**Assignment on Pandas**

Dataset E:\Work\Python for Data Science\Datasets\winemag-data-130k-v2.csv

**Indexing, slicing**

**Load the winery dataset and call it as reviews**

1. Select the description column from reviews and assign the result to the variable desc.
2. Select the first value from the description column of reviews, assigning it to variable ***first\_description***.
   1. what are the different ways to get first value from the description column
3. Select the first row of data (the first record) from reviews, assigning it to the variable ***first\_row***.
4. Select the first 10 values from the description column in reviews, assigning the result to variable ***first\_descriptions***
   1. try diff variations.
5. Select the records with index labels 1, 2, 3, 5, and 8, assigning the result to the variable ***sample\_reviews***
6. Create a variable df containing the country, province, region\_1, and region\_2 columns of the records with the index labels 0, 1, 10, and 100
7. Create a variable df containing the country and variety columns of the first 100 records. Try both loc and iloc
8. Create a DataFrame ***italian\_wines*** containing reviews of wines made in Italy.
9. Create a DataFrame `***top\_oceania\_wines***` containing all reviews with at least 95 points (out of 100) for wines from Australia or New Zealand.

**Summary Functions and Maps**

Functions like - Describe, mean, unique, value\_counts,

1. What is the median of the points column in the reviews DataFrame?
2. What countries are represented in the dataset? Print the list without duplicates
3. How often does each country appear in the dataset? Create a Series reviews\_per\_country mapping countries to the count of reviews of wines from that country.
4. Create variable centered\_price containing a version of the price column with the mean price subtracted.
5. Which wine is the "best bargain"? Create a variable bargain\_wine with the ***title*** of the wine with the highest points-to-price ratio in the dataset.

Hint use idxmax

1. There are only so many words you can use when describing a bottle of wine. Is a wine more likely to be "tropical" or "fruity"? Check how many times these words are used description.

tropical 3607

fruity 9090

Hint: use map function to check if description contain these words and then compute count

Create series to represent data

1. We'd like to host these wine reviews on our website, but a rating system ranging from 80 to 100 points is too hard to understand - we'd like to translate them into simple star ratings.
   1. A score of 95 or higher counts as 3 stars, a score of at least 85 but less than 95 is 2 stars. Any other score is 1 star.
   2. Also, the Canadian Vintners Association bought a lot of ads on the site, so any wines from Canada should automatically get 3 stars, regardless of points.
   3. Create a series star\_ratings with the number of stars corresponding to each review in the dataset

Hint: create a function to compute rating, apply it and compute new column “**star\_ratings**”