

zad1

June 8, 2024

```
[1]: 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
```

```
[2]: # Load the dataset from CSV
      df = pd.read_csv("my_data.csv")

      # One-hot encoding for categorical variable
      df_encoded = pd.get_dummies(df, columns=["Category"], drop_first=True)
```

```
[3]: # Prepare the data
      X = df_encoded.drop(columns=["Variable1", "Variable2"])
      y = df_encoded[["Variable1", "Variable2"]]
```

```
[5]: # 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
      )

      # Train the model
      model = LinearRegression()
      model.fit(X_train, y_train)
```

```
[5]: LinearRegression()
```

```
[7]: # Make predictions
      y_pred = model.predict(X_test)
      y_pred
```

```
[7]: array([[16.5 , 32.5 ],
            [15.75, 31.5 ],
            [20.  , 38.  ],
            [16.25, 31.  ],
            [17.  , 25.5 ]])
```

```
[8]: # Evaluate the model
      mse = mean_squared_error(y_test, y_pred)
```

```
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("R-squared Score:", r2)
```

Mean Squared Error: 35.61250000000001

R-squared Score: -0.4972935517549968

[]: