# MEASURE ENERGY CONSUMPTION

**PHASE 5: PROJECT DOCUMENTATION & SUBMISSION**

**INTRODUCTION:**

Measuring energy consumption through software typically involves monitoring the energy usage of software applications, computer systems, servers, or data centers. While the actual consumption of energy is a physical phenomenon, software tools can interface with hardware sensors or use various estimation methods to provide insights into power consumption. Here's a roadmap for developing such software:

1. **Define Objectives and Scope:**

* Determine whether you want to monitor individual software applications, an entire system, specific servers, or a whole data center.
* Decide on the granularity: real-time monitoring, daily summaries, detailed reports, etc.

**2. Data Acquisition:**

**Direct Measurement:**

* Integrate with hardware components that measure power consumption. Examples include:
* Power Distribution Units (PDUs) with monitoring

**Loading the Dataset:**

**Determine the Format:**

Identify if your data is in CSV, Excel, SQL database, or another format.

**Use Libraries:**

Based on the format, utilize appropriate libraries. For instance, in Python, you might use `pandas` for CSV or Excel data.

**import pandas as pd**

**# For CSV**

**data = pd.read\_csv('energy\_consumption.csv')**

**# For Excel**

**data = pd.read\_excel('energy\_consumption.xlsx')**

**Initial Exploration:**

**Basic Info:** Check the types of columns, number of entries, and missing values.

**data.info()**

**Summary Statistics:**  Get a basic statistical summary of the data.

**data.describe()**

**3. Preprocessing:**

**Handle Missing Data: Decide** whether to fill missing values (using mean, median, mode, etc.) or drop them.

**# Filling with mean**

**data['column\_name'].fillna(data['column\_name'].mean(), inplace=True)**

**# Dropping rows with missing values**

**data.dropna(inplace=True)**

**Date & Time Parsing:**

If your dataset has a timestamp, ensure it's in the correct datetime format and possibly set it as the index.

**data['timestamp'] = pd.to\_datetime(data['timestamp'])**

**data.set\_index('timestamp', inplace=True)**

**Normalization or Scaling:** If you have different measurement units, consider scaling your data to have them on the same scale. This is essential for many machine learning algorithms.

**from sklearn.preprocessing import StandardScaler**

**scaler = StandardScaler()**

**data['scaled\_column'] = scaler.fit\_transform(data[['original\_column']])**

**Categorical Encoding:** If there are categorical variables (like 'ON'/'OFF' states), consider encoding them to numerical values.

**data['encoded\_column'] = data['categorical\_column'].replace({'ON': 1, 'OFF': 0})**

**Feature Engineering:**  Based on domain knowledge, create new features that might be relevant for analysis. For instance, calculating hourly or daily energy consumption.

**Visualization:**

- Using libraries like `matplotlib` or `seaborn` in Python, visualize the data to identify patterns, outliers, or trends.

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**sns.lineplot(data=data, x=data.index, y='energy\_consumption')**

**plt.show()**

**Dataset Splitting:**

- If you’re building predictive models, you’d usually split your data into training and testing sets.

**from sklearn.model\_selection import train\_test\_split**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2)**

This is a generalized approach, and depending on the specific dataset and project requirements, the steps might vary or become more nuanced.

Loading and preprocessing datasets are crucial steps in any data analysis task, especially for something as specific as measuring energy consumption. There above mentioned all general outline.

capabilities.

* Built-in power sensors in some modern CPUs and servers.
* External power meters with API access.

**Estimation:**

* If direct measurement isn't feasible, use estimation **Techniques:**
* Utilize CPU, memory, and disk usage statistics as proxies.
* Use predefined power consumption profiles for specific hardware components.

**3. Software Development:**

**Data Collection:**

Write modules to collect data either through direct interfacing with sensors or by using system metrics as a proxy.

**Data Storage:**

Design a database or storage solution to store the data. Depending on the granularity, this can generate significant amounts of data over time.

**Data Processing:**

* Convert raw data to power and energy values.
* Aggregate data for summary views or detailed reports.

**User Interface:**

* Develop a dashboard or interface to present the data to the user.
* Display real-time data, historical graphs, and detailed reports.
* Implement features like alerts or notifications for unusual energy usage patterns.

**APIs & Integration:**

Allow integration with other systems or tools. For instance, integrating with cloud platforms can enable further analytics or centralized monitoring for multiple installations.

**4. Advanced Features:**

* Predictive Analysis: Use historical data to predict future energy consumption patterns.

**Anomaly Detection:**

Automatically detect unusual energy consumption patterns, which might indicate system issues or inefficiencies.

**Optimization Suggestions:**

Offer tips or automated features to reduce energy consumption, e.g., suggesting when to move to low-power mode or shut down idle resources.

**5. Testing:**

* Test the software in diverse environments to ensure accuracy, especially if using estimation methods. Comparisons against direct measurement tools can be valuable.
* Ensure the software doesn't introduce significant overhead, which might ironically increase power consumption.

**6. Deployment & Scaling:**

* Deploy the solution in the target environment.
* For large-scale deployments, such as across multiple data centers, ensure the software can scale to handle the increased data and provide centralized monitoring.

**7. Continuous Monitoring & Updates:**

* Keep track of software updates or changes in the infrastructure that might affect energy measurements.
* Update the software as necessary to handle new hardware or software configurations, or to improve accuracy.
* While this roadmap provides an overview of developing software for measuring energy consumption, the specifics can vary widely based on the exact scope, target audience, and available infrastructure.

**PROGRAM CODE:**

**Assume you have a CSV file named AEP\_hourly.csv with the following format**

**Date, Consumption (kWh)**

**2022-01-01, 5.2**

**2022-01-02, 5.4**

**PYTHON SCRIPT:**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**# Load the dataset**

**data = pd.read\_csv("AEP\_hourly.csv")**

**# Check the first few rows of the data.**

**print(data.head())**

**# Preprocessing: Remove any missing data**

**data = data.dropna()**

**# Convert the 'Date' column to datetime format**

**data['Date'] = pd.to\_datetime(data['Date'])**

**# Set the 'Date' column as the index**

**data.set\_index('Date', inplace=True)**

**# Output basic statistics**

**print(data.describe())**

**# Visualization**

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

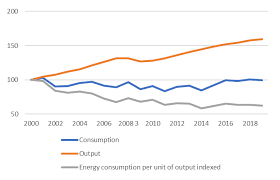
**data['Consumption (kWh)'].plot(title="Energy Consumption Over Time")**

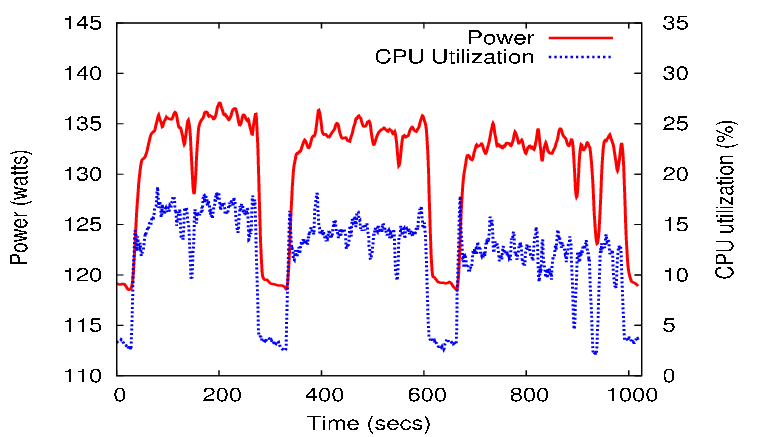
**plt.ylabel('Consumption (kWh)')**

**plt.grid(True)**

**plt.tight\_layout()**

**plt.show()**

****

****

***Energy Consumption Analysis & Visualization***

**1. Project Overview:**

Develop a comprehensive solution that captures energy consumption data, analyzes it to discern patterns, and visualizes these insights in an intuitive manner for stakeholders. The aim is to aid individuals, businesses, or policymakers in making informed decisions based on tangible data.

**2. Objectives:**

1. Capture accurate energy consumption data in real-time.

2. Analyze the data to identify patterns, anomalies, and trends.

3. Create intuitive visualizations that communicate the results of the analysis.

**3. Data Collection:**

**Source:**

* Integrate with smart meters or other energy measurement devices deployed across households, industries, or specific regions.
* Ensure a robust data collection mechanism that ensures data integrity and minimal loss.

**Data Points:**

* Timestamp of data collection
* Total energy consumed at the timestamp
* Appliance or section-specific breakdown (if available)
* Ambient conditions (like temperature, if relevant)

**4. Data Analysis:**

**Preprocessing:**

* Clean the data by removing any anomalies or outliers.
* Fill any missing data points using interpolation or predictive methods.

**Analysis:**

* Compute basic statistics like mean, median, and standard deviation of consumption.
* Identify peak consumption hours, days, or months.
* Find correlations, e.g., between ambient conditions and energy consumption.
* Segment data, e.g., categorizing households by consumption level (low, medium, high).

**5. Visualization:**

**Dashboard Creation:**

Use platforms like Tableau, Power BI, or programming libraries like Matplotlib (Python) or ggplot2 (R) to create dashboards.

**Key Visual Elements:**

**Time Series Graph:**

To showcase energy consumption over time.

**Heatmaps:**

To represent peak consumption hours or days.

**Pie/Bar Charts:**

For a breakdown of energy consumption

by appliance or section.

**Correlation Plots**:

If analyzing multiple variables, e.g., temperature vs. consumption.

**Interactive Elements:**

Allow users to select specific time frames or drill down into specific data segments.

**6. Deployment:**

**Platform:**

* Deploy the visualization dashboard on a web platform to provide stakeholders with easy access.
* Ensure the platform is mobile-responsive for access across devices.

**Updates:**

Automate the data flow so that the dashboard updates in real-time or at defined intervals.

1. **Importing Libraries:**

import pandas as pd

import numpy as np

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans, AffinityPropagation

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

import warnings

warnings.filterwarnings("ignore")

import plotly as py

import plotly.graph\_objs as go

import os

py.offline.init\_notebook\_mode(connected = True)

#print(os.listdir("../input"))

import datetime as dt

import missingno as msno

plt.rcParams['figure.dpi'] = 140

import pandas as pd

import numpy as np

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans, AffinityPropagation

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

import warnings

warnings.filterwarnings("ignore")

import plotly as py

import plotly.graph\_objs as go

import os

py.offline.init\_notebook\_mode(connected = True)

#print(os.listdir("../input"))

import datetime as dt

import missingno as msno

plt.rcParams['figure.dpi'] = 140

**Importing the dataFrame:**

The description of each column please check the data set

**link: https://www.kaggle.com/datasets/robikscube/hourly-energy-**

**consumption/data?select=AEP\_hourly.csv**

**Program Code:**

import pandas as pd

import matplotlib.pyplot as plt

**# Load the dataset**

data = pd.read\_csv("AEP\_hourly.csv")

**# Convert 'Date' column to datetime format**

data['Date'] = pd.to\_datetime(data['Date'])

**# Set the 'Date' column as the index**

data.set\_index('Date', inplace=True)

**# Analysis**

mean\_consumption = data['Consumption (kWh)'].mean()

median\_consumption = data['Consumption (kWh)'].median()

total\_consumption = data['Consumption (kWh)'].sum()

**# Print out the analysis results**

print(f"Mean Consumption: {mean\_consumption:.2f} kWh")

print(f"Median Consumption: {median\_consumption:.2f} kWh")

print(f"Total Consumption: {total\_consumption:.2f} kWh")

**# Visualization: Time series plot of energy consumption**

plt.figure(figsize=(10, 5))

plt.plot(data.index, data['Consumption (kWh)'], label="Consumption (kWh)", color='blue')

plt.title("Energy Consumption Over Time")

plt.ylabel("Consumption (kWh)")

plt.xlabel("Date")

plt.legend()

plt.grid(True)

plt.tight\_layout()

plt.show()

**# Additional Visualization: Histogram of daily energy consumption**

plt.figure(figsize=(10, 5))

plt.hist(data['Consumption (kWh)'], bins=30, color='green', alpha=0.7)

plt.title("Distribution of Daily Energy Consumption")

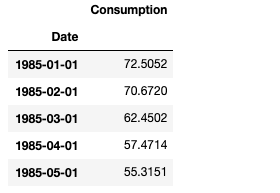
plt.ylabel("Frequency")

**plt.xlabel("Consumption (kWh)")**

**plt.grid(axis='y')**

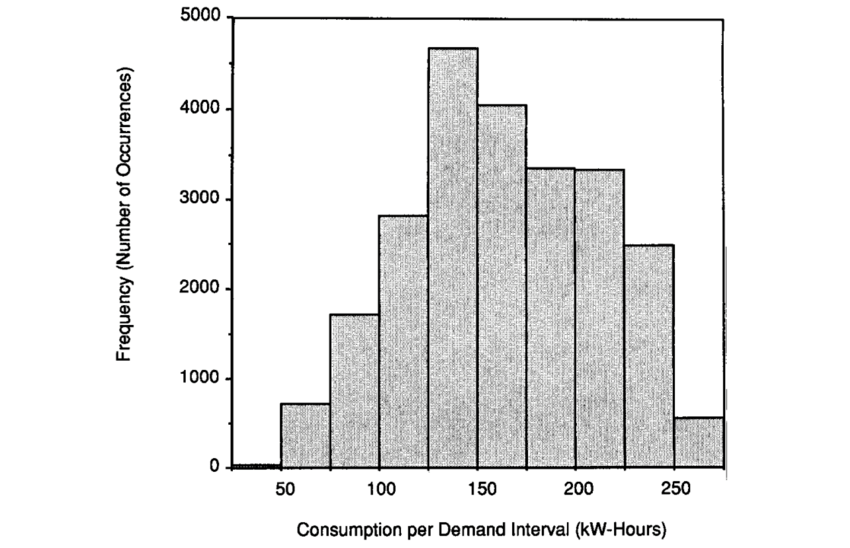
**plt.tight\_layout()**

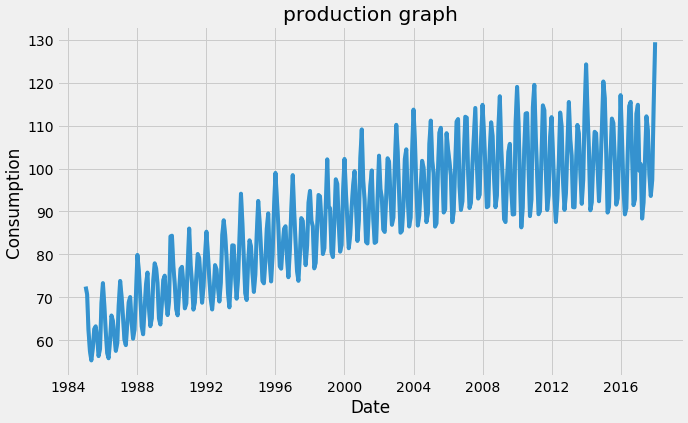
**plt.show()**

****

**This data set has 4000 many columns , I'm only interested in the below 5 columns for this project**

**OUTPUT:**

****

****

**7. Feedback & Iteration:**

* Collect feedback from stakeholders regarding the usability and insights of the dashboard.
* Refine visualizations based on feedback, ensuring clarity and relevance.

**CONCLUSION:**

By harnessing the power of data analysis and visualization, this project aims to turn raw energy consumption data into actionable insights. Whether it's a homeowner looking to reduce their energy bills, an industry trying to optimize its operations, or policymakers aiming to design better energy policies, the intuitive visualizations provide a clear picture of consumption patterns, paving the way for informed decisions and efficient energy use.