HomePro
Understanding the cost of your future home
The Team

Melanie Herscher

Yuna Kim

Nitin Swarup Sokhey

Preethi Raju
Market Opportunity
Home buying can be an Odyssey!

85% of people we surveyed specified that utility usage and environmental risks were crucial factors influencing their purchase decisions.

Industry Expects also identified first time home buyers being concerned about an asset’s “value” thinking of a home as an investment.

The Gap… Competitors are not investing in this space!
Calculate the price of your future home

Take the guesswork out of budgeting and discover the cost of your dream home with ease.
A Two Part Problem...

Predicting House Utilities Costs
Using Historical Climate & Energy Usage Data, we will be predicting future energy usage needs based on current climate projections.

Evaluating Homes Holistically
Using aggregated statistics for a given house, we will be creating a scoring algorithm to educate new homebuyers to make environmental and financially sensible decisions.
Predicting Home Price
Listings Map view

- **Purpose:**
  - Users from different State, Foreigners
  - which district / county to choose for less "Over Priced!" houses
Modelling Decisions

- **Data Preparation:**
  - 100,838 individual houses, 94 zipcodes
  - Each zipcodes: Train: 80%  Test: 20%

- **Feature Generation**
  - Base model version: 7 features + zipcode

- **Final Model**
  - Random Forest Regressor

- **Final Score Metric**
  - **Percentage difference** between Listed Price - Predicted Price
Predicting Utilities
Modelling Decisions

- **Data Sourcing**
  - NOAA and PG&E Public Datasets
  - January 2013 to Present Day
  - San Francisco Bay Area (108 Zip Codes)

- **Modelling Approach**
  - Combined Utilities Cost
  - Average Monthly Temperature Regressor
  - Used Pycat to Evaluate Multi-stream Models

- **Final Modelling Choices**
  - Light Gradient Boosting Machine
Examples of Predictions

Next 3 years predictions for zipcode: 95122

Next 3 years predictions for zipcode: 94402
Expected utilities cost change over the next 10 Years
Final Thoughts
Further Plans

- Expand Modelling outside of the SFBA
- Incorporate Further Climate & Neighborhood Data
- Deeper Analysis into Regional Utility Growth
- Rollout User Testing to Real Buyers!
Understanding your potential home’s cost should be simple!

HOMEPRO

Come See For Yourself!
Thank you!
## Breakdown of Team Responsibilities

<table>
<thead>
<tr>
<th>Team Member</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Melanie Herscher</strong></td>
<td>- Data Sourcing (Utilities)</td>
</tr>
<tr>
<td></td>
<td>- Data Engineering and Transformation (Utilities &amp; Optimization for Listings)</td>
</tr>
<tr>
<td></td>
<td>- Time Series Exploratory Data Analysis (Utilities)</td>
</tr>
<tr>
<td></td>
<td>- Initial Classical Time Series Modelling Analysis (Utilities)</td>
</tr>
<tr>
<td></td>
<td>- Exploratory Modelling Analysis (Listings)</td>
</tr>
<tr>
<td></td>
<td>- Project Management</td>
</tr>
<tr>
<td></td>
<td>- Presentation Creation &amp; Initial Script Generation</td>
</tr>
<tr>
<td></td>
<td>- Project iSchool Page Write-up &amp; Posting</td>
</tr>
<tr>
<td><strong>Yuna Kim</strong></td>
<td>- Exploratory Data Analysis (Utilities)</td>
</tr>
<tr>
<td></td>
<td>- Initial ML Time Series Modelling Analysis (Utilities)</td>
</tr>
<tr>
<td></td>
<td>- Advanced ML Modelling Analysis (Utilities)</td>
</tr>
<tr>
<td></td>
<td>- Exploratory Modelling Analysis (Listings)</td>
</tr>
<tr>
<td></td>
<td>- Final Modelling Analysis (Listings)</td>
</tr>
<tr>
<td></td>
<td>- Final Presentation Model Visualizations (Utilities &amp; Listings)</td>
</tr>
<tr>
<td><strong>Nitin Swarup Sokhey</strong></td>
<td>- Surveying Potential Buyers</td>
</tr>
<tr>
<td></td>
<td>- Final Demo Design &amp; Implementation</td>
</tr>
<tr>
<td><strong>Preethi Raju</strong></td>
<td>- Industry Expert Interviews</td>
</tr>
<tr>
<td></td>
<td>- Data Sourcing (Listings)</td>
</tr>
<tr>
<td></td>
<td>- Initial Data Engineering and Transformation (Listings)</td>
</tr>
<tr>
<td></td>
<td>- Exploratory Data Analysis (Listings)</td>
</tr>
<tr>
<td></td>
<td>- Initial Modelling Analysis (Listings)</td>
</tr>
</tbody>
</table>
Appendix
Correlation Plots for Temperature Against Usage
Baseline Modelling - Zip Code Specific Approach (Elec)

Residuals from ARIMA(3,0,4) with zero mean

Forecasts from ARIMA(3,0,4) with zero mean
Baseline Modelling - Zip Code Specific Approach (Gas)

Residuals from ARIMA(2,0,1) with non-zero mean

Forecasts from ARIMA(2,0,1) with non-zero mean
Advanced Modelling - VARs with Temperature
Autocorrelation & Partial Autocorrelations (Elec)
Autocorrelation & Partial Autocorrelations (Gas)
Utilites Cost Calculation Breakdown

- Avg Electricity Usage (population)
- Electricity cost
- Avg Gas Usage (population)
- Gas cost

Utilities Cost

Cost Prediction

Temperature Data

TS model + Forecast
House pricing Calculation Breakdown

- **Regression**
  - 94 zipcodes
  - 126,587 Individual Houses
  - Filter: 7 features
  - 100,838 individual houses
  - group by: each zipcodes
  - Train 80%  Test 20%

- **Classification**
  - Group by: 5 quantile by price
House price model Breakdown

94 zipcodes
126,587 Individual Houses

Filter:
7 features + zipcodes
100,838 individual houses

Regression

- group by: each zipcodes
- Train 80%
- Test 20%

Classification

- group by: 5 quantile by price

with Zipcode
with vs without Z

without Zipcode
Average Percentage Difference

**Actual - Prediction**

Average percentage change delta

white: negative values **underpriced**

**RED:** positive values **overpriced**

Paratemetized for individual listings on Website
Zipcode Effect Model

**House as is vs Zipcode Effect**
Price predicted without zipcode - Price predicted with zipcode model

- **white:** houses under-priced to zipcode (locational) cause than the features of the house itself
- **RED:** houses over-priced due to zipcode (locational) cause than the features of the house itself
Future model plans: Customization

94 zipcodes
126,587 Individual Houses

Features:
- Customization
- Base Model
- House Renovation Features
- Nearby school related Features

- 25 features

Sample image
Utilities Prediction Full Modelling Evaluation List

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>R2</th>
<th>RMSLE</th>
<th>MAPE</th>
<th>TT (Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lightgbm</td>
<td>17.8752</td>
<td>2260.2701</td>
<td>42.4133</td>
<td>0.7546</td>
<td>0.1266</td>
<td>0.0749</td>
<td>0.2100</td>
</tr>
<tr>
<td>rf</td>
<td>18.3278</td>
<td>2262.1368</td>
<td>42.6986</td>
<td>0.7462</td>
<td>0.1343</td>
<td>0.0777</td>
<td>0.2360</td>
</tr>
<tr>
<td>dt</td>
<td>22.1468</td>
<td>2867.2272</td>
<td>49.4165</td>
<td>0.6531</td>
<td>0.1649</td>
<td>0.0963</td>
<td>0.1680</td>
</tr>
<tr>
<td>gbr</td>
<td>25.4198</td>
<td>3165.3358</td>
<td>52.4726</td>
<td>0.6398</td>
<td>0.1714</td>
<td>0.1109</td>
<td>0.1900</td>
</tr>
<tr>
<td>et</td>
<td>25.0975</td>
<td>3366.8988</td>
<td>55.6102</td>
<td>0.5811</td>
<td>0.1719</td>
<td>0.1080</td>
<td>0.2320</td>
</tr>
<tr>
<td>knn</td>
<td>26.5964</td>
<td>3615.6036</td>
<td>58.1512</td>
<td>0.5414</td>
<td>0.1828</td>
<td>0.1139</td>
<td>0.1660</td>
</tr>
<tr>
<td>ada</td>
<td>37.5112</td>
<td>5139.4785</td>
<td>69.3230</td>
<td>0.3545</td>
<td>0.2398</td>
<td>0.1731</td>
<td>0.1700</td>
</tr>
<tr>
<td>br</td>
<td>44.7779</td>
<td>7674.2269</td>
<td>85.8472</td>
<td>-0.0063</td>
<td>0.3130</td>
<td>0.1954</td>
<td>0.1700</td>
</tr>
<tr>
<td>ridge</td>
<td>44.7822</td>
<td>7676.4847</td>
<td>85.8581</td>
<td>-0.0065</td>
<td>0.3131</td>
<td>0.1954</td>
<td>0.1660</td>
</tr>
<tr>
<td>lar</td>
<td>44.7825</td>
<td>7676.6461</td>
<td>85.8589</td>
<td>-0.0065</td>
<td>0.3131</td>
<td>0.1954</td>
<td>0.1780</td>
</tr>
<tr>
<td>lr</td>
<td>44.7825</td>
<td>7676.6461</td>
<td>85.8589</td>
<td>-0.0065</td>
<td>0.3131</td>
<td>0.1954</td>
<td>0.1800</td>
</tr>
<tr>
<td>huber</td>
<td>50.8121</td>
<td>7590.8207</td>
<td>85.7223</td>
<td>-0.0128</td>
<td>0.3226</td>
<td>0.2536</td>
<td>0.1820</td>
</tr>
<tr>
<td>lasso</td>
<td>50.8291</td>
<td>7917.9244</td>
<td>87.2857</td>
<td>-0.0419</td>
<td>0.3298</td>
<td>0.2438</td>
<td>0.1640</td>
</tr>
<tr>
<td>en</td>
<td>50.8287</td>
<td>7917.9756</td>
<td>87.2860</td>
<td>-0.0419</td>
<td>0.3298</td>
<td>0.2438</td>
<td>0.1660</td>
</tr>
<tr>
<td>llar</td>
<td>50.8291</td>
<td>7917.9244</td>
<td>87.2857</td>
<td>-0.0419</td>
<td>0.3298</td>
<td>0.2438</td>
<td>0.1800</td>
</tr>
<tr>
<td>dummy</td>
<td>50.8290</td>
<td>7917.9005</td>
<td>87.2856</td>
<td>-0.0419</td>
<td>0.3298</td>
<td>0.2438</td>
<td>0.1690</td>
</tr>
<tr>
<td>omp</td>
<td>50.3881</td>
<td>8014.9476</td>
<td>87.8201</td>
<td>-0.0549</td>
<td>0.3306</td>
<td>0.2393</td>
<td>0.1780</td>
</tr>
<tr>
<td>par</td>
<td>175.8288</td>
<td>38592.5676</td>
<td>195.8245</td>
<td>-4.5574</td>
<td>2.3152</td>
<td>0.8983</td>
<td>0.1720</td>
</tr>
</tbody>
</table>
Video: Utilities Predictions Through Time Interactive