Locating Your Home

An In-Depth Analysis on Housing Locations and Prices in King County, WA

By Melvin Garcia
May 17, 2021
Flatiron Cohort - onl01-dtsc-pt-011121

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions
- Next Steps

Business Problem

- Real Estate Agency looking to understand the Kings County Housing Market based on location
- Clientele are Washington out-of-state looking for general advice on where to potentially buy a house
- Most are targeting a vacation home

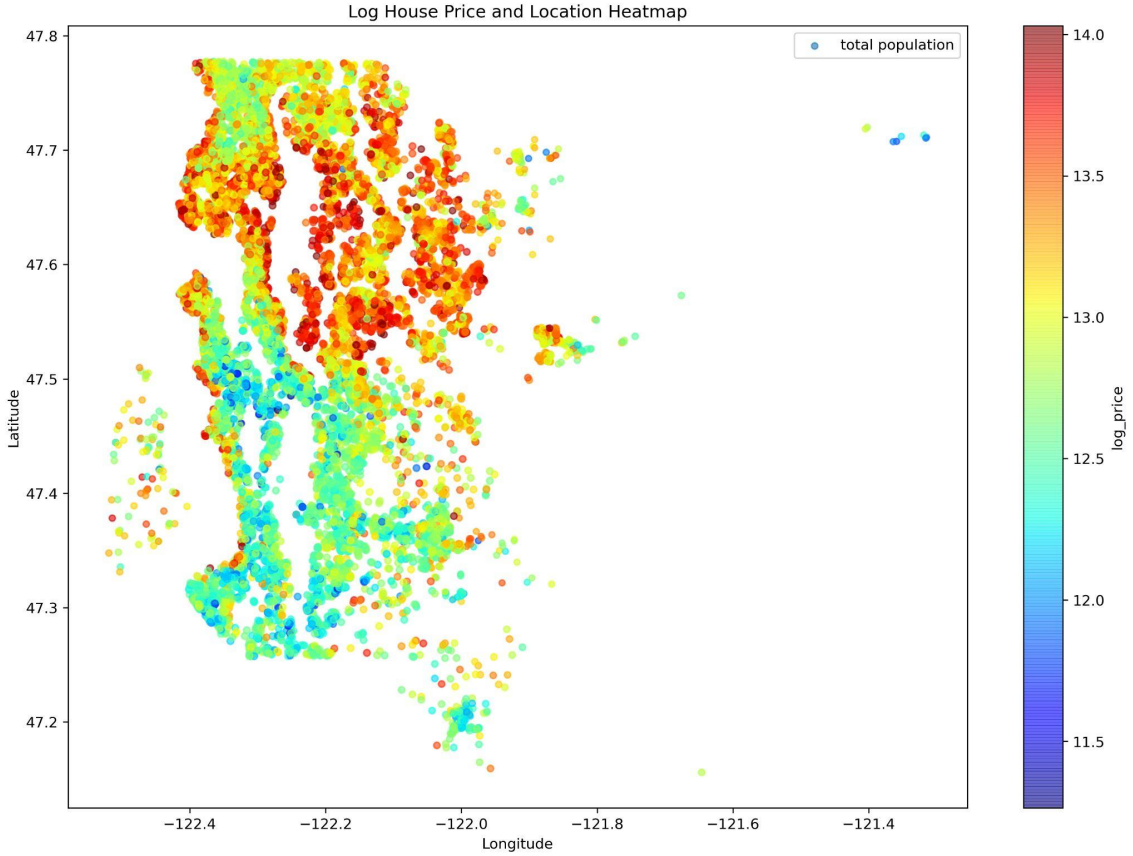
Data

- King County, Washington Housing Sales Data between May 2014 - May 2015
  - Price
  - Latitude, Longitude
  - Waterfront
  - View
  - Distance from downtown
- National Historical Geographic Information System (NHGIS) Census Data
  - Population, Education, Age, Income etc

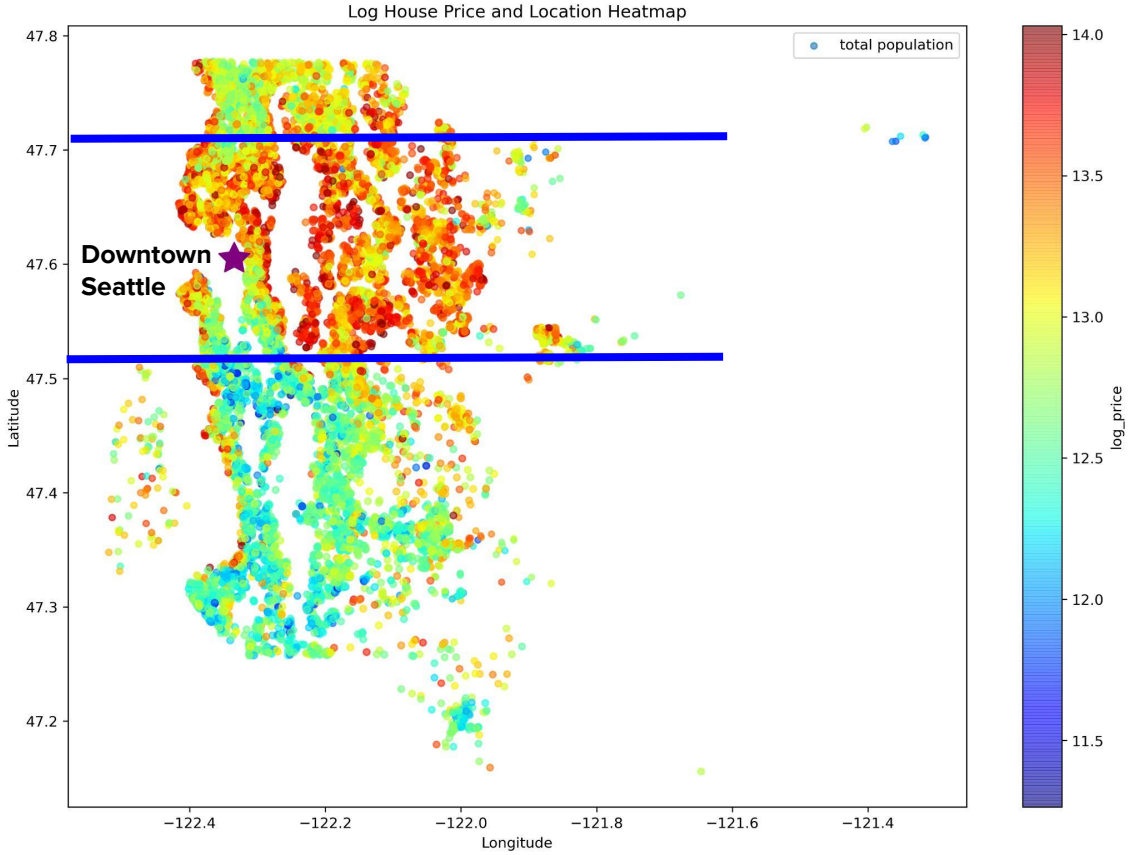
Methods

Prepare & Explore Data	Clean Data	Build Model & Check Assumptions	Feature Optimization & Iterate on Models
1. Explore and confirm main tables to use 2. Join disparate data sources 3. Data exploration!	1. Examine data quality 2. Observe skewed distributions that may affect overall analysis 3. Remove outliers	1. Prepare select features for regression modeling 2. Build our model 3. Conduct due diligence that our data is well prepared for model building	1. Learn from any errors from first model 2. Optimized the features for the model 3. Repeat step 3

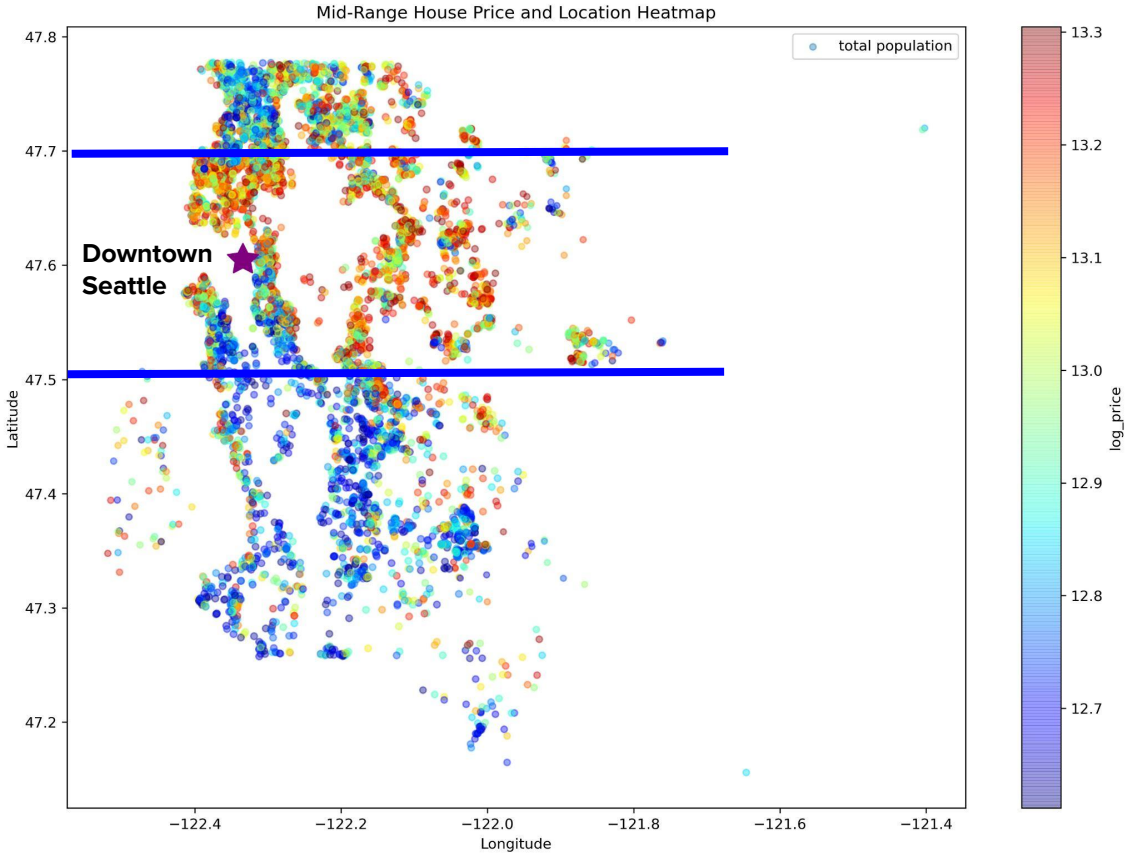
Results - Log House Price ($) and Location



Results - Log House Price ($) and Location



Results - Mid-Range House Prices ($400k-$600k) and Location



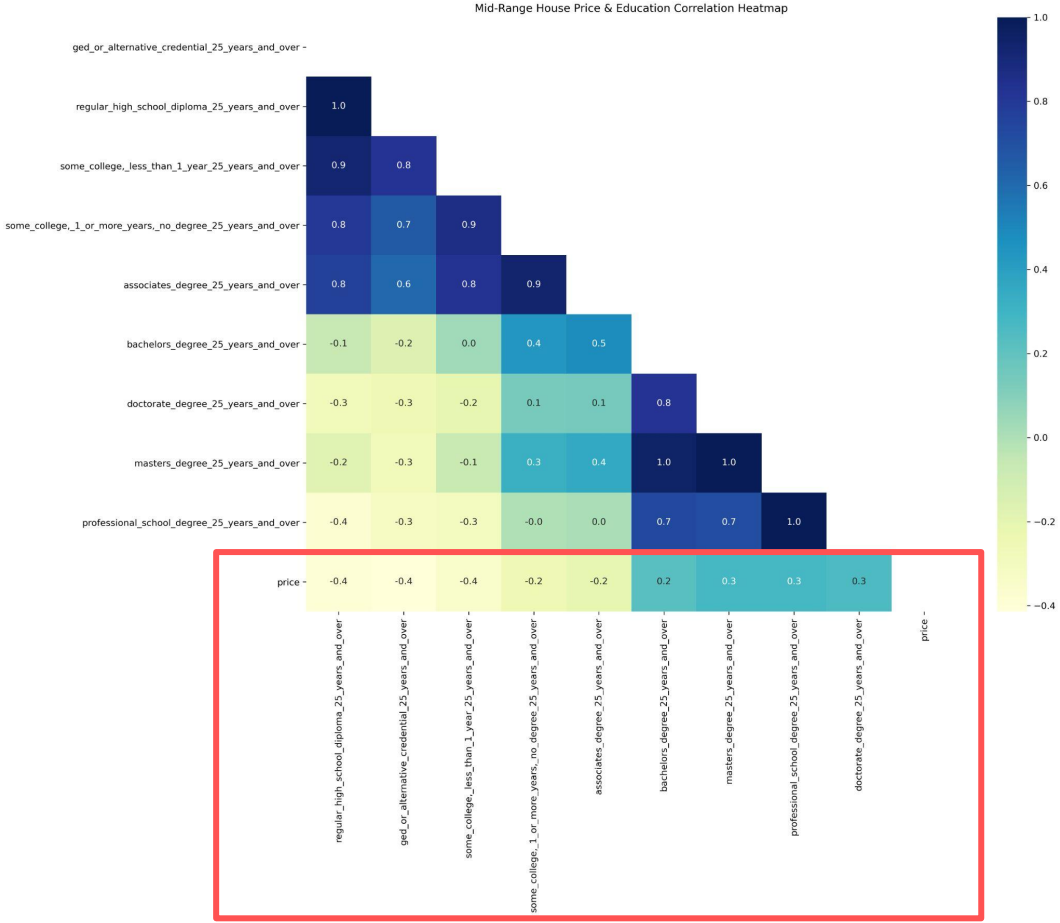
Results - Waterfront and View

~13k Total Houses

Waterfront	% of Houses
No	99.6%
Yes	0.03%

View Rating	% of Houses
0	92.1%
1	1.3%
2	3.9%
3	1.8%
4	0.8%

Results - Education and House Price Correlation Heatmap



Summary of Features

- Distance from Downtown Seattle (km)
- Regional Price Differences shaped by latitude and longitude coordinates
- Whether the property has a waterfront
- A rating (0-4) of how good the view a property has
- Education level demographic

Model Results

$R^2 = 0.457$

Feature	Coefficient	P-Value
Intercept	4.882×10^5	0.000
GED or Alternative 25 Yrs & Older	-259.3984	0.000
Masters Degree 25 Yrs & Older	22.6175	0.000
Waterfront	2.161×10^4	0.424
Distance from downtown (km)	-221.0116	0.218
View	8.383×10^4	0.000
Price Region	7.041×10^4	0.000

Model Results

$R^2 = 0.447$

Feature	Coefficient	P-Value
Intercept	5.007×10^5	0.000
GED or Alternative 25 Yrs & Older	-263.8356	0.000
Masters Degree 25 Yrs & Older	23.2083	0.000
View	8.49×10^4	0.000
Price Region	6.591×10^4	0.000

Conclusions

- A house without regard to the provided features is expected to have a house price of approximately $500,700
- The quality of the view increases the price of a house by $84,900 on average
- Buying a house within the specified region of 45.5 and 47.7 latitude increases the price of a house by $65,910 on average

Next Steps

- Experiment with logarithmic and exponential transformations of the provided features to observe any increase in our R^2
- View the house price changes over time in regards to changing demographics
  - GDP
  - Income
  - Supply of Housing
  - Unemployment
  - Age
  - Education

Source: https://core.ac.uk/download/pdf/46712962.pdf

Thank You!

Email: garciamelvin4@gmail.com
GitHub: @melvyg
LinkedIn: linkedin.com/in/melvinmgarcia/