|  |
| --- |
| > #### mini projrct of TITANIC #### > ### import the data set > library(readr)> train <- read\_csv("D:/R DOC/New folder/train.csv")> View(train) ## view the dataset> summary(train) ### summary of the datasetPassengerId Survived Pclass Name SexMin. : 1.0 Min. :0.0000 Min. :1.000 Length:891 Length:8911st Qu.:223.5 1st Qu.:0.0000 1st Qu.:2.000 Class :character Class :characterMedian :446.0 Median :0.0000 Median :3.000 Mode :character Mode :characterMean :446.0 Mean :0.3838 Mean :2.3093rd Qu.:668.5 3rd Qu.:1.0000 3rd Qu.:3.000Max. :891.0 Max. :1.0000 Max. :3.000Age SibSp Parch Ticket FareMin. : 0.42 Min. :0.000 Min. :0.0000 Length:891 Min. : 0.001st Qu.:20.12 1st Qu.:0.000 1st Qu.:0.0000 Class :character 1st Qu.: 7.91Median :28.00 Median :0.000 Median :0.0000 Mode :character Median : 14.45Mean :29.70 Mean :0.523 Mean :0.3816 Mean : 32.203rd Qu.:38.00 3rd Qu.:1.000 3rd Qu.:0.0000 3rd Qu.: 31.00Max. :80.00 Max. :8.000 Max. :6.0000 Max. :512.33NA's :177Cabin EmbarkedLength:891 Length:891Class :character Class :characterMode :character Mode :characterstr(train) ##### structure of the datasetClasses ‘tbl\_df’, ‘tbl’ and 'data.frame': 891 obs. of 12 variables:$ PassengerId: int 1 2 3 4 5 6 7 8 9 10 ...$ Survived : int 0 1 1 1 0 0 0 0 1 1 ...$ Pclass : int 3 1 3 1 3 3 1 3 3 2 ...$ Name : chr "Braund, Mr. Owen Harris" "Cumings, Mrs. John Bradley (Florence Briggs Thayer)" "Heikkinen, Miss. Laina" "Futrelle, Mrs. Jacques Heath (Lily May Peel)" ...$ Sex : chr "male" "female" "female" "female" ...$ Age : num 22 38 26 35 35 NA 54 2 27 14 ...$ SibSp : int 1 1 0 1 0 0 0 3 0 1 ...$ Parch : int 0 0 0 0 0 0 0 1 2 0 ...$ Ticket : chr "A/5 21171" "PC 17599" "STON/O2. 3101282" "113803" ...$ Fare : num 7.25 71.28 7.92 53.1 8.05 ...$ Cabin : chr NA "C85" NA "C123" ...$ Embarked : chr "S" "C" "S" "S" ...- attr(\*, "spec")=List of 2..$ cols :List of 12.. ..$ PassengerId: list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Survived : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Pclass : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Name : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Sex : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Age : list().. .. ..- attr(\*, "class")= chr "collector\_double" "collector".. ..$ SibSp : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Parch : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Ticket : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Fare : list().. .. ..- attr(\*, "class")= chr "collector\_double" "collector".. ..$ Cabin : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Embarked : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector"..$ default: list().. ..- attr(\*, "class")= chr "collector\_guess" "collector"..- attr(\*, "class")= chr "col\_spec"> ##########################################>> ### how many prople are survived> table(train$Survived)0 1549 342> #it shows the Suriverd or not Suriverd in the traning dataset> prop.table(table(train$Survived))0 10.6161616 0.3838384> ## it shows the percentage rate Suriverd or not Suriverd in the traning dataset>> ##########################################> ##import the test dataset> test <- read\_csv("D:/R DOC/New folder/test.csv")> View(test)> ### add variable in test survived> test$survived<-rep(0,418)> str(test)Classes ‘tbl\_df’, ‘tbl’ and 'data.frame': 418 obs. of 12 variables:$ PassengerId: int 892 893 894 895 896 897 898 899 900 901 ...$ Pclass : int 3 3 2 3 3 3 3 2 3 3 ...$ Name : chr "Kelly, Mr. James" "Wilkes, Mrs. James (Ellen Needs)" "Myles, Mr. Thomas Francis" "Wirz, Mr. Albert" ...$ Sex : chr "male" "female" "male" "male" ...$ Age : num 34.5 47 62 27 22 14 30 26 18 21 ...$ SibSp : int 0 1 0 0 1 0 0 1 0 2 ...$ Parch : int 0 0 0 0 1 0 0 1 0 0 ...$ Ticket : chr "330911" "363272" "240276" "315154" ...$ Fare : num 7.83 7 9.69 8.66 12.29 ...$ Cabin : chr NA NA NA NA ...$ Embarked : chr "Q" "S" "Q" "S" ...$ survived : num 0 0 0 0 0 0 0 0 0 0 ...- attr(\*, "spec")=List of 2..$ cols :List of 11.. ..$ PassengerId: list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Pclass : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Name : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Sex : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Age : list().. .. ..- attr(\*, "class")= chr "collector\_double" "collector".. ..$ SibSp : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Parch : list().. .. ..- attr(\*, "class")= chr "collector\_integer" "collector".. ..$ Ticket : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Fare : list().. .. ..- attr(\*, "class")= chr "collector\_double" "collector".. ..$ Cabin : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector".. ..$ Embarked : list().. .. ..- attr(\*, "class")= chr "collector\_character" "collector"..$ default: list().. ..- attr(\*, "class")= chr "collector\_guess" "collector"..- attr(\*, "class")= chr "col\_spec"> class(train$Sex)[1] "character"> as.factor(train$Sex)->train$Sex> class(train$Sex)[1] "factor"> summary(train$Sex)female male314 577> ### how many female are survived in total> table(train$Sex,train$Survived)0 1female 81 233male 468 109> ## prepotion> prop.table(table(train$Sex,train$Survived))0 1female 0.09090909 0.26150393male 0.52525253 0.12233446> ### row wise propotion= 1> ### column wise porpotion =2> prop.table(table(train$Sex,train$Survived),1)#### ROW WISE0 1female 0.2579618 0.7420382male 0.8110919 0.1889081> ## colunm> prop.table(table(train$Sex,train$Survived),2)#### COULUM WISE0 1female 0.1475410 0.6812865male 0.8524590 0.3187135> ### put the 1 at female potion survived and accuracy increces to 74> test$survived[test$Sex== "female"]<-1> #### AGE #############> summary(train$Age)Min. 1st Qu. Median Mean 3rd Qu. Max. NA's0.42 20.12 28.00 29.70 38.00 80.00 177> train$child<-0> train$child[train$Age<18]<-1> ###> aggregate(Survived~child+Sex,data = train,FUN = sum)child Sex Survived1 0 female 1952 1 female 383 0 male 864 1 male 23 > aggregate(Survived~child+Sex,data = train,FUN = length) child Sex Survived1 0 female 2592 1 female 553 0 male 5194 1 male 58> aggregate(Survived~child+Sex,data = train,FUN = function(x){sum(x)/length(x)})child Sex Survived1 0 female 0.75289582 1 female 0.69090913 0 male 0.16570334 1 male 0.3965517>> ####> ### fare> train$Fare2<-"30+"> train$Fare2[train$Fare<30 & train$Fare>=20]<-"20-30"> train$Fare2[train$Fare<20 & train$Fare>=10]<-"10-20"> train$Fare2[train$Fare<10]<-"<10"> aggregate(Survived~Fare2+Pclass+Sex ,data=train,FUN=function(x){sum(x)/length(x)})Fare2 Pclass Sex Survived1 20-30 1 female 0.83333332 30+ 1 female 0.97727273 10-20 2 female 0.91428574 20-30 2 female 0.90000005 30+ 2 female 1.00000006 <10 3 female 0.59375007 10-20 3 female 0.58139538 20-30 3 female 0.33333339 30+ 3 female 0.125000010 <10 1 male 0.000000011 20-30 1 male 0.400000012 30+ 1 male 0.383720913 <10 2 male 0.000000014 10-20 2 male 0.158730215 20-30 2 male 0.160000016 30+ 2 male 0.214285717 <10 3 male 0.111538518 10-20 3 male 0.236842119 20-30 3 male 0.125000020 30+ 3 male 0.2400000>> test$Survived<-0> test$survived[test$Sex=="female"]<-1> test$survived[test$Sex=="female"& test$Pclass==3 & test$Fare>=20]<-0>> ########> ######## DECISON TREE ######> library(rpart)> fit <- rpart(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked,+ data=train,+ method="class")> plot(fit)> text(fit) |
|  |
| |  | | --- | | > | |

