**Healthcare Providers Data For Anomaly Detection**

Problem statement:

Healthcare fraud is considered a challenge for many societies. Healthcare funding that could be spent on medicine, care for the elderly, or emergency room visits is lost to fraudulent activities by materialistic practitioners or patients. With rising healthcare costs, healthcare fraud is a major contributor to these increasing healthcare costs.

## About Dataset

Healthcare fraud is considered a challenge for many societies. Healthcare funding that could be spent on medicine, care for the elderly, or emergency room visits is lost to fraudulent activities by materialistic practitioners or patients. With rising healthcare costs, healthcare fraud is a major contributor to these increasing healthcare costs.

Try out various unsupervised techniques to find the anomalies in the data.

**Detailed Data File:**

The following variables are included in the detailed Physician and Other Supplier data file (see Appendix A for a condensed version of the variables included)).

**NPI** – National Provider Identifier (NPI) for the performing provider on the claim. The provider NPI is the numeric identifier registered in NPPES.

**nppes\_provider\_last\_org\_name** – When the provider is registered in NPPES as an individual (entity type code=’I’), this is the provider’s last name. When the provider is registered as an organization (entity type code = ‘O’), this is the organization's name.

**nppes\_provider\_first\_name** – When the provider is registered in NPPES as an individual (entity type code=’I’), this is the provider’s first name. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

**nppes\_provider\_mi** – When the provider is registered in NPPES as an individual (entity type code=’I’), this is the provider’s middle initial. When the provider is registered as an organization (entity type code= ‘O’), this will be blank.

**nppes\_credentials** – When the provider is registered in NPPES as an individual (entity type code=’I’), these are the provider’s credentials. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

**nppes\_provider\_gender** – When the provider is registered in NPPES as an individual (entity type code=’I’), this is the provider’s gender. When the provider is registered as an organization (entity type code = ‘O’), this will be blank.

**nppes\_entity\_code** – Type of entity reported in NPPES. An entity code of ‘I’ identifies providers registered as individuals and an entity type code of ‘O’ identifies providers registered as organizations.

**nppes\_provider\_street1** – The first line of the provider’s street address, as reported in NPPES.

**nppes\_provider\_street** – The second line of the provider’s street address, as reported in NPPES.

**nppes\_provider\_city** – The city where the provider is located, as reported in NPPES.

**nppes\_provider\_zip** – The provider’s zip code, as reported in NPPES.

**nppes\_provider\_state** – The state where the provider is located, as reported in NPPES. The state postal abbreviation reports the fifty U.S. states and the District of Columbia. The following values are used for all other areas:

'XX' = 'Unknown'  
'AA' = 'Armed Forces Central/South America'  
'AE' = 'Armed Forces Europe'  
'AP' = 'Armed Forces Pacific'  
'AS' = 'American Samoa'  
'GU' = 'Guam'  
'MP' = 'North Mariana Islands'  
'PR' = 'Puerto Rico'  
'VI' = 'Virgin Islands'  
'ZZ' = 'Foreign Country'

**nppes\_provider\_country** – The country where the provider is located, as reported in NPPES. The country code will be ‘US’ for any state or U.S. possession. For foreign countries (i.e., state values of ‘ZZ’), the provider country values include the following:  
AE=United Arab Emirates IT=Italy  
AG=Antigua JO= Jordan  
AR=Argentina JP=Japan  
AU=Australia KR=Korea  
BO=Bolivia KW=Kuwait  
BR=Brazil KY=Cayman Islands  
CA=Canada LB=Lebanon  
CH=Switzerland MX=Mexico  
CN=China NL=Netherlands  
CO=Colombia NO=Norway  
DE= Germany NZ=New Zealand  
ES= Spain PA=Panama  
FR=France PK=Pakistan  
GB=Great Britain RW=Rwanda  
GR=Greece SA=Saudi Arabia  
HU= Hungary SY=Syria  
IL= Israel TH=Thailand  
IN=India TR=Turkey  
IS= Iceland VE=Venezuela

**provider\_type** – Derived from the provider specialty code reported on the claim.

**medicare\_participation\_indicator** – Identifies whether the provider participates in Medicare and accepts the assigned assignment of Medicare-allowed amounts.

**place\_of\_service** – Identifies whether the place of service submitted on the claims is a facility (value of ‘F’) or non-facility (value of ‘O’). Non-facility is generally an office setting; however other entities are included in non-facility.

**hcpcs\_code** – HCPCS code used to identify the specific medical service furnished by the provider.

**hcpcs\_description** – Description of the HCPCS code for the specific medical service furnished by the provider.

**hcpcs\_drug\_indicator** –Identifies whether the HCPCS code for the specific service furnished by the provider is an HCPCS listed on the Medicare Part B Drug Average Sales Price (ASP) File.

**line\_srvc\_cnt** – Number of services provided; note that the metrics used to count the number provided can vary from service to service.

**bene\_unique\_cnt** – Number of distinct Medicare beneficiaries receiving the service.

**bene\_day\_srvc\_cnt** – Number of distinct Medicare beneficiaries/per day services.

**average\_Medicare\_allowed\_amt** – Average of the Medicare allowed amount for the service.

**stdev\_Medicare\_allowed\_amt** – Standard deviation of the Medicare allowed amounts.

**average\_submitted\_chrg\_amt** – Average of the charges that the provider submitted for the service.

**stdev\_submitted\_chrg\_amt** – Standard deviation of the charge amounts submitted by the provider.

**average\_Medicare\_payment\_amt** – The average amount that Medicare paid after deductible and coinsurance amounts have been deducted for the line item service.

Exploratory DataAnalysis:

Some of the unwanted columns have been dropped and accordingly, the rows performed the plots.

* Credentials of the Provider
* Gender of the Provider
* Entity Type of the Provider
* City of the Provider
* Zip Code of the Provider
* State Code of the Provider
* Country Code of the Provider
* Provider Type
* Medicare Participation Indicator
* Place of Service
* HCPCS Code
* HCPCS Drug Indicator
* Number of Services
* Number of Medicare Beneficiaries
* Number of Distinct Medicare Beneficiary/Per Day Services
* Average Medicare Allowed Amount
* Average Submitted Charge Amount
* Average Medicare Payment Amount
* Average Medicare Standardized Amount

For the heat map, we can visualize correlations, identify patterns, detect anomalies, and simplify complex data.

From the above analysis, most of the null values were found in the provider's name and the providers who belonged to the organizations.