House_price_data_vis_ana

December 11, 2017

In [1]:
```
import pandas as pd
test_df = pd.read_csv('train.csv')
test_df.head()
```

Out[1]:

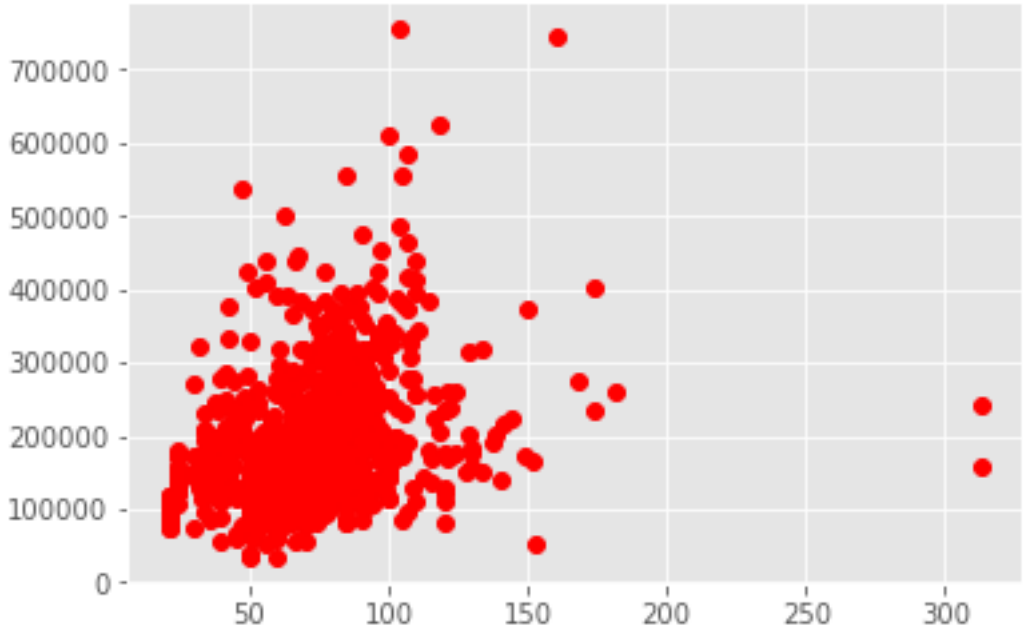
	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	

	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	\
0	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
2	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
3	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
4	Lvl	AllPub	...	0	NaN	NaN	NaN	0	

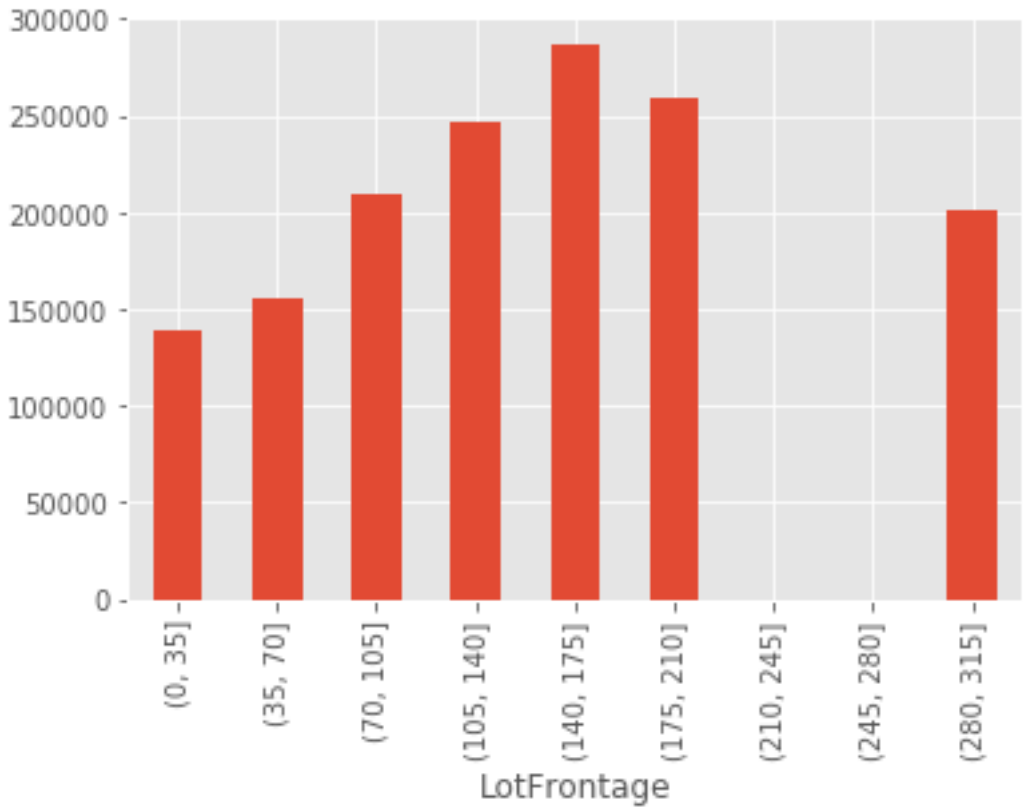
	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	2	2008	WD	Normal	208500
1	5	2007	WD	Normal	181500
2	9	2008	WD	Normal	223500
3	2	2006	WD	Abnorml	140000
4	12	2008	WD	Normal	250000

[5 rows x 81 columns]

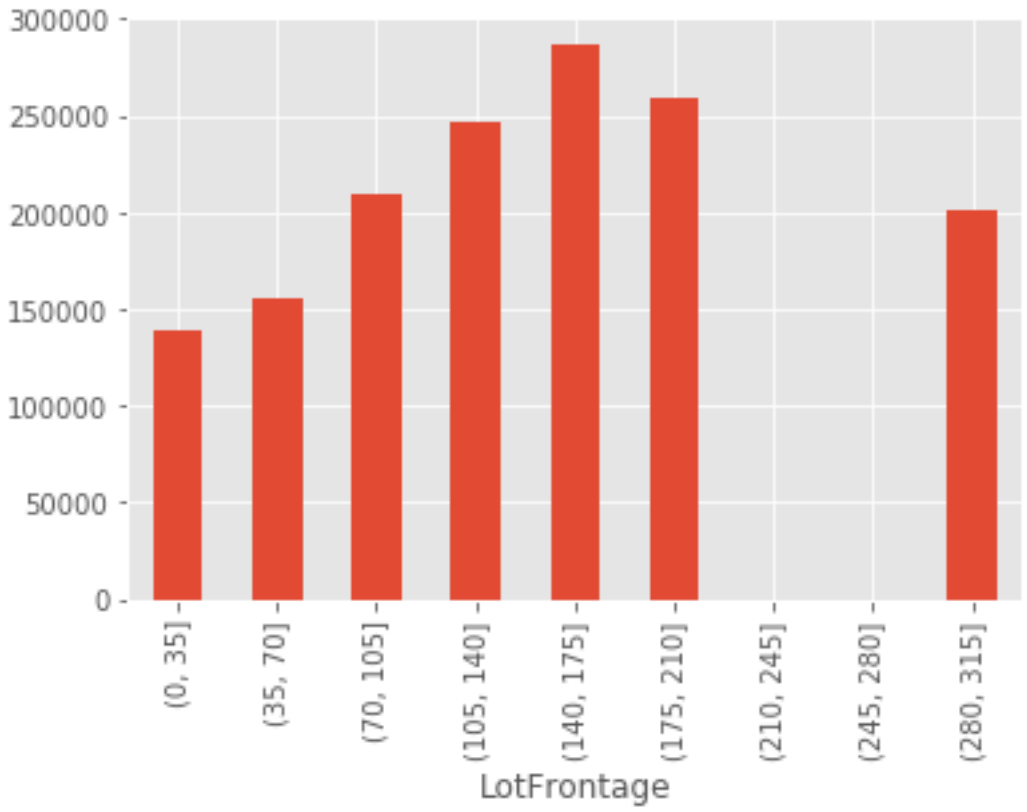
In [2]:
```
import pylab as plt
plt.style.use(style='ggplot')
plt.plot(test_df.LotFrontage,test_df.SalePrice,'ro')
plt.show()
```



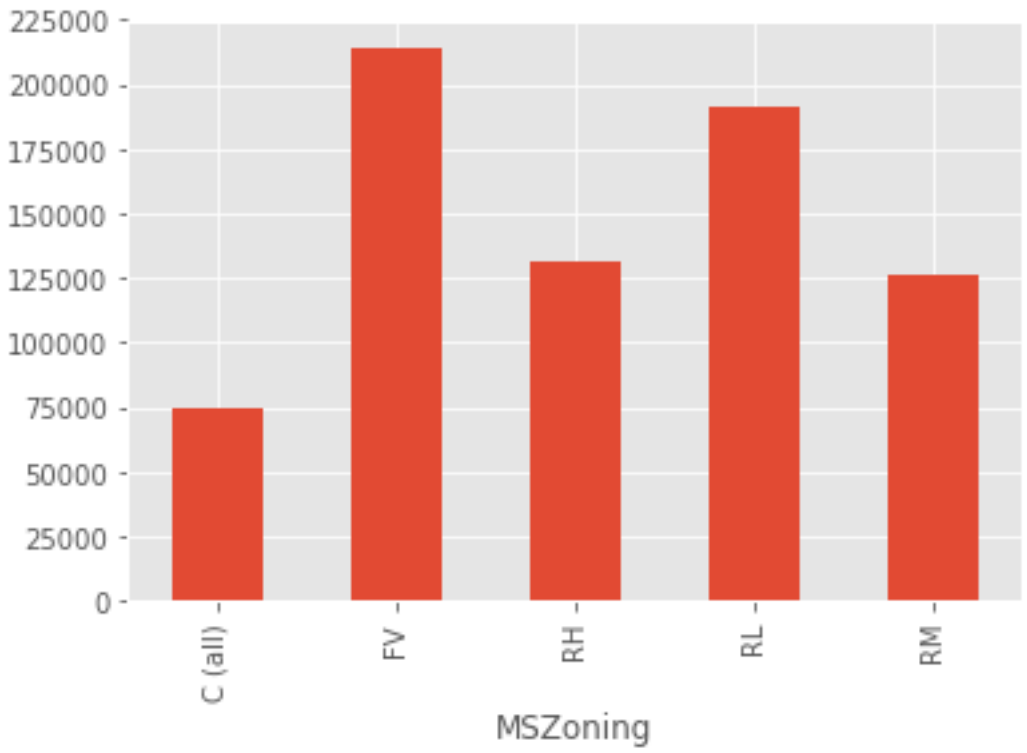
In [5]:
```
import numpy as np
import pylab as plt
group_by_price = pd.cut(test_df.LotFrontage, np.arange(0, 350, 35))
price_grouping = test_df.groupby(group_by_price).mean()
price_grouping['SalePrice'].plot.bar()
plt.show()
```



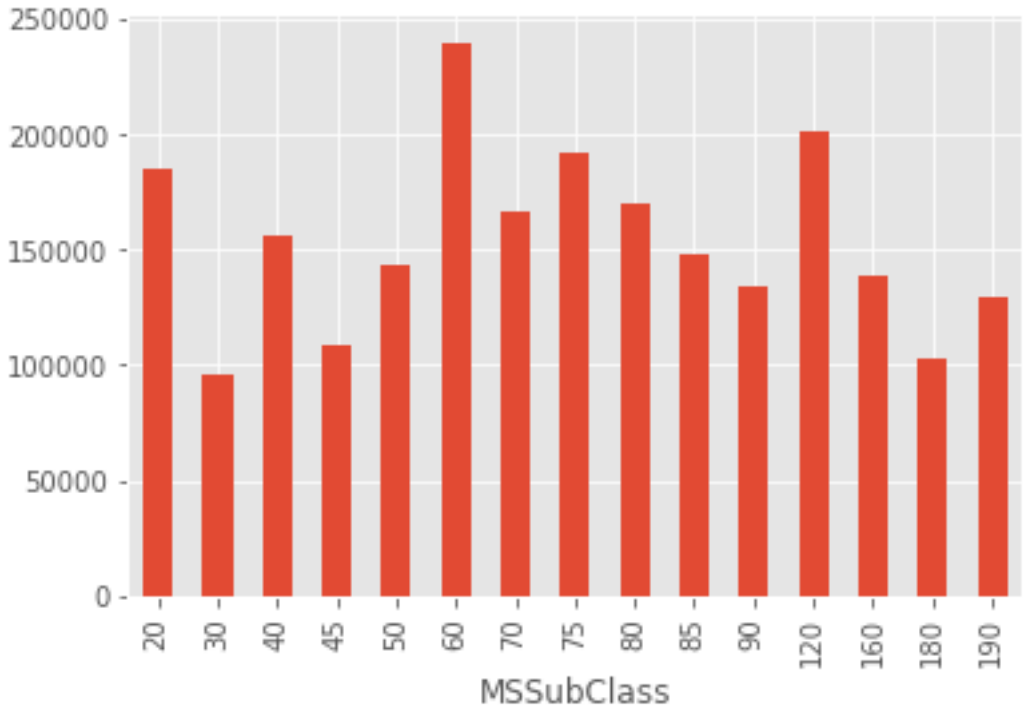

```
In [6]: import numpy as np
        import pylab as plt
        group_by_price = pd.cut(test_df.LotFrontage, np.arange(0, 350, 35))
        price_grouping = test_df.groupby(group_by_price).mean()
        price_grouping['SalePrice'].plot.bar()
        plt.show()
```




```
In [7]: zone_grouping = test_df.groupby('MSZoning').mean()
        zone_grouping['SalePrice'].plot.bar()
        plt.show()
```




```
In [8]: class_grouping = test_df.groupby('MSSubClass').mean()
        class_grouping['SalePrice'].plot.bar()
        plt.show()
```



In [9]:
```
class_grouping = test_df.groupby(['MSSubClass','MSZoning']).mean()
class_grouping
```

Out[9]:

MSSubClass	MSZoning	Id	LotFrontage	LotArea	OverallQual	\
20	C (all)	865.000000	58.000000	8856.000000	3.500000	
	FV	644.769231	73.583333	8544.846154	7.307692	
	RH	833.666667	60.000000	8366.666667	4.333333	
	RL	726.568898	78.065693	11945.566929	6.049213	
	RM	635.300000	64.800000	7871.000000	4.600000	
30	C (all)	779.000000	90.000000	12939.500000	3.500000	
	RH	1327.000000	70.000000	4270.000000	3.000000	
	RL	801.696970	59.535714	8765.575758	4.363636	
	RM	674.515152	58.906250	6814.939394	4.757576	
40	RL	1072.500000	60.000000	17869.500000	6.000000	
	RM	377.500000	45.000000	7250.000000	5.500000	
45	RH	384.000000	60.000000	9000.000000	6.000000	
	RL	591.750000	57.250000	8230.250000	5.000000	
	RM	330.428571	53.000000	5923.857143	5.714286	
50	C (all)	659.750000	72.750000	8930.500000	3.750000	
	RH	682.000000	55.000000	4500.000000	5.000000	
	RL	784.727273	67.564103	12397.693182	5.568182	
	RM	701.862745	57.244444	7407.411765	5.254902	
60	FV	636.200000	76.913043	9268.360000	7.440000	

	RL	716.575092	81.621359	12345.340659	7.025641
	RM	426.000000	60.000000	3378.000000	7.000000
70	C (all)	31.000000	50.000000	8500.000000	4.000000
	RH	916.000000	54.500000	9103.000000	5.666667
	RL	787.266667	72.037037	11849.133333	6.333333
	RM	765.038462	60.280000	8304.884615	6.000000
75	RL	594.000000	89.800000	12483.666667	6.166667
	RM	611.400000	71.400000	11120.600000	6.900000
80	RL	792.155172	80.026316	10895.482759	5.965517
85	RL	828.200000	71.571429	9317.450000	5.450000
90	RH	883.000000	74.666667	7268.666667	5.333333
	RL	815.767442	70.921053	10122.767442	4.860465
	RM	578.666667	74.333333	8059.666667	5.500000
120	FV	994.400000	45.600000	4244.200000	7.600000
	RH	904.500000	34.000000	4059.000000	6.500000
	RL	708.661017	47.062500	5810.830508	7.305085
	RM	749.761905	42.083333	4582.190476	6.285714
160	FV	759.000000	30.058824	3068.772727	6.772727
	RL	651.090909	34.700000	3630.818182	6.454545
	RM	770.700000	24.250000	2086.866667	5.566667
180	RM	798.000000	25.200000	2247.400000	4.600000
190	C (all)	94.000000	60.000000	7200.000000	6.000000
	RH	833.500000	60.000000	8989.000000	5.500000
	RL	632.375000	70.857143	22805.687500	4.687500
	RM	720.363636	62.545455	8478.909091	5.181818

		OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	\
MSSubClass	MSZoning					
20	C (all)	4.000000	1950.500000	1951.000000	0.000000	
	FV	5.000000	2007.000000	2007.230769	64.166667	
	RH	4.333333	1960.333333	1960.333333	0.000000	
	RL	5.566929	1977.811024	1985.694882	108.409901	
	RM	6.100000	1958.600000	1982.100000	0.000000	
30	C (all)	4.500000	1927.500000	1950.000000	0.000000	
	RH	6.000000	1931.000000	2006.000000	0.000000	
	RL	5.727273	1925.484848	1966.484848	3.272727	
	RM	5.848485	1927.606061	1971.272727	22.090909	
40	RL	6.500000	1964.000000	1964.500000	0.000000	
	RM	6.500000	1928.000000	1982.000000	0.000000	
45	RH	3.000000	1928.000000	1950.000000	0.000000	
	RL	6.000000	1946.250000	1965.750000	0.000000	
	RM	6.714286	1928.857143	1965.000000	0.000000	
50	C (all)	4.750000	1913.750000	1958.000000	0.000000	
	RH	5.000000	1932.000000	2000.000000	0.000000	
	RL	5.886364	1940.261364	1971.022727	45.681818	
	RM	6.352941	1930.000000	1968.176471	18.333333	
60	FV	5.080000	2005.200000	2005.560000	118.250000	
	RL	5.307692	1993.692308	1995.501832	187.387454	

	RM	8.000000	1946.000000	1992.000000	0.000000
70	C (all)	4.000000	1920.000000	1950.000000	0.000000
	RH	6.666667	1920.333333	1950.000000	0.000000
	RL	6.766667	1928.333333	1978.500000	7.466667
	RM	6.576923	1907.653846	1981.423077	0.000000
75	RL	5.166667	1924.333333	1975.166667	0.000000
	RM	6.900000	1905.600000	1977.700000	0.000000
80	RL	5.896552	1974.275862	1982.155172	128.603448
85	RL	6.100000	1975.450000	1983.050000	52.550000
90	RH	5.666667	1957.333333	1959.000000	247.666667
	RL	4.860465	1967.162791	1970.767442	91.930233
	RM	5.000000	1972.166667	1972.333333	179.333333
120	FV	5.000000	2004.000000	2004.600000	15.000000
	RH	5.000000	1998.000000	1998.500000	91.000000
	RL	5.033898	1998.152542	1998.525424	93.542373
	RM	5.095238	2000.333333	2000.476190	190.238095
160	FV	5.000000	2002.363636	2002.636364	230.090909
	RL	5.545455	1985.636364	1985.818182	28.909091
	RM	5.266667	1976.800000	1979.300000	175.600000
180	RM	5.800000	1981.200000	1984.500000	24.200000
190	C (all)	6.000000	1910.000000	1998.000000	0.000000
	RH	7.500000	1915.000000	1995.000000	0.000000
	RL	5.312500	1938.687500	1970.062500	29.375000
	RM	6.181818	1917.090909	1967.272727	0.000000

		BsmtFinSF1	BsmtFinSF2	...	WoodDeckSF	\
MSSubClass	MSZoning			...		
20	C (all)	25.000000	0.000000	...	0.000000	
	FV	390.692308	0.000000	...	56.000000	
	RH	565.666667	0.000000	...	223.666667	
	RL	577.446850	73.125984	...	99.848425	
	RM	159.200000	142.200000	...	41.400000	
30	C (all)	247.500000	0.000000	...	0.000000	
	RH	544.000000	0.000000	...	0.000000	
	RL	193.575758	18.363636	...	53.636364	
	RM	167.666667	17.878788	...	17.303030	
40	RL	816.500000	0.000000	...	134.000000	
	RM	185.000000	0.000000	...	0.000000	
45	RH	0.000000	0.000000	...	0.000000	
	RL	55.000000	0.000000	...	0.000000	
	RM	145.142857	0.000000	...	6.857143	
50	C (all)	159.250000	0.000000	...	24.500000	
	RH	182.000000	0.000000	...	0.000000	
	RL	270.125000	40.034091	...	76.204545	
	RM	157.843137	25.078431	...	52.549020	
60	FV	274.600000	0.000000	...	90.720000	
	RL	457.333333	30.534799	...	135.868132	
	RM	0.000000	0.000000	...	0.000000	

70	C (all)	0.000000	0.000000	...	0.000000
	RH	235.666667	0.000000	...	49.000000
	RL	273.400000	12.000000	...	74.866667
	RM	77.346154	0.000000	...	35.423077
75	RL	383.500000	30.666667	...	66.333333
	RM	70.900000	0.000000	...	44.700000
80	RL	486.068966	85.810345	...	127.258621
85	RL	720.900000	53.400000	...	95.900000
90	RH	256.000000	0.000000	...	0.000000
	RL	387.930233	20.255814	...	44.720930
	RM	613.500000	118.666667	...	25.833333
120	FV	679.800000	0.000000	...	50.600000
	RH	425.000000	69.500000	...	132.000000
	RL	527.084746	55.220339	...	127.372881
	RM	831.619048	0.000000	...	134.857143
160	FV	271.363636	0.000000	...	21.545455
	RL	405.818182	81.090909	...	100.818182
	RM	216.466667	43.000000	...	97.200000
180	RM	436.800000	97.900000	...	30.400000
190	C (all)	1046.000000	0.000000	...	0.000000
	RH	128.000000	0.000000	...	0.000000
	RL	707.562500	20.500000	...	122.500000
	RM	90.636364	0.000000	...	36.090909

		OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	\
MSSubClass	MSZoning					
20	C (all)	0.000000	0.000000	0.000000	0.000000	
	FV	130.769231	0.000000	0.000000	0.000000	
	RH	14.000000	8.000000	0.000000	0.000000	
	RL	40.791339	12.446850	5.535433	16.494094	
	RM	48.300000	27.000000	0.000000	0.000000	
30	C (all)	271.500000	57.500000	0.000000	0.000000	
	RH	0.000000	286.000000	0.000000	0.000000	
	RL	14.181818	49.090909	0.000000	5.030303	
	RM	26.030303	66.545455	0.000000	3.636364	
40	RL	14.000000	0.000000	0.000000	0.000000	
	RM	0.000000	91.500000	0.000000	0.000000	
45	RH	0.000000	91.000000	0.000000	0.000000	
	RL	0.000000	28.000000	0.000000	0.000000	
	RM	16.000000	112.285714	0.000000	0.000000	
50	C (all)	11.750000	134.500000	0.000000	0.000000	
	RH	0.000000	56.000000	0.000000	0.000000	
	RL	21.670455	42.215909	6.863636	23.761364	
	RM	36.745098	57.588235	0.000000	20.196078	
60	FV	106.320000	0.000000	0.000000	7.920000	
	RL	78.417582	9.128205	3.717949	17.820513	
	RM	0.000000	126.000000	0.000000	0.000000	
70	C (all)	54.000000	172.000000	0.000000	0.000000	

	RH	64.000000	84.666667	0.000000	0.000000
	RL	24.066667	52.566667	0.000000	25.700000
	RM	46.769231	100.269231	0.000000	6.461538
75	RL	58.500000	136.666667	0.000000	21.000000
	RM	174.800000	89.700000	0.000000	41.000000
80	RL	30.827586	17.396552	3.896552	16.965517
85	RL	22.100000	30.450000	0.000000	9.600000
90	RH	0.000000	0.000000	0.000000	0.000000
	RL	15.976744	3.372093	0.000000	0.000000
	RM	0.000000	0.000000	0.000000	0.000000
120	FV	132.800000	0.000000	0.000000	0.000000
	RH	54.000000	0.000000	0.000000	0.000000
	RL	43.711864	11.237288	2.847458	30.694915
	RM	58.619048	0.000000	7.285714	14.666667
160	FV	77.454545	0.000000	0.000000	0.000000
	RL	13.090909	0.000000	0.000000	0.000000
	RM	15.133333	0.000000	0.000000	0.000000
180	RM	22.500000	0.000000	0.000000	0.000000
190	C (all)	0.000000	0.000000	0.000000	99.000000
	RH	55.000000	114.000000	0.000000	0.000000
	RL	44.500000	22.375000	0.000000	0.000000
	RM	12.909091	76.727273	0.000000	24.636364

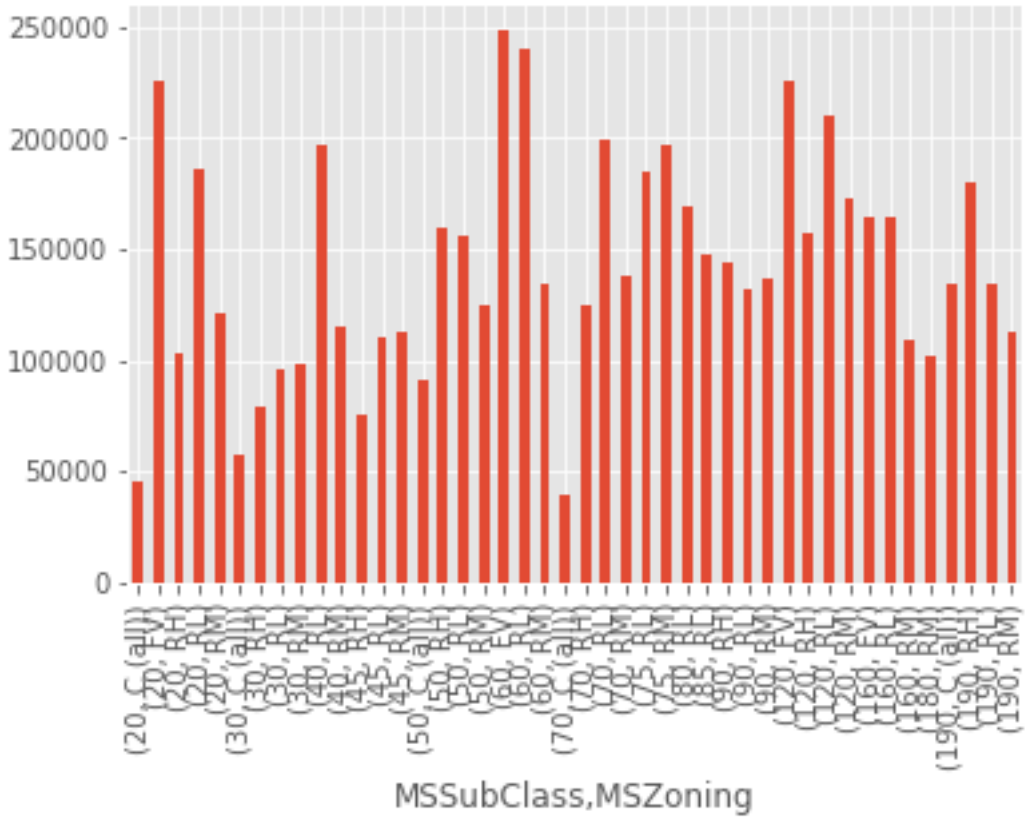
		PoolArea	MiscVal	MoSold	YrSold	\
MSSubClass	MSZoning					
20	C (all)	0.000000	27.000000	8.000000	2008.000000	
	FV	0.000000	0.000000	7.307692	2007.923077	
	RH	0.000000	0.000000	9.000000	2007.000000	
	RL	1.275591	56.358268	6.204724	2007.870079	
	RM	0.000000	0.000000	5.700000	2008.300000	
30	C (all)	0.000000	280.000000	9.500000	2008.500000	
	RH	0.000000	0.000000	5.000000	2007.000000	
	RL	0.000000	25.757576	6.000000	2007.878788	
	RM	0.000000	32.424242	5.606061	2007.636364	
40	RL	0.000000	0.000000	6.500000	2009.500000	
	RM	0.000000	300.000000	9.000000	2007.500000	
45	RH	0.000000	0.000000	10.000000	2009.000000	
	RL	0.000000	0.000000	5.250000	2007.250000	
	RM	0.000000	0.000000	6.428571	2007.285714	
50	C (all)	0.000000	0.000000	6.000000	2008.750000	
	RH	0.000000	0.000000	7.000000	2009.000000	
	RL	0.000000	65.909091	6.113636	2007.806818	
	RM	0.000000	53.921569	6.450980	2007.803922	
60	FV	0.000000	0.000000	6.720000	2008.040000	
	RL	5.692308	20.073260	6.446886	2007.695971	
	RM	0.000000	0.000000	9.000000	2009.000000	
70	C (all)	0.000000	0.000000	7.000000	2008.000000	
	RH	0.000000	0.000000	7.000000	2008.333333	

	RL	0.000000	83.333333	6.866667	2007.800000
	RM	0.000000	0.000000	7.038462	2007.769231
75	RL	85.333333	0.000000	6.000000	2007.500000
	RM	0.000000	0.000000	6.600000	2007.500000
80	RL	22.655172	20.689655	6.810345	2007.810345
85	RL	0.000000	20.000000	5.550000	2008.150000
90	RH	0.000000	0.000000	8.666667	2007.000000
	RL	0.000000	218.604651	6.604651	2007.790698
	RM	0.000000	0.000000	7.666667	2008.333333
120	FV	0.000000	0.000000	5.400000	2007.200000
	RH	0.000000	0.000000	7.000000	2007.500000
	RL	0.000000	0.000000	6.847458	2007.915254
	RM	0.000000	0.000000	5.809524	2007.714286
160	FV	0.000000	0.000000	5.363636	2007.727273
	RL	0.000000	0.000000	6.636364	2008.454545
	RM	0.000000	0.000000	5.466667	2007.633333
180	RM	0.000000	0.000000	6.800000	2007.800000
190	C (all)	0.000000	0.000000	11.000000	2007.000000
	RH	0.000000	0.000000	5.000000	2006.500000
	RL	0.000000	43.750000	5.000000	2007.625000
	RM	0.000000	318.181818	5.636364	2008.000000

		SalePrice
MSSubClass	MSZoning	
20	C (all)	45652.000000
	FV	226289.538462
	RH	102966.666667
	RL	186467.039370
	RM	121327.500000
30	C (all)	57950.000000
	RH	79000.000000
	RL	96481.212121
	RM	97983.969697
40	RL	196500.000000
	RM	115750.000000
45	RH	76000.000000
	RL	110050.000000
	RM	112414.285714
50	C (all)	91044.000000
	RH	159434.000000
	RL	156277.477273
	RM	124698.039216
60	FV	248558.600000
	RL	239544.457875
	RM	135000.000000
70	C (all)	40000.000000
	RH	124533.333333
	RL	199808.733333

```
                RM         138403.192308
        75      RL         184750.000000
                RM         197050.000000
        80      RL         169736.551724
        85      RL         147810.000000
        90      RH         144666.666667
                RL         132379.906977
                RM         136300.000000
        120     FV         226140.000000
                RH         157000.000000
                RL         210029.491525
                RM         172920.952381
        160     FV         164749.318182
                RL         164909.090909
                RM         109876.666667
        180     RM         102300.000000
        190     C (all)    133900.000000
                RH         180000.000000
                RL         134662.500000
                RM         112718.181818

        [44 rows x 37 columns]
```

```
In [10]: class_grouping = test_df.groupby(['MSSubClass','MSZoning']).mean()
         class_grouping['SalePrice'].plot.bar()
         plt.show()
```



In [11]: test_df.head()

Out[11]:

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

	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	\
0	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
2	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
3	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
4	Lvl	AllPub	...	0	NaN	NaN	NaN	0	

	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	2	2008	WD	Normal	208500
1	5	2007	WD	Normal	181500
2	9	2008	WD	Normal	223500

```
3      2   2006        WD         Abnorml     140000
4     12   2008        WD          Normal     250000

[5 rows x 81 columns]
```

```
In [13]: test_df.groupby(['PoolArea','PoolQC']).head(80)
```

Out[13]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	
5	6	50	RL	85.0	14115	Pave	NaN	IR1	
6	7	20	RL	75.0	10084	Pave	NaN	Reg	
7	8	60	RL	NaN	10382	Pave	NaN	IR1	
8	9	50	RM	51.0	6120	Pave	NaN	Reg	
9	10	190	RL	50.0	7420	Pave	NaN	Reg	
10	11	20	RL	70.0	11200	Pave	NaN	Reg	
11	12	60	RL	85.0	11924	Pave	NaN	IR1	
12	13	20	RL	NaN	12968	Pave	NaN	IR2	
13	14	20	RL	91.0	10652	Pave	NaN	IR1	
14	15	20	RL	NaN	10920	Pave	NaN	IR1	
15	16	45	RM	51.0	6120	Pave	NaN	Reg	
16	17	20	RL	NaN	11241	Pave	NaN	IR1	
17	18	90	RL	72.0	10791	Pave	NaN	Reg	
18	19	20	RL	66.0	13695	Pave	NaN	Reg	
19	20	20	RL	70.0	7560	Pave	NaN	Reg	
20	21	60	RL	101.0	14215	Pave	NaN	IR1	
21	22	45	RM	57.0	7449	Pave	Grvl	Reg	
22	23	20	RL	75.0	9742	Pave	NaN	Reg	
23	24	120	RM	44.0	4224	Pave	NaN	Reg	
24	25	20	RL	NaN	8246	Pave	NaN	IR1	
25	26	20	RL	110.0	14230	Pave	NaN	Reg	
26	27	20	RL	60.0	7200	Pave	NaN	Reg	
27	28	20	RL	98.0	11478	Pave	NaN	Reg	
28	29	20	RL	47.0	16321	Pave	NaN	IR1	
29	30	30	RM	60.0	6324	Pave	NaN	IR1	
...	...	...	...	...	...	...	...	...	
57	58	60	RL	89.0	11645	Pave	NaN	IR1	
58	59	60	RL	66.0	13682	Pave	NaN	IR2	
59	60	20	RL	60.0	7200	Pave	NaN	Reg	
60	61	20	RL	63.0	13072	Pave	NaN	Reg	
61	62	75	RM	60.0	7200	Pave	NaN	Reg	
62	63	120	RL	44.0	6442	Pave	NaN	IR1	
63	64	70	RM	50.0	10300	Pave	NaN	IR1	
64	65	60	RL	NaN	9375	Pave	NaN	Reg	
65	66	60	RL	76.0	9591	Pave	NaN	Reg	

66	67	20	RL	NaN	19900	Pave	NaN	Reg
67	68	20	RL	72.0	10665	Pave	NaN	IR1
68	69	30	RM	47.0	4608	Pave	NaN	Reg
69	70	50	RL	81.0	15593	Pave	NaN	Reg
70	71	20	RL	95.0	13651	Pave	NaN	IR1
71	72	20	RL	69.0	7599	Pave	NaN	Reg
72	73	60	RL	74.0	10141	Pave	NaN	IR1
73	74	20	RL	85.0	10200	Pave	NaN	Reg
74	75	50	RM	60.0	5790	Pave	NaN	Reg
75	76	180	RM	21.0	1596	Pave	NaN	Reg
76	77	20	RL	NaN	8475	Pave	NaN	IR1
77	78	50	RM	50.0	8635	Pave	NaN	Reg
78	79	90	RL	72.0	10778	Pave	NaN	Reg
79	80	50	RM	60.0	10440	Pave	Grvl	Reg
197	198	75	RL	174.0	25419	Pave	NaN	Reg
810	811	20	RL	78.0	10140	Pave	NaN	Reg
1170	1171	80	RL	76.0	9880	Pave	NaN	Reg
1182	1183	60	RL	160.0	15623	Pave	NaN	IR1
1298	1299	60	RL	313.0	63887	Pave	NaN	IR3
1386	1387	60	RL	80.0	16692	Pave	NaN	IR1
1423	1424	80	RL	NaN	19690	Pave	NaN	IR1

	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	\
0	Lvl	AllPub	...	0	NaN	NaN	NaN	
1	Lvl	AllPub	...	0	NaN	NaN	NaN	
2	Lvl	AllPub	...	0	NaN	NaN	NaN	
3	Lvl	AllPub	...	0	NaN	NaN	NaN	
4	Lvl	AllPub	...	0	NaN	NaN	NaN	
5	Lvl	AllPub	...	0	NaN	MnPrv	Shed	
6	Lvl	AllPub	...	0	NaN	NaN	NaN	
7	Lvl	AllPub	...	0	NaN	NaN	Shed	
8	Lvl	AllPub	...	0	NaN	NaN	NaN	
9	Lvl	AllPub	...	0	NaN	NaN	NaN	
10	Lvl	AllPub	...	0	NaN	NaN	NaN	
11	Lvl	AllPub	...	0	NaN	NaN	NaN	
12	Lvl	AllPub	...	0	NaN	NaN	NaN	
13	Lvl	AllPub	...	0	NaN	NaN	NaN	
14	Lvl	AllPub	...	0	NaN	GdWo	NaN	
15	Lvl	AllPub	...	0	NaN	GdPrv	NaN	
16	Lvl	AllPub	...	0	NaN	NaN	Shed	
17	Lvl	AllPub	...	0	NaN	NaN	Shed	
18	Lvl	AllPub	...	0	NaN	NaN	NaN	
19	Lvl	AllPub	...	0	NaN	MnPrv	NaN	
20	Lvl	AllPub	...	0	NaN	NaN	NaN	
21	Bnk	AllPub	...	0	NaN	GdPrv	NaN	
22	Lvl	AllPub	...	0	NaN	NaN	NaN	
23	Lvl	AllPub	...	0	NaN	NaN	NaN	
24	Lvl	AllPub	...	0	NaN	MnPrv	NaN	

25	Lvl	AllPub	...	0	NaN	NaN	NaN
26	Lvl	AllPub	...	0	NaN	NaN	NaN
27	Lvl	AllPub	...	0	NaN	NaN	NaN
28	Lvl	AllPub	...	0	NaN	NaN	NaN
29	Lvl	AllPub	...	0	NaN	NaN	NaN
...	...	...	...	...	...	...	...
57	Lvl	AllPub	...	0	NaN	NaN	NaN
58	HLS	AllPub	...	0	NaN	NaN	NaN
59	Bnk	AllPub	...	0	NaN	MnPrv	NaN
60	Lvl	AllPub	...	0	NaN	NaN	NaN
61	Lvl	AllPub	...	0	NaN	NaN	NaN
62	Lvl	AllPub	...	0	NaN	NaN	NaN
63	Bnk	AllPub	...	0	NaN	GdPrv	NaN
64	Lvl	AllPub	...	0	NaN	GdPrv	NaN
65	Lvl	AllPub	...	0	NaN	NaN	NaN
66	Lvl	AllPub	...	0	NaN	NaN	NaN
67	Lvl	AllPub	...	0	NaN	NaN	NaN
68	Lvl	AllPub	...	0	NaN	NaN	NaN
69	Lvl	AllPub	...	0	NaN	NaN	NaN
70	Lvl	AllPub	...	0	NaN	NaN	NaN
71	Lvl	AllPub	...	0	NaN	NaN	NaN
72	Lvl	AllPub	...	0	NaN	NaN	NaN
73	Lvl	AllPub	...	0	NaN	GdWo	NaN
74	Lvl	AllPub	...	0	NaN	NaN	NaN
75	Lvl	AllPub	...	0	NaN	GdWo	NaN
76	Lvl	AllPub	...	0	NaN	NaN	NaN
77	Lvl	AllPub	...	0	NaN	MnPrv	NaN
78	Lvl	AllPub	...	0	NaN	NaN	NaN
79	Lvl	AllPub	...	0	NaN	MnPrv	NaN
197	Lvl	AllPub	...	512	Ex	GdPrv	NaN
810	Lvl	AllPub	...	648	Fa	GdPrv	NaN
1170	Lvl	AllPub	...	576	Gd	GdPrv	NaN
1182	Lvl	AllPub	...	555	Ex	MnPrv	NaN
1298	Bnk	AllPub	...	480	Gd	NaN	NaN
1386	Lvl	AllPub	...	519	Fa	MnPrv	TenC
1423	Lvl	AllPub	...	738	Gd	GdPrv	NaN

	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	0	2	2008	WD	Normal	208500
1	0	5	2007	WD	Normal	181500
2	0	9	2008	WD	Normal	223500
3	0	2	2006	WD	Abnorml	140000
4	0	12	2008	WD	Normal	250000
5	700	10	2009	WD	Normal	143000
6	0	8	2007	WD	Normal	307000
7	350	11	2009	WD	Normal	200000
8	0	4	2008	WD	Abnorml	129900
9	0	1	2008	WD	Normal	118000

10	0	2	2008	WD	Normal	129500
11	0	7	2006	New	Partial	345000
12	0	9	2008	WD	Normal	144000
13	0	8	2007	New	Partial	279500
14	0	5	2008	WD	Normal	157000
15	0	7	2007	WD	Normal	132000
16	700	3	2010	WD	Normal	149000
17	500	10	2006	WD	Normal	90000
18	0	6	2008	WD	Normal	159000
19	0	5	2009	COD	Abnorml	139000
20	0	11	2006	New	Partial	325300
21	0	6	2007	WD	Normal	139400
22	0	9	2008	WD	Normal	230000
23	0	6	2007	WD	Normal	129900
24	0	5	2010	WD	Normal	154000
25	0	7	2009	WD	Normal	256300
26	0	5	2010	WD	Normal	134800
27	0	5	2010	WD	Normal	306000
28	0	12	2006	WD	Normal	207500
29	0	5	2008	WD	Normal	68500
...	...	...	...	...	...	...
57	0	8	2006	WD	Normal	196500
58	0	10	2006	New	Partial	438780
59	0	1	2008	WD	Normal	124900
60	0	5	2006	New	Partial	158000
61	0	3	2007	WD	Normal	101000
62	0	10	2007	WD	Normal	202500
63	0	4	2010	WD	Normal	140000
64	0	2	2009	WD	Normal	219500
65	0	10	2007	WD	Normal	317000
66	0	7	2010	WD	Normal	180000
67	0	6	2007	WD	Normal	226000
68	0	6	2010	WD	Normal	80000
69	0	7	2006	WD	Normal	225000
70	0	2	2007	WD	Normal	244000
71	0	6	2007	WD	Normal	129500
72	0	12	2009	WD	Normal	185000
73	0	5	2010	WD	Normal	144900
74	0	5	2010	WD	Normal	107400
75	0	11	2009	WD	Normal	91000
76	0	4	2008	WD	Normal	135750
77	0	1	2008	WD	Normal	127000
78	0	4	2010	WD	Normal	136500
79	0	5	2009	WD	Normal	110000
197	0	3	2006	WD	Abnorml	235000
810	0	1	2006	WD	Normal	181000
1170	0	7	2008	WD	Normal	171000
1182	0	7	2007	WD	Abnorml	745000

```
1298       0      1   2008       New        Partial     160000
1386    2000      7   2006        WD         Normal     250000
1423       0      8   2006        WD         Alloca     274970

[87 rows x 81 columns]
```

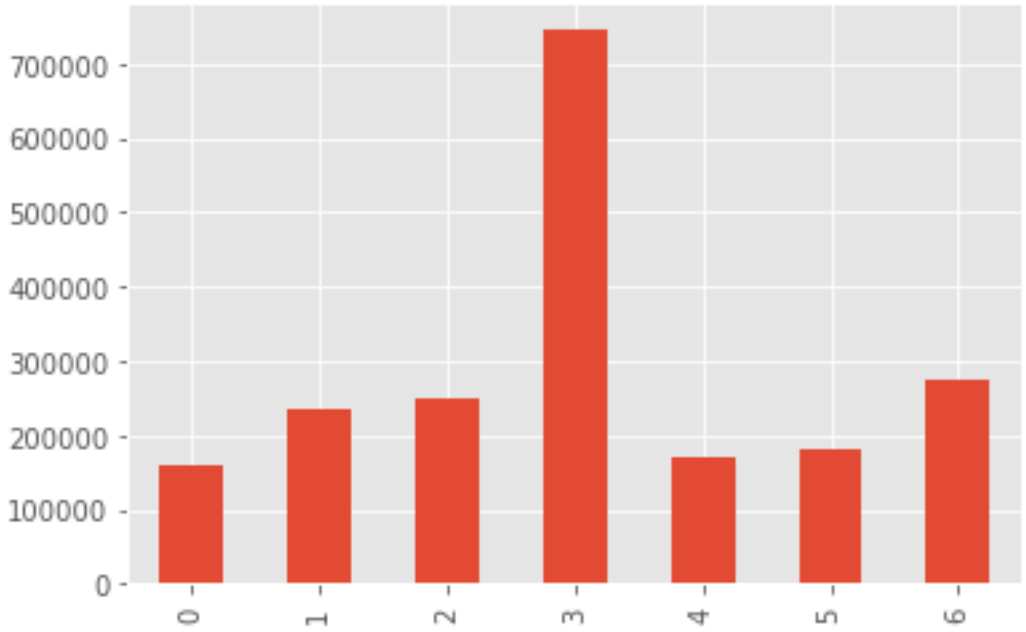
```
In [15]: test_df.groupby(['PoolArea','PoolQC']).mean()
```

Out[15]:

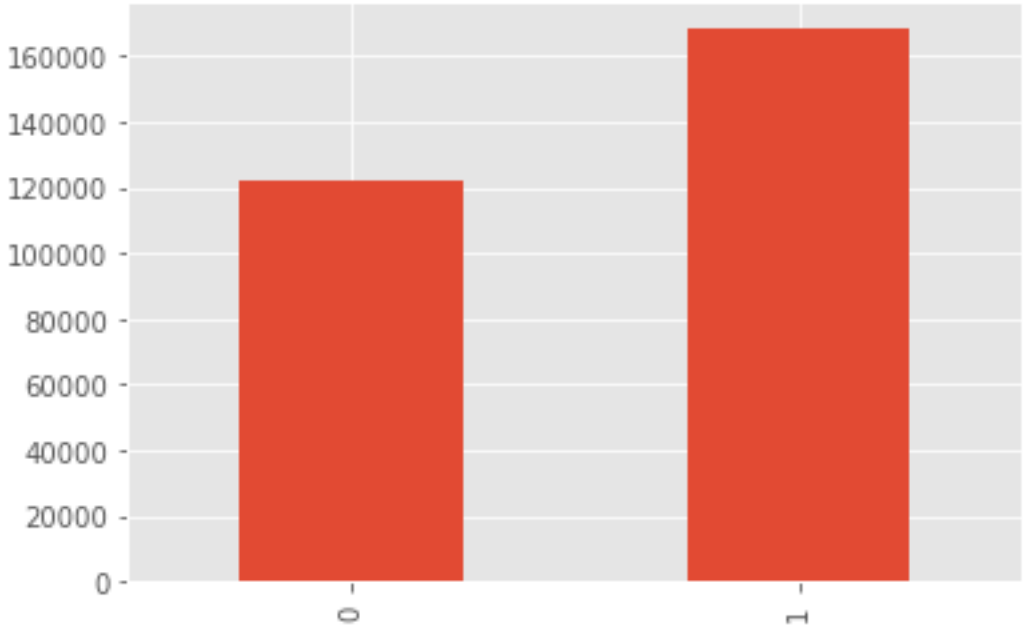
PoolArea	PoolQC	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	...	GarageArea	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	MiscVal	MoSold
480	Gd	1299	60	313.0	63887	10	5	2008	2008	796.0	5644	...	1418	214	292	0	0	0	0	1
512	Ex	198	75	174.0	25419	8	4	1918	1990	0.0	1036	...	795	0	16	552	0	0	0	3
519	Fa	1387	60	80.0	16692	7	5	1978	1978	184.0	790	...	564	0	112	0	0	440	2000	7
555	Ex	1183	60	160.0	15623	10	5	1996	1996	0.0	2096	...	813	171	78	0	0	0	0	7
576	Gd	1171	80	76.0	9880	6	6	1977	1977	0.0	522	...	358	203	0	0	0	0	0	7
648	Fa	811	20	78.0	10140	6	6	1974	1999	99.0	663	...	484	265	0	0	0	0	0	1
738	Gd	1424	80	NaN	19690	6	7	1966	1966	0.0	0	...	432	586	236	0	0	0	0	8

```
                  YrSold  SalePrice
PoolArea PoolQC
480      Gd         2008     160000
512      Ex         2006     235000
519      Fa         2006     250000
555      Ex         2007     745000
576      Gd         2008     171000
648      Fa         2006     181000
738      Gd         2006     274970

[7 rows x 37 columns]
```

In [16]:
```
pool_class = test_df.groupby(['PoolArea','PoolQC'],as_index=False).mean()
pool_class['SalePrice'].plot.bar()
plt.show()
```



In [17]:
```
pool_class = test_df.groupby('Alley',as_index=False).mean()
pool_class['SalePrice'].plot.bar()
plt.show()
```

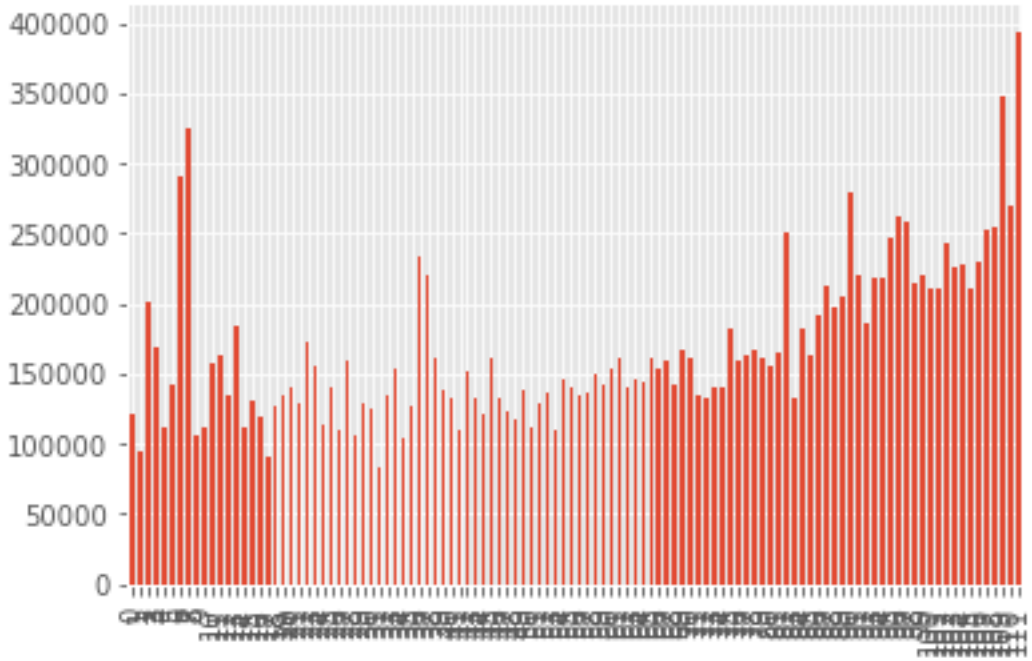



```
In [18]: pool_class = test_df.groupby('Alley',as_index=False).mean()

In [19]: test_df.groupby('Alley',as_index=False).mean()

Out[19]:   Alley          Id  MSSubClass  LotFrontage      LotArea  OverallQual  \
         0  Grvl  735.900000   63.900000    61.744681  8932.500000     5.180000
         1  Pave  726.390244  102.804878    46.666667  5953.121951     6.414634

            OverallCond    YearBuilt  YearRemodAdd  MasVnrArea       ...        \
         0     6.260000  1919.740000   1972.520000      17.500       ...
         1     5.439024  1968.317073   1993.780488     133.175       ...

            WoodDeckSF  OpenPorchSF  EnclosedPorch  3SsnPorch  ScreenPorch  PoolArea  \
         0   50.520000    52.620000      75.060000        0.0     9.600000       0.0
         1   23.902439    75.853659      51.341463        0.0    10.195122       0.0

            MiscVal   MoSold       YrSold      SalePrice
         0     58.0  6.26000  2007.660000  122219.080000
         1      0.0  5.97561  2007.804878  168000.585366

         [2 rows x 39 columns]

In [20]: test_df.groupby('Alley',as_index=False).count()

Out[20]:   Alley  Id  MSSubClass  MSZoning  LotFrontage  LotArea  Street  LotShape  \
         0  Grvl  50          50        50           47       50      50        50
```

```
1  Pave  41          41        41           39       41      41        41

   LandContour  Utilities    ...      PoolArea  PoolQC  Fence  MiscFeature  \
0           50         50    ...            50       0     14            4
1           41         41    ...            41       0      4            0

   MiscVal  MoSold  YrSold  SaleType  SaleCondition  SalePrice
0       50      50      50        50             50         50
1       41      41      41        41             41         41

[2 rows x 81 columns]
```

```
In [ ]: test_df.groupby('Alley',as_index=False).count()
```

```
In [21]: pool_class = test_df.groupby('YearBuilt',as_index=False).mean()
         pool_class['SalePrice'].plot.bar()
         plt.show()
```



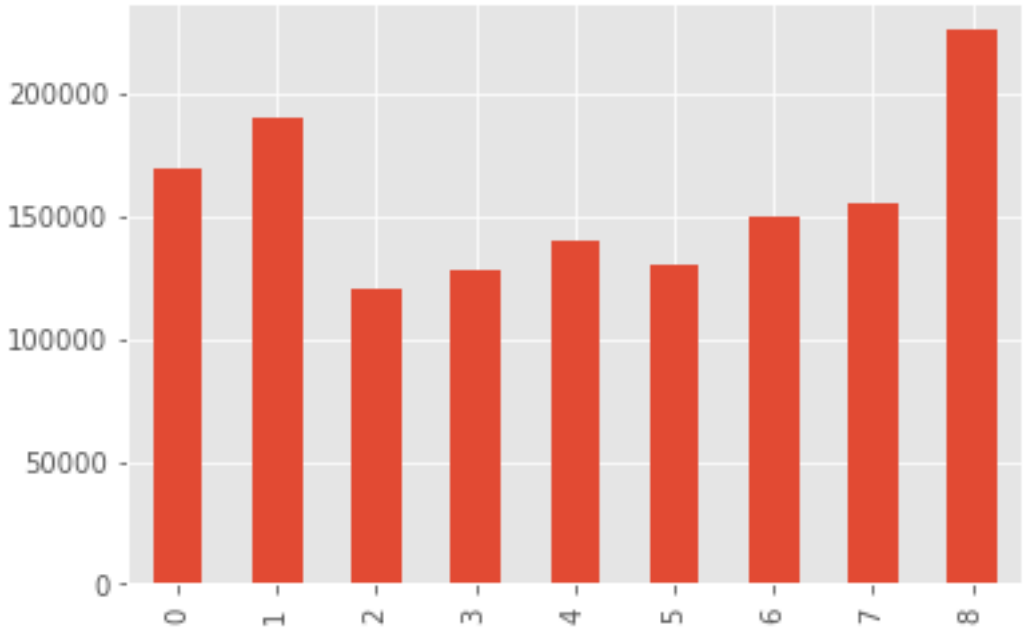

```
In [22]: test_df.YearBuilt.min(axis=0)
```

```
Out[22]: 1872
```

```
In [23]: test_df.YearBuilt.max(axis=0)
```

```
Out[23]: 2010
```

In [24]:
```
year_group = pd.cut(test_df.YearBuilt, np.arange(1870, 2010, 14))
year_class = test_df.groupby(year_group,as_index=False).mean()
year_class['SalePrice'].plot.bar()
plt.show()
```



In [25]:
```
year_group = pd.cut(test_df.YearBuilt, np.arange(1870, 2010, 14))
year_class = test_df.groupby(year_group,as_index=False).mean()
```

In [26]:
```
year_class
```

Out[26]:

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	\
0	899.571429	67.857143	73.857143	11178.428571	6.571429	
1	757.500000	79.375000	74.125000	11613.750000	6.250000	
2	746.111111	98.055556	64.833333	9507.777778	5.027778	
3	747.416667	53.583333	60.926606	8879.950000	5.316667	
4	679.453488	55.174419	61.026316	8541.081395	5.418605	
5	741.195312	42.343750	69.617391	10376.312500	4.945312	
6	730.517241	39.406130	78.958763	13152.057471	5.287356	
7	751.856502	70.493274	62.018750	10000.964126	5.623318	
8	696.726619	59.532374	75.900000	11552.172662	6.769784	

	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	\
0	6.857143	1878.428571	1987.857143	0.000000	67.857143	
1	6.750000	1890.625000	1982.750000	0.000000	136.000000	
2	6.305556	1906.888889	1979.000000	0.000000	136.138889	
3	6.066667	1920.450000	1970.458333	0.900000	186.608333	

```
4     6.337209  1934.674419   1970.279070   34.941860  205.337209
5     5.695312  1949.093750   1966.835938   39.859375  310.671875
6     5.796935  1961.555556   1973.356322   94.588462  499.766284
7     5.793722  1974.596413   1978.878924  109.825112  547.587444
8     5.302158  1991.273381   1992.841727  124.719424  661.582734

       ...        WoodDeckSF  OpenPorchSF  EnclosedPorch  3SsnPorch  \
0      ...         69.000000    38.714286     125.000000   0.000000
1      ...         24.125000   156.625000      62.625000   0.000000
2      ...         98.916667    57.444444      66.972222   0.000000
3      ...         36.833333    36.025000      97.675000   2.366667
4      ...         54.755814    15.453488      38.616279   0.000000
5      ...         43.101562    20.421875      24.812500   2.953125
6      ...         74.310345    36.141762      19.463602   4.417625
7      ...        116.766816    34.278027       7.103139   1.551570
8      ...        157.820144    56.136691      20.251799  11.597122

   ScreenPorch  PoolArea    MiscVal    MoSold       YrSold      SalePrice
0     0.000000  0.000000   0.000000  7.000000  2008.142857  169497.000000
1    51.250000  0.000000  56.250000  8.000000  2008.125000  190050.000000
2    12.277778  0.000000   0.000000  5.972222  2007.722222  120531.388889
3    15.391667  4.266667  34.750000  6.216667  2007.733333  128200.766667
4    14.337209  0.000000  85.000000  6.360465  2007.837209  139340.174419
5     9.250000  0.000000  51.203125  6.273438  2007.859375  130376.007812
6    28.678161  2.827586  96.283525  6.402299  2007.854406  150078.130268
7    15.008969  7.816143  61.883408  6.044843  2007.923767  155487.381166
8    13.187050  3.992806  26.474820  6.453237  2007.812950  225521.007194

[9 rows x 38 columns]
```

In [28]:
```
year_group = pd.cut(test_df.YearBuilt, np.arange(1870, 2010, 14))
test_df.groupby(test_df.YearBuilt,as_index=False).mean()
```

Out[28]:
```
    YearBuilt           Id  MSSubClass  LotFrontage       LotArea  \
0        1872  1350.000000   70.000000    50.000000   5250.000000
1        1875  1138.000000   50.000000    54.000000   6342.000000
2        1880   704.250000   71.250000    73.000000  12246.500000
3        1882   992.000000   70.000000   121.000000  17671.000000
4        1885   762.000000  110.000000    60.000000  11070.000000
5        1890   831.500000   60.000000    79.000000   9810.000000
6        1892   739.500000   72.500000    75.000000  14775.000000
7        1893   584.000000   75.000000    75.000000  13500.000000
8        1898   810.000000   75.000000    90.000000   8100.000000
9        1900   690.900000  126.000000    69.400000   9393.900000
10       1904   842.000000   70.000000    60.000000  10440.000000
11       1905  1394.000000  190.000000    60.000000  10800.000000
12       1906   654.000000   50.000000    60.000000  10320.000000
13       1908   517.000000   50.000000    83.500000  26519.000000
```

14	1910	789.235294	85.294118	60.647059	8106.294118
15	1911	1235.000000	70.000000	55.000000	8525.000000
16	1912	458.333333	113.333333	69.000000	5803.666667
17	1913	439.000000	30.000000	40.000000	4280.000000
18	1914	927.428571	77.142857	57.285714	9210.000000
19	1915	495.200000	60.000000	68.400000	9134.900000
20	1916	854.625000	71.250000	51.800000	8121.875000
21	1917	1249.000000	75.000000	60.000000	9600.000000
22	1918	689.571429	50.714286	80.166667	17302.142857
23	1919	886.000000	56.666667	65.000000	10551.000000
24	1920	748.333333	55.000000	61.642857	8349.666667
25	1921	681.333333	46.666667	57.500000	8763.833333
26	1922	1001.500000	39.375000	59.000000	7805.500000
27	1923	727.857143	47.142857	75.428571	11360.857143
28	1924	599.000000	43.571429	54.571429	7628.714286
29	1925	709.437500	50.000000	54.071429	7446.312500
..	...	...	...	...	...
82	1981	593.000000	20.000000	75.333333	20235.200000
83	1982	604.500000	41.666667	67.800000	7517.500000
84	1983	508.750000	43.750000	75.000000	9894.750000
85	1984	684.333333	71.111111	62.333333	8885.111111
86	1985	769.800000	68.000000	77.000000	8513.800000
87	1986	671.600000	76.000000	59.666667	10461.400000
88	1987	536.000000	90.000000	52.666667	9972.333333
89	1988	517.818182	57.272727	93.250000	11843.454545
90	1989	467.000000	66.666667	62.000000	16755.000000
91	1990	784.750000	50.000000	85.555556	11576.500000
92	1991	893.200000	48.000000	95.666667	11566.600000
93	1992	790.615385	61.153846	75.714286	13347.000000
94	1993	530.470588	73.529412	67.538462	9724.176471
95	1994	852.000000	43.157895	76.937500	13492.736842
96	1995	741.777778	56.111111	78.642857	11953.888889
97	1996	668.000000	61.666667	75.769231	11584.400000
98	1997	489.857143	56.428571	74.545455	10778.714286
99	1998	714.240000	67.200000	62.666667	10135.040000
100	1999	774.560000	90.400000	57.190476	7747.240000
101	2000	783.916667	78.333333	69.818182	8198.833333
102	2001	709.100000	48.000000	70.000000	10407.500000
103	2002	832.608696	43.478261	79.200000	10723.260870
104	2003	674.177778	67.222222	67.842105	9658.533333
105	2004	642.796296	69.259259	68.787234	8526.259259
106	2005	791.718750	73.437500	66.562500	9304.546875
107	2006	779.179104	51.940299	76.417910	10402.432836
108	2007	692.795918	42.857143	78.224490	10456.306122
109	2008	869.956522	42.608696	92.260870	14081.956522
110	2009	592.222222	52.222222	74.882353	8862.277778
111	2010	379.000000	20.000000	88.000000	11394.000000

	OverallQual	OverallCond	YearRemodAdd	MasVnrArea	BsmtFinSF1	\
0	8.000000	5.000000	1987.000000	0.000000	259.000000	
1	5.000000	8.000000	1996.000000	0.000000	0.000000	
2	6.250000	6.500000	1986.500000	0.000000	0.000000	
3	8.000000	9.000000	1986.000000	0.000000	216.000000	
4	4.000000	6.500000	1972.500000	0.000000	0.000000	
5	6.000000	7.000000	1997.000000	0.000000	544.000000	
6	7.500000	6.500000	1979.000000	0.000000	0.000000	
7	10.000000	9.000000	2000.000000	0.000000	0.000000	
8	5.000000	5.000000	1965.000000	0.000000	0.000000	
9	4.500000	5.800000	1968.900000	0.000000	242.900000	
10	5.000000	8.000000	2002.000000	0.000000	0.000000	
11	6.000000	7.000000	2000.000000	0.000000	0.000000	
12	6.000000	7.000000	1995.000000	0.000000	0.000000	
13	5.500000	8.000000	1994.000000	0.000000	0.000000	
14	5.176471	6.352941	1983.529412	0.000000	133.058824	
15	5.000000	6.000000	1950.000000	0.000000	0.000000	
16	5.000000	5.666667	1966.666667	0.000000	70.000000	
17	5.000000	6.000000	2002.000000	0.000000	365.000000	
18	5.142857	5.428571	1974.571429	0.000000	127.142857	
19	5.100000	5.700000	1968.600000	0.000000	120.100000	
20	5.500000	6.500000	1970.750000	0.000000	45.625000	
21	6.000000	5.000000	1950.000000	0.000000	319.000000	
22	6.142857	6.571429	1977.857143	0.000000	339.142857	
23	5.666667	6.666667	1981.666667	0.000000	0.000000	
24	5.100000	5.866667	1967.300000	0.000000	211.900000	
25	5.333333	6.333333	1970.333333	0.000000	221.500000	
26	4.000000	5.375000	1962.625000	0.000000	135.625000	
27	5.428571	6.428571	1980.571429	0.000000	229.857143	
28	5.285714	6.142857	1958.000000	0.000000	345.142857	
29	5.750000	6.312500	1972.375000	0.000000	202.812500	
..	...	...	...	...	...	
82	7.400000	5.800000	1982.200000	234.600000	836.000000	
83	4.833333	6.833333	1993.833333	0.000000	654.166667	
84	6.000000	6.250000	1989.500000	0.000000	706.250000	
85	6.333333	5.666667	1987.888889	8.222222	575.000000	
86	6.600000	5.000000	1985.000000	84.200000	679.800000	
87	7.000000	5.200000	1987.200000	85.400000	501.600000	
88	6.000000	6.000000	1987.333333	24.000000	891.666667	
89	6.818182	5.545455	1991.727273	115.181818	728.272727	
90	7.000000	5.333333	1989.333333	259.333333	968.000000	
91	7.000000	5.166667	1990.500000	99.166667	796.500000	
92	6.600000	5.000000	1991.600000	89.600000	292.000000	
93	6.692308	5.153846	1993.384615	47.153846	556.230769	
94	6.764706	5.529412	1994.058824	100.882353	592.352941	
95	6.684211	5.210526	1995.789474	200.842105	776.684211	
96	7.166667	5.111111	1996.222222	245.777778	575.000000	
97	6.866667	5.066667	1996.466667	139.400000	735.800000	

98	6.428571	5.071429	1997.428571	235.428571	628.000000	
99	6.960000	5.040000	1998.280000	210.000000	501.000000	
100	7.160000	5.080000	1999.600000	292.360000	501.520000	
101	6.916667	5.000000	2000.083333	114.500000	436.958333	
102	7.550000	5.050000	2001.500000	133.300000	500.400000	
103	7.391304	5.000000	2002.173913	136.619048	558.782609	
104	7.155556	5.022222	2003.377778	196.090909	482.466667	
105	7.092593	5.000000	2004.462963	93.925926	466.722222	
106	7.078125	5.000000	2005.531250	123.140625	390.937500	
107	7.567164	5.000000	2006.313433	169.707692	333.223881	
108	7.734694	5.081633	2007.306122	181.510638	500.224490	
109	8.652174	5.000000	2008.304348	320.086957	831.217391	
110	7.444444	5.000000	2009.222222	162.388889	576.944444	
111	9.000000	2.000000	2010.000000	350.000000	1445.000000	

	...	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	\
0	...	0.000000	54.000000	20.000000	0.000000	
1	...	0.000000	0.000000	176.000000	0.000000	
2	...	120.750000	12.000000	169.750000	0.000000	
3	...	0.000000	169.000000	0.000000	0.000000	
4	...	0.000000	0.000000	57.000000	0.000000	
5	...	24.000000	24.000000	193.500000	0.000000	
6	...	0.000000	273.500000	0.000000	0.000000	
7	...	105.000000	502.000000	0.000000	0.000000	
8	...	40.000000	156.000000	0.000000	0.000000	
9	...	77.400000	44.500000	62.700000	0.000000	
10	...	70.000000	78.000000	68.000000	0.000000	
11	...	0.000000	25.000000	212.000000	0.000000	
12	...	57.000000	0.000000	239.000000	0.000000	
13	...	295.000000	20.000000	0.000000	0.000000	
14	...	121.764706	42.235294	72.647059	0.000000	
15	...	0.000000	192.000000	0.000000	0.000000	
16	...	0.000000	190.000000	10.000000	0.000000	
17	...	0.000000	0.000000	34.000000	0.000000	
18	...	48.428571	111.714286	63.571429	0.000000	
19	...	63.400000	10.100000	147.600000	0.000000	
20	...	11.000000	21.625000	115.250000	17.500000	
21	...	0.000000	0.000000	259.000000	0.000000	
22	...	39.857143	41.714286	124.571429	0.000000	
23	...	90.666667	0.000000	104.000000	0.000000	
24	...	17.000000	56.400000	89.200000	0.000000	
25	...	64.500000	1.833333	125.333333	0.000000	
26	...	53.875000	36.500000	41.375000	0.000000	
27	...	13.714286	50.000000	74.857143	20.571429	
28	...	45.142857	1.142857	138.285714	0.000000	
29	...	66.750000	30.875000	80.500000	0.000000	
..	...	...	...	...	...	
82	...	436.200000	51.200000	19.800000	0.000000	

83	...	48.000000	0.000000	0.000000	0.000000
84	...	218.250000	10.000000	39.500000	0.000000
85	...	165.888889	17.444444	0.000000	0.000000
86	...	94.200000	30.800000	46.800000	0.000000
87	...	151.400000	62.600000	27.400000	0.000000
88	...	228.666667	11.666667	0.000000	0.000000
89	...	135.818182	51.272727	29.545455	16.545455
90	...	133.333333	37.000000	0.000000	65.333333
91	...	133.833333	55.500000	18.666667	42.333333
92	...	95.000000	70.600000	44.800000	0.000000
93	...	165.923077	68.230769	26.384615	0.000000
94	...	162.411765	70.411765	36.588235	42.705882
95	...	168.315789	67.473684	10.105263	0.000000
96	...	163.555556	69.722222	11.222222	0.000000
97	...	174.800000	52.600000	10.266667	0.000000
98	...	125.571429	55.785714	13.214286	0.000000
99	...	100.240000	82.000000	0.000000	0.000000
100	...	87.840000	51.240000	0.000000	6.720000
101	...	97.291667	63.458333	0.000000	0.000000
102	...	131.450000	80.450000	8.500000	12.250000
103	...	130.956522	77.043478	0.000000	0.000000
104	...	128.088889	60.422222	3.933333	0.000000
105	...	107.222222	58.185185	0.000000	2.666667
106	...	117.703125	58.968750	0.000000	2.984375
107	...	127.626866	67.701493	0.552239	0.000000
108	...	107.448980	82.244898	0.000000	0.000000
109	...	104.347826	99.913043	0.000000	19.869565
110	...	82.833333	103.944444	0.000000	0.000000
111	...	113.000000	0.000000	0.000000	0.000000

	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice
0	0.000000	0.000000	0.000000	12.000000	2008.000000	122000.000000
1	0.000000	0.000000	0.000000	5.000000	2010.000000	94000.000000
2	0.000000	0.000000	0.000000	5.250000	2007.500000	200619.750000
3	0.000000	0.000000	0.000000	11.000000	2009.000000	168000.000000
4	0.000000	0.000000	225.000000	6.000000	2008.500000	111250.000000
5	0.000000	0.000000	0.000000	8.000000	2008.000000	142200.000000
6	205.000000	0.000000	0.000000	9.000000	2007.500000	291250.000000
7	0.000000	0.000000	0.000000	7.000000	2008.000000	325000.000000
8	0.000000	0.000000	0.000000	11.000000	2009.000000	106000.000000
9	0.000000	0.000000	0.000000	5.800000	2007.800000	112492.600000
10	0.000000	0.000000	0.000000	6.000000	2008.000000	157500.000000
11	0.000000	0.000000	0.000000	4.000000	2008.000000	163000.000000
12	0.000000	0.000000	0.000000	6.000000	2008.000000	135000.000000
13	87.500000	0.000000	0.000000	7.500000	2007.000000	183500.000000
14	15.705882	0.000000	0.000000	6.000000	2007.705882	111959.058824
15	0.000000	0.000000	0.000000	11.000000	2008.000000	130000.000000
16	0.000000	0.000000	0.000000	4.333333	2007.666667	119466.666667

```
17      0.000000   0.000000    0.000000   3.000000  2007.000000   90350.000000
18      0.000000   0.000000    0.000000   6.285714  2007.428571  126055.285714
19     25.900000   0.000000  120.000000   5.100000  2008.300000  134387.500000
20      5.000000   0.000000    0.000000   6.750000  2007.625000  139800.000000
21      0.000000   0.000000    0.000000   4.000000  2008.000000  129500.000000
22     20.571429  73.142857    0.000000   6.714286  2007.000000  172421.428571
23     38.666667   0.000000    0.000000   6.000000  2008.000000  154833.333333
24     20.200000   0.000000   78.333333   6.566667  2007.866667  112903.333333
25      0.000000   0.000000    0.000000   7.000000  2008.000000  140813.333333
26      0.000000   0.000000    0.000000   5.625000  2007.750000  109237.500000
27     39.000000   0.000000    0.000000   6.000000  2007.142857  159428.571429
28      0.000000   0.000000    0.000000   5.285714  2008.000000  105985.714286
29     16.500000   0.000000   38.750000   6.375000  2007.750000  129190.625000
..           ...        ...         ...        ...          ...            ...
82      0.000000   0.000000    0.000000   5.800000  2007.200000  249880.000000
83      0.000000   0.000000    0.000000   6.500000  2008.000000  132400.000000
84      0.000000   0.000000    0.000000   8.250000  2007.250000  181400.000000
85     26.333333   0.000000   44.444444   5.444444  2007.888889  164111.111111
86     28.000000   0.000000    0.000000   6.400000  2008.200000  191400.000000
87     36.800000   0.000000    0.000000  10.400000  2007.200000  212000.000000
88      0.000000   0.000000    0.000000   5.333333  2007.000000  198000.000000
89     31.454545   0.000000   45.454545   6.363636  2007.818182  204636.363636
90    132.000000   0.000000    0.000000   4.333333  2007.333333  279500.000000
91      7.500000   0.000000    0.000000   5.333333  2007.666667  220881.666667
92      0.000000   0.000000    0.000000   7.000000  2007.000000  186140.000000
93      0.000000   0.000000   30.769231   6.076923  2007.769231  218384.615385
94      0.000000   0.000000   69.411765   6.705882  2008.176471  218058.823529
95      0.000000   0.000000   26.315789   6.052632  2008.052632  247900.000000
96     12.000000   0.000000   38.888889   6.500000  2007.611111  262194.444444
97     14.933333  37.000000    0.000000   7.200000  2008.200000  259162.666667
98     14.142857   0.000000   28.571429   6.642857  2007.285714  214925.000000
99     26.600000   0.000000    0.000000   6.200000  2007.760000  220450.000000
100     7.200000   0.000000    0.000000   6.040000  2008.320000  210614.720000
101     0.000000   0.000000    0.000000   6.833333  2007.791667  210766.666667
102    20.650000   0.000000  100.000000   6.200000  2008.450000  242630.000000
103     0.000000   0.000000    0.000000   6.434783  2007.739130  226869.565217
104     9.066667   0.000000    0.000000   6.600000  2007.955556  227408.577778
105    12.018519   0.000000    0.000000   5.925926  2007.740741  210347.722222
106     1.875000   0.000000    0.000000   5.484375  2007.062500  229680.953125
107     4.402985   0.000000    0.000000   7.000000  2006.970149  251775.447761
108    14.163265   0.000000    0.000000   7.040816  2007.836735  255362.734694
109    25.478261  20.869565    0.000000   6.304348  2008.826087  348849.130435
110     0.000000   0.000000    0.000000   6.777778  2009.444444  269220.000000
111     0.000000   0.000000    0.000000   6.000000  2010.000000  394432.000000

[112 rows x 38 columns]
```

In [29]:
```
year_group = pd.cut(test_df.YearBuilt, np.arange(1870, 2010, 14))
```

```
test_df.groupby(test_df.YearBuilt,as_index=False).mean()
```

Out[29]:

	YearBuilt	Id	MSSubClass	LotFrontage	LotArea	\
0	1872	1350.000000	70.000000	50.000000	5250.000000	
1	1875	1138.000000	50.000000	54.000000	6342.000000	
2	1880	704.250000	71.250000	73.000000	12246.500000	
3	1882	992.000000	70.000000	121.000000	17671.000000	
4	1885	762.000000	110.000000	60.000000	11070.000000	
5	1890	831.500000	60.000000	79.000000	9810.000000	
6	1892	739.500000	72.500000	75.000000	14775.000000	
7	1893	584.000000	75.000000	75.000000	13500.000000	
8	1898	810.000000	75.000000	90.000000	8100.000000	
9	1900	690.900000	126.000000	69.400000	9393.900000	
10	1904	842.000000	70.000000	60.000000	10440.000000	
11	1905	1394.000000	190.000000	60.000000	10800.000000	
12	1906	654.000000	50.000000	60.000000	10320.000000	
13	1908	517.000000	50.000000	83.500000	26519.000000	
14	1910	789.235294	85.294118	60.647059	8106.294118	
15	1911	1235.000000	70.000000	55.000000	8525.000000	
16	1912	458.333333	113.333333	69.000000	5803.666667	
17	1913	439.000000	30.000000	40.000000	4280.000000	
18	1914	927.428571	77.142857	57.285714	9210.000000	
19	1915	495.200000	60.000000	68.400000	9134.900000	
20	1916	854.625000	71.250000	51.800000	8121.875000	
21	1917	1249.000000	75.000000	60.000000	9600.000000	
22	1918	689.571429	50.714286	80.166667	17302.142857	
23	1919	886.000000	56.666667	65.000000	10551.000000	
24	1920	748.333333	55.000000	61.642857	8349.666667	
25	1921	681.333333	46.666667	57.500000	8763.833333	
26	1922	1001.500000	39.375000	59.000000	7805.500000	
27	1923	727.857143	47.142857	75.428571	11360.857143	
28	1924	599.000000	43.571429	54.571429	7628.714286	
29	1925	709.437500	50.000000	54.071429	7446.312500	
..	...	...	...	...	...	
82	1981	593.000000	20.000000	75.333333	20235.200000	
83	1982	604.500000	41.666667	67.800000	7517.500000	
84	1983	508.750000	43.750000	75.000000	9894.750000	
85	1984	684.333333	71.111111	62.333333	8885.111111	
86	1985	769.800000	68.000000	77.000000	8513.800000	
87	1986	671.600000	76.000000	59.666667	10461.400000	
88	1987	536.000000	90.000000	52.666667	9972.333333	
89	1988	517.818182	57.272727	93.250000	11843.454545	
90	1989	467.000000	66.666667	62.000000	16755.000000	
91	1990	784.750000	50.000000	85.555556	11576.500000	
92	1991	893.200000	48.000000	95.666667	11566.600000	
93	1992	790.615385	61.153846	75.714286	13347.000000	
94	1993	530.470588	73.529412	67.538462	9724.176471	
95	1994	852.000000	43.157895	76.937500	13492.736842	

96	1995	741.777778	56.111111	78.642857	11953.888889
97	1996	668.000000	61.666667	75.769231	11584.400000
98	1997	489.857143	56.428571	74.545455	10778.714286
99	1998	714.240000	67.200000	62.666667	10135.040000
100	1999	774.560000	90.400000	57.190476	7747.240000
101	2000	783.916667	78.333333	69.818182	8198.833333
102	2001	709.100000	48.000000	70.000000	10407.500000
103	2002	832.608696	43.478261	79.200000	10723.260870
104	2003	674.177778	67.222222	67.842105	9658.533333
105	2004	642.796296	69.259259	68.787234	8526.259259
106	2005	791.718750	73.437500	66.562500	9304.546875
107	2006	779.179104	51.940299	76.417910	10402.432836
108	2007	692.795918	42.857143	78.224490	10456.306122
109	2008	869.956522	42.608696	92.260870	14081.956522
110	2009	592.222222	52.222222	74.882353	8862.277778
111	2010	379.000000	20.000000	88.000000	11394.000000

	OverallQual	OverallCond	YearRemodAdd	MasVnrArea	BsmtFinSF1	\
0	8.000000	5.000000	1987.000000	0.000000	259.000000	
1	5.000000	8.000000	1996.000000	0.000000	0.000000	
2	6.250000	6.500000	1986.500000	0.000000	0.000000	
3	8.000000	9.000000	1986.000000	0.000000	216.000000	
4	4.000000	6.500000	1972.500000	0.000000	0.000000	
5	6.000000	7.000000	1997.000000	0.000000	544.000000	
6	7.500000	6.500000	1979.000000	0.000000	0.000000	
7	10.000000	9.000000	2000.000000	0.000000	0.000000	
8	5.000000	5.000000	1965.000000	0.000000	0.000000	
9	4.500000	5.800000	1968.900000	0.000000	242.900000	
10	5.000000	8.000000	2002.000000	0.000000	0.000000	
11	6.000000	7.000000	2000.000000	0.000000	0.000000	
12	6.000000	7.000000	1995.000000	0.000000	0.000000	
13	5.500000	8.000000	1994.000000	0.000000	0.000000	
14	5.176471	6.352941	1983.529412	0.000000	133.058824	
15	5.000000	6.000000	1950.000000	0.000000	0.000000	
16	5.000000	5.666667	1966.666667	0.000000	70.000000	
17	5.000000	6.000000	2002.000000	0.000000	365.000000	
18	5.142857	5.428571	1974.571429	0.000000	127.142857	
19	5.100000	5.700000	1968.600000	0.000000	120.100000	
20	5.500000	6.500000	1970.750000	0.000000	45.625000	
21	6.000000	5.000000	1950.000000	0.000000	319.000000	
22	6.142857	6.571429	1977.857143	0.000000	339.142857	
23	5.666667	6.666667	1981.666667	0.000000	0.000000	
24	5.100000	5.866667	1967.300000	0.000000	211.900000	
25	5.333333	6.333333	1970.333333	0.000000	221.500000	
26	4.000000	5.375000	1962.625000	0.000000	135.625000	
27	5.428571	6.428571	1980.571429	0.000000	229.857143	
28	5.285714	6.142857	1958.000000	0.000000	345.142857	
29	5.750000	6.312500	1972.375000	0.000000	202.812500	

```
..          ...          ...          ...          ...          ...
82     7.400000     5.800000  1982.200000   234.600000   836.000000
83     4.833333     6.833333  1993.833333     0.000000   654.166667
84     6.000000     6.250000  1989.500000     0.000000   706.250000
85     6.333333     5.666667  1987.888889     8.222222   575.000000
86     6.600000     5.000000  1985.000000    84.200000   679.800000
87     7.000000     5.200000  1987.200000    85.400000   501.600000
88     6.000000     6.000000  1987.333333    24.000000   891.666667
89     6.818182     5.545455  1991.727273   115.181818   728.272727
90     7.000000     5.333333  1989.333333   259.333333   968.000000
91     7.000000     5.166667  1990.500000    99.166667   796.500000
92     6.600000     5.000000  1991.600000    89.600000   292.000000
93     6.692308     5.153846  1993.384615    47.153846   556.230769
94     6.764706     5.529412  1994.058824   100.882353   592.352941
95     6.684211     5.210526  1995.789474   200.842105   776.684211
96     7.166667     5.111111  1996.222222   245.777778   575.000000
97     6.866667     5.066667  1996.466667   139.400000   735.800000
98     6.428571     5.071429  1997.428571   235.428571   628.000000
99     6.960000     5.040000  1998.280000   210.000000   501.000000
100    7.160000     5.080000  1999.600000   292.360000   501.520000
101    6.916667     5.000000  2000.083333   114.500000   436.958333
102    7.550000     5.050000  2001.500000   133.300000   500.400000
103    7.391304     5.000000  2002.173913   136.619048   558.782609
104    7.155556     5.022222  2003.377778   196.090909   482.466667
105    7.092593     5.000000  2004.462963    93.925926   466.722222
106    7.078125     5.000000  2005.531250   123.140625   390.937500
107    7.567164     5.000000  2006.313433   169.707692   333.223881
108    7.734694     5.081633  2007.306122   181.510638   500.224490
109    8.652174     5.000000  2008.304348   320.086957   831.217391
110    7.444444     5.000000  2009.222222   162.388889   576.944444
111    9.000000     2.000000  2010.000000   350.000000  1445.000000

          ...       WoodDeckSF  OpenPorchSF  EnclosedPorch  3SsnPorch  \
0         ...         0.000000    54.000000      20.000000   0.000000
1         ...         0.000000     0.000000     176.000000   0.000000
2         ...       120.750000    12.000000     169.750000   0.000000
3         ...         0.000000   169.000000       0.000000   0.000000
4         ...         0.000000     0.000000      57.000000   0.000000
5         ...        24.000000    24.000000     193.500000   0.000000
6         ...         0.000000   273.500000       0.000000   0.000000
7         ...       105.000000   502.000000       0.000000   0.000000
8         ...        40.000000   156.000000       0.000000   0.000000
9         ...        77.400000    44.500000      62.700000   0.000000
10        ...        70.000000    78.000000      68.000000   0.000000
11        ...         0.000000    25.000000     212.000000   0.000000
12        ...        57.000000     0.000000     239.000000   0.000000
13        ...       295.000000    20.000000       0.000000   0.000000
14        ...       121.764706    42.235294      72.647059   0.000000
```

15	...	0.000000	192.000000	0.000000	0.000000
16	...	0.000000	190.000000	10.000000	0.000000
17	...	0.000000	0.000000	34.000000	0.000000
18	...	48.428571	111.714286	63.571429	0.000000
19	...	63.400000	10.100000	147.600000	0.000000
20	...	11.000000	21.625000	115.250000	17.500000
21	...	0.000000	0.000000	259.000000	0.000000
22	...	39.857143	41.714286	124.571429	0.000000
23	...	90.666667	0.000000	104.000000	0.000000
24	...	17.000000	56.400000	89.200000	0.000000
25	...	64.500000	1.833333	125.333333	0.000000
26	...	53.875000	36.500000	41.375000	0.000000
27	...	13.714286	50.000000	74.857143	20.571429
28	...	45.142857	1.142857	138.285714	0.000000
29	...	66.750000	30.875000	80.500000	0.000000
..	...	...	...	...	...
82	...	436.200000	51.200000	19.800000	0.000000
83	...	48.000000	0.000000	0.000000	0.000000
84	...	218.250000	10.000000	39.500000	0.000000
85	...	165.888889	17.444444	0.000000	0.000000
86	...	94.200000	30.800000	46.800000	0.000000
87	...	151.400000	62.600000	27.400000	0.000000
88	...	228.666667	11.666667	0.000000	0.000000
89	...	135.818182	51.272727	29.545455	16.545455
90	...	133.333333	37.000000	0.000000	65.333333
91	...	133.833333	55.500000	18.666667	42.333333
92	...	95.000000	70.600000	44.800000	0.000000
93	...	165.923077	68.230769	26.384615	0.000000
94	...	162.411765	70.411765	36.588235	42.705882
95	...	168.315789	67.473684	10.105263	0.000000
96	...	163.555556	69.722222	11.222222	0.000000
97	...	174.800000	52.600000	10.266667	0.000000
98	...	125.571429	55.785714	13.214286	0.000000
99	...	100.240000	82.000000	0.000000	0.000000
100	...	87.840000	51.240000	0.000000	6.720000
101	...	97.291667	63.458333	0.000000	0.000000
102	...	131.450000	80.450000	8.500000	12.250000
103	...	130.956522	77.043478	0.000000	0.000000
104	...	128.088889	60.422222	3.933333	0.000000
105	...	107.222222	58.185185	0.000000	2.666667
106	...	117.703125	58.968750	0.000000	2.984375
107	...	127.626866	67.701493	0.552239	0.000000
108	...	107.448980	82.244898	0.000000	0.000000
109	...	104.347826	99.913043	0.000000	19.869565
110	...	82.833333	103.944444	0.000000	0.000000
111	...	113.000000	0.000000	0.000000	0.000000

	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice

0	0.000000	0.000000	0.000000	12.000000	2008.000000	122000.000000
1	0.000000	0.000000	0.000000	5.000000	2010.000000	94000.000000
2	0.000000	0.000000	0.000000	5.250000	2007.500000	200619.750000
3	0.000000	0.000000	0.000000	11.000000	2009.000000	168000.000000
4	0.000000	0.000000	225.000000	6.000000	2008.500000	111250.000000
5	0.000000	0.000000	0.000000	8.000000	2008.000000	142200.000000
6	205.000000	0.000000	0.000000	9.000000	2007.500000	291250.000000
7	0.000000	0.000000	0.000000	7.000000	2008.000000	325000.000000
8	0.000000	0.000000	0.000000	11.000000	2009.000000	106000.000000
9	0.000000	0.000000	0.000000	5.800000	2007.800000	112492.600000
10	0.000000	0.000000	0.000000	6.000000	2008.000000	157500.000000
11	0.000000	0.000000	0.000000	4.000000	2008.000000	163000.000000
12	0.000000	0.000000	0.000000	6.000000	2008.000000	135000.000000
13	87.500000	0.000000	0.000000	7.500000	2007.000000	183500.000000
14	15.705882	0.000000	0.000000	6.000000	2007.705882	111959.058824
15	0.000000	0.000000	0.000000	11.000000	2008.000000	130000.000000
16	0.000000	0.000000	0.000000	4.333333	2007.666667	119466.666667
17	0.000000	0.000000	0.000000	3.000000	2007.000000	90350.000000
18	0.000000	0.000000	0.000000	6.285714	2007.428571	126055.285714
19	25.900000	0.000000	120.000000	5.100000	2008.300000	134387.500000
20	5.000000	0.000000	0.000000	6.750000	2007.625000	139800.000000
21	0.000000	0.000000	0.000000	4.000000	2008.000000	129500.000000
22	20.571429	73.142857	0.000000	6.714286	2007.000000	172421.428571
23	38.666667	0.000000	0.000000	6.000000	2008.000000	154833.333333
24	20.200000	0.000000	78.333333	6.566667	2007.866667	112903.333333
25	0.000000	0.000000	0.000000	7.000000	2008.000000	140813.333333
26	0.000000	0.000000	0.000000	5.625000	2007.750000	109237.500000
27	39.000000	0.000000	0.000000	6.000000	2007.142857	159428.571429
28	0.000000	0.000000	0.000000	5.285714	2008.000000	105985.714286
29	16.500000	0.000000	38.750000	6.375000	2007.750000	129190.625000
..	...	...	...	...	...	...
82	0.000000	0.000000	0.000000	5.800000	2007.200000	249880.000000
83	0.000000	0.000000	0.000000	6.500000	2008.000000	132400.000000
84	0.000000	0.000000	0.000000	8.250000	2007.250000	181400.000000
85	26.333333	0.000000	44.444444	5.444444	2007.888889	164111.111111
86	28.000000	0.000000	0.000000	6.400000	2008.200000	191400.000000
87	36.800000	0.000000	0.000000	10.400000	2007.200000	212000.000000
88	0.000000	0.000000	0.000000	5.333333	2007.000000	198000.000000
89	31.454545	0.000000	45.454545	6.363636	2007.818182	204636.363636
90	132.000000	0.000000	0.000000	4.333333	2007.333333	279500.000000
91	7.500000	0.000000	0.000000	5.333333	2007.666667	220881.666667
92	0.000000	0.000000	0.000000	7.000000	2007.000000	186140.000000
93	0.000000	0.000000	30.769231	6.076923	2007.769231	218384.615385
94	0.000000	0.000000	69.411765	6.705882	2008.176471	218058.823529
95	0.000000	0.000000	26.315789	6.052632	2008.052632	247900.000000
96	12.000000	0.000000	38.888889	6.500000	2007.611111	262194.444444
97	14.933333	37.000000	0.000000	7.200000	2008.200000	259162.666667
98	14.142857	0.000000	28.571429	6.642857	2007.285714	214925.000000

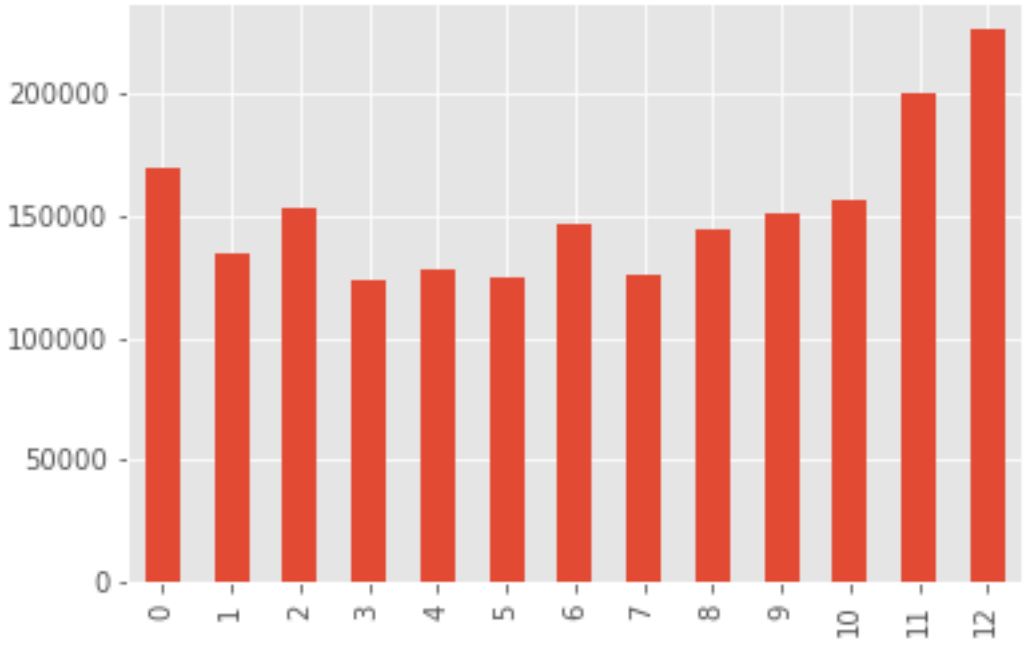
```
99     26.600000   0.000000    0.000000   6.200000  2007.760000  220450.000000
100     7.200000   0.000000    0.000000   6.040000  2008.320000  210614.720000
101     0.000000   0.000000    0.000000   6.833333  2007.791667  210766.666667
102    20.650000   0.000000  100.000000   6.200000  2008.450000  242630.000000
103     0.000000   0.000000    0.000000   6.434783  2007.739130  226869.565217
104     9.066667   0.000000    0.000000   6.600000  2007.955556  227408.577778
105    12.018519   0.000000    0.000000   5.925926  2007.740741  210347.722222
106     1.875000   0.000000    0.000000   5.484375  2007.062500  229680.953125
107     4.402985   0.000000    0.000000   7.000000  2006.970149  251775.447761
108    14.163265   0.000000    0.000000   7.040816  2007.836735  255362.734694
109    25.478261  20.869565    0.000000   6.304348  2008.826087  348849.130435
110     0.000000   0.000000    0.000000   6.777778  2009.444444  269220.000000
111     0.000000   0.000000    0.000000   6.000000  2010.000000  394432.000000

[112 rows x 38 columns]
```

In [30]:
```
year_group = pd.cut(test_df.YearBuilt, np.arange(1870, 2010, 10))
year_group
```

Out[30]:
```
0                NaN
1       (1970, 1980]
2                NaN
3       (1910, 1920]
4       (1990, 2000]
5       (1990, 2000]
6                NaN
7       (1970, 1980]
8       (1930, 1940]
9       (1930, 1940]
10      (1960, 1970]
11               NaN
12      (1960, 1970]
13               NaN
14      (1950, 1960]
15      (1920, 1930]
16      (1960, 1970]
17      (1960, 1970]
18               NaN
19      (1950, 1960]
20               NaN
21      (1920, 1930]
22               NaN
23      (1970, 1980]
24      (1960, 1970]
25               NaN
26      (1950, 1960]
27               NaN
28      (1950, 1960]
```

```
29      (1920, 1930]
            ...
1430             NaN
1431    (1970, 1980]
1432    (1920, 1930]
1433    (1990, 2000]
1434    (1970, 1980]
1435    (1960, 1970]
1436    (1970, 1980]
1437             NaN
1438    (1950, 1960]
1439    (1970, 1980]
1440    (1920, 1930]
1441             NaN
1442             NaN
1443    (1910, 1920]
1444             NaN
1445    (1960, 1970]
1446    (1960, 1970]
1447    (1990, 2000]
1448    (1900, 1910]
1449    (1960, 1970]
1450    (1970, 1980]
1451             NaN
1452             NaN
1453             NaN
1454             NaN
1455    (1990, 2000]
1456    (1970, 1980]
1457    (1940, 1950]
1458    (1940, 1950]
1459    (1960, 1970]
Name: YearBuilt, Length: 1460, dtype: category
Categories (13, interval[int64]): [(1870, 1880] < (1880, 1890] < (1890, 1900] < (1900,
```

In [31]:
```
year_group = pd.cut(test_df.YearBuilt, np.arange(1870, 2010, 10))
year_class = test_df.groupby(year_group,as_index=False).mean()
year_class['SalePrice'].plot.bar()
plt.show()
```




```
In [32]: import pandas as pd
         test_df = pd.read_csv('train.csv')
         test_df.head()
```

Out[32]:

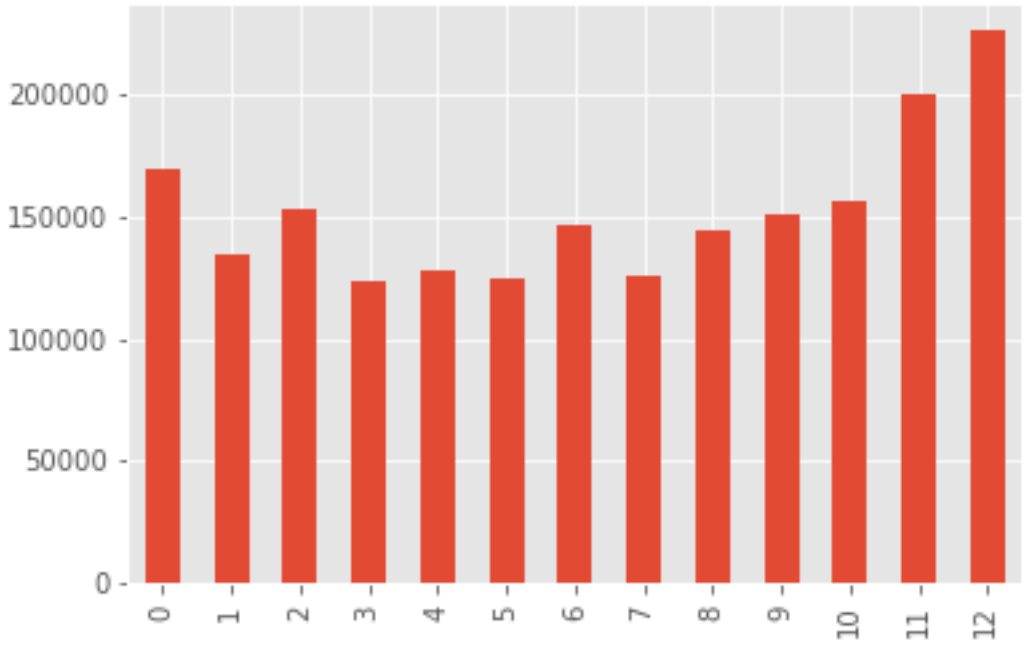
	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	

	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	\
0	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
2	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
3	Lvl	AllPub	...	0	NaN	NaN	NaN	0	
4	Lvl	AllPub	...	0	NaN	NaN	NaN	0	

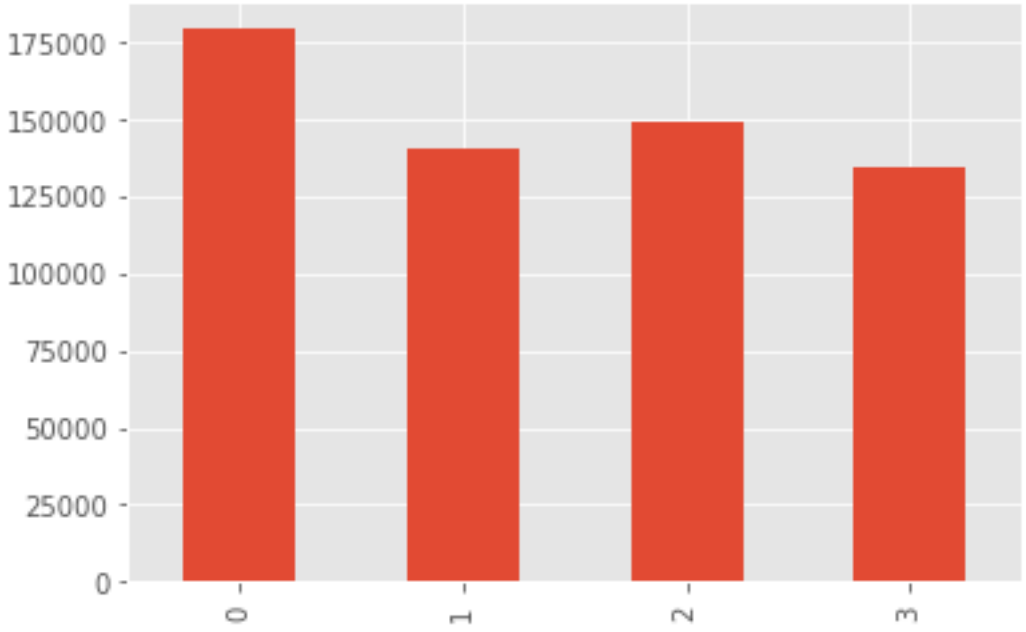
	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	2	2008	WD	Normal	208500
1	5	2007	WD	Normal	181500
2	9	2008	WD	Normal	223500
3	2	2006	WD	Abnorml	140000
4	12	2008	WD	Normal	250000

[5 rows x 81 columns]

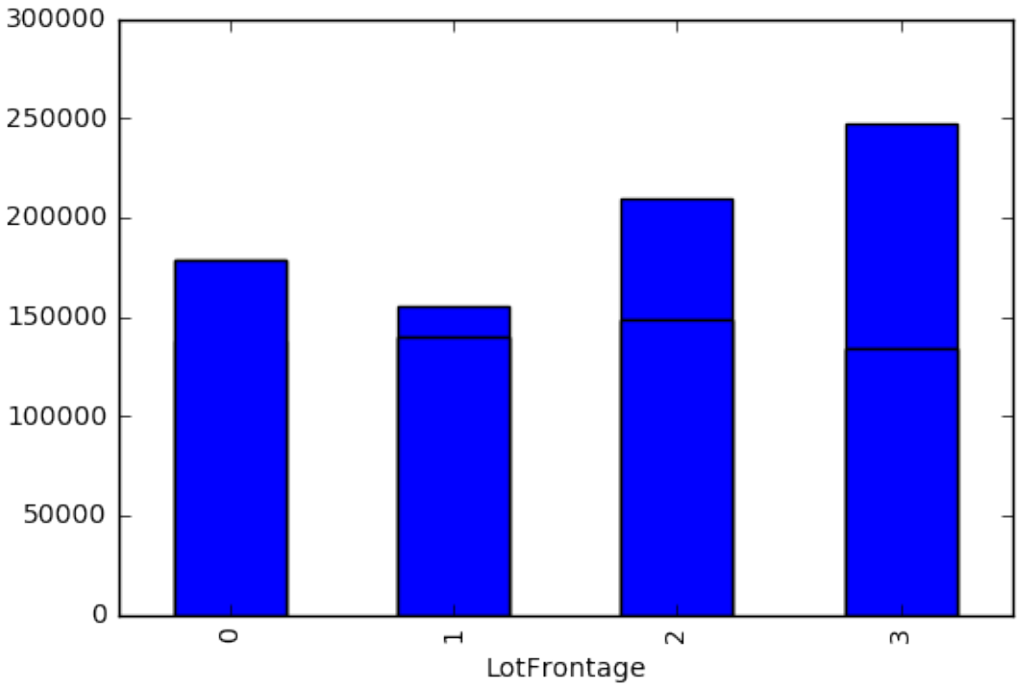
In [34]:
```
import pandas as pd
test_df = pd.read_csv('train.csv')
fence_class = test_df.groupby(test_df['Fence'],as_index=False).mean()
year_class['SalePrice'].plot.bar()
plt.show()
```



In [35]:
```
import pandas as pd
test_df = pd.read_csv('train.csv')
fence_class = test_df.groupby(test_df['Fence'],as_index=False).mean()
fence_class['SalePrice'].plot.bar()
plt.show()
```




```
In [7]: import pandas as pd
        test_df = pd.read_csv('train.csv')
        fence_class = test_df.groupby(test_df['Fence'],as_index=False).mean()
        fence_class['SalePrice'].plot.bar()
        plt.show()
```



```
In [8]: import pandas as pd
        test_df = pd.read_csv('train.csv')
        fence_class = test_df.groupby(test_df['Fence'],as_index=False).mean()
        fence_class
```

Out[8]:

	Fence	Id	MSSubClass	LotFrontage	LotArea	OverallQual	\
0	GdPrv	752.372881	57.118644	73.687500	10520.288136	6.305085	
1	GdWo	750.203704	40.000000	73.666667	9634.018519	5.166667	
2	MnPrv	704.286624	47.611465	70.562500	9204.458599	5.445860	
3	MnWw	743.454545	41.363636	69.400000	9250.909091	5.181818	

	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	...	\
0	6.084746	1961.152542	1984.169492	88.576271	...	
1	5.666667	1957.925926	1972.981481	89.314815	...	
2	5.980892	1955.942675	1977.203822	49.579618	...	
3	6.545455	1949.909091	1979.181818	48.454545	...	

	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	\
0	137.457627	52.118644	34.593220	2.847458	16.508475	41.932203	
1	87.222222	30.537037	30.888889	0.000000	13.666667	0.000000	
2	96.038217	34.942675	29.082803	4.687898	16.929936	6.840764	
3	107.454545	54.727273	32.363636	16.363636	11.090909	0.000000	

	MiscVal	MoSold	YrSold	SalePrice
0	79.661017	6.305085	2007.898305	178927.457627
1	37.962963	6.962963	2007.759259	140379.314815
2	82.929936	6.140127	2008.038217	148751.089172
3	43.636364	6.272727	2007.272727	134286.363636

[4 rows x 39 columns]

```
In [10]: s = pd.Series(list(test_df['Fence']))
         pd.get_dummies(s)
```

Out[10]:

	GdPrv	GdWo	MnPrv	MnWw
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	1	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

11	0	0	0	0
12	0	0	0	0
13	0	0	0	0
14	0	1	0	0
15	1	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	0	0	1	0
20	0	0	0	0
21	1	0	0	0
22	0	0	0	0
23	0	0	0	0
24	0	0	1	0
25	0	0	0	0
26	0	0	0	0
27	0	0	0	0
28	0	0	0	0
29	0	0	0	0
...	...	...	...	...
1430	0	0	0	0
1431	0	0	0	0
1432	0	0	0	0
1433	0	0	0	0
1434	0	0	0	0
1435	1	0	0	0
1436	0	1	0	0
1437	0	0	0	0
1438	0	0	1	0
1439	0	0	0	0
1440	0	0	0	0
1441	0	0	0	0
1442	0	0	0	0
1443	0	0	0	0
1444	0	0	0	0
1445	0	0	0	0
1446	0	0	0	0
1447	0	0	0	0
1448	0	1	0	0
1449	0	0	0	0
1450	0	0	0	0
1451	0	0	0	0
1452	0	0	0	0
1453	0	0	0	0
1454	0	0	0	0
1455	0	0	0	0
1456	0	0	1	0
1457	1	0	0	0

```
1458      0     0      0     0
1459      0     0      0     0

[1460 rows x 4 columns]
```

```
In [11]: s = pd.Series(list(test_df['Fence'])).rename(columns=lambda x: 'Fence_' + str(x))
         pd.get_dummies(s)
```

Out[11]:

	GdPrv	GdWo	MnPrv	MnWw
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	1	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0
14	0	1	0	0
15	1	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	0	0	1	0
20	0	0	0	0
21	1	0	0	0
22	0	0	0	0
23	0	0	0	0
24	0	0	1	0
25	0	0	0	0
26	0	0	0	0
27	0	0	0	0
28	0	0	0	0
29	0	0	0	0
...	...	...	...	...
1430	0	0	0	0
1431	0	0	0	0
1432	0	0	0	0
1433	0	0	0	0
1434	0	0	0	0
1435	1	0	0	0
1436	0	1	0	0
1437	0	0	0	0

```
1438      0     0      1     0
1439      0     0      0     0
1440      0     0      0     0
1441      0     0      0     0
1442      0     0      0     0
1443      0     0      0     0
1444      0     0      0     0
1445      0     0      0     0
1446      0     0      0     0
1447      0     0      0     0
1448      0     1      0     0
1449      0     0      0     0
1450      0     0      0     0
1451      0     0      0     0
1452      0     0      0     0
1453      0     0      0     0
1454      0     0      0     0
1455      0     0      0     0
1456      0     0      1     0
1457      1     0      0     0
1458      0     0      0     0
1459      0     0      0     0

[1460 rows x 4 columns]
```

```
In [12]: pd.get_dummies(test_df['Fence']).rename(columns=lambda x: 'Fence_' + str(x))
```

Out[12]:

	Fence_GdPrv	Fence_GdWo	Fence_MnPrv	Fence_MnWw
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	1	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0
14	0	1	0	0
15	1	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	0	0	1	0

```
20              0           0           0           0
21              1           0           0           0
22              0           0           0           0
23              0           0           0           0
24              0           0           1           0
25              0           0           0           0
26              0           0           0           0
27              0           0           0           0
28              0           0           0           0
29              0           0           0           0
...           ...         ...         ...         ...
1430            0           0           0           0
1431            0           0           0           0
1432            0           0           0           0
1433            0           0           0           0
1434            0           0           0           0
1435            1           0           0           0
1436            0           1           0           0
1437            0           0           0           0
1438            0           0           1           0
1439            0           0           0           0
1440            0           0           0           0
1441            0           0           0           0
1442            0           0           0           0
1443            0           0           0           0
1444            0           0           0           0
1445            0           0           0           0
1446            0           0           0           0
1447            0           0           0           0
1448            0           1           0           0
1449            0           0           0           0
1450            0           0           0           0
1451            0           0           0           0
1452            0           0           0           0
1453            0           0           0           0
1454            0           0           0           0
1455            0           0           0           0
1456            0           0           1           0
1457            1           0           0           0
1458            0           0           0           0
1459            0           0           0           0

[1460 rows x 4 columns]
```

In [20]:
```
test_df = pd.read_csv('train.csv')
dummies = pd.get_dummies(test_df['Fence']).rename(columns=lambda x: 'Fence_' + str(x))
pd.concat([test_df, dummies], axis=1)
test_df.drop(['Fence'], inplace=True, axis=1)
```

```
In [21]: test_df = pd.read_csv('train.csv')
         dummies = pd.get_dummies(test_df['Fence']).rename(columns=lambda x: 'Fence_' + str(x))
         pd.concat([test_df, dummies], axis=1)
```

Out[21]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	
5	6	50	RL	85.0	14115	Pave	NaN	IR1	
6	7	20	RL	75.0	10084	Pave	NaN	Reg	
7	8	60	RL	NaN	10382	Pave	NaN	IR1	
8	9	50	RM	51.0	6120	Pave	NaN	Reg	
9	10	190	RL	50.0	7420	Pave	NaN	Reg	
10	11	20	RL	70.0	11200	Pave	NaN	Reg	
11	12	60	RL	85.0	11924	Pave	NaN	IR1	
12	13	20	RL	NaN	12968	Pave	NaN	IR2	
13	14	20	RL	91.0	10652	Pave	NaN	IR1	
14	15	20	RL	NaN	10920	Pave	NaN	IR1	
15	16	45	RM	51.0	6120	Pave	NaN	Reg	
16	17	20	RL	NaN	11241	Pave	NaN	IR1	
17	18	90	RL	72.0	10791	Pave	NaN	Reg	
18	19	20	RL	66.0	13695	Pave	NaN	Reg	
19	20	20	RL	70.0	7560	Pave	NaN	Reg	
20	21	60	RL	101.0	14215	Pave	NaN	IR1	
21	22	45	RM	57.0	7449	Pave	Grvl	Reg	
22	23	20	RL	75.0	9742	Pave	NaN	Reg	
23	24	120	RM	44.0	4224	Pave	NaN	Reg	
24	25	20	RL	NaN	8246	Pave	NaN	IR1	
25	26	20	RL	110.0	14230	Pave	NaN	Reg	
26	27	20	RL	60.0	7200	Pave	NaN	Reg	
27	28	20	RL	98.0	11478	Pave	NaN	Reg	
28	29	20	RL	47.0	16321	Pave	NaN	IR1	
29	30	30	RM	60.0	6324	Pave	NaN	IR1	
...	...	...	...	...	...	...	...	...	
1430	1431	60	RL	60.0	21930	Pave	NaN	IR3	
1431	1432	120	RL	NaN	4928	Pave	NaN	IR1	
1432	1433	30	RL	60.0	10800	Pave	Grvl	Reg	
1433	1434	60	RL	93.0	10261	Pave	NaN	IR1	
1434	1435	20	RL	80.0	17400	Pave	NaN	Reg	
1435	1436	20	RL	80.0	8400	Pave	NaN	Reg	
1436	1437	20	RL	60.0	9000	Pave	NaN	Reg	
1437	1438	20	RL	96.0	12444	Pave	NaN	Reg	
1438	1439	20	RM	90.0	7407	Pave	NaN	Reg	
1439	1440	60	RL	80.0	11584	Pave	NaN	Reg	
1440	1441	70	RL	79.0	11526	Pave	NaN	IR1	
1441	1442	120	RM	NaN	4426	Pave	NaN	Reg	

1442	1443	60	FV	85.0	11003	Pave	NaN	Reg
1443	1444	30	RL	NaN	8854	Pave	NaN	Reg
1444	1445	20	RL	63.0	8500	Pave	NaN	Reg
1445	1446	85	RL	70.0	8400	Pave	NaN	Reg
1446	1447	20	RL	NaN	26142	Pave	NaN	IR1
1447	1448	60	RL	80.0	10000	Pave	NaN	Reg
1448	1449	50	RL	70.0	11767	Pave	NaN	Reg
1449	1450	180	RM	21.0	1533	Pave	NaN	Reg
1450	1451	90	RL	60.0	9000	Pave	NaN	Reg
1451	1452	20	RL	78.0	9262	Pave	NaN	Reg
1452	1453	180	RM	35.0	3675	Pave	NaN	Reg
1453	1454	20	RL	90.0	17217	Pave	NaN	Reg
1454	1455	20	FV	62.0	7500	Pave	Pave	Reg
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg
1456	1457	20	RL	85.0	13175	Pave	NaN	Reg
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg

	LandContour	Utilities	...	MiscVal	MoSold	YrSold	SaleType	\
0	Lvl	AllPub	...	0	2	2008	WD	
1	Lvl	AllPub	...	0	5	2007	WD	
2	Lvl	AllPub	...	0	9	2008	WD	
3	Lvl	AllPub	...	0	2	2006	WD	
4	Lvl	AllPub	...	0	12	2008	WD	
5	Lvl	AllPub	...	700	10	2009	WD	
6	Lvl	AllPub	...	0	8	2007	WD	
7	Lvl	AllPub	...	350	11	2009	WD	
8	Lvl	AllPub	...	0	4	2008	WD	
9	Lvl	AllPub	...	0	1	2008	WD	
10	Lvl	AllPub	...	0	2	2008	WD	
11	Lvl	AllPub	...	0	7	2006	New	
12	Lvl	AllPub	...	0	9	2008	WD	
13	Lvl	AllPub	...	0	8	2007	New	
14	Lvl	AllPub	...	0	5	2008	WD	
15	Lvl	AllPub	...	0	7	2007	WD	
16	Lvl	AllPub	...	700	3	2010	WD	
17	Lvl	AllPub	...	500	10	2006	WD	
18	Lvl	AllPub	...	0	6	2008	WD	
19	Lvl	AllPub	...	0	5	2009	COD	
20	Lvl	AllPub	...	0	11	2006	New	
21	Bnk	AllPub	...	0	6	2007	WD	
22	Lvl	AllPub	...	0	9	2008	WD	
23	Lvl	AllPub	...	0	6	2007	WD	
24	Lvl	AllPub	...	0	5	2010	WD	
25	Lvl	AllPub	...	0	7	2009	WD	
26	Lvl	AllPub	...	0	5	2010	WD	
27	Lvl	AllPub	...	0	5	2010	WD	

28	Lvl	AllPub	...	0	12	2006	WD
29	Lvl	AllPub	...	0	5	2008	WD
...	...	...	...	...	...	...	...
1430	Lvl	AllPub	...	0	7	2006	WD
1431	Lvl	AllPub	...	0	10	2009	WD
1432	Lvl	AllPub	...	0	8	2007	WD
1433	Lvl	AllPub	...	0	5	2008	WD
1434	Low	AllPub	...	0	5	2006	WD
1435	Lvl	AllPub	...	0	7	2008	COD
1436	Lvl	AllPub	...	0	5	2007	WD
1437	Lvl	AllPub	...	0	11	2008	New
1438	Lvl	AllPub	...	0	4	2010	WD
1439	Lvl	AllPub	...	0	11	2007	WD
1440	Bnk	AllPub	...	0	9	2008	WD
1441	Lvl	AllPub	...	0	5	2008	WD
1442	Lvl	AllPub	...	0	4	2009	WD
1443	Lvl	AllPub	...	0	5	2009	WD
1444	Lvl	AllPub	...	0	11	2007	WD
1445	Lvl	AllPub	...	0	5	2007	WD
1446	Lvl	AllPub	...	0	4	2010	WD
1447	Lvl	AllPub	...	0	12	2007	WD
1448	Lvl	AllPub	...	0	5	2007	WD
1449	Lvl	AllPub	...	0	8	2006	WD
1450	Lvl	AllPub	...	0	9	2009	WD
1451	Lvl	AllPub	...	0	5	2009	New
1452	Lvl	AllPub	...	0	5	2006	WD
1453	Lvl	AllPub	...	0	7	2006	WD
1454	Lvl	AllPub	...	0	10	2009	WD
1455	Lvl	AllPub	...	0	8	2007	WD
1456	Lvl	AllPub	...	0	2	2010	WD
1457	Lvl	AllPub	...	2500	5	2010	WD
1458	Lvl	AllPub	...	0	4	2010	WD
1459	Lvl	AllPub	...	0	6	2008	WD

	SaleCondition	SalePrice	Fence_GdPrv	Fence_GdWo	Fence_MnPrv	Fence_MnWw
0	Normal	208500	0	0	0	0
1	Normal	181500	0	0	0	0
2	Normal	223500	0	0	0	0
3	Abnorml	140000	0	0	0	0
4	Normal	250000	0	0	0	0
5	Normal	143000	0	0	1	0
6	Normal	307000	0	0	0	0
7	Normal	200000	0	0	0	0
8	Abnorml	129900	0	0	0	0
9	Normal	118000	0	0	0	0
10	Normal	129500	0	0	0	0
11	Partial	345000	0	0	0	0
12	Normal	144000	0	0	0	0

13	Partial	279500	0	0	0	0
14	Normal	157000	0	1	0	0
15	Normal	132000	1	0	0	0
16	Normal	149000	0	0	0	0
17	Normal	90000	0	0	0	0
18	Normal	159000	0	0	0	0
19	Abnorml	139000	0	0	1	0
20	Partial	325300	0	0	0	0
21	Normal	139400	1	0	0	0
22	Normal	230000	0	0	0	0
23	Normal	129900	0	0	0	0
24	Normal	154000	0	0	1	0
25	Normal	256300	0	0	0	0
26	Normal	134800	0	0	0	0
27	Normal	306000	0	0	0	0
28	Normal	207500	0	0	0	0
29	Normal	68500	0	0	0	0
...	...	...	...	...	...	...
1430	Normal	192140	0	0	0	0
1431	Normal	143750	0	0	0	0
1432	Normal	64500	0	0	0	0
1433	Normal	186500	0	0	0	0
1434	Normal	160000	0	0	0	0
1435	Abnorml	174000	1	0	0	0
1436	Normal	120500	0	1	0	0
1437	Partial	394617	0	0	0	0
1438	Normal	149700	0	0	1	0
1439	Normal	197000	0	0	0	0
1440	Normal	191000	0	0	0	0
1441	Normal	149300	0	0	0	0
1442	Normal	310000	0	0	0	0
1443	Normal	121000	0	0	0	0
1444	Normal	179600	0	0	0	0
1445	Normal	129000	0	0	0	0
1446	Normal	157900	0	0	0	0
1447	Normal	240000	0	0	0	0
1448	Normal	112000	0	1	0	0
1449	Abnorml	92000	0	0	0	0
1450	Normal	136000	0	0	0	0
1451	Partial	287090	0	0	0	0
1452	Normal	145000	0	0	0	0
1453	Abnorml	84500	0	0	0	0
1454	Normal	185000	0	0	0	0
1455	Normal	175000	0	0	0	0
1456	Normal	210000	0	0	1	0
1457	Normal	266500	1	0	0	0
1458	Normal	142125	0	0	0	0
1459	Normal	147500	0	0	0	0

```
[1460 rows x 85 columns]
```

```
In [22]: test_df = pd.read_csv('train.csv')
         dummies = pd.get_dummies(test_df['Fence']).rename(columns=lambda x: 'Fence_' + str(x))
         pd.concat([test_df, dummies], axis=1)
         test_df.drop(['Fence'],  axis=1)
```

Out[22]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	
5	6	50	RL	85.0	14115	Pave	NaN	IR1	
6	7	20	RL	75.0	10084	Pave	NaN	Reg	
7	8	60	RL	NaN	10382	Pave	NaN	IR1	
8	9	50	RM	51.0	6120	Pave	NaN	Reg	
9	10	190	RL	50.0	7420	Pave	NaN	Reg	
10	11	20	RL	70.0	11200	Pave	NaN	Reg	
11	12	60	RL	85.0	11924	Pave	NaN	IR1	
12	13	20	RL	NaN	12968	Pave	NaN	IR2	
13	14	20	RL	91.0	10652	Pave	NaN	IR1	
14	15	20	RL	NaN	10920	Pave	NaN	IR1	
15	16	45	RM	51.0	6120	Pave	NaN	Reg	
16	17	20	RL	NaN	11241	Pave	NaN	IR1	
17	18	90	RL	72.0	10791	Pave	NaN	Reg	
18	19	20	RL	66.0	13695	Pave	NaN	Reg	
19	20	20	RL	70.0	7560	Pave	NaN	Reg	
20	21	60	RL	101.0	14215	Pave	NaN	IR1	
21	22	45	RM	57.0	7449	Pave	Grvl	Reg	
22	23	20	RL	75.0	9742	Pave	NaN	Reg	
23	24	120	RM	44.0	4224	Pave	NaN	Reg	
24	25	20	RL	NaN	8246	Pave	NaN	IR1	
25	26	20	RL	110.0	14230	Pave	NaN	Reg	
26	27	20	RL	60.0	7200	Pave	NaN	Reg	
27	28	20	RL	98.0	11478	Pave	NaN	Reg	
28	29	20	RL	47.0	16321	Pave	NaN	IR1	
29	30	30	RM	60.0	6324	Pave	NaN	IR1	
...	...	...	...	...	...	...	...	...	
1430	1431	60	RL	60.0	21930	Pave	NaN	IR3	
1431	1432	120	RL	NaN	4928	Pave	NaN	IR1	
1432	1433	30	RL	60.0	10800	Pave	Grvl	Reg	
1433	1434	60	RL	93.0	10261	Pave	NaN	IR1	
1434	1435	20	RL	80.0	17400	Pave	NaN	Reg	
1435	1436	20	RL	80.0	8400	Pave	NaN	Reg	
1436	1437	20	RL	60.0	9000	Pave	NaN	Reg	
1437	1438	20	RL	96.0	12444	Pave	NaN	Reg	

1438	1439	20	RM	90.0	7407	Pave	NaN	Reg
1439	1440	60	RL	80.0	11584	Pave	NaN	Reg
1440	1441	70	RL	79.0	11526	Pave	NaN	IR1
1441	1442	120	RM	NaN	4426	Pave	NaN	Reg
1442	1443	60	FV	85.0	11003	Pave	NaN	Reg
1443	1444	30	RL	NaN	8854	Pave	NaN	Reg
1444	1445	20	RL	63.0	8500	Pave	NaN	Reg
1445	1446	85	RL	70.0	8400	Pave	NaN	Reg
1446	1447	20	RL	NaN	26142	Pave	NaN	IR1
1447	1448	60	RL	80.0	10000	Pave	NaN	Reg
1448	1449	50	RL	70.0	11767	Pave	NaN	Reg
1449	1450	180	RM	21.0	1533	Pave	NaN	Reg
1450	1451	90	RL	60.0	9000	Pave	NaN	Reg
1451	1452	20	RL	78.0	9262	Pave	NaN	Reg
1452	1453	180	RM	35.0	3675	Pave	NaN	Reg
1453	1454	20	RL	90.0	17217	Pave	NaN	Reg
1454	1455	20	FV	62.0	7500	Pave	Pave	Reg
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg
1456	1457	20	RL	85.0	13175	Pave	NaN	Reg
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg

	LandContour	Utilities	...	ScreenPorch	PoolArea	PoolQC	MiscFeature	\
0	Lvl	AllPub	...	0	0	NaN	NaN	
1	Lvl	AllPub	...	0	0	NaN	NaN	
2	Lvl	AllPub	...	0	0	NaN	NaN	
3	Lvl	AllPub	...	0	0	NaN	NaN	
4	Lvl	AllPub	...	0	0	NaN	NaN	
5	Lvl	AllPub	...	0	0	NaN	Shed	
6	Lvl	AllPub	...	0	0	NaN	NaN	
7	Lvl	AllPub	...	0	0	NaN	Shed	
8	Lvl	AllPub	...	0	0	NaN	NaN	
9	Lvl	AllPub	...	0	0	NaN	NaN	
10	Lvl	AllPub	...	0	0	NaN	NaN	
11	Lvl	AllPub	...	0	0	NaN	NaN	
12	Lvl	AllPub	...	176	0	NaN	NaN	
13	Lvl	AllPub	...	0	0	NaN	NaN	
14	Lvl	AllPub	...	0	0	NaN	NaN	
15	Lvl	AllPub	...	0	0	NaN	NaN	
16	Lvl	AllPub	...	0	0	NaN	Shed	
17	Lvl	AllPub	...	0	0	NaN	Shed	
18	Lvl	AllPub	...	0	0	NaN	NaN	
19	Lvl	AllPub	...	0	0	NaN	NaN	
20	Lvl	AllPub	...	0	0	NaN	NaN	
21	Bnk	AllPub	...	0	0	NaN	NaN	
22	Lvl	AllPub	...	0	0	NaN	NaN	
23	Lvl	AllPub	...	0	0	NaN	NaN	

24	Lvl	AllPub	...	0	0	NaN	NaN
25	Lvl	AllPub	...	0	0	NaN	NaN
26	Lvl	AllPub	...	0	0	NaN	NaN
27	Lvl	AllPub	...	0	0	NaN	NaN
28	Lvl	AllPub	...	0	0	NaN	NaN
29	Lvl	AllPub	...	0	0	NaN	NaN
...	...	...	...	...	...	...	...
1430	Lvl	AllPub	...	0	0	NaN	NaN
1431	Lvl	AllPub	...	0	0	NaN	NaN
1432	Lvl	AllPub	...	0	0	NaN	NaN
1433	Lvl	AllPub	...	0	0	NaN	NaN
1434	Low	AllPub	...	0	0	NaN	NaN
1435	Lvl	AllPub	...	0	0	NaN	NaN
1436	Lvl	AllPub	...	0	0	NaN	NaN
1437	Lvl	AllPub	...	0	0	NaN	NaN
1438	Lvl	AllPub	...	0	0	NaN	NaN
1439	Lvl	AllPub	...	0	0	NaN	NaN
1440	Bnk	AllPub	...	0	0	NaN	NaN
1441	Lvl	AllPub	...	0	0	NaN	NaN
1442	Lvl	AllPub	...	0	0	NaN	NaN
1443	Lvl	AllPub	...	40	0	NaN	NaN
1444	Lvl	AllPub	...	0	0	NaN	NaN
1445	Lvl	AllPub	...	0	0	NaN	NaN
1446	Lvl	AllPub	...	0	0	NaN	NaN
1447	Lvl	AllPub	...	0	0	NaN	NaN
1448	Lvl	AllPub	...	0	0	NaN	NaN
1449	Lvl	AllPub	...	0	0	NaN	NaN
1450	Lvl	AllPub	...	0	0	NaN	NaN
1451	Lvl	AllPub	...	0	0	NaN	NaN
1452	Lvl	AllPub	...	0	0	NaN	NaN
1453	Lvl	AllPub	...	0	0	NaN	NaN
1454	Lvl	AllPub	...	0	0	NaN	NaN
1455	Lvl	AllPub	...	0	0	NaN	NaN
1456	Lvl	AllPub	...	0	0	NaN	NaN
1457	Lvl	AllPub	...	0	0	NaN	Shed
1458	Lvl	AllPub	...	0	0	NaN	NaN
1459	Lvl	AllPub	...	0	0	NaN	NaN

	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	0	2	2008	WD	Normal	208500
1	0	5	2007	WD	Normal	181500
2	0	9	2008	WD	Normal	223500
3	0	2	2006	WD	Abnorml	140000
4	0	12	2008	WD	Normal	250000
5	700	10	2009	WD	Normal	143000
6	0	8	2007	WD	Normal	307000
7	350	11	2009	WD	Normal	200000
8	0	4	2008	WD	Abnorml	129900

9	0	1	2008	WD	Normal	118000
10	0	2	2008	WD	Normal	129500
11	0	7	2006	New	Partial	345000
12	0	9	2008	WD	Normal	144000
13	0	8	2007	New	Partial	279500
14	0	5	2008	WD	Normal	157000
15	0	7	2007	WD	Normal	132000
16	700	3	2010	WD	Normal	149000
17	500	10	2006	WD	Normal	90000
18	0	6	2008	WD	Normal	159000
19	0	5	2009	COD	Abnorml	139000
20	0	11	2006	New	Partial	325300
21	0	6	2007	WD	Normal	139400
22	0	9	2008	WD	Normal	230000
23	0	6	2007	WD	Normal	129900
24	0	5	2010	WD	Normal	154000
25	0	7	2009	WD	Normal	256300
26	0	5	2010	WD	Normal	134800
27	0	5	2010	WD	Normal	306000
28	0	12	2006	WD	Normal	207500
29	0	5	2008	WD	Normal	68500
...	...	...	...	...	...	...
1430	0	7	2006	WD	Normal	192140
1431	0	10	2009	WD	Normal	143750
1432	0	8	2007	WD	Normal	64500
1433	0	5	2008	WD	Normal	186500
1434	0	5	2006	WD	Normal	160000
1435	0	7	2008	COD	Abnorml	174000
1436	0	5	2007	WD	Normal	120500
1437	0	11	2008	New	Partial	394617
1438	0	4	2010	WD	Normal	149700
1439	0	11	2007	WD	Normal	197000
1440	0	9	2008	WD	Normal	191000
1441	0	5	2008	WD	Normal	149300
1442	0	4	2009	WD	Normal	310000
1443	0	5	2009	WD	Normal	121000
1444	0	11	2007	WD	Normal	179600
1445	0	5	2007	WD	Normal	129000
1446	0	4	2010	WD	Normal	157900
1447	0	12	2007	WD	Normal	240000
1448	0	5	2007	WD	Normal	112000
1449	0	8	2006	WD	Abnorml	92000
1450	0	9	2009	WD	Normal	136000
1451	0	5	2009	New	Partial	287090
1452	0	5	2006	WD	Normal	145000
1453	0	7	2006	WD	Abnorml	84500
1454	0	10	2009	WD	Normal	185000
1455	0	8	2007	WD	Normal	175000

```
1456       0      2   2010        WD         Normal     210000
1457    2500      5   2010        WD         Normal     266500
1458       0      4   2010        WD         Normal     142125
1459       0      6   2008        WD         Normal     147500

[1460 rows x 80 columns]
```

```
In [23]: test_df = pd.read_csv('train.csv')
         dummies = pd.get_dummies(test_df['Fence']).rename(columns=lambda x: 'Fence_' + str(x))
         test_df = pd.concat([test_df, dummies], axis=1)
         test_df =test_df.drop(['Fence'],  axis=1)
```

```
In [24]: test_df
```

Out[24]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	
5	6	50	RL	85.0	14115	Pave	NaN	IR1	
6	7	20	RL	75.0	10084	Pave	NaN	Reg	
7	8	60	RL	NaN	10382	Pave	NaN	IR1	
8	9	50	RM	51.0	6120	Pave	NaN	Reg	
9	10	190	RL	50.0	7420	Pave	NaN	Reg	
10	11	20	RL	70.0	11200	Pave	NaN	Reg	
11	12	60	RL	85.0	11924	Pave	NaN	IR1	
12	13	20	RL	NaN	12968	Pave	NaN	IR2	
13	14	20	RL	91.0	10652	Pave	NaN	IR1	
14	15	20	RL	NaN	10920	Pave	NaN	IR1	
15	16	45	RM	51.0	6120	Pave	NaN	Reg	
16	17	20	RL	NaN	11241	Pave	NaN	IR1	
17	18	90	RL	72.0	10791	Pave	NaN	Reg	
18	19	20	RL	66.0	13695	Pave	NaN	Reg	
19	20	20	RL	70.0	7560	Pave	NaN	Reg	
20	21	60	RL	101.0	14215	Pave	NaN	IR1	
21	22	45	RM	57.0	7449	Pave	Grvl	Reg	
22	23	20	RL	75.0	9742	Pave	NaN	Reg	
23	24	120	RM	44.0	4224	Pave	NaN	Reg	
24	25	20	RL	NaN	8246	Pave	NaN	IR1	
25	26	20	RL	110.0	14230	Pave	NaN	Reg	
26	27	20	RL	60.0	7200	Pave	NaN	Reg	
27	28	20	RL	98.0	11478	Pave	NaN	Reg	
28	29	20	RL	47.0	16321	Pave	NaN	IR1	
29	30	30	RM	60.0	6324	Pave	NaN	IR1	
...	...	...	...	...	...	...	...	...	
1430	1431	60	RL	60.0	21930	Pave	NaN	IR3	
1431	1432	120	RL	NaN	4928	Pave	NaN	IR1	

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

	LandContour	Utilities	...	MiscVal	MoSold	YrSold	SaleType	\
0	Lvl	AllPub	...	0	2	2008	WD	
1	Lvl	AllPub	...	0	5	2007	WD	
2	Lvl	AllPub	...	0	9	2008	WD	
3	Lvl	AllPub	...	0	2	2006	WD	
4	Lvl	AllPub	...	0	12	2008	WD	
5	Lvl	AllPub	...	700	10	2009	WD	
6	Lvl	AllPub	...	0	8	2007	WD	
7	Lvl	AllPub	...	350	11	2009	WD	
8	Lvl	AllPub	...	0	4	2008	WD	
9	Lvl	AllPub	...	0	1	2008	WD	
10	Lvl	AllPub	...	0	2	2008	WD	
11	Lvl	AllPub	...	0	7	2006	New	
12	Lvl	AllPub	...	0	9	2008	WD	
13	Lvl	AllPub	...	0	8	2007	New	
14	Lvl	AllPub	...	0	5	2008	WD	
15	Lvl	AllPub	...	0	7	2007	WD	
16	Lvl	AllPub	...	700	3	2010	WD	
17	Lvl	AllPub	...	500	10	2006	WD	

18	Lvl	AllPub	...	0	6	2008	WD
19	Lvl	AllPub	...	0	5	2009	COD
20	Lvl	AllPub	...	0	11	2006	New
21	Bnk	AllPub	...	0	6	2007	WD
22	Lvl	AllPub	...	0	9	2008	WD
23	Lvl	AllPub	...	0	6	2007	WD
24	Lvl	AllPub	...	0	5	2010	WD
25	Lvl	AllPub	...	0	7	2009	WD
26	Lvl	AllPub	...	0	5	2010	WD
27	Lvl	AllPub	...	0	5	2010	WD
28	Lvl	AllPub	...	0	12	2006	WD
29	Lvl	AllPub	...	0	5	2008	WD
...	...	...	...	...	...	...	...
1430	Lvl	AllPub	...	0	7	2006	WD
1431	Lvl	AllPub	...	0	10	2009	WD
1432	Lvl	AllPub	...	0	8	2007	WD
1433	Lvl	AllPub	...	0	5	2008	WD
1434	Low	AllPub	...	0	5	2006	WD
1435	Lvl	AllPub	...	0	7	2008	COD
1436	Lvl	AllPub	...	0	5	2007	WD
1437	Lvl	AllPub	...	0	11	2008	New
1438	Lvl	AllPub	...	0	4	2010	WD
1439	Lvl	AllPub	...	0	11	2007	WD
1440	Bnk	AllPub	...	0	9	2008	WD
1441	Lvl	AllPub	...	0	5	2008	WD
1442	Lvl	AllPub	...	0	4	2009	WD
1443	Lvl	AllPub	...	0	5	2009	WD
1444	Lvl	AllPub	...	0	11	2007	WD
1445	Lvl	AllPub	...	0	5	2007	WD
1446	Lvl	AllPub	...	0	4	2010	WD
1447	Lvl	AllPub	...	0	12	2007	WD
1448	Lvl	AllPub	...	0	5	2007	WD
1449	Lvl	AllPub	...	0	8	2006	WD
1450	Lvl	AllPub	...	0	9	2009	WD
1451	Lvl	AllPub	...	0	5	2009	New
1452	Lvl	AllPub	...	0	5	2006	WD
1453	Lvl	AllPub	...	0	7	2006	WD
1454	Lvl	AllPub	...	0	10	2009	WD
1455	Lvl	AllPub	...	0	8	2007	WD
1456	Lvl	AllPub	...	0	2	2010	WD
1457	Lvl	AllPub	...	2500	5	2010	WD
1458	Lvl	AllPub	...	0	4	2010	WD
1459	Lvl	AllPub	...	0	6	2008	WD

	SaleCondition	SalePrice	Fence_GdPrv	Fence_GdWo	Fence_MnPrv	Fence_MnWw
0	Normal	208500	0	0	0	0
1	Normal	181500	0	0	0	0
2	Normal	223500	0	0	0	0

3	Abnorml	140000	0	0	0	0
4	Normal	250000	0	0	0	0
5	Normal	143000	0	0	1	0
6	Normal	307000	0	0	0	0
7	Normal	200000	0	0	0	0
8	Abnorml	129900	0	0	0	0
9	Normal	118000	0	0	0	0
10	Normal	129500	0	0	0	0
11	Partial	345000	0	0	0	0
12	Normal	144000	0	0	0	0
13	Partial	279500	0	0	0	0
14	Normal	157000	0	1	0	0
15	Normal	132000	1	0	0	0
16	Normal	149000	0	0	0	0
17	Normal	90000	0	0	0	0
18	Normal	159000	0	0	0	0
19	Abnorml	139000	0	0	1	0
20	Partial	325300	0	0	0	0
21	Normal	139400	1	0	0	0
22	Normal	230000	0	0	0	0
23	Normal	129900	0	0	0	0
24	Normal	154000	0	0	1	0
25	Normal	256300	0	0	0	0
26	Normal	134800	0	0	0	0
27	Normal	306000	0	0	0	0
28	Normal	207500	0	0	0	0
29	Normal	68500	0	0	0	0
...	...	...	...	...	...	...
1430	Normal	192140	0	0	0	0
1431	Normal	143750	0	0	0	0
1432	Normal	64500	0	0	0	0
1433	Normal	186500	0	0	0	0
1434	Normal	160000	0	0	0	0
1435	Abnorml	174000	1	0	0	0
1436	Normal	120500	0	1	0	0
1437	Partial	394617	0	0	0	0
1438	Normal	149700	0	0	1	0
1439	Normal	197000	0	0	0	0
1440	Normal	191000	0	0	0	0
1441	Normal	149300	0	0	0	0
1442	Normal	310000	0	0	0	0
1443	Normal	121000	0	0	0	0
1444	Normal	179600	0	0	0	0
1445	Normal	129000	0	0	0	0
1446	Normal	157900	0	0	0	0
1447	Normal	240000	0	0	0	0
1448	Normal	112000	0	1	0	0
1449	Abnorml	92000	0	0	0	0

```
1450        Normal    136000          0          0          0          0
1451       Partial    287090          0          0          0          0
1452        Normal    145000          0          0          0          0
1453       Abnorml     84500          0          0          0          0
1454        Normal    185000          0          0          0          0
1455        Normal    175000          0          0          0          0
1456        Normal    210000          0          0          1          0
1457        Normal    266500          1          0          0          0
1458        Normal    142125          0          0          0          0
1459        Normal    147500          0          0          0          0

[1460 rows x 84 columns]
```

```
In [25]: df =test_df['MSZoning']
```

```
In [26]: df
```

```
Out[26]: 0       RL
         1       RL
         2       RL
         3       RL
         4       RL
         5       RL
         6       RL
         7       RL
         8       RM
         9       RL
         10      RL
         11      RL
         12      RL
         13      RL
         14      RL
         15      RM
         16      RL
         17      RL
         18      RL
         19      RL
         20      RL
         21      RM
         22      RL
         23      RM
         24      RL
         25      RL
         26      RL
         27      RL
         28      RL
         29      RM
                 ..
```

```
1430    RL
1431    RL
1432    RL
1433    RL
1434    RL
1435    RL
1436    RL
1437    RL
1438    RM
1439    RL
1440    RL
1441    RM
1442    FV
1443    RL
1444    RL
1445    RL
1446    RL
1447    RL
1448    RL
1449    RM
1450    RL
1451    RL
1452    RM
1453    RL
1454    FV
1455    RL
1456    RL
1457    RL
1458    RL
1459    RL
Name: MSZoning, dtype: object
```

```
In [29]: df.name
```

```
Out[29]: 'MSZoning'
```

```
In [31]: df.unique()
```

```
Out[31]: array(['RL', 'RM', 'C (all)', 'FV', 'RH'], dtype=object)
```

```
In [34]: df.dtypes
```

```
Out[34]: dtype('O')
```

```
In [1]: import pandas as pd
        test_df = pd.read_csv('train.csv')
        test_df.head()
```

```
Out[1]:    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
        0   1          60       RL         65.0     8450   Pave   NaN      Reg
```

```
1   2          20       RL         80.0     9600   Pave   NaN      Reg
2   3          60       RL         68.0    11250   Pave   NaN      IR1
3   4          70       RL         60.0     9550   Pave   NaN      IR1
4   5          60       RL         84.0    14260   Pave   NaN      IR1

  LandContour Utilities    ...     PoolArea PoolQC Fence MiscFeature MiscVal  \
0         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
1         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
2         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
3         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
4         Lvl    AllPub    ...            0    NaN   NaN         NaN       0

  MoSold YrSold  SaleType  SaleCondition  SalePrice
0      2   2008        WD         Normal     208500
1      5   2007        WD         Normal     181500
2      9   2008        WD         Normal     223500
3      2   2006        WD        Abnorml     140000
4     12   2008        WD         Normal     250000

[5 rows x 81 columns]
```

```
In [2]: import numpy as np
        from scipy.stats import norm
        import matplotlib.pyplot as plt
```

```
In [3]: test_df.head()
```

```
Out[3]:    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
0   1          60       RL         65.0     8450   Pave   NaN      Reg
1   2          20       RL         80.0     9600   Pave   NaN      Reg
2   3          60       RL         68.0    11250   Pave   NaN      IR1
3   4          70       RL         60.0     9550   Pave   NaN      IR1
4   5          60       RL         84.0    14260   Pave   NaN      IR1

  LandContour Utilities    ...     PoolArea PoolQC Fence MiscFeature MiscVal  \
0         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
1         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
2         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
3         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
4         Lvl    AllPub    ...            0    NaN   NaN         NaN       0

  MoSold YrSold  SaleType  SaleCondition  SalePrice
0      2   2008        WD         Normal     208500
1      5   2007        WD         Normal     181500
2      9   2008        WD         Normal     223500
3      2   2006        WD        Abnorml     140000
4     12   2008        WD         Normal     250000

[5 rows x 81 columns]
```

```
In [4]: test_val = test_df.values

In [5]: np.shape(test_val)

Out[5]: (1460, 81)

In [7]: test_df.convert_objects(convert_numeric=True)

/home/bhumihar/anaconda3/lib/python3.5/site-packages/ipykernel/__main__.py:1: FutureWarning: con
  if __name__ == '__main__':
```

Out[7]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	\
0	1	60	RL	65.0	8450	Pave	NaN	Reg	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	
5	6	50	RL	85.0	14115	Pave	NaN	IR1	
6	7	20	RL	75.0	10084	Pave	NaN	Reg	
7	8	60	RL	NaN	10382	Pave	NaN	IR1	
8	9	50	RM	51.0	6120	Pave	NaN	Reg	
9	10	190	RL	50.0	7420	Pave	NaN	Reg	
10	11	20	RL	70.0	11200	Pave	NaN	Reg	
11	12	60	RL	85.0	11924	Pave	NaN	IR1	
12	13	20	RL	NaN	12968	Pave	NaN	IR2	
13	14	20	RL	91.0	10652	Pave	NaN	IR1	
14	15	20	RL	NaN	10920	Pave	NaN	IR1	
15	16	45	RM	51.0	6120	Pave	NaN	Reg	
16	17	20	RL	NaN	11241	Pave	NaN	IR1	
17	18	90	RL	72.0	10791	Pave	NaN	Reg	
18	19	20	RL	66.0	13695	Pave	NaN	Reg	
19	20	20	RL	70.0	7560	Pave	NaN	Reg	
20	21	60	RL	101.0	14215	Pave	NaN	IR1	
21	22	45	RM	57.0	7449	Pave	Grvl	Reg	
22	23	20	RL	75.0	9742	Pave	NaN	Reg	
23	24	120	RM	44.0	4224	Pave	NaN	Reg	
24	25	20	RL	NaN	8246	Pave	NaN	IR1	
25	26	20	RL	110.0	14230	Pave	NaN	Reg	
26	27	20	RL	60.0	7200	Pave	NaN	Reg	
27	28	20	RL	98.0	11478	Pave	NaN	Reg	
28	29	20	RL	47.0	16321	Pave	NaN	IR1	
29	30	30	RM	60.0	6324	Pave	NaN	IR1	
...	...	...	...	...	...	...	...	...	
1430	1431	60	RL	60.0	21930	Pave	NaN	IR3	
1431	1432	120	RL	NaN	4928	Pave	NaN	IR1	
1432	1433	30	RL	60.0	10800	Pave	Grvl	Reg	
1433	1434	60	RL	93.0	10261	Pave	NaN	IR1	
1434	1435	20	RL	80.0	17400	Pave	NaN	Reg	

1435	1436	20	RL	80.0	8400	Pave	NaN	Reg
1436	1437	20	RL	60.0	9000	Pave	NaN	Reg
1437	1438	20	RL	96.0	12444	Pave	NaN	Reg
1438	1439	20	RM	90.0	7407	Pave	NaN	Reg
1439	1440	60	RL	80.0	11584	Pave	NaN	Reg
1440	1441	70	RL	79.0	11526	Pave	NaN	IR1
1441	1442	120	RM	NaN	4426	Pave	NaN	Reg
1442	1443	60	FV	85.0	11003	Pave	NaN	Reg
1443	1444	30	RL	NaN	8854	Pave	NaN	Reg
1444	1445	20	RL	63.0	8500	Pave	NaN	Reg
1445	1446	85	RL	70.0	8400	Pave	NaN	Reg
1446	1447	20	RL	NaN	26142	Pave	NaN	IR1
1447	1448	60	RL	80.0	10000	Pave	NaN	Reg
1448	1449	50	RL	70.0	11767	Pave	NaN	Reg
1449	1450	180	RM	21.0	1533	Pave	NaN	Reg
1450	1451	90	RL	60.0	9000	Pave	NaN	Reg
1451	1452	20	RL	78.0	9262	Pave	NaN	Reg
1452	1453	180	RM	35.0	3675	Pave	NaN	Reg
1453	1454	20	RL	90.0	17217	Pave	NaN	Reg
1454	1455	20	FV	62.0	7500	Pave	Pave	Reg
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg
1456	1457	20	RL	85.0	13175	Pave	NaN	Reg
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg

	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	\
0	Lvl	AllPub	...	0	NaN	NaN	NaN	
1	Lvl	AllPub	...	0	NaN	NaN	NaN	
2	Lvl	AllPub	...	0	NaN	NaN	NaN	
3	Lvl	AllPub	...	0	NaN	NaN	NaN	
4	Lvl	AllPub	...	0	NaN	NaN	NaN	
5	Lvl	AllPub	...	0	NaN	MnPrv	Shed	
6	Lvl	AllPub	...	0	NaN	NaN	NaN	
7	Lvl	AllPub	...	0	NaN	NaN	Shed	
8	Lvl	AllPub	...	0	NaN	NaN	NaN	
9	Lvl	AllPub	...	0	NaN	NaN	NaN	
10	Lvl	AllPub	...	0	NaN	NaN	NaN	
11	Lvl	AllPub	...	0	NaN	NaN	NaN	
12	Lvl	AllPub	...	0	NaN	NaN	NaN	
13	Lvl	AllPub	...	0	NaN	NaN	NaN	
14	Lvl	AllPub	...	0	NaN	GdWo	NaN	
15	Lvl	AllPub	...	0	NaN	GdPrv	NaN	
16	Lvl	AllPub	...	0	NaN	NaN	Shed	
17	Lvl	AllPub	...	0	NaN	NaN	Shed	
18	Lvl	AllPub	...	0	NaN	NaN	NaN	
19	Lvl	AllPub	...	0	NaN	MnPrv	NaN	
20	Lvl	AllPub	...	0	NaN	NaN	NaN	

21	Bnk	AllPub	...	0	NaN	GdPrv	NaN
22	Lvl	AllPub	...	0	NaN	NaN	NaN
23	Lvl	AllPub	...	0	NaN	NaN	NaN
24	Lvl	AllPub	...	0	NaN	MnPrv	NaN
25	Lvl	AllPub	...	0	NaN	NaN	NaN
26	Lvl	AllPub	...	0	NaN	NaN	NaN
27	Lvl	AllPub	...	0	NaN	NaN	NaN
28	Lvl	AllPub	...	0	NaN	NaN	NaN
29	Lvl	AllPub	...	0	NaN	NaN	NaN
...	...	...	...	...	...	...	...
1430	Lvl	AllPub	...	0	NaN	NaN	NaN
1431	Lvl	AllPub	...	0	NaN	NaN	NaN
1432	Lvl	AllPub	...	0	NaN	NaN	NaN
1433	Lvl	AllPub	...	0	NaN	NaN	NaN
1434	Low	AllPub	...	0	NaN	NaN	NaN
1435	Lvl	AllPub	...	0	NaN	GdPrv	NaN
1436	Lvl	AllPub	...	0	NaN	GdWo	NaN
1437	Lvl	AllPub	...	0	NaN	NaN	NaN
1438	Lvl	AllPub	...	0	NaN	MnPrv	NaN
1439	Lvl	AllPub	...	0	NaN	NaN	NaN
1440	Bnk	AllPub	...	0	NaN	NaN	NaN
1441	Lvl	AllPub	...	0	NaN	NaN	NaN
1442	Lvl	AllPub	...	0	NaN	NaN	NaN
1443	Lvl	AllPub	...	0	NaN	NaN	NaN
1444	Lvl	AllPub	...	0	NaN	NaN	NaN
1445	Lvl	AllPub	...	0	NaN	NaN	NaN
1446	Lvl	AllPub	...	0	NaN	NaN	NaN
1447	Lvl	AllPub	...	0	NaN	NaN	NaN
1448	Lvl	AllPub	...	0	NaN	GdWo	NaN
1449	Lvl	AllPub	...	0	NaN	NaN	NaN
1450	Lvl	AllPub	...	0	NaN	NaN	NaN
1451	Lvl	AllPub	...	0	NaN	NaN	NaN
1452	Lvl	AllPub	...	0	NaN	NaN	NaN
1453	Lvl	AllPub	...	0	NaN	NaN	NaN
1454	Lvl	AllPub	...	0	NaN	NaN	NaN
1455	Lvl	AllPub	...	0	NaN	NaN	NaN
1456	Lvl	AllPub	...	0	NaN	MnPrv	NaN
1457	Lvl	AllPub	...	0	NaN	GdPrv	Shed
1458	Lvl	AllPub	...	0	NaN	NaN	NaN
1459	Lvl	AllPub	...	0	NaN	NaN	NaN

	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	0	2	2008	WD	Normal	208500
1	0	5	2007	WD	Normal	181500
2	0	9	2008	WD	Normal	223500
3	0	2	2006	WD	Abnorml	140000
4	0	12	2008	WD	Normal	250000
5	700	10	2009	WD	Normal	143000

6	0	8	2007	WD	Normal	307000
7	350	11	2009	WD	Normal	200000
8	0	4	2008	WD	Abnorml	129900
9	0	1	2008	WD	Normal	118000
10	0	2	2008	WD	Normal	129500
11	0	7	2006	New	Partial	345000
12	0	9	2008	WD	Normal	144000
13	0	8	2007	New	Partial	279500
14	0	5	2008	WD	Normal	157000
15	0	7	2007	WD	Normal	132000
16	700	3	2010	WD	Normal	149000
17	500	10	2006	WD	Normal	90000
18	0	6	2008	WD	Normal	159000
19	0	5	2009	COD	Abnorml	139000
20	0	11	2006	New	Partial	325300
21	0	6	2007	WD	Normal	139400
22	0	9	2008	WD	Normal	230000
23	0	6	2007	WD	Normal	129900
24	0	5	2010	WD	Normal	154000
25	0	7	2009	WD	Normal	256300
26	0	5	2010	WD	Normal	134800
27	0	5	2010	WD	Normal	306000
28	0	12	2006	WD	Normal	207500
29	0	5	2008	WD	Normal	68500
...	...	...	...	...	...	...
1430	0	7	2006	WD	Normal	192140
1431	0	10	2009	WD	Normal	143750
1432	0	8	2007	WD	Normal	64500
1433	0	5	2008	WD	Normal	186500
1434	0	5	2006	WD	Normal	160000
1435	0	7	2008	COD	Abnorml	174000
1436	0	5	2007	WD	Normal	120500
1437	0	11	2008	New	Partial	394617
1438	0	4	2010	WD	Normal	149700
1439	0	11	2007	WD	Normal	197000
1440	0	9	2008	WD	Normal	191000
1441	0	5	2008	WD	Normal	149300
1442	0	4	2009	WD	Normal	310000
1443	0	5	2009	WD	Normal	121000
1444	0	11	2007	WD	Normal	179600
1445	0	5	2007	WD	Normal	129000
1446	0	4	2010	WD	Normal	157900
1447	0	12	2007	WD	Normal	240000
1448	0	5	2007	WD	Normal	112000
1449	0	8	2006	WD	Abnorml	92000
1450	0	9	2009	WD	Normal	136000
1451	0	5	2009	New	Partial	287090
1452	0	5	2006	WD	Normal	145000

```
1453       0      7   2006        WD         Abnorml      84500
1454       0     10   2009        WD          Normal     185000
1455       0      8   2007        WD          Normal     175000
1456       0      2   2010        WD          Normal     210000
1457    2500      5   2010        WD          Normal     266500
1458       0      4   2010        WD          Normal     142125
1459       0      6   2008        WD          Normal     147500

[1460 rows x 81 columns]
```

```
In [8]: train_df = pd.read_csv('train.csv')
```

```
In [9]: train_df.head()
```

```
Out[9]:    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
        0   1          60       RL         65.0     8450   Pave   NaN      Reg
        1   2          20       RL         80.0     9600   Pave   NaN      Reg
        2   3          60       RL         68.0    11250   Pave   NaN      IR1
        3   4          70       RL         60.0     9550   Pave   NaN      IR1
        4   5          60       RL         84.0    14260   Pave   NaN      IR1

          LandContour Utilities    ...     PoolArea PoolQC Fence MiscFeature MiscVal  \
        0         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
        1         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
        2         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
        3         Lvl    AllPub    ...            0    NaN   NaN         NaN       0
        4         Lvl    AllPub    ...            0    NaN   NaN         NaN       0

          MoSold YrSold  SaleType  SaleCondition  SalePrice
        0      2   2008        WD         Normal     208500
        1      5   2007        WD         Normal     181500
        2      9   2008        WD         Normal     223500
        3      2   2006        WD        Abnorml     140000
        4     12   2008        WD         Normal     250000

        [5 rows x 81 columns]
```

```
In [10]: train_df.convert_objects(convert_numeric=True)
```

```
/home/bhumihar/anaconda3/lib/python3.5/site-packages/ipykernel/__main__.py:1: FutureWarning: con
  if __name__ == '__main__':
```

```
Out[10]:         Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
         0        1          60       RL         65.0     8450   Pave   NaN      Reg
         1        2          20       RL         80.0     9600   Pave   NaN      Reg
         2        3          60       RL         68.0    11250   Pave   NaN      IR1
         3        4          70       RL         60.0     9550   Pave   NaN      IR1
         4        5          60       RL         84.0    14260   Pave   NaN      IR1
```

5	6	50	RL	85.0	14115	Pave	NaN	IR1
6	7	20	RL	75.0	10084	Pave	NaN	Reg
7	8	60	RL	NaN	10382	Pave	NaN	IR1
8	9	50	RM	51.0	6120	Pave	NaN	Reg
9	10	190	RL	50.0	7420	Pave	NaN	Reg
10	11	20	RL	70.0	11200	Pave	NaN	Reg
11	12	60	RL	85.0	11924	Pave	NaN	IR1
12	13	20	RL	NaN	12968	Pave	NaN	IR2
13	14	20	RL	91.0	10652	Pave	NaN	IR1
14	15	20	RL	NaN	10920	Pave	NaN	IR1
15	16	45	RM	51.0	6120	Pave	NaN	Reg
16	17	20	RL	NaN	11241	Pave	NaN	IR1
17	18	90	RL	72.0	10791	Pave	NaN	Reg
18	19	20	RL	66.0	13695	Pave	NaN	Reg
19	20	20	RL	70.0	7560	Pave	NaN	Reg
20	21	60	RL	101.0	14215	Pave	NaN	IR1
21	22	45	RM	57.0	7449	Pave	Grvl	Reg
22	23	20	RL	75.0	9742	Pave	NaN	Reg
23	24	120	RM	44.0	4224	Pave	NaN	Reg
24	25	20	RL	NaN	8246	Pave	NaN	IR1
25	26	20	RL	110.0	14230	Pave	NaN	Reg
26	27	20	RL	60.0	7200	Pave	NaN	Reg
27	28	20	RL	98.0	11478	Pave	NaN	Reg
28	29	20	RL	47.0	16321	Pave	NaN	IR1
29	30	30	RM	60.0	6324	Pave	NaN	IR1
...	...	...	...	...	...	...	...	...
1430	1431	60	RL	60.0	21930	Pave	NaN	IR3
1431	1432	120	RL	NaN	4928	Pave	NaN	IR1
1432	1433	30	RL	60.0	10800	Pave	Grvl	Reg
1433	1434	60	RL	93.0	10261	Pave	NaN	IR1
1434	1435	20	RL	80.0	17400	Pave	NaN	Reg
1435	1436	20	RL	80.0	8400	Pave	NaN	Reg
1436	1437	20	RL	60.0	9000	Pave	NaN	Reg
1437	1438	20	RL	96.0	12444	Pave	NaN	Reg
1438	1439	20	RM	90.0	7407	Pave	NaN	Reg
1439	1440	60	RL	80.0	11584	Pave	NaN	Reg
1440	1441	70	RL	79.0	11526	Pave	NaN	IR1
1441	1442	120	RM	NaN	4426	Pave	NaN	Reg
1442	1443	60	FV	85.0	11003	Pave	NaN	Reg
1443	1444	30	RL	NaN	8854	Pave	NaN	Reg
1444	1445	20	RL	63.0	8500	Pave	NaN	Reg
1445	1446	85	RL	70.0	8400	Pave	NaN	Reg
1446	1447	20	RL	NaN	26142	Pave	NaN	IR1
1447	1448	60	RL	80.0	10000	Pave	NaN	Reg
1448	1449	50	RL	70.0	11767	Pave	NaN	Reg
1449	1450	180	RM	21.0	1533	Pave	NaN	Reg
1450	1451	90	RL	60.0	9000	Pave	NaN	Reg
1451	1452	20	RL	78.0	9262	Pave	NaN	Reg

```
1452  1453         180       RM         35.0     3675   Pave   NaN      Reg
1453  1454          20       RL         90.0    17217   Pave   NaN      Reg
1454  1455          20       FV         62.0     7500   Pave  Pave      Reg
1455  1456          60       RL         62.0     7917   Pave   NaN      Reg
1456  1457          20       RL         85.0    13175   Pave   NaN      Reg
1457  1458          70       RL         66.0     9042   Pave   NaN      Reg
1458  1459          20       RL         68.0     9717   Pave   NaN      Reg
1459  1460          20       RL         75.0     9937   Pave   NaN      Reg

     LandContour Utilities    ...     PoolArea PoolQC  Fence MiscFeature  \
0            Lvl    AllPub    ...            0    NaN    NaN         NaN
1            Lvl    AllPub    ...            0    NaN    NaN         NaN
2            Lvl    AllPub    ...            0    NaN    NaN         NaN
3            Lvl    AllPub    ...            0    NaN    NaN         NaN
4            Lvl    AllPub    ...            0    NaN    NaN         NaN
5            Lvl    AllPub    ...            0    NaN  MnPrv        Shed
6            Lvl    AllPub    ...            0    NaN    NaN         NaN
7            Lvl    AllPub    ...            0    NaN    NaN        Shed
8            Lvl    AllPub    ...            0    NaN    NaN         NaN
9            Lvl    AllPub    ...            0    NaN    NaN         NaN
10           Lvl    AllPub    ...            0    NaN    NaN         NaN
11           Lvl    AllPub    ...            0    NaN    NaN         NaN
12           Lvl    AllPub    ...            0    NaN    NaN         NaN
13           Lvl    AllPub    ...            0    NaN    NaN         NaN
14           Lvl    AllPub    ...            0    NaN   GdWo         NaN
15           Lvl    AllPub    ...            0    NaN  GdPrv         NaN
16           Lvl    AllPub    ...            0    NaN    NaN        Shed
17           Lvl    AllPub    ...            0    NaN    NaN        Shed
18           Lvl    AllPub    ...            0    NaN    NaN         NaN
19           Lvl    AllPub    ...            0    NaN  MnPrv         NaN
20           Lvl    AllPub    ...            0    NaN    NaN         NaN
21           Bnk    AllPub    ...            0    NaN  GdPrv         NaN
22           Lvl    AllPub    ...            0    NaN    NaN         NaN
23           Lvl    AllPub    ...            0    NaN    NaN         NaN
24           Lvl    AllPub    ...            0    NaN  MnPrv         NaN
25           Lvl    AllPub    ...            0    NaN    NaN         NaN
26           Lvl    AllPub    ...            0    NaN    NaN         NaN
27           Lvl    AllPub    ...            0    NaN    NaN         NaN
28           Lvl    AllPub    ...            0    NaN    NaN         NaN
29           Lvl    AllPub    ...            0    NaN    NaN         NaN
...          ...       ...    ...          ...    ...    ...         ...
1430         Lvl    AllPub    ...            0    NaN    NaN         NaN
1431         Lvl    AllPub    ...            0    NaN    NaN         NaN
1432         Lvl    AllPub    ...            0    NaN    NaN         NaN
1433         Lvl    AllPub    ...            0    NaN    NaN         NaN
1434         Low    AllPub    ...            0    NaN    NaN         NaN
1435         Lvl    AllPub    ...            0    NaN  GdPrv         NaN
1436         Lvl    AllPub    ...            0    NaN   GdWo         NaN
```

1437	Lvl	AllPub	...	0	NaN	NaN	NaN
1438	Lvl	AllPub	...	0	NaN	MnPrv	NaN
1439	Lvl	AllPub	...	0	NaN	NaN	NaN
1440	Bnk	AllPub	...	0	NaN	NaN	NaN
1441	Lvl	AllPub	...	0	NaN	NaN	NaN
1442	Lvl	AllPub	...	0	NaN	NaN	NaN
1443	Lvl	AllPub	...	0	NaN	NaN	NaN
1444	Lvl	AllPub	...	0	NaN	NaN	NaN
1445	Lvl	AllPub	...	0	NaN	NaN	NaN
1446	Lvl	AllPub	...	0	NaN	NaN	NaN
1447	Lvl	AllPub	...	0	NaN	NaN	NaN
1448	Lvl	AllPub	...	0	NaN	GdWo	NaN
1449	Lvl	AllPub	...	0	NaN	NaN	NaN
1450	Lvl	AllPub	...	0	NaN	NaN	NaN
1451	Lvl	AllPub	...	0	NaN	NaN	NaN
1452	Lvl	AllPub	...	0	NaN	NaN	NaN
1453	Lvl	AllPub	...	0	NaN	NaN	NaN
1454	Lvl	AllPub	...	0	NaN	NaN	NaN
1455	Lvl	AllPub	...	0	NaN	NaN	NaN
1456	Lvl	AllPub	...	0	NaN	MnPrv	NaN
1457	Lvl	AllPub	...	0	NaN	GdPrv	Shed
1458	Lvl	AllPub	...	0	NaN	NaN	NaN
1459	Lvl	AllPub	...	0	NaN	NaN	NaN

	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	0	2	2008	WD	Normal	208500
1	0	5	2007	WD	Normal	181500
2	0	9	2008	WD	Normal	223500
3	0	2	2006	WD	Abnorml	140000
4	0	12	2008	WD	Normal	250000
5	700	10	2009	WD	Normal	143000
6	0	8	2007	WD	Normal	307000
7	350	11	2009	WD	Normal	200000
8	0	4	2008	WD	Abnorml	129900
9	0	1	2008	WD	Normal	118000
10	0	2	2008	WD	Normal	129500
11	0	7	2006	New	Partial	345000
12	0	9	2008	WD	Normal	144000
13	0	8	2007	New	Partial	279500
14	0	5	2008	WD	Normal	157000
15	0	7	2007	WD	Normal	132000
16	700	3	2010	WD	Normal	149000
17	500	10	2006	WD	Normal	90000
18	0	6	2008	WD	Normal	159000
19	0	5	2009	COD	Abnorml	139000
20	0	11	2006	New	Partial	325300
21	0	6	2007	WD	Normal	139400
22	0	9	2008	WD	Normal	230000

```
23          0      6   2007        WD         Normal     129900
24          0      5   2010        WD         Normal     154000
25          0      7   2009        WD         Normal     256300
26          0      5   2010        WD         Normal     134800
27          0      5   2010        WD         Normal     306000
28          0     12   2006        WD         Normal     207500
29          0      5   2008        WD         Normal      68500
...       ...    ...    ...       ...            ...        ...
1430        0      7   2006        WD         Normal     192140
1431        0     10   2009        WD         Normal     143750
1432        0      8   2007        WD         Normal      64500
1433        0      5   2008        WD         Normal     186500
1434        0      5   2006        WD         Normal     160000
1435        0      7   2008       COD        Abnorml     174000
1436        0      5   2007        WD         Normal     120500
1437        0     11   2008       New        Partial     394617
1438        0      4   2010        WD         Normal     149700
1439        0     11   2007        WD         Normal     197000
1440        0      9   2008        WD         Normal     191000
1441        0      5   2008        WD         Normal     149300
1442        0      4   2009        WD         Normal     310000
1443        0      5   2009        WD         Normal     121000
1444        0     11   2007        WD         Normal     179600
1445        0      5   2007        WD         Normal     129000
1446        0      4   2010        WD         Normal     157900
1447        0     12   2007        WD         Normal     240000
1448        0      5   2007        WD         Normal     112000
1449        0      8   2006        WD        Abnorml      92000
1450        0      9   2009        WD         Normal     136000
1451        0      5   2009       New        Partial     287090
1452        0      5   2006        WD         Normal     145000
1453        0      7   2006        WD        Abnorml      84500
1454        0     10   2009        WD         Normal     185000
1455        0      8   2007        WD         Normal     175000
1456        0      2   2010        WD         Normal     210000
1457     2500      5   2010        WD         Normal     266500
1458        0      4   2010        WD         Normal     142125
1459        0      6   2008        WD         Normal     147500

[1460 rows x 81 columns]
```

```
In [11]: train_df.dtypes
```

```
Out[11]: 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
                  ...
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
Length: 81, dtype: object
```

```
In [16]: train_val[:,80]
```

```
Out[16]: array([208500, 181500, 223500, ..., 266500, 142125, 147500], dtype=object)
```

```
In [17]: y = train_val[:,80]/1000
```

```
In [18]: y
```

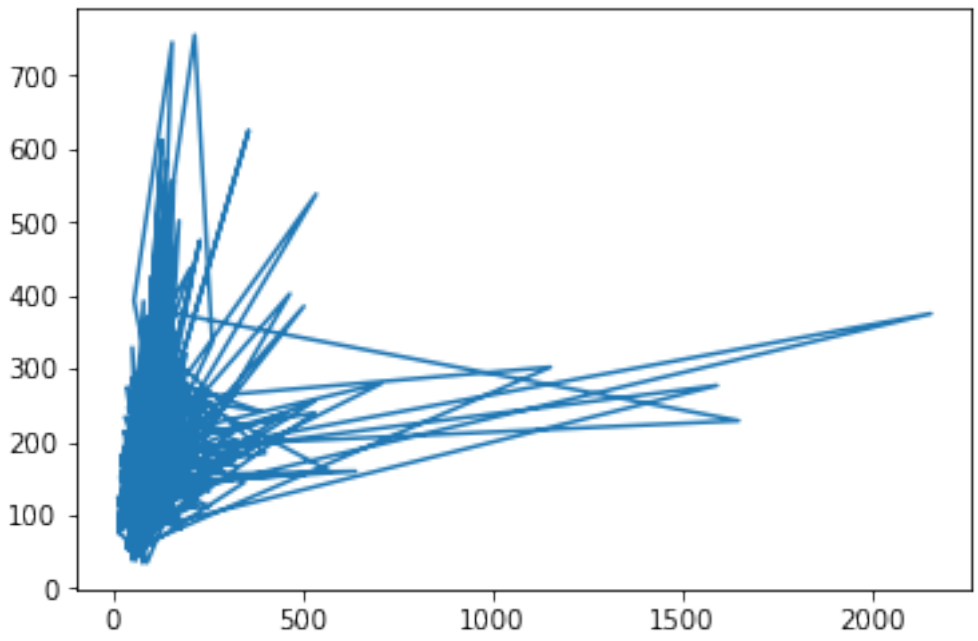
```
Out[18]: array([208.5, 181.5, 223.5, ..., 266.5, 142.125, 147.5], dtype=object)
```

```
In [19]: x1 = train_val[:,4]/100
```

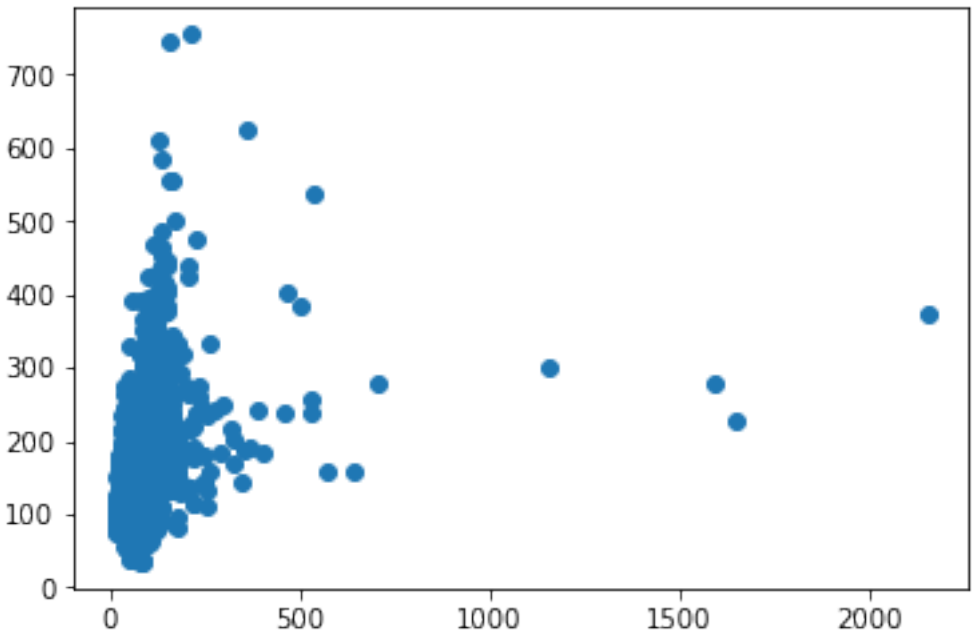
```
In [20]: x1
```

```
Out[20]: array([84.5, 96.0, 112.5, ..., 90.42, 97.17, 99.37], dtype=object)
```

```
In [21]: plt.plot(x1,y)
         plt.show()
```



```
In [22]: plt.scatter(x1,y)
         plt.show()
```




```
In [1]: import pandas as pd
        import numpy as np

In [2]: train = pd.read_csv('train.csv')
        test = pd.read_csv('test.csv')

In [3]: print ("Train data shape:", train.shape)
        print ("Test data shape:", test.shape)

Train data shape: (1460, 81)
Test data shape: (1459, 80)


In [4]: train.head()

Out[4]:    Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
        0   1          60       RL         65.0     8450   Pave   NaN      Reg
        1   2          20       RL         80.0     9600   Pave   NaN      Reg
        2   3          60       RL         68.0    11250   Pave   NaN      IR1
        3   4          70       RL         60.0     9550   Pave   NaN      IR1
        4   5          60       RL         84.0    14260   Pave   NaN      IR1

          LandContour Utilities    ...     PoolArea PoolQC Fence MiscFeature MiscVal  \
```

```
0          Lvl    AllPub    ...            0    NaN   NaN         NaN       0
1          Lvl    AllPub    ...            0    NaN   NaN         NaN       0
2          Lvl    AllPub    ...            0    NaN   NaN         NaN       0
3          Lvl    AllPub    ...            0    NaN   NaN         NaN       0
4          Lvl    AllPub    ...            0    NaN   NaN         NaN       0

  MoSold YrSold  SaleType  SaleCondition  SalePrice
0      2   2008        WD         Normal     208500
1      5   2007        WD         Normal     181500
2      9   2008        WD         Normal     223500
3      2   2006        WD        Abnorml     140000
4     12   2008        WD         Normal     250000

[5 rows x 81 columns]
```

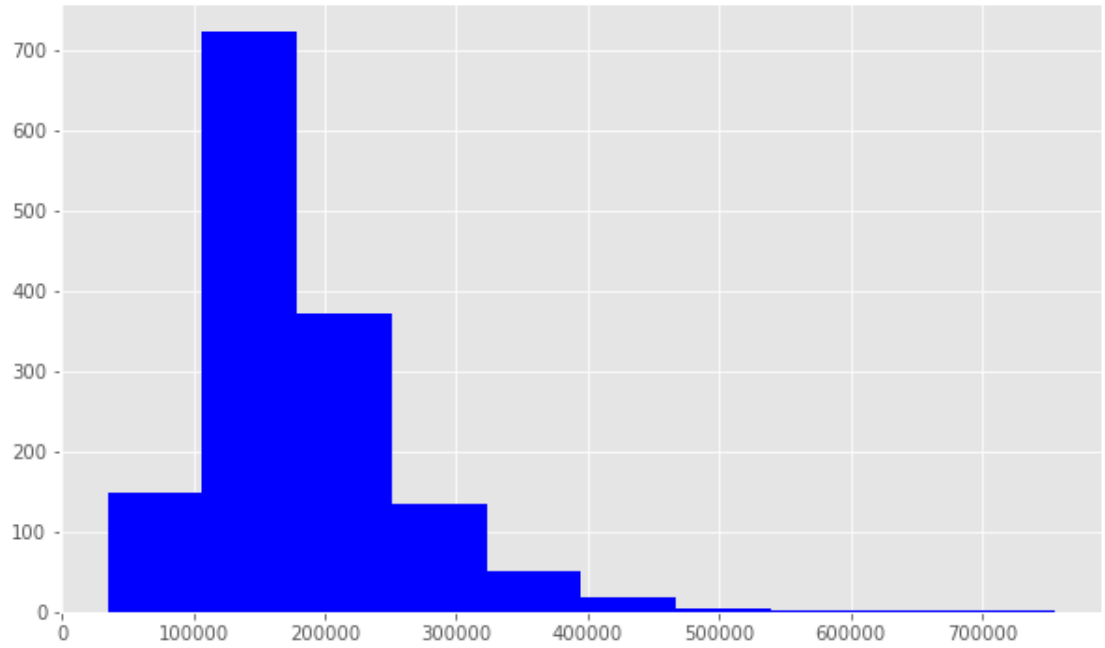
```
In [5]: import matplotlib.pyplot as plt
        plt.style.use(style='ggplot')
        plt.rcParams['figure.figsize'] = (10, 6)
```

```
In [6]: train.SalePrice.describe()
```

```
Out[6]: count      1460.000000
        mean     180921.195890
        std       79442.502883
        min       34900.000000
        25%      129975.000000
        50%      163000.000000
        75%      214000.000000
        max      755000.000000
        Name: SalePrice, dtype: float64
```

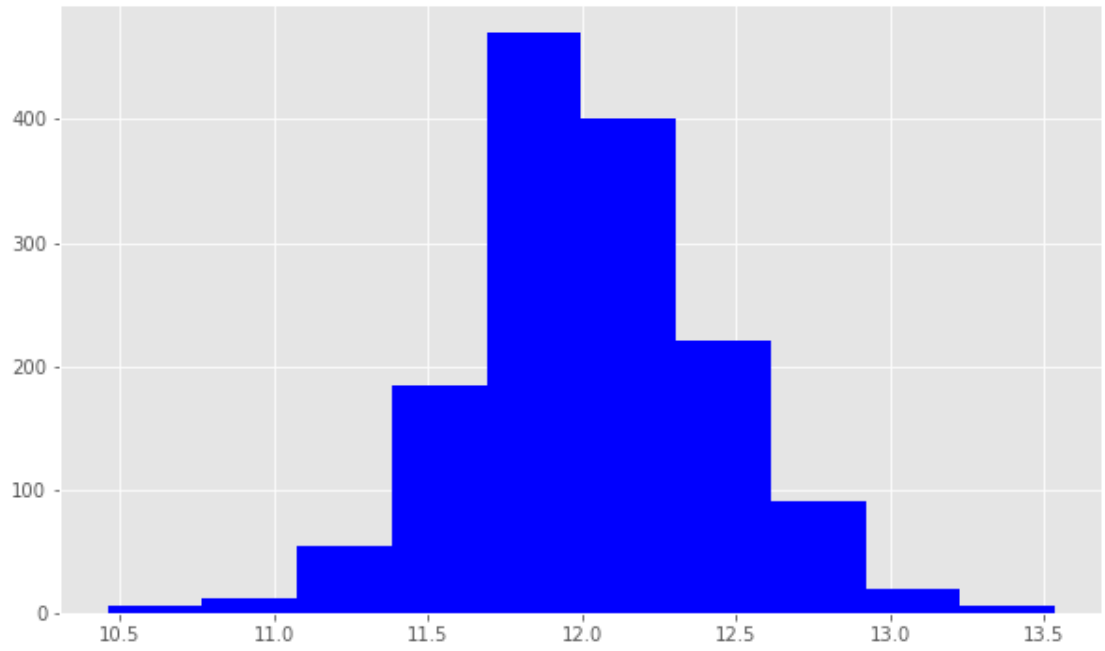
```
In [7]: print ("Skew is:", train.SalePrice.skew())
        plt.hist(train.SalePrice, color='blue')
        plt.show()
```

```
Skew is: 1.88287575977
```




```
In [8]: target = np.log(train.SalePrice)
        print ("Skew is:", target.skew())
        plt.hist(target, color='blue')
        plt.show()
```

```
Skew is: 0.121335062205
```



```
In [9]: numeric_features = train.select_dtypes(include=[np.number])
        numeric_features.dtypes
```

```
Out[9]: Id                int64
        MSSubClass        int64
        LotFrontage     float64
        LotArea           int64
        OverallQual       int64
        OverallCond       int64
        YearBuilt         int64
        YearRemodAdd      int64
        MasVnrArea      float64
        BsmtFinSF1        int64
        BsmtFinSF2        int64
        BsmtUnfSF         int64
        TotalBsmtSF       int64
        1stFlrSF          int64
        2ndFlrSF          int64
        LowQualFinSF      int64
        GrLivArea         int64
        BsmtFullBath      int64
        BsmtHalfBath      int64
        FullBath          int64
        HalfBath          int64
        BedroomAbvGr      int64
        KitchenAbvGr      int64
        TotRmsAbvGrd      int64
        Fireplaces        int64
        GarageYrBlt     float64
        GarageCars        int64
        GarageArea        int64
        WoodDeckSF        int64
        OpenPorchSF       int64
        EnclosedPorch     int64
        3SsnPorch         int64
        ScreenPorch       int64
        PoolArea          int64
        MiscVal           int64
        MoSold            int64
        YrSold            int64
        SalePrice         int64
        dtype: object
```

```
In [ ]:
```