**Capstone Project 1: Exploratory Data Analysis**

1. Are there variables that are particularly significant in terms of explaining the answer to your project question?

Calculating the correlation between each feature and the SalePrice, we were able to select 10 features with the strongest correlation with SalePrice (Corellation >50%)

|  |  |  |
| --- | --- | --- |
|  | Features | Corellation |
| 0 | SalePrice | 100 |
| 1 | OverallQual | 79.09816 |
| 2 | GrLivArea | 70.86245 |
| 3 | GarageCars | 64.04092 |
| 4 | GarageArea | 62.34314 |
| 5 | TotalBsmtSF | 61.35806 |
| 6 | 1stFlrSF | 60.58522 |
| 7 | FullBath | 56.06638 |
| 8 | TotRmsAbvGrd | 53.37232 |
| 9 | YearBuilt | 52.28973 |
| 10 | YearRemodAdd | 50.7101 |

1. Are there strong correlations between pairs of independent variables or between an independent and a dependent variable?

Yes, besides correlation with dependent variable (sale price) there are some strong correlations pairs within independent variables. These pairs can be easily visualized from the Figure 1 heatmap. Some strong and logical examples include: GrLivArea and TotRmsAbvGr, GarageArea and GarageCar. Interestingly, BsmtUnSF negatively correlate with BsmtFinSF1.

1. What are the most appropriate tests to use to analyse these relationships?

To analyze these relationships, we can use regression analysis and calculate the slope of regression lines.

**Figure 1**. Heatmap representing correlation between each features 