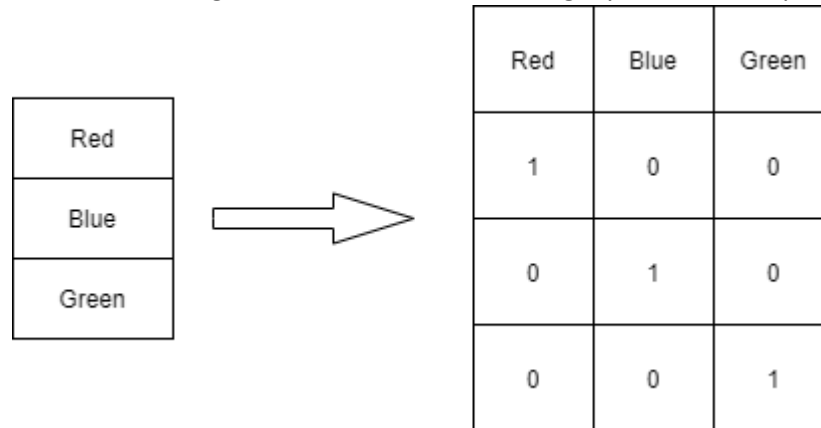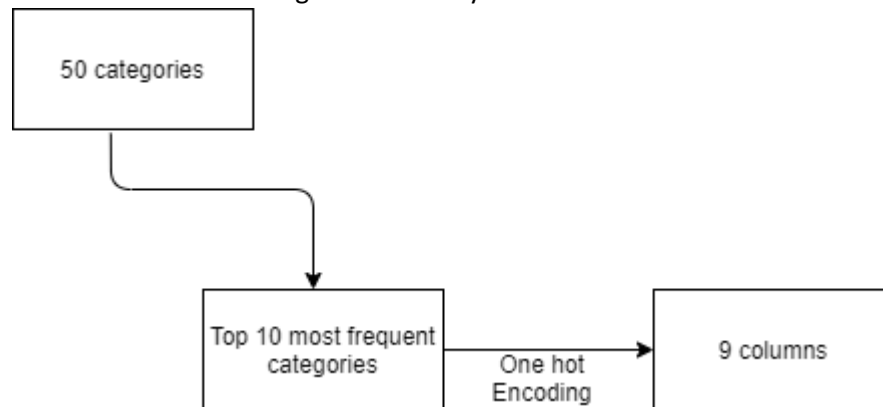Feature Engineering

1) Nominal – F/M
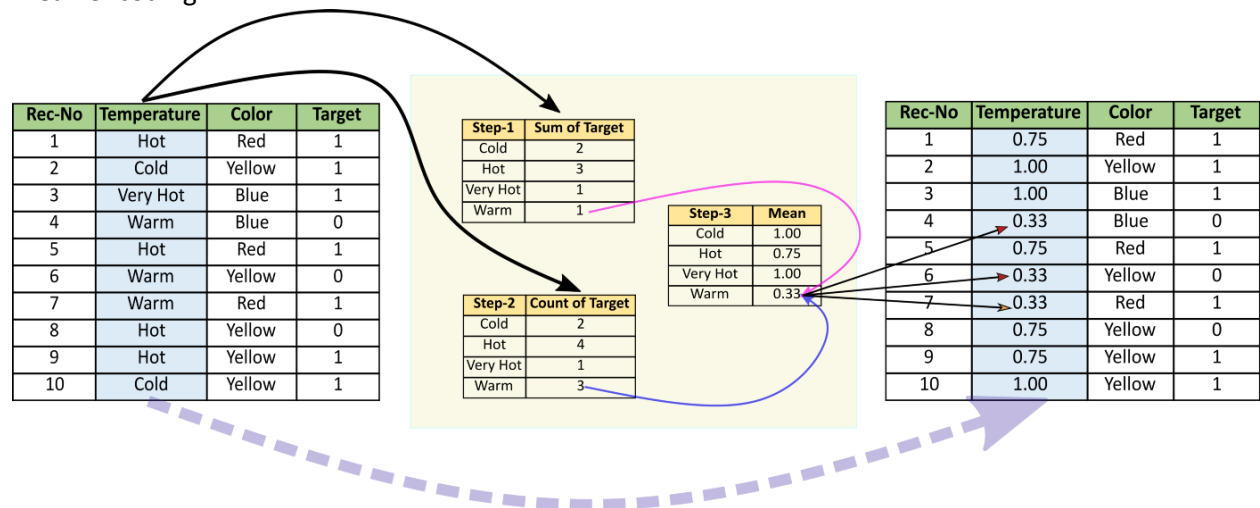   a. One hot encoding - Don't use one hot encoding if you have multiple categories

	Red	Blue	Green
Red	1	0	0
Blue	0	1	0
Green	0	0	1

   b. One hot encoding multiple categories – Pickup top value from the multiple categories and do one hot encoding for those only

50 categories → Top 10 most frequent categories → One hot Encoding → 9 columns

   c. Mean encoding –

Rec-No	Temperature	Color	Target
1	Hot	Red	1
2	Cold	Yellow	1
3	Very Hot	Blue	1
4	Warm	Blue	0
5	Hot	Red	1
6	Warm	Yellow	0
7	Warm	Red	1
8	Hot	Yellow	0
9	Hot	Yellow	1
10	Cold	Yellow	1

Step-1	Sum of Target
Cold	2
Hot	3
Very Hot	1
Warm	1

Step-2	Count of Target
Cold	2
Hot	4
Very Hot	1
Warm	3

Step-3	Mean
Cold	1.00
Hot	0.75
Very Hot	1.00
Warm	0.33

Rec-No	Temperature	Color	Target
1	0.75	Red	1
2	1.00	Yellow	1
3	1.00	Blue	1
4	0.33	Blue	0
5	0.75	Red	1
6	0.33	Yellow	0
7	0.33	Red	1
8	0.75	Yellow	0
9	0.75	Yellow	1
10	1.00	Yellow	1

2) Ordinal – B.E/M.S/B.com
   a. Label encoding

Btech	
Master's	
High School	
PHD	

→

PHD	4
Master's	3
Btech	2
High School	1
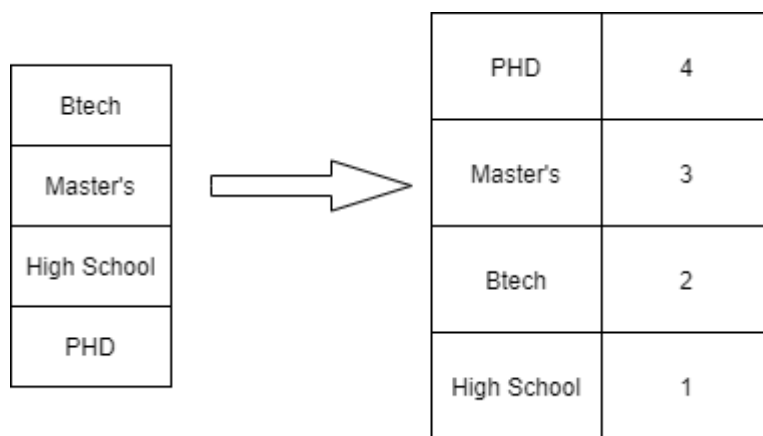
b. Target guided ordinal encoding

Column	O/P
Btech	1
PHD	1
Master's	0
High School	0
PHD	1
Btech	0
Master's	0
High School	0
High School	1

→

Mean of Btech = 0.5

Mean of PHD = 1

Mean of Master's = 0

Mean of High School = 0.33

→

Column	Rank
Btech	2
PHD	1
Master's	4
High School	3
PHD	1
Btech	2
Master's	4
High School	3
High School	3