# Breakout Activity

**1. Data Loading and Preprocessing**

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

df = pd.read_csv('fitness exercises.csv')

# Display the first few rows of the dataframe to understand its structure
print("First few rows of the dataset:")
print(df.head())
```

```
First few rows of the dataset:
        Weight     Height         BMI  Gender  Age     BMIcase  \
0    92.085190   1.760250   29.719488  Female   59  over weight
1    61.089124   1.595499   23.997776  Female   25       normal
2    82.454037   1.816538   24.987499  Female   50       normal
3   101.713306   1.790696   31.720047  Female   62        obese
4    99.609527   1.969726   25.673756    Male   57  over weight

   Exercise Recommendation Plan
0                             5
1                             4
2                             4
3                             6
4                             5
```

```python
# Select features and target variable
X = df[['Weight', 'Height', 'BMI', 'Age']]  # Features
y = df['Exercise Recommendation Plan']  # Target variable
```

```python
# Convert categorical variables to numerical (e.g., Gender)
X = pd.get_dummies(X.join(df['Gender']), drop_first=True)

# 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)

# Standardize the features (optional, but recommended for many models)
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

## 2. Model Training

```python
from sklearn.linear_model import LinearRegression

# Initialize and train the Linear Regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)
```

## 3. Evaluation Using Mean Squared Error (MSE)

```python
from sklearn.metrics import mean_squared_error

# Calculate MSE
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse:.2f}')
```

```
Mean Squared Error: 0.31
```

### Reflection

The aim of the project is to predict an individual's exercise recommendation plan based on features such as weight, height, BMI, age, and gender. This prediction helps in providing tailored fitness plans.

The first we need to perform pre-processing steps, including converting the categorical Gender feature into numerical values (using one-hot encoding) and standardizing the features to ensure consistency across different scales.

Chose Linear Regression as the model due to the continuous nature of the target variable, Exercise Recommendation Plan. This model is a good starting point, as it is simple and interpretable.

The Mean Squared Error (MSE) of the model was 0.31, which indicates how far the predicted values are from the actual values on average. The lower the MSE, the better the model fits the data. The performance seems reasonable, but improvements could be explored.