PHASE 4

DEVELOPMENT PART 2

MEASURE ENERGY CONSUMPTION

Exploratory Data Analysis (EDA):

- *Perform basic data exploration to understand the data's characteristics.*
- *Visualize the time series data to identify trends and patterns.*

Feature Engineering:

- *Create additional features, such as lag features (past values) or rolling statistics.*
- *These features can provide more information for forecasting.*

Time Series Forecasting:

- *Choose a time series forecasting model, such as ARIMA, Exponential Smoothing, or LSTM.*
- *Split the data into training and testing sets.*
- *Train the selected model on the training data and evaluate it on the test data.*

Python Script:

***Step 1:** Setup*

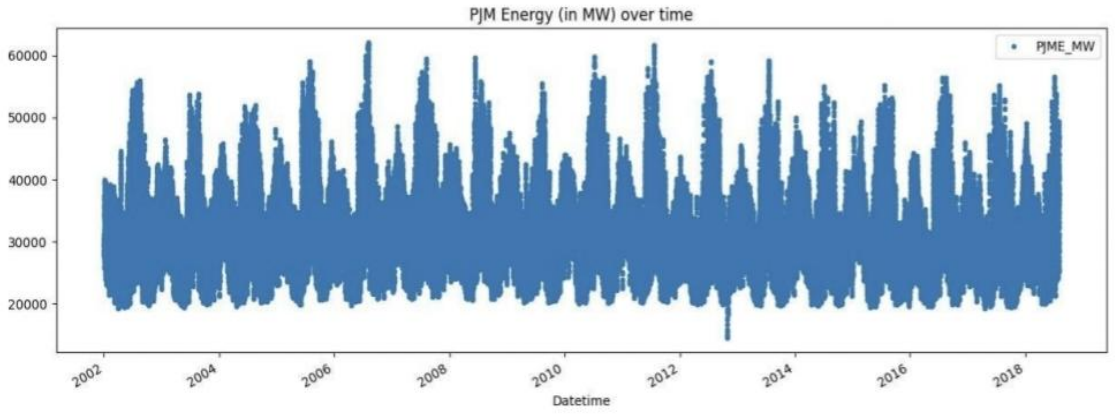
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import xgboost as xgb
from sklearn.metrics import mean_squared_error

# customize the style
pd.options.display.float_format = '{:.5f}'.format
pd.options.display.max_rows = 12
```

*#**Load the data***
filepath = '../input/hourly-energy-consumption/PJME_hourly.csv'
df = pd.read_csv(filepath)
print("Successfully Uploaded")

Step 2:** **Explore the data

*# **turn data to datetime***
df = df.set_index('Datetime')
df.index = pd.to_datetime(df.index)

*# **create the plot***
df.plot(style='.',
figsize=(15, 5),
title='PJM Energy (in MW) over time')
plt.show()

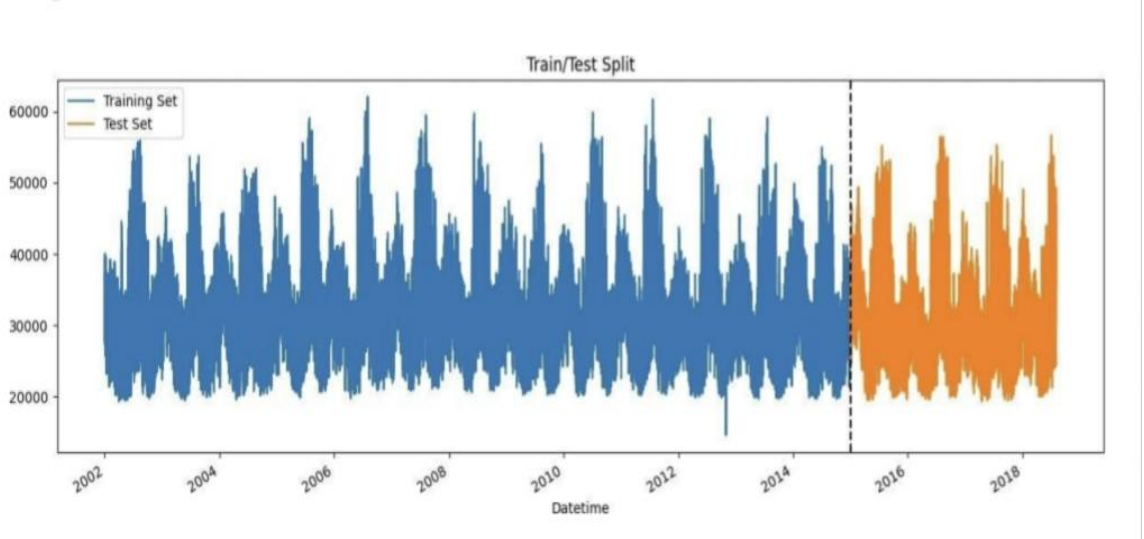


Step 2:** **Split the data

*# **train / test split***
train = df.loc[df.index < '01-01-2022']
test = df.loc[df.index >= '01-01-2022']
unfold_lessHide code
In [5]:
Linkcode

fig, ax = plt.subplots(figsize=(15, 5))
train.plot(ax=ax, label='Training Set', title='Train/Test Split')

```
test.plot(ax=ax, label='Test Set')
ax.axvline('01-01-2022', color='black', ls='--')
ax.legend(['Training Set', 'Test Set'])
plt.show()
```



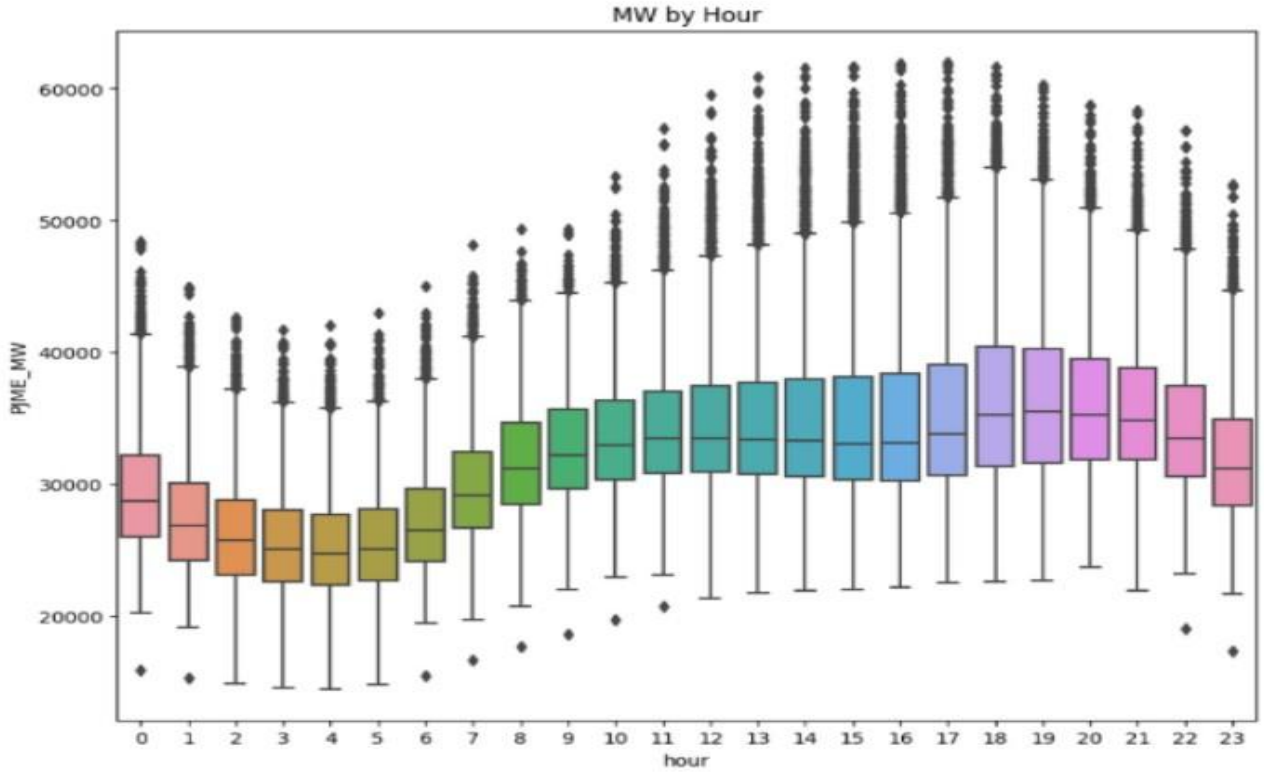
Step 3: ***Feature Engineering***

We're going to create some time features using the Datetime index. After that, we'll explore
the distributions of Hourly and Monthly megawatt usage.

```
# feature creation
def create_features(df):
 df = df.copy()
 df['hour'] = df.index.hour
 df['dayofweek'] = df.index.dayofweek
 df['quarter'] = df.index.quarter
 df['month'] = df.index.month
 df['year'] = df.index.year
 df['dayofyear'] = df.index.dayofyear
 df['dayofmonth'] = df.index.day
 df['weekofyear'] = df.index.isocalendar().week
 return df

df = create_features(df)
```

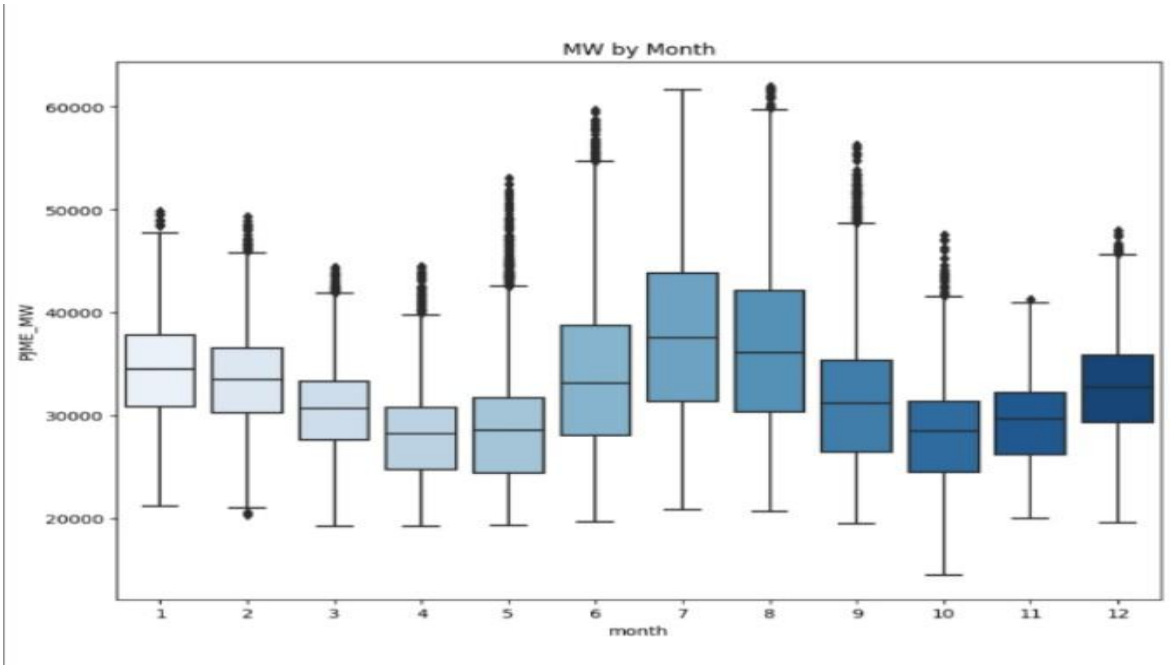
 # visualize the hourly Megawatt

```
fig, ax = plt.subplots(figsize=(10, 8))
sns.boxplot(data=df, x='hour', y='PJME_MW')
ax.set_title('MW by Hour')
plt.show()
```



 # visualize the monthly Megawatt

```
fig, ax = plt.subplots(figsize=(10, 8))
sns.boxplot(data=df, x='month', y='PJME_MW', palette='Blues')
ax.set_title('MW by Month')
plt.show()
```



Step 4: Modelling

XGBoost is good and reliable model for regression and time series analysis as well. Also, for
the metrics, we'll use mean squared error

Prepare the data

 # preprocessing
train = create_features(train)
test = create_features(test)
features = ['dayofyear', 'hour', 'dayofweek', 'quarter', 'month', 'year']
target = 'PJME_MW'
X_train = train[features]
y_train = train[target]
X_test = test[features]
y_test = test[target]

Build the model

import xgboost as xgb
from sklearn.metrics import mean_squared_error
 # build the regression model
reg = xgb.XGBRegressor(base_score=0.5, booster='gbtree',
n_estimators=1000,

```
 early_stopping_rounds=50,
 objective='reg:linear',
 max_depth=3,
 learning_rate=0.01)
reg.fit(X_train, y_train,
 eval_set=[(X_train, y_train), (X_test, y_test)],
 verbose=100)
```

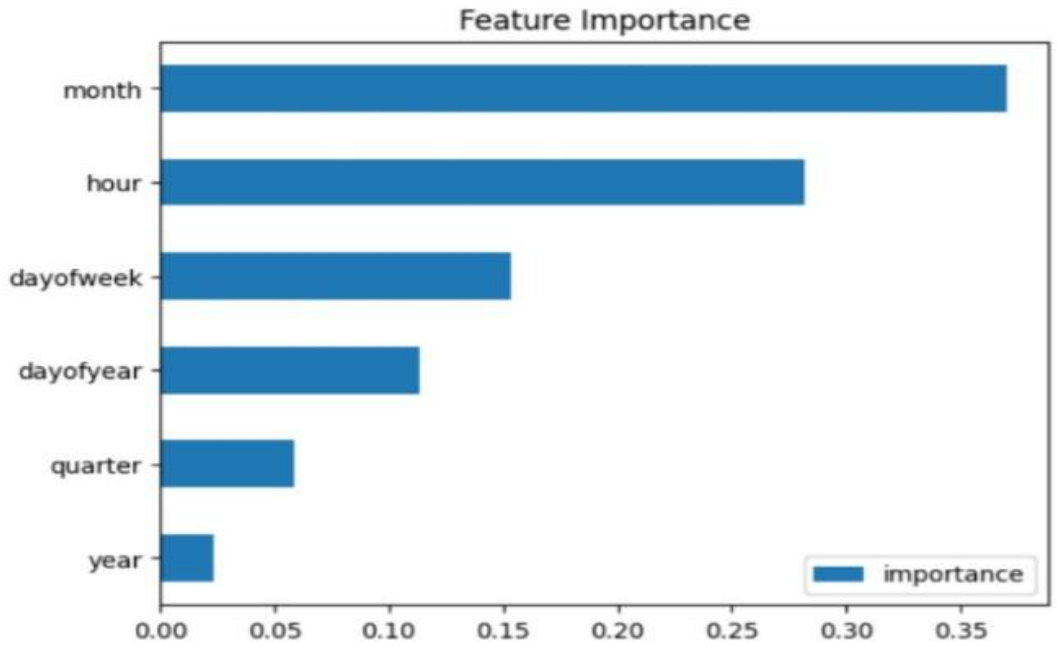
XGBRegressor

```
XGBRegressor(base_score=0.5, booster='gbtree', callbacks=None,
 colsample_bylevel=None, colsample_bynode=None,
 colsample_bytree=None, early_stopping_rounds=50,
 enable_categorical=False, eval_metric=None, feature_types=None,
 gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
 interaction_constraints=None, learning_rate=0.01, max_bin=None,
 max_cat_threshold=None, max_cat_to_onehot=None,
 max_delta_step=None, max_depth=3, max_leaves=None,
 min_child_weight=None, missing=nan, monotone_constraints=None,
 n_estimators=1000, n_jobs=None, num_parallel_tree=None,
 objective='reg:linear', predictor=None, ...)
```

Features importance

We need to see how much these features were used in each of the trees built by XGBoost
model.

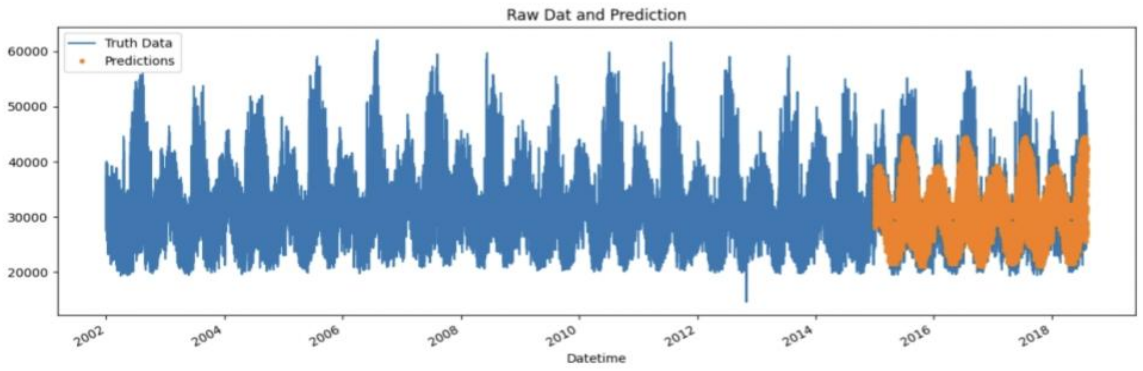
```
fi = pd.DataFrame(data=reg.feature_importances_,
 index=reg.feature_names_in_,
 columns=['importance'])
fi.sort_values('importance').plot(kind='barh', title='Feature Importance')
plt.show()
```



Forecasting on test data

compare the prediction with the actual values.

```
test['prediction'] = reg.predict(X_test)
df = df.merge(test[['prediction']], how='left', left_index=True,
right_index=True)
ax = df[['PJME_MW']].plot(figsize=(15, 5))
df['prediction'].plot(ax=ax, style='.')
plt.legend(['Truth Data', 'Predictions'])
ax.set_title('Raw Dat and Prediction')
plt.show()
```



*# **RMSE Score***
score = np.sqrt(mean_squared_error(test['PJME_MW'], test['prediction']))
print(f'RMSE Score on Test set: {score:0.2f}')

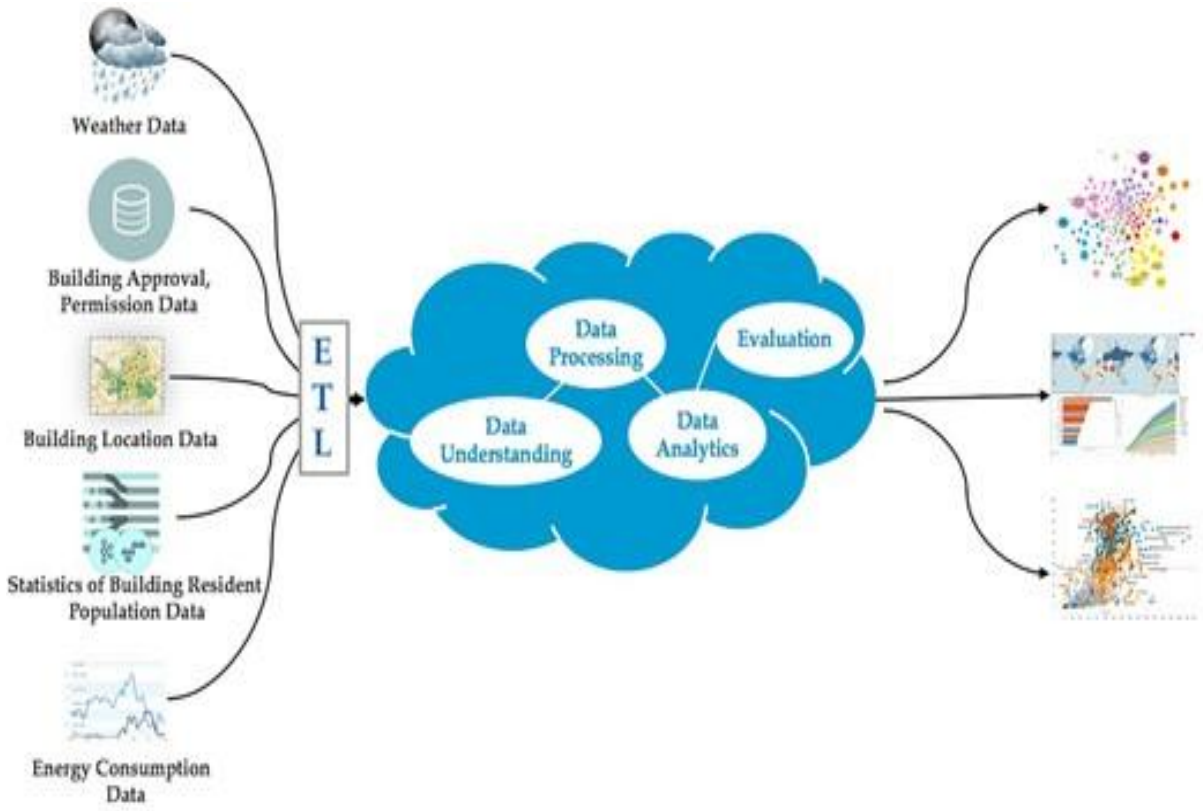
RMSE Score on Test set: 3721.75

 # R2 Score
from sklearn.metrics import r2_score
r2 = r2_score(test['PJME_MW'], test['prediction'])
print("R-squared (R2) Score:", r2)

R-squared (R2) Score: 0.6670230260104328

ANALYZING THE ENERGY CONSUMPTION DATA:

Energy consumption data can be analyzed in a variety of ways to identify trends, patterns, and areas for improvement. Here are some common steps involved in analyzing energy consumption data:



1. ***Collect and prepare the data:*** *This may involve gathering data from different sources, such as utility bills, energy monitoring systems, and building automation systems. The data should then be cleaned and formatted so that it can be easily analyzed.*

2. ***Visualize the data:*** *Creating charts and graphs can help you to identify trends and patterns in the data. For example, you could create a line chart to show how your energy consumption has changed over time, or a pie chart to show how your energy consumption is distributed across different areas of your building.*

3. ***Analyze the data:****Once you have visualized the data, you can start to analyze it to identify areas where you can reduce your energy consumption. For example, you could look for spikes in energy consumption that may be caused by inefficient equipment or processes. You could also compare your energy consumption to other similar buildings to see how you are performing.*

4. ***Identify opportunities for improvement:*** *Once you have identified areas where you can reduce your energy consumption, you can start to develop and implement energy efficiency measures. For example, you could upgrade to more efficient equipment, improve insulation, or change your operating procedures.*

5. ***Track your progress:****Once you have implemented energy efficiency measures, it is important to track your progress to see how they are impacting your energy consumption. This will help you to identify which measures are most effective and to make adjustments as needed.*

Analyzing energy consumption data for Measure Energy Consumption:

Measure Energy Consumption is a key performance indicator (KPI) that can be used to track and improve energy efficiency. It is calculated by dividing the total amount of energy consumed by a building or facility over a period of time by the total square footage of the building or facility.

To analyze energy consumption data for Measure Energy Consumption, you can follow the steps outlined above. In addition to the general steps, you may also want to consider the following:

- *Compare your Measure Energy Consumption to other similar buildings or facilities. This will help you to identify areas where you can improve your energy efficiency.*

- *Look for trends in your Measure Energy Consumption over time. This can help you to identify factors that are impacting your energy consumption and to develop strategies for reducing your consumption.*

- *Analyze your Measure Energy Consumption by building type, department, or equipment type. This can help you to identify the areas where you are using the most energy and to focus your energy efficiency efforts accordingly.*

By following these steps, you can gain valuable insights from your energy consumption data that can help you to improve your energy efficiency and reduce your energy costs.

Here are some specific examples of how you can analyze energy consumption data for Measure Energy Consumption:

- *Compare your Measure Energy Consumption to the average Measure Energy Consumption for buildings of your type. You can find this information from various sources, such as the US Energy Information Administration (EIA). If your Measure Energy Consumption is significantly higher than the average, it may indicate that you have opportunities to improve your energy efficiency.*

- *Look for trends in your Measure Energy Consumption over the past year or two. Has your consumption been increasing, decreasing, or remaining relatively constant? If it has been increasing, try to identify the factors that are contributing to the increase. For example, have you added new equipment to your building? Has the weather been colder or hotter than usual?*

- *Analyze your Measure Energy Consumption by building type, department, or equipment type. This can help you to identify the areas where you are using the most energy and to focus your energy efficiency efforts accordingly. For example, if you have a large data center, you may want to focus on improving the energy efficiency of your IT equipment.*

By analyzing your energy consumption data in these ways, you can gain valuable insights that can help you to improve your energy efficiency and reduce your energy costs.

CREATING VISUALIZATION:

Creating visualizations for measuring energy consumption can provide valuable insights and help make informed decisions. Here's a step-by-step guide:

1. ***Data Collection**: Gather data on energy consumption. This data can come from utility bills, smart meters, or sensors that monitor energy use.*

2. ***Select Visualization Tools:** Choose a visualization tool or software. Common options include Excel, Tableau, Power BI, Python libraries (Matplotlib, Seaborn), or specialized energy management software.*

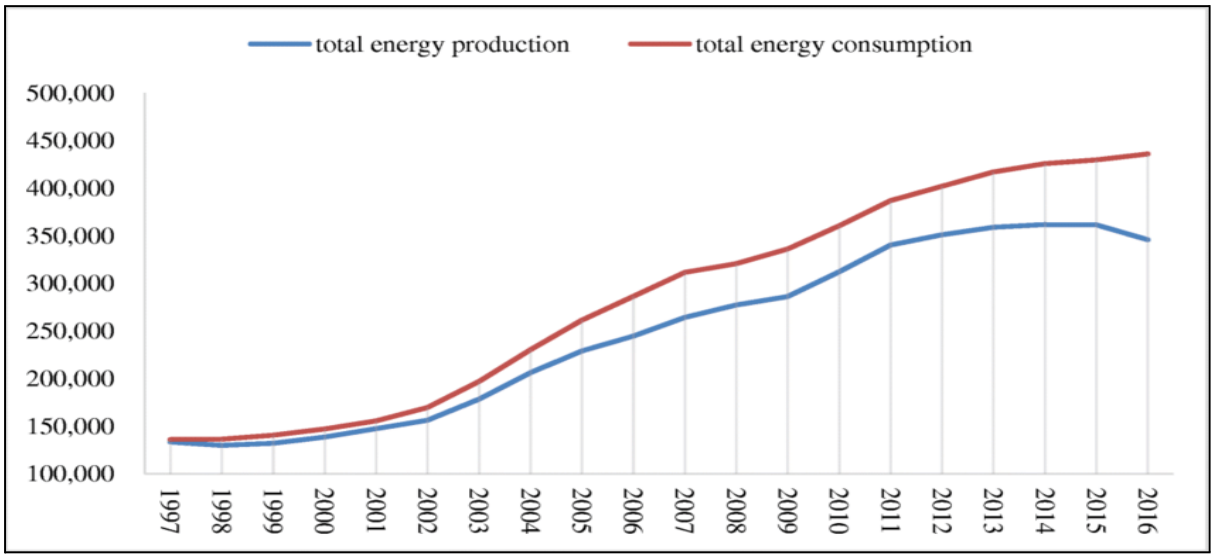
3. ***Choose the Right Chart Types:***

    - ***Line Charts:** Show energy consumption trends over time.*
    - ***Bar Charts:** Compare energy consumption among different sources or periods.*
    - ***Pie Charts:** Display the breakdown of energy use by category (e.g., lighting, HVAC).*

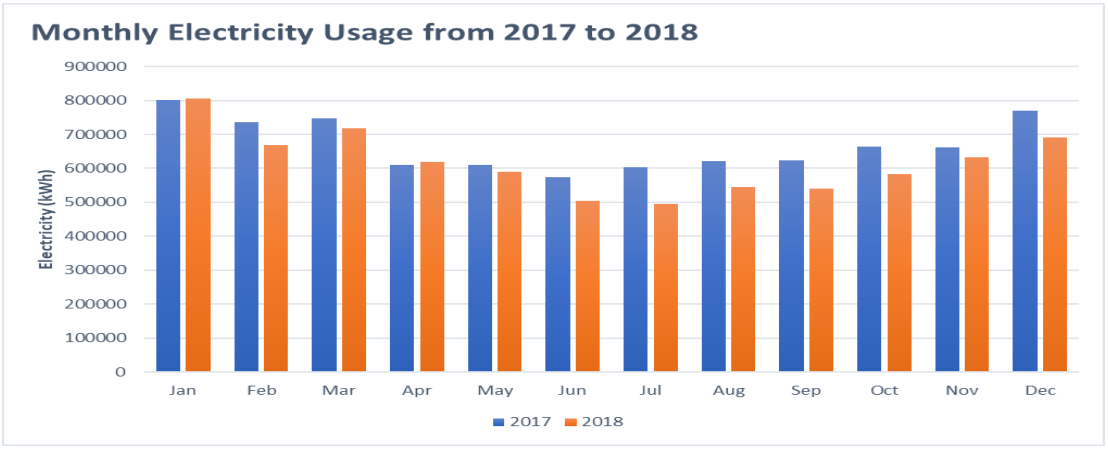
- ***Heatmaps:*** *Visualize energy consumption patterns in a specific area.*

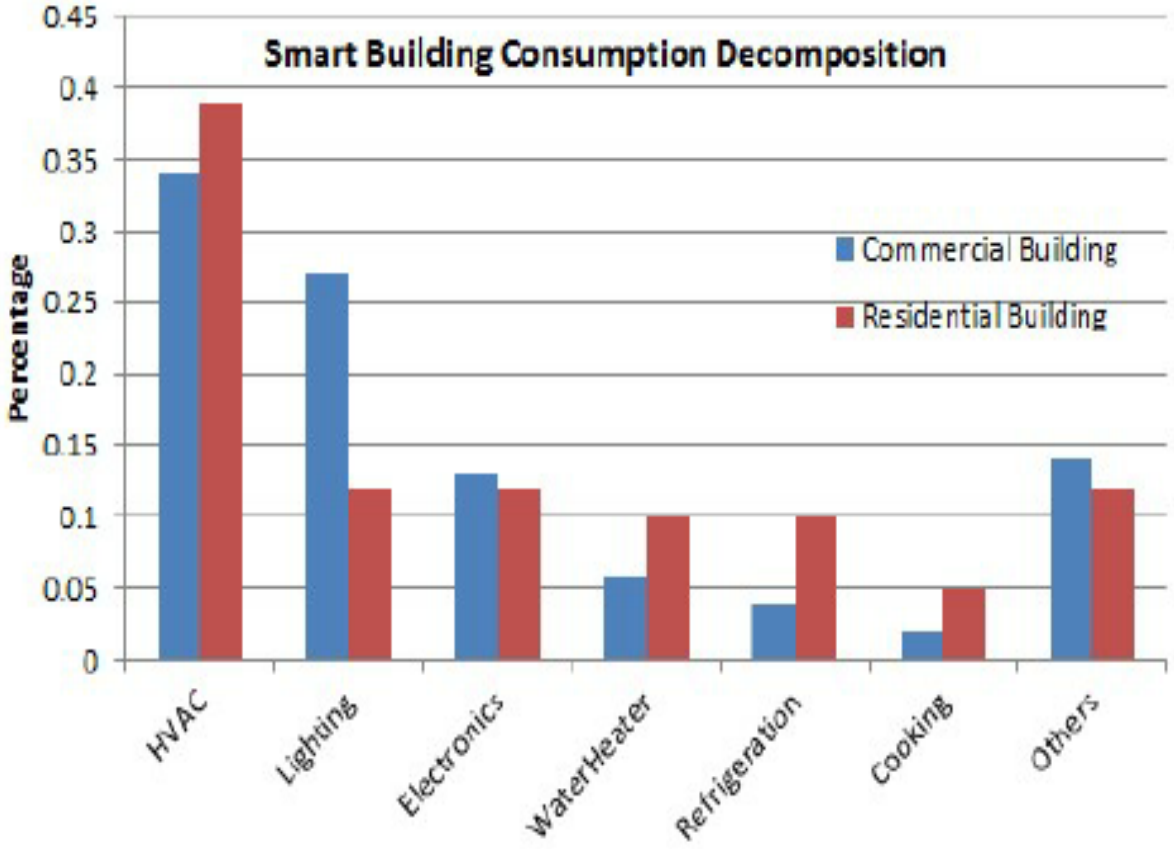
There are a number of different ways to create visualizations for energy consumption data. Some common types of visualizations include:

- ***Line charts:*** *Line charts can be used to show how energy consumption changes over time. This can be helpful for identifying trends in consumption and for tracking the impact of energy-saving measures.*

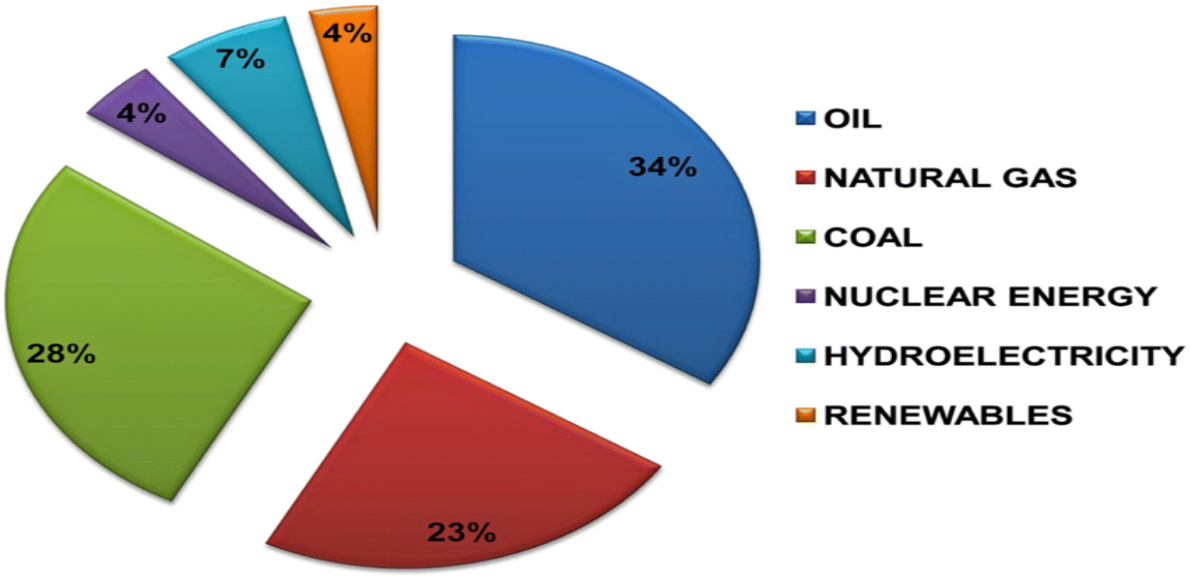


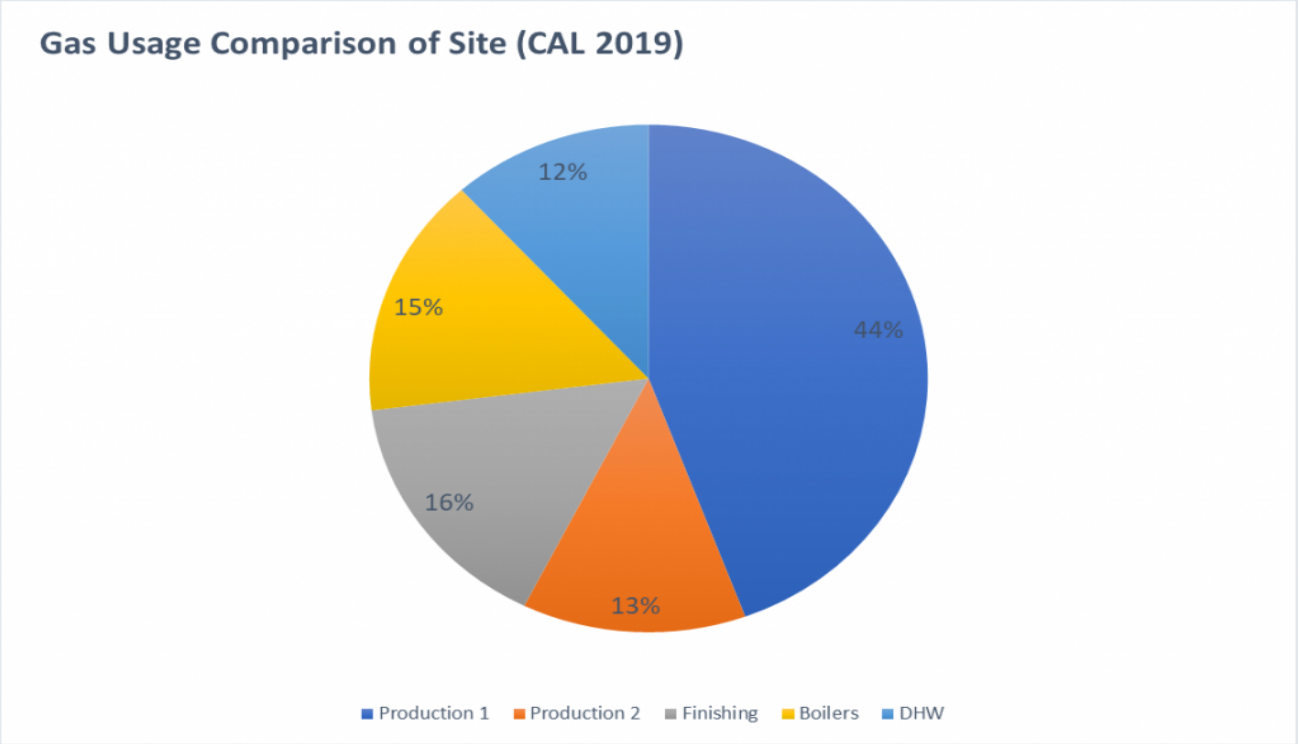
- ***Bar charts:*** *Bar charts can be used to compare energy consumption between different categories, such as different types of equipment or different areas of a facility.*



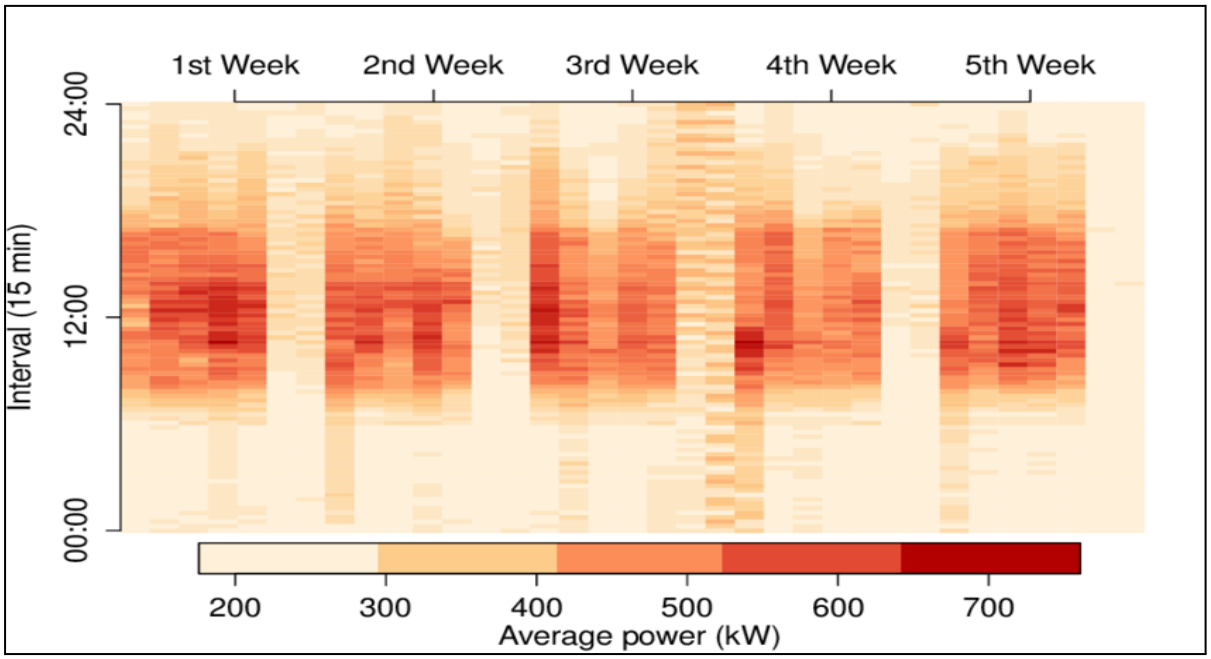


- ***Pie charts:*** *Pie charts can be used to show how energy consumption is divided between different categories.*

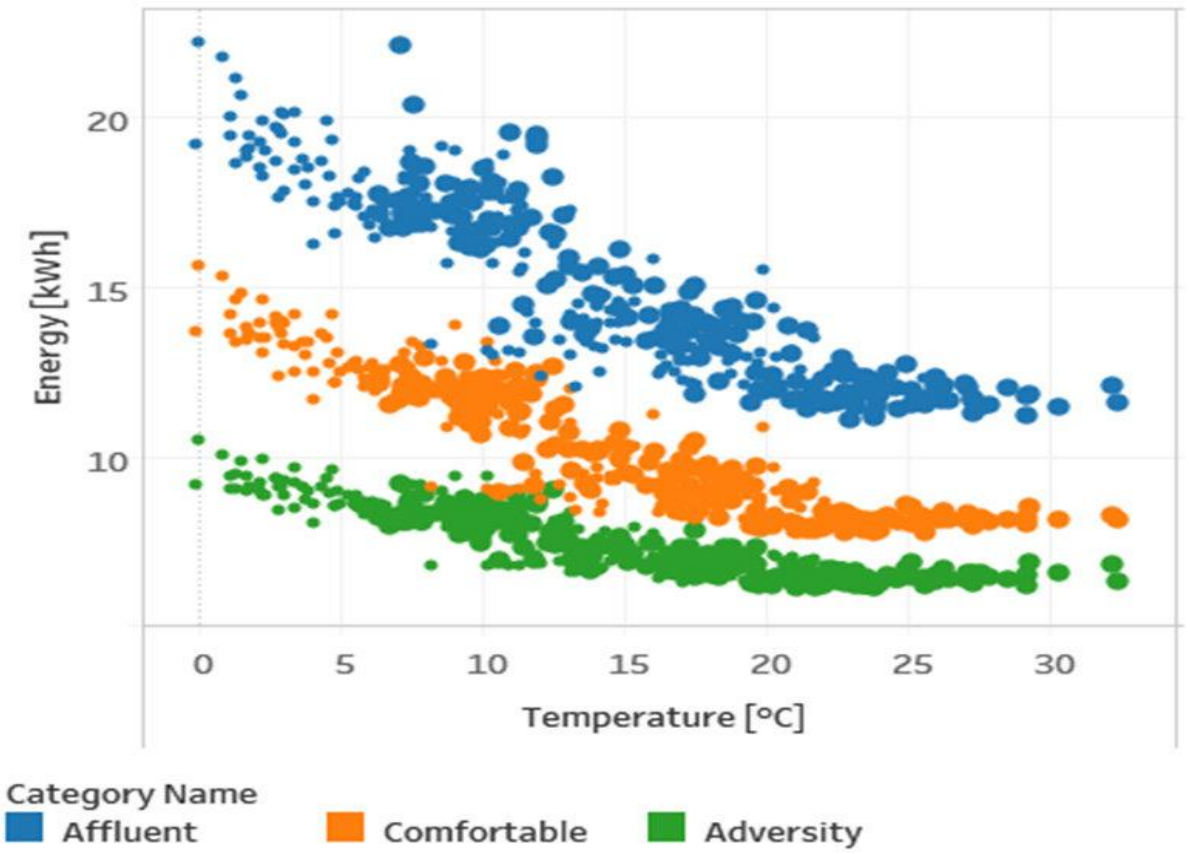


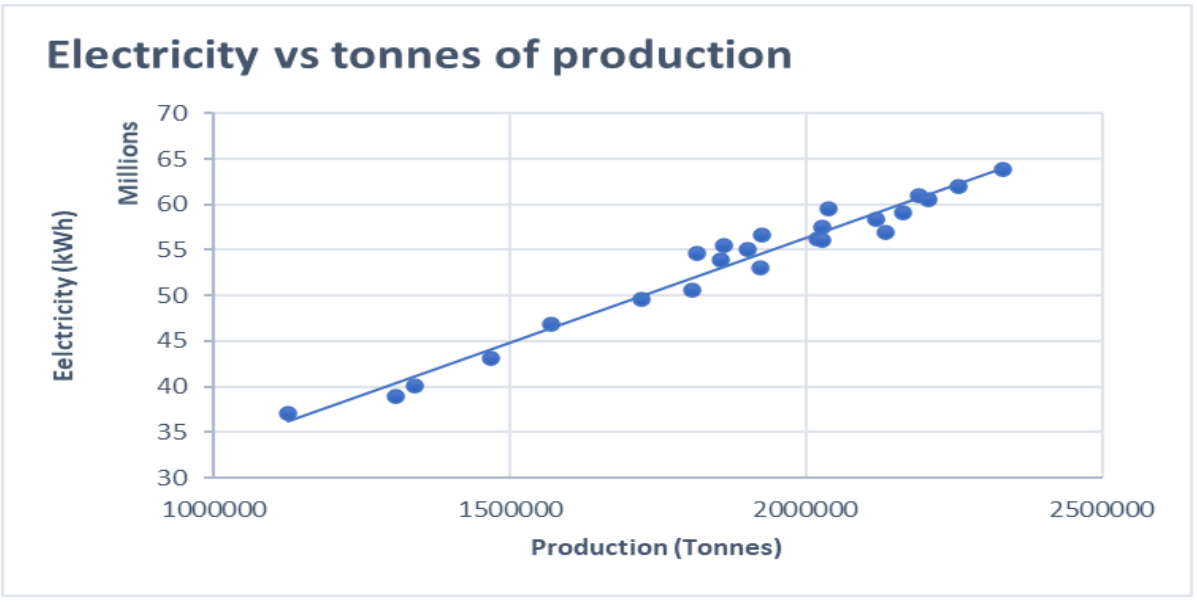


- ***Heatmaps:*** *Heatmaps can be used to show how energy consumption varies across a particular area or space.*



- ***Scatter plots:*** *Scatter plots can be used to identify relationships between different variables, such as energy consumption and weather conditions.*





- ***Indicators:*** *Indicators are great to use if your business has set targets in your energy usage, as they can then be used to track your current energy usage. They are easy to read and understand from a simple glance and can be expressed either as a gauge or ticker to create greater visual impact.*

Electricity Consumption (kWh)

Target	2000
January	1611
February	1548
March	1747
April	1614
May	2115
June	2024
July	1626
August	1662
September	2051
October	1567
November	1690
December	2124

The best type of visualization to use will depend on the specific data being analyzed and the purpose of the visualization. For example, a line chart may be the best way to show how energy consumption has changed over time, while a pie chart may be the best way to show how energy consumption is divided between different categories.

4. ***Data Preprocessing:***

- *Clean and organize your data.*
- *Calculate any necessary metrics, like daily or monthly averages.*

5. ***Create the Visualizations:***

- *Plot your data using the selected chart types.*
- *Add labels, legends, and titles for clarity.*
- *Use color coding to differentiate categories or trends.*

6. ***Analyze the Visualizations:***

- *Interpret the patterns and trends in your data.*
- *Look for spikes, seasonality, or anomalies.*

7. ***Interactive Dashboards :***

- *If you want to provide real-time monitoring, consider creating an interactive dashboard using tools like Tableau or Power BI.*

8. ***Share and Collaborate:***

    - *Share the visualizations with relevant stakeholders, such as energy managers, decision-makers, or facility operators.*

9. ***Set Alerts and Goals:***

    - *Use your visualizations to set energy consumption goals and thresholds. Implement alerts for when these are exceeded.*

10. ***Continuous Monitoring:***

    - *Update your visualizations regularly to track progress and make adjustments as needed.*

Remember to tailor your visualizations to the specific needs and goals of your energy management project. Effective visualization can help identify energy-saving opportunities and promote energy efficiency.

In addition to the type of visualization, it is also important to consider the following factors when creating visualizations for energy consumption data:

- ***Clarity**: The visualization should be clear and easy to understand.*
- ***Accuracy**: The visualization should accurately represent the data.*
- ***Completeness**: The visualization should include all of the relevant data.*
- ***Context**: The visualization should be presented in a context that helps users to understand the data.*

By following these guidelines, you can create visualizations that are both informative and engaging.

Here are some specific examples of visualizations for energy consumption data:

- *A line chart showing how energy consumption has changed over the past year. This could be used to track the impact of energy-saving measures or to identify seasonal trends in consumption.*

- *A bar chart comparing energy consumption between different departments in a company. This could be used to identify departments that are using a large amount of energy.*

- *A pie chart showing how energy consumption is divided between different types of equipment. This could be used to identify types of equipment that are using a large amount of energy.*

- *A heatmap showing how energy consumption varies across a building. This could be used to identify areas of the building that are using a large amount of energy.*

- *A scatter plot showing the relationship between energy consumption and weather conditions. This could be used to identify how weather conditions affect energy consumption.*

By creating visualizations for energy consumption data, you can gain a better understanding of how energy is being used. This information can then be used to identify areas for energy savings and to develop and implement energy management plans.