Forecasting Report: House Property Sales Time Series Analysis

Introduction and Objective

Real estate is one of the largest asset classes and it plays a pivotal role in economic stability and growth. Therefore, forecasting property sales trends is critical for investors, developers, and homeowners to make informed decisions. The ability to predict future property prices can help stakeholders identify opportunities for investment and avoid market downturns. Our group chose this topic because property sales data can reveal trends, such as rising or falling prices, that may provide valuable insights into market trends.

Data Source and Preprocessing

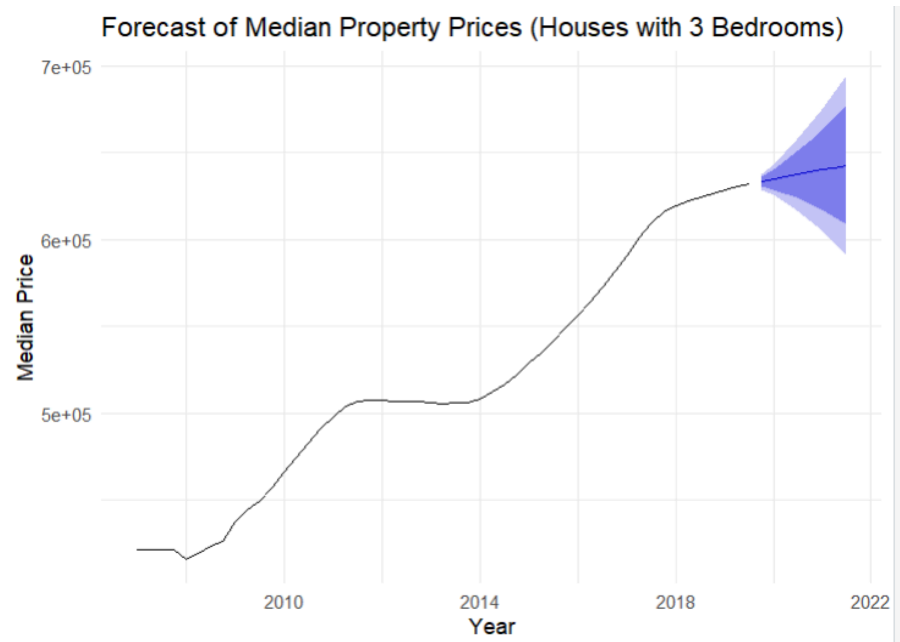
The dataset used in this analysis was sourced from Kaggle, covering quarterly property sales data between 2007 and 2019 within a specific region. The key variables include sales price, date of sale, property type (house or unit), and number of bedrooms.

Limitations

Since the dataset used in this analysis covers property sales from 2007 to 2019, we must acknowledge that it is a period preceding the onset of the COVID-19 pandemic. It is important to note that the housing market experienced shifts during the pandemic with factors including historically low interest rates, and supply chain disruptions that impacted construction costs. While our forecasts extends up to 2022, it is based purely on pre-pandemic trends, and may not fully capture the effects of the pandemic. Although our model does not incorporate pandemic related data, it can still allow stakeholders to consider both pre- and post-pandemic scenarios when making decisions. Nonetheless, the trends identified during the 2007-2019 period provide a valuable baseline for understanding the housing market under normal economic conditions.
In addition, in the analysis, we had ignored the difference on property types (house vs unit). With the quantum of the data acquired, we considered the impact from other variables is minimum.

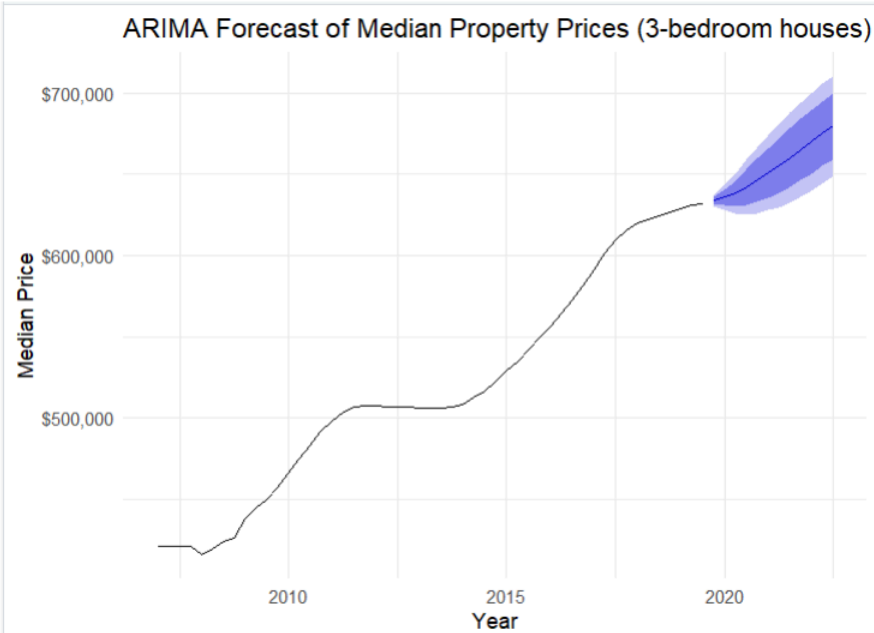
Forecasting House Prices: ETS Model (Exponential Smoothing)

In this analysis, we initially chose to focus on 3-bedroom houses due to their widespread appeal in the housing market, evidenced by 3-bedroom containing the largest dataset in the sample database. We felt that 3-bedroom homes are among the most common property types attracting a broad demographic of buyers, including families and investors. To begin forecasting, we applied the ETS model to forecast median house prices. The ETS model works by smoothing out short-term fluctuations to identify long-term trends to convey an accurate prediction of future values. The forecast shows a continuation of the upward price trend observed in previous years.



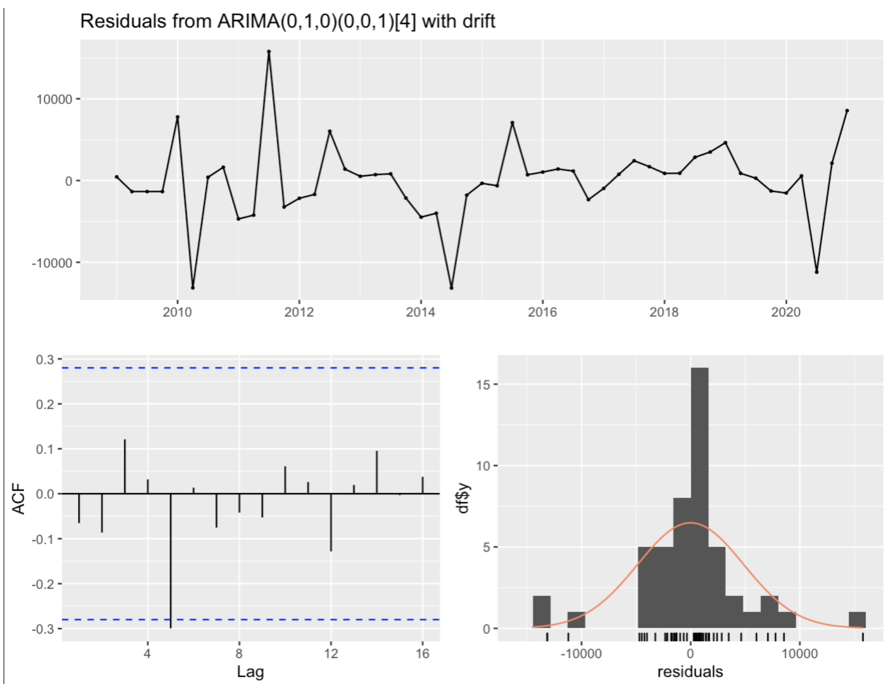
ARIMA Model Forecasting

The model's forecast projects median house prices for 3-bedroom houses for the next three years. The forecast captures a period of steady market growth, excluding any impact from the pandemic or the economic disruptions that had followed. One of the key characteristics of the model is the widening confidence interval as the forecast period extends, reflecting uncertainty about future prices due external influences. The model suggests that prices may continue rising, but it's also essential for stakeholders to consider unexpected events.



The graph below presents the forecasted median property prices through 2024, with a confidence interval shaded in blue to indicate the range of possible outcomes. The confidence interval widens over time, with increased uncertainty in the predictions as the forecast period extends. The effect here reflects potential fluctuations in the housing market that could arise from unforeseen factors.

To ensure that our model is accurate, and that it will capture the underlying patterns in the data, we conducted a residual diagnostic analysis. Top plot shows residuals over time and they should fluctuate randomly around zero without a clear pattern. This indicates that the model has efficiently captured the trend and seasonality of the data. The bottom left plot displays ACF of residuals. For a well-fitted ARIMA model, we expect the residuals to exhibit minimal autocorrelation. In this case, the ACF plot shows no significant spikes outside the confidence intervals and that the residuals are approximately white noise; making it suitable for forecasting. The bottom right shows the distribution of residuals and a normal distribution centered around zero is ideal. The histogram shows a roughly normal distribution, which aligns with the model's assumptions.

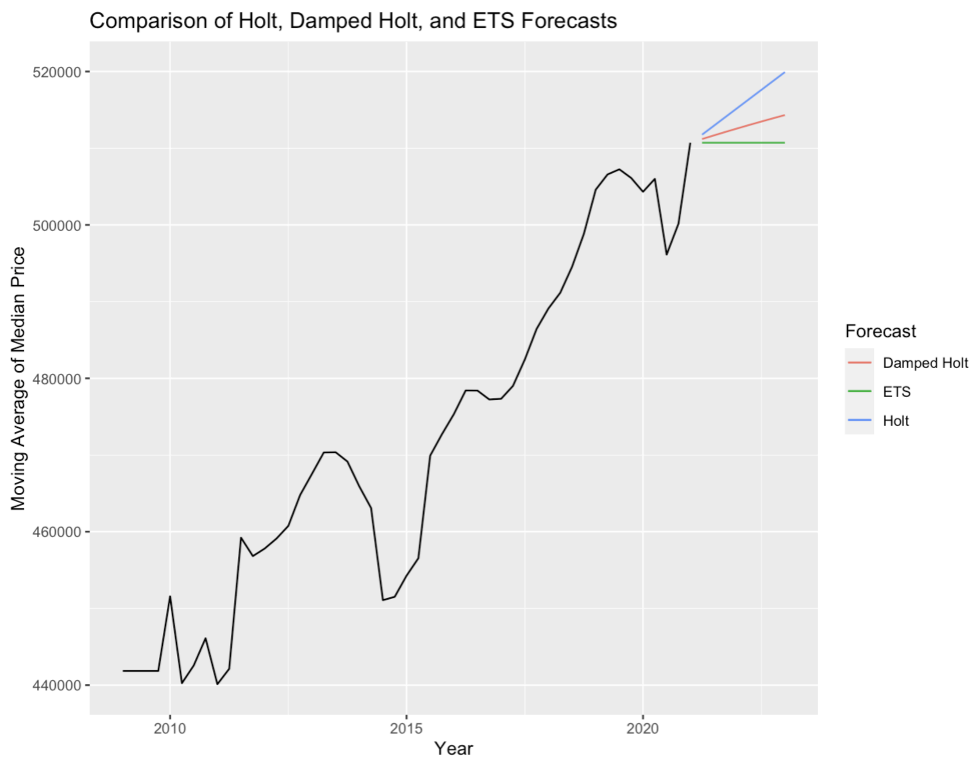


The ARIMA model (0,1,0)(0,0,1)[4] with drift has been rigorously evaluated, showing appropriate coefficient estimates, low model selection criteria, and uncorrelated residuals. This makes it a strong candidate for forecasting real estate prices, as it aligns well with statistical standards for accuracy and reliability.

```
> arima_model <- auto.arima(realestate.MA)
> print(arima_model)
Series: realestate.MA
ARIMA(0,1,0)(0,0,1)[4] with drift

Coefficients:
         sma1      drift
      -0.3693  1440.4390
s.e.   0.1853   464.6178

sigma^2 = 24513602:  log likelihood = -475.73
AIC=957.46   AICc=958.01   BIC=963.07
> arima_forecast <- forecast(arima_model, h=12)
> autoplot(arima_forecast) +
+   ggtitle("ARIMA Forecast of Real Estate Prices") +
+   xlab("Year") + ylab("Moving Average of Median Price")
> checkresiduals(arima_model)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,0)(0,0,1)[4] with drift
Q* = 7.0199, df = 7, p-value = 0.4268

Model df: 1.   Total lags used: 8
```

Comparison of Forecasting Models: Holt, Damped Holt, ETS

These models were selected due to their suitability for time series data with trends, each offering different forecasting characteristics. Holt's model represented by the blue assumes a linear trend without damping. In other words, it is projecting a more accelerated growth in median prices. This model is effective in showing consistent growth or decline without fluctuations. The Damped Holt model, represented by the red line, shows a damping parameter that slows down the trend over time. This approach is beneficial in cases where the trend is expected to persist but at a rather decreasing rate as opposed to over-forecasting in volatile markets. While Holt's model predicts a more aggressive increase in prices, Damped Holt model provides a tempered forecast. By analyzing these projections, stakeholders can assess a range of potential price trajectories and decide which forecast aligns with their market outlook or risk tolerance.



<u>Model Accuracy Comparison</u>

Holt’s model shows a moderate level of error metrics, meaning that it captures the trend in prices with reasonable accuracy. The slightly higher RMSE and MAPE suggest it may overestimate future values in some instances. The Damped Holt Model shares similar accuracy metrics when compared to Holt’s model, but demonstrates a slightly higher bias (indicated by ME). This model is appropriate for scenarios where we expect trends to decelerate over time.

```
[1] "Holt Model Accuracy:"
> print(holt_accuracy)
                   ME     RMSE      MAE        MPE      MAPE      MASE          ACF1
Training set 220.104 5047.814 3337.172 0.03695908 0.7125778 0.3448157 -0.0001888008
> print("Damped Holt Model Accuracy:")
[1] "Damped Holt Model Accuracy:"
> print(damped_holt_accuracy)
                   ME   RMSE      MAE       MPE      MAPE      MASE         ACF1
Training set 552.1828 5099.2 3437.161 0.1054991 0.7331138 0.3551471 -0.003325215
> print("ETS Model Accuracy:")
[1] "ETS Model Accuracy:"
> print(ets_accuracy)
                   ME     RMSE      MAE       MPE      MAPE      MASE        ACF1
Training set 1408.29 5227.099 3570.539 0.2899279 0.7593331 0.3689285 -0.01125242
```

```
Holt's method

Call:
holt(y = realestate.MA, h = 12)

  Smoothing parameters:
    alpha = 0.9884
    beta  = 1e-04

  Initial states:
    l = 442937.6127
    b = 1163.448

  sigma:  5267.386

     AIC     AICc      BIC
1036.317 1037.712 1045.776
```

```
Damped Holt's method

Call:
holt(y = realestate.MA, h = 12, damped = TRUE)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 1e-04
    phi   = 0.98

  Initial states:
    l = 442936.9633
    b = 1321.8389

  sigma:  5381.133

     AIC     AICc      BIC
1039.309 1041.309 1050.660
```

```
ETS(A,N,N)

Call:
ets(y = realestate.MA)

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 441711.9559

  sigma:  5337.155

     AIC     AICc      BIC
1035.737 1036.270 1041.413
```

Overall, the Holt model shows the best performance among the three models, with the lowest error metrics (RMSE and MAE), indicating it provides the most accurate forecasts. Its low Mean Error (ME) and very low autocorrelation in residuals (ACF1) suggest that it forecasts with minimal bias and captures the trend effectively without significant pattern in errors.

The Damped Holt model is slightly less accurate than the Holt model, with marginally higher errors (RMSE, MAE) and a slightly higher bias. However, it still performs reasonably well and may be useful if we expect trends to dampen over time.

The Holt and Damped Holt models focus on recent data but assume a stable and slow-changing trend. The Damped Holt model gradually reduces the trend's effect over time, which helps with long-term forecasting.

Key Takeaways and Conclusion

Our forecasting analysis covered ETS, Holt's linear Trend, Damped Holt, and ARIMA models to project the median property prices based on historical data from 2007 to 2019. We applied these models aiming to provide a view of potential price trends, capturing patterns while accommodating different scenarios, such as a gradual or accelerated growth. Although this analysis does not account for shifts that had happened during the pandemic, it still serves as a valuable baseline conveying typical market behavior in stable conditions. These kinds of insights help stakeholders prepare for expected, as well as unexpected market fluctuations, equipping them with information that they can use to make informed investment and development decisions.