

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
```

Step 1: Load and preprocess the data

```
data = pd.read_csv('/content/superstore_dataset2011-2015.csv', encoding = 'ISO-8859-1')
data
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	
0	42433	AG-2011-2040	1/1/2011	6/1/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Co
1	22253	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	JH-15985	Joseph Holt	Consumer	
2	48883	HU-2011-1220	1/1/2011	5/1/2011	Second Class	AT-735	Annie Thurman	Consumer	E
3	11731	IT-2011-3647632	1/1/2011	5/1/2011	Second Class	EM-14140	Eugene Moren	Home Office	S
4	22255	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	JH-15985	Joseph Holt	Consumer	
...	
51285	32593	CA-2014-115427	31-12-2014	4/1/2015	Standard Class	EB-13975	Erica Bern	Corporate	
51286	47594	MO-2014-2560	31-12-2014	5/1/2015	Standard Class	LP-7095	Liz Preis	Consumer	
51287	8857	MX-2014-110527	31-12-2014	2/1/2015	Second Class	CM-12190	Charlotte Melton	Consumer	I
51288	6852	MX-2014-114783	31-12-2014	6/1/2015	Standard Class	TD-20995	Tamara Dahlen	Consumer	
51289	36388	CA-2014-156720	31-12-2014	4/1/2015	Standard Class	JM-15580	Jill Matthias	Consumer	I

51290 rows × 24 columns



```
# Verify the column names in the dataset
print(data.columns)
```

```
Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'City', 'State', 'Country',
       'Postal Code', 'Market', 'Region', 'Product ID', 'Category',
       'Sub-Category', 'Product Name', 'Sales', 'Quantity', 'Discount',
       'Profit', 'Shipping Cost', 'Order Priority'],
      dtype='object')
```

```
# Extract relevant features
features = ['Quantity', 'Discount', 'Profit']
X = data[features]
y = data['Sales']
```

```
# Step 2: Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Step 3: Train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)
```

```
▼ LinearRegression
LinearRegression()
```

```
# Step 4: Make predictions on the testing set
y_pred = model.predict(X_test)
```

```
# Step 5: Evaluate the model's performance
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
```

```
print('Mean Squared Error:', mse)
print('Mean Absolute Error:', mae)
```

```
Mean Squared Error: 235030.11272933218
Mean Absolute Error: 208.39740720539112
```

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 3:05 PM

