

# House Pricing

*Chao Wang & Ling Li  
DAMG6105 Final Project*

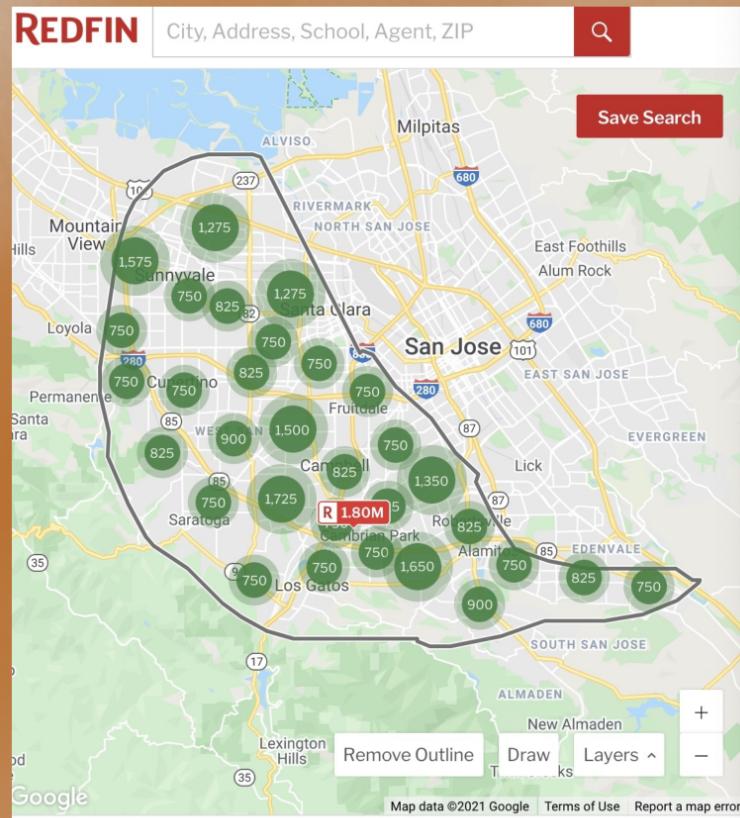


# *Introduction*

*Map*

*Influencing  
factors*

*Data*



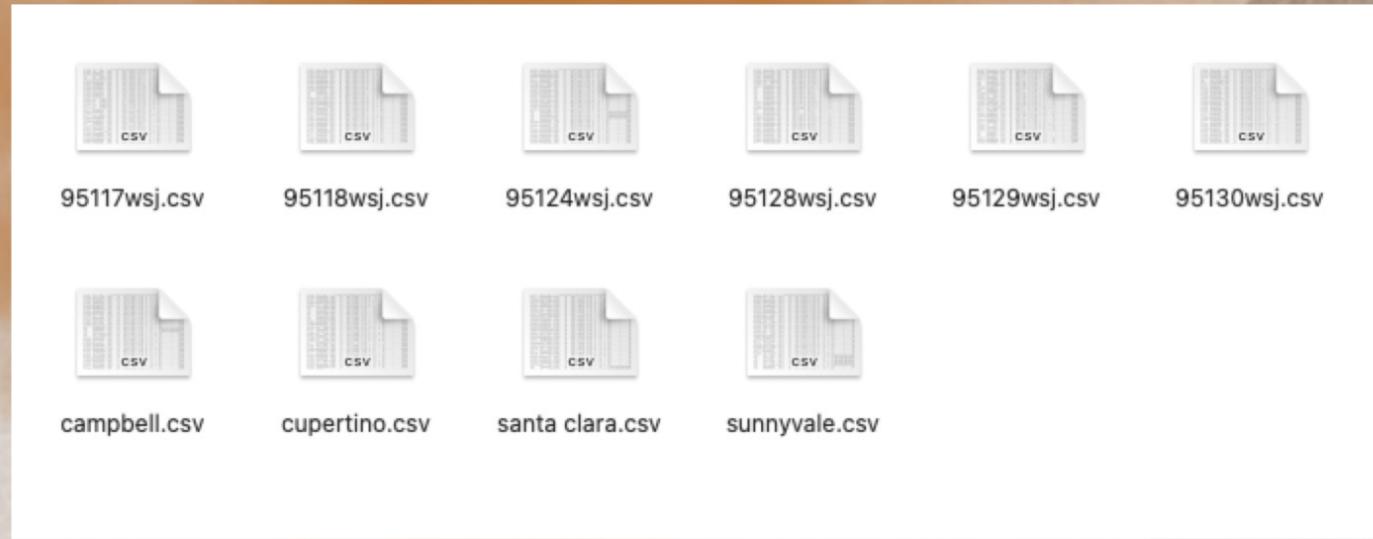
# *Introduction*

*Map*

*Influencing  
factors*

*Data*

- *West San Jose*
- *Campbell*
- *Cupertino*
- *Santa Clara*
- *Sunnyvale*



# *Introduction*

*Map*

*Influencing  
factors*

*Data*



**Property Type**

**City/ ZIP**

**Beds**

**Baths**

**Price**

**Square Feet**

**\$/Square Feet**

**HOA ( Homeowner Association )**

**Lot Size**

**Year Built**

# *Introduction*

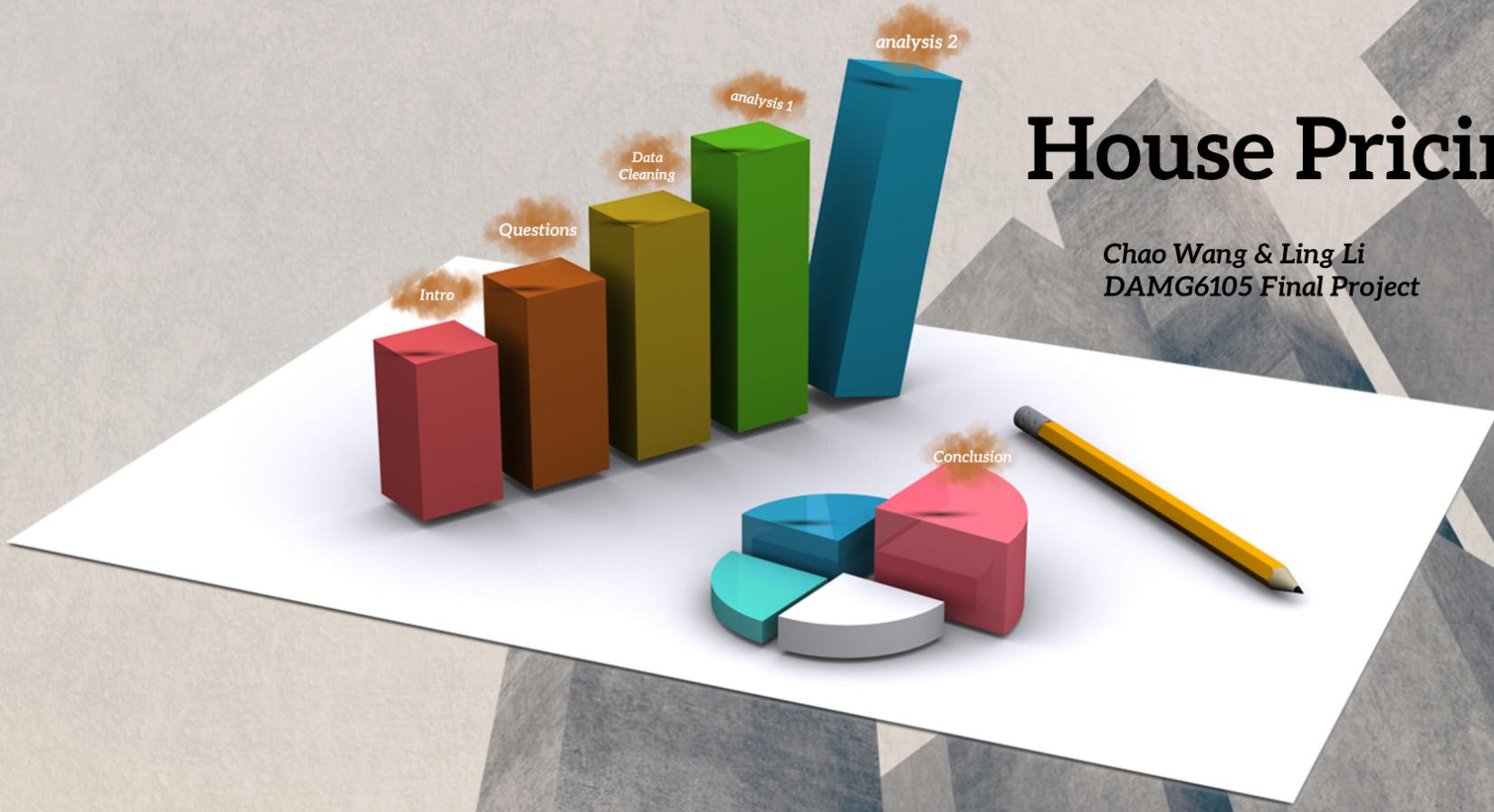
*Map*

*Influencing  
factors*

*Data*

# House Pricing

*Chao Wang & Ling Li  
DAMG6105 Final Project*





# *Research Questions*



*Self-  
attributes*

House Pricing and ?

*Covid-19*



## Influencing factors

- **Property Type**
- **City/ ZIP**
- **Beds**
- **Baths**
- **Price**
- **Square Feet**
- **\$/Square Feet**
- **HOA**
- **Lot Size**
- **Year Built**



# *Research Questions*



*Self-  
attributes*

House Pricing and ?

*Covid-19*

The background of the slide features abstract, light-grey geometric shapes resembling triangles and rectangles. A prominent, semi-transparent orange cloud or smoke effect is centered in the upper half of the slide, partially obscuring the text.

**SOLD DATE PROPERTY**

**For example:**  
**November-29-2021**



# *Research Questions*



*Self-  
attributes*

House Pricing and ?

*Covid-19*

# House Pricing

Chao Wang & Ling Li  
DAMG6105 Final Project





# Data Cleaning

1

2

3

4

5

The outlier data :

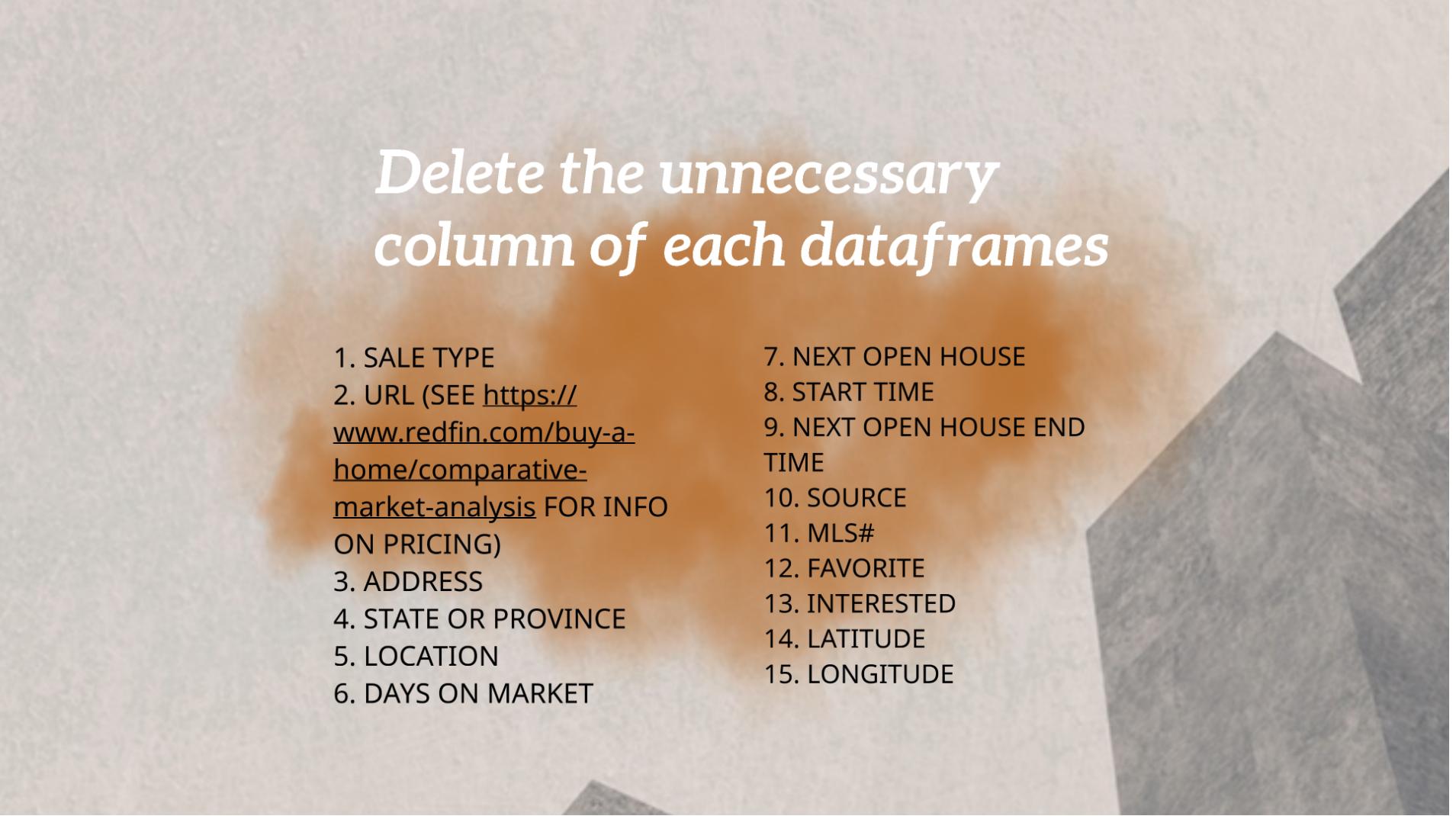
- West San Jose
- Campbell
- Cupertino
- Santa Clara
- Sunnyvale

2100 cases West San Jose  
350 cases each other city

The housing price gap between cities

wrong result

## *Delete the unnecessary column of each dataframes*

- 
- 1. SALE TYPE
  - 2. URL (SEE <https://www.redfin.com/buy-a-home/comparative-market-analysis> FOR INFO ON PRICING)
  - 3. ADDRESS
  - 4. STATE OR PROVINCE
  - 5. LOCATION
  - 6. DAYS ON MARKET
  - 7. NEXT OPEN HOUSE
  - 8. START TIME
  - 9. NEXT OPEN HOUSE END TIME
  - 10. SOURCE
  - 11. MLS#
  - 12. FAVORITE
  - 13. INTERESTED
  - 14. LATITUDE
  - 15. LONGITUDE

# *Data Cleaning*

The outlier data :

- West San Jose
- Campbell
- Cupertino
- Santa Clara
- Sunnyvale

2100 cases West San Jose  
350 cases each other city

The housing price gap between cities

wrong result

1

2

3

4

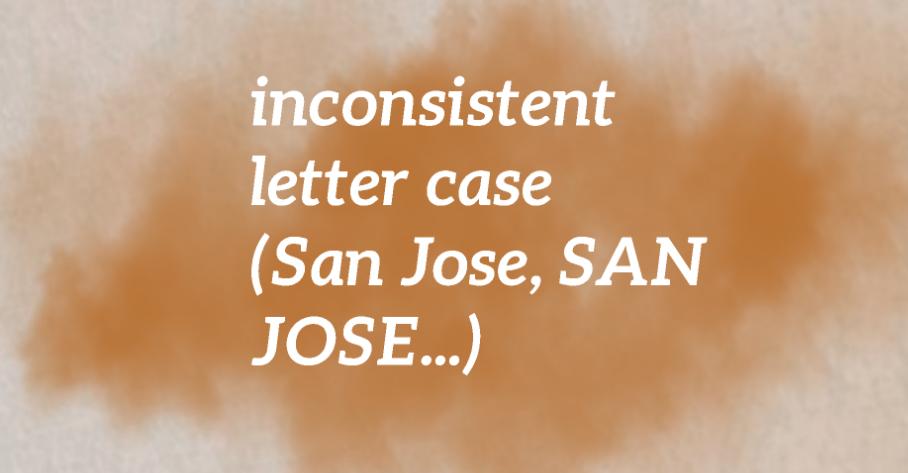
5



## *Delete the wrong data*

1. other typies of real estate  
(Cando...)
2. other cities  
(los gatos, saratoga...)

*Convert*



*inconsistent  
letter case  
(San Jose, SAN  
JOSE...)*



## *Delete the wrong data*

1. other typies of real estate  
(Cando...)
2. other cities  
(los gatos, saratoga...)

*Convert*

# Data Cleaning

The outlier data :

- West San Jose
- Campbell
- Cupertino
- Santa Clara
- Sunnyvale

2100 cases West San Jose  
350 cases each other city

The housing price gap between cities

wrong result

1

2

3

4

5



## ***delete the NaN, missing data***

Not in all column (Will lead to a significant reduction of data cases )

but the column that used as the most important attribute

# Data Cleaning

The outlier data :

- West San Jose
- Campbell
- Cupertino
- Santa Clara
- Sunnyvale

2100 cases West San Jose  
350 cases each other city

The housing price gap between cities

wrong result

1

2

3

4

5



*delete the outlier data*

On five columns:  
"PRICE", "BEDS", "BATHS",  
"SQUARE FEET", "LOT SIZE".

Z-Score

# Data Cleaning

The outlier data :

- West San Jose
- Campbell
- Cupertino
- Santa Clara
- Sunnyvale

2100 cases West San Jose  
350 cases each other city

The housing price gap between cities

wrong result

1

2

3

4

5



*Combination*

**3257**

*Sample*

	SOLD DATE	PROPERTY TYPE	CITY	ZIP OR POSTAL CODE	PRICE	BEDS	BATHS	SQUARE FEET	LOT SIZE	YEAR BUILT	\$/SQUARE FEET	HOA/MONTH	STATUS
0	November-29-2021	Townhouse	Sunnyvale	94086	1230000	2.0	2.5	1553.0	1.0	1991.0	792.0	505.0	Sold
1	September-8-2021	Single Family Residential	Sunnyvale	94086	2500000	3.0	2.5	1688.0	4964.0	2000.0	1481.0	NaN	Sold
2	February-17-2017	Townhouse	Sunnyvale	94085	849000	2.0	2.5	1277.0	867.0	2006.0	665.0	235.0	Sold
3	May-14-2018	Single Family Residential	Sunnyvale	94086	1700000	3.0	2.0	1200.0	5187.0	1944.0	1417.0	NaN	Sold
4	October-15-2020	Single Family Residential	Sunnyvale	94089	1300000	3.0	2.0	1108.0	6300.0	1958.0	1173.0	NaN	Sold

14 columns    3257 rows



*Combination*

**3257**

*Sample*



# Data Cleaning

1

2

3

4

5

The outlier data :

- West San Jose
- Campbell
- Cupertino
- Santa Clara
- Sunnyvale

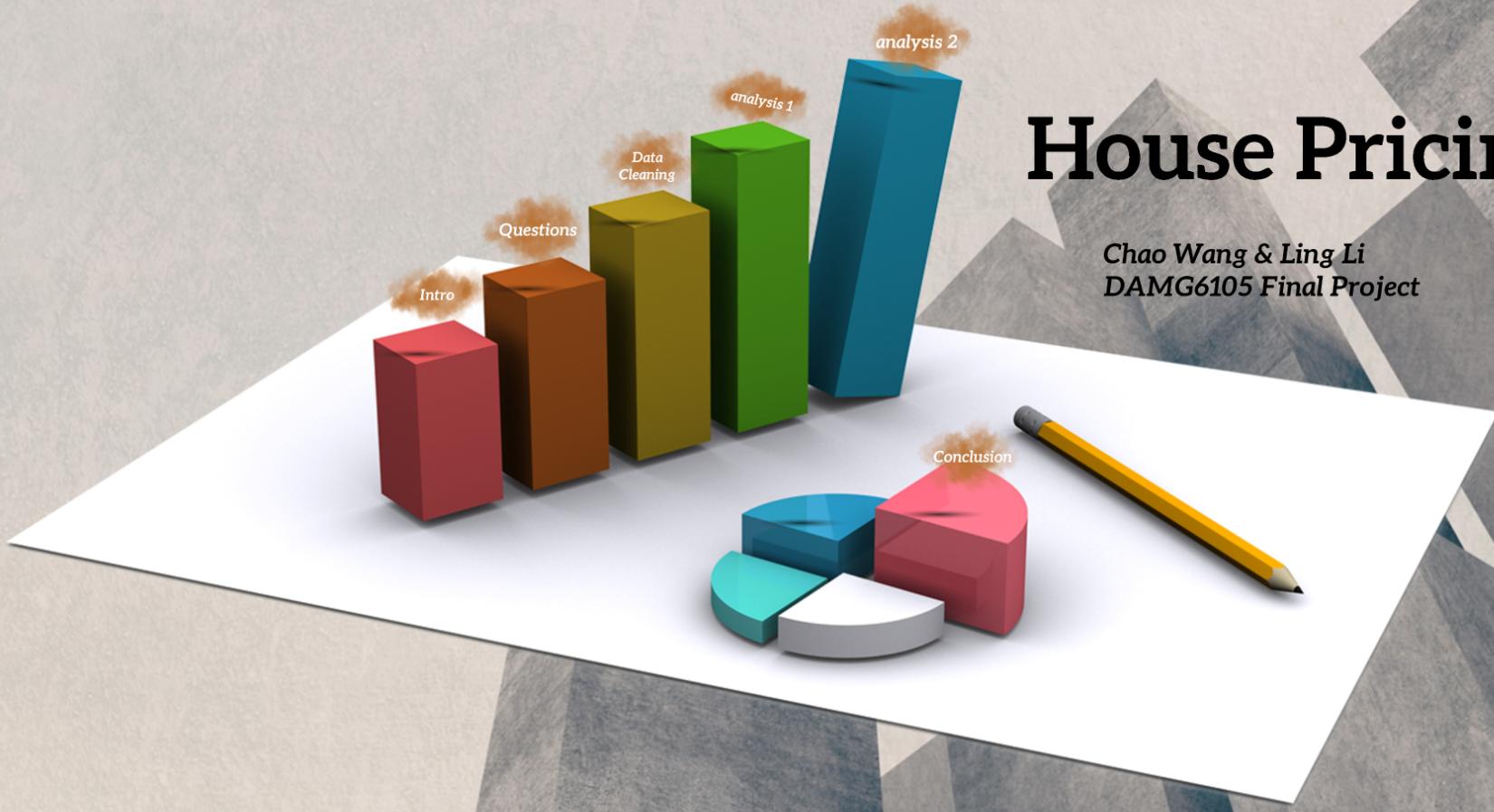
2100 cases West San Jose  
350 cases each other city

The housing price gap between cities

wrong result

# House Pricing

*Chao Wang & Ling Li  
DAMG6105 Final Project*





# *House Pricing & House Self-Attributes*



*Compare Median of Townhouse  
and Single Family Residential*

*Chart*





*Compare Median of Townhouse  
and Single Family Residential*

*Chart*



# *House Pricing & House Self-Attributes*

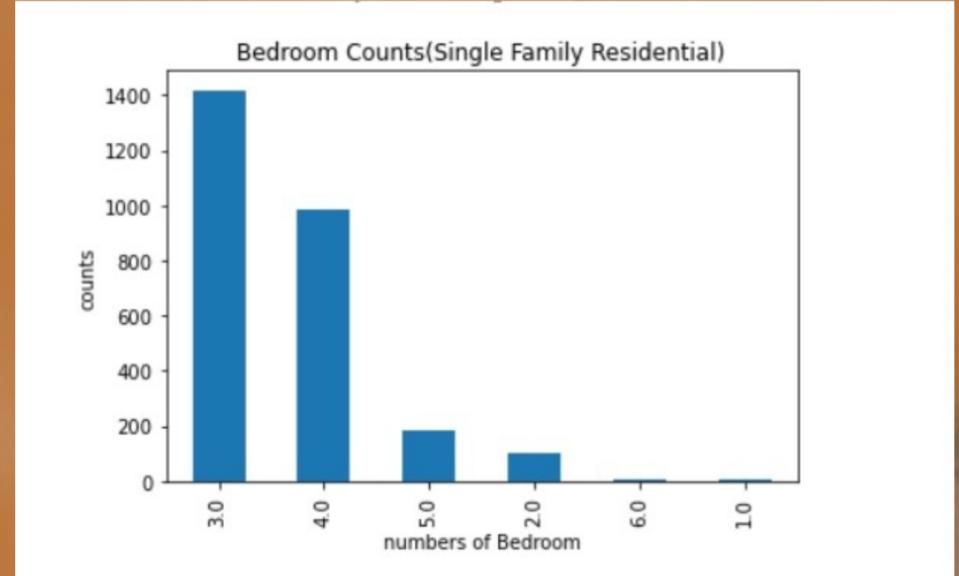
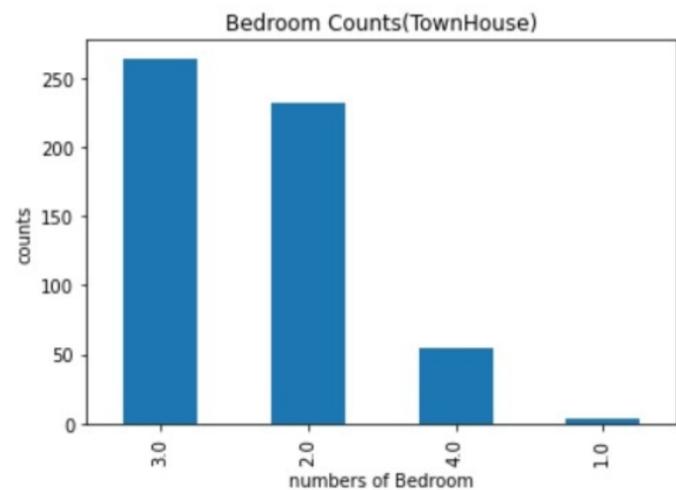


# *The relationship of Bedroom and Housing Price*

Charts

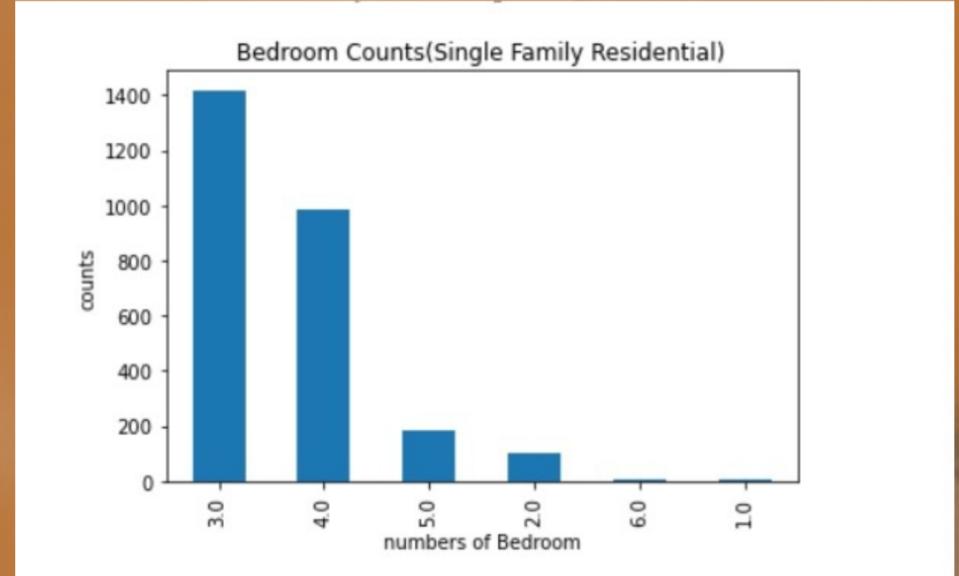
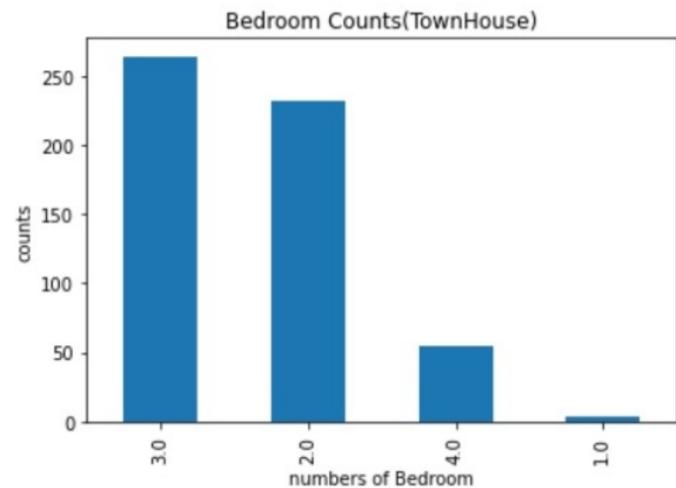
## *Bedrooms*

There are 554 Townhouses in this data.

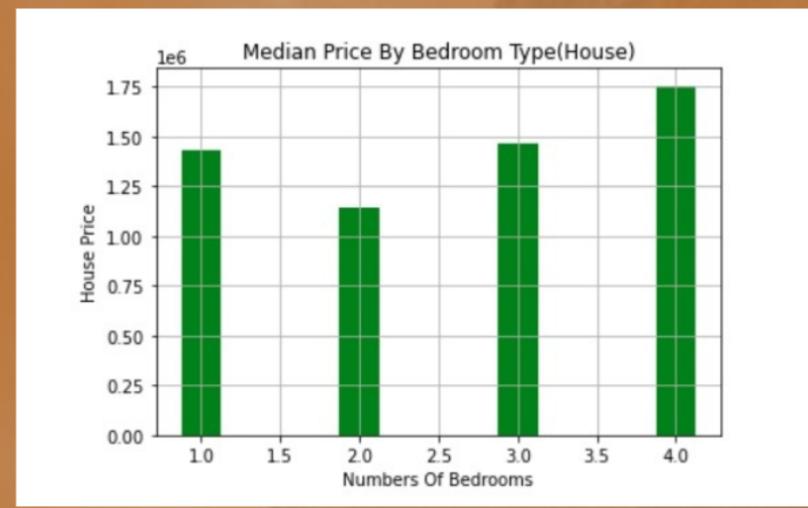


## *Bedrooms*

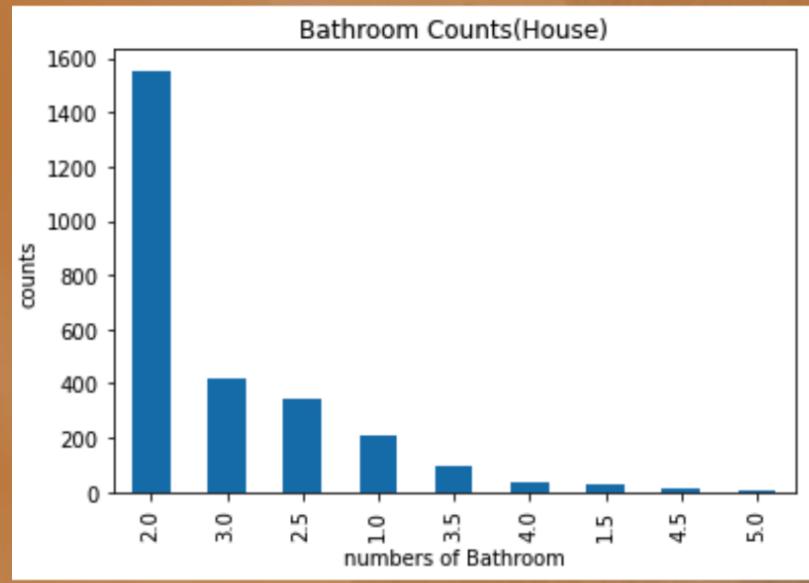
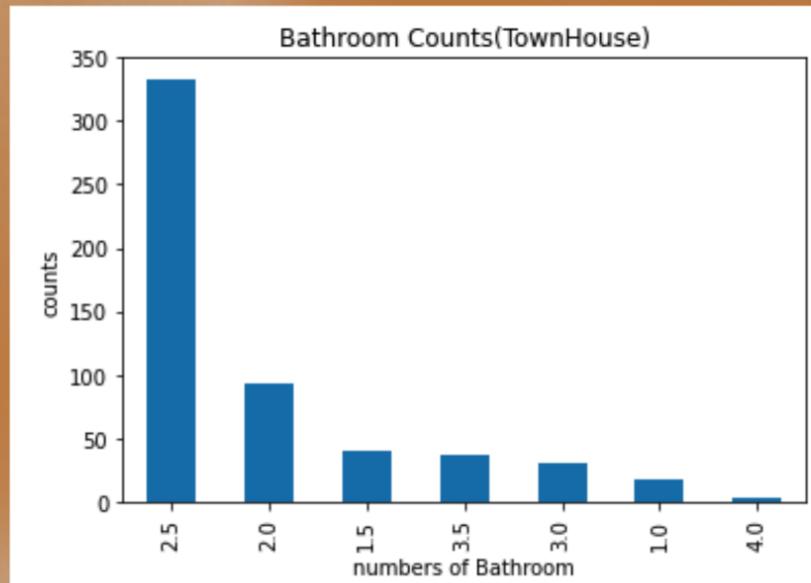
There are 554 Townhouses in this data.



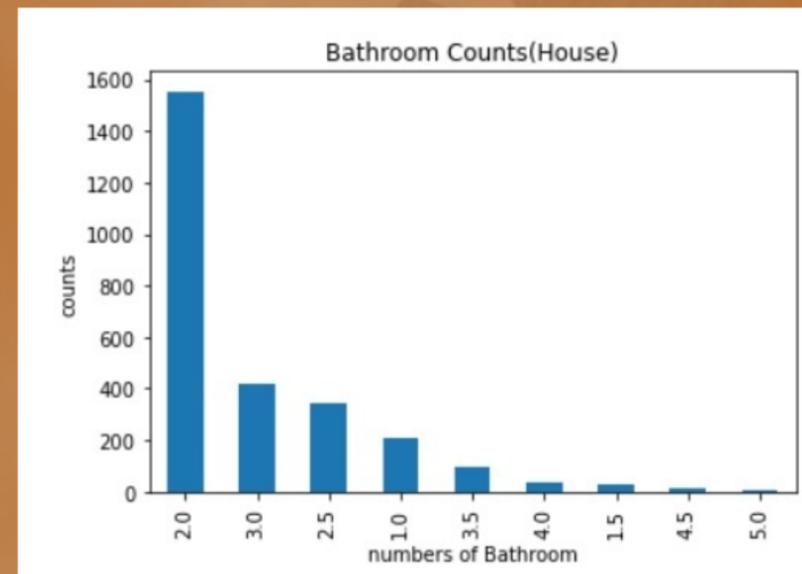
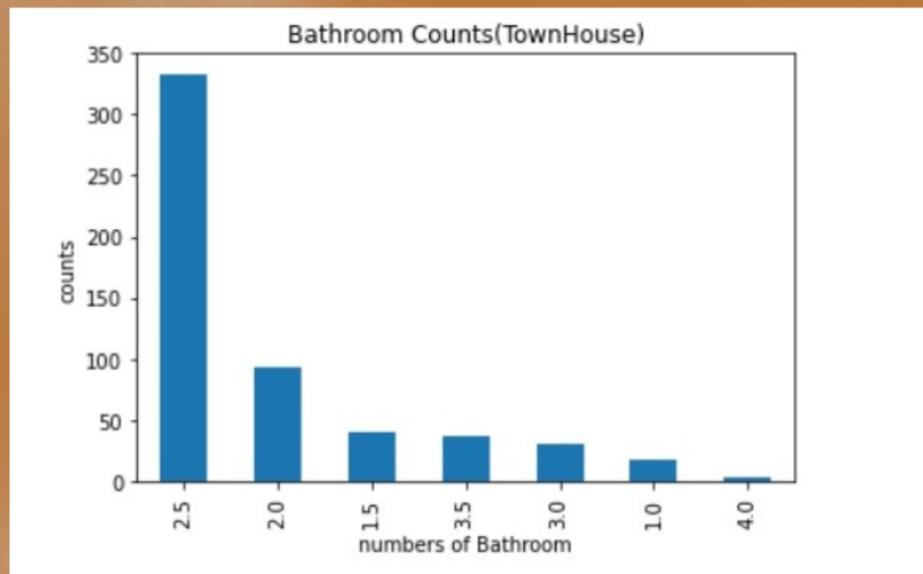
## *Median price of house with different bedroom type*



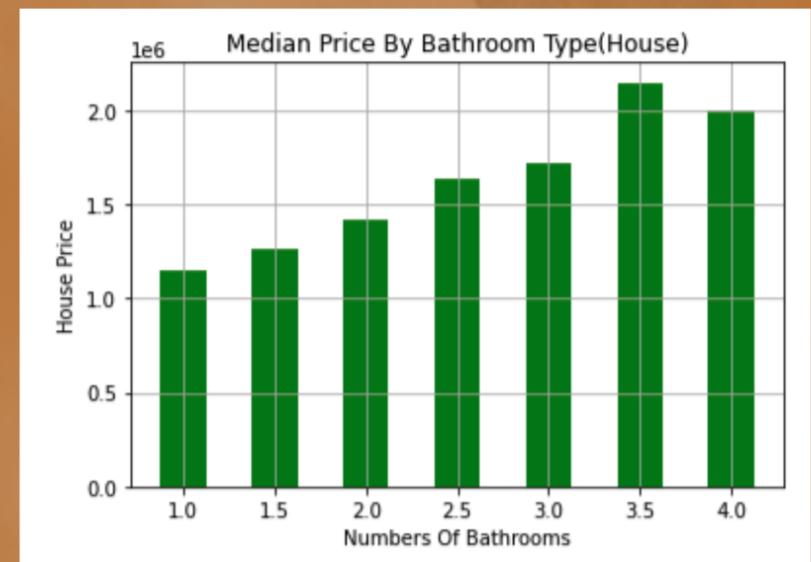
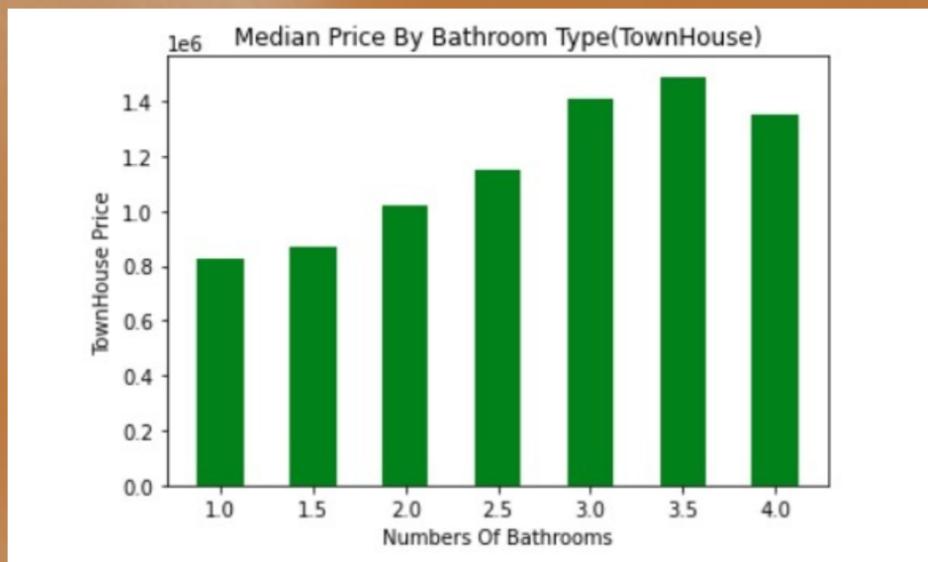
## *Bedroom Type*



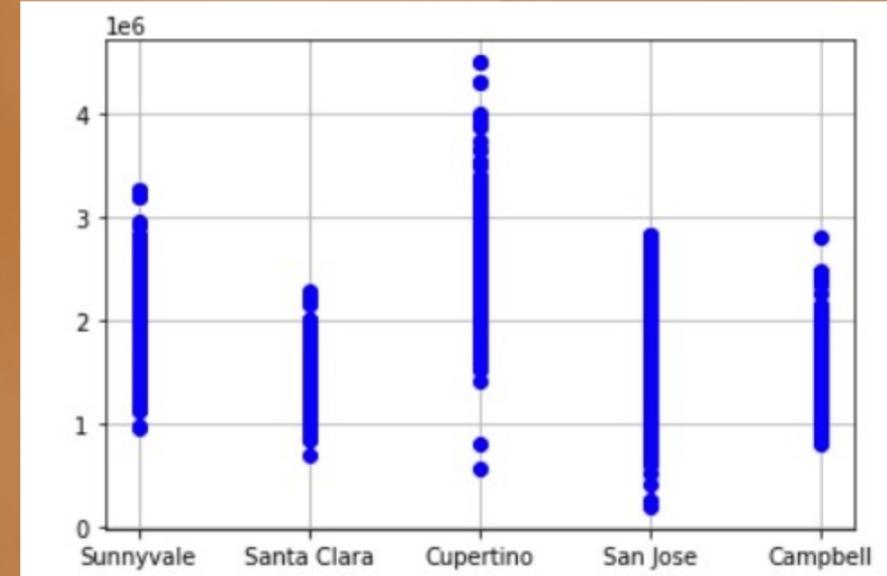
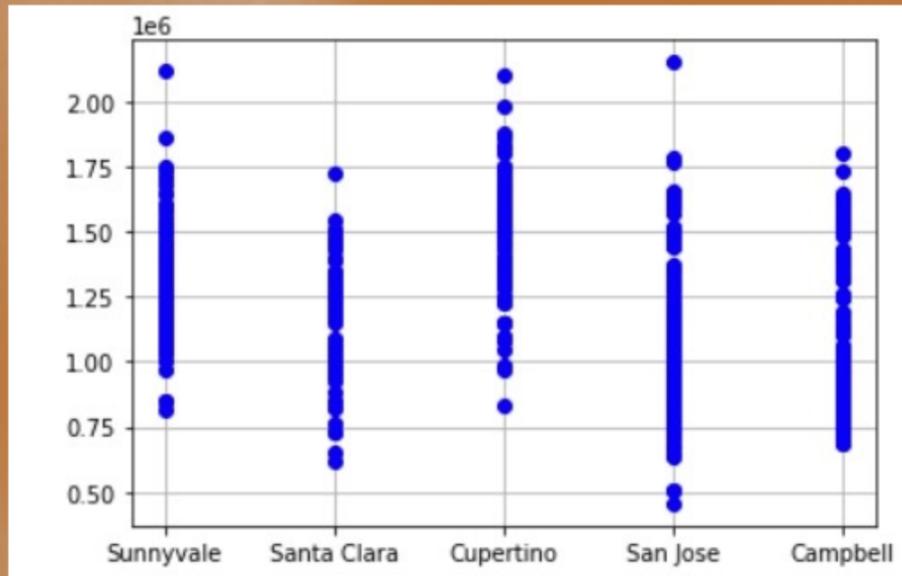
## *Different bathroom type counts among townhouse and single family residential*



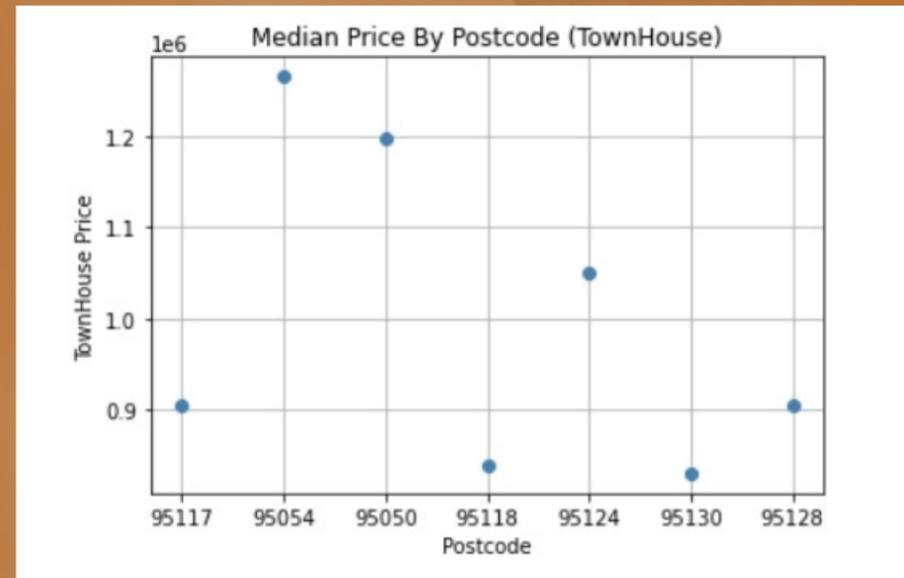
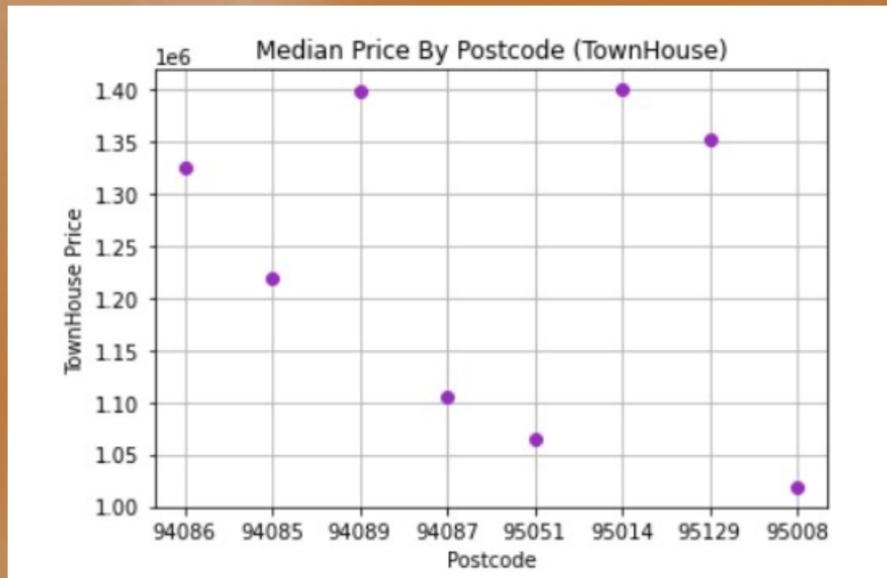
## *Median price of different bathroom type*



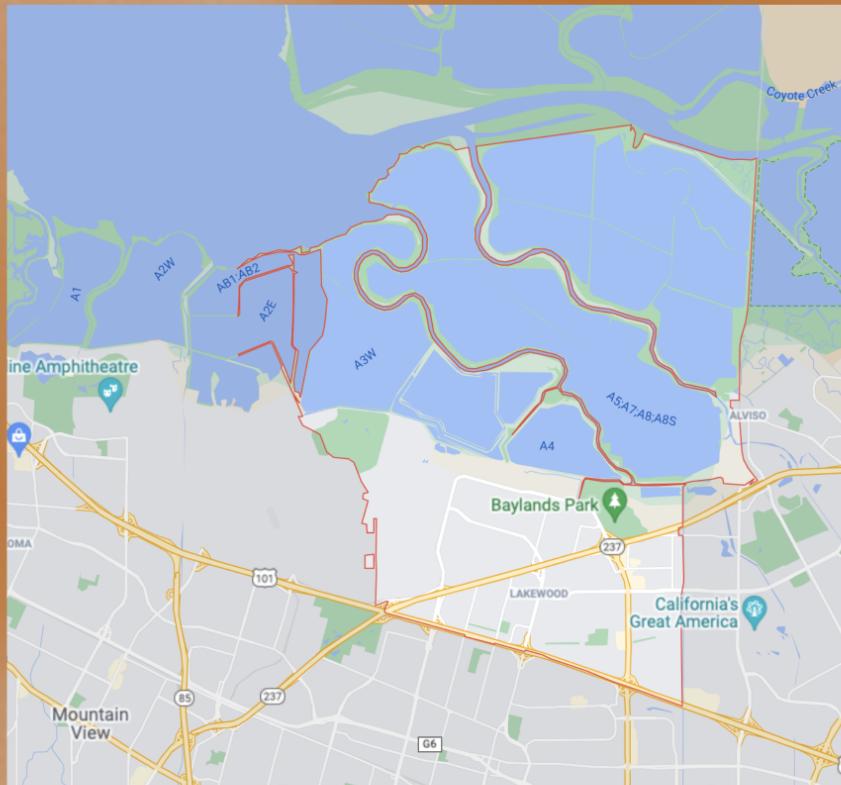
## *Different price by City*



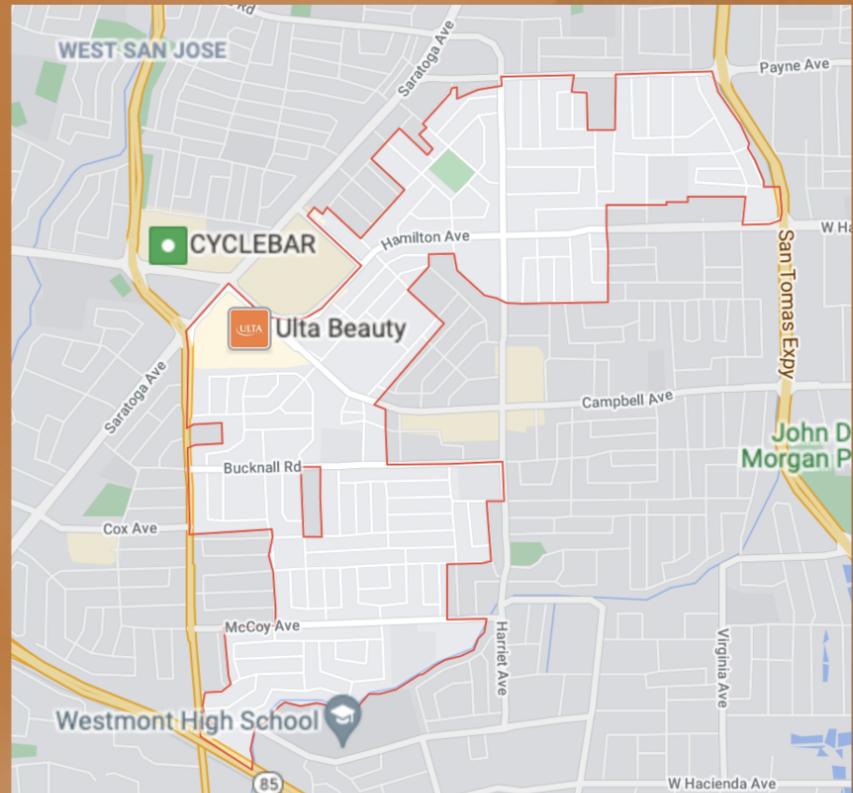
## *Median price of different postcodes of TH (separate TH into two Graphs)*



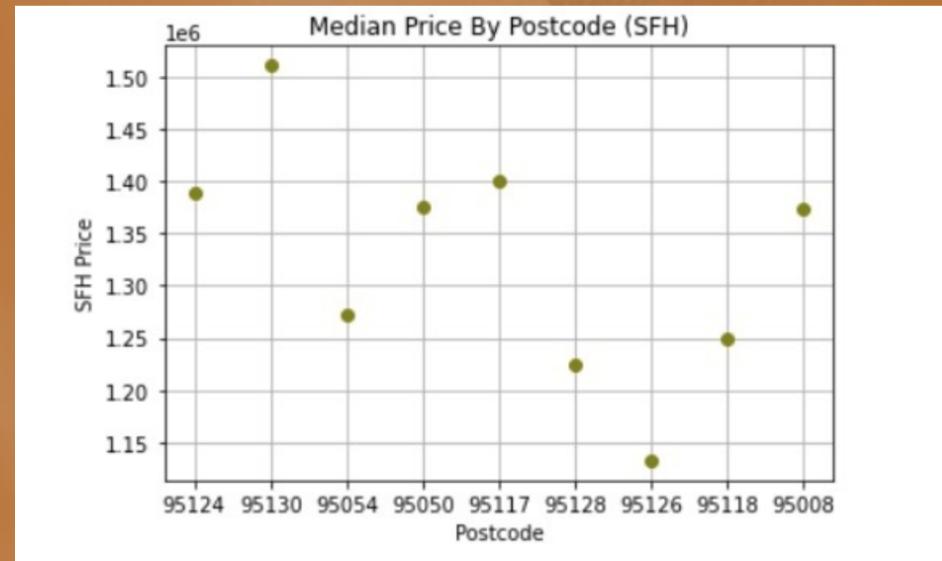
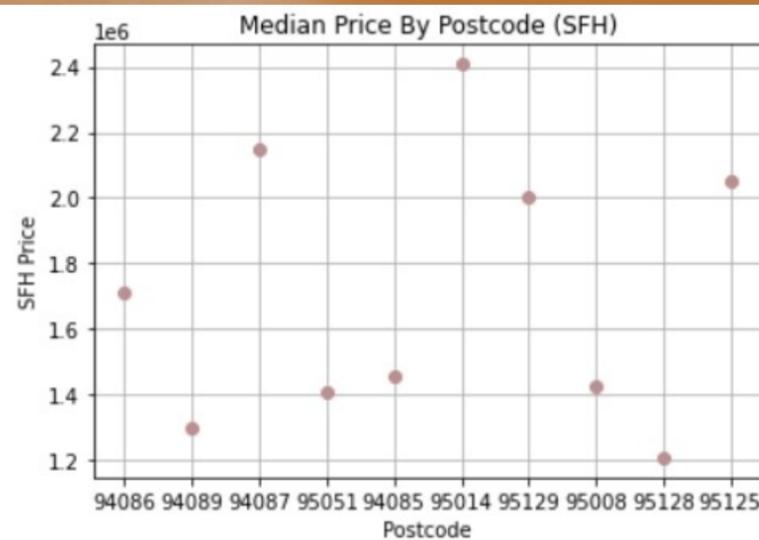
**94089**



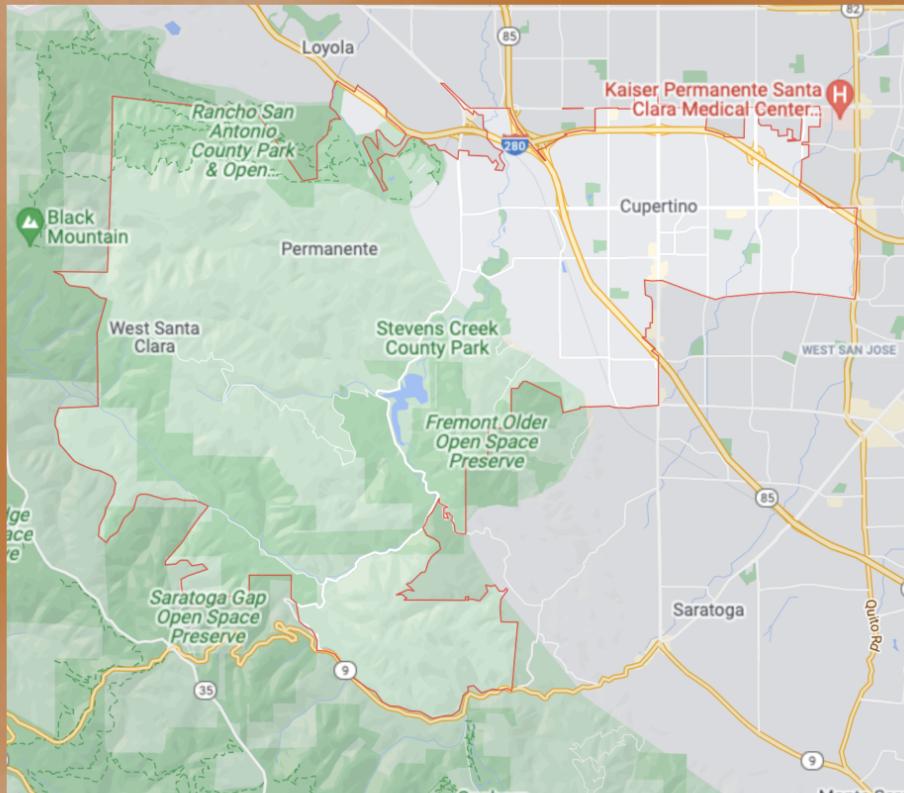
**95130**



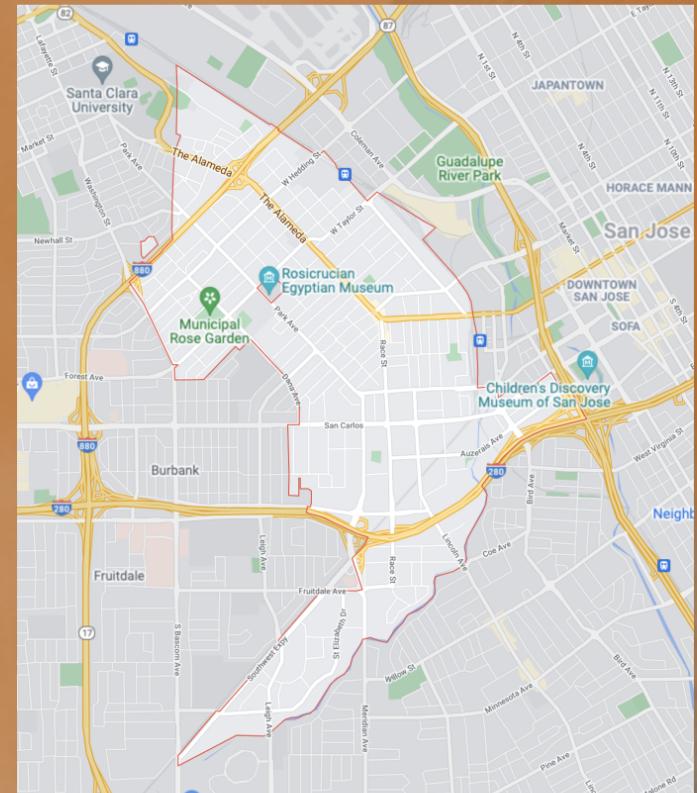
## *Median price of different postcodes of SFH (separate SFH into two Graphs)*



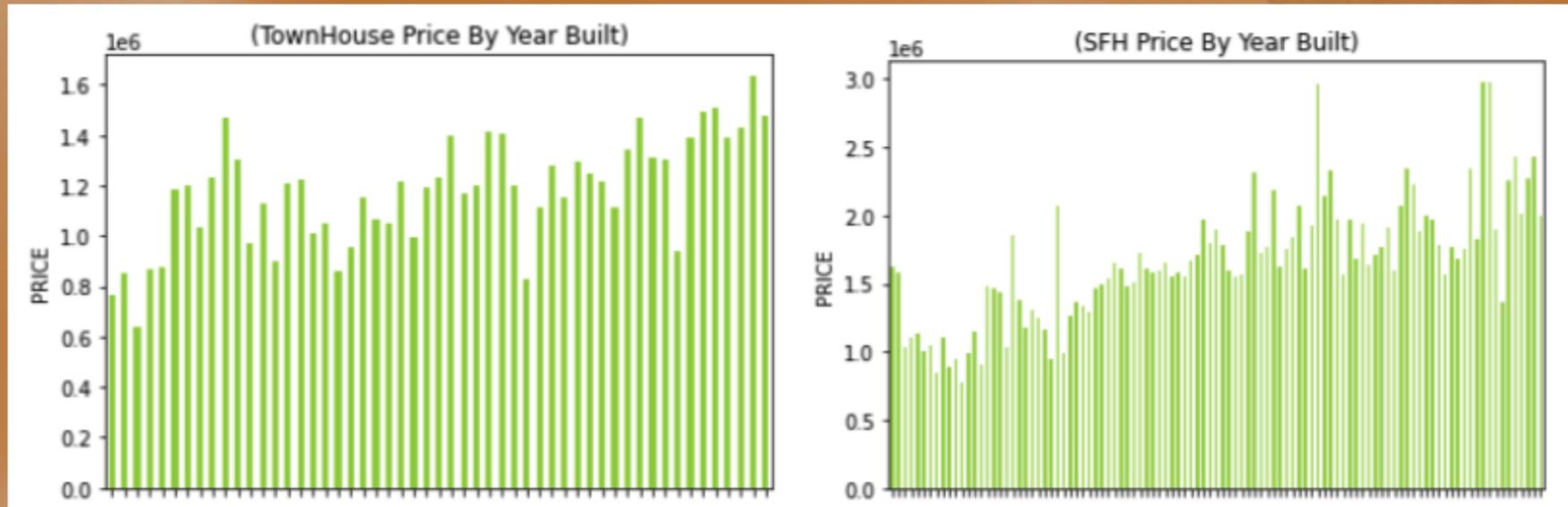
**94089**



**95126**

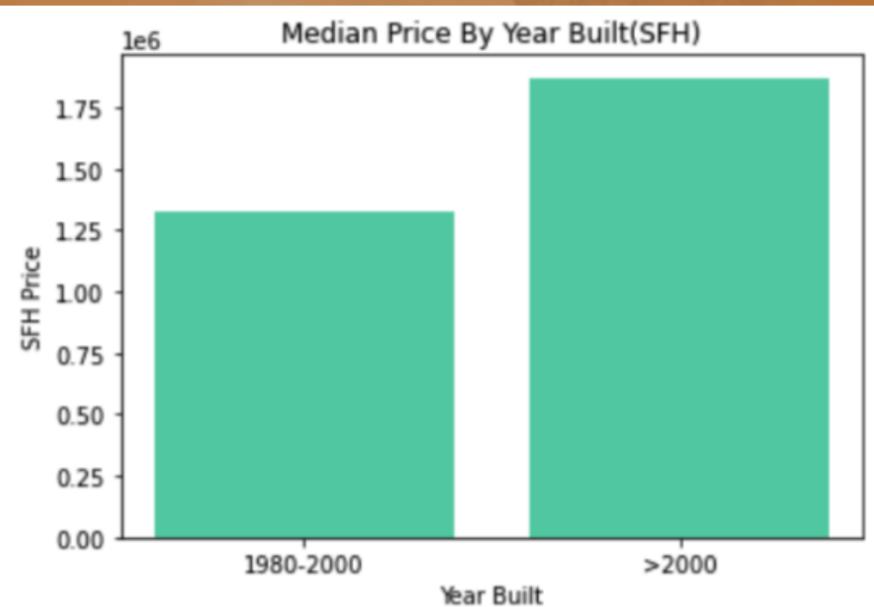
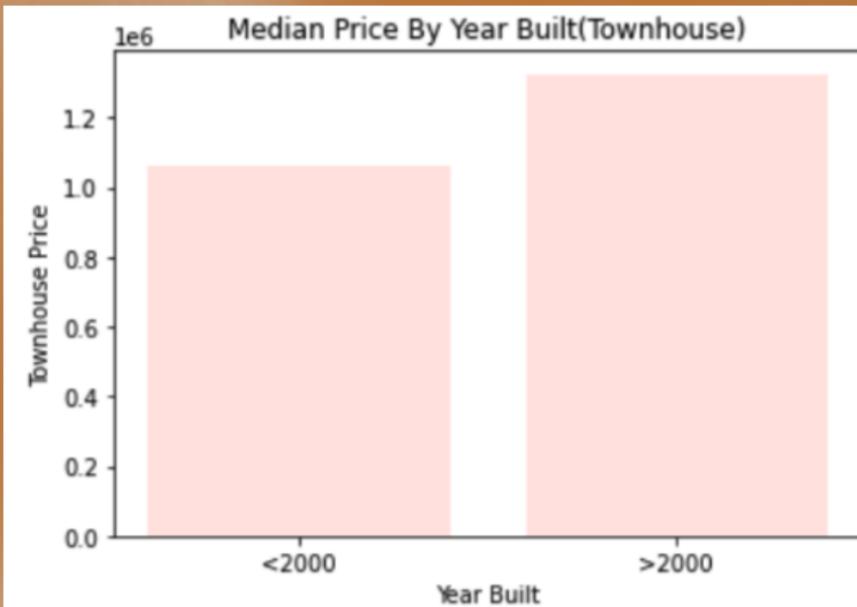


## *Mean price of different year-built*

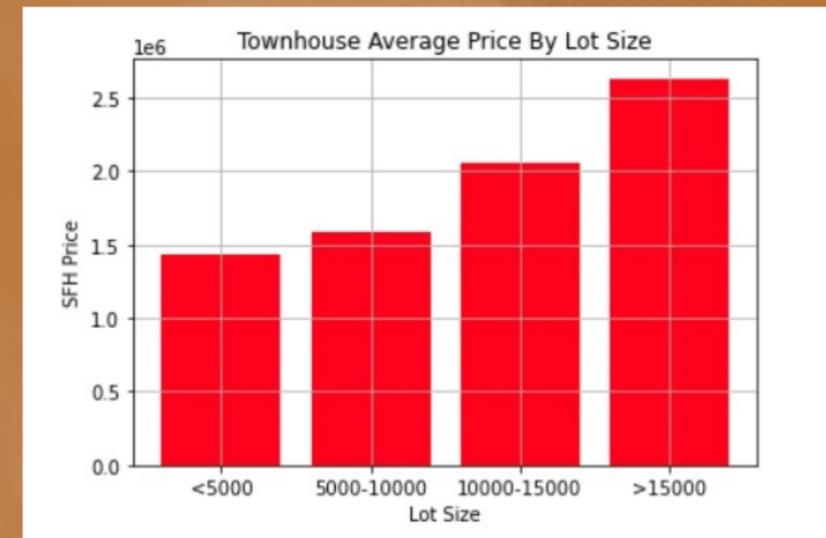
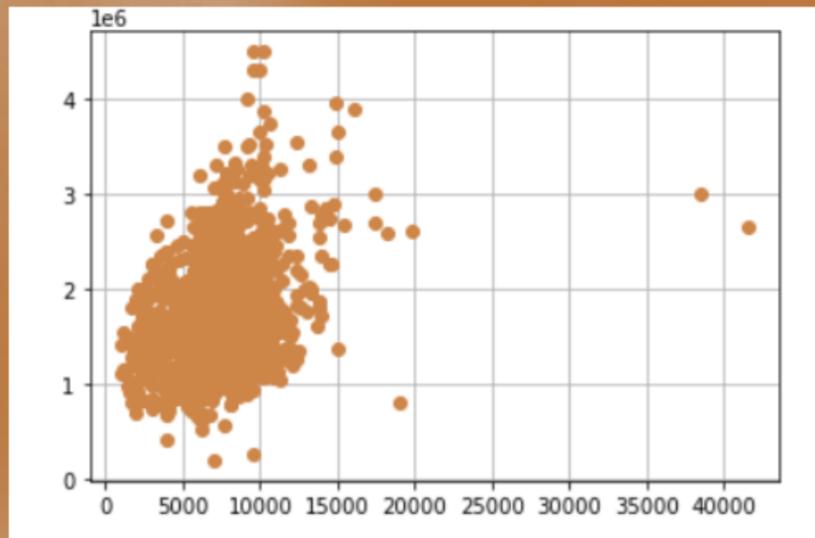


## *The relationship between year built and housing price*

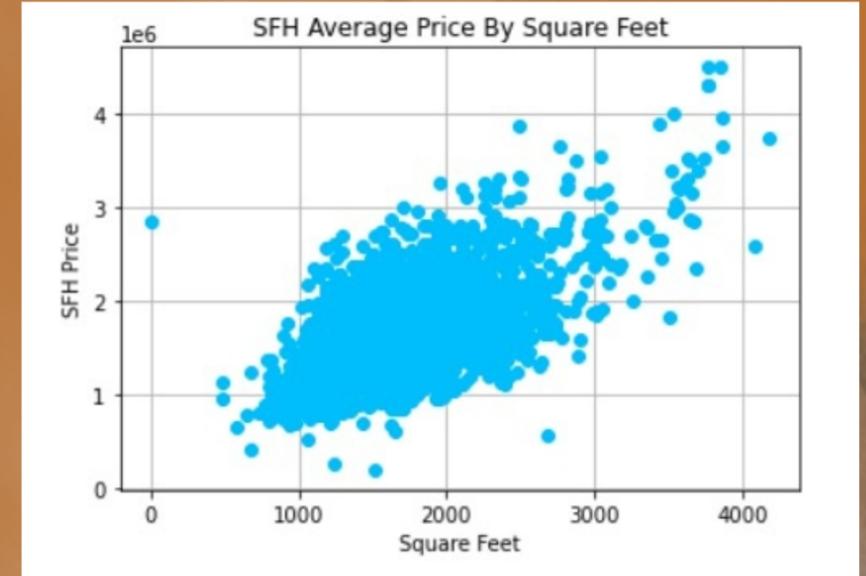
We divided year-built to 4 periods

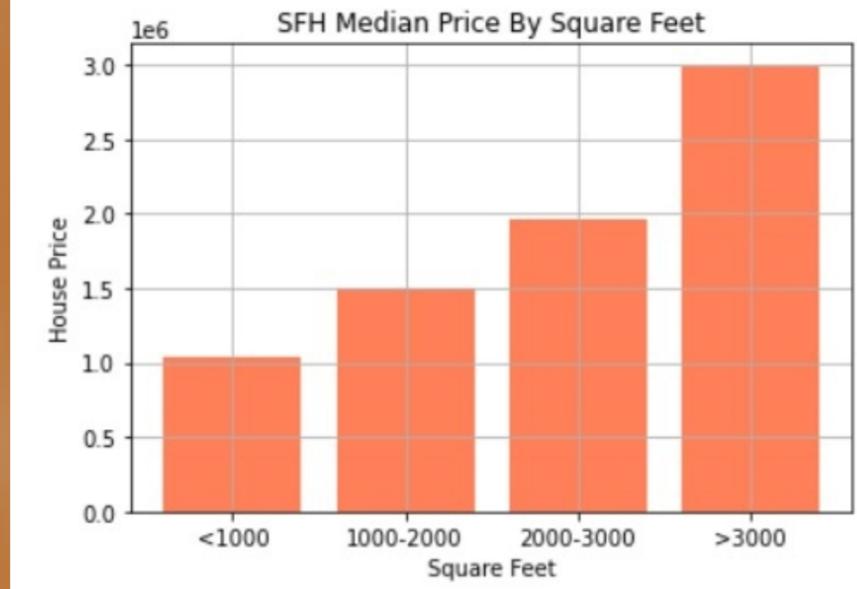
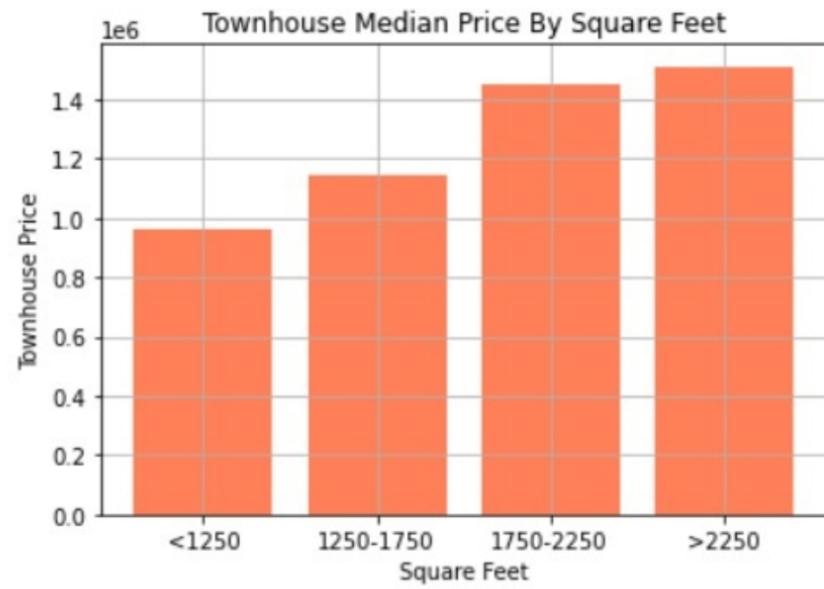


## *The relationship of Lot Size and Housing Price*

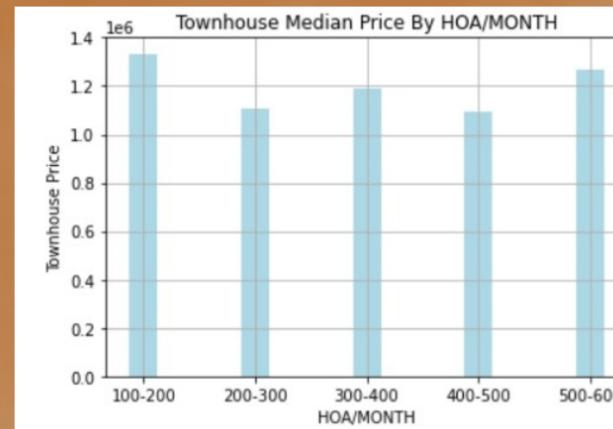
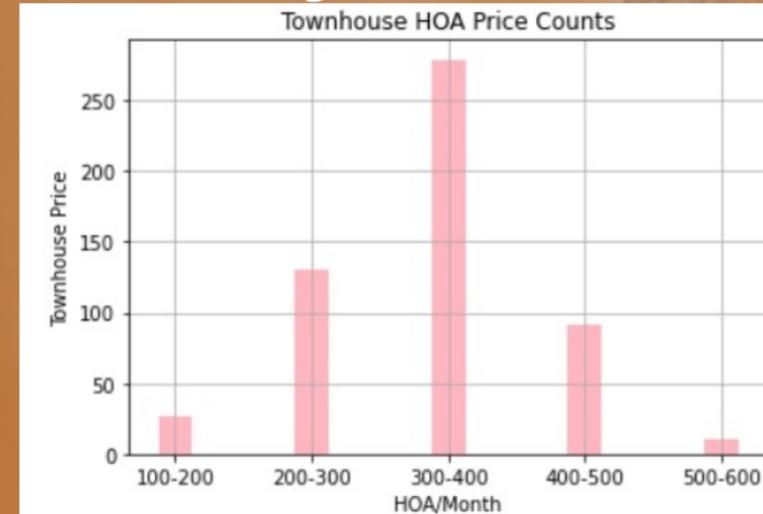
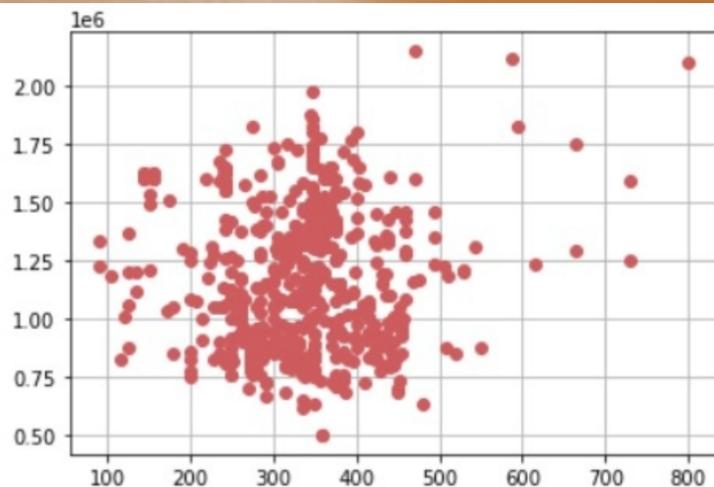


## *Scatter plot of the relationship of square feet and housing price*

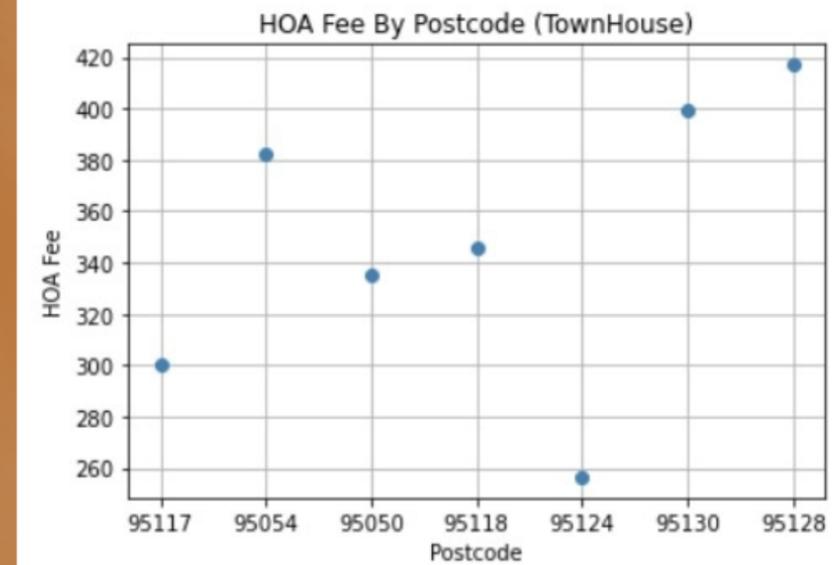
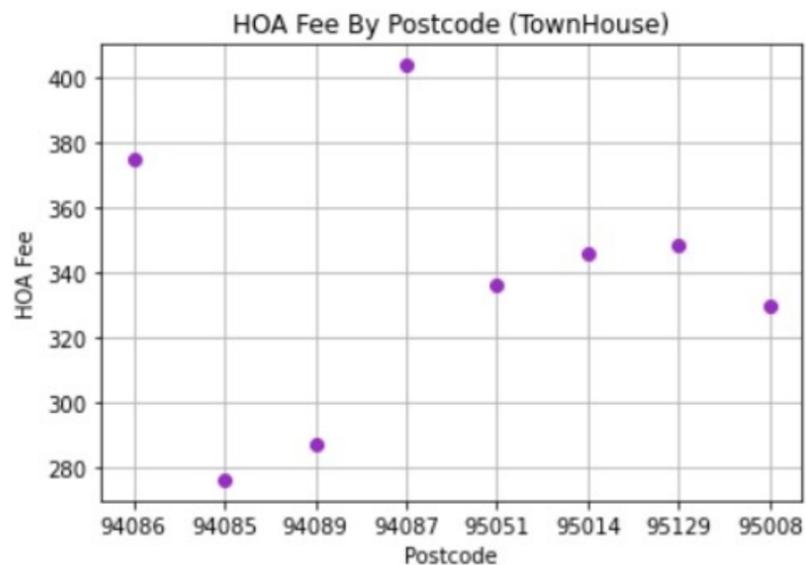


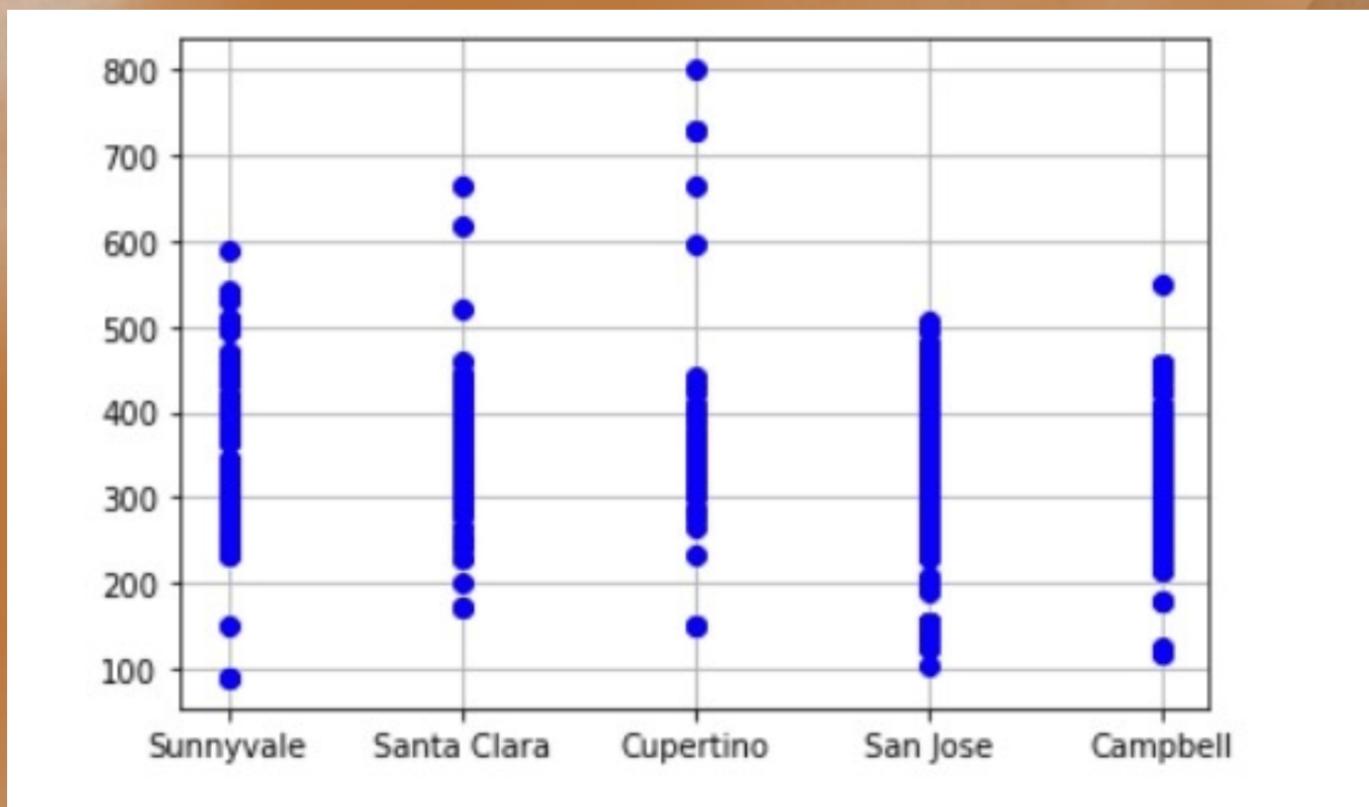


## *The relationship of HOA/MONTH and Housing Price*



## *HOA fee of different postcodes in townhouse*







# *The relationship of Bedroom and Housing Price*

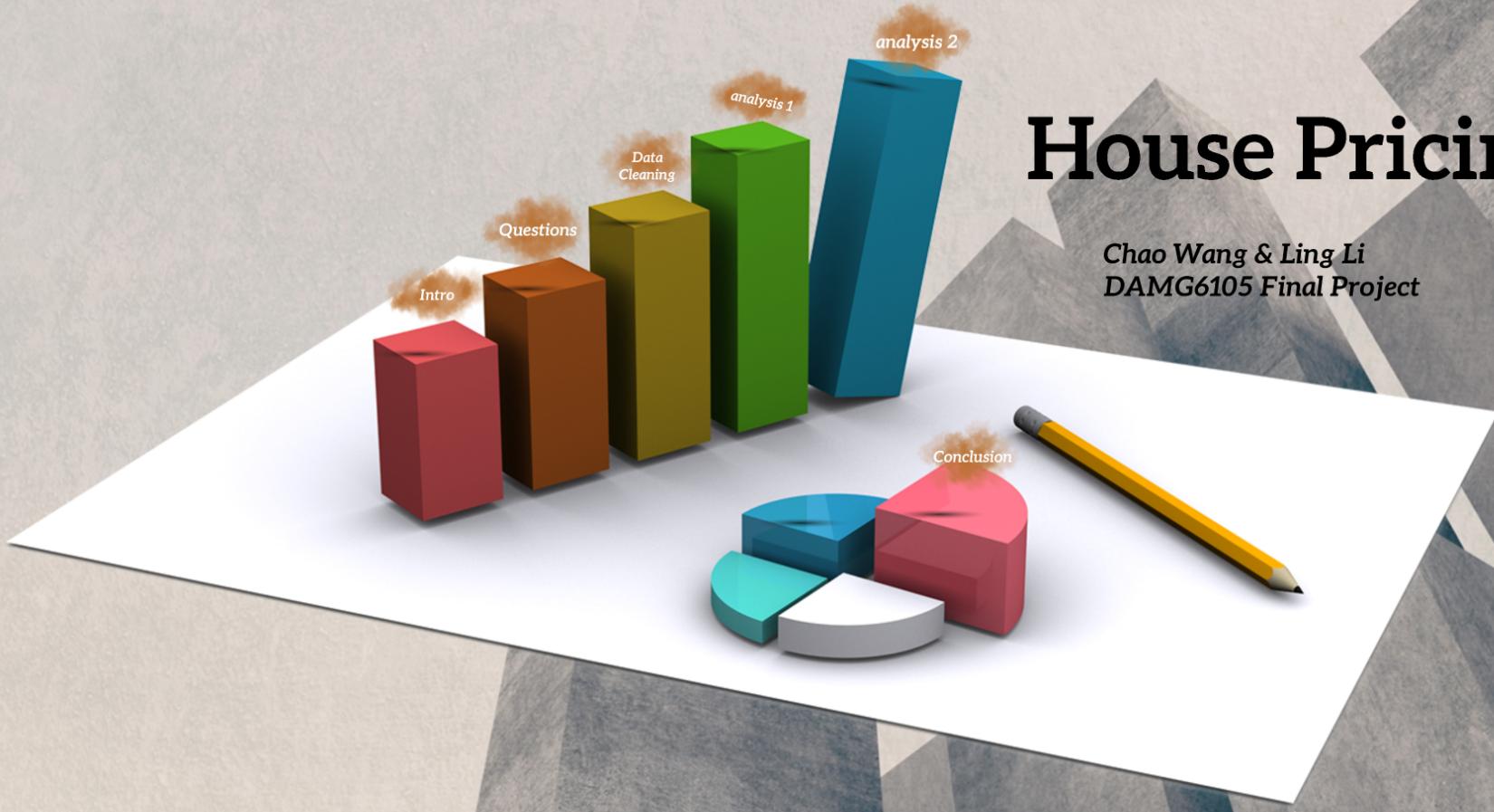
*Charts*



# *House Pricing & House Self-Attributes*

# House Pricing

Chao Wang & Ling Li  
DAMG6105 Final Project





**COVID and Housing Pricing**

**A**

**B**

**C**



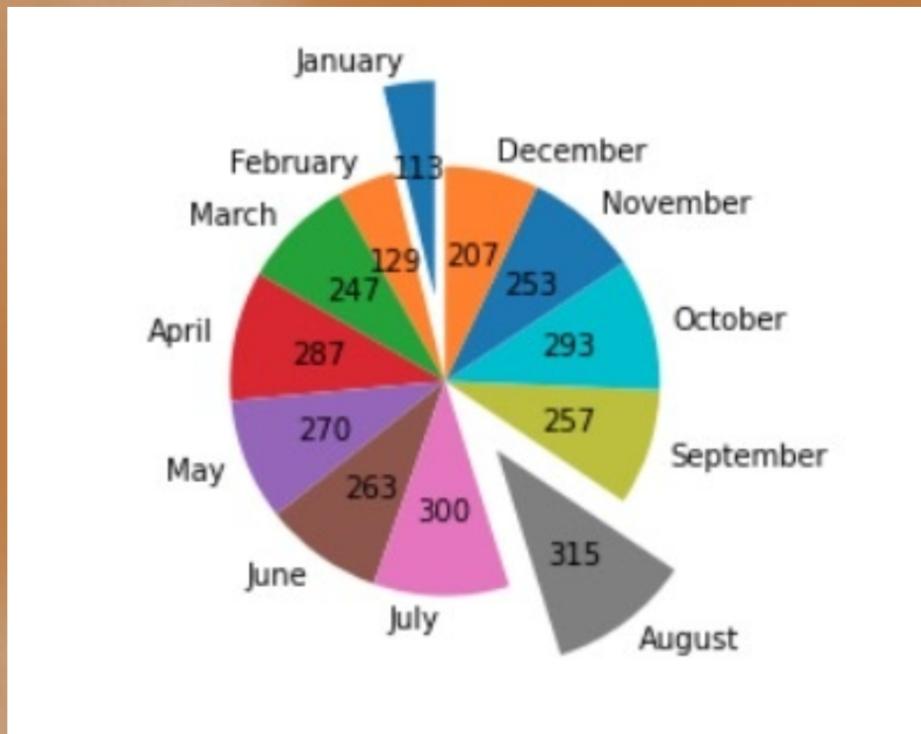
*Does the house grow faster?*



*Does the house grow faster?*

# 1.

## monthly housing sold 2017 -2021



August?

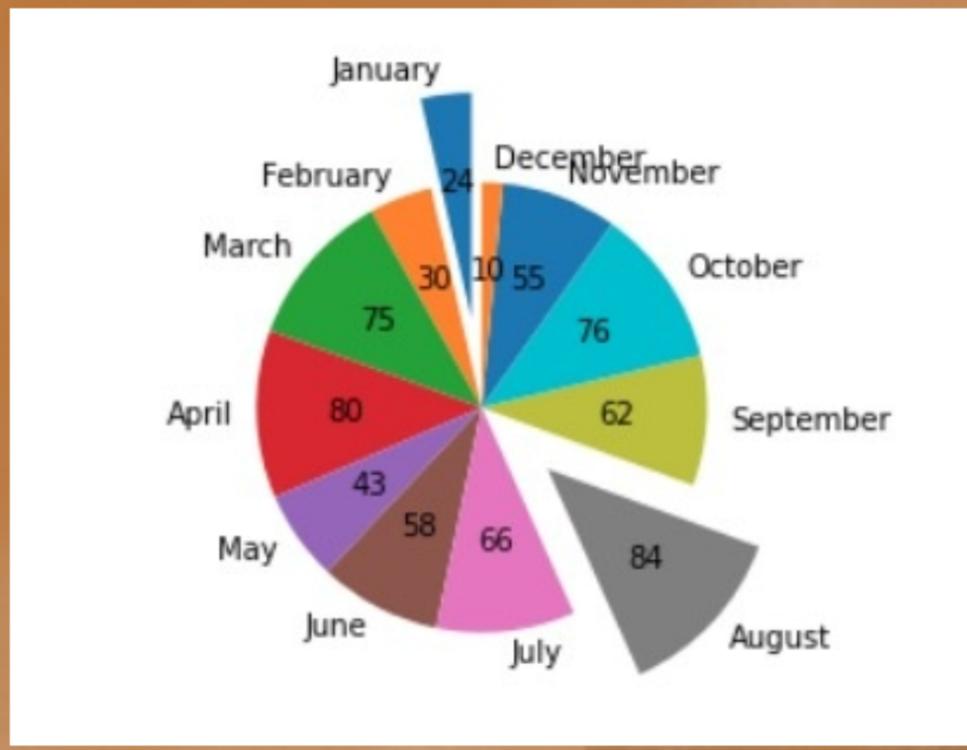
January?

sold date  
closing date  
(one month  
before)

July!

December!

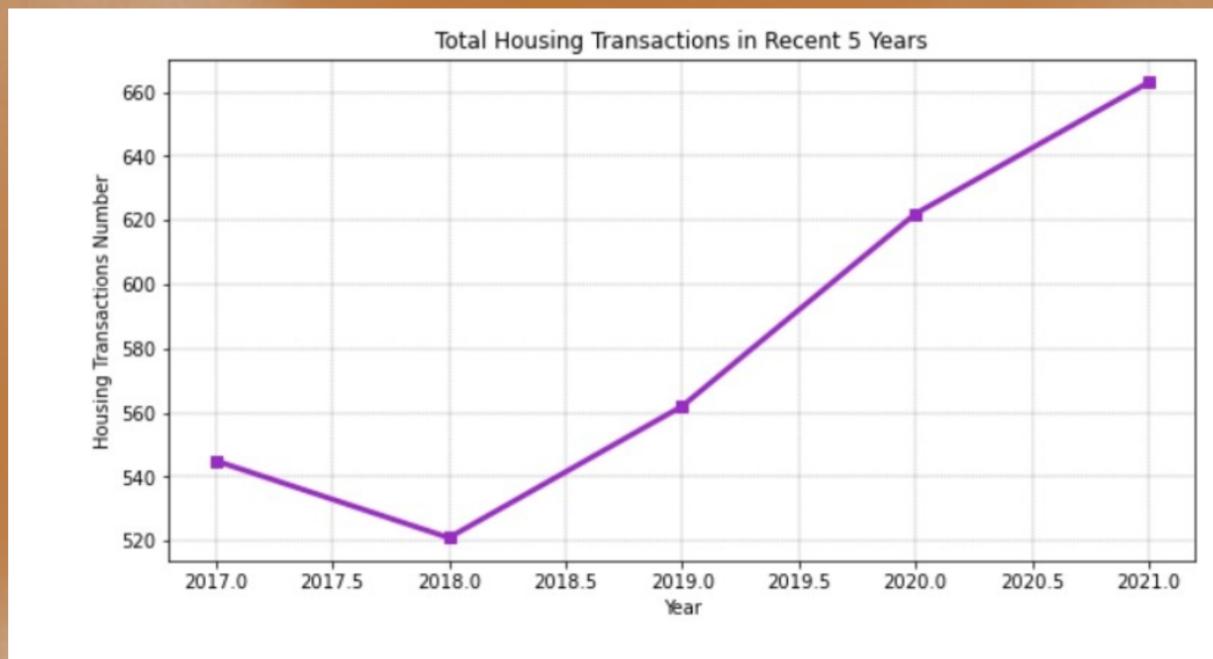
## monthly housing sold, year 2021



divide monthly housing sold 2017 -2021 with 5



## 2. Yearly Performance



2018.  
the trade war  
between China  
and American

2020.  
WFH environment  
mortgage rate

### 3. *data from the same month of different years*

2017 April 58  
2018 April 49  
2019 April 54  
2020 April 46  
2021 April 80

2017 October 50  
2018 October 46  
2019 October 48  
2020 October 73  
2021 October 76



April of 2020  
start of pandemic

April of 2021  
mortgage rate  
decreasing

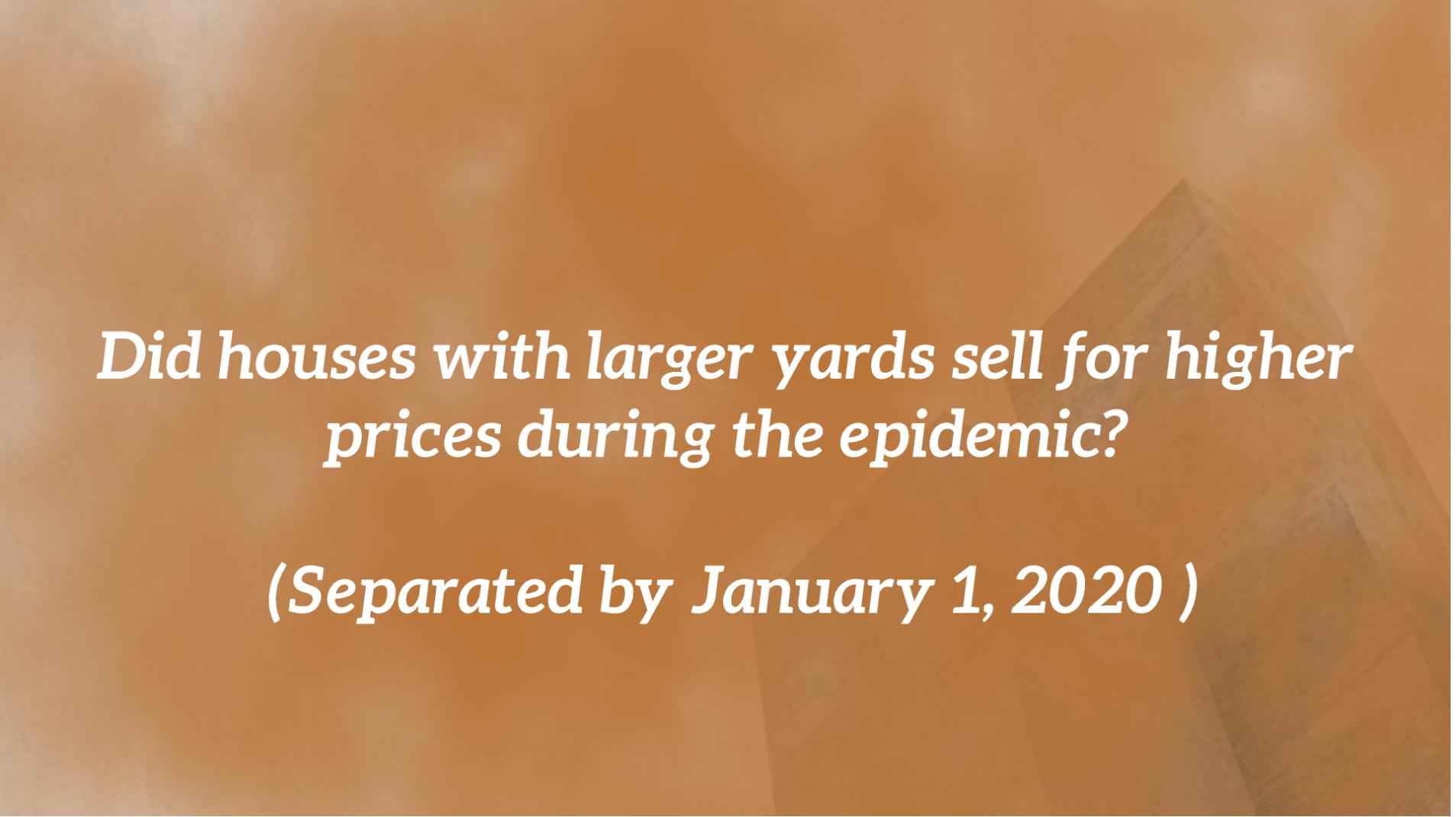


**COVID and Housing Pricing**

**A**

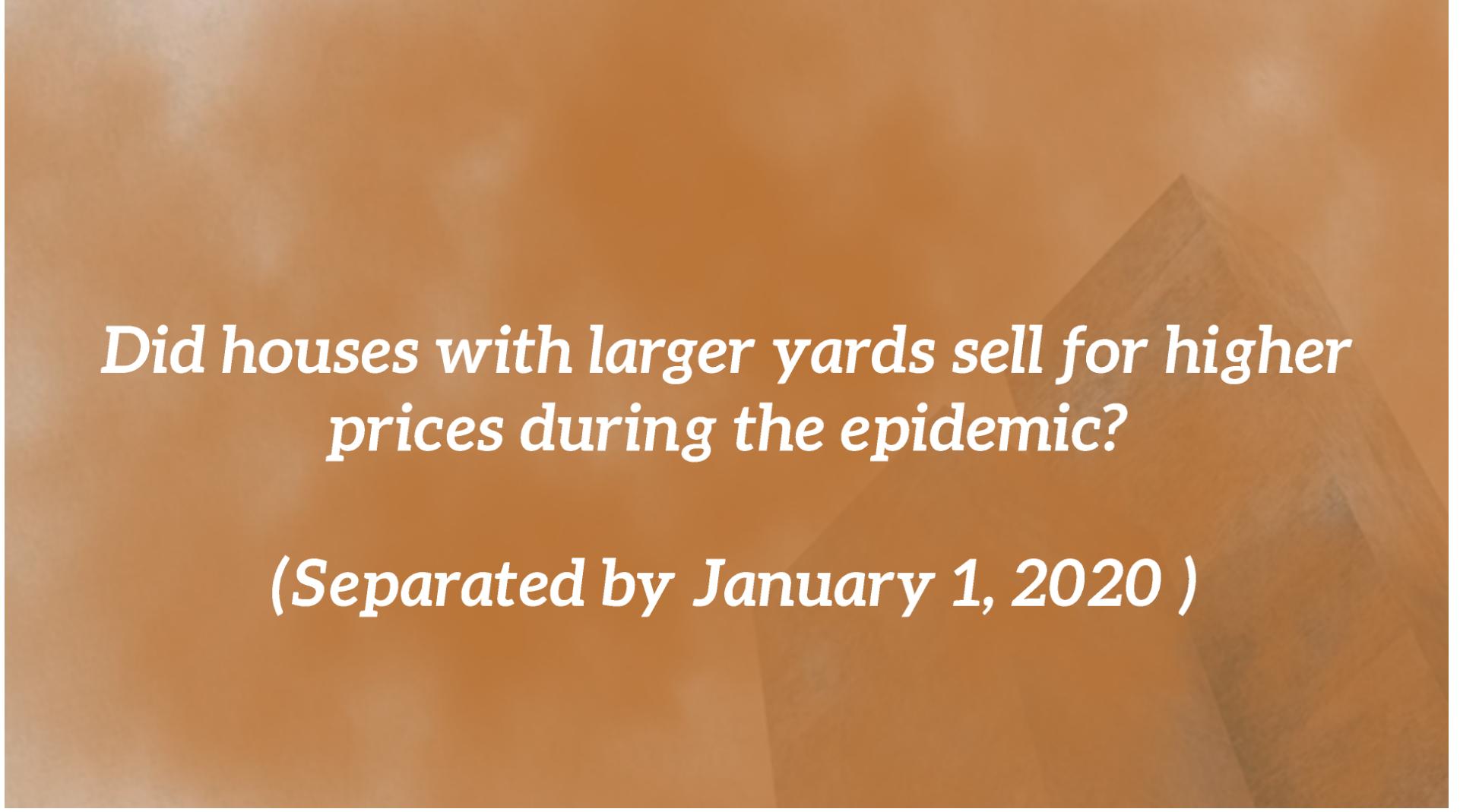
**B**

**C**



*Did houses with larger yards sell for higher  
prices during the epidemic?*

*(Separated by January 1, 2020 )*



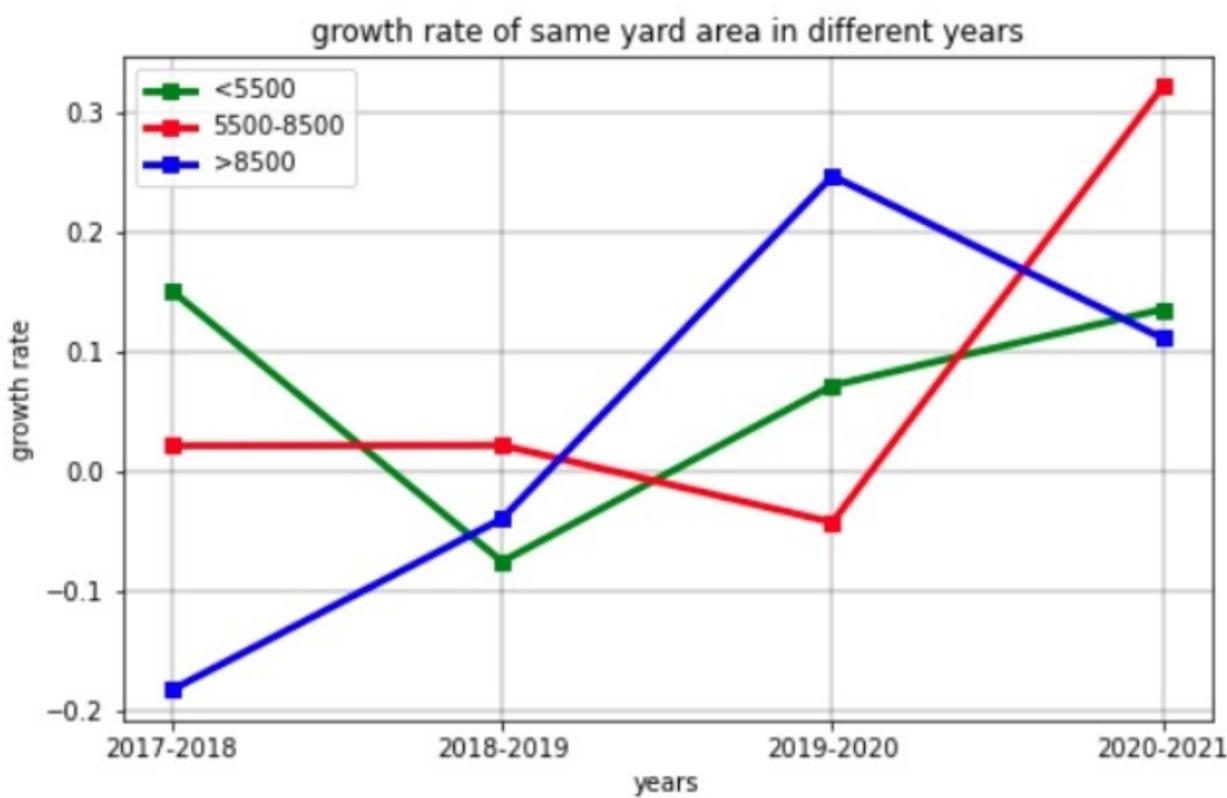
*Did houses with larger yards sell for higher  
prices during the epidemic?*

*(Separated by January 1, 2020 )*

**Control variable principle:**  
**the most common room type, 3b2b**  
**the most common indoor area, 1200-1800**  
**the only variable, the size of the yard.**

***Yard area = lot size-indoor area.***

***Divide the area of the backyard into "<5500", "5500-8500", ">8500"***



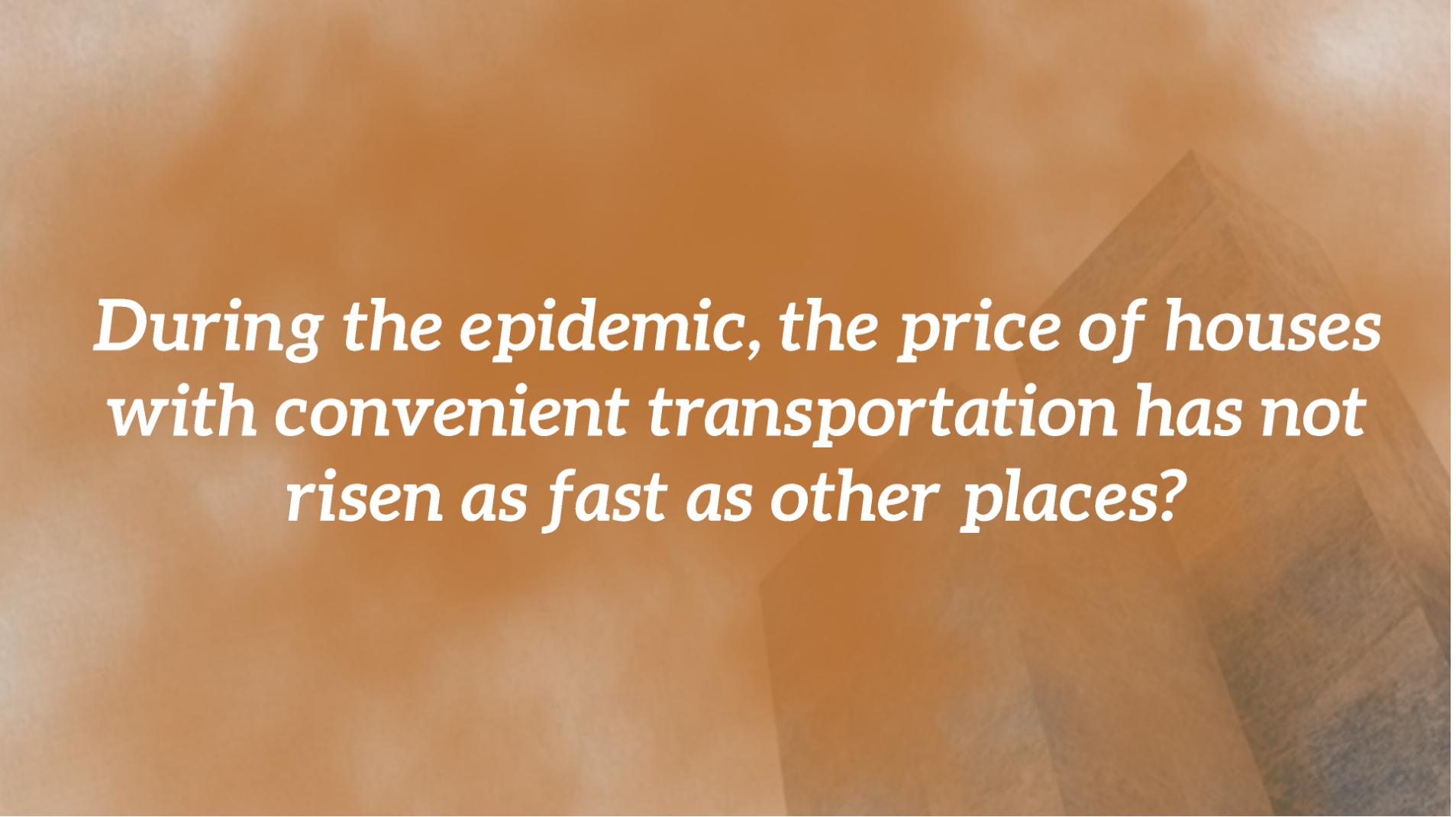


**COVID and Housing Pricing**

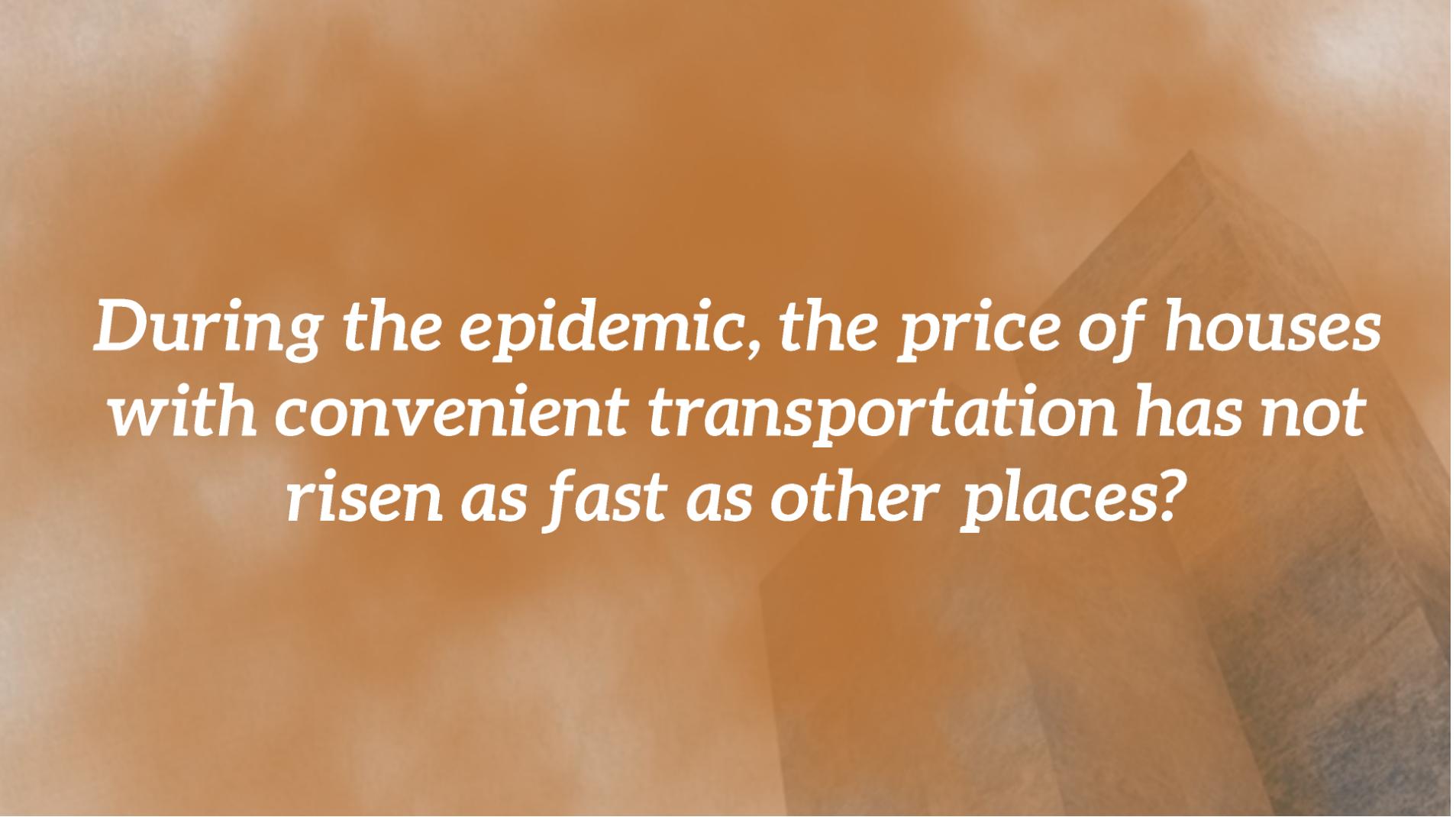
**A**

**B**

**C**



*During the epidemic, the price of houses  
with convenient transportation has not  
risen as fast as other places?*



*During the epidemic, the price of houses  
with convenient transportation has not  
risen as fast as other places?*

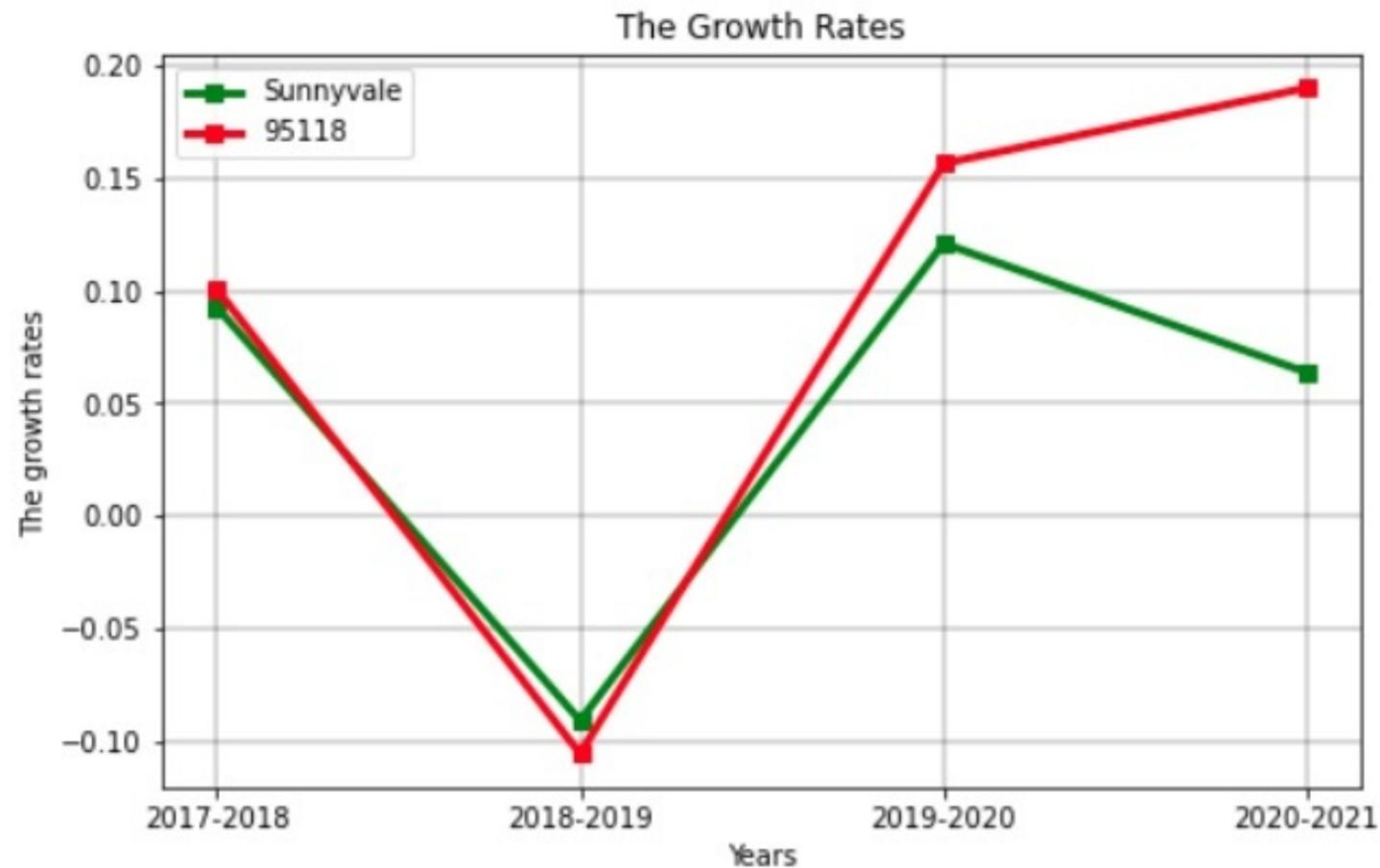


**Sunnyvale**

***The center of Silicon Valley  
nearby 101 highway (San Francisco to LA)***

**95118**

***The furthest place away from Sunnyvale in this  
research***





**COVID and Housing Pricing**

**A**

**B**

**C**

# House Pricing

Chao Wang & Ling Li  
DAMG6105 Final Project





# Conclusion

*Summary*

*Reference*



### Attributes

- Property Type, City/ZIP, Beds, Baths, Price, Square Feet, \$/Square Feet, HOA, Lot Size, Year Built

### During COVID time:

- Housing price is going up
- Transportation doesn't have preference now
- The relationship of yard with housing price is not obvious