\**After discussing with my mentor, I have decided to pick the project 3 from the three projects that I proposed in last weeks submission*.

**Objective**:

To study the impact of change in housing prices in San Francisco on the housing prices of neighboring cities in the east bay. Either Berkeley or Oakland will be chosen based on which city has most housing data available. Due to sky rocketing prices in San Francisco, the housing prices in the neighboring cities are also increase. Due to high cost of living, many residents are moving to the other part of San Francisco bay areas (renter, homeowners, business owners, etc.).

This study can help the homebuyers to access if the move to sell and/or buy the house in the neighboring city is good step in terms of how the prices are changing. In nutshell the project is a housing prediction study that would focus on understanding the change in real estate market based on the past housing price data. Housing price prediction data is quiet valuable for homebuyers, sellers, real estate agents, insurers, etc. A simple model that can provide insight in the future housing market trend can be beneficial for all parties involved.

The data used will the median housing prices of the San Francisco City obtained from Zillow.com. The Zillow website has research data of housing prices in bay area from 1996 to the present. The website has data based on the neighborhoods, cities, counties, zip codes, etc. The study would be focused on either city or county level.

This project will require understanding of data science and time series analysis. As the future predictions will be made based on the past housing prices, understanding the change in housing price trend overtime is crucial. Given the available data, the impact of real estate crisis of 2008 on the current market would be an interesting part to study. In the end, the causal-effect relationship between the housing price time series of the two selected cities will be studied. If a relationship can be established, a prediction model will be proposed based on the finding of time series. If not, the results will be reported as such.

The final deliverables will include a project report, presentation slides and the main codes written in R along with all the data files used. I is expected that the data would be able to provide enough insight to develop (relatively) accurate prediction of housing prices for at least two months in the future.

Additions:

After discussing with my mentor, it was suggested that I also include the impact of few factors other than housing prices of the neighboring cities. Impact of census data, housing inventory, and stock market data will also be considered to understand the future housing price prediction. It is expected that insight from the new features will add to improving the accuracy of the model predictions.