FHLB- SEASONAL ADJUSTMENTS OF HOUSE PRICES

1 PROBLEM

2 PROPOSED METHODS

3 EXISTING RESOURCES

4 PROCEDURE

5 RESULTS

6 CONCLUSIO
N

7 THANK YOU

GROUP - REBOOTERS

PAVANMAHAVEER SINGARA F789X254
RAJENDRA THOKALA U639D822

OBJECTIVE

To explore how adjusting for seasonal effects can provide insights into the underlying trends or changes in the market.

Problem Statement

Importance

- In this project, we will investigate the relationship between house prices and seasonal patterns in the housing market.
- By examining the fluctuations in house prices across different years and regions, we aim to uncover insights into how these variations affect the seasonal adjustment process.

- This is important in the housing market because seasonal patterns can have significant impacts on supply and demand
- By removing these seasonal patterns, analysts can better identify underlying trends and make more accurate predictions about future market behavior

Method

ARIMA (AutoRegressive Integrated Moving Average) models are useful for the objective of understanding the impact of house prices on seasonal adjustments in the housing market.

WHY ITS SUITABLE

MATHEMATICAL MODEL

- Time Series Analysis: ARIMA models are designed for time series data, making them well-suited to capture temporal patterns and trends present in house price fluctuations.
- Autoregressive and Moving Average Components: ARIMA utilizes past house prices (autoregressive) and forecast errors (moving average) to make predictions, allowing it to capture dependencies and short-term fluctuations in house prices.
- Stationarity Handling: ARIMA can transform non-stationary house price data into a stationary form through differencing, making it suitable for modeling and forecasting

The ARIMA(p, d, q) model can be defined as follows:

$$\Delta^d y_t = c + \sum_{i=1}^{p} \phi_i \Delta^d y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

where:

- $\Delta^d y_t$ represents the differenced time series of order d (i.e., d times differenced).
- c is the constant term (intercept) in the model.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients of lag orders 1 to p.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients of lag orders 1 to q.
- ε_t represents the white noise (residuals) at time t.

In this representation, we use the lag notation $\Delta^d y_{t-i}$ to denote the d-th order difference between the time series y_t at time t and its value y_{t-i} at a previous time $t - i$.

The ARIMA(1, 1, 0) model can be mathematically represented as follows:

$$\Delta y_t = c + \phi_1 \Delta y_{t-1} + \varepsilon_t$$

where:

- Δy_t represents the first-order difference of the time series y_t at time t.
- c is the constant term (intercept) in the model.
- ϕ_1 is the autoregressive coefficient of lag order 1.
- ε_t represents the white noise (residuals) at time t.

In this ARIMA(1, 1, 0) model, we have $p = 1$ for the autoregressive order (AR order), $d = 1$ for differencing (integration order), and $q = 0$ for the moving average order (MA order). The autoregressive term ϕ_1 captures the relationship between the current value and its immediately preceding value after differencing the time series once.

Method

ARIMA (AutoRegressive Integrated Moving Average) models are useful for the objective of understanding the impact of house prices on seasonal adjustments in the housing market.

WHY ITS SUITABLE

MATHEMATICAL MODEL

DATASET

Federal Home Loan Level Bank System (2009-2021)
This dataset contains information about home loans issued by the Federal Home Loan Banks over a 12-year period from 2009 to 2021.
The dataset is a public dataset.
The dataset is provided in CSV format, which can be easily loaded into popular data analysis tools and programming languages.

EXISTING DATA

FEATURES

COLUMN	DESCRIPTION
LTV	Loan-to-value ratio
UPB	Unpaid Principal Balance
Income	Income of the borrower(s)
BoCreditScore	Credit score of the borrower
CoBoCreditScor	Credit score of the co-borrower
Front	Ratio of mortgage principal and interest and housing expenses to total borrower income.
Back	Ratio of all debt payments to total borrower income.
NoteAmount	Mortgage balance at origination
PropType	Property Type
Term	Mortgage term length
Year	Year Loan Was Reported
Bank	Name of Federal Home Loan Bank District

```
RangeIndex: 743816 entries, 0 to 743815
Data columns (total 12 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   LTV             743816 non-null  float64
 1   UPB             743816 non-null  int64
 2   Income          743816 non-null  int64
 3   BoCreditScore   743816 non-null  int64
 4   CoBoCreditScor  743816 non-null  int64
 5   Front           743816 non-null  float64
 6   Back            743816 non-null  float64
 7   NoteAmount      743816 non-null  int64
 8   PropType        743816 non-null  object
 9   Term            743816 non-null  int64
 10  Year            743816 non-null  int64
 11  Bank            743816 non-null  object
dtypes: float64(3), int64(7), object(2)
memory usage: 68.1+ MB
None
```

- The HousePrice variable represents the calculated house price based on the given formula:

  HOUSE PRICE = NOTE AMOUNT / LTV

- To calculate the HousePrice, the NoteAmount is divided by the LTV. This calculation aims to estimate the house price based on the loan amount and the loan-to-value ratio. It assumes that the loan amount is proportional to the property value represented by the LTV.
- It's important to note that the HousePrice calculated using this formula is an estimation and may not reflect the actual market value of the house. It is a derived variable that can be used for analysis or modeling purposes.

Step-1 - Data Exploration

Step-2- Time Series Modelling

Data Exploration

Inspect the data structure and check for missing values.
Summarize statistics to get an overview of the data.
Visualize the time series data to understand its overall pattern and detect outliers or anomalies.

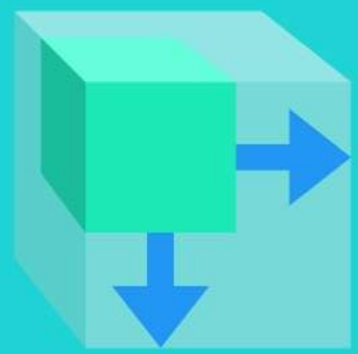
- Use seasonal decomposition to break down the time series into trend, seasonality, and residuals.
- Identify the underlying long-term trend, seasonal patterns, and unexplained fluctuations.
- Create visual plots to showcase the original data, trend, seasonality, and residuals separately.
- Understand how each component contributes to the overall behavior of the time series.
- Utilize Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to analyze autocorrelation and partial autocorrelation patterns.
- Determine suitable lag orders for the ARIMA model based on significant correlations.
- Preprocess specific columns to numeric data types, handling any non-numeric values by converting them to NaN.
- Apply the ARIMA (AutoRegressive Integrated Moving Average) model to the 'HousePrice' data.
- Set the ARIMA order (p, d, q) based on insights from ACF and PACF plots.
- Determine the number of autoregressive terms (p), differencing terms (d), and moving average terms (q).
- Fit the ARIMA model to the 'HousePrice' data using the specified order (p, d, q).
- Assess the performance of the ARIMA model in capturing the time series behavior.
- Print a summary of the fitted ARIMA model, providing essential statistical information about the model's performance and coefficients.
- Analyze the model's performance and consider further improvements if needed.

Step-1 - Data Exploration

Step-2- Time Series Modelling

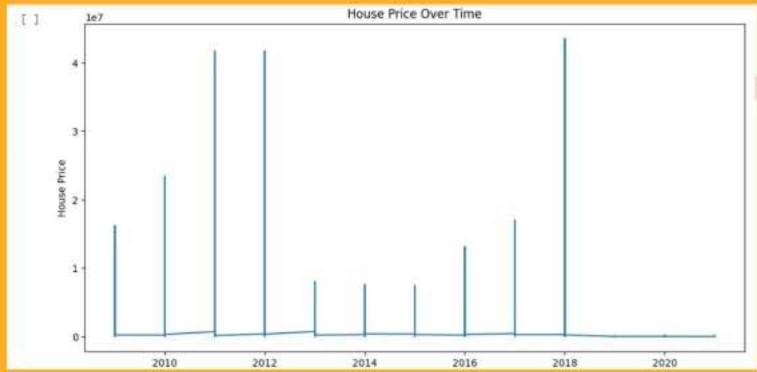
Model Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:             HousePrice   No. Observations:               743816
Model:                 ARIMA(1, 1, 0)   Log Likelihood           -10438874.739
Date:                Sat, 24 Jun 2023   AIC                       20877753.478
Time:                        23:59:21   BIC                       20877776.517
Sample:                             0   HQIC                      20877759.895
                             - 743816
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4947   2.59e-05  -1.91e+04      0.000      -0.495      -0.495
sigma2      9.068e+10   4.31e-18    2.1e+28      0.000    9.07e+10    9.07e+10
===================================================================================
Ljung-Box (L1) (Q):               15492.74   Jarque-Bera (JB):         731409110798.95
Prob(Q):                              0.00   Prob(JB):                            0.00
Heteroskedasticity (H):               0.01   Skew:                               26.68
Prob(H) (two-sided):                  0.00   Kurtosis:                         4860.66
===================================================================================
```

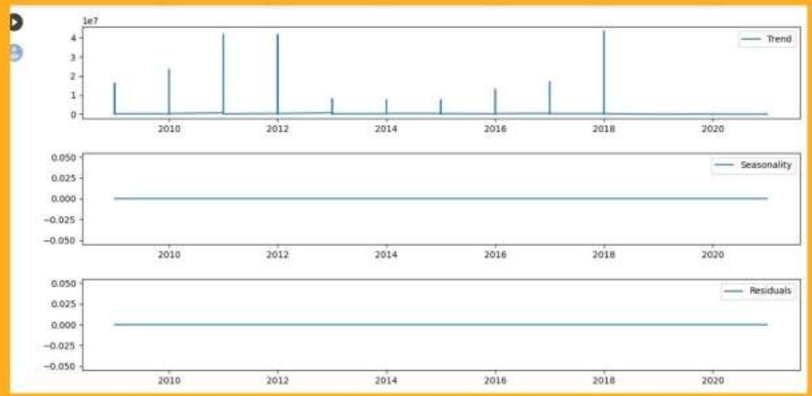


Plots

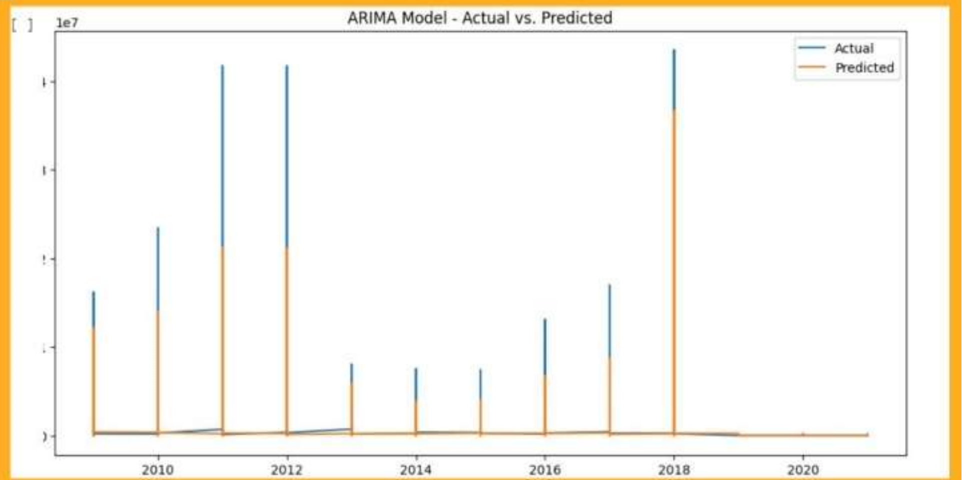
Summary



ORIGINAL DATA



SEASONALITY AND TRENDS



ARIMA MODEL

- AR.L1 coefficient has negative values and has a p-value of 0.000. It indicates ,Statistically significant relationship between current house price and its lagged value.

- Potentially supports a decreasing trend in house prices.

- Consider model limitations: autocorrelation, non-normality, and heteroskedasticity in residuals.

- The ARIMA model suggests a potential decreasing trend in house prices based on the negative coefficient for the autoregressive term (AR.L1).

- The ARIMA model forecasts a downward trend in house prices, suggesting the model anticipates a decline based on historical patterns.

- Caution should be exercised when interpreting the results due to uncertainties in predicting future prices and influences beyond the model's scope. Validation with metrics is essential to assess accuracy.

- The model offers valuable insights but should be supplemented with domain knowledge and expert insights to understand the forecasted decrease. Exploring alternative models and updating with new data can enhance accuracy.


thank you
SOLD