

CSE3212 Data Mining
Week 4 Data preparation
Additional exercise

Year	Average Grade	Database unit	Data Mining Unit	Gender	Work	AGE
2	P	Pass	Fail	F	Y	20
2	HD	??	Pass	M	N	19
2	D	Fail	Pass	F	Y	30
3	D	Pass	Pass	M	??	20
3	P	Fail	Fail	F	Y	21
3	C	Fail	Pass	M	Y	??
??	C	Pass	Pass	M	N	20
3	D	Pass	Pass	F	Y	20
3	HD	Pass	Pass	M	Y	38
3	D	Pass	??	M	N	20

Table 1. Student performance.

Perform data preparation for data mining of the data in table 1.

- (a) Fill in all the missing data and justify your choice of method.
- (b) Devise a necessary coding method to the data and justify the method.
- (c) Find an example of attributes that are independent.
- (d) What potential clusters can you see from the data (use your interpretation of the data rather than using a particular clustering method). Explain the reasons why you choose to cluster the data the way you do.