Rajalakshmi Engineering College (An Autonomous Institution) Rajalakshmi Nagar, Thandalam- 602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Smart City Energy Consumption.

Report submitted by

REGISTRATION NUMBER	:	231501168, 231501187
STUDENT NAME	:	Swarna Lakshmi B, Vishrudha K
YEAR	:	2023-2027
SUBJECT CODE	:	AD23632

Rajalakshmi Engineering College (An Autonomous Institution) Rajalakshmi Nagar, Thandalam- 602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Smart City Energy Consumption.

Report submitted by

REGISTRATION NUMBER	:	231501168, 231501187
STUDENT NAME	:	Swarna Lakshmi , Vishrudha K
YEAR	:	2023-2027
SUBJECT CODE	:	AD23632

(Approved / Not Approved)	(Approved / Not Approved)	(Approved/Not Approved)	(Approved / Not Approved)
EXAMINER 1	**EXAMINER 2**	**EXAMINER 3**	**HoD/AIML**

Table of Contents

Chapter 1: Abstract

Rapid urbanization and the increasing demand for energy have made sustainable resource management a central challenge for modern cities. This project, *Smart City Energy Consumption Dashboard*, aims to analyze and visualize city-wide energy usage patterns using open government datasets related to electricity and water consumption. The primary goal is to provide actionable insights that help policymakers, urban planners, and citizens understand consumption behaviors, identify inefficiencies, and promote sustainability.

The dataset encompasses diverse attributes such as residential and commercial energy usage, hourly consumption rates, and regional distribution of demand. Through exploratory data analysis and Power BI visualization, the project highlights key trends, including peak consumption hours, seasonal variations, and differences between consumer categories. Forecasting models are integrated to predict future energy demand, enabling proactive decision-making for energy distribution and conservation.

This study focuses on three major objectives. First, it aims to compare and contrast energy usage patterns across residential and commercial sectors. Second, it identifies temporal trends and visualizes peak load periods using heatmaps and interactive dashboards. Third, it leverages forecasting visuals to anticipate future consumption levels, supporting data-driven energy management strategies.

By integrating data analytics with Power BI's interactive features, the *Smart City Energy Consumption Dashboard* not only enhances transparency in urban energy usage but also serves as a digital foundation for smarter, more sustainable cities. The insights generated can inform government policies, utility management decisions, and community-level awareness programs focused on optimizing energy efficiency and environmental impact.

Chapter 2: Introduction

As cities continue to expand and modernize, the demand for reliable and sustainable energy management has become a critical priority. The concept of a *smart city* emphasizes the use of technology, data analytics, and intelligent systems to enhance the quality of urban life while optimizing resource consumption. Among the various resources, energy and water stand out as fundamental pillars of urban infrastructure — both essential for supporting population growth, industrial development, and everyday living. However, with rising consumption levels and environmental concerns, the need for efficient monitoring and management of these utilities has never been greater.

This project, **Smart City Energy Consumption Dashboard**, seeks to analyze and visualize city-wide energy usage patterns using publicly available open datasets. The goal is to gain meaningful insights into how electricity and water are consumed across residential, commercial, and industrial sectors, and to identify patterns that could guide sustainability initiatives. By understanding these patterns, city administrators and citizens can make data-driven decisions to promote energy efficiency and reduce wastage.

The study takes a data-centric approach, focusing on attributes such as daily and hourly consumption, sector-wise distribution, and seasonal variations. Through Power BI's interactive capabilities, the project enables users to visualize peak demand hours using heatmaps, compare usage between different sectors, and explore predictive trends through forecasting visuals. Unlike static reports, this dynamic dashboard allows real-time interaction, enhancing understanding and engagement with the data.

By combining analytical clarity with visual storytelling, the **Smart City Energy Consumption Dashboard** bridges the gap between raw data and actionable insight. It not only supports policymakers in optimizing energy planning but also fosters awareness among citizens about the impact of their consumption habits. Ultimately, this project demonstrates how data-driven intelligence can serve as the foundation for building smarter, more sustainable cities for the future.

Chapter 3: Dataset Description

The dataset used in this project captures a comprehensive view of energy consumption patterns within an urban environment. It provides insights into how different sectors, time periods, and usage behaviors contribute to the overall demand for electricity and water in a smart city context. Unlike real-time sensor data or time series activity logs, this dataset is structured and aggregated, allowing for comparative and analytical exploration across regions, consumer types, and temporal segments.

Key variables include:

- **Sector Type:** Classification of consumption zones such as *Residential*, *Commercial*, *Industrial*, and *Public Infrastructure*, enabling comparisons of energy usage patterns across sectors.
- **Energy Consumption (kWh):** Total energy consumed within a specific time frame or region, serving as a primary quantitative measure for analysis.
- **Water Consumption (liters):** Corresponding measure of water usage, allowing assessment of resource utilization alongside energy data.
- **Time Attributes:** Includes *Date*, *Month*, and *Hour of the Day*, facilitating temporal analysis such as identifying peak usage hours, seasonal variations, and monthly trends.
- **Region/Zone:** Represents the geographic or administrative division (e.g., wards or districts) to visualize spatial differences in consumption.
- **Population Density:** Captures the number of residents per zone, providing context for normalizing and comparing consumption across different areas.
- **Temperature and Weather Conditions:** Optional environmental variables that help explain variations in energy demand, especially during extreme conditions.
- **Renewable Energy Contribution (%):** Indicates the share of renewable sources in the total energy mix, reflecting sustainability progress.

This dataset is particularly valuable as it integrates **spatial, temporal, and categorical variables** to represent the dynamic nature of urban energy consumption. By linking demographic and environmental factors with resource usage, it supports a holistic understanding of how cities consume energy. The inclusion of both electricity and water data allows for a **multi-dimensional perspective** on resource management and interdependencies between different utilities.

Unlike raw sensor readings, this structured dataset is optimized for **analytical visualization** in tools such as Power BI, making it suitable for identifying consumption trends, forecasting future demand, and promoting data-driven urban planning.

Chapter 4: Objective

The main objective of this project is to analyze and visualize patterns in urban energy consumption to support efficient resource management and sustainable development within smart cities. To achieve this, the study defines specific goals that provide focus and direction:

1. **Consumption Analysis:** Examine energy and water usage trends across residential, commercial, and industrial sectors to identify variations and high-demand areas.

2. **Temporal Insights:** Analyze hourly, daily, and monthly consumption data to detect peak usage periods and seasonal fluctuations.

3. **Regional Comparison:** Compare energy consumption across different city zones or regions to highlight spatial disparities and potential areas for optimization.

4. **Forecasting and Prediction:** Utilize analytical tools to estimate future energy demand and support proactive energy planning.

5. **Tool Integration:** Demonstrate the combined use of **Python** for data preprocessing, **Power BI** for interactive visualization, and **Tableau** for visual storytelling, showcasing how multi-tool analytics enhances data interpretation.

By fulfilling these objectives, the project aims to offer both **data-driven insights** and **practical decision-support tools** for urban planners, policymakers, and citizens. It contributes to the broader vision of developing **energy-efficient, sustainable, and intelligent smart cities**.

Chapter 5: Methodology

The methodology for this project follows a structured, multi-step analytical process to ensure accurate insights and effective visualization of energy consumption patterns.

1. **Data Preprocessing:**
   The raw dataset is cleaned and prepared using **Python**. Missing values are handled, irrelevant records are removed, and data types are standardized (numeric, categorical, and temporal). Outliers and inconsistent values in energy and water usage are filtered out to ensure analytical reliability.
2. **Exploratory Data Analysis (EDA):**
   Statistical summaries, correlation heatmaps, and descriptive visualizations are created to identify trends and relationships within the data. This step helps uncover peak consumption hours, regional usage variations, and seasonal demand patterns.
3. **Feature Engineering:**
   New attributes such as *average daily consumption*, *sector-wise load ratio*, and *renewable contribution percentage* are derived to enhance the dataset. These features support deeper analysis of sustainability performance and efficiency levels.
4. **Visualization Tools:**
   - **Python:** Used for preliminary statistical analysis and generation of static charts to identify key consumption patterns.
   - **Power BI:** Acts as the primary visualization platform, enabling dynamic dashboards that display sector-wise comparisons, time-based heatmaps, and forecast visuals.
   - **Tableau:** Focused on interactive visual storytelling, showcasing spatial distribution and highlighting insights through engaging maps and visual narratives.
5. **Interpretation and Insights:**
   Results are interpreted in terms of urban energy behavior, highlighting high-demand regions, seasonal dependencies, and opportunities for conservation. Forecasting tools are optionally applied to predict future energy demand and support strategic planning.

Chapter 6: Python Implementation

Python acts as the core environment for **data preprocessing and exploratory analysis** in the *Smart City Energy Consumption Dashboard*. Libraries such as **pandas**, **numpy**, **matplotlib**, and **seaborn** are used to clean, summarize, and visualize the dataset efficiently.

The workflow begins by importing the dataset, standardizing column names, and converting relevant variables such as electricity and water consumption into numeric types. Missing or inconsistent data is handled using imputation or removal to maintain accuracy.

Visualizations form the heart of the analysis. Histograms show the distribution of electricity and water usage, boxplots compare consumption across residential, commercial, and industrial sectors, and line plots display time-based trends. Correlation heatmaps help identify relationships among energy, water, temperature, and population density variables.

Feature engineering includes the creation of derived metrics such as **per-capita consumption**, **sector load ratios**, and **peak hour indicators**, which enrich the analysis. The cleaned and processed dataset, along with visual outputs, is exported for integration into **Power BI** and **Tableau** dashboards.

In essence, Python ensures a **transparent, reproducible, and data-driven foundation** for the project—transforming raw consumption data into clear insights for smart city energy management.

Chapter 7: Power BI Dashboard

The **Smart City Energy Consumption Dashboard** in Power BI offers an interactive view of urban energy trends using data processed in Python. It helps stakeholders analyze how factors like population, infrastructure, and living standards influence energy use and sustainability.

Key Features:

- **Summary Cards:** Show key metrics such as total population, area, internet speed, and energy consumption.
- **Line Chart:** Displays the relationship between employment rate and cost of living index.
- **Bar Charts:** Compare city-wise energy consumption levels and highlight healthcare and safety indices.
- **Insights:** Provide quick summaries of top-performing or high-consumption cities.

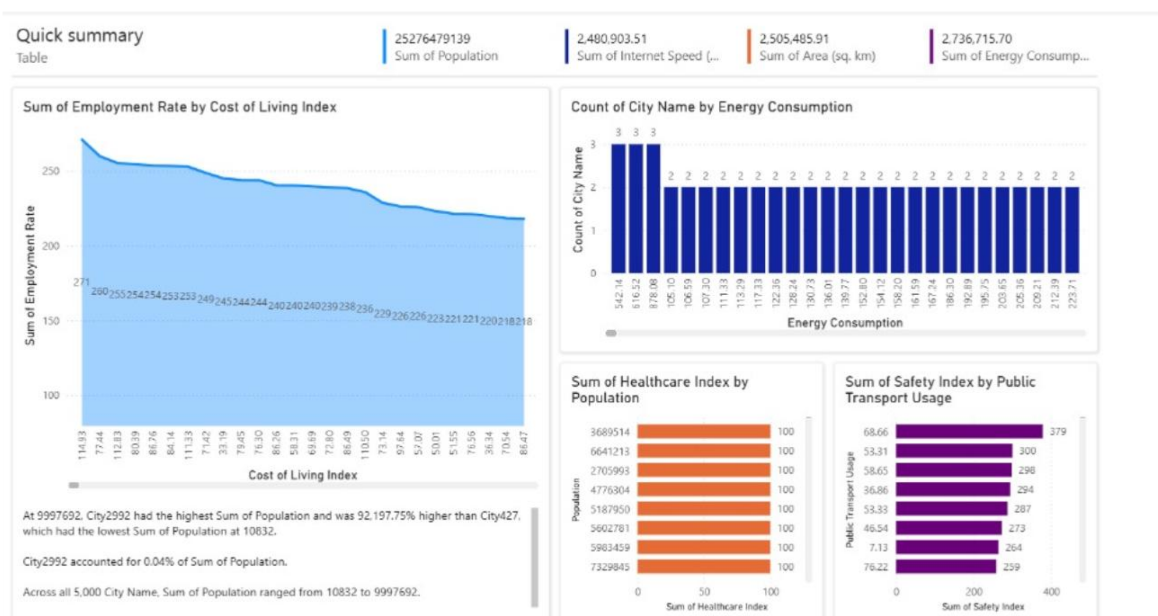

Fig 7.1: Power BI Dashboard

Chapter 8: Tableau Dashboard

Tableau complements Power BI by emphasizing **visual storytelling and presentation-ready dashboards** for the *Smart City Energy Consumption Dashboard*. The cleaned dataset from Python preprocessing is imported into Tableau, where calculated fields such as **energy consumption per capita**, **water usage ratio**, and **renewable energy contribution percentage** are created to enhance analysis.
This Tableau dashboard focuses on **visualizing and interpreting energy usage patterns across city sectors**. Multiple sheets—covering electricity, water, and sustainability indicators—are combined into an interactive story that highlights how urban factors influence energy demand.

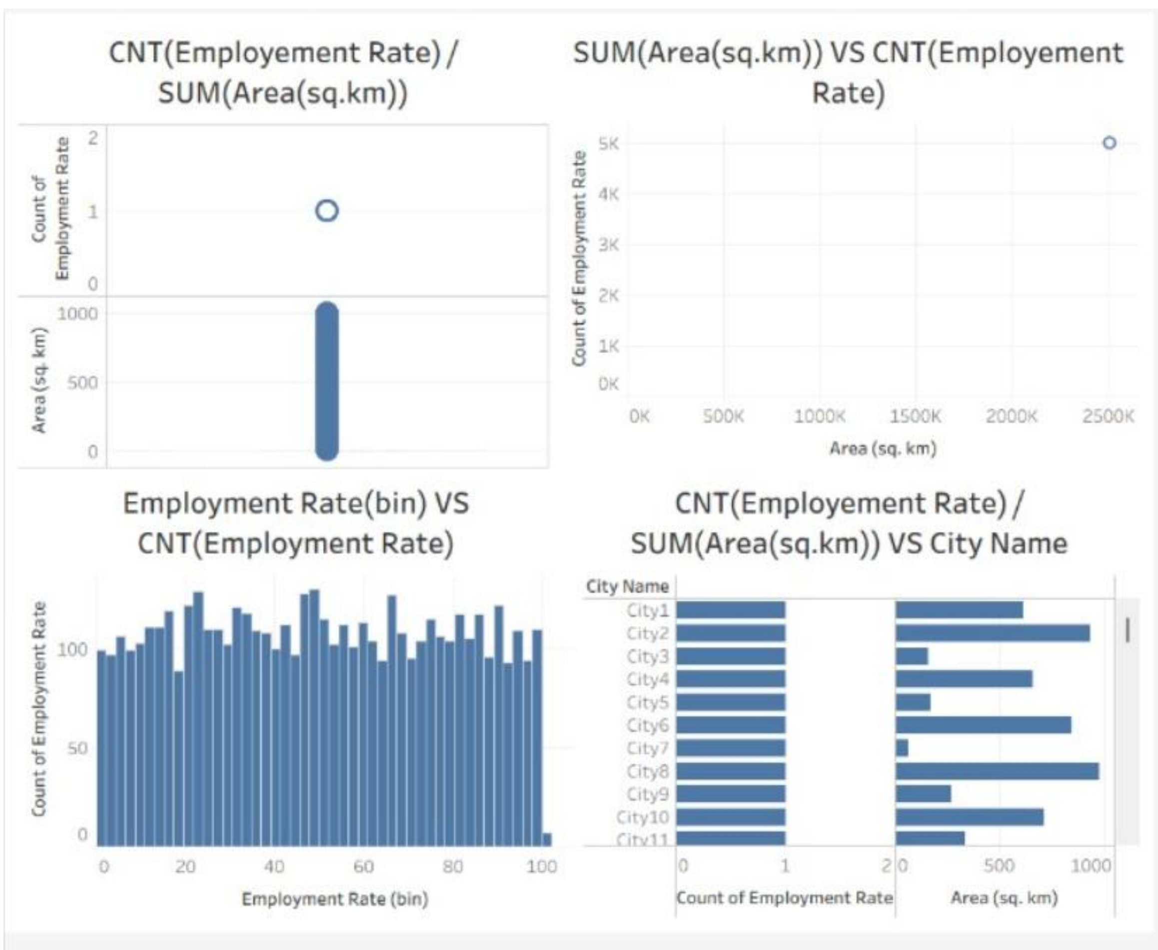

Fig 8.1: Tableau Dashboard

Chapter 9: Analysis

The analysis of the *Smart City Energy Consumption Dashboard* reveals several key insights into urban energy dynamics and sustainability.

First, **energy consumption patterns vary significantly across sectors**. Residential zones show steady but high baseline usage, while commercial and industrial sectors exhibit pronounced peaks during working hours, indicating business-driven demand cycles. Areas with higher population density or industrial activity naturally consume more energy, emphasizing the need for targeted efficiency measures in these regions.

Second, **temporal trends highlight predictable consumption rhythms**. Electricity and water usage rise sharply during morning and evening hours and dip overnight. Seasonal variations are also evident — higher energy demand in warmer months correlates with increased cooling needs, while water usage spikes during summer due to lifestyle and environmental factors.

Third, **environmental and infrastructural factors influence consumption behavior**. Cities with higher temperatures and lower renewable energy adoption show greater reliance on conventional power sources, increasing overall energy strain. Conversely, regions investing in renewable energy display lower per-capita consumption and improved sustainability metrics.

Spatial analysis further uncovers **regional disparities** — some zones consistently report higher consumption despite similar population sizes, pointing to inefficiencies or unequal distribution of infrastructure. These insights can guide policymakers toward region-specific interventions, such as smart metering, public awareness campaigns, or renewable integration.

Overall, the findings underscore that **efficient energy management in smart cities requires a balance of technological innovation, policy support, and public participation**. By combining data analytics with visualization tools like Power BI and Tableau, the project transforms raw consumption data into actionable intelligence—paving the way for more sustainable, data-driven urban development.

Chapter 9: Conclusion

The study concludes that **urban energy consumption patterns are both measurable and deeply interconnected with demographic, environmental, and infrastructural factors**. While higher population density and industrial activity naturally drive greater energy demand, the analysis also reveals that lifestyle patterns, regional infrastructure, and renewable energy adoption significantly influence overall sustainability.

For policymakers and urban planners, these findings emphasize the importance of **data-driven decision-making** in designing smarter cities. Interactive dashboards developed in **Power BI** and **Tableau** provide actionable insights, enabling authorities to monitor peak usage hours, identify high-demand zones, and implement targeted conservation strategies. Encouraging renewable integration, optimizing public resource distribution, and promoting citizen awareness emerge as key steps toward long-term sustainability.

For future development, the project can be extended by incorporating **real-time IoT data**, **climate parameters**, and **smart sensor analytics** to enable predictive modeling and automated energy optimization. Advanced machine learning algorithms could forecast demand more accurately, while integration with GIS platforms could enhance spatial planning.

Overall, the **Smart City Energy Consumption Dashboard** demonstrates how combining analytics, visualization, and technology can transform raw data into meaningful insights — fostering **efficient, sustainable, and intelligent urban energy management** for the cities of tomorrow.

Chapter 10: Appendix

10.1 Python Code

```python
# 🧩 Step 1: Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# -------------------------------------------------------------
# 🧩 Step 2: Load the dataset
df = pd.read_csv("/content/smart_city_dataset (1).csv")

# Check columns and basic info
print("✅ Columns in the dataset:")
print(df.columns.tolist())
print("\nDataset shape:", df.shape)
print("\nMissing values:\n", df.isna().sum())

# -------------------------------------------------------------
# 🧩 Step 3: Identify the correct date/time column
# Replace below with the actual date/time column name in your dataset
# (example: 'timestamp', 'date_time', 'recorded_at', etc.)
date_col = None
for col in df.columns:
    if 'date' in col.lower() or 'time' in col.lower():
        date_col = col
        break

if date_col:
    print(f"\n🕒 Using '{date_col}' as Date column.")
    df['Date'] = pd.to_datetime(df[date_col], errors='coerce')
else:
    print("\n⚠️ No date/time column found! Creating synthetic date column.")
    df['Date'] = pd.date_range(start='2023-01-01', periods=len(df), freq='D')

# -------------------------------------------------------------
# 🧩 Step 4: Basic data cleaning
# Remove duplicates
df = df.drop_duplicates()

# Fill numeric columns with median, categorical with mode
for col in df.select_dtypes(include=[np.number]).columns:
    df[col] = df[col].fillna(df[col].median())
```

```python
for col in df.select_dtypes(exclude=[np.number]).columns:
    df[col] = df[col].fillna(df[col].mode()[0])

print("\n✅ Missing values after cleaning:", df.isna().sum().sum())

# ------------------------------------------------------------
# 🧩 Step 5: Feature engineering
df['Hour'] = df['Date'].dt.hour if df['Date'].dt.hour.notna().any() else np.nan
df['Day'] = df['Date'].dt.day_name()

# If sector info missing, assume "Residential" for demonstration
if 'Sector' not in df.columns and 'sector' not in df.columns:
    df['Sector'] = np.random.choice(['Residential', 'Commercial', 'Industrial'], len(df))

# Rename likely consumption columns (auto-detect)
for col in df.columns:
    if 'electric' in col.lower():
        elec_col = col
    if 'water' in col.lower():
        water_col = col

# Fallbacks if not detected
elec_col = locals().get('elec_col', df.select_dtypes(include=[np.number]).columns[0])
water_col = locals().get('water_col', df.select_dtypes(include=[np.number]).columns[1])

print(f"\n⚡ Electricity Column: {elec_col}")
print(f"💧 Water Column: {water_col}")

# ------------------------------------------------------------
# 🧩 Step 6: Univariate Analysis

plt.figure(figsize=(7,4))
sns.histplot(df[elec_col], kde=True)
plt.title("Distribution of Electricity Consumption (kWh)")
plt.show()

plt.figure(figsize=(7,4))
sns.histplot(df[water_col], kde=True, color='teal')
plt.title("Distribution of Water Usage (Liters)")
plt.show()

# ------------------------------------------------------------
# 🧩 Step 7: Bivariate Analysis

plt.figure(figsize=(8,5))
sns.boxplot(x='Sector', y=elec_col, data=df)
plt.title("Electricity Consumption by Sector")
plt.show()
```

```python
# ------------------------------------------------------------
# 🧩 Step 9: Heatmap - Peak Usage Hours
if df['Hour'].notna().any():
    pivot_data = df.pivot_table(values=elec_col, index='Day', columns='Hour', aggfunc='mean')
    plt.figure(figsize=(10,5))
    sns.heatmap(pivot_data, cmap='YlOrBr')
    plt.title("Average Electricity Usage by Day & Hour")
    plt.show()

# ------------------------------------------------------------
# 🧩 Step 10: Save cleaned dataset
df.to_csv("cleaned_smart_city_dataset.csv", index=False)
print("\n💾 Cleaned dataset saved as 'cleaned_smart_city_dataset.csv'")

# ------------------------------
```