

# Data Visualisation & Power BI

FINANCIALEDGE

neueda

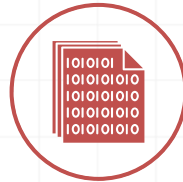
## Data Cleansing in Power BI

neueda  
FINANCIALEDGE

# What is Data Cleansing?



Process of preparing a dataset so it is in a **suitable format for analysis**



We often start our analysis with **untidy data**

**HUMAN ERROR**

**MACHINE OUTPUT**



Data cleansing steps

**IDENTIFYING ISSUES**

**RESOLVE THE ISSUES**

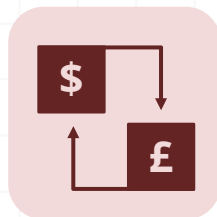
neueda<sup>7</sup>  
FINANCIALEDGE<sup>7</sup>

1



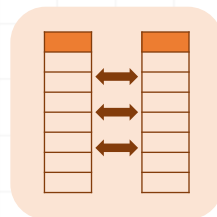
Removing **blank rows/ columns**

2



Replacing **values**

3



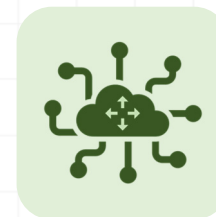
Splitting data into **separate columns**

4



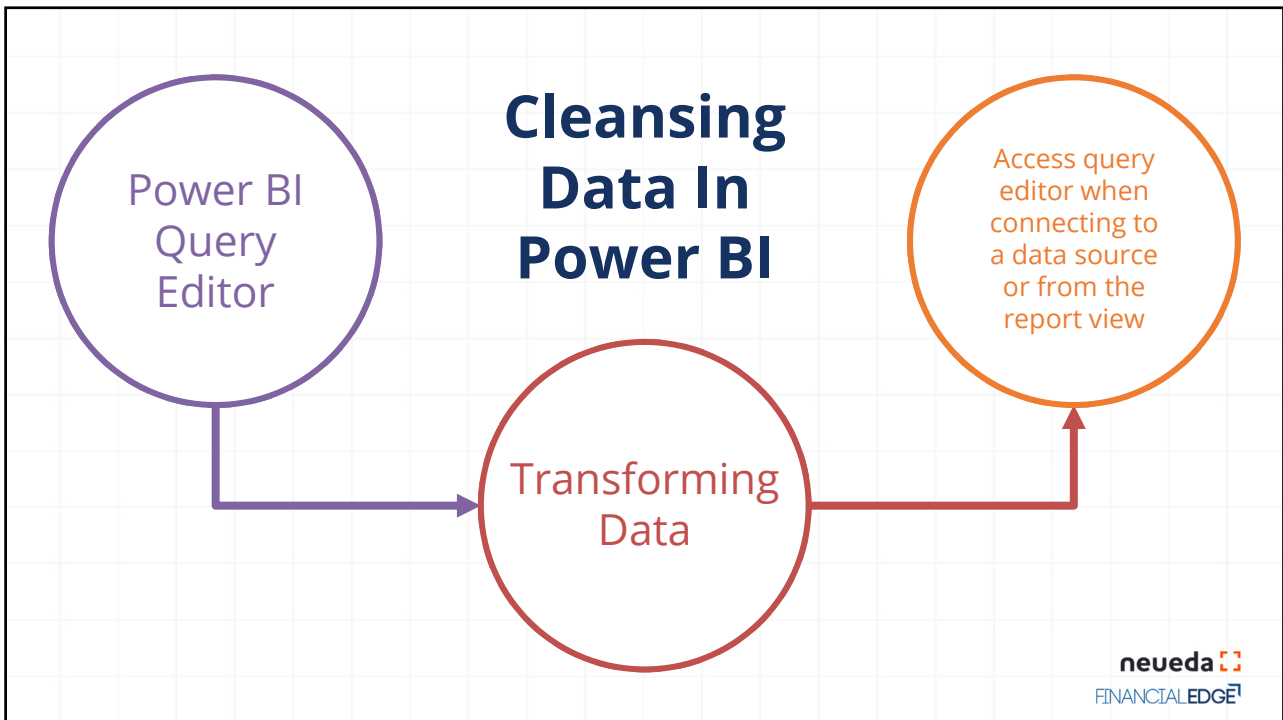
Identifying and removing **duplicates**

5



Changing the **data type**

## Common Data Cleansing Tasks



**Uncleansed Data in Query Editor**

	ABC Column1	ABC Column2	ABC Column3	ABC Column4
1		null	null	null
2	null	null	null	null
3	null	null	null	null
4	ISIN	Nominal Currency	null	Asset Class L2
5	ZKY096997709	USD	null	Real Estate,Real Estate,579621.84232
6	US5949181045	US Dollars	null	Equities,Equities,1930908.09344
7	US6792951054	USD	null	Equities,Equities,1586685.19219147
8	US6792951054	US Dolars	null	Equities,Equities,1586685.19219147
9	US7427181091	USD	null	Equities,Equities,544384.31512
10	GB00B082RF11	GBP	null	Equities,Equities,1749774.550296
11	US83304A1060	US Dolars	null	Equities,Equities,1720290.144576
12	US8552441094	US Dollars	null	Equities,Equities,1166044.454288
13	null	null	null	null
14	US06051GGU76	US Dollars	null	Fixed Income,Fixed Income,1713475.405536
15	US06051GHY89	US Dollars	null	Fixed Income,Fixed Income,584500.99924

neueda  
FINANCIALEDGE

## Cleansed Data in Query Editor

	ABC ISIN	ABC Nominal Currency	ABC Asset Class	1.2 USD Value
1	ZKY096997709	USD	Real Estate	579621.8423
2	US5949181045	USD	Equities	1930908.093
3	US6792951054	USD	Equities	1586685.192
4	US7427181091	USD	Equities	544384.3151
5	GB00B082RF11	GBP	Equities	1749774.55
6	US83304A1060	USD	Equities	1720290.145
7	US8552441094	USD	Equities	1166044.454
8	US06051GGU76	USD	Fixed Income	1713475.406
9	US06051GHY89	USD	Fixed Income	584500.9992
10	US06051GGJ22	USD	Fixed Income	1135902.808
11	US06738EBL83	USD	Fixed Income	174719.2799
12	US172967LN99	USD	Fixed Income	1132973.506
13	US38145GAF72	USD	Fixed Income	1143566.433
14	US38141GWU48	USD	Fixed Income	1395890.47

## Removing Columns

# 1 Removing Empty Columns

APC Location	APC Type	APC Column9	APC Column10	APC Column11	APC Column12	APC Column13
USA	Asset	null	null	null	null	null
USA	Asset	null	null	null	null	null
USA	Asset	null	null	null	null	null
USA	Asset	null	null	null	null	null
USA	Asset	null	null	null	null	null

Visualizations >> Fields >>>

Build visual

Search

- Assets1
  - Asset Class L1
  - Asset Class L2
  - Column10
  - Column11
  - Column12
  - Column13
  - Column9
  - ISIN
  - Location
  - Market Value USD
  - Nominal Currency

Values

Add data fields here

“Null” values in every row

Available on the field list alongside any other columns of data

# 2 Removing Columns Containing Partial Values



Columns containing a large percentage of null values which are of no use

APC Location

- Valid 8%
- Error 0%
- Empty 92%

USA

USA

null

null

Use Column Profiling to help identify these columns



### 3 Removing Columns Containing Values

Columns containing data which is **irrelevant in the report**

Columns containing data which is the same for every row – **use the filter**

Month
Apr
Apr
Apr
Apr
Apr
Apr
Apr
Apr

**FILTER**

The screenshot shows a 'Removing Columns' dialog box with the following options: Sort Ascending, Sort Descending, Clear Sort, Clear Filter, Remove Empty, and Text Filters. Under 'Text Filters', there is a search field and two checked items: '(Select All)' and 'Apr'.

## Removing Rows

The diagram illustrates the process of removing rows from a table. It shows a grid with several rows. A hand is shown reaching in from the right side to remove a row. The text 'Removing Rows' is centered in the middle of the grid.

## Removing Rows

Remove blank rows (entire row)

Not just an **empty cell**

APc ISIN	APc Product	APc Nominal Currency	APc Asset Class L1	APc Asset Class L2
ZKY096997709	BRAVO III OFF SEL FEEDER F B LTD. USD NON RDR 2ND CL	USD	Real Estate	Diversified Real Estate
US5949181045	MICROSOFT CORP	USD	Equities	Developed Equities
US6792951054	OKTA INC	USD	Equities	Developed Equities
US7427181091	PROCTER & GAMBLE CO/THE	USD		<i>null</i> Developed Equities
GB00B082RF11	RENTOKIL INITIAL PLC	GBP	Equities	Developed Equities
US83304A1060	SNAP INC - A	USD	Equities	Developed Equities
US8552441094	STARBUCKS CORP	USD	Equities	Developed Equities
		<i>null</i>	<i>null</i>	<i>null</i>
		<i>null</i>	<i>null</i>	<i>null</i>
US6051GGU76	BANK OF AMERICA CORP (M) 01-Oct-21 (R) 2.08338	USD	Fixed Income	Developed Corporate Investment Grade



Nominal Currency

- (Blank)
- CHF
- EUR
- GBP
- USD
- XAU

## Removing Rows

Top *n* Rows

Bottom *n* Rows

Trade date	Maturity date	APc Trade ID	1.2 USD Value	APc Company
26/02/2021	05/05/2022	000708829056	13620.4496	FUS
23/02/2021	03/05/2022	000858807814	1700677.37	WOR
28/02/2021	02/05/2022	000898578225	1957352.22	COU
23/02/2021	28/04/2022	000169440295	973539.6	CIR
20/02/2021	25/04/2022	000701518726	1038266.7	MED
28/02/2021	25/04/2022	000736999339	1908584.85	COG
28/02/2021	23/04/2022	000210215309	1588821.71	AEW
13/02/2021	22/04/2022	00080811746	5869.9522	ABC
27/02/2021	21/04/2022	000936513119	889674.24	CYB
11/02/2021	21/04/2022	000170534542	495816.66	HIG
18/02/2021	19/04/2022	000617264248	755466.52	GVC
<i>null</i>	<i>null</i>	Total Value	11327690.27	<i>null</i>
<i>null</i>	<i>null</i>	Average Value	1029790.025	<i>null</i>



## Promoting Row Headers

ABC 123 Column2	ABC Column3	ABC 123 Column4	ABC Column5
null	null	null	null
null	null	null	null
null	null	null	null
null	null	null	null
null	null	null	null
Maturity date	Trade ID	USD Value	Company
05/05/2022	000708829056	13620.4496	FUS
03/05/2022	000748290587	1700677.37	WOR
02/05/2022	000898578225	1957352.22	COU
28/04/2022	000169440295	973539.6	CIR
25/04/2022	000701518726	1038266.7	MED

Column names won't change even if blank rows removed

Should be column headings

ABC 123 Column2	ABC Column3	ABC 123 Column4	ABC Column5
Maturity date	Trade ID	USD Value	Company
05/05/2022	000708829056	13620.4496	FUS
03/05/2022	000748290587	1700677.37	WOR
02/05/2022	000898578225	1957352.22	COU
28/04/2022	000169440295	973539.6	CIR
25/04/2022	000701518726	1038266.7	MED

Fields >>

Search

- Assets Promote Hea... >>
- Assets
- Column2
- Column3
- Column4
- Column5
- Column6
- Column7

Promote row of data  
to **header row**

- Assets Promote Headers
- Asset Class L1
- Asset Class L2
- ISIN

Create correct  
**field names**

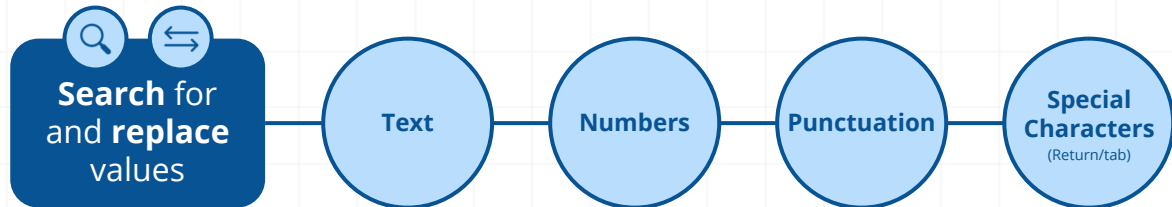
Maturity date	Trade ID	USD Value	Company
05/05/2022	708829056	13620.4496	FUS
03/05/2022	748290587	1700677.37	WOR
02/05/2022	898578225	1957352.22	COU
28/04/2022	169440295	973539.6	CIR
25/04/2022	701518726	1038266.7	MED

**Fields** >>

- Assets Promote Headers
- Asset Class L1
- Asset Class L2
- ISIN
- Location
- Market Value USD
- Nominal Currency
- Product

## Replace Values

## What is Replace Values?



Applies to the **whole column**

**Quick & easy**

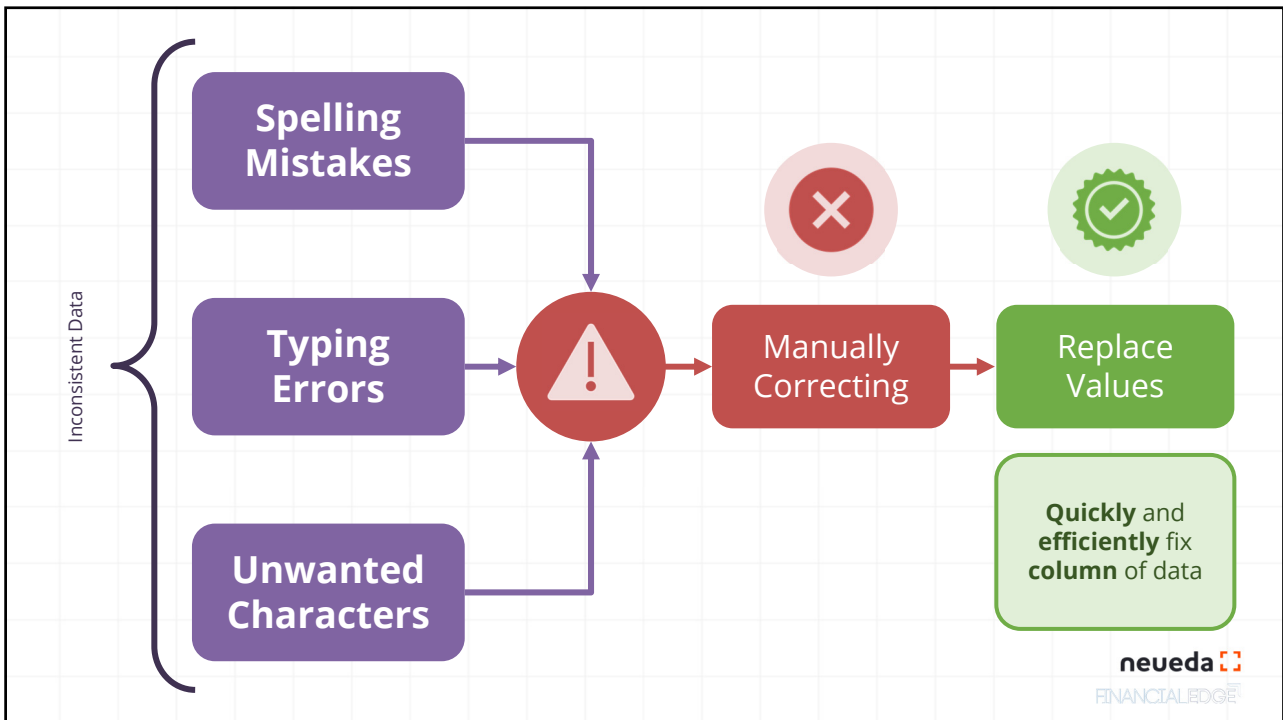
**Case-sensitive** search  
(JOE, Joe, joe)

ANALYSING

CONSISTENT

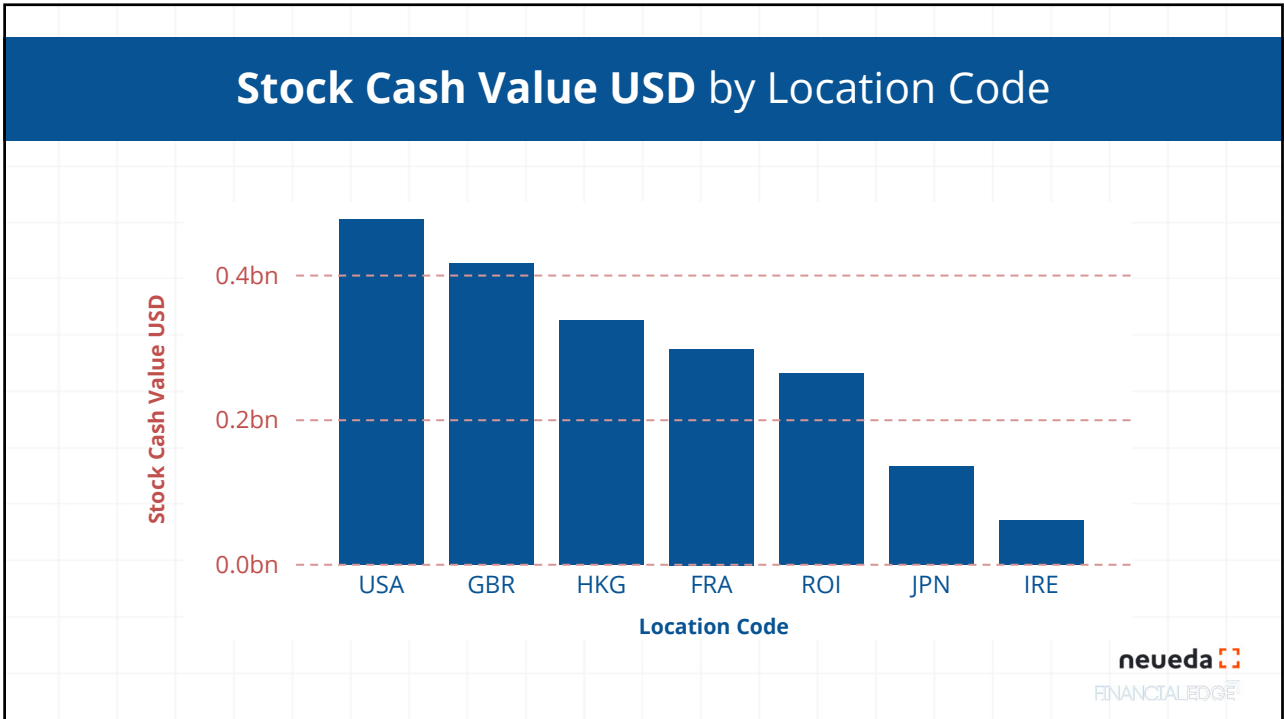
ACCURATE

VISUALISING



Actual Settlement Date	Status	Stock Cash Value USD	Location Code	Region
01 September 2021	GOOD	29736295.98	FRA	#Europe#
13 March 2021	FAIL	10418800.17	USA	#NAM#
18 August 2021	GOOD	19146481.76	HKG	#APEC#
18 June 2021	FAIL	22377650.78	ROI	#Europe#
17 October 2021	FAILED	16794155.3	IRE	#Europe#
19 October 2021	FAILED	1624893.43	HKG	#APEC#
12 May 2021	GOOD	8972586.71	ROI	#Europe#
29 October 2021	GOOD	234471.88	USA	#NAM#
29 June 2021	FAILED	31485112.43	USA	#NAM#
07 July 2021	GOOD	3319077.45	USA	#NAM#

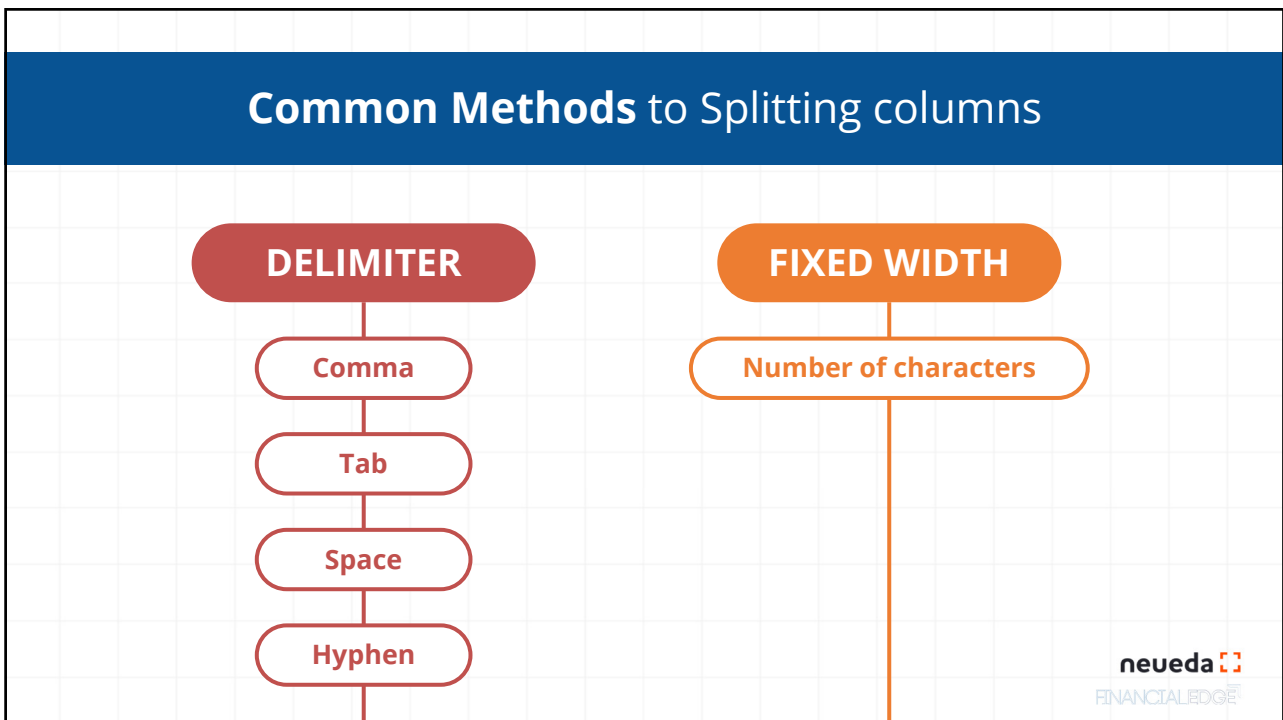
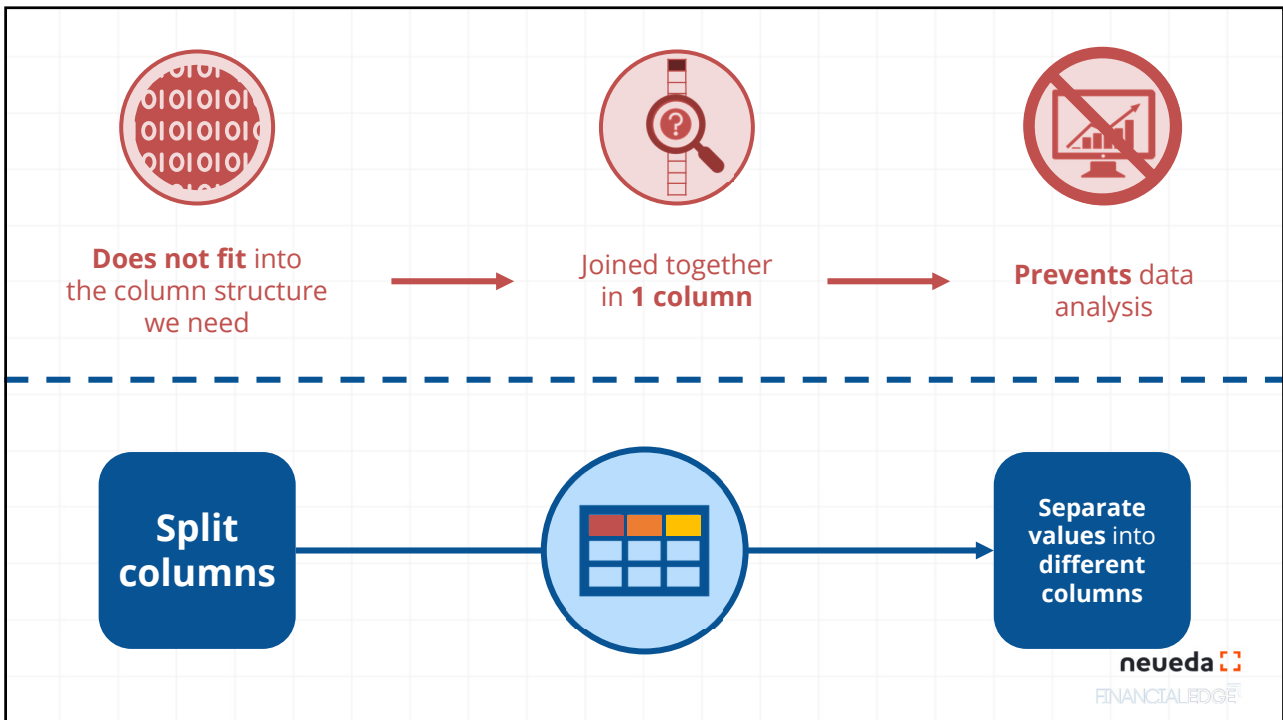
neueda  
FINANCIALEDGE




### Split Columns

NAME	SCORE 1	SCORE 2	SCORE 3
Jack	23	45	66


neueda  
FINANCIALEDGE



**EXAMPLE 1**



**Delimited**




**Fixed Width**  
(Number of characters)


A <sup>B</sup> C No of Shares Traded-Currency
253000-GBP
7000-GBP
191000-EUR
129000-EUR
377000-EUR
3810-GBP
481000-EUR
587000-EUR
555000-EUR
59500-EUR
531000-USD

**neueda**  
FINANCIALEDGE

**EXAMPLE 2**



**Delimited**  
(Right most delimiter only)




**Fixed Width**  
(Number of characters)


A <sup>B</sup> C Company Name Stock Cash Value
BlackRock 26268273
HSBC Holdings 1240377
Merril Lynch 23362865
JPMorgan 18164512
Nomura Asset Management (NAM) 25200...
BlackRock 25013095
JPMorgan 17232386
Goldman Sachs 2472248
Goldman Sachs 7634395
Barclays 1343259
Barclays 11105157

**neueda**  
FINANCIALEDGE

**EXAMPLE 3**



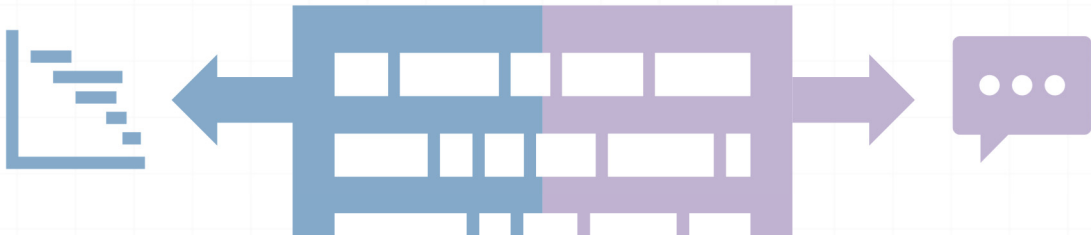
**Delimited**



**Fixed Width**  
(Number of characters 12)

A <sup>B</sup> C	ISINMaturity
GB1098963675	21/07/22
GB1098963700	11/03/22
FR123456791508	06/06/22
NL237297160906	06/05/22
IT242525967912	09/09/22
IT242525964014	09/09/22
DE120005605529	10/10/22
IT242525964622	12/12/22
SI079188858911	11/12/22
DE120005604627	12/12/22

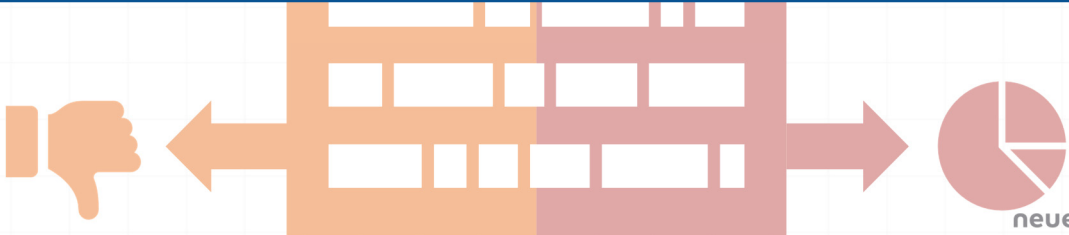
**neueda**  
FINANCIALEDGE<sup>1</sup>



---


Data Types

---




**neueda**  
FINANCIALEDGE<sup>1</sup>


## Data Types in Power BI




Each column is **automatically assigned** a data type



**Common Data Types:**  
Whole Number,  
Decimal Number,  
Date, Text



Data Type will determine how a field can be used in the **report visuals**

**neueda**   
FINANCIALEDGE

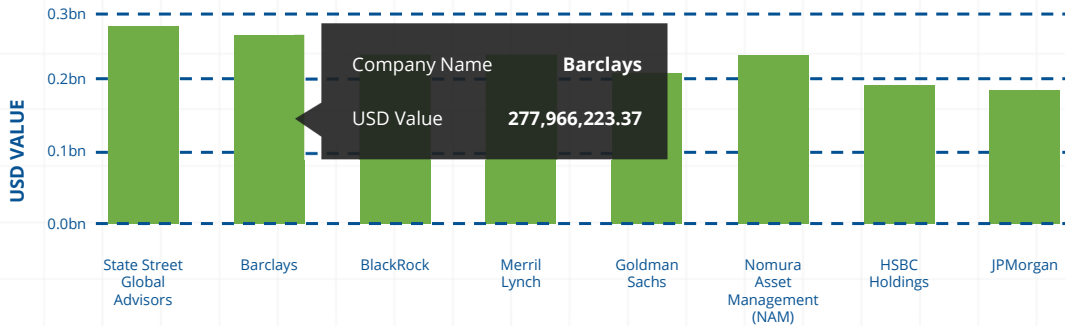
## Price of Gold Jan 2021 to Mar 2022



To create a **line chart** such as this, the **x axis** requires a field which is a **Date** or **Time Data Type**

**neueda**   
FINANCIALEDGE

## Total USD Value by Company Name



To perform any type of calculation on a field such as **Sum or Average**, a field must be **Numeric Data Type**



You cannot sum or average a **text field** – even if it contains numbers

## Changing a Data Type in Power BI

### Report View or Data View

Trade date	Trade ID	Currency	USD Value
23 Febru	858807814	GBP	1700677.37
28 Febru	898578225	GBP	1957352.22
23 Febru	169440295	GBP	973539.6
20 Febru	701518726	GBP	1038266.7
28 Febru	736999339	GBP	1908584.85
28 Febru	210215309	GBP	1588821.71
11 Febru	170534542	GBP	495816.66
18 Febru	617264248	GBP	75466.52
11 Febru	638133856	GBP	230565.24

Trade ID	Company	Sector
13620.4496	FLUS	Software & Computer S
1700677.37	WOR	Software & Computer S
1957352.22	COU	Forestry & Paper
973539.6	CIR	Health Care Equipmen
1038266.7	MED	Equity Investment Instr
1908584.85	COG	Support Services
1588821.71	AEM	Equity Investment Instr

## Identifying Data Types in a Power BI Report

The **icon** beside the **field name** in the Power BI Report **indicates a Numeric or Date/Time Data Type**

**No icon = Text**

**Calendar icon = Date/Time**

**Greek Sigma = Numeric**

The Fields pane shows the following fields and their icons:

- Company Name (No icon)
- Currency Code (No icon)
- Date (Calendar icon)
- Date Hierarchy (Calendar icon)
- Nominal (No of ... (Greek Sigma icon)
- USD Value (Greek Sigma icon)
- Value (Greek Sigma icon)

neueda  
FINANCIALEDGE

## Identifying the Data Types in Query Editor

Icon beside field name indicates the Data Type Range of Data Types:

**DATE**

**WHOLE NUMBER**

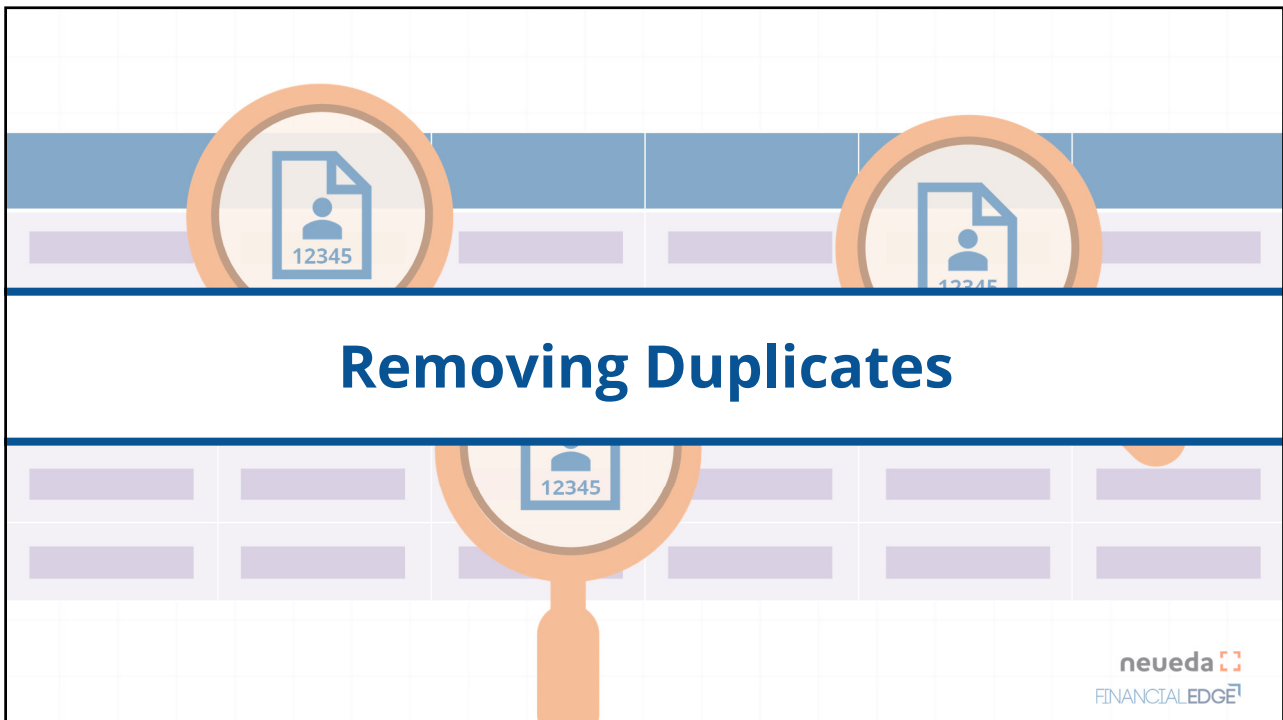
**DECIMAL NUMBER**

**TEXT**


**ANY DATA TYPE**

Date	Nominal	USD Value	Code	Settled
04/07/2022	221000	25055027.83	EUR	No
04/03/2022	255000	26323820	EUR	No
04/06/2022	289000	26903058.17	EUR	Yes
04/05/2022	319000	26841510.67	EUR	Yes
04/09/2022	357000	25992872.5	EUR	Yes

neueda  
FINANCIALEDGE



## Removing Duplicates


neueda   
FINANCIALEDGE<sup>1</sup>

## Understanding Duplicates

Most tables will have **1 column** which contains **unique IDs**:

- Transaction ID
- ISIN
- Client Reference
- Company Code

Duplicates can **exist** in a **single column** or across the **entire row**

neueda   
FINANCIALEDGE<sup>1</sup>

## Duplication in a Single Column

### EXAMPLE 1

ISIN	Company Name	Nominal (No of Shares Traded)
GB1098963700	BlackRock	221000
FR1234567915	Merril Lynch	255000
NL2372971609	Goldman Sachs	459000
IT2425259679	BlackRock	319000
IT2425259640	State Street Global Advisors	357000
DE1200056055	BlackRock	389000
IT2425259646	State Street Global Advisors	389000
NL2372971609	Goldman Sachs	459000

## Column Distribution



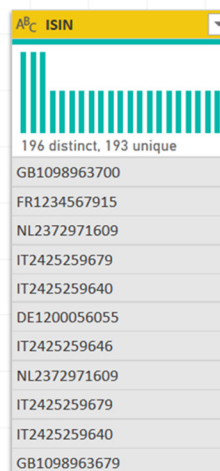
Identifies **distinct** and **unique values** in a column

**Distinct** – number of different values in that column

**Unique** – how many of those values are unique



**Less unique values than distinct = duplication**



193 unique ISIN values

196 different ISIN values

3 values used more than once

## Duplication Across All Columns

### EXAMPLE 2

Datasets don't always have a column which **uniquely identifies each row**

Company Name	Nominal (No of Shares Traded)	Market Price	Currency Code
BlackRock	221000	113.3711667	EUR
Merril Lynch	255000	103.2306667	EUR
Merril Lynch	255000	103.2306667	EUR
Merril Lynch	319000	84.14266667	EUR
State Street Global Advisors	357000	72.80916667	EUR

Check for duplication