Learning regression model to predict house price

...

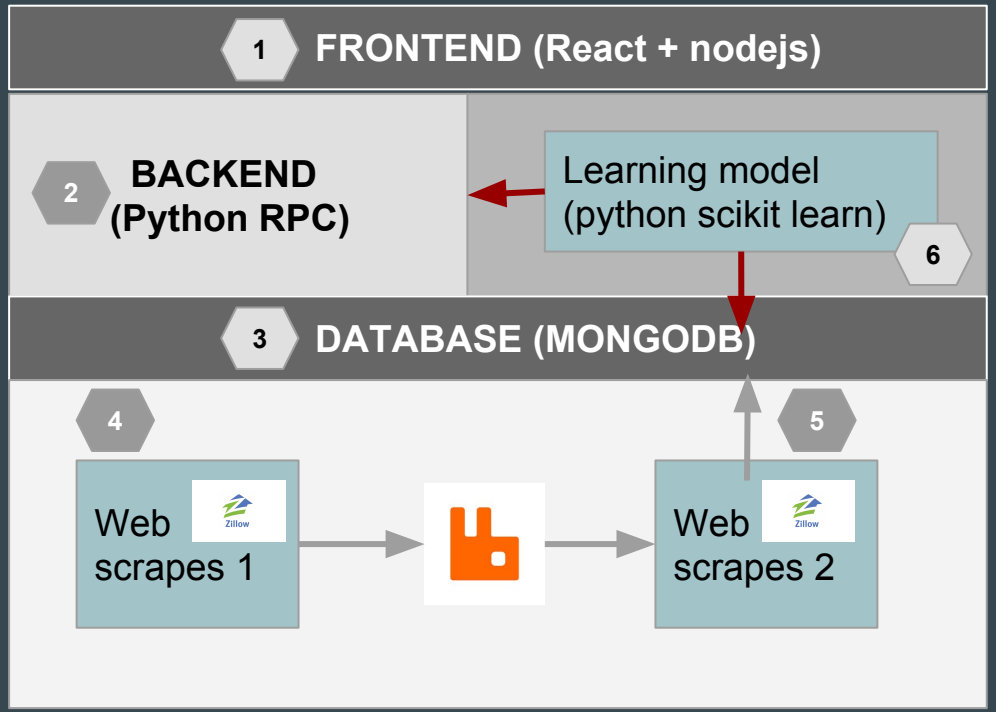
Capstone CS503
Larissa Liu
2017-11

Overview

This application will help sellers to predict the sale price of their house based on those house-level features they provided. Seller will also be able to add/update certain house renovation ideas there to see how much those renovation could help on improving the sale price.

In order to achieve a high level prediction accuracy, the application will scrape Zillow.com, an online real estate database to extract real estate listings available. This real estate scraper will extract details of property listings based on zip code. (for demo purpose, it will only focus on real estate market at Dallas, TX). this application will also use linear regression , one of most popular technique , to implement the learning model.

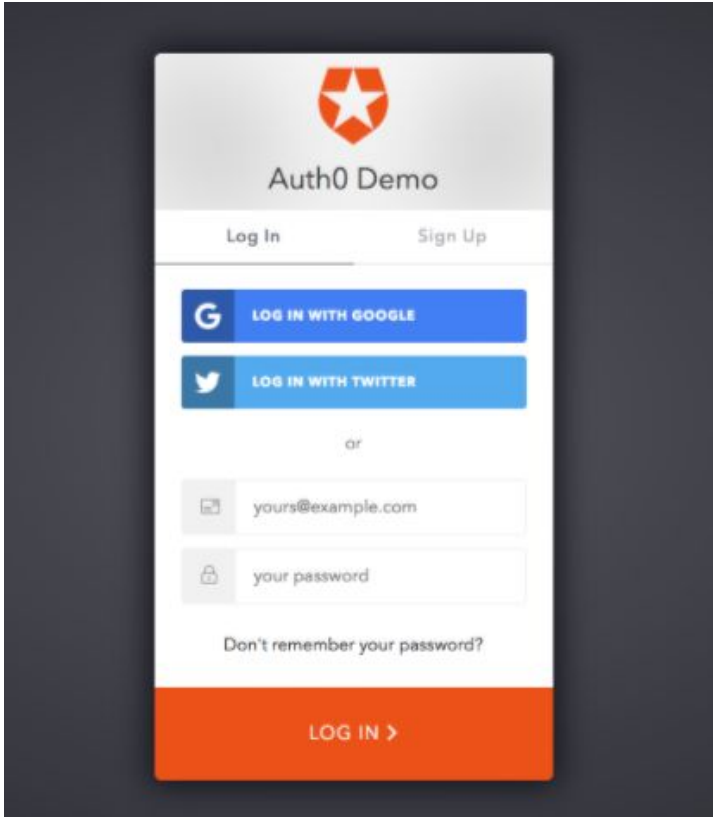
Architecture



1. Use react and nodejs to host a simple web form and house market metrics. It will use auth0 as Identity provider to enable social login
2. Provide a backend with rpc apis for storing data into DB and return predicated sale price.
3. Use mongoDB to store historical house-level features and sale price
4. Scrape zillow.com listing page to get list of urls based on zip code.
5. Scrape zillow.com detail page to get house features and sale price.
6. Do linear regression using numpy, scipy, stats model and scikit learn

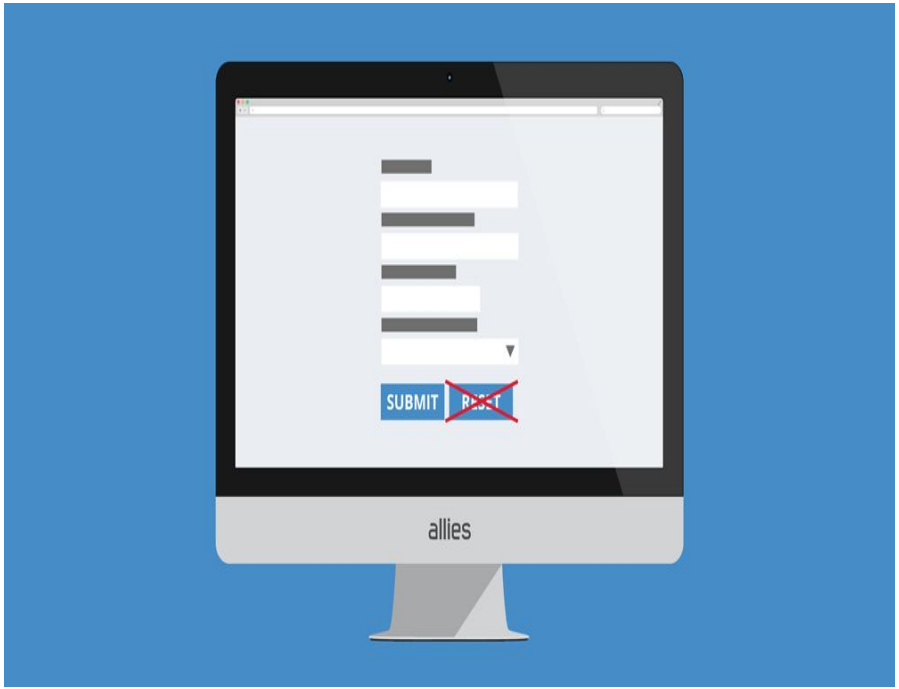
Social login

User have to login first using social credential (FB, Google, Twitter etc.)



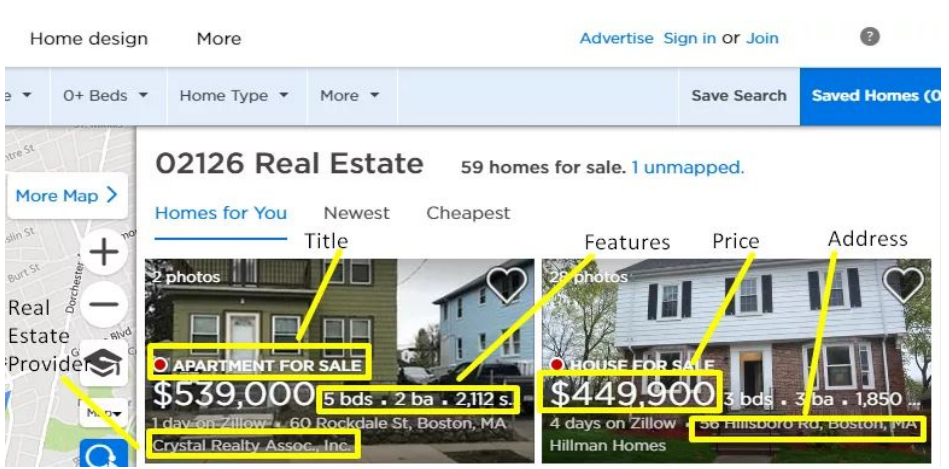
Add/update house features

User can fill/update a web form in order to let the application help on predicating the house price



Scrape zillow listing page

a. Using python requests to make requests and download the HTML content of the page
b. Using python LXML, for parsing the HTML Tree Structure using Xpaths
c. Populate house basic feature and urls into Rabbitmq



Title	Address	City	State	Zip Code	Price	Facts and Features	Real Estate Provider	URL
House For Sale	45 Faunce Rd	Boston	MA	2126	$349,000	3 bds , 2 ba , 956 sqft	Pondside Realty	https://www.zillow.com/homedetails/45-Faunce-Rd-Boston-MA-02126/59137709_zpid/
House For Sale	60 Temple St	Boston	MA	2126	$540,000	4 bds , 2 ba , 1,696 sqft	Gibson Sotheby's International Realty	https://www.zillow.com/homedetails/60-Temple-St-Boston-MA-02126/59138896_zpid/
House For Sale	29 Violante St	Boston	MA	2126	$259,000	3 bds , 2 ba , 1,632 sqft		https://www.zillow.com/homedetails/29-Violante-St-Boston-MA-02126/59139388_zpid/
Condo For Sale	53 River St	Boston	MA	2126	$529,000	3 bds , 3 ba , 1,589 sqft	The Galvin Group, LLC	https://www.zillow.com/homedetails/53-River-St-Boston-MA-02126/2092600080_zpid/
Condo For Sale	6 Temple St	Boston	MA	2126	$659,000	3 bds , 3 ba , 1,697 sqft	The Galvin Group, LLC	https://www.zillow.com/homedetails/6-Temple-St-Boston-MA-02126/2124803167_zpid/
Condo For Sale	15 Groveland St # 3	Boston	MA	2126	$349,900	3 bds , 1 ba , 830 sqft	Newbury Street Real Estate	https://www.zillow.com/homedetails/15-Groveland-St-3-Boston-MA-02126/2092921175
House For Sale	20 Almont St	Boston	MA	2126	$550,000	8 bds , 3 ba , 3,770 sqft		https://www.zillow.com/homedetails/20-Almont-St-Boston-MA-02126/59138113_zpid/
House For Sale	21 Kennebec St	Boston	MA	2126	$420,000	2 bds , 2 ba , 908 sqft	Success! Real Estate	https://www.zillow.com/homedetails/21-Kennebec-St-Boston-MA-02126/59138042_zpid/
Apartment For Sale	40 Tennis Rd	Boston	MA	2126	$629,900	6 bds , 2 ba , 4,452 sqft	M7 Real Estate Group	https://www.zillow.com/homedetails/40-Tennis-Rd-Boston-MA-02126/59137978_zpid/
House For Sale	3 Duxbury Rd	Boston	MA	2126	$475,000	4 bds , 1.5 ba , 1,296 sqft		https://www.zillow.com/homedetails/3-Duxbury-Rd-Boston-MA-02126/59137470_zpid/
Condo For Sale	51 River St	Mattapan	MA	2126	$599,000	3 bds , 3 ba , 1,485 sqft	The Galvin Group, LLC	https://www.zillow.com/homedetails/51-River-St-Mattapan-MA-02126/2092719719_zpi
Condo For Sale	51A River St	Boston	MA	2126	$579,000	3 bds , 3 ba , 1,491 sqft	The Galvin Group, LLC	https://www.zillow.com/homedetails/51A-River-St-Boston-MA-02126/2092760832_zpid/
Apartment For Sale	4-6 Greendale Rd	Boston	MA	2126	$599,900	6 bds , 3 ba , 2,496 sqft	WEICHERT, REALTORSÂ® - Briarwood Real Estate	https://www.zillow.com/homedetails/4-6-Greendale-Rd-Boston-MA-02126/2092835887_
Condo For Sale	27 Hosmer St APT 1	Boston	MA	2126	$199,000	2 bds , 1 ba , 896 sqft		https://www.zillow.com/homedetails/27-Hosmer-St-APT-1-Boston-MA-02126/81378796_
House For Sale	64 Tampa St	Boston	MA	2126	$230,000	1 bd , 1 ba , 729 sqft	Rudy & Associates REALTORSÂ®	https://www.zillow.com/homedetails/64-Tampa-St-Boston-MA-02126/2118499072_zpid/
House For Sale	151 Hebron St	Boston	MA	2126	$299,900	2 bds , 1 ba , 960 sqft	Pondside Realty	https://www.zillow.com/homedetails/151-Hebron-St-Boston-MA-02126/59138658_zpid/
Condo For Sale	622 Morton St # 1	Boston	MA	2126	$325,000	3 bds , 1 ba , 1,200 sqft	Alpha Realty	https://www.zillow.com/homedetails/622-Morton-St-1-Boston-MA-02126/2093444356_

Scrape zillow detail page

a. Using python requests to make requests and download the HTML content of the page
b. Using python LXML, for parsing the HTML Tree Structure using Xpaths
c. Transform categorized data type into integer.