Zillow

Regression Project

James Allen

Our plans and where we're headed →

Let's Talk

TODAY'S MEETING AGENDA

01 Executive Summary

02 Data Discussion

03 County Taxes

04 Conclusion

Executive Summary

- **Single Unit Properties:** Linear Regression model to predict property values using 2017 data
- **Baseline: $363,349**
- **Prediction: $269,196**
- **Tax Counties:**
  - **Los Angeles: 1.38 %**
  - **Orange: 1.21 %**
  - **Ventura: 1.19%**

Data Science Pipeline

01	**Project Planning**	Initial thoughts and outline
02	**Acquire**	Getting necessary zillow data
03	**Prepare**	Cleaning and filtering data
04	**Exploration**	Visualizing and hypothesizing
05	**Modeling**	Conduct multiple model tests

Zillow Data

- **Zillow Database**
  - **properties_2017**
- **Acquire.py**
  - **Bring in data**
- **Prepare.py**
  - **Clean data**

Tools

- SQL
- Python
- Jupyter Notebook

Data Exploration

Stats Test: T-Test and Correlation Test

- # of Bedrooms positively related to tax value
- # of Bathrooms positively related to tax value

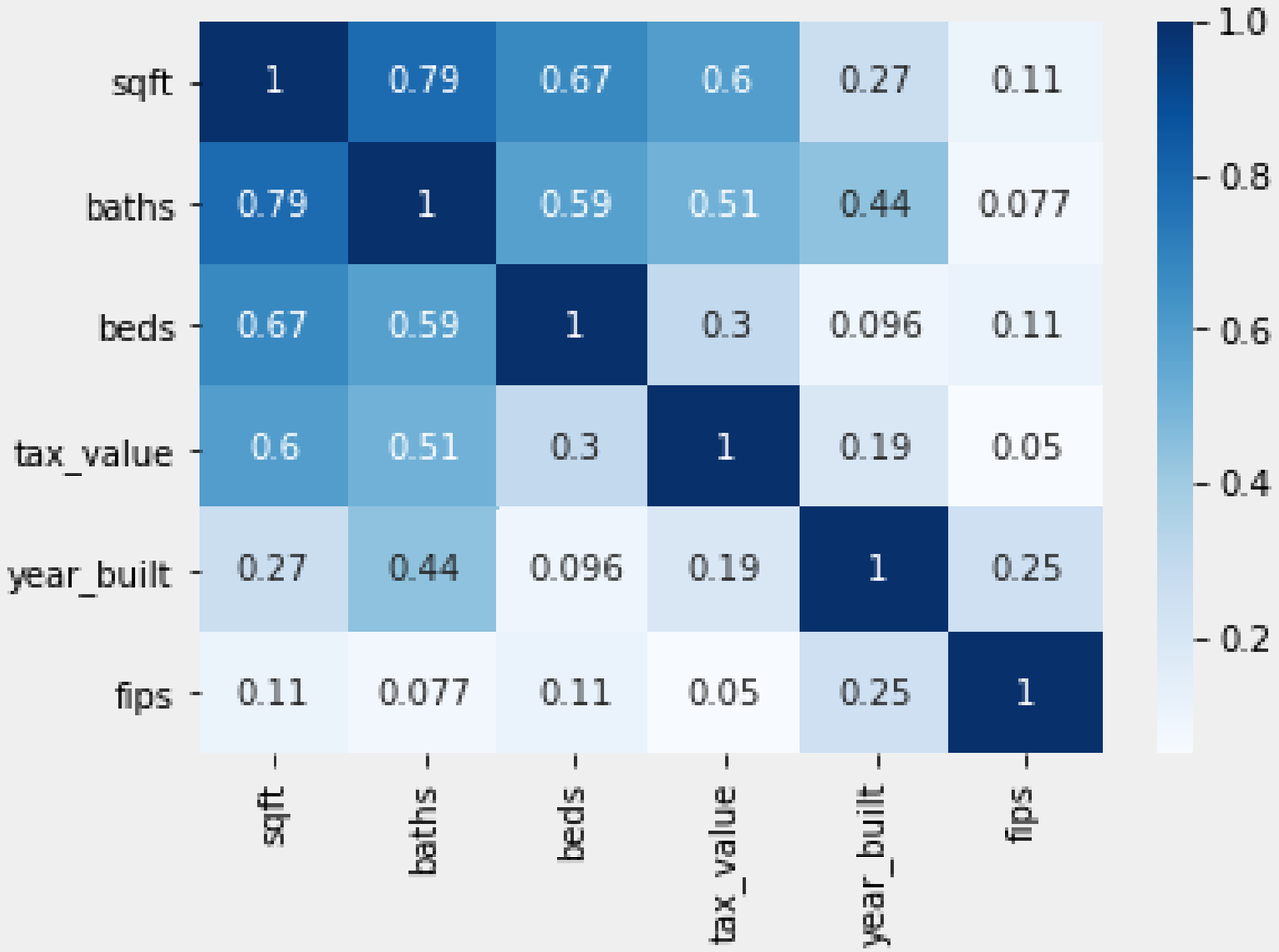
Data Visualization:

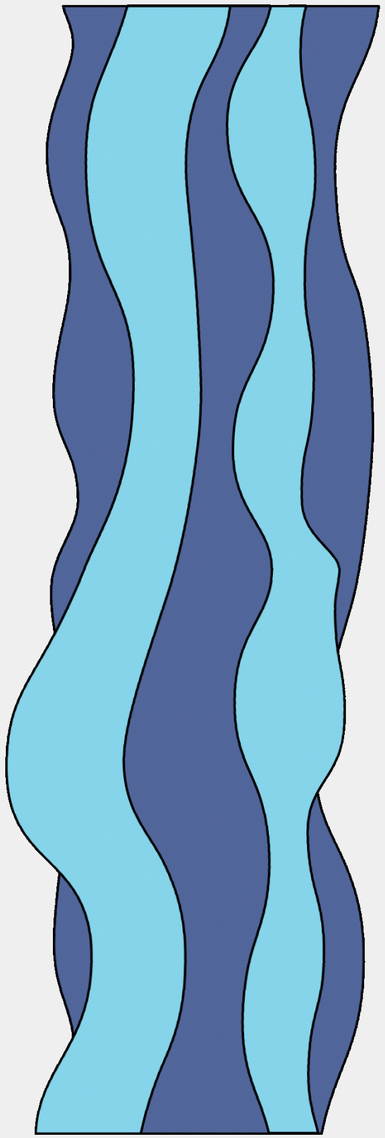
- Pairplots, Histograms, Heatmaps

Distributions:

- Los Angeles County has the largest amount of Properties
- 3 Bedrooms, 2 Bathrooms most popular

HeatMap Correlation





Modeling

Baseline

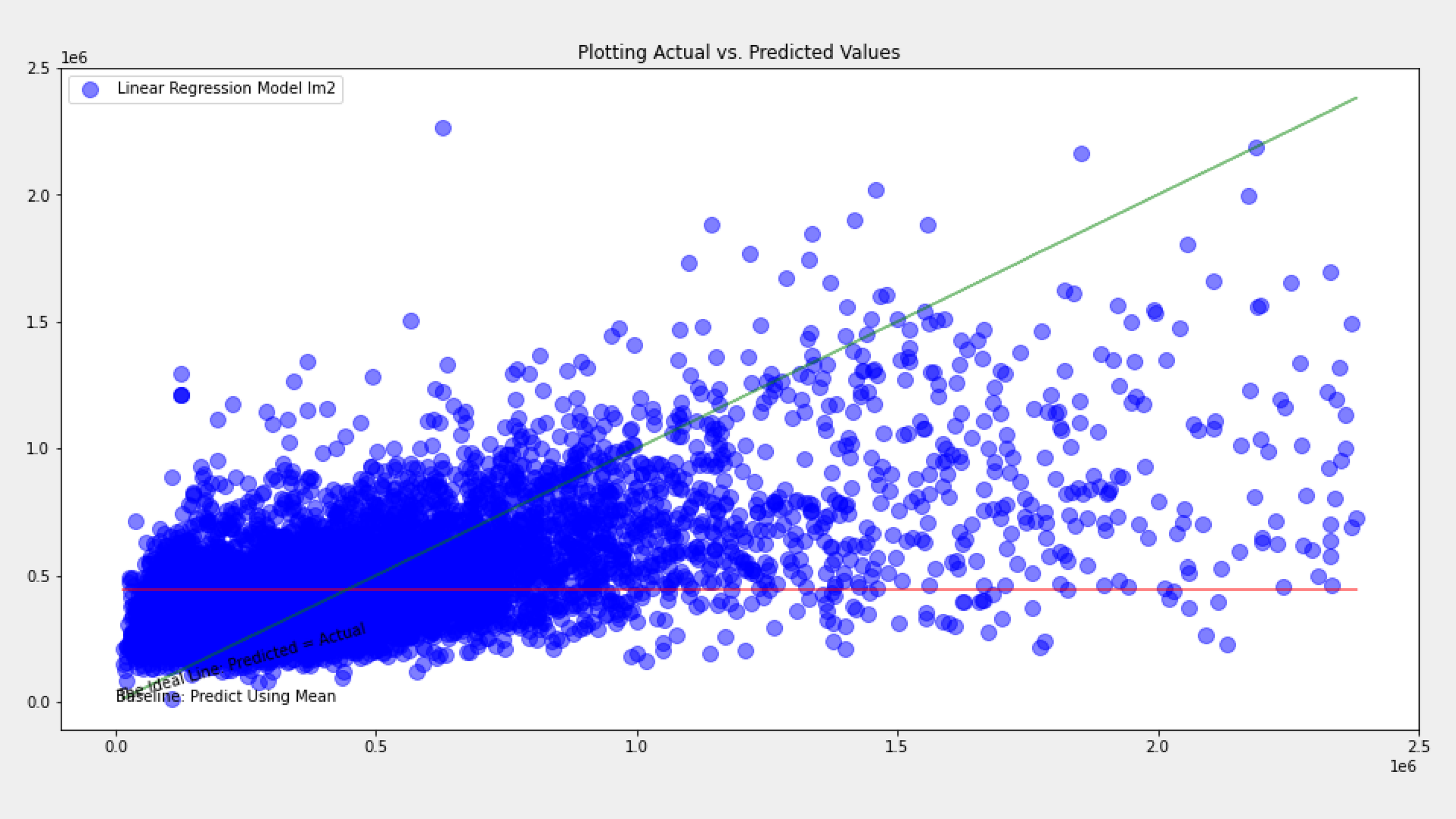
RMSE: $363,349

r^2: -0.0003

Polynomial Regression

RMSE: $269,196

r^2: 0.41

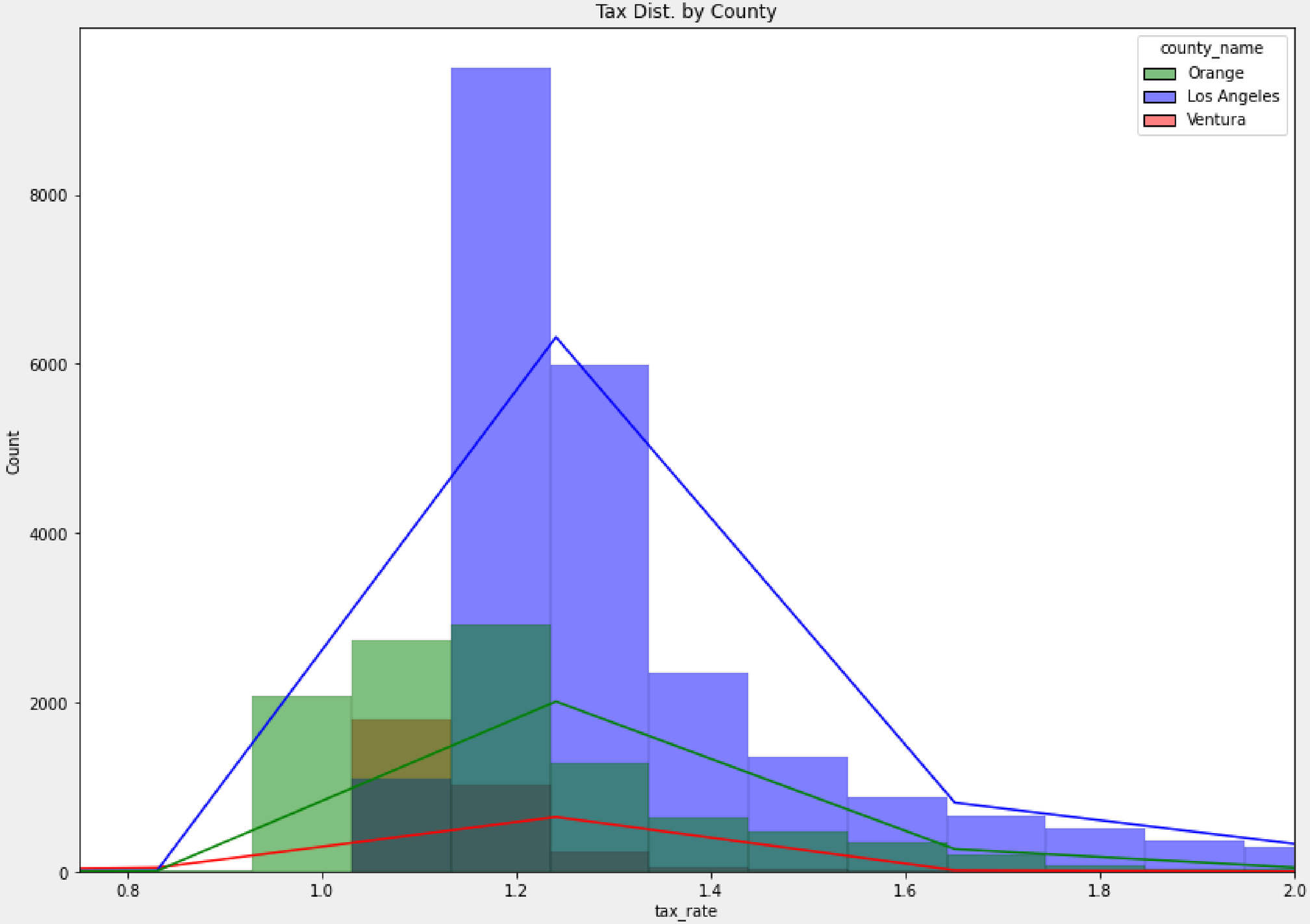

Plotting Actual vs. Predicted Values
1e6
Linear Regression Model lm2
The Ideal Line: Predicted = Actual
Baseline: Predict Using Mean
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
1e6

Tax Counties

Los Angeles: 1.38 %

Orange: 1.21 %

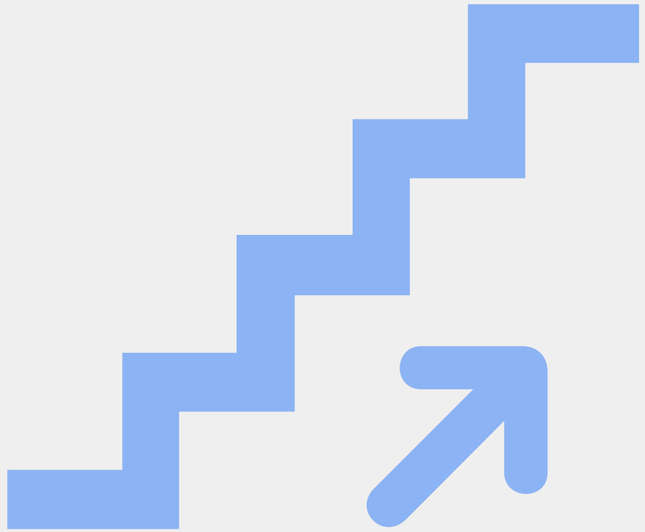
Ventura: 1.19%


Tax Dist. by County
county_name
Orange
Los Angeles
Ventura
8000
6000
4000
2000
0
Count
0.8
1.0
1.2
1.4
1.6
1.8
2.0
tax_rate

Conclusion

- Single Unit Properties: 2017 Data
- Linear Regression Model
- Baseline: $363,349
- Prediction: $269,196
- Tax Counties:
  - Los Angeles: 1.38 %
  - Orange: 1.21 %
  - Ventura: 1.19%

Next Steps



With more time I would like to:

- **Explore the year built and age of property compared to property value**
- **Refine my models with different parameters and run additional tests**
- **Map out county locations and run models on specific counties**

Thanks for listening!

Reach out to us if you have any questions or concerns.