import pandas as pd

air = pd.read\_csv(‘airquality.csv’)

air

air.shape

air.head(10)

air.count()

air.isnull().sum()

air.info()

air.describe()

a = air.dropna()

a.shape

a = air.fillna(0)

a.shape

a.head(10)

a = air.fillna(method=’pad’)

a.head(10)

a = air.fillna(method=’backfill’)

a.head(10)

import numpy as np

a = air[‘Ozone’].fillna(air[‘Ozone’].mean())

a.head()

a = air[‘Ozone’].fillna(air[‘Ozone’].median())

a.head()

from sklearn.impute import SimpleImputer

imp = SimpleImputer(missing\_values = np.nan, strategy=’mean’)

A = imp.fit\_transform(air)

A

imp = SimpleImputer(missing\_values = np.nan, strategy=’most\_frequent’)

A = imp.fit\_transform(air)

A

A = pd.DataFrame(A, columns = air.colums)

A.head()

from sklearn.model\_selection import train\_test\_split

len(A)

train, test = train\_test\_split(A)

len(train)

len(test)

train.head()

train, test = train\_test\_split(A, test\_size = 0.20)

len(test)

len(train)

A.describe()

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

sc = scaler.fit\_transform(A)

pd.DataFrame(sc).describe()

from sklearn.linear\_model import LinearRegression

X = A[‘Ozone’].values

X = X.reshape(-1, 1)

X

Y = A[‘Temp’]

model = LinearRegression()

model.fit(X,Y)

model.score(X,Y)\*100

import matplotlib.pyplot as plt

plt.scatter(X,Y)