

Student: MEHER ARASADAPercentage: 52%

QUESTION	EXCEPTIONAL 5	MERITORIOUS 4	COMPETENT 3	PASS 2	FAIL 1
Dataset description (15%)	Infers multiple pre-processing AND analysis techniques from descriptive statistics	Infers pre-processing OR analysis techniques from descriptive statistics	Provides descriptive statistics for the dataset	Correctly defines the data types; describes the original data source	Misinterprets data types and meanings
Data pre-processing (10%+10%)	Performs multiple complex primary pre-processing operations	Codes data sensibly and discusses any data removal or manipulation	Performs obvious coding (Yes/No) OR justifies data manipulation	Describes the data used but does not perform any significant pre-processing	Performs no primary pre-processing
	Performs multiple secondary pre-processing operations	Recodes data to allow further analyses AND adds new calculated fields	Recodes data to allow further analyses OR adds new calculated fields	Performs appropriate secondary pre-processing	Performs no secondary pre-processing
Data analysis - data exploration (15%+10%)	Uses multiple methods other than decision trees Full and detailed discussion of the selected model	Uses multiple types of algorithm AND multiple options with each of these.	Uses multiple types of algorithm appropriately and shows a logical progression within an analysis	Uses the same type of algorithm in each analysis	Incorrect, misleading or missing
	Fully tests each model in each analysis	Uses MVT methodology correctly	Includes prediction and comparison of algorithms	Only creates models on the total dataset	Incorrect, misleading or missing
Reporting of results (20%)	Four complete analyses with discussion of the implications of the results	Three complete analyses with discussion of the implications of the results	Two complete analyses with discussion of the implications of the results	One complete analysis with discussion of the implications of the results	Incorrect, misleading or missing
Reflection (10%)	Full and detailed reflection with discussion of how the analysis might be repeated and extended	Discussion of how the research might be repeated OR extended.	Discussion of satisfying analyses and problems encountered.	List of skills learned.	Incorrect, misleading or missing
Report writing (10%)	Reads like a technical review article	Good use of signposting	Clear sectioning of content	Basic use of introduction, body and conclusion; correctly referenced	No clear structure; incorrectly referenced

Analyses are linked to research questions

You should make it clear that this is an artificial dataset, which is why it needs so little preprocessing, and that as a result it may not show the full range of random behaviour of a real-world dataset.

7. Classifier — You need to say what question you're trying to answer by using a classifier. Why did you choose this technique, and what parameters did you use? Just taking the first 300 records is not a good idea — you need to take as large a group as possible, selected for specific reasons.

8. Clustering — Again, what is the research question? Why is cluster analysis the best way to answer it? ~~Also~~ This suggests that you did multiple versions ~~with~~ with different numbers of clusters and identified the one with the least errors, which is good, but why not use other algorithms? Or other combinations of attributes? This could then be presented as a table showing the effectiveness of each algorithm. What do the clusters actually represent?

9. Association Analysis — Again, no research question stated. Some good business suggestions made based upon this data.

10. Time Series — No research question stated. This explores the capabilities of the time series analysis functionality rather than looking for specific things. Dropping to 300 records is not a good idea unless you have a good reason for selecting those 300 records.

11. Tableau — No research question stated. This has some general material about displaying sales figures and creating forecasts.

Overall this shows a reasonable attempt to explore a dataset with a set of data mining tools, but it would have been much easier to gain marks by exploring specific research questions.