

```
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.columns)
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)
```