



Weather report forecasting Analysis & Graphical Speculation on Dashboard

About the project:

Weather prediction is one of the most certainly required information in all over the regions

It involves collecting global meteorological surface and upper-air observations, preparing global surface and upper air pressure, temperature, moisture, and wind analyses at frequent time intervals based upon these observations we predict some data for upcoming days weather conditions

Project Description:

This data contains day wise weather attributes from 2022 to July 2033 (predicted data)

Columns are as follows :

Date

Average temperature (°F)

Average humidity (%)

Average dewpoint (°F)

Average barometer (in)

Average windspeed (mph)

Average gust speed (mph)

Average direction (°degree)

Rainfall for month (in)

Rainfall for year (in)

Maximum rain per minute

Maximum temperature (°F)

Minimum temperature (°F)

Maximum humidity (%)

Minimum humidity (%)

Maximum pressure
Minimum pressure
Maximum wind speed (mph)
Maximum gust speed (mph)
Maximum heat index (°F)

NOTE: *the given data requires cleaning and analysis to proceed further for any analysis.*

The Project will consist of 2 modules:

Module 1: Pre-processing, Analysing data using Python and SQL.

Module 2: Visualizing data using Power bi.

Prerequisites for the Project (mandatory):

1. SQL (MYSQL)
2. POWER BI/ Tableau
3. Excel
4. Python

Module 1: Pre-processing, Analysing data using SQL

In this module, you will query the dataset using structured query language to gain insights from the database. The problem statements to be solved will be provided to you and you need to provide the solution for the same using your logic. Different concepts of SQL will be used in this process such as aggregating the data, grouping the data, ordering the data, etc. Module 1 consists of subtasks which are as follows

➤ **Task 1: Pre-processing the data**

Data preprocessing is one of the important steps in data analytics because data that is not processed can lead to different unwanted results when the data is used for further applications. This task includes sub-tasks such as handling null values, deletion or transformation of irrelevant values, data type transformation, removing duplicates, etc. The tasks to be performed for cleaning the data set are given below:

***Here is the link to the data set for the module: [Click here](#)**

- ***Subtask 1: Correct years for given data set***
- ***Subtask 2: removal of duplicate rows and duplicate Columns***
- ***Subtask 3: fix a few labels in the given data set***
- ***Subtask 4: Encoding data into suitable format***

The dataset we used in this project was populated using a scraping technique, the encoding method may be different; we prefer UTF-8 encoding as it is majorly used in all the database servers. To do so, perform the below tasks:

1. In the menu bar of excel, select option data. Then select from text/CSV and choose the file in which you made changes using excel
2. The default file origin will be UTF -8. Keep all the field as it is and select load
3. Save the file as CSV(comma-delimited) file.
4. Import the file into your SQL database

(Analyze the dataset with respect to your database system and then change it, if necessary).

Go through both cleaned datasets for a perfect understanding:

Since you have analyzed the complete data set accurately, Let's answer the following requirements to proceed further for a better understanding

1. Give the count of the minimum number of days for the time when temperature reduced
2. Find the temperature as Cold / hot by using the case and avg of values of the given data set
3. Can you check for all 4 consecutive days when the temperature was below 30 Fahrenheit
4. Can you find the maximum number of days for which temperature dropped
5. Can you find the average of average humidity from the dataset

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

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

10. Create another table with a “Foreign key” relation with the existing given data set.

Please make the Module 1 submission in the form: [click here](#).

Module 2: Visualizing data using Power BI.

Here comes the next module. You will use the weather datasets and the data table you created to visualize what the numbers in and behind the data want to convey. You have to create a dashboard for the same using different statistical graphs and diagrams for a visual understanding and analysis. Given below are the requirements, create a dashboard consisting of mentioned visuals.

1. Design a dashboard of your choice and imagination with attractive wallpapers and design visuals
2. The dashboard must consist of basic power bi visuals like stacked bar charts, cards, line charts, pie charts, etc.
3. You can use any visual capable of representing the given dataset.
4. Columns having numeric data types are to be majorly used in visuals
5. Include the Weather dataset and the data table you created in the above task (question 10).

Weekly Module Progress Submission Process:

Each module is to be completed in one week. Submissions will be on a weekly basis. Module 1 submission is to be done in the first week, and module 2 submission in the next consecutive week.

Guidelines for submissions are mentioned below:

1. A [google form](#) will be circulated for module file submissions.
2. For module 1, you must submit 4 files as mentioned below:
 1. A Text file consisting of the solutions to each SQL requirement.
 2. .py file used for cleaning the dataset
 3. Cleaned .csv file of playstore_apps dataset
 4. Cleaned .csv file of playstore_reviews dataset

**create a zip file of all 4 files and then submit.*

3. A Youtube video link that consists of a screen-recorded solution of the query. (Run the query and display its solution).

For module 2, you have to submit:

1. A power bi file (.pbix)
2. A Youtube video link consists of a screen-recorded solution of the designed dashboard.

Evaluation Criteria:

1. The project submission guidelines and prerequisites are mandatory and non-negotiable.
2. The submissions that are completed will be considered for evaluation.

Session Schedule & Timelines:

Day	Date	Time	Session
Monday	6th February	6:00 PM PST	Project Kickoff Call/ Introduction to Module 1
Sunday	12 th February	6:00 PM PST	Module 1 Submission Deadline
Monday	13 th February	6:00 PM PST	Introduction to Module 2
Sunday	19 th February	6:00 PM PST	Module 2 submission Deadline
Wednesday	24th Feb	6:00 PM PST	Top Teams Announcement

**Any changes in the dates and time of the sessions will be informed in advance.*

Important Links related to the Live Project:

1. Data Live Project Kick Off/ Introduction Call 6th Feb '23 6:00 PM PST- [click here to register and join](#)
2. Module Submission google form: [click here](#)
3. Live Project Doubts and queries: [click here](#)
4. Doubt Clearing Session for Module 1:
5. Introduction to Module 2:
6. Deck used during the Kickoff Session: Click here

