

Connecting and Preparing Data

Prepared by Eng. Ahmed Hussein, MBA,
PfMP®, PMP®, CMP APMG

Course: From Data to Decisions using Power BI

Lecture 02



Oriental
Lecture

What we will cover

1. Connecting to different data sources
2. Using Power Query
3. Clean and Transform data

Lecture

02

1. Connecting to different data sources

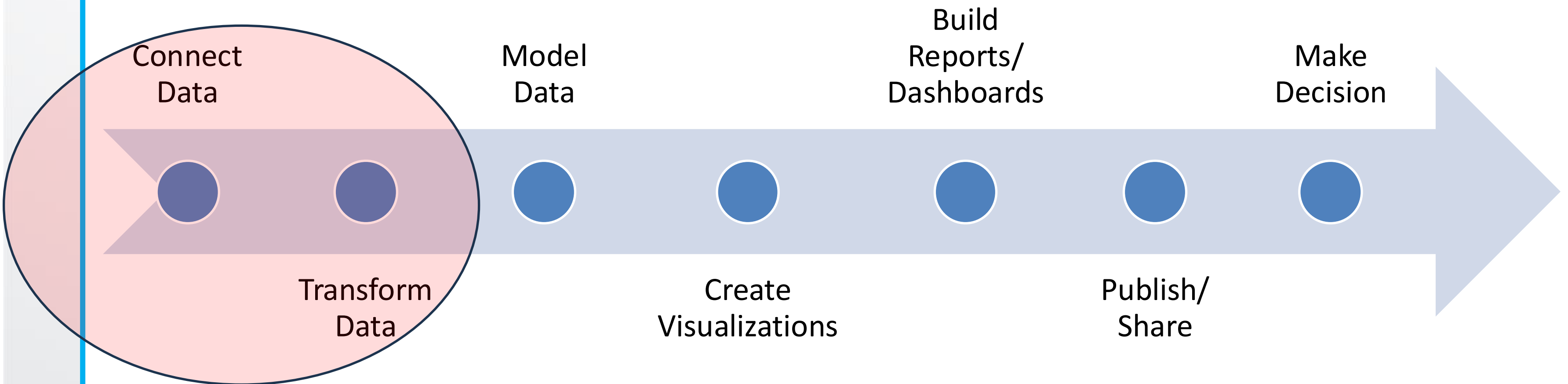


"Garbage In, Garbage Out"

Poor data leads to poor decisions.

Lecture
02

Power BI Workflow (The Analytics Journey)



Each stage adds value — skip one and the chain breaks

Lecture
02

What is Data Quality?

- Data quality refers to four key dimensions:

1

Accuracy

Data reflects the
real world
correctly

2

Completeness

No critical values
are missing

3

Consistency

Same data, same
answer
everywhere

4

Reliability

Trustworthy over
time and sources

Lecture
02

Why Data Quality Matters?

- Poor data causes real business damage:



Incorrect KPIs

Metrics mislead
leadership



Bad Decisions

Actions based on
wrong data



Loss of Trust

Stakeholders lose
confidence



Reporting Delays

Teams waste time
fixing errors

Lecture
02

Common Data Problems

- Issues that compromise your analysis:

?

Missing
Values

2x

Duplicates

A/a

Inconsistent
Formats

31/13

Incorrect
Dates

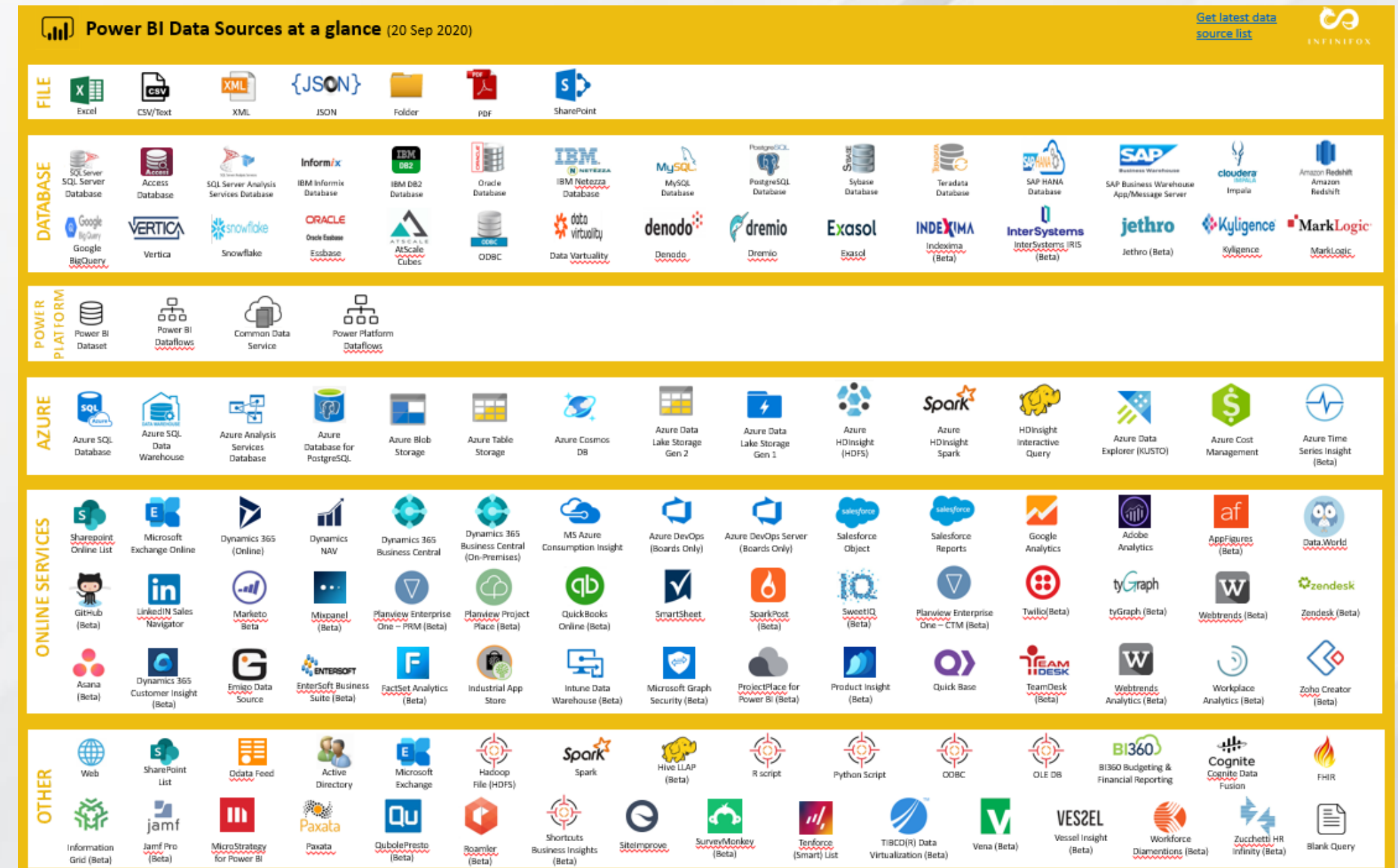
#!

Wrong
Data Types

Power BI Connectivity

Connecting to data through Power BI is easy, and can be achieved with a variety of data source formats:

- Files
- Databases
- Online Services
- SaaS
- Other



Power BI Connectivity

Most Common Data Sources

1

Files

Excel, CSV,
XML

2

Database

SQL, Oracle,
MySQL

3

Cloud

Azure,
SharePoint

4

Websites

Web
scraping

5

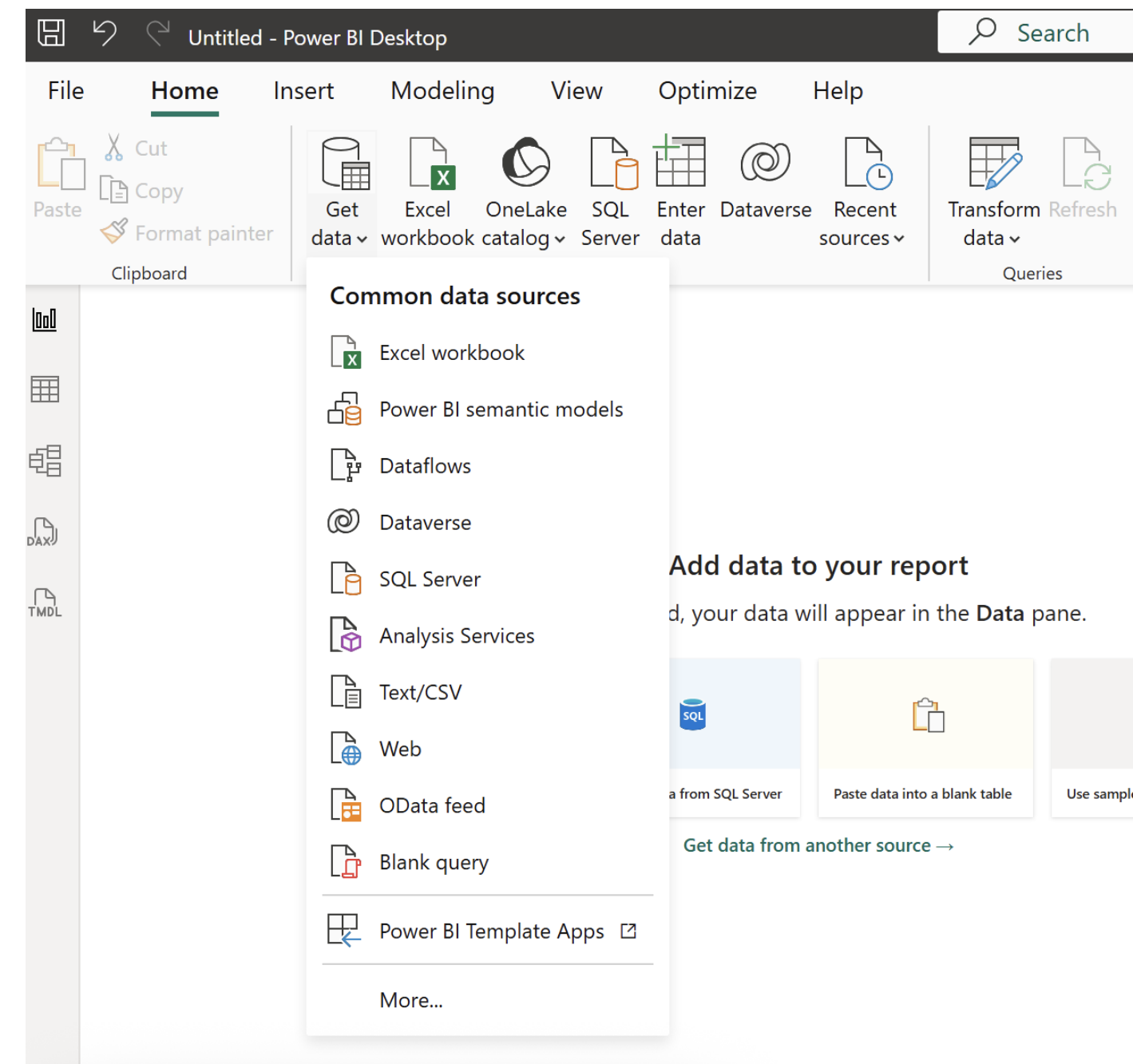
APIs

REST, OData

Lecture 02

Importing Data to Power BI

- Open Power BI Desktop → New Report
- Click Home → Get Data.
- Select the Source
- Connect to your file/database.



Lecture 02

Previewing Data

- Before loading data into Power BI, we should always inspect and prepare it.
- Loading poor quality data directly into the report can lead to misleading dashboards.
- Navigator Window Preview tables and sheets before loading into your model
- . Choose Transform Data instead of Load to open Power Query

Navigator

Display Options ▾

☐ Sales_Transactions_Q1.xlsx [2]
☒ SalesTransactionsQ1
☐ Sales_Transactions_Q1

SalesTransactionsQ1

Transaction ID	Date	Customer ID	Product ID	Region
T0001		46029 C294	P206	Dubai
T0002		46090 C108	P178	alexandria
T0003		46055 C233	P344	singapore
T0004		46082 C327	P273	Berlin
T0005		46110 C433	P223	new york
T0006		46055 C121	P208	Dubai
T0007		46078 C464	P019	dubai
T0008		46028 C249	P401	Cairo
T0009		46069 C454	P047	NY
T0010		46097 C477	P269	Riyadh
T0011		46072 C381	P121	NY
T0012	2026/03/17	C274	P192	alexandria
T0013		46041 C367	P189	Riyadh
T0014		46053 C197	P159	London
T0015		46073 C470	P131	Riyadh
T0016		46105 C018	P141	ALEXANDRIA
T0017		46035 C358	P134	Singapore
T0018		46026 C329	P329	Giza
T0019	11.01.2026	C128	P265	Giza
T0020		46064 C102	P044	new york
T0021		46094 C321	P118	ALEXANDRIA
T0022		46102 C134	P077	alexandria
T0023		46081 C254	P061	dubai

Lecture

02

2. Using Power Query



Lecture
02

What is Power Query?

- Power BI's data preparation engine
- "The workshop where data gets fixed."
- Benefits:

No coding required

Automated Transformations

Faster reporting

Reduced manual work

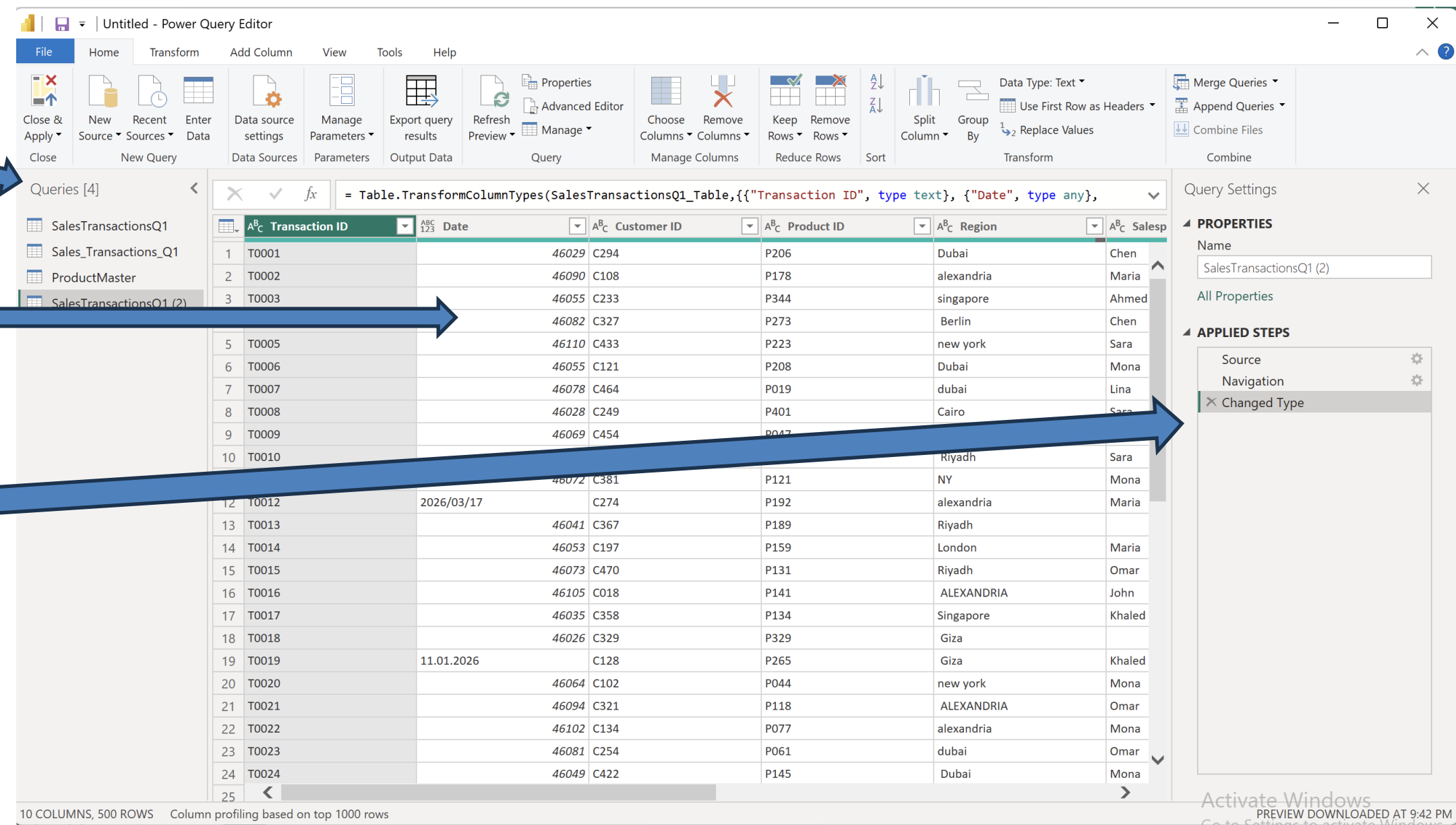
Lecture 02

Power Query Interface Main Area

- Queries Pane

- Data Preview

- Applied Steps



Lecture 02

What is Applied Steps in Power Query?

- Applied Steps in Power Query are the list of transformations Power BI records every time you clean or change your data.
- The most important benefit is that Power BI remembers your cleaning process as:
 - Removing duplicate transactions
 - Fixing fields
 - Changing formats
 - Replacing blank values
- Applied Steps make data preparation repeatable, transparent, and reliable.

Action	Example
Connecting to data	Importing an Excel file
Promoting headers	Making the first row become column names
Changing data types	Converting Revenue from text to number
Removing duplicates	Keeping only one record per Transaction ID
Replacing values	Changing "Alex" to "Alexandria"
Trimming spaces	Removing extra spaces from names
Filtering rows	Keeping only Q1 transactions

Lecture 02

What is Applied Steps in Power Query?

- Applied Steps in Power Query are the list of transformations Power BI records every time you clean or change your data.
- The most important benefit is that Power BI remembers your cleaning process as:
 - Removing duplicate transactions
 - Fixing fields
 - Changing formats
 - Replacing blank values
- Applied Steps make data preparation repeatable, transparent, and reliable.

Action	Example
Connecting to data	Importing an Excel file
Promoting headers	Making the first row become column names
Changing data types	Converting Revenue from text to number
Removing duplicates	Keeping only one record per Transaction ID
Replacing values	Changing "Alex" to "Alexandria"
Trimming spaces	Removing extra spaces from names
Filtering rows	Keeping only Q1 transactions

Lecture 02

Using SQL Queries in Power Query

- When connecting Power BI to a database, Power Query can use **SQL queries** to retrieve specific data instead of importing entire tables.
- This helps users bring only the data needed for analysis.
- **When to Use SQL Queries**

Use Case	Example
Filter data before import	Import only sales transactions for 2026
Select specific columns	Import Date, Customer ID, Product ID, and Revenue only
Join database tables	Combine Sales and Customer tables before loading
Reduce data volume	Avoid importing millions of unnecessary rows
Improve performance	Let the database process part of the work

Lecture
02

Using SQL Queries in Power Query

SQL queries are mainly used when connecting to database sources such as:

- SQL Server
- Oracle
- MySQL
- PostgreSQL

SQL EXAMPLE

```
SELECT
    TransactionDate,
    CustomerID,
    ProductID,
    Region,
    Revenue
FROM SalesTransactions
WHERE TransactionDate >= '2026-01-01';
```

SQL queries help Power Query import the right data, not all data.

Lecture
02

Refreshing Data in Power BI

- Power BI enables online and on-premises data refresh for datasets uploaded from:
 - Power BI Desktop files
 - Excel workbooks, where Power Query or Power Pivot is used to query data
- Data refresh can be
 - on-demand, or
 - scheduled
- Minimum schedule frequency is hourly
- Achieved via a gateway
 - Personal
 - Enterprise (Available only on Pro License)

Lecture
02

What Happens during Refresh

When you click refresh, Power BI repeats the process:

Step	What Power BI Does
1	Connects to the original data source
2	Imports the latest data
3	Applies the saved Power Query steps
4	Updates relationships and calculations
5	Updates all visuals in the report

Lecture

02

3. Clean and Transform Data

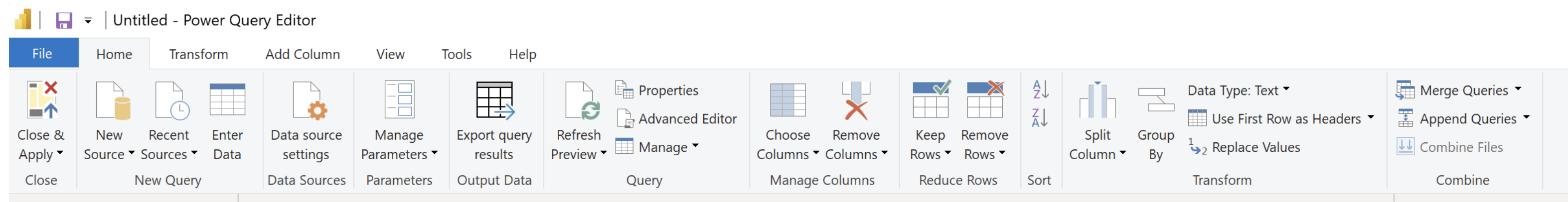


Lecture
02

Data Preparation in Power Query

- Power Query is mainly used to clean, transform, combine, and prepare data before loading it into Power BI.
- The three most important tabs are:

Tab	Main Purpose
Home	Import, manage, combine, and load data
Transform	Change existing columns and clean existing data
Add Column	Create new columns based on existing data



Lecture 02

Key Functions Under Home

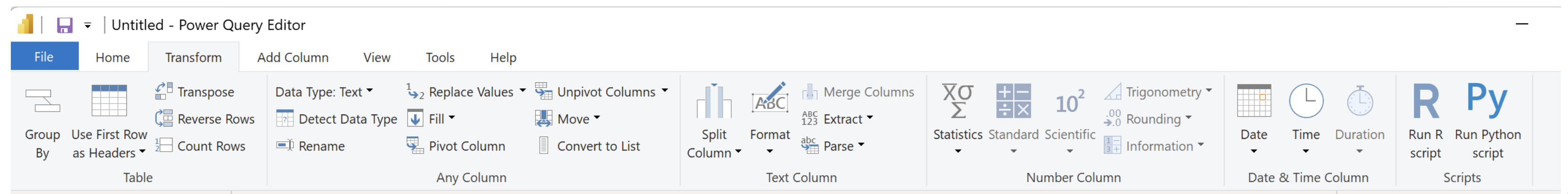
- The Home tab contains the most commonly used functions in Power Query.

Function	What It Does	Business Example
Close & Apply	Saves Power Query changes and loads the cleaned data into Power BI.	Load the cleaned sales dataset into the report view.
New Source	Connects to a new data source such as Excel, CSV, database, or folder.	Import sales data from Excel and customer data from CSV.
Choose / Remove Columns	Keeps only the columns needed for analysis and removes unnecessary ones.	Remove unused notes or system columns from a sales file.
Remove Rows / Remove Duplicates	Removes unwanted rows, blank rows, errors, or duplicate records.	Remove repeated sales transactions using Transaction ID.
Merge Queries	Combines two tables side by side using a common field.	Merge Sales with Customer Master using Customer ID.
Append Queries	Combines similar tables by stacking them together.	Combine January, February, and March sales files into one table.

Lecture 02

Key Functions Under Transform

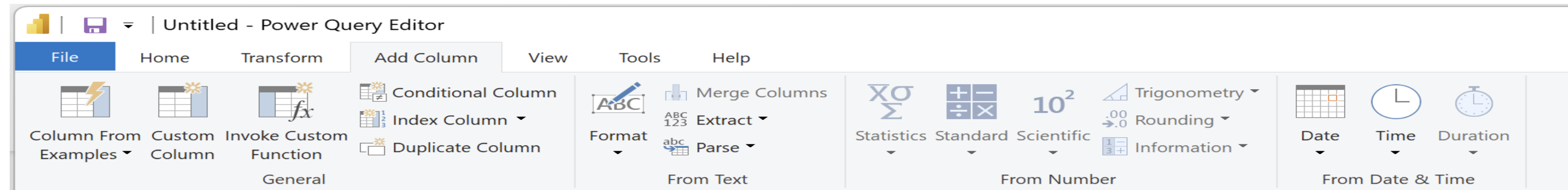
Function	What It Does	Business Example
Data Type	Changes columns to the correct format such as Date, Text, Number, or Currency.	Convert Revenue from text to decimal number.
Format	Cleans text by trimming spaces, cleaning hidden characters, or capitalizing words.	Convert “ cairo ”, “CAIRO”, and “cairo” into “Cairo”.
Replace Values	Replaces specific values to standardize data.	Replace “Alex” with “Alexandria” or blanks with “Unknown”.
Split Column	Splits one column into multiple columns.	Split Full Name into First Name and Last Name.
Unpivot Columns	Converts multiple columns into rows for better reporting.	Convert Jan, Feb, Mar columns into Month and Sales rows.



Lecture 02

Key Functions Under Add Column

Function	What It Does	Business Example
Custom Column	Creates a new calculated column using a formula.	Create Profit = Revenue - Cost.
Conditional Column	Creates a new column based on business rules.	Classify customers as High, Medium, or Low Value based on revenue.
Column From Examples	Creates a new column by learning from examples you type.	Extract first name from a full customer name.
Index Column	Adds a sequence number to the rows.	Add row numbers to transaction records for tracking.
Date Columns	Extracts date parts such as Year, Month, Quarter, or Day.	Create Month and Quarter columns from the Sales Date.



Lecture
02

Examples

- Removing Duplicates:

BEFORE

Name	Revenue
Ahmed	\$5,000
Sara	\$3,000
Ahmed	\$5,000

>>>

AFTER

Name	Revenue
Ahmed	\$5,000
Sara	\$3,000

- Standardizing Text

BEFORE

CAIRO
cairo
Cairo

>>>

AFTER

Cairo

Lecture
02

Examples

- Handling Missing Values (Options):

1

Remove rows

2

Replace values

3

Fill missing data

- Correcting Data Types:

ABC

Text

123

Number

01/06

Date

\$\$\$

Currency

Lecture
02

Examples

- Splitting Columns

Ahmed Mohamed Hassan

SPLIT

First Name

Ahmed

Middle Name

Mohamed

Last Name

Hassan

- Adding New Columns

Full Name

First + Last

Profit

Revenue - Cost

Region Category

IF logic

Lecture 02

Examples

- Unpivot Columns

Product	Jan	Feb	Mar
Widget A	\$5K	\$3K	\$7K

PROBLEM

Difficult to analyze, filter, or visualize

Product	Month	Sales
Widget A	Jan	\$5K
Widget A	Feb	\$3K
Widget A	Mar	\$7K

Benefits

Easier reporting

Easier filtering

Better visualizations

Lecture 02

Examples

- Merge Query
Sales Table

Transaction ID	Customer ID	Revenue
T001	C001	5000
T002	C002	3000

Customer Table

Customer ID	Customer Name	Segment
C001	Delta Trading	Corporate
C002	Nile Foods	SME

After Merge

Transaction ID	Customer ID	Customer Name	Segment	Revenue
T001	C001	Delta Trading	Corporate	5000
T002	C002	Nile Foods	SME	3000

Lecture 02

Moving from Data Preparation to Reporting

After cleaning & transforming data in **Power Query**, you need to apply the changes so Power BI can use the prepared data for reports, visuals, and dashboards.

- In Power BI, the main report file is saved as a: **.pbix file**

Home Tab → Close & Apply
What “Close & Apply” Does:

Action	What Happens
Closes Power Query Editor	Takes you back to the main Power BI report screen.
Applies all transformation steps	Runs all cleaning steps recorded in Applied Steps.
Loads the cleaned data	Makes the prepared tables available in Power BI.
Updates the data model	Allows tables, columns, and relationships to be used in visuals.
Refreshes report fields	The cleaned columns appear in the Fields/Data pane.

Lecture 02

Validating Data After Power Query

- Before building reports and dashboards, we need to confirm that the cleaned data is accurate, complete, and ready for analysis.

Validation Area	What to Check	Example
Row Count	Confirm the number of records is reasonable after cleaning.	Did removing duplicates reduce the sales records correctly?
Data Types	Ensure each column has the correct format.	Date = Date, Revenue = Decimal Number, Customer ID = Text
Missing Values	Check for blanks or null values in important fields.	Are any Customer IDs, Product IDs, or Revenue values missing?
Duplicates	Confirm duplicate records have been removed.	No repeated Transaction IDs should remain.
Standardized Values	Check that inconsistent names have been corrected.	Cairo, cairo, and CAIRO should appear as Cairo.
Merged Data	Confirm related tables were combined correctly.	Each sales transaction should show the correct Customer Name and Product Name.

Lecture
02

Quick Validation Visuals in Power BI

Card: Total Revenue

Bar Chart: Revenue by Region

Bar Chart: Revenue by Product Category

Table: Customer Name, Product Name, Revenue, Profit

Clean data is not enough — validated data is what makes reports trustworthy.

Lecture

02

Demo on Power BI and using Power Query



Exercise: Preparing Quarterly Sales Data in Power BI

- **Using the Provided Data Set:**
PowerBI_Session2_Demo_Dataset
- **Includes 3 files**
 - Sales_Transactions_Q1.xlsx
 - Product_Master.xlsx
 - Customer_Master.csv
- **What needs to be Done:**

Clean, transform, and combine sales, customer, and product data using Power BI Power Query to prepare a reliable dataset for reporting.



Lecture 02

Exercise

Exercise: Preparing Quarterly Sales Data in Power BI

Step	Task	What to Do
1	Import the data files	Open Power BI Desktop and import the three files: Sales Transactions , Customer Master , and Product Master . Choose Transform Data to open Power Query instead of loading the files directly.
2	Clean the sales data	Remove duplicate transactions, trim extra spaces, standardize region names, fix missing values, and correct data types such as Date, Quantity, Unit Price, Discount, and Revenue.
3	Clean the master data files	In the Customer Master file, remove duplicate customers, standardize city names, and replace missing segments with Unclassified . In the Product Master file, remove duplicate products, standardize product names, fix unit cost data type, and replace missing categories with Uncategorized .
4	Combine the datasets	Merge the Sales table with the Customer Master table using Customer ID , then merge it with the Product Master table using Product ID . Expand the required customer and product fields into the sales table.
5	Validate the prepared data	Load the cleaned data into Power BI and create quick validation visuals such as Total Revenue , Revenue by Region , Revenue by Product Category , and Revenue by Customer Segment to confirm the data is ready for reporting.



Lecture 02

Exercise

Lecture
02**Key Points**

Power BI connects to multiple data sources.

Power Query prepares data for analysis.

Data quality drives reporting quality.

Clean data leads to trustworthy insights.

Data preparation is the foundation of every successful Power BI dashboard.

Lecture 03:

- Data Modelling for Business Reporting



Lecture
02

Next Lecture



Thank You