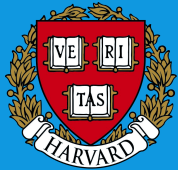


TEAM 1 (REX): Milestone 2

Preston Ching, Marcel Hedman,
Nam Luu, Owen Schafer



REX



Agenda



01

Problem Statement

Motivation, Scope of Work,
Lit Review

03

Baseline Model

Non-Submarket and
Submarket Approaches

02

Diving into the Data

Data Sources and Integration

04

Developed Model

Submarket Hierarchical
Approach

01

Problem Statement

Motivation, Scope of Work

Motivation



Industry Opportunity

\$9.6 trillion real estate market has limited technology disruption.



Price Indices

The NAR and Zillow only provide price indices which are noisy metrics due to spreads and approvals.



Non-Granular

Current solutions only provide granularity at the state or city level.

Scope of Work

Proposed Model

Our aim is to predict **demand and supply** by building a Bayesian **submarkets hierarchical model** in the Denver Market

Challenges

1. No clear definition of demand and supply **target variables**
2. Fixing or varying the **number of submarkets**

Success Metrics

1. Benchmark against non-Bayesian baseline models (with and without submarkets)
2. Accuracy and AUC on testing sets

Literature Review

01

Fall 2020 IACS Project



- Motivates measures of demand and supply
- **Top-down** approach (macroeconomic / census)

02

Hedonic Valuation
with Submarkets



- Motivates use of latent variable submarkets
- **Bottom-up** approach but still a **price index**

03

Our Model: Hedonic
Demand and Supply



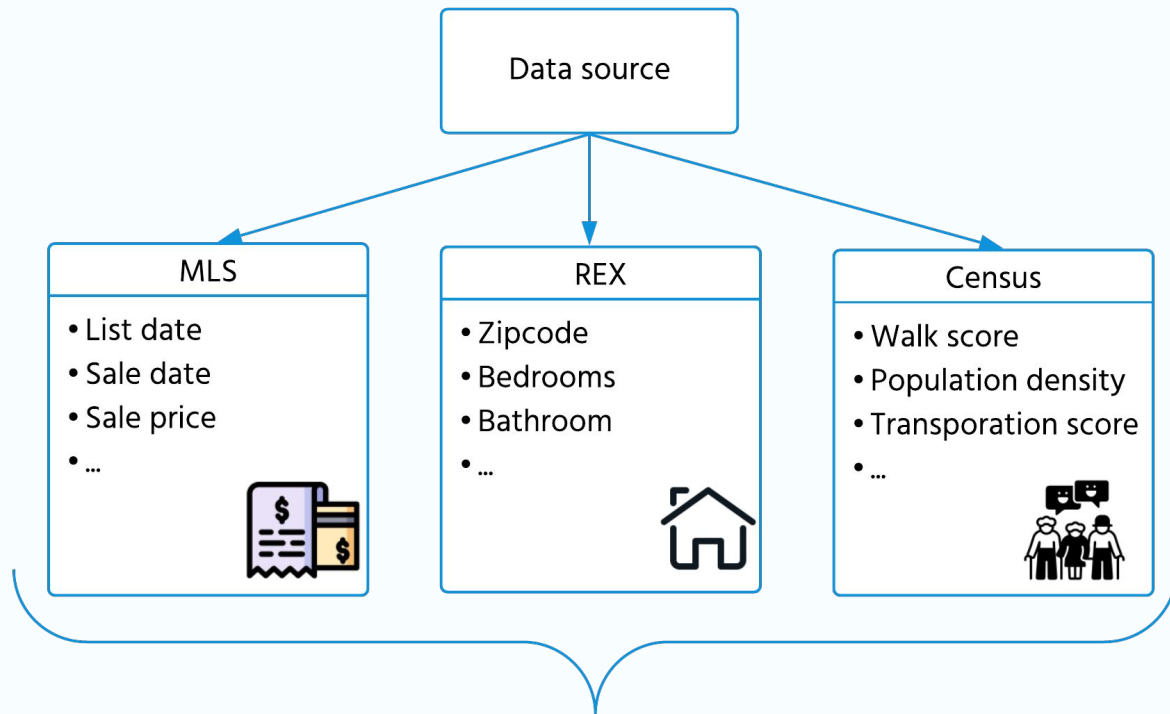
- Assumes measures of demand and supply are functions of the homes' attributes

02

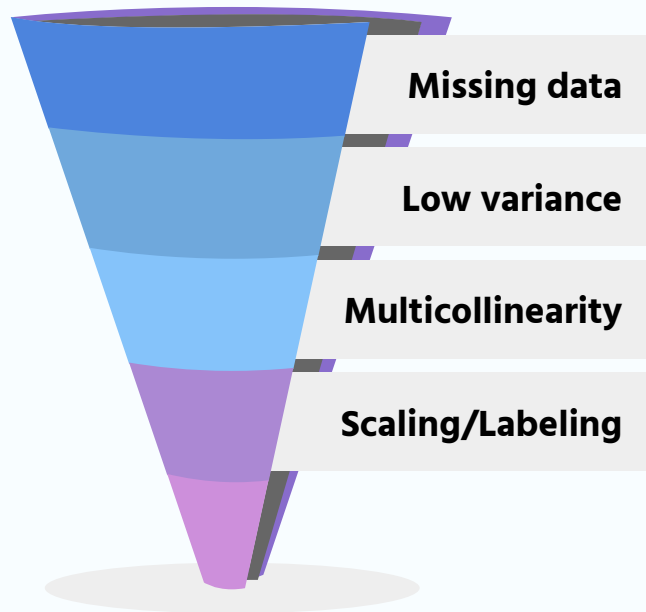
Diving Into the Data

Data Sources and Integration

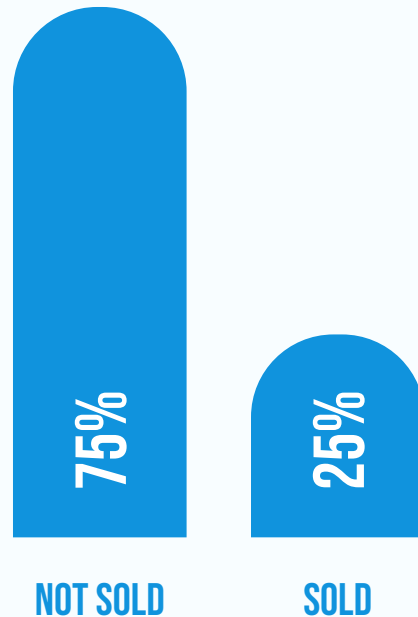
Datasets



Data Processing



Imbalanced dataset



03

Baseline Model

Non-Submarket and Submarket

What We are Predicting

GOAL

Demand

DETAIL

Predicted probabilities of house
being sold given that it was
listed in a time period

Evaluated using accuracy and
ROC-AUC

Overview of Models Developed

1

Non-submarket baseline

Parametric and
non-parametric
approaches

Overview of Models Developed

1

Non-submarket baseline

Parametric and
non-parametric
approaches

2

Submarket baseline

Naively assigning
to submarkets by
feature similarity

Overview of Models Developed

1

Non-submarket baseline

Parametric and
non-parametric
approaches

2

Submarket baseline

Naively assigning
to submarkets by
feature similarity

3

Developed models

Bayesian
approaches

Non-submarket Baseline



Incorporate new dataset
from REX

3 layers fully-connected
network

Low ROC-AUC

	Test accuracy	ROC-AUC
Logistic Regression	0.71	0.54
Random Forest	0.71	0.5
Neural Network	0.81	0.51

Submarket Baseline



5 distinct clusters from
K-means



5 distinct models in line
with **hedonic** assumptions

	Test accuracy	ROC-AUC
Logistic Regression	0.72	0.55
Neural Network	0.68	0.58

Weighted average across
submarkets

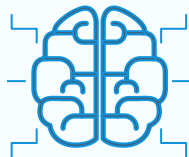
04

Developed Model

Submarket Hierarchical Approach



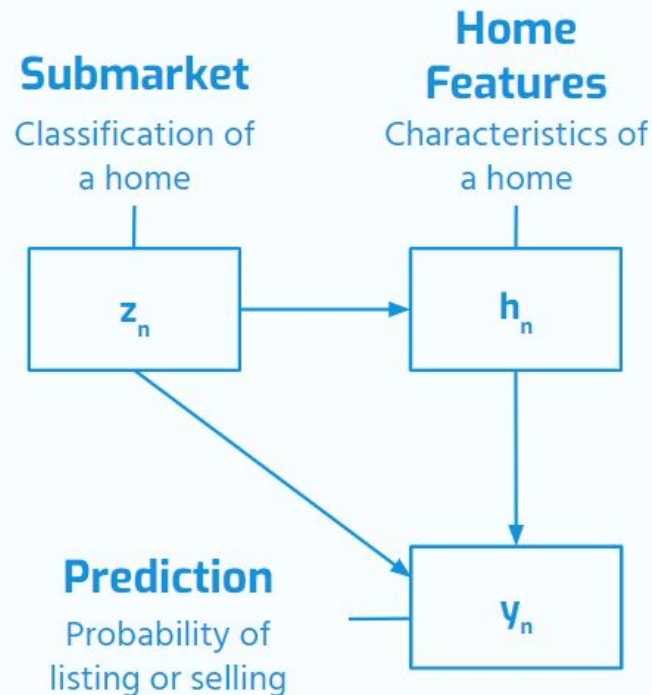
Model Hierarchy



Submarket Latent Variable

Model learns submarket classification and fit to observed data simultaneously

Identifies optimal submarket classification for most accurate prediction



Developed Model



Same dataset as baseline models

Bayesian and frequentist hierarchy implementations

Similar accuracies and ROC-AUC to baselines

	Test accuracy	ROC-AUC
pymc3	0.72	0.54
EM	0.73	0.57

Next Steps

01

Summarize Lessons Learned



- Imbalanced dataset is a barrier to accurate prediction
- **Hedonic assumption may not hold**

02

Augment data with macro features



- Known from past work to predict supply and demand well
- Allows submarkets to weight each macro factor differently

03

Handle Imbalance



- Simulate new data for less class imbalance

Q&A

Acknowledgements:

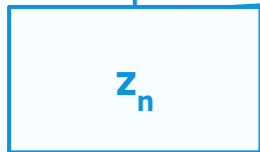
- Zona Kostic
- Will Fried
- Chris Tanner



Developed Model

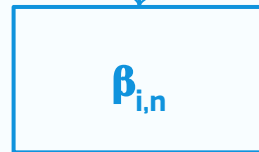
Submarket

Classification of
a home



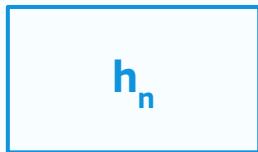
Weights

Used in supply or
demand model



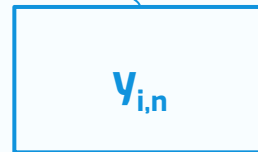
Home Features

Characteristics of
a home



Prediction

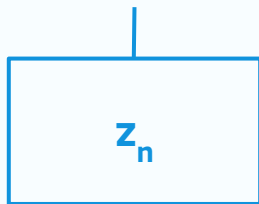
Probability of
listing or selling



Developed Model

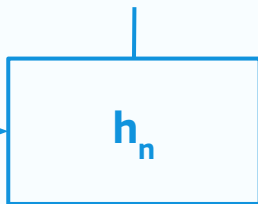
Submarket

Classification of
a home



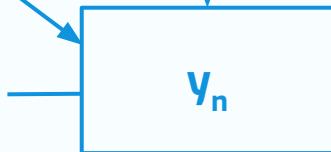
Home Features

Characteristics of
a home



Prediction

Probability of
listing or selling



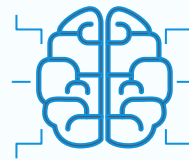
Baseline Model



Days on Market

We modelled discretized DOM
as a demand proxy

Non-time series approach



Time interval

Predicted probabilities of house
being sold given that it was
listed in a time period

In-line with developed model
approach