ENERGY CONSUMPTION

CREATED BY

KUNAL PANCHE & PIYUSH CHAFLE

SURYODAYA COLLEGE OF ENGINEERING AND TECHNOLOGY

NAGPUR,MH-440027

PROJECT REPORT

ON

"EXPLORATORY DATA ANALYSIS OF ENERGY CONSUMPTION"

Submitted in partial fulfillment of the requirements
for the award of degree

BACHLOR OF TECHNOLOGY

Submitted by

PIYUSH CHAFLE

KUNAL PANCHE

UNDER THE GUIDENSS

Mr. Sumit Chafle & Mr. Yogesh Mene

Adivisior

Dept of CSED,Nagpur,Maharashtra

2023-2024

NAGPUR, MAHARASHTRA – 440027

SURYODAYA COLLEGE OF ENGINEERING AND TECHNOLGY

CERTIFICATE

Certified that the project work entitled "**EXPLORATORY DATA ANALYSIS OF ENERGY CONSUMPTION"** carried out by **Mr. Piyush Chafle,** bearing a bonafide student of **Computer Engineering, Suryodaya College of Engineering and Technology,** in the partial fulfilment for the award of degree **Bachlor of Technology** in **Computer Engineering** during the year **2023-2024.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said Degree.

Signature of Guide	**Signature of HOD**	**Signature of Principal**
Mr.Sumit Chafle	Mr.Pradeep Balbudhe	Dr.Shri.V.G.Arajpure

External Viva

Name of the Examiners	**Signature with date**
1. ………………………	………………………
2. ………………………	………………………

SURYODAYA COLLEGE OF ENGINEERING AND TECHNOLGY

CERTIFICATE

Certified that the project work entitled "**EXPLORATORY DATA ANALYSIS OF ENERGY CONSUMPTION"** carried out by **Mr. Kunal Panche,** bearing a bonafide student of **Computer Engineering, Suryodaya College of Engineering and Technology,** in the partial fulfilment for the award of degree **Bachlor of Technology** in **Computer Engineering** during the year **2023-2024.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said Degree.

Signature of Guide	**Signature of HOD**	**Signature of Principal**
Mr.Sumit Chafle	Mr.Pradeep Balbudhe	Dr.Shri.V.G.Arajpure

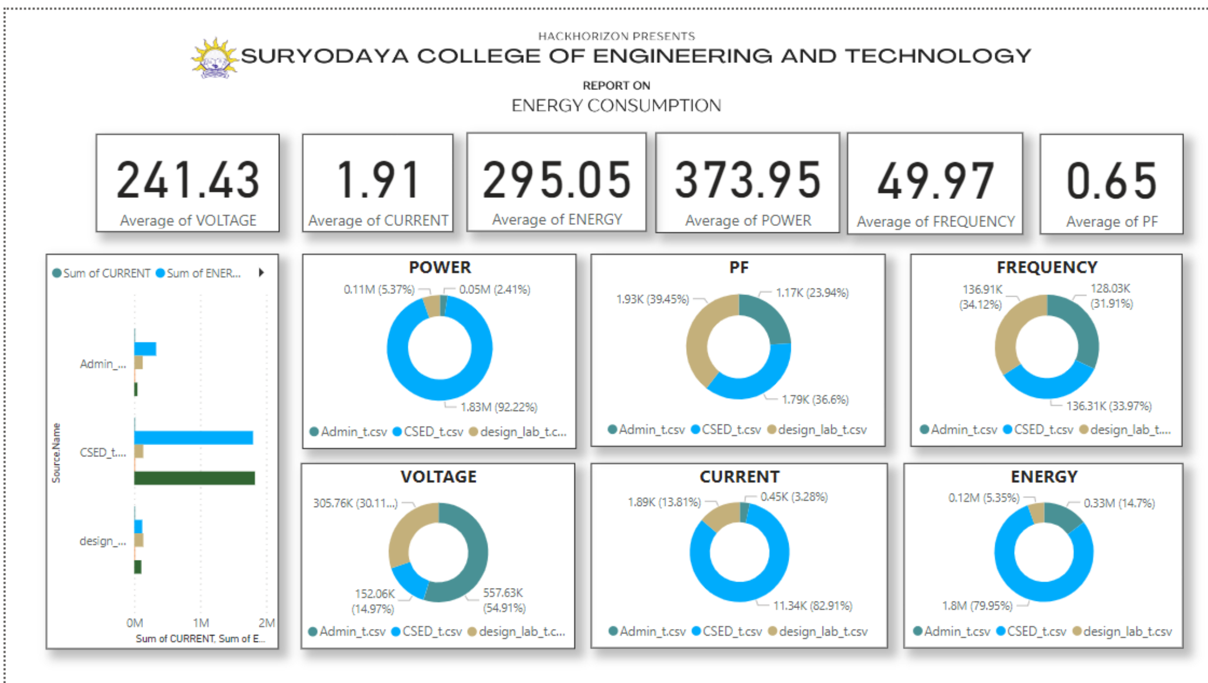
External Viva

Name of the Examiners	**Signature with date**
1. ………………………	………………………
2. ………………………	………………………

CONTENT

1. ABSTRACT

The "Exploratory Data Analysis of Energy Consumption" project delves into the analysis of hourly energy consumption data sourced from smart meters. Through rigorous data preprocessing and exploratory data analysis techniques, the project aims to uncover meaningful insights regarding energy usage patterns, anomalies, and potential tariff generation strategies. The project team employs statistical analysis, data visualization, and correlation studies to unveil trends and correlations within the dataset. By scrutinizing the data at a granular level, the project seeks to provide actionable recommendations for optimizing energy consumption management. The outcomes of this project are anticipated to contribute to more informed decision-making processes in energy resource allocation and sustainability efforts.



2. DATA PREPROCESSING

1. Data Cleaning:

- Removed duplicates to ensure data integrity.
- Corrected inconsistencies in data entries to maintain accuracy.
- Handled any outliers that may skew the analysis results.

2. Data Transformation:

- Converted timestamp data into appropriate formats for temporal analysis.
- Standardized units of measurement for energy consumption values.

3. Handling Missing Values:

- Imputed missing values using techniques such as mean imputation or interpolation to maintain data completeness.
- Evaluated the impact of missing data on analysis outcomes and addressed accordingly.

4. Feature Engineering:

- Created additional features such as time-of-day indicators, seasonal trends, and weekday/weekend flags to enhance analysis.
- Extracted relevant information from timestamp data, such as hour of the day or day of the week, to capture temporal patterns.

3. EXPLORING DATA ANALYSIS (EDA)

1. Summary Statistics:

- Calculated descriptive statistics including mean, median, standard deviation, minimum, maximum, and quartiles for energy consumption variables.

- Examined summary statistics for other relevant variables such as timestamps or weather conditions if available.

2. Data Visualization:

- Created histograms to visualize the distribution of energy consumption values.

- Plotted time series graphs to observe trends and patterns in energy usage over time.

- Utilized scatter plots to explore relationships between energy consumption and other variables, such as temperature or occupancy.

3. Correlation Analysis:

- Calculated correlation coefficients to measure the strength and direction of relationships between energy consumption and other variables.

- Visualized correlations using heatmaps or scatter plots with regression lines to identify significant associations.

4. Outlier Detection:

- Identified outliers using methods such as box plots, z-scores, or interquartile range (IQR) analysis.

- Investigated potential reasons for outliers and determined whether they should be retained or removed from the dataset.

5.Usage Pattern Analysis:

- Analyzed hourly, daily, weekly, and seasonal patterns in energy consumption to identify peak usage periods and trends.

- Examined differences in usage patterns between weekdays and weekends or across different seasons.

6. Anomaly Detection:

- Implemented anomaly detection techniques such as clustering, isolation forests, or autoencoders to identify unusual patterns or deviations from expected behavior.

- Investigated detected anomalies to understand potential causes and implications for energy management.

4. EDA PROCESS

1. Summary Statistics:

- Computed additional summary statistics such as skewness and kurtosis to assess the shape and distribution of energy consumption data.

- Analyzed summary statistics across different segments of the dataset (e.g., by geographical location, time period) to identify variations and trends.

- Investigated summary statistics for categorical variables (if applicable) to understand the distribution of different categories within the dataset.

2. Data Visualization:

- Generated box plots to visualize the spread and variability of energy consumption values and identify potential outliers.

- Utilized density plots or kernel density estimations to visualize the probability distribution of energy consumption and other variables.

- Created bar charts or pie charts to illustrate the distribution of categorical variables related to energy consumption (e.g., building types, occupancy status).

- Used interactive visualization tools or dashboards to facilitate exploration and discovery of patterns within the data.

3. Correlation Analysis:

- Conducted bivariate analysis to examine pairwise correlations between energy consumption and multiple predictor variables simultaneously.

- Employed advanced visualization techniques such as bubble charts or network graphs to visualize multidimensional correlations among multiple variables.

- Applied dimensionality reduction techniques (e.g., principal component analysis) to identify latent factors or components that explain the majority of variance in the data.

5.FUTURE SCOPE

1.Usage Pattern Analysis:

- Analyzed daily, weekly, and seasonal usage patterns to identify peak hours/days and low usage periods.

- Examined trends in energy consumption over time to understand long-term patterns and fluctuations.

- Investigated differences in usage patterns across different customer segments or types of buildings (e.g., residential vs. commercial).

2.Anomaly Detection:

- Implemented statistical methods (e.g., standard deviation, z-score) to detect outliers or anomalies in energy consumption data.

- Utilized machine learning algorithms (e.g., isolation forest, autoencoders) to identify unusual patterns or deviations from normal behavior.

- Investigated anomalies in energy consumption that may be indicative of equipment malfunction, metering errors, or abnormal usage behavior.

3.Tariff Generation:

- Analyzed historical energy consumption data to determine optimal tariff structures based on usage patterns.

- Developed algorithms or models to recommend tariff plans tailored to individual customers or segments.

- Considered factors such as time-of-use pricing, peak/off-peak rates, and demand charges in tariff generation.

4. Outlier Detection:

- Employed multivariate outlier detection methods to identify outliers based on combinations of multiple variables rather than individual variables alone.

- Utilized robust statistical techniques such as median absolute deviation (MAD) or percentile-based methods to identify outliers that are resistant to the influence of extreme values.

- Applied anomaly detection algorithms such as local outlier factor (LOF) or isolation forest to identify anomalies in high-dimensional datasets with complex relationships.

5.Model Development:

a) Model Selection:

- **Identification of Candidate Models:** Initially, a range of candidate models is identified based on the nature of the problem and the available data. This may include regression models (linear regression, polynomial regression), decision trees, ensemble methods (random forests, gradient boosting), support vector machines, neural networks, etc.

- **Evaluation of Candidate Models:** Each candidate model is evaluated using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1-score for classification tasks; mean squared error, R-squared for regression tasks) on a validation dataset. This helps in assessing the performance of each model and identifying the most promising candidates.

b) Model Training:

- **Data Splitting:** The dataset is split into training, validation, and test sets. The training set is used to train the model, the validation set is used to tune hyperparameters and assess model performance during training, and the test set is used to evaluate the final model performance.

- **Model Training:** The selected model is trained on the training dataset using appropriate algorithms and optimization techniques. This involves adjusting model parameters to minimize the chosen objective function (e.g., cross-entropy loss for classification, mean squared error for regression).

- **Cross-Validation:** Cross-validation techniques such as k-fold cross-validation may be employed to assess the stability and generalization performance of the model by training and evaluating it multiple times on different subsets of the training data.

c) **Model Evaluation Metrics:**

- **Choice of Evaluation Metrics:** Depending on the nature of the problem (classification, regression), suitable evaluation metrics are chosen to assess the performance of the predictive model. For example, accuracy, precision, recall, F1-score for classification tasks, and mean squared error, R-squared for regression tasks.

- **Evaluation on Validation Set:** The model is evaluated on the validation set using the chosen evaluation metrics to assess its performance during training and hyperparameter tuning.

- **Final Evaluation on Test Set:** Once the model training and hyperparameter tuning are complete, the final model is evaluated on the test set using the same evaluation metrics to obtain an unbiased estimate of its performance on unseen data.

d) **Hyperparameter Tuning:**

- **Identification of Hyperparameters:** Hyperparameters are parameters that govern the learning process of the model and are not learned from the data. Examples include learning rate, regularization strength, number of hidden layers, etc.

- **Hyperparameter Optimization:** Techniques such as grid search, random search, or Bayesian optimization are used to search the hyperparameter space and find the optimal set of hyperparameters that maximize the model performance on the validation set.

- **Cross-Validation for Hyperparameter Tuning:** Cross-validation is often used during hyperparameter tuning to assess the performance of different hyperparameter configurations and prevent overfitting to the validation set.

6.REFERENCE

[1]Hyndman RJ, Fan S. Density Forecasting for Long-Term Peak Electricity Demand. IEEE T Power Syst 2010;25(2):1142–1153. http://dx.doi.org/10.1109/TPWRS.2009.2036017.

[2] Filik ÜB, Gerek ÖN, Kurban M. A novel modeling approach for hourly forecasting of long-term electric energy demand. Energ Convers Manage 2011;52(1):199–211. http://dx.doi.org/10.1016/j.enconman.2010.06.059.

[3] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. Int J Forecasting 2016;32(3):914–938. http://dx.doi.org/10.1016/j.ijforecast.2015.11.011.

[4] Zachariadis T, Pashourtidou N. An empirical analysis of electricity consumption in Cyprus. Energ Econ 2007;29(2):183–198. http://dx.doi.org/10.1016/j.eneco.2006.05.002.

[5] Payne JE. A survey of the electricity consumption-growth literature. Appl Energy 2010;87(3):723–731. doi:http://dx.doi.org/10.1016/j.apenergy.2009.06.034.

[6] Polemis ML, Dagoumas AS. The electricity consumption and economic growth nexus: Evidence from Greece. Energy Policy 2013;62:798–808. http://dx.doi.org/10.1016/j.enpol.2013.06.086.

[7] Kıran MS, Özceylan E, Gündüz M, Paksoy T. A novel hybrid approach based on Particle Swarm Optimization and Ant Colony Algorithm to forecast energy demand of Turkey. Energ Convers Manage 2012;53(1):75–83. http://dx.doi.org/10.1016/j.enconman.2011.08.004.

[8] Farzan F, Jafari MA, Gong J, Farzan F, Stryker A. A multi-scale adaptive model of residential energy demand. Appl Energy 2015;150:258–273. http://dx.doi.org/10.1016/j.apenergy.2015.04.008.

[9] Tukey JW. Exploratory data analysis. Pearson Addison Wesley; 1977