

Visualize and Forecast House Value in US

Yarui Wang, Ke Zhang, Taoyouwei Gao



CSE6242 Project

About Instruction Contact

Visualize and Forecast House Value in US

Start Now

Summary

- As more foreign investors target on the US real estate market, there is an increasing need for them to learn more about the US market, to discover more places like Atlanta or Austin which has a relatively low investment cost but possesses large potential of earnings.
- Current popular online real estate websites like Zillow or Redfin do not provide general aggregated statistical information, while local real estate agents usually do not have the vision of the entire country.
- We can help. Through our website, foreign investors can get information progressively and find their most optimal investment location across the U.S

METHODOLOGY

Key Features

- D3 choropleth map visualization
- Two-level zooming functionality
- Interactive customizable filtering options
- Forecast future growth potential through ML algorithm
- Pop-out line chart showing historical data and predictive results

Data

Source

- Data:Zillow Research
- Choropleth map: US.json

Size

- 25MB raw data in CSV
- 200MB processed data in sqlite3

Preliminary Cleaning

- Filter raw data with Openrefine
- Regulate data with Python pandas package

Data Processing and Storage

Database designed based on frontend usecases

- Region (id, name, size_rank)
- HouseType (id, category, house_type)
- MedianHouseValueByMonth (regionId, housetypeld, time, price)

Model

Features:

- **Raw features:** RegionName, Median Sale Price, Median Listing Price, Median Listing price per sqft, Days on Zillow, Inventory, Percentage of Listings with pricecut, Pricecut percentage, sale count
- **Engineered features:** Median Sale price next month, sale price change %, growth potential

Experiments:

- **Models:**Linear Regression, Random Forest, SVM, Multilayer Perceptron
- Stepwise variable selection
- Hyper-parameter tuning

Test&Evaluation

Model Evaluation:

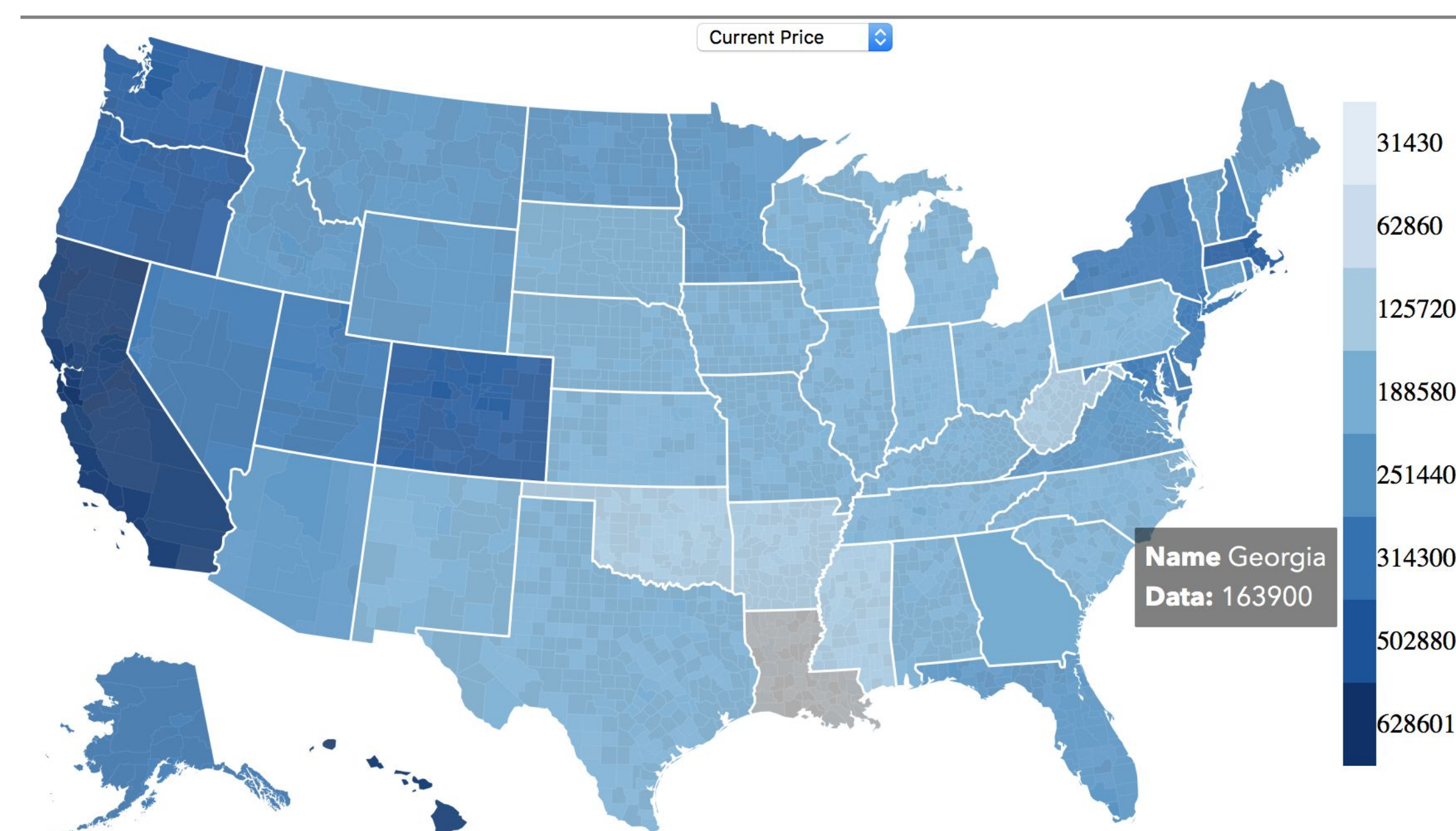
- Evaluation measures: testing accuracy, mean absolute error, coefficient of determination
- Best model and accuracy: Linear SVM, mean absolute error = 1500

Visualization Evaluations:

- AB testing on two workflows of frontend and compare functionality
- Functional tests and usability test
- *Heuristic evaluation:*
 - Current price map shows that house values in East and West coast are higher.
 - Growth potential map indicates midwest has larger growth potential.

VISUALIZATION

House price visualization in the USA



- Built on top of Flask framework
- Data fetched in real-time from backend database
- D3 choropleth map shows distribution of current price or prediction value. A tooltip will appear showing its name and value.
- When selecting a certain state or county, interactive selection boxes would emerge.The line chart could be changed as time, housetype and number of number vary.

