House Price Prediction Using Machine learning

BATCH MEMBERS

Matchimai.K.E (matchierumporai@gmail.com)

preethi R (preethipriya2128@gmail.com)

mirudula V (mirudula2503@gmail.com)

parkavi K.S (parkavisathya982@gmail.com)

phase 3 submission document

project Title ： House price prediction

Topic : Start building the house price prediction model by loading and pre-processing the dataset

House Price Prediction

Introduction:

⬚ Whether you're a homeowner looking to estimate the value of your property, a real estate investor seeking profitable opportunities, or a data scientist aiming to build a predictive model, the foundation of this endeavor lies in loading and preprocessing the dataset.

⬚ Building a house price prediction model is a data-driven process that involves harnessing the power of machine learning to analyze historical housing data and make informed price predictions. This journey begins with the fundamental steps of data loading and preprocessing.

⬚ This introduction will guide you through the initial steps of the process. We'll explore how to import essential libraries, load the housing dataset, and perform critical preprocessing steps. Data preprocessing is crucial as it helps clean, format, and prepare the data for further analysis. This includes handling missing values, encoding categorical variables, and ensuring that the data is appropriately scaled.

Given data set :

Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
79545.45857	5.682861322	7.009188143	4.09	23086.8005	1059033.558	208 Michael Ferry Apt. 674 Laurabury, NE 37010-5101

79248.64245	6.002899808	6.730821019	3.09	40173.07217	1505890.915	188 Johnson Views Suite 079 Lake Kathleen, CA 48958
61287.06718	5.86588984	8.51272743	5.13	36882.1594	1058987.988	9127 Elizabeth Stravenue Danieltown, WI 06482-3489
63345.24005	7.188236095	5.586728665	3.26	34310.24283	1260616.807	USS Barnett FPO AP 44820
59982.19723	5.040554523	7.839387785	4.23	26354.10947	630943.4893	USNS Raymond FPO AE 09386
80175.75416	4.988407758	6.104512439	4.04	26748.42842	1068138.074	06039 Jennifer Islands Apt. 443 Tracyport, KS 16077
64698.46343	6.025335907	8.147759585	3.41	60828.24909	1502055.817	4759 Daniel Shoals Suite 442 Nguyenburgh, CO 20247
78394.33928	6.989779748	6.620477995	2.42	36516.35897	1573936.564	972 Joyce Viaduct Lake William, TN 17778-6483
59927.66081	5.36212557	6.393120981	2.3	29387.396	798869.5328	USS Gilbert FPO AA 20957
81885.92718	4.42367179	8.167688003	6.1	40149.96575	1545154.813	Unit 9446 Box 0958 DPO AE 97025
80527.47208	8.093512681	5.0427468	4.1	47224.35984	1707045.722	6368 John Motorway

						Suite 700 Janetbury, NM 26854
50593.6955	4.496512793	7.467627404	4.49	34343.99189	663732.3969	911 Castillo Park Apt. 717 Davisborough, PW 78603
39033.80924	7.671755373	7.250029317	3.1	39220.36147	1042814.098	209 Natasha Stream Suite 961 Huffmanland, NE 52457
73163.66344	6.919534825	5.993187901	2.27	32326.12314	1291331.518	829 Welch Track Apt. 992 North John, AR 26532-5136
69391.38018	5.344776177	8.406417715	4.37	35521.29403	1402818.21	PSC 5330, Box 4420 APO AP 08302
73091.86675	5.443156467	8.517512711	4.01	23929.52405	1306674.66	2278 Shannon View North Carriemouth, NM 84617
79706.96306	5.067889591	8.219771123	3.12	39717.81358	1556786.6	064 Hayley Unions Nicholsborough, HI 44161-1887
61929.07702	4.788550242	5.097009554	4.3	24595.9015	528485.2467	5498 Rachel Locks New Gregoryshire, PW 54755
63508.1943	5.94716514	7.187773835	5.12	35719.65305	1019425.937	Unit 7424 Box 2786 DPO AE 71255
62085.2764	5.739410844	7.091808104	5.49	44922.1067	1030591.429	19696 Benjamin Cape

						Stephentown, ME 36952-
86294.99909	6.62745694	8.011897853	4.07	47560.77534	2146925.34	030 Larry Park Suite 665 Thomashaven, HI 87941-5197
60835.08998	5.551221592	6.517175038	2.1	45574.74166	929247.5995	USNS Brown FPO AP 85833
64490.65027	4.21032287	5.478087731	4.31	40358.96011	718887.2315	95198 Ortiz Key Port Sara, TN 24541-2855
60697.35154	6.170484091	7.150536572	6.34	28140.96709	743999.8192	9003 Jay Plains Suite 838 Lake Elizabeth, IN 90622-0804
59748.85549	5.339339881	7.748681606	4.23	27809.98654	895737.1334	24282 Paul Valley West Perry, MI 03169-5806
56974.47654	8.287562194	7.312879971	4.33	40694.86951	1453974.506	61938 Brady Falls Lewisfort, DE 61227
82173.62608	4.018524685	6.992698757	2.03	38853.91807	1125692.507	3599 Ramirez Springs Jacksonhaven, AZ 72798
64626.88098	5.44335959	6.988753539	4	27784.74228	975429.4928	073 Christopher Falls Suite 882 West Cynthia, MA 89075-2814
90499.05745	6.384358921	4.242191302	3.04	33970.16499	1240763.766	6531 Chase Prairie Apt. 245

Susanshire.

Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
59323.7921	6.97782794	8.273697078	4.07	37520.65773	1577017.76	17124 Johnson Squares Lake Robertfurt, AL 61811-3832
77811.52203	5.31446021 5	6.68668613 3	3.24	33754.7378	1246830.18 8	1359 Mitchell Vista Courtneyfort, NY 93065-7224
68652.6092 1	6.1243421	6.29082016 7	4.42	39355.62573	1170720.894	4343 Joshua Lake Suite 096 Pierceview, NV 62614-9110
55041.3515 8	7.127128553	8.59192329 4	5.36	30122.47348	1071279.21	0057 Jacob Coves Apt. 932 Lake Emily, VA 39465-6041

Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
79545.45857	5.682861322	7.009188143	4.09	23086.8005	1059033.55 8	208 Michael Ferry Apt. 674 Laurabury, NE 37010-5101
79248.64245	6.00289980 8	6.73082101 9	3.09	40173.07217	1505890.91 5	188 Johnson Views Suite 079 Lake Kathleen, CA 48958
61287.06718	5.86588984	8.51272743	5.13	36882.1594	1058987.988	9127 Elizabeth Stravenue Danieltown,

						WI 06482-3489
63345.24005	7.188236095	5.586728665	3.26	34310.24283	1260616.807	USS Barnett FPO AP 44820
59982.19723	5.040554523	7.839387785	4.23	26354.10947	630943.4893	USNS Raymond FPO AE 09386
80175.75416	4.988407758	6.104512439	4.04	26748.42842	1068138.074	06039 Jennifer Islands Apt. 443 Tracyport, KS 16077
64698.46343	6.025335907	8.147759585	3.41	60828.24909	1502055.817	4759 Daniel Shoals Suite 442 Nguyenburgh, CO 20247
78394.33928	6.989779748	6.620477995	2.42	36516.35897	1573936.564	972 Joyce Viaduct Lake William, TN 17778-6483
59927.66081	5.36212557	6.393120981	2.3	29387.396	798869.5328	USS Gilbert FPO AA 20957
81885.92718	4.42367179	8.167688003	6.1	40149.96575	1545154.813	Unit 9446 Box 0958 DPO AE 97025
80527.47208	8.093512681	5.0427468	4.1	47224.35984	1707045.722	6368 John Motorway Suite 700 Janetbury, NM 26854
50593.6955	4.496512793	7.467627404	4.49	34343.99189	663732.3969	911 Castillo Park Apt. 717 Davisborough, PW 78603
39033.80924	7.671755373	7.250029317	3.1	39220.36147	1042814.098	209 Natasha Stream Suite 961

						Huffmanland,
73163.66344	6.919534825	5.993187901	2.27	32326.12314	1291331.518	829 Welch Track Apt. 992 North John, AR 26532-5136
69391.38018	5.344776177	8.406417715	4.37	35521.29403	1402818.21	PSC 5330, Box 4420 APO AP 08302
73091.86675	5.443156467	8.517512711	4.01	23929.52405	1306674.66	2278 Shannon View North Carriemouth, NM 84617
79706.96306	5.067889591	8.219771123	3.12	39717.81358	1556786.6	064 Hayley Unions Nicholsborough, HI 44161-1887
61929.07702	4.788550242	5.097009554	4.3	24595.9015	528485.2467	5498 Rachel Locks New Gregoryshire, PW 54755
63508.1943	5.94716514	7.187773835	5.12	35719.65305	1019425.937	Unit 7424 Box 2786 DPO AE 71255
62085.2764	5.739410844	7.091808104	5.49	44922.1067	1030591.429	19696 Benjamin Cape Stephentown, ME 36952-4733
86294.99909	6.62745694	8.011897853	4.07	47560.77534	2146925.34	030 Larry Park Suite 665 Thomashaven, HI 87941-5197
60835.08998	5.551221592	6.517175038	2.1	45574.74166	929247.5995	USNS Brown FPO AP 85833

64490.6502 7	4.21032287	5.478087731	4.31	40358.9601 1	718887.2315	95198 Ortiz Key Port Sara, TN 24541-2855
60697.3515 4	6.17048409 1	7.150536572	6.34	28140.9670 9	743999.8192	9003 Jay Plains Suite 838 Lake Elizabeth, IN 90622-0804
59748.8554 9	5.33933988 1	7.748681606	4.23	27809.9865 4	895737.1334	24282 Paul Valley West Perry, MI 03169-5806
56974.47654	8.287562194	7.312879971	4.33	40694.8695 1	1453974.50 6	61938 Brady Falls Lewisfort, DE 61227
82173.62608	4.01852468 5	6.992698757	2.03	38853.91807	1125692.507	3599 Ramirez Springs Jacksonhave n, AZ 72798
64626.8809 8	5.44335959	6.98875353 9	4	27784.74228	975429.4928	073 Christopher Falls Suite 882 West Cynthia, MA 89075-2814
90499.0574 5	6.38435892 1	4.242191302	3.04	33970.1649 9	1240763.766	6531 Chase Prairie Apt. 245 Susanshire, MN 22365
59323.7921	6.97782794	8.273697078	4.07	37520.65773	1577017.76	17124 Johnson Squares Lake Robertfurt, AL 61811-3832
77811.52203	5.31446021 5	6.68668613 3	3.24	33754.7378	1246830.18 8	1359 Mitchell Vista

							Courtneyfort, NY 93065-7224
68652.60921	6.1243421	6.290820167	4.42		39355.62573	1170720.894	4343 Joshua Lake Suite 096 Pierceview, NV 62614-9110
55041.35158	7.127128553	8.591923294	5.36		30122.47348	1071279.21	0057 Jacob Coves Apt. 932 Lake Emily, VA 39465-6041

3000 Rows *9 Columns

preprocessing code :

```python
from google.colab import drive

drive.mount("/content/drive")


 import pandas as pd

data = pd.read_csv('/content/sample_data/california_housing_test.csv')

missing_values = data.isnull().sum()

numeric_columns = data.select_dtypes(include=['number']).columns

data[numeric_columns] = data[numeric_columns].fillna(data[numeric_columns].mean())

missing_values_after = data.isnull().sum()

data.to_csv('preprocessed_dataset.csv', index=False)


import numpy as np

import pandas as pd

import os
```

```python
for dirname, _, filenames in os.walk('/content/preprocessed_dataset.csv'):

    for filename in filenames:

        print(os.path.join(dirname, filename))


import matplotlib.pyplot as plt

import seaborn as sb


data = pd.read_csv('/content/preprocessed_dataset.csv')

data.head()

data.shape

data.info()

data.isna().sum()

data.duplicated().sum()

sb.pairplot(data = data)
```

visualization