ISSN (Online) 2278-1021

IJARCCE



International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 10, October 2018

Introduce Website about Insights into Housing Markets in Viet Nam

Hoang Tuan Long

People's Police University - Ho Chi Minh City, Vietnam

Abstract: Nowadays, data has become one of the most important resources in the world, and according to some reports, data have even overtaken oil. In Vietnam, there are still many data resources which still have not been utilized fully such as real estate data. In recent years, people have witnessed a boom in the real estate market with a great number of property projects that can provide the middle class a house to live, and this can also boost the economy in Vietnam. However, despite hundreds of thousands real estate items are created every day, they are just considered as trivial properties posted on a website. Moreover, there are few organizations that are willing to make effort to find out an original approach for utilizing these data. Because of that, in this article, the author introduce a website that can take advantages of real estate sources in Vietnam and provided people who have business in the real estate market a detailed insight into the market.

Keywords: Data mining, data visualization, housing markets.

I. INTRODUCTION

Despite the global financial crisis in the period between 2007 and 2008, Vietnam has witnessed a reasonable increase in the real estate market recently. There are many reasons leading to this trend such as the increase of citizens' income, the rise of population in Vietnam, together with it the demand of homeownership, and the flow of foreign investment into Vietnamese real estate market. More specifically, the proclamation of Housing Law and Law on the Real Estate Business allowing the foreigners from various countries around the world to own, sell and transfer property legally has caused a boom in Vietnam real estate price [1]. Due to the hike of investment and real estate projects, customers now have many choices ranging from affordable to premium. Moreover, with the banking and financial supports from real estate companies, owning a house is not too difficult anymore and customers can have a desired house without paying a lot of money in advance. Despite the benefits, this also leads to a serious dilemma. There are a variety of information sources, but not all of them can be trusted. Today, most of the estate activities are usually done on websites specialized in real estate field. There are thousands of properties for sale and rent posted on these housing advertisements. However, they cannot offer the customers detailed information about real estate market such as the trend of price in recent five years, the distribution of real properties through Vietnam, the popularity of a current house type, or the average price in a specific area. Instead, they are just the websites used for advertising properties, and their purpose is mainly to offer the customers real estate items that they can buy or rent. Regardless of the above drawback, these websites still have many promising applications because of the large volume of data that they own. As a result, by utilizing the useful data from these commercial websites, the author would like to introduce a solution to the present problem.

II. INTRODUCE WEBSITE

This section consists of a number of figures with descriptions to help users know how to use the website [2].



Figure 1: Graph type menu

IJARCCE



International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 10, October 2018

At first, when users access the website, they need to choose a type of graph that they want to view. There is a total of 10 categories of graph that the user can choose from, including Listing Distribution (Heat Map), Listing Distribution (Bar Chart), Price Distribution, Density Probability, Mean & Median, Box Plot Price, Posting Trend, House Type, Price Trend, and Price Prediction. To choose graph type, users can click on any graph in the red frame in Figure 1. Moreover, there are other kinds of graphs and users can scroll to the right to choose the desired one. Regarding the graph types, each type has different information fields which allow users to filter data. For example, Figure 2 and Figure 3 demonstrate how the real estate listings are distributed throughout Vietnam. Even though the 2 graphs show the same information, they have totally different approaches. While distribution listings heat map has only two filters (transaction type and house type), distribution listings chart has an additional filter named Select Province to narrow the data range. Furthermore, users will have a number of options to filter the data in detailed. Price trend chart provides users a particular filter called Date step, which determines how frequent the data is reassessed.

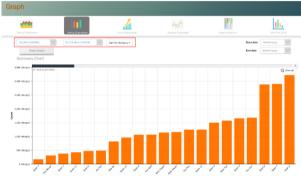


Figure 2: Distribution of listings showed by bar chart



Figure 3: Distribution of listings showed by heat map

In Figure 4, users want to view the graph showing information of mean and median in price of Ho Chi Minh. So, first of all, they will choose the graph type name Mean & Median, then a group of filters will be shown. After choosing appropriate values for the filters, users need to press the button Show Graph so that the graph will be generated by the website (Figure 5).



Figure 4: Button Confirm

Besides showing information from the beginning, users can also choose a specific time interval to view the data. For example, Figure 6 showing the mean and median price of apartments for sale in Hanoi from May 4th, 2017 to June 23th, 2017.

IJARCCE



International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 10, October 2018



Figure 5: Mean and median price of apartments for sale in Hanoi



Figure 6: Mean and median price of apartments for sale in Hanoi from 04/05/2017 to 23/06/2017

CONCLUSION

The authors have learned the basic knowledge of web scraping and gain a background in data mining, including data cleaning, anomaly detection, and pattern detection. For data cleaning, the authors analysed multiple record linkage methods including Support Vector Machine, Decision Tree, Bayes, Neural network, and rule - based record linkage to understand their characteristic and find a method that is suitable for our case. For anomaly detection, the authors experimented with three methods to detect outlier, including standard deviation, interquartile range, and modified z - score. After many test, the authors concluded that modified z – score is a robust solution to detect outlier values that is not affected by extreme value and is suitable to our current need. Finally, from the processed dataset that the authors produced, the authors tried applying multiple data visualization methods to find the pattern underlying the real estate market. The authors have built a system of crawlers to scrape data from six big real estate websites in Vietnam. Using this system, the authors have generated a dataset with around 350,000 listings of real estate, both for rent and for sale between November 2016 and June 2017. From this dataset, the authors extract the pattern and characteristic of the real estate market, including average price, highest, lowest price, and the deviation in data of each region, how price change over time, how location and real estate type affect a property's price, how likely is it that a property in a region has its price fall in a certain range, what is the percentage in the market of each real estate type, and the pattern that brokers and real estate's owners post their listings. Finally, the authors developed a website to visualize these pattern and characteristic to the public, together with extensive filtering options covering multiple aspect of a property like location, transaction type, and real estate type. Currently, dataset consists of only listings between November 2016 and June 2017, since the authors have only started scraping from November 2016, to a total of around 350,000 listings. Even though dataset can be used to get a close insight to many regions (Ho Chi Minh and Hanoi), there are still a lot of regions that our dataset could not cover. Because of that, target in the future is to continue scraping to further expand dataset in other regions. Furthermore, while the authors did store the description of each real estate listings, which could hold many information about the target, its current use is only as a weight to decide whether two listings are duplicate. For further progress, the authors aim to utilize this source of information using text mining and natural language processing to extract other details of a property.

REFERENCES

- [1]. S. R. O. V. NAM. [Trực tuyến]. Available: http://moj.gov.vn/vbpq/en/lists/vn%20bn%20php%20lut/view_detail.aspx?itemid=4769. [Đã truy câp 13 7 2017].
- [2]. Vo Thanh An, Nguyen Hoang Phuc Huy (2017), "Insights Into Housing Markets In Ho Chi Minh City", Bachelor Of Science In Computer Science.