**House Price**

**REPRICE**

**Written Requirements**

**Revision History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Authors** | **Description of Change** | **Sections** | **Rev** | **Date** |
| Mohammad H., Kunal M., and Don F. | Initial Release | All | 0 | 3/20/2019 |
| Mohammad H. | Train and test error values added to the document | Modeling | 1 | 4/24/2019 |

**Table of Contents**

[1 Team Description 4](#_Toc3971113)

[2 Terminology 5](#_Toc3971114)

[3 REPRICE 6](#_Toc3971115)

[3.1 Overview 6](#_Toc3971116)

[3.1.1 Data Extraction 6](#_Toc3971117)

[3.1.2 Modeling 6](#_Toc3971118)

[3.1.3 User Interaction 8](#_Toc3971119)

# Team Description

|  |  |
| --- | --- |
| **Team Member Name** | **Email Address** |
| Mohammadreza Hajy Heydary | Mheydary@csu.fullerton.edu |
| Kunal Matthews | Amos-m2345@csu.fullerton.edu |
| Don Feng | donfeng97@csu.fullerton.edu |

# Terminology

The following table defined terms used within this document.

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Predictor | An interface class designed to simplify the prediction process and hide modeling details |
| State-Level Prediction | A point estimate of the current median property value for a given state using the past *m* observations in the state timeseries |
| County-Level Prediction | A point estimate of the current median property value for a given county using the past *m* observations in the county timeseries and current prediction of state median values |
| City-Level Prediction | A point estimate of the current median property value for a given city using the past *m* observations in the city timeseries and current prediction of county median values |
| Neighborhood Level Prediction | A point estimate of the current median property value for a given neighborhood using the past *m* observations in the neighborhood timeseries and current prediction of zip-code median values |
| Stored Database | Database containing the stored house record. |

# REPRICE

## Overview

REPRICE is a housing value predictor software which allows users to search for housing prices based geographical criteria.

### Data Extraction

The following describes the data management systems requirements

The database shall retrieve data from Zillow’s website at the end of every week RID-1001

If the data manager fails to retrieve the records, it shall display the message “Out dated prediction warning!” RID-1002

Given an input, the data manager shall check if it exists in the database RID-1003

Given *m*, the data manager shall return the vector of past m true observations RID-1004

### Modeling

The following section details the requirements of the modeling process.

The train-test phase development indicated that all the proposed models are capable of recalling and predicting values with a maximum error of $3.0. Thus, *Etrain\_max* = *Etest\_max* = 3.0.

#### Predictor

If the data management fails to retrieve the records, then the error message “Data Retrieval Failure” shall be displayed RID-2001

Predictor shall be able to predict the train-set entries with a maximum error of *Etrain\_max* RID-2002

Predictor shall be able to predict the test-set entries with a maximum error of *Etest\_max*  RID-2003

Predictor shall scale the final prediction based on the property square footage provided by the user RID-2004

#### State-Level Predictor

State-Level Predictor shall be able to predict the median property value at state-level for a random subset of entries from the train-set with a maximum error of *Etrain\_max* RID-3001

State-Level Predictor shall be able to predict the median property value at state-level for a random subset of entries from the test-set with a maximum error of *Etest\_max*  RID-3002

#### County-Level Predictor

County-Level Predictor shall be able to predict the median property value at county-level for a random subset of entries from the train-set with a maximum error of *Etrain\_max* RID-4001

County-Level Predictor shall be able to predict the median property value at county-level for a random subset of entries from the test-set with a maximum error of *Etest\_max*  RID-4002

#### City-Level Predictor

City-Level Predictor shall be able to predict the median property value at city-level for a random subset of entries from the train-set with a maximum error of *Etrain\_max* RID-5001

City-Level Predictor shall be able to predict the median property value at city-level for a random subset of entries from the test-set with a maximum error of *Etest\_max*  RID-5002

#### Neighborhood-Level Predictor

Neighborhood-Level Predictor shall be able to predict the median property value at neighborhood-level for a random subset of entries from the train-set with a maximum error of *Etrain\_max* RID-6001

Neighborhood-Level Predictor shall be able to predict the median property value at neighborhood-level for a random subset of entries from the test-set with a maximum error of *Etest\_max*  RID-6002

### User Interaction

The following describes the user input and the resulting output

#### User Interface

The interface shall require user to input state, county, city, zip and neighborhood in a valid range, which will be used by the prediction model to output a result. RID-7001

IF user does not enter input for all fields THEN the system shall return an error RID-7002

#### User Input

The user shall only input variables that will be checked against the data that is in our dataset and if found the predictor will output a result. The data should exist in the dataset or error will be displayed. RID-7003

#### System output

Results from the Predictor shall be displayed to the user. If the dataset does not exist or other errors occur, the output will display an error and ask another set of inputs form the user. RID-7004