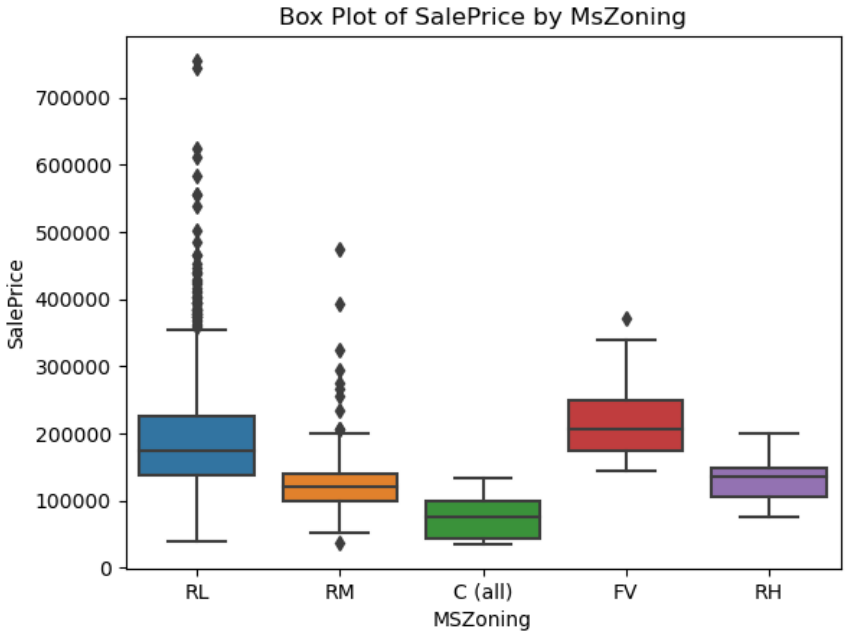
Summary ‘house prices’ project

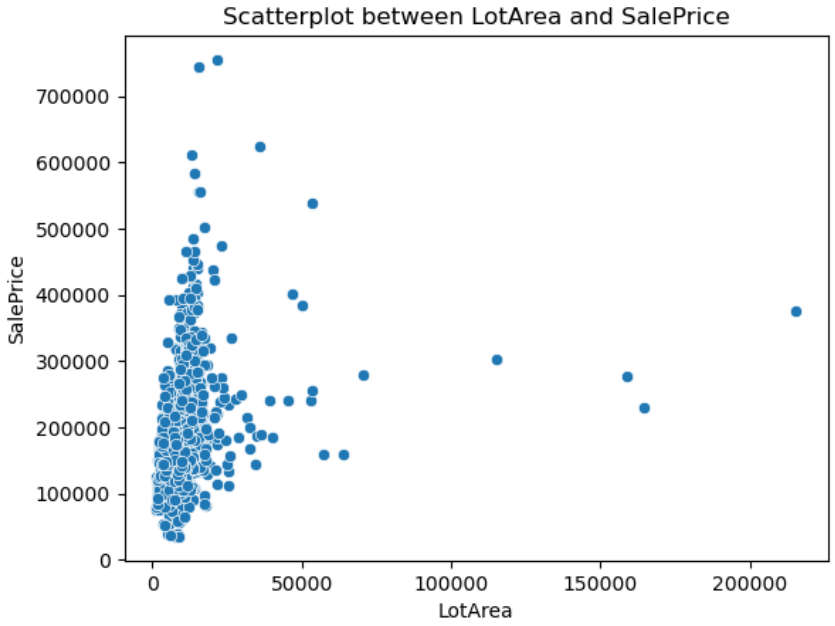
Exploring the relation between 'MSZoning' and 'SalePrice'



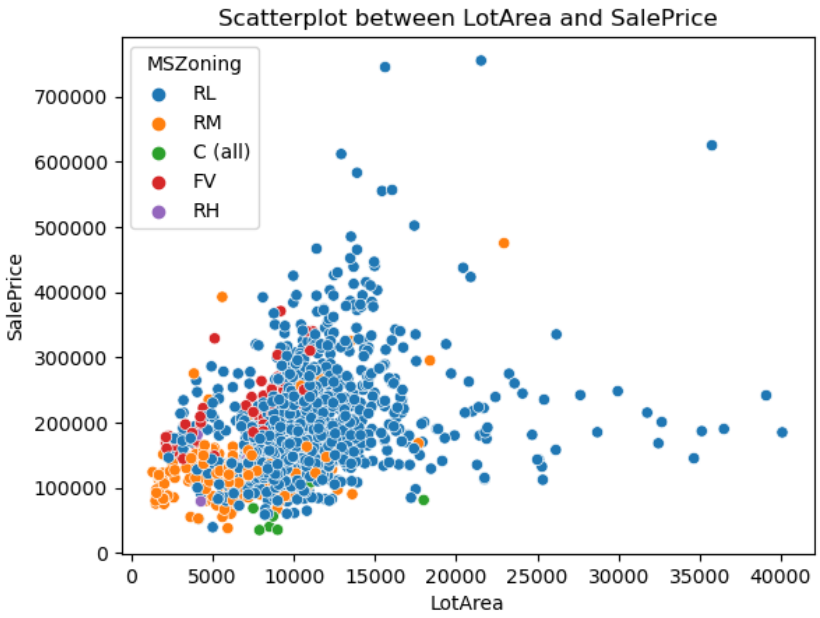
MSZoning	Mean	Q1	Median	Q3
C	74528.000000	43998.25	74700.0	98332.0
FV	214014.061538	174000.00	205950.0	250000.0
RH	131558.375000	106150.00	136500.0	148608.5
RL	191004.994787	137700.00	174000.0	224700.0
RM	126316.830275	100000.00	120500.0	140000.0

Despite outliers, houses in 'Floating Village Residential' are the most expensive ones.

Exploring the relation between 'LotArea' and 'SalePrice'



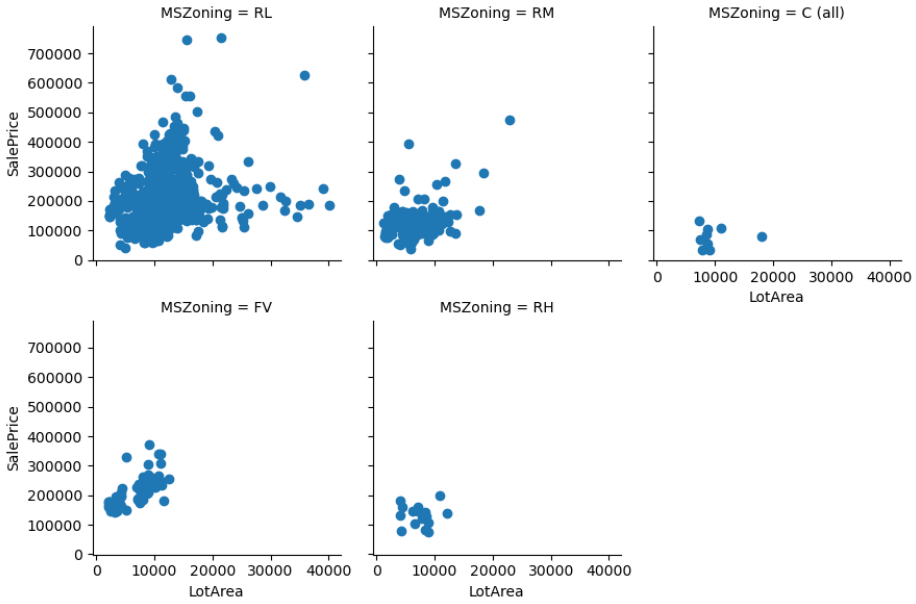
After cleaning outliers, let's plot again.



$$Corr = 0.376477$$

According to the plot and Pearson correlation. There is not a strong correlation between 'LotArea' and 'SalePrice'.

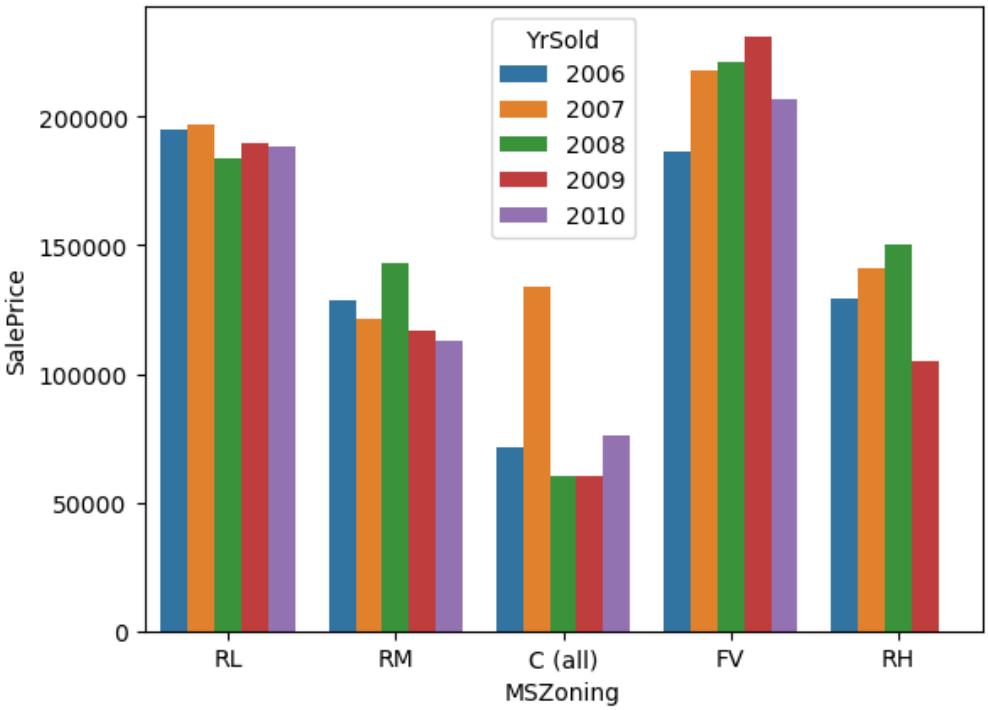
Let's explore correlation between 'LotArea' and 'SalesPrice' by each value of 'MSZoning'.



MSZoning	Corr
RL	0.312308
RM	0.419281
C	0.085421
FV	0.700305
RH	0.013779

The only linear correlation between LotArea and SalePrice was found within 'FV' houses (CORR = 0.70). It's positive and strong.

Exploring the relation between 'Year' and 'SalePrice'



According to the plot, prices have been rising between 2006 and 2010 for 'FV' houses, but apart from that we cant establish any relation between 'YrSold' and 'SalePrice'.