By: Vicky Clayton, Kate Youngman, Maragatham Natarajan, Charissa Rentier

Which are the upcoming neighborhoods for renting or buying in New York?

- VKMC Consulting -

Problem Statement

At the end of every year, real-estate company StreetEasy puts together a list of ten NYC neighborhoods to watch in the coming year, based on changes in median sales and rent prices, number of listing views on StreetEasy and certificates of occupancy filed (indicating completion of new construction). For 2018, West Harlem topped the list, with Fort Greene, Flushing, Elmhurst, East New York, Norwood, Prospect Park South, Chinatown, Brighton Beach, and Woodside as runners up. They asked us to choose which of these selected neighbourhoods was most “hot” using a data-driven approach, and which they should most recommend to their potential buyers.

Approach

We reasoned that neighborhoods that are considered currently great places to go out in the evening may be early indications of good places to live in a few years' time. Based on this reasoning, we used open data from the Municipal Transport Authority of New York on the number of exits from turnstiles in New York subway stations as a proxy for the popularity of the neighborhoods. The turnstiles count the cumulative number of people who have going through each turnstile and record the count at particular snapshots in time. This allows us to derive a rate of number of people exiting the turnstile per hour so that we get an indication of the neighborhoods with the highest rate of exits in the evening.

We selected the turnstiles at stations within the neighborhoods StreetEasy identified and took data from April 2015 - December 2017. We then focused only on the evenings (5pm - Midnight) so that we'd capture people going out after work, or heading home (also an indication of it being a good place to live, though one which is already recognised!) We conducted our data cleaning by turnstile so that we'd be able to identify problematic recordings and deal with them at the most granular level but we then aggregated up the turnstiles to analyse the average exits per hour over the course of the evening for each neighborhood. We chose exits per hour as the gap in between recordings varies. We mapped the 40 stations to the ten neighborhoods using this list from NY Mag.

Preliminary Results

We investigated whether the rate of exits ("the footfall") had increased in these neighborhoods from 2015 to 2017. We were interested in the trend of footfall rather than the absolute numbers to see whether neighborhoods are changing in popularity. We hypothesise that neighborhoods with already high absolute numbers would have such popularity already incorporated into market rates for housing.

April Results

For preliminary data exploration, we chose a subset of the data. We suspected that the changing weather conditions with the season may have differential effects on subway use and so we also investigated the rate of exits for these 10 neighbourhoods in 2015, 2016 and 2017 focused exclusively on April (which doesn't have heavy snow or muggy heat) to avoid the influence of seasonality (Table 1). This showed a surprising amount of variability across the neighborhoods, and we decided to investigate using the whole dataset of recordings.

Table 1: Change in rate of evening exits for each neighbourhood between 2015 and 2017 (April only)

	NEIGHBORHOOD	4/2015_EXITS	4/2016_EXITS	4/2017_EXITS	April Delta %
0	Brighton Beach	1074.543907	1016.413916	965.019828	-10.19
1	Chinatown	1199.867340	1111.524294	1234.848795	2.92
2	East New York	5555.488717	5422.964794	5639.360316	1.51
3	Elmhurst	10440.525845	7309.265851	7333.101961	-29.76
4	Flushing	2151.234568	NaN	3943.922694	83.33
5	Fort Greene	4304.206490	4137.928692	650.022431	-84.90
6	Norwood	827.841365	743.113382	2676.251422	223.28
7	Prospect Park South	2708.161765	2642.391073	2905.911383	7.30
8	West Harlem	3098.856470	3085.665869	5506.148112	77.68
9	Woodside	5851.333857	5603.536415	NaN	NaN

Whole Dataset

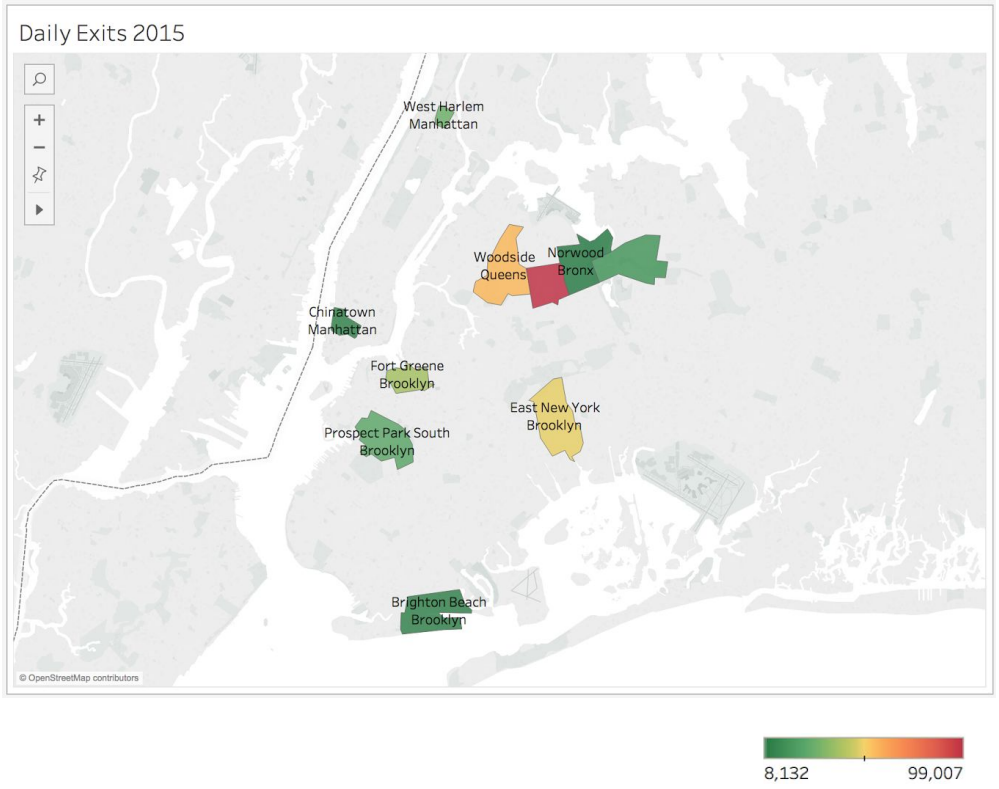
We take the three neighborhoods with the highest increase in the rate of exits as our “up and coming” neighborhoods particularly to watch from StreetEasy’s list (Table 2). These were: East New York, Woodside and Fort Greene. We saw the highest increase in rate of exits in East New York with 45.7% increase in evening subway exits in 2017 compared to 2015, and a decrease in the rate of exits in Elmhurst (-5.95%).

Table 2: Change in rate of evening exits for each neighbourhood between 2015 and 2017 (full year)

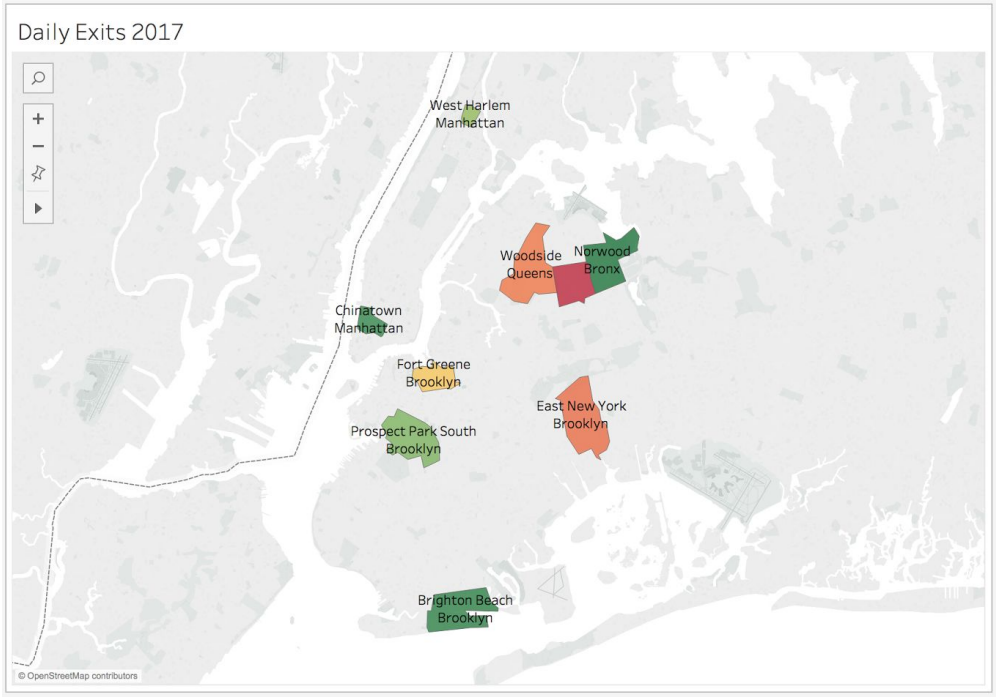
	NEIGHBORHOOD	YR15_EVENING_EXITS	YR17_EVENING_EXITS	EXIT_DELTA	DELTA DIV	DELTA %
2	East New York	47985.353284	69913.343597	21927.990313	0.456973	45.70
9	Woodside	53986.234393	67343.892336	13357.657943	0.247427	24.74
5	Fort Greene	38426.816177	49560.496834	11133.680656	0.289737	28.97
7	Prospect Park South	24904.371637	31876.496072	6972.124434	0.279956	28.00
8	West Harlem	28534.881660	35282.678985	6747.797325	0.236475	23.65
1	Chinatown	10839.837502	14669.107225	3829.269723	0.353259	35.33
0	Brighton Beach	10546.194480	12323.968579	1777.774099	0.168570	16.86
6	Norwood	7540.153603	8088.892026	548.738423	0.072775	7.28
3	Elmhurst	93249.239466	87702.020503	-5547.218962	-0.059488	-5.95
4	Flushing	18988.966533	NaN	NaN	NaN	NaN

Chinatown also jumps out with a 35.3% increase over the same period. Fort Greene and Prospect South Park are not far behind.

Map 1: Heat map of rate of exits in StreetEasy's 10 neighbourhoods (2015)



Map 2: Heat map of rate of exits in StreetEasy's 10 neighbourhoods (2017)

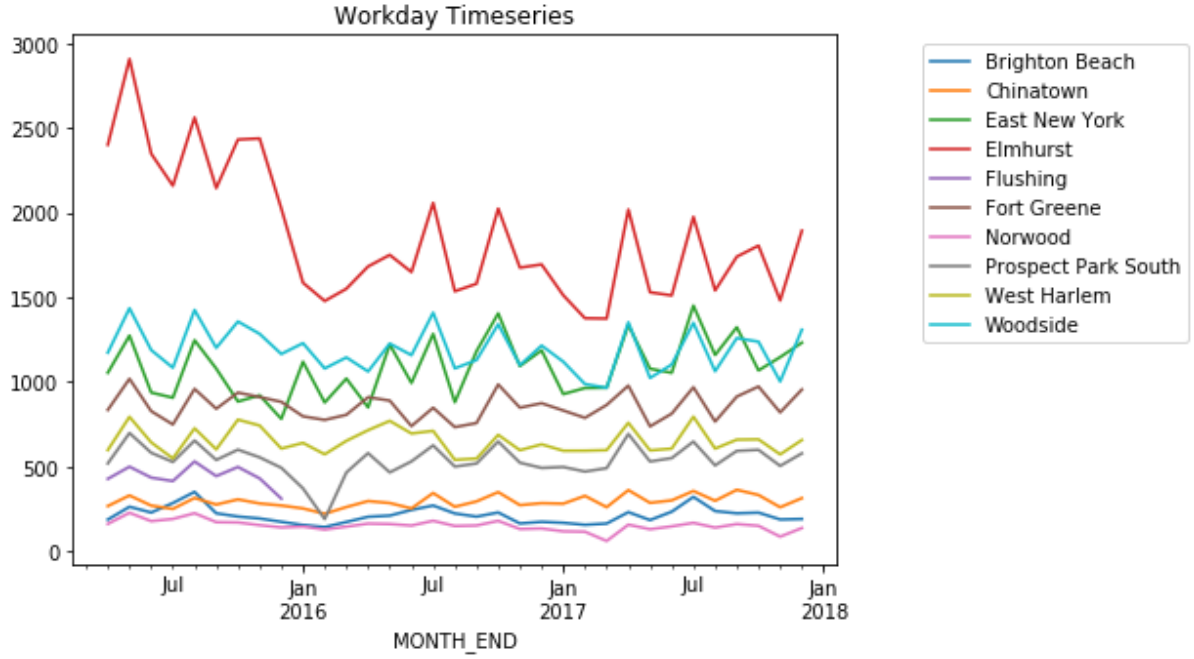


From the heat maps displayed above, we can see that in 2015 (Map 1), the neighborhood Elmhurst, which is bright red on the map, has considerable more foot traffic than the other neighborhoods in the sample, with only Woodside as an obvious runner up. However, looking at the map for 2017 (Map 2), we see that other neighborhoods have crept up on Elmhurst in terms of ranking based on subway exits. Over time, Woodside has gotten even closer than it was before, but now also East New York and Fort Greene have gotten higher up in the ranking. A note of caution to add on these heat maps is that we cannot compare the neigbhorhoods in the individual heat maps directly to each other over time: the color scales are based on the comparison between neighborhoods at that particular point in time, not over time. For a better overview of traffic coming into the neighborhoods by subway over time, we constructed time series which are displayed below.

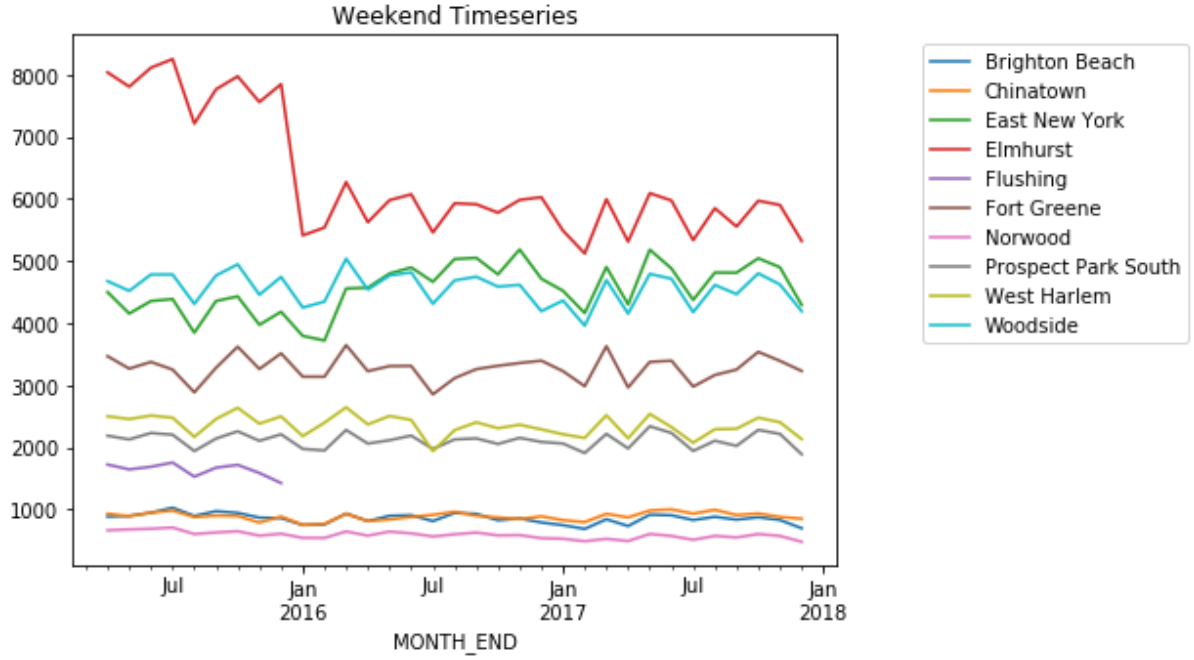
Weekends and Weekdays

We broke down the analysis by whether the evening was a weekday or a weekend. We can see the trends are similar with Elmhurst dropping significantly early 2016. It would be interesting to explore further the reason for this. Please note that data was not available for Flushing after January 2016. In terms of the overall trends (considering weekends and weekdays together), it is not clear from the graph the top 3 we've identified.

Graph 1: Number of exits per neighborhood from April 2016 to December 2017 (workday)

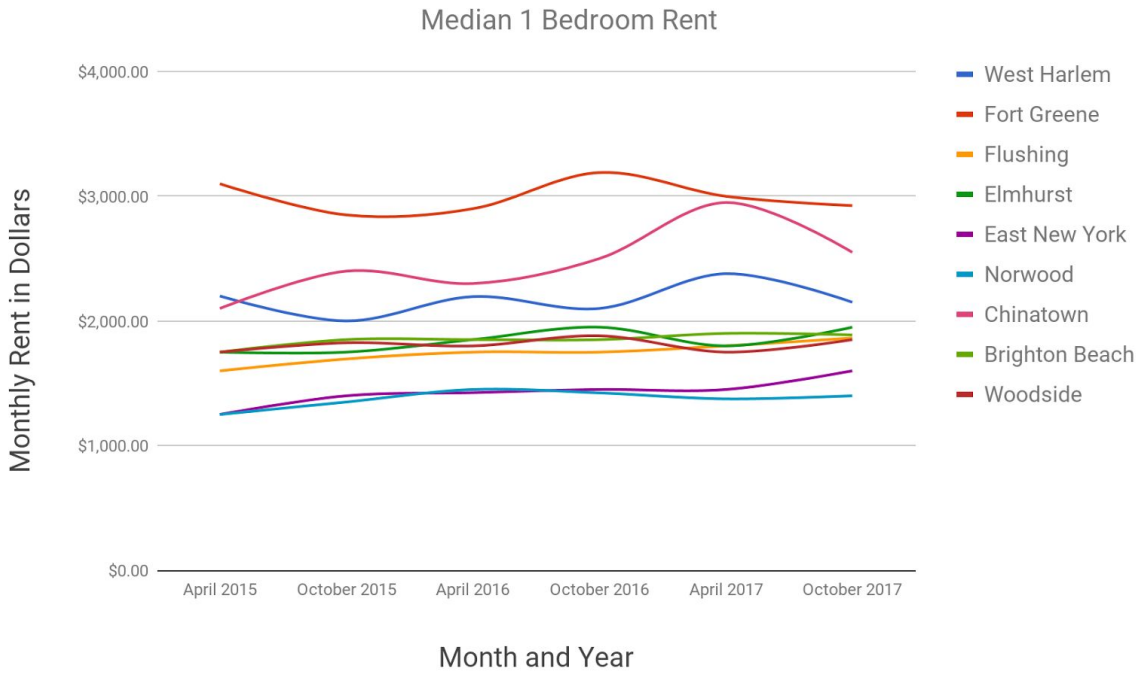


Graph 2: Number of exits per neighborhood from April 2016 to December 2017 (weekend)



We also compared rental price data (median price for one bedroom apartments) in each neighborhood for the same time period to see whether the trends in rates of subway exits is reflected in rental prices in subsequent months or years. The trend of increased footfall in East New York is also reflected in an upwards trend in rental prices (although we have not tested this formally and this says nothing of causality). It is interesting that the large drop in footfall we see in Elmhurst in Graphs 1 and 2 is not reflected in rental price. This is something we'd like to explore further. For future analysis, we could analyse this

relationship between footfall and rental prices more formally through regressing rental prices on lagging footfall and appropriate control variables.



(Data source: https://www.zumper.com)

What else we could do

With MTA data

Another area of interest would be to look at how neighborhoods close to the ones pinpointed by Streeteasy are developing. It could be that instead of moving to the rising stars identified, people might look for adjacent neighborhoods seeking more affordable housing.

With additional data

As mentioned, turnstile data does not tell us a lot about what actually goes on on the ground. As mentioned, we could conduct regression analysis to formally analyse the relationship between footfall as a proxy for neighborhood popularity and rental prices. We could analyse the correlation between the numbers of restaurants, bars and clubs in those areas to test our assumption of footfall being a proxy of “hipness” (if we agree that this makes a hip area!)

Caveats

Please keep in mind that the outcome represents an average number of exits per hour over the course of the evening and snapshots of footfall are unevenly spread throughout the

evening so commuter stations may feature higher in the results if the snapshot is taken around rush-hour. In the future, we would take a slightly more sophisticated approach where we could estimate the footfall for each hour in between the snapshots through interpolation.

Conclusion

Subject to our caveats and further analysis, we would recommend to StreetEasy's buyers and renters to focus on East New York, Woodside and Fort Green in their search for an apartment, and to avoid Elmhurst.

Appendix

Data Cleaning

We dropped recovery audits which allowed us to get rid of duplicate entries, and dropped recordings with timestamps with non-zero seconds. Our reasoning on this was that it would be unusual to schedule an audit for such a specific time other than on the hour, and so such off-hour. According to our diagnostics, there were some circumstances where the counters recorded a lower number of exits than the previous recording. Where the numbers of in line with expectation but negative, we assumed that the counter had clicked into reverse, and took the modulus of these counts. Occasionally, the number of exits per recording was much larger than expected (in the order of hundreds of thousands or millions, and we reasoned that the counter may have been reset, and so set numbers of exits greater than a rate of 2 per second.