Report

Q. 1) Find all the unique 'Wind Speed' values in the data.

Input : - wd1["Wind Speed\_km/h"].unique()

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

dtype=int64)

* Here, the code n unique () gives 34 number of unique values of a weather data of a windspeed column. The unique () gives the unique values of wind speed.

Q. 2) Find the number of times when the 'Weather is exactly Clear'.

Input :- wd1.Weather.value\_counts().Clear

Output :- 1326

* Here, the code represents the number of occurences of weather condition when “weather is exactly Clear”.

Q. 3) Find the number of times when the 'Wind Speed was exactly 4 km/h'.

​

Input :- wd1["Wind Speed\_km/h"].value\_counts()[4]

Output :- 474

* This code tells to select rows when the wind speed is 4 km/h and then counts the number of such rows, storing the count in the variable windspeed count.

Q. 4) Find out all the Null Values in the data.

Input :- wd1.isnull()

Output :-

|  | Date/Time | Temp\_C | Dew Point Temp\_C | Rel Hum\_% | Wind Speed\_km/h | Visibility\_km | Press\_kPa | Weather |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | False | False | False | False | False | False | False | False |
| 1 | False | False | False | False | False | False | False | False |
| 2 | False | False | False | False | False | False | False | False |
| 3 | False | False | False | False | False | False | False | False |
| 4 | False | False | False | False | False | False | False | False |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8779 | False | False | False | False | False | False | False | False |
| 8780 | False | False | False | False | False | False | False | False |
| 8781 | False | False | False | False | False | False | False | False |
| 8782 | False | False | False | False | False | False | False | False |
| 8783 | False | False | False | False | False | False | False | False |

8784 rows × 8 columns

* In code it provides to find out how many null values are there in each column of the data frame and counts the null values of each column. Here it returns all sum of null value columns in weather data.

Q. 5) Rename the column name 'Weather' of the dataframe to 'Weather Condition'.

Input :- dfwd=wd1.rename(columns={"Weather":"Weather Condition"})

Dfwd

Output :-

|  | Date/Time | Temp\_C | Dew Point Temp\_C | Rel Hum\_% | Wind Speed\_km/h | Visibility\_km | Press\_kPa | Weather Condition |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 01-01-2012 00:00 | -1.3 | -3.5 | 18 | 9 | 25.0 | 98.67 | Clear |
| 1 | 01-01-2012 01:00 | 7.4 | 2.8 | 20 | 24 | 24.1 | 99.37 | Rain |
| 2 | 01-01-2012 02:00 | 15.7 | 13.4 | 21 | 26 | 25.0 | 99.84 | Cloudy |
| 3 | 01-01-2012 03:00 | 4.9 | -2.6 | 27 | 15 | 24.1 | 100.94 | Mainly Clear |
| 4 | 01-01-2012 04:00 | -13.4 | -19.7 | 30 | 4 | 25.0 | 102.32 | Mostly Cloudy |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8779 | 9/30/2012 5:00 | 1.4 | -3.7 | 97 | 22 | 48.3 | 100.16 | Cloudy |
| 8780 | 9/30/2012 6:00 | -4.6 | -9.5 | 98 | 11 | 48.3 | 101.46 | Mostly Cloudy |
| 8781 | 9/30/2012 7:00 | 1.5 | -6.3 | 99 | 30 | 24.1 | 101.48 | Clear |
| 8782 | 9/30/2012 8:00 | -6.3 | -13.5 | 99 | 15 | 24.1 | 101.90 | Cloudy |
| 8783 | 9/30/2012 9:00 | 24.5 | 14.7 | 100 | 11 | 25.0 | 102.98 | Mostly Cloudy |

8784 rows × 8 columns

* This code tells renaming operation does not modify the original data frame but creates a new data frame with the modified column name. Here weather column is renamed as weather condition.

Q. 6) What is the mean 'Visibility' ?

Input :- wd1["Visibility\_km"].mean()

Output :- 27.664446721311478

* It computes the average visibility in kilometeres based on values present in the visibility column.

Q. 7) What is the Standard Deviation of 'Pressure' in this data?

Input :- wd1["Press\_kPa"].std()

Output :-0.8440047459486459

* In above code , it tells about the standard deviation of atmospheric pressure values in “pressure” column.

Q. 8) What is the Variance of 'Relative Humidity' in this data ?

Input :-wd1["Rel Hum\_%"].var()

Output :-286.2485501985015

* It calculates the variance of the relative humidity values in the “ Rel Hum\_%” column of the weather data. Variance provides insight into how much the humidity values deviate from their mean value , giving a measure of the dispersion or spread of the relative humidity datapoints.

Q. 9) Find all instances when 'Snow' was recorded.

Input :- wd2 = dfwd[dfwd["Weather Condition"] == "Snow"]

wd2

output :-

|  | Date/Time | Temp\_C | Dew Point Temp\_C | Rel Hum\_% | Wind Speed\_km/h | Visibility\_km | Press\_kPa | Weather Condition |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | 01-01-2012 11:00 | -6.2 | -9.6 | 37 | 35 | 4.8 | 101.56 | Snow |
| 70 | 03-01-2012 22:00 | -4.0 | -6.6 | 62 | 22 | 16.1 | 100.48 | Snow |
| 73 | 04-01-2012 01:00 | 2.3 | -3.4 | 64 | 35 | 25.0 | 103.43 | Snow |
| 105 | 05-01-2012 09:00 | -1.8 | -4.2 | 73 | 15 | 6.4 | 101.28 | Snow |
| 112 | 05-01-2012 16:00 | 1.7 | -0.3 | 75 | 6 | 9.7 | 101.47 | Snow |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8573 | 9/22/2012 13:00 | -6.0 | -10.2 | 82 | 19 | 16.1 | 101.71 | Snow |
| 8650 | 9/25/2012 18:00 | -4.6 | -6.6 | 52 | 4 | 12.9 | 100.48 | Snow |
| 8671 | 9/26/2012 15:00 | -5.9 | -10.5 | 60 | 13 | 16.1 | 101.01 | Snow |
| 8713 | 9/28/2012 1:00 | -5.2 | -7.8 | 72 | 33 | 4.0 | 101.33 | Snow |
| 8734 | 9/28/2012 8:00 | 0.7 | -1.2 | 79 | 30 | 8.0 | 101.22 | Snow |

390 rows × 8 columns

* In above it computes to find all rows where snow was recorded in weather column.

10) Find all instances when 'Wind Speed is above 24' and 'Visibility is 25'.

Input :- wd3 = dfwd[(dfwd["Wind Speed\_km/h"]>24) & (dfwd["Visibility\_km"]==25)]

wd3

output :-

|  | Date/Time | Temp\_C | Dew Point Temp\_C | Rel Hum\_% | Wind Speed\_km/h | Visibility\_km | Press\_kPa | Weather Condition |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 01-01-2012 02:00 | 15.7 | 13.4 | 21 | 26 | 25.0 | 99.84 | Cloudy |
| 73 | 04-01-2012 01:00 | 2.3 | -3.4 | 64 | 35 | 25.0 | 103.43 | Snow |
| 126 | 06-01-2012 06:00 | 10.0 | 5.4 | 77 | 39 | 25.0 | 101.30 | Cloudy |
|  |  |  |  |  |  |  |  |  |
| 158 | 07-01-2012 14:00 | 1.9 | -2.1 | 87 | 26 | 25.0 | 100.87 | Rain,Snow Grains |
| 184 | 08-01-2012 16:00 | 14.2 | 9.2 | 35 | 44 | 25.0 | 99.49 | Mostly Cloudy |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8707 | 9/27/2012 5:00 | -1.0 | -6.0 | 70 | 33 | 25.0 | 98.56 | Mostly Cloudy |
| 8714 | 9/28/2012 10:00 | 2.6 | 0.3 | 72 | 26 | 25.0 | 101.60 | Rain |
| 8738 | 9/29/2012 10:00 | 22.8 | 12.3 | 80 | 28 | 25.0 | 101.60 | Mostly Cloudy |
| 8745 | 9/29/2012 17:00 | -10.3 | -12.9 | 82 | 28 | 25.0 | 102.16 | Cloudy |
| 8776 | 9/30/2012 23:00 | 19.2 | 13.2 | 93 | 43 | 25.0 | 101.60 | Mainly Clear |

308 rows × 8 columns

* The resulting data frame contains only the rows where the windspeed is greater than 24 km/h and visibility is 25km.

Q. 11) What is the Mean value of each column against each 'Weather Condition ?

Input :- dfwd.mean(numeric\_only = True)

Output :-

Temp\_C 8.798144

Dew Point Temp\_C 2.555294

Rel Hum\_% 67.431694

Wind Speed\_km/h 14.945469

Visibility\_km 27.664447

Press\_kPa 101.051623

dtype: float64

* It is a data frame with column would calculate the mean only and provide a Series with the mean values for those columns..

Q. 12) What is the Minimum & Maximum value of each column against each 'Weather Condition ?

Input :-dfwd.iloc[:,0:7].agg([max,min])

Output :-

|  | Date/Time | | Temp\_C | | Dew Point Temp\_C | | Rel Hum\_% | | Wind Speed\_km/h | | Visibility\_km | | Press\_kPa |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| max | 9/30/2012 9:00 | | 33.0 | | 24.4 | | 100 | | 83 | | 48.3 | | 103.65 |
| min | 01-01-2012 00:00 | | -23.3 | | -28.5 | | 18 | | 0 | | 0.2 | | 97.52 |
| * The result is a data frame where each column contains two rows-one representing the minimum value and the other representing the maximum value for corresponding column “weather”. | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |

Q. 13) Show all the Records where Weather Condition is Fog.

Input :-df5=dfwd[dfwd["Weather Condition"]=="Fog"]

df5

output :-

|  | Date/Time | Temp\_C | Dew Point Temp\_C | Rel Hum\_% | Wind Speed\_km/h | Visibility\_km | Press\_kPa | Weather Condition |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 13 | 01-01-2012 13:00 | 9.5 | 7.8 | 40 | 13 | 6.4 | 100.90 | Fog |
| 53 | 03-01-2012 05:00 | -3.6 | -4.3 | 57 | 7 | 9.7 | 101.32 | Fog |
| 136 | 06-01-2012 16:00 | 14.8 | 13.5 | 80 | 19 | 9.7 | 100.86 | Fog |
| 197 | 09-01-2012 05:00 | 2.1 | 0.7 | 43 | 11 | 8.0 | 101.44 | Fog |
| 278 | 12-01-2012 14:00 | 1.2 | 0.6 | 70 | 13 | 6.4 | 103.22 | Fog |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8475 | 9/18/2012 11:00 | 6.2 | 5.4 | 56 | 7 | 4.8 | 102.03 | Fog |
| 8511 | 9/19/2012 22:00 | 15.7 | 15.4 | 66 | 7 | 8.0 | 101.93 | Fog |
| 8518 | 9/19/2012 8:00 | -2.9 | -4.5 | 68 | 6 | 6.4 | 100.41 | Fog |
| 8537 | 9/20/2012 3:00 | -0.5 | -2.1 | 74 | 7 | 4.0 | 100.81 | Fog |
| 8771 | 9/30/2012 19:00 | 12.8 | 12.2 | 91 | 19 | 4.8 | 100.60 | Fog |

150 rows × 8 columns

* It interprets a data frame containing only the rows where the weather condition is “fog”. It shows all the rows where weather condition is fog.

Q. 14) Find all instances when 'Weather is Clear' or 'Visibility is above 40'.

Input :- df6 =wd1[["Weather","Visibility\_km"]][(wd1["Weather"] == "Clear") & (wd1["Visibility\_km"] > 40)]

df6

output :-

|  | Weather | Visibility\_km |
| --- | --- | --- |
| 9 | Clear | 48.3 |
| 19 | Clear | 48.3 |
| 40 | Clear | 48.3 |
| 51 | Clear | 48.3 |
| 87 | Clear | 48.3 |
| ... | ... | ... |
| 8640 | Clear | 48.3 |
| 8644 | Clear | 48.3 |
| 8658 | Clear | 48.3 |
| 8693 | Clear | 48.3 |
| 8719 | Clear | 48.3 |

ws × 2 columns

* The resulting data frame contains only the rows where the weather condition is “Clear” or the visibility is greater than 40 km

Q. 15) Find all instances when :

A. 'Weather is Clear' and 'Relative Humidity is greater than 50'

or

B. 'Visibility is above 40'

Input :-df7 = wd1[["Weather","Rel Hum\_%","Visibility\_km"]][(wd1["Weather"] == "Clear") & (wd1["Rel Hum\_%"] > 50) | (wd1["Visibility\_km"] > 40) ]

df7

output :-

|  | Weather | Rel Hum\_% | Visibility\_km |
| --- | --- | --- | --- |
| 9 | Clear | 35 | 48.3 |
| 17 | Mainly Clear | 42 | 48.3 |
| 18 | Cloudy | 42 | 48.3 |
| 19 | Clear | 43 | 48.3 |
| 23 | Mainly Clear | 45 | 48.3 |
| ... | ... | ... | ... |
| 8774 | Mostly Cloudy | 92 | 48.3 |
| 8777 | Mainly Clear | 95 | 48.3 |
| 8779 | Cloudy | 97 | 48.3 |
| 8780 | Mostly Cloudy | 98 | 48.3 |
| 8781 | Clear | 99 | 24.1 |

s × 3 columns

​

* It represents a data frame containing only the rows where either the weather condition is “ Clear ” and relative humidity is greater than 50% or the visibility is greater than 40km.
* It computes data frame containing only the rows from the "Weather", " Rel Hum\_%", and "Visibility\_ km" columns where either the weather is 'Clear' and relative humidity is greater than 50% or visibility is greater than 40 km.