HOME PRICE INDEX ANALYSIS

1. **OBJECTIVE**:

To find **key factors** which will **influence Home Prices** in **US** on a **national level** from time period **2003 till 2022**.

2. **KEY FACTORS SELECTED**:

1. **GDP Growth** - The **Gross Domestic Product growth** is a **strong** indicator of **economic growth** for a country. When **GDP** is **growing** it will **shoot** up **economy** with **rising income levels** and **job opportunities**.This will **boost public** to **buy house**s thereby **increasing deman**d for houses which leads to **higher home prices** due to **demand**.

2. **Unemployment Rate** - When there is **Unemployment Rate** then **people** will **not** be having any **reliable source of income** to **buy** houses, which will **reduce** the **demand** for **houses** thereby **reducing home prices**.

3. **Inflation Rate** - As the name suggests **Inflation** will lead to **higher home prices** due to **raw materials cost** for **construction** and other **miscellaneous costs**.

4. **Mortgage Interest Rates** - **Mortgage Interest Rates** are crucial factor to **provide loans**, when the **interest rates are less** then it will be very much **easier** for the **public to borrow more money** to **buy homes** which will **positively influence home prices**.

5. **Housing Starts** - **Supply of new homes** is Housing Starts, an **increase** in **Housing Starts will stabilize or lower home prices**, likewise **decrease** in housing starts will **increase prices due to high demand**.

6. **Existing Home Sales** - Existing House Sales is very **good** indicator to get to know the **Real Estate current demand** based on which **prices are framed**.

7. **Consumer Confidence Index** - As the name suggests Consumer Confidence Index **influences** in **capturing** the **confidence levels and sentiments of people** in **deciding** to buy houses.

8. **Housing Credit Availability Index** - With **more availability** of Housing Credit to people, they can **easily lend loans** to purchase properties.

9. **Interest Rate Policy** - **Lower Interest Rates** from Banks will **stimulate** more **purchases** eventually increasing home prices.

These data are collected from US government websites, placing it here for your reference:

I. **Gross Domestic Product (GDP) Growth**:

Source Website: https://fred.stlouisfed.org/series/GDP

II. **Unemployment Rate**:

Source Website : https://data.bls.gov/pdq/SurveyOutputServlet

III. **Inflation Rate**:

Source Website : https://data.oecd.org/price/inflation-cpi.htm

IV. **Mortgage Interest Rates**:

Source Website : https://fred.stlouisfed.org/series/MORTGAGE30US

V. **Housing Starts**:

Source Website : https://fred.stlouisfed.org/series/HOUSTNSA

VI. **Existing Home Sales**:

Source Website: https://fred.stlouisfed.org/series/EXHOSLUSM495S

VII. **Consumer Confidence**:

Source Website:
https://data.oecd.org/leadind/consumer-confidence-index-cci.htm

VIII. **Housing Credit Availability Index**:

Source Website:
https://datacatalog.urban.org/dataset/housing-credit-availability-index-hcai

IX. **Interest Rate Policy**:

Source Website: https://fred.stlouisfed.org/series/FEDFUNDS

3. DATA PREPARATION AND PROCESSING:

Linear Interpolation:

In the **data attributes extracted**, some where in **Monthly format** and **some in Quarterly format**.

As we will be performing **Monthly level analysis** on **Home Price index**, we needed to **perform Linear Interpolation** to **convert** from **Quarterly** to **Monthly**.

```
In [135]: data_monthly['VALUE'].interpolate(method='linear', inplace=True)
```

Noise Detection Test:

Noise Detection Tests are being **performed** on all the features to **check** for the **presence of White Noise**.

Ljung–Box test which is primarly **performed** for **detecting White Noise** is being taken and **results** being arrived at:

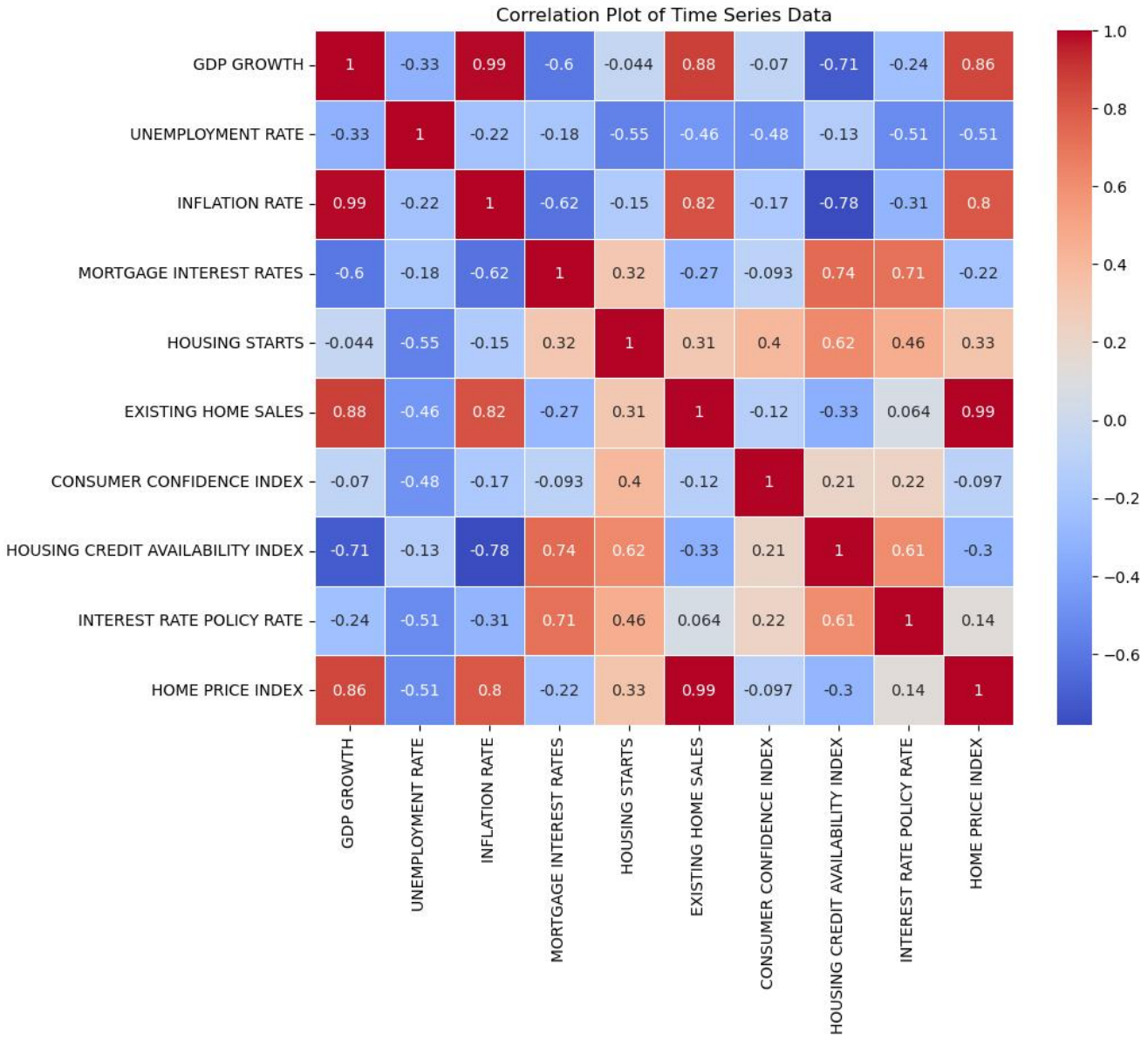
```
Column: GDP GROWTH
Result: Not White Noise

Column: UNEMPLOYMENT RATE
Result: Not White Noise

Column: INFLATION RATE
Result: Not White Noise

Column: MORTGAGE INTEREST RATES
Result: Not White Noise

Column: HOUSING STARTS
Result: Not White Noise

Column: EXISTING HOME SALES
Result: Not White Noise

Column: CONSUMER CONFIDENCE INDEX
Result: Not White Noise

Column: HOUSING CREDIT AVAILABILITY INDEX
Result: Not White Noise

Column: INTEREST RATE POLICY RATE
Result: Not White Noise

Column: HOME PRICE INDEX
Result: Not White Noise
```

As per the results we could see **none** of the **columns** contains **White Noise**.

4. DATA VISUALIZATION AND ANALYSIS:

Correlation Plot:

Correlation plot is created to analyze the **correlation** of **key factors** with **Home Price Index**.



Based on plot we could see **GDP GROWTH, INFLATION RATE, EXISTING HOME SALES** are **positively** impacting **HOME PRICE INDEX**, so basically

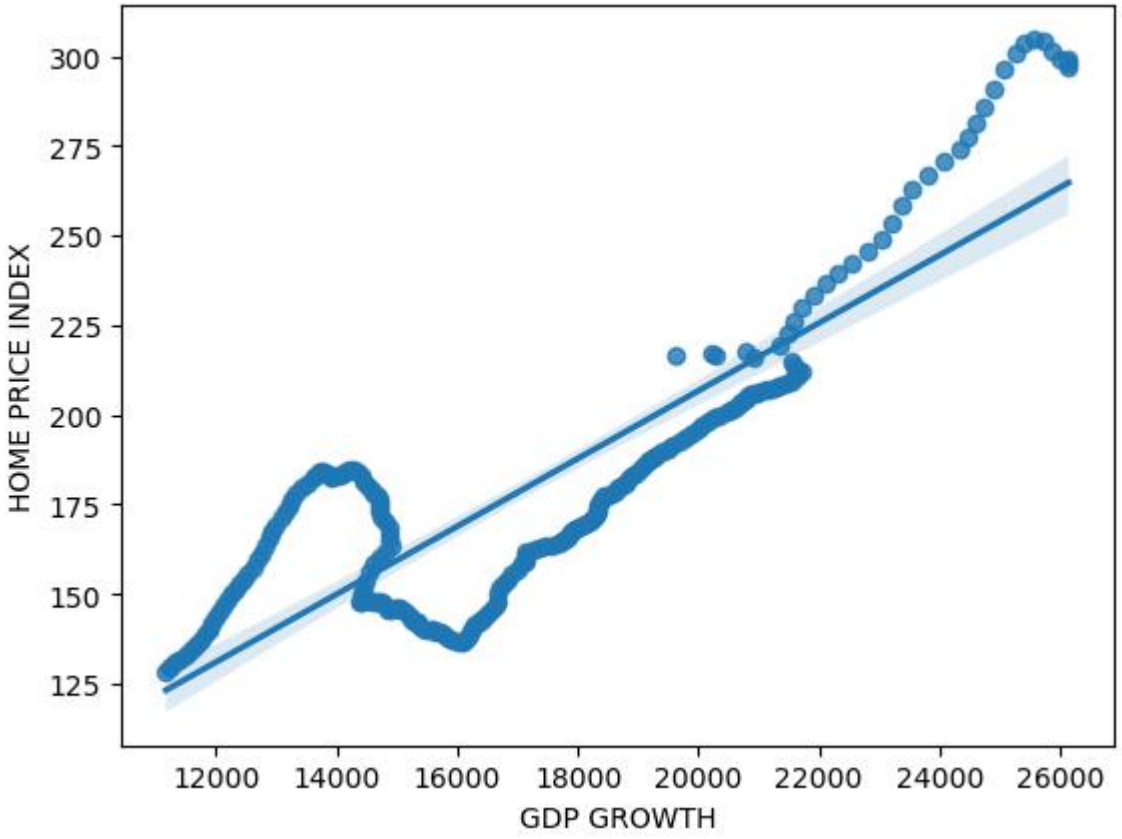
when these attributes **increase** then **HOME PRICE INDEX** also **increases**.

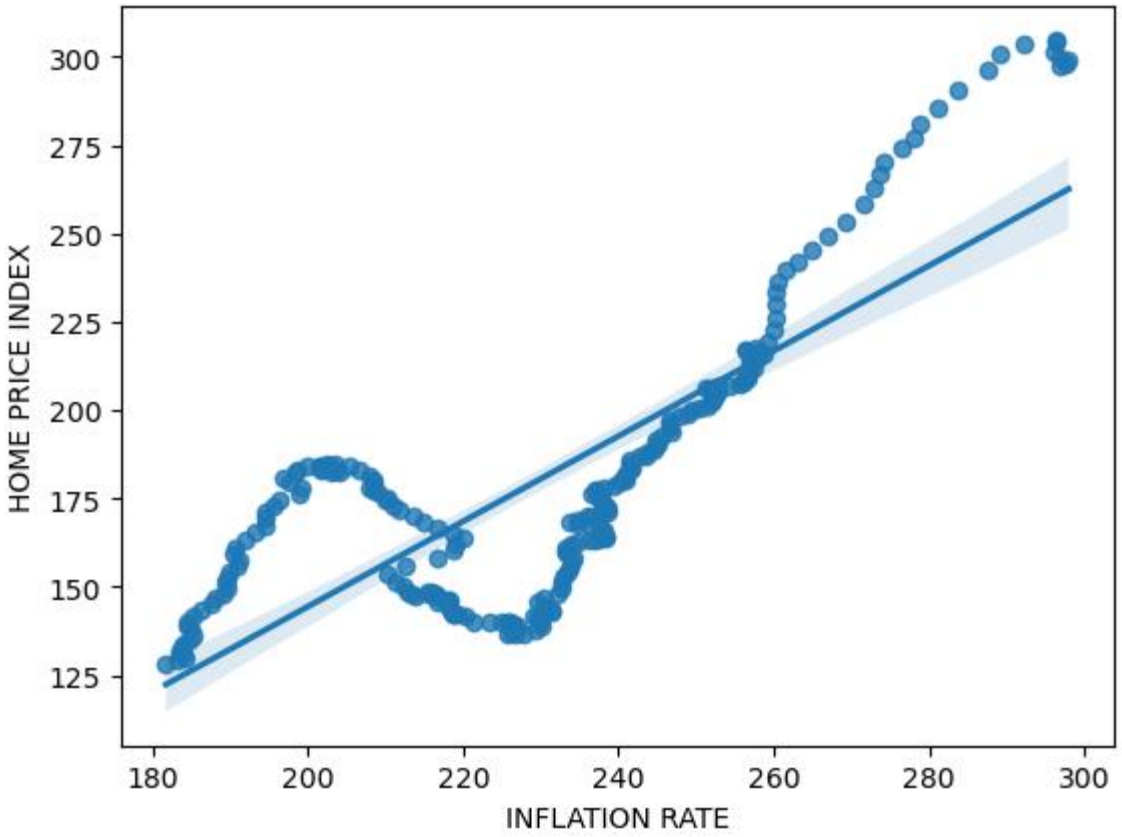
On the other hand **UNEMPLOYMENT RATE, MORTGAGE INTEREST RATES, CONSUMER CONFIDENCE INDEX AND HOUSING CREDIT AVAILABILITY INDEX** are **negatively** impacting **HOME PRICE INDEX**,so when these attributes **increases** then **HOME PRICE INDEX decreases**

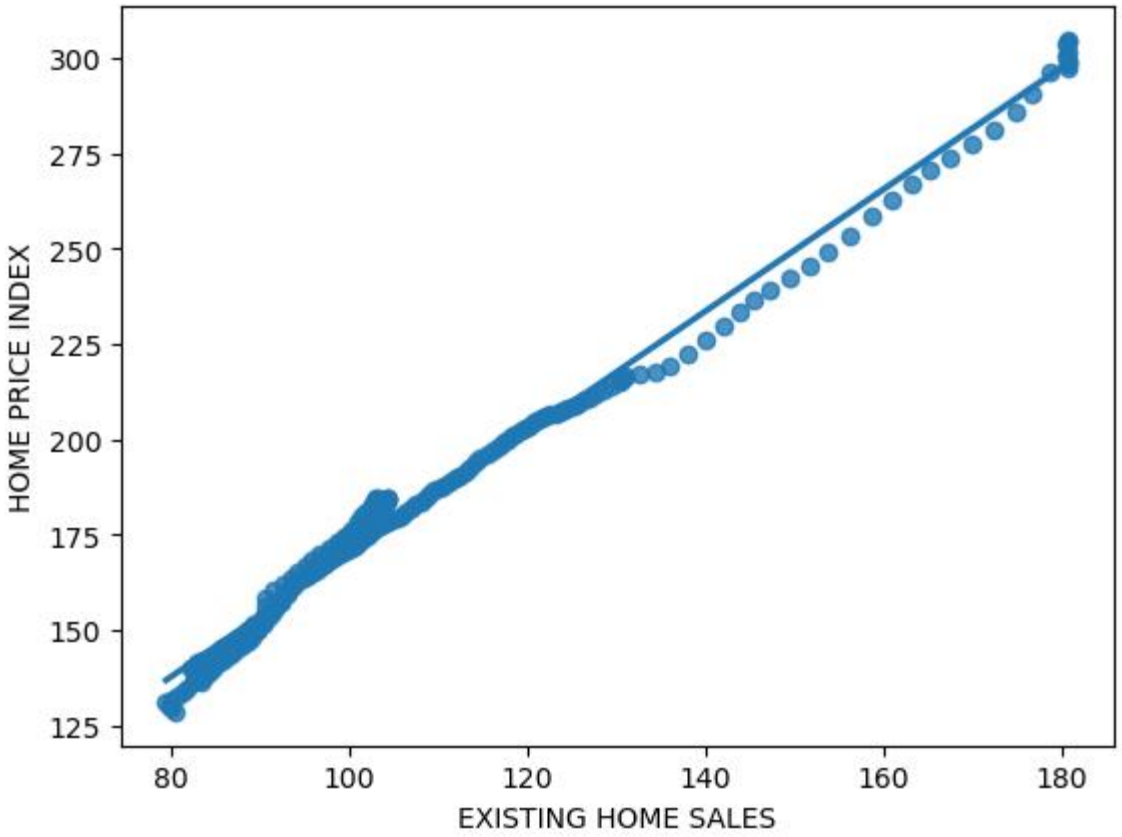
HOUSING STARTS and **INTEREST RATE POLICY RATE doesn't** influence **HOME PRICE INDEX**

Detailed plots:

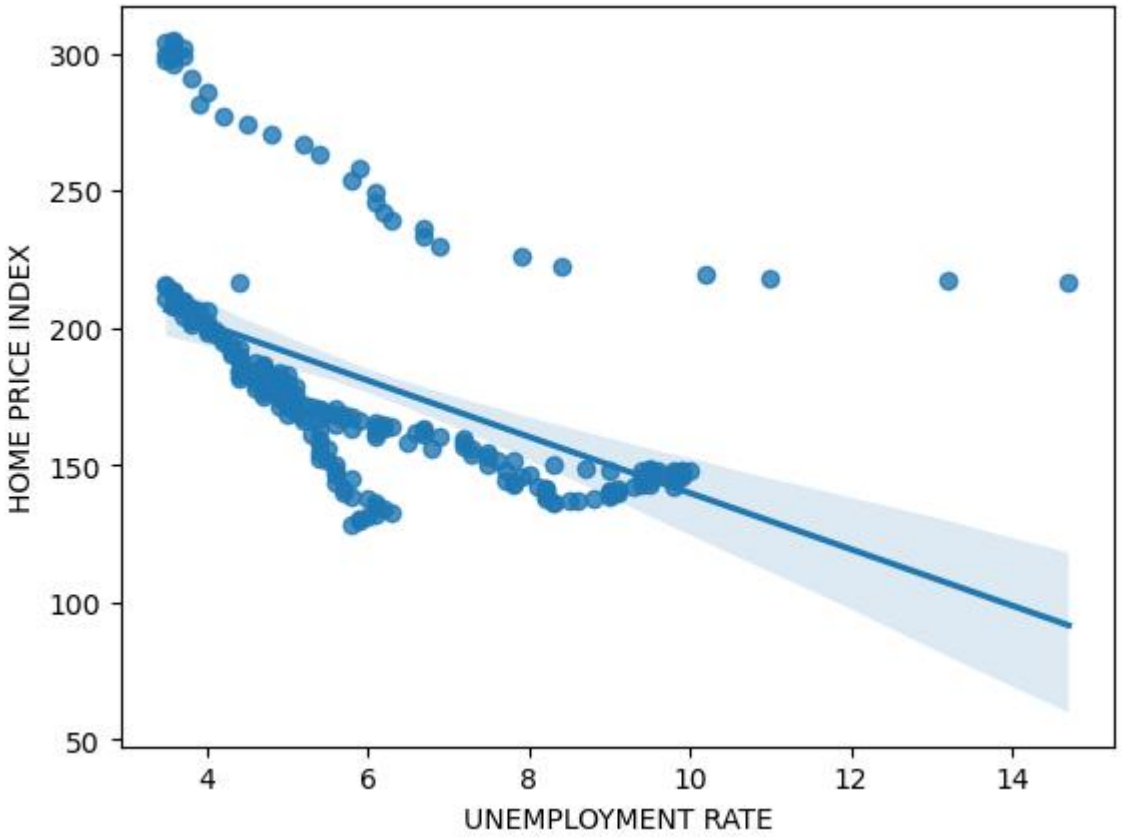
We could also see in detail manner the **Positive** and **Negative** influence.

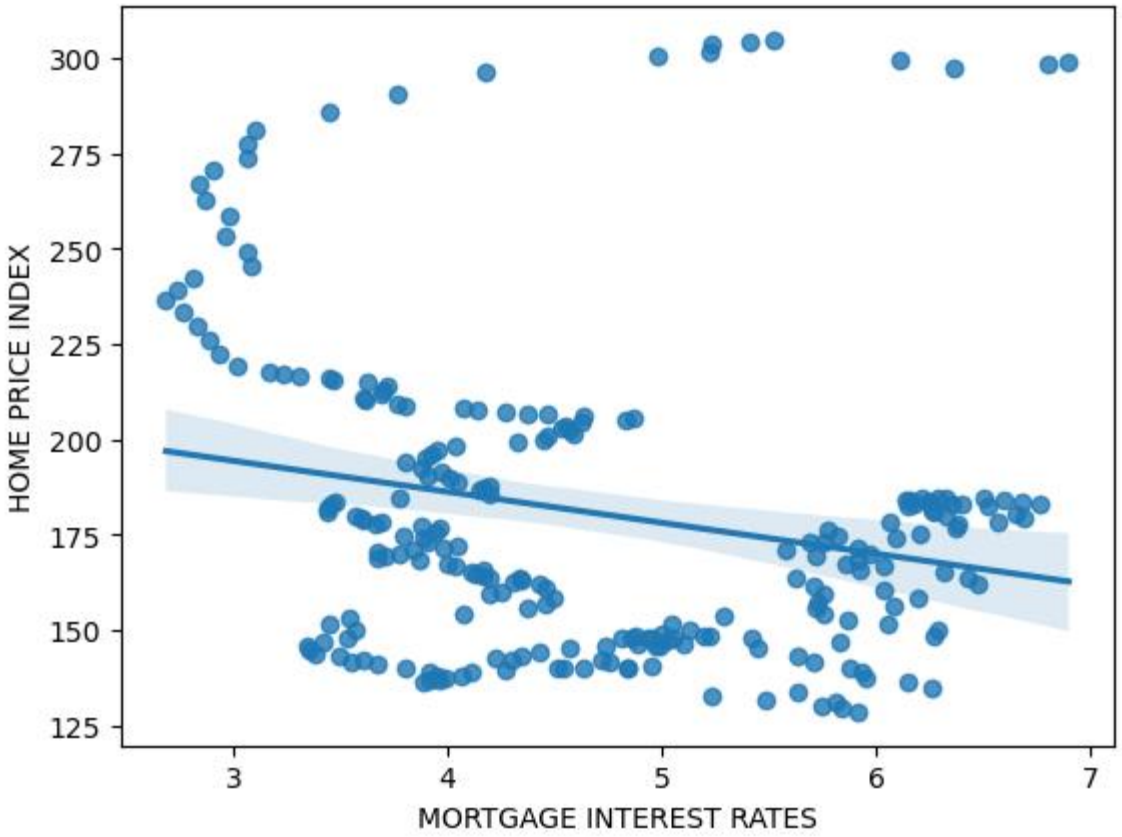


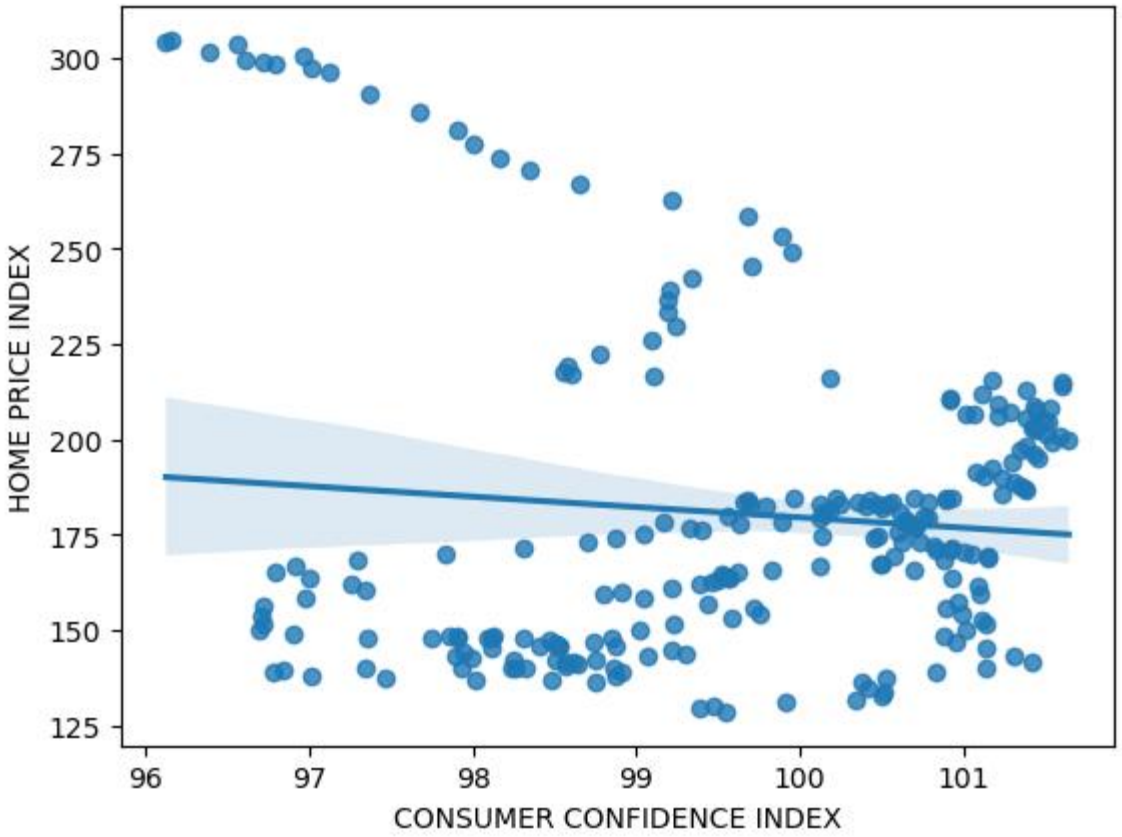

HOME PRICE INDEX
300
275
250
225
200
175
150
125
180
200
220
240
260
280
300
INFLATION RATE


HOME PRICE INDEX
300
275
250
225
200
175
150
125
80
100
120
140
160
180
EXISTING HOME SALES

As seen from Correlation plot these attributes are Positively influencing Home Price Index.




HOME PRICE INDEX
300
275
250
225
200
175
150
125
3
4
5
6
7
MORTGAGE INTEREST RATES



We could see these attributes are Negatively impacting Home Price Index as seen in Correlation Plot.

We could see that **Consumer Confidence Index** is **negatively** influencing **Home Price Index** which is **not the case** in **real time** scenario, so lets try to **analyze** this **deeper**.

Past Lags Analyse:

As far as **time series data** is being concerned not all attributes or features in the dataset might always **positively influence** with **present** data alone.

It can also **influence** with **past** data as well.

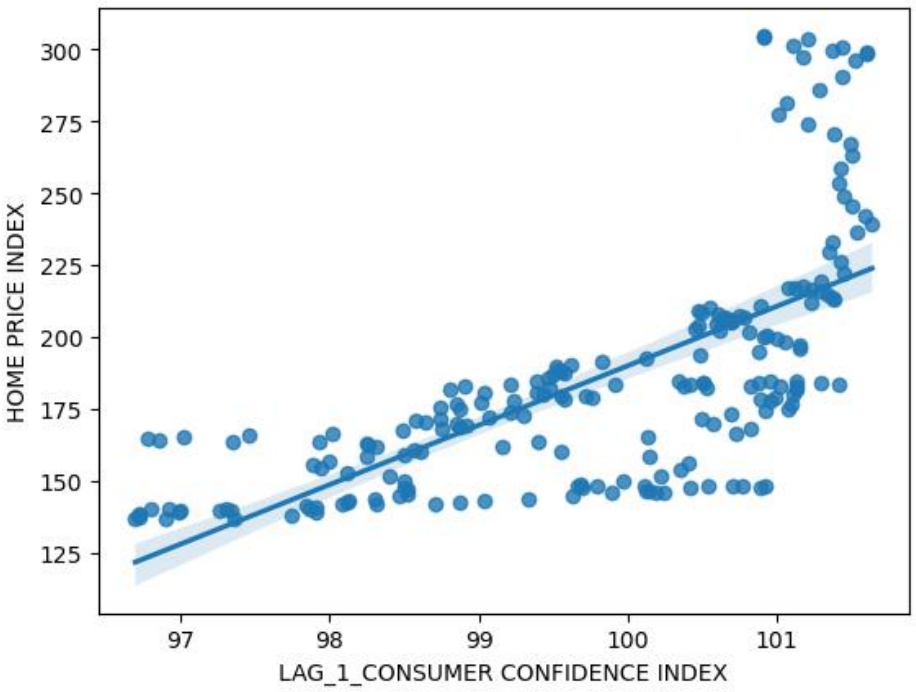
So lets try to **check** the **correlations for 40 lags** and see in **which lag** we are **observing** a **positive influence of consumer confidence index** upon home price index.

	Lag	Correlation
0	1.0	-0.067422
1	2.0	-0.035531
2	3.0	-0.002772
3	4.0	0.031346
4	5.0	0.068183
5	6.0	0.108740
6	7.0	0.150632
7	8.0	0.189979
8	9.0	0.226228
9	10.0	0.261245
10	11.0	0.293747
11	12.0	0.322625
12	13.0	0.348632
13	14.0	0.373426
14	15.0	0.396600
15	16.0	0.418010
16	17.0	0.436532
17	18.0	0.449555
18	19.0	0.458220
19	20.0	0.465017
20	21.0	0.471477
21	22.0	0.480605
22	23.0	0.493940
23	24.0	0.509632
24	25.0	0.526341
25	26.0	0.543684
26	27.0	0.560668
27	28.0	0.578696
28	29.0	0.599189
29	30.0	0.621109

30	31.0	0.642639
31	32.0	0.663168
32	33.0	0.677804
33	**34.0**	**0.681514**
34	35.0	0.676793
35	36.0	0.668481
36	37.0	0.659353
37	38.0	0.651027
38	39.0	0.644027
39	40.0	0.638061
40	41.0	0.631717
41	42.0	0.622770
42	43.0	0.611545
43	44.0	0.598808
44	45.0	0.585692
45	46.0	0.572773
46	47.0	0.560671
47	48.0	0.548357
48	49.0	0.534073
49	50.0	0.518300

At **34th lag level** we could see **Positive influence** with **Home Price Index**.

Lets try to visualize this



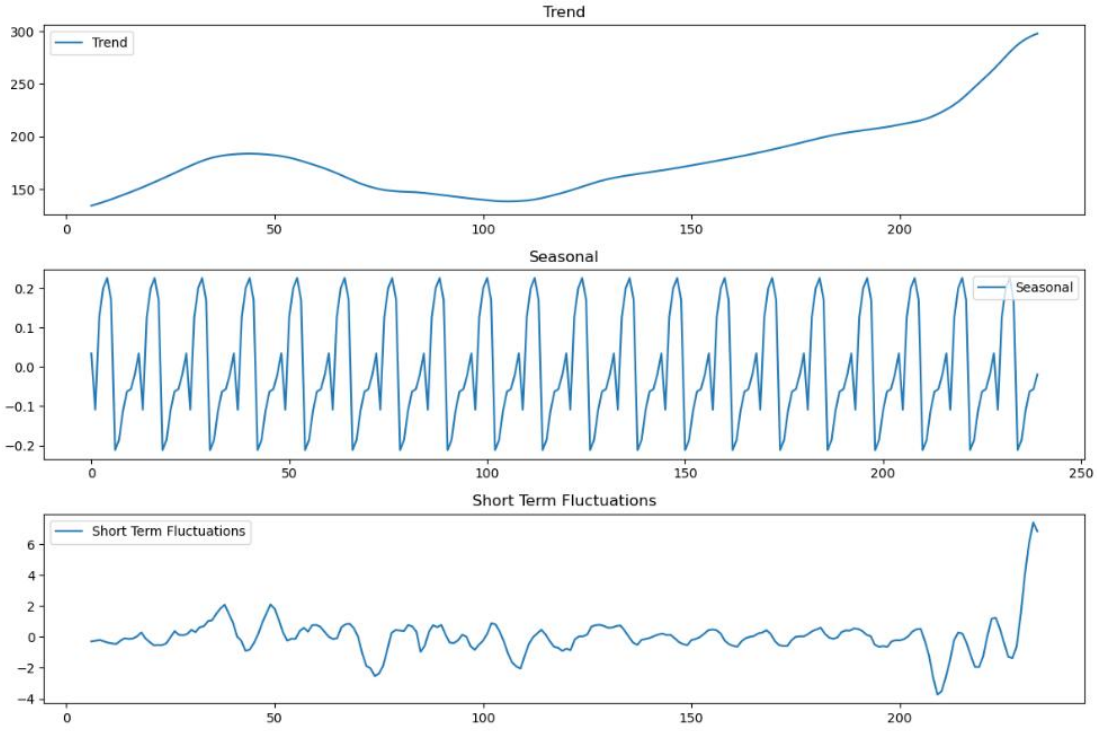
We could **consumer confidence index** at **34 time lag** is **positively** influencing **home price index**.

This can mean that **Consumer Confidence surveys** may have been conducted **34 month interval** prior to **House Sales**, so a **person** would have **appeared** to a **Consumer Confidence survey 34 months before purchasing a House or Home**.

As **Consumer Confidence Index** is **helping** only with a **34 time lag**, it will be **difficult** to use this **data** for **forecasting** as we cannot be able to **determine** in which **time level** the **consumer confidence index** will **influence future home price index**.

So its **better** to **not consider Consumer Confidence Index** as a feature for our dataset.

<u>Trend,Seasonality and Decomposition</u>:



TREND:

1. We could see a **slight increase from 2004** and a **flatness** from 2**005 to 2008**.

2. From **2009 to 2013** there is **flatness** in **lower** level and **rise** from **2014.**

3. **After 2014** we could see **significant rise continuously.**

SEASONALITY:

The data **doesn't show** any **changes** in **behaviour or trend** on a **seasonal level** and its pretty much **constant**.

SHORT TERM FLUCTUATIONS:

We could see **many small high low dip fluctuations** and **2022** we could see **significant increase level.**

5. MODEL BUILDING:

We will be **building FB Prophet model** where it will be **trained** with data from **January 2003 to December 2021** and **forecasted** for J**anuary 2022 to December 2022 data**.

Here **Home Price Index** will be **forecasted** by the **model**.

Initially **train** and **test data** is being **prepared**.

```
train_data = df_model[df_model['ds'] <= '2021-12-01']
test_data = df_model[(df_model['ds'] >= '2022-01-01') & (df_model['ds'] <= '2022-12-01')]
```

Model is **built**

```
model = Prophet(growth='linear',)
for i in train_data.drop(['y','ds'],1):
    model.add_regressor(i,mode='multiplicative')

model.fit(train_data)

future = test_data

forecast = model.predict(future)
```

Model is **evaluated** with **Actual Home Price Index.**


```
metrics.r2_score(actual,forecast['yhat'])
```

```
0.6719489563230197
```

```
metrics.mean_absolute_error(actual,forecast['yhat'])
```

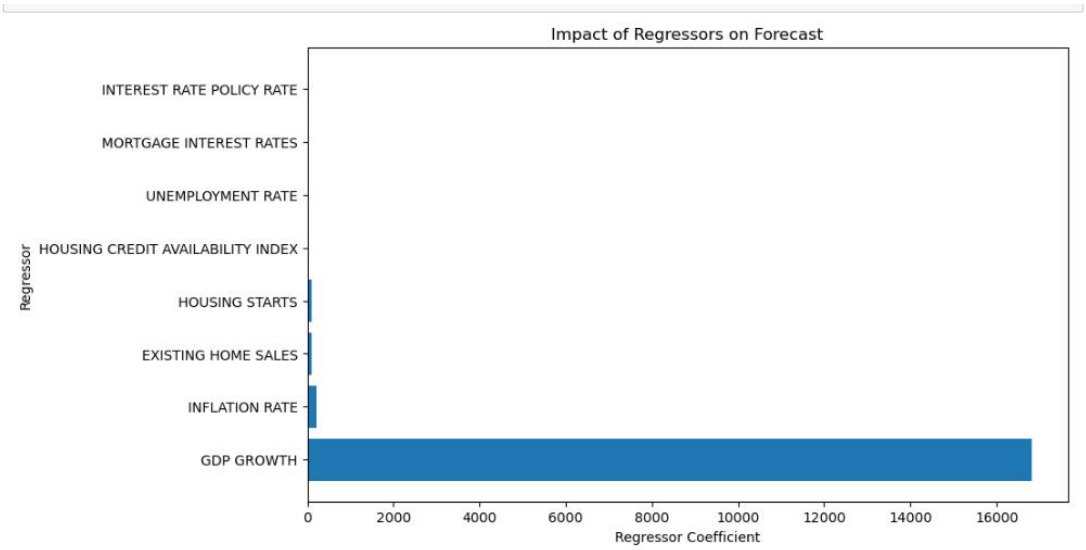
```
2.658352165569402
```

We could see only **67% accuracy level** is achieved, this is **due** to **low records data**, as **totally** including **train and test data** we have only **240 records**, **splitting** it into **train and test** will be **228** for training and **12** for testing.

Due to low records the model is **Underfitting** if **more data** is collected it will **perform well**.

6. FEATURE IMPORTANCE FOR THE MODEL:

Feature Importance plot is **created** based on **coefficients** from the **model**.



We could see that **GDP GROWTH, INFLATION RATE,EXISTING HOME SALES and HOUSING STARTS** are **influencing** the **model** pretty much.

7. FINAL INFERENCE:

Finally we can **conclude** that **GDP GROWTH,INFLATIONRATE and EXISTING HOME SALES** are **key factors** which **influence HOME PRICE INDEX** very much.