

```
In [1]: import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

/Users/youneshosseini/anaconda3/lib/python3.11/site-packages/pandas/core/array/s/masked.py:61: UserWarning: Pandas requires version '1.3.6' or newer of 'bottleneck' (version '1.3.5' currently installed).

```
from pandas.core import (
```

```
In [3]: df_train = pd.read_csv('house_price/train.csv')
df_train.head()
```

```
Out[3]:
```

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub

5 rows × 81 columns

```
In [5]: df_test = pd.read_csv('house_price/test.csv')
df_test.head()
```

```
Out[5]:
```

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
0	1461	20	RH	80.0	11622	Pave	NaN	Reg	Lvl	AllPub
1	1462	20	RL	81.0	14267	Pave	NaN	IR1	Lvl	AllPub
2	1463	60	RL	74.0	13830	Pave	NaN	IR1	Lvl	AllPub
3	1464	60	RL	78.0	9978	Pave	NaN	IR1	Lvl	AllPub
4	1465	120	RL	43.0	5005	Pave	NaN	IR1	HLS	AllPub

5 rows × 80 columns

```
In [6]: df_train['SalePrice'].mean()
```

```
Out[6]: 180921.19589041095
```

```
In [7]: df_test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1459 entries, 0 to 1458
```

```
Data columns (total 80 columns):
```

#	Column	Non-Null Count	Dtype
0	Id	1459 non-null	int64
1	MSSubClass	1459 non-null	int64
2	MSZoning	1455 non-null	object
3	LotFrontage	1232 non-null	float64
4	LotArea	1459 non-null	int64
5	Street	1459 non-null	object
6	Alley	107 non-null	object
7	LotShape	1459 non-null	object
8	LandContour	1459 non-null	object
9	Utilities	1457 non-null	object
10	LotConfig	1459 non-null	object
11	LandSlope	1459 non-null	object
12	Neighborhood	1459 non-null	object
13	Condition1	1459 non-null	object
14	Condition2	1459 non-null	object
15	BldgType	1459 non-null	object
16	HouseStyle	1459 non-null	object
17	OverallQual	1459 non-null	int64
18	OverallCond	1459 non-null	int64
19	YearBuilt	1459 non-null	int64
20	YearRemodAdd	1459 non-null	int64
21	RoofStyle	1459 non-null	object
22	RoofMatl	1459 non-null	object
23	Exterior1st	1458 non-null	object
24	Exterior2nd	1458 non-null	object
25	MasVnrType	565 non-null	object
26	MasVnrArea	1444 non-null	float64
27	ExterQual	1459 non-null	object
28	ExterCond	1459 non-null	object
29	Foundation	1459 non-null	object
30	BsmtQual	1415 non-null	object
31	BsmtCond	1414 non-null	object
32	BsmtExposure	1415 non-null	object
33	BsmtFinType1	1417 non-null	object
34	BsmtFinSF1	1458 non-null	float64
35	BsmtFinType2	1417 non-null	object
36	BsmtFinSF2	1458 non-null	float64
37	BsmtUnfSF	1458 non-null	float64
38	TotalBsmtSF	1458 non-null	float64
39	Heating	1459 non-null	object
40	HeatingQC	1459 non-null	object
41	CentralAir	1459 non-null	object
42	Electrical	1459 non-null	object
43	1stFlrSF	1459 non-null	int64
44	2ndFlrSF	1459 non-null	int64
45	LowQualFinSF	1459 non-null	int64
46	GrLivArea	1459 non-null	int64
47	BsmtFullBath	1457 non-null	float64
48	BsmtHalfBath	1457 non-null	float64
49	FullBath	1459 non-null	int64
50	HalfBath	1459 non-null	int64
51	BedroomAbvGr	1459 non-null	int64
52	KitchenAbvGr	1459 non-null	int64
53	KitchenQual	1458 non-null	object
54	TotRmsAbvGrd	1459 non-null	int64

```

55 Functional      1457 non-null object
56 Fireplaces      1459 non-null int64
57 FireplaceQu     729 non-null object
58 GarageType      1383 non-null object
59 GarageYrBlt     1381 non-null float64
60 GarageFinish    1381 non-null object
61 GarageCars      1458 non-null float64
62 GarageArea      1458 non-null float64
63 GarageQual      1381 non-null object
64 GarageCond      1381 non-null object
65 PavedDrive      1459 non-null object
66 WoodDeckSF      1459 non-null int64
67 OpenPorchSF     1459 non-null int64
68 EnclosedPorch   1459 non-null int64
69 3SsnPorch       1459 non-null int64
70 ScreenPorch     1459 non-null int64
71 PoolArea        1459 non-null int64
72 PoolQC          3 non-null object
73 Fence           290 non-null object
74 MiscFeature     51 non-null object
75 MiscVal         1459 non-null int64
76 MoSold          1459 non-null int64
77 YrSold          1459 non-null int64
78 SaleType        1458 non-null object
79 SaleCondition   1459 non-null object
dtypes: float64(11), int64(26), object(43)
memory usage: 912.0+ KB

```

In [8]: `df_train.describe()`

Out[8]:

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	
count	1460.000000	1460.000000	1201.000000	1460.000000	1460.000000	1460.000000	1460
mean	730.500000	56.897260	70.049958	10516.828082	6.099315	5.575342	197
std	421.610009	42.300571	24.284752	9981.264932	1.382997	1.112799	30
min	1.000000	20.000000	21.000000	1300.000000	1.000000	1.000000	187
25%	365.750000	20.000000	59.000000	7553.500000	5.000000	5.000000	195
50%	730.500000	50.000000	69.000000	9478.500000	6.000000	5.000000	197
75%	1095.250000	70.000000	80.000000	11601.500000	7.000000	6.000000	200
max	1460.000000	190.000000	313.000000	215245.000000	10.000000	9.000000	201

8 rows × 38 columns

In [10]: `df_train['LotFrontage']=df_train['LotFrontage'].fillna(df_train['LotFrontage'])`

In [11]: `df_train.drop(['Id', 'Alley', 'PoolQC', 'Fence', 'MiscFeature', 'FireplaceQu'], axis=1, inplace=True)`

In [12]: `df_train.drop(['Street', 'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandCondition1', 'Condition2', 'BldgType', 'HouseStyle', 'RoofStyle', 'RoofMaterial1', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType2', 'BsmtFinSF2', 'Heating', 'HeatingQC', 'CentralAir', 'Electrical', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual', 'TotRmsAbvGrd', 'Functional', 'GarageType', 'GarageYrBlt', 'GarageFinish', 'GarageCars', 'PoolArea',`

```
'GarageQual','GarageCond','PavedDrive','WoodDeckSF','OpenPorchSF','EnclosedPorch','MiscVal','MoSold','SaleType','SaleCondition'], axis=1, inplace=True)
```

```
In [13]: df_train.select_dtypes(np.number)
```

```
Out[13]:
```

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd
0	60	65.0	8450	7	5	2003	2003
1	20	80.0	9600	6	8	1976	1976
2	60	68.0	11250	7	5	2001	2002
3	70	60.0	9550	7	5	1915	1970
4	60	84.0	14260	8	5	2000	2000
...
1455	60	62.0	7917	6	5	1999	2000
1456	20	85.0	13175	6	6	1978	1988
1457	70	66.0	9042	7	9	1941	2006
1458	20	68.0	9717	5	6	1950	1996
1459	20	75.0	9937	5	6	1965	1965

1460 rows x 19 columns

```
In [14]: df_train['MasVnrArea']=df_train['MasVnrArea'].fillna(df_train['MasVnrArea'].mean())
```

```
In [15]: df_train.MSZoning.value_counts()
```

```
Out[15]:
```

MSZoning	
RL	1151
RM	218
FV	65
RH	16
C (all)	10
Name: count, dtype: int64	

```
In [17]: df_train
```

Out[17]:

	MSSubClass	MSZoning	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearF
0	60	RL	65.0	8450	7	5	2003	
1	20	RL	80.0	9600	6	8	1976	
2	60	RL	68.0	11250	7	5	2001	
3	70	RL	60.0	9550	7	5	1915	
4	60	RL	84.0	14260	8	5	2000	
...
1455	60	RL	62.0	7917	6	5	1999	
1456	20	RL	85.0	13175	6	6	1978	
1457	70	RL	66.0	9042	7	9	1941	
1458	20	RL	68.0	9717	5	6	1950	
1459	20	RL	75.0	9937	5	6	1965	

1460 rows x 20 columns

In [18]: `df_train.isnull().sum()`

Out[18]:

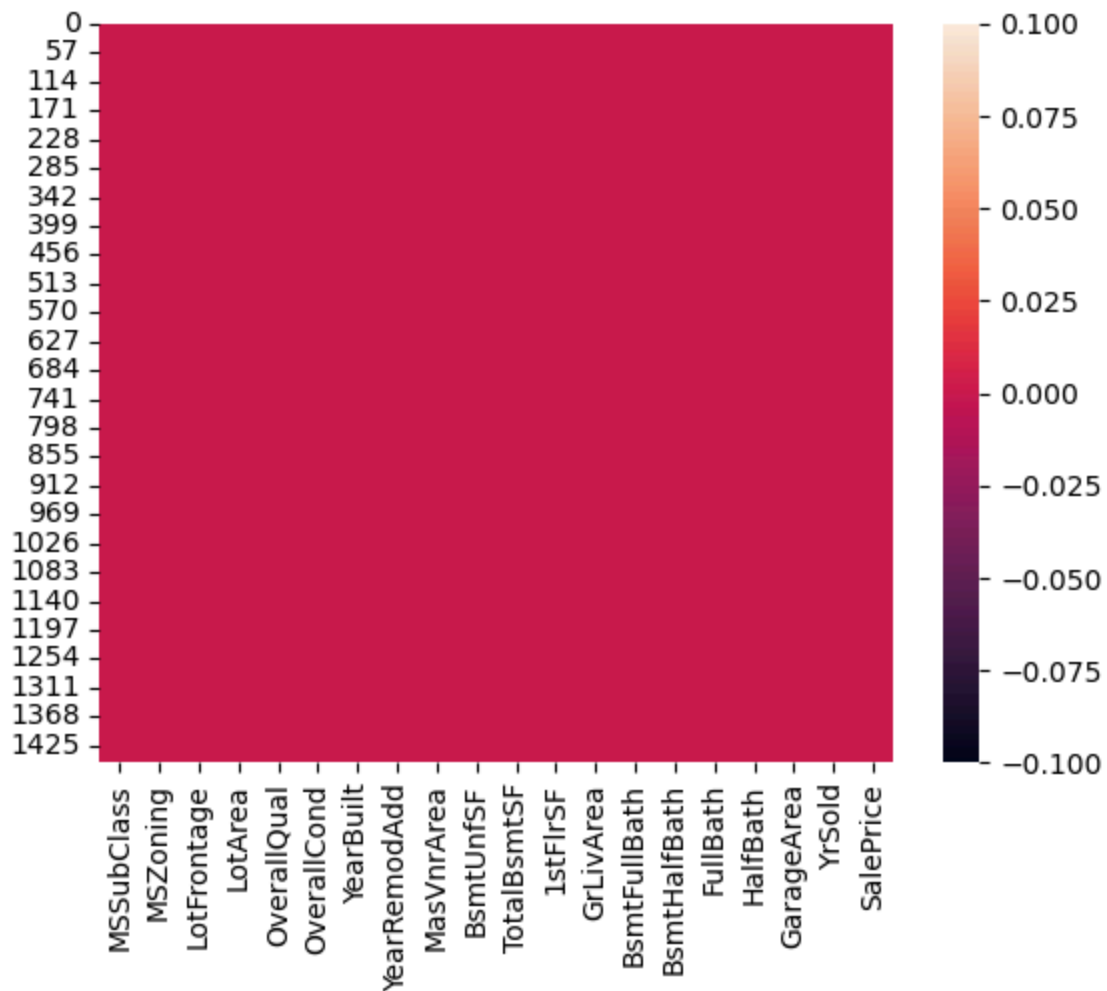
MSSubClass	0
MSZoning	0
LotFrontage	0
LotArea	0
OverallQual	0
OverallCond	0
YearBuilt	0
YearRemodAdd	0
MasVnrArea	0
BsmtUnfSF	0
TotalBsmtSF	0
1stFlrSF	0
GrLivArea	0
BsmtFullBath	0
BsmtHalfBath	0
FullBath	0
HalfBath	0
GarageArea	0
YrSold	0
SalePrice	0

dtype: int64

In [19]: `df_train.shape`

Out[19]: (1460, 20)

In [21]: `sns.heatmap(df_train.isnull())`
`plt.show()`



```
In [23]: from sklearn.model_selection import train_test_split
```

```
X = df_train.drop(['SalePrice'], axis=1)
y = df_train['SalePrice']
```

```
In [24]: X
```

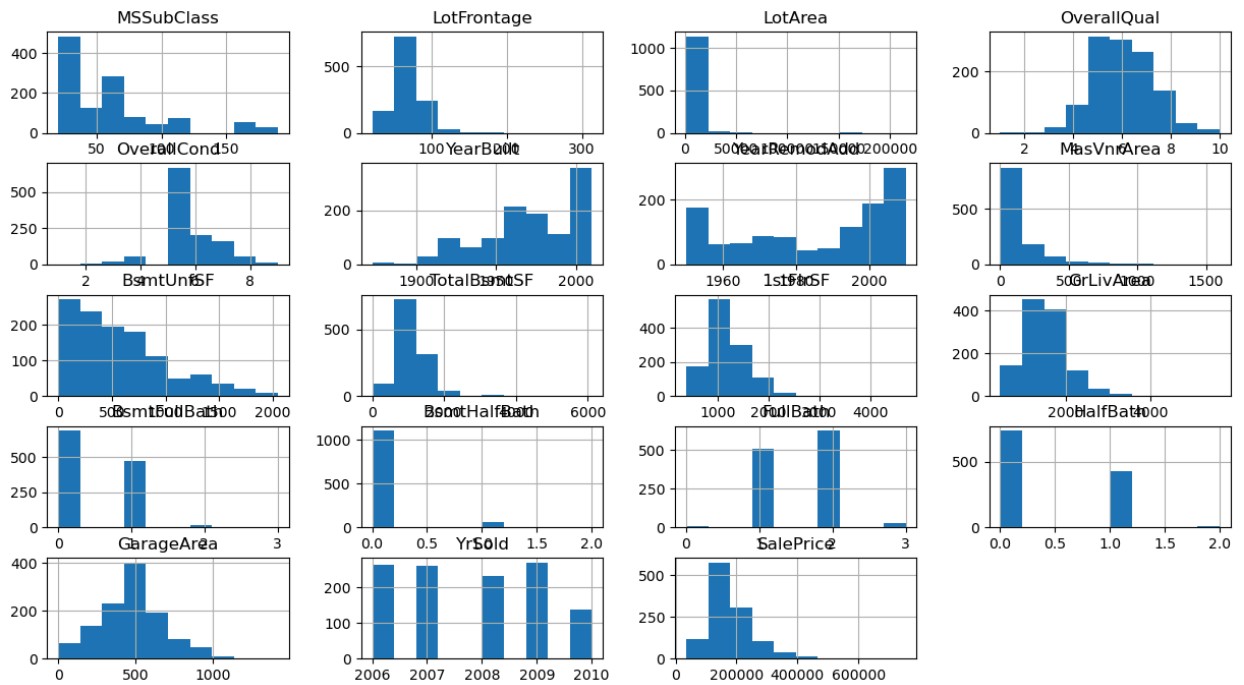
Out [24]:

	MSSubClass	MSZoning	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearF
0	60	RL	65.0	8450	7	5	2003	
1	20	RL	80.0	9600	6	8	1976	
2	60	RL	68.0	11250	7	5	2001	
3	70	RL	60.0	9550	7	5	1915	
4	60	RL	84.0	14260	8	5	2000	
...
1455	60	RL	62.0	7917	6	5	1999	
1456	20	RL	85.0	13175	6	6	1978	
1457	70	RL	66.0	9042	7	9	1941	
1458	20	RL	68.0	9717	5	6	1950	
1459	20	RL	75.0	9937	5	6	1965	

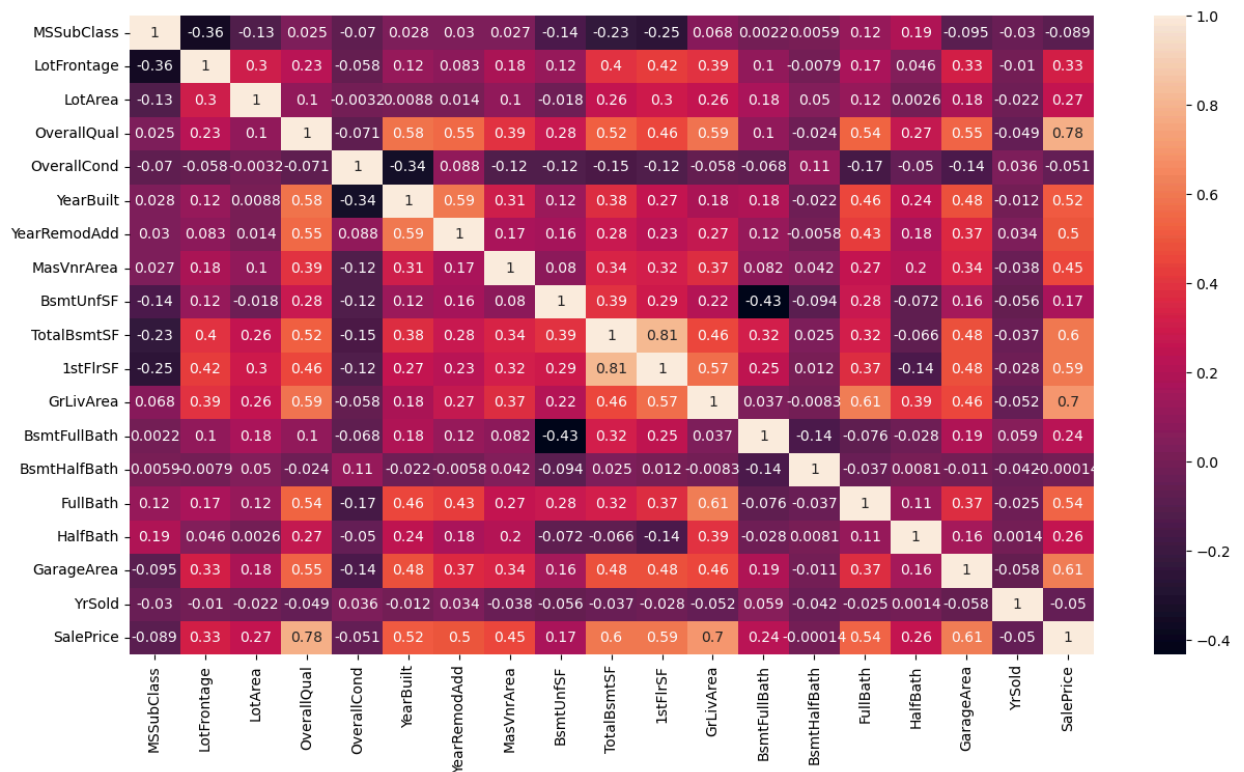
1460 rows x 19 columns

In [25]: `X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)`In [26]: `train_data = X_train.join(y_train)`In [27]: `train_data.hist(figsize=(15,8))`

Out [27]: array([[<Axes: title={'center': 'MSSubClass'}>,
 <Axes: title={'center': 'LotFrontage'}>,
 <Axes: title={'center': 'LotArea'}>,
 <Axes: title={'center': 'OverallQual'}>],
 [<Axes: title={'center': 'OverallCond'}>,
 <Axes: title={'center': 'YearBuilt'}>,
 <Axes: title={'center': 'YearRemodAdd'}>,
 <Axes: title={'center': 'MasVnrArea'}>],
 [<Axes: title={'center': 'BsmtUnfSF'}>,
 <Axes: title={'center': 'TotalBsmtSF'}>,
 <Axes: title={'center': '1stFlrSF'}>,
 <Axes: title={'center': 'GrLivArea'}>],
 [<Axes: title={'center': 'BsmtFullBath'}>,
 <Axes: title={'center': 'BsmtHalfBath'}>,
 <Axes: title={'center': 'FullBath'}>,
 <Axes: title={'center': 'HalfBath'}>],
 [<Axes: title={'center': 'GarageArea'}>,
 <Axes: title={'center': 'YrSold'}>,
 <Axes: title={'center': 'SalePrice'}>], dtype=object)



```
In [28]: plt.figure(figsize=(15,8))
sns.heatmap(train_data.corr(numeric_only=True), annot=True)
plt.show()
```



```
In [29]: train_data = train_data.join(pd.get_dummies(train_data.MSZoning)).drop(['MSZoning'])
```

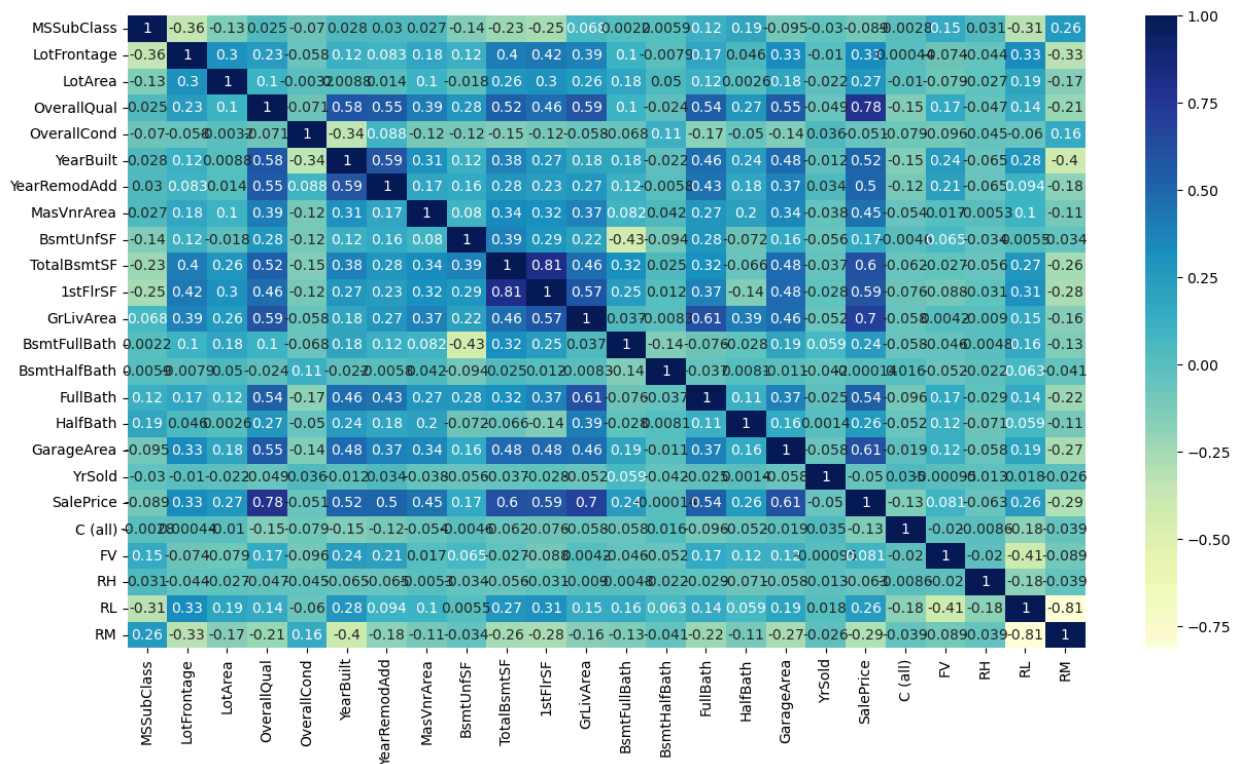
```
In [30]: train_data
```


Out[30]:

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd
1290	80	70.049958	14112	5	7	1964	1964
852	75	53.000000	7128	7	5	1941	1950
463	70	74.000000	11988	6	7	1934	1995
1111	60	80.000000	10480	7	6	1976	1976
1333	50	60.000000	7200	5	6	1938	1995
...
159	60	134.000000	19378	7	5	2005	2006
1259	20	65.000000	9750	6	8	1969	1969
222	60	85.000000	11475	6	6	1975	1975
1402	20	64.000000	6762	7	5	2006	2006
88	50	105.000000	8470	3	2	1915	1982

1168 rows x 24 columns

```
In [31]: plt.figure(figsize=(15,8))
sns.heatmap(train_data.corr(numeric_only=True), annot=True, cmap='YlGnBu')
plt.show()
```



In []:

In []:

In []: