Experiment: Linear Regression for Ozone Hole Forecasting

Introduction

This experiment aims to develop a linear regression model for forecasting the ozone hole area using historical data.

Data Loading

The dataset is loaded and inspected to understand its structure.

```python
import pandas as pd

# Load the dataset
file_path = "OzoneHole_Data.csv"  # Update the file path if necessary
df = pd.read_csv(file_path)

# Display basic information
print(df.info())
print(df.head())
```

Explanation: This segment reads the dataset and prints basic information, helping us verify the data's integrity.

Data Preprocessing

We extract relevant features and split the dataset for training and testing.

```python
from sklearn.model_selection import train_test_split

# Extract features and target variable
X = df[['Year']]
y = df['Hole Area']

# Split the data into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Explanation: The data is split to ensure a fair evaluation of the model.

Model Training

We initialize and train a linear regression model.

```python
from sklearn.linear_model import LinearRegression

# Initialize and train the model
model = LinearRegression()
model.fit(X_train, y_train)
```

Explanation: This step trains the linear regression model to predict the ozone hole area.

Model Evaluation

We test the model and evaluate its performance.

```python
from sklearn.metrics import mean_absolute_error, mean_squared_error

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = mse ** 0.5

# Print evaluation metrics
print(f"Mean Absolute Error (MAE): {mae}")
print(f"Root Mean Squared Error (RMSE): {rmse}")
```

Explanation: The model's performance is assessed using MAE and RMSE, which indicate prediction accuracy.

Visualization

We plot actual vs. predicted data for better understanding.

```python
import matplotlib.pyplot as plt
```

```
import seaborn as sns

plt.figure(figsize=(10, 5))
sns.scatterplot(x=df["Year"], y=df["Hole Area"], label="Actual Data")
sns.lineplot(x=df["Year"], y=model.predict(X), color="red", label="Regression Line")
plt.xlabel("Year")
plt.ylabel("Ozone Hole Area (Million sq. km)")
plt.title("Linear Regression for Ozone Hole Forecasting")
plt.legend()
plt.show()
```

Diagram: The plot shows how well the model's predictions align with actual data.

Future Predictions

We use the model to predict ozone hole area for future years.

```python
# Future prediction example
future_years = pd.DataFrame({'Year': range(2025, 2035)})
future_predictions = model.predict(future_years)

print("Future Predictions:")
print(pd.DataFrame({'Year': future_years['Year'], 'Predicted Hole Area':
future_predictions}))
```

Explanation: This step extends predictions beyond the available data, helping anticipate trends.

Conclusion

The **Linear Regression for Ozone Hole Forecasting** experiment has been successfully completed. The model demonstrates reasonable accuracy in predicting ozone hole area, and future predictions can provide insights into upcoming trends.