```
import pandas as pd
```

```
import numpy as np
```

```
path = "/content/city_hour.csv"
```

```
df = pd.read_csv(path)
```

```
path1 = "/content/city_day.csv"
```

```
df1 = pd.read_csv(path1)
```

```
df
```

	City	Datetime	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2
0	Ahmedabad	2015-01-01 01:00:00	NaN	NaN	1.00	40.01	36.37	NaN	1.00	122.07
1	Ahmedabad	2015-01-01 02:00:00	NaN	NaN	0.02	27.75	19.73	NaN	0.02	85.90
2	Ahmedabad	2015-01-01 03:00:00	NaN	NaN	0.08	19.32	11.08	NaN	0.08	52.83
3	Ahmedabad	2015-01-01 04:00:00	NaN	NaN	0.30	16.45	9.20	NaN	0.30	39.53
4	Ahmedabad	2015-01-01 05:00:00	NaN	NaN	0.12	14.90	7.85	NaN	0.12	32.63
...	...	...	...	...	...	...	...	...	...	...

```
df1
```

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Tolue
0	Ahmedabad	01-01-2015	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	0.00	0.
1	Ahmedabad	02-01-2015	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	3.68	5.
2	Ahmedabad	03-01-2015	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	6.80	16.
3	Ahmedabad	04-01-2015	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4.43	10.
4	Ahmedabad	05-01-2015	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7.01	18.
...	...	...	...	...	...	...	...	...	...	...	...	...	
		27-											

```
df.isna().any()
```

```
City          False
Datetime      False
PM2.5          True
PM10           True
NO             True
NO2            True
NOx            True
NH3            True
CO             True
SO2            True
O3             True
Benzene        True
Toluene        True
Xylene         True
```

```
AQI            True
AQI_Bucket     True
dtype: bool
```

```
df.isnull().sum()
```

```
City               0
Datetime           0
PM2.5         145088
PM10          296737
NO            116632
NO2           117122
NOx           123224
NH3           272542
CO             86517
SO2           130373
O3            129208
Benzene       163646
Toluene       220607
Xylene        455829
AQI           129080
AQI_Bucket    129080
dtype: int64
```

```
#data cleaning
df.any()
```

```
City          True
Datetime      True
PM2.5         True
PM10          True
NO            True
NO2           True
NOx           True
NH3           True
CO            True
SO2           True
O3            True
Benzene       True
Toluene       True
Xylene        True
AQI           True
AQI_Bucket    True
dtype: bool
```

```
df.isnull().sum()
```

```
City               0
Datetime           0
PM2.5         145088
PM10          296737
NO            116632
NO2           117122
NOx           123224
NH3           272542
CO             86517
SO2           130373
O3            129208
Benzene       163646
Toluene       220607
Xylene        455829
AQI           129080
AQI_Bucket    129080
dtype: int64
```

```
df.duplicated().sum()
```

```
0
```

```
temp=df.drop_duplicates(subset=None,keep=False,ignore_index=True)
temp.duplicated().sum()
```

```
0
```

```
df=df.merge(df1)
```

```
df.head()
```

	City	Datetime	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	A
0	Ahmedabad	2015-01-10 03:00:00	NaN	NaN	NaN	NaN	0.0	NaN	NaN	NaN	NaN	0.0	0.0	0.0	Na

```
df.shape
```

```
(13780039, 17)
```

```
df1.shape
```

```
(29531, 16)
```

```
df.isnull().sum()
```

```
City                0
Datetime            0
PM2.5          13779783
PM10           13779993
NO             13779774
NO2            13779397
NOx            13081924
NH3            13779691
CO             10165508
SO2            13779778
O3             13779935
Benzene        10315808
Toluene        10315838
Xylene         10329155
AQI            13779727
AQI_Bucket     13779727
Date                0
dtype: int64
```

```
df.isna()
```

	City	Datetime	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene
0	False	False	True	True	True	True	False	True	True	True	True	False	False
1	False	False	True	True	True	True	False	True	True	True	True	False	False
2	False	False	True	True	True	True	False	True	True	True	True	False	False
3	False	False	True	True	True	True	False	True	True	True	True	False	False
4	False	False	True	True	True	True	False	True	True	True	True	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...
13780034	False	False	True	True	True	True	True	True	True	True	True	True	True
13780035	False	False	True	True	True	True	True	True	True	True	True	True	True
13780036	False	False	True	True	True	True	True	True	True	True	True	True	True
13780037	False	False	True	True	True	True	True	True	True	True	True	True	True
13780038	False	False	False	False	False	False	False	False	False	False	False	True	True

13780039 rows × 17 columns

```
df.isnull().sum()
```

```
City                0
Datetime            0
PM2.5          13779783
PM10           13779993
NO             13779774
NO2            13779397
NOx            13081924
NH3            13779691
CO             10165508
SO2            13779778
O3             13779935
Benzene        10315808
Toluene        10315838
Xylene         10329155
AQI            13779727
AQI_Bucket     13779727
Date                0
dtype: int64
```

```
df.shape
```

```
(13780039, 17)
```

```
df.isnull().sum()
```

```
City                13
Datetime             0
PM2.5         13779783
PM10          13779993
NO            13779774
NO2           13779397
NOx           13081924
NH3           13779691
CO            10165508
SO2           13779778
O3            13779935
Benzene       10315808
Toluene       10315838
Xylene        10329155
AQI           13779727
AQI_Bucket    13779727
Date                 0
dtype: int64
```

```
df.dropna()
```

	City	Datetime	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	To
6025042	Amritsar	2020-07-01 00:00:00	57.67	100.99	32.81	15.11	30.20	17.73	0.59	3.48	16.48	1.30	
6656116	Chandigarh	2020-07-01 00:00:00	32.90	72.38	0.56	9.87	5.92	36.63	0.33	14.91	34.33	3.31	
6672816	Delhi	2020-07-01 00:00:00	54.01	128.66	6.33	21.05	16.81	29.06	0.97	11.15	29.73	2.03	
6850209	Gurugram	2020-07-01	61.64	174.08	5.09	19.58	13.27	2.91	1.11	7.57	44.14	2.67	

```
df.isnull().sum()
```

```
City                13
Datetime             0
PM2.5         13779783
PM10          13779993
NO            13779774
NO2           13779397
NOx           13081924
NH3           13779691
CO            10165508
SO2           13779778
O3            13779935
Benzene       10315808
Toluene       10315838
Xylene        10329155
AQI           13779727
AQI_Bucket    13779727
Date                 0
dtype: int64
```

```
df1.isnull().sum()
```

```
City             0
Date             0
PM2.5         4598
PM10         11140
NO            3582
NO2           3585
NOx           4185
NH3          10328
CO            2059
SO2           3854
O3            4022
Benzene       5623
Toluene       8041
Xylene       18109
AQI           4681
AQI_Bucket    4681
dtype: int64
```

```
df.dropna()
```

	City	Datetime	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	To
6025042	Amritsar	2020-07-01 00:00:00	57.67	100.99	32.81	15.11	30.20	17.73	0.59	3.48	16.48	1.30	
6656116	Chandigarh	2020-07-01 00:00:00	32.90	72.38	0.56	9.87	5.92	36.63	0.33	14.91	34.33	3.31	
6672816	Delhi	2020-07-01 00:00:00	54.01	128.66	6.33	21.05	16.81	29.06	0.97	11.15	29.73	2.03	

```
df.isnull().sum()
```

```
City                0
Datetime            0
PM2.5         13779783
PM10          13779993
NO            13779774
NO2           13779397
NOx           13081924
NH3           13779691
CO            10165508
SO2           13779778
O3            13779935
Benzene       10315808
Toluene       10315838
Xylene        10329155
AQI           13779727
AQI_Bucket    13779727
Date                0
dtype: int64
```

```
//log, square root, cube root
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
data_log = np.sqrt(data)
```

```
axs[0].hist(data, edgecolor='black')
axs[1].hist(data_log, edgecolor='black')
```

```
(array([15., 37., 45., 53., 51., 50., 24., 16.,  6.,  3.]),
 array([0.01861567, 0.1054491 , 0.19228253, 0.27911596, 0.36594939,
        0.45278282, 0.53961625, 0.62644968, 0.71328312, 0.80011655,
        0.88694998]),
 <BarContainer object of 10 artists>)
```

```
axs[0].set_title('Original Data')
axs[1].set_title('Square Transformed Data')
```

```
Text(0.5, 1.0, 'Square Transformed Data')
```

```
np.log(data)
```

```
array([-2.20998732, -2.8409356 , -2.6980004 , -1.67829818, -4.37724684,
       -0.94348954, -1.98882615, -3.56012945, -1.32515317, -1.92022742,
       -6.11817214, -1.40321266, -3.12466348, -1.44044308, -2.85840551,
       -1.5623299 , -0.68382208, -1.61543261, -1.86268483, -1.49728708,
       -3.56645563, -1.11785188, -0.91881767, -1.48768416, -2.54958407,
       -2.27724699, -4.53880223, -2.64617888, -1.78107988, -3.73914211,
       -2.82936484, -0.87083791, -1.96694937, -2.43289769, -1.23579847,
       -2.3351472 , -1.90974045, -0.9742796 , -1.50367411, -1.17604047,
       -2.14635126, -1.6921895 , -1.34068862, -1.79370813, -2.3441471 ,
       -1.51749849, -2.25079156, -3.86236585, -1.53626923, -3.61432221,
       -0.68020863, -2.09679006, -2.00353509, -1.65416265, -2.78155695,
       -1.1297571 , -3.78700205, -1.54467027, -2.53663772, -2.56809719,
       -2.39233274, -4.34524735, -2.18094853, -5.85376324, -2.26138824,
       -0.23993339, -1.36705126, -2.248029  , -1.59295205, -3.43547892,
       -1.26462785, -2.33026974, -2.08993835, -2.20186351, -1.22126966,
       -1.16243158, -1.75481288, -2.32524082, -1.5769426 , -2.0126293 ,
       -0.62311649, -5.25081961, -1.38352926, -4.66232511, -0.7130207 ,
       -3.26981305, -1.12347947, -3.1518734 , -1.60139925, -0.67951052,
       -1.32608415, -3.13957409, -0.65292568, -2.74064612, -3.63791614,
       -1.9870483 , -0.90827581, -4.815991  , -3.5396552 , -2.50342768,
       -3.61216591, -3.10259975, -1.51634458, -1.53679214, -1.3738105 ,
       -1.53037908, -1.67651472, -1.75607364, -2.60693995, -2.69502954,
```

```
            -2.32375722, -1.48929684, -1.43732044, -1.8259452 , -3.07101023,
            -1.95002283, -1.42236678, -1.29243413, -1.45916214, -3.76874392,
            -2.75729896, -5.81797774, -1.4089723 , -1.57143593, -5.34441291,
            -2.92658723, -3.66125299, -2.31984016, -1.82925158, -1.39725461,
            -2.31717176, -2.29864882, -2.29190609, -1.97202428, -1.35057227,
            -0.71599472, -2.96864982, -1.12510013, -2.04708981, -2.21579931,
            -2.5140061 , -1.83592778, -2.03687613, -0.8790403 , -3.78386131,
            -3.31209036, -1.57287118, -1.22853652, -1.78426551, -2.01362731,
            -2.79428939, -1.41801192, -0.90414953, -1.44969076, -1.39168984,
            -1.87679223, -2.14735783, -2.69702838, -5.74857658, -2.2943435 ,
            -2.25855309, -7.96750343, -1.98779887, -1.62580854, -2.44012884,
            -1.93035314, -1.54356847, -1.66984837, -1.76751743, -2.35500734,
            -1.46317569, -3.26342777, -1.87906098, -2.25972755, -4.86506757,
            -1.79419779, -2.3836143 , -2.352501  , -2.1754623 , -1.72948029,
            -2.61787298, -3.38775948, -1.40152606, -0.75364573, -1.02742671,
            -1.37754089, -3.83831729, -3.23924903, -0.26313296, -4.36171348,
            -1.15454549, -1.8623814 , -1.54642036, -4.87138742, -3.23282897,
            -3.85724625, -0.76592347, -1.10952324, -2.06629322, -2.8493753 ,
            -2.75400539, -4.39779276, -2.35170566, -1.53696563, -1.17643532,
            -2.14075366, -4.3608602 , -3.71876026, -2.76967351, -6.00947573,
            -2.99920146, -3.85540468, -4.15253685, -1.29508876, -1.63350137,
            -3.53396058, -0.51870566, -1.30197116, -0.66319936, -1.29138269,
            -2.80828874, -1.69659788, -2.41722848, -0.65431926, -1.61224106,
            -2.20486378, -1.20748753, -0.91351713, -1.66956123, -1.99832795,
            -3.61673492, -2.49409189, -0.31742202, -2.38169082, -3.51293619,
            -3.64859443, -2.78244726, -2.94032242, -1.84151497, -1.02488415,
            -1.07413095, -2.55817415, -1.88277318, -1.10653045, -2.17911103,
            -1.95647915, -4.27700227, -1.14216281, -3.1598393 , -4.01250958,
            -2.6145129 , -4.97527583, -2.76601322, -1.6621946 , -2.2575241 ,
            -2.131749  , -1.78600452, -1.53666871, -1.8411557 , -0.76087372,
            -0.55244727, -2.80612622, -3.61360928, -2.31013952, -1.56126483,
            -1.16863255, -1.75517778, -2.79599468, -1.47822163, -2.85369344,
            -1.35369609, -1.1444606 , -1.64230286, -4.18436174, -3.28043541,
            -4.4798374 , -2.92830261, -2.69105963, -1.74629356, -1.05909753,
            -4.97317911, -4.36556433, -3.33950122, -1.19620677, -1.38204653,
            -2.18202992, -1.73681569, -0.69522787, -2.1226223 , -2.21009168,
```

```
import numpy as np
```

```
from sklearn.linear_model import LinearRegression
```

```
x = np.array([5, 15, 25, 35, 45, 55]).reshape((-1, 1))
```

```
y = np.array([5, 20, 14, 32, 22, 38])
```

```
x
```

```
array([[ 5],
       [15],
       [25],
       [35],
       [45],
       [55]])
```

```
y
```

```
array([ 5, 20, 14, 32, 22, 38])
```

```
model = LinearRegression()
```

```
model.fit(x, y)
```

```
▾ LinearRegression
LinearRegression()
```

```
model = LinearRegression().fit(x, y)
```

```
r_sq = model.score(x, y)
```

```
print(f"coefficient of determination: {r_sq}")
```

```
coefficient of determination: 0.7158756137479542
```