Evaluating Variance in Hospital Pricing

Shaan Penmetsa

DAT-03-20-17

Project Background

- What are some of the factors relating to the variation in hospital charges for similar services provided?
- Breaking down hospital charges and payments:
    - Hospitals have a master charge list--the total price tag for your care is the summed price of the services you receive
    - But, no one actually pays those rates:
        - Government payors (Medicare and Medicaid) pay fixed prices based on the patient diagnosis (DRG), with variation for geography and other hospital factors
        - Private payors negotiate discounts off the master charge list with each hospital
        - Even uninsured patients get a discount rate
- So, if no one actually pays the rates predicted in the model, how valuable can it be?

Project Background

- Other drawbacks:
  - Limits predictive power of model
  - Payors can negotiate very different rates and there is little price transparency when choosing insurance, or even choosing a hospital
  - Tough to measure quality in healthcare
- A great dataset to work with and provides some interesting insights nonetheless
- I hypothesize several different features will have an impact in predicting a higher Average Covered Charge per DRG: 1) Hospital for-profit status; 2) Hospital teaching status; 3) Proportion of patients as Medicare patients; 4) lower market concentration. What might be more telling that the features hypothesized to have an impact is the set of features that are not included: quality and outcomes measures for the hospital's patients.

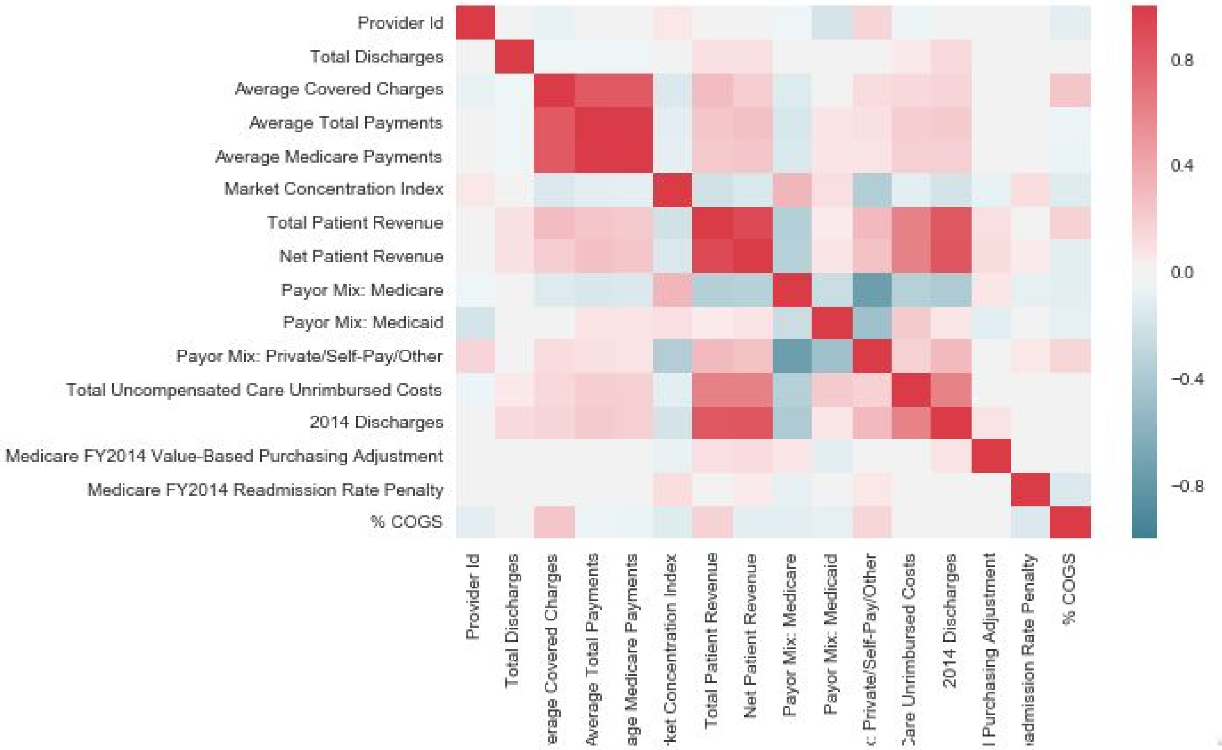
Datasets

- 2014 Medicare Provider Charge data
  - DRG - Provider combination is the index
  - Average Covered Charge: The sum of the prices of goods/services from the hospital’s master charge list that the hospital renders to treat the patient, per DRG
- Definitive Healthcare Hospital Data
  - Ownership
  - Academic Medical Center
  - Market Concentration Index
  - Payor Mix
  - Medicare FY2014 Value-Based Purchasing Adjustment
  - Medicare FY2014 Readmission Rate Penalty

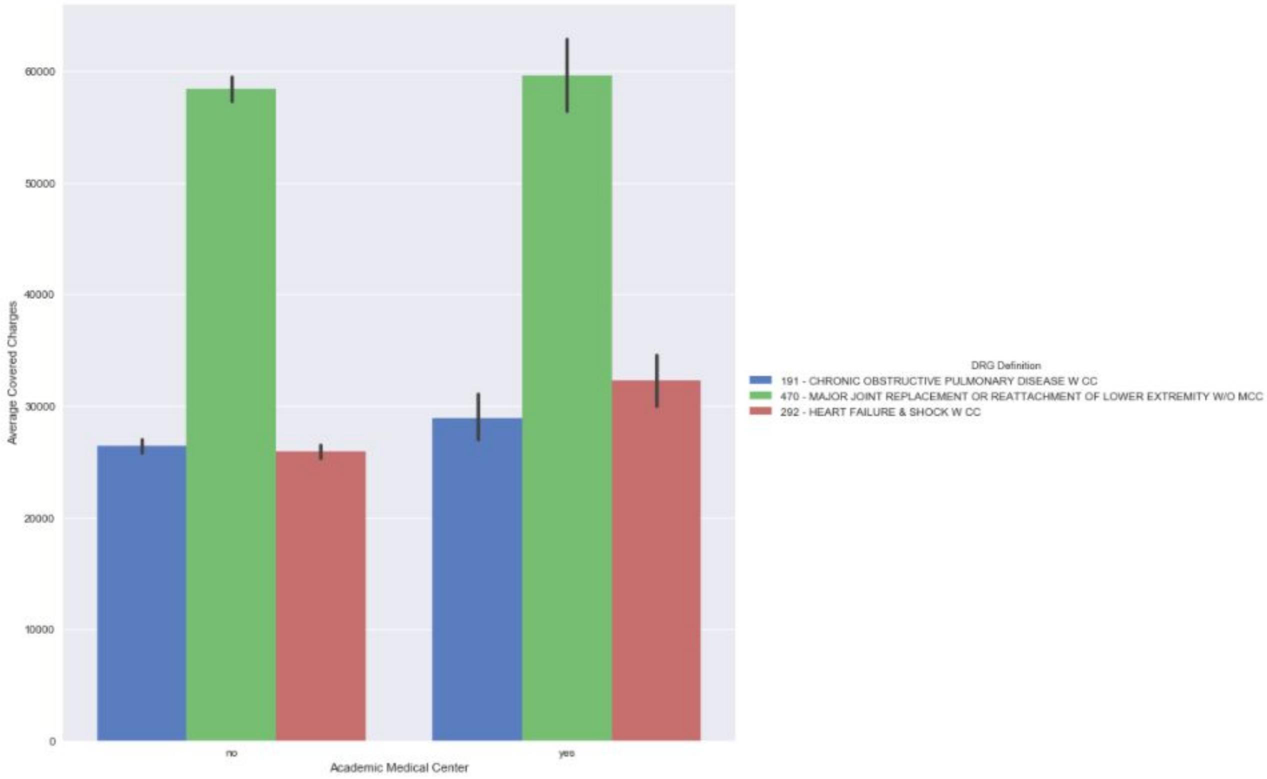
Merging the Datasets & Creating New Variables

- % COGS = (Gross Patient Revenue - Net Patient Revenue) / Gross Patient Revenue
- % Charge Difference: creating for each Provider a single metric to represent how much they "over-" or "under-" charge for treatment of DRGs
  - (Provider's Average Covered Charge for a DRG - Average Average Covered Charge for a DRG) / Average Average Covered Charge for a DRG
  - Average this % across all the DRGs a hospital treats, after dropping less common DRGs
  - Simpson's paradox...
- Many dummy variables were created

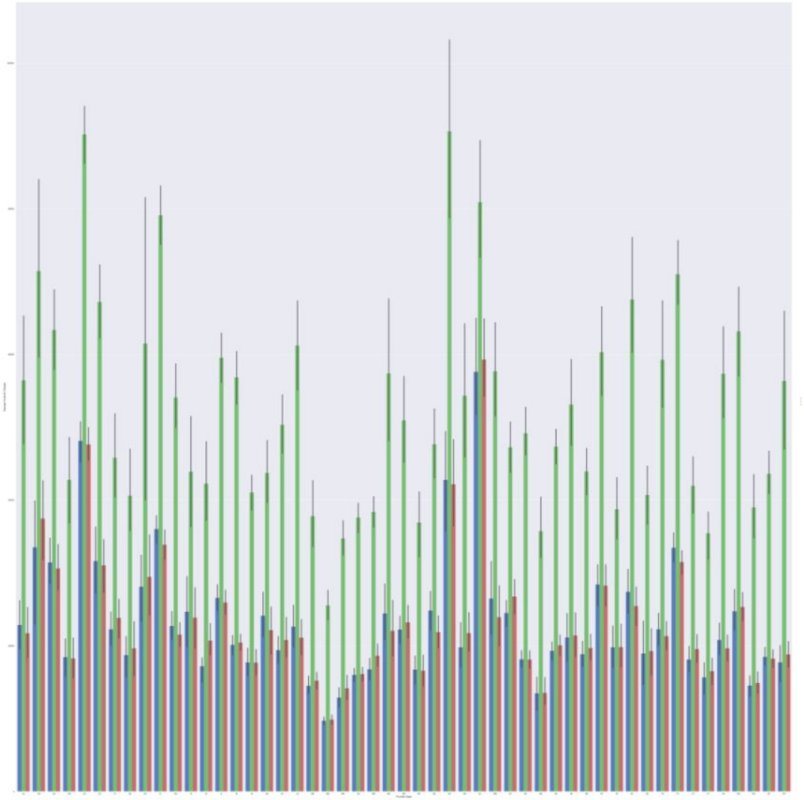
Correlation Matrix



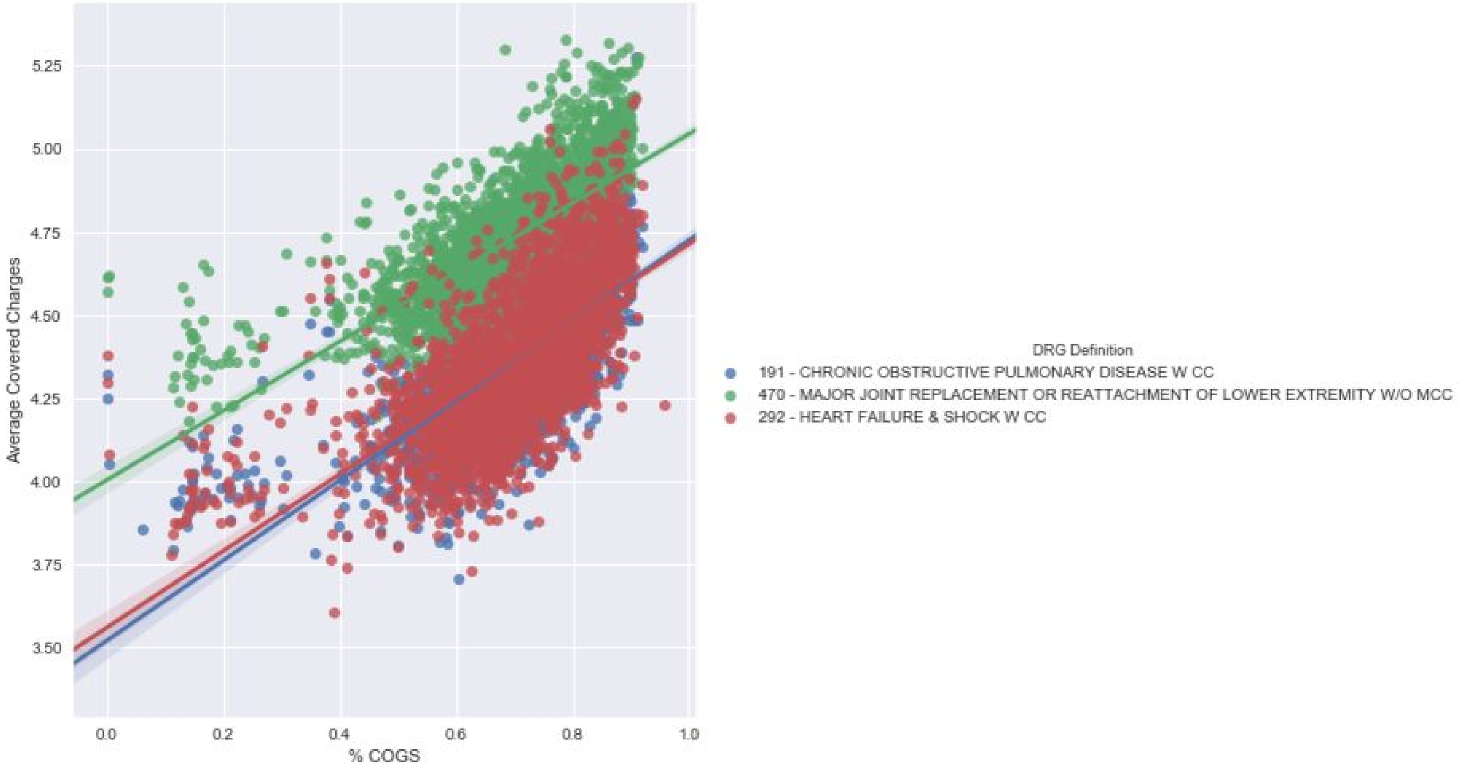
Visualizations for 3 Key DRGs



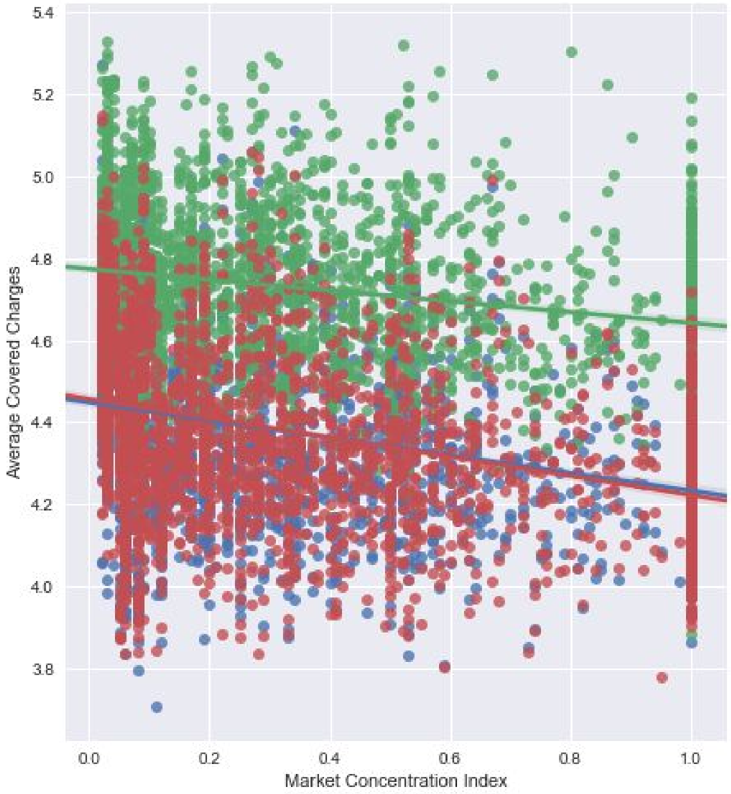
Visualizations for 3 Key DRGs - State



Visualizations for 3 Key DRGs - % COGS



Visualizations for 3 Key DRGs



Model Results - Aggregated Price Score

```
print gs_Ridge.best_score_
print
print gs_Ridge.best_estimator_
print
print zip(x.columns,gs_Ridge.best_estimator_.coef_)
```

```
0.420048144886

Ridge(alpha=100.0, copy_X=True, fit_intercept=True, max_iter=None,
   normalize=False, random_state=None, solver='auto', tol=0.001)

[('Market Concentration Index', -12.438047742780482), ('Payor Mix: Medicare', -19.576979
166842754), ('Payor Mix: Private/Self-Pay/Other', -2.185609612543026), ('2014 Discharge
s', 0.00026076800364352215), ('Medicare FY2014 Value-Based Purchasing Adjustment', 7.133
9973548149), ('Medicare FY2014 Readmission Rate Penalty', 0.29541056415666067), ('% COG
S', 222.09289895150448), ('state_AK', 36.080329102353367), ('state_AL', -15.640857785282
121), ('state_AR', -25.790578392293327), ('state_AZ', -11.099693948401738), ('state_CA',
62.816877248898578), ('state_CO', 11.903318507646013), ('state_CT', -2.692004684436202
2), ('state_DC', 7.3036587321501072), ('state_DE', 3.163351233666599), ('state_FL', -3.2
429390367701298), ('state_GA', -12.564002107052554), ('state_HI', 10.945308591479058),
 ('state_IA', -8.68824385524373), ('state_ID', 1.8970989532417182), ('state_IL', -4.1300
839881032241), ('state_IN', -7.1533378948327071), ('state_KS', -4.3980814364691554), ('s
tate_KY', -23.690194305779603), ('state_LA', -9.0740950467272175), ('state_MA', -15.1413
01739841818), ('state_ME', 4.9825030008629909), ('state_MI', -19.914304282973227), ('sta
te_MN', -0.58129027446590287), ('state_MO', -11.269173710215563), ('state_MS', -5.020858
8138865036), ('state_MT', 8.770053288146336), ('state_NC', -14.156087692824432), ('state
_ND', 0.35669878887895073), ('state_NE', 3.246588007076797), ('state_NH', -1.45164245098
37509), ('state_NJ', 56.690838812344296), ('state_NM', -2.5973845152980486), ('state_N
V', 7.1408886529685791), ('state_NY', 10.822470718883435), ('state_OH', -16.919358851486
759), ('state_OK', -16.133636309478181), ('state_OR', 7.3342628361532389), ('state_PA',
 6.6065178375433158), ('state_RI', -13.6674152365644), ('state_SC', -3.303058990709925
3), ('state_SD', 2.4263378670330118), ('state_TN', -29.342042155600446), ('state_TX', -
3.0109187885407791), ('state_UT', -1.8036243555067655), ('state_VA', -10.4919485077759
3), ('state_VT', 6.0974088567779781), ('state_WA', 8.6401160464928051), ('state_WI', 0.4
6589193718128585), ('state_WV', -15.029831015880296), ('own_Governmental - City', 24.041
746719896061), ('own_Governmental - City-County', -0.23472733962203857), ('own_Governmen
tal - County', -13.440528398558595), ('own_Governmental - Hospital District', -15.082153
443580863), ('own_Governmental - Other', -4.1433955986503221), ('own_Governmental - Stat
e', 3.9406592954062822), ('own_Proprietary - Corporation', 14.842859248437048), ('own_Pr
oprietary - Other', 0.20557902400456871), ('own_Proprietary - Partnership', 27.079680740
401955), ('own_Voluntary Nonprofit - Church', -8.7705265850285272), ('own_Voluntary Nonp
rofit - Other', -5.3169615993328385), ('sch_Yes', 8.8726712098075975), ('amc_yes', 12.00
0886416861926), ('aco_yes', 5.2279125352987483)]
```

Lasso would not converge when the predictor was the aggregated % difference.

Using Ridge, was left with all the features.

Notice the fairly high alpha.

Model Results - Single DRG Price

```
print gs_Lasso.best_score_
print
print gs_Lasso.best_estimator_
print
output2 = zip(x2.columns,gs_Lasso.best_estimator_.coef_)
labels2 = ['Feature','Coefficient']
Feature_Summary = pd.DataFrame.from_records(output2, columns=labels2)
print Feature_Summary
```

```
0.405031170384

Lasso(alpha=10.0, copy_X=True, fit_intercept=True, max_iter=1000,
   normalize=True, positive=False, precompute=False, random_state=None,
   selection='cyclic', tol=0.0001, warm_start=False)
```

	Feature	Coefficient
0	Total Discharges	-2.871504
1	Market Concentration Index	-0.000000
2	Payor Mix: Medicare	-0.000000
3	Payor Mix: Private/Self-Pay/Other	-0.000000
4	Medicare FY2014 Value-Based Purchasing Adjustment	22325.266131
5	Medicare FY2014 Readmission Rate Penalty	0.000000
6	% COGS	135738.546483
7	state_AK	19312.513815
8	state_AL	-10649.747174
9	state_AR	-13504.824317
10	state_AZ	-0.000000
11	state_CA	28219.845772
12	state_CO	8123.967489
13	state_CT	-1567.935207
14	state_DC	0.000000
15	state_DE	0.000000
16	state_FL	0.000000
17	state_GA	-4639.146165
18	state_HI	0.000000
19	state_IA	-506.803613
20	state_ID	-0.000000
21	state_IL	-0.000000
22	state_IN	1958.558734
23	state_KS	-5435.534763
24	state_KY	-6801.027025
25	state_LA	0.000000
26	state_MA	-1745.612704
27	state_ME	0.000000
28	state_MI	-7359.568634
29	state_MN	0.000000

Running Lasso for just one DRG, thereby avoiding Simpson's paradox (chose hip/knee replacement, a common and pricey surgery).

Left with 40 features (mostly the different states).

R-squared = .405

Model Results - Feature Importance

	Feature	Coefficient	abs Scaled Coefficient
	% COGS	135738.546483	14103.804283
	state_CA	28219.845772	7610.307892
	own_Proprietary - Corporation	15786.618455	5690.679644
	state_TN	-13619.257568	1650.495293
	state_WA	10605.613188	1500.358361
	state_NJ	8974.844037	1498.797816
	state_MI	-7359.568634	1327.754712
	state_AR	-13504.824317	1302.285391
	amc_yes	4121.154918	1291.704755
	state_AL	-10649.747174	1290.625241
	state_MO	-6838.070719	1131.244705
	state_CO	8123.967489	1036.912986
	own_Governmental - Hospital District	-5444.343164	989.960328
	state_KY	-6801.027025	882.159059
	own_Governmental - City	9232.099392	836.721285
	state_AK	19312.513815	785.380651
	state_OH	-3468.617195	780.960734
	state_GA	-4639.146165	760.123721
	state_TX	3037.781551	752.682381
	state_OK	-6559.592428	668.350394
	state_KS	-5435.534763	568.050070
	sch_Yes	1828.385258	536.047112
	Total Discharges	-2.871504	525.807211
	state_RI	-7350.948718	516.920304
	own_Governmental - State	4086.226172	512.928463
	own_Proprietary - Partnership	8211.010503	430.845967
	state_WV	-3944.389452	369.108083
	state_IN	1958.558734	336.094767
	state_MS	-3071.167076	312.918180
	state_MA	-1745.612704	271.725909
	state_OR	1846.823155	206.801905
	state_CT	-1567.935207	179.298873
	state_NH	957.044573	80.794209
	own_Proprietary - Other	1024.447667	75.915103
	state_SC	571.414981	72.933282
	Medicare FY2014 Value-Based Purchasing Adjustment	22325.266131	62.363806
	own_Governmental - County	-344.051024	61.079780
	state_IA	-506.803613	56.750400
19	state_MT	584.964629	41.134839
20	aco_yes	72.600269	35.513386

Scaled features and re-ran Lasso. Created two Pandas dataframes on feature coefficients in model and merged.

- A 1% increase in % COGS results in a $1,357 increase in surgery price
- Having surgery in California (as opposed to Wyoming) results in a $28,219 increase in surgery price
- Having surgery at a for-profit hospital run by a corporation (as opposed to a individually owned for-profit hospital) results in a $15,786 increase in surgery price

Model Results - OLS

```
from sklearn import cross_validation
kf = cross_validation.KFold(len(data_jr), n_folds=10, shuffle=True)

z = data_jr[Features2.Feature]

lm = smf.ols(formula='y2 ~ stats.zscore(z)', data=data_jr).fit()
lm.summary()
```

OLS Regression Results

Dep. Variable:	y2	**R-squared:**	0.627
Model:	OLS	**Adj. R-squared:**	0.619
Method:	Least Squares	**F-statistic:**	74.41
Date:	Wed, 24 May 2017	**Prob (F-statistic):**	0.00
Time:	08:26:06	**Log-Likelihood:**	-20153.
No. Observations:	1811	**AIC:**	4.039e+04
Df Residuals:	1770	**BIC:**	4.061e+04
Df Model:	40		
Covariance Type:	nonrobust		

Using features identified from Lasso.

R-squared = 0.627

Model Results - Feature Importance II

	Feature	Coefficient	abs Scaled Coefficient
	% COGS	135738.546483	14103.804283
	state_CA	28219.845772	7610.307892
	own_Proprietary - Corporation	15786.618455	5690.679644
	state_TN	-13619.257568	1650.495293
	state_WA	10605.613188	1500.358361
	state_NJ	8974.844037	1498.797816
	state_MI	-7359.568634	1327.754712
	state_AR	-13504.824317	1302.285391
	amc_yes	4121.154918	1291.704755
	state_AL	-10649.747174	1290.625241
	state_MO	-6838.070719	1131.244705
	state_CO	8123.967489	1036.912986
	own_Governmental - Hospital District	-5444.343164	989.960328
	state_KY	-6801.027025	882.159059
	own_Governmental - City	9232.099392	836.721285
	state_AK	19312.513815	785.380651
	state_OH	-3468.617195	780.960734
	state_GA	-4639.146165	760.123721
	state_TX	3037.781551	752.682381
	state_OK	-6559.592428	668.350394
	state_KS	-5435.534763	568.050070
	sch_Yes	1828.385258	536.047112
	Total Discharges	-2.871504	525.807211
	state_RI	-7350.948718	516.920304
	own_Governmental - State	4086.226172	512.928463
	own_Proprietary - Partnership	8211.010503	430.845967
	state_WV	-3944.389452	369.108083
	state_IN	1958.558734	336.094767
	state_MS	-3071.167076	312.918180
	state_MA	-1745.612704	271.725909
	state_OR	1846.823155	206.801905
	state_CT	-1567.935207	179.298873
	state_NH	957.044573	80.794209
	own_Proprietary - Other	1024.447667	75.915103
	state_SC	571.414981	72.933282
	Medicare FY2014 Value-Based Purchasing Adjustment	22325.266131	62.363806
	own_Governmental - County	-344.051024	61.079780
	state_IA	-506.803613	56.750400
19	state_MT	584.964629	41.134839
20	aco_yes	72.600269	35.513386

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	5.919e+04	391.448	151.219	0.000	5.84e+04 6e+04
stats.zscore(z)[0]	1.472e+04	506.811	29.041	0.000	1.37e+04 1.57e+04
stats.zscore(z)[1]	7846.3225	423.011	18.549	0.000	7016.670 8675.975
stats.zscore(z)[2]	5896.4041	445.394	13.239	0.000	5022.851 6769.957
stats.zscore(z)[3]	-2151.7334	399.934	-5.380	0.000	-2936.126 -1367.341
stats.zscore(z)[4]	1969.0574	400.668	4.914	0.000	1183.226 2754.889
stats.zscore(z)[5]	1718.7000	409.751	4.194	0.000	915.052 2522.348
stats.zscore(z)[6]	-1772.5837	402.734	-4.401	0.000	-2562.469 -982.699
stats.zscore(z)[7]	-1769.6286	395.823	-4.471	0.000	-2545.959 -993.299
stats.zscore(z)[8]	1915.5227	423.108	4.527	0.000	1085.679 2745.367
stats.zscore(z)[9]	-1751.5300	405.036	-4.324	0.000	-2545.928 -957.132
stats.zscore(z)[10]	-1688.6779	400.760	-4.214	0.000	-2474.690 -902.666
stats.zscore(z)[11]	1365.1082	396.842	3.440	0.001	586.779 2143.437
stats.zscore(z)[12]	-1513.6844	403.158	-3.755	0.000	-2304.401 -722.968
stats.zscore(z)[13]	-1376.1674	397.401	-3.463	0.001	-2155.592 -596.742
stats.zscore(z)[14]	1487.0127	402.377	3.696	0.000	697.828 2276.197
stats.zscore(z)[15]	1208.0739	392.984	3.074	0.002	437.313 1978.835
stats.zscore(z)[16]	-1319.4707	408.669	-3.229	0.001	-2120.996 -517.945
stats.zscore(z)[17]	-1234.7030	403.354	-3.061	0.002	-2025.803 -443.603
stats.zscore(z)[18]	911.4704	419.552	2.172	0.030	88.601 1734.340
stats.zscore(z)[19]	-1133.5494	396.327	-2.860	0.004	-1910.867 -356.231
stats.zscore(z)[20]	-1114.9561	397.203	-2.807	0.005	-1893.991 -335.921
stats.zscore(z)[21]	1273.9006	427.576	2.979	0.003	435.294 2112.507
stats.zscore(z)[22]	-1039.1941	413.468	-2.513	0.012	-1850.132 -228.256
stats.zscore(z)[23]	-999.1872	395.003	-2.530	0.012	-1773.909 -224.465
stats.zscore(z)[24]	824.6130	414.443	1.990	0.047	11.764 1637.462
stats.zscore(z)[25]	758.9312	394.727	1.923	0.055	-15.249 1533.112
stats.zscore(z)[26]	-897.7537	400.924	-2.239	0.025	-1684.087 -111.420
stats.zscore(z)[27]	749.9851	403.057	1.861	0.063	-40.532 1540.503
stats.zscore(z)[28]	-715.1815	400.636	-1.785	0.074	-1500.951 70.588
stats.zscore(z)[29]	-695.5869	415.361	-1.675	0.094	-1510.236 119.062
stats.zscore(z)[30]	639.1269	401.648	1.591	0.112	-148.628 1426.882
stats.zscore(z)[31]	-643.7815	397.016	-1.622	0.105	-1422.450 134.887
stats.zscore(z)[32]	454.3489	394.951	1.150	0.250	-320.270 1228.968
stats.zscore(z)[33]	525.7381	394.334	1.333	0.183	-247.672 1299.148
stats.zscore(z)[34]	408.3478	401.234	1.018	0.309	-378.594 1195.290
stats.zscore(z)[35]	565.0579	405.564	1.393	0.164	-230.376 1360.492
stats.zscore(z)[36]	-439.3672	411.964	-1.067	0.286	-1247.355 368.620
stats.zscore(z)[37]	-537.9039	398.602	-1.349	0.177	-1319.683 243.876
stats.zscore(z)[38]	455.3900	398.033	1.144	0.253	-325.275 1236.055
stats.zscore(z)[39]	312.0659	416.680	0.749	0.454	-505.171 1129.303

Next Steps

- Creating a better Aggregated Charge feature
  - Weighting based on volume per DRG
- Train model on each DRG one at a time
  - Include more features on the Provider-DRG level
- Evaluate differences in costs across hospitals
  - Related to higher quality or more amenities?
- Link to Notebook