

# HOUSING PRICES- ADVANCED REGRESSION TECHNIQUES

CAPSTONE 1 PROJECT

# PROBLEM

## **BASIS FOR ESTIMATING SALE PRICE OF A HOUSE**

The price estimation can be based on few factors or external sources such as real estate agencies. The problem for the buyer is knowing the exact amount for the purchase price of the house.

For a real estate company, which can also pose as a buyer or broker, the problem is to negotiate for the best deal.

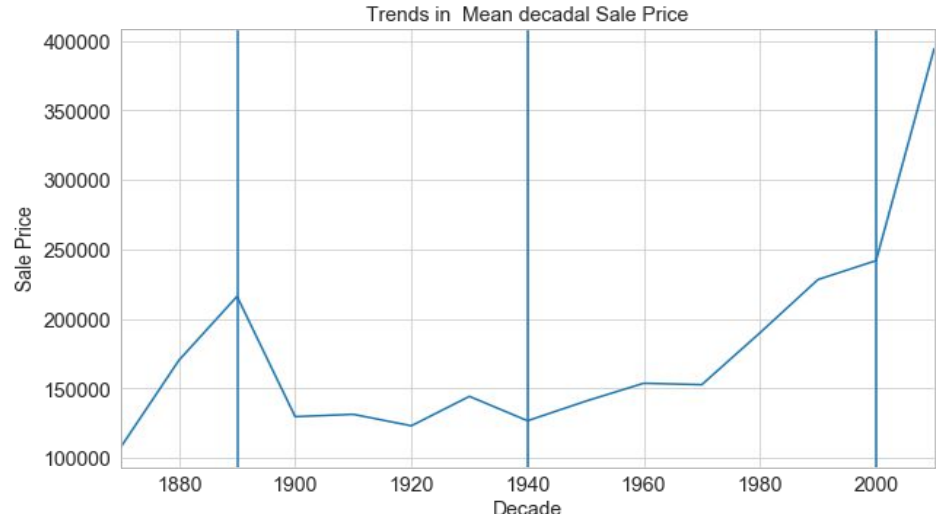
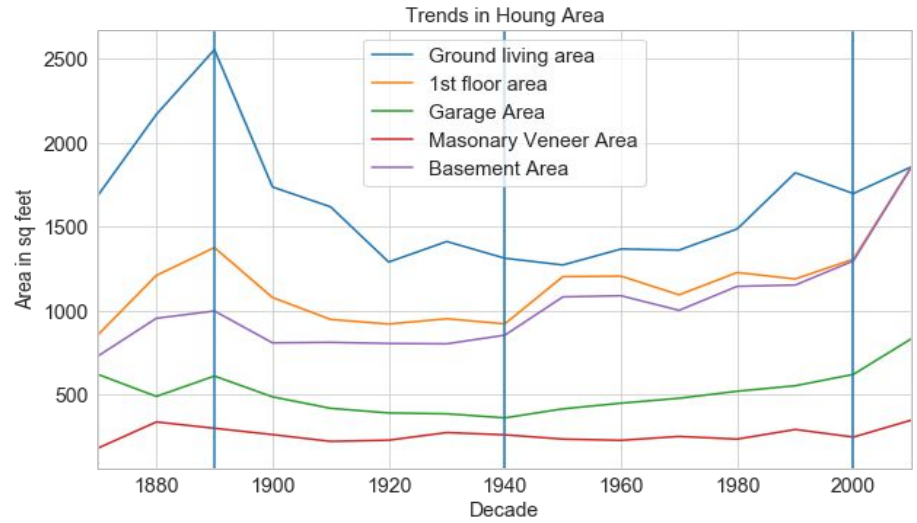
This dataset has several factors.

It becomes crucial to know the levers that drive the price and develop a model to predict them with best accuracy.

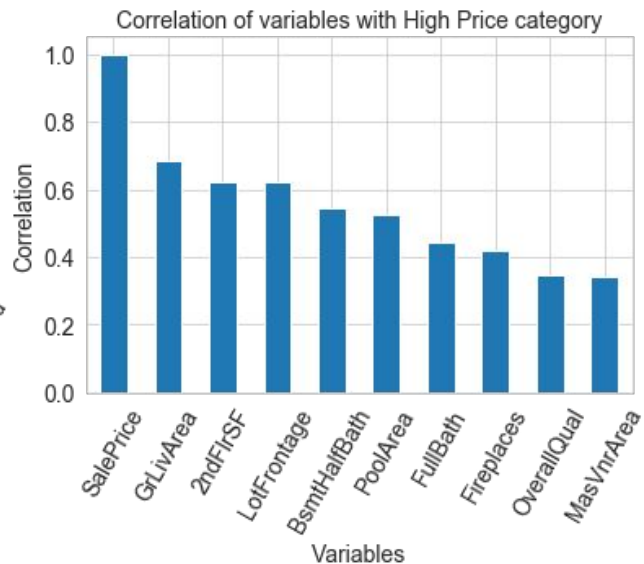
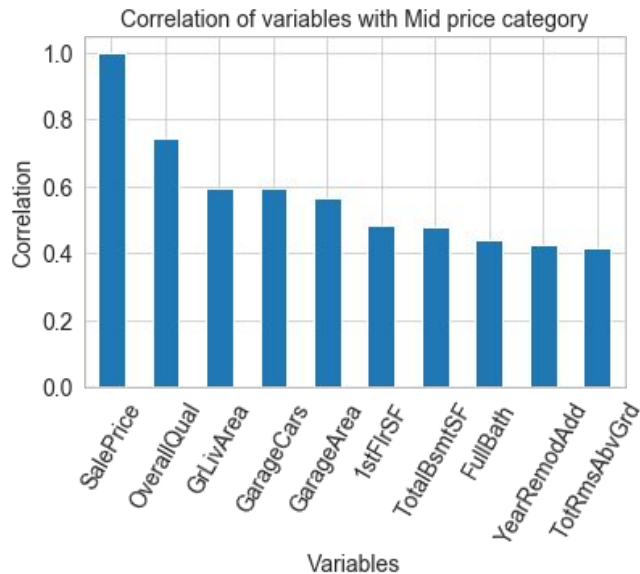
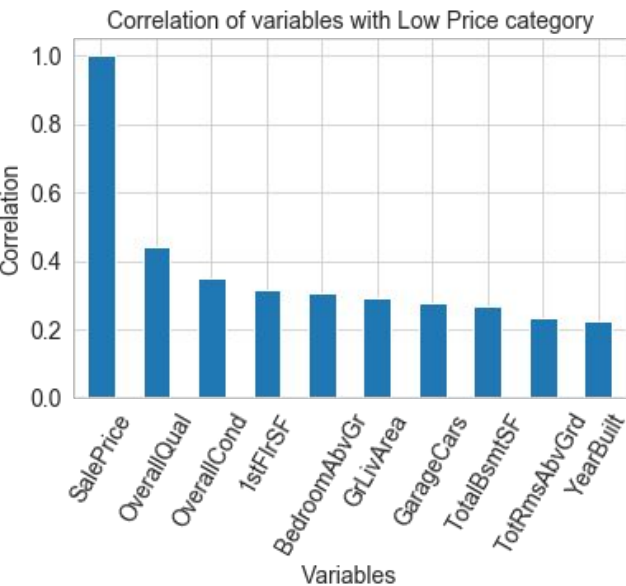
# Trends in Housing Area

Observations: We can see 3 phases: an upward phase(till 1890), a downward and stable phase(till 1940) and an upward again (from 1940)

In 1890 the avg area of houses were big. Avg Ground living area in 1890 were biggest, which we don't see today. We see a downward to a more stable phase till 1940. Increase in avg areas take place from 1940 with steep increase from 2000 onwards. Masonary Veneer Area shows development from 1920's.



# Correlation of variables with different Sale price ranges



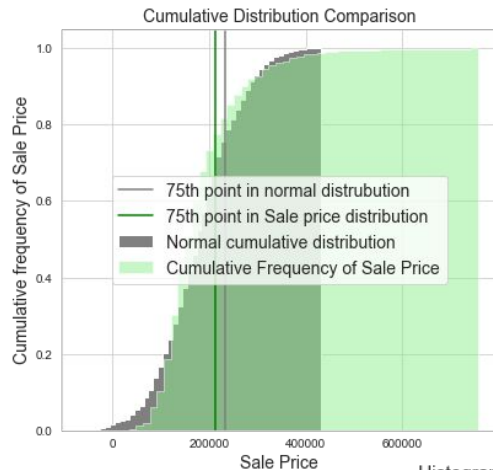
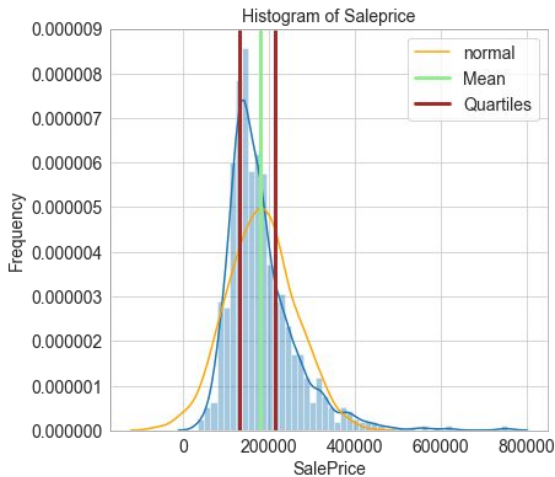
**What price can be expected with these numeric features?**

	Low Sale Price	Mid Sale Price	High Sale Price
<b>SalePrice</b>	106539	200382	512751
<b>1stFlrSF</b>	912	1234	2013
<b>BsmtFinSF1</b>	541	682	1292
<b>BsmtUnfSF</b>	566	634	813
<b>GarageArea</b>	395	530	842
<b>GrLivArea</b>	1120	1628	2863
<b>LotArea</b>	7724	11349	18066
<b>TotalBsmtSF</b>	842	1152	2016
<b>WoodDeckSF</b>	177	198	222

**What price can be expected with these categorical features?**

	Low Sale Price	Mid Sale Price	High Sale Price
<b>BedroomAbvGr</b>	6	8	4
<b>BsmtFullBath</b>	2	3	1
<b>Condition2</b>	RRNn	RRAn	Norm
<b>Foundation</b>	Stone	Wood	PConc
<b>Heating</b>	Wall	GasW	GasA
<b>KitchenAbvGr</b>	3	2	1
<b>Neighborhood</b>	SawyerW	Veenker	StoneBr
<b>PoolArea</b>	0	738	555
<b>PoolQC</b>	Absent	Gd	Ex
<b>RoofStyle</b>	Mansard	Shed	Hip
<b>TotRmsAbvGrd</b>	11	14	12

# Log transformation of Target Variable 'SalePrice'

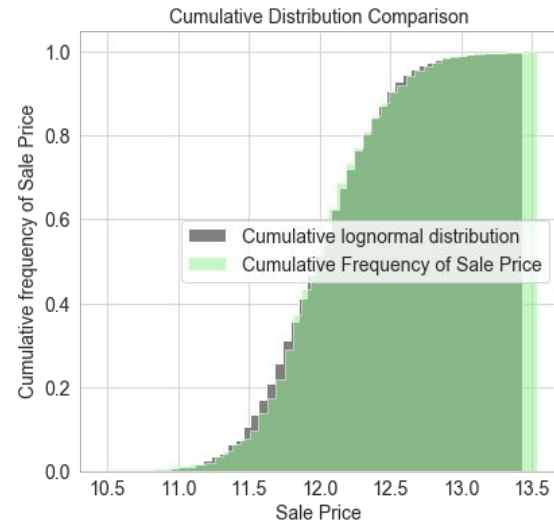
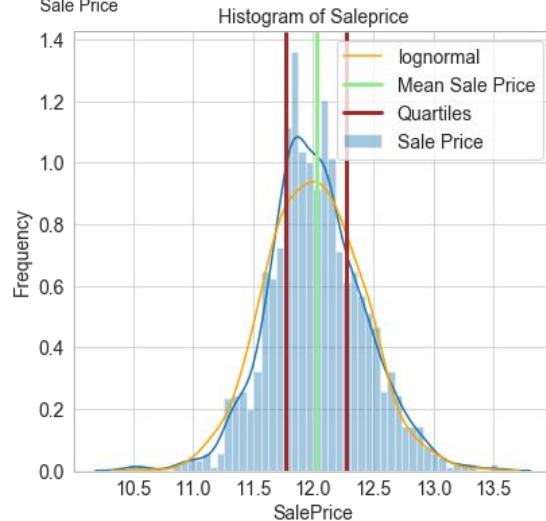


Observations:

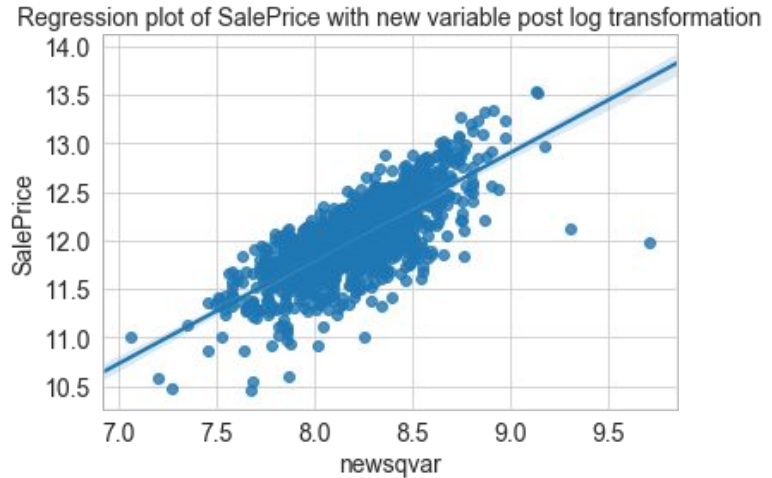
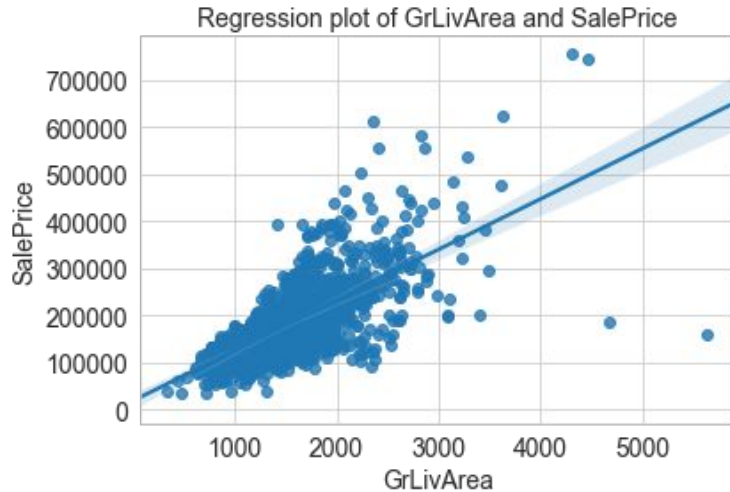
1. The distribution is not normal.
2. Distribution of SalePrice is leptokurtic.
3. The distribution is right skewed.
4. Mean Sale Price is not a good representation and there are quite a number of outliers.

Observations:

The distribution of 'SalePrice' is very close to lognormal distribution.  
The tails are matching, though 'SalePrice' appears to be bimodal.  
The range above the upper quartile has normalised to quite an extent.



# Independent variable GrLivArea and its transformation:

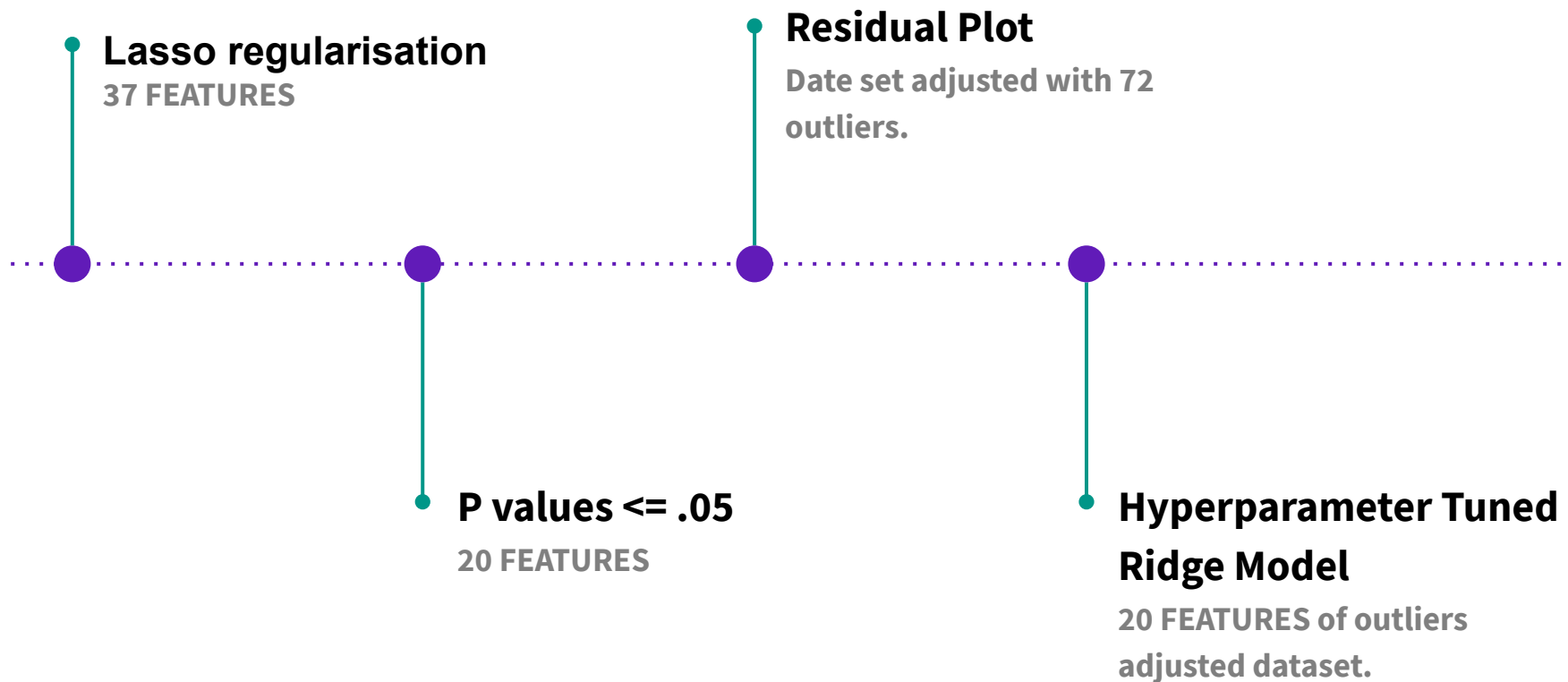


We create a new variable- 'newsqvar' which is a combination of 'TotalBsmtSF', '1stFlrSF' and 'GrLivArea'. This new variable has a strong linear relation with SalePrice and a constant variance as well. The correlation also has improved from .70 to .76.

## Transforming labels of categorical variables.

Basis our observations we find that there are categories, sensitive to Average Sale Price. For including these variables in the prediction models we need to assign them numerical labels. Here, have assigned the labels based on average SalePrice.

# FEATURE SELECTION - RIDGE MODEL





# RIDGE MODEL- SCORE & FEATURES

## SCORE

(RMSE on Train set is Cross validated )	Root Mean Squared Error between log of Sale Price and log of predictions	
	TRAIN SET	TEST SET
LINEAR REGRESSION MODEL	0.096	0.12
RIDGE MODEL	0.112	0.08

## SELECTED FEATURES

['newsqvar', 'OverallQual', 'Neighborhood', 'HalfBath',  
'CentralAir', 'Functional', 'YearRemodAdd', 'BsmtExposure',  
'MSZoning', 'TotRmsAbvGrd', 'KitchenQual', 'GarageCars',  
'Condition1', 'LotArea', 'HeatingQC', 'FullBath',  
'SaleCondition', 'WoodDeckSF', 'BedroomAbvGr', 'BsmtUnfSF']

# RANDOM FOREST MODEL- SCORE & FEATUERS

## SCORE

<i>(RMSE on Train set is Cross validated )</i>	Root Mean Squared Error between log of Sale Price	
	TRAIN SET	TEST SET
RANDOM FOREST MODEL	0.14	0.086

## SELECTED FEATURES

Newsqvar	Neighborhood	OverallQual	YearBuilt	ExterQual
GarageYrBlt	BsmtQual	KitchenQual	GarageCars	YearRemodAdd
FireplaceQu	GarageArea	FullBath	GarageType	GarageFinish
Fireplaces	BsmtFinSF1	MSSubClass	Foundation	LotArea

# Contact

## Your Name

aroonima12@gmail.com