

Big Data Analytics OL5 24-25

Individual Summative Assessment.

Marks & Feedback

cs-online-programmes@york.ac.uk



Student Number: S197

Group: 6

Marker: farid.bello@york.ac.uk

The following shows which section of the marking criteria (as presented in the assessment brief) you have achieved. Where there are multiple statements defining that criterium you may have met all or the majority of them. The specific tutor feedback aims to highlight the key areas that you have/have not been successful at, as well as any aspects, such as exceeding word/page limits, that have constrained your grade. Their feedback may not align directly with the generic assessment criteria statements.

Marking Criteria	Feedback	Marks
Submission		
Supporting Files	A single .zip file was provided containing all the supporting evidence as required. The included files covered the requirements for supplying both data (.arff, .csv and/or .ipynb files) and code/models (from WEKA, Java or Python) .	
Task 1: Rental Price Investigation		
a. Approach & Results	The evidence provided is both complete and consistent and clearly presented/organised (from supplementary files; data set or code/models) and the discussion indicates that the three required areas of investigation have been addressed with appropriate techniques applied and discussed and some clear justification for the selection of those techniques with some critical thinking evident. The results presented are clear and fully consistent with the techniques discussed in relation to the investigation.	15
b. Justification	There is a clear and appropriate justification for the techniques/approach used that fully covers all three of the investigation areas and is well supported by wider reading.	8
c. Critical Evaluation	The critical evaluation provides an accurate and rational appraisal of the effectiveness of the techniques/approach used for all three of the areas of investigation, which is fully and clearly aligned with the presented results and is well supported by wider reading.	7
Task 1 Tutor Feedback	You showed some good knowledge of statistical techniques and machine learning, with clear justification for method choices. Discrete variables were correctly identified and explored with both feature importance and visual plots. Your language is often overly complex, which makes some findings harder to interpret. For example, you used RFC and MDI, etc without first stating what they are. To improve, simplify the language, include all model outputs, and ensure variable selections are better aligned with prediction goals.	
Task 2: Storing data and scalable solutions		
1a. Database Design	The requested ERD has either not been presented, is not readable or demonstrates little or no effective relational database structure (entity and attribute names are not rational or appropriate), with little or no effective normalisation, and significant errors in notation, and/or there is little or no discussion of the decisions made in the design of the relational database.	3
1b. SQL statements	The presented SQL includes all three statements and demonstrates correct construction (with no errors) for at least two statements (for the input and at least one output)	6

2. Scalable Solution	A fully valid and robust solution has been presented which fully considers the potential gains and any significant limitations. The discussion is complete with clear justification given for the proposed solution that fully considers both strengths and weaknesses and has clear evidence of comparisons being drawn with a range of alternative solutions, that is well supported by wider reading.	15
Task 2 Tutor Feedback	You provide a rich narrative on 3NF and database design principles. The ERD is well thought out and described in detail, but it is not presented in UML-standard form. There are no relationship lines, cardinalities, or explicit key markers, which are required. SQL queries are present and explained. The proposed AWS cloud-based architecture is sound, with reasoning for using EMR and Lambda. To improve, use UML-compliant ER diagrams and avoid too technical terms and abbreviations.	
Task 3: Considering public-facing application		
a. Privacy Issues	At least three valid privacy issues have been identified that are relevant to the context of the task scenario and the discussion is both complete and comprehensive, and is well supported by wider reading.	7
b. Mitigation Strategies	At least three mitigation strategies were presented that were feasible and relevant in the context of the issues discussed in the first part of the task and in the context of the task scenario, and these were well supported by wider reading.	7
Task 3 Tutor Feedback	You have discussed ethical and legal risks including overcollection, model inversion, and SQL injection. Each is supported by references and discussed across technical and managerial layers. However, the mitigations are also described but conceptually. You may give strong ties to the database. Again, you should have made it easier by discussing the risks followed by the mitigations without being overly complex.	
Overall Tutor Feedback	Overall, your responses are thoughtful, and technically capable. However, it misses important formatting and clarity requirements. To improve, simplify your language and expression. You should also strictly comply with word limits and structure your work accordingly.	
Total:		68/ 100