Decision Tree Model to Predict Rental Days Vacant

The Goal: Predict if a rental will take more than 15 days to rent

1. Data used includes daily advertised rents and property specs, more on this later

2. Scraped between 07-06-2018 and 12-09-2019

3. The final dataset: approximately 18,000 unique rentals.

Why predict rental days vacant?

To inform best practices for professional property management firms to reduce vacancy times

To know which property attributes and prices lead to problem properties with long vacancies

Limited features to work with...

Number of rent changes

Total rent change amount

Initial rental price

Acres,

Square footage

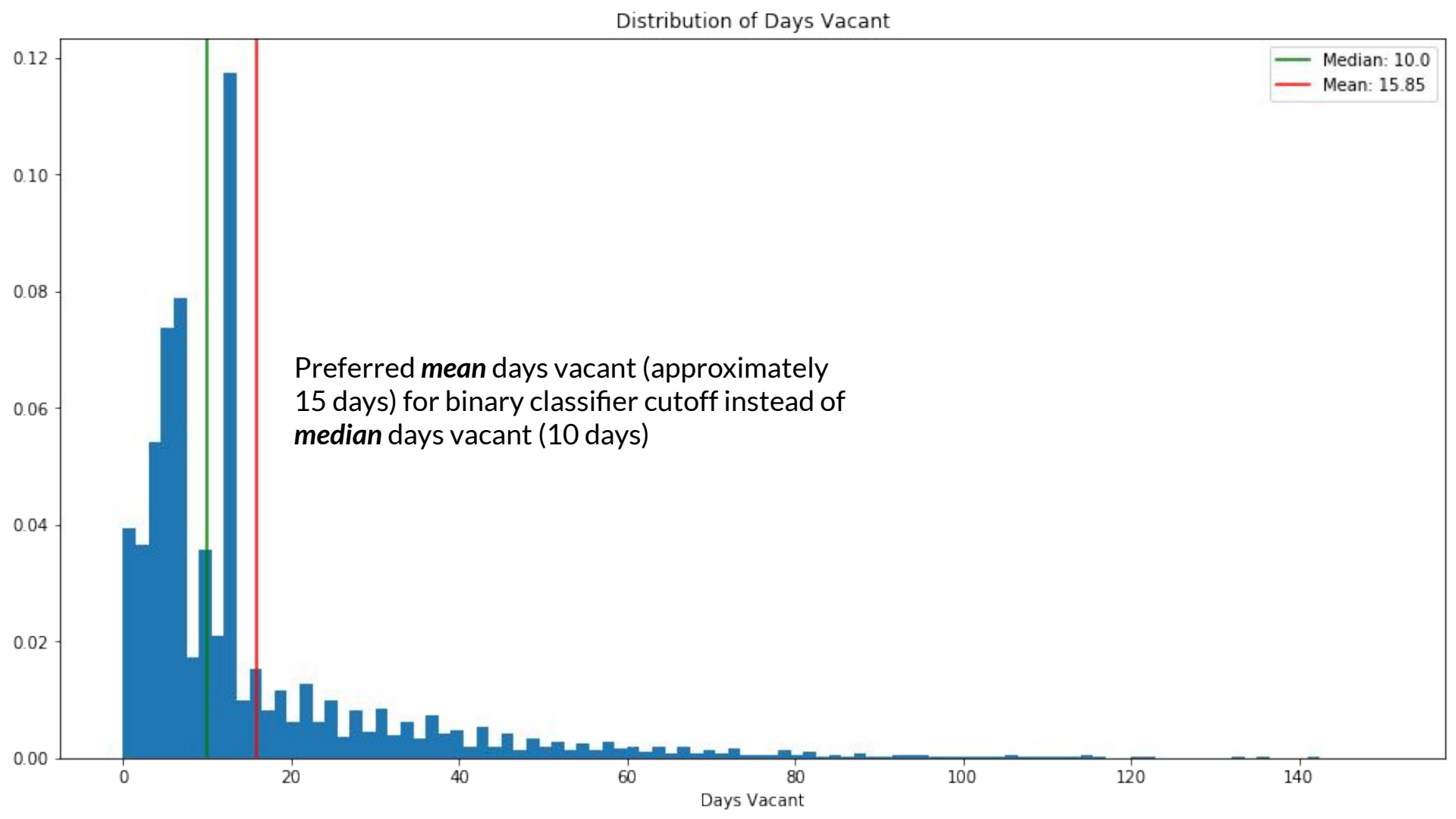
Bedroom count

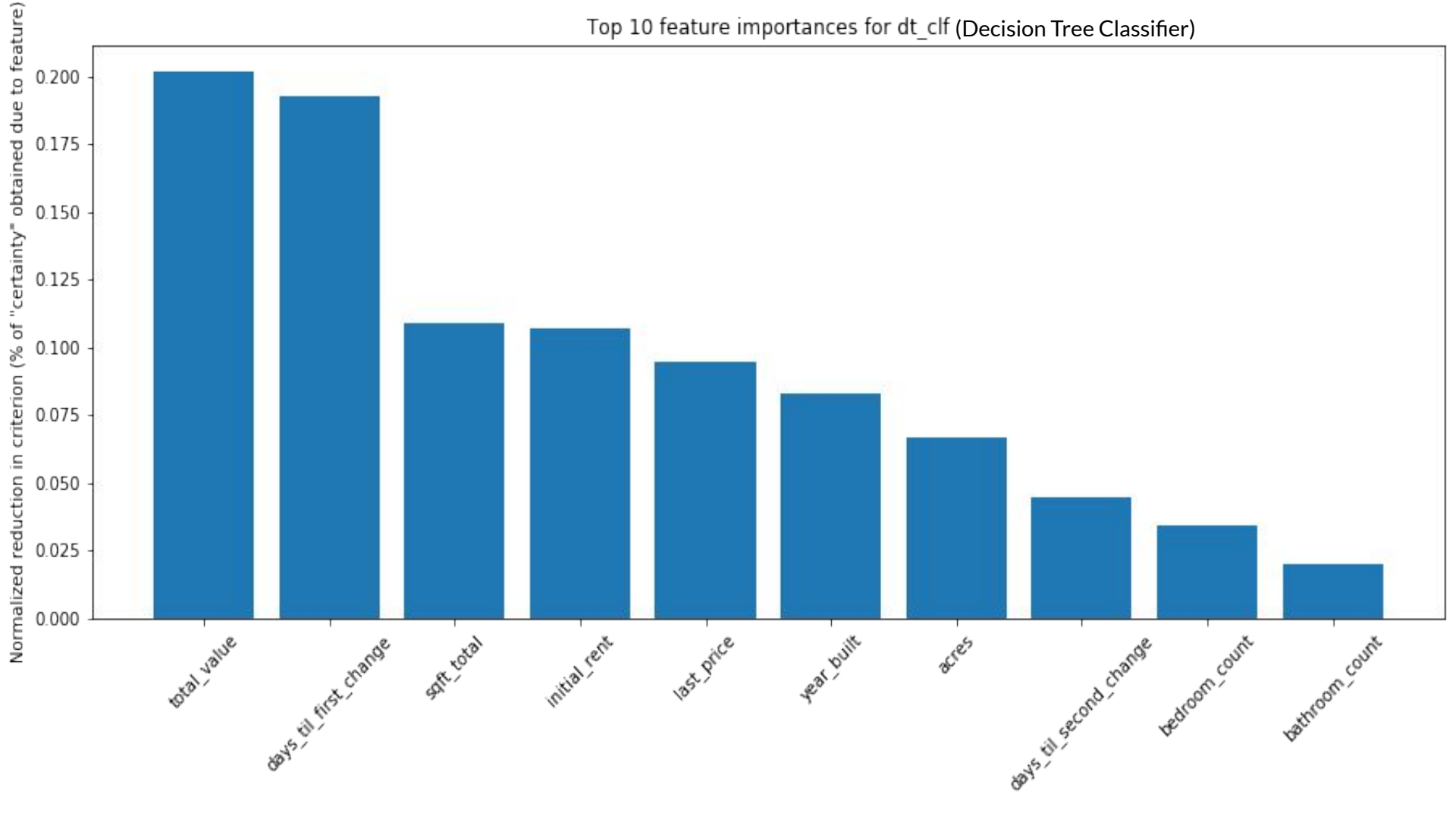
Bathroom count

Year built

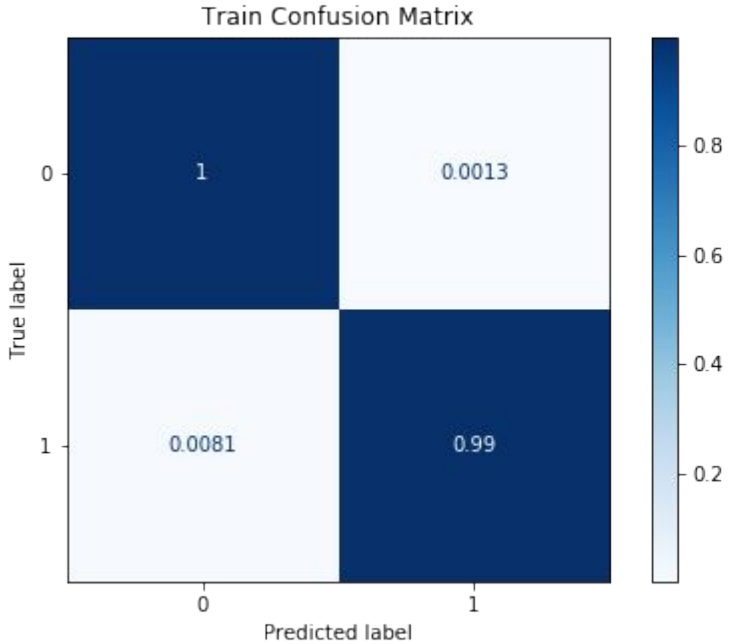
Last price

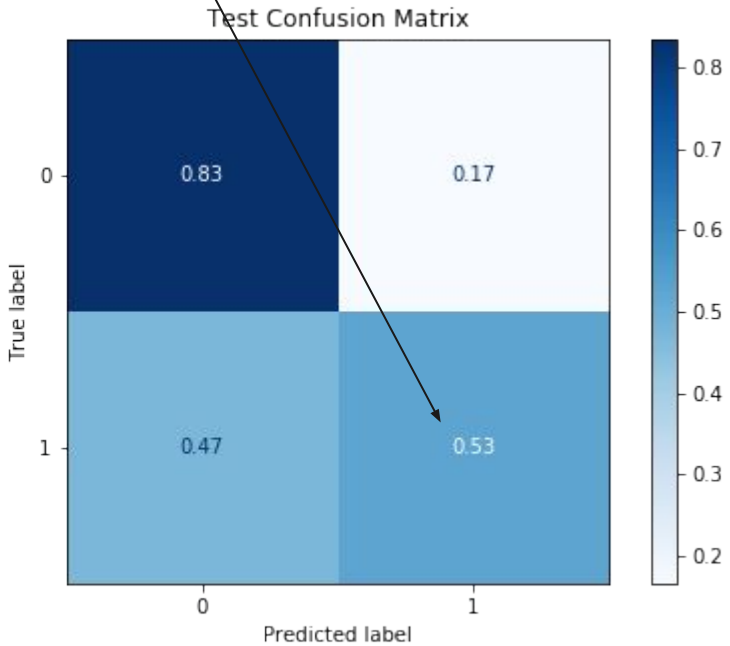
Total value


Distribution of Days Vacant
Median: 10.0
Mean: 15.85
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
120
140
Days Vacant
Preferred ***mean*** days vacant (approximately 15 days) for binary classifier cutoff instead of ***median*** days vacant (10 days)


Top 10 feature importances for dt_clf (Decision Tree Classifier)
Normalized reduction in criterion (% of "certainty" obtained due to feature)
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
total_value
days_til_first_change
sqft_total
initial_rent
last_price
year_built
acres
days_til_second_change
bedroom_count
bathroom_count

Surprisingly, the best performing model (with a **high recall** being the most important target metric) was an overfit decision tree classifier. Recall was .53 which mean that our model caught truly **over 15 day** rentals only half the time. This isn't ideal.





Future work: How to improve model?

Multiple methods were tested including XGBoost, Random Forest, and Decision Tree. Similar results for test recall each time (about .50)

Gridsearch tuning of hyperparameters didn't substantially increase models ability to catch true positives

More data with more features may be necessary to separate **over 15 day vacant** rentals

THANK YOU!

Jupyter Notebook at:
https://github.com/benratkin/module_3_project