NAAN MUTHALVAN

IBM COLLABARATE

ARTIFICIAL INTELLIGENCE

PROJECT TITLE

MEASURE ENERGY CONSUMPTION

NAME : PRAVEEN KUMAR M

DEPT & YEAR : CSE & III yr

REG.NO : 712221104014

COLLEGE : PARK CLLEGE OF ENGINEERING AND TECHNOLOGY

DATA ANALYSIS AND PREPROCESSING

**STEP 1:**

**Import library**

**Code :**

**import** **pandas** **as** **pd**

**import** **numpy** **as** **np**

**import** **matplotlib.pyplot** **as** **plt**

**import** **seaborn** **as** **sns**

**import** **pprint**

%**matplotlib** inline

df = pd.read\_csv("AEP\_hourly.csv")

print("="\*50)

print("First Five Rows ","**\n**")

print(df.head (5),"**\n**")

print("="\*50)

print("Information About Dataset","**\n**")

print(df.info(),"**\n**")

print("="\*50)

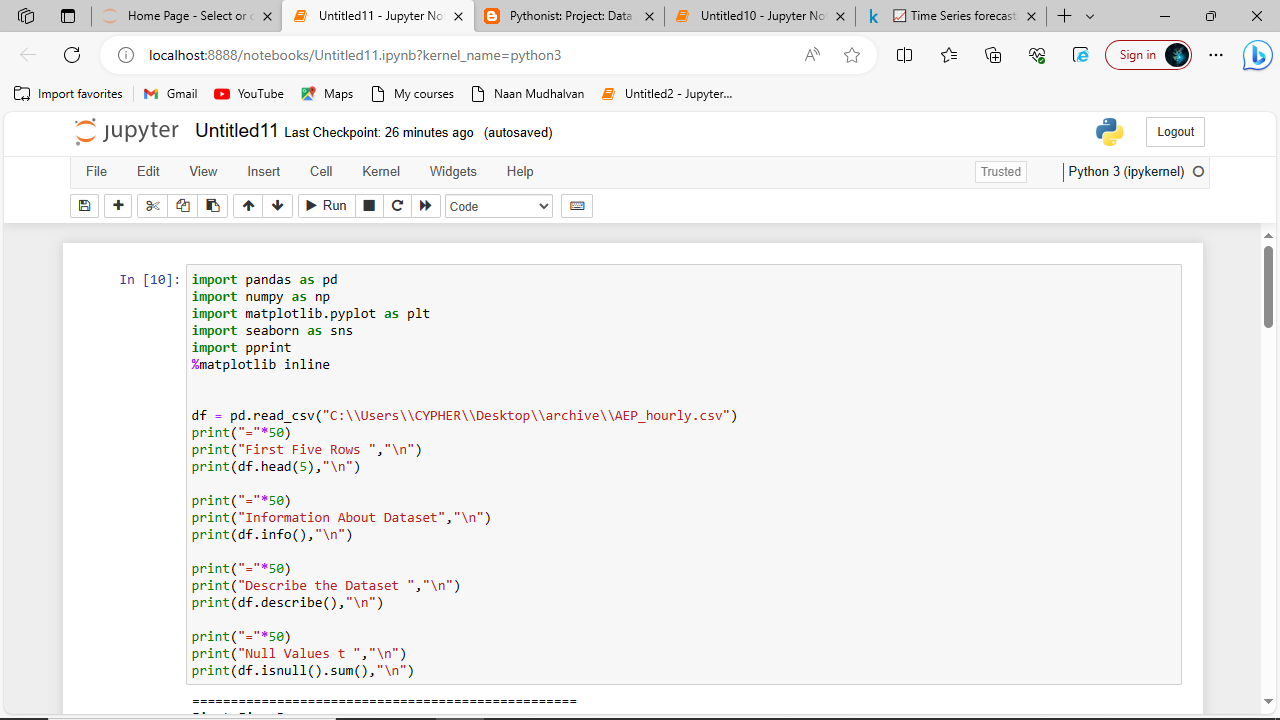
print("Describe the Dataset ","**\n**")

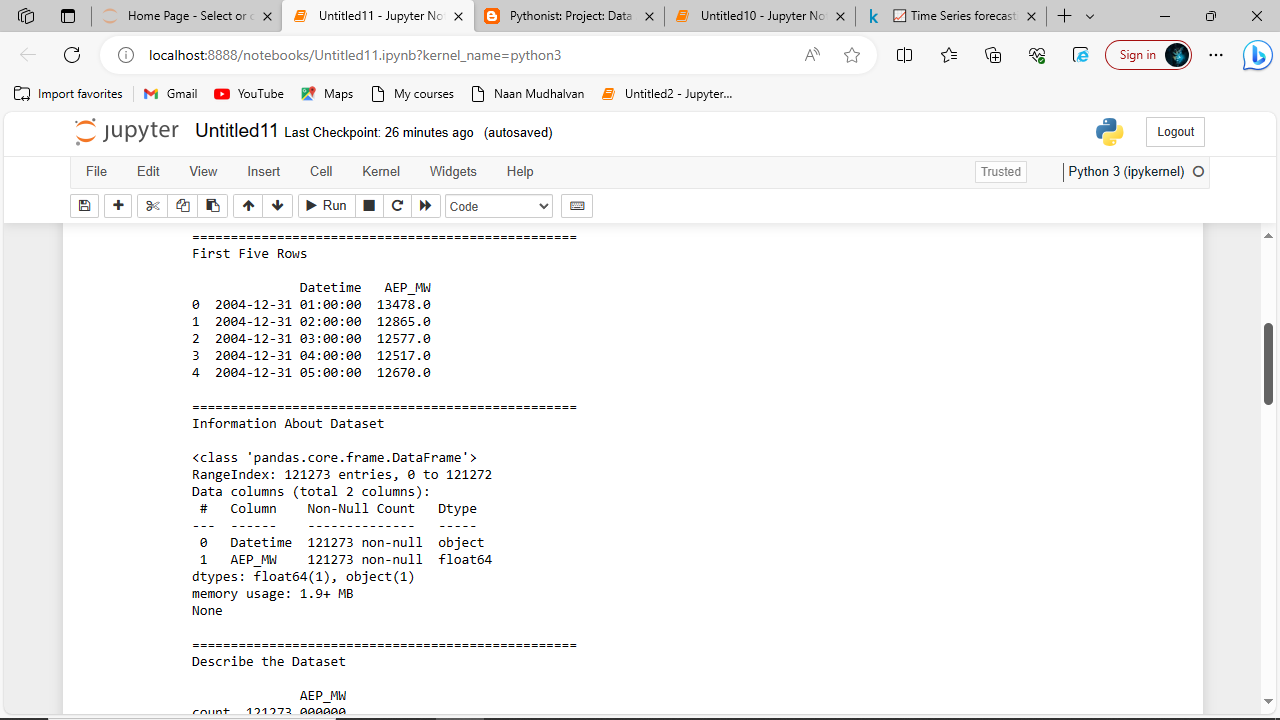
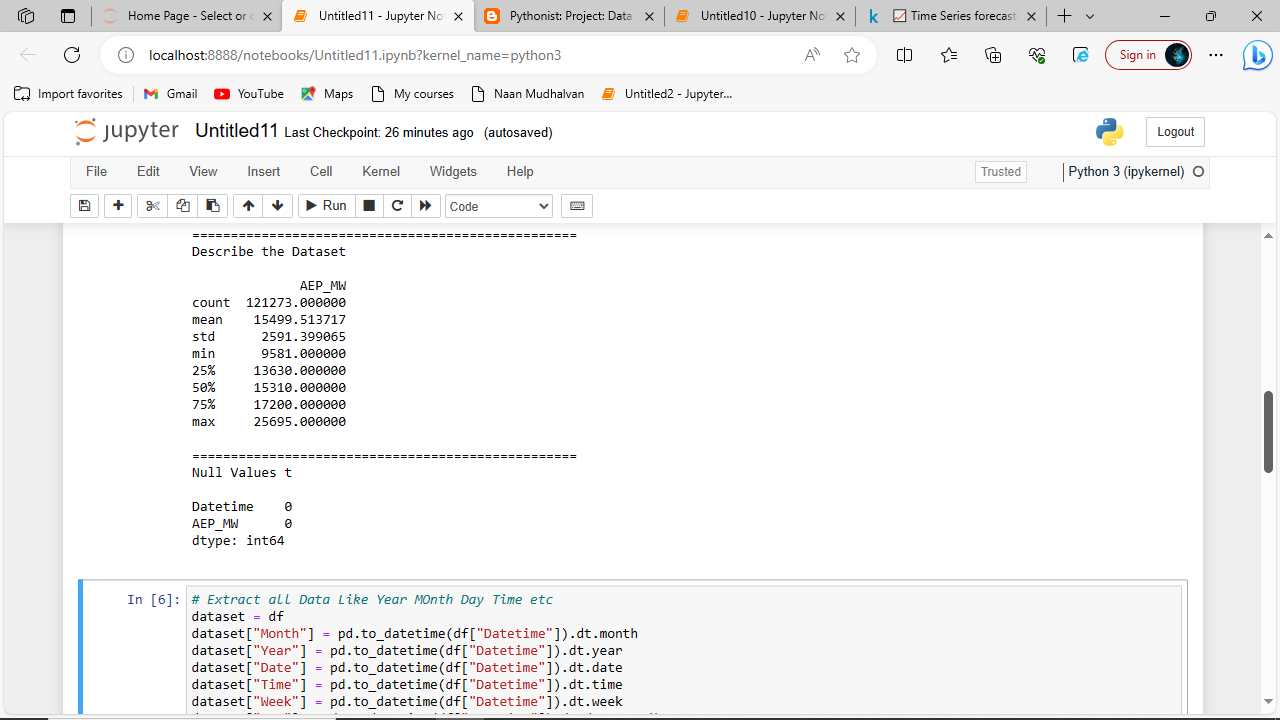
print(df.describe (),"**\n**")

print("="\*50)

print("Null Values t ","**\n**")

print(df.isnull ().sum (),"**\n**")

****

**Output :**

**STEP 2:**

**Reformat the Date Time Columns**

**Code :**

*# Extract all Data Like Year MOnth Day Time etc*

dataset = df

dataset["Month"] = pd.to\_datetime(df["Datetime"]).dt.month

dataset["Year"] = pd.to\_datetime(df["Datetime"]).dt.year

dataset["Date"] = pd.to\_datetime(df["Datetime"]).dt.date

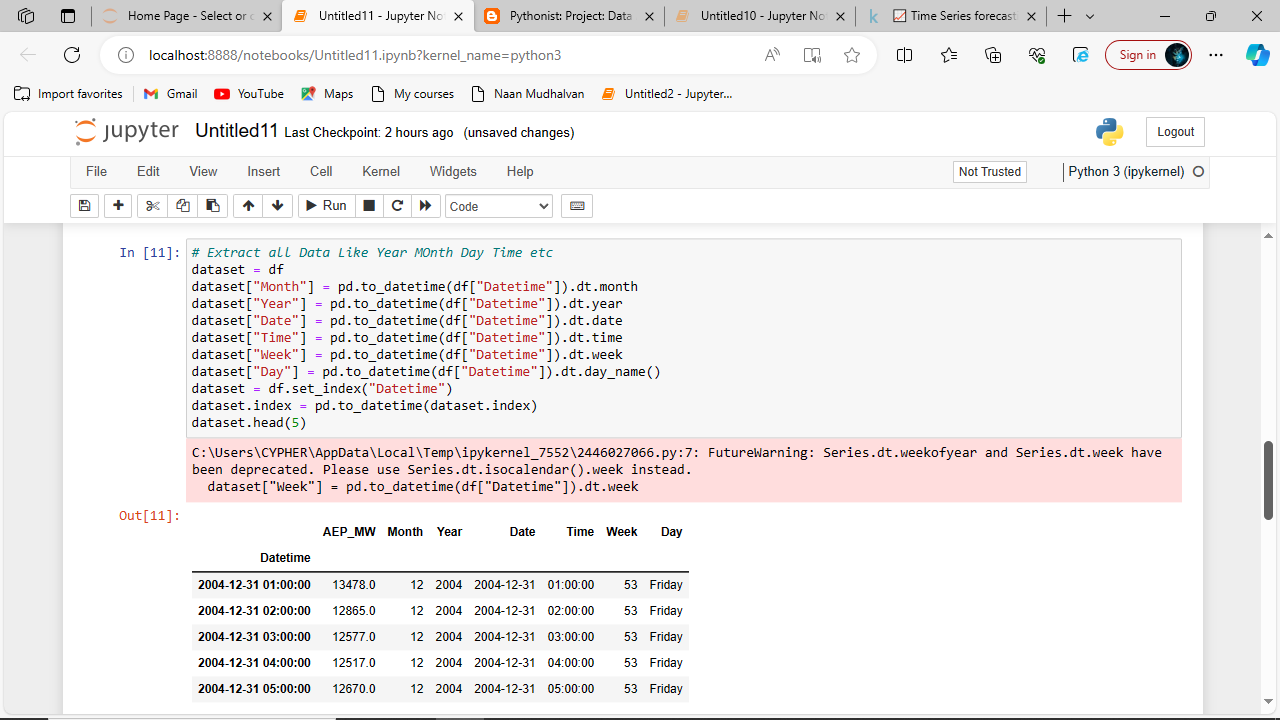
dataset["Time"] = pd.to\_datetime(df["Datetime"]).dt.time

dataset["Week"] = pd.to\_datetime(df["Datetime"]).dt.week

dataset["Day"] = pd.to\_datetime(df["Datetime"]).dt.day\_name()dataset = df.set\_index("Datetime")

dataset.index = pd.to\_datetime(dataset.index)

dataset.head(5)

**Output :**

**STEP 3:**

**Show the Energy Consumption Each Year**

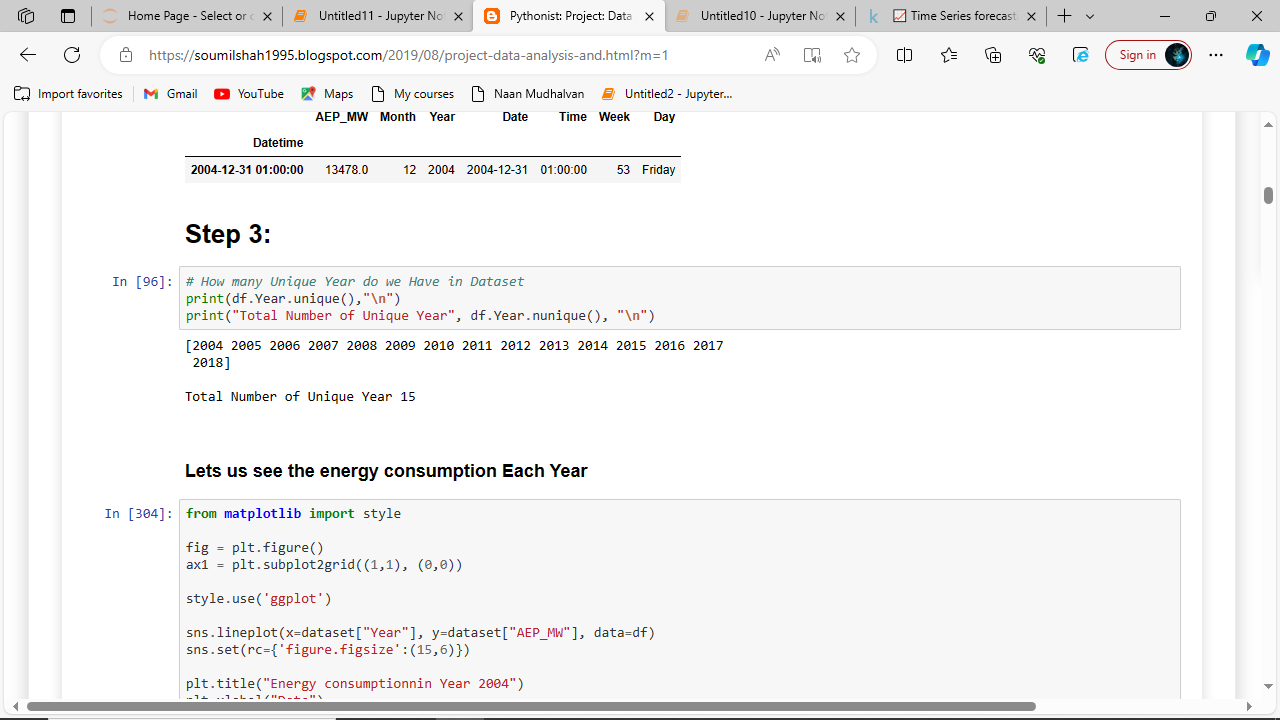
**Code :**

*# How many Unique Year do we Have in Dataset*

print(df.Year.unique(),"**\n**")

print("Total Number of Unique Year", df.Year.nunique(), "**\n**")

**Output :**

**Code :**

**from** **matplotlib** **import** style

fig = plt.figure()

ax1 = plt.subplot2grid((1,1), (0,0))

style.use('ggplot')

sns.lineplot(x=dataset["Year"], y=dataset["AEP\_MW"], data=df)sns.set(rc={'figure.figsize':(15,6)})

plt.title("Energy consumptionnin Year 2004")

plt.xlabel("Date")

plt.ylabel("Energy in MW")

plt.grid(**True**)

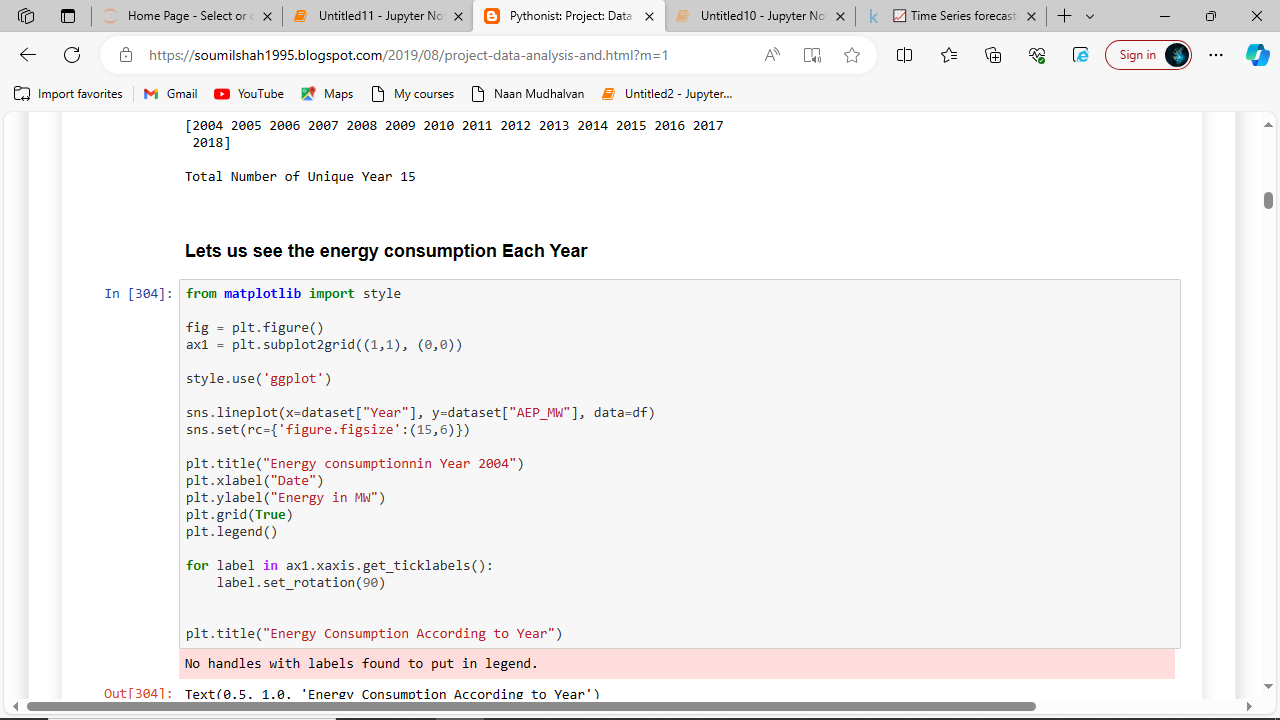
plt.legend()

**for** label **in** ax1.xaxis.get\_ticklabels():

label.set\_rotation(90)

plt.title("Energy Consumption According to Year")

**Output :**

****

**Code :**

**from** **matplotlib** **import** style

fig = plt.figure()

ax1= fig.add\_subplot(311)

ax2= fig.add\_subplot(312)

ax3= fig.add\_subplot(313)

style.use('ggplot')

y\_2004 = dataset["2004"]["AEP\_MW"].to\_list()

x\_2004 = dataset["2004"]["Date"].to\_list()

ax1.plot(x\_2004,y\_2004, color="green", linewidth=1.7)

y\_2005 = dataset["2005"]["AEP\_MW"].to\_list()

x\_2005 = dataset["2005"]["Date"].to\_list()

ax2.plot(x\_2005, y\_2005, color="green", linewidth=1)

y\_2006 = dataset["2006"]["AEP\_MW"].to\_list()

x\_2006 = dataset["2006"]["Date"].to\_list()

ax3.plot(x\_2006, y\_2006, color="green", linewidth=1)

plt.rcParams["figure.figsize"] = (18,8)

plt.title("Energy consumptionnin")

plt.xlabel("Date")

plt.ylabel("Energy in MW")

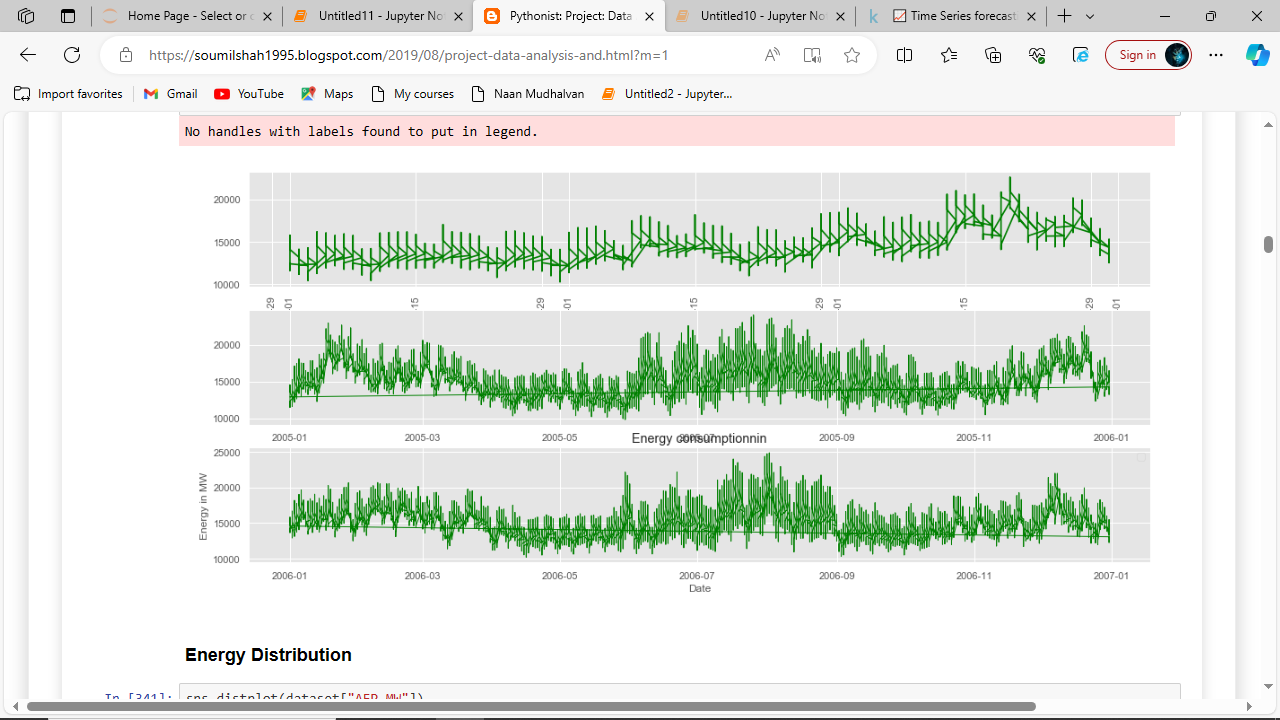
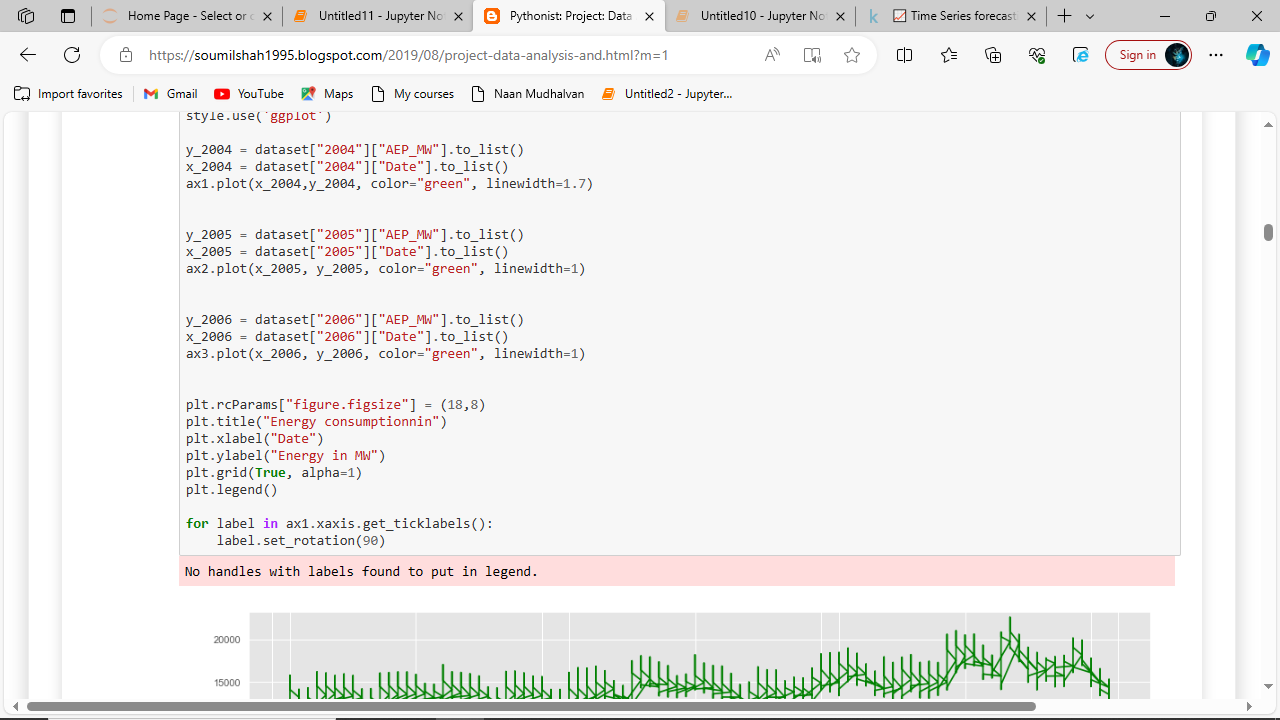
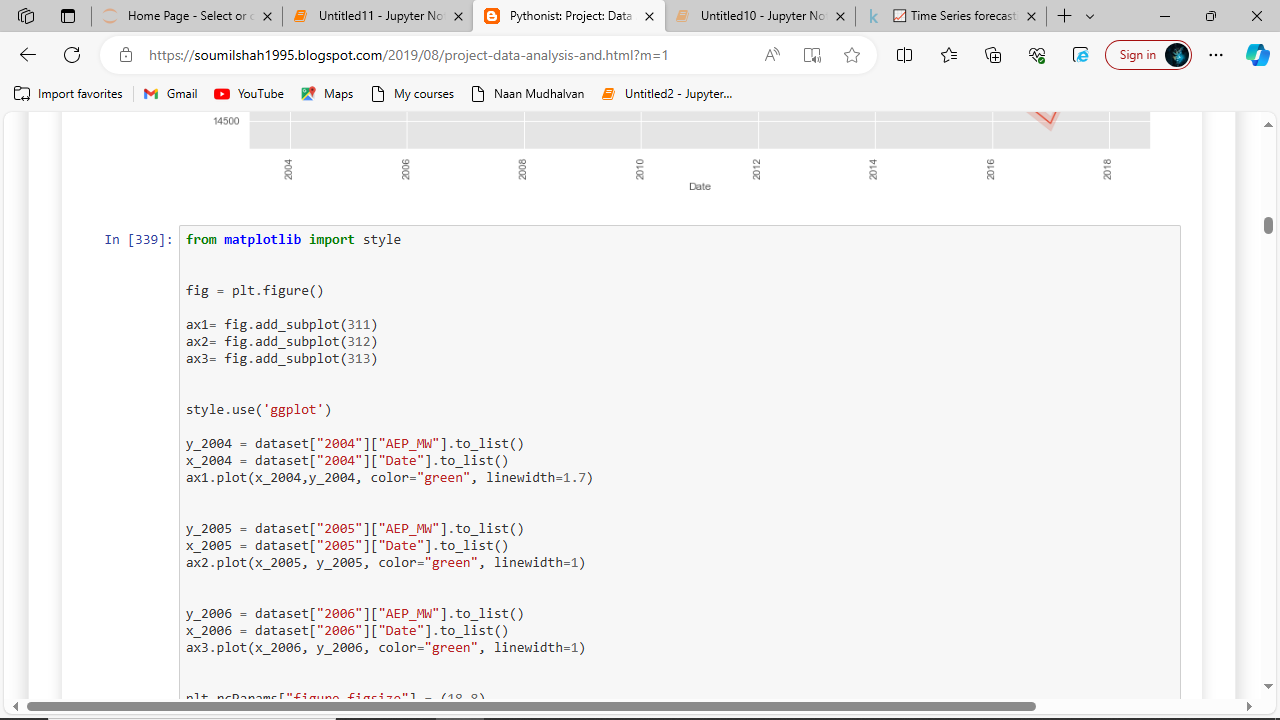
plt.grid(**True**, alpha=1)

plt.legend()

**for** label **in** ax1.xaxis.get\_ticklabels():

label.set\_rotation(90)

**Output :**

****

**STEP 4:**

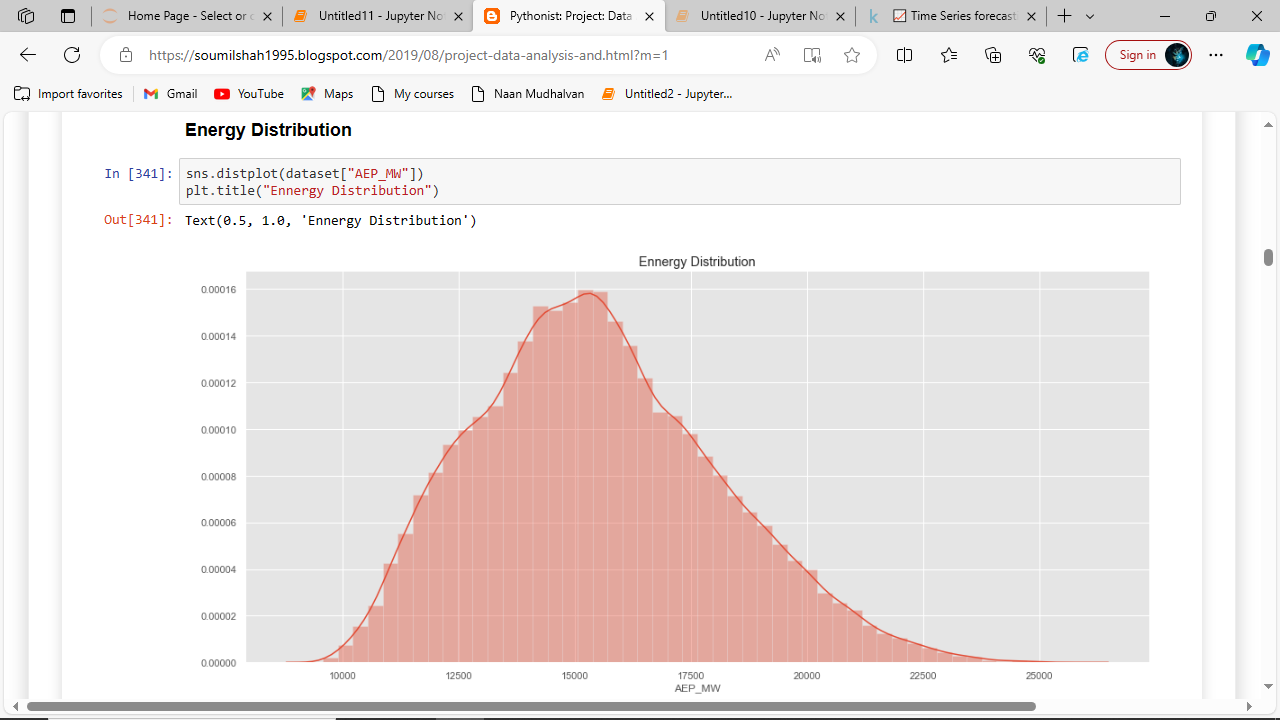
**Energy with Respect To Time**

**Code :**

sns.distplot(dataset["AEP\_MW"])

plt.title("Energy Distribution")

**Output :**

****

**Code :**

fig = plt.figure()

ax1= fig.add\_subplot(111)

sns.lineplot(x=dataset["Time"],y=dataset["AEP\_MW"], data=df)plt.title("Energy Consumption vs Time ")

plt.xlabel("Time")

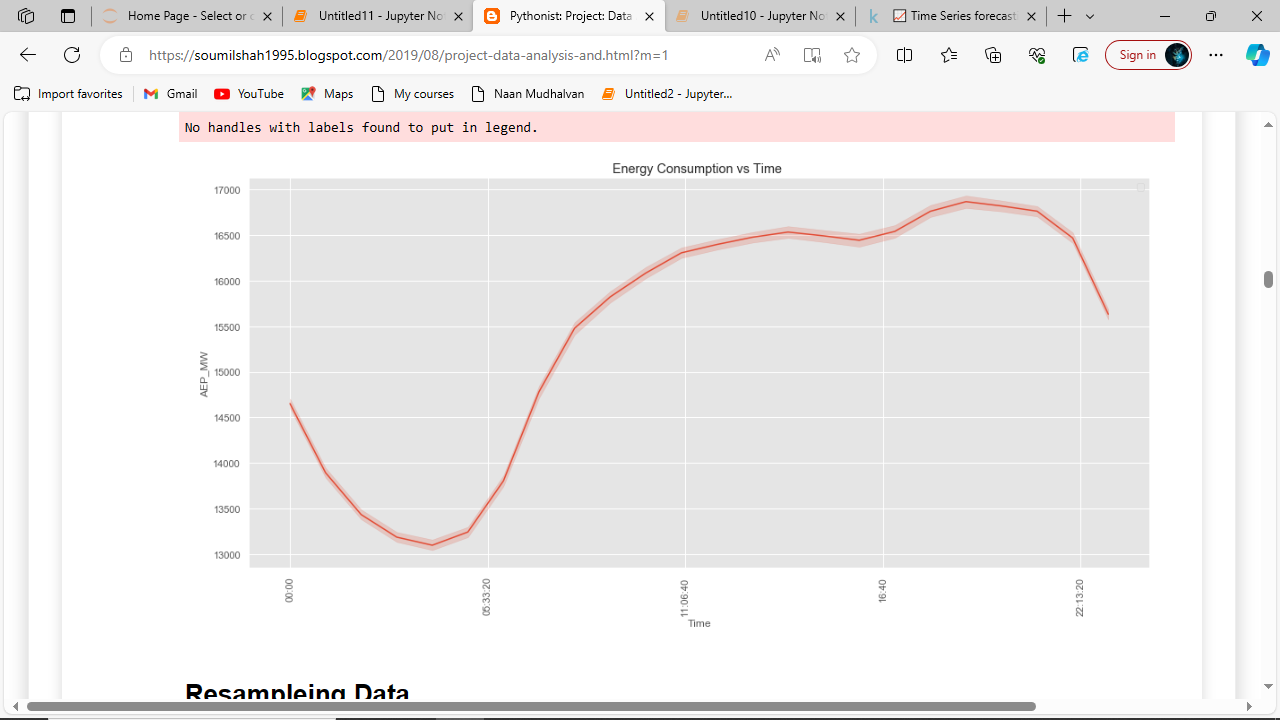
plt.grid(**True**, alpha=1)

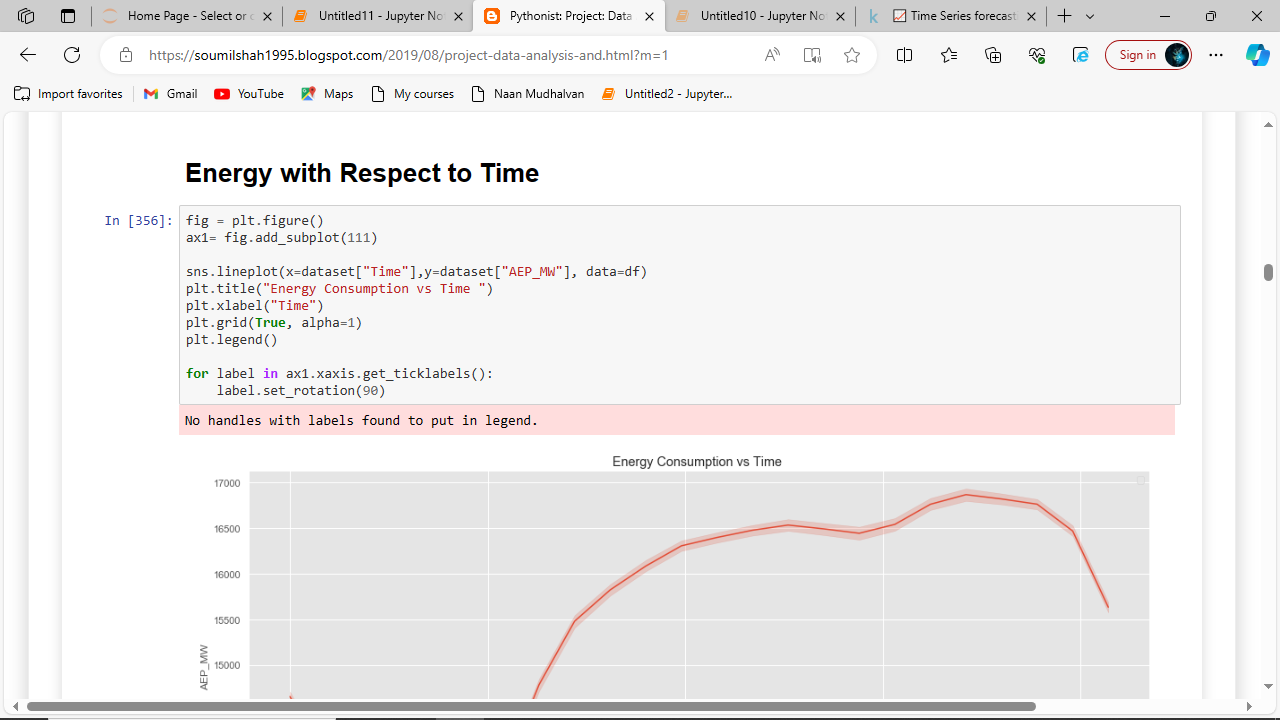
plt.legend()

**for** label **in** ax1.xaxis.get\_ticklabels():

label.set\_rotation(90)

**Output :**

****

****