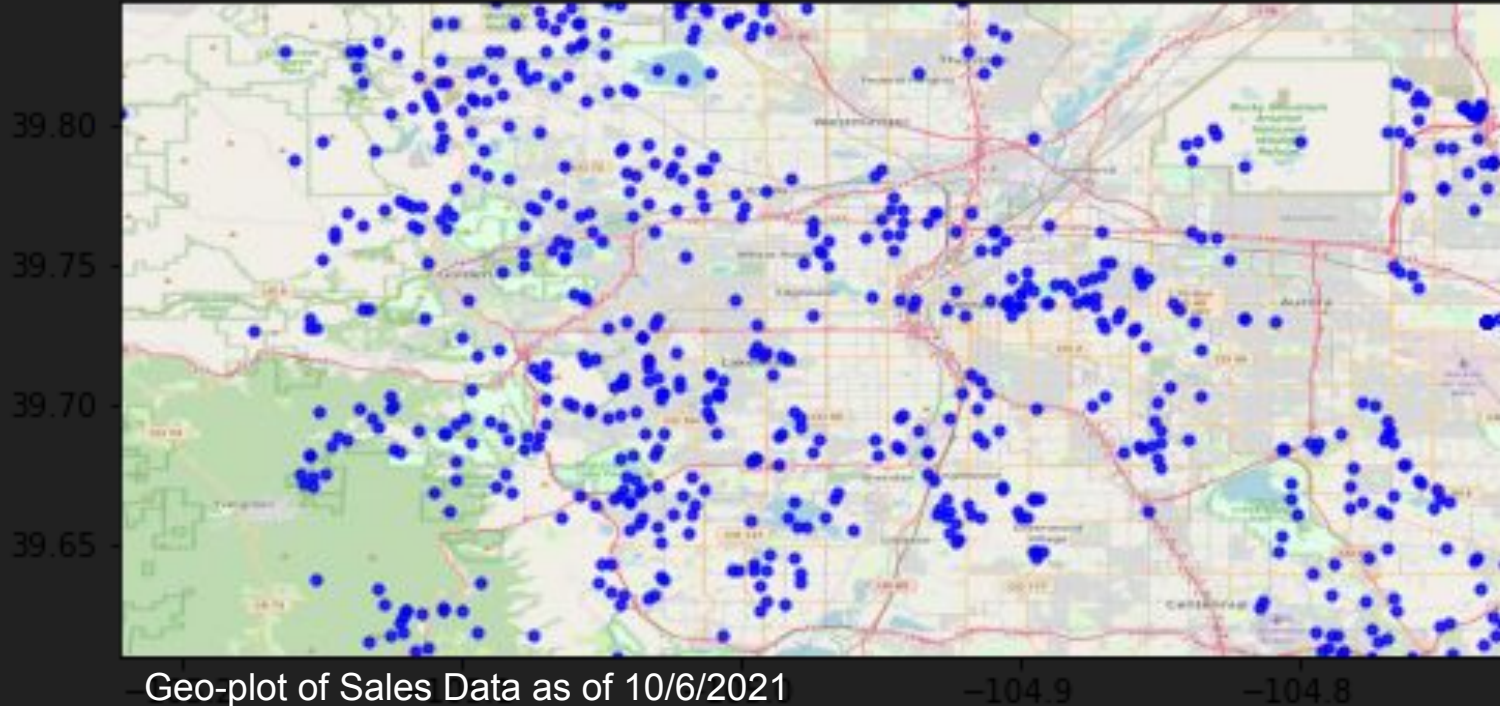# Predicting Success of Real Estate Investments
Melody Cefalo

Data Science Capstone Project, Springboard
October 11, 2021

With Thanks to Juanjo Milla,
Head Data & AI at Jeff, and
President of ANBAN

Where to invest??
Can we use an algorithm to predict which houses are a 'good' investment?

Geo-plot of Sales Data as of 10/6/2021

What makes a 'good' investment?

➢ Capitalization rate which exceeds investor's threshold

➢ Includes all costs of ownership, including opportunity cost of money on *current value of investment*

➢ Does not include potential increase in value of investment!

Estimated Cap Rate of Invitation Homes' Properties:

2.96% ← Our 'Buy' Indicator

# Primary Goal of Model - Capitalization Rate

What this model will do:

Predict market rent, calculate capitalization rate, and select appropriate investments from a list of sales data

What this model will not do:

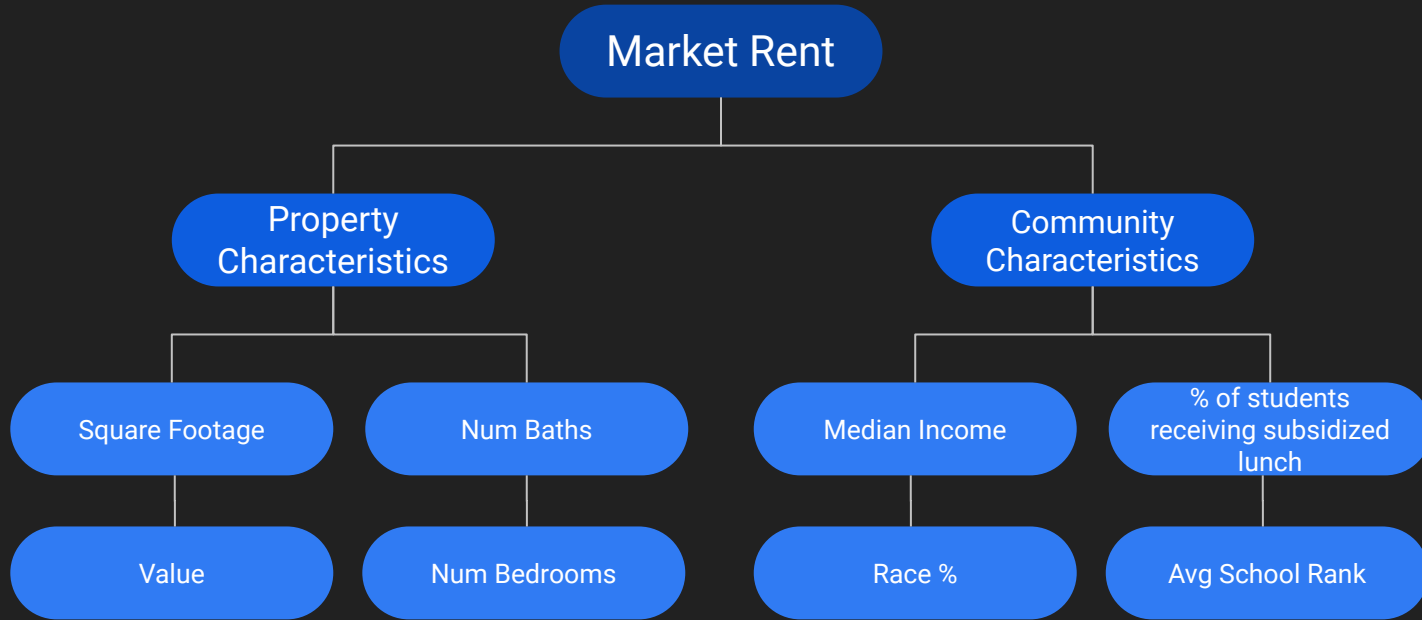Predict who will pay their rent, or predict which investments will appreciate/depreciate over time

Who can use a predictive model?

Individual Investors need an advantage:
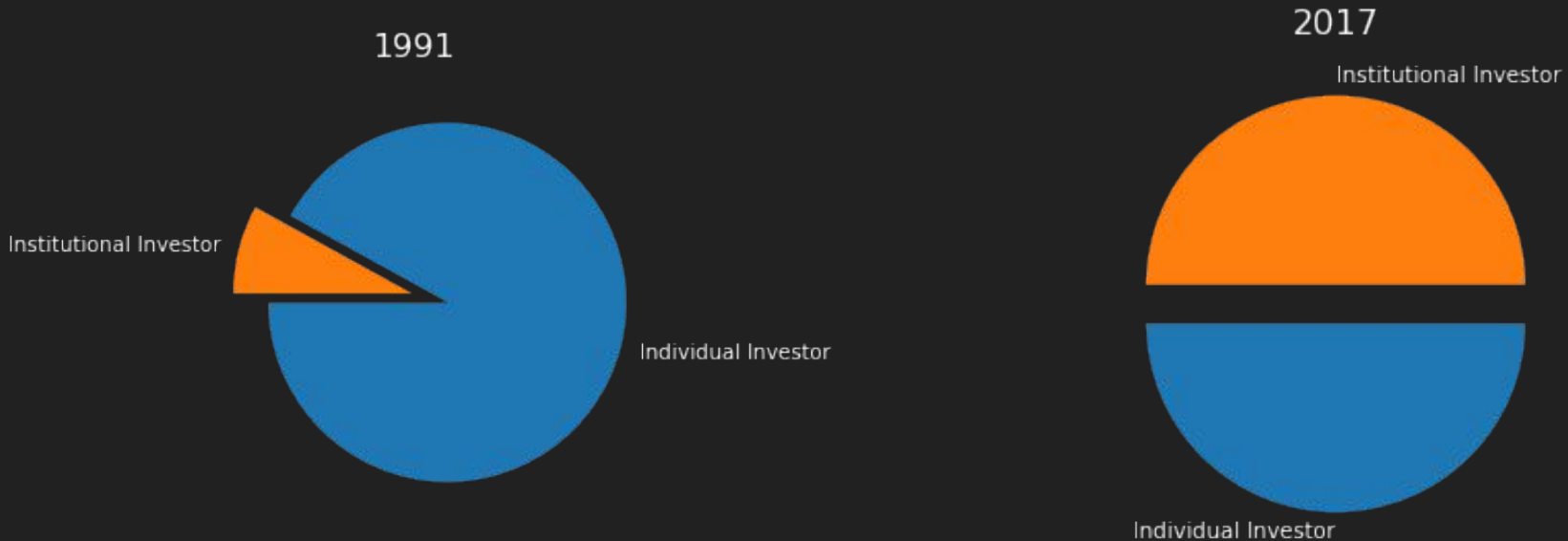
Data taken from https://www.census.gov/prod/1/statbrief/sb96_01.pdf,
'who Owns the Nation's Rental Properties?"

What might affect market rent?

Data Sources: Invitation Homes, Denver MLS listings, RealtyMole API

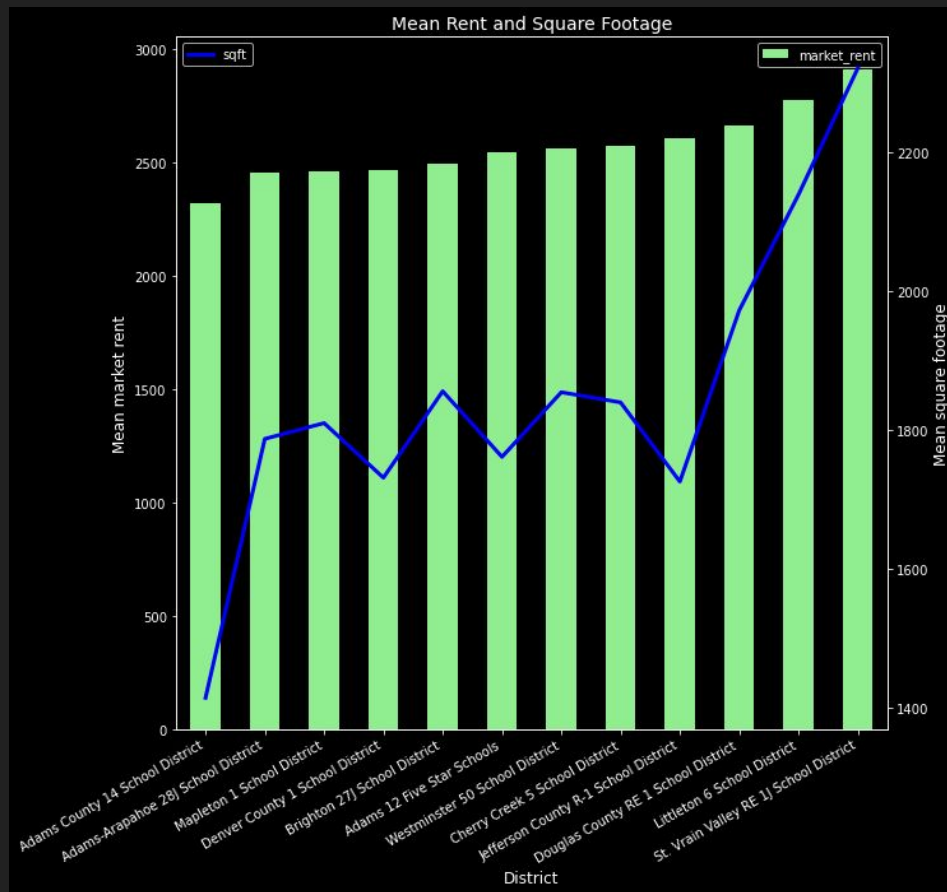Data Sources: SchoolDigger, US Census, and State of Colorado (data.colorado.gov)

Data Information

Invitation Homes ~ 250 Properties in geographical boundary, scraped from website over a period of three months
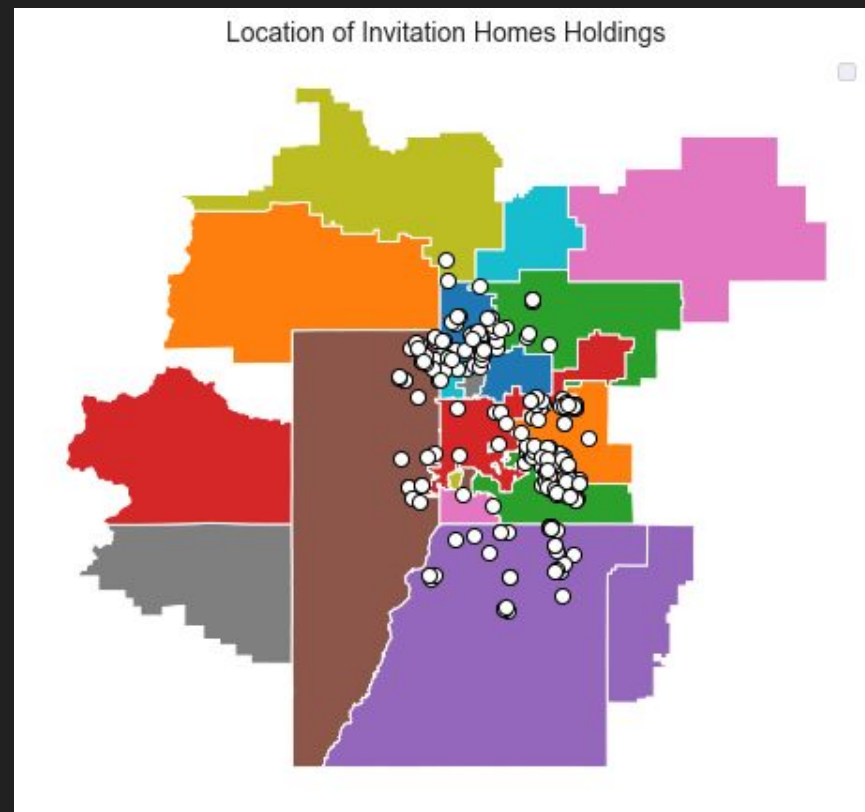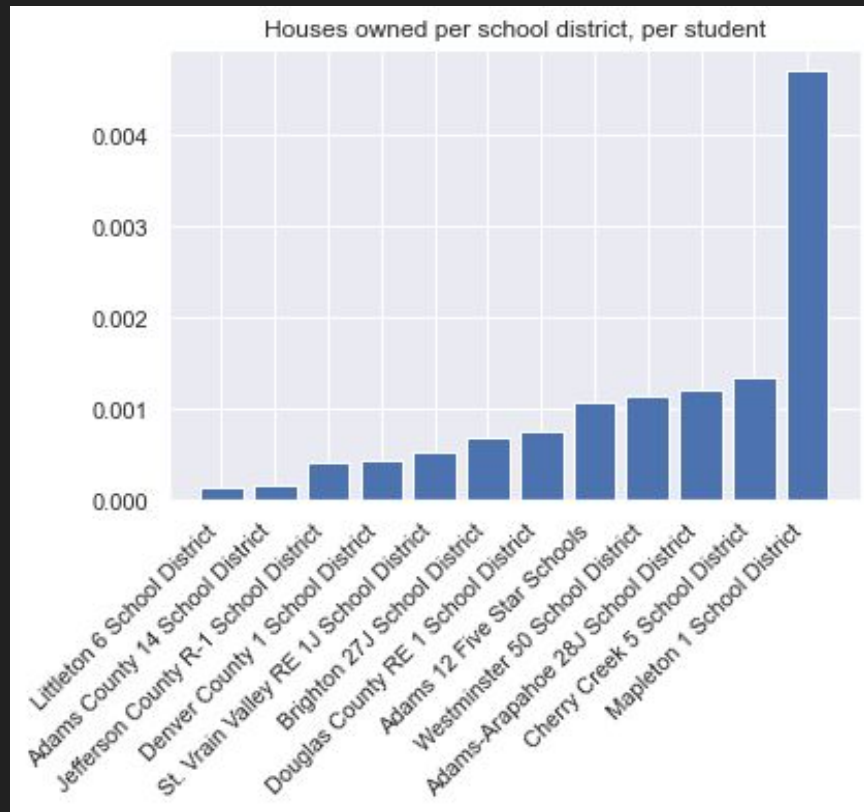
Sales ~ 700 Properties for sale in boundary, taken 10/6/21

EDA: Clear Correlations

Market Rent and Square Footage:
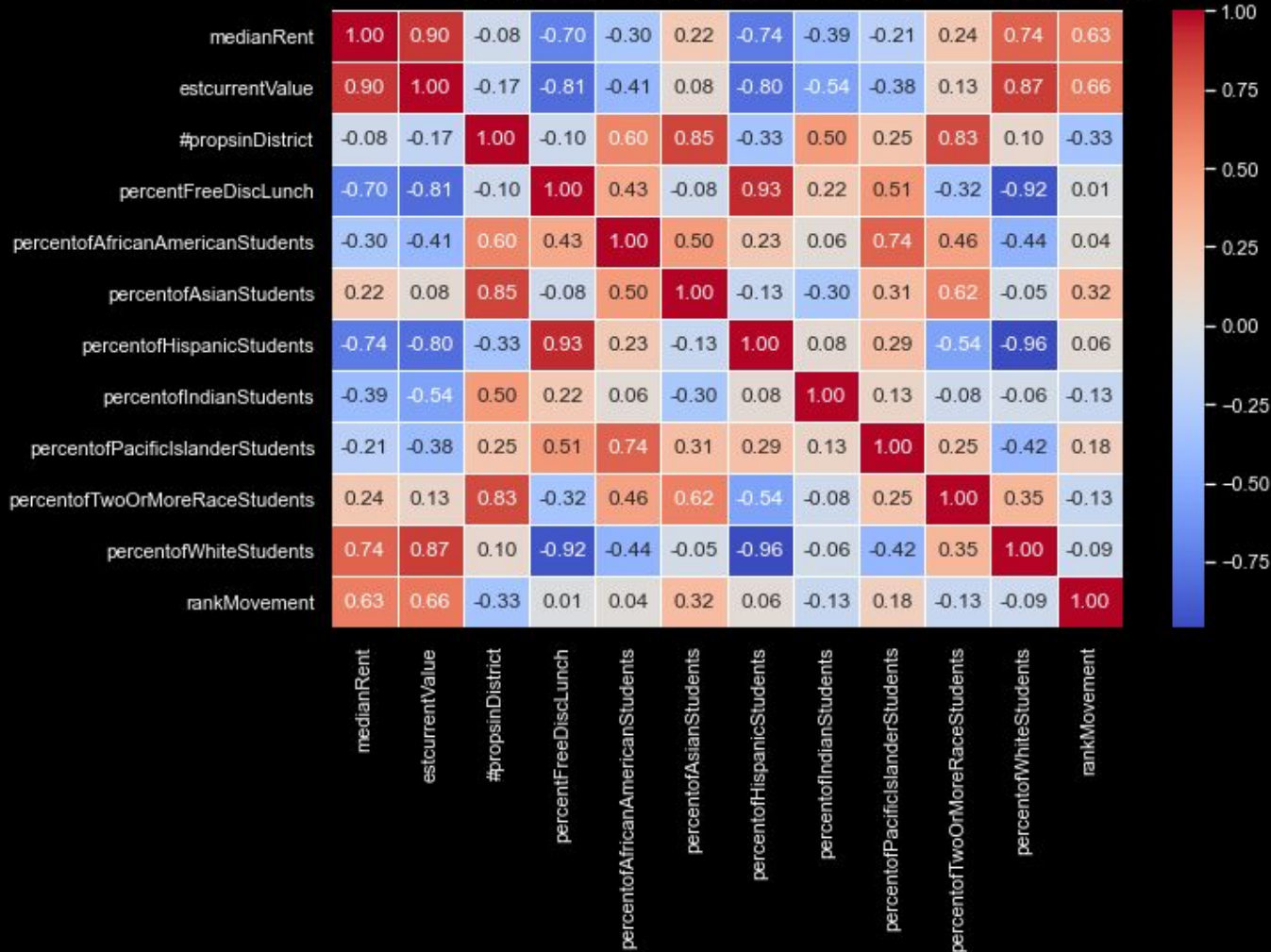
Positively Correlated!

EDA: Invitation Homes are Clustered - Geographically, and by Student Population

Dropping the 'Value' feature: Correlation Coefficient of .90

House Inc Attributes Grouped by District: Correlation Heatmap

Model Overview

Supervised Learning

Regression Model

Tool: Scikit-learn
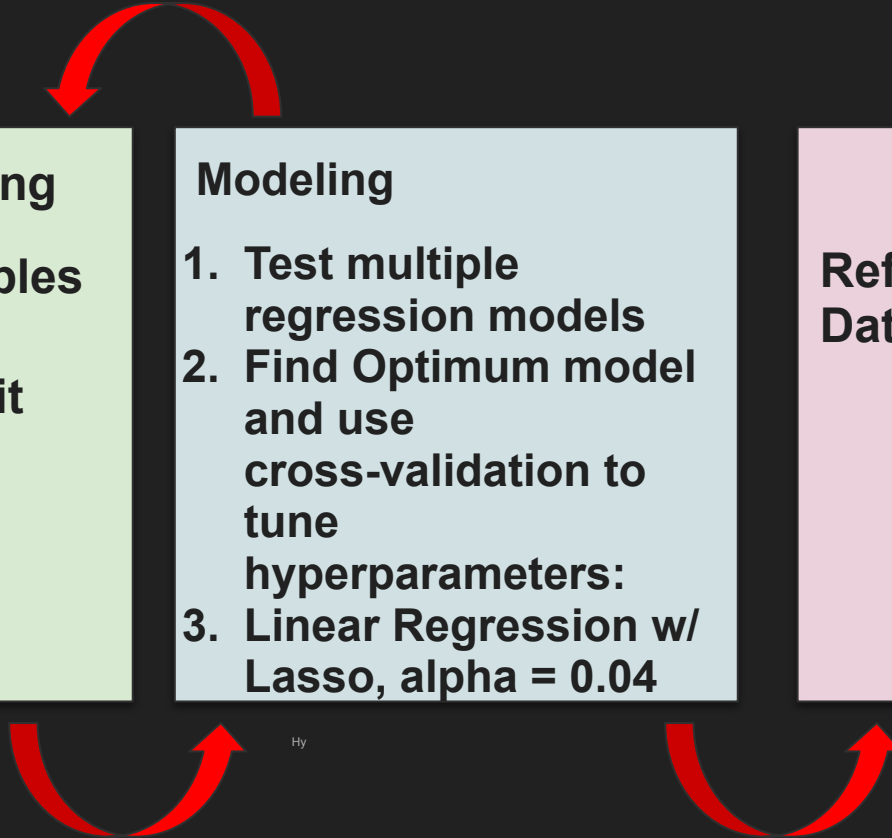
Data Handling and Modeling

Data Pre-Processing

1. **Dummy Variables**
2. **Scaling**
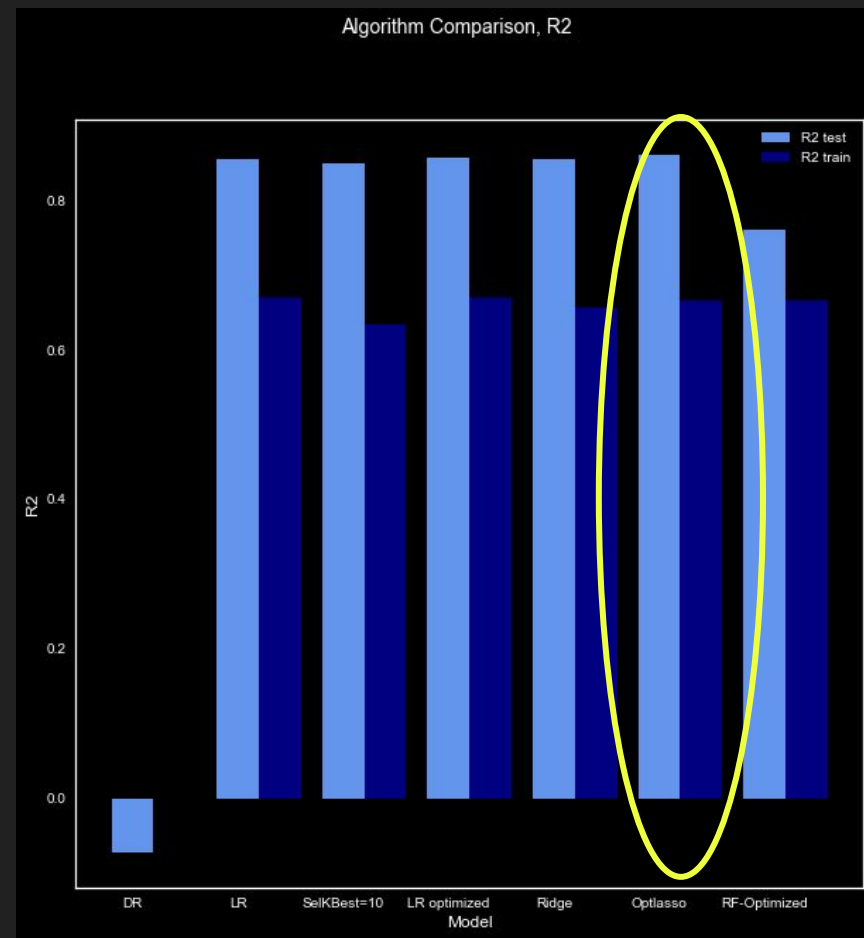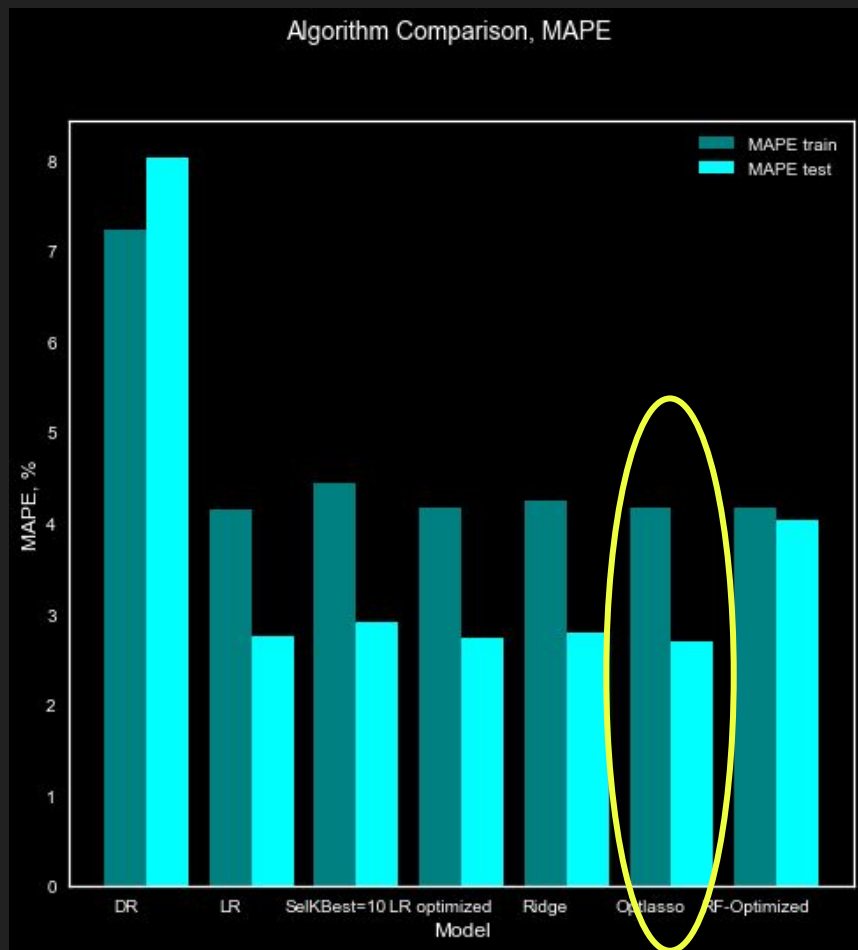3. **Train/Test Split (95% - 5%)**

Modeling

1. **Test multiple regression models**
2. **Find Optimum model and use cross-validation to tune hyperparameters:**
3. **Linear Regression w/ Lasso, alpha = 0.04**

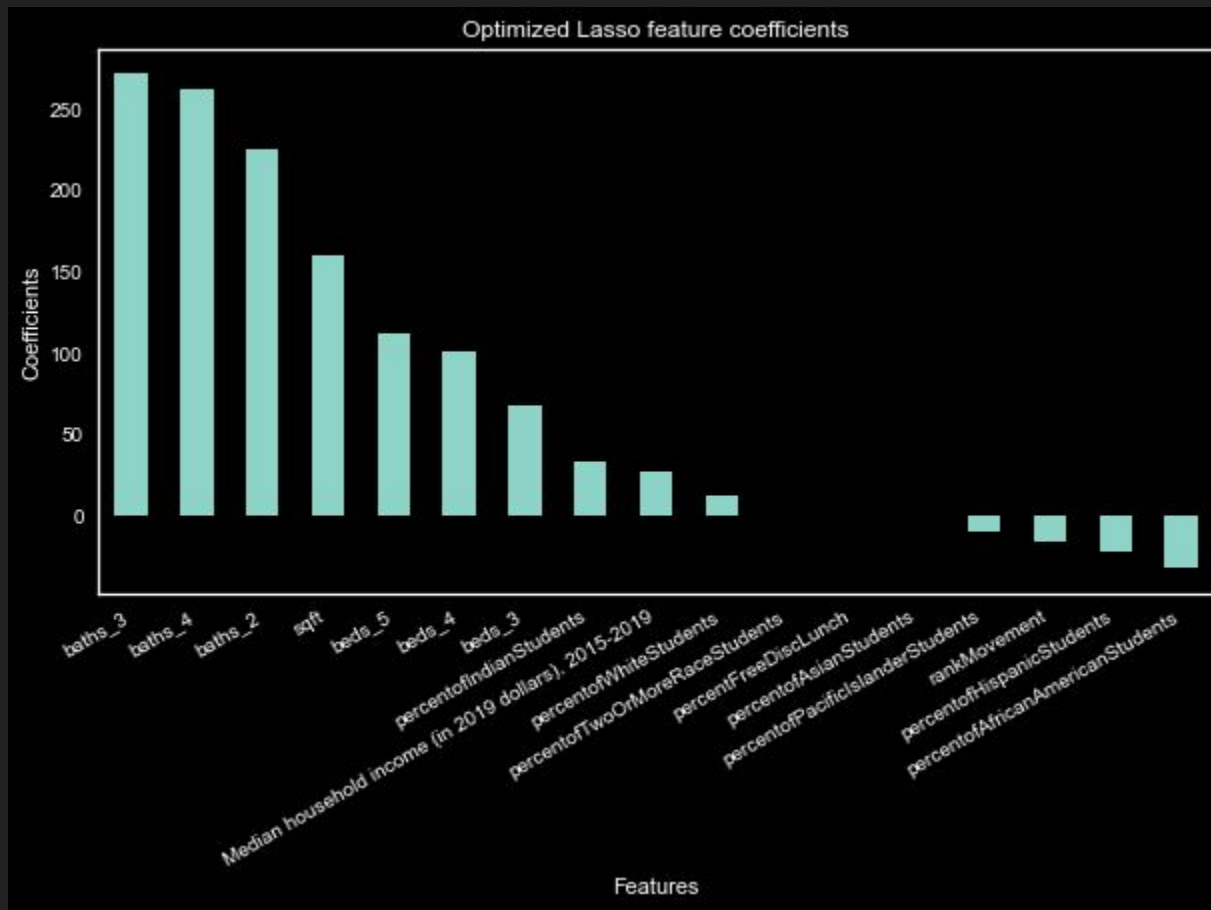**Refit on All Available Data!**

Hy

Model Comparisons

Linear Regression w/ Lasso Regularization

Eliminated Features:

17 → 3

- % 2 or more race students
- % receiving subsidized meals
- % asian students

Optimized Lasso feature coefficients

Heteroscedasticity / "Spray"??
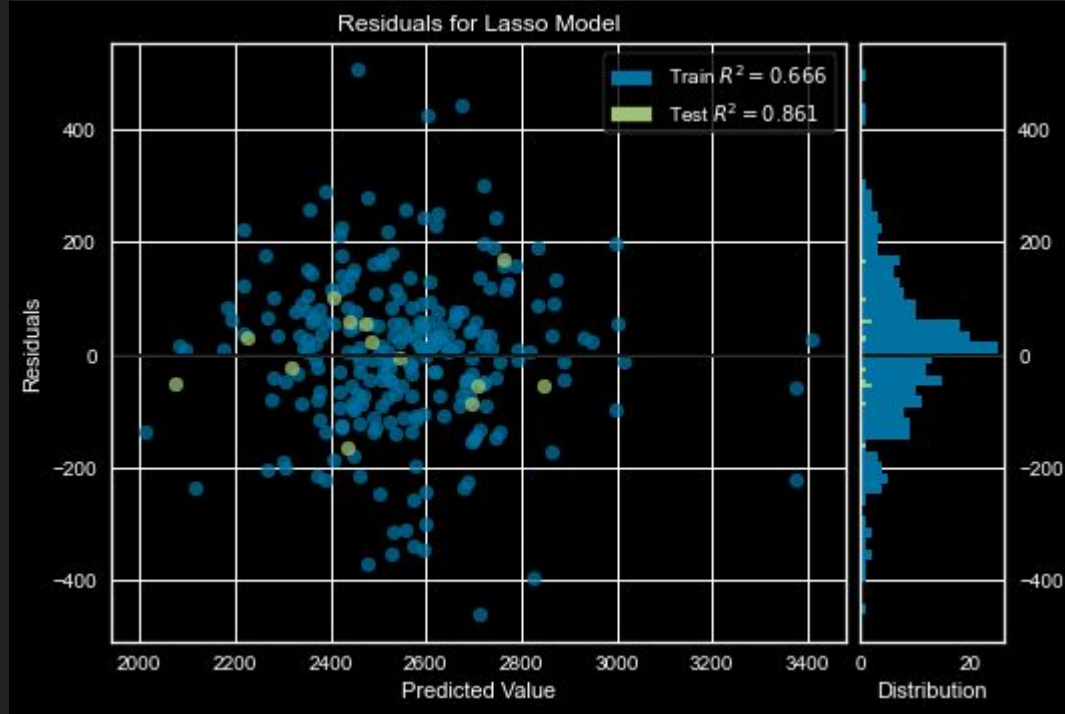
Error is Randomly Scattered Around Zero -

Variance of output (y) does NOT depend on input (X)

Helping Our Investor - Applying the model to sales data

1. Predict market rent for all sales listings

Example of output:

Address	beds	baths	market_rent
19632 W. 59th Ave	4	3	2822.42
2202 Braun Court	5	3	2884.40
642 Deframe	4	3	2801.49

2. Calculate Capitalization Rate

Assuming we want the SAME or BETTER cap rate as Invitation Homes -

Estimated Cap Rate for Sales Properties vs Corporate Desired Cap Rate
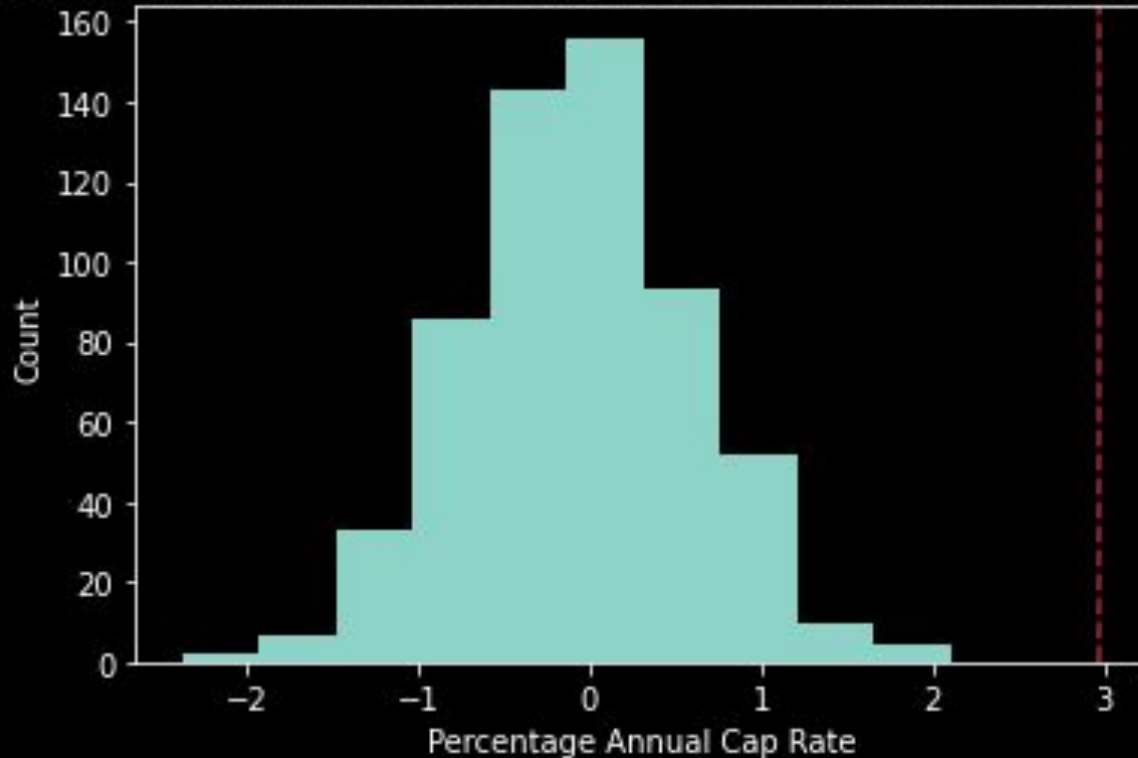
NO PROPERTIES meet Cap Rate

Assumptions and Limitations

- We don't know if Invitation Homes priced their rentals consistently - and their pricing could be based on confidential information

- The fidelity of income and race data is by school district, NOT by neighborhood

- We only have access to their properties which have gone through 'turn-over'; that is, they have been leased out since May, 2021

Conclusions

- Our data is likely still experiencing collinearity; for example, there are only so many races, and everyone is in a category, so all races shouldn't be included in the analysis
- More interesting data can likely be found for school performance including school discipline, PSAT scores, or perhaps police activity in proximity to rental house location
- Mainly, rent depends on number of bathrooms, number of bedrooms, size, and race components of school district - though this brings up many social concerns, this project is observational and cannot answer the 'why'

It's probably not a great time to buy in the Denver metro area, unless you find something "off-market" - Estates, word-of-mouth, cash sales - ZERO properties listed on MLS meet the mean cap rate established by Invitation Homes!