CO2 Emissions and Energy Consumption Analysis

This notebook provides a comprehensive analysis of CO2 emissions and energy consumption across five major countries over the period of 10 years (2010-2019).

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import f_oneway
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

Data Creation

Generating a sample dataset for the analysis.

In [ ]:
```
data = {
    "Country": ["USA", "China", "India", "Russia", "Japan"] * 10,
    "Year": np.tile(np.arange(2010, 2020), 5),
    "CO2_Emissions": np.random.randint(4000, 10000, 50),
    "Energy_Consumption": np.random.randint(500, 2000, 50)
}
df = pd.DataFrame(data)
df.head()
```

Descriptive Statistics

Calculating summary statistics for the dataset.

In [ ]:
```
df.describe()
```

In [ ]:
```
df.groupby('Country').describe()
```

Trend Analysis

Visualizing the trends in CO2 emissions and energy consumption over time.

In [ ]:
```
plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='Year', y='CO2_Emissions', hue='Country', marker='o', palette='tab10')
plt.title('Trend of CO2 Emissions (2010-2019)')
plt.ylabel('CO2 Emissions (Million Metric Tons)')
plt.xlabel('Year')
plt.legend(title='Country')
plt.grid(True)
plt.show()
```

In [ ]:
```
plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='Year', y='Energy_Consumption', hue='Country', marker='o', palette='tab10')
plt.title('Trend of Energy Consumption (2010-2019)')
plt.ylabel('Energy Consumption (Million Tonnes of Oil Equivalent)')
plt.xlabel('Year')
plt.legend(title='Country')
plt.grid(True)
plt.show()
```

Correlation Analysis

Examining the relationship between energy consumption and CO2 emissions.

In [ ]:
```
correlation = df[['Energy_Consumption', 'CO2_Emissions']].corr().iloc[0, 1]
correlation
```

In [ ]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='Energy_Consumption', y='CO2_Emissions', hue='Country', style='Country', s=100, palette='tab10')
plt.title('Correlation between Energy Consumption and CO2 Emissions')
plt.xlabel('Energy Consumption (Million Tonnes of Oil Equivalent)')
plt.ylabel('CO2 Emissions (Million Metric Tons)')
plt.grid(True)
plt.show()
```

Inferential Analysis: ANOVA Test

Performing ANOVA to check for significant differences in CO2 emissions among countries.

In [ ]:
```
anova_result = f_oneway(
    df[df['Country'] == 'China']['CO2_Emissions'],
    df[df['Country'] == 'USA']['CO2_Emissions'],
    df[df['Country'] == 'India']['CO2_Emissions'],
    df[df['Country'] == 'Russia']['CO2_Emissions'],
    df[df['Country'] == 'Japan']['CO2_Emissions']
)
anova_result
```

Predictive Analysis: Linear Regression

Building a linear regression model to predict CO2 emissions based on energy consumption.

In [ ]:
```
X = df[['Energy_Consumption']]
y = df['CO2_Emissions']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)

r2, mae, mse
```

Additional Visualizations

Exploring the distribution of CO2 emissions and comparing them across countries.

In [ ]:
```
plt.figure(figsize=(12, 6))
sns.boxplot(data=df, x='Country', y='CO2_Emissions', palette='tab10')
plt.title('Boxplot of CO2 Emissions by Country')
plt.xlabel('Country')
plt.ylabel('CO2 Emissions (Million Metric Tons)')
plt.grid(True)
plt.show()
```

In [ ]:
```
plt.figure(figsize=(12, 6))
sns.histplot(df['CO2_Emissions'], kde=True, bins=15, color='blue')
plt.title('Distribution of CO2 Emissions')
plt.xlabel('CO2 Emissions (Million Metric Tons)')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```

Summary, Findings, and Recommendations

Concluding the analysis with key insights and actionable recommendations.

- **Trends:** The trend analysis showed variability in CO2 emissions and energy consumption across different countries, with some showing increasing trends while others remained stable.
- **Correlation:** The correlation between energy consumption and CO2 emissions was found to be very weak, indicating that other factors may also play significant roles in CO2 emissions.
- **ANOVA Test:** The ANOVA test did not find statistically significant differences in CO2 emissions among the five countries, suggesting that emissions levels are relatively comparable across these nations.
- **Predictive Modeling:** The simple linear regression model performed poorly, indicating that energy consumption alone is not a strong predictor of CO2 emissions. This suggests that a more complex model incorporating other variables may be needed.

Recommendations:

- Incorporate additional variables such as industrial output, energy efficiency, and renewable energy adoption into the analysis.
- Explore more sophisticated predictive models such as multiple regression, decision trees, or machine learning algorithms.
- Conduct a deeper dive into specific countries to understand the unique factors driving their CO2 emissions.