Load dataset:

titanic.dataset <- read.csv("C:/Users/Raymond Rochioli/Desktop/titanic dataset.csv")

Find the missing values and replace them with S. in embarked column:

titanic.dataset[titanic.dataset[,11] == "", 11] <- "S"

Calculate the mean of the Age column and use that value to populate the missing values:

mean(titanic.dataset$age, na.rm = T)

titanic.dataset[is.na(titanic.dataset[,5] ,5]<-29.88113

Fill these empty slots in *boat* column with a dummy value e.g. the string 'None' or 'NA'

titanic.dataset[titanic.dataset[,12] == "", 12]<-"NA"

Create a new column has\_cabin\_number which has 1 if there is a cabin number, and 0 otherwise.

titanic.dataset[paste("has\_cabin\_number", NULL, sep = "")] <- ifelse(titanic.dataset$cabin =="", 0, 1)

Save titanic\_clean file as a csv

write.table(titanic.dataset, file="titanic\_clean.csv", quote=TRUE, append=FALSE, col.names=NA, row.names=TRUE, sep=",", dec = ".", qmethod = c("escape", "double"), fileEncoding="")