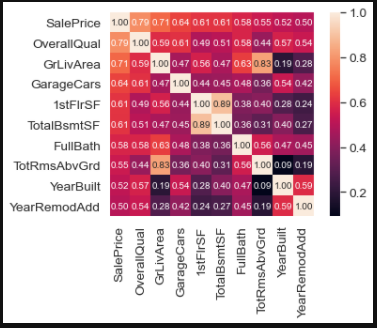
**House Pricing in the next 10 years:**

There are many factors that I think will affect residential house prices. I will split the reasons into two broad categories:

1. Category 1: These include the physical and meta-attributes (pertaining to the construction) of the house in question such as the living area, number of rooms, the size of the Garage, year built etc. These will largely show the same trend as they have in the past as the rationale behind why these factors affect house pricing will largely remain the same even in the future.
2. Category 2: These are factors that are governed by socio-economic norms including changes in policy, investor activity, the economy (buying capability of the public)

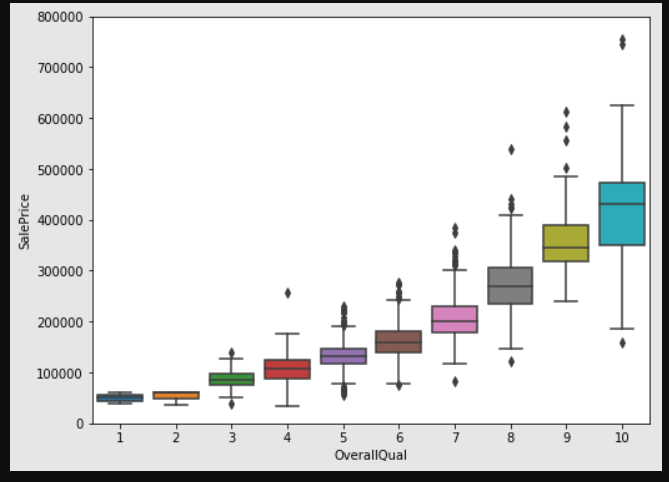
Category 1:

To fully understand the trends in variables pertaining the physical attributes of a house, I leveraged the house-pricing dataset of Ames, Iowa (as available on Kaggle). This dataset had data for 1470 houses with a whopping 79 features documented for each (barring missing data), a lot of which I deemed redundant after some exploratory analysis.

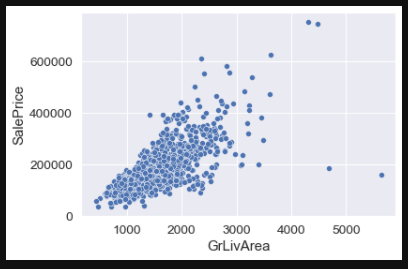


As a preliminary step, I used a correlation matrix to see which of the variables were closely related to the sale price. This revealed the following:

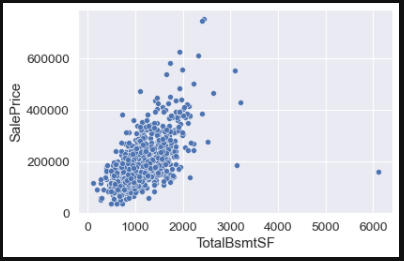
* Overall Quality showed high correlation with the price. Although this was a categorical variable, its classes indicated that there was a notion of order (with 1 being poor in quality and 10 being excellent), therefore to treat as a continuous variable just for the exploration made sense.



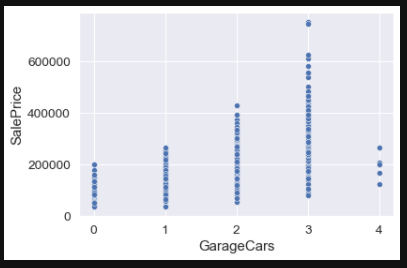
* The second was the general living area, which rendered the variables total number of rooms and first floor area (measured in square feet) redundant. The price generally increases with the living area, barring a few outliers, which were agricultural plots.



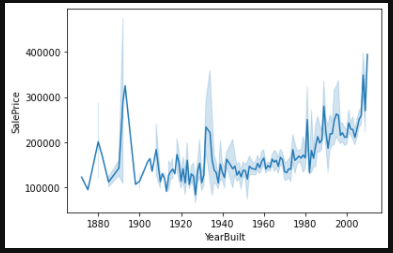
* The next was total area of basement (also renders the 1st floor area redundant). This too showed a strong positive linear trend with respect to Sale Price with a slight exponential curve towards very high values for this variable.



* Then was the number of cars that could be housed in the garage. This rendered the square feet area of Garage redundant. Although this is also a discrete variable, the fact that there is a notion of ordering, makes it acceptable to treat it as a continuous variable for the sake of analysis. Generally, the more cars that can fit in the garage, the more expensive the property. Although this doesn’t to a great job of separating the data into distinct clusters.



* Finally, the year it was built. This shows a general upward trend; that is to say, the more recent the construction was completed, the more expensive the property is.



It may be important to note that location did not have much impact on the price of the house, I believe that this is due to the fact that not much data was recorded for the same. The only variable that was included was amenities and most of the houses in the dataset shared the same ones (schools, recreational centers).

**Category 2:**

These include dynamic factors such as government policy, investor activity, rate of migration etc. These would change rapidly with time and predicting the behavior as they pertain to house pricing in the distant future is not feasible. These can include reasons such as tax laws as they pertain to property or comparable sales in neighborhoods. To leverage these as possible variables, a very small time frame must be taken into account at which point in time the predictions may be accurate and not obsolete.

