DATA DESCRIPTION:

1. For housing data:

- Date: House sold on the following date
- Price: Price of the house
- Bedrooms: Number of bedrooms
- Bathrooms: Number of bathrooms
- Sqft_living: Total area of the house
- Sqft_lot: Area around the house
- Floors: Number of floors in the house
- View: View from the house
- Condition: Condition of the house
- Grade: Measure of quality of construction
- Sqft_above: Sqft above ground
- Sqft_basement: Sqft in basement
- Yr_built: Year in which the house of built
- Zipcode: Zipcode in which the house is present
- Lat: Latitude
- Long: Longitude
- Sqft_living15: Average sqft of 15 nearest neighbors
- Sqft_lot15: Average lot size of 15 nearest neighbors

2. For crime_data:

We have combined the crimes in a neighborhood and added the zipcode for all the neighborhoods.Then we have combined all the locations by zipcode and produced another dataframe with yearly crime rate of individual zipcodes.

3. Combining housing and crime_data:

We have added crime rates for the neighborhoods with the same zip codes are merged together. The various crimes under a given zip code are consolidated and matched with the same corresponding zip codes in the housing dataset. The unmatched zip codes in the housing dataset are removed.

4. Adding interaction terms

An interaction term between the sqft_living and bedrooms is added. For a given sqft_living value, the increase in the number of bedrooms is negatively correlated to the price.