

# Individual Assignment Specification

## Contents

|  |   |
|--|---|
| Contents .....   | 1 |
| 1. Purpose.....  | 2 |
| 2. Tasks .....   | 2 |
| 3. Submission.....   | 4 |
| 3.1 Weighting.....   | 4 |
| 3.2 Academic honesty and integrity.....                    | 4 |
| 3.3 Submission Guidelines.....                             | 4 |
| 1. Submission 1: Checkpoint Submission (Checkpoint 1)..... | 4 |
| 2. Submission 2: Final Submission .....                    | 5 |
| 3. Guidelines for LLM use.....                             | 6 |
| 3.4 Marking Guidance.....                                  | 7 |

## 1. Purpose

Setting this assignment is intended for you to obtain the course learning outcomes 1 and 3. Attempting the assignment is more valuable than the mark received for completing the assignment.

In attempting and completing this assignment correctly, it is intended that you will improve your:

- Data-based problem-solving skills.  
The answers to most questions will likely require you to tackle the problem bit-by-bit (a form of abstraction or “divide and conquer” problem-solving).
- Ability to write structured query language (SQL) queries to retrieve and manipulate data stored in relational databases.

## 2. Tasks

You are an analyst within the data management team at Worldwide Importers. Your manager wants to understand the Customer Order Patterns and Sales Performance of Worldwide Importers based on the data in the current database. Descriptive analytics is required to understand these details.

You can access the Worldwide Importers database by following the instructions [here](#). Restore the database in SQL Server Management Studio using the instructions in the file.

This [link](#) provides a detailed documentation for the database. As an analyst, you are expected to explore the documentation and understand the database structure, given that the database has multiple schemas, which would be helpful to you in your tasks. The documentation contains descriptions of the table and, most importantly, database diagrams of the different schemas.

a) **[34 marks]** Write SQL statements for the following to perform this descriptive analysis.

1. [4 marks] What is the average order value for each customer who made purchases in 2016? Include the customer's ID, full name, and average order value (rounded to two decimal places) in the results. Order the results appropriately to quickly identify the customers with the highest average spending per order.
2. [4 marks] Which stock groups have generated the highest total sales between January 1, 2014, and December 31, 2016? Include the stock group ID, stock group name, and total sales amount in your results. Order the results suitably to identify the top-performing stock groups.
3. [5 marks] List all suppliers, displaying the total sales amount for their items (if any), and order the suppliers by the total sales amount in descending order, ensuring that suppliers with no sales are shown with a total sales amount of zero.
4. [7 marks] List all delivery methods and usage counts in sales invoices and purchase orders. Return the delivery method ID, delivery method name, and the counts of their usage in both sales and purchasing.
5. [7 marks] Identify which customers purchased the most diverse range of products in 2016, and the total amount they spent. Include the number of unique products each customer has bought, and the total amount spent in the results to demonstrate the diversity of products. Order and filter the result set in a suitable manner to find the top 10 high-value customers.

6. [7 marks] Modify your query from question 5) above to display the details of these purchases for the top 5 high-value customers. Include in your results the customer's ID and full name, the product IDs and names, the number of orders for each product, the total quantity ordered, and the total amount spent on each product.

b) **[16 marks]** For each SQL statement in part a) above, provide the following as a comment.

- [4 marks] Briefly describe your steps, the thought process behind your decisions, and any references you used.
- [4 marks] Briefly describe the challenges or errors you encountered while writing the query.
- [4 marks] Provide any drafts or iterations of the query to explain your challenges and briefly describe the steps you took to resolve those challenges.

**NOTE:** Comments unaccompanied by queries will not receive any marks. Queries unaccompanied by comments will not receive any marks. ALL comments MUST INCLUDE ALL 3 sections for comments indicated above to receive any marks.

c) **[25 marks]** Your manager has also requested that you perform a detailed Exploratory Data Analysis (EDA) on this database based on your findings from Task a). EDA involves analysing the dataset to summarise its main characteristics. It helps discover patterns and spot anomalies. You are required to use SQL to perform the EDA. Your manager has suggested that if you wish, you can involve an LLM (e.g. ChatGPT) as a team member in this phase.

Your EDA can focus on any **ONE** of the following two contexts:

1. Customers
2. Products

### Requirements for EDA

1. [1 mark] Choose a Data Context, either Customers or Products, for your EDA, indicating your justification for selection.
2. Formulate EDA Questions
  - [4 marks] Propose any **4 (FOUR)** questions that could be part of your EDA.
  - [20 marks] Write the SQL queries required to answer your questions comprehensively.
  - Note that answering a specific question might require writing multiple queries and comparing/contrasting result sets. It might also involve a series of queries incrementally driving into more detail within the data.
3. Examples of EDA Questions
  1. What characteristics describe our top-spending customers vs. the low-spending customers?
  2. Can we identify different segments of customers based on their purchasing behaviour, such as preferred delivery methods and payment methods, etc.?
  3. What are the characteristics of products that have the highest sales?
  4. Do sales vary across different demographics?

This task c) will require planning.

To aid in planning, we will have one checkpoint for this assignment. The checkpoint requirements are detailed in the **Submission** section below.

- d) **[12.5 marks]** Write a concise 1000-word summary of your findings from the above EDA. Focus on your result sets from the above analysis and what insights may be obtained from them. Provide a conclusive answer for each of your 4 EDA questions. Include a section informing your manager of how these findings could be helpful for Worldwide Importers in any future decision-making.
- e) **[12.5 marks]** Provide a reflection that evaluates the contributions of your LLM teammate. In the reflection, you will rate the LLMs in tasks such as explaining concepts and code, correcting syntax, and debugging. A full transcript of all interactions with the LLMs (e.g., prompts in ChatGPT) is required as an appendix for the reflection submission. A documentation template for your reflection is included in the **Submission** section below.

### 3. Submission

#### 3.1 Weighting

This assignment is worth 25% of your final grade.

#### 3.2 Academic honesty and integrity

This assignment is an individual assessment. You must complete all the work yourself. Do not submit work that you did not produce. Do not work in a way that could result in parties producing the same or very similar work.

In attempting this assignment, you agree to adhere to all the principles and practices of academic honesty and integrity for the University of Auckland outlined here: <https://www.auckland.ac.nz/en/about/learning-and-teaching/policies-guidelines-and-procedures/academic-integrity-info-for-students.html>. Any form of cheating, plagiarism, assistance in cheating, unfair collaboration, or other behaviour deemed to be academic misconduct will not be tolerated. Academic misconduct will be dealt with according to the University's Student Academic Conduct Statute.

#### 3.3 Submission Guidelines

##### 1. Submission 1: Checkpoint Submission (Checkpoint 1)

This checkpoint submission will be evaluated, and you will receive feedback. You can modify, enhance, or continue with your questions based on feedback.

You will submit ONE key document for the checkpoint—a .pdf file.

Upload your **.pdf file** to Canvas before the deadline for the assignment checkpoint, as stated on Canvas.

This document (.pdf) is up to two A4 pages long.

A proposed structure is below.

## Proposed Structure for Checkpoint 1 Submission document

### Introduction

List your chosen context (Customers or Products).

### Formulated Questions

List and explain the 4 EDA questions you have formulated.

Explain what queries you believe you might have to write to answer your questions.

You do not have to provide any SQL here. This section describes your plan/ logic for answering the questions.

### LLM Usage Decision

Acknowledge your decision on using LLMs and which LLM you will use.

If you decide to use an LLM, explain how you plan to utilise it.

Consider Table 1 below, which provides some guidelines (non-exhaustive) on the proper and improper use of LLMs in this assignment.

## 2. Submission 2: Final Submission

You will be submitting **TWO** key files. A **.sql file** and a **.pdf file**.

The submission requirements for each file are listed below.

- 1) The .sql file:
    - For questions **1- 6** in Task a), comments as required in Task b) and **ALL queries** you write to answer your 4 EDA questions in Task c), type up your answers in a **single** .sql file. Ensure that the file is appropriately commented to clearly show the question numbers and answers for each question for Task a) and all queries in Task c).
    - Include all the SQL code and all comments you write for the SQL within this .sql file.
    - Upload your **.sql file** to Canvas before the deadline for the assignment as stated on Canvas.
    - Keep a record of all intermediate files and work as evidence of your progress.
  - 2) The .pdf file
    - This document is where you report on this assignment. It has two sections.
1. **Assignment Report**
    - Include your answer for Task b.)
    - Include your answer for Task c) the 4 EDA questions.
    - Provide the SQL queries used to answer each of those EDA questions.
    - Provide screenshots of your result sets (partial screenshot is acceptable if there are too many rows in the result set)
    - Include your answer for Task d).
  2. **Conclusive Reflection**
    - Include your answer for Task e).
      - Evaluate the contributions of the LLM teammate.
      - Rate the LLM's overall helpfulness and proficiency in explaining concepts and code, correcting syntax, debugging, etc.
      - Discuss how the LLM contributed to the project and any improvements or changes

made based on the LLM's assistance.

- A full transcript of all interactions with the LLMs (e.g., prompts in ChatGPT) is required as an appendix for submission of this reflection.
- A proposed structure is below.

### Proposed Structure for Conclusive Reflection

#### LLM Contributions

Concisely describe how the LLM was used throughout the project.

Provide specific examples of tasks where the LLM was helpful.

#### Evaluation of LLM

| Criteria                             | Rating [1- poor, 5 – excellent] | Comments   |
|--------------------------------------|---------------------------------|--|
| Helpfulness in arriving at questions | [1-5]                           | Provide detailed comments on how helpful the LLM was.  |
| Proficiency in Explaining Concepts   | [1-5]                           | Comment on the clarity and thoroughness of the explanations provided by the LLM.             |
| Proficiency in Correcting Syntax     | [1-5]                           | Discuss the accuracy and reliability of the LLM in identifying and correcting syntax errors. |
| Proficiency in Debugging             | [1-5]                           | Discuss the LLM's ability to debug complex queries and provide helpful debugging strategies. |

**Table 1: Possible Criteria for Evaluating your LLM Teammate (Feel free to extend)**

#### Strengths and Weaknesses (i.e. how did the LLM help you/ fail you)

List and explain the critical strengths observed when using the LLM.

List and explain the critical weaknesses observed when using the LLM.

#### Reflection on Improvements

Reflect on any changes or improvements made based on the LLM's assistance.

Discuss any additional insights gained from using the LLM.

### 3. Guidelines for LLM use

| Perspective       | Proper Uses  | Improper Uses   |
|-------------------|--|---|
| Critical Thinking | Using LLMs to generate diverse viewpoints or questions to challenge your understanding of the dataset.       | Relying solely on LLMs to form your questions and conclusions without critically evaluating the output. |
|                   | Employing LLMs to explore different perspectives of the dataset and assess the validity of your assumptions. | Using LLMs to circumvent the need for thorough analysis and understanding of the data.                  |

|                                  |   |  |
|----------------------------------|---|--|
|                                  | Seeking explanations and rationales for SQL functions and constructs to deepen your understanding.      |  |
| <b>Disciplinary Knowledge</b>    | Using LLMs to help identify key learning resources and tutorials relevant to SQL and EDA.               | Copying SQL queries generated by LLMs without attempting to understand them.                   |
|                                  | Leveraging LLMs for syntax checks, debugging, and understanding complex queries.                        | Using LLMs to generate entire SQL queries or answers without engaging in the learning process. |
|                                  | Requesting detailed explanations and learning about best practices in SQL.                              |  |
| <b>Ethical Perspective</b>       | Using LLMs to enhance your learning and understanding while acknowledging the assistance received.      | Presenting LLM-generated SQL code, reports, or analysis as your original work.                 |
|                                  | Requesting LLMs to help format, proofread, and edit your report to improve clarity and professionalism. | Using LLMs to paraphrase and disguise others' work as your own.                                |
|                                  |   | Engaging in any form of academic dishonesty by misrepresenting the contributions of LLMs.      |
| <b>Solution Seeking with SQL</b> | Explore different approaches to data analysis by asking LLMs for alternative methods or perspectives.   | Depend on LLMs to complete assignments without engaging in critical thinking and analysis.     |
|                                  | Look up SQL functions, best practices, and documentation.   | Present AI-generated SQL code as your original work.   |
|                                  | Proofread and edit your report for better clarity and presentation.                                     | Present AI-generated reports as your work.   |
|                                  | Convert pseudocode or logic descriptions into SQL code.   | Use AI paraphrasing tools to disguise others' work.  |
|                                  | Use LLM for code explanation, syntax error fixing, debugging, and documentation.                        |  |

**Table 2. Proper vs. Improper Uses of LLMs in the Assignment**

### 3.4 Marking Guidance

All marks are awarded on a five-point scale, generally representing:

| 100% of the marks   | 75%   | 50%  | 25%   | 0%  |
|---|---|--|---|---|
| There are no errors in the answer.<br>The answer reflects a proficient level of competency. | There is a minor error in the answer.<br>The answer reflects a satisfactory | The answer is weak but passable.<br>The answer reflects a basic level of | There are too many errors in the answer to reflect a satisfactory level of competency.<br>Significant corrections are | There is no answer, or the answer shows no competency.<br>Significant corrections are |

|   |                         |             |           |           |
|---|-------------------------|-------------|-----------|-----------|
|   | level of<br>competency. | competency. | required. | required. |
| <p>Each SQL statement (Task a), c)) will be marked considering:</p> <ul style="list-style-type: none"> <li>a. Its technical correctness.</li> <li>b. The correctness of your logic.</li> <li>c. Penalties for unprofessional work or work that is otherwise difficult to mark.</li> </ul> |                         |             |           |           |