

Student: PAIGE HARRISPercentage: 65%

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

good data description, with clear descriptive data. Good discussion of the pre-processing procedure. Curious that there are no invoices for Saturday - perhaps the company is owned by observant Jews who close everything down for the Shabbat? There are easier ways than data mining to answer some of the research questions, but as example problems they work okay. Ideally you don't just want to show the results of one algorithm, but to try several different algorithms and compare the results. This allows you to get a sense of what algorithms work under different circumstances. The layout of problem - investigation - results works well!

Data Mining – Final Poster			
Student	PAIGE HARRIS		
Dataset	Online Retail Dataset		
Marking			
Abstract (Clarity, readability, flow, accuracy, completeness) Ideally the abstract should also contain a brief outline of the results.			(10 marks) 7
Dataset Description (Where did the data come from? What does the data describe? Is the sample large enough? What data formats and types are involved? What were the objectives of the project?) Good description with data attributes, types and sizes			(10 marks) 8
Data Processing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Good clear discussion of multiple pre-processing activities.			(20 marks) 16
Analysis (What analyses were carried out? Are they appropriate?) Problems stated, tool/analysis techniques identified, with comments on specific data requirements.			(20 marks) 16
Results (Are the results clearly stated? Do they meet the objectives?) Clearly presented in the right place. Results relate to each other.			(20 marks) 16
Graphic Design (Is it balanced and readable? Do the colours work? Are the graphics appropriate?) Good layout and colouring (although the green could be slightly paler :)).			(20 marks) 15
General comments Good linkage between analysis and results.			
Date	20/5/19	Agreed Final Grade (%)	78
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		

$$15^{28} - 15^{43} + 3$$

Data Mining – Final Presentation			
Student	PAIGE HARRIS		
Dataset	Online Retail Set.		
Marking			
Presentation/Slides (Clarity, readability, flow, accuracy, completeness, performance) Clear slides - text a bit small at times. Be careful with colour choices - make sure everything is visible			(10 marks) ~~7~~
Dataset Description (Where did the data come from? What does the data describe? Is the sample large enough? What data formats and types are involved? What were the objectives of the project?) Clear description of the dataset - good comparison of UK/not-UK.			(10 marks) ~~7~~
Preprocessing (What primary and secondary preprocessing was needed? How was data integrity maintained?) Good description of the calculations. Good level of detail.			(25 marks) ~~20~~
Analysis (What analyses were carried out? Are they appropriate?) Clear statement of questions. All covered within the presentation.			(20 marks) 15
Results (Are the results clearly stated? Do they meet the objectives?) Good use of visualisation to get initial ideas. Nice discussion of association test that didn't go as planned.			(25 marks) 18
Results (What further analyses could be carried out if you had more time?) Some good ideas for going further.			(20 marks) ~~14~~ 14
General comments Good clear presentation style, but try not to read from the slides!			
Date	20/5/19	Agreed Final Grade (%)	81
Tutor Signature	Nik Whitehead		
Second Marker Signature	 		