



Analysis of Property values

Ian Johnson
September 15, 2021

Today's Agenda



Key takeaways:

- Executive Summary
- The Dataset
- Key findings
- Summary

Executive Summary



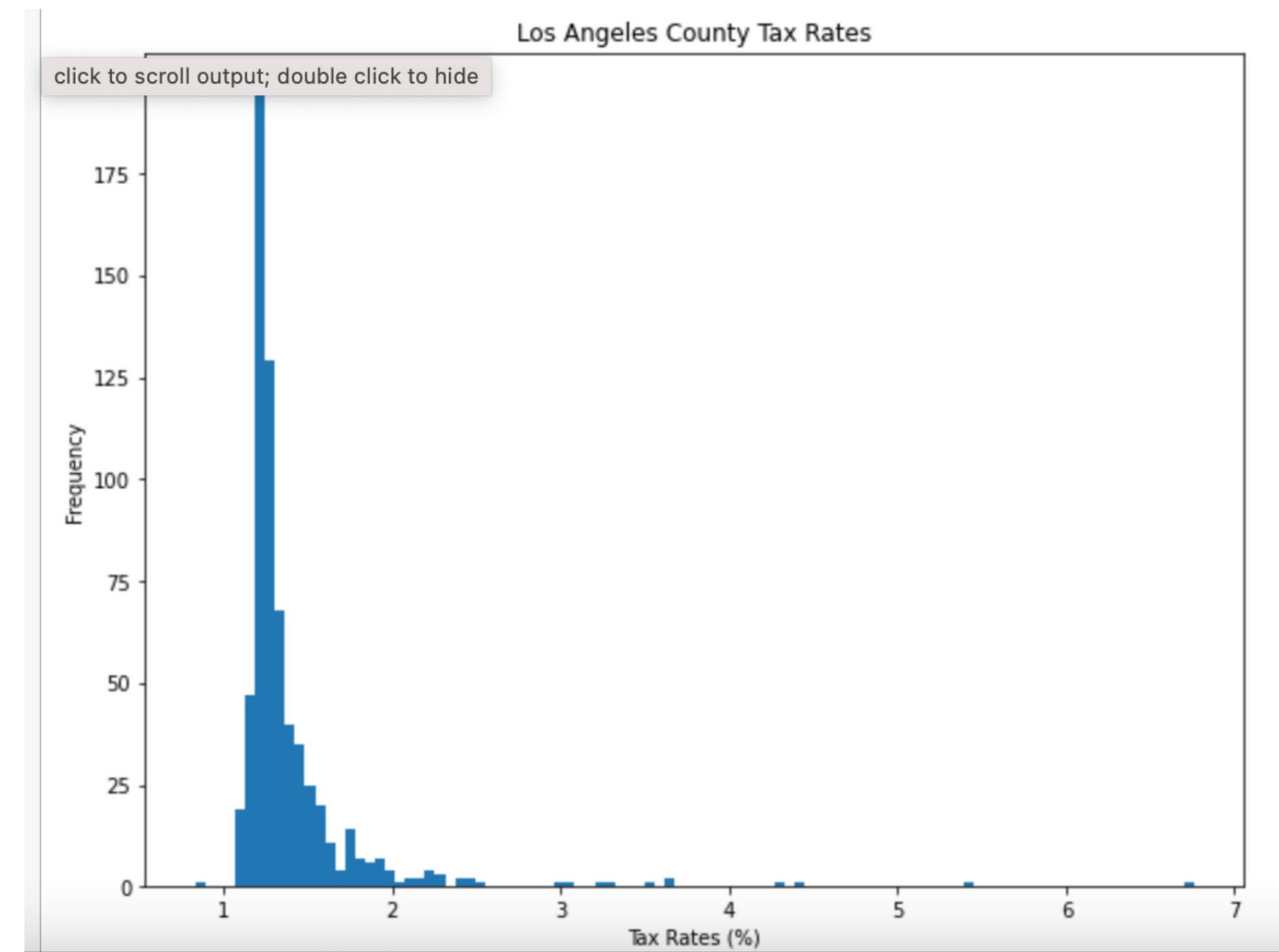
Key takeaways:

- A method of identifying property values is needed
- Machine learning can be used to predict property values
- The best predictors of property are:
 - Area (sq ft) of property
 - Number of bedrooms
 - Number of bathrooms

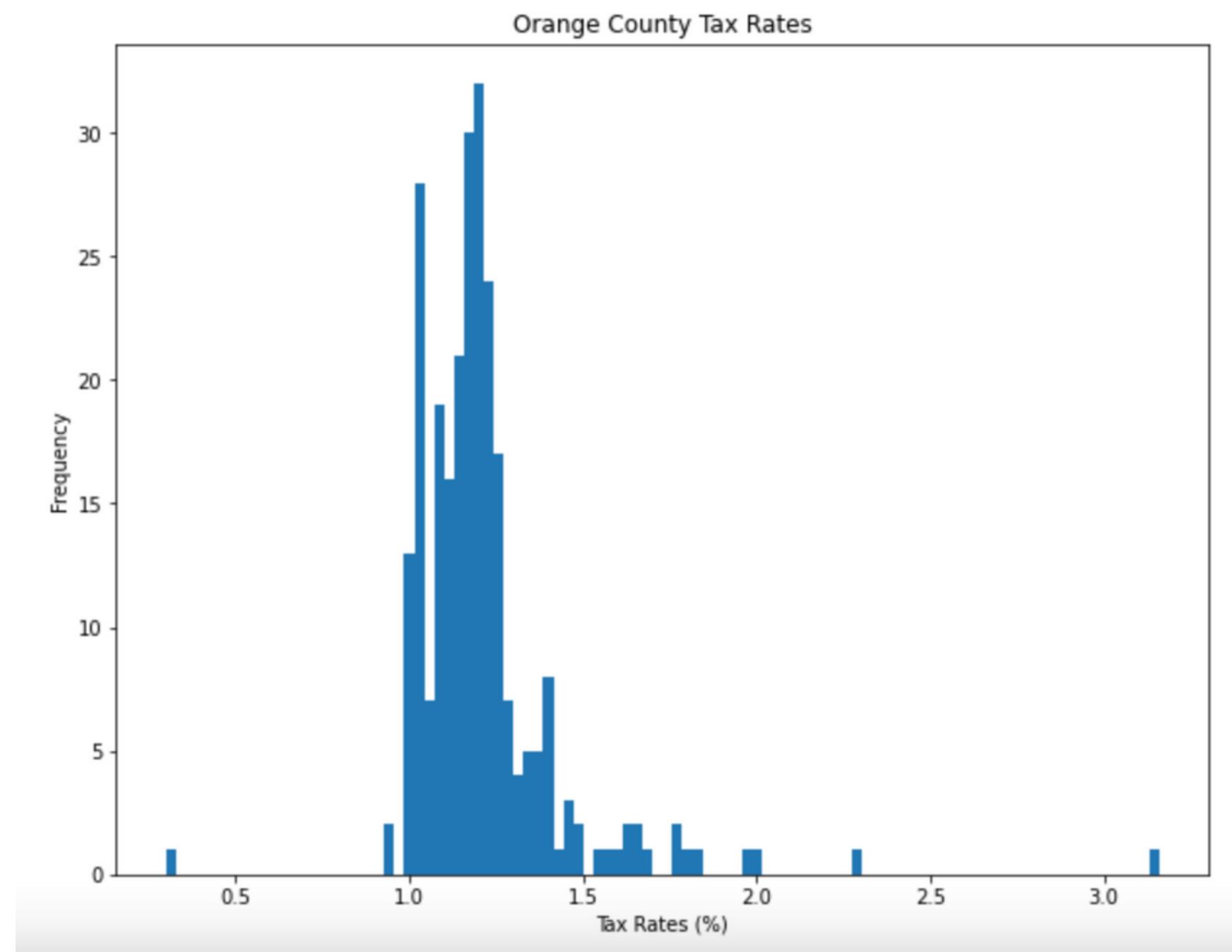
Zillow Dataframe Containing Property Information

	bedroom	bathroom	area	taxvalue	yearbuilt	taxamount	fips
231660	2.0	1.0	810.0	303199.0	1950.0	3740.97	6037.0
205263	4.0	2.0	1875.0	243858.0	1964.0	2993.74	6111.0
350760	3.0	2.0	1768.0	208122.0	1960.0	2834.69	6037.0
283943	3.0	2.0	1527.0	275999.0	1929.0	3661.24	6037.0
292376	3.0	1.0	1191.0	60014.0	1949.0	892.84	6037.0

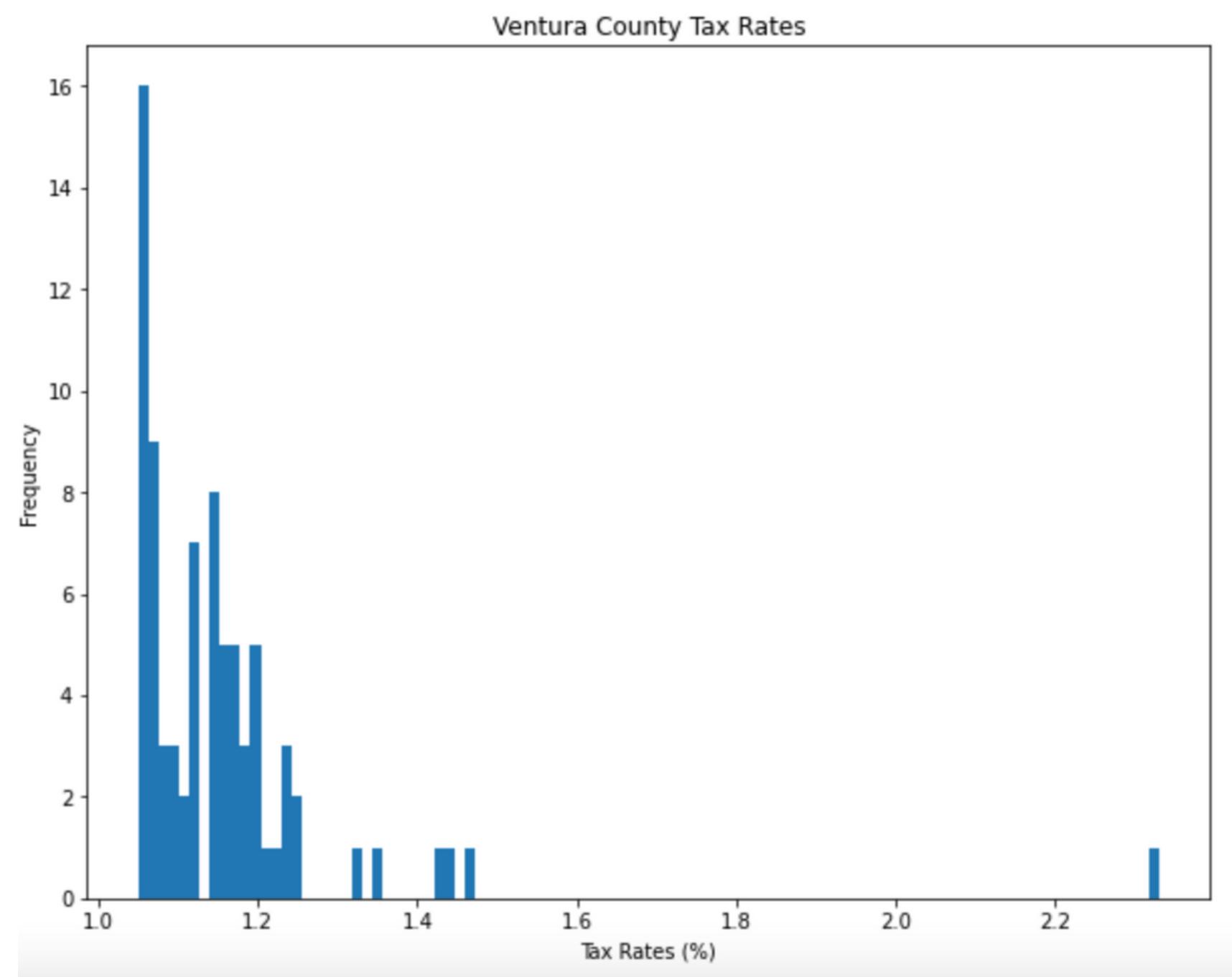
Tax Value of Properties in Los Angelos County



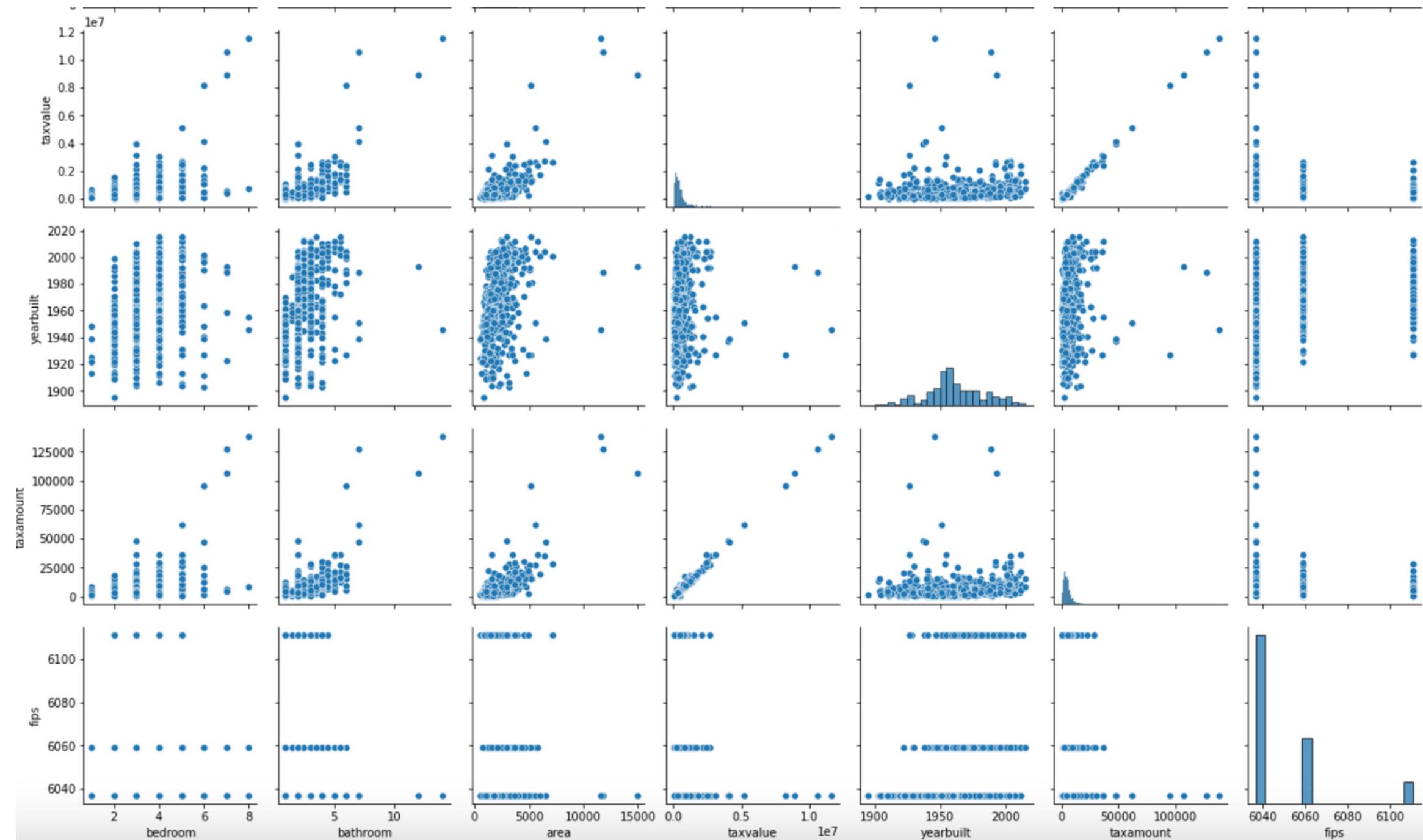
Tax Value of Properties in Orange County



Tax Value of Properties in Ventura County



Pair Plot of Features



Correlation Table of Features

	bedroom	bathroom	area	taxvalue	yearbuilt	taxamount	fips
bedroom	1.000000	0.683797	0.642334	0.394968	0.375808	0.392913	0.142556
bathroom	0.683797	1.000000	0.864450	0.656396	0.464806	0.654048	0.082278
area	0.642334	0.864450	1.000000	0.750119	0.410388	0.750703	0.081865
taxvalue	0.394968	0.656396	0.750119	1.000000	0.136755	0.995941	0.004068
yearbuilt	0.375808	0.464806	0.410388	0.136755	1.000000	0.139309	0.300508
taxamount	0.392913	0.654048	0.750703	0.995941	0.139309	1.000000	-0.017461
fips	0.142556	0.082278	0.081865	0.004068	0.300508	-0.017461	1.000000

Features



Select Best

Bedroom

Bathroom

Area

RFE

Bedroom

Bathroom

Area

Models on Train Data



Model Selection

- Median Baseline
RMSE = 689912
 $R^2 = 0$
- Linear Regression
RMSE = 317697739864
 $R^2 = -66539514414$
- LassoLars
RMSE= 562670
 $R^2 = 0.31$
- Generalized Linear Model
RMSE = 677117
 $R^2 = 0$



Summary and Conclusions

THE MOST PREDICTIVE MODEL

LassLars

THE MOST IMPORTANT FEATURES

The three best predictive features for property value are:
area (soft), bedrooms, and bathrooms.

County does affect property values, but is not one of the
three top drivers

Many of the features have correlation with each other.

FUTURE WORK

Impute null values instead of dropping them.

Feature engineering.

Examine features from other tables on SQL database.