```
#Leitura dos arquivos
titanic.train <- read.csv(file = "train.csv", stringsAsFactors = FALSE, header = TRUE)
titanic.test <- read.csv(file = "test.csv", stringsAsFactors = FALSE, header = TRUE)

#Adicionando o campo IsTrainSet e definindo como TRUE o arquivo train e FALSE o arquivo test
titanic.train$IsTrainSet <- TRUE
titanic.test$IsTrainSet <- FALSE

#Adicionando o campo Survived no arquivo test e setando como not available
titanic.test$Survived <- NA

#Criando um df com o train e test
titanic.completo <- rbind(titanic.train, titanic.test)

#Alterando os valores vazios da coluna Embarked para "S"
titanic.completo[titanic.completo$Embarked == '', "Embarked"] <- 'S'

#Alterando os campos NA da coluna de idade para a média de idade dos passageiros
age.median <- median(titanic.completo$Age, na.rm = TRUE)
titanic.completo[is.na(titanic.completo$Age), "Age"] <- age.median

#Alterando os campos NA da coluna de tarifa para a média de tarifa dos passageiros
#fare.median <- median(titanic.completo$Fare, na.rm = TRUE)
#titanic.completo[is.na(titanic.completo$Fare), "Fare"] <- fare.median

#boxplot(titanic.completo$Fare)
#boxplot.stats(titanic.completo$Fare)

#Remover o maior valor da tarifa da consulta, pois somente 1 é muito acima e joga a média para cima
upper.whisker <- boxplot.stats(titanic.completo$Fare)$stats[5]
outlier.filter <- titanic.completo$Fare < upper.whisker
#titanic.completo[outlier.filter,]

fare.equation = "Fare ~ Pclass + Sex + Age + SibSp + Parch + Embarked"
fare.model <- lm(
  formula = fare.equation,
  data = titanic.completo[outlier.filter,]
)

fare.row <- titanic.completo[
  is.na(titanic.completo$Fare),
  c("Pclass", "Sex", "Age", "SibSp", "Parch", "Embarked")
]

fare.predictions <- predict(fare.model, newdata = fare.row)
titanic.completo[is.na(titanic.completo$Fare), "Fare"] <- fare.predictions
#titanic.completo[is.na(titanic.completo$Fare), "Fare"]

#Valores categóricos
titanic.completo$Pclass <- as.factor(titanic.completo$Pclass)
titanic.completo$Sex <- as.factor(titanic.completo$Sex)
titanic.completo$Embarked <- as.factor(titanic.completo$Embarked)

#Devolver os valores para as tabelas originais
titanic.train <- titanic.completo[titanic.completo$IsTrainSet == TRUE,]
titanic.test <- titanic.completo[titanic.completo$IsTrainSet == FALSE,]

#Valores categóricos
titanic.train$Survived <- as.factor(titanic.train$Survived)

#Modelo
survived.equation <- "Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked"
survived.formula <- as.formula(survived.equation)
library(randomForest)
titanic.model <- randomForest(formula = survived.formula, data = titanic.train, ntree = 500, mtry = 3, nodesize =
0.01 * nrow(titanic.test))
#titanic.model

#Preparando arquivo para enviar no Kaggle
Survived <- predict(titanic.model, newdata = titanic.test)
PassengerId <- titanic.test$PassengerId
output.df <- as.data.frame(PassengerId)
output.df$Survived <- Survived
write.csv(output.df, file = "Kaggle_submission.csv", row.names = FALSE)
```