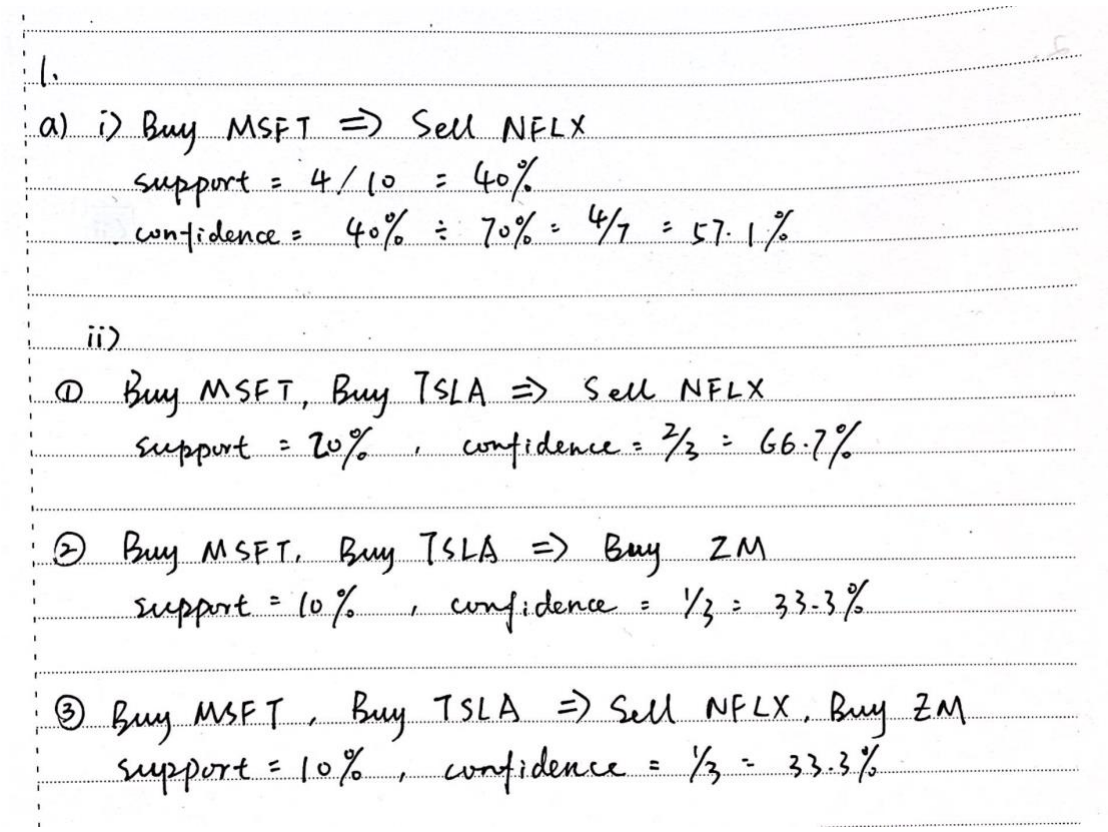
Question 1:



b)

1-itemset	count
Buy MSFT	**7**
~~Sell MSFT~~	~~0~~
Buy NFLX	**2**
Sell NFLX	**5**
Buy TSLA	**6**
~~Sell TSLA~~	~~1~~
Buy ZM	**3**
Sell ZM	**2**

2-itemset	count
~~Buy MSFT, Buy NFLX~~	~~1~~
Buy MSFT, Sell NFLX	**4**
Buy MSFT, Buy TSLA	**3**
~~Buy MSFT, Buy ZM~~	~~1~~
~~Buy MSFT, Sell ZM~~	~~1~~
~~Buy NFLX, Buy TSLA~~	~~1~~
~~Buy NFLX, Buy ZM~~	~~1~~
~~Buy NFLX, Sell ZM~~	~~0~~
Sell NFLX, Buy TSLA	**3**
Sell NFLX, Buy ZM	**2**

~~Sell NFLX, Sell ZM~~	~~0~~
Buy TSLA, Buy ZM	**3**
~~Buy TSLA, Sell ZM~~	~~1~~

3-itemset	count
Buy MSFT, Sell NFLX, Buy TSLA	2
Sell NFLX, Buy TSLA, Buy ZM	2

Question 2:

	Have smart balance	Like selfie	Both	Total
Have driver license	3250	3750	2500	5000
No driver license	2750	4000	2250	5000

a)
min_supp = 10 % i.e. 1000 people

1-itemset	count
Have license	5000
No license	5000
Like selfie	7750
Not selfie	2250
Have BW	6000
No BW	4000

2-itemset	count
Have license, Like selfie	3750
Have license, Not selfie	1250
Have license, Have BW	3250
Have license, No BW	1750
No license, Like selfie	4000
No license, Not selfie	1000
No license, Have BW	2750
No license, No BW	2250
Like selfie, Have BW	4750
Like selfie, No BW	3000
Not selfie, Have BW	1250
Not selfie, No BW	1000

3-itemset	count
Have license, Like selfie, Have BW	**2500**
Have license, Like selfie, No BW	**1250**
~~Have license, Not selfie, Have BW~~	~~750~~
~~Have license, Not selfie, No BW~~	~~500~~
No license, Like selfie, Have BW	**2250**
No license, Like selfie, No BW	**1750**
~~No license, Not selfie, Have BW~~	~~500~~
~~No license, Not selfie, No BW~~	~~500~~

There are 4 frequent itemsets. For each, the strong rules are listed here:

1. { Have license, Like selfie, Have BW} support = 25%

Rule	Confidence
Have license => Like selfie, Have BW	50%
Have license, Like selfie => Have BW	66.7%
Have license, Have BW => Like selfie	76.9%
Like selfie, Have BW => Have license	52.6%

2. { Have license, Like selfie, No BW} support = 12.5%

Rule	Confidence
Have license, No BW => Like selfie	71.4%

3. { No license, Like selfie, Have BW} support = 22.5%

Rule	Confidence
No license, Like selfie => Have BW	56.3
No license, Have BW => Like selfie	81.8

4. { No license, Like selfie, No BW} support = 17.5%

Rule	Confidence
No license, No BW => Like selfie	77.8
Like selfie, No BW => No license	58.3

b)

2 - itemsets	count	support
Have license, Like selfie	3750	37.5%
Have license, Not selfie	1250	12.5%
Have license, Have BW	3250	32.5%
Have license, No BW	1750	17.5%
No license, Like selfie	4000	40%
No license, Not selfie	1000	10%
No license, Have BW	2750	27.5%
No license, No BW	2250	22.5%
Like selfie, Have BW	4750	47.5%
Like selfie, No BW	3000	30%
Not selfie, Have BW	1250	12.5%
Not selfie, No BW	1000	10%

Q3:

a)

Step 1:

User ID	Sequence
A	<(E, F), (B, E, F), (E, F)>
B	<(C, D, E), (A, C, E), (C, D)>
C	<(B, D), (B, D), (B, D)>
D	<(B, C, E, F), (C, E, F), (B, C, E)>
E	<(A, B, D, F) , (A, B, D, F) , (A, D, F)>
F	<(A, D, E) , (A, D, E) , (A, E)>

The largest itemset size is related to the min_sup:

Min_sub	Largest itemset size
>= 0/6	4
>= 2/6	2
>= 4/6	1

However, no matter in which situation, the longest sequence length should be 3.

b)

Step 2: min_sup = 40% i.e. 3 users

Frequent itemset	Mapped to
(A)	1
(B)	2
(D)	3
(E)	4
(F)	5
(B F)	6

Step 3:

User ID	Sequence	Mapping
D	<(B, C, E, F), (C, E, F), (B, C, E)>	<{2, 4, 5, 6}, {4, 5}, {2, 4}>
E	<(A, B, D, F) , (A, B, D, F) , (A, D, F)>	<{1, 2, 3, 5, 6}, {1, 2, 3, 5, 6}, {1, 3, 5}>

c)

Sequences:

length 3: $7 * 7 * 3 = 147$

length 2: $7 * 7 + 7 * 2 + 4 * 3 = 75$

length 1: A, C, D, E, AC, AE, CD, CE, DE, CDE, ACE: totally 11

Therefore, there are $147 + 75 + 11 = 233$ possible sequences.

d)
Length 1: <(A)>, <(B)>
Length 2: <(B), (F)>, <(B, F), (F)>,
Length 3: <(E), (E), (E)> only one