

# Weather prediction analysis...

(Whether to go to picnic or not..)

**Step1-** Loaded both the data csv in python code.

```
[65] import pandas as pd
df1 = pd.read_csv('weather_prediction_dataset.csv')
df1.head()

[66] df2 = pd.read_csv('weather_prediction_picnic_labels.csv')
df2.head()
```

**Step2-** since I to work with both the csv while analysing whether the given different values of variable make any impact on the decision of whether to go to picnic for not and how much. Therefore, we will merge the two-csv making a third dataframe (df3).

```
df3 = pd.merge(df1, df2).copy()
df3.head()
```

ity	BASEL_pressure	BASEL_global_radiation	BASEL_precipitation	BASEL_sunshine	BASEL_temp_mean	BASEL_temp_min	...	LJUBLJANA_picnic_weather
1.89	1.0286	0.20	0.03	0.0	2.9	1.6	...	False
1.87	1.0318	0.25	0.00	0.0	3.6	2.7	...	False
1.81	1.0314	0.50	0.00	3.7	2.2	0.1	...	False
1.79	1.0262	0.63	0.35	6.9	3.9	0.5	...	False
1.90	1.0246	0.51	0.07	3.7	6.0	3.8	...	False

**Step3-** For now i will work for one location and few of the variables as of beginner level but after completing the code for analysing after that it will be easy to work for other locations as well by making some changes in code.

```
[101] df4 = df3[['DATE', 'BASEL_precipitation', 'BASEL_sunshine', 'BASEL_temp_mean', 'BASEL_temp_min', 'BASEL_temp_max', 'BASEL_pressure', 'BASEL_humidity',
df4.head()
```

DATE	BASEL_precipitation	BASEL_sunshine	BASEL_temp_mean	BASEL_temp_min	BASEL_temp_max	BASEL_pressure	BASEL_humidity	BASEL_picnic_weather
00101	0.03	0.0	2.9	1.6	3.9	1.0286	0.89	False
00102	0.00	0.0	3.6	2.7	4.8	1.0318	0.87	False
00103	0.00	3.7	2.2	0.1	4.8	1.0314	0.81	False
00104	0.35	6.9	3.9	0.5	7.5	1.0262	0.79	False
00105	0.07	3.7	6.0	3.8	8.6	1.0246	0.90	False

**Step4-** I have used a function from pandas library for co-relation, which gives as relation on how different variable affected the BASEL\_picnic\_weather and you will be able to see that DATE has

nearly no co-relation with it, while pressure also has a quiet low impact.

```
corr_matrix= df4.corr()
corr_matrix["BASEL_picnic_weather"]
```

DATE	-0.017003
BASEL_precipitation	-0.257213
BASEL_sunshine	0.613136
BASEL_temp_mean	0.561821
BASEL_temp_min	0.455671
BASEL_temp_max	0.600784
BASEL_pressure	0.040971
BASEL_humidity	-0.442578
BASEL_picnic_weather	1.000000

Name: BASEL\_picnic\_weather, dtype: float64

**Step5-** Then I have written the code for co relation using the library seaborn whose regplot uses linear regression model to plot co-

```
import seaborn as sns
plt.figure(figsize=(10, 6))
for column in df4.columns:
    a=df4[column]
    sns.scatterplot(x='BASEL_sunshine', y=a, data=df4, color='royalblue', marker='o')
    if column=='BASEL_picnic_weather':
        continue

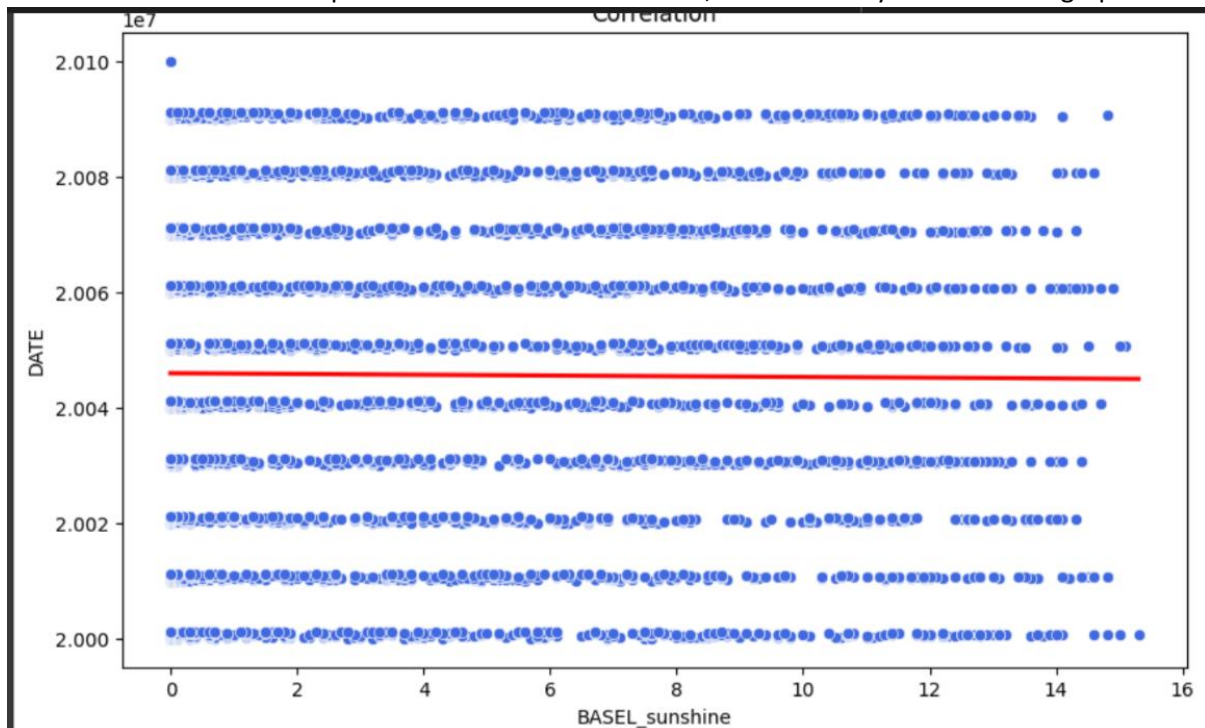
    sns.regplot(x='BASEL_sunshine', y=a, data=df4, scatter=False, color='red', ci=None)

plt.title('Correlation')
plt.xlabel('BASEL_sunshine')
plt.ylabel(column)
plt.show()
```

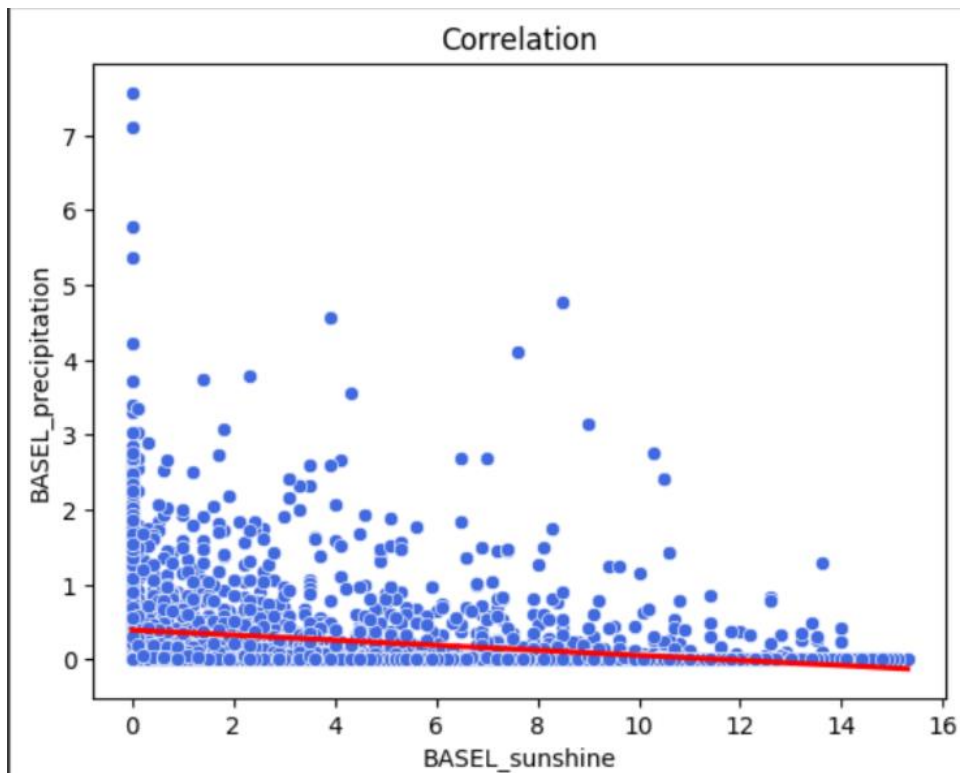
relation graph, I have used a loop for making multiple graphs.

## Analysis-

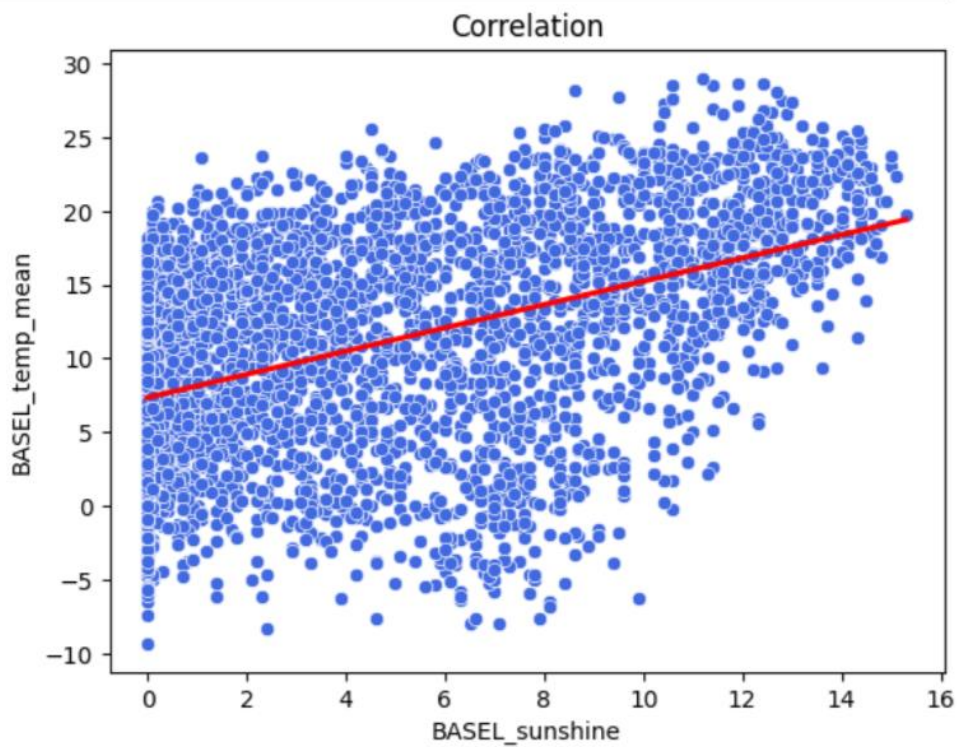
As we know date has no impact and relation with sunshine, which is easily visible on the graph



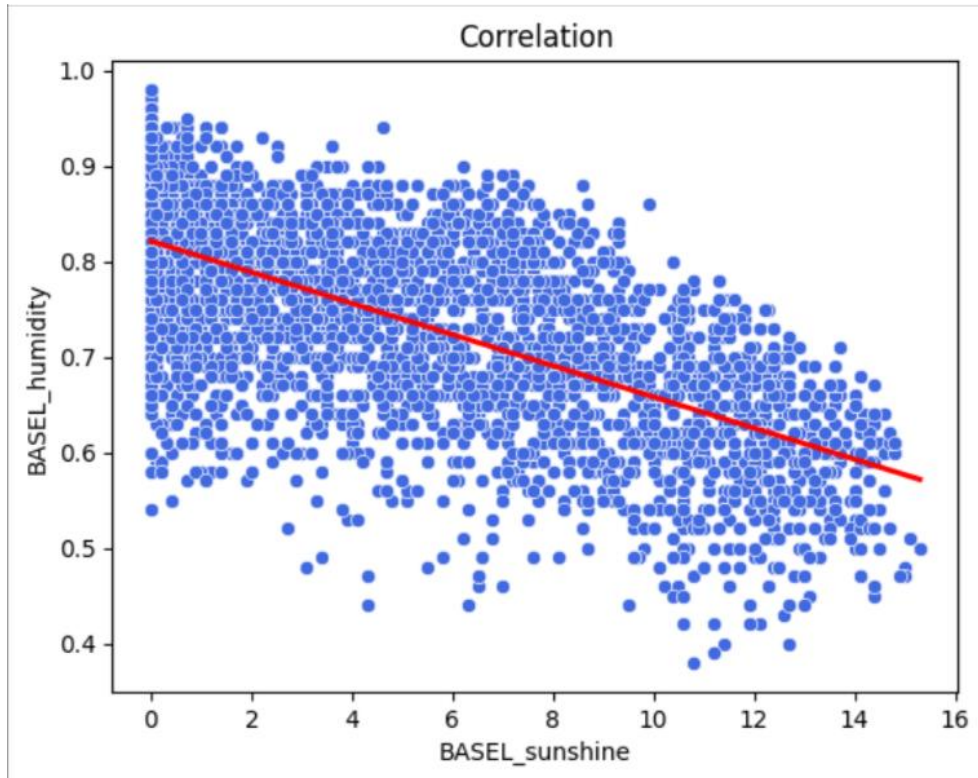
Next we have sunshine versus precipitation graph, precipitation has very less impact on sunshine.



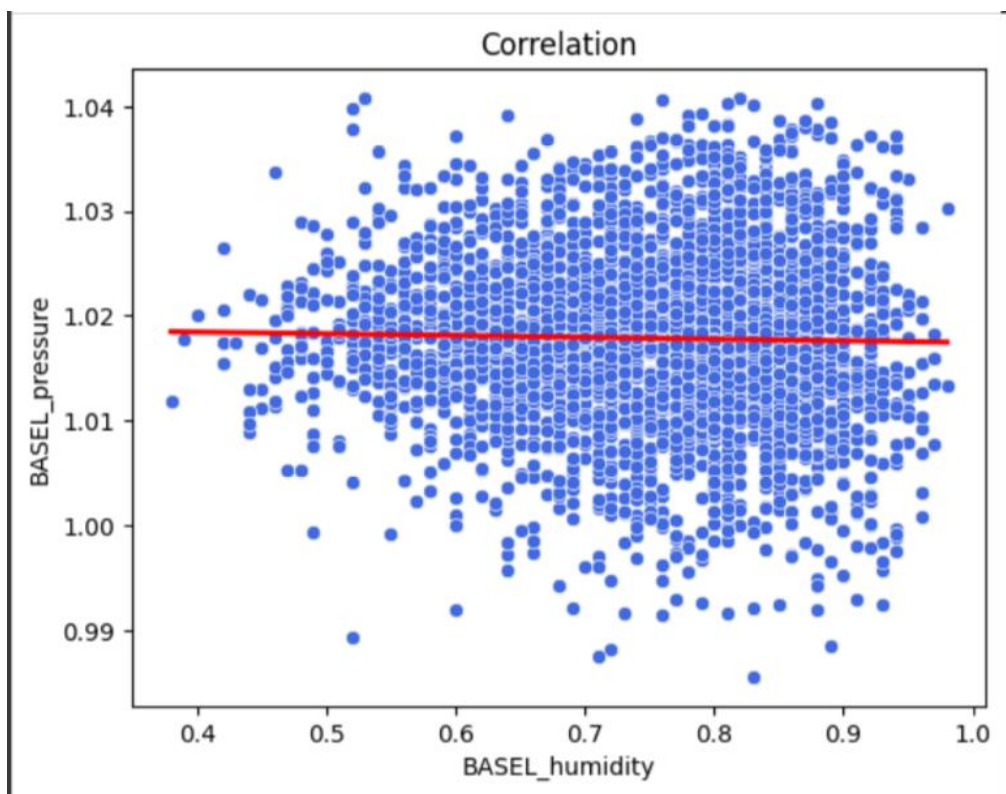
Now we have mean temp versus sunshine, as sunshine increases mean temp also starts to increase.



Next we have humidity vs sunshine, as sunshine increases humidity decreases.



Pressure vs humidity, as seen from the graph we can say that there is no relation between pressure and humidity.



This is how we analyse a data set and find the co relation between different variables and after this we can even include some new features with the combination of given features to enhance the accuracy and to be easily understandable, we then Train the machine and take prediction from it for the given data and find how much accuracy and precision it has.