

CALIFORNIA HOUSE PRICING

WAMBUI WAMAGATA

21-2129

MAT 271 A

LIST OF CONTENTS

INTRODUCTION

LITERATURE REVIEW

RESEARCH METHODOLOGY

BACKGROUND OF STUDY

House

- Difference in the house prices
- House price prediction

House price

- Machine learning
- Factors



SCOPE

LOCATION
California

DATASET
California
house pricing

FACTORS
Ocean
proximity, Popula
tion, Households

DATASET: CALIFORNIA HOUSE PRICING

housing - Excel

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do... Share

Clipboard: Paste Cut Copy Format Painter

Font: Calibri 11 Bold Italic Underline Color Fill Background Color

Alignment: Wrap Text Merge & Center

Number: General \$ % .00 .00

Styles: Conditional Formatting Format as Table Cell Styles

Cells: Insert Delete Format

Editing: AutoSum Fill Clear Sort & Filter Find & Select

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	longitude	latitude	housing_r	total_roor	total_bed	populatio	household	median_i	median_ho	cean_proximity											
2	-122.23	37.88	41	880	129	322	126	8.3252	452600	NEAR BAY											
3	-122.22	37.86	21	7099	1106	2401	1138	8.3014	358500	NEAR BAY											
4	-122.24	37.85	52	1467	190	496	177	7.2574	352100	NEAR BAY											
5	-122.25	37.85	52	1274	235	558	219	5.6431	341300	NEAR BAY											
6	-122.25	37.85	52	1627	280	565	259	3.8462	342200	NEAR BAY											
7	-122.25	37.85	52	919	213	413	193	4.0368	269700	NEAR BAY											
8	-122.25	37.84	52	2535	489	1094	514	3.6591	299200	NEAR BAY											
9	-122.25	37.84	52	3104	687	1157	647	3.12	241400	NEAR BAY											
10	-122.26	37.84	42	2555	665	1206	595	2.0804	226700	NEAR BAY											
11	-122.25	37.84	52	3549	707	1551	714	3.6912	261100	NEAR BAY											
12	-122.26	37.85	52	2202	434	910	402	3.2031	281500	NEAR BAY											
13	-122.26	37.85	52	3503	752	1504	734	3.2705	241800	NEAR BAY											
14	-122.26	37.85	52	2491	474	1098	468	3.075	213500	NEAR BAY											
15	-122.26	37.84	52	696	191	345	174	2.6736	191300	NEAR BAY											
16	-122.26	37.85	52	2643	626	1212	620	1.9167	159200	NEAR BAY											
17	-122.26	37.85	50	1120	283	697	264	2.125	140000	NEAR BAY											
18	-122.27	37.85	52	1966	347	793	331	2.775	152500	NEAR BAY											
19	-122.27	37.85	52	1228	293	648	303	2.1202	155500	NEAR BAY											
20	-122.26	37.84	50	2239	455	990	419	1.9911	158700	NEAR BAY											
21	-122.27	37.84	52	1503	298	690	275	2.6033	162900	NEAR BAY											
22	-122.27	37.85	40	751	184	409	166	1.3578	147500	NEAR BAY											
23	-122.27	37.85	42	1639	367	929	366	1.7135	159800	NEAR BAY											
24	-122.27	37.84	52	2436	541	1015	478	1.725	113900	NEAR BAY											
25	-122.27	37.84	52	1688	337	853	325	2.1806	99700	NEAR BAY											

Ready | housing | 100%

DOMAIN

Topic	Findings
House	House is one of the most important components of one's existence because it gives shelter,safety and warmth to the owner and is a necessity for every human.
House Price	Each house has it's own price based on it's location,ocean proximity,households,,total rooms,total bedrooms and and population.
House Price Prediction	House price prediction gives a description of the features and current market price which can be used to make effective predictions and gain a better understanding of the market.This is very important to potential buyers and sellers.

FACTORS

Topic	Findings
Location	A house placed close to the ocean is expected to be of more value compared to the house that is far away from the ocean.
House Structure	One with more rooms occupies more space hence the structure varies and influences the price.

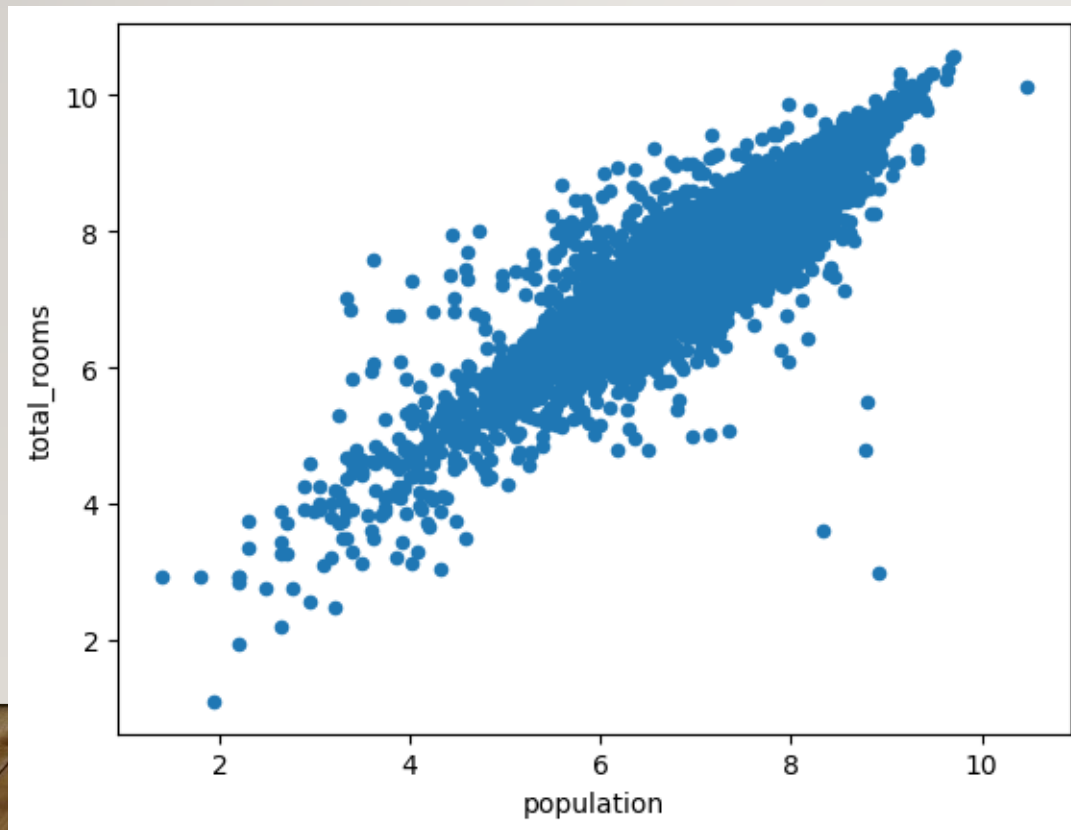
MACHINE LEARNING

Topic	Findings
Machine Learning	<p>Consists of adaptive processes that allow computers to learn from their own experiences. It has different approaches which are supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning.</p>

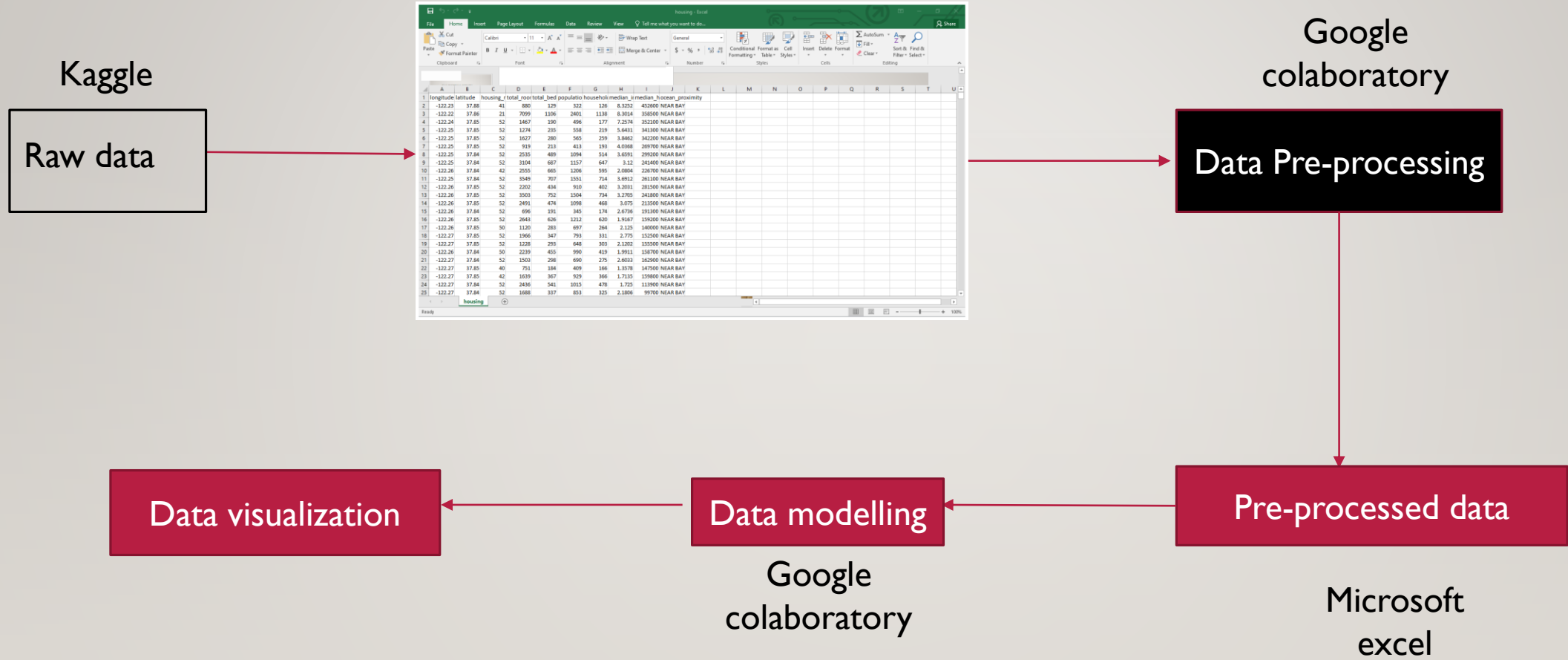
TECHNIQUE

Linear Regression is a regression analysis when a statistical procedure is used to estimate the relationship between distinct variables. It is a simple and basic way but has a degree of error that is slightly above the average. An example is as shown below

```
#checking if the data is linear  
train_data.plot(kind='scatter', x='population', y='total_rooms')  
plt.show()
```



DASHBOARD ARCHITECTURE



THE END