

**House Price Prediction**

Submitted by:

Pankaj Pagare

**ACKNOWLEDGMENT**

I would like to express my gratitude towards FlipRobo Technologies for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons and my mentor (Ms. Shristi Maan) for giving me such attention and time as and whenever required.

**INTRODUCTION**

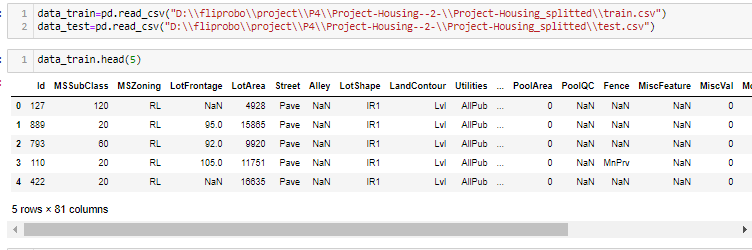
* **Business Problem Statement**
* A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia.
* The company is looking at prospective properties to buy houses to enter the market. You are required to build a model using Machine Learning in order to predict the actual value of the prospective properties and decide whether to invest in them or not. For this company wants to know:
* Which variables are important to predict the price of variable?
* How do these variables describe the price of the house?
* **Conceptual Background of the Domain Problem**
* Houses are one of the necessary needs of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market and there are various companies working in the domain.
* Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases.
* Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

**Review**

* We are required to model the price of houses with the available independent variables.
* This model will then be used by the management to understand how exactly the prices vary with the variables.
* They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high returns.
* Further, the model will be a good way for the management to understand the pricing dynamics of a new market.
* Motivation for the Problem Undertaken
* Having lived in India for so many years if there is one thing that I had been taking for granted, it’s those housing and rental prices continue to rise. Housing prices have recovered remarkably well, especially in major housing markets.
* So, to maintain the transparency among customers and also the comparison can be made easy through this model. If customer finds the price of house at some given website higher than the price predicted by the model, so he can reject that house.
* So we have to predict the pricing as per customers requirement and needs.

**Analytical Problem Framing**

* Dataset Representation:

****

**Observation:**

1. Seeing the data, we have to build a model which can be used to predict the SalePrice.
2. The data seems to be a combination of both numerical and categorical features.

**So clearly it is a regression problem.**

* **Data Sources and their formats & inferences**

MSSubClass: Identifies the type of dwelling involved in the sale.

20 1-STORY 1946 & NEWER ALL STYLES

30 1-STORY 1945 & OLDER

40 1-STORY W/FINISHED ATTIC ALL AGES

45 1-1/2 STORY - UNFINISHED ALL AGES

50 1-1/2 STORY FINISHED ALL AGES

60 2-STORY 1946 & NEWER

70 2-STORY 1945 & OLDER

75 2-1/2 STORY ALL AGES

80 SPLIT OR MULTI-LEVEL

85 SPLIT FOYER

90 DUPLEX - ALL STYLES AND AGES

120 1-STORY PUD (Planned Unit Development) - 1946 & NEWER

150 1-1/2 STORY PUD - ALL AGES

160 2-STORY PUD - 1946 & NEWER

180 PUD - MULTILEVEL - INCL SPLIT LEV/FOYER

190 2 FAMILY CONVERSION - ALL STYLES AND AGES

MSZoning: Identifies the general zoning classification of the sale.

A Agriculture

C Commercial

FV Floating Village Residential

I Industrial

RH Residential High Density

RL Residential Low Density

RP Residential Low Density Park

RM Residential Medium Density

LotFrontage: Linear feet of street connected to property

LotArea: Lot size in square feet

Street: Type of road access to property

Grvl Gravel

Pave Paved

Alley: Type of alley access to property

Grvl Gravel

Pave Paved

NA No alley access

LotShape: General shape of property

Reg Regular

IR1 Slightly irregular

IR2 Moderately Irregular

IR3 Irregular

LandContour: Flatness of the property

Lvl Near Flat/Level

Bnk Banked - Quick and significant rise from street grade to building

HLS Hillside - Significant slope from side to side

Low Depression

Utilities: Type of utilities available

AllPub All public Utilities (E,G,W,& S)

NoSewr Electricity, Gas, and Water (Septic Tank)

NoSeWa Electricity and Gas Only

ELO Electricity only

LotConfig: Lot configuration

Inside Inside lot

Corner Corner lot

CulDSac Cul-de-sac

FR2 Frontage on 2 sides of property

FR3 Frontage on 3 sides of property

LandSlope: Slope of property

Gtl Gentle slope

Mod Moderate Slope

Sev Severe Slope

Neighbourhood: Physical locations within Ames city limits

Blmngtn Bloomington Heights

Blueste Bluestem

BrDale Briardale

BrkSide Brookside

ClearCr Clear Creek

CollgCr College Creek

Crawfor Crawford

Edwards Edwards

Gilbert Gilbert

IDOTRR Iowa DOT and Rail Road

MeadowV Meadow Village

Mitchel Mitchell

Names North Ames

NoRidge Northridge

NPkVill Northpark Villa

NridgHt Northridge Heights

NAmes Northwest Ames

OldTown Old Town

SWISU South & West of Iowa State University

Sawyer Sawyer

SawyerW Sawyer West

Somerst Somerset

StoneBr Stone Brook

Timber Timberland

Veenker Veenker

Condition1: Proximity to various conditions

Artery Adjacent to arterial street

Feedr Adjacent to feeder street

Norm Normal

RRNn Within 200' of North-South Railroad

RRAn Adjacent to North-South Railroad

PosN Near positive off-site feature--park, greenbelt, etc.

PosA Adjacent to positive off-site feature

RRNe Within 200' of East-West Railroad

RRAe Adjacent to East-West Railroad

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

Artery Adjacent to arterial street

Feedr Adjacent to feeder street

Norm Normal

RRNn Within 200' of North-South Railroad

RRAn Adjacent to North-South Railroad

PosN Near positive off-site feature--park, greenbelt, etc.

PosA Adjacent to positive off-site feature

RRNe Within 200' of East-West Railroad

RRAe Adjacent to East-West Railroad

BldgType: Type of dwelling

1Fam Single-family Detached

2FmCon Two-family Conversion; originally built as one-family dwelling

Duplx Duplex

TwnhsE Townhouse End Unit

TwnhsI Townhouse Inside Unit

HouseStyle: Style of dwelling

1Story One story

1.5Fin One and one-half story: 2nd level finished

1.5Unf One and one-half story: 2nd level unfinished

2Story Two story

2.5Fin Two and one-half story: 2nd level finished

2.5Unf Two and one-half story: 2nd level unfinished

SFoyer Split Foyer

SLvl Split Level

OverallQual: Rates the overall material and finish of the house

10 Very Excellent

9 Excellent

8 Very Good

7 Good

6 Above Average

5 Average

4 Below Average

3 Fair

2 Poor

1 Very Poor

OverallCond: Rates the overall condition of the house

10 Very Excellent

9 Excellent

8 Very Good

7 Good

6 Above Average

5 Average

4 Below Average

3 Fair

2 Poor

1 Very Poor

YearBuilt: Original construction date

YearRemodAdd: Remodel date (same as construction date if no remodelling or additions)

RoofStyle: Type of roof

Flat Flat

Gable Gable

Gambrel Gabrel (Barn)

Hip Hip

Mansard Mansard

Shed Shed

RoofMatl: Roof material

ClyTile Clay or Tile

CompShg Standard (Composite) Shingle

Membran Membrane

Metal Metal

Roll Roll

Tar&Grv Gravel & Tar

WdShake Wood Shakes

WdShngl Wood Shingles

Exterior1st: Exterior covering on house

AsbShng Asbestos Shingles

AsphShn Asphalt Shingles

BrkComm Brick Common

BrkFace Brick Face

CBlock Cinder Block

CemntBd Cement Board

HdBoard Hard Board

ImStucc Imitation Stucco

MetalSd Metal Siding

Other Other

Plywood Plywood

PreCast PreCast

Stone Stone

Stucco Stucco

VinylSd Vinyl Siding

Wd Sdng Wood Siding

WdShing Wood Shingles

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

AsbShng Asbestos Shingles

AsphShn Asphalt Shingles

BrkComm Brick Common

BrkFace Brick Face

CBlock Cinder Block

CemntBd Cement Board

HdBoard Hard Board

ImStucc Imitation Stucco

MetalSd Metal Siding

Other Other

Plywood Plywood

PreCast PreCast

Stone Stone

Stucco Stucco

VinylSd Vinyl Siding

Wd Sdng Wood Siding

WdShing Wood Shingles

MasVnrType: Masonry veneer type

BrkCmn Brick Common

BrkFace Brick Face

CBlock Cinder Block

None None

Stone Stone

MasVnrArea: Masonry veneer area in square feet

ExterQual: Evaluates the quality of the material on the exterior

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

ExterCond: Evaluates the present condition of the material on the exterior

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

Foundation: Type of foundation

BrkTil Brick & Tile

CBlock Cinder Block

PConc Poured Contrete

Slab Slab

Stone Stone

Wood Wood

BsmtQual: Evaluates the height of the basement

Ex Excellent (100+ inches)

Gd Good (90-99 inches)

TA Typical (80-89 inches)

Fa Fair (70-79 inches)

Po Poor (<70 inches

NA No Basement

BsmtCond: Evaluates the general condition of the basement

Ex Excellent

Gd Good

TA Typical - slight dampness allowed

Fa Fair - dampness or some cracking or settling

Po Poor - Severe cracking, settling, or wetness

NA No Basement

BsmtExposure: Refers to walkout or garden level walls

Gd Good Exposure

Av Average Exposure (split levels or foyers typically score average or above)

Mn Mimimum Exposure

No No Exposure

NA No Basement

BsmtFinType1: Rating of basement finished area

GLQ Good Living Quarters

ALQ Average Living Quarters

BLQ Below Average Living Quarters

Rec Average Rec Room

LwQ Low Quality

Unf Unfinshed

NA No Basement

BsmtFinSF1: Type 1 finished square feet

BsmtFinType2: Rating of basement finished area (if multiple types)

GLQ Good Living Quarters

ALQ Average Living Quarters

BLQ Below Average Living Quarters

Rec Average Rec Room

LwQ Low Quality

Unf Unfinshed

NA No Basement

BsmtFinSF2: Type 2 finished square feet

BsmtUnfSF: Unfinished square feet of basement area

TotalBsmtSF: Total square feet of basement area

Heating: Type of heating

Floor Floor Furnace

GasA Gas forced warm air furnace

GasW Gas hot water or steam heat

Grav Gravity furnace

OthW Hot water or steam heat other than gas

Wall Wall furnace

HeatingQC: Heating quality and condition

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

CentralAir: Central air conditioning

N No

Y Yes

Electrical: Electrical system

SBrkr Standard Circuit Breakers & Romex

FuseA Fuse Box over 60 AMP and all Romex wiring (Average)

FuseF 60 AMP Fuse Box and mostly Romex wiring (Fair)

FuseP 60 AMP Fuse Box and mostly knob & tube wiring (poor)

Mix Mixed

1stFlrSF: First Floor square feet

2ndFlrSF: Second floor square feet

LowQualFinSF: Low quality finished square feet (all floors)

GrLivArea: Above grade (ground) living area square feet

BsmtFullBath: Basement full bathrooms

BsmtHalfBath: Basement half bathrooms

FullBath: Full bathrooms above grade

HalfBath: Half baths above grade

Bedroom: Bedrooms above grade (does NOT include basement bedrooms)

Kitchen: Kitchens above grade

KitchenQual: Kitchen quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

Functional: Home functionality (Assume typical unless deductions are warranted)

Typ Typical Functionality

Min1 Minor Deductions 1

Min2 Minor Deductions 2

Mod Moderate Deductions

Maj1 Major Deductions 1

Maj2 Major Deductions 2

Sev Severely Damaged

Sal Salvage only

Fireplaces: Number of fireplaces

FireplaceQu: Fireplace quality

Ex Excellent - Exceptional Masonry Fireplace

Gd Good - Masonry Fireplace in main level

TA Average - Prefabricated Fireplace in main living area or Masonry Fireplace in basement

Fa Fair - Prefabricated Fireplace in basement

Po Poor - Ben Franklin Stove

NA No Fireplace

GarageType: Garage location

2Types More than one type of garage

Attchd Attached to home

Basment Basement Garage

BuiltIn Built-In (Garage part of house - typically has room above garage)

CarPort Car Port

Detchd Detached from home

NA No Garage

GarageYrBlt: Year garage was built

GarageFinish: Interior finish of the garage

Fin Finished

RFn Rough Finished

Unf Unfinished

NA No Garage

GarageCars: Size of garage in car capacity

GarageArea: Size of garage in square feet

GarageQual: Garage quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

NA No Garage

GarageCond: Garage condition

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

NA No Garage

PavedDrive: Paved driveway

Y Paved

P Partial Pavement

N Dirt/Gravel

WoodDeckSF: Wood deck area in square feet

OpenPorchSF: Open porch area in square feet

EnclosedPorch: Enclosed porch area in square feet

3SsnPorch: Three season porch area in square feet

ScreenPorch: Screen porch area in square feet

PoolArea: Pool area in square feet

PoolQC: Pool quality

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

NA No Pool

Fence: Fence quality

GdPrv Good Privacy

MnPrv Minimum Privacy

GdWo Good Wood

MnWw Minimum Wood/Wire

NA No Fence

MiscFeature: Miscellaneous feature not covered in other categories

Elev Elevator

Gar2 2nd Garage (if not described in garage section)

Othr Other

Shed Shed (over 100 SF)

TenC Tennis Court

NA None

MiscVal: $Value of miscellaneous feature

MoSold: Month Sold (MM)

YrSold: Year Sold (YYYY)

SaleType: Type of sale

WD Warranty Deed - Conventional

CWD Warranty Deed - Cash

VWD Warranty Deed - VA Loan

New Home just constructed and sold

COD Court Officer Deed/Estate

Con Contract 15% Down payment regular terms

ConLw Contract Low Down payment and low interest

ConLI Contract Low Interest

ConLD Contract Low Down

Oth Other

SaleCondition: Condition of sale

Normal Normal Sale

Abnorml Abnormal Sale - trade, foreclosure, short sale

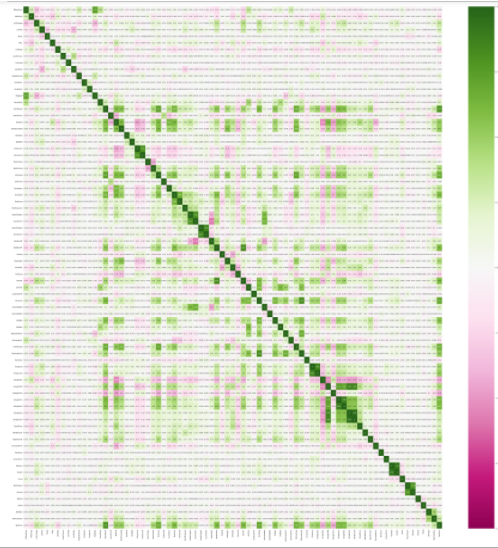
AdjLand Adjoining Land Purchase

Alloca Allocation - two linked properties with separate deeds, typically condo with a garage unit

Family Sale between family members

Partial Home was not completed when last assessed (associated with New Homes)

* **Data Pre-processing**
  1. Removed Null- values from the dataset by replacing null values with not available.
  2. Dropped Utilities and Id features as Id Consist for houses is unique and the utilities for each house were same.
  3. GarageBuilt this type feature has null values so we replaced it by 1800 because such null values are because of no garage in the houses and replaced by 1800 so that no- high variance will occur in the dataset.
  4. LotFrontage: This feature is of object as there were some methods in it because of which it transformed to object type so we replaced it by its mean.
  5. We transformed Categorical values into Integer type by replacing for some features because there some features whose values was not in the test dataset while it was in train dataset. And for some features that were telling about condition and quality so I replace it by their respective no. in ordered way like Excellent we give 10 while for poor 0. Otherwise for all categorical feature I used Label Encoder technique.
* **Data Inputs- Logic- Output Relationships**
* **Multi-Colinearity :**

****

**Observations:**

* Exterior1st & Exterior2nd ---> 80% collinear to each other.
* GrLivArea & TotRmsAbvGrd are 82% collinear to each other.
* TotalBsmtSF&1stFlrSF--->81% collinear with each other.
* BsmtFinSF2&BsmtFinType2 ----> -81% collinear.
* BsmtFinSF1&BsmtFinType1 ----> -73% collinear.
* GarageCars & GarageArea ----> 88% collinear to each other.
* FirePlace&FireQual are 72% collinear to each other.
* MiscFeature & MiscVal are 78% collinear to each other.
* PoolArea & PoolQC are collinear about -93%.
* Assumption for the problem:

So clearly it is a Regression problem.

* Hardware and Software Requirements and Tools Used

Software Used:

* + Jupyter Notebook
  + Ms-Paint
  + MS-PowerPoint
  + MS-Word

Hardware used:

* + Laptop
  + Good internet connectivity

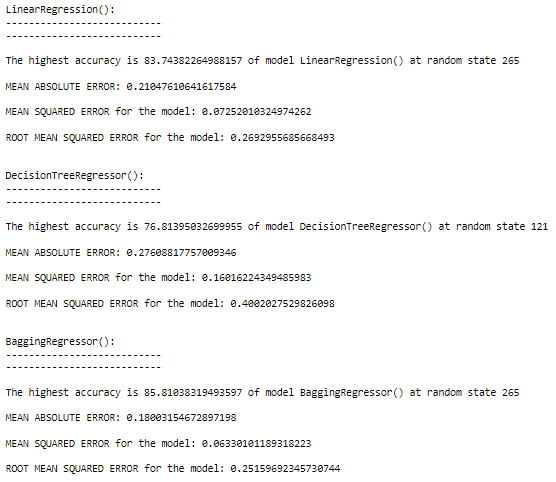
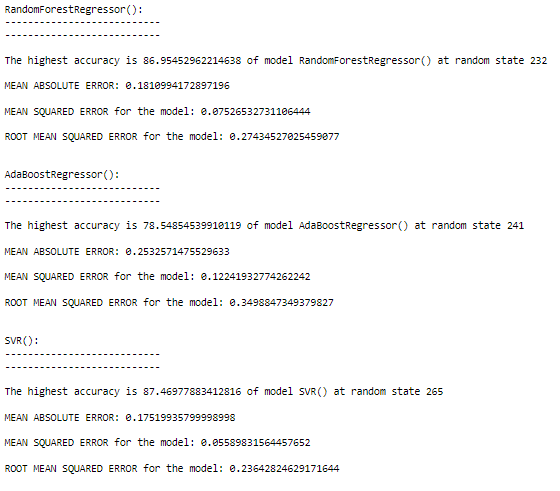
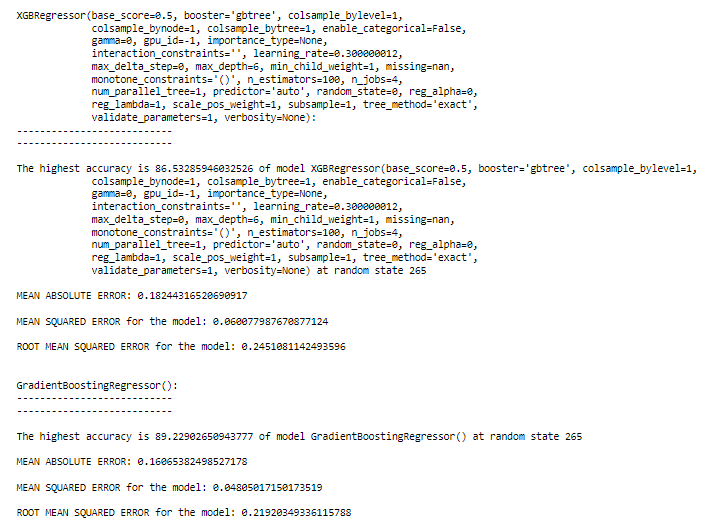
**Model/s Development and Evaluation**

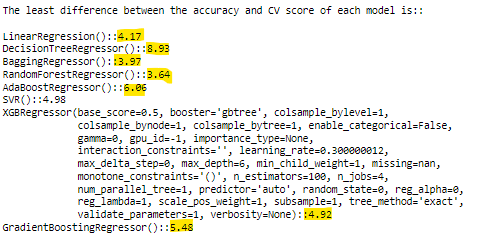
* Identification of possible problem-solving approaches (methods)
* Used PCA i.e., Principal Component Analysis for feature Selection.
* Reduced Skewness from the dataset.
* Reduced Outliers using Z-Score technique
* Dropped columns on the basis multi-Collinearity
* Used Square root for some feature to reduce the variance.
* **Testing of Identified Approaches (Algorithms)**
* Used 8 models that are:
  + Linear Regression
  + Decision Tree
  + Random Forest
  + Gradient Boosting
  + AdaBoost
  + Bagging
  + Support Vector Machine
  + Xtreme Gradient Boost
* Run and Evaluate selected models

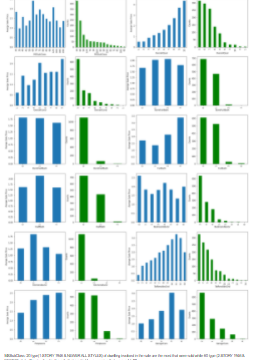
**Code:**

****

**Output:**

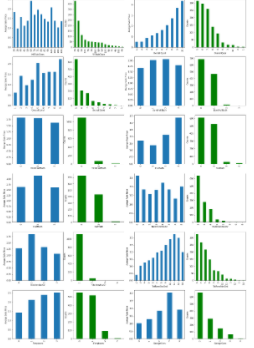
**** **** 

* Key Metrics for success in solving problem under consideration
* As we get the r2 score for each model but we didn’t know the finalized model.
* So, we find CV Score for each model with k-fold CV between the 3 to 9.
* Best cv score for each model with the best cv we gathered.
* After gathering cv score for each model we get difference between Cv Score and r2 score for each model
* The least difference for the model Random Forest we get 3.64.
* 
* Visualizations



Observations:

* + MSZoning: Floating Village Residential (FV) has highest average Sale Price of the houses and RL (Residential Low Density) types of Zoning has highest types of houses and very low amount of houses for commercial(C) types of Zoning.
  + Street: Paved Type of road access to property has highest average Sale Prices for the houses and Low amount of houses are there that are Gravel Type of road access to property.
  + Alley: No alley access types of such houses are very common and their prices for Sale is highest.
  + Lot Shape : Not much info we gathered from Lo Shape but Moderate type of irregularities has highest averge sale Price while Irregular type of records are not much available in our dataset.
  + LandContour: HLS(Hillside - Significant slope from side to side) has highest Sales Price while Flat type buildings has higher o. of records in the dataset.
  + LotConfig: Cul-de-sac & Frontage on 3 sides of property (CullDSac & Fr3) having highest Sale Prices for the houses while such type of houses is rarely available.
  + LandSlope: No such impact n the SalePrice for the houses while Gentle Slope (Gtl) has highest no. of records among them.
  + Neighbourhood: North Ames (NAmes) such type houses are mostly available.
  + Condition1: It shows the Proximity to various conditions in which we found that RRNn & PosA (Within 200' of North-South Railroad and Adjacent to positive off-site feature) such type has highest average Sale Prices for the houses. While Norm (NOrmal0 type of houses are easily available for Sale.
  + Condition2: It is Proximity to various conditions in which PosA (Adjacent to positive off-site feature) types has highest SalePrice while Norm i.e., Normal proximity to various condition such type are easily available in Sale.
  + Bldg Type: it is type of dwelling in the house in which Single-family Detached & Townhouse End Unit (1Farm & TwnhsE) has highest SalePrice and 2FmCon type are rarely found to Sale.
  + HouseStyle: Two and one-half story: 2nd level finished (2.5 Fin) such houses are rarely available to sale and has highest prices for Sale.
  + RoofStyle: Shed type of roof in the houses are very rarely available for sale and are expensive.
  + RoofMatl: Wood Shingles type of material used for roofs in the houses are rarely available and such houses are expensive.
  + Exterior1st: Exterior covering on house in that Imitation Stucco of exterior used in the houses are rarely available and such houses are costly.
  + Exterior2nd: Other type Exterior covering on house are rarely available in the sale are costly such houses.
  + MasVnrType: Stone type Masonry veneer type of houses are expensive and Brick Common type of masonry veneer type of houses are rarely available in the Sales.
  + ExterQual: Evaluates the quality of the material on the exterior in which Excellent quality of houses are raely available and are expensive ones in the Sale.
  + ExterCond: Evaluates the present condition of the material on the exterior in which Excellent quality of houses are rarely available and are expensive ones in the Sale.
  + Foundation: Poured Contrete Type of foundation are costly in Sale of houses while wood type are rarely available in Sale of Houses.
  + BsmtQual: Excellent Quality of basement in the houses are expensive while fair type of basement quality are rarely available in the sale of houses.
  + BsmtCond: The basement condition is good then such houses are exensive while Poor type of basement Condition are rare.
  + Basement Exposure: The exposure of the basement is better their prices for sale is better while No basement in houses are very easy eaily vailable in the sale.
  + BsmtFinType1: Better the Living Quarters better the sale Prices for the houses and there bery less houses that doen't have basement in it.
  + BsmtFinType2: Better the Living Quarters better the sale Prices for the houses and mostly houses have unfinished basement in it.
  + Heating: GasA(Gas forced warm air furnace) type of heating in the houses are mostly in the houses and such type of heating container has high sale Price.
  + HeatingQC: Better the heating Quality and its conditions better the Sale Price for the houses.
  + CentralAir : Mostly houses has Central Air in it and are Costly.
  + Electrical: Mostly houses has Electical system of SBrkr(Standard Circuit Breakers & Romex) in it and are Costly.
  + KitchenQual: Better the quality of kitchen of the houses better their Sale Price.
  + Functional: Home functionality Typ(typical) type are in most of houses and such houses have high salePrice.
  + FirePlaceQual: Better the quality for fire place having higher Sale Prices for such houses.
  + GarageType: Builtin Garage types of houses are expensive in Sale Price and 2Types are rarely available for sale.
  + Garage finish: better the finishing of garagee better the Sale Price and vary rare houses that has not Garage in it.
  + Garage Qual: Better the quality of garage in the houses higher the prices of houses. And mostly houses their garage quality is typical/Average.
  + GarageCond: Good and Typical/Average type of garage condition in the houses has hhigher SalePrice of the houses.
  + Paved drive: Paved(Y) driveway has highest average Sale Price of the houses and mostly founded during Sale of houses.
  + PoolQC: Better the qaulity and conditions for the pool of the houses better the Sale Price of the houses. Most of houses hasn't have pool in it.
  + Fence: Fence is not impacting much to Sle Price for the houses but mostly houses doen't have fencing in it.
  + MiscFeature: Miscellaneous feature not covered in other categories in which houses Tennis Court is available such houss has higher Sale Pice. Mostly houses hases has no ther miscellaneaous features in it.
  + SaleType: con(Contract 15% Down payment regular terms) and New(Home just constructed and sold) such types of sales has highest SalePrices. mostly houses are those whose Sale type WD (Warranty Deed - Conventional).
  + Sale COndition: Partial [Home was not completed when last assessed (associated with New Homes)] such condition of Sales are having High Sale prices for the houses. While most of houses has normal Sale Condition.



Observations:

MSSubClass: 20 type(1-STORY 1946 & NEWER ALL STYLES) of dwelling involved in the sale are the most that were sold while 60 type (2-STORY 1946 & NEWER) of dwelling involved in the sale that are highly expensive that were sold.

OverallQual: As in data description there we set a no. from 1 to 10 that specifies the quality so in accordance with that higher the quality higher the price for sale. And there are very less no. for records for quality of houses 1 7 2 that are very poor/ poor

OverallCond: Here I analyse that the average overall condition of houses i.e. 5 were more in sale from both graph as beteer the conditions of houses their prices for sale in increasing.

BsmFullBath: there are zero no. of full basement bathrrom records are high from both for these grapd i aalysed that more the no. of basement Full bathrooms more the seles price for the house.

BsmtHalfBath: from both graphs there is ionly analyse that there are more no. of houss that has no half bathroom in basement.

FullBath: Increased in the no. of full bathrooms will result in the increase in sale price for the houses and there are very less records for the the O & 3 no. of bathrooms.

HalfBath: there are very no. of houses that has 2 no. of Half bathrooms but the average sale price for having only 1 halfbathroom is maximum.

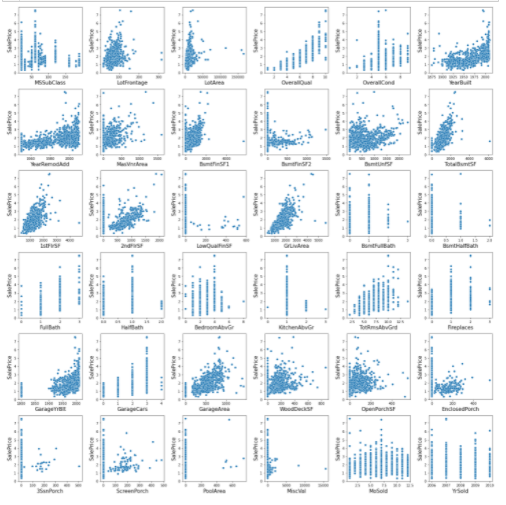
BedrooomAbvGr: there no such impacts for the no. of bedroooms to the salePrice but there very less no of records for 6,0,8 no. of bedroooms that lead model to underfiited.

KitchenAbvGr: very less records for the houses having 3 or no kitchen above gorund and the sale price are high for having only 1 kitchen in the house.

TotRmsAbvGrd: here we analyse that the salePrice is increasing w.r.t. increase in the no. of rooms above ground while there are very less records for having 2 & 14 no. of rooms in the house.

Fireplaces: Increased no. of fireplaces is impacting the increase in the SalePrice of the houses.

GarageCars: More No. of cars capacity in the house more its sale Price.



Observations:

LotFrontage: Linear feet of street connected to property is somehow seems that there is some linear relation between them.

LotArea: Not so specific but the lot area somehow seems like to be very slight increase in the area leads to extra amount of change in increasing manner of SalesPrice.

MasVnArea: Masonry veneer area except zero area there is linear reltions that tells that increase in the masonary ares lead to increase the sales Price of the houses.

BsmtFinSF1: Basement finished Type 1 rather than zero square feet there is some min amount of increase in the area there is good amount of increase in SalePrice of House.

BsmtFinSF2: Not much impacts it showing to the target variable.

BsmtUnfSf: Not much impacts it showing to the target variable.

TotalBsmtSF: Shows that there is increase in the total areas for the basement in the house leads to increase in the SalePrice.

1stFlrSF: In this more area in on the 1st Floor of houses will impact to increase in the sale price of the houses.

2ndFlrSF: In this more area in on the 2nd Floor of houses will impact to increase in the sale price of the houses.

LowQualFinSF: Not much impacts it showing to the target variable.

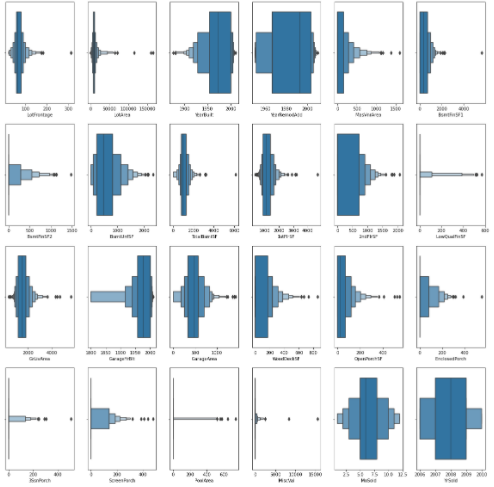
GrLivArea: In this more increase in above grade (ground) living area will impact to increase in the sale price of the houses.

GarageArea: Except with the 0 area there is some linear relation that stats that increase in the Garage area increase the Sale price of the House.

WoodDeckSF: Except with the 0 area sq. ft. there is some linear relation that stats that increase in the Wood Deck area increase the Sale price of the House. And seems to be outliers in it.

OpenPorchSF:Except with the 0 area sq. ft. there is some linear relation that stats that increase in the open Porch area increase the Sale price of the House. And seems to be outliers in it.

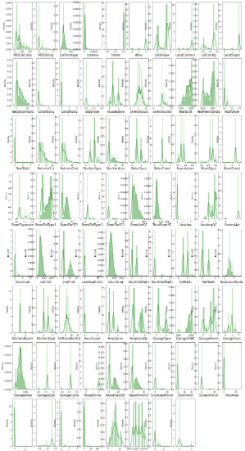
Enclosed Porch, 3SnPorch,ScreenPorch, PoolArea,MiscVal these area very less founded in the houses due to which not much info we gathered from graph.

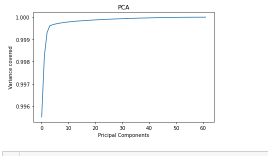


**Observations:**

LotArea, BsmtFinSF1, BsmtFinSF2, TotalBsmtSF, 1stFlrSF, GrLivArea, EnclosedPorch, 3SsnPorch, MiscVal.

These features consisting outliers in it.





Observations:

From above figure we analyze that after selecting 20 feature there is should be very minute variance that lead to predict the label.

* Interpretation of the Results
* Firstly, we removed null values by their respective values with their respective features.
* We know that the over all quality and Condition, the area for different type like for basement, floors, features in the houses and their conditions, Sale type and its Conditions, Garages areas, No. Of cars parking in that houses, remodelling of the houses, etc such types of features are impacting more to predict the sale price for the houses.
* We get to know about the problem existence like outliers, skewness, high- variance, Corelation of each feature with the target variable.
* Transformed Columns with their respective values or by the Label Encoder techniques.
* Used PCA for feature selection
* Find the best r2 score with their respective random state.
* Selected Random Forest as the finalized model.
* Performed Hyper tunning for the best model.
* Predicted values for the test dataset by Random Forest default parameterized model.
* Save The Model.

**CONCLUSION**

* Learning Outcomes of the Study in respect of Data Science
* Our customers requirements are our highest priority so the project was built to satisfy their needs so the project works well and there is no customer churn
* We should maintain the transparency among customers and also the comparison can be made easy through this model. If customer finds the price of house at some given website higher than the price predicted by the model, so he can reject that house.
* So we have to predict the pricing as per customers requirement and needs.
* Limitations of this work and Scope for Future Work
* This model will then be used by the management to understand how exactly the prices vary with the variables.
* They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high returns.
* Further, the model will be a good way for the management to understand the pricing dynamics of a new market.
* But still customers are always comparing the prices hence we should keep on updating our project to meet their necessity.

--------------------------------------------------------------

**Thank You!!**