

1. Consider the traffic accident data set shown in [Table 6.10](#).

Table 6.10. Traffic accident data set.

WeatherCondition	Driver'sCondition	TrafficViolation	Seat Belt	CrashSeverity
Good	Alcohol-impaired	Exceed speed limit	No	Major
Bad	Sober	None	Yes	Minor
Good	Sober	Disobey stop sign	Yes	Minor
Good	Sober	Exceed speed limit	Yes	Major
Bad	Sober	Disobey traffic signal	No	Major
Good	Alcohol-impaired	Disobey stop sign	Yes	Minor
Bad	Alcohol-impaired	None	Yes	Major
Good	Sober	Disobey traffic signal	Yes	Major
Good	Alcohol-impaired	None	No	Major
Bad	Sober	Disobey traffic signal	No	Major
Good	Alcohol-impaired	Exceed speed limit	Yes	Major
Bad	Sober	Disobey stop sign	Yes	Minor

- a. Show a binarized version of the data set.
- b. What is the maximum width of each transaction in the binarized data?
- c. Assuming that the support threshold is 30%, how many candidate and frequent itemsets will be generated?
- d. Create a data set that contains only the following asymmetric binary attributes: (Weather = Bad, Driver's condition = Alcohol-impaired, Traffic violation = Yes, Seat Belt = No, Crash Severity = Major). For Traffic violation, only None has a value of 0. The rest of the attribute values are assigned to 1. Assuming that the support threshold is 30%, how many candidate and frequent itemsets will be generated?
- e. Compare the number of candidate and frequent itemsets generated in parts (c) and (d).