

Weather Data Analysis (Part of Big Data Analysis)

The Weather Dataset here is a time series dataset with per hour information about the weather conditions at a particular location. It records Temperature , Dew Point Temperature, Relative Humidity , Wind Speed , Visibility , Pressure and Conditions

```
In [40]: # IMPORTING A FILE
import pandas as pd

df = pd.read_csv('C:\Users\Administrator\OneDrive\Documents\Weather_data.csv')
df

Out[40]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	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	86	6	4.0	101.27	Freezing Drizzle,Fog
4	1/1/2012 4:00	-1.5	-3.3	86	7	4.8	101.23	Fog
...
8779	12/9/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/9/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/9/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/9/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/9/2012 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

8784 rows x 8 columns

```
In [101]: # ANALYZING THE DATA
# 1) .head() = shows the first n rows in a data, here n = 10
print(df.head(10))

Out[101]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	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	86	6	4.0	101.27	Freezing Drizzle,Fog
4	1/1/2012 4:00	-1.5	-3.3	86	7	4.8	101.23	Fog
5	1/1/2012 5:00	-1.4	-3.2	89	6	6.4	101.29	Fog
6	1/1/2012 6:00	-1.5	-3.1	89	7	8.0	101.26	Fog
7	1/1/2012 7:00	-1.4	-3.6	85	7	6.0	101.20	Fog
8	1/1/2012 8:00	-1.4	-3.4	85	9	4.0	101.20	Fog
9	1/1/2012 9:00	-1.3	-3.1	88	15			

```
In [102]: # 2) .shape = shows the total number of rows and columns of the dataframe
print(df.shape)

Out[102]:
```

(8784, 9)

```
In [103]: # 3) .index = attributes provide the index to the dataframe
df.index

Out[103]:
```

RangeIndex(start=0, stop=8784, step=1)

```
In [104]: # 4) .columns = shows the column name
df.columns

Out[104]:
```

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

```
In [105]: # 5) .dtypes = shows the data types
df.dtypes

Out[105]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	float64	float64	float64	float64	float64	float64	float64	object
1	float64	float64	float64	float64	float64	float64	float64	object
2	float64	float64	float64	float64	float64	float64	float64	object
3	float64	float64	float64	float64	float64	float64	float64	object
4	float64	float64	float64	float64	float64	float64	float64	object
5	float64	float64	float64	float64	float64	float64	float64	object
6	float64	float64	float64	float64	float64	float64	float64	object
7	float64	float64	float64	float64	float64	float64	float64	object
8	float64	float64	float64	float64	float64	float64	float64	object
9	float64	float64	float64	float64	float64	float64	float64	object

```
In [106]: # 6) .dtypes = shows the data types
df.dtypes

Out[106]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	float64	float64	float64	float64	float64	float64	float64	object
1	float64	float64	float64	float64	float64	float64	float64	object
2	float64	float64	float64	float64	float64	float64	float64	object
3	float64	float64	float64	float64	float64	float64	float64	object
4	float64	float64	float64	float64	float64	float64	float64	object
5	float64	float64	float64	float64	float64	float64	float64	object
6	float64	float64	float64	float64	float64	float64	float64	object
7	float64	float64	float64	float64	float64	float64	float64	object
8	float64	float64	float64	float64	float64	float64	float64	object
9	float64	float64	float64	float64	float64	float64	float64	object

```
In [107]: # 7) .unique() = shows the unique values for a particular column
df['Weather'].unique()

Out[107]:
```

array(['Fog', 'Freezing Drizzle,Fog', 'Mostly Cloudy', 'Cloudy', 'Rain', 'Rain Showers', 'Mainly Clear', 'Snow Showers', 'Snow', 'Clear', 'Freezing Rain,Fog', 'Freezing Rain', 'Freezing Drizzle', 'Rain,Snow', 'Moderate Snow', 'Freezing Drizzle,Snow', 'Freezing Rain,Snow Grains', 'Snow,Blowing Snow', 'Freezing Fog', 'Freezing Drizzle,Rain', 'Freezing Rain,Rain', 'Snow,Rain', 'Snow,Rain', 'Snow,Fog', 'Snow,Ice Pellets', 'Rain,Rain', 'Thunderstorms,Rain', 'Thunderstorms,Rain Showers', 'Thunderstorms,Heavy Rain Showers', 'Thunderstorms,Rain Showers,Fog', 'Thunderstorms', 'Thunderstorms,Rain,Fog', 'Thunderstorms,Moderate Rain Showers,Fog', 'Rain Showers,Fog', 'Rain Showers,Snow Showers', 'Snow Pellets', 'Rain,Snow,Fog', 'Moderate Rain,Fog', 'Freezing Rain,Ice Pellets,Fog', 'Drizzle,Ice Pellets,Fog', 'Drizzle,Snow', 'Main,Ice Pellets', 'Drizzle,Snow,Fog', 'Rain,Snow Grains', 'Rain,Snow,Ice Pellets', 'Snow Showers,Fog', 'Moderate Snow,Blowing Snow'], dtype=object)

```
In [108]: # 8) .nunique() = shows the unique values for each column. Can also apply on whole data frame
df.nunique()

Out[108]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	8784	533	489	83	34	24	518	50
1	8784	533	489	83	34	24	518	50
2	8784	533	489	83	34	24	518	50
3	8784	533	489	83	34	24	518	50
4	8784	533	489	83	34	24	518	50
5	8784	533	489	83	34	24	518	50
6	8784	533	489	83	34	24	518	50
7	8784	533	489	83	34	24	518	50
8	8784	533	489	83	34	24	518	50
9	8784	533	489	83	34	24	518	50

```
In [109]: # 9) .count() = shows the total of non null values in each column. Can also apply on whole data frame
df.count()

Out[109]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	8784	8784	8784	8784	8784	8784	8784	8784
1	8784	8784	8784	8784	8784	8784	8784	8784
2	8784	8784	8784	8784	8784	8784	8784	8784
3	8784	8784	8784	8784	8784	8784	8784	8784
4	8784	8784	8784	8784	8784	8784	8784	8784
5	8784	8784	8784	8784	8784	8784	8784	8784
6	8784	8784	8784	8784	8784	8784	8784	8784
7	8784	8784	8784	8784	8784	8784	8784	8784
8	8784	8784	8784	8784	8784	8784	8784	8784
9	8784	8784	8784	8784	8784	8784	8784	8784

```
In [110]: # 10) .value_counts = shows the unique values with their count. Can only apply on a single column
df.value_counts('Press_kPa')

Out[110]:
```

Press_kPa	count
101.33	69
101.36	57
100.69	56
101.06	54
100.65	54
...	...
98.45	3
102.70	1
98.50	1
98.53	1
97.52	1
Name: count, dtype: int64	

```
In [111]: # 11) .info = provides teh basic info of the dataframe
df.info()

Out[111]:
```

```
<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   float64  
 4   Wind Speed_kmh/h     8784 non-null   float64  
 5   Visibility_km         8784 non-null   float64  
 6   Press_kPa            8784 non-null   float64  
 7   Weather              8784 non-null   object  
dtypes: float64(6), int64(1), object(1)
memory usage: 549.1+ KB
```

```
In [112]: # 11) Finding the "Wind Speed" unique values in the data.
df['Wind Speed_kmh/h'].unique()

Out[112]:
```

array([4, 7, 6, 9, 15, 13, 20, 22, 19, 24, 30, 35, 39, 32, 33, 26, 44, 43, 45, 37, 28, 17, 11, 0, 85, 70, 57, 46, 41, 52, 50, 63, 54, 21], dtype=int64)

```
In [113]: df['Wind Speed_kmh/h'].nunique()

Out[113]:
```

34

```
In [114]: # 11) Find the number of times when the weather is exactly "Clear".
# a) First way
df[df.Weather == "Clear"]

Out[114]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
67	1/9/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
114	1/9/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	1/9/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	1/9/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	1/9/2012 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 x 8 columns

```
In [115]: # b) Second Way
df.groupby("Weather").get_group("Clear")

Out[115]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
67	1/9/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
114	1/9/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	1/9/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	1/9/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	1/9/2012 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 x 8 columns

```
In [116]: # 11) Find the number of times when the wind speed was exactly "6km/h".
df.groupby("Wind Speed_kmh/h").get_group(6)

Out[116]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	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	1/9/2012 0:00	-8.8	-19.7	79	4	9.7	100.32	Snow
101	1/9/2012 5:00	-7.0	-9.5	82	4	4.0	100.19	Snow
146	1/7/2012 2:00	-6.1	-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	86	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 x 8 columns

```
In [117]: df.isnull().sum()

Out[117]:
```

Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
...
8783	0	0	0	0	0	0	0
8784	0	0	0	0	0	0	0

```
In [118]: # 12) Renaming a column "Weather" to "Weather Condition"
df.rename(columns = {'Weather':"Weather Condition"}, inplace = True)
df.head(2)
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	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

```
In [119]: # 13) What is the mean visibility?
df.Visibility_km.mean()

Out[119]:
```

21.664446723311478

```
In [120]: # 14) What is the standard deviation of pressure in this data?
df.Press_kPa.std()

Out[120]:
```

0.6440047459486474

```
In [121]: # 15) What is the variance of relative humidity in this data?
df['Rel Hum_%'].var()

Out[121]:
```

286.2485501984988

```
In [122]: # 16) Find all the instances when snow was recorded?
# first way
df["Weather Condition"].value_counts()

Out[122]:
```

Weather Condition	count
Mainly Clear	2105
Mostly Cloudy	2059
Cloudy	1728
Clear	1326
Snow	330
Rain	305
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	16
Thunderstorms,Rain Showers	16
Rain	15
Rain Showers	15
Drizzle,Snow,Fog	14
Freezing Rain	14
Freezing Drizzle,Snow	12
Freezing Drizzle	7
Freezing Rain,Ice Pellets	6
Freezing Drizzle,Fog	6
Snow,Rain	5
Freezing Fog	4
Snow Showers,Fog	4
Moderate Snow	4
Rain,Snow,Ice Pellets	4
Freezing Rain,Fog	4
Freezing Drizzle,Rain	3
Thunderstorms,Rain	3
Thunderstorms,Rain Showers,Fog	3
Freezing Rain,Rain	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
Rain,Ice Pellets	1
Thunderstorms,Rain,Fog	1
Rain,Ice Pellets	1
Thunderstorms,Heavy Rain Showers	1
Freezing Rain,Snow Grains	1
Name: count, dtype: int64	

```
In [123]: # second way is filtering
df[df["Weather Condition"] == "Snow"]

Out[123]:
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_kmh	Visibility_km	Press_kPa	Weather Condition
55	1/2/2012 7:00	-14.0	-19.5	63	19	25.0	100.35	Snow
84	1/4/2012 12:00	-13.7	-21.7	51	11	24.1		

