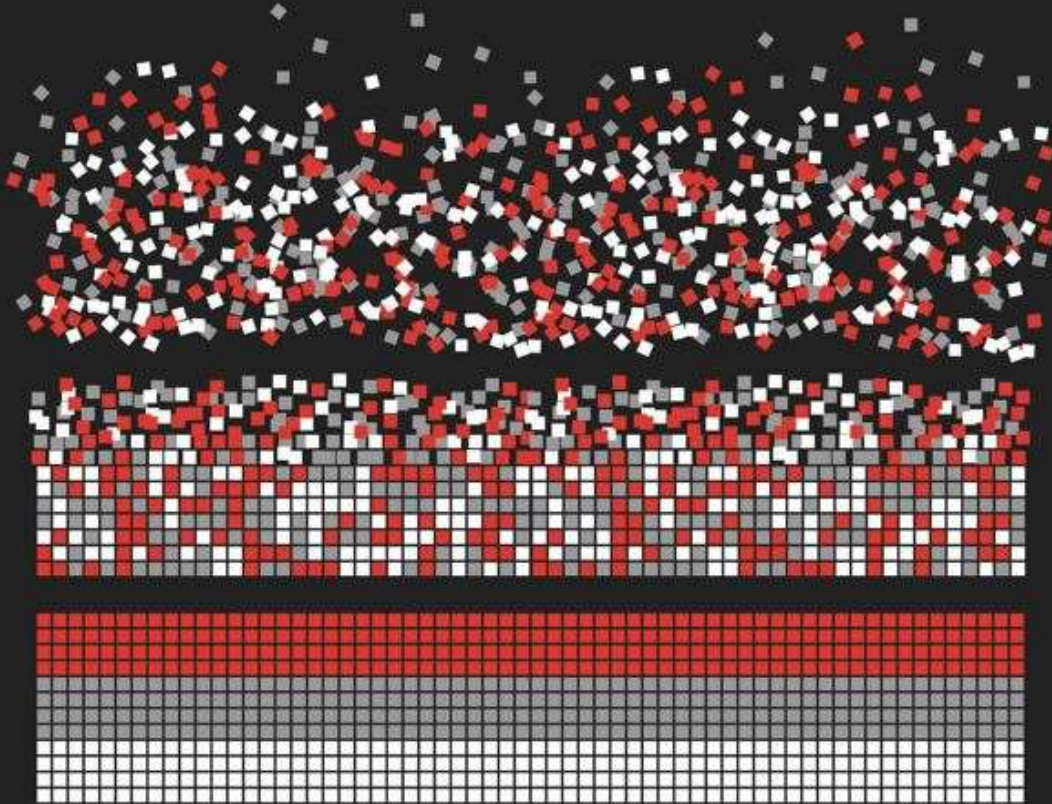


# BIG DATA



Assessor Guide

BSBXBD402

## Test big data samples

Assessment 3 of 4

Project



## Assessment Instructions

### Task overview

This assessment task is divided into three parts having eight (8) 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 PowerBI Desktop App - Download and install the free **PowerBI 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 PowerBI – 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

## 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 C: Validate big data sample process and business logic

As preparation for this task, do the following first.

1. Save a copy of the *AUS Retail\_STM&TestCase\_template.xlsx* in the 'Phase 2 – MapReduce 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 20<sup>th</sup> April 2022 by John Smith should have a filename as follows:*

  - **'AUS Retail\_STM&TestCase\_Dataset1(Sales)\_JS\_20042022.xlsx'**
2. Place a copy of the previously validated sample dataset files from the 'Phase 1 – Data validation' folder into the 'Phase 2 – MapReduce validation' folder as you will be performing process validation tasks on these sample datasets in this part of the assessment.
  - *AUS Retail Sales\_sample*
  - *AUS Retail Products\_sample*

## C1. Align datasets to relevant parts of the organisation

In this task, you are required to further evaluate the contents of each sample dataset and align the source fields (column names) to specific entities and relevant parts of the organisation.

### Instructions:

Refer to the following documents, specifications, and advice from your supervisor to understand AUS Retail's business logic and reporting requirements.

- *AUS Retail\_ Data flow and schemas.pdf* – outlines the operational data types, sources, flows and recommended schemas to be implemented.
- *AUS Retail\_Reporting requirements.pdf* – outlines the requirements for reporting as relevant for the sales and production departments.
- *AUS Retail\_ Big data sample testing policy.pdf* > section 5.1 **Target table field/column alignment with source systems** – outlines specific requirements to consider when aligning source system dataset fields/columns with the target output table fields/columns.
- Advice received from your supervisor as shown below:

*"Please consider the following additional requirements for sales related target output fields that need to be captured in the source to target mapping table.*

- *A new 'Profit' column to calculate profit from each sales order. (Profit = Revenue – Cost)*
- *A new 'Location' column that combines the County, State details to indicate the location from where each order is placed*
- *The Sales column should be renamed as Revenue.*
- *The Management had also informed that any shipping related data is not required to be included in the sales reports."*

### Task:

Complete the *AUS Retail\_STM&TestCase\_template.xlsx* > **Source to Target Mapping** tab for each dataset by:

- using the *AUS Retail\_STM&TestCase\_template.xlsx* documents to record source to mapping details separately for each dataset

- identifying and recording the following details for each dataset in the *Source to Target Mapping* tab
  - source system details such as table name and source field names - record the sample dataset table name and its associated column names for each dataset
  - target output table details such as the table names and target field names - align each Source Field (Column Name) recorded for each dataset to the relevant entities of the organisation by filling the Target Output (Table Name) and Target Fields (Column Name) columns
  - transformation logic – this includes details of any queries, functions, expressions, filters or calculation formulas etc. that can be used to generate the required result to create the target fields/columns.
  - use the comments column to:
    - record any discrepancies between the source field data and the target field data considering the target output requirements for PowerBI reporting
    - make notes of any columns that are not required for reporting.
  - including any new target output fields that are required to get the desired output according to the reporting requirements and advice from your supervisor.

### Evidence of performing the task:

Your assessment submission must include the following documents in the 'Phase 2 – MapReduce validation' sub-folder. The *Source to Target Mapping* 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:** Samples of the completed Source to Mapping information for both datasets are given below. Also refer to the contents in the **BSBXBD402\_AG\_03\_Project\_Exemplar (student submission folder)** sample work files.

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

Source System (Table Name)	Source Field (Column Name)	Transformation Logic (Dirty, Formatting) - if applicable	Target Output (Table Name)	Target Field (Column Name)	Comments on source table:
Sample of Sales	Row_ID		Order Detail (Fact table)	Row ID	Discrepancy exists between source field name and target field name.
Sample of Sales	Order_ID		Order Detail (Fact table)	Order ID	Discrepancy exists between source field name and target field name.
Sample of Sales	Order_Date		Order (Dimension table)	Order Date	Discrepancy exists between source field name and target field name.
Sample of Sales	Ship_Code			Ship Date	Need to format (ISO8601)
Sample of Sales	Ship_Mode			Ship Mode	Not required in the report.
Sample of Sales	Customer_ID		Order Detail (Fact table)	Customer ID	Discrepancy exists between source field name and target field name.
Sample of Sales	Country		Customer (Dimension table)	Country	
Sample of Sales	State		Customer (Dimension table)	State	
Sample of Sales	Segment		Customer (Dimension table)	Customer Segment	
Sample of Sales	Category_ID		Product (Dimension table)	Category ID	Discrepancy exists between source field name and target field name.
Sample of Sales	Category		Product (Dimension table)	Category	
Sample of Sales	Sub-Category		Product (Dimension table)	Subcategory	Discrepancy exists between source field name and target field name.
Sample of Sales	Product_ID		Order Detail (Fact table)	Product ID	Discrepancy exists between source field name and target field name.
Sample of Sales	Product_Name		Product (Dimension table)	Product Name	Discrepancy exists between source field name and target field name.
Sample of Sales	Sales		Order Detail (Fact table)	Revenue	
Sample of Sales	Quantity		Order Detail (Fact table)	Quantity	
Sample of Sales	Discount		Order Detail (Fact table)	Discount	
Sample of Sales	Cost		Order Detail (Fact table)	Cost	
Sample of Sales	Country State	Location = (Country & " " & State)	Order (Dimension table)	Location	Need to combine Country and State fields from source table and set target field name as 'Location'.
Sample of Sales	Sales Cost	Profit = Sales - Cost	Order Detail (Fact table)	Profit	Need to calculate and set target field name as 'Profit'.

Figure 1 - Screenshot for task C1 Dataset 1 using Microsoft Excel © Microsoft

*AUS Retail\_STM&TestCase\_Dataset2(Products)\_YourNameInitials\_ddmmyyy.xlsx*



Source System (Table Name)	Source Field (Column Name)	Transformation Logic (Query, Formatting) - if applicable	Target Output (Table Name)	Target Field (Column Name)	Comments on issues noted:
Sample of Products	Category_ID		Product (Fact table) Category (Dimension table)	Category ID	Discrepancy exists between source field name and target field name
Sample of Products	Category		Category (Dimension table)	Category	
Sample of Products	Sub-Category		Category (Dimension table)	Subcategory	Discrepancy exists between source field name and target field name
Sample of Products	Product_ID		Product (Fact table)	Product ID	Discrepancy exists between source field name and target field name
Sample of Products	Product_Name		Product (Fact table)	Product Name	Discrepancy exists between source field name and target field name
Sample of Products	Product_Price		Product (Fact table)	Product Price	Discrepancy exists between source field name and target field name

Figure 2 - Screenshot for task C1 Dataset 2 using Microsoft Excel © Microsoft

## C2. Implement data segregation rules

In this task, you are required to implement data segregation rules to create the required target output tables from the sample datasets loaded in *PowerBI* according to the **Source to Target Mapping** table (included in *AUS Retail\_STM&TestCase\_template.xlsx*) created in the previous task.

### Instructions:

Open the *PowerBI Desktop* application and save a blank *PowerBI* file in the 'Phase 2 – MapReduce validation' folder for each dataset as:

- 'Dataset1\_MapReduce validation\_YourNameInitials\_DDMMYYYY'
- 'Dataset2\_MapReduce validation\_YourNameInitials\_DDMMYYYY'.

E.g. A file saved on the 12<sup>th</sup> April 2022 by John Smith should have the name: 'Dataset1\_MapReduce validation\_JS\_12042022'

When providing screenshots, ensure that they clearly show the *Report* view tabs that are named appropriately to indicate which type of data is displayed in the report.

### Tasks:

C2.1 Load the validated sample datasets (AUS Retail Sales\_sample.xlsx and AUS Retail Products\_sample.xlsx) into the associated *PowerBI* files.

- Rename the loaded sample datasets tables in *PowerBI Desktop* accordingly. (e.g. Sample of Sales, Sample of Products)

C2.2 Create new tables to segregate the sample datasets into separate target tables using DAX queries according to the *Source to Target Mapping* you've completed in task C1. In doing so, ensure that you:

- rename each new table with the relevant target output table name
  - use the correct DAX function to select required columns from the validated sample dataset whilst renaming and creating new target field/column names as required
- Important note: If you notice any anomalies or inconsistencies in the output data in the tables, do not try to fix them at this stage. You will be reporting on and fixing these issues and anomalies at a later task.**
- Once all the target tables have been created, select the option to **Hide** the sample dataset table (Sample of Sales, Sample of Product) from Report View.
  - for each new table created, provide screenshots separately for each dataset using the answer tables,
    - Table 6: Target output tables for Dataset1 (Transactional)
    - Table 7: Target output tables for Dataset2 (Non-transactional)
  - The screenshots should clearly show:
    - an expanded view of the *Fields* column in *PowerBI Desktop* showing the new tables created
    - the name of the PowerBI file in the “BSBXBD402 – Firstname\_Lastname\_DDMMYYYY” format on the title bar of PowerBI window.

C2.3 In the *model* view in *PowerBI* for each dataset,

- create a new tab and rename it to reflect the correct department name of the dataset
  - drag and drop the tables relevant for each department within each data model view tab
  - create the appropriate relationships between the tables
- Important: Select the recommended relationship type in PowerBI at this stage. If you notice any anomalies make a note of them as these will need to be addressed at a later stage.**
- provide a screenshot of each new data model view tab created in **Table 8: New data model views for each department**. Your screenshots should clearly show:
    - the *Data model* view tabs for each department with the department name.
    - the name of the PowerBI file in the “BSBXBD402 – Firstname\_Lastname\_DDMMYYYY” format on the title bar of PowerBI window.
  - the name of the PowerBI file displayed on the title bar as 'Dataset#\_MapReduce validation\_YourNameInitials\_DDMMYYYY'.

### Evidence of performing the tasks:

In addition to the screenshots you will include in **Table 6**, **Table 7** and **Table 8** given below, your assessment submission must include the following documents in the 'Phase 2 – MapReduce validation' sub-folder. The PowerBI work files should contain evidence of implementing data segregation rules.

- 'Dataset1\_MapReduce validation\_YourNameInitials\_DDMMYYYY'
- 'Dataset2\_MapReduce validation\_YourNameInitials\_DDMMYYYY'.

**Assessor instructions:** Refer to the contents in the **BSBXBD402\_AG\_03\_Project\_Exemplar (student submission folder)** sample work files.



Table 1 - Target output tables for Dataset 1 (Transactional)

New table name:

Evidence of performing the task:  
*(Screenshots showing DAX functions/queries used to create the tables from the source table)*

Order Detail

```

1 Order Detail = SELECTCOLUMNS('Sample of Sales',
2   "Row ID", [Row_ID],
3   "Order ID", [Order_ID],
4   "Order Date", [Order_Date],
5   "Customer ID", [Customer_ID],
6   "Product ID", [Product_ID],
7   "Quantity", [Quantity],
8   "Cost", [Cost],
9   "Discount", [Discount],
10  "Revenue", [Sales],
11  "Profit", [Sales]-[Cost]
12 )
    
```

Order ID	Customer ID	Product ID	Cost	Revenue	Profit
AU-2018-147627	HL-15040	FUR-FU-10003194	27.02	38.6	11.58
AU-2018-143637	MS-17710	FUR-FU-10002813	25.9072	40.48	14.5728
AU-2018-101462	BP-11230	FUR-FU-10000409	32.3568	59.92	27.5632
AU-2018-103429	LW-16825	TEC-PH-10003505	329.44	464	134.56
AU-2018-141796	JG-15160	TEC-PH-10001578	862.5435	1214.85	352.3065
AU-2018-124478	MA-17560	TEC-PH-10001128	215.9856	299.98	83.9944
AU-2018-131002	TB-21400	FUR-FU-10004665	608.1912	821.88	213.6888
AU-2018-109456	LS-17245	TEC-AC-10003610	93.5844	179.97	86.3856
AU-2018-106439	GG-14650	TEC-AC-10004568	204.0471	251.91	47.8629
AU-2018-151005	OH-18715	FUR-FU-10002818	27.0207	60.71	33.6893

Table: Order Detail (480 rows)

Figure 3 - Screenshot for Dataset1 Order Detail table using PowerBI Desktop © Microsoft

Order

```

1 Order = SELECTCOLUMNS('Sample of Sales',
2   "Order ID", [Order_ID],
3   "Location", [Country]&" "&[State]
4 )
    
```

Order ID	Location
AU-2018-156349	Australia,NSW
AU-2018-105767	Australia,NSW
AU-2018-140858	Australia,NSW
AU-2018-115812	Australia,NSW
AU-2018-165862	Australia,NSW
AU-2018-146283	Australia,NSW
AU-2018-138240	Australia,NSW
AU-2018-159618	Australia,NSW
AU-2018-126340	Australia,NSW

Table: Order (480 rows)

Figure 4 - Screenshot for Dataset1 Order table using PowerBI Desktop © Microsoft

New table name:

Evidence of performing the task:  
(Screenshots showing DAX functions/queries used to create the tables from the source table)

Product

The screenshot shows the Power BI Desktop interface with the 'Table tools' ribbon active. The 'Name' field is set to 'Product'. The DAX query in the formula bar is:

```
1 Product = SELECTCOLUMNS('Sample of Sales',  
2 "Product ID", [Product_ID],  
3 "Product Name", [Product_Name]  
4 )
```

The data view below the query shows a table with two columns: 'Product ID' and 'Product Name'. The table contains 360 rows of data, including items like 'Dixon Ticonderoga Core-Lock Colored Pencils' and 'Kensington 6 Outlet MasterPiece HOMEOFFICE Power Control Center'. The 'Fields' pane on the right shows the 'Product' table selected, with 'Product ID' and 'Product Name' listed as fields.

Figure 5 - Screenshot for Dataset1 Product table using PowerBI Desktop © Microsoft

Customer

The screenshot shows the Power BI Desktop interface with the 'Table tools' ribbon active. The 'Name' field is set to 'Customer'. The DAX query in the formula bar is:

```
1 Customer = SELECTCOLUMNS('Sample of Sales',  
2 "Customer ID", [Customer_ID],  
3 "Customer Segment", [Segment]  
4 )
```

The data view below the query shows a table with two columns: 'Customer ID' and 'Customer Segment'. The table contains 480 rows of data, all with the 'Customer' segment. The 'Fields' pane on the right shows the 'Customer' table selected, with 'Customer ID' and 'Customer Segment' listed as fields.

Figure 6 - Screenshot for Dataset1 Customer table using PowerBI Desktop © Microsoft

Table 2 - Target output tables for Dataset 2 (Non-transactional)

New table name:

Evidence of performing the task:  
*(Screenshots showing DAX functions/queries used to create the tables from the source table)*

Product

The screenshot shows the Power BI Desktop interface with the 'Table tools' ribbon active. The 'Name' field is set to 'Product'. The DAX formula bar contains the following code:

```
1 Product = SELECTCOLUMNS('Sample of Products',
2   "Product ID", [Product_ID],
3   "Category ID", [Category_ID],
4   "Product Name", [Product_Name],
5   "Product Price", [Product_Price]
6 )
```

The resulting table view shows the following data:

Product ID	Category ID	Product Name	Product Price
OFF-AP-10002495	1OFS-APL	Acco Smartsocket Table Surge Protector, 6 Color-Coded Ad	43.435
OFF-AP-10003040	1OFS-APL	Fellowes 8 Outlet Superior Workstation Surge Protector w	24.8788
OFF-AP-10004487	1OFS-APL	Kensington 4 Outlet MasterPiece Compact Power Control	49.69758584
OFF-AP-10002118	1OFS-APL	1.7 Cubic Foot Compact "Cube" Office Refrigerators	151.9568
OFF-AP-10001634	1OFS-APL	Hoover Commercial Lightweight Upright Vacuum	2.5752
OFF-AP-10001394	1OFS-APL	Harmony Air Purifier	104.4717521
OFF-AP-10001469	1OFS-APL	Fellowes 8 Outlet Superior Workstation Surge Protector	24.90404273
OFF-AP-10000692	1OFS-APL	Fellowes Mighty 8 Compact Surge Protector	12.11
OFF-AP-10001947	1OFS-APL	Acco 6 Outlet Guardian Premium Plus Surge Suppressor	13.0072
OFF-AP-10001492	1OFS-APL	Acco Six-Outlet Power Strip, 4' Cord Length	6.3788

Table: Product (340 rows)

Figure 7 - Screenshot for Dataset2 Product table using PowerBI Desktop © Microsoft

Category

The screenshot shows the Power BI Desktop interface with the 'Table tools' ribbon active. The 'Name' field is set to 'Category'. The DAX formula bar contains the following code:

```
1 Category = SELECTCOLUMNS('Sample of Products',
2   "Category ID", [Category_ID],
3   "Category", [Category],
4   "Subcategory", [Sub-Category]
5 )
```

The resulting table view shows the following data:

Category ID	Category	Subcategory
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances
1OFS-APL	Office Supplies	Appliances

Table: Category (340 rows)

Figure 8 - Screenshot for Dataset2 Category table using PowerBI Desktop © Microsoft

Table 3 - New data model views for each department

New data model view name:

Evidence of performing the task:  
(Screenshots)

Sales Department

Assessor guidelines:  
The transactional dataset *Sample of Sales* should be aligned with the data types in the Sales Department

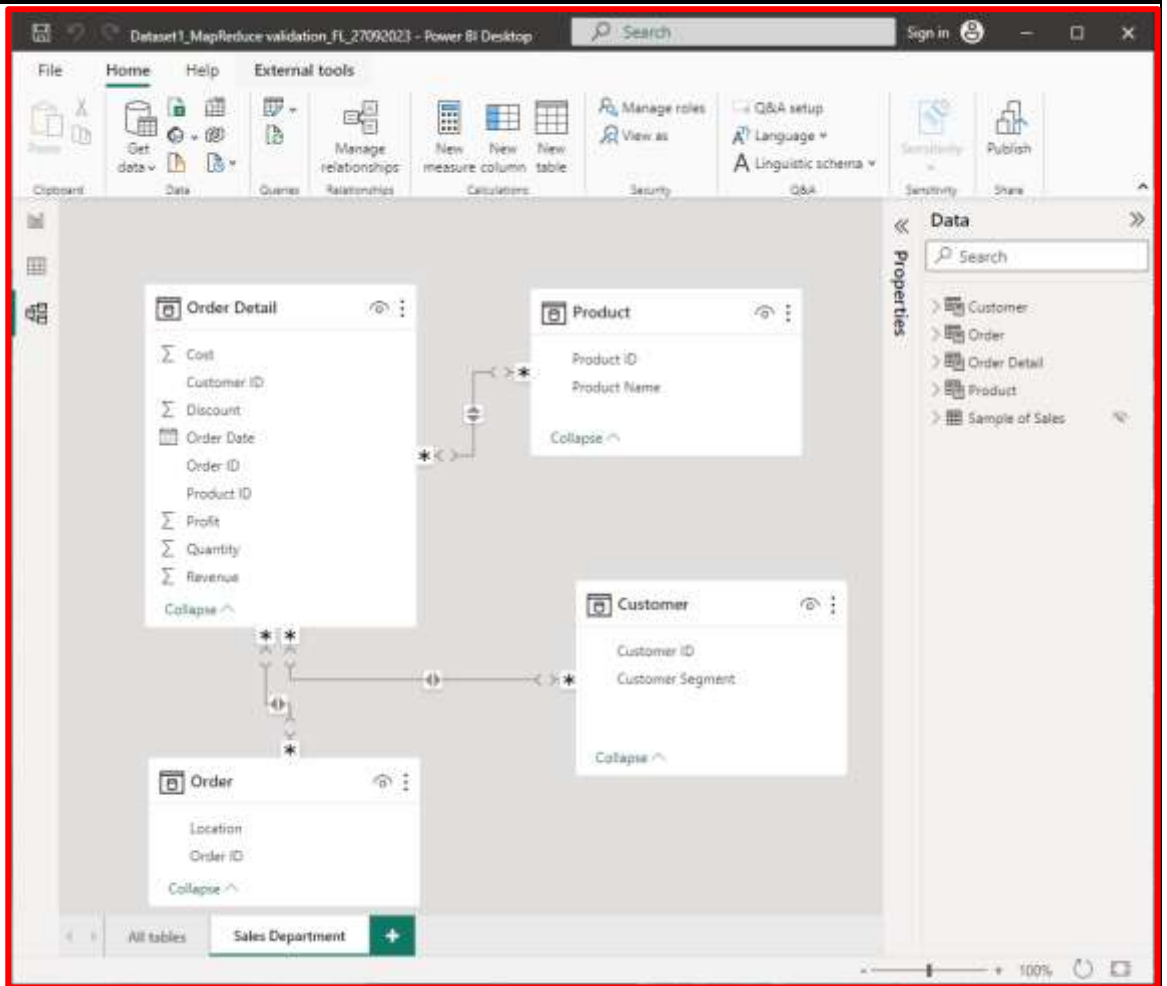


Figure 9 - Screenshot for Dataset1 Data model view using PowerBI Desktop © Microsoft



New data model view name:

Evidence of performing the task:  
(Screenshots)

Production Department

Assessor guidelines:  
Non-transactional dataset *Sample of Products* aligned with the data types in the Production Department

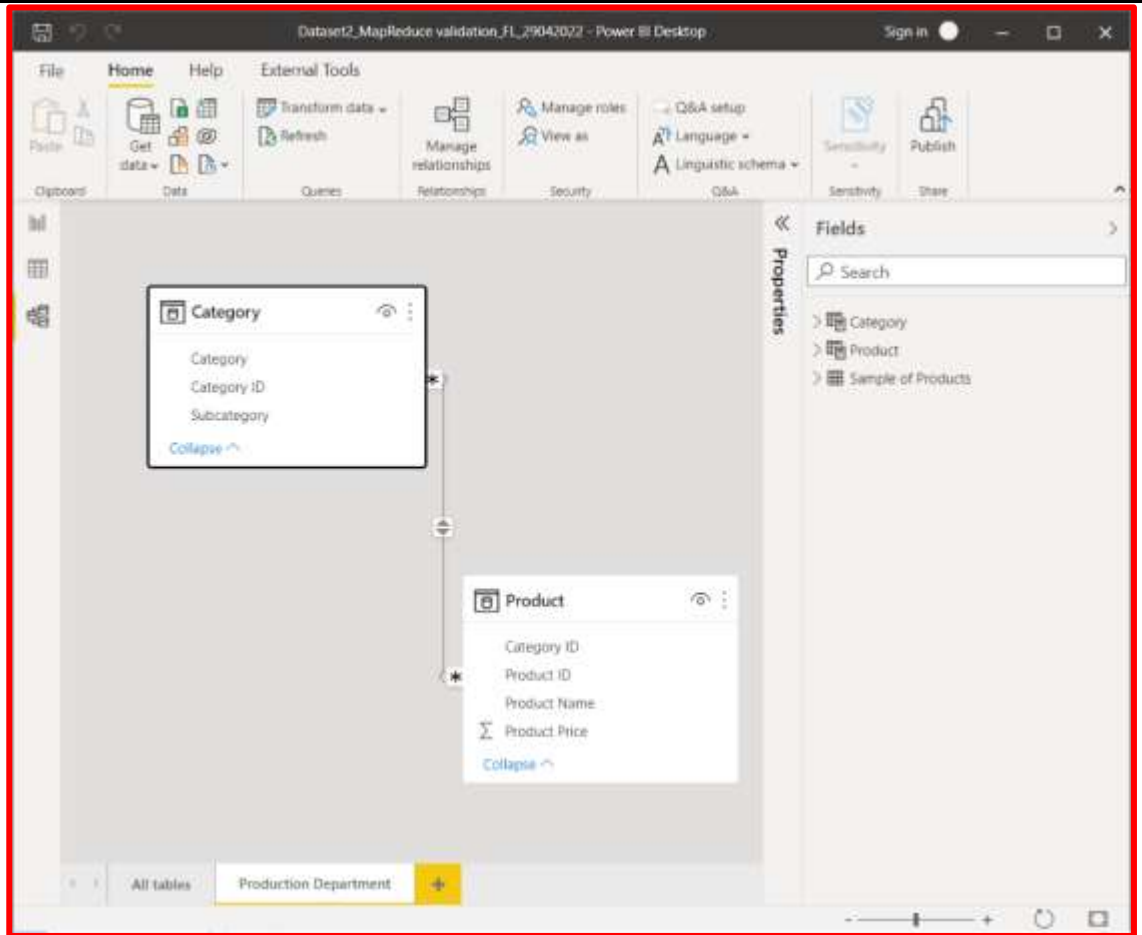


Figure 10 - Screenshot for Dataset2 Data model view using PowerBI Desktop © Microsoft

### C3. Implement data aggregation rules

In this task, you are required to implement data aggregation and segregation rules on the small set of sample data and datasets with appropriate visualisations to display the required data.

#### Instructions:

Do this task using the PowerBI files created in task C1.

Refer to the **reporting requirements** outlined for *Sales* and *Production* departments in the *AUS Retail\_Reporting requirements.pdf* document.

The screenshots you provide as evidence should clearly show the *Report* view tabs that are named appropriately to indicate which type of data that is displayed in the report.

#### Tasks:

C3.1 Do the following using the Dataset 1 PowerBI file:

- rename *Page 1* of the report view tab as '**Sales Report 1**'
- use the correct visualisations and measures to implement the data aggregation rules for the *Sales* department according to the reporting requirements provided.

**Assessor instructions:** Students must do the following to implement data aggregation and segregation rules on the transactional dataset.

- Add a **Card** visual to display: *Cost, Revenue* and *Profit*.
- Add a **Clustered bar chart** visual to display: *Cost, Revenue* and *Profit*.
- Add a **Slicer** visual to filter report data based on yearly and quarterly.
- Add a **Map** visual to display *Revenue* details by *Location*.
- Add a **Donut chart** to display cost, revenue and profit for each product category.
- Add a **Stacked area chart** displaying total revenue and profit for each customer segment.

C3.2 Do the following using the Dataset 2 PowerBI file:

- rename *Page 1* of the report view tab as 'Product Report 1'
- use the correct visualisations and measures to implement the data aggregation rules for the *Production* department according to the reporting requirements provided.

**Assessor instructions:** Students must do the following to implement data aggregation and segregation rules on the non-transactional dataset.

- Add a **Scorecard** visual to display the total number of distinct products.
- Add a **Pie chart** visual to display (the percentage of distinct products in each sub-category).
- Add a **Matrix** visual to list the total number of distinct products in each category and Sub-category.
- Add a **Slicer** visual to filter report data based on the product category.

The student should create a measure called **Distinct Products – to get the distinct no. of products in each category**.

#### Evidence of performing the tasks:

In addition to the screenshots you will include in **Table 9** given below, your assessment submission must include the following documents in the 'Phase 2 – MapReduce validation' sub-folder. The PowerBI work files should contain evidence of implementing data aggregation rules.


- 'Dataset1\_MapReduce validation\_YourNameInitials\_DDMMYYYY'
- 'Dataset2\_MapReduce validation\_YourNameInitials\_DDMMYYYY'.

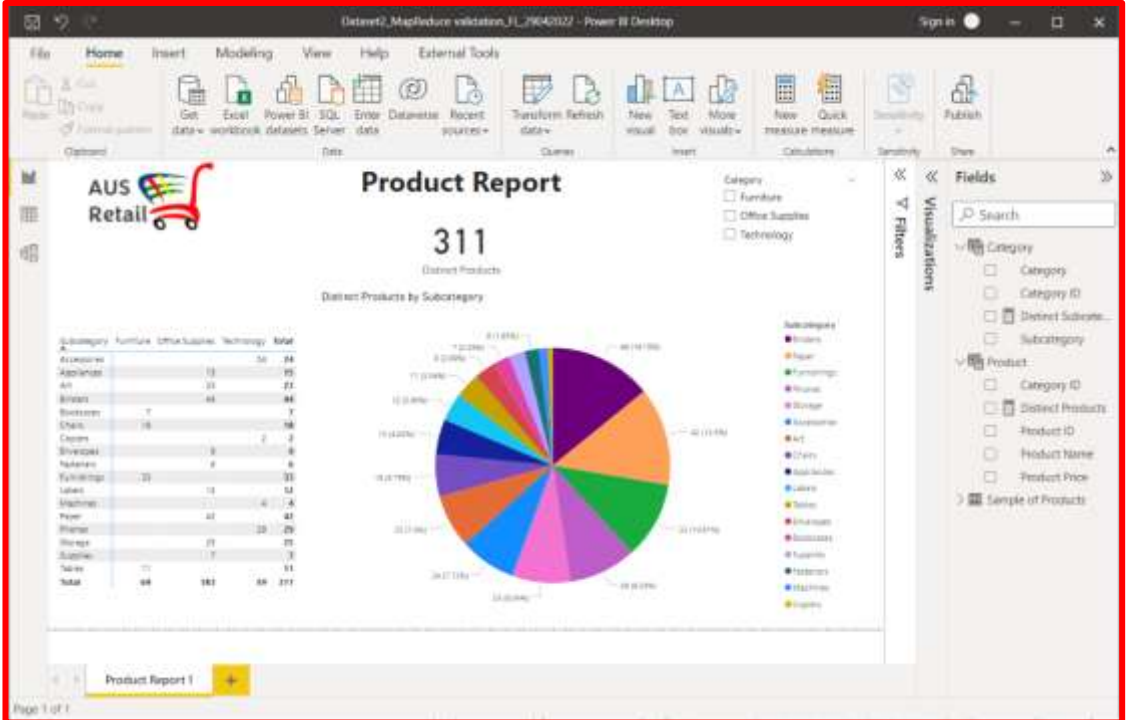
**Assessor instructions:** The student should provide two screenshots showing aggregated and segregated data in each department. Sample screenshots are given below.

Refer to the contents in the **BSBXBD402\_AG\_03\_Project\_Exemplar (student submission folder)** sample work files.



Table 4 - Evidence of performing demonstration task C3

<p><b>Tasks:</b></p> <p><b>Sales Report 1 report:</b>  <i>(The screenshot should show the required visualisations as per the reporting requirements provided)</i></p>	<p><b>Evidence of performing the tasks:</b>  <i>(Screenshots)</i></p>  <p>Figure 11 - Screenshot for Dataset1 Sales report view using PowerBI Desktop © Microsoft</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>Product Report 1 report:</b>  <i>(The screenshot should show the required visualisations as per the reporting requirements provided)</i></p>	 <p>Figure 12 - Screenshot for Dataset2 Product report view using PowerBI Desktop © Microsoft</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## C4. Identify anomalies

In this task, you are required to further evaluate Dataset 1 (Sales) and Dataset 2 (Products) to identify anomalies in the aggregated data.

**Tasks:**

C4.1 Check for anomalies in Dataset 1 (Sales) by doing the following.

- a. Create a new tab in PowerBI Report mode called 'Anomaly Detection (Sales)'
- b. Create line charts to display each of the following time series data from Dataset1 (Sales)
  - **Cost** by Year, Quarter and Month
  - **Revenue** by Year, Quarter and Month
  - **Profit** by Year, Quarter and Month
- c. Use the *Find Anomalies* feature in PowerBI to detect any anomalies in the sales data.
- d. Provide a screenshot of the 'Anomaly Detection (Sales)' tab showing the detected anomalies for each line chart visualisation for **Cost**, **Revenue** and **Profit** in the answer table given below.

C4.2. Check for anomalies in Dataset 2 (Products) by doing the following.

1. Create a new tab in PowerBI Report mode called 'Anomaly Detection (Products).
2. Add Matrix visualisation to list the Product Name details by the number of Distinct Products. Ensure the Product IDs' and Product Price values are grouped within the Product Name lists so that any anomalies can be identified with specific product IDs and their prices.
3. If there is a value greater than 1 displayed for *Distinct Products*, that indicates an anomaly in the data.  
**Business logic:** *One product ID should have one distinct product name with one standard price.*
4. Provide screenshot(s) of the detected anomalies in the Matrix visualisation for in the answer table given below.  
**Note:** Expand the items that have an anomaly in the Matrix visual to obtain further details of the Product IDs and its price.

#### Evidence of performing the tasks:

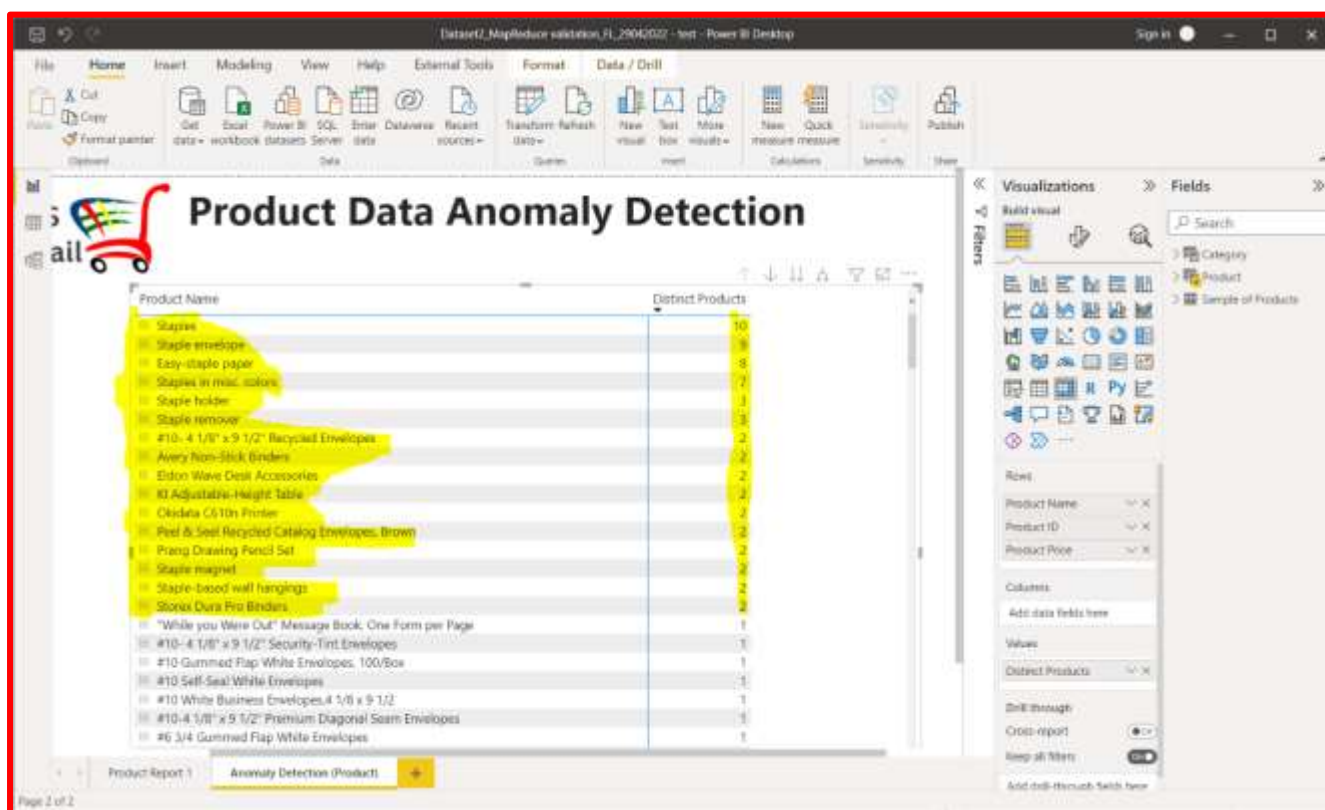
In addition to the screenshots you will include in **Table 10** given below, your assessment submission must include the following documents in the 'Phase 2 – MapReduce validation' sub-folder. The PowerBI work files should contain evidence of identifying anomalies in each dataset.

- 'Dataset1\_MapReduce validation\_YourNameInitials\_DDMMYYYY'
- 'Dataset2\_MapReduce validation\_YourNameInitials\_DDMMYYYY'.

**Assessor instructions:** In addition to the sample screenshots provided in the answer table below, refer to the contents in the **BSBXBD402\_AG\_03\_Project\_Exemplar (student submission folder)** sample work files.



Assessor guidelines: The following variety of anomalies may be detected by the student. The screenshot below shows all possible anomalies in product-related data.



## C5. Consult supervisor to clarify and resolve identified anomalies

In this task, you are required to consult your supervisor to clarify and receive advice on resolving the identified anomalies in the previous task C3.

### Task:

Write a draft email addressed to your supervisor for the purpose of clarifying and obtaining advice on how to resolve the identified anomalies. When drafting the email, you must:

- briefly outline the details of the anomalies detected in the sales data and product data
- include relevant screenshots to clearly indicate the detected anomalies (highlight, circle, to draw attention to specific issues you've identified)
- use clear, specific and industry-related terminology when presenting your validation test results in your email
- use the email template given below.

### Answer: Drafted email to Supervisor

Lastname, Firstname

From: Lastname, Firstname

Sent: Monday, 14 February 2022 10:44 AM

To: Lastname, Firstname

Subject: Sample Email Template

Dear [Name]

Email body goes here.

List Bullet

List Bullet

Kind regards

Firstname Lastname

Your role

[Firstname.Lastname@ausretail.com.au](mailto:Firstname.Lastname@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.*

**Assessor instructions:** The email drafted by the student should indicate that:

- the email was addressed to the supervisor, **Chief Data Officer (CDO), Mia Gonzales**
- it contains all necessary information regarding the data anomalies identified in the previous task for the purpose of consultation to clarify and resolve issues.

Please note that the anomalies detected by the student may have differences depending on the representative sample they've chosen to conduct testing.

A sample answer is provided below.

**Lastname, Firstname**

**From:** Lastname, Firstname

**Sent:** Monday, 05 May 2022 12:53 PM

**To:** Lastname, Firstname

**Subject:** Sample Email Template

---

Hi Mia,

I have conducted some validation checks on the sales and product related datasets and have found the following anomalies in the data.

Refer to the circled areas in the screenshot given below, where it indicates that the total profit has an unexpectedly low figure (minus value) on the 25<sup>th</sup> of November 2020. This is resulting from the unexpectedly high costs on the same day and very low sales figures on the same day. A similar issue occurred on the 02<sup>nd</sup> of October 2021 as well.



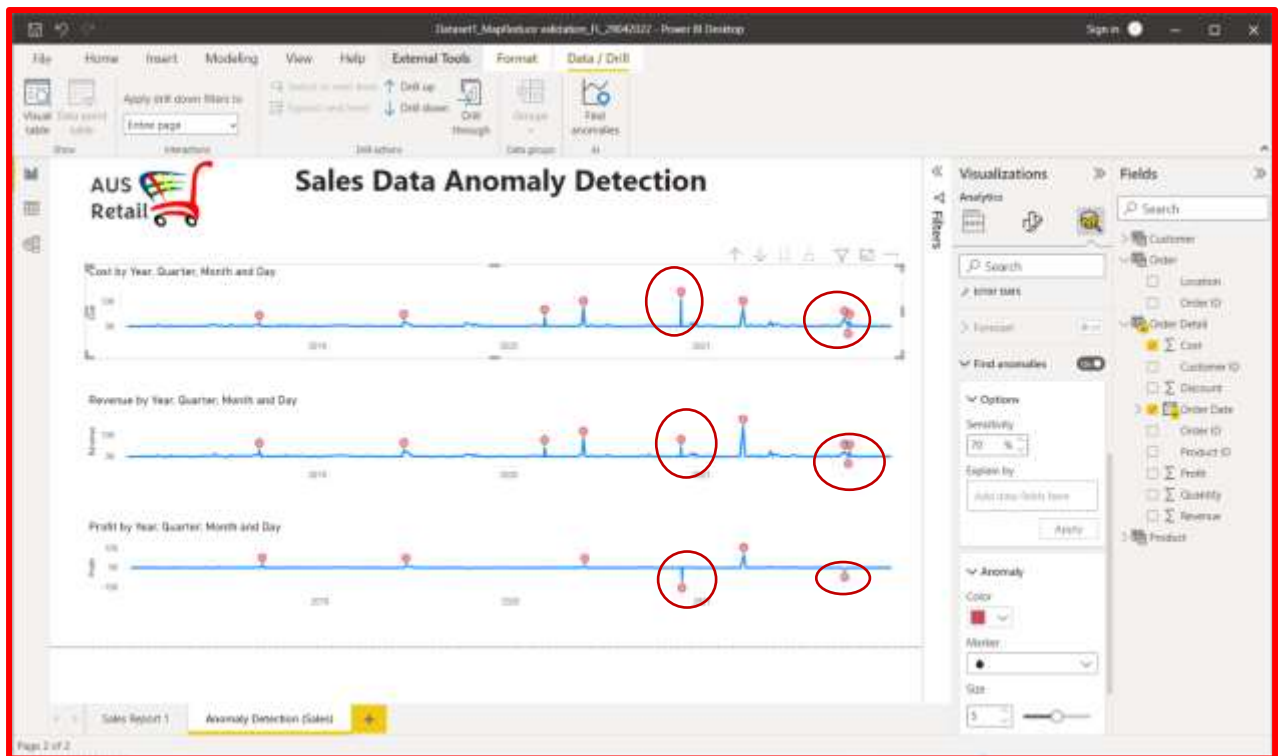


Figure 15 - Anomaly detection in Sales Data using PowerBI Desktop © Microsoft

Concerning the product related data, there are multiple products exist that have the same exact name but different Product IDs and Prices.

The product names that show this anomaly are as follows.:

- Avery Non-Stick Binders
- Staple envelope
- Staple remover

The anomalies found in the product dataset are highlighted in the screenshot given below.



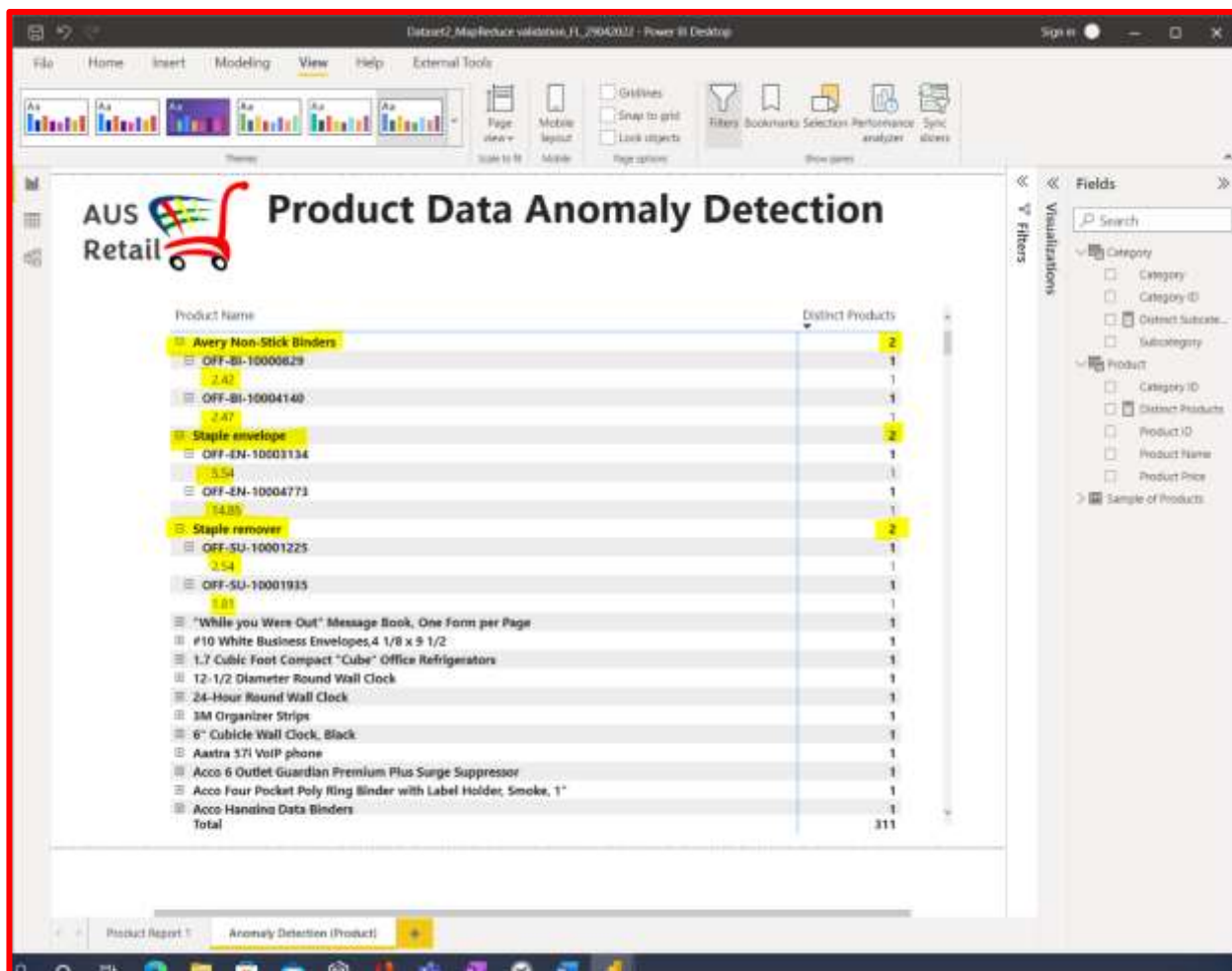


Figure 16 - Anomaly detection in Product Data using PowerBI Desktop © Microsoft

Please do let me know how these anomalies can be resolved before moving forward in the data validation process.

Thanks and kind regards

Firstname Lastname

**Trainee analyst**

[Firstname.Lastname@ausretail.com.au](mailto:Firstname.Lastname@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.*

# Assessment checklist:

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

Part B: Validate assembled or obtained big data sample		
B1	Table 1- Sampling strategy for Dataset 1 (Transactional) Table 2 – Sampling strategy for Dataset 2 (Non-transactional)	<input type="checkbox"/>
B2	Table 3 – Evidence of performing demonstration task B2	<input type="checkbox"/>
B3	Table 4 – Evidence if validating Dataset 1 (Transactional) Table 5 – Evidence if validating Dataset 2 (Non-transactional)	<input type="checkbox"/>
Part C: Validate big data sample process and business logic		
C1	Excel templates, <i>Source to Target Mapping</i> tab: <ul style="list-style-type: none"><li>• <i>AUS Retail_STM&amp;TestCase_Dataset1(Sales)_YourNameInitials_ddmmyyyy.xlsx</i></li><li>• <i>AUS Retail_STM&amp;TestCase_Dataset2(Products)_YourNameInitials_ddmmyyyy.xlsx</i></li></ul>	<input type="checkbox"/>
C2	Table 6 – Target output for Dataset 1 (Transactional) Table 7 – Target output for Dataset 1 (Non-transactional) Table 8 – New data model views for each department (Sales and Production)	<input type="checkbox"/>
C3	Table 9 – Evidence of performing demonstration task C3	<input type="checkbox"/>
C4	Table 10 – Evidence of performing demonstration task C4	<input type="checkbox"/>
C5	Email to Supervisor – email draft for clarification and resolution advice.	<input type="checkbox"/>



**Congratulations you have reached the end of Assessment [3]!**

© UP Education Online Pty Ltd 2022

Except as permitted by the copyright law applicable to you, you may not reproduce or communicate any of the content on this website, including files downloadable from this website, without the permission of the copyright owner.

## 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].