**New York City Housing Guide Business Questions**

How is this different from my original project in :
https://public.tableau.com/app/profile/winson.tom/viz/NewYorkCityHousingGuideFinal/Story1

In my original I was emphasizing how location is the primary factor determining pricing. I gave per zip code pricing, differences in pricing per neighborhood & per property type, price ranges for the addresses, price per square foot, and neighborhood safety rating. While this was good to explore the data that was already provided, I found this lens to be for "data collectors" rather than a functional tool for actual buyers. People who would use the previous dashboard would include developers, people working for the Census, or writers looking for a broader point of view. However, this was intended to serve buyers wanting to move into New York City or wanting to re-locate within New York City. With that you need a lot more granular image. It requires answering more day to day living, personal questions. This dashboard and 2nd round of the project looks to solve the problem of "I don't know anything about the different parts of NYC and what it's like to live there. Give me a summary ".

Specific Problem:
**"I want to buy a personal residence in NYC, but I don't know anything about NYC neighborhoods. What property is right for me?" - Buyers**

*This is a very common question that buyers will ask their real estate agent. There are a few layers to this question. First, we identified this is for the buyer to actually live in. Second to identify what property is right by understanding the standardized costs such as the mortgage price, property tax price, insurance price, and utility price.; We are going to assume the property is in functional condition and no major renovations are included in this- Regular monthly costs. Then the question of what property type is right includes price, # of occupants, neighborhood's safety rating, neighborhood's perception, and commute time to the "city".*

Questions that support this Specific Problem:
1. What is the estimated mortgage rate per month?
2. What is the estimated home owner insurance price per month?
3. What is the estimated utility price per month?
4. How many people can live here?
5. What Is the neighborhood safety rating?
6. What are the common perceptions of this neighborhood?
7. How far is the commute to Mid-town Manhattan?
8. How "Dense" is my neighborhood?

1. What is the estimated mortgage rate per month?
   - *Mortgage rates are primarily determined by interest rate, duration of loan, down payment amount, and credit score.*

   To standardized the rates across the data we will be using
   - *6.75% interest rate*
   - *Fixed 30 years loan*
   - *20% downpayment*
   - *750 credit score*

2. What is the estimated home owner insurance price per month?
   - *Most lenders will require owners to have homeowner insurance.*
   - *Homeowner insurance vary based on property value, crime rate of the neighborhood, coverage, and proximity to water*

   To standardized the homeowner insurance, we will attribute the same cost across all properties within a certain neighborhood.
   - *Upper East Side, Upper West Side, Tribeca, Chelsea: $2,500*
   - *Harlem, Washington Heights, Inwood: $1,850*
   - *Brooklyn Heights, Park Slope, Williamsburg, DUMBO: $2,250*
   - *Bushwick, Bedford-Stuyvesant, East New York: $1,850*
   - *Astoria, Long Island City, Forest Hills: $2,050*
   - *Jamaica, Flushing, Far Rockaway: $1,850*
   - *Riverdale, Pelham Bay, Throgs Neck: $1,850*
   - *South Bronx, Hunts Point, Soundview: $1,600*
   - *St. George, Tottenville, Great Kills: $1,850*

3. What is the estimated utility price per month?

Utility costs will consist of the necessities such as electricity, gas, water & sewage
   - *Utility rates will vary based on the city's utility company's electricity, gas, water rates.*
   - *Total usage as well as location will cause the costs to vary*

To standardized the utility price per month, we will create Square Footage ranges, then use this size of the property to predict the utility costs:
   - 0 to 500 - $160
   - 500 to 1000 - $205
   - 1000 to 1500 - $375
   - 1501 to 2000 - $550
   - 2001 to 3000 - $600
   - 3001 to 4000 - $1000

- 4001 to 5000 - $1,400
- 5000 to 6000 - $1,700
- 6000 to 7000 - $2,000

4. How many people can live here?

**Maximum Occupancy** states "The maximum number of persons who may occupy any such apartment shall be determined by dividing the total livable floor area of the apartment by eighty square feet." This excludes private halls, foyers, bathrooms or closet space, but it does include the kitchen. As we cannot predict what area for every house is livable and not livable , we add an additional 40 square feet to the 80 square feet rule for every bathroom. To summarize we will use the equation:
- Square Feet of property / ((80 + (X*40)).
- X being the number of bathrooms.

Example: 1 bathroom apartment that is 750 square feet.  750 / ((80+(1*40)) = 6.25

Maximum 6 residents.

**Recommended Occupancy:**

- NYC is generally more expensive. Many people will have roommates sharing to reduce costs of living.
- NYS Housing Recommendations
1. Studio – 2 people
2. 1 bed – 2 people
3. 2 bed – 4 people
4. 3 bed – 6 people
5. 4 bed – 8 people
6. 5 bed – 10 people
7. 6 bed – 12 people
8. 7 bed – 14 people
9. 8 bed – 16 people
10. 9 bed – 18 people
11. 10 bed – 20 people
12. 10+ bed – (2x # of beds)= people living

- Using the NYS government housing recommendations of how many residents should the space be comfortably accommodated is a good reference point for the housing.

5. What is the neighborhood safety rating?

Data pulled from Niche.com. Using Python, input all the info into Safety_Rating_Updated_Neighborhoods.xlsx

6. What are common perceptions about this neighborhood?

Perceptions on the Excel Sheet. Clean_Data with Perceptions.csv

7. How far from the "city"?
*Most common NYC problem as most people will need to commute to Manhattan for work or entertainment. There is always a need to measure your distance from city, which will give an idea of the transportation time required to commute there.*

Excel Sheet, Python formula to find distance in miles between coordinates.

Python file: 4.0 Distance to "Work".ipynb
Excel File: Distance_To_Midtown.xlsx

8. What is the population density per neighborhood?

Population Density is important to know as home buyers as you want to know just how packed is your neighborhood and how this compare to other areas. The measure of this in the United States is measured by the Population / Square Miles. Although borough information is important, we ultimately spend a lot of time within our neighborhoods. Knowing the specific neighborhood population, then finding the neighborhood square miles , and then calculating the population density.

What visualizations would best depict the information we have gained?

Vertical Bar Chart – <u>Category comparison</u> – What Boroughs & total listings available. Y axis , # of listings, X axis – 5 boroughs. "Page 0 intro duction"

Horizontal Bar Chart – <u>Show Rankings</u> – Average prices of each neighborhood. *Can assist with Q1

Stacked Bar Chart- <u>Part to Whole; How much it contributes to total</u> – Total listings and amount that are houses , apartments, condos

Histogram – <u>Showing distribution –</u> Mortgage price per month range & how much of the diff listings in each.

GeoJson File – Map of NYC

Steps:

1.  8 new columns were created. These are "Down Payment, Mortgage Rate, Years, Loan Amount, Loan Payment per month, Property Tax Per Month, Homeowner Insurance, Utility".
    a.  Down Payment, Mortgage Rate, Years are all fixed values of 20% down payment, 6.75% interest rate – the interest rate in January 2024 , 30 year mortgage.

2.  Loan Amount was calculated as 80% of the Price. ***This does not provide the loan payment per month. It is not just as simple as Loan Amount / # of Payments.***

3.  Loan Payment Per Month otherwise known as PMT.

    -   Calculated through Excel's PMT function.
    -   P = Principal is the amount calculated in the Loan Amount
    -   R = Rate 6.75%
    -   T = 30 years on the loan
    -   N = number of payments

4.  Property Tax Per Month
    -   NYC annual property tax is a unique calculation. The annual amount is 6% of the assessed value * 20.309% if it is a class 1 property.
    -   Since we do not have the assessed value by a NYC appraiser, we are to use the market value.
    -   Class 1 property is defined as a residential property of up to three units, which is majority of the properties. Therefore, we will be using 20.309% the % rate.
    -   Excel input: ((Sale Price * 6%) * 20.309%)/12

5.  Homeowner insurance
    -   Primarily determined by the neighborhood safety, crime rate, history of property, location, and the coverage chosen
    -   We do not have all the sufficient information to customize each and every address
    -   For ease of standardizing this, a formula commonly used is Sale Price * .0025 / 12 to find the monthly home owner insurance amount.

6.  Utility
    -   Utility costs will also greatly vary based on an individual's property size and individual usage.

- Within Utility we included electricity, gas, water & sewage

To standardized the utility price per month, we will create Square Footage ranges, then use this size of the property to predict the utility costs:
- 0 to 500 - $160
- 500 to 1000 - $205
- 1000 to 1500 - $375
- 1501 to 2000 - $550
- 2001 to 3000 - $600
- 3001 to 4000 - $1000
- 4001 to 5000 - $1,400
- 5000 to 6000 - $1,700
- 6000 to 7000 - $2,000

- Excel created If And function for the various square feet ranges.

7. Create 2 columns. $1^{st}$ for maximum occupancy. $2^{nd}$ for recommended occupancy.

8. Create neighborhood safety column. Used If then statements on Excel to categorize the rating of each neighborhood.

9. Scour the internet for various perceptions of the neighborhood. Add as an Excel Sheet. Input column

10. Identifying distance to "Midtown" a very central location in which a large portion of people work around; Good reference point and concern of those living in NYC. Since everyone's mode of transportation is different and NYC traffic/delays are extremely variable I decided on distance rather than a specific commute time. Create excel sheet with all the "starting point" latitudes and longitudes. Another 2 column for the "ending point" which was Empire State Building. Having created a distances sheet I went on Python to create a script that determines the distance in miles from "First coordinates" to "Second coordinates"?

11. Python – Create column for population, create column for area. Use Excel to divide the two columns to get Population Density. Resolve the "Errors" as some were lacking proper neighborhood.

12. **In an attempt to answer Question # 8, we had to carefully review each neighborhood & the population density within them. However many of the neighborhoods were too general "South West Brooklyn", "Central Harlem", "North-East Bronx". I decided to dissect each of these generalized neighborhoods further to get greater understanding of what each neighborhood's population density is like.

*As a result we will have to make edits to the columns: Population, Square Miles, Population Density*

13. As all things population density is now addressed, we now have to <mark>add Perceptions</mark> and <mark>Neighborhood Safety Rating</mark> for the updated neighborhoods. Previously there was 41 neighborhoods. Now there are now 137 neighborhoods & perceptions. Perceptions_2.0 completed.

14. Matching Addresses, with the proper updated Neighborhoods.

15. Updating Newly updated Neighborhoods with the perceptions & safety rating.

16. Returning back to Step 10, incorporating the data of "distance from Midtown" into the primary data set.

17. Incorporating Population Density into the whole data set. (Full Data Set after 6.0 .xlsx)

18. Identify what visualizations would best depict the information.

19. Editing the full data set to become smaller data sets , easier to load and create visualizations.

20. Creating vertical bar chart, horizontal bar chart on Python.

21. In the process of creating the stacked bar chart, I discover Property Types column was uncleaned, a lot of "for sale, foreclosures, pending, N/A". Fixed all of these up and added it to the full dataset. StackedBarChartData2.0.xlsx . Created Stacked Bar chart with the updated information

22. Create Histogram on Python.

23. Incorporating data into Tableau Worksheets

24. Created Stacked Bar Chart on Tableau showing total listings per borough and property types

25. Created Histogram on Tableau showing count of bedrooms per property

26. Created Histogram on Tableau showing listings per price range

27. Created chart on Tableau showing maximum occupancy per property

28. Created GeoJson map on Tableau showing distance to Midtown Manhattan per neighborhood.

29. Created Horizontal Bar chart on Tableau showing Safety Ratings and # of available listings.

30. Created Population Density Map on Tableau showing each neighborhood's population density.

31. Realized Population & Square miles column were not fully cleaned and therefore population density has some error. Going back to standardize both population and square miles for ALL neighborhoods. Making sure every neighborhood has the same size and population and therefore population density.

32. Clean Staten Island Data, neighborhoods are too broad. Only change it to North Shore, Mid-Island, South Shore. This is how Staten Islanders categorize the regions.

33. Upkeep for Tableau Dashboards organizing spacing, titles, legends.

34. Finalize all Tableau storyboards.


Final Tableau Storyboard:
https://public.tableau.com/shared/YR4ZKFB3R?:display_count=n&:origin=viz_share_link


Data Sources:

https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market

https://communityprofiles.planning.nyc.gov/staten-island/1