In [24]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [25]:
```
```

In [26]:
```
dftrain=pd.read_csv(r"C:\Users\user\Downloads\C10_air\csvs_per_year\csvs_per_y
```

Out[26]:

	date	BEN	CO	EBE	MXY	NMHC	NO_2	NOx	OXY	O_3	
0	2001-08-01 01:00:00	NaN	0.37	NaN	NaN	NaN	58.400002	87.150002	NaN	34.529999	10
1	2001-08-01 01:00:00	1.50	0.34	1.49	4.10	0.07	56.250000	75.169998	2.11	42.160000	10
2	2001-08-01 01:00:00	NaN	0.28	NaN	NaN	NaN	50.660000	61.380001	NaN	46.310001	10
3	2001-08-01 01:00:00	NaN	0.47	NaN	NaN	NaN	69.790001	73.449997	NaN	40.650002	6
4	2001-08-01 01:00:00	NaN	0.39	NaN	NaN	NaN	22.830000	24.799999	NaN	66.309998	7
...	...	...	...	...	...	...	...	...	...	...	
217867	2001-04-01 00:00:00	10.45	1.81	NaN	NaN	NaN	73.000000	264.399994	NaN	5.200000	4
217868	2001-04-01 00:00:00	5.20	0.69	4.56	NaN	0.13	71.080002	129.300003	NaN	13.460000	2
217869	2001-04-01 00:00:00	0.49	1.09	NaN	1.00	0.19	76.279999	128.399994	0.35	5.020000	4
217870	2001-04-01 00:00:00	5.62	1.01	5.04	11.38	NaN	80.019997	197.000000	2.58	5.840000	3
217871	2001-04-01 00:00:00	8.09	1.62	6.66	13.04	0.18	76.809998	206.300003	5.20	8.340000	3

217872 rows × 16 columns

In [27]:
```
```

Out[27]:
```
Index(['date', 'BEN', 'CO', 'EBE', 'MXY', 'NMHC', 'NO_2', 'NOx', 'OXY', 'O_3
',
       'PM10', 'PXY', 'SO_2', 'TCH', 'TOL', 'station'],
      dtype='object')
```

In [28]:
```
a=dftrain[['CO','NO_2','NOx','O_3','PM10','station']]
```

Out[28]:

	CO	NO_2	NOx	O_3	PM10	station
0	0.37	58.400002	87.150002	34.529999	105.000000	28079001
1	0.34	56.250000	75.169998	42.160000	100.599998	28079035
2	0.28	50.660000	61.380001	46.310001	100.099998	28079003
3	0.47	69.790001	73.449997	40.650002	69.779999	28079004
4	0.39	22.830000	24.799999	66.309998	75.180000	28079039
...	...	...	...	...	...	...
217867	1.81	73.000000	264.399994	5.200000	47.880001	28079022
217868	0.69	71.080002	129.300003	13.460000	26.809999	28079023
217869	1.09	76.279999	128.399994	5.020000	40.770000	28079024
217870	1.01	80.019997	197.000000	5.840000	37.889999	28079025
217871	1.62	76.809998	206.300003	8.340000	35.369999	28079099

217872 rows × 6 columns

In [29]:
```
b=dftrain.head(10)
```

Out[29]:

	date	BEN	CO	EBE	MXY	NMHC	NO_2	NOx	OXY	O_3	PM1
0	2001-08-01 01:00:00	NaN	0.37	NaN	NaN	NaN	58.400002	87.150002	NaN	34.529999	105.00000
1	2001-08-01 01:00:00	1.50	0.34	1.49	4.10	0.07	56.250000	75.169998	2.11	42.160000	100.59999
2	2001-08-01 01:00:00	NaN	0.28	NaN	NaN	NaN	50.660000	61.380001	NaN	46.310001	100.09999
3	2001-08-01 01:00:00	NaN	0.47	NaN	NaN	NaN	69.790001	73.449997	NaN	40.650002	69.77999
4	2001-08-01 01:00:00	NaN	0.39	NaN	NaN	NaN	22.830000	24.799999	NaN	66.309998	75.18000
5	2001-08-01 01:00:00	2.11	0.63	2.48	5.94	0.05	66.260002	118.099998	3.15	33.500000	122.69999
6	2001-08-01 01:00:00	NaN	0.28	NaN	NaN	NaN	35.799999	39.590000	NaN	68.250000	124.90000
7	2001-08-01 01:00:00	NaN	0.67	NaN	NaN	NaN	74.830002	112.000000	NaN	26.410000	113.00000
8	2001-08-01 01:00:00	NaN	0.41	NaN	NaN	NaN	33.209999	37.299999	NaN	62.299999	125.30000
9	2001-08-01 01:00:00	NaN	0.17	NaN	NaN	0.13	24.129999	36.970001	NaN	46.200001	95.58999

In [30]:
```
a=b[['CO','NO_2','NOx','O_3','PM10','station']]
```

Out[30]:

	CO	NO_2	NOx	O_3	PM10	station
0	0.37	58.400002	87.150002	34.529999	105.000000	28079001
1	0.34	56.250000	75.169998	42.160000	100.599998	28079035
2	0.28	50.660000	61.380001	46.310001	100.099998	28079003
3	0.47	69.790001	73.449997	40.650002	69.779999	28079004
4	0.39	22.830000	24.799999	66.309998	75.180000	28079039
5	0.63	66.260002	118.099998	33.500000	122.699997	28079006
6	0.28	35.799999	39.590000	68.250000	124.900002	28079007
7	0.67	74.830002	112.000000	26.410000	113.000000	28079009
8	0.41	33.209999	37.299999	62.299999	125.300003	28079038
9	0.17	24.129999	36.970001	46.200001	95.589996	28079011

In [37]:
```
c=a.iloc[:,0:6]
```

In [38]:

Out[38]:
```
(10, 6)
```

In [39]:

Out[39]:
```
(10,)
```

In [40]:

In [41]:

In [42]:
```
logr=LogisticRegression()
```

Out[42]:
```
LogisticRegression()
```

In [45]:

In [46]:
```
prediction=logr.predict(observation)
```

Out[46]:
```
array([28079038], dtype=int64)
```

In [47]:

Out[47]:
```
array([28079001, 28079003, 28079004, 28079006, 28079007, 28079009,
       28079011, 28079035, 28079038, 28079039], dtype=int64)
```

In [48]:

Out[48]:
```
1.5371127557724223e-13
```

In [ ]: