Descriptive statistics (TDS by drinkable vs non-drinkable)

To support the analysis, we created a binary categorical variable called Drinkable from the potability column (0 = non drinkable, 1 = drinkable). this helped us compare Total Dissolved Solids (TDS) between the two groups. the dataset was then summarised using basic descriptive statistics like count minimum maximum and mean for TDS overall and also by group.

In the cleaned dataset used for this study, there were 1998 non drinkable samples and 1278 drinkable samples. overall the TDS values range from 320.9 to 61,227.2 with a mean of about 22,014.1, which shows a very wide spread of values and suggests there are some extreme high values present. for the non drinkable group, TDS ranges from 320.9 to 61,227.2 with a mean of 21,777.5. for the drinkable group, the minimum TDS is 728.8, the maximum is 56,488.7 and the mean is around 22,384.

When we compare the group means, the difference between drinkable and non drinkable water is quite small. both groups have large ranges and a lot of overlap, which suggests that TDS alone may not clearly seperate the two classes. the presence of very high TDS values, especially near the upper end, also shows that the data is right skewed, which is common in enviromental datasets. these descriptive statistics help give background context before using the inferential test (Welch's two sample t test) later to check if the mean difference is statistically signifcant.

```
summary(data_rq$TDS)
```

```
   Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
  320.9   13213.7   20428.4   22014.1   27973.0   61227.2 
```

```
summary(data_rq$TDS[data_rq$Drinkable == "Non-drinkable"])
```

```
   Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
  320.9   13055.3   20125.6   21777.5   27642.1   61227.2 
```

```
summary(data_rq$TDS[data_rq$Drinkable == "Drinkable"])
```

```
   Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
  728.8   13402.6   20890.7   22384.0   28315.4   56488.7 
```

```
table(data_rq$Drinkable)
```

```
Non-drinkable     Drinkable 
         1998          1278 
```