In [3]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#prep
from sklearn.model_selection import train_test_split, GridSearchCV, Ran
from sklearn.preprocessing import LabelEncoder, StandardScaler, MaxAbsS

#models
from sklearn.linear_model import LogisticRegression, LogisticRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor

#validation libraries
#from sklearn.cross_validation import KFold, StratifiedKFold
from IPython.display import display
from sklearn import metrics

```

In [4]:
```python
df_train = pd.read_csv('train.csv')
#df_test =  pd.read_csv('test.csv')
```

In [5]:
```python
#frames = [df_train,df_test ]
#result = pd.concat(frames, sort = True)
```

In [6]:
```python
df_train.head()
```

Out[6]:

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

5 rows × 81 columns

In [7]:
```python
df_train['MSZoning'].value_counts() # categorical values
```

Out[7]:
```
RL         1151
RM          218
FV           65
RH           16
C (all)      10
Name: MSZoning, dtype: int64
```

In [70]:
```python
#plt.scatter_plot()
```

```
In [9]:   1   df_train['MSZoning'].isnull()
```

```
Out[9]:  0        False
         1        False
         2        False
         3        False
         4        False
         5        False
         6        False
         7        False
         8        False
         9        False
         10       False
         11       False
         12       False
         13       False
         14       False
         15       False
         16       False
         17       False
         18       False
         19       False
         20       False
         21       False
         22       False
         23       False
         24       False
         25       False
         26       False
         27       False
         28       False
         29       False
                  ...
         1430     False
         1431     False
         1432     False
         1433     False
         1434     False
         1435     False
         1436     False
         1437     False
         1438     False
         1439     False
         1440     False
         1441     False
         1442     False
         1443     False
         1444     False
         1445     False
         1446     False
         1447     False
         1448     False
         1449     False
         1450     False
         1451     False
         1452     False
         1453     False
```

```
1454     False
1455     False
1456     False
1457     False
1458     False
1459     False
Name: MSZoning, Length: 1460, dtype: bool
```

In [10]:
```
1  df_train.shape
```

Out[10]: (1460, 81)

In [11]:   1 df_train.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
Id                1460 non-null int64
MSSubClass        1460 non-null int64
MSZoning          1460 non-null object
LotFrontage       1201 non-null float64
LotArea           1460 non-null int64
Street            1460 non-null object
Alley             91 non-null object
LotShape          1460 non-null object
LandContour       1460 non-null object
Utilities         1460 non-null object
LotConfig         1460 non-null object
LandSlope         1460 non-null object
Neighborhood      1460 non-null object
Condition1        1460 non-null object
Condition2        1460 non-null object
BldgType          1460 non-null object
HouseStyle        1460 non-null object
OverallQual       1460 non-null int64
OverallCond       1460 non-null int64
YearBuilt         1460 non-null int64
YearRemodAdd      1460 non-null int64
RoofStyle         1460 non-null object
RoofMatl          1460 non-null object
Exterior1st       1460 non-null object
Exterior2nd       1460 non-null object
MasVnrType        1452 non-null object
MasVnrArea        1452 non-null float64
ExterQual         1460 non-null object
ExterCond         1460 non-null object
Foundation        1460 non-null object
BsmtQual          1423 non-null object
BsmtCond          1423 non-null object
BsmtExposure      1422 non-null object
BsmtFinType1      1423 non-null object
BsmtFinSF1        1460 non-null int64
BsmtFinType2      1422 non-null object
BsmtFinSF2        1460 non-null int64
BsmtUnfSF         1460 non-null int64
TotalBsmtSF       1460 non-null int64
Heating           1460 non-null object
HeatingQC         1460 non-null object
CentralAir        1460 non-null object
Electrical        1459 non-null object
1stFlrSF          1460 non-null int64
2ndFlrSF          1460 non-null int64
LowQualFinSF      1460 non-null int64
GrLivArea         1460 non-null int64
BsmtFullBath      1460 non-null int64
BsmtHalfBath      1460 non-null int64
FullBath          1460 non-null int64
HalfBath          1460 non-null int64
BedroomAbvGr      1460 non-null int64
```

```
KitchenAbvGr       1460 non-null int64
KitchenQual        1460 non-null object
TotRmsAbvGrd       1460 non-null int64
Functional         1460 non-null object
Fireplaces         1460 non-null int64
FireplaceQu        770 non-null object
GarageType         1379 non-null object
GarageYrBlt        1379 non-null float64
GarageFinish       1379 non-null object
GarageCars         1460 non-null int64
GarageArea         1460 non-null int64
GarageQual         1379 non-null object
GarageCond         1379 non-null object
PavedDrive         1460 non-null object
WoodDeckSF         1460 non-null int64
OpenPorchSF        1460 non-null int64
EnclosedPorch      1460 non-null int64
3SsnPorch          1460 non-null int64
ScreenPorch        1460 non-null int64
PoolArea           1460 non-null int64
PoolQC             7 non-null object
Fence              281 non-null object
MiscFeature        54 non-null object
MiscVal            1460 non-null int64
MoSold             1460 non-null int64
YrSold             1460 non-null int64
SaleType           1460 non-null object
SaleCondition      1460 non-null object
SalePrice          1460 non-null int64
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB
```

```
In [12]:  1  df_train.columns
```

```
Out[12]:  Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Stree
          t',
                 'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
                 'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgTyp
          e',
                 'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemo
          dAdd',
                 'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrTyp
          e',
                 'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
                 'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
                 'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heatin
          g',
                 'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
                 'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'Full
          Bath',
                 'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
                 'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'Garage
          Type',
                 'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'Garage
          Qual',
                 'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
                 'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
                 'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
                 'SaleCondition', 'SalePrice'],
                dtype='object')
```

```
In [13]:  1  len(df_train.columns)
```

```
Out[13]:  81
```

In [14]:
```python
for x,y in zip (df_train.dtypes.index, df_train.dtypes):
    print(" {} || {} " .format(x,y))
```

```
Id ||  int64
MSSubClass ||  int64
MSZoning ||  object
LotFrontage ||  float64
LotArea ||  int64
Street ||  object
Alley ||  object
LotShape ||  object
LandContour ||  object
Utilities ||  object
LotConfig ||  object
LandSlope ||  object
Neighborhood ||  object
Condition1 ||  object
Condition2 ||  object
BldgType ||  object
HouseStyle ||  object
OverallQual ||  int64
OverallCond ||  int64
YearBuilt ||  int64
YearRemodAdd ||  int64
RoofStyle ||  object
RoofMatl ||  object
Exterior1st ||  object
Exterior2nd ||  object
MasVnrType ||  object
MasVnrArea ||  float64
ExterQual ||  object
ExterCond ||  object
Foundation ||  object
BsmtQual ||  object
BsmtCond ||  object
BsmtExposure ||  object
BsmtFinType1 ||  object
BsmtFinSF1 ||  int64
BsmtFinType2 ||  object
BsmtFinSF2 ||  int64
BsmtUnfSF ||  int64
TotalBsmtSF ||  int64
Heating ||  object
HeatingQC ||  object
CentralAir ||  object
Electrical ||  object
1stFlrSF ||  int64
2ndFlrSF ||  int64
LowQualFinSF ||  int64
GrLivArea ||  int64
BsmtFullBath ||  int64
BsmtHalfBath ||  int64
FullBath ||  int64
HalfBath ||  int64
BedroomAbvGr ||  int64
KitchenAbvGr ||  int64
KitchenQual ||  object
```

```
TotRmsAbvGrd ||   int64
Functional ||   object
Fireplaces ||   int64
FireplaceQu ||   object
GarageType ||   object
GarageYrBlt ||   float64
GarageFinish ||   object
GarageCars ||   int64
GarageArea ||   int64
GarageQual ||   object
GarageCond ||   object
PavedDrive ||   object
WoodDeckSF ||   int64
OpenPorchSF ||   int64
EnclosedPorch ||   int64
3SsnPorch ||   int64
ScreenPorch ||   int64
PoolArea ||   int64
PoolQC ||   object
Fence ||   object
MiscFeature ||   object
MiscVal ||   int64
MoSold ||   int64
YrSold ||   int64
SaleType ||   object
SaleCondition ||   object
SalePrice ||   int64
```

```
In [15]:    1  df_train.isna().sum().sort_values(ascending=False) / len(df_train) * 10
```

```
Out[15]:  PoolQC          99.520548
          MiscFeature     96.301370
          Alley           93.767123
          Fence           80.753425
          FireplaceQu     47.260274
          LotFrontage     17.739726
          GarageCond       5.547945
          GarageType       5.547945
          GarageYrBlt      5.547945
          GarageFinish     5.547945
          GarageQual       5.547945
          BsmtExposure     2.602740
          BsmtFinType2     2.602740
          BsmtFinType1     2.534247
          BsmtCond         2.534247
          BsmtQual         2.534247
          MasVnrArea       0.547945
          MasVnrType       0.547945
          Electrical       0.068493
          Utilities        0.000000
          YearRemodAdd     0.000000
          MSSubClass       0.000000
          Foundation       0.000000
          ExterCond        0.000000
          ExterQual        0.000000
          Exterior2nd      0.000000
          Exterior1st      0.000000
          RoofMatl         0.000000
          RoofStyle        0.000000
          YearBuilt        0.000000
                             ...
          GarageArea       0.000000
          PavedDrive       0.000000
          WoodDeckSF       0.000000
          OpenPorchSF      0.000000
          3SsnPorch        0.000000
          BsmtUnfSF        0.000000
          ScreenPorch      0.000000
          PoolArea         0.000000
          MiscVal          0.000000
          MoSold           0.000000
          YrSold           0.000000
          SaleType         0.000000
          Functional       0.000000
          TotRmsAbvGrd     0.000000
          KitchenQual      0.000000
          KitchenAbvGr     0.000000
          BedroomAbvGr     0.000000
          HalfBath         0.000000
          FullBath         0.000000
          BsmtHalfBath     0.000000
          BsmtFullBath     0.000000
          GrLivArea        0.000000
          LowQualFinSF     0.000000
          2ndFlrSF         0.000000
```

```
1stFlrSF          0.000000
CentralAir        0.000000
SaleCondition     0.000000
Heating           0.000000
TotalBsmtSF       0.000000
Id                0.000000
Length: 81, dtype: float64
```

In [44]:   1   #null_columns.shape

```
In [17]:   1  null_columns=df_train.columns[df_train.isnull().any()]
           2  df_train[null_columns].isnull().sum()
           3
           4  labels = []
           5  values = []
           6  for col in null_columns:
           7      labels.append(col)
           8      values.append(df_train[col].isnull().sum())
           9  ind = np.arange(len(labels))
          10  width = 0.1
          11  fig, ax = plt.subplots(figsize=(12,10))
          12  rects = ax.barh(ind, np.array(values), color='violet')
          13  ax.set_yticks(ind+((width)/2.))
          14  ax.set_yticklabels(labels, rotation='horizontal')
          15  ax.set_xlabel("Count of missing values")
          16  ax.set_ylabel("Column Names")
          17  ax.set_title("Variables with missing values");
          18
```
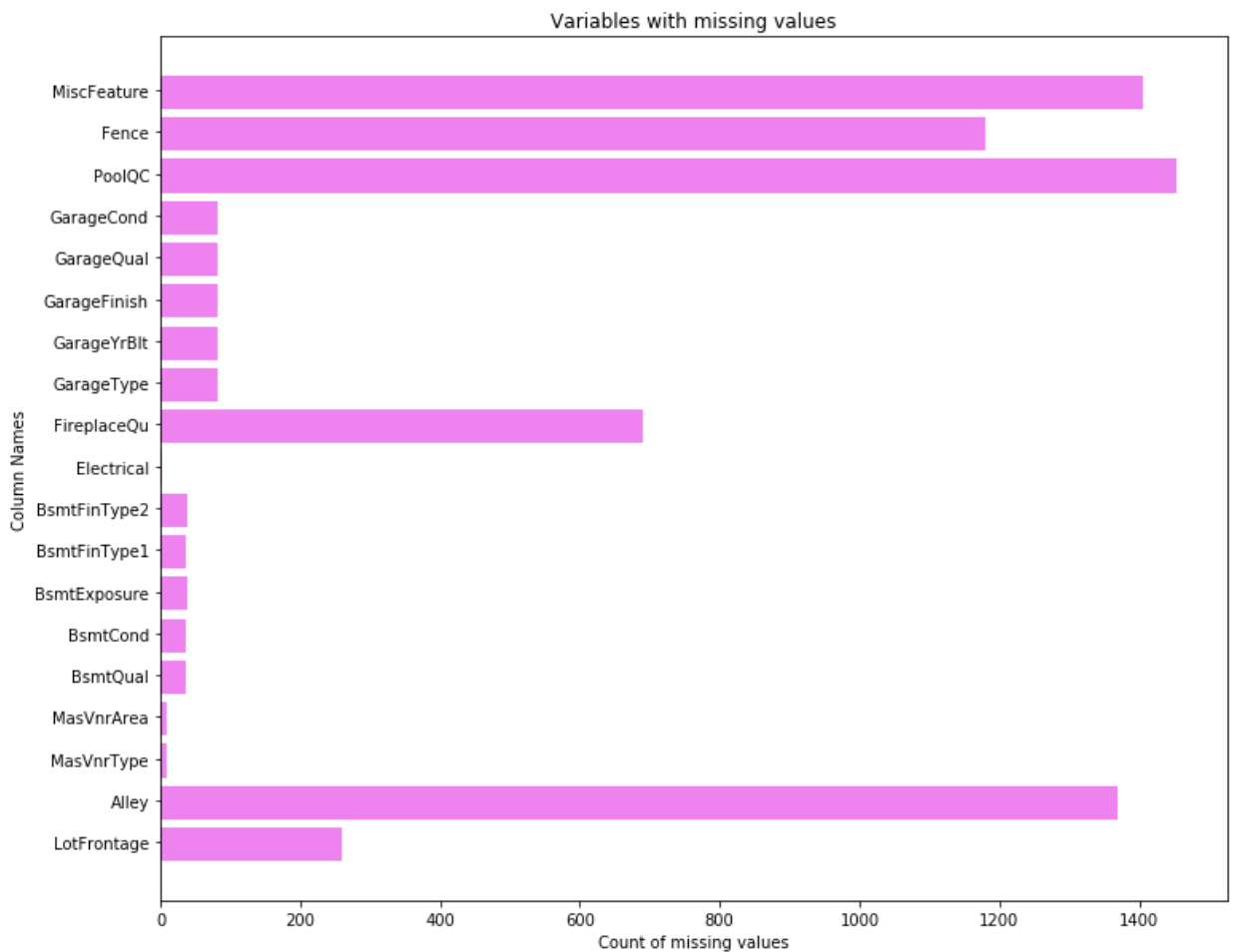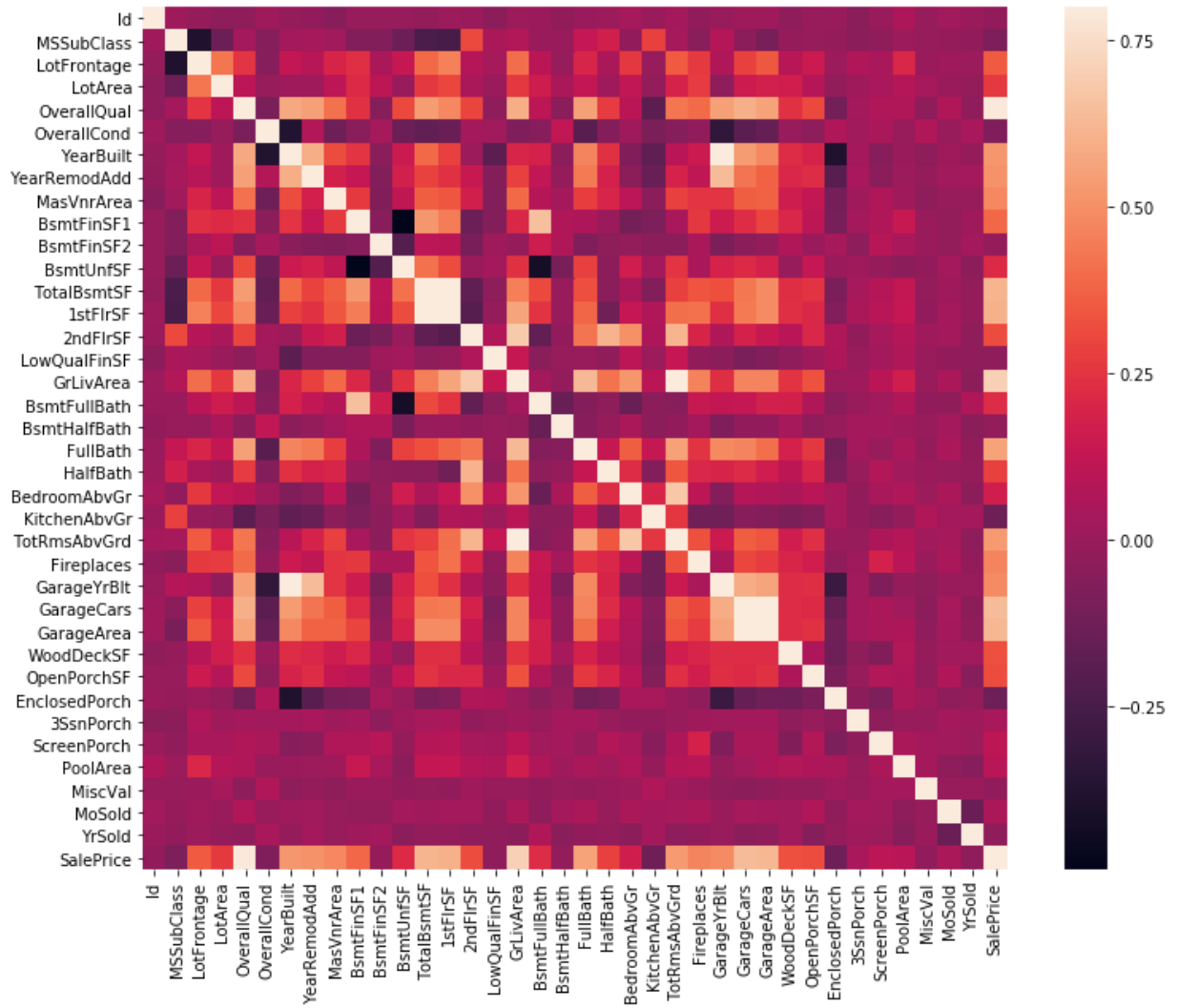


```
In [18]:   1  #df_train#.isna().sum().sort_values(ascending=False) / len(df_train) *
```
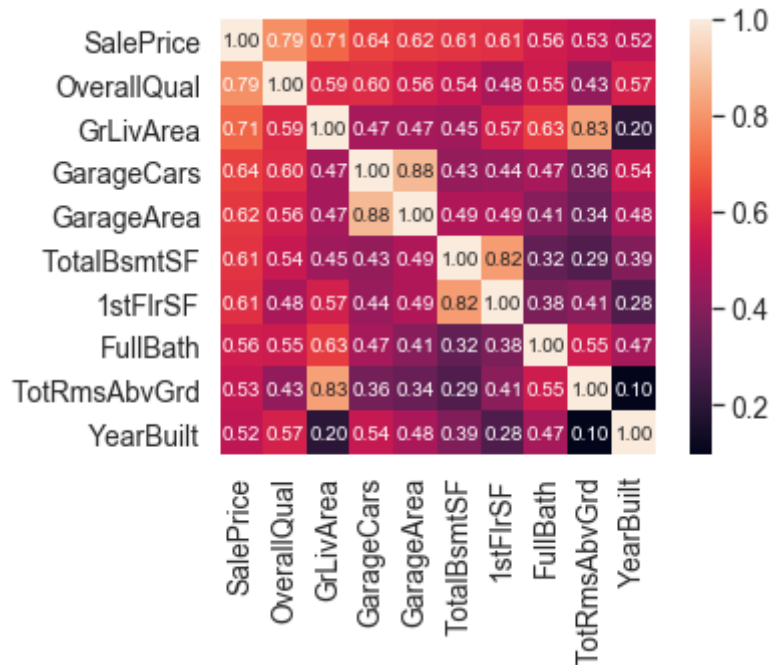
In [19]:
```python
corrmat = df_train.corr()
f, ax = plt.subplots(figsize=(13, 10))
sns.heatmap(corrmat, vmax=.8, square=True);
```

In [20]:
```python
#saleprice correlation matrix
k = 10 #number of variables for heatmap
cols = corrmat.nlargest(k, 'SalePrice')['SalePrice'].index
cm = np.corrcoef(df_train[cols].values.T)
sns.set(font_scale=1.25)
hm = sns.heatmap(cm, cbar=True, annot=True, square=True, fmt='.2f', ann
plt.show()
```

In [21]:
```python
#plotting correlations
num_feat=df_train.columns[df_train.dtypes!=object]
num_feat=num_feat[1:-1]
labels = []
values = []
for col in num_feat:
    labels.append(col)
    values.append(np.corrcoef(df_train[col].values, df_train.SalePrice.

ind = np.arange(len(labels))
width = 0.5
fig, ax = plt.subplots(figsize=(12,15))
rects = ax.barh(ind, np.array(values), color='red')
ax.set_yticks(ind+((width)/2.))
ax.set_yticklabels(labels, rotation='horizontal')
ax.set_xlabel("Correlation coefficient")
ax.set_title("Correlation Coefficients w.r.t Sale Price");
```

Correlation Coefficients w.r.t Sale Price
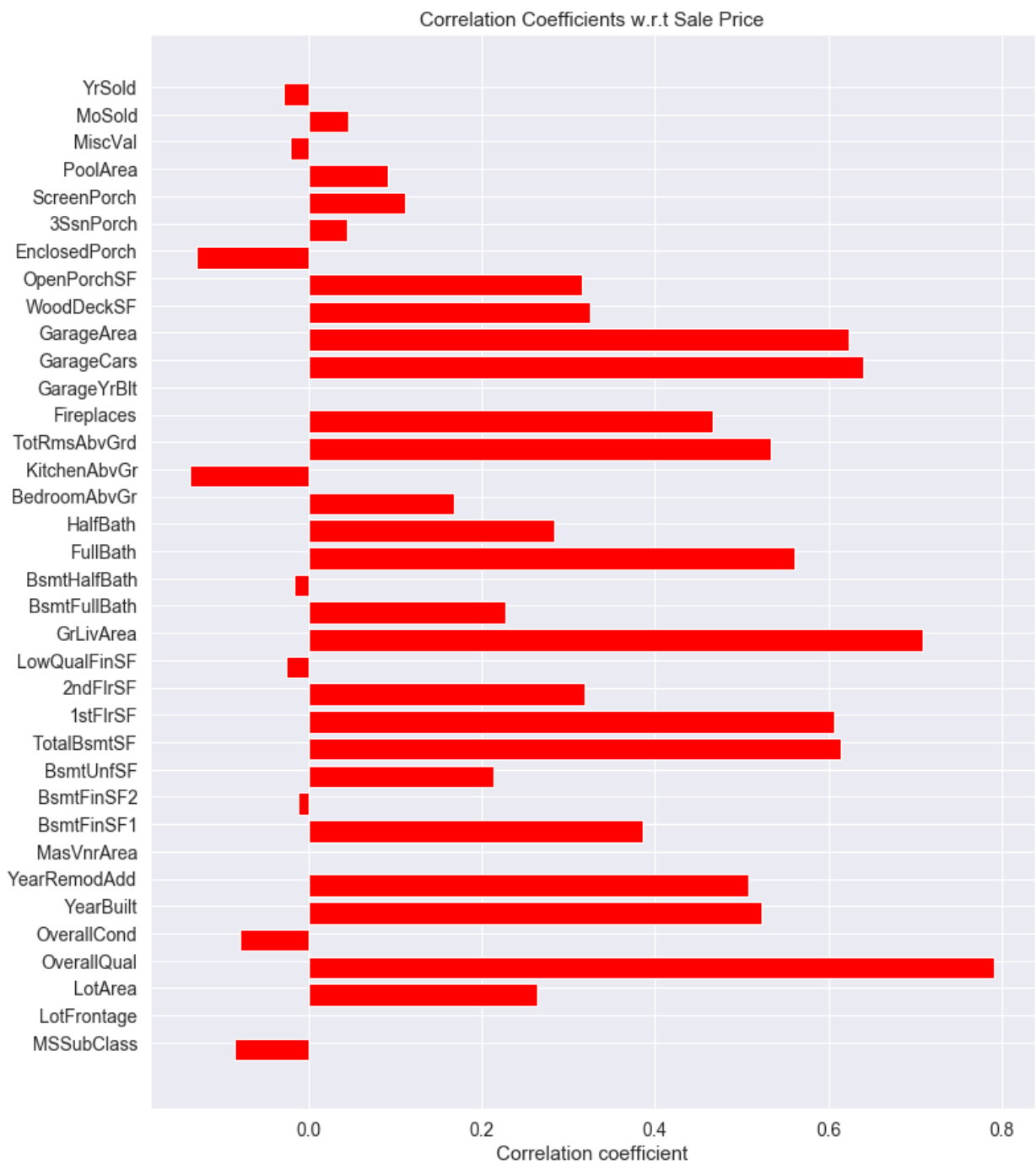
In [22]:

```python
correlations=df_train.corr()
attrs = correlations.iloc[:-1,:-1] # all except target

threshold = 0.5
important_corrs = (attrs[abs(attrs) > threshold][attrs != 1.0]) \
    .unstack().dropna().to_dict()

unique_important_corrs = pd.DataFrame(
    list(set([(tuple(sorted(key)), important_corrs[key]) \
    for key in important_corrs])),
        columns=['Attribute Pair', 'Correlation'])

    # sorted by absolute value
unique_important_corrs = unique_important_corrs.ix[
    abs(unique_important_corrs['Correlation']).argsort()[::-1]]

unique_important_corrs
```

```
/Users/shradhitsubudhi/anaconda3/envs/python37charm/lib/python3.7/site-pa
ckages/ipykernel_launcher.py:14: DeprecationWarning:
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-d
eprecated (http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-i
ndexer-is-deprecated)
```

Out[22]:

	Attribute Pair	Correlation
25	(GarageArea, GarageCars)	0.882475
4	(GarageYrBlt, YearBuilt)	0.825667
11	(GrLivArea, TotRmsAbvGrd)	0.825489
5	(1stFlrSF, TotalBsmtSF)	0.819530
16	(2ndFlrSF, GrLivArea)	0.687501
27	(BedroomAbvGr, TotRmsAbvGrd)	0.676620
13	(BsmtFinSF1, BsmtFullBath)	0.649212
0	(GarageYrBlt, YearRemodAdd)	0.642277
21	(FullBath, GrLivArea)	0.630012
12	(2ndFlrSF, TotRmsAbvGrd)	0.616423
1	(2ndFlrSF, HalfBath)	0.609707
9	(GarageCars, OverallQual)	0.600671
19	(GrLivArea, OverallQual)	0.593007
26	(YearBuilt, YearRemodAdd)	0.592855
24	(GarageCars, GarageYrBlt)	0.588920

	Attribute Pair	Correlation
20	(OverallQual, YearBuilt)	0.572323
23	(1stFlrSF, GrLivArea)	0.566024
7	(GarageArea, GarageYrBlt)	0.564567
15	(GarageArea, OverallQual)	0.562022
17	(FullBath, TotRmsAbvGrd)	0.554784
6	(OverallQual, YearRemodAdd)	0.550684
14	(FullBath, OverallQual)	0.550600
2	(GarageYrBlt, OverallQual)	0.547766
22	(GarageCars, YearBuilt)	0.537850
8	(OverallQual, TotalBsmtSF)	0.537808
10	(BsmtFinSF1, TotalBsmtSF)	0.522396
18	(BedroomAbvGr, GrLivArea)	0.521270
3	(2ndFlrSF, BedroomAbvGr)	0.502901

In [23]:

```
 1  df_train['LotArea_norm'] = df_train['LotArea']
 2
 3  ss = StandardScaler()
 4  mas = MaxAbsScaler()
 5  qs = QuantileTransformer()
 6
 7  df_train['LotArea_norm'] = ss.fit_transform(df_train[['LotArea']])
 8  df_train['LotArea_mas'] = mas.fit_transform(df_train[['LotArea']])
 9  df_train['LotArea_qs'] = qs.fit_transform(df_train[['LotArea']])
10
11
12  df_train[['LotArea_norm','LotArea_mas','LotArea_qs', 'LotArea']].head(5
```

Out[23]:

	LotArea_norm	LotArea_mas	LotArea_qs	LotArea
0	-0.207142	0.039258	0.349349	8450
1	-0.091886	0.044600	0.525025	9600
2	0.073480	0.052266	0.720220	11250
3	-0.096897	0.044368	0.509667	9550
4	0.375148	0.066250	0.897123	14260

In [24]:
```python
1  plt.figure(figsize=(8, 6))
2  sns.distplot(df_train['SalePrice'], color="r", kde=False, )
3  plt.title("Distribution of Sale Price")
4  plt.ylabel("Number of Occurences")
5
6  plt.xlabel("Sale Price");
```


In [25]:
```python
1  sns.distplot?
```

```
In [26]:    1  plt.scatter(df_train["TotalBsmtSF"],df_train["SalePrice"])
            2  plt.title("Total Basement area in Square Feet Vs SalePrice ")
            3  plt.ylabel("SalePrice")
            4  plt.xlabel("Total Basement area in Square Feet");
```
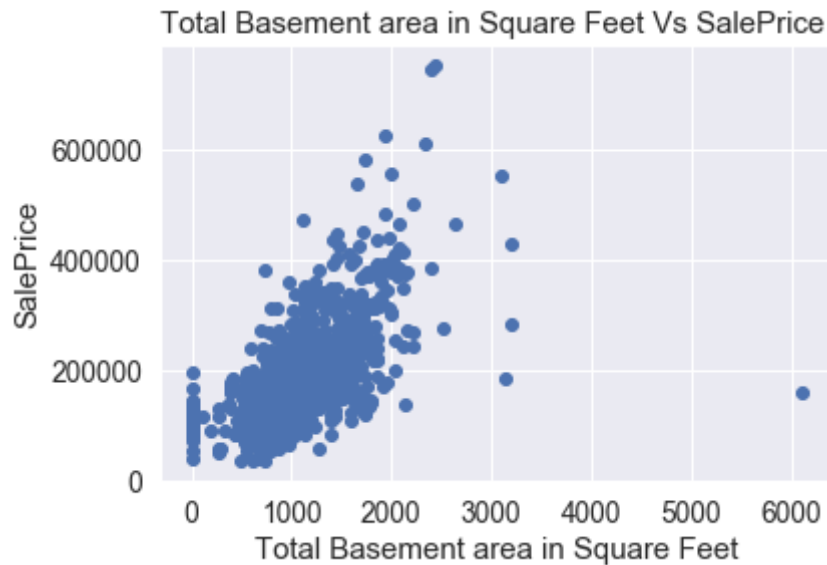


```
In [27]:    1  #sns.factorplot("Fireplaces","SalePrice",data=df_train);
            2
```

In [28]:
```python
#GarageArea has got some outliers lets remove them.
upperlimit = np.percentile(df_train.GarageArea.values, 99.5)
df_train['GarageArea'].ix[df_train['GarageArea']>upperlimit] = upperlim

plt.scatter(df_train.GarageArea, df_train["SalePrice"].values,color='vi
plt.title("Garage Area Vs SalePrice ")
plt.ylabel("SalePrice")
plt.xlabel("Garage Area in sq feet");
```

```
/Users/shradhitsubudhi/anaconda3/envs/python37charm/lib/python3.7/site-pa
ckages/ipykernel_launcher.py:3: DeprecationWarning:
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-d
eprecated (http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-i
ndexer-is-deprecated)
  This is separate from the ipykernel package so we can avoid doing impor
ts until
/Users/shradhitsubudhi/anaconda3/envs/python37charm/lib/python3.7/site-pa
ckages/pandas/core/indexing.py:190: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy (http://pandas.pydata.or
g/pandas-docs/stable/indexing.html#indexing-view-versus-copy)
  self._setitem_with_indexer(indexer, value)
```
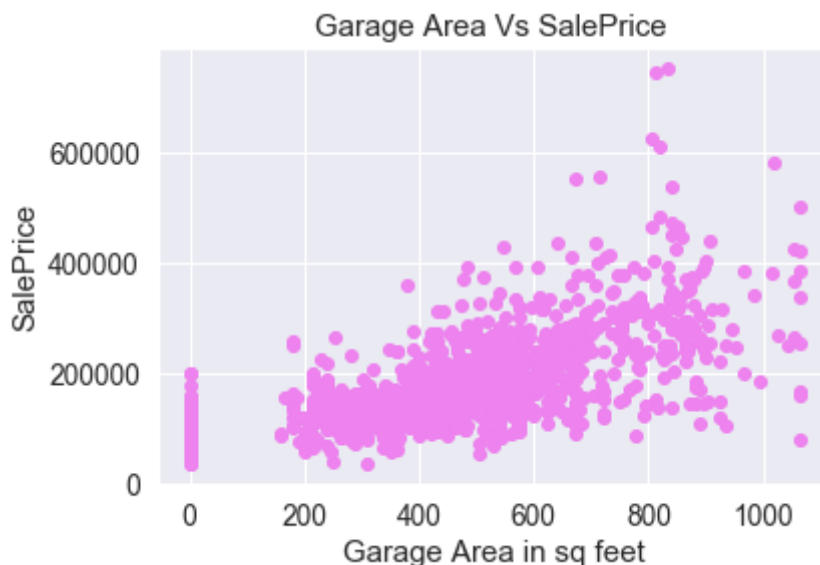

In [29]:
```python
upperlimit = np.percentile(df_train.GarageArea.values, 99.5)
#upperlimit
```

In [30]:
```python
plt.figure(figsize=(10, 8))
plt.xticks(rotation=45)
sns.barplot(df_train["Neighborhood"],df_train["SalePrice"])
plt.title("Sale Price vs Neighborhood");
```


Sale Price vs Neighborhood
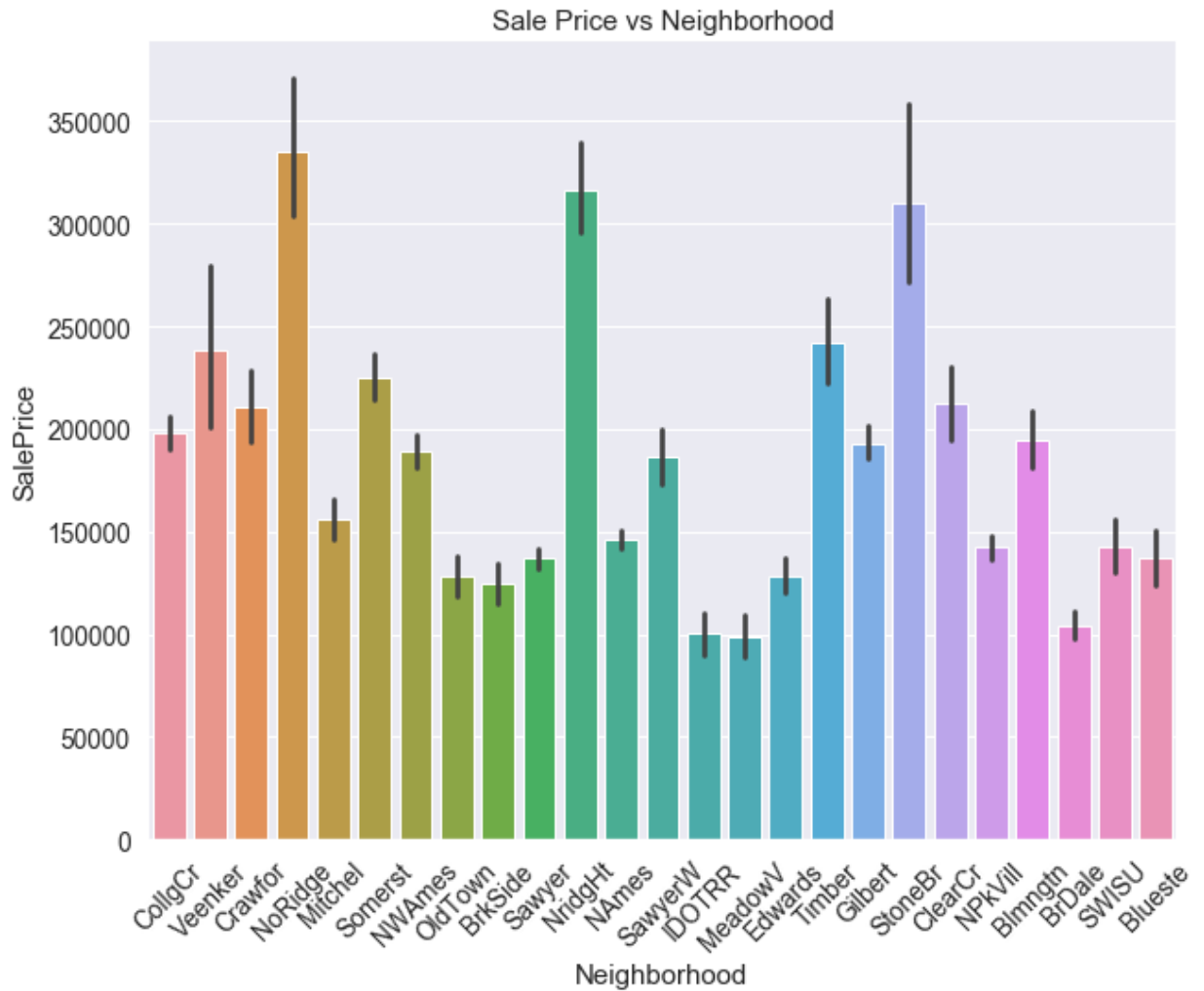
In [42]:
```python
df_train.set_index('Id', inplace=True)
#df.set_index('month')
```

In [43]: `1  df_train`

Out[43]:

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPul
2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPul
3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPul
4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPul
5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPul
6	50	RL	85.0	14115	Pave	NaN	IR1	Lvl	AllPul
7	20	RL	75.0	10084	Pave	NaN	Reg	Lvl	AllPul
8	60	RL	NaN	10382	Pave	NaN	IR1	Lvl	AllPul
9	50	RM	51.0	6120	Pave	NaN	Reg	Lvl	AllPul
10	190	RL	50.0	7420	Pave	NaN	Reg	Lvl	AllPul
11	20	RL	70.0	11200	Pave	NaN	Reg	Lvl	AllPul
12	60	RL	85.0	11924	Pave	NaN	IR1	Lvl	AllPul
13	20	RL	NaN	12968	Pave	NaN	IR2	Lvl	AllPul
14	20	RL	91.0	10652	Pave	NaN	IR1	Lvl	AllPul
15	20	RL	NaN	10920	Pave	NaN	IR1	Lvl	AllPul
16	45	RM	51.0	6120	Pave	NaN	Reg	Lvl	AllPul
17	20	RL	NaN	11241	Pave	NaN	IR1	Lvl	AllPul
18	90	RL	72.0	10791	Pave	NaN	Reg	Lvl	AllPul
19	20	RL	66.0	13695	Pave	NaN	Reg	Lvl	AllPul
20	20	RL	70.0	7560	Pave	NaN	Reg	Lvl	AllPuli
21	60	RL	101.0	14215	Pave	NaN	IR1	Lvl	AllPul
22	45	RM	57.0	7449	Pave	Grvl	Reg	Bnk	AllPul
23	20	RL	75.0	9742	Pave	NaN	Reg	Lvl	AllPul
24	120	RM	44.0	4224	Pave	NaN	Reg	Lvl	AllPul
25	20	RL	NaN	8246	Pave	NaN	IR1	Lvl	AllPul
26	20	RL	110.0	14230	Pave	NaN	Reg	Lvl	AllPul
27	20	RL	60.0	7200	Pave	NaN	Reg	Lvl	AllPul
28	20	RL	98.0	11478	Pave	NaN	Reg	Lvl	AllPul
29	20	RL	47.0	16321	Pave	NaN	IR1	Lvl	AllPul
30	30	RM	60.0	6324	Pave	NaN	IR1	Lvl	AllPul
...	...	...	...	...	...	...	...	...	...
1431	60	RL	60.0	21930	Pave	NaN	IR3	Lvl	AllPul

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
1432	120	RL	NaN	4928	Pave	NaN	IR1	Lvl	AllPul
1433	30	RL	60.0	10800	Pave	Grvl	Reg	Lvl	AllPul
1434	60	RL	93.0	10261	Pave	NaN	IR1	Lvl	AllPul
1435	20	RL	80.0	17400	Pave	NaN	Reg	Low	AllPul
1436	20	RL	80.0	8400	Pave	NaN	Reg	Lvl	AllPul
1437	20	RL	60.0	9000	Pave	NaN	Reg	Lvl	AllPul
1438	20	RL	96.0	12444	Pave	NaN	Reg	Lvl	AllPul
1439	20	RM	90.0	7407	Pave	NaN	Reg	Lvl	AllPul
1440	60	RL	80.0	11584	Pave	NaN	Reg	Lvl	AllPul
1441	70	RL	79.0	11526	Pave	NaN	IR1	Bnk	AllPul
1442	120	RM	NaN	4426	Pave	NaN	Reg	Lvl	AllPul
1443	60	FV	85.0	11003	Pave	NaN	Reg	Lvl	AllPul
1444	30	RL	NaN	8854	Pave	NaN	Reg	Lvl	AllPul
1445	20	RL	63.0	8500	Pave	NaN	Reg	Lvl	AllPul
1446	85	RL	70.0	8400	Pave	NaN	Reg	Lvl	AllPul
1447	20	RL	NaN	26142	Pave	NaN	IR1	Lvl	AllPul
1448	60	RL	80.0	10000	Pave	NaN	Reg	Lvl	AllPul
1449	50	RL	70.0	11767	Pave	NaN	Reg	Lvl	AllPul
1450	180	RM	21.0	1533	Pave	NaN	Reg	Lvl	AllPul
1451	90	RL	60.0	9000	Pave	NaN	Reg	Lvl	AllPul
1452	20	RL	78.0	9262	Pave	NaN	Reg	Lvl	AllPul
1453	180	RM	35.0	3675	Pave	NaN	Reg	Lvl	AllPul
1454	20	RL	90.0	17217	Pave	NaN	Reg	Lvl	AllPul
1455	20	FV	62.0	7500	Pave	Pave	Reg	Lvl	AllPul
1456	60	RL	62.0	7917	Pave	NaN	Reg	Lvl	AllPul
1457	20	RL	85.0	13175	Pave	NaN	Reg	Lvl	AllPul
1458	70	RL	66.0	9042	Pave	NaN	Reg	Lvl	AllPul
1459	20	RL	68.0	9717	Pave	NaN	Reg	Lvl	AllPul
1460	20	RL	75.0	9937	Pave	NaN	Reg	Lvl	AllPul

1460 rows × 83 columns

- PoolQC 99.520548

- MiscFeature 96.301370
- Alley 93.767123
- Fence 80.753425
- FireplaceQu 47.260274
- LotFrontage 17.739726
- GarageCond 5.547945
- GarageType 5.547945
- GarageYrBlt 5.547945
- GarageFinish 5.547945
- GarageQual 5.547945
- BsmtExposure 2.602740
- BsmtFinType2 2.602740
- BsmtFinType1 2.534247
- BsmtCond 2.534247
- BsmtQual 2.534247
- MasVnrArea 0.547945
- MasVnrType 0.547945
- Electrical 0.068493

```
In [51]:    1  sum(df_train['FireplaceQu'].isna())
```

Out[51]: 690

```
In [52]:    1  len(df_train['FireplaceQu'].isna())
```

Out[52]: 1460

```
In [53]:    1  df_train['FireplaceQu'].value_counts()
```

Out[53]: Gd   380
         TA   313
         Fa   33
         Ex   24
         Po   20
         Name: FireplaceQu, dtype: int64

```
In [55]:    1  from sklearn.preprocessing import LabelEncoder
            2  le = LabelEncoder()
```

```
In [61]:    1  #le.fit(df_train['FireplaceQu'])
            2  df['Alley'].fillna(value=df['MSZoning'].value_counts().index[0],inplace
            3
```

```
In [58]:   1   #df_train['FireplaceQu']
```

```
Out[58]:   Id
           1       NaN
           2        TA
           3        TA
           4        Gd
           5        TA
           6       NaN
           7        Gd
           8        TA
           9        TA
           10       TA
           11      NaN
           12       Gd
           13      NaN
           14       Gd
           15       Fa
           16      NaN
           17       TA
           18      NaN
           19      NaN
           20      NaN
           21       Gd
           22       Gd
           23       Gd
           24       TA
           25       TA
           26       Gd
           27      NaN
           28       Gd
           29       Gd
           30      NaN
                   ...
           1431     Gd
           1432    NaN
           1433    NaN
           1434     TA
           1435     Gd
           1436     Gd
           1437    NaN
           1438     Gd
           1439    NaN
           1440     TA
           1441     Gd
           1442     TA
           1443     Ex
           1444     Gd
           1445    NaN
           1446    NaN
           1447    NaN
           1448     TA
           1449    NaN
           1450    NaN
           1451    NaN
           1452     Gd
           1453    NaN
```

```
1454    NaN
1455    NaN
1456     TA
1457     TA
1458     Gd
1459    NaN
1460    NaN
Name: FireplaceQu, Length: 1460, dtype: object
```

In [63]:   1 len(df_train.columns)

Out[63]: 83

```
In [68]:   1  for x in df_train.dtypes:
           2      print(x)
```

```
int64
object
float64
int64
object
object
object
object
object
object
object
object
object
object
object
object
int64
int64
int64
int64
object
object
object
object
object
float64
object
object
object
object
object
object
int64
object
int64
int64
int64
object
object
object
object
int64
int64
int64
int64
int64
int64
int64
int64
int64
int64
object
int64
```

```
object
int64
object
object
float64
object
int64
float64
object
object
object
int64
int64
int64
int64
int64
int64
object
object
object
int64
int64
int64
object
object
int64
float64
float64
float64
```

In [ ]:   1