**Discussion**

We first explored each attribute (quality, pH, acidity, sulfur dioxide, *etc*.) in the dataset to do univariate analysis, the most relevant and valuable features are consolidated. Then, we use correlation matrix to study the relationship among these attributes to figure out the most related attributes to quality score. After we find the most related several attributes, we build a regression model to estimate the wine quality based on our model, and a model evaluation was also conducted

We implemented logistic regression combined with clustering on the wine quality data and successfully developed an evaluation model to help the customers to learn about the wine quality and wine manufactures to improve wine production.

**Citation:**

Modeling wine preferences by data mining from physicochemical properties

*https://www.sciencedirect.com/science/article/pii/S0167923609001377?via%3Dihub*