CS109b Spring 2018
Tommy Hill - Andrea Porelli

"LOCATION, LOCATION, LOCATION: CONSIDERING SPACE AND DISTANCE IN HOME PREDICTION"

0 WHAT PROBLEM ARE WE TRYING TO SOLVE?

PROBLEM:

ACCURATELY PREDICTING HOME SALE PRICE

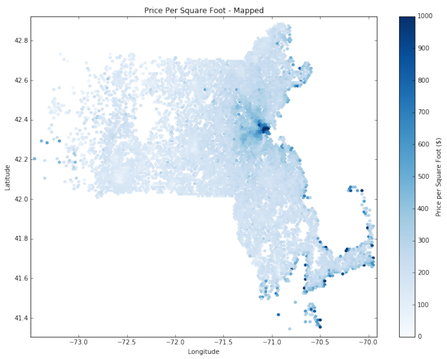
POSSIBLE SOLUTIONS:

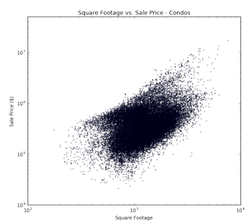
ADD NEIGHBORHOOD CHARACTERISTICS
ADD UNSTRUCTURED DATA
ADD ENGINEERED FEATURES
USE HEDONIC REGRESSION

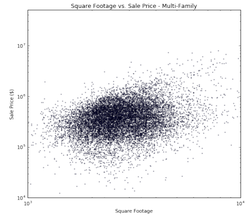
OUR APPROACH:

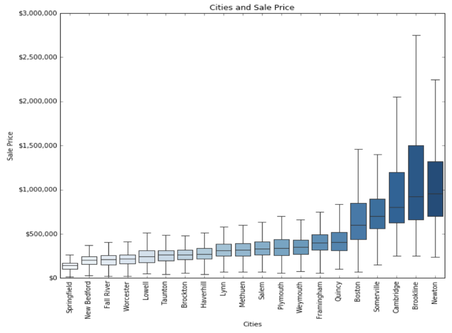
CONSIDER LOCATION AND DISTANCES AMONG PROPERTIES

1 EXPLORING OUR DATA





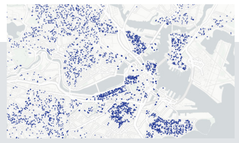




2 DATA PREPROCESSING

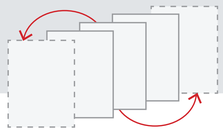
DATA COLLECTION:

GEOCODE LOCATION
ArcGIS



DATA CLEANING:

COMBINE DATASETS
REMOVE DUPLICATES
DIVIDE INTO 2016 AND 2017, ACTIVE AND SOLD
FEATURE EXTRACTION
DUMMY VARIABLES

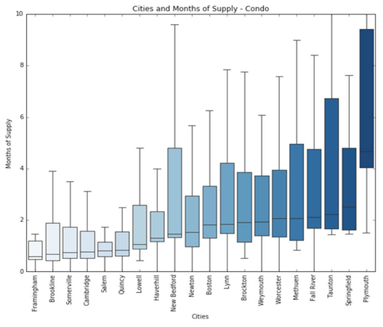


3 FEATURE ENGINEERING

MONTHS OF SUPPLY:

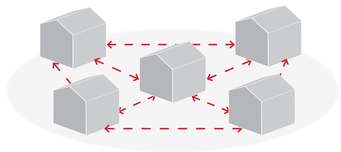
- SHARE OF ACTIVE PROPERTIES OVER PROPERTIES SOLD IN THE PREVIOUS YEAR FOR A GIVEN CATEGORY
- APPROXIMATES HOUSING MARKET SUPPLY AND DEMAND CONDITIONS
- INTUITIVE AND WIDELY USED IN PROPERTY APPRAISAL

$$MOS = \frac{\#ACTIVE\ PROPERTIES}{\#PROPERTIES\ SOLD\ IN\ PREVIOUS\ YEAR}$$



DISTANCE MATRIX:

- USE PYSAL TO COMPUTE DISTANCE AMONG ALL PROPERTIES SOLD IN 2016 AND 2017
- INTENSIVE COMPUTATION - $O(n^2)$



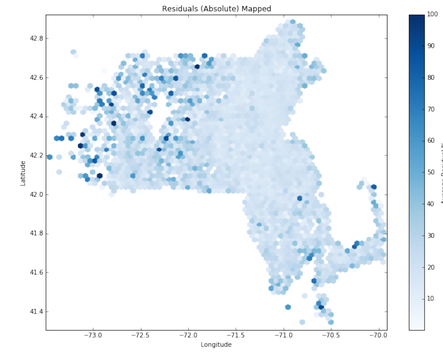
	Sample Distance Matrix - Distance between each property (Km)				
	1 Earhart St	33 Fulkerson St	31 Fulkerson St	17 Otis St	32 Sciarappa St
1 Earhart St	0	1.043	1.035	0.251	0.542
33 Fulkerson St	1.043	0	0.007	0.829	0.531
31 Fulkerson St	1.035	0.007	0	0.821	0.524
17 Otis St	0.251	0.829	0.821	0	0.389
32 Sciarappa St	0.542	0.531	0.524	0.389	0

4 MODELING & ANALIZING

PIPELINE:

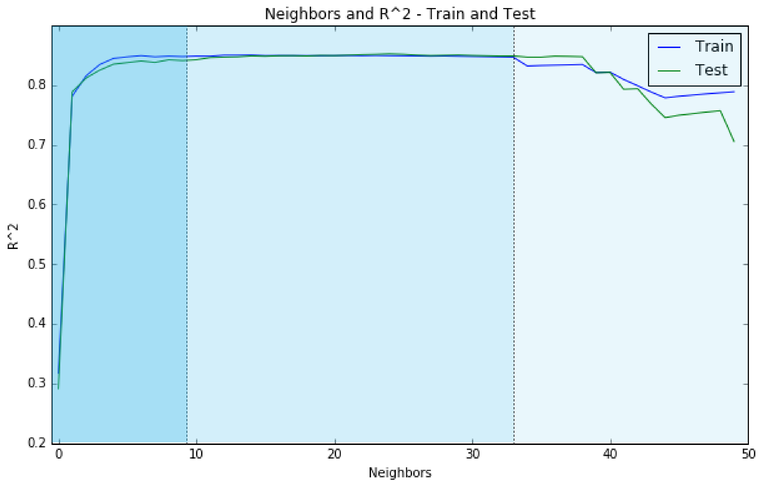
START
- SPLIT INTO 2016 AND 2017 SALES
- BUILD BASELINE REGRESSION MODEL
- BUILD DISTANCE MATRIX
- OPTIMIZE NUMBER OF NEIGHBORS
- PREDICT

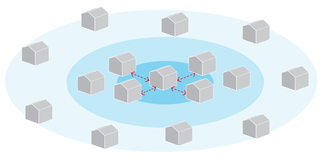
END



$$PRICE = \beta_0 + \beta_1 * AGE + \beta_2 * BATHS + \beta_3 * BEDS + \beta_4 * GARAGE + \beta_5 * SQFT + \beta_6 * MOS + \beta_7 * PREDICTED\_PRICES + \beta_8 * REVISED\_PRICE$$

5 RESULTS





OLS Regression Results

Dep. Variable:	SOLDPRICE	R-squared:	0.869
Model:	OLS	Adj. R-squared:	0.869
Method:	Least Squares	F-statistic:	1.381e+04
Date:	Mon, 30 Apr 2018	Prob (F-statistic):	0.00
Time:	12:02:52	Log-Likelihood:	-2.2303e+05
No. Observations:	16655	AIC:	4.461e+05
Df Residuals:	16646	BIC:	4.461e+05
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.209e+04	3896.106	3.104	0.002	4457.132	1.97e+04
AGE	5.7261	6.382	0.897	0.370	-6.784	18.236
BATHS	4.355e+04	2206.592	19.735	0.000	3.92e+04	4.79e+04
BEDS	2.336e+04	1283.623	18.201	0.000	2.08e+04	2.59e+04
GARAGE	2.353e+04	1507.415	15.611	0.000	2.06e+04	2.65e+04
SQFT	-114.7648	2.636	-43.541	0.000	-119.931	-109.598
Category_MOS	1.512e+04	1.05e+04	1.443	0.149	-5412.258	3.57e+04
Predicted_Prices	1.0738	0.005	230.878	0.000	1.065	1.083
Revised_Price	-0.0030	0.001	-2.068	0.039	-0.006	-0.000

Institute for Applied Computational Science
HARVARD SCHOOL OF ENGINEERING AND APPLIED SCIENCES

HARVARD John A. Paulson School of Engineering and Applied Sciences

LITERATURE REVIEW:
Basu, Sabyasachi, and Thomas G. Thibodeau. "Analysis of Spatial Autocorrelation in House Prices." The Journal of Real Estate Finance and Economics 17, no. 1 (1998): 61–85.
Dubin, Robin A. "Spatial Autocorrelation and Neighborhood Quality." Regional Science and Urban Economics 22, no. 3 (1992): 433–452.
Dubin, Robin, Kelley Pace, and Thomas Thibodeau. "Spatial Autoregression Techniques for Real Estate Data." Journal of Real Estate Literature 7, no. 1 (1999): 79–95.
Harring, Robert. "Spatial Data Analysis: Theory and Practice." Cambridge University Press (2003): 289–378.
Legendre, Pierre. "Spatial Autocorrelation: Trouble or New Paradigm?" Ecology 74, no. 6 (1993): 1659–1673.
Pace, R. Kelley, Ronald Barry, and C. F. Sirmans. "Spatial Statistics and Real Estate." The Journal of Real Estate Finance and Economics 17, no. 1 (1998): 5–13.
Schernthanner, Harald, Hartmut Asche, Julia Gonschorek, and Lasse Scheele. "Spatial Modeling and Geovisualization of Rental Prices for Real Estate Portals." In International Conference on Computational Science and Its Applications, 120–133. Springer, 2016.