Part 1 summary:

The first step is to analyse the data set. This involves sorting through each variable and producing information about each one like below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Type | Segment | Expectation | Conclusion | Comments |
| The name of the variable. | The type of variable. There are two types, numerical and categorical. | This identifies the variable’s segment. In this project this is between building, space, location. | This is the expectation of the influence on ‘Saleprice’ the variable has. | This is the decided importance of the variable after a quick look at the data. | Any general comments made. |

Expectation:

Is this variable thought about when buying a house?

Is this information described in another variable?

Type:

Numerical = variables which are numbers

Categorical = variables for which the values are categories

Each of these columns is important to deciding the importance of each variable and giving us a general idea and expectation about how the sales price will be affected.

Variables with the highest importance:

OverallQual

YearBuilt

TotalBsmtSF

GrLivArea