

TEAM 1 (REX): Final Presentation

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** by building a Bayesian **submarkets hierarchical model** in the Denver Market

Challenges

1. No clear definition of demand **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, AUC and R^2 on testing sets

Literature Review

01

Fall 2020 IACS Project



- Motivates measures of demand
- **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: Augmented
Hedonic Demand



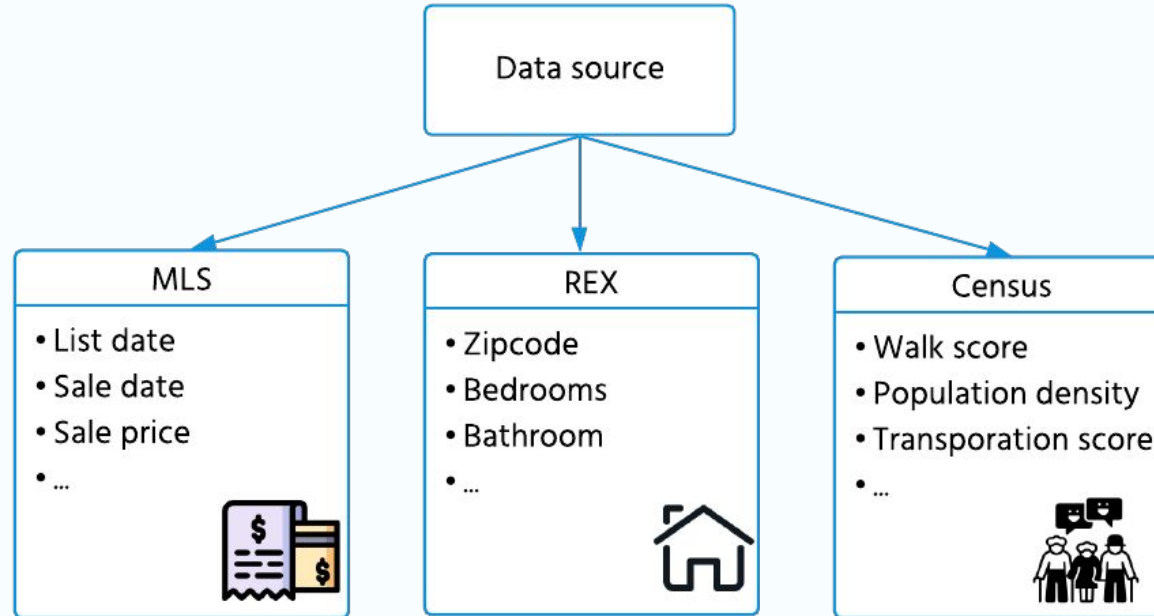
- Assumes main measures of demand are functions of the homes' attributes along with census-tract characteristics

02

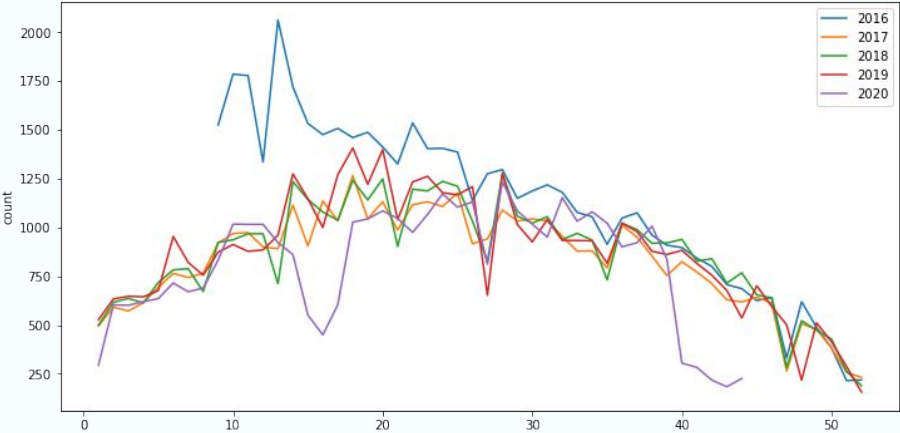
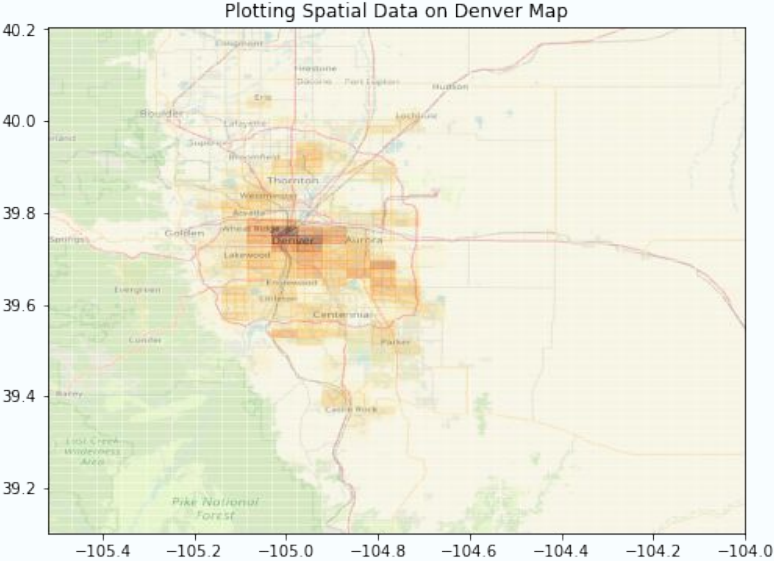
Diving Into the Data

Data Sources and Integration

Datasets

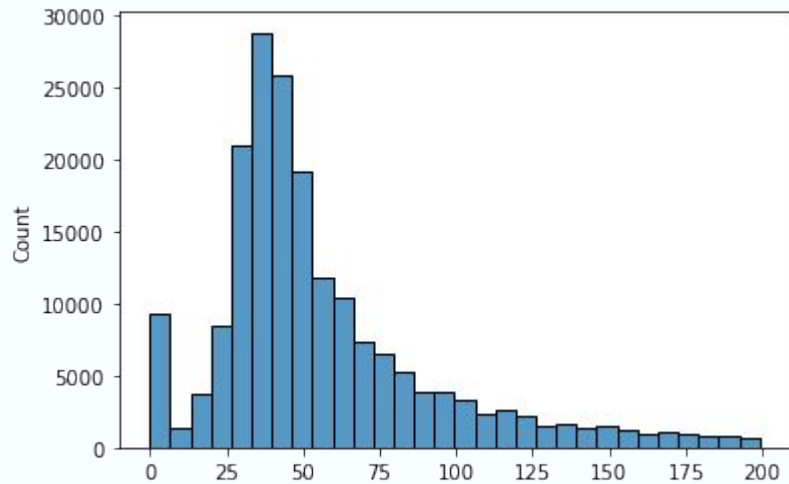


Data exploration



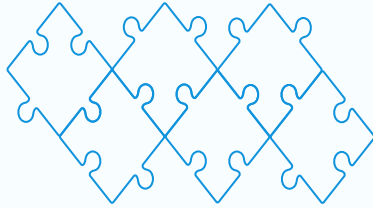
Target variable

$$y = \mathbf{P} \left(\frac{\text{property is sold}}{\text{property is listed}} \right)$$

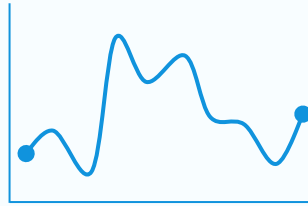


Distribution of days-on-market

Data Processing



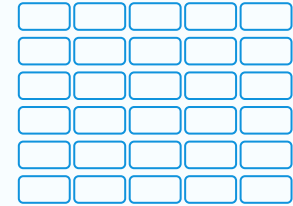
Combine features



Remove low
variance

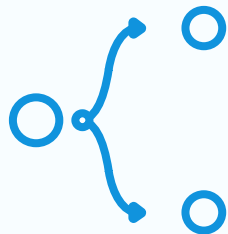
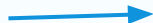


Handpick features



Correlation matrix

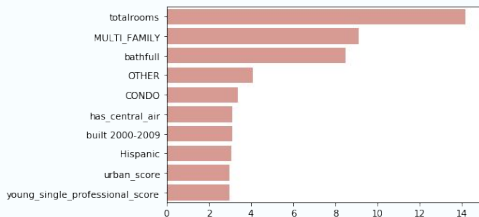
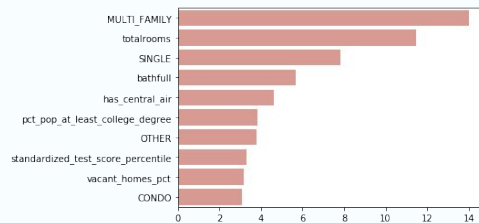
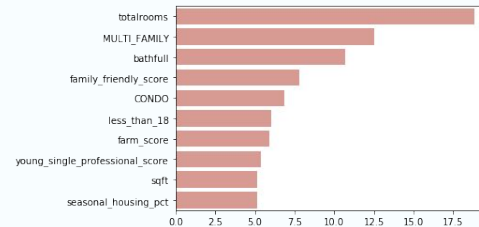
Data Processing



K-means



XGBoost



03

Baseline Model

Non-Submarket and Submarket

What We are Predicting

GOAL

Demand

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

DETAIL

Evaluated using

1. Accuracy and ROC-AUC
2. Predicted number homes sold and MSE/R^2

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 model

Bayesian
Hierarchical
approach



Non-submarket Baseline



Incorporate new dataset
from REX

Previously ran with
approaches such as Random
Forest and neural networks

**Test
accuracy**

ROC-AUC

**Predicted
number
of sales
/
Actual
number
of sales**

**Logistic
Regression**

0.63

0.54

2.07

Submarket Baseline



Cluster using K-Means



Cross validation to find
Optimal cluster number



Fit K distinct models in line
with **hedonic** assumptions



Ran for logistic regression
and XGBoost

Submarket Baseline

	No. Submarkets	Test accuracy	ROC-AUC	Predicted number of sales / Actual number of sales
Logistic Regression	11	0.63	0.55	0.9964
XGBoost	7	0.65	0.69	1.0004

Weighted average across
submarkets



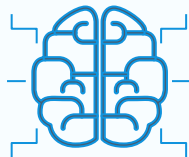
REX

04

Developed Model

Submarket Hierarchical Approach

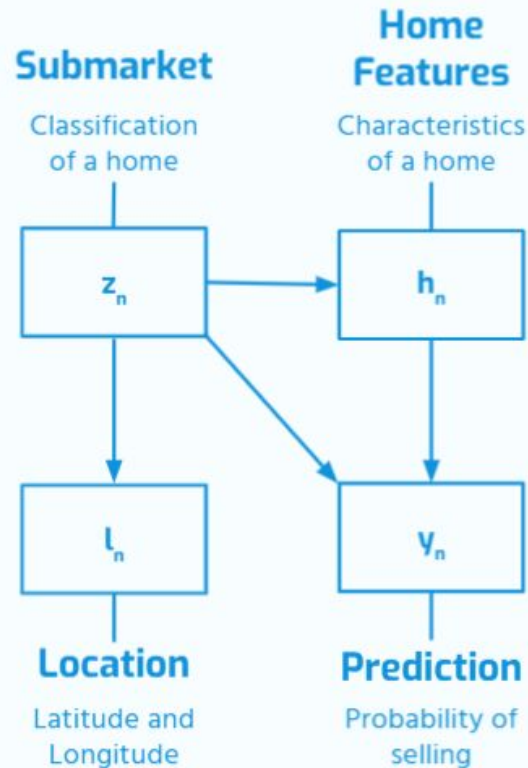
Model Hierarchy



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

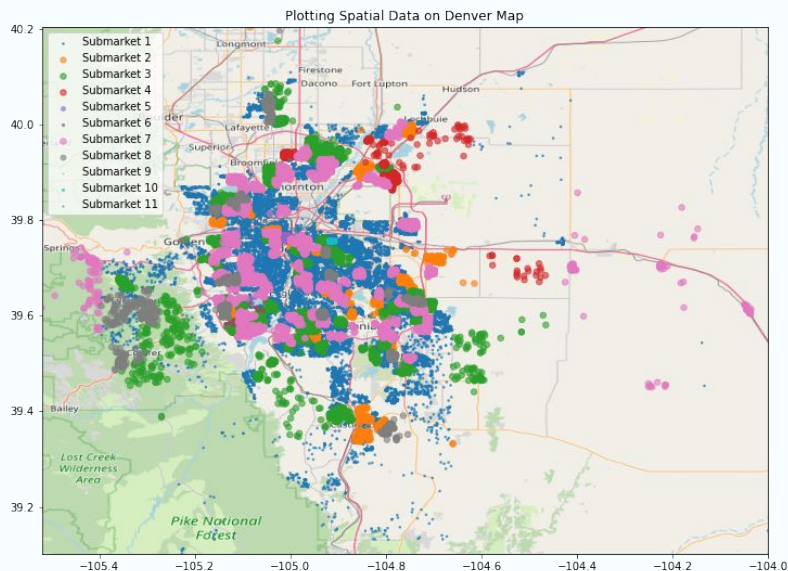
EM used to train models with different hedonic functions

Similar accuracies and ROC-AUC to baselines

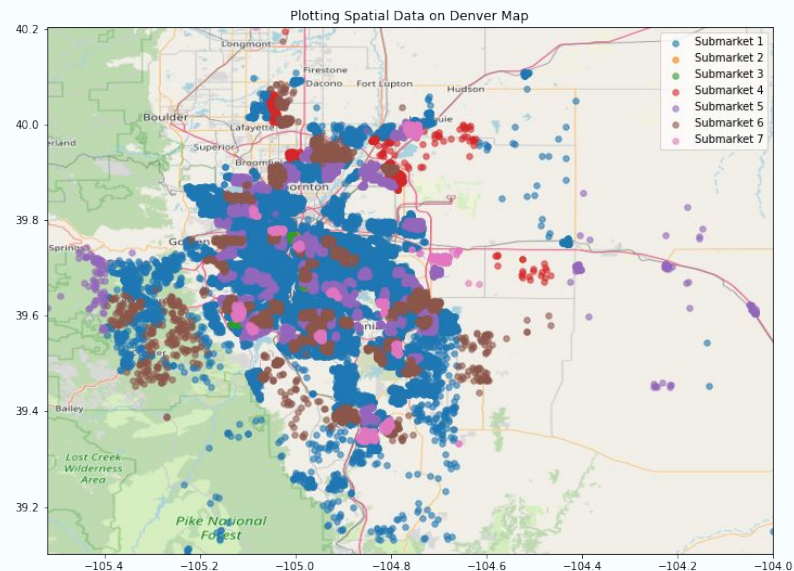
	Logistic Regression	XGBoost
No. Submarkets	11	7
Test Accuracy	0.62	0.66
ROC-AUC	0.50	0.62
MSE	383.0	511.9

Developed Model

Logistic Regression



XGBoost



Next Steps

01

Summarize Lessons Learned



- Aggregate prediction better than granular
- Prefer XGBoost hedonic function

02

Test Other Hedonic Demand Functions



- Logistic Regression and XGBoost are interpretable
- Prioritize balance between interpretability and performance

03

**Stretch Goal:
Extend to Time Series**



- Currently only focused on prediction for a fixed time frame

Q&A

Acknowledgements:

- Zona Kostic
- Will Fried
- Chris Tanner

