

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics

# Load the dataset
dataset = pd.read_csv('/home/avcoe/temperatures.csv')

# Data exploration
print(dataset.shape)
print(dataset.describe())

# Data visualization
dataset.plot(x='JAN', y='FEB', style='o', title='JAN vs FEB', xlabel='mintemp', ylabel='maxtemp')
plt.show()

plt.figure(figsize=(15, 10))
sns.distplot(dataset['FEB'])
plt.show()

# Data preparation
X, y = dataset['JAN'].values.reshape(-1, 1), dataset['FEB'].values.reshape(-1, 1)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

# Model training
model = LinearRegression().fit(X_train, y_train)
```

```
# Model evaluation
```

```
y_pred = model.predict(X_test)
```

```
df = pd.DataFrame({'Actual': y_test.flatten(), 'Predicted': y_pred.flatten()})
```

```
print(df.head(25))
```

```
# Visualization of results
```

```
df.head(25).plot(kind='bar', figsize=(16, 10))
```

```
plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
```

```
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
```

```
plt.show()
```

```
plt.scatter(X_test, y_test, color='gray')
```

```
plt.plot(X_test, y_pred, color='red', linewidth=2)
```

```
plt.show()
```

```
# Model evaluation metrics
```

```
print('Mean absolute error:', metrics.mean_absolute_error(y_test, y_pred))
```

```
print('Mean squared error:', metrics.mean_squared_error(y_test, y_pred))
```

```
print('Root mean squared error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```