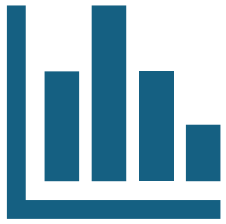
Case Study – Chickpea Harvesting

By – Md Javed Momin

Version Used



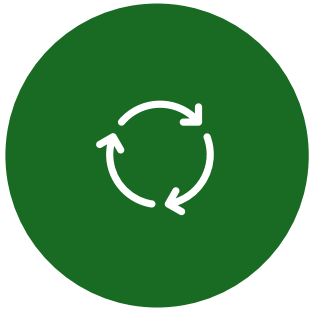
Data Cleaning : MS Excel 365

Database : MYSQL Server 8.0

Visualization : Power BI

Data Cleaning Steps – Columm Renaming

I USED CONCATENATE FUNCTION TO COMBINE TWO CELLS WHERE I COMBINED DISTRICT AND PARAMETERS IN ONE CELL WITHOUT WHICH IT WOULD BE DIFFICULT TO ANALYSE THE DATA.



= CONCATENATE(CELL NUMBER, CELL NUMBER, CELL NUMBER ...)

= CONCATENATE(F2, “ ”, “-”, “ ”,F3)

OUTPUT : DISTRICT 1 - T2M_MIN

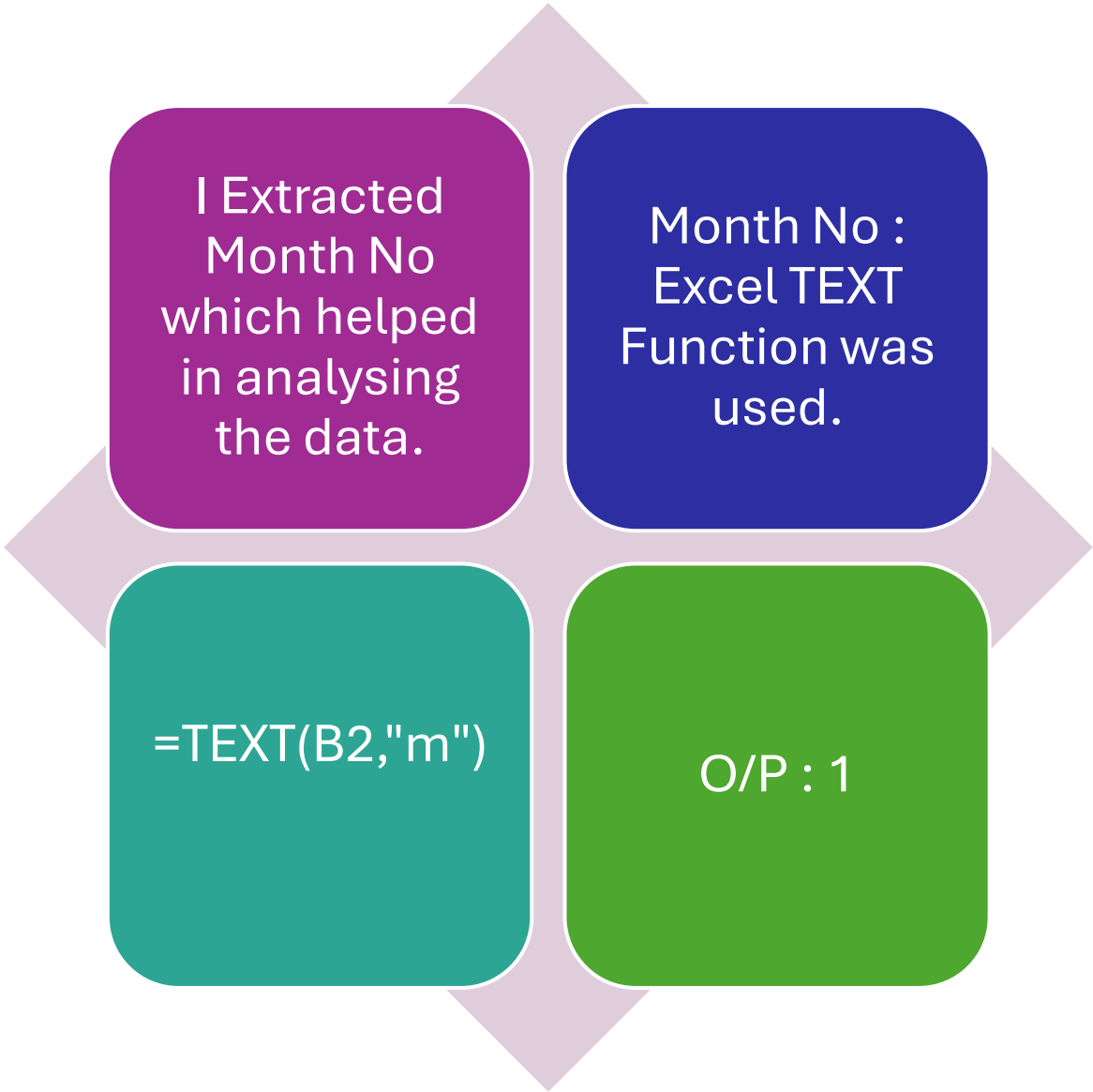
Data Cleaning Steps – Inconsistent Date & Time Values

In Date & Time Column, Inconsistent Data was noticed.

Ex : 01-01-2013 12:00:00 AM and 14-01-2013 00:00

So, I used Flash Fill (Ctrl + E) where I manually entered First Two Rows so that Excel can understand the Pattern and then Used Flash Fill to get Consistent Data Formatting (Only Date Ignoring Time)

Data Cleaning Steps : Extracting Month No



Data Cleaning Steps : Extracting Month Name

I Extracted Month Name which helped in analysing the data.

Month Name : Excel TEXT Function was used.

=TEXT(B2,“MMM")

O/P : Jan

Analysing MAX Temperature

I used Excel's Pivot Functions for Analysing the Data.

I calculated Average MAX Temperature of each district and grouped it by months.

Later, I took the average of those 32 districts and grouped it by months.

Months	Max Temperature (Average)
Jan	24.07
Feb	28.64
Mar	33.64
Apr	39.1
May	41.65
Jun	39.83
Jul	35.32
Aug	32.89
Sep	33.2
Oct	33.19
Nov	29.46
Dec	25.36

MAX Temperature

- The MAX Temperature was recorded in May having Temperature of 41.65

Analysing MIN Temperature

I used Excel's Pivot Functions for Analysing the Data.

I calculated Average MIN Temperature of each district and grouped it by months.

Later, I took the average of those 32 districts and grouped it by months.

Months	Min Temperature (Average)
Jan	8.38
Feb	11.78
Mar	17.08
Apr	22.60
May	26.82
Jun	28.07
Jul	26.56
Aug	24.78
Sep	23.35
Oct	19.90
Nov	15.08
Dec	10.43

MIN Temperature

- The MAX Temperature was recorded in Jan having Temperature of 8.38

Analysing Rainfall (mm)

I used Excel's Pivot Functions for Analysing the Data.

I calculated Average Rainfall (mm) of each district and grouped it by months.

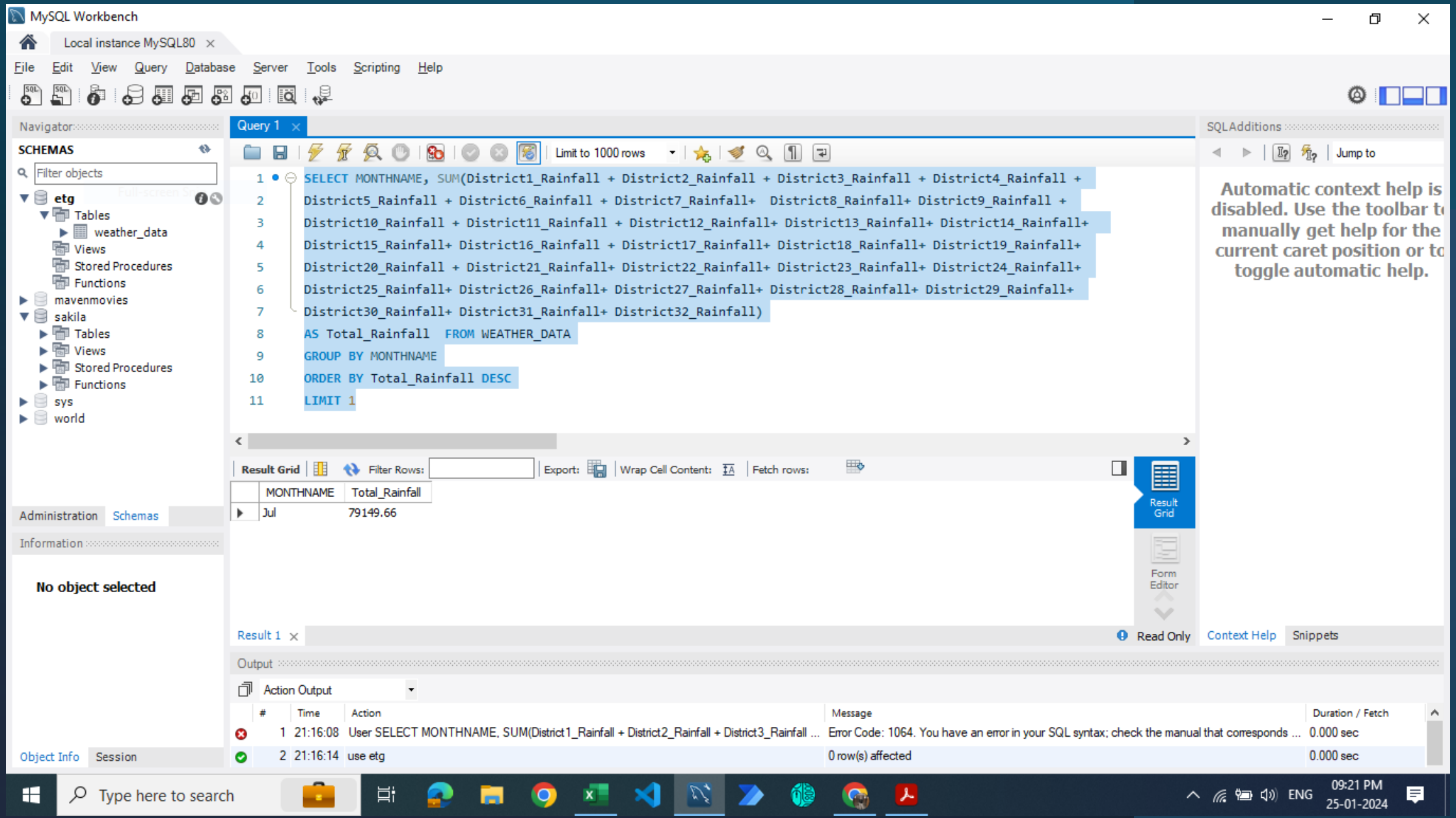
Later, I took the average of those 32 districts and grouped it by months.

Months	Rainfall (mm) (Average)
Jan	0.33
Feb	0.17
Mar	0.38
Apr	0.25
May	0.63
Jun	2.56
Jul	7.25
Aug	6.21
Sep	2.98
Oct	0.51
Nov	0.25
Dec	0.12

SQL Query : Write a query to find which month got the highest rainfall in the State.

```
SELECT MONTHNAME, SUM(District1_Rainfall + District2_Rainfall + District3_Rainfall +
District4_Rainfall +
District5_Rainfall + District6_Rainfall + District7_Rainfall+  District8_Rainfall+ District9_Rainfall +
District10_Rainfall + District11_Rainfall + District12_Rainfall+ District13_Rainfall+ District14_Rainfall+
District15_Rainfall+ District16_Rainfall + District17_Rainfall+ District18_Rainfall+ District19_Rainfall+
District20_Rainfall + District21_Rainfall+ District22_Rainfall+ District23_Rainfall+ District24_Rainfall+
District25_Rainfall+ District26_Rainfall+ District27_Rainfall+ District28_Rainfall+ District29_Rainfall+
District30_Rainfall+ District31_Rainfall+ District32_Rainfall)
AS Total_Rainfall  FROM WEATHER_DATA
GROUP BY MONTHNAME
ORDER BY Total_Rainfall DESC
LIMIT 1
```

Max_Rainfall_Month	Max_Rainfall (mm)
Jul	79149.66


MySQL Workbench
Local instance MySQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
etg
Tables
weather_data
Views
Stored Procedures
Functions
mavenmovies
sakila
Tables
Views
Stored Procedures
Functions
sys
world
Administration Schemas
Information
No object selected
Object Info Session
Query 1
Limit to 1000 rows
1 SELECT MONTHNAME, SUM(District1_Rainfall + District2_Rainfall + District3_Rainfall + District4_Rainfall +
2 District5_Rainfall + District6_Rainfall + District7_Rainfall+ District8_Rainfall+ District9_Rainfall +
3 District10_Rainfall + District11_Rainfall + District12_Rainfall+ District13_Rainfall+ District14_Rainfall+
4 District15_Rainfall+ District16_Rainfall + District17_Rainfall+ District18_Rainfall+ District19_Rainfall+
5 District20_Rainfall + District21_Rainfall+ District22_Rainfall+ District23_Rainfall+ District24_Rainfall+
6 District25_Rainfall+ District26_Rainfall+ District27_Rainfall+ District28_Rainfall+ District29_Rainfall+
7 District30_Rainfall+ District31_Rainfall+ District32_Rainfall)
8 AS Total_Rainfall FROM WEATHER_DATA
9 GROUP BY MONTHNAME
10 ORDER BY Total_Rainfall DESC
11 LIMIT 1
Result Grid Filter Rows: Export: Wrap Cell Content: Fetch rows:
MONTHNAME Total_Rainfall
Jul 79149.66
Result Grid
Form Editor
Result 1
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
1 21:16:08 User SELECT MONTHNAME, SUM(District1_Rainfall + District2_Rainfall + District3_Rainfall ... Error Code: 1064. You have an error in your SQL syntax; check the manual that corresponds ... 0.000 sec
2 21:16:14 use etg 0 row(s) affected 0.000 sec
Type here to search
ENG
09:21 PM
25-01-2024

Yield Prediction

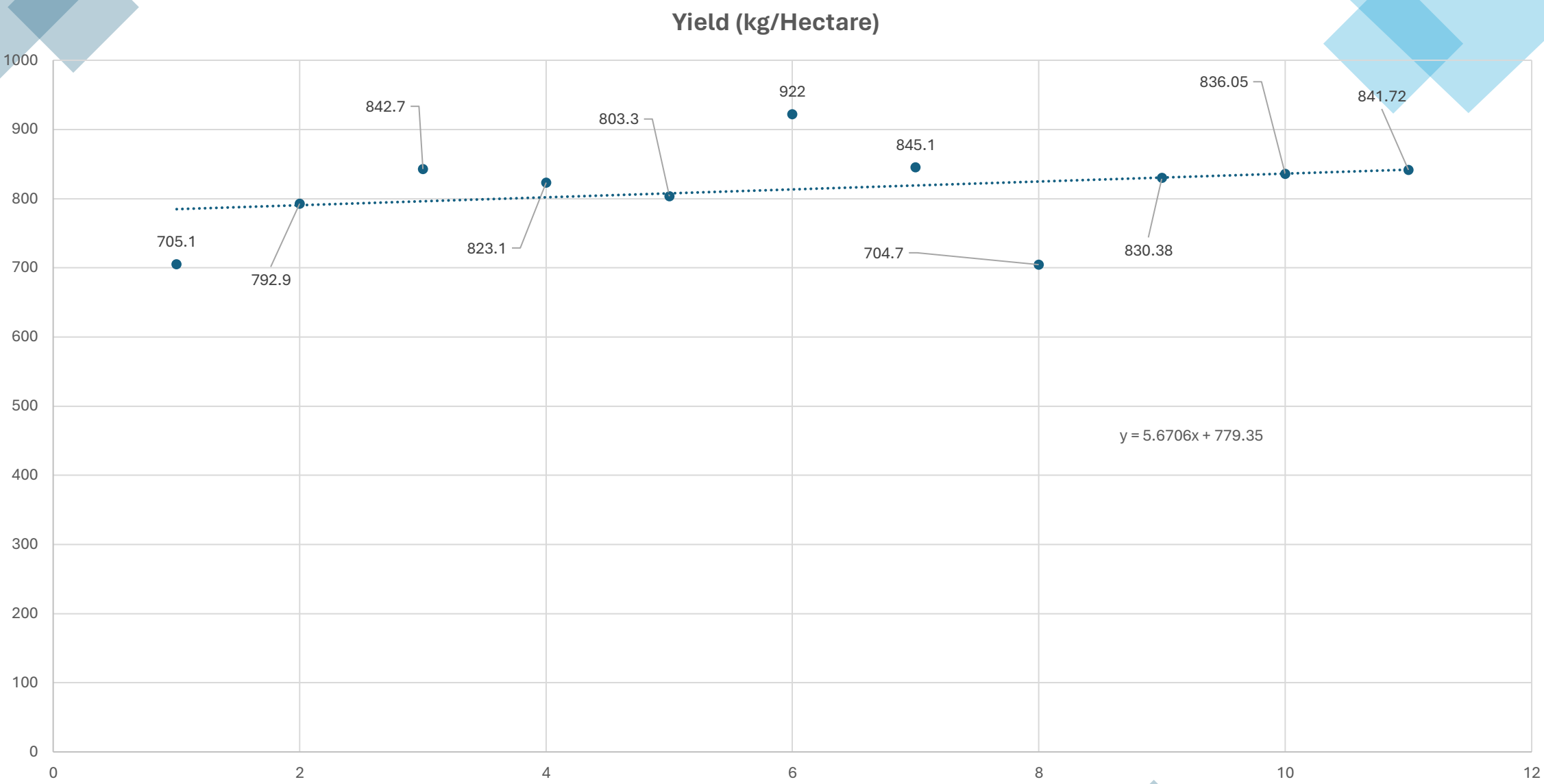
I analysed in Excel as I was unable to perform in Python/R.

I used Linear Regression Method for the same.

The Equation of Line was : $y = 5.6702x + 779.35$

Here, 'x' represents the numerical years, and 'y' represents the predicted yield.

Years	Yield (kg/Hectare)	X
2013-14	705.1	
2014-15	792.9	
2015-16	842.7	
2016-17	823.1	
2017-18	803.3	
2018-19	922	
2019-20	845.1	
2020-21	704.7	
2021-22	830.38	x=9
2022-23	836.05	x=10
2023-24	841.72	x=11


Yield (kg/Hectare)
1000
900
800
700
600
500
400
300
200
100
0
0
2
4
6
8
10
12
705.1
792.9
842.7
823.1
803.3
922
845.1
704.7
830.38
836.05
841.72
y = 5.6706x + 779.35

Power BI Analysis

I analysed the Data in Power BI as well & created meaningful visualizations.

Parameters Analysed :

A) MIN Rainfall (mm)

B) MAX Rainfall (mm)

C) MIN Temperature (Average)

D) MAX Temperature (Average)

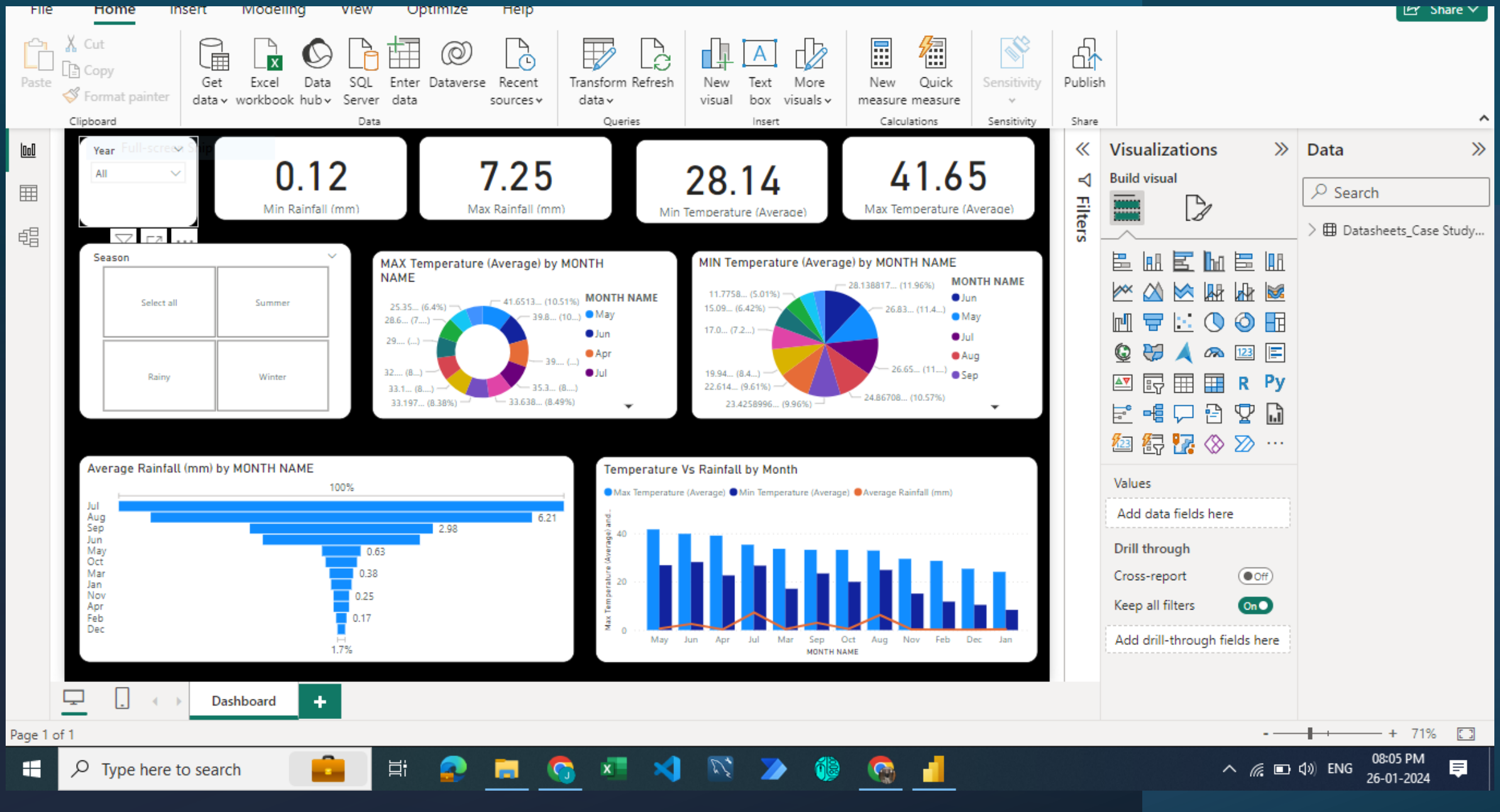
E) MIN Temperature (Average) by month

F) MAX Temperature (Average) by month

G) Average Rainfall (mm) by month

H) Temperature Vs Rainfall by month

Filters Used : Seasons & Years


File
Home
Insert
Modeling
View
Optimize
Help
Share
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
Data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New measure
Quick measure
Calculations
Sensitivity
Publish
Share
Year
All
0.12
Min Rainfall (mm)
7.25
Max Rainfall (mm)
28.14
Min Temperature (Average)
41.65
Max Temperature (Average)
Season
Select all
Summer
Rainy
Winter
MAX Temperature (Average) by MONTH NAME
MONTH NAME
May
Jun
Apr
Jul
MIN Temperature (Average) by MONTH NAME
MONTH NAME
Jun
May
Jul
Aug
Sep
Average Rainfall (mm) by MONTH NAME
100%
Jul
Aug
Sep
Jun
May
Oct
Mar
Jan
Nov
Apr
Feb
Dec
6.21
2.98
0.63
0.38
0.25
0.17
1.7%
Temperature Vs Rainfall by Month
Max Temperature (Average)
Min Temperature (Average)
Average Rainfall (mm)
MONTH NAME
May
Jun
Apr
Jul
Mar
Sep
Oct
Aug
Nov
Feb
Dec
Jan
Filters
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Off
Keep all filters
On
Add drill-through fields here
Data
Search
Datasheets_Case Study...
Dashboard
Page 1 of 1
71%
Type here to search
ENG
08:05 PM
26-01-2024

Thank You !!!