US Housing Price Affordability Analysis

John E. Hickman

Northwest Missouri State University, Maryville MO 64468, USA
s276335@nwmissouri.edu

Abstract. US housing values have accelerated at paces that has priced many homebuyers out of the market. In addition, incomes have risen on average at a slower annual pace and mortgage interest rates are the highest seen in a generation. According to Urban Institute[6], homebuyer affordability is at an all-time low. This project examined the inputs into housing affordability for US homebuyers, examining 22 years of housing values and attempts to use data analytics capabilities of MS Excel to predict where homeowner affordability is going based on the housing values, wages and mortgage interest rates.

Keywords: Data Analytics · US Housing Prices · Predictive Analytics · Housing Affordability

1 Introduction

Currently home affordability is at at a historical low, between high interest rates and historically high home prices, low supply, high consumer demand are all making it very difficult for first time home buyers and home buying in general. Housing is an engine for the US Economy representing 20 percent of GDP.

1.1 Goals of this Project

This project intendeds to examine and to predict where home price affordability will come back into focus for the majority of home buyers.

1.2 Data Sources

Data from this project has been sourced from Zillow via Kaggle,St. Louis Federal Reserve Bank (FRED) and the Organization of Economic Co-operation and Development (OECD).

2 Approach

This project will look at the relationship between the housing market and price drivers at the aggregate US market level. It will look at the relationship between housing supply and housing price as well as the impact that interest rates have on home affordability. Lastly, it will examine US home buyer income to look at the relationship between home price and borrower affordability in the US.

2.1 Project Phases

To complete this project, specific data needs to be obtained to provide enough history to build an appropriate model on. There are two sets of data that need to be sourced, one to address aggregate US home prices (including pricing drivers) and average US incomes. Once the data is located, analysis will need to be conducted to determine features in the data set that have a correlation to the desired output. Data will need to prepped and examined for missing data elements and decisions will be need to be made to determine how missing, inaccurate or incomplete data will be addressed. Appropriate models will need to be tested for suitability with the data. Regression modeling for example may be suitable given the time series data the project will be working with.

3 Data Collection

This project uses data from three sources to obtain the needed correlation in the model.

Table 1. Data Sources

Short Data Description	Source of Data
Historical US Housing Value Data	Zillow
Historical US Annual Average Salary	OECD
Historical US 10 Year Treasury Yield	St. Louis Federal Reserve Bank

The primary set of data is from Zillow[4], as cleansed and provided via Kaggle .This dataset is a time series set of data tracking the monthly median home value of a given state produced by Zillow, a real estate data firm. The data is a composite of each state for each period's housing value with a focus on single family housing. It uses a bread-basket approach of a mix of different property types and property attributes to arrive at this aggregate value versus accumulating actual home sales for the period which may skew housing prices if too many lower value or higher value properties are sold in any given period. The data is structured in a CSV format that has 285 rows representing 23 years, calculated monthly, of housing price data by state. Each row has 51 records, representing all 50 states and the District of Columbia. The full data set is about 270kb. This is time series data with a date in representing what period the data sample is from. Column headers are alpha characters for the state name. The rest of the data is numeric representing US dollars in xxxxxx.xx format with no thousands separating commas. The decimal exceeds two places. There are some values missing from early years from some states. The missing data is not expected to impact the analysis. Since the project focuses on aggregate affordability in the US, it is expected that the null values can be ignored without impact to the outcome. The missing state data is infrequent and in less populous states. All other data looks in a fairly standard format.

In order to assess affordability, a salary data point is needed to determine how the much income the average US home buyer can afford. The salary data has been sourced from The Organisation for Economic Co-operation and Development (OECD) [2]. Per the OECD's website it "is an international organisation that works to build better policies for better lives. Our goal is to shape policies that foster prosperity, equality, opportunity and well-being for all. We draw on 60 years of experience and insights to better prepare the world of tomorrow." The OECD provides a hub for data and analysis on a variety of economic data. This salary dataset contains data on average annual wages for full-time and full-year equivalent employees in the total US economy. The data is calculated by taking all wages earned and dividing it by the average number of employees in the total economy. The total data set includes salary data for countries outside the United States, but for the purpose of this exercise, only the US data will be leveraged. The data is available in a variety of formats, but will be used in CSV format for simplicity of joining the salary data with the housing price data. No data for the period called upon by the housing data was observed to be missing. The data set has more historical records than the Zillow data set. Older records will be ignored for this analysis.

The final data set used within the model is pricing data on the average 10 year Treasury yields as obtained via the St. Louis Federal Reserve data website, FRED[3]. Treasury yields have direct correlation on the 30 year fixed mortgage rate[1] with the rate spread mirroring the average mortgage interest rate. The 10-year Treasury yield refers to the interest rate on the 10-year U.S. Treasury bond. It is a key indicator in financial markets and is closely watched by investors, economists, and policymakers. The data is available in a variety of formats. For this purpose, we'll leverage a CSV for purposes of being able to join the data easily with the other data sets described above. The data does not have any missing values and is fairly clean. The data set covers a much larger period than used for this analysis. Older records will be ignored for this analysis.

For this project, we'll be using an aggregated view of home prices by year and by state, so there is little in this data set that won't be used. Data for incomes and average mortgage interest rates will be used in addition to home price to reflect additional drivers on affordability.

Since the data is already in a clean CSV form, we expect to retain the data in this format and ingest it directly into the model. Some manipulation will be required to merge the data into one row per time period. For the purposes of the analysis, we'll use a monthly dimension to the data, both the Zillow and Treasury data is already in the time series monthly. Annual incomes will need to appended to the monthly data to reflect the monthly period for annual income. Also since the Zillow data is by state, a composite price will be calculated across all values to arrive an an average price in the US, so that it can be compared to salary and interest rate.

4 Data Prep and Cleansing

The data has been curated from multiple sources as mentioned above.

4.1 Attributes Used in Analysis

The attributes being leveraged for the analysis are:

1. US Average Home Value – This attribute is a calculated average value across all states with reported data for the given month. In most months, there are 51 observations of Home Value from each state and the District of Columbia, with some limited exceptions. The calculated value represents an observation for each state/number of states reporting that period. 99.9 percent of the rows of data contain a value for each of the states, when that data is not observed, the denominator is modified to represent the actual number of observations for that row. The absence of the values is not deemed statistically significant enough to impact the analysis.

2. Average US Wages – This attribute is the value provided by OECD and represents the median salary for a worker for the period the value covers. This attribute as chosen over other measures of income such as household income, as it represents the least possible input to paying a monthly mortgage note. Additionally, measuring household income across the US creates other challenges and varies widely across demographics.

3. 10 Year Treasury Rate – This attribute measures the yield on U.S. Treasury Securities at 10-Year Constant Maturity as a percentage. The 10 Year Treasury Rate has a close relationship with the 30 Year Fixed Mortgage Rate. Per Price Mortgage, "the 10-year Treasury yield and mortgage rates have had a strong correlation of about 0.85 (1 being perfect) over the last decade" suggesting that they move together most of the time.

4. 30 Year Mortgage Rate – The prevailing interest rate for a 30-year fixed rate mortgage for the period.

5. Monthly Principal and Interest Payments – A calculated field based on the median home value x .9 (assumes a 10 percent down payment) x (30 Year Mortgage Rate/12)/ (1-(1/(1+interest rate/12/360)) to arrive at an average monthly payment based on the prevailing home value and interest rate of the period.

6. Total Monthly Payment – A calculated field that leverages the Monthly Principal and Interest Payment result x 1.75 percent of Home Value to estimate Taxes and Insurance to arrive at a total monthly payment a borrower might pay. The 1.75 percent estimate was a suggested value based on Urban Institute's similar analysis to estimate home affordability.

7. Affordability Index – a calculated field that assesses if median wages are sufficient to cover the monthly mortgage payments given the periods wages, home values and interest rate. A value less than 100 means that the salary is insufficient and a value over 100 means that the salary is sufficient to cover the monthly payment. This index is using a value of 30 percent of salary to cover the monthly payment, a measure used by [5]HUD to determine mortgage eligibility. The lower the number below 100, is an assessment of the gap between affordability and unaffordability.

4.2 Data Prep

This project is attempting to determine both at what future level of wages and/or monthly payment, affordability becomes back and an attempt to predict at what level of wages and/or monthly payment changes would be needed for this occur. Affordability is determined by wages and monthly payment (which is a factor of home value and mortgage interest rate). Total Monthly Payment and Average US Wages are the independent variables, and the Affordability Index is the dependent variable.

4.3 Data Challenges and Approach

Since this is time series data, all the data must be set to be normalized to the same period to be able to perform analysis on the data. The time series we're using for this analysis is monthly, with data from January 2000 to December 2022. The Zillow Home Value data is already in a monthly format and includes data through October 2023. We have dropped the records from 2023 as there is incomplete data on Salary information for 2023 available from OECD to compare home affordability to. The OECD data was annual and since the data from Zillow is monthly, the annual value is used for the monthly values of the same year. The Treasury Bill rate obtained from St. Louis Federal Reserve is already in monthly format and covers the subject period of this analysis and requires no modification. No values were missing. The last main input to the analysis is the 30-year fixed mortgage rate obtained from Freddie Mac's weekly survey (retrieved from St. Louis Federal Reserve Bank). Given that this data is weekly and the periods for the analysis are monthly, the rate for the first week of each month was used. Given that rates do not fluctuate widely week over week and when they do, a monthly snap shot will reflect this swing. All of the attributes from the above have been collected into a single Excel workbook to easily compare the data visually and for the ease of creating the calculations described above.

5 Exploratory Data Analysis

5.1 Exploratory Data Analysis Objective

The objective of performing exploratory data analysis for this project is to gain insights into the factors influencing home affordability, utilizing a comprehensive dataset that spans a 22 year period for the years 2000-2022. The dataset includes monthly home value data from Zillow, annual income data (expressed monthly) provided by OECD, monthly 30 year mortgage interest rate data and 10 year Treasury yield data over the 22 year period.. The goal of the analysis is to understand the trends in home values, income, and interest rates to inform the development of a predictive model for home affordability.

5.2 Analysis Techniques

Multiple analysis techniques were employed to extract meaningful information and insights from the data.

Descriptive Statistics Descriptive statistics were used to provide an overview of the central tendencies, dispersion of the data, and distributions of the variables 1 .Histograms were leveraged to understand distribution of the variables 2 . Figures below provide insight into distribution and statistics about the data leveraged within the workbook.

Time Series And Graphical Correlation Analysis Time series analysis was conducted comparing home value, annual wages, and interest rates to identify trends and patterns in affordability over the 22-year period examined. Charts were generated to depict the trends and correlations within the data 3.

5.3 Preliminary Results

The analysis revealed interesting insights into the dynamics of home affordability. Time series analysis in Figure 5 indicated some month to month and year to year fluctuations in home values and its relationship with affordability 6 , suggesting potential cyclical patterns or external influences impacting the real estate market. Correlation analysis visuals demonstrated the expected inverse relationship between interest rates and home affordability, with higher interest rates leading to decreased affordability. 3 Additionally, correlations between home values and wages were explored to understand the income elasticity of home prices. 5

5.4 Useful Findings and Next Steps

Several useful findings have emerged. The trends in home values shown over time provide context for understanding the value changes of the housing market. The correlation between the 30 year mortgage interest rate and home affordability reaffirms the significance of rate fluctuations in influencing housing market dynamics. Preliminary insights into the relationship between income and home values demonstrate how widely the two can be swing apart from each other with some influence on affordability. Interestingly wages are the most steady of the data, recognizing that the data is only refreshed annually, it never-the-less trend-wise rises consistently and steadily each year. The hypothesis is that future projections on income growth will also be steady and have less an influence on affordability than home prices and interest rate.

6 Model Design and Selection

For this project, we are leveraging predictive analytics capabilities in Microsoft Excel. This choice was made based on the number of transformations and calculations needed to arrive at the Affordability Index [?], which is derived from Home Values, Wages and Interest Rate as described in the data section of this document.

6.1 Data Collection/Sheet Organization

: Organized the data from the previously disclosed sources on home values, wages and interest rates, organizing them into a logical manner across a monthly time horizon. Since the project involves predicting affordability, prepare the calculations and workbook to understand how the variables work together to predict affordability.

6.2 Data Prep and Pre-processing

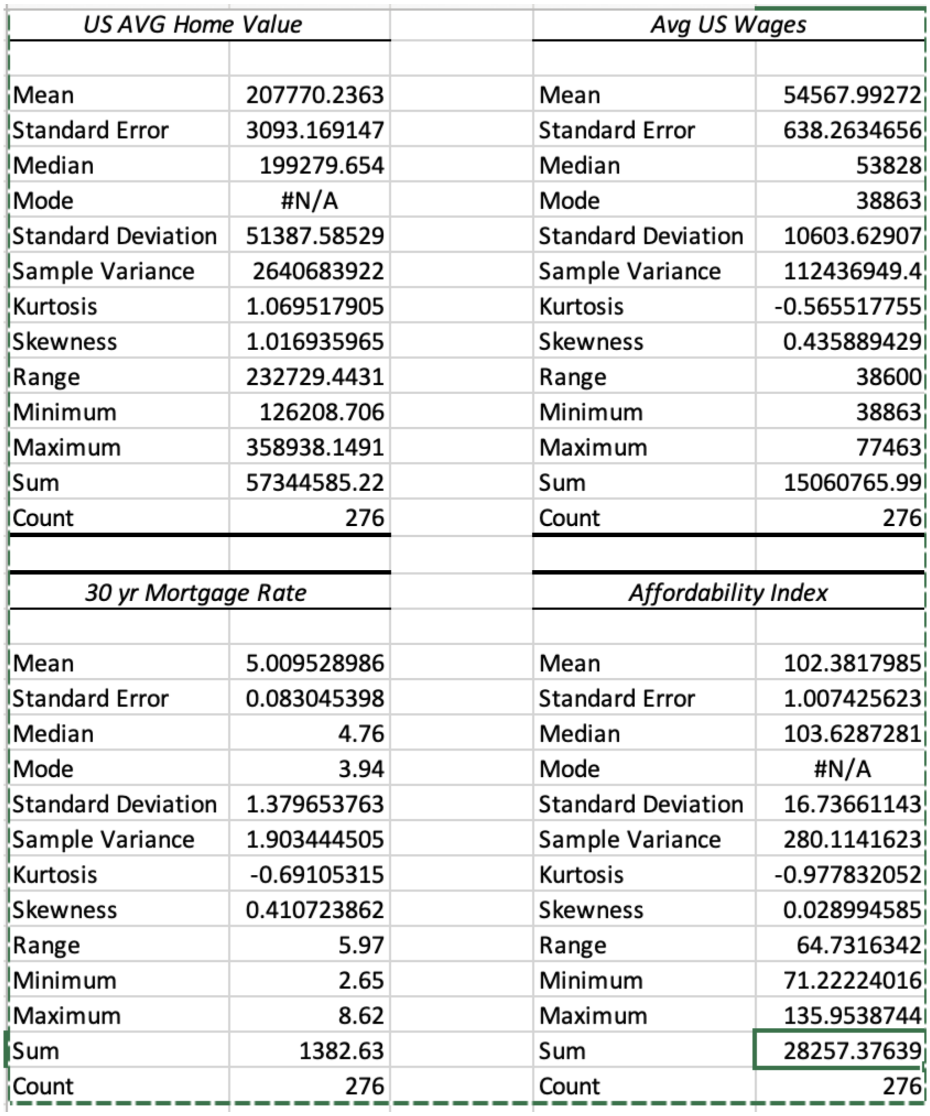
Adjusted the calculations for missing data and formatted the data. For example, interest rate had to be adjusted to convert from 5 percent to .05 for the calculations to be accurate.

6.3 Feature Engineering

Identified the relevant features that are inputs into housing affordability.Calculated the housing affordability index, which is how much of an average monthly mortgage payment an average income could pay assuming a 30 percent of monthly income were applied to housing payment.

6.4 Model Selection

Experimented with a few different methods of forecasting. With the project, we're attempting to predict housing affordability, which is the output of housing value, wages and interest rate. Each one requires calculations and then calculations have to be performed on the predicted values for each feature that is



US AVG Home Value			*Avg US Wages*	
Mean	207770.2363		Mean	54567.99272
Standard Error	3093.169147		Standard Error	638.2634656
Median	199279.654		Median	53828
Mode	#N/A		Mode	38863
Standard Deviation	51387.58529		Standard Deviation	10603.62907
Sample Variance	2640683922		Sample Variance	112436949.4
Kurtosis	1.069517905		Kurtosis	-0.565517755
Skewness	1.016935965		Skewness	0.435889429
Range	232729.4431		Range	38600
Minimum	126208.706		Minimum	38863
Maximum	358938.1491		Maximum	77463
Sum	57344585.22		Sum	15060765.99
Count	276		Count	276
30 yr Mortgage Rate			*Affordability Index*	
Mean	5.009528986		Mean	102.3817985
Standard Error	0.083045398		Standard Error	1.007425623
Median	4.76		Median	103.6287281
Mode	3.94		Mode	#N/A
Standard Deviation	1.379653763		Standard Deviation	16.73661143
Sample Variance	1.903444505		Sample Variance	280.1141623
Kurtosis	-0.69105315		Kurtosis	-0.977832052
Skewness	0.410723862		Skewness	0.028994585
Range	5.97		Range	64.7316342
Minimum	2.65		Minimum	71.22224016
Maximum	8.62		Maximum	135.9538744
Sum	1382.63		Sum	28257.37639
Count	276		Count	276

Fig. 1. Workbook Descriptive Statistics

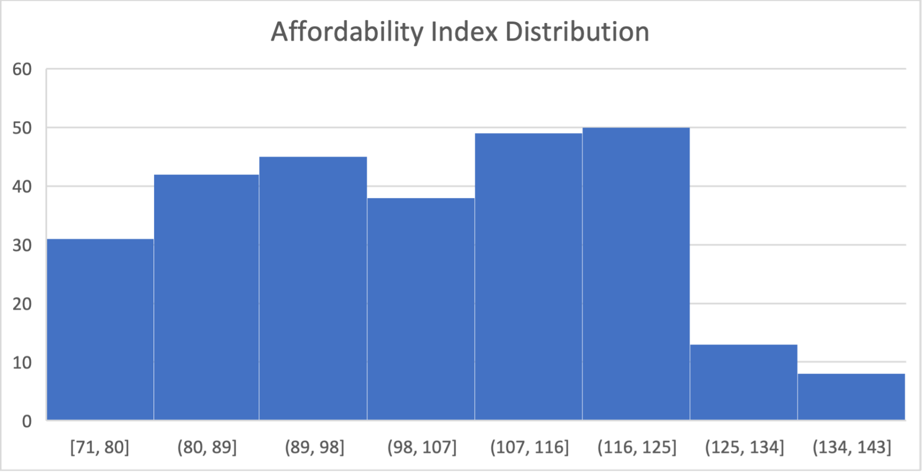


Fig. 2. Affordability Index Histogram

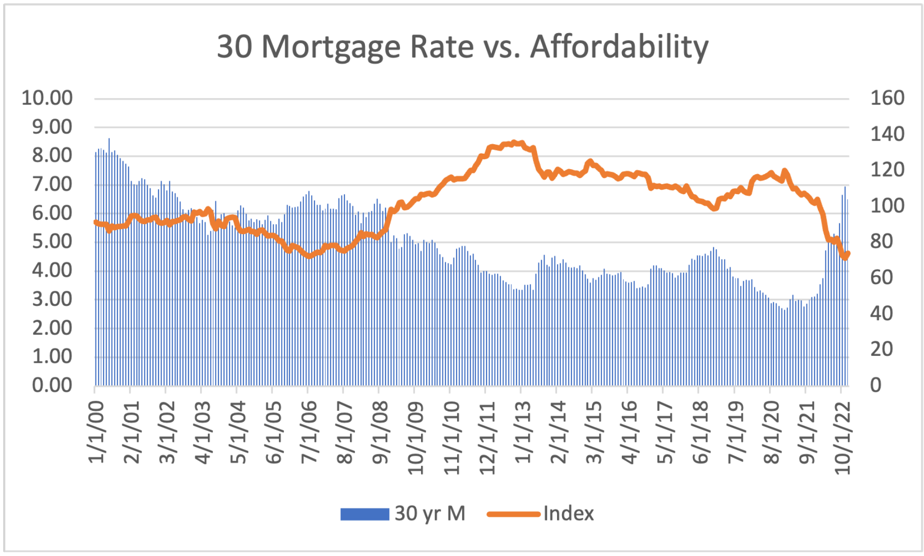


Fig. 3. 30 Year Mortgage Rate vs. Affordability 2000-2022

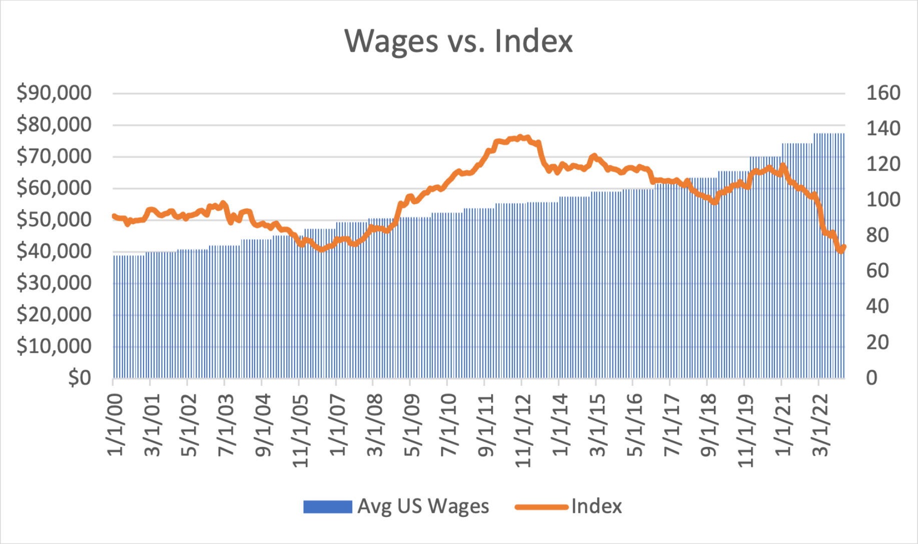


Fig. 4. Average US Annual Wages vs. Affordability Index 2000-2022

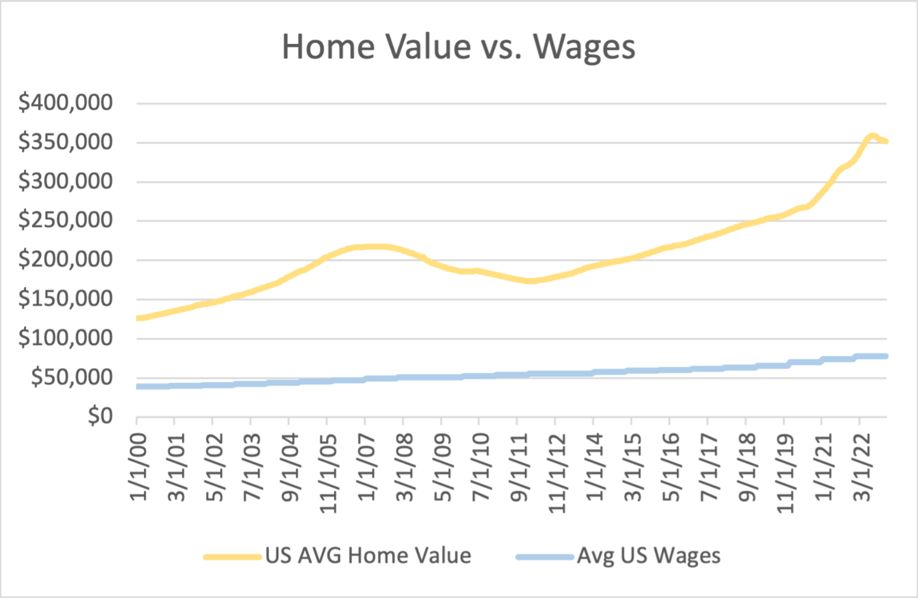


Fig. 5. Average US Home Values vs. Average US Annual Wages 2000-2022

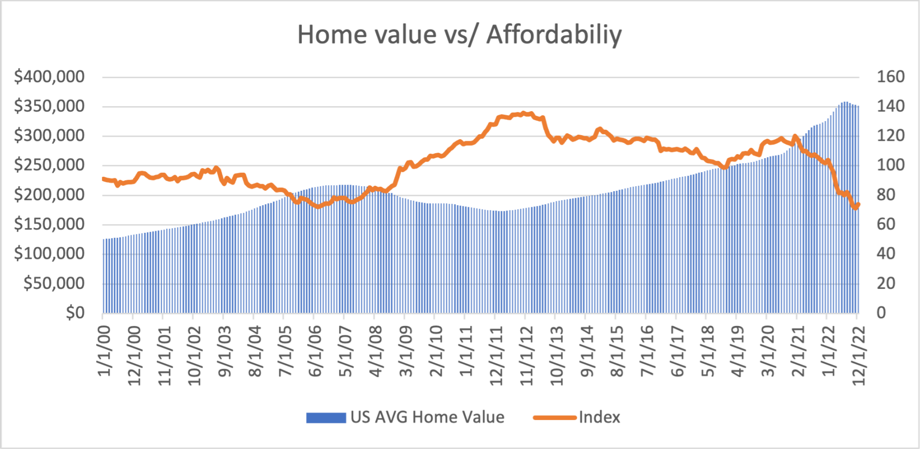


Fig. 6. Average US Home Values vs. Affordability Index 2000-2022

an input into affordability. Models evaluated were linear regression, exponential smoothing, double exponential smoothing, quadratic models. The models leveraged the function and built-in capabilities of Excel, Data Analysis Toolpaks and Statistical Analytics tools. 7

6.5 Model Selection

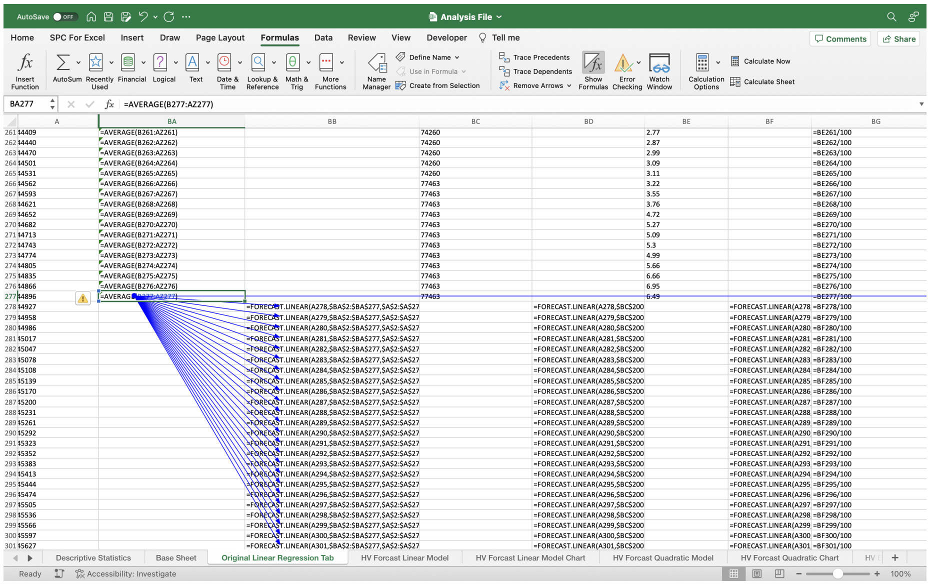
We ran several models starting with Housing Values based on the 2000-2022 dataset. Linear Regression was the original model that was tested with the full dataset and then reduced datasets. We found better results with smaller datasets limited to 50-80 observations of housing values. 6

6.6 Testing Process

Model forecasts for housing values were compared to actual results for 2023 YTD from the dataset on hand.

6.7 Limitations

There may be variables that are difficult to model in this analysis, or have features and attributes that are influenced by data sources outside those in scope for this analysis. For example, unemployment rate may be a variable that is hard to predict and its relationship may be challenging to model with the selected data sets.



BA277 =AVERAGE(B277:AZ277)

	A	BA	BB	BC	BD	BE	BF	BG
261	44409	=AVERAGE(B261:AZ261)		74260		2.77		=BE261/100
262	44440	=AVERAGE(B262:AZ262)		74260		2.87		=BE262/100
263	44470	=AVERAGE(B263:AZ263)		74260		2.99		=BE263/100
264	44501	=AVERAGE(B264:AZ264)		74260		3.09		=BE264/100
265	44531	=AVERAGE(B265:AZ265)		74260		3.11		=BE265/100
266	44562	=AVERAGE(B266:AZ266)		77463		3.22		=BE266/100
267	44593	=AVERAGE(B267:AZ267)		77463		3.55		=BE267/100
268	44621	=AVERAGE(B268:AZ268)		77463		3.76		=BE268/100
269	44652	=AVERAGE(B269:AZ269)		77463		4.72		=BE269/100
270	44682	=AVERAGE(B270:AZ270)		77463		5.27		=BE270/100
271	44713	=AVERAGE(B271:AZ271)		77463		5.09		=BE271/100
272	44743	=AVERAGE(B272:AZ272)		77463		5.3		=BE272/100
273	44774	=AVERAGE(B273:AZ273)		77463		4.99		=BE273/100
274	44805	=AVERAGE(B274:AZ274)		77463		5.66		=BE274/100
275	44835	=AVERAGE(B275:AZ275)		77463		6.66		=BE275/100
276	44866	=AVERAGE(B276:AZ276)		77463		6.95		=BE276/100
277	44896	=AVERAGE(B277:AZ277)		77463		6.49		=BE277/100
278	44927		=FORECAST.LINEAR(A278,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A278,$BC$200		=FORECAST.LINEAR(A278	=BF278/100
279	44958		=FORECAST.LINEAR(A279,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A279,$BC$200		=FORECAST.LINEAR(A279	=BF279/100
280	44986		=FORECAST.LINEAR(A280,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A280,$BC$200		=FORECAST.LINEAR(A280	=BF280/100
281	45017		=FORECAST.LINEAR(A281,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A281,$BC$200		=FORECAST.LINEAR(A281	=BF281/100
282	45047		=FORECAST.LINEAR(A282,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A282,$BC$200		=FORECAST.LINEAR(A282	=BF282/100
283	45078		=FORECAST.LINEAR(A283,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A283,$BC$200		=FORECAST.LINEAR(A283	=BF283/100
284	45108		=FORECAST.LINEAR(A284,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A284,$BC$200		=FORECAST.LINEAR(A284	=BF284/100
285	45139		=FORECAST.LINEAR(A285,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A285,$BC$200		=FORECAST.LINEAR(A285	=BF285/100
286	45170		=FORECAST.LINEAR(A286,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A286,$BC$200		=FORECAST.LINEAR(A286	=BF286/100
287	45200		=FORECAST.LINEAR(A287,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A287,$BC$200		=FORECAST.LINEAR(A287	=BF287/100
288	45231		=FORECAST.LINEAR(A288,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A288,$BC$200		=FORECAST.LINEAR(A288	=BF288/100
289	45261		=FORECAST.LINEAR(A289,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A289,$BC$200		=FORECAST.LINEAR(A289	=BF289/100
290	45292		=FORECAST.LINEAR(A290,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A290,$BC$200		=FORECAST.LINEAR(A290	=BF290/100
291	45323		=FORECAST.LINEAR(A291,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A291,$BC$200		=FORECAST.LINEAR(A291	=BF291/100
292	45352		=FORECAST.LINEAR(A292,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A292,$BC$200		=FORECAST.LINEAR(A292	=BF292/100
293	45383		=FORECAST.LINEAR(A293,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A293,$BC$200		=FORECAST.LINEAR(A293	=BF293/100
294	45413		=FORECAST.LINEAR(A294,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A294,$BC$200		=FORECAST.LINEAR(A294	=BF294/100
295	45444		=FORECAST.LINEAR(A295,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A295,$BC$200		=FORECAST.LINEAR(A295	=BF295/100
296	45474		=FORECAST.LINEAR(A296,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A296,$BC$200		=FORECAST.LINEAR(A296	=BF296/100
297	45505		=FORECAST.LINEAR(A297,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A297,$BC$200		=FORECAST.LINEAR(A297	=BF297/100
298	45536		=FORECAST.LINEAR(A298,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A298,$BC$200		=FORECAST.LINEAR(A298	=BF298/100
299	45566		=FORECAST.LINEAR(A299,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A299,$BC$200		=FORECAST.LINEAR(A299	=BF299/100
300	45597		=FORECAST.LINEAR(A300,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A300,$BC$200		=FORECAST.LINEAR(A300	=BF300/100
301	45627		=FORECAST.LINEAR(A301,$BA$2:$BA$277,$A$2:$A$27		=FORECAST.LINEAR(A301,$BC$200		=FORECAST.LINEAR(A301	=BF301/100

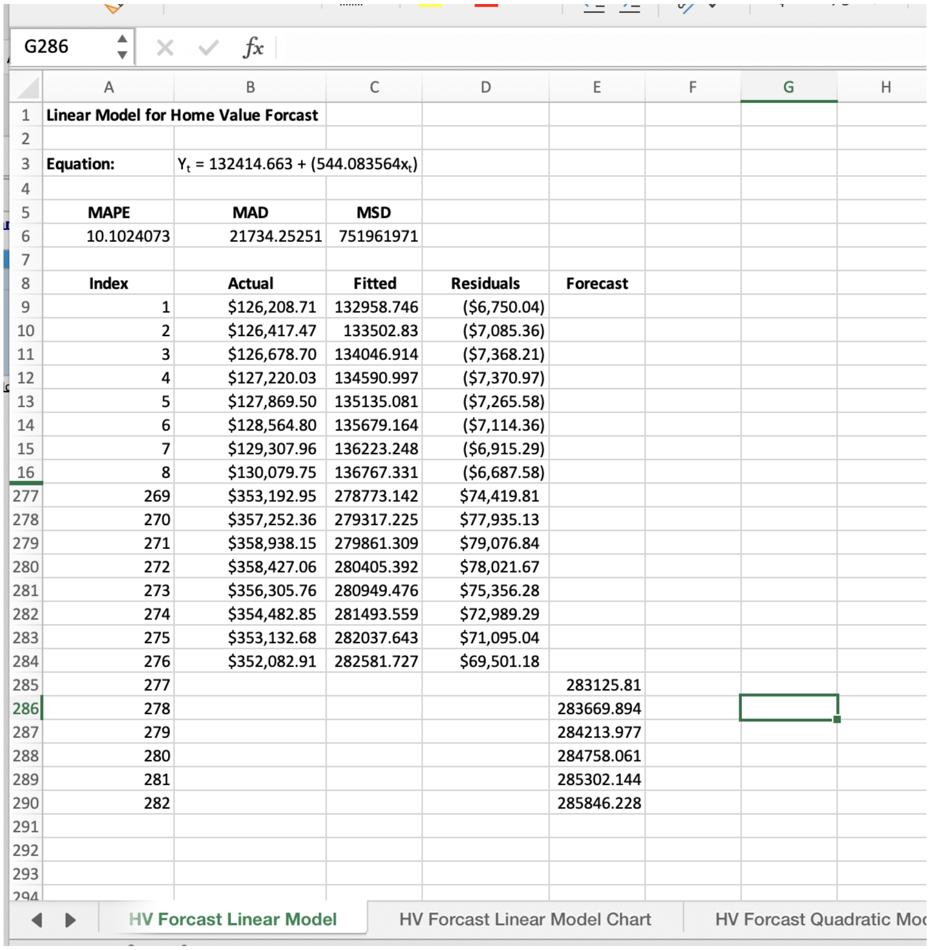
Fig. 7. Regression Model Formulas

7 Model Output and Analysis

Outputs of the models were compared to each other and the 2023 housing values that were not included as inputs into the model. None of the results were close to actual nor did they follow the actual trend lines. Results were especially poor for interest rates, a large input driver to affordability.

We leveraged several models mentioned above and reduced the data sets in an attempt to get better results, however all models and reduced data sets produced unacceptable results. For example, using Linear Regression we used the full dataset for housing values (278 monthly observations) and reduced to 250, 200, 150 and 100 all of which produced marginally better results. However, using the same reduction of observations for interest rate created poorer results. Given that both sets of data need to be synced over the same time horizon in order to product credible results, the time periods evaluated need to be consistent.

The models used did not prove to be statistically significant and produced less than ideal outputs and while, we didn't take the work further to calculate the affordability index, given the poor results of the input attribute forecasts. All had MASE results of 10 which is deemed inadequate. All outputs were compared to 2023 actual results and found to not follow the trend of actual results. Any affordability index calculations based on this output would provide overly optimistic results. The exponential smoothing model had the best results for housing values but produced poor outputs for the mortgage interest rate predictions, which would have skewed the affordability results.



	A	B	C	D	E	F	G	H
1	**Linear Model for Home Value Forcast**							
2								
3	**Equation:**	Y_t = 132414.663 + (544.083564x_t)						
4								
5	**MAPE**	**MAD**	**MSD**					
6	10.1024073	21734.25251	751961971					
7								
8	**Index**	**Actual**	**Fitted**	**Residuals**	**Forecast**			
9	1	$126,208.71	132958.746	($6,750.04)				
10	2	$126,417.47	133502.83	($7,085.36)				
11	3	$126,678.70	134046.914	($7,368.21)				
12	4	$127,220.03	134590.997	($7,370.97)				
13	5	$127,869.50	135135.081	($7,265.58)				
14	6	$128,564.80	135679.164	($7,114.36)				
15	7	$129,307.96	136223.248	($6,915.29)				
16	8	$130,079.75	136767.331	($6,687.58)				
277	269	$353,192.95	278773.142	$74,419.81				
278	270	$357,252.36	279317.225	$77,935.13				
279	271	$358,938.15	279861.309	$79,076.84				
280	272	$358,427.06	280405.392	$78,021.67				
281	273	$356,305.76	280949.476	$75,356.28				
282	274	$354,482.85	281493.559	$72,989.29				
283	275	$353,132.68	282037.643	$71,095.04				
284	276	$352,082.91	282581.727	$69,501.18				
285	277				283125.81			
286	278				283669.894			
287	279				284213.977			
288	280				284758.061			
289	281				285302.144			
290	282				285846.228			
291								
292								
293								
294								

Fig. 8. Housing Value Linear Regression Forecast Model (Truncated)

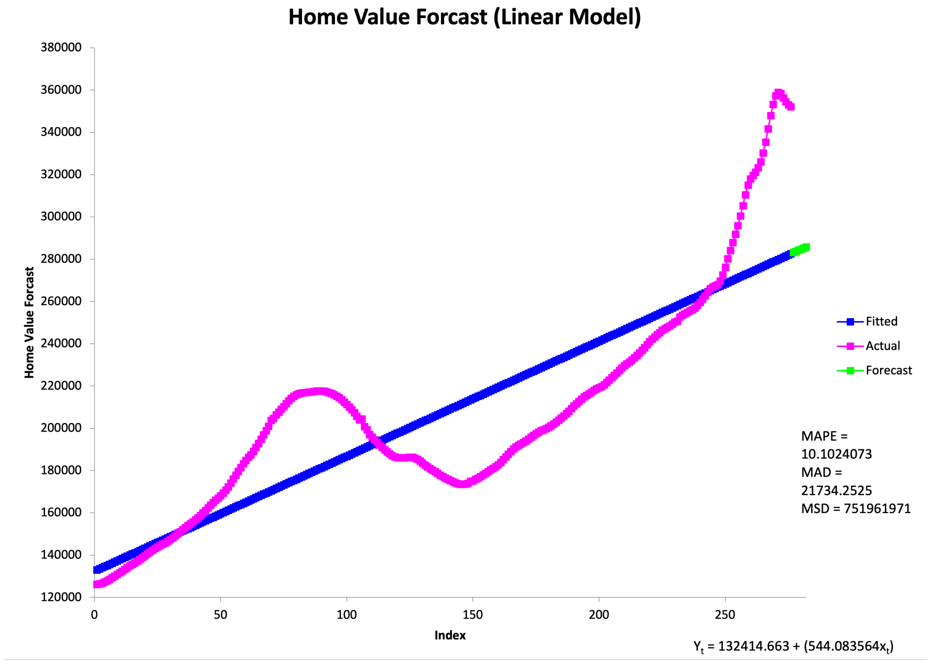


Fig. 9. Housing Values Linear Regression Forecast Chart

8 Conclusions

The results are unacceptable for a prediction model. I have attached screenshots of both chart and model output for the Linear Regression model (truncated) to share the results. Other models had similar results. We did see some improved results for exponential smoothing models when predicting interest rates but still fell short of expectations.

Based the model results, we have concluded that none of these models will produce acceptable results given the prediction results will ultimately product an inaccurate prediction on housing affordability. All models used are contained within the workbook for interrogation. Project Link

9 Future Work

This project requires significant investments in time and additional machine learning techniques in order to make the affordability index predictions work. Different modeling techniques are required to model each variable separately with more credible output and then calculate the results. This particular project attempted to look at housing as a whole system and it may be more meaningful to look at individual markets or metros to prove out other models. Given more time, we would recommend spending more time on housing value predictions first. Also, we would explore other attributes that may have more influence on mortgage interest rates. We examined Treasury Bills, but it's possible in a high inflationary period like the one experienced in 2021-2023, other drivers may need to be examined.

▯

References

1. 30-year fixed rate mortgage average in the united states. https://fred.stlouisfed.org/series/MORTGAGE30US, accessed: November 12, 2023
2. Average annual wages dataset. https://stats.oecd.org/Index.aspx?DataSetCode=AV_AN_WAGE#, accessed: November 12, 2023
3. Market yield on u.s. treasury securities at 10-year constant maturity, quoted on an investment basis dataset. https://fred.stlouisfed.org/series/GS10, accessed: November 12, 2023
4. Zillow home value index dataset. https://www.kaggle.com/datasets/robikscube/zillow-home-value-index, accessed: November 12, 2023
5. HUD: Defining housing affordability (2017), https://www.huduser.gov/portal/pdredge/pdr-edge-featd-article-081417.html#:~:text=Keeping%20housing%20costs%20below%2030,to%20be%20housing%20cost%20burdened., accessed: Novermber 19, 2023
6. Institute, U.: Housing finance at a glance monthly chartbook october 2023 (2023), https://www.urban.org/sites/default/files/2023-10/Housing%20Finance-At%20a%20Glance%20Monthly%20Chartbook-October%202023.pdf, accessed: Novermber 19, 2023

7. Mortgage, P.: How the 10 year us treasury note impacts mortgage rates (2023), https://pricemortgage.com/10-year-treasury-mortgage-rates/#:~:text=Key%20stats%20are%20telling%20too,around%2085%25%20of%20the%20time., accessed: Novermber 19, 2023
8. Vidhya, A.: Predictive modeling in excel – how to create a linear regression model from scratch (2020), https://www.analyticsvidhya.com/blog/2020/06/predictive-modeling-excel-linear-regression/., accessed: Novermber 24, 2023