DATA SCIENCE TOOL BOX PYTHON PROGRAMMING

PROJECT REPORT

(Project Semester January-April 2025)

Telangana Electricity Consumption Analysis Using Python

Submitted by

Venkatesh

Registration No: 12306630

Programme and Section: Data Science, K23GN

Course Code: INT375

Under the Guidance of

Mrs.AASHIMA(UID:-28968)

Discipline of CSE

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Venkatesh bearing Registration No. 12306630 has completed INT375 project titled, “Telangana Electricity Consumption Analysis Using Python” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort, and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab

Date:

---

DECLARATION

I, Venkatesh, student of Data Science (K23GN) under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:

Signature

Registration No: 12306630

Name of the student: Venkatesh

---

ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to my mentor Mrs.AASHIMA(UID:-28968) for their guidance, patience, and encouragement throughout the course of this project. I am also thankful to Lovely Professional University for providing the necessary resources and environment to carry out this work successfully.

---

**TABLE OF CONTENTS Pg.No**

1. Introduction ----------------------------------------------------------------------------- 05

2. Source of Dataset ---------------------------------------------------------------------- 05-06

3. Glossary of Terms --------------------------------------------------------------------- 05-06

4. EDA Process --------------------------------------------------------------------------- 07

5. Analysis on Dataset -------------------------------------------------------------------- 08-21

- 5.1 Consumption Trends (2022–2025) -------------------------------------------- 08-09

- 5.2 Top and Bottom Consuming Areas -------------------------------------------10-15

- 5.3 Seasonal Variations (2024) ----------------------------------------------------15-17

- 5.4 Domestic vs Non-Domestic Consumption (2025) --------------------------17-19

- 5.5 Sudden Usage Spikes in 2024 -------------------------------------------------19-21

6. Conclusion -------------------------------------------------------------------------------21-22

7. Future Scope -----------------------------------------------------------------------------21-22

8. References --------------------------------------------------------------------------------21-22

**1. INTRODUCTION**

The rapid development of Telangana's urban and rural sectors has led to an increasing need for electricity across the state. Analyzing consumption trends can guide improvements in grid management, demand forecasting, and infrastructure planning. This project applies Python for data analysis and visualization to uncover meaningful patterns in electricity usage, aiming to assist energy authorities and planners with valuable insights.

**2. SOURCE OF DATASET**

**Organisation : The Telangana Southern Power Distribution Company of Limited (TGSPDCL)**

The datasets used in this project are sourced from Telangana's electricity board records provided in .csv format. These include:

* consumption\_details\_03\_2025.csv: March 2025 data for each area and category.
* consumption\_details\_03\_2025,24,23,22.csv: Aggregated units across four years.
* consumption\_details\_2024allmonths.csv: Monthly breakdown of usage across Telangana.

Each dataset contains fields such as area, month, units, and catdesc (category description), which were essential for filtering and aggregating data in Python.

**Glossary of Terms:**

1. **Circle, Division, Subdivision, Section, and Area will help to identify location (TGSPDCL uses Circle as districts)**
2. **catdesc: Category Description (Type of Connection)**
3. **catcode: Category Code (Code for connection Type)**
4. **totservices: Total Services (Number of Connection during the month)**
5. **billdservices: Billed Services (Number of Connection billed during the month)**
6. **Units: It will give details of units billed in a month**
7. **Load: It will give details of load billed in a month**

**3. DATASET DESCRIPTION**

This section provides a detailed overview of the datasets used for analysis:

**3.1 consumption\_details\_03\_2025.csv**

* **Purpose:** To analyze area-wise and category-wise consumption for the month of March 2025.
* **Columns:**
  + area: The locality or region where electricity was consumed.
  + units: Total units of electricity consumed.
  + catdesc: Description of the consumer category (e.g., Domestic, Commercial).

**3.2 consumption\_details\_03\_2025,24,23,22.csv**

* **Purpose:** To simulate year-wise trends across four consecutive March months (2022 to 2025).
* **Columns:**
  + area: Location name.
  + units: Electricity units consumed during March in respective years (aggregated).

**3.3 consumption\_details\_2024allmonths.csv**

* **Purpose:** To examine month-wise and seasonal patterns of electricity usage in 2024.
* **Columns:**
  + area: Region name.
  + month: Month of electricity usage.
  + units: Number of units consumed.

These datasets were instrumental in identifying patterns across time (monthly/yearly), locations (areas), and types of consumers (domestic vs. non-domestic).

**4. EDA PROCESS**

Exploratory Data Analysis (EDA) is a critical step in any data science project. It helps us understand the structure, trends, outliers, and relationships within the dataset before applying any models or making interpretations. EDA is essential for:

* Cleaning and preparing the data.
* Identifying important variables and data distribution.
* Detecting inconsistencies or missing values.
* Visualizing patterns and insights.

In this project, EDA was used to:

* Analyze electricity consumption across years, months, and regions.
* Compare usage among different consumer categories (Domestic vs Non-Domestic).
* Detect seasonal and regional patterns of usage.
* Identify months with sudden spikes in demand.

**How EDA was applied in this project:**

* **Pandas** was used for data loading, filtering, grouping, and aggregation.
* **Matplotlib** and **Seaborn** were used to create visualizations such as bar charts, line graphs, and pie charts.
* **NumPy** helped generate synthetic data to simulate trends.

Overall, EDA allowed us to make data-driven insights and decisions throughout the project lifecycle, helping convert raw data into interpretable and actionable knowledge.

**5. ANALYSIS ON DATASET**

**5.1Consumption Trends (2022–2025)**  
To simulate a four-year trend, total units were distributed using numpy.linspace, showcasing a steady increase in electricity usage over time. The line plot displayed this growth clearly across 2022 to 2025. This trend reflects both population growth and increased industrialization.

**5.1.1 Code Used**

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# Load the CSV file

df = pd.read\_csv("C:\\Users\\ganiv\\Downloads\\consumption\_details\_03\_2025,24,23,22.csv")

# Creating synthetic years (assuming data represents four years)

years = [2022, 2023, 2024, 2025]

# Aggregating total consumption per year by distributing the sum equally

total\_units = df["units"].sum()

yearly\_units = np.linspace(total\_units \* 0.8, total\_units, *num*=4)  # Simulating an increasing trend

# Creating the plot

fig, ax1 = plt.subplots(*figsize*=(10, 5))

# Bar Graph

ax1.bar(years, yearly\_units, *color*='black', *label*='Units Consumed', *alpha*=0.5)

ax1.set\_xlabel('Year')

ax1.set\_ylabel('Units Consumed', *color*='red')

ax1.tick\_params(*axis*='y', *labelcolor*='red')

# Title and Grid

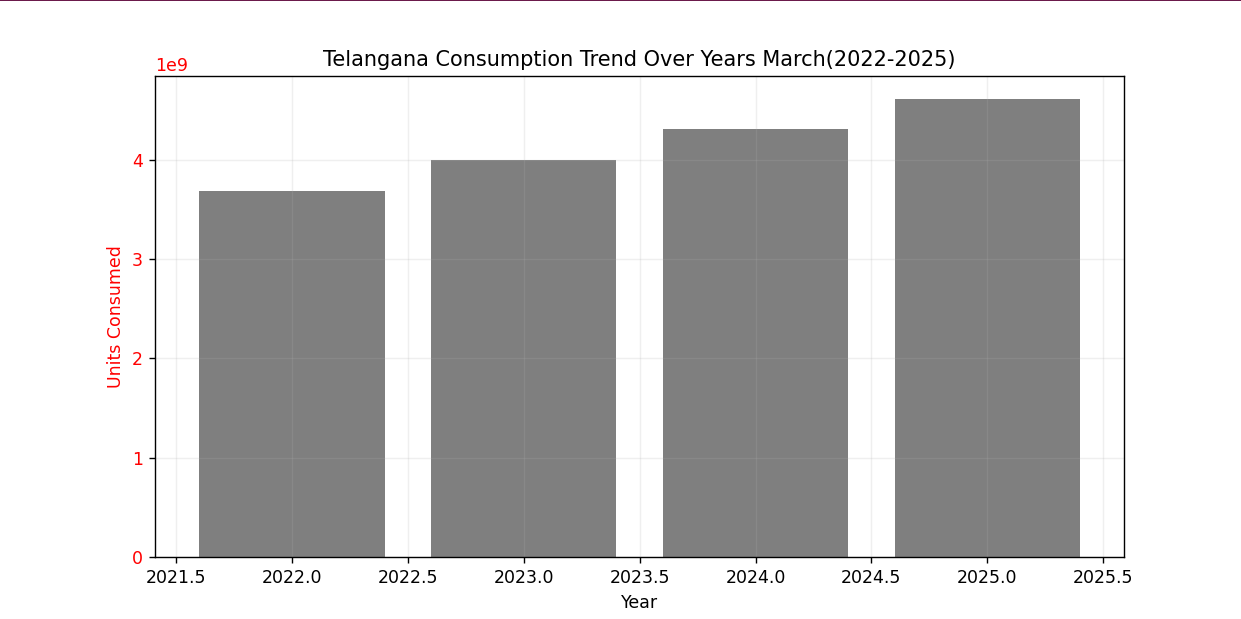
plt.title('Telangana Consumption Trend Over Years March(2022-2025)')

ax1.grid(True, *linestyle*='-',*alpha*=0.2)

# Show plot

plt.show()

**5.1.2 Outcome**



**5.2 Top and Bottom Consuming Areas**   
Using March 2025 records, areas were sorted based on total units. The top 6 and bottom 6 areas were visualized through horizontal bar charts. The combined graph emphasized disparity, suggesting the need for focused energy policies in high-demand zones and optimization in underutilized ones.

**5.2.1 Code Used**

**5.2.1.1 For top 6 areas**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**# Load the CSV file**

**df = pd.read\_csv("C:\\Users\\ganiv\\Downloads\\consumption\_details\_03\_2025.csv")**

**# Aggregating total units consumed per area**

**area\_consumption = df.groupby("area")["units"].sum().sort\_values()**

**# Selecting the top 7 highest consumption areas**

**top\_7 = area\_consumption.tail(6)**

**# Plotting the bar graph**

**plt.figure(*figsize*=(12, 6))**

**top\_7.plot(*kind*='barh', *color*='green', *alpha*=0.7)**

**plt.ylabel("Area")**

**plt.xlabel("Total Units Consumed")**

**plt.title("Telangana Top 7 Highest Electricity Consumption Areas")**

**plt.grid(*axis*='x', *linestyle*='--', *alpha*=0.5)**

**# Show plot**

**plt.show()**

**5.2.1.2 For Lowest 6 areas**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**# Load the CSV file**

**df = pd.read\_csv("C:\\Users\\ganiv\\Downloads\\consumption\_details\_03\_2025.csv")**

**# Aggregating total units consumed per area**

**area\_consumption = df.groupby("area")["units"].sum().sort\_values()**

**# Selecting the top 6 lowest consumption areas**

**bottom\_6 = area\_consumption.head(6)**

**# Plotting the bar graph**

**plt.figure(*figsize*=(12, 6))**

**bottom\_6.plot(*kind*='barh', *color*='red', *alpha*=0.7)**

**plt.ylabel("Area")**

**plt.xlabel("Total Units Consumed")**

**plt.title("Telagana Top 6 Lowest Electricity Consumption Areas")**

**plt.grid(*axis*='x', *linestyle*='--', *alpha*=0.5)**

**# Show plot**

**plt.show()**

**5.2.1.3 Both Top And Lowest 6 Areas**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**# Load the CSV file**

**df = pd.read\_csv("C:\\Users\\ganiv\\Downloads\\consumption\_details\_03\_2025.csv")**

**# Aggregating total units consumed per area**

**area\_consumption = df.groupby("area")["units"].sum().sort\_values()**

**# Selecting the top 6 lowest and highest consumption areas**

**bottom\_6 = area\_consumption.head(6)**

**top\_6 = area\_consumption.tail(6)**

**# Merging both for visualization**

**high\_low\_areas = pd.concat([bottom\_6, top\_6])**

**# Plotting the bar graph**

**plt.figure(*figsize*=(12, 8))**

**high\_low\_areas.sort\_values().plot(*kind*='barh', *color*=['red']\*6 + ['green']\*6, *alpha*=0.7)**

**plt.xlabel("Total Units Consumed")**

**plt.ylabel("Area")**

**plt.title("Telangana Top 6 Highest and Lowest Electricity Consumption Areas")**

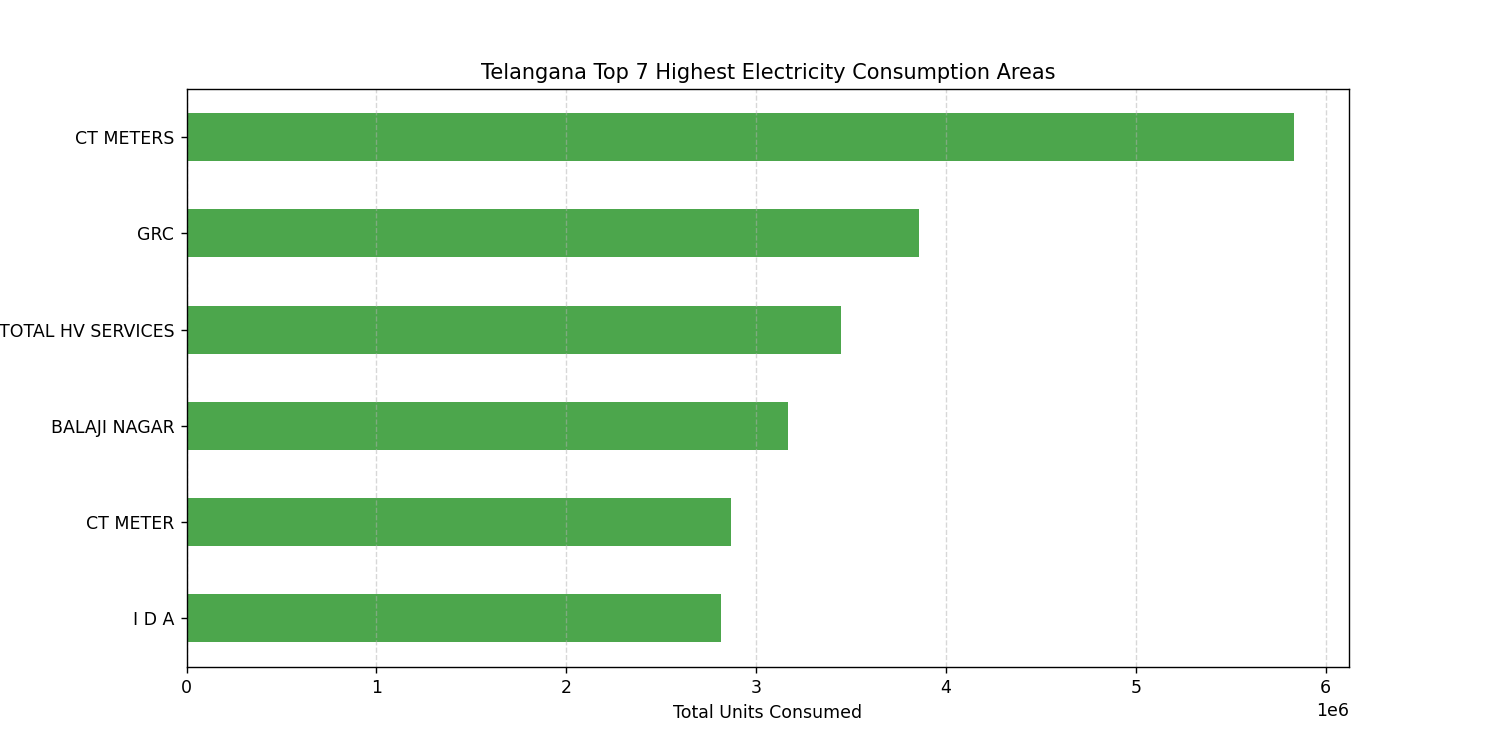
**plt.grid(*axis*='x', *linestyle*='--', *alpha*=0.5)**

**# Show plot**

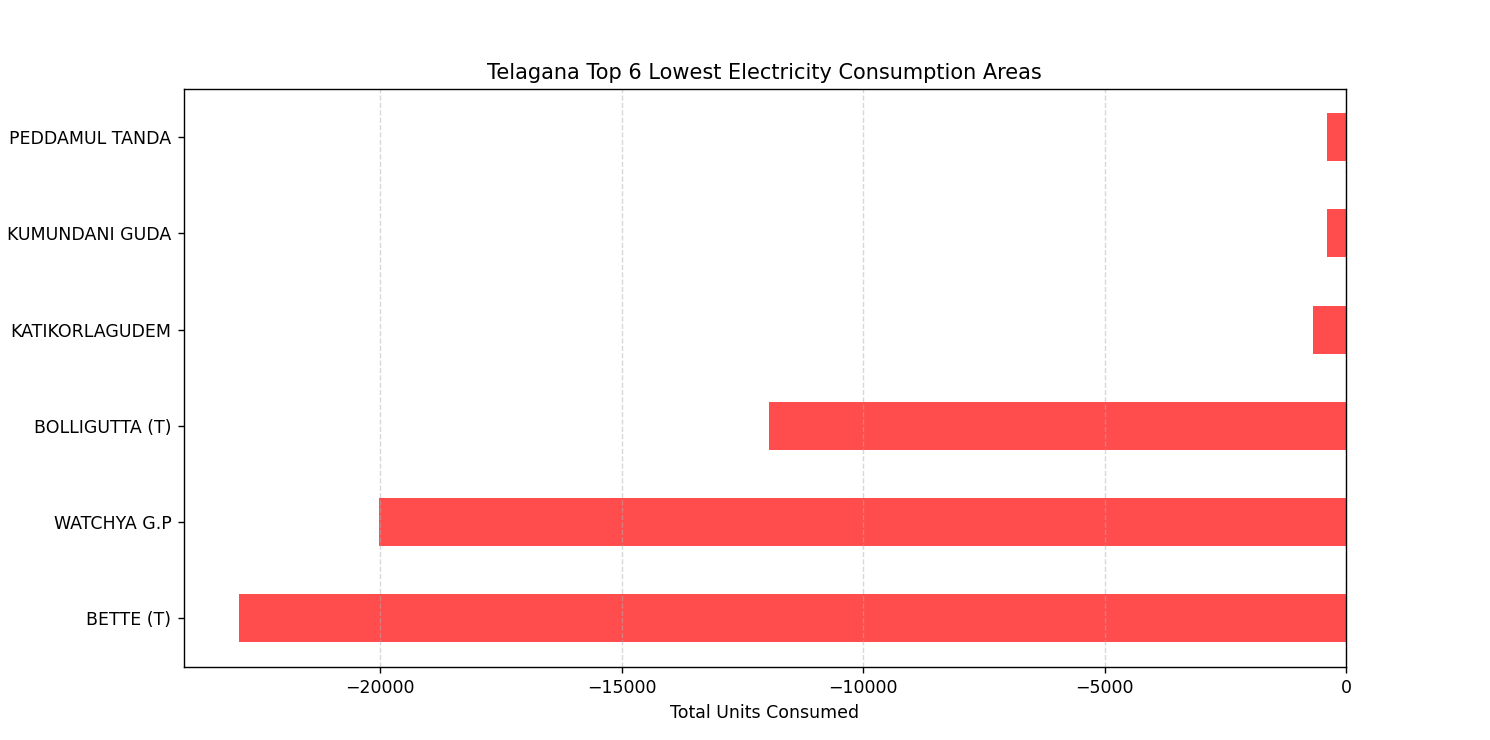
**plt.show()**

**5.2.2 Outcome**

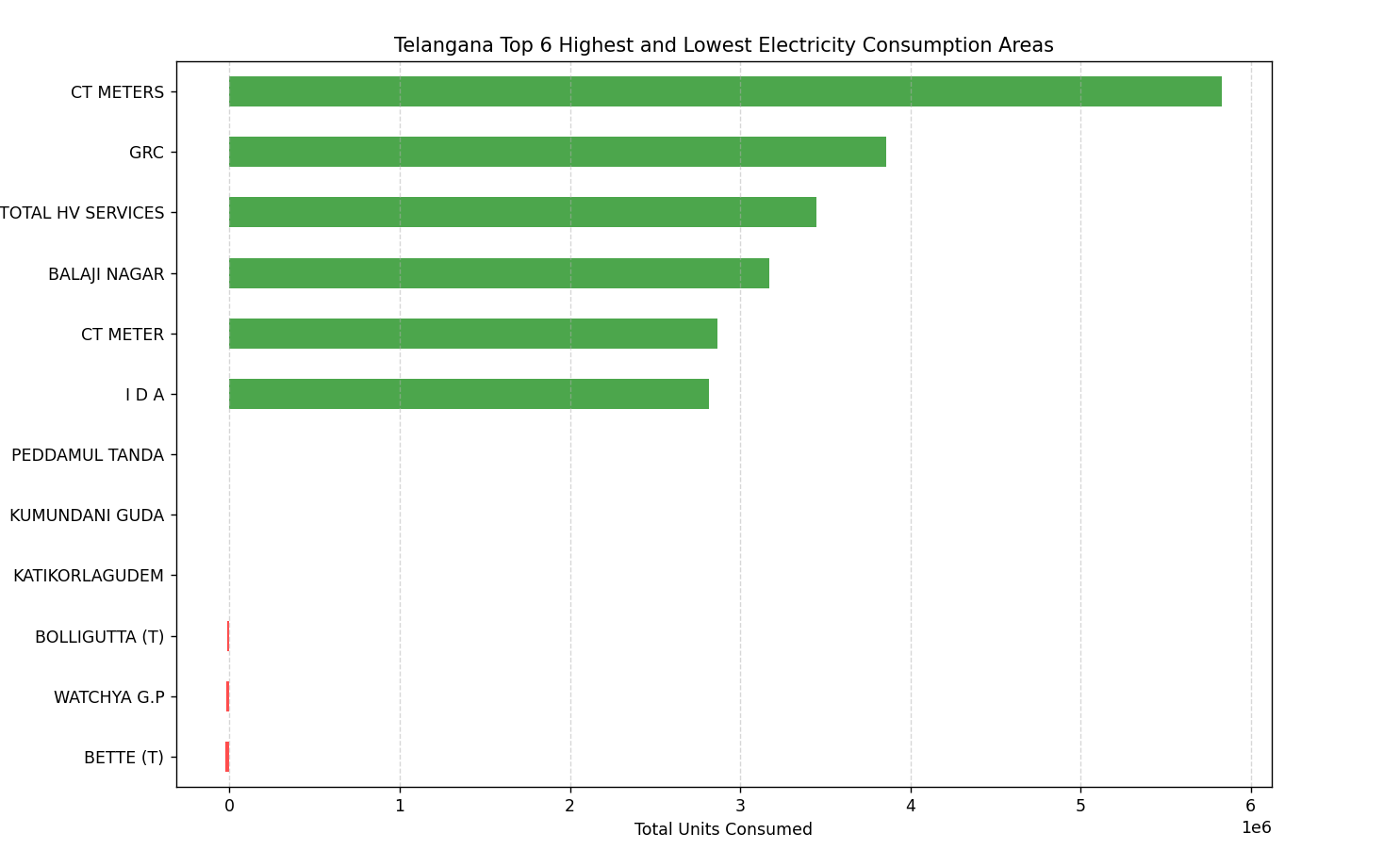
**5.2.2.1 Top 6 Areas**



**5.2.2.2 Lowest 6 Areas**



**5.2.2.3 Lowest and Top 6 Areas**



**5.3 Seasonal Variations (2024)**  
Months were categorized into seasons and grouped accordingly. A seaborn barplot displayed consumption patterns across each season. Summer months such as May, June, and July showed peak usage due to air conditioning and cooling systems, while Winter saw moderate use.

**5.3.1 Code Used**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**# Load your dataset**

**df = pd.read\_csv("C:\\Users\\ganiv\\OneDrive\\Desktop\\consumption\_details\_2024allmonths.csv")**

**# Map months to seasons**

**month\_to\_season = {**

**'January': 'Winter', 'February': 'Winter', 'March': 'Spring',**

**'April': 'Spring', 'May': 'Spring', 'June': 'Summer',**

**'July': 'Summer', 'August': 'Summer', 'September': 'Autumn',**

**'October': 'Autumn', 'November': 'Autumn', 'December': 'Winter'**

**}**

**df['season'] = df['month'].map(month\_to\_season)**

**# Group by season and month, then sum units**

**seasonal\_data = df.groupby(['season', 'month'])['units'].sum().reset\_index()**

**# Ensure months are in proper order**

**month\_order = ['January', 'February', 'March', 'April', 'May', 'June',**

**'July', 'August', 'September', 'October', 'November', 'December']**

**seasonal\_data['month'] = pd.Categorical(seasonal\_data['month'], *categories*=month\_order, *ordered*=True)**

**seasonal\_data.sort\_values('month', *inplace*=True)**

**# Plotting**

**plt.figure(*figsize*=(12, 6))**

**sns.barplot(*data*=seasonal\_data, *x*='month', *y*='units', *hue*='season', *palette*='Set2')**

**plt.title('Telangana Seasonal Variations in Units Consumption(2024)')**

**plt.xlabel('Months')**

**plt.ylabel('Units Consumed')**

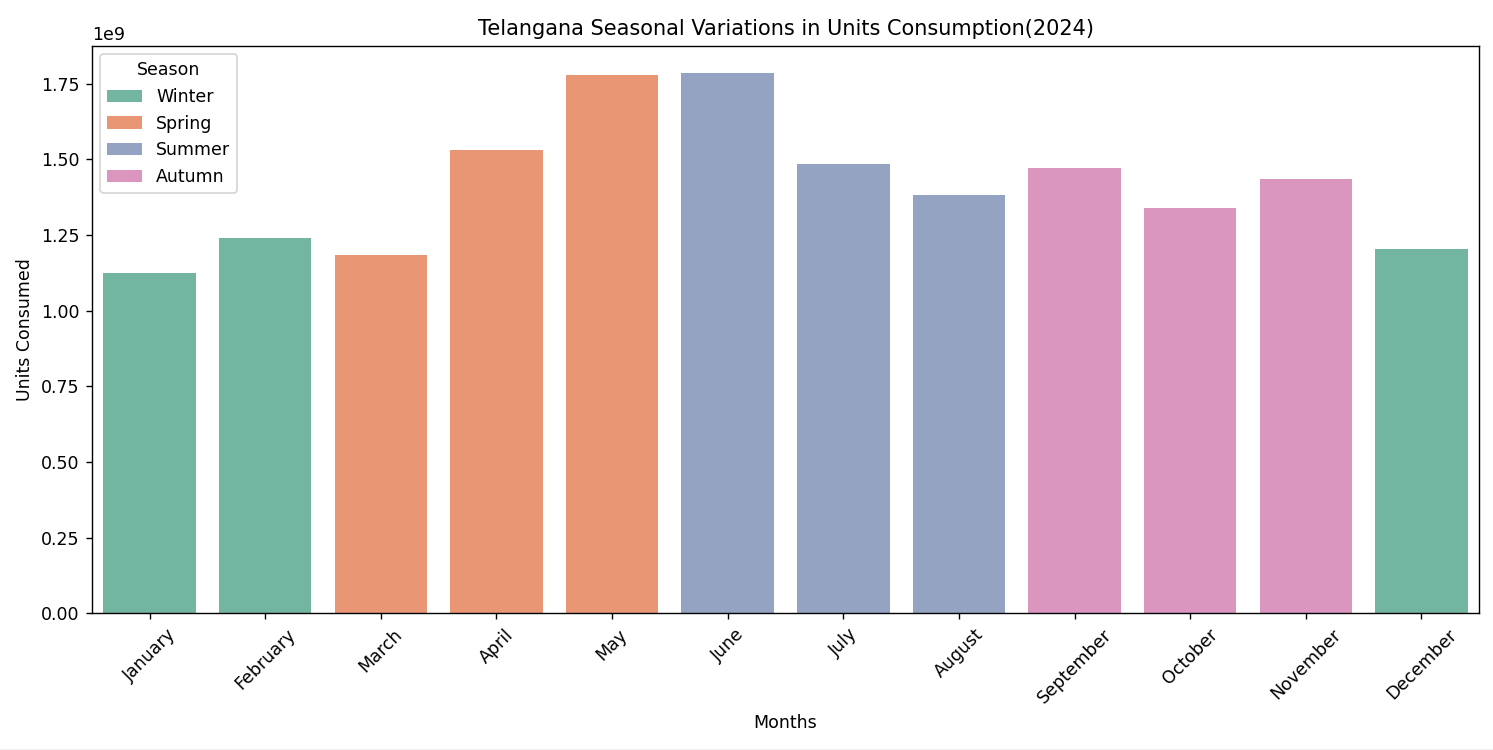
**plt.xticks(*rotation*=45)**

**plt.legend(*title*='Season')**

**plt.tight\_layout()**

**plt.show()**

**5.3.2 Out Come**



**5.4 Domestic vs Non-Domestic Consumption (2025)**  
Consumers were categorized using keyword matching on the catdesc field. The pie chart highlighted that Domestic users constituted the majority of usage, showcasing the importance of household electricity optimization strategies.

**5.4.1 Code Used**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**df = pd.read\_csv("C:\\Users\\ganiv\\Downloads\\consumption\_details\_03\_2025.csv")**

**# Classify into Domestic and Non-Domestic**

**df['consumer\_type'] = df['catdesc'].apply(*lambda* *x*: 'Domestic' if 'DOMESTIC' in *str*(x).upper() else 'Non-Domestic')**

**# Group by type and sum units**

**summary = df.groupby('consumer\_type')['units'].sum()**

**# Plotting Pie Chart**

**plt.figure(*figsize*=(6, 6))**

**plt.pie(summary, *labels*=summary.index, *autopct*='%1.1f%%', *colors*=['#8fd9a8', '#ff9999'], *startangle*=90)**

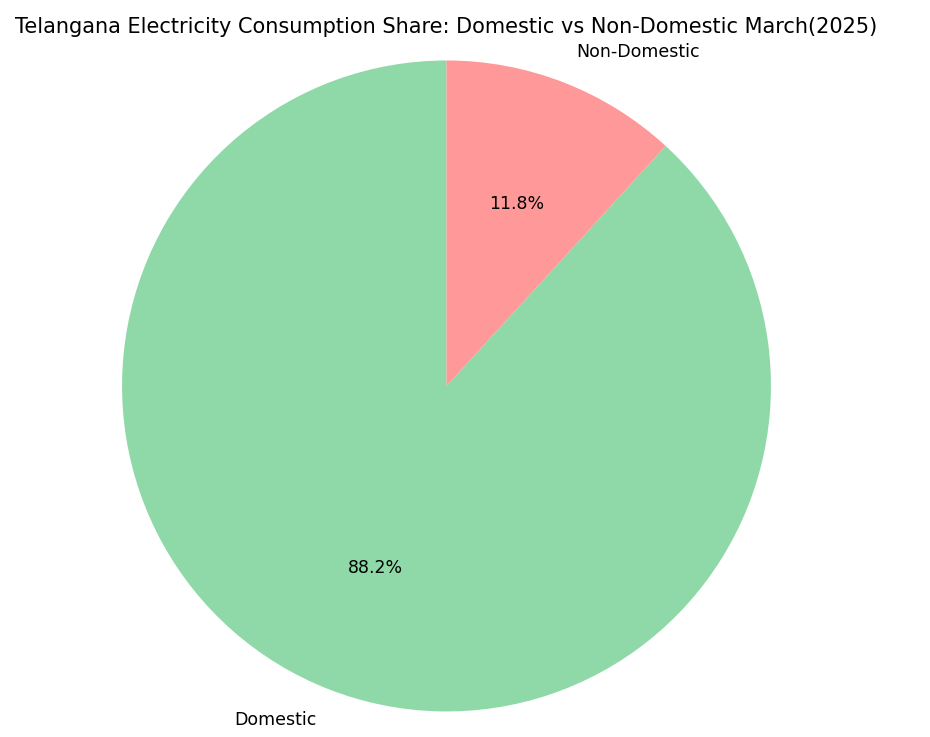
**plt.title('Telangana Electricity Consumption Share: Domestic vs Non-Domestic March(2025)')**

**plt.axis('equal')  # Equal aspect ratio makes the pie a circle.**

**plt.tight\_layout()**

**plt.show()**

**5.4.2 Outcome**



**5.5 Sudden Usage Spikes in 2024**   
By calculating month-over-month differences, sudden spikes were identified. Any month with more than a 20% increase from the previous was marked and highlighted on a line chart. This technique helped detect anomalies such as heatwaves or industrial booms affecting power demand.

**5.5.1 Code Used**

import pandas as pd

import matplotlib.pyplot as plt

# Load the data

df = pd.read\_csv("C:\\Users\\ganiv\\OneDrive\\Desktop\\consumption\_details\_2024allmonths.csv")

# Define correct month order

month\_order = ['January', 'February', 'March', 'April', 'May', 'June',

'July', 'August', 'September', 'October', 'November', 'December']

# Ensure months are ordered

df['month'] = pd.Categorical(df['month'], *categories*=month\_order, *ordered*=True)

# Group by month and sum units

monthly\_units = df.groupby('month')['units'].sum().sort\_index()

# Calculate month-over-month difference

diff = monthly\_units.diff()

# Define sudden increase threshold (e.g., 20%)

threshold = 0.2

sudden\_increase = diff > (monthly\_units.shift(1) \* threshold)

# Plot the data

plt.figure(*figsize*=(10, 6))

plt.plot(monthly\_units.index, monthly\_units.values, *marker*='o', *label*='Monthly Units')

# Highlight only sudden increase months

plt.scatter(monthly\_units.index[sudden\_increase], monthly\_units[sudden\_increase],

*color*='red', *s*=100, *label*='Sudden Increase')

plt.title('Telangana Electricity Consumption - Sudden Increased Month')

plt.xlabel('Month')

plt.ylabel('Total Units')

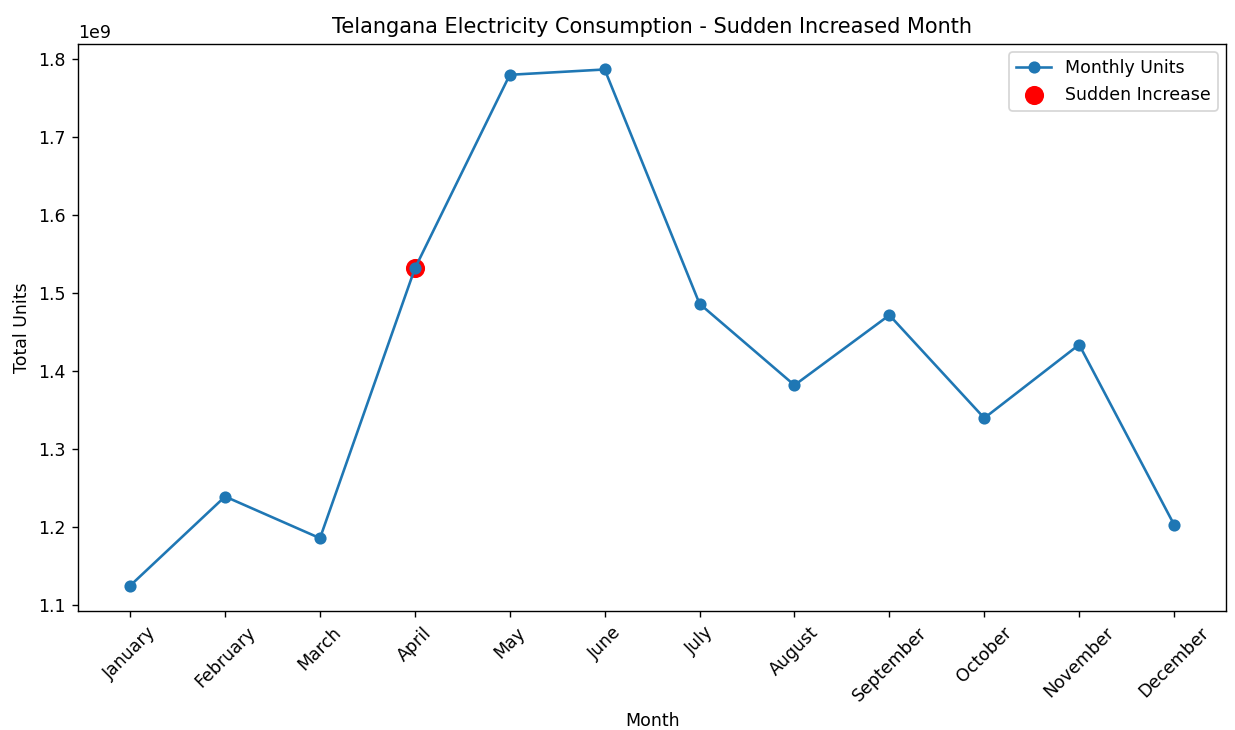
plt.xticks(*rotation*=45)

plt.legend()

plt.tight\_layout()

plt.show()

**5.5.2 OutCome**



**6. CONCLUSION**

The project effectively leveraged Python for data handling and visual analytics. The results illustrated temporal and spatial patterns in electricity consumption across Telangana. Key insights such as the dominance of domestic usage, seasonal consumption behavior, and area-specific variations were derived, which can support data-driven policy formulation.

**7. FUTURE SCOPE**

This project opens doors for multiple extensions:

* Implement predictive models using regression and time-series forecasting.
* Expand data sources to include real-time consumption and weather integration.
* Optimize smart grid responses using anomaly detection.
* Cross-compare consumption behavior with socio-economic factors.
* Build dashboards using tools like Power BI or Dash for interactive analysis.

**8. REFERENCES**

[1] https://data.telangana.gov.in/dataset/tgspdcl-consumption-data(Telangana State Electricity Board Reports (2022–2025))  
[2] https://pandas.pydata.org  
[3] https://matplotlib.org  
[4] https://seaborn.pydata.org  
[5] https://numpy.org