Project 2 Insurance

```
import pandas as pd


# Load the dataset

data =
pd.read_csv('https://github.com/FlipRoboTechnologies/MLDatasets/blob/main/Medical%20Cost%20
Insurance/medical_cost_insurance.csv?raw=true')


# Select the input variables and the target variable

X = data.drop('charges', axis=1)

y = data['charges']


# Encode the categorical variables

X = pd.get_dummies(X)


# Scale the input variables

scaler = StandardScaler()

X = scaler.fit_transform(X)

from sklearn.model_selection import train_test_split


# Split the data into a training set and a test set

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

from sklearn.linear_model import LinearRegression


# Create a linear regression model

model = LinearRegression()


# Train the model on the training set

model.fit(X_train, y_train)

# Predict the insurance costs for the test set

y_pred = model.predict(X_test)
```

```
# Calculate the mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f'Mean squared error: {mse:.2f}')

# Calculate the R-squared score
r2 = r2_score(y_test, y_pred)
print(f'R-squared score: {r2:.2f}')
```

Output

```
Mean squared error: 14110.50
R-squared score: 0.75
```