ARPAN CHATTERJEE
2312RES162
IITP000646
DATA SCIENCE PROJECT 2 (GUVI HCL)
arpan_2312res162@iitp.ac.in
4th Semester BSc CSDA
Github: https://github.com/ArpanC03/GuviProject2DataScience
Colab: https://colab.research.google.com/drive/1xApbaj9PFJoNHeqZisF6BLhwWF4SXGV0#scrollTo=zvJ9zB6yMtI0

Weather Trends Visualizer

View in Power BI ↗

Overview

This project focuses on analyzing and visualizing monthly weather patterns using historical data. By leveraging popular Python libraries such as Pandas for data manipulation and Plotly for interactive plotting, the project demonstrates how to extract meaningful insights from daily records of temperature and rainfall. The key objective is to provide an easy-to-understand, graphical representation of weather trends for each month of the year.

Data Preparation

The dataset used in this analysis consists of daily entries capturing the date, recorded temperature, and precipitation values. To ensure robust insights, preprocessing steps included converting date columns to datetime format, managing missing or inconsistent data, and identifying the specific month for each entry. This structured approach allowed for effective grouping and filtering of data, setting the stage for accurate monthly trend analysis.

Visualization Methodology

For visualization, the project employs Plotly’s powerful charting capabilities in combination with ipywidgets for interactivity. Users can interactively select a month from a dropdown menu, prompting the system to display that month’s temperature and rainfall trends. The temperature values are depicted using a line chart, while rainfall data is presented as overlayed bars. This dual representation helps reveal correlations and anomalies within each month’s weather patterns.

AutoSave Off weather_2013_to_2024.csv • Saved to this PC

Search

File Home Insert Draw Page Layout Formulas Data Review View Help

Comments Share

Paste Cut Copy Format Painter Clipboard

Calibri 11 Font

Wrap Text Merge & Center Alignment

General Number

Conditional Formatting Format as Table Cell Styles Styles

Insert Delete Format Cells

AutoSum Fill Clear Sort & Filter Find & Select Editing

Add-ins Add-ins

BACK UP THIS DOCUMENT Allow Copilot on this and other files backing using OneDrive (License required). Open OneDrive

A1 fx

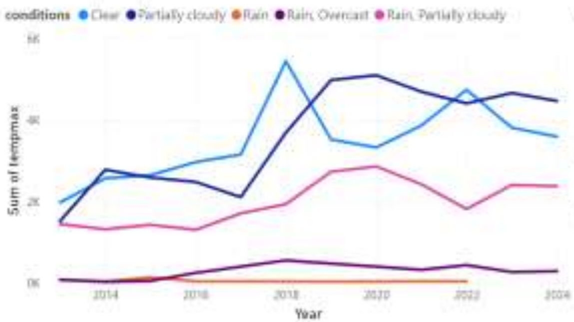
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1		DATE	tempmax	tempmin	temp	feelslikem	feelslikemi	feelslike	humidity	precip	precipprob	precipcove	windspeed	sealevelpr	conditions	Year	month	dayofweel	dayofyear	year-2000	weekofyea	tempmax_	tempmin_
2	0	4/14/2013	37.7	23.1	28.7	35.4	23.1	28.1	39.7	2	100	4.17	22.3	1008.3	Rain, Parti	2013	4	6	104	13	15	1496.69	917.07
3	1	4/15/2013	37.5	21.1	28.6	35.3	21.1	28	41.7	0	0	0	17.6	1005.1	Clear	2013	4	0	105	13	16	1563.75	879.87
4	2	4/16/2013	40.1	21.9	31.7	37.5	21.9	30.4	30.7	0	0	0	27.7	1001.7	Partially cl	2013	4	1	106	13	16	1231.07	672.33
5	3	4/17/2013	36.4	21	29.9	34	21	28.5	27.4	0	0	0	22.3	1002.5	Clear	2013	4	2	107	13	16	997.36	575.4
6	4	4/18/2013	37.5	21.7	30.6	35.2	21.7	29.2	23.7	0	0	0	12.4	1003.4	Clear	2013	4	3	108	13	16	888.75	514.29
7	5	4/19/2013	37.3	21.4	30.3	35	21.4	28.9	27.2	0	0	0	14.8	1004	Clear	2013	4	4	109	13	16	1014.56	582.08
8	6	4/20/2013	37.7	22.2	30.6	36	22.2	29.4	27.8	0	0	0	16	1003.9	Partially cl	2013	4	5	110	13	16	1048.06	617.16
9	7	4/21/2013	37	23.9	28.4	35.4	23.9	28.1	39.5	0	0	0	27.7	1007.2	Partially cl	2013	4	6	111	13	16	1461.5	944.05
10	8	4/22/2013	31.7	21.1	26.9	31.3	21.1	26.7	49.5	0	0	0	22.3	1010.4	Partially cl	2013	4	0	112	13	17	1569.15	1044.45
11	9	4/23/2013	34	23.3	29.1	33.2	23.3	28.6	42.8	0	0	0	14	1010.4	Partially cl	2013	4	1	113	13	17	1455.2	997.24
12	10	4/24/2013	36	23.8	30.8	35.7	23.8	30.3	38.8	0	0	0	11.2	1010.2	Clear	2013	4	2	114	13	17	1396.8	923.44
13	11	4/25/2013	37.7	24.8	31.5	37	24.8	31.2	39.1	0	0	0	20.5	1009	Partially cl	2013	4	3	115	13	17	1474.07	969.68
14	12	4/26/2013	34.3	25.4	30.9	36	25.4	30.9	41.5	0	0	0	27.5	1009	Partially cl	2013	4	4	116	13	17	1423.45	1054.1
15	13	4/27/2013	37.7	24.2	31.8	37.3	24.2	30.9	32.7	0	0	0	24.1	1008.4	Partially cl	2013	4	5	117	13	17	1232.79	791.34
16	14	4/28/2013	37	23.4	31.2	34.6	23.4	29.9	29.6	0	0	0	25.9	1006.5	Partially cl	2013	4	6	118	13	17	1095.2	692.64
17	15	4/29/2013	39.1	26.5	32.8	36.8	26.5	31.3	24.3	0	0	0	17.3	1005.8	Partially cl	2013	4	0	119	13	18	950.13	643.95
18	16	4/30/2013	40.1	24.9	32.9	38.9	24.9	32.1	31.3	0	0	0	18.4	1005.9	Partially cl	2013	4	1	120	13	18	1255.13	779.37
19	17	5/13/2013	37.1	24.1	30.8	36.5	24.1	30.4	45.3	0	0	0	14.5	1004.8	Partially cl	2013	5	0	133	13	20	1680.63	1091.73
20	18	5/14/2013	42.1	25.6	33.7	40.2	25.6	33.1	31.8	0	0	0	46.4	1002.7	Partially cl	2013	5	1	134	13	20	1338.78	814.08
21	19	5/15/2013	41.1	25.9	33.8	39.5	25.9	32.9	29.6	0	0	0	16.3	1001.6	Partially cl	2013	5	2	135	13	20	1216.56	766.64
22	20	5/16/2013	40.6	27.2	34.4	39.6	27.8	33.2	27	0.1	100	4.17	21.2	1000	Rain	2013	5	3	136	13	20	1096.2	734.4
23	21	5/17/2013	42.1	25.4	34.4	39.3	25.4	32.8	25	0	0	0	20.9	1001.3	Clear	2013	5	4	137	13	20	1052.5	635
24	22	5/18/2013	44.1	26.2	35.9	44.1	26.2	34.4	22.2	0	0	0	38.2	1000.8	Clear	2013	5	5	138	13	20	979.02	581.64
25	23	5/19/2013	45	27.2	36.7	45	26.6	35.5	19.2	0	0	0	25.8	1000.4	Clear	2013	5	6	139	13	20	864	522.24
26	24	5/20/2013	44	30.9	37.4	44	29	35.8	17.7	0	0	0	22.7	1000.9	Clear	2013	5	0	140	13	21	778.8	546.93

weather_2013_to_2024

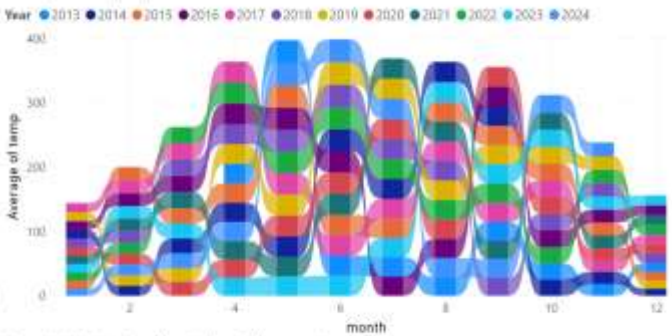
Ready Accessibility: Unavailable 100%

27°C Cloudy Quick search ENG IN 22:01 18-08-2025

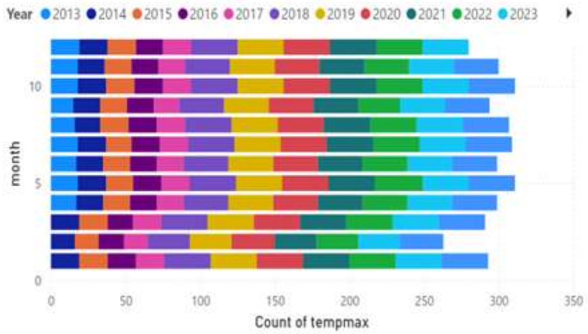
Sum of tempmax by Year and conditions



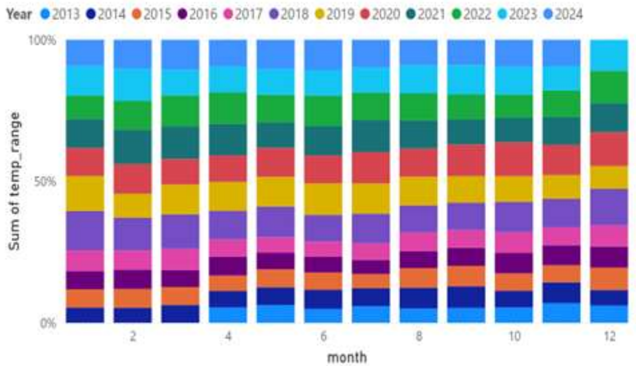
Average of temp by month and Year



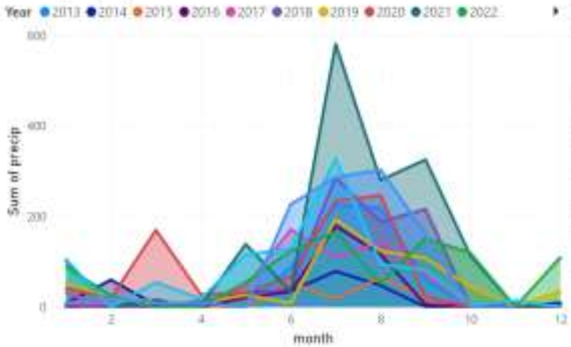
Count of tempmax by month and Year



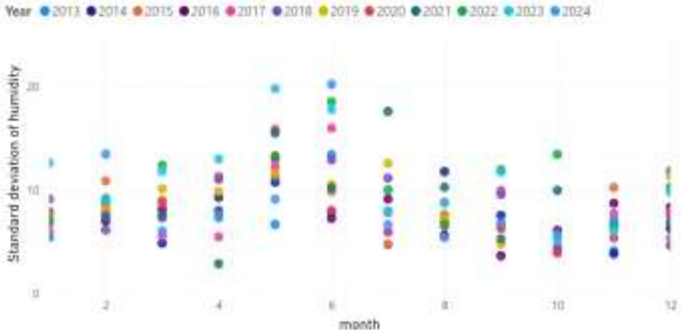
Sum of temp_range by month and Year



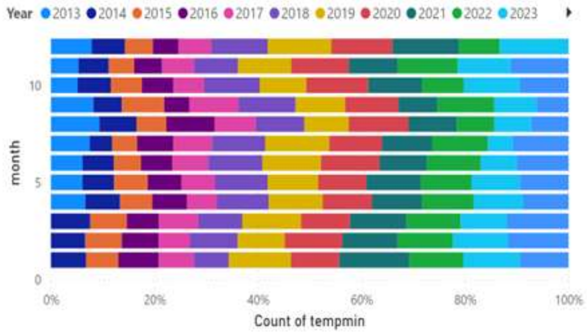
Sum of precip by month and Year



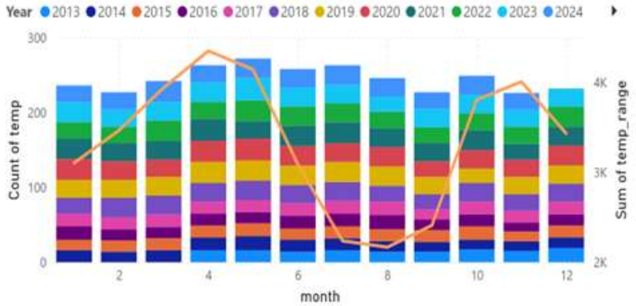
Standard deviation of humidity by Year and month



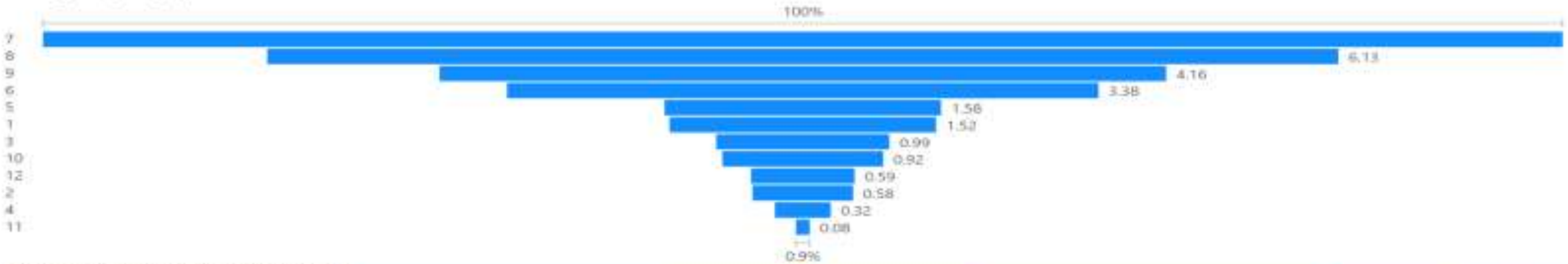
Count of tempmin by month and Year



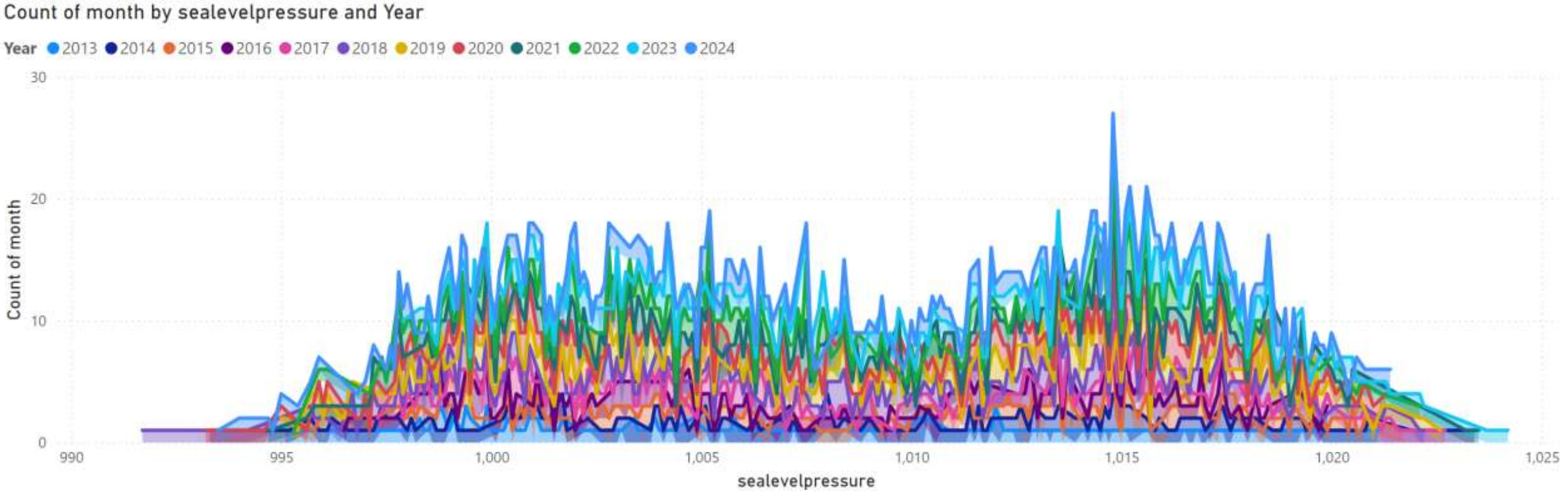
Count of temp and Sum of temp_range by month and Year



Average of precip by month

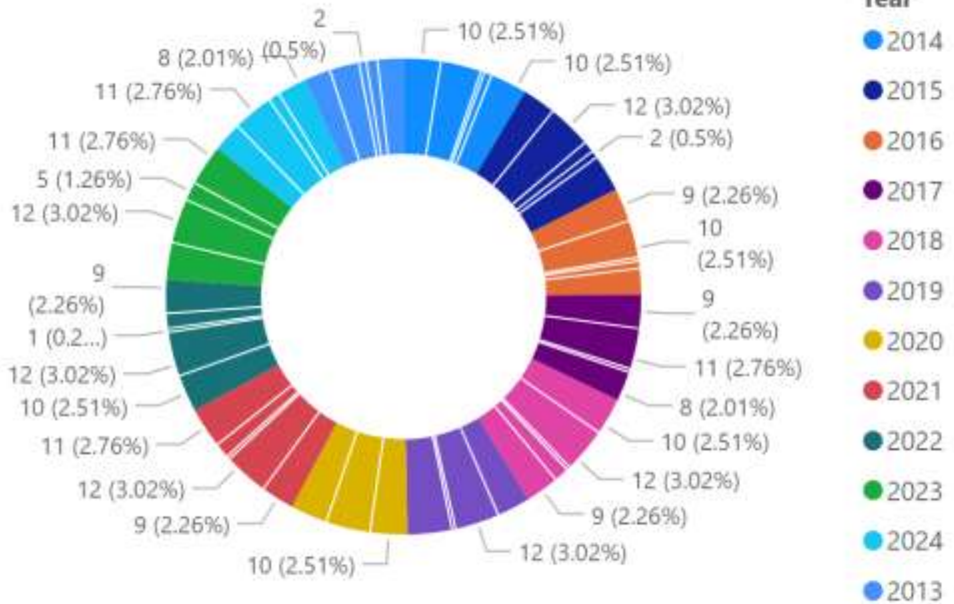


Sum of humidity by Year and month



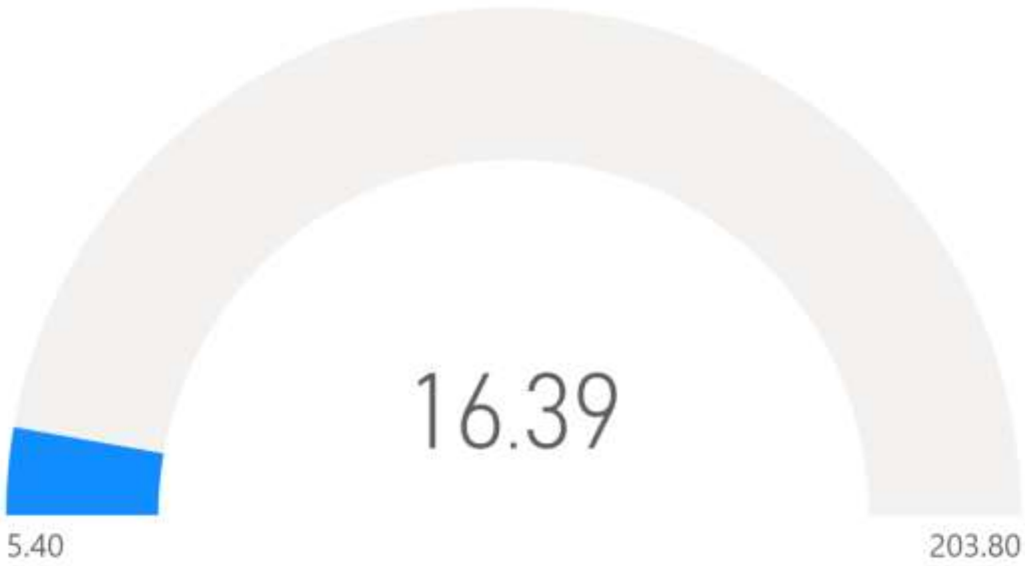
Year	2013							2014			
precip	Sum of precip	Sum of humidity	Sum of temp	Sum of tempmax	Sum of tempmin	Sum of windspeed	Sum of sealevelpressure	Sum of precip	Sum of humidity	Sum of temp	Sum of tempmax
0.00	0.00	6,056.30	2,798.70	3,495.10	2,169.80	1,850.00	107,864.40	0.00	10,613.80	4,237.20	5,346.40
0.02											
0.03											
0.05											
0.06											
0.08											
0.09											
0.09											
0.10	0.60	437.50	184.50	209.60	159.60	109.90	6,001.00	0.30	194.60	87.40	103.90
0.12											
0.13											
Total	**600.19**	**9,977.20**	**4,184.80**	**5,100.50**	**3,393.70**	**2,655.40**	**156,949.70**	**285.81**	**13,627.80**	**5,390.90**	**6,728.90**

Count of month by Year and conditions



Sum of sealevelpressure by sealevelpressure and month

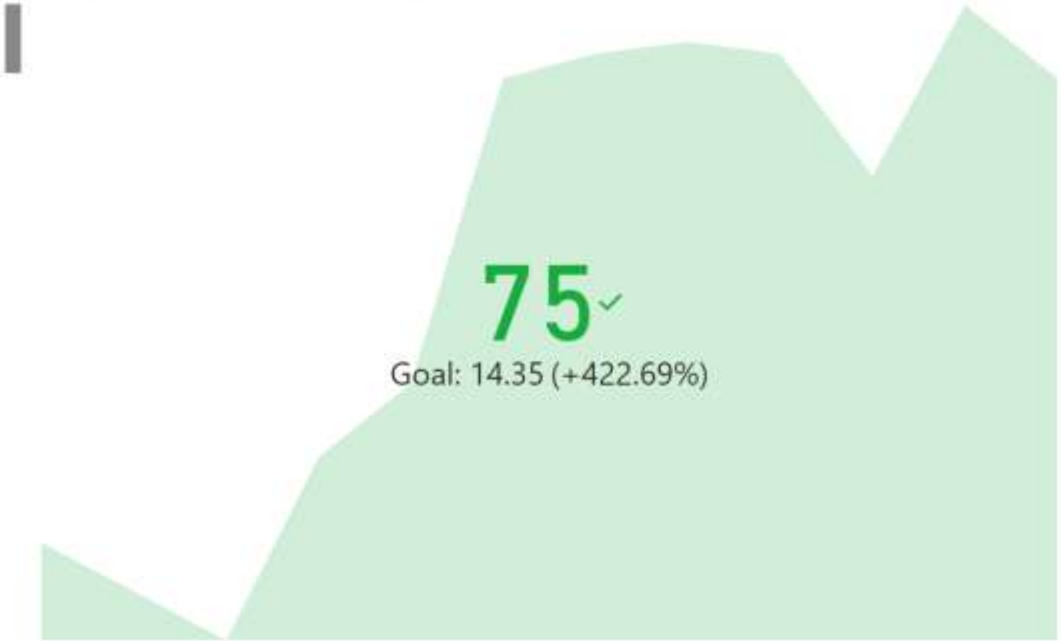
Average of windspeed, Min of windspeed and Max of windspeed

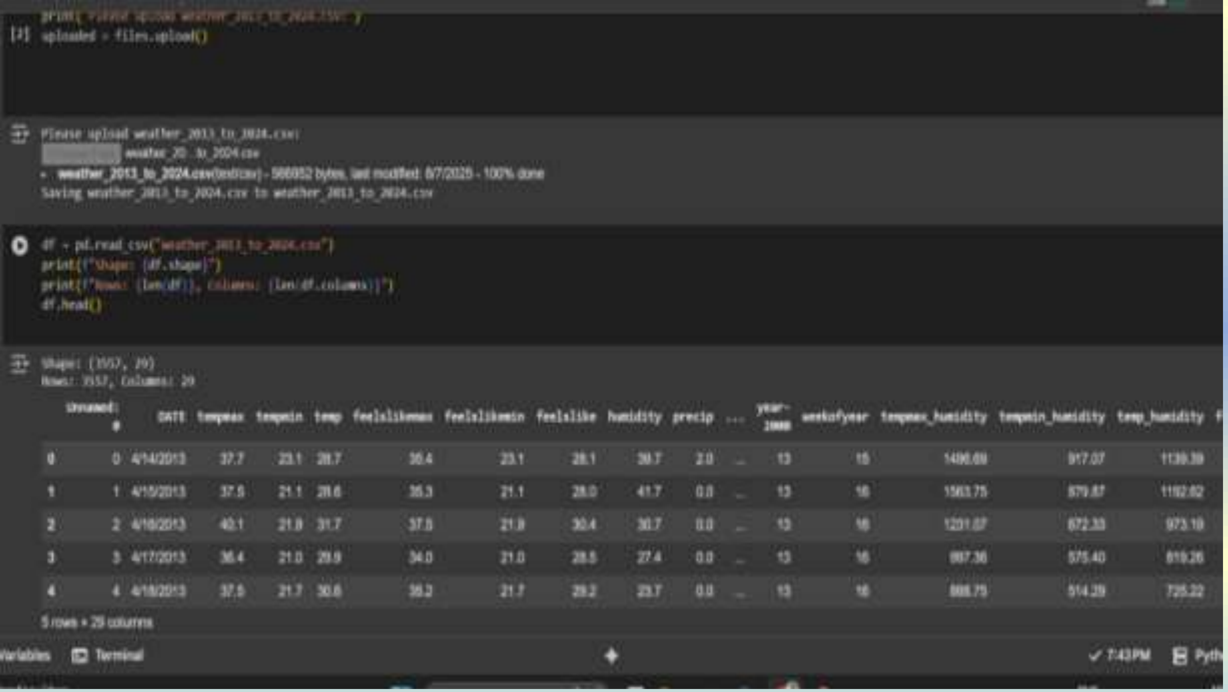


temp_range

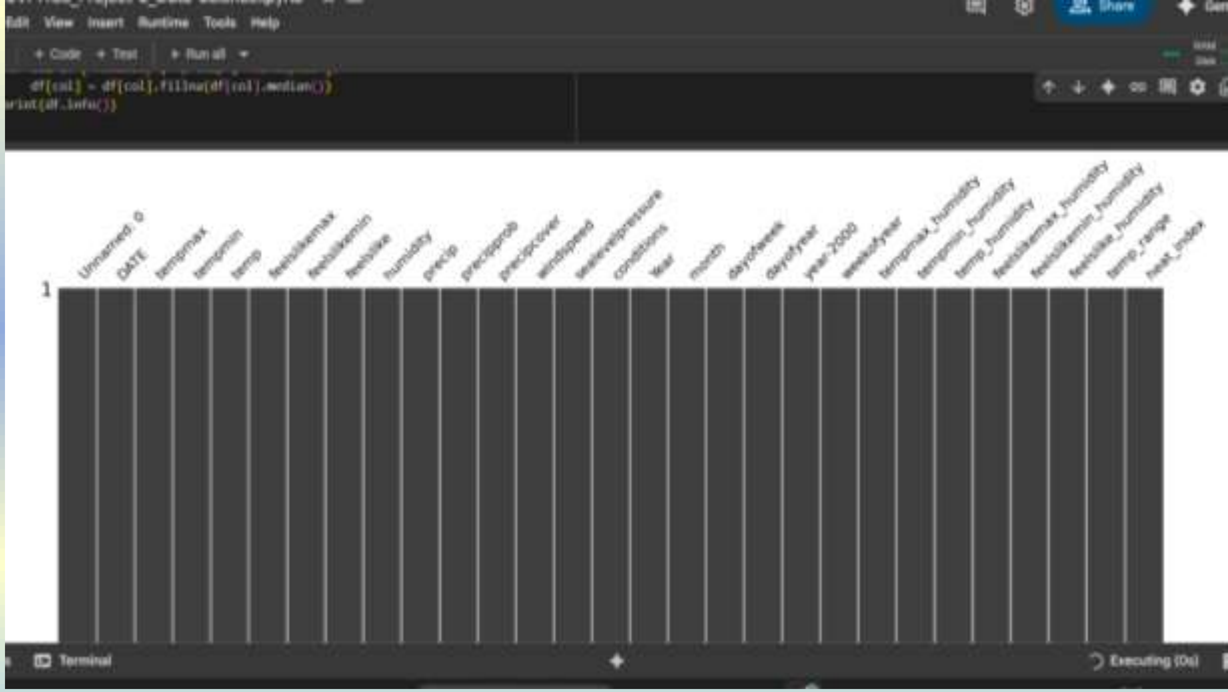
- [ ] 1.50
- [ ] 1.70
- [ ] 1.80
- [ ] 1.90
- [ ] 2.00
- [ ] 2.20
- [ ] 2.30
- [ ] 2.50
- [ ] 2.60
- [ ] 2.70
- [ ] 2.80

Count of precip and Standard deviation of precip by Year

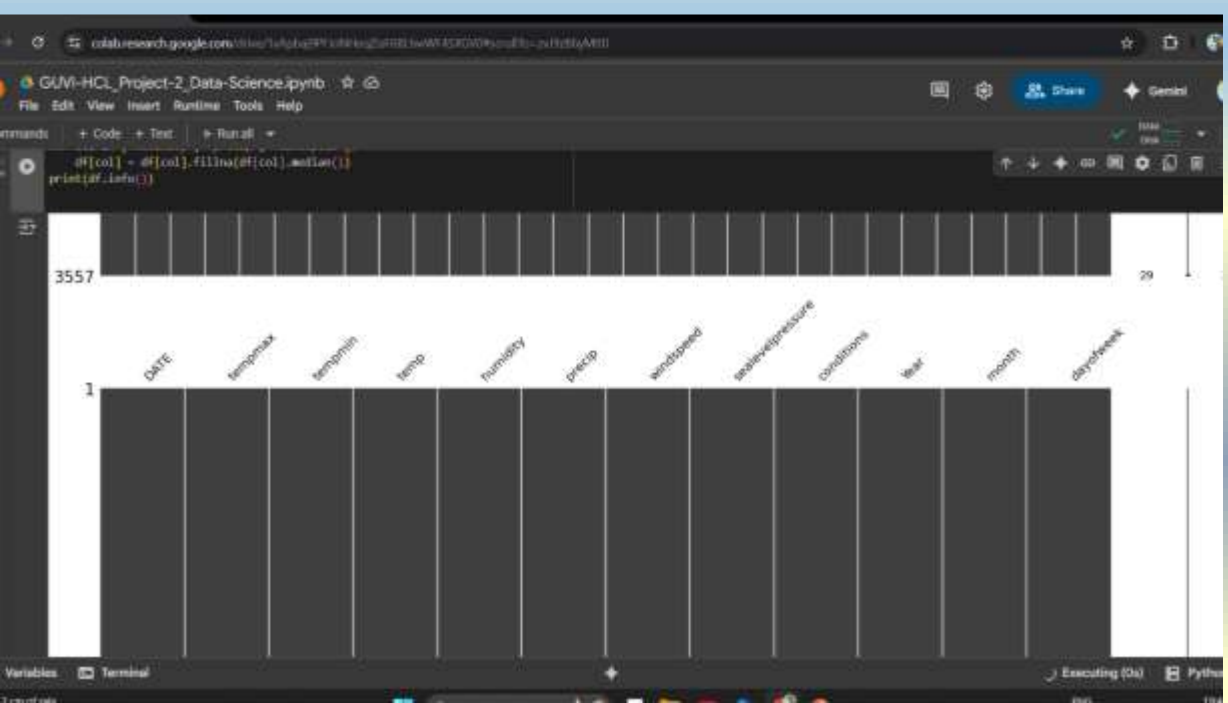




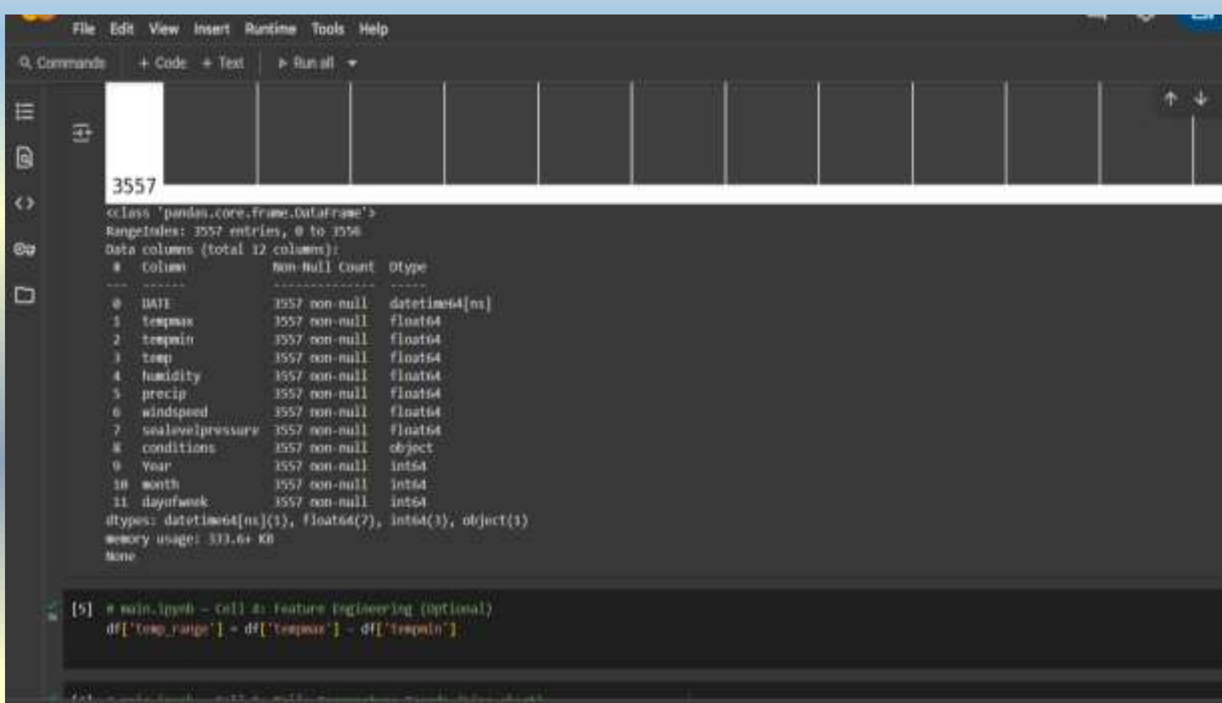
uploaded = files.upload()
Please upload weather_2013_to_2024.csv:
Saving weather_2013_to_2024.csv to weather_2013_to_2024.csv
df = pd.read_csv("weather_2013_to_2024.csv")
df.head()
Shape: (3557, 29)
Rows: 3557, Columns: 29
5 rows × 29 columns
Variables
Terminal
7:43PM



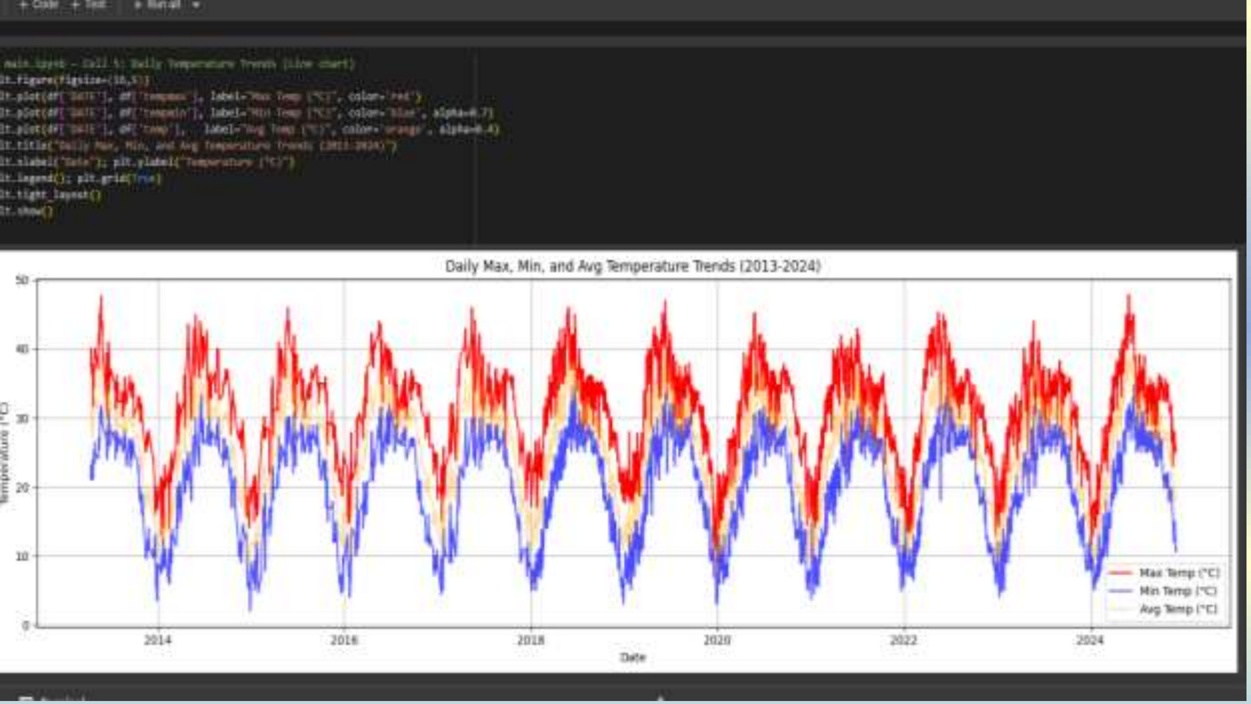
Edit View Insert Runtime Tools Help
+ Code + Text
Run all
df[col] = df[col].fillna(df[col].median())
print(df.info())
Unnamed: 0
DATE
tempmax
tempmin
temp
feelslikemax
feelslikemin
feelslike
humidity
precip
precipprob
precipcover
windspeed
sealevelpressure
conditions
Year
month
dayofweek
dayofyear
year-2000
weekofyear
tempmax_humidity
tempmin_humidity
temp_humidity
feelslikemax_humidity
feelslikemin_humidity
feelslike_humidity
temp_range
heat_index
1
Terminal
Executing (0s)



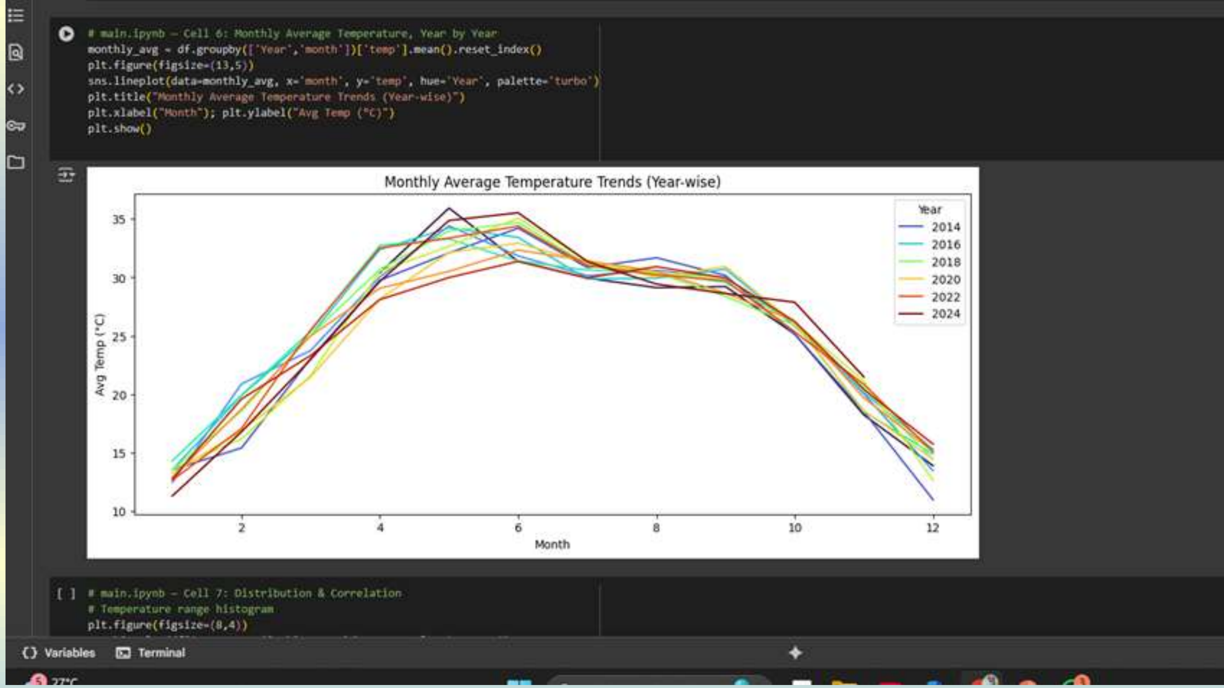
GUVI-HCL_Project-2_Data-Science.ipynb
File Edit View Insert Runtime Tools Help
+ Code + Text
Run all
Share
Gemini
df[col] = df[col].fillna(df[col].median())
print(df.info())
3557
29
DATE
tempmax
tempmin
temp
humidity
precip
windspeed
sealevelpressure
conditions
Year
month
dayofweek
1
Variables
Terminal
Executing (0s)



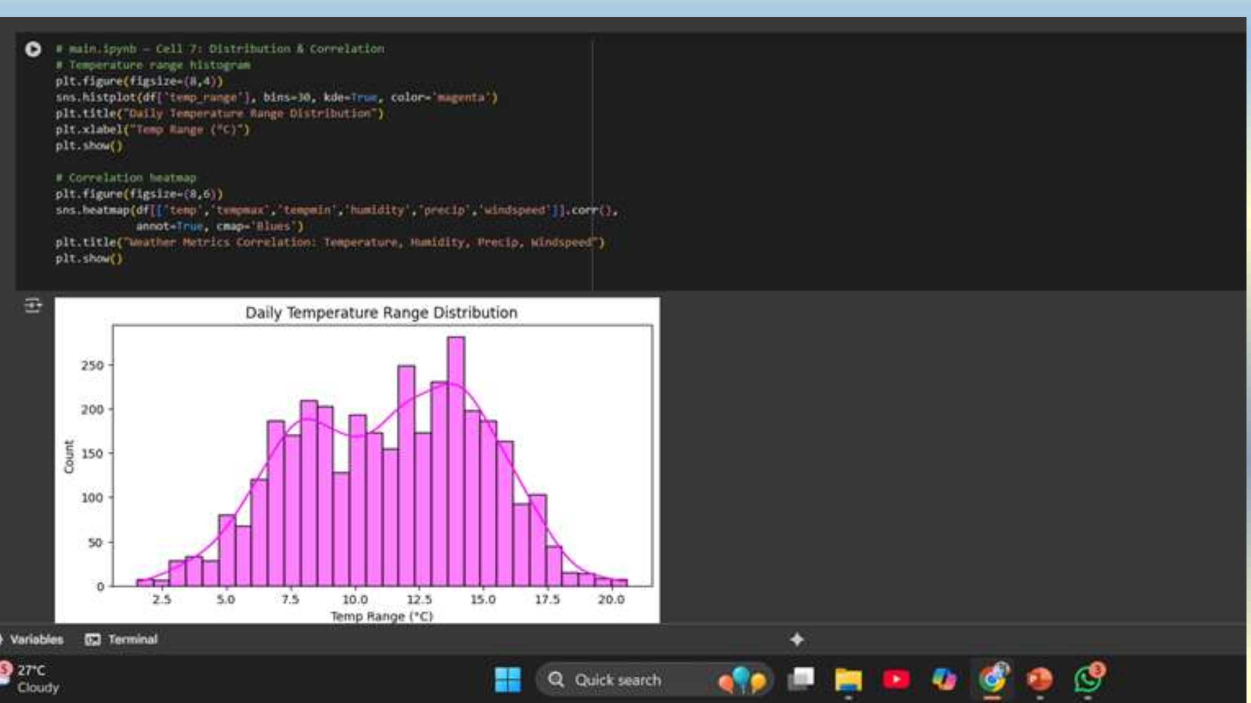
File Edit View Insert Runtime Tools Help
Commands
+ Code + Text
Run all
3557
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3557 entries, 0 to 3556
Data columns (total 12 columns):
Column Non-Null Count Dtype
0 DATE 3557 non-null datetime64[ns]
1 tempmax 3557 non-null float64
2 tempmin 3557 non-null float64
3 temp 3557 non-null float64
4 humidity 3557 non-null float64
5 precip 3557 non-null float64
6 windspeed 3557 non-null float64
7 sealevelpressure 3557 non-null float64
8 conditions 3557 non-null object
9 Year 3557 non-null int64
10 month 3557 non-null int64
11 dayofweek 3557 non-null int64
dtypes: datetime64[ns](1), float64(7), int64(3), object(1)
memory usage: 333.6+ KB
None
[5] # main.ipynb - Cell 4: Feature Engineering (Optional)
df['temp_range'] = df['tempmax'] - df['tempmin']



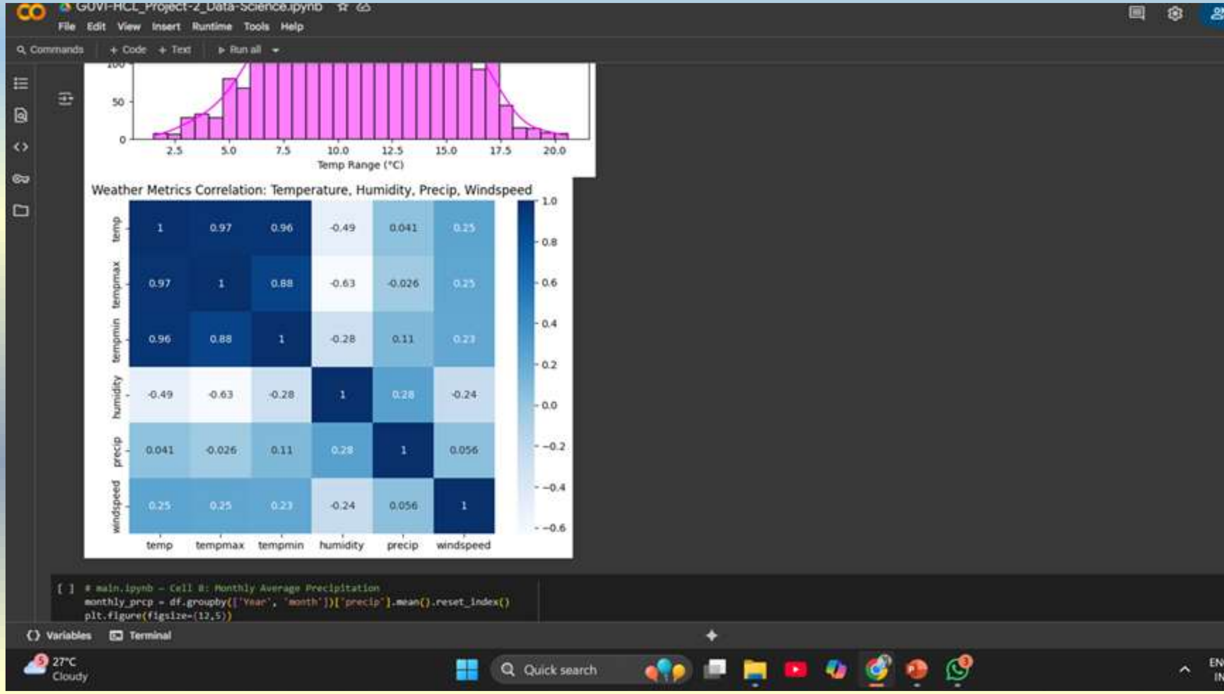
Daily Max, Min, and Avg Temperature Trends (2013-2024)
Max Temp (°C)
Min Temp (°C)
Avg Temp (°C)
Date



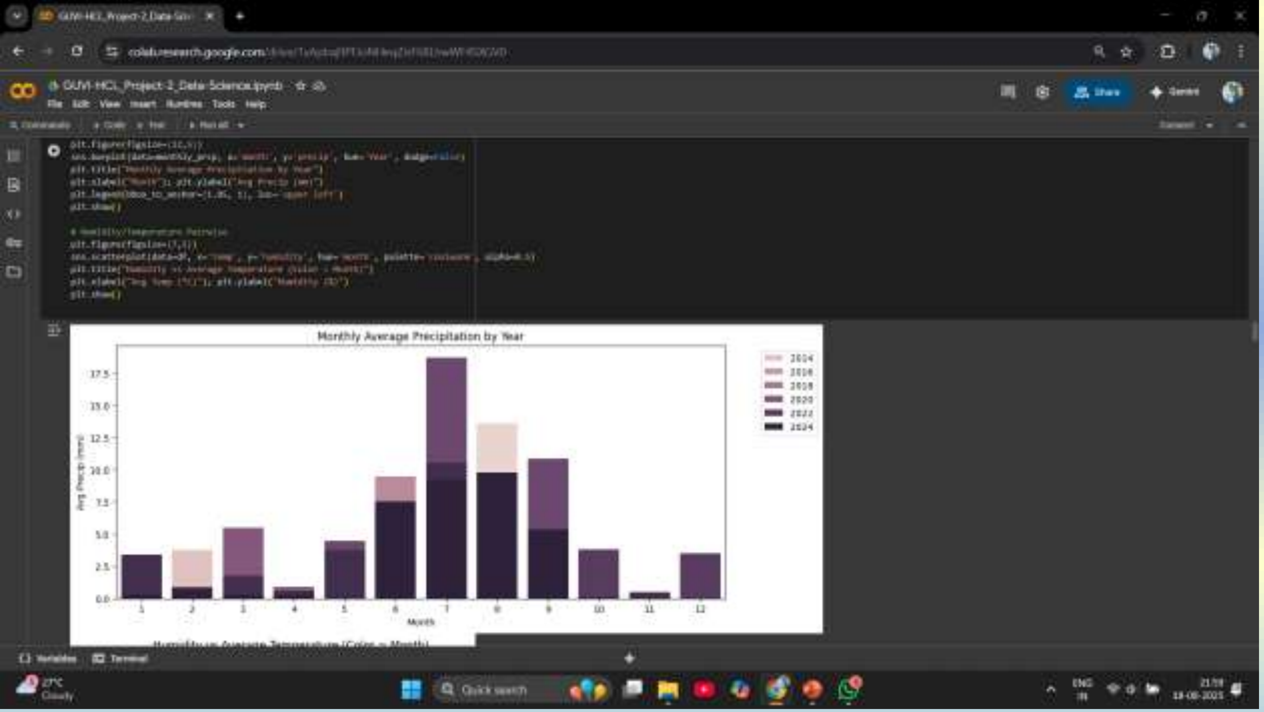
# main.ipynb – Cell 6: Monthly Average Temperature, Year by Year
monthly_avg = df.groupby(['Year','month'])['temp'].mean().reset_index()
plt.figure(figsize=(13,5))
sns.lineplot(data=monthly_avg, x='month', y='temp', hue='Year', palette='turbo')
plt.title("Monthly Average Temperature Trends (Year-wise)")
plt.xlabel("Month"); plt.ylabel("Avg Temp (°C)")
plt.show()
Monthly Average Temperature Trends (Year-wise)
Year
2014
2016
2018
2020
2022
2024
Avg Temp (°C)
Month
# main.ipynb – Cell 7: Distribution & Correlation
# Temperature range histogram
plt.figure(figsize=(8,4))
Variables
Terminal

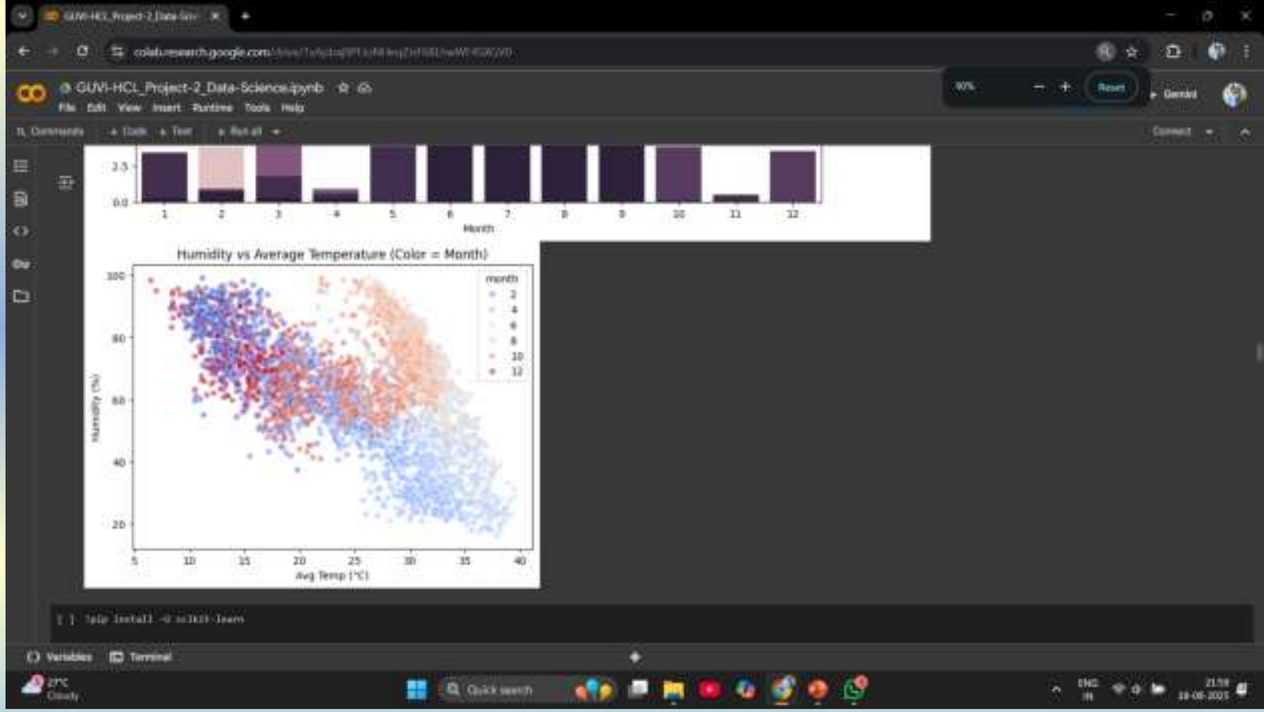


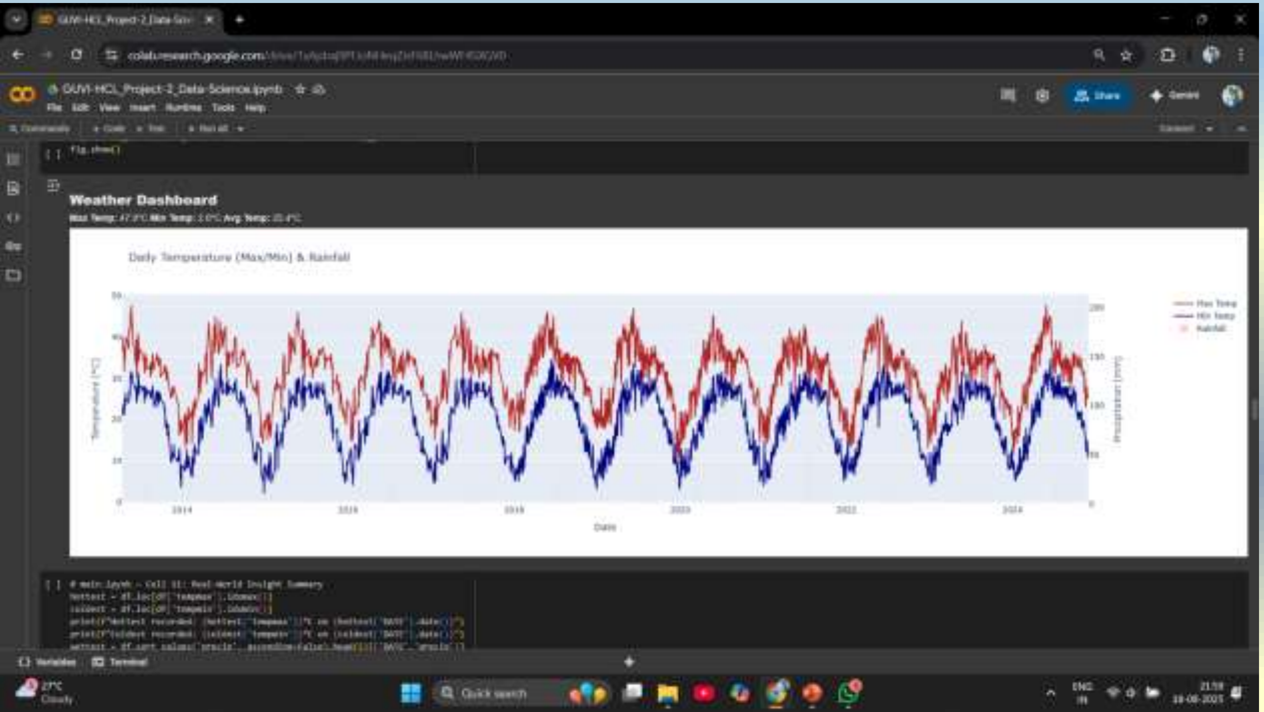
# main.ipynb – Cell 7: Distribution & Correlation
# Temperature range histogram
plt.figure(figsize=(8,4))
sns.histplot(df['temp_range'], bins=30, kde=True, color='magenta')
plt.title("Daily Temperature Range Distribution")
plt.xlabel("Temp Range (°C)")
plt.show()
# Correlation heatmap
plt.figure(figsize=(8,6))
sns.heatmap(df[['temp','tempmax','tempmin','humidity','precip','windspeed']].corr(),
annot=True, cmap='Blues')
plt.title("Weather Metrics Correlation: Temperature, Humidity, Precip, Windspeed")
plt.show()
Daily Temperature Range Distribution
Count
Temp Range (°C)
Variables
Terminal
27°C
Cloudy
Quick search

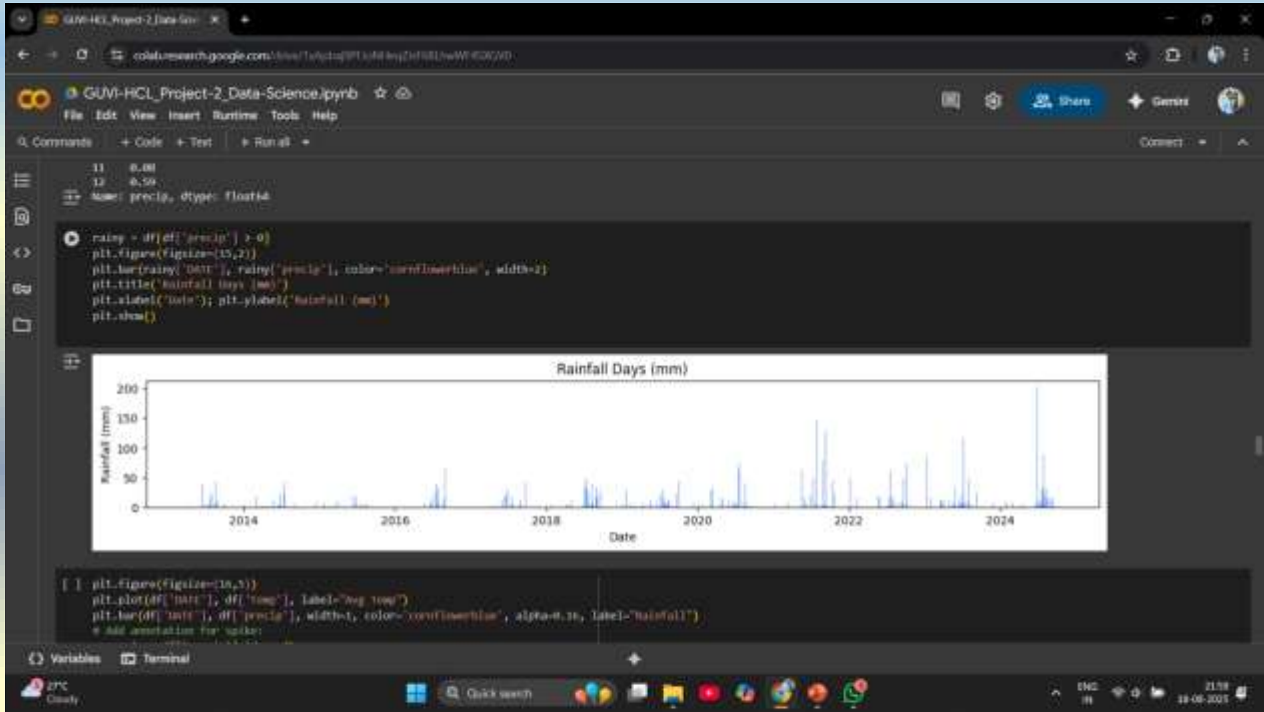


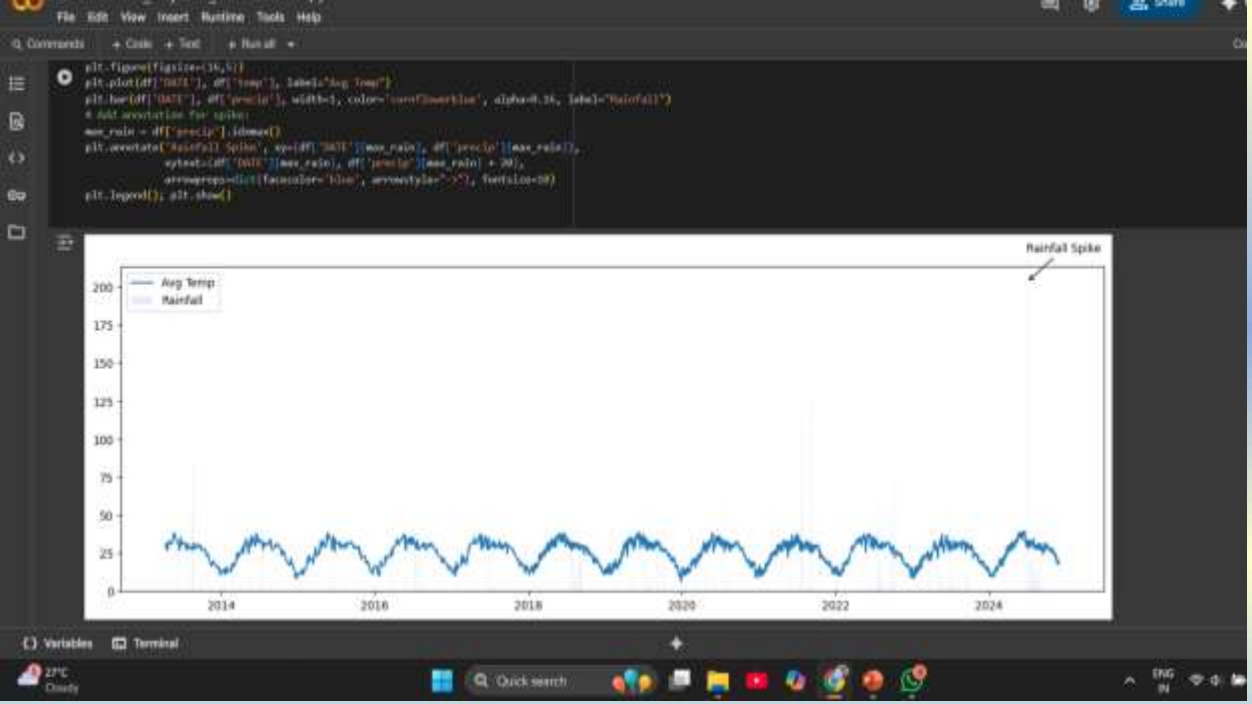
GUVI-HCL_Project-2_Data-Science.ipynb
File Edit View Insert Runtime Tools Help
Commands
+ Code
+ Text
Run all
Temp Range (°C)
Weather Metrics Correlation: Temperature, Humidity, Precip, Windspeed
temp
tempmax
tempmin
humidity
precip
windspeed
# main.ipynb – Cell 8: Monthly Average Precipitation
monthly_prcp = df.groupby(['Year', 'month'])['precip'].mean().reset_index()
plt.figure(figsize=(12,5))
Variables
Terminal
27°C
Cloudy
Quick search


GUVI-HCL_Project-2_Data-Science.ipynb
Monthly Average Precipitation by Year
2014
2016
2018
2020
2022
2024
Avg Precip (mm)
Month

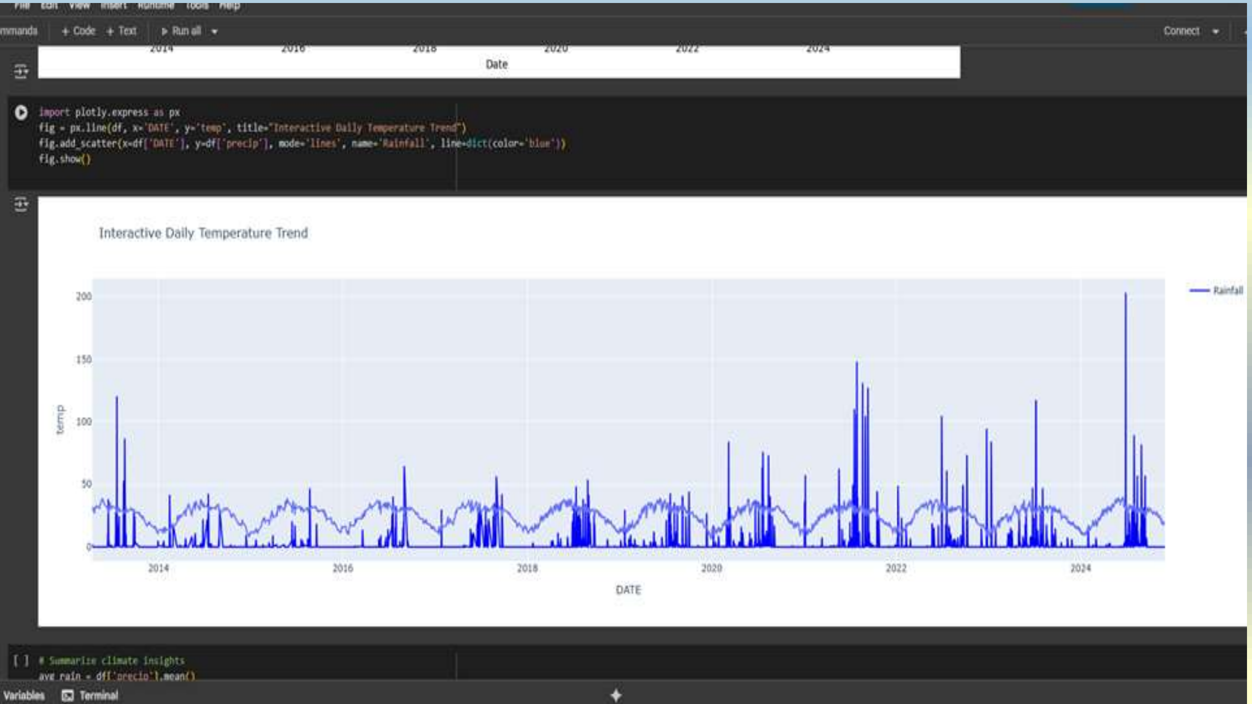

GUVI-HCL_Project-2_Data-Science.ipynb
Humidity vs Average Temperature (Color = Month)
month
Humidity (%)
Avg Temp (°C)
!pip install -U scikit-learn

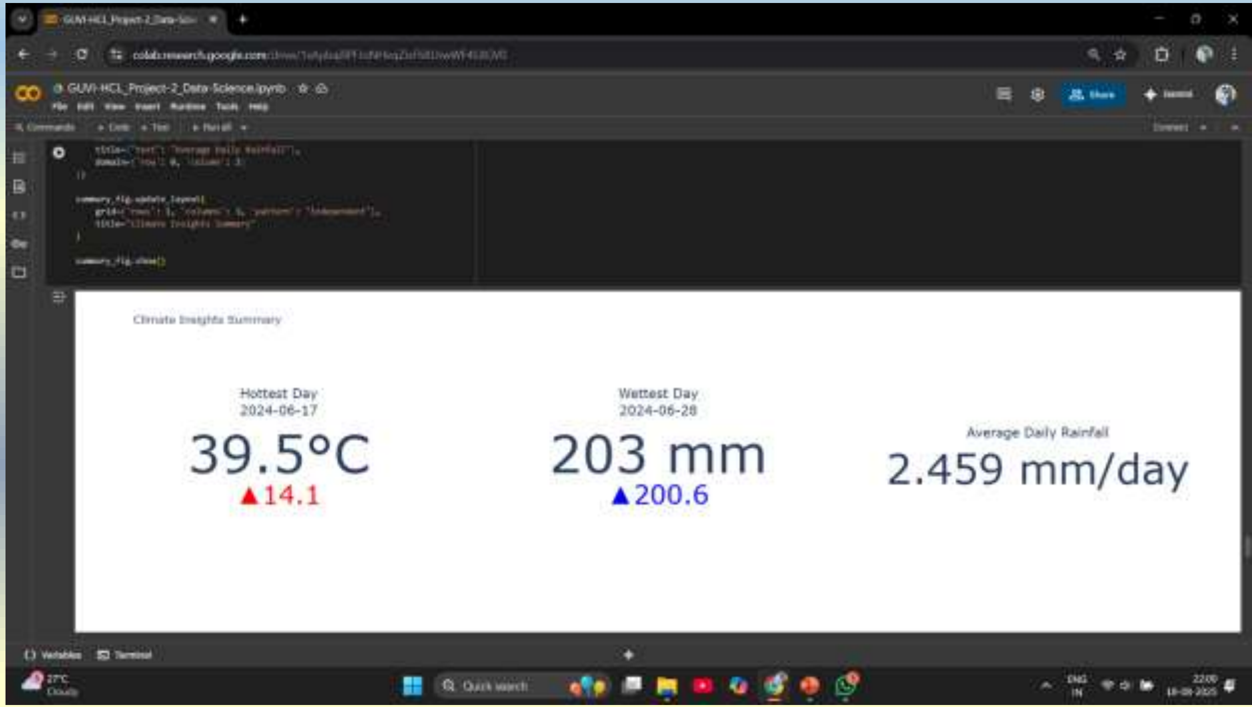

GUVI-HCL_Project-2_Data-Science.ipynb
Weather Dashboard
Daily Temperature (Max/Min) & Rainfall
Max Temp
Min Temp
Rainfall
Temperature (°C)
Precipitation (mm)
Date

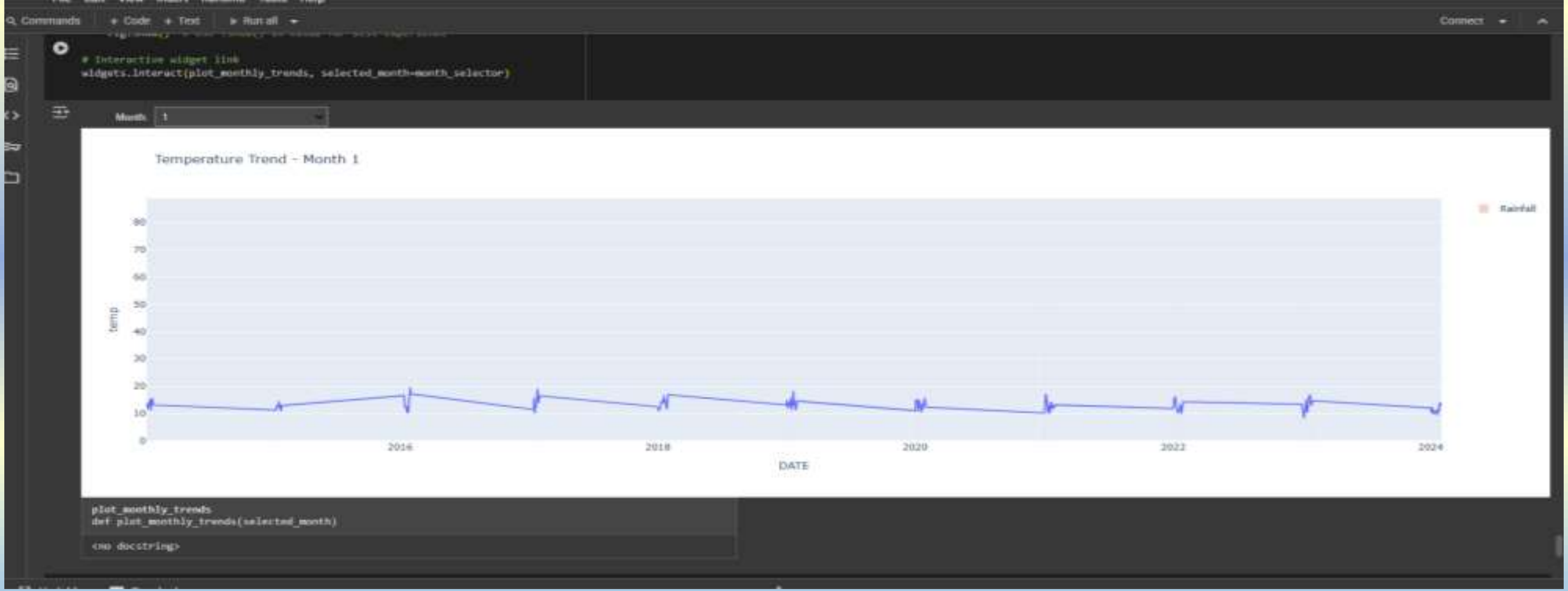

GUVI-HCL_Project-2_Data-Science.ipynb
Name: precip, dtype: float64
rainy = df[df['precip'] > 0]
plt.figure(figsize=(15,2))
plt.bar(rainy['DATE'], rainy['precip'], color='cornflowerblue', width=2)
plt.title('Rainfall Days (mm)')
plt.xlabel('Date'); plt.ylabel('Rainfall (mm)')
plt.show()
Rainfall Days (mm)
Rainfall (mm)
Date


Rainfall Spike
Avg Temp
Rainfall


Temperature Spikes Highlighted
Avg Temp


import plotly.express as px
fig = px.line(df, x='DATE', y='temp', title='Interactive Daily Temperature Trend')
fig.add_scatter(x=df['DATE'], y=df['precip'], mode='lines', name='Rainfall', line=dict(color='blue'))
fig.show()
Interactive Daily Temperature Trend
Rainfall
temp
DATE
Summarize climate insights


GUVI-HCL_Project-2_Data-Science.ipynb
Climate Insights Summary
Hottest Day
2024-06-17
39.5°C
▲14.1
Wettest Day
2024-06-28
203 mm
▲200.6
Average Daily Rainfall
2.459 mm/day



Insights from the Analysis

The generated plots provide clear evidence of seasonal weather shifts. For instance, months typically associated with summer show an upward trend in temperature and relatively minimal rainfall, whereas monsoon months feature pronounced spikes in precipitation. These visualizations enable quick identification of weather anomalies and can inform planning for agricultural cycles, outdoor activities, and resource allocation.

Conclusion

This project successfully demonstrates how data science techniques can transform raw meteorological data into actionable insights for a wide range of users. By collecting daily temperature and rainfall records and systematically preprocessing the data, we ensured that the subsequent analysis was both accurate and meaningful. The use of interactive tools such as Plotly and ipywidgets enabled the creation of dynamic visualizations, allowing users to explore trends month by month and gain a granular understanding of seasonal weather patterns.

The resulting visualizations highlighted clear cycles in temperature and rainfall, revealing trends such as peak summer heat and intensified monsoon precipitation. These insights are not only valuable for academic study but also have practical implications for agriculture, urban planning, event management, and personal scheduling. For instance, farmers could use these monthly patterns to optimize planting schedules, while city planners might better anticipate periods of heavy rainfall to improve infrastructure readiness.

Moreover, this approach underscores the importance of interactive, user-friendly data analysis in making complex information accessible to a broader audience. The ability to select specific months and instantly visualize changes empowers users to dig deeper into the data, uncovering correlations and anomalies that might otherwise go unnoticed in static reports or spreadsheets. Such methods foster data-driven decision-making and promote a culture of openness and exploration around environmental information.

Overall, this project highlights the critical role of data science in bridging the gap between environmental data and everyday applications. The workflow established here can be extended to other datasets or tailored to specific local needs, providing a scalable foundation for future weather analytics and decision support systems. As climate continues to impact daily life, accessible analytical tools like these will become ever more essential in helping society adapt and thrive.