**Data analysis report of the weather dataset**

**Aim:**

The goal of this project is to get a experience of using python and it’s libraries like numpy,pandas and matplotlib to import, clean, analysis and visualize the data.

**Data Description:**

The name of the dataset is ‘Weather Data’. This dataset contains about the weather information per hour of a certain location .In this dataset the table contains the date & time, temperature, dew point temperature, relative humidity, pressure, visibility, wind speed and weather type of each hour .The data is imported from kaggle .The dataset is in .csv format.

The data set contains 8784 rows and 8 columns .The dataset contains 70272 values .There is no missing and null values present in the data. The table of columns and it’s data type is following:

Date/Time object

Temp\_C float64

Dew Point Temp\_C float64

Rel Hum\_% int64

Wind Speed\_km/h int64

Visibility\_km float64

Press\_kPa float64

Weather object

Here we can see that date/time is not in datatime type so we have changed the datatype of this column.

**Methodology:**

We have used many function of the python,it’s libraries and their functions to import ,clean, analyses and visualize the data. They are following:

1. Numpy,pandas and matplotlib libraries 18. get\_group()
2. pd.read\_csv() : to import the dataset 19. rename()
3. head() 20. mean()
4. tail() 21. sum()
5. shape 22. iloc[]
6. size 23. round()
7. columns 24. std()
8. dtypes 25.var()
9. astype() 26.str.contains()
10. nuinque 27.min()
11. unique 28.max()
12. isnull() 29.dt.month
13. value\_counts() 30.dt.date
14. info() 31.plt.plot()
15. describe() 32.plt.bar()
16. duplicated() 33.plt.title()
17. groupby() 34.plt.xlabel()

35.plt.ylabel()

36.plt.show()

37.plt.figures()

38.plt.grid()

**Findings:**

After cleaning and exploring the data we find the result of the following questions from the dataset:

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

2. Find no. of times when weather was exactly clear.

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

4. Rename the column 'weather' to 'weather condition'.

5. What is mean of the visibility?

6. What is the standard deviation of the pressure column?

7. What is the variance of the "Relative Humidity"?

8. Find all the instance when "Snow" was recorded.

9. Find all the instance where "Wind speed is above than 24" and "Visibility is 25".

10. Find the mean of each column on the basis of the weather.

11. What is the maximum and minimum of each column against each weather condition?

12. Find all the records where weather condition is fog.

13. Find all the instance where "Weather is Clear" or 'Visibility is above 40'.

14. Find the instances where weather is clear and relative humidity is above 50 or visibility is above 40 .

15. Plot a graph to show the average temperature of the each month.

16. Plot the graph for the average of the wind speed for each month.

17. Find the hottest day in July.

18. Find the coolest day of the year.

19. Find the hottest day of the year.

**Results:**

All the results of the above questions have been defined in this pdf : [result\_with\_code](https://drive.google.com/file/d/1yeFmvnKKR_zQW-gjkj4VHca8tSUiBkps/view?usp=sharing)

**Discussion:**

From the data we get to know that there are 533 values for temperature, 50 unique weather conditions, 83 values for relative humidity %, 34 unique values for wind speed have been recorded in the data set. The weather condition was clear 2106 times, snow for 390 times and rain for 306 times have occurred. The average temperature throughout the year was 8.798 degree Celsius and minimum temperature in the whole year was -23.3 that occurred at 15 January 2012 at 8:00 AM and the maximum temperature was 33 degree Celsius that was recorded at 21 June 2012 at 9:00 PM .July was the month when the average temperature was the highest and in January we were having the lowest average temperature .In the July the hottest day was 14th July with temperature 28 degree Celsius .Wind speed was highest in the month of January and lowest in July so we can say that when temperature was lowest then wind speed was highest and when temperature was highest then wind speed was lowest.