

Name

RUTENDO GORE

Mark

40/70

QUESTION	EXCEPTIONAL 5	MERITORIOUS 4	COMPETENT 3	PASS 2	FAIL 1	NON SUBMISSION
Schema (20%)	Good schema with full discussion of data and of alternatives	Workable schema with description of tables and attributes	Workable schema with descriptions of tables or attributes	Schema is workable but unexplained	Unworkable schema	Missing
ETL process (25%)	Full and detailed discussion of ETL process	Discusses data transformations and missing values	Discusses data transformations or missing values	Basic description of which data goes into each table	Incorrect or misleading	Missing
Report generation (25%)	Full and detailed discussion of report generation	Discusses the analytical process used	Explains the screenshots	Provides screenshots only	Incorrect or misleading	Missing
Dashboard design (20%)	Elegant design with full discussion of decisions made	Good design of relevant data; uses multiple encoding methods well	Shows relevant data and uses more than one encoding method	Shows relevant data in an understandable format	Incorrect or misleading	Missing
Report writing (10%)	Reads like a technical review article	Good use of signposting; good use of English; tables of figures and tables	Clear sectioning of content; labelling of tables and figures; reasonable English	Basic use of introduction, body and conclusion; English is comprehensible; some labelling of tables and figures	No clear structure	Missing
Data governance elements (20%)	Clear identification of all elements and related industry standard approaches	Relation to relevant external regulations	Consideration of management and operational levels in practice	Identification of obvious data governance processes	Incorrect or misleading	Missing
Problems identified (25%)	Full discussion of a wide range of potential issues	Identifies more subtle issues as well as the obvious ones	Identifies 2-3 issues other than basic data quality	Identifies basic data quality issues	Incorrect or misleading	Missing
Proposed solutions (25%)	Full discussion with reference to professional toolkits	Provides strategic level solutions to problems; may include toolkits	Provides tactical level solutions to problems; may involve tools	Provides simple data cleaning solutions	Incorrect or misleading	Missing
Illustrations and examples (20%)	All points illustrated to a professional level	All points illustrated referring back to issues raised in part 1 ETL	Most points illustrated; some refer back to issues raised in part 1 ETL	Limited examples; some relate to specified dataset	Incorrect or misleading	Missing
Report writing (10%)	Reads like a business report	Good use of signposting; generally good language use; spellchecked and grammar-checked	Clear sectioning of content; requires spellchecking and/or grammar-checking; labelling of tables and figures	Basic use of introduction, body and conclusion; language is understandable; some labelling of tables and figures	No clear structure	Missing

clear introduction.

Why do you need to remove the duplicates?

You've ~~repeatedly~~ copied all of the columns into individual tables without deciding if they should be part of a fact or a dimension. This means that you've lost the links between each invoice amount and its data. Amount should be part of the UKAEA transaction fact, while the other tables are all dimensions.

You should have included a full ETL document.

The reports and dashboard could have been done in Tableau.

There's a lot of good stuff about what data governance is and how it should work - great idea to use existing toolkits - but it needs to link to the UKAEA example more closely. How does all of this apply to the UKAEA? What problems did you find in their data that could have been fixed by better data governance.