# Project Origin and Scope

This repository is setup for SMU Data Science Master program, Doing the Data Science course team project. Nitin and YuMei are the two team members, their contact info as follows.

YuMei Bennett YuMei.Bennett@yahoo.com

Nitin Agarwal [nitina@mail.smu.edu](mailto:nitina@mail.smu.edu)

The rest of the readme is written as if the team is executing a client data analytic project scenario, for the purpose of course requirement.

# Introduction

Machine Churning consulting is a new startup, our mission is to provide business intelligence to convert data into information and knowledge. Integrating large volumes of data with advanced analytics techniques, modern computer technology, and domain expertise within specific business sectors, our data analytic service looking inside the crystal ball of Big Data to predict the future of their enterprise.

Cobell Entertainment is one of the largest brewery and restaurant corporation in the US. They have been entertaining customers for over 50 years. Recent market down turn created difficult business environment for them, management is looking for innovative ideas on all front to create new revenue stream. They have contracted Machine Churning consulting on a beer and brewery data analytic project.

# Raw Data Input

There are two historical data files are provided as the analysis input.

The Beers dataset contains a list of 2410 US craft beers and Breweries dataset contains 558 US breweries. The datasets descriptions are as follows.

Beers.csv:

Name: Name of the beer.

Beer\_ID: Unique identifier of the beer.

ABV: Alcohol by volume of the beer.

IBU: International Bitterness Units of the beer.

Brewery\_ID: Brewery id associated with the beer.

Style: Style of the beer.

Ounces: Ounces of beer.

Breweries.csv:

Brew\_ID: Unique identifier of the brewery.

Name: Name of the brewery.

City: City where the brewery is located.

State: U.S. State where the brewery is located.

# Data Transformation

We merged two data table together by Breweries Identification. We had to align the Names for Breweries ID to be the same as Brew\_ID.

Merged is the data file after merge.

# Exploratory Data Analysis

The following new data frames are created:

Breweriesperstate table shows how many breweries per state.

NApercolumn table shows how many NAs per variable.

medianABVperstate table shows alcohol content median per state

medianIBUperstate table shows international bitterness median per state

MergedsortABV table sort on ABV in descending order

MergedsortIBU table sort on IBU in descending order

The following graphs are created:

Bar chart for Median IBU per State

Bar chart for Median ABV per State

Scatterplot IBU vs. ABV with regression line