Residential Real Estate Sales in King County WA, circa 2015: Fluctuations in sale price by geolocation

Michael Collins, Flatiron School
2020-10-30

Residential Real Estate Sales Transactions that occurred in King County, Washington, USA during the dates of May 2, 2014 through May 27, 2015, were studied.

A total of 21,597 transactions were included in the study. The supplied Information about each transaction is summarized on the next slide.

id
date
price
bedrooms
bathrooms
sqft_living
sqft_lot
floors
waterfront
view
condition
grade
sqft_above
sqft_basement
yr_built
yr_renovated
zipcode
lat
long
sqft_living15
sqft_lot15

My goal:

Figure out the marginal financial value of certain physical features and "Scores" assigned to a particular residential property

My approach:

- Consider the (geographic) location of the property that corresponds to each Transaction T_i
- Determine the mean sales price of the 15 transactions nearest to Transaction T_i (excluding transactions that occurred at the SAME location as T_i)
- Use OLS to measure the (marginal) effect of Features F_ij on the relative sales price P_i

For each transaction i, the percent deviation in the price p(i) from the local mean price $p^{bar}(i,k)$

is modeled as the linear combination of the DEVIATION of each of that property's features $f_j(i)$ from the local mean of that feature, $f^{bar}_j(i,k)$.

$$\left\{100\left(\frac{p_i - \overline{p}_{ik}}{\overline{p}_{ik}}\right)\right\} = \sum_{j=0}^{J-1} \beta_j \left[f_j(i) - \overline{f}_j(i,k)\right]$$

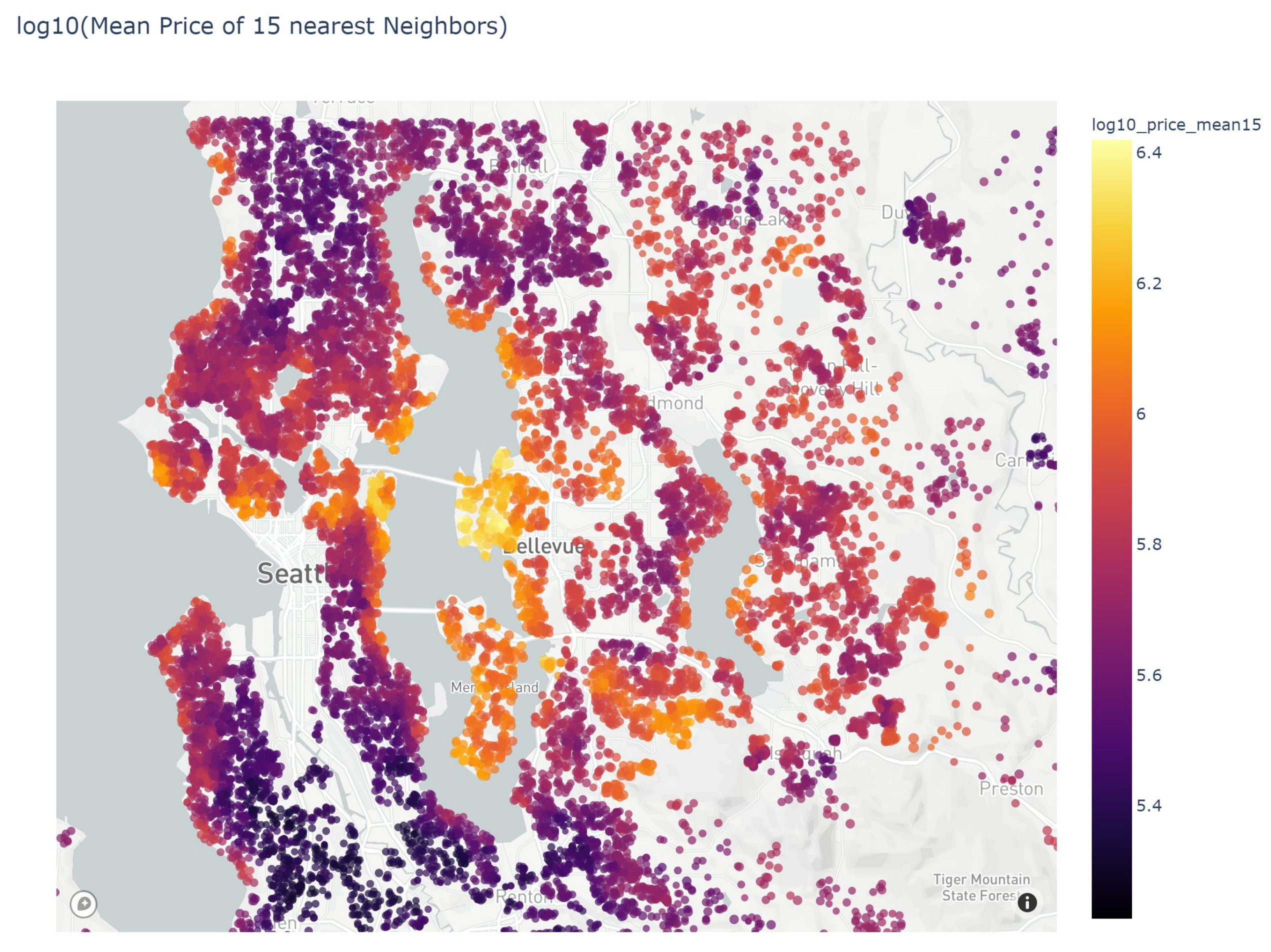
Dividing each side of the previous equation by 100 does the following:

It constrains “beta”, so that B(j) measures the LOCAL marginal effect of feature fj on the RELATIVE value of property i.

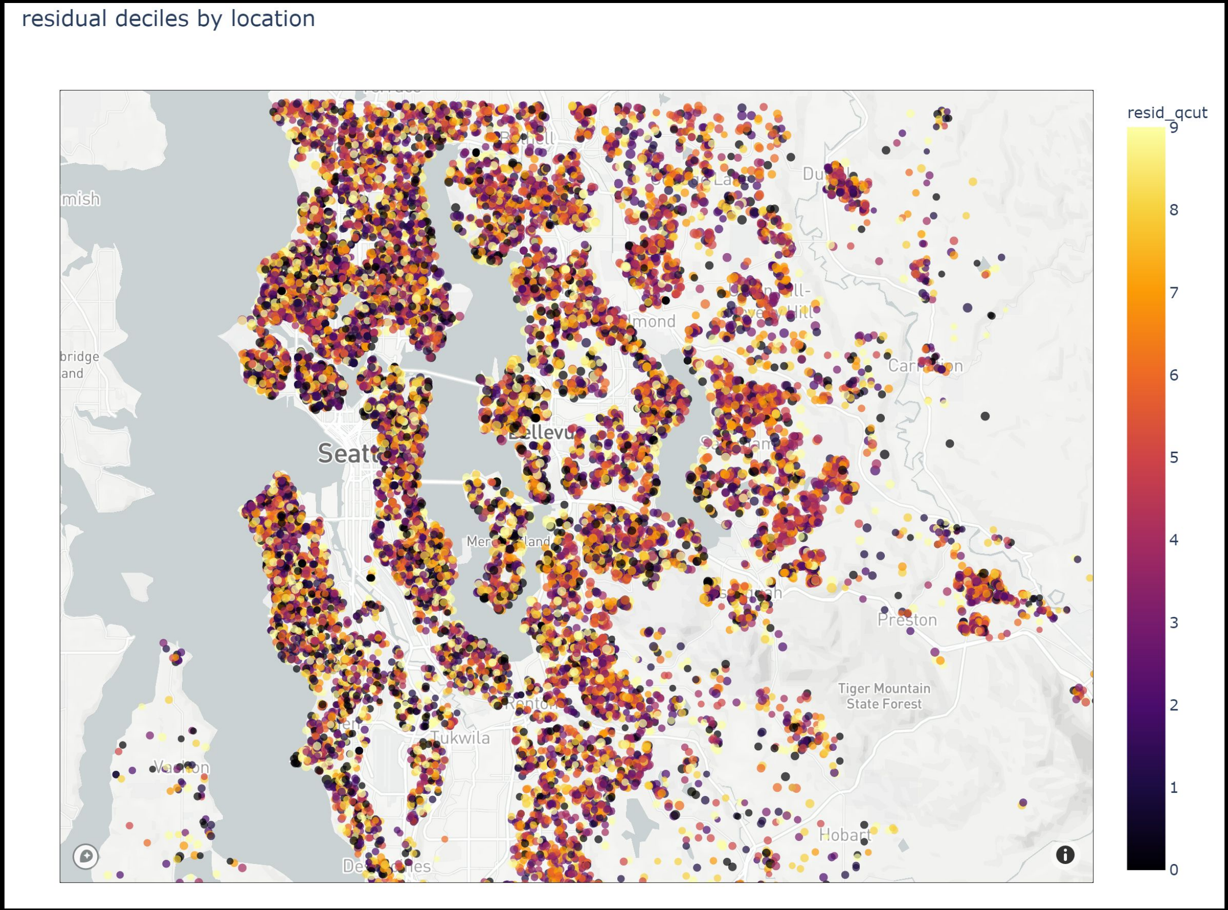
The units of B(j) are “percent change in p_i per marginal change in feature j”:

$$\left(\frac{p_i - \overline{p_{ik}}}{\overline{p_{ik}}} \right) = \sum_{j=0}^{J-1} \beta_j \left(\frac{1}{100} \left[f_j(i) - \overline{f_j(i,k)} \right] \right)$$

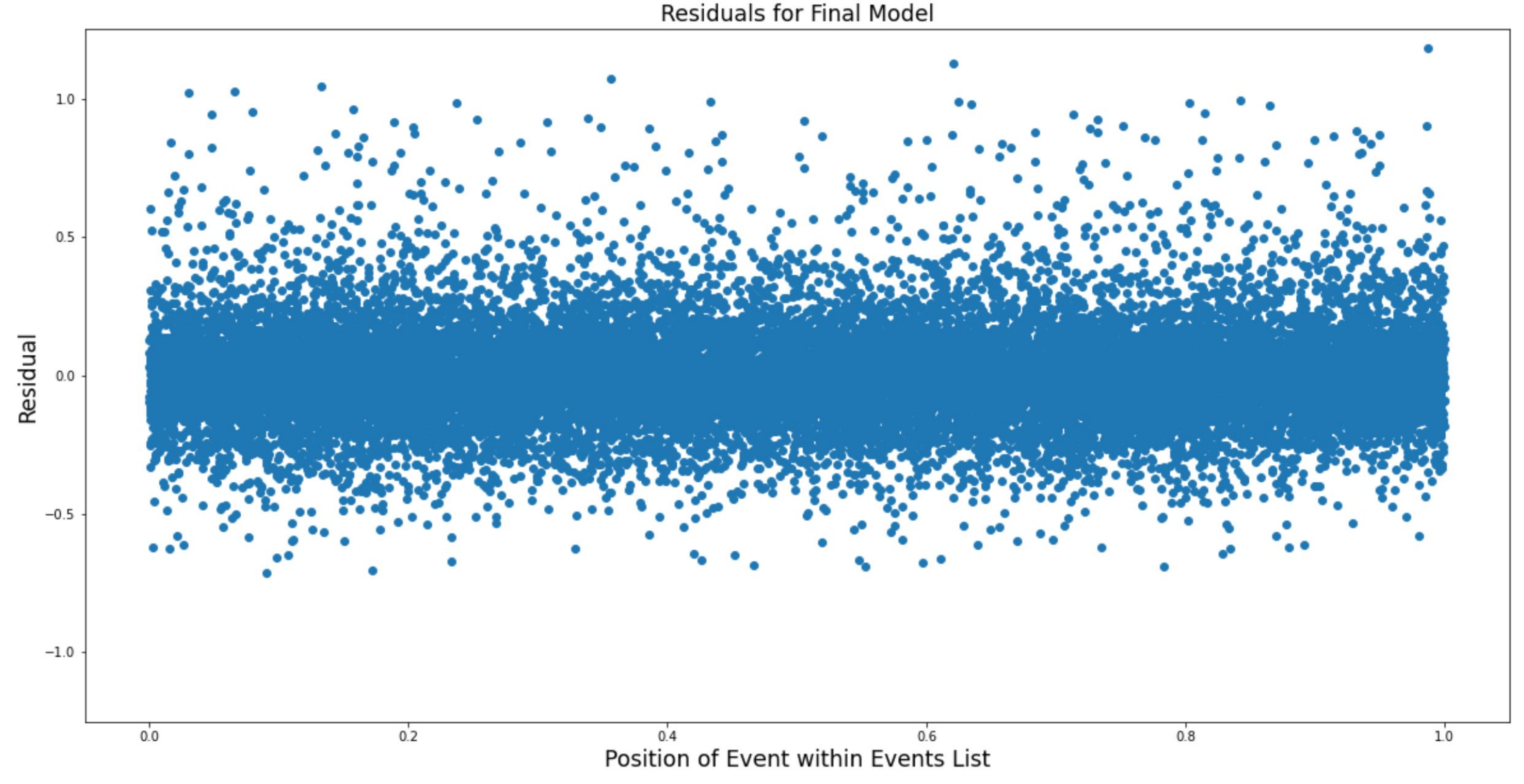
Local Mean Price: varies according to geographic position



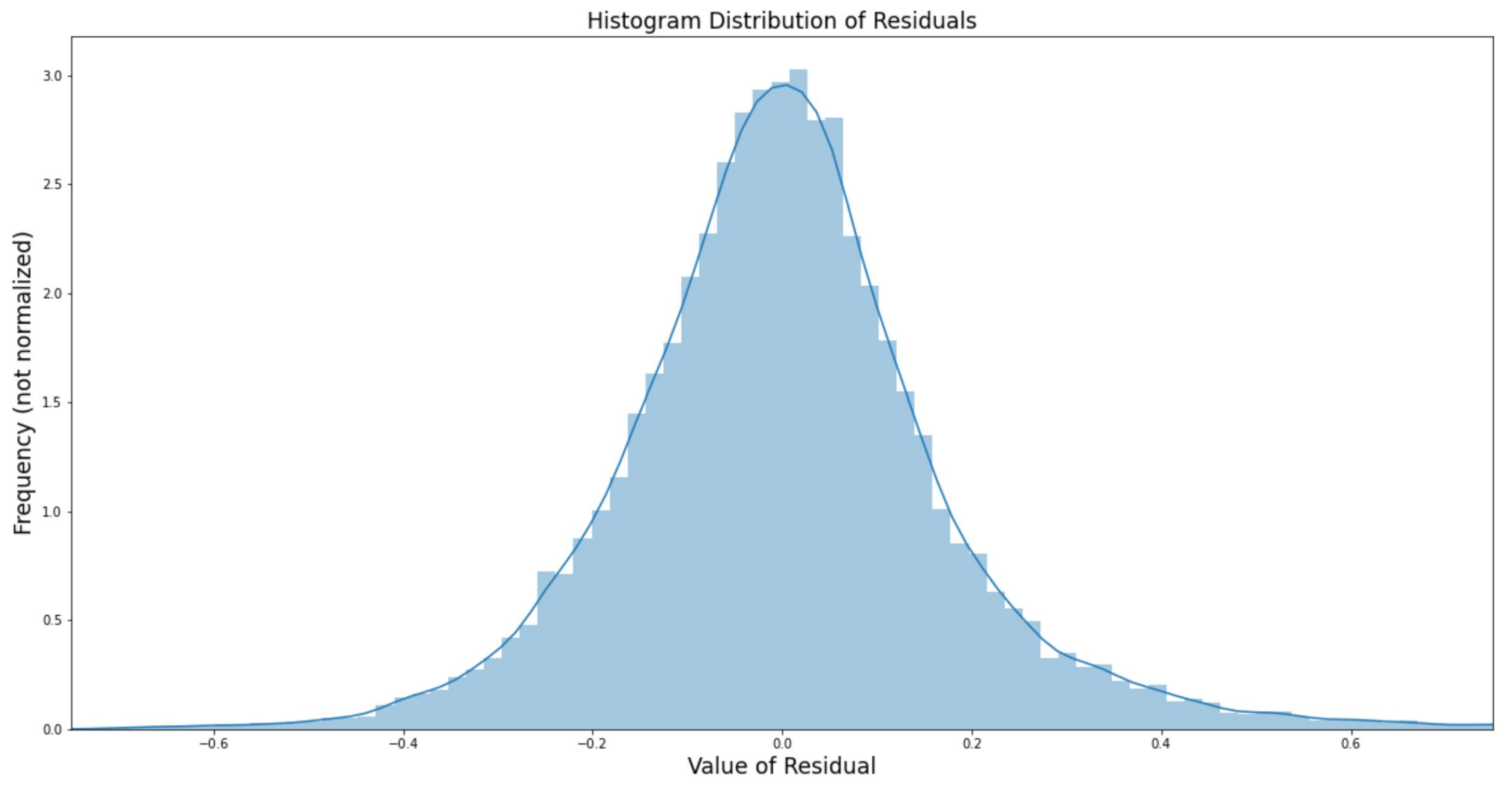
Residual of Model:
sufficiently HETEROSKEDASTIC by geographic location



Residuals:
sufficiently HETEROSKEDASTIC by position in event list



Distribution of Residual values: sufficiently GAUSSIAN



Model Results:

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:            price_zed15   R-squared (uncentered):                  0.683
Model:                            OLS   Adj. R-squared (uncentered):             0.683
Method:                 Least Squares   F-statistic:                             6611.
Date:                Fri, 30 Oct 2020   Prob (F-statistic):                       0.00
Time:                        06:49:49   Log-Likelihood:                         6716.8
No. Observations:               21507   AIC:                                -1.342e+04
Df Residuals:                   21500   BIC:                                -1.336e+04
Df Model:                           7
Covariance Type:            nonrobust
======================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
bedrooms_zed15       -1.1833      0.175     -6.765      0.000      -1.526      -0.840
bathrooms_zed15       2.2330      0.283      7.899      0.000       1.679       2.787
floors_zed15          0.3517      0.315      1.118      0.264      -0.265       0.968
waterfront_zed15     86.2473      1.724     50.027      0.000      82.868      89.626
metascore_zed15       6.3519      0.118     53.798      0.000       6.120       6.583
sqft_living_zed15     0.0255      0.000     95.633      0.000       0.025       0.026
building_age_zed15   -0.0241      0.007     -3.639      0.000      -0.037      -0.011
==============================================================================
Omnibus:                     3143.009   Durbin-Watson:                   2.012
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11168.543
Skew:                           0.722   Prob(JB):                         0.00
Kurtosis:                       6.222   Cond. No.                     1.04e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.04e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
Residuals failed test/tests
```

- Having one more bathroom than the “local mean number of bathrooms” (the mean number of bathrooms for the 15 nearest neighbors) is associated with a 2.2 percent increase in sale price for that dwelling.

- An offset by 1 in the value of the feature “Waterfront” (compared to the MEAN of the 15 nearest neighbors’ “Waterfront” designations) is associated with an 86 percent increase in the value of the property at that location.

Additional Model Results:

- The Beta parameter associated with the AGE of the dwelling (years elapsed since construction) indicates a 1% reduction in value for every 50 years of "excess" age compared to the home's neighbors.
- Each additional 100 square feet of living space (compared to one's neighbors) is associated with a 2% increase in value for that property.

Conclusions:

- My approach (to express the TARGET and the FEATURES for the OLS model according to Equation 2) was successful.
- The "Beta Coefficients" are meaningful, and form the basis for relevant insights into the housing market of Seattle.

Future Work:

- It would be interesting to apply this same approach to OTHER METRO AREAS, and determine whether similar relationships between homes' prices and features can be observed.

- It would also be interesting to monitor time series of housing sales data, to determine whether and how the Beta coefficients might vary with time.

Thank You

For willingly experiencing the HORROR of having to see equations on the day before HALLOWEEN.