Phase 3 : Development

Code:

```
#Loading the dataset into python

import pandas as pd

df = pd.read_csv('electricity_price_prediction.csv')

#Exploring the dataset

# Display the first few rows of the dataset

print("First 5 rows of the dataset:")

print(df.head())

# Display the shape of the dataset (number of rows, number of columns)

print("\nShape of the dataset:")

print(df.shape)

# Display summary statistics of numerical columns

print("\nSummary statistics:")

print(df.describe())

# Check for missing values in each column

print("\nMissing values:")

print(df.isnull().sum())

# Check data types of each column

print("\nData types:")

print(df.dtypes)

# Explore unique values in categorical columns

categorical_columns = ['column1', 'column2']  # Replace with actual column names from your dataset

for col in categorical_columns:

    print(f"\nUnique values in {col}:")

    print(df[col].unique())

#Cleaning the dataset
```

```
# Check for missing values
Missing_values = df.isnull().sum()
print(missing_values)
# Handle missing values
Df = df.dropna()  # Drop rows with missing values
# Verify if missing values have been handled
Missing_values_after_handling = df.isnull().sum()
print(missing_values_after_handling)
Df = df.drop_duplicates()
# Verify if duplicate rows have been removed
Duplicates_removed = df.duplicated().sum()
print(duplicates_removed)
# Convert columns to appropriate data types (if needed)
Df['date'] = pd.to_datetime(df['date'])
# Remove unnecessary columns (if any)
Df = df.drop(['column_name'], axis=1)
# Perform other cleaning and transformation operations as required
Df.to_csv('cleaned_dataset.csv', index=False)
#Perform data analysis
import matplotlib.pyplot as plt
# Perform data analysis tasks here...
# For example, you can calculate statistics, visualize data, etc.
# Calculate descriptive statistics
Statistics = df.describe()
print(statistics)
# Visualize data using a line plot
plt.plot(df['Date'], df['Price'])
plt.xlabel('Date')
```

```
plt.ylabel('Price')
plt.title('Electricity Price Over Time')
plt.show()
```