Assignment 2 (B)

```
import pandas as pd
import math
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error

# Load dataset
df = pd.read_csv("CarPrice_Assignment.csv")

# Encode categorical variables
df = pd.get_dummies(df, columns=['fueltype'])
df.drop("fueltype_gas", axis=1, inplace=True)  # drop redundant dummy column

label_encoder = preprocessing.LabelEncoder()
df["enginetype"] = label_encoder.fit_transform(df["enginetype"])
df["carbody"] = label_encoder.fit_transform(df["carbody"])

# Define features (X) and target (Y)
X = df[["horsepower", "fueltype_diesel", "enginesize", "enginetype", "carbody"]]
Y = df[["price"]]

# Train-test split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3, random_state=42)

# Train model
model = LinearRegression()
model.fit(X_train, Y_train)

# Predictions
y_pred = model.predict(X_test)

# Evaluation metrics
print('Mean Squared Error      : ', mean_squared_error(Y_test, y_pred))
print('Mean Absolute Error     : ', mean_absolute_error(Y_test, y_pred))
print('Root Mean Squared Error : ', math.sqrt(mean_squared_error(Y_test, y_pred)))
```