Introduction to Data Science, HW #3 – Map Reduce

In the following question you need to provide pseudo-code or an implementation in Python. You need to write both Map() and Recude() function. In some cases, there may be more then 1 MapReduce operations.

1. Implement K-Means algorithm in MapReduce paradigm
    - **Data – points (x,y) in [0,1]x[0,1]:**
      0.72 0.44
      0.16 0.82
      0.42 0.37
      0.19 0.65
      …
    - **Desired output:**
      (0.72 0.44) (0.55 0.83)
      (0.16 0.82) (0.55 0.83)
      (0.42 0.37) (0.29 0.16)
      …

2. Implement CheckClique function, that given an undirected graph returns if the graph is a full clique
    - **Data:**
      A -> B C
      B -> A C
      C -> A B
      …
    - **Desired output:**
      YES

3. Implement pseudo-synonyms detection algorithm in MapReduce paradigm

    - **Data – queries:**
      buy cheap house
      buy big house
      buy new house
      rent cheap car
      rent new car
      rent Volkswagen car
      …
    - **Desired output:**
      cheap – new (2)

4. Find triangles. Given a graph, find all triangles in the graph:
    - **Data:**
      A -> B C D
      B -> A C E
      C -> A B D
      D -> A C
      E -> B
    - **Desired output:**
      A, B, C
      A, C, D