Michael Ayala

INFO 523

1. Link: <https://arizona.zoom.us/rec/share/5WU4KAWUsgODhfaS_XvEy4eZh9XTpKJ2qBYbVRrZVz5Y9YfTekl5IC27KKPTZp_V.M3AiOcIcV9PAMH7s?startTime=1660631195000>
2. If given housing parameters in a dataset from a city in the state of Iowa, can we determine what is the most abundant and affordable housing based on the type and layout of the house?

* I have chosen parameters that I care about out of a list available in git-hub: [data-descriptions](https://github.com/mayala77/Final_Project/blob/main/data_description.txt)
  + LotArea – Lot Area (Sqr Feet)
  + FullBath – Full Bathrooms
  + HalfBath – Half Bathrooms
  + BsmtFullBath – Basement Bathrooms
  + HouseStyle – House Style
  + BedroomAbvGr – Bedrooms Above Ground
  + MSZoning – The type of Zoning the house is in
  + SalePrice – Sale Price of the Property

1. Worked alone
2. <https://github.com/mayala77/Final_Project>
3. I have taken an interest in housing over the years and wanted to do a project on the topic. I purchased a house about 3 years ago and have really enjoyed my purchase. I would like to make another home purchase in the future and wanted to make some determinations based on the parameters listed above. These are the typical care abouts when deciding for myself what house I should buy. In Arizona there are no basements, but the data that was used is a city in Iowa and found it to be relevant as I went through the data. I began a simple run through of some of my main comparisons: [Cost Per Lot Size](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Cost_Per_Lot_Size.png), [Cost vs Zone Type](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Cost_Per_Zone_Type.png), and [Housing Style Availability](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Availability_of_Specific_House_Style.png). This information then drove some simple filtering, [Single Story – Cost Per Lot Size](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Single_Story_Lot_Prices.png). Considering that the results of the previous comparisons I could move forward with Clustering the data:  
   LotArea < 50,000 & MSZoning == 'RL' & HouseStyle =='1Story' & SalePrice < 600,000  
   I separated into 5 grouping to show the correlation between [Lot Area vs Sale Price](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Lot_Area_vs_Sale_Price_5_Cluster.png).   
   When involving housing we also need to add in Bedrooms and Bathrooms as this is important in buying a house. I wanted to determine the types amenities the houses have involving [Bedroom\_vs\_Bathroom\_In\_Homes](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Bedroom_vs_Bathroom_Types_in_Home.png). It was clear that a specific amount bedrooms and bathrooms were present and so I filtered the data more to show a different clustering plot of 5 groups. This time it was [Bed\_vs\_Bath\_Filter\_on\_Lot\_Area\_vs\_Sale\_Price](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Bed_vs_Bath_Filter_on_Lot_Area_vs_Sale_Price.png) which was narrowing down the search for the home of interest. Based on the original parameter of unfiltered data I wanted to determine solely based on the 1 Story parameter if the unsupervised machine learning could tell me the same answers that I had already determined. I was not sure what support level to put so I guessed with 25%. To my surprise the analysis showed me that my determinations were correct of the data with the [8 Rules Scatter Plot](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Scatterplot_for_8_Rules.png) and [25% Support Graph](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Graph_Machine_Learned_Data_Sup_25.png). The support level was changed to be 20% and the data in [21 Rules Scatter Plot](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Scatterplot_for_21_Rules.png) and [20% Support Graph](https://github.com/mayala77/Final_Project/blob/main/Data_Images/Graph_Machine_Learned_Data_Sup_20.png) showed just how close other data involving Bedrooms and Bathrooms were to the data I originally selected. This analysis proved to me that based on the selection of the parameters of interest, the machine learning Algorithm can find the correlation of interest based on the support level.

(Images in the Hyperlinks are also in the #4 Repository)