Bài 5: KHAI THÁC MẪU KẾT HỢP (TT)

1. Thuật toán Vertical Apriori:

BEGIN:

Items = {0: 'Apple', 1: 'Bread', 2: 'Butter', 3: 'Chips', 4: 'Milk', 5: 'Wine'}

TID_list = {'Apple': [1, 4, 5, 7, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21],
       'Bread': [1, 2, 3, 5, 7, 9, 10, 11, 13, 14, 15, 16, 17, 19, 20, 21, 22],
       'Butter': [1, 2, 3, 4, 5, 7, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20],
       'Chips': [1, 4, 5, 6, 7, 8, 11, 13, 16, 17, 18, 19, 21, 22],
       'Milk': [1, 2, 3, 5, 6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22],
       'Wine': [1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 19, 20, 21]}

TID_list_temp = {'Apple': [1, 4, 5, 7, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21],
              'Bread': [1, 2, 3, 5, 7, 9, 10, 11, 13, 14, 15, 16, 17, 19, 20, 21, 22],
              'Butter': [1, 2, 3, 4, 5, 7, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20],
              'Chips': [1, 4, 5, 6, 7, 8, 11, 13, 16, 17, 18, 19, 21, 22],
              'Milk': [1, 2, 3, 5, 6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22],
              'Wine': [1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 19, 20, 21]}

F[0] = {'Apple': [1, 4, 5, 7, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21],
      'Bread': [1, 2, 3, 5, 7, 9, 10, 11, 13, 14, 15, 16, 17, 19, 20, 21, 22],
      'Butter': [1, 2, 3, 4, 5, 7, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20],
      'Chips': [1, 4, 5, 6, 7, 8, 11, 13, 16, 17, 18, 19, 21, 22],
      'Milk': [1, 2, 3, 5, 6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22],
      'Wine': [1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 19, 20, 21]}

LOOP 1:

Items = {0: ['Apple', 'Bread'],
       1: ['Apple', 'Butter'],
       2: ['Apple', 'Chips'],
       3: ['Apple', 'Milk'],
       4: ['Apple', 'Wine'],
       5: ['Bread', 'Butter'],
       6: ['Bread', 'Chips'],
       7: ['Bread', 'Milk'],
       8: ['Bread', 'Wine'],
       9: ['Butter', 'Chips'],
       10: ['Butter', 'Milk'],
       11: ['Butter', 'Wine'],
       12: ['Chips', 'Milk'],
       13: ['Chips', 'Wine'],
       14: ['Milk', 'Wine']}

TID_list = {"['Apple', 'Bread']": [1, 5, 7, 9, 11, 14, 15, 16, 17, 19, 20, 21],
        "['Apple', 'Butter']": [1, 4, 5, 7, 11, 12, 15, 16, 17, 18, 19, 20],
        "['Apple', 'Chips']": [1, 4, 5, 7, 11, 16, 17, 18, 19, 21],
        "['Apple', 'Milk']": [1, 5, 12, 14, 15, 16, 17, 18, 19, 20, 21],
        "['Apple', 'Wine']": [1, 5, 7, 9, 12, 14, 15, 16, 19, 20, 21],
        "['Bread', 'Butter']": [1, 2, 3, 5, 7, 10, 11, 13, 15, 16, 17, 19, 20],
        "['Bread', 'Chips']": [1, 5, 7, 11, 13, 16, 17, 19, 21, 22],
        "['Bread', 'Milk']": [1, 2, 3, 5, 10, 13, 14, 15, 16, 17, 19, 20, 21, 22],
        "['Bread', 'Wine']": [1, 2, 5, 7, 9, 10, 13, 14, 15, 16, 19, 20, 21],
        "['Butter', 'Chips']": [1, 4, 5, 7, 11, 13, 16, 17, 18, 19],
        "['Butter', 'Milk']": [1, 2, 3, 5, 10, 12, 13, 15, 16, 17, 18, 19, 20],
        "['Butter', 'Wine']": [1, 2, 5, 7, 10, 12, 13, 15, 16, 19, 20],
        "['Chips', 'Milk']": [1, 5, 6, 8, 13, 16, 17, 18, 19, 21, 22],
        "['Chips', 'Wine']": [1, 5, 6, 7, 8, 13, 16, 19, 21],
        "['Milk', 'Wine']": [1, 2, 5, 6, 8, 10, 12, 13, 14, 15, 16, 19, 20, 21]}

F[1] =   {"['Apple', 'Bread']": [1, 5, 7, 9, 11, 14, 15, 16, 17, 19, 20, 21],
        "['Apple', 'Butter']": [1, 4, 5, 7, 11, 12, 15, 16, 17, 18, 19, 20],
        "['Apple', 'Chips']": [1, 4, 5, 7, 11, 16, 17, 18, 19, 21],
        "['Apple', 'Milk']": [1, 5, 12, 14, 15, 16, 17, 18, 19, 20, 21],
        "['Apple', 'Wine']": [1, 5, 7, 9, 12, 14, 15, 16, 19, 20, 21],
        "['Bread', 'Butter']": [1, 2, 3, 5, 7, 10, 11, 13, 15, 16, 17, 19, 20],
        "['Bread', 'Chips']": [1, 5, 7, 11, 13, 16, 17, 19, 21, 22],
        "['Bread', 'Milk']": [1, 2, 3, 5, 10, 13, 14, 15, 16, 17, 19, 20, 21, 22],
        "['Bread', 'Wine']": [1, 2, 5, 7, 9, 10, 13, 14, 15, 16, 19, 20, 21],
        "['Butter', 'Chips']": [1, 4, 5, 7, 11, 13, 16, 17, 18, 19],
        "['Butter', 'Milk']": [1, 2, 3, 5, 10, 12, 13, 15, 16, 17, 18, 19, 20],
        "['Butter', 'Wine']": [1, 2, 5, 7, 10, 12, 13, 15, 16, 19, 20],
        "['Chips', 'Milk']": [1, 5, 6, 8, 13, 16, 17, 18, 19, 21, 22],
        "['Chips', 'Wine']": [1, 5, 6, 7, 8, 13, 16, 19, 21],
        "['Milk', 'Wine']": [1, 2, 5, 6, 8, 10, 12, 13, 14, 15, 16, 19, 20, 21]}

LOOP 2:

Items = {0: ['Butter', 'Bread', 'Apple'],
        1: ['Chips', 'Bread', 'Apple'],
        2: ['Milk', 'Bread', 'Apple'],
        3: ['Wine', 'Bread', 'Apple'],
        14: ['Chips', 'Butter', 'Apple'],
        15: ['Butter', 'Milk', 'Apple'],
        16: ['Wine', 'Butter', 'Apple'],
        27: ['Chips', 'Milk', 'Apple'],
        28: ['Wine', 'Chips', 'Apple'],
        33: ['Chips', 'Apple', 'Butter'],
        39: ['Wine', 'Milk', 'Apple'],
        60: ['Chips', 'Butter', 'Bread'],

61: ['Butter', 'Milk', 'Bread'],
62: ['Wine', 'Butter', 'Bread'],
69: ['Chips', 'Milk', 'Bread'],
70: ['Wine', 'Chips', 'Bread'],
71: ['Chips', 'Bread', 'Butter'],
77: ['Wine', 'Milk', 'Bread'],
90: ['Butter', 'Milk', 'Chips'],
91: ['Wine', 'Butter', 'Chips'],
95: ['Wine', 'Butter', 'Milk'],
102: ['Wine', 'Milk', 'Chips']}

TID_list = {"['Butter', 'Bread', 'Apple']": [1, 5, 7, 11, 15, 16, 17, 19, 20],
　　　　　"['Chips', 'Bread', 'Apple']": [1, 5, 7, 11, 16, 17, 19, 21],
　　　　　"['Milk', 'Bread', 'Apple']": [1, 5, 14, 15, 16, 17, 19, 20, 21],
　　　　　"['Wine', 'Bread', 'Apple']": [1, 5, 7, 9, 14, 15, 16, 19, 20, 21],
　　　　　"['Chips', 'Butter', 'Apple']": [1, 4, 5, 7, 11, 16, 17, 18, 19],
　　　　　"['Butter', 'Milk', 'Apple']": [1, 5, 12, 15, 16, 17, 18, 19, 20],
　　　　　"['Wine', 'Butter', 'Apple']": [1, 5, 7, 12, 15, 16, 19, 20],
　　　　　"['Chips', 'Milk', 'Apple']": [1, 5, 16, 17, 18, 19, 21],
　　　　　"['Wine', 'Chips', 'Apple']": [1, 5, 7, 16, 19, 21],
　　　　　"['Chips', 'Apple', 'Butter']": [1, 4, 5, 7, 11, 16, 17, 18, 19],
　　　　　"['Wine', 'Milk', 'Apple']": [1, 5, 12, 14, 15, 16, 19, 20, 21],
　　　　　"['Chips', 'Butter', 'Bread']": [1, 5, 7, 11, 13, 16, 17, 19],
　　　　　"['Butter', 'Milk', 'Bread']": [1, 2, 3, 5, 10, 13, 15, 16, 17, 19, 20],
　　　　　"['Wine', 'Butter', 'Bread']": [1, 2, 5, 7, 10, 13, 15, 16, 19, 20],
　　　　　"['Chips', 'Milk', 'Bread']": [1, 5, 13, 16, 17, 19, 21, 22],
　　　　　"['Wine', 'Chips', 'Bread']": [1, 5, 7, 13, 16, 19, 21],
　　　　　"['Chips', 'Bread', 'Butter']": [1, 5, 7, 11, 13, 16, 17, 19],
　　　　　"['Wine', 'Milk', 'Bread']": [1, 2, 5, 10, 13, 14, 15, 16, 19, 20, 21],
　　　　　"['Butter', 'Milk', 'Chips']": [1, 5, 13, 16, 17, 18, 19],
　　　　　"['Wine', 'Butter', 'Chips']": [1, 5, 7, 13, 16, 19],
　　　　　"['Wine', 'Butter', 'Milk']": [1, 2, 5, 10, 12, 13, 15, 16, 19, 20],
　　　　　"['Wine', 'Milk', 'Chips']": [1, 5, 6, 8, 13, 16, 19, 21]}

F[2] = {"['Butter', 'Bread', 'Apple']": [1, 5, 7, 11, 15, 16, 17, 19, 20],
　　　　"['Chips', 'Bread', 'Apple']": [1, 5, 7, 11, 16, 17, 19, 21],
　　　　"['Milk', 'Bread', 'Apple']": [1, 5, 14, 15, 16, 17, 19, 20, 21],
　　　　"['Wine', 'Bread', 'Apple']": [1, 5, 7, 9, 14, 15, 16, 19, 20, 21],
　　　　"['Chips', 'Butter', 'Apple']": [1, 4, 5, 7, 11, 16, 17, 18, 19],
　　　　"['Butter', 'Milk', 'Apple']": [1, 5, 12, 15, 16, 17, 18, 19, 20],
　　　　"['Wine', 'Butter', 'Apple']": [1, 5, 7, 12, 15, 16, 19, 20],
　　　　"['Chips', 'Milk', 'Apple']": [1, 5, 16, 17, 18, 19, 21],
　　　　"['Wine', 'Chips', 'Apple']": [1, 5, 7, 16, 19, 21],
　　　　"['Chips', 'Apple', 'Butter']": [1, 4, 5, 7, 11, 16, 17, 18, 19],
　　　　"['Wine', 'Milk', 'Apple']": [1, 5, 12, 14, 15, 16, 19, 20, 21],
　　　　"['Chips', 'Butter', 'Bread']": [1, 5, 7, 11, 13, 16, 17, 19],

"['Butter', 'Milk', 'Bread']": [1, 2, 3, 5, 10, 13, 15, 16, 17, 19, 20],
"['Wine', 'Butter', 'Bread']": [1, 2, 5, 7, 10, 13, 15, 16, 19, 20],
"['Chips', 'Milk', 'Bread']": [1, 5, 13, 16, 17, 19, 21, 22],
"['Wine', 'Chips', 'Bread']": [1, 5, 7, 13, 16, 19, 21],
"['Chips', 'Bread', 'Butter']": [1, 5, 7, 11, 13, 16, 17, 19],
"['Wine', 'Milk', 'Bread']": [1, 2, 5, 10, 13, 14, 15, 16, 19, 20, 21],
"['Butter', 'Milk', 'Chips']": [1, 5, 13, 16, 17, 18, 19],
"['Wine', 'Butter', 'Chips']": [1, 5, 7, 13, 16, 19],
"['Wine', 'Butter', 'Milk']": [1, 2, 5, 10, 12, 13, 15, 16, 19, 20],
"['Wine', 'Milk', 'Chips']": [1, 5, 6, 8, 13, 16, 19, 21]}

LOOP 3:

Items = {0: ['Chips', 'Butter', 'Apple', 'Bread'],
     1: ['Butter', 'Apple', 'Bread', 'Milk'],
     2: ['Wine', 'Butter', 'Apple', 'Bread'],
     21: ['Apple', 'Bread', 'Chips', 'Milk'],
     22: ['Wine', 'Apple', 'Bread', 'Chips'],
     23: ['Butter', 'Apple', 'Bread', 'Chips'],
     41: ['Wine', 'Apple', 'Milk', 'Bread'],
     43: ['Butter', 'Apple', 'Milk', 'Bread'],
     45: ['Chips', 'Apple', 'Milk', 'Bread'],
     64: ['Wine', 'Chips', 'Apple', 'Bread'],
     66: ['Wine', 'Apple', 'Bread', 'Milk'],
     78: ['Butter', 'Apple', 'Chips', 'Milk'],
     79: ['Wine', 'Butter', 'Apple', 'Chips'],
     95: ['Wine', 'Butter', 'Milk', 'Apple'],
     96: ['Chips', 'Butter', 'Milk', 'Apple'],
     101: ['Butter', 'Milk', 'Bread', 'Apple'],
     112: ['Wine', 'Butter', 'Chips', 'Apple'],
     117: ['Wine', 'Butter', 'Bread', 'Apple'],
     126: ['Wine', 'Apple', 'Milk', 'Chips'],
     127: ['Butter', 'Apple', 'Milk', 'Chips'],
     132: ['Apple', 'Milk', 'Bread', 'Chips'],
     141: ['Wine', 'Apple', 'Chips', 'Milk'],
     171: ['Wine', 'Milk', 'Bread', 'Apple'],
     175: ['Wine', 'Chips', 'Milk', 'Apple'],
     176: ['Butter', 'Milk', 'Bread', 'Chips'],
     177: ['Wine', 'Butter', 'Bread', 'Chips'],
     186: ['Wine', 'Butter', 'Milk', 'Bread'],
     187: ['Chips', 'Butter', 'Milk', 'Bread'],
     196: ['Wine', 'Butter', 'Chips', 'Bread'],
     203: ['Wine', 'Milk', 'Bread', 'Chips'],
     224: ['Wine', 'Chips', 'Milk', 'Bread'],
     225: ['Wine', 'Butter', 'Milk', 'Chips']}

TID_list ={"['Chips', 'Butter', 'Apple', 'Bread']": [1, 5, 7, 11, 16, 17, 19],
          "['Butter', 'Apple', 'Bread', 'Milk']": [1, 5, 15, 16, 17, 19, 20],
          "['Wine', 'Butter', 'Apple', 'Bread']": [1, 5, 7, 15, 16, 19, 20],
          "['Apple', 'Bread', 'Chips', 'Milk']": [1, 5, 16, 17, 19, 21],
          "['Wine', 'Apple', 'Bread', 'Chips']": [1, 5, 7, 16, 19, 21],
          "['Butter', 'Apple', 'Bread', 'Chips']": [1, 5, 7, 11, 16, 17, 19],
          "['Wine', 'Apple', 'Milk', 'Bread']": [1, 5, 14, 15, 16, 19, 20, 21],
          "['Butter', 'Apple', 'Milk', 'Bread']": [1, 5, 15, 16, 17, 19, 20],
          "['Chips', 'Apple', 'Milk', 'Bread']": [1, 5, 16, 17, 19, 21],
          "['Wine', 'Chips', 'Apple', 'Bread']": [1, 5, 7, 16, 19, 21],
          "['Wine', 'Apple', 'Bread', 'Milk']": [1, 5, 14, 15, 16, 19, 20, 21],
          "['Butter', 'Apple', 'Chips', 'Milk']": [1, 5, 16, 17, 18, 19],
          "['Wine', 'Butter', 'Apple', 'Chips']": [1, 5, 7, 16, 19],
          "['Wine', 'Butter', 'Milk', 'Apple']": [1, 5, 12, 15, 16, 19, 20],
          "['Chips', 'Butter', 'Milk', 'Apple']": [1, 5, 16, 17, 18, 19],
          "['Butter', 'Milk', 'Bread', 'Apple']": [1, 5, 15, 16, 17, 19, 20],
          "['Wine', 'Butter', 'Chips', 'Apple']": [1, 5, 7, 16, 19],
          "['Wine', 'Butter', 'Bread', 'Apple']": [1, 5, 7, 15, 16, 19, 20],
          "['Wine', 'Apple', 'Milk', 'Chips']": [1, 5, 16, 19, 21],
          "['Butter', 'Apple', 'Milk', 'Chips']": [1, 5, 16, 17, 18, 19],
          "['Apple', 'Milk', 'Bread', 'Chips']": [1, 5, 16, 17, 19, 21],
          "['Wine', 'Apple', 'Chips', 'Milk']": [1, 5, 16, 19, 21],
          "['Wine', 'Milk', 'Bread', 'Apple']": [1, 5, 14, 15, 16, 19, 20, 21],
          "['Wine', 'Chips', 'Milk', 'Apple']": [1, 5, 16, 19, 21],
          "['Butter', 'Milk', 'Bread', 'Chips']": [1, 5, 13, 16, 17, 19],
          "['Wine', 'Butter', 'Bread', 'Chips']": [1, 5, 7, 13, 16, 19],
          "['Wine', 'Butter', 'Milk', 'Bread']": [1, 2, 5, 10, 13, 15, 16, 19, 20],
          "['Chips', 'Butter', 'Milk', 'Bread']": [1, 5, 13, 16, 17, 19],
          "['Wine', 'Butter', 'Chips', 'Bread']": [1, 5, 7, 13, 16, 19],
          "['Wine', 'Milk', 'Bread', 'Chips']": [1, 5, 13, 16, 19, 21],
          "['Wine', 'Chips', 'Milk', 'Bread']": [1, 5, 13, 16, 19, 21],
          "['Wine', 'Butter', 'Milk', 'Chips']": [1, 5, 13, 16, 19]}

F[3] =   {"['Chips', 'Butter', 'Apple', 'Bread']": [1, 5, 7, 11, 16, 17, 19],
          "['Butter', 'Apple', 'Bread', 'Milk']": [1, 5, 15, 16, 17, 19, 20],
          "['Wine', 'Butter', 'Apple', 'Bread']": [1, 5, 7, 15, 16, 19, 20],
          "['Apple', 'Bread', 'Chips', 'Milk']": [1, 5, 16, 17, 19, 21],
          "['Wine', 'Apple', 'Bread', 'Chips']": [1, 5, 7, 16, 19, 21],
          "['Butter', 'Apple', 'Bread', 'Chips']": [1, 5, 7, 11, 16, 17, 19],
          "['Wine', 'Apple', 'Milk', 'Bread']": [1, 5, 14, 15, 16, 19, 20, 21],
          "['Butter', 'Apple', 'Milk', 'Bread']": [1, 5, 15, 16, 17, 19, 20],
          "['Chips', 'Apple', 'Milk', 'Bread']": [1, 5, 16, 17, 19, 21],
          "['Wine', 'Chips', 'Apple', 'Bread']": [1, 5, 7, 16, 19, 21],
          "['Wine', 'Apple', 'Bread', 'Milk']": [1, 5, 14, 15, 16, 19, 20, 21],
          "['Butter', 'Apple', 'Chips', 'Milk']": [1, 5, 16, 17, 18, 19],

"['Wine', 'Butter', 'Apple', 'Chips']": [1, 5, 7, 16, 19],
"['Wine', 'Butter', 'Milk', 'Apple']": [1, 5, 12, 15, 16, 19, 20],
"['Chips', 'Butter', 'Milk', 'Apple']": [1, 5, 16, 17, 18, 19],
"['Butter', 'Milk', 'Bread', 'Apple']": [1, 5, 15, 16, 17, 19, 20],
"['Wine', 'Butter', 'Chips', 'Apple']": [1, 5, 7, 16, 19],
"['Wine', 'Butter', 'Bread', 'Apple']": [1, 5, 7, 15, 16, 19, 20],
"['Wine', 'Apple', 'Milk', 'Chips']": [1, 5, 16, 19, 21],
"['Butter', 'Apple', 'Milk', 'Chips']": [1, 5, 16, 17, 18, 19],
"['Apple', 'Milk', 'Bread', 'Chips']": [1, 5, 16, 17, 19, 21],
"['Wine', 'Apple', 'Chips', 'Milk']": [1, 5, 16, 19, 21],
"['Wine', 'Milk', 'Bread', 'Apple']": [1, 5, 14, 15, 16, 19, 20, 21],
"['Wine', 'Chips', 'Milk', 'Apple']": [1, 5, 16, 19, 21],
"['Butter', 'Milk', 'Bread', 'Chips']": [1, 5, 13, 16, 17, 19],
"['Wine', 'Butter', 'Bread', 'Chips']": [1, 5, 7, 13, 16, 19],
"['Wine', 'Butter', 'Milk', 'Bread']": [1, 2, 5, 10, 13, 15, 16, 19, 20],
"['Chips', 'Butter', 'Milk', 'Bread']": [1, 5, 13, 16, 17, 19],
"['Wine', 'Butter', 'Chips', 'Bread']": [1, 5, 7, 13, 16, 19],
"['Wine', 'Milk', 'Bread', 'Chips']": [1, 5, 13, 16, 19, 21],
"['Wine', 'Chips', 'Milk', 'Bread']": [1, 5, 13, 16, 19, 21],
"['Wine', 'Butter', 'Milk', 'Chips']": [1, 5, 13, 16, 19]}

Items = {0: ['Chips', 'Apple', 'Milk', 'Butter', 'Bread'],
      1: ['Chips', 'Apple', 'Wine', 'Butter', 'Bread'],
      31: ['Apple', 'Wine', 'Milk', 'Butter', 'Bread'],
      90: ['Chips', 'Apple', 'Wine', 'Milk', 'Bread'],
      286: ['Chips', 'Apple', 'Wine', 'Milk', 'Butter'],
      468: ['Chips', 'Wine', 'Milk', 'Butter', 'Bread']}

LOOP 4:

TID_list = {"['Chips', 'Apple', 'Wine', 'Milk', 'Butter', 'Bread']": [1, 5, 16, 19]}

F[4] = {"['Chips', 'Apple', 'Wine', 'Milk', 'Butter', 'Bread']": [1, 5, 16, 19]}

Items = {}

END.

2. CODE :

2.1 Nhập dữ liệu từ file csv:

```
[1]: import pandas as pd
     import numpy as np
```

```
[2]: df=pd.read_csv('CSDL.csv')
     df
```

[2]:

	TID	Itemset
0	1	Wine, Chips, Bread, Butter, Milk, Apple
1	2	Wine, Bread, Butter, Milk
2	3	Bread, Butter, Milk
3	4	Chips, Butter, Apple
4	5	Wine, Chips, Bread, Butter, Milk, Apple
5	6	Wine, Chips, Milk
6	7	Wine, Chips, Bread, Butter, Apple
7	8	Wine, Chips, Milk
8	9	Wine, Bread, Apple
9	10	Wine, Bread, Butter, Milk
10	11	Chips, Bread, Butter, Apple
11	12	Wine, Butter, Milk, Apple
12	13	Wine, Chips, Bread, Butter, Milk
13	14	Wine, Bread, Milk, Apple
14	15	Wine, Bread, Butter, Milk, Apple
15	16	Wine, Chips, Bread, Butter, Milk, Apple
16	17	Chips, Bread, Butter, Milk, Apple
17	18	Chips, Butter, Milk, Apple
18	19	Wine, Chips, Bread, Butter, Milk, Apple
19	20	Wine, Bread, Butter, Milk, Apple
20	21	Wine, Chips, Bread, Milk, Apple
21	22	Chips, Bread, Milk

2.2 Xử lý dữ liệu và chia thành Itemset và TID:

```
[3]: data=df['Itemset'].values
```

```
[4]: np.savetxt("temp.txt",data, fmt='%s')
```

```
[5]: # open a file
     Itemset = pd.read_csv("temp.txt", sep=', ',header=None)
     Itemset
```

C:\Users\maith\AppData\Local\Temp\ipykernel_24356\324984109.py:2: ParserWarning: Falling
  Itemset = pd.read_csv("temp.txt", sep=', ',header=None)

[5]:

	0	1	2	3	4	5
0	Wine	Chips	Bread	Butter	Milk	Apple
1	Wine	Bread	Butter	Milk	None	None
2	Bread	Butter	Milk	None	None	None
3	Chips	Butter	Apple	None	None	None
4	Wine	Chips	Bread	Butter	Milk	Apple
5	Wine	Chips	Milk	None	None	None
6	Wine	Chips	Bread	Butter	Apple	None
7	Wine	Chips	Milk	None	None	None
8	Wine	Bread	Apple	None	None	None
9	Wine	Bread	Butter	Milk	None	None
10	Chips	Bread	Butter	Apple	None	None
11	Wine	Butter	Milk	Apple	None	None
12	Wine	Chips	Bread	Butter	Milk	None
13	Wine	Bread	Milk	Apple	None	None
14	Wine	Bread	Butter	Milk	Apple	None
15	Wine	Chips	Bread	Butter	Milk	Apple
16	Chips	Bread	Butter	Milk	Apple	None
17	Chips	Butter	Milk	Apple	None	None
18	Wine	Chips	Bread	Butter	Milk	Apple
19	Wine	Bread	Butter	Milk	Apple	None
20	Wine	Chips	Bread	Milk	Apple	None
21	Chips	Bread	Milk	None	None	None

```
[6]: TID=df.iloc[:,:-1]
     TID
```

[6]:

	TID
0	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10
10	11
11	12
12	13
13	14
14	15
15	16
16	17
17	18
18	19
19	20
20	21
21	22

2.3 Hàm chuyển dữ liệu thành dạng thẳng đứng:

```python
[7]: def convert_to_vertical(TID,Itemset):
    Items = {}
    TID_list = {}

    Items_temp=np.unique(Itemset.astype(dtype='str').values.flatten())
    Items_temp=np.array(Items_temp)
    Items_temp=Items_temp[Items_temp!='None']

    for k in range(len(Items_temp)):
        Items[k]=Items_temp[k]
        TID_list[Items[k]]=[]
    TID_key=list(TID_list.keys())

    for l in range(len(TID_key)):
        for i in range(Itemset.shape[0]):
            for j in range(Itemset.shape[1]):
                if pd.isna(Itemset.iloc[i][j])==False:
                    if TID_key[l]==Itemset.iloc[i][j]:
                        TID_list[TID_key[l]].append(int(TID.values[i]))

    return Items,TID_list
```

2.4 Hàm kiểm tra minsup:

```python
[8]: def check_minsup(Items,TID_list,minsup):
    n=len(Items)
    Items_value=list(Items.values())
    Items_key=list(Items.keys())
    for i in range(n):
        if  len(Items)>0:
            if len(TID_list[str(Items_value[i])])< minsup :
                del TID_list[str(Items_value[i])]
                del Items[Items_key[i]]
        if len(Items)==0:
            return {},{}

    return Items,TID_list
```

2.5 Hàm kết hợp Items:

```python
[9]: def join_Itemset(Items):
        Items_key=list(Items.keys())
        Items_value=list(Items.values())
        Items_temp={}
        k=0

        for i in range(len(Items)):
            for j in range(i+1,len(Items)):
                Items_temp[k]=[]
                if str(Items_value[0]).count(',')==0:
                    Items_temp[k].append(Items[Items_key[i]])
                    Items_temp[k].append(Items[Items_key[j]])
                else:
                    Items_temp[k]= [*set(Items[Items_key[i]]+Items[Items_key[j]])]

                    if str(Items_temp[k]).count(',')!=str(Items_value[0]).count(',')+1:
                        del Items_temp[k]
                k=k+1
        for v in Items_temp.values():
            v.sort()
        result = {}
        for key,value in Items_temp.items():
            if value not in result.values():
                result[key] = value
        return result
```

2.6 Hàm kết hợp TID_list:

Hàm bổ trợ so sánh 2 array:

```python
[10]: def compare(arr1,arr2):
        arr=[]
        k=0
        for i in range(len(arr1)):
            for j in range(len(arr2)):
                if arr1[i]==arr2[j]:
                    arr.append(arr1[i])
        return arr
```

```
[11]: def join_TID_list(Items,TID_list):
          Items_value=list(Items.values())
          TID_value=list(TID_list.values())
          TID_list_temp={}
          #temp1=[]
          for i in range(len(Items_value)):
              for j in range(len(Items_value[i])):
                  for k in range(len(Items_value[i])):
                      if k==0:
                          temp1=TID_list[Items_value[i][0]]
                      else:
                          temp1=temp
                      temp2=TID_list[Items_value[i][k]]
                      temp=(compare(temp1,temp2))

                  TID_list_temp[str(Items_value[i])]=temp
          return TID_list_temp
```

2.7 Thuật toán Vertical Apriori gọi các hàm trên:

```
[12]: def VerticalApriori(TID,Itemset,minsup):
          F={}
          k=0
          Items,TID_list=convert_to_vertical(TID,Itemset)
          TID_list_temp=TID_list
          F[k]=TID_list

          while 1:
              Items=join_Itemset(Items)
              TID_list=join_TID_list(Items,TID_list_temp)
              Items,TID_list=check_minsup(Items,TID_list,minsup)
              if len(Items)>0:
                  k=k+1
                  F[k]=TID_list
              else:
                  return F
      res=VerticalApriori(TID,Itemset,3)
      res
```

2.8 Kết quả:

```
[12]: {0: {'Apple': [1, 4, 5, 7, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21],
       'Bread': [1, 2, 3, 5, 7, 9, 10, 11, 13, 14, 15, 16, 17, 19, 20, 21, 22],
       'Butter': [1, 2, 3, 4, 5, 7, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20],
       'Chips': [1, 4, 5, 6, 7, 8, 11, 13, 16, 17, 18, 19, 21, 22],
       'Milk': [1, 2, 3, 5, 6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22],
       'Wine': [1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 19, 20, 21]},
      1: {"['Apple', 'Bread']": [1, 5, 7, 9, 11, 14, 15, 16, 17, 19, 20, 21],
       "['Apple', 'Butter']": [1, 4, 5, 7, 11, 12, 15, 16, 17, 18, 19, 20],
       "['Apple', 'Chips']": [1, 4, 5, 7, 11, 16, 17, 18, 19, 21],
       "['Apple', 'Milk']": [1, 5, 12, 14, 15, 16, 17, 18, 19, 20, 21],
       "['Apple', 'Wine']": [1, 5, 7, 9, 12, 14, 15, 16, 19, 20, 21],
       "['Bread', 'Butter']": [1, 2, 3, 5, 7, 10, 11, 13, 15, 16, 17, 19, 20],
       "['Bread', 'Chips']": [1, 5, 7, 11, 13, 16, 17, 19, 21, 22],
       "['Bread', 'Milk']": [1, 2, 3, 5, 10, 13, 14, 15, 16, 17, 19, 20, 21, 22],
       "['Bread', 'Wine']": [1, 2, 5, 7, 9, 10, 13, 14, 15, 16, 19, 20, 21],
       "['Butter', 'Chips']": [1, 4, 5, 7, 11, 13, 16, 17, 18, 19],
       "['Butter', 'Milk']": [1, 2, 3, 5, 10, 12, 13, 15, 16, 17, 18, 19, 20],
       "['Butter', 'Wine']": [1, 2, 5, 7, 10, 12, 13, 15, 16, 19, 20],
       "['Chips', 'Milk']": [1, 5, 6, 8, 13, 16, 17, 18, 19, 21, 22],
       "['Chips', 'Wine']": [1, 5, 6, 7, 8, 13, 16, 19, 21],
       "['Milk', 'Wine']": [1, 2, 5, 6, 8, 10, 12, 13, 14, 15, 16, 19, 20, 21]},
      2: {"['Apple', 'Bread', 'Butter']": [1, 5, 7, 11, 15, 16, 17, 19, 20],
       "['Apple', 'Bread', 'Chips']": [1, 5, 7, 11, 16, 17, 19, 21],
       "['Apple', 'Bread', 'Milk']": [1, 5, 14, 15, 16, 17, 19, 20, 21],
       "['Apple', 'Bread', 'Wine']": [1, 5, 7, 9, 14, 15, 16, 19, 20, 21],
       "['Apple', 'Butter', 'Chips']": [1, 4, 5, 7, 11, 16, 17, 18, 19],
       "['Apple', 'Butter', 'Milk']": [1, 5, 12, 15, 16, 17, 18, 19, 20],
       "['Apple', 'Butter', 'Wine']": [1, 5, 7, 12, 15, 16, 19, 20],
       "['Apple', 'Chips', 'Milk']": [1, 5, 16, 17, 18, 19, 21],
       "['Apple', 'Chips', 'Wine']": [1, 5, 7, 16, 19, 21],
       "['Apple', 'Milk', 'Wine']": [1, 5, 12, 14, 15, 16, 19, 20, 21],
       "['Bread', 'Butter', 'Chips']": [1, 5, 7, 11, 13, 16, 17, 19],
       "['Bread', 'Butter', 'Milk']": [1, 2, 3, 5, 10, 13, 15, 16, 17, 19, 20],
       "['Bread', 'Butter', 'Wine']": [1, 2, 5, 7, 10, 13, 15, 16, 19, 20],
       "['Bread', 'Chips', 'Milk']": [1, 5, 13, 16, 17, 19, 21, 22],
       "['Bread', 'Chips', 'Wine']": [1, 5, 7, 13, 16, 19, 21],
       "['Bread', 'Milk', 'Wine']": [1, 2, 5, 10, 13, 14, 15, 16, 19, 20, 21],
       "['Butter', 'Chips', 'Milk']": [1, 5, 13, 16, 17, 18, 19],
       "['Butter', 'Chips', 'Wine']": [1, 5, 7, 13, 16, 19],
       "['Butter', 'Milk', 'Wine']": [1, 2, 5, 10, 12, 13, 15, 16, 19, 20],
       "['Chips', 'Milk', 'Wine']": [1, 5, 6, 8, 13, 16, 19, 21]},
```

```
3: {"['Apple', 'Bread', 'Butter', 'Chips']": [1, 5, 7, 11, 16, 17, 19],
 "['Apple', 'Bread', 'Butter', 'Milk']": [1, 5, 15, 16, 17, 19, 20],
 "['Apple', 'Bread', 'Butter', 'Wine']": [1, 5, 7, 15, 16, 19, 20],
 "['Apple', 'Bread', 'Chips', 'Milk']": [1, 5, 16, 17, 19, 21],
 "['Apple', 'Bread', 'Chips', 'Wine']": [1, 5, 7, 16, 19, 21],
 "['Apple', 'Bread', 'Milk', 'Wine']": [1, 5, 14, 15, 16, 19, 20, 21],
 "['Apple', 'Butter', 'Chips', 'Milk']": [1, 5, 16, 17, 18, 19],
 "['Apple', 'Butter', 'Chips', 'Wine']": [1, 5, 7, 16, 19],
 "['Apple', 'Butter', 'Milk', 'Wine']": [1, 5, 12, 15, 16, 19, 20],
 "['Apple', 'Chips', 'Milk', 'Wine']": [1, 5, 16, 19, 21],
 "['Bread', 'Butter', 'Chips', 'Milk']": [1, 5, 13, 16, 17, 19],
 "['Bread', 'Butter', 'Chips', 'Wine']": [1, 5, 7, 13, 16, 19],
 "['Bread', 'Butter', 'Milk', 'Wine']": [1, 2, 5, 10, 13, 15, 16, 19, 20],
 "['Bread', 'Chips', 'Milk', 'Wine']": [1, 5, 13, 16, 19, 21],
 "['Butter', 'Chips', 'Milk', 'Wine']": [1, 5, 13, 16, 19]},
4: {"['Apple', 'Bread', 'Butter', 'Chips', 'Milk']": [1, 5, 16, 17, 19],
 "['Apple', 'Bread', 'Butter', 'Chips', 'Wine']": [1, 5, 7, 16, 19],
 "['Apple', 'Bread', 'Butter', 'Milk', 'Wine']": [1, 5, 15, 16, 19, 20],
 "['Apple', 'Bread', 'Chips', 'Milk', 'Wine']": [1, 5, 16, 19, 21],
 "['Apple', 'Butter', 'Chips', 'Milk', 'Wine']": [1, 5, 16, 19],
 "['Bread', 'Butter', 'Chips', 'Milk', 'Wine']": [1, 5, 13, 16, 19]},
5: {"['Apple', 'Bread', 'Butter', 'Chips', 'Milk', 'Wine']": [1, 5, 16, 19]}}
```

CHECK WITH ECLAT

```
[26]:   # importing the ECLAT module
        from pyECLAT import ECLAT
        # loading transactions DataFrame to ECLAT class
        eclat = ECLAT(data=Itemset)
        # DataFrame of binary values
        eclat.df_bin
```

[26]:

	Chips	Wine	Milk	Apple	Butter	None	Bread
0	1	1	1	1	1	0	1
1	0	1	1	0	1	0	1
2	0	0	1	0	1	0	1
3	1	0	0	1	1	0	0
4	1	1	1	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	1	1	0	1
7	1	1	1	0	0	0	0
8	0	1	0	1	0	0	1
9	0	1	1	0	1	0	1
10	1	0	0	1	1	0	1
11	0	1	1	1	1	0	0
12	1	1	1	0	1	0	1
13	0	1	1	1	0	0	1
14	0	1	1	1	1	0	1
15	1	1	1	1	1	0	1
16	1	0	1	1	1	0	1
17	1	0	1	1	1	0	0
18	1	1	1	1	1	0	1
19	0	1	1	1	1	0	1
20	1	1	1	1	0	0	1
21	1	0	1	0	0	0	1

```
[30]: import pandas as pd
      result = pd.DataFrame(rule_supports.items(),columns=['Item', 'Support'])
      result
```

[30]:

	Item	Support
0	Chips & Wine	0.409091
1	Chips & Milk	0.500000
2	Chips & Apple	0.454545
3	Chips & Butter	0.454545
4	Chips & Bread	0.454545
5	Wine & Milk	0.636364
6	Wine & Apple	0.500000
7	Wine & Butter	0.500000
8	Wine & Bread	0.590909
9	Milk & Apple	0.500000
10	Milk & Butter	0.590909
11	Milk & Bread	0.636364
12	Apple & Butter	0.545455
13	Apple & Bread	0.545455
14	Butter & Bread	0.590909
15	Chips & Wine & Milk	0.363636
16	Chips & Wine & Apple	0.272727
17	Chips & Wine & Butter	0.272727
18	Chips & Wine & Bread	0.318182
19	Chips & Milk & Apple	0.318182
20	Chips & Milk & Butter	0.318182
21	Chips & Milk & Bread	0.363636
22	Chips & Apple & Butter	0.409091
23	Chips & Apple & Bread	0.363636
24	Chips & Butter & Bread	0.363636
25	Wine & Milk & Apple	0.409091
26	Wine & Milk & Butter	0.454545
27	Wine & Milk & Bread	0.500000
28	Wine & Apple & Butter	0.363636

29	Wine & Apple & Bread	0.454545
30	Wine & Butter & Bread	0.454545
31	Milk & Apple & Butter	0.409091
32	Milk & Apple & Bread	0.409091
33	Milk & Butter & Bread	0.500000
34	Apple & Butter & Bread	0.409091
35	Chips & Wine & Milk & Apple	0.227273
36	Chips & Wine & Milk & Butter	0.227273
37	Chips & Wine & Milk & Bread	0.272727
38	Chips & Wine & Apple & Butter	0.227273
39	Chips & Wine & Apple & Bread	0.272727
40	Chips & Wine & Butter & Bread	0.272727
41	Chips & Milk & Apple & Butter	0.272727
42	Chips & Milk & Apple & Bread	0.272727
43	Chips & Milk & Butter & Bread	0.272727
44	Chips & Apple & Butter & Bread	0.318182
45	Wine & Milk & Apple & Butter	0.318182
46	Wine & Milk & Apple & Bread	0.363636
47	Wine & Milk & Butter & Bread	0.409091
48	Wine & Apple & Butter & Bread	0.318182
49	Milk & Apple & Butter & Bread	0.318182
50	Chips & Wine & Milk & Apple & Butter	0.181818
51	Chips & Wine & Milk & Apple & Bread	0.227273
52	Chips & Wine & Milk & Butter & Bread	0.227273
53	Chips & Wine & Apple & Butter & Bread	0.227273
54	Chips & Milk & Apple & Butter & Bread	0.227273
55	Wine & Milk & Apple & Butter & Bread	0.272727
56	Chips & Wine & Milk & Apple & Butter & Bread	0.181818