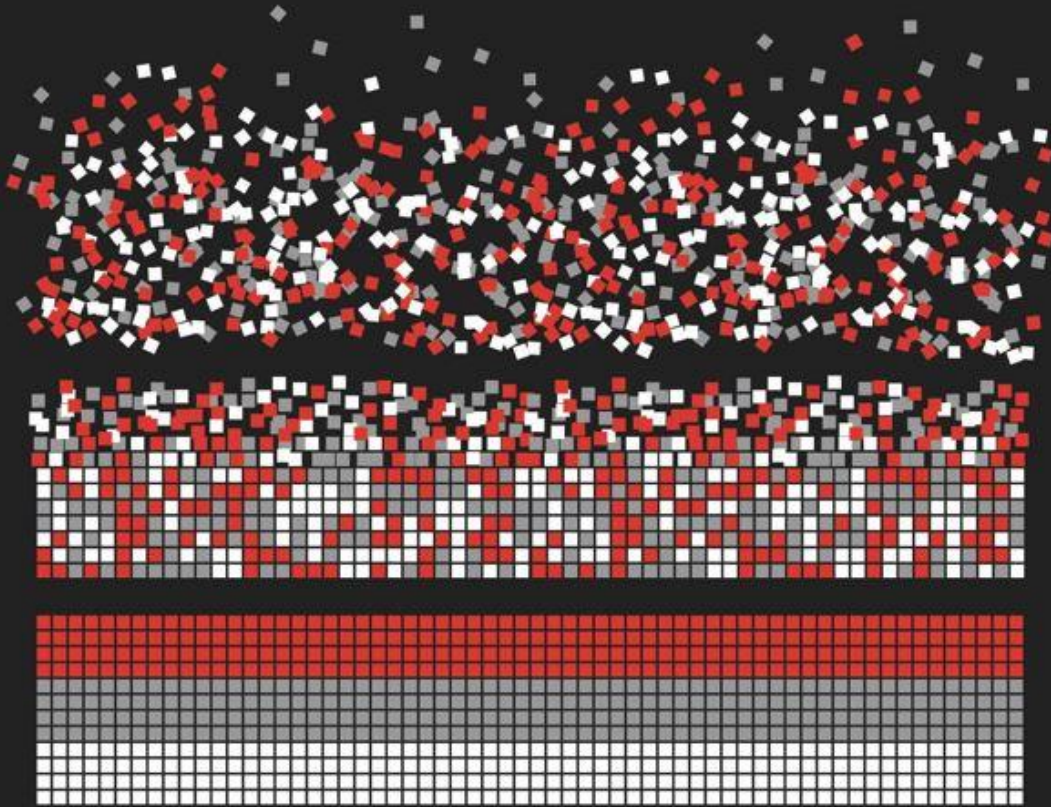


# BIG DATA



BSBXBD402

Assessor Guide – Part 1 of 2

## Test big data samples

Assessment 4 of 4

Project



## Assessment Instructions

### Task overview

This assessment task is divided into four (4) parts having a total of six (6) demonstration activities. Read each question carefully before typing your response in the space provided.

To complete this assessment, you will need the following:

### Information and telecommunications equipment

- A computer installed with the Windows operating system.
- Microsoft Power BI Desktop App - Download and install the free **Power BI Desktop** App from Microsoft Store: [Downloads | Microsoft Power BI](https://powerbi.microsoft.com/en-au/downloads/) (Long URL: <https://powerbi.microsoft.com/en-au/downloads/>)
- Latest version of DAX Studio – An external tool that can be used for running queries and test scripts for Power BI – Download and install the free **DAX Studio** App from [Downloads | DAX Studio.org](https://DAX Studio.org/downloads/) (Long URL: <https://DAX Studio.org/downloads/>)

### Additional resources and supporting documents

Assessment supporting documents (zipped folder) - This folder contains the following sub-folders, documents and templates required for reference and use when performing the tasks in this assessment.

- AUS Retail\_Raw datasets (folder)
  - AUS Retail\_Products (.csv)
  - AUS Retail\_Sales 2018-2021 (.xlsx)
- AUS Retail\_ Data flow and dataset schemas.pdf
- AUS Retail\_Big data sample testing policy.pdf
- AUS Retail\_Reporting requirements.pdf
- AUS Retail\_STM&TestCase\_template.xlsx

**Important note:** The above resources were already provided to you as part of the Assessment 3 Project.

### Work files from the previous Assessment 3 Project

As this project assessment is a continuation of the work tasks from the Assessment 3 Project, you will need to have access to the following files you have worked on previously.

- Phase 2 – MapReduce validation (folder)
  - AUS Retail\_Sales\_sample (.xlsx)
  - AUS Retail\_Products\_sample (.xlsx)
  - AUS Retail\_STM&TestCases\_Dataset1(Sales)\_NameInitials\_ddmmyyyy (.xlsx)
  - AUS Retail\_STM&TestCases\_Dataset2(Products)\_NameInitials\_ddmmyyyy (.xlsx)
  - Dataset1\_MapReduce validation\_NameInitials\_ddmmyyyy (.pbix)
  - Dataset2\_MapReduce validation\_NameInitials\_ddmmyyyy (.pbix)

For all parts of this assessment, refer to the Project Assessment 3, Part A: Project scenario as the same scenario is applicable to this assessment.

## Assessment Information



### Submission

You are entitled to three (3) attempts to complete this assessment satisfactorily. Incomplete assessments will not be marked and will count as one of your three attempts.

All questions must be responded to correctly to be assessed as satisfactory for this assessment.

Answers must be typed into the space provided and submitted electronically via the LMS. Hand-written assessments will not be accepted unless previously arranged with your assessor.



### Reasonable adjustment

Students may request a reasonable adjustment for assessment tasks.

Reasonable adjustment usually involves varying:

- the processes for conducting the assessment (e.g. allowing additional time)
- the evidence gathering techniques (e.g. oral rather than written questioning, use of a scribe, modifications to equipment)



However, the evidence collected must allow the student to demonstrate all requirements of the unit.

Refer to the Student Handbook or contact your Trainer for further information.



Please consider the environment before printing this assessment.

# Part A: Resolve anomalies and test performance

To complete this part of the assessment, you are required to continue to work on the following PowerBI work files and sample dataset files that you have worked on as part of the previous Assessment 3 Project to perform tasks A1 and A2.

- AUS Retail\_Sales\_sample [.xlsx]
- AUS Retail\_Products\_sample [.xlsx]
- Dataset1\_MapReduce validation\_NameInitials\_ddmmyyyy [.pbix]
- Dataset2\_MapReduce validation\_NameInitials\_ddmmyyyy [.pbix]

## A1. Resolve identified anomalies according to consultation advice

Read the scenario carefully and perform the following tasks.

### Scenario:

You have received consultation advice via the following email response from your supervisor.

**Gonzales, Mia**

**Sent:** Today, Present date and time

**To:** Lastname, Firstname

**Subject:** Resolving identified anomalies

Hi <Firstname Lastname>,

Thank you for informing me of the anomalies detected in the sales and product-related data. I can confirm that some of the anomalies are due to errors made when entering data into the database. It has been escalated to the sales and product departments to update these details in the system. However, for the purpose of carrying out the sample data test, I would recommend that you make the following changes to your sample dataset source files.

If your sales data sample contains any records from the state of Victoria for the following months, that display minus [-] figures for sales/revenue please correct them as positive values.

- 2018 March
- 2019 September
- 2020 May
- 2021 January

If your product data sample contains any duplicate product names having different product IDs and prices, apply the following fix:

- Retain the record for the Product ID, that has the highest product price and remove all other records of that product.

The rest of the records are correct although it is identified as an anomaly in PowerBI.

Thanks and kind regards,

**Mia Gonzales**

*Chief Data Officer (CDO)*



[Mia.gonzales@ausretail.com.au](mailto:Mia.gonzales@ausretail.com.au)

*Before printing this email please consider the environment.*

*This message may contain privileged information or confidential information or both and is intended for the recipient named. If you are not the intended addressee, please delete it and notify the sender.*

## Tasks:

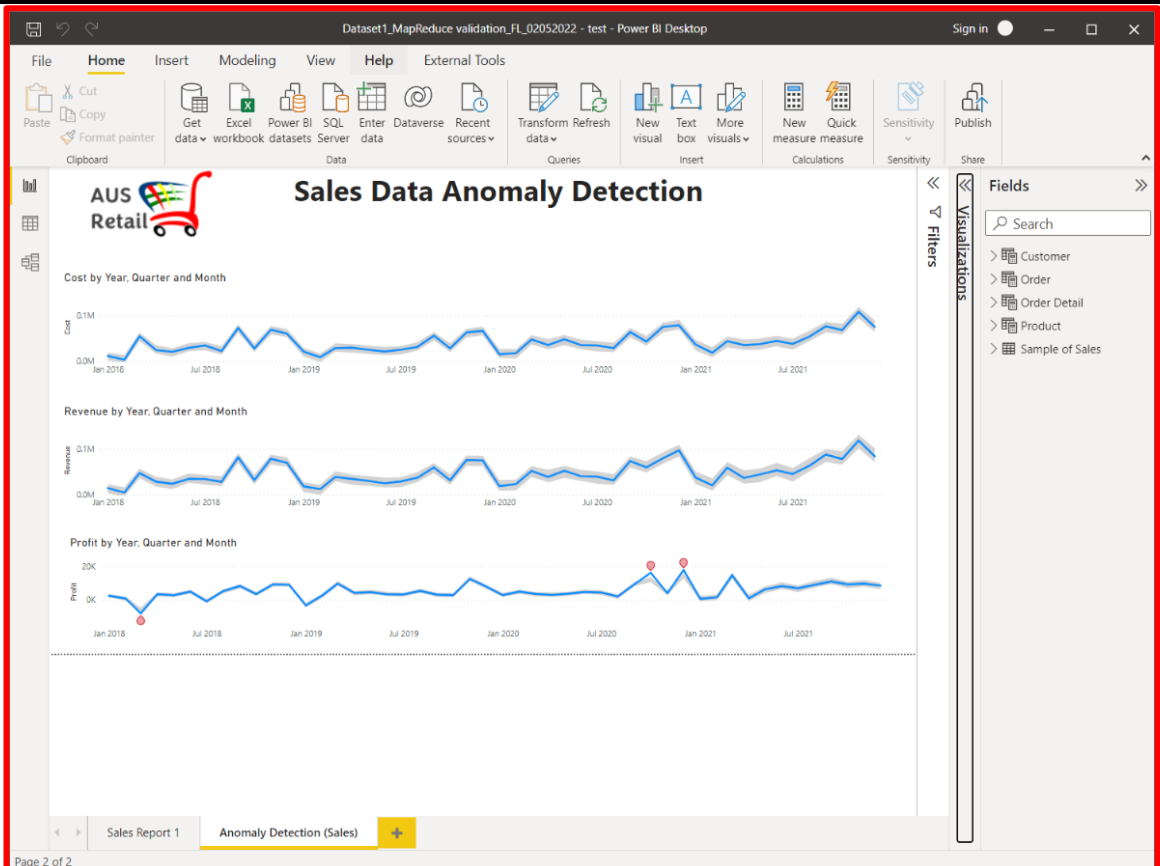
According to the advice received from your supervisor do the following,

1. Resolve the anomalies in the sample dataset source files (AUS Retail Sales\_sample and AUS Retail Products\_sample)
2. Refresh the data in your PowerBI working file and ensure that the anomalies have been resolved.
3. Provide screenshot(s) of the following report views showing the absence of the resolved anomalies:
  - Anomaly Detection [Sales]
  - Anomaly Detection [Products]

## Evidence of performing the task:

**Assessor instructions:** The screenshots provided should clearly indicate the absence of the resolved anomalies.

Table 1 – Evidence of performing demonstration task A1

| Report views:  | Evidence of resolved anomalies<br>(Screenshots)   |
|--|---|
| <p><b>Anomaly Detection [Sales]</b> – showing that the previous anomalies have been resolved.</p> <p><b>Assessor guidelines:</b> The anomalies fixed or still remaining may display differently than the given screenshot due to the variations in the sample dataset chosen by the student.</p> |  <p>Figure 1 – Screenshot of resolved anomalies for Dataset1 using PowerBI Desktop © Microsoft</p> |

## Report views: Evidence of resolved anomalies [Screenshots]

**Anomaly Detection (Products)** – showing that the previous anomalies have been resolved.

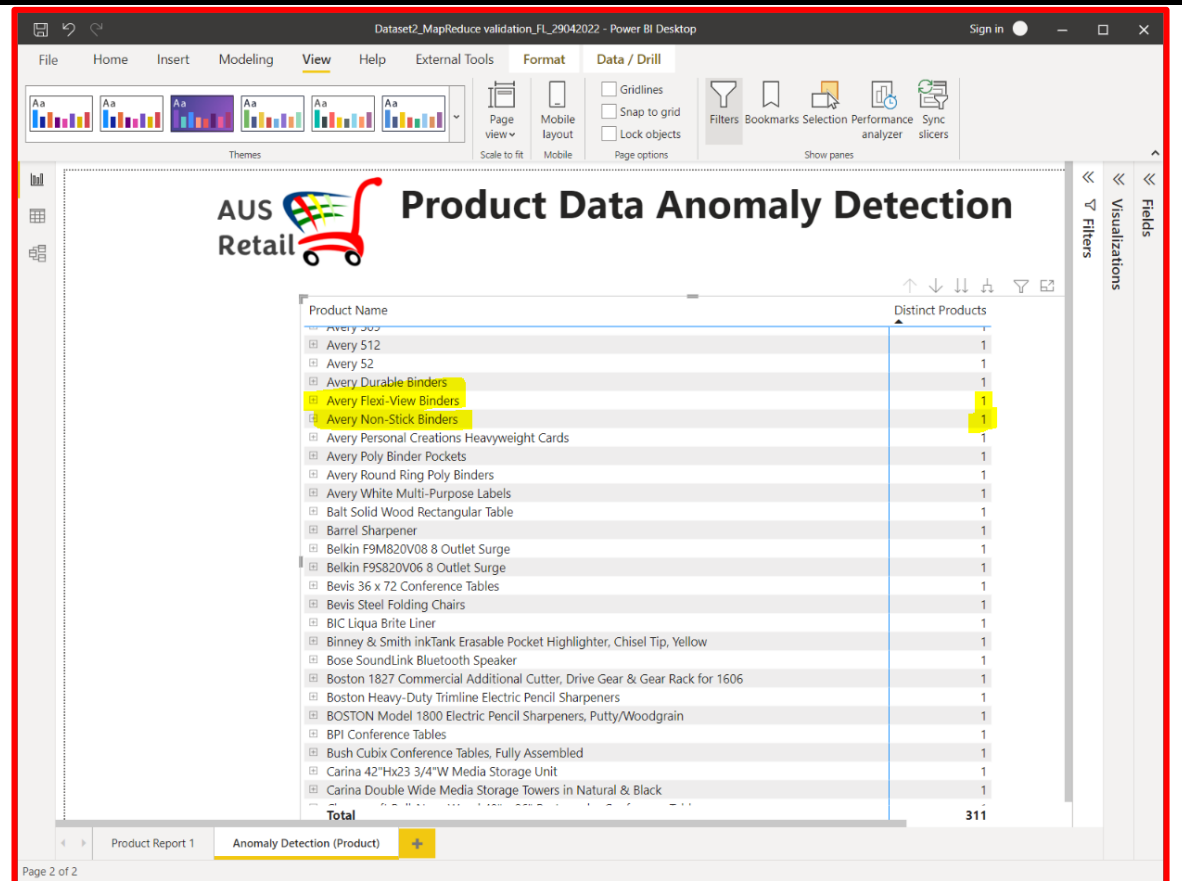


Figure 2 – Screenshot of resolved anomalies for Dataset2 using PowerBI Desktop © Microsoft

## A2. Conduct performance testing

In this task, you are required to conduct performance tests on visualisation outputs from both datasets [transactional and non-transactional] using the in-built *Performance analyser* tool in *PowerBI Desktop* and by using the external tool *DAX Studio*.

### Instructions:

As preparation for this task, do the following first.

- Create a new folder in your local computer called 'BSBXBD402\_Firstname\_Lastname'. – All the files you will be working on in this assessment should be saved in this folder location.
- Within the *BSBXBD402\_Firstname\_Lastname* folder, create the following sub-folders
  - Phase 3 – Output validation
  - Results optimised

### Tasks:

Do the following tasks for both datasets and provide relevant screenshots as evidence of performing the tasks in the answer tables provided:

#### A2.1 Capture data throughput performance

- a. Start recording performance data using the Performance analyser
- b. Clear any previously captured data
- c. Refresh all visuals and capture performance data.
- d. Perform some filters within the data visualisations in the report



- e. Provide a screenshot of the *PowerBI* report view in the answer table [Table 9], clearly showing the data captured by the *Performance analyser*.

## A2.2 Capture data processing (query) performance

- Export the previously recorded data from the Performance analyser into a JSON File, with a meaningful name that relates to the type of data output report tested (e.g. *PowerBIPerformanceData – Sales Report 1, PowerBIPerformanceData – Product Report 1*)  
Note: Save these exported JSON files in the 'Phase 3 – Output validation' folder.
- Open *DAX Studio*
- Load the performance data JSON File into *DAX Studio*.
- Provide a screenshot that clearly shows the data processing/query related performance information captured from the loaded report in *DAX Studio*

## A2.3 Capture sub-component performance

- In *DAX Studio* go to Advanced > View Metrics
- Expand the relevant datasets (tables) for Sales Output view to obtain sub-component performance data.
- Provide a screenshot of the *DAX Studio* window clearly showing all sub-component performance metrics for the relevant tables associated with the data output report]

### Evidence of performing the tasks:

In addition to the screenshots you will include in **Table 2** and **Table 3** given below, your assessment submission must include the following performance test files in the 'Phase 3 – Output validation' sub-folder.

- *PowerBIPerformanceData – Sales Report 1*
- *PowerBIPerformanceData – Product Report 1*

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

Table 2 – Performance testing of Dataset 1 [Transactional]

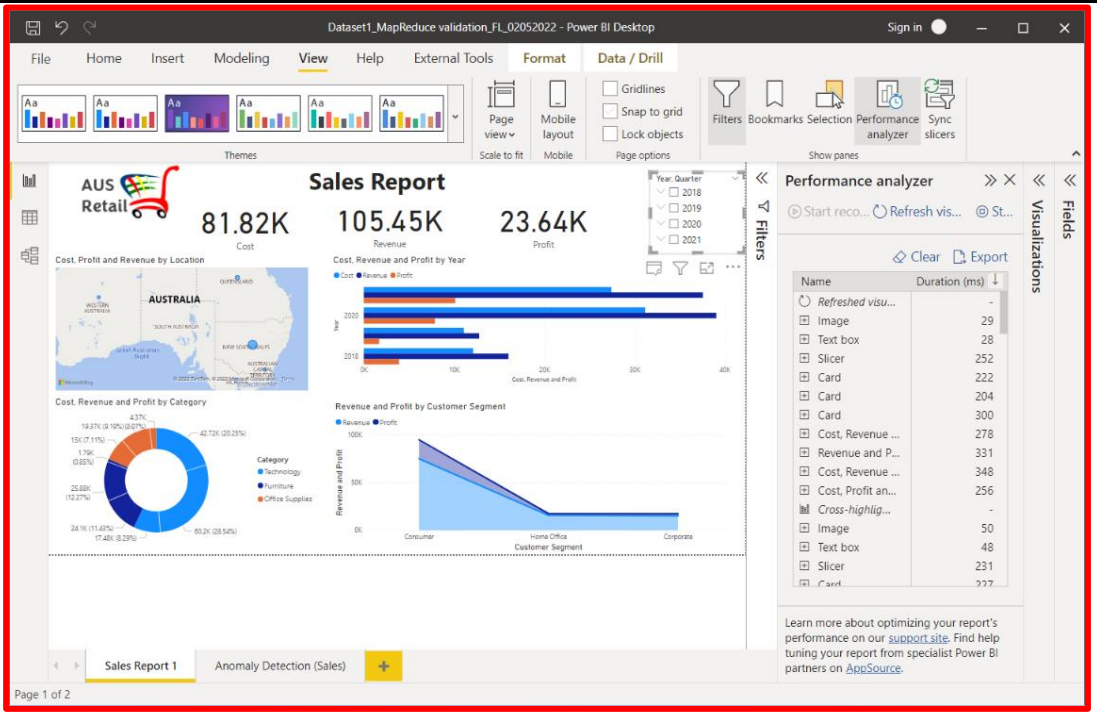
| Performance tests:   | Evidence of test results: [Screenshots]  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
|--|--|------|---------------|-------------------|----|-------|----|----------|-----|--------|-----|------|-----|------|-----|------|-----|-------------------|-----|------------------|-----|-------------------|-----|--------------------|---|------------------|----|-------|----|----------|-----|--------|-----|------|-----|
| <p><b>A2.1 Data throughput</b><br/>(The screenshot should show the data throughput information captured in the Performance analyser column in PowerBI for Sales Output report]</p> |  <p>The screenshot shows the PowerBI Desktop interface for a report titled 'Sales Report'. The main area contains several visualizations: a map of Australia, a bar chart for 'Cost, Revenue and Profit by Year', a donut chart for 'Cost, Revenue and Profit by Category', and a line chart for 'Revenue and Profit by Customer Segment'. On the right side, the 'Performance analyzer' pane is open, showing a table of visualization names and their durations in milliseconds. The table includes columns for 'Name' and 'Duration (ms)'. The 'Refreshed visu...' row is highlighted.</p> <table border="1"> <thead> <tr> <th>Name</th> <th>Duration (ms)</th> </tr> </thead> <tbody> <tr> <td>Refreshed visu...</td> <td>29</td> </tr> <tr> <td>Image</td> <td>28</td> </tr> <tr> <td>Text box</td> <td>252</td> </tr> <tr> <td>Slicer</td> <td>222</td> </tr> <tr> <td>Card</td> <td>204</td> </tr> <tr> <td>Card</td> <td>300</td> </tr> <tr> <td>Card</td> <td>278</td> </tr> <tr> <td>Cost, Revenue ...</td> <td>331</td> </tr> <tr> <td>Revenue and P...</td> <td>348</td> </tr> <tr> <td>Cost, Revenue ...</td> <td>256</td> </tr> <tr> <td>Cost, Profit an...</td> <td>-</td> </tr> <tr> <td>Cross-highlig...</td> <td>50</td> </tr> <tr> <td>Image</td> <td>48</td> </tr> <tr> <td>Text box</td> <td>231</td> </tr> <tr> <td>Slicer</td> <td>231</td> </tr> <tr> <td>Card</td> <td>227</td> </tr> </tbody> </table> | Name | Duration (ms) | Refreshed visu... | 29 | Image | 28 | Text box | 252 | Slicer | 222 | Card | 204 | Card | 300 | Card | 278 | Cost, Revenue ... | 331 | Revenue and P... | 348 | Cost, Revenue ... | 256 | Cost, Profit an... | - | Cross-highlig... | 50 | Image | 48 | Text box | 231 | Slicer | 231 | Card | 227 |
| Name   | Duration (ms)  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Refreshed visu...  | 29   |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Image  | 28   |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Text box   | 252  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Slicer   | 222  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Card   | 204  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Card   | 300  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Card   | 278  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Cost, Revenue ...  | 331  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Revenue and P...   | 348  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Cost, Revenue ...  | 256  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Cost, Profit an...   | -  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Cross-highlig...   | 50   |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Image  | 48   |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Text box   | 231  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Slicer   | 231  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |
| Card   | 227  |      |               |                   |    |       |    |          |     |        |     |      |     |      |     |      |     |                   |     |                  |     |                   |     |                    |   |                  |    |       |    |          |     |        |     |      |     |

Figure 3 – Screenshot of task A2.1 for Dataset1 using PowerBI Desktop © Microsoft

**Performance tests: Evidence of test results: [Screenshots]**

**A2.2 Data processing**  
 [The screenshot should clearly show the data processing/query related performance information - captured in the loaded report in DAX Studio]

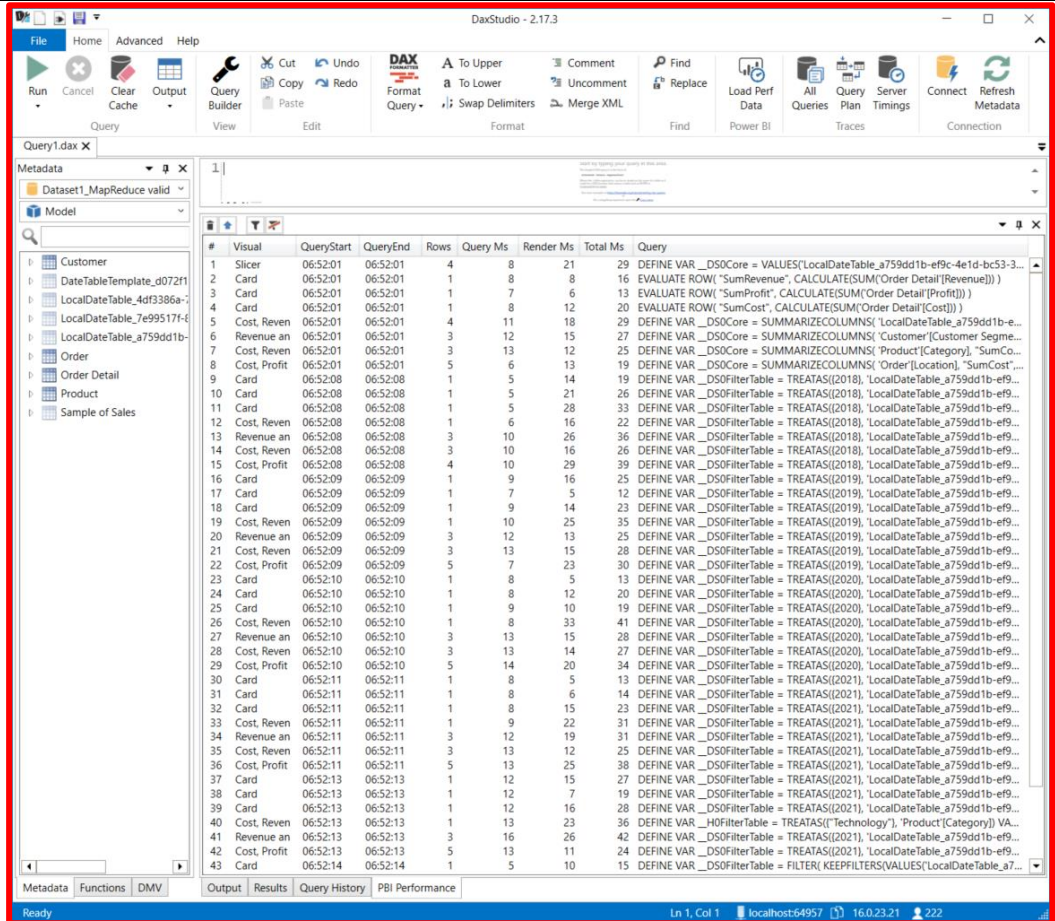


Figure 4 - Screenshot of task A2.2 for Dataset1 using DAX Studio © DAX Studio

**A2.3 Sub-component performance**

[The screenshot should clearly show all sub-component performance metrics for the relevant tables in DAX Studio]

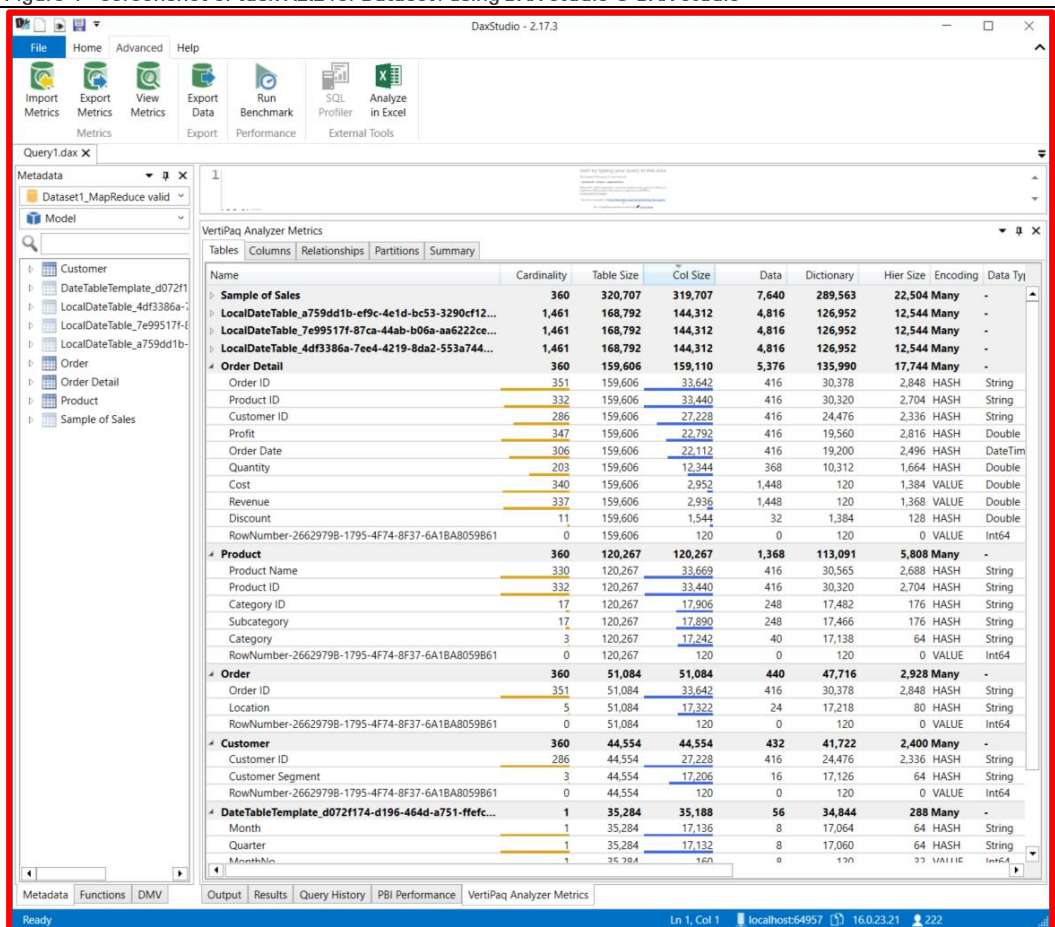


Figure 5 - Screenshot of task A2.3 for Dataset1 using DAX Studio © DAX Studio



Table 3 – Performance testing of Dataset 2 [Non-transactional]

**Performance tests:**

**A2.1 Data throughput**  
*(The screenshot should show the data throughput information captured in the Performance analyser column in PowerBI for Product Output report)*

**Evidence of test results:**  
*(Screenshots)*

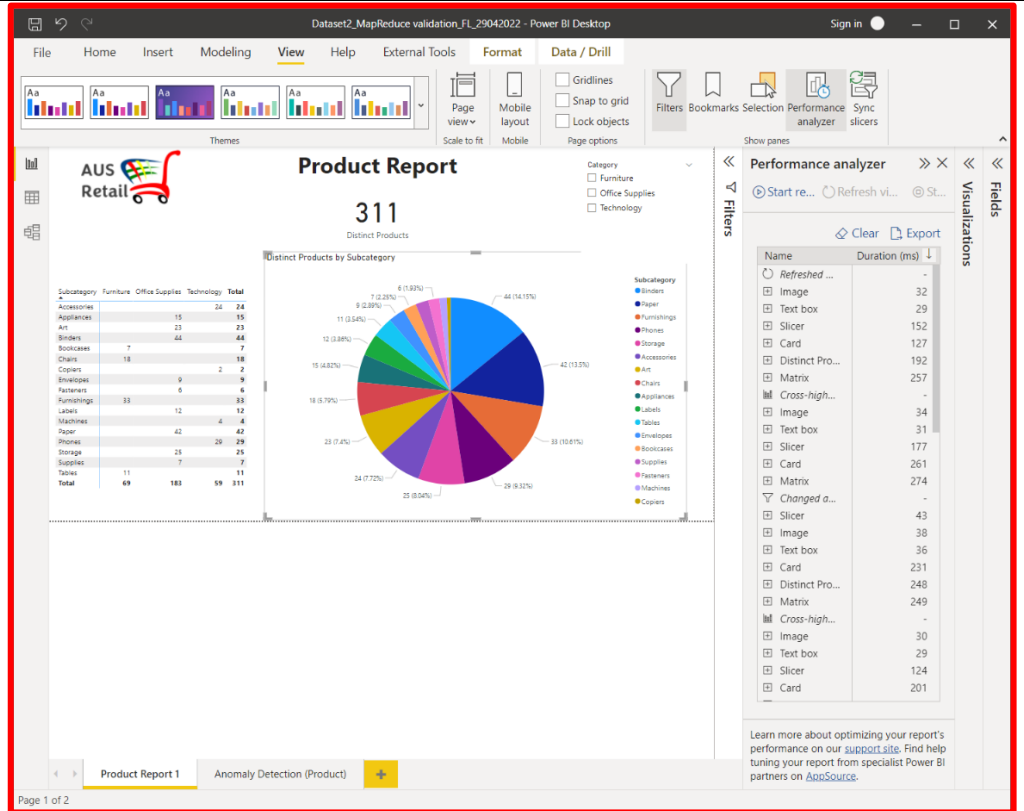


Figure 6 – Screenshot of task A2.1 for Dataset2 using PowerBI Desktop © Microsoft

**A2.2 Data processing**  
*(The screenshot should clearly show the data processing/query related performance information - captured in the loaded report in DAX Studio)*

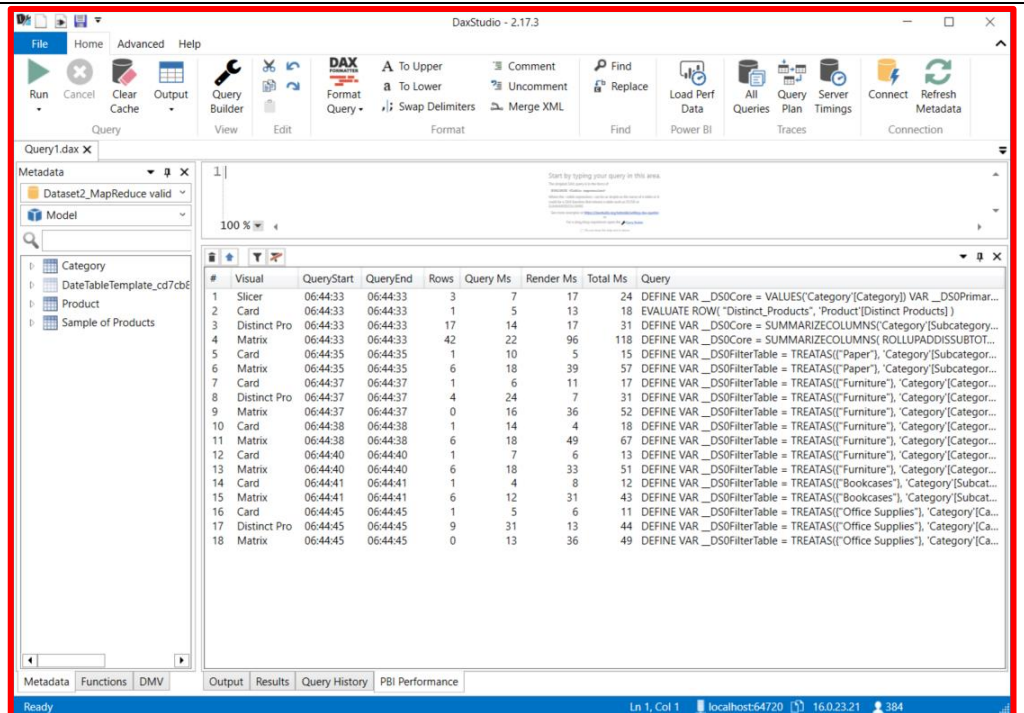


Figure 7 – Screenshot of task A2.2 for Dataset2 using DAX Studio © DAX Studio

**Performance tests:**

**Evidence of test results:  
(Screenshots)**

**A3.3 Sub-component performance**  
*[The screenshot should clearly show all sub-component performance metrics for the relevant tables in DAX Studio]*

| Name   | Cardinality | Table Size       | Col Size         | Data         | Dictionary       | Hier Size    | Encoding | Data Type |
|--|-------------|------------------|------------------|--------------|------------------|--------------|----------|-----------|
| <b>Sample of Products</b>                                | <b>340</b>  | <b>5,362,872</b> | <b>5,362,872</b> | <b>1,656</b> | <b>5,353,056</b> | <b>8,160</b> | Many     | -         |
| Product  | 340         | 3,230,992        | 3,230,992        | 1,408        | 3,221,664        | 7,920        | Many     | -         |
| Product Name   | 312         | 3,230,992        | 1,071,048        | 392          | 1,068,112        | 2,544        | HASH     | String    |
| Product ID   | 311         | 3,230,992        | 1,071,024        | 392          | 1,068,104        | 2,528        | HASH     | String    |
| Category ID  | 17          | 3,230,992        | 1,066,160        | 232          | 1,065,752        | 176          | HASH     | String    |
| Product Price  | 329         | 3,230,992        | 22,640           | 392          | 19,576           | 2,672        | HASH     | Double    |
| RowNumber-2662979B-1795-4F74-8F37-6A18A8059B61           | 0           | 3,230,992        | 120              | 0            | 120              | 0            | VALUE    | Int64     |
| <b>Category</b>  | <b>340</b>  | <b>3,198,160</b> | <b>3,198,160</b> | <b>480</b>   | <b>3,197,264</b> | <b>416</b>   | Many     | -         |
| Category ID  | 17          | 3,198,160        | 1,066,160        | 232          | 1,065,752        | 176          | HASH     | String    |
| Subcategory  | 17          | 3,198,160        | 1,066,160        | 232          | 1,065,752        | 176          | HASH     | String    |
| Category   | 3           | 3,198,160        | 1,065,720        | 16           | 1,065,640        | 64           | HASH     | String    |
| RowNumber-2662979B-1795-4F74-8F37-6A18A8059B61           | 0           | 3,198,160        | 120              | 0            | 120              | 0            | VALUE    | Int64     |
| <b>DateTableTemplate_cd7cb857-5220-4d6e-885e-b88a...</b> | <b>1</b>    | <b>35,284</b>    | <b>35,188</b>    | <b>56</b>    | <b>34,844</b>    | <b>288</b>   | Many     | -         |
| Month  | 1           | 35,284           | 17,136           | 8            | 17,064           | 64           | HASH     | String    |
| Quarter  | 1           | 35,284           | 17,132           | 8            | 17,060           | 64           | HASH     | String    |
| QuarterNo  | 1           | 35,284           | 160              | 8            | 120              | 32           | VALUE    | Int64     |
| MonthNo  | 1           | 35,284           | 160              | 8            | 120              | 32           | VALUE    | Int64     |
| Year   | 1           | 35,284           | 160              | 8            | 120              | 32           | VALUE    | Int64     |
| Date   | 1           | 35,284           | 160              | 8            | 120              | 32           | VALUE    | DateTime  |
| Day  | 1           | 35,284           | 160              | 8            | 120              | 32           | VALUE    | Int64     |
| RowNumber-2662979B-1795-4F74-8F37-6A18A8059B61           | 0           | 35,284           | 120              | 0            | 120              | 0            | VALUE    | Int64     |

Figure 8 – Screenshot of task A2.3 for Dataset2 using DAX Studio © DAX Studio

# Part B: Validate output of captured big data sample and record results

As preparation for this task, do the following first.

- Save a copy of the *AUS Retail\_STM&TestCase\_template.xlsx* files you have created in the previous assessment 3 Project, 'Phase 2 – MapReduce validation' folder into the 'Phase 3 – Output validation' folder for each dataset and rename it with the current date 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 20<sup>th</sup> April 2022 by John Smith should have a filename as follows:
  - '*AUS Retail\_STM&TestCase\_Dataset1(Sales)\_JS\_20042022.xlsx*'

**Note:** You will be further updating these templates with more information in this section of the assessment.

## B1. Design, formulate and select test cases and scenarios

In this task, you are required to design, formulate and select suitable test scenarios and test cases to validate the big data output tables specified in the *AUS Retail\_STM&TestCase\_template.xlsx* > **Source to Target Mapping** tab and the report visualisations you've implemented previously.

### Tasks:

Formulate the test case scenarios for testing the big data samples according to the data quality standards outlined in *AUS Retail\_Big data sample testing policy.pdf* > section 6.1 **Target output table data quality standards for PowerBI reports**.

Document each test case and test scenario clearly, in a logical sequence using the *AUS Retail\_STM&TestCase\_template.xlsx* document > **Test Cases** tab.

### Evidence of performing the task:

Your assessment submission must include the following documents in the 'Phase 3 – Output validation' sub-folder. The *Test Cases* tab should be completed with the required information.

- *AUS Retail\_STM&TestCase\_Dataset1(Sales)\_YourNameInitials\_ddmmyyyy.xlsx*
- *AUS Retail\_STM&TestCase\_Dataset2(Products)\_YourNameInitials\_ddmmyyyy.xlsx*

### Assessor instructions:

The student should use the *AUS Retail\_STM&TestCase\_template.xlsx* > **Test Cases** tab to formulate suitable test case scenarios for testing all target output tables as specified in the **Source to Target Mapping** tab in the Excel spreadsheet template provided. Sample screenshots of the completed **Source to Target Mapping** tab for both datasets are given below.

| Test Case | Test Description   | Test Steps and Method  | Test Data / Query  | Expected Result  |
|-----------|--|--|--|--|
| SD_001    | Verify <b>Order ID</b> column in <b>Order Detail</b> table | 1. Check the datatype and data format <b>Order ID</b> in <b>Order Detail</b> table<br><b>Method:</b> Select the <b>Order ID</b> column in PowerBI Desktop > Data view; then check the Data type and Format details under Column Tools.<br>2. Validate <b>Order ID</b> data correctness<br><b>Method:</b> Run query in DAX Studio | Evaluate Filter ('Order Detail', [Customer ID] = "FM-14290")   | Text<br><br>AU-2020-114867<br>AU-2021-121160   |
|           |  | 3. Check for missing values or blank cells<br><b>Method:</b> Run query in DAX Studio   | DEFINE<br>MEASURE 'Order Detail'(EmptyID) =<br>CALCULATE (<br>COUNTROWS ( 'Order Detail' ),<br>'Order Detail'[Order ID] = BLANK ()<br>)<br>EVALUATE<br>SUMMARIZECOLUMNS (<br>'Order Detail'[Order ID],<br>"Order Detail", COUNTROWS ( 'Order Detail'<br>),<br>"Order Detail with blank Order ID", [EmptyID]<br>) | No values returned for <b>Order Detail with blank Order ID</b> in the test output.                                 |
| SD_002    | Verify <b>Cost</b> column in <b>Order Detail</b> table     | 1. Check the datatype and data format <b>Cost</b> in <b>Order Detail</b> table<br><b>Method:</b> Select the <b>Cost</b> column in PowerBI<br>2. Validate <b>Cost</b> data correctness<br><b>Method:</b> Run query in DAX Studio  | Evaluate Filter ('Order Detail', [Order ID] = "AU-2020-114867")  | Currency<br>Two decimals<br><br>\$1,874.96   |
|           |  | 3. Check for missing values or blank cells<br><b>Method:</b> Run query in DAX Studio   | DEFINE<br>MEASURE 'Order Detail'(EmptyCost) =<br>CALCULATE (<br>COUNTROWS ( 'Order Detail' ),<br>'Order Detail'[Cost] = BLANK ()<br>)<br>EVALUATE<br>SUMMARIZECOLUMNS (<br>'Order Detail'[Order ID],<br>"Order Detail", COUNTROWS ( 'Order Detail'<br>),<br>"Orders with blank Cost", [EmptyCost]<br>)           | No values returned for <b>Orders with blank Cost</b> in the test output.   |
|           |  | 4. Check for negative cost values  | Evaluate Filter ('Order Detail', [Cost] < 0)   | No values returned for negative costs.   |
| SD_003    | Validate <b>Order Detail</b> table                         | 1. Check table fields/columns<br><b>Method:</b> Run query in DAX Studio<br><br>2. Check for duplicate records<br><b>Method:</b> Run query in DAX Studio  | Evaluate 'Order Detail'  | Row ID<br>Order ID<br>Customer ID<br>Product ID<br>Cost<br>Revenue<br>Profit<br>Order Date<br>Quantity<br>Discount |
|           |  |  | Evaluate SUMMARIZECOLUMNS("Total rows", Countrows('Order Detail'), "Distinct rows", Countrows(DISTINCT('Order Detail')))   | Same value displayed for both Total rows and Distinct rows.  |

Figure 9 - Screenshot of task B1 for Dataset 1 using Microsoft Excel © Microsoft

| Test Case | Test Description  | Test Steps and Method   | Test Data / Query   | Expected Result   |
|-----------|---|---|---|---|
| PD_001    | Verify <b>Product ID</b> column in <b>Product</b> table | 1. Check the datatype and data format <b>Product ID</b> in <b>Product</b> table<br><b>Method:</b> Select the <b>Product</b> column in PowerBI Desktop > Data view; then check the Data type and   |   | Text  |
|           |   | 2. Validate <b>Product ID</b> data correctness<br><b>Method:</b> Run query in DAX Studio  | Evaluate Filter ('Product', [Product ID] = "OFF-AP-10004487")   | Category ID: 10FS-APL<br>Product Name: Kensington 4 Outlet MasterPiece Compact Power Control Center |
|           |   | 3. Check for missing values or blank cells<br><b>Method:</b> Run query in DAX Studio  | DEFINE<br>MEASURE 'Product'[EmptyID] =<br>CALCULATE (<br>COUNTROWS ( 'Product' ),<br>'Product'[Product ID] == BLANK ()<br>)<br>EVALUATE<br>SUMMARIZECOLUMNS (<br>'Product'[Product ID],<br>"Product", COUNTROWS ( 'Product' ),<br>"Product with blank Product ID",<br>[EmptyID]<br>)                  | No values returned for <b>Product with blank Product ID</b> in the test output.                     |
| PD_002    | Verify <b>Product Price</b> column in <b>Product</b>    | 1. Check the datatype and data format <b>Product Price</b> in <b>Product</b> table<br><b>Method:</b> Select the <b>Product Price</b> column in PowerBI Desktop > Data view; then check the Data type and Format details under Column Tools. |   | Currency<br>Two decimals  |
|           |   | 2. Validate <b>Product Price</b> data correctness<br><b>Method:</b> Run query in DAX Studio   | Evaluate Filter ('Product', [Product ID] = "OFF-AP-10004487")   | \$49.69   |
|           |   | 3. Check for missing values or blank cells<br><b>Method:</b> Run query in DAX Studio  | DEFINE<br>MEASURE<br>'Product'[EmptyProductID] =<br>CALCULATE (<br>COUNTROWS ( 'Product' ),<br>'Product'[Product ID] == BLANK ()<br>)<br>EVALUATE<br>SUMMARIZECOLUMNS (<br>'Product'[Product ID],<br>"Product", COUNTROWS ( 'Product' ),<br>"Product with blank Product ID",<br>[EmptyProductID]<br>) | No values returned for <b>Products with blank Product ID</b> in the test output.                    |
|           |   | 4. Check for negative cost values   | Evaluate Filter ('Order Detail', [Cost] < 0 )   | No values returned for negative costs.  |
| PD_003    | Validate <b>Product</b> table                           | 1. Check table fields/columns<br><b>Method:</b> Run query in DAX Studio   | Evaluate 'Product'  | Category ID<br>Product ID<br>Product Name<br>Product Price  |
|           |   | 2. Check for duplicate records<br><b>Method:</b> Run query in DAX Studio  | Evaluate<br>SUMMARIZECOLUMNS(<br>"Total rows", Countrows('Product'),<br>"Distinct rows",<br>Countrows(DISTINCT('Product'))  | Same value displayed for both Total rows and Distinct rows.   |

Figure 10 - Screenshot of task B1 for Dataset 2 using Microsoft Excel © Microsoft



## B2. Implement selected test cases and scenarios

In this task, you are required to use the test case scenarios you have formulated in task B1 and implement the selected test cases by testing the sample data using the recommended tools and recording the test results according to organisational procedures.

### Tasks:

- Refer to the *AUS Retail\_Big data sample testing policy.pdf* > section 6.2 Procedure for implementing big data sample test cases and scenarios.
- Use the recommended testing tools and process to implement the test cases and scenarios formulated in Task D1.
- Document the actual results and the final test results [Pass/Fail] using the *AUS Retail\_STM&TestCase\_template.xlsx* > *Validation* tab, columns [Actual Result 1 and Test Result 1] for Test Run 1. You must use clear, specific and industry-related terminology when representing test results.

### Evidence of performing the task:

Your assessment submission 'BSBXBD402\_Firstname\_Lastname' folder must include the following Excel template documents within the 'Phase 3 – Output validation' sub-folder. The Excel templates should include Test Run 1 Results documented in the *Validation* tab.

- *AUS Retail\_STM&TestCase\_Dataset1(Sales)\_YourNameInitials\_ddmmyyyy.xlsx*
- *AUS Retail\_STM&TestCase\_Dataset2(Products)\_YourNameInitials\_ddmmyyyy.xlsx*

### Assessor instructions:

The student should use the *AUS Retail\_STM&TestCase\_template.xlsx* > *Test Cases* tab to document the actual results of the tests using the Excel spreadsheet template provided. A screenshot of the sample test case implementation for dataset 1 is given below. Similarly, students must provide evidence of implementing test cases for dataset 2.

*AUS Retail\_STM&TestCase\_Dataset1(Sales)\_YourNameInitials\_ddmmyyyy.xlsx*

| Test Run 1 - 02052022 |  |  |  |  |  |               |                  |
|-----------------------|--|--|--|--|--|---------------|------------------|
| Test Case             | Test Description                             | Test Steps and Method  | Test Data / Query  | Expected Result  | Actual Result 1  | Test Result 1 | Notes/Comments   |
| SD_001                | Verify Order ID column in Order Detail table | 1. Check the datatype and data format Order ID in Order Detail table<br>Method: Select the Order ID column in PowerBI Desktop > Data view; then check the Data type and Format details under Column Tools. | 0  | Text   | Text   | Pass          |                  |
| 0                     | 0  | 2. Validate Order ID data correctness<br>Method: Run query in DAX Studio   | Evaluate Filter ('Order Detail', [Customer ID] = "FM-14290")   | AU-2020-114867<br>AU-2021-121160   | AU-2020-114867<br>AU-2021-121160   | Pass          |                  |
| 0                     | 0  | 3. Check for missing values or blank cells<br>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.  | 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<br>Method: Select the Cost column in PowerBI Desktop > Data view; then check the Data type and Format details under Column Tools.         | 0  | Currency<br>Two decimals   | General, Decimal number having more than two values  | Fail          |                  |
| 0                     | 0  | 2. Validate Cost data correctness<br>Method: Run query in DAX Studio   | Evaluate Filter ('Order Detail', [Order ID] = "AU-2020-114867")  | 1874.96  | 1874.96  | Pass          |                  |
| 0                     | 0  | 3. Check for missing values or blank cells<br>Method: Run query in DAX Studio  | DEFINE MEASURE 'Order Detail'[EmptyCost] = CALCULATE ( COUNTROWS ( 'Order Detail' ),                                       | No values returned for Orders with blank Cost in the test output.  | No values returned   | Pass          |                  |
| 0                     | 0  | 4. Check for negative cost values  | Evaluate Filter ('Order Detail', [Cost] < 0 )  | No values returned for negative costs.   | No values returned   | Pass          |                  |
| SD_003                | Validate Order Detail table                  | 1. Check table fields/columns<br>Method: Run query in DAX Studio   | Evaluate 'Order Detail'  | Row ID<br>Order ID<br>Customer ID<br>Product ID<br>Cost<br>Revenue<br>Profit<br>Order Date<br>Quantity<br>Discount | Row ID<br>Order ID<br>Customer ID<br>Product ID<br>Cost<br>Revenue<br>Profit<br>Order Date<br>Quantity<br>Discount | Pass          |                  |
| 0                     | 0  | 2. Check for duplicate records<br>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.  | Distinct rows < Total rows   | Fail          | Duplicates exist |

Figure 11 – Screenshot of task B2 for Dataset 1 using Microsoft Excel © Microsoft

## B3. Isolate sub-standard data

In this task, you are required to isolate the sub-standard data from the sample dataset source files.

### Instructions:

Do the following first.

Save a copy of the PowerBI work files [*Dataset1\_MapReduce validation\_YourNameInitials\_DDMMYYYY* and *Dataset2\_MapReduce validation\_YourNameInitials\_DDMMYYYY*] you have created in the 'Phase 2 – MapReduce validation' folder into the 'Phase 3 – Output validation' folder and rename it as follows with the current date.

- Dataset1\_Output validation\_YourNameInitials\_DDMMYYYY'
- Dataset2\_Output validation\_YourNameInitials\_DDMMYYYY'

For example, a file saved on the 28<sup>th</sup> April 2022 by John Smith for Dataset1 should have the file name: 'Dataset1\_Output validation \_JS\_28042022'

Read through the scenario carefully before doing the tasks.

### Scenario:

Your supervisor had advised you of the following types of sub-standard data that need to be isolated in the departmental reports for AUS Retail.

#### Dataset 1 [transactional]

- Sales records that have **minus [-]** values for *Profit* should be isolated from the sample dataset

#### Dataset 2 [non-transactional]

The products that belong to the subcategory 'Copiers' are no longer in production and have been phased out. Therefore, they need to be isolated from the sample dataset.

### Tasks:

For each of the following tasks, provide evidence in the form of screenshots in the answer table below. The screenshots should capture details of the new tables created and queries used to isolate the sub-standard data.

B3.1 Create a table called 'Minus Profit' to isolate the sub-standard sales records relevant to Dataset 1.

**Assessor instructions:** Students need to create a profit column in the sample dataset source table (Sample of Sales) in order to do this task. They will need to apply this by what they've learned so far.

B3.2 Create a new sample table for sales called 'Sample of Sales2' which does not contain the records in 'Minus Profit' table.

B3.3 Create a table called 'Copier Products' to isolate the sub-standard product records relevant to Dataset 2.

B3.4 Create a new sample table for products called 'Sample of Products2' which does not contain the records in 'Copier Products' table.

### Evidence of performing the task:

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

Table 4 – Evidence of performing the demonstration task B3

New tables names:

Evidence of creating the table and query used:

*(Screenshot clearly showing the table contents/columns and DAX query used)*

**B3.1 Minus Profit table**

**Assessor Instructions:**  
Students must use the correct DAX query to isolate all records that have a minus value as shown in the screenshot.

The screenshot shows the Power BI Desktop interface. The main view is a table with the following columns: Product\_Name, Sales, Quantity, Discount, Cost, Year, Month, and Profit. The table contains 72 rows of data. A DAX query is applied to the table: `1 Minus Profit = FILTER('Sample of Sales', 'Sample of Sales'[Profit] < 0)`. A 'Sort ascending' dialog box is open over the Profit column, showing a list of values to be sorted, including (Select all), -383.031, -374.9925, -122.877, -122.3928, -116.844, -97.72, -88.784, -84.448, -82.99, -81.3065, -80.178, -68.1856, -66.3916, -59.373, -53.7432, -52.8908, -49.92, -44.1552, -43.848, and -42.657R. The 'Fields' pane on the right shows the 'Sample of Sales' table with columns like Customer, Category, Category\_ID, Cost, Country, Customer\_ID, Discount, Month, Order\_Date, Order\_ID, Product\_ID, Product\_Name, Profit, Quantity, Row\_ID, Sales, Segment, Ship\_Date, Ship\_Mode, State, Sub-category, and Year.

Figure 12 – Screenshot of task B3.1 for Dataset 1 using PowerBI Desktop © Microsoft

**New tables names:**

**Evidence of creating the table and query used:**  
*[Screenshot clearly showing the table contents/columns and DAX query used]*

**B3.2 Sample of Sales2 table**

**Assessor Instructions:**  
 Students must use the correct DAX query to exclude the records in the 'Minus Profit' table from the new 'Sample of Sales2' table as shown in the screenshot.

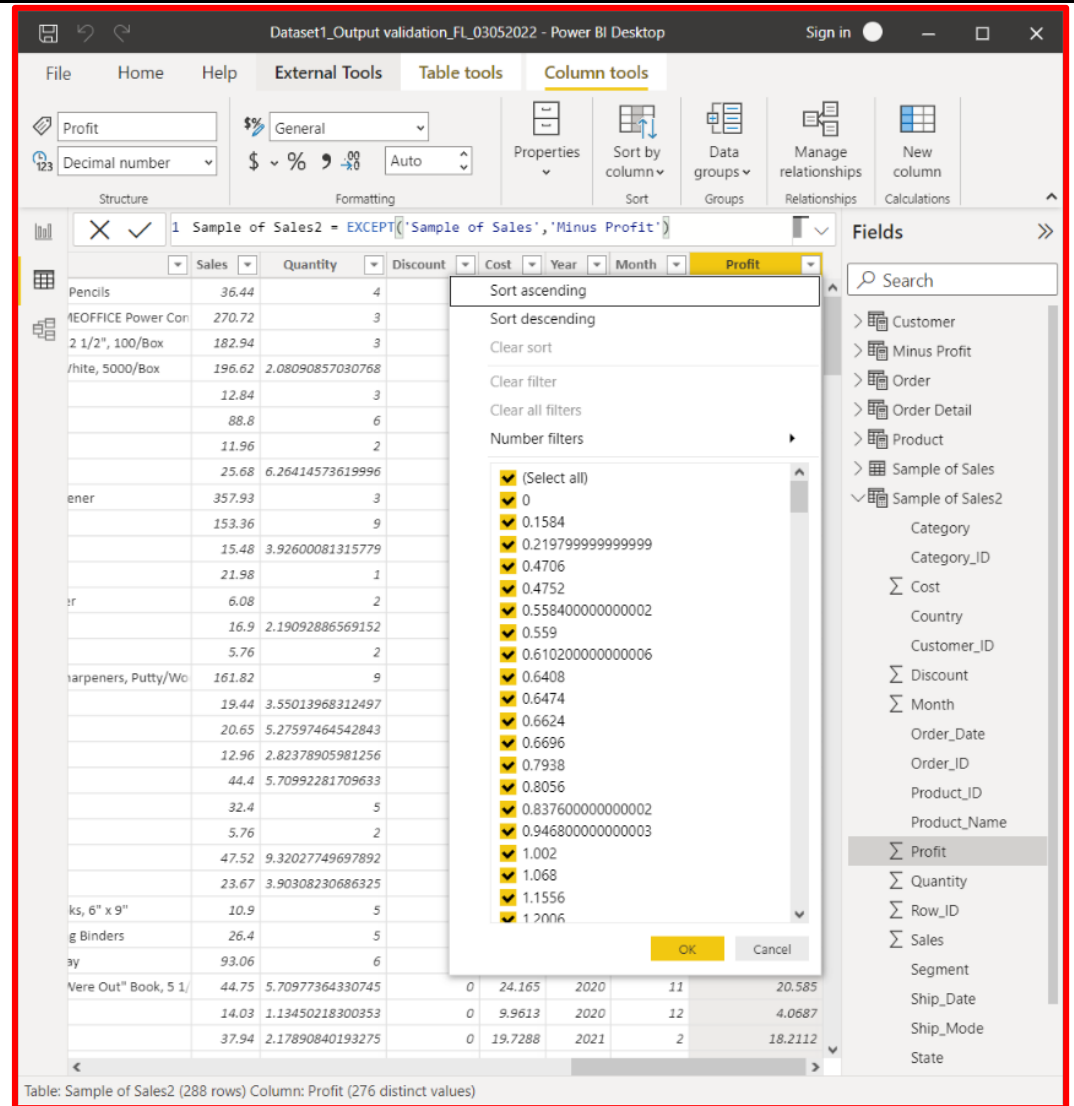


Figure 13 – Screenshot of task B3.2 for Dataset 1 using PowerBI Desktop © Microsoft

**B3.3 Copier Products table**

**Assessor Instructions:**  
 Students must use the correct DAX query to isolate all product records that belong to the Copiers sub-category as shown in the screenshot.

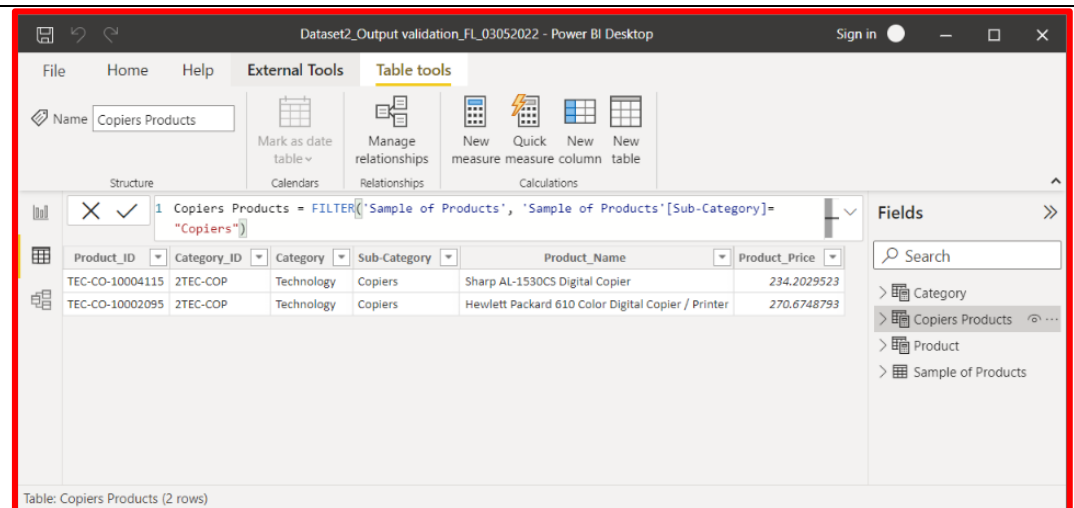


Figure 14 – Screenshot of task B3.3 for Dataset 2 using PowerBI Desktop © Microsoft



New tables names:

Evidence of creating the table and query used:

[Screenshot clearly showing the table contents/columns and DAX query used]

### B3.4 Sample of Products2 table

#### Assessor

#### Instructions:

Students must use the correct DAX query to exclude the records in the 'Copier Products' table from the new 'Sample of Products2' table as shown in the screenshot.

| Product_ID      | Category_ID | Category        | Sub-Category | Product_Name   | Product_Price |
|-----------------|-------------|-----------------|--------------|--|---------------|
| OFF-AP-10002495 | 1OFS-APL    | Office Supplies | Appliances   | Acco Smartsocket Table Surge Protector, 6 Color-Coded Ac | 2             |
| OFF-AP-10003040 | 1OFS-APL    | Office Supplies | Appliances   | Fellowes 8 Outlet Superior Workstation Surge Protector w | 2             |
| OFF-AP-10004487 | 1OFS-APL    | Office Supplies | Appliances   | Kensington 4 Outlet MasterPiece Compact Power Control    | 49.69         |
| OFF-AP-10002118 | 1OFS-APL    | Office Supplies | Appliances   | 1.7 Cubic Foot Compact "Cube" Office Refrigerators       | 15            |
| OFF-AP-10001634 | 1OFS-APL    | Office Supplies | Appliances   | Hoover Commercial Lightweight Upright Vacuum             |               |
| OFF-AP-10001394 | 1OFS-APL    | Office Supplies | Appliances   | Harmony Air Purifier                                     | 104.4         |
| OFF-AP-10001469 | 1OFS-APL    | Office Supplies | Appliances   | Fellowes 8 Outlet Superior Workstation Surge Protector   | 24.90         |
| OFF-AP-10000692 | 1OFS-APL    | Office Supplies | Appliances   | Fellowes Mighty 8 Compact Surge Protector                | 12.13         |
| OFF-AP-10001947 | 1OFS-APL    | Office Supplies | Appliances   | Acco 6 Outlet Guardian Premium Plus Surge Suppressor     | 1             |
| OFF-AP-10001492 | 1OFS-APL    | Office Supplies | Appliances   | Acco Six-Outlet Power Strip, 4' Cord Length              |               |
| OFF-AP-10003849 | 1OFS-APL    | Office Supplies | Appliances   | Hoover Shoulder Vac Commercial Portable Vacuum           | 225.6         |

Figure 15 – Screenshot of task B3.4 for Dataset 2 using PowerBI Desktop © Microsoft

## B4. Correct data acquisition paths as required

In this task, you are required to make the necessary corrections to the data acquisition paths of the reports to reflect the changes after isolating sub-standard data in the previous task.

### Tasks:

Correct the data acquisition paths for the visualisations in the reports for both datasets by doing the following:

- Rename the 'Sales Report 1' and 'Product Report 1' tabs in the PowerBI workfiles and to 'Sales Report 2' and 'Product Report 2'.
- Correct all the target output table queries in each dataset to retrieve data from the new sample table created in task B3 (Sample of Sales2, Sample of Products2).
- Ensure that the visualisations in the 'Sales Report 2' and 'Product Report 2' reflect data from the new sample tables.
- As evidence of performing this task, provide screenshots of the reports in the 'Sales Report 2' and 'Product Report 2' reports in the answer table given below. The screenshots should show the changes done in the DAX query to correct the source data table name in at least one of the target output tables in each dataset.

### Evidence of performing the task:

**Assessor Instructions:** Students must provide two screenshots showing the changes in the reports after isolating sub-standard data and changing the data acquisition paths. Sample screenshots are given below.

Table 5 – Evidence of performing demonstration task B4

**New report view name:** Evidence of performing the tasks: *[Screenshots]*

**Sales Output 2 report:**  
*[The screenshot should show the changes in the visualisations and the changes made to the DAX query to correct the data acquisition paths in the target output tables.]*

Figure 16 – Screenshot of task B4 for Dataset 1 using PowerBI Desktop © Microsoft

**Product Output 2 report:**  
*[The screenshot should show the changes in the visualisations and the changes made to the DAX queries to correct the data acquisition paths in the target output tables.]*

Figure 17 – Screenshot of task B4 for Dataset 2 using PowerBI Desktop © Microsoft

<Assessors, please refer to Part 2 of this Assessor Guide for parts B5 onwards.>