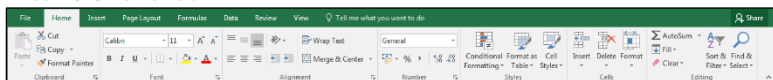


Data Analytics - Foundation

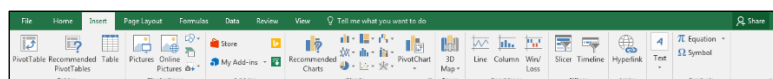
This quick reference is for Microsoft Excel 2016 and Microsoft Power BI on Windows 10.

Quick Reference

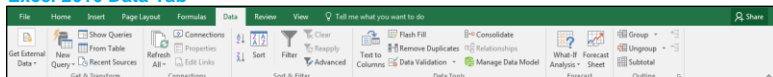
Excel 2016 Home Tab



Excel 2016 Insert Tab



Excel 2016 Data Tab



Power BI Desktop Home Tab



Power BI Service Report Ribbon



Certification Test Goals

This module sets out foundational knowledge and skills in data analytics that can be used to get, prepare and analyse data to generate business intelligence that can be used to achieve business goals.

Successful candidates will be able to:

- Understand the key concepts relating to the application of data analytics in business.
- Understand and apply key statistical analysis concepts.
- Import data into a spreadsheet and prepare it for analysis using data cleansing and filtering techniques.
- Summarise data sets using pivot tables and pivot charts.
- Understand and apply data visualization techniques and tools.
- Create and share reports and dashboards in a data visualization tool.

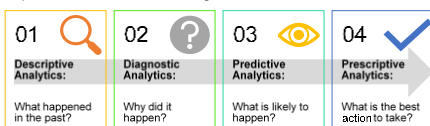
1 Concepts and Statistical Analysis

1.1 Key Concepts

1.1.1 Identify the main types of data analytics:

Data analytics refers to the techniques and processes used to collect, organise and examine data sets to create meaningful and useful information. Four types of data analytics are:

- Descriptive analytics** - Used to find out what has happened in the past by summarising data from multiple sources.
- Diagnostic analytics** - Used to determine why something happened in the past by looking at the data to understand root causes and contributing factors.
- Predictive analytics** - Used to forecast the probabilities of future outcomes using the findings of descriptive and diagnostic analytics.
- Prescriptive analytics** - Used to identify suitable actions to take now, for example, to eliminate future problems or take advantage of trends.



Two broad categories of data analytics are:

- Quantitative analytics** - Used to analyse information that can be measured and written down with numbers, such as, profits, costs, and sales.
- Qualitative analytics** - Used to analyse information that cannot be easily measured or expressed as numbers, such as attitudes, opinions and behaviours.

1.1.2 Outline the business benefits of data analytics:

- Identifies patterns/trends** - Evaluating large sets of data containing varying types of data enables businesses to uncover hidden data patterns, market trends and customer preferences and behaviour.
- Improves efficiency** - Insight gained by data analytics enables businesses to automate and optimize their business processes thus improving operational efficiency.
- Supports decision making** - Access to large volumes of analysed data allows businesses to make better and more informed business decisions leading to more effective business strategies.
- Presents information effectively** - Information presented in easy-to-understand formats such as tables, charts, maps, and combined in reports and dashboards, help businesses to better interpret and understand their data.

1.1.3 Identify the main phases of data analysis:

The phases vary depending on the type of analytics and the business implementation. The following phases relate to data mining, a type of analytics that explores large data sets and generates predictions:

- Business understanding** - The aim of the first phase is to define the data analytics objectives and project plan to achieve the specified business goals.
- Data understanding** - The aim of the second phase is to consider the data requirements, which includes collecting, exploring and assessing the quality of the data.
- Data preparation** - The aim of the third phase is to complete the final data set, which includes acquiring, cleaning and transforming the data.
- Modelling** - The aim of the fourth phase is to select and apply appropriate data modelling techniques.
- Evaluation** - The aim of the fifth phase is to evaluate the model to ensure that the results are useful to the business.
- Deployment** - The aim of the sixth phase is to deploy the model, which includes planning and monitoring the deployment of results.

1.1.4 Recognise data protection considerations when analysing data like:

- Anonymise personal data if possible** - If personal data is not needed to achieve data analytics objectives, then avoid the effort of complying with data protection laws by not collecting it or anonymising it. Anonymisation techniques include removing direct identifiers, such as names and addresses, and using aggregation to reduce details.
- Comply with applicable data protection regulations** - Data protection regulations, which protect the privacy rights of individuals in relation to the processing of their personal data, vary across the world. Businesses should comply with applicable data protection regulations in the relevant jurisdiction or risk penalties for non-compliance.

1.2 Statistical Analysis

1.2.1 Describe measures of central tendency of a data set:

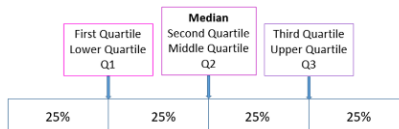
- Mean** - This measure describes the average of a data set and is calculated by summing all the values in the data set and dividing by the number of values in the data set.
- Median** - This measure describes the middle value in a data set distribution when it is arranged in ascending order. In a data set with an even number of values it is the average of the two middle values. In a data set with an odd number of values, it is the middle value in the data set.
- Mode** - This measure is the value that occurs most frequently in a data set. There is no mode when all the values in a data set are unique.

1.2.2 Calculate the central tendency value of a data set using a function:

- Mean** - To calculate the mean, the syntax is `AVERAGE(number1,[number2], ...)`
- Median** - To calculate the median, the syntax is `MEDIAN(number1,[number2], ...)`
- Mode** - To calculate the mode, the syntax is `MODE((number1,[number2],...))`

1.2.3 Describe measures of variation of a data set:

- Quartiles** - These are values that divide the data set into quarters according to where the values fall on the number line.
- The 1st quartile is the median of the lower half of the data set, the 2nd quartile is the median of the data set, and the 3rd quartile is the median of the upper half of the data set.



- The formula to find a quartile, by determining the value in the appropriate position in a ranked discrete data set, are outlined below, where n is the number of values in the data set:

Quartile Position	Formula
First, Q1	$(n + 1) / 4$
Second, Q2 (the median)	$(n + 1) / 2$
Third, Q3	$3(n + 1) / 4$
Fourth, Q4	Maximum value

- Variance** - This measure describes how far a data set is spread out by showing how representative the mean is of each value in the data set.
- The smaller the variance, the closer each value is to the mean; the greater the variance, the farther each value is from the mean; and a variance of zero indicates that there is no variation from the mean.
- Technically variance is the average of the squared differences from the mean.
- Range** - This measure describes how the data is dispersed or spread out in a data set and is calculated by subtracting the minimum value from the maximum value.

1.2.4 Calculate the variation of a data set:

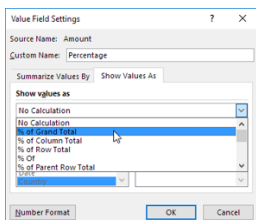
- Quartile** - To calculate a quartile, the syntax is `QUARTILE(array, quart)`
The `array` argument is the range of cells containing the data set and the `quart` argument is the position of the required quartile to return (0: minimum, 1: first, 2: second, 3: third, 4: maximum).
- Variance** - To calculate a variance, the syntax is `VAR.P(number1,[number2], ...)`

- Click on any cell in the **PivotTable**.
- In the **PivotTable Fields** pane, drag and drop the required fields, which can include multiple instances of the same value field, into the **Σ Values** area.
- To change the aggregation method of a value field, right-click any cell in the field and select **Summarize Values By**.
- Select the required aggregation method to change to: **Sum**, **Count**, **Average**, **Max**, or **Min**.
- To change the name of the value field, right click the field and select **Value Field Settings**.

- Enter a new name for the value field in the **Custom Name** text box.
- Click **OK**.

3.1.3 Display values as: % calculation, difference from specific values, running total, ranked.

- Click on any value cell in the **PivotTable**.
- On the **Analyze** tab, in the **Active Field** group, click **Field Settings**.
- Select the **Show Values As** tab.
- Select the **Show values as** drop-down list.
- Select the required display value.
- Depending on the display value selected, select appropriate options in the **Base field** and **Base item** lists.
- If required, enter a new name in the **Custom Name** text box.



- If required, click on **Number Format**, set the format for the number and click **OK**.
- Click **OK**.

3.2 Pivot Table Frequency Analysis

3.2.1 Automatically group data.

- Click on any date/time or numeric value cell in the **PivotTable**.
- On the **Analyze** tab, in the **Group** group, click **Group Field**.
- If grouping **date/time** values, in the **Grouping** dialog box, in the **By** list, select the period of time to group by. If grouping by **Days**, set the **Number of days**.
- If grouping **numeric** values, in the **Grouping** dialog box, in the **By** text box, type the **interval** to group by.
- If required, adjust the **Starting at** and **Ending at** values. By default, these are the first and last dates or the lowest and highest values in the pivot table.
- Click **OK**.

3.2.2 Manually group data.

- In the **PivotTable**, select the items to add to the group. (To select multiple non-contiguous items, press **CTRL** and click on the items.)
- On the **Analyze** tab, in the **Group** group, click **Group Selection**.

3.2.3 Rename groups.

- In the **PivotTable**, click on a **group name**.
- Type a **new name**.
- Press **Enter**.

3.2.4 Ungroup data.

- In the **PivotTable**, select a **group label**.
- On the **Analyze** tab, in the **Group** group, click **Ungroup**.

3.3 Filtering Pivot Tables

3.3.1 Use the report filter.

- Click on any cell in the **PivotTable**.
- To insert a report filter, drag and drop the required fields to the **Filters** area in the **PivotTable Fields** panel.
- To use the report filter to display items in the **PivotTable**, select the **report filter** drop-down arrow and select the item or items to display and click **OK**.
- To separate the report into different worksheets based on the specified filter field, on the **Analyze** tab, in the **PivotTable** group, click the **Options** down-pointing arrow.
- Click **Show Report Filter Pages**.
- In the **Show Report Filter Pages** dialog box, select the required filter field listed and click **OK**.
- To clear the filter from the **PivotTable**, select **All** from the **report filter** drop-down list and click **OK**.
- To remove report filters from the **PivotTable**, remove the fields from the **Filters** area.

3.3.2 Insert and use slicers to filter single pivot tables.

- Click on any value in the **PivotTable**.
- On the **Analyze** tab, in the **Filter** group, click **Insert Slicer**.
- Choose the required **field** from the list of fields in the **Insert Slicer** dialog box and
- Click **OK**.

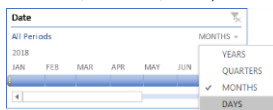
- Click on the items shown in the slicer to filter the table according to the items selected.
- To clear the filter, click the **Clear Filter** icon at the top right of the slicer.
- To delete the slicer, click on the slicer and press **Delete**.

3.3.2 Insert and use slicers to filter multiple pivot tables.

- Select the required **slicer**.
- On the **Options** tab, in the **Slicer** group, click **Report Connections**.
- In the **Report Connections** dialog box, select the **pivot tables** to connect to the slicer.
- Click **OK**.
- Click on the **items** shown in the slicer to filter the tables according to the items selected.

3.3.3 Insert and filter a timeline.

- Click on any value in the **PivotTable**.
- On the **Analyze** tab, in the **Filter** group, click **Insert Timeline**.
- In the **Insert Timelines** dialog box, select the **field** on which to base the timeline.
- Click **OK**.
- Click the **time level down arrow** and select the **time level** to display on the timeline (**YEARS**, **QUARTERS**, **MONTHS**, **DAYS**).

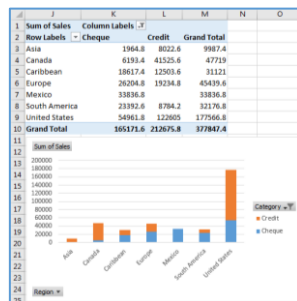


- To select the time period to display, click a **period tile** and drag at either end of the tile to include the time range required.

3.4 Using Pivot Charts

3.4.1 Insert a pivot chart for an existing pivot table.

- Click on any cell in the **PivotTable**.
- On the **Analyze** tab, in the **Tools** group, click **PivotChart**.
- Select the chart type and variant in the **Insert Chart** dialog box and click **OK**.
- To customise the appearance and layout, click on the **PivotChart** and select the **Chart Elements** button or the **Chart Styles** button to set their respective options as required.
- To show or hide the field buttons, on the **Analyze** tab, in the **Show/Hide** group, select or deselect the **Field Buttons** in the **Field Buttons** drop down menu as required.



3.4.2 Create a pivot chart from fields in a data set.

- Select any cell in the table.
- On the **Insert** tab, in the **Charts** group, click **PivotChart**.
- In the **Create PivotChart** dialog box, select where to locate the pivot chart and click **OK**.
- Drag and drop the **fields** as required to the **Filters**, **Legend (Series)**, **Axis (Categories)**, and **Values** areas in the **PivotChart Fields** panel.
- To change the default chart type, on the **Design** tab, in the **Type** group, click **Change Chart Type**, select the required **chart type** and **variant** in the **Change Chart Type** dialog box, and click **OK**.
- To customise the appearance and layout, click on the **PivotChart** and select the **Chart Elements** button or the **Chart Styles** button to set their respective options as required.
- To show or hide the field buttons, on the **Analyze** tab, in the **Show/Hide** group, select or deselect the **Field Buttons** in the **Field Buttons** drop down menu as required.

4 Data Visualization

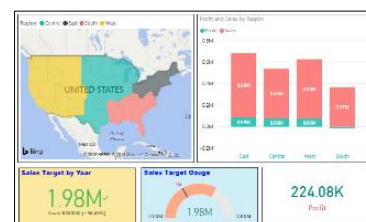
4.1 Concepts and Setup

4.1.1 Understand the concept of data visualization using reports and dashboards.

- Data visualization** provides a visual way to access data – for example, by displaying summary information or trends and patterns within data in a visual form.
- There are many types of visualizations to choose from, depending on the data and what you want to present or find out.
- Data visualizations can provide a visual snapshot of data and can be interactive for ongoing analysis and querying.
- Reports** include data visualizations and results of the data analytic process. A report can consist of one or more pages of visualizations. They can provide summary information and key insights from business data and can also include interactive features. They provide a visual format to create and share business intelligence.



- Dashboards** combine multiple visualizations to organise and display key information and metrics on a single screen that can be viewed at a glance. They are useful for monitoring changes and gaining business insights in real-time. Visualizations are pinned (added) to dashboards from reports.



4.1.1 Outline common visualizations like:

- Charts** - Useful for conveying numerical data across categories or time scales, visually revealing similarities, differences and trends among multiple values. Column, line, pie/donut and bar charts are popular types.
- Key performance indicators (KPIs)** - Useful for highlighting critical data and measuring performance or progress towards a goal using predefined conditions. KPIs and gauge charts are popular types, which show the current status, the goal to be achieved and the distance from the goal.
- Maps** - Useful for presenting quantitative information according to location to give an overview of data distribution across geographic locations. They give an insight into data sets by visually highlighting the best and worst performing areas, trends and outliers. Bubble maps and filled maps are popular types, where bubble size and shading, respectively, show how values differ in proportion across geographical locations.

4.1.2 Recognise common data visualization tools.

- Tableau Public**: A free web-based platform for creating data visualizations, which are uploaded to its public website for sharing.
- Microsoft Power BI**: A suite of tools consisting of **Power BI Desktop**, a free desktop application and **Power BI service**, an online platform with different licensing options.

4.1.2 Recognise common data visualization tools' functions like:

- Visualise data** - Data visualization tools allow you to connect to a wide range of data sources to analyse data sets, through the use of data visualizations, to find and visually present business intelligence in reports and dashboards.
- Publish and share business intelligence** - Data visualization tools allow you to publish and share business intelligence online in reports and dashboards. In **Microsoft Power BI** the sharing functionality and creation of dashboards is available in the online platform, **Power BI service**, only. Reports can be created in **Power BI Desktop** or **Power BI service** but those created in **Power BI Desktop** need to be published to **Power BI service** for sharing.

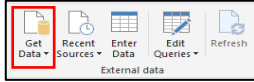
4.1.3 Understand good design practice in reports and dashboards like:

- Clean and uncluttered layout** - Include only the most relevant data on reports and dashboards so the reader can easily see the key insights.

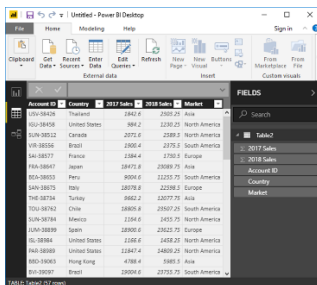
- **Descriptive titles** - Use descriptive titles to ensure data visualizations are easy to understand. Add labels using text boxes to describe the page, grouping of visuals, or describe an individual visual.
- **Consistent fonts and colour** - Use one or two font types and up to three font sizes to create visual hierarchy among the content (titles, labels, data).
- **Use of colour for emphasis and understanding** - Use mainly natural colours and keep bright or dark colours for highlighting outliers or critical data; use colour shades or gradients to show the spread of values across data; and use colour to highlight variance around a central value, for example, green to show positive values and red to show negative values.

4.1.4 Import a data set from a spreadsheet into a data visualization tool and save the file.

- Double-click on the **Power BI Desktop** app icon to launch the application.
- On the **Home** tab, in the **External Data** group, click **Get Data**.



- Select **Excel** in the dropdown menu.
- Locate and select the required **Excel** file and click **Open**.
- Check the required data table under **Display options** in the **Navigator** window and click **Load**.
- In the **Report View**, the **Fields** pane displays all the column headings in the data set and the report canvas page is empty.
- Click the **Data View** icon on the left to view the data set that has been imported.

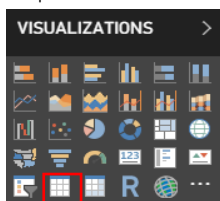


- Click **File** and choose **Save**.
- Type the name of the file in the **File name** box in the **Save As** dialog box.
- Click **Save** to save the data as a **Power BI** file.

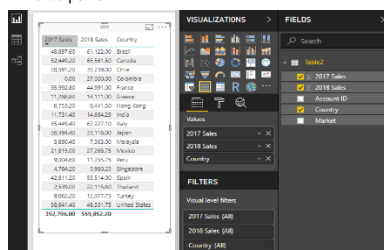
4.2 Visualization

4.2.1 Create tables in a report.

- In **Power BI Desktop**, select **File** and select **Open**.
- Select the relevant **Power BI** file and click **Open**.
- In **Report view**, click on the **Table** icon in the **Visualizations** pane.



- Select the required fields to add to the table in the **Fields** pane.



- Select the table on the report canvas page and click the **Format** icon.
- Format the table as required using the formatting options.
- Click outside the table on the report canvas page and add additional visualizations, as required, to the report.
- **Save** and **close** the report.

4.2.2 Visualize data as a chart: column, bar, line, pie.

- In **Report view** in **Power BI Desktop**, click on the appropriate chart icon in the **Visualizations** pane.



- Select the required fields to add to the chart in the **Fields** pane.
- For the **line** chart, also select the time scale to display by deselecting the options not required.
- Click the **Format** icon and format the chart as required using the formatting options.

4.2.3 Apply font and background conditional formatting to show: high/low values, above/below average values.

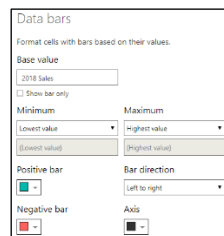
- In **Report view** in **Power BI Desktop**, select the relevant visualization.
- Click the down arrow next to the relevant field in the **Values** area in the **Visualizations** pane.
- Select **Conditional Formatting**.
- Choose **Font color** or **Background color**.
- To set a colour for the lowest value and the highest value, set a colour in the **Minimum** area and the **Maximum** area respectively and click **OK**.
- To set rules, choose the **Format by Rules** option, set one or more value ranges and related colours in the **Rules** area by entering the relevant **If value condition**, and **value condition** and related **colour**, and click **OK**.

4.2.3 Edit font and background conditional formatting to show: high/low values, above/below average values.

- To edit font and background conditional formatting, select the down arrow next to the relevant field, select **Conditional Formatting**, click **Font color** or **Background color**, as appropriate, and edit as required.
- To remove font and background conditional formatting, select the down arrow next to the relevant field, select **Remove Conditional Formatting**, and select **All**, **Font color** or **Background color**, as appropriate.

4.2.4 Apply data bars.

- In **Report view** in **Power BI Desktop**, select the relevant visualization.
- Click the down arrow next to the relevant field in the **Values** area in the **Visualizations** pane.
- Select **Conditional Formatting** and choose **Data bars**.



- The default range for the data bars is the **Lowest value** in the data set for **Minimum** and **Highest value** for **Maximum**.
- Change the option below **Minimum** and **Maximum** to **Number** and specify a value for each option, to set the range of the data bars.
- Set colours for **Axis**, **Positive bar**, and **Negative bar** and **Bar direction** as required.

4.2.4 Edit data bars.

- To edit data bars, select the down arrow next to the relevant field and select **Conditional Formatting**, click **Data bars** and edit as required.
- To remove data bars, select the down arrow next to the relevant field, select **Remove Conditional Formatting** and select **Data bars**.

4.2.5 Apply, edit visual level filters.

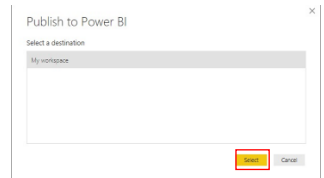
- In **Report view** in **Power BI Desktop**, select the relevant table visualization.
- The value fields selected in the visualization are included automatically under **Visual level filters** in the **Filters** pane.
- To add other fields for filtering, click the items under the **Field** pane or drag and drop the **Field** items under **Visual level filters**.
- Click on the down arrow next to the relevant field under **Visual level filters** to display the visual level filter options for that field.

- For numerical values, select the items to display by setting rules (for example, is less than or equal to) and specifying values as required.
- For qualitative values, select and deselect the items to display from the items displayed under **Basic filtering**.
- Toggle between selecting all items and unselecting all items by checking and unchecking **Select All**.

4.3 Publishing and Sharing

4.3.1 Publish a report.

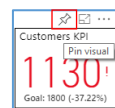
- To publish reports from **Power BI Desktop** to **Power BI service**, first setup **Power BI service** with a **Power BI Pro** license.
- In **Power BI Desktop**, save the report to publish and click the **Publish** icon.
- Select **Publish to Power BI**.
- Sign in to your **Power BI** account.
- Select **My workspace** as the destination to publish the report and click **Select**.



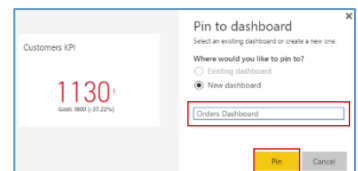
- To open the report in your **Power BI** site, click the link received or click **Got It** to continue working in **Power BI Desktop**.

4.3.2 Create a dashboard.

- Go to <https://app.powerbi.com> and sign in to your **Power BI** account.
- Select **My Workspace** on the left menu pane, select the **Reports** heading and select the report containing the visualizations to add to the dashboard.
- Hover over the required **visualization** in the report and click on the **pushpin** icon.



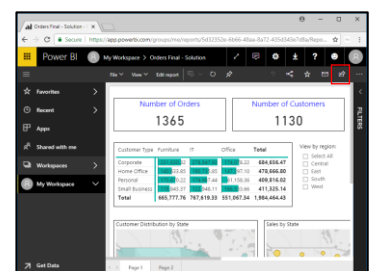
- In the **Pin to dashboard** window, select **New dashboard**.
- Type the name of the new dashboard in the text box and click **Pin**.



- Pin new visualizations, as required, to the **Existing dashboard**, from this report or other reports.

4.3.3 Share a report, dashboard using a link.

- In an open **report** or **dashboard** in **Power BI service**, select **Share** in the menu bar.



- Under **Grant access to**, enter the email addresses of other users and click **Share**.
- Or under **Report Link**, copy the link to send via email or another communication tool.
- Reports and dashboards can only be shared with users with Power BI Pro licenses.

4.3.3 Share a report to web.

- In an open report in **Power BI service**, select **File** in the menu bar and select **Publish to web**.
- Click on **Create embed code** and click **Publish**.
- In the **Success** pane, copy the link that can be sent in email or the HTML that can be pasted directly into your web page or blog, and click **Close**.

For more information, visit: www.icdl.org