Data Cleaning Project

import pandas as pd

import numpy as np

df=pd.read\_csv('Pfizer\_1.csv')

df.columns

df\_melt=pd.melt(df,id\_vars=['Date','Drug\_Name','Parameter'],var\_name='Time',value\_name='Readings')

df\_pivot=df\_melt.pivot(index=['Date','Drug\_Name','Time'],columns='Parameter',values='Readings')

df\_pivot=df\_pivot.reset\_index()

def avg\_temp(x):

  x['Average\_Temperature']=x['Temperature'].mean()

  return x

df\_final=df\_pivot.groupby(['Drug\_Name']).apply(avg\_temp)

def avg\_pressure(y):

  y['Average\_Pressure']=y['Pressure'].mean()

  return y

#Ther are 26 NaN values before data cleaning

df\_final['Temperature'].fillna(df\_final['Average\_Temperature'],inplace=True)

df\_final['Pressure'].fillna(df\_final['Average\_Pressure'],inplace=True)

Parameter Date 0 Drug\_Name 0 Time 0 Pressure 0 Temperature 0 Average\_Temperature 0 Average\_Pressure 0