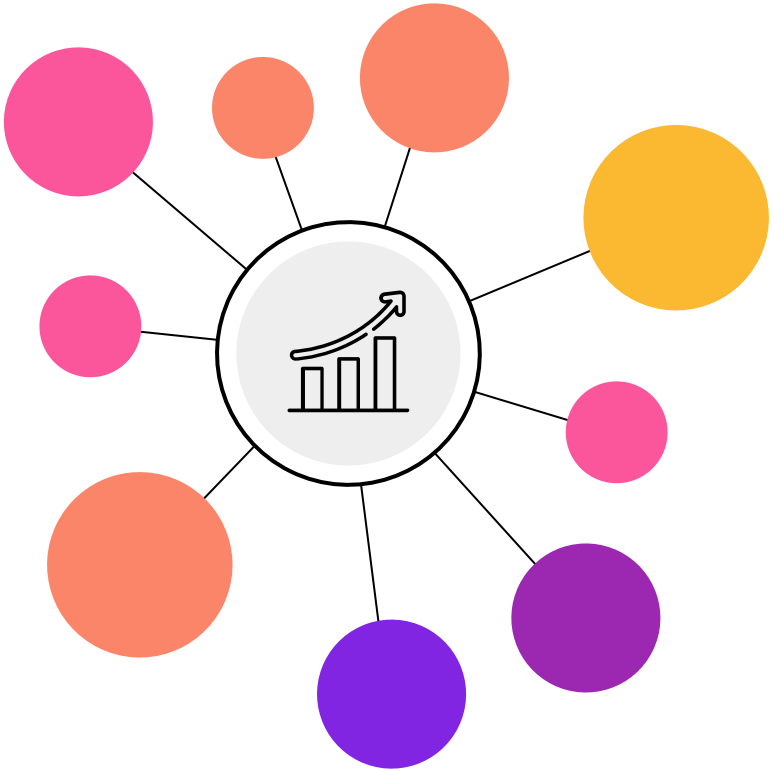
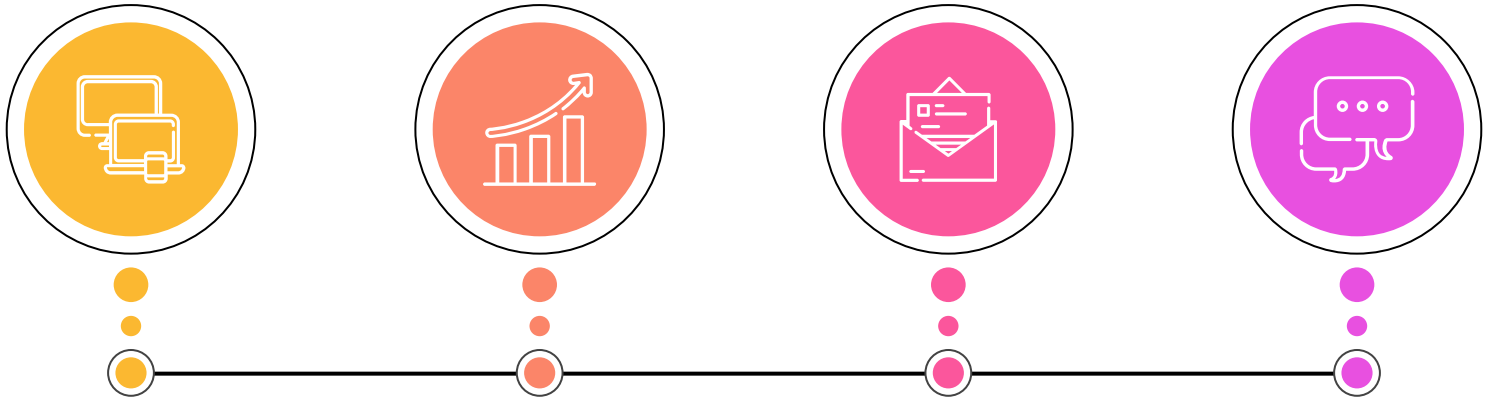


Real Estate Value Prediction for Different Locations

1. Muthoni Kahuko
2. Ndanu Mwatu
3. Denis Ochieng

Project Workflow



Business understanding

Data preparation

Modeling

Recommendation

Project Overview

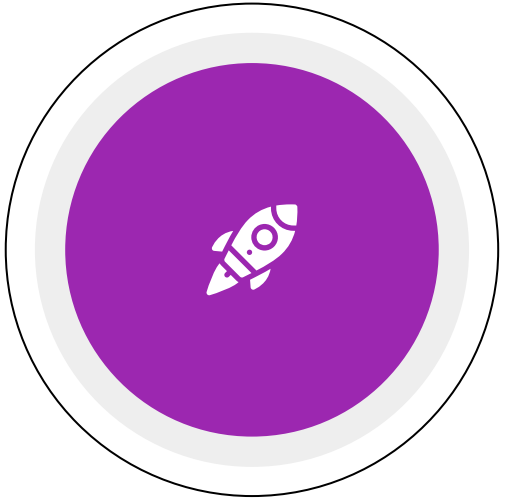
MDN real estate investment firm is currently looking to expand their horizons and invest in different locations that are **predicted to do well in the next few years**. The company wants to do their due diligence and have data drive their decisions

Due to their need for data, the company requested us,the data analyst consultants, develop a model that can predict **the future of real estate in different states**. MDN seeks to gain insight into the best locations to invest in and help advise clients where to buy property.

This project uses **Time Series modelling** to build predictive models for Real Estate prices in USA.

Business problem

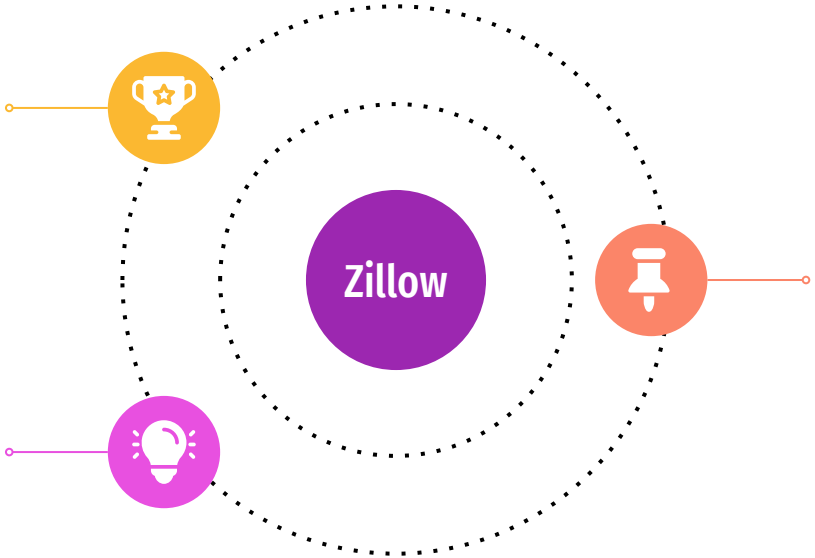
MDN real estate investment firm is currently looking to expand their horizons and invest in different locations that are predicted to do well in the next few years



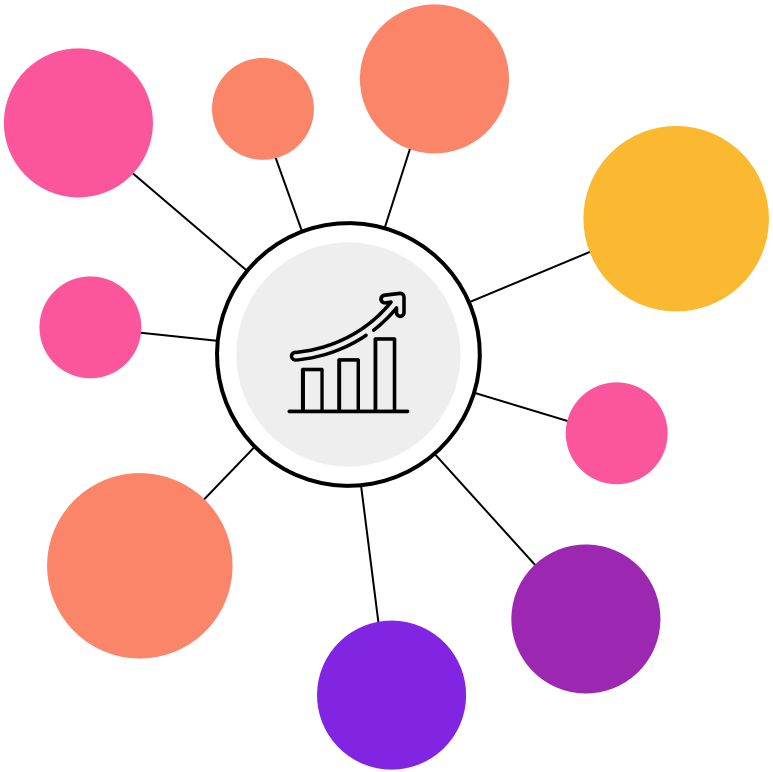
Specific Objectives

Identify prime investment locations by pinpointing the top 10 states with well to do real estate

Identify prime investment locations by highlighting the top 10 zip codes with highest value of property



Develop Time series models that can forecast the future of real estate prices and identify the best performing one based on lowest RMSE.



Data preparation

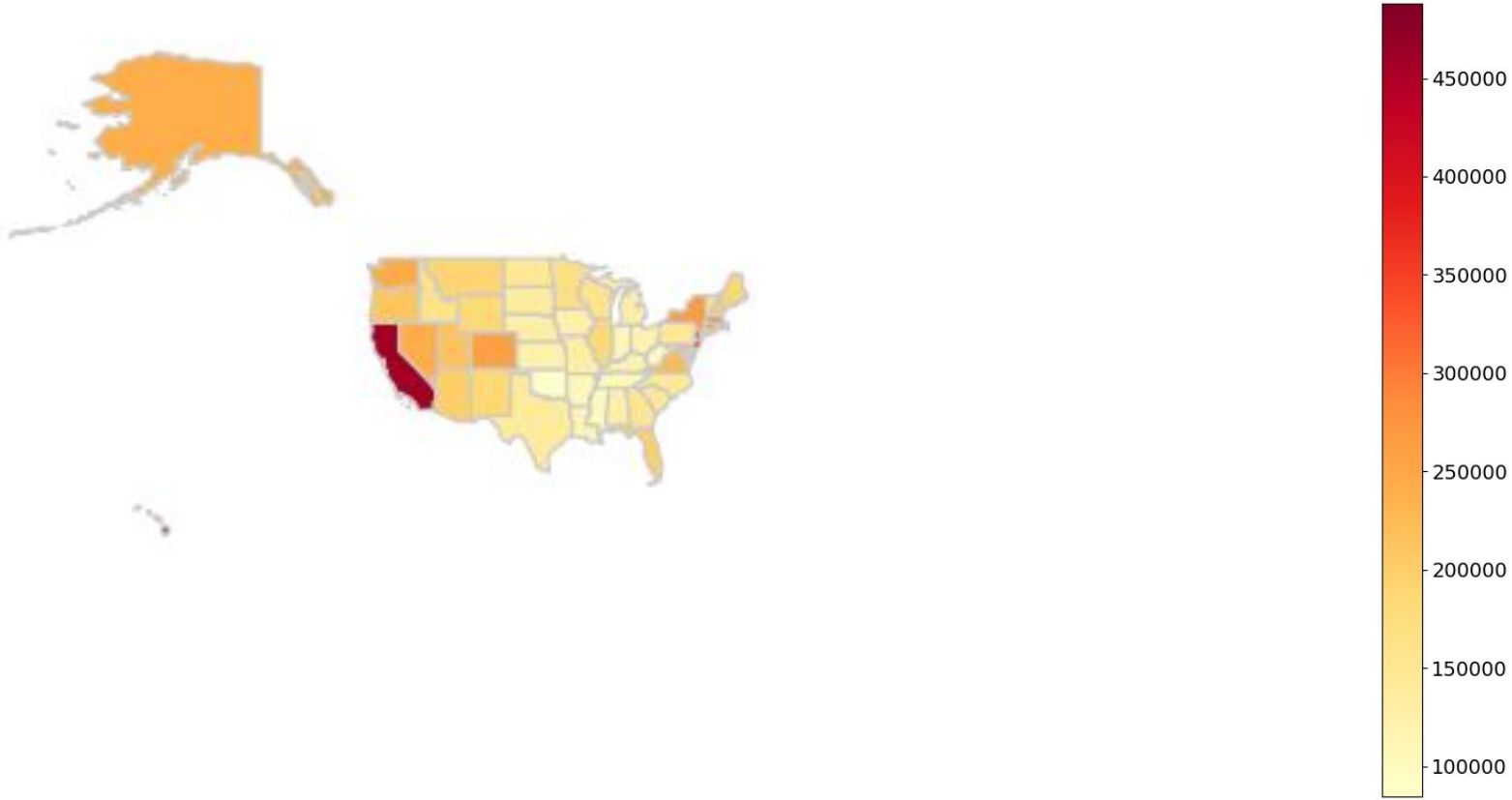
Data understanding

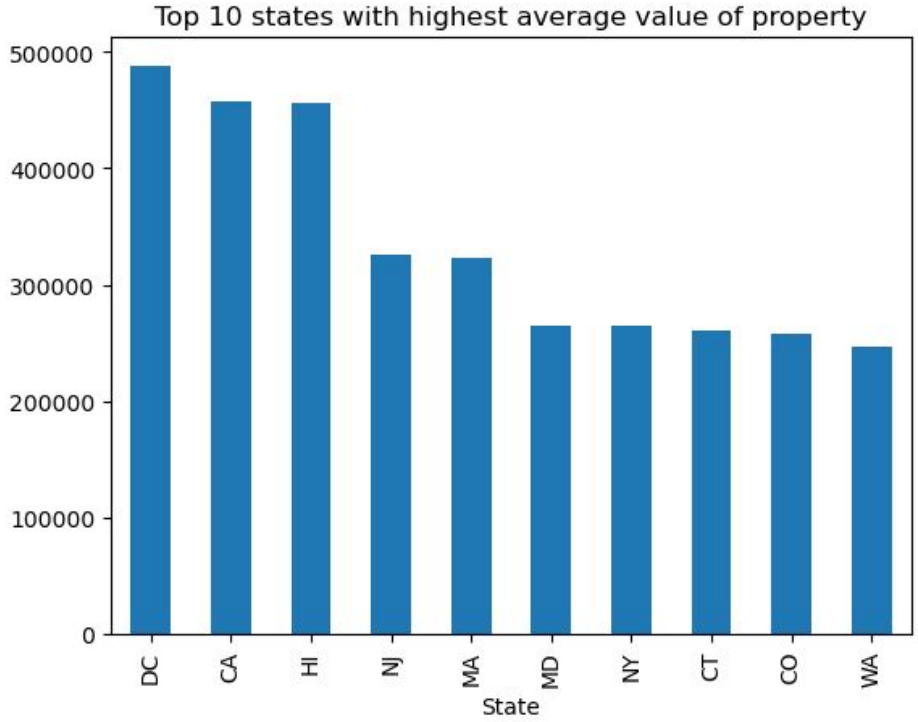
We will be using a data set called Zillow dataset that contains 272 rows Region Id, Region name, City, State, Metro, CountyName, Size rank, and the remaining 265 are the Time series values

They are 14723 rows

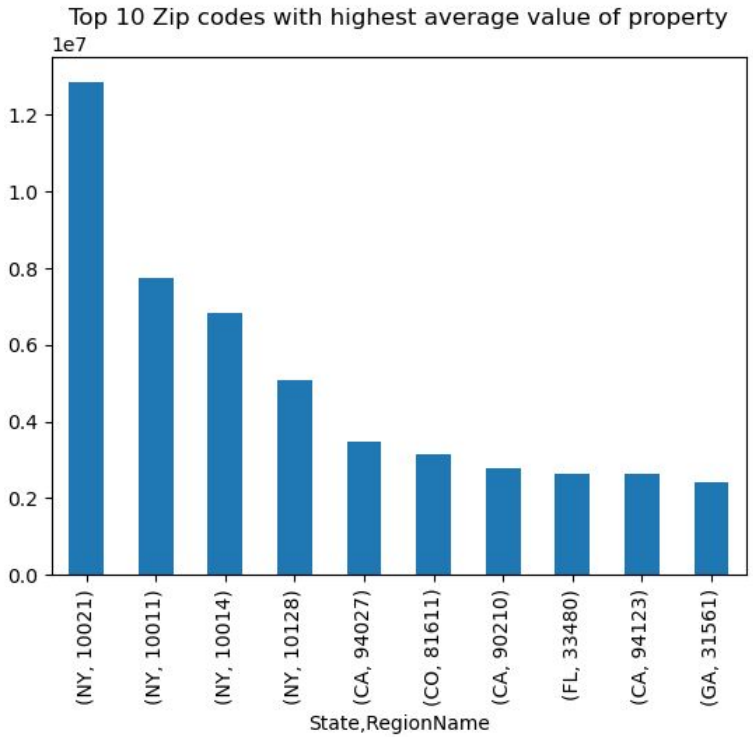
We decided to melt the dataset to allow the 265 columns to become rows

Average Property Value by State





After running through the code we realised that the DC has the highest average when it comes to the value of property

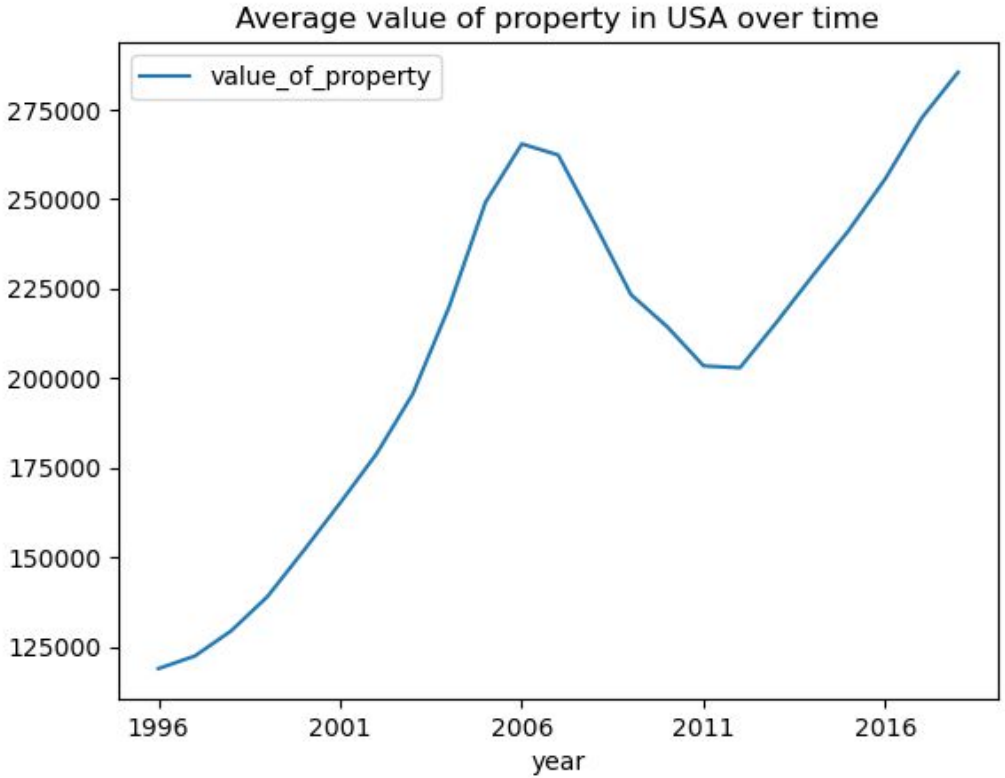


We also realised that the New york has the highest average when it comes to the value of property in the zipcodes

From this graph we noticed that from **1996 to about 2007** there was a great rise in the real estate market and according to the economic climate

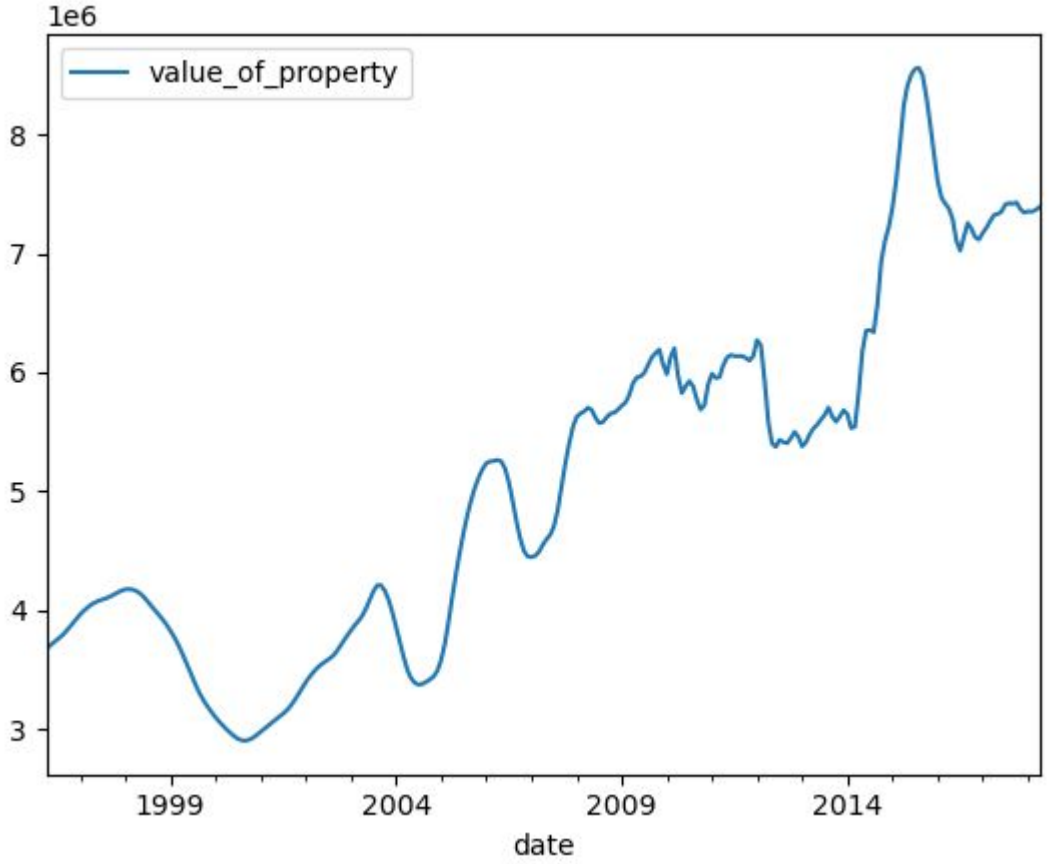
The drop was constituted by the **great recession of 2008**. From 2008 to about 2012 we can therefore see from the plot, a steep decline of the real estate market.

From **2012 to 2018**, we realize the real estate market recovered from the great recession.

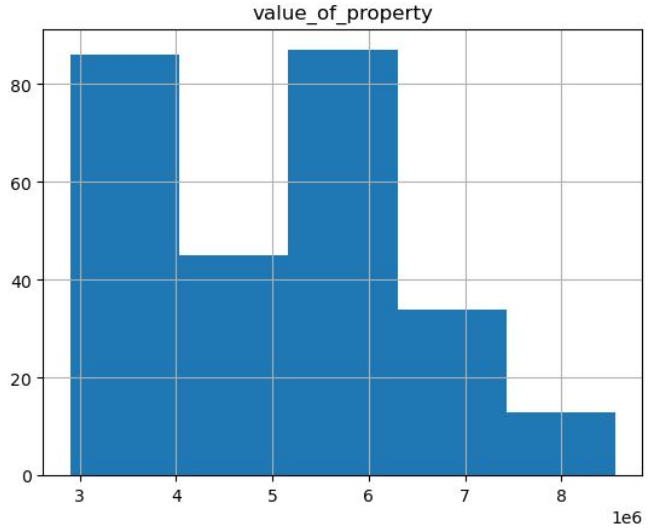


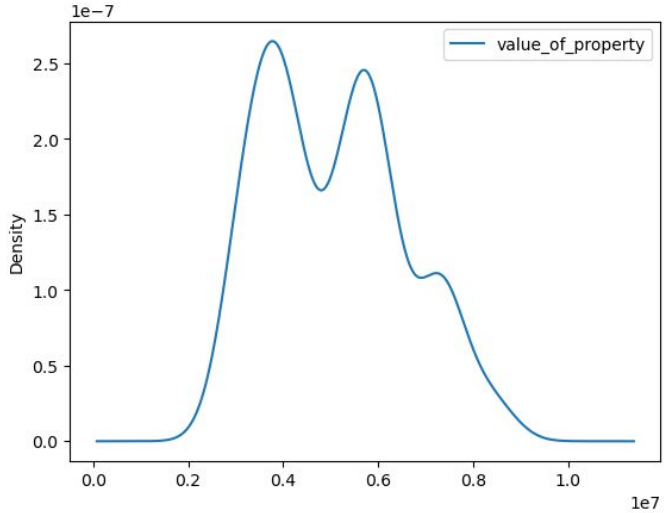
Zip code 10128 in New York

We selected zip code **10128** in New York, for data processing and modeling.



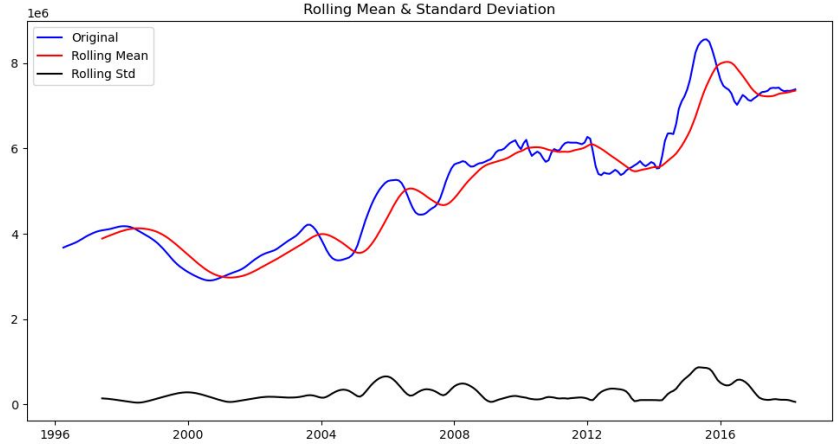
Zip code 10128 in New York

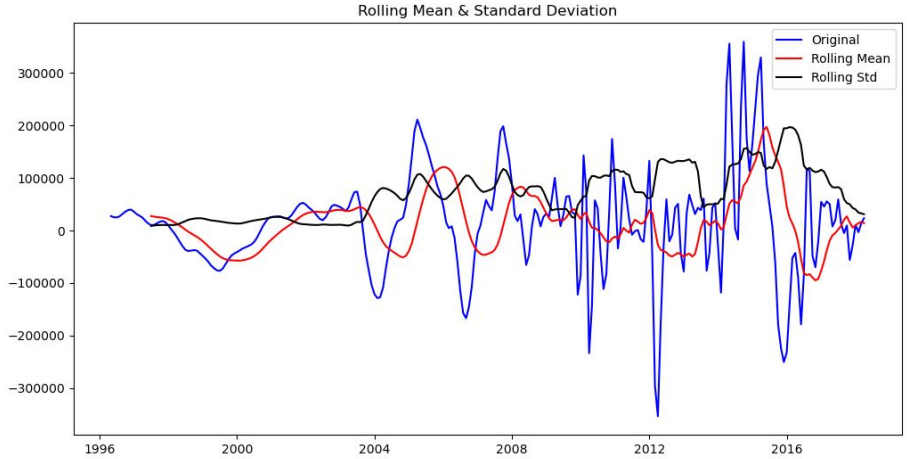




The line plot suggests a **linear trend,** while the histogram and density plot suggests the data is not normally distributed

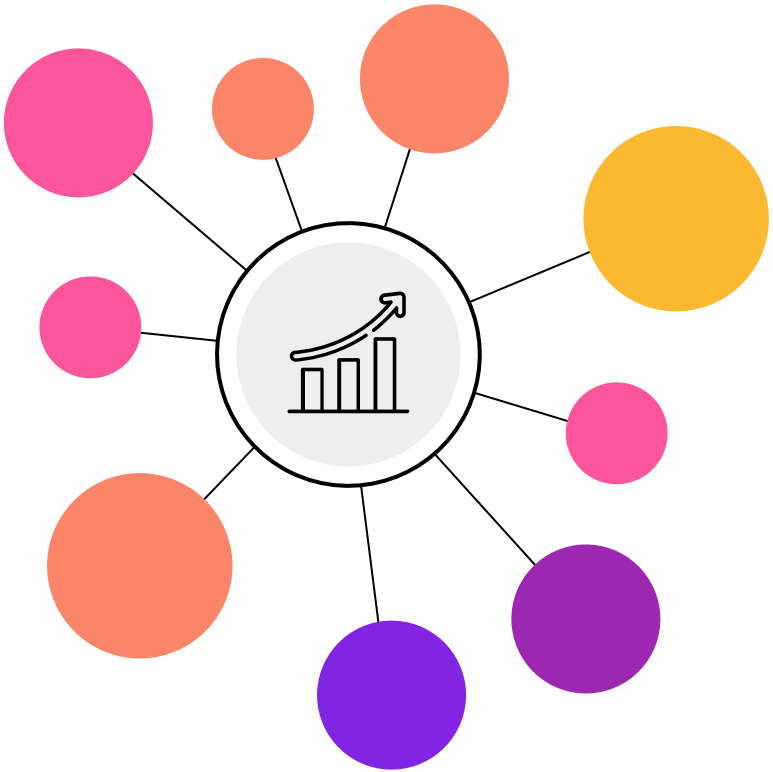
Stationarity





Data before differencing: Dickey Fuller result **p-value of 0.93**

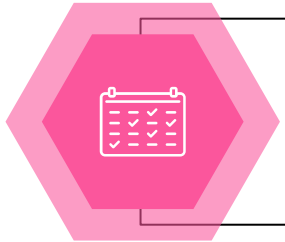
Data after differencing: Dickey Fuller result p-value of **0.000147**



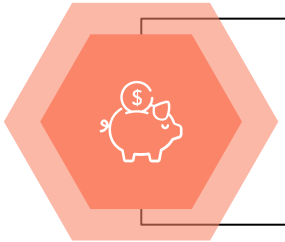
Modeling

Different models

Arima model as baseline



Arima model using PMDArima to get best parameter



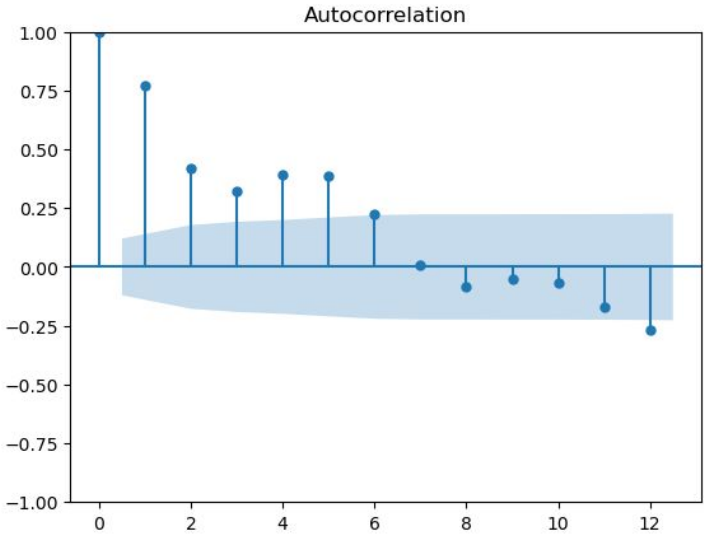
FB Prophet model

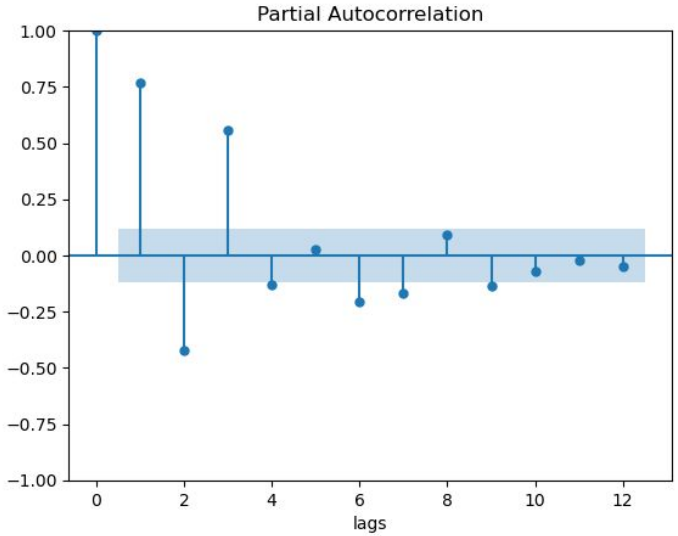
Arima model as baseline

```
[ ]  # Evaluation of baseline model

     from sklearn.metrics import r2_score, mean_squared_error
     from math import sqrt

     print('Model evaluation for NY_10128 : ','\n')
     print('R2 Score for NY_10128 : {0:.2f} %'.format(100*r2_score(NY_10128_diff['value_of_property'],predictions_base_NY_10128)),'\n')
     print('Root Mean Squared Error  for NY_10128 : ',sqrt(mean_squared_error(NY_10128_diff['value_of_property'],predictions_base_NY_10128)),'\n')
```

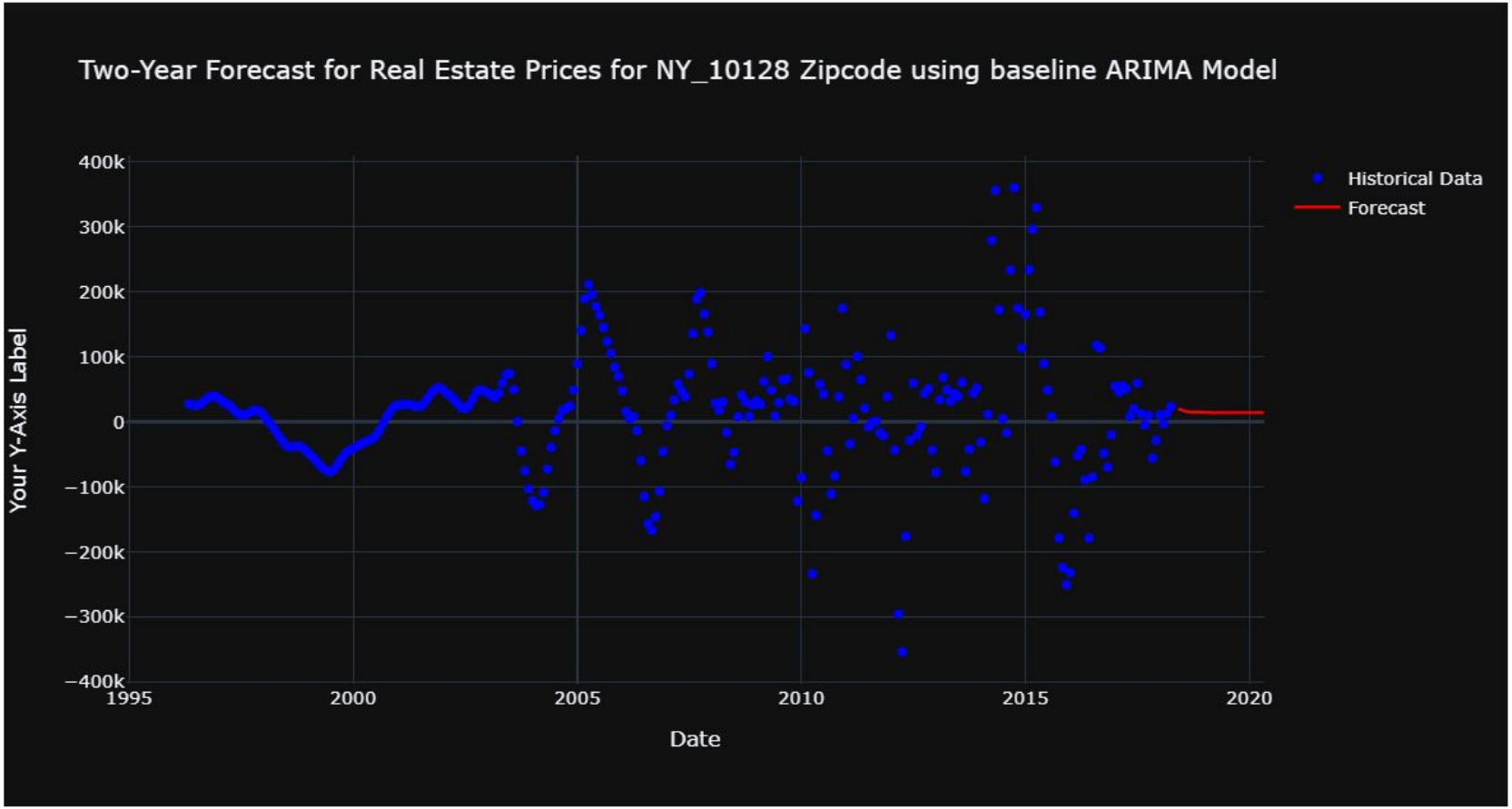
```
Model evaluation for NY_10128 :

R2 Score for NY_10128 : 73.58 %

Root Mean Squared Error  for NY_10128 :  48865.984827292436
```

- The baseline ARIMA model has an AIC of 6460 and a BIC of 6474. TheR2 score is about 73.6% and the RMSE is about 48,866
- We then proceed to investigate optimal valus of the parameters using PMD ARIMA for the second iterative model.





From ACF and PACF, baseline model order
1,0,1 is selected

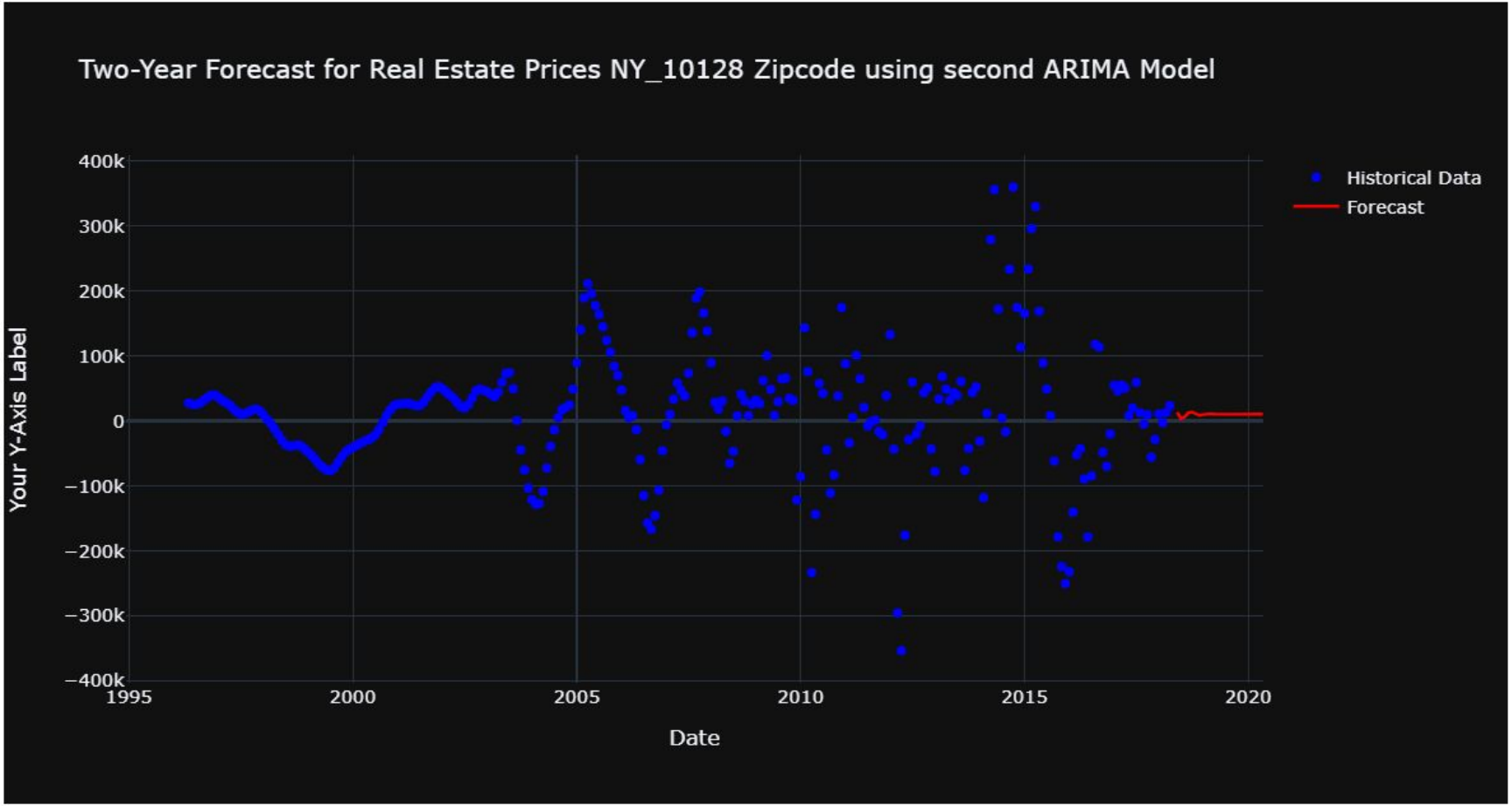

Two-Year Forecast for Real Estate Prices for NY_10128 Zipcode using baseline ARIMA Model
Historical Data
Forecast
Your Y-Axis Label
400k
300k
200k
100k
0
−100k
−200k
−300k
−400k
1995
2000
2005
2010
2015
2020
Date

Arima model using PMDArima to get best parameter

```
[ ]  print('Evaluation Result for whole data : ','\n')
     print('R2 Score for whole data : {0:.2f} %'.format(100*r2_score(NY_10128_diff['value_of_property'],predictions_PMDarma_NY_10128)),'\n')
     print('Root Mean Squared Error : ',sqrt(mean_squared_error(NY_10128_diff['value_of_property'],predictions_PMDarma_NY_10128)),'\n')
```

```
Evaluation Result for whole data :

R2 Score for whole data : 75.63 %

Root Mean Squared Error :  46931.60311641508
```

- The second ARIMA model has an AIC of 6415 and a BIC of 6433. TheR2 score is about 75.6% and the RMSE is about 46,8931
- This is a better performing ARIMA model than the baseline model.

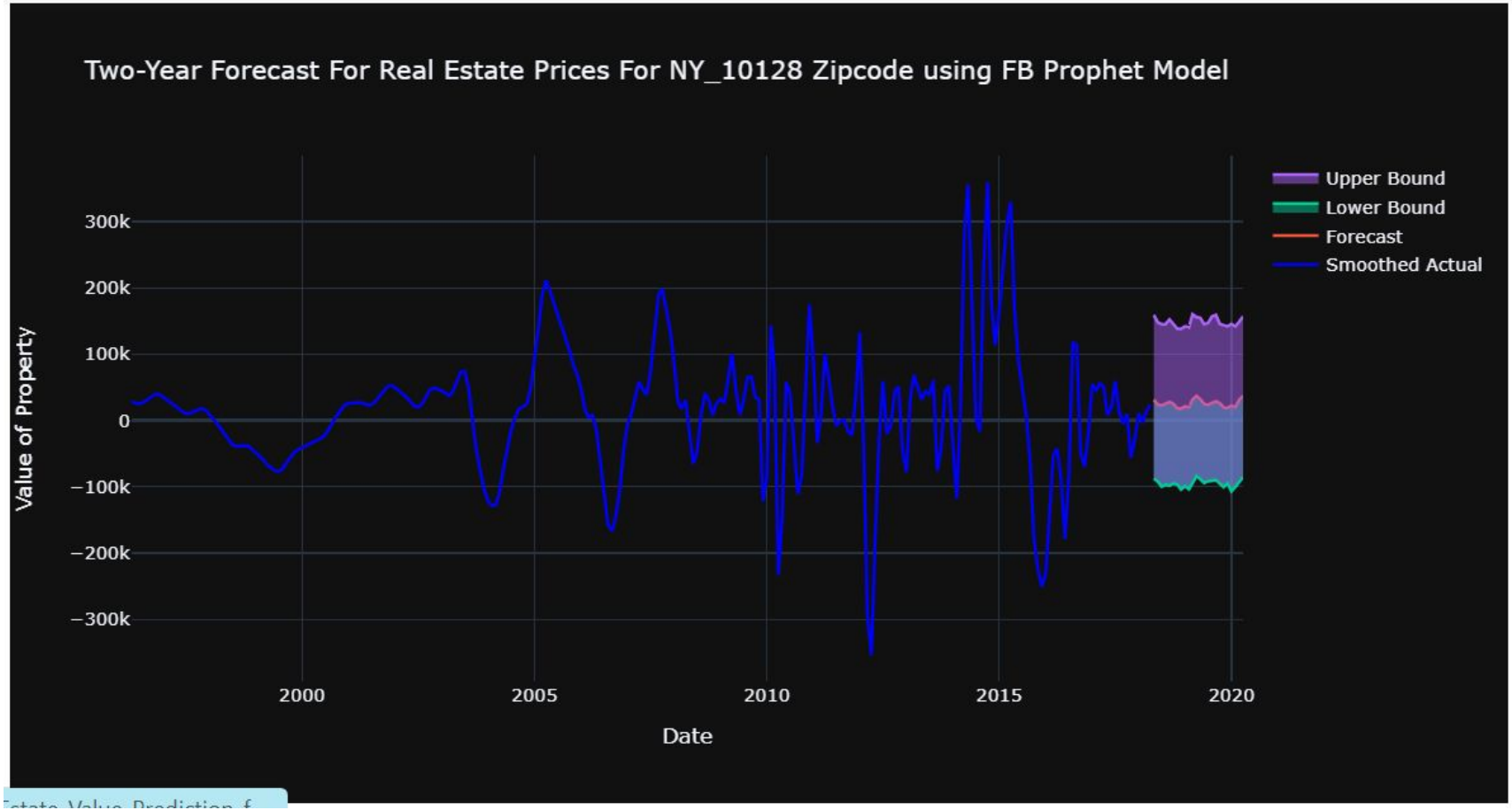

Two-Year Forecast for Real Estate Prices NY_10128 Zipcode using second ARIMA Model
Historical Data
Forecast
Your Y-Axis Label
400k
300k
200k
100k
0
−100k
−200k
−300k
−400k
1995
2000
2005
2010
2015
2020
Date

FB Prophet model

```
[ ]  print('Evaluation Result for whole data : ','\n')
     print('R2 Score for whole data : {0:.2f} %'.format(100*r2_score(NY_10128_diff['value_of_property'],predictions_PMDarma_NY_10128)),'\n')
     print('Root Mean Squared Error : ',sqrt(mean_squared_error(NY_10128_diff['value_of_property'],predictions_PMDarma_NY_10128)),'\n')
```

```
Evaluation Result for whole data :

R2 Score for whole data : 75.63 %

Root Mean Squared Error :  46931.60311641508
```

- The second ARIMA model has an AIC of 6415 and a BIC of 6433. TheR2 score is about 75.6% and the RMSE is about 46,8931
- This is a better performing ARIMA model than the baseline model.

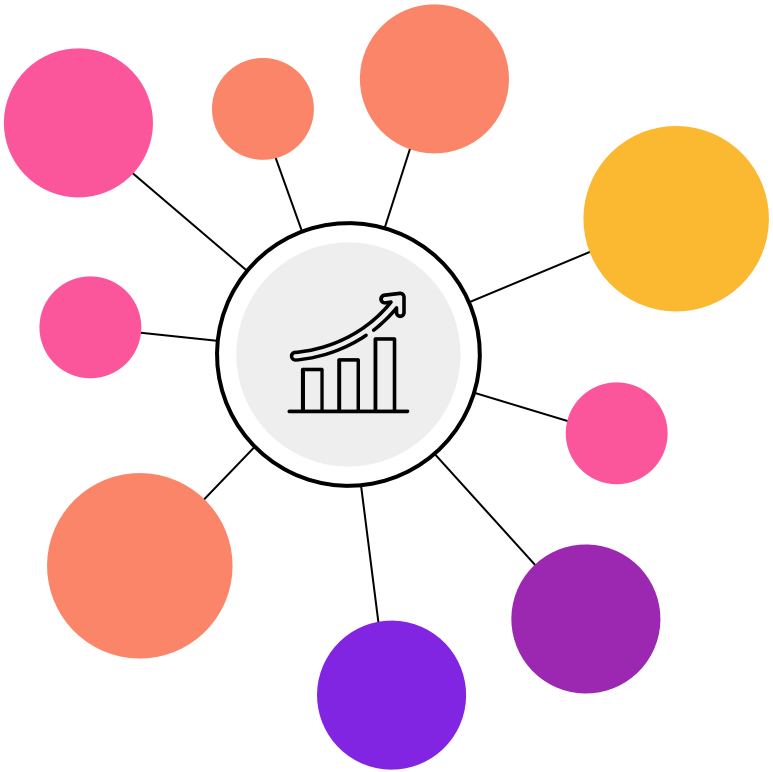

Two-Year Forecast For Real Estate Prices For NY_10128 Zipcode using FB Prophet Model
Upper Bound
Lower Bound
Forecast
Smoothed Actual
300k
200k
100k
0
−100k
−200k
−300k
Value of Property
2000
2005
2010
2015
2020
Date

Findings

- The three models have been evaluated for the Zip code 10128 in New York using RMSE as the metric of evaluation. The resultant RMSEs are:

    - 48,866(baseline ARIMA Model of orde 1,0,1),
    - 46,931 (Second Arima Model of order 2,1,2)
    - 16,821 (the final model usingFB prophet)

- From the foregoing results, the **FB prophet model** performs best in predicting future real estate prices of the Zip code 10128 in New York.



Recommendations

Recommendations

We would advise the client, MDN real estate investment firm to focus on the top 10 highest value states and top 10 zip codes in assessing the areas with a higher future value of property.

We would further advise the client to mitigate against risk, by not investing in one area only. For example, the top 4 zip codes with highest value of property are all in New York. We would advise spreading of risk by looking into other zip codes and states within the top 10.

We also advise MDN consultants to make use of the FB Prophet model in predicting the future value of property as it performed best. Nonetheless, each different zip code should be assessed independently on the three models.

More data is needed to improve model performance.

Other research may also be needed eg PESTEL analysis to complement the model findings.

Next steps

We propose **further work** in creating and running predictive models for all the top 10 zip codes to compare and assess which zipcodes are likely to have best performance in the future based on the predictive models.

In the future, apart from using value of property for the Time Series modelling, we can also incorporate into the models, **other factors** such as proximity to facilities like schools and hospitals; social & economic factors such as pandemics; crime rate of the area; development rate of the area just to mention a few.

Deployment of the best performing model could also be done, to aid investors in forecasting and visualising the future of real estate for particular zipcodes of interest.

Thank you

For more information kind contact us on:

muthoni.kahuko@student.moringaschool.com

mwatu.ndanu@student.moringaschool.com

denis.ochieng@student.moringaschool.com