## LEARNING OUTCOMES

The two parts of this assignment address the following learning outcomes:

1. Use SQL and XML to define data applications, appropriate to a specified problem
2. Use a conceptual modelling language to specify and analyse data requirements and apply the principles of database design to map a set of system requirements to an efficient (e.g. normalized) database
3. Explain and design transaction-based processing in database systems
4. Exploit techniques for storying and querying XML data

## OBJECTIVES

The objectives of this part of the assignment are:

- to develop skills in the design of data driven applications

- to develop skills in applying SQL to implement the relational design

- to develop skills in using SQL to define data applications

- to present an opportunity to practice problem solving and communication skills.

## TASKS

You work for a software sales company, doing data analysis. You have been sent three files – transactions, products and customers, along with a file description. You will first need to create a data model and load the data into a database, you will then need to create some simple reports from them.

## DELIVERABLES

1. Examine the files and determine an appropriate relationship model between them.  Create and document a physical data model detailing the field types and relationships. **[3 marks]**
2. Recommend and design one additional table that would be appropriate for reporting purposes, update your model accordingly. **[3 marks]**
3. Load all four tables into a new database schema, and provide evidence of loading scripts and row counts. **[2 marks]**

The three tables show data redundancies – the reporting system needs to be designed to accommodate at least 10 years of data, and it is unlikely that the current data model will support performance with the anticipated volume.

1. Redesign the data model to become more efficient for querying purposes. by detailing a logical data model showing elements of first and second normal forms where appropriate **[4 marks]**
2. Ensure that the data is clean, describe the steps taken, and if you need to remove some of it, explain why that is the case. **[2 marks]**
3. Answer the following questions, using database queries – and include your SQL statements.
   * 1. How many **different** products do the company sell?
     2. How many products have the company sold in October?
     3. How many customers have bought products from the company in October?
     4. Which customer has spent the most money with the company in October?
     5. How many countries did the company ship products to in October?
     6. Which state in America has the highest number of registered customers?
     7. Which solution area has the most product sku’s within it?
     8. On which day were the most products sold?
     9. Which development group had the most sales in October?
     10. Which industry vertical had the lowest number of sales in October?
     11. Which solution area has the lowest number of sales in the Retail industry vertical?

**[11 marks]**

## SUBMISSION

A pdf file with incorporated narratives, screen grabs, figures, schemes, code, test runs and conclusions delivered via MOODLE. Throughout your work clearly state any assumptions made.

(**25 marks** total for part 2)