|  |
| --- |
| ADVANCED DATA ANALYTICS |
| Analyzing Trends in Housing PRICING: A Comparative Study on NY, CA, TX, and IL |
|  |

|  |
| --- |
| AVINASH APPINENI  5-4-2024 |

Table of Contents

[**1.** **Abstract** 2](#_Toc165756242)

[**2.** **Introduction** 3](#_Toc165756243)

[**A. Context and Background:** 3](#_Toc165756244)

[**B. Purpose of Study:** 3](#_Toc165756245)

[**C. Justification and Methodology:** 4](#_Toc165756246)

[**D. Significance of the Problem:** 4](#_Toc165756247)

[**3.** **Literature Review** 4](#_Toc165756248)

[**4.** **Methodology** 6](#_Toc165756249)

[**A.** **Data Collection:** 6](#_Toc165756250)

[**B.** **Data Preprocessing:** 6](#_Toc165756251)

[**Cleaning and Transformation:** 6](#_Toc165756252)

[**Variable Selection:** 7](#_Toc165756253)

[**C.** **Analytical Techniques:** 7](#_Toc165756254)

[**Trend Analysis:** 7](#_Toc165756255)

[**Regression Modelling:** 7](#_Toc165756256)

[**Comparative Visualization:** 7](#_Toc165756257)

[**D.** **Data Analysis:** 8](#_Toc165756258)

[**Quantitative Analysis:** 8](#_Toc165756259)

[**Graphical Representation:** 8](#_Toc165756260)

[**E.** **Logical Flow:** 8](#_Toc165756261)

[**5.** **Results & Discussions** 8](#_Toc165756262)

[**A.** **Key Findings:** 9](#_Toc165756263)

[**Trend Analysis:** 9](#_Toc165756264)

[**Regression Modelling:** 9](#_Toc165756265)

[**Comparative Visualization:** 9](#_Toc165756266)

[**B.** **Interpretation:** 10](#_Toc165756267)

[**C.** **Implications:** 10](#_Toc165756268)

[**D.** **Conclusion:** 10](#_Toc165756269)

[**6.** **Output** 11](#_Toc165756270)

[**7.** **References:** 15](#_Toc165756271)

**Analysing Trends in Housing Pricing:**

**A Comparative Study on NY, CA, TX, and IL**

# **Abstract**

This study explores the variable growth in the home prices within four key U.S. states which includes Texas (TX), Illinois (IL), New York (NY), and California (CA). This study helps to identify and analyse the suitable patterns that have been present in the various properties which markets over the past few years by using the Large Zillow House Price Dataset. So, we have a set of approaches which includes, Regression Modelling, Trend Analysis, and Visualization techniques with comparisons.

After the analysis that we have done, we have found out the major differences in the prices of house trends for the states that we have chosen. There are numerous factors which includes change in population, legal frameworks, and economic conditions which have impact on these differences. We are aiming to provide useful information for all the investors, politicians and stakeholders, by using solid statistical analysis helps to clarify the underlying factors influencing these different trends. So, these analysis and factors that we considered will be useful for all the people in the states who are real estate enthusiast.

The results that we obtain are helpful in saying how crucial to understand the dynamics of the housing market within the context of certain geographical settings. Stakeholders can allocate assets, invest in housings and change the policies with more knowledge of the various socioeconomic factors that exists in each state. Our study and research contribute wide range of knowledge within the regional economic inequality, urban development, and affordable housings.

# **Introduction**

In this study, the paper states us the complex dynamics of prices of homes and their trends with respect to four major U.S. states includes: Texas (TX), Illinois (IL), New York (NY), and California (CA). In addition, it represents the regions economic health, prices of housings, and have a significant impact on the lives of residents in those states. As such, they are very much crucial in understanding the socioeconomic environment.

## **A. Context and Background:**

As, the Housing background and their markets affects the economies of localities and quality life of people, New York, Los Angeles, Houston, and Chicago are the major urban centres which will frequently subject to criticism. Diverse groups from various areas and states come together for search of a place to live, work and for experiencing cultural activities. Grasping the opportunity of the diverse information and other trends in property pricings for all the investors, individuals, politicians and other real estate agents which helps in their future tasks.

## **B. Purpose of Study:**

The main goal and purpose of the study is to have a performance evaluation of comparative analysis on the variable growth in the prices of homes and properties of the selected states like NY, CA, TX and IL. By utilizing the data set provided by the Zillow, our objective is to detect the new trends, pointing out the variations and to derive the understanding of the fundamental elements influencing the swings and movements of the prices structure of prices in these selected states in US.

## **C. Justification and Methodology:**

By understanding the complexity of the pricings of homes and properties and their variable trends is such an important thing to deal with, since the housing is a vital component of the people’s lives, we want to have a light on the rise and fall variability of home prices over the last few years by utilizing the trend analysis, comparison visualisation and regression modelling. Utilizing a strong statistical method to draw significant and important conclusions from the data that we have taken and our approach is quite grounded in the empirical analysis.

## **D. Significance of the Problem:**

There are several reasons encountered while examining the patterns in the house prices which is very important to discuss. The first benefit is that, it offers very insightful information to the people who have their work on the real estates and other land dealers and who debate housing supply and its affordability. And second benefit is, it really educates the real estate experts and investors who invests on land and on possible investment dangers and opportunities. Lastly, for the people and families to make the well-informed choices about the rental agreements, financial planning, and other ownerships related to housings and properties, as they have a solid grasping of knowledge and information about how the housing market works and how the trends vary within terms.

# **Literature Review**

Many studies have been conducted and implemented on the housings and property markets in both the academic background and the business-based background due to the complicated connections between the social, economic and regulatory aspects. An extensive analysis of literature offers very insightful information about the dynamics of falling and raising of the variable house prices and the factors which are influencing the trends of the property and housing markets.

The studies that have conducted in the past that have pointed out the importance of the housings and properties as a major fundamental element and as a significant indicator of the state of the economy as a whole (Case & Shiller, 2003). According to the Glaeser et al. (2008), The housings and property markets will have an outcome which are shaped by the supply and dynamics od demand, regulatory policies and demographic trends of houses. The empirical analysis of the data that we have collected indicates that a number of variables which includes affordability of housings, interest rates, income levels and population growths which further have a significant impact on housing prices (Gyourko & Molloy, 2015).

The studies that have been conducted on the regional housing and property markets have highlighted and have focus on the significance of the regional employment trends, patterns on the people who are planning to migrate from one state to another, and other local economic conditions which can have a drastic influence on the costs of houses among various geographical areas (Henderson & loannides, 1983). According to Saiz (2010), The research has started and it says that the significance of the amenities, quality life of people, and the infrastructure are the variables in influencing housing demand and by inflation of prices.

The main effect of the developments in technology are the growth of the big data analytics and developing real estate platforms in online which further helps for efficiency and transparency of the housing markets and has also been the subject of recent literature (loannides & Zabel, 2008). In addition, the research on how the changes in the housings and property markets that affects the economic phenomena which further includes the inequality in wealth, financial and economical stability and macroeconomic performance that has become more well known in academic circles (Mian & Sufi. 2009).

There is a need of empirical studies which offers a detailed insights of the specific dynamics and growth of the housing prices with respect to region, and even with the large amount of research on housing markets. The objective of the study is to add to the existing body of the literature by providing a break through of analysis of the factors influencing the prices of houses and properties and their trends in New York, Texas, Illinois and California through a comparative analysis of their housing markets.

# **Methodology**

Texas (TX), Illinois (IL), New York (NY), and California (CA) were the major states that we have taken into consideration in US on which we further concentrated on conducting the research. The importance of these states are the main reasons why we chose them majorly are due to the population, gross domestic product, and market size of housing which led for the selection. The above is the description of analysis where there are steps included in the methodology,

## **Data Collection:**

The data that we have collected is from the Zillow house price dataset, which offers a detailed information and knowledge about the house prices across the various areas and time periods with respect to the price data of housings.

## **Data Preprocessing:**

### **Cleaning and Transformation:**

The cleaning procedure are applied to the raw dataset which is taken directly from the source to eliminate the missing values or discrepancies. The main aim of the cleaning and transforming the data is to make the data consistent and compatible for analysis and make it uniform for further processing.

### **Variable Selection:**

The variable selection includes the variables that needs to be included from the dataset. The four states of interest were the primary and major focus of our selection of the variables from the dataset. The variables that we include are RegionID, SizeRank, RegionName, StateName, and Dates.

## **Analytical Techniques:**

### **Trend Analysis:**

To find the average housing price for each state during the certain time periods we use trend analysis. This can help in finding the variable trends in prices of houses and properties. Also, this entails figuring out the trend lines and evaluating the frequency and direction change in price movements.

### **Regression Modelling:**

As to find the mathematical correlations between the years and prices, that is exactly time which means years and average house prices, we tend to use the Linear Regression Models. So, as we are using it we are able to evaluate the periodical factors of house prices and forecast and predict the future prices and patterns of the variable prices.

### **Comparative Visualization:**

So, for the visualization we are using regression line scatter plots which makes it to get easier data to compare between states and time periods (Years). In addition to pointing out any major variations in pricing patterns, these visualizations will offer insights into the respective performance of various housing markets.

## **Data Analysis:**

### **Quantitative Analysis:**

As to extract very valuable insights from the dataset we use quantitative analysis. The importance of the correlations between the variables have been determined using the statistical techniques like regression analysis. So, that’s why we use quantitative analysis for extraction of insights.

### **Graphical Representation:**

The main purpose of the graphical representation is the display the studied data properly and is very much useful to scatter plots and line plots. Due to these representations, it will be easy to understand titles, axis labels, legends which make them self-explanatory.

## **Logical Flow:**

The Methodology part is arranged logically, where the starting part is with the preprocessing and collecting information and data and moving to a description of the analytical approaches used. To provide the understanding and consistency in the presentation of the parts of technicalities of study, where sub headings are used to indicate the differences in stages of analysis process. This could be the logical flow that we used in the paper.

# **Results & Discussions**

Here in the results and discussions section we are going to discuss about the key findings, implementations and interpretations and finally about the conclusion. The below are the sub sections of discussion.

## **Key Findings:**

### **Trend Analysis:**

Here we discuss about the trend analysis with respect to states that we have chosen:

New York: Over the period of time, according to the trend analysis there was a noticeable acceleration and variation in house prices in NY

California: Over a period of time, there was a notable increasing trend in the house prices in CA, which have a strong growth in the states real estate sector.

Texas: Over a period of time, although at a slower pace than NY and CA, Texas also having significant increases in prices.

Illinois: Over a Period of time, the raise and fall of house prices was comparable to Texas, which shows that it is having steady growth over time.

### **Regression Modelling:**

The Regression Analysis supported the overall growth and spike in the trend in data by establishing a positive association between the time period (Years) and average price of homes in all states that we have selected. The regression models have coefficients to provide data about the increase in prices size and which is connected to each unit rise over time.

### **Comparative Visualization:**

There is an evaluation in trends of house prices across the four states, this is known through the scatter plots and line plot. The Visual inspection tells us the similarities between the patterns of pricing and the dynamics of the housing markets. These similarities also act as a proof in the general pattern of price fluctuations. Easy to find the market dynamics.

## **Interpretation:**

The results that we obtained are the hypothesis of telling us that there has been an increase in prices of homes in NY, TX, CA, and IL in the overall time period which is calculated in years acted as supporting in findings. The results also say about the real estate markets which will rise continuously and consistently with the other research showing a positive relation between time and prices of properties and houses. Any ways our findings can confirm that the current market system of housings and their patters of pricing, and also offers some important insights into the dynamics that we see in each state that we have selected. These are the interpretations of the paper.

## **Implications:**

There are implications gathered from this paper which further discuss about the Patterns that are noted by the owners, real estate people, investors and politicians in the housing prices that have important effects. Making the decisions about investments and policy creations and also economic planning requires an understanding the major elements which can influence housing market dynamics. These are the important implications that have been extracted from this paper.

## **Conclusion:**

Concluding this paper, All the states that we have selected and included shows that there is a rise in prices oof houses over the period of time which is years, according to the study of trends and analysis of housing prices in all the states the results tell us that the politicians, real estate people, investors, and other people who are interested in putting money in lands can be beneficiary and very much crucial to understand the dynamics of the social and economical aspects which further drives the markets of housing. As the regression model is used and supported in a positive correlation between the years and average prices of properties, trend analysis and rising prices. The market dynamics information is obtained by using comparative visualization, which further highlights the trends of pricing which are similar across the states. These findings help the necessity of well informed to solve housing market difficulties and have implications for economic planning, formulating policy and decision making in investment. In summary, this study helps in advancing the knowledge on affordable housing, urban development, regional economic inequality. Also provides very useful information for stakeholders to navigate the complicated structure of the markets.

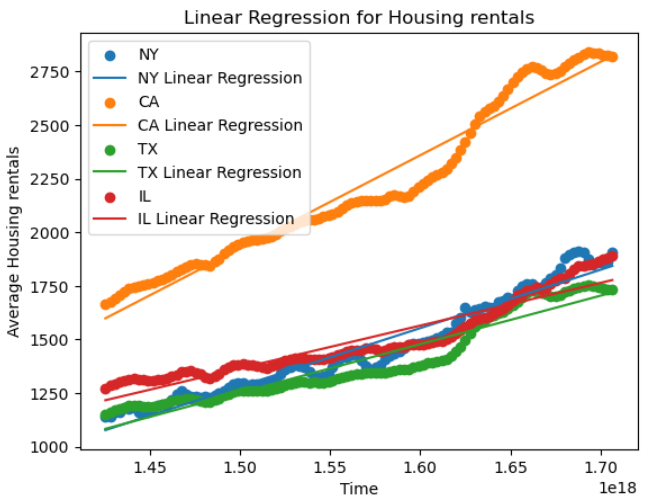
# **Output**

House rentals:

Code:



Output:

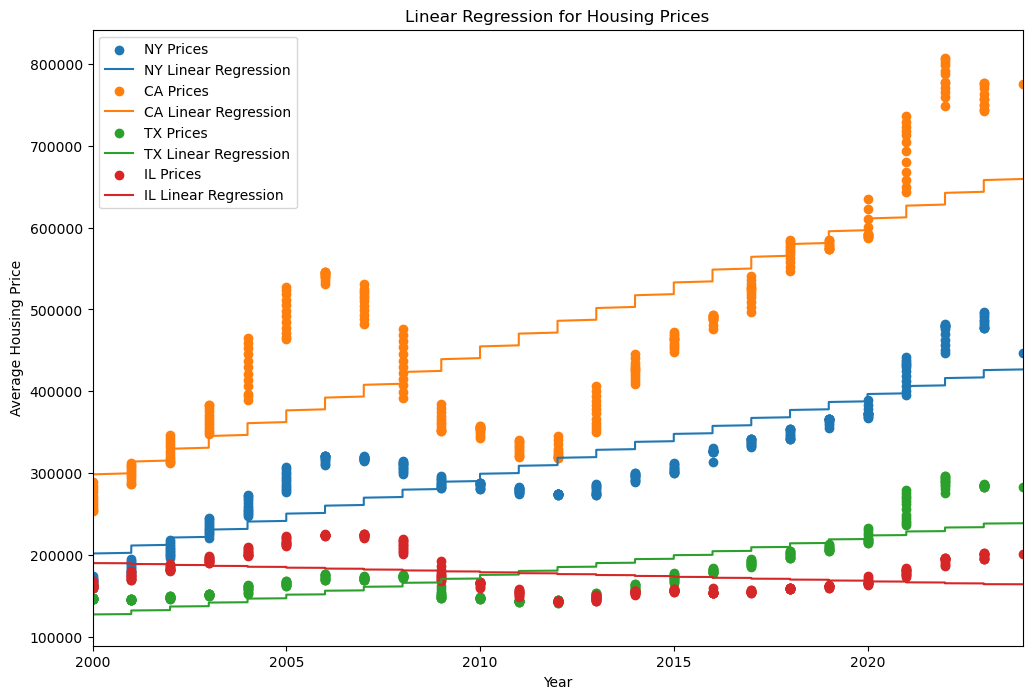


House prices:

Code:

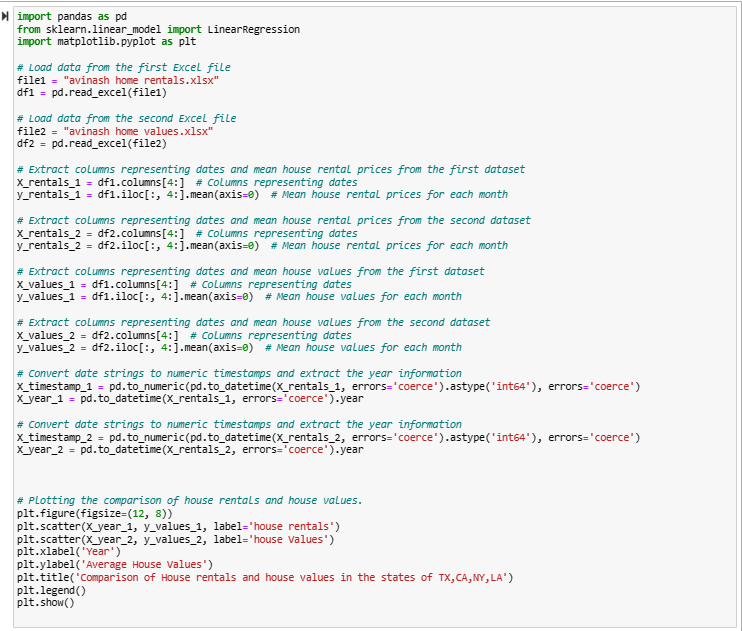


Output:

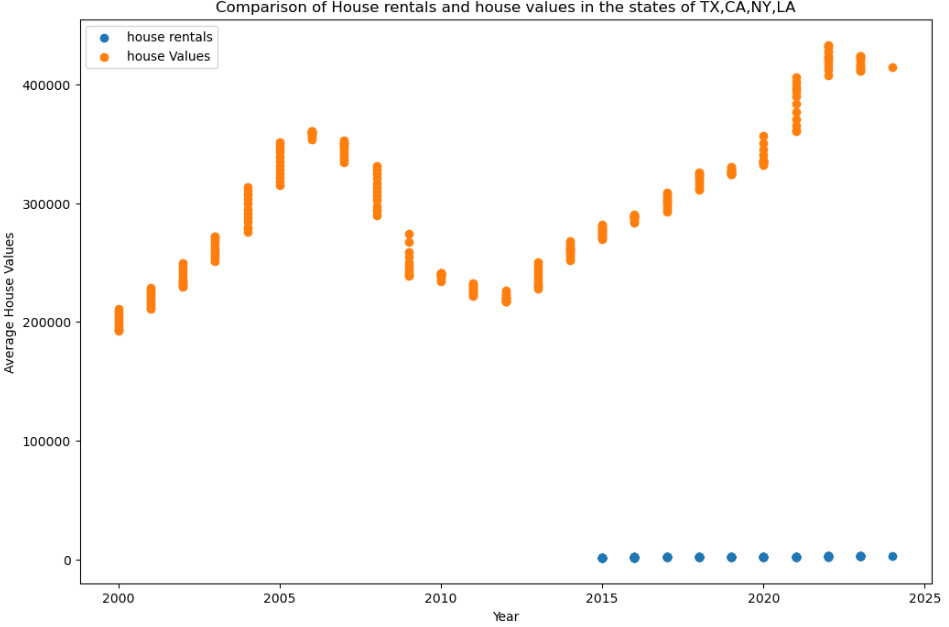


Comparision:

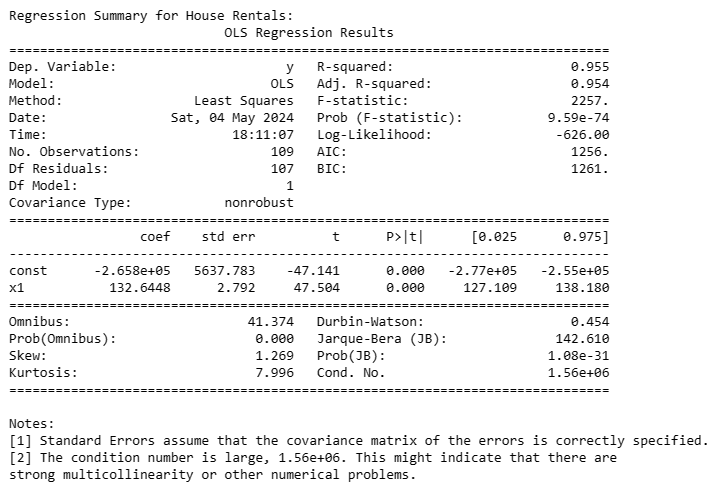
Code:

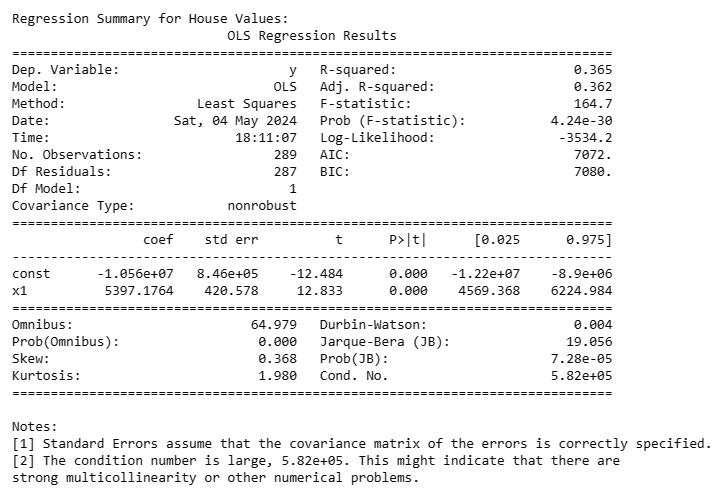


Output:



OLS:





# **References:**

Case, K. E., & Shiller, R. J. (2003). *Is There a Bubble in the Housing Market?* Brookings Papers on Economic Activity, 2003(2), 299–362.

Glaeser, E. L., Gyourko, J., & Saks, R. E. (2008). *Why Have Housing Prices Gone Up?* American Economic Review, 98(2), 311–315.

Gyourko, J., & Molloy, R. (2015). *Regulation and Housing Supply.* National Bureau of Economic Research.

Henderson, J. V., & Ioannides, Y. M. (1983). *A Model of Housing Tenure Choice.* American Economic Review, 73(1), 98–113.

Ioannides, Y. M., & Zabel, J. E. (2008). *Interactions, Neighborhood Selection, and Housing Demand.* Journal of Urban Economics, 63(1), 229–252.

Mian, A., & Sufi, A. (2009). *The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis.* Quarterly Journal of Economics, 124(4), 1449–1496.

Saiz, A. (2010). *The Geographic Determinants of Housing Supply.* The Quarterly Journal of Economics, 125(3), 1253–1296.