Zillow Regression Project

Anna Vu

NEXT

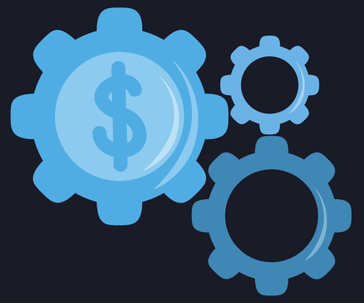
Agenda

1. Executive Summary
2. Data Wrangling
3. Data Exploration
4. Modeling
5. County Tax Rates
6. Conclusion/ Next Steps

Executive Summary

Goals:

Build a regression model to predict values of single unit properties using a Zillow data set



Best Features

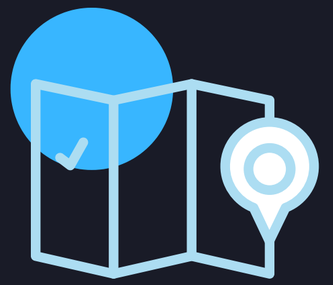
- Square footage
- Bathrooms
- Bedrooms
- Property Age

Takeaways

Polynomial Regression (degree=3):

RMSE = $272,168
R^2 = .403

Baseline:
RMSE = 357,185.61
R^2 = -3.07



Tax Rate Distribution

LA County: 1.4%

Orange County: 1.2%

Ventura County: 1.2%

Data Wrangling

Acquire + Prepare

Bring in 2017 Zillow data for single unit properties sold during May-Aug

- Observe data types
- Fill nulls or drop
- Remove outliers
- Rename columns
- Create new features

Data Exploration

Correlations

H0: No correlation between feature and tax value

Ha: Evident correlation between feature and tax value

Since p is 0.00, we reject the null

Feature	r	p
sqft	0.60	0.00
baths	0.51	0.00
beds	0.30	0.00
age	-0.19	0.00

Distributions

- Mostly normal after removing outliers
- Most properties are located in Los Angeles County

Stats Tests

Using t tests, reject that the size of features (except for age) and tax value are independent.

Modeling

Baseline

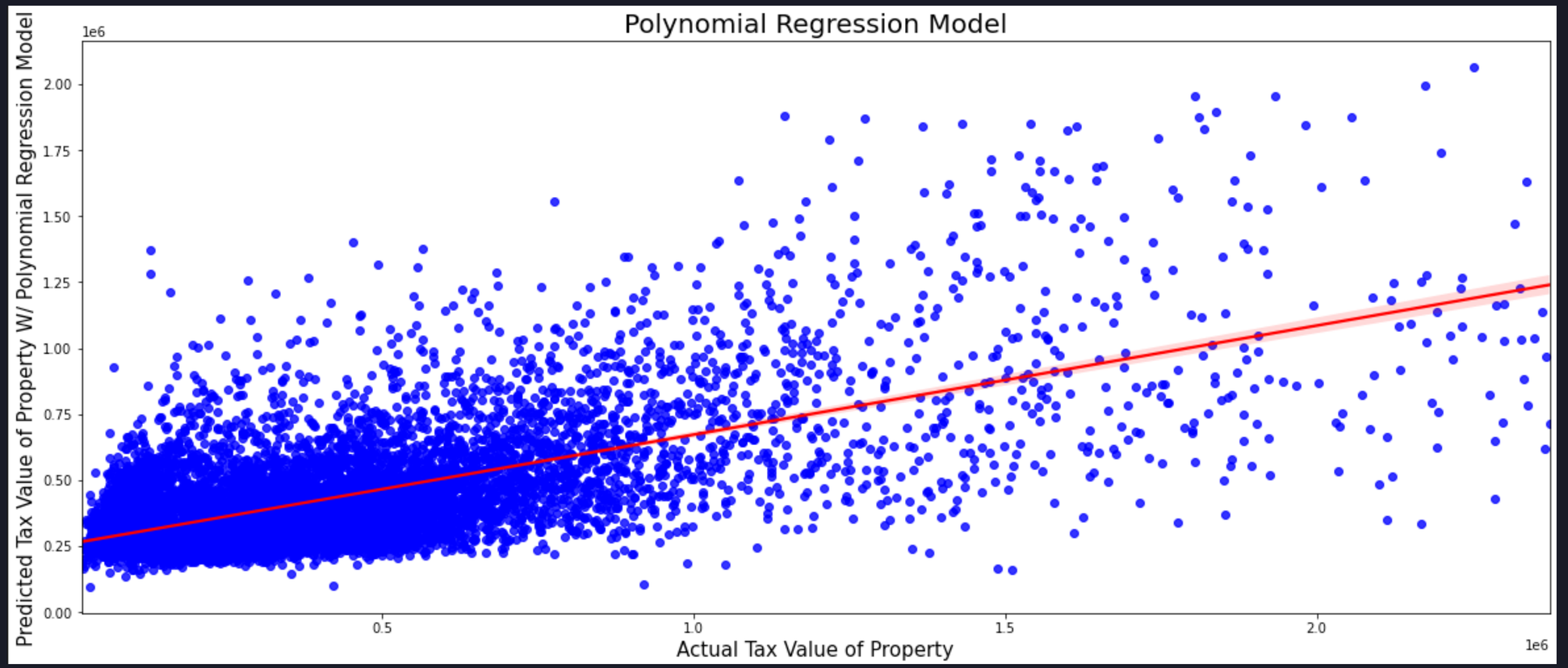
RMSE: 359,454.06

R^2 = -3.07

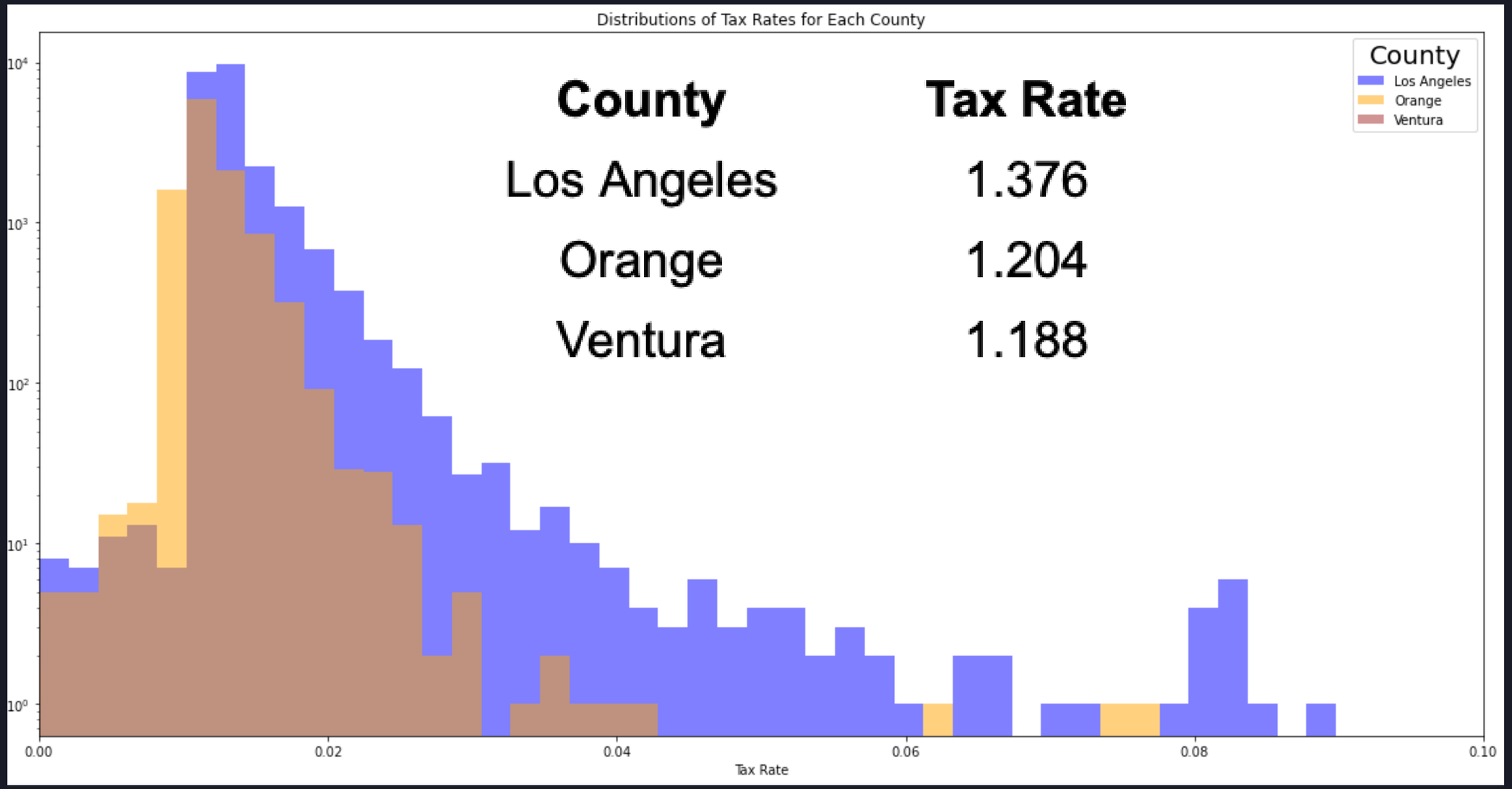
Polynomial Regression

RMSE: 272,168.27

R^2 =.403



County Tax Rates



County	Tax Rate
Los Angeles	1.376
Orange	1.204
Ventura	1.188

Conclusion

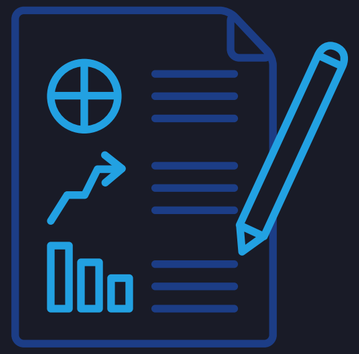
Best Drivers for Value:

- Square footage
- No. of Bathrooms

Other Drivers:

- No. of Bedrooms
- Age

Location:
Most of these properties were in LA, which also has the highest tax rates



Polynomial Model:
RMSE of 272,168.27, ~40% of variance can be explained by the model

Next Steps

- Look into more property features
- Research other ways to handle nulls/outliers
- Implement more useful location feature with latitude/longitude, zip code, etc.

Thank you!

For more information, resources, and a complete walk through of my process---visit my Github @annah-vu

NEXT