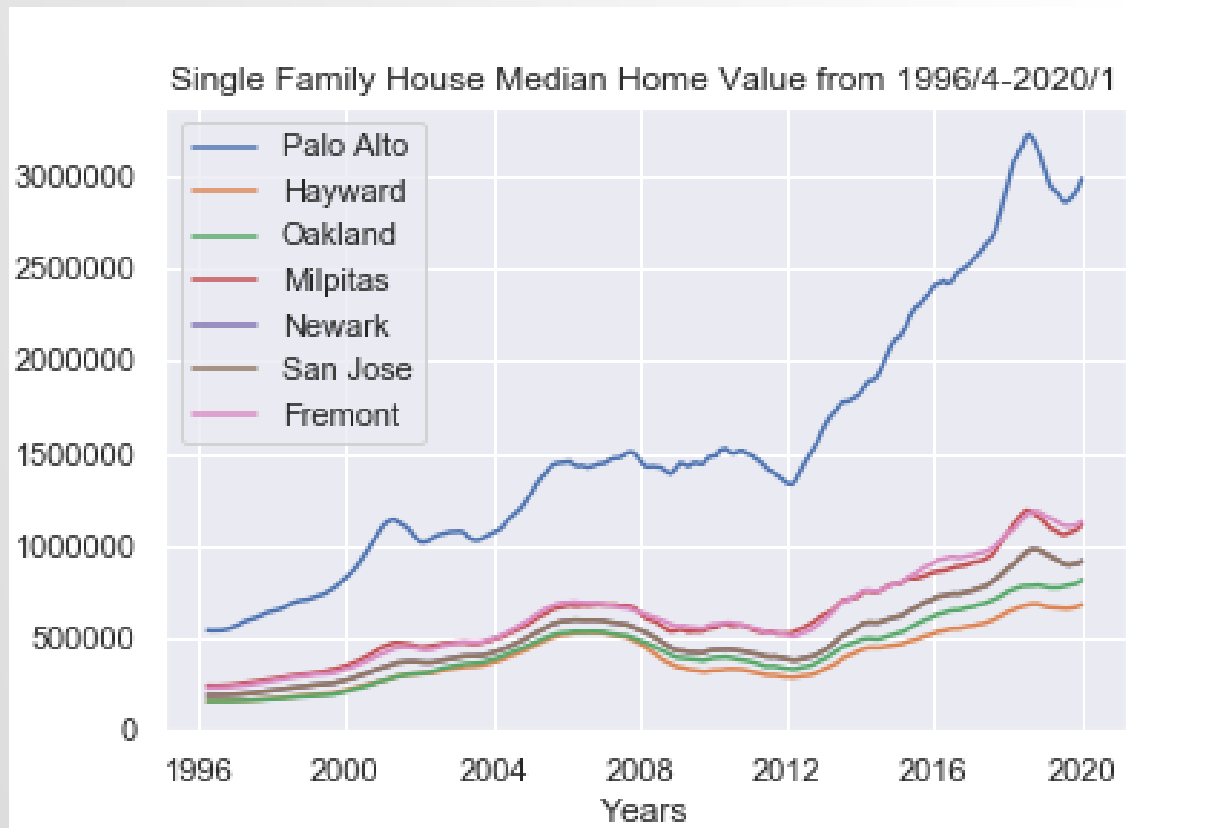# SINGLE FAMILY HOUSE VALUE ANANLYSIS IN FREMONT

By Qi Zhang
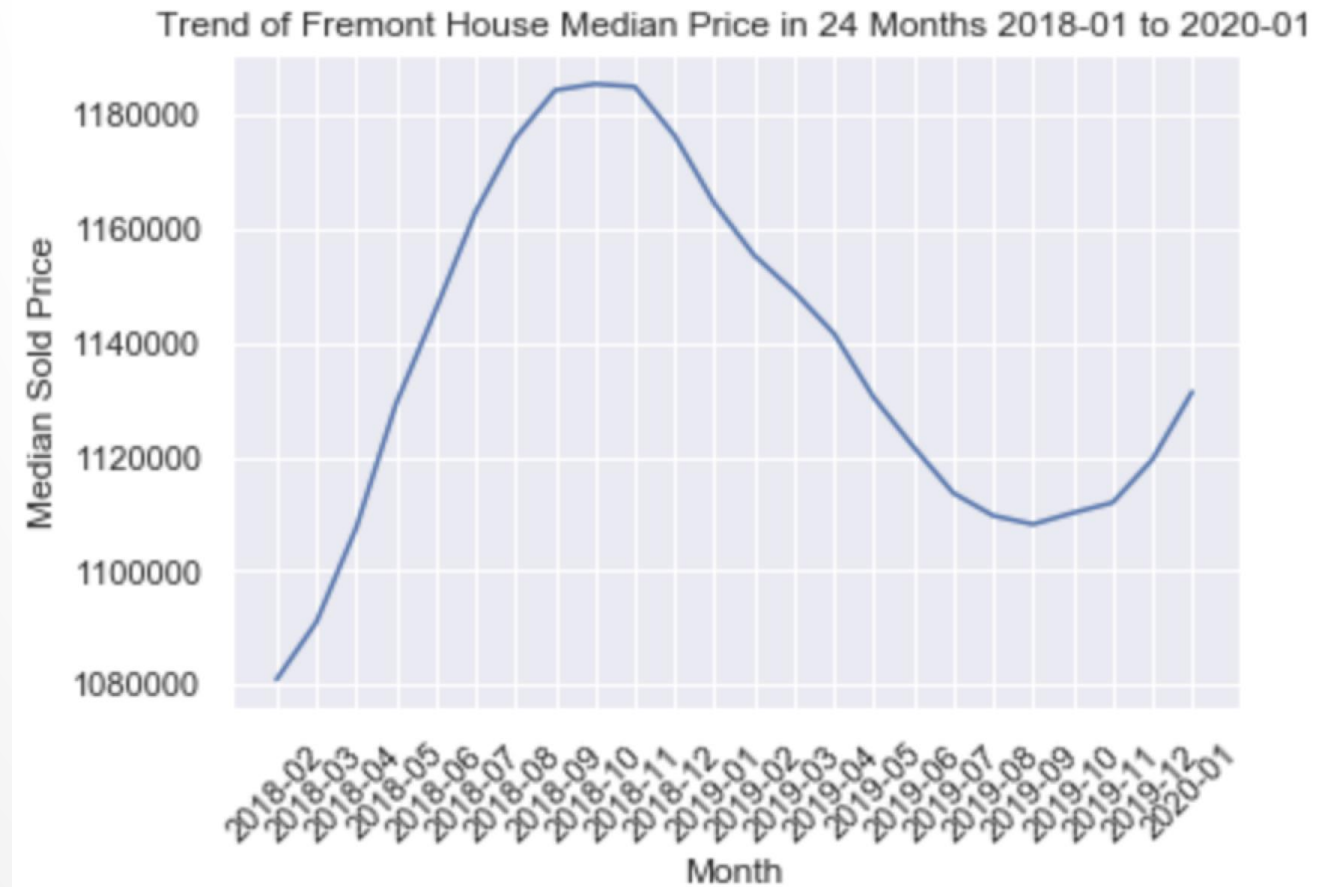
RESEARCH QUESTION 1:
WHAT IS MEDIAN PRICE TREND AMONG 7 CITIES BETWEEN 1996 TO 2020

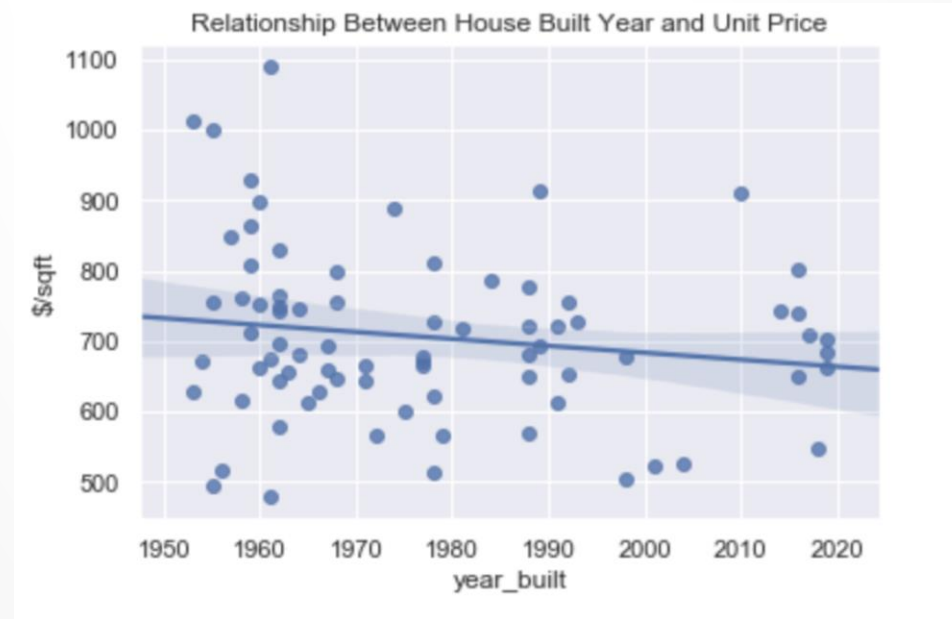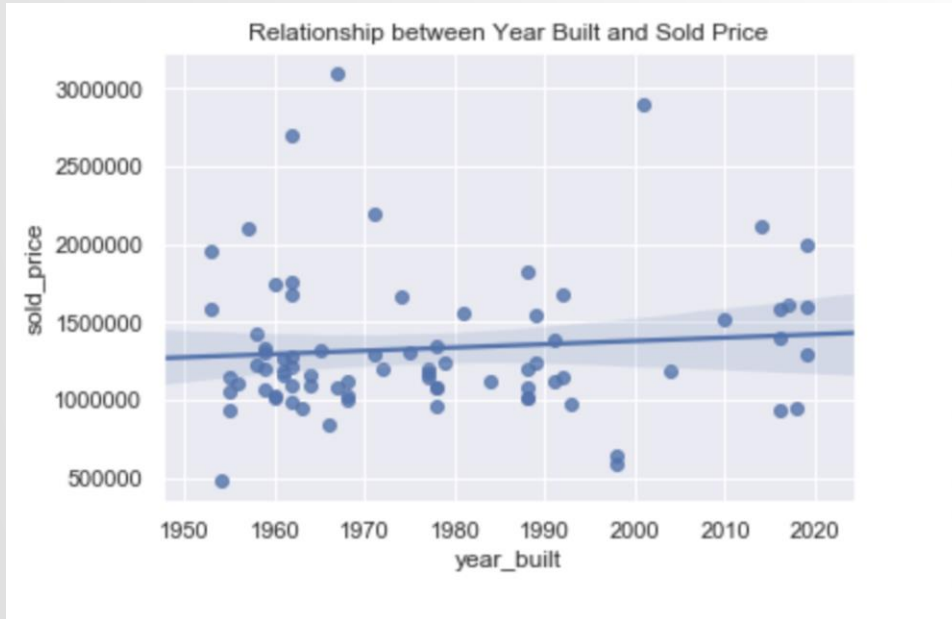Single Family House Median Home Value from 1996/4-2020/1

- House price of all 7 cities from 1996 to 2020, trend is going up as whole

- Palo Alto increased the most.

- Fremont and Milpitas house prices are close

- tendency: Milpitas recently is catching up Fremont.

RESEARCH QUESTION 2:
TREND OF FREMONT HOUSE (1996-2020)

- Soared from around 0.25M to 0.75M between 1996 to 2006 before it started decreasing to local minimum value of roughly 0.5M in 2012.

- It climbed quickly after 2012 to the local maximum value around 1.18 in 2018.

Trend of Fremont House Median Price in 24 Months 2018-01 to 2020-01

RESEARCH QUESTION 3:
WHICH HOUSE HAVE A BETTER POTENTIAL(BASED ON BUILT YEAR)

Relationship between Year Built and Sold Price

Relationship Between House Built Year and Unit Price

- it shows a little positive relatationship between year_built and sold_price.
- negative relatationship between year_built and $/sqft.

BETWEEN YEAR_BUILT AND OTHER FEATURES

Relationship Between House Yard Size and Sold Price

Relationship Between House Living Footage and Year Built
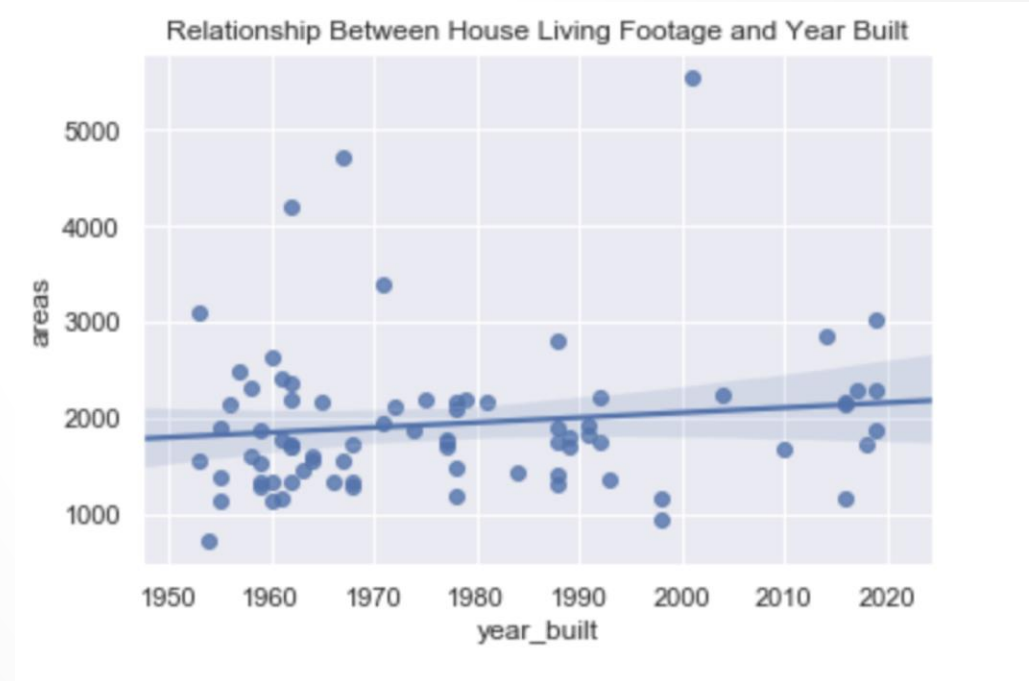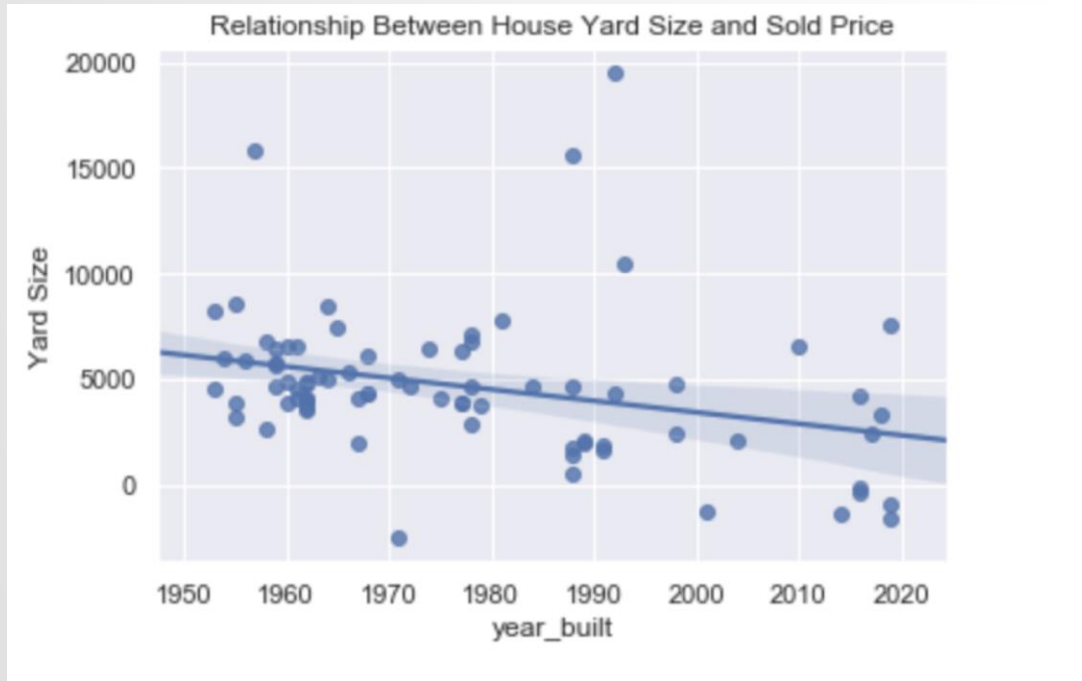
The price of the house is increasing, but the unit price is decreasing yearly. The reason is that living sqft is increased, house yard shrinked. It can be derived that in the future, if the old houses are enlarged, it will have a higher value.

RESEARCH QUESTION 4:

Popularity of Neighborhood

According to badget of a buyer and purpose of the purchase, which neighborhood to choose for buying a SFH by considering title price, $/sqft, school scores?
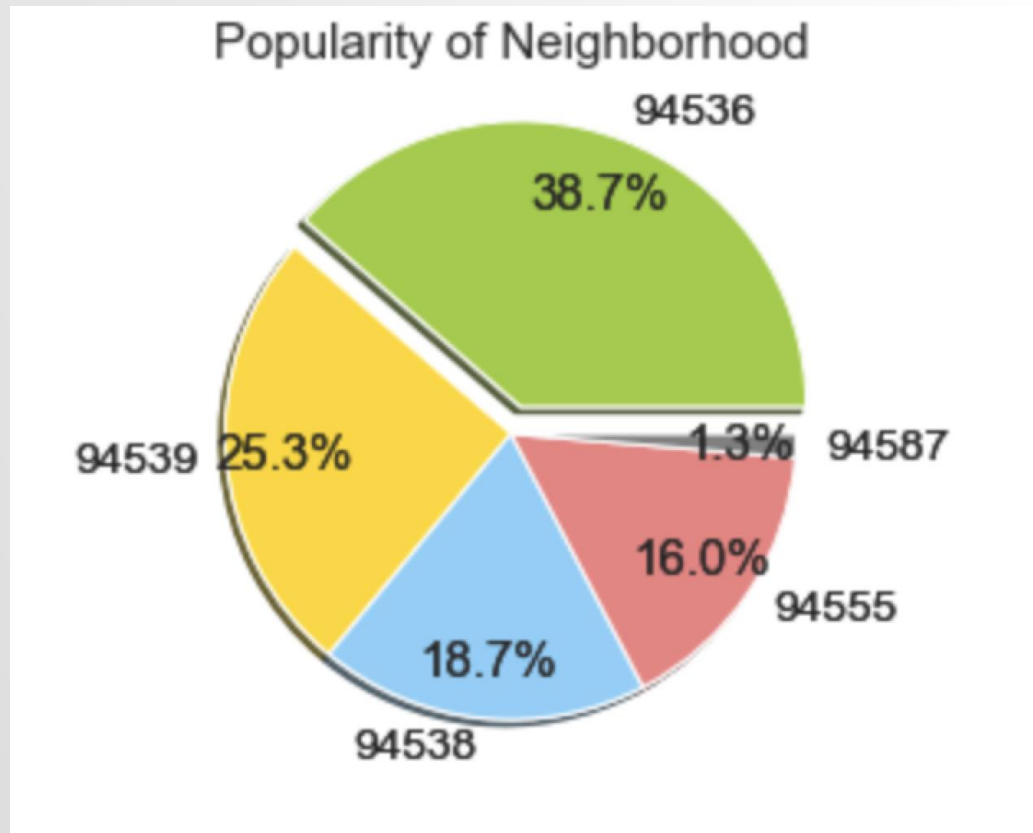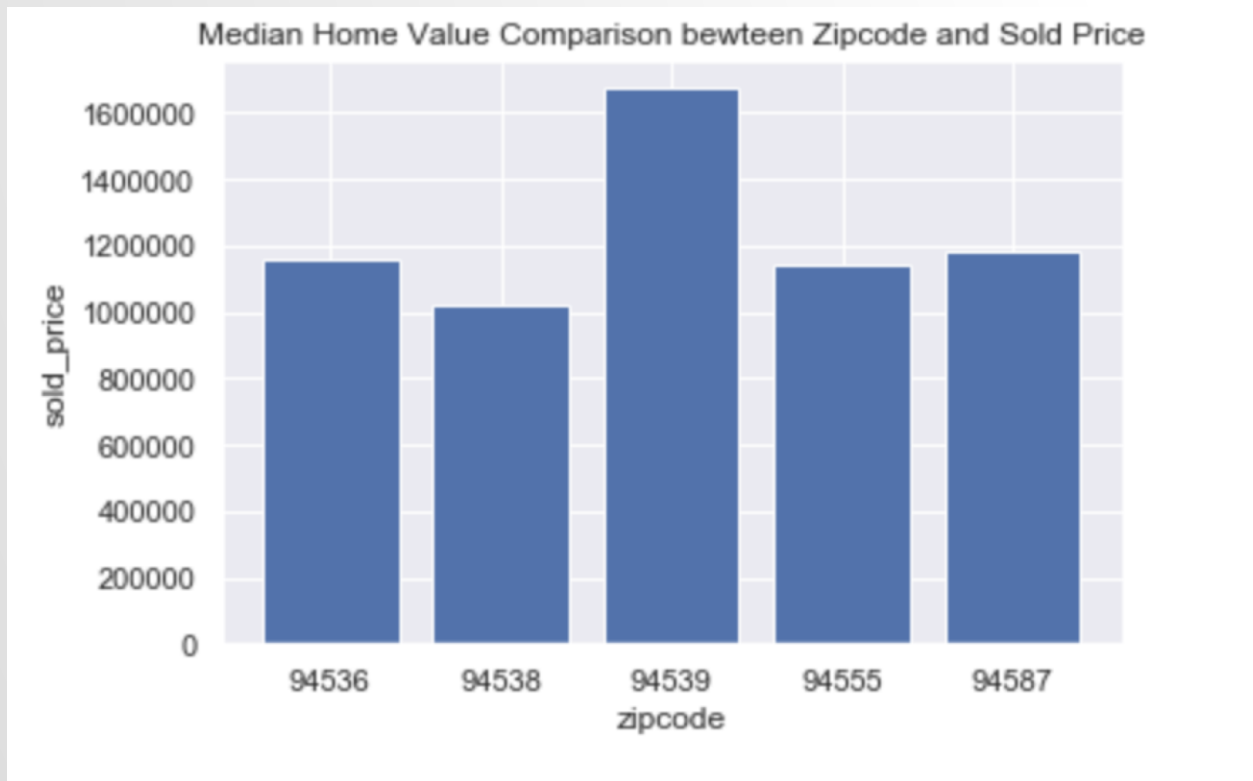
94536: Central Fremont area
94539: South East Fremont area
94538: South West Fremont area
94555: North West Fremont area
93587: North East Fremont area

MEDIAN HOME VALUE AND ZIPCODE

Median Home Value Comparison bewteen Zipcode and Sold Price

94539, south east Fremont, the second most popular area have the highest median price in the Fremont. it may be because the home living footage is larger in the area or simply the commumity and other factor makes it worth more.

$/SQFT AND ZIPCODE

$/sqft Comparison bewteen Zipcode and Area
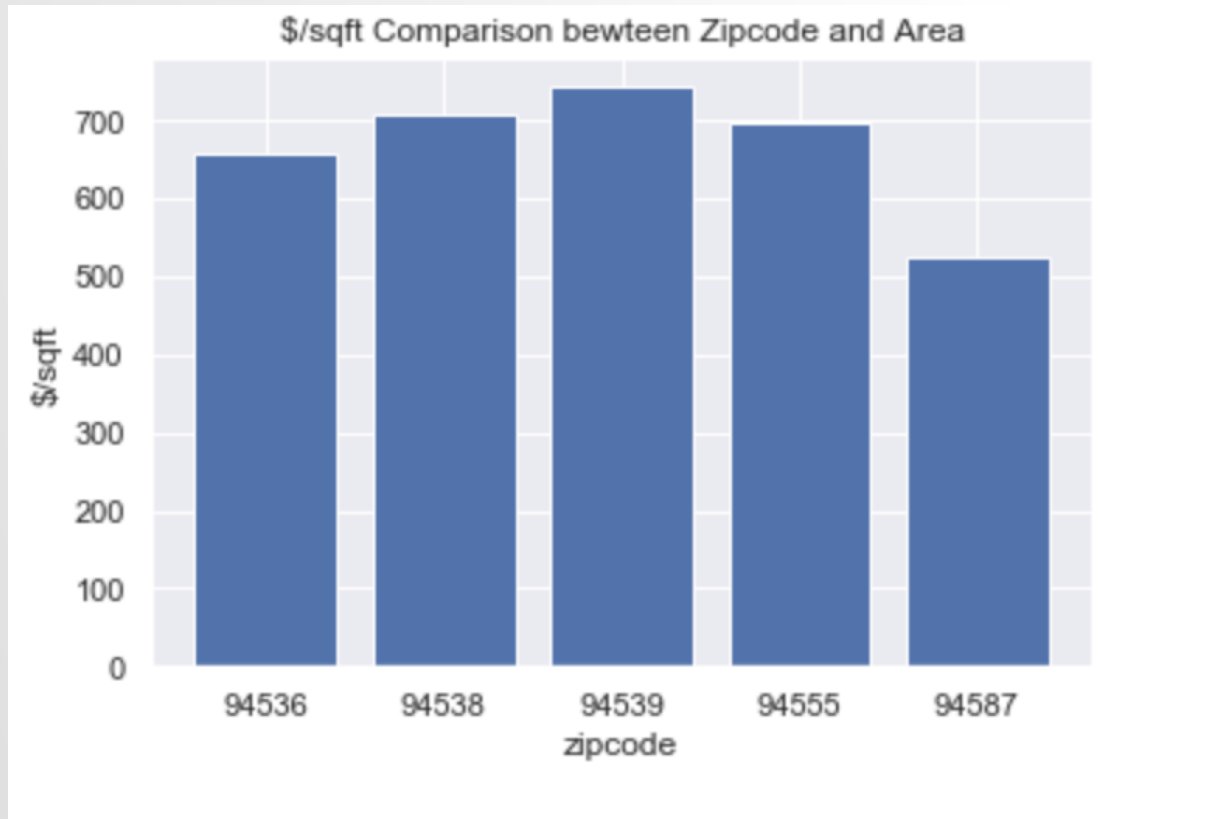
- 94539 has highest unit value and the most expensive total price in Fremont. Second highest is 94538, however, the total price of the house is the lowest. So we can derive from it that the house area is smaller, but if renovate and expand living area, seller can sell a good price. And for buyer, it is more affordable than other area, and the potential is better.

SCHOOL SCORES AND ZIPCODE
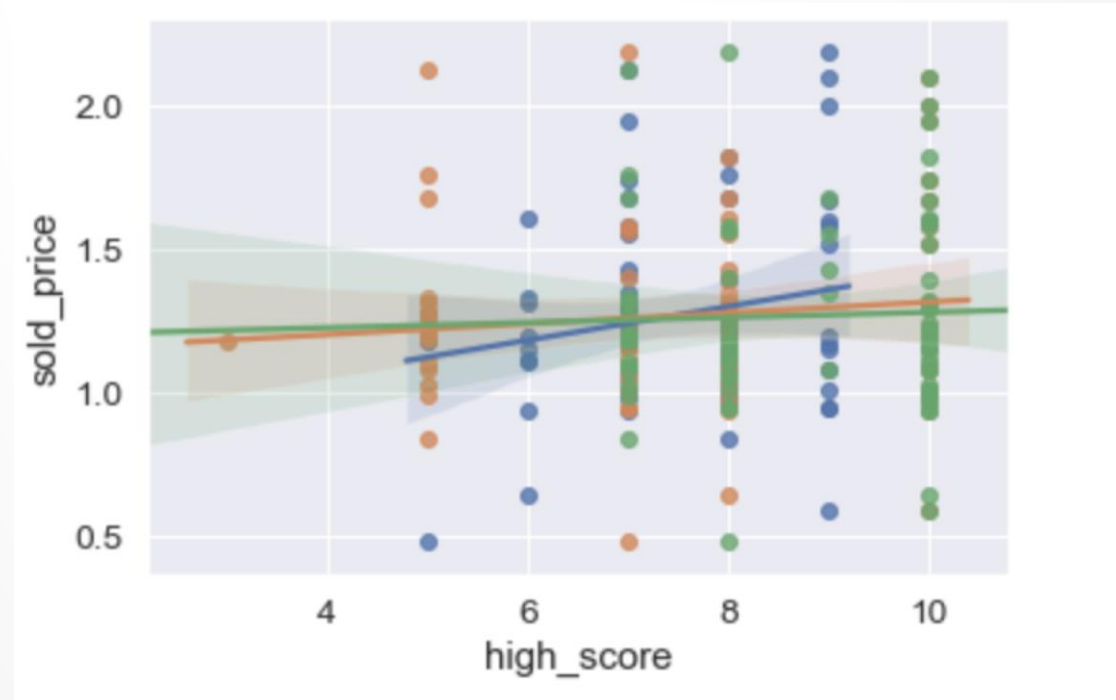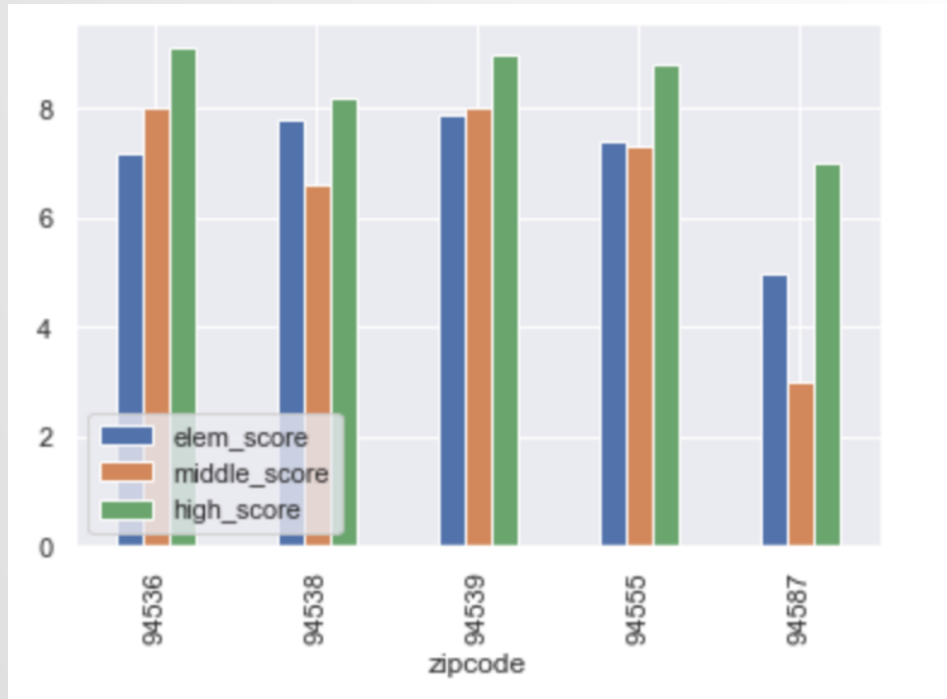
94539 has top schools in Fremont
Better school, hogher the price, or vice versa

ANALYSIS: QUESTION 4

- 94539, top schools, large sqft and highest value, 2nd popular (For family has more budget)

- 94536, top high middle school, and elementary school score is a bit lower, large sqft, most popular, 2nd least expensive $/sqft, large living area. (second choice for family with less budget, prefer good cost performance.)

- 94538, ok schools, smaller house, 2nd expensive. (for family with less people, with limited budget and prefer investment property)

- 94555, good schools, price are in the middle. silimar with 94536.

MODEL DESIGN:

Multicollinearity
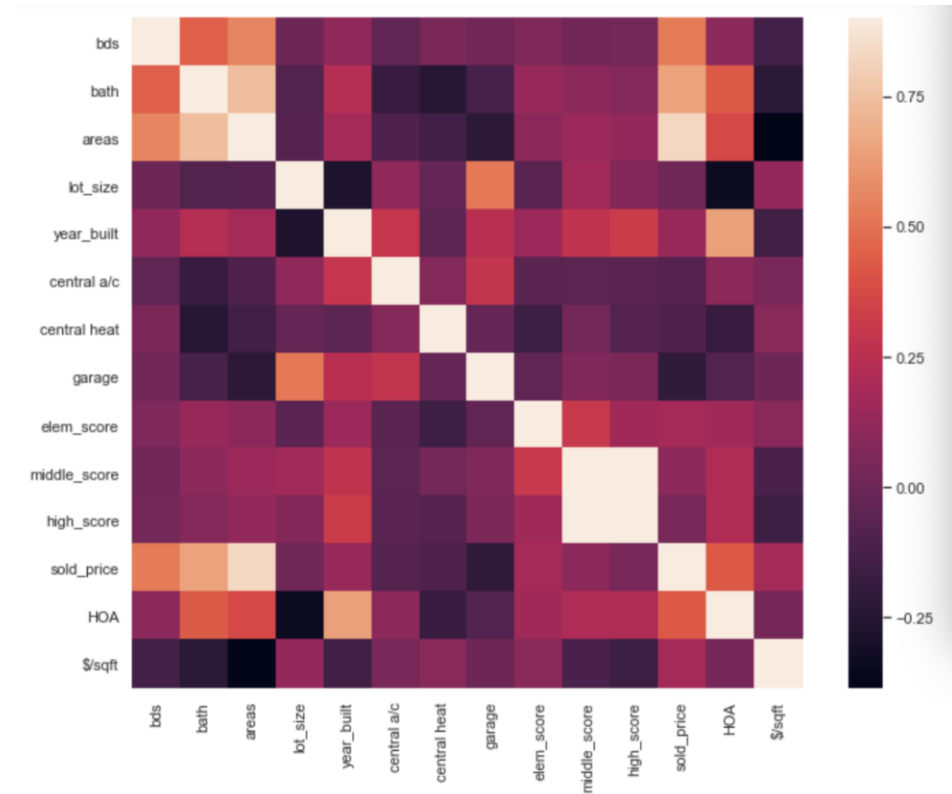
'square feet' and 'baths', 'year built' and 'HOA/month', 'lot_size' and 'garage' indicate a situation of multicollinearity. It also shows that 'bds', 'bath', 'areas' are highly related to sold price.

LINEAR REGRESSION MODEL

```
reg = LinearRegression()
reg.fit(X_train, y_train)
predict = reg.predict(X_test)
```

```
reg.coef_
```

```
array([ 7149.77377955, 53106.139028
        352.89701801])
```

```
reg.intercept_
```

```
65925.04922272707
```

```
reg.score(X_test, y_test)
```

```
0.2694866992503747
```

```
reg.score(X_train, y_train)
```

```
0.7700963545796342
```

- Ression model doesn't fit well

- Need to try some other models on the dataset.

- Maybe the dataset is not large enough, require more data to try again.