id - Unique ID for each home sold  
date - Date of the home sale  
price - Price of each home sold  
bedrooms - Number of bedrooms  
bathrooms - Number of bathrooms, where .5 accounts for a room with a toilet but no shower  
sqft\_living - Square footage of the apartments interior living space  
sqft\_lot - Square footage of the land space  
floors - Number of floors  
waterfront - A dummy variable for whether the apartment was overlooking the waterfront or not  
view - An index from 0 to 4 of how good the view of the property was  
condition - An index from 1 to 5 on the condition of the apartment, 分数越高评价越好吗？  
grade - An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.（是不是需要分段进行分析1-3；极差。4-6:差，7：一般，8-10：好 11-13:极好）  
sqft\_above - The square footage of the interior housing space that is above ground level  
sqft\_basement - The square footage of the interior housing space that is below ground level  
yr\_built - The year the house was initially built  
yr\_renovated - The year of the house’s last renovation  
zipcode - What zipcode area the house is in  
lat - Lattitude  
long - Longitude  
sqft\_living15 - The square footage of interior housing living space for the nearest 15 neighbors（是平均值吗？）  
sqft\_lot15 - The square footage of the land lots of the nearest 15 neighbors

verified from 2 sources:  
<https://www.slideshare.net/PawanShivhare1/predicting-king-county-house-prices>  
<https://rstudio-pubs-static.s3.amazonaws.com/155304_cc51f448116744069664b35e7762999f.htm>

Exploration questions

In the opening of the data exploration journey, our group decides to choose the housing dataset as our final project. This dataset includes 21 variables and 21614 records revolved in the numerical, string, as well time, geographic data. In this process, we found that some data is involved in decimal or not consistent, which is difficult for us to understand or analyze. For instance, the number of bathrooms is 0.75, which means that 0.75 or ¾ bath is a bathroom that contains one sink, one toilet, and a shower or a bath. Traditionally, a full bath contains at least one sink, one toilet, a shower, and a bath, so a 0.75 bathroom only has either a shower or a bath. It is the key process to figure out the meanings of each variable and data before data analysis.

Our group will discuss What kinds of variables are aligned with the house price: positive correlation or negative correlation? Explore variables relationship with a price: sqfl\_lot, floor, sqft\_basement, grade, yr\_renovated, yr\_build. Also, If the larger the square footage of the house which includes lot, basement variables, the higher the price. Will the number of floors affects the house price? Which years is the best-selling house built? When are people willing to buy a house? We also want to know if the recently renovated or constructed houses are more popular for the buyer?