HealthCare Cost Analysis

Q1) To record the patient statistics, the agency wants to find the age category of people who frequently visit the hospital and has the maximum expenditure.

Ans) The below mention is the code in R which gives an overview of the age category of people who frequently visit the hospital and has the maximum expenditure. The excel file has been converted to csv file for the analysis.

```
print("Healthcare cost Analysis")

health_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/Hospitals/main/Hospital.csv")

print(health_data)

str(health_data)

max(health_data$TOTCHG)

thedata<-filter(health_data,TOTCHG==48388)

print(thedata)


thedata<-filter(health_data,LOS==41)

print(thedata)
```

Below attached are the outputs

```
print(health_data)
    AGE FEMALE LOS RACE TOTCHG APRDRG
1    17      1   2    1   2660    560
2    17      0   2    1   1689    753
3    17      1   7    1  20060    930
4    17      1   1    1    736    758
5    17      1   1    1   1194    754
6    17      0   0    1   3305    347
7    17      1   4    1   2205    754
8    16      1   2    1   1167    754
9    16      1   1    1    532    753
10   17      1   2    1   1363    758
11   17      1   2    1   1245    758
12   15      0   2    1   1656    753
13   15      1   2    1   1379    751
14   15      1   4    1   2346    758
15   15      1   7    1   4006    753
16   15      1   4    1   2181    758
17   14      1   1    1    628    754
18   14      1   4    1   2463    758
```

19	15	1	3	1	1956	753
20	14	1	3	1	1802	758
21	13	1	1	1	3188	812
22	17	1	2	1	2129	566
23	12	0	1	1	7421	249
24	15	1	1	1	1122	422
25	13	1	2	4	1173	754
26	12	0	2	1	3625	812
27	11	1	2	1	3908	50
28	15	0	1	1	3994	139
29	11	0	0	1	1033	753
30	10	0	2	1	2860	141
31	11	0	2	1	3814	420
32	7	0	0	1	1132	139
33	16	1	2	6	1163	751
34	17	1	1	1	610	751
35	6	0	3	1	9530	97
36	15	1	1	1	1268	811
37	17	1	4	1	2582	753
38	16	1	2	1	1287	755
39	17	1	3	1	6594	930
40	13	1	0	1	909	755
41	7	0	0	1	2530	347
42	11	1	2	2	1534	753
43	3	0	5	1	14243	720
44	16	1	3	1	1699	754
45	2	0	2	1	7298	53
46	16	1	1	1	636	754
47	15	1	1	1	626	754
48	1	0	2	1	3782	53
49	14	1	2	1	1444	753
50	14	1	2	1	1183	754
51	14	1	5	1	3045	754
52	14	1	5	1	3624	754
53	14	1	12	1	6810	760
54	1	0	1	1	1409	249
55	13	0	2	1	1211	754
56	1	0	4	1	9606	53
57	1	1	1	1	1411	249
58	15	1	0	1	607	754
59	1	0	1	1	2932	249
60	1	0	3	1	5075	139
61	14	1	1	1	762	753
62	16	1	6	1	6329	753
63	17	1	1	1	1226	753
64	3	1	4	1	8223	710
65	17	0	2	1	1193	776
66	13	1	2	1	1076	754
67	12	1	6	1	17434	115
68	12	1	2	1	1647	753
69	14	1	7	1	3865	754
70	13	1	1	1	628	754
71	15	1	1	1	806	755
72	0	1	41	1	29188	602
73	0	0	2	1	4717	138
74	0	0	12	1	15129	137

75	0	1	2	1	1085	640
76	0	0	3	1	1607	640
77	0	1	3	1	1499	640
78	0	1	3	1	7648	53
79	0	1	2	1	1527	640
80	0	0	2	1	1483	640
81	0	1	4	1	2844	640
82	0	1	3	1	3124	640
83	0	0	3	1	1760	640
84	0	1	2	1	1278	640
85	0	1	2	1	1620	640
86	0	1	2	1	1220	640
87	0	1	2	1	1134	640
88	16	1	0	1	1235	754
89	0	0	3	1	1656	640
90	0	0	4	5	4072	639
91	0	0	2	5	1393	143
92	0	0	0	5	615	254
93	16	1	1	1	779	755
94	0	0	2	1	1385	640
95	0	0	2	1	1224	640
96	0	1	3	1	1779	640
97	0	0	2	1	1526	640
98	15	1	1	1	882	754
99	0	0	1	1	2075	581
100	0	0	17	1	12042	633
101	0	0	2	1	1309	640
102	0	0	2	1	1290	640
103	0	0	2	1	1280	640
104	0	0	3	1	1719	640
105	0	1	2	1	1102	640
106	0	1	3	1	1543	640
107	0	1	2	1	1174	640
108	0	1	2	1	1105	640
109	0	0	2	1	1335	640
110	0	0	2	1	1550	640
111	0	0	4	1	2473	640
112	0	0	2	1	1322	640
113	0	0	4	1	2553	640
114	15	0	5	1	2835	753
115	0	1	2	1	1191	640
116	0	0	2	1	1439	640
117	0	1	2	1	1237	640
118	0	0	2	1	1265	640
119	0	1	4	1	2280	640
120	0	0	2	1	1096	640
121	0	1	2	1	1156	640
122	0	0	2	1	1199	640
123	13	1	10	1	5615	754
124	0	1	4	1	2518	640
125	15	0	0	1	625	754
126	0	1	2	1	1246	640
127	0	1	3	1	1821	640
128	0	0	5	1	3101	626
129	12	1	2	1	1293	754
130	0	1	2	1	1176	640

```
131   0      0   3    1   1891    640
132   5      1   2    1  10584     53
133  13      1   3    1   2373    754
134   0      0   1    1    935    640
135   0      0   2    1   1395    640
136   0      0   2    1   1561    640
137   0      1   7    1   6912    636
138  12      1   2    1   1157    754
139   0      0   3    1   2197    640
140   0      0   4    1   2288    640
141  16      1   4    1   2348    754
142   0      0   2    1   1320    640
143   0      1   2    1   1139    640
144   0      1   4    1   2134    639
145   0      0   2    1   1407    640
146   0      0   2    1   1982    640
147   0      0   4    1   2539    640
148   0      0   2    1   1528    640
149   0      1   2    1   1513    640
150   0      1   2    1   1191    640
151   0      0   2    1   1280    640
152   0      0   2    1   3977    139
153   0      1   2    1   1269    640
154   0      0   2    1   1501    640
155   0      1   2    1   1396    640
156   0      0   3    1   1777    640
157   0      1   1    1    833    640
158   0      1   1    1    715    640
159  17      1   5    1   2936    751
160   0      0   2    1   1375    640
161   0      0   2    1   1330    640
162   0      0   2    1   1628    640
163   0      0   2    1   1368    640
164  12      1   1    1    622    755
165  17      0   2    1  14174     23
166   7      0   1    1   6425     57
 [ reached 'max' / getOption("max.print") -- omitted 334 rows ]
```

The above mentioned screenshot represent the tabular view of Hospital Records

```
thedata<-filter(health_data,TOTCHG==48388)
> print(thedata)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1  17      1   7    1  48388    911
```

```
>
```

The above mention screenshot gives an overview of the Hospital Record with highest expenditure

```
thedata<-filter(health_data,LOS==41)
```

```
> print(thedata)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1   0      1  41    1  29188    602


>
```

The above mention screenshot gives an overview of the Hospital Record with longest length of stay in days

Hence,the age category of people who frequently visit the hospital and has the maximum expenditure falls under 0 and in 17 age groups

Q2) In order of severity of the diagnosis and treatments and to find out the expensive treatments, the agency wants to find the diagnosis-related group that has maximum hospitalization and expenditure.

Ans) The below mention is the code in R which helps in diagnosing the diagnosis related group that has maximum hospitalization and expenditure

```
print("Healthcare cost Analysis")

health_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/Hospitals/main/Hospital.csv")

print(health_data)


thedata<-filter(health_data,LOS==41 &TOTCHG==48388)

print(thedata)


thedata<-filter(health_data,LOS==41)

print(thedata)


thedata<-filter(health_data,TOTCHG==48388)

print(thedata)
```

Below attached are the screenshots of the output

```
thedata<-filter(health_data,LOS==41)
> print(thedata)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1   0      1  41    1  29188    602
```

The above mention screenshot represent the overview of the hospital data which has longest Length of stay in days

```
thedata<-filter(health_data,TOTCHG==48388)
> print(thedata)
  AGE FEMALE LOS RACE TOTCHG APRDRG
1  17      1   7    1  48388    911


>
```

The above mention screenshot represent the overview of the hospital data with maximum expenditure

Hence,the diagnosis related groups for maximum hospital and expenditure comes under 911 and 602 groups

Q3) To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.

Ans) For analysing the race of the patient wrt hospitalization costs, I am using the concept of Simple Linear Regression. Below mention is the code in R

```
print("Healthcare cost Analysis")

health_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/Hospitals/main/Hospital.csv")

print(health_data)

View(health_data)

health_data$RACE<-as.integer(health_data$RACE)

health_results<-lm(formula=RACE~TOTCHG,data=health_data)

print(health_results)

print(summary(health_results))
```

```
print(health_results)

Call:
lm(formula = RACE ~ TOTCHG, data = health_data)

Coefficients:
(Intercept)       TOTCHG
  1.085e+00   -2.403e-06
```

The above mention screenshot represent the relationship between Race of the patient and Hospital Discharge Cost. It is clear from the above mention output that there is a negative correlation between Race and Hospital Discharge Cost as the value of Hospital Discharge Cost is negative wrt Race of the Patient

```
print(summary(health_results))

Call:
lm(formula = RACE ~ TOTCHG, data = health_data)

Residuals:
    Min      1Q  Median      3Q     Max
-0.0836 -0.0819 -0.0810 -0.0786  4.9189

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.085e+00  2.834e-02  38.274   <2e-16 ***
TOTCHG      -2.403e-06  5.932e-06  -0.405    0.686
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5152 on 497 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.0003299, Adjusted R-squared:  -0.001681
F-statistic: 0.164 on 1 and 497 DF,  p-value: 0.6856
```

If,I look at the summary of the health results it is clear that maximum value of residuals is 5(approx) which is basically the difference between the dependent variable and predicted variable. Almost 33% approx. values from the above mention formula fit to the model. Hence there is negligible dependency between Race and Hospital Discharge Cost

Q4) To properly utilize the costs, the agency has to analyze the severity of the hospital costs by age and gender for the proper allocation of resources.

Ans) For analysing the severity of the hospital costs by age and gender for the proper allocation or resources, I am using the concept of Regression Analysis with multiple variables

Below mention is the code in R

```
print("Healthcare cost Analysis")

health_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/Hospitals/main/Hospital.csv")

print(health_data)

health_results<-lm(formula=TOTCHG~AGE+RACE,data=health_data)
```

print(health_results)

summary(health_results)

Below attached are the output

```
health_results<-lm(formula=TOTCHG~AGE+RACE,data=health_data)
> print(health_results)

Call:
lm(formula = TOTCHG ~ AGE + RACE, data = health_data)

Coefficients:
(Intercept)          AGE         RACE
    2567.63        73.59      -153.08
```

The above mention screenshot represent the relationship between the Hospital Discharge Cost wrt Age and Race.It is clear from the above mention output that there is a positive linear regression between age and hospital discharge cost and negative linear regression between Race and Hospital Discharge Cost

```
summary(health_results)

Call:
lm(formula = TOTCHG ~ AGE + RACE, data = health_data)

Residuals:
   Min     1Q Median     3Q    Max
 -3060  -1319  -1002   -291  44722

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2567.63     419.79   6.116 1.94e-09 ***
AGE            73.59      24.91   2.954  0.00329 **
RACE         -153.08     336.51  -0.455  0.64937
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3865 on 496 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.01761,   Adjusted R-squared:  0.01365
F-statistic: 4.446 on 2 and 496 DF,  p-value: 0.01219



>
```

If I look at the summary of the health results of the hospital,it is clear that age has positive impact on Hospital Discharge while Race has negative impact on hospital discharge. Even from the residuals it is clear that maximum value of residuals is 44722 ie the difference between the dependent variable and predictor

Q5) Since the length of stay is the crucial factor for inpatients, the agency wants to find if the length of stay can be predicted from age, gender, and race.

Ans) For determining the relationship of length of stay wrt age,gender and race, I am using the concept of Naïve Bayes and Decision Tree Model. Below mentioned is the code in R

```
print("Healthcare cost Analysis")

health_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/Hospitals/main/Hospital.csv")

print(health_data)

str(health_data)

health_data$LOS<-sapply(health_data$LOS,factor)

# Build the model

naive_model<-naiveBayes(LOS~.,data=health_data)

print(naive_model)



# Predicting the model


naive_predict<-predict(naive_model,health_data)

naive_predict

table(naive_predict,health_data$LOS)



# Decision Tree

tree_model<-rpart(LOS~.,data=health_data,method="class")

print(tree_model)
```

summary(tree_model)

Below attached are the output of screenshots

```
library(e1071)
> naive_model<-naiveBayes(LOS~.,data=health_data)
> print(naive_model)

Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
    2     7     1     0     4     3     5    12     6    41    17    10
0.448 0.022 0.158 0.030 0.076 0.196 0.028 0.004 0.016 0.002 0.002 0.002
   39     8    18    15     9    23    24
0.002 0.002 0.004 0.002 0.002 0.002 0.002

Conditional probabilities:
    AGE
Y         [,1]      [,2]
  2   3.120536  5.943586
  7   6.636364  7.839295
  1  10.493671  6.539571
  0  10.533333  6.300416
  4   6.210526  7.637595
  3   2.734694  5.721277
  5   9.714286  7.194320
  12  7.000000  9.899495
  6  10.875000  6.895910
  41  0.000000        NA
  17  0.000000        NA
  10 13.000000        NA
  39  0.000000        NA
  8   0.000000        NA
  18  7.500000 10.606602
  15  0.000000        NA
  9  15.000000        NA
  23  0.000000        NA
  24  0.000000        NA

    FEMALE
Y         [,1]      [,2]
  2  0.4687500 0.5001401
  7  0.7272727 0.4670994
  1  0.5443038 0.5012157
  0  0.4666667 0.5163978
  4  0.6578947 0.4807829
  3  0.5000000 0.5025707
  5  0.5714286 0.5135526
  12 0.5000000 0.7071068
  6  0.6250000 0.5175492
  41 1.0000000        NA
  17 0.0000000        NA
```

```
  10 1.0000000        NA
  39 0.0000000        NA
  8  0.0000000        NA
  18 0.5000000 0.7071068
  15 0.0000000        NA
  9  0.0000000        NA
  23 1.0000000        NA
  24 1.0000000        NA

    RACE
Y        [,1]      [,2]
  2  1.094170 0.5890228
  7  1.000000 0.0000000
  1  1.012658 0.1125088
  0  1.266667 1.0327956
  4  1.184211 0.7298746
  3  1.030612 0.3030458
  5  1.214286 0.8017837
  12 1.000000 0.0000000
  6  1.000000 0.0000000
  41 1.000000        NA
  17 1.000000        NA
  10 1.000000        NA
  39 1.000000        NA
  8  1.000000        NA
  18 1.000000 0.0000000
  15 1.000000        NA
  9  1.000000        NA
  23 1.000000        NA
  24 1.000000        NA

    TOTCHG
Y         [,1]       [,2]
  2   1707.987  1582.1172
  7  12307.273 13139.1781
  1   1907.722  2336.5807
  0   1606.200  1031.3062
  4   3415.526  2264.4783
  3   2537.367  1844.2687
  5   5372.500  3914.7095
  12 10969.500  5882.4213
  6   8370.500  6564.3777
  41 29188.000         NA
  17 12042.000         NA
  10  5615.000         NA
  39 26356.000         NA
  8   5014.000         NA
  18 11167.000   732.5626
  15  8631.000         NA
  9  16520.000         NA
  23 13112.000         NA
  24 13040.000         NA

    APRDRG
Y        [,1]      [,2]
```

```
  2  620.9018 150.12787
  7  678.5455 164.93536
  1  607.6582 236.69741
  0  578.8000 228.39602
  4  629.3421 180.07090
  3  605.1735 168.36718
  5  687.5714 112.00049
  12 448.5000 440.52752
  6  557.1250 298.13872
  41 602.0000        NA
  17 633.0000        NA
  10 754.0000        NA
  39 421.0000        NA
  8  640.0000        NA
  18 689.5000  89.80256
  15 614.0000        NA
  9  225.0000        NA
  23 614.0000        NA
  24 863.0000        NA


>
```

The above mention are the conditional probabilities for Length of Stay(LOS).The Apriori Probabilties for LOS is also mention above

```
summary(tree_model)
Call:
rpart(formula = LOS ~ ., data = health_data, method = "class")
  n= 500

          CP nsplit rel error    xerror       xstd
1 0.20289855      0 1.0000000 1.0000000 0.04028881
2 0.15942029      1 0.7971014 0.8007246 0.04023501
3 0.03623188      2 0.6376812 0.6449275 0.03879215
4 0.02355072      3 0.6014493 0.6159420 0.03837844
5 0.01811594      5 0.5543478 0.6231884 0.03848638
6 0.01449275      7 0.5181159 0.6050725 0.03821081
7 0.01000000      8 0.5036232 0.5652174 0.03753595

Variable importance
TOTCHG APRDRG    AGE FEMALE
    71     18      9      1

Node number 1: 500 observations,    complexity param=0.2028986
  predicted class=2  expected loss=0.552  P(node) =1
    class counts:   224    11    79    15    38    98    14     2     8
1     1     1     1     1     2     1     1     1     1
   probabilities: 0.448 0.022 0.158 0.030 0.076 0.196 0.028 0.004 0.016 0.00
2 0.002 0.002 0.002 0.002 0.004 0.002 0.002 0.002 0.002
  left son=2 (265 obs) right son=3 (235 obs)
```

```
  Primary splits:
      TOTCHG < 1653.5 to the left,  improve=62.5164100, (0 missing)
      AGE    < 0.5    to the left,  improve=21.3989600, (0 missing)
      APRDRG < 675    to the left,  improve=14.4239300, (0 missing)
      FEMALE < 0.5    to the left,  improve= 1.1393180, (0 missing)
      RACE   < 1.5    to the right, improve= 0.7992354, (1 missing)
  Surrogate splits:
      APRDRG < 639.5  to the right, agree=0.674, adj=0.306, (0 split)
      AGE    < 0.5    to the left,  agree=0.568, adj=0.081, (0 split)
      FEMALE < 0.5    to the right, agree=0.542, adj=0.026, (0 split)

Node number 2: 265 observations,    complexity param=0.1594203
  predicted class=2  expected loss=0.2792453  P(node) =0.53
    class counts:   191     0    56     9     0     9     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.721 0.000 0.211 0.034 0.000 0.034 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
  left son=4 (210 obs) right son=5 (55 obs)
  Primary splits:
      TOTCHG < 1058   to the right, improve=60.307880, (0 missing)
      AGE    < 0.5    to the left,  improve=26.807390, (0 missing)
      APRDRG < 695.5  to the left,  improve=20.899350, (0 missing)
      FEMALE < 0.5    to the left,  improve= 1.764631, (0 missing)
      RACE   < 1.5    to the right, improve= 0.647964, (1 missing)
  Surrogate splits:
      APRDRG < 754.5  to the left,  agree=0.823, adj=0.145, (0 split)
      AGE    < 13.5   to the left,  agree=0.819, adj=0.127, (0 split)

Node number 3: 235 observations,    complexity param=0.03623188
  predicted class=3  expected loss=0.6212766  P(node) =0.47
    class counts:    33    11    23     6    38    89    14     2     8
1     1     1     1     1     2     1     1     1     1
   probabilities: 0.140 0.047 0.098 0.026 0.162 0.379 0.060 0.009 0.034 0.00
4 0.004 0.004 0.004 0.004 0.009 0.004 0.004 0.004 0.004
  left son=6 (94 obs) right son=7 (141 obs)
  Primary splits:
      TOTCHG < 2229   to the left,  improve=21.218440, (0 missing)
      APRDRG < 620    to the left,  improve= 9.507665, (0 missing)
      AGE    < 0.5    to the right, improve= 8.423164, (0 missing)
      FEMALE < 0.5    to the left,  improve= 2.473021, (0 missing)
      RACE   < 1.5    to the right, improve= 1.011006, (0 missing)
  Surrogate splits:
      AGE    < 0.5    to the left,  agree=0.672, adj=0.181, (0 split)
      APRDRG < 637.5  to the right, agree=0.630, adj=0.074, (0 split)

Node number 4: 210 observations
  predicted class=2  expected loss=0.1  P(node) =0.42
    class counts:   189     0    10     2     0     9     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.900 0.000 0.048 0.010 0.000 0.043 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 5: 55 observations
  predicted class=1  expected loss=0.1636364  P(node) =0.11
```

```
    class counts:     2     0    46     7     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.036 0.000 0.836 0.127 0.000 0.000 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 6: 94 observations
  predicted class=3  expected loss=0.2765957  P(node) =0.188
    class counts:    13     0     6     0     7    68     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.138 0.000 0.064 0.000 0.074 0.723 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 7: 141 observations,    complexity param=0.02355072
  predicted class=4  expected loss=0.7801418  P(node) =0.282
    class counts:    20    11    17     6    31    21    14     2     8
1     1     1     1     1     2     1     1     1     1
   probabilities: 0.142 0.078 0.121 0.043 0.220 0.149 0.099 0.014 0.057 0.00
7 0.007 0.007 0.007 0.007 0.014 0.007 0.007 0.007 0.007
  left son=14 (26 obs) right son=15 (115 obs)
  Primary splits:
      TOTCHG < 2646   to the left,  improve=6.1443720, (0 missing)
      APRDRG < 560.5  to the left,  improve=6.0891220, (0 missing)
      FEMALE < 0.5    to the left,  improve=2.4946230, (0 missing)
      AGE    < 0.5    to the right, improve=2.0253920, (0 missing)
      RACE   < 1.5    to the right, improve=0.4269534, (0 missing)

Node number 14: 26 observations
  predicted class=4  expected loss=0.3846154  P(node) =0.052
    class counts:     2     0     1     3    16     4     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.077 0.000 0.038 0.115 0.615 0.154 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 15: 115 observations,    complexity param=0.02355072
  predicted class=2  expected loss=0.8434783  P(node) =0.23
    class counts:    18    11    16     3    15    17    14     2     8
1     1     1     1     1     2     1     1     1     1
   probabilities: 0.157 0.096 0.139 0.026 0.130 0.148 0.122 0.017 0.070 0.00
9 0.009 0.009 0.009 0.009 0.017 0.009 0.009 0.009 0.009
  left son=30 (50 obs) right son=31 (65 obs)
  Primary splits:
      APRDRG < 591.5  to the left,  improve=5.0985950, (0 missing)
      TOTCHG < 3624.5 to the left,  improve=2.2984840, (0 missing)
      FEMALE < 0.5    to the left,  improve=1.8350810, (0 missing)
      AGE    < 0.5    to the right, improve=1.7617800, (0 missing)
      RACE   < 1.5    to the right, improve=0.7864504, (0 missing)
  Surrogate splits:
      TOTCHG < 6377   to the right, agree=0.687, adj=0.28, (0 split)
      FEMALE < 0.5    to the left,  agree=0.670, adj=0.24, (0 split)
      AGE    < 0.5    to the right, agree=0.635, adj=0.16, (0 split)

Node number 30: 50 observations,    complexity param=0.01449275
  predicted class=2  expected loss=0.7  P(node) =0.1
    class counts:    15     1    12     3     3    10     1     1     2
0     0     0     1     0     0     0     1     0     0
```

```
   probabilities: 0.300 0.020 0.240 0.060 0.060 0.200 0.020 0.020 0.040 0.00
0 0.000 0.000 0.020 0.000 0.000 0.000 0.020 0.000 0.000
  left son=60 (15 obs) right son=61 (35 obs)
  Primary splits:
      APRDRG < 55     to the left,  improve=2.4800000, (0 missing)
      TOTCHG < 13099  to the left,  improve=2.0401330, (0 missing)
      FEMALE < 0.5    to the left,  improve=0.8431579, (0 missing)
      AGE    < 9.5    to the left,  improve=0.6532689, (0 missing)
  Surrogate splits:
      AGE < 15.5   to the right, agree=0.72, adj=0.067, (0 split)

Node number 31: 65 observations,    complexity param=0.01811594
  predicted class=5  expected loss=0.8  P(node) =0.13
    class counts:     3    10     4     0    12     7    13     1     6
1     1     1     0     1     2     1     0     1     1
   probabilities: 0.046 0.154 0.062 0.000 0.185 0.108 0.200 0.015 0.092 0.01
5 0.015 0.015 0.000 0.015 0.031 0.015 0.000 0.015 0.015
  left son=62 (27 obs) right son=63 (38 obs)
  Primary splits:
      TOTCHG < 3640.5 to the left,  improve=3.4300490, (0 missing)
      APRDRG < 715    to the right, improve=1.9434730, (0 missing)
      AGE    < 1.5    to the right, improve=1.6461250, (0 missing)
      FEMALE < 0.5    to the right, improve=0.4871795, (0 missing)
  Surrogate splits:
      RACE < 1.5    to the right, agree=0.631, adj=0.111, (0 split)

Node number 60: 15 observations
  predicted class=2  expected loss=0.5333333  P(node) =0.03
    class counts:     7     0     0     0     3     4     0     0     1
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.467 0.000 0.000 0.000 0.200 0.267 0.000 0.000 0.067 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 61: 35 observations
  predicted class=1  expected loss=0.6571429  P(node) =0.07
    class counts:     8     1    12     3     0     6     1     1     1
0     0     0     1     0     0     0     1     0     0
   probabilities: 0.229 0.029 0.343 0.086 0.000 0.171 0.029 0.029 0.029 0.00
0 0.000 0.000 0.029 0.000 0.000 0.000 0.029 0.000 0.000

Node number 62: 27 observations,    complexity param=0.01811594
  predicted class=4  expected loss=0.7037037  P(node) =0.054
    class counts:     3     0     2     0     8     6     8     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.111 0.000 0.074 0.000 0.296 0.222 0.296 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
  left son=124 (15 obs) right son=125 (12 obs)
  Primary splits:
      AGE    < 6      to the right, improve=3.3444440, (0 missing)
      APRDRG < 681.5  to the right, improve=3.3444440, (0 missing)
      TOTCHG < 3113   to the right, improve=1.2444440, (0 missing)
      FEMALE < 0.5    to the left,  improve=0.1777778, (0 missing)
  Surrogate splits:
      APRDRG < 681.5  to the right, agree=1.000, adj=1.000, (0 split)
      RACE   < 2.5    to the left,  agree=0.593, adj=0.083, (0 split)
```

```
      TOTCHG < 2890   to the right, agree=0.593, adj=0.083, (0 split)

Node number 63: 38 observations
  predicted class=7  expected loss=0.7368421  P(node) =0.076
    class counts:     0    10     2     0     4     1     5     1     6
1     1     1     0     1     2     1     0     1     1
   probabilities: 0.000 0.263 0.053 0.000 0.105 0.026 0.132 0.026 0.158 0.02
6 0.026 0.026 0.000 0.026 0.053 0.026 0.000 0.026 0.026

Node number 124: 15 observations
  predicted class=5  expected loss=0.5333333  P(node) =0.03
    class counts:     3     0     2     0     3     0     7     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.200 0.000 0.133 0.000 0.200 0.000 0.467 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 125: 12 observations
  predicted class=3  expected loss=0.5  P(node) =0.024
    class counts:     0     0     0     0     5     6     1     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.000 0.000 0.000 0.417 0.500 0.083 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000


>
```

From the above mention decision tree, it is clear that most important factors affecting the length of stay are Hospital Discharge Costs and All Patient Refined Diagnosis Related Groups

Q6) To perform a complete analysis, the agency wants to find the variable that mainly affects hospital costs.

Ans) Decision Tree Algorithm helps in analysing the variable that mainly affects the hospital costs. Below attached is the code in R which gives an overview of the decision tree algorithm

```
print("Healthcare cost Analysis")

health_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/Hospitals/main/Hospital.csv")

print(health_data)

View(health_data)

health_data$APRDRG<-sapply(health_data$APRDRG,factor)

mytree<-rpart(APRDRG~.,data=health_data,method="class")

print(mytree)
```

printcp(mytree)

plotcp(mytree)

summary(mytree)

Below attached are the screenshots of the decision tree

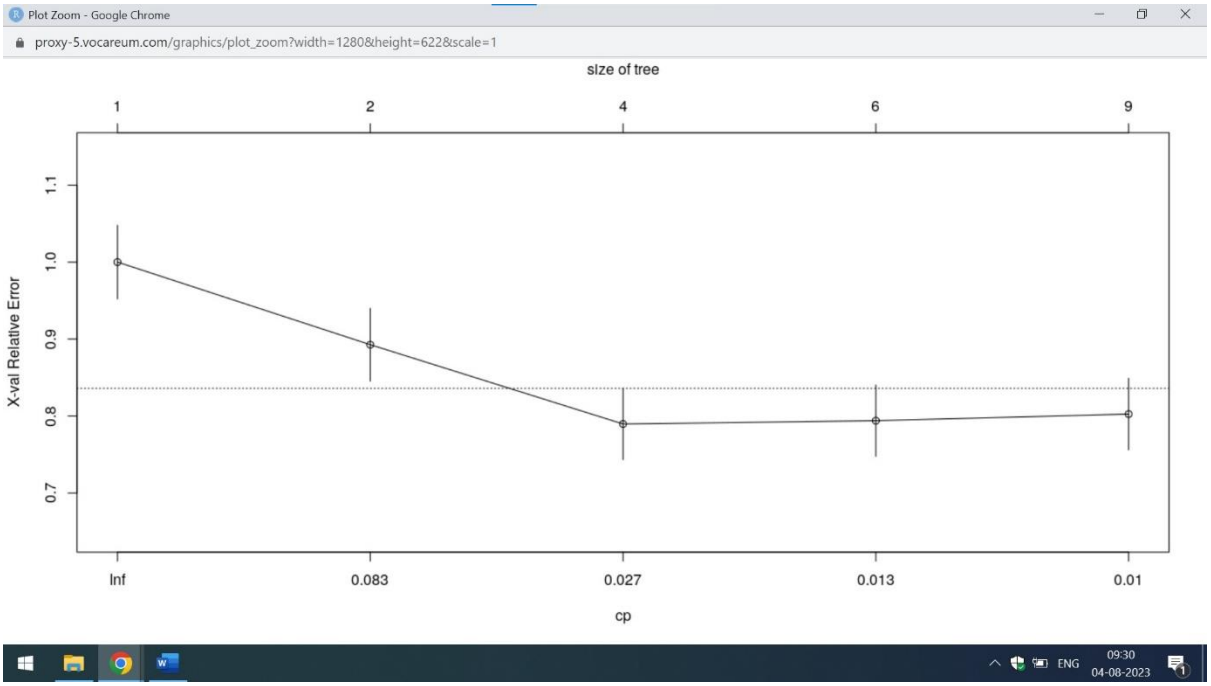
```
mytree<-rpart(APRDRG~.,data=health_data,method="class")
> print(mytree)
n= 500

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

  1) root 500 233 640 (0.004 0.072 0.004 0.04 0.074 0.006 0.028 0.006 0.002 0.0
12 0.006 0.002 0.01 0.002 0.004 0.002 0.004 0.026 0.002 0.02 0.004 0.002 0.002
0.004 0.002 0.008 0.002 0.53 0.008 0.002 0.002 0.006 0.008 0.012 0.006 0.002 0.
004 0.002 0.002 0.002 0.002 0.002 0.002 0.006 0.004 0.002 0.002 0.002 0.004 0.0
02 0.004 0.002 0.002 0.004 0.002 0.002 0.004 0.002 0.002 0.002 0.002 0.002 0.00
2)
    2) AGE>=0.5 193 156 754 (0.01 0.19 0.01 0.1 0.19 0.016 0.073 0.016 0.0052 0
.026 0.01 0.0052 0.021 0.0052 0.01 0.0052 0.01 0.067 0.0052 0.047 0.01 0.0052 0
.0052 0.01 0 0 0 0 0 0 0 0 0 0 0 0.0052 0.01 0 0 0.0052 0.0052 0.0052 0.0052 0
0 0.0052 0.0052 0.0052 0.01 0.0052 0.01 0.0052 0.0052 0.01 0.0052 0.0052 0.01 0
.0052 0.0052 0.0052 0.0052 0.0052 0)
      4) TOTCHG< 6377 157 120 754 (0.013 0.22 0 0.13 0.24 0.013 0.089 0.019 0.0
064 0.025 0.013 0.0064 0.025 0.0064 0.013 0 0.013 0.083 0 0.0064 0.0064 0 0.006
4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.0064 0 0.0064 0 0 0 0.0064 0 0 0.013 0 0.00
64 0 0 0.013 0 0.0064 0 0.0064 0 0 0 0.0064 0)
        8) TOTCHG< 1308 70  45 754 (0 0.16 0 0.11 0.36 0 0.086 0 0 0 0.029 0 0.
014 0 0 0 0.014 0.17 0 0 0 0 0.014 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0.029 0 0 0 0 0 0 0.014 0 0 0 0 0 0 0)
         16) AGE>=11.5 62  37 754 (0 0.11 0 0.11 0.4 0 0.097 0 0 0 0.016 0 0 0
0 0 0.016 0.19 0 0 0 0 0.016 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.016 0 0 0 0 0 0 0.016 0 0 0 0 0 0 0)
           32) AGE< 16.5 52  28 754 (0 0.077 0 0.077 0.46 0 0.096 0 0 0 0.019 0
0 0 0 0 0.019 0.23 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.019 0 0 0 0 0 0 0 0 0 0 0 0 0 0)
             64) LOS>=1.5 14   3 754 (0 0 0 0 0.79 0 0.14 0 0 0 0 0 0 0 0 0 0 0
.071 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0) *
             65) LOS< 1.5 38  25 754 (0 0.11 0 0.11 0.34 0 0.079 0 0 0 0.026 0
0 0 0 0 0.026 0.29 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.026 0 0 0 0 0 0 0 0 0 0 0 0 0 0)
              130) AGE>=13.5 29  17 754 (0 0.14 0 0.1 0.41 0 0.069 0 0 0 0.034
0 0 0 0 0 0.034 0.17 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0.034 0 0 0 0 0 0 0 0 0 0 0 0 0) *
              131) AGE< 13.5 9   3 755 (0 0 0 0.11 0.11 0 0.11 0 0 0 0 0 0 0 0
0 0 0.67 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0) *
           33) AGE>=16.5 10   7 753 (0 0.3 0 0.3 0.1 0 0.1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0.1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
.1 0 0 0 0 0 0 0) *
```

```
         17) AGE< 11.5 8   4 753 (0 0.5 0 0.12 0 0 0 0 0 0 0.12 0 0.12 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.12 0 0 0 0 0 0
0 0 0 0 0 0 0 0) *
        9) TOTCHG>=1308 87  63 753 (0.023 0.28 0 0.14 0.14 0.023 0.092 0.034 0.
011 0.046 0 0.011 0.034 0.011 0.023 0 0.011 0.011 0 0.011 0.011 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0.011 0 0.011 0 0 0 0.011 0 0 0 0 0.011 0 0 0.023 0 0 0 0.0
11 0 0 0 0.011 0)
         18) AGE>=10.5 75  51 753 (0.027 0.32 0 0.16 0.16 0.013 0.11 0.04 0.013
0 0 0.013 0.013 0 0.027 0 0.013 0.013 0 0 0.013 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0.013 0 0.013 0 0 0 0.013 0 0 0 0 0 0 0 0.013 0 0 0 0 0 0 0 0.013 0) *
         19) AGE< 10.5 12   8 249 (0 0 0 0 0 0.083 0 0 0 0.33 0 0 0.17 0.083 0
0 0 0 0 0.083 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.083
0 0 0.083 0 0 0 0.083 0 0 0 0 0) *
      5) TOTCHG>=6377 36  28 53 (0 0.028 0.056 0 0 0.028 0 0 0 0.028 0 0 0 0 0
0.028 0 0 0.028 0.22 0.028 0.028 0 0.056 0 0 0 0 0 0 0 0 0 0 0 0.028 0.056 0 0
0 0.028 0 0.028 0 0 0 0.028 0.028 0 0.028 0.028 0.028 0.028 0 0.028 0 0.056 0 0
.028 0.028 0.028 0 0) *
    3) AGE< 0.5 307  40 640 (0 0 0 0 0 0 0 0 0 0.0033 0.0033 0 0.0033 0 0 0 0 0
0 0.0033 0 0 0 0 0.0033 0.013 0.0033 0.87 0.013 0.0033 0.0033 0.0098 0.013 0.02
0.0098 0 0 0.0033 0.0033 0 0 0 0 0.0098 0.0065 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0.0033) *
```

The above attached is the screenshot of the decision tree which gives an overview of the important variables for the analysis

```
printcp(mytree)

Classification tree:
rpart(formula = APRDRG ~ ., data = health_data, method = "class")

Variables actually used in tree construction:
[1] AGE    LOS    TOTCHG

Root node error: 233/500 = 0.466

n= 500

        CP nsplit rel error  xerror     xstd
1 0.158798      0   1.00000 1.00000 0.047873
2 0.042918      1   0.84120 0.89270 0.047302
3 0.017167      3   0.75536 0.78970 0.046282
4 0.010014      5   0.72103 0.79399 0.046334
5 0.010000      8   0.69099 0.80258 0.046436


>
```

From the above mention screenshot,it is clear that variables used in tree construction are AGE,LOS and TOTCHG.Hence, the variable that mainly affects the hospital costs are Age,Length of stay in days(LOS) and Hospital Discharge Costs (Totchg)



This screenshot gives an idea when the size of tree then relative error first decreases and then b ecome stagnant to specific point

```
summary(mytree)
Call:
rpart(formula = APRDRG ~ ., data = health_data, method = "class")
  n= 500

          CP nsplit rel error    xerror       xstd
1 0.15879828      0 1.0000000 1.0000000 0.04787322
2 0.04291845      1 0.8412017 0.8927039 0.04730229
3 0.01716738      3 0.7553648 0.7896996 0.04628194
4 0.01001431      5 0.7210300 0.7939914 0.04633405
5 0.01000000      8 0.6909871 0.8025751 0.04643571

Variable importance
   AGE    LOS TOTCHG
    64     20     16

Node number 1: 500 observations,    complexity param=0.1587983
  predicted class=640  expected loss=0.466  P(node) =1
    class counts:     2    36     2    20    37     3    14     3     1     6
3     1     5     1     2     1     2    13     1    10     2     1     1     2
1     4     1   267     4     1     1     3     4     6     3     1     2     1
1     1     1     1     1     3     2     1     1     1     2     1     2     1
1     2     1     1     2     1     1     1     1     1     1
```

```
   probabilities: 0.004 0.072 0.004 0.040 0.074 0.006 0.028 0.006 0.002 0.012 0
.006 0.002 0.010 0.002 0.004 0.002 0.004 0.026 0.002 0.020 0.004 0.002 0.002 0.
004 0.002 0.008 0.002 0.534 0.008 0.002 0.002 0.006 0.008 0.012 0.006 0.002 0.0
04 0.002 0.002 0.002 0.002 0.002 0.002 0.006 0.004 0.002 0.002 0.002 0.004 0.00
2 0.004 0.002 0.002 0.004 0.002 0.002 0.004 0.002 0.002 0.002 0.002 0.002 0.002
  left son=2 (193 obs) right son=3 (307 obs)
  Primary splits:
      AGE    < 0.5    to the right, improve=101.301800, (0 missing)
      TOTCHG < 2106   to the left,  improve= 25.191560, (0 missing)
      LOS    < 1.5    to the left,  improve= 22.945390, (0 missing)
      FEMALE < 0.5    to the right, improve=  5.062709, (0 missing)
      RACE   < 1.5    to the right, improve=  2.320454, (1 missing)
  Surrogate splits:
      LOS    < 1.5    to the left,  agree=0.730, adj=0.301, (0 split)
      TOTCHG < 2853.5 to the right, agree=0.678, adj=0.166, (0 split)

Node number 2: 193 observations,    complexity param=0.04291845
  predicted class=754  expected loss=0.8082902  P(node) =0.386
    class counts:     2    36     2    20    37     3    14     3     1     5
2     1     4     1     2     1     2    13     1     9     2     1     1     2
0     0     0     0     0     0     0     0     0     0     1     2     0
0     1     1     1     1     0     0     1     1     1     2     1     2     1
1     2     1     1     2     1     1     1     1     1     0
   probabilities: 0.010 0.187 0.010 0.104 0.192 0.016 0.073 0.016 0.005 0.026 0
.010 0.005 0.021 0.005 0.010 0.005 0.010 0.067 0.005 0.047 0.010 0.005 0.005 0.
010 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.005 0.0
10 0.000 0.000 0.005 0.005 0.005 0.005 0.000 0.000 0.005 0.005 0.005 0.010 0.00
5 0.010 0.005 0.005 0.010 0.005 0.005 0.010 0.005 0.005 0.005 0.005 0.005 0.000
  left son=4 (157 obs) right son=5 (36 obs)
  Primary splits:
      TOTCHG < 6377   to the left,  improve=5.8193280, (0 missing)
      AGE    < 9.5    to the right, improve=4.9418450, (0 missing)
      FEMALE < 0.5    to the right, improve=2.5924420, (0 missing)
      LOS    < 1.5    to the right, improve=2.3146360, (0 missing)
      RACE   < 1.5    to the left,  improve=0.5973253, (0 missing)
  Surrogate splits:
      AGE < 6.5    to the right, agree=0.834, adj=0.111, (0 split)
      LOS < 6.5    to the left,  agree=0.829, adj=0.083, (0 split)

Node number 3: 307 observations
  predicted class=640  expected loss=0.1302932  P(node) =0.614
    class counts:     0     0     0     0     0     0     0     0     0     1
1     0     1     0     0     0     0     0     0     1     0     0     0     0
1     4     1   267     4     1     1     3     4     6     3     0     0     1
1     0     0     0     0     3     2     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     1
   probabilities: 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0
.003 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.000 0.000 0.000 0.
000 0.003 0.013 0.003 0.870 0.013 0.003 0.003 0.010 0.013 0.020 0.010 0.000 0.0
00 0.003 0.003 0.000 0.000 0.000 0.000 0.010 0.007 0.000 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.003

Node number 4: 157 observations,    complexity param=0.04291845
  predicted class=754  expected loss=0.7643312  P(node) =0.314
    class counts:     2    35     0    20    37     2    14     3     1     4
2     1     4     1     2     0     2    13     0     1     1     0     1     0
```

```
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     1     0     1     0     0     0     1     0     0     2     0     1     0
0     2     0     1     0     1     0     0     0     1     0
   probabilities: 0.013 0.223 0.000 0.127 0.236 0.013 0.089 0.019 0.006 0.025 0
.013 0.006 0.025 0.006 0.013 0.000 0.013 0.083 0.000 0.006 0.006 0.000 0.006 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.006 0.000 0.006 0.000 0.000 0.000 0.006 0.000 0.000 0.013 0.00
0 0.006 0.000 0.000 0.013 0.000 0.006 0.000 0.006 0.000 0.000 0.000 0.006 0.000
  left son=8 (70 obs) right son=9 (87 obs)
  Primary splits:
      TOTCHG < 1308   to the left,  improve=3.787893, (0 missing)
      AGE    < 9      to the right, improve=3.775906, (0 missing)
      LOS    < 1.5    to the right, improve=2.481353, (0 missing)
      FEMALE < 0.5    to the right, improve=1.772522, (0 missing)
  Surrogate splits:
      LOS  < 1.5    to the left,  agree=0.764, adj=0.471, (0 split)
      RACE < 3      to the right, agree=0.561, adj=0.014, (0 split)

Node number 5: 36 observations
  predicted class=53   expected loss=0.7777778  P(node) =0.072
    class counts:     0     1     2     0     0     1     0     0     0     1
0     0     0     0     0     1     0     0     1     8     1     1     0     2
0     0     0     0     0     0     0     0     0     0     0     1     2     0
0     0     1     0     1     0     0     0     1     1     0     1     1     1
1     0     1     0     2     0     1     1     1     0     0
   probabilities: 0.000 0.028 0.056 0.000 0.000 0.028 0.000 0.000 0.000 0.028 0
.000 0.000 0.000 0.000 0.000 0.000 0.028 0.000 0.000 0.028 0.222 0.028 0.028 0.000 0.
056 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.028 0.0
56 0.000 0.000 0.000 0.028 0.000 0.028 0.000 0.000 0.000 0.028 0.028 0.000 0.02
8 0.028 0.028 0.028 0.000 0.028 0.000 0.056 0.000 0.028 0.028 0.028 0.000 0.000

Node number 8: 70 observations,    complexity param=0.01716738
  predicted class=754  expected loss=0.6428571  P(node) =0.14
    class counts:     0    11     0     8    25     0     6     0     0     0
2     0     1     0     0     0     1    12     0     0     0     0     1     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     2     0     0     0
0     0     0     1     0     0     0     0     0     0     0
   probabilities: 0.000 0.157 0.000 0.114 0.357 0.000 0.086 0.000 0.000 0.000 0
.029 0.000 0.014 0.000 0.000 0.000 0.014 0.171 0.000 0.000 0.000 0.000 0.014 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.029 0.00
0 0.000 0.000 0.000 0.000 0.000 0.014 0.000 0.000 0.000 0.000 0.000 0.000 0.000
  left son=16 (62 obs) right son=17 (8 obs)
  Primary splits:
      AGE    < 11.5   to the right, improve=2.8308760, (0 missing)
      TOTCHG < 1163.5 to the right, improve=1.9726080, (0 missing)
      LOS    < 1.5    to the right, improve=1.8301930, (0 missing)
      FEMALE < 0.5    to the right, improve=0.5275946, (0 missing)

Node number 9: 87 observations,    complexity param=0.01716738
  predicted class=753  expected loss=0.7241379  P(node) =0.174
    class counts:     2    24     0    12    12     2     8     3     1     4
0     1     3     1     2     0     1     1     0     1     1     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
```

```
0     1     0     1     0     0     0     1     0     0     0     0     1     0
0     2     0     0     0     1     0     0     0     1     0
   probabilities: 0.023 0.276 0.000 0.138 0.138 0.023 0.092 0.034 0.011 0.046 0
.000 0.011 0.034 0.011 0.023 0.000 0.011 0.011 0.000 0.011 0.011 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.011 0.000 0.011 0.000 0.000 0.000 0.011 0.000 0.000 0.000 0.00
0 0.011 0.000 0.000 0.023 0.000 0.000 0.000 0.011 0.000 0.000 0.000 0.011 0.000
  left son=18 (75 obs) right son=19 (12 obs)
  Primary splits:
      AGE    < 10.5   to the right, improve=3.535862, (0 missing)
      LOS    < 1.5    to the right, improve=3.101949, (0 missing)
      TOTCHG < 2100.5 to the left,  improve=2.854809, (0 missing)
      FEMALE < 0.5    to the right, improve=1.712604, (0 missing)

Node number 16: 62 observations,    complexity param=0.01001431
  predicted class=754  expected loss=0.5967742  P(node) =0.124
    class counts:     0     7     0     7    25     0     6     0     0     0
1     0     0     0     0     0     1    12     0     0     0     0     1     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     1     0     0     0
0     0     0     1     0     0     0     0     0     0     0
   probabilities: 0.000 0.113 0.000 0.113 0.403 0.000 0.097 0.000 0.000 0.000 0
.016 0.000 0.000 0.000 0.000 0.000 0.016 0.194 0.000 0.000 0.000 0.000 0.016 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.016 0.00
0 0.000 0.000 0.000 0.000 0.000 0.016 0.000 0.000 0.000 0.000 0.000 0.000 0.000
  left son=32 (52 obs) right son=33 (10 obs)
  Primary splits:
      AGE    < 16.5   to the left,  improve=2.5548390, (0 missing)
      LOS    < 1.5    to the right, improve=2.5287520, (0 missing)
      TOTCHG < 1141   to the right, improve=2.2679350, (0 missing)
      FEMALE < 0.5    to the right, improve=0.5669599, (0 missing)

Node number 17: 8 observations
  predicted class=753  expected loss=0.5  P(node) =0.016
    class counts:     0     4     0     1     0     0     0     0     0     0
1     0     1     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     1     0     0     0
0     0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.500 0.000 0.125 0.000 0.000 0.000 0.000 0.000 0.000 0
.125 0.000 0.125 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.125 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 18: 75 observations
  predicted class=753  expected loss=0.68  P(node) =0.15
    class counts:     2    24     0    12    12     1     8     3     1     0
0     1     1     0     2     0     1     1     0     0     1     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     1     0     1     0     0     0     1     0     0     0     0     0     0
0     1     0     0     0     0     0     0     0     1     0
   probabilities: 0.027 0.320 0.000 0.160 0.160 0.013 0.107 0.040 0.013 0.000 0
.000 0.013 0.013 0.000 0.027 0.000 0.013 0.013 0.000 0.000 0.013 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
```

```
00 0.000 0.000 0.013 0.000 0.013 0.000 0.000 0.000 0.013 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.013 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.013 0.000

Node number 19: 12 observations
  predicted class=249  expected loss=0.6666667  P(node) =0.024
    class counts:     0     0     0     0     0     1     0     0     0     4
0     0     2     1     0     0     0     0     0     1     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     1     0     0
0     1     0     0     0     1     0     0     0     0     0
   probabilities: 0.000 0.000 0.000 0.000 0.000 0.083 0.000 0.000 0.000 0.333 0
.000 0.000 0.167 0.083 0.000 0.000 0.000 0.000 0.000 0.083 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.00
0 0.083 0.000 0.000 0.083 0.000 0.000 0.000 0.083 0.000 0.000 0.000 0.000 0.000

Node number 32: 52 observations,    complexity param=0.01001431
  predicted class=754  expected loss=0.5384615  P(node) =0.104
    class counts:     0     4     0     4    24     0     5     0     0     0
1     0     0     0     0     0     1    12     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     1     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.077 0.000 0.077 0.462 0.000 0.096 0.000 0.000 0.000 0
.019 0.000 0.000 0.000 0.000 0.000 0.019 0.231 0.000 0.000 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.019 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
  left son=64 (14 obs) right son=65 (38 obs)
  Primary splits:
      LOS    < 1.5    to the right, improve=2.7894740, (0 missing)
      TOTCHG < 1141   to the right, improve=2.4666670, (0 missing)
      AGE    < 14.5   to the right, improve=0.6474074, (0 missing)
      FEMALE < 0.5    to the right, improve=0.4005168, (0 missing)
  Surrogate splits:
      TOTCHG < 1141   to the right, agree=0.942, adj=0.786, (0 split)
      RACE   < 2.5    to the right, agree=0.769, adj=0.143, (0 split)

Node number 33: 10 observations
  predicted class=753  expected loss=0.7  P(node) =0.02
    class counts:     0     3     0     3     1     0     1     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     1     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     1     0     0     0     0     0     0     0
   probabilities: 0.000 0.300 0.000 0.300 0.100 0.000 0.100 0.000 0.000 0.000 0
.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.100 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.100 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 64: 14 observations
  predicted class=754  expected loss=0.2142857  P(node) =0.028
    class counts:     0     0     0     0    11     0     2     0     0     0
0     0     0     0     0     0     0     1     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
```

```
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.000 0.000 0.000 0.786 0.000 0.143 0.000 0.000 0.000 0
.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.071 0.000 0.000 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 65: 38 observations,    complexity param=0.01001431
  predicted class=754  expected loss=0.6578947  P(node) =0.076
    class counts:     0     4     0     4    13     0     3     0     0     0
1     0     0     0     0     0     1    11     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     1     0     0     0
0     0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.105 0.000 0.105 0.342 0.000 0.079 0.000 0.000 0.000 0
.026 0.000 0.000 0.000 0.000 0.000 0.026 0.289 0.000 0.000 0.000 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.026 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
  left son=130 (29 obs) right son=131 (9 obs)
  Primary splits:
      AGE    < 13.5   to the right, improve=2.474894, (0 missing)
      TOTCHG < 692.5  to the right, improve=1.366082, (0 missing)
      FEMALE < 0.5    to the right, improve=0.408683, (0 missing)
  Surrogate splits:
      FEMALE < 0.5    to the right, agree=0.789, adj=0.111, (0 split)

Node number 130: 29 observations
  predicted class=754  expected loss=0.5862069  P(node) =0.058
    class counts:     0     4     0     3    12     0     2     0     0     0
1     0     0     0     0     0     1     5     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     1     0     0     0
0     0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.138 0.000 0.103 0.414 0.000 0.069 0.000 0.000 0.000 0
.034 0.000 0.000 0.000 0.000 0.000 0.034 0.172 0.000 0.000 0.000 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.034 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Node number 131: 9 observations
  predicted class=755  expected loss=0.3333333  P(node) =0.018
    class counts:     0     0     0     1     1     0     1     0     0     0
0     0     0     0     0     0     0     6     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0     0     0     0     0
0     0     0     0     0     0     0     0     0     0
   probabilities: 0.000 0.000 0.000 0.111 0.111 0.000 0.111 0.000 0.000 0.000 0
.000 0.000 0.000 0.000 0.000 0.000 0.667 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.
000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.0
00 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.00
0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```

The most important factor for the hospital costs are the Age of the patient discharged,Length o f stay in days(LOS) and Hospital Discharge Costs(Totchg)