[Confidential]

Sound Realty
House Price Model

Ryan Mahtab

Agenda

4 Business Objective

5 Base Model Details

6 v2 Model Details

8 Technical Architecture

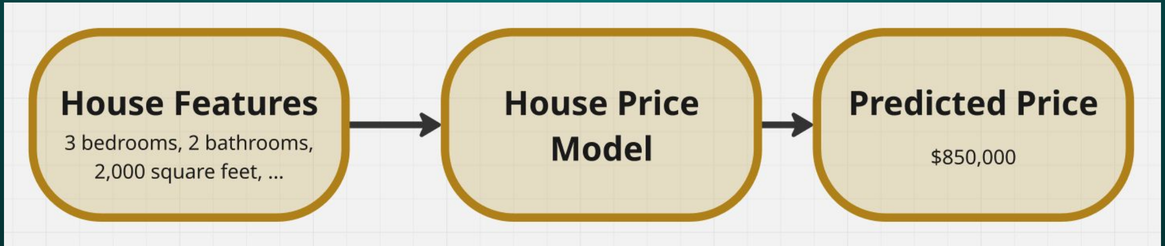
10 Areas of Improvement

Model Overview

Business Objective

[Confidential]

- Model that predicts house prices
- Trained on sales data from 5/2/14 - 5/27/15
- Model inputs
  - House features
  - Zip Code level demographic features
- Predictions returned to users in real-time



House Features

3 bedrooms, 2 bathrooms, 2,000 square feet, ...

House Price Model

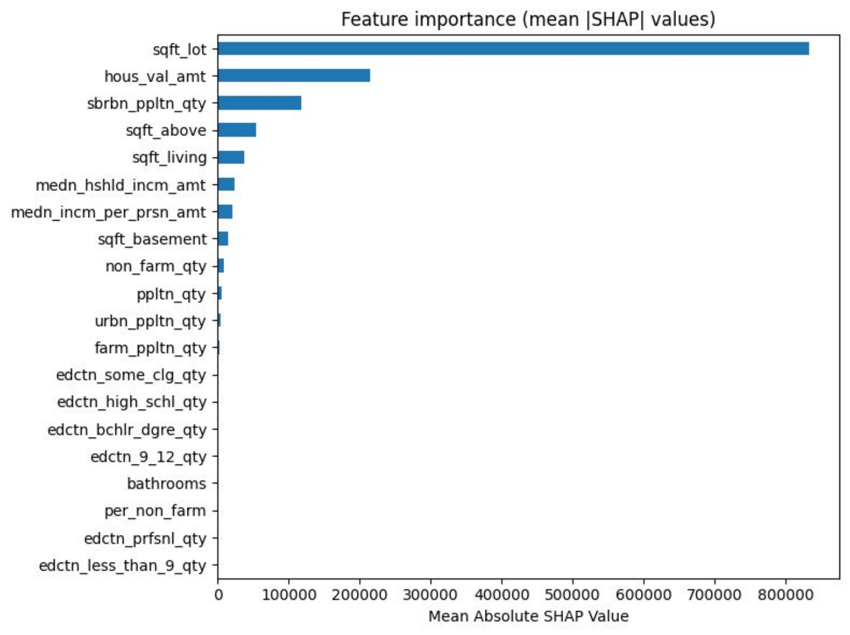
Predicted Price

$850,000

Base Model Details

[Confidential]

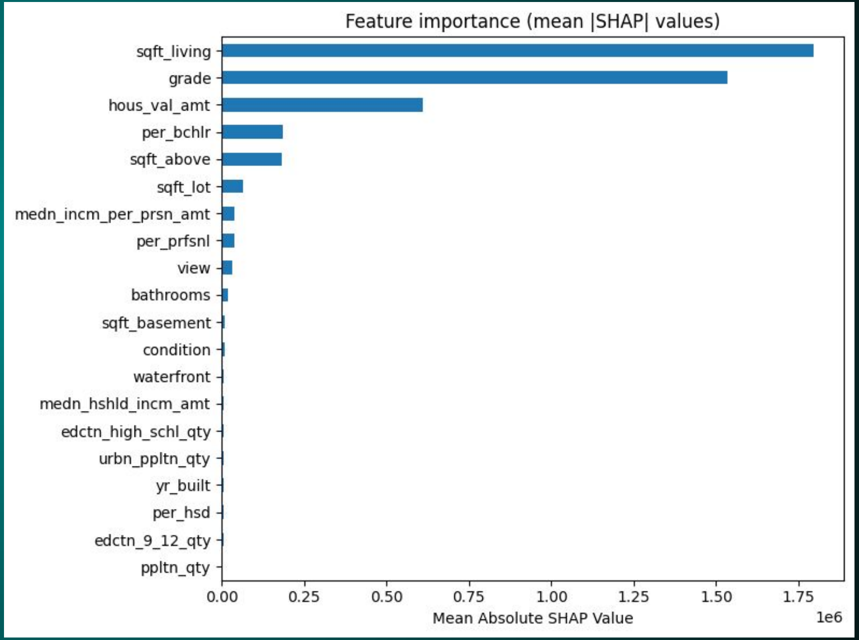
- Using 33 features out of 18 available
  - 7/18 property features
  - 26/26 demographic features
- Using Mean Absolute Error (MAE) as training metric
- Performance
  - $102,069 MAE
- Important features: property square footage, zip code house values, & suburban population



v2 Model Details

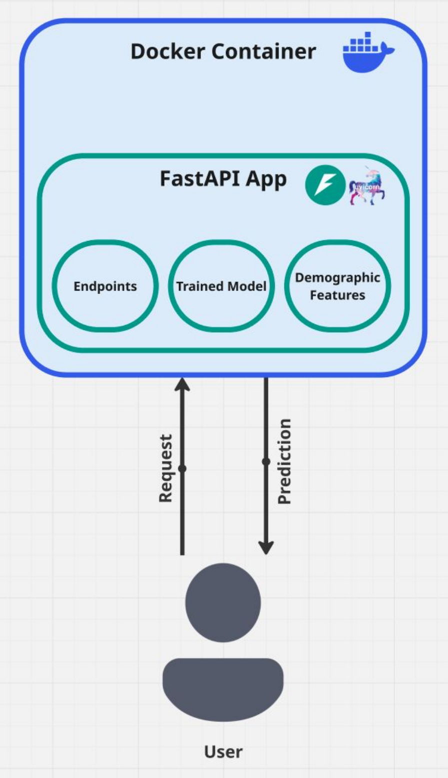
[Confidential]

- Using 39 features out of 18 available
  - 13/18 property features
  - 26/26 demographic features
- Performance
  - $75,071 MAE
- Important features: living square footage, property grade, & zip code house values



Technical Details

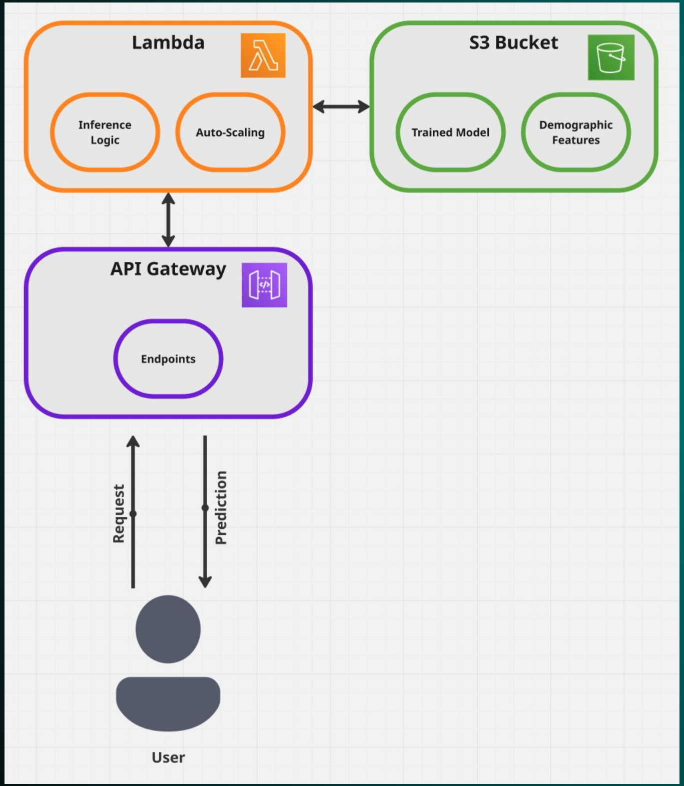
Technical Architecture



- Dockerized FastAPI service for real-time model inference
    - Loads preprocessing + model on startup
- Uvicorn server handles HTTP requests to endpoints
    - /predict and /predict-lite
- Pydantic schemas validate incoming payloads
- Seamlessly hot-swap models with /reload-model endpoint

Technical Architecture: Future State

[Confidential]



- API Gateway routes HTTP requests to Lambda function
- Lambda loads model artifacts from S3 bucket
- Lambda executes inference and returns JSON response
- Lambda concurrency handles horizontal auto-scaling
- S3 stores versioned ML models and features

Areas of Improvement

[Confidential]

- Supplement training data
  - More recent sales data (2015-2025)
  - Multi-year range of data
- Enrich feature set
  - Macroeconomic (mortgage rates, inflation indices, property tax rates)
  - Use lat/long to generate neighborhood clusters
- Expand model training
  - More extensive hyperparameter tuning
  - Experiment with more model algorithms (linear, boosting)