REAL ESTATE ANALYSIS:

Data:

Real Estate Sales data from 2003 to 2014 were analyzed. Only residential and mixed use buildings have been accounted for in this analysis.

Date set was cleaned to remove Null Values, all numbers that were in string format in the csv files were converted to integers. Outliers were removed by filtering out any sale price less than 10,000 or more than 3 standard deviation of the mean price. The sales data were aggregated at a monthly level.

Analysis:

We performed regression analysis on median monthly sales data of Chelsea neighborhood vs. rest of Manhattan for two time periods: before the opening of Highline in 2003 to June 2009 and after June 2009 till December 2016 (See Plot 1). Before Highline was opened the slopes of the regression line for the monthly aggregated median sale prices in Chelsea and Manhattan had the values of 10.69 and 10.44, respectively. This indicated that prices were increasing at roughly the same rate. But after the Highline was built the rate of the change of the real estate prices in Chelsea have increased, which is indicated by the higher slope of 12.66 whereas for Manhattan the pace became slower, the slope is 2.29. The pre-Highline median price for Chelsea ($814,000) was $23,000 lower than that for Manhattan ($837,000). But the post-Highline median price in Chelsea ($1,384,000) was $208,000 more compared to Manhattan ($1,176,000).

We performed similar regression analysis on mean monthly sales data of Chelsea neighborhood vs the rest of Manhattan (See Plot 2). Before Highline was opened the mean monthly sale price in both Chelsea and Manhattan were increasing at the same pace, with the slopes of 10.51 and 10.6 respectively. But after Highline was built the rate of change of the mean real estate price in Chelsea increased 2.7 times while for Manhattan it only increased 1.5 times compared to the pre-Highline slopes.

The pre-Highline mean price for Chelsea ($1,748,000) was $220,000 higher than that of Manhattan ($1,528,000). The post-Highline price for Chelsea ($2,558,000) was $240,000 higher than that of Manhattan ($2,318,000). Although the latter change is not so much higher than the prior one, we do see that the value of the mean price in Chelsea has increased by $810,000 indicating that the number of luxury building sales increased in Chelsea neighborhood after Highline had been opened compared to the prior time. Also the higher post-Highline mean prices for both Chelsea and Manhattan indicate a significant growth in luxury apartment sales.

It shall be noted that the post-Highline sales regression lines have been shifted for better visualization. The actual lines (dotted lines) for both mean and median price comparison were shifted down compared to the pre-Highline sales regression lines. This is because the opening of Highline coincided with the time of the great recession. The housing prices were impacted by the recession and the sale prices just started picking up the trend in the beginning of post-Highline period.

In order to minimize the effect of recession on the sale prices, we decomposed the sale price time series

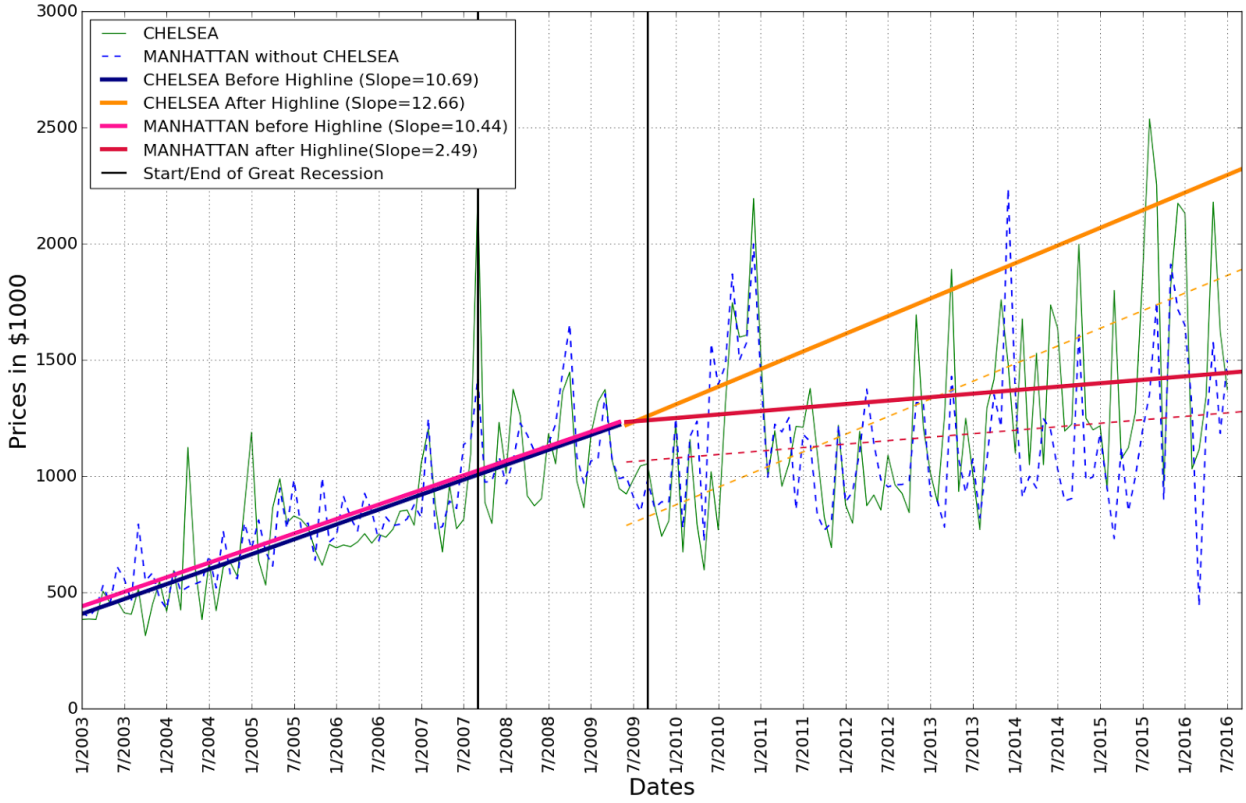
corresponding to the period before recession (before July 2007) using fast Fourier transform and extrapolated the result to predict the monthly sales pattern until December 2016. We then compared our extrapolation with the actual sale prices.

We used Fourier extrapolation as it is much more accurate at exploration of time series for longer periods of time compared to other standard methods like ARIMA, neural networks or Markov chains because they add too much noise over time. But overfitting the number of harmonics for Fourier transformation generates decent extrapolation even when the extrapolation period is of comparable length with the training period. Other techniques would start significantly deviate their predictions from the actual trend having the predicted period of a similar length with the training period.

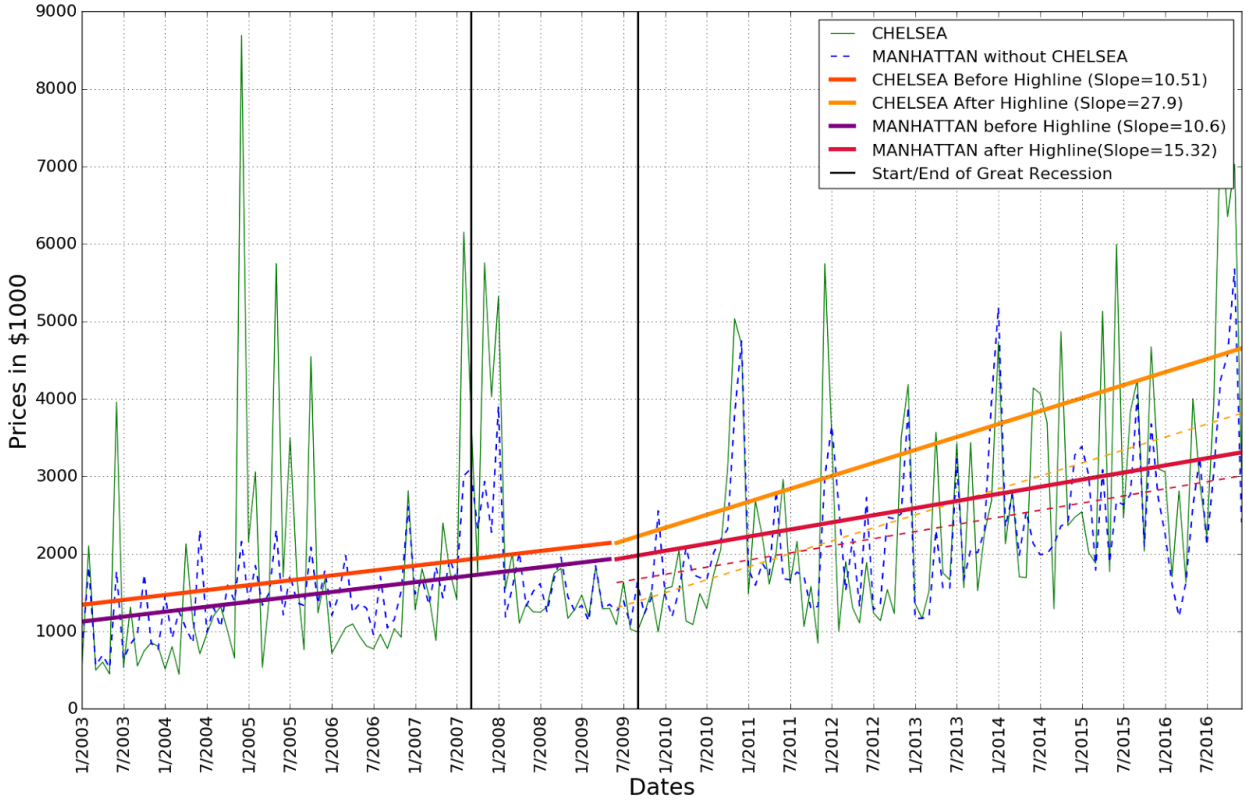
From our four models (table 1) we observed that for all the cases except one our models predicted the sale prices to be higher than the actual price as suggested by the negative slope of residual regression lines. Only for the mean sales in Chelsea neighborhood the actual prices are higher than the ones predicted by the model. This yet again supports our previous assumption that during the post-Highline opening period the number of luxury building sales has increased.

SHORTCOMING OF THE ANALYSIS:

The analysis was performed on the total sale prices which treats all buildings as if they had similar attributes. In order to account for this we initially wanted to use sale prices normalized by area but about 80% of the data had missing building area values removing which would result in a poor analysis. We also tried performing analysis by normalizing with the number of units in the building but that also resulted in a lot of data loss.

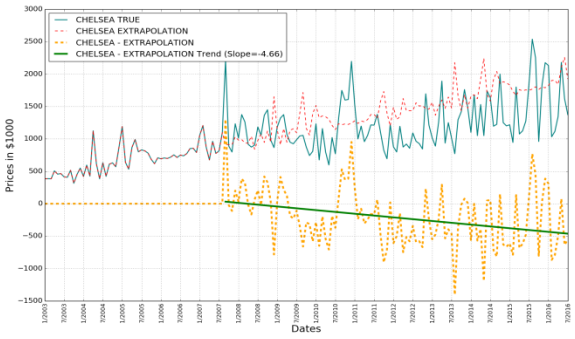
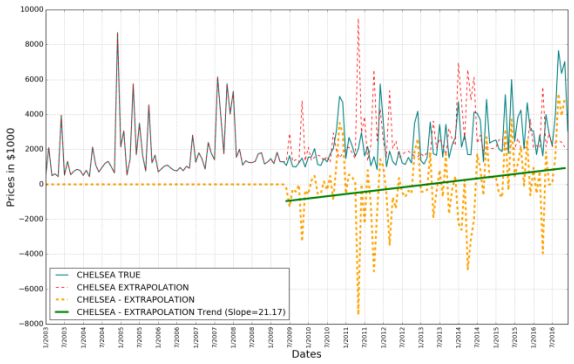
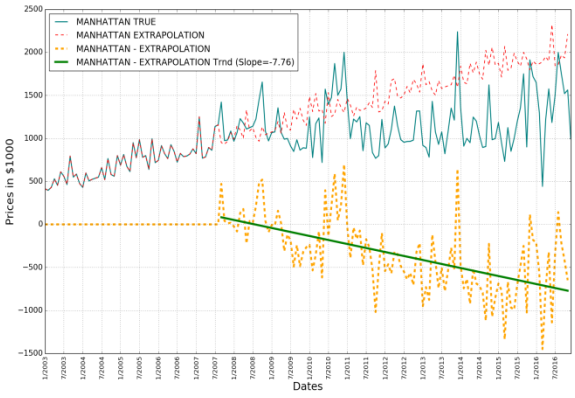
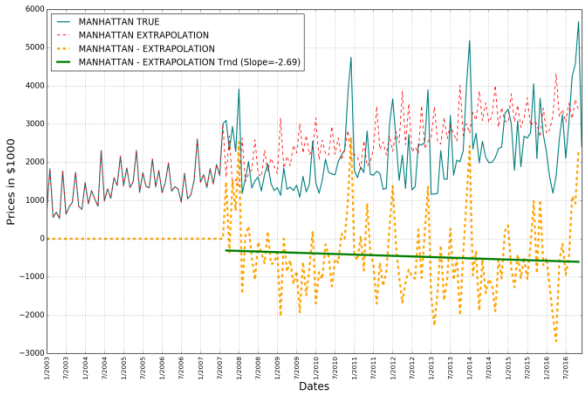

CHELSEA
MANHATTAN without CHELSEA
CHELSEA Before Highline (Slope=10.69)
CHELSEA After Highline (Slope=12.66)
MANHATTAN before Highline (Slope=10.44)
MANHATTAN after Highline(Slope=2.49)
Start/End of Great Recession
Prices in $1000
Dates
0
500
1000
1500
2000
2500
3000
1/2003
7/2003
1/2004
7/2004
1/2005
7/2005
1/2006
7/2006
1/2007
7/2007
1/2008
7/2008
1/2009
7/2009
1/2010
7/2010
1/2011
7/2011
1/2012
7/2012
1/2013
7/2013
1/2014
7/2014
1/2015
7/2015
1/2016
7/2016

PLOT 1.


CHELSEA
MANHATTAN without CHELSEA
CHELSEA Before Highline (Slope=10.51)
CHELSEA After Highline (Slope=27.9)
MANHATTAN before Highline (Slope=10.6)
MANHATTAN after Highline(Slope=15.32)
Start/End of Great Recession
Prices in $1000
Dates
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
1/2003
7/2003
1/2004
7/2004
1/2005
7/2005
1/2006
7/2006
1/2007
7/2007
1/2008
7/2008
1/2009
7/2009
1/2010
7/2010
1/2011
7/2011
1/2012
7/2012
1/2013
7/2013
1/2014
7/2014
1/2015
7/2015
1/2016
7/2016

PLOT 2

TABLE 1.

	MEDIAN SALES (Model vs. Actual)	MEAN SALES (Model vs. Actual)
CHELSEA		
	SLOPE OF RESIDUAL : -4.66	SLOPE OF RESIDUAL : 4.30
MANHATTAN		
	SLOPE OF RESIDUAL : -7.76	SLOPE OF RESIDUAL : -2.63