Predicting Solar Energy Production Using Python

Presented By:Abhishek Sharma

Introduction

Overview

- Importance of solar energy production forecasting.
- Benefits of using data-driven methods for predictions.

Data Collection

Key Points

- Historical solar production data.
- Weather data (irradiance, temperature, etc.).
- System specifications (PV size, storage size).

Visual: Diagram showing data sources

```
data= pd.read_csv("Solar Energy.csv")
```

```
# Check the columns of the DataFrame
print("Columns in DataFrame:", data.columns)

# Print the first few rows of the DataFrame
data.head().style.set_properties(
    **{
        'background-color': 'OliveDrab',
        'color': 'white',
        'border-color': 'darkblack'
    })
```

```
Columns in DataFrame: Index(['Data Through Date', 'Project ID', 'Interconnection Date', 'Utility',
       'City/Town', 'County', 'Zip', 'Division', 'Substation', 'Circuit ID',
       'Developer', 'Metering Method', 'Estimated PV System Size (kWdc)',
       'PV System Size (KWac)', 'Estimated Annual PV Energy Production (kWh)',
       'Energy Storage System Size (kWac)', 'Number of Projects'],
      dtype='object')
```

	Data Through Date	Project ID	Interconnection Date	Utility	City/Town	County	Zip	Division	Substation	Circuit ID	Developer	Metering Method	Estim Sy (k
0	12/31/2023	SDG-66301	12/29/2023	Con Ed	Richmond Hill	Queens	11418.000000	CENY-BK	Brownsville_2	9B05	Kamtech Solar Solutions	NM	6.05
1	12/31/2023	SDG-66299	12/29/2023	Con Ed	Bronx	Bronx	10473.000000	CENY-BX	Parkchester_2	5X67	Kamtech Solar Solutions	NM	6.74
2	12/31/2023	SDG-66288	12/29/2023	Con Ed	Brooklyn	Kings	11225.000000	CENY-BK	Bensonhurst_2	4B08	SUNCO	NM	3.05
3	12/31/2023	SDG-66284	12/29/2023	Con Ed	Brooklyn	Kings	11236.000000	CENY-BK	Bensonhurst_2	3037	Kamtech Solar Solutions	NM	5.62
4	12/31/2023	SDG-66277	12/28/2023	Con Ed	Springfield Gardens	Queens	11413.000000	CENY-Q	Jamaica	9112	Kamtech Solar Solutions	NM	6.05

```
data.describe().style.background_gradient(cmap='rainbow')
```

	Zip	Estimated PV System Size (kWdc)	PV System Size (kWac)	Estimated Annual PV Energy Production (kWh)	Energy Storage System Size (kWac)
count	218019.000000	218115.000000	218115.000000	218115.000000	4091.000000

Data Preparation

Key Steps:

- Data cleaning (handling missing values, outliers).
- Normalization and scaling.

Code Snippet:

```
# Handling missing values
features_cleaned = features.dropna()
```

Feature Engineering

Key Points:

- Creating relevant features (time-based, lag features).
- Correlation analysis to identify important features.

Model Selection

Key Points:

- Choosing appropriate models (e.g., Random Forest, Linear Regression).
- Importance of hyperparameter tuning.

Code Snippet:

```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
```

Model Training

Key Steps:

- Splitting data into training and testing sets.
- Fitting the model to the training data.

Code Snippet:

```
model.fit(X_train, y_train)
```

Evaluation

Key Points:

- Evaluating model performance using metrics (MAE, MSE).

Code Snippet:

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, y_pred)
```

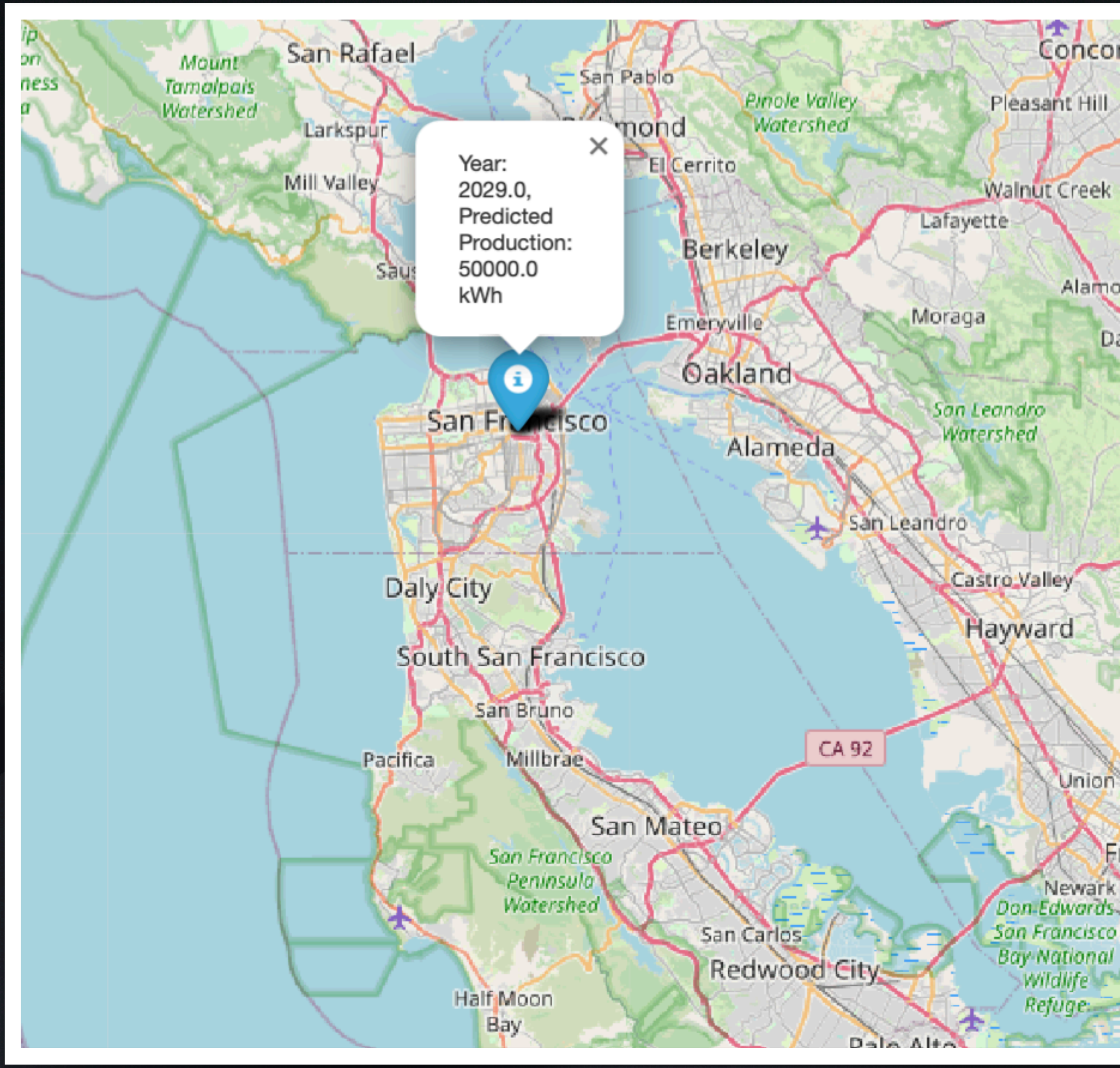
Future Predictions

Key Steps

- Preparing data for the next 5 years.
- Making predictions using the trained model.

Code Snippet:

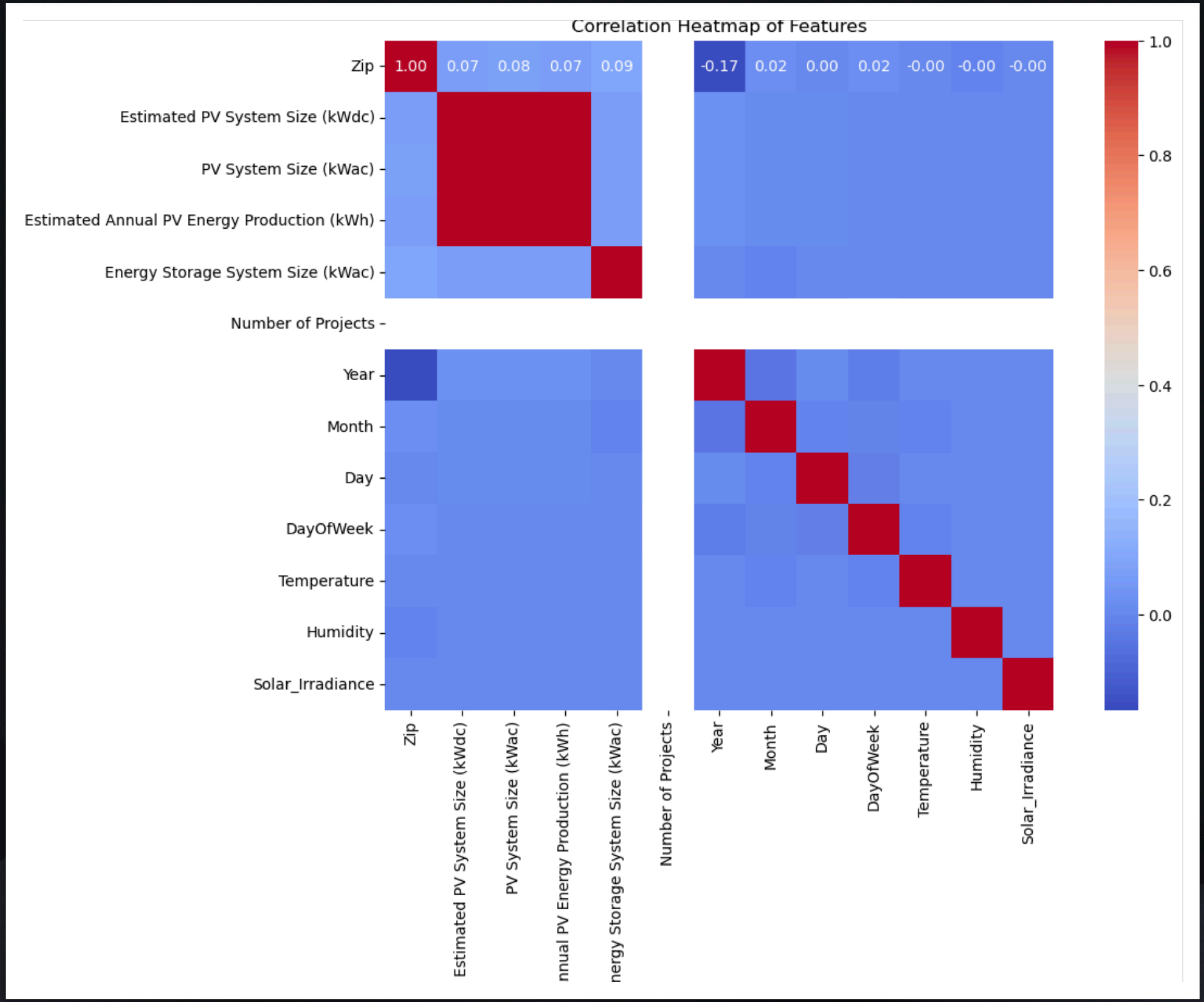
```
future_years = pd.DataFrame({
    'PV System Size (kW)': [data['PV System Size (kW)'].mean()] * 5
    'Year': [2025 + i for i in range(5)]
})
predictions = model.predict(future_years)
```



Visualization

Key Points:

1. **Enhanced Understanding:**
    - Visualization transforms complex data into intuitive graphical representations, making it easier for stakeholders to grasp trends, patterns, and relationships within the data. This clarity aids in better decision-making and facilitates communication of insights to both technical and non-technical audiences.
2. **Identifying Insights and Anomalies:**
    - Effective visualizations can highlight key insights and anomalies that may not be immediately apparent in raw data. By using charts, graphs, and dashboards, users can quickly identify outliers, seasonal trends, and correlations, enabling proactive responses to potential issues and informed strategic planning.



- Tools for visualization (Tableau).

Conclusion

Summary:

Comprehensive Approach:

- The process of predicting solar energy production involves a systematic approach that includes data collection, preparation, feature engineering, model selection, training, evaluation, and future predictions. Each step is crucial for building a robust predictive model that accurately reflects solar energy generation potential.

Significance of Accurate Forecasting:

- Accurate forecasting is essential for optimizing resource allocation, enhancing financial planning, and ensuring effective grid management. It plays a vital role in supporting the transition to renewable energy sources and helps stakeholders make informed decisions.

Impact on Sustainability:

- By improving the accuracy of solar energy production forecasts, we can facilitate better integration of solar power into the energy grid, reduce reliance on fossil fuels, and contribute to environmental sustainability. This ultimately supports the global effort to combat climate change and promote clean energy solutions.