- Narrative that tells a story with the data as you complete this section.
  - ⇨ With the in-demand housing market and increasing prices, Real estate / Housing prices topic has gotten my interest. Along with external financial market conditions in a country, various factors contribute to the sale prices of houses / residential properties like size of property, living area, the localities / neighborhood, age of property, number of rooms, building quality to name a few. As a part of the project work for DSC 520, I would like to build a possible model to be able to predict the prices based on various factors and get as much closer as possible in terms of predictive power.

- Summarize the problem statement you addressed.
  - ⇨ Traditionally, the factors like lot size, living area, number of rooms, bathrooms, neighborhood, building quality, building age etc. have been known to affect the real estate prices. However, I wanted to explore the impact of additional factors like Foundation, Basements, open area, masonry, building types and style, garage characteristics and finishing, Remodeled age,

- Summarize how you addressed this problem statement (the data used and the methodology employed).
  - ⇨ In addition to the set of factors considered in the first stage of analysis which involved mostly numeric variables like area / size, number of rooms / bathrooms, overall condition and quality etc., I had to iteratively keep refining model to include the categorical variables around Housing Zones, Open areas available, Foundation and Roofing styles, Masonry types, Exteriors and Basement Finishing. This was performed in step by step, iterative manner by looking at the various Histograms, Scatter Plots as well deriving correlation coefficients and checking the R-Squared and Adjusted R-squared values for the model, p-values, Analysis of Variance, Durbin-Watson test statistics, autocorrelation and multicollinearity aspects. Also, the VIF analysis, Residuals Fitting, Q-Q plots and studentized residuals, Cooks distance were other helpful factors in verifying the accuracy of the model built.

- Summarize the interesting insights that your analysis provided.
  - ⇨ Overall, people interested in buying a house need to consider not only the basic factors like area / size, number of rooms / bathrooms, neighborhood, garage area, year of initial built but also the overall quality of the construction, Foundations well Roof types, when the house is remodeled, if applicable and Exteriors, open areas and masonry work along with Basement finishing. Based on different geographic locations, the proximity to key essentials and amenities like hospitals, schools, shopping malls, commercial spaces, airports, city / town centers, crime rates etc. are also important. So, people need to consider all the aspects and some more before determining fair price of the property and purchasing it.

- Summarize the implications to the consumer (target audience) of your analysis.
  - ⇨ This study has helped uncover many different angles to look at the available housing market data and should be helpful to an extent for understanding the factors driving the prices. However, this study was done on a limited sized data, which may hold some level of distortions and did not have some of the above mentioned factors. So, buyers need to be

cognizant of these points and perform their own research or consult local market experts as well.

- Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.
  - ⇨ Looking at the model, it is fairly close representation of the sample and a generalizable model to the larger population. It can be improved further by deleting outliers for model building purpose. We may need to consider additional predictor variables or more cleaning of data is needed. Another way, perhaps could be finding additional / larger size of data, which can smoothen out the normal distribution even further and help improve accuracy of the model. Current data size after filtering became 1344 records. So, possibly a little larger data size could help. Also, this study dataset did not cover any other aspects like proximity to commercial as well as essential amenities, airports, oceans, city centers