



Assessment Tool

Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Assessment type (☑):

- ☐ Questioning (Oral/Written)
☐ Practical Demonstration
☐ 3rd Party Report
☒ Other – Project/Portfolio *(please specify)*

Portfolio (POR) Assessment

Assessment Resources:

*Marking key available for lecturer via Blackboard.
Students may refer to the lecture material in formulating their answers.*

Assessment Instructions:

*Students must attempt all questions. Answer succinctly using full sentences. At most two paragraphs are expected per answer.
All answers must be at the student's own words – copying generated code or answers from ChatGPT or other AI tools is **strictly** prohibited.*



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Overview

Provides an opportunity to summarise and demonstrate your learning from the last three sessions.

Objective

Demonstrate that you:

1. Created views from the data
2. Used Power Query to store a procedure for transforming the data
3. Extracted features from the data using DAX

Reflective Assessment Questions

All questions refer to the PL-300-based labs 2 – 8.

Q 01. Data Analytics and Feature Extraction

Reflect on a calculated table and calculated column you created.

- Identify the calculation and its purpose.
- What feature about the data did these calculations extract?
- How do these feature extraction procedures aid in data analytics?

Brief answer only

I have created a calculated table named Salesperson.

The calculation was very simple like Salesperson = 'Salesperson (Performance)'

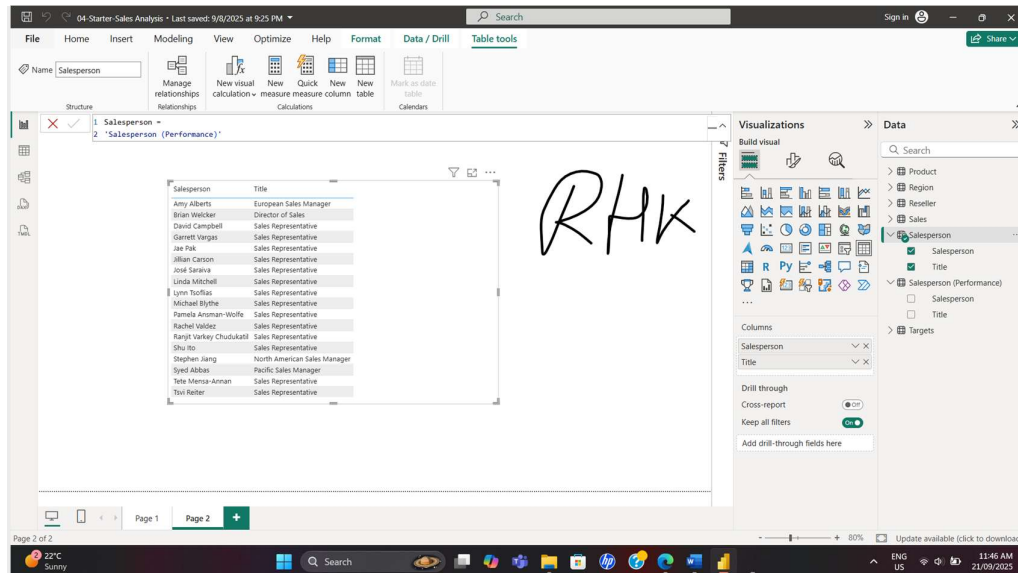
The purpose was to create a copy of the *Salesperson (Performance) table*

This calculation extract the following feature “Salesperson related to sales”

A salesperson's sales can be analysed using the Salesperson table.



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data



I have created a calculated column named Year.

The calculation was Year =

"FY" & YEAR('Date'[Date]) + IF(MONTH('Date'[Date]) > 6, 1)

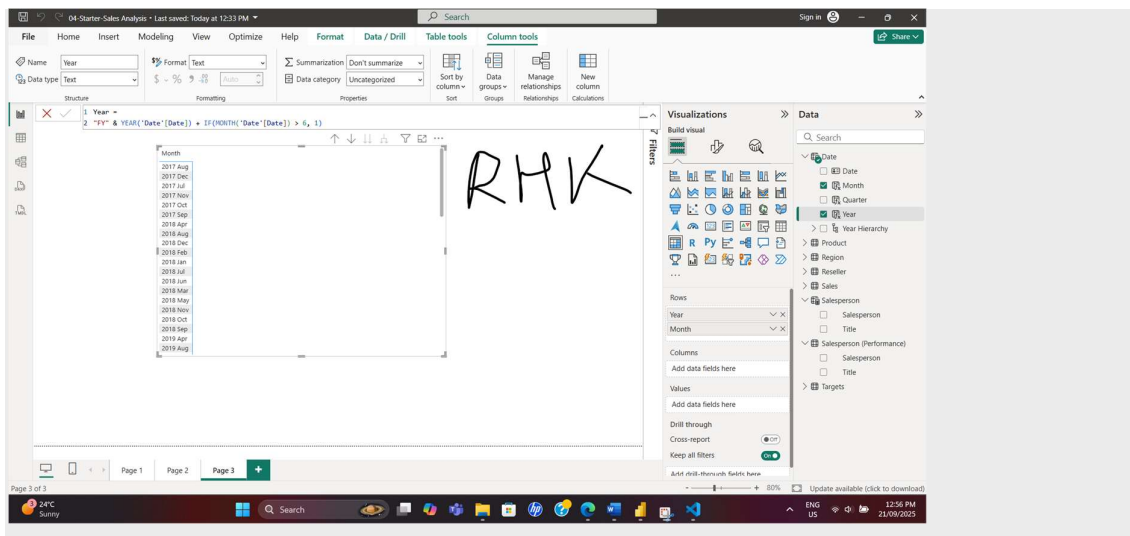
The purpose was to enable filtering and grouping by different time periods

This calculation extracted the date's year value but added one to the year value when the month is after June

It will aid in analysing sales data on yearly basis



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data





Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Q 02. Understanding CALCULATE() Function

- Can you explain the purpose and advantage of using the CALCULATE() function in your DAX measures?
- How did it help you manipulate the filter context, especially for time-bound data?

The CALCULATE function is a powerful DAX (Data Analysis Expressions) function that enables us to modify the filter context to evaluate an expression, allowing for flexible and complex calculations by applying, overwriting, or removing filters.

Calculate function helped me to manipulate the filter context for time-bound data while I was calculating the YoY sales growth. The formula was as follows:

Sales YoY Growth =

VAR SalesPriorYear =

```

CALCULATE(
    SUM(Sales[Sales]),
    PARALLELPERIOD(
        'Date'[Date],
        -12,
        MONTH
    )
)

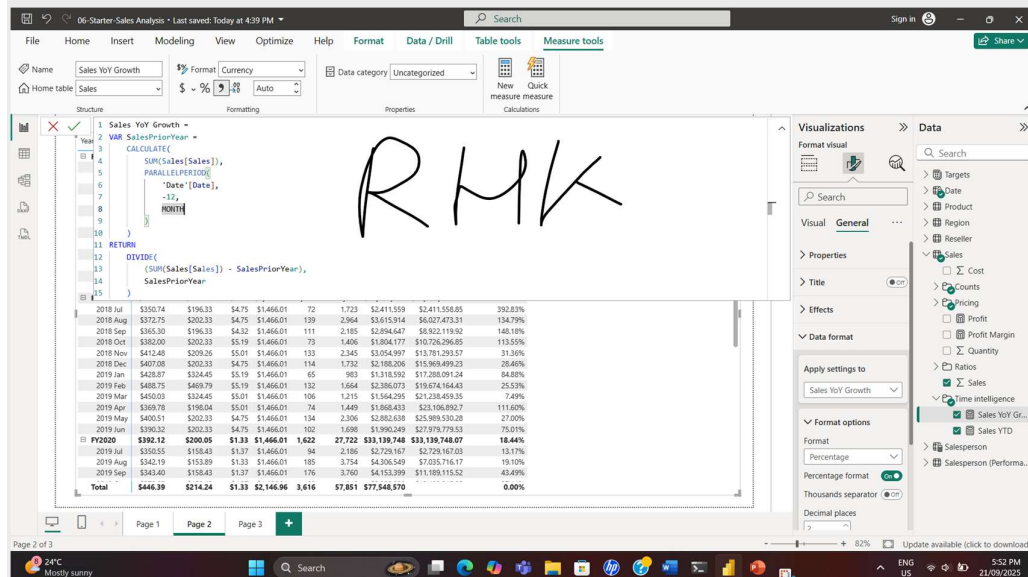
```

RETURN

Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

```
DIVIDE(  
    (SUM(Sales[Sales]) - SalesPriorYear),  
    SalesPriorYear  
)
```

By using this formula I calculated the sum of the Sales column in a modified filtered context which, used the PARALLELPERIOD function to shift 12 months back from each date in that filtered context





Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Q 03. Data Source Characteristics

- Provide example characteristics of data sources you used? Consider data types, update frequency, and quality.
- Give an example of how these characteristics impacted your data analysis and reporting?

I have basically used excel/CSV and SQL/Relational Database as the data sources.

Characteristics of excel file: Structured data but mostly implicit (No strict data type) data types, excel data could be updated very frequently and needed to load each time the data file changes. The quality can vary and might needs transformation (formatting) to enhance the quality.

Characteristics of SQL/Relational Database data: Structured data with explicit data types for example varchar, integer, float, double, datetime, etc. and data updates & quality can be ensured through automated processes and integration.

SQL is appropriate for dynamic, often-changing datasets as it pulls data from the database in real time and it also permits scheduled refreshes.

Power BI scans for updates in OneDrive or SharePoint roughly every hour, with the option to schedule refreshes.

I performed some analysis on **AdventureWorks** data and loaded it from SQL server.

I have also used **ResellerSalesTargets.csv** file to perform data analysis and reporting.

I have checked column quality, column distribution & column values, on the **View** ribbon tab, from inside the **Data Preview** group.



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

The column quality feature allowed me to easily determine the percentage of valid, error, or empty values found in columns while column distribution helped me to know the no of unique and distinct values for a distribution. Column profile showed the column values. Column header provided me the data type for the column.

These characteristics help me to determine my strategy for the analysis as well as in determining if any transformative action is required prior to the analysis.



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Q 04. Data Transformation and Feature Analysis

- Reflect on the Power Query (transform data) procedure you stored to filter out data and the DAX measure you created for month-over-month growth.
- How do these procedures and measures aid in data analysis and business decision-making?
- Give an example of a feature you extracted (calculated measure) in the activities. How did creating the measure aid in analysis?

I have used the following formula while doing DAX measure for YoY growth to convert the type of the Sales column.

`Table.TransformColumnTypes(#"Renamed Columns",({{"Quantity", Int64.Type}, {"Unit Price", Currency.Type}, {"Sales", Currency.Type}, {"Cost", Currency.Type}}))`

I have also renamed the column SalesAmount to Sales by using the following formula:

`= Table.RenameColumns(#"Changed Type",({{"OrderQuantity", "Quantity"}, {"UnitPrice", "Unit Price"}, {"SalesAmount", "Sales"}}))`



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

This transformation helped me while creating the DAX measure for YoY growth as Sales represents the SalesAmount more precisely and the currency type helped me to represent the sales amount in terms of the actual currency.

After the transformation I created the following measure to calculate the yearly sales growth in a modified filtered context which, used the PARALLELPERIOD function to shift 12 months back from each date in that filtered context by using the following formula:

Sales YoY Growth =

VAR SalesPriorYear =

```

CALCULATE(
    SUM(Sales[Sales]),
    PARALLELPERIOD(
        'Date'[Date],
    
```



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

```

-12,

MONTH

)

)

RETURN

DIVIDE(

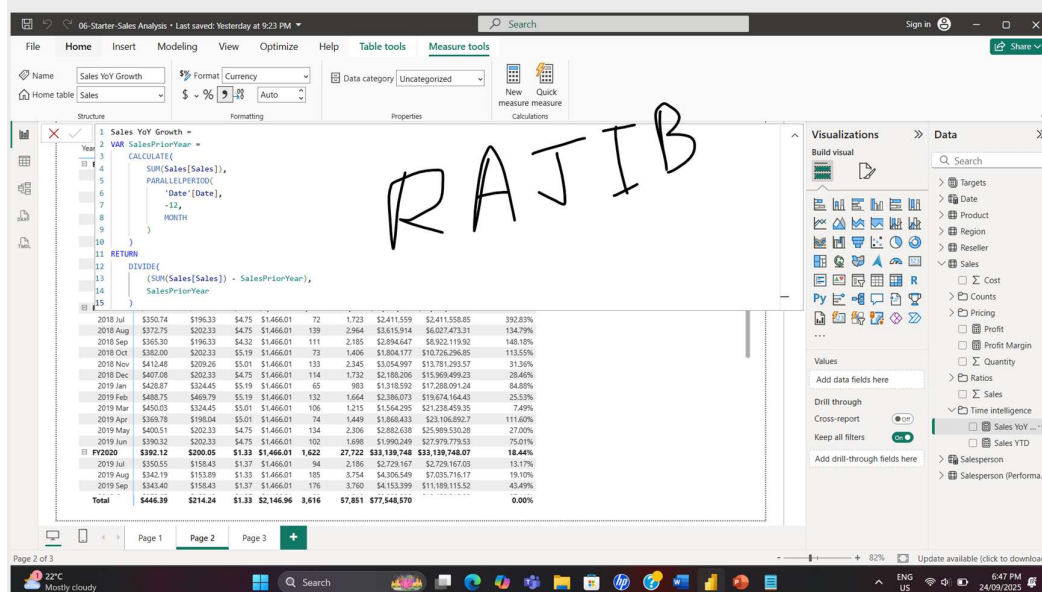
(SUM(Sales[Sales]) - SalesPriorYear),

SalesPriorYear

)

```

This measure provided the yearly sales growth which helps the stakeholders to make decisions regarding the business based on time intelligence information extracted from the measure or get further information regarding which factors boosted the yearly sales growth.





Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Q 05. Provide a brief explanation for the following terms:

Visualisation

It is also sometimes called a visual, and it is a visual representation of data, such as a color-coded map, chart, or other visually appealing creation. There are numerous visualization kinds available in Power BI, and new ones are constantly being added. A variety of various visualizations made with the Power BI service are displayed in the following image.

Datasets

A dataset is a collection of data from a variety of sources that may be combined and filtered to produce a special set of data that is used to make visualizations.

Dashboard

A collection of images from one page that you can distribute to other people. It's usually a chosen collection of images that offer a brief overview of the information or narrative you're attempting to convey.

Tiles

A single dashboard or report visualization. It is the rectangular container that contains a single image.

Q 06. Describe the difference between digitally and manually entered data. List at least two advantages of the former and one advantage of the latter.

While automated solutions give speed and efficiency in the database along with flexibility and adaptation, manual data processing delivers control and accuracy with papers.

When it comes to error rate, manual data entry is more likely to make mistakes.

The speed of processing is faster with automated data entry.

Automated Data Entry involves initial setup cost like software, hardware and training expense but manual data entry don't have that extra cost.



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Q 07. In Power BI or SQL, how would you determine if values for a given column are between a valid upper and lower threshold?

In Power BI we can use the following DAX formula to determine if values for a given column are between a valid upper and lower threshold:

OutlierFlag =

VAR MeanValue = CALCULATE(AVERAGE('Table'[Value]))

VAR StdDev = CALCULATE(STDEV.P('Table'[Value]))

VAR ThresholdHigh = MeanValue + 2 * StdDev

VAR ThresholdLow = MeanValue - 2 * StdDev

RETURN

IF('Table'[Value] > ThresholdHigh || 'Table'[Value] < ThresholdLow, "Outlier", "Normal")

In the above formula:

MeanValue: is the Average of the selected column.

StdDev: is the Standard deviation of selected the column.

Thresholds: Defines the upper and lower bounds (± 2 * standard deviations).

OutlierFlag: Labels each row as "Outlier" or "Normal".

If we consider the example of Sales column of Sales table the formula should be as follows:



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

SalesOutlierFlag =

VAR MeanSales =

```
CALCULATE(
    AVERAGE('Sales'[Sales])
)
```

VAR StdDevSales =

```
CALCULATE(
    STDEV.P('Sales'[Sales])
)
```

VAR UpperBound = MeanSales + 2 * StdDevSales

VAR LowerBound = MeanSales - 2 * StdDevSales

RETURN

IF('Sales'[Sales] > UpperBound || 'Sales'[Sales] < LowerBound, "Outlier", "Normal")

Use Cases of outliers

Finding irregularities in sensor data, sales, or expenses.

Highlighting odd dashboard patterns.

Visualization Tip: To draw attention to outliers in visualizations, we can use conditional formatting or filters.



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

Q 08. Examine the following explanation of what a sign off document is in the context of data analytics:

In the context of data analytics, a sign-off document is a formal, written acknowledgment that a project, report, or analysis meets the specified requirements and has been reviewed and approved by the stakeholders or authorities involved. This document is crucial for several reasons:

1. **Confirmation of Requirements Fulfillment:** It confirms that the delivered data analytics work fulfills the predefined requirements or objectives. This might include aspects like data quality, analysis accuracy, compliance with data privacy laws, and alignment with the project's goals.
2. **Accountability and Responsibility:** The sign-off document typically includes the names and signatures of the parties responsible for the project. This clarifies accountability and ensures that key stakeholders have reviewed and are satisfied with the work.
3. **Formal Closure of Projects:** It serves as a formal closure of a phase or the entire project, indicating that the work has been completed and is ready to move to the next stage, be it implementation, further analysis, or archiving.
4. **Reference for Future Projects:** The document can serve as a reference point for future projects, showcasing what was agreed upon, delivered, and how it was received by the stakeholders.
5. **Legal and Compliance Reasons:** In some cases, especially where data is sensitive or subject to regulatory compliance, a sign-off document can be a legal requirement to ensure all parties are aware of and agree to how the data is used and presented.
6. **Feedback and Improvement:** It often includes feedback or comments from the stakeholders, providing insights into areas of strength and potential improvement for future projects.

This document might include details like the scope of the data analyzed, methodologies used, key findings, limitations, and any assumptions made during the analysis.

- a. In your own words, what is the purpose of a sign off document in a data analytics context?

The purpose of sign-off document, in data analytics, is to provide a formal written acknowledgement that a project, report, or study satisfies the requirements and has been examined and accepted by the relevant authorities or stakeholders.



Qualification national code and title	ICT50220-SS-A Applied Data Analysis Skillset
Unit/s national code/s and title/s	ICTPRG431 Apply query language in relational databases ICTDAT401 Evaluate organisational compliance with data ethics legislation ICTDAT402 Clean and verify data

- b. Thinking of your particular work context, which two justifications for a sign-off document are most relevant to you and why?

I think the following two justifications are most relevant to me because as a data analysts I will be closely working with the data and I must make sure that these two justifications are met.

Confirmation of Requirements Fulfillment: While working at Synergy we need to make sure that fulfilling the requirements for which we are collecting our customers data

Legal and Compliance Reasons: We also need to make sure we follow the privacy regulations regarding our customer data