Strategic Data Management (STRADAM)

Stockholm University

Department of Computer and Systems Sciences

Title: Profiling, Cleaning, Transforming and Analyzing Data Using Power Query

HT2025

By Workneh Y. Ayele (Ph.D.)

Lab Goals

After completing this module, students should understand why and how Power Query is used for strategic data management. The goals of this module are therefore:

- *Familiarize and understand the role of Power Query in data cleaning, profiling, and preparation for analysis and decision making.*
- *Develop practical skills in transforming raw, inconsistent data into a structured, reliable dataset using a smaller dataset, and students should be able to practice on massive datasets after completing this exercise.*
- *Improve data quality and consistency to enable trustworthy business insights.*
- *Prepare datasets for downstream analysis (PivotTables, dashboards, reporting)*
- *Briefly introduce AI-assisted data transformation to improve efficiency and scalability.*

Learning Outcomes

- *To understand why Power Query is needed in strategic data management (data cleaning, profiling, automation, reproducibility).*
- *Ability to practice data import, query creation, and use the Query Settings pane for data transformation*
- *Ability to do data profiling and quality assessment using Column Quality, Column Distribution, and Column Profile to identify missing values, detect duplicates, spot errors, and inconsistencies*
- *Ability to make data-quality-related decisions*
- *Ability to do data cleaning and standardization ß change data types, identify errors, and clean text by removing leading and trailing spaces, standardize letter cases (e.g., upper case, lower case, sentence case)*
- *Ability to practice data transformation and feature engineering. For example, create new columns using built-in and custom transformations, using calculated fields, and merge columns using custom separators.*
- *Ability to practice data standardization for business logic. For example, standardizing categorical values (e.g., Transfer, Online Transfer, and Wire Money à Transfer).*
- *Ability to practice aggregation and analytical preparation by duplicating queries, grouping using group transformations for analytics tasks, and preparing transformed data suitable for dashboards, strategic decision support, and reporting*
- *Ability to practice AI-assisted data transformation.*

1 About Power Query

Power Query is Microsoft's standard data preparation and transformation engine used across Excel, Power BI, Enterprise ETL (Extraction, Transformation, and Loading) in SSIS and Azure, and Low-code platforms (Power Platform). In this tutorial, we demonstrate how data transformation is done using a smaller dataset accessible through the link provided in the instruction. The instructions also apply to larger datasets.

1.1 Power Query Supporting Version

Power Query in Microsoft Excel is used mainly to:

- Import data from many sources (CSV, Excel files, Access Files, SharePoint, APIs, etc.)
- Clean and preprocess data without having to code or with low-code/no-code capabilities to filter, merge, split, reshape, change data types, remove duplicates, etc.
- Automate repeatable tasks so that data can be refreshed with a single click.
- Combine multiple data sources into a single, analysis-ready table.

Excel versions that support Power Query:

- Excel 2010 and 2013 on Windows ← Power Query must be downloaded and installed separately from Microsoft.
- Excel 2016, 2019, 2021, Excel 365/Microsoft 365 ← it is built-in in these versions (not a plugin), in these versions, it appears under the ribbon Data → Get & Transform Data

1.2 Data pre-processing and profiling

80% of the time spent on data analysis/analytics is allocated to cleaning and pre-processing data. It is therefore essential to learn how to prepare data for analytics. It is possible to use Python for pre-processing data; however, Power Query offers a faster and more user-friendly approach to clean, transform, and reshape data without requiring code. Power Query has an intuitive interface and supports a step-by-step transformation approach, enabling the efficient handling of everyday data preparation tasks, including removing duplicates, correcting missing values, splitting and merging columns, changing data types, filtering rows, and combining data from multiple sources. Additionally, Power Query automatically records every action, making the cleaning process transparent, repeatable, and easy to update when new data is added. Therefore, Power Query is a powerful tool for efficiently and reliably preparing data before performing analytics or building reports and dashboards for businesses.

2 Power Query for cleaning, standardization and transforming small dataset

In this lab, you will use Power Query in Excel to clean, standardize, and transform a small dataset.

The goal is to prepare analysis-ready data that can support strategic and managerial decision-making.

Before starting, reflect on the following:

- Why is **data cleaning** important for strategic analysis?
- What problems can inconsistent data cause in dashboards and decisions?
- How does Power Query help automate and document data preparation?

2.1 Tools Required

- Microsoft Excel (Windows recommended; Mac has limitations)
- Power Query (built-in in Excel 2016+ / Microsoft 365)

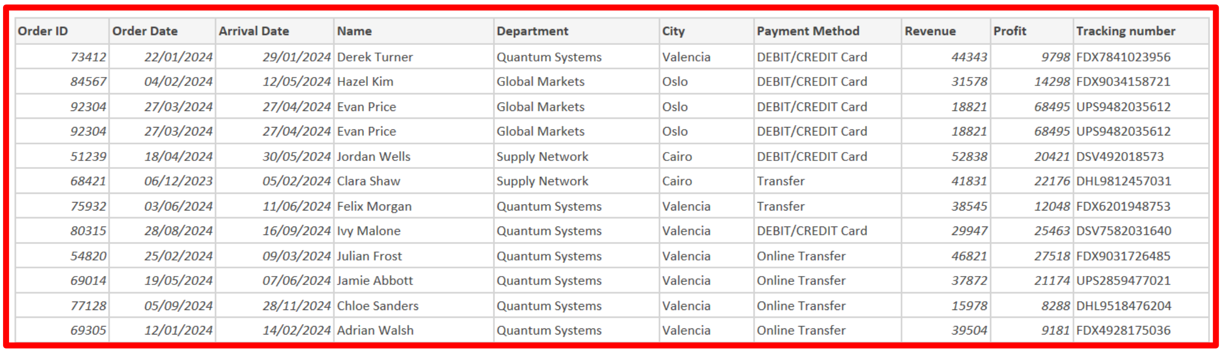
2.2 Dataset Description

You will find the dataset (an Excel file) on GitHub → https://github.com/worknehy/Course-Materials---Computer-and-Systems-Sciences-/commit/c19da1b81da774475773ba4acccd3fb3ed17d441

⇨ Download the file

The dataset contains the following columns:

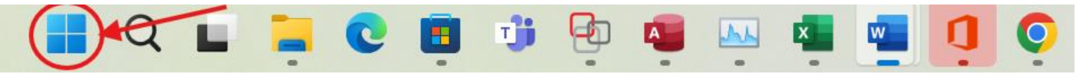
- Order ID ← Unique order identification number
- Order Date ← the date when the order was placed
- Arrival Date ← the date when the ordered items/products arrived
- Name ← contains first and last names, some of the names have leading and trailing spaces
- Department ← the department providing the items/products
- City ← shipment destination city
- Payment Method ← inconsistencies (For example, Transfer and Online Transfer are the same thing)
- Revenue ← revenue from the order
- Profit ← profit earned from the order
- Tracking number ← the first three characters contain the shipping company.



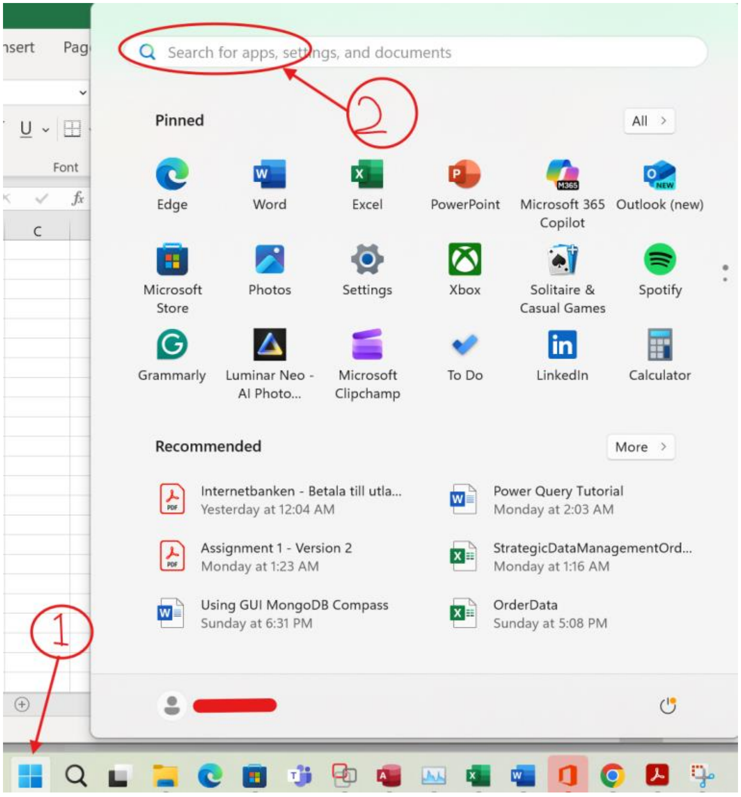
Order ID	Order Date	Arrival Date	Name	Department	City	Payment Method	Revenue	Profit	Tracking number
73412	22/01/2024	29/01/2024	Derek Turner	Quantum Systems	Valencia	DEBIT/CREDIT Card	44343	9798	FDX7841023956
84567	04/02/2024	12/05/2024	Hazel Kim	Global Markets	Oslo	DEBIT/CREDIT Card	31578	14298	FDX9034158721
92304	27/03/2024	27/04/2024	Evan Price	Global Markets	Oslo	DEBIT/CREDIT Card	18821	68495	UPS9482035612
92304	27/03/2024	27/04/2024	Evan Price	Global Markets	Oslo	DEBIT/CREDIT Card	18821	68495	UPS9482035612
51239	18/04/2024	30/05/2024	Jordan Wells	Supply Network	Cairo	DEBIT/CREDIT Card	52838	20421	DSV492018573
68421	06/12/2023	05/02/2024	Clara Shaw	Supply Network	Cairo	Transfer	41831	22176	DHL9812457031
75932	03/06/2024	11/06/2024	Felix Morgan	Quantum Systems	Valencia	Transfer	38545	12048	FDX6201948753
80315	28/08/2024	16/09/2024	Ivy Malone	Quantum Systems	Valencia	DEBIT/CREDIT Card	29947	25463	DSV7582031640
54820	25/02/2024	09/03/2024	Julian Frost	Quantum Systems	Valencia	Online Transfer	46821	27518	FDX9031726485
69014	19/05/2024	07/06/2024	Jamie Abbott	Quantum Systems	Valencia	Online Transfer	37872	21174	UPS2859477021
77128	05/09/2024	28/11/2024	Chloe Sanders	Quantum Systems	Valencia	Online Transfer	15978	8288	DHL9518476204
69305	12/01/2024	14/02/2024	Adrian Walsh	Quantum Systems	Valencia	Online Transfer	39504	9181	FDX4928175036

2.3 Load Data into Power Query

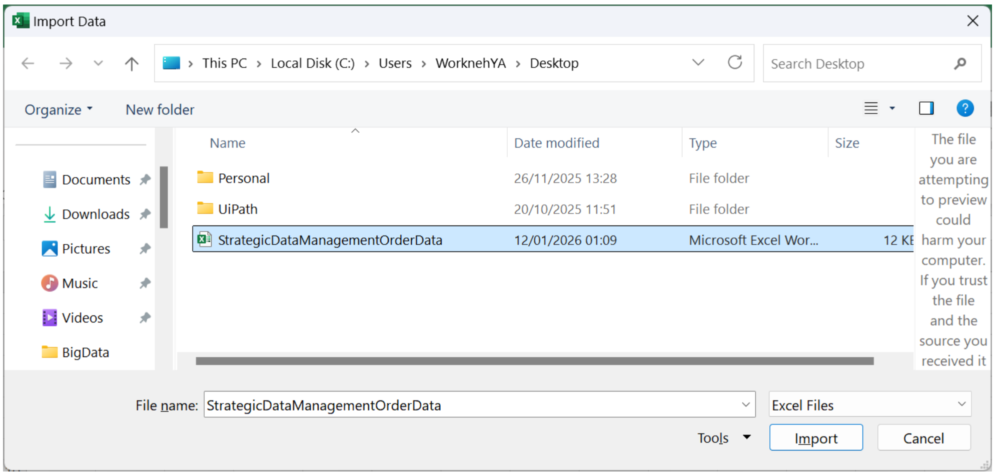
- Open Excel and load the dataset.
  - Click on **Start** Menu



  - Select **Microsoft Excel** from your **Start** menu (or you can type **Excel** as illustrated below under (2) text box to search for Excel and select it when it shows up.



  -
- Select **Data** to display the **Data** ribbon.
- Select **Get Data**, and the **Get Data** commands will pop up as illustrated below.
  - You can view a list of options to import data from a plethora of sources.
- Select **From File** and you can select **From Workbook** to import from an Excel file, **From Text/CSV** to import from a flat file, **From XML** to import an XML file, etc.
  - Select **From Workbook** ←
- Locate the file, you downloaded

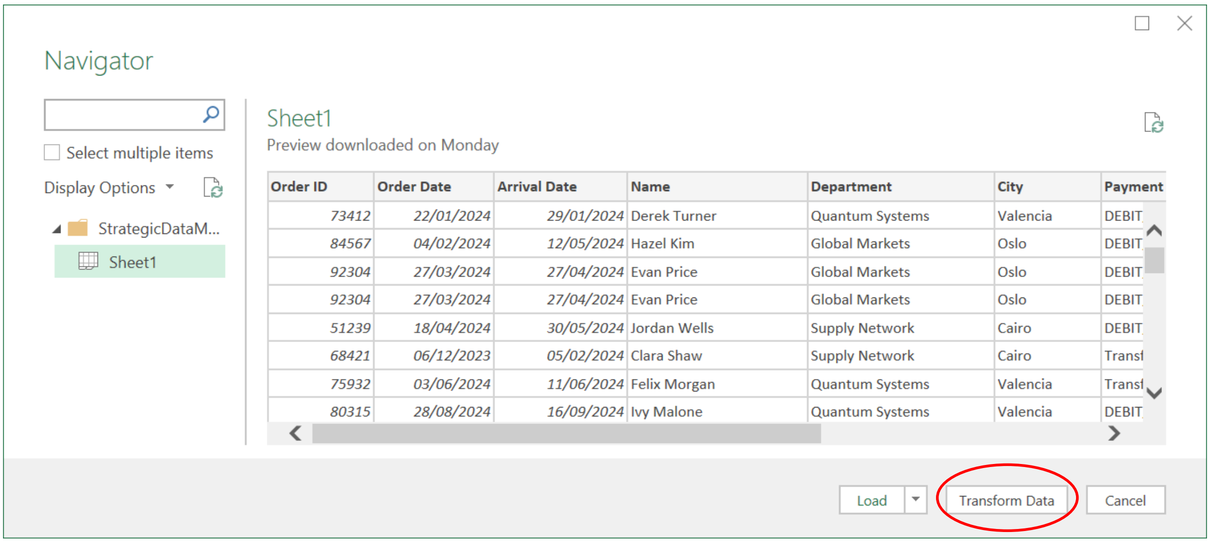

Import Data
This PC › Local Disk (C:) › Users › WorknehYA › Desktop
Search Desktop
Organize
New folder
Documents
Downloads
Pictures
Music
Videos
BigData
The file you are attempting to preview could harm your computer. If you trust the file and the source you received it
Tools
Import
Cancel

Name	Date modified	Type	Size
Personal	26/11/2025 13:28	File folder	
UiPath	20/10/2025 11:51	File folder	
StrategicDataManagementOrderData	12/01/2026 01:09	Microsoft Excel Wor...	12 KE

File name: StrategicDataManagementOrderData

Excel Files

- Click on **Import**
- Select the worksheet that contains the file as illustrated below
- Click on **Transform Data**

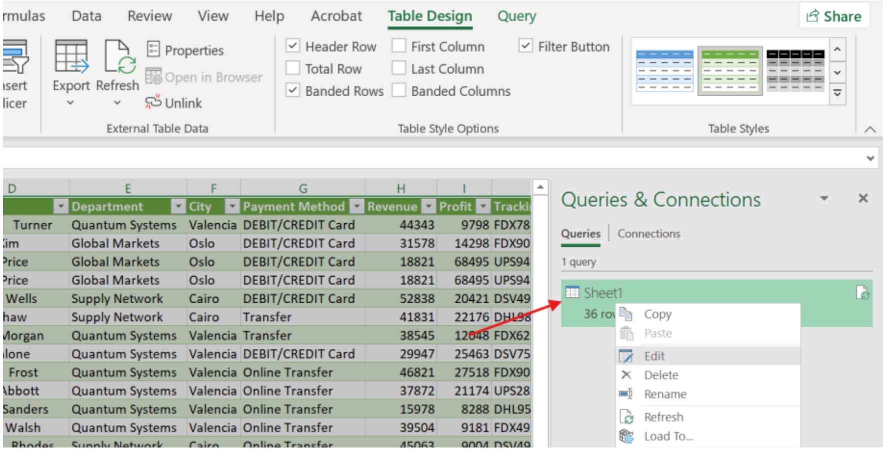


Sheet1

Preview downloaded on Monday

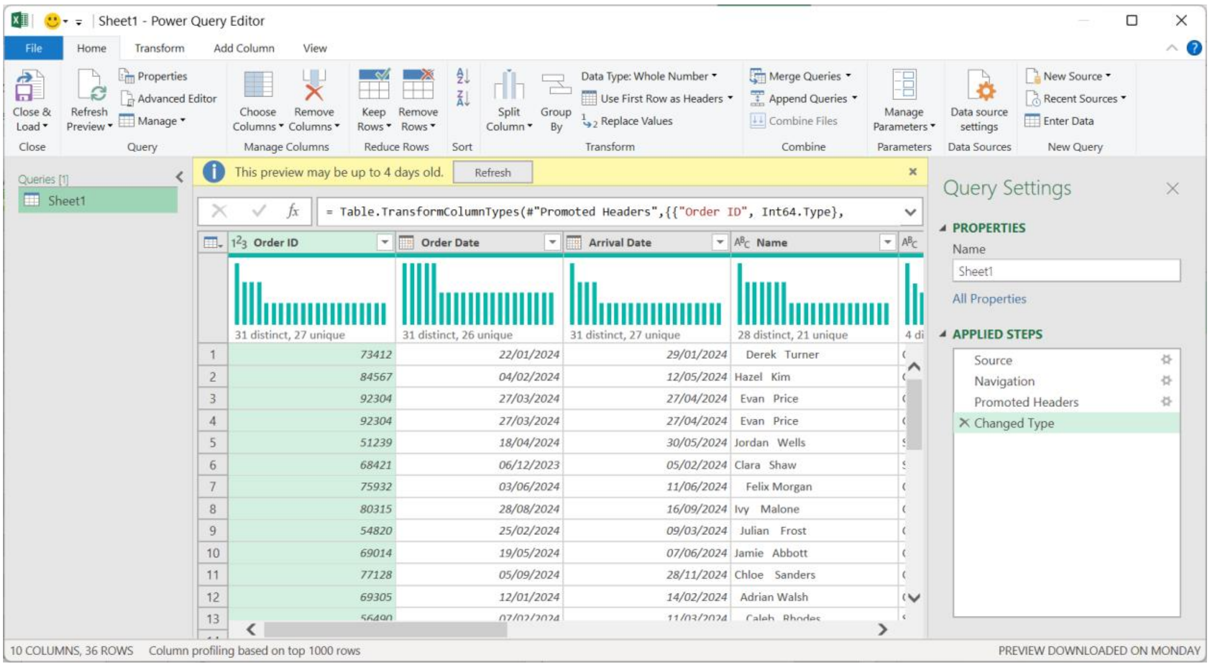
Order ID	Order Date	Arrival Date	Name	Department	City	Payment
73412	22/01/2024	29/01/2024	Derek Turner	Quantum Systems	Valencia	DEBIT
84567	04/02/2024	12/05/2024	Hazel Kim	Global Markets	Oslo	DEBIT
92304	27/03/2024	27/04/2024	Evan Price	Global Markets	Oslo	DEBIT
92304	27/03/2024	27/04/2024	Evan Price	Global Markets	Oslo	DEBIT
51239	18/04/2024	30/05/2024	Jordan Wells	Supply Network	Cairo	DEBIT
68421	06/12/2023	05/02/2024	Clara Shaw	Supply Network	Cairo	Transf
75932	03/06/2024	11/06/2024	Felix Morgan	Quantum Systems	Valencia	Transf
80315	28/08/2024	16/09/2024	Ivy Malone	Quantum Systems	Valencia	DEBIT

- ○ **Note:** If you mistakenly clicked on **Load,** the **Power Query Window** will not show up, but if you did, right-click on your Worksheet (**Sheet1**) from the **Queries & Connections** panel and select **Edit**, as illustrated below.



D	E	F	G	H	I	
	Department	City	Payment Method	Revenue	Profit	Tracki
Turner	Quantum Systems	Valencia	DEBIT/CREDIT Card	44343	9798	FDX78
Kim	Global Markets	Oslo	DEBIT/CREDIT Card	31578	14298	FDX90
Price	Global Markets	Oslo	DEBIT/CREDIT Card	18821	68495	UPS94
Price	Global Markets	Oslo	DEBIT/CREDIT Card	18821	68495	UPS94
Wells	Supply Network	Cairo	DEBIT/CREDIT Card	52838	20421	DSV49
haw	Supply Network	Cairo	Transfer	41831	22176	DHL98
Morgan	Quantum Systems	Valencia	Transfer	38545	12048	FDX62
lone	Quantum Systems	Valencia	DEBIT/CREDIT Card	29947	25463	DSV75
Frost	Quantum Systems	Valencia	Online Transfer	46821	27518	FDX90
Abbott	Quantum Systems	Valencia	Online Transfer	37872	21174	UPS28
Sanders	Quantum Systems	Valencia	Online Transfer	15978	8288	DHL95
Walsh	Quantum Systems	Valencia	Online Transfer	39504	9181	FDX49
Rhodes	Supply Network	Cairo	Online Transfer	45063	9004	DSV49

- **Power Query Editor** will open; see the illustration below.



This preview may be up to 4 days old.

`= Table.TransformColumnTypes(#"Promoted Headers",{{"Order ID", Int64.Type},`

	Order ID	Order Date	Arrival Date	Name
	31 distinct, 27 unique	31 distinct, 26 unique	31 distinct, 27 unique	28 distinct, 21 unique
1	73412	22/01/2024	29/01/2024	Derek Turner
2	84567	04/02/2024	12/05/2024	Hazel Kim
3	92304	27/03/2024	27/04/2024	Evan Price
4	92304	27/03/2024	27/04/2024	Evan Price
5	51239	18/04/2024	30/05/2024	Jordan Wells
6	68421	06/12/2023	05/02/2024	Clara Shaw
7	75932	03/06/2024	11/06/2024	Felix Morgan
8	80315	28/08/2024	16/09/2024	Ivy Malone
9	54820	25/02/2024	09/03/2024	Julian Frost
10	69014	19/05/2024	07/06/2024	Jamie Abbott
11	77128	05/09/2024	28/11/2024	Chloe Sanders
12	69305	12/01/2024	14/02/2024	Adrian Walsh
13	56490	07/02/2024	11/03/2024	Caleb Rhodes

Name: Sheet1

Applied steps: Source; Navigation; Promoted Headers; Changed Type

10 COLUMNS, 36 ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED ON MONDAY

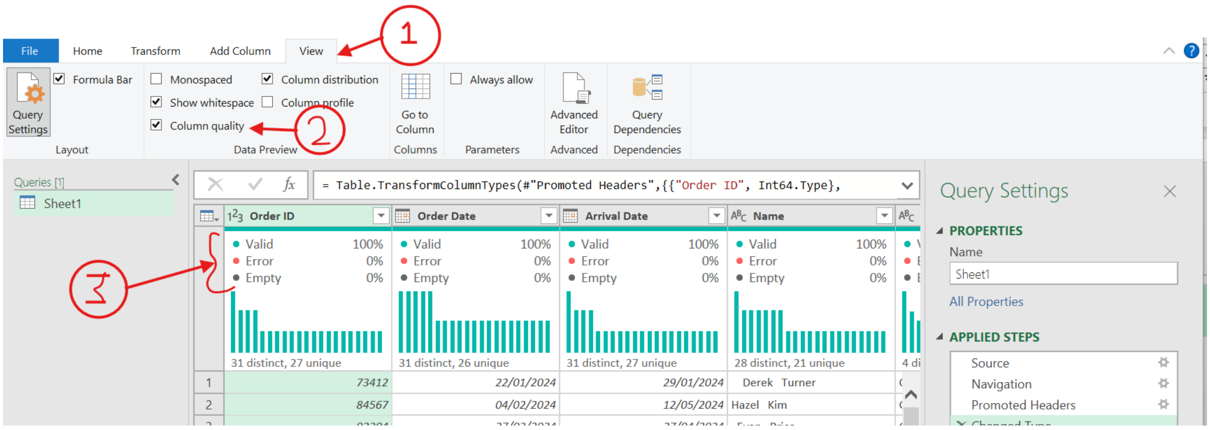
2.4 Data Profiling (Understand Your Data)

Purpose: Understand data completeness, and reliability before cleaning.

2.4.1 Column Quality

Purpose: check column quality, check valid values, errors and empty values

- In Power Query, go to:
  - Click on the **View** tab → check **Column Quality** from the ribbon



☑ Formula Bar
☐ Monospaced
☑ Show whitespace
☑ Column quality
☑ Column distribution
☐ Column profile

`= Table.TransformColumnTypes(#"Promoted Headers",{{"Order ID", Int64.Type},`

	Order ID	Order Date	Arrival Date	Name
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	31 distinct, 27 unique	31 distinct, 26 unique	31 distinct, 27 unique	28 distinct, 21 unique
1	73412	22/01/2024	29/01/2024	Derek Turner
2	84567	04/02/2024	12/05/2024	Hazel Kim

Name: Sheet1

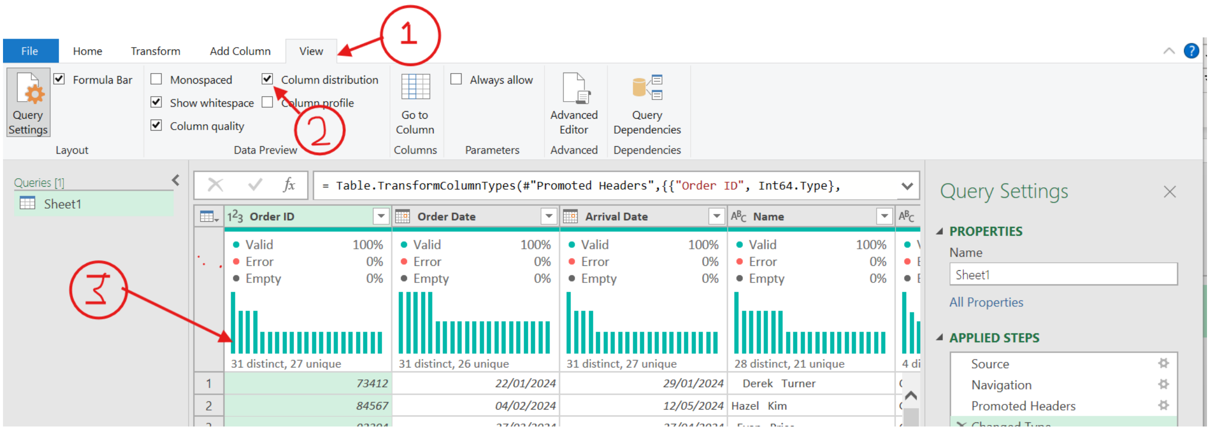
Applied steps: Source; Navigation; Promoted Headers; Changed Type

- You can check:
  - Valid values
  - Errors
  - Empty values

2.4.2 Column Distribution

Purpose: You can check the data distribution in your columns

- In Power Query, go to:
  - Click on the **View** tab → check **Column Distribution** from the ribbon



☑ Formula Bar
□ Monospaced
☑ Show whitespace
☑ Column quality
☑ Column distribution
□ Column profile
□ Always allow

```
= Table.TransformColumnTypes(#"Promoted Headers",{{"Order ID", Int64.Type},
```

	Order ID	Order Date	Arrival Date	Name
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	31 distinct, 27 unique	31 distinct, 26 unique	31 distinct, 27 unique	28 distinct, 21 unique
1	73412	22/01/2024	29/01/2024	Derek Turner
2	84567	04/02/2024	12/05/2024	Hazel Kim

Name: Sheet1

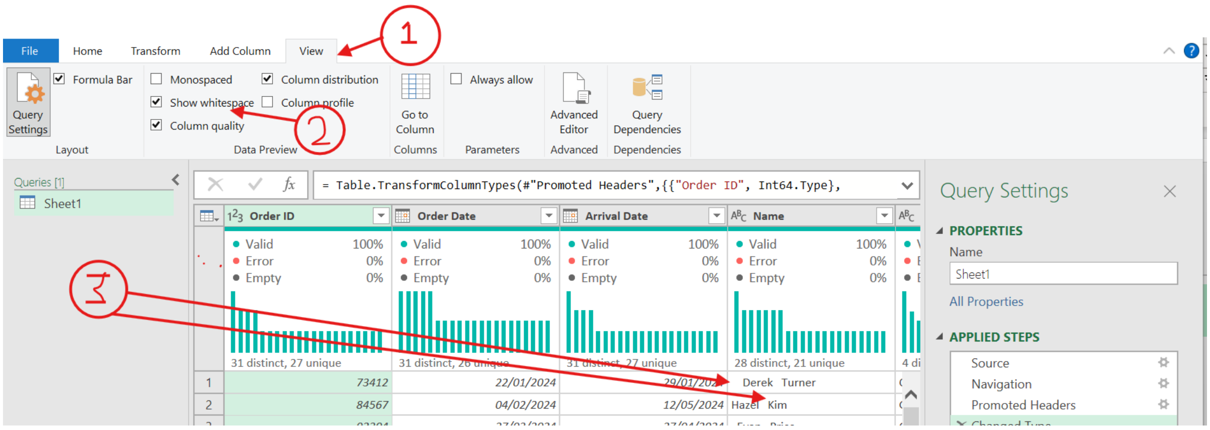
Applied steps: Source, Navigation, Promoted Headers, Changed Type

- You can check:
  - Distinct and unique values under each column

2.4.3 Show whitespace

Purpose: You can check unnecessary and extra whitespaces before word strings and between them, as illustrated below

- In Power Query, go to:
  - Click on the **View** tab → check **Whitespaces** from the ribbon



☑ Formula Bar
□ Monospaced
☑ Show whitespace
☑ Column quality
☑ Column distribution
□ Column profile
□ Always allow

```
= Table.TransformColumnTypes(#"Promoted Headers",{{"Order ID", Int64.Type},
```

	Order ID	Order Date	Arrival Date	Name
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	31 distinct, 27 unique	31 distinct, 26 unique	31 distinct, 27 unique	28 distinct, 21 unique
1	73412	22/01/2024	29/01/2024	Derek Turner
2	84567	04/02/2024	12/05/2024	Hazel Kim

Name: Sheet1

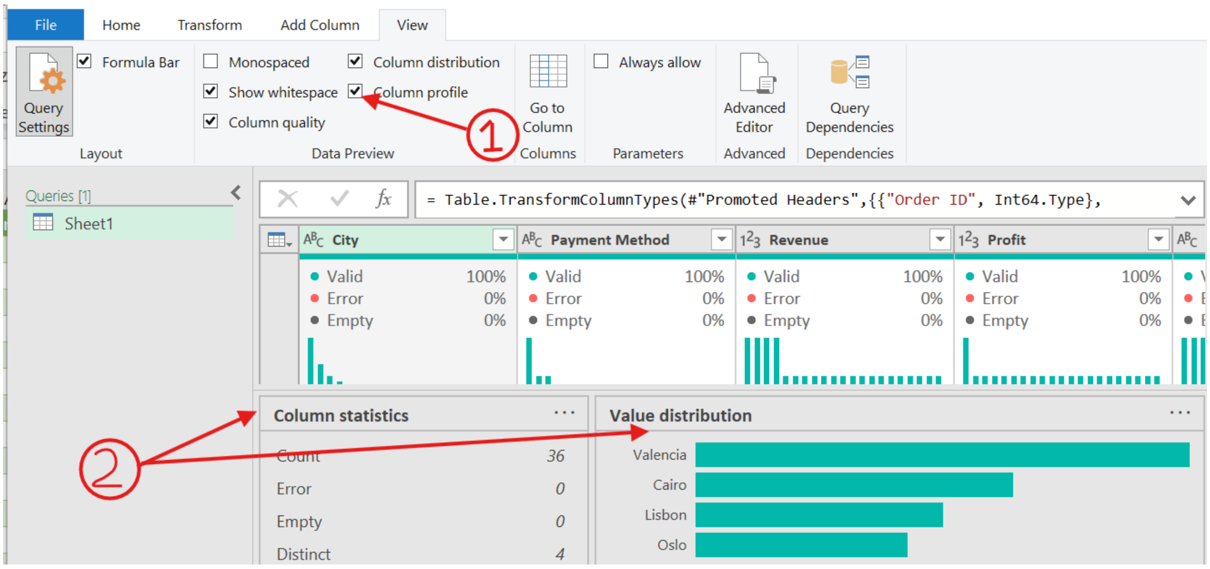
Applied steps: Source, Navigation, Promoted Headers, Changed Type

- You can check:
  - Extra whitespaces in your columns

2.4.4 Column Profile

Purpose: Understand data composition, such as column statistics, minimum, maximum, distinct, error, empty, and count. Additionally, you can see the column value distribution.

- In Power Query, go to:
  - Click on the **View** tab → check **Column profile** from the ribbon



☑ Formula Bar
□ Monospaced
☑ Show whitespace
☑ Column quality
☑ Column distribution
☑ Column profile
□ Always allow

```
= Table.TransformColumnTypes(#"Promoted Headers",{{"Order ID", Int64.Type},
```

	City	Payment Method	Revenue	Profit
Valid	100%	100%	100%	100%
Error	0%	0%	0%	0%
Empty	0%	0%	0%	0%

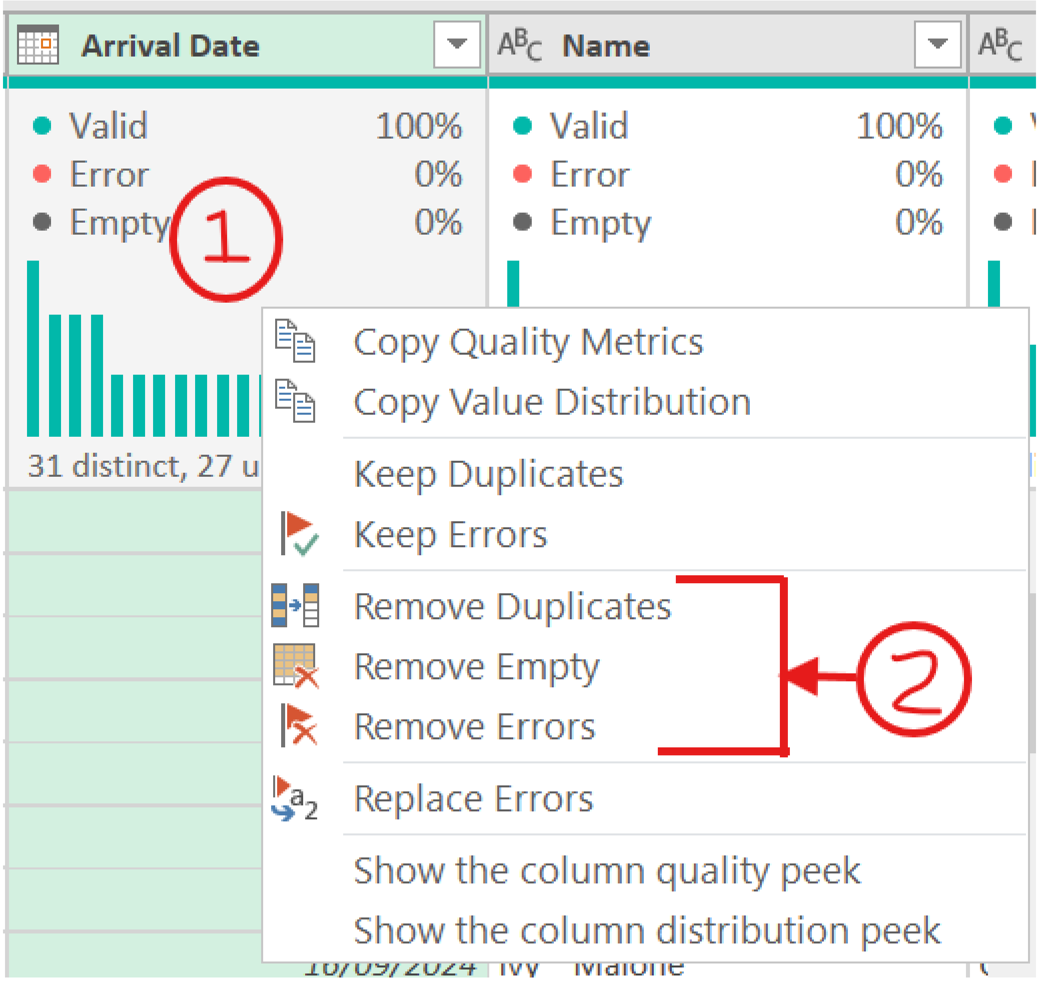
Column statistics	
Count	36
Error	0
Empty	0
Distinct	4

- You can check:
  - Column statistics
    - Count, Error, Empty, Distinct, Unique, Empty string, Min, and Max.
  - Value distribution
    - The distribution of unique values, as illustrated in the snapshot

2.4.5 Remove Empty, Errors, or Duplicates (Use Column Distribution)

Purpose: Ensure uniqueness and consistency

- In Power Query, go to:
  - **View → Column Distribution**
- Identify duplicates and unusual values:
  - Right-click on a **Column Distribution**:
  - You can select > **Remove Empty**, **Remove Duplicates**, or **Remove Errors** where appropriate

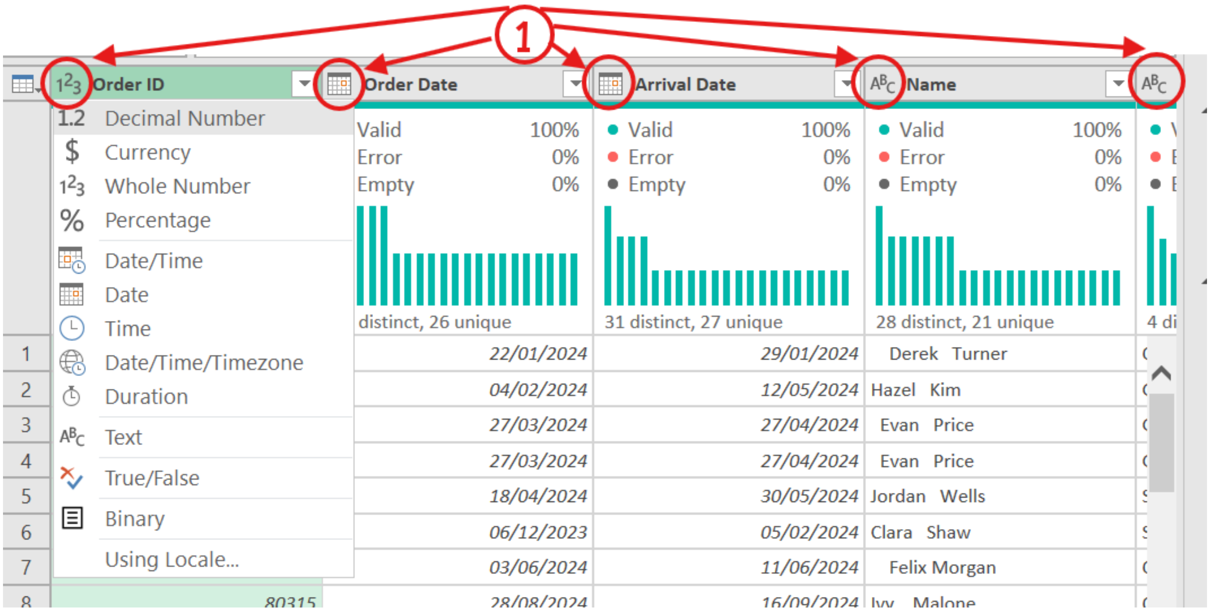


2.5 Data Type Validation

2.5.1 Change Data Type

Purpose: Prevent calculation and aggregation errors later.

- Select a column and change its data type.
  - Click on the top left corner of the columns as illustrated below.
  - Select datatype (e.g., Decimal Number, Currency, Whole Number, Percentage, Date/Time, Date, Time, Date/Time/TimeZone, etc.).

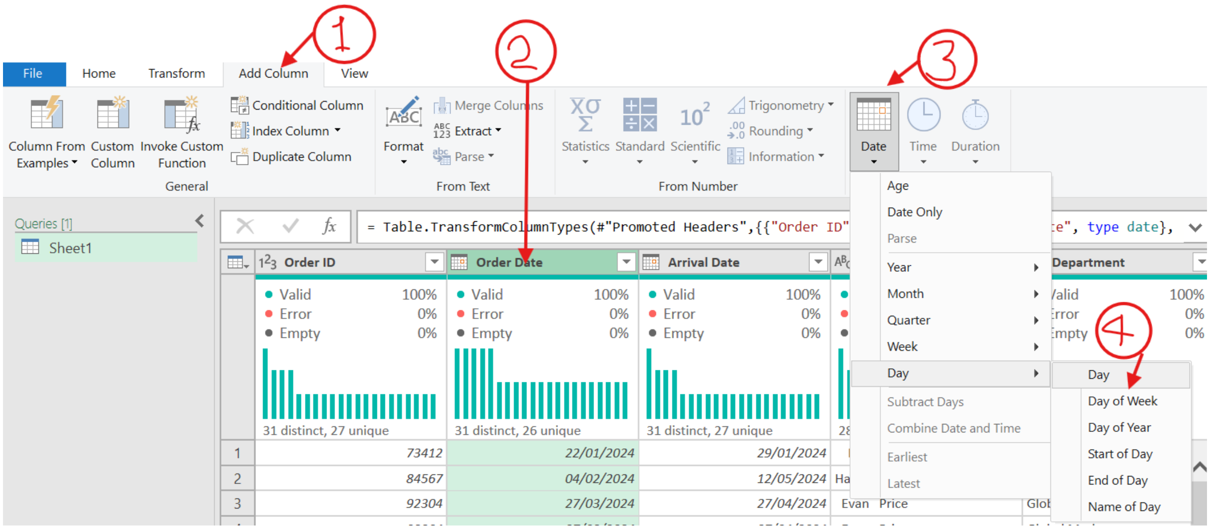


	Order ID	Order Date	Arrival Date	Name
1		22/01/2024	29/01/2024	Derek Turner
2		04/02/2024	12/05/2024	Hazel Kim
3		27/03/2024	27/04/2024	Evan Price
4		27/03/2024	27/04/2024	Evan Price
5		18/04/2024	30/05/2024	Jordan Wells
6		06/12/2023	05/02/2024	Clara Shaw
7		03/06/2024	11/06/2024	Felix Morgan
8	80315	28/08/2024	16/09/2024	Ivy Malone

2.6 Creating New Columns

2.6.1 Delivery Days Columns

- Click on the **Add Column (1)** tab,
- Select **Order Date (2)**.
- Click on **Date (3)** Commands
- Select **Day (4)**
- Rename the new column by right-clicking on the new column name and selecting **Rename...**:
  - Type new name: **Delivery Days**



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"Order ID" e", type date},
```

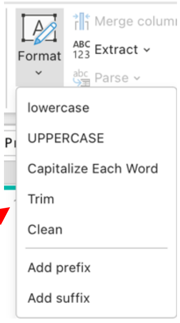
	Order ID	Order Date	Arrival Date	
1	73412	22/01/2024	29/01/2024	
2	84567	04/02/2024	12/05/2024	Ha
3	92304	27/03/2024	27/04/2024	Evan Price

2.7 Text Cleaning and Formatting

Purpose: Standardize text for grouping and matching.

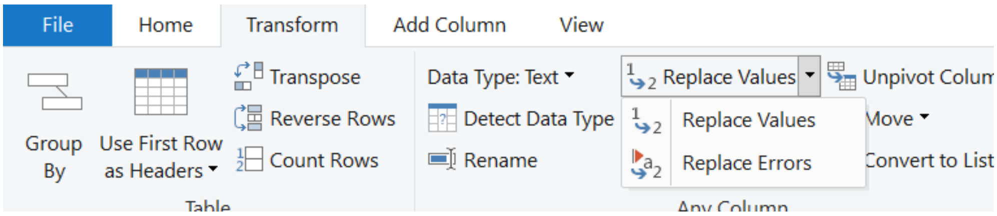
2.7.1 Formatting Name Column

- Select the **Name** column.
- Go to:
  - Select the **Transform** tab,
  - Select **Format** command,
  - Click on **Trim** ← remove Trailing and Leading spaces
- You can also:
  - Change case (e.g., UPPERCASE, lowercase, Capitalize Each Work)
  - Clean
  - Add prefix (e.g., Dr., Mr., Miss, etc.)
  - Add suffix

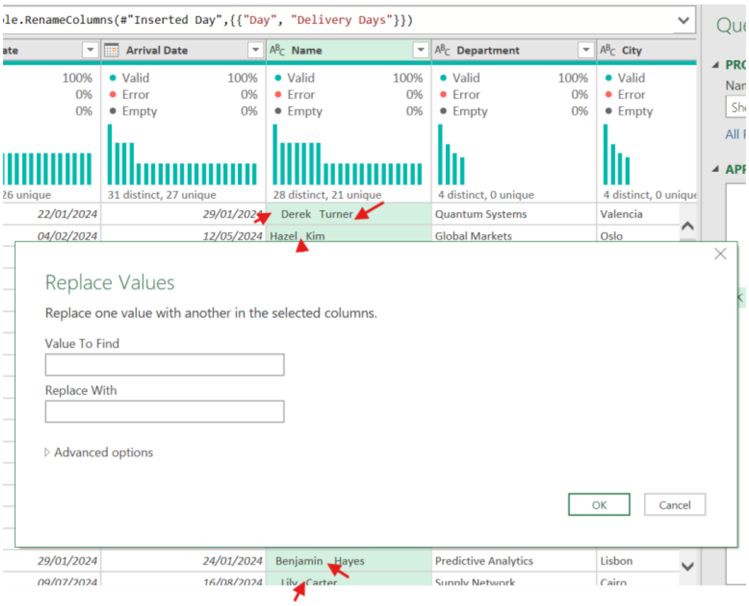


2.7.2 Removing Extra Spaces (spaces between words)

- Select the **Name** column
- Go to:
  - Select the **Transform** tab,
  - **Replace Values**



  - Enter two whitespaces under **Value To Find**
  - Enter one whitespace under **Replace With**
    - Click on the **OK** button



```
le.RenameColumns(#"Inserted Day",{{"Day", "Delivery Days"}})
```

ate	Arrival Date	Name	Department	City
22/01/2024	29/01/2024	Derek Turner	Quantum Systems	Valencia
04/02/2024	12/05/2024	Hazel Kim	Global Markets	Oslo
29/01/2024	24/01/2024	Benjamin Hayes	Predictive Analytics	Lisbon
09/07/2024	16/08/2024	Lily Carter	Supply Network	Cairo

Replace Values

Replace one value with another in the selected columns.

Value To Find

Replace With

- Repeat the previous steps:
    - Enter three whitespaces under **Value To Find**
    - Enter one whitespace under **Replace With**
    - Click on the **OK** button
- The previous actions will replace:
    - Double spaces with single spaces
    - Triple spaces with single spaces

2.8 Splitting Columns

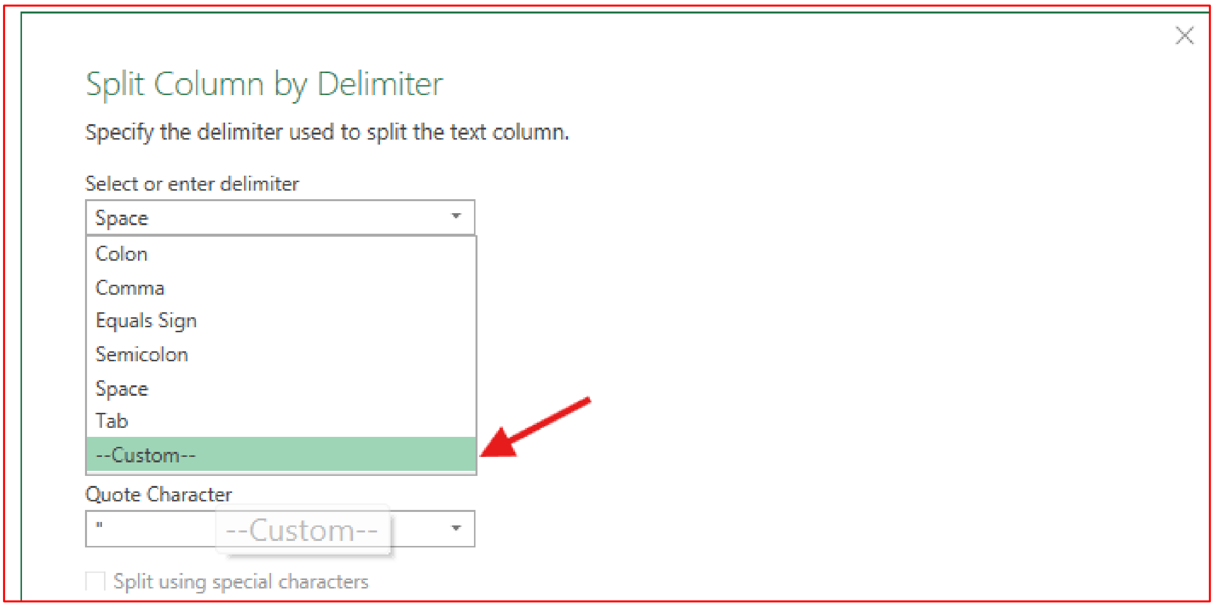
2.8.1 Split Name Column

Purpose: Separate first and last names for analysis.

- Select the **Name** column.
- Go to:
    - Select the **Transform** tab**,**
    - Click on the **Split Column** command,
    - Select **By Delimiter,**
    - Select **Space** and check the ,
    - Click OK
- Rename Name.1 to FirstName and Name.2 to LastName.

Note: Custom delimiter:

- Or if you have to a use custom delimiter (which includes ",", ";", ":" and space)



Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter: Space

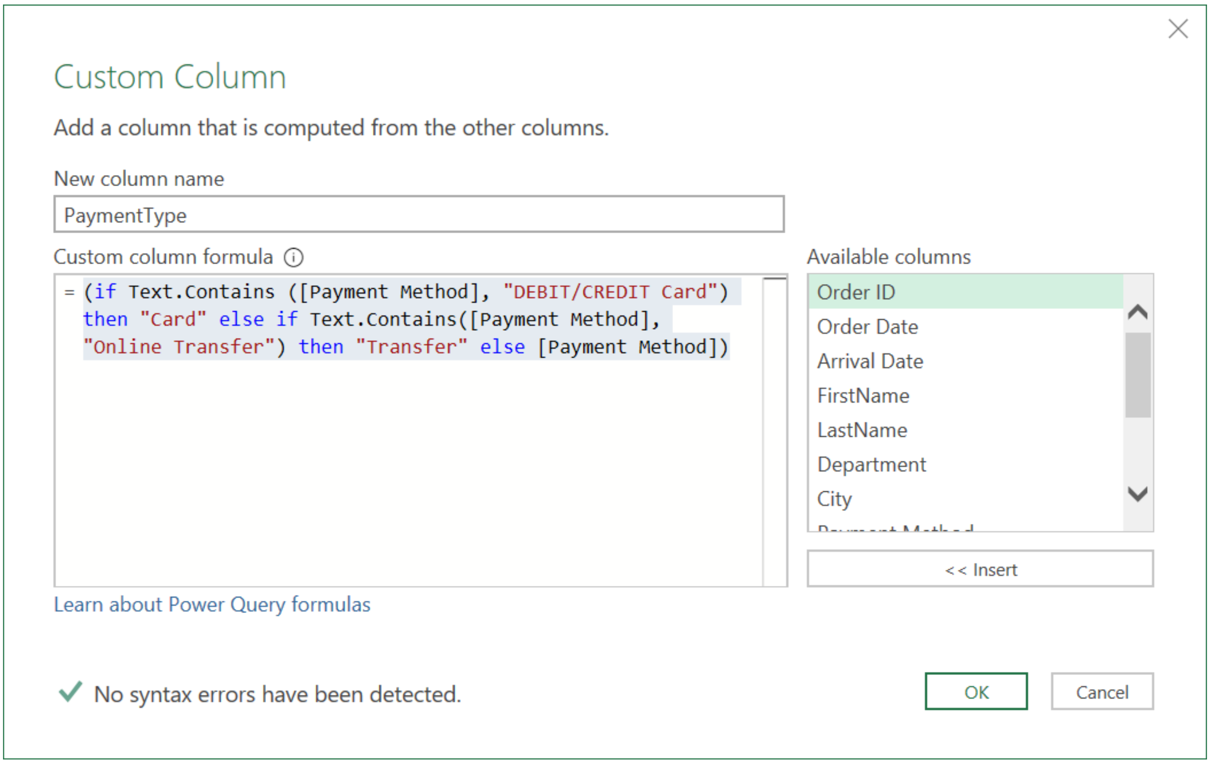
- Choose:
    - Custom delimiter
    - Enter Space (which is a whitespace and shows nothing)
- Click **OK**

2.9 Standardizing Categorical Data (By Adding a new column)

Purpose: Enforce consistent business definitions.

2.9.1 Payment Method Standardization

- Select the **Payment Method** column,
- Select the **Add Column** tab,
- Click on **Custom Column**
- Create a new column:
  - Enter **PaymentType** under the **New column name**
- Change:
  - DEBIT/CREDIT Card to Card
  - Online Transfer to Transfer
  - Add the following script as illustrated below:
  - The added code:
    - “(if Text.Contains ([Payment Method], "DEBIT/CREDIT Card") then "Card" else if Text.Contains([Payment Method], "Online Transfer") then "Transfer" else [Payment Method])”



Custom Column

Add a column that is computed from the other columns.

New column name

PaymentType

Custom column formula

```
= (if Text.Contains ([Payment Method], "DEBIT/CREDIT Card")
  then "Card" else if Text.Contains([Payment Method],
  "Online Transfer") then "Transfer" else [Payment Method])
```

✓ No syntax errors have been detected.

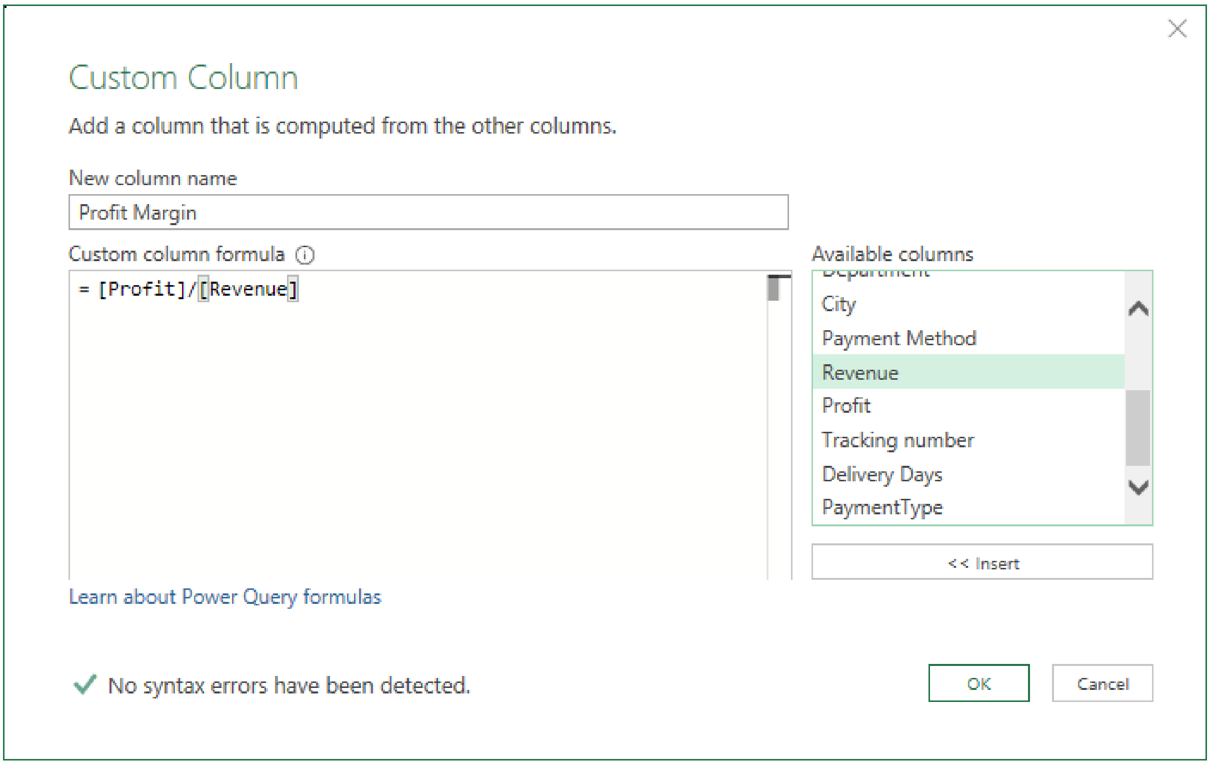
2.10 Calculated Fields

2.10.1 Profit Margin Calculation from Profit and Revenue

Purpose: Create metrics needed for strategic analysis.

- Select **Add Column** tab,

- Click on **Custom Column**
- Create a new column:
    - Enter **Profit Margin** under the **New column name**
        - Which will be calculated: **Profit Margin = [Profit]/[Revenue]**, as illustrated below.



Custom Column

Add a column that is computed from the other columns.

New column name: Profit Margin

Custom column formula:

```
= [Profit]/[Revenue]
```

No syntax errors have been detected.

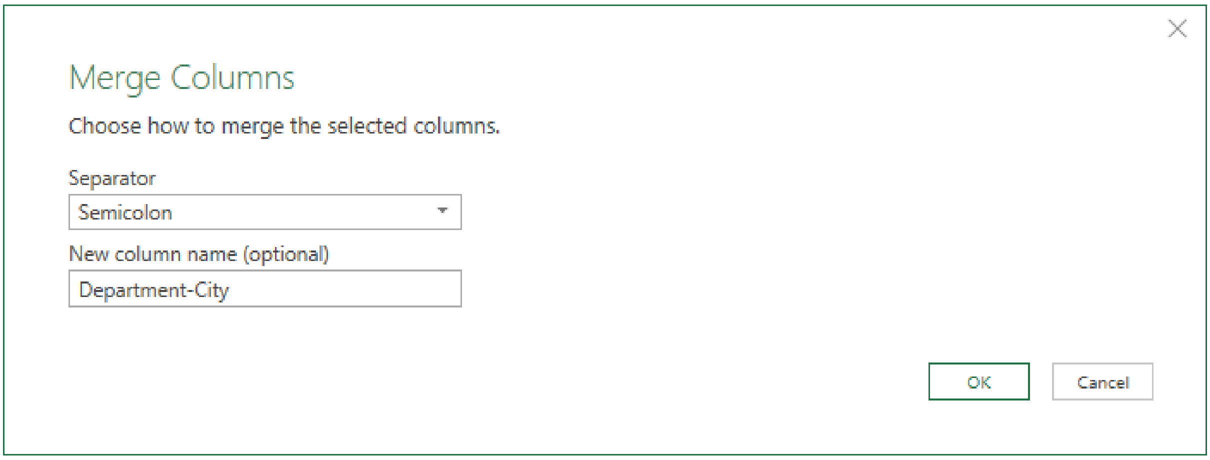
- Click **OK**

2.11 Merging Columns

Purpose: Create composite dimensions for analysis.

2.11.1 Merge Department and City

- Select **Department** and **City** columns (use Shift + clicking).
- Select **Add Column** tab,
- Click on the **Merge Columns** command
- Separator: semicolon (;)
- Under the **New column** name, enter “Department-City”



Merge Columns

Choose how to merge the selected columns.

Separator: Semicolon

New column name (optional): Department-City

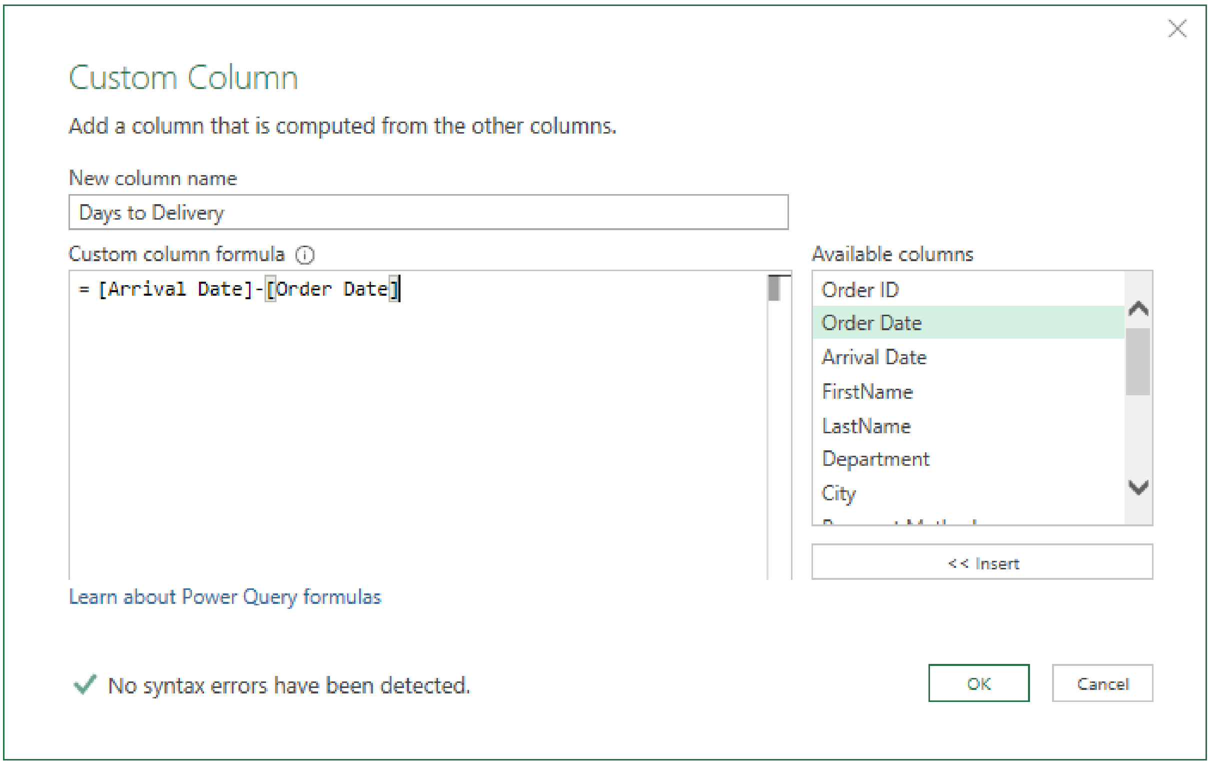
- Click **OK**

2.12 Delivery Time Analysis

2.12.1 New Delivery Time Column

Create a column that represents days to delivery, name it **Days to Delivery**.

- Select **Add Column** tab,
- Click on the **Custom Column** command
- Under the **New column** name, enter “**Days to Delivery**”
- Enter [Arrival Date]-[Order Date] under the formula.



Custom Column

Add a column that is computed from the other columns.

New column name: Days to Delivery

Custom column formula:

```
= [Arrival Date]-[Order Date]
```

Available columns: Order ID, Order Date, Arrival Date, FirstName, LastName, Department, City

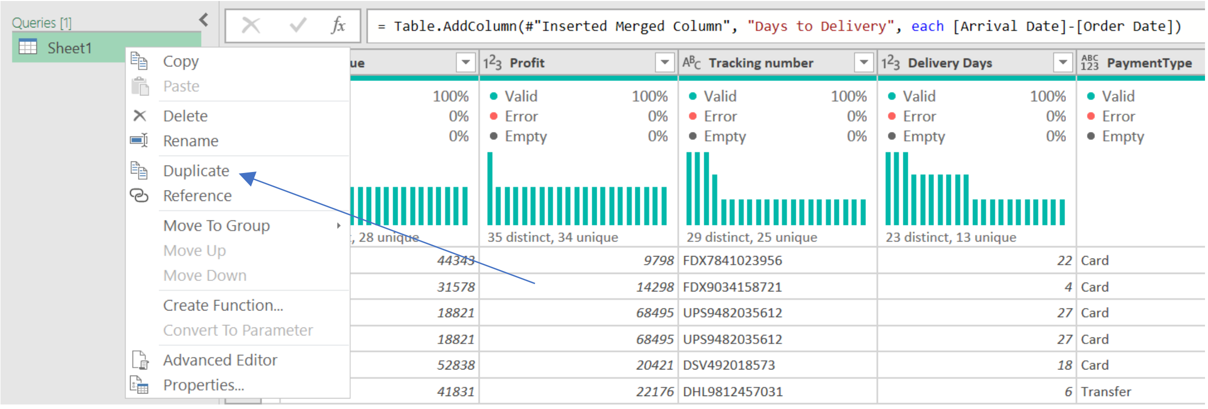
Learn about Power Query formulas

✓ No syntax errors have been detected.

2.12.2 Analysis by Department and Region

Duplicate the current query

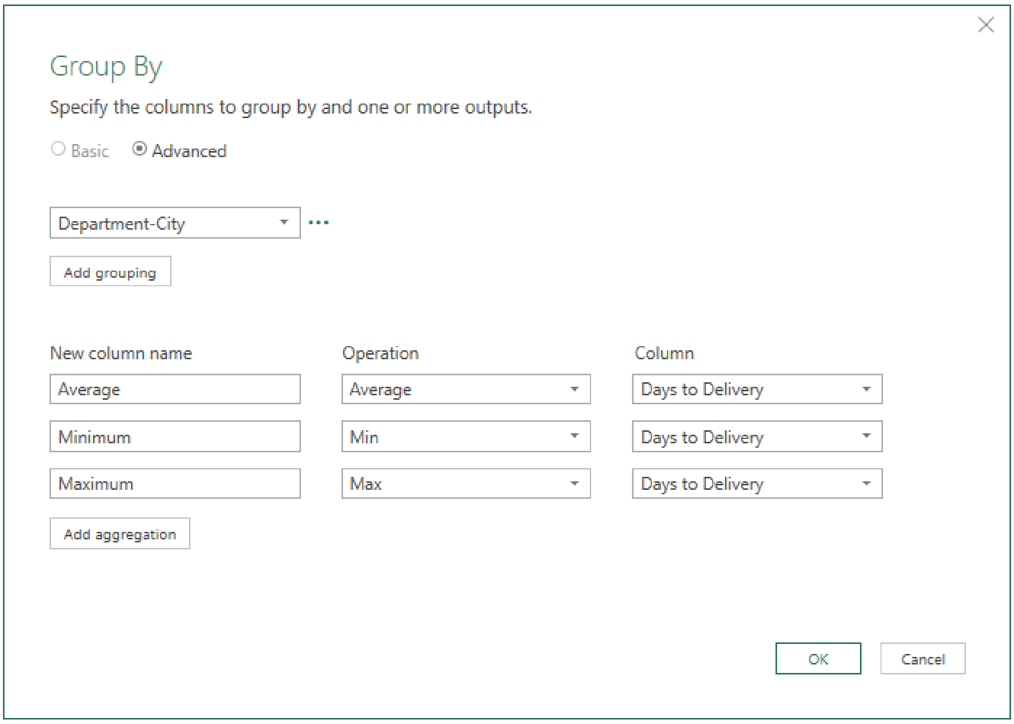
- In **Queries**, right-click your cleaned query:
  - Click **Duplicate**



```
= Table.AddColumn(#"Inserted Merged Column", "Days to Delivery", each [Arrival Date]-[Order Date])
```

ue	1²3 Profit	A^B_C Tracking number	1²3 Delivery Days	ABC 123 PaymentType
100% 0% 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid ● Error ● Empty
, 28 unique	35 distinct, 34 unique	29 distinct, 25 unique	23 distinct, 13 unique	
44343	9798	FDX7841023956	22	Card
31578	14298	FDX9034158721	4	Card
18821	68495	UPS9482035612	27	Card
18821	68495	UPS9482035612	27	Card
52838	20421	DSV492018573	18	Card
41831	22176	DHL9812457031	6	Transfer

- In the duplicated query (Sheet1 (2) → rename it to Delivery Days by Department):
  - Select the **Transform** tab,
  - Select **the Group By** command,
  - Click on the **Advanced** button,
  - Under Group by Select the **Department-City** column:
    - Under the **New column name:** enter **Average,** under **Operations:** select **Average,** under **Column: select Days to Delivery**
  - Click **Add aggregate**
    - Under the **New column name:** enter **Minimum,** under **Operations:** select **Min,** under **Column:** select **Days to Delivery**
    - Under **New column name:** enter **Maximum,** under **Operations:** select **Max,** under **Column: select Days to Delivery**
  - Your dialog box should be similar to the following snapshot:



Group By

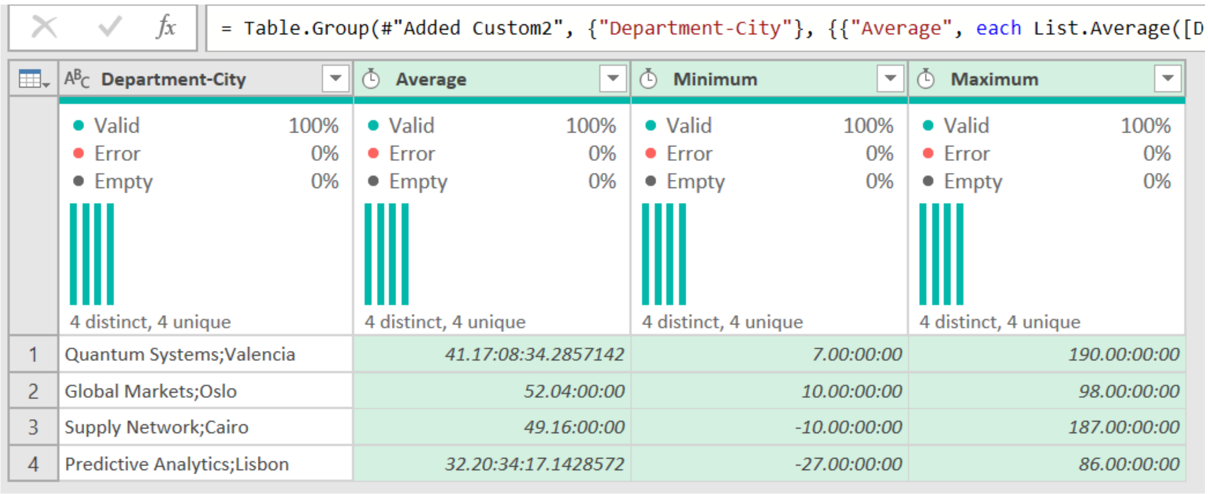
Specify the columns to group by and one or more outputs.

○ Basic ● Advanced

Department-City ▾

New column name	Operation	Column
Average	Average	Days to Delivery
Minimum	Min	Days to Delivery
Maximum	Max	Days to Delivery

- Aggregate delivery days by:
    - Department
    - Region
    - The result (which can be used for decision making (dashboard)):



```
= Table.Group(#"Added Custom2", {"Department-City"}, {{"Average", each List.Average([D
```

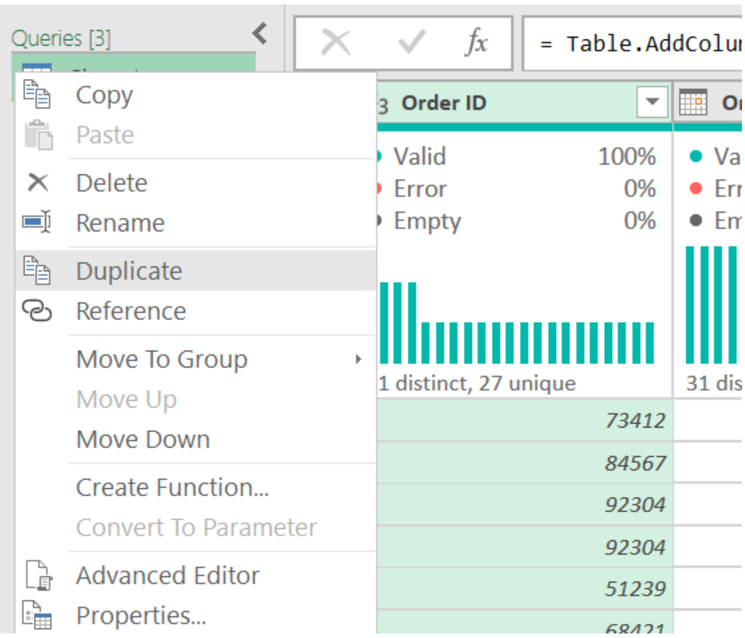
	Department-City	Average	Minimum	Maximum
	● Valid 100% ● Error 0% ● Empty 0% 4 distinct, 4 unique	● Valid 100% ● Error 0% ● Empty 0% 4 distinct, 4 unique	● Valid 100% ● Error 0% ● Empty 0% 4 distinct, 4 unique	● Valid 100% ● Error 0% ● Empty 0% 4 distinct, 4 unique
1	Quantum Systems;Valencia	41.17:08:34.2857142	7.00:00:00	190.00:00:00
2	Global Markets;Oslo	52.04:00:00	10.00:00:00	98.00:00:00
3	Supply Network;Cairo	49.16:00:00	-10.00:00:00	187.00:00:00
4	Predictive Analytics;Lisbon	32.20:34:17.1428572	-27.00:00:00	86.00:00:00

2.13 Analysis by Payment Method

Purpose: Support strategic comparisons between payment methods.

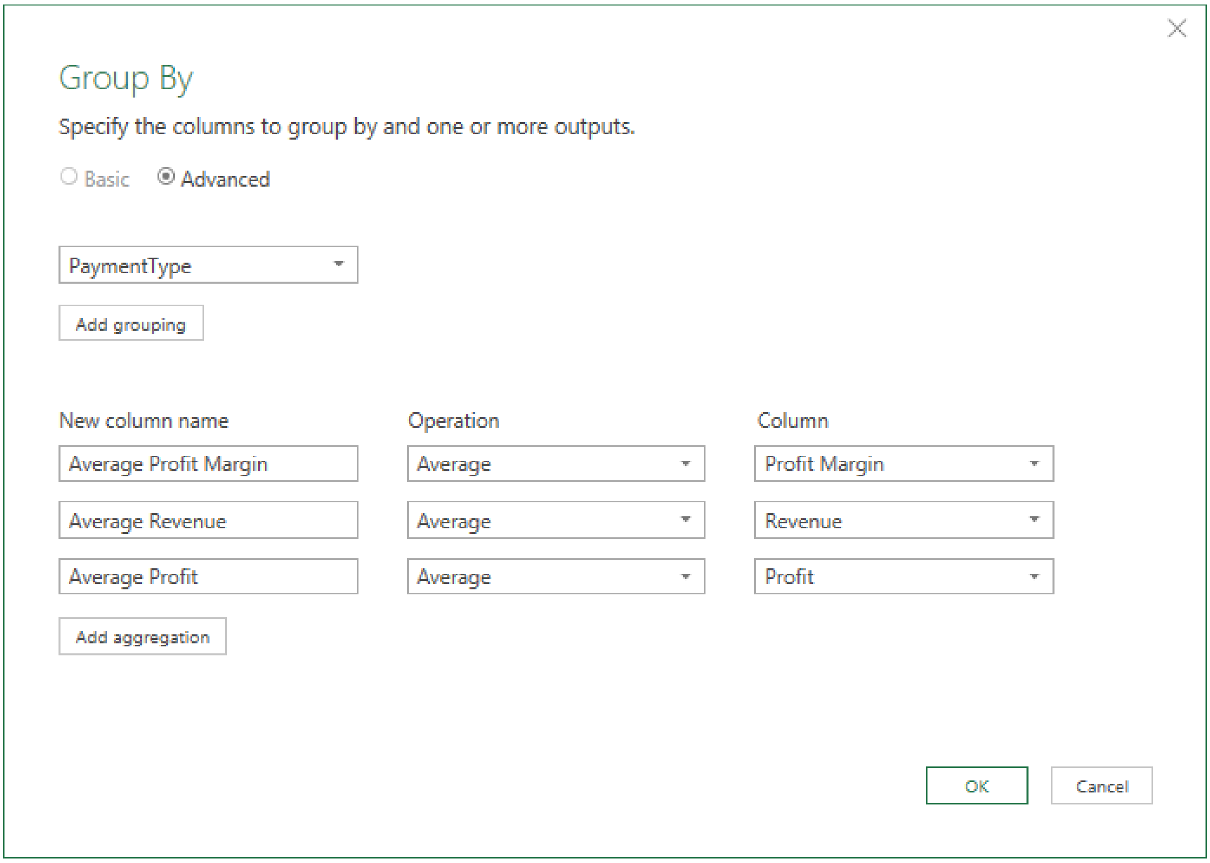
2.13.1 Payment Method Analysis

- Duplicate the first query (the query that was cleaned for analysis), rename it Analysis of Payment.



○

- Click on the **Transform** ribbon,
- Click on the **Group By** command,
    - Under the **Group by** list, select: Payment Type
- Analyze metrics such as:
    - Average delivery days
    - Revenue or profit
- Your aggregation should look the same as the snapshot below.



Group By

Specify the columns to group by and one or more outputs.

○ Basic ● Advanced

PaymentType

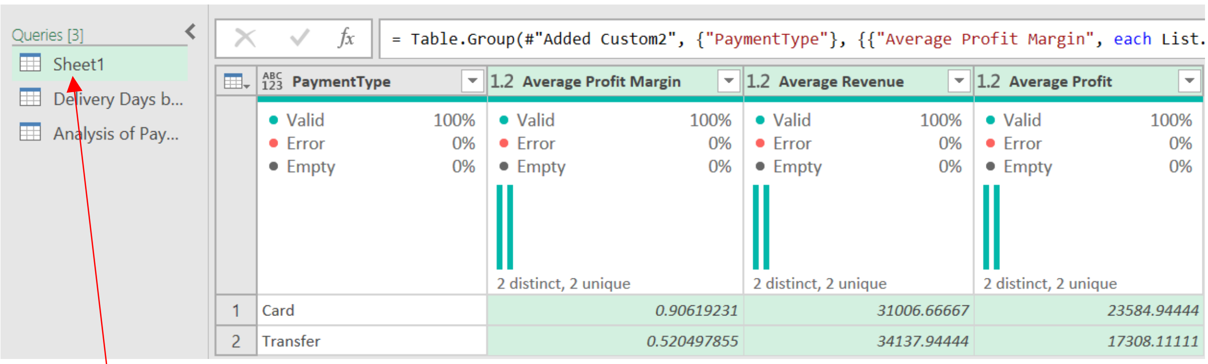
Add grouping

New column name	Operation	Column
Average Profit Margin	Average	Profit Margin
Average Revenue	Average	Revenue
Average Profit	Average	Profit

Add aggregation

OK Cancel

Your result:



```
= Table.Group(#"Added Custom2", {"PaymentType"}, {{"Average Profit Margin", each List.
```

	PaymentType	Average Profit Margin	Average Revenue	Average Profit
1	Card	0.90619231	31006.66667	23584.94444
2	Transfer	0.520497855	34137.94444	17308.11111

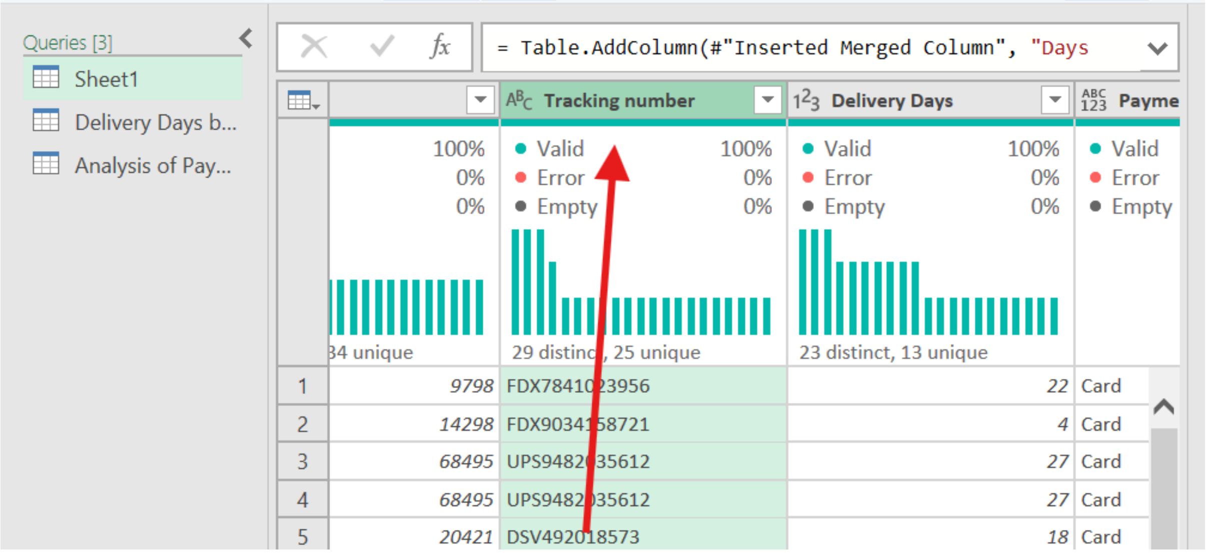
2.14 AI-Assisted Transformation

2.14.1 Extract Shipping Provider (AI Feature)

UPS, DHL, FDX, and DSV provide shipping services. For Example, FDX7841023956, DSV492018573, and DHL5928143 are tracking numbers for FDX, DSV, and DHL, respectively.

Steps:

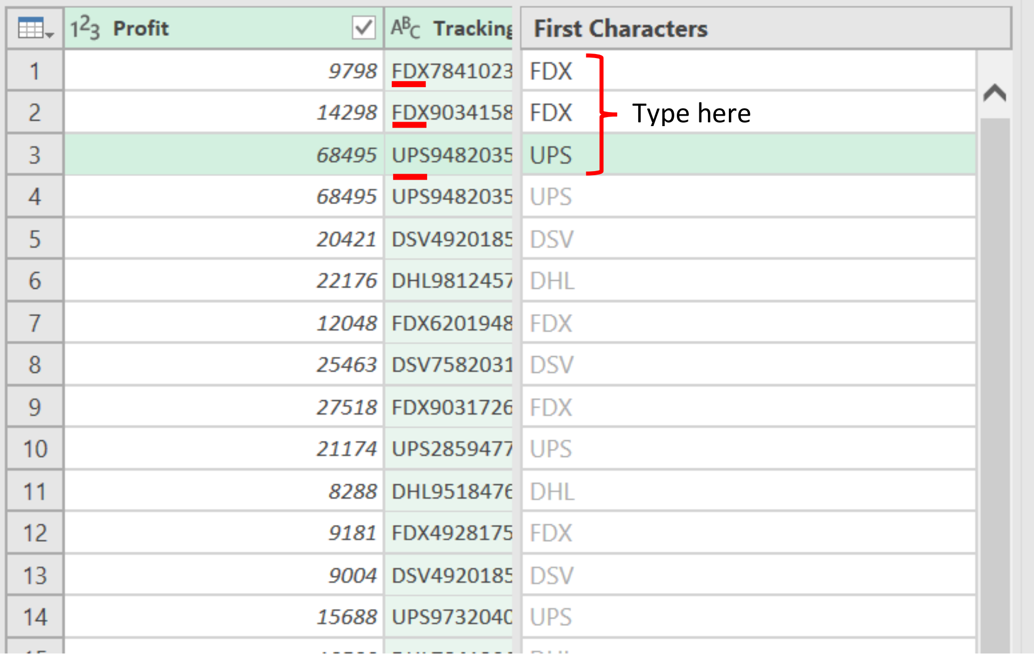
- Select Sheet1 query,
- Select the **Tracking Number** column.



```
= Table.AddColumn(#"Inserted Merged Column", "Days
```

		Tracking number	Delivery Days	Payme
1	9798	FDX7841023956	22	Card
2	14298	FDX9034158721	4	Card
3	68495	UPS9482035612	27	Card
4	68495	UPS9482035612	27	Card
5	20421	DSV492018573	18	Card

- Click on the **Add Column** tab,
- Select **Column from Examples**
- Manually type examples (UPS, DHL, FDX, DSV), in the new column corresponding to the first three characters in the Tracking column:



	Profit	Tracking	First Characters
1	9798	FDX7841023	FDX
2	14298	FDX9034158	FDX
3	68495	UPS9482035	UPS
4	68495	UPS9482035	UPS
5	20421	DSV4920185	DSV
6	22176	DHL9812457	DHL
7	12048	FDX6201948	FDX
8	25463	DSV7582031	DSV
9	27518	FDX9031726	FDX
10	21174	UPS2859477	UPS
11	8288	DHL9518476	DHL
12	9181	FDX4928175	FDX
13	9004	DSV4920185	DSV
14	15688	UPS9732040	UPS

Type here

- Power Query automatically generates the logic.

When you are done, close the Power Query and keep the changes. It is simple to return to Power Query by double-clicking on your queries, under **Queries & Connections.**

3 Final Reflection

Answer briefly:

- How did Power Query improve data quality?
- Which transformations were most important for analysis?
- How can this workflow scale to larger datasets?