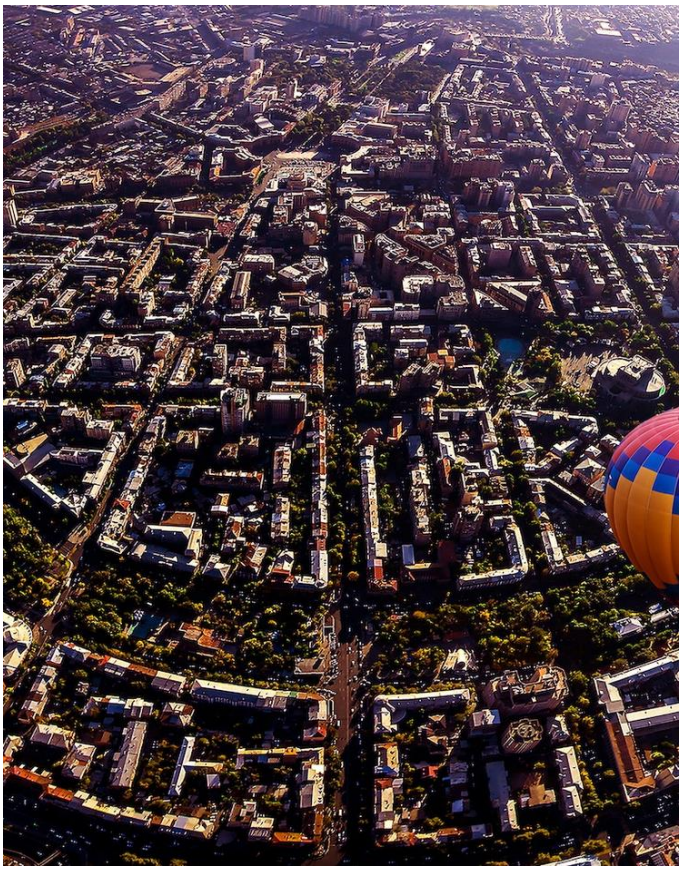

REAL ESTATE PRICE CHANGE PREDICTION IN YEREVAN BASED ON LOCATION

RESEARCH PROPOSAL|

Lecturer: Arthur Pokrikyan

Authors: Karakhanyan Aram, Lachinyan Izabella, Hakobyan Hakob

Date: 19/12/2022

Table of Contents

Abstract

The aim of this study is to analyze the dynamics of the housing market in Armenia's economy and to examine the impact of variables related to housing prices. Preferred by many international housing investors, Armenia hosts profitable real estate investments as one of the developing countries with a shining housing market. This study applies the dynamic model averaging methodology to predict house price growth. We suggest to do it in a phased manner, firstly data mining, then counting the future value of dinamic variables, after that inputting the variables into the final model and counting the price change. With the increasing use of information technologies, Google online searches are incorporated into the study. For this purpose, twelve independent variables, with the Residential Property Price Index as the dependent variable, are suggested to be used in the period of January 2010–December 2022.

Introduction

Importance

House is one of human life's most essential needs, along with other fundamental needs such as food, water etc. Demand for houses in or near the big cities grows rapidly over time, which may be related to both natural factors such as population growth and urbanization and non-natural factors such as unexpected change of migration or governmental decisions. While most people buy or construct houses for living, some people do it as an investment expecting higher returns. The benefits of investing in real estate are numerous. With well-chosen assets, investors can enjoy predictable cash flow, excellent returns, tax advantages, and diversification—and it's possible to leverage real estate to build wealth.[1] Changes in the real estate price can

[1] https://www.investopedia.com/articles/mortgages-real-estate/11/key-reasons-invest-real-estate.asp

affect various household investors, bankers, policymakers, and so on. Thus, predicting the rate of the price change for real estate is important both for investors and other entities.

Problem Statement

Real estate industry is among the fast-growing financial industries topping the list. Though there are several models in predicting apartment prices abroad, in Yerevan investors tend to rely on their own foresight and information on future development of the particular district. Prices and valuations of apartments have significant unprecedented changes over time. Both the investors and the users of the apartments and the entire real estate industry come across with significant challenge of price fluctuations depending on seasons, location, economy and more factors. There is a need to develop a reliable predictive model to give an expected price of an apartment soon. Having a predetermined future help developer plan their future investment trends per demand and economic viability.

Objective

House price is considered to be related to various house features. In general, these features can be grouped by two categories: structural features, such as the number of bedrooms and floor space area; locational features, such as the distance to the city center and the quality of nearby schools. As the goal of our research is to find out how much the change of real estate prices will be on a certain date and in a certain area, we are going to use only locational variables. The primary aim of this paper is to use these Machine Learning Techniques and curate them into ML models which can then serve the users. Our model will answer the question about how much the change of real estate price may be in a certain location based on the information that an investor inputs about the future.

Literature review

House price prediction can be done by using multiple prediction models such as support vector regression, artificial neural network, and so on. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them

determine house prices. The main disadvantage of the current systems is that they are concentrating on predicting the price of a certain house based on its structural parameters mostly without considering market trends which results in price increase. In this paper, literature review focuses on estimation of house price change based on the locational variables using differrent models such as multiple linear regrassion and machine learning. In this paper we analyze different research papers for constructing better models in our case.

Shang Mei researched the connection between real estate value and investment in the construction industry, unemployment level, average income rate, output of industrial products, interest level of bank loans. By calculating the correlation between these index and real estate series, Shang Mei mentions that these indexes are strongly related to the variables of real estate prices.

Wang Xia researched the impact of the traffic environment on real estate projects, which is based on the residential areas along the Beijing Metro, especially studied the change of housing price in the residential area around Beijing Metro Line 13. The scholar found that the change of the price of the residential area around Line 13 is obviously related to the location of the region. The worse the location of the district is, the farther away from the city center is, the greater the price will rise and the better the location of the district lies. On the contrary, the closer to the city center is, the smaller the prices will rise.[2]

Rafiei and Adeli have used SVR to determine whether a property developer should build a new development or stop the construction at the beginning of a project based on the prediction of future housing prices.[3] Using data from 350 condos built in **Tehran (Iran)** for the period between 1993 and 2008, their study trained a model with 26 features, such as ZIP code, gross floor area, lot area, estimated construction cost, duration of construction, property prices of nearby housing developments, currency exchange rate and demographic factors. Their results

[2] Wang, X., Zhu, D. and Zhang, M. (2004). The impact of urban rail transit on real estate price--taking beijing light rail line 13 as an example. Urban, (6): 39-42. Available: https://doi.org/10.13239/j.bjsshkxy.cswt.2004.06.010
[3] https://www.tandfonline.com/doi/full/10.1080/09599916.2020.1832558

have demonstrated that SVR is an appropriate method to make predictions of housing prices since the prediction loss (error) is as low as 3.6% of the test data.

The idea of boosting algorithms has been developed by Breiman as numerical optimisation techniques in function space, and these algorithms can provide a solver to many classification and regression problems. Gradient boosted classifiers belong to a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model, typically decision trees. Gradient boosting models have gained their popularity because they are effective in classifying complex datasets.[4]

In spite of above-mentioned ML methods there are number of other methods to foresee future price of real estate. We should take into account that every country has its special characteristics, reachable data, development strategies, therefore one model can be accurate for several countries, whereas it could show big estimation errors for others. Hence there is no common way that could be equally effective for all countries or regions.

Research questions and hypothesis

Research questions

Based on the problem set in the introduction of this research proposal, we are going to answer the following questions:

- What factors affect the natural trends of the real estate price changes, and what are the factors that may affect the prices not naturally?
- What is the relation between the real estate prices and development of infrastructures in an administrative district in Yerevan?

[4] https://www.tandfonline.com/doi/full/10.1080/09599916.2020.1832558

- How does the change of urbanization level in Armenia affect the change of the real estate prices?
- How do the force majors inside and outside the borders of Armenia (e.g. state of war, pandemic, political instability etc.) affect the real estate prices?
- How can the real estate market supply and demand be evaluated based on the available data in Armenia?

The answers of these questions will help us develop a model to evaluate the change of real estate prices at any location in Yerevan.

Hypothesis

Based on our objectives we set these hypotheses, which we are going to test:

- If new projects aimed at improving infrastructure (for example, road construction, improving tourist centers, establishing factories or other companies in the region that provide jobs) are implemented in the region, then the price increase in the region should be more than where there are no infrastructure improvement plans.
- Inflation that is not related to natural trends (immigration, war, epidemic, etc.) has a temporary nature, and its effect is neutral in the long run.
- If the city expands in any direction, then the price increase will be the largest in those directions during the observed period.

Implementation

Data

In order to evaluate the real estate prices of each area of the city of Yerevan we should possess accurate data to pass it to our ML models. First we choose the websites specialized in buying and selling apartments. Then we extract the data, scrapping it with the help of R and cleaning it removing outliers and any kind of suspicious data. As a dependent variable we select the mean

price of flats of every area or district in the city. With regards to Dependent variables or features we try to take the significant ones wich must have strong impact on apartment prices.

For better result independent variables should have as less correlation between each other as it is possible. In fact features are not completely independent, and there are many explicit or implicit connections between them. For example, two independent features in geographical space may belong to the same school district. As a result, when analyzing the impact of the school on house price, the two parts should be merged rather than separated.

In the great majority of cases, consumers are more concerned about general locations than with detailed addresses ,so our prediction model will decide the mean price of apartment or house in a selected Geographical area.

Since some of our features have categorical characteristics, we have to apply some transformetions converting them into numbers. In case of the numeric variables we do not face such problems and we are able to count metrics such as mean ,median ,Standart deviation etc.

Education. In recent years, educational resources have received more and more attention, thus we mention elementary and secondary schools ranking them giving a number which will determine their quality,such type of data we can collect from the **Ministry of Education https://escs.am/en**. find the accessible schools of the particular district and count the mean school rank of the area, which definitely have the impact on houses of area.

Transportation. Since transportation networks have always been of a great concern, we set up several features:

1. the accessibility of metro station. if a district does not have metro station we give it 0 rank if there is metro station we divide the the district into parts where walking distance to the station will separate them. For example 0-5 minutes to station we rank 3 points,5-10 minutes 2 points 10-15 minutes 1 point and above 15 minutes 0.This data is possible to collect by using **Google map**.

2.The distance and self-driving time from the center of the distinct to the city center using **Google map**.

Facility. Proximity to facilities such as trade centers,shops, hospitals, clinics, etc. may affect the area house price too. Therefore, we introduce six different features based on these six facilities to describe the distance between a given house and the accessibility of nearest six facilities, where the distance calculation uses **Google Maps API**. These are supermarkets, hospitals, clinics, parks, entertainment areas and shopping centres.

Urbanisation rate. Another descriptive factor to estimate price of area is urbanization rate. Such data can be easily collected from the following website called **armstat.am**. The datum represented here is numeric ,so we do not have to transform it.

Migration rate. Recent years in Yerevan there is a significant growth on demend on real estate prices, conditioned by the War in Republic of Artsakh and Russian-Ukranian conflict. Many families migrate to the capital to ensure peacefull lifestyle for their children. This has a dramatic influence on apartment prices and this tendency also worth including as an important feature to determine it price. Migration rate is a measurable quantity and the data is possible to scrap from **migration.am**.

Job opportunities and companies nearby. Job opportunities are another significant factor when determining the housing future prices for a certain location. The thing is that people tend to arrange their workplace to be as close to their houses as possible, or they need their new house to be as close to their workplace as possible. So considering how many job opportunities and companies there are in a certain distinct, or how the tendence is going over the years, may be correlated with the real estate future price in that region. Such a data may be scrapped from **spyur.am** based on the location of each company. We can take how many job opportunities each company offers, where it is located, when it is founded etc.

Methods and models

In order to achieve our goal we have to go through several stages. For the first step some time series modeling is to be implemented to predict the future values for some of our input features that change over time (e.g. migration, urbanization). There are also some variables that are static (e.g. the distance of a distinct from the city center) which we will use with their absolute present value. This little classification is important as without this we won't be able to input

proper variable values in our final prediction model. All we need to do is to take time series for each necessary variable, and predict its value based on its trends. The article of Bonnie Ma (2020) gives a well-designed time series modeling with ARIMA to predict future house prices. We can use the same method to predict necessery features which we are going to use in our final model.

For the second stage we should insert all our inputs to our price prediction ML or multiple linear regression model, which will compute the average future price for each district in Yerevan having the features mentioned above as independent variables and the average price of all transactions during a year as the independent variable. Getting this output we should check the significance of the influence for each of our features, and take only those variables which have significant influence on our output, and then we will get our model rid of unnecessary data. The paper of Guangliang Gao (2019) detailly describes the comparison of different methods of doing the location-centered house price prediction with a multitask learning approach. After all we consider the multiple linear regression model as a relevant method to acchieve our goal.

After that, having the furure valure of real estatein a certain distinct as a semifinal output we will be able to calculate the overall predicted price change for a certain time period. Then the app will visualise the output for different distincts and show the users the final procent of price change for each distinct.

Assumptions

The main assumptions of the implementation are the following:

- We assume that it is possible to generate the necessary data from the sources we have mentioned above;
- We assume that the average price change of a certain location also applies to the price of each land or building taken separately;
- We assume that every force major will accordingly change some of the features implemented in our model and affect the prices, so the users (investors) will have opportunity to mention in the app the information they have about the future, and they will get different outputs based on the information each of them has.

Timeline

Sequence of implementation stages

Stage	Activity	Estimated duration
Research design and planning	Finalise research problem/questions	two day
	Develop research design	four day
	Prepare research proposal	one week
Literature review	Search, capture and synthesise relevant literature	one week
	Prepare draft literature review	five day
Data collection	Finalise sampling plan	two week
	Develop data collection instrument	one and half week
	Pre-test/pilot data collection instrument	one week
	Carry out data collection	two week
	Write up data collection	one week
Data analysis	Prepare data for analysis	one month
	Analyse data	two month
	Draw conclusions/ recommendations	five day
Writing up	Final draft of application	three month
	Review draft	one week
	Final editing	one week
	Submit to extramural funder	six month

The team

Our team members are Karakhanyan Aram, Lachinyan Izabela, Hakobyan Hakob. We are Masters students of Data Science for Business, and have also a lot of support from our professional lecturers, so the realization of our research proposal is just a matter of time.

The budget

The budget of the realization of our research is 0 dollars, as we consider it to be our own investment in our final project. But to create an app which will be our MVP we will surely need budget for about $5000.

Risks and expected results

Risks that may prevent the project from being implemented

1. Lack of some historical data for assessment
2. Not consideering emergency situations (earthquake, political events, war) in our model
3. Low performance of the model on real data. Performance risk occurs when a project no longer performs as originally expected
4. Lack of resources. Resource risk occurs when you don't have enough resources to complete a project. Resources include time, skills, money, and tools, among others.
5. Lack of clarity. Lack of clarity can be expressed in misunderstanding among stakeholders, uncertainty in the scope or timing of the project. This uncertainty can result in lack of control due to fragmented workflows, over budget, missed project deadlines, changing project requirements, the need to adjust project progress, or disappointing project results.

We assume the features we have taken to have significant influens on our dependent variable. So we expect our model will find a solution to our problem with the highest possible accuracy.

References

1. Key Reasons to Invest in Real Estate, BARCLAY PALMER (2022)
2. Shang, M. (2004). Analysis of macroeconomic factors affecting the price level of construction products. Northern Economy, (4): 39-40.
3. Wang, X., Zhu, D. and Zhang, M. (2004). The impact of urban rail transit on real estate price--taking beijing light rail line 13 as an example. Urban, (6): 39-42.

4. Predicting property prices with machile learning algorythms, Winky K.O. Ho, Bo-Sin Tang, Siu Wai Wong, (2020)
5. Time Series Modeling with ARIMA to Predict Future House Price, Bonnie Ma (2020)
6. Location-Centered House Price Prediction: A Multi-Task Learning Approach, Guangliang Gao, Zhifeng Bao, Jie Cao, A. K. Qin, Timos Sellis, Fellow, IEEE, Zhiang Wu, (2019)