Deliverable 1

Select Dataset “Wine Quality” on Kaggle, which includes a complete set of statistics for features such fixed acidity, volatile acidity, residual sugar, etc. Select wines, analyze the quality of wine, according to different features, and try to predict the price of a wine based on these features.

I selected this dataset because there are many wines to analyze with, and 13 columns with many features needed to evaluate a wine. With this dataset I should have enough statistics to build an algorithm and relate the price of wine to different features.

My Plan on approaching this project is to first do some research on the quality of wine and trend according to different features and decide on a few features to focus on since it is a large dataset. I will use supervised learning regression model that generalizes well for most of data points. The reason for choosing this model is simple: there is one target (quality) and many features, regression is good for predicting and forecasting. The next step is to regroup the data and perform train-test split. Next is to find out how to apply linear regression. The last step is to graph the results and evaluate the model.

For demo project, I will build a poster presentation. I will present my ML code and try to match the standard baseline online provided by other users.

Dataset link:

https://www.kaggle.com/danielpanizzo/wine-quality

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553. ISSN: 0167-9236.