Using MongoDB Analyse weather data from [VALENTIA OBSERVATORY](https://www.met.ie/about-us/our-history/valentia-observatory) co. Kerry

Bharathi Ramapatnam

StudentNo:A00304671

A house near a lake

Description automatically generated

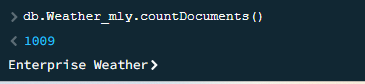
* **Dataset is taken form** [**gov.ie**](https://data.gov.ie/dataset/valentia-observatory-monthly-data)**. csv contains 1009 documents.**
* **Dataset contains weather information from year 1939 till 2023.**
* **1939 and 2023 are not full year, only three months of data in 1939 and 10 months in 2023.**
* **There are 128 records missing in the sunchine\_duration, mean\_wind and grass\_min\_temp.**
* **Colum names are re-named using MongoDB to be more meaningful.**

1. **Import csv file through mongo shell :**

**A black screen with white text

Description automatically generated**

1. **Find how many records in the collection :**

****

1. **Rename Column name using UpdateMany**

A screen shot of a computer code

Description automatically generated

1. **Find a single record to check column re-name :**

Before re-name column names :

A computer screen shot of numbers and letters

Description automatically generated

After Re-name column names :

A screen shot of a computer program

Description automatically generated

1. **How many years of Data we have in the csv**? : We have monthly data from year 1939 till 2023

A screenshot of a computer code

Description automatically generated

1. **Count number of rows in each year to identify if we have full 12 months in each year.**

Year 1939 is not a full year. We only have 3 months of data in this year.

Year 2023 is not a full year we only have 10 months of data in this year.

**A computer code with red and blue text

Description automatically generated**

**A screenshot of a computer code

Description automatically generated**

**A screen shot of a computer code

Description automatically generated**

1. **What is the max month covered in Y2023 ? :** Dataset has data up Until month 10 in the current year. We can see there is null value in the Sunshine\_duration field. Let find all the null values.

**A screen shot of a computer program

Description automatically generated**

1. **Find all the documents in year 2023 order by month ASC:** to see all the months data in year 2023. Output gives 10 records.

**A computer screen shot of a program

Description automatically generated**

1. **Find the count of no value items in sunshine\_duration in the entire collection.**

There are 124 documents in the entire collection with no values in sunshine\_duration.

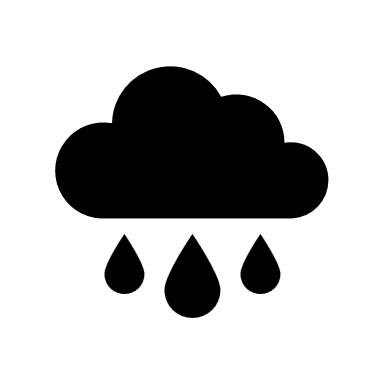
**A blue background with white text

Description automatically generated**

1. **Similarly find the missing values in grass min and mean windspeed:** There are two observations in each column with no values in grass\_min\_temp and mean\_wind\_speed .

**A blue screen with white text

Description automatically generated**

****

**Lets analyse the Rain fall with in the dataset.**

1. **In which year and month was the lowest and maximum rainfall recorded ?** for this filter records for year, month and rain sort by rain asc.

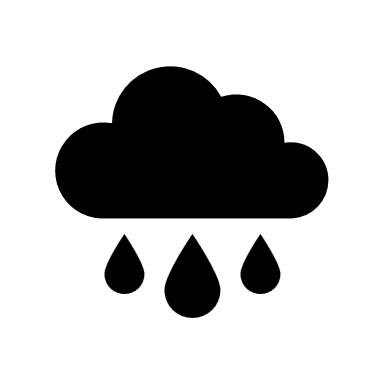
Lowest rainfall of 2.3mm was recorded in February 1965 .

Highest rainfall of 360.3mm was recorded in November 2009.

**A computer screen shot of a computer code

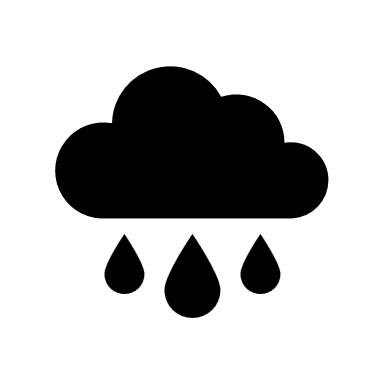
Description automatically generated**

1. **What is highest rainfall in the past 5 years? :** In the last five years Y2022/ November received highest rainfall of 354.8mm

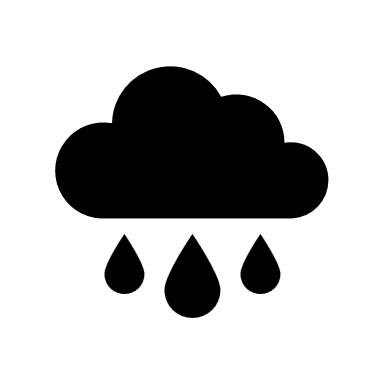
**A computer screen shot of a computer code

Description automatically generated**

1. **What is the average rainfall for each in the last 5 years:** Year 2020 has the highest average rainfall in the last five years. Interesting fact is 2023 we have only 10 months and its already second highest in the avg maximum rainfall**.**

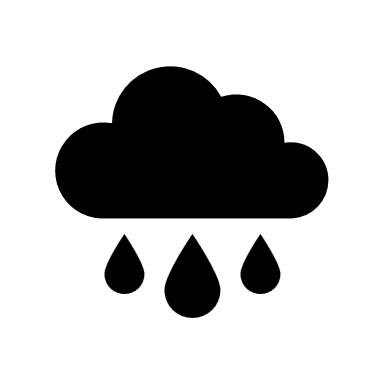
**A screenshot of a computer code

Description automatically generated**

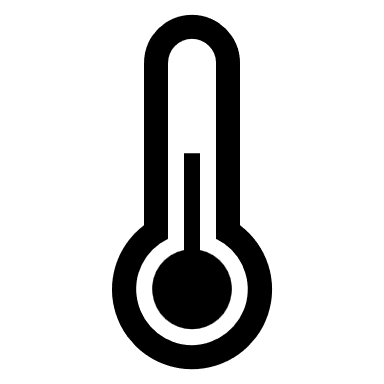
1. ** Calculate average rainfall monthly and yearly for all the years in the dataset and output the value into new collection.**

**A computer screen with text

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

****

**Let’s Analyse the mean max temperature.**

1. **What is the mean maximum temperature in the year 2022 and 2023**

In year 2022 , month of August was recorded with highest temperature of 19.7C

**A computer screen shot of a computer code

Description automatically generated**

In year 2023 month 6 was recorded with highest temperature of 19.8C

**A computer screen shot of a computer code

Description automatically generated**

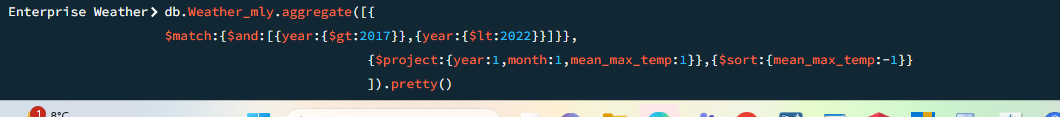
1. **Find the highest maximum temperatures in the last five years excluding year 2023 and 2022**

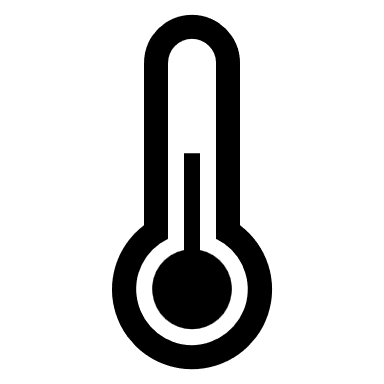
Between the years 2018 and 2021 : July 2018 has the highest mean temp

Of 19.5c

Hence highest mean temperatures has been consistent in the range of 19.7 to 19.5c since 2018

Its mostly summer months ex : July till September shows maximum mean temperatures, except in year 2020 month of May has hit the highest mean temperature of 15.9

****

****

1. **Output the results of the previous query to a new collection**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screen shot of a computer

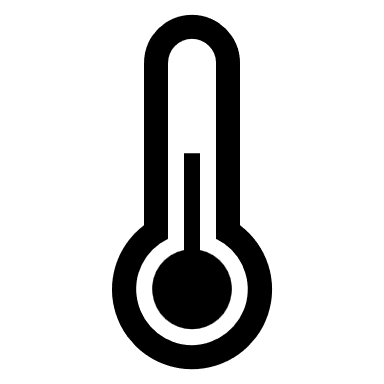
Description automatically generated**

1. **Calculate the average mean temp for all the years in the dataset and output the results into new collection.**

Year 2023 has highest average mean temperatures of 15.01 which is not a full year.

**A computer screen shot of code

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

1. **Calculate the monthly average temperatures and output the results into new collection.**

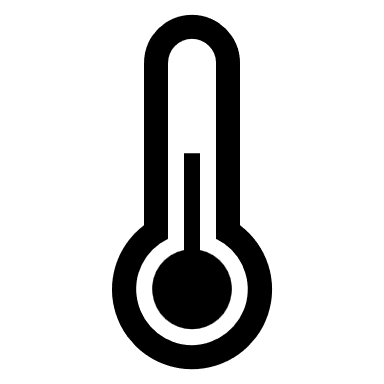
Month of August has seen the highest average temperature of 18.08c, second highest is July with 17.8c .

January and February are colder than December .

There is three degree drop in average temperatures between October and November

**A computer screen shot of a code

Description automatically generated**

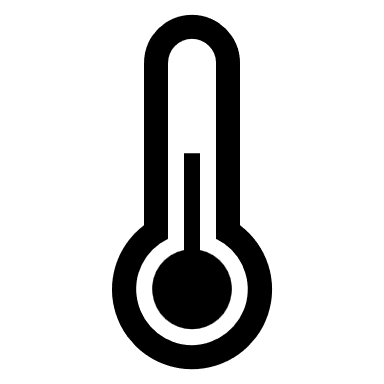
****Average monthly temperatures round to two decimal places.

**A computer screen with text

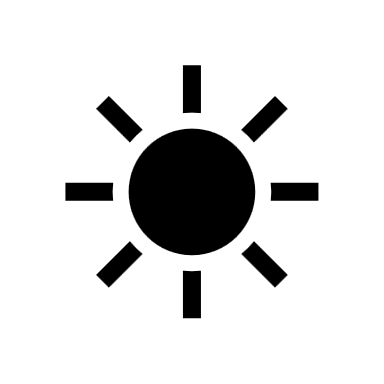
Description automatically generated**

**A screen shot of a computer code

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

1. **Insert missing records for sunshine\_duration with the median value of the field. Only for the year 2023 and 2022 .**

Sunshine\_duration field is updated for year 2023 and 2022 with median value of the field which is 121.21

**A screen shot of a computer

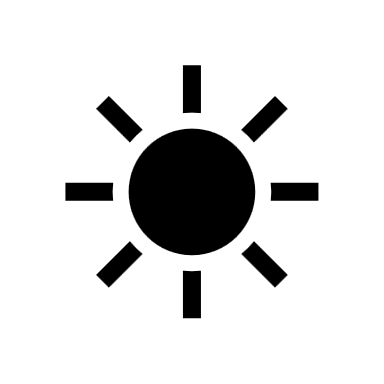
Description automatically generated**

**A screen shot of a computer

Description automatically generated**

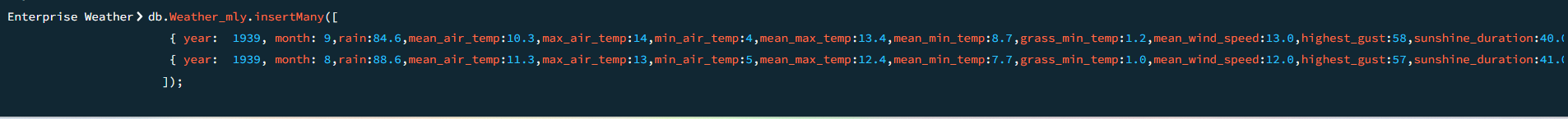
**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

1. **Insert two months of data for year 1939 : Month 8 and 9 which was previously missing.**

****

**A screenshot of a computer program

Description automatically generated**

1. **Insert Document manually using GUI:**

**A screenshot of a computer program

Description automatically generated**

**Find the inserted data**

**A screenshot of a computer

Description automatically generated**

**Mongo Compass only :**

1. **Find year, Month , rain, mean\_max\_temperature, highest\_gust for the year 2021 and 2022**

**Order by year desc and month asc**

A screenshot of a computer

Description automatically generated

1. **Find average rainfall for year the current year.**

A screenshot of a computer

Description automatically generated

1. **Find the maximum average rainfall : Year 2009 received maximum rainfall of 181.24mm**

A screenshot of a computer

Description automatically generated

1. **What is the yearly min average of rain ?**

Year 1971 received min average rainfall of 82.616 mm. Year 2009 received double the min rainfall.

A screenshot of a computer

Description automatically generated

1. **Which month receives min average rainfall?**

Month of June receives min avg rainfall.

A screenshot of a computer

Description automatically generated

1. Update records using Composs GUI

Update windspeed and sunshine duration for February 2018 and January 2012. Change datatype from string to Double.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated