

HOUSING: PRICE PREDICTION

Submitted by:

Chitrak Satyarthi

**ACKNOWLEDGMENT**

I would like to express my special thanks of gratitude to my mentor as well as the company for giving me the golden opportunity to do this wonderful project on the topic **HOUSING: PRICE PREDICTION**, which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.

**INTRODUCTION**

* Business Problem Framing

Houses are one of the necessary need 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. 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.

Whenever a company is investing for business in new country or even in a new city this is all the analysis which is required to make for higher profit.

* Conceptual Background of the Domain Problem

There are few basic keywords in real estate which will be easy to understand the problem statement.

Realtor, Realtors, finding a Realtor, how to find a Realtor, find real estate agents, how to find a real estate agent, best Realtor near \_\_\_\_\_\_\_\_, top realtor for buying a new home, top real estate agents in \_\_\_\_\_\_\_\_\_\_\_, selling a home, sell a home fast, fastest way to sell a home in \_\_\_\_\_\_\_\_, home selling tips, cost of selling your home, marketing your home, ways to sell your home, house staging tips, how to stage your home, foreclosure, foreclosures, short sales, foreclosure or short sale, short selling, short sale process, cons of a short sale, short selling your home, for sale by owner, fsbo, why to use a Realtor, property for sale by owner, for sale by owner listings, listing property for sale by owner, selling your home by owner, selling real estate without a Realtor, tips for selling your home, short selling your home, top tips to get the best offer, best improvements for home valuation, process for selling a home, tips to sell your \_\_\_\_\_ home, tips to sell in under a month, do open houses sell houses, best Realtor in [neighborhood], top rated Realtor in [area], house for sale on [area], buy real estate, buy home, process of buying a home, best real estate listings, find real estate, foreclosures for sale, buy a home, houses for sale, house 4 sale, real estate agent listings, condos for sale, townhomes for sale, town houses for sale, mls listings, real estate listing mls, mls real estate listings, multiple listing services, first time home buyer guide, best homes for first time home buyer, first time home buyer programs, home buyer help, list of real estate agents, reviews of Realtor, best schools near [area], horse farms outside of [area], best homes for [big business] employees, relocation Realtor in [area].

* Review of Literature

This research is biased on predicting the house price where a US based company is investing in Australia and wanted to know the price bracket where they should know under which slab they should buy the property.

* Motivation for the Problem Undertaken

Usually when a new company is making an investment they do a lot of research and the team which works on that is really putting a lot of efforts in making the things right. So, the motivating factor is being in the team where you can be in the foundation of the new project.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

* Data Sources and their formats

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1168 entries, 0 to 1167

Data columns (total 81 columns):

# Column Non-Null Count Dtype

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

0 Id 1168 non-null int64

1 MSSubClass 1168 non-null int64

2 MSZoning 1168 non-null object

3 LotFrontage 954 non-null float64

4 LotArea 1168 non-null int64

5 Street 1168 non-null object

6 Alley 77 non-null object

7 LotShape 1168 non-null object

8 LandContour 1168 non-null object

9 Utilities 1168 non-null object

10 LotConfig 1168 non-null object

11 LandSlope 1168 non-null object

12 Neighborhood 1168 non-null object

13 Condition1 1168 non-null object

14 Condition2 1168 non-null object

15 BldgType 1168 non-null object

16 HouseStyle 1168 non-null object

17 OverallQual 1168 non-null int64

18 OverallCond 1168 non-null int64

19 YearBuilt 1168 non-null int64

20 YearRemodAdd 1168 non-null int64

21 RoofStyle 1168 non-null object

22 RoofMatl 1168 non-null object

23 Exterior1st 1168 non-null object

24 Exterior2nd 1168 non-null object

25 MasVnrType 1161 non-null object

26 MasVnrArea 1161 non-null float64

27 ExterQual 1168 non-null object

28 ExterCond 1168 non-null object

29 Foundation 1168 non-null object

30 BsmtQual 1138 non-null object

31 BsmtCond 1138 non-null object

32 BsmtExposure 1137 non-null object

33 BsmtFinType1 1138 non-null object

34 BsmtFinSF1 1168 non-null int64

35 BsmtFinType2 1137 non-null object

36 BsmtFinSF2 1168 non-null int64

37 BsmtUnfSF 1168 non-null int64

38 TotalBsmtSF 1168 non-null int64

39 Heating 1168 non-null object

40 HeatingQC 1168 non-null object

41 CentralAir 1168 non-null object

42 Electrical 1168 non-null object

43 1stFlrSF 1168 non-null int64

44 2ndFlrSF 1168 non-null int64

45 LowQualFinSF 1168 non-null int64

46 GrLivArea 1168 non-null int64

47 BsmtFullBath 1168 non-null int64

48 BsmtHalfBath 1168 non-null int64

49 FullBath 1168 non-null int64

50 HalfBath 1168 non-null int64

51 BedroomAbvGr 1168 non-null int64

52 KitchenAbvGr 1168 non-null int64

53 KitchenQual 1168 non-null object

54 TotRmsAbvGrd 1168 non-null int64

55 Functional 1168 non-null object

56 Fireplaces 1168 non-null int64

57 FireplaceQu 617 non-null object

58 GarageType 1104 non-null object

59 GarageYrBlt 1104 non-null float64

60 GarageFinish 1104 non-null object

61 GarageCars 1168 non-null int64

62 GarageArea 1168 non-null int64

63 GarageQual 1104 non-null object

64 GarageCond 1104 non-null object

65 PavedDrive 1168 non-null object

66 WoodDeckSF 1168 non-null int64

67 OpenPorchSF 1168 non-null int64

68 EnclosedPorch 1168 non-null int64

69 3SsnPorch 1168 non-null int64

70 ScreenPorch 1168 non-null int64

71 PoolArea 1168 non-null int64

72 PoolQC 7 non-null object

73 Fence 237 non-null object

74 MiscFeature 44 non-null object

75 MiscVal 1168 non-null int64

76 MoSold 1168 non-null int64

77 YrSold 1168 non-null int64

78 SaleType 1168 non-null object

79 SaleCondition 1168 non-null object

80 SalePrice 1168 non-null int64

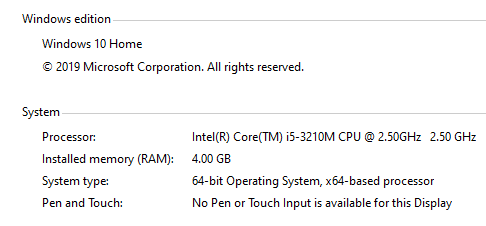
dtypes: float64(3), int64(35), object(43)

memory usage: 739.2+ KB

* Data Pre-processing Done

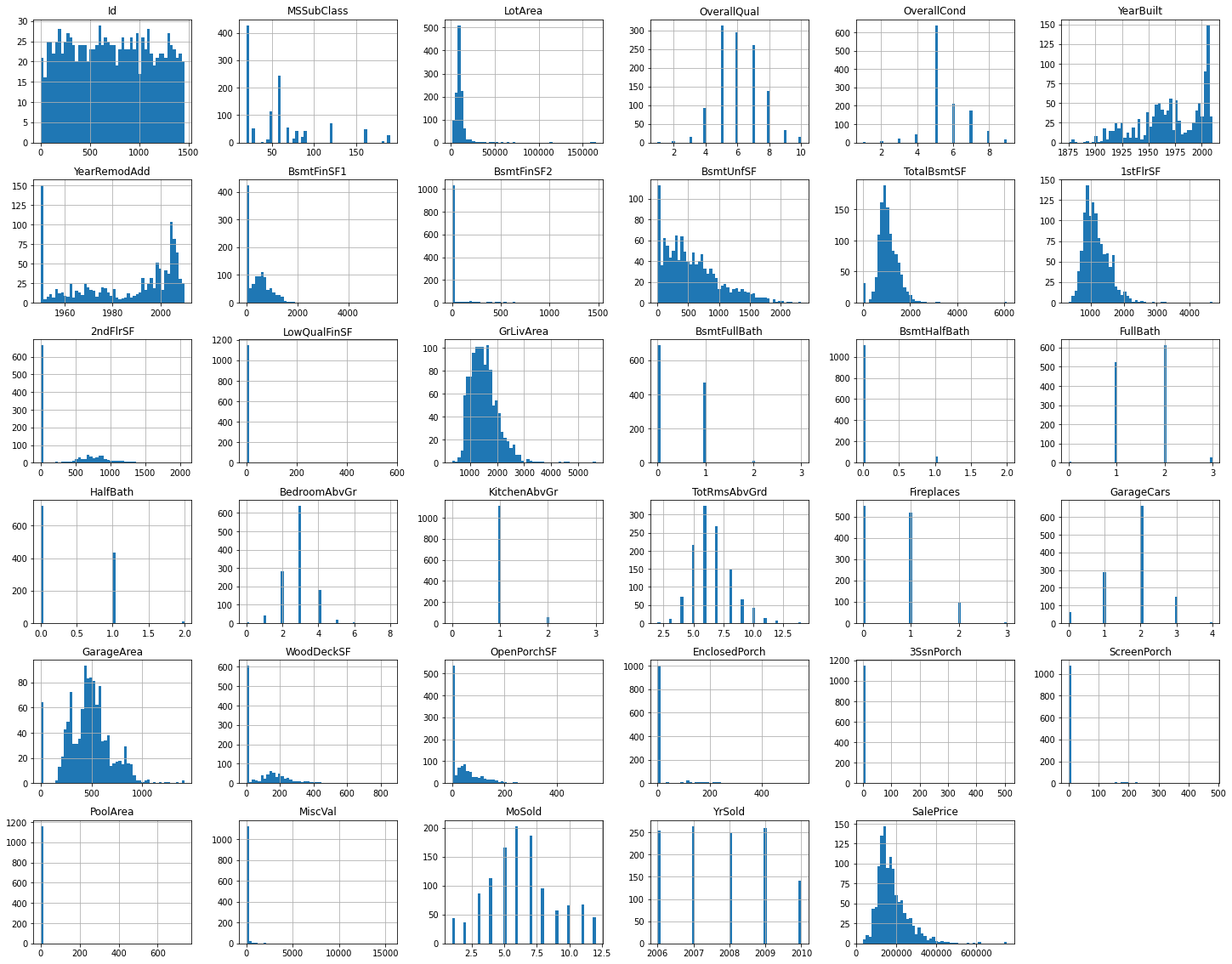
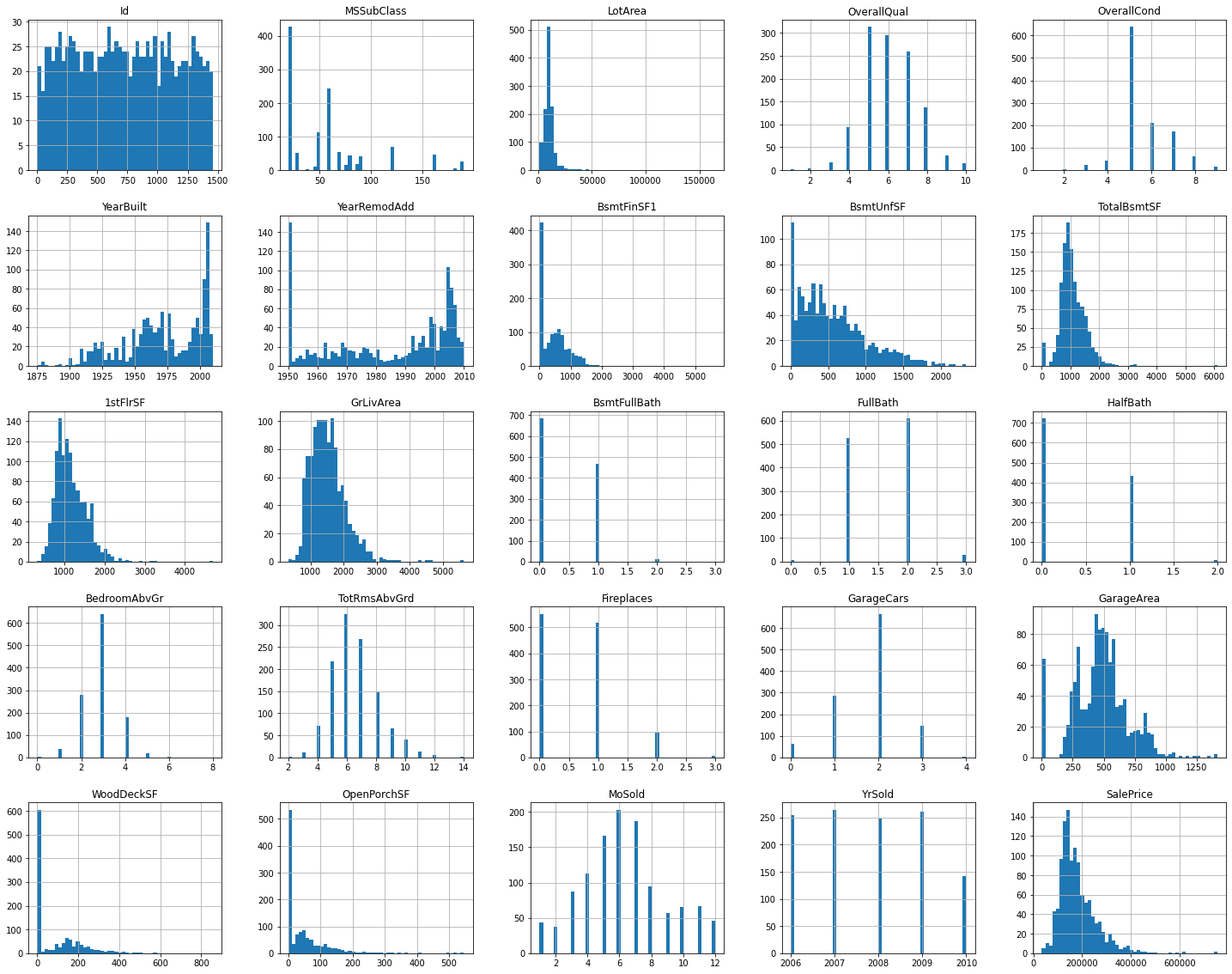
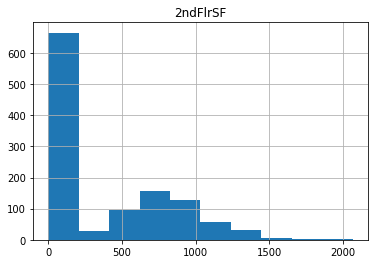
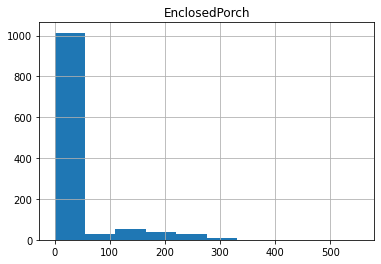
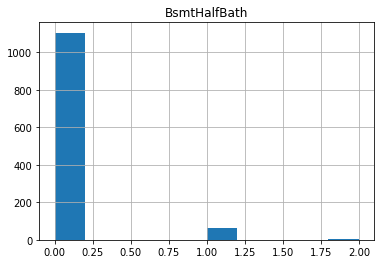
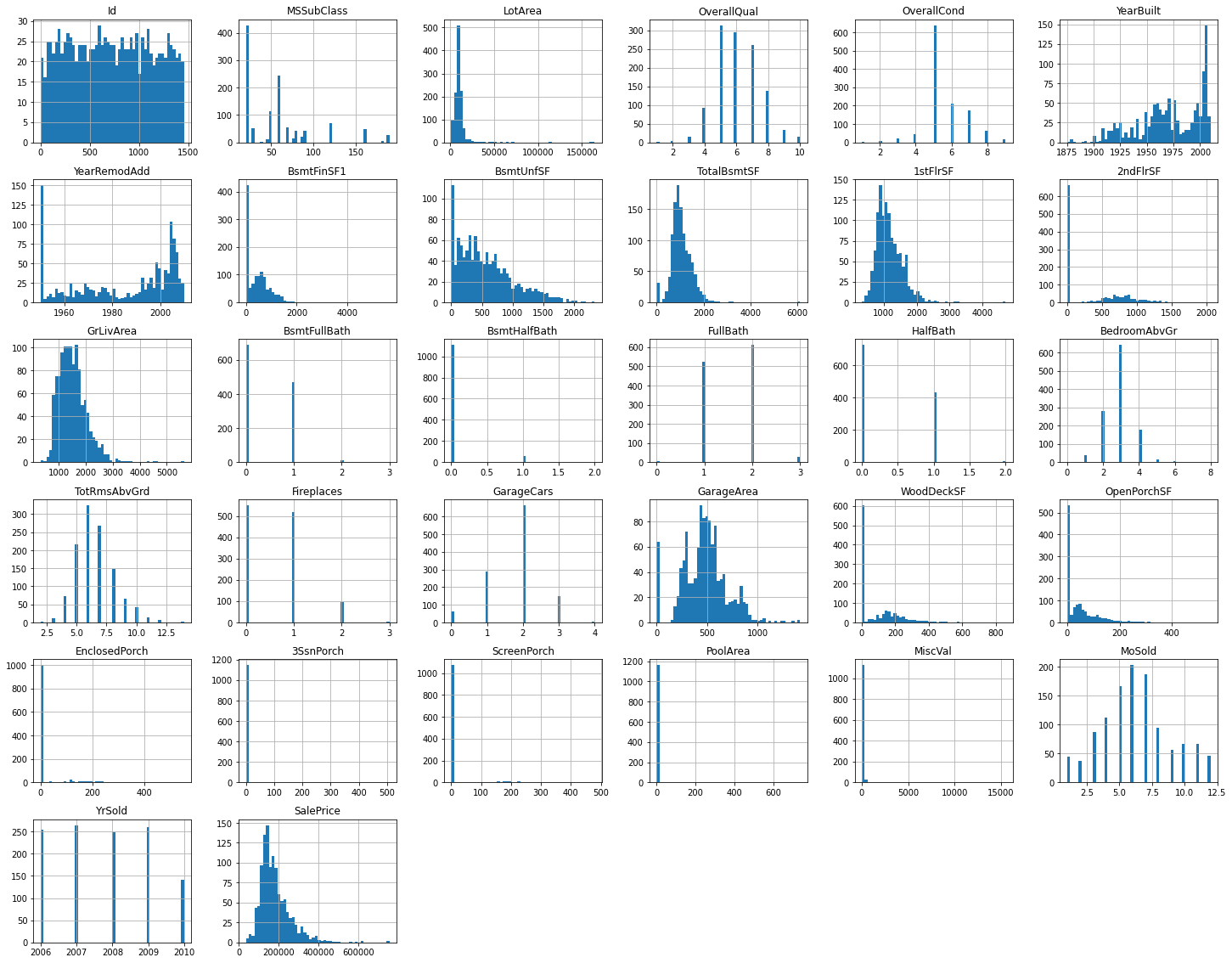
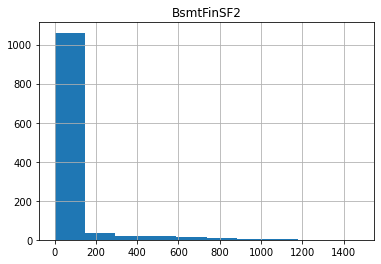
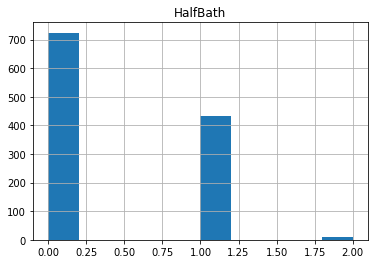
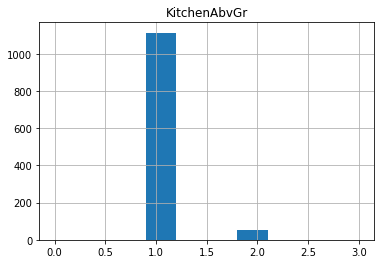
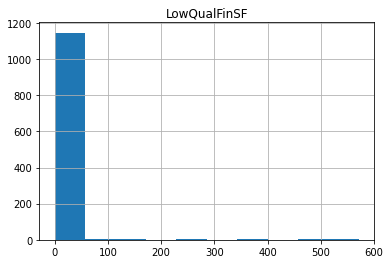
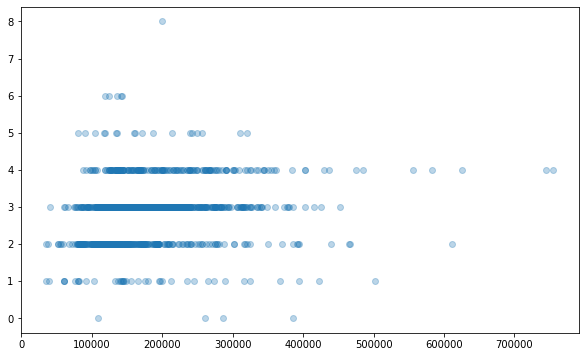
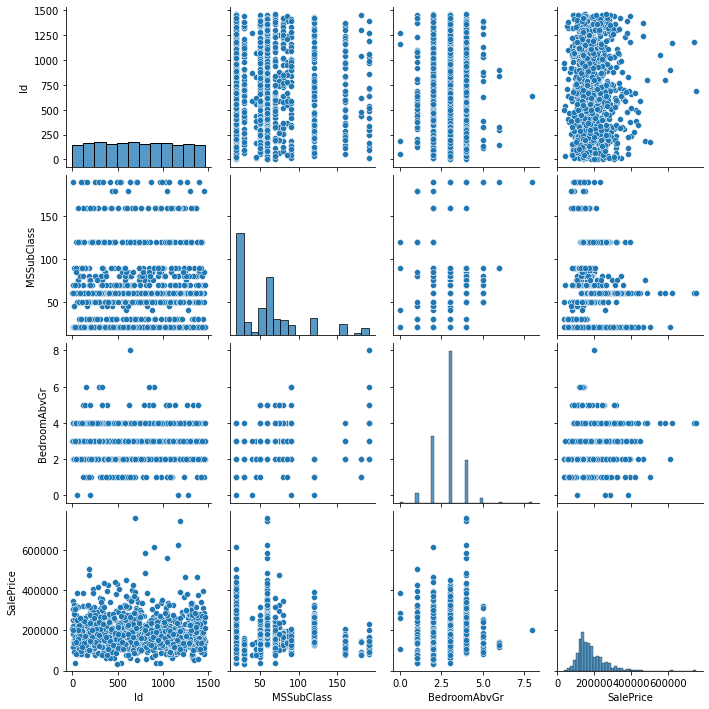
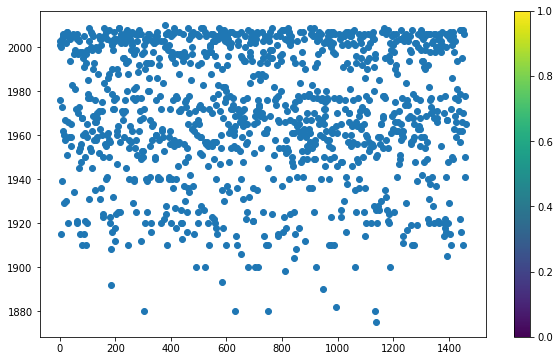
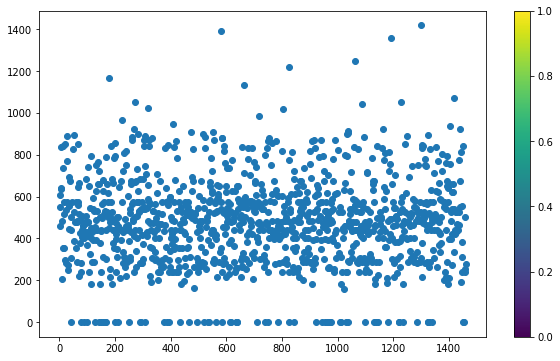
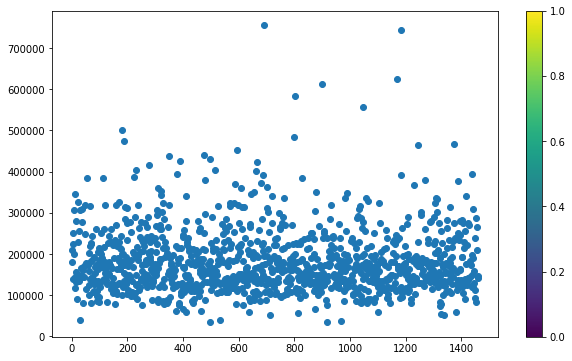
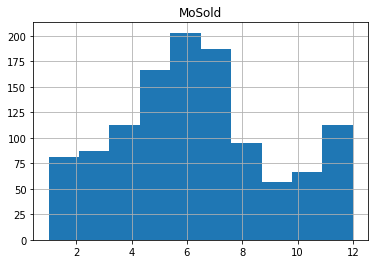
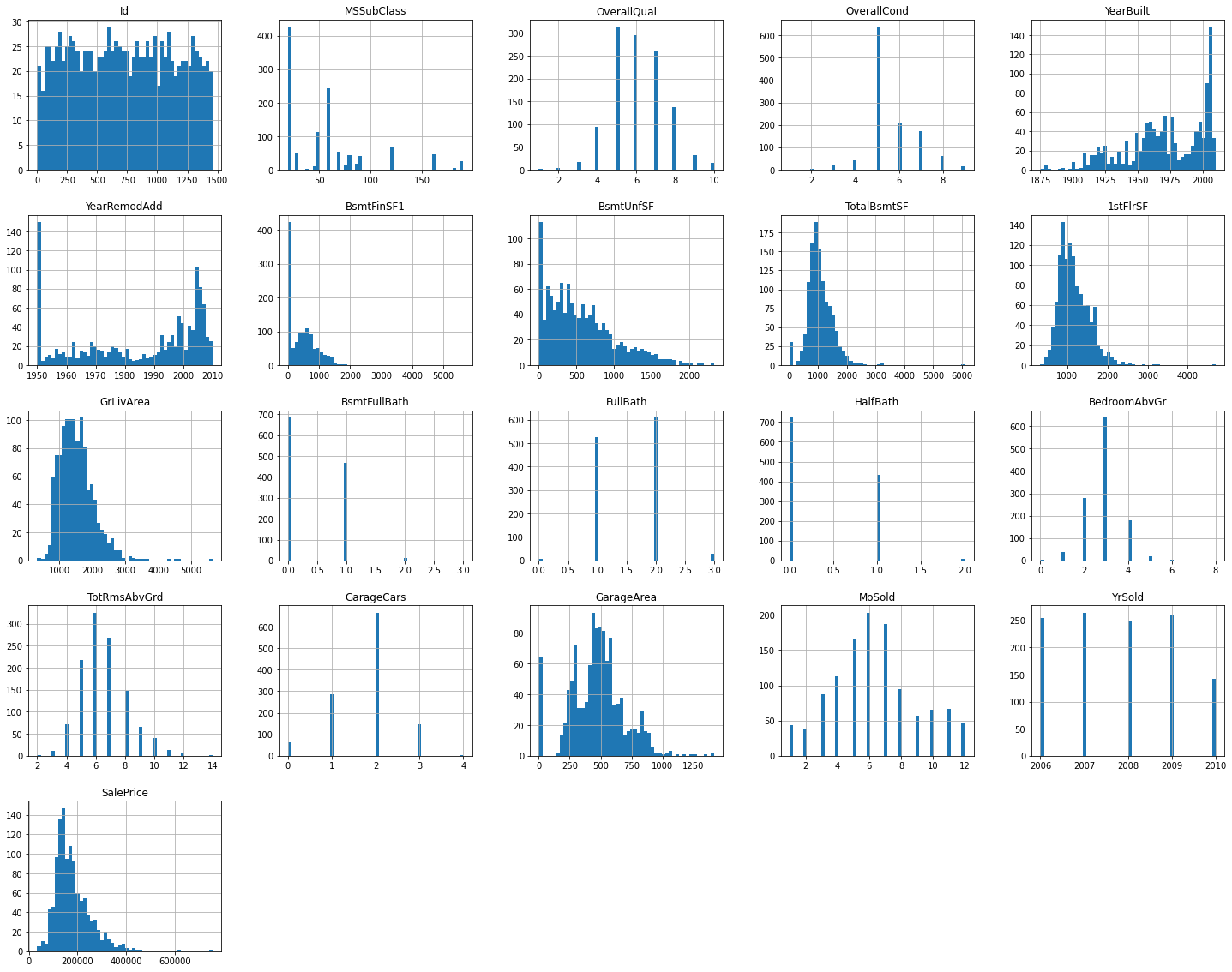
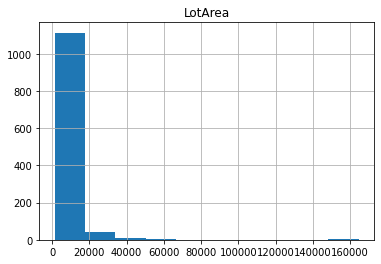
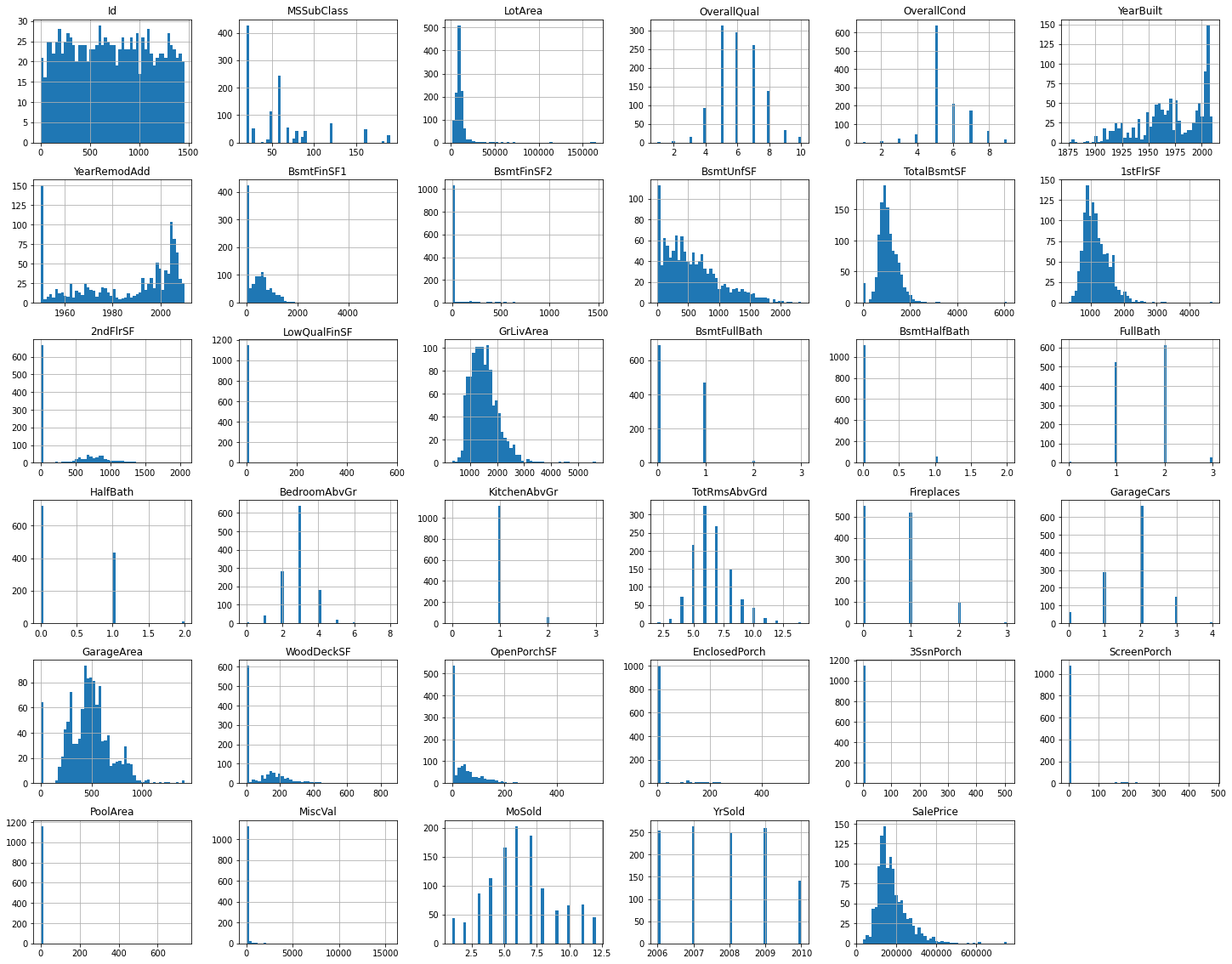
Usually the data was analysed on the basic knowledge, which were required, basically the data with the null values were removed. Rest of the data was analysed using histogram and other charts which were required for the EDA and were considered or removed biased on the analysis.

* Hardware and Software Requirements and Tools Used



**Model/s Development and Evaluation**

* Possible Histograms



**CONCLUSION**

* Key Findings and Conclusions of the Study

linreg.intercept\_ - -4881.4902574024745

linreg.coef\_ - array([26679.73785731, 50021.26293769, 25019.54707213, 2060.9627429 ,

22245.49334955, 86.10549404])

[('BsmtFullBath', 26679.737857310847),

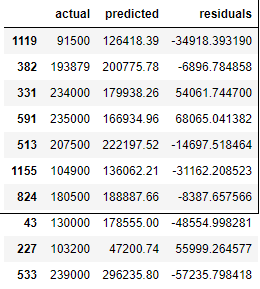
('FullBath', 50021.262937693864),

('HalfBath', 25019.54707213473),

('BedroomAbvGr', 2060.962742900453),

('GarageCars', 22245.493349545206),

('GarageArea', 86.10549404029007)]



R squared: 0.5619816094110293

RMSE: 54245.3

