US HOME VALUE FORECAST

HOLT'S EXPONENTIAL FORECASTING

BACKGROUND & PROBLEM STATEMENT

The real estate market is heavily influenced by various factors that affect home values over time. Zillow's Home Value Index (ZHVI) provides a reliable measure of home values across cities in the US, but accurately predicting future trends remains a challenge. Using historical ZHVI data from 1996 to 2020, this project aims to develop a forecasting model based on Holt's Exponential Smoothing method to predict future home values for multiple US cities beyond 2020.

Data Set - City_AllHomes

The dataset provides **city-wise time series home** values across the USA using Zillow's Home Value Index (**ZHVI**) - an index that measures the typical home value in a given area based on various influencing factors, such as market conditions, location, and economic indicators.

Data Description-

1. **RegionID** A unique identifier for each city.
2. **RegionName** The name of the city or location.
3. **State The state** in which the city is located.
4. **ZHVI -** typical home value for that quarter of te year for respective city
5. **Metro** The metropolitan area to which the city belongs, if applicable.
6. **SizeRank** The ranking of the city based on population size, with lower ranks representing larger cities.

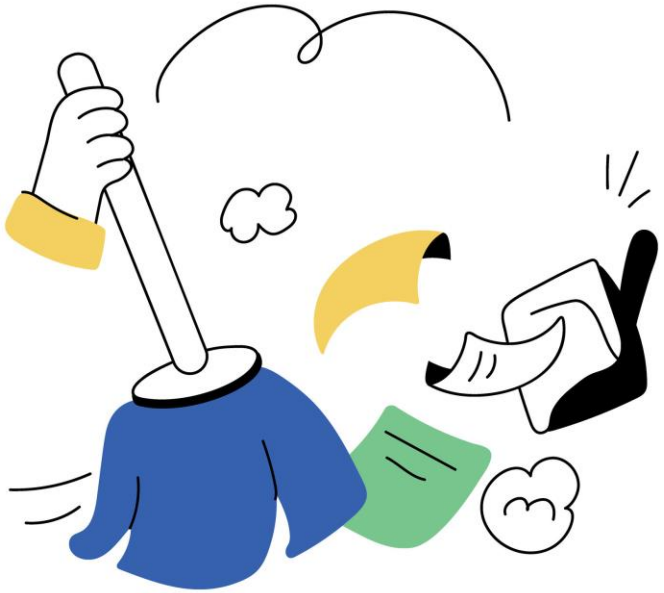
DATASET (Before Data Cleaning)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		RegionID	SizeRank	RegionName	RegionType	StateName	State	Metro	CountyName	1/31/1996	2/29/1996	3/31/1996	4/30/1996	5/31/1996	6/30/1996	7/31/1996	8/31/1
2	0	6181	0	New York	City	NY	NY	New York-Newark-Jersey City	Queens County	196258	195693	195383	194836	194652	194520	194447	194
3	1	12447	1	Los Angeles	City	CA	CA	Los Angeles-Long Beach-Anaheim	Los Angeles County	185649	185625	185645	185706	185755	185852	186014	185
4	2	39051	2	Houston	City	TX	TX	Houston-The Woodlands-Sugar Land	Harris County	93518	93633	93636	93811	93966	94079	94126	94
5	3	17426	3	Chicago	City	IL	IL	Chicago-Naperville-Elgin	Cook County	130920	130772	130511	130495	130172	130291	130143	130
6	4	6915	4	San Antonio	City	TX	TX	San Antonio-New Braunfels	Bexar County	94041	94007	93973	93957	93895	93880	93880	93
7	5	13271	5	Philadelphia	City	PA	PA	Philadelphia-Camden-Wilmington	Philadelphia County	70764	70673	70592	70445	70333	70231	70188	70
8	6	40326	6	Phoenix	City	AZ	AZ	Phoenix-Mesa-Scottsdale	Maricopa County	98258	98556	98876	99501	100096	100639	101151	101
9	7	18959	7	Las Vegas	City	NV	NV	Las Vegas-Henderson-Paradise	Clark County	148835	148713	148809	148805	148966	148990	149166	149
10	8	54296	8	San Diego	City	CA	CA	San Diego-Carlsbad	San Diego County	194249	193854	193590	193063	192736	192486	192219	191
11	9	38128	9	Dallas	City	TX	TX	Dallas-Fort Worth-Arlington	Dallas County	101339	101335	101526	101895	102282	102536	102533	102
12	10	10221	10	Austin	City	TX	TX	Austin-Round Rock	Travis County	183868	182798	181167	178841	178270	178821	180823	181
13	11	33839	11	San Jose	City	CA	CA	San Jose-Sunnyvale-Santa Clara	Santa Clara County	218575	217984	217791	217482	217594	217845	218067	218
14	12	25290	12	Jacksonville	City	FL	FL	Jacksonville	Duval County	84888	84999	85111	85318	85502	85699	85911	86
15	13	32149	13	Indianapolis	City	IN	IN	Indianapolis-Carmel-Anderson	Marion County								
16	14	20330	14	San Francisco	City	CA	CA	San Francisco-Oakland-Hayward	San Francisco County	297118	296155	295503	294359	293536	293161	292873	293
17	15	24043	15	Charlotte	City	NC	NC	Charlotte-Concord-Gastonia	Mecklenburg County	127476	127747	128008	128602	129158	129816	130460	131
18	16	18172	16	Fort Worth	City	TX	TX	Dallas-Fort Worth-Arlington	Tarrant County	91307	91450	91592	92007	92380	92649	92839	93
19	17	7481	17	Tucson	City	AZ	AZ	Tucson	Pima County	105070	105415	105817	106408	106859	107304	107694	108
20	18	10920	18	Columbus	City	OH	OH	Columbus	Franklin County	93008	93127	93261	93543	93891	94240	94599	94
21	19	12455	19	Louisville	City	KY	KY	Louisville/Jefferson County	Jefferson County	88266	88546	88755	89214	89594	89977	90370	90
22	20	13121	20	Orlando	City	FL	FL	Orlando-Kissimmee-Sanford	Orange County	100361	100463	100652	100939	101342	101672	101946	102
23	21	17933	21	El Paso	City	TX	TX	El Paso	El Paso County	85886	85804	85787	85764	85867	85993	86089	86
24	22	17762	22	Detroit	City	MI	MI	Detroit-Warren-Dearborn	Wayne County								
25	23	11093	23	Denver	City	CO	CO	Denver-Aurora-Lakewood	Denver County	126922	127499	127991	129031	129940	130819	131687	132

DATA PRE-PROCESSING

	A	B	C	D	E	F	G	H
1	date	New York zhvi	Los Angeles	Houston	Chicago	Phoenix		
2	1/31/1996	196258	185649	93518	130920	98258		
3	2/29/1996	195693	185625	93633	130772	98556		
4	3/31/1996	195383	185645	93636	130511	98876		
5	4/30/1996	194836	185706	93811	130495	99501		
6	5/31/1996	194652	185755	93966	130172	100096		
7	6/30/1996	194520	185852	94079	130291	100639		
8	7/31/1996	194447	186014	94126	130143	101151		
9	8/31/1996	194313	185977	94220	130523	101637		
10	9/30/1996	194271	185812	94436	131041	102107		
11	10/31/1996	194341	185482	94701	131790	102559		
12	11/30/1996	194496	185346	95017	132635	103021		
13	12/31/1996	194755	185219	95344	133214	103504		
14	1/31/1997	195144	184629	95284	133819	103992		
15	2/28/1997	196135	184056	95039	134352	104474		
16	3/31/1997	197185	183653	94827	135028	104907		
17	4/30/1997	198143	183835	94820	135143	105318		
18	5/31/1997	198601	184073	94841	135857	105741		
19	6/30/1997	199185	184301	94967	135900	106169		
20	7/31/1997	199877	184759	95173	136201	106571		
21	8/31/1997	200789	185678	95360	135303	106967		
22	9/30/1997	201707	186633	95427	133972	107431		
23	10/31/1997	202620	187795	95474	132069	107926		

ibm project, zillow city zhvi a | clean data | NY | LA | +



WHAT IS HOLT'S FORECASTING ??

Holt's Forecasting, also known as Holt's Exponential Smoothing or Double Exponential Smoothing, is a statistical forecasting method used to predict future values of a time series that exhibits a linear trend

Key components of Holt's Forecasting:

- Level (Lt): The average value of the time series at a particular point in time.
- Trend (Tt): Rate at which the level is changing over time.
- Seasonality (St) : The seasonality component accounts for recurring patterns in the data that repeat over a specific period.
- Smoothing Parameter (**α**): Controls the weighting given to recent observations in calculating the level component.
- Trend Smoothing Parameter (**β**): Controls the weighting given to recent observations in calculating the trend component.
- Seasonal Smoothing Parameter (**γ**): Controls the weighting given to recent observations in calculating the seasonal component.

METHODOLOGY

- CONNECT THE DATASET
- READ THE DATASET
- INITIALIZE THE FORECASTING VARIABLES

```
[ ]  import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
```

```
[ ]  ny=pd.read_csv('//content/drive/MyDrive/NY_ZHVI.csv')
     ny.head()
```

```
[ ]  def initial_T_L():
       current=ny['New York zhvi'][:23]
       next=ny['New York zhvi'][:23].shift(-1)

       dif=next-current
       s=dif.sum()
       T0=abs((ny['New York zhvi'].iat[23]-ny['New York zhvi'].iat[0])/23) # 24= number of months from 1996-1997
       L0=ny['New York zhvi'].iat[23]

       print("Initial Trend of 1996 is: ", T0)
       print("The starting level of Forecast is: ",L0)
       return L0,T0
```

```
[ ]  L0, T0 = initial_T_L()
     ny['Lt']=0.000
     ny['Tt']=0.00000

     ny['Lt'].iat[23]=L0
     ny['Tt'].iat[23]=T0

     print(ny['Lt'].iat[23],ny['Tt'].iat[23])
```

```
Initial Trend of 1996 is:  379.8695652173913
The starting level of Forecast is:  204995.0
204995.0 379.8695652173913
```

APPLY HOLT'S FORECASTING

- LEVEL EQUATION

  $Lt = \alpha Yt + (1 - \alpha)(Lt\text{-}1 + Tt\text{-}1)$

- TREND EQUATION

  $Tt = \beta(Lt - Lt\text{-}1) + (1 - \beta)Tt\text{-}1$

- FORECAST EQUATION

  $Ft+h = Lt + hTt$

```
[ ]  def Level(i):
       a=0.9
       ny['Lt'].iat[i]= ((a* ny['New York zhvi'].iat[i])+ ((1-a)*(ny['Lt'].iat[i-1] + ny['Tt'].iat[i-1])))
```

```
[ ]  def Trend(i):
       b=0.9
       ny['Tt'].iat[i]=(b*((ny['Lt'].iat[i])-(ny['Lt'].iat[i-1])))+ ((1-b)*( ny['Tt'].iat[i-1]))
```

```
[ ]  def Forecast_train():
       ny['Ft']=0.0000
       for i in range(24, 241):
         ny['Ft'].iat[i] = ny['Lt'].iat[i-1] + ny['Tt'].iat[i-1]

     Forecast_train()
     ny.head(30)
```

Testing Future Forecast and Calculating error

```
[ ]  def Forecast_test_future():
       j=2 #the preiod for forecasting starting 2/29/2016
       for i in range(241,420):
         ny['Ft'].iat[i]=ny['Lt'].iat[239]+(j * ny['Tt'].iat[239])
         j=j+1

     Forecast_test_future()
```

```
[ ]  def error():
       ny["Error"]=0.0000
       for i in range(24, 291): #uptill 240
         ny["Error"].iat[i] = abs(ny['New York zhvi'].iat[i] - ny['Ft'].iat[i])

     error()
     ny.head(290)
```

	date	New York zhvi	Lt	Tt	Ft	Error
0	1/31/1996	196258.0	0.0	0.0	0.000000	0.000000
1	2/29/1996	195693.0	0.0	0.0	0.000000	0.000000
2	3/31/1996	195383.0	0.0	0.0	0.000000	0.000000
3	4/30/1996	194836.0	0.0	0.0	0.000000	0.000000
4	5/31/1996	194652.0	0.0	0.0	0.000000	0.000000
...	...	...	...	...	...	...
285	10/31/2019	655613.0	0.0	0.0	703986.542236	48373.542236
286	11/30/2019	654394.0	0.0	0.0	707417.132568	53023.132568
287	12/31/2019	653930.0	0.0	0.0	710847.722900	56917.722900
288	1/31/2020	653901.0	0.0	0.0	714278.313232	60377.313232
289	2/29/2020	653565.0	0.0	0.0	717708.903564	64143.903564

290 rows × 6 columns

Calculating Mean Absolute Deviation (MAD)

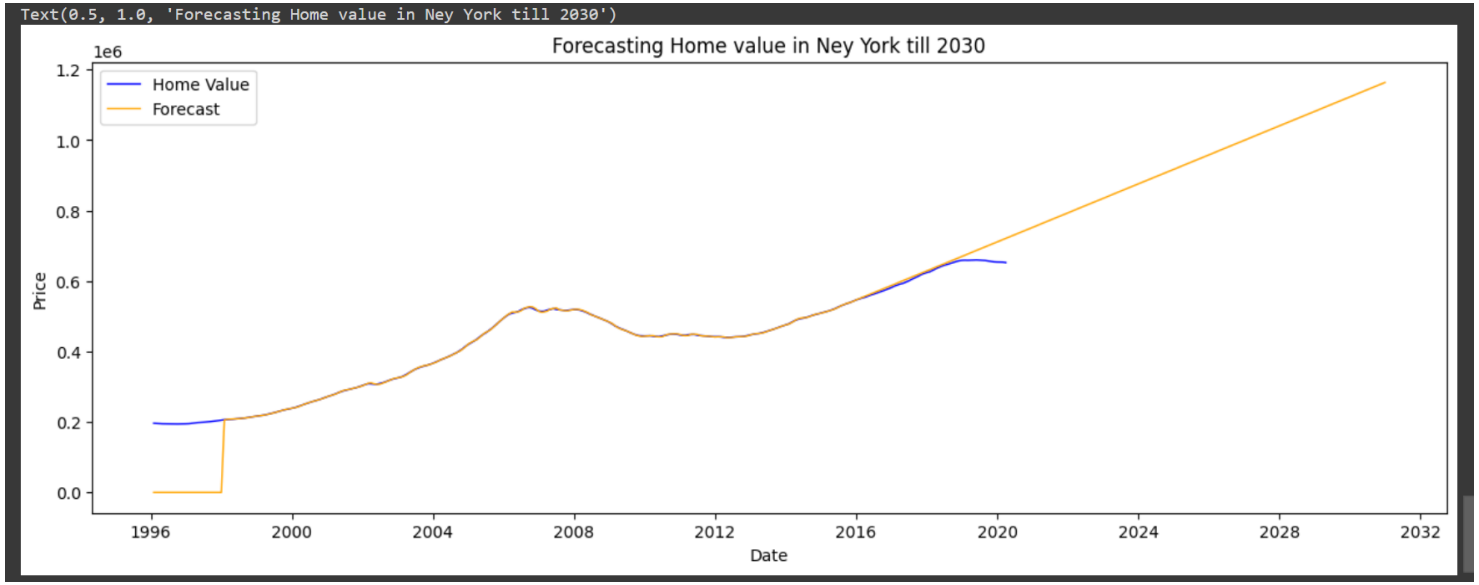
```
[ ]  def MAD():
       sum_error=0
       for x in range(24,291):
         sum_error += ny['Error'].iat[x]

       print("Sum of error: ",sum_error)
       mad=sum_error/267   #(290-23) = 267 = number of eror calculations
       print("Mean Absolute Deviation of the model is: ",mad)

     MAD()
```

```
Sum of error:  949476.2449425324
Mean Absolute Deviation of the model is:  3556.090805028211
```

Graphical Representation & Analysis

- There is a slight increasing trend in the forecasted value for house price prediction.
- The Forecasted Home values have been accurate and close to the real values from the year 2015 to the year 2020.
- After that, the real value stagnates, and the forecast continues in a linear trend based on its previous training.

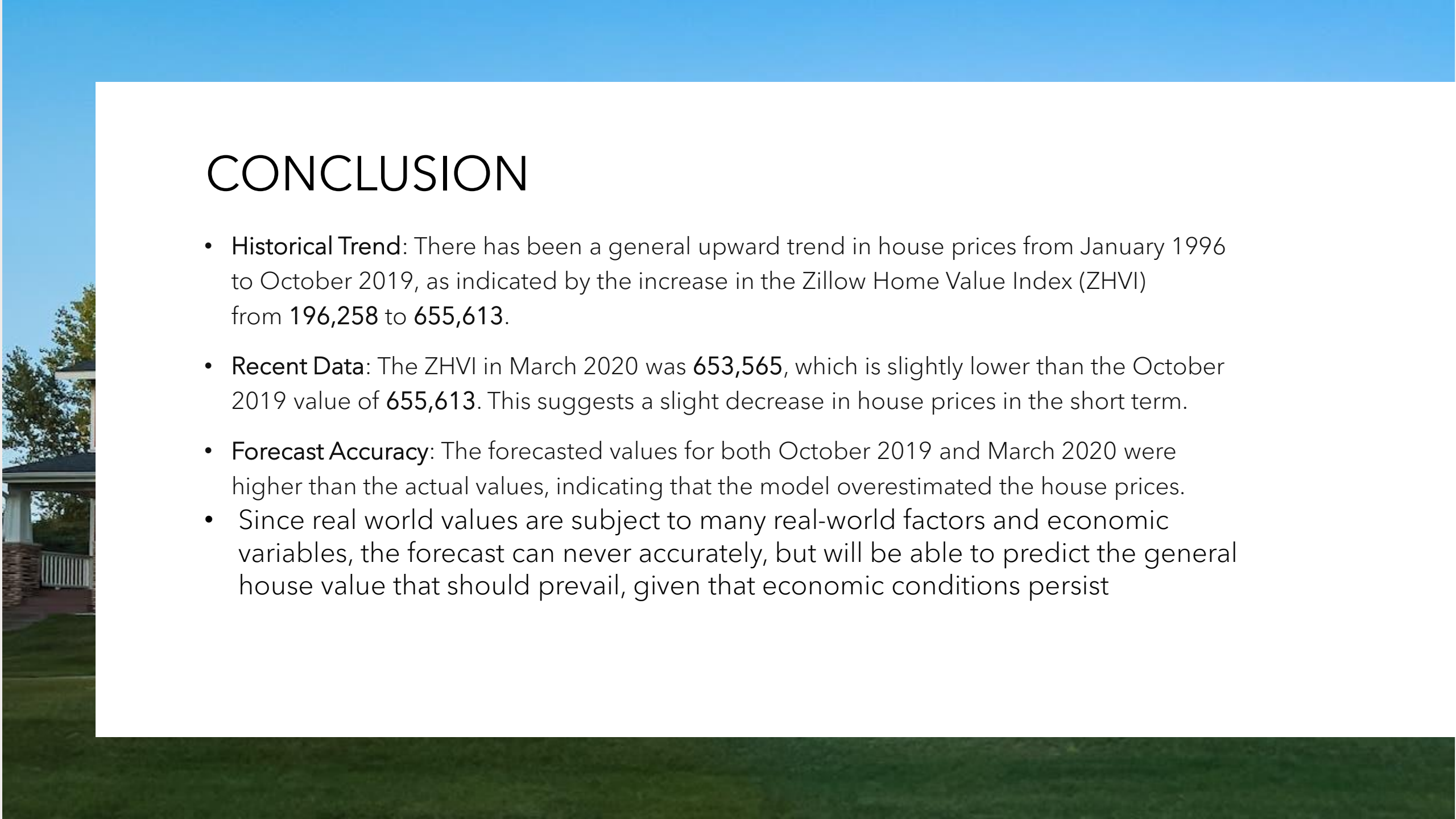
```
[ ]  ny['date'] = pd.to_datetime(ny['date'])

     plt.figure(figsize=(15,5))
     plt.plot(ny['date'],ny['New York zhvi'],'-',linewidth=1, color='b', label= 'Home Value')
     plt.plot(ny['date'],ny['Ft'],'-',linewidth=1, color ='orange', label="Forecast", )
     plt.xlabel('Date')
     plt.ylabel('Price')
     plt.legend()
     plt.title("Forecasting Home value in Ney York till 2030")
```



TESTING MODEL IN EXCEL

L_0	203785	
T_0	379.8696	
a	0.9	
b	0.9	
l_t	$\alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$	
T_t	$\gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$	
F_t	$Y_{t+1} = L_t + (h)T_t$	
mad	3556.091	

A	B	C	D	E	F	G
date	New York zhvi	L_t	T_t	F_t	E_t	
11/30/1997	203785			$\gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$		
12/31/1997	204995	204995	379.869565			
1/31/1998	206282	206191.3	1114.64522	205374.8696	907.13	907.1304
2/28/1998	206895	206936.1	781.790157	207305.9322	410.93	410.9322
3/31/1998	207400	207431.8	524.304624	207717.8834	317.88	317.8834
4/30/1998	207784	207801.2	384.909325	207956.093	172.09	172.093
5/31/1998	208585	208545.1	708.003242	208186.1186	398.88	398.8814
6/30/1998	209417	209400.6	840.750008	209253.1151	163.88	163.8849
7/31/1998	210201	210205	808.057178	210241.3615	40.36	40.36152
8/31/1998	211248	211224.5	998.331581	211013.0933	234.91	234.9067
9/30/1998	212434	212412.9	1169.37044	212222.8409	211.16	211.1591
10/31/1998	213887	213856.5	1416.21427	213582.2545	304.75	304.7455
11/30/1998	215089	215107.4	1267.38509	215272.7397	183.74	183.7397
12/31/1998	216323	216328.2	1225.46025	216374.7591	51.76	51.75907
1/31/1999	217391	217407.3	1093.72496	217553.6362	162.64	162.6362
2/28/1999	218756	218730.5	1300.28421	218500.9886	255.01	255.0114
3/31/1999	220205	220187.6	1441.39993	220030.7831	174.22	174.2169
4/30/1999	222080	222034.9	1806.72756	221628.9782	451.02	451.0218
5/31/1999	224080	224056.2	1999.811	223841.6254	238.37	238.3746



CONCLUSION

- **Historical Trend**: There has been a general upward trend in house prices from January 1996 to October 2019, as indicated by the increase in the Zillow Home Value Index (ZHVI) from **196,258** to **655,613**.
- **Recent Data**: The ZHVI in March 2020 was **653,565**, which is slightly lower than the October 2019 value of **655,613**. This suggests a slight decrease in house prices in the short term.
- **Forecast Accuracy**: The forecasted values for both October 2019 and March 2020 were higher than the actual values, indicating that the model overestimated the house prices.
- Since real world values are subject to many real-world factors and economic variables, the forecast can never accurately, but will be able to predict the general house value that should prevail, given that economic conditions persist

REFERENCE

DATASET LINK:

zillow-all-homes-data

Saanvi – L014
Arpita - L005