```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LinearRegression
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv("/content/drive/MyDrive/StudentsPerformance.csv")
print("First 5 Rows of Dataset:")
print(df.head())
```

```
First 5 Rows of Dataset:
   gender race/ethnicity parental level of education         lunch  \
0  female        group B           bachelor's degree      standard
1  female        group C                some college      standard
2  female        group B            master's degree      standard
3    male        group A          associate's degree  free/reduced
4    male        group C                some college      standard

  test preparation course  math score  reading score  writing score
0                    none          72             72             74
1               completed          69             90             88
2                    none          90             95             93
3                    none          47             57             44
4                    none          76             78             75
```

```
df = df.dropna()

gender_enc = LabelEncoder()
race_enc = LabelEncoder()
edu_enc = LabelEncoder()

df["gender_enc"] = gender_enc.fit_transform(df["gender"])
df["race_enc"] = race_enc.fit_transform(df["race/ethnicity"])
df["parentEdu_enc"] = edu_enc.fit_transform(df["parental level of
education"])

X = df[["math score", "reading score", "gender_enc"]]
y = df["writing score"]

LR = LinearRegression()
LR.fit(X, y)
```

```
LinearRegression()
```

```
example_prediction = LR.predict([[72, 70, 1]])  # sample input
print("\nPredicted Writing Score:", example_prediction[0])
```

```
Predicted Writing Score: 67.45975972119355
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/
validation.py:2739: UserWarning: X does not have valid feature names,
but LinearRegression was fitted with feature names
  warnings.warn(
```

```
sns.set(style="whitegrid")

sns.lmplot(
    x="reading score",
    y="writing score",
    hue="gender",
    data=df,
    height=6,
    aspect=1.5,
    scatter_kws={"s":60},
    line_kws={"lw":3}
)

plt.title("Writing Score vs Reading Score (Regression Line by
Gender)")
plt.xlabel("Reading Score")
plt.ylabel("Writing Score")
plt.show()
```