1. 导入训练集数据，并观察数据内容
2. 销售价格数据的观察
3. 描述性统计

count 1460.000000

mean 180921.195890

std 79442.502883

min 34900.000000

25% 129975.000000

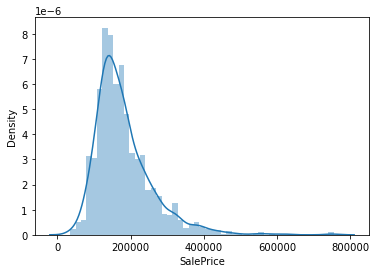
50% 163000.000000

75% 214000.000000

max 755000.000000

Name: SalePrice, dtype: float64

1. 柱状图

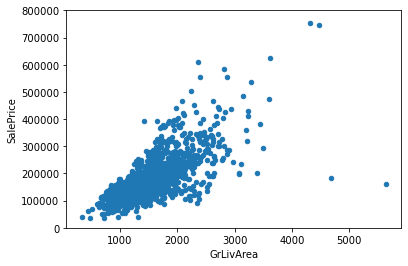


偏态分布

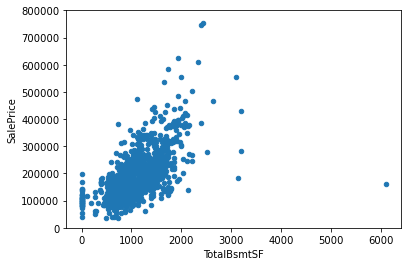
偏度:1.882876

峰度:6.536282

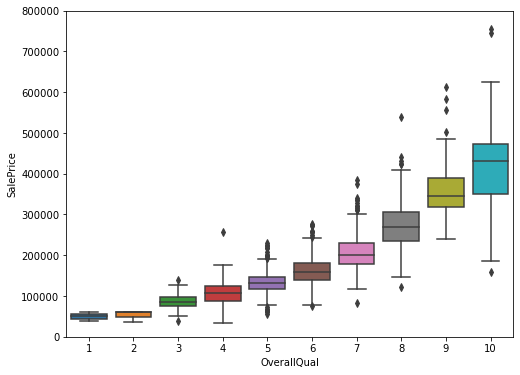
1. 探索价格与其他变量的关系
2. 居住面积和价格的关系



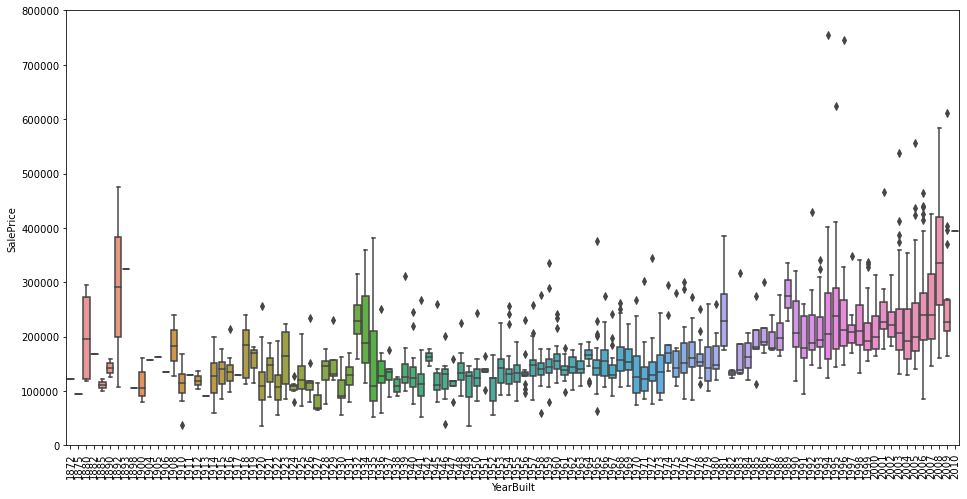
二）地下室面积和价格的关系



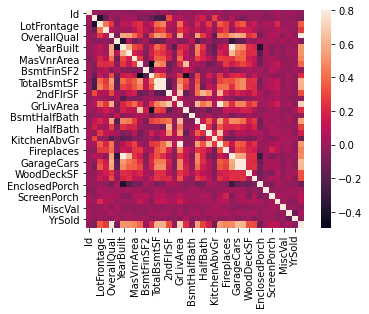
1. 整体材料和便面质量与价格的关系



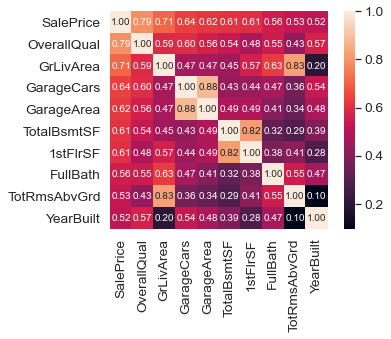
1. 建筑年份与价格的关系



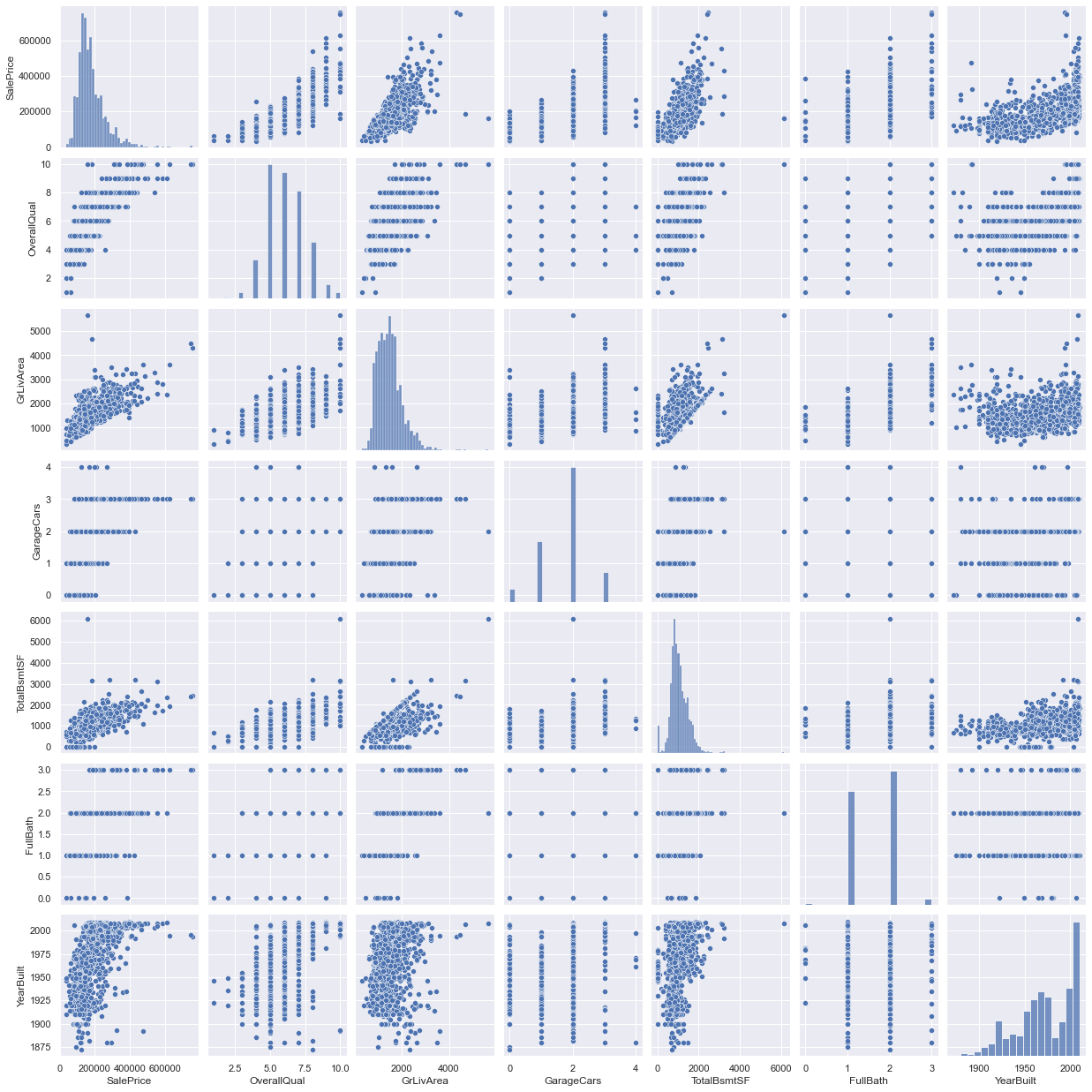
1. 客观分析
2. 所有变量之间相关性分析热力矩阵图



1. 价格与前十个关联度较大的变量的热力图



1. 销售价格和相关变量的散点图



1. 处理缺失数据、异常值
2. 单变量分析（标准化数据）

为了减少离散值对数据的影响，将数据标准化，得到数据顶部和底部的范围如下：

底部范围的分布：

[[-1.83820775]

[-1.83303414]

[-1.80044422]

[-1.78282123]

[-1.77400974]

[-1.62295562]

[-1.6166617 ]

[-1.58519209]

[-1.58519209]

[-1.57269236]]

顶部范围的分布：

[[3.82758058]

[4.0395221 ]

[4.49473628]

[4.70872962]

[4.728631 ]

[5.06034585]

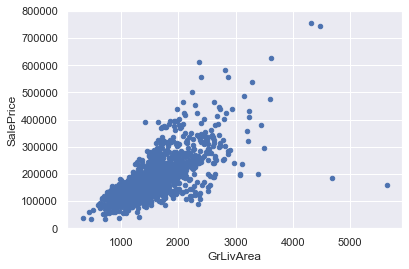
[5.42191907]

[5.58987866]

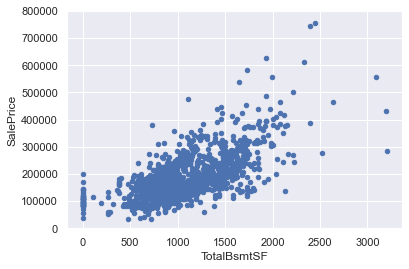
[7.10041987]

[7.22629831]]

1. 双变量分析
2. 价格和居住面积的分析

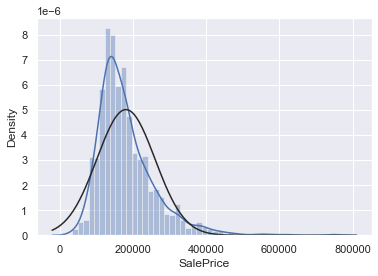


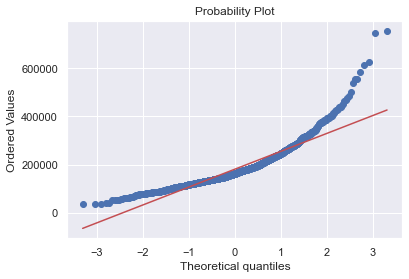
1. 价格和地下室总面积的分析



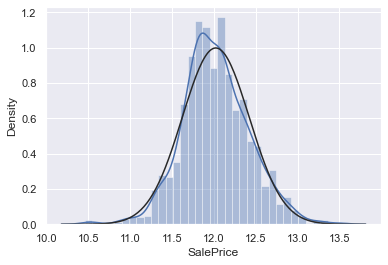
1. 价格的分布直方图和正态概率图

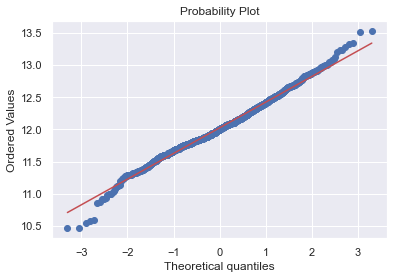
1）





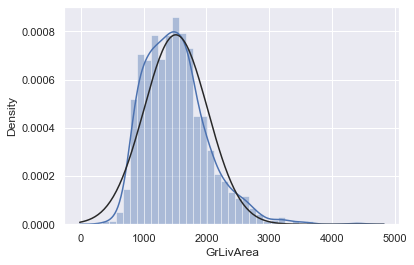
1. 对数变换，正态化后直方图和正态概率图

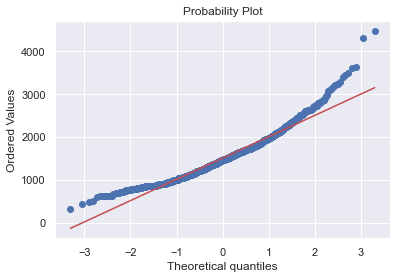




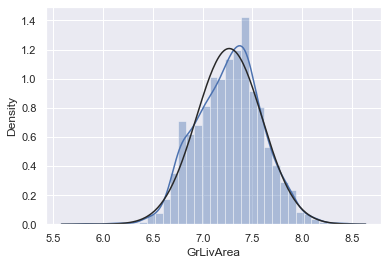
1. 居住面积的直方图和正态概率图

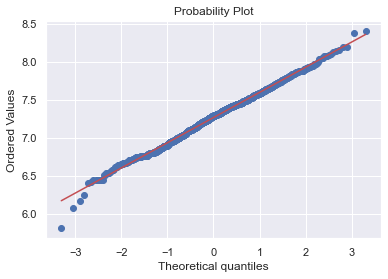
1）





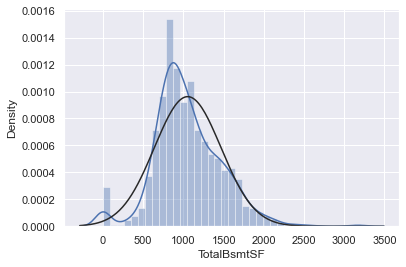
1. 对数变换后的直方图和正态概率图

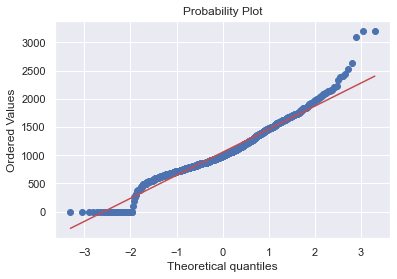




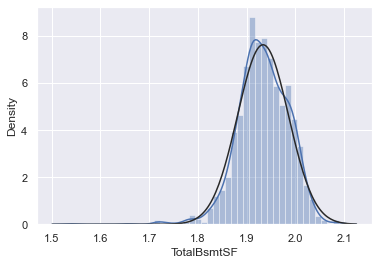
1. 地下室居住面积

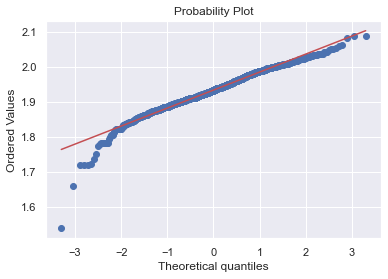
1）



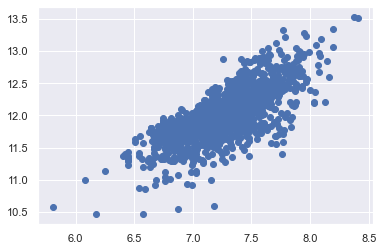


1. 对数变换后





1. 价格和居住面积的分析



1. 价格和地下室面积的分析

