

Apply Machine Learning to Capstone

Approach

- Explore Avg. Provider Charges for 1 DRG (039) by the following factors:
 - Overall pop
 - 65+ year old pop
 - Pop density
 - # Discharges

Explore Correlations



Overall pop and 65+ pop are colinear

Population

As Overall Population Increased, So Did Charges

```
Call:
lm(formula = Provider_Avg_Covered_Charges ~ Overall_Pop_Thousands,
    data = .)
```

Residuals:

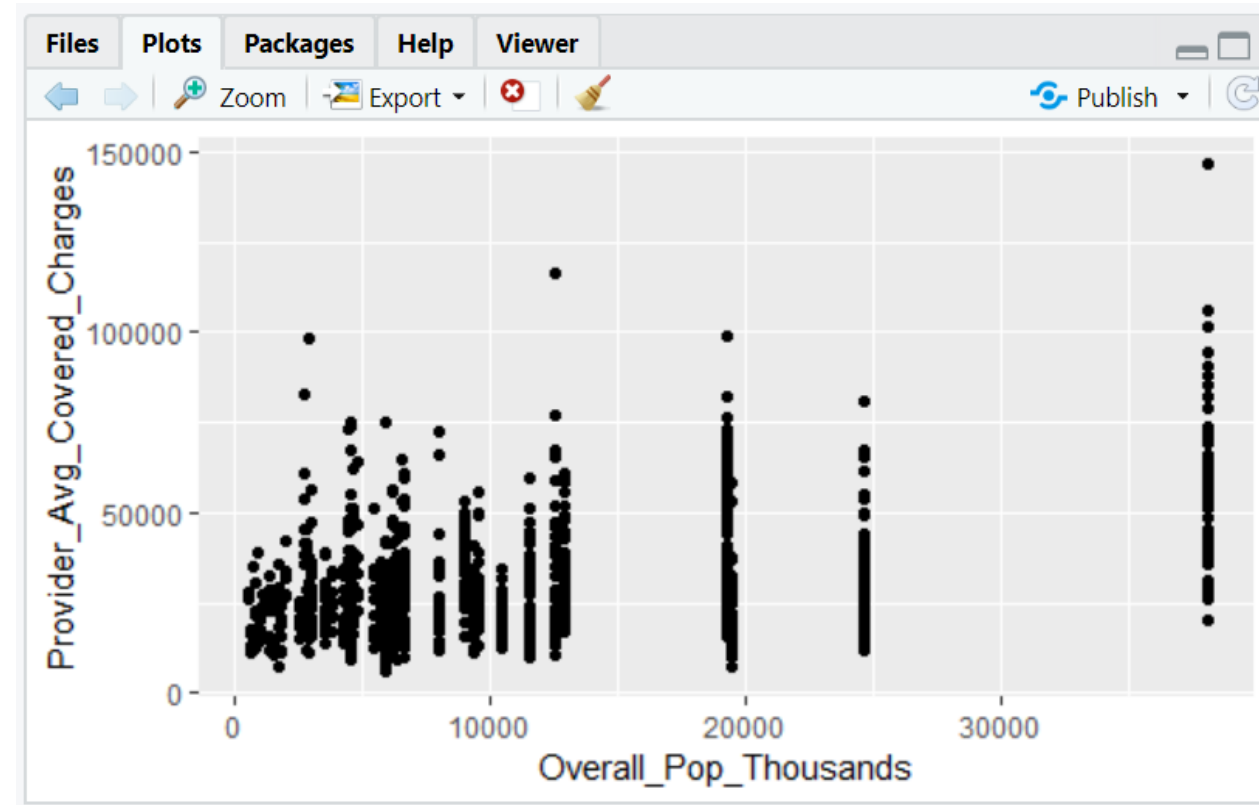
Min	1Q	Median	3Q	Max
-29135	-9545	-2501	6558	97529

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.217e+04	6.971e+02	31.80	<2e-16 ***
Overall_Pop_Thousands	7.143e-01	4.671e-02	15.29	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14440 on 1074 degrees of freedom
Multiple R-squared: 0.1788, Adjusted R-squared: 0.178
F-statistic: 233.8 on 1 and 1074 DF, p-value: < 2.2e-16



65+ Year Old Pop Shows Similar Trend to Overall Pop Because of Collinearity

```
call:
lm(formula = Provider_Avg_Covered_Charges ~ Sum_65_Plus_Thousands,
    data = .)
```

Residuals:

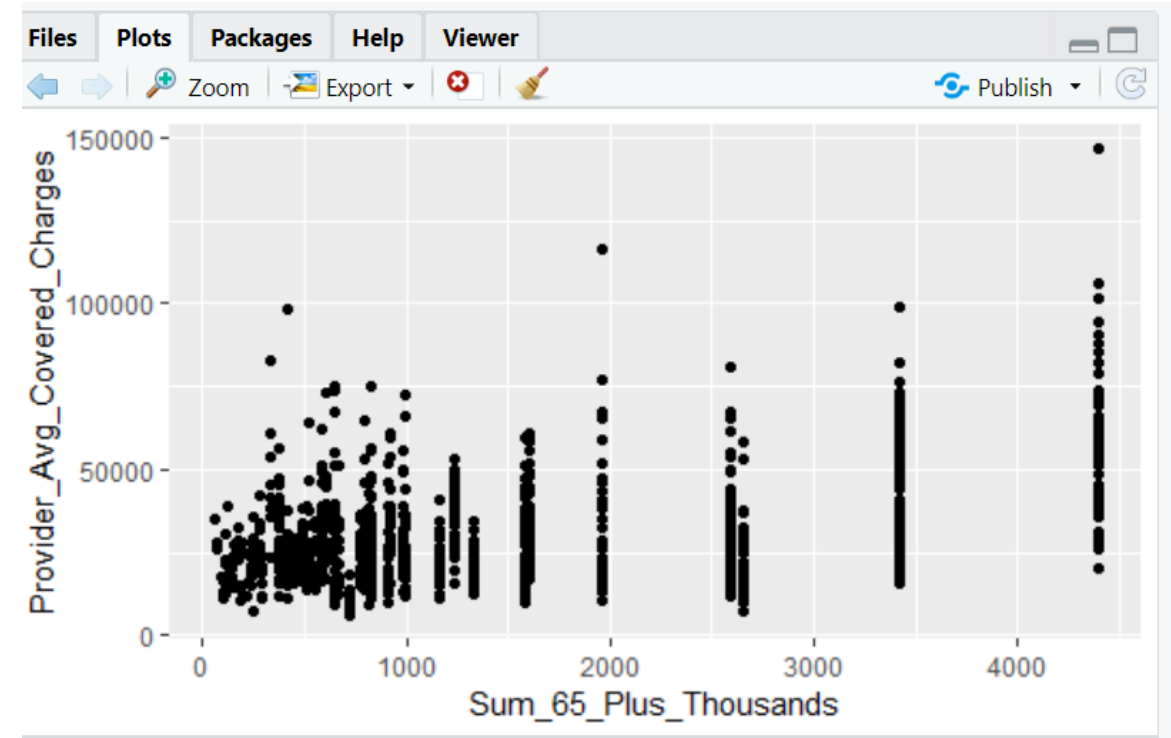
Min	1Q	Median	3Q	Max
-29750	-9282	-2412	6541	99353

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.150e+04	7.118e+02	30.20	<2e-16 ***
Sum_65_Plus_Thousands	5.928e+00	3.726e-01	15.91	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14340 on 1074 degrees of freedom
Multiple R-squared: 0.1908, Adjusted R-squared: 0.19
F-statistic: 253.2 on 1 and 1074 DF, p-value: < 2.2e-16



No Association Between Charges and Pop Density

```
Call:
lm(formula = Provider_Avg_Covered_Charges ~ Pop_Per_Sq_Mi, data = .)
```

Residuals:

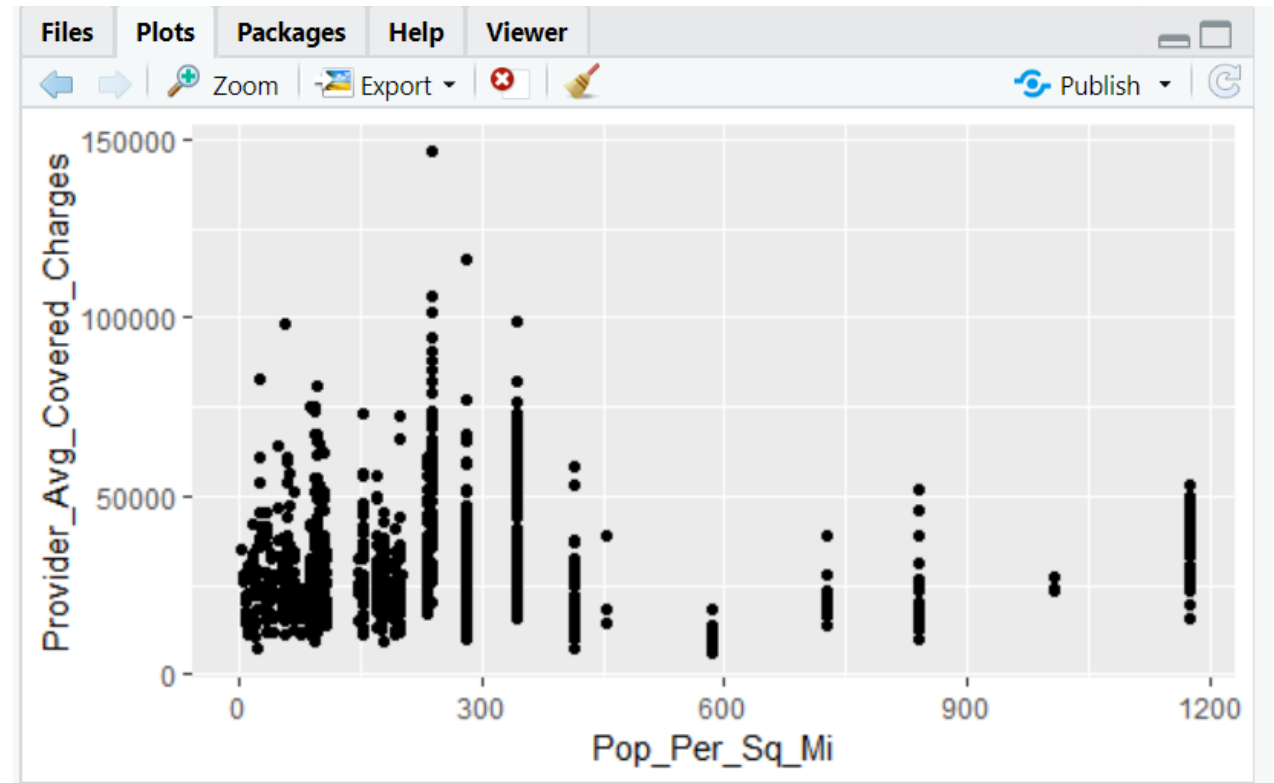
Min	1Q	Median	3Q	Max
-24666	-10670	-3825	6137	116457

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.029e+04	6.882e+02	44.012	<2e-16 ***
Pop_Per_Sq_Mi	6.118e-01	2.063e+00	0.296	0.767

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15940 on 1074 degrees of freedom
Multiple R-squared: 8.185e-05, Adjusted R-squared: -0.0008492
F-statistic: 0.08791 on 1 and 1074 DF, p-value: 0.7669



Discharges

As Discharges Go Up, Cost Goes Down

```
Call:
lm(formula = Provider_Avg_Covered_Charges ~ Provider_Total_Discharges,
    data = .)
```

Residuals:

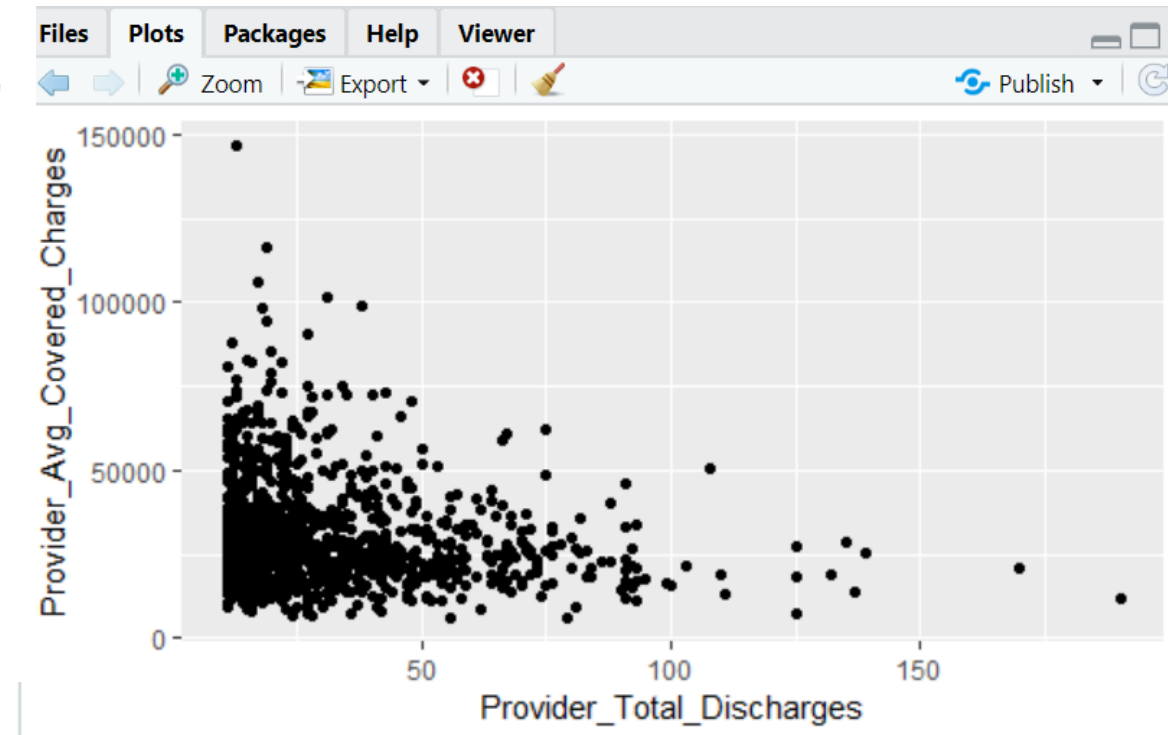
Min	1Q	Median	3Q	Max
-24564	-9978	-3575	6363	114172

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	34353.11	832.35	41.273	< 2e-16	***
Provider_Total_Discharges	-125.63	21.83	-5.754	1.14e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15700 on 1074 degrees of freedom
Multiple R-squared: 0.0299, Adjusted R-squared: 0.029
F-statistic: 33.11 on 1 and 1074 DF, p-value: 1.137e-08



Summary

Factor	Units	Estimate	P-value	Adjusted R2	F-Statistic	Residual Standard Error
Pop Overall	Thousands	0.71	***	0.178	233	14,440
Pop: 65+ Years	Thousands	5.9	***	0.19	253	14,340
Pop Density	People_Per_Sq_Mi	0.61		-0.0008	0.09	15,940
# Discharges	1	-125	***	0.029	33	15,700