

MACHINE LEARNING FOR RENTALS

Predicting rental prices in Auckland

LUCAS HANSON
CAPSTONE PROJECT

AGENDA

Project outline

Data overview and analysis

Feature engineering

Modeling

Model Prototype

Final thoughts



HOW MUCH SHOULD I RENT MY PROPERTY FOR ?

This is Jenny, our real estate agent

She wants to avoid arguing with landlords about how much to rent thier properties for...



Our solution is to create an ML tool to produce a rental price guideline based on past rental data.

RENTAL DATA OVERVIEW

Data sources:

Ministry of Social Development - OIA requests

Realestate.co.nz - Direct data requests

Timeline:

MSD : 2018-2022

Realestate.co.nz: Jan 2021- June 2023

Limitations of data:

No precise location of individual properties

No size of the property

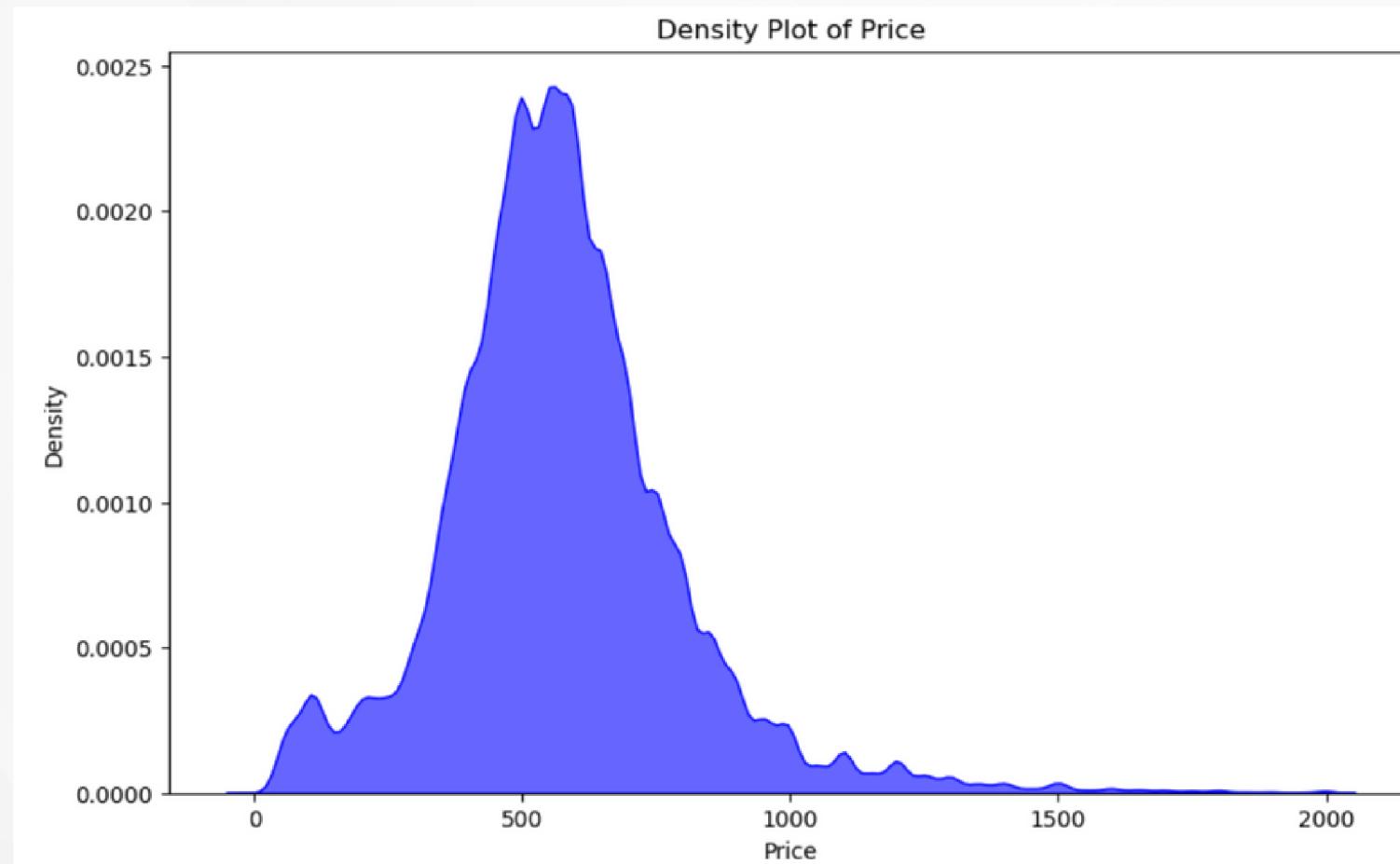
No property description data

Ensuite's feature is somewhat unreliable

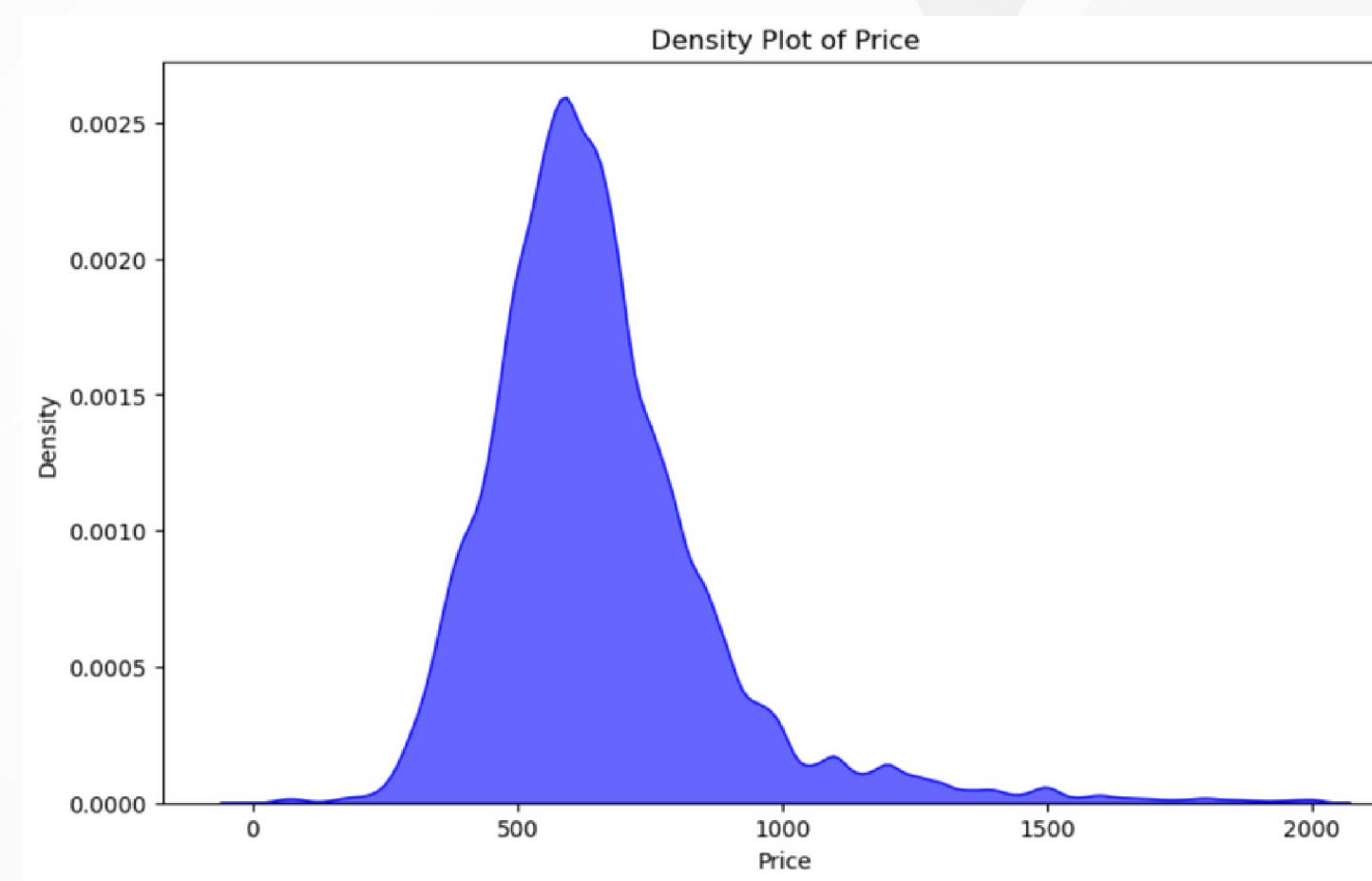
MSD DATA	REALESTATE.CO.NZ
Publish Date	Publish Date
Suburb	Suburb
No. of Bedrooms	No. of Bedrooms
	Listing Type
	District
	No. of Full Bathrooms
	No. of Ensuites
	Listing title
Price	Price

PRICE DISTRIBUTION

MSD data

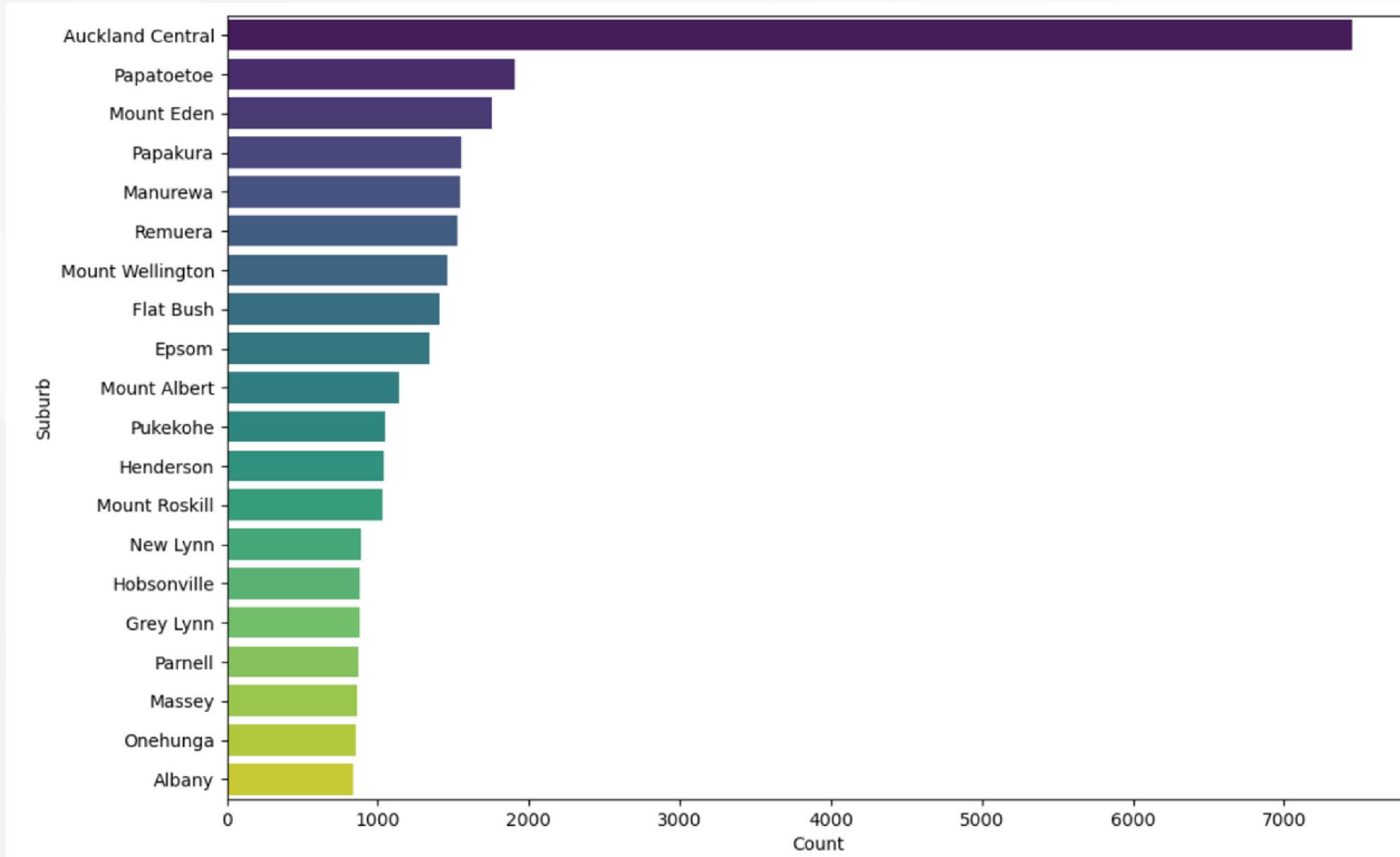


Realestate.co.nz data

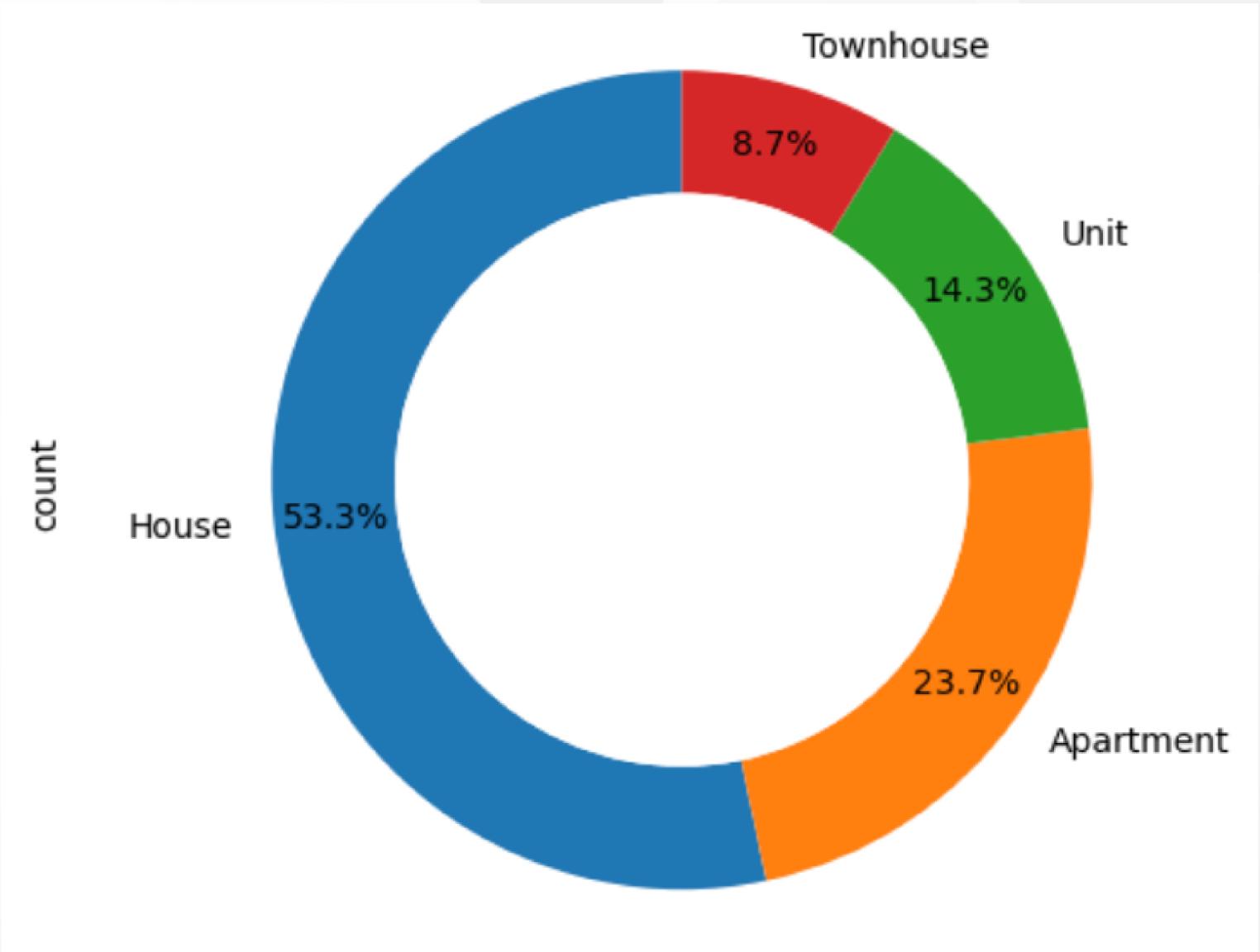


LISTING TYPE AND SUBURB

Top 20 Suburbs

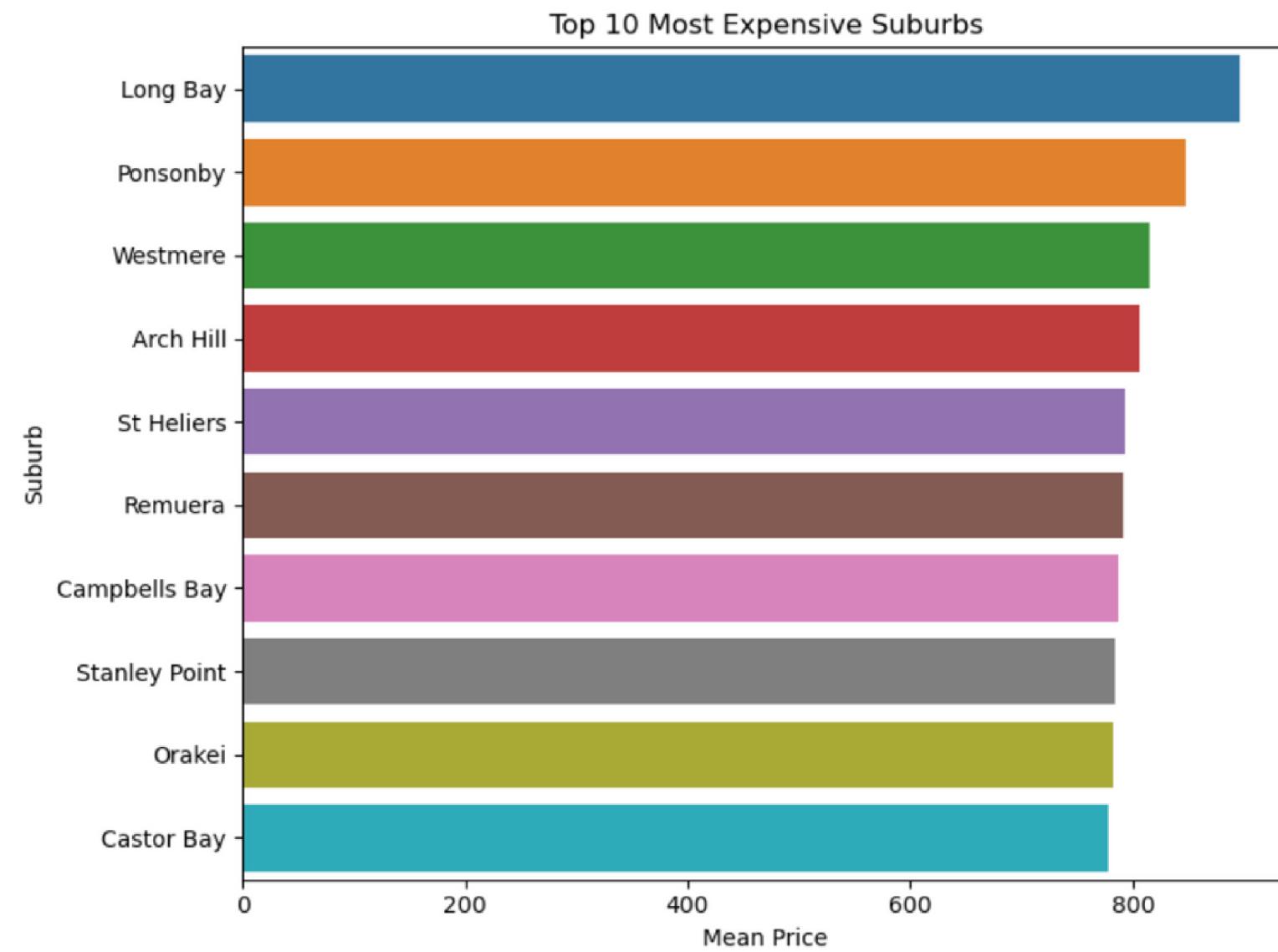


Listing Type



FEATURE ENGINEERING

- Encoding Listing Types and district
- Using MSD data to create an 'Expensive Suburb' feature merged by Suburb
- Creating lag features to capture signal from suburbs
- Creating lag features for Apartments, Bathrooms and No. of rooms to take advantage of our time element of our data.



LAG FEATURES



- Our lag features take the previous 3 months of listings with similar features excluding the listing itself and any recorded on the same day and create an average price of those listings.
- Lag features that carried the best signal were lag by bedrooms , lag by bathrooms, lag by suburbs, lag by apartment and lag by bedrooms by bathrooms

TEXT FEATURE



“Modern family home in Unsworth Heights”

- We created a multinomial naive bayes model to predict listings in the top 10th percentile for 'No. of bedrooms' values using our count vectorised text from the listing title feature
- Extracting the priors of this model to create a list of words associated with premium properties

**'home', 'furnished', 'new', 'views', 'family', 'fully', 'brand',
'luxury'**

- This list of words became a 'Premium Properties' feature in our regression model that checks if any of these words are in the title. This helps provide more signal for these more expensive listings

MODELING

Filtering the data

1-7 Bedrooms

1-4 Full Bathrooms

1-4 Ensuites

AUCKLAND REGION

'North Shore City' 'Auckland City'

'Waitakere City' 'Rodney'

'Manukau City' 'Franklin' 'Papakura'

'Waiheke Island'

Remove Carpark,

"Flatmate wanted" and
Holiday section listings

Approach to modeling

Regression model to predict
price of the property



Model should output a
sensible range for price
based on a confidence interval



Create a user interface where
agents can enter features of the property
and receive a price recommendation



Model should be tested on the most recent
data available to ensure robustness

PROPOSED MODEL

Baseline model

**Linear regression
RMSE: 105.1**

Tuned LGB model

**N-estimators = 3000
Learning rate = 0.01
Subsample = 56000
Max depth = 10**

RMSE: 100.9

Grid searched using Time series split, fixed window with no overlapping

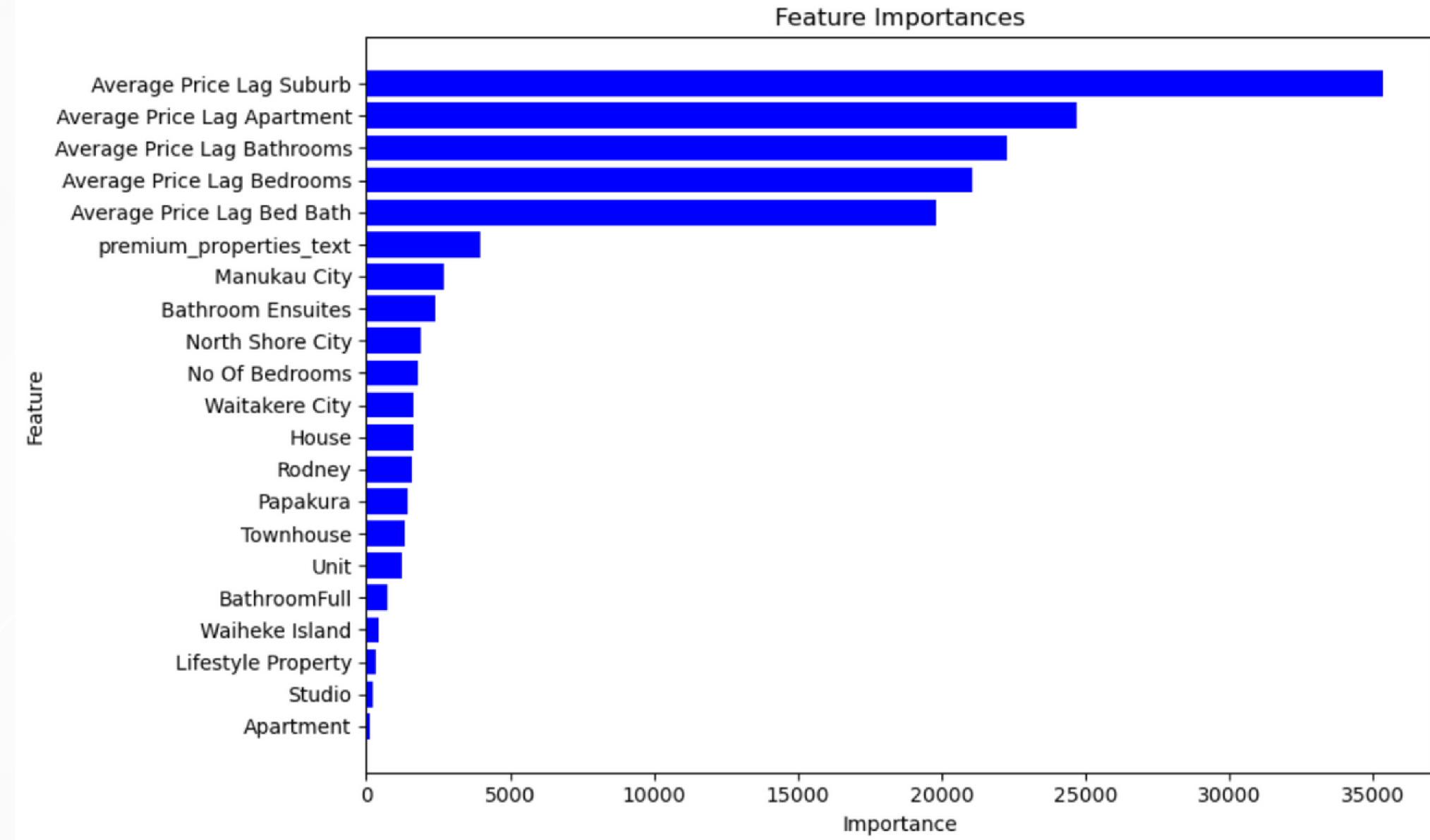
3000 trees built.

85% subsample of the data which introduces randomness and helps with overfitting.

stopping the depth of the trees at 10 also helps not to overfit the model.

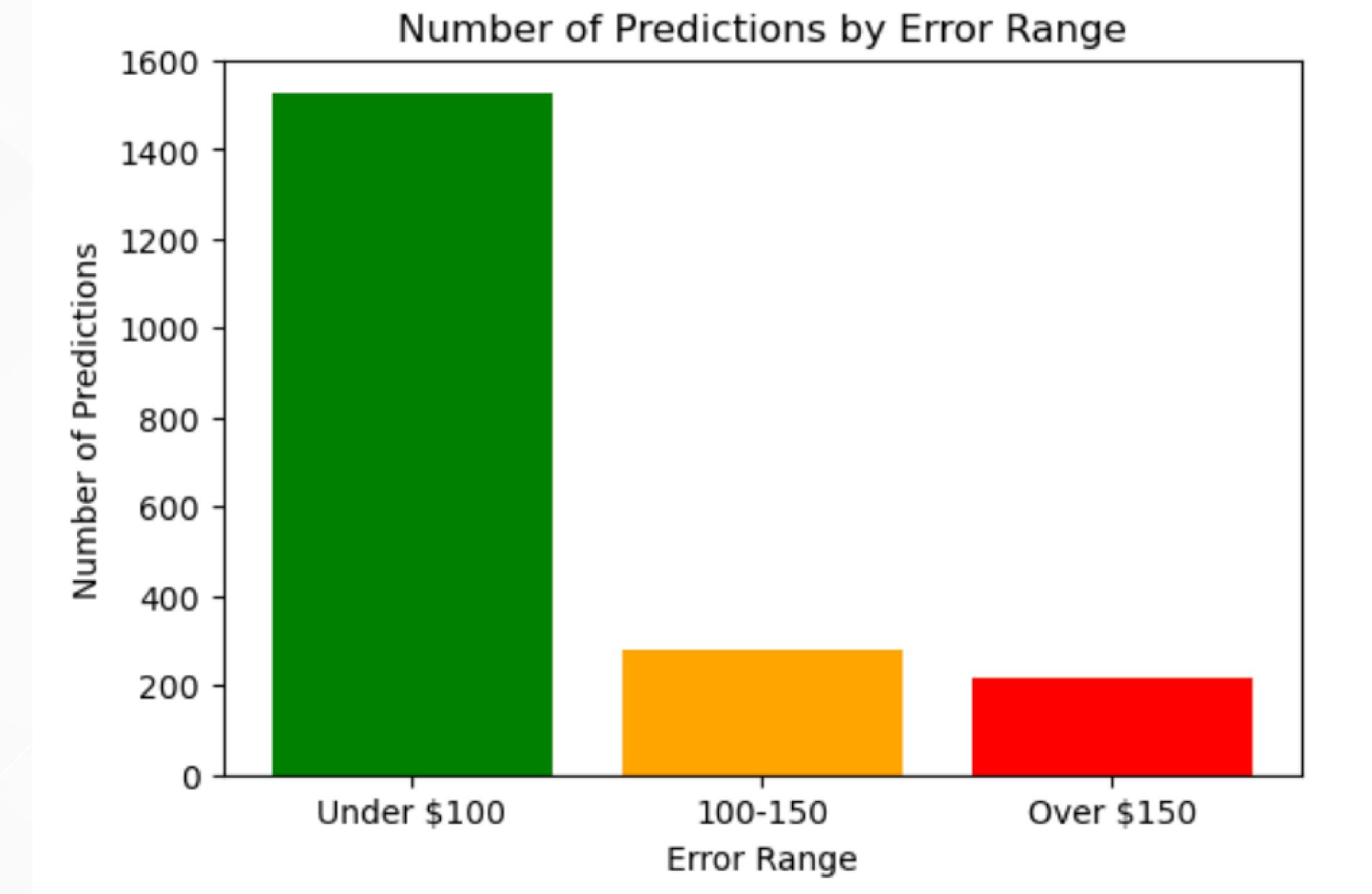
MODEL EVALUATION

This graph shows "split gain" feature importances which is a measure of how much a feature contributes to the reduction of the loss function. Our lag features have the highest signal for the model. This is expected as signal from most of our other features have been encoded into these lag features.



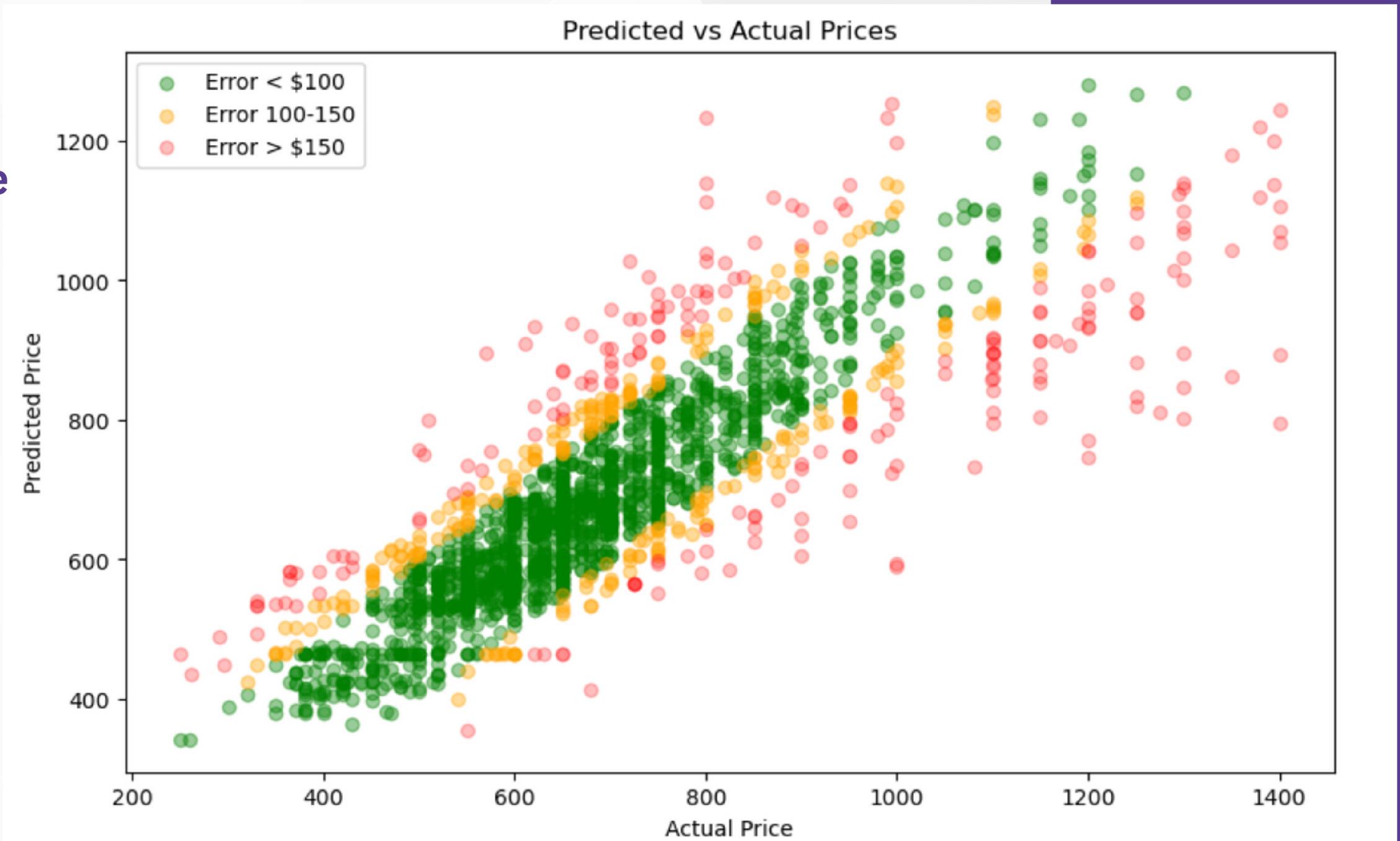
MODEL EVALUATION

- 75% of the models predictions are under \$100 of error
- 15% have an error between \$100 and \$150
- 10% have an error over \$150



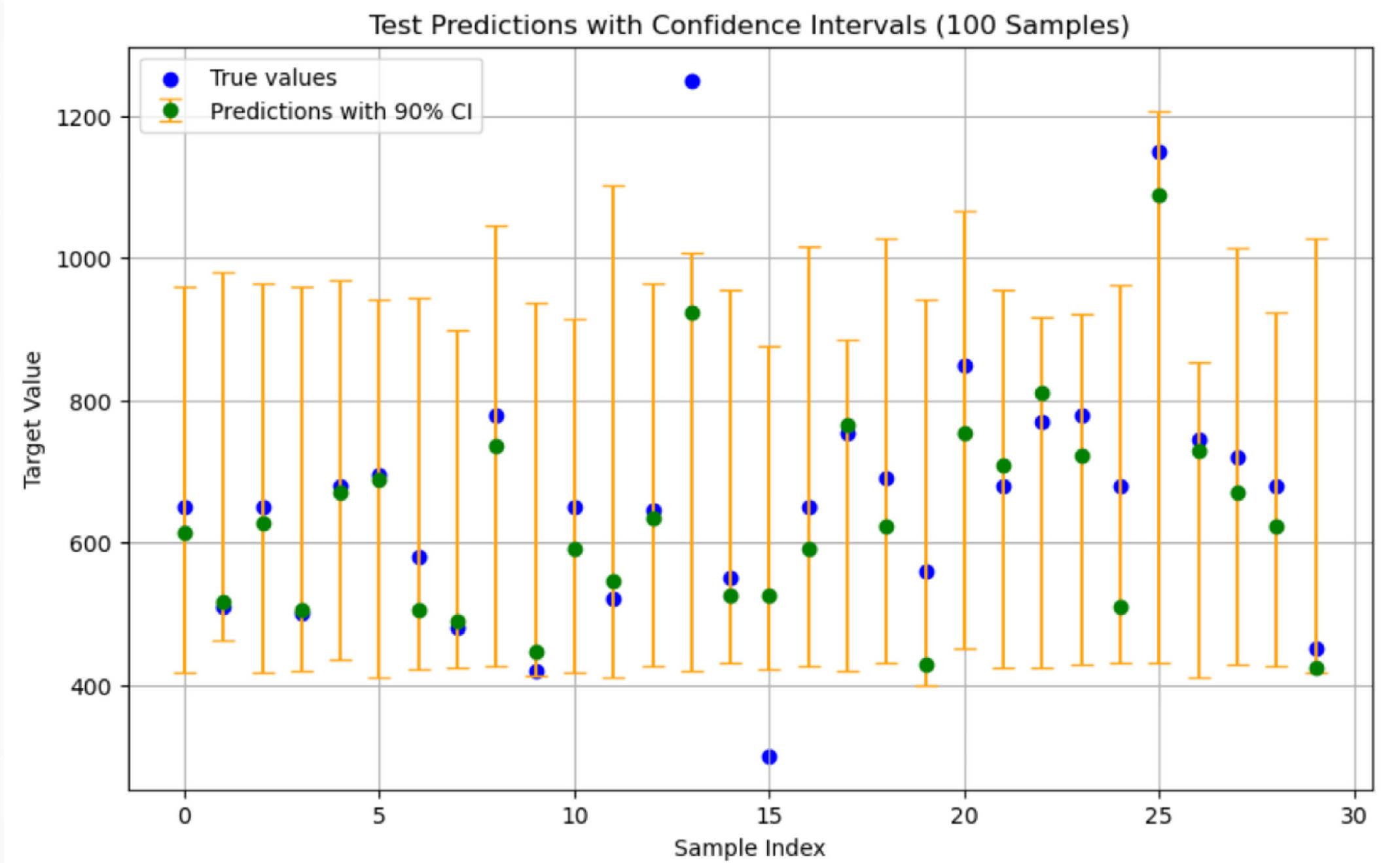
MODEL EVALUATION

- The model performs effectively for the majority of our data, capturing the common price patterns. However, it lacks the necessary signal to accurately predict prices for "premium properties" which encompass features like coveted views, exceptionally large properties, or prime locations that significantly impact the price.



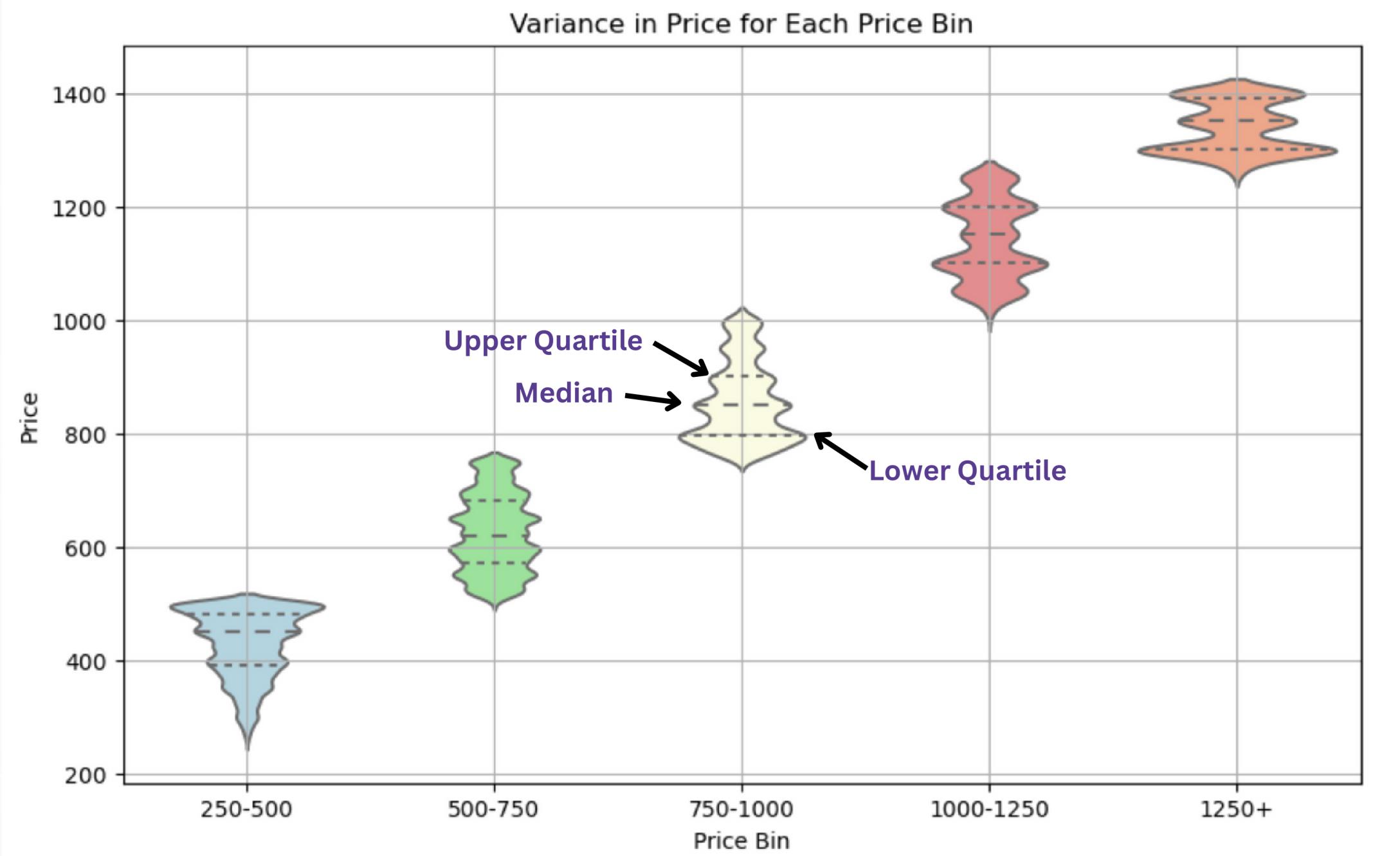
IMPLEMENTING CONFIDENCE INTERVALS

While implementing a prediction range, we've encountered an issue where the intervals are excessively large, making them unreliable for generating a prediction range.

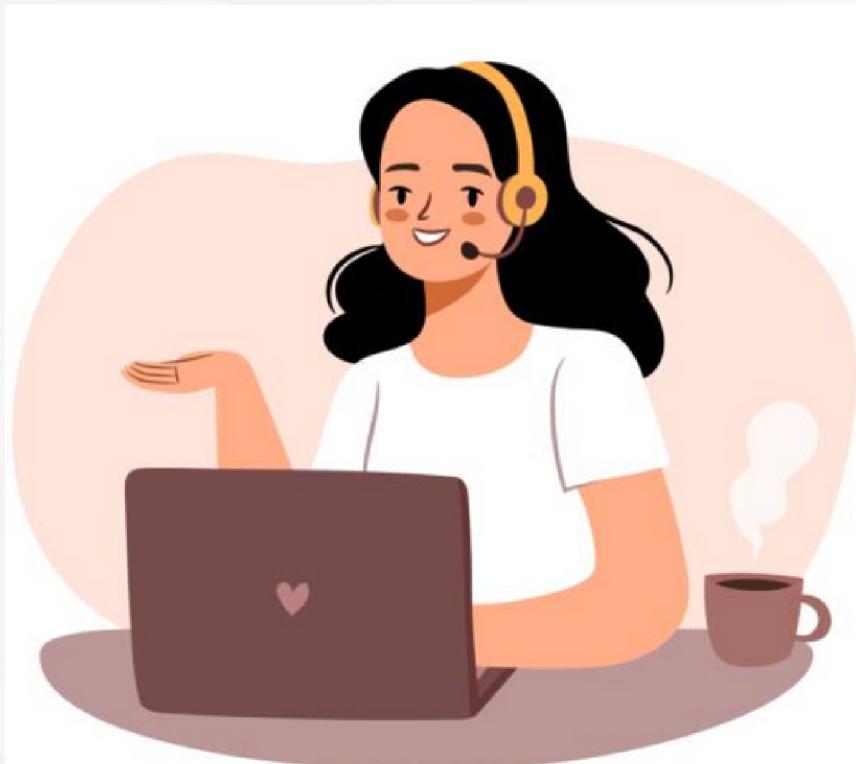


IMPLEMENTING CONFIDENCE INTERVALS

The wide price variance in our data contributes to the large confidence intervals in our predictions. Notably, the abundance of observations beyond the upper and lower quartiles, as depicted in the violin plots, creates this issue.



MODEL PROTOTYPE



Auckland Rental Price Prediction

Suburb

Mount Eden

Number of bathroom ensuites

0

Number of full bathrooms

2

Number of bedrooms

2

Apartment

False

Make Prediction

This image shows a user interface for a rental price prediction model. The title 'Auckland Rental Price Prediction' is at the top. Below it are several input fields: 'Suburb' (set to 'Mount Eden'), 'Number of bathroom ensuites' (set to '0'), 'Number of full bathrooms' (set to '2'), 'Number of bedrooms' (set to '2'), 'Apartment' (set to 'False'), and a 'Make Prediction' button. Each input field has a minus and plus sign for adjustment.

IMPROVEMENTS

- Additional data: number of photos, utilities provided, furnished
- Using the listing description text as input to a text model to create keywords that correlate to price could capture those premium listings that there is no signal for in our current dataset.



Rent includes Water/Hot Water, Power and unlimited premium wi-fi.

If you're looking to rent out luxury penthouse in arguably Auckland's best further than 1 Enfield!

The penthouse (603) is perched on the top of this iconic building with expansive views over the city and Mt Eden. This is the ultimate lock up & leave, investment or bolthole, comprising 3 bedrooms, 2 baths plus 2 car spaces (stacker) apartment.

Brand new, high-end spec building and ready to move into when it suits. Approx total area 150m² (inc 53m² of deck).

Luxury & Brand New Penthouse, Bills Included

\$ **\$1,999 per week**

Apartment 3 2 2



FINAL THOUGHTS

- Our regression model demonstrates consistent predictive capability for rental prices in Auckland, particularly for properties that are not classified as premium.
- Incorporating additional information from the listing description could enhance the model's ability to capture signals related to premium properties.
- Our prototype shows productionising the model creates great usability for real estate agents such as Jenny.





THANK YOU
Q & A