In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats

from sklearn import svm
from sklearn.covariance import EllipticEnvelope
%matplotlib inline
```

```
/Users/swang/anaconda/lib/python2.7/site-packages/matplotlib/font_
manager.py:273: UserWarning: Matplotlib is building the font cache
using fc-list. This may take a moment.
  warnings.warn('Matplotlib is building the font cache using fc-li
st. This may take a moment.')
```

In [2]:

```
train = pd.read_csv("./input/train.csv")
test = pd.read_csv("prediction_training.csv").drop('Id',axis=1,inplace=False)
origin = pd.DataFrame(train['SalePrice'])
```

In [3]:

```
dif = np.abs(test-origin) > 12000
```

In [4]:

```
idx = dif[dif['SalePrice']].index.tolist()
```

In [5]:

```
train.drop(train.index[idx],inplace=True)
```

In [6]:

```
train.shape
```

Out[6]:

```
(1408, 81)
```

In [7]:

```
idx
```

```
Out[7]:

[4,
 11,
 13,
 20,
 46,
 66,
 70,
 167,
 178,
 185,
 199,
 224,
 261,
 309,
 313,
 318,
 349,
 412,
 423,
 440,
 454,
 477,
 478,
 523,
 540,
 581,
 585,
 588,
 595,
 654,
 688,
 691,
 774,
 798,
 875,
 898,
 926,
 970,
 987,
 1027,
 1109,
 1169,
 1182,
 1239,
 1256,
 1298,
 1324,
 1353,
 1359,
 1405,
 1442,
 1447]
```

In [ ]:

```
#Conclusion:Thus we done preprocssing on the given datset with outlier detectio and feature engineering
```