Predicting the power generation in a solar plant

Yaswanth Kottana 110120056; Tom Joseph 110120112; Harshmeet Singh Saluja 110120040

The data set is used to predict the power generation of a solar plant according to the irradiating from the Sum, ambient temperature of the atmosphere around and the module temperature.

Github link for the code: https://github.com/TomJosephKavalam/CSOE18Proj

Visualising Data

The follwing code uses seaborn to visualise the dataset and how the features vary with each other.

In [29]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sn
```

In [30]:

```
df=pd.read_csv('Dataset/P1.csv')
df
```

Out[30]:

	Unnamed: 0	DC_POWER	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
0	0	0.0	23.128673	20.464305	0.0
1	1	0.0	23.032562	20.341429	0.0
2	2	0.0	22.967493	20.269493	0.0
3	3	0.0	22.810594	20.198918	0.0
4	4	0.0	22.611436	20.085866	0.0
...	...	...	...	...	...
3152	3152	0.0	23.670292	21.691071	0.0
3153	3153	0.0	23.795434	22.067778	0.0
3154	3154	0.0	23.727901	21.662972	0.0
3155	3155	0.0	23.497284	21.051402	0.0
3156	3156	0.0	23.244698	20.774560	0.0

3157 rows × 5 columns

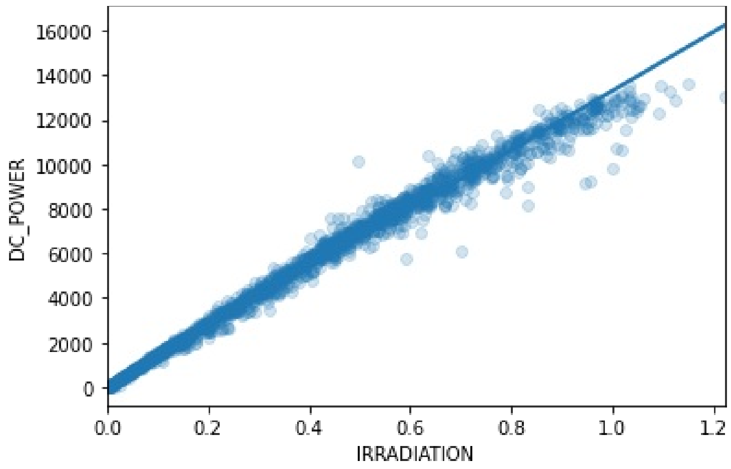
Visualising the trend between IRRADIATION and DC_POWER using scatter plot

In [31]:

```
sn.regplot(x = "IRRADIATION", y="DC_POWER", data=df, fit_reg = True, scatter_kws={"alpha
": 0.2})
```

Out[31]:

```
<AxesSubplot:xlabel='IRRADIATION', ylabel='DC_POWER'>
```



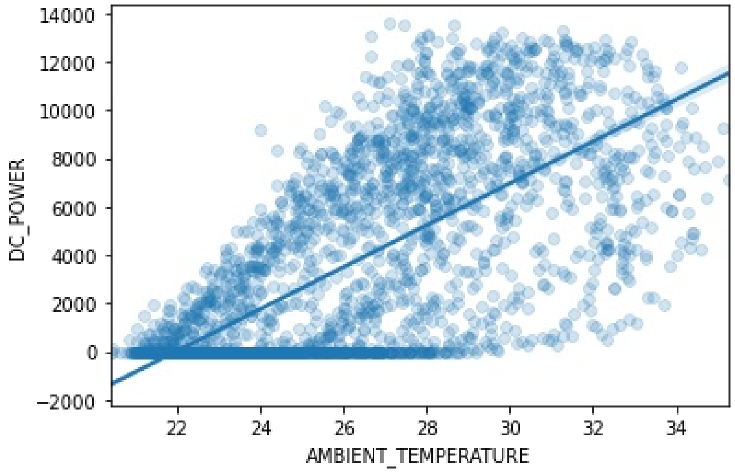
Visualising the trend between AMBIENT_TEMPERATURE and DC_POWER using scatter plot

In [32]:

```
sn.regplot(x = "AMBIENT_TEMPERATURE", y="DC_POWER", data=df, fit_reg = True, scatter_kws
={"alpha": 0.2})
```

Out[32]:

```
<AxesSubplot:xlabel='AMBIENT_TEMPERATURE', ylabel='DC_POWER'>
```



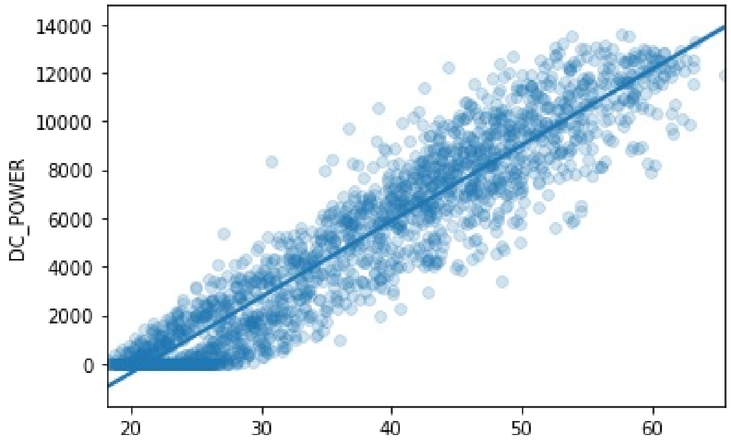
Visualising the trend between MODULE_TEMPERATURE and DC_POWER using scatter plot

In [33]:

```
sn.regplot(x = "MODULE_TEMPERATURE", y="DC_POWER", data=df, fit_reg = True, scatter_kws=
{"alpha": 0.2})
```

Out[33]:

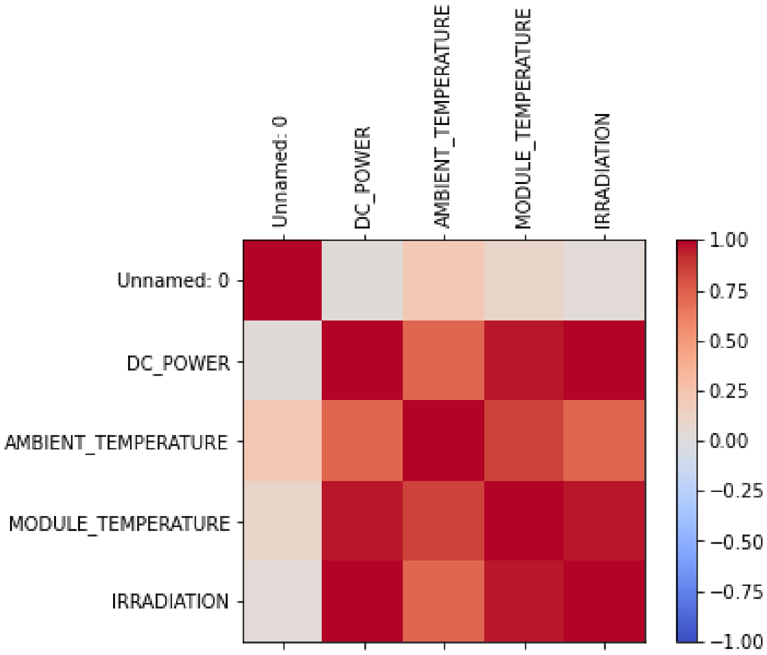
```
<AxesSubplot:xlabel='MODULE_TEMPERATURE', ylabel='DC_POWER'>
```



Visualising the correlation between the four variables (consider the column 'Unamed: 0' null as it is used for indexing the dataset).

In [34]:

```
columns=[]
corr = df.corr()
fig = plt.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(corr,cmap='coolwarm', vmin=-1, vmax=1)
fig.colorbar(cax)
ticks = np.arange(0,len(df.columns),1)
ax.set_xticks(ticks)
plt.xticks(rotation=90)
ax.set_yticks(ticks)
ax.set_xticklabels(df.columns)
ax.set_yticklabels(df.columns)
plt.show()
```


Unnamed: 0
DC_POWER
AMBIENT_TEMPERATURE
MODULE_TEMPERATURE
IRRADIATION
Unnamed: 0
DC_POWER
AMBIENT_TEMPERATURE
MODULE_TEMPERATURE
IRRADIATION