Charlotte Housing Market: Time Series Analysis & Forecasting

Presented by Marvin Mills

March 8th, 2021

Agenda

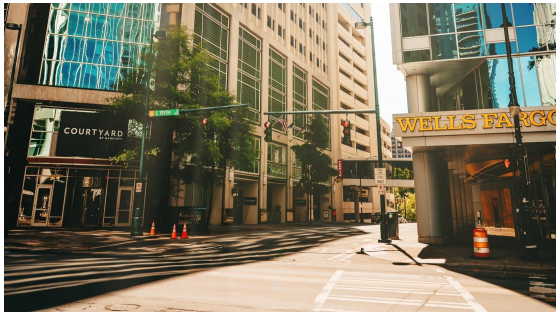
- Business Context
- About The Data
- Analysis + Modeling
- Recommendation
- Conclusion

Business Context

"Where should we build-to-rent 5-10 properties?"

Real estate investment and asset management firm specializing in **Build-to-Rent** real estate assets

Charlotte in 2020:
28% shrink in inventory
1.0 months of supply



The best zip code(s) for a **5-10 year investment** in Charlotte or surrounding

About The Data

Zillow Dataset

14,723 zip codes
FOCUS: Charlotte, NC and surrounding areas
SOURCE: https://www.zillow.com/research/data/

Created Time Series

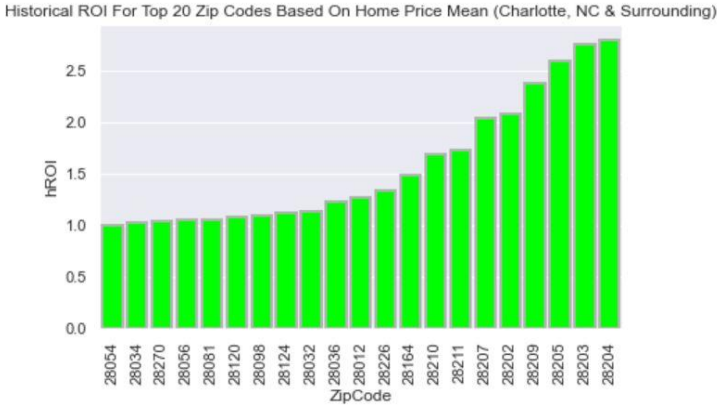
Two Periods:
1996-2018
2010-2018

Culled by County

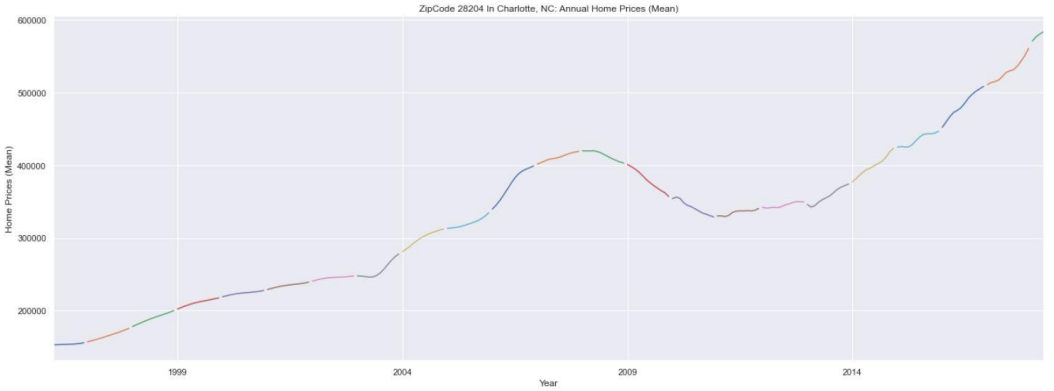
Mecklenburg, Cabarrus and Gaston counties

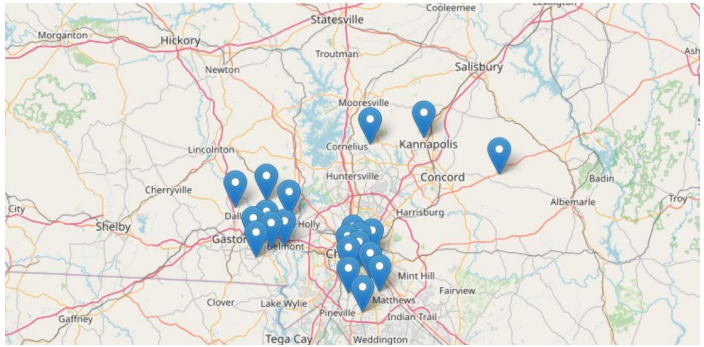
Feature Engineering

Created column for **Historical ROI**
Focused on top 21 zip codes initially



Historical ROI + Map (Top 21 Zip Codes)



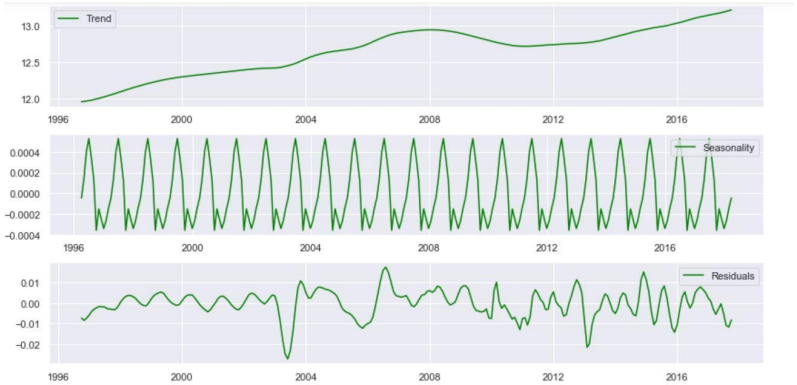


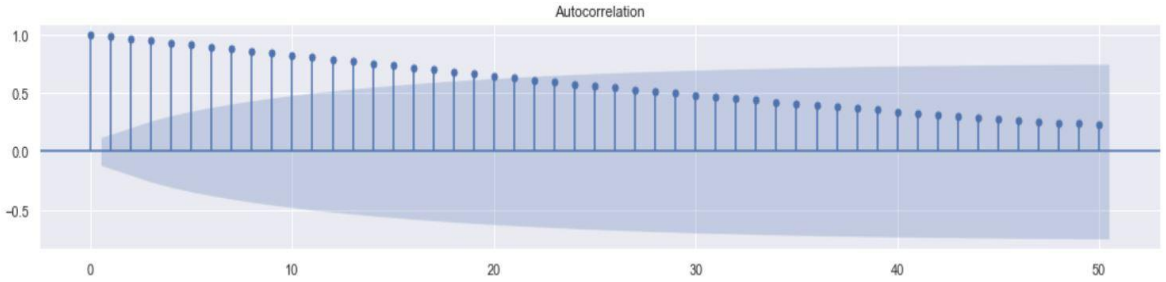
Initial Focus - 28204

Stationarity Challenges

Seasonal Decompose Residuals reveal Stationarity, **ultimately**.

Evident seasonality





28204 - 1 Difference

Covariance,
Homoscedasticity, etc.

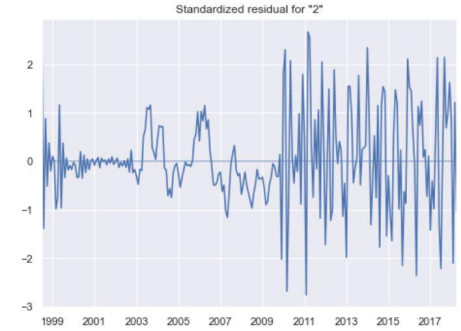
Analysis Results: High order of differencing is likely needed

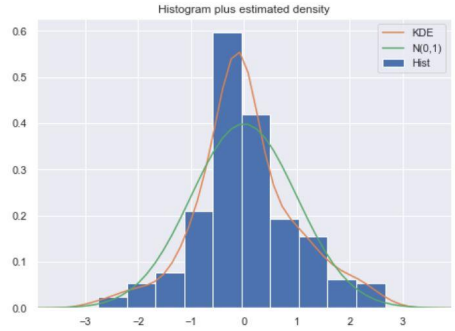
Many positive
correlations beyond 10

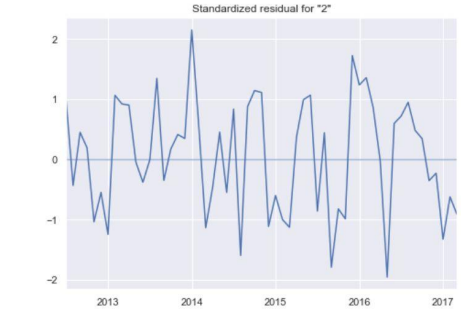
Modeling - SARIMAX

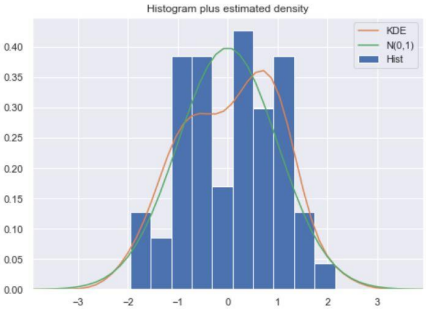
Focus: Lowest AIC Score

1996-2018









28205 - Charlotte

```
In [172]: 1 comparing_series(series_28205_shorter, series_28205_full)

FOR SERIES -- 2010- (end)2017
**********
The winner is: an AIC of 866.6166810120076 with (1, 1, 1) for order terms and (0, 1, 1, 12) for seasonal parameters
----------
```

```
In [173]: 1 comparing_series(series_28012_shorter, series_28012_full)

FOR SERIES -- 2010- (end)2017
**********
The winner is: an AIC of 873.1922582342307 with (1, 1, 1) for order terms and (0, 1, 1, 12) for seasonal parameters
----------
```

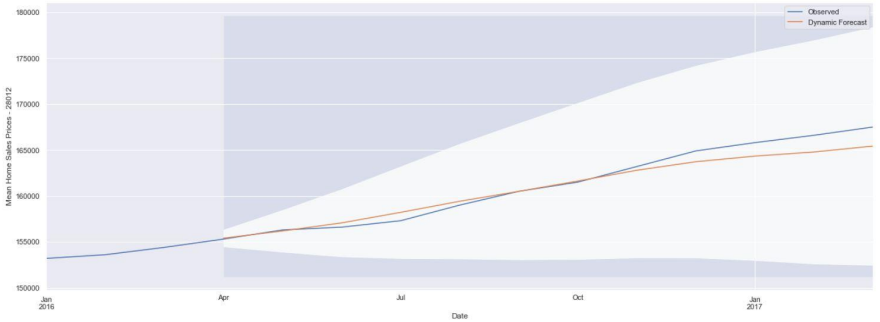
28012 - Belmont

```
In [171]: 1 comparing_series(series_28203_shorter, series_28203_full)

FOR SERIES -- 2010- (end)2017
**********
The winner is: an AIC of 1028.3520778139166 with (1, 1, 1) for order terms and (1, 1, 1, 12) for seasonal parameters
----------
                                 SARIMAX Results
==========================================================================================
Dep. Variable:                          28203   No. Observations:                   84
Model:             SARIMAX(1, 1, 1)x(1, 1, 1, 12)   Log Likelihood                -509.176
Date:                            Sat, 06 Mar 2021   AIC                           1028.352
Time:                                   11:17:22   BIC                           1038.567
Sample:                                 04-01-2010   HQIC                          1032.322
                                      - 03-01-2017
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6932      0.166      4.186      0.000       0.369       1.018
ma.L1          0.1415      0.071      2.006      0.045       0.003       0.280
ar.S.L12      -0.5893      0.178     -3.306      0.001      -0.939      -0.240
```

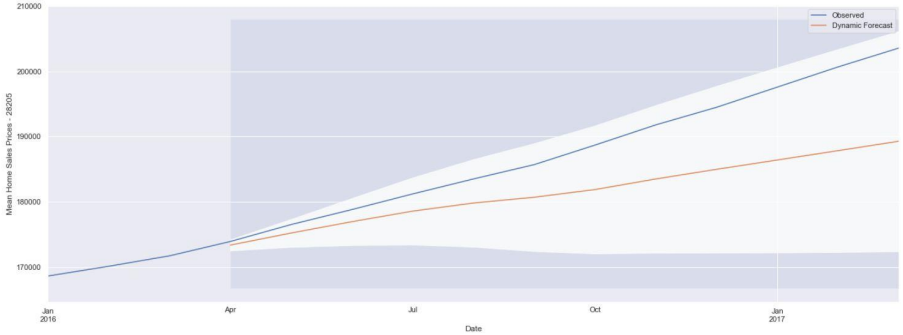
2010-2018

In [205]:
```
# Extract the predicted and true values of the time series - 28012
forecasted_vals = dynamic_pred.predicted_mean
true_vals = series_28012_shorter['2016-04-01':]

# Determining mean square error
mse = ((forecasted_vals - true_vals) ** 2).mean()
print('The Mean Squared Error is {}'.format(round(mse, 2)))
print('The Root Mean Squared Error is {}'.format(round(np.sqrt(mse), 2)))
```

```
The Mean Squared Error is 1047714.78
The Root Mean Squared Error is 1023.58
```



28012 vs. 28205

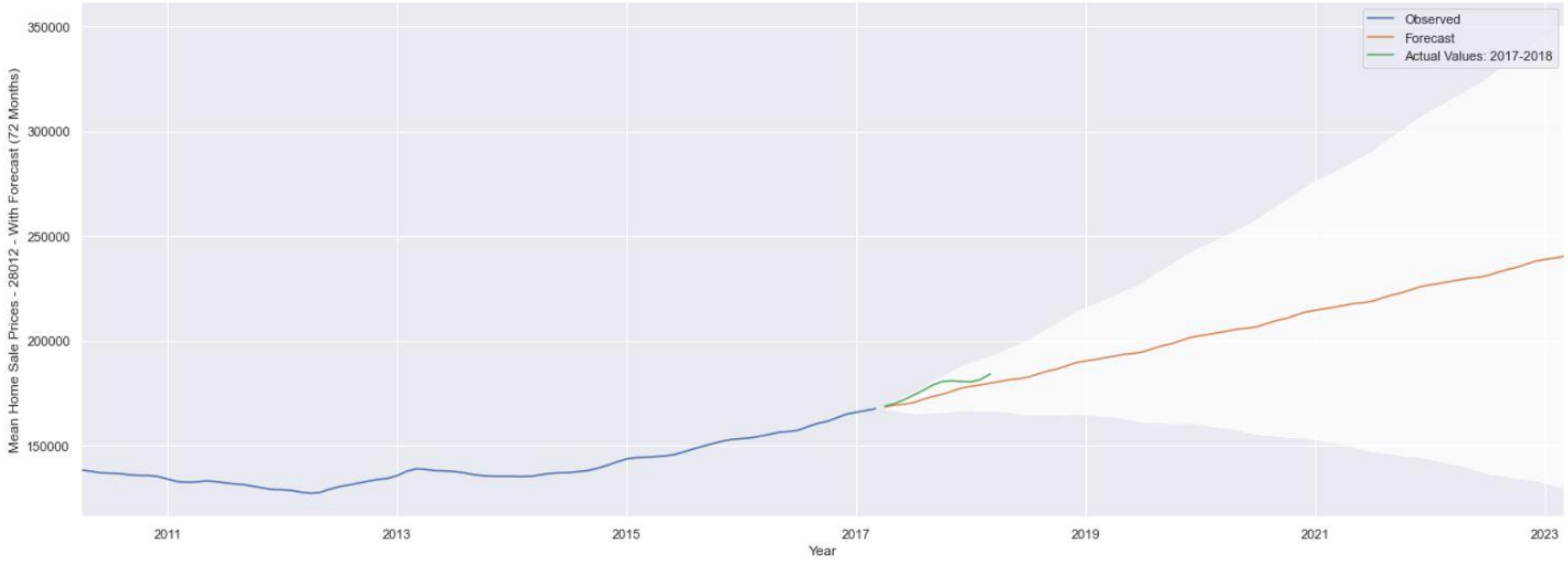


Best Models

In [187]:
```
# Extract the predicted and true values of the time series - 28205
forecasted_vals = dynamic_pred.predicted_mean
true_vals = series_28205_shorter['2016-04-01':]

# Determining mean square error
mse = ((forecasted_vals - true_vals) ** 2).mean()
print('The Mean Squared Error is {}'.format(round(mse, 2)))
print('The Root Mean Squared Error is {}'.format(round(np.sqrt(mse), 2)))
```

```
The Mean Squared Error is 62703869.64
The Root Mean Squared Error is 7918.58
```

Winner: Belmont, NC
28012



Future Recommendations + Conclusion

Thank you!

github.com/emel333 ~ for the repository

marvinlee_3@outlook.com ~ for any questions

Future Steps:

- Incorporate 2019-2021 housing data (vitally important) with a focus on the top 21 hROI zip codes
- Consider migration trends for Charlotte and surrounding areas
- Productionize model that analyzes ROI for individual home given the zip code, purchase price and purchase date
- Investigate renter's market in Charlotte and surrounding
- Investigate the cost of land itself in Charlotte and surrounding areas

Credits

Shapes & Icons

Vectorial Shapes in this Template were created by **Free Google Slides Templates** and downloaded from **pexels.com** and **unsplash.com**.

Icons in this Template are part of Google® Material Icons and **1001freedownloads.com**.

Backgrounds

The backgrounds were created by **Free Google Slides Templates**.

Images

Photos in this template were downloaded from **pixabay.com**. Attribution is located in each slide notes and the Credits slide.

Fonts

The fonts used in this template are taken from **Google** fonts. (Muli)
You can download the fonts from the following url: https://www.google.com/fonts/

Color Palette

The Template provides a theme with four basic colors:

#4e6e9aff	#4e6e9acc
#5477a7cc	#eeeeeeff

Trademarks

Microsoft® and PowerPoint® are trademarks or registered trademarks of Microsoft Corporation.

Google Drive® is a registered trademark of Google Inc.