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 $f_j(i,k)$ from the local mean value of that feature.

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

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

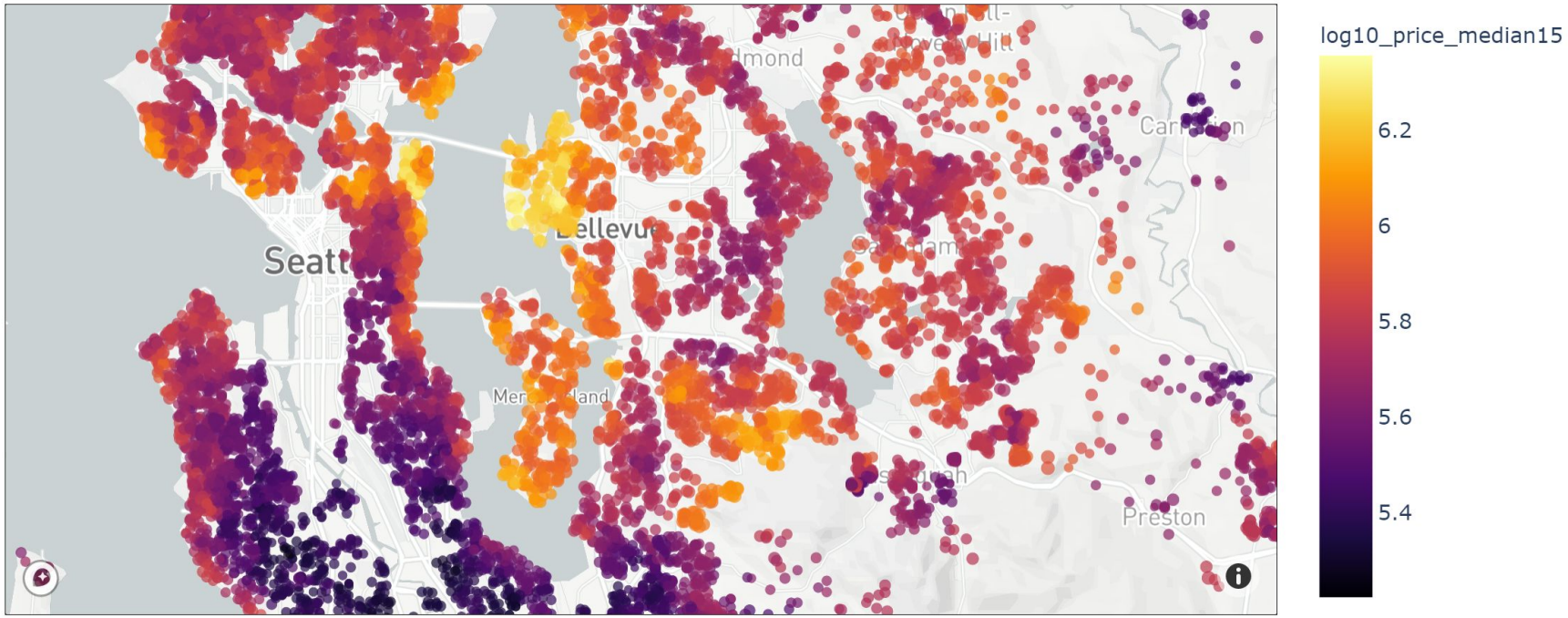
It constrains “beta”, so that B(j) measures the LOCAL marginal effect of feature fj on the RELATIVE value of property i.

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

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

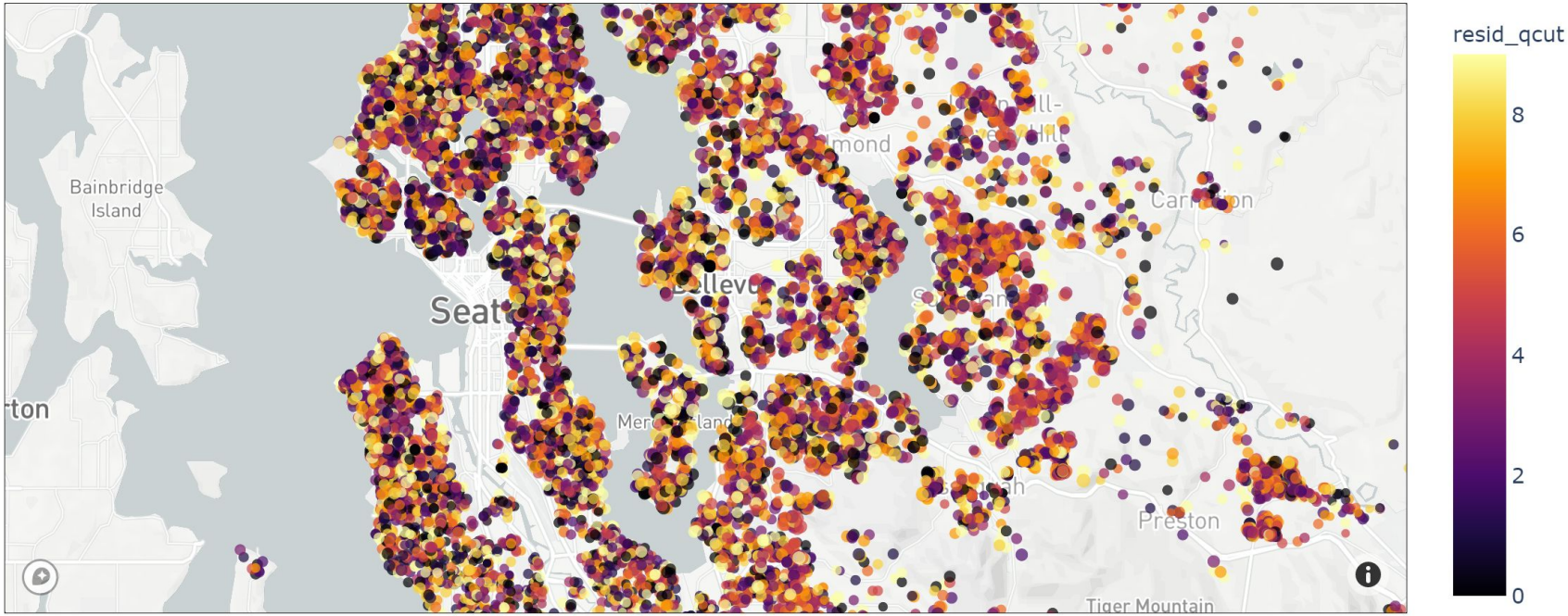
Local Median Price: varies gradually by geographic location

log10(Median Price of 15 nearest Neighbors)

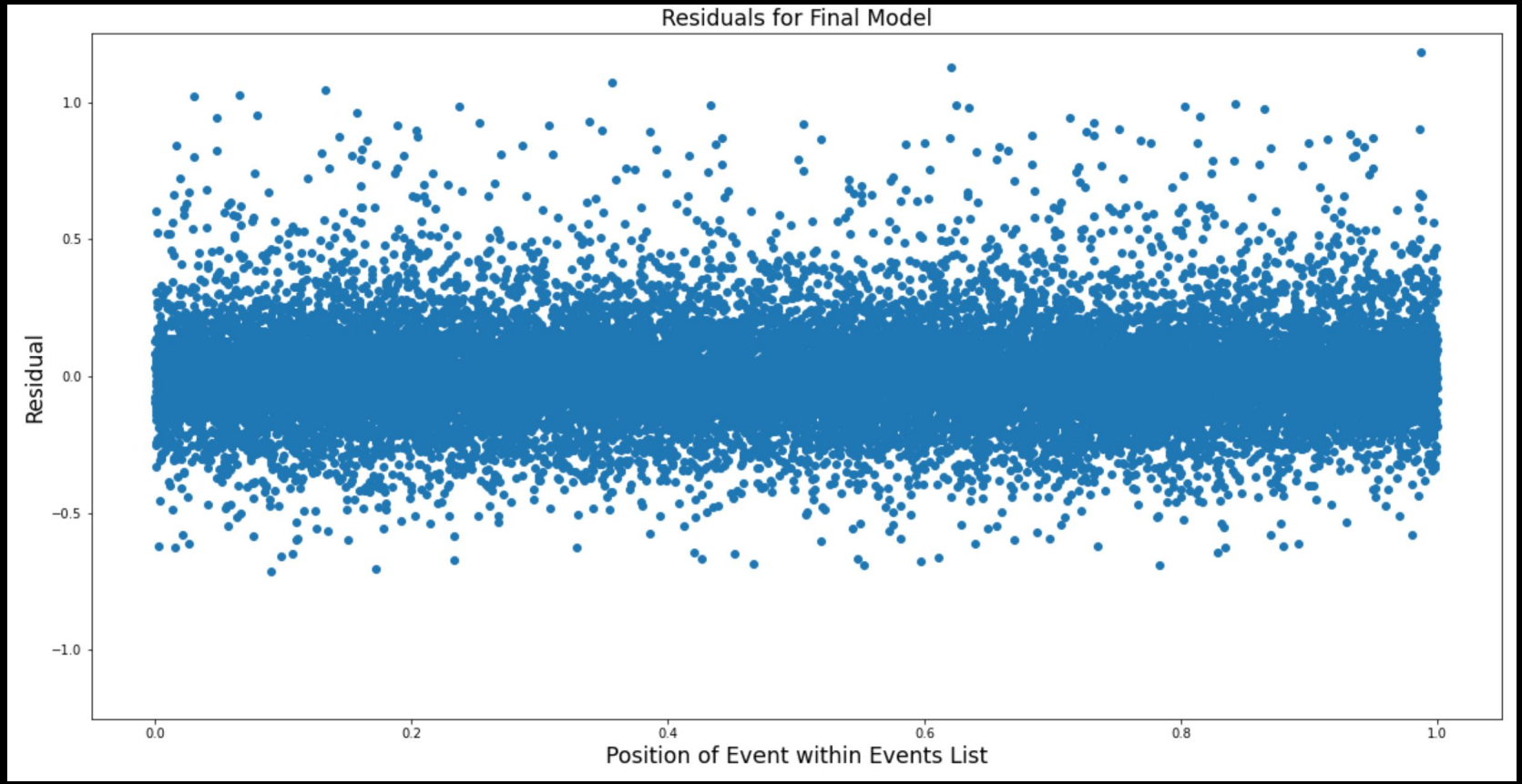


Residual of Model:
sufficiently HETEROSKEDASTIC by geographic location

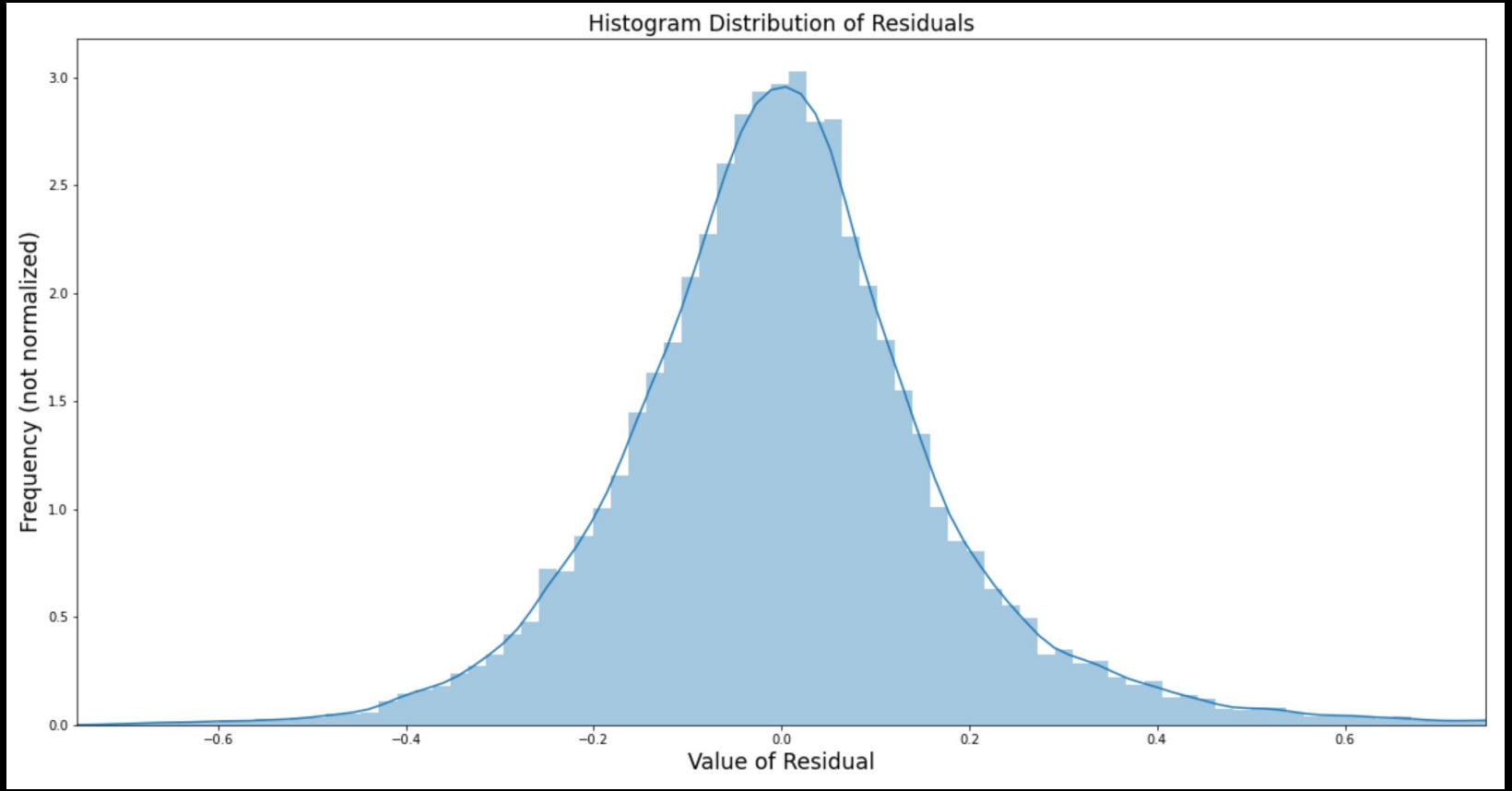
residual deciles by location



Residuals:
sufficiently HETEROSKEDASTIC by position in event list



Distribution of Residual values: sufficiently GAUSSIAN



Model Results:

An offset by 1 in the value of the feature “Waterfront” (compared to one’s 15 nearest neighbors) is worth an 89 percent increase in the value of the property at that location.

Conclusions:

- My approach (to express the TARGET and the FEATURES for the OLS model according to Equation 2) was successful.
- Similar results can be obtained from the same multivariate OLS model for other Features j.

Future Work:

Update this slide presentation to include more information about the OLS results including R^2, etc.

Thank You

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