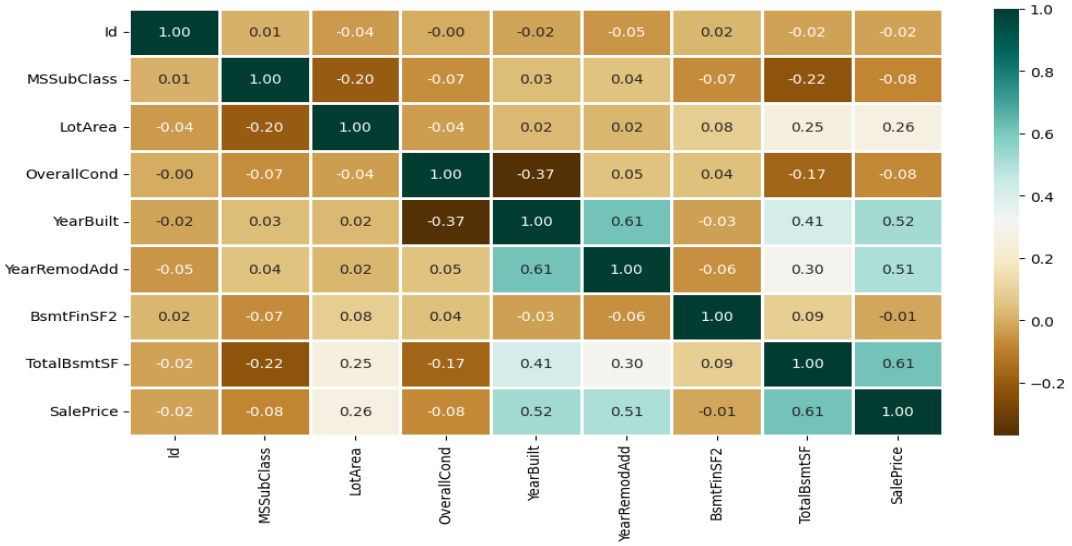
CONSIDERING 5 DATASET:

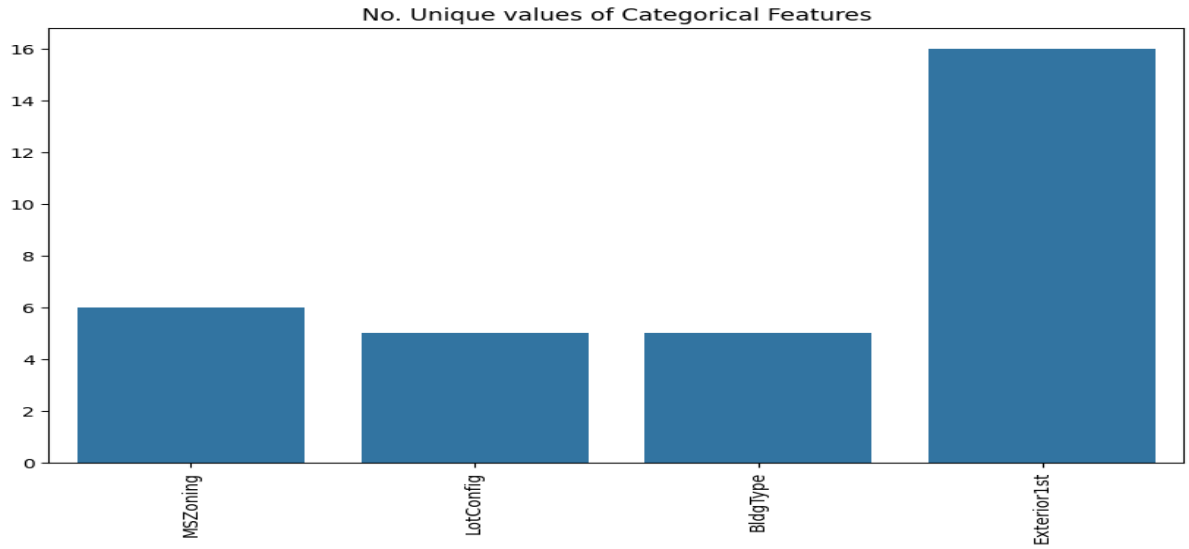
	MSSubClass	MSZoning	LotArea	LotConfig	BldgType	OverallCond	YearBuilt
0	60	RL	8450	Inside	1Fam	5	2003
1	20	RL	9600	FR2	1Fam	8	1976
2	60	RL	11250	Inside	1Fam	5	2001
3	70	RL	9550	Corner	1Fam	5	1915
4	60	RL	14260	FR2	1Fam	5	2000

	YearRemodAdd	Exterior1st	BsmtFinSF2	TotalBsmtSF	SalePrice
0	2003	VinylSd	0.0	856.0	208500.0
1	1976	MetalSd	0.0	1262.0	181500.0
2	2002	VinylSd	0.0	920.0	223500.0
3	1970	Wd Sdng	0.0	756.0	140000.0
4	2000	VinylSd	0.0	1145.0	250000.0

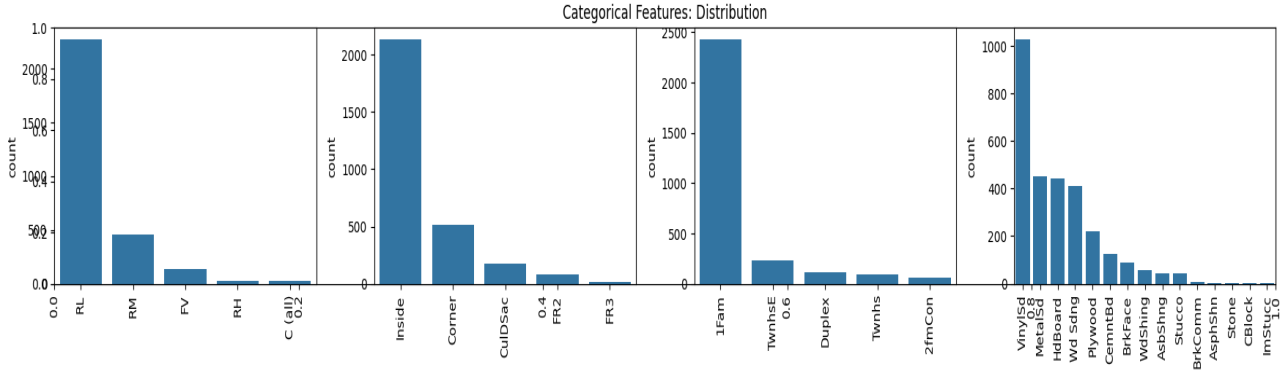
HEAT MAP:



Bar Plot 1: Visualizes the number of unique values for each categorical feature in the dataset



Bar Plot 2: Plotted bar charts for the distribution of categorical features in the dataset. Each subplot shows how frequently each unique category appears in a particular categorical column



Checking features which have null values in the new dataframe

```
MSSubClass      0
MSZoning        0
LotArea         0
LotConfig       0
BldgType        0
OverallCond     0
YearBuilt       0
YearRemodAdd    0
Exterior1st     0
BsmtFinSF2      0
TotalBsmtSF     0
SalePrice       0
dtype: int64
```

To Calculate Mean Absolute Percentage Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- Mean absolute percentage error using **SVM**: **0.18705129**
- Mean absolute percentage error using **Random Forest Regression: 0.1929469**
- Mean absolute percentage error using **Linear Regression: 0.187416838**

Conclusion: Clearly SVM model is giving better accuracy as the mean absolute error is the least among all the other regressor models i.e. 0.18 approx.