

```
In [246... import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("train.csv")
data.head()
```

```
Out[246...   Id  MSSubClass  MSZoning  LotFrontage  LotArea  Street  Alley  LotShape  LandContour  Utili
0    1           60        RL           65.0    8450    Pave   NaN      Reg          Lvl     All
1    2           20        RL           80.0    9600    Pave   NaN      Reg          Lvl     All
2    3           60        RL           68.0   11250    Pave   NaN      IR1          Lvl     All
3    4           70        RL           60.0    9550    Pave   NaN      IR1          Lvl     All
4    5           60        RL           84.0   14260    Pave   NaN      IR1          Lvl     All
```

5 rows × 81 columns

```
In [247... data.shape #checks total row*column
```

```
Out[247... (1460, 81)
```

```
In [248... data.info() #non-null count, keep columns closer/equal to 1460
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Id                    1460 non-null  int64
1   MSSubClass            1460 non-null  int64
2   MSZoning              1460 non-null  object
3   LotFrontage           1201 non-null  float64
4   LotArea               1460 non-null  int64
5   Street                1460 non-null  object
6   Alley                 91 non-null    object
7   LotShape              1460 non-null  object
8   LandContour           1460 non-null  object
9   Utilities             1460 non-null  object
10  LotConfig             1460 non-null  object
11  LandSlope             1460 non-null  object
12  Neighborhood          1460 non-null  object
13  Condition1            1460 non-null  object
14  Condition2            1460 non-null  object
15  BldgType              1460 non-null  object
16  HouseStyle            1460 non-null  object
17  OverallQual           1460 non-null  int64
18  OverallCond           1460 non-null  int64
19  YearBuilt              1460 non-null  int64
20  YearRemodAdd          1460 non-null  int64
21  RoofStyle             1460 non-null  object
22  RoofMatl              1460 non-null  object
23  Exterior1st           1460 non-null  object
24  Exterior2nd           1460 non-null  object
25  MasVnrType            1452 non-null  object
26  MasVnrArea            1452 non-null  float64
27  ExterQual             1460 non-null  object
28  ExterCond             1460 non-null  object
```

```

29 Foundation      1460 non-null object
30 BsmtQual         1423 non-null object
31 BsmtCond         1423 non-null object
32 BsmtExposure     1422 non-null object
33 BsmtFinType1     1423 non-null object
34 BsmtFinSF1       1460 non-null int64
35 BsmtFinType2     1422 non-null object
36 BsmtFinSF2       1460 non-null int64
37 BsmtUnfSF        1460 non-null int64
38 TotalBsmtSF      1460 non-null int64
39 Heating          1460 non-null object
40 HeatingQC        1460 non-null object
41 CentralAir       1460 non-null object
42 Electrical       1459 non-null object
43 1stFlrSF         1460 non-null int64
44 2ndFlrSF         1460 non-null int64
45 LowQualFinSF     1460 non-null int64
46 GrLivArea        1460 non-null int64
47 BsmtFullBath     1460 non-null int64
48 BsmtHalfBath     1460 non-null int64
49 FullBath         1460 non-null int64
50 HalfBath         1460 non-null int64
51 BedroomAbvGr    1460 non-null int64
52 KitchenAbvGr    1460 non-null int64
53 KitchenQual      1460 non-null object
54 TotRmsAbvGrd    1460 non-null int64
55 Functional       1460 non-null object
56 Fireplaces       1460 non-null int64
57 FireplaceQu      770 non-null object
58 GarageType       1379 non-null object
59 GarageYrBlt      1379 non-null float64
60 GarageFinish     1379 non-null object
61 GarageCars       1460 non-null int64
62 GarageArea       1460 non-null int64
63 GarageQual       1379 non-null object
64 GarageCond       1379 non-null object
65 PavedDrive       1460 non-null object
66 WoodDeckSF       1460 non-null int64
67 OpenPorchSF      1460 non-null int64
68 EnclosedPorch    1460 non-null int64
69 3SsnPorch        1460 non-null int64
70 ScreenPorch      1460 non-null int64
71 PoolArea         1460 non-null int64
72 PoolQC           7 non-null object
73 Fence            281 non-null object
74 MiscFeature       54 non-null object
75 MiscVal          1460 non-null int64
76 MoSold           1460 non-null int64
77 YrSold            1460 non-null int64
78 SaleType         1460 non-null object
79 SaleCondition     1460 non-null object
80 SalePrice        1460 non-null int64
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB

```

```
In [249... # Columns for which null values are greater than 80% are PoolQC, Fence, MiscFeat
```

```
In [250... #create function for calculating missing values percentage > 30%

def missing_value_percentage(data):
    """ A function that returns the columns of the data file and their respective
    percent_missing = (data.isna().sum()/len(data.index))*100
    return percent_missing
```

```
df = missing_value_percentage(data)
print(df)
```

```
Id          0.000000
MSSubClass  0.000000
MSZoning    0.000000
LotFrontage 17.739726
LotArea     0.000000
...
MoSold      0.000000
YrSold      0.000000
SaleType    0.000000
SaleCondition 0.000000
SalePrice   0.000000
Length: 81, dtype: float64
```

```
In [251... import pandas as pd
data = pd.read_csv("train.csv")
df = missing_value_percentage(data)

def greaterThanToDict(n, data):
    """A function thats takes a number n, and creates a dictionary of all column
    that have a null percentage > n"""
    d = {}
    i = 0
    while i < len(data):
        if data[i] > n:
            d[data.index[i]] = data[i]
            i = i + 1

    return d

greaterThanToDict(30, df)
```

```
Out[251... {'Alley': 93.76712328767123,
'FireplaceQu': 47.26027397260274,
'PoolQC': 99.52054794520548,
'Fence': 80.75342465753424,
'MiscFeature': 96.30136986301369}
```

```
In [252... import pandas as pd
data = pd.read_csv("train.csv")
update_data = data.drop(["Alley", "PoolQC", "MiscFeature"], axis=1)

update_data['FireplaceQu'] = update_data['FireplaceQu'].fillna('NoFireplace')
update_data['Fence'] = update_data['Fence'].fillna('NoFence')

print(update_data.shape) #we can see that we dropped 3 columns
print(update_data)
#we have updated our data file to drop 3 columns which had an extremely high nul
#additionally, we updated null values to help create categorical response
```

```
(1460, 78)
   Id  MSSubClass MSZoning  LotFrontage  LotArea  Street  LotShape  \
0    1         60      RL         65.0     8450    Pave      Reg
1    2         20      RL         80.0     9600    Pave      Reg
2    3         60      RL         68.0    11250    Pave      IR1
3    4         70      RL         60.0     9550    Pave      IR1
4    5         60      RL         84.0    14260    Pave      IR1
...  ...      ...      ...      ...      ...      ...
1455 1456         60      RL         62.0     7917    Pave      Reg
1456 1457         20      RL         85.0    13175    Pave      Reg
1457 1458         70      RL         66.0     9042    Pave      Reg
```

1458	1459	20	RL	68.0	9717	Pave	Reg
1459	1460	20	RL	75.0	9937	Pave	Reg

	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	\
0	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
1	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	
2	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
3	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	
4	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	
...	...	...	...	...	...	...	
1455	Lvl	AllPub	Inside	Gtl	Gilbert	Norm	
1456	Lvl	AllPub	Inside	Gtl	NWAmes	Norm	
1457	Lvl	AllPub	Inside	Gtl	Crawfor	Norm	
1458	Lvl	AllPub	Inside	Gtl	NAmes	Norm	
1459	Lvl	AllPub	Inside	Gtl	Edwards	Norm	

	Condition2	BldgType	HouseStyle	OverallQual	OverallCond	YearBuilt	\
0	Norm	1Fam	2Story	7	5	2003	
1	Norm	1Fam	1Story	6	8	1976	
2	Norm	1Fam	2Story	7	5	2001	
3	Norm	1Fam	2Story	7	5	1915	
4	Norm	1Fam	2Story	8	5	2000	
...	...	...	...	...	...	...	
1455	Norm	1Fam	2Story	6	5	1999	
1456	Norm	1Fam	1Story	6	6	1978	
1457	Norm	1Fam	2Story	7	9	1941	
1458	Norm	1Fam	1Story	5	6	1950	
1459	Norm	1Fam	1Story	5	6	1965	

	YearRemodAdd	RoofStyle	RoofMatl	Exterior1st	Exterior2nd	MasVnrType	\
0	2003	Gable	CompShg	VinylSd	VinylSd	BrkFace	
1	1976	Gable	CompShg	MetalSd	MetalSd	None	
2	2002	Gable	CompShg	VinylSd	VinylSd	BrkFace	
3	1970	Gable	CompShg	Wd Sdng	Wd Shng	None	
4	2000	Gable	CompShg	VinylSd	VinylSd	BrkFace	
...	...	...	...	...	...	...	
1455	2000	Gable	CompShg	VinylSd	VinylSd	None	
1456	1988	Gable	CompShg	Plywood	Plywood	Stone	
1457	2006	Gable	CompShg	CemntBd	CmentBd	None	
1458	1996	Hip	CompShg	MetalSd	MetalSd	None	
1459	1965	Gable	CompShg	HdBoard	HdBoard	None	

	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtQual	BsmtCond	\
0	196.0	Gd	TA	PConc	Gd	TA	
1	0.0	TA	TA	CBlock	Gd	TA	
2	162.0	Gd	TA	PConc	Gd	TA	
3	0.0	TA	TA	BrkTil	TA	Gd	
4	350.0	Gd	TA	PConc	Gd	TA	
...	...	...	...	...	...	...	
1455	0.0	TA	TA	PConc	Gd	TA	
1456	119.0	TA	TA	CBlock	Gd	TA	
1457	0.0	Ex	Gd	Stone	TA	Gd	
1458	0.0	TA	TA	CBlock	TA	TA	
1459	0.0	Gd	TA	CBlock	TA	TA	

	BsmtExposure	BsmtFinType1	BsmtFinSF1	BsmtFinType2	BsmtFinSF2	...	\
0	No	GLQ	706	Unf	0	...	
1	Gd	ALQ	978	Unf	0	...	
2	Mn	GLQ	486	Unf	0	...	
3	No	ALQ	216	Unf	0	...	
4	Av	GLQ	655	Unf	0	...	
...	...	...	...	...	...	...	
1455	No	Unf	0	Unf	0	...	
1456	No	ALQ	790	Rec	163	...	
1457	No	GLQ	275	Unf	0	...	

1458	Mn	GLQ	49	Rec	1029	...
1459	No	BLQ	830	LwQ	290	...
	1stFlrSF	2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath	BsmtHalfBath \
0	856	854	0	1710	1	0
1	1262	0	0	1262	0	1
2	920	866	0	1786	1	0
3	961	756	0	1717	1	0
4	1145	1053	0	2198	1	0
...	...	...	...	...	...	...
1455	953	694	0	1647	0	0
1456	2073	0	0	2073	1	0
1457	1188	1152	0	2340	0	0
1458	1078	0	0	1078	1	0
1459	1256	0	0	1256	1	0

	FullBath	HalfBath	BedroomAbvGr	KitchenAbvGr	KitchenQual	\
0	2	1	3	1	Gd	
1	2	0	3	1	TA	
2	2	1	3	1	Gd	
3	1	0	3	1	Gd	
4	2	1	4	1	Gd	
...	...	...	...	...	...	
1455	2	1	3	1	TA	
1456	2	0	3	1	TA	
1457	2	0	4	1	Gd	
1458	1	0	2	1	Gd	
1459	1	1	3	1	TA	

	TotRmsAbvGrd	Functional	Fireplaces	FireplaceQu	GarageType	\
0	8	Typ	0	NoFireplace	Attchd	
1	6	Typ	1	TA	Attchd	
2	6	Typ	1	TA	Attchd	
3	7	Typ	1	Gd	Detchd	
4	9	Typ	1	TA	Attchd	
...	...	...	...	...	...	
1455	7	Typ	1	TA	Attchd	
1456	7	Min1	2	TA	Attchd	
1457	9	Typ	2	Gd	Attchd	
1458	5	Typ	0	NoFireplace	Attchd	
1459	6	Typ	0	NoFireplace	Attchd	

	GarageYrBlt	GarageFinish	GarageCars	GarageArea	GarageQual	GarageCond	\
0	2003.0	RFn	2	548	TA	TA	
1	1976.0	RFn	2	460	TA	TA	
2	2001.0	RFn	2	608	TA	TA	
3	1998.0	Unf	3	642	TA	TA	
4	2000.0	RFn	3	836	TA	TA	
...	...	...	...	...	...	...	
1455	1999.0	RFn	2	460	TA	TA	
1456	1978.0	Unf	2	500	TA	TA	
1457	1941.0	RFn	1	252	TA	TA	
1458	1950.0	Unf	1	240	TA	TA	
1459	1965.0	Fin	1	276	TA	TA	

	PavedDrive	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	\
0	Y	0	61	0	0	0	
1	Y	298	0	0	0	0	
2	Y	0	42	0	0	0	
3	Y	0	35	272	0	0	
4	Y	192	84	0	0	0	
...	...	...	...	...	...	...	
1455	Y	0	40	0	0	0	
1456	Y	349	0	0	0	0	
1457	Y	0	60	0	0	0	

1458	Y	366	0	112	0	0
1459	Y	736	68	0	0	0

	PoolArea	Fence	MiscVal	MoSold	YrSold	SaleType	SaleCondition	\
0	0	NoFence	0	2	2008	WD	Normal	
1	0	NoFence	0	5	2007	WD	Normal	
2	0	NoFence	0	9	2008	WD	Normal	
3	0	NoFence	0	2	2006	WD	Abnorml	
4	0	NoFence	0	12	2008	WD	Normal	
...	...	...	...	...	...	...	...	
1455	0	NoFence	0	8	2007	WD	Normal	
1456	0	MnPrv	0	2	2010	WD	Normal	
1457	0	GdPrv	2500	5	2010	WD	Normal	
1458	0	NoFence	0	4	2010	WD	Normal	
1459	0	NoFence	0	6	2008	WD	Normal	

	SalePrice
0	208500
1	181500
2	223500
3	140000
4	250000
...	...
1455	175000
1456	210000
1457	266500
1458	142125
1459	147500

[1460 rows x 78 columns]

In [253...

```
import pandas as pd
data = pd.read_csv("train.csv")
df = missing_value_percentage(data)

def greaterThanToDict(n, data):
    d = {}
    i = 0
    while i < len(data):
        if data[i] > n:
            d[data.index[i]] = data[i]
        i = i + 1

    return d

update_data = data.drop(["Alley", "PoolQC", "MiscFeature"], axis=1)

update_data['FireplaceQu'] = update_data['FireplaceQu'].fillna('NoFireplace')
update_data['Fence'] = update_data['Fence'].fillna('NoFence')

update_data = update_data.fillna(value='NONE')

lst = []
#loop through all the values of the GarageType column. Add the values to the list
i = 0
while i < len(update_data):
    lst.append(update_data["GarageType"][i])
    i = i + 1

#print("NONE" in lst) #This double checks that our update_data = update_data.fillna('NONE')
#lst = ['Attchd', 'Attchd', 'Attchd' ...]
```

```
#These are the category columns that we can replace null values with "NONE"
noneCols = ['BsmtQual','BsmtCond','BsmtExposure','BsmtFinType1','BsmtFinType2']

for i in noneCols:
    update_data[i] = update_data[i].fillna(value='NONE')

print(update_data)
```

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	LotShape	\
0	1	60	RL	65	8450	Pave	Reg	
1	2	20	RL	80	9600	Pave	Reg	
2	3	60	RL	68	11250	Pave	IR1	
3	4	70	RL	60	9550	Pave	IR1	
4	5	60	RL	84	14260	Pave	IR1	
...	...	...	...	...	...	...	...	
1455	1456	60	RL	62	7917	Pave	Reg	
1456	1457	20	RL	85	13175	Pave	Reg	
1457	1458	70	RL	66	9042	Pave	Reg	
1458	1459	20	RL	68	9717	Pave	Reg	
1459	1460	20	RL	75	9937	Pave	Reg	

	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	\
0	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
1	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	
2	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
3	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	
4	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	
...	...	...	...	...	...	...	
1455	Lvl	AllPub	Inside	Gtl	Gilbert	Norm	
1456	Lvl	AllPub	Inside	Gtl	NWAmes	Norm	
1457	Lvl	AllPub	Inside	Gtl	Crawfor	Norm	
1458	Lvl	AllPub	Inside	Gtl	NAMES	Norm	
1459	Lvl	AllPub	Inside	Gtl	Edwards	Norm	

	Condition2	BldgType	HouseStyle	OverallQual	OverallCond	YearBuilt	\
0	Norm	1Fam	2Story	7	5	2003	
1	Norm	1Fam	1Story	6	8	1976	
2	Norm	1Fam	2Story	7	5	2001	
3	Norm	1Fam	2Story	7	5	1915	
4	Norm	1Fam	2Story	8	5	2000	
...	...	...	...	...	...	...	
1455	Norm	1Fam	2Story	6	5	1999	
1456	Norm	1Fam	1Story	6	6	1978	
1457	Norm	1Fam	2Story	7	9	1941	
1458	Norm	1Fam	1Story	5	6	1950	
1459	Norm	1Fam	1Story	5	6	1965	

	YearRemodAdd	RoofStyle	RoofMatl	Exterior1st	Exterior2nd	MasVnrType	\
0	2003	Gable	CompShg	VinylSd	VinylSd	BrkFace	
1	1976	Gable	CompShg	MetalSd	MetalSd	None	
2	2002	Gable	CompShg	VinylSd	VinylSd	BrkFace	
3	1970	Gable	CompShg	Wd Sdng	Wd Shng	None	
4	2000	Gable	CompShg	VinylSd	VinylSd	BrkFace	
...	...	...	...	...	...	...	
1455	2000	Gable	CompShg	VinylSd	VinylSd	None	
1456	1988	Gable	CompShg	Plywood	Plywood	Stone	
1457	2006	Gable	CompShg	CemntBd	CmentBd	None	
1458	1996	Hip	CompShg	MetalSd	MetalSd	None	
1459	1965	Gable	CompShg	HdBoard	HdBoard	None	

	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtQual	BsmtCond	BsmtExposure	\
0	196	Gd	TA	PConc	Gd	TA	No	
1	0	TA	TA	CBlock	Gd	TA	Gd	
2	162	Gd	TA	PConc	Gd	TA	Mn	

3	0	TA	TA	BrkTil	TA	Gd	No
4	350	Gd	TA	PConc	Gd	TA	Av
...	...	...	...	...	...	...	...
1455	0	TA	TA	PConc	Gd	TA	No
1456	119	TA	TA	CBlock	Gd	TA	No
1457	0	Ex	Gd	Stone	TA	Gd	No
1458	0	TA	TA	CBlock	TA	TA	Mn
1459	0	Gd	TA	CBlock	TA	TA	No

	BsmtFinType1	BsmtFinSF1	BsmtFinType2	BsmtFinSF2	...	1stFlrSF	\
0	GLQ	706	Unf	0	...	856	
1	ALQ	978	Unf	0	...	1262	
2	GLQ	486	Unf	0	...	920	
3	ALQ	216	Unf	0	...	961	
4	GLQ	655	Unf	0	...	1145	
...	...	...	...	...	...	...	
1455	Unf	0	Unf	0	...	953	
1456	ALQ	790	Rec	163	...	2073	
1457	GLQ	275	Unf	0	...	1188	
1458	GLQ	49	Rec	1029	...	1078	
1459	BLQ	830	LwQ	290	...	1256	

	2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath	BsmtHalfBath	FullBath	\
0	854	0	1710	1	0	2	
1	0	0	1262	0	1	2	
2	866	0	1786	1	0	2	
3	756	0	1717	1	0	1	
4	1053	0	2198	1	0	2	
...	...	...	...	...	...	...	
1455	694	0	1647	0	0	2	
1456	0	0	2073	1	0	2	
1457	1152	0	2340	0	0	2	
1458	0	0	1078	1	0	1	
1459	0	0	1256	1	0	1	

	HalfBath	BedroomAbvGr	KitchenAbvGr	KitchenQual	TotRmsAbvGrd	\
0	1	3	1	Gd	8	
1	0	3	1	TA	6	
2	1	3	1	Gd	6	
3	0	3	1	Gd	7	
4	1	4	1	Gd	9	
...	...	...	...	...	...	
1455	1	3	1	TA	7	
1456	0	3	1	TA	7	
1457	0	4	1	Gd	9	
1458	0	2	1	Gd	5	
1459	1	3	1	TA	6	

	Functional	Fireplaces	FireplaceQu	GarageType	GarageYrBlt	\
0	Typ	0	NoFireplace	Attchd	2003	
1	Typ	1	TA	Attchd	1976	
2	Typ	1	TA	Attchd	2001	
3	Typ	1	Gd	Detchd	1998	
4	Typ	1	TA	Attchd	2000	
...	...	...	...	...	...	
1455	Typ	1	TA	Attchd	1999	
1456	Min1	2	TA	Attchd	1978	
1457	Typ	2	Gd	Attchd	1941	
1458	Typ	0	NoFireplace	Attchd	1950	
1459	Typ	0	NoFireplace	Attchd	1965	

	GarageFinish	GarageCars	GarageArea	GarageQual	GarageCond	PavedDrive	\
0	RFn	2	548	TA	TA	Y	
1	RFn	2	460	TA	TA	Y	
2	RFn	2	608	TA	TA	Y	



3	Unf	3	642	TA	TA	Y
4	Rfn	3	836	TA	TA	Y
...	...	...	...	...	...	...
1455	Rfn	2	460	TA	TA	Y
1456	Unf	2	500	TA	TA	Y
1457	Rfn	1	252	TA	TA	Y
1458	Unf	1	240	TA	TA	Y
1459	Fin	1	276	TA	TA	Y

	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	\
0	0	61	0	0	0	0	
1	298	0	0	0	0	0	
2	0	42	0	0	0	0	
3	0	35	272	0	0	0	
4	192	84	0	0	0	0	
...	...	...	...	...	...	...	
1455	0	40	0	0	0	0	
1456	349	0	0	0	0	0	
1457	0	60	0	0	0	0	
1458	366	0	112	0	0	0	
1459	736	68	0	0	0	0	

	Fence	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	NoFence	0	2	2008	WD	Normal	208500
1	NoFence	0	5	2007	WD	Normal	181500
2	NoFence	0	9	2008	WD	Normal	223500
3	NoFence	0	2	2006	WD	Abnorml	140000
4	NoFence	0	12	2008	WD	Normal	250000
...	...	...	...	...	...	...	...
1455	NoFence	0	8	2007	WD	Normal	175000
1456	MnPrv	0	2	2010	WD	Normal	210000
1457	GdPrv	2500	5	2010	WD	Normal	266500
1458	NoFence	0	4	2010	WD	Normal	142125
1459	NoFence	0	6	2008	WD	Normal	147500

[1460 rows x 78 columns]

In [254...

```
def calnullpercentage(df):
    print(df.isnull().sum()/len(df.index) *100)

calnullpercentage(update_data)
```

```
Id          0.0
MSSubClass  0.0
MSZoning    0.0
LotFrontage 0.0
LotArea     0.0
...
MoSold      0.0
YrSold      0.0
SaleType    0.0
SaleCondition 0.0
SalePrice   0.0
Length: 78, dtype: float64
```

In [255...

```
update_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 78 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Id              1460 non-null  int64
1   MSSubClass      1460 non-null  int64
2   MSZoning        1460 non-null  object
```

3	LotFrontage	1460	non-null	object
4	LotArea	1460	non-null	int64
5	Street	1460	non-null	object
6	LotShape	1460	non-null	object
7	LandContour	1460	non-null	object
8	Utilities	1460	non-null	object
9	LotConfig	1460	non-null	object
10	LandSlope	1460	non-null	object
11	Neighborhood	1460	non-null	object
12	Condition1	1460	non-null	object
13	Condition2	1460	non-null	object
14	BldgType	1460	non-null	object
15	HouseStyle	1460	non-null	object
16	OverallQual	1460	non-null	int64
17	OverallCond	1460	non-null	int64
18	YearBuilt	1460	non-null	int64
19	YearRemodAdd	1460	non-null	int64
20	RoofStyle	1460	non-null	object
21	RoofMatl	1460	non-null	object
22	Exterior1st	1460	non-null	object
23	Exterior2nd	1460	non-null	object
24	MasVnrType	1460	non-null	object
25	MasVnrArea	1460	non-null	object
26	ExterQual	1460	non-null	object
27	ExterCond	1460	non-null	object
28	Foundation	1460	non-null	object
29	BsmtQual	1460	non-null	object
30	BsmtCond	1460	non-null	object
31	BsmtExposure	1460	non-null	object
32	BsmtFinType1	1460	non-null	object
33	BsmtFinSF1	1460	non-null	int64
34	BsmtFinType2	1460	non-null	object
35	BsmtFinSF2	1460	non-null	int64
36	BsmtUnfSF	1460	non-null	int64
37	TotalBsmtSF	1460	non-null	int64
38	Heating	1460	non-null	object
39	HeatingQC	1460	non-null	object
40	CentralAir	1460	non-null	object
41	Electrical	1460	non-null	object
42	1stFlrSF	1460	non-null	int64
43	2ndFlrSF	1460	non-null	int64
44	LowQualFinSF	1460	non-null	int64
45	GrLivArea	1460	non-null	int64
46	BsmtFullBath	1460	non-null	int64
47	BsmtHalfBath	1460	non-null	int64
48	FullBath	1460	non-null	int64
49	HalfBath	1460	non-null	int64
50	BedroomAbvGr	1460	non-null	int64
51	KitchenAbvGr	1460	non-null	int64
52	KitchenQual	1460	non-null	object
53	TotRmsAbvGrd	1460	non-null	int64
54	Functional	1460	non-null	object
55	Fireplaces	1460	non-null	int64
56	FireplaceQu	1460	non-null	object
57	GarageType	1460	non-null	object
58	GarageYrBlt	1460	non-null	object
59	GarageFinish	1460	non-null	object
60	GarageCars	1460	non-null	int64
61	GarageArea	1460	non-null	int64
62	GarageQual	1460	non-null	object
63	GarageCond	1460	non-null	object
64	PavedDrive	1460	non-null	object
65	WoodDeckSF	1460	non-null	int64
66	OpenPorchSF	1460	non-null	int64
67	EnclosedPorch	1460	non-null	int64

```

68 3SsnPorch      1460 non-null  int64
69 ScreenPorch   1460 non-null  int64
70 PoolArea      1460 non-null  int64
71 Fence         1460 non-null  object
72 MiscVal       1460 non-null  int64
73 MoSold        1460 non-null  int64
74 YrSold        1460 non-null  int64
75 SaleType      1460 non-null  object
76 SaleCondition 1460 non-null  object
77 SalePrice     1460 non-null  int64

```

```
dtypes: int64(35), object(43)
```

```
memory usage: 889.8+ KB
```

```
In [256... update_data.head(10)
```

```
Out[256...
   Id  MSSubClass  MSZoning  LotFrontage  LotArea  Street  LotShape  LandContour  Utilities  L
0    1           60        RL           65     8450    Pave      Reg          Lvl        AllPub
1    2           20        RL           80     9600    Pave      Reg          Lvl        AllPub
2    3           60        RL           68    11250    Pave      IR1          Lvl        AllPub
3    4           70        RL           60     9550    Pave      IR1          Lvl        AllPub
4    5           60        RL           84    14260    Pave      IR1          Lvl        AllPub
5    6           50        RL           85    14115    Pave      IR1          Lvl        AllPub
6    7           20        RL           75    10084    Pave      Reg          Lvl        AllPub
7    8           60        RL          NONE    10382    Pave      IR1          Lvl        AllPub
8    9           50        RM           51     6120    Pave      Reg          Lvl        AllPub
9   10          190        RL           50     7420    Pave      Reg          Lvl        AllPub

```

10 rows × 78 columns

```
In [257... update_data['ScreenPorch'].value_counts()
#majority of data in ScreenPorch column is 0. Do not think there is enough data
```

```
Out[257...
0      1344
192        6
224         5
120         5
189         4
...
182         1
440         1
178         1
312         1
480         1
Name: ScreenPorch, Length: 76, dtype: int64
```

```
In [258... update_data['3SsnPorch'].value_counts()
#majority of data in 3SsnPorch column is 0. Do not think there is enough data to
```

```
Out[258...
0      1436
168         3
216         2
144         2
180         2
245         1
```

```

238      1
290      1
196      1
182      1
407      1
304      1
162      1
153      1
320      1
140      1
130      1
96       1
23       1
508      1
Name: 3SsnPorch, dtype: int64

```

```

In [259... update_data['PoolArea'].value_counts()
#majority of data in PoolArea column is 0. Do not think there is enough data to

```

```

Out[259... 0      1453
738      1
648      1
576      1
555      1
519      1
512      1
480      1
Name: PoolArea, dtype: int64

```

```

In [260... update_data['MiscVal'].value_counts()
#majority of data in MiscVal column is 0. Do not think there is enough data to c

```

```

Out[260... 0      1408
400      11
500       8
700       5
450       4
2000      4
600       4
1200       2
480       2
1150      1
800       1
15500     1
620       1
3500      1
560       1
2500      1
1300      1
1400      1
350       1
8300      1
54        1
Name: MiscVal, dtype: int64

```

```

In [261... #drop 4 columns which don't have enough meaningful data for analysis
data1 = update_data.drop(['3SsnPorch', 'ScreenPorch', 'PoolArea', 'MiscVal', 'Id

```

```

In [262... print(data1)

```

```

      MSSubClass MSZoning LotFrontage  LotArea Street LotShape LandContour \
0             60      RL           65     8450   Pave      Reg         Lvl
1             20      RL           80     9600   Pave      Reg         Lvl
2             60      RL           68    11250   Pave     IR1         Lvl

```

3	70	RL	60	9550	Pave	IR1	Lvl
4	60	RL	84	14260	Pave	IR1	Lvl
...	...	...	...	...	...	...	...
1455	60	RL	62	7917	Pave	Reg	Lvl
1456	20	RL	85	13175	Pave	Reg	Lvl
1457	70	RL	66	9042	Pave	Reg	Lvl
1458	20	RL	68	9717	Pave	Reg	Lvl
1459	20	RL	75	9937	Pave	Reg	Lvl

	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	Condition2	\
0	AllPub	Inside	Gtl	CollgCr	Norm	Norm	
1	AllPub	FR2	Gtl	Veenker	Feedr	Norm	
2	AllPub	Inside	Gtl	CollgCr	Norm	Norm	
3	AllPub	Corner	Gtl	Crawfor	Norm	Norm	
4	AllPub	FR2	Gtl	NoRidge	Norm	Norm	
...	...	...	...	...	...	...	
1455	AllPub	Inside	Gtl	Gilbert	Norm	Norm	
1456	AllPub	Inside	Gtl	NWAmes	Norm	Norm	
1457	AllPub	Inside	Gtl	Crawfor	Norm	Norm	
1458	AllPub	Inside	Gtl	NAmes	Norm	Norm	
1459	AllPub	Inside	Gtl	Edwards	Norm	Norm	

	BldgType	HouseStyle	OverallQual	OverallCond	YearBuilt	YearRemodAdd	\
0	1Fam	2Story	7	5	2003	2003	
1	1Fam	1Story	6	8	1976	1976	
2	1Fam	2Story	7	5	2001	2002	
3	1Fam	2Story	7	5	1915	1970	
4	1Fam	2Story	8	5	2000	2000	
...	...	...	...	...	...	...	
1455	1Fam	2Story	6	5	1999	2000	
1456	1Fam	1Story	6	6	1978	1988	
1457	1Fam	2Story	7	9	1941	2006	
1458	1Fam	1Story	5	6	1950	1996	
1459	1Fam	1Story	5	6	1965	1965	

	RoofStyle	RoofMatl	Exterior1st	Exterior2nd	MasVnrType	MasVnrArea	\
0	Gable	CompShg	VinylSd	VinylSd	BrkFace	196	
1	Gable	CompShg	Metalsd	Metalsd	None	0	
2	Gable	CompShg	VinylSd	VinylSd	BrkFace	162	
3	Gable	CompShg	Wd Sdng	Wd Shng	None	0	
4	Gable	CompShg	VinylSd	VinylSd	BrkFace	350	
...	...	...	...	...	...	...	
1455	Gable	CompShg	VinylSd	VinylSd	None	0	
1456	Gable	CompShg	Plywood	Plywood	Stone	119	
1457	Gable	CompShg	CemntBd	CmentBd	None	0	
1458	Hip	CompShg	Metalsd	Metalsd	None	0	
1459	Gable	CompShg	HdBoard	HdBoard	None	0	

	ExterQual	ExterCond	Foundation	BsmtQual	BsmtCond	BsmtExposure	\
0	Gd	TA	PConc	Gd	TA	No	
1	TA	TA	CBlock	Gd	TA	Gd	
2	Gd	TA	PConc	Gd	TA	Mn	
3	TA	TA	BrkTil	TA	Gd	No	
4	Gd	TA	PConc	Gd	TA	Av	
...	...	...	...	...	...	...	
1455	TA	TA	PConc	Gd	TA	No	
1456	TA	TA	CBlock	Gd	TA	No	
1457	Ex	Gd	Stone	TA	Gd	No	
1458	TA	TA	CBlock	TA	TA	Mn	
1459	Gd	TA	CBlock	TA	TA	No	

	BsmtFinType1	BsmtFinSF1	BsmtFinType2	BsmtFinSF2	BsmtUnfSF	\
0	GLQ	706	Unf	0	150	
1	ALQ	978	Unf	0	284	
2	GLQ	486	Unf	0	434	

3	ALQ	216	Unf	0	540
4	GLQ	655	Unf	0	490
...	...	...	...	...	...
1455	Unf	0	Unf	0	953
1456	ALQ	790	Rec	163	589
1457	GLQ	275	Unf	0	877
1458	GLQ	49	Rec	1029	0
1459	BLQ	830	LwQ	290	136

	TotalBsmtSF	Heating	HeatingQC	CentralAir	Electrical	1stFlrSF	2ndFlrSF	\
0	856	GasA	Ex	Y	SBrkr	856	854	
1	1262	GasA	Ex	Y	SBrkr	1262	0	
2	920	GasA	Ex	Y	SBrkr	920	866	
3	756	GasA	Gd	Y	SBrkr	961	756	
4	1145	GasA	Ex	Y	SBrkr	1145	1053	
...	...	...	...	...	...	...	...	
1455	953	GasA	Ex	Y	SBrkr	953	694	
1456	1542	GasA	TA	Y	SBrkr	2073	0	
1457	1152	GasA	Ex	Y	SBrkr	1188	1152	
1458	1078	GasA	Gd	Y	FuseA	1078	0	
1459	1256	GasA	Gd	Y	SBrkr	1256	0	

	LowQualFinSF	GrLivArea	BsmtFullBath	BsmtHalfBath	FullBath	HalfBath	\
0	0	1710	1	0	2	1	
1	0	1262	0	1	2	0	
2	0	1786	1	0	2	1	
3	0	1717	1	0	1	0	
4	0	2198	1	0	2	1	
...	...	...	...	...	...	...	
1455	0	1647	0	0	2	1	
1456	0	2073	1	0	2	0	
1457	0	2340	0	0	2	0	
1458	0	1078	1	0	1	0	
1459	0	1256	1	0	1	1	

	BedroomAbvGr	KitchenAbvGr	KitchenQual	TotRmsAbvGrd	Functional	\
0	3	1	Gd	8	Typ	
1	3	1	TA	6	Typ	
2	3	1	Gd	6	Typ	
3	3	1	Gd	7	Typ	
4	4	1	Gd	9	Typ	
...	...	...	...	...	...	
1455	3	1	TA	7	Typ	
1456	3	1	TA	7	Min1	
1457	4	1	Gd	9	Typ	
1458	2	1	Gd	5	Typ	
1459	3	1	TA	6	Typ	

	Fireplaces	FireplaceQu	GarageType	GarageYrBlt	GarageFinish	GarageCars	\
0	0	NoFireplace	Attchd	2003	RFn	2	
1	1	TA	Attchd	1976	RFn	2	
2	1	TA	Attchd	2001	RFn	2	
3	1	Gd	Detchd	1998	Unf	3	
4	1	TA	Attchd	2000	RFn	3	
...	...	...	...	...	...	...	
1455	1	TA	Attchd	1999	RFn	2	
1456	2	TA	Attchd	1978	Unf	2	
1457	2	Gd	Attchd	1941	RFn	1	
1458	0	NoFireplace	Attchd	1950	Unf	1	
1459	0	NoFireplace	Attchd	1965	Fin	1	

	GarageArea	GarageQual	GarageCond	PavedDrive	WoodDeckSF	OpenPorchSF	\
0	548	TA	TA	Y	0	61	
1	460	TA	TA	Y	298	0	
2	608	TA	TA	Y	0	42	

3	642	TA	TA	Y	0	35
4	836	TA	TA	Y	192	84
...	...	...	...	...	...	...
1455	460	TA	TA	Y	0	40
1456	500	TA	TA	Y	349	0
1457	252	TA	TA	Y	0	60
1458	240	TA	TA	Y	366	0
1459	276	TA	TA	Y	736	68

	EnclosedPorch	Fence	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	0	NoFence	2	2008	WD	Normal	208500
1	0	NoFence	5	2007	WD	Normal	181500
2	0	NoFence	9	2008	WD	Normal	223500
3	272	NoFence	2	2006	WD	Abnorml	140000
4	0	NoFence	12	2008	WD	Normal	250000
...	...	...	...	...	...	...	...
1455	0	NoFence	8	2007	WD	Normal	175000
1456	0	MnPrv	2	2010	WD	Normal	210000
1457	0	GdPrv	5	2010	WD	Normal	266500
1458	112	NoFence	4	2010	WD	Normal	142125
1459	0	NoFence	6	2008	WD	Normal	147500

[1460 rows x 73 columns]

In [263... *#create copy of data1*

```
data2 = data1.copy()
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 73 columns):
#   Column                Non-Null Count  Dtype
---  -
0   MSSubClass             1460 non-null   int64
1   MSZoning                1460 non-null   object
2   LotFrontage            1460 non-null   object
3   LotArea                 1460 non-null   int64
4   Street                  1460 non-null   object
5   LotShape                1460 non-null   object
6   LandContour             1460 non-null   object
7   Utilities               1460 non-null   object
8   LotConfig               1460 non-null   object
9   LandSlope               1460 non-null   object
10  Neighborhood            1460 non-null   object
11  Condition1              1460 non-null   object
12  Condition2              1460 non-null   object
13  BldgType                1460 non-null   object
14  HouseStyle              1460 non-null   object
15  OverallQual             1460 non-null   int64
16  OverallCond             1460 non-null   int64
17  YearBuilt               1460 non-null   int64
18  YearRemodAdd            1460 non-null   int64
19  RoofStyle               1460 non-null   object
20  RoofMatl                1460 non-null   object
21  Exterior1st             1460 non-null   object
22  Exterior2nd             1460 non-null   object
23  MasVnrType              1460 non-null   object
24  MasVnrArea              1460 non-null   object
25  ExterQual               1460 non-null   object
26  ExterCond               1460 non-null   object
27  Foundation              1460 non-null   object
28  BsmtQual                1460 non-null   object
29  BsmtCond                1460 non-null   object
30  BsmtExposure            1460 non-null   object
```

```

31 BsmtFinType1    1460 non-null    object
32 BsmtFinSF1     1460 non-null    int64
33 BsmtFinType2   1460 non-null    object
34 BsmtFinSF2     1460 non-null    int64
35 BsmtUnfSF      1460 non-null    int64
36 TotalBsmtSF    1460 non-null    int64
37 Heating        1460 non-null    object
38 HeatingQC      1460 non-null    object
39 CentralAir     1460 non-null    object
40 Electrical     1460 non-null    object
41 1stFlrSF       1460 non-null    int64
42 2ndFlrSF       1460 non-null    int64
43 LowQualFinSF   1460 non-null    int64
44 GrLivArea      1460 non-null    int64
45 BsmtFullBath   1460 non-null    int64
46 BsmtHalfBath   1460 non-null    int64
47 FullBath       1460 non-null    int64
48 HalfBath       1460 non-null    int64
49 BedroomAbvGr  1460 non-null    int64
50 KitchenAbvGr  1460 non-null    int64
51 KitchenQual    1460 non-null    object
52 TotRmsAbvGrd  1460 non-null    int64
53 Functional     1460 non-null    object
54 Fireplaces     1460 non-null    int64
55 FireplaceQu    1460 non-null    object
56 GarageType     1460 non-null    object
57 GarageYrBlt    1460 non-null    object
58 GarageFinish   1460 non-null    object
59 GarageCars     1460 non-null    int64
60 GarageArea     1460 non-null    int64
61 GarageQual     1460 non-null    object
62 GarageCond     1460 non-null    object
63 PavedDrive     1460 non-null    object
64 WoodDeckSF     1460 non-null    int64
65 OpenPorchSF    1460 non-null    int64
66 EnclosedPorch  1460 non-null    int64
67 Fence          1460 non-null    object
68 MoSold         1460 non-null    int64
69 YrSold         1460 non-null    int64
70 SaleType       1460 non-null    object
71 SaleCondition  1460 non-null    object
72 SalePrice      1460 non-null    int64

```

dtypes: int64(30), object(43)

memory usage: 832.8+ KB

In [264...

```

#numeric columns
display(data2.select_dtypes("int64"))
pd.options.display.max_columns = 73

```

	MSSubClass	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	BsmtFinSF1	Bs
0	60	8450	7	5	2003	2003	706	
1	20	9600	6	8	1976	1976	978	
2	60	11250	7	5	2001	2002	486	
3	70	9550	7	5	1915	1970	216	
4	60	14260	8	5	2000	2000	655	
...	...	...	...	...	...	...	...	
1455	60	7917	6	5	1999	2000	0	
1456	20	13175	6	6	1978	1988	790	



	MSSubClass	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	BsmtFinSF1	Bs
1457	70	9042	7	9	1941	2006	275	
1458	20	9717	5	6	1950	1996	49	
1459	20	9937	5	6	1965	1965	830	

1460 rows × 30 columns

```
In [265... convert_to_categ = ['MSSubClass', 'OverallQual', 'OverallCond', 'BsmtFullBath', '
data2[convert_to_categ] = data2[convert_to_categ].astype('object')
```

```
In [266... data2.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 73 columns):
#   Column                Non-Null Count  Dtype
---  -
0   MSSubClass            1460 non-null   object
1   MSZoning              1460 non-null   object
2   LotFrontage          1460 non-null   object
3   LotArea               1460 non-null   int64
4   Street               1460 non-null   object
5   LotShape              1460 non-null   object
6   LandContour          1460 non-null   object
7   Utilities             1460 non-null   object
8   LotConfig            1460 non-null   object
9   LandSlope            1460 non-null   object
10  Neighborhood          1460 non-null   object
11  Condition1            1460 non-null   object
12  Condition2           1460 non-null   object
13  BldgType              1460 non-null   object
14  HouseStyle            1460 non-null   object
15  OverallQual           1460 non-null   object
16  OverallCond           1460 non-null   object
17  YearBuilt             1460 non-null   int64
18  YearRemodAdd          1460 non-null   int64
19  RoofStyle            1460 non-null   object
20  RoofMatl             1460 non-null   object
21  Exterior1st          1460 non-null   object
22  Exterior2nd          1460 non-null   object
23  MasVnrType           1460 non-null   object
24  MasVnrArea           1460 non-null   object
25  ExterQual             1460 non-null   object
26  ExterCond            1460 non-null   object
27  Foundation           1460 non-null   object
28  BsmtQual             1460 non-null   object
29  BsmtCond             1460 non-null   object
30  BsmtExposure         1460 non-null   object
31  BsmtFinType1         1460 non-null   object
32  BsmtFinSF1           1460 non-null   int64
33  BsmtFinType2         1460 non-null   object
34  BsmtFinSF2           1460 non-null   int64
35  BsmtUnfSF            1460 non-null   int64
36  TotalBsmtSF          1460 non-null   int64
37  Heating              1460 non-null   object
38  HeatingQC            1460 non-null   object
39  CentralAir           1460 non-null   object
40  Electrical            1460 non-null   object
41  1stFlrSF             1460 non-null   int64
```

```

42  2ndFlrSF          1460 non-null   int64
43  LowQualFinSF      1460 non-null   int64
44  GrLivArea         1460 non-null   int64
45  BsmtFullBath      1460 non-null   object
46  BsmtHalfBath      1460 non-null   object
47  FullBath          1460 non-null   object
48  HalfBath          1460 non-null   object
49  BedroomAbvGr      1460 non-null   object
50  KitchenAbvGr      1460 non-null   object
51  KitchenQual       1460 non-null   object
52  TotRmsAbvGrd      1460 non-null   object
53  Functional        1460 non-null   object
54  Fireplaces        1460 non-null   object
55  FireplaceQu       1460 non-null   object
56  GarageType        1460 non-null   object
57  GarageYrBlt       1460 non-null   object
58  GarageFinish      1460 non-null   object
59  GarageCars        1460 non-null   object
60  GarageArea        1460 non-null   int64
61  GarageQual        1460 non-null   object
62  GarageCond        1460 non-null   object
63  PavedDrive        1460 non-null   object
64  WoodDeckSF        1460 non-null   int64
65  OpenPorchSF       1460 non-null   int64
66  EnclosedPorch     1460 non-null   int64
67  Fence            1460 non-null   object
68  MoSold            1460 non-null   int64
69  YrSold            1460 non-null   int64
70  SaleType          1460 non-null   object
71  SaleCondition     1460 non-null   object
72  SalePrice         1460 non-null   int64

```

dtypes: int64(18), object(55)

memory usage: 832.8+ KB

In [ ]:

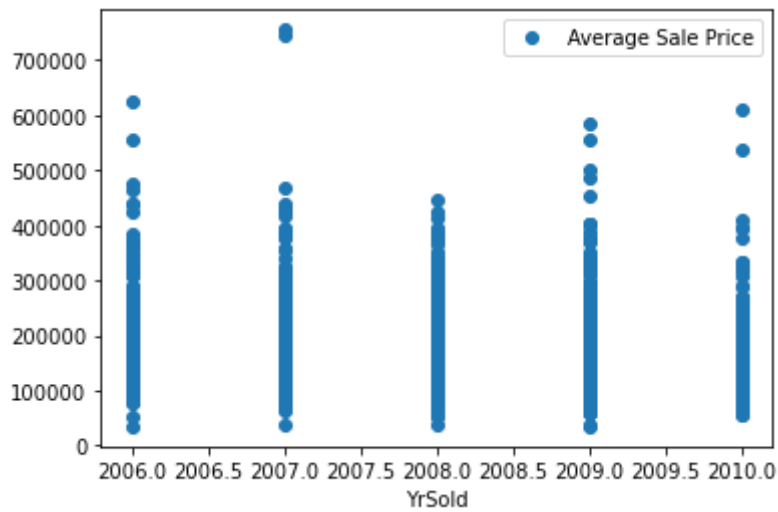
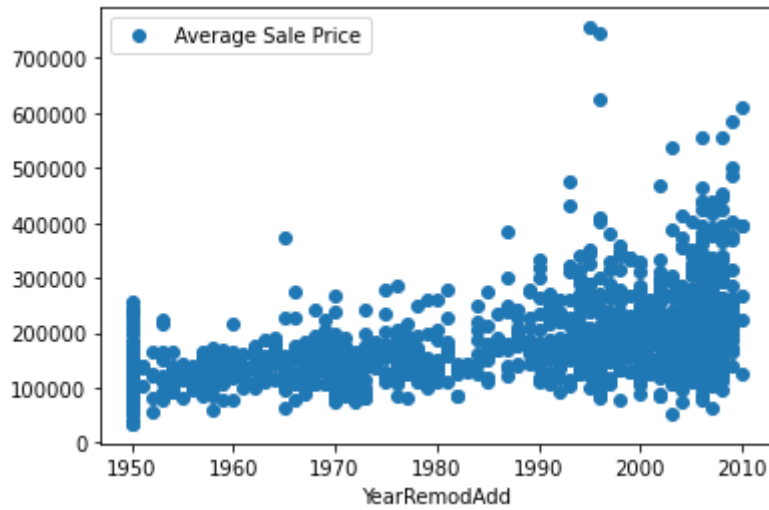
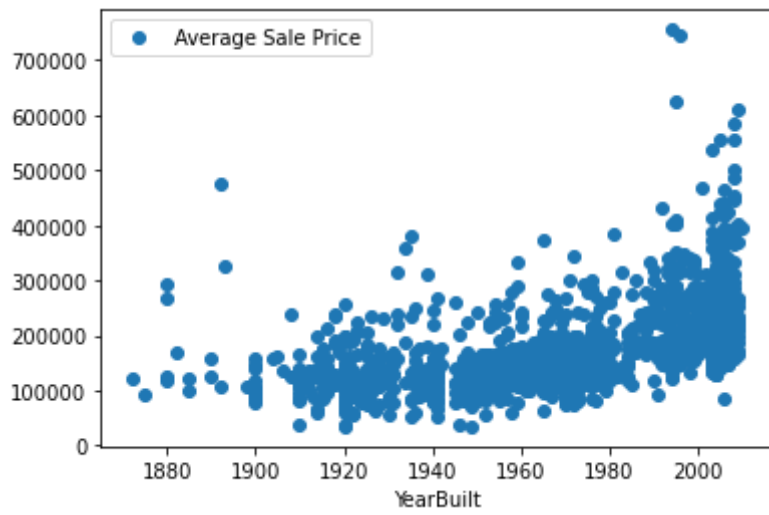
In [274...

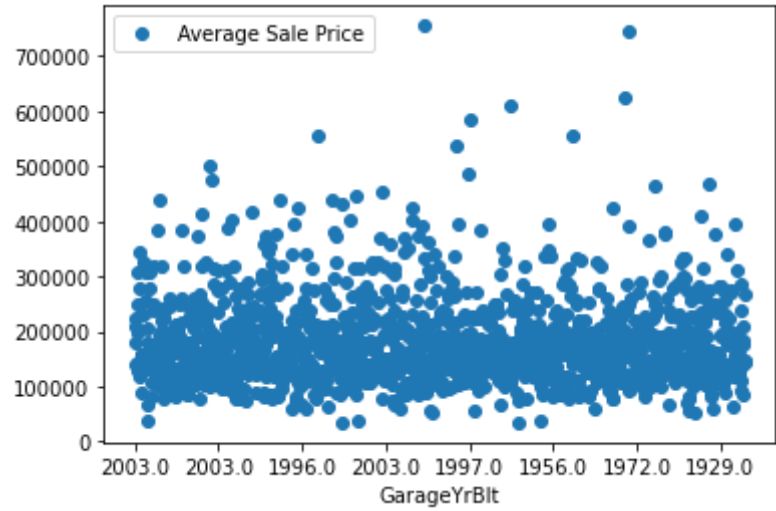
*#Year Features*

```

col = ['YearBuilt', 'YearRemodAdd', 'YrSold', 'GarageYrBlt']
i = 1
for col_name in col:
    d = {col_name : data2[col_name], 'Average Sale Price': data2['SalePrice']}
    df = pd.DataFrame(d)
    df.plot(col_name, y=['Average Sale Price'], style = 'o')
    #y = data2['SalePrice']
    #data2[x, y].plot( x = 'col_name', y = 'Average Sale Price')
    #df.plot(x= x_data, y='Average Sale Price', style='o')
    i += 1
plt.show

```





In [ ]:

In [ ]: