T-test

```
data <- read.csv("water_potability.csv")

data$Drinkable <- ifelse(data$Potability == 1,"Drinkable","Non-drinkable")

data$Drinkable <- factor(data$Drinkable,levels = c("Non-drinkable", "Drinkable"))

data_RQ <- data.frame(Drinkable = data$Drinkable,TDS = data$Solids)

t_result <- t.test(TDS ~ Drinkable,data = data_rq,var.equal = FALSE)

t_result
```

```
Welch Two Sample t-test

data:  TDS by Drinkable

t = -1.9052, df = 2595, p-value = 0.05686

alternative hypothesis: true difference in means between group Non-drinkable and group Drinkable is not equal to 0

95 percent confidence interval:

-1230.72363    17.72317

sample estimates:

mean in group Non-drinkable     mean in group Drinkable

                21777.49                  22383.99
```

The analysis tested if drinkable and non drinkable water samples actually differ in their mean TDS levels. The Welch t test and the boxplot showed that the means don't differ much, meaning TDS alone may not be a strong factor for separating the two groups. these results also need to be understood with the dataset limits in mind, like the missing values and the imbalance between potability classes which maybe affects the outcome a bit.