In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
```

In [2]:
```
%matplotlib inline
```

Binning of data

In [3]:
```
from scipy.stats import binned_statistic
from sklearn.preprocessing import LabelEncoder
```

Regression ouput

In [4]:
```
from sklearn.linear_model import LinearRegression
import statsmodels.formula.api as smf
```

In [5]:
```
os.getcwd()
```

Out[5]:
```
'C:\\Users\\Administrator\\Documents\\Master\\MSIS-5223-70250 - Programming f
or Data Sci - 8282017 - 159 PM\\Homework'
```

Import Personnel data

In [6]:
```
Personnel = pd.read_table('CaliforniaHospitalData_Personnel.txt', sep = '\t')
```

Import Hospital Data

In [7]:
```
Hospital = pd.read_csv('CaliforniaHospitalData.csv',sep=',')
```

Joint Two Table

In [8]:
```
df = pd.concat([Personnel, Hospital], axis=1, join='inner')
```

In [9]:
```
df.dtypes
```

Out[9]:
```
HospitalID         int64
Work_ID            int64
LastName          object
FirstName         object
Gender            object
PositionID         int64
PositionTitle     object
Compensation       int64
MaxTerm            int64
StartDate         object
HospitalID         int64
Name              object
Zip               object
Website           object
TypeControl       object
Teaching          object
DonorType         object
NoFTE            float64
NetPatRev        float64
InOperExp        float64
OutOperExp       float64
OperRev            int64
OperInc            int64
AvlBeds            int64
dtype: object
```

In [10]:
```
dfnew = df
```

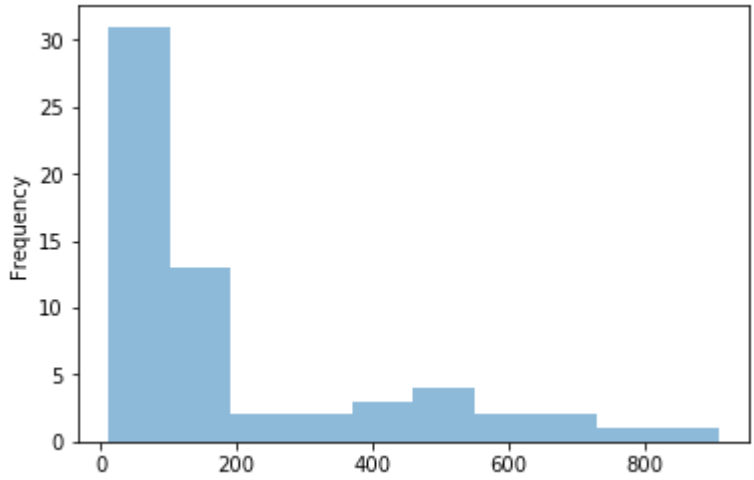
In [11]:
```
dfnew = dfnew[['Teaching', 'DonorType', 'Gender', 'PositionTitle', 'Compensati
on','TypeControl','AvlBeds', 'OperInc', 'OperRev']]
```

Bin the variable AvlBeds and create a categorical variable. (2 pts.)

In [12]:
```
dfnew_AvlBeds = dfnew[['AvlBeds']]
#ascending = dfnew_AvlBeds.sort_values(by='AvlBeds')
#ascending
```

In [13]:
```
dfnew_AvlBeds['AvlBeds'].plot.hist(alpha =0.5)
```

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x20b4724e898>



In [14]:
```
dfnew_AvlBeds['AvlBeds'].max()
```

Out[14]: 909

In [15]:
```
dfnew_AvlBeds['AvlBeds'].min()
```

Out[15]: 12

In [16]:
```
bin_interval=[10,50,200,910]
bin_counts,bin_edges,binnum = binned_statistic(dfnew_AvlBeds['AvlBeds'],dfnew_
AvlBeds['AvlBeds'],statistic='count',bins = bin_interval)
```

In [17]:
```
#bins = np.array(dfnew_AvlBeds['AvlBeds']);
```

In [18]:
```
dfnew_AvlBeds['AvlBeds'].value_counts()
```

Out[18]:
```
25     4
121    2
99     2
48     2
42     2
35     2
119    2
15     2
146    1
38     1
59     1
186    1
55     1
563    1
45     1
523    1
37     1
527    1
12     1
28     1
60     1
909    1
152    1
150    1
20     1
26     1
62     1
445    1
107    1
373    1
363    1
234    1
103    1
102    1
530    1
353    1
606    1
730    1
87     1
658    1
84     1
211    1
179    1
462    1
77     1
76     1
715    1
453    1
66     1
65     1
131    1
Name: AvlBeds, dtype: int64
```

In [19]:
```
#bin_counts,bin_edges,binnum = binned_statistic(dfnew_AvlBeds['AvlBeds'],dfnew
_AvlBeds['AvlBeds'],statistic='count',bins=3)
```

bin_interval now looks normal distribution

In [20]:
```
bin_counts
```

Out[20]:
```
array([ 19.,  25.,  17.])
```

In [21]:
```
bin_edges
```

Out[21]:
```
array([  10.,   50.,  200.,  910.])
```

In [22]:
```
#binlabels = ['BedRange_12-311','BedRange_312-610','BedRange_611-909']
```

Recode the values in the age column based on the beginning

In [23]:
```
binlabels = ['BedRange_10-50','BedRange_51-200','BedRange_201-910']
```

In [24]:
```
Bed_categ = pd.cut(dfnew_AvlBeds['AvlBeds'],bins =
3,right=False,retbins=False, labels=binlabels)
```

In [25]:
```
Bed_categ
```

Out[25]:

0	BedRange_10-50
1	BedRange_10-50
2	BedRange_10-50
3	BedRange_10-50
4	BedRange_10-50
5	BedRange_10-50
6	BedRange_10-50
7	BedRange_10-50
8	BedRange_10-50
9	BedRange_10-50
10	BedRange_10-50
11	BedRange_10-50
12	BedRange_10-50
13	BedRange_10-50
14	BedRange_10-50
15	BedRange_10-50
16	BedRange_10-50
17	BedRange_10-50
18	BedRange_10-50
19	BedRange_10-50
20	BedRange_10-50
21	BedRange_10-50
22	BedRange_10-50
23	BedRange_10-50
24	BedRange_10-50
25	BedRange_10-50
26	BedRange_10-50
27	BedRange_10-50
28	BedRange_10-50
29	BedRange_10-50
...	...
31	BedRange_10-50
32	BedRange_10-50
33	BedRange_10-50
34	BedRange_10-50
35	BedRange_10-50
36	BedRange_10-50
37	BedRange_10-50
38	BedRange_10-50
39	BedRange_10-50
40	BedRange_10-50
41	BedRange_10-50
42	BedRange_10-50
43	BedRange_10-50
44	BedRange_51-200
45	BedRange_10-50
46	BedRange_201-910
47	BedRange_51-200
48	BedRange_10-50
49	BedRange_51-200
50	BedRange_51-200
51	BedRange_51-200
52	BedRange_51-200
53	BedRange_51-200
54	BedRange_51-200
55	BedRange_201-910
56	BedRange_51-200

```
57     BedRange_51-200
58     BedRange_51-200
59    BedRange_201-910
60    BedRange_201-910
Name: AvlBeds, Length: 61, dtype: category
Categories (3, object): [BedRange_10-50 < BedRange_201-910 < BedRange_51-200]
```

In [26]:
```
Bed_categ.name = 'Bed_categ'
```

Take the binning data and add it as a column to the dataframe

In [27]:
```
dfnew1 = dfnew.join(pd.DataFrame(Bed_categ))
dfnew1
```

Out[27]:

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	AvlBeds
0	Small/Rural	Charity	F	Regional Representative	46978	District	15
1	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	99
2	Small/Rural	Charity	F	Regional Representative	46978	District	107
3	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	55
4	Small/Rural	Charity	M	Regional Representative	46978	Investor	42
5	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	66
6	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	65
7	Small/Rural	Charity	F	Regional Representative	46978	District	35
8	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	48
9	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	186
10	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	119
11	Small/Rural	Charity	F	State Board Representative	89473	Non Profit	119
12	Small/Rural	Charity	F	State Board Representative	89473	Non Profit	48
13	Small/Rural	Charity	F	State Board Representative	89473	Non Profit	25
14	Small/Rural	Charity	M	State Board Representative	89473	Non Profit	60
15	Small/Rural	Charity	M	Acting Director	248904	Non Profit	25
16	Small/Rural	Charity	F	Acting Director	248904	Non Profit	121
17	Small/Rural	Charity	M	Acting Director	248904	District	62
18	Small/Rural	Charity	M	Acting Director	248904	Investor	25

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	AvlBeds
19	Small/Rural	Charity	F	Acting Director	248904	District	131
20	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	42
21	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	59
22	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	35
23	Small/Rural	Charity	M	Safety Inspection Member	23987	Non Profit	26
24	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	25
25	Small/Rural	Charity	M	Regional Representative	46978	Investor	84
26	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	152
27	Small/Rural	Charity	M	Regional Representative	46978	District	146
28	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	121
29	Small/Rural	Charity	M	Acting Director	248904	Non Profit	20
...	...	...	...	...	...	...	...
31	Small/Rural	Charity	F	Safety Inspection Member	23987	District	28
32	Small/Rural	Charity	F	Safety Inspection Member	23987	District	179
33	Small/Rural	Charity	F	Safety Inspection Member	23987	District	77
34	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	102

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	AvlBeds
35	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	45
36	Small/Rural	Charity	F	State Board Representative	89473	District	150
37	Small/Rural	Charity	F	State Board Representative	89473	District	37
38	Small/Rural	Charity	M	Acting Director	248904	District	76
39	Small/Rural	Charity	M	Regional Representative	46978	City/County	87
40	Small/Rural	Charity	F	Regional Representative	46978	District	38
41	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	12
42	Small/Rural	Charity	M	Regional Representative	46978	District	99
43	Small/Rural	Charity	F	Acting Director	248904	Non Profit	103
44	Teaching	Alumni	F	Regional Representative	46978	Non Profit	363
45	Teaching	Alumni	F	State Board Representative	89473	Non Profit	234
46	Teaching	Alumni	F	Safety Inspection Member	23987	City/County	715
47	Teaching	Alumni	M	Acting Director	248904	City/County	353
48	Teaching	Alumni	M	Regional Representative	46978	City/County	211
49	Teaching	Alumni	F	State Board Representative	89473	City/County	373
50	Teaching	Alumni	M	State Board Representative	89473	City/County	523
51	Teaching	Alumni	M	Safety Inspection Member	23987	City/County	453
52	Teaching	Alumni	M	State Board Representative	89473	Non Profit	606

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	AvlBeds
53	Teaching	Alumni	M	Safety Inspection Member	23987	Non Profit	563
54	Teaching	Alumni	M	Acting Director	248904	Non Profit	527
55	Teaching	Alumni	F	Acting Director	248904	Non Profit	658
56	Teaching	Alumni	M	Safety Inspection Member	23987	Non Profit	530
57	Teaching	Alumni	M	Safety Inspection Member	23987	Non Profit	462
58	Teaching	Alumni	F	Acting Director	248904	Non Profit	445
59	Teaching	Alumni	F	Regional Representative	46978	Non Profit	730
60	Teaching	Alumni	M	State Board Representative	89473	Non Profit	909

61 rows × 10 columns

In [28]:
```
dfnew1[['AvlBeds','Bed_categ']].sort_values(by='AvlBeds')
```

Out[28]:

	AvlBeds	Bed_categ
41	12	BedRange_10-50
0	15	BedRange_10-50
30	15	BedRange_10-50
29	20	BedRange_10-50
24	25	BedRange_10-50
13	25	BedRange_10-50
15	25	BedRange_10-50
18	25	BedRange_10-50
23	26	BedRange_10-50
31	28	BedRange_10-50
7	35	BedRange_10-50
22	35	BedRange_10-50
37	37	BedRange_10-50
40	38	BedRange_10-50
4	42	BedRange_10-50
20	42	BedRange_10-50
35	45	BedRange_10-50
8	48	BedRange_10-50
12	48	BedRange_10-50
3	55	BedRange_10-50
21	59	BedRange_10-50
14	60	BedRange_10-50
17	62	BedRange_10-50
6	65	BedRange_10-50
5	66	BedRange_10-50
38	76	BedRange_10-50
33	77	BedRange_10-50
25	84	BedRange_10-50
39	87	BedRange_10-50
1	99	BedRange_10-50
...	...	...
34	102	BedRange_10-50

	AvlBeds	Bed_categ
43	103	BedRange_10-50
2	107	BedRange_10-50
10	119	BedRange_10-50
11	119	BedRange_10-50
16	121	BedRange_10-50
28	121	BedRange_10-50
19	131	BedRange_10-50
27	146	BedRange_10-50
36	150	BedRange_10-50
26	152	BedRange_10-50
32	179	BedRange_10-50
9	186	BedRange_10-50
48	211	BedRange_10-50
45	234	BedRange_10-50
47	353	BedRange_51-200
44	363	BedRange_51-200
49	373	BedRange_51-200
58	445	BedRange_51-200
51	453	BedRange_51-200
57	462	BedRange_51-200
50	523	BedRange_51-200
54	527	BedRange_51-200
56	530	BedRange_51-200
53	563	BedRange_51-200
52	606	BedRange_51-200
55	658	BedRange_201-910
46	715	BedRange_201-910
59	730	BedRange_201-910
60	909	BedRange_201-910

61 rows × 2 columns

In [29]:
```
df_Bed = pd.get_dummies(dfnew1['Bed_categ'],sparse=True)
#dummies= pd.get_dummies(dfnew1['Bed_categ'])
 #df_with_dummy = df[['data1']].join(dummies)
# df_with_dummy
#dummies['BedRange_10-50'] = dummies['BedRange_10-50'].apply(dummies.get).asty
pe('flo')
#dummies.dtypes
```

In [30]:
```
df_Bed.head()
#dummies.head()
df_Bed.dtypes
```

Out[30]:
```
BedRange_10-50      uint8
BedRange_201-910    uint8
BedRange_51-200     uint8
dtype: object
```

In [31]:
```
#df_Bed['BedRange_10-50'] = number.fit_transform(df_Bed['BedRange_10-50'].asty
pe('str'))
```

In [32]:
```
dfnew1 = dfnew1.join(df_Bed)
```

In [33]:
```
dfnew1 = dfnew1.drop('Bed_categ', 1)
```

In [34]:
```
dfnew1 = dfnew1.drop('AvlBeds', 1)
```

In [35]:
```
dfnew1
```

Out[35]:

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	Ope
0	Small/Rural	Charity	F	Regional Representative	46978	District	-550593
1	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	-444391
2	Small/Rural	Charity	F	Regional Representative	46978	District	-215292
3	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	304041
4	Small/Rural	Charity	M	Regional Representative	46978	Investor	140011
5	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	123863
6	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	272400
7	Small/Rural	Charity	F	Regional Representative	46978	District	-451423
8	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	21025
9	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	7943967
10	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	568387
11	Small/Rural	Charity	F	State Board Representative	89473	Non Profit	568387
12	Small/Rural	Charity	F	State Board Representative	89473	Non Profit	878929
13	Small/Rural	Charity	F	State Board Representative	89473	Non Profit	165732
14	Small/Rural	Charity	M	State Board Representative	89473	Non Profit	135555
15	Small/Rural	Charity	M	Acting Director	248904	Non Profit	-380431
16	Small/Rural	Charity	F	Acting Director	248904	Non Profit	978071
17	Small/Rural	Charity	M	Acting Director	248904	District	93097
18	Small/Rural	Charity	M	Acting Director	248904	Investor	619833
19	Small/Rural	Charity	F	Acting Director	248904	District	906670

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	Ope
20	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	-125478
21	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	203421
22	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	554638
23	Small/Rural	Charity	M	Safety Inspection Member	23987	Non Profit	278395
24	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	325224
25	Small/Rural	Charity	M	Regional Representative	46978	Investor	123211
26	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	110520
27	Small/Rural	Charity	M	Regional Representative	46978	District	260069
28	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	279100
29	Small/Rural	Charity	M	Acting Director	248904	Non Profit	735057
...	...	...	...	...	...	...	...
31	Small/Rural	Charity	F	Safety Inspection Member	23987	District	-512923
32	Small/Rural	Charity	F	Safety Inspection Member	23987	District	591431
33	Small/Rural	Charity	F	Safety Inspection Member	23987	District	-506146
34	Small/Rural	Charity	F	Safety Inspection Member	23987	Non Profit	-785086
35	Small/Rural	Charity	F	Regional Representative	46978	Non Profit	-110817

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	Ope
36	Small/Rural	Charity	F	State Board Representative	89473	District	157269
37	Small/Rural	Charity	F	State Board Representative	89473	District	-791156
38	Small/Rural	Charity	M	Acting Director	248904	District	-158599
39	Small/Rural	Charity	M	Regional Representative	46978	City/County	-356389
40	Small/Rural	Charity	F	Regional Representative	46978	District	-122923
41	Small/Rural	Charity	M	Regional Representative	46978	Non Profit	-191048
42	Small/Rural	Charity	M	Regional Representative	46978	District	-179274
43	Small/Rural	Charity	F	Acting Director	248904	Non Profit	210180
44	Teaching	Alumni	F	Regional Representative	46978	Non Profit	538609
45	Teaching	Alumni	F	State Board Representative	89473	Non Profit	281075
46	Teaching	Alumni	F	Safety Inspection Member	23987	City/County	-321163
47	Teaching	Alumni	M	Acting Director	248904	City/County	-119949
48	Teaching	Alumni	M	Regional Representative	46978	City/County	-810220
49	Teaching	Alumni	F	State Board Representative	89473	City/County	572973
50	Teaching	Alumni	M	State Board Representative	89473	City/County	-314000
51	Teaching	Alumni	M	Safety Inspection Member	23987	City/County	187793
52	Teaching	Alumni	M	State Board Representative	89473	Non Profit	200563
53	Teaching	Alumni	M	Safety Inspection Member	23987	Non Profit	537812
54	Teaching	Alumni	M	Acting Director	248904	Non Profit	146047

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	Ope
55	Teaching	Alumni	F	Acting Director	248904	Non Profit	4861612
56	Teaching	Alumni	M	Safety Inspection Member	23987	Non Profit	5990693
57	Teaching	Alumni	M	Safety Inspection Member	23987	Non Profit	3850812
58	Teaching	Alumni	F	Acting Director	248904	Non Profit	5927399
59	Teaching	Alumni	F	Regional Representative	46978	Non Profit	1444459
60	Teaching	Alumni	M	State Board Representative	89473	Non Profit	164872

61 rows × 11 columns

In [36]:
```
# Create dummies for gender
#df_gender = pd.get_dummies(dfnew1['Gender'],prefix='Gender_')
#df_gender.head()
#dfnew1 =dfnew1.join(df_gender)
#dfnew1
```

In [37]:
```
number = LabelEncoder()
dfnew1['Gender'] = number.fit_transform(dfnew1['Gender'].astype('str'))
dfnew1.dtypes
```

Out[37]:
```
Teaching            object
DonorType           object
Gender               int64
PositionTitle       object
Compensation         int64
TypeControl         object
OperInc              int64
OperRev              int64
BedRange_10-50       uint8
BedRange_201-910     uint8
BedRange_51-200      uint8
dtype: object
```

In [38]:
```
dfnew1.dtypes
```

Out[38]:
```
Teaching            object
DonorType           object
Gender               int64
PositionTitle       object
Compensation         int64
TypeControl         object
OperInc              int64
OperRev              int64
BedRange_10-50       uint8
BedRange_201-910     uint8
BedRange_51-200      uint8
dtype: object
```

Create Teaching column

In [39]:
```
#df_Teaching = pd.get_dummies(dfnew1['DonorType'],prefix='Donor_')
#df_Teaching.head()
#dfnew1 =dfnew1.join(df_Teaching )
number = LabelEncoder()
dfnew1['DonorType'] = number.fit_transform(dfnew1['DonorType'].astype('str'))
#dfnew1
```

In [40]:
```
dfnew1['Teaching'] = number.fit_transform(dfnew1['Teaching'].astype('str'))
```

In [41]:
```
dfnew1
```

Out[41]:

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
0	0	1	0	Regional Representative	46978	District	-5505933
1	0	1	1	Regional Representative	46978	Non Profit	-444391
2	0	1	0	Regional Representative	46978	District	-2152920
3	0	1	1	Regional Representative	46978	Non Profit	3040415
4	0	1	1	Regional Representative	46978	Investor	14001153
5	0	1	1	Regional Representative	46978	Non Profit	12386360
6	0	1	1	Regional Representative	46978	Non Profit	2724001
7	0	1	0	Regional Representative	46978	District	-4514239
8	0	1	0	Regional Representative	46978	Non Profit	21025
9	0	1	0	Regional Representative	46978	Non Profit	7943967
10	0	1	0	Regional Representative	46978	Non Profit	5683870
11	0	1	0	State Board Representative	89473	Non Profit	5683870
12	0	1	0	State Board Representative	89473	Non Profit	878929
13	0	1	0	State Board Representative	89473	Non Profit	1657323
14	0	1	1	State Board Representative	89473	Non Profit	1355554
15	0	1	1	Acting Director	248904	Non Profit	-3804313
16	0	1	0	Acting Director	248904	Non Profit	9780718
17	0	1	1	Acting Director	248904	District	93097
18	0	1	1	Acting Director	248904	Investor	6198330
19	0	1	0	Acting Director	248904	District	906670

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
20	0	1	0	Safety Inspection Member	23987	Non Profit	-1254783
21	0	1	0	Safety Inspection Member	23987	Non Profit	2034214
22	0	1	0	Safety Inspection Member	23987	Non Profit	5546385
23	0	1	1	Safety Inspection Member	23987	Non Profit	2783953
24	0	1	0	Regional Representative	46978	Non Profit	3252242
25	0	1	1	Regional Representative	46978	Investor	12321176
26	0	1	0	Regional Representative	46978	Non Profit	1105205
27	0	1	1	Regional Representative	46978	District	2600695
28	0	1	1	Regional Representative	46978	Non Profit	2791000
29	0	1	1	Acting Director	248904	Non Profit	735057
...	...	...	...	...	...	...	...
31	0	1	0	Safety Inspection Member	23987	District	-5129238
32	0	1	0	Safety Inspection Member	23987	District	591431
33	0	1	0	Safety Inspection Member	23987	District	-5061461
34	0	1	0	Safety Inspection Member	23987	Non Profit	-7850861
35	0	1	0	Regional Representative	46978	Non Profit	-1108177

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIn
36	0	1	0	State Board Representative	89473	District	1572695
37	0	1	0	State Board Representative	89473	District	-791156
38	0	1	1	Acting Director	248904	District	-1585993
39	0	1	1	Regional Representative	46978	City/County	-3563893
40	0	1	0	Regional Representative	46978	District	-1229238
41	0	1	1	Regional Representative	46978	Non Profit	-191048
42	0	1	1	Regional Representative	46978	District	-179274
43	0	1	0	Acting Director	248904	Non Profit	210180
44	1	0	0	Regional Representative	46978	Non Profit	53860958
45	1	0	0	State Board Representative	89473	Non Profit	2810751
46	1	0	0	Safety Inspection Member	23987	City/County	-3211633
47	1	0	1	Acting Director	248904	City/County	-1199493(
48	1	0	1	Regional Representative	46978	City/County	-8102204
49	1	0	0	State Board Representative	89473	City/County	5729737
50	1	0	1	State Board Representative	89473	City/County	-3140006(
51	1	0	1	Safety Inspection Member	23987	City/County	18779323
52	1	0	1	State Board Representative	89473	Non Profit	20056379
53	1	0	1	Safety Inspection Member	23987	Non Profit	53781218
54	1	0	1	Acting Director	248904	Non Profit	14604710

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
55	1	0	0	Acting Director	248904	Non Profit	48616127
56	1	0	1	Safety Inspection Member	23987	Non Profit	5990693
57	1	0	1	Safety Inspection Member	23987	Non Profit	38508125
58	1	0	0	Acting Director	248904	Non Profit	59273999
59	1	0	0	Regional Representative	46978	Non Profit	14444599
60	1	0	1	State Board Representative	89473	Non Profit	16487241

61 rows × 11 columns

Create dummies for Position column

In [42]:
```
#df_PositionTitles = pd.get_dummies(dfnew1['PositionTitles'],prefix='Donor_')
#df_PositionTitles.head()
#dfnew1 =dfnew1.join(df_PositionTitles )
df_Position = pd.get_dummies(dfnew1['PositionTitle'],prefix ='_')
df_Position.dtypes
dfnew1 = dfnew1.join(df_Position)
dfnew1
```

Out[42]:

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
0	0	1	0	Regional Representative	46978	District	-5505933
1	0	1	1	Regional Representative	46978	Non Profit	-444391
2	0	1	0	Regional Representative	46978	District	-2152920
3	0	1	1	Regional Representative	46978	Non Profit	3040415
4	0	1	1	Regional Representative	46978	Investor	14001153
5	0	1	1	Regional Representative	46978	Non Profit	12386360
6	0	1	1	Regional Representative	46978	Non Profit	2724001
7	0	1	0	Regional Representative	46978	District	-4514239
8	0	1	0	Regional Representative	46978	Non Profit	21025
9	0	1	0	Regional Representative	46978	Non Profit	7943967
10	0	1	0	Regional Representative	46978	Non Profit	5683870
11	0	1	0	State Board Representative	89473	Non Profit	5683870
12	0	1	0	State Board Representative	89473	Non Profit	878929
13	0	1	0	State Board Representative	89473	Non Profit	1657323
14	0	1	1	State Board Representative	89473	Non Profit	1355554
15	0	1	1	Acting Director	248904	Non Profit	-3804313
16	0	1	0	Acting Director	248904	Non Profit	9780718
17	0	1	1	Acting Director	248904	District	93097
18	0	1	1	Acting Director	248904	Investor	6198330
19	0	1	0	Acting Director	248904	District	906670

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
20	0	1	0	Safety Inspection Member	23987	Non Profit	-1254783
21	0	1	0	Safety Inspection Member	23987	Non Profit	2034214
22	0	1	0	Safety Inspection Member	23987	Non Profit	5546385
23	0	1	1	Safety Inspection Member	23987	Non Profit	2783953
24	0	1	0	Regional Representative	46978	Non Profit	3252242
25	0	1	1	Regional Representative	46978	Investor	12321176
26	0	1	0	Regional Representative	46978	Non Profit	1105205
27	0	1	1	Regional Representative	46978	District	2600695
28	0	1	1	Regional Representative	46978	Non Profit	2791000
29	0	1	1	Acting Director	248904	Non Profit	735057
...	...	...	...	...	...	...	...
31	0	1	0	Safety Inspection Member	23987	District	-5129238
32	0	1	0	Safety Inspection Member	23987	District	591431
33	0	1	0	Safety Inspection Member	23987	District	-5061461
34	0	1	0	Safety Inspection Member	23987	Non Profit	-7850861
35	0	1	0	Regional Representative	46978	Non Profit	-1108177

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
36	0	1	0	State Board Representative	89473	District	1572695
37	0	1	0	State Board Representative	89473	District	-791156
38	0	1	1	Acting Director	248904	District	-1585993
39	0	1	1	Regional Representative	46978	City/County	-3563893
40	0	1	0	Regional Representative	46978	District	-1229238
41	0	1	1	Regional Representative	46978	Non Profit	-191048
42	0	1	1	Regional Representative	46978	District	-179274
43	0	1	0	Acting Director	248904	Non Profit	210180
44	1	0	0	Regional Representative	46978	Non Profit	53860958
45	1	0	0	State Board Representative	89473	Non Profit	2810751
46	1	0	0	Safety Inspection Member	23987	City/County	-32116333
47	1	0	1	Acting Director	248904	City/County	-1199493(
48	1	0	1	Regional Representative	46978	City/County	-8102204
49	1	0	0	State Board Representative	89473	City/County	5729737
50	1	0	1	State Board Representative	89473	City/County	-3140006(
51	1	0	1	Safety Inspection Member	23987	City/County	18779323
52	1	0	1	State Board Representative	89473	Non Profit	20056379
53	1	0	1	Safety Inspection Member	23987	Non Profit	53781218

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
54	1	0	1	Acting Director	248904	Non Profit	14604710
55	1	0	0	Acting Director	248904	Non Profit	48616127
56	1	0	1	Safety Inspection Member	23987	Non Profit	5990693
57	1	0	1	Safety Inspection Member	23987	Non Profit	38508125
58	1	0	0	Acting Director	248904	Non Profit	59273999
59	1	0	0	Regional Representative	46978	Non Profit	14444599
60	1	0	1	State Board Representative	89473	Non Profit	16487241

61 rows × 15 columns

Create dummies for column Type Control

In [43]:
```
#df_TyoeControl = pd.get_dummies(dfnew1['TypeControl'],prefix='Donor_')
#df_TypeControl.head()
df_Control = pd.get_dummies(dfnew1['TypeControl'],prefix='DonorType_')
df_Control.head()
dfnew1 =dfnew1.join(df_Control)
dfnew1
```

Out[43]:

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
0	0	1	0	Regional Representative	46978	District	-5505933
1	0	1	1	Regional Representative	46978	Non Profit	-444391
2	0	1	0	Regional Representative	46978	District	-2152920
3	0	1	1	Regional Representative	46978	Non Profit	3040415
4	0	1	1	Regional Representative	46978	Investor	14001153
5	0	1	1	Regional Representative	46978	Non Profit	12386360
6	0	1	1	Regional Representative	46978	Non Profit	2724001
7	0	1	0	Regional Representative	46978	District	-4514239
8	0	1	0	Regional Representative	46978	Non Profit	21025
9	0	1	0	Regional Representative	46978	Non Profit	7943967
10	0	1	0	Regional Representative	46978	Non Profit	5683870
11	0	1	0	State Board Representative	89473	Non Profit	5683870
12	0	1	0	State Board Representative	89473	Non Profit	878929
13	0	1	0	State Board Representative	89473	Non Profit	1657323
14	0	1	1	State Board Representative	89473	Non Profit	1355554
15	0	1	1	Acting Director	248904	Non Profit	-3804313
16	0	1	0	Acting Director	248904	Non Profit	9780718
17	0	1	1	Acting Director	248904	District	93097
18	0	1	1	Acting Director	248904	Investor	6198330
19	0	1	0	Acting Director	248904	District	906670

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
20	0	1	0	Safety Inspection Member	23987	Non Profit	-1254783
21	0	1	0	Safety Inspection Member	23987	Non Profit	2034214
22	0	1	0	Safety Inspection Member	23987	Non Profit	5546385
23	0	1	1	Safety Inspection Member	23987	Non Profit	2783953
24	0	1	0	Regional Representative	46978	Non Profit	3252242
25	0	1	1	Regional Representative	46978	Investor	12321176
26	0	1	0	Regional Representative	46978	Non Profit	1105205
27	0	1	1	Regional Representative	46978	District	2600695
28	0	1	1	Regional Representative	46978	Non Profit	2791000
29	0	1	1	Acting Director	248904	Non Profit	735057
...	...	...	...	...	...	...	...
31	0	1	0	Safety Inspection Member	23987	District	-5129238
32	0	1	0	Safety Inspection Member	23987	District	591431
33	0	1	0	Safety Inspection Member	23987	District	-5061461
34	0	1	0	Safety Inspection Member	23987	Non Profit	-7850861
35	0	1	0	Regional Representative	46978	Non Profit	-1108177

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
36	0	1	0	State Board Representative	89473	District	1572695
37	0	1	0	State Board Representative	89473	District	-791156
38	0	1	1	Acting Director	248904	District	-1585993
39	0	1	1	Regional Representative	46978	City/County	-3563893
40	0	1	0	Regional Representative	46978	District	-1229238
41	0	1	1	Regional Representative	46978	Non Profit	-191048
42	0	1	1	Regional Representative	46978	District	-179274
43	0	1	0	Acting Director	248904	Non Profit	210180
44	1	0	0	Regional Representative	46978	Non Profit	53860958
45	1	0	0	State Board Representative	89473	Non Profit	2810751
46	1	0	0	Safety Inspection Member	23987	City/County	-32116333
47	1	0	1	Acting Director	248904	City/County	-1199493(
48	1	0	1	Regional Representative	46978	City/County	-8102204
49	1	0	0	State Board Representative	89473	City/County	5729737
50	1	0	1	State Board Representative	89473	City/County	-3140006(
51	1	0	1	Safety Inspection Member	23987	City/County	18779323
52	1	0	1	State Board Representative	89473	Non Profit	20056379
53	1	0	1	Safety Inspection Member	23987	Non Profit	53781218

	Teaching	DonorType	Gender	PositionTitle	Compensation	TypeControl	OperIr
54	1	0	1	Acting Director	248904	Non Profit	14604710
55	1	0	0	Acting Director	248904	Non Profit	48616127
56	1	0	1	Safety Inspection Member	23987	Non Profit	5990693
57	1	0	1	Safety Inspection Member	23987	Non Profit	38508125
58	1	0	0	Acting Director	248904	Non Profit	59273999
59	1	0	0	Regional Representative	46978	Non Profit	14444599
60	1	0	1	State Board Representative	89473	Non Profit	16487241

61 rows × 19 columns

In [44]:
```
dfnew1 = dfnew1.drop('PositionTitle', 1)
```

In [45]:
```
dfnew1 = dfnew1.drop('TypeControl', 1)
```

In [46]:
```
#df_Bed['BedRange_201-910'] = number.fit_transform(df_Bed['BedRange_201-910'].
astype('str'))
```

In [47]:
```
#dfnew1['BedRange_51-200'] = number.fit_transform(dfnew1['BedRange_51-200'].as
type('str'))
```

After I created dummies for categorical variables. Some of the data type pf variables are wierd. I try convert to int or float using

df_Bed['BedRange_10-50'] = number.fit_transform(df_Bed['BedRange_10-50'].astype('str'))

or #dfnew1['BedRange_51-200'] = number.fit_transform(dfnew1['BedRange_51-200'].astype('str'))

df_Bed['BedRange_201-910'] = number.fit_transform(df_Bed['BedRange_201-910'].astype('str'))

so my models is denpence on OperRev ~ Teaching + DonorType since they has interger data type.

Please tell me what is happened.

Base on model below I see the model Operation Revenue is better since Rsquare is better and P value is smaller

In [48]:
```
dfnew1.dtypes
```

Out[48]:
```
Teaching                        int64
DonorType                       int64
Gender                          int64
Compensation                    int64
OperInc                         int64
OperRev                         int64
BedRange_10-50                  uint8
BedRange_201-910                uint8
BedRange_51-200                 uint8
__Acting Director               uint8
__Regional Representative       uint8
__Safety Inspection Member      uint8
__State Board Representative    uint8
DonorType__City/County          uint8
DonorType__District             uint8
DonorType__Investor             uint8
DonorType__Non Profit           uint8
dtype: object
```

In [50]:
```
Model_OperRev= smf.ols(formula = 'OperRev ~ Teaching + DonorType',data = dfnew
1).fit()
Model_OperRev.summary()
```

Out[50]:

OLS Regression Results

Dep. Variable:	OperRev	**R-squared:**	0.592
Model:	OLS	**Adj. R-squared:**	0.585
Method:	Least Squares	**F-statistic:**	85.54
Date:	Sun, 08 Oct 2017	**Prob (F-statistic):**	4.41e-13
Time:	22:58:58	**Log-Likelihood:**	-1270.6
No. Observations:	61	**AIC:**	2545.
Df Residuals:	59	**BIC:**	2550.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	2.775e+08	2.61e+07	10.648	0.000	2.25e+08	3.3e+08
Teaching	5.004e+08	4.64e+07	10.793	0.000	4.08e+08	5.93e+08
DonorType	-2.228e+08	3.53e+07	-6.309	0.000	-2.93e+08	-1.52e+08

Omnibus:	36.140	**Durbin-Watson:**	1.191
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	126.160
Skew:	1.622	**Prob(JB):**	4.02e-28
Kurtosis:	9.254	**Cond. No.**	2.17e+17

In [51]:
```
Model_OperRev= smf.ols(formula = 'OperInc ~ Teaching + DonorType',data = dfnew
1).fit()
Model_OperRev.summary()
```

Out[51]:

OLS Regression Results

Dep. Variable:	OperInc	**R-squared:**	0.001
Model:	OLS	**Adj. R-squared:**	-0.016
Method:	Least Squares	**F-statistic:**	0.07846
Date:	Sun, 08 Oct 2017	**Prob (F-statistic):**	0.780
Time:	22:59:01	**Log-Likelihood:**	-1189.2
No. Observations:	61	**AIC:**	2382.
Df Residuals:	59	**BIC:**	2387.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-9.545e+05	6.86e+06	-0.139	0.890	-1.47e+07	1.28e+07
Teaching	-3.361e+06	1.22e+07	-0.275	0.784	-2.78e+07	2.11e+07
DonorType	2.406e+06	9.3e+06	0.259	0.797	-1.62e+07	2.1e+07

Omnibus:	54.386	**Durbin-Watson:**	1.320
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	355.347
Skew:	-2.348	**Prob(JB):**	6.88e-78
Kurtosis:	13.852	**Cond. No.**	2.17e+17