

Assignment 3.2

Summary of House Dataset

Name: Mohammad Nazim Rehman

Student Id: 100415501

Summary:

Variables: 81, Rows: 1460

Size of dataset: 118,260

The houses dataset discusses the various components of a housing market and what goes on to decide the price of the house, e.g., Carpet Area, Amenities, Neighbourhood, when it was built, etc.

As we can observe from the correlation matrix the sale price of a house is highly dependent upon the area of the first floor, the number of car garages and which year it was built

- The **SalePrice** distribution is skewed, which may affect certain regression models. Transformations like the **log transformation** might help normalize the data.
- There are strong correlations between **GrLivArea**, **OverallQual**, and **SalePrice**, which means these features will likely be strong predictors for house pricing.
- Handling outliers and missing values is crucial for clean data and better model accuracy.
- **Size and Price**: Generally, a notable positive correlation is expected between the size of a home (its area) and its price. Larger homes tend to be priced higher due to the increased space and amenities they offer.
- **Location and Price**: The geographical location plays a significant role in the price of a property. Homes in more central or urban areas often fetch higher prices compared to those in suburban or rural settings due to factors like accessibility, proximity to services, and demand.
- **Condition and Price**: The condition of the property should influence its market value. Well-maintained homes with fewer repair needs are usually priced higher, as they require less immediate investment from the buyer.
- **Age and Price**: Newer homes are often more expensive because of modern construction standards, updated infrastructure, and reduced maintenance needs. However, older homes in excellent condition or with historical significance can also retain substantial value.
- **Garage and Price**: Homes with a garage generally have a higher price tag compared to those without, as the added space for storage and parking is seen as a valuable feature for many buyers.

Summary Statistics for Numerical Columns:

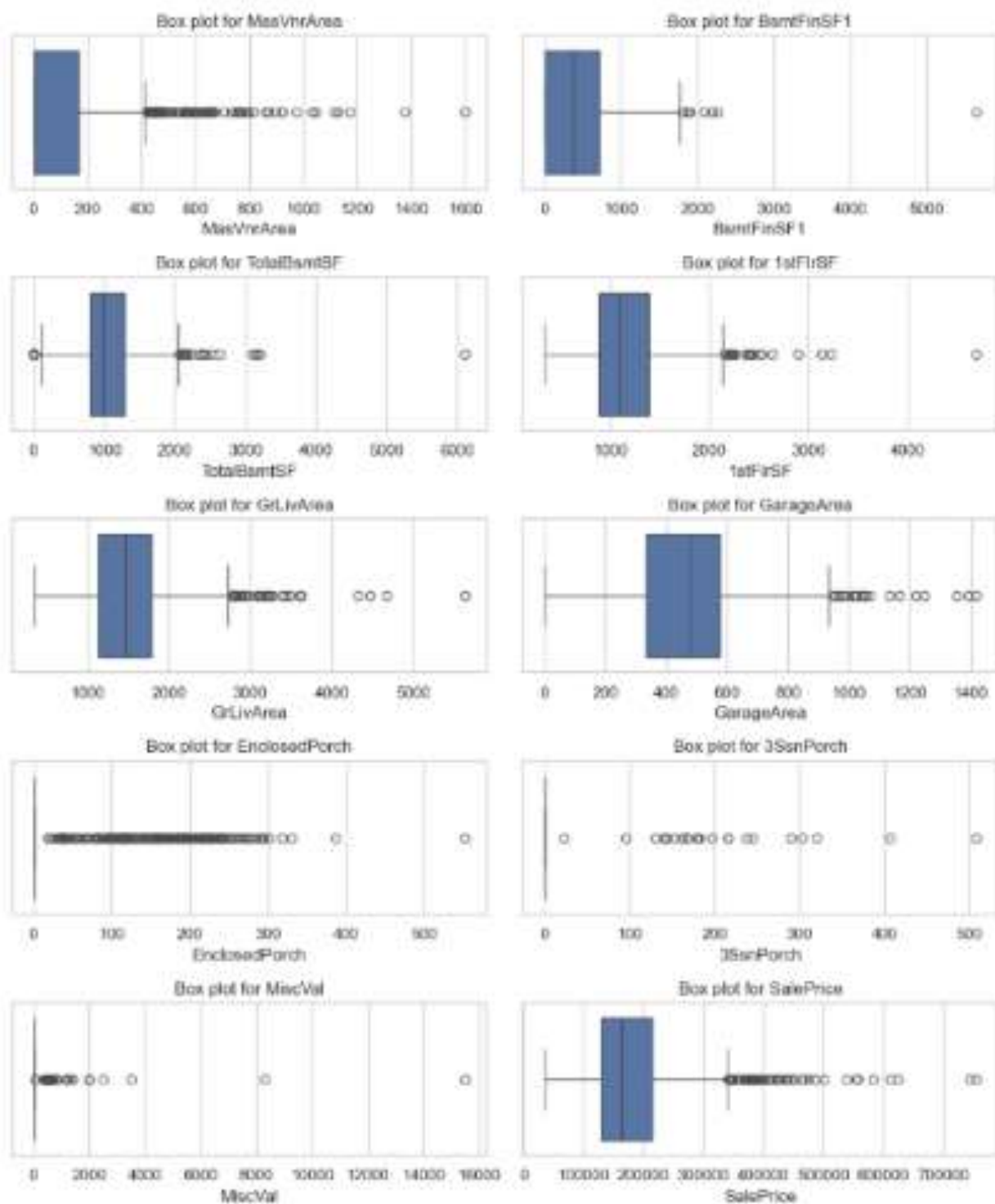
	Id	LotFrontage	LotArea	MasVnrArea	BsmtFinSF1 \
count	1460.000000	1460.000000	1460.000000	1452.000000	1460.000000
mean	730.500000	70.049958	10516.828082	103.685262	443.639726
std	421.610009	22.024023	9981.264932	181.066207	456.098091
min	1.000000	21.000000	1300.000000	0.000000	0.000000
25%	365.750000	60.000000	7553.500000	0.000000	0.000000
50%	730.500000	70.049958	9478.500000	0.000000	383.500000
75%	1095.250000	79.000000	11601.500000	166.000000	712.250000
max	1460.000000	313.000000	215245.000000	1600.000000	5644.000000

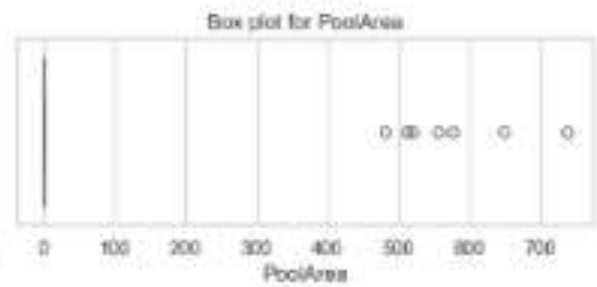
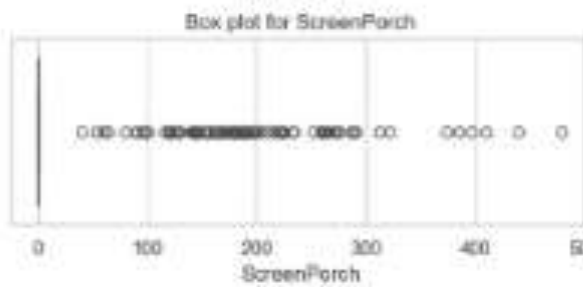
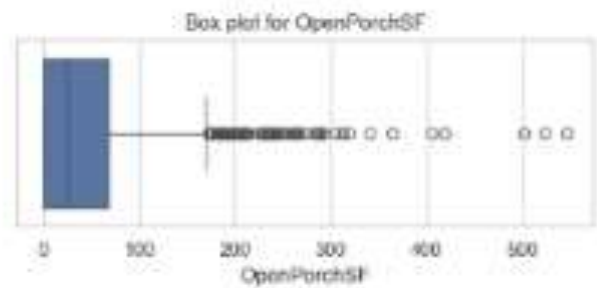
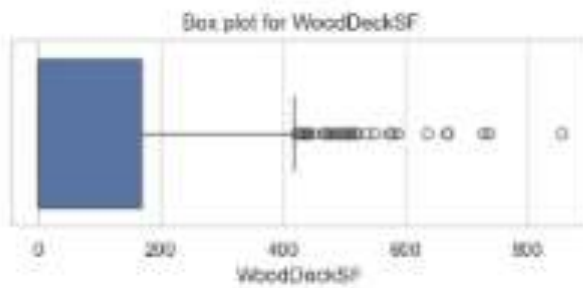
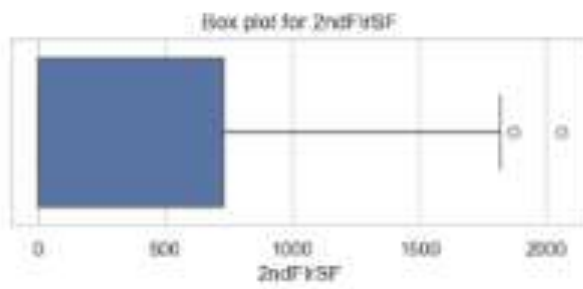
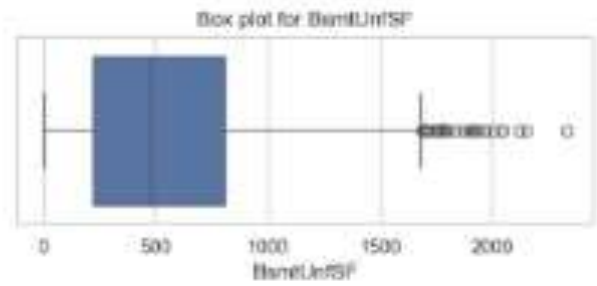
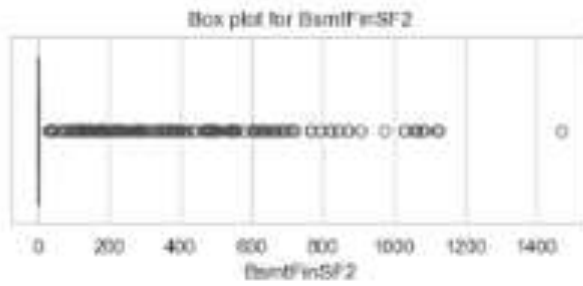
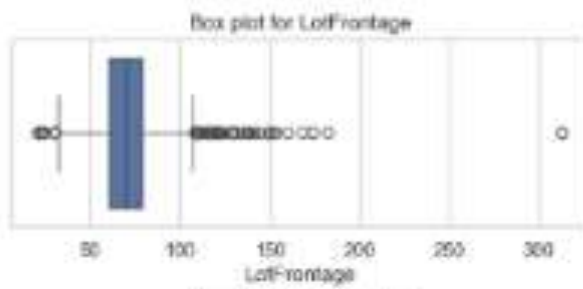
	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF	1stFlrSF	2ndFlrSF \
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
mean	46.549315	567.240411	1057.429452	1162.626712	346.992466
std	161.319273	441.866955	438.705324	386.587738	436.528436
min	0.000000	0.000000	0.000000	334.000000	0.000000
25%	0.000000	223.000000	795.750000	882.000000	0.000000
50%	0.000000	477.500000	991.500000	1087.000000	0.000000
75%	0.000000	808.000000	1298.250000	1391.250000	728.000000
max	1474.000000	2336.000000	6110.000000	4692.000000	2065.000000

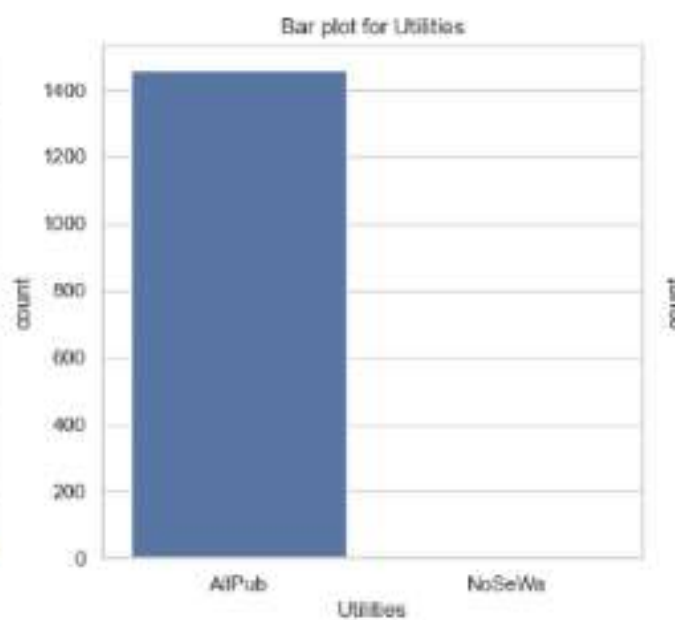
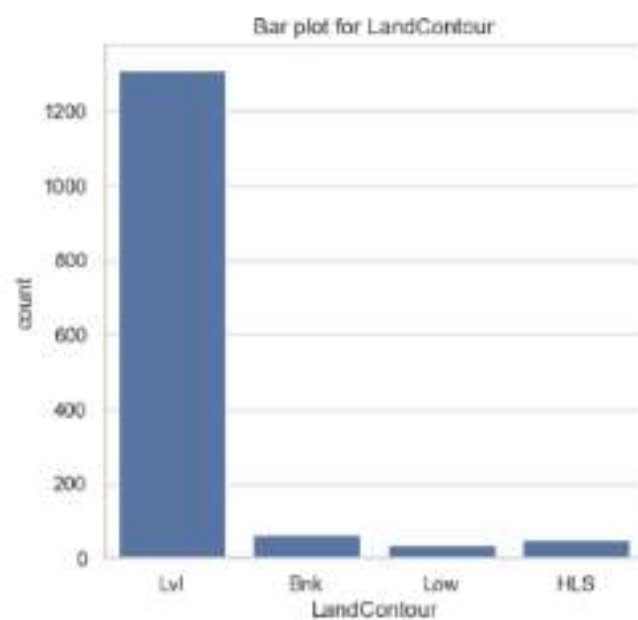
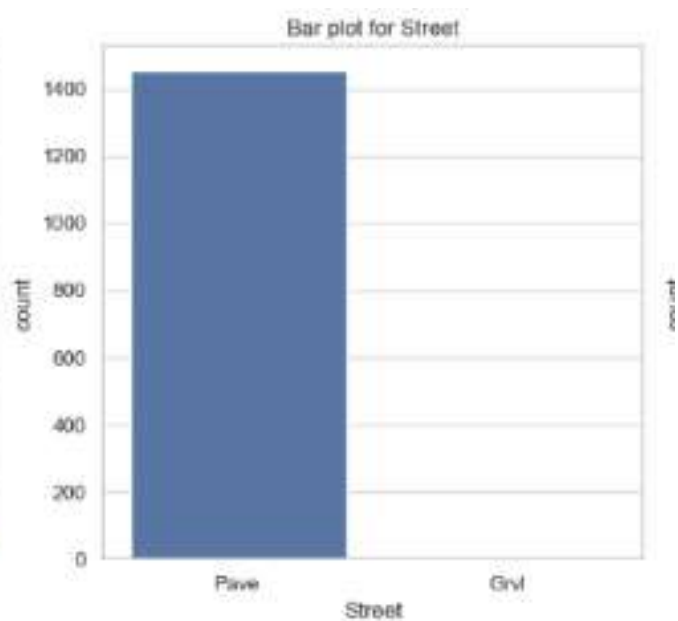
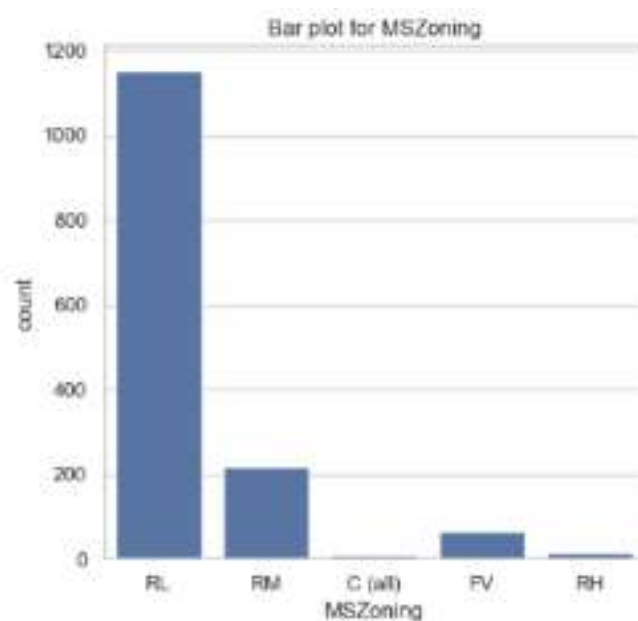
	LowQualFinSF	GrLivArea	GarageArea	WoodDeckSF	OpenPorchSF \
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
mean	5.844521	1515.463699	472.980137	94.244521	46.660274
std	48.623081	525.480383	213.804841	125.338794	66.256028
min	0.000000	334.000000	0.000000	0.000000	0.000000
25%	0.000000	1129.500000	334.500000	0.000000	0.000000
50%	0.000000	1464.000000	480.000000	0.000000	25.000000
75%	0.000000	1776.750000	576.000000	168.000000	68.000000
max	572.000000	5642.000000	1418.000000	857.000000	547.000000

	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	MiscVal \
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
mean	21.954110	3.409589	15.060959	2.758904	43.489041
std	61.119149	29.317331	55.757415	40.177307	496.123024
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000	0.000000	0.000000
max	552.000000	508.000000	480.000000	738.000000	15500.000000

	SalePrice
count	1460.000000
mean	180921.195890
std	79442.502883
min	34900.000000
25%	129975.000000
50%	163000.000000
75%	214000.000000
max	755000.000000

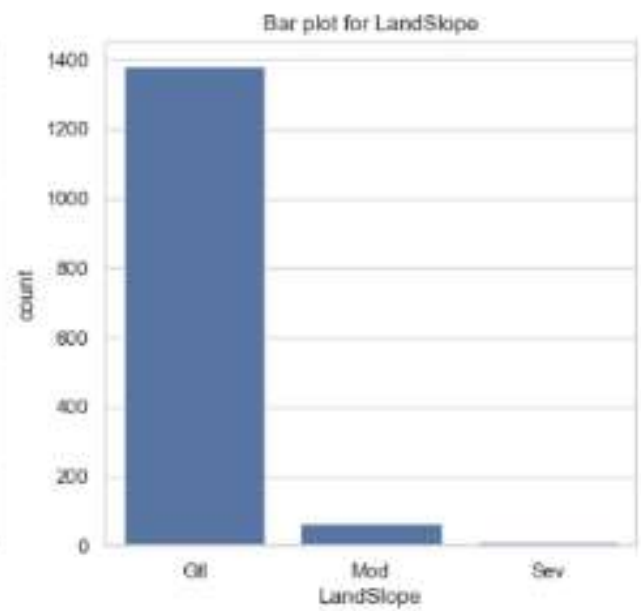
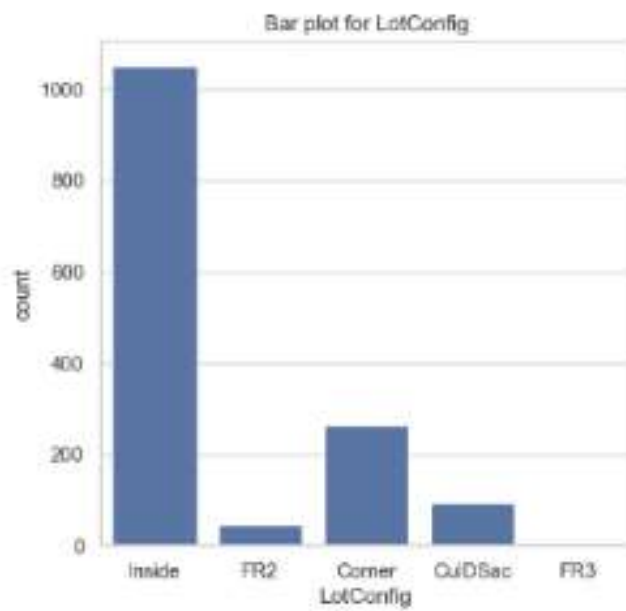
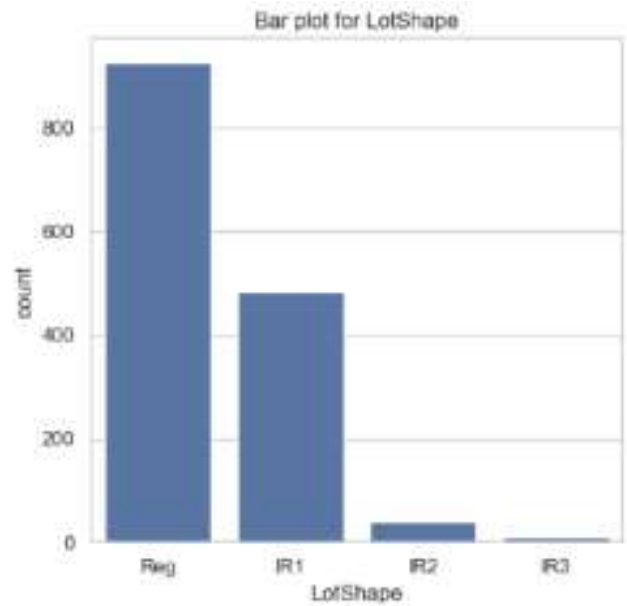
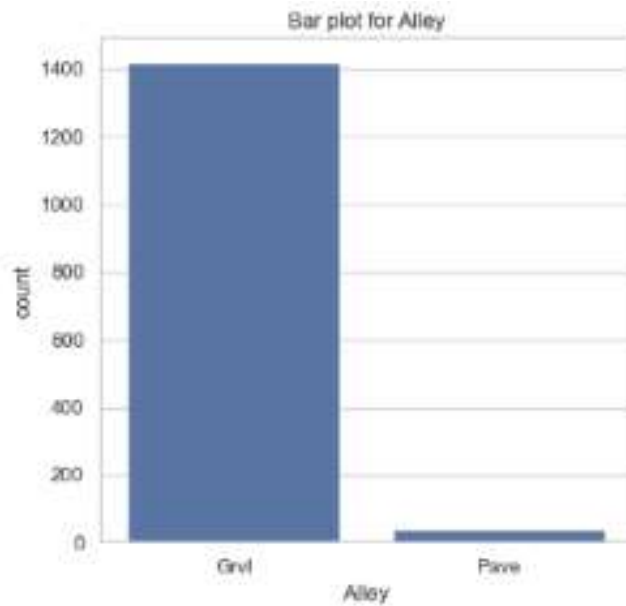


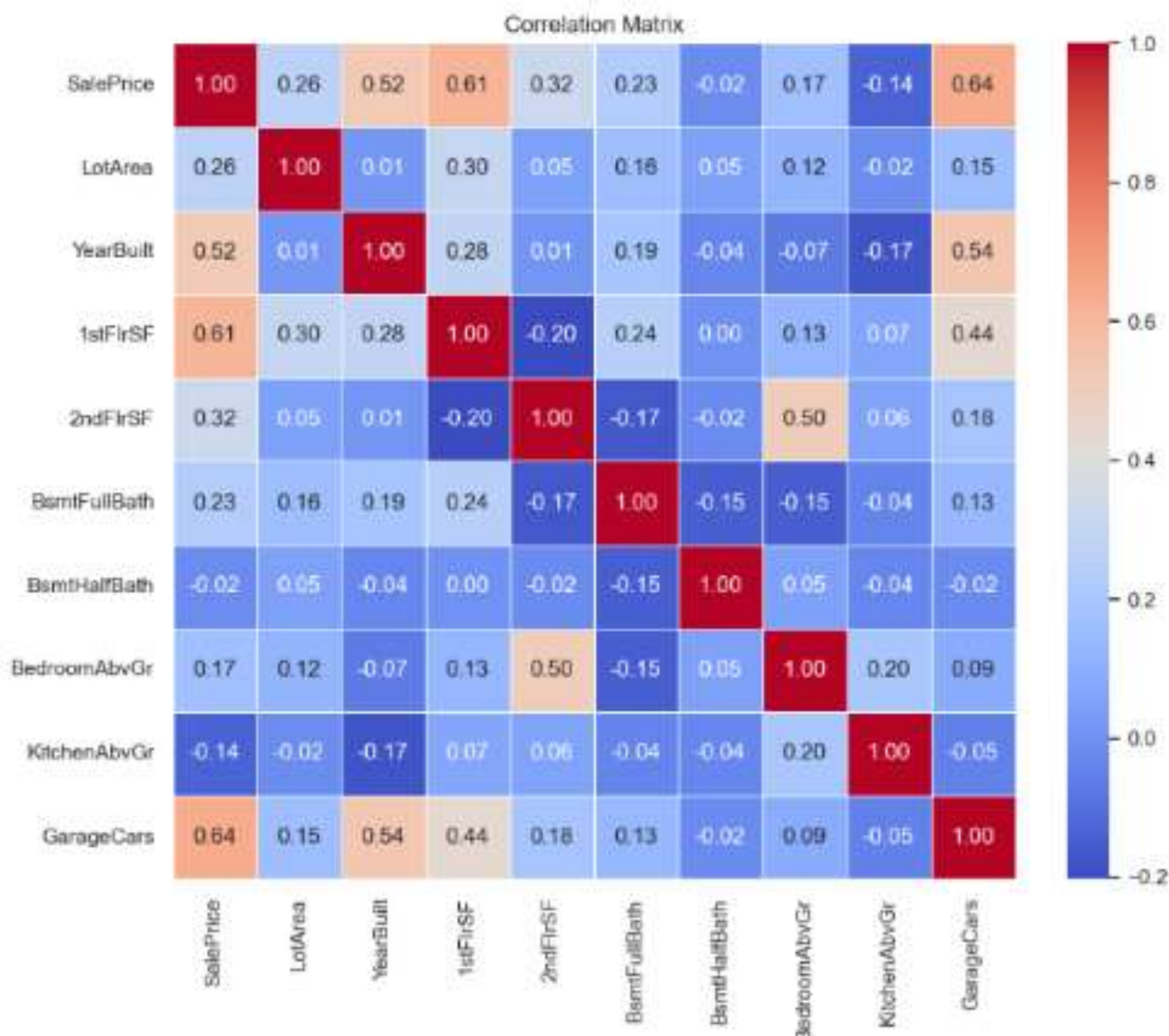
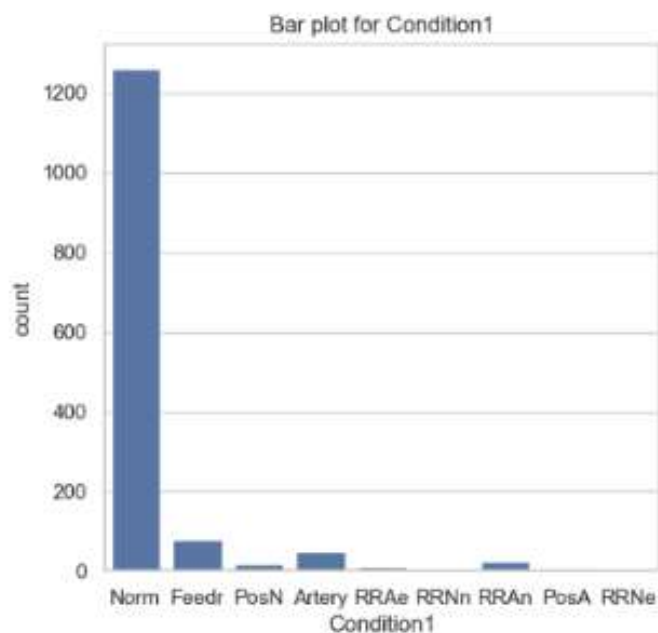
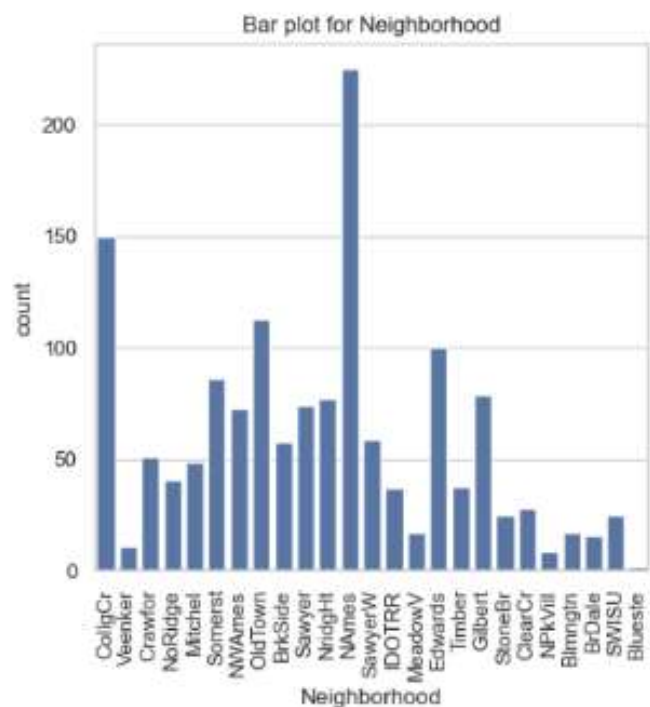




Bar plot for Neighborhood

Bar plot for Condition1





Scatter Plots of Numerical Variables

