IntelliPaat

Microsoft Power BI
Certification Training (DA-100)

Module 3 – Shaping and Combining Data

Agenda

01 Shaping Data

02 Formatting Data

03 Transforming Data

04 Combining Data

05 Merging Data

06 Appending Data

Agenda

07 Power BI Data Model

08 Optimizing Models for Reports

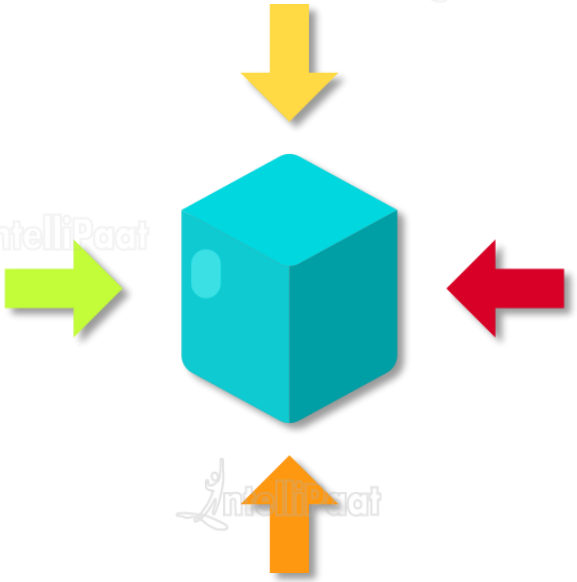
09 What are Hierarchies?

10 Creating Hierarchies

Shaping Data

Shaping Data

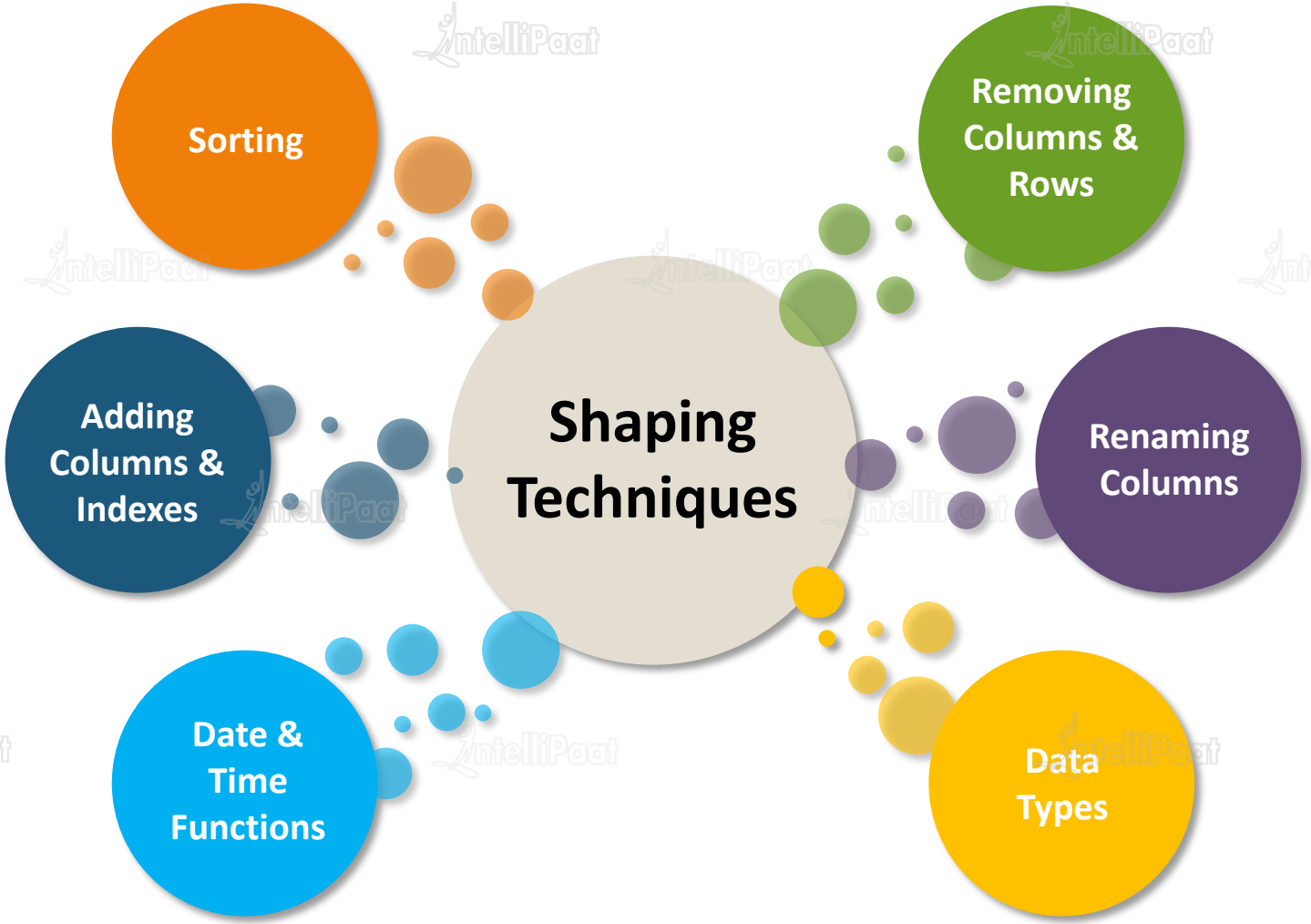
Shaping data is the process of transforming data, in which Power Query Editor loads and presents the data in the best way



Here, we only deal with **View** in Power BI Desktop, so the original data remains unchanged

All the steps that we perform will be captured by Applied Steps in the Query Settings pane

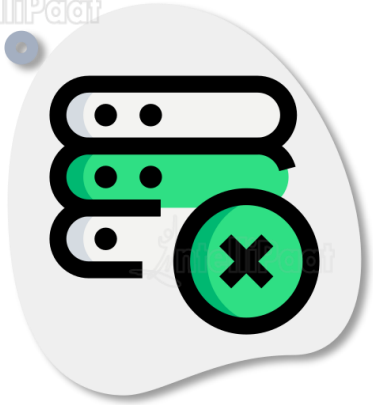
Shaping Techniques



Shaping Data

Removing Rows & columns

If we have a large dataset, we need to remove redundant and unnecessary data to make it as small as possible, which improves the performance of data handling in Power BI



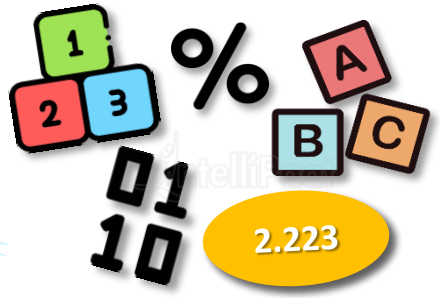
Renaming Columns

Columns should have names that make them easy to work with, and each column should give an adequate description of the data that it contains

Shaping Data

Data Types

Power Query Editor predicts the data type of each column when loaded with the data, but it is always advisable to check the given column data types and then format if necessary



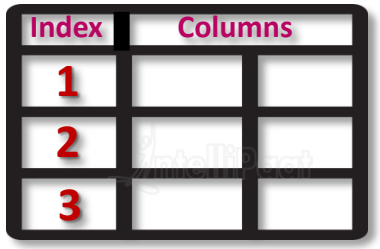
Date & Time Functions

We should also format date-time columns with the correct data type and use these columns to extract the year, quarter, month, week, day, time, hour, etc.

Shaping Data

Adding Columns & Indexes

We can create columns by duplicating an existing column, splitting a column into multiple ones, or creating calculated columns. Indexes can be created with the seed value, starting at 1 or 0, or we can also create a custom index



Sorting

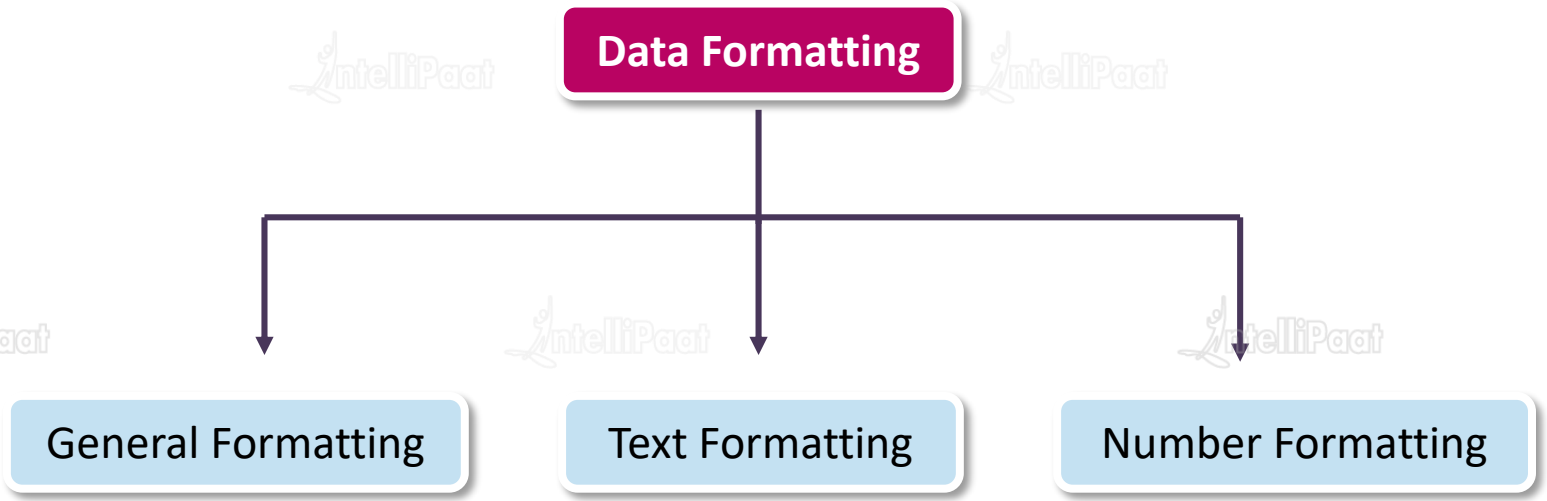
We can also apply the sorting technique based on a particular column. The Home tab of Power Query Editor includes a Sort group with which we can sort A-Z or Z-A

Demo: Shaping Data

Formatting Data

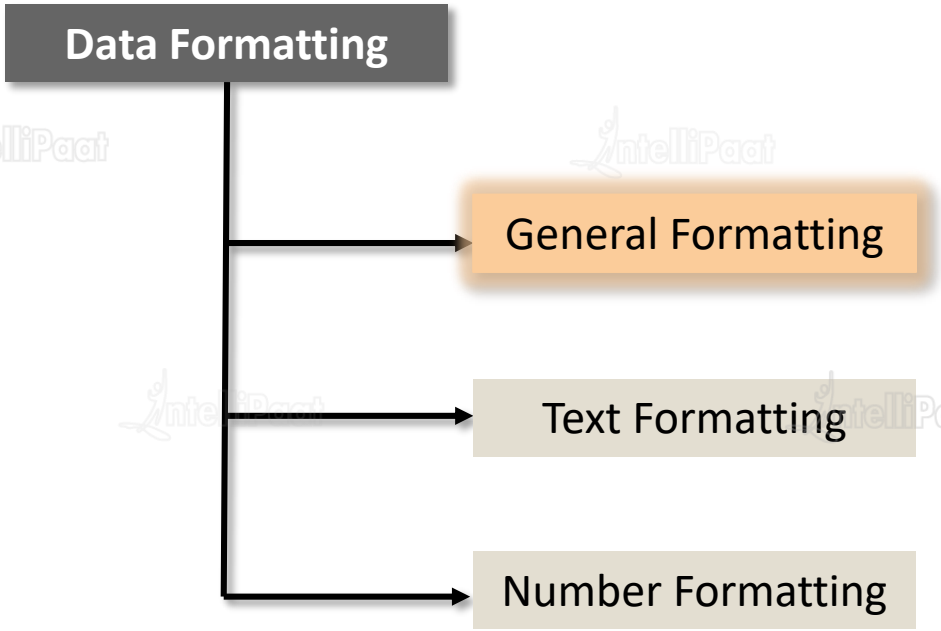
Formatting Data

Formatting data helps in categorizing and identifying the data and making it much easier to work with



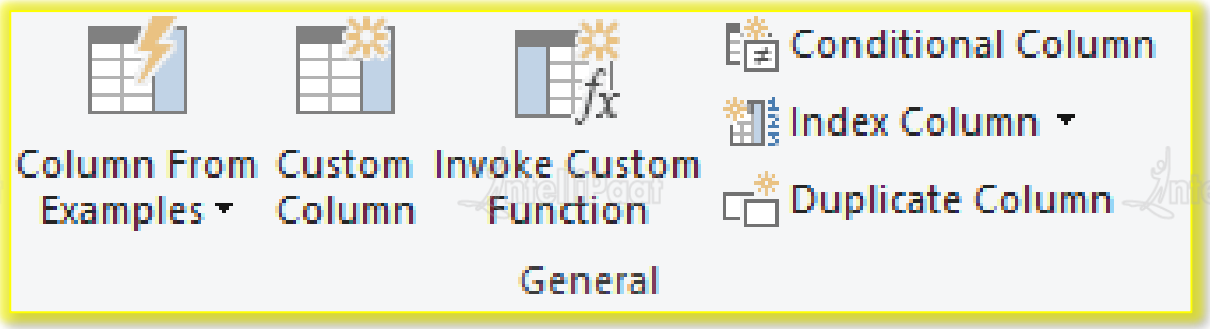
In Power Query Editor, we apply formatting functions to our text and number columns to create consistency and ensuring that the data is well presented

Formatting Data

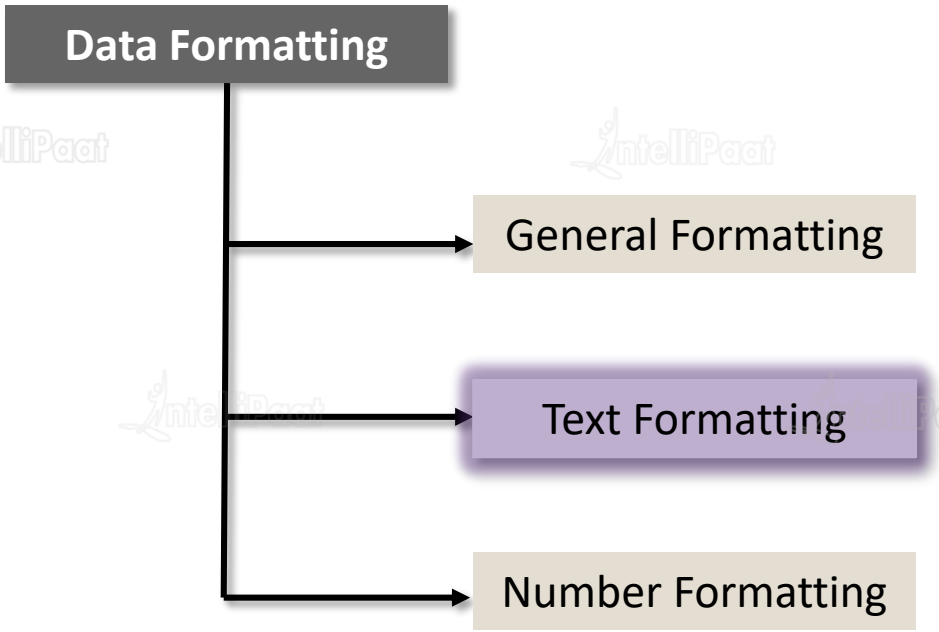


General Column in Power Query Editor helps us create a custom column or a duplicate column. It also lets us add an index column to the table

General formatting includes:

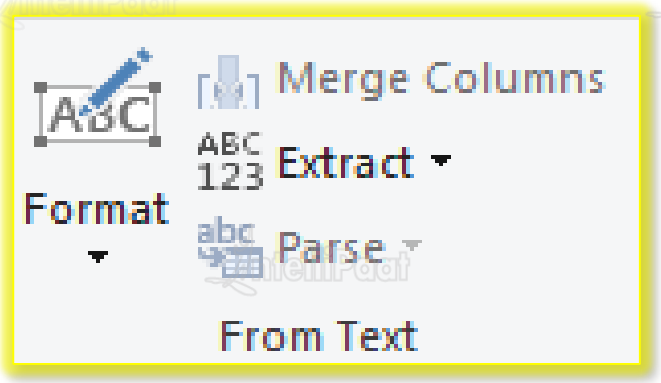


Formatting Data

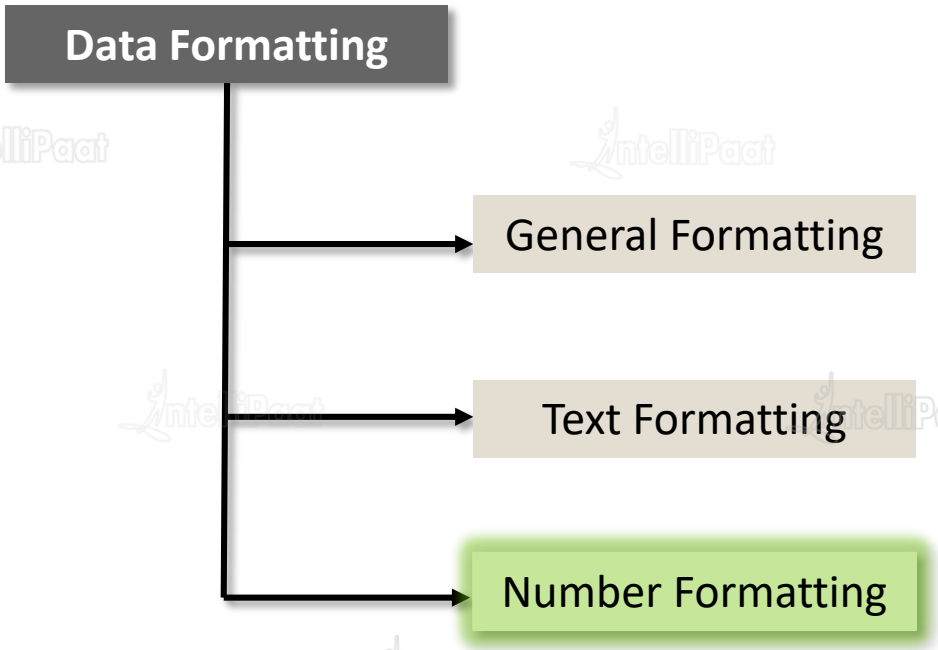


From Text provides options for formatting string values, merging columns, extracting values, and parsing to other formats

Text formatting includes:

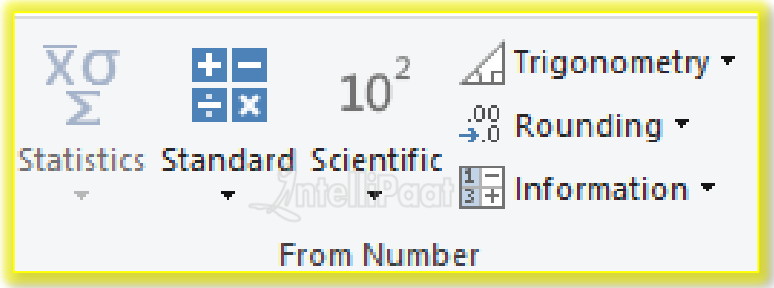


Formatting Data



From Number offers a wide range of formatting functions for number columns, such as
Statistics, Standard, Scientific, Trigonometry, Rounding, and Information

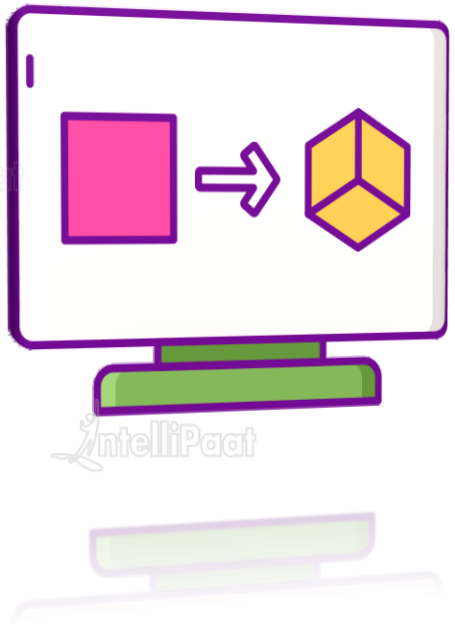
Number formatting includes:



Transforming Data

Transforming Data

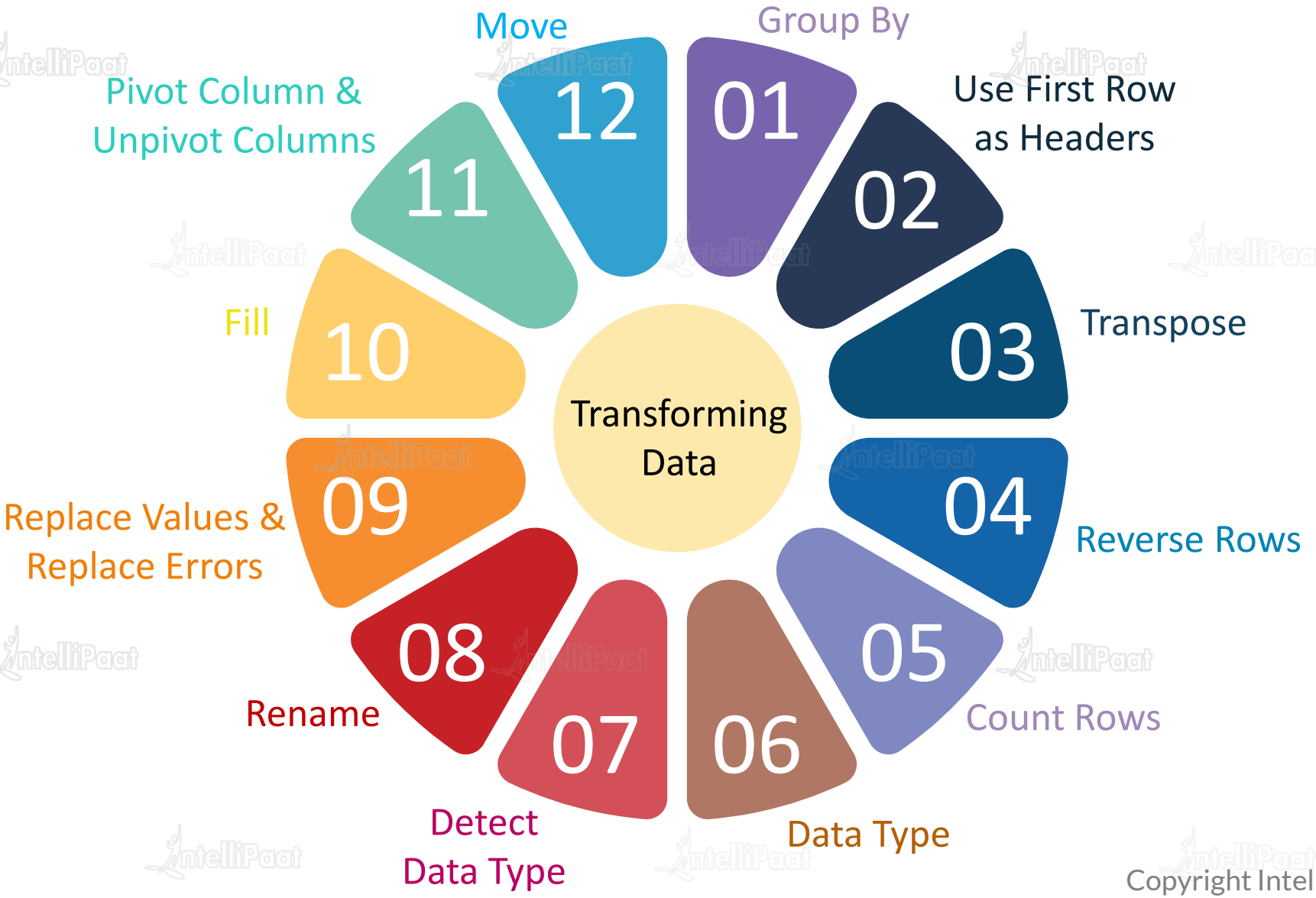
While Power BI is flexible with a variety of data sources that we can import data from, visualizations work best with the data that is in a columnar format



For example, data imported from Excel may be easy for the human eye to visualize but might not be structurally appropriate for Power BI to translate the values into visuals

Power Query Editor offers plenty of functions to transform data into a structure that Power BI can use effectively to build reports

Transforming Data



Transforming Data

1

Use First Row as Headers

If we import columns with numeric values that include a header, Power BI can detect that the first row is a string and that it is the header. This is not so obvious when all of the columns contain string values. We use this function to make the values of the first row as the header

Group By

We can aggregate one or more columns in a table. For that, we have to click on Group By on the Table group and select the columns we want to include

2

Transforming Data

3

Transpose

With Transpose, we can treat rows as columns and columns as rows. This is useful if we import a table from a spreadsheet in a matrix format but do not translate it into a format that Power BI can use easily

Reverse Rows

This function reverses the order of the rows in the table such that the bottom rows come at the top and the top rows go at the bottom

4

Transforming Data

5

Count Rows

We use this function to return the number of rows in the current table. The rows are replaced with the count of the rows

Data Type

The Data Type function is useful for formatting the columns where Power BI has incorrectly guessed the data type

6

Transforming Data

7

Detect Data Type

We can select one or more columns and use the built-in data type detection function. Power BI uses this function to automatically correct the wrongly guessed data types

Rename

To rename a column, we select the column in the table and click on Rename from the Any Column group, or we can right-click on the column and then click on Rename

8

Transforming Data

9

Replace Values & Replace Errors

With these two functions, we can very quickly replace a value or an error in a column with another value. Both functions can work on one or more columns

10

Fill

We can use the Fill function to fill in the null values with values from adjacent cells either in an upward or in a downward direction. This function works only at the column level

Transforming Data

11

Pivot Column & Unpivot Columns

The Pivot Column function takes in values from a selected column and uses them to create a new column. The Unpivot Columns function can also help with this but by converting the selected columns into attribute–value pairs

12

Move

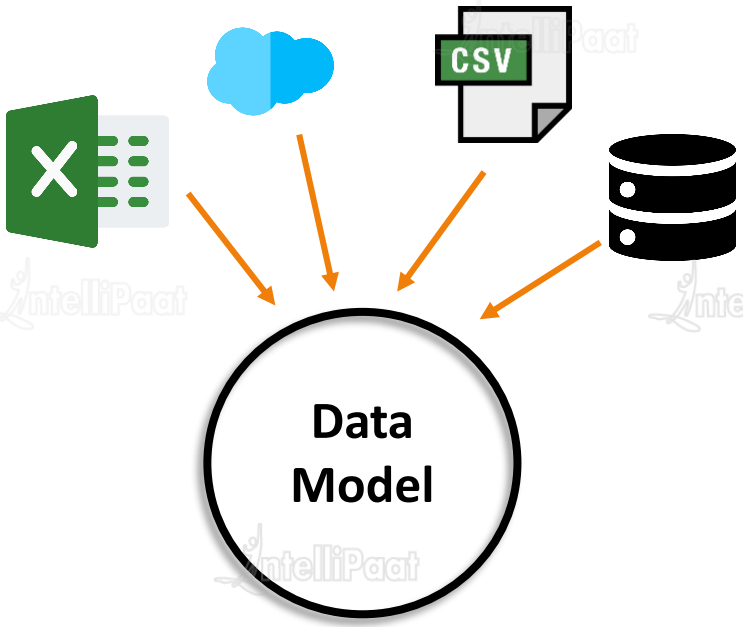
The Move function moves one or more columns to another location in the table. We can move the columns to the left, to the right, to the beginning, or to the end

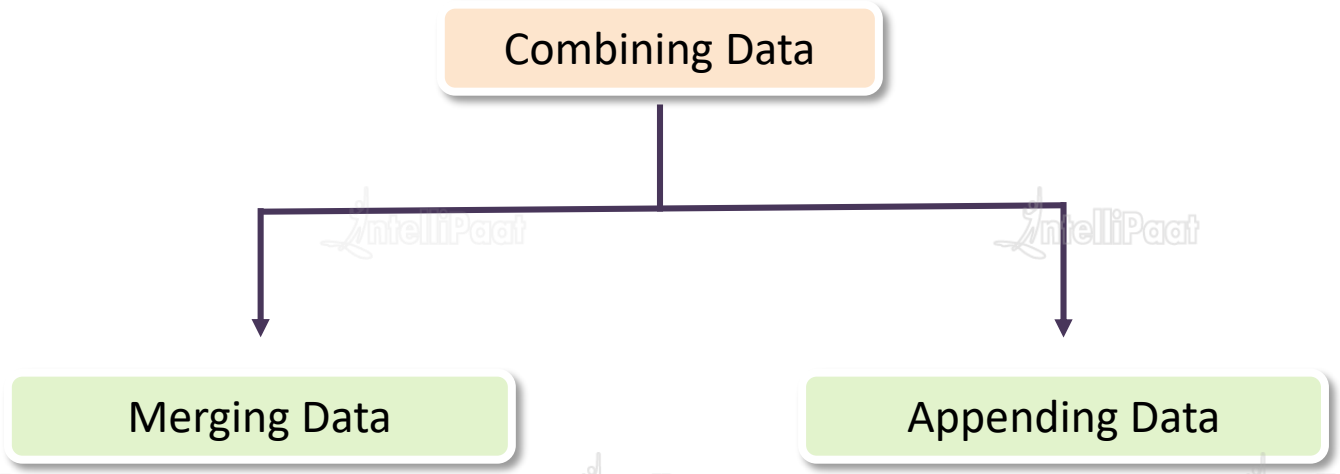
Demo: Transforming Data with Power Query Editor

Combining Data

Combining Data

By using Power BI, we can gather data from different sources and different types into a single dataset to build reports

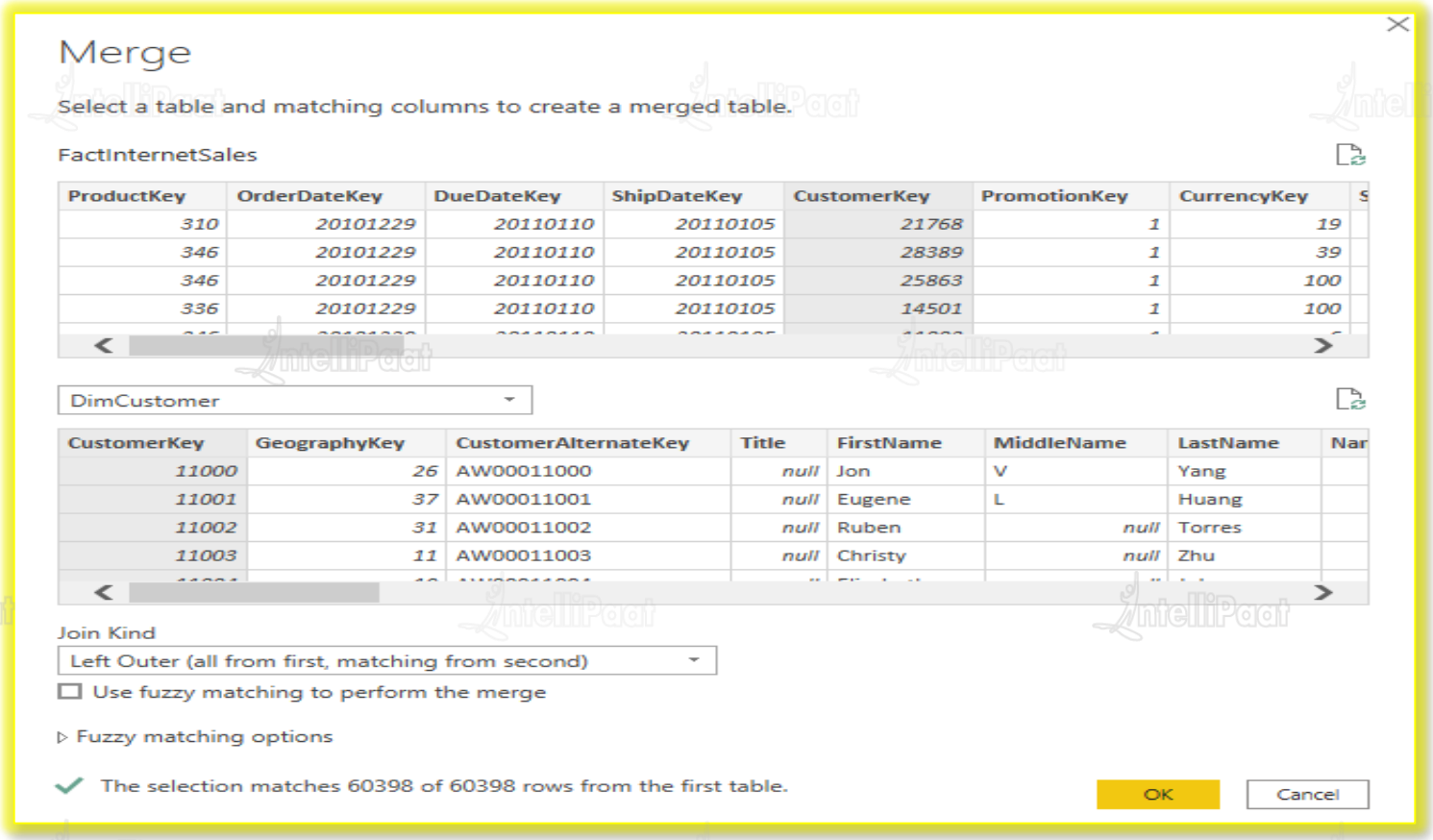




Combining Data - Merging Data

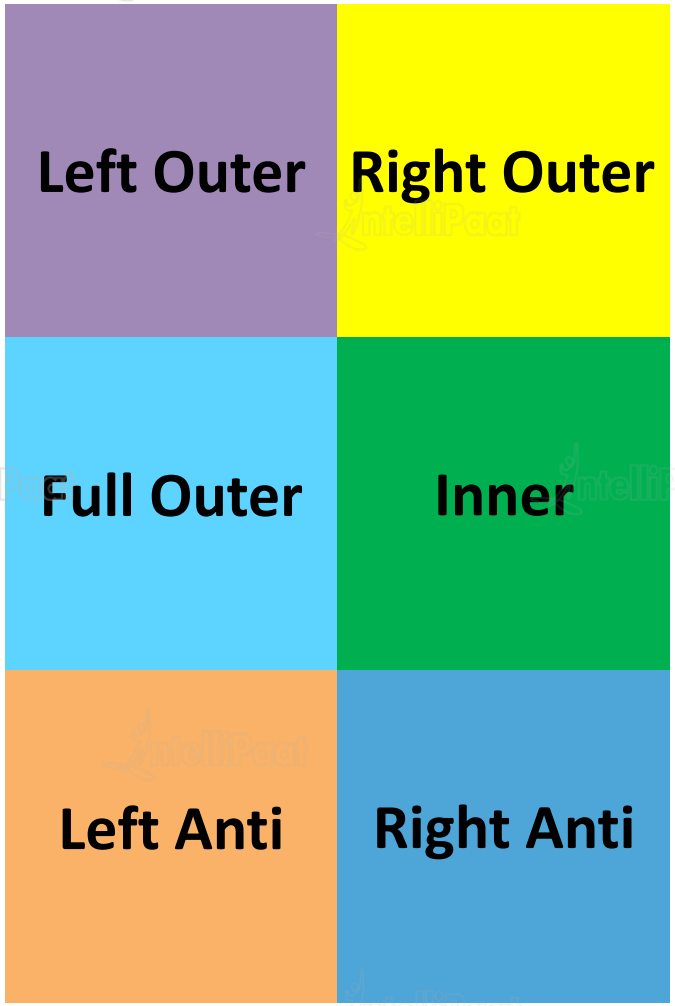
Combining Data - Merging Data

To merge columns of two tables, the two tables must have a joining column, where the values match in order so that they can be combined



Combining Data - Merging Data

These are the Joins used in Merging data



All rows from the first table and the matching rows from the second table

All rows from the second table and the matching rows from the first table

All rows from both tables

Matching rows from both tables

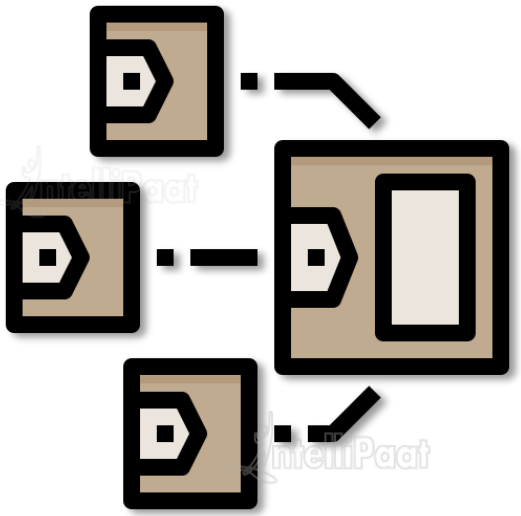
Rows only from the first table

Rows only from the second table

Combining Data - Appending Data

Combining Data - Appending Data

When we append rows, we take rows from one or more tables and add them to the first table



In most situations, the columns and the data types will match. However, we can also append rows between two tables that have all different columns, but the result will be unclean. There will be no values when the number of columns of these tables does not match

If we are appending rows from multiple sources and if the table contains index values that overlap when the data is combined, we combine the data and then create a new index column on the table into which the rows are appended

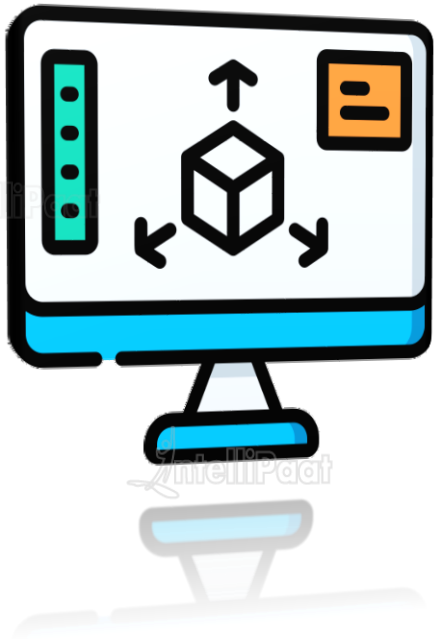
Combining Data - Appending Data

Demo: Combining Data in Power Query Editor

Power BI Data Model

Power BI Data Model

A data model is typically associated with a relational database, such as Microsoft SQL Server. In Power BI Desktop, we connect to multiple data sources and bring data together in the data model by creating relationships between them



In a data model, we can create calculated tables and columns, relationships, hierarchies, etc. and change data types, defaults, and properties

If we do all these right in our model, then the creation of reports will be a much smoother process, and it will produce more accurate results

Power BI Data Model

Things to keep in mind while developing a model in Power BI:

Data types

A wrong data type might cause incorrect results in our visuals. When importing numbers, we have to be accurate and determine whether we need precise rounding

Fact and Dimension Tables

The data is stored in fact tables, and the descriptive information about the data is stored in dimension tables

Cross-filtering

When we turn on bi-directional cross-filtering in our relationships, it enables the tables in our star schema to operate as if held in a single table, and we can join and aggregate values between the dimension tables

Reduced Size of the Dataset

To reduce the volume of the data, we can join the tables, apply filters, or remove the data that will not be used in visualization

Optimizing Models for Reports

Optimizing Models for Reports

Data is frequently in a raw format when we import it into Power BI, especially if we have taken it directly from a database

When we combine data from different data sources into the Power BI model, there is a high chance that the data will have different formats and data types

Optimizing the data makes it more consistent and helps us work with it more efficiently, focusing on the information we need

Optimizing Models for Reports

Three techniques to optimize our data in the model:

Hide Fields

It is good to hide fields that we are not going to use in our visuals. Hiding a field removes the field name from the Fields pane, but neither the column nor the underlying data will be deleted

Sort Data

Power BI automatically sorts data, but when we want the data to be ordered in a particular way, this can also be done. It makes data analysis much easier for users

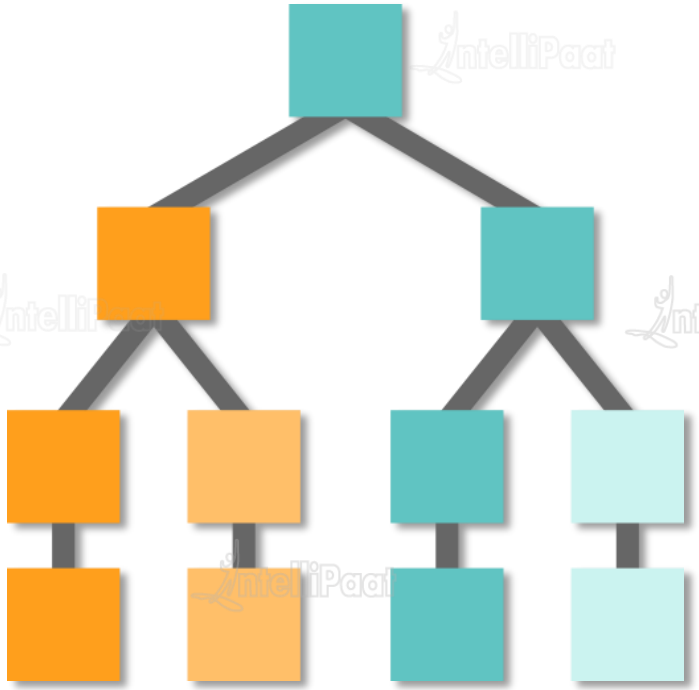
Format Data

Changing data types and formatting the data are good ways of optimizing our data, and it helps in presenting the same with clarity in reports and dashboards

What are Hierarchies?

What are Hierarchies?

A hierarchy is a set of related fields grouped together to drill down from one level to the next

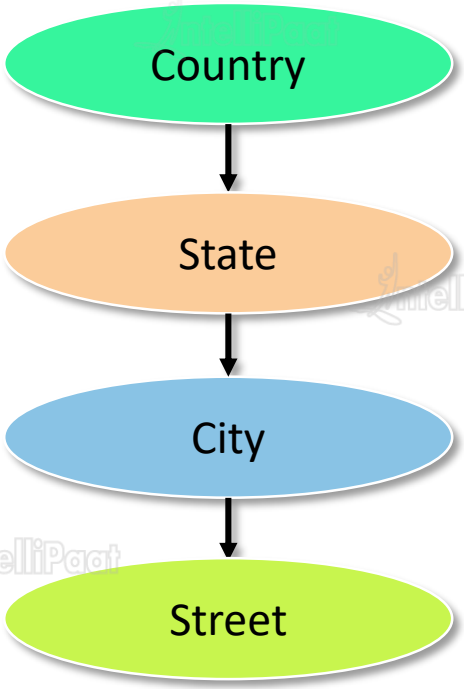


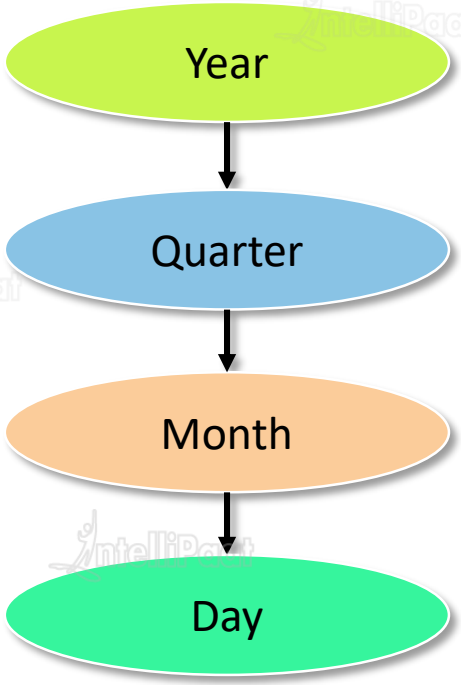
Each level of the hierarchy is contained within the next level, and it cannot exist independently

Power BI automatically creates a hierarchy on date-time fields. However, we can also create our own hierarchies within the model to suit our requirements for analysis

What are Hierarchies?

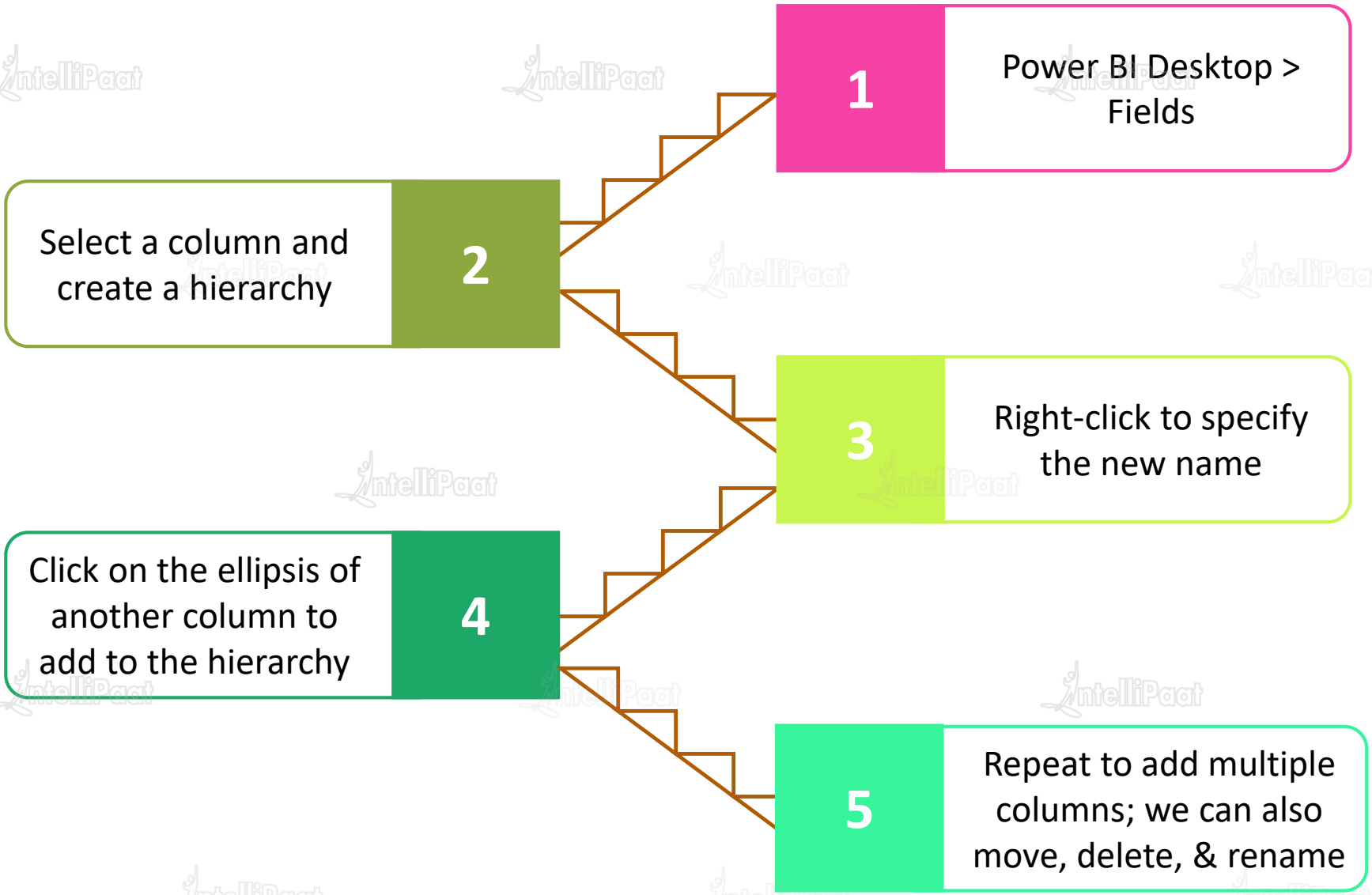
Examples of hierarchies that we can create in Power BI to drill down from top to bottom:





Creating Hierarchies

Creating Hierarchies



Demo: Creating Hierarchies

India: +91-7847955955

US: 1-800-216-8930 (TOLL FREE)

sales@intellipaat.com

24/7 Chat with Our Course Advisor