

```

# =====
# Task 1: Linear Regression - California Housing
# Google Colab | Phone Safe | ERROR-FREE
# =====

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.datasets import fetch_openml
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# 1. Load dataset from OpenML (SAFE)
housing = fetch_openml(name="california_housing", version=1, as_frame=True)
df = housing.frame

print("First 5 rows:")
print(df.head())

# 2. Handle missing values
df["total_bedrooms"] = df["total_bedrooms"].fillna(df["total_bedrooms"].median())

# 3. Encode categorical column
df = pd.get_dummies(df, drop_first=True)

# 4. Split features and target
X = df.drop("median_house_value", axis=1)
y = df["median_house_value"]

# 5. Train-test split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# 6. Train model
model = LinearRegression()
model.fit(X_train, y_train)

# 7. Predictions
y_pred = model.predict(X_test)

# 8. Evaluation
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print("\nModel Performance:")
print(f"MAE : {mae}")
print(f"RMSE : {rmse}")
print(f"R² : {r2}")

# 9. Visualization
plt.scatter(y_test, y_pred, alpha=0.5)
plt.xlabel("Actual House Values")
plt.ylabel("Predicted House Values")
plt.title("Actual vs Predicted House Prices")

```

```
plt.show()
```

```
# 10. Conclusion
```

```
print("""
```

```
Conclusion:
```

```
Linear Regression provides a solid baseline model for predicting  
California house prices. Performance can be improved using  
regularization or ensemble models.
```

```
""")
```

First 5 rows:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms
0	-122.23	37.88	41	880	129.0
1	-122.22	37.86	21	7099	1106.0
2	-122.24	37.85	52	1467	190.0
3	-122.25	37.85	52	1274	235.0
4	-122.25	37.85	52	1627	280.0

	population	households	median_income	median_house_value	ocean_proxim
0	322	126	8.3252	452600	NEAR
1	2401	1138	8.3014	358500	NEAR
2	496	177	7.2574	352100	NEAR
3	558	219	5.6431	341300	NEAR
4	565	259	3.8462	342200	NEAR

Model Performance:

MAE : 50670.73824097194

RMSE : 70060.5218447353

R<sup>2</sup> : 0.6254240620553595

Actual vs Predicted House Prices

