1、sklearn中包含比较多的数据预处理函数，提供的文档主要对类别型特征进行了LabelEncoder()处理。因为数据本身的格式是DataFrame类型的，可查看各个特征的基本统计特征、以及缺失状况。并能通过pandas.get\_dummies()函数对类别型特征进行one-hot编码。

2、模型一般只能拟合数值型的特征，所以通过one-hot编码对类别型特征进行转换。

3、在该数据集中，1stFlrSF、2ndFlrSF等特征的取值比较大，而其它一些特征的取值比较小。此外由于预测模型用的是LinearRegression，需消除数据特征之间量纲的影响，使各个特征处于同一个数量级。此外，归一化之后，特征之间处于同一个数量级，在使用梯度下降法更新参数时，其更新速度更为一致，更容易找到最优解。