



Austin Real Estate

Data Scrappy, Analysis and Interactive Visualization

Computational Optimization

Advisor: Dr. Ned Dimitrov

Xiaoyu Qian (UT ORIE)

Daqi Xu (UT ME)

Chunqi Wang (UT ME)

November 17 2015



Outline

- Data Acquisition
(scrapy, pandana)
- Analysis and Prediction
(sklearn, xgboost)
- Interactive Visualization
(Bokeh)

Data Acquisition

- Zestimate[®] home value

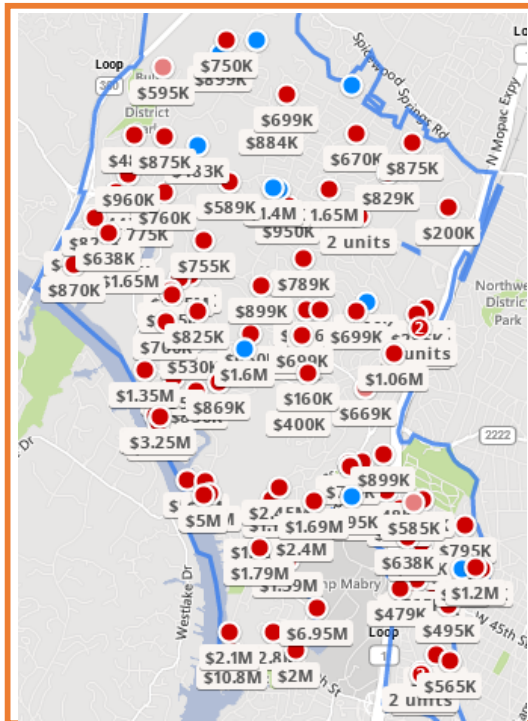
Data Coverage and Zestimate Accuracy Table
Choose a location type below to change data:
Texas Counties

		Zestimate Accuracy	Homes on Zillow	Homes With Zestimates	Within 5% of Sale Price	Within 10% of Sale Price	Within 20% of Sale Price	Median Price
Crosby	★	2.7K	2.0K	--	--	--	--	--
Dallam	★	2.9K	1.8K	--	--	--	--	--
Dallas	★★★★★	677.1K	625.1K	58.1%	79.8%	91.9%	4.0%	--

- Elements influence price
 - Geography
 - Size

Data Acquisition

- Zillow.com (Scrapy)



- Geography
 - Latitude
 - Longitude
 - Zip Code
 - Address
- House
 - Area
 - Bathroom
 - Bedroom
 - Price

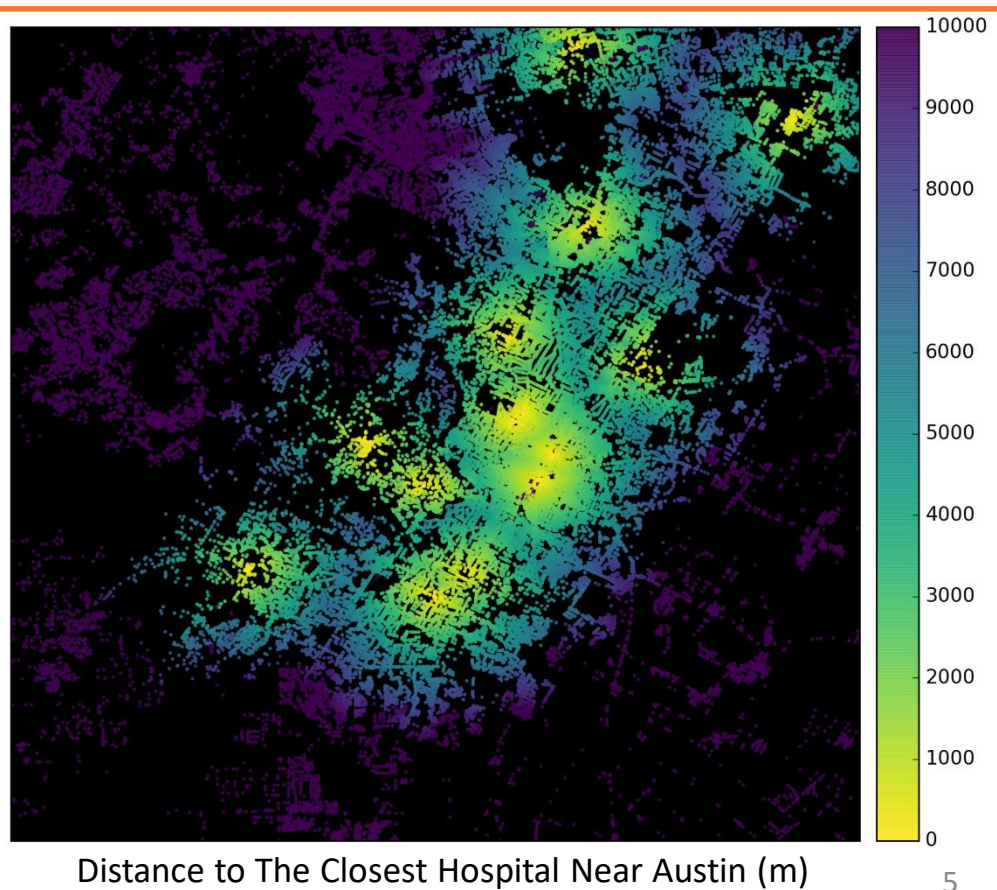
http://www.zillow.com/homes/for_sale/Austin-TX-78731_rb/

Data Acquisition

- OpenStreetMap API (Pandana)



★supermarket ★ restaurant
★hospital ★ bar ★ school



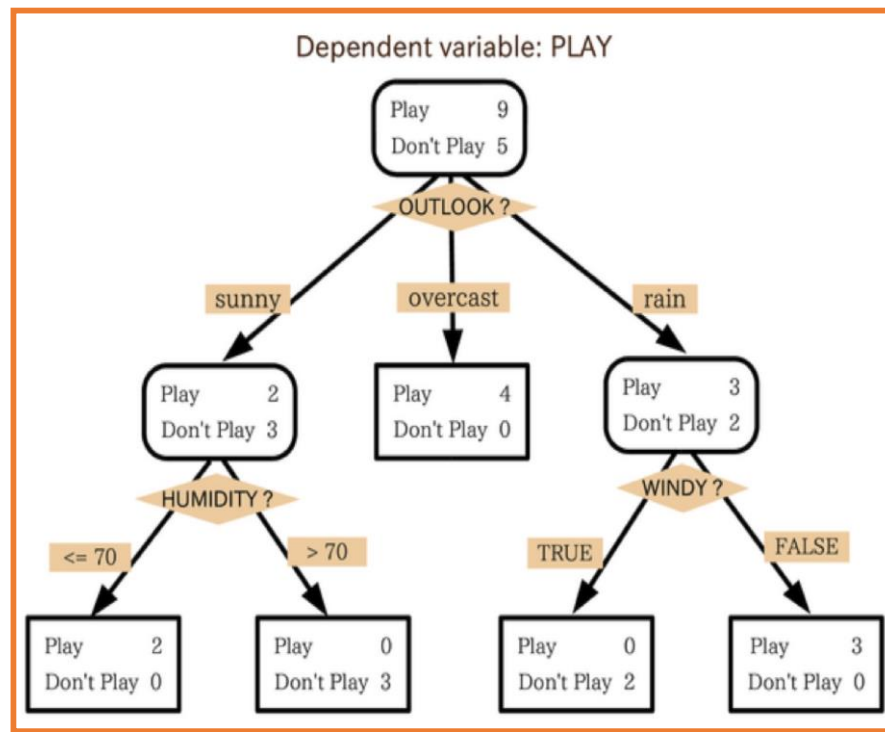


Analysis and Prediction

- Preprocessing
 - Format
 - `data.bedrooms = "studio" -> 1 ;`
 - `data.price = "132K" -> 132000;`
 - `Data.price = 495000+ -> 495000;`
 - Missing Value
 - Clustering:
 - Location
 - Size
 - Amenities

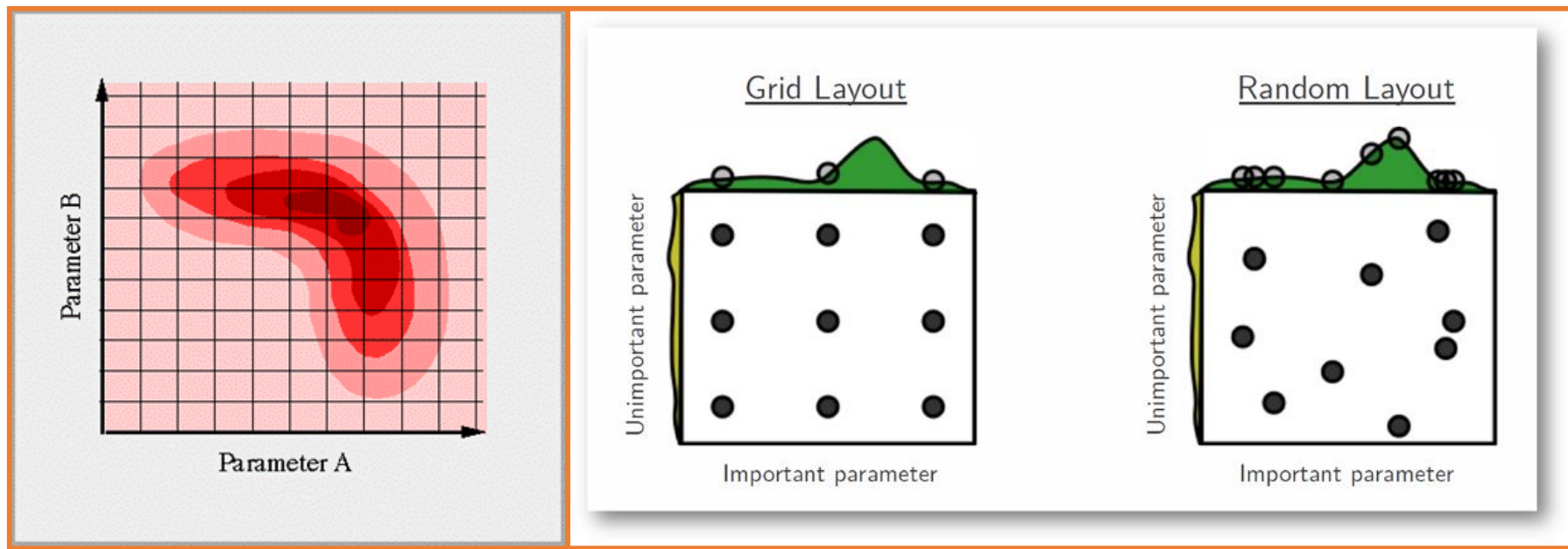
Analysis and Prediction

- Model
 - Linear Regression
 - Random Forest Tree
 - Xgboost



Analysis and Prediction

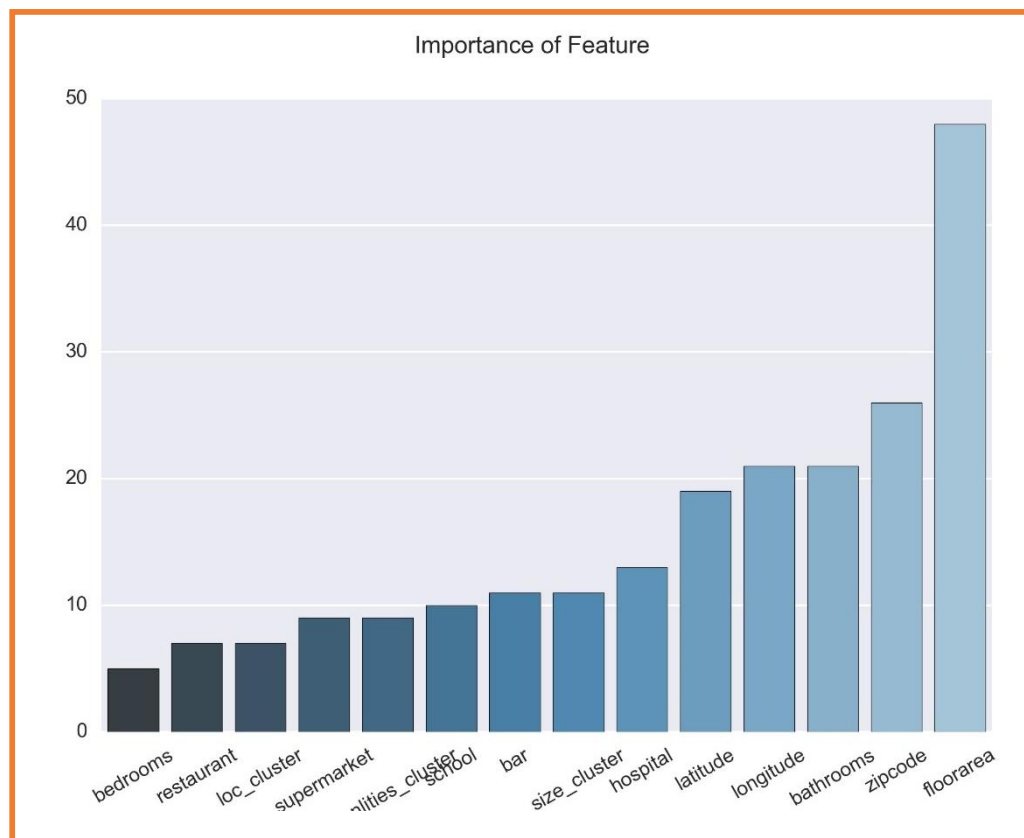
- Tuning Parameters
 - GridSearch





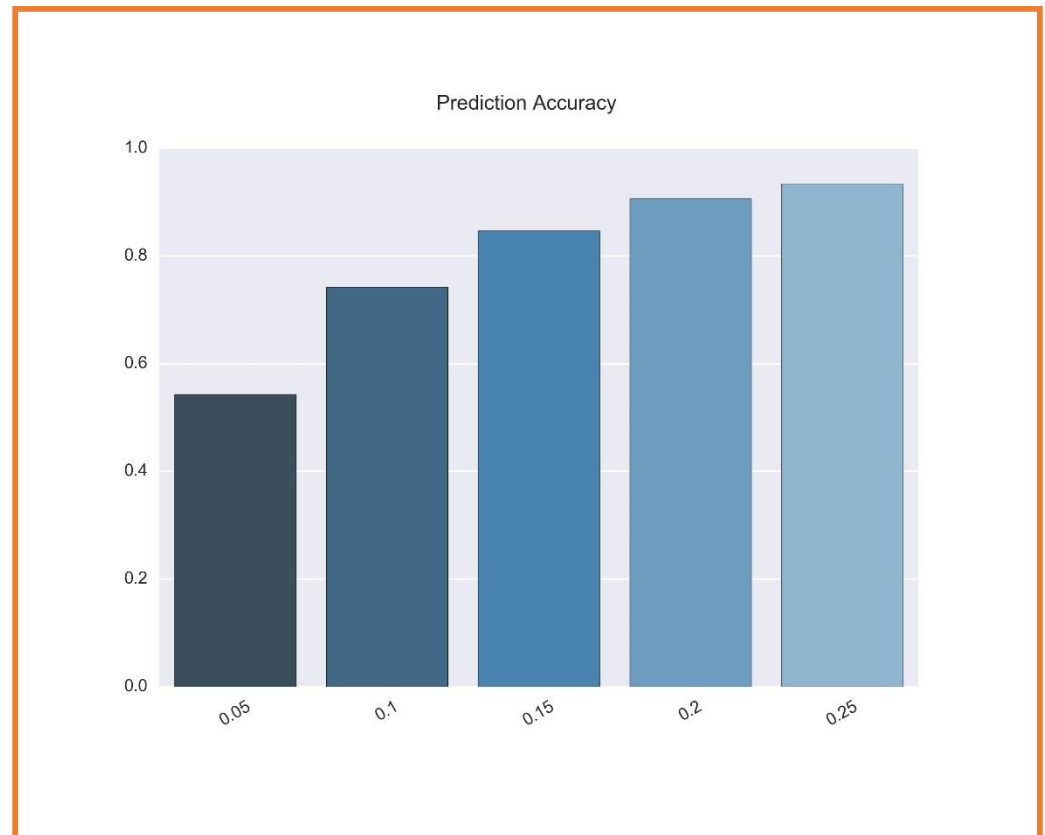
Analysis and Prediction

- Result
 - Important Feature
 - Prediction Result
 - Error Analysis



Analysis and Prediction

- Result
 - Important Feature
 - Prediction Result
 - Error Analysis



Analysis and Prediction

- Result
 - Important Feature
 - Prediction Result
 - Error Analysis

916 Barton Creek Blvd,
Austin, TX 78746
5 beds · 7 baths · 8,538 sqft



● FOR SALE
\$1,935,000
Zestimate*: \$2,760,235

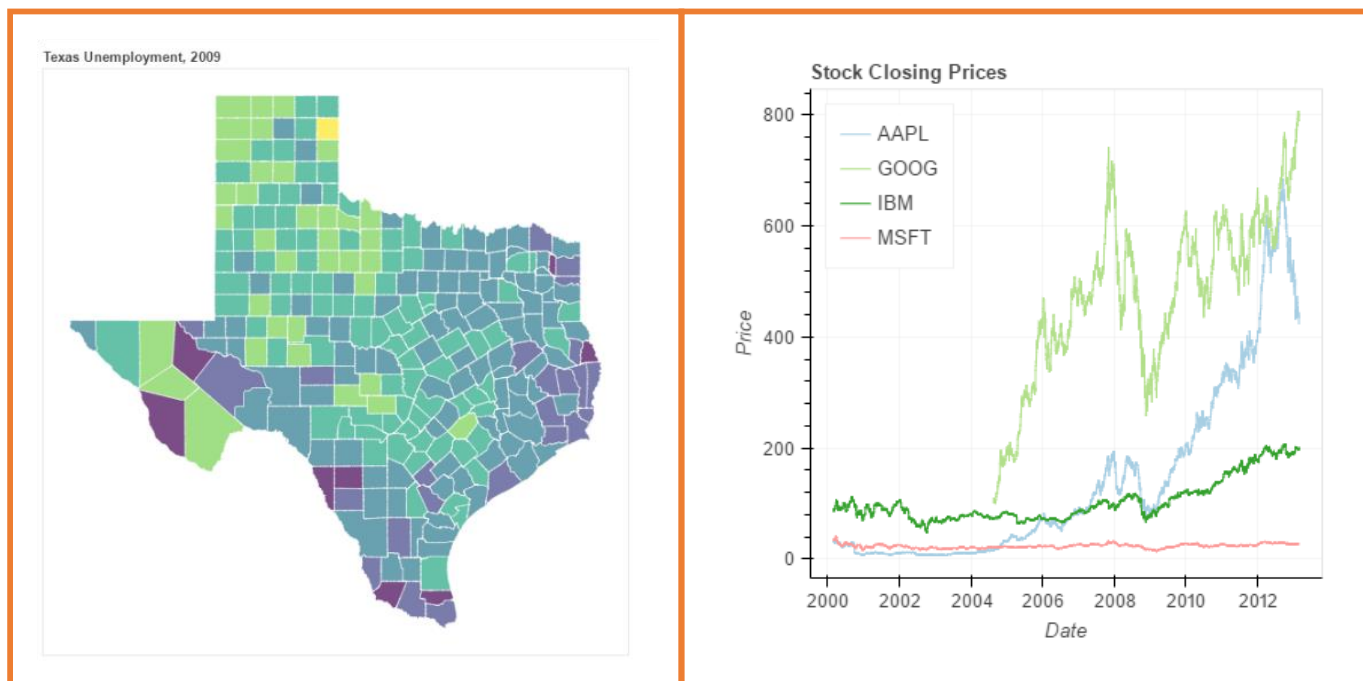


Visualization

- Key for a good data report:
 - Effective
 - Attractive
 - Comprehensive
- Interactive report
 - Interesting
 - One-for-All

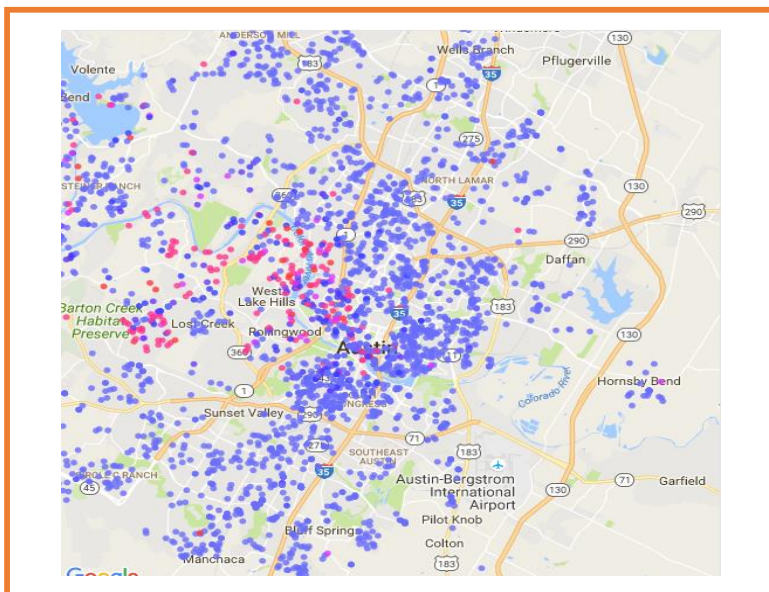
Visualization

- Bokeh: plotting, ColumnDataSource, Circle
- Examples:

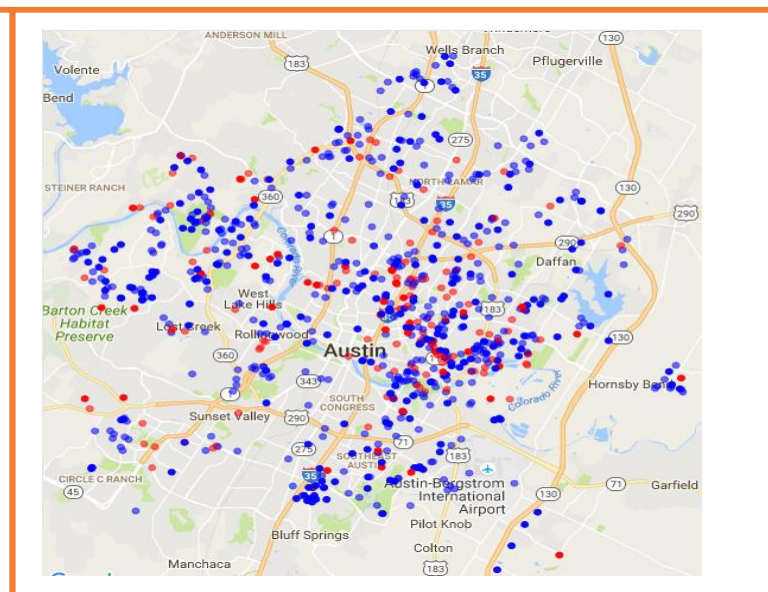


Visualization

Austin Houses for Sale:



Austin Predictions:





Future Work

- Scrapping more information (year built) from other website (redfin.com)
- Collecting more data from other Texas counties
- D3 interactive analysis



Q&A