HomePro

Understanding the cost of your future home

The Team

Melanie Herscher

Yuna Kim

Nitin Swarup Sokhey

Preethi Raju

Market Opportunity

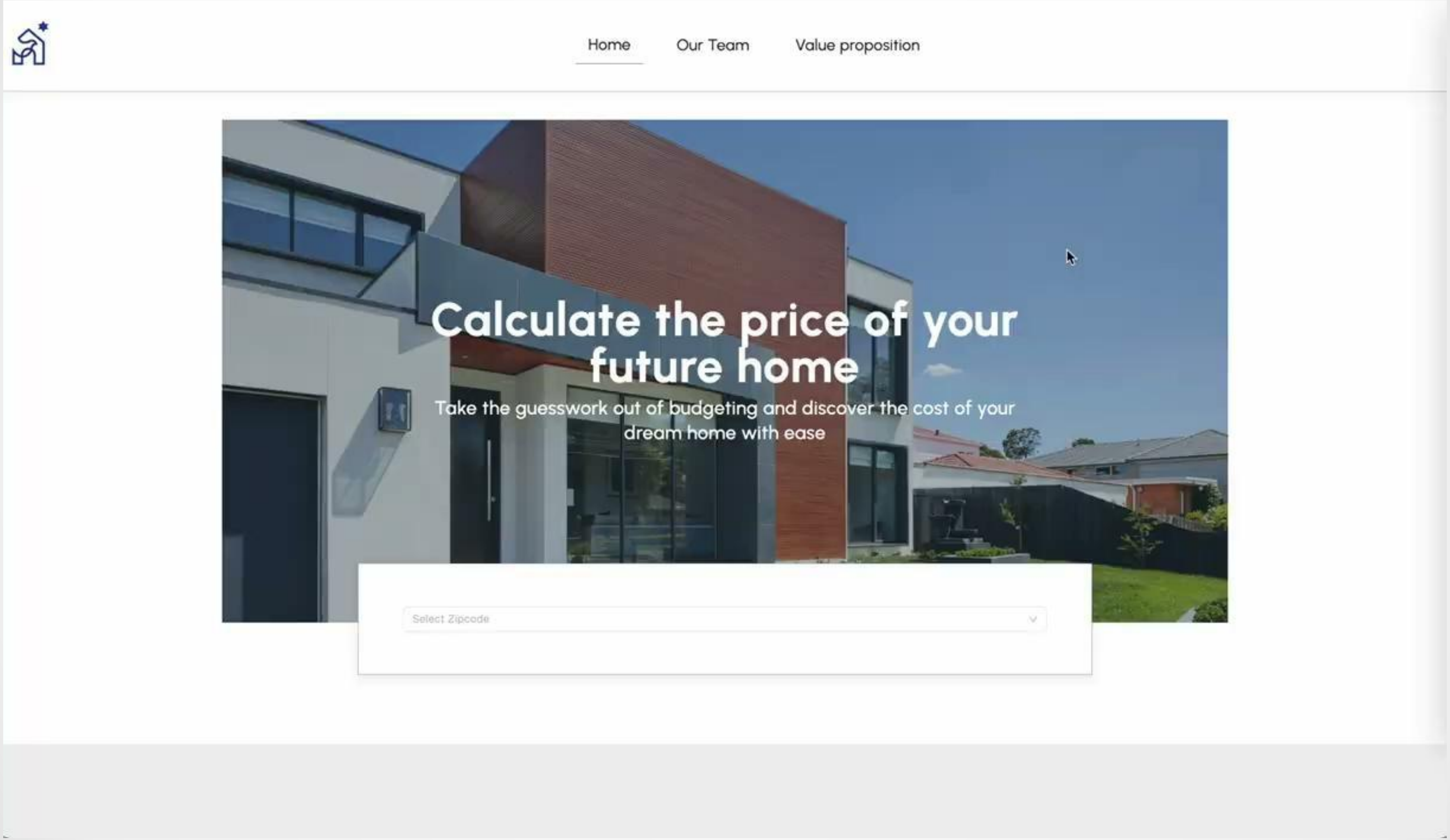
Home buying can be an Odyssey!

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!

DEMO


Home
Our Team
Value proposition
Calculate the price of your future home
Take the guesswork out of budgeting and discover the cost of your dream home with ease
Select Zipcode

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.

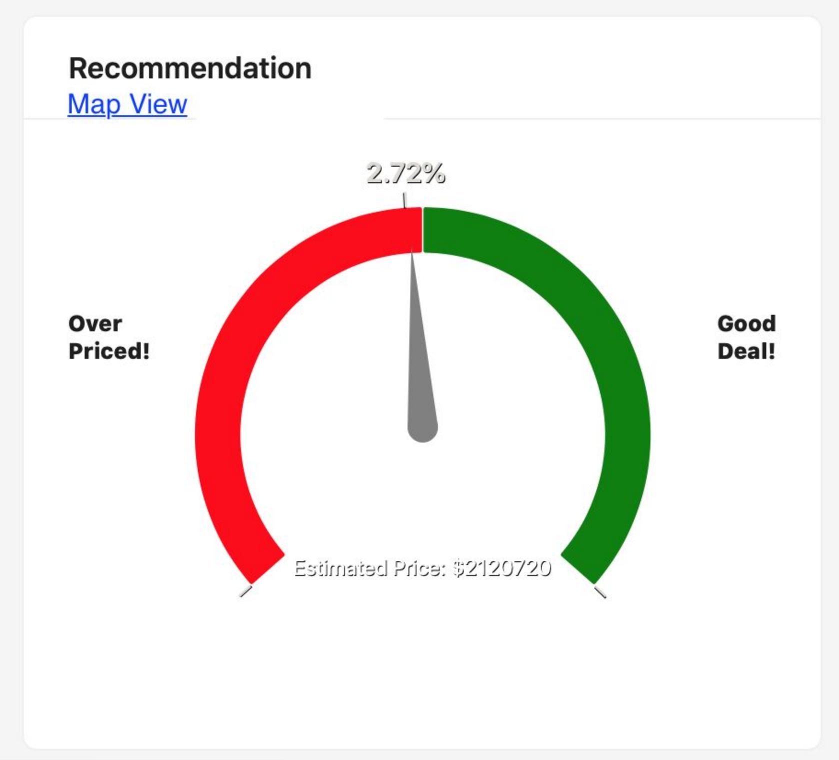
Evaluating Homes Holistically

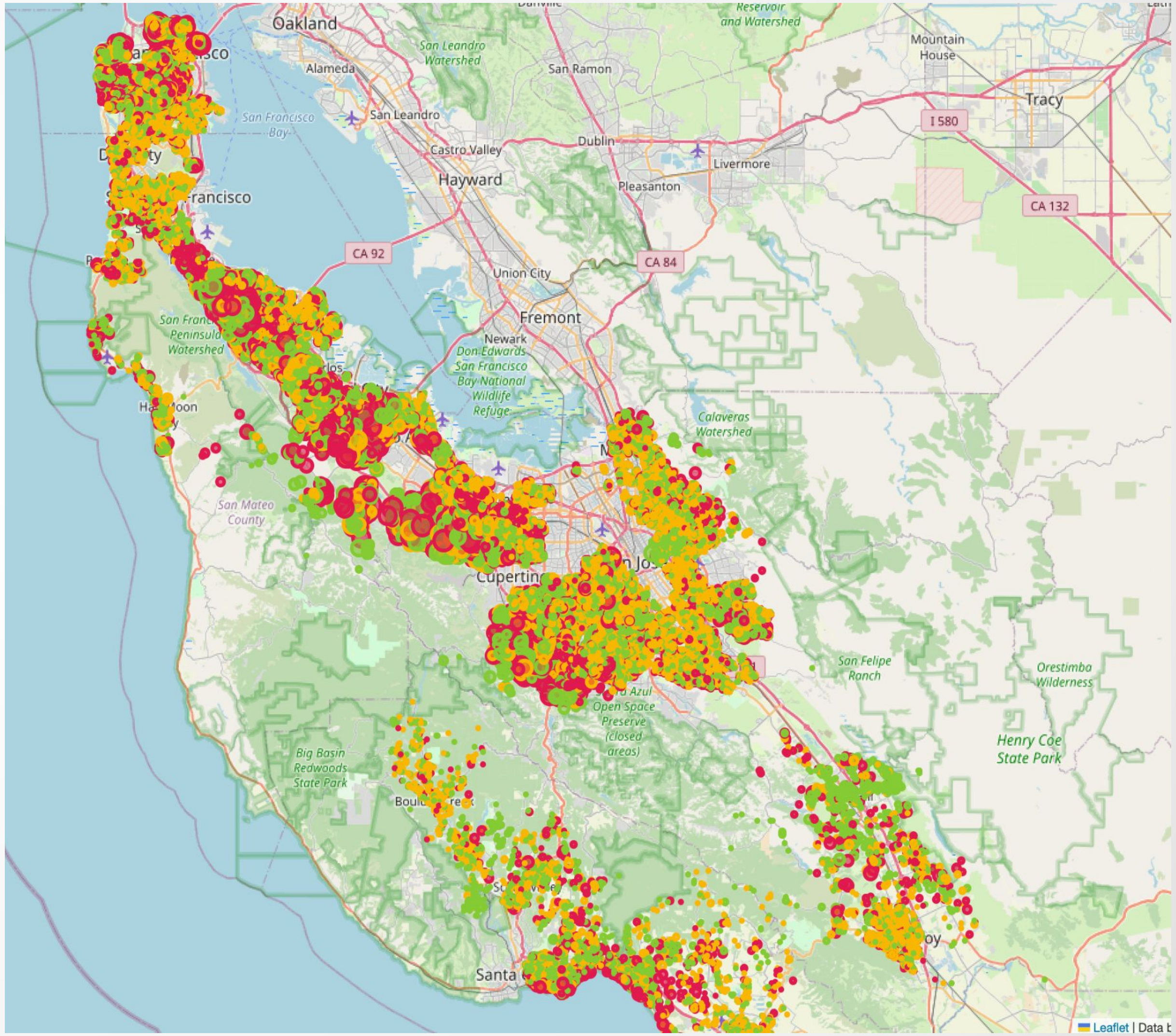
Using aggregated statistics for a given house, we will be creating a scoring algorithm to educate new homebuyers make environmental and financially sensible decisions.

Predicting Home Price

Listings Map view

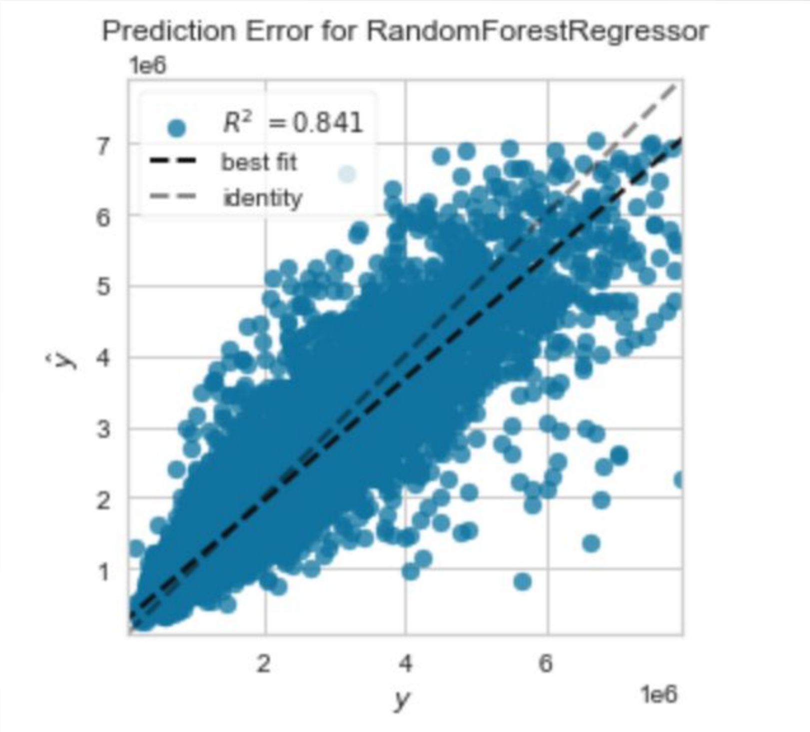
- **Purpose:**
  - Users from different State, Foreigners
  - which district / county to choose
    for less "Over Priced!" houses

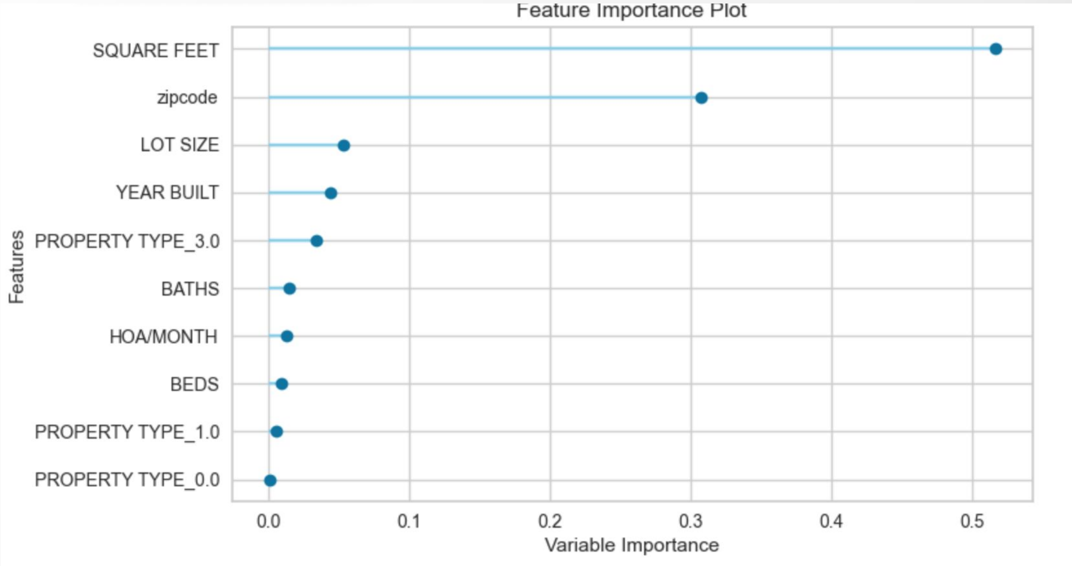




Modelling Decisions

- **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





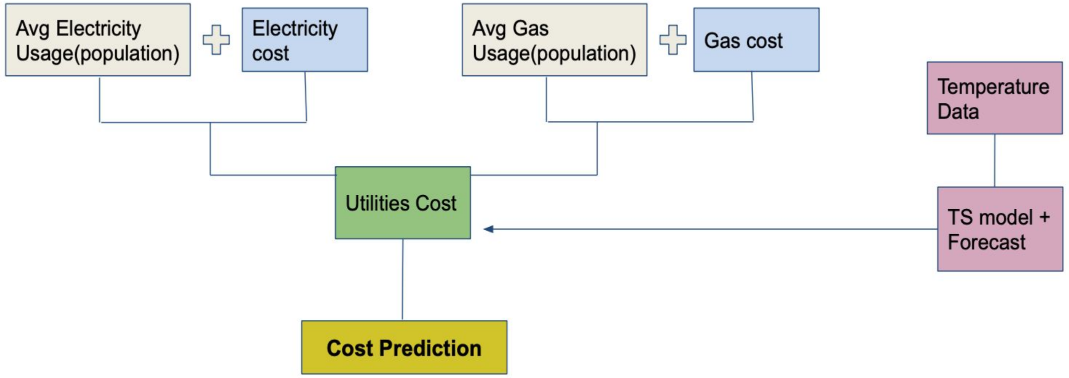
Predicting Utilities

Modelling Decisions

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

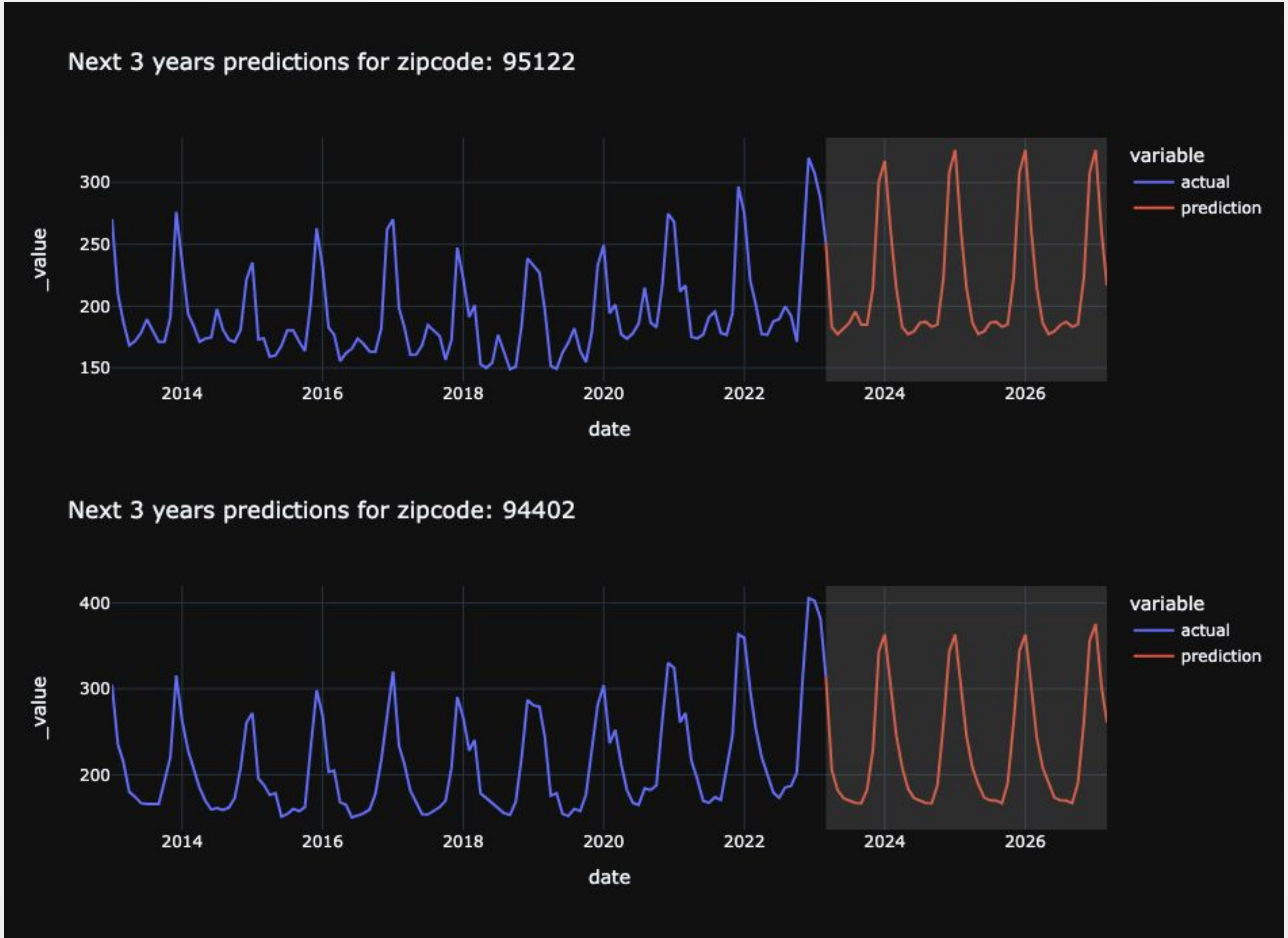
- **Modelling Approach**
  - Combined Utilities Cost
  - Average Monthly Temperature Regressor
  - Used Pycat to Evaluate Multi-stream Models

- **Final Modelling Choices**
  - Light Gradient Boosting Machine

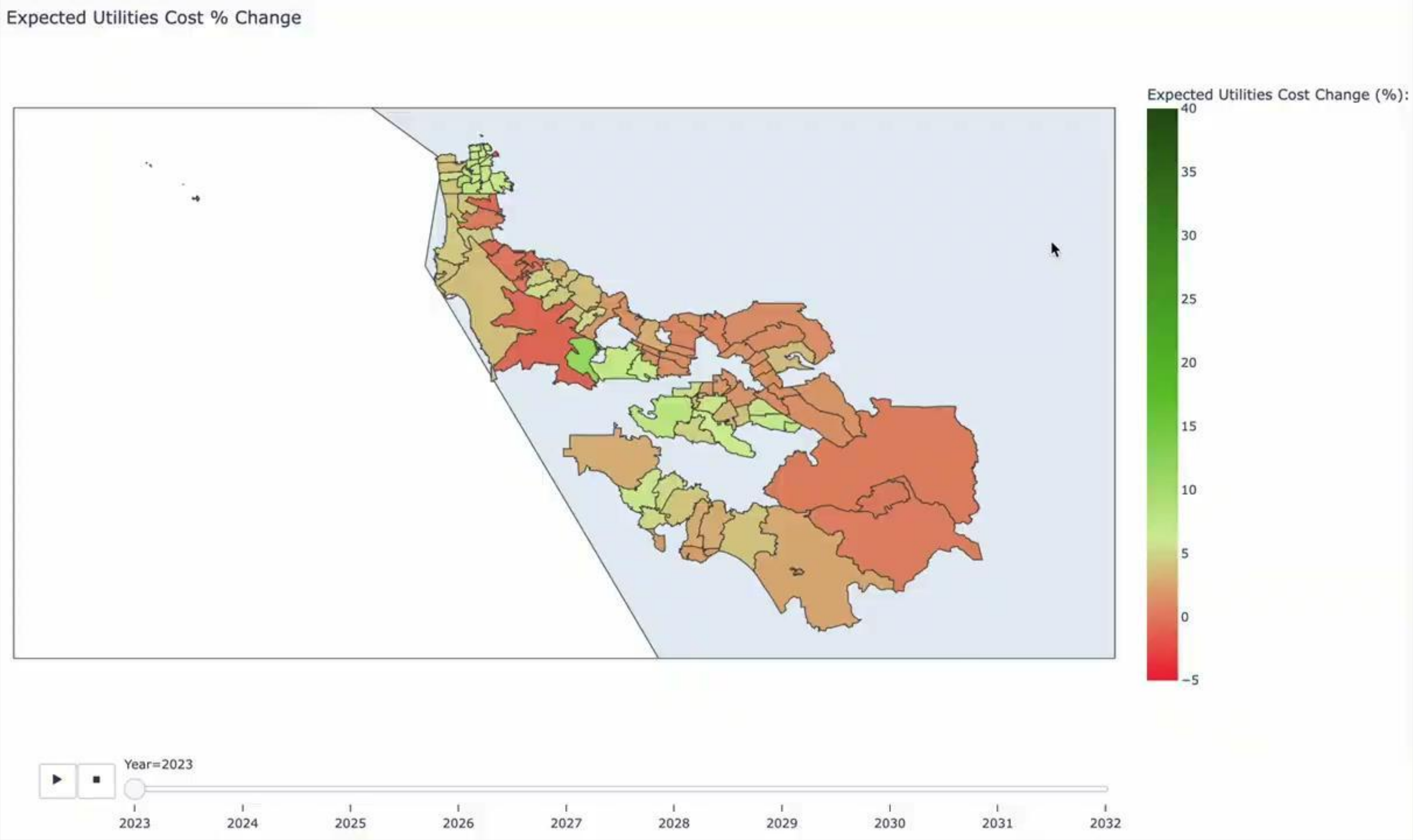


	Model	MAE	MSE	RMSE	R2
lightgbm	Light Gradient Boosting Machine	17.8752	2260.2701	42.4133	0.7546
rf	Random Forest Regressor	18.3278	2262.1368	42.6986	0.7462
dt	Decision Tree Regressor	22.1468	2867.2272	49.4165	0.6531
gbr	Gradient Boosting Regressor	25.4198	3165.3358	52.4726	0.6336
et	Extra Trees Regressor	25.0975	3366.8988	55.6102	0.5811

Examples of Predictions



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!

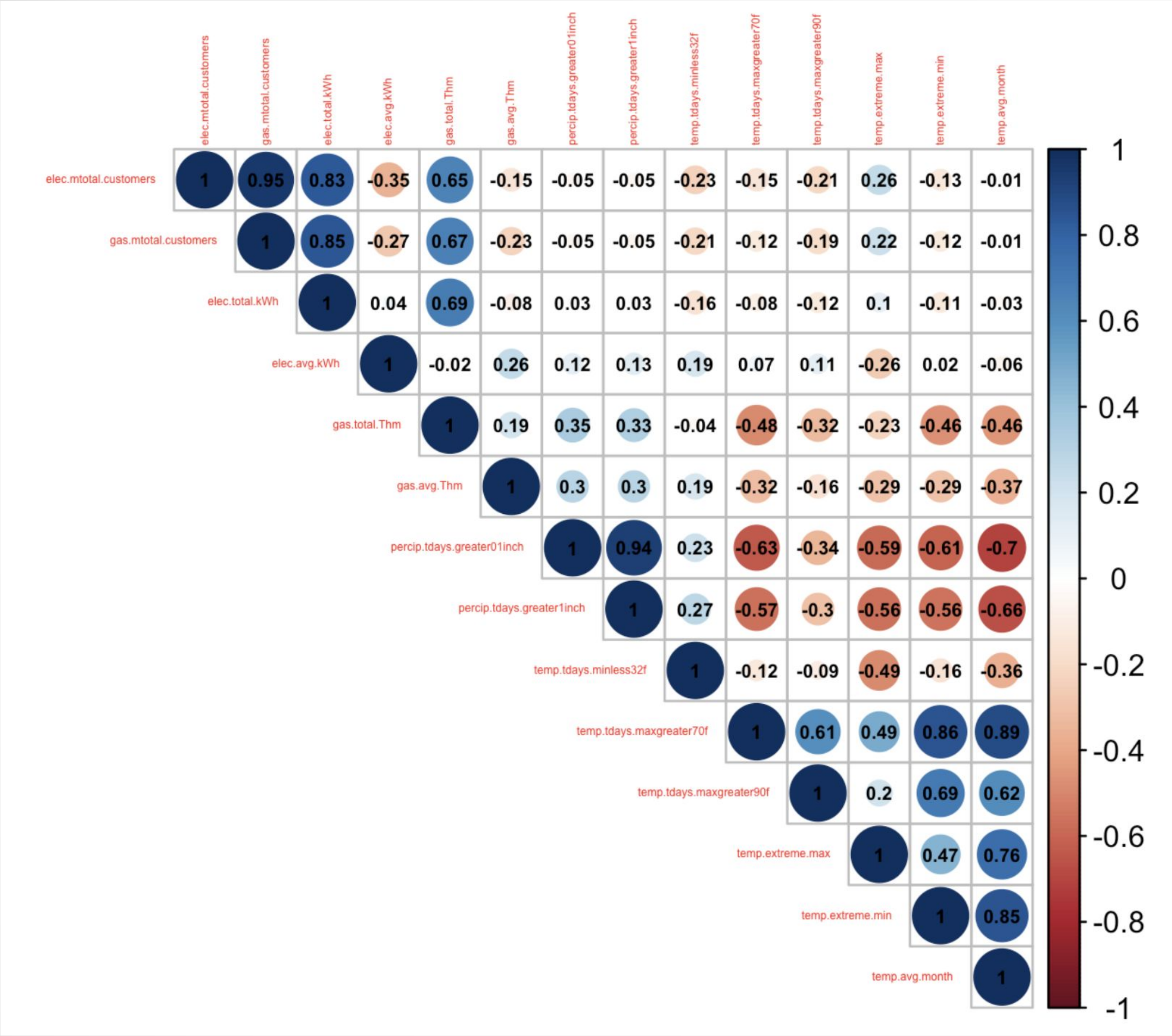
Thank you!

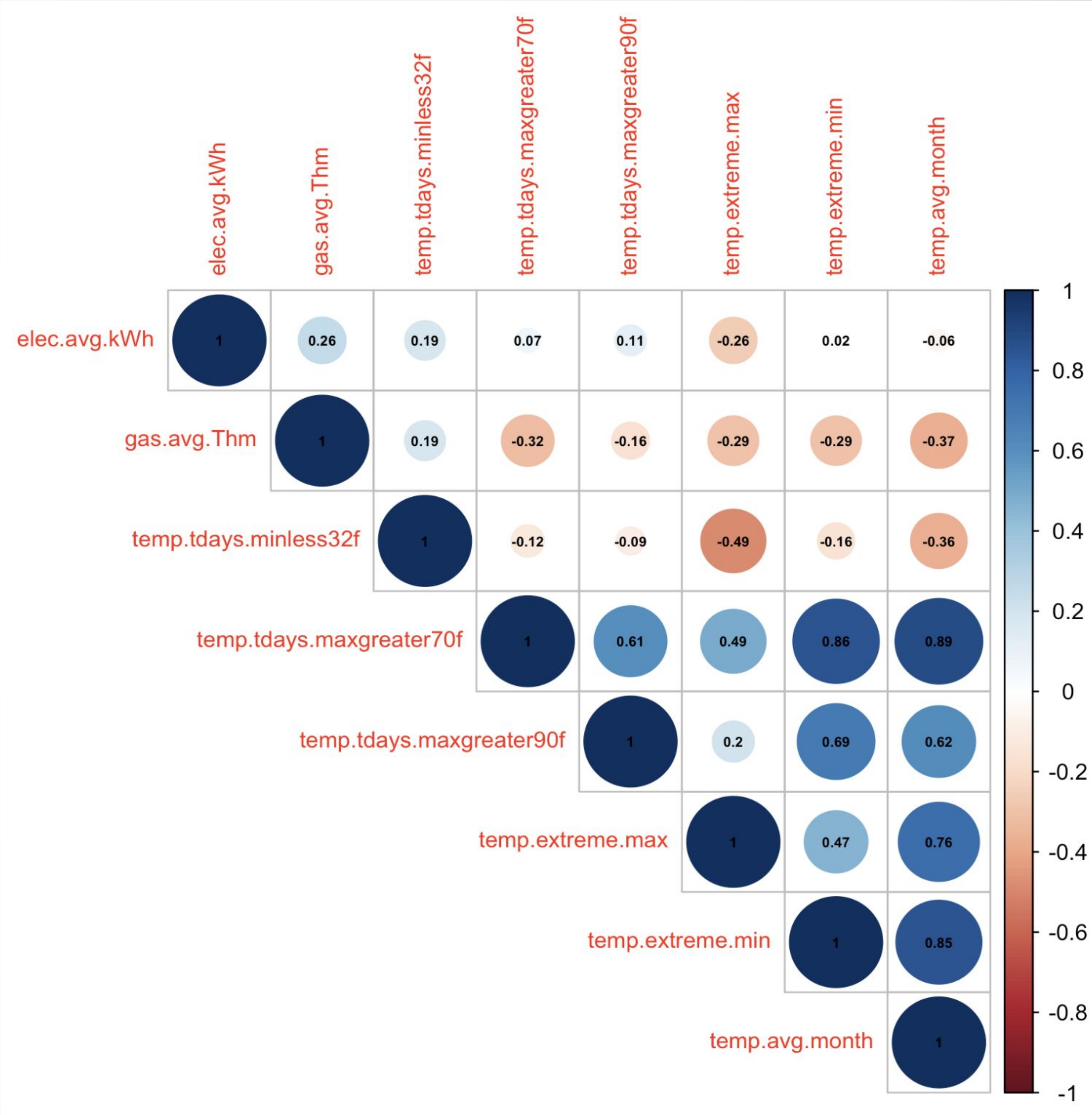
Breakdown of Team Responsibilities

Team Member	Contribution
Melanie Herscher	- Data Sourcing (Utilities) - Data Engineering and Transformation (Utilities & Optimization for Listings) - Time Series Exploratory Data Analysis (Utilities) - Initial Classical Time Series Modelling Analysis (Utilities) - Exploratory Modelling Analysis (Listings) - Project Management - Presentation Creation & Initial Script Generation - Project iSchool Page Write-up & Posting
Yuna Kim	- Exploratory Data Analysis (Utilities) - Initial ML Time Series Modelling Analysis (Utilities) - Advanced ML Modelling Analysis (Utilities) - Exploratory Modelling Analysis (Listings) - Final Modelling Analysis (Listings) - Final Presentation Model Visualizations (Utilities & Listings)
Nitin Swarup Sokhey	- Surveying Potential Buyers - Final Demo Design & Implementation
Preethi Raju	- Industry Expert Interviews - Data Sourcing (Listings) - Initial Data Engineering and Transformation (Listings) - Exploratory Data Analysis (Listings) - Initial Modelling Analysis (Listings)

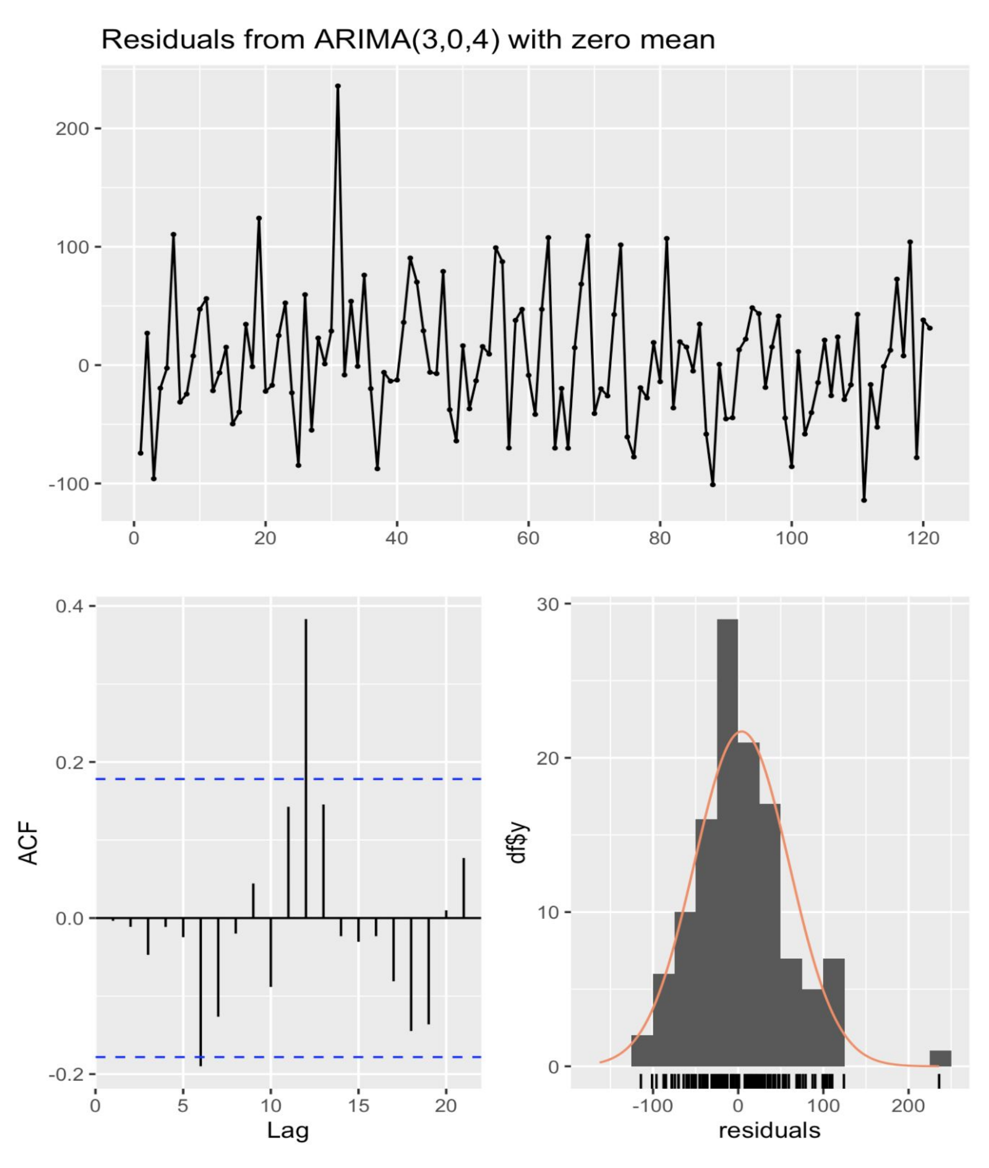
Appendix

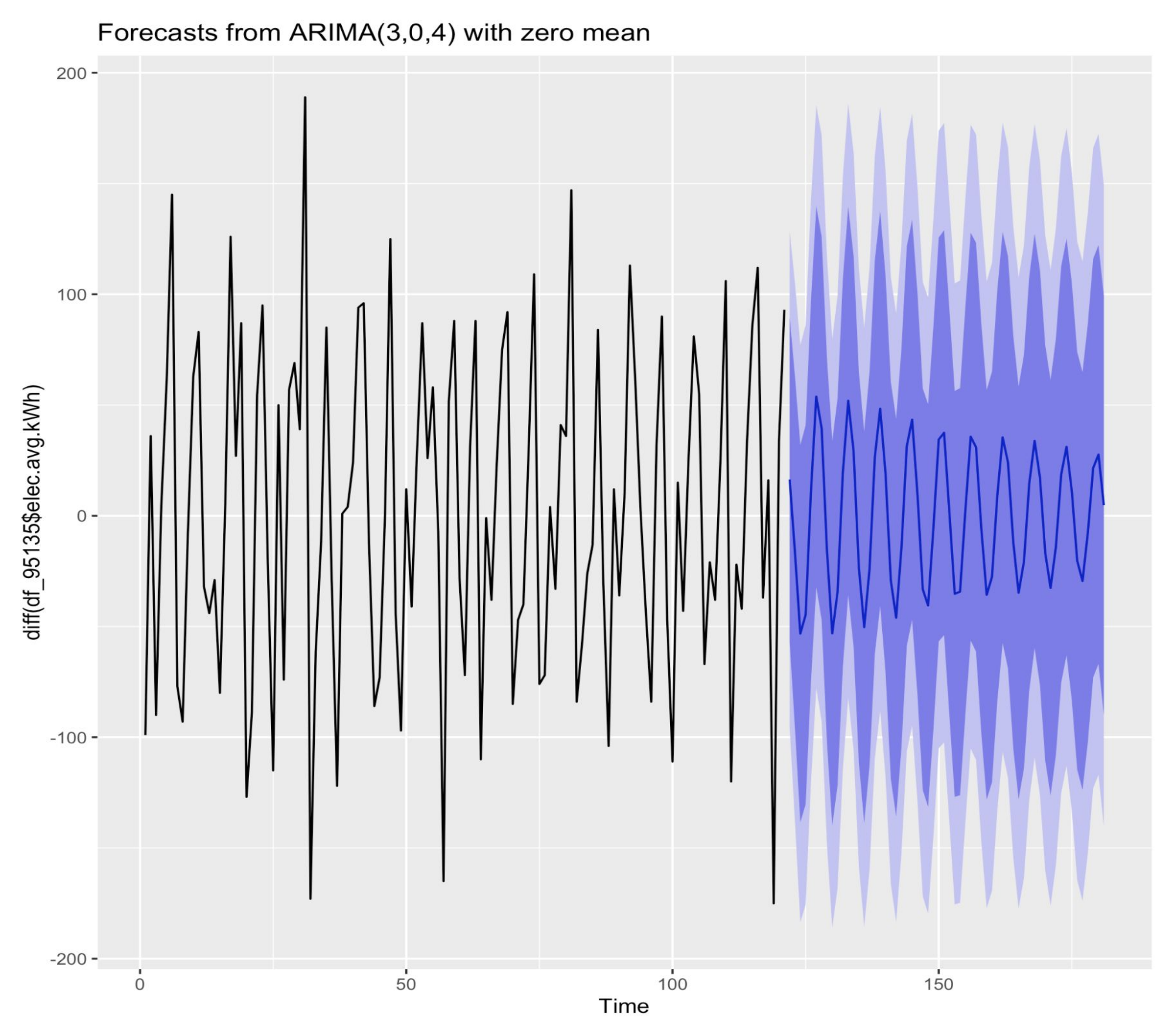
Correlation Plots for Temperature Against Usage



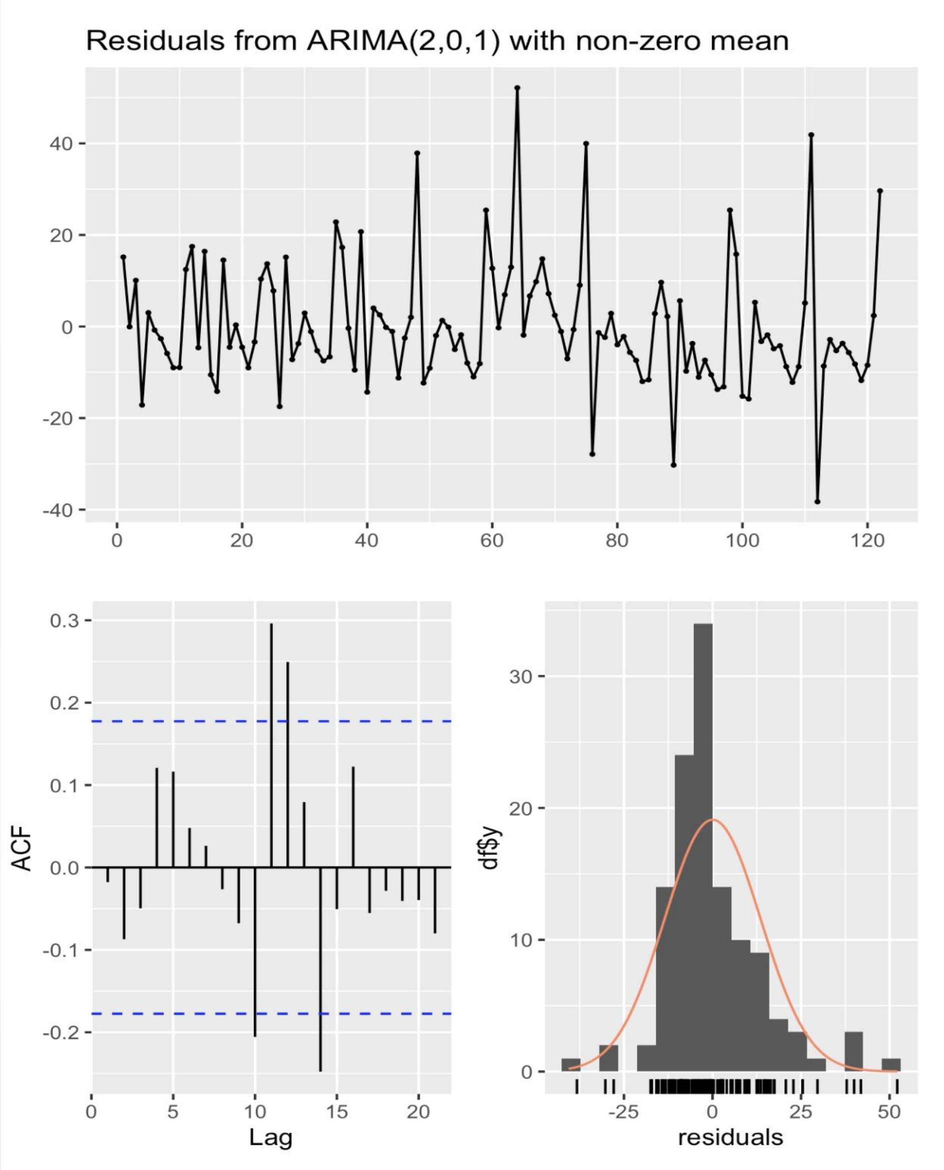


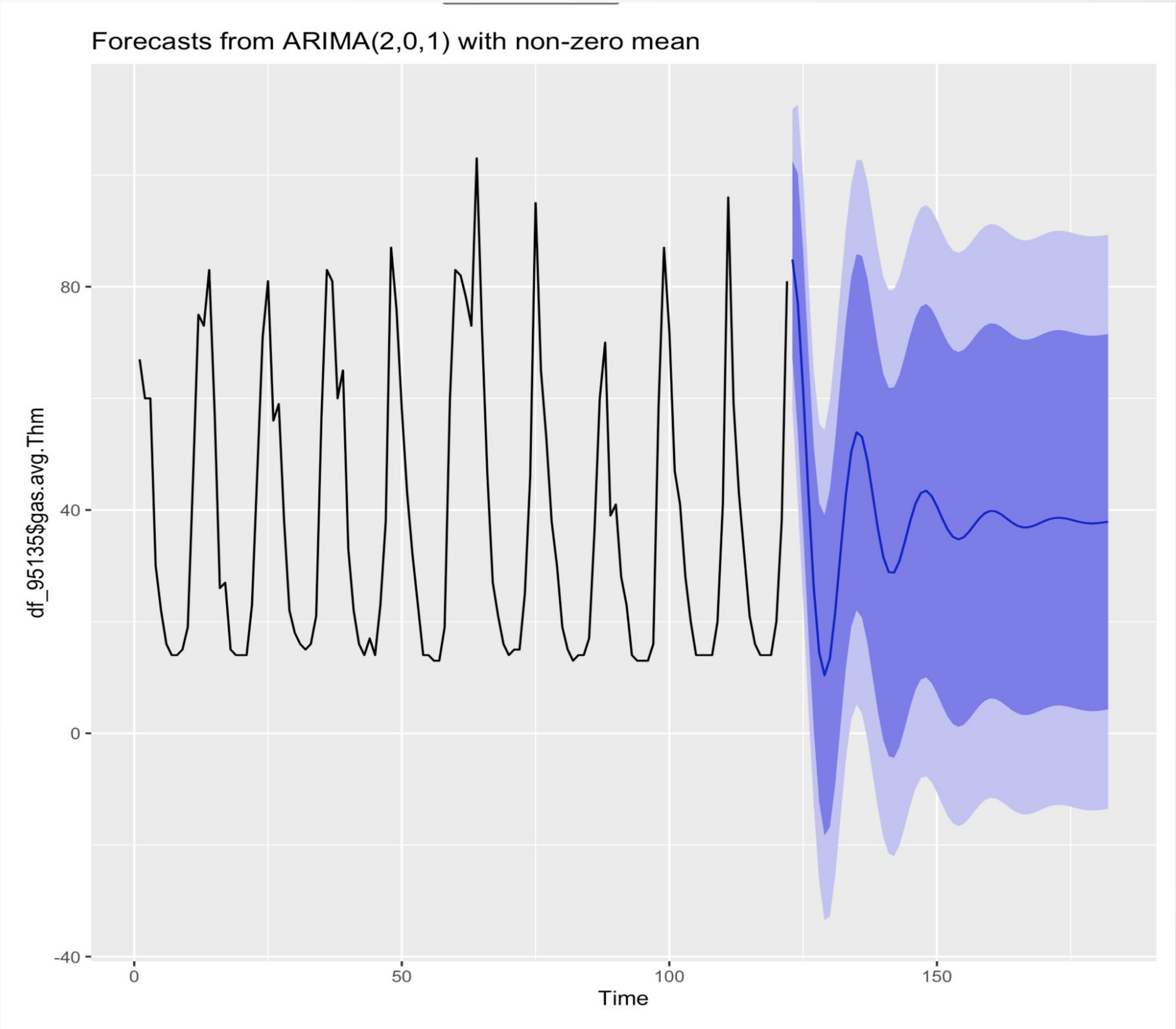
Baseline Modelling - Zip Code Specific Approach (Elec)



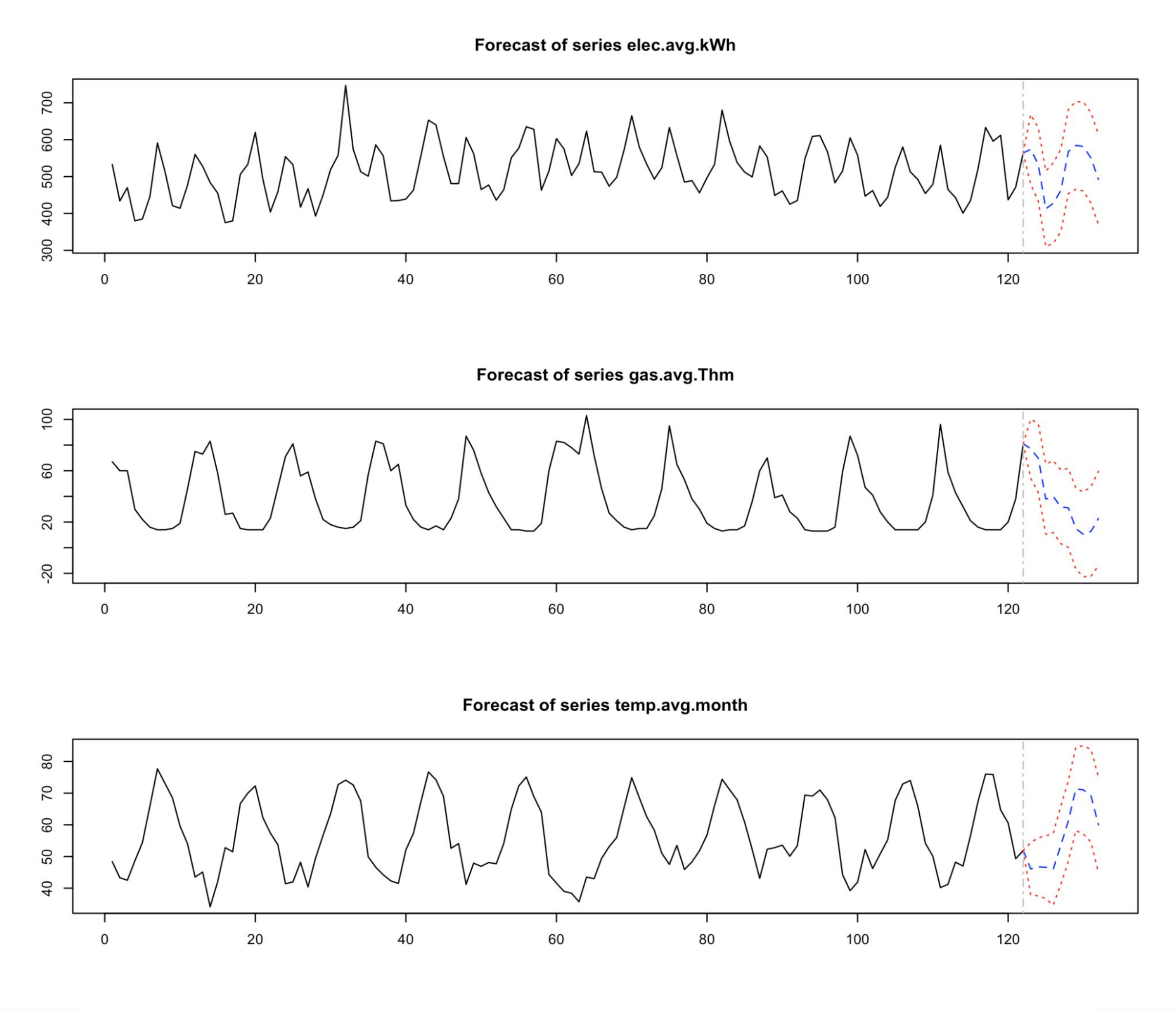


Baseline Modelling - Zip Code Specific Approach (Gas)

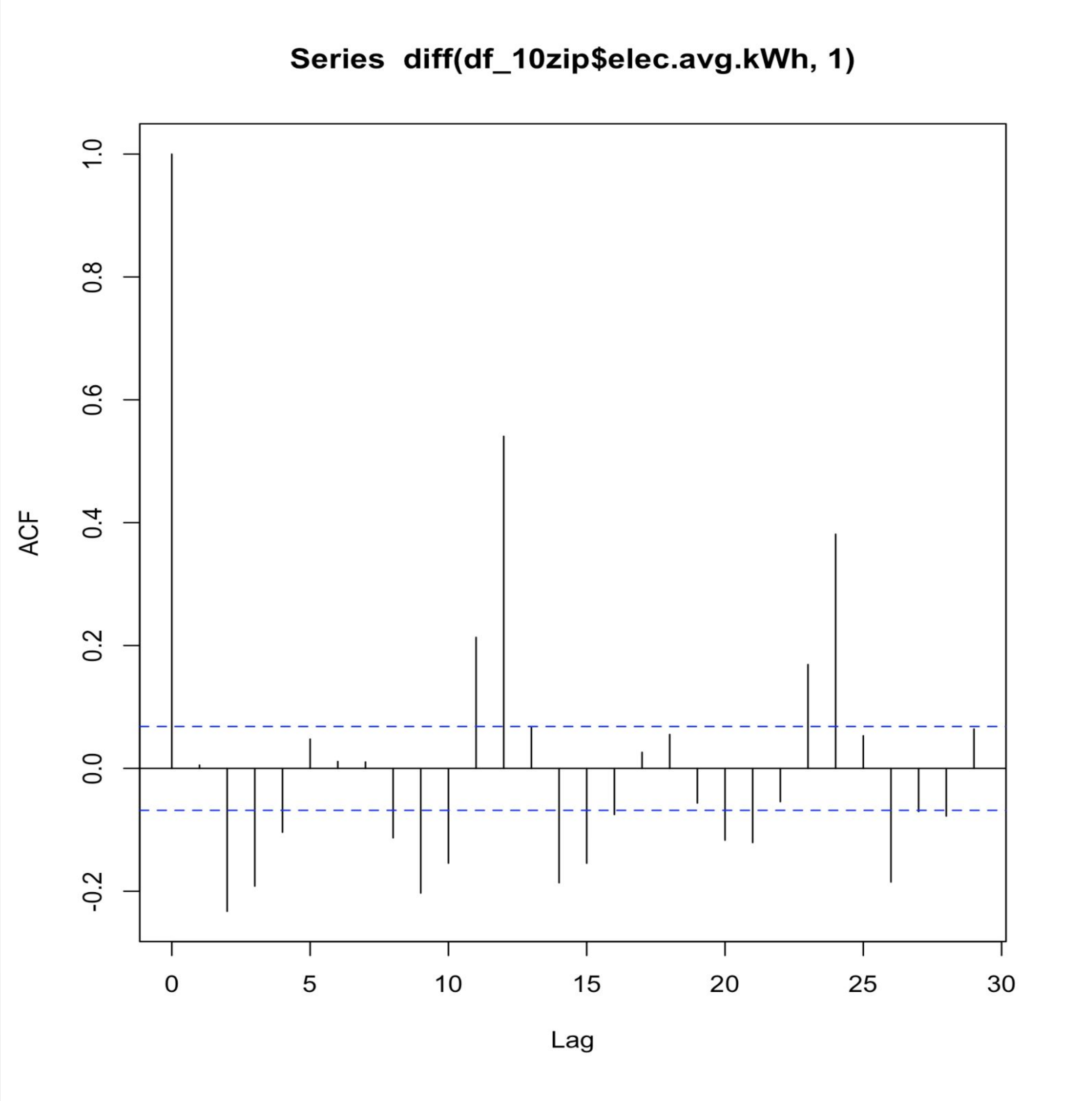


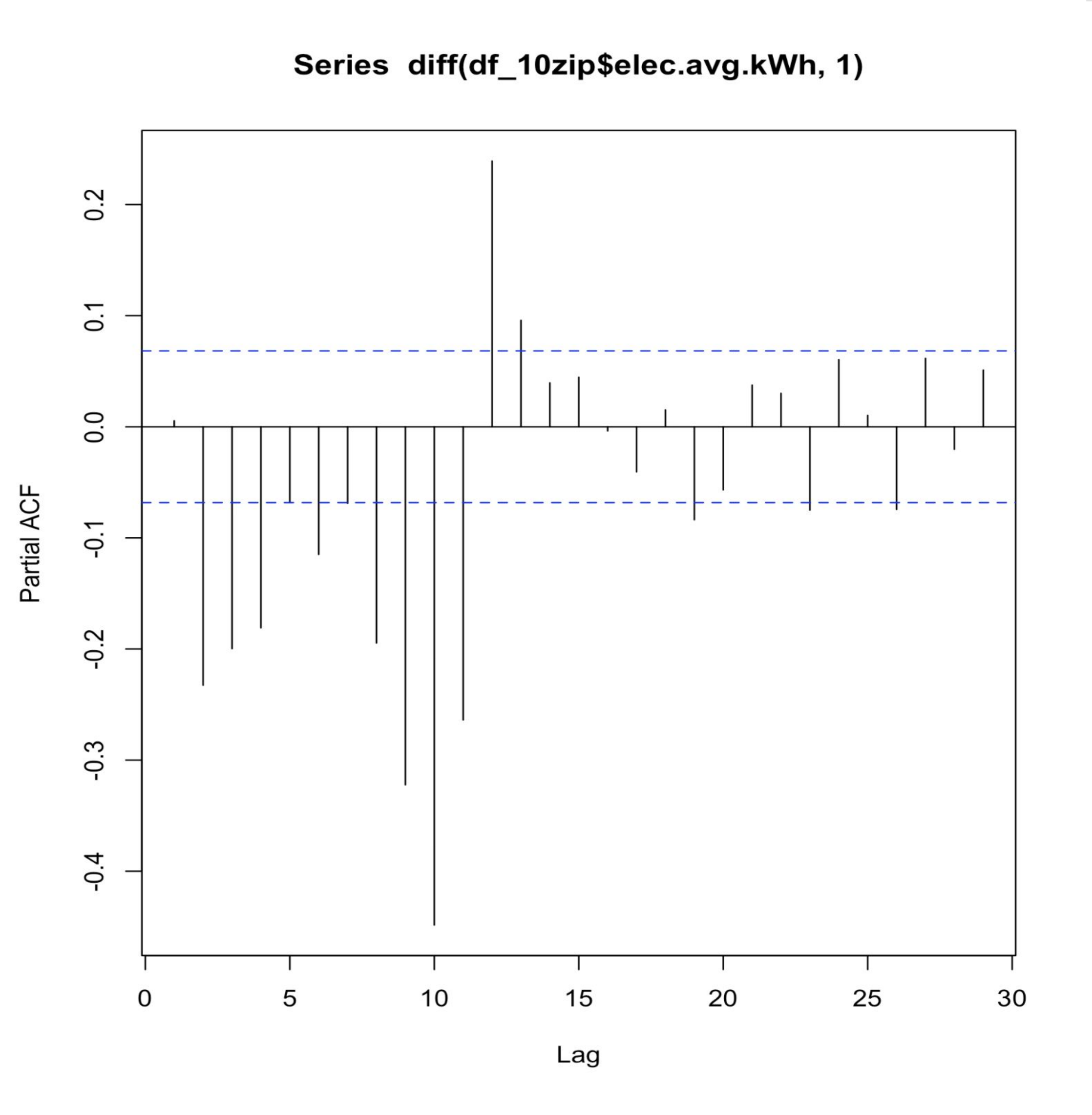


Advanced Modelling - VARs with Temperature

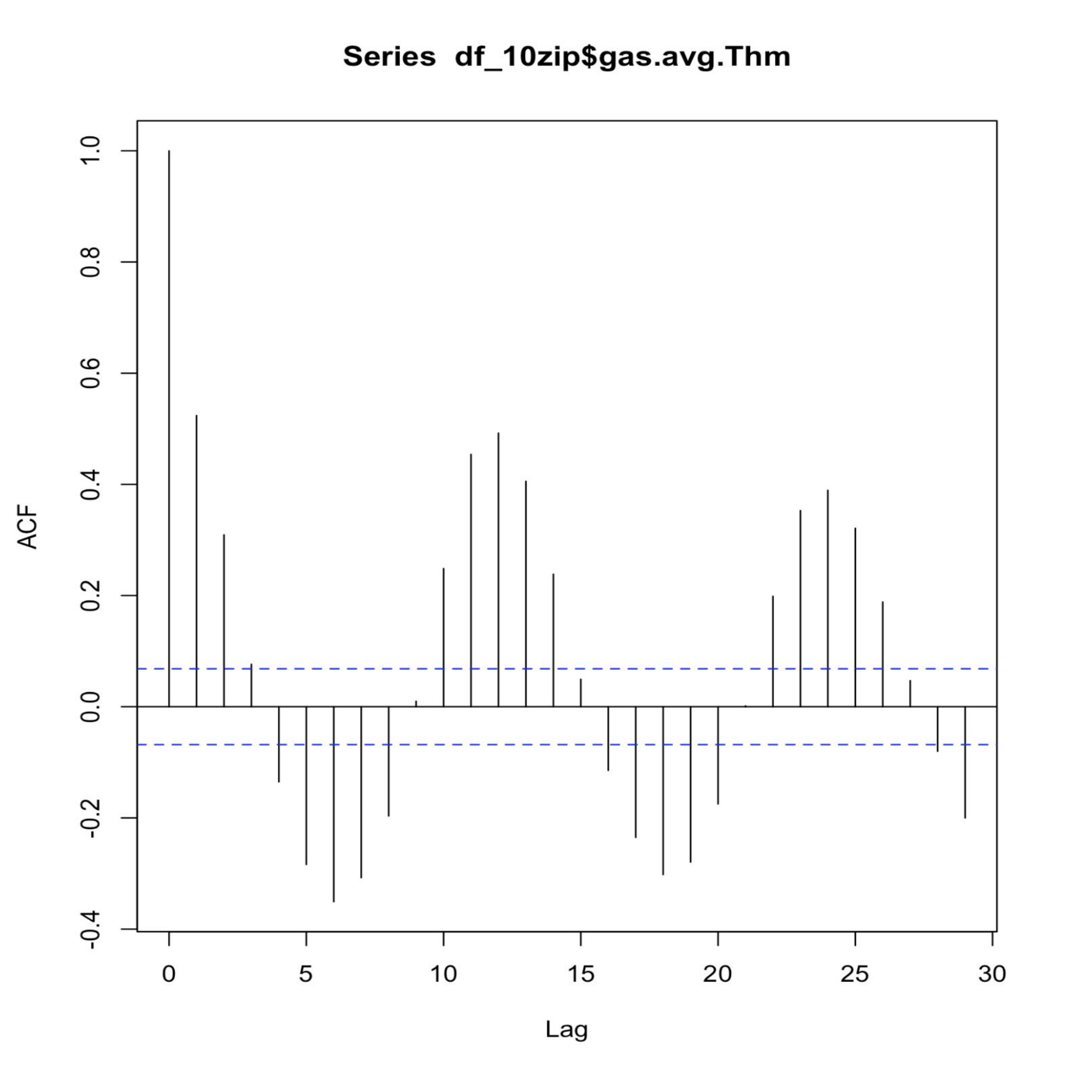


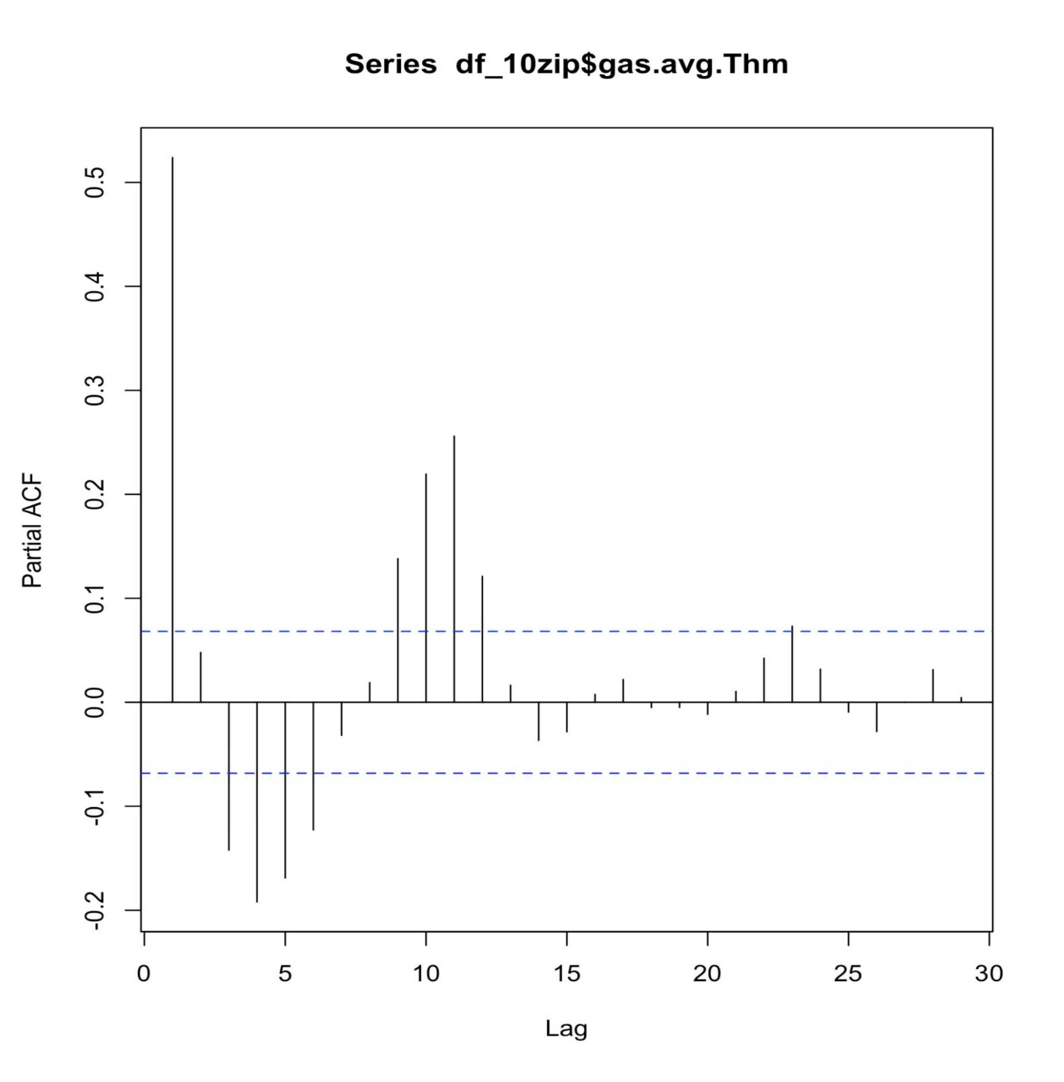
Autocorrelation & Partial Autocorrelations (Elec)



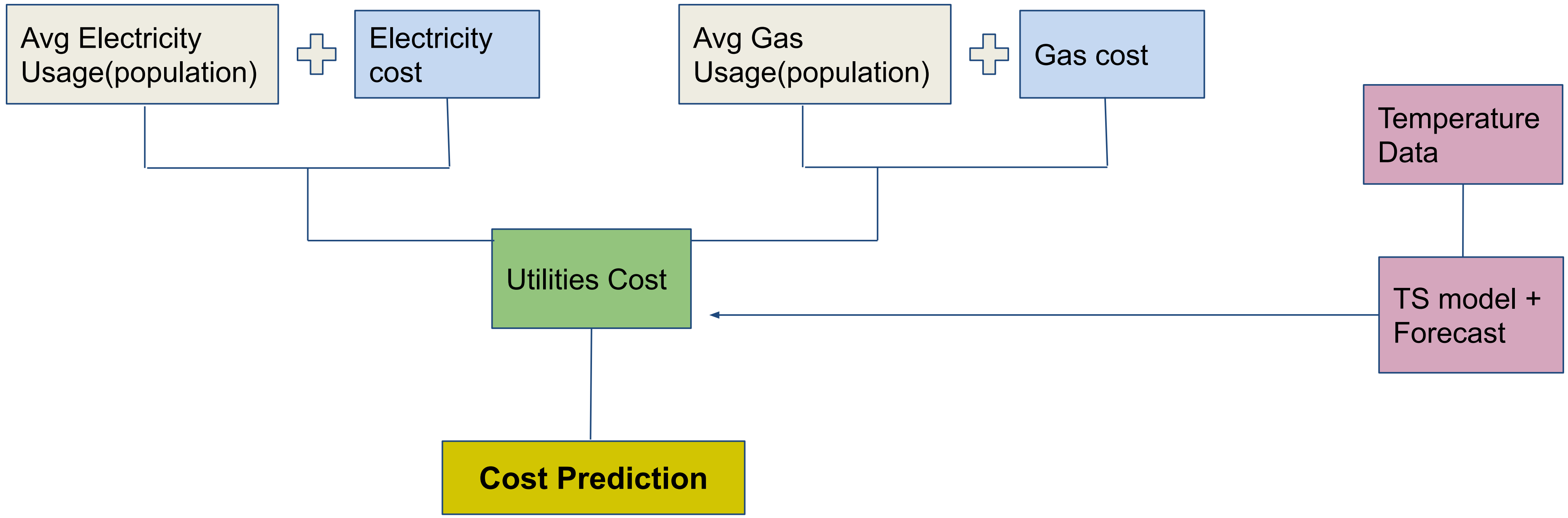


Autocorrelation & Partial Autocorrelations (Gas)

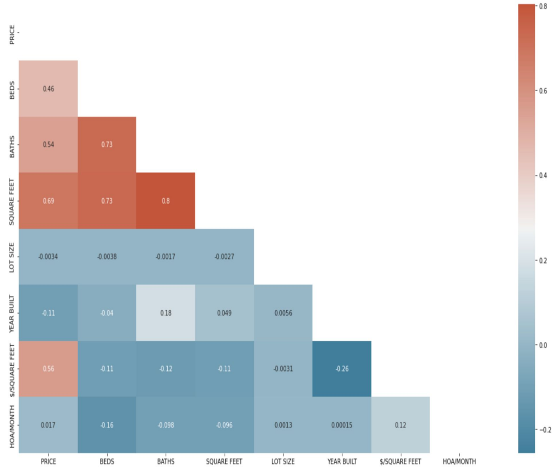




Utilites Cost Calculation Breakdown



House pricing Calculation Breakdown



94 zipcodes
126,587 Individual Houses

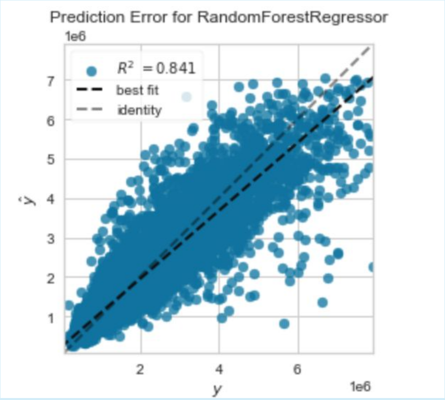
Filter:
7 features
100,838 individual houses

Regression

group by: each zipcodes

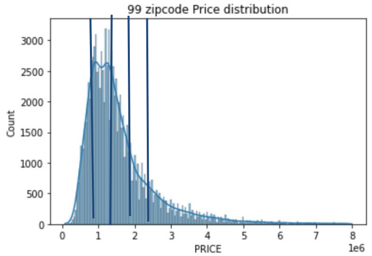
Train 80%

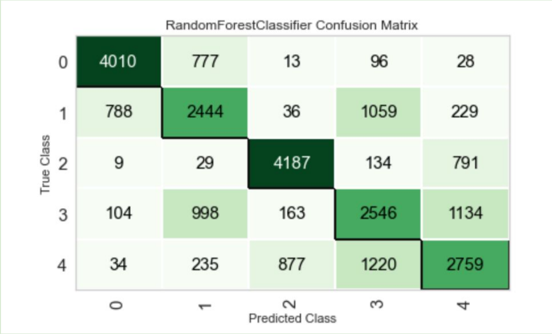
Test 20%



Classification

group by: 5 quantile by price

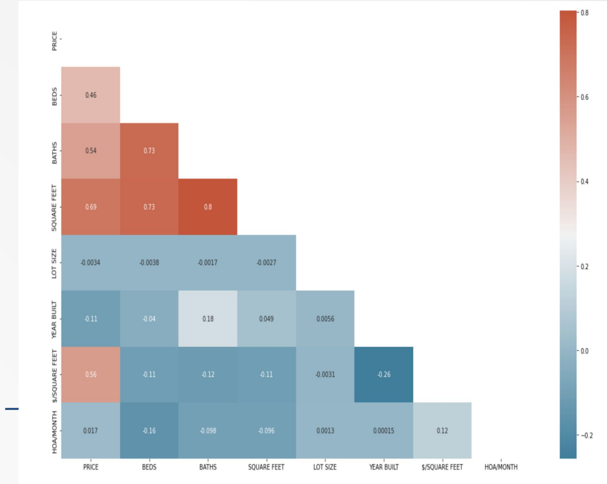




House price model Breakdown

94 zipcodes
126,587 Individual Houses

Filter:
7 features + zipcodes
100,838 individual houses



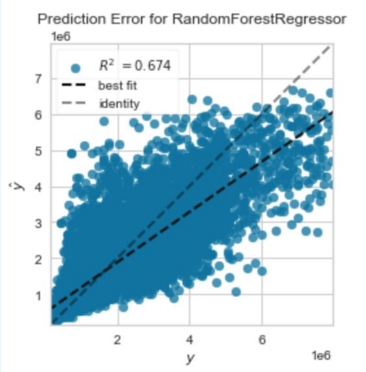
Regression

group by: each zipcodes

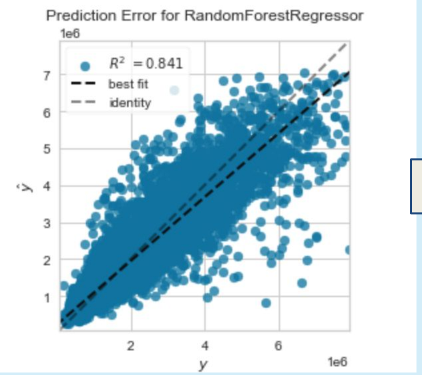
Train 80%

Test 20%

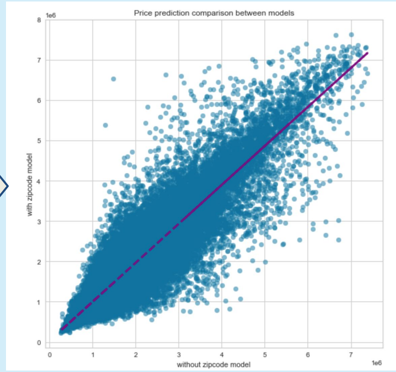
without Zipcode



with Zipcode

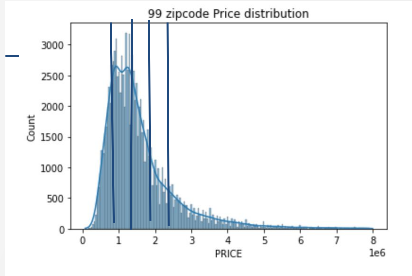


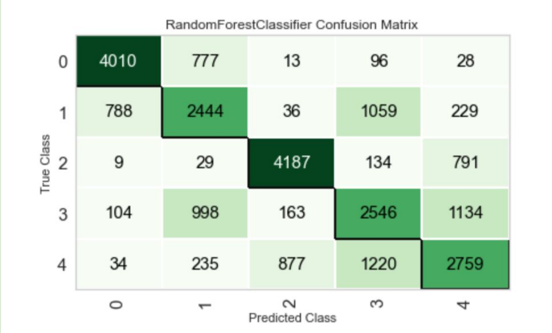
with vs without Z



Classification

group by: 5 quantile by price





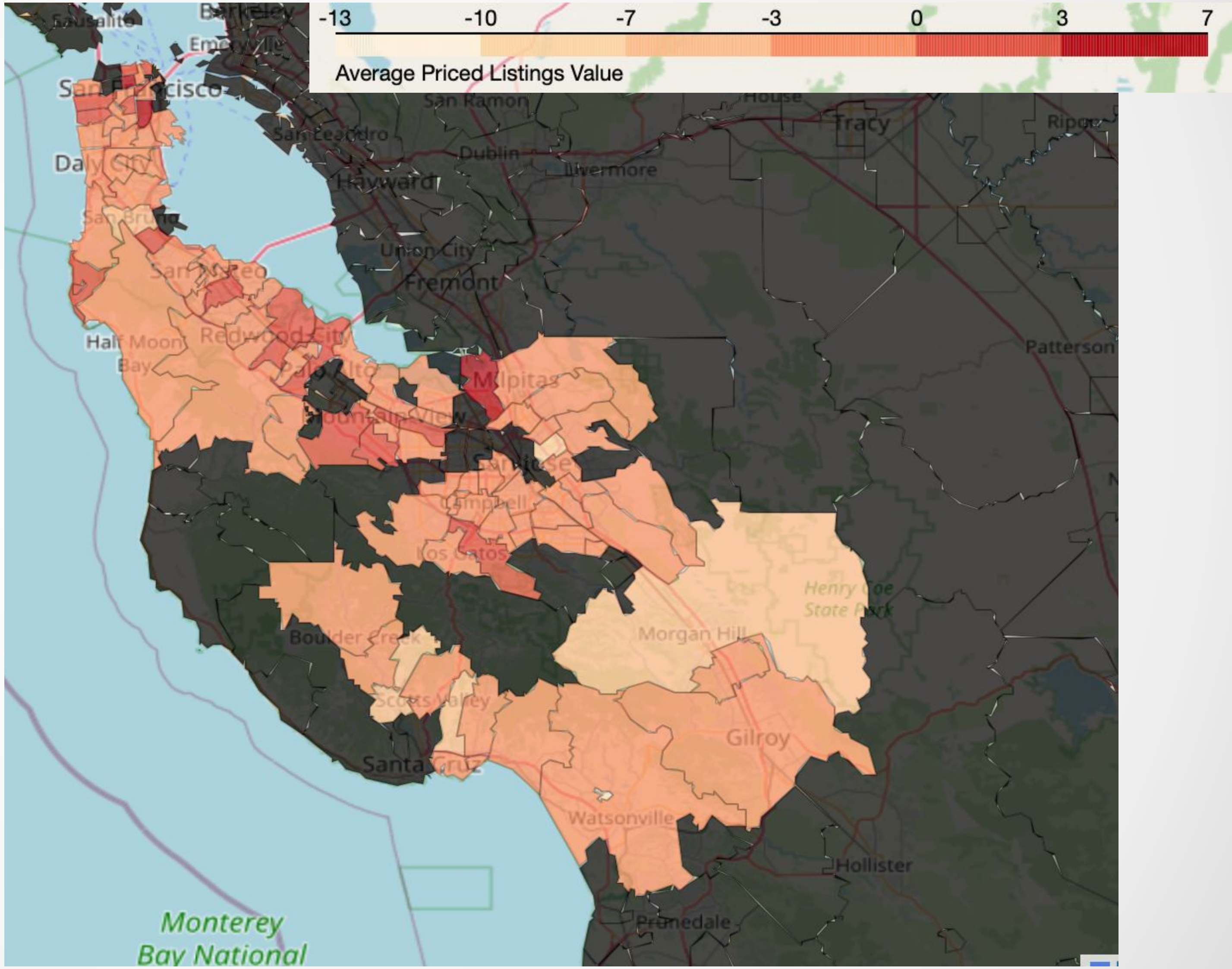
Average Percentage Difference

Actual - Prediction

Average percentage change delta

white: negative values underpriced

RED: positive values overpriced



Paratemetized for individual listings on Website

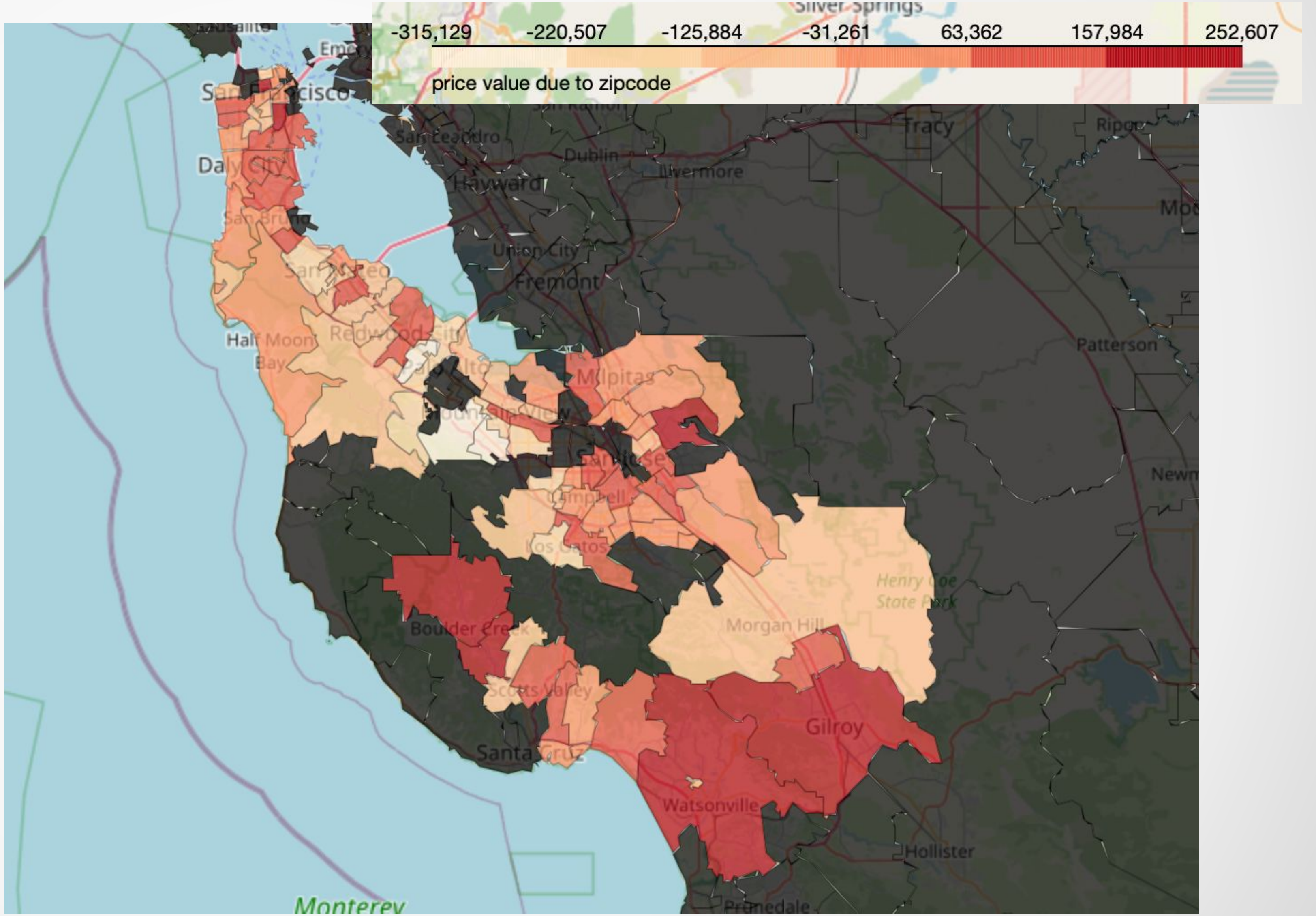
Zipcode Effect Model

House as is vs Zipcode Effect

Price predicted without zipcode - Price predicted with zipcode model

white: houses underpriced to zipcode (locational) cause than the features of the house itself

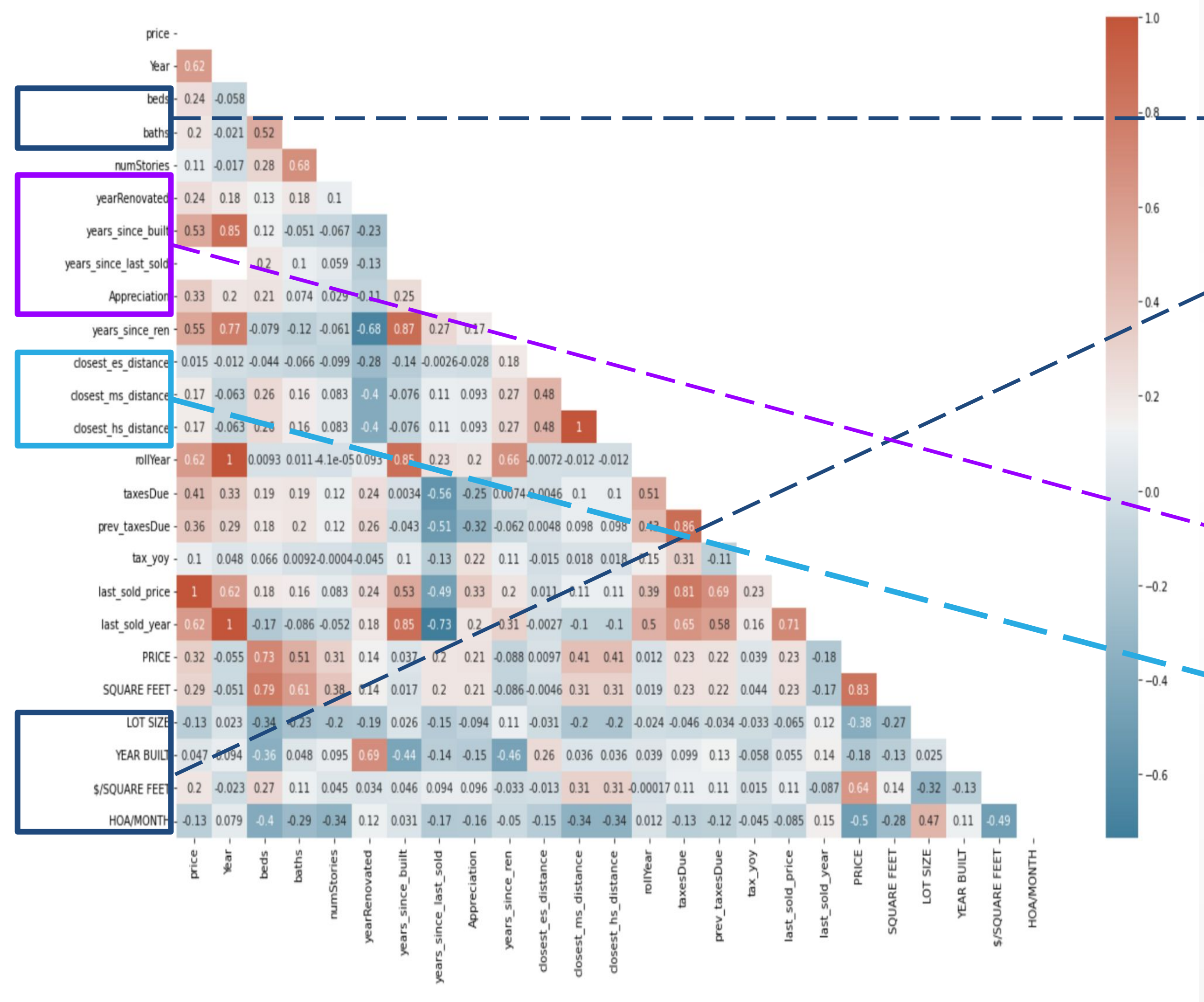
RED: houses overpriced due to zipcode (locational) cause than the features of the house itself



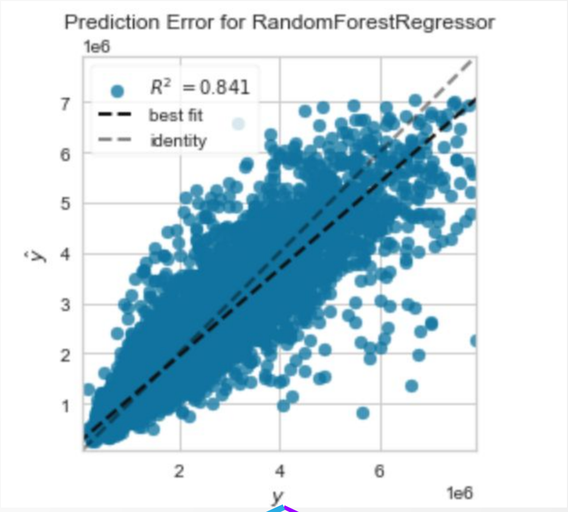
Future model plans: Customization

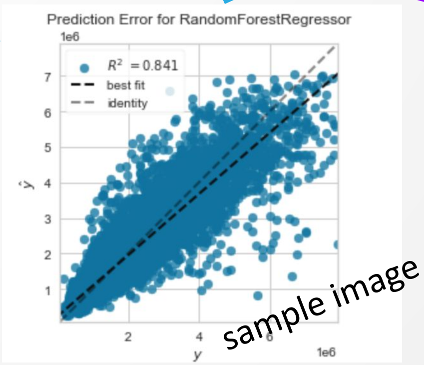
94 zipcodes
126,587 Individual Houses

Features:
25 features

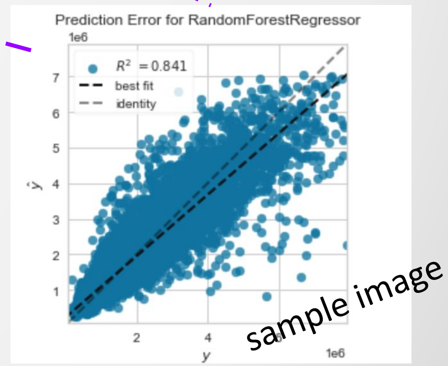


Base Model





Nearby school related Features



House Renovation Features

Utilities Prediction Full Modelling Evaluation List

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lightgbm	Light Gradient Boosting Machine	17.8752	2260.2701	42.4133	0.7546	0.1266	0.0749	0.2100
rf	Random Forest Regressor	18.3278	2262.1368	42.6986	0.7462	0.1343	0.0777	0.2360
dt	Decision Tree Regressor	22.1468	2867.2272	49.4165	0.6531	0.1649	0.0963	0.1680
gbr	Gradient Boosting Regressor	25.4198	3165.3358	52.4726	0.6336	0.1714	0.1109	0.1900
et	Extra Trees Regressor	25.0975	3366.8988	55.6102	0.5811	0.1719	0.1080	0.2320
knn	K Neighbors Regressor	26.5964	3615.6036	58.1512	0.5414	0.1828	0.1139	0.1660
ada	AdaBoost Regressor	37.5112	5139.4785	69.3230	0.3545	0.2398	0.1731	0.1700
br	Bayesian Ridge	44.7779	7674.2269	85.8472	-0.0063	0.3130	0.1954	0.1700
ridge	Ridge Regression	44.7822	7676.4847	85.8581	-0.0065	0.3131	0.1954	0.1660
lar	Least Angle Regression	44.7825	7676.6461	85.8589	-0.0065	0.3131	0.1954	0.1780
lr	Linear Regression	44.7825	7676.6461	85.8589	-0.0065	0.3131	0.1954	0.1800
huber	Huber Regressor	50.8121	7590.8207	85.7223	-0.0128	0.3226	0.2536	0.1820
lasso	Lasso Regression	50.8291	7917.9244	87.2857	-0.0419	0.3298	0.2438	0.1640
en	Elastic Net	50.8287	7917.9756	87.2860	-0.0419	0.3298	0.2438	0.1660
llar	Lasso Least Angle Regression	50.8291	7917.9244	87.2857	-0.0419	0.3298	0.2438	0.1800
dummy	Dummy Regressor	50.8290	7917.9005	87.2856	-0.0419	0.3298	0.2438	0.1600
omp	Orthogonal Matching Pursuit	50.3881	8014.9476	87.8201	-0.0549	0.3306	0.2393	0.1780
par	Passive Aggressive Regressor	175.8288	38592.5676	195.8245	-4.5574	2.3152	0.8983	0.1720

Video: Utilites Predicitons Through Time Interactive