

TEAM REX: Milestone 1

Preston Ching, Marcel Hedman,
Nam Luu, Owen Schafer



REXX



Agenda



01

Problem Statement

Motivation, Scope of Work

02

Initial Process

Collaboration Infrastructure,
Learning Goals, Lit Review

03

Diving into the Data

Datasets, Exploratory Data
Analysis

04

Building the Model

Baseline Model, Developed
Model

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 **supply and demand** by building a Bayesian sub-markets **hierarchical model** in the **Denver Market**

Challenges

- No clear definition of supply and demand **target variables** as a function of listings / sales
- Fixing or varying the **number of sub-markets**

Success Metrics

- Accuracy and AUC on testing sets
- Benchmark against non-Bayesian ML models and last semester's model

02

Initial Process

Collaboration Infrastructure,
Learning Goals, Lit Review

Collaboration Infrastructure



Marcel

Data Cleaning +
Baseline Model



Nam

EDA + Baseline
Model



Preston

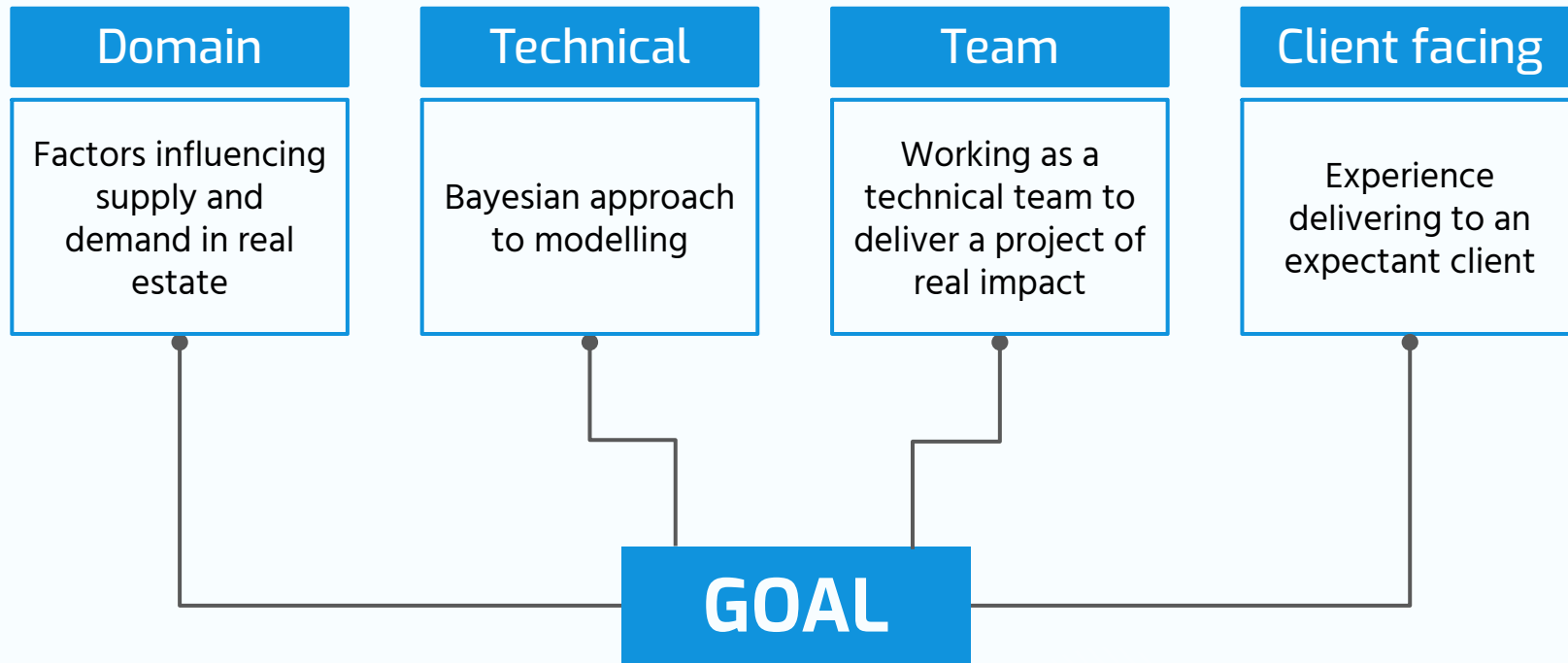
EDA + Developed
Model



Owen

Lit Review +
Developed Model

Learning Goals



Literature Review / Background

“Toward a Revamped Real Estate Index”

Previous AC 297R REX partnership, motivates domain and measures of supply and demand

“Modeling Submarket Effect for Real Estate Hedonic Valuation: A Probabilistic Approach”

Motivates use of latent variable submarket classification and hedonic supply and demand models

Hedonic Supply and Demand Model:

Assumes measures of supply and demand for a home are functions of the home's attributes



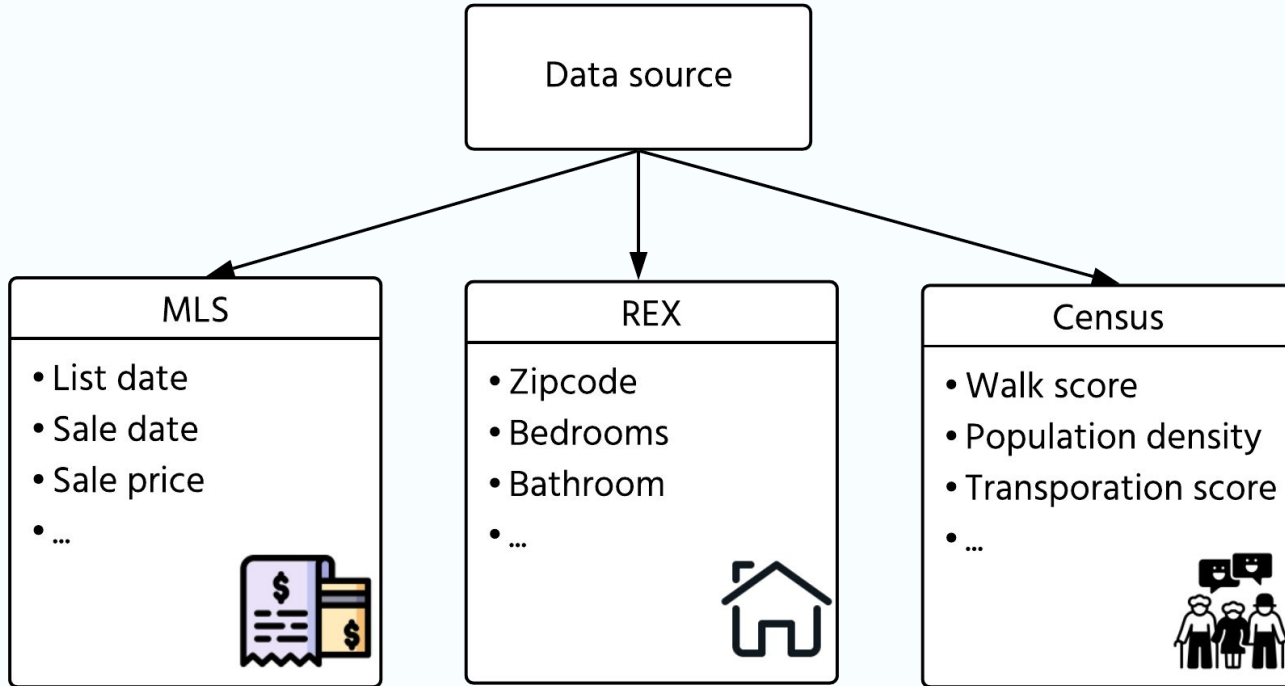
03

Diving Into the Data

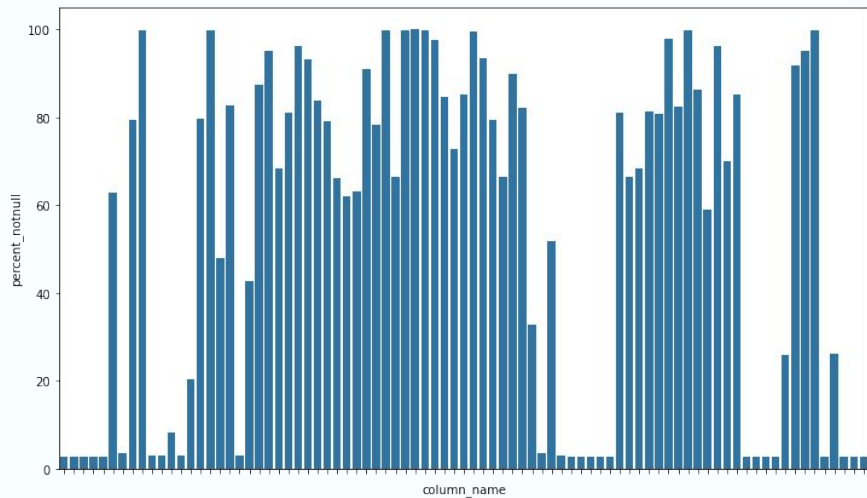
Datasets, Exploratory Data Analysis



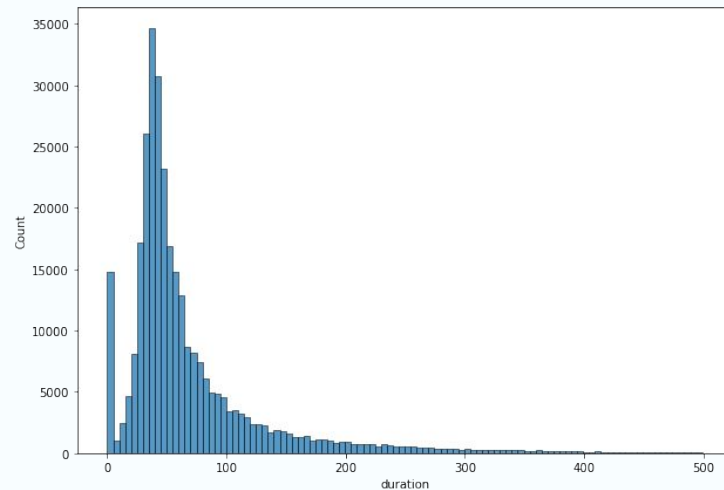
Datasets



Data Cleaning



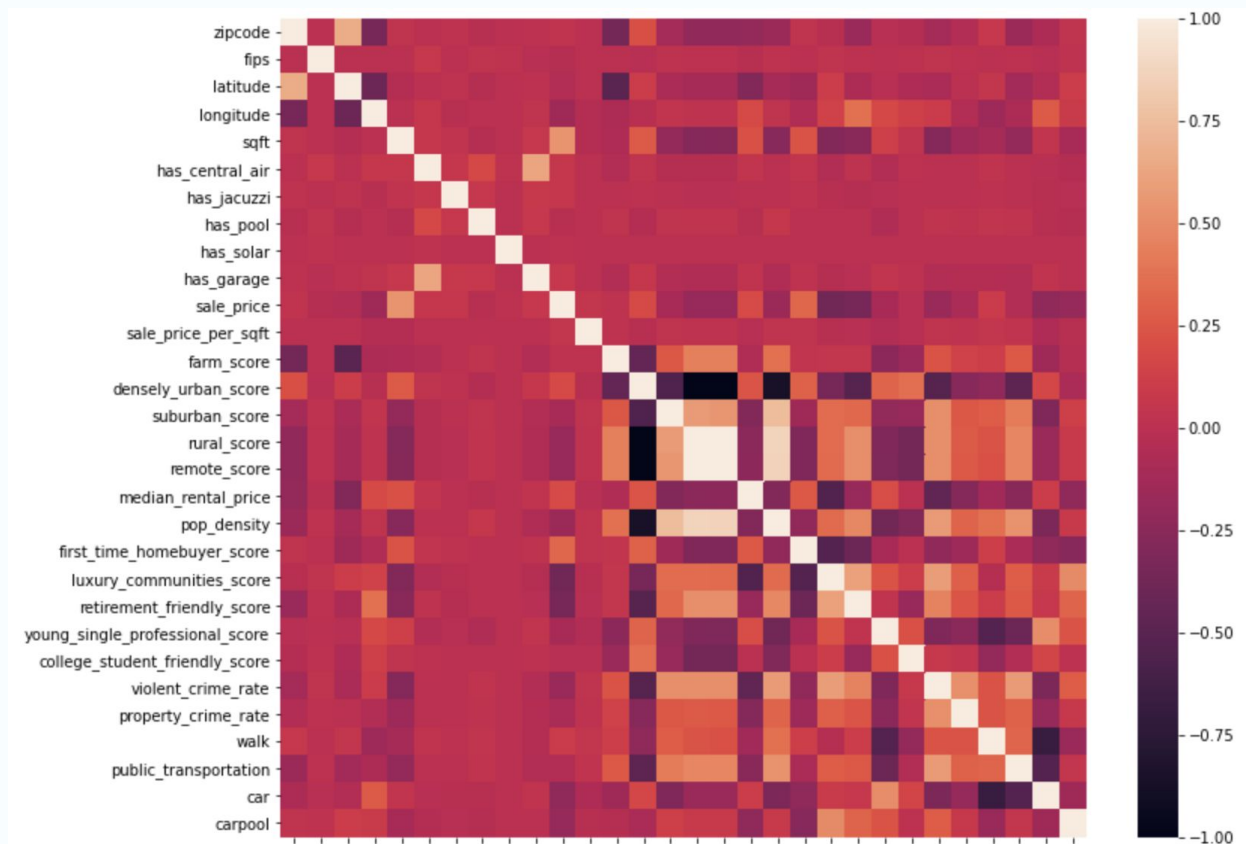
Missing values



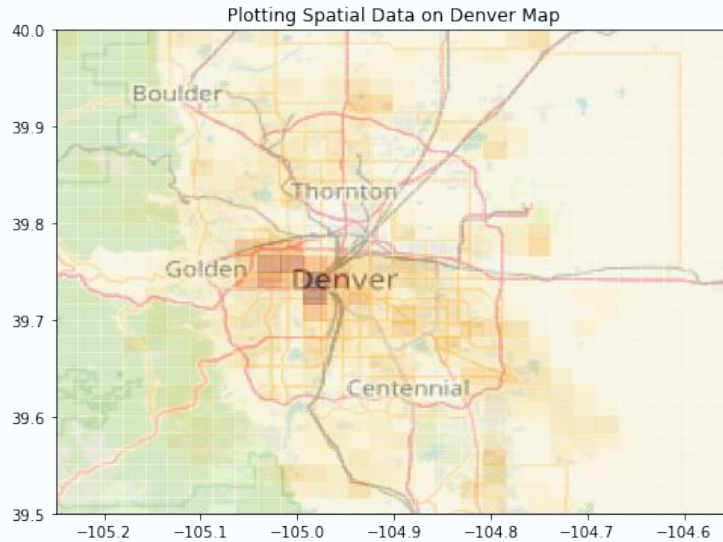
Outliers



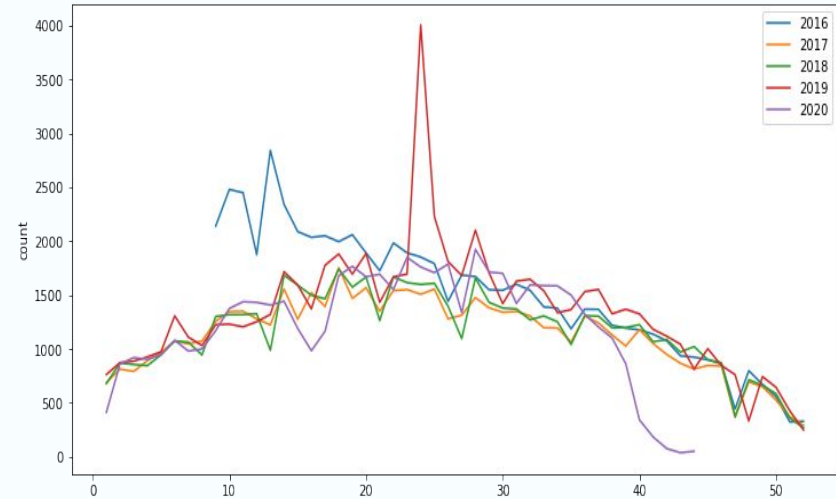
Data Exploration - Correlation



Data Exploration

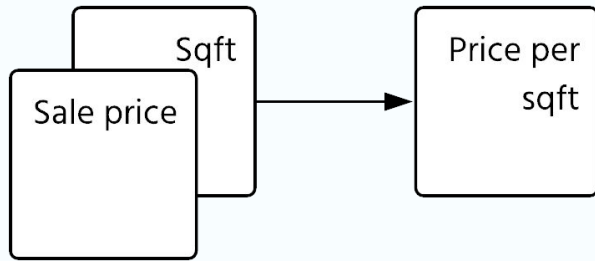


Geographical

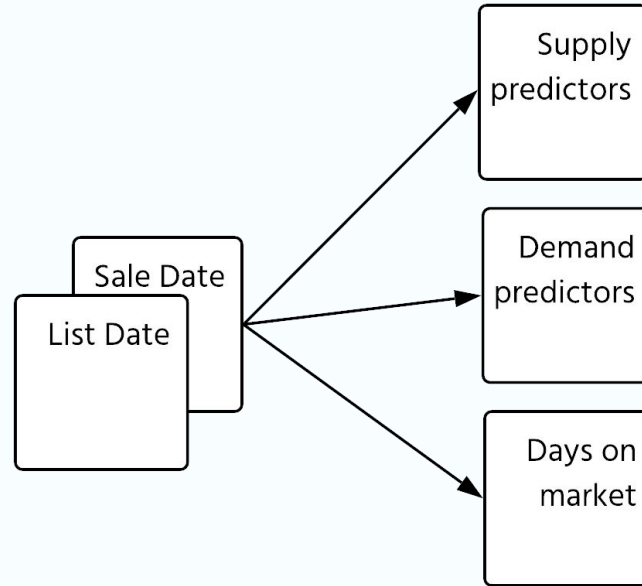


Temporal

Feature Engineering



Feature modifications



Target variables

04

Building The Model

Baseline Model, Developed Model

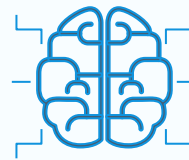
Baseline Model



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

Baseline Model

Days on Market

Mediocre model performance
but gave feedback for future
feature engineering/ handling
imbalance

	Test accuracy	AUC
Logistic Regression	0.63	0.57
Random Forest	0.63	0.59
Extra Trees Classifier	0.55	0.55

Baseline Model

Time interval

Saw major differences in total number predicted sold vs true number

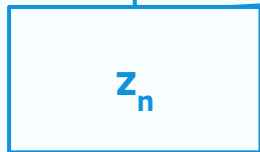
0.17 - ratio of predicted number of listings sold to true number sold

	Test accuracy	AUC
Logistic Regression	0.73	0.54

Developed Model

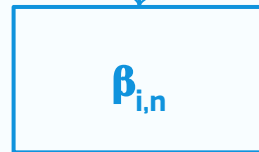
Submarket

Classification of
a home



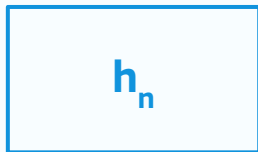
Weights

Used in supply or
demand model



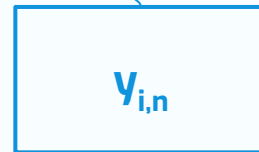
Home Features

Characteristics of
a home



Prediction

Probability of
listing or selling



Q&A

Acknowledgements:

- Zona Kostic
- Will Fried
- Chris Tanner

