Synthetic Dataset

Item	Min	Max	Alsc	TRUE
1	A	A	A	**A**
2	A	B	B	**B**
3	B	C	C	**C**
4	D	D	D	**D**
6	E	E	E	**E**
7	B	F	F	**F**
8	A	G	G	**G**
9	A	H	H	**H**
10	J	J	J	**J**
11	K	K	K	**K**
12	L	L	A	**L**
...	…	…	…	**...**

a record

Filter

New Filtered Synthetic Dataset

Item	Min	Max	Alsc	TRUE
1	A	A	A	**A**
2	A	B	B	**B**
3	B	C	C	**C**
...	…	…	…	**...**

New Filtered Synthetic Dataset have classes A,B,C.

Synthetic Dataset have classes A,B,C,D,E,F,G,H,I,J,K,L.

Each class represents a Q-vector as follows:

A	11001	G	11100
B	00110	H	00001
C	10010	I	10000
D	11111	J	10001
E	11100	K	11000
F	01101	L	11011

Real Dataset

Item	Min	Max	Alsc	TRUE
1	A	A	A	**?**
2	F	B	B	**?**
3	B	A	A	**?**
4	E	C	C	**?**
5	A	A	B	**?**
6	K	B	B	**?**
7	A	C	C	**?**

Calculate frequency in this color area

Frequency table	
A	8
B	6
C	4
E	1
F	1
K	1

1) Calculate frequencies of Q-vectors in Real Dataset

2) We assume those most frequent classes (>5%)are potential classes (A,B,C) for TRUE value in Real Dataset.

3) We just only allow records or rows with classes A,B,C in TRUE in new Filtered synthetic data. Delete rest of the records or rows with classes D,E,F,G,H,I,J,K,L in TRUE value.