import pandas as pd
data = pd.read\_csv('/content/drive/MyDrive/Dataset/Project+1+-+Weather+Dataset.csv')
data.head()

<b>→</b>		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	11.
	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·nn	-1.5	-3.3	88	7	4.8	101.23	Fog	•

Étapes suivantes :

Générer du code avec data

Afficher les graphiques recommandés

New interactive sheet

head() --> shows the first N raws in data N=5 by default

data.head()

<b>→</b>		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	11.
	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	•

Étapes suivantes : (

Générer du code avec data

Afficher les graphiques recommandés

New interactive sheet

shape shows num of lines and colonnes of the dataframe

data.shape

**→** (8784, 8)

Index shows the index of the dataframe

data.index

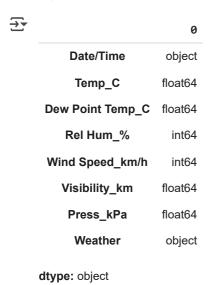
→ RangeIndex(start=0, stop=8784, step=1)

Columns shows the name of the columns

```
data.columns
```

Dtypes shows the datatype of each column

## data.dtypes

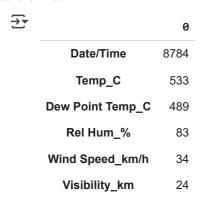


In a column it shows all the unique values, it can be applied on a single column only not on the whole dataframe

```
data['Weather'].unique()
```

nunique shows the total number of unique values in a columns or in a dataframe

```
data.nunique()
```



Press\_kPa

Weather

518

50

dtype: int64

data['Weather'].nunique()

**→** 50

count shows the number of non-null in each column as wel as on whole dataframe

data['Dew Point Temp\_C'].count()

₹ 8784

data.count()



value\_count() Shows all the unique values with their count on single column only

# Ce texte est au format code

data['Visibility\_km'].value\_counts()



count

Visibility_km	
25.0	3324
48.3	2014
24.1	1921
19.3	281
16.1	224
9.7	181
12.9	174
6.4	129
8.0	122
4.8	79
4.0	55
2.4	50
3.2	42
11.3	36
1.2	28
2.0	28
1.6	25
0.6	16
2.8	15
1.0	11
3.6	11
0.2	8
0.8	7
0.4	3

dtype: int64

info() provide a basic information about the dataframe

data.info()



<pr 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
d+vn	oc. £100+64/4\ in	+C1(2) object(2	\

dtypes: float64(4), int64(2), object(2)

memory usage: 549.1+ KB

## Qestion 1: find the all unique "Wind speed" values in the data

data['Wind Speed\_km/h'].nunique()

**→** 34

Find the number of times when the weather is exactly clear

data

	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
	12/31/2012							

Étapes suivantes :

Générer du code avec data

Afficher les graphiques recommandés

New interactive sheet

# values count

data.Weather.value\_counts()



## count

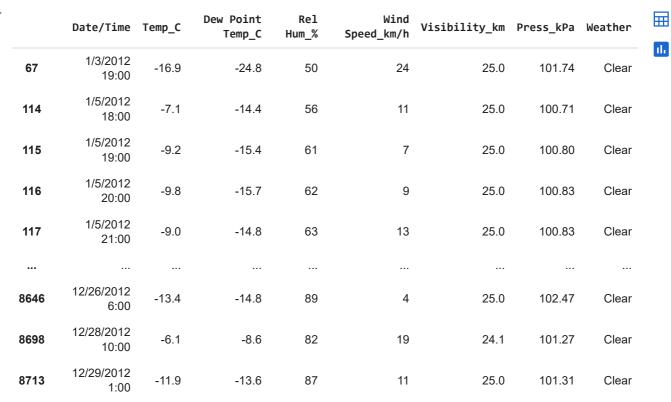
Weather	
Mainly Clear	2106
Mostly Cloudy	2069
Cloudy	1728
Clear	1326
Snow	390
Rain	306
Rain Showers	188
Fog	150
Rain,Fog	116
Drizzle,Fog	80
Snow Showers	60
Drizzle	41
Snow,Fog	37
Snow,Blowing Snow	19
Rain,Snow	18
Thunderstorms, Rain Showers	16
Haze	16
Drizzle,Snow,Fog	15
Freezing Rain	14
Freezing Drizzle,Snow	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 Thunderstorms, Heavy Rain Showers 1 Freezing Rain, Snow Grains 1

dtype: int64

data.groupby('Weather').get\_group('Clear')

	_	_
		÷
-	7	_



Find the number of times when the wind speed was exactly 4km/h

Commencez à coder ou à <u>générer</u> avec l'IA.

data[data['Wind Speed\_km/h'] == 4]



	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	-5 A	<sub>-</sub> 7 5	88	А	19 0	100 78	Snow

count = data['Wind Speed\_km/h'].value\_counts().get(4, 0)
print(count)

**→** 474

La propriété .shape retourne un tuple (nombre\_de\_lignes, nombre\_de\_colonnes).

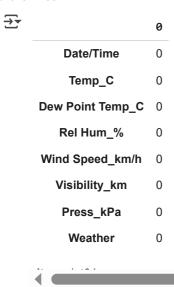
- $.shape[0] \rightarrow Nombre total de lignes$
- $.shape[1] \rightarrow Nombre total de colonnes$

count = data[data['Wind Speed\_km/h'] == 4].shape[0]
print(count)

**→** 474

Find out all the null values in the data

data.isnull().sum()



Rename the column name 'Weather' of the dataframe to 'Weather condition'

data.rename(columns={'Weather' : 'Weather Conditions'}, inplace= True)

data.head()

<b>→</b>		Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Conditions	
	0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog	11.
	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	•

Étapes suivantes :

Générer du code avec data

Afficher les graphiques recommandés

New interactive sheet

What is the mean visibility?

data.Visibility\_km.mean()

**27.664446721311478** 

The question "Find all the instances when 'Snow' was recorded" means:

Identify and extract all the rows in a dataset where the 'Weather' column contains the word "Snow"

data.groupby('Weather Conditions').get\_group('Snow')



	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Conditions
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	0.0	-2.1	86	30	11.3	99.89	Snow

by doing a filter

data[data['Weather Conditions'] == 'Snow']



	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Conditions
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	0.0	-2.1	86	30	11.3	99.89	Snow

using contains

data[data["Weather Conditions"].str.contains("Snow")]



	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Conditions
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
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	0.0	-2.1	86	30	11.3	99.89	Snow

Find all instances where wind speed si above 24 and visibility is 25

data[(data['Wind Speed\_km/h'] > 24) & (data['Visibility\_km'] == 25)]

<b>→</b>		Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Conditions	11.
	23	1/1/2012 23:00	5.3	2.0	79	30	25.0	99.31	Cloudy	
	24	1/2/2012 0:00	5.2	1.5	77	35	25.0	99.26	Rain Showers	
	25	1/2/2012 1:00	4.6	0.0	72	39	25.0	99.26	Cloudy	
	26	1/2/2012 2:00	3.9	-0.9	71	32	25.0	99.26	Mostly Cloudy	
	27	1/2/2012 3:00	3.7	-1.5	69	33	25.0	99.30	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	-12.1	-15.1	78	28	25.0	101.52	Mostly Cloudy	

What is the minimum and maximum value of each column against each weather condition

data.groupby('Weather Conditions').min()



	Date/Time	Temp_C	Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Weather Conditions							
Clear	1/11/2012 1:00	-23.3	-28.5	20	0	11.3	99.52
Cloudy	1/1/2012 17:00	-21.4	-26.8	18	0	11.3	98.39
Drizzle	1/23/2012 21:00	1.1	-0.2	74	0	6.4	97.84
Drizzle,Fog	1/23/2012 20:00	0.0	-1.6	85	0	1.0	98.65
Drizzle,Ice Pellets,Fog	12/17/2012 9:00	0.4	-0.7	92	20	4.0	100.79
Drizzle,Snow	12/17/2012 15:00	0.9	0.1	92	9	9.7	100.63
Drizzle,Snow,Fog	12/18/2012 21:00	0.3	-0.1	92	7	2.4	97.79
Fog	1/1/2012 0:00	-16.0	-17.2	80	0	0.2	98.31
Freezing Drizzle	1/13/2012 10:00	-9.0	-12.2	78	6	4.8	98.44
Freezing Drizzle,Fog	1/1/2012 2:00	-6.4	-9.0	82	6	3.6	98.74
Freezing Drizzle,Haze	2/1/2012 11:00	-5.8	-8.3	81	9	2.0	100.28
Freezing Drizzle,Snow	1/13/2012 3:00	-8.3	-10.4	79	6	2.4	99.19
Freezing Fog	1/22/2012 6:00	-19.0	-22.9	71	0	0.2	101.97
Freezing Rain	1/13/2012 11:00	-6.5	-9.0	81	7	2.8	98.22
Freezing Rain,Fog	1/17/2012 23:00	-6.1	-8.7	82	7	2.8	98.32
Freezing Rain,Haze	2/1/2012 14:00	-4.9	-7.5	82	6	2.0	100.34
Freezing Rain,Ice Pellets,Fog	12/17/2012 3:00	-2.6	-3.7	92	28	8.0	100.95
Freezing Rain,Snow Grains	1/13/2012 9:00	-5.0	-7.3	84	32	4.8	98.56
Haze	1/22/2012 12:00	-11.5	-16.0	68	0	4.8	100.35
Mainly Clear	1/10/2012 11:00	-22.8	-28.0	20	0	12.9	98.67
Moderate Rain,Fog	12/10/2012 8:00	1.7	0.8	94	17	6.4	99.98
Moderate Snow	1/12/2012 15:00	-6.3	-7.6	83	26	0.6	99.88
Moderate Snow,Blowing Snow	12/27/2012 10:00	-5.5	-6.6	92	39	0.6	100.50
h receive google com/drive/102h	1/1/2012	22 2 AlvO S IvV	20 E	40 0#00rollTr	^ Nr-218ITy_0\/s	44.2	UO 3E