```python
import pandas as pd, numpy as np, matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

df = pd.read_csv('/content/housing.csv')

# Drop rows with any missing values
df.dropna(inplace=True)

print(df.head())
print(df.info())

# Separate features (X) and target (y)
X = df.drop(columns=['median_house_value'])
y = df['median_house_value']

# Apply one-hot encoding to 'ocean_proximity' in X
X = pd.get_dummies(X, columns=['ocean_proximity'], drop_first=True)

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

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

y_pred = model.predict(X_test)
print("MSE:", mean_squared_error(y_test, y_pred))

plt.figure()
plt.scatter(y_test, y_pred)
plt.xlabel("Actual")
plt.ylabel("Predicted")
plt.show()
```

```
       longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
0       -122.23     37.88                41.0        880.0           129.0
1       -122.22     37.86                21.0       7099.0          1106.0
2       -122.24     37.85                52.0       1467.0           190.0
3       -122.25     37.85                52.0       1274.0           235.0
4       -122.25     37.85                52.0       1627.0           280.0

       population  households  median_income  median_house_value ocean_proximity
0          322.0       126.0         8.3252            452600.0        NEAR BAY
1         2401.0      1138.0         8.3014            358500.0        NEAR BAY
2          496.0       177.0         7.2574            352100.0        NEAR BAY
3          558.0       219.0         5.6431            341300.0        NEAR BAY
```

Start coding or generate with AI.

```
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   longitude           20433 non-null  float64
 1   latitude            20433 non-null  float64
 2   housing_median_age  20433 non-null  float64
 3   total_rooms         20433 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20433 non-null  float64
 6   households          20433 non-null  float64
 7   median_income       20433 non-null  float64
 8   median_house_value  20433 non-null  float64
 9   ocean_proximity     20433 non-null  object
dtypes: float64(9), object(1)
memory usage: 1.7+ MB
None
MSE: 4802173538.60416
```