# CAPSTONE Intelligent abnormal electricity usage

*Energy and Power*

## Team Details :

Project Lead              : Tejaswini Palwai
Data Lead                 : Utkarsh Jain
Strategy Lead             : Anwesha Adhikari
PPT & Quality Lead : Sankalp M Tellur
Dashboard Lead        : Kashika Agarwal
Analysis Lead             : Lokendra Singh

## Project :

DVA (DATA VISUALIZATION  & ANALYSIS)

# Problem :

The objective of this project is to design a district-level anomaly detection framework for January consumption data in Kerala, enabling electricity authorities to identify high-risk consumers, prioritize inspection resources, and prevent operational losses proactively.

# Our Approach :

We analyse January residential electricity data for Kerala, ensuring clean and reliable records. By calculating usage deviations and cluster-level patterns, we detect anomalies and highlight high-risk consumers. An interactive dashboard enables authorities to prioritise inspections and prevent operational losses proactively.

# Key Insights and Recommendations :

1. **Data-Driven Monitoring:**
   Leverage cleaned data, deviation metrics, and cluster analysis through an interactive dashboard to continuously identify abnormal consumption patterns.
2. **Implement Real-Time Detection:**
   Deploy continuous monitoring to flag high deviation cases early and proactively reduce operational losses.
3. **Prioritize High-Risk Clusters:**
   Focus inspections and audits on clusters with consistently high average consumption to optimize resource allocation.
4. **Promote Energy Efficiency:**
   Encourage adoption of energy-efficient appliances and optimised connected load to reduce excessive usage at the household level.

## Sector Overview

This project falls within the energy and utilities sector, focusing on residential electricity consumption monitoring and anomaly detection.

## Current Challenges

Utilities face abnormal usage, demand fluctuations, inefficient consumption patterns, data inconsistencies, and limited real-time anomaly detection.

## Why This Problem

To identify high-risk consumers, detect abnormal usage patterns, and enable data-driven decision-making for improved energy efficiency and operational control.

## Formal Problem Definition

Analyse residential electricity consumption data to detect abnormal usage, measure deviations between expected and actual consumption, and identify key drivers of variability.

## Project Scope

Includes data cleaning, KPI computation (usage deviation, cluster averages), pivot-based summaries, and development of an interactive dashboard for monitoring and analysis. Excludes predictive model deployment.

## Success Criteria

- Accurate KPI and deviation calculations
- Clear detection of abnormal consumption
- Meaningful cluster segmentation
- Insightful interactive dashboard
- Actionable recommendations for energy management

# DATA DESCRIPTION :

The dataset used in this project is a residential electricity consumption dataset obtained for academic analysis purposes only.

## SOURCE LINK:   https://deepdatalake.com/details.php?dataset_id=112

The dataset includes household characteristics, environmental factors, energy consumption values, calculated KPIs, and clustering outputs.

## Final Cleaned Column Explanation :

| | |
|---|---|
| Meter_Id (prefix removed) | Numeric meter ID for simplified analysis |
| Cleaned_Date | Date of energy record |
| Week_number | Week of the month |
| District | Region name |
| Labeled_Dwelling | Type of residence in labelled format |
| Num_Occupants | Number of residents |
| House_Area (sqft) | Area of the house |
| Appliance_Score | Indicator of appliance usage intensity |
| Appliance_Score_Grouped | Category segregation for the |
| Connected_Load (kW) | Maximum electrical load capacity |
| Load_category | Categorised load level |
| Temperature_C | Recorded temperature |
| Temperature_category | Categorised temperature range |
| Humidity (%) | Humidity level |
| Cleaned_Expected_Energy (kWh) | Estimated energy consumption |
| Cleaned_Actual_Energy (kWh) | Actual recorded consumption |
| Usage_Deviation (%) | Percentage difference between expected and actual |
| Cluster_Avg_Energy (kWh) | Average energy of the assigned cluster |
| Abnormal_Usage | Binary indicator (1 = abnormal, 0 = normal) |
| Pattern_Usage | Usage behavior pattern classification |

**Data Size :** The cleaned dataset contains **10,800 records** with **20 columns**, covering multiple districts and residential types.

**Data Limitations :** Presence of missing values in key energy fields before cleaning , Limited time range (January 2023 only) , Geographic coverage restricted to Kerala.

# Data Cleaning

## Handling Missing Values

- Identified missing values in critical fields such as Expected Energy and Actual Energy.
- Removed or corrected incomplete records to ensure accurate KPI computation.
- Ensured no null values remained in key analytical columns.

## Outlier Treatment

- Identified outliers in Usage Deviation (%) to detect abnormal consumption patterns.
- Validated extreme values to differentiate genuine anomalies from data errors.
- Retained verified outliers as key indicators of abnormal usage.

## Transformations

- Standardized Meter_ID and date formats; extracted weekday information.
- Categorized temperature ranges (<25, 25–30, >30).
- Converted deviation values into percentage format.
- Created load and temperature categories for better segmentation.

## Feature Engineering

- Calculated Usage Deviation (%).
- Created Abnormal_Usage flag based on deviation thresholds.
- Generated Cluster_Avg_Energy (kWh).
- Derived Load_Category and Pattern_Usage for behavior analysis.

## Assumptions

- Expected Energy represents baseline consumption.
- High deviation indicates abnormal usage.
- Temperature and connected load influence consumption patterns.
- Each record represents an independent daily observation.

# KPIs & Its Metrics

**Total Unique Meters:** 360
**Total Abnormal Activities:** 4739
**Average Energy Consumption:** 16.78 kWh
**District with Maximum Abnormal Usage:** Ernakulam
**Abnormal Activity Count – Ernakulam:** 1930

### 1. Total Unique Meters

**Formula:** COUNT(DISTINCT Meter_ID)
**Purpose:** Represents total monitored consumers.
**Impact:** Ensures complete analysis coverage.

### 2. Total Abnormal Activities

**Formula:** SUM(Abnormal_Usage)
**Purpose:** Measures irregular consumption events.
**Impact:** Supports anomaly detection and risk monitoring.

### 3. Average Energy Consumption

**Formula:** AVERAGE(Cleaned_Actual_Energy)
**Purpose:** Indicates overall consumption level.
**Impact:** Helps in demand forecasting and load planning.

### 4. District with Maximum Abnormal Usage

**Formula:** MAX(SUM(Abnormal_Usage) by District)
**Purpose:** Identifies high-risk regions.
**Impact:** Enables targeted intervention.

### 5. Abnormal Activity Count – Ernakulam

**Formula:** SUM(Abnormal_Usage WHERE District = 'Ernakulam')
**Purpose:** Quantifies risk concentration.
**Impact:** Supports focused monitoring and corrective action.

# Exploratory Data Analysis (EDA)

## Trend Analysis

### Week of the Month vs Avg Actual Energy
Average energy consumption remains stable across weeks (16.65–16.86 kWh), indicating consistent usage patterns.

### District-wise Abnormal Usage by Week
Ernakulam recorded the highest abnormal usage (1930), followed by Thiruvananthapuram (1467) and Alappuzha (1342), highlighting higher irregular activity in Ernakulam according to the weekly data.

## Comparison Analysis

### District-wise Average Energy Consumption
Actual consumption is slightly higher than expected across all districts, showing consistent positive deviation.

### Dwelling Type vs Usage Deviation
Independent Houses show the highest deviation (5.99%), followed by Villas and Apartments.

### Temperature Category vs Usage Deviation
Usage deviation is highest when temperature exceeds 30°C (6.37%), confirming climate impact on energy consumption.

## Distribution Analysis

### Appliance Score vs Energy Load Distribution
Higher appliance scores align with higher load categories, indicating increased energy demand.

### Occupants vs Actual Energy Consumption vs Connected Load
Energy consumption increases steadily with household size (13.75 kWh for 1 occupant to 19.40 kWh for 6 occupants).Connected load remains relatively stable across occupants, suggesting consumption changes are usage-driven rather than infrastructure-driven.

## Correlation Analysis

### Load Category vs Actual Energy
Energy consumption rises with the load category (14.74 → 17.69 kWh), showing a positive relationship.

# Advanced Analysis

## Forecasting:

Weekly consumption is stable (~16.7 kWh), indicating steady short-term demand. Higher temperatures (>30°C) may slightly increase future usage.

## Segmentation:

Higher consumption observed in Load Category 3 and households with 4–6 occupants. Ernakulam shows the highest abnormal activity segment.

## Root Cause Analysis:

Abnormal usage is concentrated in Ernakulam (1930 cases), likely influenced by higher load intensity and usage behavior.

## Risk / Anomaly Analysis:

A total of 4739 abnormal activities indicate monitoring importance. High load category and high temperature increase deviation risk.

## Scenario Analysis:

Increase in occupants, load category, or temperature directly leads to higher energy consumption and deviation levels.

# Dashboard Analysis

## Energy Consumption Trends and Efficiency Dashboard

| Total Unique Meters | Total Abnormal Activities | Maximum Abnormal Activity | Average Energy | Leading Abnormal Usage District |
|---|---|---|---|---|
| 360 | 4739 | 1930 | 16.78 | Ernakulam |

Week_Number — All ▾    District — All ▾    Dwelling_Type — All ▾

### Weekly Average Electricity Consumption (kWh)



AVERAGE of Actual_Energy(kwh) vs Week of January

### Energy Benchmark vs Actual Usage



- AVERAGE of Cleaned_Expected_Energy(kwh)
- AVERAGE of Cleaned_Actual_Energy(kwh)

Average of Expected and Actual Energies vs District
(Alappuzha, Ernakulam, Thiruvananthapuram)

### Usage Variance by Dwelling Category



Dwelling_Type (Apartment, Independent House, Villa) vs Average of Usage_Deviation(%)

District-wise abnormal … — All ▾    Temparature_category — All ▾

### Weekly Abnormal Energy Usage by District



- Alappuzha
- Ernakulam
- Thiruvananthapuram

Count of Abnormal Activities vs Week of January(1 to 5)

### Impact of Temperature on Energy Usage Deviation



Average usage Deviation vs Temperature Category (<25, >30, 25-30)

Num_Occupants — All ▾

### Consumption vs Occupancy



AVERAGE of Cleaned_Actual_Energ… vs Num_Occupants

### Occupancy vs Connected Load



Num_Occupants

### Electricity Usage by Load Category



Load_category(labeled) (1, 2, 3) vs AVERAGE of Cleaned_Actual_Energy(kwh)

### Energy Load Distribution by Appliance Score



- Load_category(labeled)

Meters with Abnormal reading vs Appliance Score (Appliance_Score_Grouped, High, Low, Medium)

# Dashboard Implementation

The dashboard was developed in **Google Sheets** using pivot tables, calculated KPIs, structured formulas, and interactive slicers.

All analytical transformations and visualizations were built directly within Sheets, ensuring compliance with capstone requirements.

## Dashboard Objective

The objective of the dashboard is to:

- Detect abnormal electricity consumption patterns
- Identify high-risk districts and consumer segments
- Compare expected vs. actual energy usage
- Enable data-driven inspection prioritization
- Support proactive operational loss prevention

The dashboard acts as a district-level anomaly detection framework for January consumption data in Kerala.

### Trend Analysis Section : Identifies time-based abnormal patterns.

- Weekly Average Electricity Consumption
- Weekly Abnormal Usage by District

### Comparative & Behavioral Analysis : Segments consumers and highlights deviation drivers.

- Expected vs Actual Energy by District
- Usage Deviation by Dwelling Type
- Temperature Impact on Deviation
- Consumption vs Occupancy

### Load & Appliance Intelligence : Identifies structural factors influencing abnormal consumption.

- Energy Usage by Load Category
- Appliance Score vs Load Distribution
- Occupancy vs Connected Load

## Filters & Drilldowns

The dashboard includes interactive filters for:

Week Number , District , Dwelling Type , Temperature Category , Number of Occupants

# Insights Summary

### Total Abnormal Incidents Are Significantly High

4,739 abnormal activities were identified across 360 meters in January.
**Decision:** Implement structured anomaly tracking rather than reactive complaint-based inspection.

### Actual Usage Consistently Exceeds Expected Benchmark

Across districts, actual consumption is higher than expected energy benchmarks.
**Decision:** Strengthen meter audit checks and validate billing accuracy.

### Independent Houses Show Highest Usage Deviation

Independent houses exhibit higher deviation compared to apartments and villas.
**Decision:** Focus inspection and awareness campaigns on independent house consumers.

### High Temperature (>30°C) Increases Energy Deviation

Usage deviation peaks 6.37% during high-temperature periods.
**Decision:** Introduce seasonal monitoring controls and load advisory during peak summer weeks.

### Abnormal Activities Are Consistently Present Across Weeks

Abnormal counts remain steady across Weeks 1–4, indicating systematic issues rather than random spikes.
**Decision:** Implement continuous monitoring instead of one-time corrective action.

### Higher Occupancy Drives Higher Consumption

Energy usage increases steadily with the number of occupants.
**Decision:** Integrate occupancy factor into energy benchmark modeling.

### Higher Load Category Consumers Use More Energy

Load Category 3 consumers show the highest average energy usage.
**Decision:** Introduce tier-based consumption monitoring for high load categories.

### Ernakulam Shows Highest Risk Concentration

Ernakulam records the highest abnormal usage, indicating district-level consumption irregularity.
**Decision:** Prioritize Ernakulam for focused inspections and continuous monitoring.
*Urban density, high appliance penetration, AC/heater usage, high-rise load systems, and seasonal occupancy variations likely contribute to this trend.*

# Recommendations

## Prioritize High-Risk Districts for Targeted Inspections

**Insight :** Emakulam shows the highest abnormal usage indicating concentrated risk.

**Recommendation :** Allocate inspection teams primarily to high-risk districts

**Business Impact :**

- Faster detection of irregular consumption
- Reduced operational revenue losses
- Optimized manpower utilization

**Feasibility:** High – District-level risk segmentation already available in the dashboard.

## Implement Continuous Deviation Based Monitoring

**Insight:** Actual energy consistently exceeds expected benchmarks across districts.

**Recommendation:** Adopt usage deviation (%) as a formal anomaly detection KPI and monitor it continuously instead of relying on periodic audits.

**Business Impact :**

- Early anomaly identification
- Reduced cumulative financial leakage
- Improved billing accuracy

**Feasibility:** High – Deviation metrics and abnormal flags already calculated.

## Deploy Seasonal Monitoring During High Temperature Periods

**Insight:** Deviation peaks when temperature exceeds 30°C (6.37%).

**Recommendation:** Introduce temperature-triggered alerts and peak-season monitoring controls.

**Business Impact:**

- Better demand management
- Reduced grid stress
- Improved forecasting accuracy

**Feasibility:** Medium–High – Temperature categorization already structured.

# Impact Estimation

## Cost Savings

With 4,739 abnormal cases identified in January, even if only 15% represent recoverable billing inefficiencies, and an average correction of ₹400 per case:

**Estimated annual recovery ≈ ₹30–35 lakhs (conservative estimate).**

Early detection prevents cumulative revenue leakage over time.

## Improved Operational Efficiency

Risk-based targeting allows authorities to focus on high-deviation districts and high-risk consumers instead of random inspections.

This can reduce inspection effort by **40–50%**, while increasing detection accuracy.

## Improved Service Quality

Deviation monitoring enables:

- Faster identification of billing inconsistencies
- Reduced consumer disputes
- More transparent consumption benchmarking

This improves consumer trust and service reliability.

## Risk Reduction

The framework reduces:

- Revenue leakage risk
- Peak-season grid stress (>30°C impact identified)
- High-load misuse
- Undetected abnormal consumption clusters

This shifts monitoring from reactive correction to proactive control.

Resource Links ( THESE LINKS BELOW ARE GENERALISED ARTICLES FROM WHICH INSIGHTS WERE CONSIDERED )

[Resource article 1](#) , [Resource article 2](#) , [Resource article 3](#)

# Limitations

## Data Issues

- The analysis is based only on **January data**, limiting seasonal comparison.

- The dataset includes residential consumers only; commercial and industrial patterns are excluded.

- Expected Energy values are assumed to be accurate baseline estimates.

- Data quality depends on meter recording accuracy and manual data entry reliability.

## Assumption Risks

- High Usage Deviation is assumed to indicate abnormal consumption, but deviation may also result from genuine lifestyle changes.

- Temperature and occupancy are assumed to directly influence consumption though other behavioral factors are not measured.

- Each record is treated as an independent observation, without accounting for historical consumer behavior trends.

## What Cannot Be Concluded

- The analysis does not confirm electricity theft , it only flags statistical anomalies.

- Long-term consumption trends cannot be established due to single-month data.

- Causal relationships cannot be definitively proven, only associations are observed.

- The framework identifies high-risk consumers but does not predict future anomalies.

# Future Scope

## Further Analysis

- Extend analysis to multi-month and seasonal data for stronger trend validation.

- Develop predictive models to forecast high-risk consumers.

- Build a composite risk score combining deviation %, load category, and appliance score.

- Conduct peak-hour consumption analysis for better demand management.

## Additional Data Required

- Historical consumption data (6–12 months)

- Smart meter real-time readings

- Billing and payment history

- Transformer/feeder-level load data

- Appliance-level consumption details

# Conclusion

This project developed a district-level anomaly detection framework for residential electricity consumption in Kerala. Through deviation analysis, clustering, and dashboard visualization, high-risk consumers and abnormal usage patterns were clearly identified.

The solution enables data-driven inspection prioritization, improved operational efficiency, reduced revenue leakage, and proactive monitoring.

Overall, the framework transforms electricity consumption data into actionable decision support for better energy governance and risk management.

# Section_D_G-3 Contribution matric

| Team Member | Dataset & Sourcing | Cleaning | KPI & Analysis | Dashboard | Report Writing | PPT | Overall Role |
|---|---|---|---|---|---|---|---|
| Tejaswini | ✔ | | ✔ | ✔ | ✔ | ✔ | Project Lead |
| Utkarsh Jain | ✔ | ✔ | | | | | Data Lead |
| Anwesha | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | Strategy Lead |
| Kashika | | | ✔ | | | | Dashboard Lead |
| Lokendra | | | ✔ | ✔ | | | Analysis Lead |
| Sankalp | ✔ | | | | ✔ | ✔ | PPT & Quality Lead |

**Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.**

**Team Signature Block:** SECTION_D_GROUP_3