

Import Database

```
In [12]: import pandas as pd

In [13]: datapoint_read_csv(r'\\E:\V\Project Data\London Housing\file.csv')
data.head()
```

```
Out[13]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	01-01-2012 00:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	01-01-2012 01:00	-1.8	-3.7	87	4	8.0	101.24	Fog
2	01-01-2012 02:00	-1.8	-3.4	89	7	4.0	101.26	Freezing Drizzle,Fog
3	01-01-2012 03:00	-1.5	-3.2	88	6	4.0	101.27	Freezing Drizzle,Fog
4	01-01-2012 04:00	-1.5	-3.3	88	7	4.8	101.23	Fog

How to analyze data

```
In [15]: data.shape

Out[15]: (8784, 8)

In [16]: data.describe()
```

```
Out[16]:
```

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa
count	8784.000000	8784.000000	8784.000000	8784.000000	8784.000000	8784.000000
mean	8.786144	2.555294	67.431694	14.945469	27.664447	101.051623
std	11.687883	10.883072	16.918881	8.688896	12.622688	0.844005
min	-23.300000	-28.500000	18.000000	0.000000	0.200000	97.520000
5%	0.100000	-5.900000	56.000000	9.000000	24.100000	100.560000
25%	9.300000	3.300000	68.000000	13.000000	25.000000	101.070000
75%	18.800000	11.800000	81.000000	20.000000	25.000000	101.590000
max	33.000000	24.400000	100.000000	83.000000	48.300000	103.650000

```
In [19]: data.index
data.columns
data.dtypes
data['weather'].unique()
data.nunique()
data.count()
data['weather'].value_counts()
data.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 datetime64[ns]
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_kmh 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(1)
memory usage: 549.1+ KB
```

Change data type of 'Date/Time' column as datetime format

```
In [19]: data['date/Time'] = pd.to_datetime(data['Date/Time'],format='mixed')
data.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 datetime64[ns]
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_kmh 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: datetime64[ns](1), float64(4), int64(2), object(1)
memory usage: 549.1+ KB
```

1. Find all the unique'wind speed' value in the data.

```
In [37]: data['Wind Speed_kmh']

Out[37]: array([ 4,  7,  6,  9, 15, 13, 26, 22, 19, 24, 30, 35, 39, 32, 33, 26, 44,
        43, 48, 37, 28, 37, 11,  8, 83, 70, 57, 46, 41, 52, 90, 63, 54, 2],
      dtype=int64)
```

2. Find the number of time when the 'weather' is exactly clear?

```
In [43]: # value_count()
data.Weather.value_counts()

# Filtering
data[data.Weather == 'Clear']

# groupby
data.groupby('Weather').get_group('Clear')
```

```
Out[43]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
614	2012-01-03 19:00:00	-16.9	-24.8	50	24	25.0	101.74	Clear
117	2012-01-05 18:00:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	2012-01-05 19:00:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	2012-01-05 20:00:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	2012-01-05 21:00:00	-9.0	-14.8	63	13	25.0	100.83	Clear
...
8646	2012-12-26 06:00:00	-13.4	-14.8	89	4	25.0	102.47	Clear
8698	2012-12-28 10:00:00	-6.1	-8.6	82	19	24.1	101.27	Clear
8713	2012-12-29 01:00:00	-11.9	-13.6	87	11	25.0	101.31	Clear
8756	2012-12-29 02:00:00	-11.8	-13.1	90	13	25.0	101.33	Clear
8714	2012-12-30 20:00:00	-13.8	-16.5	80	24	25.0	101.52	Clear

1326 rows x 8 columns

3. Find the number of times when the 'wind speed was exactly 4 Km/h'

```
In [44]: data[data['Wind Speed_kmh'] == 4]

Out[44]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	2012-01-01 00:00:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	2012-01-01 01:00:00	-1.8	-3.7	87	4	8.0	101.24	Fog
96	2012-01-05 00:00:00	-8.8	-11.7	79	4	9.7	100.32	Snow
101	2012-01-05 05:00:00	-7.0	-9.5	82	4	4.0	100.19	Snow
146	2012-01-07 02:00:00	-8.1	-11.1	79	4	19.3	100.15	Cloudy
...
8766	2012-12-31 08:00:00	-8.6	-10.3	87	4	3.2	101.14	Snow Showers
8769	2012-12-31 09:00:00	-8.1	-9.6	89	4	2.4	101.09	Snow
8770	2012-12-31 10:00:00	-7.4	-8.9	89	4	6.4	101.05	Snow,Fog
8772	2012-12-31 12:00:00	-5.8	-7.5	88	4	12.9	100.78	Snow
8773	2012-12-31 13:00:00	-4.6	-6.6	86	4	12.9	100.63	Snow

474 rows x 8 columns

4. Find the null values in dataset.

```
In [46]: data.isnull().sum()
# or
data.notnull().sum()

Out[46]:
```

Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
8784	8784	8784	8784	8784	8784	8784	8784
Dew Point Temp_C	8784	8784	8784	8784	8784	8784	8784
Rel Hum_%	8784	8784	8784	8784	8784	8784	8784
Wind Speed_kmh	8784	8784	8784	8784	8784	8784	8784
Press_kPa	8784	8784	8784	8784	8784	8784	8784
Weather	8784	8784	8784	8784	8784	8784	8784
dtype: int64							

5 . Rename the column Name 'Weather' as 'Weather Condition'

```
In [48]: data.rename( columns= {'Weather':'Weather Condition'},inplace=True)
data

Out[48]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather Condition
0	2012-01-01 00:00:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	2012-01-01 01:00:00	-1.8	-3.7	87	4	8.0	101.24	Fog
2	2012-01-01 02:00:00	-1.8	-3.4	89	7	4.0	101.26	Freezing Drizzle,Fog
3	2012-01-01 03:00:00	-1.5	-3.2	88	6	4.0	101.27	Freezing Drizzle,Fog
4	2012-01-01 04:00:00	-1.5	-3.3	88	7	4.8	101.23	Fog
...
8779	2012-12-31 19:00:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	2012-12-31 20:00:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	2012-12-31 21:00:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	2012-12-31 22:00:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	2012-12-31 23:00:00	0.0	-2.1	86	30	11.3	99.89	Snow

8784 rows x 8 columns

7. What is the mean() of 'Visibility'?

```
In [49]: data.Visibility_km.mean()

Out[49]: 27.664446721311478
```

8. What is the SD of 'pressure'?

```
In [50]: data.Press_kPa.std()

Out[50]: 0.8449847458486474
```

9. Find all the instance whom 'Snow' was recorded?

```
In [54]: # value_count()
data['weather condition'].value_counts()

# Filtering
data[data['Weather Condition'] == 'Snow']

# groupby
data.groupby('Weather Condition').get_group('Snow')
```

```
Out[54]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather Condition
55	2012-01-03 07:00:00	-14.0	-19.5	63	19	25.0	100.95	Snow
84	2012-01-04 12:00:00	-13.7	-21.7	51	11	24.1	101.25	Snow
86	2012-01-04 14:00:00	-11.3	-19.0	53	7	19.3	100.97	Snow
87	2012-01-04 15:00:00	-10.2	-16.3	61	11	9.7	100.89	Snow
88	2012-01-04 16:00:00	-9.4	-15.5	61	13	19.3	100.79	Snow
...
8779	2012-12-31 19:00:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	2012-12-31 20:00:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	2012-12-31 21:00:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	2012-12-31 22:00:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	2012-12-31 23:00:00	0.0	-2.1	86	30	11.3	99.89	Snow

390 rows x 8 columns

10. find the instance when'wind speed is above 24' and 'visibility is 25'.

```
In [57]: data[(data['Wind Speed_kmh']> 24) & (data['Visibility_km'] == 25)]

Out[57]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather Condition
23	2012-01-01 23:00:00	5.3	2.0	79	30	25.0	99.31	Cloudy
24	2012-01-02 00:00:00	5.2	1.5	77	35	25.0	99.26	Rain Showers
25	2012-01-02 01:00:00	4.6	0.0	72	39	25.0	99.26	Cloudy
26	2012-01-02 02:00:00	3.9	-0.9	71	32	25.0	99.26	Mostly Cloudy
27	2012-01-02 03:00:00	3.7	-1.5	69	33	25.0	99.30	Mostly Cloudy
...
8795	2012-12-28 17:00:00	-8.6	-12.0	76	26	25.0	101.34	Mainly Clear
8753	2012-12-30 17:00:00	-12.1	-15.8	74	26	25.0	101.26	Mainly Clear
8755	2012-12-30 19:00:00	-13.4	-16.5	77	26	25.0	101.47	Mainly Clear
8759	2012-12-30 23:00:00	-12.1	-15.1	78	28	25.0	101.52	Mostly Cloudy
8760	2012-12-31 00:00:00	-11.1	-14.4	77	26	25.0	101.51	Cloudy

308 rows x 8 columns

12. What is the mean value of each column against 'weather Condition'?

```
In [59]: data.groupby('Weather Condition').mean()

Out[59]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa
Clear	2012-06-25 18:03:12.7601889982	6.825716	0.089367	64.487738	10.557315	30.153243	101.971441
Cloudy	2012-07-01 16:51:29.58333376	7.970544	2.375810	69.592593	16.127315	26.625752	100.581441
Drizzle	2012-09-17 15:00:00.000000000	7.359599	5.504818	88.243502	16.097561	17.931707	100.430366
Drizzle,Fog	2012-06-30 00:44:15.000000000	8.867500	7.033750	93.275000	11.862500	5.257500	100.786625
Drizzle,Peleets,Fog	2012-12-17 09:00:00.000000000	0.400000	-0.700000	92.000000	20.000000	4.000000	100.795000
Drizzle,Snow	2012-12-18 16:30:00.000000000	1.690000	0.150000	93.500000	14.000000	10.500000	100.890000
Drizzle,Snow,Fog	2012-12-19 20:56:00.000000000	0.893333	0.120000	96.866667	15.533333	5.513333	99.281333
Fog	2012-06-24 21:08:00.000000000	4.303333	3.159033	92.286667	7.946667	6.249000	101.194067
Freezing Drizzle	2012-03-05 23:25:42.857142784	-5.657143	-8.000000	83.571429	16.571429	9.200000	100.202857
Freezing Drizzle,Fog	2012-04-30 06:20:00.000000000	-2.533333	-4.183333	88.500000	17.000000	5.266667	100.441667
Freezing Drizzle,Haze	2012-02-01 12:00:00.000000000	-5.433333	-8.000000	82.000000	10.333333	2.666667	100.316667
Freezing Drizzle,Snow	2012-06-27 12:32:43.636363620	-5.109991	-7.072727	86.090909	16.272727	5.872727	100.520909
Freezing Fog	2012-02-22 06:45:00.000000000	-7.575000	-9.250000	87.750000	4.750000	0.650000	102.320000
Freezing Rain	2012-02-14 01:34:17.142857216	-3.885714	-6.078571	84.643507	19.214286	8.242857	99.647143
Freezing Rain,Fog	2012-06-28 15:45:00.000000000	-2.225000	-3.750000	89.500000	15.500000	7.550000	99.945000
Freezing Rain,Haze	2012-02-01 14:30:00.000000000	-4.900000	-7.450000	82.500000	7.500000	2.400000	100.375000
Freezing Rain,Peleets,Fog	2012-12-17 03:00:00.000000000	-2.600000	-3.700000	92.000000	28.000000	8.000000	100.950000
Freezing Rain,Snow Grains	2012-12-13 08:00:00.000000000	-5.000000	-7.300000	84.000000	32.000000	4.800000	98.560000
Haze	2012-06-06 00:37:30.000000000	-0.200000	-2.975000	81.625000	10.437500	7.831250	101.482500
Mainly Clear	2012-07-12 06:48:08.888888832	12.558927	4.581671	60.687142	14.144824	34.264962	101.248832
Moderate Rain,Fog	2012-12-10 08:00:00.000000000	1.700000	0.800000	94.000000	17.000000	6.400000	99.980000
Moderate Snow	2012-07-05 12:00:00.000000000	-5.525000	-7.250000	87.750000	33.750000	0.750000	100.275000
Moderate Snow,Blowing Snow	2012-07-02 11:00:00.000000000	-5.450000	-6.500000	92.500000	18.50000		

