Analysis of some data sets

In [1]:

```
import os
os.getcwd
```

Out[1]:

```
<function nt.getcwd()>
```

In [2]:

```
import pandas as pd
df =pd.read_csv(r'C:\Users\PC-chetan\Downloads\1. Weather Data.csv')
df
```

Out[2]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog
2	1/1/2012 2:00	-1.8	-3.4	89	7	4.0	101.26	Freezing Drizzle,Fog
3	1/1/2012 3:00	-1.5	-3.2	88	6	4.0	101.27	Freezing Drizzle,Fog
4	1/1/2012 4:00	-1.5	-3.3	88	7	4.8	101.23	Fog
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/31/2012 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

8784 rows × 8 columns

In [3]:

```
df[['Temp_C','Rel Hum_%']].mean()
```

Out[3]:

```
Temp_C        8.798144
Rel Hum_%    67.431694
dtype: float64
```

In [4]:

```
df.columns
```

Out[4]:

```
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%',
       'Wind Speed_km/h', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')
```

In [5]:

```
df.Weather.nunique()
```

Loading [MathJax]/extensions/Safe.js

Out[5]:
```
50
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8784 entries, 0 to 8783
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Date/Time         8784 non-null   object 
 1   Temp_C            8784 non-null   float64
 2   Dew Point Temp_C  8784 non-null   float64
 3   Rel Hum_%         8784 non-null   int64  
 4   Wind Speed_km/h   8784 non-null   int64  
 5   Visibility_km     8784 non-null   float64
 6   Press_kPa         8784 non-null   float64
 7   Weather           8784 non-null   object 
dtypes: float64(4), int64(2), object(2)
memory usage: 549.1+ KB
```

In [7]:
```
df.notnull().sum()
```

Out[7]:
```
Date/Time           8784
Temp_C              8784
Dew Point Temp_C    8784
Rel Hum_%           8784
Wind Speed_km/h     8784
Visibility_km       8784
Press_kPa           8784
Weather             8784
dtype: int64
```

In [8]:
```
df.columns
```

Out[8]:
```
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%',
       'Wind Speed_km/h', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')
```

In [9]:
```
df['Wind Speed_km/h'].unique()
```

Out[9]:
```
array([ 4,  7,  6,  9, 15, 13, 20, 22, 19, 24, 30, 35, 39, 32, 33, 26, 44,
       43, 48, 37, 28, 17, 11,  0, 83, 70, 57, 46, 41, 52, 50, 63, 54,  2],
      dtype=int64)
```

In [10]:
```
df['Weather'][df['Weather'] == 'Clear'].shape
```

Out[10]:
```
(1326,)
```

In [11]:
```
df.Weather.value_counts()
```

Out[11]:
```
Mainly Clear                               2106
Mostly Cloudy                              2069
Cloudy                                     1728
Clear                                      1326
Snow                                        390
Rain                                        306
Rain Showers                                188
Fog                                         150
Rain,Fog                                    116
```

Loading [MathJax]/extensions/Safe.js

```
Drizzle,Fog                                 80
Snow Showers                                60
Drizzle                                     41
Snow,Fog                                    37
Snow,Blowing Snow                           19
Rain,Snow                                   18
Thunderstorms,Rain Showers                  16
Haze                                        16
Drizzle,Snow,Fog                            15
Freezing Rain                               14
Freezing Drizzle,Snow                       11
Freezing Drizzle                             7
Snow,Ice Pellets                             6
Freezing Drizzle,Fog                         6
Snow,Haze                                    5
Freezing Fog                                 4
Snow Showers,Fog                             4
Moderate Snow                                4
Rain,Snow,Ice Pellets                        4
Freezing Rain,Fog                            4
Freezing Drizzle,Haze                        3
Rain,Haze                                    3
Thunderstorms,Rain                           3
Thunderstorms,Rain Showers,Fog               3
Freezing Rain,Haze                           2
Drizzle,Snow                                 2
Rain Showers,Snow Showers                    2
Thunderstorms                                2
Moderate Snow,Blowing Snow                   2
Rain Showers,Fog                             1
Thunderstorms,Moderate Rain Showers,Fog      1
Snow Pellets                                 1
Rain,Snow,Fog                                1
Moderate Rain,Fog                            1
Freezing Rain,Ice Pellets,Fog                1
Drizzle,Ice Pellets,Fog                      1
Thunderstorms,Rain,Fog                       1
Rain,Ice Pellets                             1
Rain,Snow Grains                             1
Thunderstorms,Heavy Rain Showers             1
Freezing Rain,Snow Grains                    1
Name: Weather, dtype: int64
```

In [12]:

```
df[df.Weather == 'Clear']
```

Out[12]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	1/5/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	1/5/2012 21:00	-9.0	-14.8	63	13	25.0	100.83	Clear
...	...	...	...	...	...	...	...	...
8646	12/26/2012 6:00	-13.4	-14.8	89	4	25.0	102.47	Clear

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
8698	12/28/2012 10:00	-6.1	-8.6	82	19	24.1	101.27	Clear
8713	12/29/2012 1:00	-11.9	-13.6	87	11	25.0	101.31	Clear
8714	12/29/2012 2:00	-11.8	-13.1	90	13	25.0	101.33	Clear
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Clear

1326 rows × 8 columns

In [13]:

```
df.groupby('Weather').get_group('Clear')
```

Out[13]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	1/5/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	1/5/2012 21:00	-9.0	-14.8	63	13	25.0	100.83	Clear
...	...	...	...	...	...	...	...	...
8646	12/26/2012 6:00	-13.4	-14.8	89	4	25.0	102.47	Clear
8698	12/28/2012 10:00	-6.1	-8.6	82	19	24.1	101.27	Clear
8713	12/29/2012 1:00	-11.9	-13.6	87	11	25.0	101.31	Clear
8714	12/29/2012 2:00	-11.8	-13.1	90	13	25.0	101.33	Clear
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Clear

1326 rows × 8 columns

In [14]:

```
df.columns
```

Out[14]:

```
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%',
       'Wind Speed_km/h', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')
```

In [15]:

```
df[df['Wind Speed_km/h'] == 4]
```

Out[15]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather

Loading [MathJax]/extensions/Safe.js

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog
96	1/5/2012 0:00	-8.8	-11.7	79	4	9.7	100.32	Snow
101	1/5/2012 5:00	-7.0	-9.5	82	4	4.0	100.19	Snow
146	1/7/2012 2:00	-8.1	-11.1	79	4	19.3	100.15	Cloudy
...	...	...	...	...	...	...	...	...
8768	12/31/2012 8:00	-8.6	-10.3	87	4	3.2	101.14	Snow Showers
8769	12/31/2012 9:00	-8.1	-9.6	89	4	2.4	101.09	Snow
8770	12/31/2012 10:00	-7.4	-8.9	89	4	6.4	101.05	Snow,Fog
8772	12/31/2012 12:00	-5.8	-7.5	88	4	12.9	100.78	Snow
8773	12/31/2012 13:00	-4.6	-6.6	86	4	12.9	100.63	Snow

474 rows × 8 columns

In [16]:

```
df.groupby('Wind Speed_km/h').get_group(4)
```

Out[16]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog
96	1/5/2012 0:00	-8.8	-11.7	79	4	9.7	100.32	Snow
101	1/5/2012 5:00	-7.0	-9.5	82	4	4.0	100.19	Snow
146	1/7/2012 2:00	-8.1	-11.1	79	4	19.3	100.15	Cloudy
...	...	...	...	...	...	...	...	...
8768	12/31/2012 8:00	-8.6	-10.3	87	4	3.2	101.14	Snow Showers
8769	12/31/2012 9:00	-8.1	-9.6	89	4	2.4	101.09	Snow
8770	12/31/2012 10:00	-7.4	-8.9	89	4	6.4	101.05	Snow,Fog
8772	12/31/2012 12:00	-5.8	-7.5	88	4	12.9	100.78	Snow
8773	12/31/2012 13:00	-4.6	-6.6	86	4	12.9	100.63	Snow

474 rows × 8 columns

In [17]:

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

Out[17]:

```
Date/Time             0
Temp_C                0
Dew Point Temp_C      0
Rel Hum_%             0
Wind Speed_km/h       0
Visibility_km         0
Press_kPa             0
Weather               0
dtype: int64
```

In [18]:

```
df.rename(columns={'Weather': 'Weather Condition'})
```

Out[18]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog
2	1/1/2012 2:00	-1.8	-3.4	89	7	4.0	101.26	Freezing Drizzle,Fog
3	1/1/2012 3:00	-1.5	-3.2	88	6	4.0	101.27	Freezing Drizzle,Fog
4	1/1/2012 4:00	-1.5	-3.3	88	7	4.8	101.23	Fog
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/31/2012 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

8784 rows × 8 columns

In [ ]:

In [19]:

```
df.columns
```

Out[19]:

```
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%',
       'Wind Speed_km/h', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')
```

In [20]:

```
df['Visibility_km'].mean()
```

Loading [MathJax]/extensions/Safe.js

Out[20]:

```
27.664446721311478
```

In [21]:

```
df['Visibility_km'].std()
```

Out[21]:

```
12.622688245171492
```

In [22]:

```
df['Rel Hum_%'].var()
```

Out[22]:

```
286.24855019850196
```

In [23]:

```
df.Weather.value_counts()
```

Out[23]:

```
Mainly Clear                                2106
Mostly Cloudy                               2069
Cloudy                                      1728
Clear                                       1326
Snow                                         390
Rain                                         306
Rain Showers                                 188
Fog                                          150
Rain,Fog                                     116
Drizzle,Fog                                   80
Snow Showers                                  60
Drizzle                                       41
Snow,Fog                                      37
Snow,Blowing Snow                             19
Rain,Snow                                     18
Thunderstorms,Rain Showers                    16
Haze                                          16
Drizzle,Snow,Fog                              15
Freezing Rain                                 14
Freezing Drizzle,Snow                         11
Freezing Drizzle                               7
Snow,Ice Pellets                               6
Freezing Drizzle,Fog                           6
Snow,Haze                                      5
Freezing Fog                                   4
Snow Showers,Fog                               4
Moderate Snow                                  4
Rain,Snow,Ice Pellets                          4
Freezing Rain,Fog                              4
Freezing Drizzle,Haze                          3
Rain,Haze                                      3
Thunderstorms,Rain                             3
Thunderstorms,Rain Showers,Fog                 3
Freezing Rain,Haze                             2
Drizzle,Snow                                   2
Rain Showers,Snow Showers                      2
Thunderstorms                                  2
Moderate Snow,Blowing Snow                     2
Rain Showers,Fog                               1
Thunderstorms,Moderate Rain Showers,Fog        1
Snow Pellets                                   1
Rain,Snow,Fog                                  1
Moderate Rain,Fog                              1
Freezing Rain,Ice Pellets,Fog                  1
Drizzle,Ice Pellets,Fog                        1
Thunderstorms,Rain,Fog                         1
Rain,Ice Pellets                               1
Rain,Snow Grains                               1
```

Loading [MathJax]/extensions/Safe.js

```
Thunderstorms,Heavy Rain Showers              1
Freezing Rain,Snow Grains                     1
Name: Weather, dtype: int64
```

In [24]:

```
df.groupby('Weather').get_group('Snow')
```

Out[24]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
55	1/3/2012 7:00	-14.0	-19.5	63	19	25.0	100.95	Snow
84	1/4/2012 12:00	-13.7	-21.7	51	11	24.1	101.25	Snow
86	1/4/2012 14:00	-11.3	-19.0	53	7	19.3	100.97	Snow
87	1/4/2012 15:00	-10.2	-16.3	61	11	9.7	100.89	Snow
88	1/4/2012 16:00	-9.4	-15.5	61	13	19.3	100.79	Snow
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/31/2012 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

390 rows × 8 columns

In [25]:

```
df[df['Weather'].str.contains('Snow')]
```

Out[25]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
41	1/2/2012 17:00	-2.1	-9.5	57	22	25.0	99.66	Snow Showers
44	1/2/2012 20:00	-5.6	-13.4	54	24	25.0	100.07	Snow Showers
45	1/2/2012 21:00	-5.8	-12.8	58	26	25.0	100.15	Snow Showers
47	1/2/2012 23:00	-7.4	-14.1	59	17	19.3	100.27	Snow Showers
48	1/3/2012 0:00	-9.0	-16.0	57	28	25.0	100.35	Snow Showers
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/31/2012 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

583 rows × 8 columns

In [26]:

```
df[(df['Wind Speed_km/h'] > 24) & (df['Visibility_km'] >25)]
```

Out[26]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
109	1/5/2012 13:00	-4.4	-9.7	66	26	48.3	100.40	Mainly Clear
111	1/5/2012 15:00	-4.3	-12.0	55	26	48.3	100.52	Mainly Clear
350	1/15/2012 14:00	-16.0	-23.4	53	26	48.3	102.66	Mainly Clear
422	1/18/2012 14:00	-10.3	-17.6	55	28	48.3	101.19	Mainly Clear
423	1/18/2012 15:00	-10.4	-18.0	54	30	48.3	101.32	Mainly Clear
...	...	...	...	...	...	...	...	...
8748	12/30/2012 12:00	-12.2	-15.7	75	26	48.3	100.91	Mostly Cloudy
8749	12/30/2012 13:00	-12.4	-16.2	73	37	48.3	100.92	Mostly Cloudy
8750	12/30/2012 14:00	-11.8	-16.1	70	37	48.3	100.96	Mainly Clear
8751	12/30/2012 15:00	-11.3	-15.6	70	32	48.3	101.05	Mainly Clear
8752	12/30/2012 16:00	-11.4	-15.5	72	26	48.3	101.15	Mainly Clear

232 rows × 8 columns

In [ ]:

```

```

In [27]:

```
df[(df['Wind Speed_km/h'] > 24) & (df['Visibility_km'] == 25)& (df['Press_kPa'] > 100)]
```

Out[27]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
45	1/2/2012 21:00	-5.8	-12.8	58	26	25.0	100.15	Snow Showers
48	1/3/2012 0:00	-9.0	-16.0	57	28	25.0	100.35	Snow Showers

Loading [MathJax]/extensions/Safe.js

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
51	1/3/2012 3:00	-11.3	-18.7	54	33	25.0	100.61	Snow Showers
168	1/8/2012 0:00	0.6	-3.2	76	32	25.0	100.72	Cloudy
169	1/8/2012 1:00	-0.6	-4.6	74	32	25.0	100.80	Mostly Cloudy
...	...	...	...	...	...	...	...	...
8705	12/28/2012 17:00	-8.6	-12.0	76	26	25.0	101.34	Mainly Clear
8753	12/30/2012 17:00	-12.1	-15.8	74	28	25.0	101.26	Mainly Clear
8755	12/30/2012 19:00	-13.4	-16.5	77	26	25.0	101.47	Mainly Clear
8759	12/30/2012 23:00	-12.1	-15.1	78	28	25.0	101.52	Mostly Cloudy
8760	12/31/2012 0:00	-11.1	-14.4	77	26	25.0	101.51	Cloudy

227 rows × 8 columns

In [28]:

```
df.groupby('Weather').mean()
```

Out[28]:

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Weather						
Clear	6.825716	0.089367	64.497738	10.557315	30.153243	101.587443
Cloudy	7.970544	2.375810	69.592593	16.127315	26.625752	100.911441
Drizzle	7.353659	5.504878	88.243902	16.097561	17.931707	100.435366
Drizzle,Fog	8.067500	7.033750	93.275000	11.862500	5.257500	100.786625
Drizzle,Ice Pellets,Fog	0.400000	-0.700000	92.000000	20.000000	4.000000	100.790000
Drizzle,Snow	1.050000	0.150000	93.500000	14.000000	10.500000	100.890000
Drizzle,Snow,Fog	0.693333	0.120000	95.866667	15.533333	5.513333	99.281333
Fog	4.303333	3.159333	92.286667	7.946667	6.248000	101.184067
Freezing Drizzle	-5.657143	-8.000000	83.571429	16.571429	9.200000	100.202857
Freezing Drizzle,Fog	-2.533333	-4.183333	88.500000	17.000000	5.266667	100.441667
Freezing Drizzle,Haze	-5.433333	-8.000000	82.000000	10.333333	2.666667	100.316667
Freezing Drizzle,Snow	-5.109091	-7.072727	86.090909	16.272727	5.872727	100.520909
Freezing Fog	-7.575000	-9.250000	87.750000	4.750000	0.650000	102.320000
Freezing Rain	-3.885714	-6.078571	84.642857	19.214286	8.242857	99.647143
Freezing Rain,Fog	-2.225000	-3.750000	89.500000	15.500000	7.550000	99.945000
Freezing Rain,Haze	-4.900000	-7.450000	82.500000	7.500000	2.400000	100.375000
Freezing Rain,Ice Pellets,Fog	-2.600000	-3.700000	92.000000	28.000000	8.000000	100.950000

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Weather						
Freezing Rain,Snow Grains	-5.000000	-7.300000	84.000000	32.000000	4.800000	98.560000
Haze	-0.200000	-2.975000	81.625000	10.437500	7.831250	101.482500
Mainly Clear	12.558927	4.581671	60.667142	14.144824	34.264862	101.248832
Moderate Rain,Fog	1.700000	0.800000	94.000000	17.000000	6.400000	99.980000
Moderate Snow	-5.525000	-7.250000	87.750000	33.750000	0.750000	100.275000
Moderate Snow,Blowing Snow	-5.450000	-6.500000	92.500000	40.000000	0.600000	100.570000
Mostly Cloudy	10.574287	3.131174	62.102465	15.813920	31.253842	101.025288
Rain	9.786275	7.042810	83.624183	19.254902	18.856536	100.233333
Rain Showers	13.722340	9.187766	75.159574	17.132979	22.816489	100.404043
Rain Showers,Fog	12.800000	12.100000	96.000000	13.000000	6.400000	99.830000
Rain Showers,Snow Showers	2.150000	-1.500000	76.500000	22.500000	21.700000	101.100000
Rain,Fog	8.273276	7.219828	93.189655	14.793103	6.873276	100.500862
Rain,Haze	4.633333	2.066667	83.333333	11.666667	6.700000	100.540000
Rain,Ice Pellets	0.600000	-0.600000	92.000000	24.000000	9.700000	100.120000
Rain,Snow	1.055556	-0.566667	89.000000	28.388889	11.672222	99.951111
Rain,Snow Grains	1.900000	-2.100000	75.000000	26.000000	25.000000	100.600000
Rain,Snow,Fog	0.800000	0.300000	96.000000	9.000000	6.400000	100.730000
Rain,Snow,Ice Pellets	1.100000	-0.175000	91.500000	23.250000	6.000000	100.105000
Snow	-4.524103	-7.623333	79.307692	20.038462	11.171795	100.536103
Snow Pellets	0.700000	-6.400000	59.000000	35.000000	2.400000	99.700000
Snow Showers	-3.506667	-7.866667	72.350000	19.233333	20.158333	100.963500
Snow Showers,Fog	-10.675000	-11.900000	90.750000	13.750000	7.025000	101.292500
Snow,Blowing Snow	-5.410526	-7.621053	84.473684	34.842105	4.105263	99.704737
Snow,Fog	-5.075676	-6.364865	90.675676	17.324324	4.537838	100.688649
Snow,Haze	-4.020000	-6.860000	80.600000	5.000000	4.640000	100.782000
Snow,Ice Pellets	-1.883333	-3.666667	87.666667	23.833333	7.416667	100.548333
Thunderstorms	24.150000	19.750000	77.000000	7.500000	24.550000	100.230000
Thunderstorms,Heavy Rain Showers	10.900000	9.000000	88.000000	9.000000	2.400000	100.260000
Thunderstorms,Moderate Rain Showers,Fog	19.600000	18.500000	93.000000	15.000000	3.200000	100.010000
Thunderstorms,Rain	20.433333	18.533333	89.000000	15.666667	19.833333	100.420000
Thunderstorms,Rain Showers	20.037500	17.618750	86.375000	18.312500	15.893750	100.233750
Thunderstorms,Rain Showers,Fog	21.600000	18.700000	84.000000	19.666667	9.700000	100.063333
Thunderstorms,Rain,Fog	20.600000	18.600000	88.000000	19.000000	4.800000	100.080000

In [29]:

```
df.groupby('Weather').max()
```

Out[29]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kP
Weather							
Clear	9/9/2012 5:00	32.8	20.4	99	33	48.3	103.6
Cloudy	9/9/2012 23:00	30.5	22.6	99	54	48.3	103.6
Drizzle	9/30/2012 3:00	18.8	17.7	96	30	25.0	101.5
Drizzle,Fog	9/30/2012 2:00	19.9	19.1	100	28	9.7	102.0
Drizzle,Ice Pellets,Fog	12/17/2012 9:00	0.4	-0.7	92	20	4.0	100.7
Drizzle,Snow	12/19/2012 18:00	1.2	0.2	95	19	11.3	101.1
Drizzle,Snow,Fog	12/22/2012 3:00	1.1	0.6	98	32	9.7	100.1
Fog	9/22/2012 0:00	20.8	19.6	100	22	9.7	103.0
Freezing Drizzle	2/1/2012 5:00	-2.3	-3.3	93	26	12.9	101.0
Freezing Drizzle,Fog	12/10/2012 5:00	-0.3	-2.3	94	33	8.0	101.2
Freezing Drizzle,Haze	2/1/2012 13:00	-5.0	-7.7	83	11	4.0	100.3
Freezing Drizzle,Snow	3/2/2012 12:00	-3.3	-4.6	94	24	12.9	101.1
Freezing Fog	3/17/2012 6:00	-0.1	-0.3	99	9	0.8	102.8
Freezing Rain	2/1/2012 7:00	0.3	-1.7	92	28	16.1	101.0
Freezing Rain,Fog	12/17/2012 1:00	0.1	-0.9	93	26	9.7	101.0
Freezing Rain,Haze	2/1/2012 15:00	-4.9	-7.4	83	9	2.8	100.4
Freezing Rain,Ice Pellets,Fog	12/17/2012 3:00	-2.6	-3.7	92	28	8.0	100.9
Freezing Rain,Snow Grains	1/13/2012 9:00	-5.0	-7.3	84	32	4.8	98.5
Haze	3/13/2012 23:00	14.1	11.1	86	17	9.7	102.9
Mainly Clear	9/9/2012 9:00	33.0	21.2	99	63	48.3	103.5
Moderate Rain,Fog	12/10/2012 8:00	1.7	0.8	94	17	6.4	99.9
Moderate Snow	12/27/2012 9:00	-4.9	-6.7	93	39	0.8	100.6
Moderate Snow,Blowing Snow	12/27/2012 12:00	-5.4	-6.4	93	41	0.6	100.6

Loading [MathJax]/extensions/Safe.js

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kP
Weather							
Mostly Cloudy	9/9/2012 2:00	32.4	24.4	100	83	48.3	103.6
Rain	9/5/2012 2:00	22.8	20.4	99	52	48.3	102.2
Rain Showers	9/8/2012 16:00	26.4	23.0	97	41	48.3	102.3
Rain Showers,Fog	10/20/2012 3:00	12.8	12.1	96	13	6.4	99.8
Rain Showers,Snow Showers	12/5/2012 10:00	2.2	-1.2	78	28	24.1	101.1
Rain,Fog	9/30/2012 23:00	21.7	19.5	100	46	9.7	101.7
Rain,Haze	3/13/2012 9:00	5.5	2.9	86	17	9.7	100.6
Rain,Ice Pellets	12/18/2012 5:00	0.6	-0.6	92	24	9.7	100.1
Rain,Snow	4/23/2012 3:00	1.7	0.5	94	52	25.0	101.0
Rain,Snow Grains	12/21/2012 0:00	1.9	-2.1	75	26	25.0	100.6
Rain,Snow,Fog	12/8/2012 21:00	0.8	0.3	96	9	6.4	100.7
Rain,Snow,Ice Pellets	12/21/2012 5:00	1.3	0.1	94	28	6.4	100.4
Snow	4/27/2012 9:00	3.7	0.3	96	57	25.0	102.7
Snow Pellets	11/24/2012 15:00	0.7	-6.4	59	35	2.4	99.7
Snow Showers	3/4/2012 21:00	2.9	-0.7	94	37	48.3	102.5
Snow Showers,Fog	12/29/2012 13:00	-10.0	-11.1	92	22	9.7	102.5
Snow,Blowing Snow	2/25/2012 9:00	-1.4	-2.9	91	48	9.7	100.6
Snow,Fog	3/14/2012 19:00	1.1	0.8	99	35	9.7	102.0
Snow,Haze	2/1/2012 21:00	-3.6	-6.4	81	15	6.4	100.9
Snow,Ice Pellets	3/3/2012 4:00	0.8	-1.7	92	33	11.3	100.9
Thunderstorms	7/4/2012 16:00	26.7	20.1	87	15	25.0	100.6
Thunderstorms,Heavy Rain Showers	5/29/2012 6:00	10.9	9.0	88	9	2.4	100.2
Thunderstorms,Moderate Rain Showers,Fog	7/17/2012 6:00	19.6	18.5	93	15	3.2	100.0
Thunderstorms,Rain	7/23/2012 18:00	21.3	19.1	93	30	24.1	100.8

Loading [MathJax]/extensions/Safe.js

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kP
Weather							
Thunderstorms,Rain Showers	9/8/2012 4:00	25.5	23.1	98	32	25.0	101.0
Thunderstorms,Rain Showers,Fog	7/31/2012 20:00	22.9	21.3	91	35	9.7	100.6
Thunderstorms,Rain,Fog	7/17/2012 5:00	20.6	18.6	88	19	4.8	100.0

In [30]:

```
df.groupby('Weather').min()
```

Out[30]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kP
Weather							
Clear	1/11/2012 1:00	-23.3	-28.5	20	0	11.3	99.5
Cloudy	1/1/2012 17:00	-21.4	-26.8	18	0	11.3	98.3
Drizzle	1/23/2012 21:00	1.1	-0.2	74	0	6.4	97.8
Drizzle,Fog	1/23/2012 20:00	0.0	-1.6	85	0	1.0	98.6
Drizzle,Ice Pellets,Fog	12/17/2012 9:00	0.4	-0.7	92	20	4.0	100.7
Drizzle,Snow	12/17/2012 15:00	0.9	0.1	92	9	9.7	100.6
Drizzle,Snow,Fog	12/18/2012 21:00	0.3	-0.1	92	7	2.4	97.7
Fog	1/1/2012 0:00	-16.0	-17.2	80	0	0.2	98.3
Freezing Drizzle	1/13/2012 10:00	-9.0	-12.2	78	6	4.8	98.4
Freezing Drizzle,Fog	1/1/2012 2:00	-6.4	-9.0	82	6	3.6	98.7
Freezing Drizzle,Haze	2/1/2012 11:00	-5.8	-8.3	81	9	2.0	100.2
Freezing Drizzle,Snow	1/13/2012 3:00	-8.3	-10.4	79	6	2.4	99.1
Freezing Fog	1/22/2012 6:00	-19.0	-22.9	71	0	0.2	101.9
Freezing Rain	1/13/2012 11:00	-6.5	-9.0	81	7	2.8	98.2
Freezing Rain,Fog	1/17/2012 23:00	-6.1	-8.7	82	7	2.8	98.3
Freezing Rain,Haze	2/1/2012 14:00	-4.9	-7.5	82	6	2.0	100.3
Freezing Rain,Ice Pellets,Fog	12/17/2012 3:00	-2.6	-3.7	92	28	8.0	100.9

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kP
Weather							
Freezing Rain,Snow Grains	1/13/2012 9:00	-5.0	-7.3	84	32	4.8	98.5
Haze	1/22/2012 12:00	-11.5	-16.0	68	0	4.8	100.3
Mainly Clear	1/10/2012 11:00	-22.8	-28.0	20	0	12.9	98.6
Moderate Rain,Fog	12/10/2012 8:00	1.7	0.8	94	17	6.4	99.9
Moderate Snow	1/12/2012 15:00	-6.3	-7.6	83	26	0.6	99.8
Moderate Snow,Blowing Snow	12/27/2012 10:00	-5.5	-6.6	92	39	0.6	100.5
Mostly Cloudy	1/1/2012 16:00	-23.2	-28.5	18	0	11.3	98.3
Rain	1/1/2012 18:00	0.3	-5.7	40	0	4.0	97.5
Rain Showers	1/1/2012 22:00	1.6	-7.2	37	0	6.4	98.5
Rain Showers,Fog	10/20/2012 3:00	12.8	12.1	96	13	6.4	99.8
Rain Showers,Snow Showers	11/4/2012 8:00	2.1	-1.8	75	17	19.3	101.0
Rain,Fog	1/23/2012 18:00	0.0	-1.2	83	0	2.0	98.6
Rain,Haze	3/13/2012 7:00	4.0	1.0	81	7	4.0	100.5
Rain,Ice Pellets	12/18/2012 5:00	0.6	-0.6	92	24	9.7	100.1
Rain,Snow	1/10/2012 5:00	0.6	-1.7	81	13	2.4	98.1
Rain,Snow Grains	12/21/2012 0:00	1.9	-2.1	75	26	25.0	100.6
Rain,Snow,Fog	12/8/2012 21:00	0.8	0.3	96	9	6.4	100.7
Rain,Snow,Ice Pellets	12/21/2012 1:00	0.9	-0.7	88	17	4.8	99.8
Snow	1/10/2012 1:00	-16.7	-24.6	41	0	1.0	97.7
Snow Pellets	11/24/2012 15:00	0.7	-6.4	59	35	2.4	99.7
Snow Showers	1/12/2012 7:00	-13.3	-19.3	52	0	2.4	99.4
Snow Showers,Fog	12/26/2012 9:00	-11.3	-12.7	89	7	4.0	100.6
Snow,Blowing Snow	1/13/2012 21:00	-12.0	-16.2	70	24	0.6	98.1
Snow,Fog	12/16/2012 15:00	-10.1	-12.0	77	4	1.2	99.3

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kP
Weather							
Snow,Haze	2/1/2012 17:00	-4.3	-7.2	80	0	4.0	100.6
Snow,Ice Pellets	12/10/2012 3:00	-4.3	-5.9	76	19	2.8	99.4
Thunderstorms	7/16/2012 1:00	21.6	19.4	67	0	24.1	99.8
Thunderstorms,Heavy Rain Showers	5/29/2012 6:00	10.9	9.0	88	9	2.4	100.2
Thunderstorms,Moderate Rain Showers,Fog	7/17/2012 6:00	19.6	18.5	93	15	3.2	100.0
Thunderstorms,Rain	5/25/2012 20:00	19.4	18.2	83	4	16.1	100.1
Thunderstorms,Rain Showers	5/29/2012 16:00	11.0	7.0	68	7	6.4	99.6
Thunderstorms,Rain Showers,Fog	6/29/2012 3:00	19.5	16.1	80	7	9.7	99.7
Thunderstorms,Rain,Fog	7/17/2012 5:00	20.6	18.6	88	19	4.8	100.0

In [ ]:

In [ ]:

In [31]:

```
df1 = pd.read_csv(r'C:\Users\PC-chetan\Downloads\2. Cars Data1.csv')
df1
```

Out[31]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepow
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	26
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	20
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	20
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	27
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	22
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	19
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	24
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	26
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	17
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	20

Loading [MathJax]/extensions/Safe.js mns

In [32]:

```
df1.notnull().sum()
```

Out[32]:

```
Make           428
Model          428
Type           428
Origin         428
DriveTrain     428
MSRP           428
Invoice        428
EngineSize     428
Cylinders      426
Horsepower     428
MPG_City       428
MPG_Highway    428
Weight         428
Wheelbase      428
Length         428
dtype: int64
```

In [33]:

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 432 entries, 0 to 431
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Make         428 non-null    object 
 1   Model        428 non-null    object 
 2   Type         428 non-null    object 
 3   Origin       428 non-null    object 
 4   DriveTrain   428 non-null    object 
 5   MSRP         428 non-null    object 
 6   Invoice      428 non-null    object 
 7   EngineSize   428 non-null    float64
 8   Cylinders    426 non-null    float64
 9   Horsepower   428 non-null    float64
 10  MPG_City     428 non-null    float64
 11  MPG_Highway  428 non-null    float64
 12  Weight       428 non-null    float64
 13  Wheelbase    428 non-null    float64
 14  Length       428 non-null    float64
dtypes: float64(8), object(7)
memory usage: 50.8+ KB
```

In [34]:

```
df1.describe()
```

Out[34]:

	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	
count	428.000000	426.000000	428.000000	428.000000	428.000000	428.000000	428.000000	42
mean	3.196729	5.807512	215.885514	20.060748	26.843458	3577.953271	108.154206	18
std	1.108595	1.558443	71.836032	5.238218	5.741201	758.983215	8.311813	1
min	1.300000	3.000000	73.000000	10.000000	12.000000	1850.000000	89.000000	14
25%	2.375000	4.000000	165.000000	17.000000	24.000000	3104.000000	103.000000	17
50%	3.000000	6.000000	210.000000	19.000000	26.000000	3474.500000	107.000000	18
75%	3.900000	6.000000	255.000000	21.250000	29.000000	3977.750000	112.000000	19
max	8.300000	12.000000	500.000000	60.000000	66.000000	7190.000000	144.000000	23

In [35]:

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

Out[35]:

```
Make           4
Model          4
Type           4
Origin         4
DriveTrain     4
MSRP           4
Invoice        4
EngineSize     4
Cylinders      6
Horsepower     4
MPG_City       4
MPG_Highway    4
Weight         4
Wheelbase      4
Length         4
dtype: int64
```

In [36]:

```
df1['Cylinders'].fillna(df1['Cylinders'].mean(), inplace=True)
```

In [37]:

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

Out[37]:

```
Make           4
Model          4
Type           4
Origin         4
DriveTrain     4
MSRP           4
Invoice        4
EngineSize     4
Cylinders      0
Horsepower     4
MPG_City       4
MPG_Highway    4
Weight         4
Wheelbase      4
Length         4
dtype: int64
```

In [38]:

```
df1['Cylinders'].fillna(df1['Cylinders'].mean(), inplace=True)
```

In [39]:

```
df1['Cylinders'].fillna(df1['Cylinders'].mean(), inplace=True)
```

In [40]:

```
df1.head()
```

Out[40]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	M
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265.0	
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200.0	
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	200.0	
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	270.0	

Loading [MathJax]/extensions/Safe.js

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MI
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	225.0	

In [41]:

```
df1.Make.describe()
```

Out[41]:

```
count        428
unique        38
top       Toyota
freq          28
Name: Make, dtype: object
```

In [42]:

```
df1.groupby('Make').get_group('Make')
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
C:\Users\PC-CHE~1\AppData\Local\Temp/ipykernel_8680/3313579405.py in <module>
----> 1 df1.groupby('Make').get_group('Make')

c:\python\python39\lib\site-packages\pandas\core\groupby\groupby.py in get_group(self, nam
e, obj)
    751         inds = self._get_index(name)
    752         if not len(inds):
--> 753             raise KeyError(name)
    754 
    755         return obj._take_with_is_copy(inds, axis=self.axis)

KeyError: 'Make'
```

In [95]:

```
df1.Make.value_counts()
```

Out[95]:

```
Toyota           28
Chevrolet        27
Mercedes-Benz    26
Ford             23
BMW              20
Audi             19
Honda            17
Nissan           17
Volkswagen       15
Chrysler         15
Dodge            13
Mitsubishi       13
Volvo            12
Jaguar           12
Hyundai          12
Subaru           11
Pontiac          11
Mazda            11
Lexus            11
Kia              11
Buick             9
Mercury           9
Lincoln           9
Saturn            8
Cadillac          8
Suzuki            8
Infiniti          8
GMC               8
Acura             7
```

Loading [MathJax]/extensions/Safe.js

```
Porsche          7
Saab             7
Land Rover       3
Oldsmobile       3
Jeep             3
Scion            2
Isuzu            2
MINI             2
Hummer           1
Name: Make, dtype: int64
```

In [43]:
```
df1.head()
```

Out[43]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MI
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265.0	
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200.0	
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	200.0	
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	270.0	
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	225.0	

In [48]:
```
df1[df1['Origin'].isin(['Asia','Europe'])]
```

Out[48]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepow
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	26
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	20
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	20
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	27
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	22
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	19
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	24
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	26
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	17
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	20

281 rows × 15 columns

In [50]:
```
df1[~(df1['Weight'] > 4000)]
```

Out[50]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsep

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsep
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	
5	Acura	3.5 RL w/Navigation 4dr	Sedan	Asia	Front	$46,100	$41,100	3.5	6.0	
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	

329 rows × 15 columns

In [51]:

```

```

Out[51]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepow
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	26
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	20
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	20
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	27
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	22
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	19
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	24
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	26
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	17
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	20

432 rows × 15 columns

In [53]:

```
df1['MPG_City']=df1['MPG_City'].apply(lambda x:x+3)
```

In [54]:

```
df1
```

Loading [MathJax]/extensions/Safe.js

Out[54]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepov
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	26
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	20
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	20
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	27
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	22
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	19
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	24
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	26
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	17
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	20

432 rows × 15 columns

In [56]:

```
df1['MPG_City']=df1['MPG_City'].apply(lambda x:x-3)
```

In [57]:

```
df1
```

Out[57]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepow
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	26
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	20
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	20
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	27
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	22
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	19
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	24
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	26
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	17
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	20

432 rows × 15 columns

In [ ]:

Loading [MathJax]/extensions/Safe.js

In [58]:

```
df2 = pd.read_csv(r'C:\Users\PC-chetan\Downloads\3. Police Data.csv')
```

In [59]:

```
df2
```

Out[59]:

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race
0	1/2/2005	1:55	NaN	M	1985.0	20.0	White
1	1/18/2005	8:15	NaN	M	1965.0	40.0	White
2	1/23/2005	23:15	NaN	M	1972.0	33.0	White
3	2/20/2005	17:15	NaN	M	1986.0	19.0	White
4	3/14/2005	10:00	NaN	F	1984.0	21.0	White
...	...	...	...	...	...	...	...
65530	12/6/2012	17:54	NaN	F	1987.0	25.0	White
65531	12/6/2012	22:22	NaN	M	1954.0	58.0	White
65532	12/6/2012	23:20	NaN	M	1985.0	27.0	Black
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	NaN
65534	12/7/2012	0:30	NaN	F	1985.0	27.0	White

65535 rows × 15 columns

In [60]:

```
df2.isnull().sum()
```

Out[60]:

```
stop_date                 0
stop_time                 0
country_name          65535
driver_gender          4061
driver_age_raw         4054
driver_age             4307
driver_race            4060
violation_raw          4060
violation              4060
search_conducted          0
search_type           63056
stop_outcome           4060
is_arrested            4060
stop_duration          4060
drugs_related_stop        0
dtype: int64
```

In [62]:

```
df2.shape
```

Out[62]:

```
(65535, 15)
```

In [67]:

```
df2.drop( columns = 'country_name' , inplace=True)
```

In [68]:

```
df2
```

Out[68]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe

Loading [MathJax]/extensions/Safe.js

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe
2	1/23/2005	23:15	M	1972.0	33.0	White	Spe
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for Se
4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
...	...	...	...	...	...	...	
65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
65531	12/6/2012	22:22	M	1954.0	58.0	White	Spe
65532	12/6/2012	23:20	M	1985.0	27.0	Black	Equipment/Inspe Vio
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	
65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe

65535 rows × 14 columns

In [69]:

```
df2.isnull().sum()
```

Out[69]:

```
stop_date               0
stop_time               0
driver_gender        4061
driver_age_raw       4054
driver_age           4307
driver_race          4060
violation_raw        4060
violation            4060
search_conducted        0
search_type         63056
stop_outcome         4060
is_arrested          4060
stop_duration        4060
drugs_related_stop      0
dtype: int64
```

In [70]:

```
df2.drop(columns = 'search_type', inplace =True)
```

In [71]:

```
df2
```

Out[71]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe
2	1/23/2005	23:15	M	1972.0	33.0	White	Spe
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for Se
4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
...	...	...	...	...	...	...	
65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
65531	12/6/2012	22:22	M	1954.0	58.0	White	Spe
65532	12/6/2012	23:20	M	1985.0	27.0	Black	Equipment/Inspe Vio

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	
65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe

65535 rows × 13 columns

In [78]:

```
df2[(df2['violation'] == 'Speeding') & (df2['violation_raw'] == 'Speeding')].value_counts
```

Out[78]:

```
37119
```

In [83]:

```
df2[df2.violation == 'Speeding'].driver_gender.value_counts()
```

Out[83]:

```
M    25517
F    11686
Name: driver_gender, dtype: int64
```

In [86]:

```
df2.groupby('driver_gender').search_conducted.sum()
```

Out[86]:

```
driver_gender
F     366
M    2113
Name: search_conducted, dtype: int64
```

In [89]:

```
df2.stop_duration.describe()
```

Out[89]:

```
count         61475
unique            4
top        0-15 Min
freq          47379
Name: stop_duration, dtype: object
```

In [99]:

```
df2
```

Out[99]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe
2	1/23/2005	23:15	M	1972.0	33.0	White	Spe
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for Se
4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
...	...	...	...	...	...	...	
65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
65531	12/6/2012	22:22	M	1954.0	58.0	White	Spe
65532	12/6/2012	23:20	M	1985.0	27.0	Black	Equipment/Inspe Vio
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	
65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe

65535 rows × 13 columns

Loading [MathJax]/extensions/Safe.js

```
df2['stop_duration'].value_counts()
```

Out[101…

```
0-15 Min     47379
16-30 Min    11448
30+ Min       2647
2                1
Name: stop_duration, dtype: int64
```

In [102…

```
df2['stop_duration']=df2['stop_duration'].map({'0-15 Min': 7.5, '16-30 Min': 24, '30+ Min
```

In [103…

```
df2
```

Out[103…

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe
2	1/23/2005	23:15	M	1972.0	33.0	White	Spe
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for Se
4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
...	...	...	...	...	...	...	
65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
65531	12/6/2012	22:22	M	1954.0	58.0	White	Spe
65532	12/6/2012	23:20	M	1985.0	27.0	Black	Equipment/Inspe Vio
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	
65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe

65535 rows × 13 columns

In [105…

```
df2['stop_duration'].mean()
```

Out[105…

```
12.187420698181345
```

In [112…

```
df2.groupby('driver_age').violation.describe()
```

Out[112…

	count	unique	top	freq
driver_age				
15.0	5	2	Moving violation	4
16.0	34	5	Speeding	18
17.0	449	5	Speeding	338
18.0	1344	5	Speeding	980
19.0	2388	5	Speeding	1655
...	...	...	...	...
83.0	2	2	Speeding	1
84.0	3	1	Speeding	3
85.0	1	1	Moving violation	1

Loading [MathJax]/extensions/Safe.js

	count	unique	top	freq
driver_age				
86.0	6	3	Speeding	3
88.0	2	1	Speeding	2

73 rows × 4 columns

In [172…

```
df3 = pd.read_csv(r'C:\Users\PC-chetan\Downloads\covid_19_data.csv')
```

In [173…

```
df3
```

Out[173…

	Date	State	Region	Confirmed	Deaths	Recovered
0	4/29/2020	NaN	Afghanistan	1939	60	252
1	4/29/2020	NaN	Albania	766	30	455
2	4/29/2020	NaN	Algeria	3848	444	1702
3	4/29/2020	NaN	Andorra	743	42	423
4	4/29/2020	NaN	Angola	27	2	7
...	...	...	...	...	...	...
316	4/29/2020	Wyoming	US	545	7	0
317	4/29/2020	Xinjiang	Mainland China	76	3	73
318	4/29/2020	Yukon	Canada	11	0	0
319	4/29/2020	Yunnan	Mainland China	185	2	181
320	4/29/2020	Zhejiang	Mainland China	1268	1	1263

321 rows × 6 columns

In [174…

```
df3.Region.value_counts()
```

Out[174…

```
US                  58
Mainland China      31
Canada              15
France              11
UK                  11
                    ..
Guinea               1
Guinea-Bissau        1
Guyana               1
Haiti                1
Macau                1
Name: Region, Length: 187, dtype: int64
```

In [175…

```
df3.count()
```

Out[175…

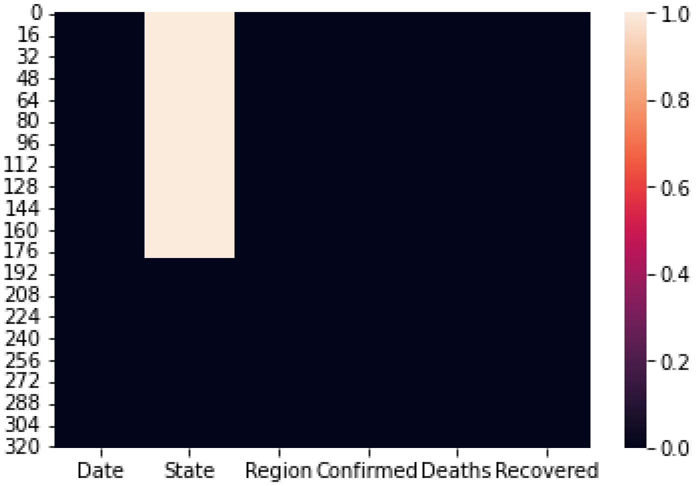
```
Date          321
State         140
Region        321
Confirmed     321
Deaths        321
Recovered     321
```

```
In [176…  df3.isnull().sum()
```

```
Out[176…  Date           0
          State        181
          Region         0
          Confirmed      0
          Deaths         0
          Recovered      0
          dtype: int64
```

```
In [177…  import seaborn as sns
          import matplotlib.pyplot as plt
```

```
In [178…  sns.heatmap(df3.isnull())
          plt.show()
```



```
In [179…  df3.groupby('Region')['Deaths','Confirmed'].sum().head(20)
```

```
C:\Users\PC-CHE~1\AppData\Local\Temp/ipykernel_8680/2284753542.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
  df3.groupby('Region')['Deaths','Confirmed'].sum().head(20)
```

Out[179…

	Deaths	**Confirmed**
Region		
Afghanistan	60	1939
Albania	30	766
Algeria	444	3848
Andorra	42	743
Angola	2	27
Antigua and Barbuda	3	24
Argentina	214	4285
Armenia	30	1932
Australia	91	6752
Austria	580	15402
Azerbaijan	23	1766

Loading [MathJax]/extensions/Safe.js

	Deaths	Confirmed
Region		
Bahamas	11	80
Bahrain	8	2921
Bangladesh	163	7103
Barbados	7	80
Belarus	84	13181
Belgium	7501	47859
Belize	2	18
Benin	1	64
Bhutan	0	7

In [180…

```
df3.groupby('Region').sum().head(20)
```

Out[180…

	Confirmed	Deaths	Recovered
Region			
Afghanistan	1939	60	252
Albania	766	30	455
Algeria	3848	444	1702
Andorra	743	42	423
Angola	27	2	7
Antigua and Barbuda	24	3	11
Argentina	4285	214	1192
Armenia	1932	30	900
Australia	6752	91	5715
Austria	15402	580	12779
Azerbaijan	1766	23	1267
Bahamas	80	11	23
Bahrain	2921	8	1455
Bangladesh	7103	163	150
Barbados	80	7	39
Belarus	13181	84	2072
Belgium	47859	7501	11283
Belize	18	2	9
Benin	64	1	33
Bhutan	7	0	5

In [181…

```
df3.groupby('Region')['Confirmed'].sum().sort_values(ascending = False)
```

Out[181…

```
Region
US                     1039909
                        236899
```

```
Italy                        203591
France                       166543
UK                           166441
                              ...
Sao Tome and Principe             8
Papua New Guinea                  8
Bhutan                            7
Western Sahara                    6
Yemen                             6
Name: Confirmed, Length: 187, dtype: int64
```

In [182…

```
df3.groupby('Region')['Deaths'].sum().sort_values(ascending = True).head(40)
```

Out[182…

```
Region
Laos                                    0
Mongolia                                0
Mozambique                              0
Cambodia                                0
Fiji                                    0
Namibia                                 0
Nepal                                   0
Madagascar                              0
Macau                                   0
Papua New Guinea                        0
Rwanda                                  0
Saint Kitts and Nevis                   0
Bhutan                                  0
Dominica                                0
Central African Republic                0
Saint Lucia                             0
Holy See                                0
Sao Tome and Principe                   0
Yemen                                   0
Western Sahara                          0
Eritrea                                 0
Vietnam                                 0
Saint Vincent and the Grenadines        0
Timor-Leste                             0
Uganda                                  0
Grenada                                 0
South Sudan                             0
Seychelles                              0
Liechtenstein                           1
Maldives                                1
Gambia                                  1
Eswatini                                1
Guinea-Bissau                           1
Equatorial Guinea                       1
Mauritania                              1
Cabo Verde                              1
Benin                                   1
Burundi                                 1
Suriname                                1
Brunei                                  1
Name: Deaths, dtype: int64
```

In [183…

```
df3
```

Out[183…

	Date	State	Region	Confirmed	Deaths	Recovered
0	4/29/2020	NaN	Afghanistan	1939	60	252
1	4/29/2020	NaN	Albania	766	30	455

Loading [MathJax]/extensions/Safe.js

	Date	State	Region	Confirmed	Deaths	Recovered
2	4/29/2020	NaN	Algeria	3848	444	1702
3	4/29/2020	NaN	Andorra	743	42	423
4	4/29/2020	NaN	Angola	27	2	7
...	...	...	...	...	...	...
316	4/29/2020	Wyoming	US	545	7	0
317	4/29/2020	Xinjiang	Mainland China	76	3	73
318	4/29/2020	Yukon	Canada	11	0	0
319	4/29/2020	Yunnan	Mainland China	185	2	181
320	4/29/2020	Zhejiang	Mainland China	1268	1	1263

321 rows × 6 columns

In [190…

```
df3[df3['Region'] == 'India']
```

Out[190…

	Date	State	Region	Confirmed	Deaths	Recovered
74	4/29/2020	NaN	India	33062	1079	8437

In [191…

```
df3.sort_values(by =['Confirmed'], ascending = True)
```

Out[191…

	Date	State	Region	Confirmed	Deaths	Recovered
285	4/29/2020	Recovered	US	0	0	120720
284	4/29/2020	Recovered	Canada	0	0	20327
203	4/29/2020	Diamond Princess cruise ship	Canada	0	1	0
305	4/29/2020	Tibet	Mainland China	1	0	1
289	4/29/2020	Saint Pierre and Miquelon	France	1	0	0
...	...	...	...	...	...	...
57	4/29/2020	NaN	France	165093	24087	48228
168	4/29/2020	NaN	UK	165221	26097	0
80	4/29/2020	NaN	Italy	203591	27682	71252
153	4/29/2020	NaN	Spain	236899	24275	132929
265	4/29/2020	New York	US	299691	23477	0

321 rows × 6 columns

In [193…

```
df3.sort_values(by =['Deaths'], ascending = True).head(50)
```

Out[193…

	Date	State	Region	Confirmed	Deaths	Recovered
126	4/29/2020	NaN	Papua New Guinea	8	0	0
279	4/29/2020	Prince Edward Island	Canada	27	0	0
135	4/29/2020	NaN	Rwanda	225	0	98
272	4/29/2020	Northwest Territories	Canada	5	0	0

Loading [MathJax]/extensions/Safe.js

	Date	State	Region	Confirmed	Deaths	Recovered
271	4/29/2020	Northern Territory	Australia	28	0	25
178	4/29/2020	NaN	Yemen	6	0	1
267	4/29/2020	Ningxia	Mainland China	75	0	75
90	4/29/2020	NaN	Laos	19	0	7
260	4/29/2020	New Caledonia	France	18	0	17
259	4/29/2020	New Brunswick	Canada	118	0	0
184	4/29/2020	Anguilla	UK	3	0	3
192	4/29/2020	Bonaire, Sint Eustatius and Saba	Netherlands	5	0	0
29	4/29/2020	NaN	Cambodia	122	0	119
244	4/29/2020	Macau	Macau	45	0	34
204	4/29/2020	Diamond Princess cruise ship	US	49	0	0
237	4/29/2020	Jiangsu	Mainland China	653	0	648
206	4/29/2020	Falkland Islands (Malvinas)	UK	13	0	11
51	4/29/2020	NaN	Eritrea	39	0	19
207	4/29/2020	Faroe Islands	Denmark	187	0	181
210	4/29/2020	French Polynesia	France	58	0	50
55	4/29/2020	NaN	Fiji	18	0	12
214	4/29/2020	Gibraltar	UK	141	0	131
70	4/29/2020	NaN	Holy See	10	0	2
215	4/29/2020	Grand Princess	Canada	13	0	0
217	4/29/2020	Greenland	Denmark	11	0	11
45	4/29/2020	NaN	Dominica	16	0	13
177	4/29/2020	NaN	Western Sahara	6	0	5
31	4/29/2020	NaN	Central African Republic	50	0	10
281	4/29/2020	Qinghai	Mainland China	18	0	18
318	4/29/2020	Yukon	Canada	11	0	0
136	4/29/2020	NaN	Saint Kitts and Nevis	15	0	4
137	4/29/2020	NaN	Saint Lucia	17	0	15
138	4/29/2020	NaN	Saint Vincent and the Grenadines	16	0	8
140	4/29/2020	NaN	Sao Tome and Principe	8	0	4
144	4/29/2020	NaN	Seychelles	11	0	6
305	4/29/2020	Tibet	Mainland China	1	0	1
115	4/29/2020	NaN	Nepal	57	0	16
114	4/29/2020	NaN	Namibia	16	0	8
113	4/29/2020	NaN	Mozambique	76	0	12
152	4/29/2020	NaN	South Sudan	34	0	0
99	4/29/2020	NaN	Madagascar	128	0	90

	Date	State	Region	Confirmed	Deaths	Recovered
294	4/29/2020	Shanxi	Mainland China	197	0	164
286	4/29/2020	Reunion	France	420	0	300
175	4/29/2020	NaN	Vietnam	270	0	222
284	4/29/2020	Recovered	Canada	0	0	20327
285	4/29/2020	Recovered	US	0	0	120720
169	4/29/2020	NaN	Uganda	81	0	52
110	4/29/2020	NaN	Mongolia	38	0	10
288	4/29/2020	Saint Barthelemy	France	6	0	6
18	4/29/2020	NaN	Bhutan	7	0	5

In [ ]:

```
```

In [194…

```
df4 = pd.read_csv(r'C:\Users\PC-chetan\Downloads\5. London Housing Data.csv')
df4
```

Out[194…

	date	area	average_price	code	houses_sold	no_of_crimes
0	1/1/1995	city of london	91449	E09000001	17.0	NaN
1	2/1/1995	city of london	82203	E09000001	7.0	NaN
2	3/1/1995	city of london	79121	E09000001	14.0	NaN
3	4/1/1995	city of london	77101	E09000001	7.0	NaN
4	5/1/1995	city of london	84409	E09000001	10.0	NaN
...	...	...	...	...	...	...
13544	9/1/2019	england	249942	E92000001	64605.0	NaN
13545	10/1/2019	england	249376	E92000001	68677.0	NaN
13546	11/1/2019	england	248515	E92000001	67814.0	NaN
13547	12/1/2019	england	250410	E92000001	NaN	NaN
13548	1/1/2020	england	247355	E92000001	NaN	NaN

13549 rows × 6 columns

In [195…

```
df4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13549 entries, 0 to 13548
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   date           13549 non-null  object 
 1   area           13549 non-null  object 
 2   average_price  13549 non-null  int64  
 3   code           13549 non-null  object 
 4   houses_sold    13455 non-null  float64
 5   no_of_crimes   7439 non-null   float64
dtypes: float64(2), int64(1), object(3)
memory usage: 635.2+ KB
```

In [197…

```
df4.count()
```

Out[197…

```
date              13549
area              13549
average_price     13549
code              13549
houses_sold       13455
no_of_crimes       7439
dtype: int64
```

In [198…

```
df4.isnull().sum()
```

Out[198…

```
date                0
area                0
average_price       0
code                0
houses_sold        94
no_of_crimes     6110
dtype: int64
```

In [200…

```
df4.head(10
        )
```

Out[200…

	date	area	average_price	code	houses_sold	no_of_crimes
0	1/1/1995	city of london	91449	E09000001	17.0	NaN
1	2/1/1995	city of london	82203	E09000001	7.0	NaN
2	3/1/1995	city of london	79121	E09000001	14.0	NaN
3	4/1/1995	city of london	77101	E09000001	7.0	NaN
4	5/1/1995	city of london	84409	E09000001	10.0	NaN
5	6/1/1995	city of london	94901	E09000001	17.0	NaN
6	7/1/1995	city of london	110128	E09000001	13.0	NaN
7	8/1/1995	city of london	112329	E09000001	14.0	NaN
8	9/1/1995	city of london	104473	E09000001	17.0	NaN
9	10/1/1995	city of london	108038	E09000001	14.0	NaN

In [234…

```
df4[df4['no_of_crimes'] == df4['no_of_crimes'].isnull()]
```

Out[234…

	date	area	average_price	code	houses_sold	no_of_crimes
72	1/1/2001	city of london	284262	E09000001	24.0	0.0
73	2/1/2001	city of london	198137	E09000001	37.0	0.0
74	3/1/2001	city of london	189033	E09000001	44.0	0.0
75	4/1/2001	city of london	205494	E09000001	38.0	0.0
76	5/1/2001	city of london	223459	E09000001	30.0	0.0
...	...	...	...	...	...	...
178	11/1/2009	city of london	397909	E09000001	11.0	0.0
179	12/1/2009	city of london	411955	E09000001	16.0	0.0
180	1/1/2010	city of london	464436	E09000001	20.0	0.0
181	2/1/2010	city of london	490525	E09000001	9.0	0.0

	date	area	average_price	code	houses_sold	no_of_crimes
182	3/1/2010	city of london	498241	E09000001	15.0	0.0

104 rows × 6 columns

In [235…

```
df4.groupby('no_of_crimes').describe()
```

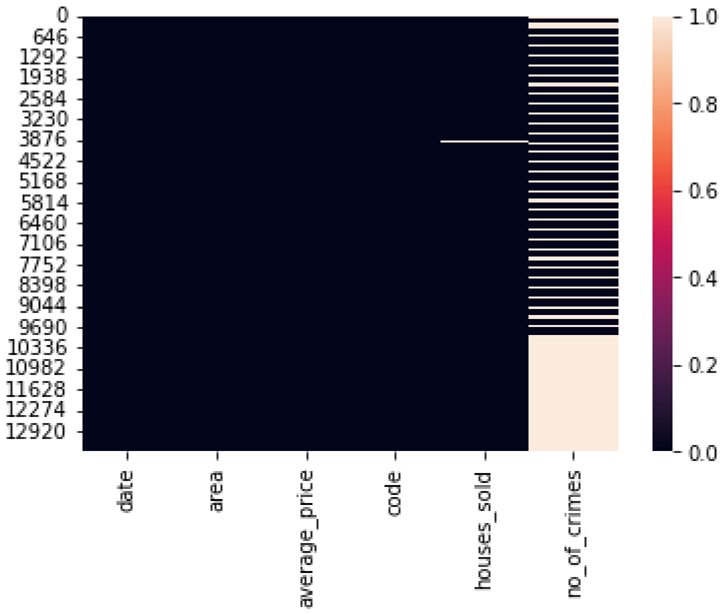
Out[235…

	average_price							
	count	mean	std	min	25%	50%	75%	max
no_of_crimes								
0.0	104.0	329678.913462	72750.364469	189033.0	276289.25	316759.5	379974.5	498241.0
3.0	1.0	467348.000000	NaN	467348.0	467348.00	467348.0	467348.0	467348.0
5.0	1.0	411183.000000	NaN	411183.0	411183.00	411183.0	411183.0	411183.0
7.0	3.0	407195.333333	10394.067170	399437.0	401290.50	403144.0	411074.5	419005.0
8.0	1.0	473887.000000	NaN	473887.0	473887.00	473887.0	473887.0	473887.0
...	...	...	...	...	...	...	...	...
7076.0	1.0	331670.000000	NaN	331670.0	331670.00	331670.0	331670.0	331670.0
7208.0	1.0	927864.000000	NaN	927864.0	927864.00	927864.0	927864.0	927864.0
7215.0	1.0	960161.000000	NaN	960161.0	960161.00	960161.0	960161.0	960161.0
7227.0	1.0	992834.000000	NaN	992834.0	992834.00	992834.0	992834.0	992834.0
7461.0	1.0	968404.000000	NaN	968404.0	968404.00	968404.0	968404.0	968404.0

2669 rows × 16 columns

In [236…

```
sns.heatmap(df4.isnull())
plt.show()
```



In [ ]: