Weather Report Forecasting Analysis

Import Required Libraries

In [4]:
```
import pandas as pd
import numpy as np
```

Upload Data

In [5]:
```
Weather_D=pd.read_csv("weather_data.csv",sep=",")
```

Checking The Data

In [3]:
```
Weather_D.head()
```

Out[3]:

	Date	Temperature	Average humidity	Average dewpoint	Average barometer	Average windspeed	Average gustspeed	Average direction	Rainfall for month	Rainfall for year	...	Maximum humidity	Minimum humidity	Maximum pressure
0	01-01-2022	37.8	35	12.7	29.7	26.4	36.8	274	0.0	0.0	...	4	27	29.762
1	02-01-2022	43.2	32	14.7	29.5	12.8	18.0	240	0.0	0.0	...	4	16	29.669
2	03-01-2022	25.7	60	12.7	29.7	8.3	12.2	290	0.0	0.0	...	8	35	30.232
3	04-01-2022	9.3	67	0.1	30.4	2.9	4.5	47	0.0	0.0	...	7	35	30.566
4	05-01-2022	23.5	30	-5.3	29.9	16.7	23.1	265	0.0	0.0	...	5	13	30.233

5 rows × 23 columns

In [4]:
```
Weather_D.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3903 entries, 0 to 3902
Data columns (total 23 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Date                     3903 non-null   object
 1    Temperature             3903 non-null   float64
 2    Average humidity        3903 non-null   int64
 3    Average dewpoint        3903 non-null   float64
 4    Average barometer       3903 non-null   float64
 5    Average windspeed       3903 non-null   float64
 6    Average gustspeed       3903 non-null   float64
 7    Average direction       3903 non-null   int64
 8    Rainfall for month      3903 non-null   float64
 9   Rainfall for year        3903 non-null   float64
 10   Maximum rain per minute 3903 non-null   int64
 11   Maximum temperature     3903 non-null   float64
 12   Minimum temperature     3903 non-null   float64
 13   Maximum humidity        3903 non-null   int64
 14   Minimum humidity        3903 non-null   int64
 15   Maximum pressure        3903 non-null   float64
 16   Minimum pressure        3903 non-null   float64
 17   Maximum windspeed       3903 non-null   float64
 18   Maximum gust speed      3903 non-null   float64
 19   Maximum heat index      3903 non-null   float64
 20  Date_N                   3903 non-null   object
 21   Month                   3903 non-null   int64
 22   diff_pressure           3903 non-null   float64
dtypes: float64(15), int64(6), object(2)
memory usage: 701.4+ KB
```

In [5]:
```
Weather_D.dtypes
```

Out[5]:
```
Date                         object
 Temperature                float64
 Average humidity             int64
 Average dewpoint           float64
 Average barometer          float64
 Average windspeed          float64
 Average gustspeed          float64
 Average direction            int64
 Rainfall for month         float64
Rainfall for year           float64
 Maximum rain per minute      int64
 Maximum temperature        float64
 Minimum temperature        float64
 Maximum humidity             int64
 Minimum humidity             int64
 Maximum pressure           float64
 Minimum pressure           float64
 Maximum windspeed          float64
 Maximum gust speed         float64
 Maximum heat index         float64
Date_N                       object
 Month                        int64
 diff_pressure              float64
dtype: object
```

In [6]:
```
Weather_D.columns
```

Out[6]:
```
Index(['Date', ' Temperature', ' Average humidity ', ' Average dewpoint ',
       ' Average barometer ', ' Average windspeed ', ' Average gustspeed ',
       ' Average direction ', ' Rainfall for month ', 'Rainfall for year ',
       ' Maximum rain per minute', ' Maximum temperature ',
       ' Minimum temperature ', ' Maximum humidity ', ' Minimum humidity ',
       ' Maximum pressure', ' Minimum pressure', ' Maximum windspeed ',
       ' Maximum gust speed ', ' Maximum heat index', 'Date_N', ' Month',
       ' diff_pressure'],
      dtype='object')
```

In [13]:
```
Weather_D.isnull().sum()
```

Out[13]:
```
Date                         0
 Temperature                 0
 Average humidity            0
 Average dewpoint            0
 Average barometer           0
 Average windspeed           0
 Average gustspeed           0
 Average direction           0
 Rainfall for month          0
Rainfall for year            0
 Maximum rain per minute     0
 Maximum temperature         0
 Minimum temperature         0
 Maximum humidity            0
 Minimum humidity            0
 Maximum pressure            0
 Minimum pressure            0
 Maximum windspeed           0
 Maximum gust speed          0
 Maximum heat index          0
Date_N                       0
 Month                       0
 diff_pressure               0
dtype: int64
```

Cleaning The Data

Subtask 2: removal of duplicate rows and duplicate Columns

In [14]:
```
Weather_D = Weather_D.drop_duplicates()
```

In [15]:
```
Weather_D.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3902 entries, 0 to 3901
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Date                      3902 non-null   object 
 1    Temperature              3902 non-null   float64
 2    Average humidity         3902 non-null   int64  
 3    Average dewpoint         3902 non-null   float64
 4    Average barometer        3902 non-null   float64
 5    Average windspeed        3902 non-null   float64
 6    Average gustspeed        3902 non-null   float64
 7    Average direction        3902 non-null   int64  
 8    Rainfall for month       3902 non-null   float64
 9   Rainfall for year         3902 non-null   float64
 10   Maximum rain per minute  3902 non-null   int64  
 11   Maximum temperature      3902 non-null   float64
 12   Minimum temperature      3902 non-null   float64
 13   Maximum humidity         3902 non-null   int64  
 14   Minimum humidity         3902 non-null   int64  
 15   Maximum pressure         3902 non-null   float64
 16   Minimum pressure         3902 non-null   float64
 17   Maximum windspeed        3902 non-null   float64
 18   Maximum gust speed       3902 non-null   float64
 19   Maximum heat index       3902 non-null   float64
 20  Date_N                    3902 non-null   object 
 21   Month                    3902 non-null   int64  
 22   diff_pressure            3902 non-null   float64
dtypes: float64(15), int64(6), object(2)
memory usage: 731.6+ KB
```

Subtask 3: fix a few labels in the given data set

It can be clearly seen that all the labels are not fixed. Most of them are missing ')' in the end . We need to fix it.

In [16]:
```
cols = Weather_D.columns.tolist()
```

In [17]:
```
cols
```

Out[17]:
```
['Date',
 ' Temperature',
 ' Average humidity ',
 ' Average dewpoint ',
 ' Average barometer ',
 ' Average windspeed ',
 ' Average gustspeed ',
 ' Average direction ',
 ' Rainfall for month ',
 'Rainfall for year ',
 ' Maximum rain per minute',
 ' Maximum temperature ',
 ' Minimum temperature ',
 ' Maximum humidity ',
 ' Minimum humidity ',
 ' Maximum pressure',
 ' Minimum pressure',
 ' Maximum windspeed ',
 ' Maximum gust speed ',
 ' Maximum heat index',
 'Date_N',
 ' Month',
 ' diff_pressure']
```

In [18]:
```
for i, col in enumerate (cols):
    if '('in col:
        cols[i] = col + ")"
        Weather_D.columns = cols
        print(Weather_D.columns)
```

In [20]:
```
Weather_D.columns
```

Out[20]:
```
Index(['Date', ' Temperature', ' Average humidity ', ' Average dewpoint ',
       ' Average barometer ', ' Average windspeed ', ' Average gustspeed ',
       ' Average direction ', ' Rainfall for month ', 'Rainfall for year ',
       ' Maximum rain per minute', ' Maximum temperature ',
       ' Minimum temperature ', ' Maximum humidity ', ' Minimum humidity ',
       ' Maximum pressure', ' Minimum pressure', ' Maximum windspeed ',
       ' Maximum gust speed ', ' Maximum heat index', 'Date_N', ' Month',
       ' diff_pressure'],
      dtype='object')
```

In [21]:
```
Weather_D.to_csv('Clean_weather_Dataset.csv' , encoding = 'utf-8' , index=False)
```

In [22]:
```
Weather_D.to_csv('Clean_weather_Dataset.xlsx' , index=False)
```

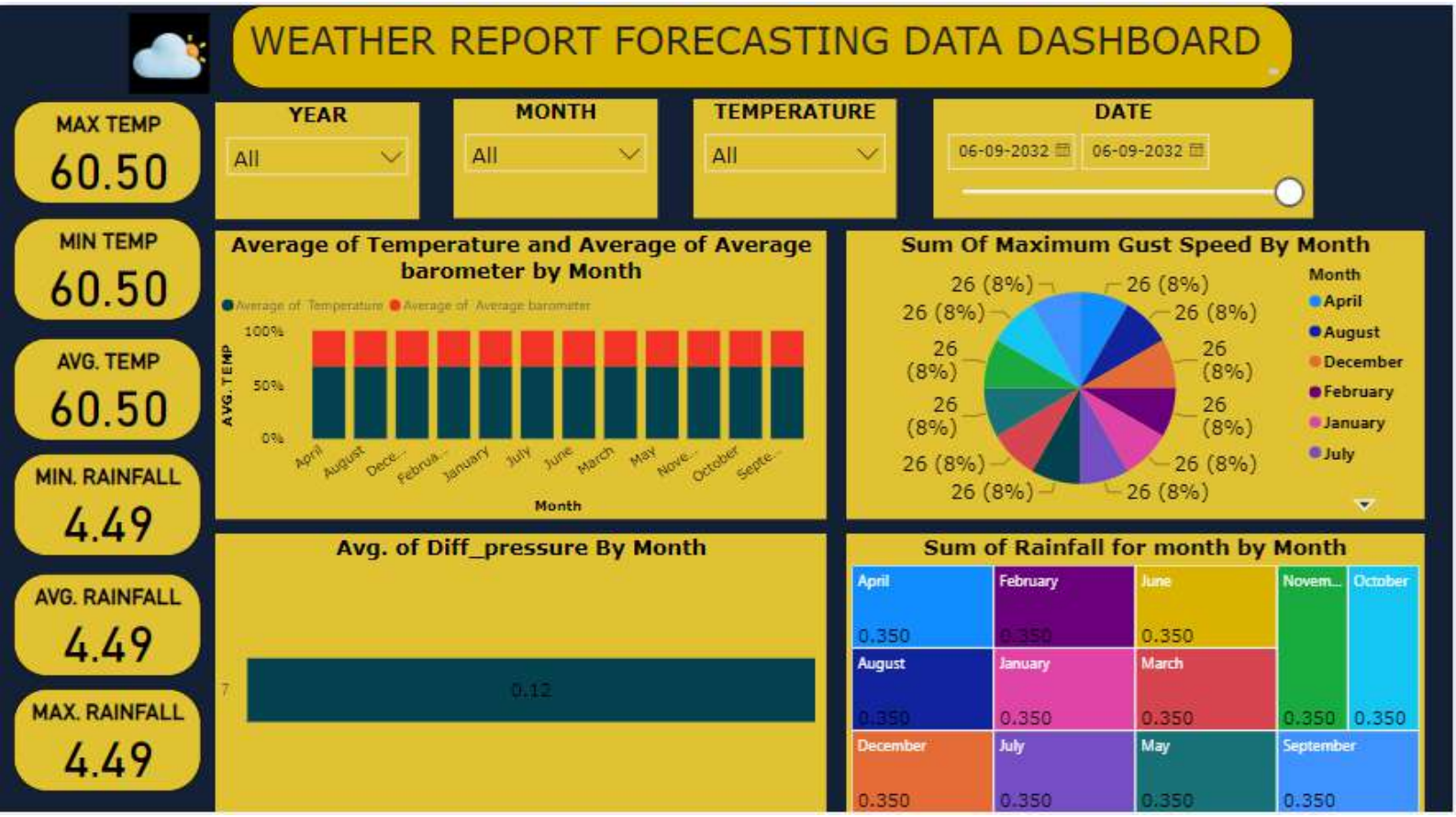
In [ ]:

```
CREATE DATABASE WEATHER;
USE  WEATHER;
DROP Table weather ;
CREATE TABLE WEATHER(
`Date` VARCHAR(20) ,
Temperature  FLOAT NOT NULL,
Average_humidity INT NOT NULL,
Average_dewpoint FLOAT NOT NULL,
Average_barometer      FLOAT NOT NULL ,
Average_windspeed  FLOAT NOT NULL,
Average_gustspeed  FLOAT NOT NULL,
Average_direction       INT NOT NULL,
Rainfall_for_month FLOAT NOT NULL,
Rainfall_for_year  FLOAT NOT NULL,
Maximum_rain_per_minute       INT NOT NULL,
Maximum_temperature  FLOAT NOT NULL,
Minimum_temperature     FLOAT NOT NULL,
Maximum_humidity INT NOT NULL,
Minimum_humidity INT NOT NULL,
Maximum_pressure  FLOAT NOT NULL,
Minimum_pressure  FLOAT NOT NULL,
Maximum_windspeed       FLOAT NOT NULL,
Maximum_gust_speed      FLOAT NOT NULL,
Maximum_heat_index     FLOAT NOT NULL,
`Month`      INT NOT NULL,
diff_pressure INT NOT NULL);

ALTER TABLE  WEATHER MODIFY Maximum_heat_index DECIMAL (5,2);
ALTER TABLE  WEATHER MODIFY Minimum_temperature DECIMAL (5,2);
ALTER TABLE  WEATHER MODIFY Maximum_temperature DECIMAL (5,2);
ALTER TABLE  WEATHER MODIFY Average_dewpoint DECIMAL (5,2);
ALTER TABLE  WEATHER MODIFY Temperature DECIMAL (5,2);

load data infile
'D:\Clean_weather_Dataset.csv'
into table WEATHER
fields terminated by ','
enclosed by '"'
lines terminated by '\n'
ignore 1 rows;
SET SESSION sql_mode = '';
SELECT * FROM WEATHER;

-- 1.Give the count of the minimum number of days for the time when
temperature reduced --

SELECT COUNT(*) FROM (
  SELECT Date, Temperature, LAG(Temperature) OVER(ORDER BY DATE) AS
prev_temp
  FROM WEATHER
) subquery_alias
WHERE Temperature < prev_temp;
```

```
-- 2.Find the temperature as Cold / hot by using the case and avg of
values of the given data set --
SELECT
date,
Temperature,
Avg(Temperature) As avg_temp,
CASE
WHEN Temperature > 25 THEN 'HOT'
ELSE 'COLD'
END AS HOTorCOLD
FROM WEATHER
GROUP BY date;

-- 3.Can you check for all 4 consecutive days when the temperature was
below 30 Fahrenheit --

CREATE TEMPORARY TABLE t1 SELECT Date, Temperature,
      SUM( CASE WHEN Temperature < 30 THEN 1 ELSE 0 END)
        OVER(ORDER BY Date ROWS BETWEEN 3 preceding AND CURRENT ROW) AS
below_30_count
FROM WEATHER;
SELECT date , Temperature FROM T1 WHERE below_30_count = 4;

-- 4.Can you find the maximum number of days for which temperature dropped
--

 SELECT MAX(count_days) FROM (
    SELECT t1.Date, t1.temperature,
        (
            SELECT COUNT(*)
            FROM WEATHER t2
            WHERE t2.Date < t1.Date AND t2.Temperature < t1.Temperature
        ) AS count_days
    FROM WEATHER t1
) AS temp_diff
WHERE Temperature < (
    SELECT Temperature FROM WEATHER t2
    WHERE t2.Date < temp_diff.Date
    ORDER BY Date DESC
    LIMIT 1
);

with weather_data_temp as (select date, temperature, lag(temperature, 1)
over (order by date) as prev_temp from weather)
select Date, temperature, prev_temp, case when prev_temp <= temperature
then 0 else 1
end as drop_temp_indicator,
sum(case when prev_temp <= temperature then 0 else 1
end) over (order by date rows between unbounded preceding and current row
) as drop_temp from weather order by date;


-- 5.Can you find the average of average humidity from the dataset --
```

```
-- ( NOTE: should contain the following clauses: group by, order by, date
) --

select AVG(avg_humidity) AS avg_of_avg_hum
from
    (select
    date,Avg(Average_Humidity) as avg_humidity
    from
    WEATHER
    GROUP BY date
    ORDER BY date) as derivetable;

   --  6.Use the GROUP BY clause on the Date column and make a query to
fetch details for average windspeed --
  --  ( which is now windspeed done in task 3 ) --

select
date, Average_Windspeed as avg_wind_speed
from
WEATHER
GROUP BY date
limit 25;

-- 7.Please add the data in the dataset for 2034 and 2035 as well as
forecast predictions for these years --
-- ( NOTE: data consistency and uniformity should be maintained ) --


 -- 8.If the maximum gust speed increases from 55mph, fetch the details
for the next 4 days --

 SELECT *
FROM WEATHER
WHERE date >= (
    SELECT MIN(Date)
    FROM Weather
    WHERE Maximum_gust_speed >= 5
)
AND date <= (
    SELECT MIN(Date) + INTERVAL 4 DAY
    FROM WEATHER
    WHERE Maximum_gust_speed >= 55
);


-- 9.Find the number of days when the temperature went below 0 degrees
Celsius --

select
count(temperature) as number_of_days
from
(select date, temperature
from WEATHER
where temperature < 0 ) as temp;
```

```
-- 10.Create another table with a â€œForeign keyâ€ relation with the
existing given data set.--

ALTER TABLE WEATHER ADD COLUMN id INT AUTO_INCREMENT PRIMARY KEY FIRST;

create table WEATHER_D (
OrderId int NOT NULL ,
OrderNumber int NOT NULL,
id int,
Primary Key (OrderId),
Foreign Key (id) REFERENCES WEATHER(id)
);
```


WEATHER REPORT FORECASTING DATA DASHBOARD
MAX TEMP
60.50
MIN TEMP
60.50
AVG. TEMP
60.50
MIN. RAINFALL
4.49
AVG. RAINFALL
4.49
MAX. RAINFALL
4.49
YEAR
All
MONTH
All
TEMPERATURE
All
DATE
06-09-2032
06-09-2032
Average of Temperature and Average of Average barometer by Month
Average of Temperature
Average of Average barometer
100%
50%
0%
AVG. TEMP
April
August
Dece...
Februa...
January
July
June
March
May
Nove...
October
Septe...
Month
Sum Of Maximum Gust Speed By Month
26 (8%)
Month
April
August
December
February
January
July
Avg. of Diff_pressure By Month
0.12
Sum of Rainfall for month by Month
April
0.350
February
0.350
June
0.350
Novem...
October
0.350
August
0.350
January
0.350
March
0.350
December
0.350
July
0.350
May
0.350
September
0.350