**OBSERVATIONS FROM HOUSE PRICES DATASET**

After performing exploratory data analysis on the house prices dataset, I observed the following:

* From the heat map, some of the independent variables are correlated with each other. One of the basic assumptions of linear regression is that the independent variables are not correlated with each other, if they are then we will not know which of the two variables is imparting the correct information about the dependent variable.
* The first feature which has the maximum correlation with the Sale Price is OverallQual. As we can see in the scatter plot below, that when the OverallQual is increasing, the average Sale Price for each OverallQual is increasing which actually makes sense.
* The second most important feature is the GrLivArea. We can see in the graph that as the living area increases there is an increase in the SalePrice of the house. But we can also see that there are couple of outliers where GrLivArea>4500 but Sale Price<200000.
* The third most important feature is the GarageCars. We can see in the graph that as the GarageCars increases there is an increase in the SalePrice of the house. But we can also see that there are couple of outliers where GarageCars >3 but Sale Price<300000
* The fourth most important feature is the GarageArea. We can see in the graph that as the GarageArea increases there is an increase in the SalePrice of the house. But we can also see that there are couple of outliers where GarageArea >1200 but Sale Price<300000.
* The fifth most important feature is the TotalBsmtSF. We can see in the graph that as the TotalBsmtSF increases there is an increase in the SalePrice of the house. But we can also see that there are couple of outliers where TotalBsmtSF >6000 but Sale Price<200000
* Another point to be noted here is that GarageCars and GarageArea are correlated with each other in a sense that, larger the area, larger is the number of cars that can be accommodated. This should be investigated in the correlation matrix.
* Houses that were built recently have higher prices as compared to others. Most of the houses were built after 1950. A good third of all properties was built 1990 and later. Newer houses tend to yield a little higher mean of sale price. Many houses were built in 2005 and 2006 as compared to other years.
* Fewer sales in 2010, this could be either because less samples were collected. Or the financial crisis of 2009 that hit the market. Most sales were in the summer months. On average the properties were 37 years old at the time of sale (with a mean of 35 very close to that). The oldest house was 136 years old and we have sales of house that were built in the year of sale. I observed a notable difference for newer houses. E.g. the median of price for houses between 0 and 19 years of age is almost the same now.
* Most of the houses are 1Story/2Story. Some of the houses have One and Half story. The wide majority of properties are Single-family Detached. Townhouse End Units (TwhnsE) come second pricewise and are more expensive than Inside Units (Twhns).
* Most houses have Poured Concrete or Concrete Block as their Foundation. Poured Concrete has higher values of Sale Prices. Most of the houses are Inside Lot type followed by Corner plot.
* More than 50% of the houses have average (rating of 5) overall condition of the house. There is no house with a rating of 10. Very few have a rating of 9. OverallCond does not have the same effect as that of OverallQual towards SalePrice variable.
* Most of the houses have Gable type of roof followed by Hip type. Houses with Gable type of roof also have the highest sale price. This tells us that there is a relationship between roof type and sale price.
* Most of the houses have Average / Typical external quality of the house; this is followed by houses that are rated Good. No Missing values here. The prices of houses that have excellent external quality rating are much higher as compared to others.
* Most of the houses have either Typical or Good condition of the material on the exterior.