WEATHER ANALYSIS

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sea
from matplotlib.pyplot import rcParams as rcp


import csv
import requests
url = "https://www.kaggle.com/muthuj7/weather-dataset"

df = requests.get(url)

df.headers
```

```
    {'Date': 'Fri, 19 Mar 2021 10:32:05 GMT', 'Content-Type': 'text/html; chars
```

```
df=pd.read_csv('weather_analysis/weatherHistory.csv')
```

```
df.head()
```

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	(
0	2006-04-01 00:00:00.000 +0200	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	
1	2006-04-01 01:00:00.000 +0200	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	
2	2006-04-01 02:00:00.000 +0200	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	
3	2006-04-01 03:00:00.000 +0200	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	
4	2006-04-01 04:00:00.000 +0200	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	

```
df.tail()
```

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Win Spee (km/h
96448	2016-09-09 19:00:00.000 +0200	Partly Cloudy	rain	26.016667	26.016667	0.43	10.996
96449	2016-09-09 20:00:00.000 +0200	Partly Cloudy	rain	24.583333	24.583333	0.48	10.094
96450	2016-09-09 21:00:00.000 +0200	Partly Cloudy	rain	22.038889	22.038889	0.56	8.983
96451	2016-09-09 22:00:00.000 +0200	Partly Cloudy	rain	21.522222	21.522222	0.60	10.529
96452	2016-09-09 23:00:00.000 +0200	Partly Cloudy	rain	20.438889	20.438889	0.61	5.876

```
df.shape
```

```
(96453, 12)
```

```
df.describe()
```

	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visib
count	96453.000000	96453.000000	96453.000000	96453.000000	96453.000000	96453.0
mean	11.932678	10.855029	0.734899	10.810640	187.509232	10.3
std	9.551546	10.696847	0.195473	6.913571	107.383428	4.1
min	-21.822222	-27.716667	0.000000	0.000000	0.000000	0.0
25%	4.688889	2.311111	0.600000	5.828200	116.000000	8.3
50%	12.000000	12.000000	0.780000	9.965900	180.000000	10.0
75%	18.838889	18.838889	0.890000	14.135800	290.000000	14.8
max	39.905556	39.344444	1.000000	63.852600	359.000000	16.1

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96453 entries, 0 to 96452
Data columns (total 12 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Formatted Date            96453 non-null  object 
 1   Summary                   96453 non-null  object 
 2   Precip Type               95936 non-null  object 
 3   Temperature (C)           96453 non-null  float64
 4   Apparent Temperature (C)  96453 non-null  float64
 5   Humidity                  96453 non-null  float64
 6   Wind Speed (km/h)         96453 non-null  float64
 7   Wind Bearing (degrees)    96453 non-null  float64
 8   Visibility (km)           96453 non-null  float64
 9   Loud Cover                96453 non-null  float64
 10  Pressure (millibars)      96453 non-null  float64
 11  Daily Summary             96453 non-null  object 
dtypes: float64(8), object(4)
memory usage: 8.8+ MB
```

```
df.dtypes
```

```
Formatted Date               object
Summary                      object
Precip Type                  object
Temperature (C)             float64
Apparent Temperature (C)    float64
Humidity                    float64
Wind Speed (km/h)           float64
Wind Bearing (degrees)      float64
Visibility (km)             float64
Loud Cover                  float64
Pressure (millibars)        float64
Daily Summary                object
dtype: object
```

```
categories = df.select_dtypes(include = ["object"]).keys()
print(categories)
```

```
Index(['Formatted Date', 'Summary', 'Precip Type', 'Daily Summary'], dtype=
```

```
quantitative = df.select_dtypes(include = ["int64","float64"]).keys()
print(quantitative)
```

```
Index(['Temperature (C)', 'Apparent Temperature (C)', 'Humidity',
       'Wind Speed (km/h)', 'Wind Bearing (degrees)', 'Visibility (km)',
       'Loud Cover', 'Pressure (millibars)'],
      dtype='object')
```

```
df.isnull().sum()
```

```
Formatted Date               0
Summary                      0
Precip Type                517
Temperature (C)              0
Apparent Temperature (C)     0
Humidity                     0
Wind Speed (km/h)            0
Wind Bearing (degrees)       0
Visibility (km)              0
Loud Cover                   0
Pressure (millibars)         0
Daily Summary                0
dtype: int64
```

```
df['Precip Type'].value_counts()
```

```
rain    85224
snow    10712
Name: Precip Type, dtype: int64
```

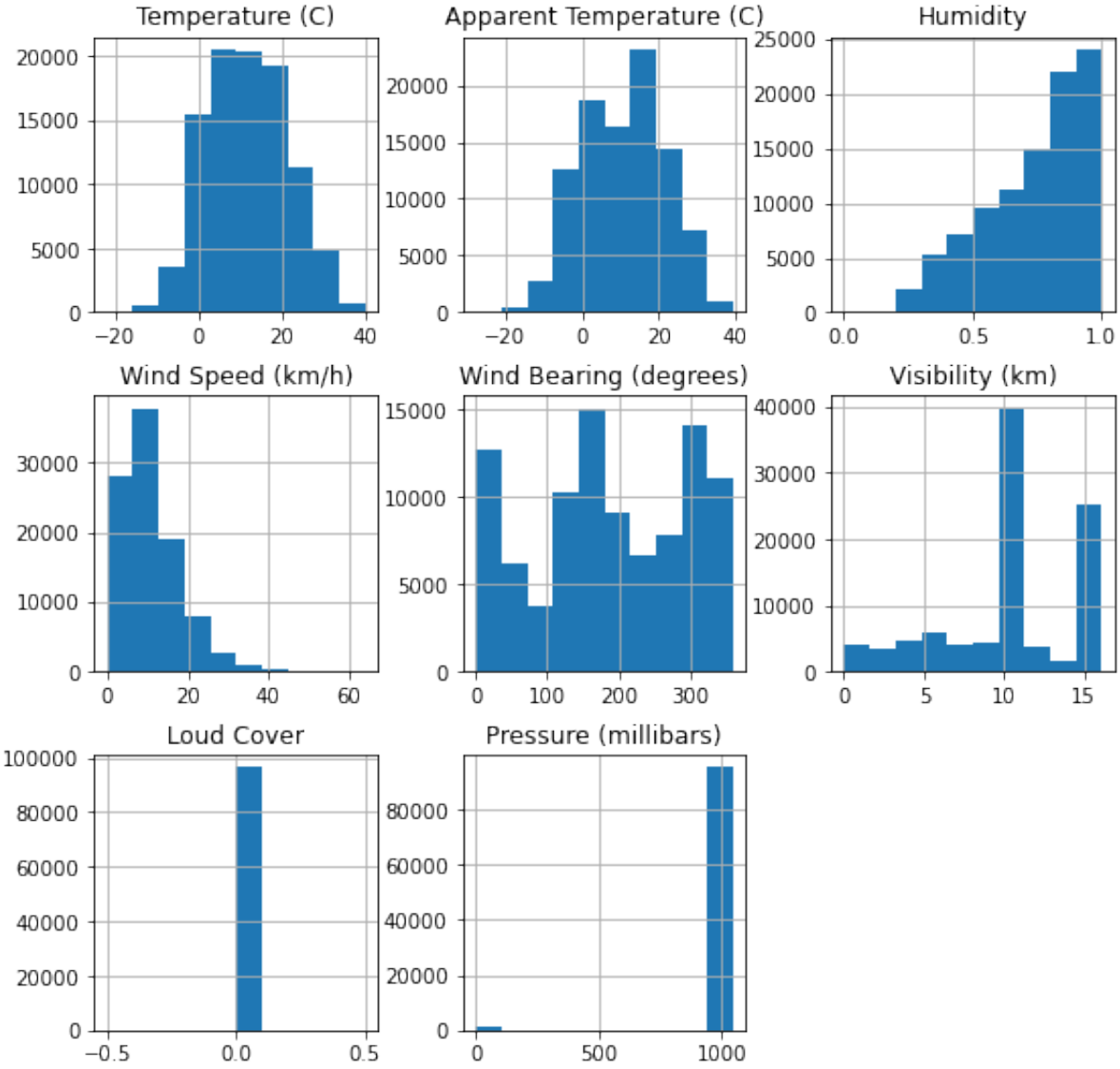
```
df['Precip Type'].fillna(method='ffill',inplace=True,axis=0)
df['Precip Type'].value_counts()
```

```
rain    85741
snow    10712
Name: Precip Type, dtype: int64
```

HISTOGRAMS

```
rcp['figure.figsize'] = 9, 9
df[quantitative].hist()
```

```
array([[<AxesSubplot:title={'center':'Temperature (C)'}>,
        <AxesSubplot:title={'center':'Apparent Temperature (C)'}>,
        <AxesSubplot:title={'center':'Humidity'}>],
       [<AxesSubplot:title={'center':'Wind Speed (km/h)'}>,
        <AxesSubplot:title={'center':'Wind Bearing (degrees)'}>,
        <AxesSubplot:title={'center':'Visibility (km)'}>],
       [<AxesSubplot:title={'center':'Loud Cover'}>,
        <AxesSubplot:title={'center':'Pressure (millibars)'}>,
        <AxesSubplot:>]], dtype=object)
```




```
df=df.drop('Loud Cover',axis=1)
```

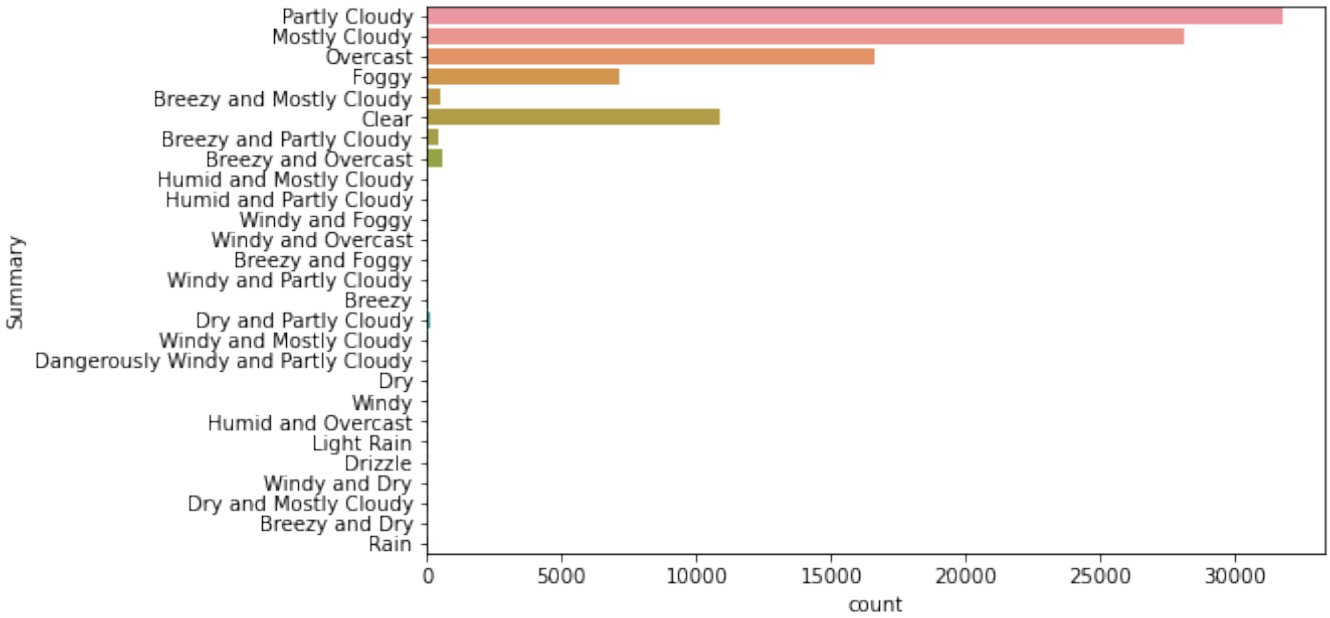
```
df['Summary'].value_counts()
```

```
Partly Cloudy                             31733
Mostly Cloudy                             28094
Overcast                                  16597
Clear                                     10890
Foggy                                      7148
Breezy and Overcast                         528
Breezy and Mostly Cloudy                    516
Breezy and Partly Cloudy                    386
Dry and Partly Cloudy                        86
Windy and Partly Cloudy                      67
Light Rain                                   63
Breezy                                       54
Windy and Overcast                           45
Humid and Mostly Cloudy                      40
Drizzle                                      39
Breezy and Foggy                             35
Windy and Mostly Cloudy                      35
Dry                                          34
Humid and Partly Cloudy                      17
Dry and Mostly Cloudy                        14
Rain                                         10
Windy                                         8
Humid and Overcast                            7
Windy and Foggy                               4
Dangerously Windy and Partly Cloudy           1
Breezy and Dry                                1
Windy and Dry                                 1
Name: Summary, dtype: int64
```

COUNTPLOT

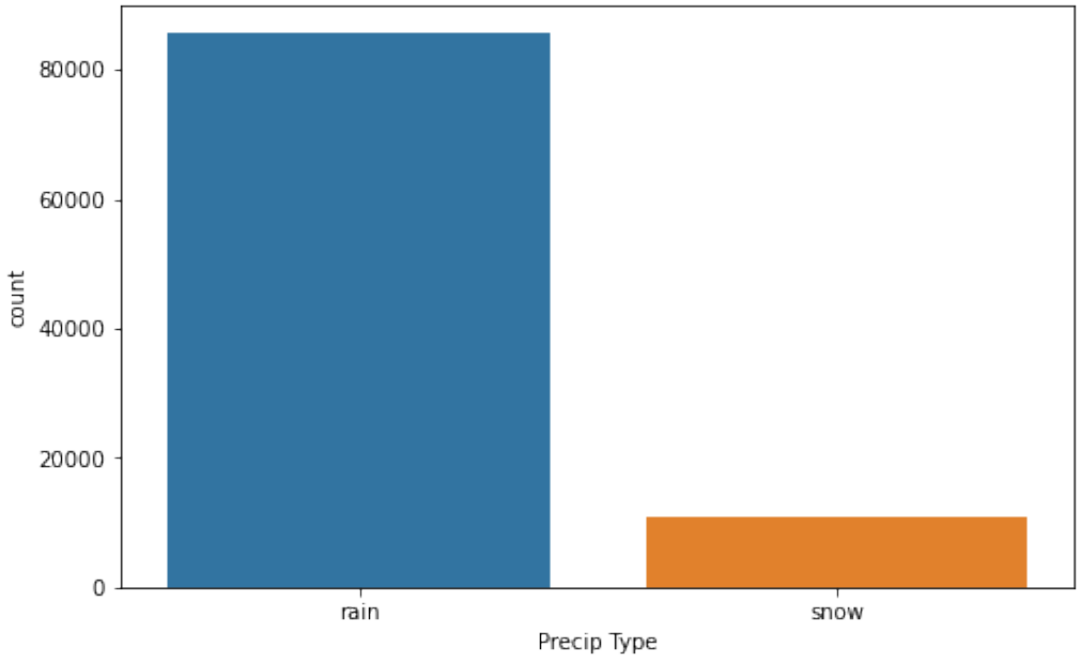
```
rcp['figure.figsize'] = 8, 5
sns.countplot(y=df['Summary'])
```

```
<AxesSubplot:xlabel='count', ylabel='Summary'>
```



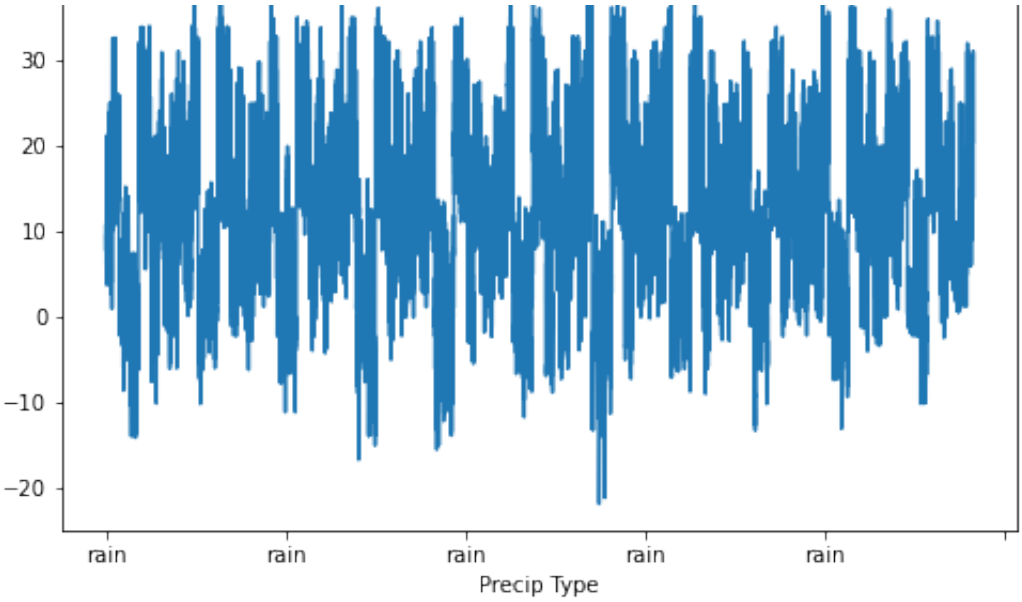

```
sea.countplot(x=df['Precip Type'])
```

```
<AxesSubplot:xlabel='Precip Type', ylabel='count'>
```

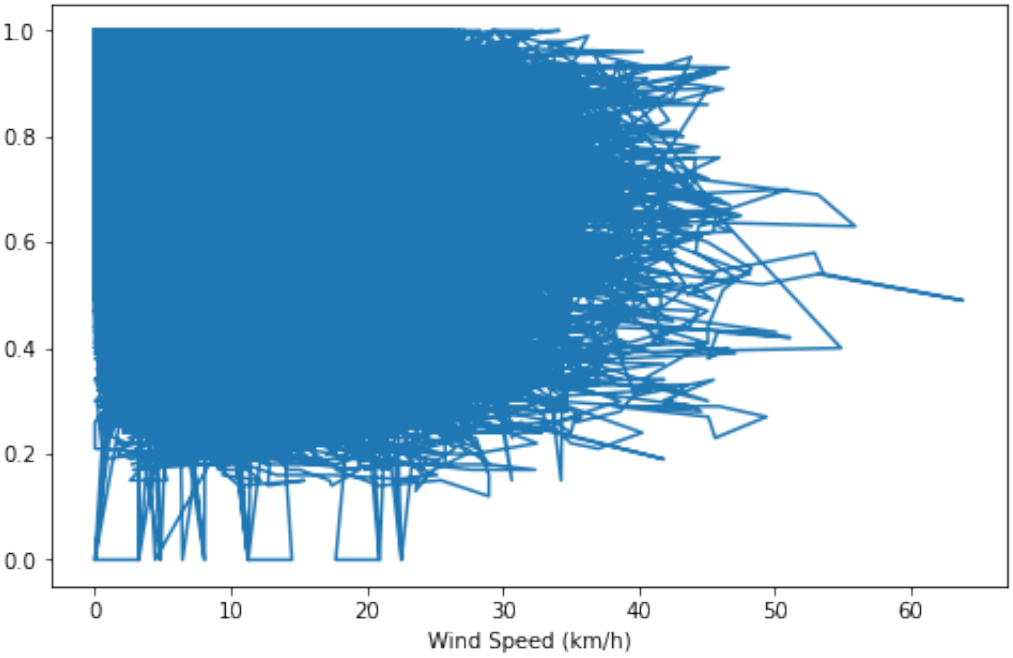



```
for col in df.columns:
  print(col)
ts = pd.Series(df['Temperature (C)'].values, index=df['Precip Type'])
fig = plt.figure()
plt.suptitle('Time Series for temp vs precip type')
ts.plot()
```

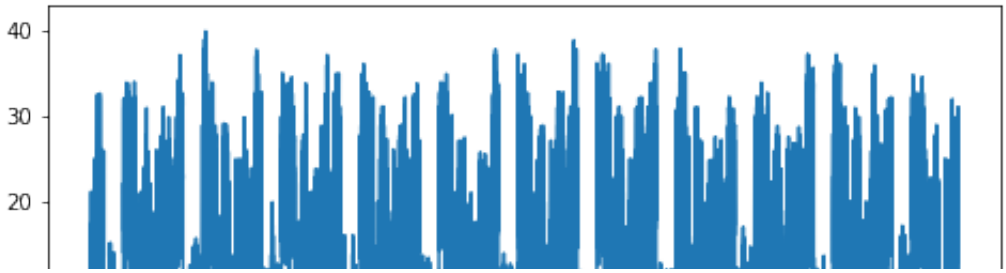
```
ts1 = pd.Series(df['Humidity'].values, index=df['Wind Speed (km/h)'])
fig = plt.figure()
plt.suptitle('Time Series for humidity vs wind speed')
ts1.plot()

ts2 = pd.Series(df['Temperature (C)'].values, index=df['Daily Summary'])
fig = plt.figure()
plt.suptitle('Time Series for temp vs Daily summary')
ts2.plot()
```

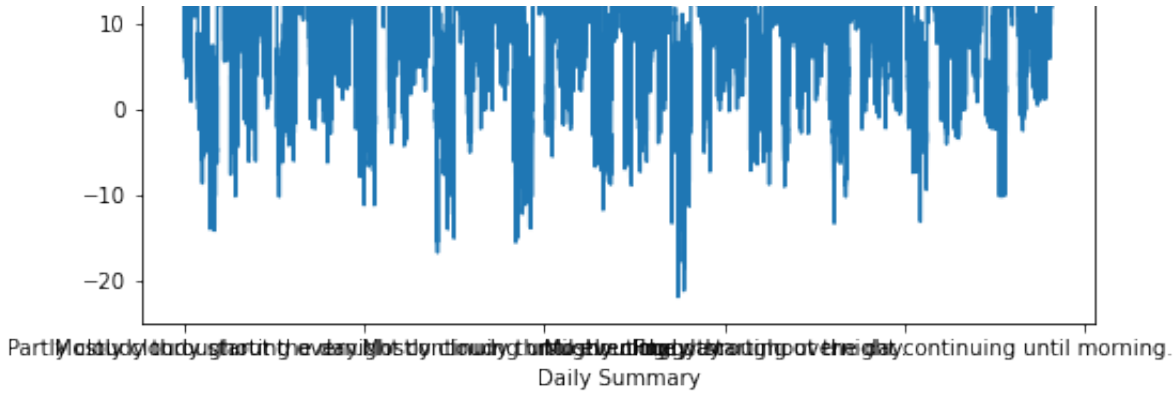


Time Series for humidity vs wind speed



Time Series for temp vs Daily summary

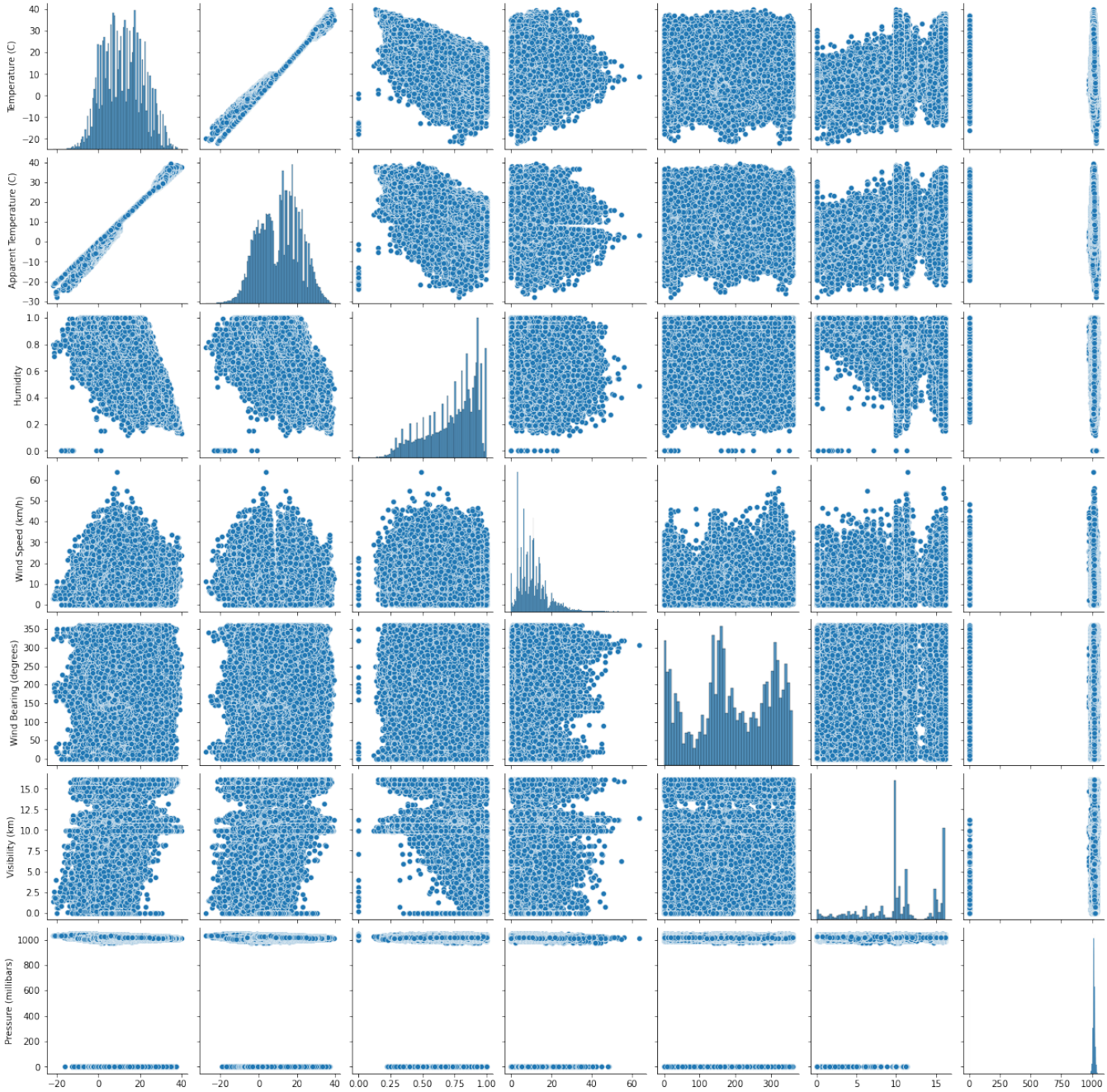




PAIRPLOT

```
sea.pairplot(df,palette="coolwarm")
```

```
<seaborn.axisgrid.PairGrid at 0x7fbdeeae9970>
```



Temperature (C)
Apparent Temperature (C)
Humidity
Wind Speed (km/h)
Wind Bearing (degrees)
Visibility (km)
Pressure (millibars)