Why is Rent so High? (West Coast Edition)

By: John Lassetter
Sept. 31 2021

Question

How much will an apartment rent for based on its description?
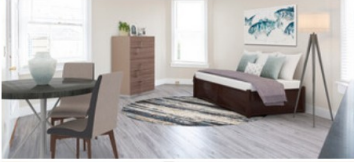
Ex. features:

- Location
- Square Footage
- Number of bedrooms/bathrooms

Data

- Address

- Monthly rent

- #Bedrooms/Baths

- Sq. Feet

- Build Date

- Nearby Colleges

- Nearby Airports

- More: pets allowed, Amenities

Data – Multiple Listings per Page

Luma

1440 Columbia St, San Diego, CA 92101

Little Italy

★ ★ ★ ★ ★ (0 reviews) ⌄   ⊘ Verified Listing

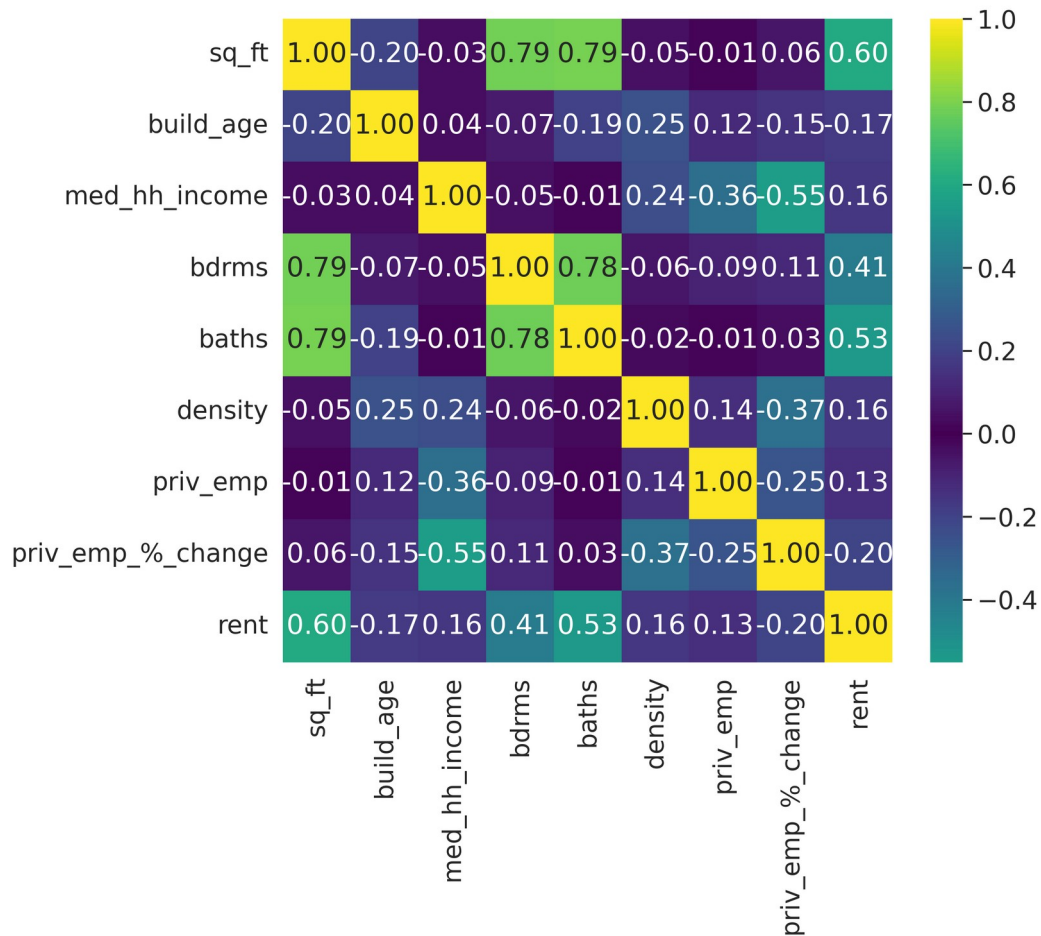| Monthly Rent	Bedrooms	Bathrooms	Square Feet
$2,643 - $14,754	Studio - 3 bd	1 - 3 ba	546 - 2,648 sq ft

Assumption: Lowest/highest rent corresponds to combination of all lowest/highest values
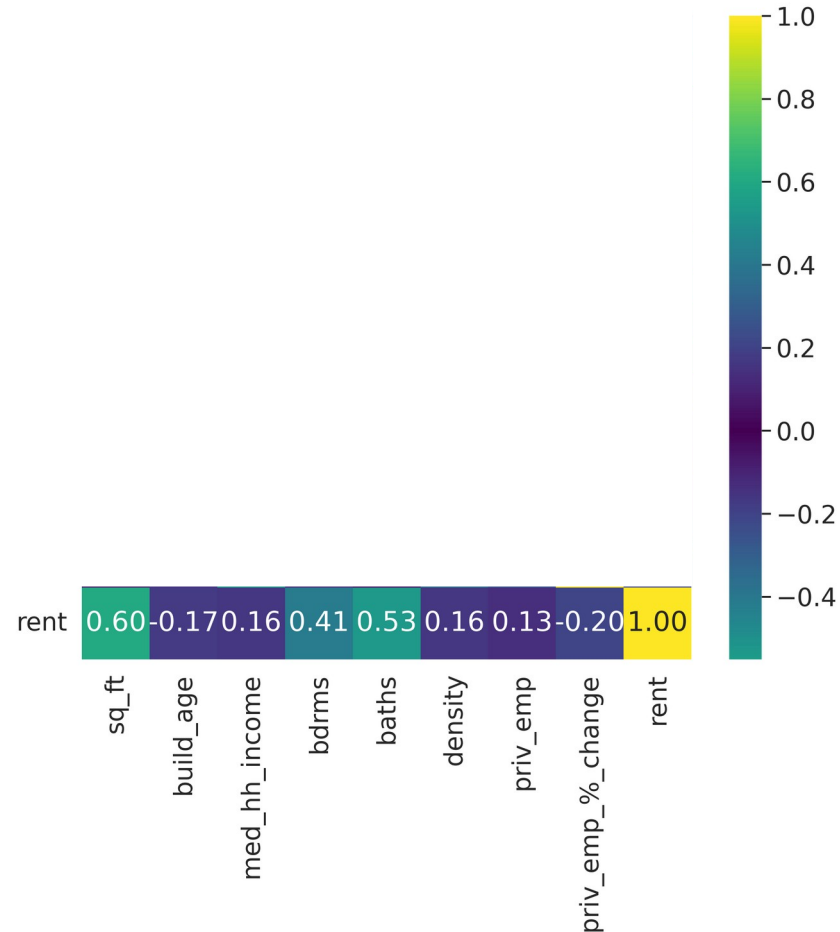
Consequences:
- Apt complexes with more than 2 listings are under-represented
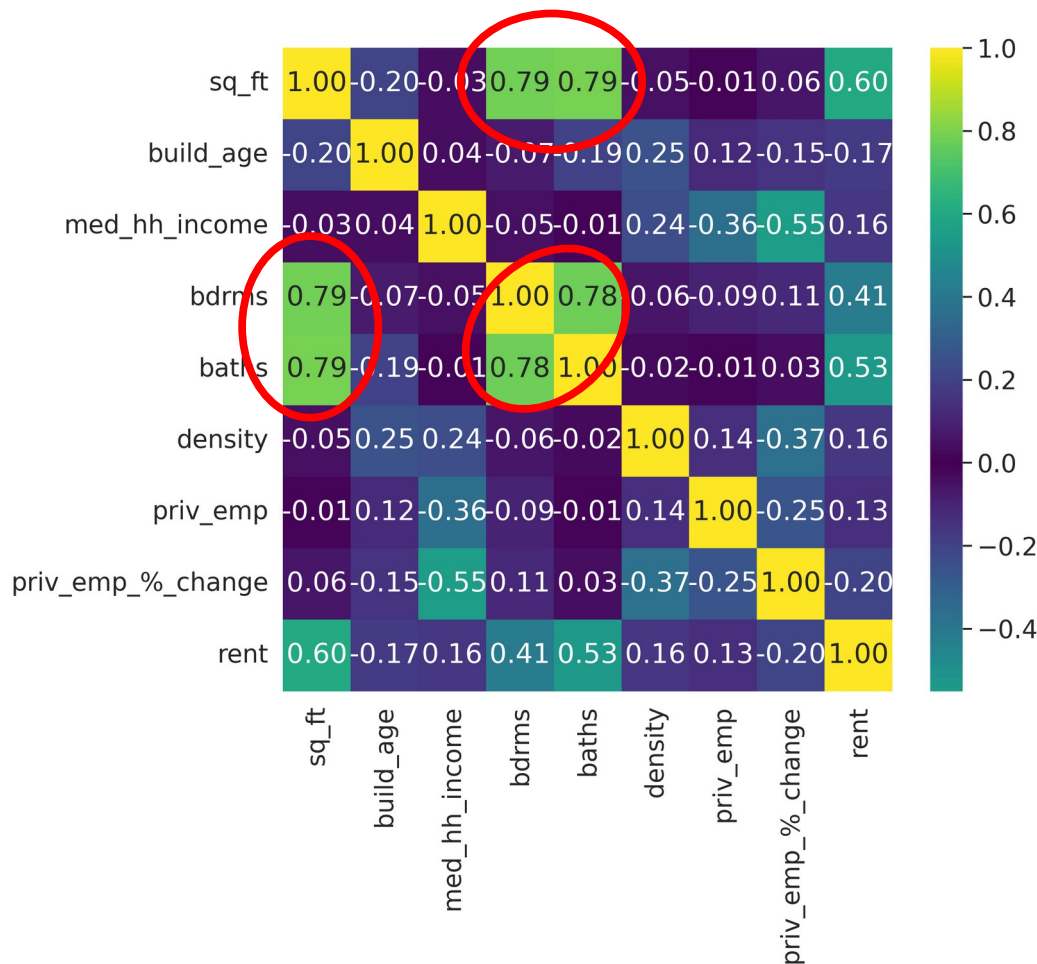- Middling apt's are under-represented

EDA

EDA

- Sq. ft. and # of baths have highest correlation

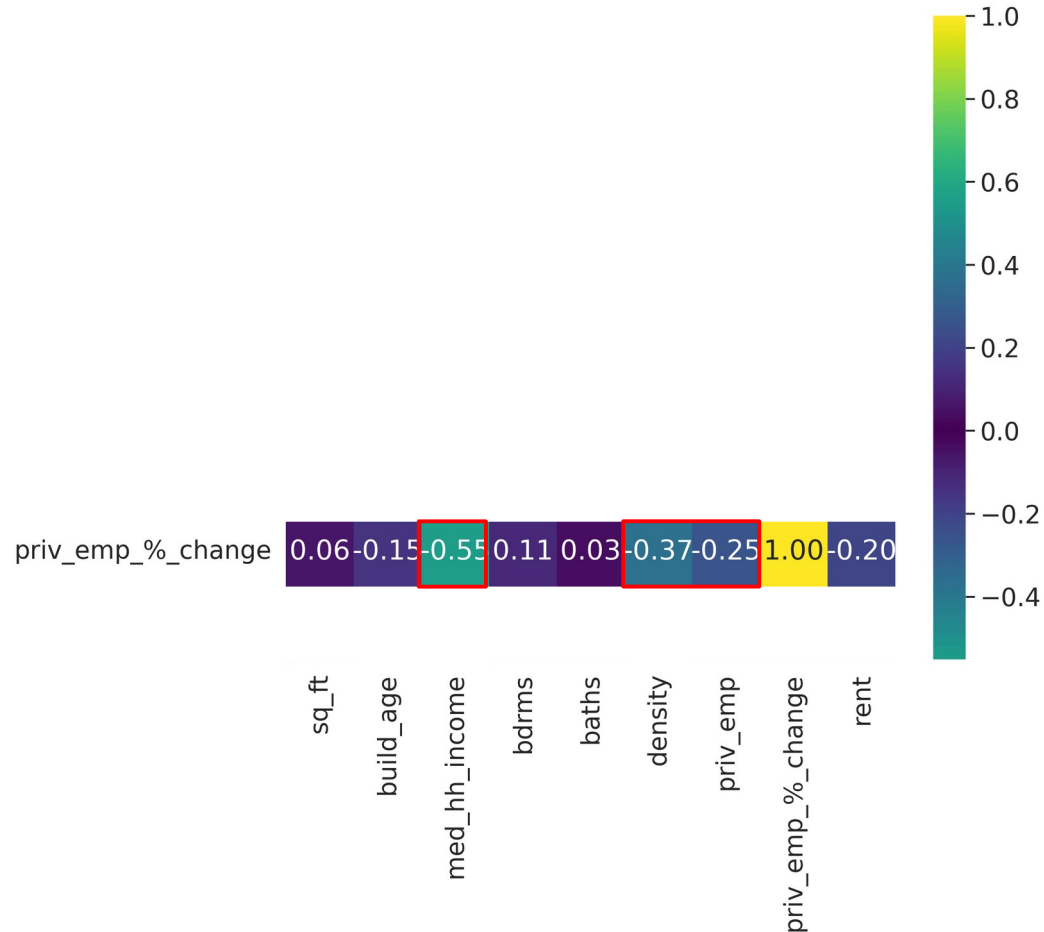- % change in employment has negative correlation with rent

EDA

- Sq. ft., # baths, and # bdrms all highly correlated

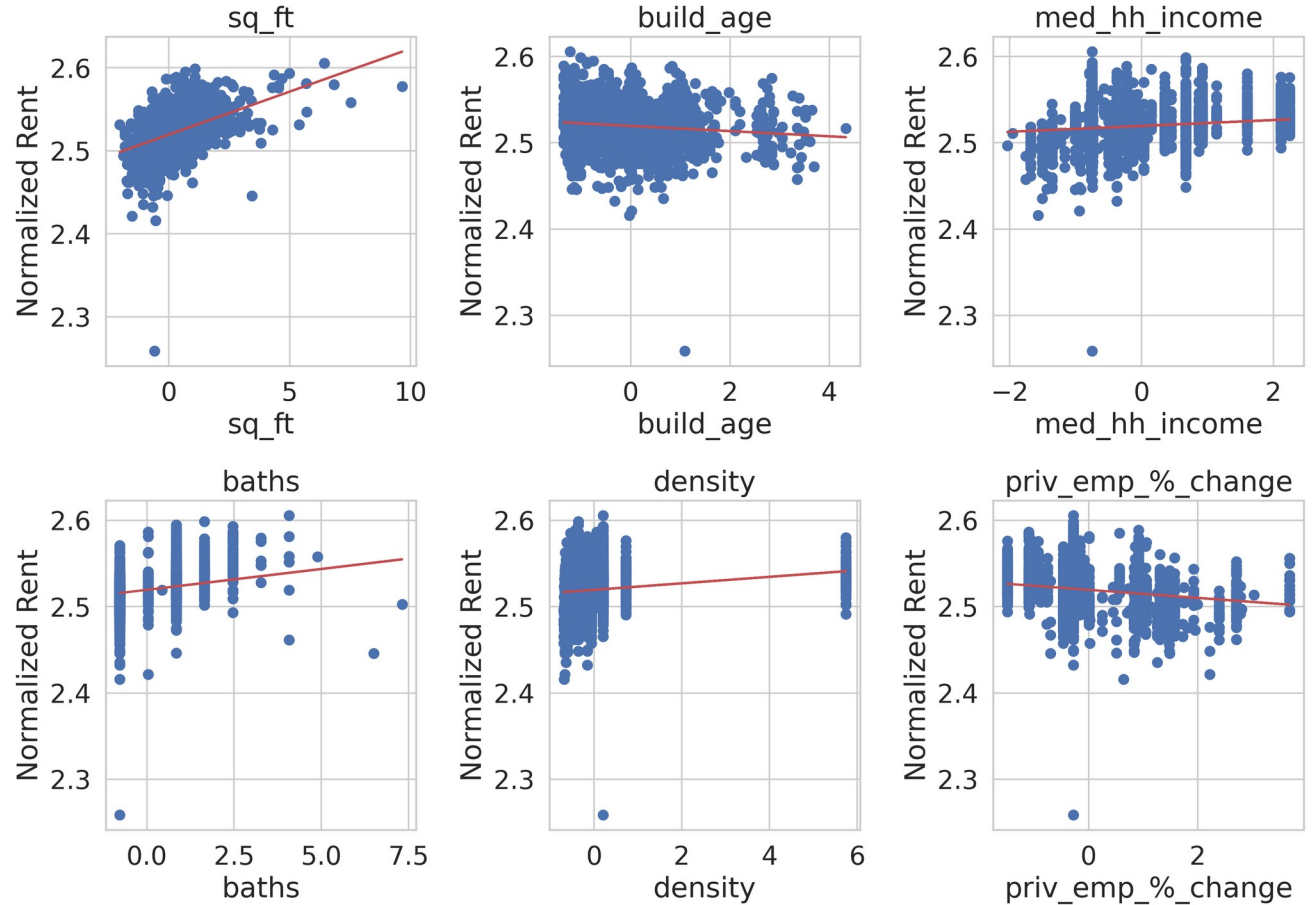- Demographic features are all correlated with each other

EDA

- Sq. ft., # baths, and # bdrms all highly correlated

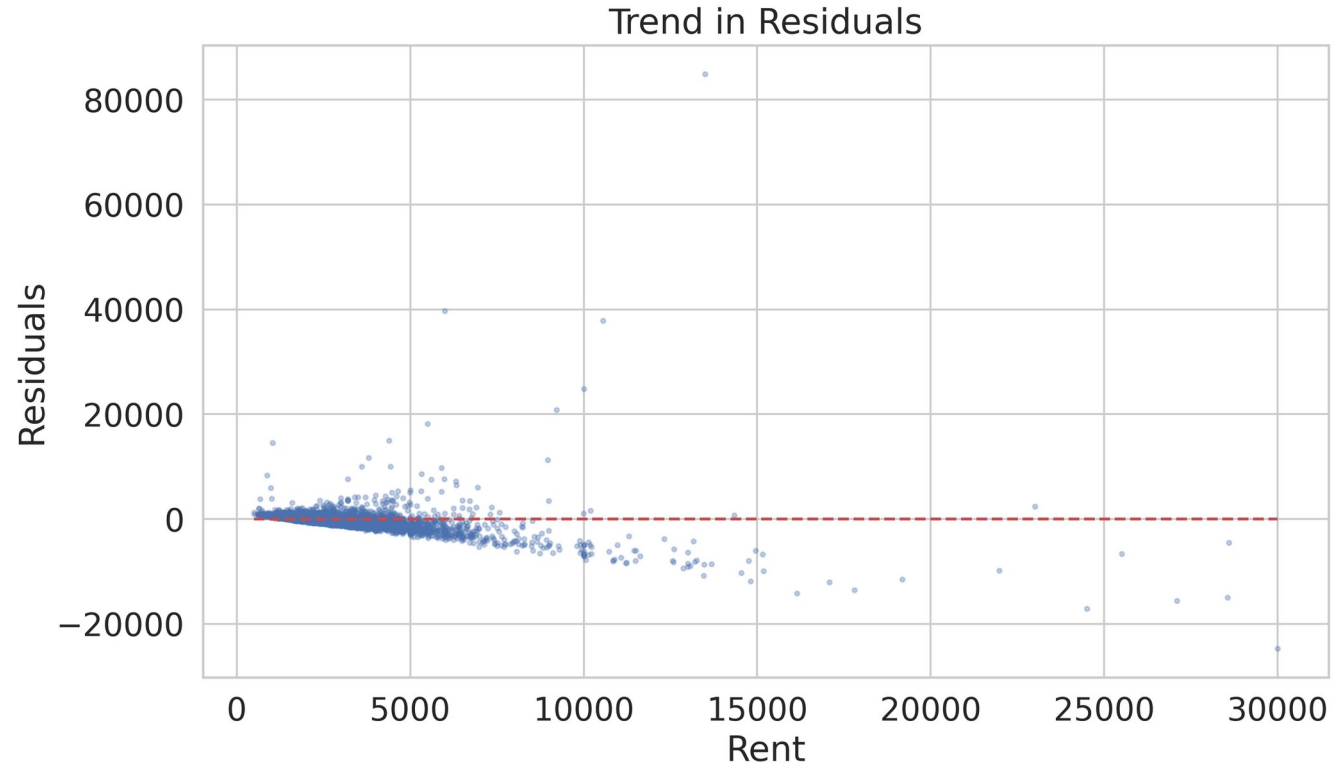- Demographic features are all correlated with each other

Results

- Model: LASSO

- MAE: $772/mo

- Median Rent: $2,295/mo

- R^2=0.44

Results

- Room for improvement

- Model underestimates expensive apartments
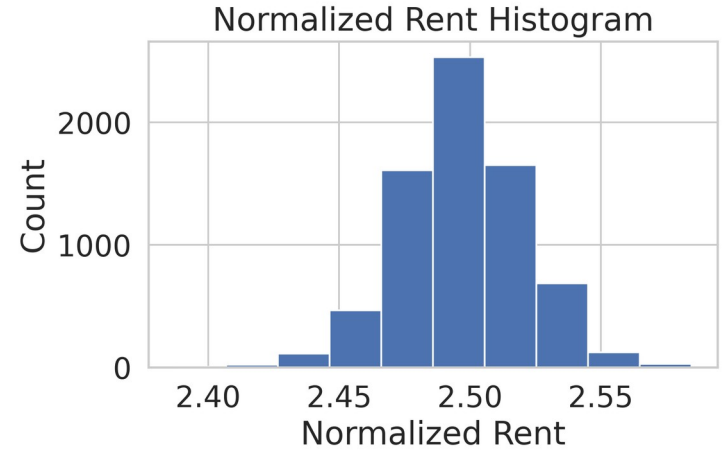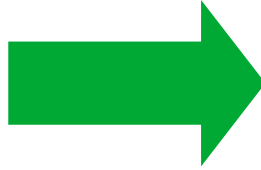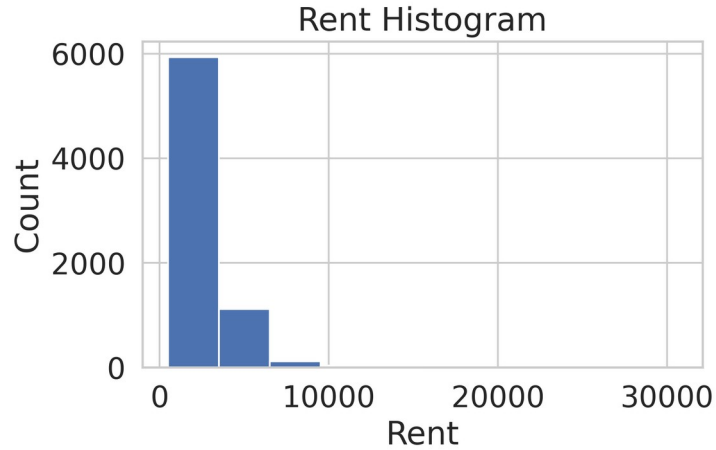

Trend in Residuals

Possible Improvements

- Add additional features (such as distance to ocean/city center)

- Try nonlinear regressors like a random forest

- Scrape more data (especially from outside LA county)

Questions?

Appendix

$$y_{box-cox} = \frac{y^{\lambda} - 1}{\lambda}$$

Appendix

Rent vs features scatterplots