Assignment 2 - Spark – 50 points

# REQUIRED

You are provided with Craft Brewery data to perform basic data analysis using Spark. The data includes three files:

* Beers.csv
* Breweries.csv
* Dictionary.doc

Use the above csv files and perform the following:

1. Import the brewery.csv into Sqoop. Move the brewery.csv into Spark from Sqoop (intermediate movement to another component in Hadoop is fine)
2. Create RDD for Beers.csv
3. For the following questions use Spark SQL
   1. Determine the number of breweries in each state
   2. Determine the cities with most breweries
   3. Determine the most brewed beer style

**OPTIONAL**

Using MlLib, predict either beerstyle or abv content. You can use any statistical or machine learning model for your prediction. You can also come up with your own problem statement and predict using any model available in MlLib. Sample code using MlLib is available on Canvas.

**Deliverables:**

1. You must provide the screenshots of your solution along the answers on the terminal.
2. Please provide a few line explanation under each screenshot provided.
3. Final deliverable must be submitted in a Word document.