Housing is a universal problem that affects everyone and can be tricky trying to understand what makes housing prices go up and down. With the goal of our project we wanted to help individuals make more informed decisions when buying or selling their house. There are multiple factors that contribute to housing prices like building square feet, area code/neighborhood, and age of the property. We are going to focus on the building square feet of a house and its sale price, we are going to create a normalized model then create an interpolating polynomial equation to help us see where a house with varying square feet would fit on our graph. Important information to know about our project is that our data set is based on Cook County Illinois, we chose this specific data because it encompasses such a diverse area with the northshore, parts of Chicago and surrounding neighborhoods. It was also under investigation for having unfair housing prices with certain northshore areas being evaluated at lower than market price while lower income housing was being overvalued causing less fortunate families to pay more in taxes based on their properties and the rich less. All of our data was collected by the Cook County Assessor's office, between 2013 till 2019 and was collected digitally by accessing storage locations but that information was not available. The main question we hope to answer is if you know your building square feet or land square feet can you accurately determine a reasonable price for your house.

Summary & FCQ

Our data collection comes from the Cook County Assessor's office which is a government website that manages housing appraisals. For our data set we had a lot of outliers that needed to be taken into consideration, a house being sold for 1 dollar and another being sold for 7.1 million. This skewed our data so we had to take them out and make our data fit between on a 95% confidence interval dismissing outliers. After that we had to take the log of our sale

price to normalize our graph. To find this data, one of our group members is directly affected by this as he lives in Chicago, and this has been a known problem. To make sure that data worked for us, we retrieved relevant information from the .csv files, to be specific, the price and square footage of buildings. We used these two variables because they're directly related, since generally, the price of a building is directly proportional to its size in square feet. The original data retrieved was unnormalized and there were many houses that were sold at or close to 1 dollar (various reasons, like selling house to family member), we got the log of the values and plotted them on a graph for both the sale price and the square footage of any given building. Then, once we received the two graphs, we merged them into one graph that would allow us to predict values of a given square footage visually.

Overall, through this project we were able to find evidence that Cook County unfairly priced the housing in their area. How we found this out, we normalized the unreliable data and combined them to give us a prediction chart. Using this chart, we interpolated the data by segmenting it and then using that information to predict a price value at a given square footage. We did manage to get a fair price for the building square feet, however due to other factors like area code, and additions done to the home our price was off compared to other houses with similar square footage. To conclude, we were able to answer the question if you could get a fair price for your house based on building square footage, but you still need to consider many other factors to get a realistic price.

Important links:

https://datacatalog.cookcountyil.gov/stories/s/p2kt-hk36
https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/assessments.html
https://chicago.suntimes.com/politics/2019/11/7/20954286/unfair-assessments-lawsuit-dismissed-assessor-kae
https://www.chicagotribune.com/2017/12/14/lawsuit-targets-berrios-over-biased-residential-assessments-in-coo