**Presentation Speaker Notes**

Carissa | Samantha | Brandon | Sara

1. Intro question
2. Project Description
   1. \*\*\* Sparked by curiosity about the housing boom that seemed to occur during the covid-19 pandemic, we took a look at the real estate and population data in New York through the years 2019 - 2021 to find the relationship between housing prices and population. We will input this data into a Linear Regression Machine Learning Model to see if we can predict future housing prices for New York in both low and high population cities. We will use this data to determine:
      1. If there was an effect on housing prices and populations during 2019-2021
      2. If other features affect the price of a house: such as bedrooms, bathrooms, square footage, and acreage
      3. If the change in population has an impact on housing prices.
3. Data Exploration - population cities
   1. We found population data for 2019-2021 to add to the real estate data, to determine if population was one of the factors as well
   2. This led us to focus in on
      1. 5 higher population cities: Bronx, Brooklyn, Queens, Staten Island, and Yonkers
      2. 5 lower population cities: Scarsdale, Tuckahoe, Suffern, Nyack, and Stony Point
4. Initial Analysis – beds/baths
   1. We used the next few charts to help us visually determine outliers in the data, such as houses with 12 bedrooms, or 10 bathrooms, which would skew our findings
   2. We found that the majority of houses in all of New York had

3-5 bedrooms and

2-3 bathrooms

1. Initial Analysis – acreage
   1. Most of the houses we had data for, were located on lots of less than 1 acre
2. Initial Analysis – square footage
   1. We also looked at the square footage of each house, and determined that the majority of them were between 1 and 4000 square feet
3. Initial Analysis – price by zip code
   1. and that the average cost of the houses was significantly more expensive in the southeastern corner of New York, around New York City and the Long Island areas
   2. This helped us to break our data down into smaller chunks based on population, to try to provide more accuracy in our population analysis, since the prices appeared to be much higher in the higher populated areas
4. Results - Impact of Population – Scroll down to Sara’s Tableau Story
   1. Chart 1 – city populations 2019
      1. The first few slides show the difference in populations between high population and low population cities.
      2. In 2019, of the high population cities chosen, Yonkers had the lowest population at 200,370 while the most populous city was Brooklyn which had a population of over 2 million 500 thousand.
      3. In comparison, Tuckahoe is the least populous of the low population cities chosen with a population of less than 7,000, and the low population city with the highest population was Scarsdale which had a population of 17,871.
   2. Chart 2 – city populations 2020
      1. In 2020, populations in high population cities rose by 5.43% on average and by 6.64% on average in low populated cities.
      2. Hovering over the pie chart, you will see that Yonkers' population rose by over 10,000, Brooklyn’s population rose by over 100,000, Tuckahoe’s population rose by over 500, and Scarsdale’s population rose by over 300.
   3. Chart 3 – city populations 2021
      1. Fast forward to 2021, and we see a decline in populations.
      2. High population cities experienced a (2.78%) drop in populations while low population cities populations dropped by (.81%)
   4. Chart 4 – average # of acres, home price – high population
      1. The following two bar charts display the difference in average acreage vs average home prices of the high populated cities chosen for the analysis. The darker in color the bar is the higher the average number of beds.
      2. As you can see, the Bronx and Staten Island have a higher average number of acres. However, two of the cities with the lowest average acreage – Brooklyn and Queens – have the highest average home price. Viewers can hover over the bars for population statistics and average number of beds, baths, and acres.
   5. Chart 5 – average # of acres, home price – low population
      1. Again, you see the darker the bar, the higher the average number of beds. The low population cities have higher average acreage across the board but have less average number of beds. Average home prices are also lower in these cities. This bar chart also includes the tooltip.
5. Machine Learning Model Accuracy
   1. For our linear regression models
      1. The high pop cities have an accuracy of 50%
      2. And the low pop cities have an accuracy of 77%
6. Results - Predicting Prices for Cities
   1. Results show as expected for the traditional factors of a house
      1. But high pop shows a weak impact
      2. And low pop shows the opposite of expected
      3. This points to the inaccuracy of our model
   2. However, the change in pop from 2019 does show a decrease in price in high pop
      1. And an increase in price for low pop
      2. The inverse is seen for 2020
      3. This supports the idea of people leaving larger cities during the height of the pandemic and having to return to them once offices were no longer offering remote work
7. Summary  – Was there an effect on housing prices and populations during the years 2019-2021?
   1. high population cities: All of the cities had a population spike in 2020 followed by a small drop in 2021. It appears as though Queens and Staten Island had price drops after 2020. There does not seem to be a pattern or correlation between housing prices and population between 2019 and 2021.
8. Summary  – Was there an effect on housing prices and populations during the years 2019-2021?
   1. low population cities: There were a few minor increases in population in Stony Point, Suffern, and Tuckahoe. Our dataset only had housing price data from all 3 years for Nyack and Scarsdale. All we are able to determine is that housing prices dropped in these 2 cities from 2020 to 2021, but we are unable to determine if it’s related to the pandemic.
9. Summary – Does the number of bedrooms matter?
   1. When looking at whether or not the number of bedrooms affect the price of the house, we can see that in high population cities, it does not appear to have any impact by itself, but in low population cities it does, especially when the number of bedrooms exceeds 5
10. Summary – bathrooms
    1. As far as the number of bathrooms is concerned, it follows a similar pattern as the bedrooms with there appearing to not be a correlation for the higher populated cities, but the lower populations are somewhat affected, especially when the number of bathrooms exceeds 5
11. Summary – square footage
    1. Does the square footage of the house affect the price? Again, higher populated cities do not appear to matter with regard to the cost of the house, but lower populated cities do show a direct correlation, especially when the square footage exceeds 4000
12. Summary – acreage
    1. However, the size of the property’s acre lot does not appear to have any effect on the price of the house for both higher and lower populated cities
13. Summary – Does population have an impact on housing prices? - high population cities
    1. The graph here shows our actual data (blue) vs our predicted outcomes (pink) and shows the margin of error for this difference
       1. The left is only population
       2. While the right is all the features used
       3. This shows that population and change in population do not seem to be strong indicators for price change for high population cities
14. Summary – Does population have an impact on housing prices? - low population cities
    1. The same can be said for lower population cities
15. Recommendations for Future Analysis
    1. If we were to recommend changes to this model it would be
       1. To aim for more specific markets (smaller areas)
       2. And find better ways to evaluate location itself
       3. This would be done by evaluating crime rate, school quality, and job opportunity to better evaluate location.