```
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 house price dataset
data = {
    "Area": [800, 1000, 1200, 1500, 1800, 2000, 2200, 2500],
    "Bedrooms": [2, 2, 3, 3, 4, 4, 4, 5],
    "Bathrooms": [1, 2, 2, 2, 3, 3, 3, 4],
    "Age": [20, 15, 10, 8, 5, 3, 2, 1],
    "Price": [2000000, 2800000, 3500000, 4500000,
              5500000, 6500000, 7200000, 8500000]
}

df = pd.DataFrame(data)
df
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"Area\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 601,\n        \"min\": 800,\n
\"max\": 2500,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          1000,\n          2000,\n          800\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Bedrooms\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
1,\n        \"min\": 2,\n        \"max\": 5,\n
\"num_unique_values\": 4,\n        \"samples\": [\n          3,\n
5,\n          2\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Bathrooms\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0,\n        \"min\": 1,\n
\"max\": 4,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          2,\n          4,\n          1\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 6,\n        \"min\": 1,\n
\"max\": 20,\n        \"num_unique_values\": 8,\n        \"samples\":
[\n          15,\n          3,\n          20\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Price\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 2264595,\n
\"min\": 2000000,\n        \"max\": 8500000,\n
\"num_unique_values\": 8,\n        \"samples\": [\n          2800000,\
n          6500000,\n          2000000\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```

```
X = df[["Area", "Bedrooms", "Bathrooms", "Age"]]
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: 8712087697.347504
R2 Score: 0.9974544667063996
```

```
# New house details
new_house = pd.DataFrame(
    [[1600, 3, 2, 6]],
    columns=["Area", "Bedrooms", "Bathrooms", "Age"]
)

predicted_price = model.predict(new_house)
print("Predicted House Price:", int(predicted_price[0]))
```

```
Predicted House Price: 4856177
```