

```

# Importing necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import seaborn as sns
import matplotlib.pyplot as plt

# Fixing the issue by ensuring proper data splitting and aligning
# residual computation
# Extract X and y for the model
data = pd.read_csv('LRA.csv')

# Target variable and predictors
y = data['Earnings_1978']
X = data.drop(columns=['Earnings_1978'])

# Convert categorical variables to dummy variables
categorical_cols = ['Education', 'Race', 'Hisp', 'MaritalStatus']
X = pd.get_dummies(X, columns=categorical_cols, drop_first=True)

# Splitting 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)

# Train the Linear Regression model
lr_model = LinearRegression()
lr_model.fit(X_train, y_train)

LinearRegression()

# Predictions and residual calculations
y_train_pred = lr_model.predict(X_train)
residuals = y_train - y_train_pred

# Predictions on test data for evaluation
y_test_pred = lr_model.predict(X_test)

# Verify updates for training and residuals
mse = mean_squared_error(y_test, y_test_pred)
r2 = r2_score(y_test, y_test_pred)

mse, r2, residuals[:5] # Display key metrics and residuals sample
(48625764.00113674,
 0.4767136132388897,
 4561    2480.397578
 921     -1242.364809
 8784    -1567.146627
 1182    -322.548334

```

```
3910      5149.904575  
Name: Earnings_1978, dtype: float64)
```