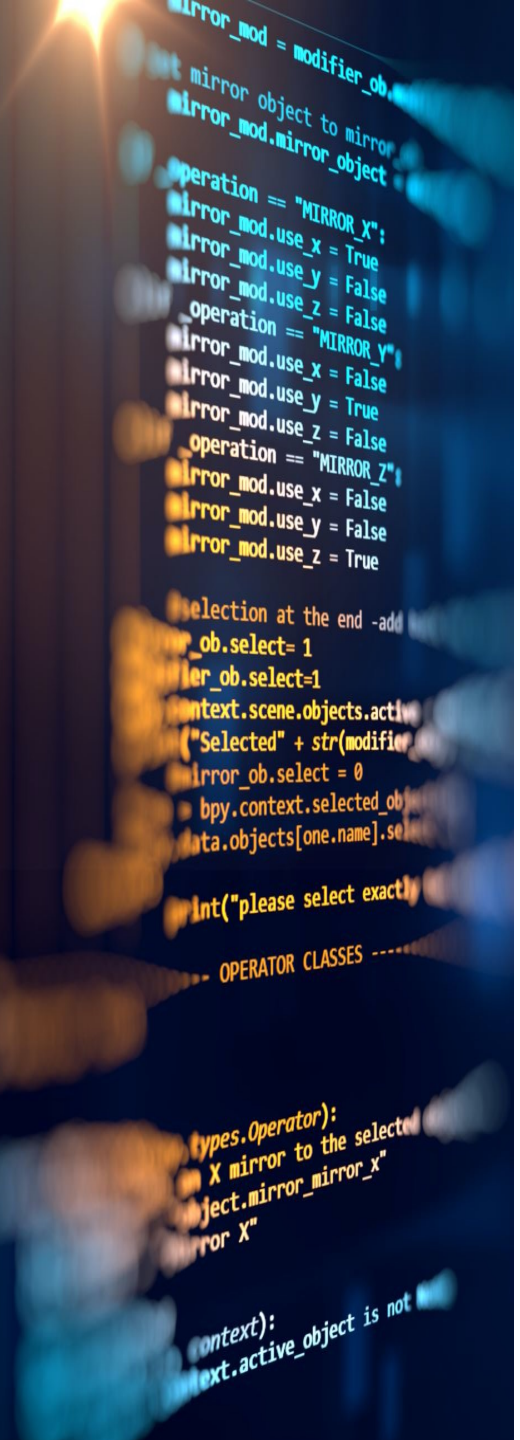


Detect Healthcare Fraud with Machine Learning

Jan Ruffner
Team StoneCap

Motivation

Top Healthcare Fraud Takedowns of 2020

(Courtesy: iStock)

HEALTHCARE FRAUD

Texas Doctor Accused of Defrauding $3.2 Million From Taxpayer-Funded Healthcare Program

A clinic allegedly offered TRICARE members $50 gift cards in exchange for urine and DNA to run fraudulent tests.

BY WILL MADDOX | PUBLISHED IN HEALTHCARE BUSINESS | APRIL 9, 2021 | 9:00 AM

4 DOCTORS FOUND GUILTY IN $150M HEALTHCARE FRAUD SCHEME

A federal judge found four Detroit, Michigan-area doctors guilty for participating in a Medicare fraud scheme that cost the federal healthcare program $150 million.

According to the February **announcement**, the doctors practiced at Tri-County Group where they would bill Medicare for medically unnecessary services, including facet joint injections, urinary drug screens, and home healthcare. Specifically, the doctors would administer the back injections in exchange for prescriptions of over 6.6 million doses of medically unnecessary opioids, the Justice Department reported.

Tri-County received higher rates for facet joint injections than any other medical clinic in the US at the time, and the four doctors all ranked in the top 25 doctors for dollars paid by Medicare for the injections despite working just a few hours a week, according law enforcement authorities.

$681M SUBSTANCE ABUSE TREATMENT FRAUD CASE IN FLORIDA

The Justice Department **announced** in July charges in a years-long healthcare scheme involving substance abuse treatment in Palm Beach County, Florida.

The department stated that Michael J. Ligotti, DO, was charged with conspiracy to commit healthcare fraud and wire fraud after allegedly engaging in fraudulent billing for tests and treatments involving patients seeking drug and/or alcohol addiction help.

According to the criminal complaint, Ligotti authorized "standing orders" for hundreds of millions of dollars in medically unnecessary urinalysis tests as the owner of Whole Health in Delray Beach, Florida. Ligotti also reportedly paid kickbacks to sober homes or addiction treatment facilities in exchange for having their patients treated by Whole Health providers.

The scheme resulted in $681 million in fraudulent laboratory testing claims. Some patients were also billed between $10,000 and $20,000 by Ligotti and Whole Health for a single day's visit, the Justice Department stated.

TEXAS PHYSICIAN SENTENCED 84 MONTHS FOR HEALTHCARE FRAUD

In an October **announcement**, the Justice Department shared that a Houston, Texas-area physician and anesthesiologist at two registered pain clinics, Texas Pain Solutions and Integra Medical Clinic, received seven years in prison for his role in fraudulently billing healthcare programs.

According to the evidence, Rezik Saqer would lure vulnerable patients to the clinics by prescribing opioids, then provide unnecessary and potentially dangerous procedures and tests, many of which were performed by unlicensed staff.

Seqer fraudulently billed healthcare providers for nearly $5 million for these services, which also resulted in multiple patient deaths, the Justice Department reported.

In addition to the prison sentence, the judge also required Seqer to pay $5 million in restitution.

FEDS CHARGE 10 INDIVIDUALS IN $1.4B RURAL HOSPITAL BILLING SCHEME

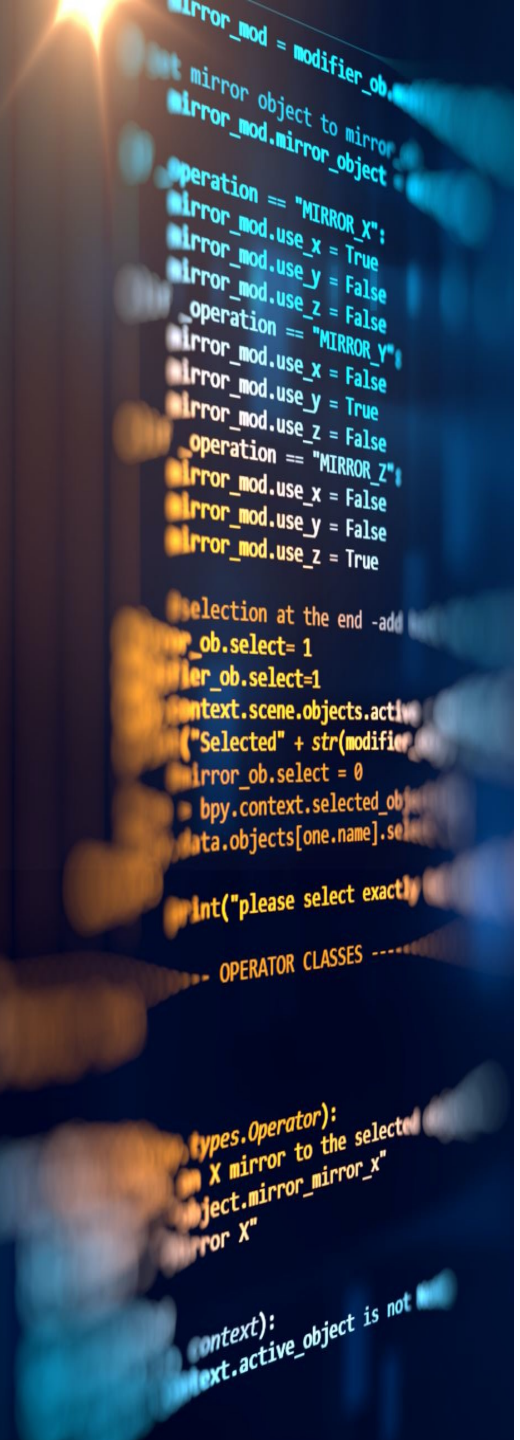
One of the largest healthcare fraud takedowns in 2020 involved charges against 10 individuals, including hospital managers, laboratory owners, billers, and recruiters, for their alleged involvement in a pass-through billing scheme using struggling rural hospitals in the South.

According to the indictment **announced** in June, the conspirators billed private payers approximately $1.4 billion for laboratory testing claims, out of which they were paid about $400 million.

The scheme allegedly involved the takeover of small, rural hospitals that were often experiencing financial hardships and using them as a means to bill private payers for expensive urinalysis drug and blood tests that were largely performed at outside laboratories despite claims stating they were done in-house.

The conspirators also reportedly negotiated higher reimbursement rates for tests done within the rural hospitals versus outside laboratories.

All defendants have been charged with conspiracy to commit healthcare fraud and wire fraud.



Players in the Healthcare System

Health Care Consumer
- Consumes healthcare services
- Pays premium for insurance
- Pays deductible
- After paying the deductible and the premium there is no incentive to consume less healthcare services

Physicians
- Provides the healthcare services
- Gets paid a certain amount from the patient
- Earns more if he provides expensive services

Incentives lead to high healthcare costs

Insurance Company:
- Pays for healthcare service if consumer is insured
- Gets paid premium for insurance
- Has strong incentive to keep costs under control

Tries to keep costs under control
One measure is: Find potential fraud

Dataset

Dataset

Dataset 1/2: Inpatient Data/Outpatient Data

- Claim ID, Claim Start Date, Claim End Date
- Provider, Attending Physician, Operating Physician, Other Physicion
- Claim Diagnosis Code
- Claim Procedure Code

Dataset 3: Beneficiary Data

- Date of Birth, DOD, Gender, Race, State, County
- Chronic Condition: Alzheimer, Heart failure, Kidney Disease, Cancer, Obstr Pulmonary, Depression, Diabetes, Ischemic Heart, Osteoporasis, Rheumatoidarthritis, Stroke
- Annual Reimbursement Amt, IP Annual Deductible, OP Annual Reimbursement, OP Annual Deductible

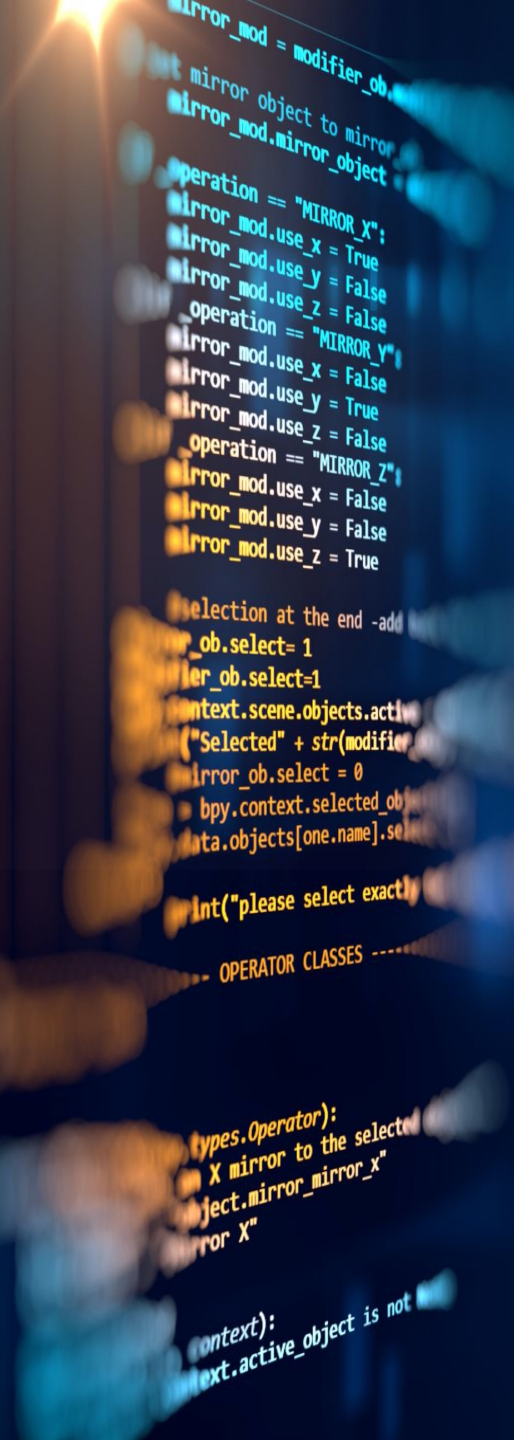
Dataset 4: Potential Fraud Data

- Provider
- Potential Fraud (Yes/No)

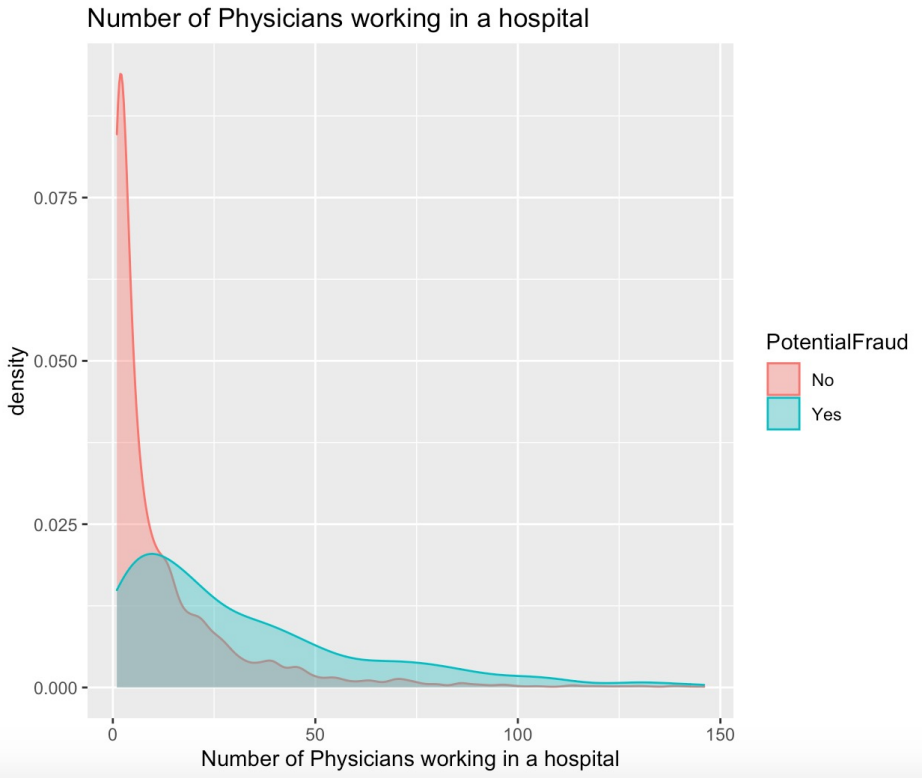
Steps to combine datasets:

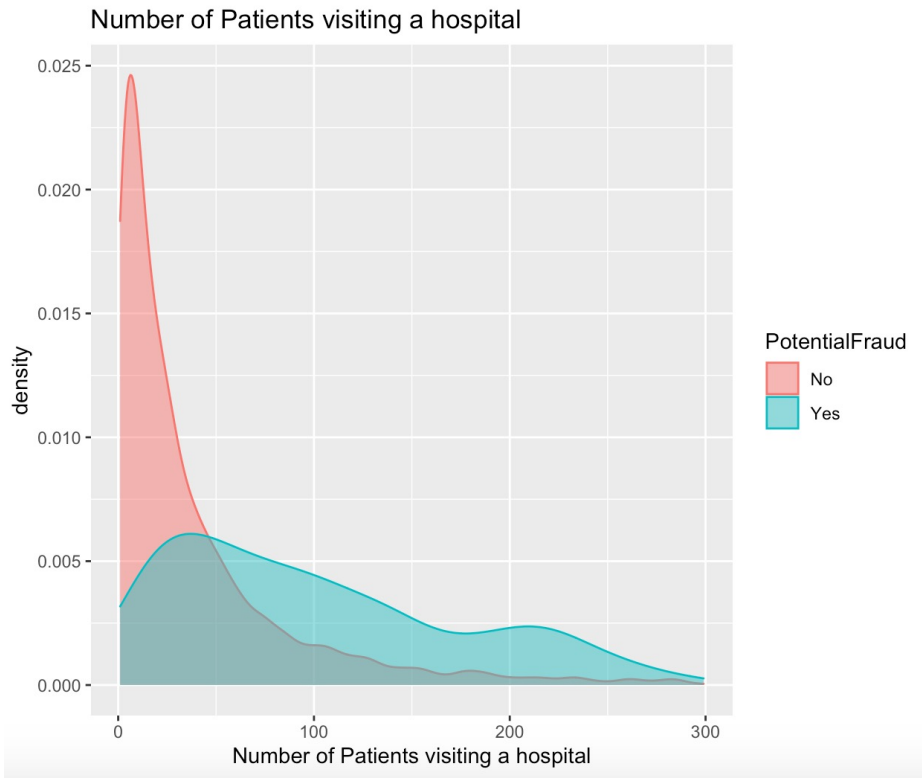
1. Combine four datasets – provides data on claim level (558’221 claims)
2. Aggregate the dataset on the provider level (5’410 provider)

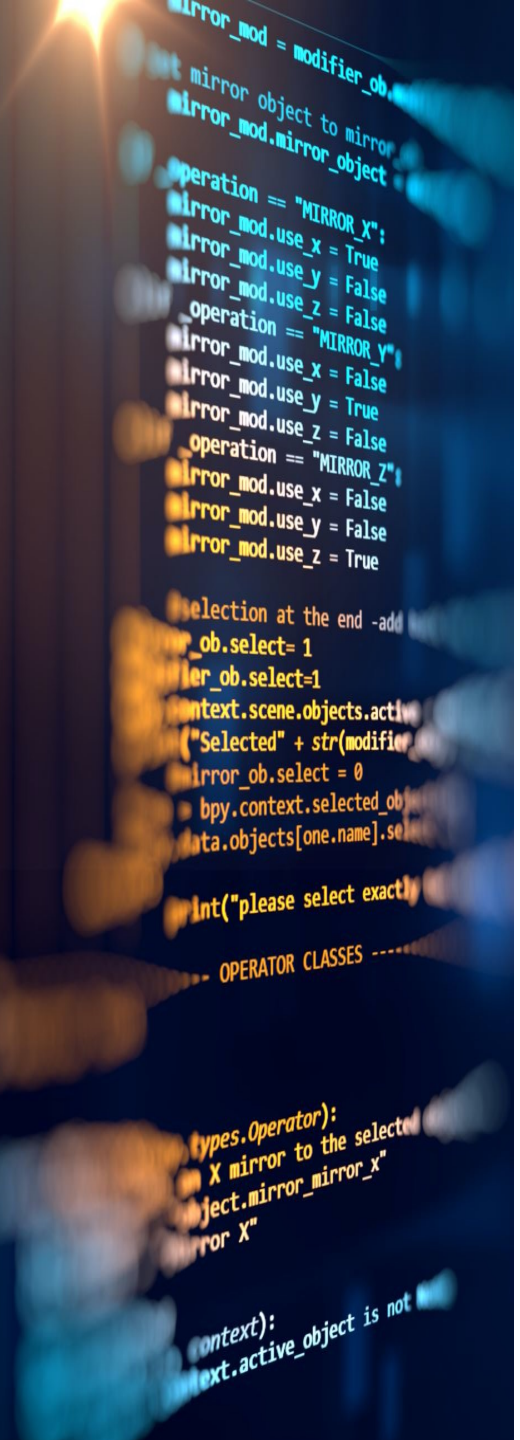
Explanatory Data Analysis



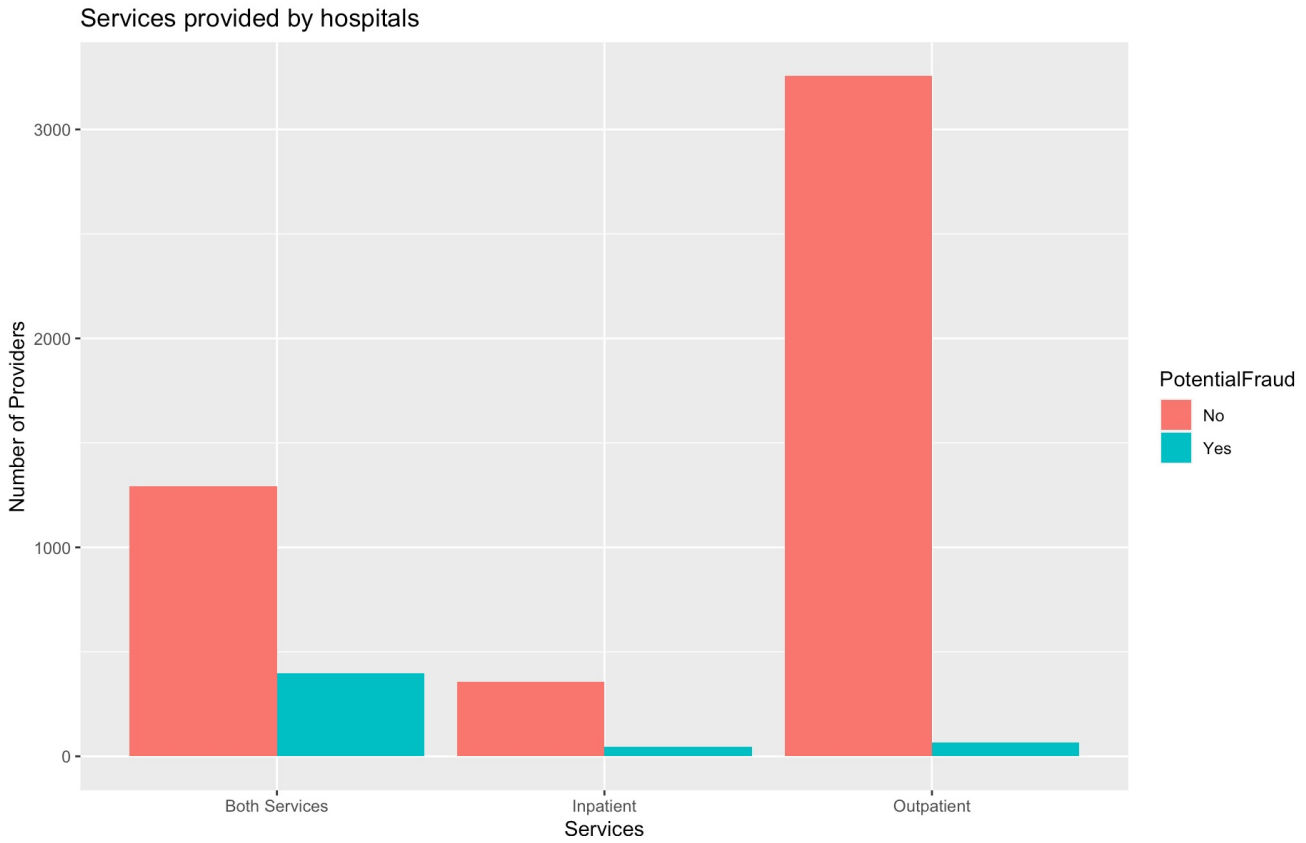
Number of Physicians or Patients in a hospital







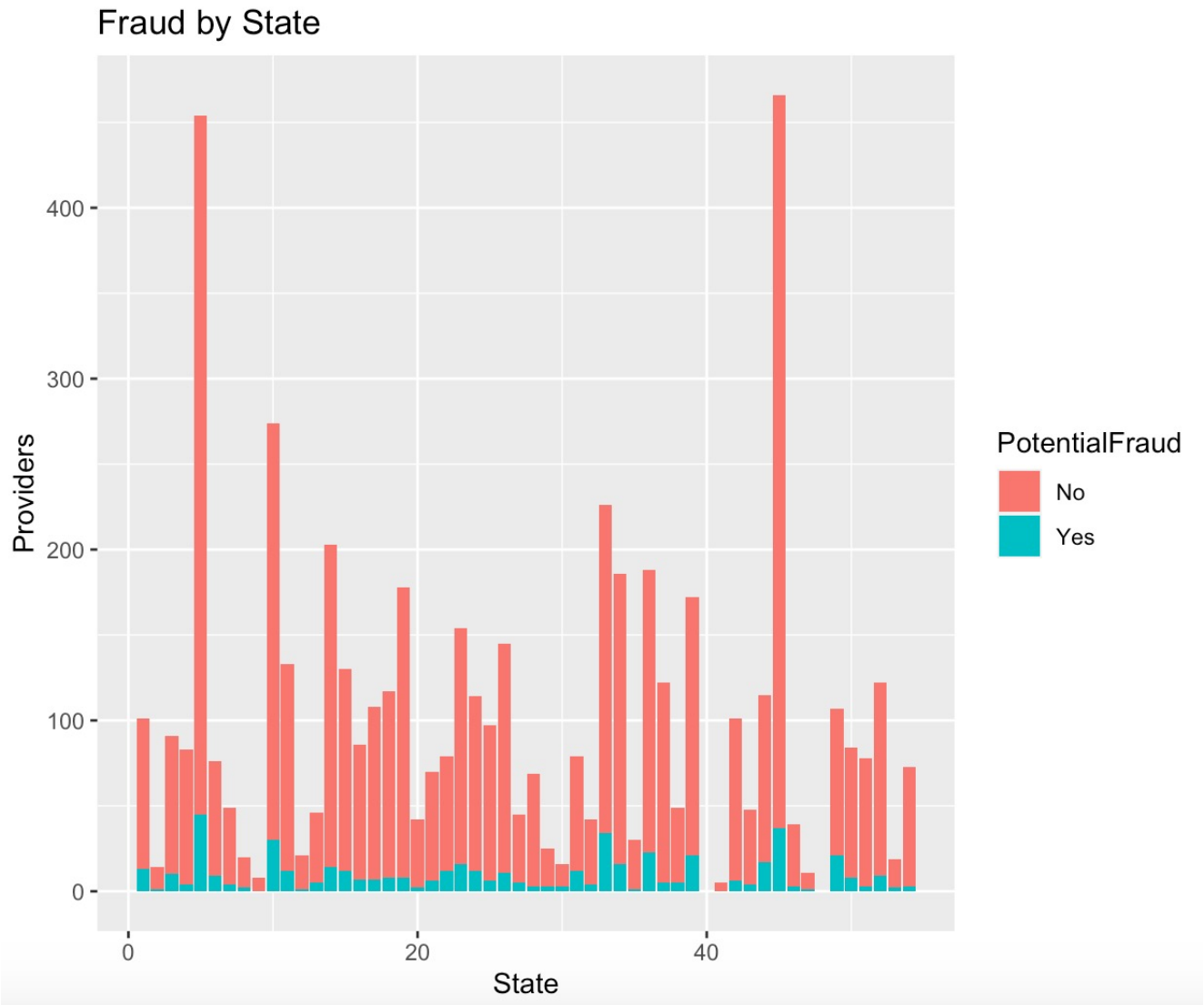
Type of Service and Fraud

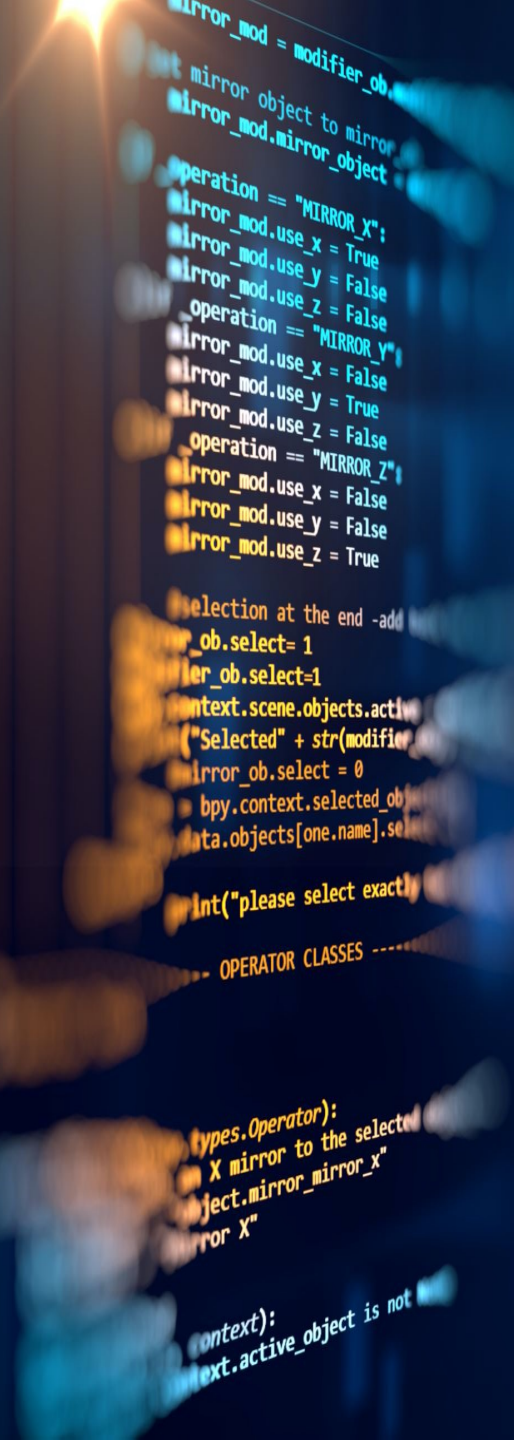


Potential Fraud:

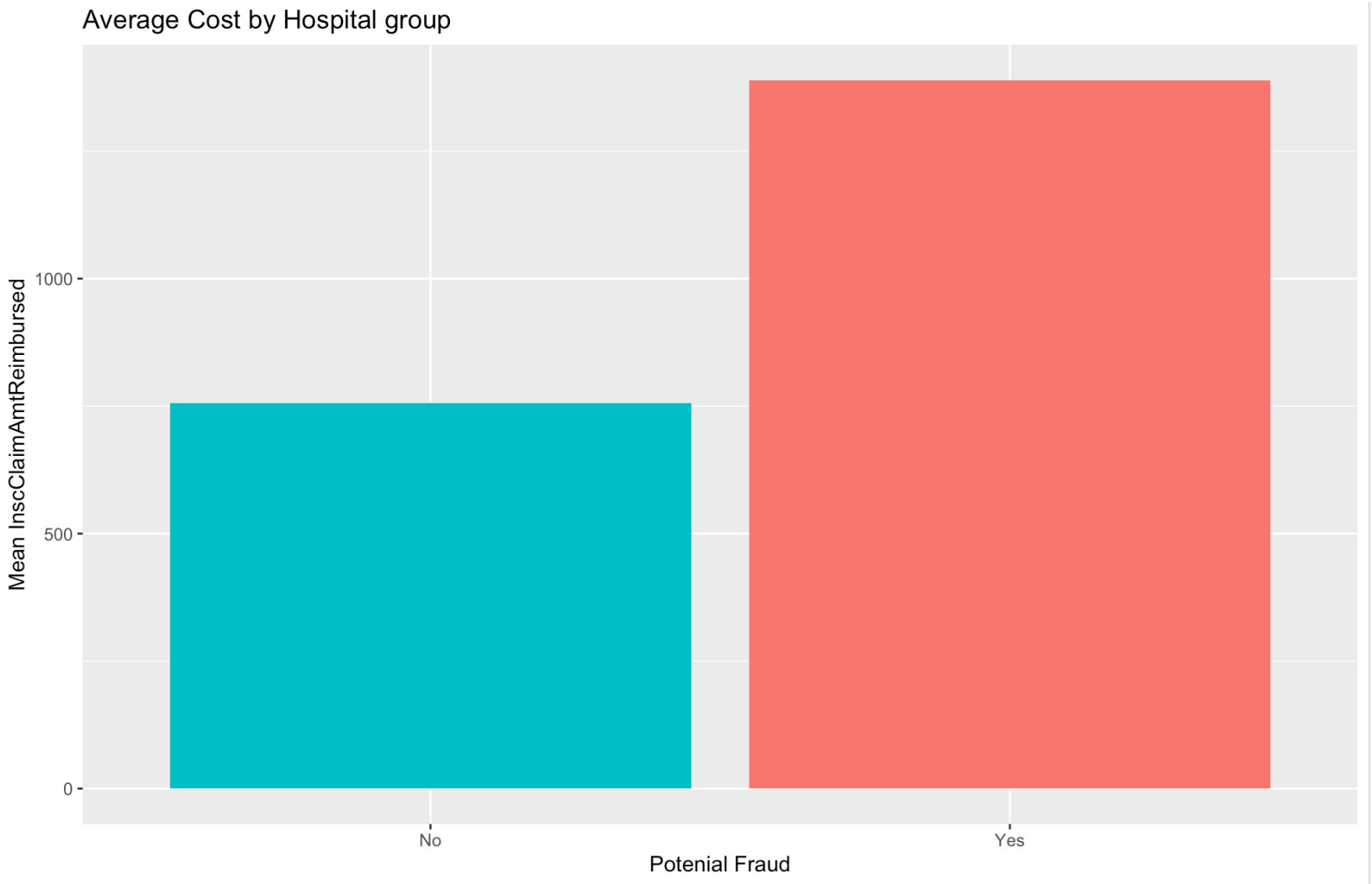
- Both Services: 23.4%
- Inpatient Services: 11.0%
- Outpatient Service: 1.9%

Number of hospitals commiting fraud, by state

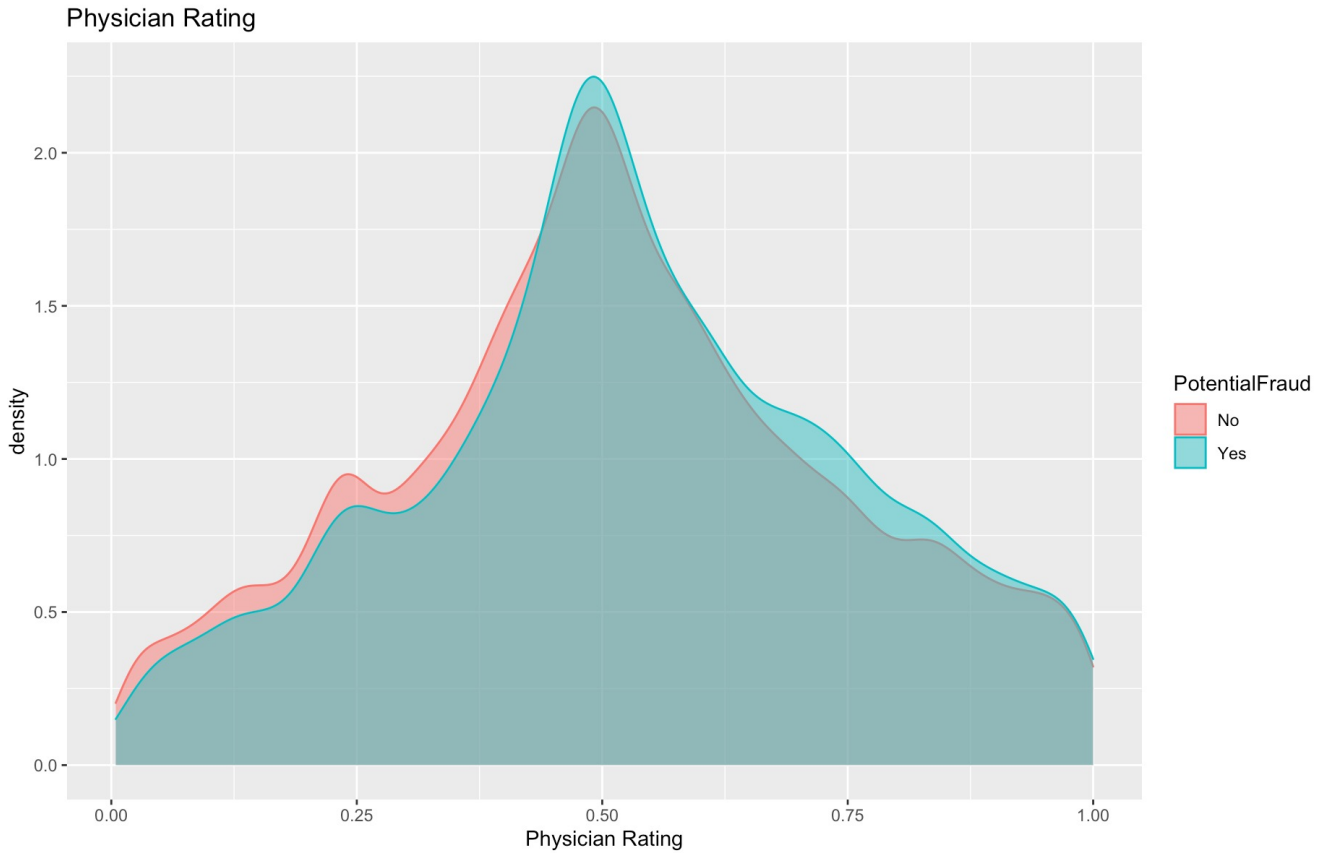


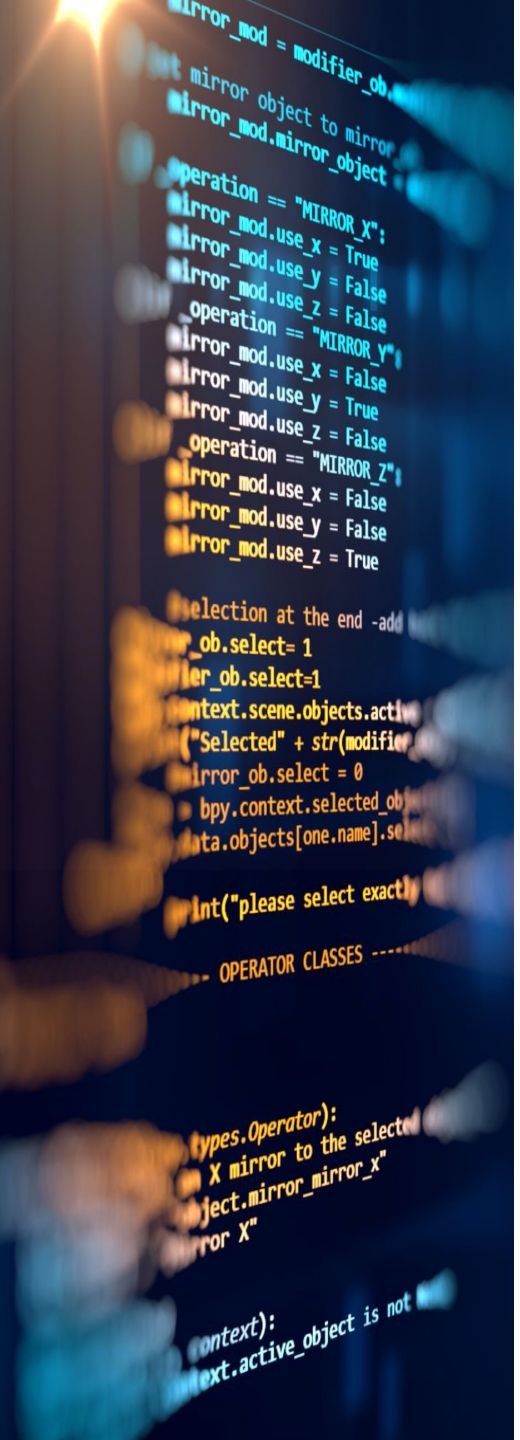


Average Insurance Claim Amount Reimbursed

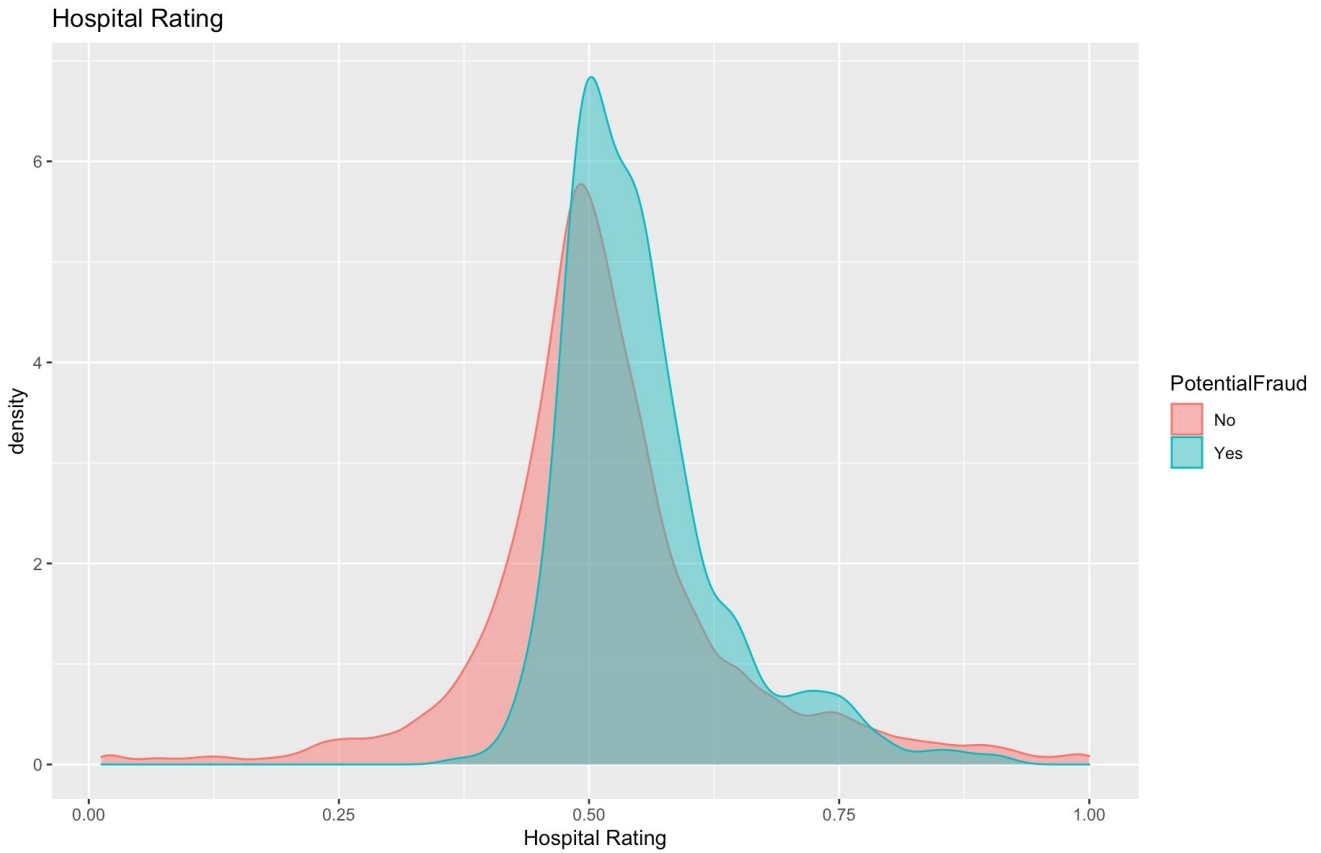


Physician Ranking



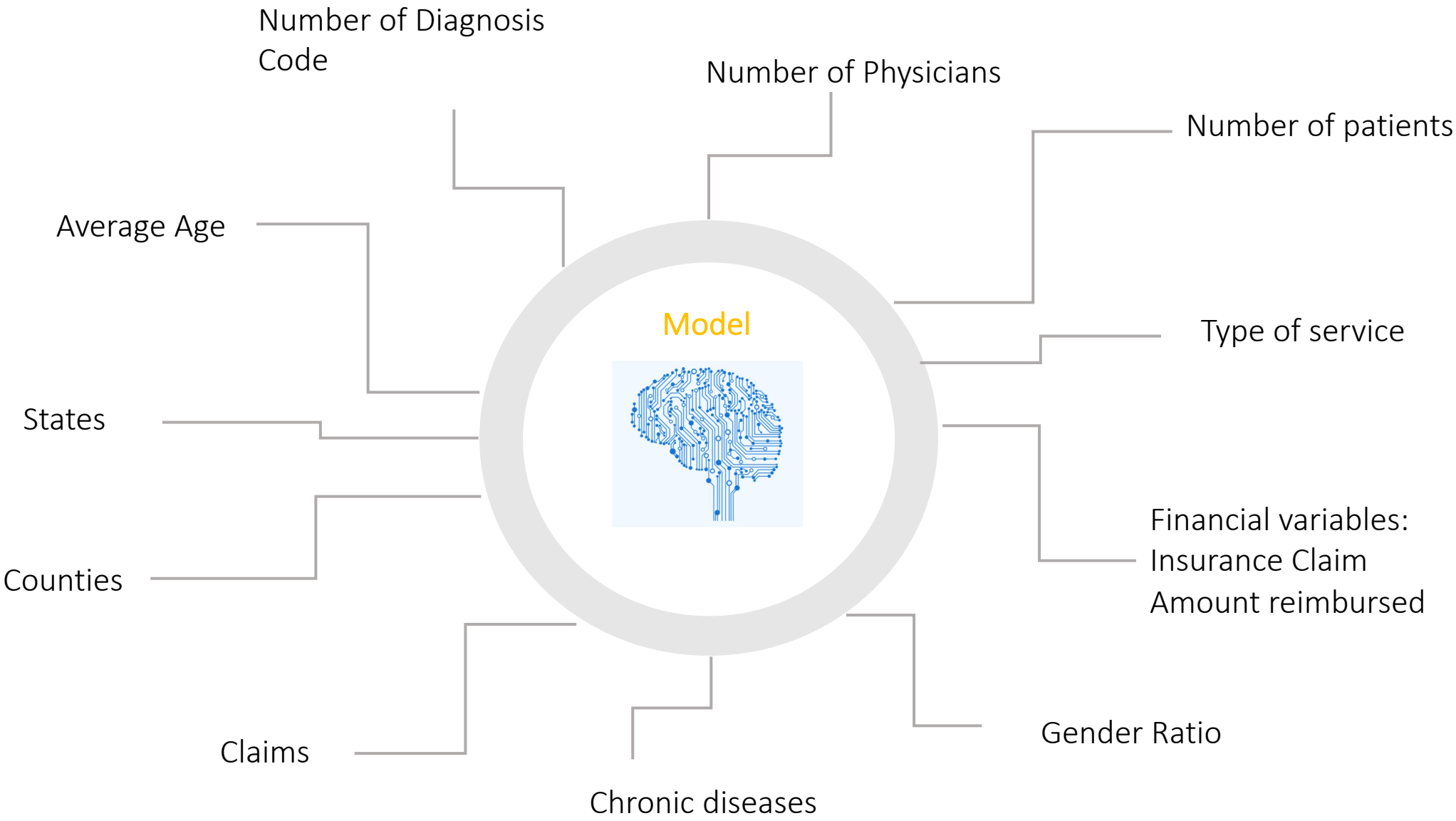


Hospital Ranking



Model Methodology

Feature Selection



Variable Selection

Run Lasso with Cross Validation

Get best lambda

Estimate model with best lambda

Variables selected:
- Sum of insurance claim amount reimbursed
- Outpatient to inpatient ratio
- Sum of distinct diagnosis codes
- State (dummies)

Adjust lamba

Estimate model with new lambda

Newly selected variables:
- Sum of insurance claim amount reimbursed
- Outpatient to inpatient ratio
- Sum of distinct diagnosis codes
- State (dummies)
- Sum of different attending physicians
- Sum of different beneficiaries
- Mean Age
- Hospital Rating

Comment: Performance of model with these variables was mediocre

Logistic Regression

	Estimate	Std. Error	Z value	Pr(>|z|)	
Intercept	-2.01E+00	1.67E+00	-1.205	2.28E-01	
sum.BeneID	-7.59E-03	1.85E-03	-4.109	3.97E-05	***
sum.InscClaimAmtREimbursed	6.73E-06	4.81E-07	13.973	2.00E-16	***
OutpatientInpatient2Ratio	-1.35E+03	3.09E-01	-4.35	1.36E-05	***
sum.AttendingPhysicians	6.32E-03	2.66E-03	2.376	1.75E-02	*
sum.Diagnosis	7.78E-03	1.40E-03	5.555	2.77E-08	***
summean_perc	1.13E+00	8.24E-01	1.376	0.1687	
mean.Age	1.93E-01	1.91E-02	-0.687	0.4919	
State Dummies	Y	-	-	-	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*'

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 2666.0 on 4322 degrees of freedom
Residual deviance: 1493 on 4264 degrees of freedom
AIC: 1611

Parameter Tuning

Logistic Regression	Random Forest	XGBoost	Support Vector Machine
No Parameter Tuning	Parameter Tuning bootstrap: True, max_depth: 6, max_features: 8, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 10	Parameter Tuning: colsample_bytree:1, gamma:5, learning_rates:0.001, max_depth:7, n_estimators: 40	Parameter Tuning: C:0.1, gamma:1, kernel: rbf
Train Score: 0.930	Train Score: 0.941	Train Score: 0.946	Train Score: 1
Test Score: 0.941	Test Score: 0.939	Test Score: 0.941	Test Score: 0.907

Conclusion

Conclusion

Best performing model:

- XGBoost

Performance Test:

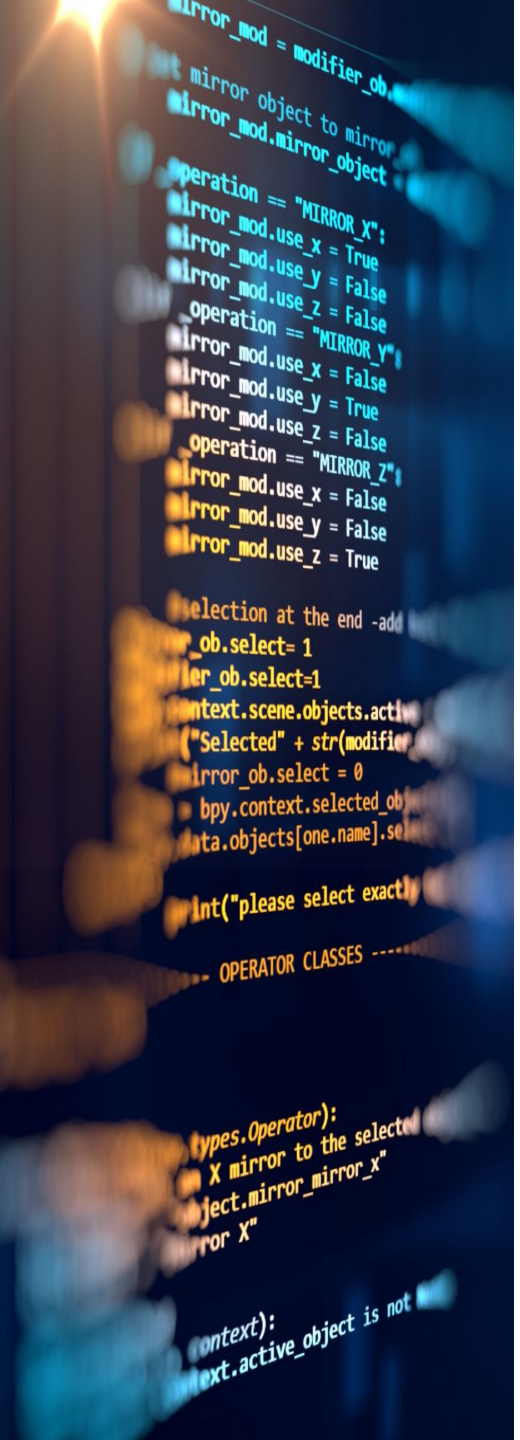
- 0.941 Accuracy

Best performing variables:

- Inpatient Outpatient Ratio
- Sum of insurance amount reimbursed
- Sum of different diagnosis code
- Hospital Rating
- Sum of different attending physicians
- Sum of different beneficiaries
- Mean age
- State

Comments:

- It is very difficult to detect fraud in the health care system.
- This is a case where black box models are not a problem.



Number of Claims in Hospitals which potentially commit Fraud