

# Anticipated Home Buying Trends by MBA Graduates Using SAC

Authors: Madeleine Dela Cruz, AndyMike Emata, Claudia Perez De La Cruz,  
Kelly Templeton, Kelly Young

California State University Los Angeles BUS5100-93: Introduction To Business Analytics  
{aemata, mdelac65, cperezd, ktemple2, kyoung49}@calstatela.edu

**Abstract:** The purpose of this paper is to analyze a real estate dataset to assess the best location to buy a home for post-MBA graduates, based on salary. Our aim is to implement SAP Analytics Cloud tools and principles from the BUS 5100-93 MBA course at California State University Los Angeles.

The dataset is from Zillow Data<sup>[1]</sup>. It is a compilation of three CSV files, along with a file from US Census, totaling 14.8MB of data. The compiled datasets include real estate sales in the US over the last 20+ years by city, with a variability in home size (number of bedrooms).

Specifically, we will create geo-mapping, charts, time series analysis, and a predictive regression model to evaluate best locations in the US considering the following: 1) Average salaries for MBA graduates; 2) Types of residences, including different sized homes; 3) Different household scenarios such as single income, dual income, and those with children.

## 1. Introduction

This paper utilizes SAP Analytics Cloud (SAC) to process and visualize five datasets of real estate sales within the US. The datasets retrieved from Zillow consist of information concerning densely populated metropolitan statistical areas (MSAs), states, home size, and average prices. The dataset from US Census includes longitude and latitude supplemental data to enable geolocation mapping.

Homeownership is one of the most widely discussed topics across generations, including Generation X, millennials and Generation Z. Previously, these demographics were raised by baby boomers who had more access to affordable housing within their means. However, with variables such as inflation, rising costs of living, and a fluctuating economy, many feel that homeownership is nearly impossible. One might pursue a master's degree, such as an MBA, to have an easier path to a mid-level or executive job, resulting in career path with a higher salary and higher net worth. According to a Corporate Recruiters Survey conducted by Graduate Management Admission Council, the expected average salary for an MBA graduate in 2024 is \$125,000.<sup>[2]</sup>

We hope this information will assist our classmates after graduation and help to keep them motivated for the remainder of the FEMBA program. It is directly applicable as this would be a major life milestone and decision

contemplated by our classmates after receiving their diplomas.

## 2. Related Work

One related study was published to LinkedIn Pulse with a user sharing the work entitled, Exploring House Price Prediction: A Data Analysis Journey with Tableau<sup>[3]</sup>. As indicated in the title, the analysis was performed using Tableau, and evaluated several variables, featuring number of bedrooms similar to our analysis. Also considered were variables including total area, number of bathrooms, number of stories, connection to the main road, availability of a guest room and basement, presence of hot water heating and air conditioning systems, parking spaces, preferred area status, and furnishing status. This went beyond what we were evaluating and aimed to predict pricing based on these factors. This study found that house prices were correlated most to air conditioning and main road access. Their data agreed with ours that 2-bedroom homes have a disadvantage to favorable price growth trends. They were not able to complete a time series analysis, as the dataset was missing historical price trends.

Another related study of a similar dataset was published on Kaggle.com: Data Analysis on US House Prices<sup>[4]</sup>. The data was analyzed using Python and examined the relationship of housing prices to average area income, average house age, average number of rooms, average number of bedrooms, area population, and address. They found that the data set was normally distributed and continued with a heatmap showing correlational impact of all variables to house price. Also included in the analysis were several linear charts evaluating the relationship between house price and each variable individually, as well as a pairplot comparing house price to each. Like our analysis, this study found that factors like city and number of bedrooms impact the price for residential houses.

## 3. Specifications

The data was retrieved from Zillow, a real estate marketplace company that offers information and services related to selling, buying, renting, and financing property through its web and mobile platforms. Three datasets were acquired for one-, two-, and three-bedroom properties, from the years 2010 through 2024. Average prices for each home size were provided on a monthly basis, ensuring highly reliable and consistent data. The size of each dataset is listed in Figure 1 below.

Dataset	Size
Zillow 1 Bedroom CSV File	2.6 MB
Zillow 2 Bedroom CSV File	3.6 MB
Zillow 3 Bedroom CSV File	3.8 MB
US Census Data	0.002 MB
Simple Maps Latitude Longitude	4.8 MB
<b>TOTAL</b>	<b>14.802 MB</b>

Figure 1. Dataset Overview

For the purposes of this study, we are making several assumptions regarding the study's potential homebuyers:

- The subject has graduated with an MBA in 2024;
- The subject has been offered a job, at a starting salary of \$125,000, which is the national average salary for an MBA graduate in 2024;
- The job opportunity allows the subject to work remotely from any US state of their choice;
- The subject will spend 30% of their monthly household gross income on homeowner's expenses, including mortgage, property taxes, homeowner's insurance, and utilities; and
- The subjects are a single person household (budget \$380K house price), a double income household (budget \$900K house price), and a double income household with children (budget \$900K house price).

This will serve to give MBA candidates in 2024 some ideal locations and home prices to consider when graduating.

#### 4. Implementation Flowchart

The process followed for data retrieval, formatting, uploading, cleaning, analyzing and formatting, is shown below. Zillow data required significant reformatting in Excel. As shown in Figure 2, first, the three spreadsheets were compiled into one spreadsheet. The data file was reduced by removing irrelevant information. Regions [5], divisions, latitude and longitude were then added for each city [6]. The reformatting and new data transformed our final CSV dataset to 1.2 MB. After reformatting the data, the CSV file was uploaded to SAC. Using SAC tools, the data was further cleaned in order to create the stories and models used in our Powerpoint for analysis presentation.

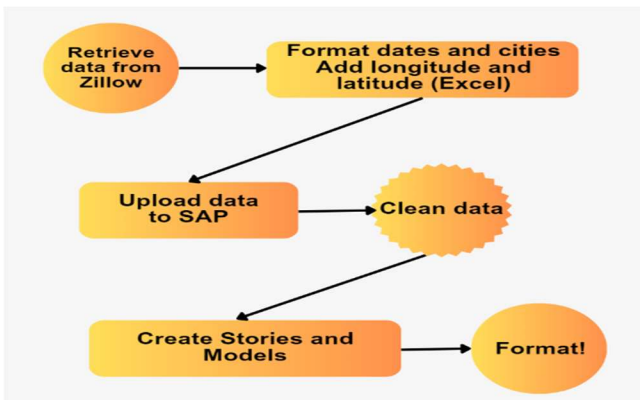


Figure 2. Implementation Flowchart

#### 5. Data Cleaning

The data was cleaned in SAC first by highlighting the dates and transposing them from columns to rows and displaying the Key as a Date and Value as HousePrice. Then, the housing price, latitude and longitude were designated as measures. Most significantly, MetroName was split into respective cities and states.

#### 6. Analysis and Visualization

After cleaning and validating the data, it was determined that the stories and predictive models would include an average price per year and location, a timeseries analysis, prices by geographic region, and regression analysis for price by location.

##### 6.1 National Average Price per Type

The first visualization as a bar chart (Figure 3), was created in SAC and shows the difference in average home price in 2024 using the Zillow Home Value Index by the number of bedrooms and by region as defined by the US Census Bureau. The Midwest region has the lowest average home price across each house category (\$126,386.74 for one-bedroom, \$169,407.97 for two-bedrooms, and \$240,253.06 for three-bedrooms) with an average 25.9% increase in the South, 56.2% increase in the Northeast, and 157.6% increase in the West compared to the Midwest.

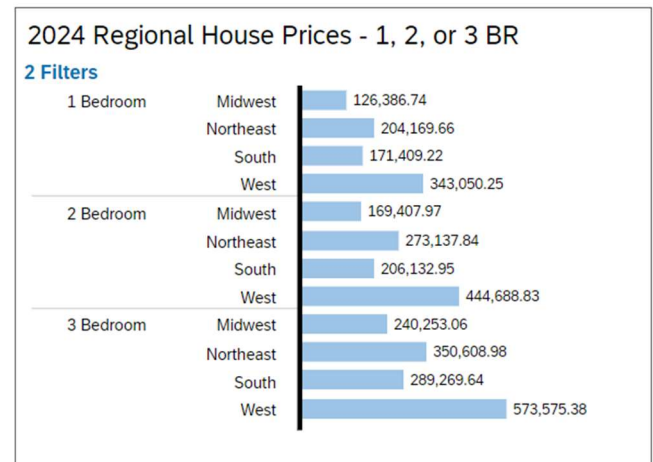


Figure 3. One-, Two-, or Three-Bedroom Average House Price by Regional Location

The second set of bar charts were created in SAC comparing home prices between 2014 and 2024 in five selected MSAs: Boston (Northeast), Chicago (Midwest), Denver (West), Houston (South), and Los Angeles (West). Denver had the highest average home price increase across home categories of 133.4%, followed by Houston with 113.1%, Los Angeles with 95.9%, Boston with 87.9%, and Chicago with 82.0%. Surprisingly, in the one-bedroom analyses, shown in Figure 4 and Figure 5, Chicago one-

bedrooms saw the lowest increase of 65.6% and Denver 1 Bedrooms saw the highest increase of 139.7%.

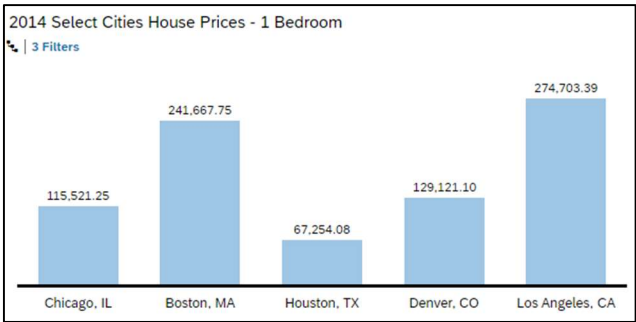


Figure 4. 2014 Select Cities House Prices, One BR

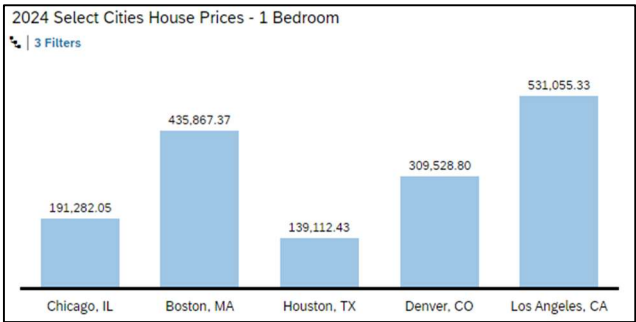


Figure 5. 2024 Select Cities House Prices, One BR

6.2 Timeseries Analysis

The timeseries analysis, shown in Figure 6, was generated through SAC. The analysis includes forecasting the 2026 average house pricing per selected metropolitan area (Boston, Chicago, Denver, Houston, Los Angeles). We can see using the different filters for number of bedrooms that the demand for two- and three-bedroom properties are expected to increase in 2026, as these attract competitive prices and offers from dual income households of an assumed \$250,000, with a budget house price of \$900K; single bedrooms will still increase, but on a lower level, as they may not be as appealing to a growing family or dual income household. Instead, these will remain within reach for single income households and/or young couples.

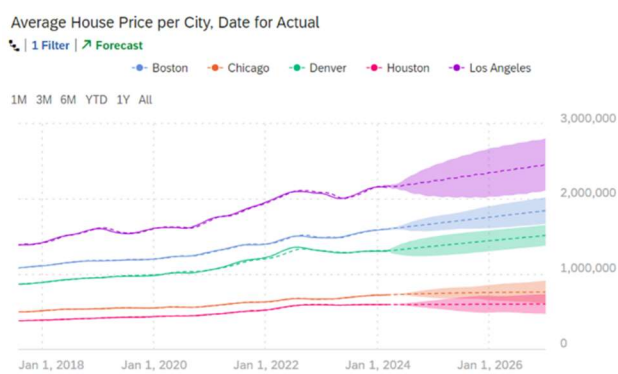


Figure 6. Timeseries Analysis, Five MSAs

6.3 Average House Prices by Geography

The geo map of the US shown in Figure 7 was generated in SAC. The legend demonstrates the house pricing difference by color with blue green being the affordable houses and yellow-orange-red being the most expensive. By switching the filter for the number of bedrooms we can see the different average home prices for a one-, two- and three-bedroom house. As we can see in the map below, the most affordable locations for a three-bedroom home are in the Midwest and the South. The West (Southern California), Hawaii, Florida and the Northeast (New York) are the highest priced homes. For Figure 7 all the bubbles shown are prime locations to purchase a \$900,000 three-bedroom home.

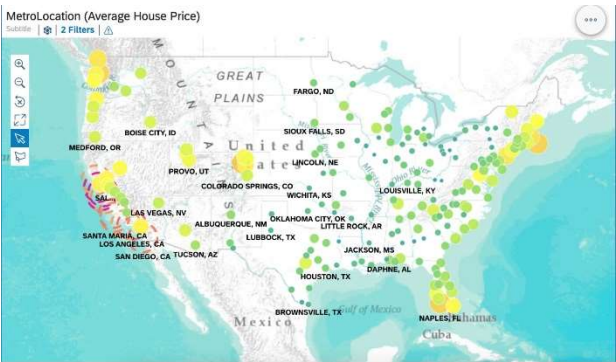


Figure 7. Geo Mapping of an Average - 3 Bedroom Home

6.4 Regression Analysis

SAC features a tool for regression analysis under predictive scenarios. Regression analysis relates one or more independent variables to one dependent variable. We established a regression model through SAC using our Zillow data with the dependent variable as house price for prediction. In Figure 8, we see that the prediction confidence at 99.97% is close to 100%, the highest possible value, indicating that the model is reliable. The root mean square error (RMSE) is 8,993.65, which appears high, but considering the data value of house prices is in dollars, and the scale is in the tens or hundreds of thousands, an RMSE of less than 10,000 is very low, indicating high reliability.

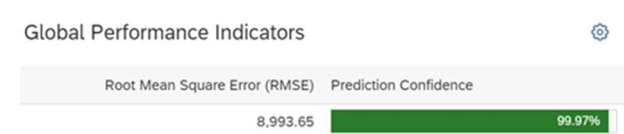


Figure 8. Global Performance Indicators RSME and Prediction Confidence

Below in Figure 9, the influencer contributions are shown. Once the model is built, the prediction is based on the values of the independent variables, as related to the dependent variable. Here, the dependent variable is housing prices, and the independent variables or influencers are shown in order of contribution. The highest impacts are from

Date\_Y at 20.88%, No. of Bedrooms at 17.39%, Longitude at 16.77%, Region at 12.67%, and Region ID at 9.12%.

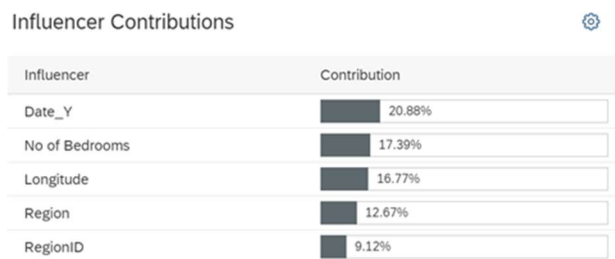


Figure 9. Influencer Contributions to House Price

When the predictive model is established and validated in SAC, a portion of the data is used to assess variation from the regression line or average data, with actual data used as confirmation of the predictive model. The predicted house price vs. actual house price chart (below, Figure 10) provides an indication of the variation in the data set, and accuracy of the predictive model. The range of error during the validation of the model is shown in the chart. The validation – error min in our model is a min of 505,874.21 and the validation – error max is 858,396.54, or a range of +/- 25.8% from the predicted value of 681,983.14. the validation – actual, or perfect model, is 682,121.87, with a difference from the predicted value of 138.73, a fraction of a percentage. The predictive model was confirmed to accurately predict house price values.

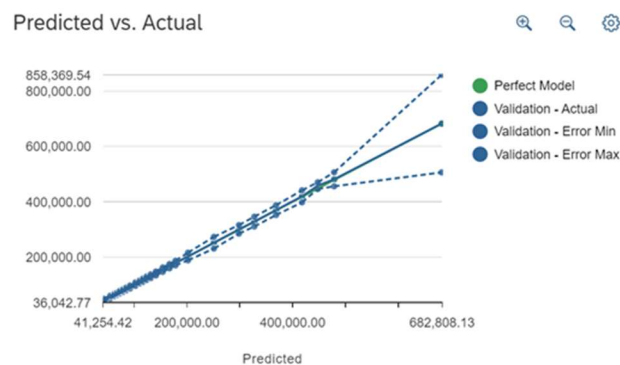


Figure 10. Predicted House Price vs. Actual House Price

### 7. Conclusion

In embarking on this project, we looked at several possible datasets, and the real estate data from Zillow offered broad options. We narrowed our focus to affordability of homes in five key cities across the US for recent MBA graduates in 2024. As we worked through this project, we found that data wrangling is critical for functionality of SAC. Once we cleaned the data, SAC was a useful tool for the analysis and our dataset yielded high accuracy in the regression model, with a prediction confidence of 99.7%.

In the analysis, we saw major variances in housing prices within the US, showing clearly that Los Angeles is not in the affordable range for recent MBA graduates, even for double income households. We were able to identify the regions and cities in the US where affordable housing is available.

Through geo-mapping, charts, time series and predictive regression model, Alaska is most affordable, though the major cities of Houston and Denver are the most appealing in terms of lifestyle and professional opportunities. With an average salary of \$125,000, there are many options outside of Los Angeles for a recent MBA graduate to buy a house and build a life.

### References

[1] Housing Data (n.d.) Zillow. <https://www.zillow.com/research/data/>

[2] Bao, K., Nelson, K., Rampino, N., Walker, A., Williams, A., Yuan, K., White, S. (2023, July). Corporate Recruiters Survey- 2023 Summary Report. Graduate Management Admission Council.

[3] Ng, Wai Yan. Exploring House Price Prediction: A Data Analysis Journey with Tableau. July 28, 2023. <https://www.linkedin.com/pulse/exploring-house-price-prediction-data-analysis-journey-raymond-ng/>

[4] SARTHAK BOBDE, Data Analysis on US House price, March 2021. Kaggle.com. <https://www.kaggle.com/code/sarthakbobde/data-analysis-on-us-house-price>

[5] U.S. Department of Commerce. (2000, Sep 7). Census Regions and Divisions of the United States. U.S. Census Bureau. [https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us\\_regdiv.pdf](https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf)

[6] Latitudes and Longitudes by City SimpleMaps.com. (2024, Jan 30). US Cities Database. SimpleMaps.com. <https://simplemaps.com/data/us-cities>