

```

data_clean<-read.csv("C:\\Users\\viswa\\Downloads\\Dataset.csv")
onlycolumns<-c("ICULOS", "Age", "Gender", "HospAdmTime","HR", "O2Sat", "SBP", "MAP",
"DBP", "SepsisLabel")

df_clean<-data_clean[,onlycolumns]
sum(is.na(df_clean$ICULOS))
sum(is.na(df_clean$Age))
sum(is.na(df_clean$Gender))
sum(is.na(df_clean$HospAdmTime))
sum(is.na(df_clean$SepsisLabel))
sum(is.na(df_clean$HR))
sum(is.na(df_clean$O2Sat))
sum(is.na(df_clean$SBP))
sum(is.na(df_clean$MAP))
sum(is.na(df_clean$MAP))

final_df<-na.omit(df_clean)
str(final_df)
dt<-final_df[1:9]

sum(is.na(final_df))
nrow(final_df)
str(final_df)

max(final_df$Age)
min(final_df$Age)

# numeric_cols <- sapply(dt,is.numeric)
# df_numeric <- dt[, numeric_cols]

# scaled_data <- scale(df_numeric)
# scaled_df <- as.data.frame(scaled_data)
# final_sepsis <- cbind(final_df[!numeric_cols], scaled_df)
# final_sepsis

set.seed(123)

# num_rows <- nrow(final_df)
# sample_size <- min(15000, num_rows)
# sampled_df <- final_df[sample(num_rows, sample_size), ]
# str(sampled_df)

final_df$Patient_ID <- paste0("P", seq_len(nrow(final_df)))
str(final_df)

datafinal<-final_df[, c("Patient_ID", "Gender", "Age", "ICULOS", "HospAdmTime", "HR",
"O2Sat", "SBP", "MAP", "DBP", "SepsisLabel")]
str(datafinal)
table(datafinal$SepsisLabel)
library(sqldf)
tab<-sqldf("SELECT *
FROM datafinal
WHERE SepsisLabel = 0
LIMIT 5000")
tab1<-sqldf("SELECT * FROM datafinal WHERE SepsisLabel = 1 LIMIT 5000;")

last<-rbind(tab1,tab)
names(last)[names(last) == "SepsisLabel"] <- "Sepsis_Result"
last <- last[, -which(names(last) == "Patient_ID")]

str(last)

```

```
write.csv(last, file = "data.csv", row.names = FALSE)
```