```python
1  from sklearn import preprocessing
2  from sklearn.model_selection import train_test_split
3  from sklearn import linear_model
4  import quandl
5  import numpy as np
```

```python
1  def prepare_data(df,forecast_col,forecast_out,test_size):
2      label = df[forecast_col].shift(-forecast_out);#creating new column called label wi
3      X = np.array(df[[forecast_col]]); #creating the feature array
4      X = preprocessing.scale(X) #processing the feature array
5      X_lately = X[-forecast_out:] #creating the column i want to use later in the predi
6      X = X[:-forecast_out] # X that will contain the training and testing
7      label.dropna(inplace=True); #dropping na values
8      y = np.array(label)  # assigning Y
9      X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size=test_size) #cr
10
11     response = [X_train,X_test , Y_train, Y_test , X_lately];
12     return response;
```

```python
1  df = quandl.get('WIKI/AMZN')
```

```python
1  df.head()
```

	Open	High	Low	Close	Volume	Ex-Dividend	Split Ratio	Adj. Open	Adj. High	
Date										
1997-05-16	22.38	23.75	20.50	20.75	1225000.0	0.0	1.0	1.865000	1.979167	1.7(
1997-05-19	20.50	21.25	19.50	20.50	508900.0	0.0	1.0	1.708333	1.770833	1.62
1997-05-20	20.75	21.00	19.63	19.63	455600.0	0.0	1.0	1.729167	1.750000	1.63

```python
1  forecast_col = 'Close'#choosing which column to forecast
2  forecast_out = 5 #how far to forecast
3  test_size = 0.2; #the size of my test set
4
5  X_train, X_test, Y_train, Y_test , X_lately =prepare_data(df,forecast_col,forecast_out
6
7  learner = linear_model.LinearRegression(); #initializing linear regression model
8
9  learner.fit(X_train,Y_train); #training the linear regression model
10 score=learner.score(X_test,Y_test);#testing the linear regression model
11
12 forecast= learner.predict(X_lately); #set that will contain the forecasted data
13
14 response={};#creting json object
15 response['test_score']=score;
16 response['forecast_set']=forecast;
17
18 print(response);
```

```
{'test_score': 0.9966751954766299, 'forecast_set': array([1595.31891496, 1557.23093
        1509.77226061])}
```

```python
1
```