**Problem Statement**

You are the sales manager for "BeerMart", an online beer store in the United States. You want to build a recommendation system (collaborative) for your store, where customers will be recommended the beer that they are most likely to buy. You collect data about the ratings that the customers have given in the past. You can download the dataset from the link below.

**[Beer Dataset](https://cdn.upgrad.com/UpGrad/temp/3f1237da-dbc5-434b-8ea4-c4bf43dbfd42/beer_data.csv" \o "beer_data.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/3f1237da-dbc5-434b-8ea4-c4bf43dbfd42/beer_data.csv" \o "beer_data.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/3f1237da-dbc5-434b-8ea4-c4bf43dbfd42/beer_data.csv" \o "beer_data.csv" \t "_blank)**

Description: Each record includes a beer's name and the user's name, along with the ratings he/she has given to the beer. All ratings are on a scale from 1 to 5, with 5 being the best rating.

As you solve the case study, you need to complete the following tasks:

1. Data preparation

Choose only those beers that have at least N number of reviews.

* 1. Figure out an appropriate value of N using EDA; this may not have one correct answer, but you should not choose beers that have an extremely low number of ratings.

1. Data exploration
   1. What are the unique values of ratings?
   2. Visualise the rating values and notice:
      * The average beer ratings
      * The average user ratings
      * The average number of ratings given to the beers
      * The average number of ratings given by the users
2. Recommendation Models
   1. Divide your data into training and testing dataset.
   2. Build user-based and item-based models.
   3. Determine how similar the first 10 users are to each other and visualise it.
   4. Compute and visualise the similarity between the first 10 beers.
   5. Compare the performance of the two models using test data and suggest the one that should be deployed.
   6. Give the names of the top 5 beers that you would recommend to the users 'cokes', 'genog' and 'giblet' using both the models.