Throughout the data collection,cleaning and analysis process the group has learned several things that will alter the details of our final draft as well as our model. The data collected was based on housing prices annually, the location of houses sold and the amount of houses available at certain price ranges. One of the first things recognized in the data collection process is how it will take more than just the analysis of the years 2020-2022 because that data set is too small and doesn’t offer enough information. Instead it was decided to look at data from the past 50 years and document that instead in order to truly understand what impacts the price of housing in the United States. One impact discovered in the data collection alone was location. It seems that houses in the South and in the West are sold much more frequently than houses in the North and in the Midwest. This may play a factor in the prices of housing since the price of a house can fluctuate often with different owners. With this being taken into consideration location will have to be one of the variables that impacts the model. Another realization that was found in the EDA process was that there is a positive correlation between the year a house was sold and the price of the house. Obviously this is most likely due to inflation however for our research question it’s important to see the average growth in this price year to year compared to the years after and during the global pandemic. This may not be something that plays a factor in our model as much as it will be something that will be heavily covered in the final report. In order to come to these conclusions there obviously had to be some data reduction done in order to view the data in its cleanest form. One major action taken was to make sure there were no null values of course. By doing this alone it made it a lot easier to view the data in different ways through graphs, charts and other models. Another action taken was to make sure that values outside of the 3 standard deviations weren’t in the data since those would be considered outliers. Lastly, after obtaining the data, cleaning the data and performing the EDA it is now time to begin developing the mode which will be followed with modeling refinements. Finally there will be a professor review of the model and give the model its final touches.