BSBXBD402

Assessor Guide – Part 2 of 2

Test big data samples

Assessment 4 of 4

Project

SWINBURNE OPEN EDUCATION

Version 1.1

B5. Generate and store results of validation activity and associated supporting evidence

In this task, you are required to generate and store results of the validation activity and associated supporting evidence according to the organisational procedures.

Instructions:

Refer to the *AUS Retail_ Big data sample testing policy.pdf* > section **5.4 Procedure for generating and storing results of validation activity and associated supporting evidence** which also includes legislative requirements that you should consider.

Tasks:

- Use the recommended testing tools and process to validate the reports and the data quality by performing **Test Run 2** against the test cases formulated before.
- Document the actual results and the final test results (Pass/Fail) using the ***AUS Retail_STM&TestCase_template.xlsx > Validation*** tab, columns (Actual Result 2 and Test Result 2) for **Test Run 2.**

- Copy the new sample data table contents (*Sample of Sales2* and *Sample of Products2*) from PowerBI table view and paste it onto new Excel worksheets and save them in the *'Phase 3 – Output validation'* folder using the following format for each dataset sample.
    - For dataset 1 (transactional) name the new sample dataset as:
        - *AUS Retail Sales_sample2*
    - For dataset 2 (non-transactional) name the new sample dataset as:
        - *AUS Retail Products_sample2*

- Check that you have generated and stored the results of the validation activities and the supporting evidence using the appropriate folder structure as specified in *AUS Retail_ Big data sample testing policy.pdf* > section **6.3 Procedure for generating and storing results of validation activity and associated supporting evidence** which also includes legislative requirements that you should consider. Provide screenshots of the folder structure you've maintained for storing the validation results from Phase 1 to Phase 3 validation stages.

Evidence of performing the tasks:

Provide screenshots in the space given below to demonstrate that you have maintained a folder structure on your local computer to store evidence of the validation activities for Phase 1, Phase 2 and Phase 3 according to the organisational requirements.

In addition to the screenshots, you will be providing below, your assessment submission 'BSBXBD402_Firstname_Lastname' folder must:

- be organised according to the recommended folder structure as outlined in the organisational procedure
- contain the required supporting documents as evidence of performing the validation activities.
- include the following excel template documents with the *Validation* **Test Run 2** results completed.
    - *AUS Retail_STM&TestCase_Dataset1(Sales]_YourNameInitials_ddmmyyyy.xlsx*
    - *AUS Retail_STM&TestCase_Dataset2(Products]_YourNameInitials_ddmmyyyy.xlsx*

Assessor instructions: Students must create the folder structure required to store validation activity evidence at each phase of the big data validation process. Sample screenshots of the folder structure are given below.

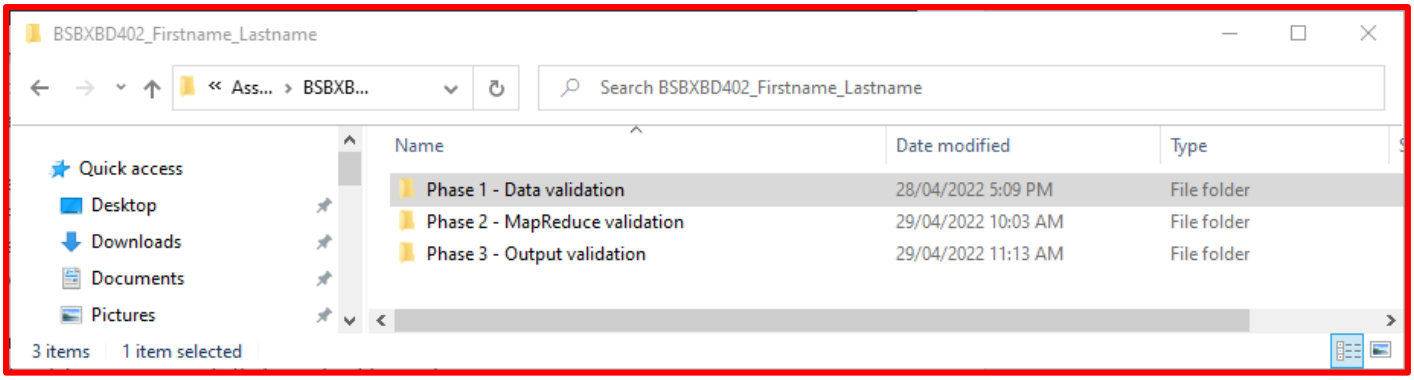


Figure 1 – Screenshot1 of task B5 using File Explorer © Microsoft

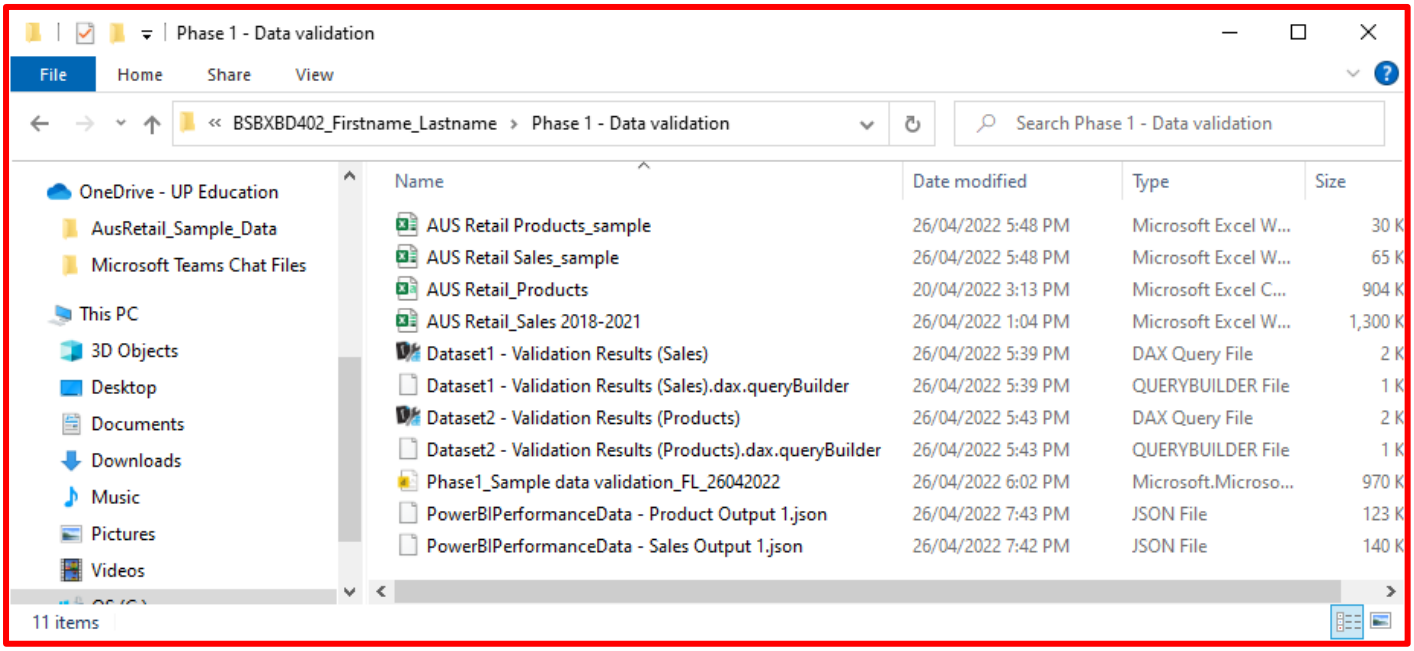


Figure 2 – Screenshot2 of task B5 using File Explorer © Microsoft

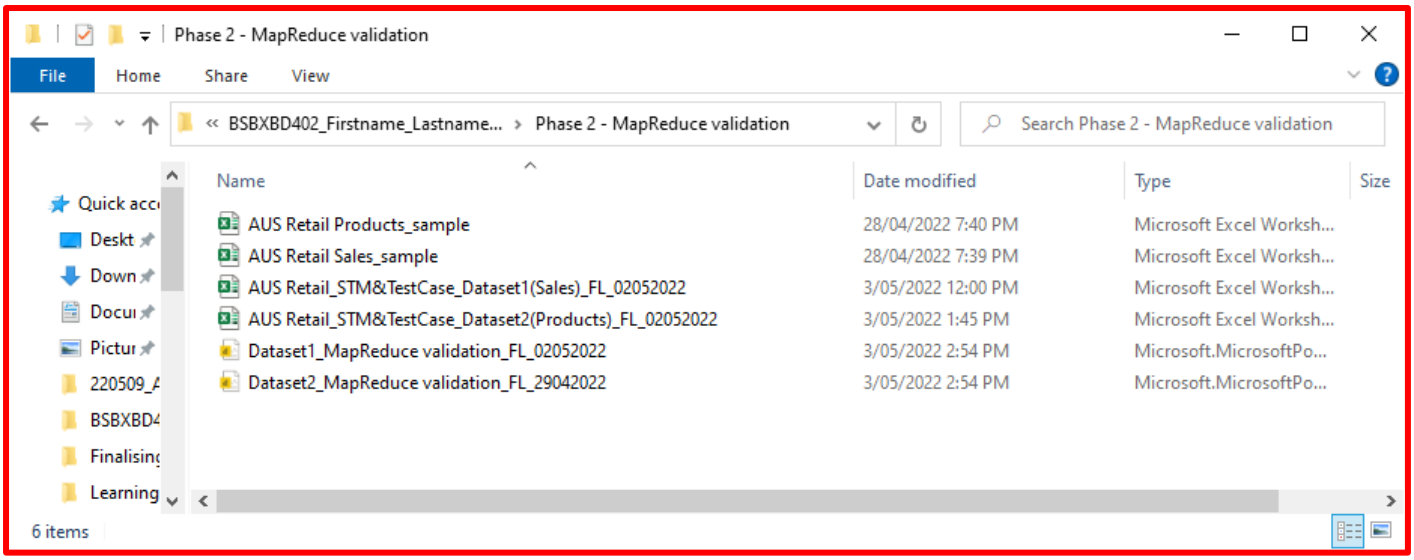


Figure 3 – Screenshot3 of task B5 using File Explorer © Microsoft

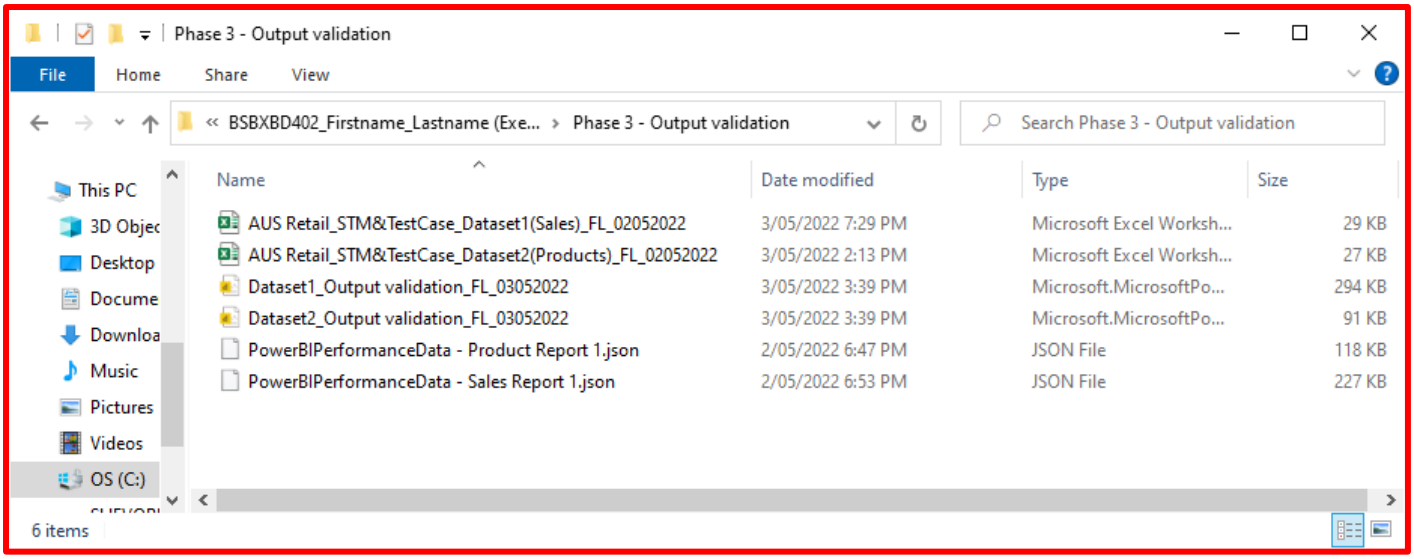


Figure 4 – Screenshot4 of task B5 using File Explorer © Microsoft

Students must use the ***AUS Retail_STM&TestCase_template.xlsx > Validation*** tab to document the actual results of the tests for **Test Run 2** using the excel spreadsheet template provided.

A screenshot of the completed **Test Run 2** information for one dataset is given below. Similarly, students must provide evidence of completing Test Run 2 for the second dataset as well.

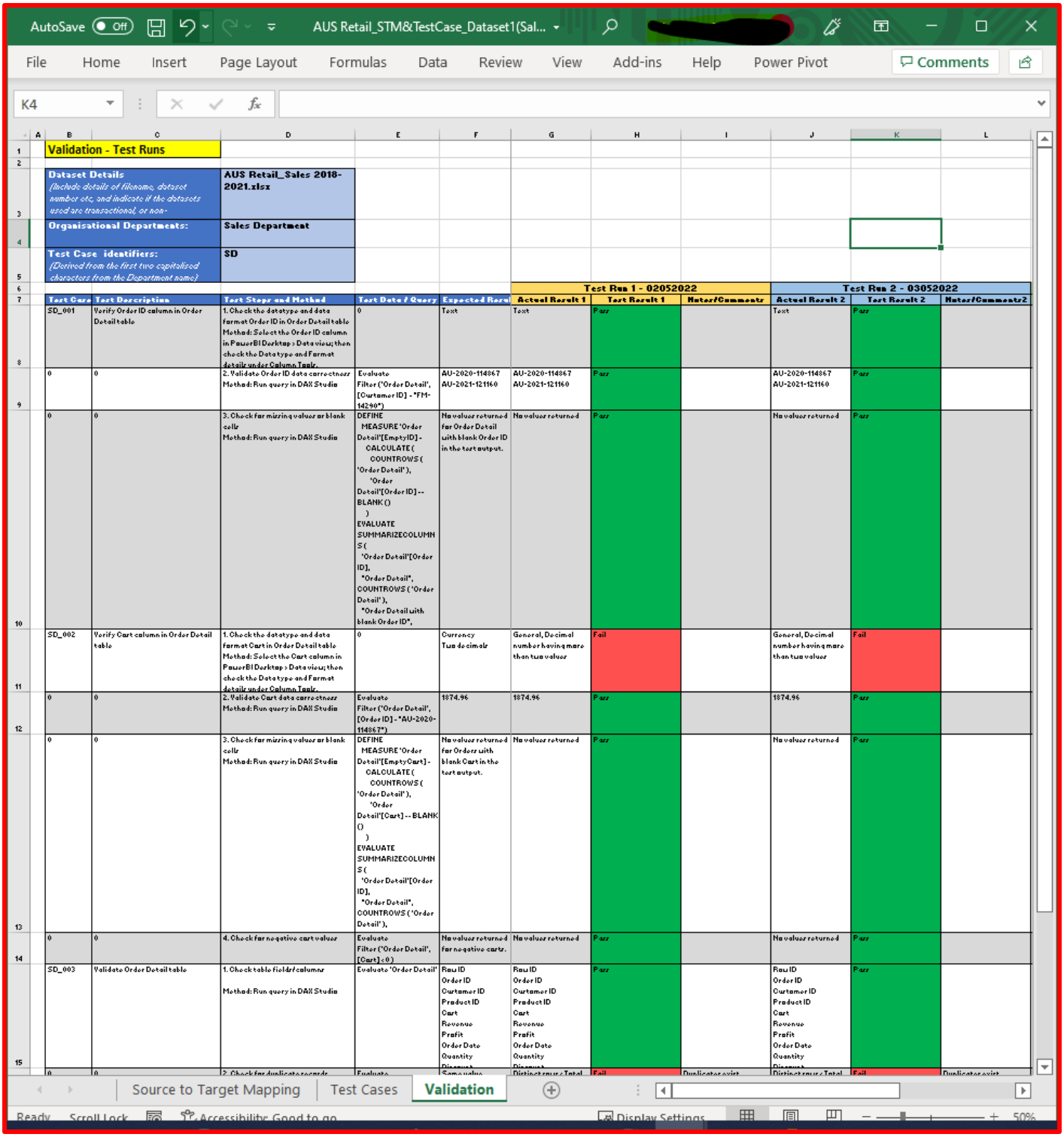


Figure 5 – Screenshot5 of task B5 for using Microsoft Excel © Microsoft

Part C: Optimise big data sample results and documentation

In this part of the assessment, you are required to create a final copy of the AUS Retail reports with optimised results, by performing data cleansing on the validated sample datasets and confirming the absence of issues that were present previously.

As preparation for this task, do the following first.

- Within the '*BSBXBD402_Firstname_Lastname'* folder, create a sub-folder called 'Results Optimised' – all final documents you will be working on should be saved in this folder as the optimised solution.

C1. Perform data cleansing and optimise big data sample results

Instructions:

Save a copy of the PowerBI work files (*Dataset1_Output validation_YourNameInitials_DDMMYYYY* and *Dataset2_Output validation_YourNameInitials_DDMMYYYY*) you have created in the 'Phase 3 – Output validation' folder into the 'Results Optimised' folder and rename it as follows with the current date.

- Dataset1_Results optimised_*YourNameInitials*_DDMMYYYY'
- Dataset2_Results optimised_*YourNameInitials*_DDMMYYYY'

For example, a file saved on the 02nd May 2022 by John Smith for Dataset1 should have the file name: 'Dataset1_Results optimised _JS_02052022'

Refer to the following industry practices and organisational procedures as outlined in the *AUS Retail_ Big data sample testing policy.pdf* document sections:

- **7.1 Perform data cleansing on big data sample**
- **7.2 Perform optimisations on big data sample**

Tasks:

Perform data cleansing on big data sample according to industry practices and organisational procedures by doing the following:

1. Plan for the task: List the details of the data cleansing tasks that need to be carried out in a logical order. Use the answer tables *Table 6* and *Table 7* to document your plan for cleansing each dataset.
2. Carry out the data cleansing tasks according to the plan and provide screenshots in the answer tables *Table 6* and *Table 7* that show the applied steps and modified queries.
3. Perform the following optimisations on the big data samples and provide the necessary screenshots in Table 8.
   - The data model should be optimised to establish appropriate relationships between the dimension tables and the fact tables in both datasets.
   - A *DateTable* should be created to optimise the report for time intelligence for the transactional dataset related to sales data.

Evidence of planning and performing the tasks:

Assessor Instructions: Refer to the sample screenshots provided in the answer table below.

Table 1 – Data cleansing plan for Dataset1 (Transactional)

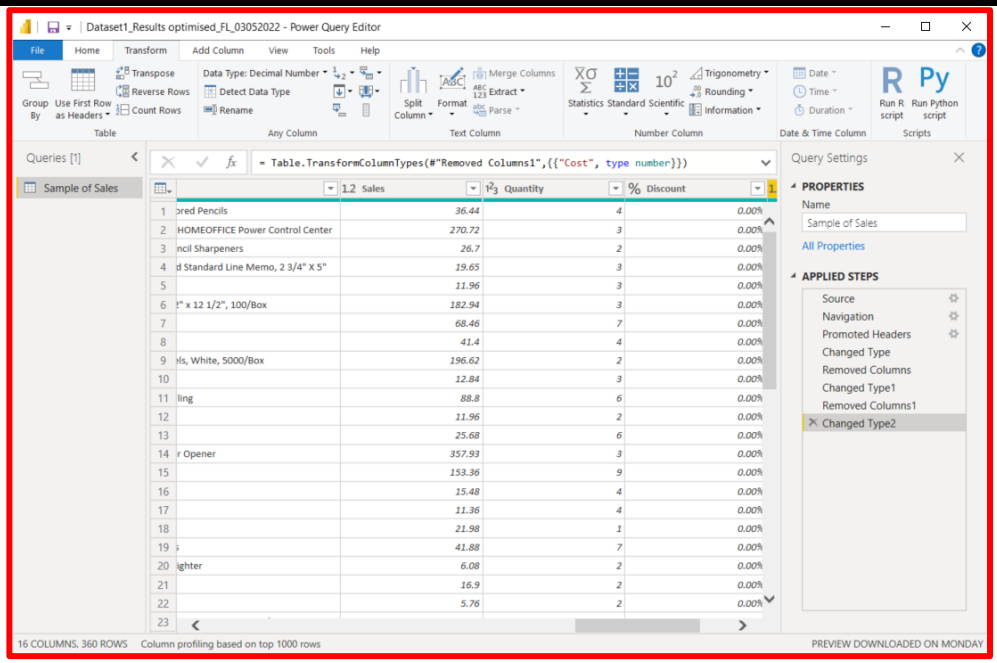
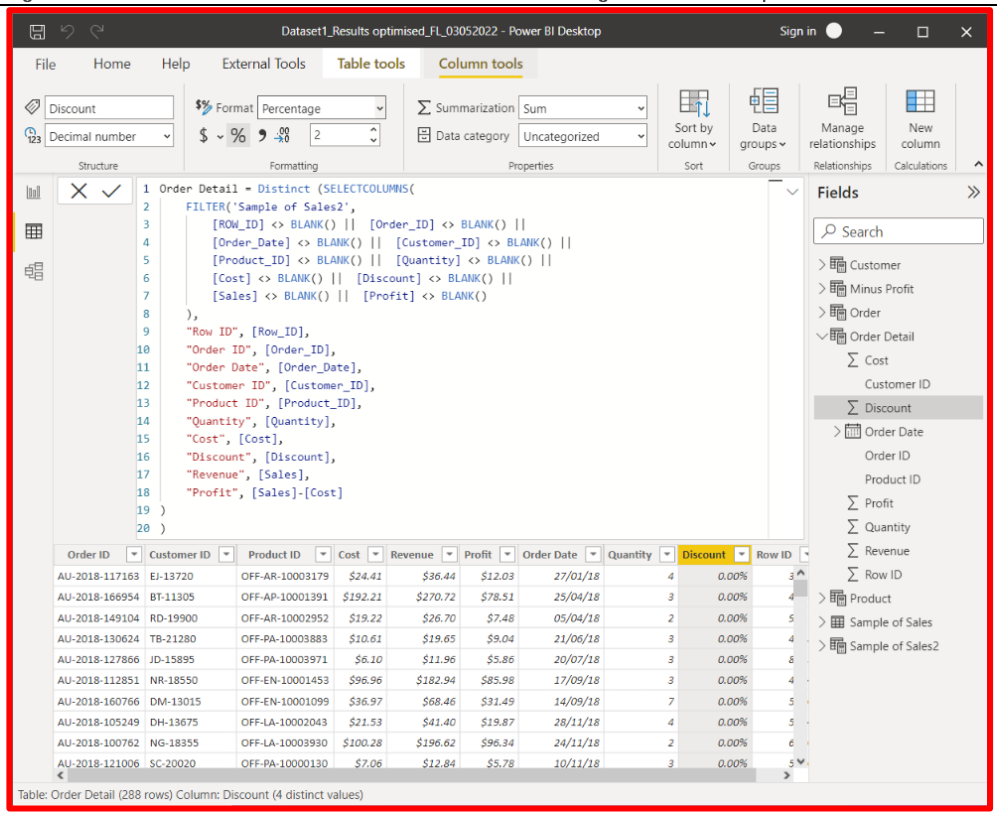
Table name:	Data cleansing task details:	Screenshot showing applied steps or modified queries
Sample of Sales (Source table)	Remove unwanted columns (Ship_Date, Ship_Mode)	 Figure 6 – Screenshot of task C1 source table for Dataset 1 using PowerBI Desktop © Microsoft
Order Detail (Fact Table)	Change the query to ensure distinct values are returned in the table. Ensure any blank records are filtered out. Ensure data type and format is correctly displayed	 Figure 7 – Screenshot of task C1 Order Detail table for Dataset 1 using PowerBI Desktop © Microsoft

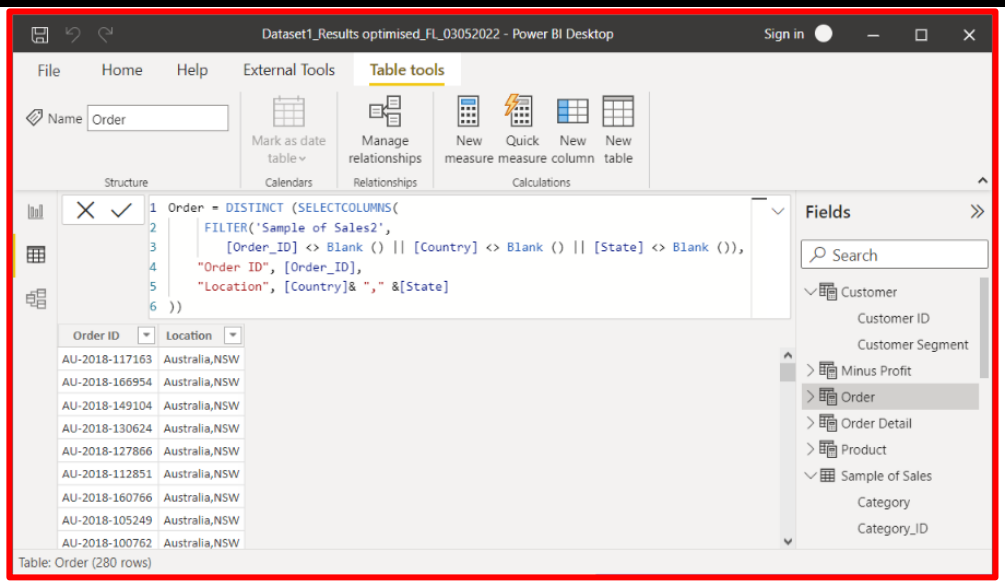
Table name:	Data cleansing task details:	Screenshot showing applied steps or modified queries
Order (Dimension Table)	Change the query to ensure distinct values are returned in the table. Ensure any blank records are filtered out.	

Figure 8 – Screenshot of task C1 Order table for Dataset 1 using PowerBI Desktop © Microsoft

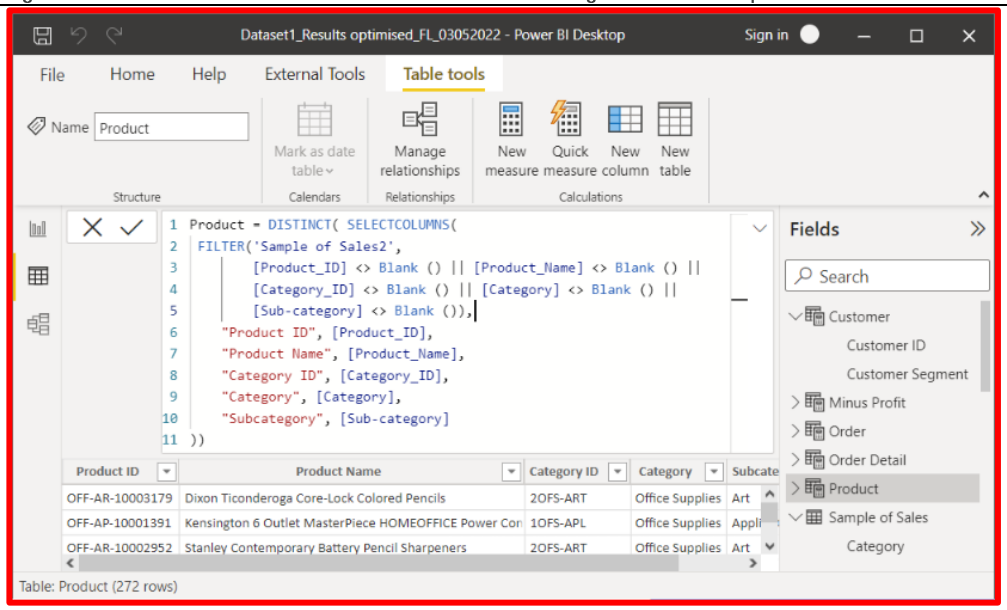
Product (Dimension Table)	Change the query to ensure distinct values are returned in the table. Ensure any blank records are filtered out.	

Figure 9 – Screenshot of task C1 Product table for Dataset 1 using PowerBI Desktop © Microsoft

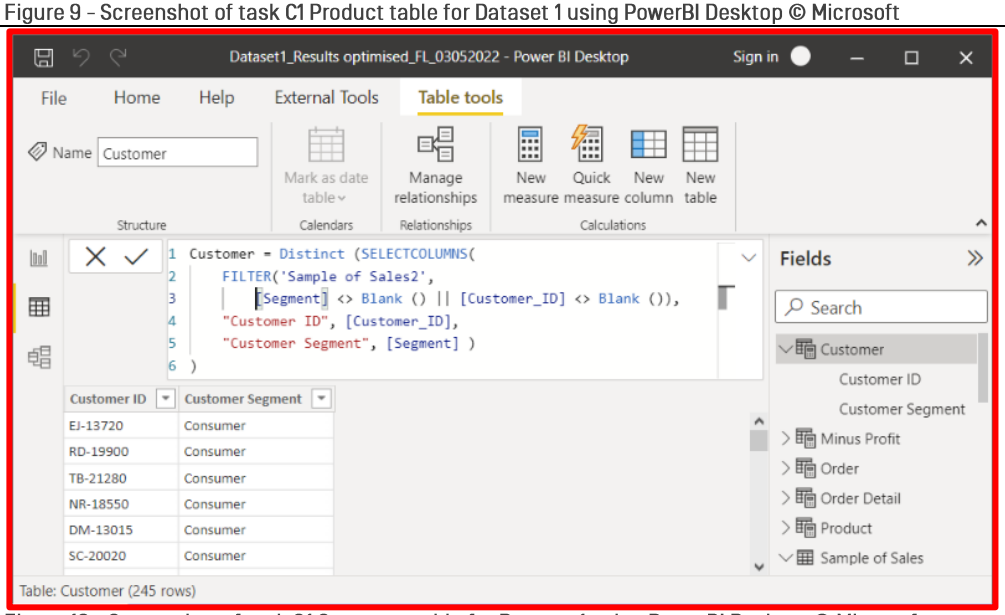
Customer (Dimension Table)	Change the query to ensure distinct values are returned in the table. Ensure any blank records are filtered out.	

Figure 10 – Screenshot of task C1 Customer table for Dataset 1 using PowerBI Desktop © Microsoft

Table 2 – Data cleansing plan for Dataset2 (Non-transactional)

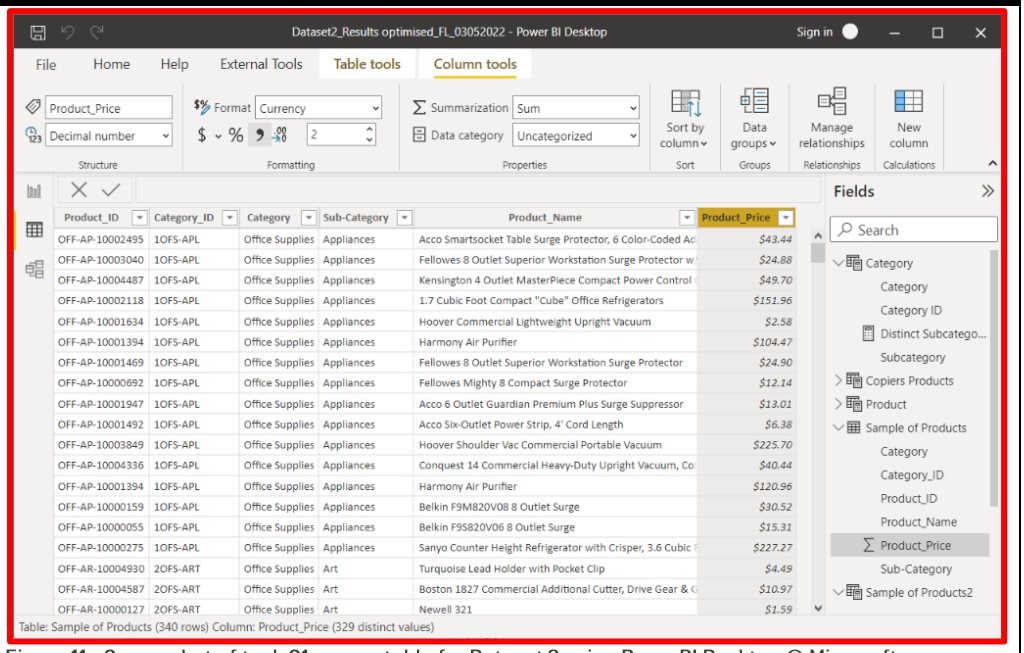
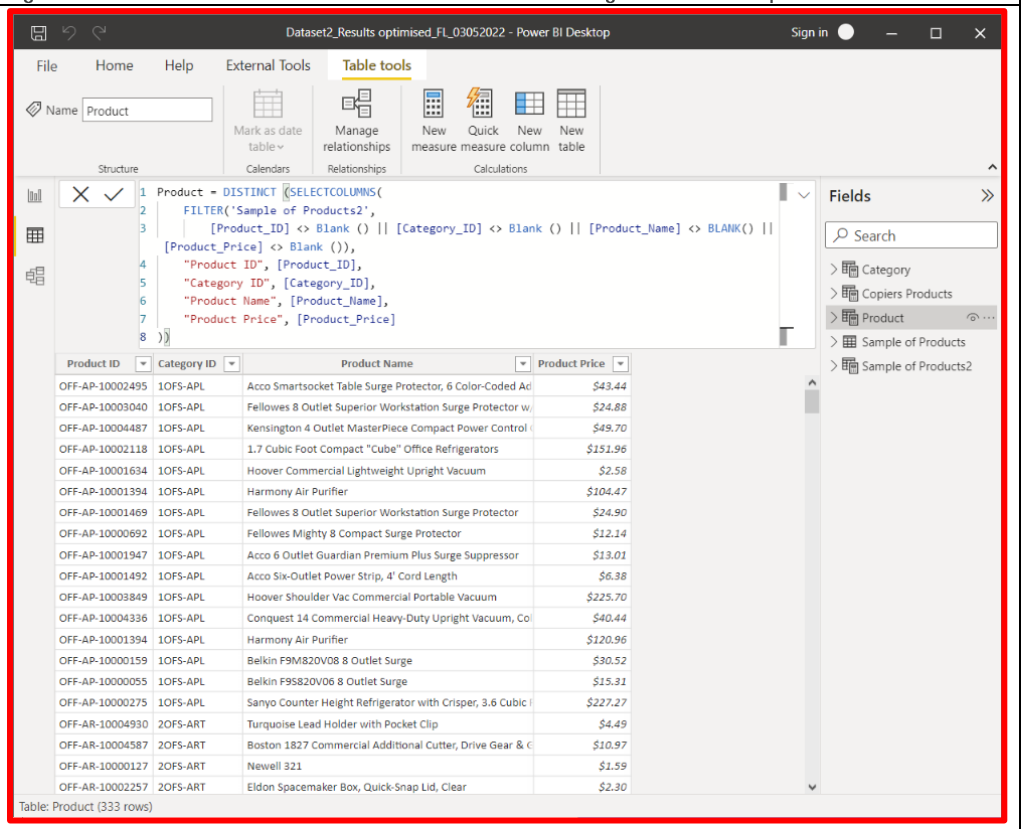
Table name	Data cleansing task details:	Screenshot showing applied steps or modified queries
Sample of Products (Source table)	Correct the data type and format of the Product_Price column to Currency with two decimal values according to the requirement	
Product (Fact Table)	Change the query to ensure distinct values are returned in the table. Ensure any blank records are filtered out.	

Figure 11 – Screenshot of task C1 source table for Dataset 2 using PowerBI Desktop © Microsoft

Figure 12 – Screenshot of task C1 Product table for Dataset 2 using PowerBI Desktop © Microsoft

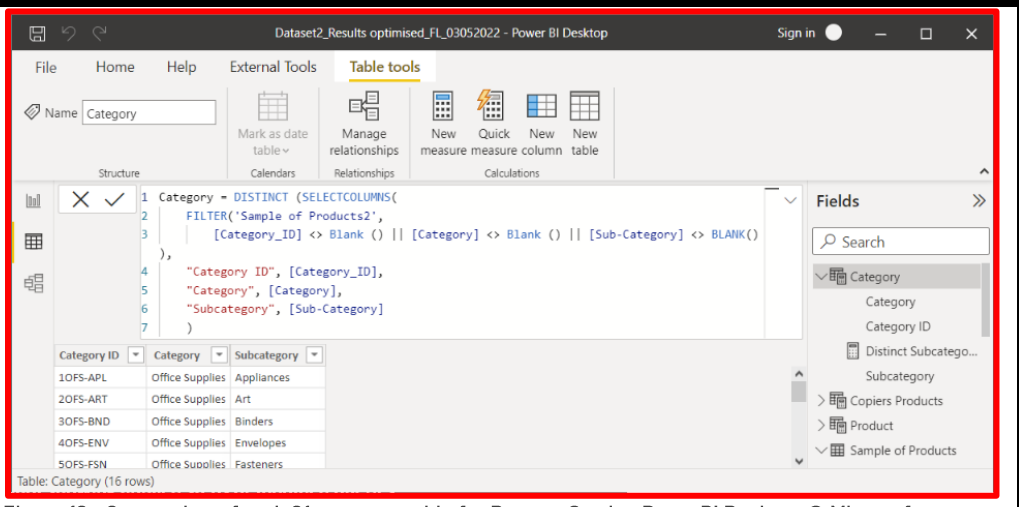
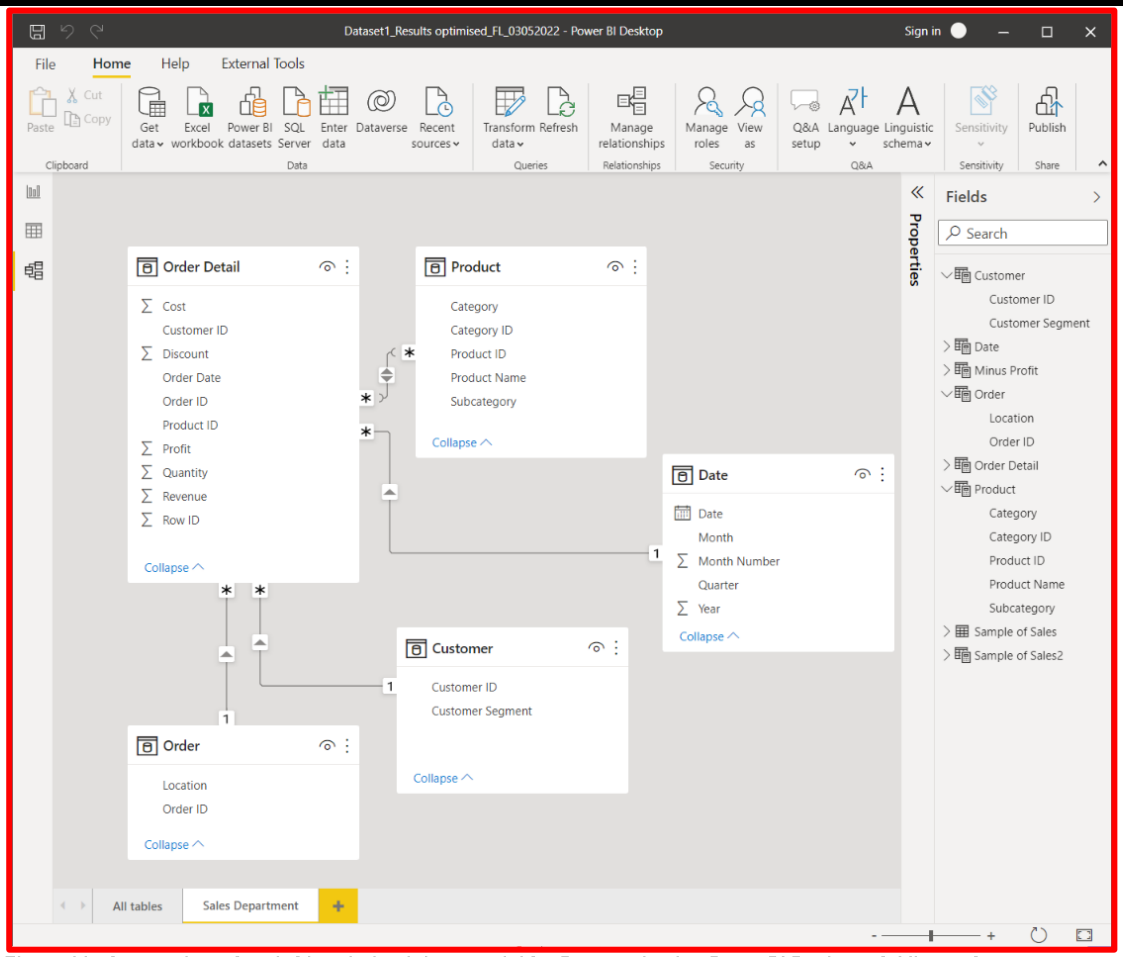
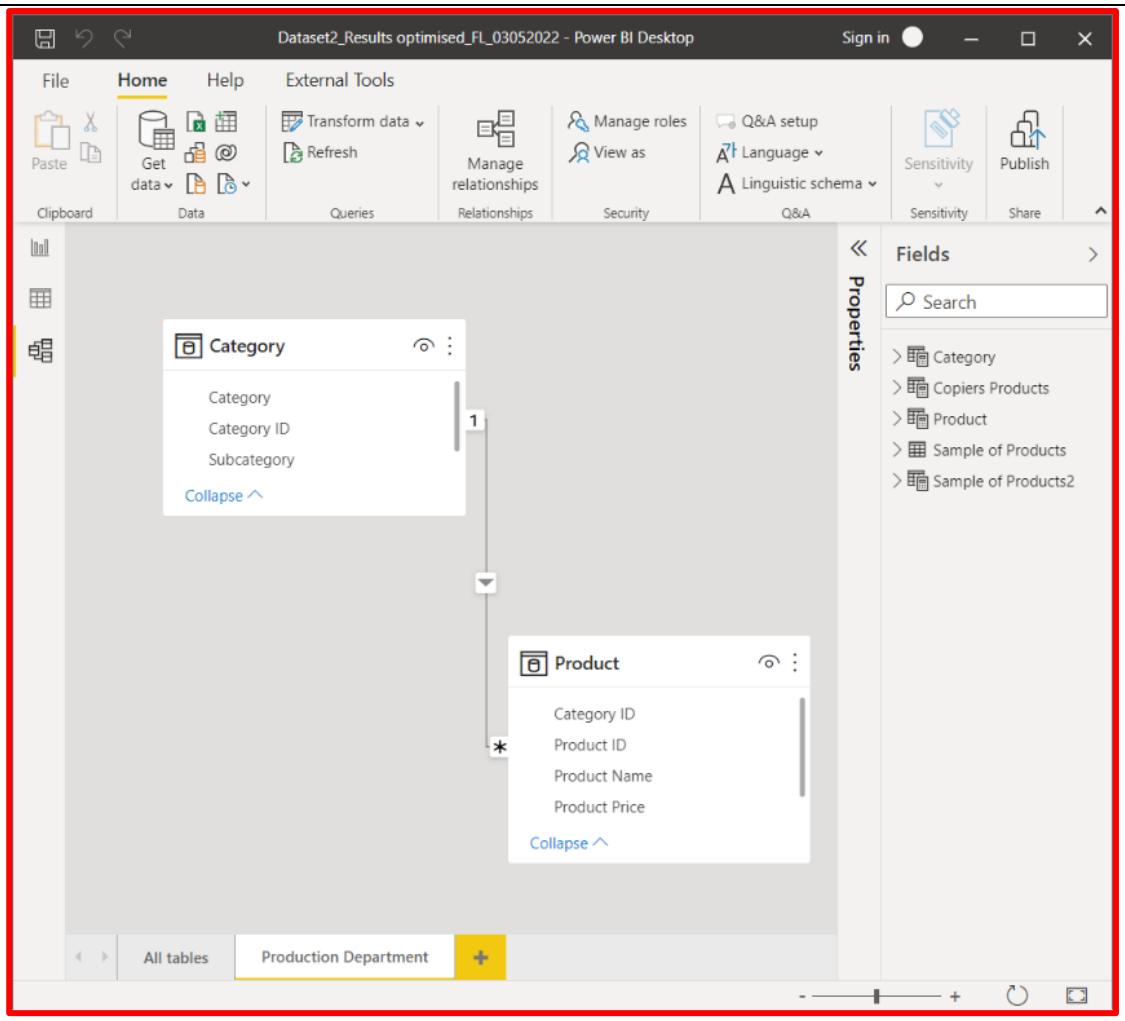
Table name	Data cleansing task details:	Screenshot showing applied steps or modified queries
Category (Dimension Table)	Change the query to ensure distinct values are returned in the table. Ensure any blank records are filtered out. Ensure data type and format is correctly displayed	 Figure 13 - Screenshot of task C1 category table for Dataset 2 using PowerBI Desktop © Microsoft

Table 3 - Evidence of performing optimisations

Optimised dataset model views	Evidence of performing the tasks: *(Screenshots)*
Dataset1 – Sales *(The screenshot should show the model view optimised to have a star schema design with fact table and dimension tables)*	 Figure 14 - Screenshot of task C1 optimised data model for Dataset 1 using PowerBI Desktop © Microsoft **Assessor instructions:** All the dimension tables should ideally form one-to-many relationships with the **Order Detail** fact table. However, there is a possibility of having one

Optimised dataset model views	Evidence of performing the tasks: *(Screenshots)*
	many-to-many relationship between the Product table and the Order Detail table as shown in the screenshot. The student may choose to fix this issue by further breakdown the Product table into a separate Category Table to form a one-to-many relationship with the Fact table.
Dataset2 – Products *(The screenshot should show the model view optimised to have a star schema design with fact table and dimension table)*	 Figure 15 - Screenshot of task C1 optimised data model for Dataset 2 using PowerBI Desktop © Microsoft **Assessor instructions:** The dimention table should ideally form a one-to-many relationship with the **Product** fact table.

C2. Collate validated output testing

Conduct performance testing on the final report output and record results.

Instructions:

As preparation for this task, do the following first.

Save a copy of the *AUS Retail_STM&TestCase_template.xlsx* in the 'Phase 3 – Output validation' folder for each dataset and rename the files as follows:

- *AUS Retail_STM&TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx*
- *AUS Retail_STM&TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx*

E.g. A file saved on the 03rd May 2022 by John Smith for Dataset 1 should have a filename as follows:

- 'AUS Retail_STM&TestCase_Dataset1(Sales)_JS_03052022.xlsx'

Task:

- Use the recommended testing tools and process to validate the reports and the data quality by performing **Test Run 3** against the test cases formulated before.
- Document the actual results and the final test results (Pass/Fail) using the ***AUS Retail_STM&TestCase_template.xlsx > Validation*** tab, columns (Actual Result 3 and Test Result 3) for **Test Run 3.**
  - Carry out performance tests on the final reports after optimising the PowerBI work files for each dataset. Save the performance test JSON files as:
    - *PowerBIPerformanceData – Sales Report 2*
    - *PowerBIPerformanceData – Product Report 2*

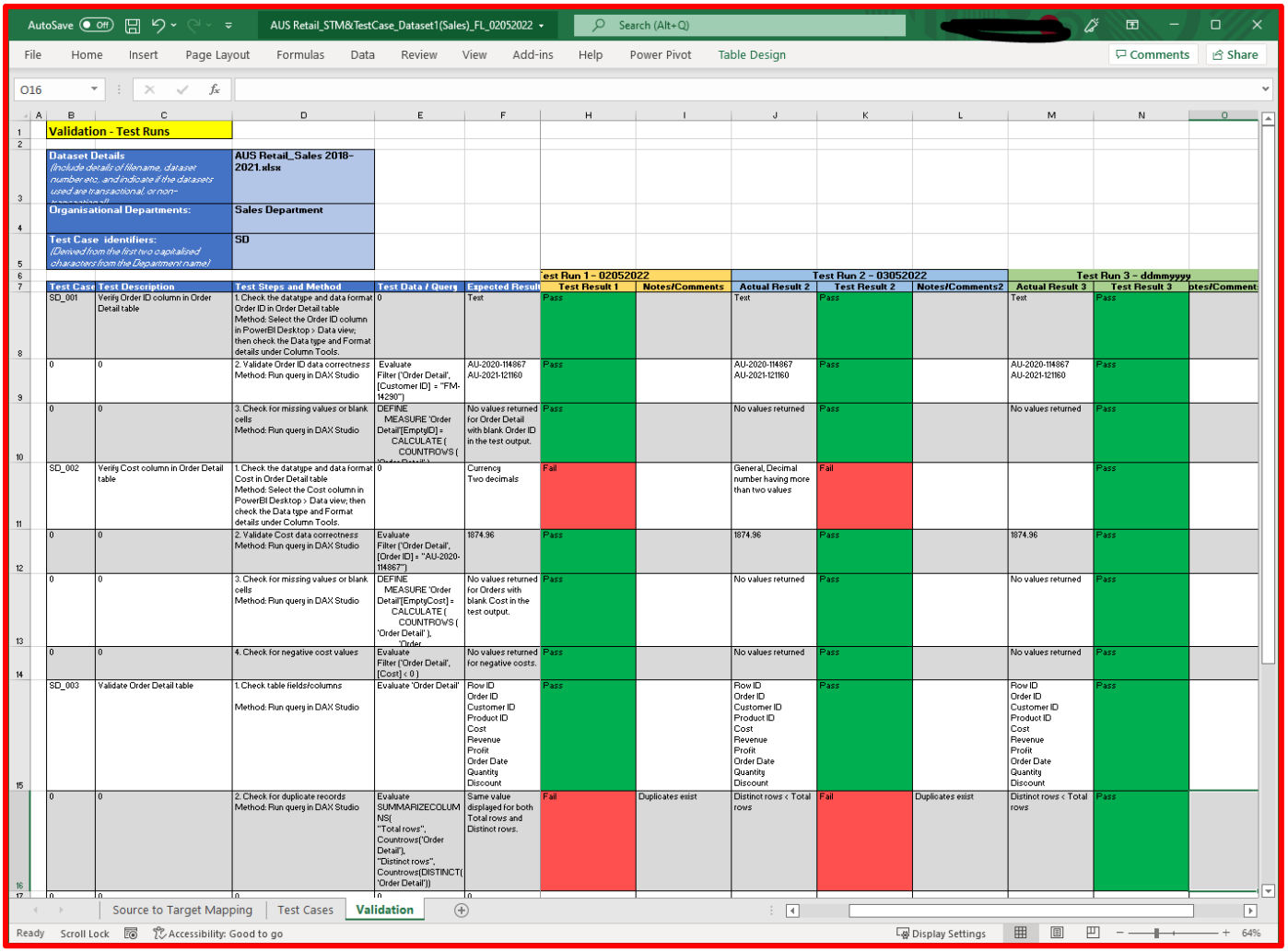
Evidence of performing the task:

Your assessment submission 'BSBXBD402_Firstname_Lastname' folder must:

- contain the required supporting documents as evidence of performing the validation activities (performance result JSON files for each dataset).
  - *PowerBIPerformanceData – Sales Report 2*
  - *PowerBIPerformanceData – Product Report 2*
- include the following excel template documents with the *Validation* **Test Run 3** results completed.
  - *AUS Retail_STM&TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx*
  - *AUS Retail_STM&TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx*

Assessor guidelines: Students must use the ***AUS Retail_STM&TestCase_template.xlsx > Validation*** tab to document the actual results of the tests for **Test Run 3** using the excel spreadsheet template provided. Ideally, at this stage all test cases should achieve a **Pass** result.

A screenshot of the completed **Test Run 3** information for dataset 1 (sales) is given below. Similarly, students must provide evidence of completing **Test Run 3** for dataset 2 (products) as well.



Validation - Test Runs														
Dataset Details (Include details of filename, dataset number etc, and indicate if the datasets used are transactional, or non-transactional)	AUS Retail_Sales 2018-2021.xlsx													
Organisational Departments:	Sales Department													
Test Case identifiers: (Derived from the first two capitalised characters from the Department name)	SD													
						Test Run 1 - 02052022		Test Run 2 - 03052022			Test Run 3 - ddmmyyyy			
Test Case	Test Description	Test Steps and Method	Test Data / Query	Expected Result		Test Result 1	Notes/Comments	Actual Result 2	Test Result 2	Notes/Comments2	Actual Result 3	Test Result 3	otes/Comment	
SD_001	Verify Order ID column in Order Detail table	1. Check the datatype and data format Order ID in Order Detail table Method: Select the Order ID column in PowerBI Desktop > Data view; then check the Data type and Format details under Column Tools.	0	Text		Pass		Text	Pass		Text	Pass		
0	0	2. Validate Order ID data correctness Method: Run query in DAX Studio	Evaluate Filter ('Order Detail', [Customer ID] = "FM-14290")	AU-2020-114867 AU-2021-121160		Pass		AU-2020-114867 AU-2021-121160	Pass		AU-2020-114867 AU-2021-121160	Pass		
0	0	3. Check for missing values or blank cells Method: Run query in DAX Studio	DEFINE MEASURE 'Order Detail'[EmptyID] = CALCULATE (COUNTROWS (	No values returned for Order Detail with blank Order ID in the test output.		Pass		No values returned	Pass		No values returned	Pass		
SD_002	Verify Cost column in Order Detail table	1. Check the datatype and data format Cost in Order Detail table Method: Select the Cost column in PowerBI Desktop > Data view; then check the Data type and Format details under Column Tools.	0	Currency Two decimals		Fail		General_Decimal number having more than two values	Fail			Pass		
0	0	2. Validate Cost data correctness Method: Run query in DAX Studio	Evaluate Filter ('Order Detail', [Order ID] = "AU-2020-114867")	1874.96		Pass		1874.96	Pass		1874.96	Pass		
0	0	3. Check for missing values or blank cells Method: Run query in DAX Studio	DEFINE MEASURE 'Order Detail'[EmptyCost] = CALCULATE (COUNTROWS ('Order Detail'), 'Order	No values returned for Orders with blank Cost in the test output.		Pass		No values returned	Pass		No values returned	Pass		
0	0	4. Check for negative cost values	Evaluate Filter ('Order Detail', [Cost] < 0)	No values returned for negative costs.		Pass		No values returned	Pass		No values returned	Pass		
SD_003	Validate Order Detail table	1. Check table fields/columns Method: Run query in DAX Studio	Evaluate 'Order Detail'	Row ID Order ID Customer ID Product ID Cost Revenue Profit Order Date Quantity Discount		Pass		Row ID Order ID Customer ID Product ID Cost Revenue Profit Order Date Quantity Discount	Pass		Row ID Order ID Customer ID Product ID Cost Revenue Profit Order Date Quantity Discount	Pass		
0	0	2. Check for duplicate records Method: Run query in DAX Studio	Evaluate SUMMARIZECOLUMNS("Total rows", Countrows('Order Detail'), "Distinct rows", Countrows(DISTINCT('Order Detail'))	Same value displayed for both Total rows and Distinct rows.		Fail	Duplicates exist	Distinct rows < Total rows	Fail	Duplicates exist	Distinct rows < Total rows	Pass		

Figure 16 - Screenshot of task C2 Test Run 3 results using Microsoft Excel © Microsoft

C3. Recommend configuration optimisation changes

Based on the performance testing results from the previous task, what additional recommendations can you provide regarding the configuration optimisation changes for AUS Retail's big data sample testing?

Outline your recommendations using the space given below.

(Word count: 55-75 words)

Assessor instructions: Students must provide recommendations on configuration optimisation changes according to their evaluation of the current situation and performance results obtained in the previous tasks.

A sample answer is provided below.

- Memory optimisations by choosing different visualisation options
- Certain DAX functions that are memory intensive can be optimised for efficiency using various query optimization techniques.
- Not having multi-directional or cross-directional filters unnecessarily.
- Disable unwanted options, such as automatic Date hierarchy option.

Part D: Roleplay – Big data testing feedback

To complete this part of the assessment, you are required to communicate the resolved anomalies in task A1 and the final big data sample test results from **Part C** of this assessment to your supervisor.

Roleplay instructions:

The role-play/meeting must include at least two (2) participants, must not exceed five (5) minutes in duration and must address all elements of the **Observation Checklist** below.

In this task, you will participate in a role-play meeting with your supervisor. The supervisor may be resourced using one of the following options:

1. Peer/s who you are already working with, in the industry your qualification relates to.
2. Fellow student/s who will play the role of a team member. Please contact your fellow student/s via the Discussion Forum and coordinate your role play with them directly.

If you are unable to find a participant/s to play the role of the other team member/s, contact your assessor via the Discussion Forum who will discuss options for pairing up with another student/s to complete this task.

- Option 1: Peer/s participant
  - Should you complete this task with your Peer/s, you must fully brief all participant/s, providing them with the context to the role play/meeting, a role outline to play and a copy of the observation checklist so that they can prepare for the recording.
  - Peer/s will need to state their name and job title at the start of the recording to inform consent.

- Option 2: Fellow student/s participant
  - Fellow student/s participating in the recording must be provided with context to their role and responsibilities in the session and have reviewed the assessment activity and observation checklist so that they can prepare for the recording.
  - Student/s will need to state their name and that they are a student (as their job title) at the start of the recording to inform consent.
- Recording instructions
  - Your role play must be recorded with all participant/s captured in a virtual room using a system such as Zoom, Skype or Teams.
  - Consent to participate in the recording must be captured for all participant/s at the start of the meeting. This is achieved by the student reading the following statement at the start of the recording, with all participants replying their name and job title to inform consent.

    "This session/presentation is being recorded for assessment purposes for my course with Swinburne Open Education. This session will be recorded and submitted through my course online learning platform to my Assessor for grading. All participant/s in this session indicate their consent to be included in this recording by stating their name and job title."

  - The time taken to capture consent at the start of the recording does not count towards the recording time limit.
  - Include this recording as part of your assessment submission.

Roles and responsibilities for each party:

Trainee (You):

- presents yourself in a professional manner
- asks appropriate questions (open and closed) to
  - discuss findings of the big data testing results
  - seek clarification on the resolved anomalies

- actively listens to the feedback
- responds appropriately to the given feedback.

Supervisor (Peer or Fellow student participant):

- provides feedback on the final big data sample test results.

Refer to the list of criteria in the Observation Checklist below to understand what you need to demonstrate in this section of the assessment. This Checklist outlines the assessment criteria used to assess your performance.

Observation checklist:

Students must demonstrate the following:

Table 4 - Observation checklist for Part D

OBSERVATION		Assessors are to indicate the result as Satisfactory (S) or Unsatisfactory (US)	
		Satisfactory (S)	Unsatisfactory (US)
1.	The student communicated the final big data sample results **Assessor guidelines:** • The student provides a brief outline of the final results of the test carried out using clear language	☐	☐
2.	The student clarified the resolved anomalies **Assessor guidelines:** • The student communicates the resolutions applied to solve the anomalies using clear language	☐	☐
3.	The student asked open questions **Assessor guidelines:** • Specifically asked for feedback on the test results • Specifically asked for clarifications on the resolved anomalies	☐	☐
4.	The student asked closed probing questions **Assessor guidelines:** • Repeats or summaries what they have said to confirm it is correct	☐	☐
5.	The student listened actively to the feedback given **Assessor guidelines:** • Eye contact • Note-taking • Non-verbal responses - uses appropriate facial expressions such as nodding and smiling to show they are listening • Verbal clarification (sounds) • Focuses on the words spoken by the other person • Refrains from interrupting the person speaking	☐	☐
COMMENTS (Assessor to add comments related to the achievement of the observation requirements above)			

Assessment checklist:

Students must have completed all activities within this assessment before submitting. This includes:

Part A: Resolve anomalies and test performance		
A1	Table 1 – Evidence of performing demonstration task A1	☐
A2	Table 2 – Performance testing of Dataset 1 (Transactional) Table 3 - Performance testing of Dataset 2 (Non-transactional)	☐
Part B: Validate output of captured big data sample and record results		
B1	Excel templates, *Test Cases* tab: • *AUS Retail_STM&TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx* • *AUS Retail_STM&TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx*	☐
B2	Excel templates, *Validation* tab – **Test Run 1**: • *AUS Retail_STM&TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx* • *AUS Retail_STM&TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx*	☐
B3	Table 4 – Evidence of performing demonstration task B3	☐
B4	Table 5 – Evidence of performing demonstration task B4	☐
B5	BSBXBD402_Firstname_Lastname' folder with required sub-folders and completed templates Excel templates, *Validation* tab – **Test Run 2**: • *AUS Retail_STM&TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx* • *AUS Retail_STM&TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx*	☐
Part C: Optimise big data sample results and documentation		
C1	Table 6 – Data cleansing plan for Dataset 1 (Transactional) Table 7– Data cleansing plan for Dataset 2 (Non-transactional) Table 8 – Evidence of performing optimisations	☐
C2	Excel templates, *Validation* tab – **Test Run 3**: • *AUS Retail_STM&TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx* • *AUS Retail_STM&TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx*	☐
C3	Provide a written answer in the space provided.	☐
Part D: Roleplay – Big data testing feedback		
D1	Communicate final big data sample test results to Supervisor Table 9 - Observation Checklist for roleplay demonstration	☐

Congratulations you have reached the end of Assessment [4]!

References:

Learning Container. 2020. *Sample sales data excel xls.* [online] Available at: https://www.learningcontainer.com/download/sample-sales-data-excel-xls/ [Accessed 04 April 2022].