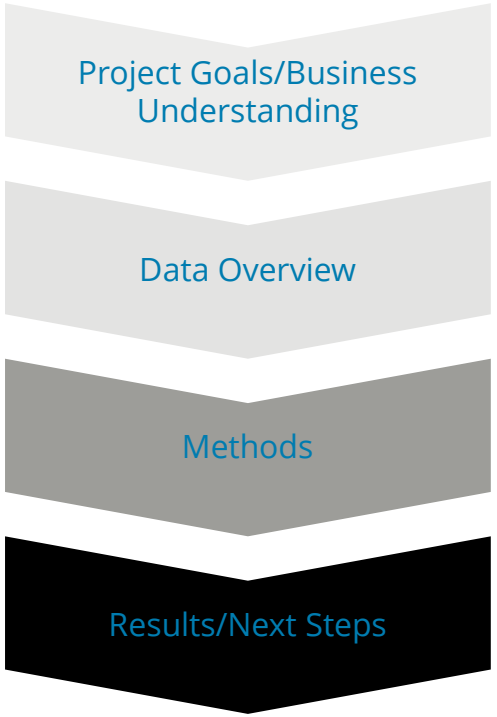
Deloitte.

Predicting Median Home Values in the California Housing Market

By Matthew Rodberg

Agenda



- Exploring the goals of the project and the business understanding/relevance
- Source of the data, limitations of data, understanding of data
- What particular methods were used to find our answers
- Explaining the results and any possible next steps to take

Business Problem

Market Uncertainty

No one wants to overpay

Imperfect information

Business Understanding

Stakeholders

Consumers	• Zillow Zestimate feature • Ability to estimate value before even looking at a home
Real Estate Firms	• Real Estate investing • Development
Mortgage Underwriters	• Ability to instantly generate market value of a home before underwriting

Data Understanding

Data limitations, applicability, and general information

Data Information

Data was gathered from Kaggle

- Subsidiary of Google LLC
- Data Science Hub
- Data taken from 1990 census
- Measures on a block-by-block basis

Data Features

To successfully estimate the price of a home, we must look to the features of the home

- Bedrooms
- Distance to ocean
- Bathrooms
- Etc

Applicability

California's housing market is notoriously high in comparison to the rest of the nation

- Value estimation may be inaccurate applied elsewhere
- Features identified as driving value will however remain somewhat applicable

Data Exploration/ Manipulation

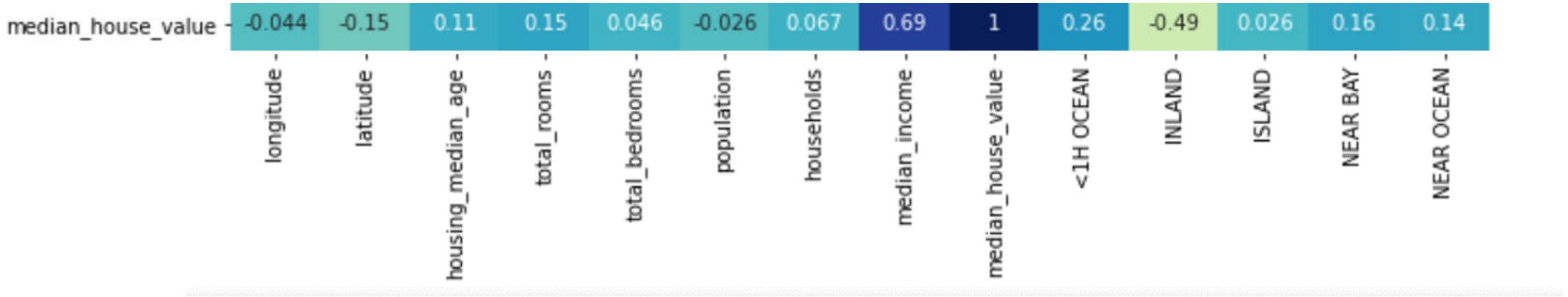
Insights

One Hot Encoding

- **Distance to ocean**
  - Close to ocean helps value
  - Inland detracts from value

Normalization

- **Log transformation**



Modeling

How I reached my conclusions

Linear Regression

- **Most apparent technique**
  - Yielded a good result but could be better
  - Required data preparation
  - 60% score

Decision Trees

- **Best left to classification problems**
  - Worst result out of the techniques used

Random Forest

- **Built the best model with score of around 83%**
- Best of the non NN options
- Need to know features over accuracy
- Less ability to apply elsewhere

Results

MEDIAN INCOME

PROXIMITY TO OCEAN

RANDOM FOREST
SCORE OF 83%

Next Steps

- More state data
- Tuning/feature engineering
- Try external data

Questions?