* change variables quality and condition to integers
* Compare if it is better to treat as factor or as continuous
* Normalize continuous variables
* Check R2 change for discrete variables
* Remove similar variables that might impact the analysis below.

Questions: which part of the house most increases their price with quality?

How does area and price correlate for different parts of the house?

How does the number of rooms affect price?

**Id**

Manter

**MSSubClass** - Identifies the type of dwelling involved in the sale.

Tratar como categórico. Ver no R2 se vale a pena ser preservado

**MSZoning -** Identifies the general zoning classification of the sale.

Tratar como categórico. Ver no R2 se vale a pena ser preservado

**LotFrontage** - Linear feet of street connected to property.

Tratar como continuo. Ver no R2 primário se vale a pena. Fazer trim dos outliers

**LotArea -** Lot size in square feet

Tratar como continuo. Ver no R2 primário se vale a pena. Fazer trim dos outliers

**Street -** Type of road access to property

Categórico. Apenas 6 dos 1500 estrada é de gravel. Ver no R2 a importância

**Alley -** Type of alley access to property

Categorico. Maior parte sem alley. Ver no R2 a importância

**LotShape-** General shape of property

Categorico. Agrupar como irregular versus regular. Ver R2

**LandContour -** Flatness of the property

Categorico. Ver R2

**Utilities** - Type of utilities available

Deletar: apenas 1 não está na categoria principal

**LotConfig -** Lot configuration

Categorico. Ver R2

**LandSlope -** Slope of property

Categorico. Ver R2

**Neighborhood -** Physical locations within Ames city limits

**Condition1 -** Proximity to various conditions

**Condition2 -** Proximity to various conditions (if more than one is present)

Create binary variables with both these variables. Group adjancet and within 200 together

**BldgType -** Type of dwelling

Categorico. Ver R2

**HouseStyle -** Style of dwelling

Categorico. Ver R2

**OverallQual -** Rates the overall material and finish of the house

Continous

**OverallCond -** Rates the overall condition of the house

Continous

**YearBuilt -** Original construction date

Continous

**YearRemodAdd -** Remodel date (same as construction date if no remodeling or additions)

Continous

**RoofStyle -** Type of roof

Categorico. Ver R2

**RoofMatl -** Roof material

Drop, 1400+ In single category

**Exterior1st -** Exterior covering on house

Categorico. Ver R2. Group in smaller groups

**Exterior2nd -** Exterior covering on house (if more than one material)

Categorico. Ver R2. Group in smaller groups. Drop?

**MasVnrType -** Masonry veneer type

Categorico. Ver R2. Group in smaller groups. Drop? (Obs: 8 NaN)

**MasVnrArea -** Masonry veneer area in square feet

Continuous. Create variable for values equal to 0.

**ExterQual -** Evaluates the quality of the material on the exterior

Change to continous

**ExterCond -** Evaluates the present condition of the material on the exterior

Change to continous

**Foundation -** Type of foundation

Categorico. Ver R2. Group in smaller groups.

**BsmtQual -** Evaluates the height of the basement

Change to continous

**BsmtCond -** Evaluates the general condition of the basement

Change to continuous. deal with na**.**

**BsmtExposure -** Refers to walkout or garden level walls

Categorical. See R2

**BsmtFinType1 -** Rating of basement finished area

Categorical. See R2

**BsmtFinSF1 -** Type 1 finished square feet

Continous

**BsmtFinType2 -** Rating of basement finished area (if multiple types)

Drop

**BsmtFinSF2 -** Type 2 finished square feet

Drop

**BsmtUnfSF -** Unfinished square feet of basement area

Continuous without 0

**TotalBsmtSF -** Total square feet of basement area

Continuous without 0

**Heating -** Type of heating

Drop. Too little values outside gas.

**HeatingQC -** Heating quality and condition

Change to continous

**CentralAir -** Central air conditioning

Binary data. Keep as is

**Electrical -** Electrical system

Change to Normal vs Other

**X1stFlrSF -** First Floor square feet

Continuous. Cut outliers

**X2ndFlrSF -** Second floor square feet

Continuous. Create var has\_2nd\_floor

**LowQualFinSF -** Low quality finished square feet (all floors)

Change to binary: has unfinished square feet (>0) or not (0)

**GrLivArea -** Above grade (ground) living area square feet

Continuous. Cut outliers

**BsmtFullBath -** Basement full bathrooms

Change to binary: 0 or more than 1

**BsmtHalfBath -** Basement half bathrooms

Change to binary: 0 or more than 1

**FullBath -** Full bathrooms above grade

Change to <=1, 2,3

**HalfBath -** Half baths above grade

Binary: 0 or >=1

**BedroomAbvGr -** Bedrooms above grade (does NOT include basement bedrooms)

Change to: <=2, 3, >=4

**KitchenAbvGr -** Kitchens above grade

Drop: 1392 had 1 kitchen. From the plots, it did not seem there was a huge difference.

**KitchenQual -** Kitchen quality

Change to continous

**TotRmsAbvGrd -** Total rooms above grade (does not include bathrooms)

Change to continous

**Functional -** Home functionality (Assume typical unless deductions are warranted)

Change to binary: normal vs others

**Fireplaces -** Number of fireplaces

Change to: 0, 1 >=2

**FireplaceQu -** Fireplace quality

Change to continuous: 1 to 5. If NaN, assign as 0.

**GarageType -** Garage location

Change to: Attchd, Builtin, Detchd, Others

**GarageYrBlt -** Year garage was built

Change to continnous

**GarageFinish -** Interior finish of the garage

Categorical. Keep as is

**GarageCars -** Size of garage in car capacity

Change to: 0, 1, 2, 3 or more

**GarageArea -** Size of garage in square feet

Continous

**GarageQual -** Garage quality

Change to continnous. Nan as 0

**GarageCond -** Garage condition

Change to continnous. Nan as 0

**PavedDrive -** Paved driveway

Categorical. Keep as is

**WoodDeckSF -** Wood deck area in square feet

Continuous. Create variable has\_wooddeck. Drop outliers

**OpenPorchSF -** Open porch area in square feet

Continuous. Create variable has\_wooddeck. Drop outliers

**EnclosedPorch -** Enclosed porch area in square feet

Drop. Not enough values of enclosedporch

**X3SsnPorch -** Three season porch area in square feet

Drop. Not enough values

**ScreenPorch -** Screen porch area in square feet

Drop. Not enough values

**PoolArea -** Pool area in square feet

Drop. Not enough values

**PoolQC -** Pool quality

Change to continuous. Value equal to NaN assigned as 0

**Fence -** Fence quality

Change to continuous. Value equal to NaN assigned as 0

**MiscFeature -** Miscellaneous feature not covered in other categories

Drop. Not enough values

**MiscVal -** Value of miscellaneous feature

Continuous. Test R2

**MoSold -** Month Sold (MM)

Categorical. Test R2

**YrSold -** Year Sold (YYYY)

Categorical. Test R2

**SaleType -** Type of sale

Categorical. Change: WD, New, Contract. Test R2

**SaleCondition -** Condition of sale

Categorical. Group category in others. Test R2

**SalePrice**