Profile Evaluation data by Denials

Analysis by Narendra Prasad K

Inferences Observed in Denials Profile data

Data Observations

- Denials Profile data received for 14 months from Jan 2016 to 2017 Feb
- Denials Profile data contains dimensions and facts information, there are over 13439 records with 41 variables.
- Observed that, the dataset contains 20 Providers information, Billed information with Approval Amount details and Payer insurance details, also data set has Category wise denials, Diagnosis Description details along with their Denial Reason and description code.
- Data has Insurance payer information and allowed amount by the insurance.
- Identified the Missing value information is below.

Variable_Name	% Count Missing
Modality	100%
Rv.Cd	100%
Diag.2	71%
Diag.Desc.2	71%
Ded.Amt	44%
Alwd.Amt	29%

Objectives

- Trace out significant tests between top 5 Payer and CPT4.
- Create the slicing and dicing for Billed Amount with related factors.
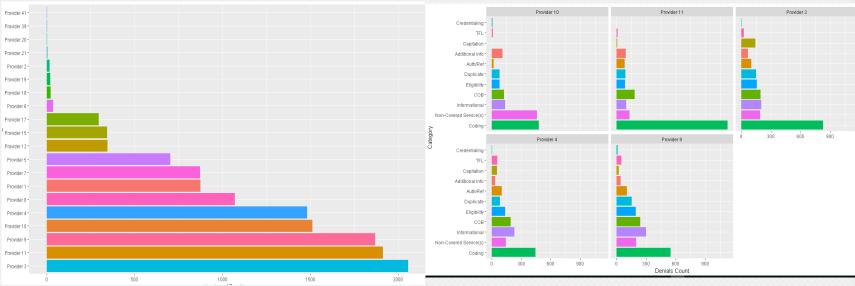- Identify the decision tree rules for denials by Providers

TOP 5 : Provider Denials Vs category

sort(xtabs(~ Enc.Rendering, data=denials_report),decreasing = TRUE)

```
 Provider 3  Provider 11    Provider 9  Provider 10    Provider 4    Provider 8
       2059         1915          1870         1512          1483          1070
 Provider 1    Provider 7    Provider 5  Provider 12  Provider 15  Provider 17
        874          872           703          345          344           295
 Provider 6  Provider 18    Provider 19    Provider 2  Provider 21  Provider 20
         35           21            19           16            3            1
Provider 39  Provider 41
          1            1
```

x11()

ggplot(denials_report, aes(x=reorder(Enc.Rendering, -table(Enc.Rendering)[Enc.Rendering]),fill=Enc.Rendering)) + geom_bar()+coord_flip()+
  labs (x = "Provider_id", y = "Number of Providers")
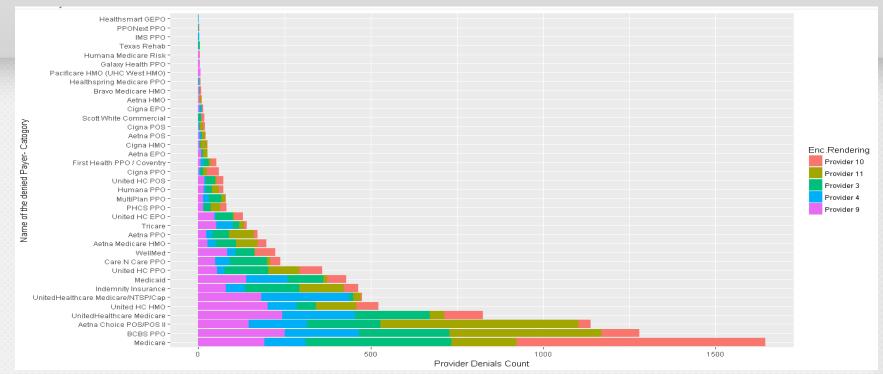

Provider wise Insights :
- Top5 Denials - Provider 3,11,9,10 and 4
- Above Top5 denials covered nearly 65% in overall data
- Top5 Provider categories are : Coding , Non-Covered Services , Informational, COB and Eligibility

TOP 5 : Provider Vs Insurance Category

x11()
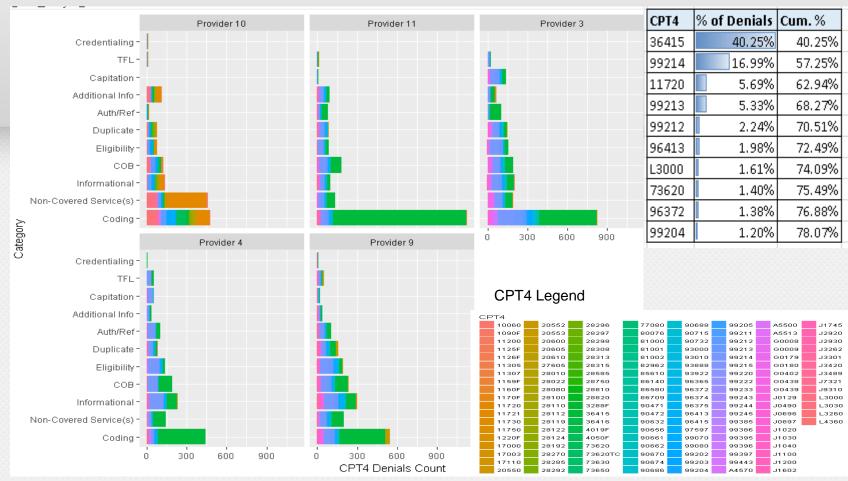
ggplot(Top5_Providers, aes(x=reorder(Fin.Class, -table(Fin.Class)[Fin.Class]),fill=Enc.Rendering)) + geom_bar()+

coord_flip()+ labs (x = "FIN.Class", y = "Denials Count")

Insights

- Insurance Company Category wise denials :**Medicare, BCBS PPO, Aetna Choice POS/POS2 , United Healthcare Medicare**

- **Medicare** : Most of the denials are Provider 3 and Provider 10

- **BCBS PPO** : Most of the denials are Provider 11 and Provider 9

- **Aetna Choice POS/POS2:** Most of the denials are Provider 11 , Provider 9 and Provider3

- **United Healthcare Medicare** : Most of the denials are Provider 9, Provider 4 and Provider 3

TOP 5 : Provider Vs CPT4 and Category

CPT4	% of Denials	Cum. %
36415	40.25%	40.25%
99214	16.99%	57.25%
11720	5.69%	62.94%
99213	5.33%	68.27%
99212	2.24%	70.51%
96413	1.98%	72.49%
L3000	1.61%	74.09%
73620	1.40%	75.49%
96372	1.38%	76.88%
99204	1.20%	78.07%

Insights :

- CPT4-36415 is highly denials in TOP5 Providers, which is covered nearly 40%
- CPT4-36415 is covered under category of "Coding".
- In top 5 providers, top10 CPT4 Coverage is 78%
- CPT 36415 is majorly effecting on providers : 11, 3,4 and 9

TOP 5 : Significant Test for Provider and CPT4

Top5_Providers = subset(denials_report,denials_report$Enc.Rendering %in% c("Provider 3","Provider 11","Provider 9","Provider 10","Provider 4"))

- ❑ **Null Hypothesis : H0 = There is No Sig. between Provider and CPT4**
- ❑ **Alternative Hypothesis : H1 = There is Sig. between Provider and CPT4**
- ❑ **Sig level 95% with 0.05**

test_Top5_Providers = table(Top5_Provider_id$**Enc.Rendering**,Top5_Provider_id$CPT4)

print(test_Top5_Providers)

print(chisq.test(test_Top5_Providers))

Test Statistics :

> Pearson's Chi-squared test
> data: test_Top5_Providers
> X-squared = 11534, df = 580, p-value < 2.2e-16

Conclusion : Accept alternative Hypothesis (H1) i.e. "strong sig. between Provider and CPT"

Top5_CPT4_CPT = subset(denials_report,denials_report$CPT4 %in% c("36415","99214","99213","11720","G0439"))

- ❑ **Null Hypothesis : H0 = There is No Sig. between Provider r and CPT4**
- ❑ **Alternative Hypothesis : H1 = There is Sig. between Provider and CPT4**
- ❑ **Sig level 95% with 0.05**

test_Top_CPT4 = table(Top5_CPT4$**CPT4**,Top5_CPT4$Enc.Rendering)

print(test_Top_CPT4)

print(chisq.test(test_Top_CPT4))

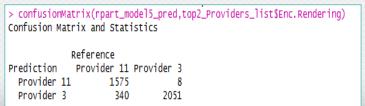**Test Statistics :**

> Pearson's Chi-squared test
> data: test_Top_CPT4
> X-squared = 10187, df = 72, p-value < 2.2e-16

Conclusion : Accept alternative Hypothesis (H1) i.e. "strong sig. between Provider and CPT"

Decision Tree : Provider3 and Provider11(Top-2)

- Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems.

- Goal : Identify the denials reason for Provider 3 and Provider 11 using decision tree.

- To identify the denials reason we are using the "**rpart**" algorithm. as this is binary splitting with Gini index.

  - **Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets.

  - **Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node.

  - **Leaf/ Terminal Node:** Nodes do not split is called Leaf or Terminal node.

  - **Splitting:** It is a process of dividing a node into two or more sub-nodes.

- Denials splitting by GINI is done based on (0.5,0.5), as assumed there are equal number of records from each class. (Probability is 1 if classes are pure)

- For Model evaluation process we are using caret Library trainControl method for cross validation with iterations count as 5, So there is no need to split the data for "Training and Validation".

- After building the model we done sample prediction on same dataset, and compare the Prediction , actual using the Confusion matrix.

- We build the Final tree model using 10 variables with 90% of Accuracy.

- Verified the Accuracy using confusion matrix, out of 3974 records, 348 records are misclassified. See below Confusion Matrix and Statistics

```
> confusionMatrix(rpart_model5_pred,top2_Providers_list$Enc.Rendering)
Confusion Matrix and Statistics

              Reference
Prediction    Provider 11 Provider 3
  Provider 11        1575          8
  Provider 3          340       2051
```

Decision Tree :R code

- **Train Control : Control parameters for train**
  - fitControl <- trainControl(method = "cv", number = 5)
  - tune_grid = expand.grid(.cp = seq(0.01,0.1,0.01))
- **Decision Tree :**

set.seed(1005)

rpart_model5=train(Enc.Rendering~Billed.Amt+Aprv.Amt+Tran.Cd+Tran.Status+Rsn.Cds.Remarks+Category+Sub.Category+CPT4+Diag.Desc.1+Payer.Name,data = top2_Providers_list,**method = "rpart**",
trControl = fitControl, tuneGrid=tune_grid)

Output :

```
> rpart_model5
CART

3974 samples
  10 predictor
   2 classes: 'Provider 11', 'Provider 3'

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 3179, 3179, 3179, 3179, 3180
Resampling results across tuning parameters:

  cp    Accuracy    Kappa
  0.01  0.9073983   0.8133502
  0.02  0.8779570   0.7534532
  0.03  0.8563196   0.7092488
  0.04  0.8379491   0.6716396
  0.05  0.7941644   0.5815754
  0.06  0.7941644   0.5815754
  0.07  0.7941644   0.5815754
  0.08  0.7632134   0.5175029
  0.09  0.7576788   0.5060623
  0.10  0.7425655   0.4746347

Accuracy was used to select the optimal model using  the
 largest value.
The final value used for the model was cp = 0.01.
```

Decision Tree :Rules

```
> rpart_model5$finalModel
n= 3974

node), split, n, loss, yval, (yprob)
      * denotes terminal node

   1) root 3974 1915 Provider 3 (0.481882235 0.518117765)
     2) Diag.Desc.1Rheu arthritis w rheu factor mult site w/o org/sys involv>=0.5 747     3 Provider 11 (0.995983936 0.004016064) *
     3) Diag.Desc.1Rheu arthritis w rheu factor mult site w/o org/sys involv< 0.5 3227 1171 Provider 3 (0.362875736 0.637124264)
       6) Diag.Desc.1Rheumatoid arthritis w/o rheumatoid factor, multiple sites>=0.5 211     0 Provider 11 (1.000000000 0.000000000) *
       7) Diag.Desc.1Rheumatoid arthritis w/o rheumatoid factor, multiple sites< 0.5 3016  960 Provider 3 (0.318302387 0.681697613)
        14) Diag.Desc.1Polyosteoarthritis, unspecified>=0.5 145     0 Provider 11 (1.000000000 0.000000000) *
        15) Diag.Desc.1Polyosteoarthritis, unspecified< 0.5 2871  815 Provider 3 (0.283873215 0.716126785)
          30) Diag.Desc.1Arthropathic psoriasis, unspecified>=0.5 92     0 Provider 11 (1.000000000 0.000000000) *
          31) Diag.Desc.1Arthropathic psoriasis, unspecified< 0.5 2779  723 Provider 3 (0.260165527 0.739834473)
            62) Diag.Desc.1Other psoriatic arthropathy>=0.5 82     0 Provider 11 (1.000000000 0.000000000) *
            63) Diag.Desc.1Other psoriatic arthropathy< 0.5 2697  641 Provider 3 (0.237671487 0.762328513)
             126) Diag.Desc.1Ankylosing spondylitis of unspecified sites in spine>=0.5 65     0 Provider 11 (1.000000000 0.000000000) *
             127) Diag.Desc.1Ankylosing spondylitis of unspecified sites in spine< 0.5 2632  576 Provider 3 (0.218844985 0.781155015)
               254) Diag.Desc.1Systemic involvement of connective tissue, unspecified>=0.5 54     0 Provider 11 (1.000000000 0.000000000) *
               255) Diag.Desc.1Systemic involvement of connective tissue, unspecified< 0.5 2578  522 Provider 3 (0.202482545 0.797517455)
                 510) Diag.Desc.1Inflammatory polyarthropathy>=0.5 47     3 Provider 11 (0.936170213 0.063829787) *
                 511) Diag.Desc.1Inflammatory polyarthropathy< 0.5 2531  478 Provider 3 (0.188858159 0.811141841)
                  1022) Diag.Desc.1Rheumatoid Arthritis>=0.5 37     0 Provider 11 (1.000000000 0.000000000) *
                  1023) Diag.Desc.1Rheumatoid Arthritis< 0.5 2494  441 Provider 3 (0.176824379 0.823175621)
                    2046) Diag.Desc.1Unspecified osteoarthritis, unspecified site>=0.5 31     1 Provider 11 (0.967741935 0.032258065) *
                    2047) Diag.Desc.1Unspecified osteoarthritis, unspecified site< 0.5 2463  411 Provider 3 (0.166869671 0.833130329)
                      4094) Diag.Desc.1Systemic lupus erythematosus, unspecified>=0.5 30     1 Provider 11 (0.966666667 0.033333333) *
                      4095) Diag.Desc.1Systemic lupus erythematosus, unspecified< 0.5 2433  382 Provider 3 (0.157007809 0.842992191)
                        8190) Diag.Desc.1Enteropathic arthropathies, unspecified site>=0.5 22     0 Provider 11 (1.000000000 0.000000000) *
                        8191) Diag.Desc.1Enteropathic arthropathies, unspecified site< 0.5 2411  360 Provider 3 (0.149315637 0.850684363)
                         16382) Diag.Desc.1Polymyalgia rheumatica>=0.5 20     0 Provider 11 (1.000000000 0.000000000) *
                         16383) Diag.Desc.1Polymyalgia rheumatica< 0.5 2391  340 Provider 3 (0.142199916 0.857800084) *
> |
```

- **Insights :**
  - Final Decision tree generate the rules based on the Diag.Desc Variable
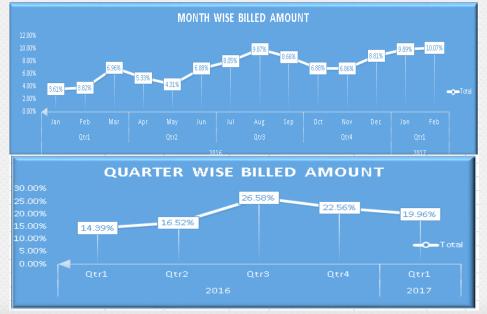 - For Decision map pick the "**Provider ID Treemap.pdf**" as below link.

Provider ID Tree map.pdf

Insights based on Data Analysis

➢ Nearly **50%** of total billed amount denials for providers are" **Provider 9 and Provider 11**"

➢ Overall **20%** of total billed amount denials for Payers are "**Medicare and BCBS 84980**"

➢ Overall **39%** of total billed amount denials for Fin class Category are "**Medicare, United HC HMO and BCBS PPO**"

➢ Nearly **53%** of total billed amount denials for Category are "**Coding, Informational and Non-Covered Service(s)**"

➢ Nearly 38% of total billed amount denials for CPT4 are **"J9310, J0129 and 99214"**

➢ Nearly 23 % of total billed amount denials for Diag Desc 1 is "**Rheu arthritis w rheu factor mult site w/o org/sys involve**".

➢ See **Month wise Trend** and **Quarter wise trends**

MONTH WISE BILLED AMOUNT — Jan 3.61%, Feb 3.82%, Mar 6.96%, Apr 5.33%, May 4.31%, Jun 6.88%, Jul 8.05%, Aug 9.87%, Sep 8.66%, Oct 6.88%, Nov 6.86%, Dec 8.81%, Jan 9.89%, Feb 10.07%

QUARTER WISE BILLED AMOUNT — Qtr1 14.39%, Qtr2 16.52%, Qtr3 26.58%, Qtr4 22.56%, Qtr1 19.96%

Year : 2016
Highest Denials : Aug, Dec, and Sept
Least Denials : Jan, Feb and May
==============================
Year : 2017
• **Compare to 2016, 2017 are more denials in Jan and Feb**
==============================
2016 Quarter Analysis
• Quarter 3 is more denials
• Quarter 4 is 2nd Highest

Thank you.