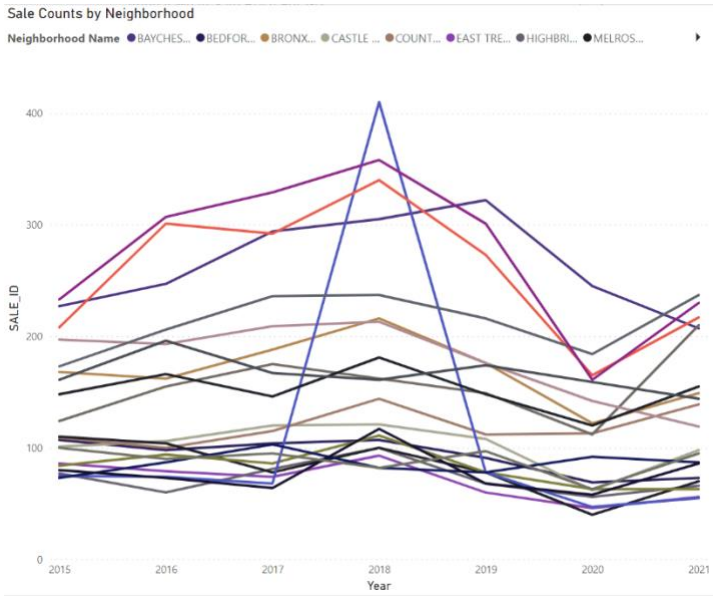
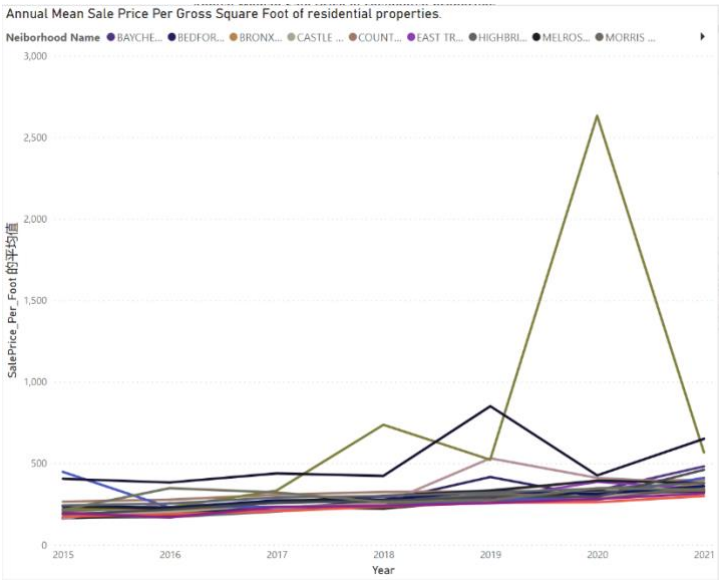
NYC Real Estate Analytics Project

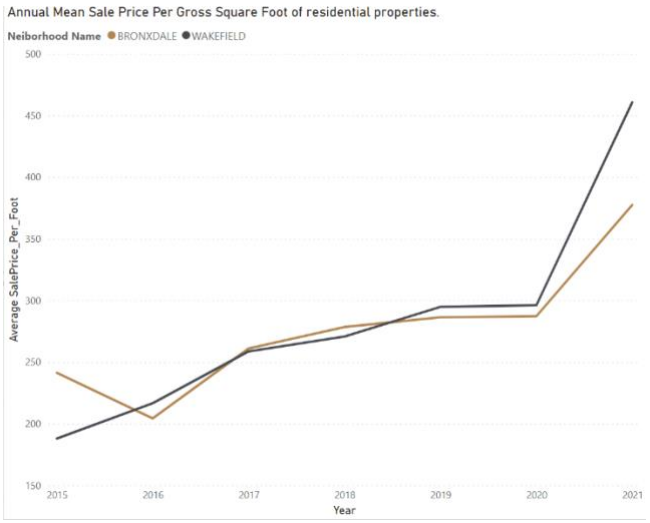
Introduction

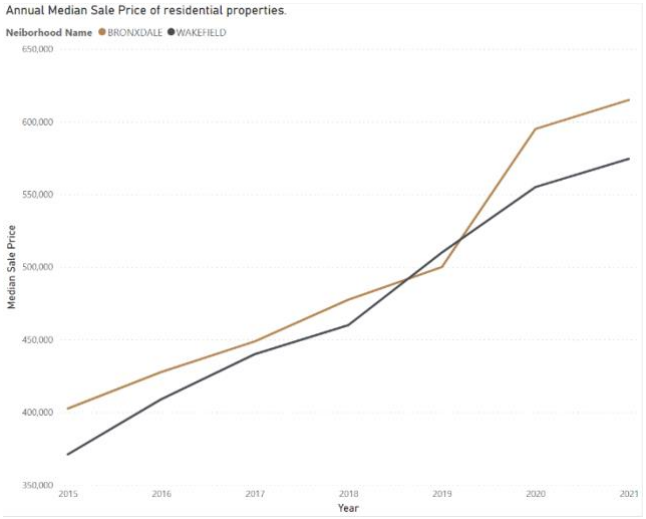
Real estate analytics is a multidisciplinary field that applies data analysis, statistical techniques, and machine learning to gain insights and make informed decisions in the real estate industry. In the past, real estate analytics focused on traditional metrics. It has gained significant traction in recent years due to advancements in data collection, technology, and a growing recognition of the importance of data-driven decision-making in real estate. Real estate has long been a data-driven industry, with professionals relying on market data, property assessments, and financial models for decision-making. Data analytics should have its own strategic direction with long-term roles and goals beyond just a few pilot projects and use cases. Researchers have developed predictive models for property prices, rental rates, and market trends. These models use historical data and factors such as location, economic indicators, and demographics to forecast future property values. Machine learning algorithms, including regression models, decision trees, and neural networks, are increasingly used to automate property valuation and predict market dynamics. Natural language processing (NLP) is employed to analyze sentiment in property listings and social media data. Advanced data visualization techniques help in presenting complex real estate data in an understandable and actionable format. Dashboards and interactive maps enhance decision-making. The integration of data analytics and artificial intelligence continues to reshape the real estate industry by providing more accurate insights and enhancing decision-making processes.

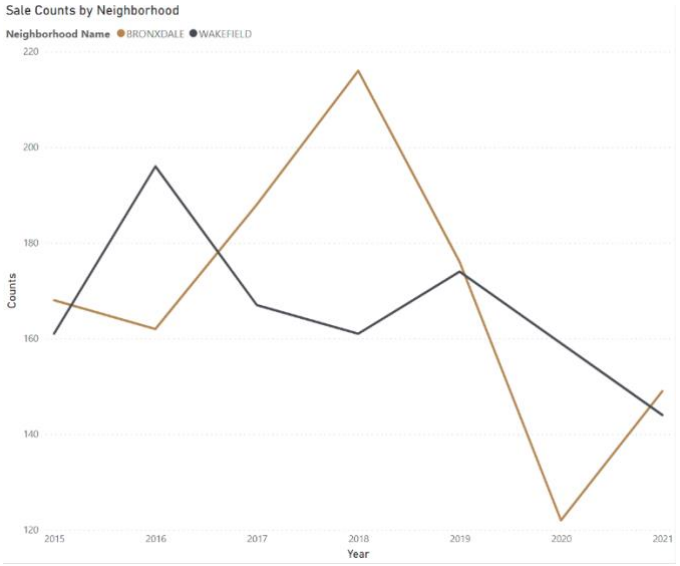
First, I chose to filter out areas where the number of house sales was higher than 500 in the past few years. So I chose three different charts, Sales counts, Median and Average. I'm interested in these two areas, Bronxdale and Wakefield because their median and average sales prices per gross square foot are generally relatively high and stable. And then there's also the middle-of-the-road segment in terms of sales numbers.

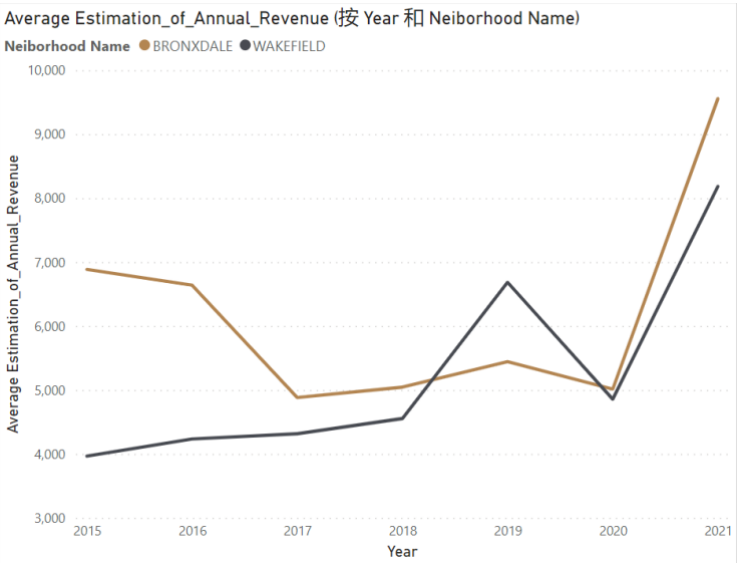


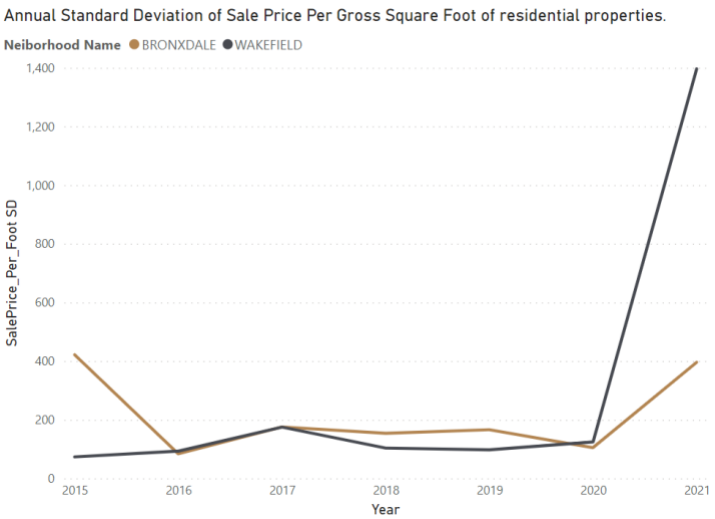



Annual Mean Sale Price Per Gross Square Foot of residential properties.
Neiborhood Name BRONXDALE WAKEFIELD
Average SalePrice_Per_Foot
Year


Annual Median Sale Price of residential properties.
Neiborhood Name BRONXDALE WAKEFIELD
Median Sale Price
Year


Sale Counts by Neighborhood
Neighborhood Name BRONXDALE WAKEFIELD
Counts
Year


Average Estimation_of_Annual_Revenue (按 Year 和 Neiborhood Name)
Neiborhood Name BRONXDALE WAKEFIELD
Average Estimation_of_Annual_Revenue
Year


Annual Standard Deviation of Sale Price Per Gross Square Foot of residential properties.
Neiborhood Name BRONXDALE WAKEFIELD
SalePrice_Per_Foot SD
Year


Annual Standard Deviation of Sales Prices of residential properties.
Neiborhood Name BRONXDALE WAKEFIELD
SALE_PRICE SD
Year

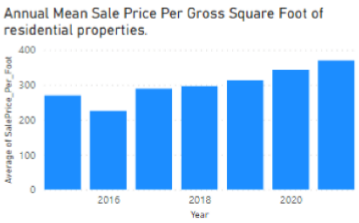
It is obvious that Estimation of Annual Revenue is very likely to have an upward trend year by year in the future. But the annual standard deviations of sale price per gross square foot are getting higher which means sale prices vary widely across regions. This has advantages and disadvantages. The advantage is that there is cultural diversity, but the disadvantage is that areas with too low prices may have higher crime rates due to safety issues.

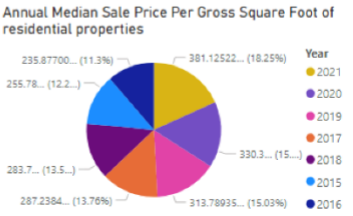
KPI analysis

PELHAM PARKWAY NORTH

The chart above shows the six KPIs for PELHAM PARKWAY NORTH separately, i ncluding Annual Sales Count of residential properties;
Annual Median Sale Price of residential properties;
Annual Standard Deviation of Sales Prices of residential properties;
Annual Mean Sale Price Per Gross Square Foot of residential properties;
Annual Standard Deviation of Sale Price Per Gross Square Foot of residential properti es;
Annual Median Sale Price Per Gross Square Foot of residential properties. Based on t he chart above, we can easily get that the price per feet in this area is increasing year b y year. House sales in the area peaked in 2017.

PELHAM PARKWAY SOUTH

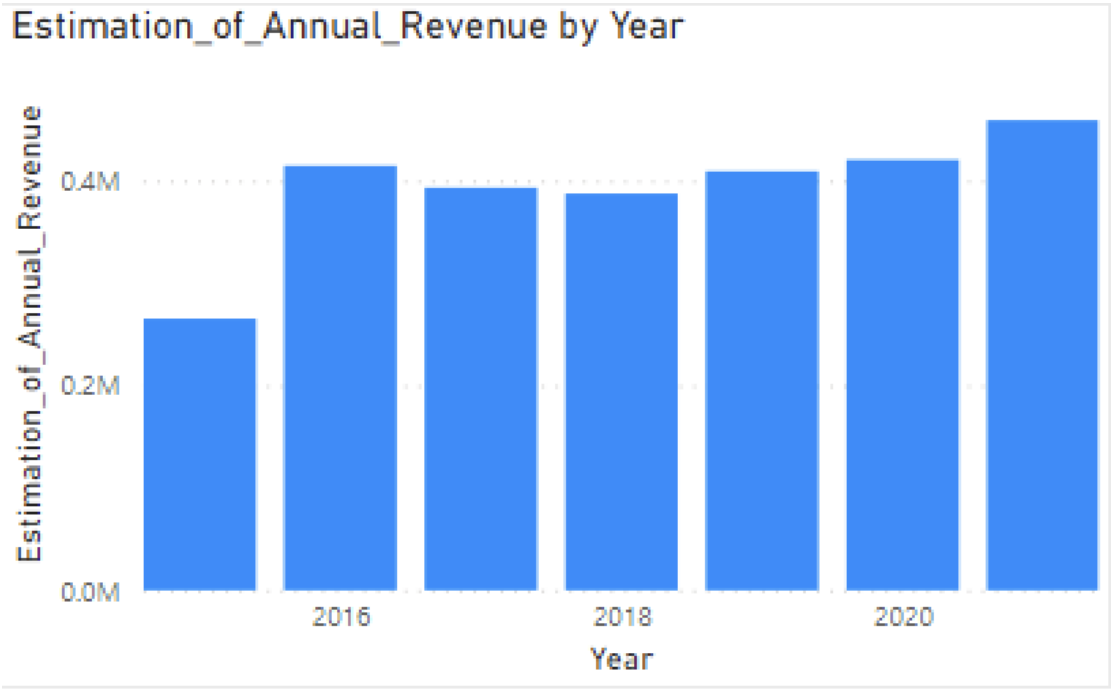




The chart above shows the six KPIs for PELHAM PARKWAY NORTH separately, i ncluding Annual Sales Count of residential properties;
Annual Median Sale Price of residential properties;
Annual Standard Deviation of Sales Prices of residential properties;
Annual Mean Sale Price Per Gross Square Foot of residential properties;
Annual Standard Deviation of Sale Price Per Gross Square Foot of residential properti es;
Annual Median Sale Price Per Gross Square Foot of residential properties. Based on t he chart above, we can easily get that the price per feet in this area is increasing year b y year.
House sales in the area peaked in 2018; the standard deviation is largest in 2015 and g radually stabilizes thereafter

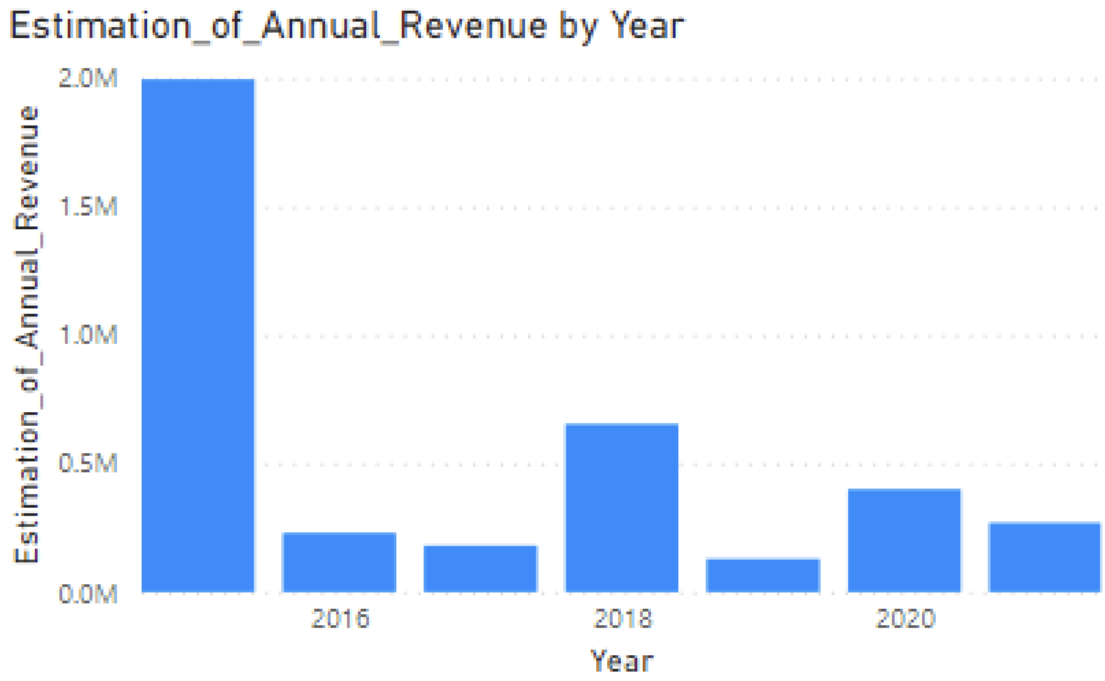
Estimation of Annual Revenue

PELHAM PARKWAY NORTH



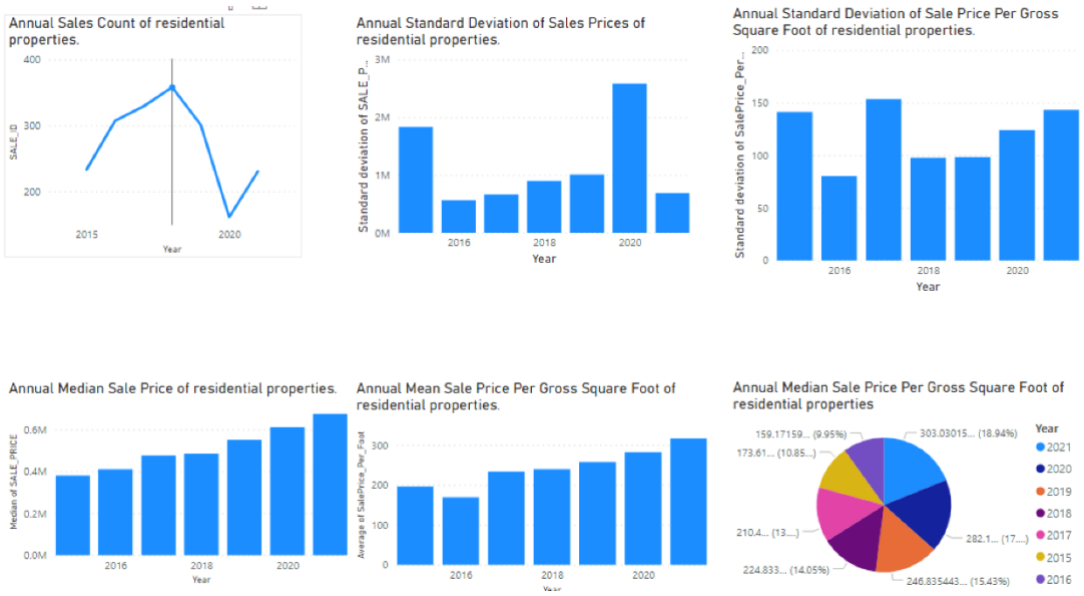
According to the graph above, the income from the sale of houses in the area has increased almost every year. That's why there's a good prospect for real estate business in the area.

PELHAM PARKWAY SOUTH

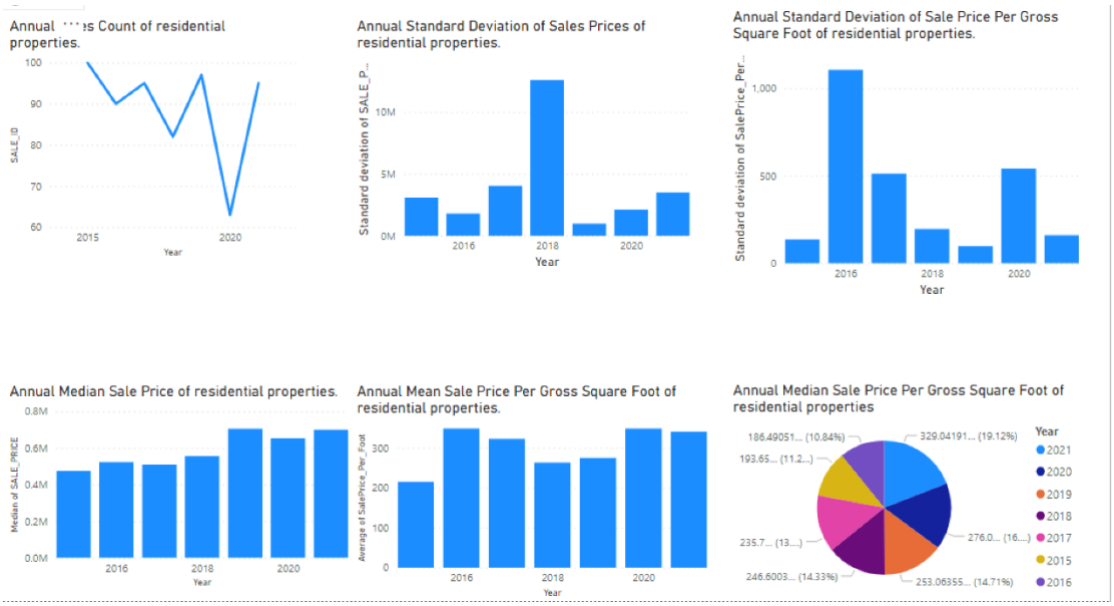


According to the above chart, the total income from housing transactions in the area is decreasing, which leads to the conclusion that the area is not suitable for housing transactions.

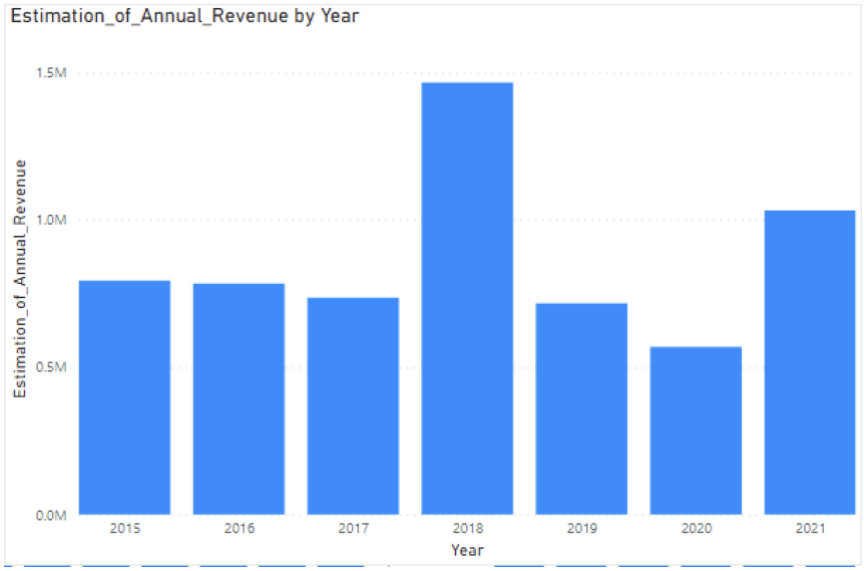
KPI Analysis (Westchester and Williamsbridge)



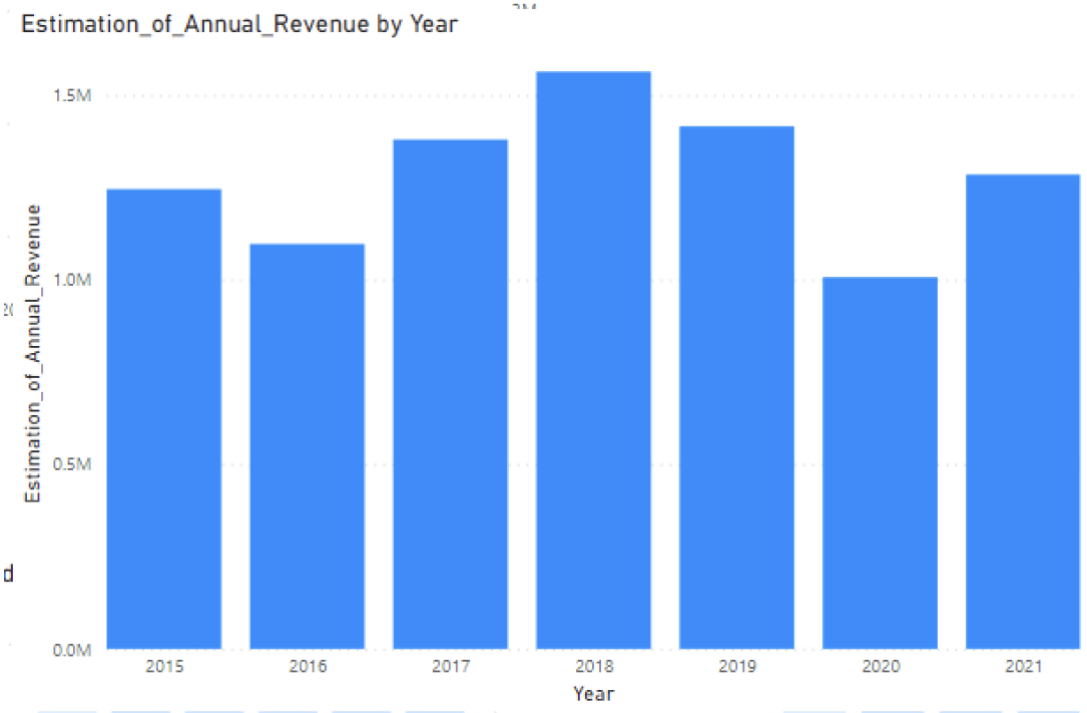
From charts above, we can see that house sales peaked in Westchester in 2018 and there was a huge decrease in 2020, probably because of the pandemic. Price per feet in this area grew steadily.



Williamsbridge experienced more changing intervals than Westchester. We can see a steady decline in house sales in this area.



Predicted annual incomes in Westchester are generally same except in 2018 and 2020. It rose to a high point and peaked in 2018, and continually fell in the following two years.



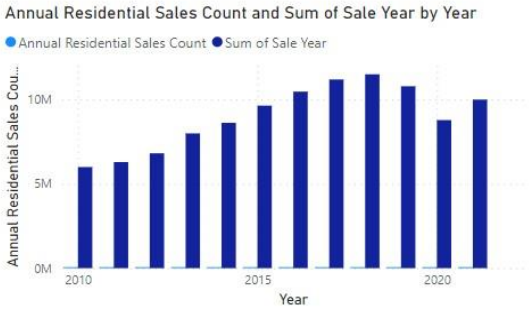
In general, predicted annual incomes in Williamsbridge are higher than Westchester. It also rose to a high point and peaked in 2018, and continually fell in the following two years. Like Westchester, it increased from 2020 to 2021.

Reference

Asaftei, G.M., Doshi, S., Means, J., and Sanghvi, A. (2018, October 8). *Getting ahead of the market: How big data is transforming real estate.* McKinsey. https://www.mckinsey.com/industries/real-estate/our-insights/getting-ahead-of-the-market-how-big-data-is-transforming-real-estate

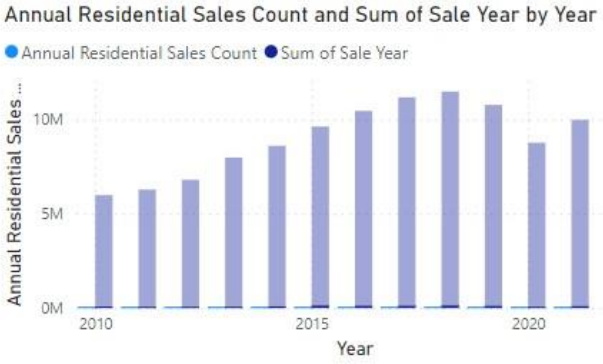
KPI Analysis (Bathway & Baychester)

Annual Residential Sales Count	NEIGHBORHOOD_NAME
387	BATHGATE
2844	BAYCHESTER
1672	BEDFORD PARK/NORWOOD
522	BELMONT
9	BRONX PARK
2136	BRONXDALE
2	BRONX-UNKNOWN
1087	CASTLE HILL/UNIONPORT
606	CITY ISLAND
20	CITY ISLAND-PELHAM STRIP
49089	



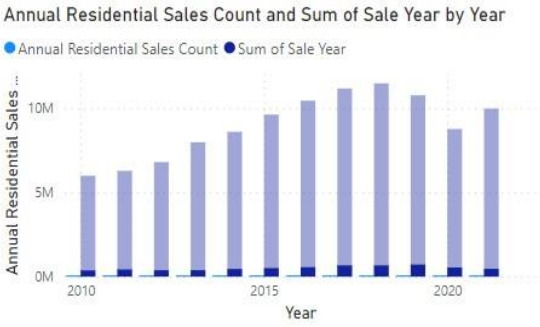
The data depicts the annual sales count in the Bronx over a seven-year period. It reveals that the highest number of sales occurred in the year 2018, followed by a gradual increase in sales from 2015 to 2021. However, it's noteworthy that there was a substantial decline in sales from the peak year 2018 to the year 2021.

Annual Residential Sales Count	NEIGHBORHOOD_NAME
387	BATHGATE
2844	BAYCHESTER
1672	BEDFORD PARK/NORWOOD
522	BELMONT
9	BRONX PARK
2136	BRONXDALE
2	BRONX-UNKNOWN
1087	CASTLE HILL/UNIONPORT
606	CITY ISLAND
20	CITY ISLAND-PELHAM STRIP
49089	

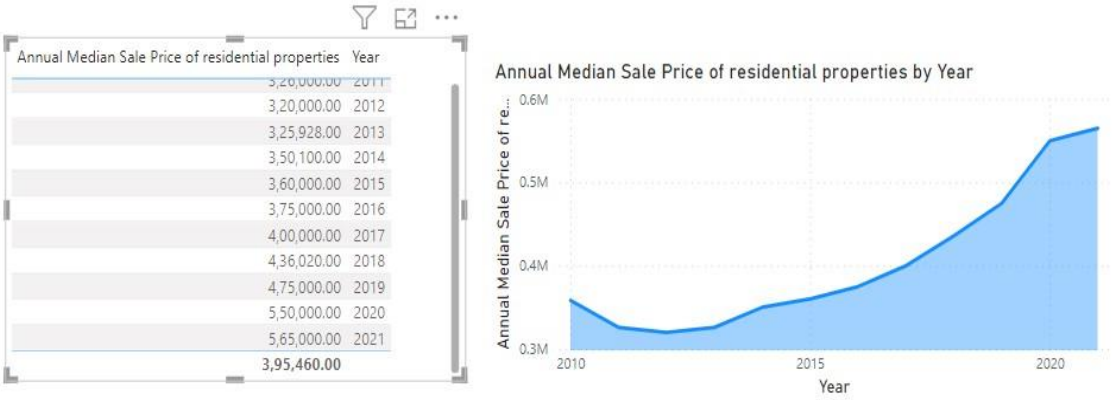


I selected the **Bathgate** neighborhood for analysis because it consistently contributed an average sales count to the data. However, upon closer examination, it's apparent that Bathgate's sales count remained relatively stable over time. Despite its consistent contribution, the neighborhood's sales trend doesn't show significant variations, making it less prominent at first glance.

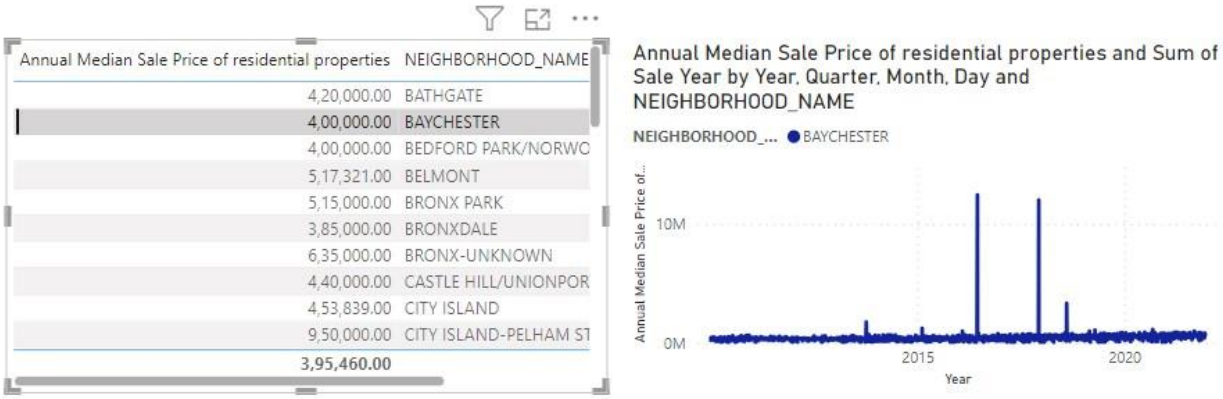
Annual Residential Sales Count	NEIGHBORHOOD_NAME
387	BATHGATE
2844	BAYCHESTER
1672	BEDFORD PARK/NORWOOD
522	BELMONT
9	BRONX PARK
2136	BRONXDALE
2	BRONX-UNKNOWN
1087	CASTLE HILL/UNIONPORT
606	CITY ISLAND
20	CITY ISLAND-PELHAM STRIP
49089	



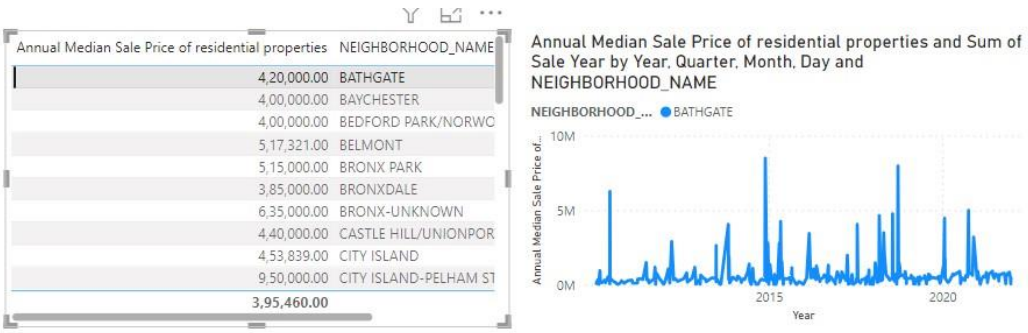
I chose to focus on the **Baychester** neighborhood because it stood out as one of the highest contributors to the dataset. In the data analysis, it's evident that Baychester's sales count reached its peak in the year 2018 and then gradually declined, but the decline was minimal up to 2021. This suggests that Baychester played a significant role in the dataset, particularly during the peak year of 2018, and maintained relatively stable sales in the following years.



In this dataset, we observe a notable surge in the annual median sales price of all residential properties. It increased sharply from 0.35 million dollars in 2015 to 0.55 million dollars in 2021. This indicates a significant and rapid rise in property prices over the six-year period.

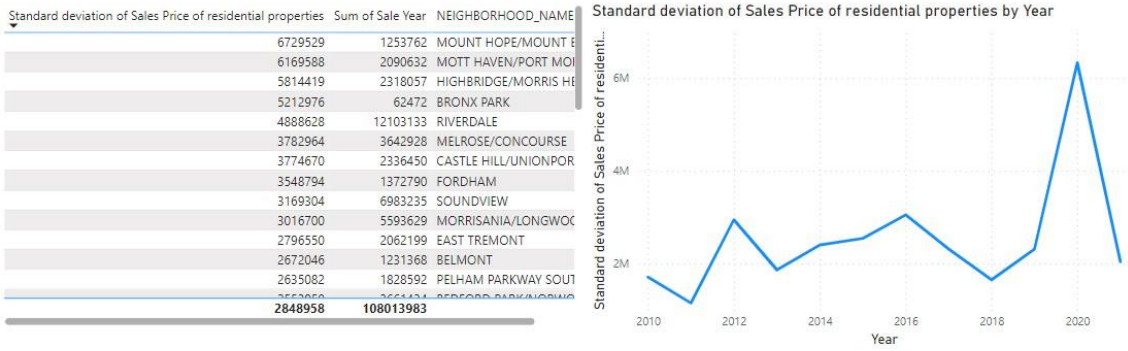


Looking at the Baychester dataset, we can see that in certain years, such as 2016 and 2018, the median property prices experienced increases, reaching higher levels. However, overall, Baychester consistently maintained its median prices in the thousands range. This suggests that while there were occasional spikes in property prices, they generally remained within a certain range during the specified years.



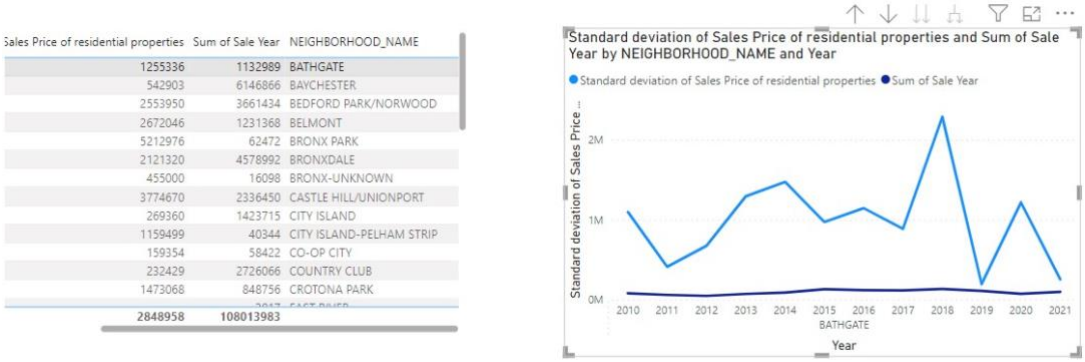
Annual Median Sale Price of residential properties	NEIGHBORHOOD_NAME
4,20,000.00	BATHGATE
4,00,000.00	BAYCHESTER
4,00,000.00	BEDFORD PARK/NORWO
5,17,321.00	BELMONT
5,15,000.00	BRONX PARK
3,85,000.00	BRONXDALE
6,35,000.00	BRONX-UNKNOWN
4,40,000.00	CASTLE HILL/UNIONPOR
4,53,839.00	CITY ISLAND
9,50,000.00	CITY ISLAND-PELHAM ST
3,95,460.00	

The Bathgate neighborhood exhibited an erratic pattern of both increases and decreases in property values from 2015 to 2018. Specifically, it experienced rises in the years 2015, 2018, and 2021. This indicates a somewhat unpredictable trend in property values during that period, with fluctuations occurring in those specific years.



Standard deviation of Sales Price of residential properties	Sum of Sale Year	NEIGHBORHOOD_NAME
6729529	1253762	MOUNT HOPE/MOUNT E
6169588	2090632	MOTT HAVEN/PORT MOI
5814419	2318057	HIGHBRIDGE/MORRIS HE
5212976	62472	BRONX PARK
4888628	12103133	RIVERDALE
3782964	3642928	MELROSE/CONCOURSE
3774670	2336450	CASTLE HILL/UNIONPOR
3548794	1372790	FORDHAM
3169304	6983235	SOUNDVIEW
3016700	5593629	MORRISANIA/LONGWOO
2796550	2062199	EAST TREMONT
2672046	1231368	BELMONT
2635082	1828592	PELHAM PARKWAY SOUT
2848958	108013983	

This chart illustrates the standard deviation of sales prices for all residential properties in the Bronx. Notably, there is a sharp spike in the year 2020, followed by an upward trend in 2019. This indicates a period of increased variability or dispersion in property prices during 2020, with prices moving in different directions, and a potential trend towards higher variability in 2019.



Sales Price of residential properties	Sum of Sale Year	NEIGHBORHOOD_NAME
1255336	1132989	BATHGATE
542903	6146866	BAYCHESTER
2553950	3661434	BEDFORD PARK/NORWOOD
2672046	1231368	BELMONT
5212976	62472	BRONX PARK
2121320	4578992	BRONXDALE
455000	16098	BRONX-UNKNOWN
3774670	2336450	CASTLE HILL/UNIONPORT
269360	1423715	CITY ISLAND
1159499	40344	CITY ISLAND-PELHAM STRIP
159354	58422	CO-OP CITY
232429	2726066	COUNTRY CLUB
1473068	848756	CROTONA PARK
2848958	108013983	

In the Bathgate dataset, we can observe a distinct pattern. In the year 2018, there was a noticeable upward trend in the data line, indicating an increase. However, this was followed by a sharp decrease in 2019. Subsequently, there was a dramatic increase, followed by another decrease. These fluctuations in the data demonstrate significant shifts over these specific years.

Sales Price of residential properties	Sum of Sale Year	NEIGHBORHOOD_NAME
1255336	1132989	BATHGATE
542903	6146866	BAYCHESTER
2553950	3661434	BEDFORD PARK/NORWOOD
2672046	1231368	BELMONT
5212976	62472	BRONX PARK
2121320	4578992	BRONXDALE
455000	16098	BRONX-UNKNOWN
3774670	2336450	CASTLE HILL/UNIONPORT
269360	1423715	CITY ISLAND
1159499	40344	CITY ISLAND-PELHAM STRIP
159354	58422	CO-OP CITY
232429	2726066	COUNTRY CLUB
1473068	848756	CROTONA PARK
2848958	108013983	

The standard deviation of Bathgate's data exhibited a spike in 2015, followed by a decrease in 2016. After this initial variation, it maintained a consistent pattern. This suggests that there was a period of increased variability or dispersion in the data in 2015, which then stabilized in the subsequent years.