**Part 1 – Discover Data Analysis**

**Intro**:

-You must be able to tell a story within data in order to use it to its potential.

*-An example is businesses using data to tell sale trends, track inventory, purchase habits etc*

*-This can hep businesses make fast paced decisions and craft reports that tell a story and help*

*Leaders take action to have a competitive advantage*

*-You must act on data! Data alone is not enough!*

Data Analysis

-*The process of identifying, cleaning, transforming, and modeling data to discover useful meaning*

**Overview of Data Analysis:**

Components of analytics

-*Descriptive (helps answers question about whats happened based on historical data, summarizes large data sets to describe outcomes)*

*-Diagnostic (Helps answer questions about why events have happened. They do this by identifying anomalies, collecting data that’s related to the anomalies, and using statistical techniques to discover relationships and trends)*

*-Predictive (helps answer questions on what will happen in the future using historical data to identify trends and the likeliness of reoccurance)*

*-Prescriptive (helps answer questions about what actions should be taken to achieve a specific goal, uses machine learning to find patterns)*

*-Cognitive (Draws inferences from existing data and patterns to draw conclusions and adds there findings back for future use)*

**Roles In Data**

*Buisness Analyst – Similar to a data analyst. Difference is a business analyst is closer to the business and is a specialist in interepreting the data from the viz*

*Data Analyst – Empowers organizations to maximize the use of their data through viz’s and reporting. Responsible for profiling, cleaning, and transforming data as well as designing and building data models. Responsible for managing the different tools like dashboards*

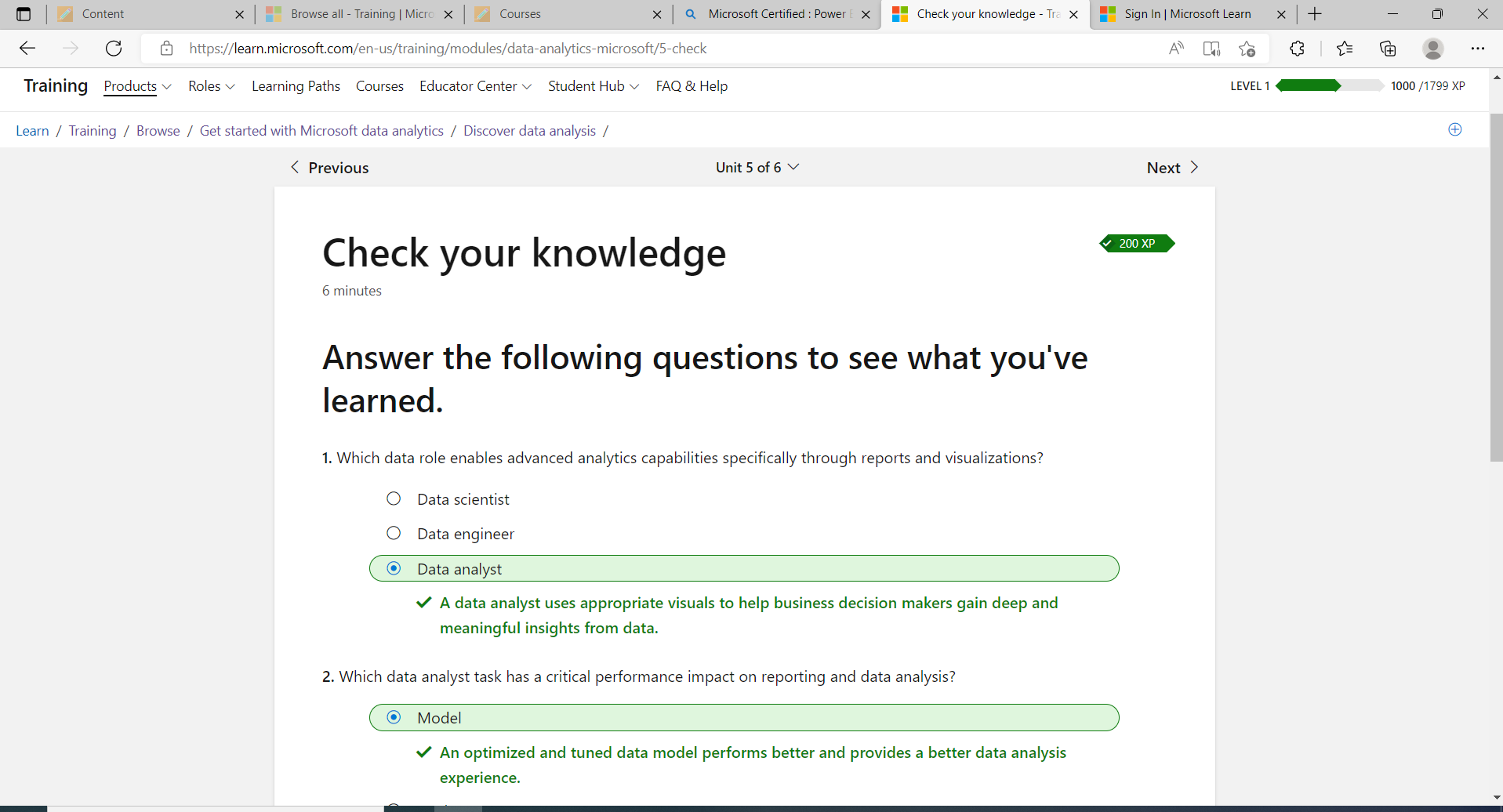
*Data Engineer – Set up technologies both on premise and in the cloud, they manage and secure the flow of data from different sources.*

*Data Scientist – Perform advanced analytics to extract value. They can produce descriptive and predictive analytics*

*Data Base Administrator – Implements and manages the operational aspects of the cloud*

**Tasks Of A Data Analyst**

*To prepare, model, visualize, analyze, and manage data*



**Getting Started With Power BI**

**Introduction**

*Introduction – Power BI is a tool that connects multiple platforms to work and run together to create interactive insights. Can be used extensively or for personal use.*

*Power BI Desktop – Online SaaS, can produce reports*

**Use Power BI**

*Common Flow Of Activity In Power BI*

1. *Bring data into Power BI desktop and create a report*
2. *Publihs to Power BI service*
3. *Share dashboards with others*
4. *View and interact with shared dashboard and reports*

**Building Blocks of Power BI**

1. *Visualizations – visual representation of data*
2. *Datasets – collection of data that power BI uses to create its visualizations (can be a combo Of different sources*
3. *Reports – A collection of visualizations that appear together, can be many on multiple pages*
4. *Dashboards – Collection of viz’s, one page*
5. *Tiles – one single viz on a dashboard*

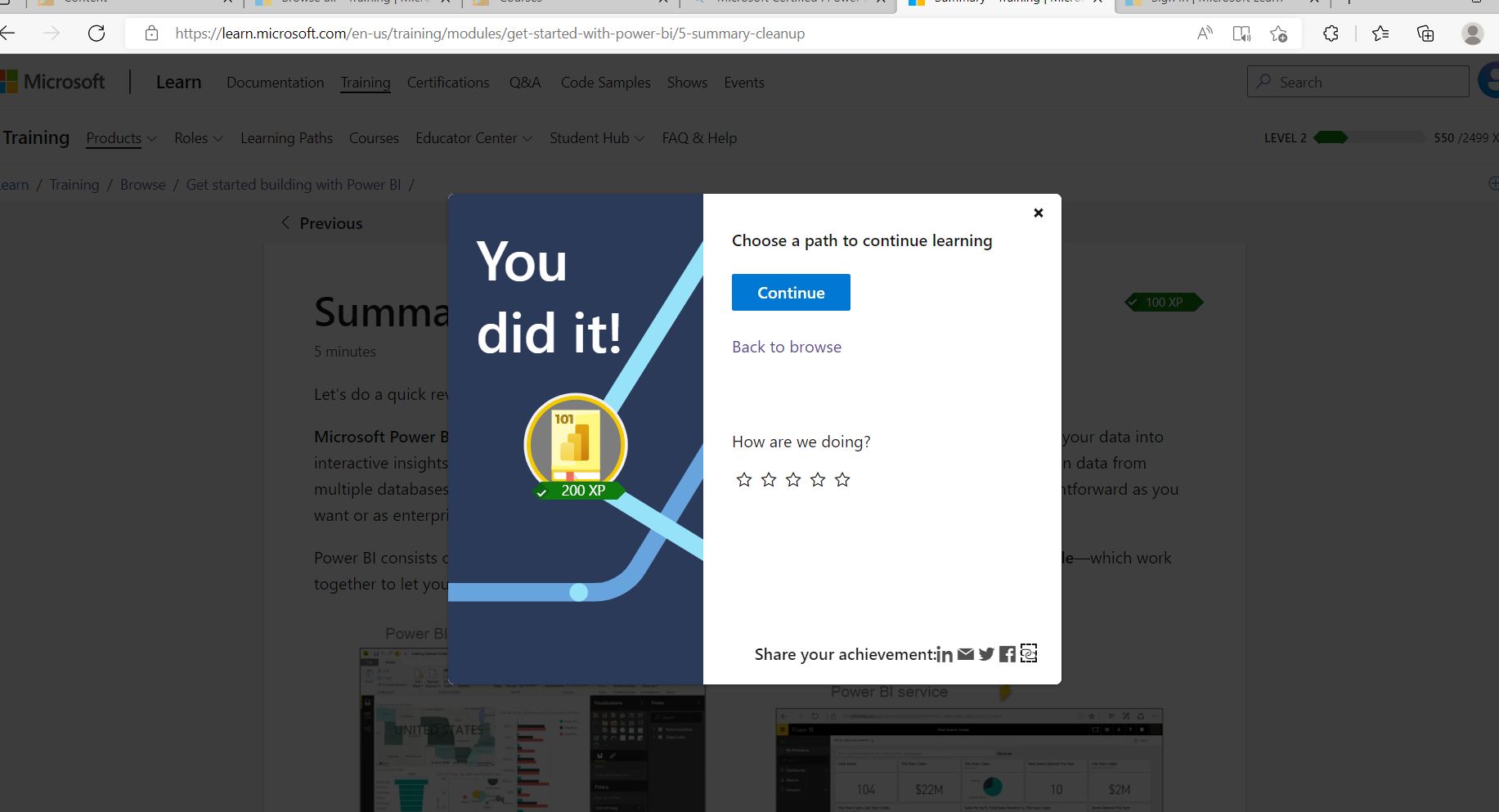
**Create Out Of The Box Dashboards**

*Canvas –* shows you the available sources of data in the Power BI

-There are ready made visuals that are pre-arranged on dashboards

-The update setting to update the dataset in the dashboard is in the update icon in the shape of

A rectangle



**Get Data From Files**

*-Flat file – type of file that has only one data table and each row has the same structure, does not contain hierarchies*

*-You can find a list of available data options when you use the get data feature*

*-Different Locations you can use to export and store in your data (Local, OneDrive for Business,*

*OneDrive Personal, SharePoint)*

**Connect Data In a File**

*-Home tab, Get Data, Select file type option*

*-Then select the file data to import (select the check boxes that you want to bring in to Power BI*

*This activates the load and Transform Data buttons)*

*-To change the source file go to – Data sources settings, Query settings, advanced editor*

**Sign In**

*-You can sign in using three options – Windows, database, or Microsoft account*

**Transform Data**

-Opening data in Microsoft Power Query you can delete unnecessary rows or columns, group your data, remove errors

**Import Data By Writing A SQL Query**

*-Go on the SQL Server Database, select the arrow next to advanced options, in the SQL statement box write your query, and select OK*

**Change the Data Source Settings**

* *Home tab, transform data, data source settings, change source. With power query select the table, select the Data Source Settings on the home settings, close and apply*

**Get data from a NoSQL database**

*-Get data feature in power BI, select more, select Azure category, select Azure cosmos DB, select connect.*

*- On the preview connector window select continue, enter database credentials*

**Import JSON File**

*Connect to the database account, select the table that you want to import from the navigator window, select pop up from the preview pan*

**Get Data From Online Services**

*-Select the Get data feature in Power BI desktop, select the option online services category, select connect, enter your sharepoint URL, when the connection with sharepoint is made in the navigator window, select the list that you want to load in power BI, select load*

**Select A Storage Mode**

*-You can choose from three different types of storage modes – Import (Allows you to create a local power BI copy of your data sheet from the data source, DirectQuery (useful when you do not want to save local copies of your data because your data will not be cached), or Dual (can select some data to be directly imported and some that must be queried)*

**Get Data From Azure Analysis Services**

*-Azure Analysis Services can authenticate to the server, pick the cube you want to use, select what tables you need*

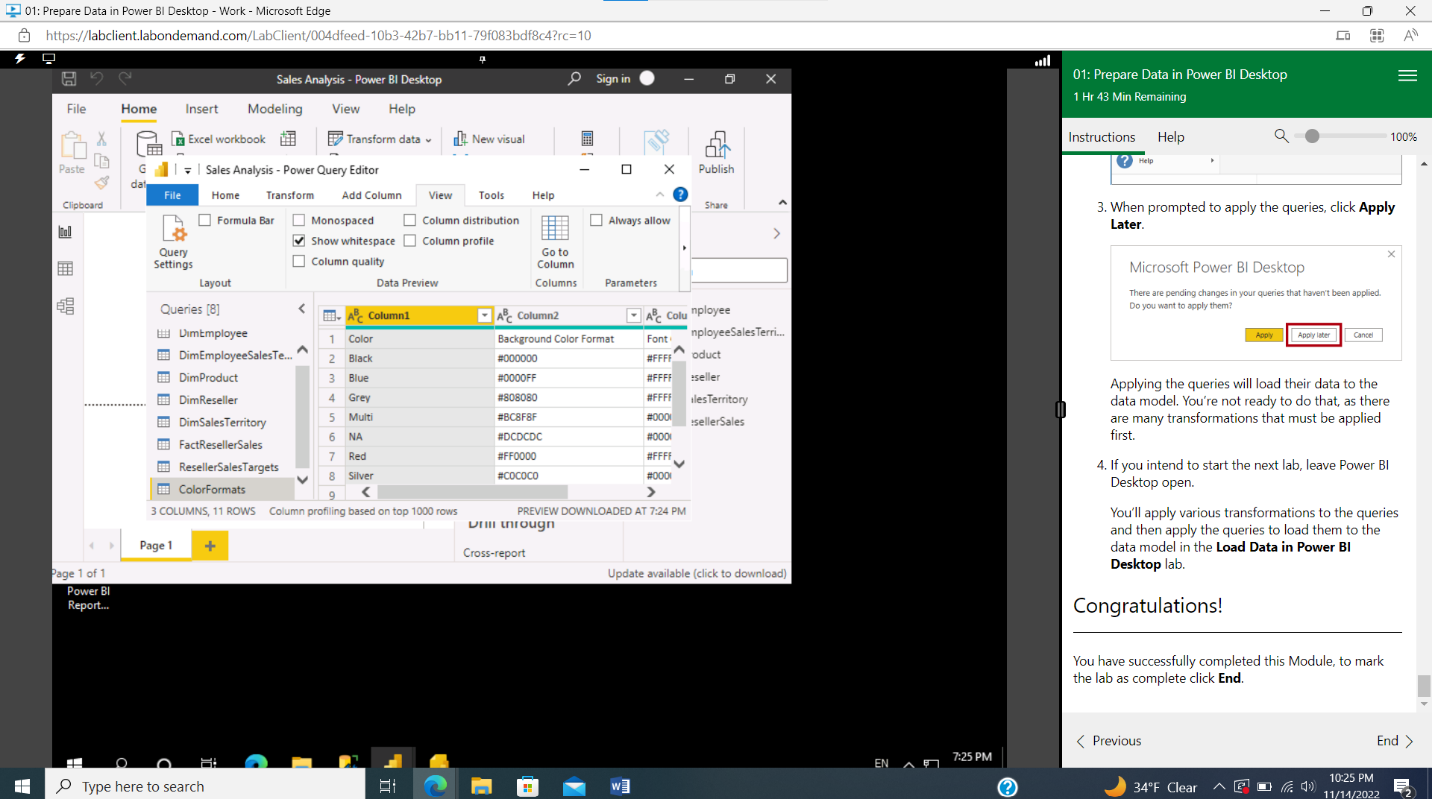
*-Difference – Analysis services cubes have calcuations already in the cube and if you don’t need the entire table, you can query the data directly and you can use multi dimensional expressions*

**Resolve Data Import Errors**

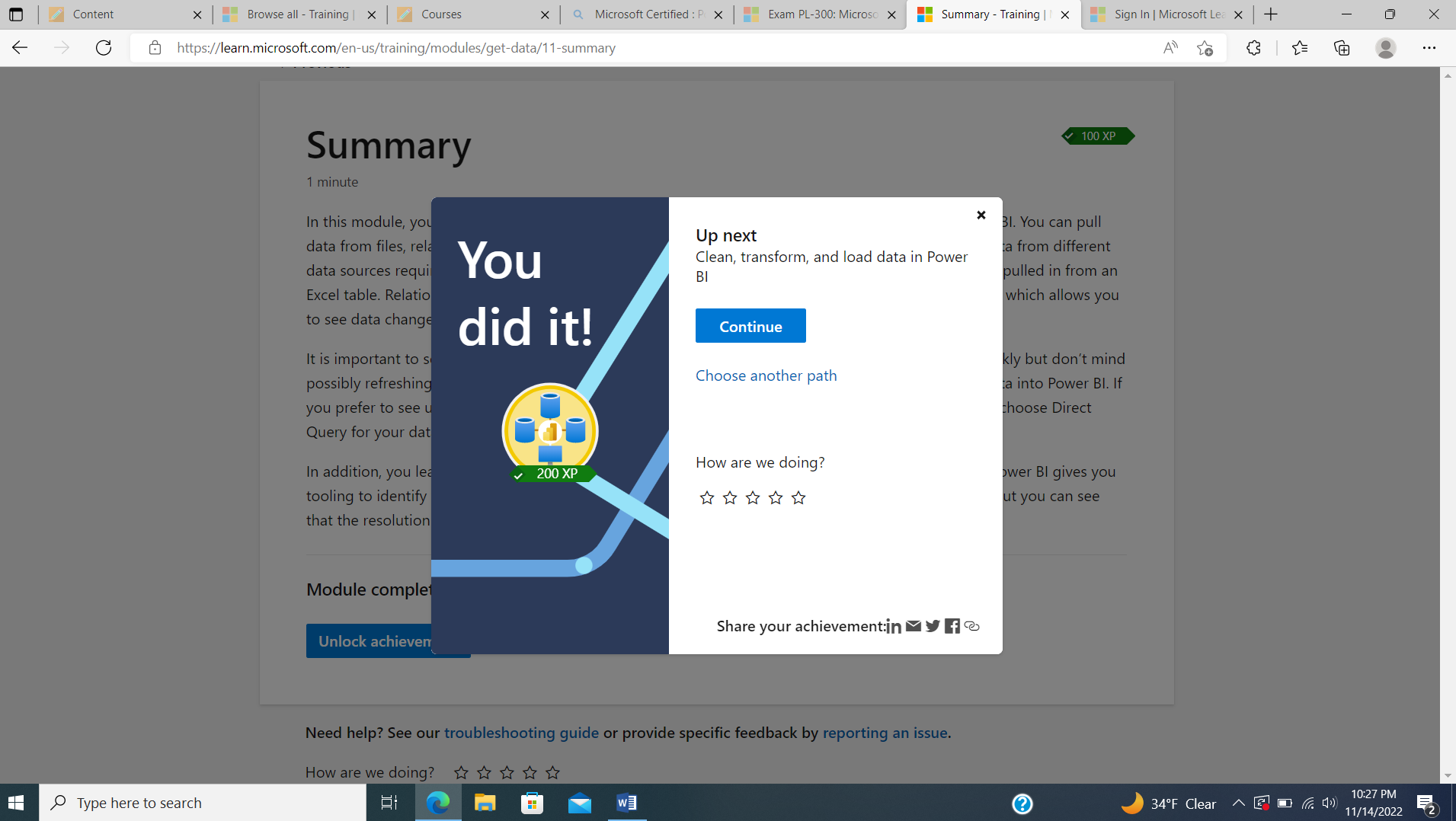
*-“Query timeout expired” – fixed by pulling fewer columns or rows from a single table/*

*- “We couldn’t find any data formatted as table” – soluion is to open your excel workbook, highlight the data that you want to import, press Ctrl-T, verify the oclummn headers and try to import data again*

*- “could not find file” Solution – open power query by selecting the transform data button in power BI, highlight the query that is creating the error, under query settings on the left select the gear icon next to source, change the file location to the new location*

****Lab**

**Achievment**

****

**USING POWER QUERY EDITOR**

**-Query Settings -** You can see a list of your steps on the right side of the screen, many buttons you can use to select, view, and shape your data.

-You can identify column headers and names

-Promote headers by lecting the **Use First Row as Headers** option on the **Home** tab or by selecting the drop-down button next to **Column1** and then selecting **Use First Row as Headers**.

-Rename Columns, Remove top rows, unpivot columns, pivot columns, rename a query, replace values, replace null values, remove duplicates,

**Incorrect Data Types**

**-To change the column data type you:** select **Data Type** in the **Transform** tab, and then select the correct data type from the list.

**Table Combining -**  *merging and appending* - **Home** tab on the Power Query Editor ribbon, select the drop-down list for **Append Queries**. You can select **Append Queries as New**, which means that the output of appending will result in a new query or table, or you can select **Append Queries**, which will add the rows from an existing table into another.

**Column Distribution** - shows you the distribution of the data within the column and the counts of distinct and unique values, both of which can tell you details about the data counts

**Column Profile -**  gives you a more in-depth look into the statistics within the columns for the first 1,000 rows of data. This column provides several different values, including the count of rows, which is important when verifying whether the importing of your data was successful.

**Value Distribution** - graph tells you the counts for each distinct value in that specific column. When looking at the graph in the previous image, notice that the value distribution indicates that "Anthony Grosse"

Graphical user interface, text, application, email

Description automatically generated**Sales Person -** column and that "Lily Code" appears the least amount of times. This information is particularly important because it identifies outliers.

Graphical user interface, application

Description automatically generated

***Charts***

**Matrix**  - looks similar to the table visualization; however, it allows you to select one or more elements (rows, columns, values) in the matrix to cross-highlight other visuals on the report page

**Pie chart / donut chart / treemap -** visualizations show you the relationship of parts to the whole by dividing the data into segments

**Combo Charts -** a combination of a column chart and a line chart that can have one or two Y axes. The combination of the two charts into one lets you:

**Card Visualization -** displays a single value: a single data point. This type of visualization is ideal for visualizing important statistics that you want to track on your Power BI dashboard or report, such as total value, YTD sales, or year-over-year change.

**Multi Row Card Vis -** displays one or more data points, with one data point for each row

***Format and Configure Visualization’s***

**Title -** you can add a title to the visual,

**Background -** can set any color or image as the background for the visual. If you plan to use an image as a background, try to select an image that won't have lines or shapes that would make it difficult for the user to read the data. It is best to keep a white background so the presented data can be clearly seen

**Border -**  you can set a border around the visual to isolate the visual from other elements on the canvas, which helps make it easier for the user to read and understand the data

Diagram

Description automatically generated**General** - be able to set the precise size and place for your visual on your canvas.

**Visuals**

Number of visuals - Using more visuals might have the opposite effect to what you are trying to achieve. More visuals might make your report look too busy, causing users to feel overwhelmed because they don't know where to focus their attention

Position of visuals - It is best practice to place the most important visual in the upper-left corner of your report because your users most likely read left-to-right and top-to-bottom.

Report Accessibility - It is essential that you consider the possibility that your users might have hearing, motor, cognitive, or visual impairments.

Navigation buttons - you can set

Edit Interactions - turns grey to show that it's enabled, and **Filter**, **Highlight**, and/or **None** icons are added to the other visualizations on the report page

Cross Report Drill Through - allows you to contextually jump from one report to another report in the same Power BI service workspace or app

-to enable:  Go to **File > Options and settings > Options**, then scroll down the **Current File** settings and select **Report settings**. In the **Cross-report drillthrough** section, select the check box for **Allow visuals in this report to use drillthrough targets from other reports** and then select **OK**.

Add a Slicer - Provide quicker access to commonly used or important filters.

* Simplify a user's ability to see the current filtered state without having to open a drop-down list.
* Filter by columns that are unneeded and hidden in the data tables.
* Create more focused reports (by putting slicers next to important visuals).
* Defer queries to the data model by using a drop-down slicer, particularly when you are using DirectQuery.
* Slicers are not supported for input fields and drilldown functions.
* **Filters on this visual** - Filters that apply to the selected visual and nothing else. This section only displays if you have a visual selected.
* **Filters on this page** - Filters that apply to the whole page that you currently have open.
* **Filters on all pages** - Filters that apply to all the pages in your report.
* **Drillthrough** - Filters that apply to a single entity in a report.
* **Sort descending** - Sorts the visual by the selected column in the order of greatest value to smallest value.
* **Sort ascending** - Sorts the visual by the selected column in the order of smallest value to greatest value.
* **Sort by** - Sorts the data by a specific column. Hover over this option to display the list of columns that you can select from.

Graphical user interface, text, application, email

Description automatically generatedPerformance Analyzer- allows you to discover how each of your report elements, such as visuals and DAX formulas, are performing. **Performance analyzer** provides you with logs that measure (in time duration) how each of your report elements performs when users interact with them.

Diagram

Description automatically generated

**Dashboards vs. reports**

* Dashboards can be created from multiple datasets or reports.
* Dashboards do not have the **Filter**, **Visualization**, and **Fields** panes that are in Power BI Desktop, meaning that you can't add new filters and slicers, and you can't make edits.
* Dashboards can only be a single page, whereas reports can be multiple pages.
* You can't see the underlying dataset directly in a dashboard, while you can see the dataset in a report under the **Data** tab in Power BI Desktop.
* Both dashboards and reports can be refreshed to show the latest data.

## Q&A feature

The Q&A feature is a tool within Power BI Desktop that allows you to ask natural-language questions about the data.

To locate the Q&A feature, go to your dashboard in Power BI service. Along the top ribbon is the **Ask a question about your data** search box.

- **Question box**

- **Pre-populated suggestion tiles**

## Pin a live page

When you pin a visual, you can add it to a new or an existing dashboard. You can do the same with entire reports; when you pin a report page, all visuals on the report will be pinned to a dashboard and they are also live, meaning that any changes you make on the report will be immediately reflected on the dashboard that you have pinned the report to.

Graphical user interface, diagram, text, application

Description automatically generated

# **Create a paginated report**

Insert table, select insert table, report data, run (home tab), Design’

## Best practices

* What purpose is this report for?
* Who is using the report?
* How can I help people do a better job?
* What is the most important information and how can I highlight it?
* Is this report readable?
* Can people change the elements that they need to if their questions change?
* Do I have visuals that are distracting from the core message of the report?
* Is this report staying focused in a single topic or only a few topics?
* Am I providing all information that the user expects to see in the report?

Diagram

Description automatically generated

## Create a common date table

* Source data
* DAX
* Power Query

## Source data

* Identify company holidays
* Separate calendar and fiscal year
* Identify weekends versus weekdays

## Power Query

You can use M-language, the development language that is used to build queries in Power Query, to define a common date table.

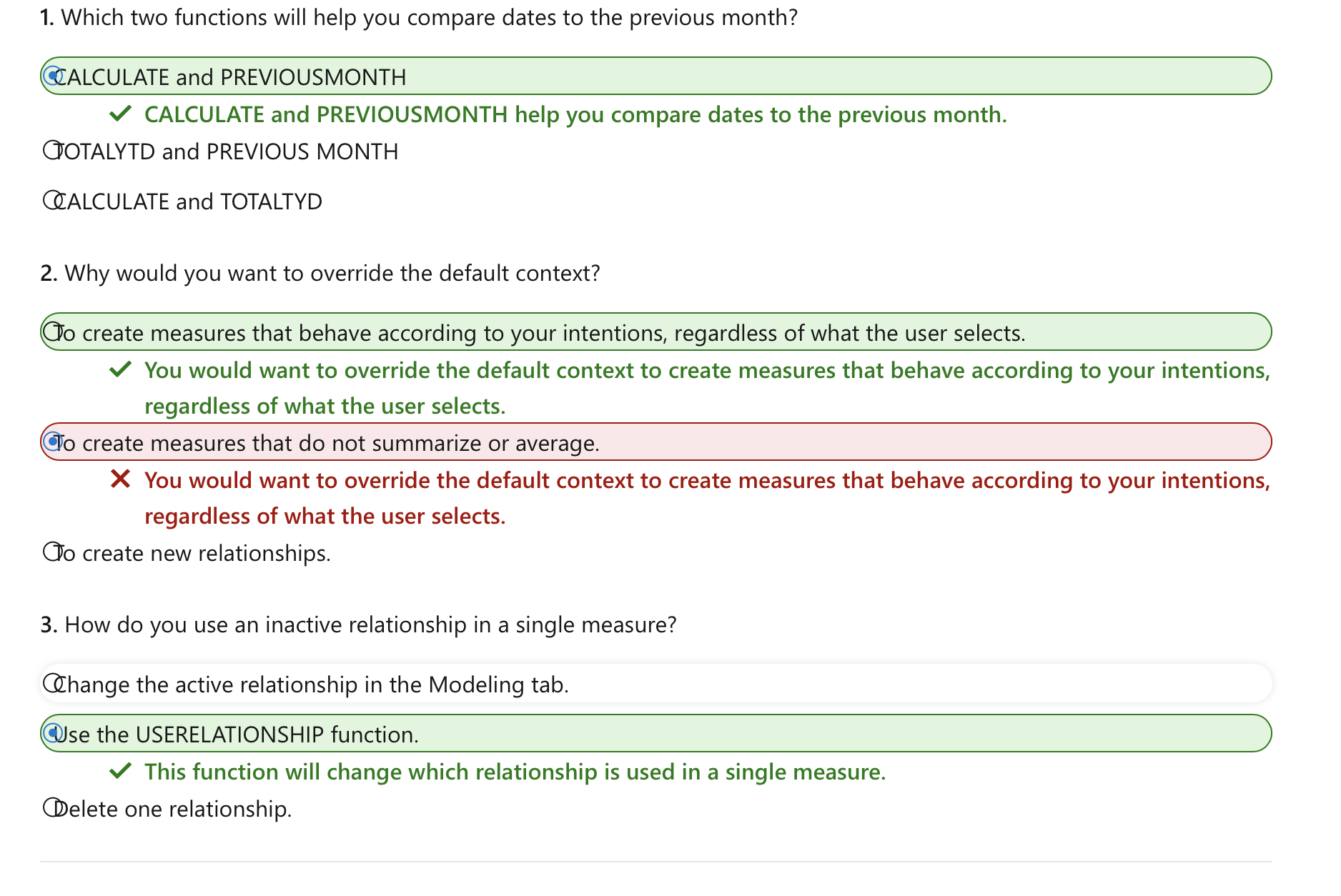
Select **Transform Data** in Power BI Desktop, which will direct you to Power Query. In the blank space of the left **Queries**pane, right-click to open the following drop-down menu, where you will select **New Query > Blank Query**.

## Build your visual

 To complete this task, go to **Model** tab **>** **Manage Relationships**, where you can create relationships between the common date table and the Orders and Sales tables by using the **OrderDate** colum

Graphical user interface, text, application, letter

Description automatically generated



# **Review performance of measures, relationships, and visuals**

Visual cache - When you load a visual, you can't clear this visual cache without closing Power BI Desktop and opening it again. To avoid any caching in play, you need to start your analysis with a clean visual cache.

* Data Engine Cache - When a query is run, the results are cached, so the results of your analysis will be misleading. You need to clear the data cache before rerunning the visual.

To clear the data cache, you can either restart Power BI Desktop or connect DAX Studio to the data model and then call Clear Cache.

**Review results**

* **DAX query** - The time it took for the visual to send the query, along with the time it took Analysis Services to return the results.
* **Visual display** - The time it took for the visual to render on the screen, including the time required to retrieve web images or geocoding.
* **Other** - The time it took the visual to prepare queries, wait for other visuals to complete, or perform other background processing tasks. If this category displays a long duration, the only real way to reduce this duration is to optimize DAX queries for other visuals, or reduce the number of visuals in the report.

## Identify cardinality levels in columns

* **Distinct values count** - The total number of different values found in a given column.
* **Unique values count** - The total number of values that only appear once in a given column.

### **Limitations of DirectQuery connections**

* **Performance** - As previously discussed, your overall user experience depends heavily on the performance of the underlying data source.
* **Security** - If you use multiple data sources in a DirectQuery model, it is important to understand how data moves between the underlying data sources and the associated security implications. You should also identify if security rules are applicable to the data in your underlying source because, in Power BI, every user can see that data.
* **Data transformation** - Compared to imported data, data that is sourced from DirectQuery has limitations when it comes to applying data transformation techniques within Power Query Editor. For example, if you connect to an OLAP source, such as SAP BW, you can't make any transformations at all; the entire external model is taken from the data source. If you want to make any transformations to the data, you will need to do this in the underlying data source.
* **Modeling** - Some of the modeling capabilities that you have with imported data aren't available, or are limited, when you use DirectQuery.
* **Reporting** -- Almost all the reporting capabilities that you have with imported data are also supported for DirectQuery models, provided that the underlying source offers a suitable level of performance. However, when the report is published in Power BI service, the Quick Insights and Q&A features are not supported. Also, the use of the Explore feature in Excel will likely result in poorer performance.

Diagram

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated

Diagram

Description automatically generated

**16/17 hours completed!**