## Column Names and descriptions for Kings County Data Set

- [x] id - unique identified for a house

- [x] date Date - house was sold

- [ ] price Price - is prediction target

- [ ] bedrooms Number - of Bedrooms/House

- [ ] bathrooms Number - of bathrooms/bedrooms

- [ ] sqft\_living square - footage of the home

- [ ] sqft\_lot square - footage of the lot

- [ ] floors Total - floors (levels) in house

- [ ] waterfront - House which has a view to a waterfront - has na values

- [ ] view - Has been viewed - has na values

- [ ] condition - How good the condition is ( Overall )

- [ ] grade - overall grade given to the housing unit, based on King County grading system

- [ ] sqft\_above - square footage of house apart from basement

- [ ] sqft\_basement - square footage of the basement

- [ ] yr\_built - Built Year

- [ ] yr\_renovated - Year when house was renovated -has na values

- [ ] zipcode - zip

- [ ] lat - Latitude coordinate

- [ ] long - Longitude coordinate

- [ ] sqft\_living15 - The square footage of interior housing living space for the nearest 15 neighbors

- [ ] sqft\_lot15 - The square footage of the land lots of the nearest 15 neighbors

MODULE 1 PROJECT – BLOG NOTES

Please also write a blog post about one element of the project - it could be the EDA, the feature selection, the choice of visualizations or anything else technical relating to the project. It should be between 800-1500 words and should be targeted at your peers - aspiring data scientists.

OUR GOAL IS TO PREDICT SALES PRICE AS ACCURATELY AS POSSIBLE

* Preliminary guesses as to factors that may be important
  + Bedroom
  + bathrooms
  + Sqft\_living
  + Sqft\_living15
  + Sqft\_lot
  + Sqft\_lot15
  + Waterfront
  + condition

1. Cleaning the data
   1. Before recast all of the data columns, used scatter\_matrix to do preliminary observations
   2. Possible relationships between:
   3. sqft\_living and price