Utah Single Family Residence Median Sales and Future Median Sales Projections

JDee Warren - 001463291

Western Governors University

Table of Contents

A. Proposal Overview

A.1 Research Question or Organizational Need

Based on Zillow's last ten years of housing sales data, will the median sales price for single-family residences in Utah for the five largest cities increase over the next twelve months?

A.2 Context and Background

A real estate firm (fictitious) has requested data that examines past single-family median sales versus future values. The data, which includes housing sales data collected since 2008 by Zillow for most major cities in the United States, will be used to extract the information needed for the Utah real estate market. The firm would like to understand how the next twelve months of single-family median sales will affect the Utah market and assist in making appropriate marketing decisions for the next twelve months. They would like to target the five largest cities in the state with their marketing campaigns.

A.3 and A3A Summary of Published Works and Their Relation to the Project

An article written by Curt Gresseth, a contributor with KSL radio, provided insight into Utah's housing outlook through 2024. Gresseth writes that high demand, low inventory, and high interest rates are the main reasons for increasing values. A Utah real estate stated, "…lower home prices aren't likely again without the inventory of housing units being replenished to meet demand." For the demand to drop, many more homes must be on the market. Additionally, the agent shared his thoughts on how interest rates would fall, saying, "…for interest rates to drop, the economy has to worsen, as was the case during the COVID-19 pandemic and the U.S. financial crisis of 2007-08". (Gresseth, 2024)

On March 13, 2024, Kinsey Love of Altabank, a local Utah bank, wrote a piece on the current housing market in Utah. Kinsey discusses that demand is still a driving factor of home

values in Utah. And while there has been some relief through slightly lower interest rates and more buyers, home values will remain stable or increase throughout the year. Kinsey goes on to share the thoughts of Altabanks mortgage division President Eric Fisher on the probability of the home values declining. "Barring an unforeseen geopolitical event or something that would cause home prices to drop, I don't see that happening right now," Eric says. "Utah is an extremely resilient state, people are moving here, and demand is huge." (Kinsey, 2024)

And finally, a post by Bhagyesh Behere with Houzeo.com, also makes a case for increasing home values. Behere points out that while demand seems to be easing slightly in the last half of 2023 and early 2024, the demand remains high to keep values increasing. When asked whether it is a buyer's or seller's market, Behere states, "Home prices in Utah are rising and will continue to rise till the supply-demand dynamic changes." He states that Utah is continuing to grow, bringing in new residents due to its low crime and a property tax rate that is much lower than the national average. (Bhere, 2024)

All of the articles above point to one common theme: values will continue to increase. The Utah housing market is still solid, and values should continue to rise for the foreseeable future, barring any economic shifts or significant interest rate changes. As long as there is a low to moderate inventory, housing demand will increase housing values.

A.4 Summary of Data Analytics Solution

To meet the needs of the real estate firm, we will analyze single-family median sales prices from data collected over 16 years. The data is made available by Zillow.com through a .csv file and is open and accessible to the public. We will use Jupyter Notebook and Python coding to clean and analyze the data. To achieve future median sales projections, we will use predictive analytics to calculate what the future twelve months will look like compared to

historical data. The output will include a linear chart outlining historical median sales information and forecasted median sales values. Using forecasting, a bar chart will lay out the future median sales data by cities spread over the twelve months. Finally, to validate the future median sales data, we will perform a null hypothesis test to verify the integrity of the data.

A.5 Benefits and Support of Decision-Making Process

The benefit of the data provided is that it allows stakeholders to make marketing decisions for the upcoming months. If the data shows that median single-family sales are increasing, they can start a marketing campaign for buyers looking to add equity to their homes quickly. On the other hand, the agencies can adjust marketing to show a buyer's market if the median sales price declines.

B. Data Analytics Project Plan

B.1 Goals, Objectives, and Deliverables

The goal of the request from the real estate firm is to use free and public data provided by Zillow to analyze and determine the future median single-family housing sales for upcoming months.

- Objective one is to clean Zillow's housing data. The real estate firm wants to analyze Utah's single-family sales for the five most significant regions (cities). This will be accomplished by removing all other states and regions. We will verify that no remaining null information could distort the data.
    - Deliverable 1.1.1 will be a clean data source that only includes Utah's five largest regions (cities) for the last ten years of sales and no null information.

- Objective two is to provide future sales data to the real estate firm regularly and consistently as data becomes available. This will be done using a linear graph, bar chart, and data source indicating the historical and future projections.
    - Deliverable 2.1.1: An easy-to-read graph showing future median pricing models for the five largest cities in Utah that can be used in marketing materials for prospective buyers and sellers. The first will be a linear graph that contains historical median sales information from 2014-2024. It will also include the projections for the next twelve months of median sales.
    - Deliverable 2.1.2: Provide a bar graph with the projected median sales for the next twelve months. The bar chart will include all five regions with their projected median sales price.
    - Deliverable 2.1.3 Create a data source of future projections illustrating the monthly price increase.
- Objective three will test null hypotheses to confirm our future projection results. The null hypothesis test will examine a two-sample set for different periods (2014-2016 and 2022-2024). The first null test will state that there has been no value increase during given times. The second test will show that there has been a significant increase during given times. They will provide a t-test, p-value, and alpha values. If the p-value is less than the α = alpha value of 0.05, the null hypothesis is rejected and supports the initial hypothesis.
    - Deliverable 3.1.1: Provide findings from the null hypothesis test confirming or rejecting the outcome.

B.2 Scope of Project

The project's scope will be using Jupyter Notebooks, which uses Python coding to analyze historical median single-family home sales. The data will be cleaned, transformed, and broken down into months and years to decrease the size of the dataset to a manageable size. I will then create graphs, charts, and a small dataset to illustrate the projected data for the real estate firm that they may use for their marketing campaigns.

One item that will not be included in the scope will be data surrounding interest rates or economic changes. Due to the infrequent changes in interest rates and economic changes, it would be difficult to include those items due to their uncertainty.

B.3 Standard Methodology

The methodology that will be used for this case is Waterfall. The Waterfall methodology has six steps: requirements, design, implementation, testing, delivery, and maintenance. The Waterfall method is just like a waterfall. You complete one phase and move down to the next phase. This is one of the oldest methodologies for project management and would be a perfect fit for this question.

- **Requirements**

The first phase of the waterfall methodology is the requirements. This requires sitting with the stakeholders and discussing their needs and questions. In this case, the real estate firm has requested data showing if there will be an increase in the median single-family sales values over the next twelve months. They would like to focus their marketing campaign on prospective buyers and sellers with the information. Other requirements will include data being supplied by Zillow.

- **Design**

The next phase is to take the requirements provided by the stakeholders and create a design for their request. This phase will focus on the whiteboard aspects of the requirements, listing out how we will go out gathering, cleaning, selecting pertinent fields, and sorting the data necessary for the request. In this case, the stakeholders would like to see graphs and data showing whether the housing values will continue to increase in the upcoming twelve months.

- **Implementation**

Once the design has been selected, I will utilize Python coding in Jupyter Notebooks to analyze Zillow's data. I will begin by cleaning and transforming the data, selecting the last ten years of data for the five largest cities in Utah. Using the cleaned data, I will create a line graph showing the ten years of data by city, including a trendline. Using predictive analytics, I will then use the data to forecast the sales data for the next twelve months, which will be included in the graph. Finally, I will create a bar graph displaying the future sales values for each city for the next twelve months.

- **Testing**

The testing phase will consist of monitoring Jupyter Notebooks for errors in coding. Additionally, all graphs and charts will be delivered to stakeholders for review and sign-offs from all stakeholders.

- **Delivery**

Once all parties have reviewed and signed off on the final product, a delivery method and timing will be selected. As the information from Zillow is updated monthly, a monthly delivery should be recommended.

- **Maintenance**

Maintenance should occur regularly to verify that the data source and data are accurate. Documentation and versioning control should also be implemented, detailing how the graphs were created and tracking updates for future changes.

B.4 Timeline and Milestones

Milestone	Duration (hours or days)	Projected start date	Anticipated end date
Project Selection / Requirements	3 Days	*08/04/2024*	*08/07/2024*
Project Approved / Design	1 Day	*08/07/2024*	*08/08/2024*
Zillow Data Download	1 Hour	*08/08/2024*	*08/08/2024*
Implementation • Clean • Transform • Predictive Analytics • Graphs • Data	3 Days	*08/09/2024*	*08/14/2024*
Testing	1 Day	*08/14/2024*	*08/16/2024*
Delivery	1 Day	*08/16/2024*	*08/16/2024*
Maintenance	1 Day	*08/19/2024*	*Ongoing/Quarterly*

B.5 Resources and Costs

Resource	**Cost**
Computer	N/A – Used Personal Laptop
Jupyter Notebook (Python)	N/A – Free Application
Workspace	N/A – Home Office
Billed Hours	N/A – Student Hours

B.6 Criteria for Success

The project's success should be determined by three items: the ability to collect usable data, clean and transform the data without any errors, and generate functional graphs and charts that can be used for marketing.

For the first item, it is imperative to select data that can be used on a regular basis and comes from a trusted source. In this case, the data provided is coming from Zillow, which has been a proven source of information in the housing industry and provides this information monthly. The data must also remain relevant, meaning it should always be able to provide new data. If the data is not available regularly, in this case monthly, it will be unusable for the needs of the real estate firm.

The second item, cleaning and transforming the data without any errors, relies on accurate coding. This translates to accurately removing duplicates, outliers, and inconsistencies in the data. Along with accuracy, consistency should be a form of success. Is the information consistent, or are there significant fluctuations in the data that cannot be addressed or answered? If these items are addressed during this phase, the cleaning and transformation of data can be deemed a success.

And lastly, generating functional graphs and charts that the real estate company can use in its marketing campaign. Success for this item is measured in the ability to provide predictive data analysis and generate graphs showing twelve months of future data. The graphs will identify the trajectory of the upcoming months and illustrate a positive or negative increase in median sales values.

C. Design of Data Analytics Solution

In this part, you will discuss the design details of your Capstone data analytics solution.

C.1 Hypothesis

Based on the ten years of data provided by Zillow, the future projections for Utah will show that the median sales values will continue to increase over the next twelve months (barring any economic shifts or significant interest rate changes).

C.2 and C.2.A Analytical Method

The analytical method for this project will use predictive analysis. Catherine Cote of Harvard Business School online provided this description of predictive analysis. “Predictive analytics uses data to predict future trends and events. It uses historical data to forecast potential scenarios that can help drive strategic decisions”. (Cote, 2024)

In this scenario, we will analyze ten years of median single-family home sales to determine how the next twelve months will trend. With predictive analytics, Python uses linear regression to help us determine future trends for the housing market in Utah.

C.3 Tools and Environments

I will be using Python coding in Jupyter Notebooks to complete the analysis. Python offers a free programming platform with access to libraries that range from statistical analysis to data visualization and machine learning, to name just a few.

C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance

The methods and metrics used to analyze the Utah housing market include a linear chart, bar graph, and hypothesis testing using t-stat, p-value, and alpha value.

The linear and bar graphs represent the ten-year median sales data for single-family residences. Additionally, using the regression model, we can apply forecast fit and plot a projection of the next twelve months, which is included in the graphs for representation. This analysis does not take into account any economic or interest rate changes that could occur during the projected outline due to the unpredictability and frequency of those changes.

For the statistical test on the housing data, we will employ the null hypothesis theory to determine if there are any differences if placed into a different scenario. As a tremendous amount of data spread over ten years, we will look at two sample tests with smaller and separate time frames in our hypothesis. Our first hypothesis will state that the values have not increased during the 2014-2016 time frame and the 2022-2024 time frame. An alternative hypothesis will look at the same time frames but indicate that there is a significant difference in the sales prices. If the p-value is less than the α = alpha value of 0.05, the null hypothesis is rejected and supports the initial hypothesis that the median single-family home sales will increase over the next twelve months.

C.5 Practical Significance

The data, visualizations, and statistics will allow the real estate firm to make educated decisions on their marketing moves. All tests and graphs, if done monthly as new data is made available, provide the information they are seeking to meet future financial and business needs.

C.6 Visual Communication

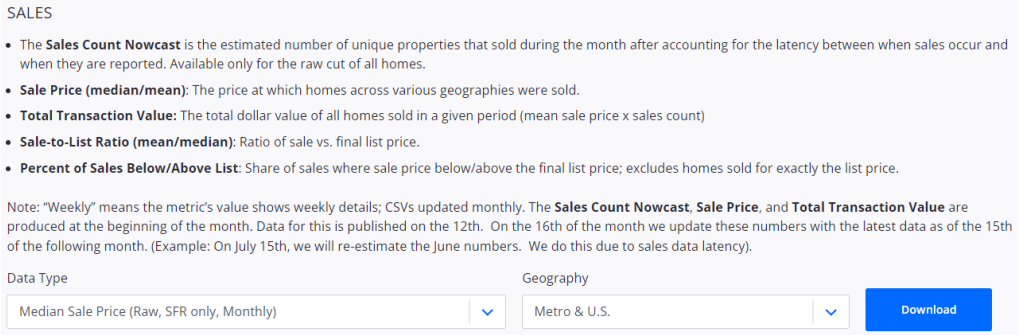
The project will include a linear graph (Median SFR Sales Price Trends and twelve-month Forecast for Utah Cities) that indicates median single-family values for Salt Lake City, Ogden, Provo, Saint George, and Logan regions. The values, which are averaged for the month over ten years, will indicate the increases and decreases in value and provide a future trendline over twelve months for each city. Additionally, a bar graph (twelve-Month Forecast of Median Sale Prices for Utah Cities) will highlight each city's future values, where the y-axis shows the value and the x-axis the city. This can indicate any pricing increases or decreases based on the predictive modeling.

For the null hypothesis testing, a dataset will indicate if the null hypothesis will be rejected, reinforcing our original hypothesis that values will increase over the next twelve months.

D. Description of Dataset

D.1 Source of Data

The data collected for this project is provided by Zillow.com and is accessible to the public. It can be located at https://www.zillow.com/research/data/.



D.2 Appropriateness of Dataset

The dataset encompasses 16 years of housing data for the United States. The dataset contains median sales prices, regions (cities), dates, and rankings for all regions. Along with the data, Zillow offers several research principles. The first research principle listed states, "Zillow Research provides unbiased data and analysis about the housing market in a transparent way." (Zillow, n.d)

D.3 Data Collection Methods

The data was collected by filtering the specific criteria: median sales price (Raw, SFR only, Monthly) from https://www.zillow.com/research/data/. The data is downloaded as a .csv file.

D.4 Observations on Quality and Completeness of Data

The data provided by Zillow is a quality product and provides the necessary fields to complete a thorough analysis of the housing data. While some regions (cities) are missing median sales value, they appear absent in the earliest years, which could be due to how recently those regions made the data available. The data used in this project is complete and available. According to the research principles provided on their About page they state, "Zillow Research benchmarks findings against outside datasets whenever possible to ensure accuracy and appropriate context." (Zillow, n.d)

D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

The data and information used for this project are free and open source to the public. There are no data governance, privacy, security, ethical, legal, or regulatory compliance considerations required for the data being used or stored.

D.5.A

There are no precautions. There is no personally identifiable information contained in any of the datasets, and the data is free and open source to any parties.

References

Gresseth, Curt (2024, February 28). *Utah home prices not likely to fall soon, says real-estate Expert. KSLnewsradio.com.* Retrieved August 8, 2024, from https://kslnewsradio.com/2083844/utah-home-prices-staying-high/

Love, Kinsey (2024, March 13). *Utah Housing Market Trends in 2024.* AltaBank. altabank.com retrieved August 8, 2024, from https://www.altabank.com/utah-housing-market-trends-in-2024

Behere, Bhagyesh (2024, March, 14). *Utah Real Estate Market: Predictions for 2024.* Houzeo.com. retrieved August 8, 2024, from https://www.houzeo.com/blog/utah-real-estate-market/

Cote, Catherine (2021, October 26). *What is predictive analytics? 5 Examples.* Harvard Business School retrieved August 13, 2024, from https://online.hbs.edu/blog/post/predictive-analytics

Zillow (n.d). *About Zillow Research.* Zillow.com Retrieved August 18, 2024, from https://www.zillow.com/research/about-us/