

Q2. Analyze the Aircrash dataset and find out the following: (Dataset: Airplane_Crashes_Since_1908.csv)

c. 'Summary' field has the details about the crashes. Find the reasons of the crash and categorize them in different clusters i.e Fire, shot down, weather (for the 'Blanks' in the data category can be UNKNOWN) you are open to make clusters of your choice but they should not exceed 7.

Answer:

Matrix of first 10 transaction (true/false matrix) to develop the 7 overlapping clusters

id	Unknown	fire	shot-down	weather	storm	engine failure	pilot error
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	1	0	0
4	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	1	0	0	0	0
10	0	0	1	0	0	0	0

Airline Crash Reasons

