Modeling Miami Housing Costs

Tatev Gomtsyan, Ethan Assefa, Thomas Burrell

Purpose

Motivation

To gain insights into the real estate market, understand trends in property values, and assess affordability given the impact of economic factors like inflation, interest rates, and changing work dynamics

Stakeholders

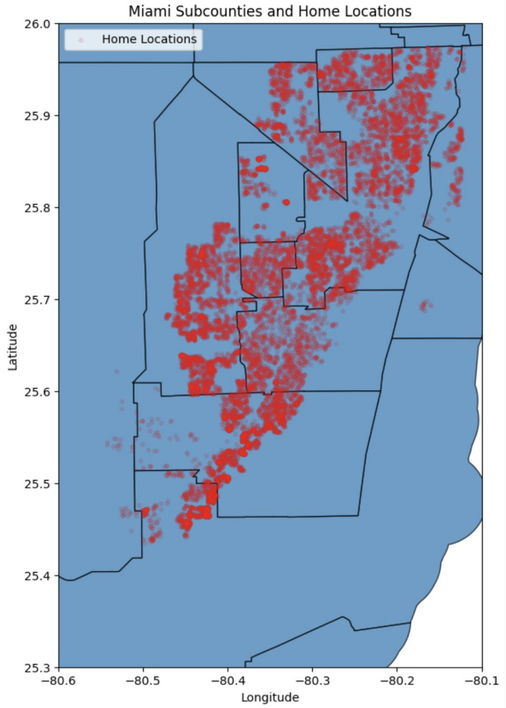
Such analysis can help **individuals** make informed decisions regarding home buying, selling, or investment, and provide valuable information to **policymakers and economists** in monitoring economic stability and housing market resilience

Goal

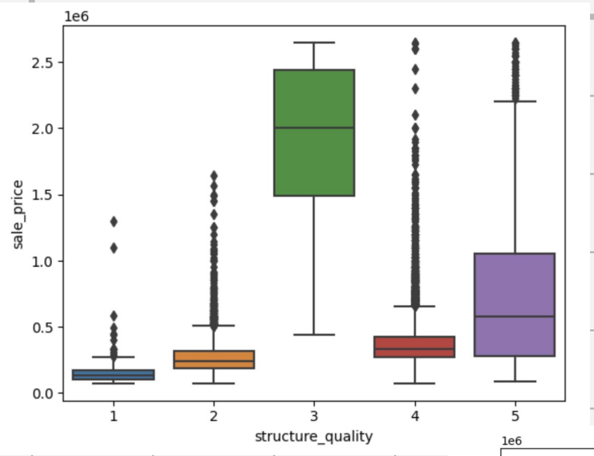
Clearly quantify & describe the uncertainty around predictions to a potential home-owner; communicating the range of expected sales prices (rather than simple point-estimate) will allow them to better prepare for the home-buying process

Our Data

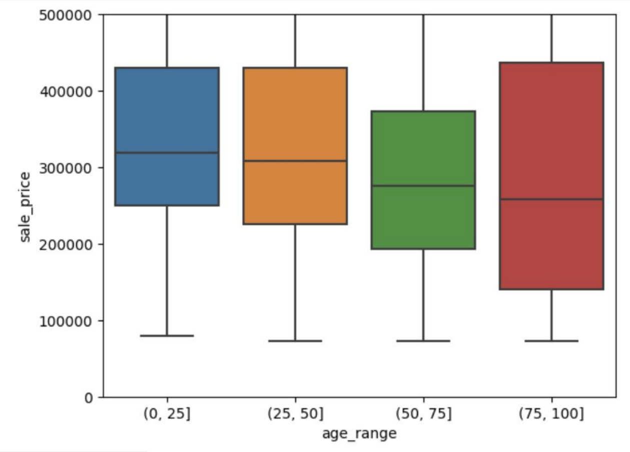
- **Miami Housing Dataset:** Link to Data
    - 13,932 rows (Single-Family Homes)
    - 17 columns
        - Response Variable: Sale Price
        - Predictor Variable: Square Footage
        - Other Variables of Interest
            - Structure Quality
            - Age
            - Distance to Ocean
    - All numeric variables scaled & standardized
- **2022 TIGER/Line® Shapefiles Dataset:** Link to Data
    - Contains geometries for sub-county boundaries
    - Determine sub-county of home using its latitude & longitude
        - Categorical sub-county group variable created



Exploratory Data Analysis

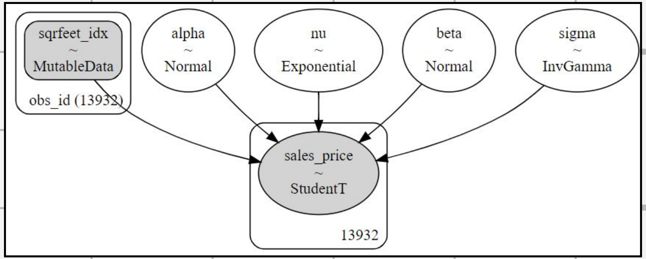


Weak positive relationship (corr=0.363) between sq. footage & sales price

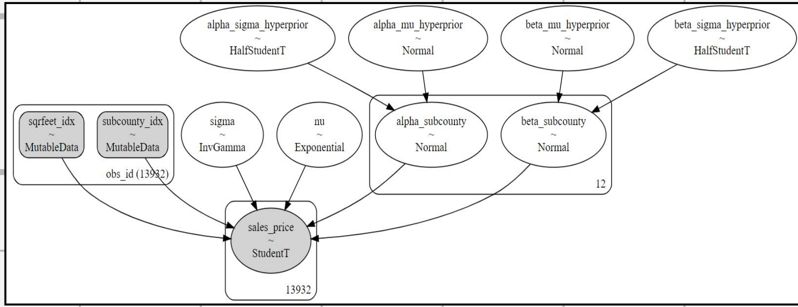


Houses with avg structure quality have the highest sale price

The oldest houses have the lowest price points



Models



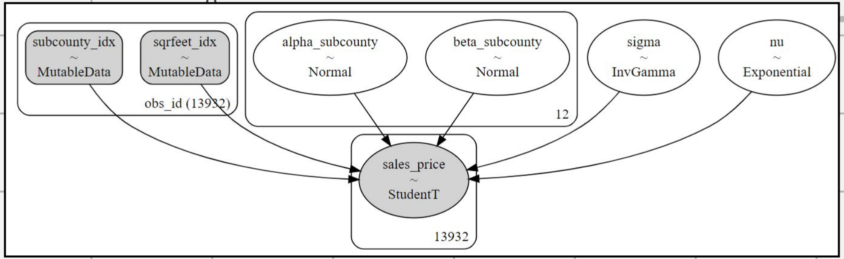
Unpooled/Pooled

Unpooled: Separate models on sub-county groups predicting square footage against sales price. 12 groups total.

Pooled: Single model of all data, predicting square footage against sales price.

Hierarchical

Hierarchical: Create a nested model including sub-county groupings in the predictor-response relationship. We used all data points, while incorporating group differences.

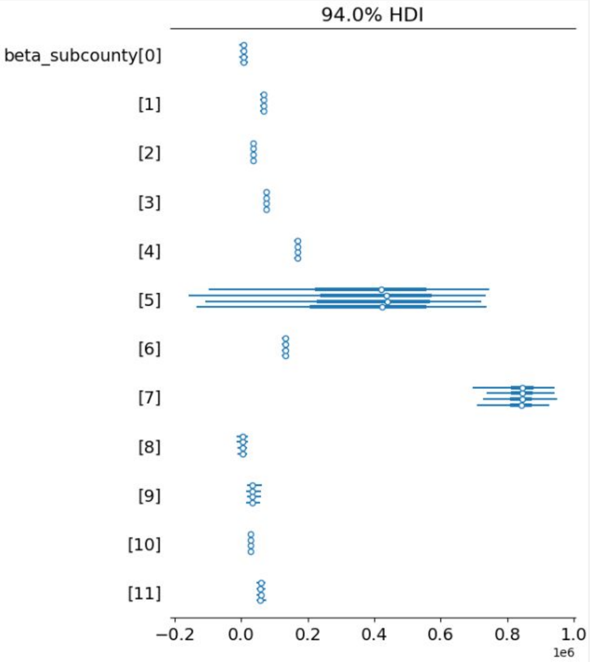


Evaluation

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Partial Pooled	0	-186490.005844	56.814192	0.000000	0.944301	170.621313	0.000000	False	log
Full Pooled	1	-192352.885439	1.839233	5862.879595	0.055699	140.665467	111.365543	False	log
No Pooled	2	-192371.011483	1.776372	5881.005639	0.000000	140.633785	111.444180	False	log

- *Partial-Pooled:* Best-performing model
  - The elpd_diff is 0 compared to itself (since it's the best model)
  - It has the highest model weight, indicating it should be weighted more in a ensemble
- *Pooled:* Second ranked model
  - The model weight is relatively low, indicating it has a little information to contribute
- *Unpooled:* Worst ranked model
  - The model weight is zero, it should not be a used in an averaging of the models

Results



	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
beta_subcounty[0]	6512.04	6345.73	-5654.65	18064.18	113.43	84.13	3133.57	2967.38	1.0
beta_subcounty[1]	66547.59	5988.78	55968.62	77562.76	99.87	70.91	3586.99	3269.67	1.0
beta_subcounty[2]	34382.87	3450.10	27842.05	40707.80	50.61	35.79	4668.71	3272.39	1.0
beta_subcounty[3]	74150.36	3653.29	67336.23	81041.33	62.51	44.35	3414.44	3219.18	1.0
beta_subcounty[4]	168045.98	4244.89	160416.96	176246.58	72.71	51.42	3459.75	2616.03	1.0
beta_subcounty[5]	379574.22	243828.26	-114917.99	746631.80	5796.14	4099.16	2073.02	1909.79	1.0
beta_subcounty[6]	131482.55	4748.61	122790.26	140344.39	78.18	55.29	3711.72	2949.84	1.0
beta_subcounty[7]	838149.18	59919.76	712455.40	940093.48	1463.21	1034.82	2088.78	1424.05	1.0
beta_subcounty[8]	2318.91	8581.66	-13553.15	18208.40	125.37	120.69	4835.24	2791.81	1.0
beta_subcounty[9]	35096.63	11851.72	15233.17	58611.17	236.30	191.58	3349.69	2046.24	1.0
beta_subcounty[10]	27503.91	2680.63	22557.22	32502.21	41.23	29.66	4282.73	2680.13	1.0
beta_subcounty[11]	58610.59	7244.47	45323.33	71954.33	114.09	86.05	5161.16	2599.27	1.0

subcounty 5 = Kendall-Palmetto Bay
subcounty 7 = Kendale Lakes-Tamiama

Conclusions

- We achieved our goal of creating hierarchical models that can be used for predicting house prices in Miami
- Limitations
  - Data only represents a very small portion of homes in the United States.
  - Housing market fluctuates year to year; this model may not be appropriate for future use
  - Use and choice of priors could be adjusted
  - Computational power and resources
  - Non-linear models may better explain relationship

Thank you!
Questions?