



# KING COUNTY HOUSING PRICES

**project 1 - EDA**

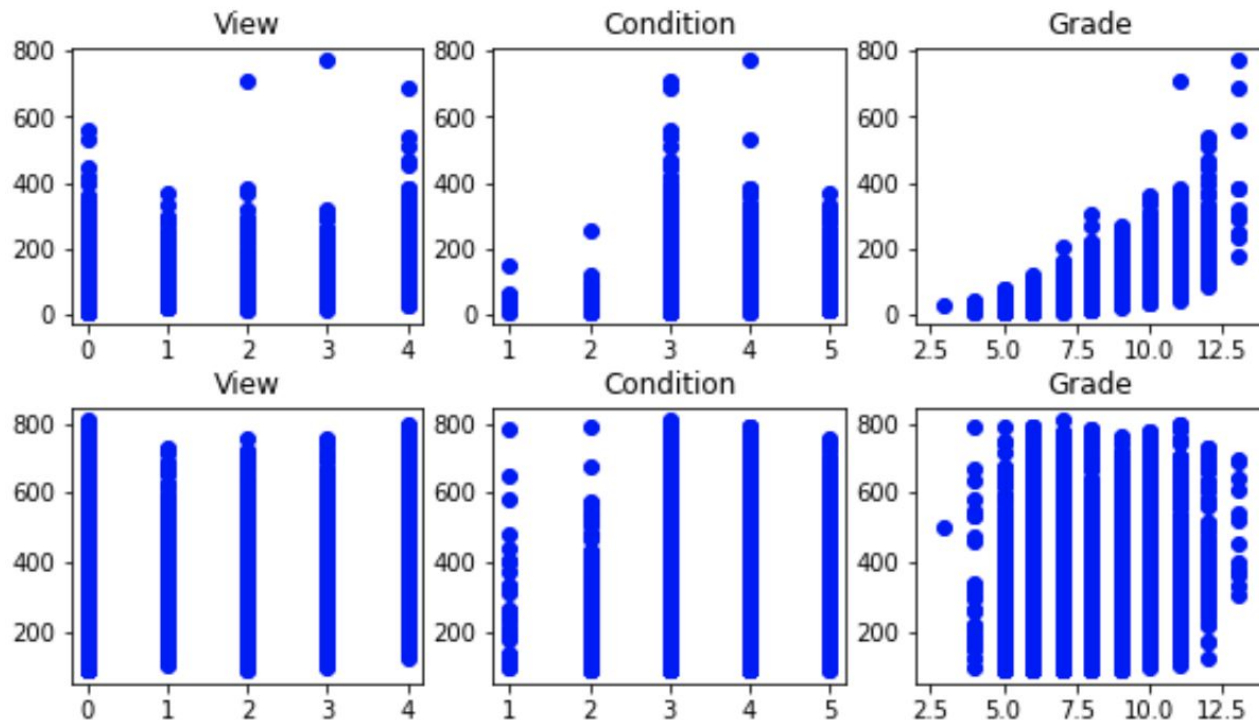
# WHAT THE DATA LOOKS LIKE

- ❑ 21597 properties sold in King County 05/2014-05/2015
- ❑ 21 variables describing each property
- ❑ price in USD as response variable (label)

# DATA CLEANING AND PREPARATION FOR EDA

- ❑ in *waterfront* and *view* replace NaN with zero
- ❑ in *sqr\_living* replace ?'s with calculated values
- ❑ replace *yr\_renovated* with time since last renovation
- ❑ replace *yr\_built* with age of building

# HANDLING CATEGORICAL FEATURES



# BASIC EDA BY LOOKING AT CORRELATIONS

|                |           |
|----------------|-----------|
| price          | 1.000000  |
| sqft_living    | 0.701917  |
| grade          | 0.667951  |
| sqft_above     | 0.605368  |
| sqft_living15  | 0.585241  |
| price_per_sqft | 0.556056  |
| bathrooms      | 0.525906  |
| view           | 0.393497  |
| sqft_basement  | 0.323799  |
| bedrooms       | 0.308787  |
| lat            | 0.306692  |
| waterfront     | 0.264306  |
| floors         | 0.256804  |
| sqft_lot       | 0.089876  |
| sqft_lot15     | 0.082845  |
| condition      | 0.036056  |
| long           | 0.022036  |
| id             | -0.016772 |
| zipcode        | -0.053402 |
| age            | -0.053890 |
| time_since_ren | -0.097462 |

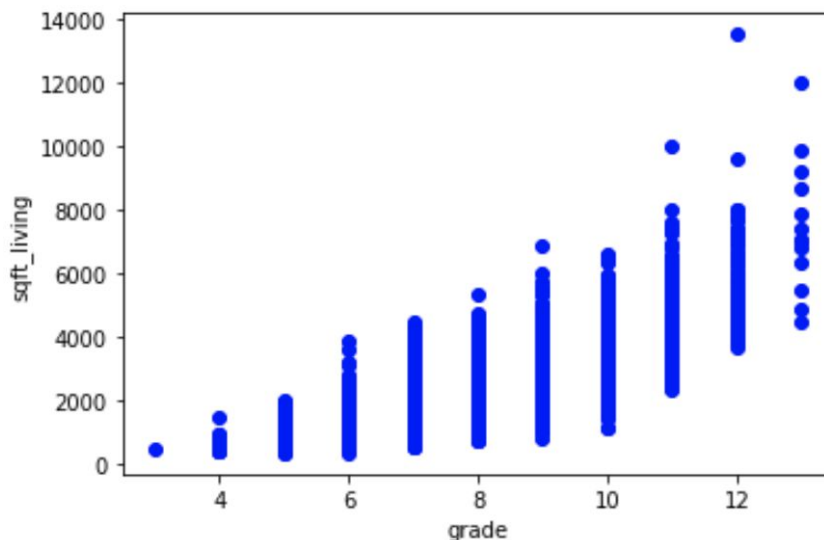
- main drivers of price:
  - sqft related features
  - grade

# BASIC EDA BY LOOKING AT CORRELATIONS

|                |           |
|----------------|-----------|
| price          | 1.000000  |
| sqft_living    | 0.701917  |
| grade          | 0.667951  |
| sqft_above     | 0.605368  |
| sqft_living15  | 0.585241  |
| price_per_sqft | 0.556056  |
| bathrooms      | 0.525906  |
| view           | 0.393497  |
| sqft_basement  | 0.323799  |
| bedrooms       | 0.308787  |
| lat            | 0.306692  |
| waterfront     | 0.264306  |
| floors         | 0.256804  |
| sqft_lot       | 0.089876  |
| sqft_lot15     | 0.082845  |
| condition      | 0.036056  |
| long           | 0.022036  |
| id             | -0.016772 |
| zipcode        | -0.053402 |
| age            | -0.053890 |
| time_since_ren | -0.097462 |

main drivers of price:

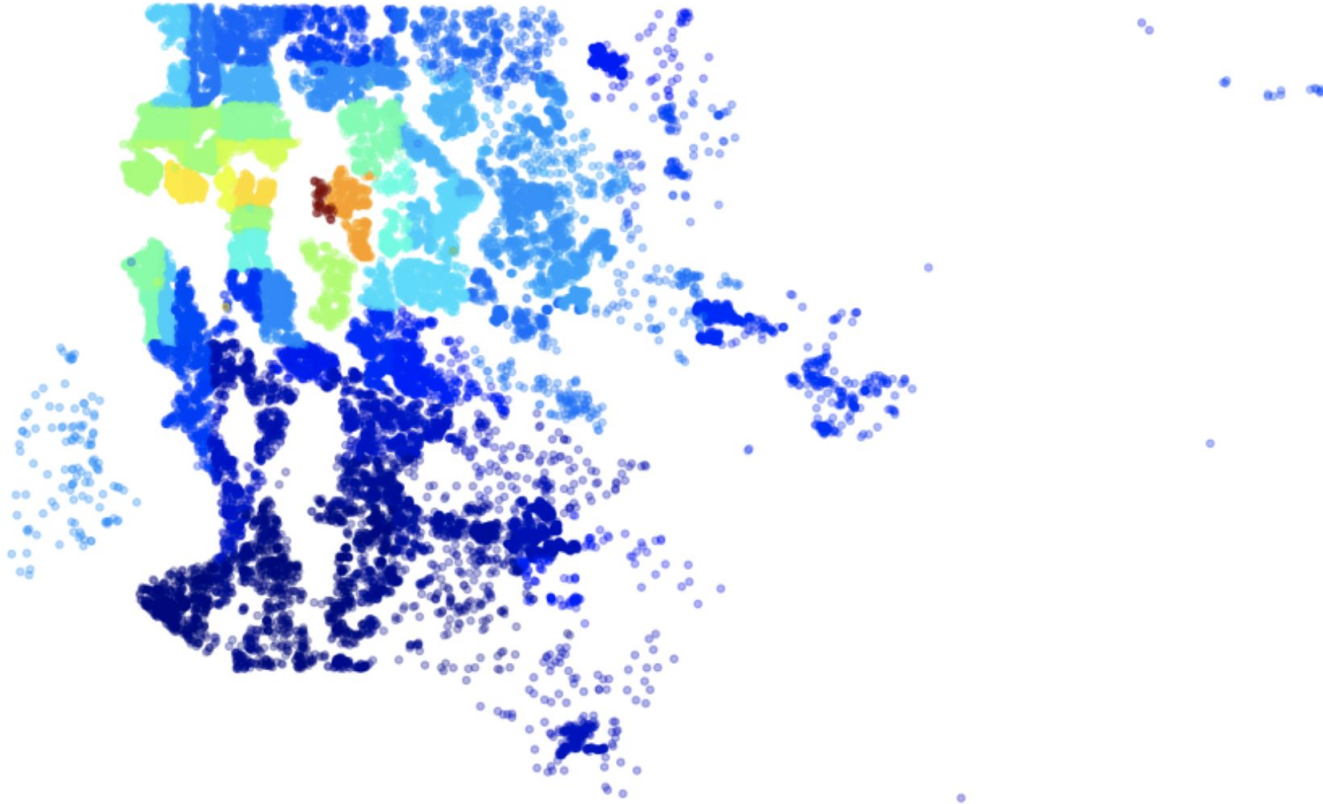
- sqft related features
- grade



# ALTERNATIVE- USE PRICE/SQFT INSTEAD OF PRICE

|                |           |                |           |
|----------------|-----------|----------------|-----------|
| price          | 1.000000  | price_per_sqft | 1.000000  |
| sqft_living    | 0.701917  | price          | 0.556056  |
| grade          | 0.667951  | lat            | 0.472516  |
| sqft_above     | 0.605368  | age            | 0.290756  |
| sqft_living15  | 0.585241  | time_since_ren | 0.255977  |
| price_per_sqft | 0.556056  | view           | 0.220944  |
| bathrooms      | 0.525906  | waterfront     | 0.186646  |
| view           | 0.393497  | zipcode        | 0.172752  |
| sqft_basement  | 0.323799  | grade          | 0.125200  |
| bedrooms       | 0.308787  | condition      | 0.103213  |
| lat            | 0.306692  | sqft_living15  | 0.038999  |
| waterfront     | 0.264306  | floors         | 0.004651  |
| floors         | 0.256804  | id             | -0.005244 |
| sqft_lot       | 0.089876  | sqft_basement  | -0.025360 |
| sqft_lot15     | 0.082845  | sqft_lot       | -0.034749 |
| condition      | 0.036056  | sqft_lot15     | -0.059063 |
| long           | 0.022036  | sqft_above     | -0.087555 |
| id             | -0.016772 | bathrooms      | -0.090285 |
| zipcode        | -0.053402 | sqft_living    | -0.091167 |
| age            | -0.053890 | bedrooms       | -0.205905 |
| time_since_ren | -0.097462 | long           | -0.236594 |

# A MAJOR FACTOR OF PRICE/SQFT IS LOCATION





# FEATURE ENGINEERING

- ❑ zipcode based features
  - ❑ average grade
  - ❑ average condition
  - ❑ average size of properties
  - ❑ average price/sqft
- ❑ location based features
  - ❑ distance to Seattle
  - ❑ ...

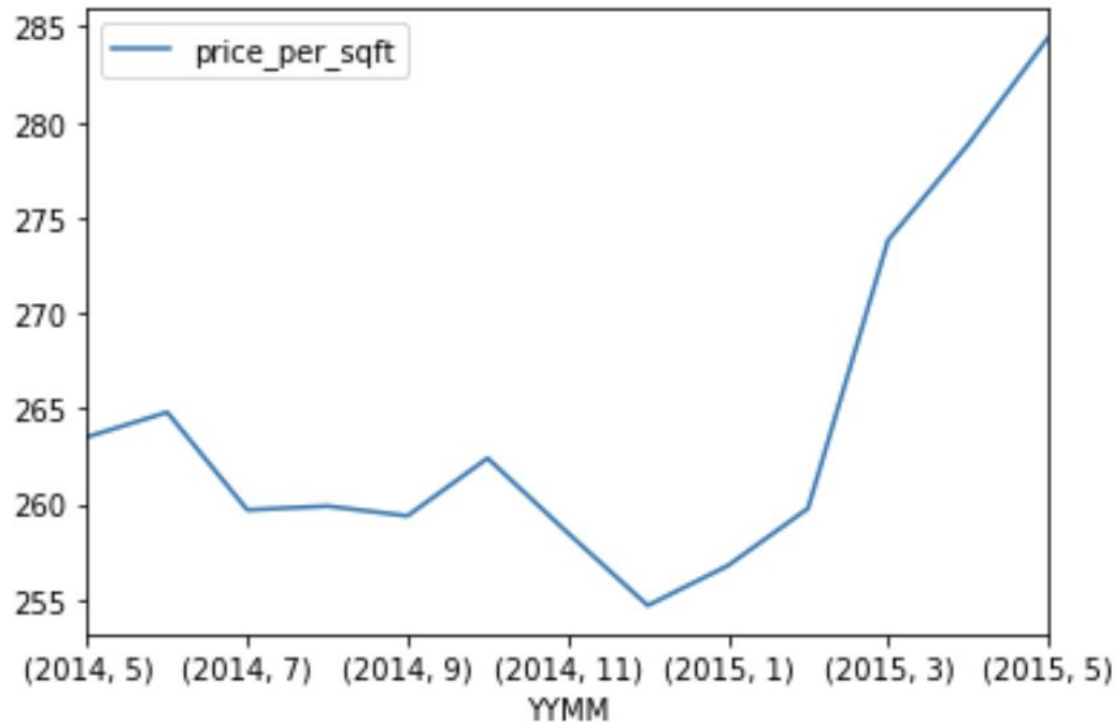
|                 |           |
|-----------------|-----------|
| price_per_sqft  | 1.000000  |
| price_in_zip    | 0.743671  |
| price           | 0.556056  |
| zip_liv_lot     | 0.476935  |
| lat             | 0.472516  |
| zip_grade       | 0.293809  |
| age             | 0.290756  |
| time_since_ren  | 0.255977  |
| zip_condition   | 0.209179  |
| .               | - - - - - |
| comp_livlot     | -0.183996 |
| liv_bigger      | -0.194377 |
| bedrooms        | -0.205905 |
| long            | -0.236594 |
| dist_to_seattle | -0.537458 |
| zipdistance     | -0.538267 |

# PRICE INCREASE OF PROPERTIES

- ❑ 175 properties in the data set have been sold twice
- ❑ *grade, condition or renovation* did not change
- ❑ possible reasons:
  - ❑ strong price fluctuations
  - ❑ people sell their properties cheap

|              |            |
|--------------|------------|
| <b>count</b> | 175.000000 |
| <b>mean</b>  | 0.568851   |
| <b>std</b>   | 0.467654   |
| <b>min</b>   | -0.054054  |
| <b>25%</b>   | 0.234590   |
| <b>50%</b>   | 0.545455   |
| <b>75%</b>   | 0.744355   |
| <b>max</b>   | 3.218009   |

# PRICE FLUCTUATIONS



THANK YOU!

HAPPY TO ANSWER AND  
DISCUSS QUESTIONS

