**Sensegrass Internship**

The Wine Land store has various types of wine, this data is stored in the feature ‘variety’ of the dataset.

The different types of wine are as follows:

*‘Pinot Noir', 'Chardonnay', 'Cabernet Sauvignon', 'Red Blend', 'Bordeaux-style Red Blend', 'Riesling', 'Sauvignon Blanc', 'Syrah', 'Rosé', 'Merlot', 'Nebbiolo', 'Zinfandel', 'Sangiovese', 'Malbec', 'Portuguese Red', 'White Blend', 'Sparkling Blend', 'Tempranillo', 'Rhône-style Red Blend', 'Pinot Gris', 'Champagne Blend', 'Cabernet Franc', 'Grüner Veltliner', 'Portuguese White', 'Bordeaux-style White Blend', 'Pinot Grigio', 'Gamay', 'Gewürztraminer'*

The Exploratory Data Analysis was done with two approaches :

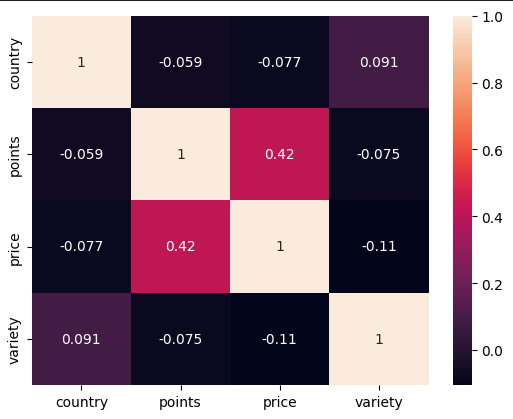
1. All features were considered and explored.
2. Only reviews were considered and explored.

**Actionable Insights :**

1. ­The ‘review\_title’ is the primary and the most important feature as it gives a direct relation with the ‘variety’ i.e. the dependent feature for the predictive model.
2. Upon further exploration, the features : ‘designation’, ‘province’, ‘region\_1’, ‘region\_2’ and ‘winery’ were derived from the feature : ‘review\_title’.
3. One of the insights from the EDA when considering all features was that the features other than ‘review\_title’ and ‘review\_description’ doesn’t have an influence in dictating the ‘variety’ of the wine.
4. On exploring the feature : ‘review\_title’, the language were mostly English, but a considerable number of records were in other languages too. This information was deduced from the feature : ‘country’.
5. There were instances of same words(most of them directly but some in other languages as well)being used in the feature : ‘review\_title’ and the dependent feature : ‘variety’ that can be exploited to get an effective model.

\* On performing sentiment analysis on the feature : ‘review\_description’ , the inference was made that it does not dictate the variety of the wine.

\*\* The inference that the feature ‘points’ and ‘price’ does not dictate the variety of the wine was made on the results from correlation matrices between these features and the feature ‘variety’ of the wine.



\*\*\* The reviews given by the ‘user\_name’ is considered credible and as such the user\_name itself does not dictate the variety of the wine.