

# Analytics with **Alteryx**: Classifying **Medicare** Claims **Fraud**



*IDS.506 - Robert Minh Duc Bui - mbui7@uic.edu*

# Executive Summary



- We can perform ingestion, engineering, and modeling using **Alteryx** to classify potential **fraudulent** providers based on aggregated **claims** metrics to **Medicare**.
- Final models at-a-glance:
  - **Logistic Regression** or **Linear Support Vector Machine**
  - (Implicit) dummy variables & derived numeric features
  - Hyperparameter tuning
- Models **Performance Metrics** compared to **assuming all fraud**:
  - **GLM:**

- Positive Class Accuracy:	<b>Acc_Yes</b>	=	<b>0.6122</b>	(+145% lift)
- F1 Measure:	<b>F1</b>	=	<b>0.6742</b>	(+68.5% lift)
  - **L-SVM:**

- Positive Class Accuracy:	<b>Acc_Yes</b>	=	<b>0.6122</b>	(+145% lift)
- F1 Measure:	<b>F1</b>	=	<b>0.6977</b>	(+74.4% lift)

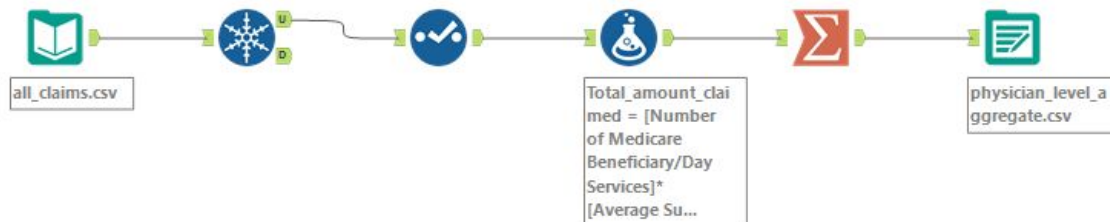
# Research **Mission**



- To develop a model that:
  - Uses aggregated claims data from medical providers to:
    - **Classify** fraudulent providers.
- Pitfalls:
  - Small, shortened dataset.



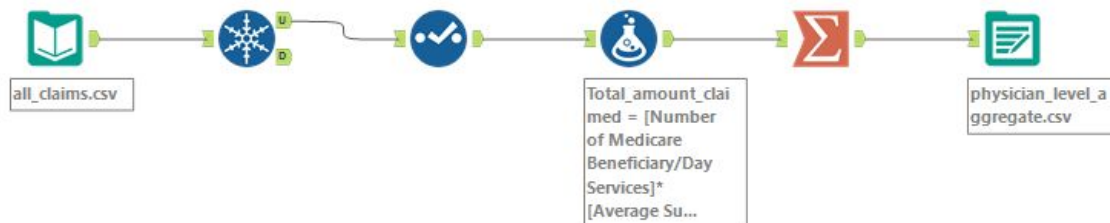
# Data Processing Workflow



Output Column	Data Preview
Total_amount_claimed	5509.52 $[\text{Number of Medicare Beneficiary/Day Services}] * [\text{Average Submitted Charge Amount}]$
Total_amount_paid	918.56 $[\text{Number of Medicare Beneficiary/Day Services}] * [\text{Average Medicare Payment Amount}]$
Total_amount_allowed	1263.55 $[\text{Number of Medicare Beneficiary/Day Services}] * [\text{Average Medicare Allowed Amount}]$
Payout_Ratio	0.1667223 $[\text{Average Medicare Payment Amount}] / [\text{Average Submitted Charge Amount}]$

Allowance_Ratio	0.2293394 $[\text{Average Medicare Allowed Amount}] / [\text{Average Submitted Charge Amount}]$
Final_amount_received	4590.96 $[\text{Total\_amount\_claimed}] - [\text{Total\_amount\_paid}]$
Excess_amount_claimed	4245.97 $[\text{Total\_amount\_claimed}] - [\text{Total\_amount\_allowed}]$

# Data Processing Workflow #2



Field	Action	Output Field Name
National Provid...	Group By	National Provider Identifier
Gender	Group By	Gender
Provider Type	Group By	Provider Type
HCPCS Code	Count	Count_HCPCS
Payout_Ratio	Average	Avg_Payout_Ratio
Allowance_Ratio	Average	Avg_Allowance_Ratio
Final_amount_r...	Average	Avg_Final_amount_received
Excess_amount_...	Average	Avg_Excess_amount_claimed
Number of Med...	Average	Avg_Number of Medicare Beneficiaries
Number of Med...	Average	Avg_Number of Medicare Beneficiary/Day Services
Total_amount_cl...	Sum	Sum_Total_amount_claimed
Total_amount_p...	Sum	Sum_Total_amount_paid
Total_amount_a...	Sum	Sum_Total_amount_allowed
Fraud_Label	Group By	Fraud_Label
Payout_Ratio	Min	Min_Payout_Ratio
Payout_Ratio	Max	Max_Payout_Ratio
Allowance_Ratio	Min	Min_Allowance_Ratio
Allowance_Ratio	Max	Max_Allowance_Ratio

# Student's T-tests

**Welch's Two Sample t-test(s) of Avg\_Number.of.Medicare.Beneficiary.Day.Services by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-3.8642	295.5	0.00013703

**Welch's Two Sample t-test(s) of Sum\_Total\_amount\_allowed by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-5.20062	276.5	3.8706e-07

**Welch's Two Sample t-test(s) of Min\_Payout\_Ratio by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	2.43704	464.01	0.015183

**Welch's Two Sample t-test(s) of Max\_Payout\_Ratio by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-9.3885	453.48	2.9665e-19

**Welch's Two Sample t-test(s) of Min\_Allowance\_Ratio by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	0.818269	408.87	0.41368

**Welch's Two Sample t-test(s) of Avg\_Number.of.Medicare.Beneficiaries by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-3.06344	345.56	0.0023599

**Welch's Two Sample t-test(s) of Avg\_Excess\_amount\_claimed by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	0.5137	453.75	0.60771

**Welch's Two Sample t-test(s) of Avg\_Final\_amount\_received by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-0.432438	434.34	0.66564

**Welch's Two Sample t-test(s) of Avg\_Allowance\_Ratio by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-6.58843	525.6	1.0813e-10

**Welch's Two Sample t-test(s) of Avg\_Payout\_Ratio by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-4.83207	581.13	1.7312e-06

## Student's **T-tests** #2

**Welch's Two Sample t-test(s) of Count\_HCPCS by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-3.90556	398.16	0.00011041

**Welch's Two Sample t-test(s) of Sum\_Total\_amount\_claimed by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-2.47924	282.57	0.01375

**Welch's Two Sample t-test(s) of Sum\_Total\_amount\_paid by Fraud\_Label**

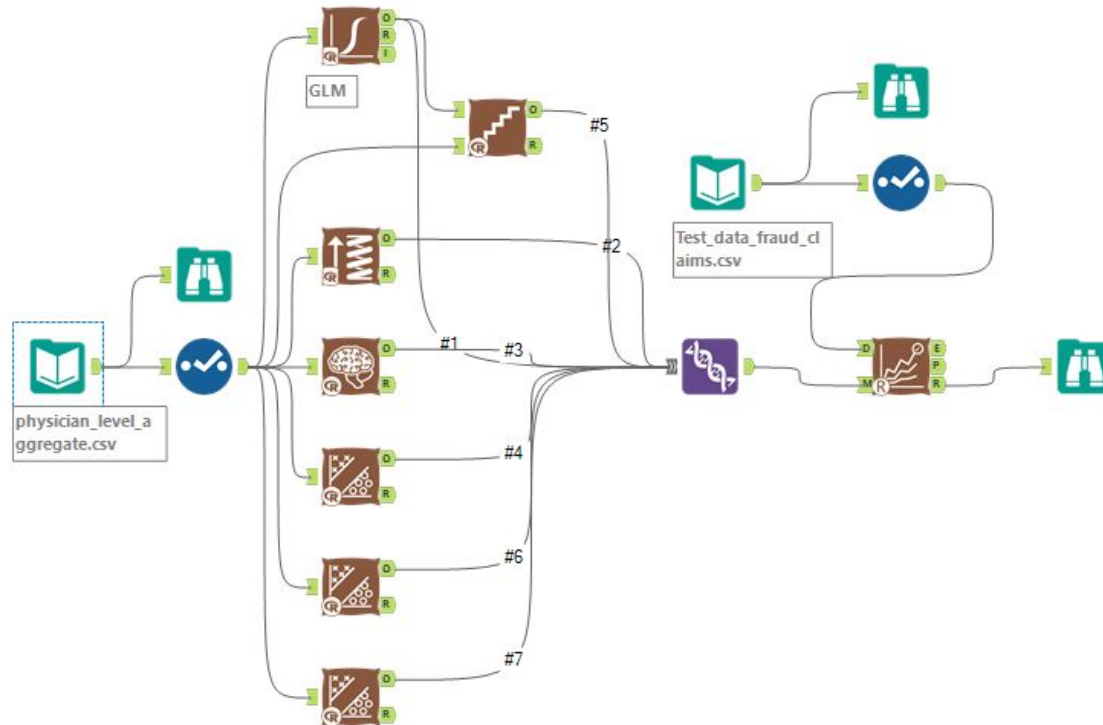
Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-4.94088	275.82	1.3506e-06

**Welch's Two Sample t-test(s) of Max\_Allowance\_Ratio by Fraud\_Label**

Test	t-Statistic	Degrees of Freedom	p-Value
No vs Yes	-11.3646	524.21	6.6151e-27



# Models Workflow



# Models Comparison Report

## Model Comparison Report

### Fit and error measures

Model	Accuracy	Accuracy_No	Accuracy_Yes	F1	AUC
GLM	0.7958	0.8925	0.6122	0.6742	0.8113
xGB	0.7606	0.8602	0.5714	0.6222	0.7874
SLP	0.6761	0.9570	0.1429	0.2333	0.6193
Polynom_SVM	0.6690	0.9892	0.0612	0.1132	0.7275
Stepwise_GLM	0.7958	0.8925	0.6122	0.6742	0.8122
Radial_SVM	0.7324	0.9032	0.4082	0.5128	0.8183
Linear_SVM	0.8169	0.9247	0.6122	0.6977	0.8510

**Model:** model names in the current comparison.

**Accuracy:** overall accuracy, number of correct predictions of all classes divided by total sample number.

**Accuracy\_[class name]:** accuracy of Class [class name] is defined as the number of cases that are **correctly** predicted to be Class [class name] divided by the total number of cases that actually belong to Class [class name], this measure is also known as *recall*.

**AUC:** area under the ROC curve, only available for two-class classification.

**F1:** F1 score,  $2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$ . The *precision* measure is the percentage of actual members of a class that were predicted to be in that class divided by the total number of cases predicted to be in that class. In situations where there are three or more classes, average precision and average recall values across classes are used to calculate the F1 score.

# Models Comparison Report #2

Confusion matrix of GLM		
	Actual_No	Actual_Yes
Predicted_No	83	19
Predicted_Yes	10	30

Confusion matrix of Linear_SVM		
	Actual_No	Actual_Yes
Predicted_No	86	19
Predicted_Yes	7	30

Confusion matrix of Polynom_SVM		
	Actual_No	Actual_Yes
Predicted_No	92	46
Predicted_Yes	1	3

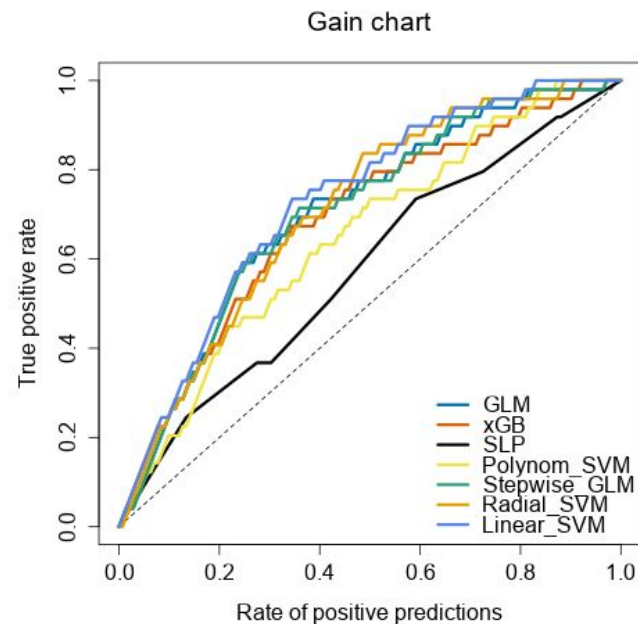
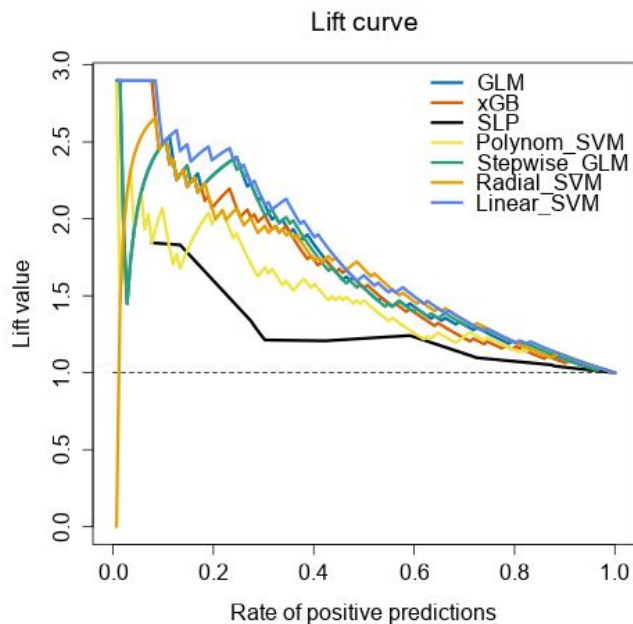
Confusion matrix of Radial_SVM		
	Actual_No	Actual_Yes
Predicted_No	84	29
Predicted_Yes	9	20

Confusion matrix of SLP		
	Actual_No	Actual_Yes
Predicted_No	89	42
Predicted_Yes	4	7

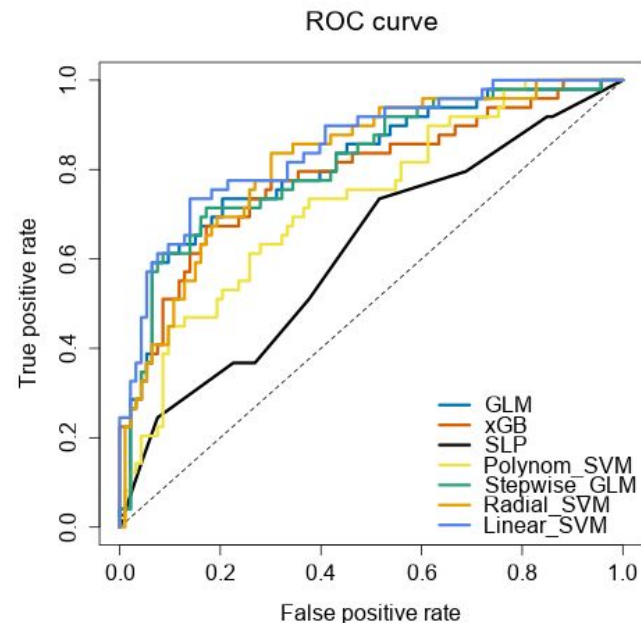
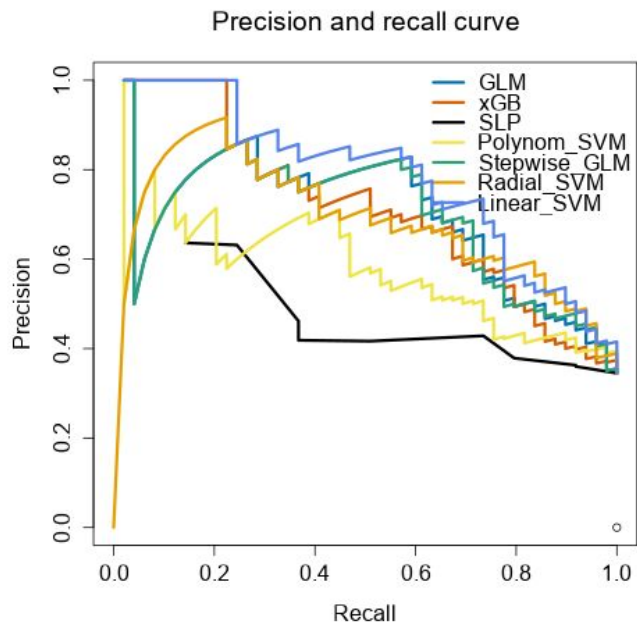
Confusion matrix of Stepwise_GLM		
	Actual_No	Actual_Yes
Predicted_No	83	19
Predicted_Yes	10	30

Confusion matrix of xGB		
	Actual_No	Actual_Yes
Predicted_No	80	21
Predicted_Yes	13	28

# Models Comparison Report #3



# Models Comparison Report #4



# Coefficients: Logistic Regression

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.778e+00	1.138e+00	-2.442171	0.0146 *
Avg_Allowance_Ratio	7.020e+00	4.443e+00	1.580175	0.11407
Avg_Excess_amount_claimed	2.247e-06	4.258e-05	0.052778	0.95791
Avg_Final_amount_received	-7.588e-06	3.777e-05	-0.200897	0.84078
Avg_Number.of.Medicare.Beneficiary.Day.Services	1.228e-03	5.492e-04	2.235927	0.02536 *
Avg_Payout_Ratio	-5.771e+00	5.510e+00	-1.047387	0.29492
Count_HCPCS	-2.264e-02	1.954e-02	-1.158837	0.24652
GenderM	1.869e+00	2.632e-01	7.101246	1.23e-12 ***
Max_Allowance_Ratio	1.832e+00	2.068e+00	0.885823	0.37571
Max_Payout_Ratio	-2.618e-01	2.092e+00	-0.125109	0.90044
Min_Allowance_Ratio	-1.045e+00	2.739e+00	-0.381371	0.70293
Min_Payout_Ratio	-2.885e+00	3.656e+00	-0.789156	0.43002
Provider.TypeAnesthesiologist Assistants	-1.658e+01	3.956e+03	-0.004191	0.99666
Provider.TypeAnesthesiology	-1.992e+00	1.195e+00	-1.667080	0.0955 .
Provider.TypeAudiologist (billing independently)	-1.595e+01	1.346e+03	-0.011845	0.99055
Provider.TypeCRNA	-1.593e+00	1.466e+00	-1.086647	0.27719
Provider.TypeCardiac Surgery	-1.693e+01	3.956e+03	-0.004279	0.99659
Provider.TypeCardiology	-1.414e+00	1.183e+00	-1.195635	0.23184
Provider.TypeCertified Clinical Nurse Specialist	-1.635e+01	3.956e+03	-0.004134	0.9967
Provider.TypeChiropractic	-1.774e+00	1.101e+00	-1.611867	0.10699
Provider.TypeClinical Psychologist	-1.538e+00	1.299e+00	-1.183302	0.23669
Provider.TypeCritical Care (Intensivists)	-1.813e+01	2.099e+03	-0.008635	0.99311
Provider.TypeDermatology	-5.066e+00	2.228e+00	-2.273555	0.02299 *
Provider.TypeDiagnostic Radiology	-2.715e+00	1.563e+00	-1.737201	0.08235 .
Provider.TypeEmergency Medicine	-1.371e+00	1.109e+00	-1.235625	0.2166
Provider.TypeEndocrinology	-1.377e+00	1.573e+00	-0.875316	0.3814
Provider.TypeFamily Practice	-6.270e-01	1.032e+00	-0.607629	0.54343
Provider.TypeGastroenterology	-1.575e+00	1.310e+00	-1.202184	0.22929
Provider.TypeGeneral Practice	1.945e+00	1.618e+00	1.202284	0.22925
Provider.TypeGeneral Surgery	-3.051e+00	1.452e+00	-2.101154	0.03563 *
Provider.TypeGeriatric Medicine	-9.611e-01	1.571e+00	-0.611614	0.54079
Provider.TypeHand Surgery	-1.619e+01	3.956e+03	-0.004093	0.99673
Provider.TypeHematology/Oncology	-2.376e+00	1.589e+00	-1.495639	0.13475
Provider.TypeHospice and Palliative Care	-1.527e+01	3.956e+03	-0.003859	0.99692
Provider.TypeInfectious Disease	-3.152e-01	1.742e+00	-0.180950	0.85641
Provider.TypeInternal Medicine	-6.995e-01	1.028e+00	-0.680393	0.49626

Provider.TypeInterventional Pain Management	1.556e+01	1.492e+03	0.010430	0.99168
Provider.TypeLicensed Clinical Social Worker	-1.765e+00	1.224e+00	-1.442027	0.14929
Provider.TypeNephrology	-1.934e+00	1.498e+00	-1.291099	0.19667
Provider.TypeNeurology	-1.359e+00	1.227e+00	-1.108224	0.26776
Provider.TypeNeurosurgery	-2.792e+00	1.580e+00	-1.767537	0.07714 .
Provider.TypeNuclear Medicine	1.498e+01	3.956e+03	0.003788	0.99698
Provider.TypeNurse Practitioner	-1.221e+00	1.133e+00	-1.077918	0.28107
Provider.TypeObstetrics/Gynecology	-8.025e-01	1.237e+00	-0.648608	0.51659
Provider.TypeOccupational therapist	-2.050e+00	2.194e+00	-0.934373	0.35011
Provider.TypeOphthalmology	-3.907e+00	1.326e+00	-2.946096	0.00322 **
Provider.TypeOptometry	-1.830e+01	8.143e+02	-0.022470	0.98207
Provider.TypeOral Surgery (dentists only)	1.764e+01	3.956e+03	0.004460	0.99644
Provider.TypeOrthopedic Surgery	-1.364e+00	1.177e+00	-1.159098	0.24642
Provider.TypeOsteopathic Manipulative Medicine	-1.515e+01	3.956e+03	-0.003830	0.99694
Provider.TypeOtolaryngology	-1.254e+00	1.354e+00	-0.926067	0.35441
Provider.TypePain Management	-6.979e-01	2.557e+00	-0.273004	0.78485
Provider.TypePathology	-1.507e+00	1.505e+00	-1.001131	0.31676
Provider.TypePhysical Medicine and Rehabilitation	-5.448e-01	1.171e+00	-0.465316	0.64171
Provider.TypePhysical Therapist	-2.027e+00	1.131e+00	-1.792080	0.07312 .
Provider.TypePhysician Assistant	-1.880e+00	1.195e+00	-1.573478	0.11561
Provider.TypePlastic and Reconstructive Surgery	1.663e+01	2.529e+03	0.006576	0.99475
Provider.TypePodiatry	-5.380e-01	1.128e+00	-0.476918	0.63342
Provider.TypePsychiatry	-2.323e-01	1.093e+00	-0.212578	0.83166
Provider.TypePulmonary Disease	-1.864e+01	1.847e+03	-0.010093	0.99195
Provider.TypeRadiation Oncology	-2.669e-01	1.880e+00	-0.141971	0.8871
Provider.TypeRegistered Dietician/Nutrition Professional	-1.428e+01	2.278e+03	-0.006267	0.995
Provider.TypeRheumatology	6.932e-01	1.991e+00	0.348233	0.72767
Provider.TypeSpeech Language Pathologist	1.921e+01	3.956e+03	0.004855	0.99613
Provider.TypeSurgical Oncology	-1.553e+01	3.956e+03	-0.003925	0.99687
Provider.TypeThoracic Surgery	-7.201e-02	1.751e+00	-0.041128	0.96719
Provider.TypeUrology	-1.820e+01	1.309e+03	-0.013904	0.98891
Provider.TypeVascular Surgery	-1.784e+01	3.956e+03	-0.004510	0.9964
Sum_Total_amount_allowed	5.140e-05	2.236e-05	2.298838	0.02151 *
Sum_Total_amount_claimed	3.968e-07	9.087e-07	0.436696	0.66233
Sum_Total_amount_paid	-6.392e-05	2.871e-05	-2.226054	0.02601 *