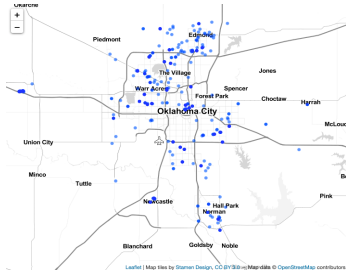


Understanding the OKC Rental Market through Analysis of Craigslist Data

Karley Nadolski

April 29, 2021

Introductions - Craigslist as a Data Source



Craigslist is an important data source for rental markets all over the country

Each Craigslist listing is required to include:

- ▶ advertised rent price
- ▶ user-generated description of the unit
- ▶ Optional: square footage, number of bedrooms/bathrooms, and other descriptions about amenities

Does the extra information in Craigslist descriptions matter? What's the premium of extra information in the Craigslist rental market?

Data Collection - Webscraping

- ▶ Step 1: fill out baseurl
(`"https://oklahomacity.craigslist.org/search/apa"`)
- ▶ Step 2: build queries (bedrooms=2, bathrooms=1, sqft=900)
- ▶ Step 3: choose which attributes to extract from each listing
(ID, title, price, sqft, etc...)
- ▶ Step 4: coerce code into a for loop to iterate through all listings within the query
- ▶ Step 5: organize and clean data

Methods

$$Y_{price} = \left(\begin{array}{l} \beta_0 + \beta_1 X_{sqft} + \beta_2 X_{bdrms} + \beta_3 X_{p.QDAP} + \beta_4 X_{wordcount} \\ + \beta_5 X_{t.wordcount} + \beta_6 X_{t.direction} + \varepsilon, \end{array} \right) \quad (1)$$

where Y_{price} is a continuous variable describing the price of each rental listing, X_{sqft} is the square footage of the rental unit, X_{bdrms} is the number of bedrooms, $X_{p.QDAP}$ is the positivity index from the listing's description, $X_{wordcount}$ and $X_{t.wordcount}$ are the word counts for the description and the title, respectively, and finally $X_{t.direction}$ describes the sentiment of the listing title.

Research Findings

Findings

Table: OLS estimates

	Estimate	Std. Error	P-value
sqft	0.89505***	(0.03317)	$< 2e - 16$
bdrms	-234.539***	(18.02865)	$< 2e - 16$
p.QDAP	229.54856	(184.04089)	0.21255
wordcount	1.87407***	(0.10400)	$< 2e - 16$
t.wordcount	-12.57888**	(4.26128)	0.00322
t.directionneutral	86.17754*	(41.89659)	0.03992
t.directionpositive	100.19908*	(41.44415)	0.01577

Notes: Standard errors in parentheses. ***Significantly different from zero at the .1% level; **Significantly different from zero at the 1% level. *Significantly different from zero at the 5% level.

What did we learn? Why was it important?

- ▶ For every additional word in a unit description, the rental price rises by about \$2 per month
- ▶ For every additional word in a unit title, the rental price falls by about \$13
- ▶ Positive titles have a higher premium than neutral titles (all else equal)
- ▶ Units with higher rent prices often have detailed unit descriptions and snappy, positive listing titles

What did we learn? Why was it important?

In Oklahoma City, like in many cities across the United States, there exists a rental market that Craigslist is uniquely able to summarize. Traditional data sources that cover rental markets in the United States aren't able to release data at a fast enough pace to accurately characterize the market as it changes. With an introductory knowledge of webscraping, it's possible to analyze data from a local Craigslist subdomain in real time.