# **STORYTELLING W/R SHINY: IMPORTANCE OF CONTEXT & APPLICATION**

**ARVIND PAWAR**

**ASHISH SALUNKHE**

**NORTHEASTERN UNIVERSITY**

**ALY6070 COMMUNICATION AND VISUALIZATION FOR DATA ANALYTICS**

**INSTRUCTOR NAME- MYKHAYLO TRUBSKYY**

**DATE- 03/21/2019**

****

**INTRODUCTION**

In this assignment, we are given the house price data which contains the price of the properties based on various factors. We are asked to address a business problem, exploratory visualization and based on that provide solutions. We explored the various variables in the dataset and found out that this data is regarding the type of variables that determine the price of the property. We are focusing on variables such as SalePrice and Neighborhood to help the NextGen Real Estate company to help their customers to have a look at the rates of the properties of previous years of a particular neighborhood so that they can decide whether to invest in property in particular neighborhood or not.

**ANALYSIS**

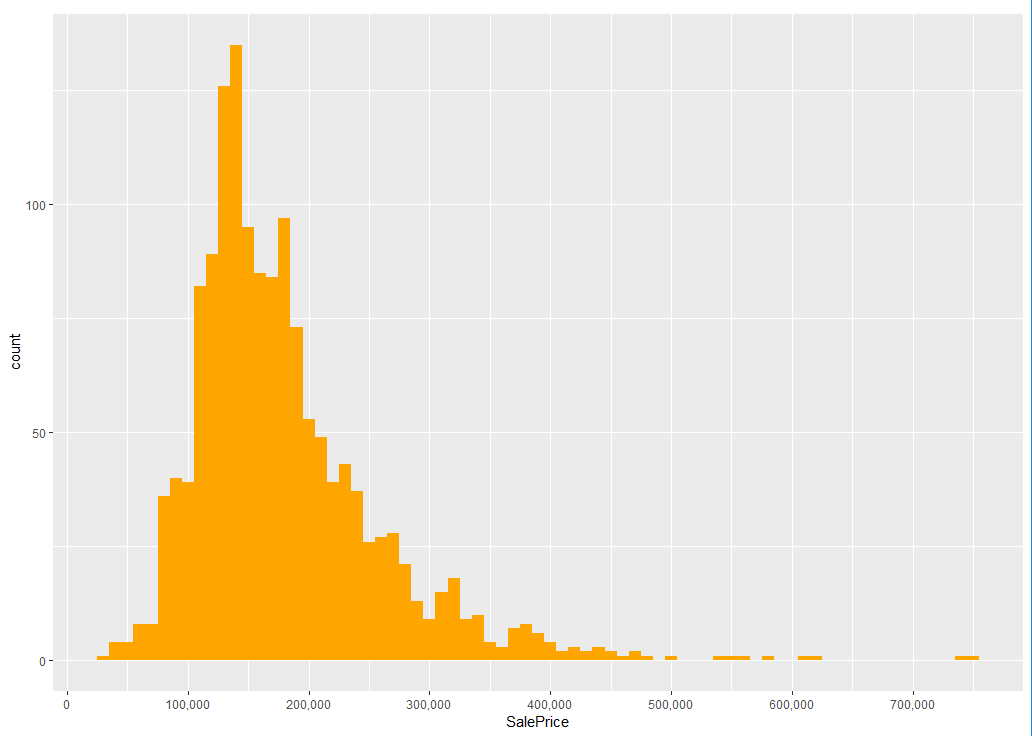


Fig. 1

After observing the variables, we tried to find out the distribution of sale price by using the ggplot. We can say that in the histogram shown above the sale prices are right-skewed. We find out that the maximum number of houses had a sale price ranging from $100,000 to $250,000. We found out that as the sale price increases the sale of a house decreased depending on the price. As few people were willing to buy houses with such a high sale price.



Fig. 2

As we are focusing on SalePrice, we tried to find out what all variables are affecting the SalePrice. So that this will help the Real Estate company to explain to their customers why the SalePrice of the Property Differ and what factors decide the price of the property. We tried to find out the correlation between the data variables by using the corrplot function, we found out that sale price was highly correlated with overall quality, garage area, and X1stflrSF.

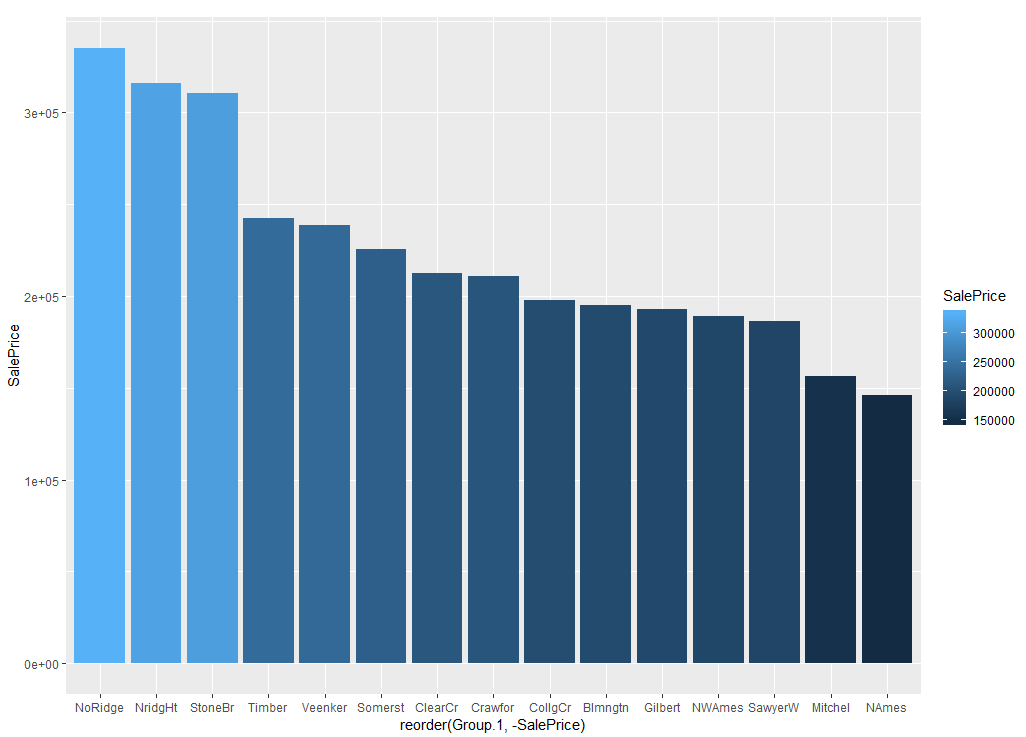


Fig. 3

Now, we want to focus on the neighborhoods, we want to check which neighborhood has the highest sales price. We used the bar plot to find out the top neighborhoods with the highest sale price. We found out that NoRidge, NridgHt, StoneBr were the top 3 neighborhoods with maximum SalePrice.

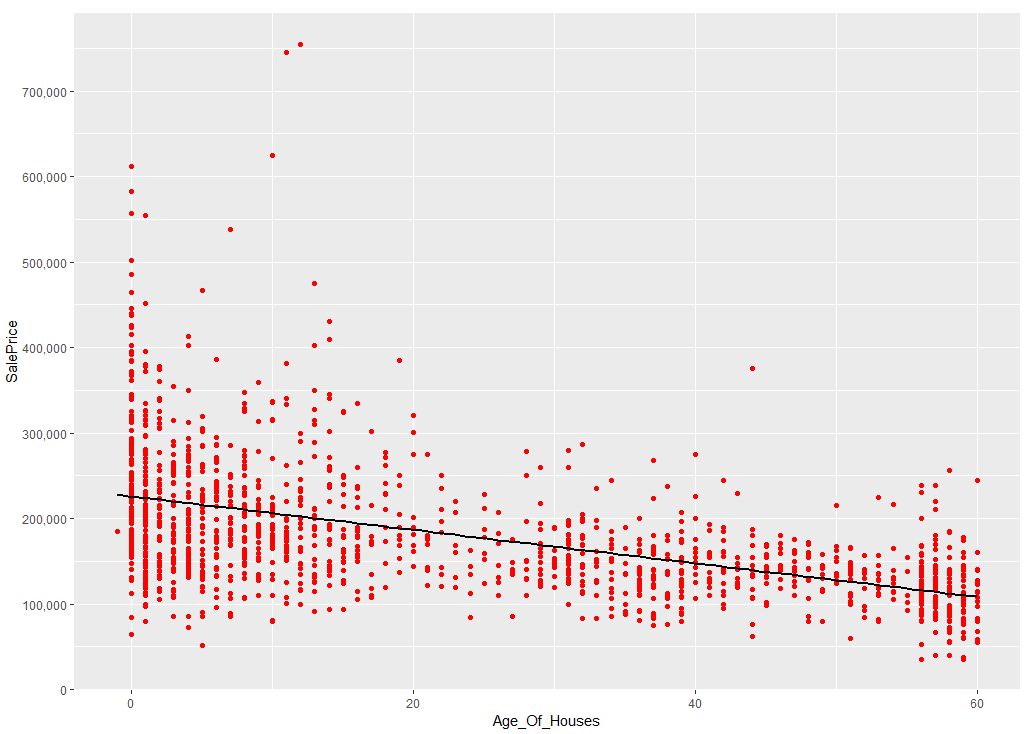


Fig. 4

In order to give a detailed insight into the property prices, we used a scatter plot and observed that newly constructed property have a high sale price compared to old properties.

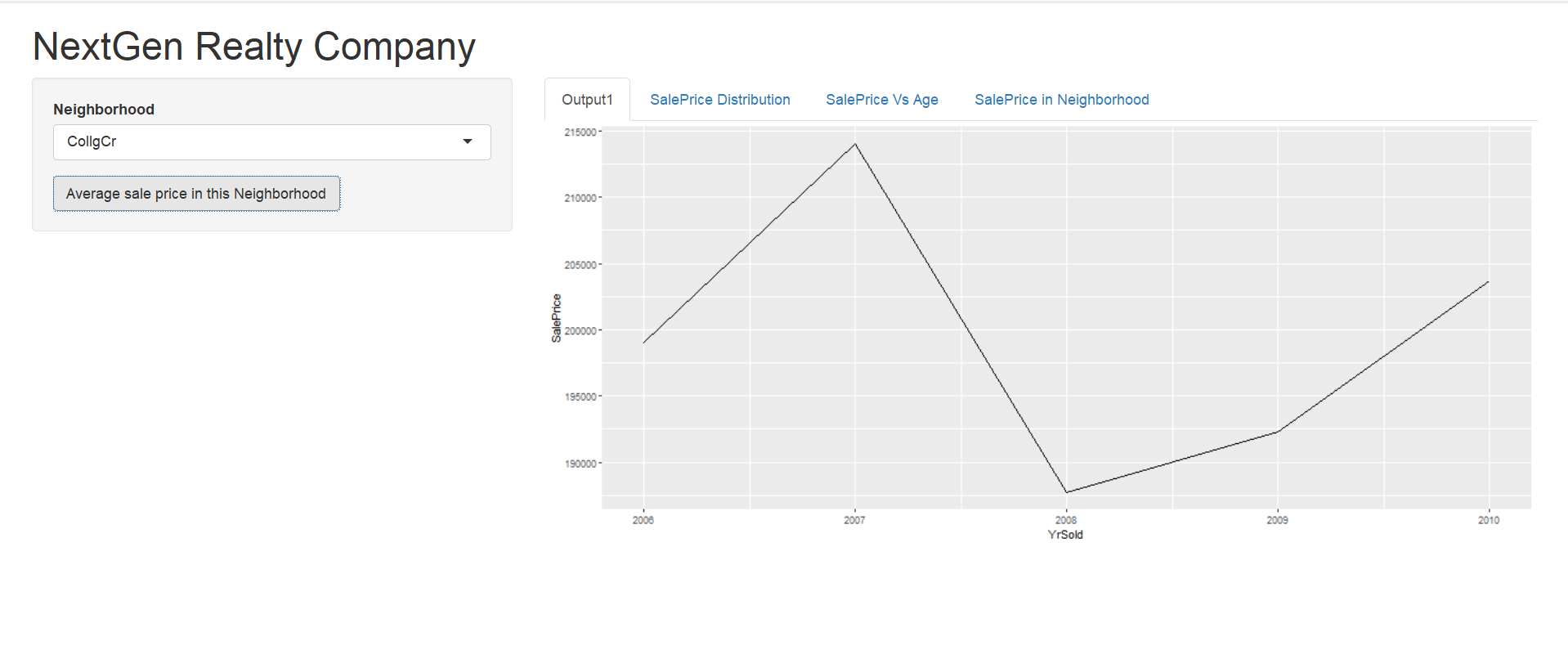


Fig. 5

After finding out about the top neighborhood properties, we tried to create a UI for the Real Estate company using R shiny, so that their customers will be able to decide to look at the prices of the neighborhood and decide if they want to invest in the neighborhood or not.

**CONCLUSION**

By analyzing the above visualization, we can say that SalePrice was highest in the newly constructed properties in the neighborhood. We observed the Top 3 Neighborhood with highest SalePrice. Finally, we created a UI using R shiny for the NExtGen Realty Company. This UI will help NextGen’s Customers to find out what SalePrice to expect from their desired Neighborhood. By looking at the previous prices of the Neighborhood in the previous years. The customers will decide whether they want to invest in that particular property or not.

**REFERENCES**

* R Overview. (2019). www.tutorialspoint.com. Retrieved 22 March 2019, from https://www.tutorialspoint.com/r/r\_overview.htm
* Bar Plot in R Using barplot() Function. (2017). DataMentor. Retrieved 22 March 2019, from https://www.datamentor.io/r-programming/bar-plot/
* Shiny - Interactive plots. (2019). Shiny.rstudio.com. Retrieved 22 March 2019, from https://shiny.rstudio.com/articles/plot-interaction.html