When to Buy a House in San Francisco

Eric Castellanos, Russell Chan, Katelyn Wu

College of Business, San Francisco State University

DS 604: Business Forecasting

Professor Yabing Zhao

December 16, 2022

Introduction

For this forecasting project, we will look into the housing market. We are interested in investing in the real estate market so we analyzed the prices of homes in San Francisco and the potential influences on the fluctuation of home prices over time. The objective of this project is to figure out when it is the best month in a given year to purchase a home in San Francisco.

The Data

Our data was taken from Zillow, a real-estate company. We decided to use the raw data of median sales prices for our project. The data provided was from January 2018 to October 2022 of all homes sold in San Francisco.

The Problem

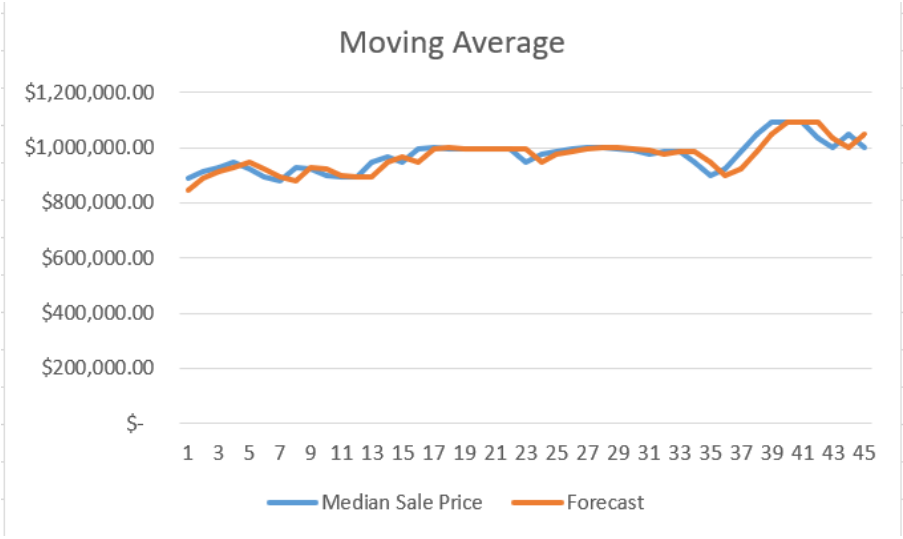
The problem for our project is to figure out which forecasting method is best for determining which month in a given year to invest in purchasing a San Francisco home.

Forecasting Methods

The forecasting methods we used in our project included simple moving average, simple exponential smoothing, Holts forecast, Winters forecast, and Decomposition. We computed and compared the Mean Squared Error (MSE) and Root Mean Square Error (RMSE) for each forecasting method to help us draw our conclusion of the best forecasting method for our dataset.

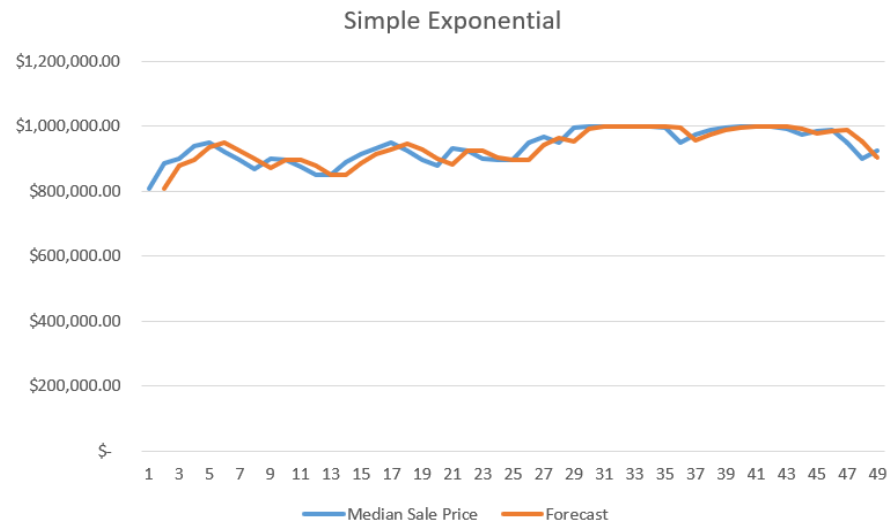
Moving Average

The first forecast we applied to our data is the simple moving average. We want to smooth the data to make the trend more visible. The analysis of the moving average is there is a slight positive trend in the data. The RMSE is 30,264.62 which is extremely high, so we wouldn't consider the moving average to be a significant model.



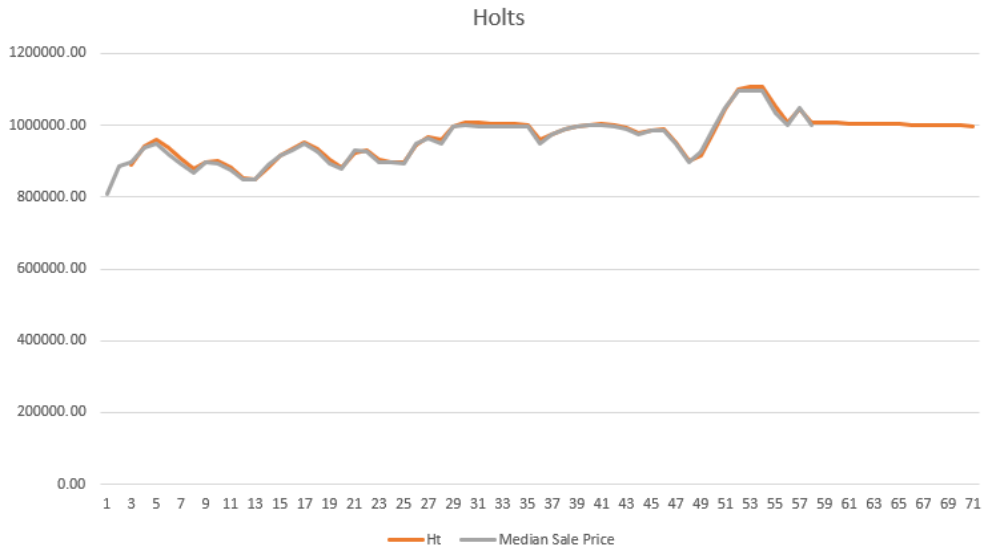
Simple Exponential Smoothing

Second, we tried simple exponential smoothing (SES). This forecasting method is appropriate for data with no seasonality, and no trend. SES is computed using two predictors: the last observed result and the last forecast. It uses a single parameter alpha, to weigh each predictor to output the most accurate forecast. Our data found the best alpha is 0.9, which means it mostly relies on the last observation to forecast. The RMSE is high at 26,716.34, which is also not an acceptable error rate.



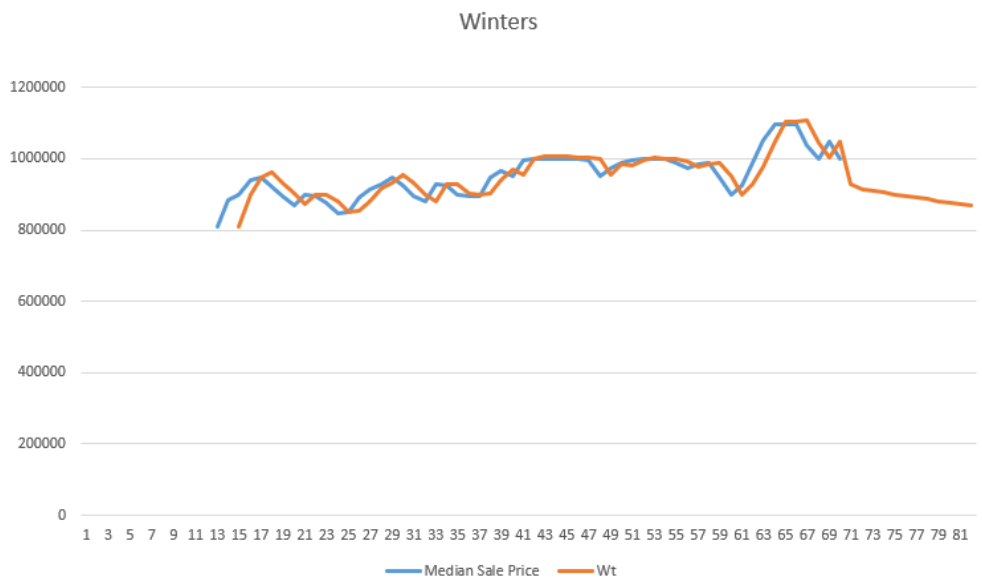
Holts Forecast

Next we applied Holt's exponential smoothing, which is similar to the last forecast with an added trend factor. Holt's forecast uses two parameters, alpha and beta to estimate. There is a noticeable positive trend in our data, so we expect to see some improvement in our forecast. Visually, the forecast is extremely accurate, it looks as if it's tracing the observed data. The RMSE is 7075.69, which is much smaller than the previous forecasts. The best parameters found are alpha = 0.9 like from the previous forecast, and beta = 0.1.



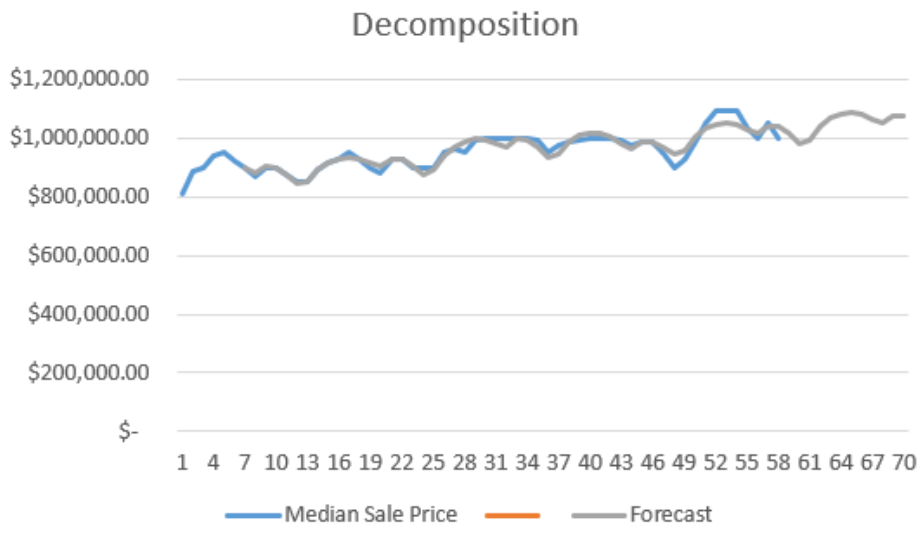
Winters Forecast

Winter's forecast is similar to Holt's forecast with an additional measure of seasonality. Winters needs three parameters to compute forecasts, alpha, beta, and gamma. They are necessary to estimate the level, trend and seasonality, respectively. We notice there is slight seasonality in 2018 and 2019, but then it becomes random, so we don't have many expectations for this forecast. The RMSE is 33105.44, which is high. We can assume the high error rate is caused by the lack of seasonality in the middle and end of the data.



Decomposition

Finally, we apply Decomposition forecasting. Decomposition looks after trend, seasonality, cycle, and any irregularities. Fitting all those categories, we expect this method to give us the lowest RMSE, however the error rate is moderate at 19686.07. We believe that this high rating is due to the absence of cycles during the pandemic. Decomposition failed to predict an accurate forecast because it fails to account for the unpredictable cycle the pandemic poised on the housing market.

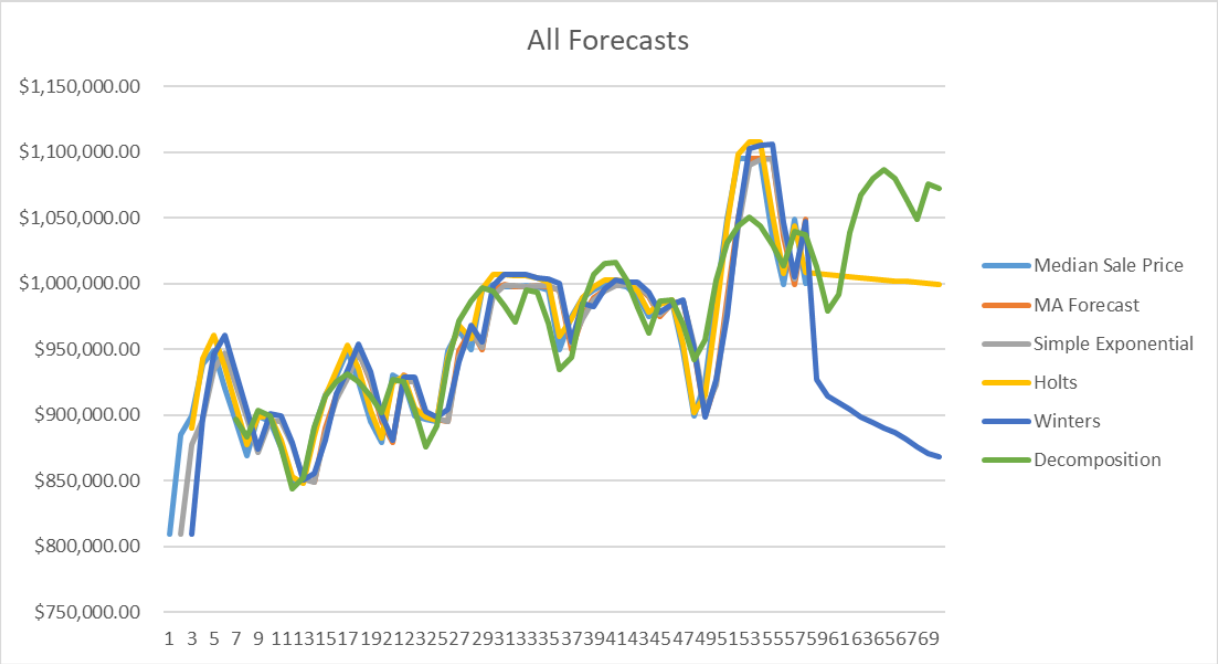


Summary and Analysis

After running every forecast, we see that Holt's delivers the best RMSE. Holt's accurately depicts the upward trend and seasonality of the graph, however, we expected decomposition to give us the best forecast because it can predict not only seasonality and trend, but also irregularities and cycles. Simple exponential only relies on the previous forecast, so it won't be able to predict an accurate forecast especially for an unstable data we have.

Forecast	**RMSE**
Moving Average	30164.62
Simple Exponential Smoothing	26716.34
Holt's Forecast	7075.69
Winters Forecast	33105.44
Decomposition	19686.07

The best forecast found is Holt's forecast, which predicts exponentially with a trend factor. A slight positive trend is observed in our data, and our RMSE is the lowest at 7075.69. The second best forecast we tested is decomposition with an RMSE of 19686.07. Even though decomposition was the next lowest RMSE, we won't count it because it is still drastically bigger than Holts with a difference of 12,610. Decomposition takes trend, seasonality, cycles, and randomness into account to predict. Our data lacks cycles and seasonality due to the pandemic around the start of 2020, so we have slight error issues. The rest of the forecasts we applied had a high error rate, and we would not recommend using them to forecast the housing market.



Conclusion and Recommendations

From the data and the forecasts we collected, we determined that November, December and January are the best times to purchase a house in San Francisco from an average rating of less than 1 SI for those months. It should be noted that this is heavily exaggerated due to only observing the past four years with a pandemic affecting half of it. If we were to observe data from the past decade we may get a better forecast and conclusion on which months are the best months to purchase a house. However, just like any investment, the housing market is unpredictable due to so many factors; the economy's health, climate change, and housing demand, just to name a few.