

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Sample car dataset
data = {
    "Year": [2010, 2012, 2015, 2018, 2020, 2011, 2016, 2019],
    "Mileage": [80000, 60000, 40000, 30000, 20000, 75000, 35000,
25000],
    "EngineSize": [1.2, 1.5, 1.6, 2.0, 2.2, 1.3, 1.8, 2.0],
    "Price": [300000, 400000, 550000, 800000, 1000000, 320000, 700000,
900000]
}

df = pd.DataFrame(data)
df

{"summary": "{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n    {\n      \"column\": \"Year\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 3,\n        \"min\": 2010,\n        \"max\": 2020,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          2012,\n          2011,\n          2010\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Mileage\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 23058,\n        \"min\": 20000,\n        \"max\": 80000,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          60000,\n          75000,\n          80000\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"EngineSize\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0.35856858280031817,\n        \"min\": 1.2,\n        \"max\": 2.2,\n        \"num_unique_values\": 7,\n        \"samples\": [\n          1.2,\n          1.5,\n          1.3\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Price\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 269201,\n        \"min\": 300000,\n        \"max\": 1000000,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          400000,\n          320000,\n          300000\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    }\n  ]\n}", "type": "dataframe", "variable_name": "df"}

X = df[["Year", "Mileage", "EngineSize"]]
y = df["Price"]

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.25, random_state=42
)

```

```
model = LinearRegression()
model.fit(X_train, y_train)

LinearRegression()

y_pred = model.predict(X_test)

print("Mean Squared Error:", mean_squared_error(y_test, y_pred))
print("R2 Score:", r2_score(y_test, y_pred))

Mean Squared Error: 1741315806.092752
R2 Score: -0.08832237880796989

# New car details
new_car = pd.DataFrame(
    [[2017, 45000, 1.8]],
    columns=["Year", "Mileage", "EngineSize"]
)

predicted_price = model.predict(new_car)
print("Predicted Car Price:", int(predicted_price[0]))

Predicted Car Price: 777966
```