SUMMARY

Tasks:

1. Visualize the two datasets on a map. What can you observe?

From the plot it is obvious that most of the gardens are concentrated in the upper Manhattan, and in the Brooklyn area. There is a very small region of concentration in lower Manhattan.

”MANHATTAN CONDOMINIUM PROPERTY Gross Income per SqFt” seems to be highest in the Centre of the Manhattan with upper Manhattan having lowest values.

“MANHATTAN CONDOMINIUM PROPERTY Market Value per SqFt” also seems to be highest in the center of the Manhattan with upper Manhattan having lowest values. NOTE – Central area of Manhattan have very few gardens.

*Many of the relatively new property were built along the eastern strip of Manhattan.

****

2. What is the most expensive Neighborhood Tabulation Area (NTA) in New York? What else can you tell about this NTA based on the data?

"park-cemetery-etc-Manhattan" NTA is the most expensive Neighborhood.

Usually expensive neighborhoods have smaller “Mean_area_perSqFt”, and cheaper Neighborhood have larger “Mean_area_perSqFt”. But in case of "park-cemetery-etc-Manhattan", both are in the higher spectrum. (data from **NTA_info** table)

*Mean_property_value_perSqFt 299.0000 Highest among all NTA
*Mean_income_perSqFt 61.00000
*Mean_area_perSqFt 377444.00
*Mean_year_built 1955.000

****

3. How do the Greenthumb community gardens affect the value of the property? And what about rental income?

<u>Assumption - Post code having large number of gardens have properties which are on average closer to the gardens.</u>

One logical way to analyze is to compare Postal codes which are very close to each other and have large variations in number of gardens. (data from **Postcode_info** table) :

For example:

Postcode	No. of Gardens	Mean_property_value_perSqFt	Mean_income_perSqFt
10034	**1**	**110.25**	**26.08**
10035	**17**	**132.35**	**29.56**
10036	**2**	**188.77**	**39.31**
10037	**1**	**110.25**	**26.08**

Note-A lot of data is missing, some of the data can be calculated based on the given fields if there is more time.

CONCLUSION – There is no clear evidence that Greenthumb community gardens are influencing the rate. Although this analysis is preliminary.

Scatter plots and correlation matrix shows very little evidence that gardens are affecting the value of the property or rental income. Although detailed analysis might be different.

****

NOTE –

Best Approach: (Can be done if given more time)

According to me the best way to solve this problem is to perform following steps:

1) **Compute the distance between each garden coordinates and property in a particular region.**
2) **If the distance is less that a threshold (assume a threshold, example 0.2 miles), then that property is closer to that garden,**
3) **Group all the properties in a region closer to the gardens and assign a variable 1.**
4) **Group all the property that are far away (greater than 0.2 miles) from any of the gardens and assign value 0.**
5) **Compare the prices of two groups in each geographic region and come up with a hypothesis, take any other possible factor influencing the results into the account.**