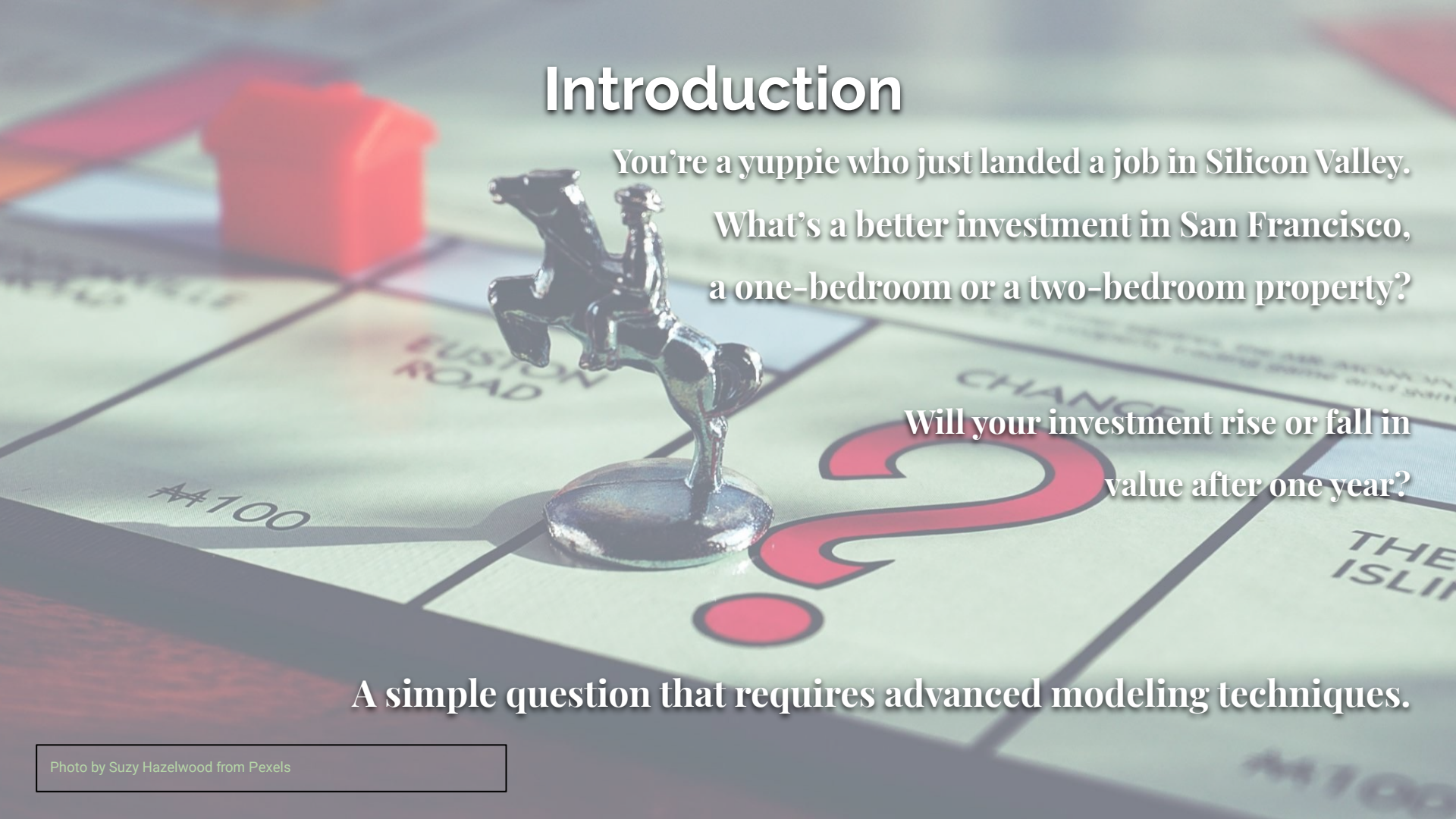
Future Prices of Property

Forecasting with Time Series
by Alexander Xin & Jonathan Silverman

Introduction



You're a yuppie who just landed a job in Silicon Valley.

What's a better investment in San Francisco,
a one-bedroom or a two-bedroom property?

Will your investment rise or fall in
value after one year?

A simple question that requires advanced modeling techniques.

Photo by Suzy Hazelwood from Pexels

Photo by Ahmed Rabea from Pexels

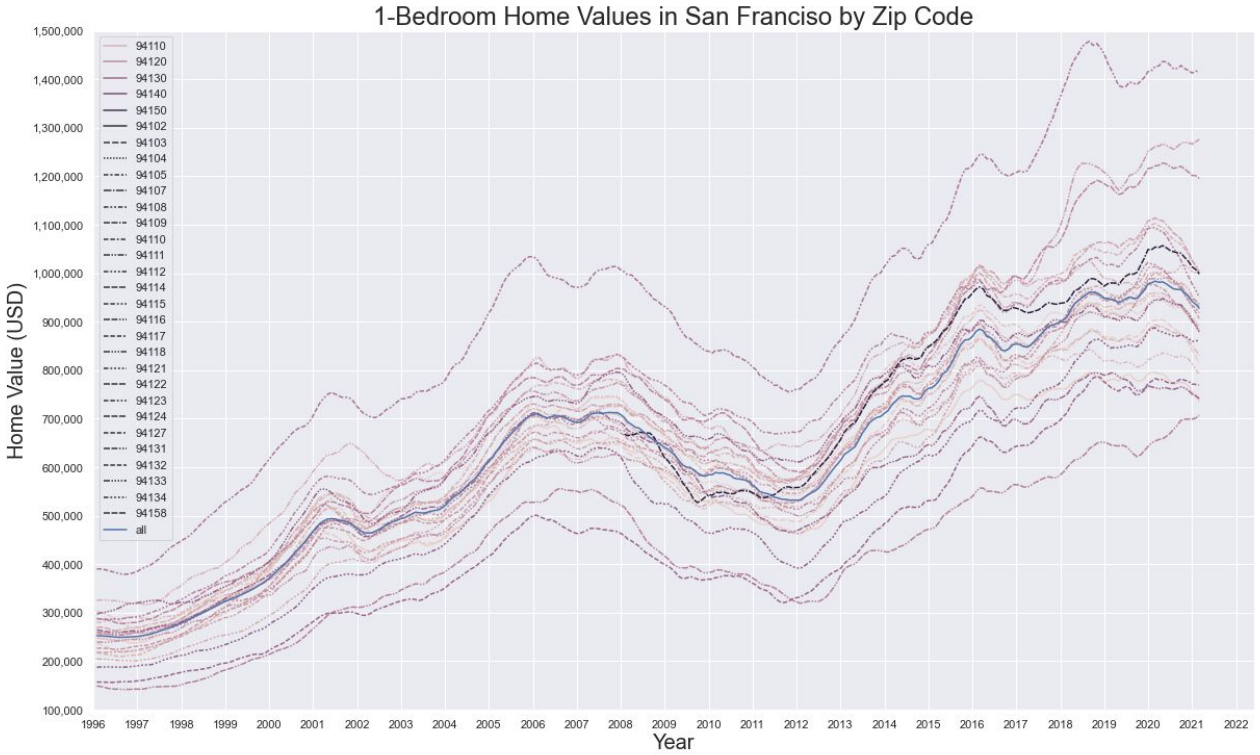
Overview

- Dataset is Home Value Index Data from the real estate company Zillow.
- Home value for each month is a separate column, i.e. data is in wide format.
- However, for doing time-series analysis, all the data must be in a single column, i.e. long format
- Therefore the dataset must be transformed via “melting”.
- This was achieved with a customized function using the *pandas* `melt` method.

Data Prep

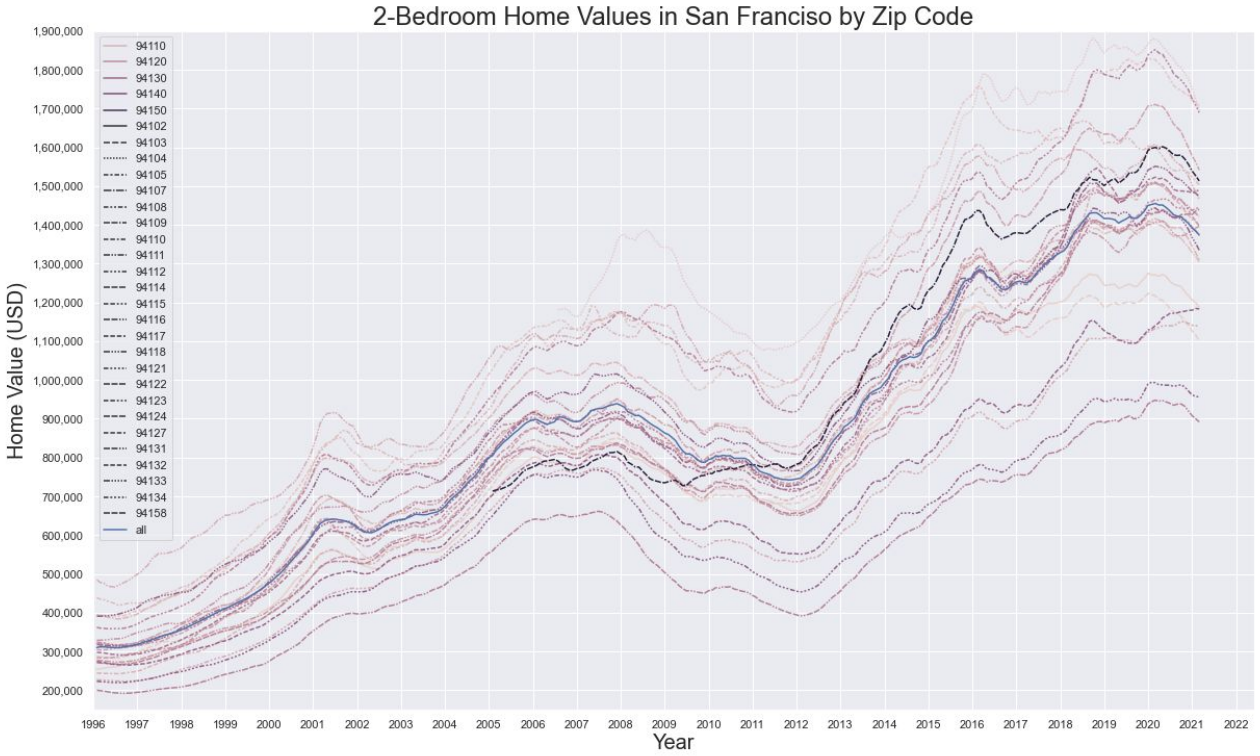
- Each zip code ('RegionName') with monthly data on home value is a separate time series.
- Stationarity was evaluated: data required differencing twice.
- A dictionary structure was implemented to store time series data across all zip codes.
- A 12 month seasonality was found in the data.

Data Visualization: 1-Bedroom Homes



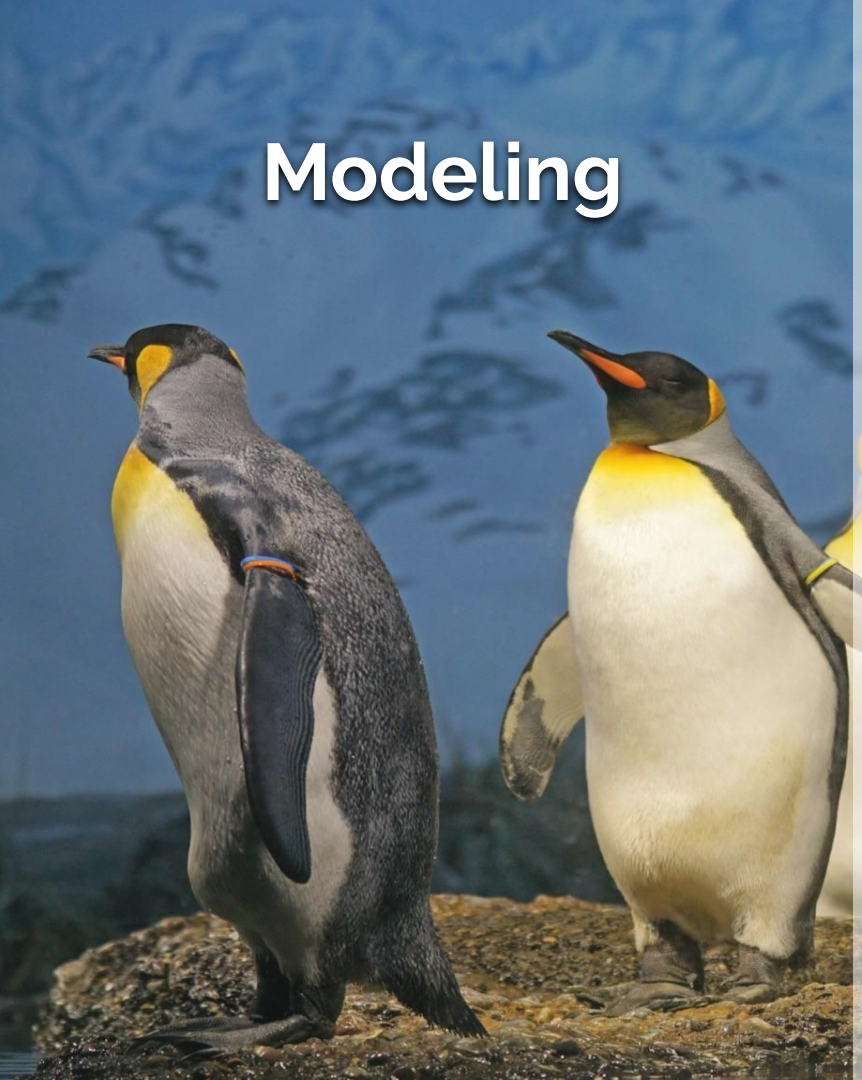
- We see a general trend upward with some cyclical movements.
- Not all Zip Codes moved in the same direction during certain market shocks.

Data Visualization: 2-Bedroom Homes



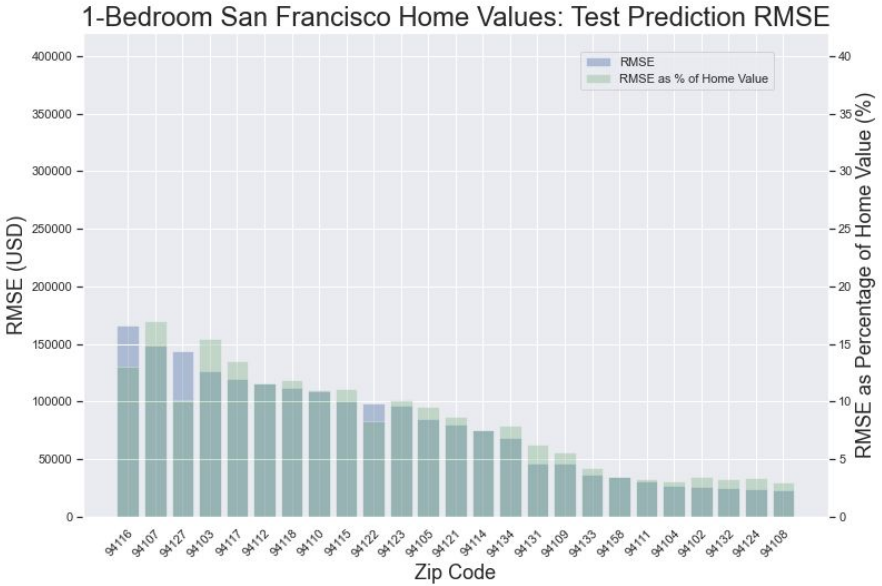
- Same story with 2-bedroom homes.
- After market crashes such as in 2000, 2008, and 2016, it took at least year, sometimes several, for prices to recover.

Modeling



- Each zip code is its own beast and requires a separate model.
- It's a reasonable expectation that models may have different parameters — this was borne out by our investigation.
- A triple `for` loop and AIC score were used to score models to determine the best parameters for each zip code.
- Using an 80/20 train-test split, we implemented a stepwise test prediction algorithm and generated RMSE scores to evaluate model fit.

- We used optimal SARIMAX parameters of ARIMA (2, 1, 0) and SARIMA (2, 1, 0, 12).
- AICs of models were in the high 3000 range.
- RMSEs represent between 2% to 16% of home values.



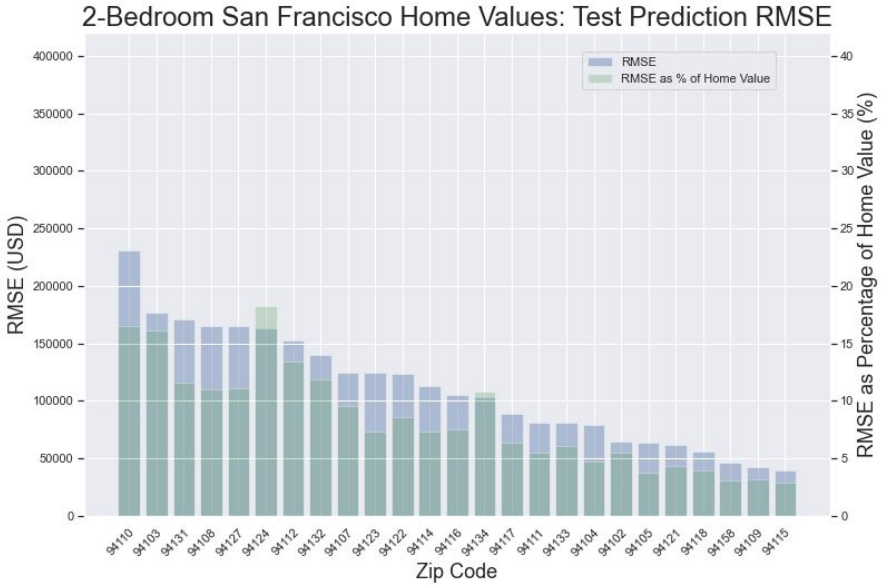


Photo by WENCHENG JIANG from Pexels

Forecasting

Forecast 12 months into the future ⟶ Feb 2022.

Calculate the percentage change in home value from last observed to last prediction for each zip code.

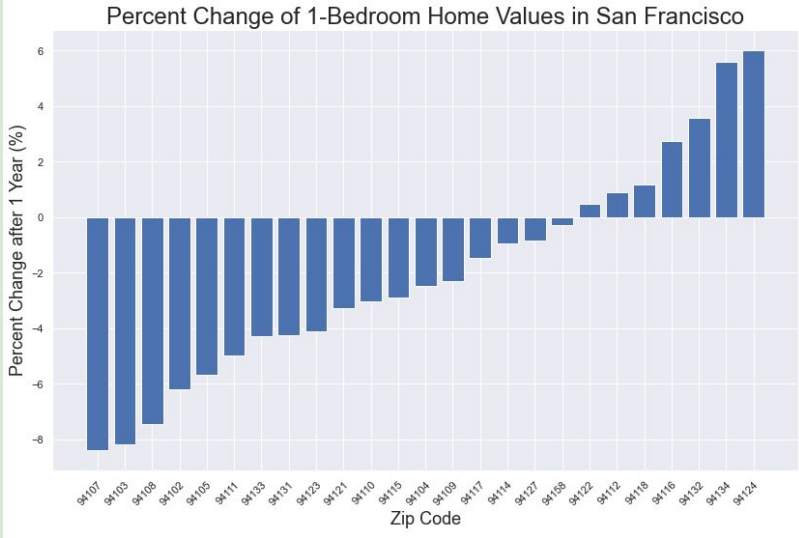
Most zip codes for 1-bedrooms showed negative growth, whereas 2-bedrooms were close to evenly split.

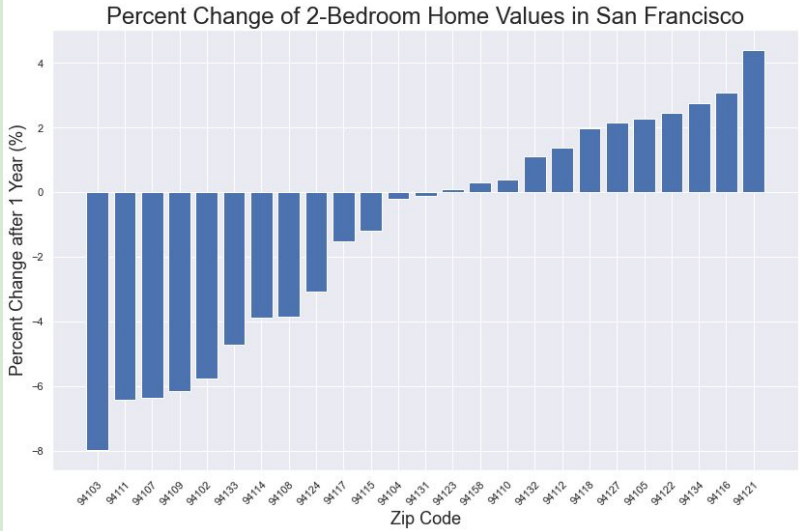
Maybe a market correction is about to occur?

. . .

Growth Charts

- Some zip codes that are projected to grow for 1-bedroom home show negative growth for 2-bedroom homes.
- The reverse is also true.
- Systemic shocks and black-swan events are impossible to predict (by definition), and may affect outcomes.
- Further investigation of exogenous variables such as crime rate, proximity to major transportation hubs, etc. can improve our predictions.

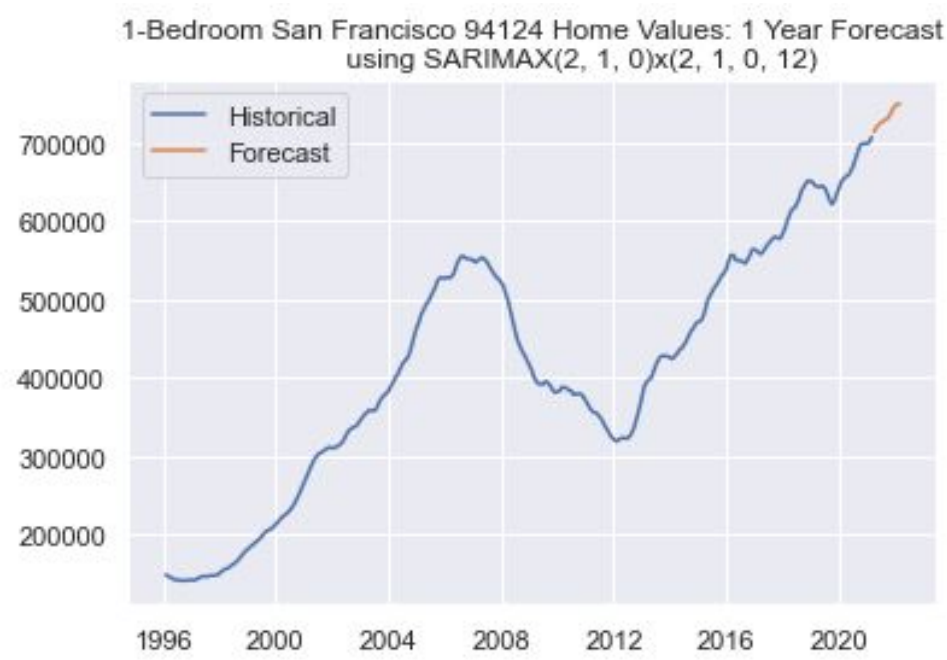


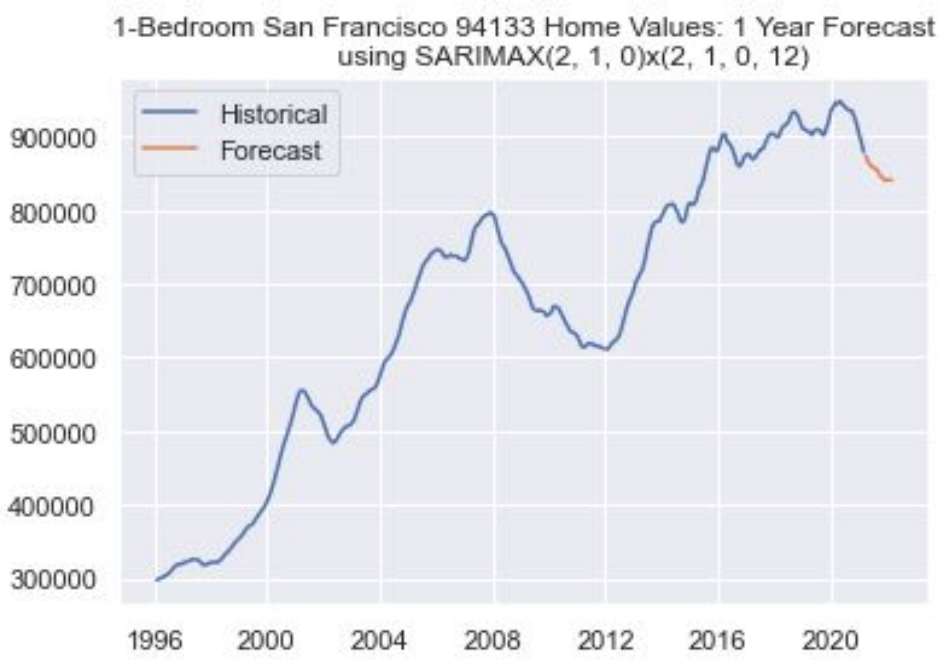


12-Month Forecasts

The blue series are historical data. The orange series are the prediction going out a year.

Can you guess simply from the data that one market will shrink while the other will grow?





Results

Best zip codes for investing in mid-tier homes, looking at a 1 year time-horizon...

One-bedroom

1. 94124 +6.0%
2. 94134 +5.6%
3. 94132 +3.6%

Two-bedroom

1. 94121 +4.4%
2. 94116 +3.1%
3. 94134 +2.7%

Results not guaranteed. This is not financial advice. We are not financial advisers. This analysis is provided for entertainment purposes only.

Conclusion

Without time-series modeling, you could be just rolling the dice with financial decisions.

Authors

Jonathan Silverman | Wei Alexander Xin

Jonathan's GitHub | Alexander's GitHub

Jonathan's LinkedIn | Alexander's LinkedIn

Happy flipping!