**Exercise:**

1. Download Amazon Fine Food Reviews dataset from Kaggle. You may have to create a Kaggle account to download data. (<https://www.kaggle.com/snap/amazon-fine-food-reviews>)
2. Split data into train and test using time based slicing as 70% train & 30% test.
3. Perform featurization BoW,TFIDF, Avg Word2Vec, tf-idf-Word2Vec.
4. **K-Means ++ Clustering :**
   * Find the best ‘k’ using elbow-knee method.
   * Also find k sets and read through all the individual points in sets and evaluate that reviews make some sense or not.
   * Also apply k-medoids algorithm as well.
5. **Agglomerative Clustering :**
   * Apply agglomerative algorithm and try different number of clusters like 2,5 etc.
   * Also read through all the individual points in the clusters and evaluate that reviews make some sense or not.
   * Note: Only take ~5000 reviews to perform hierarchical clustering because of high training complexity.
6. **DBSCAN Clustering :**
   * Apply DBSCAN algorithm and take MinPts = 2\*d.
   * Find the best ‘Eps’ using elbow-knee method.
   * After finding the optimal ‘Eps’, try different values of ‘Eps’ to check the variations in clustering.
   * Also read through all the individual points in the clusters and evaluate that reviews make some sense or not.
   * Note: Take complete ~364K dataset and perform DBSCAN algorithm on Word2Vec vectorizers with 100-dimension.
7. Write your observations in English as crisply and unambiguously as possible. Always quantify your results.