STAT40800 Midterm Assignment - K.SAKETH SAI NIGAM - 22201204

In [1]:
```
# Load in necessary packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

1. Load in the weather.csv dataset into Python as a pandas DataFrame. Describe the data. How many years of recordings are included? What is the temporal resolution of the data? Which weather measurements are reported?

->loading the ***weatherdataset*** (with skipping 11 rows)

In [2]:
```
weatherdataset = pd.read_csv('weather.csv', skiprows=11)
```

-> Describing the 'weatherdataset'

In [3]:
```
weatherdataset.head()
```

Out[3]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun
0	1	jan	2021	Dublin Airport	5.0	-1.9	0.0	8.8	19.0	6.6
1	2	jan	2021	Dublin Airport	3.5	-2.4	0.1	10.1	19.0	2.4
2	3	jan	2021	Dublin Airport	3.9	-2.5	4.1	9.2	30.0	0.6
3	4	jan	2021	Dublin Airport	5.3	2.2	0.5	11.5	27.0	0.6
4	5	jan	2021	Dublin Airport	5.2	1.2	1.0	9.5	25.0	1.0

-> to check how many years of recording is present in the 'weatherdataset'

In [4]:
```
reqyear = weatherdataset.year.unique()
print("The year of recording is present in the 'weatherdataset': ",
```

```
The year of recording is present in the 'weatherdataset':  [2021]
```

Consequently, there is only one year of data in the 'weatherdataset', which is "2021."

-> Temporial Resolution of 'weatherdataset'

The temporal resolution is the time frame necessary to return to the same location and gather data. It depends on the time unit, as evidenced by this. Since the dataset "**sun: Sunshine duration**" returns values in hours, this is the case. As a result, various airports temporal resolutions in 2021 are provided by the category of 'sun' in the offered 'weatherdataset': ***SUNSHINE DURATION***

-> Weather Measurements in 'weatherdataset'

The weather Measurements in weatherdataset are ***THE MAXIMUM AIR TEMPETATURE(C), THE MINIMUM AIR TEMPETATURE(C), THE AMOUNT OF PRECIPITATION(mm), THE MEAN WIND SPEED(knot), THE HIGHEST GUST(knot)*** *are reported from the 'weatherdataset'*

2.Determine how many missing values there are in each column of the dataset. Can you think of a reason why these values are missing? Discuss different strategies for filling the missing values, highlighting the advantages and disadvantages of each strategy, in the context of this dataset.

Note: You do not need to implement any of your suggested strategies.

->determining missing values are there in each column of the 'weatherdataset'

In [5]:
```
weatherdataset.isnull().sum()
```

Out[5]:
```
day        0
month      0
year       0
station    0
maxtp      2
mintp      2
rain       4
wdsp       1
hg         3
sun        2
dtype: int64
```

From above the missing values are ***THE MAXIMUM AIR TEMPETATURE(C), THE MINIMUM AIR TEMPETATURE(C), THE AMOUNT OF PRECIPITATION(mm), THE MEAN WIND SPEED(knot), THE HIGHEST GUST(knot), SUNSHINE DURATION(hours)*** *are from the 'weatherdataset'*

->Reason of Data Missing:

1. When you don't have data stored for specific variables or participants
2. improper data entry

3. system malfunctions
4. lost files

->Strategies for filling the missing values:

1. Rows with missing values are deleted.
2. Put in a constant value in its stead.
3. Use the median or mean in its place.
4. Utilizing missing value-supporting algorithms.

->The advantages and disadvantages of each strategy:

1. Rows with missing values are deleted

Advatages:

1. A robust and more accurate model is produced when all missing value data is completely removed.
2. It is a good idea to remove a specific row or column with no specific information because it does not have any significance. Disadvatages:
3. The dataset might end up with some useful information removed.

2. Put in a constant value in its stead

Advantage:

1. Quicker and convenient way to access the entire dataset

Disadvantage:

1. observant of anomalies
2. Modification of the initial variance
3. influences correlation

3. Use the median or mean in its place

Advantage:

1. A rapid, easy, and approximate imputation method for continuous variables that substitutes the mean or median for any missing values.
2. facilitates straightforward analysis and understanding of data.

Disadvantage:

1. misinterpretation of the initial heterogeneity

4. Utilizing missing value-supporting algorithms

Advantage:

1. Has a high level of performance.

2. Outliers don't matter as much.

Disadvantage:

1. Processing takes a very long time for larger datasets. Poor performance, perhaps.

3. Write code to answer the following questions: (15 marks)
   A. How many different weather stations are included in the data set?
   B. At what station and on what date was the lowest minimum air temperature recorded?
   C. At what station and on what date was the largest amount of rain recorded?

->A. To determine different weather stations are included in 'weatherdataset'

In [6]:
```
uniqueweatherstation = weatherdataset.station.unique()
print("The Different Weather Stations are: ",uniqueweatherstation)
```

```
The Different Weather Stations are:  ['Dublin Airport' 'Shannon Ai
rport' 'Cork Airport']
```

->B. At what station and on what date was the lowest minimum air temperature recorded?

In [7]:
```
weatherdataset.loc[weatherdataset['mintp'] == min(weatherdataset['m
```

Out[7]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun
8	9	jan	2021	Dublin Airport	2.4	-5.9	0.0	6.5	19.0	6.8

At Dublin Airport, the lowest minimum air temperature was recorded on January 9, 2021 in degrees celsius.

-> C. At what station and on what date was the largest amount of rain recorded?

In [8]:
```
weatherdataset.loc[weatherdataset['rain'] == max(weatherdataset['ra
```

Out[8]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun
1029	27	oct	2021	Cork Airport	14.7	12.4	50.1	16.4	36.0	0.0

At Cork Airport, the most rainfall was observed on October 27, 2021.

4. Create a numerical summary (mean, standard deviation, minimum, maximum, etc.) for each of the weather measurements. Discuss and interpret your results.

-> Summary of the weatherdataset{(mean, standard deviation, minimum, maximum, etc.} for each of the weather measurements

In [25]:
```
weatherdataset.describe()
```

Out[25]:

	day	year	maxtp	mintp	rain	wdsp	
count	1095.000000	1095.0	1093.000000	1093.000000	1091.000000	1094.000000	1092.0000
mean	15.720548	2021.0	13.732662	6.950046	2.596242	8.587020	22.8635
std	8.800266	0.0	5.107554	4.687423	4.864790	3.705469	8.3967
min	1.000000	2021.0	0.600000	-5.900000	0.000000	1.900000	7.0000
25%	8.000000	2021.0	10.200000	3.400000	0.000000	6.000000	17.0000
50%	16.000000	2021.0	13.400000	7.100000	0.300000	7.900000	21.0000
75%	23.000000	2021.0	17.600000	10.700000	3.050000	10.600000	28.0000
max	31.000000	2021.0	29.600000	18.700000	50.100000	26.300000	64.0000

Interpreting the results of weather measurements:-

MAXTP:THE MAXIMUM AIR TEMPETATURE(C)

There are 2 missing values out of the 1093 values in maxtp. The maxtp's mean temperature is roughly 13.7 degrees Celsius (C), and the data's standard deviation is 5. The range of the highest temperature is 29 C since the minimum and maximum values of the maximum temperature are 0.6 C and 29.6 C, respectively. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%. MaxTP's 25%, 50%, and 75% values are 10.2, 13.4, and 17.6, respectively.

MINTP:THE MINIMUM AIR TEMPETATURE(C)

Additionally, mintp contains 1093 values, which shows that there are 2 missing values. The data spread is roughly 4.68 C, while the mean of the mintp is approximately 6.95 C. -5.9 C and 18.7 C are the minimum and highest values of mintp, respectively. Consequently, the range of mintp is 24.6 C. The percentiles (25%, 50%, and 75%) are also included in the summary. The bottom quartile, or 25% of the dataset, is made up of

the first 25%. The center number, 50%, is referred to as the median, while the top quartiles, which make up 75% of the data, are referred to as the upper quartiles. The respective 25%, 50%, and 75% for mintp are 3.4 C, 7.1 C, and 10.7 C.

RAIN:THE AMOUNT OF PRECIPITATION(mm)

There are 4 missing values in rain, which has 1091 values. These missing values must be filled either with test statistics or arbitrarily chosen values. The spread of the data is roughly 4.86 mm, while the mean of the rain is 2.59 mm. Rainfall ranges from 0 millimeters to 50 millimeters, respectively. This also implies that there will be 50mm of rain. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%.It rains 0 mm, 0.3 mm, and 3.05 mm, respectively, for the 25%, 50%, and 75%.

WDSP:THE MEAN WIND SPEED(knot)

There are 1 missing values according to the 1094 values in wdsp. These missing values must be filled either with test statistics or arbitrarily chosen values. The data spread is roughly 3.7 knots, while the wdsp mean is 8.58 knots. The value of mintp ranges from 1.9 knot to 26.3 knot, accordingly. Additionally, this indicates that the wdsp's range is 24.4 knot. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%.For wdsp, the respective 25%, 50%, and 75% are 6, 7.9, and 10.6 knot.

HG:THE HIGHEST GUST(knot)

The fact that hg has 1092 values means there are three missing values, which should either be discarded or filled in with any arbitrary number or test statistic. The data spread is roughly 8.39 knots, while the mean of the hg is approximately 22.86 knots. The range of hg is 57 knot since the minimum and highest values of mintp are 7 knot and 64 knot, respectively. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%.For hg, the respective 25%, 50%, and 75% are 17, 21, and 28 knot.

SUN:SUNSHINE DURATION(hours)

Sun contains 1093 values, thus there are 2 missing values that should either be discarded or replaced in with any arbitrary number or test statistic. The spread of the data is approximately 3.95 hours, and the average amount of sunshine each day is 4 hours. The range of sunshine is 15 hours because there are 15 hours between the minimum and highest sunshine duration of 0 hours. The percentiles (25%, 50%, and 75%) are also included in the summary. The bottom quartile, or 25% of the dataset, is

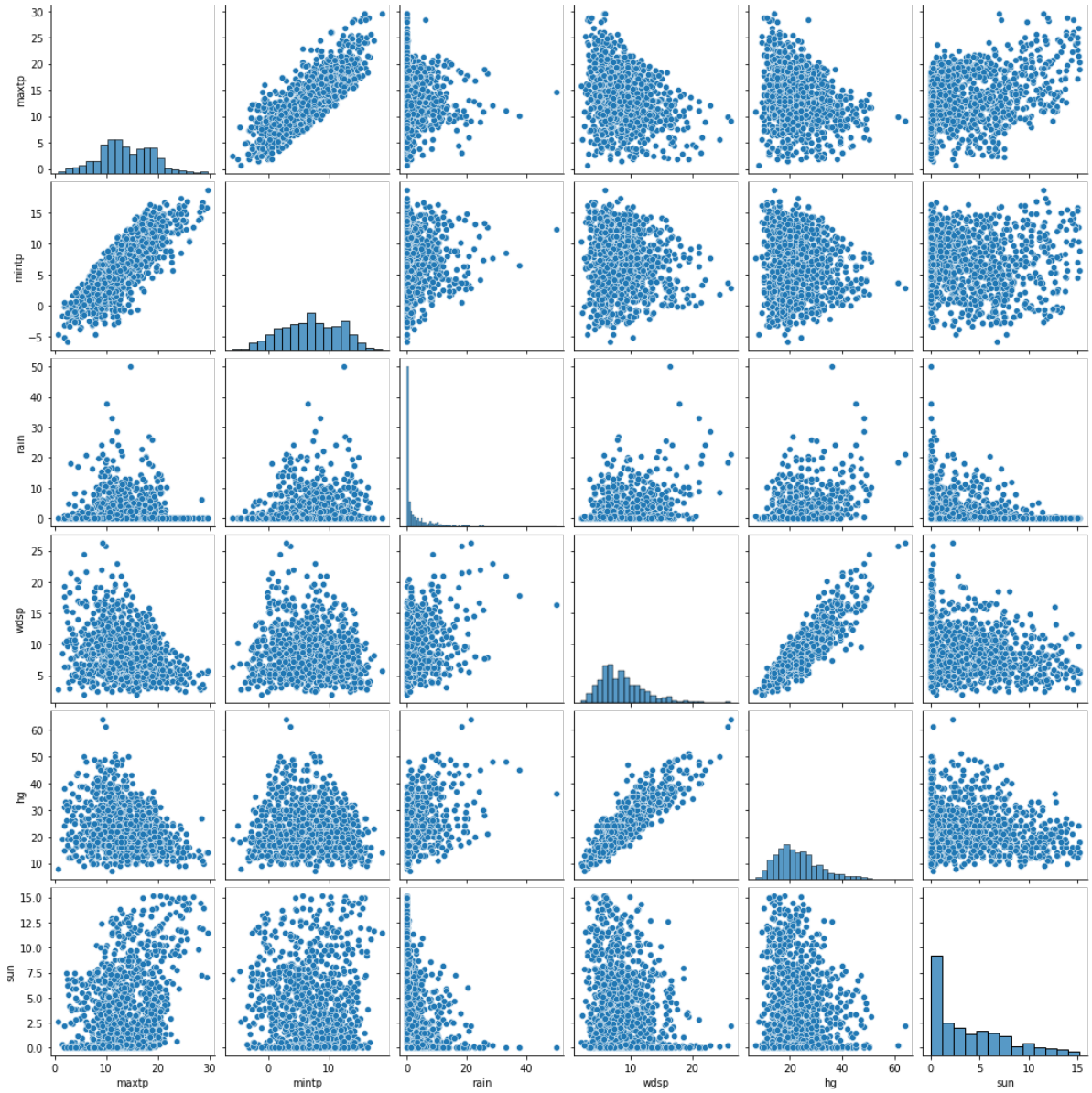
made up of the first 25%. The center number, 50%, is referred to as the median, while the top quartiles, which make up 75% of the data, are referred to as the upper quartiles.For sun, the corresponding percentages are 25, 50, and 75 are 0.3, 2.9, and 6.6.

5. Create a graphical summary for each of the weather measurements. Discuss your plots in relation to the summary statistics found in question 4.

-> Graphical Summary for each of the weather measurements

In [12]:
```
graphicalsummaryofweather = weatherdataset
graphicalsummaryofweather = graphicalsummaryofweather.drop(['day','
sns.pairplot(graphicalsummaryofweather)
```

Out[12]:
```
<seaborn.axisgrid.PairGrid at 0x7fe084a5e9d0>
```



Discussing about plots:-

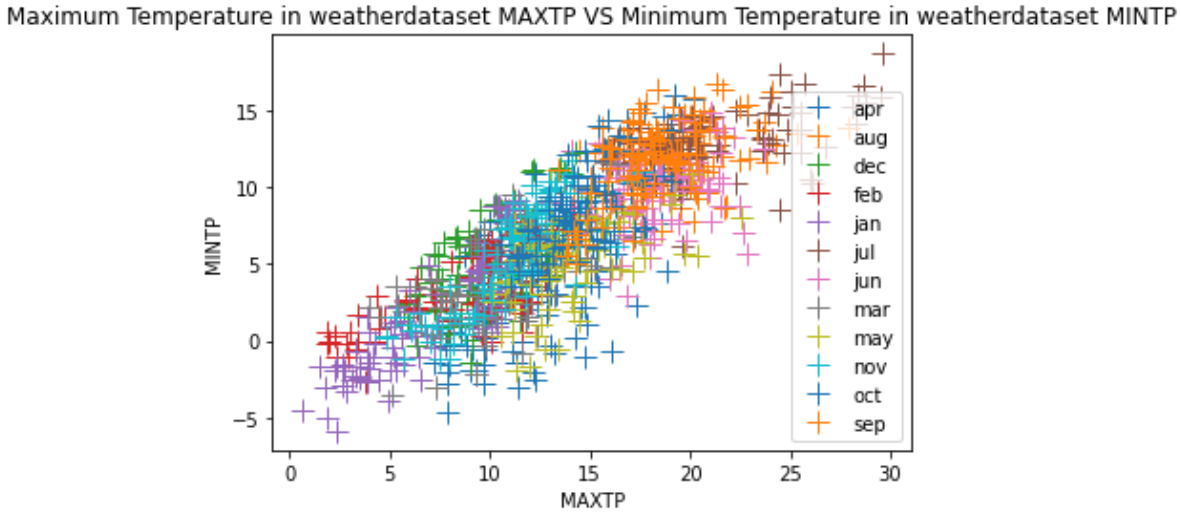
In a dataset, pairwise relationships are plotted using a pairplot. Each variable in the data will be shared in a single row and column on the y-axis and a single column on the x-axis thanks to the pairplot function's creation of a grid of axes. Thus, plots are produced, as above. It makes use of a set of weather data. Six meteorological measurements are included in the data set: the maximum and minimum air temperatures, the amount of precipitation, the mean wind speed, the strongest gust, and the number of hours of sunshine. The grid above displays a map of the data. There are 6 measurements, hence a 6x6 plot is produced. We may use pairplot to examine the non-diagonal linear relationship and diagonal distribution of these measurements. We receive the results as a 6X6 shape based on the various values of each measurement.

6. Produce a scatter plot of the maximum temperature versus the minimum temperature and colour your points based on month. Interpret your plot.

In [13]:
```
monthsofweatherdataset = weatherdataset.groupby('month')
for nameofweatherdataset, monthofweatherdataset in monthsofweatherd
    plt.plot(monthofweatherdataset.maxtp, monthofweatherdataset.min
             label=nameofweatherdataset)

plt.xlabel("MAXTP")
plt.ylabel("MINTP")
plt.title("Maximum Temperature in weatherdataset MAXTP VS Minimum T
plt.legend()
```

Out[13]:
```
<matplotlib.legend.Legend at 0x7fe084da6b80>
```



Interpreting the plot:-

Here, the relation between the highest and lowest temperatures is being studied. The stations, months, years, maximum air temperature, minimum air temperature, precipitation amount, mean wind speed, highest gust, and sunshine duration make up the meteorological/weather data collection. The minimum and highest temperatures are displayed on the vertical and horizontal axes, respectively. The locations of each "+" on the scatterplot are determined by the maximum and minimum temperatures for the 12 different months from January to December.Because the minimum temperature in various months rises along with the maximum temperature, these two elements are positively correlated. This case is the best illustration of a linear relationship. The relationship is said to be linear if one variable increases roughly at the same pace as the other variables change by one unit.

7. Compute the daily temperature range, and add this as an additional variable to your DataFrame. Print out the first few rows of your DataFrame to show that the column has been added correctly.

->Determining the daily temperature range, adding this to weatherdataset and showing the first five rows

In [14]:
```
weatherdataset['DailyTemperatureRange'] = weatherdataset['maxtp'] -
weatherdataset.head()
```

Out[14]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	DailyTemperatureRange
0	1	jan	2021	Dublin Airport	5.0	-1.9	0.0	8.8	19.0	6.6	6.9
1	2	jan	2021	Dublin Airport	3.5	-2.4	0.1	10.1	19.0	2.4	5.9
2	3	jan	2021	Dublin Airport	3.9	-2.5	4.1	9.2	30.0	0.6	6.4
3	4	jan	2021	Dublin Airport	5.3	2.2	0.5	11.5	27.0	0.6	3.1
4	5	jan	2021	Dublin Airport	5.2	1.2	1.0	9.5	25.0	1.0	4.0

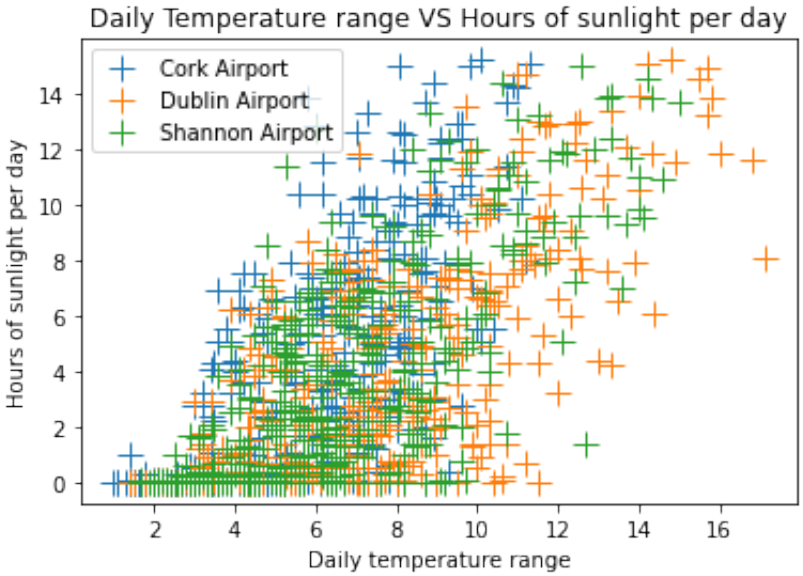
8. Plot the daily temperature range versus the hours of sunlight per day, colouring the points based on the weather station. Interpret your plot.

-> Based on a weather station, a plotting of the daily temperature range against the number of hours of sunlight per day.

In [15]:
```
weatherstations = weatherdataset.groupby('station')
for weathername, weatherstation in weatherstations:
    plt.plot(weatherstation.DailyTemperatureRange, weatherstation.s
             label=weathername)

plt.xlabel("Daily temperature range")
plt.ylabel("Hours of sunlight per day")
plt.title("Daily Temperature range VS Hours of sunlight per day")
plt.legend()
```

Out[15]:
```
<matplotlib.legend.Legend at 0x7fe087350c40>
```



Interpreting the plot:-

Here, the association between daily sunshine hours and temperature range is being researched. The weather data collection comprises STATIONS, MONTHS, YEARS, MAXIMUM AIR TEMPETATURE(C), MINIMUM AIR TEMPETATURE(C), AMOUNT OF PRECIPITATION(mm), MEAN WIND SPEED(knot), HIGHEST GUST(knot), and SUNSHINE DURATION (hours). The vertical axis shows the number of hours of sunlight per day, while the horizontal axis shows the daily temperature range. The placement of each "+" on the scatterplot depends on the daily temperature range at 3 independent locations (Dublin, Cork, and Shannon) Airports as well as the number of hours of sunlight each day. These two factors are positively interconnected because the number of hours of sunlight per day grows as the range of daily temperatures does. This particular instance best exemplifies a linear relationship. If one variable increases roughly at the same rate as the other variables change by one unit, the relationship is said to be linear.

9. Perform a comparative analysis of the weather at Dublin Airport, Shannon Airport and Cork Airport.

For full marks on this question you should create numerical and graphical summaries of the weather measurements at each weather station and discuss how the weather differs (or is similar) across these locations.

Dublin Airport:-

In [16]:
```
weatherofdublinairport = weatherdataset[weatherdataset['station']==
weatherofdublinairport.head()
```

Out[16]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	DailyTemperatureRange
0	1	jan	2021	Dublin Airport	5.0	-1.9	0.0	8.8	19.0	6.6	6.9
1	2	jan	2021	Dublin Airport	3.5	-2.4	0.1	10.1	19.0	2.4	5.9
2	3	jan	2021	Dublin Airport	3.9	-2.5	4.1	9.2	30.0	0.6	6.4
3	4	jan	2021	Dublin Airport	5.3	2.2	0.5	11.5	27.0	0.6	3.1
4	5	jan	2021	Dublin Airport	5.2	1.2	1.0	9.5	25.0	1.0	4.0

Numerical Summary of Dublin Airport:-

In [17]:
```
weatherofdublinairport.describe()
```

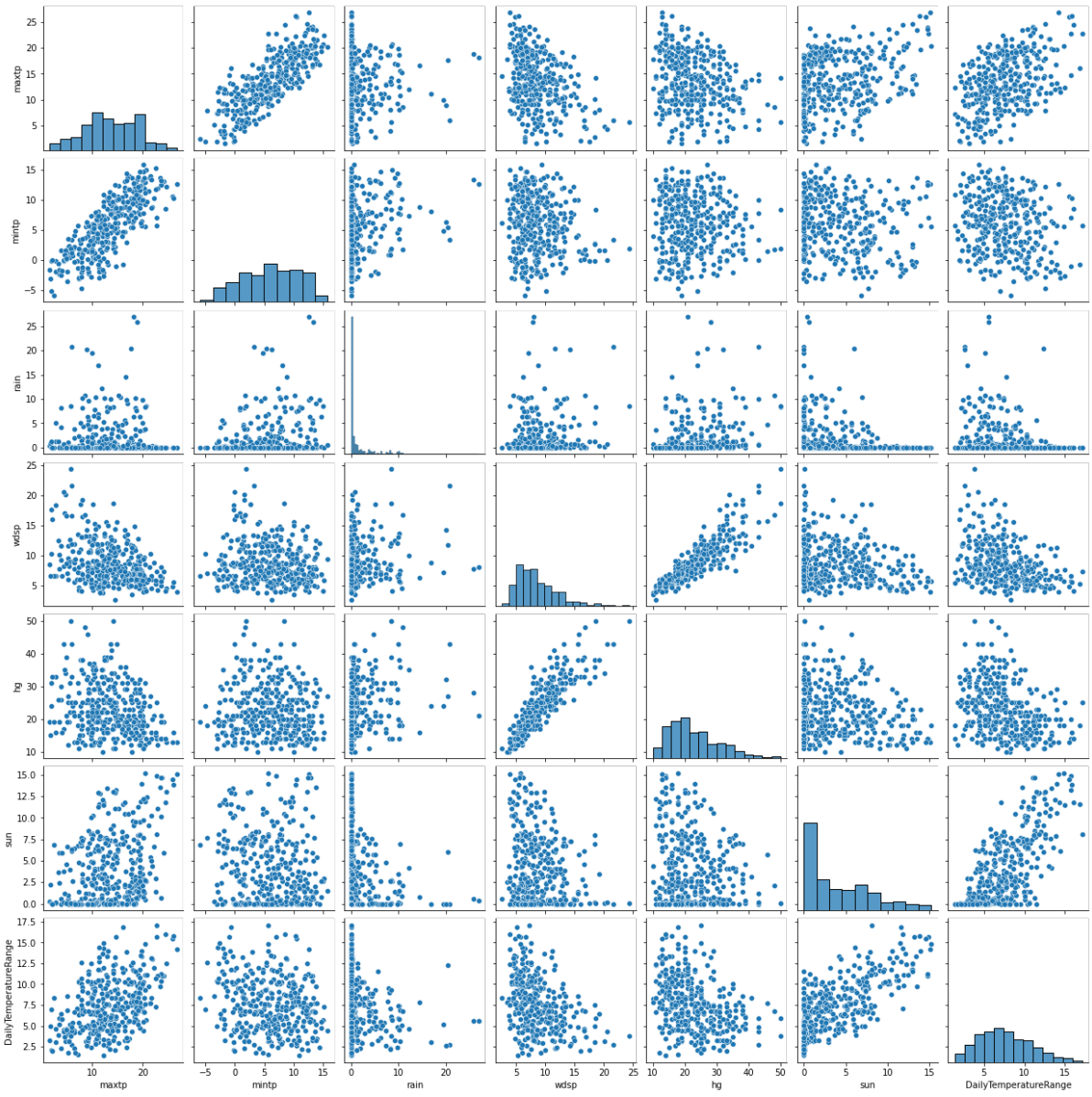
Out[17]:

	day	year	maxtp	mintp	rain	wdsp	hg	
count	365.000000	365.0	364.000000	364.000000	364.000000	365.000000	364.000000	36
mean	15.720548	2021.0	13.590934	6.001099	1.828846	8.625753	22.623626	
std	8.808321	0.0	5.258806	4.748774	3.878082	3.456376	7.790196	
min	1.000000	2021.0	1.500000	-5.900000	0.000000	2.600000	10.000000	
25%	8.000000	2021.0	10.000000	2.000000	0.000000	6.100000	17.000000	
50%	16.000000	2021.0	13.300000	6.200000	0.100000	7.900000	21.000000	
75%	23.000000	2021.0	18.100000	9.900000	1.325000	10.500000	27.250000	
max	31.000000	2021.0	26.800000	15.800000	26.900000	24.400000	50.000000	1

Graphical Summary of Dublin Airport:-

In [18]:
```
weather_ofdublinairport = weatherofdublinairport
weather_ofdublinairport = weather_ofdublinairport.drop(['day','year
sns.pairplot(weather_ofdublinairport)
```

Out[18]: <seaborn.axisgrid.PairGrid at 0x7fe06847e8b0>



Cork Airport:-

In [19]:
```
weatherofcorkairport = weatherdataset[weatherdataset['station']== '
weatherofcorkairport.head()
```

Out[19]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	DailyTemperatureRar
730	1	jan	2021	Cork Airport	5.0	0.9	0.0	12.1	28.0	6.9	
731	2	jan	2021	Cork Airport	5.1	-0.4	0.1	8.8	17.0	0.2	
732	3	jan	2021	Cork Airport	2.6	-1.6	0.1	10.8	21.0	6.3	
733	4	jan	2021	Cork Airport	3.3	-0.7	0.0	10.8	22.0	4.0	
734	5	jan	2021	Cork Airport	4.0	0.6	0.0	10.2	20.0	2.8	

Numerical Summary of Cork Airport:-

In [20]:
```
weatherofcorkairport.describe()
```

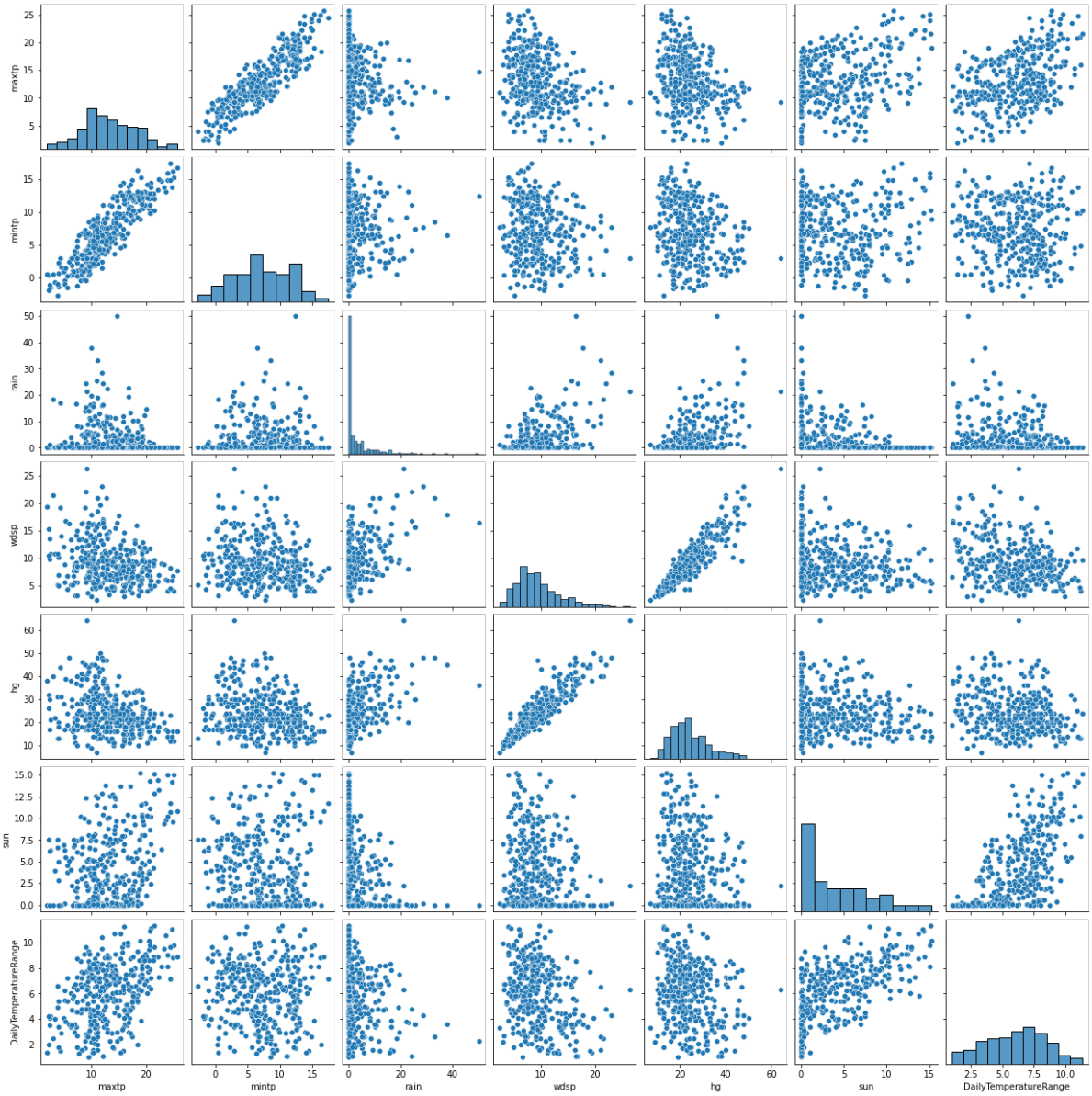
Out[20]:

	day	year	maxtp	mintp	rain	wdsp	hg	
count	365.000000	365.0	365.000000	365.000000	362.000000	364.000000	365.000000	36
mean	15.720548	2021.0	13.239726	7.236438	3.415470	9.321978	24.164384	
std	8.808321	0.0	4.826949	4.339971	6.234737	3.924194	8.843284	
min	1.000000	2021.0	1.900000	-2.700000	0.000000	2.400000	7.000000	
25%	8.000000	2021.0	9.900000	3.800000	0.000000	6.600000	18.000000	
50%	16.000000	2021.0	12.800000	6.900000	0.600000	8.500000	23.000000	
75%	23.000000	2021.0	16.800000	11.100000	4.300000	11.200000	29.000000	
max	31.000000	2021.0	25.700000	17.400000	50.100000	26.300000	64.000000	1

Graphical Summary of Cork Airport:-

In [21]:
```
weather_ofcorkairport = weatherofcorkairport
weather_ofcorkairport = weather_ofcorkairport.drop(['day','year'],a
sns.pairplot(weather_ofcorkairport)
```

Out[21]: <seaborn.axisgrid.PairGrid at 0x7fe06a6ceb80>



Shannon Airport:-

In [22]:
```
weatherofshannonairport = weatherdataset[weatherdataset['station']=
weatherofshannonairport.head()
```

Out[22]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	DailyTemperatureRa
365	1	jan	2021	Shannon Airport	6.5	-1.1	0.0	6.0	16.0	5.2	
366	2	jan	2021	Shannon Airport	6.6	-2.5	0.1	4.2	14.0	0.1	
367	3	jan	2021	Shannon Airport	2.6	-2.9	0.1	4.0	12.0	5.3	
368	4	jan	2021	Shannon Airport	3.1	-2.5	0.1	7.7	19.0	2.5	
369	5	jan	2021	Shannon Airport	4.2	-1.0	0.0	6.2	17.0	4.8	

Numerical Summary of Shannon Airport:-

In [23]:
```
weatherofshannonairport.describe()
```

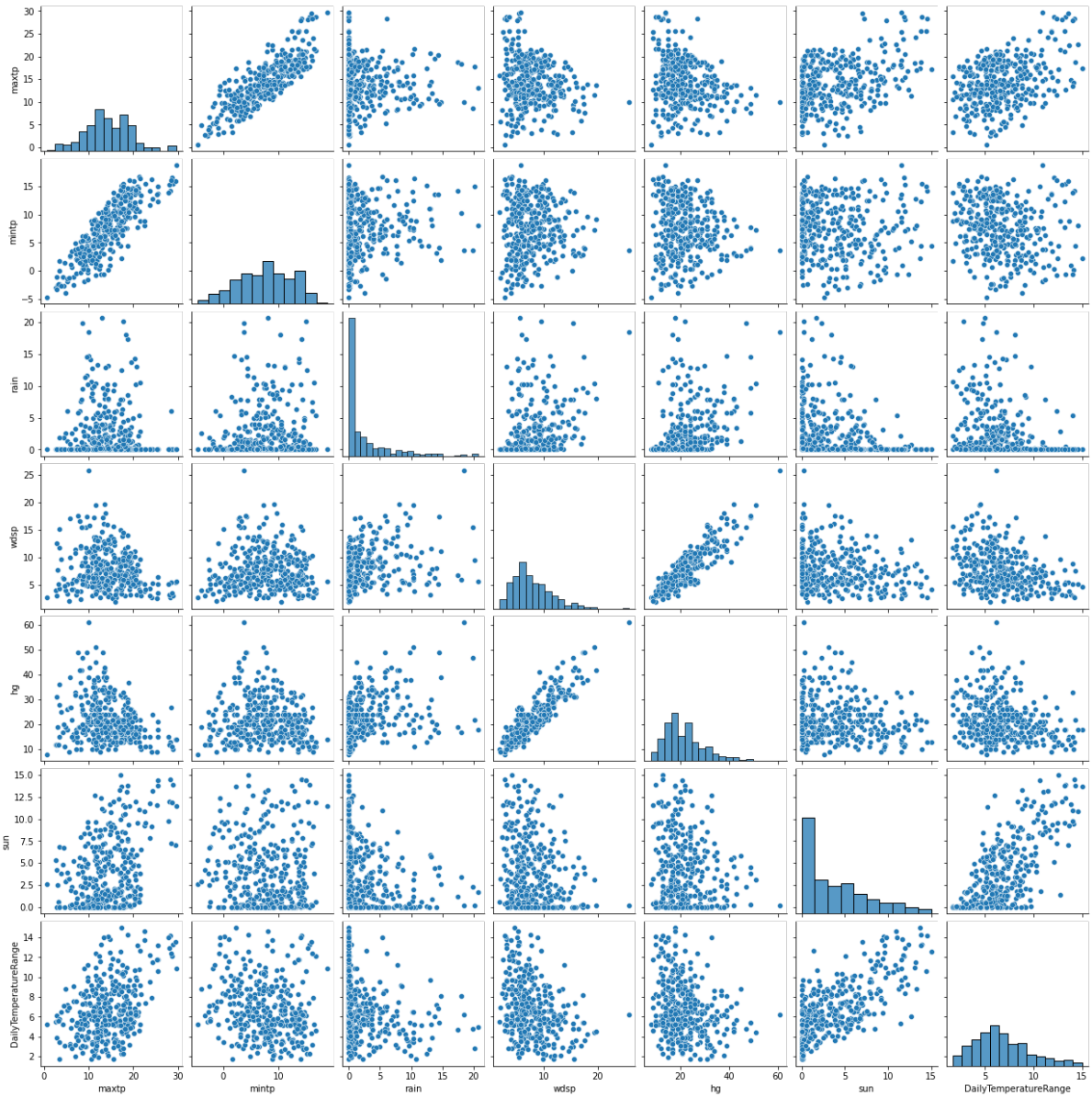
Out[23]:

	day	year	maxtp	mintp	rain	wdsp	hg	
count	365.000000	365.0	364.000000	364.000000	365.000000	365.000000	363.000000	36
mean	15.720548	2021.0	14.368681	7.611813	2.549041	7.815342	21.796143	
std	8.808321	0.0	5.175728	4.820239	4.007511	3.575878	8.372032	
min	1.000000	2021.0	0.600000	-4.600000	0.000000	1.900000	8.000000	
25%	8.000000	2021.0	11.100000	4.275000	0.000000	5.400000	16.000000	
50%	16.000000	2021.0	14.100000	7.650000	0.500000	6.900000	20.000000	
75%	23.000000	2021.0	18.000000	11.525000	3.400000	9.700000	26.000000	
max	31.000000	2021.0	29.600000	18.700000	20.700000	25.800000	61.000000	1

Graphical Summary of Shannon Airport:-

In [24]:
```
weather_ofshannonairport = weatherofshannonairport
weather_ofshannonairport = weather_ofshannonairport.drop(['day','ye
sns.pairplot(weather_ofshannonairport)
```

Out[24]: <seaborn.axisgrid.PairGrid at 0x7fe06d927fa0>



Discussing how the weather differs (or is similar) across these Dublin Airport,Cork Airport,Shannon Airport

MAXTP:THE MAXIMUM AIR TEMPETATURE(C)

There are 2 missing values out of the 364 values in maxtp.The maxtp's mean temperature in (Dublin,Cork,Shannon)Airports are roughly (13.59,13.23,14.36) degrees Celsius (C), and the data's standard deviation is (5.25,4.82,5.17). The range of the highest temperature (Dublin,Cork,Shannon)Airports are almost (27,26,30) C since the minimum and maximum values of the maximum temperature are (1.5,1.9,0.6) C and (26.8,25.7,29.6) C, respectively. Additionally, the summary gives the percentiles (25%,

50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%. MaxTP's 25%, 50%, and 75% values of(Dublin,Cork,Shannon)Airports are (10.0,13.3,18.1)C,(9.9,12.8,16.8)C,(11.1,14.1,18)C respectively.

MINTP:THE MINIMUM AIR TEMPETATURE(C)

Additionally, mintp contains 364 values, which shows that there are 2 missing values.The mintp's mean temperature in (Dublin,Cork,Shannon)Airports are roughly (6.0,7.2,7.6) degrees Celsius (C), and the data's standard deviation is (4.7,4.3,4.8). The range of the lowest temperature (Dublin,Cork,Shannon)Airports are almost (-5.9,-2.7,-4.6) C since the minimum and maximum values of the minimum temperature are (-5.9,-2.7,-4.6) C and (15.8,17.4,18.7) C, respectively. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%. MaxTP's 25%, 50%, and 75% values of(Dublin,Cork,Shannon)Airports are (2.0,6.2,9.9), (3.8,6.9,11.1),(4.2,7.6,11.5) respectively.

RAIN:THE AMOUNT OF PRECIPITATION(mm)

There are 4 missing values in rain, which has 365 values.These missing values must be filled either with test statistics or arbitrarily chosen values. The spread of the data in (Dublin,Cork,Shannon)Airports are roughly (3.87,6.23,4.00) mm, while the mean of the rain is (1.8,3.4,2.5) mm. Rainfall in (Dublin,Cork,Shannon)Airports ranges from (0,0,0) millimeters to (26.9,50.1,20.7) millimeters, respectively. This also implies that there will be (26.9,50.1,20.7)mm of rain. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%.It rains (0,0,0) mm, (0,0.6,0.5) mm, and (0.1,4.3,3.4) mm, respectively, for the 25%, 50%, and 75% in (Dublin,Cork,Shannon)Airports.

WDSP:THE MEAN WIND SPEED(knot)

There are 1 missing values according to the 365 values in wdsp. These missing values must be filled either with test statistics or arbitrarily chosen values. The data spread in (Dublin,Cork,Shannon)Airports are roughly (3.45,3.92,3.57)knots, while the wdsp mean is (8.6,9.3,7.8) knots. The value of min ranges from (2.6,2.4,1.9) knot to (24.4,26.3,25.8) knot, accordingly. Additionally, this indicates that the wdsp's range is (24.4,26.3,25.8) knot. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%.For wdsp, the respective 25%, 50%, and 75% in (Dublin,Cork,Shannon)Airports are (6.1,6.6,5.4)knots, (7.9,8.5,6.9)knots,(10.5,11.2,9.7)knots.

HG:THE HIGHEST GUST(knot)

The fact that hg has 363 values means there are three missing values, which should either be discarded or filled in with any arbitrary number or test statistic. The data spread in (Dublin,Cork,Shannon)Airports are roughly (7.7,8.8,8.3)knots, while the mean of the hg is approximately (22.6,24.1,21.7) knots. The range of hg is (50,64,61) knot since the minimum and highest values in (Dublin,Cork,Shannon)Airports are (10,7,8) knot and (50,64,61) knot, respectively. Additionally, the summary gives the percentiles (25%, 50%, and 75%). Lower quartile refers to the 25% of the dataset that is comprised of the first 25%. The higher quartiles, which make up 75% of the data, are referred to as the upper quartiles, while the median, which is the center number, is referred to as 50%.For hg, the respective 25%, 50%, and 75% are (17.0,18.0,16)knot,(21.0,23.0,20)knot, (25.2,29.0,26)knot in (Dublin,Cork,Shannon)Airports.

SUN:SUNSHINE DURATION(hours)

Sun contains 364 values, thus there are 2 missing values that should either be discarded or replaced in with any arbitrary number or test statistic. The spread of the data in (Dublin,Cork,Shannon)Airports are approximately (3.98,4.00,3.81)hours, and the average amount of sunshine each day is (4,4,3.8) hours. The range of sunshine is (15.2,15.2,15) hours because there are (15.2,15.2,15) hours between the minimum and highest sunshine duration of (0,0,0) hours. The percentiles (25%, 50%, and 75%) are also included in the summary. The bottom quartile, or 25% of the dataset, is made up of the first 25%. The center number, 50%, is referred to as the median, while the top quartiles, which make up 75% of the data, are referred to as the upper quartiles.For sun, the corresponding percentages are 25%, 50%, and 75% are (0.5,0.2,0.4)hours, (2.9,3.1,2.7)hours,(6.8,7.0,6.2)hours in (Dublin,Cork,Shannon)Airports..