Task 1:

ID	Transactions
1	A,B,C,D,F,G,H
2	A,B,C,D
3	A,B,C,D
4	A,B,D
5	B,C,E

|T| = Total num of transactions = 5

Minsup = 0.25
Minconf = 0.80

Set	Sup
{A}	$\frac{4}{5} = 0.8$
{B}	$\frac{5}{5} = 1$
{C}	$\frac{4}{5} = 0.8$
{D}	$\frac{4}{5} = 0.8$
{E}	$\frac{1}{5} = 0.2$
{F}	$\frac{1}{5} = 0.2$
{G}	$\frac{1}{5} = 0.2$
{H}	$\frac{1}{5} = 0.2$

→

Set	Sup	Conf
{A,B}	$\frac{4}{5} = 0.8$	A→B $\frac{4}{4} = 1$ B→A $\frac{4}{5} = 0.8$
{A,C}	$\frac{3}{5} = 0.6$	A→C C→A $\frac{3}{4} = 0.75$
{A,D}	$\frac{4}{5} = 0.8$	A→D D→A $\frac{4}{4} = 1$
{B,C}	$\frac{4}{5} = 0.8$	B→C $\frac{4}{5} = 0.8$ C→B $\frac{4}{4} = 1$
{B,D}	$\frac{4}{5} = 0.8$	B→D $\frac{4}{5} = 0.8$ D→B $\frac{4}{4} = 1$
{C,D}	$\frac{3}{5} = 0.6$	C→D D→C $\frac{3}{4} = 0.75$

→

Set	Sup	Conf
{A,B,C}	$\frac{3}{5} = 0.6$	A→B,C $\frac{3}{4} = 0.75$ B,C→A $\frac{3}{4} = 0.75$ B→A,C $\frac{3}{4} = 0.75$ A,C→B $\frac{3}{3} = 1$ C→A,B $\frac{3}{4} = 0.75$ A,B→C $\frac{3}{4} = 0.75$
{A,B,D}	$\frac{4}{5} = 0.8$	A→B,D $\frac{4}{4} = 1$ B,D→A $\frac{4}{4} = 1$ B→A,D $\frac{4}{4} = 1$ A,D→B $\frac{4}{4} = 1$ D→A,B $\frac{4}{4} = 1$ A,B→D $\frac{4}{4} = 1$
{A,C,D}	$\frac{3}{5} = 0.6$	A→C,D $\frac{3}{4} = 0.75$ C,D→A $\frac{3}{3} = 1$ C→A,D $\frac{3}{4} = 0.75$ A,D→C $\frac{3}{4} = 0.75$ D→A,C $\frac{3}{4} = 0.75$ A,C→D $\frac{3}{3} = 1$
{B,C,D}	$\frac{3}{5} = 0.6$	B→C,D $\frac{3}{5} = 0.6$ C,D→B $\frac{3}{3} = 1$ C→B,D $\frac{3}{4} = 0.75$ B,D→C $\frac{3}{4} = 0.75$ D→B,C $\frac{3}{4} = 0.75$ B,C→D $\frac{3}{3} = 1$

→

Set	Sup	Conf
{A,B,C,D}	$\frac{3}{5} = 0.6$	A→BCD $\frac{3}{4} = 0.75$ BCD→A $\frac{3}{3} = 1$ B→ACD $\frac{3}{4} = 0.75$ ACD→B $\frac{3}{3} = 1$ C→ABD $\frac{3}{4} = 0.75$ ABD→C $\frac{3}{4} = 0.75$ D→ABC $\frac{3}{4} = 0.75$ ABC→D $\frac{3}{3} = 1$ AB→CD $\frac{3}{4} = 0.75$ CD→AB $\frac{3}{3} = 1$ AC→BD $\frac{3}{3} = 1$ BD→AC $\frac{3}{4} = 0.75$ AD→BC $\frac{3}{4} = 0.75$ BC→AD $\frac{3}{4} = 0.75$

Rules produced:

- A→B
- B→A
- A→D
- D→A
- B→C
- C→B
- B→D
- D→B
- AC→B
- A→BD
- BD→A
- B→AD
- AD→B
- D→AB
- AB→D
- CD→A
- AC→D
- CD→A
- ABC→D
- ACD→B
- BCD→A
- AC→BD
- CD→AB

Frequent itemsets from first iteration are {A}, {B}, {C}, {D}, {E}, {F}, {G}, {H} with only itemsets {A}, {B}, {C}, {D} that meet the requirement of minsup≥0.25 and can proceed to 2nd iteration which are {AB}, {AC}, {AD}, {BC}, {BD}, {CD}, the rest of the itemsets are pruned.

All itemsets in 2nd iteration meet the requirement of minsup≥0.25 however not all the rules of the 2nd iteration meet the minconf≥0.8 requirement, but since the support values are ≥0.25 they can proceed to the 3rd iteration. The rules that are lower than the specified minconf are excluded from the rules produced list.

All itemsets in 3rd iteration with itemsets {ABC}, {ABD}, {ACD}, {BCD} have support values ≥0.25 so they will continue to the 4th iteration. Some rules don't meet the minconf≥0.8 requirement so they are discarded from the rules produced list.

4th iteration consist of only {ABCD} and will be the last iteration as there is no further combination to create a 5th iteration. It has support value of 1 which is ≥0.25 so it will be classified as frequent as well. Some rules are discarded as the confidence values of the rules are less than the specified minconf.

*note: Itemsets' support values and rules' confidence values that don't meet the minsup and minconf are highlighted in red.

Task 2:

→ Done in Jupyter notebook.

*note: In my program the minsup is `minsup*length of dataframe` so to calculate the support in % divide the support found by length of dataframe. This method is more simple to calculate the support for me, that's why I did that, having to divide it will require extra steps which might take longer time for the program to run.

Task 3:

```
   itemset  sup
0  {B, A}    4
1  {C, A}    3
2  {D, A}    4
3  {B, C}    4
4  {B, D}    4
5  {C, D}    3
{'A->B': 1.0, 'B->A': 0.8, 'A->D': 1.0, 'D->A': 1.0, 'B->C': 0.8, 'C->B': 1.0, 'B->D': 0.8, 'D->B': 1.0}

        itemset  sup
0     {B, C, A}    3
1     {B, D, A}    4
2  {B, C, D, A}    3
3     {C, D, A}    3
4     {B, C, D}    3
{'A, C->B': 1.0, 'A->B,D': 1.0, 'B, D->A': 1.0, 'B->A,D': 0.8, 'A, D->B': 1.0, 'D->A,B': 1.0, 'A, B->D': 1.0, 'B, C, D-
>A': 1.0, 'A, C, D->B': 1.0, 'A, B, C->D': 1.0, 'C, D->A': 1.0, 'A, C->D': 1.0, 'C, D->B': 1.0}
```

*note: In my program the minsup is `minsup*length of dataframe` so to calculate the support in % divide the support found by length of dataframe. Also the screenshot image is just to show a bit of result I got from my program using Task1's data, there are some more results that aren't in the screenshot. I also have a filtering of data where those < minsup are pruned, the image above is only to show that my program was able to calculate the support and confidence values correctly.

Task 4:

There are so many data that goes with "baking needs → bread and cake" and "baking needs → milk cream" and other baking related transactions. It's kind of interesting since it seems that the transactions in supermarket.csv have a lot of people who are fond of baking cakes, thus the high number of support values related to baking, cake ingredients, etc.

Task 5:

Propose a sector or application for which you might use association rule mining, and discuss why it could be useful.

→ An example application in real-world that can use association rule mining would be Medicine. Association rules in medicine can help doctors to find the pattern of what medicines to use for patients with a particular sickness. Many diseases shares the same systems and this way doctors may be able to compare the symptoms relationship with other cases, determine its conditional probability and which medicine to treat the disease. As new diagnoses are made, this can help to adapt rules to to reflect new updated data.