

```
# -*- coding: utf-8 -*-  
"""Untitled1.ipynb
```

Automatically generated by Colaboratory.

Original file is located at

```
https://colab.research.google.com/drive/1VX5SLtxA5LyOFIqFlnERr9gM\_rTyyHqq  
"""
```

```
import numpy as np #Numpy:It is used to perform mathematical operations and contains multi-dimensional arrays a  
nd matrix data structure.  
import pandas as pd #Pandas : It provides a flexible way to merge, concatenate or reshape the data.  
import matplotlib.pyplot as plt #Matplotlib : It is used to plot data in graphs in Python.  
import seaborn as sns #Seaborn : It is used for statistical plotting.
```

```
pd.read_csv("Statewise electricity consumption.csv") #Statewise Electricity Consumption data
```

```
data=pd.read_csv("Statewise electricity consumption.csv")  
data['Total'].max() #Found out the state i.e. Maharashtra which consumes the maximum electricity in India
```

```
import numpy as np #Numpy:It is used to perform mathematical operations and contains multi-dimensional arrays a  
nd matrix data structure.  
import pandas as pd #Pandas : It provides a flexible way to merge, concatenate or reshape the data.  
import matplotlib.pyplot as plt #Matplotlib : It is used to plot data in graphs in Python.  
import seaborn as sns #Seaborn : It is used for statistical plotting.  
sns.set()
```

```
from sklearn.linear_model import LinearRegression #sklearn : It is used to capture dependencies between electricity  
consumption (Y) and electricity price (X1), unemployment rate (X2) and per capita GDP(X3).
```

```
data = pd.read_csv("Electricity Consumption 2021.csv") #This file contains the data of Maharashtra from year 2010-  
2021
```

```
data.axes #Used for segregating the X and Y variables
```

```
data=pd.read_csv("Electricity Consumption 2021.csv")
```

```
X = data[['Electricity Price (per unit) X1',"Unemployment rate(in %) X2',"Per capita GDP(in Rs.) X3"]]  
y = data["Electricity Consumption( in GWh) (Y)"]
```

```
reg = LinearRegression()  
reg.fit(X,y) #reg.fit is used to fit the model to the input training instances.
```

```
reg.coef_ #To predict the coefficient of X1,X2,X3.
```

```
reg.intercept_ #To predict the constant term of equation.
```

```
reg.score(X,y) #It predicts the future value based on independent variables.
```

```
r2 = reg.score(X,y)  
n= X.shape[0] #Showing the number of rows in data.  
p= X.shape[1] #Showing number of columns on X's data. (X1,X2,X3)  
adjusted_r2 = 1-(1-r2)*(n-1)/(n-p-1)  
print (n)  
print (p)
```

adjusted_r2

```
from sklearn.feature_selection import f_regression
result = f_regression(X,y)
p_values = result[1]
```

p_values.round(3) #p values of all the three factors

```
data=pd.read_csv("Electricity Consumption 2021 PLOT.csv")
data
```

```
data=pd.read_csv #Output for Regression model
```

```
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.formula.api as smf
model = smf.ols(formula='ElectricityConsumption ~ ElectricityPrice + UnemploymentRate + PerCapitaGDP',data=
pd.read_csv('Electricity Consumption 2021 PLOT.csv'))
results_formula=model.fit()
results_formula.params
```

```
Result = pd.read_csv("AllAppliances.csv")
print(Result)
```

```
Result['Avg power consumption per household/year'].max() #Found out that maximum power is consumed by Air C
onditioner
```

```
pd.read_csv("Appliance Final.csv")
```

```
df.corr() #Correlation of top three most used appliances
```