```
install.packages("dplyr")

library(dplyr)


wine <- read.csv("wine_cleaned.csv")
head(wine, 5)
summary(wine)

t.test(alcohol ~ type, data = wine)
#After running the t-test, the p-value we observed of 0.004278 is less
than 0.05
#Hence, we can conclude that there is a significance difference in
alcohol level for white and red wine; with white wine containing about
0.09% more than red wines
#At first glance, the difference in the mean alcohol level may seem
relatively small. However, since we are working with a large dataset of
more than 6000 rows, the difference is significant.

t.test(pH ~ type, data = wine)
#The t-test shows us that p-value for PH is lower than 0.05
#Similar to the alcohol test, since the p value is lower than 0.05, we
have enough evidence to suggest that there is significant difference in
pH levels for red and white wines.
#Red wines appear to to have more acidity compared to white wines

t.test(density ~ type, data = wine)
#Again, the t-test shows us that p-value for density is lower than 0.05
#This suggests that there is a significance difference in density for red
and white wines. With a t value of 42.709, we can conclude that the
difference in density is considerbly large given the nature of low
#variances in density itself.
#Here, we can also see that red wines have a higher density than white
wines.

t.test(residual.sugar ~ type, data = wine)
#Similar to the previous t-tests, the p value for residual sugar is less
than 0.05
#This shows us that there is significant difference in residual sugar
levels between the red and the white wine. The t value of -47.802 also
suggests that the difference is considerable.
#It leads us to believe that the white wines are sweeter than red wines.

str(wine$type)
table(wine$type)

wine$type <- as.factor(wine$type)
wine$quality <- as.factor(wine$quality)


log_model <- glm(type ~ alcohol + residual.sugar + density + pH,
                 data = wine,
                 family = binomial)
```

```
summary(log_model)

pred_prob <- predict(log_model, type = "response")
pred_class <- ifelse(pred_prob > 0.5, "white", "red")

mean(pred_class == wine$type)

table(Predicted = pred_class, Actual = wine$type)
```