ISYS2038 Database Design and Development Semester 2, 2021



Individual Assignment – Specifications

Released Date: Tuesday 26 October, 10.00am

Due Date : Wednesday 3 November, 11.59pm

Mark : 50% of the total assessment of this course

1) Background

In this assignment, you are required to demonstrate and apply the concepts and knowledge covered in Week 1 – Week 11. For the report, assume that you are employed as a business analyst at Best Innovative Solution (BIS) Pty Ltd. You are assigned to investigate the questions from a new client and provide the solution.

The dataset about Costello Pty. Ltd. has been collected, and you are given the following two files to apply data analytics:

- dataset (CostelloData.xlsx)
- header description (CostelloDataset-header-description.txt)

These two files are available for you to download on Canvas | Assignments | Assessment Task 3: Individual Final Assessment | Individual Assignment - Case Study, Specifications and Dataset. Further information about the dataset is listed below:

- The dataset and header description are stored in two separate files.
- The dataset consists of binary, categorical and numerical data.
- There are nine different attributes and 1000 cases (or instances).
- There are some missing values in the dataset.
- The dataset covers mainly the aspect of customer-related data.

2) Questions from the new client

You need to provide:

- (a) answers to all the following questions based on the case study
- (b) mention the word count on the cover page of your report

Question 1: Data Modelling

- Provide an Entity Relationship Diagram (ERD). For each entity, provide <u>at least one</u> attribute/ field with a <u>maximum of two</u> attributes (excluding primary key).
- Provide a Relational Model and show the most appropriate attributes/ fields, primary keys, and foreign keys. In addition, include the data type for each attribute/ field listed in each table/ entity.
- You should apply the third normal form (3NF) and show consistency in the use of notation.
- If you discover while drawing the diagrams that the narrative of the case study
 is incomplete, then provide justifications, assumptions, or reasonable
 explanations to complete the case study along with your diagrams. You could
 include the justifications, assumptions, or reasonable explanations under the
 Appendix section.

Question 2: SQL Scripts

- Based on data modelling from Question 1, create 3 queries that have the following requirements. Note that each of the query can cover one or more of the requirements.
 - o One of the 3 queries should have a calculation.
 - o One of the 3 queries should make use of Group By.
 - One of the 3 queries should include a scalar function (one that returns a value).
 - One of the 3 queries should demonstrate Nested Query (could be either standard or correlated sub-query).
 - o One of the 3 gueries should demonstrate joining tables.
 - One of the 3 queries should demonstrate the creation and testing of Trigger with test data.
- For each query,
 - o Explain and justify its business purpose and business value or impact.
 - o Provide the SQL script.
 - Explain the design of SQL script.
 - Make use of a table(s) with sample data to show the potential result/ outcome of the query.

Question 3: Data Analytics with Orange

After you have dealt with the missing values, discuss how you would predict the spending score of customers for Costello Pty. Ltd. using the knowledge of data visualisation and regression. In your explanations, include the screenshots of using Orange to produce the visualisation and prediction model. In addition, please explain how your results from data visualisation and regression could help Costello Pty. Ltd. expand its business.

Question 4: Database Administration Approaches

Describe two potential impacts on Costello Pty. Ltd. if data is not appropriately protected. For each type of impact, discuss a recommended security feature of a database management system that Costello Pty. Ltd. can use to protect its data.

Question 5: Big Data and Analytics

Suggest two applications of big data and analytics that can improve Costello Pty. Ltd. in understanding its business. Discuss the benefits where each of the suggested application would deliver to the company.

3) Report

Word Limit: 1500 – 2000 words +/- 10% (excluding Table of Content, Appendix, figures, tables and screenshots, anything beyond word limit will not be marked)

Font Size: 11pt or 12pt

Font Style: Calibri or Times New Roman

Spacing: Single or 1.5 Spacing

Document Margins: 2.5cm for top, bottom, left and right.

Some notes and guidelines (in addition to Lecture 12 slides):

- Include and discuss all the questions given in *Part 2*) Questions from the new client in your report.
- Do not need to access MySQL server to work on Question 2 SQL Scripts.
 You can make use of tables with sample data (or examples of records) to
 support your explanations. The tables could be created by typing in MS Word
 or MS Excel.

- Use a diagramming tool(s) to produce your diagrams for Question 1: Data Modelling. Do not submit diagrams that are drawn using pens/ pencils and papers.
- Organise the report with sections and subsections.
- Include justifications and explanations of the SQL queries in your report.
- Consider that the report covers a wider audience, including management and business users as well as developers.
- Label the print screens, figures, diagrams, and tables in the report properly.
- Include References and Appendix sections in your report as you see fit.
- Use Harvard referencing style if you are citing references from web resources to support your discussions.

Submission Format:

For each individual, the following documents must be submitted via Canvas. You can merge the following files and submit as a zip file.

- a single copy of the final report (PDF version)
- Assignment Cover Sheet
- a Turnitin generated report (PDF version or receipt/ evidence of submission to Turnitin if it takes too long to generate the report)
- Orange related files (Orange loadable dataset(s) in .xlsx, Orange workflow(s) in .ows)

You can go to Assignments | Assessment Task 3: Individual Final Assessment | Assessment Task 3 – Individual Assignment and click on Submit Assignment.

Use of Turnitin:

Please note that you need to use Turnitin to self-check your report for compliance of academic integrity and plagiarism detection. Report that is not checked by Turnitin will not be marked.

Turnitin is *not* the submission link of your final report (PDF version). After you have self-checked your report, you need to submit the final version of your report together with its Turnitin report (or receipt/ evidence of submission to Turnitin if it takes too long to generate the report) via the designated submission link.

To self-check your report, you can go to Assignments | Important – Other Assignment Related Info. | Turnitin - Self-Check Your Report *Not Final Submission* and click on Load Turnitin - Self-Check Your Report *Not Final Submission* in a new window.

4) Marking Rubric:

Note: Marking rubric shows the mark out of 100 points and the total mark will then be converted to 50%.

Assessed	Exceptional	Very Good	Acceptable	Needs	Poor
Components				Improvement	
Data	Outstanding	Sound	Satisfactory	Limited	No
Modelling	demonstratio	demonstrati	demonstrati	demonstration	demonstration
with ERD and	n of using	on of using	on of using	of using	or very little
Relational	appropriate	appropriate	appropriate	appropriate	demonstration
Model	data	data	data	data modelling	of using
(30 points)	modelling		modelling	techniques.	appropriate

	techniques. Diagrams produced are outstanding, elegant and have very few or no mistakes. (24.00 – 30.00 points)	modelling techniques. Diagrams produced are sound and have few mistakes. (21.00 – 23.99 points)	techniques. Diagrams are produced at satisfactory level but still have some mistakes. (18.00 – 20.99 points)	Diagrams are somewhat complete but still have a lot of mistakes. (15.00 – 17.99 points)	data modelling techniques. Diagrams are incomplete and have a lot of mistakes. (0.00 – 14.99 points)
SQL Scripts (15 points)	SQL scripts are sophisticated and logically well structured, show great complexity, use appropriate syntax, contain no error and are able to address all business purposes. Discussions/justifications are critical, well-articulated and rigorous. (12.00 – 15.00 points)	SQL scripts are logically well structured, show good complexity, use correct syntax but still contain very little errors and can address most/ all business purpose. Discussions / justifications are sound. (10.50 – 11.99 points)	SQL scripts use appropriate syntax, show some complexity but contain reasonable number of errors and can address some business purposes. Discussions / justifications are sufficient. (9.00 – 10.49 points)	SQL scripts somehow use appropriate syntax, show minimal complexity but full of errors and can address limited business purposes. Discussions/ justifications are limited or insufficient. (7.50 – 8.99 points)	SQL scripts do not use appropriate syntax, overly simplistic and are unable to address any/ very limited business purpose. Discussions/ justifications are missing or very limited. (0.00 – 7.49 points)
Data Analytics with Orange (15 points)	Excellent model chosen to produce accurate prediction and visualisation. Discussions/justifications are insightful, critical, well-articulated and rigorous. (12.00 – 15.00 points)	Good model chosen to produce accurate prediction and visualisation. Discussions / justifications are insightful and sound. (10.50 – 11.99 points)	Appropriate model chosen to produce satisfactory prediction and visualisation. Discussions / justifications are somewhat insightful and sufficient.	Appropriate model chosen to produce sensible prediction and visualisation. Discussions/ justifications are limited or insufficient and barely have insights. (7.50 – 8.99 points)	Inappropriate model chosen and unable to produce sensible prediction and visualisation. Discussions/ justifications are missing or very limited and have no insights. (0.00 – 7.49 points)

			(9.00 – 10.49 points)		
Database Administratio n Approaches (20 points)	Convincing, critical, well-articulated and rigorous discussions/justifications for the required recommenda tions. (16.00 – 20.00 points)	Sound discussions/ justifications for the required recommend ations. (14.00 – 15.99 points)	Satisfactory discussions/ justifications for the required recommend ations. (12.00 – 13.99 points)	Limited or insufficient discussions/ justifications for the required recommendati ons. (10.00 – 11.99 points)	No or very limited discussions/ justifications for the required recommendati ons. (0.00 – 9.99 points)
Big Data and Analytics (20 points)	Convincing, critical, well-articulated and rigorous discussions/justifications for the required suggestions. (16.00 – 20.00 points)	Sound discussions/ justifications for the required suggestions . (14.00 – 15.99 points)	Satisfactory discussions/ justifications for the required suggestions . (12.00 – 13.99 points)	Limited or insufficient discussions/ justifications for the required suggestions. (10.00 – 11.99 points)	No or very limited discussions/ justifications for the required suggestions. (0.00 – 9.99 points)
Overall	High Distinction (HD)	Distinction (D)	Credit (C)	Pass (P)	Fail (N)
	80 or more	70 or more	60 or more	50 or more	less than 50