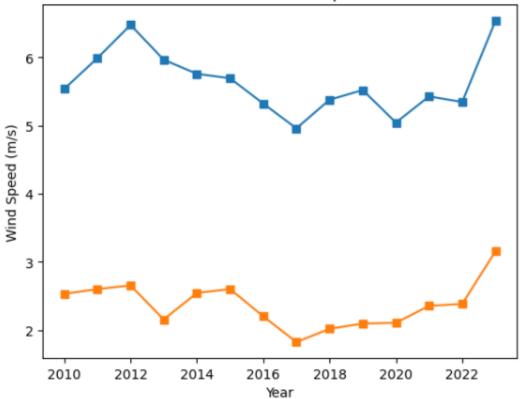




```
var = ['Maximum Wind Speed', 'Average Wind Speed']
dfy1 = df.groupby(['Year']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Wind Speed (m/s)')
plt.title('The annual wind speed')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.3),ncol=3)
```

<ipython-input-36-22ece1f8e126>:2: FutureWarning: The default value of nu
 dfy1 = df.groupby(['Year']).mean()[var]
<matplotlib.legend.Legend at 0x7f6ce20dff70>

#### The annual wind speed











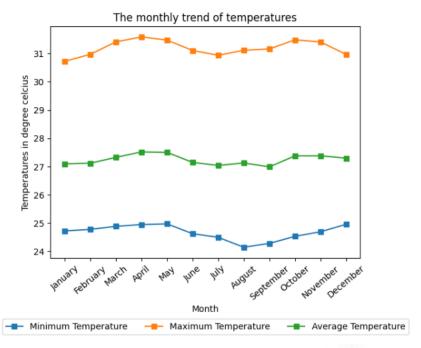








```
var = ['Minimum Temperature', 'Maximum Temperature', 'Average Temperature']
dfy1 = df.groupby(['Month']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Temperatures in degree celcius')
plt.title('The monthly trend of temperatures')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.35),ncol=3)
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12], ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'], rotation=40)
```







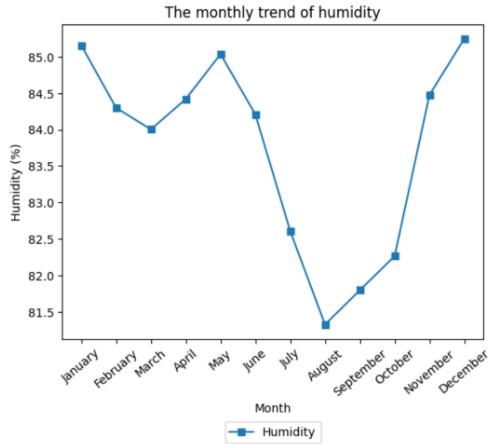








```
var = ['Humidity']
dfy1 = df.groupby(['Month']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Humidity (%)')
plt.title('The monthly trend of humidity')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.35),ncol=3)
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12], ['January', 'February', 'March', 'April', 'May', 'June',
                                            'July', 'August', 'September', 'October', 'November', 'December'], rotation=40)
```









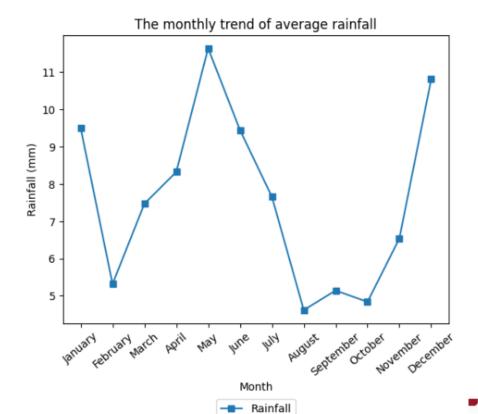








```
var = ['Rainfall']
dfy1 = df.groupby(['Month']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Rainfall (mm)')
plt.title('The monthly trend of average rainfall')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.35),ncol=3)
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12], ['January', 'February', 'March','April','May','June','July','August','September','October','November','December'], rotation=40)
```









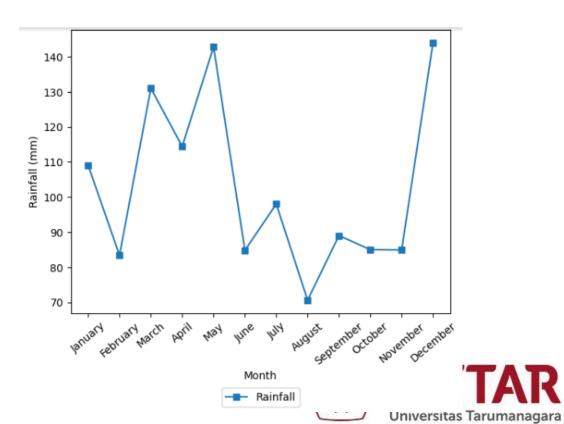






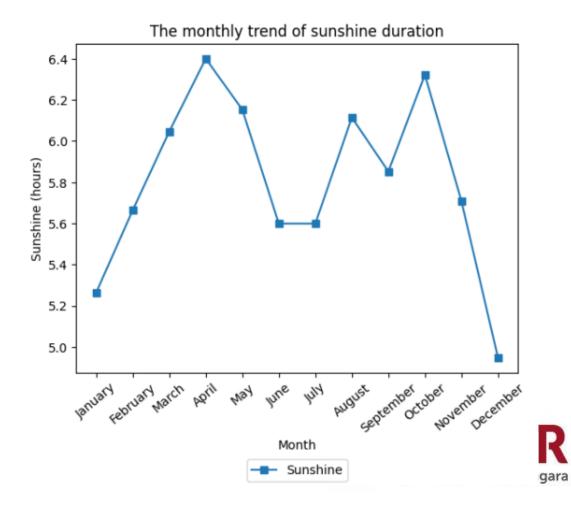


```
var = ['Rainfall']
dfy1 = df.groupby(['Month']).max()[var]
dfy1.plot(marker = 's')
plt.ylabel('Rainfall (mm)')
plt.title('The monthly trend of maximum rainfall')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.35),ncol=3)
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12], ['January', 'February', 'March','April','May','June','July','August','September','October','November','December'], rotation=40)
```



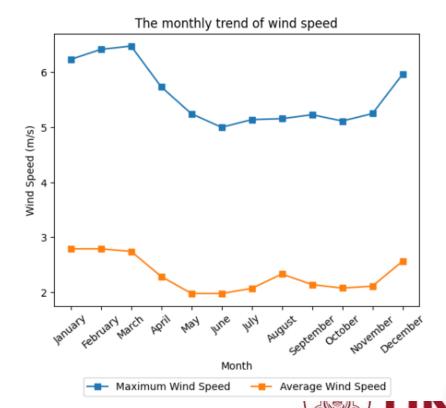


```
var = ['Sunshine']
dfy1 = df.groupby(['Month']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Sunshine (hours)')
plt.title('The monthly trend of sunshine duration')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.35),ncol=3)
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12], ['January', 'February', 'March','April','May','June','July','August','September','October','November','December'], rotation=40)
```





```
var = ['Maximum Wind Speed', 'Average Wind Speed']
dfy1 = df.groupby(['Month']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Wind Speed (m/s)')
plt.title('The monthly trend of wind speed')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.35),ncol=3)
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12], ['January', 'February', 'March','April','May','June','July','August','September','October','November','December'], rotation=40)
```





Universitas Tarumanagara







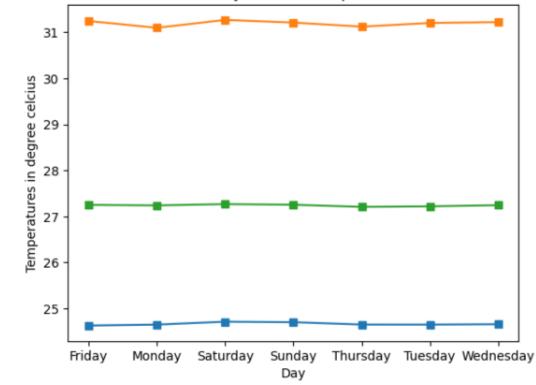




```
var = ['Minimum Temperature','Maximum Temperature','Average Temperature']
dfy1 = df.groupby(['Day']).mean()[var]
dfy1.plot(marker = 's')
plt.ylabel('Temperatures in degree celcius')
plt.title('The daily trend of temperatures')
plt.legend(loc='lower center', bbox_to_anchor=(0.5, -0.25),ncol=3)
```

<ipython-input-43-ad93f8611161>:2: FutureWarning: The default value of numeric\_only in
 dfy1 = df.groupby(['Day']).mean()[var]
<matplotlib.legend.Legend at 0x7f6ce23bbbe0>

#### The daily trend of temperatures











```
df2 = df.drop(['Date','Day','Month','Year','Wind Mode','Wind Direction'], axis = 1)
df2.head()
```

	Minimum Temperature	Maximum Temperature	Average Temperature	Humidity	Rainfall	Sunshine	Maximum Wind Speed	Average Wind Speed
0	25.0	32.6	27.4	86.0	2.0	7.4	6.0	2.0
1	24.0	32.8	27.5	84.0	2.0	8.0	5.0	3.0
2	24.0	32.0	27.0	85.0	0.0	6.5	7.0	4.0
3	23.0	32.8	28.0	80.0	0.0	6.7	7.0	4.0
4	21.0	30.8	26.7	86.0	5.0	2.0	8.0	6.0















c = df2.corr()

	Minimum Temperature	Maximum Temperature	Average Temperature	Humidity	Rainfall	Sunshine	Maximum Wind Speed	Average Wind Speed
Minimum Temperature	1.000000	0.161591	0.437941	-0.125919	-0.151194	0.154081	0.130903	0.051249
Maximum Temperature	0.161591	1.000000	0.611433	-0.444706	-0.266628	0.399245	-0.019838	-0.012138
Average Temperature	0.437941	0.611433	1.000000	-0.558051	-0.310120	0.362798	0.047026	0.108333
Humidity	-0.125919	-0.444706	-0.558051	1.000000	0.291074	-0.319253	-0.100928	-0.237841
Rainfall	-0.151194	-0.266628	-0.310120	0.291074	1.000000	-0.278796	0.034241	-0.003079
Sunshine	0.154081	0.399245	0.362798	-0.319253	-0.278796	1.000000	0.001135	0.008192
Maximum Wind Speed	0.130903	-0.019838	0.047026	-0.100928	0.034241	0.001135	1.000000	0.567636
Average Wind Speed	0.051249	-0.012138	0.108333	-0.237841	-0.003079	0.008192	0.567636	1.000000









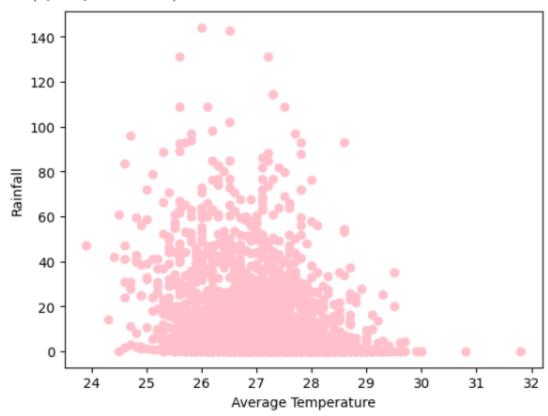






```
plt.scatter(df['Average Temperature'], df['Rainfall'], color = 'pink')
plt.xlabel('Average Temperature')
plt.ylabel('Rainfall')
```

Text(0, 0.5, 'Rainfall')











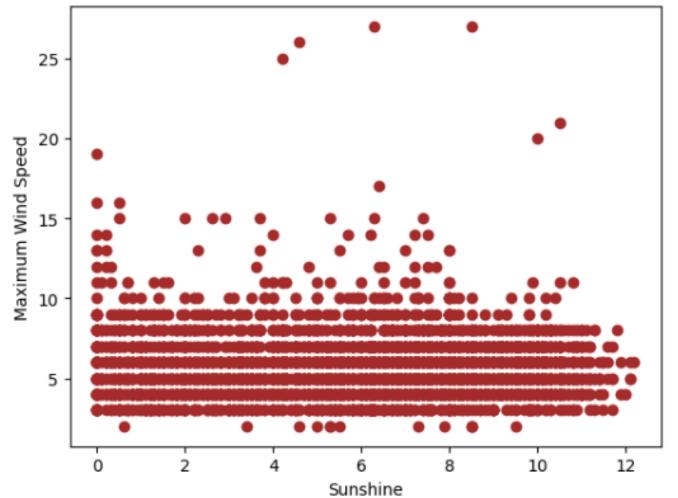






```
plt.scatter(df['Sunshine'], df['Maximum Wind Speed'], color = 'brown')
plt.xlabel('Sunshine')
plt.ylabel('Maximum Wind Speed')
```

Text(0, 0.5, 'Maximum Wind Speed')







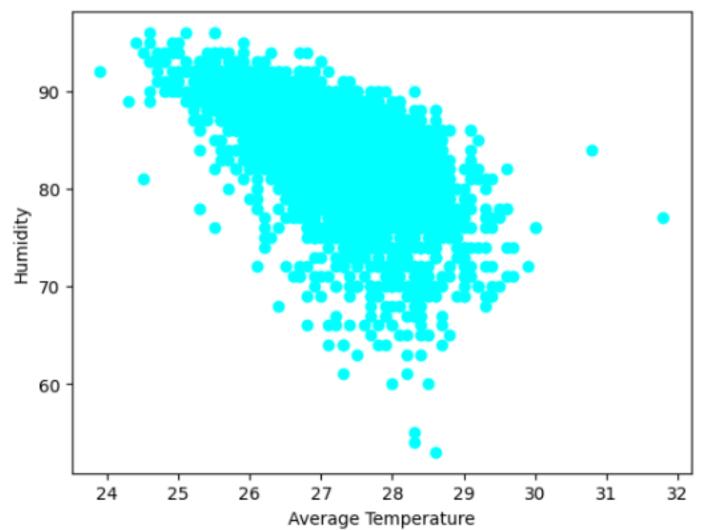






```
plt.scatter(df['Average Temperature'], df['Humidity'], color = 'cyan')
plt.xlabel('Average Temperature')
plt.ylabel('Humidity')
```

Text(0, 0.5, 'Humidity')







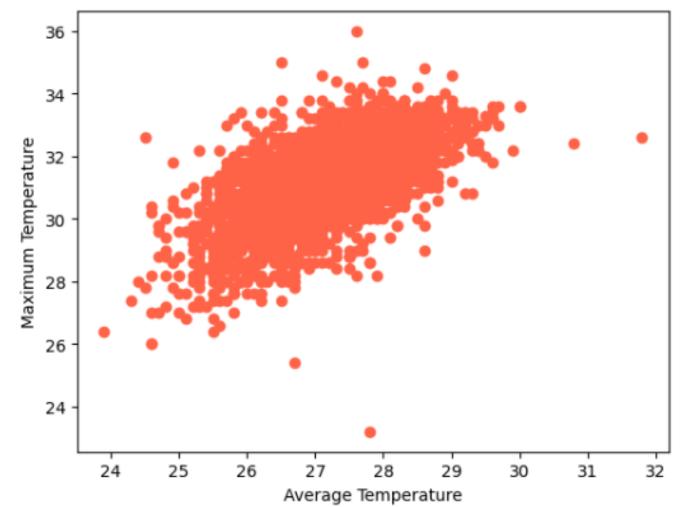






```
plt.scatter(df['Average Temperature'], df['Maximum Temperature'], color = 'tomato')
plt.xlabel('Average Temperature')
plt.ylabel('Maximum Temperature')
```

Text(0, 0.5, 'Maximum Temperature')











```
plt.scatter(df['Sunshine'], df['Maximum Temperature'], color = 'red', marker = '*')
plt.xlabel('Sunshine')
plt.ylabel('Maximum Temperature')
```

Text(0, 0.5, 'Maximum Temperature')

