```
[12] 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, r2_score, root_mean_squared_error
     from sklearn.preprocessing import LabelEncoder
```

```
[2]  data = pd.read_csv("data.csv")
```

```
[3]  print(data.head())
```

```
   Age          Eduacation      Race         Hisp MaritalStatus  Nodeg  \
0   45  LessThanHighSchool  NotBlack  NotHispanic       Married      1
1   21        Intermediate  NotBlack  NotHispanic    NotMarried      0
2   38          HighSchool  NotBlack  NotHispanic       Married      0
3   48  LessThanHighSchool  NotBlack  NotHispanic       Married      1
4   18  LessThanHighSchool  NotBlack  NotHispanic       Married      1

   Earnings_1974  Earnings_1975  Earnings_1978
0      21516.670      25243.550      25564.670
1       3175.971       5852.565      13496.080
2      23039.020      25130.760      25564.670
3      24994.370      25243.550      25564.670
4       1669.295      10727.610       9860.869
```

```
[4]  label_encoder = LabelEncoder()
     data['Education'] = label_encoder.fit_transform(data['Eduacation'])
     data['Race'] = label_encoder.fit_transform(data['Race'])
     data['Hisp'] = label_encoder.fit_transform(data['Hisp'])
     data['MaritalStatus'] = label_encoder.fit_transform(data['MaritalStatus'])
```

```
[5]  X = data[['Age', 'Education', 'Race', 'Hisp', 'MaritalStatus', 'Nodeg', 'Earnings_1974', 'Earnings_1975']]
     y = data['Earnings_1978']
```

```
[6]  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
[7]  model = LinearRegression()
     model.fit(X_train, y_train)
```

```
LinearRegression
LinearRegression()
```

```
[8]  y_pred = model.predict(X_test)
```

```
[14] mse = mean_squared_error(y_test, y_pred)
     rmse=root_mean_squared_error(y_test,y_pred)
     r2 = r2_score(y_test, y_pred)
```

```
[15] print(f"Mean Squared Error: {mse}")
     print(f"Root Mean Squared Error:{rmse}")
     print(f"R-squared: {r2}")
```

```
Mean Squared Error: 48625781.822785094
Root Mean Squared Error:6973.218899675034
R-squared: 0.4767134214511377
```