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

In [12]:
```
import numpy as np
```

In [46]:
```
df_train = pd.read_csv('train.csv')
df_test = pd.read_csv('test.csv')
```

In [87]:
```
df_train.columns
```

Out[87]:
```
Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Stree
t',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgTy
pe',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRe
modAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrTy
pe',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQua
l',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heati
ng',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'Fu
llBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'Gara
geType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'Gara
geQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQ
C',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleTyp
```

```
e',
       'SaleCondition', 'SalePrice'],
      dtype='object')
```

In [42]:
```
df_test.head()
```

Out[42]:

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

5 rows × 80 columns

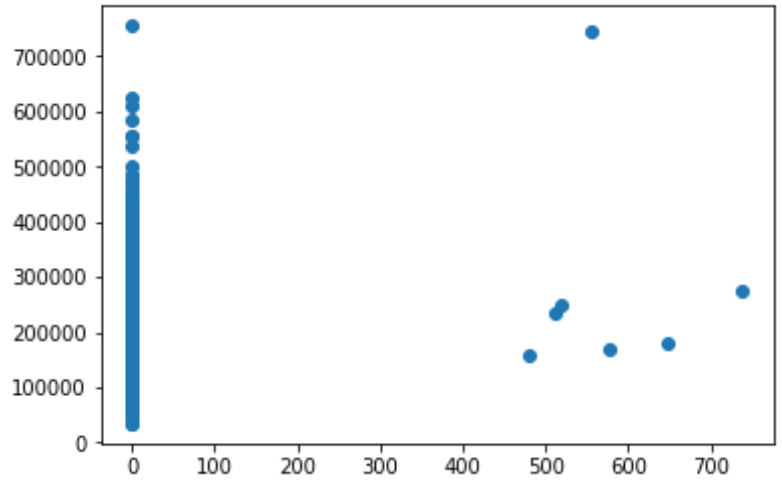
In [86]:
```
```

In [ ]:
```
```

In [19]:
```
import matplotlib.pyplot as plt
%matplotlib inline
```

In [32]:
```
plt.scatter(df_train['PoolArea'],df_train['SalePrice'])
```

Out[32]:
```
<matplotlib.collections.PathCollection at 0x1e409c324e0>
```



In [139]:
```
x = df_train.drop(['Id','SalePrice'], axis=1)
y = df_train['SalePrice']
x.columns
```

Out[139]:
```
Index(['MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street', 'Alley',
       'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandSlope',
       'Neighborhood', 'Condition1', 'Condition2', 'BldgType', 'HouseStyle',
       'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'RoofStyle',
       'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType', 'MasVnrArea',
       'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond',
       'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1', 'BsmtFinType2',
       'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating', 'HeatingQC',
       'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF', 'LowQualFinSF',
       'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath', 'HalfBath',
       'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual', 'TotRmsAbvGrd',
```

```
       'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType', 'Garage
YrBlt',
       'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual', 'Garag
eCond',
       'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3Ss
nPorch',
       'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeature', 'Mi
scVal',
       'MoSold', 'YrSold', 'SaleType', 'SaleCondition'],
      dtype='object')
```

In [140]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
c = ['MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street', 'All
ey',
       'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandSlope'
,
       'Neighborhood', 'Condition1', 'Condition2', 'BldgType', 'HouseSt
yle',
       'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'Roof
Style',
       'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType', 'MasVnrA
rea',
       'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond',
       'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1', 'BsmtFinType2',
       'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating', 'HeatingQC'
,
       'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF', 'LowQualFinS
F',
       'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath', 'HalfBa
th',
       'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual', 'TotRmsAbvGrd',
       'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType', 'Garage
YrBlt',
       'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual', 'Garag
eCond',
       'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3Ss
nPorch',
       'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeature', 'Mi
```

```
scVal',
       'MoSold', 'YrSold', 'SaleType', 'SaleCondition']
for i in c:

    s = le.fit_transform(x[i].astype(str))
    x[i] = s
x.head()
#from sklearn.preprocessing import OneHotEncoder
#neHotEncoder().fit_transform(x)
```

Out[140]:

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
0	9	3	75	859	1	2	3	3	0
1	4	3	90	1030	1	2	3	3	0
2	9	3	78	161	1	2	0	3	0
3	10	3	70	1021	1	2	0	3	0
4	9	3	94	386	1	2	0	3	0

5 rows × 79 columns

In [141]:
```
from sklearn.model_selection import train_test_split
```

In [142]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2, ran
dom_state=42)
```

In [143]:
```
x_train.head()
```

Out[143]:

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilit
254	4	3	80	852	1	2	3	3	
1066	9	3	69	789	1	2	0	3	

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilit
638	5	3	77	906	1	2	3	3	
799	8	3	70	734	1	2	3	3	
380	8	3	60	620	1	1	3	3	

5 rows × 79 columns

In [ ]:

In [144]:
```
from sklearn.linear_model import LinearRegression
```

In [145]:
```
clf = LinearRegression()
```

In [147]:
```
clf.fit(x_train, y_train)
```

Out[147]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [148]:
```
clf.score(x_test,y_test)
```

Out[148]:
```
0.7857307099868154
```

In [ ]: