Power Query for Data Analysis

Satyam Govila

To Do:-

- Apply all the concepts learned till date in a mini in-class project
- The project should include loading of a csv file in MS Excel
- Cleaning of data
- Data Quality Checks
- Data Visualisation
- Data Entry Validation
- Creating informative features from existing data

Challenges of MS Excel

- Challenge No. 1: Built for Individuals, Not Collaboration
- Challenge No. 2: Unavoidable Errors
- Challenge No. 3: Lack of Data Protection (Data Security , Governance , Backups)
- Challenge No. 4: Not Optimum for Robust Applications

Need of advanced tool: PowerQuery

Power Query

What is Power Query?
When should I use it?

- Power Query is a user-friendly and powerful **Excel add-in** that allows you to import, clean and consolidate data.
- Uses of Power Query:

  Processing large amounts of data (where Excel is limited to one million rows)

  Importing data from different sources (e.g. Excel, CSV, SharePoint, SQL, and Salesforce)

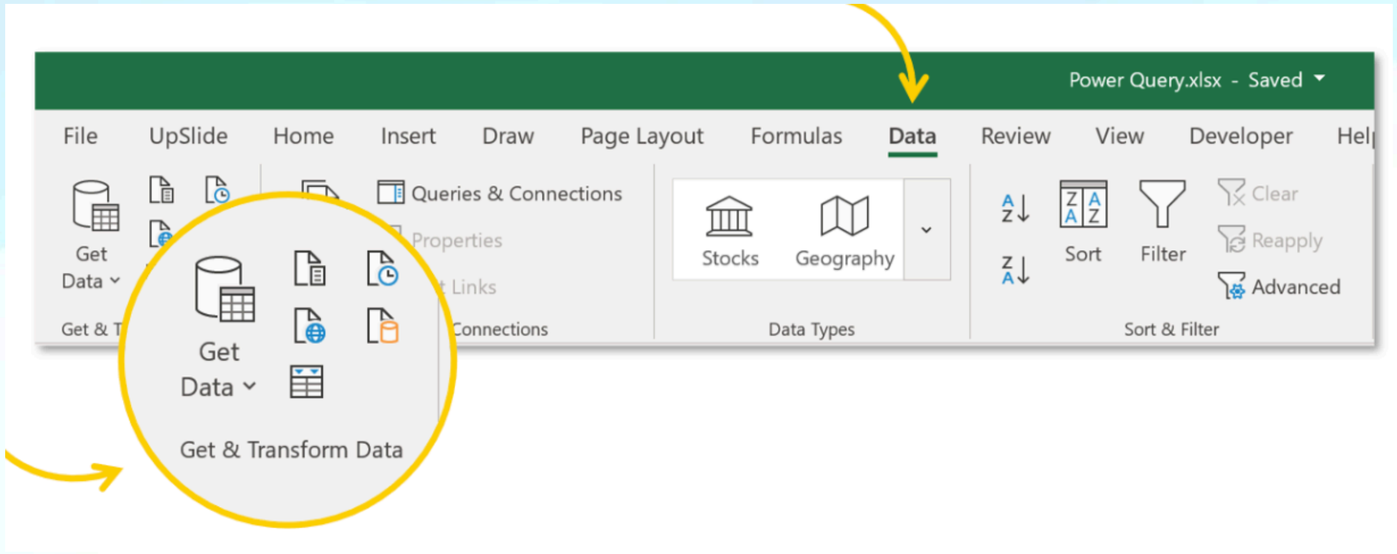
  Manipulating, cleaning and combining several tables, files or even folders!

Power Query

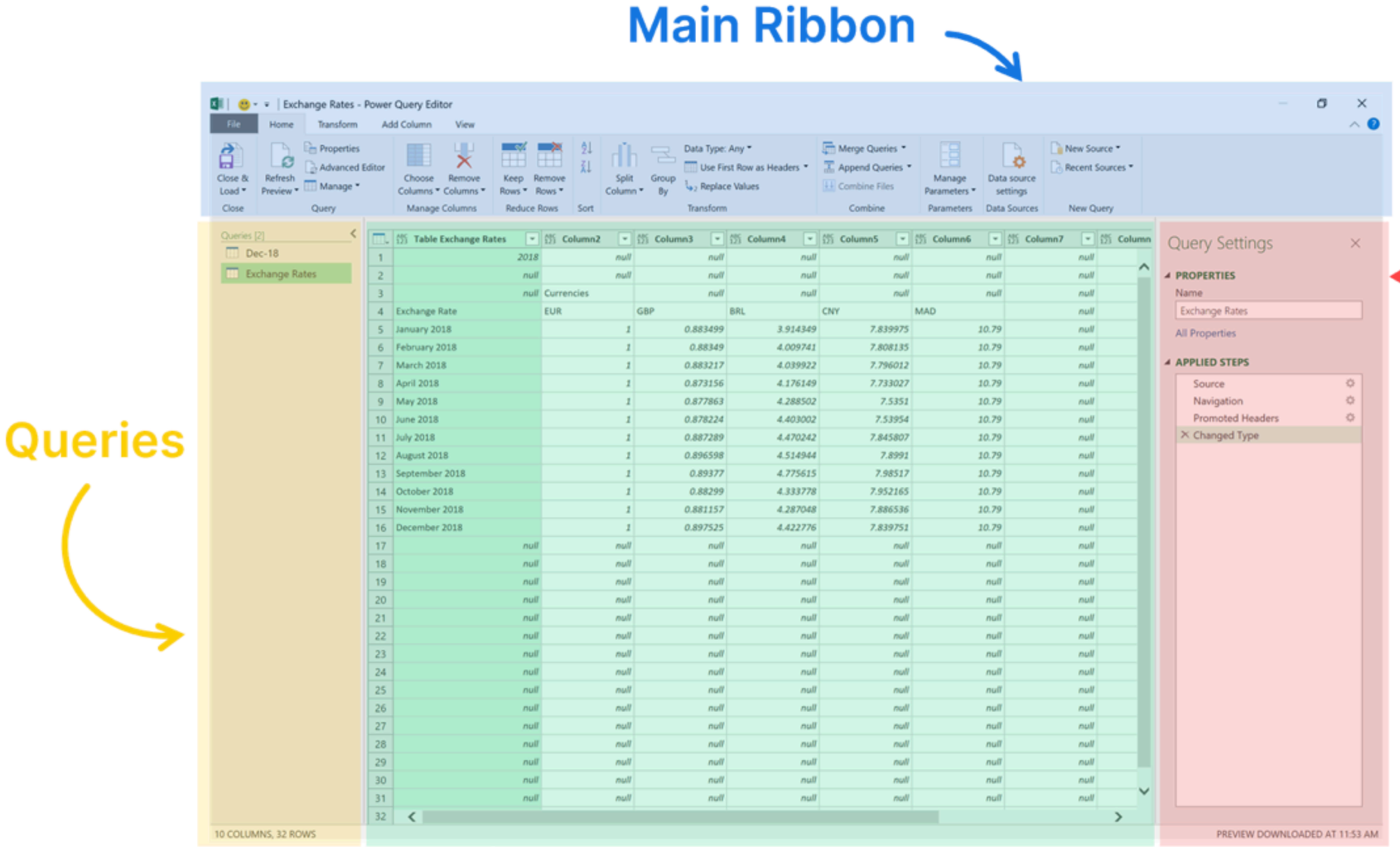
- **Pre-requisite:** Excel 2016 and later versions. You can use 2010 or 2013 versions by installing external add-in.
- Get & Transform section in Data tab using ‘New Query’
- Used for ETL (Extract, Transform and Load) in data analysis
- Multiple sources of data from web, database, files, etc
- After we get data, transform the data.
- Remove additional columns, duplicate data, blank data, etc to create tables used for further analysis is done using PowerQuery.
- After transformation, load the data.
- ‘Refresh’ is the most powerful feature of PowerQuery. This is important for recurring tasks.

Get started with Power Query

Download / Setup Power Query



Working of Power Query



Working of Power Query

- **Main ribbon:** to create queries from scratch thanks to its buttons (if you are familiar with the M language, you can also access the Advanced Editor to tweak them manually).
- **Queries:** to keep track of all connections to data sources and streams.
- **Data Table preview:** to visualize your data result after a specific step.
- **Query Settings (Applied Steps):** to navigate through all the steps you made. Power Query automatically records all transformations made.

Query Editor

- Where the transformation is done.
- Opening blank Query for Query Editor
- Close & Load option to a file, connection, table, etc
- Options to split columns, remove columns, use first row as headers, etc
- Transform tab – Changing data type, transpose, etc
- Add column tab – adding new columns including custom columns
- View tab – Viewing query settings, formula bar, etc
- Applied Steps is where tracking is done and acts as a recorder for all transformation steps.
- Changing source and keeping other steps intact is how refresh can be done for new data (for recurring tasks)

Import Data

- Importing data through web in the form of tables
- Transform/Edit data in Query Editor – Applied steps
- Remove unwanted columns, first row as headers, and check data types for all columns
- Remove wrong steps in Applied Steps section
- Close and Load to get into Excel worksheet
- Refresh the report by connecting with the link
- Importing data through files – CSV, Excel workbook
- Change source and refresh to update data
- Shows objects in Excel workbook including tables

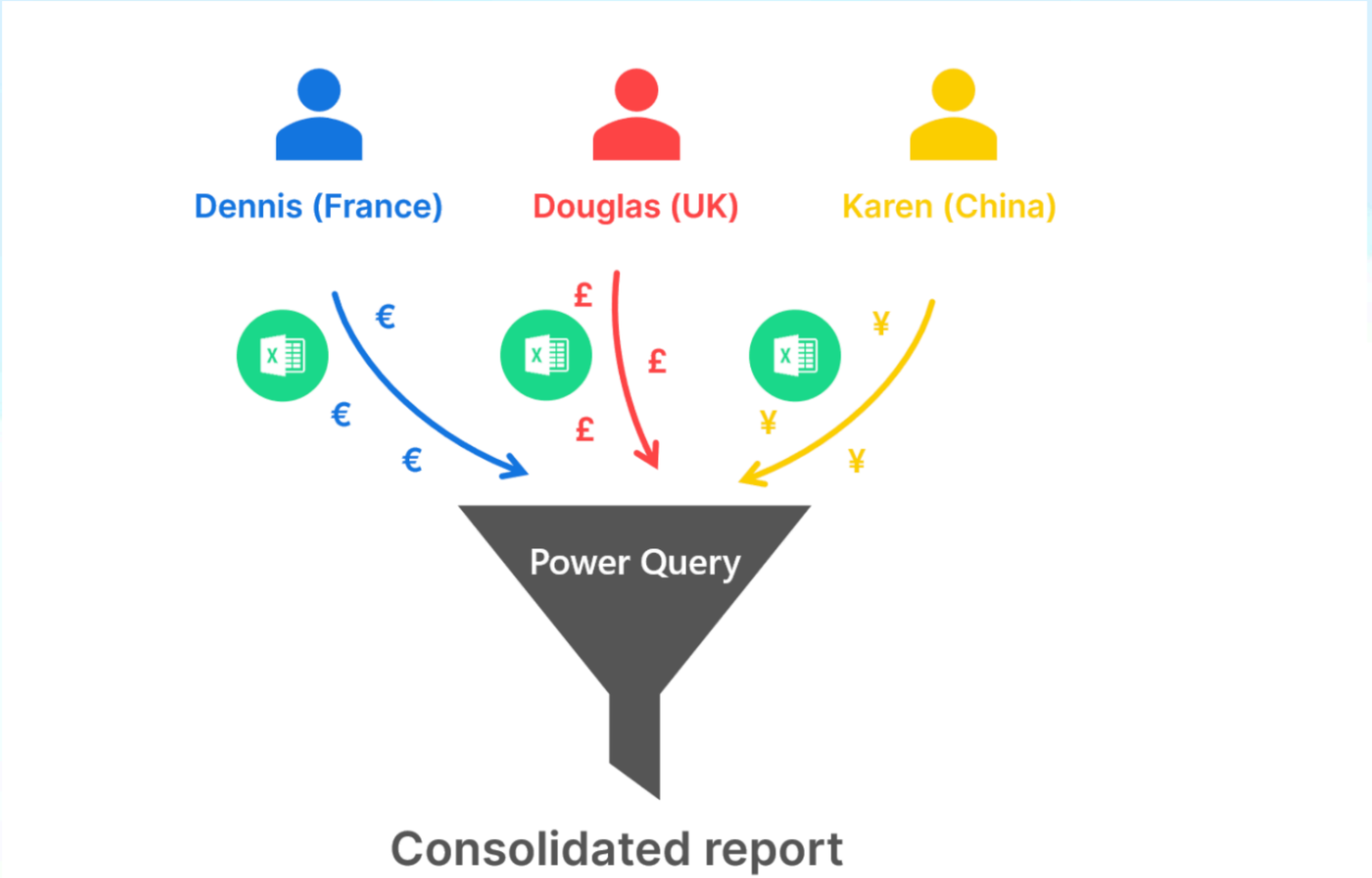
Merge and Append

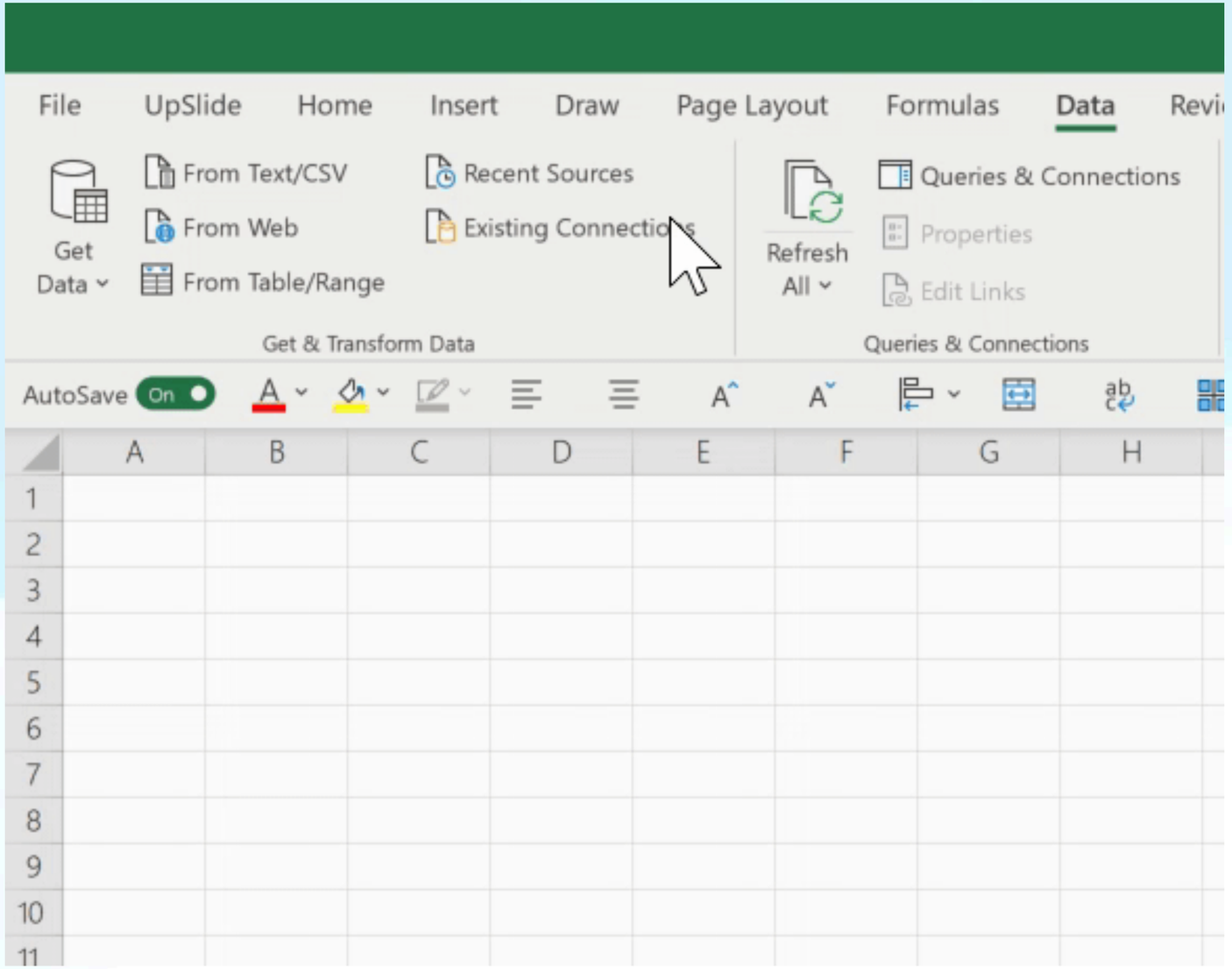
- Create Connections for append and merge instead of load
- Append – Adding more data/horizontal rows of same structure
- Merge – Adding new columns to the existing data using joins (Left outer)
- While merging - selecting the columns that we need, prefix from original column option
- Helps in creation of pivot tables
- Append using different Excel Workbooks, split/merge columns, etc can be done

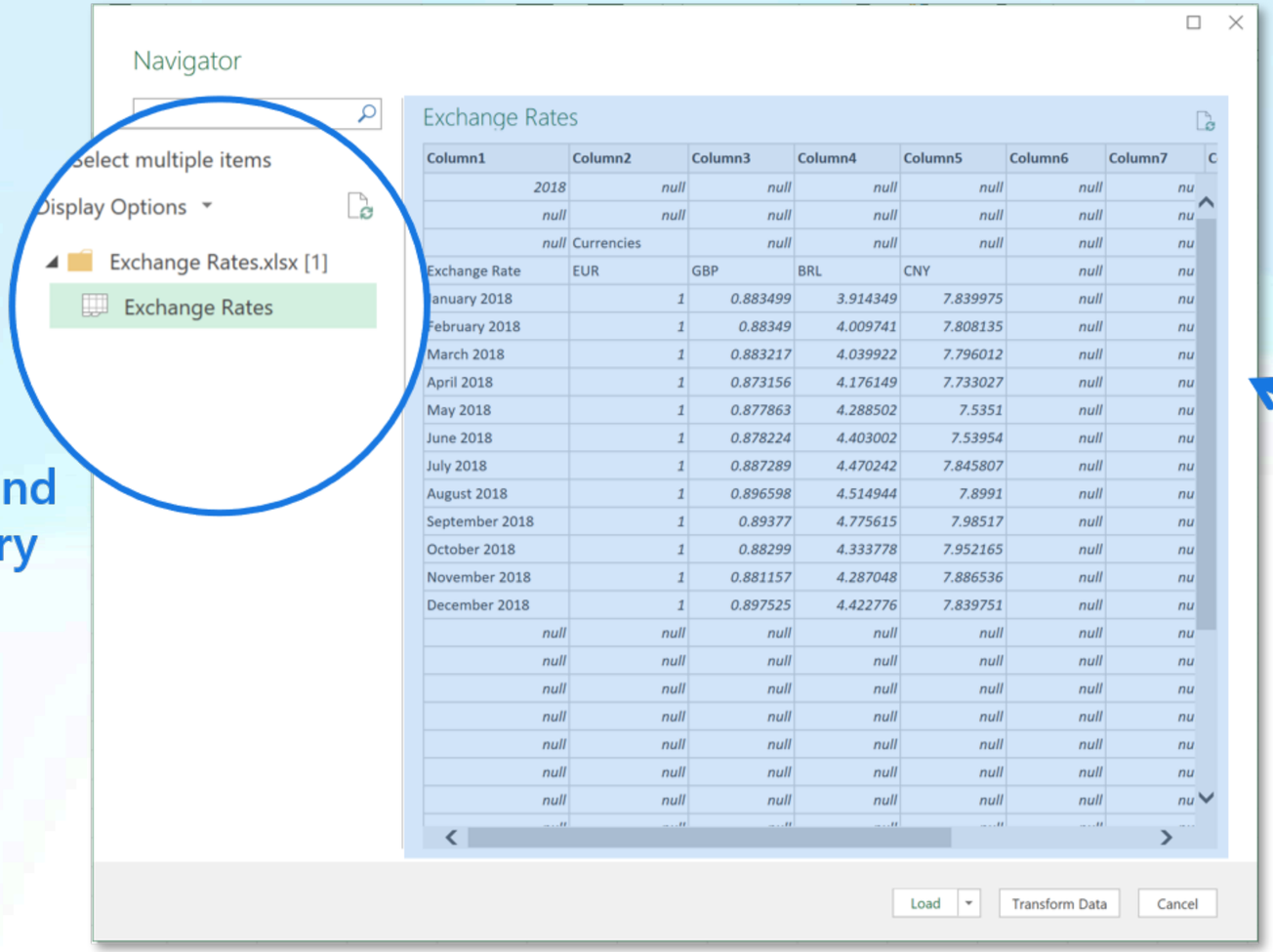
Merge and Append

- Create Connections for append and merge instead of load
- Append – Adding more data/horizontal rows of same structure
- Merge – Adding new columns to the existing data using joins (Left outer)
- While merging - selecting the columns that we need, prefix from original column option
- Helps in creation of pivot tables
- Append using different Excel Workbooks, split/merge columns, etc can be done

Demo : Use Case



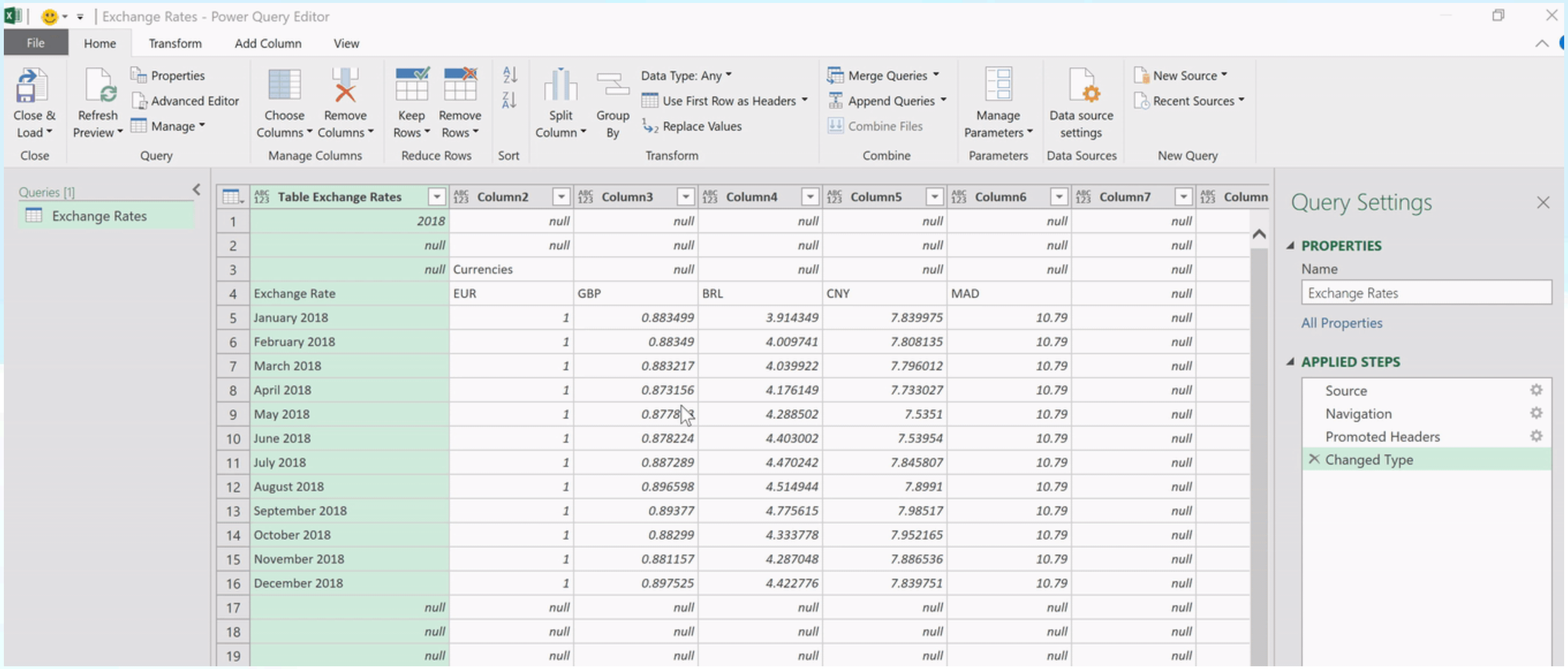

File
UpSlide
Home
Insert
Draw
Page Layout
Formulas
Data
Get Data
From Text/CSV
From Web
From Table/Range
Recent Sources
Existing Connections
Get & Transform Data
Refresh All
Queries & Connections
Properties
Edit Links
Queries & Connections
AutoSave
On
A
B
C
D
E
F
G
H
1
2
3
4
5
6
7
8
9
10
11

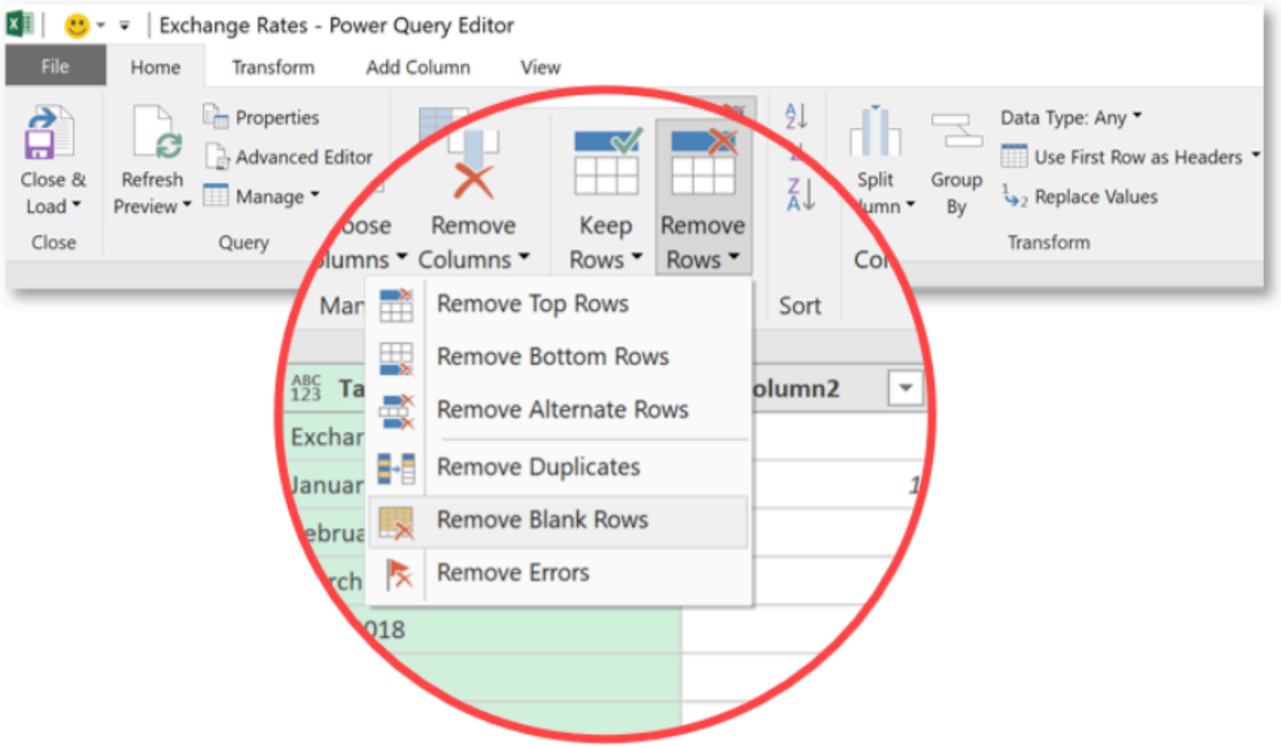


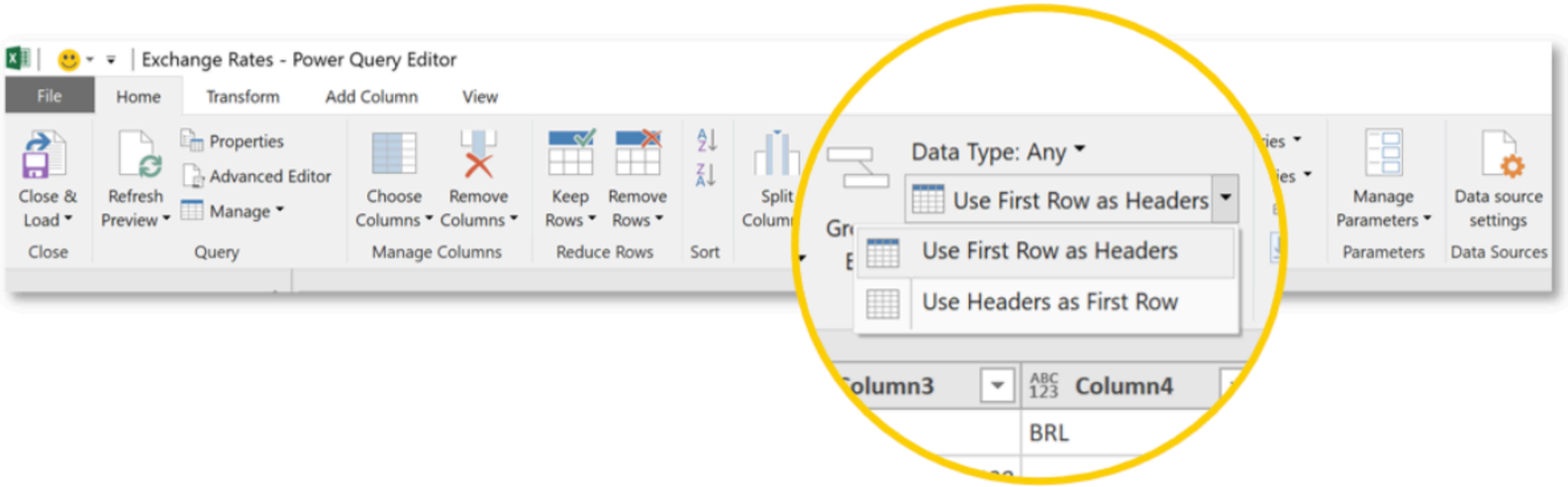
Column1	Column2	Column3	Column4	Column5	Column6	Column7	C
2018	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu
null	Currencies	null	null	null	null	null	nu
Exchange Rate	EUR	GBP	BRL	CNY	null	null	nu
anuary 2018	1	0.883499	3.914349	7.839975	null	null	nu
February 2018	1	0.88349	4.009741	7.808135	null	null	nu
March 2018	1	0.883217	4.039922	7.796012	null	null	nu
April 2018	1	0.873156	4.176149	7.733027	null	null	nu
May 2018	1	0.877863	4.288502	7.5351	null	null	nu
June 2018	1	0.878224	4.403002	7.53954	null	null	nu
July 2018	1	0.887289	4.470242	7.845807	null	null	nu
August 2018	1	0.896598	4.514944	7.8991	null	null	nu
September 2018	1	0.89377	4.775615	7.98517	null	null	nu
October 2018	1	0.88299	4.333778	7.952165	null	null	nu
November 2018	1	0.881157	4.287048	7.886536	null	null	nu
December 2018	1	0.897525	4.422776	7.839751	null	null	nu
null	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu
null	null	null	null	null	null	null	nu

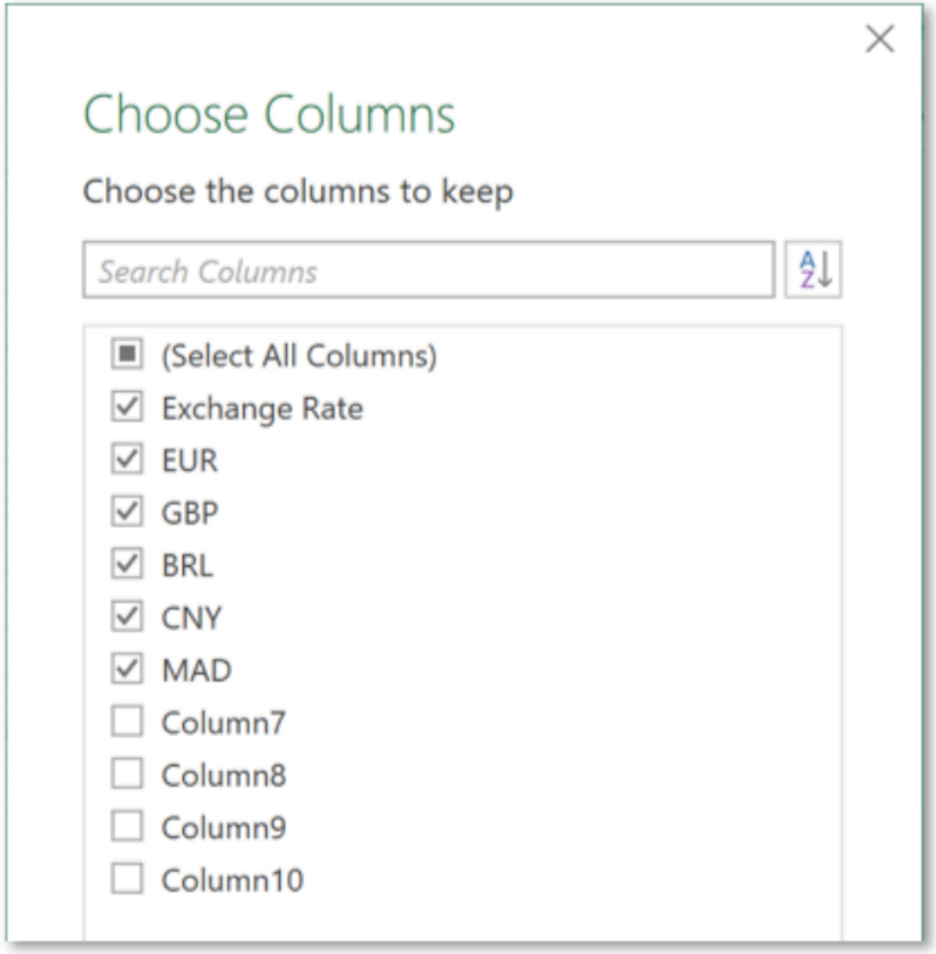
Data tables found by Power Query

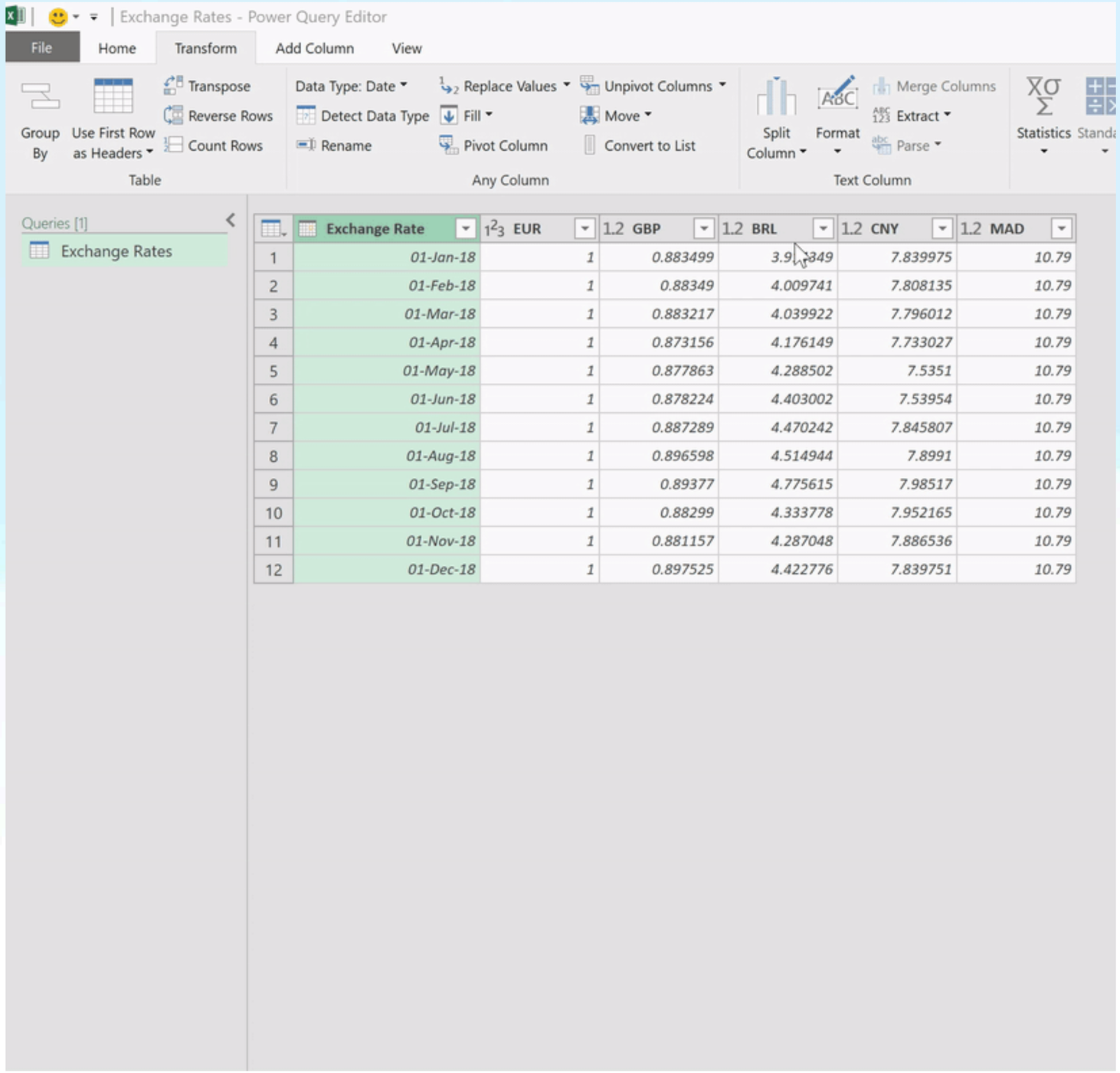
Data preview


Exchange Rates - Power Query Editor
File
Home
Transform
Add Column
View
Close & Load
Refresh Preview
Properties
Advanced Editor
Manage
Close
Query
Choose Columns
Remove Columns
Manage Columns
Keep Rows
Remove Rows
Reduce Rows
Sort
Split Column
Group By
Data Type: Any
Use First Row as Headers
Replace Values
Transform
Merge Queries
Append Queries
Combine Files
Combine
Manage Parameters
Parameters
Data source settings
Data Sources
New Source
Recent Sources
New Query
Queries [1]
Exchange Rates
Table Exchange Rates
Column2
Column3
Column4
Column5
Column6
Column7
Column
1 2018 null null null null null null
2 null null null null null null null
3 null Currencies null null null null null
4 Exchange Rate EUR GBP BRL CNY MAD null
5 January 2018 1 0.883499 3.914349 7.839975 10.79 null
6 February 2018 1 0.88349 4.009741 7.808135 10.79 null
7 March 2018 1 0.883217 4.039922 7.796012 10.79 null
8 April 2018 1 0.873156 4.176149 7.733027 10.79 null
9 May 2018 1 0.8778 4.288502 7.5351 10.79 null
10 June 2018 1 0.878224 4.403002 7.53954 10.79 null
11 July 2018 1 0.887289 4.470242 7.845807 10.79 null
12 August 2018 1 0.896598 4.514944 7.8991 10.79 null
13 September 2018 1 0.89377 4.775615 7.98517 10.79 null
14 October 2018 1 0.88299 4.333778 7.952165 10.79 null
15 November 2018 1 0.881157 4.287048 7.886536 10.79 null
16 December 2018 1 0.897525 4.422776 7.839751 10.79 null
17 null null null null null null null
18 null null null null null null null
19 null null null null null null null
Query Settings
PROPERTIES
Name
Exchange Rates
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type


Exchange Rates - Power Query Editor
File
Home
Transform
Add Column
View
Close & Load
Refresh Preview
Properties
Advanced Editor
Manage
Close
Query
Remove Columns
Keep Rows
Remove Rows
Sort
Split
Group By
Data Type: Any
Use First Row as Headers
Replace Values
Transform
Remove Top Rows
Remove Bottom Rows
Remove Alternate Rows
Remove Duplicates
Remove Blank Rows
Remove Errors
Column2


Exchange Rates - Power Query Editor
File
Home
Transform
Add Column
View
Close & Load
Refresh Preview
Properties
Advanced Editor
Manage
Close
Query
Choose Columns
Remove Columns
Manage Columns
Keep Rows
Remove Rows
Reduce Rows
Sort
Split
Data Type: Any
Use First Row as Headers
Use First Row as Headers
Use Headers as First Row
Manage Parameters
Parameters
Data source settings
Data Sources
Column3
Column4
BRL

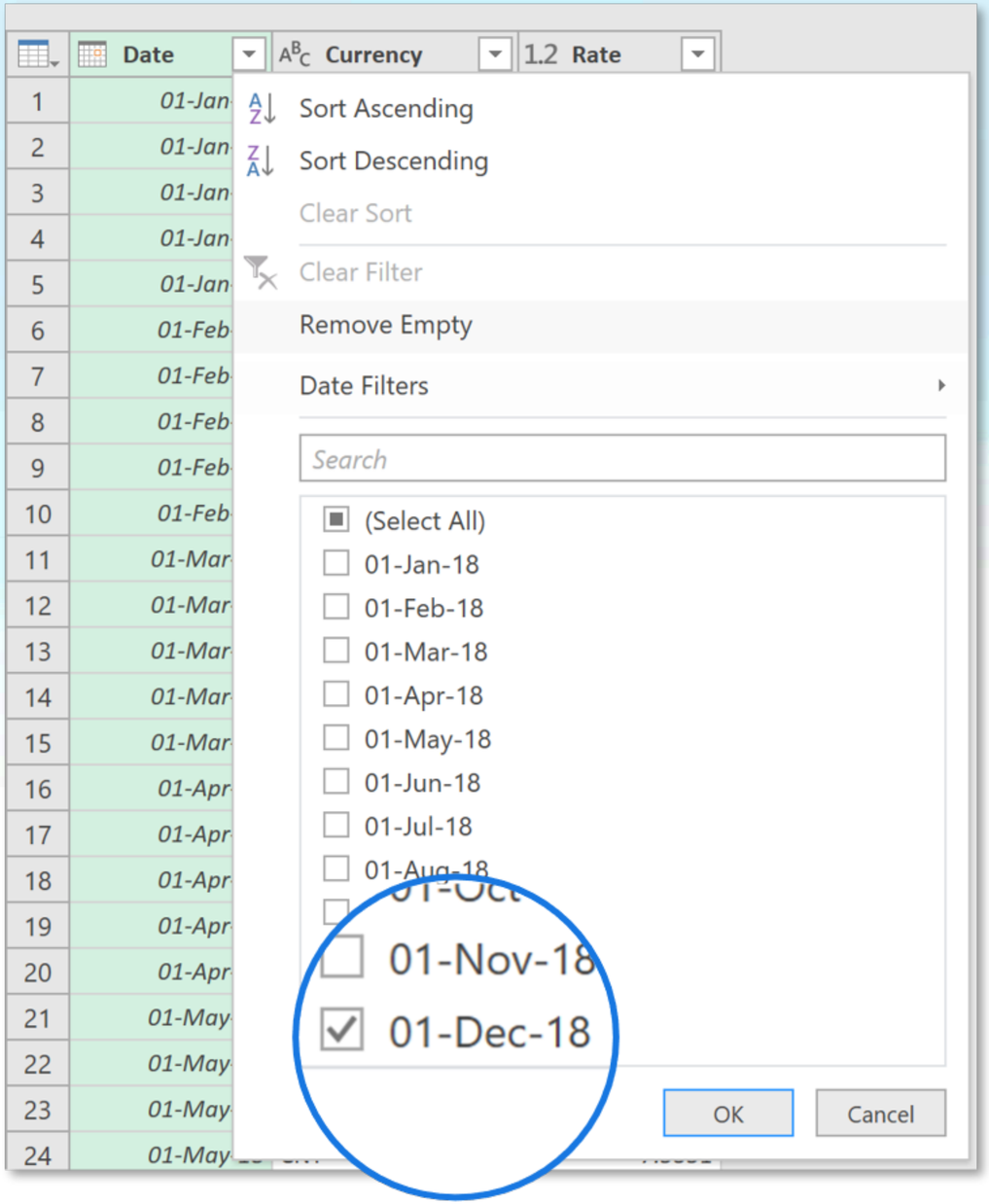

Choose Columns
Choose the columns to keep
Search Columns
(Select All Columns)
Exchange Rate
EUR
GBP
BRL
CNY
MAD
Column7
Column8
Column9
Column10



	Exchange Rate	EUR	GBP	BRL	CNY	MAD
1	01-Jan-18	1	0.883499	3.9[illegible]349	7.839975	10.79
2	01-Feb-18	1	0.88349	4.009741	7.808135	10.79
3	01-Mar-18	1	0.883217	4.039922	7.796012	10.79
4	01-Apr-18	1	0.873156	4.176149	7.733027	10.79
5	01-May-18	1	0.877863	4.288502	7.5351	10.79
6	01-Jun-18	1	0.878224	4.403002	7.53954	10.79
7	01-Jul-18	1	0.887289	4.470242	7.845807	10.79
8	01-Aug-18	1	0.896598	4.514944	7.8991	10.79
9	01-Sep-18	1	0.89377	4.775615	7.98517	10.79
10	01-Oct-18	1	0.88299	4.333778	7.952165	10.79
11	01-Nov-18	1	0.881157	4.287048	7.886536	10.79
12	01-Dec-18	1	0.897525	4.422776	7.839751	10.79

Any Column

	Exchange Rate	Attribute	Value
1	01-Jan-18	EUR	1
2	01-Jan-18	GBP	0.883499
3	01-Jan-18	BRL	3.914349
4	01-Jan-18	CNY	7.839975
5	01-Jan-18	MAD	10.79
6	01-Feb-18	EUR	1
7	01-Feb-18	GBP	0.88349
8	01-Feb-18	BRL	4.009741
9	01-Feb-18	CNY	7.808135


Date
Currency
1.2 Rate
Sort Ascending
Sort Descending
Clear Sort
Clear Filter
Remove Empty
Date Filters
Search
(Select All)
01-Jan-18
01-Feb-18
01-Mar-18
01-Apr-18
01-May-18
01-Jun-18
01-Jul-18
01-Aug-18
01-Nov-18
01-Dec-18
OK
Cancel