Bruin Apartment Finder

Matthew Chuang, Aryan Sunkersett, Darsh Vema, Stella Koh, Skylar Nguyen, Sumedha Goyal, Ria Kundu

01. Data Collection

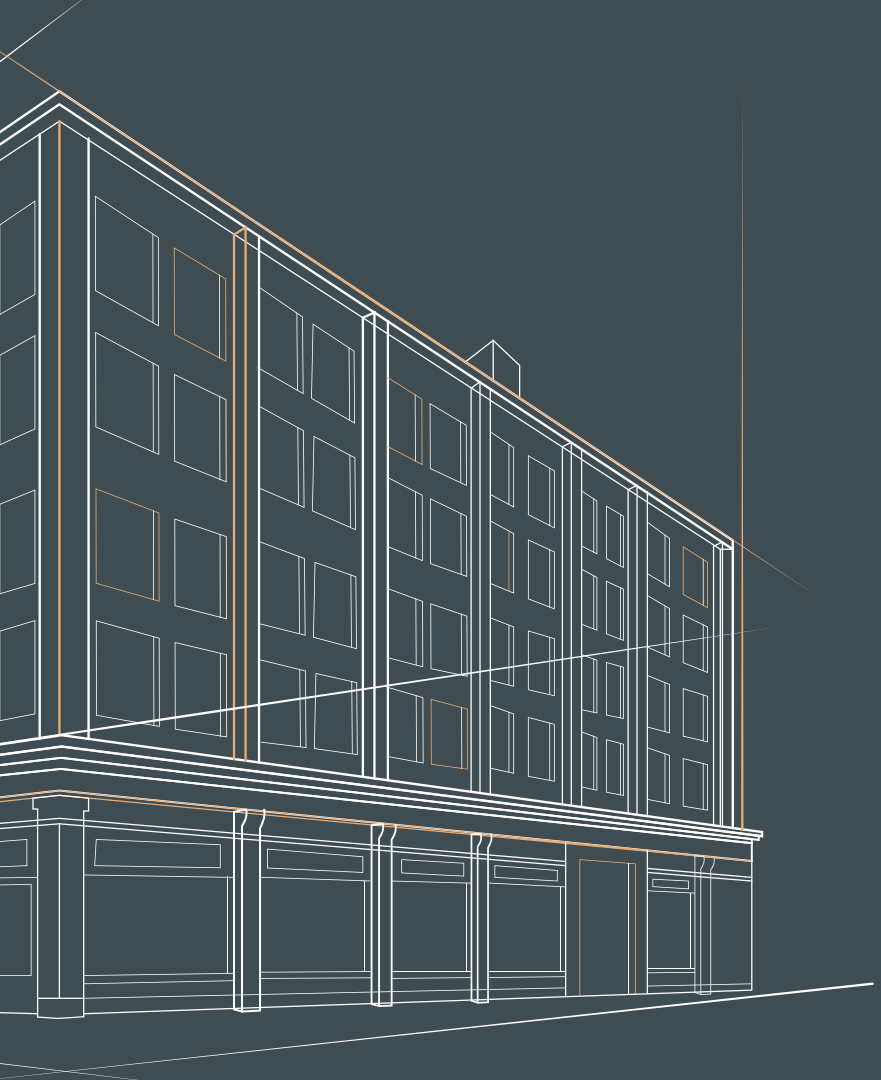
02. Data Cleaning

03. Data Visualization

04. Machine Learning

05. Conclusions

06. Future Ideas

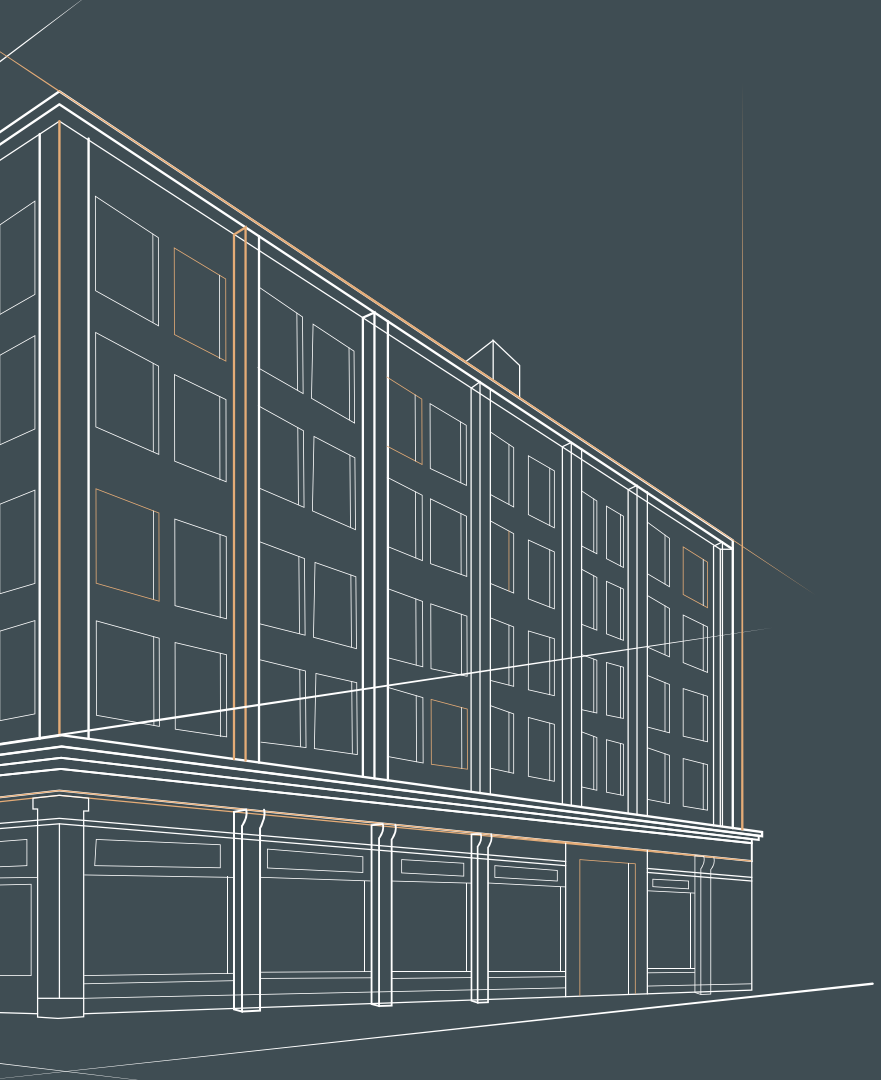


01

Data Collection

Collection Methods

- Obtaining data
  - Webscraped data from apartments.com with a Bright Data tool
- Dataset
  - Analyzed ~380 apartments in the Westwood area
  - Relevant data points include: monthly price, sq ft, location, distance, rating, bedrooms
  - Distance is calculated from the apartment to Bruin Bear
  - Other (irrelevant) data points: internet, ice, intercom, granite, type of floor, accessibility
- Goals
  - To what extent do apartment features contribute to the price?
  - To what extent does the distance to UCLA impact the price?

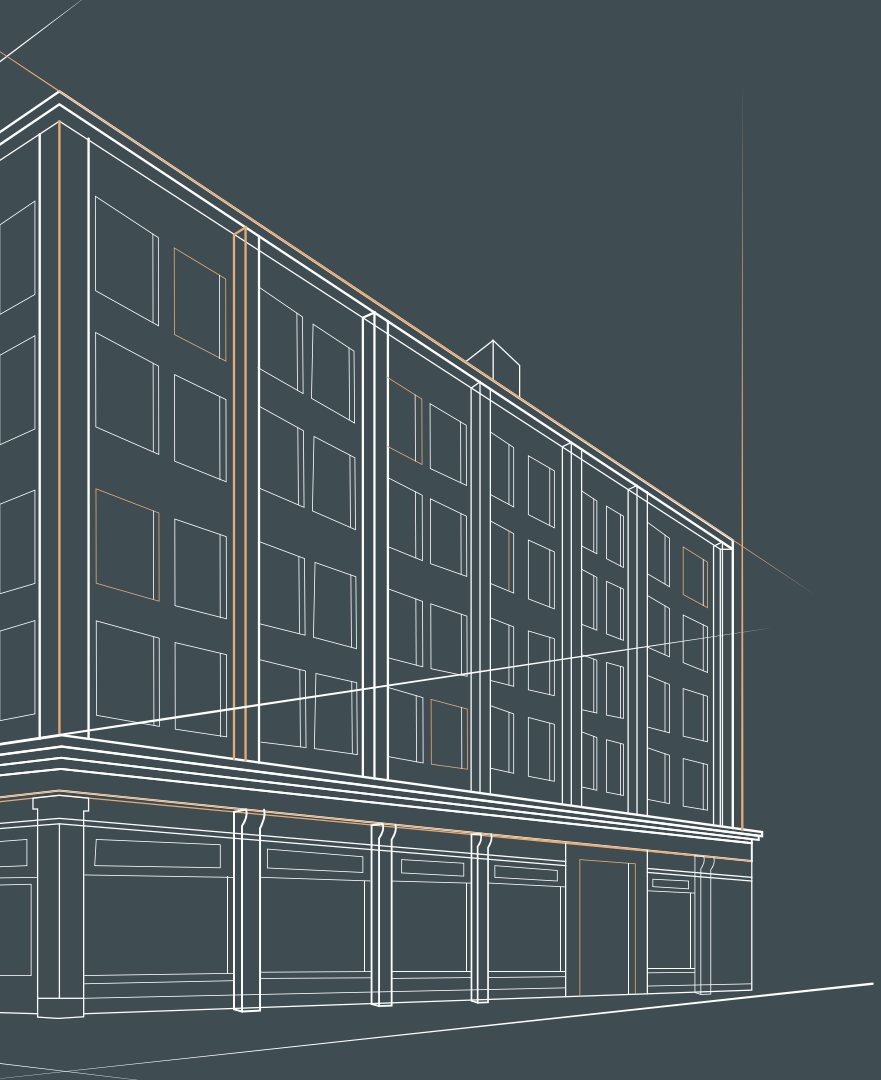


02

Data Cleaning

Data Cleaning

- Standardized prices and square feet
- Deleted observations with missing prices
- Identified key data from qualitative descriptions
- Geocoded locations

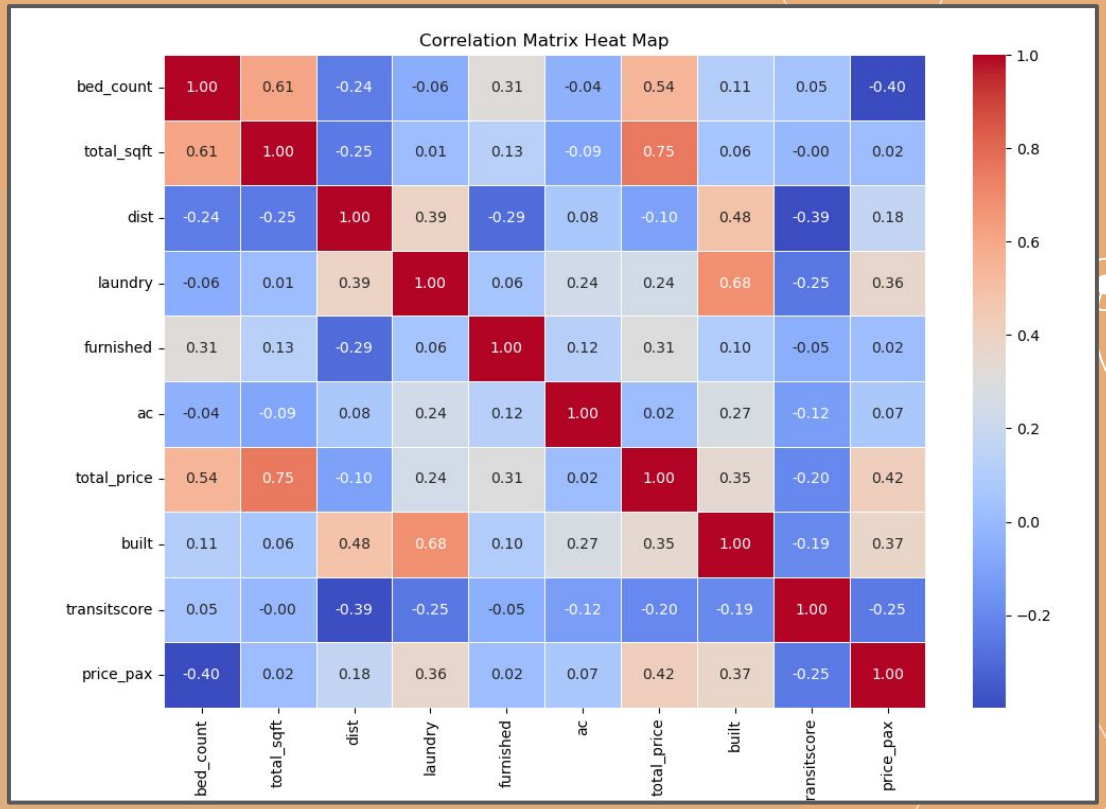


03

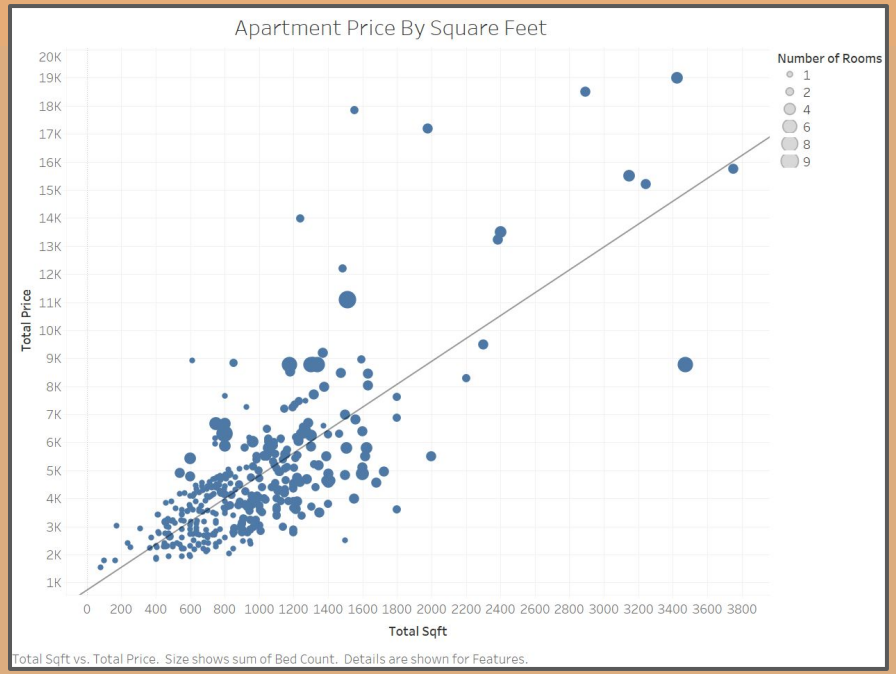
Data Visualization

Correlation Coefficients

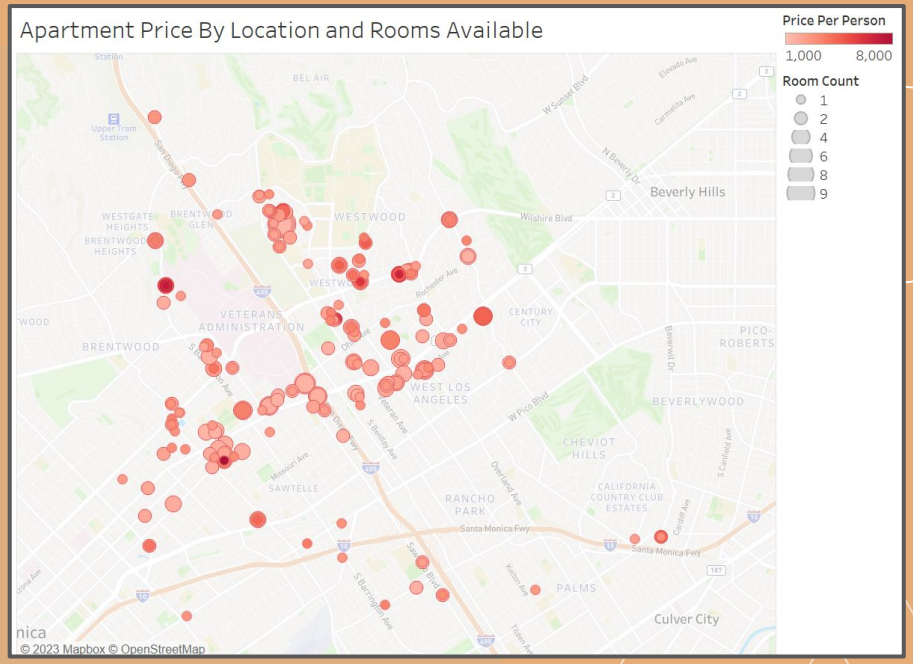
- Seaborn Correlation matrix heat map
- Help us see relationship between features (closer to one, stronger the relationship)
- In this case, orange-red squares are stronger



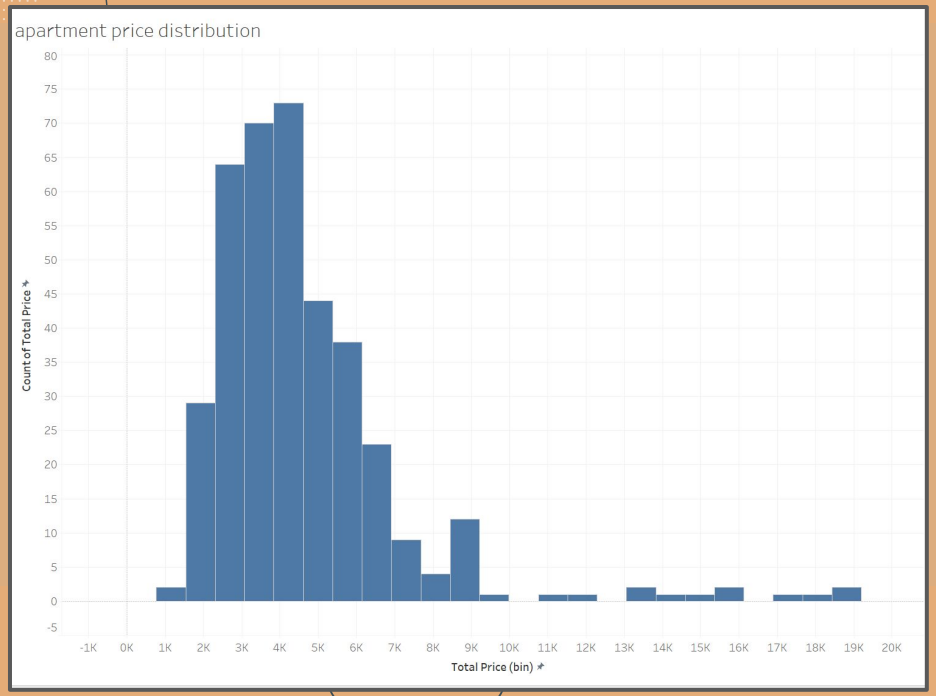
Square feet and total price relationship

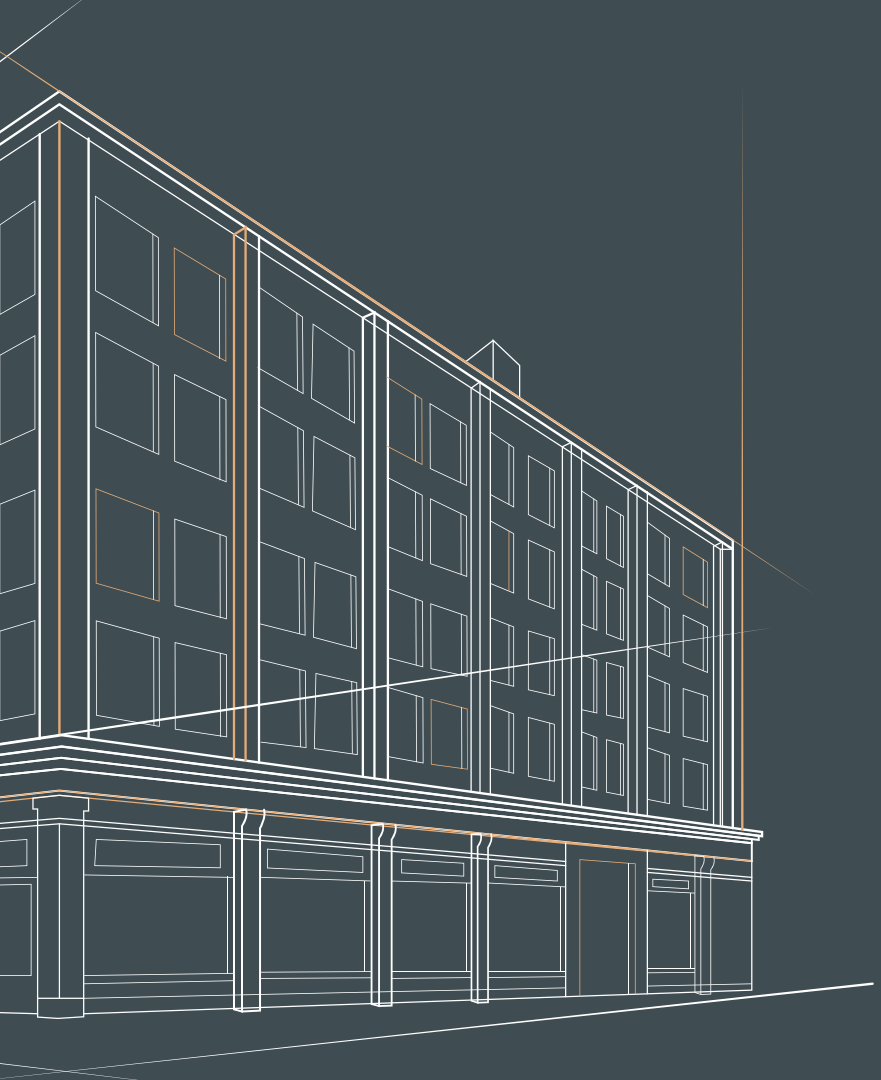


Geographical (price and #rooms) relationship



Price Distribution

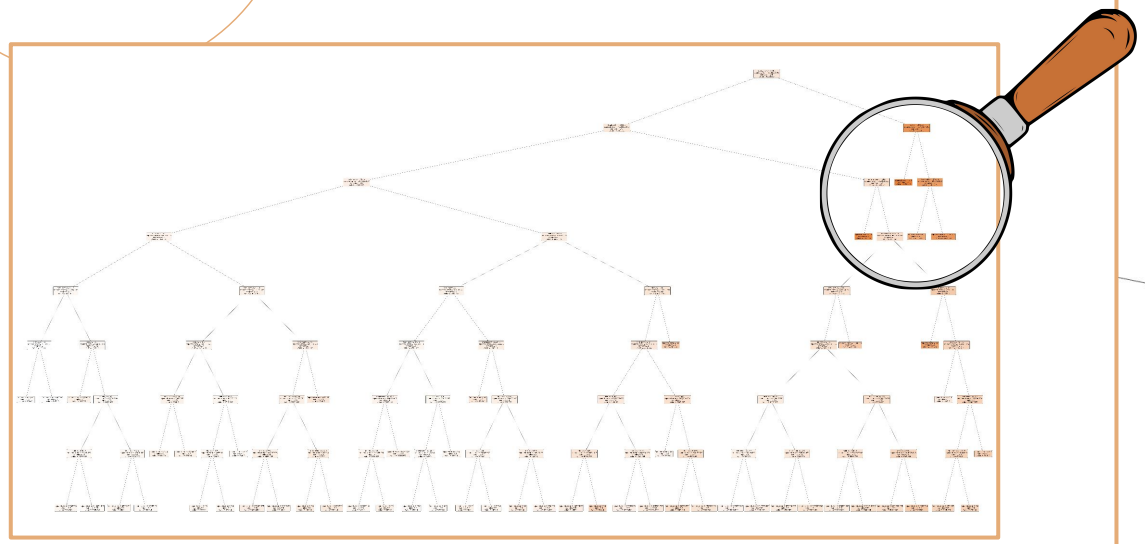


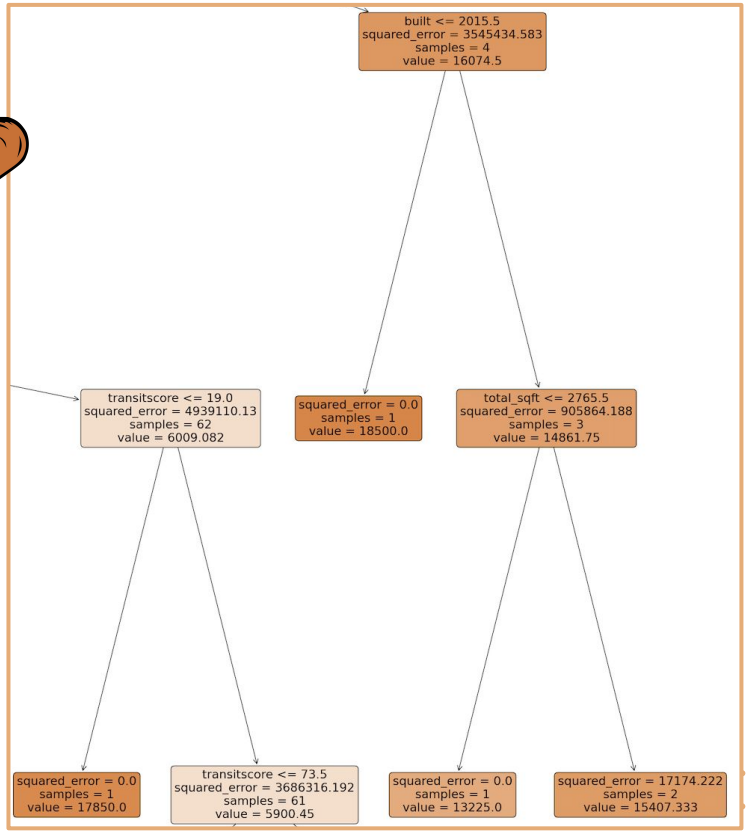


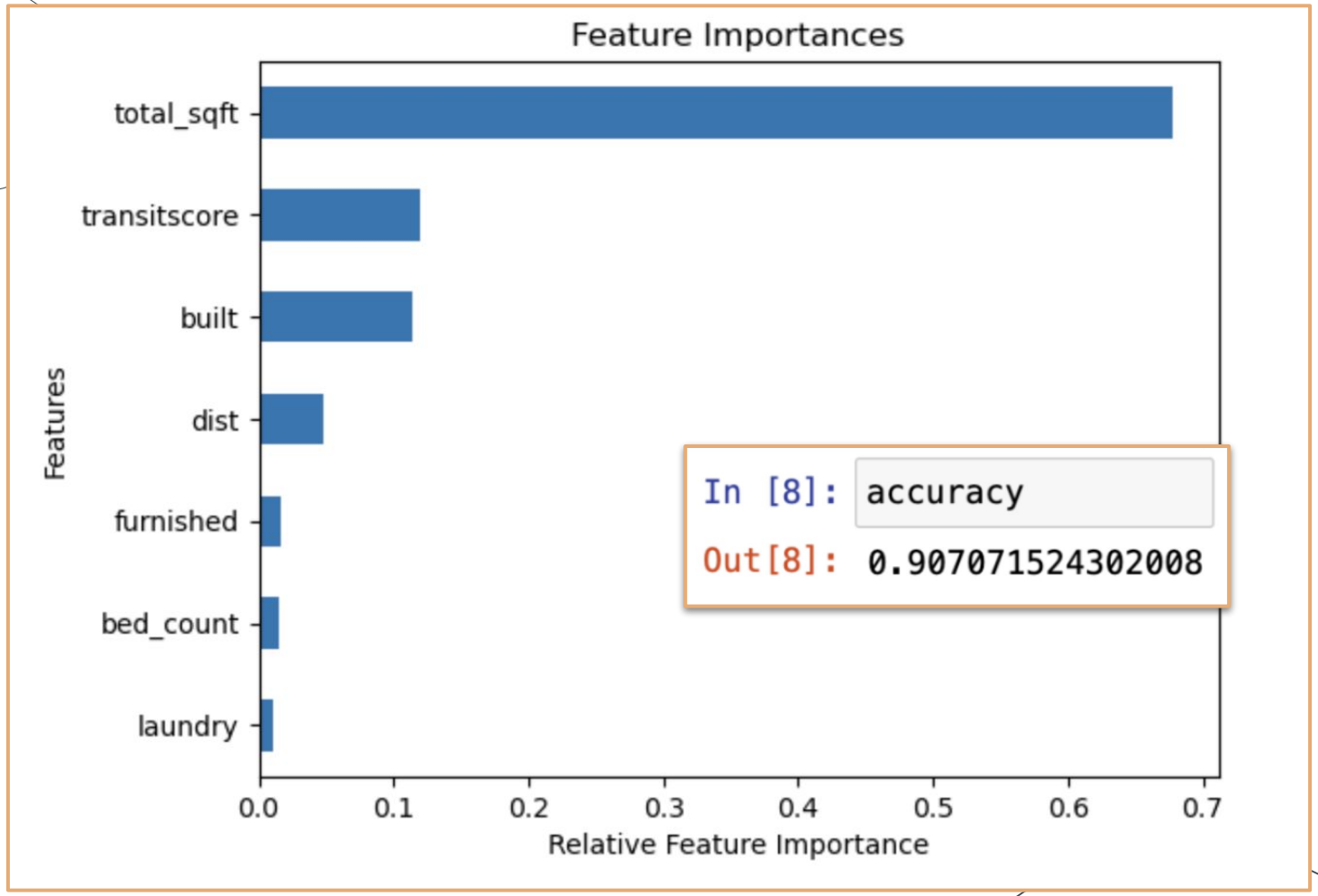
04

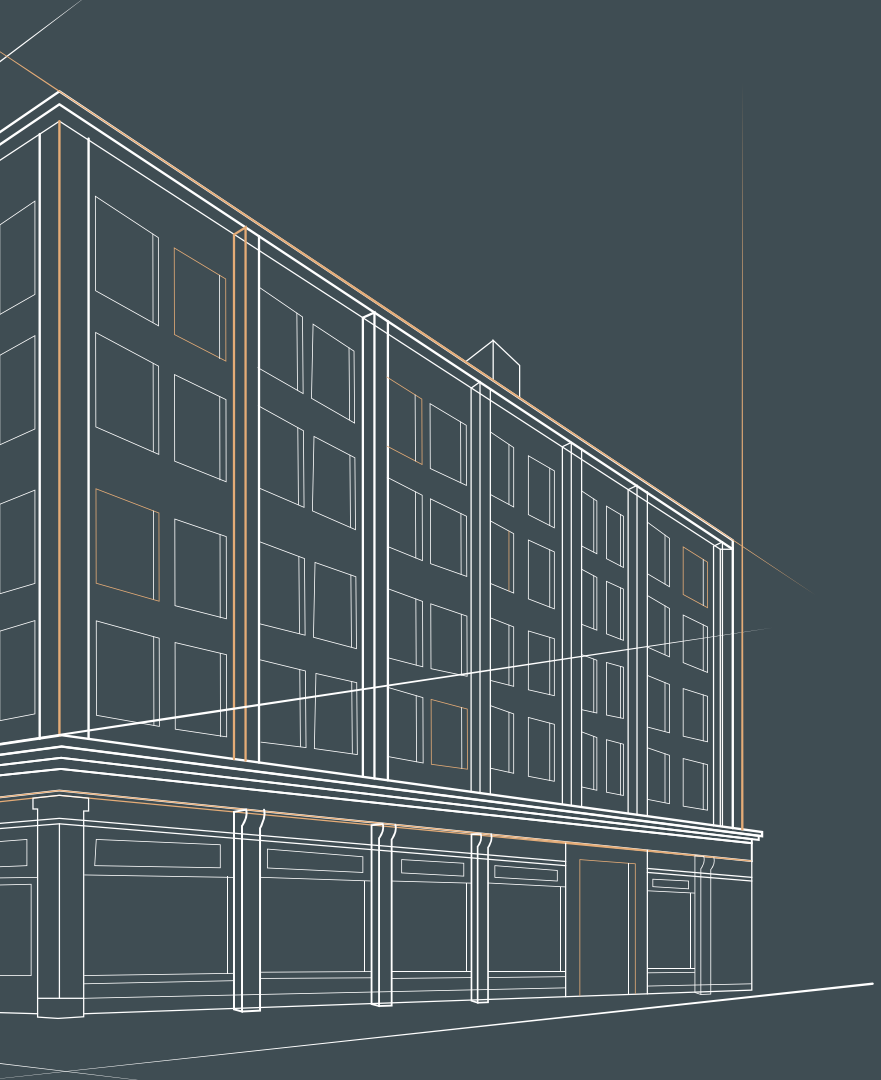
Machine Learning

Random Forest Regressor






Feature Importances
total_sqft
transitscore
built
dist
furnished
bed_count
laundry
Features
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Relative Feature Importance
In [8]: accuracy
Out[8]: 0.907071524302008



05

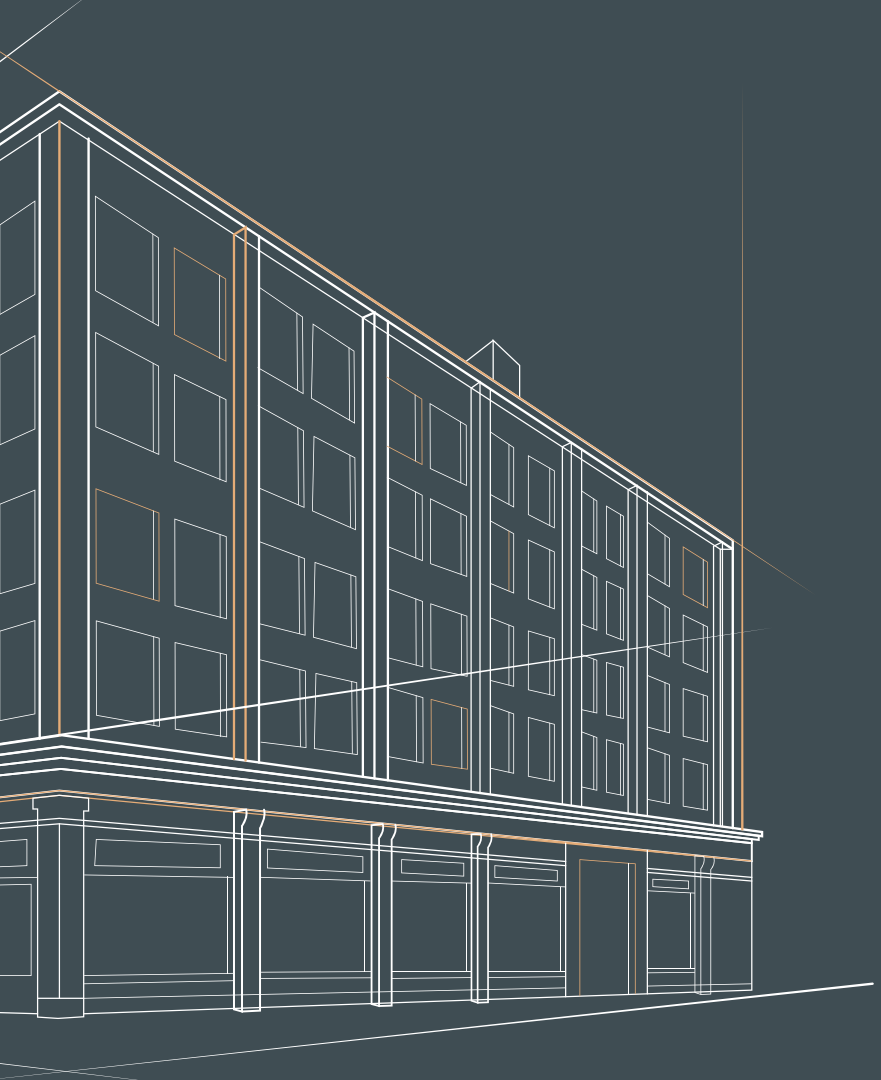
Conclusions

Conclusions

- Most impactful feature against total price was the the total square feet (as expected) This also matches what we noticed from our visualizations

- Surprisingly, the built date and transit score accounted for a good chunk of feature importance while the distance ended up being less than we thought.

- Keep in mind this observation used price per person, thus these features matter the most to individuals. Looking for the best apartments for multiple people would require more metrics and retraining of the model to accommodate for different sizes



06

Future Improvements

Future Improvements

- Failed NLP, Lack of sentiment words in Reviews
- Had idea of looking for “eye-catching” words in listing Titles
  - Didn’t have enough consistency or quantity
- Scrape more data in a smaller range around westwood/UCLA
  - A lot of the listings were too far and “unrealistic” for students (upwords to 9k for rent, 4+ miles away)