

MEASURE ENERGY CONSUMPTION

PROBLEM STATEMENT:

This project aims to analyze historical power consumption data from various sources and create meaningful visualizations to understand consumption patterns and trends. The project helps in making informed decisions related to power generation and distribution.

PHASES OF DEVELOPMENT:

Data Collection: Gathered data from multiple sources, including AEP, COMED, DAYTON, DEOK, DOM, DUQ, EKPC, FE, NI, and more.

DATA PREPROCESSING:

Data Collection:

Gathered data from multiple sources, including AEP, COMED, DAYTON, DEOK, DOM, DUQ, EKPC, FE, NI, and more.

The data is collected in the form of CSV files, each containing hourly power consumption data for different regions.

Data Merging:

Merged data from multiple CSV files to consolidate it into a single dataset.

The merging was performed by concatenating the datasets along the columns and ensuring that there were no duplicate columns.

Handling Missing Values:

Checked for missing values in the dataset.

Dealt with any missing values through methods such as removal or imputation, ensuring the dataset is ready for analysis.

Datetime Conversion:

Converted the 'Datetime' column to a datetime format. This is essential for time-based analysis and visualization.

Program:

```
import pandas as pd

dataset_filenames = [ 'AEP_hourly.csv', 'COMED_hourly.csv', 'DAYTON_hourly.csv', 'DEOK_hourly.csv',
'DOM_hourly.csv','DUQ_hourly.csv','EKPC_hourly.csv','FE_hourly.csv','NI_hourly.csv',
'pjm_hourly_est.csv','PJM_Load_hourly.csv', 'PJME_hourly.csv','PJM_W_hourly.csv']

datasets = []

for dataset_filename in dataset_filenames:

    data = pd.read_csv(dataset_filename)
```

```

datasets.append(data)

merged_dataset = pd.concat(datasets, axis=1)

merged_dataset = merged_dataset.loc[:, ~merged_dataset.columns.duplicated()]

print(merged_dataset)

```

Output:

```

Datetime AEP_MW COMED_MW DAYTON_MW DEOK_MW DOM_MW \ 0 2004-12-31 01:00:00
13478.0 9970.0 1596.0 2945.0 9389.0 1 2004-12-31 02:00:00 12865.0 9428.0 1517.0 2868.0 9070.0 2 2004-
12-31 03:00:00 12577.0 9059.0 1486.0 2812.0 9001.0 3 2004-12-31 04:00:00 12517.0 8817.0 1469.0
2812.0 9042.0 4 2004-12-31 05:00:00 12670.0 8743.0 1472.0 2860.0 9132.0 ... ... ... 178257 NaN
NaN NaN NaN NaN NaN 178258 NaN NaN NaN NaN NaN NaN 178259 NaN NaN NaN NaN NaN NaN
178260 NaN NaN NaN NaN NaN NaN NaN 178261 NaN NaN NaN NaN NaN NaN DUQ_MW EKPC_MW
FE_MW NI_MW ... DUQ EKPC FE NI \ 0 1458.0 1861.0 6222.0 9810.0 ... NaN NaN NaN NaN 1 1377.0
1835.0 5973.0 9001.0 ... NaN NaN NaN NaN 2 1351.0 1841.0 5778.0 8509.0 ... NaN NaN NaN NaN 3
1336.0 1872.0 5707.0 8278.0 ... NaN NaN NaN NaN 4 1356.0 1934.0 5691.0 8089.0 ... NaN NaN NaN NaN
... ... ... 178257 NaN NaN NaN NaN ... 1962.0 2866.0 9378.0 NaN 178258 NaN NaN NaN
NaN ... 1940.0 2846.0 9255.0 NaN 178259 NaN NaN NaN NaN ... 1891.0 2883.0 9044.0 NaN 178260 NaN
NaN NaN NaN ... 1820.0 2880.0 8676.0 NaN 178261 NaN NaN NaN NaN ... 1721.0 2846.0 8393.0 NaN
PJME PJMW PJM_Load PJM_Load_MW PJME_MW PJMW_MW 0 NaN NaN 29309.0 29309.0 26498.0
5077.0 1 NaN NaN 28236.0 28236.0 25147.0 4939.0

2 NaN NaN 27692.0 27692.0 24574.0 4885.0 3 NaN NaN 27596.0 27596.0 24393.0 4857.0 4 NaN NaN
27888.0 27888.0 24860.0 4930.0 ... ... ... 178257 44284.0 8401.0 NaN NaN NaN NaN 178258
43751.0 8373.0 NaN NaN NaN NaN 178259 42402.0 8238.0 NaN NaN NaN NaN 178260 40164.0 7958.0
NaN NaN NaN NaN 178261 38608.0 7691.0 NaN NaN NaN NaN [178262 rows x 25 columns]

```

VISUALIZATION:

Line Plot: Power Consumption Over Time:

Objective: To visualize and understand how power consumption changes over time.

Description: A line plot was created to show the power consumption of AEP and COMED over time. The x-axis represents time (Datetime), and the y-axis represents power consumption in Megawatts (MW).

Insights: This visualization helps in identifying trends and patterns in power consumption for AEP and COMED.

Scatter Plot: PJM_Load vs. PJM_Load_MW:

Objective: To explore the relationship between 'PJM_Load' and 'PJM_Load_MW'.

Description: A scatter plot was used to display data points where the x-axis represents 'PJM_Load', and the y-axis represents 'PJM_Load_MW'. Each point on the plot represents a specific observation.

Insights: This visualization helps in understanding how 'PJM_Load' relates to the actual power consumption 'PJM_Load_MW'. It can reveal trends and correlations.

Histogram: AEP_MW Distribution:

Objective: To visualize the distribution of 'AEP_MW' values.

Description: A histogram was created to display the frequency distribution of 'AEP_MW' values. The x-axis represents the 'AEP_MW' values, and the y-axis represents the frequency of occurrence.

Insights: This visualization helps in understanding the spread and concentration of power consumption levels for AEP.

Dataset

Dataset Description: The dataset contains hourly power consumption data from different regions, including the 'Datetime' and 'MW' columns.

Data Preprocessing: Merged data from multiple CSV files, handled missing values, and converted the 'Datetime' column to a datetime format.

Visualization Techniques

Line Plot:

Power Consumption Over Time: Visualizes power consumption of AEP and COMED over time.

Scatter Plot:

PJM_Load vs. PJM_Load_MW: Displays the relationship between 'PJM_Load' and 'PJM_Load_MW'.

Histogram:

AEP_MW Distribution: Shows the distribution of 'AEP_MW'.

Program:

```
import pandas as pd
import matplotlib.pyplot as plt

dataset_filenames = [
    'AEP_hourly.csv', 'COMED_hourly.csv', 'DAYTON_hourly.csv', 'DEOK_hourly.csv',
    'DOM_hourly.csv', 'DUQ_hourly.csv', 'EKPC_hourly.csv', 'FE_hourly.csv', 'NI_hourly.csv',
    'pjm_hourly_est.csv', 'PJM_Load_hourly.csv', 'PJME_hourly.csv', 'PJM_W_hourly.csv'
]

datasets = []
for dataset_filename in dataset_filenames:
    data = pd.read_csv(dataset_filename)
    datasets.append(data)

merged_dataset = pd.concat(datasets, axis=1)
merged_dataset = merged_dataset.loc[:, ~merged_dataset.columns.duplicated()]
merged_dataset = merged_dataset.dropna(subset=['Datetime'])
merged_dataset['Datetime'] = pd.to_datetime(merged_dataset['Datetime'])

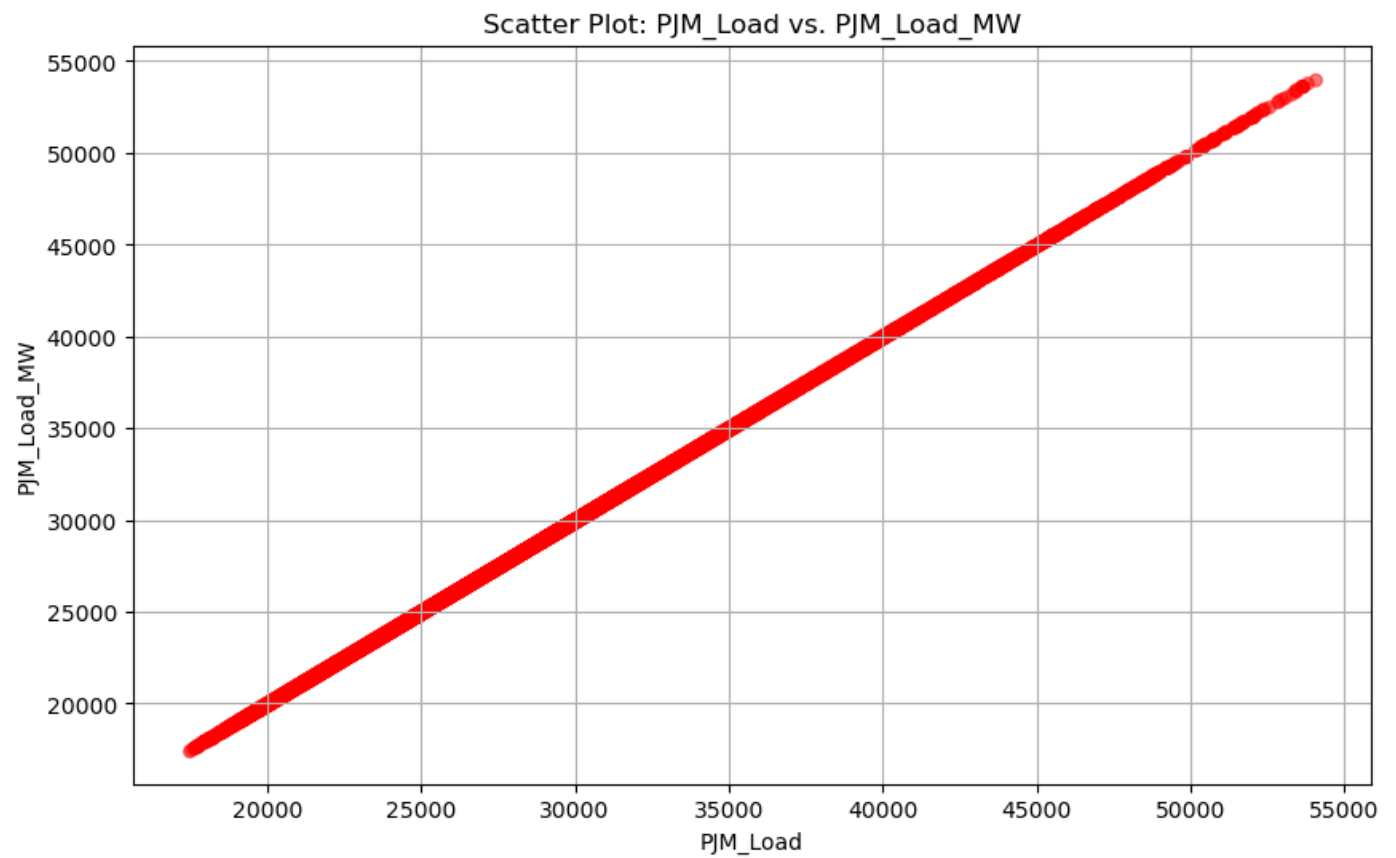
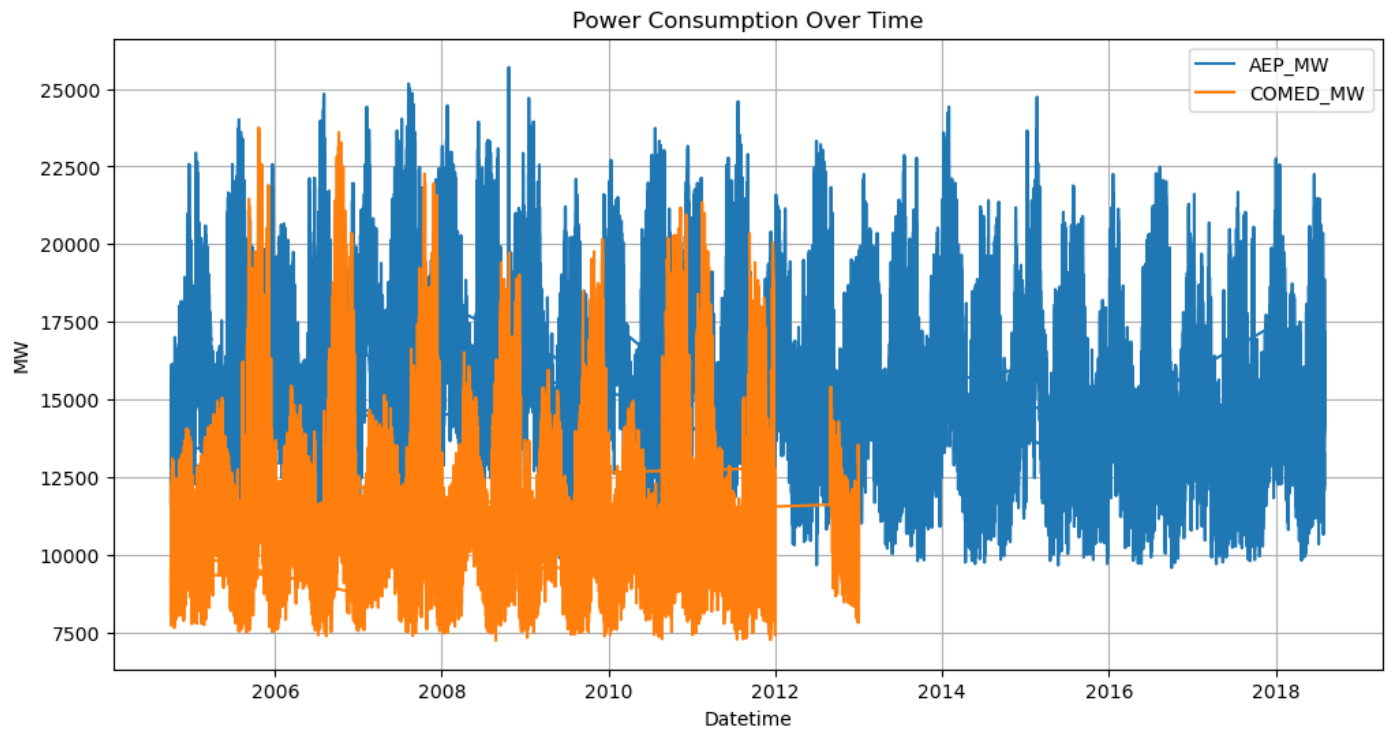
plt.figure(figsize=(12, 6))
```

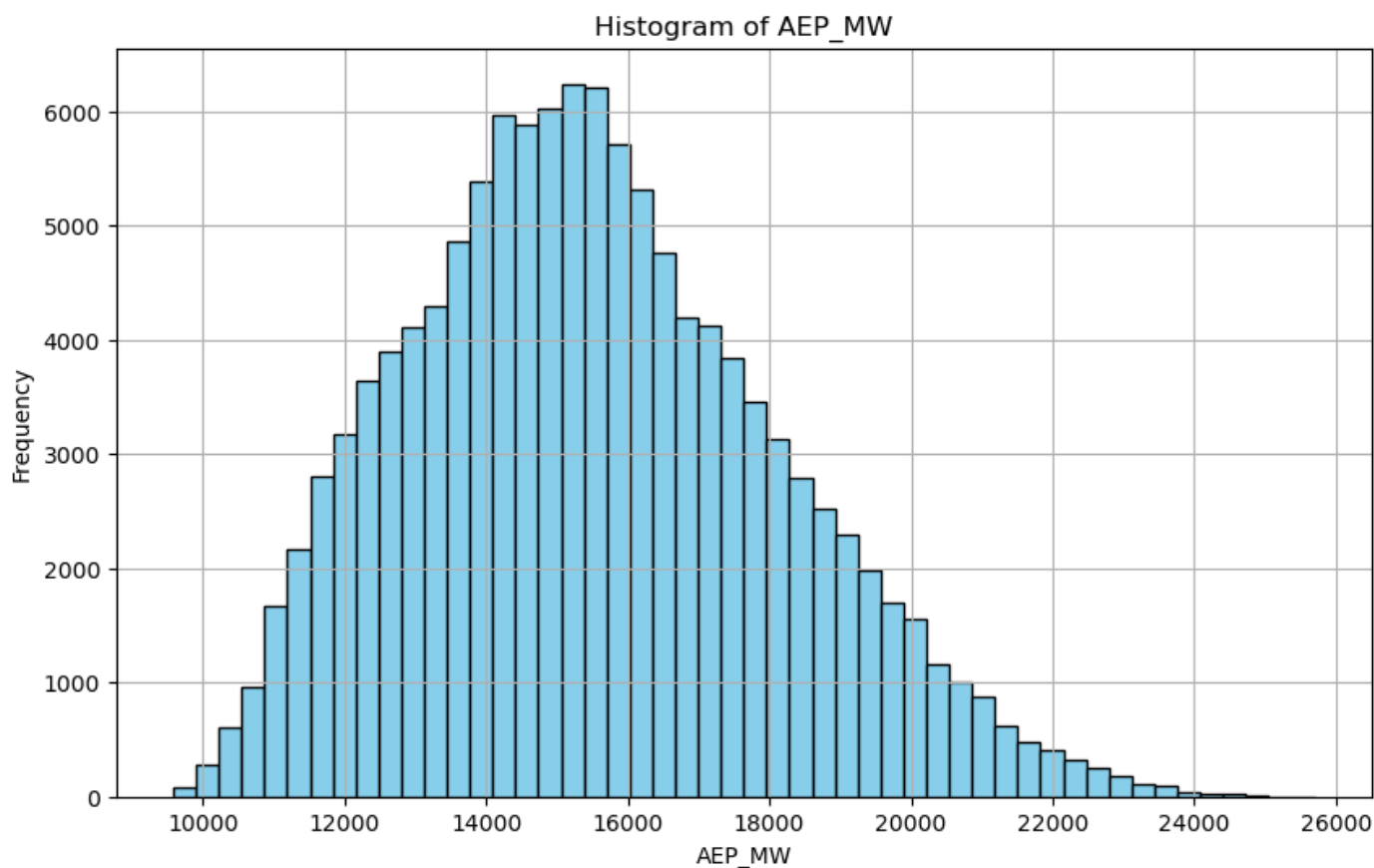
```
plt.plot(merged_dataset['Datetime'], merged_dataset['AEP_MW'], label='AEP_MW')
plt.plot(merged_dataset['Datetime'], merged_dataset['COMED_MW'], label='COMED_MW')
plt.xlabel('Datetime')
plt.ylabel('MW')
plt.legend()
plt.title('Power Consumption Over Time')
plt.grid(True)
plt.show()

plt.figure(figsize=(10, 6))
plt.scatter(merged_dataset['PJM_Load'], merged_dataset['PJM_Load_MW'], s=30, alpha=0.5,
marker='o',color='red')
plt.xlabel('PJM_Load')
plt.ylabel('PJM_Load_MW')
plt.title('Scatter Plot: PJM_Load vs. PJM_Load_MW')
plt.grid(True)
plt.show()

plt.figure(figsize=(10, 6))
plt.hist(merged_dataset['AEP_MW'], bins=50, color='skyblue', edgecolor='black')
plt.xlabel('AEP_MW')
plt.ylabel('Frequency')
plt.title('Histogram of AEP_MW')
plt.grid(True)
plt.show()
```

output:





CONCLUSION:

In this project, we looked at power consumption data from different sources and made cool graphs to understand how electricity is used.

We cleaned up the data to make it nice and put it all together. Then, we used line graphs, scatter plots, and histograms to make pictures that show the patterns in the data.