**Extract:**

For this project we have chosen the data that we have pulled from the Kaggle. Two csv files that had the same information for white and red wines.

**The Hypothesis:**

* Does the difference in the amount of sulfate in wine change the quality?
  + If yes, does the quality of wine result in lower alcohol content?
  + If no, does the quality of wine result in higher alcohol content?
* Is there a difference in the amount of sulfate to the quality of white or red wine?
  + If yes, is the difference because of the amount of sulfate?
  + If no, is the difference because of the alcohol content?
* The difference is neither the sulfate nor the alcohol content – specify dependent on the data

**Transform:**

Our purpose was to find out which components of wine can predict the quality of either white or red wine.

To clean up the data we first took a look at what the data embodied and if there’s a commonality that the data can be merged as one.

From examining the data, we found out that the two tables contain exactly the same 12 columns thus we have created the column named “Type” to indicate either white or red wine.

Going forward we dropped the cells that did not have any data on it by using the “.dropna”.

We chose SQL pgAdmin to load both tables for further analysis.

After further analysis, we realized that the quality of wine is not dependent and had no direct correlation with each other but rather the quality of wine is dependent on acidity, chloride, and sulphates that determine the quality of wine.