**BrewBuddy Analytics-Powered Beer Recommendation System**

Task A: Data Extraction

* Extract approximately 5,000–6,000 reviews from [BeerAdvocate Top 250](https://www.beeradvocate.com/beer/top-rated/).
* Retain only reviews with text, resulting in 1,700–2,000 reviews.
* Output file: 3 columns – product\_name, product\_review, user\_rating.

Task B: Attribute Identification

* Identify attributes important to beer descriptions using word frequency analysis and lift analysis. Examples:
  + Aggressive: Bold aroma/taste.
  + Balanced: Equal malt sweetness and hop bitterness.
  + Complex: Multidimensional flavors.
  + Fruity/Hoppy/Malty: Fruit-like or hop-derived aromas.
  + Crisp: Effervescent carbonation.
  + Robust: Rich and full-bodied.
* Ensure identified attributes frequently co-occur in reviews.

Task C: Similarity Analysis

* Accept 3 user-specified attributes as input.
* Calculate similarity scores using cosine similarity (bag-of-words model).
* Output file: 3 columns – product\_name, product\_review, similarity\_score.

Task D: Sentiment Analysis

* Perform sentiment analysis on reviews using VADER or another tool.
* Optionally customize the lexicon for better contextual understanding.

Task E: Evaluation and Recommendation

* Combine similarity and sentiment scores into an evaluation score for each beer.
* Recommend 3 products based on these scores.

Task F: Word Vectors vs. Bag-of-Words

* Replace cosine similarity (bag-of-words) with word vectors (e.g., spaCy medium-sized pretrained vectors).
* Compare % of reviews mentioning user-specified attributes across both methods.
* Analyze differences in recommendations and explain findings.

Task G: Ratings-Based Recommendations

* Recommend the 3 highest-rated products in the dataset without similarity or sentiment analysis.
* Compare these recommendations with attribute-based ones.
* Justify whether they meet user requirements based on analysis.

Task H: Similar Beer Analysis

* Select any 10 beers and find the most similar beer to one chosen beer among the remaining 9.
* Explain the methodology used to determine similarity.