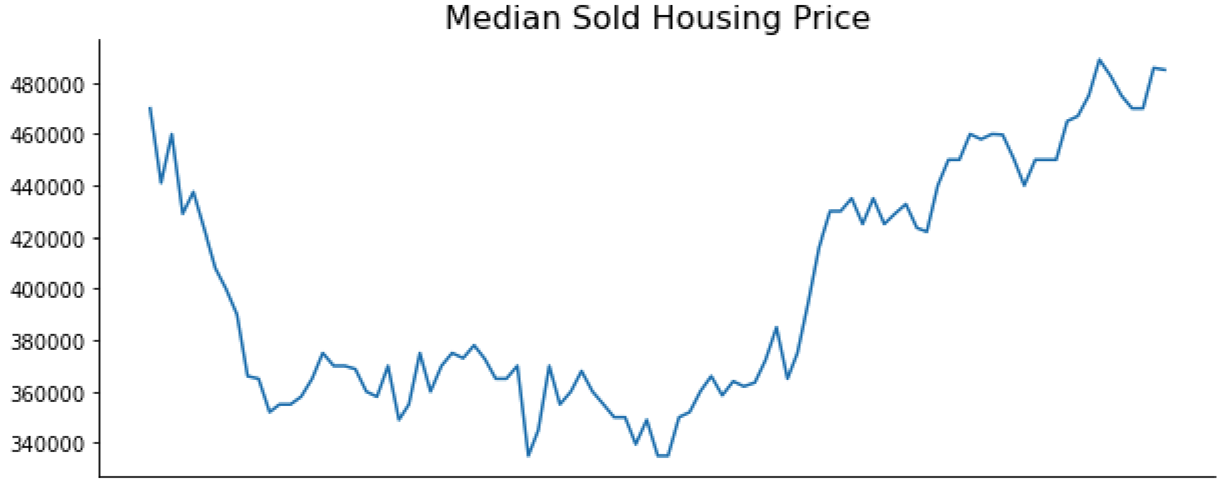
Time Series Analysis On the Housing Price in California

Group 3: Tian Qi, Phillip Navo, Danh Nguyen

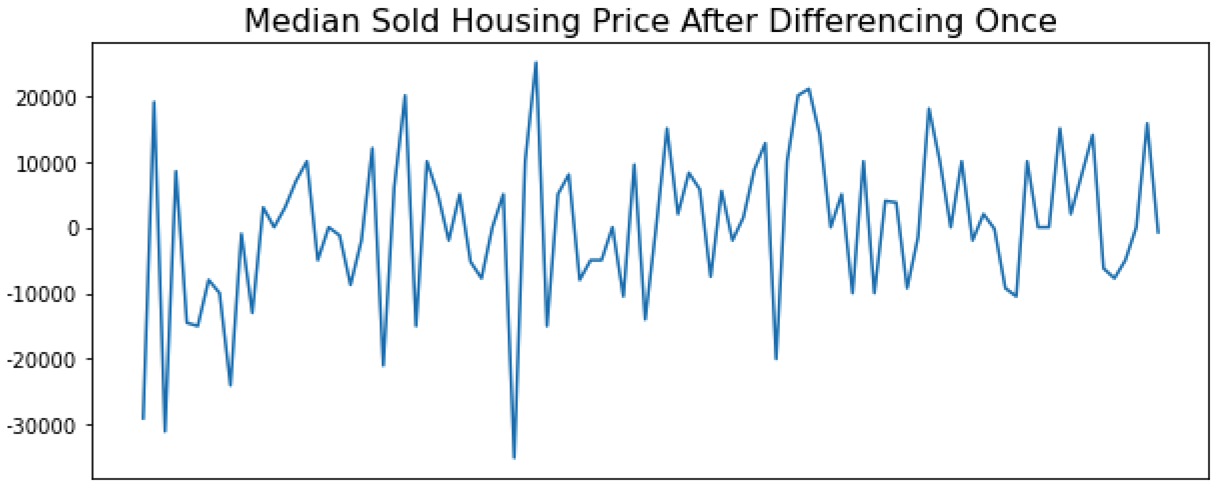
Introduction

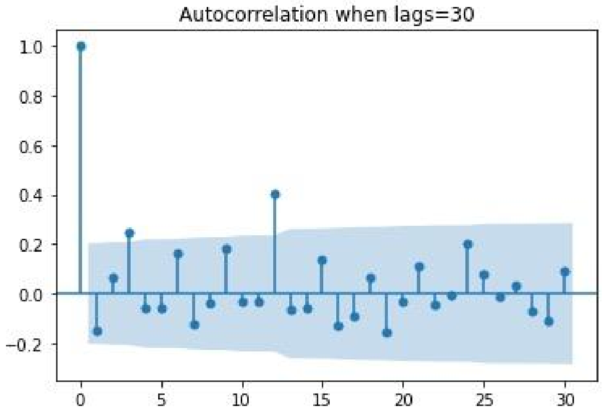
In this project, we aim to predict the monthly median sold price for housing in California in 2016 based on the price between 2008 and 2015. The dataset contains 107 records of median sold price, median mortgage rate and unemployment rate between Feb 2008 and Dec 2016. We split records before 2016 as the training set and records in 2016 as the test set. We first explored the dataset through visualizations to understand the trend and seasonality of price over time. Then we tried univariate and multivariate time series models including SARIMA, ETS, Prophet, VAR, and LSTM.

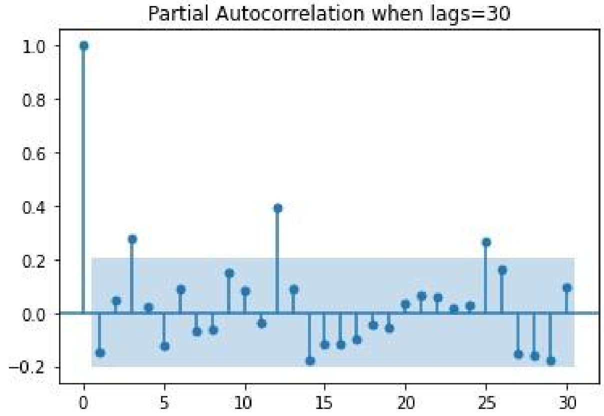
Exploratory Data Analysis

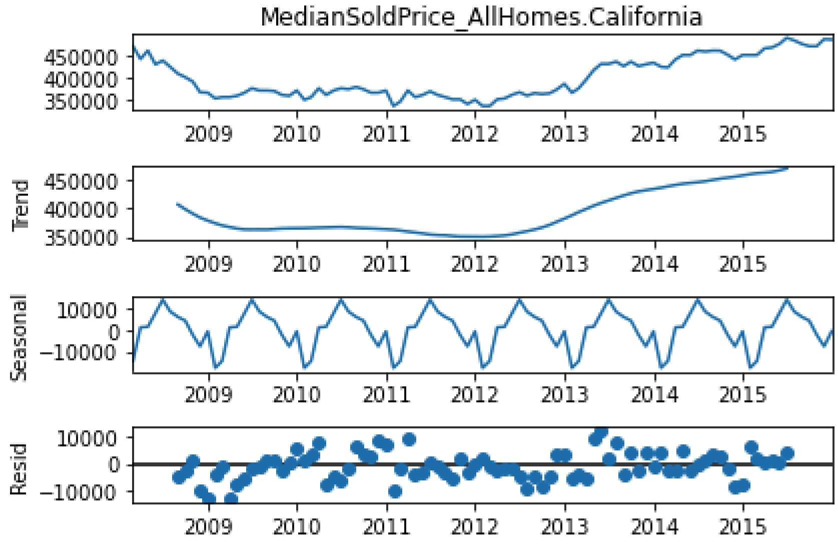


From the plot we can see the price went down and then went up. Therefore the price is not stationary, which is confirmed by the ADF test. After we differenced the data once, it becomes stationary. In addition, we see a yearly pattern so assume there is a yearly seasonality.









Based on the seasonal decomposition above, we confirmed that seasonality is additive so we do not need to do a logistic transformation for our data before applying the SARIMA model.

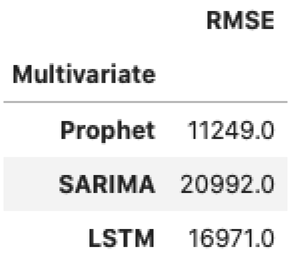
Model Selection

In general, we applied testing to both univariate and multivariate models and select the best model among them based on the following steps:

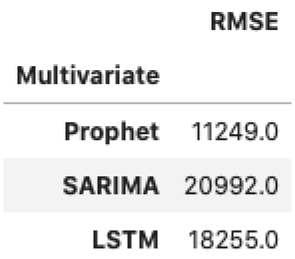
- We plotted the original data, ACF, and PACF to check for stationary conditions.
- Visual inspection of plots showed data is not stationary, confirmed with ADF test, p-value: 0.953391, this is much higher than alpha which indicates not stationary.
- Difference the data once and recheck, this is enough to make the data stationary.
- Confirmed again with ADF test, p-value: 0.027443 this is less than alpha indicating stationary.
- Wet set d = 1 and m=12 to auto search the other parameters among the (S)ARIMA models to select the best model based on BIC, the best univariate (S)ARIMA model we found has trend order (2,1,2) and season order (0,1,0,12).
- Next, we tried all the combinations of ETS models based on season, trend, damped, and m=12 (yearly seasonality) and found the best model within this family is the one with multiplicative trend and seasonality, and with damped as True.
- Then we used Prophet to produce another univariate candidate model based on frequency as “MS”, (we even tried setting seasonality to multiplicative but we got worse results) which stands for Month Start for monthly data point, and period as 12 for a year.
- In addition, we tried a variation of multivariate Prophet models with both unemployment rate and mortgage rate as exogenous variables. The best Prophet model in this case is the one with just the Mortgage Rate.
- We decided not to use the VAR model to apply the multivariate regression because we see that the unemployment rate might affect housing prices but not the other way around.
- We also tried univariate and multivariate LSTM with 10 hidden layers, 200 epochs and batch size 12.

Findings

We found that the univariate ETS model with a multiplicative trend and multiplicative seasonality had the lowest RMSE value when comparing the training and validation data sets. Therefore we chose it as our final model.

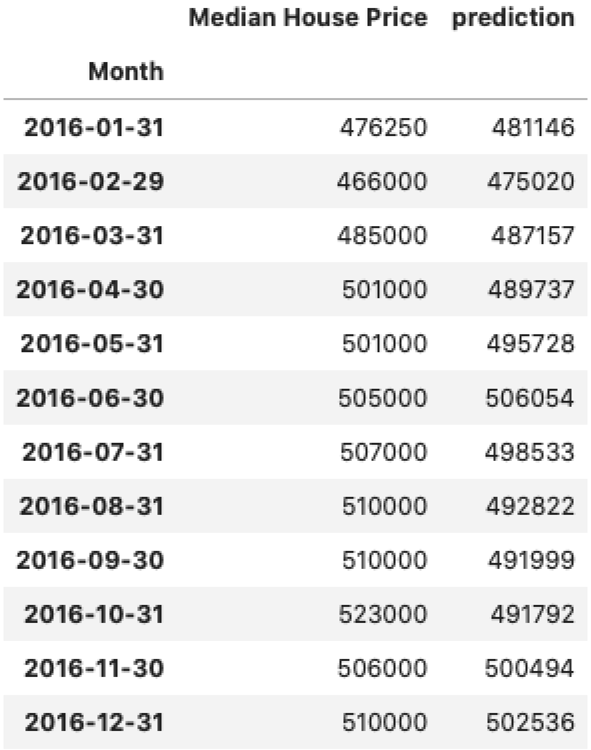


	RMSE
Multivariate	
Prophet	11249.0
SARIMA	20992.0
LSTM	16971.0



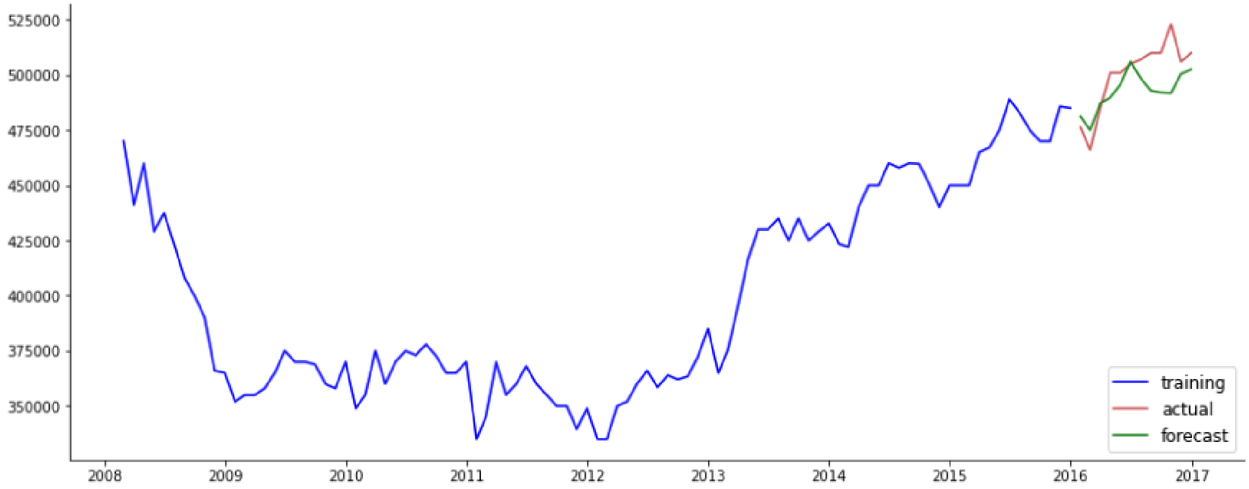
	RMSE
Multivariate	
Prophet	11249.0
SARIMA	20992.0
LSTM	18255.0

The RMSE and MAPE on the test set is 12963 and 0.02 respectively, which is a fair performance. The plot shows that our model captures the general trend of the housing price.

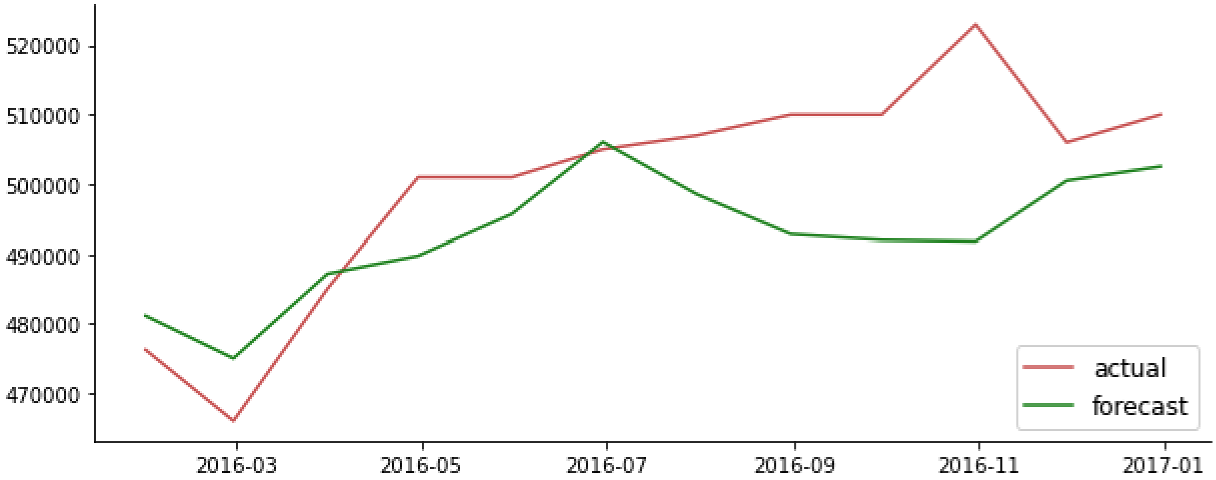


	Median House Price	prediction
Month		
2016-01-31	476250	481146
2016-02-29	466000	475020
2016-03-31	485000	487157
2016-04-30	501000	489737
2016-05-31	501000	495728
2016-06-30	505000	506054
2016-07-31	507000	498533
2016-08-31	510000	492822
2016-09-30	510000	491999
2016-10-31	523000	491792
2016-11-30	506000	500494
2016-12-31	510000	502536

Plot of data over time:



Forecast VS Actual for test data:


actual
forecast