## What is the problem you want to solve?

I will create a model to predict wine demand for a medium-sized wine supplier, which has approximately 35 wines brands under the enterprise umbrella. Specifically, the predictive model will be developed, at least initially, to forecast wine over a 12-month period for only a subset of wine brands, excluding niche and high-end wines. The model will be created to predict demand of wine sold to distributors and will ignore the negligible volume of wine demanded and sold directly to consumers.

## Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

My client is a wine supplier located in Northern California. Having accurate demand and projected sales is important to the company for sales and operation planning. Sales levels drive marketing initiatives, so this information could be used for quarterly campaigning and sales “blitzes”. Operationally, the company shares a few bottling facilities amongst the separate wineries located in Northern California and knowing demand levels in advance helps with the scheduling of these facilities and the logistics around the movement of wine from a winery to the bottling facility. Other uses of this information would include projection analysis for borrowing purposes and decisions around future production.

## What data are you going to use for this? How will you acquire this data?

Data will come from several sources including company data, used by permission, industry data found online or available through market research companies (IRI, for example), and U.S census data.

## In brief, outline your approach to solving this problem (knowing that this might change later).

I will attempt to solve this problem step-by-step, using these general steps as a guide:

1. Exploratory data analysis: get the data and begin to review it for outliers, trends, correlation, missing data, type of data and the distribution of the data. Log and take notes of this process.
2. Determine key attributes
3. Data cleaning and preprocessing: Manipulate and standardize data, decide how missing values will be handled,
4. Model measurement: Determine how the accuracy and performance of the models will be measured, set this aside
5. Split the data into a training set and a test set
6. Train several different models, keeping track of the performance of each. Start with a basic model and refine it through several iterations.

## What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Deliverables will include code (Python scripts), a paper explaining the process and results, and a more succinct slide deck summary.