**Summary**

Housing market in different cities in the US fluctuates a lot. Housing price is not always related to a single factor. In this project, we uncovered patterns on housing prices in different US cities. We looked on how the city population, income, poverty, and different other factors that we get from census API like education, transportation affects the housing price. And from all these different factors we nominated the best 5 US cities to live in.

1. **Understanding the Barriers of Affordable Housing**
2. **How Housing Prices related to other factors?**

At first, we want to uncover patterns on housing prices in different US cities and want to see how the housing price is related to other factors. For this we looked on the most important parameters for year 2019 which are as follows:

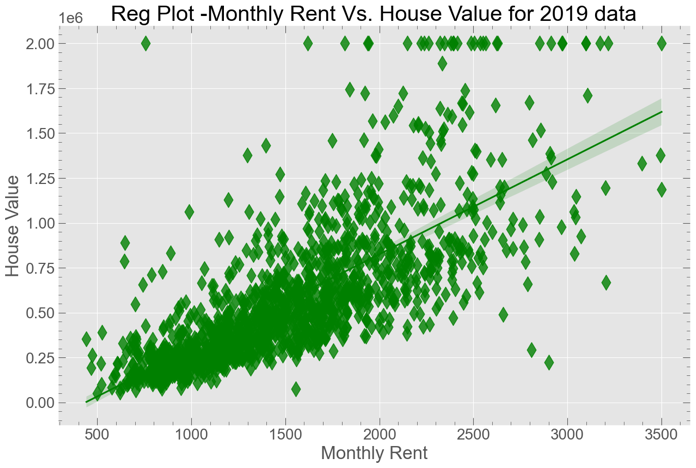
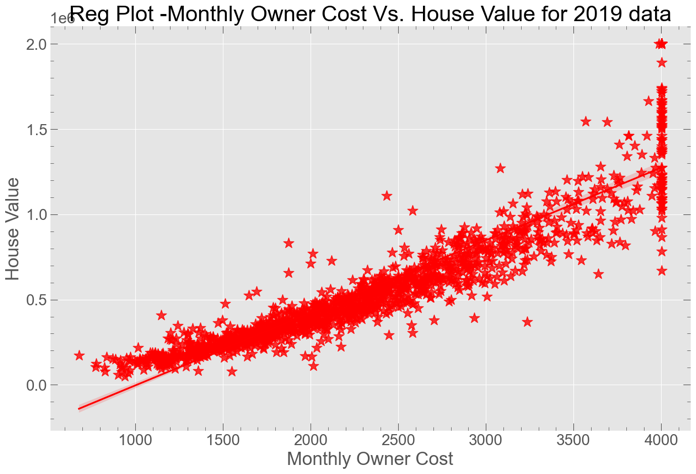
1. **Economy**

Fig 1 shows how the housing price in each city with population size. This figure gives us an idea to see how expensive each city is.

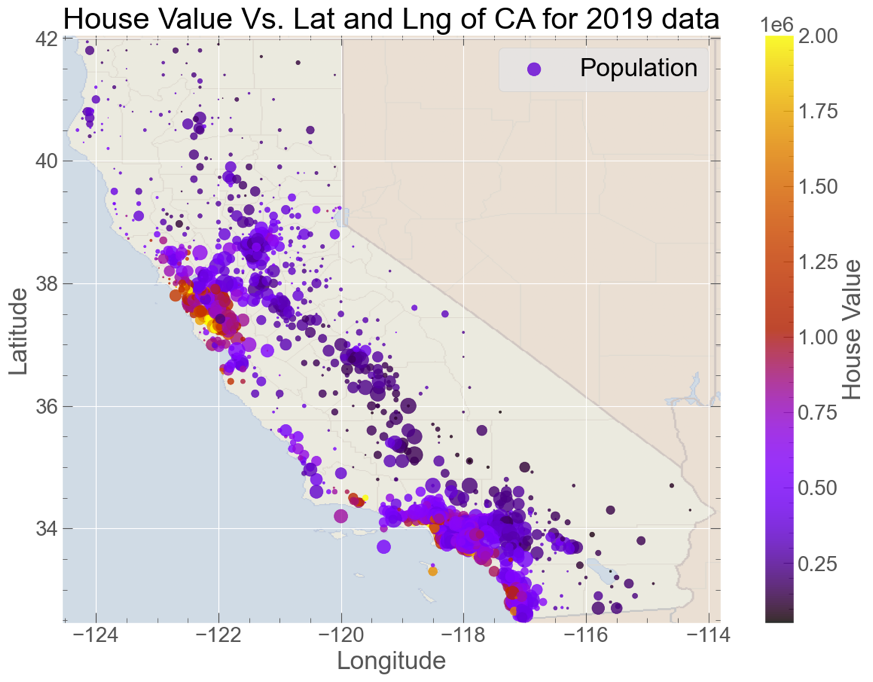
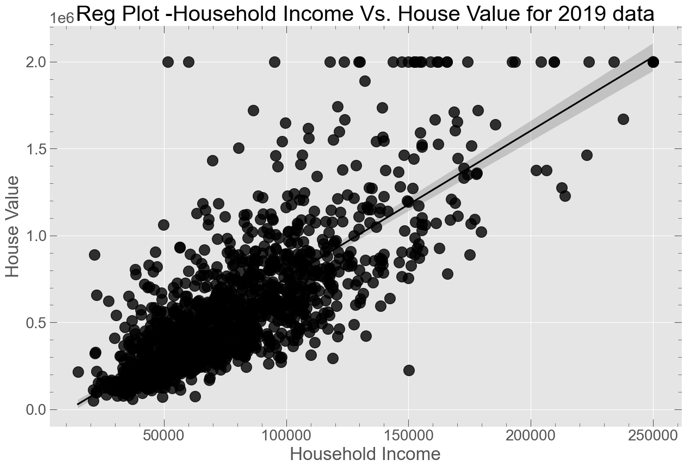
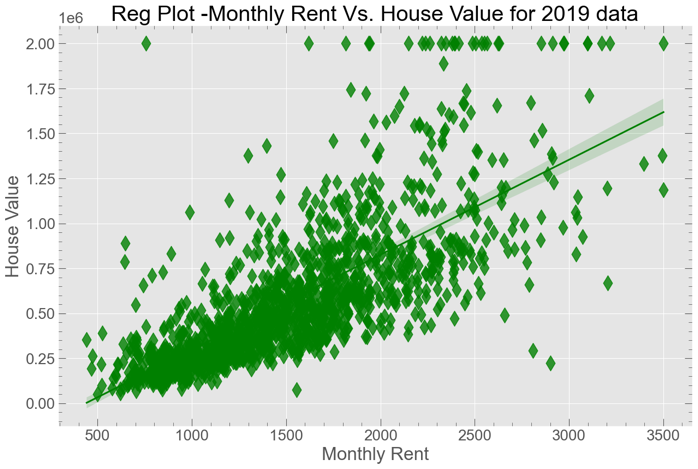


Fig. 1: Latitude and Longitude plot of housing price for each cit

We looked how the household income, monthly rent, poverty rate and monthly owner cost effects the housing value. We saw that there is positive relation with household income, monthly owner cost and monthly rent and negative relation with poverty as shown in Fig. 2.

**** ****

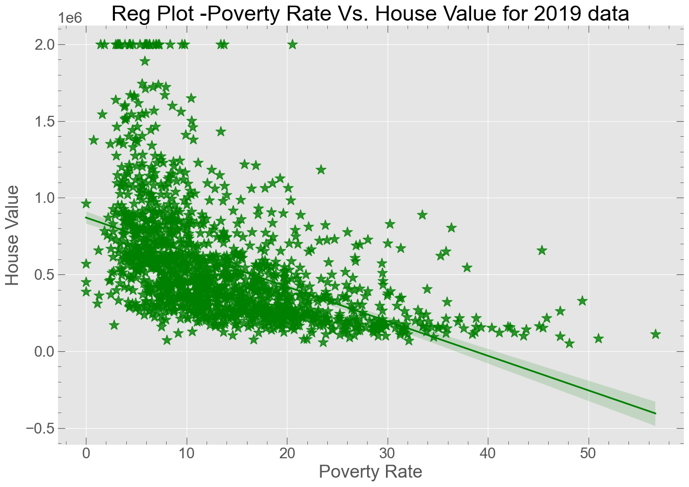
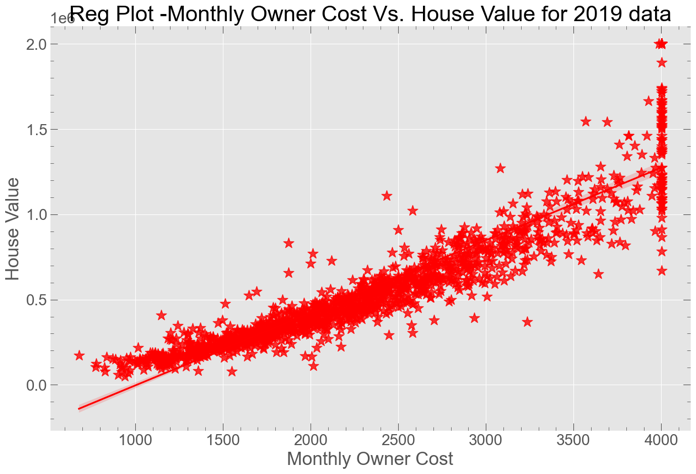
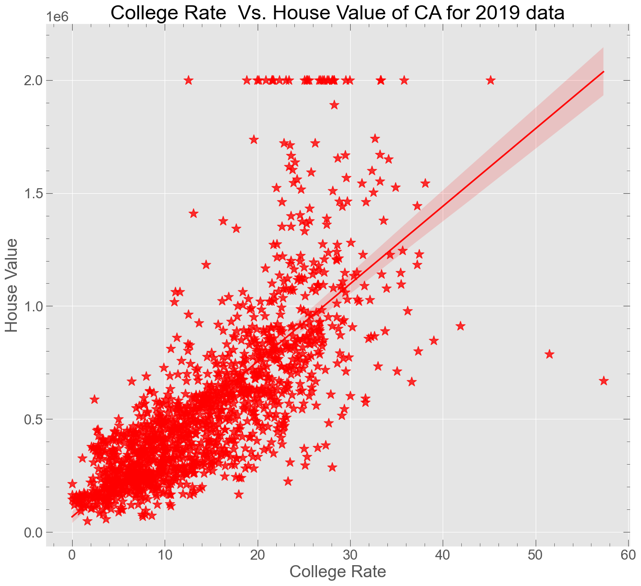
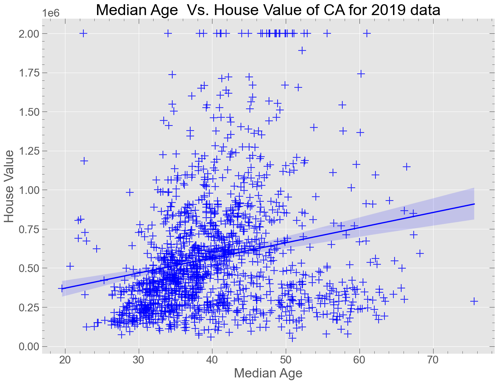
**** ****

Fig. 2: Scatter plots of housing value with household income, monthly rent, poverty rate, and monthly owner cost. The straight line is the fitted data.

1. **Education**

We also explored how the with college rate, Median Age of people, high school rate, and college rate effects the housing value. We saw that there is positive relation with college degree people and median age of people however there is negative relation with high school people rate and uneducated people rate. (see Fig.3).

**** ****

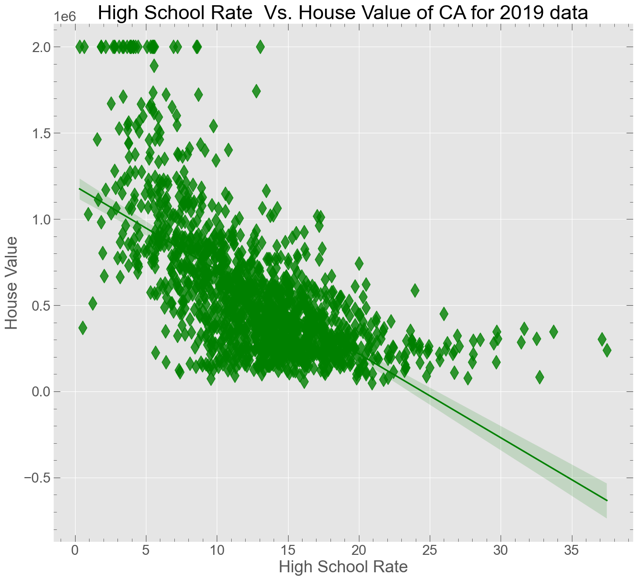
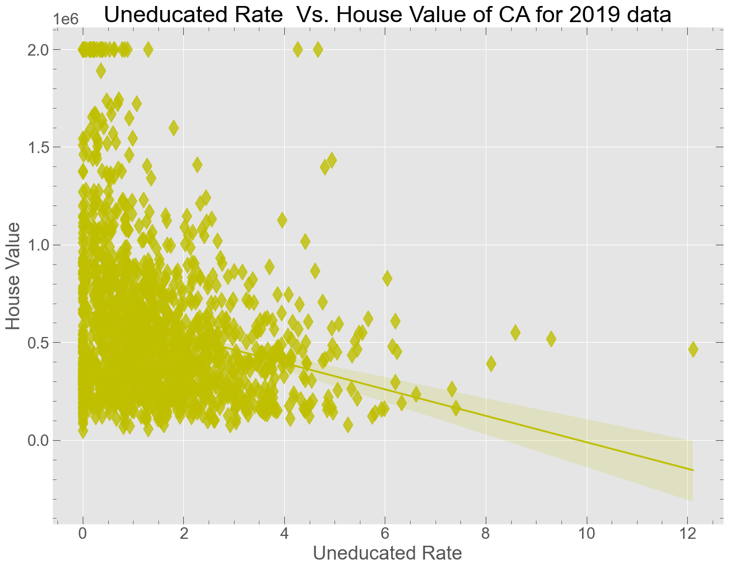
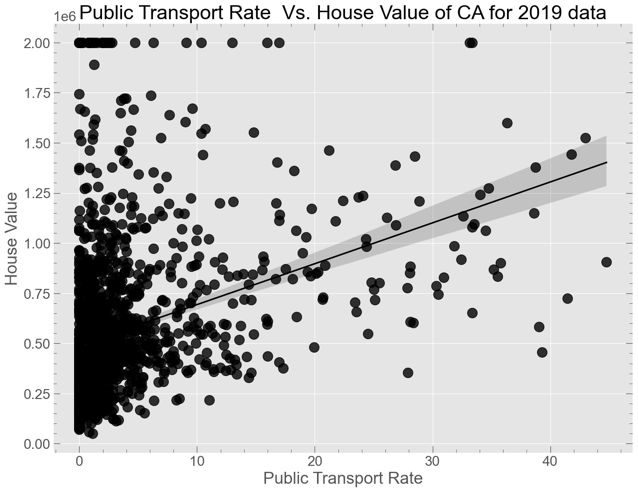
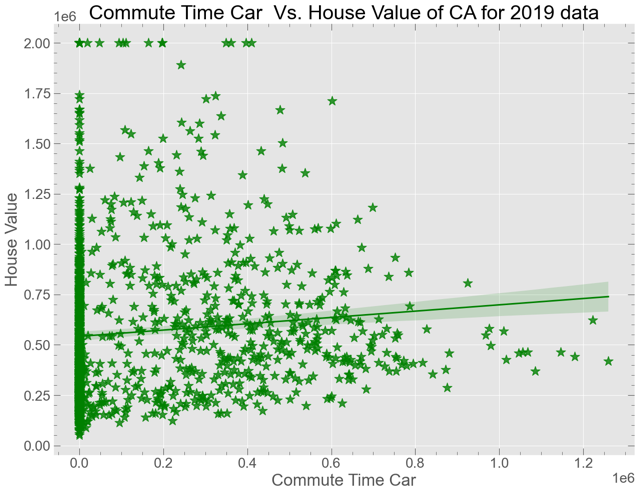
**** ****

Fig. 3: Scatter plots of housing value with college rate, Median Age of people, high school rate, and college rate. The straight line is the fitted result.

1. **Transportation**

For the transportation related parameters, we explored how the public transportation, personal transportation, and commute time of the city effects the housing value. We saw that there is positive relation with public transportation. So that means city having good public transportation has a higher house value. Public transportation and commute time has no effects on the housing value as shown in see Fig.4.

**** ****

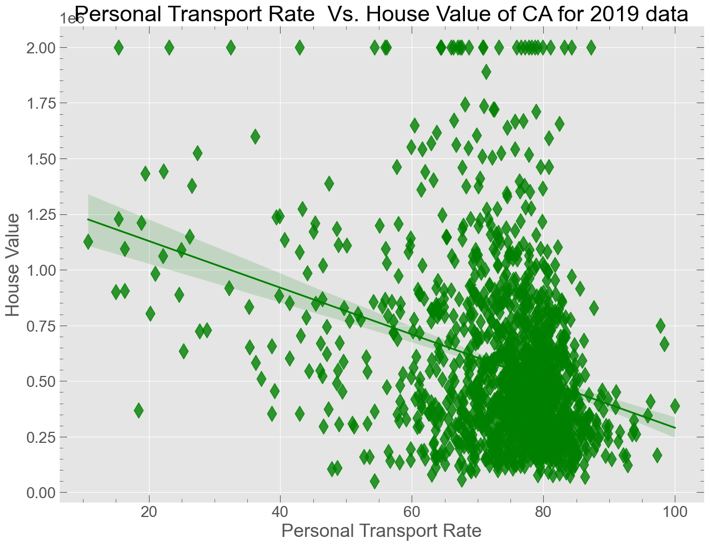
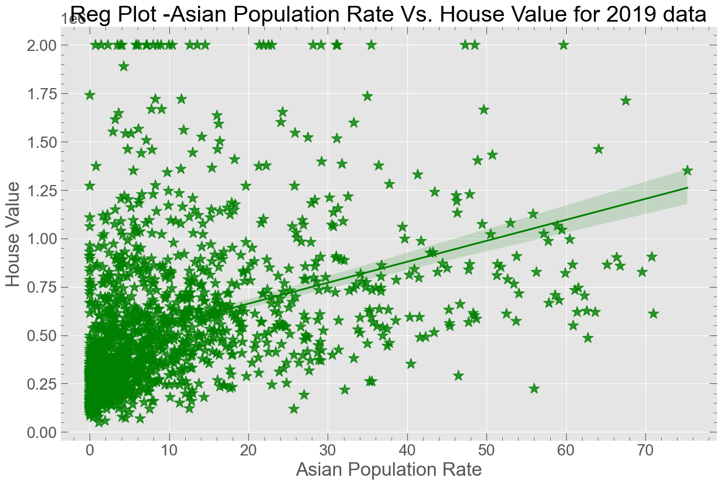
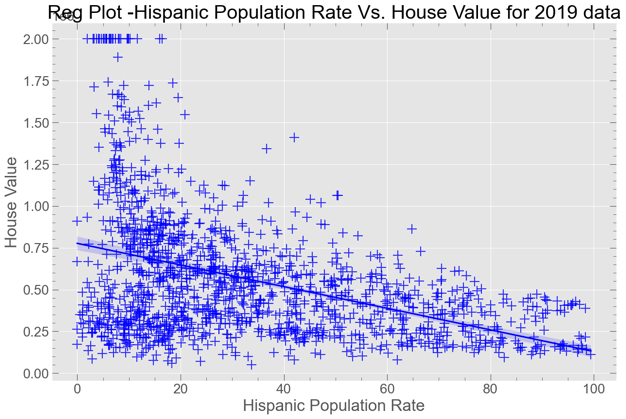
****

Fig. 4: Scatter plots of housing value with public transportation, personal transportation, and commute time of different cities.

1. **Race**

For the race related parameters, we explored how the race of people like, White, Black, Asian, Hispanic race effects the housing value. We saw that there is positive relation with Asian Race people and negative relation with Hispanic race to the housing value. Other race has no effects on the housing value as shown in see Fig.4.

**** ****

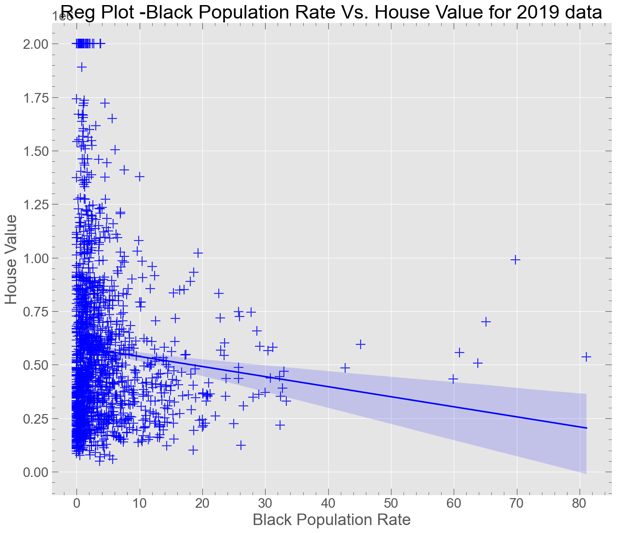
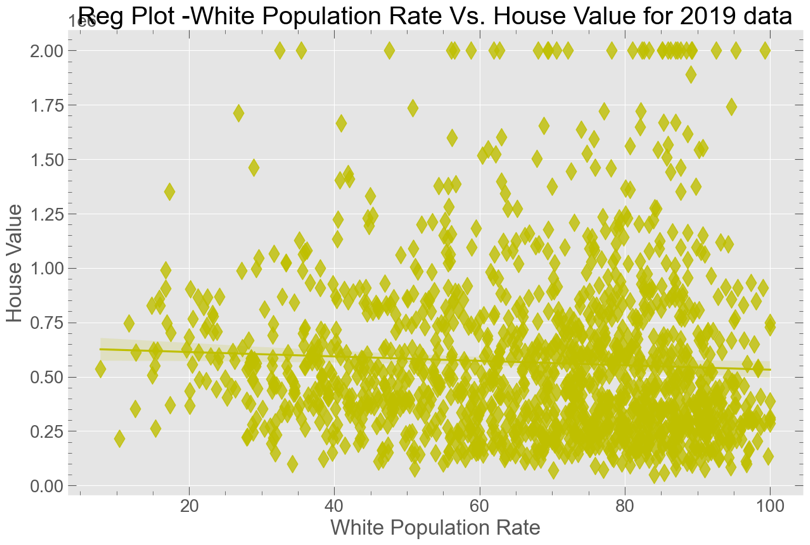
**** ****

Fig. 5: Scatter plots of housing value with different race for different cities.

1. **Understanding the housing price correlation between different terms.**

We also explored on housing price for different years from 2012 to 2019, and see slight increase in the California county house value for each year as shown in Fig. 6. We plotted the correlation of housing price of different years data to different other factors as shown in Fig. 7. From the figure we see a positive correlation with household income, population, monthly owner cost, monthly rent, public transportation, median people age, per capita income, and college rate. However, there is negative correlation of housing price with poverty rate, unemployment rate, and uneducated rate. We also see an important trend for different years data. The correlation of housing price decrease for people median age and college rate for 2019 in compression to 2012 data, which means younger and less educated people are buying the house more.

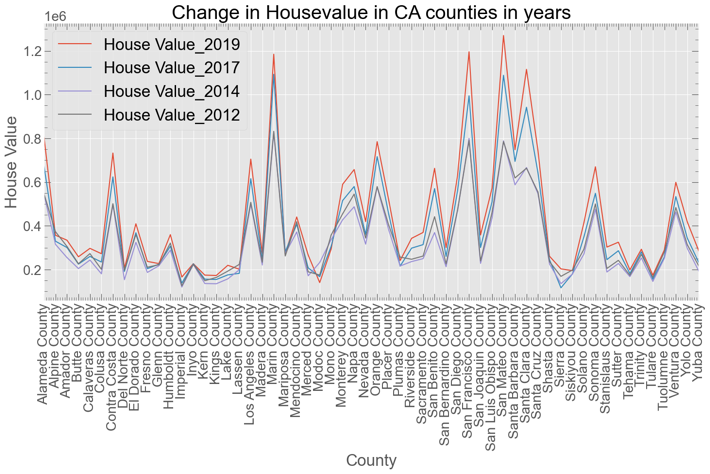
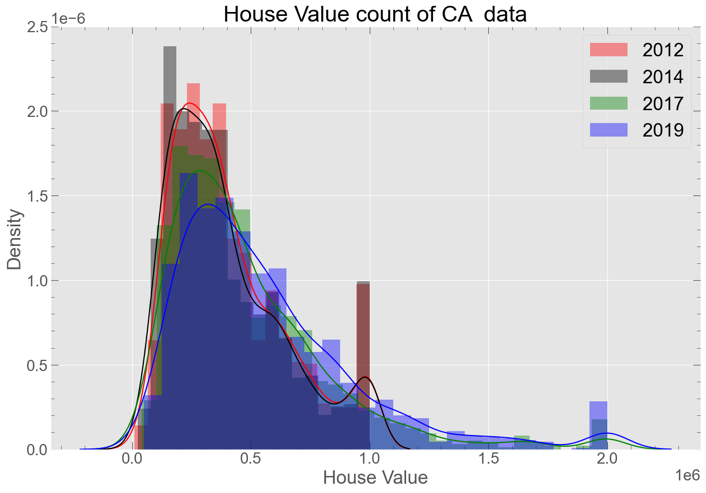
**** ****

Fig. 6: a) Line plot of housing value with County for different years. b) Distplot of house value for different years.

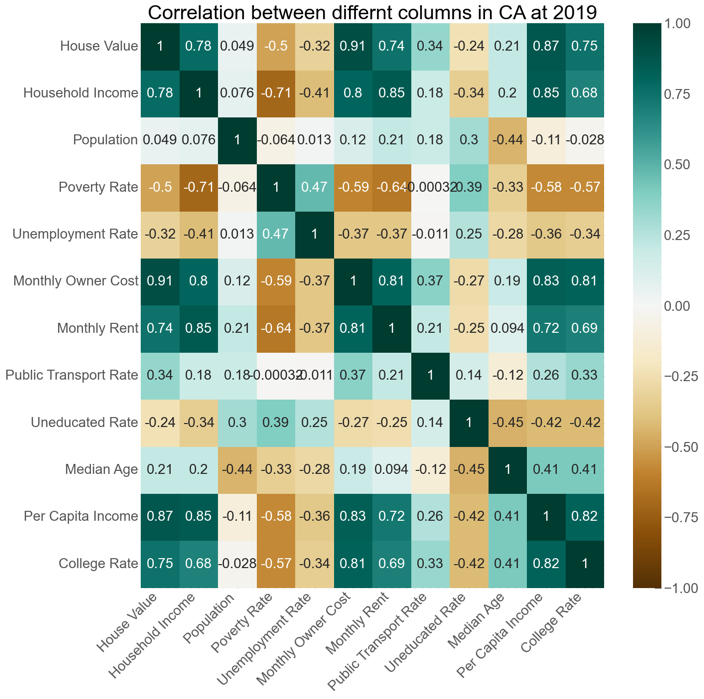
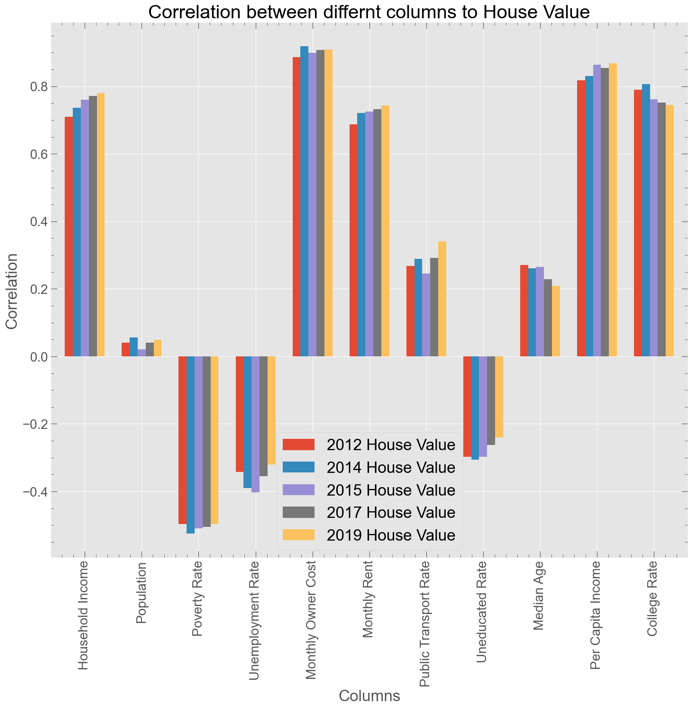
**** ****

Fig. 7: Correlation of house value with other parameters.

1. **Most Affordable 5 Cities of California**

Our team analyzed the best cities to live in based on certain criteria, for some it could be weather, cuisine or endless entertainment for family. Hence, we chose to analyze on some of the most important and common factors an individual would consider. After exploring all these different parameters, we nominated the best cities with the most common factors which are as follows:

* + **House Value**
  + **Income**
  + **Cost/Rent**
  + **Poverty**
  + **Public Transport**
  + **Employment/unemployment rate**

We gave each parameter equal weight and nominated the best cities which satisfy most of the parameters above. The table for best cities for different years is as follows:

We also plotted the best city using google maps API. As we can see, most best cities are in bay area and southern part of California.

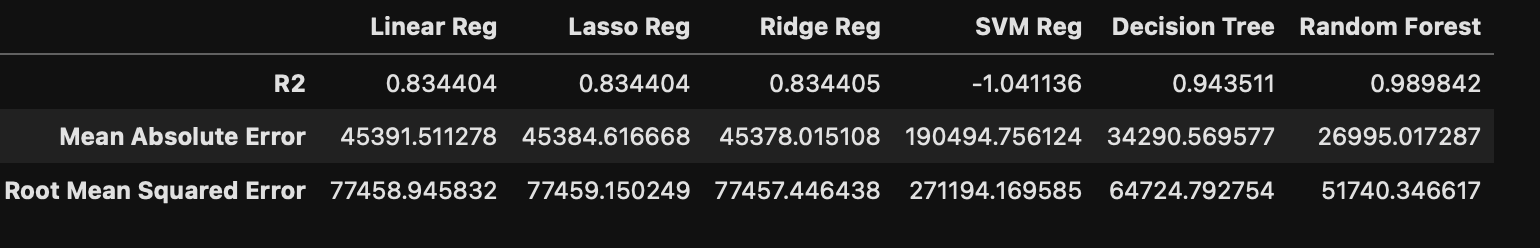
Fig. 8: Our best nominated cities for different years

1. **Model Building**

To predict the house value, we built different supervised linear models from scikit learn. Different models we worked on are as follows:

* **Linear Regression**
* **Lasso Regression**
* **Ridge Regression**
* **Support Vector Machine**
* **Decision Tree Regressor**
* **Random Forest Regressor**

As we can see from the evaluation metrices below, our best model is Random Forest Regressor which has a mean absolute error of around $27000.



From our best model, we predicted the house value of unseen data for 2017. The actual house value and predicted house value from our best model is shown in Fig 9. So our best model successfully predict the housing value of unseen data from 2017 Census with an prediction error of around $27,000

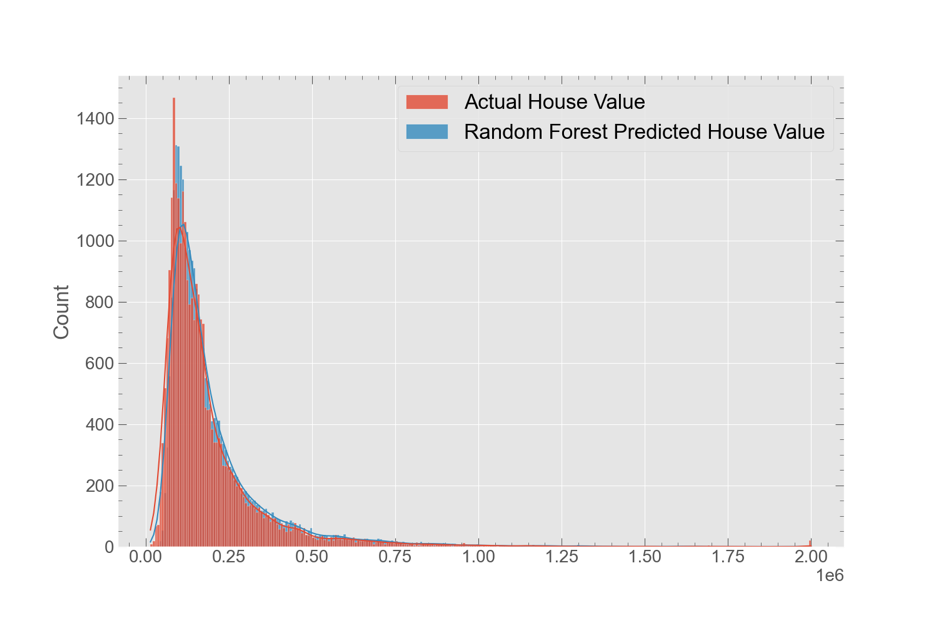
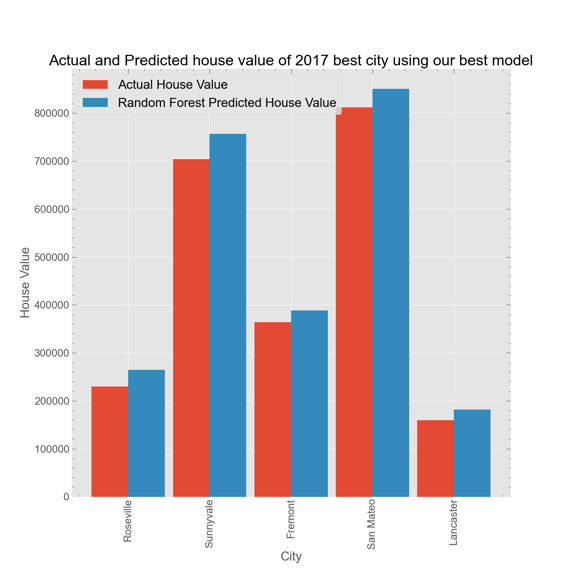
 

Fig. 8: Actual house value and predicted house value from our best model for 2017 Census data.

**Conclusion:**

We used Census API to download the 5 years of US data which has 33000 rows and 30 columns. We looked the relationship of house price to other factors like economy, education, transportation and race. We get some interesting trend of house value on different years census data. We also analyzed the data and recommended the best city of CA based on these different parameters. We also build a model to predict the house value and our best model predicts the house value perfectly with an error of around $30,000. Hence, from this project, we found the most important parameters for the housing price and which parameters are surging the price of houses in the city.  From all these different variables, we will also find the best 5 CA cities to live in. So from our project we can saw that economy, education, and transportation are the important parameters for the housing value and what recommended that these parameters other cities need to improve to be the best city.