**Portfolio Project Option 1**

**Data Analysis of United States Data**

Joshua Humphreys

Colorado State University Global

Course Code:MIS540

Prof. Mark Bateh

December 4, 2022

**Portfolio Project Option 1**

For my portfolio project I analyzed multiple datasets from the Federal Reserve Economic Data website to answer the following business question: Which month is the most favorable for a home buyer? To answer this question, I collected several datasets for analysis and merged them into a single dataset. In the following pages I will discuss my business question and hypotheses, descriptive analysis, predictive analysis, business analysis, and recommendations for further research.

**Business Questions and Hypotheses**

Again, the initial business problem I set out to answer was to determine the best month to purchase a home in Colorado. To address this problem, I collected data from the Federal Reserve Economic Data website that captures active real estate listings, new real estate listings, median number of days real estate listings spent on the market, median real estate listing prices, and average mortgage rates. The data used in my analysis was grouped by month ranging from July 2017 to September 2022. To address the business problem, I developed the following business questions and corresponding null and alternate hypotheses:

**Q1:** What month do active listings reach their annual high?

**H10**: Active listings do not cycle annually.

**H1A**: Active listings follow an annual cycle and therefore attain an annual maximum.

**Q2:** What month do median days on market reach their annual high?

**H20**: Median days on market do not cycle annually.

**H2A**: Median days on market follow an annual cycle and therefore attain an annual maximum.

**Q3:** What month do median listing prices reach their annual low?

**H30**: Median listing prices do not cycle annually.

**H3A**: Median listing prices follow an annual cycle and therefore attain an annual minimum.

**Q4:** What month do new listings reach their annual high?

**H40**: New listings do not cycle annually.

**H4A**: New listings follow an annual cycle and therefore attain an annual maximum.

**Q5:** What month do mortgage rates reach their annual low?

**H50**: Mortgage rates do not cycle annually.

**H5A**: Mortgage rates follow an annual cycle and therefore attain an annual minimum.

**Q6:** What month is most favorable to home buyers?

**H60**:Real estate market favorability does not cycle annually

**H6A**: The real estate market cycles annually and therefore reaches a most favorable state annually.

Stated differently, my intent was to find which month of the year best balances real estate volume and price. More specifically, I used active listings and new listings to measure real estate supply. I used the median number of days on the market to measure real estate demand. I used median real estate listing prices as an aggregate measure of supply and demand.

**Descriptive Analysis**

In this case, the purpose of my descriptive analysis was to determine, first, whether or not each of these variables cycle annually. Assuming the variables did cycle annually, the next task was to determine if the variables’ cycles were synchronous. Figures 1 through 5below demonstrate the patterns found in the dataset. Each figure contains a series plot of the behavior of each variable over the entire timespan of the dataset as well as series plot of each variable’s behavior grouped by year.

**Figure 1.**

*Series plots of Active Listings*

*Chart, line chart

Description automatically generated Chart

Description automatically generated*

**Figure 2.**

*Series plots of New Listings*

*Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated*

**Figure 3.**

*Series plots of Median Days on Market*

*Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated*

**Figure 4.**

*Series plots of Median Listing Price*

*Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated*

**Figure 5.**

*Series plots of Mortgage Rates*

*Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated*

The figures above clearly indicate that not only is there a regularly occurring cycle of active listings, new listings, days on the market, and listing price, but that cycle occurs annually. Therefore, based on these observations the null hypotheses associated with business questions 1-4 could be rejected. However, Figure 5 clearly indicates that the behavior of mortgage rates does not cycle annually. Therefore, I was unable to reject the null hypothesis for business question 5.

The next step in descriptive analysis is to determine whether the annual cycles of the remaining four variables are synchronized. The simplest way to do this, was simply to produce bar graphs of the average value for each month for each variable. Figures 6 through 9 below demonstrate that active listings reach their average high in August. New listings reach their annual high in June. Median days on the market reaches its average annual low in June. Listing prices reach their low in December. Importantly, Figures 6-9 demonstrate that the variables do not cycle synchronously. This means that predicting the best time to purchase a house in Colorado will require more sophisticated means that simply identifying minimums and maximum values.

**Figure 6.**

*Bar graph of average active listings by month*

*Chart, bar chart, histogram

Description automatically generated*

**Figure 7.**

*Bar graph of average new listings by month*

*Chart, bar chart, histogram

Description automatically generated*

**Figure 8.**

*Bar graph of average median days on market by month*

*Chart, bar chart, histogram

Description automatically generated*

**Figure 9.**

*Bar graph of average of median listing price by month*

*Chart, bar chart, histogram

Description automatically generated*

**Predictive Analysis**

Since the variable cycles are asynchronous, to predict market favorability towards buyers, I elected to create what I called a buyer friendliness value (BFV). The function to calculate BFV takes standardized values for each predictor variable and calculates a weighted average. The theory behind the function is that active listings and new listings are equally weighted measurements of supply. Median number of days on the market is a measurement of demand. This value is then weighted equal to the two measures of supply. These three measurements, then, form an aggregate measurement of supply vs. demand. Since listing price is also an aggregate measurement of supply and demand, it is equally weighted to the other aggregate measure. This yields the following function:

Where AL, NL, DM, LP are standardized values of Active listings, new listings, median days on market, and median listing prices respectively.

The *STDIZE* procedure was first used to calculate standardized values. Then, a data step was used to calculate the BFV for each month in the dataset. Series plots were used to evaluate the BFV (as demonstrated in Figure 10 below). Finally, the monthly mean BFVs were computed to determine the best month to buy a home (as demonstrated in figure 11).

**Figure 10.**

*Series plot for BFV*

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

**Figure 11.**

*Bar graph of average BFV by month*

*Chart, waterfall chart

Description automatically generated*

**Business Insights**

The implications of this analysis for any business involved in real-estate is that the market most favors buyers in January. October was a close second. Additionally, we can also say the market favors sellers in the summer. This analysis, of course, attempts to balance price, demand, and inventory. It’s not surprising, given the disproportionate increase in housing prices vs other variables, that the BFV is more influenced by price than other variables. Also, the series plot demonstrates an accurate picture of what the housing market is doing in Colorado.

**Further Analysis**

My recommendation for further analysis would involve two parts. I think it would be useful to break the data in to separate one-year segments and use those values to calculate standardized values. In that way, we could further isolate annual movement from global movement. Additionally, multiple regression could be used with the original dataset and the calculated BFV to produce a function that takes raw values as input to produce a friendliness value. This value could then be used in real world applications to assess current markets.

**Figure 12.**

*Screenshot of final data manipulation steps*

Graphical user interface, text, application, email

Description automatically generated

**References**

Cody, R. (2019). *A gentle Introduction to Statistics Using SAS Studio.* SAS Institute Inc

Elliott, A.C. & Woodward, W.A. (2016) *SAS essentials: mastering SAS for data analytics (2nd ed.)*. John Wiley & Sons, Inc.

Federal Reserve Bank of St. Louis (n.d.) *FRED economic data.* Economic Research: Federal Reserve Bank of St. Louis. Retrieved on October 28, 2022, from <https://fred.stlouisfed.org/>

SAS Institute. (2022). *Programming Documentation for SAS® 9.4 and SAS® Viya® 3.5.* SAS® Help Center. https://documentation.sas.com/doc/en/pgmsascdc/9.4\_3.5/pgmsaswlcm/home.htm