RESKILLING LAB
DT-03 TRAINING & SELF-LEARNING PACK

DATA ANALYTICS

From Spreadsheet User to Junior Data Analyst

Beginner → Intermediate | Recommended learning time: 50–70 hours
Expanded Learner Handbook | Version 1.0 | September 2026

How to Use This Handbook

DT-03 is a practical learner handbook for self-learning, blended learning and trainer-supported delivery. It develops the core analytical workflow from business question to data acquisition, cleaning, exploration, modelling, visualisation, interpretation and communication. The pathway uses spreadsheets and Power BI as accessible practice environments while introducing transferable concepts that also apply to SQL, Python and other analytics tools.

The purpose is not to teach button-clicking alone. Learners should understand why an analytical step is performed, what assumptions it introduces, how data quality affects conclusions and how to communicate uncertainty. All exercises should use fictional, open, anonymised or explicitly authorised data.

1. Pathway Overview

Item	DT-03 Specification
Pathway	Data Analytics
Target level	Beginner → Intermediate
Recommended workload	50–70 hours
Target groups	Employees; SME staff; VET learners; jobseekers; career changers; administrative, finance, operations and junior digital staff
Entry requirements	Basic digital literacy; familiarity with spreadsheets is helpful but not mandatory
Possible target roles	Junior Data Analyst; Business Data Analyst; Reporting Assistant; BI Assistant; Data-enabled Business Professional
Core tools	Spreadsheet software; Power BI Desktop; optional SQL/Python exposure
Main evidence	Data-quality audit; cleaned dataset; exploratory analysis; KPI definitions; dashboard; insight report; capstone analytics project

1.1 Learning Outcomes

- Explain the role of data analytics in evidence-based decision-making.
- Distinguish descriptive, diagnostic, predictive and prescriptive analytics.
- Translate a business problem into answerable analytical questions and measurable KPIs.
- Recognise data types, structures, measurement scales and common sources of bias.
- Import, profile, clean, transform and document data.
- Apply basic descriptive statistics and interpret distributions, variation and relationships.
- Use spreadsheets for structured analysis, formulas, pivot tables and quality checks.
- Explain relational data concepts and use basic SQL logic conceptually.
- Create a simple analytical data model and understand table relationships.
- Design clear, accessible and non-misleading charts and dashboards.
- Use Power BI to connect, transform, model and visualise data.
- Interpret results without confusing correlation with causation.
- Apply privacy, security, ethics and data-governance principles.
- Use generative AI as an analytical assistant while independently verifying outputs.
- Complete an end-to-end SME data analytics project and communicate actionable insights.

1.2 Competence and Career Context

Data literacy is part of modern digital competence. DigComp 2.2 treats information and data literacy as a core competence area covering browsing, evaluating and managing data, information and digital content. In professional analytics, the data analyst role typically combines business requirements, data preparation, modelling, visualisation, analysis and communication.

Microsoft's current Power BI data-analyst pathway similarly organises work around preparing data, modelling data, visualising and analysing data, and managing and securing analytics assets. DT-03 remains vendor-neutral at the conceptual level, but Power BI is used as a practical environment because it supports the complete beginner-to-intermediate workflow.

2. Initial Skills Check

#	Statement	0 No	1 Partly	2 Yes
1	I can turn a business problem into a clear analytical question.	☐	☐	☐
2	I understand rows, columns, records, variables and data types.	☐	☐	☐
3	I can identify missing values, duplicates and inconsistent categories.	☐	☐	☐
4	I can use spreadsheet formulas for basic calculations.	☐	☐	☐
5	I can create and interpret pivot tables.	☐	☐	☐
6	I understand mean, median, percentage, rate and standard deviation at a basic level.	☐	☐	☐
7	I can choose an appropriate chart for a business question.	☐	☐	☐
8	I understand why correlation does not automatically imply causation.	☐	☐	☐
9	I can explain the difference between a dimension and a measure.	☐	☐	☐
10	I can connect and transform data in a BI tool.	☐	☐	☐
11	I can define a KPI with numerator, denominator, time period and target.	☐	☐	☐
12	I can identify privacy, bias or quality risks in a dataset.	☐	☐	☐
13	I can explain an analytical finding in plain business language.	☐	☐	☐
14	I verify AI-generated calculations, formulas and interpretations before using them.	☐	☐	☐

Indicative formative interpretation: 0–10 Foundation needed; 11–18 Developing; 19–23 Functional; 24–28 Ready for intermediate practice. This is not a formal qualification.

3. Data Analytics: From Business Question to Decision

Data analytics is the systematic process of examining data to answer questions, identify patterns, measure performance and support decisions. The analytical process begins before a dataset is opened. A useful analysis starts with a decision context: what problem exists, who needs the answer, what action may follow and what evidence would be sufficient.

Descriptive analytics asks what happened. Diagnostic analytics asks why it happened. Predictive analytics estimates what may happen under defined assumptions. Prescriptive analytics considers what action might

be taken. These categories overlap, but they help prevent a common mistake: using descriptive data to make unsupported causal or predictive claims.

A business question should be specific enough to guide data selection. 'How are sales?' is weak. 'How did monthly revenue and gross margin change by product category during the last 12 months, and which categories account for the largest decline?' is analytically useful because it identifies measures, dimensions, period and decision focus.

An analyst creates value by reducing uncertainty, not by producing the largest possible dashboard. The strongest analysis connects a clearly defined question to trustworthy data, transparent methods and an interpretation that a stakeholder can act upon.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

4. Data, Variables and Measurement

A dataset is a structured representation of observations. Rows often represent records or events; columns represent variables or attributes. Data can be numerical, categorical, textual, temporal, geographic, binary or more complex. Correctly identifying data type affects calculations, sorting, aggregation and visualisation.

Measurement scales matter. Nominal variables represent categories without inherent order. Ordinal variables have an order but not necessarily equal intervals. Interval scales have meaningful differences but no true zero, while ratio scales have a meaningful zero and support ratio comparisons. In everyday business analytics, the practical lesson is to avoid calculations that the measurement scale cannot support.

Dimensions describe the context of an observation, such as product, region, customer segment or date. Measures represent quantities that can often be aggregated, such as revenue, quantity or cost. This distinction becomes especially important in BI modelling.

Granularity describes the level of detail represented by each record. Mixing daily and monthly data, customer-level and transaction-level data, or product and product-category data without understanding granularity can produce duplicate counting and misleading results.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

5. Data Sources, Collection and Sampling

Data may come from operational systems, spreadsheets, databases, surveys, sensors, web analytics, public datasets, APIs or third-party providers. Each source has strengths and limitations. Operational data may be detailed but designed for transactions rather than analysis; survey data may capture attitudes but suffer from response bias.

Primary data is collected for the current purpose; secondary data already exists for another purpose. Reusing existing data is efficient, but the analyst must check definitions, collection methods, time periods and whether the source population matches the analytical question.

A population is the complete group of interest; a sample is the subset observed. Sampling methods influence representativeness. Convenience samples are easy to collect but may be biased. Random or stratified approaches can improve representativeness when implemented correctly.

Selection bias, non-response bias, survivorship bias and measurement bias can distort results even when calculations are technically correct. Data quality therefore begins with understanding how the data was produced.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

6. Data Quality and Data Profiling

Data quality is multidimensional. Accuracy concerns whether values reflect reality. Completeness concerns missing required values. Consistency asks whether definitions and formats agree across records or systems. Timeliness concerns whether data is sufficiently current. Validity checks whether values follow rules. Uniqueness concerns unwanted duplicates.

Data profiling is the systematic examination of a dataset before deeper analysis. Useful profiling includes row counts, data types, missing-value rates, unique-value counts, minimum and maximum values, frequency distributions and detection of suspicious outliers.

Cleaning should not be silent. If an analyst removes records, replaces missing values, changes categories or caps outliers, the decision should be documented. A reproducible cleaning log makes the analysis auditable and helps another person understand how the final dataset differs from the source.

Missing data requires judgement. Deleting incomplete rows may bias the sample; replacing values with averages may artificially reduce variation. Sometimes the correct treatment is to leave the value missing and explicitly report the limitation.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.

- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

7. Spreadsheet Analytics Foundations

Spreadsheets remain a major analytics tool because they combine data entry, formulas, tables, pivoting and visualisation in an accessible environment. Their flexibility is also a risk: manual edits, hidden formulas and inconsistent ranges can make errors difficult to detect.

Good spreadsheet practice separates raw data, transformation logic, calculations and presentation. Source data should be preserved. Structured tables, meaningful names, consistent data types and documented formulas improve reliability.

Core analytical functions include arithmetic, percentages, conditional logic, lookup functions, text and date functions, aggregation and error handling. The objective is not memorising formulas but selecting a calculation that correctly represents the business definition.

Pivot tables summarise data by dimensions and measures. They are especially useful for exploring totals, averages, counts and percentages across categories and time. Analysts should check whether the aggregation makes sense; summing a percentage or averaging an already aggregated average can be misleading.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

8. Descriptive Statistics and Distribution

Descriptive statistics summarise the main characteristics of data. The mean is the arithmetic average; the median is the middle value; the mode is the most frequent value. The mean is sensitive to extreme values, while the median can better represent a skewed distribution.

Range, variance, standard deviation and interquartile range describe dispersion. Two groups can have the same average but very different variability. In operational analysis, variability may be as important as the average because unstable processes create planning risk.

Percentiles show the position of a value relative to a distribution. The 90th percentile, for example, is a value at or below which approximately 90% of observations fall. Percentiles are useful for response times, delivery times and performance distributions.

An outlier is an unusually distant observation. It may indicate an error, a rare but legitimate event or an important business signal. Outliers should be investigated rather than automatically removed.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

9. Relationships, Correlation and Causation

Analysts often examine whether variables move together. Correlation measures the strength and direction of association between variables under specific assumptions. A positive correlation means higher values of one variable tend to occur with higher values of another; a negative correlation means the opposite.

Correlation does not establish causation. Two variables may move together because one influences the other, because a third factor influences both, because of selection effects or simply by chance. Causal claims require stronger designs and assumptions than ordinary descriptive analysis.

Confounding occurs when another variable affects both the apparent cause and outcome. Seasonality, customer mix, region or economic conditions can create misleading associations if not considered.

Analysts should communicate relationships precisely: 'is associated with' is different from 'causes'. This distinction is essential in business, policy and educational data.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

10. KPI Design and Performance Measurement

A Key Performance Indicator is a measure selected to track progress toward an objective. Not every metric is a KPI. A good KPI has a clear business purpose, precise definition, owner, data source, frequency, target and interpretation.

Ratios and rates often provide more useful comparisons than raw totals. Complaint count may rise simply because sales increased; complaints per 1,000 orders may reveal whether service quality actually changed.

Targets need context. Comparing actual performance with prior period, budget, benchmark or threshold can produce different interpretations. Dashboards should make the comparison basis explicit.

Good KPI systems avoid metric overload and gaming. When people are strongly rewarded for one measure, behaviour may optimise the measure rather than the underlying objective. Balanced indicators and qualitative review can reduce this risk.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

11. Data Visualisation Principles

Data visualisation converts quantitative information into visual form so that patterns, comparisons and exceptions can be perceived efficiently. Chart selection should follow the analytical question rather than decorative preference.

Bar charts are effective for comparing categories. Line charts are generally suitable for trends over time. Scatter plots reveal relationships between two numerical variables. Histograms show distributions. Maps should be used only when geography is analytically relevant.

Visual encodings differ in perceptual accuracy. Position and length are generally easier to compare precisely than area or angle. Three-dimensional effects, excessive decoration and unnecessary colour can obscure differences and create false emphasis.

Axes, scales and baselines affect interpretation. Truncated axes can exaggerate small differences. Dual axes can imply relationships that are not meaningful. Visualisation should help the viewer understand the data rather than persuade through distortion.

Accessibility matters. Do not rely only on colour to communicate categories. Use readable labels, sufficient contrast, logical reading order and alternative descriptions where required.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

12. Data Storytelling and Communicating Insight

An analytical report should distinguish data, interpretation and recommendation. Data states what was observed. Interpretation explains what the pattern may mean. Recommendation proposes an action based on evidence and business context.

A useful insight is specific, evidence-based and relevant to a decision. 'Sales fell' is an observation. 'Revenue fell 8% year-on-year, with 70% of the decline concentrated in two product categories after March' is more actionable.

Storytelling does not mean dramatizing the data. It means organising information around a question, presenting the most relevant evidence, explaining uncertainty and guiding the audience through the reasoning.

Executives often need a concise summary with key indicators, exceptions, implications and required decisions. Analysts should be able to explain the same analysis at different levels of technical detail without changing the underlying evidence.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

13. Relational Data and SQL Concepts

Relational databases organise data into tables connected through keys. A primary key uniquely identifies a record. A foreign key links a record to another table. Relationships reduce duplication and support consistent data management.

SQL is a language for querying relational databases. Core concepts include SELECT for choosing fields, WHERE for filtering, GROUP BY for aggregation, ORDER BY for sorting and JOIN for combining related tables.

Joins require careful reasoning. An inner join returns matching records; a left join preserves all records from the left table and adds matches from the right. Incorrect joins can duplicate records and inflate totals.

Even when learners use Power BI rather than writing SQL, relational thinking is essential because data models depend on table relationships, keys and granularity.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

14. Data Modelling and the Star Schema

A data model organises tables and relationships so that analytical calculations behave correctly. In a star schema, a central fact table contains events or measurements and connects to dimension tables describing entities such as date, product, customer or region.

Fact tables typically contain foreign keys and numerical measures at a defined grain. Dimension tables contain descriptive attributes. Keeping dimensions separate improves consistency and simplifies filtering.

Relationships have cardinality, such as one-to-many. Incorrect relationship direction or duplicate keys can create ambiguous filters and inaccurate results. Before building visuals, the analyst should confirm the grain and key structure of each table.

Semantic models add business meaning through measures, hierarchies, formats and relationships. A well-designed model allows many reports to use the same trusted definitions.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

15. Power Query and Data Transformation

Power Query provides a repeatable transformation process for connecting to sources, changing data types, removing errors, splitting or merging columns, combining tables, pivoting and unpivoting data, and preparing information for modelling.

The key advantage over manual spreadsheet editing is reproducibility. Transformation steps can be refreshed when new data arrives. This reduces repeated manual work and makes the preparation logic visible.

Data types should be set deliberately. A date stored as text will not behave like a date; a numerical identifier should not necessarily be summed. Profiling tools can reveal errors, nulls and unexpected distributions before loading.

Transformations should preserve traceability. Rename steps meaningfully, avoid unnecessary complexity and document business rules that are not obvious from the technical operation.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

16. Measures, Calculated Columns and DAX Concepts

In BI models, calculated columns are evaluated for rows and stored in the model, while measures are evaluated dynamically in the filter context of a visual or query. This distinction affects performance and analytical behaviour.

DAX, or Data Analysis Expressions, is the formula language used in Power BI semantic models. Common measures include SUM, COUNT, DISTINCTCOUNT, AVERAGE and ratio calculations. More advanced DAX modifies filter context to answer questions such as year-to-date performance or share of total.

Filter context is central to understanding BI calculations. A measure can return different values depending on selected date, region, product or other filters. Analysts should test measures under multiple filter combinations.

A measure should represent a documented business definition. If 'Active Customer' or 'On-time Delivery' is not clearly defined, a technically correct formula can still produce organisational disagreement.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

17. Dashboard and Report Design in Power BI

A dashboard or report should be designed around user decisions. Start by identifying the audience, questions and frequency of use. Only then select visuals and layout.

Overview pages should prioritise a small number of important indicators and exceptions. Detail pages can support exploration. Filters and slicers should be useful but not force the user to configure the report before understanding it.

Consistency improves comprehension. Use stable terminology, number formats, date conventions and visual conventions. Titles should describe what a visual shows and, when useful, communicate the main finding.

Interactive reports can encourage exploration through filtering, drill-through and tooltips. Interaction should not hide the analytical logic. Users need to understand which filters are active and what population a visual represents.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

18. Privacy, Ethics, Bias and Data Governance

Data analytics can affect people even when no machine-learning model is used. Decisions based on customer, employee, learner or citizen data can create privacy, fairness and discrimination risks. Analysts should ask whether the data is necessary, lawfully available, sufficiently protected and appropriate for the intended use.

De-identification reduces direct identification but may not eliminate re-identification risk, especially when datasets contain rare combinations of attributes. Analysts should avoid sharing row-level personal data when aggregated information is sufficient.

Bias can enter through collection, missing data, historical practices, category definitions, sampling and interpretation. A dashboard may accurately reproduce biased historical data without being fair as a basis for future decisions.

Governance defines ownership, access, definitions, quality responsibilities, retention and approved use. A data dictionary and KPI catalogue are simple but powerful governance tools for SMEs.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

19. AI-Assisted Data Analytics

Generative AI can assist with formula explanations, SQL drafts, data-cleaning suggestions, chart ideas, documentation and interpretation prompts. It can accelerate work, especially for learners, but it should not become an unverified analytical authority.

AI-generated formulas and code can be syntactically plausible but logically wrong. The analyst must test results on known cases, inspect assumptions and compare outputs with independent calculations.

Sensitive datasets should not be uploaded to unapproved AI services. Where AI is used with organisational data, privacy, contractual terms, access control and retention need to be considered.

Natural-language interfaces in BI tools can help users explore data, but their answers depend on the semantic model and business definitions. Poorly modelled data cannot be repaired merely by adding an AI interface.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

20. Reproducibility, Documentation and Analytical Quality

Reproducibility means another competent person can understand how source data became the reported result. It requires preserved source data, documented transformations, versioned calculations and clear definitions.

An analytical quality review asks whether the question is appropriate, data is fit for purpose, transformations are justified, calculations are correct, visuals are not misleading and conclusions match the evidence.

Peer review is valuable because analysts can become accustomed to their own assumptions. A reviewer should attempt to reproduce key figures and challenge ambiguous definitions.

Good analytics also communicates limitations. Missing data, small samples, changing definitions, uncertain causal interpretation and unobserved factors should be visible rather than hidden.

Practical Application

- Apply the concepts to a fictional SME dataset containing Date, Product, Region, Customer Segment, Units, Revenue, Cost, Discount and Satisfaction Score.
- Record one data-quality risk, one analytical question and one management decision that could be supported by this chapter.
- Add relevant terms and business definitions to your learner data dictionary.

21. Practical Lab Series

Lab	Focus	Task	Evidence
Lab 1	Business Question & KPI Design	Convert a vague management request into five analytical questions and define six KPIs.	Question sheet + KPI catalogue
Lab 2	Data Profiling	Inspect a messy spreadsheet for missing values, duplicates, invalid types and inconsistent categories.	Data-quality audit
Lab 3	Spreadsheet Cleaning	Clean and standardise the dataset while preserving the original source.	Clean dataset + change log
Lab 4	Spreadsheet Analysis	Use formulas, conditional logic and pivot tables to analyse revenue, margin and volume.	Analysis workbook
Lab 5	Descriptive Statistics	Calculate and interpret central tendency, variation, percentiles and outliers.	Statistics worksheet
Lab 6	Visualisation	Create four charts matched to four different business questions.	Chart portfolio
Lab 7	Power Query	Import, transform, combine and profile two related datasets.	Transformation log
Lab 8	Data Model	Build a simple star schema with a fact table, Date, Product and Region dimensions.	Model diagram
Lab 9	Power BI Report	Create an interactive report with KPIs, trends, category comparisons and filters.	PBIX/report evidence
Lab 10	Insight Writing	Write a one-page management brief separating findings, interpretation, limitations and recommendations.	Insight brief
Lab 11	AI-Assisted Analysis Review	Use an approved AI assistant on fictional data/tasks and independently verify formulas and claims.	AI verification log
Lab 12	Capstone	Complete an end-to-end SME analytics project.	Dashboard + report + portfolio

22. Core Spreadsheet Practice

Skill	Example practice
Data validation	Restrict category entries and identify invalid records.
IF logic	Classify transactions by margin or service level.
SUMIFS / COUNTIFS	Aggregate results using multiple business conditions.

XLOOKUP or equivalent	Retrieve reference attributes from a master table.
Date functions	Create month, quarter and ageing fields.
Text functions	Standardise names, codes and categories.
Pivot tables	Summarise measures by time, product, region and segment.
Conditional formatting	Highlight exceptions without replacing analytical judgement.
Error checking	Detect divide-by-zero, missing lookups and inconsistent formulas.

23. Basic SQL Learning Block

The following examples are conceptual and should be practised only on an authorised training database.

```
SELECT Region, SUM(Revenue) AS TotalRevenue
FROM Sales
WHERE OrderDate >= '2026-01-01'
GROUP BY Region
ORDER BY TotalRevenue DESC;
```

Interpretation: select a dimension (Region), aggregate a measure (Revenue), filter the time period, group the records and sort the result. Learners should be able to explain the logic before attempting more complex queries.

- Practice SELECT and column selection.
- Apply WHERE filters.
- Use COUNT, SUM and AVG.
- Use GROUP BY for category summaries.
- Understand INNER JOIN and LEFT JOIN.
- Check row counts before and after joins to detect duplication.

24. Data Visualisation Decision Guide

Question	Recommended starting visual	Avoid
Compare categories	Bar/column chart	3D effects; too many categories
Show change over time	Line chart	Unordered category charts
Show distribution	Histogram / box plot where supported	Only reporting the mean
Show relationship	Scatter plot	Claiming causation from association
Show composition	Stacked bar for limited categories	Many-slice pie charts
Show one KPI	Card + comparison context	Large decorative gauges without context
Show geography	Map only when location matters	Maps for non-geographic comparisons

25. SME Case Studies

Case 1 – Sales Decline

A small manufacturer reports that annual revenue is down. Management wants to know which products, regions and months explain the change.

Task: Define the analytical questions, required fields, comparisons and visuals. Identify what additional data would be needed before recommending action.

Case 2 – Training Completion

A VET provider wants to compare completion rates across programmes.

Task: Define numerator and denominator carefully. Consider cohort size, withdrawals, missing outcomes and whether programmes are comparable.

Case 3 – Customer Satisfaction

An SME has satisfaction scores from only 35% of customers.

Task: Explain non-response bias and why the average score may not represent all customers. Propose a cautious reporting approach.

Case 4 – Delivery Performance

Average delivery time improved, but complaints increased.

Task: Examine distribution, percentiles, late-delivery rate and customer segments rather than relying on the average alone.

Case 5 – Marketing Campaign

Website visits and sales both increased after a campaign.

Task: Explain why the data shows association but does not automatically prove that the campaign caused all of the sales increase.

Case 6 – AI-Generated Insight

An AI assistant claims that one region is 'underperforming because customers dislike the product'.

Task: Identify which part is data, which part is inference, and what evidence would be required to support the causal explanation.

26. Capstone Project – SME Data Analytics Dashboard

Complete an end-to-end analytics project using a fictional, open, anonymised or authorised dataset. The project should demonstrate analytical reasoning, not only visual design.

Required Deliverables

1. Business context and decision problem.
2. Five to eight analytical questions.
3. Data-source description and data dictionary.
4. Data-quality audit and cleaning log.
5. Clean analytical dataset.
6. At least six documented KPI definitions.
7. Descriptive statistical analysis.
8. Simple analytical data model.
9. Interactive dashboard/report with appropriate visualisations.
10. Five evidence-based insights.

11. Limitations and uncertainty statement.
12. Three practical recommendations linked to evidence.
13. One-page management summary.
14. Portfolio reflection describing what was learned and what should be improved.

Assessment Rubric

Criterion	Weight
Business question and analytical scope	10%
Data quality and preparation	20%
Calculations and statistical reasoning	15%
Data model and KPI definitions	15%
Visualisation and dashboard usability	15%
Interpretation and evidence	15%
Ethics, privacy and limitations	5%
Communication and reproducibility	5%

27. Final Knowledge Assessment

1. What is descriptive analytics?

2. Why is the median useful?

3. What is granularity?

4. Why should data cleaning be documented?

5. What is a KPI?

6. Why does correlation not prove causation?

7. What is a primary key?

8. What is a star schema?

9. What is Power Query used for?

10. What is the difference between a measure and a calculated column in Power BI?

11. Why can a truncated axis be misleading?

12. What should be verified when AI generates a formula or analytical conclusion?

Answer Key

1. Analysis focused on what happened or what is observed.

2. It is less sensitive than the mean to extreme values and skewed distributions.

3. The level of detail represented by each record.

4. To make transformations transparent and reproducible.

5. A defined measure used to track progress toward an objective.

6. Association can arise from confounding, selection, reverse influence or chance.

7. A field or combination of fields that uniquely identifies a record.

8. A model with a central fact table linked to dimension tables.

9. Connecting, profiling, cleaning and transforming data reproducibly.

10. A measure is evaluated dynamically in filter context; a calculated column is evaluated by row and stored.

11. It can visually exaggerate small differences.

12. Logic, calculations, assumptions, source data and resulting claims.

28. Post-Training Skills Check and Reflection

Repeat the Initial Skills Check and compare results with the practical evidence in your portfolio.

- Which analytical task can you now perform independently?
- Which data-quality problem do you find most difficult to detect?
- Which KPI definition in your organisation is currently ambiguous?
- Which visualisation mistake are you most likely to avoid after this training?
- What evidence would you require before making a causal claim?
- Which next skill should you develop: advanced Excel, SQL, Power BI/DAX, Python, statistics or data engineering?

29. Career Transition Guide

Starting profile	Possible direction	Portfolio evidence
Administrative / office staff	Reporting Assistant	Clean workbook; KPI table; management report
Finance / operations staff	Business Data Analyst	Pivot analysis; KPI definitions; dashboard
Spreadsheet user	Junior Data Analyst	Cleaning log; statistics; Power BI report
Marketing / sales staff	Marketing Data Analyst	Campaign analysis; funnel KPIs; visual story
VET / education staff	Education Data Analyst / Reporting role	Completion/participation analysis; dashboard
Junior digital staff	BI / Power BI pathway	Power Query; star schema; measures; interactive report

30. Glossary

Term	Definition
Analytics	Systematic examination of data to answer questions and support decisions.
Dataset	A structured collection of observations and variables.
Variable	A characteristic or field that can take different values.
Dimension	A descriptive attribute used to group or filter analysis.

Measure	A quantitative value used in calculations and aggregations.
Granularity	The level of detail represented by a record.
Data Profiling	Systematic examination of data structure, quality and distributions.
Missing Value	An expected value that is absent.
Outlier	An observation unusually distant from the main distribution.
Mean	Arithmetic average.
Median	Middle value of an ordered distribution.
Standard Deviation	A measure of dispersion around the mean.
Percentile	A value below which a given percentage of observations falls.
Correlation	Statistical association between variables.
Causation	A relationship in which a change in one factor contributes to a change in another under appropriate evidence.
KPI	Key Performance Indicator.
Primary Key	A unique identifier for a record.
Foreign Key	A field linking a record to a key in another table.
Join	An operation combining related tables.
Fact Table	A table containing events or measurements at a defined grain.
Dimension Table	A table containing descriptive attributes used to analyse facts.
Star Schema	A dimensional model connecting a fact table to dimensions.
Power Query	Microsoft data connection and transformation technology used in products including Power BI.
DAX	Data Analysis Expressions, the formula language used in Power BI semantic models.
Filter Context	The set of filters affecting evaluation of a measure.
Dashboard	A visual interface summarising important metrics and analytical information.
Data Governance	Rules, responsibilities and processes for managing data quality, access, definitions and use.
Reproducibility	Ability to understand and repeat the process that produced an analytical result.

31. Curated Learning Resources

European Commission JRC – DigComp 2.2
https://publications.jrc.ec.europa.eu/repository/handle/JRC128415
Use for information and data literacy, critical evaluation and broader digital competence.

Microsoft Learn – Training for Data Analysts
https://learn.microsoft.com/en-us/training/career-paths/data-analyst
Current Microsoft role-based pathway covering data preparation, modelling, visualisation and analysis.

Microsoft Learn – Get started with Microsoft data analytics
https://learn.microsoft.com/en-us/training/paths/data-analytics-microsoft/
Introductory path to analytics and Power BI.

Microsoft Learn – Prepare data for analysis with Power BI
https://learn.microsoft.com/en-us/training/paths/prepare-data-power-bi
Hands-on learning for connecting, cleaning, transforming and loading data.

Microsoft Learn – Model data with Power BI
https://learn.microsoft.com/en-us/training/paths/model-power-bi/
Intermediate learning on semantic models, relationships and DAX.

Microsoft Learn – Prepare and visualize data with Power BI
https://learn.microsoft.com/en-us/training/paths/prepare-visualize-data-power-bi/
Beginner-oriented end-to-end Power BI learning.

Microsoft Learn – Power BI learning path directory
https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-learning-path-directory
Scenario-based routes for new users, Excel users, analysts and report creators.

Microsoft Learn – PL-300 Study Guide
https://learn.microsoft.com/en-us/credentials/certifications/resources/study-guides/pl-300
Current certification blueprint; useful for learners who want to progress toward Power BI Data Analyst Associate.

Google Data Analytics Professional Certificate
https://www.coursera.org/professional-certificates/google-data-analytics

Optional structured beginner career programme covering cleaning, analysis, visualisation, spreadsheets, SQL, Python and Tableau; access/cost conditions should be checked.

IBM SkillsBuild
https://skillsbuild.org/
Optional skills platform; check current data and analytics learning offerings and credential conditions.

32. Certification and Progression

DT-03 completion should be presented as a ReskillingLab pathway completion record, not as equivalent to an external professional certification. Learners seeking an industry credential can use DT-03 as preparation for further study.

- Microsoft Certified: Power BI Data Analyst Associate (PL-300) is a relevant next-step credential for learners specialising in Power BI.
- Microsoft Learn currently structures PL-300 skills around preparing data, modelling data, visualising and analysing data, and managing/securing Power BI.
- Google's Data Analytics Professional Certificate is a broader beginner career programme and currently includes spreadsheets, SQL, Python and Tableau.
- Credential availability, pricing, exam requirements and provider conditions should be checked immediately before publication on ReskillingLab.

33. Suggested 50–70 Hour Delivery Plan

Component	Indicative hours
Orientation and initial skills check	2
Analytics foundations, data and measurement	6
Data collection, quality and cleaning	8
Spreadsheet analytics	8
Statistics, relationships and KPI design	8
Visualisation and data storytelling	7
Relational data and SQL concepts	5

Power Query, modelling and DAX	9
Power BI report development	7
Ethics, AI-assisted analytics and quality	4
Capstone, assessment and reflection	6
Total (can be compressed toward 50 hours by reducing guided practice)	70 hours;

34. Trainer / Facilitator Notes

- Use realistic but fictional or authorised datasets and preserve a clean source copy.
- Require learners to explain the business question before opening an analytics tool.
- Assess definitions and reasoning as well as formulas and visual appearance.
- Ask learners to verify totals after joins, filters and transformations.
- Do not allow a polished dashboard to compensate for weak data quality or unsupported conclusions.
- Make learners distinguish observation, interpretation and recommendation.
- Require disclosure and verification when generative AI is used for formulas, code or interpretation.
- Encourage a portfolio containing the data dictionary, cleaning log, KPI catalogue, dashboard and capstone report.

35. References

European Commission, Joint Research Centre. (2022). DigComp 2.2: The Digital Competence Framework for Citizens. Publications Office of the European Union.

Microsoft Learn. (2026). Training for Data Analysts.

Microsoft Learn. (2026). Get started with Microsoft data analytics.

Microsoft Learn. (2026). Prepare data for analysis with Power BI.

Microsoft Learn. (2026). Model data with Power BI.

Microsoft Learn. (2026). Prepare and visualize data with Microsoft Power BI.

Microsoft Learn. (2026). Study guide for Exam PL-300: Microsoft Power BI Data Analyst.

Google Career Certificates. (2026). Google Data Analytics Professional Certificate.

IBM SkillsBuild. Data and analytics learning resources.

36. Website Publishing Structure

For ReskillingLab, DT-03 should be published as modular learning content. Each learning unit should use a consistent interface: Learn | Practice | Check Your Understanding | Resources.

- Overview & Initial Skills Check
- Analytics Foundations
- Data & Measurement
- Data Collection & Sampling
- Data Quality & Cleaning

- Spreadsheet Analytics
- Descriptive Statistics
- Relationships & Correlation
- KPI Design
- Data Visualisation
- Data Storytelling
- Relational Data & SQL
- Data Modelling
- Power Query
- DAX & Measures
- Power BI Reporting
- Privacy, Ethics & Governance
- AI-Assisted Analytics
- Reproducibility & Quality
- Labs & Cases
- Capstone
- Assessment
- Resources, Credentials & Glossary