Data exploration and storytelling with hypothesis testing

1) How Sales Price looks like by MSzoning? MSzoning contains "Agriculture", "Commercial", "Floating Village Residential", "Industrial", "Residential High Density", "Residential low density", "Residential low density park", "Residential Medium density".

   The above figure shows that the median housing price for "Floating Village Residential" is the highest. The next highest median housing price is "Residential low density", while the lowest median housing price is "Commercial". The box plot presents that the housing price for "Floating Village Residential" is more dispersed. The least dispersed housing prices is the "Residential Medium Density". Medium housing prices: FV>RL>RH>RM>C. Mean housing prices: FV>RL>RH>RM>C

2) Explore how does housing price look like by neighborhood?

   The mean housing price by neighborhood shows that on average, the most expensive neighborhood is "NoRidge", followed by "NridgHt", "StoneBr", "Timber" and "Veenker".

3) Does "Month sold" show any pattern in housing prices? This aims to see whether there is some seasonality of the housing market.

   I looked at the mean housing prices for houses sold in September and November versus the mean housing prices sold for other months. The method I used is the independent T-test. The p-value for the test is 0.04, this shows that the difference between the mean housing price for September November and the mean housing price for other months is statistically significant. So seasonality does exist.

4) Want to explore whether garage type makes a difference in housing prices

   Build in and attached garages have very high mean housing price comparing to other types of garage. I did a hypothesis testing to see whether this is statistically significant. I used the independent T-test again to compare the mean housing prices with the building and attached garages and mean housing prices with other garages. The hypothesis test result shows that the difference between the mean housing prices with buildin and attached garage and the mean housing prices with other garage types is statistically significant.

5) Exploring some correlation between the features and the housing prices. I am specifically interested in the following features: exterior quality, exterior condition, kitchen quality, house square feet, total property square feet, number of total bathrooms, age of the house, overall quality and overall condition.

Given the hypothesis test results, all the p-values of the feature shows the correlation between the feature and housing price is statistically significant. Thus, housing prices is highly correlated with (absolute $r>0.5$) exterior quality, exterior condition, kitchen quality, house square feet, property square feet, total number of bathrooms, and overall quality. Specially the most correlated three features are total property square feet, house square feet and overall quality.