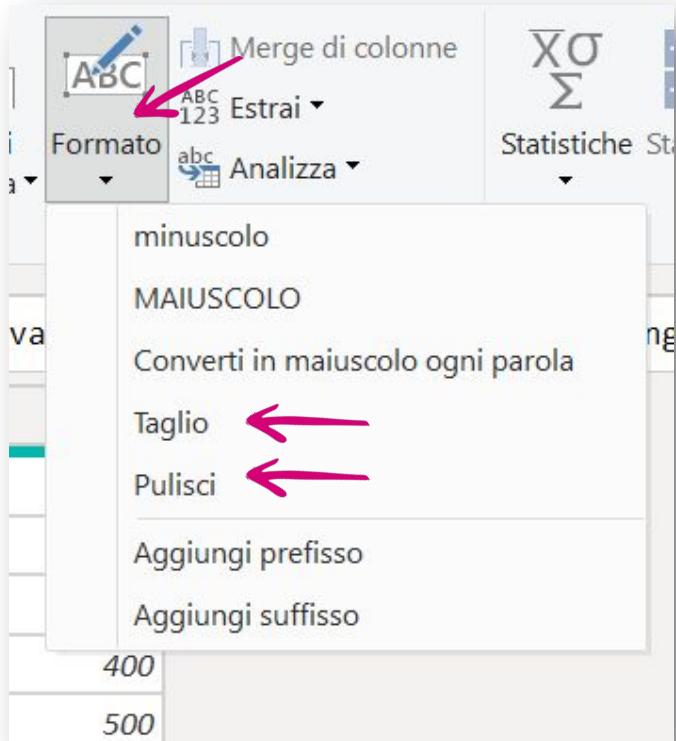




**Power Query : Aggiungere colonne
e trasformazioni sui dati**

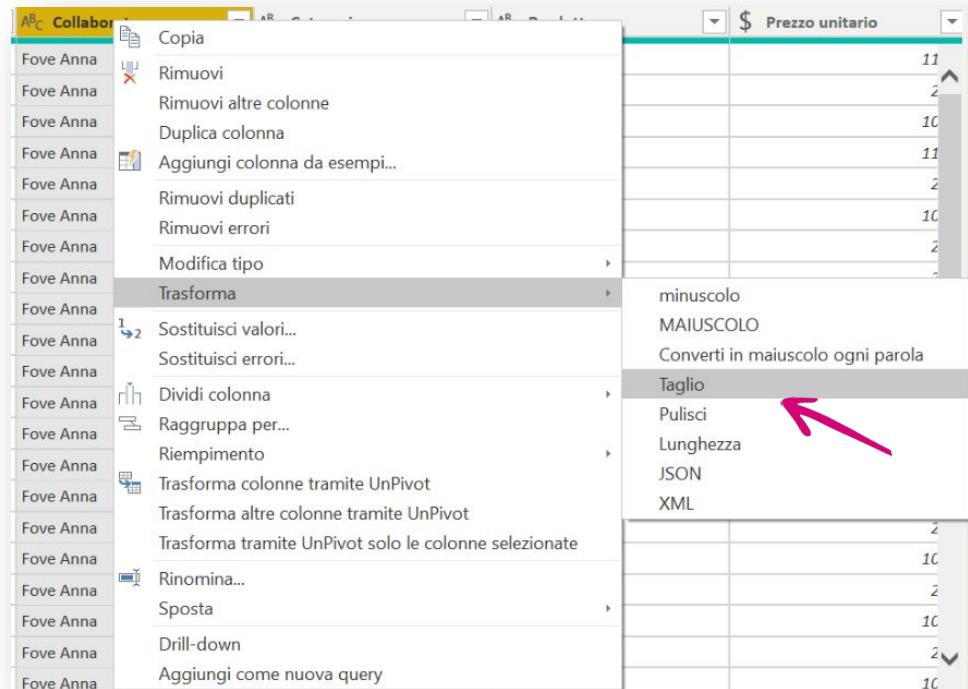
Power Query – Trasforma Formato



Nella scheda **Trasforma** si trova il menù **Formato** con i comandi utili a:

- convertire le lettere maiuscole in minuscole e viceversa
- i comandi **Pulisci** e **Taglio** per rimuovere caratteri non stampabili e i caratteri “spazio” inutili presenti nelle celle
- possibilità di inserire un prefisso o un suffisso alle celle in alternativa a concatena

Power Query – Taglio e pulisci

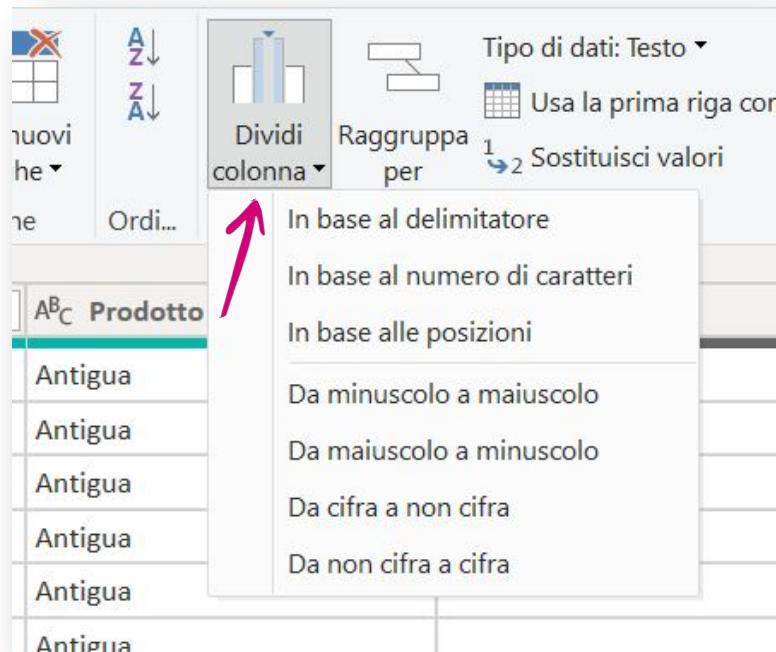


Spesso nelle colonne troviamo dati «errati» per colpa di caratteri SPAZIO prima o dopo il testo. Problema difficile da trovare visto che il carattere spazio non si vede. La parola «CIAO» non è uguale alla parola «CIAO » seguita da uno spazio.

Cliccando col tasto destro su una colonna di tipo testo, compare un menù dove troviamo la voce **TRASFORMA** → **TAGLIO**. Con questo comando siamo sicuri di eliminare tutti i caratteri SPAZIO inutili.

Il comando **PULISCI** elimina tutti i caratteri non stampabili

Power Query – Dividi colonna



Cliccando sul comando **Dividi Colonna** si aprirà un elenco con diverse opzioni. La scelta dipende da come sono scritti i dati: nel nostro esempio il Cognome e il Nome sono separati dal carattere SPAZIO quindi possiamo scegliere l'opzione **In base al delimitatore**.

A seconda dell'opzione scelta si aprirà una finestra per personalizzare ulteriormente l'operazione: nel nostro caso quale carattere separa le parole.

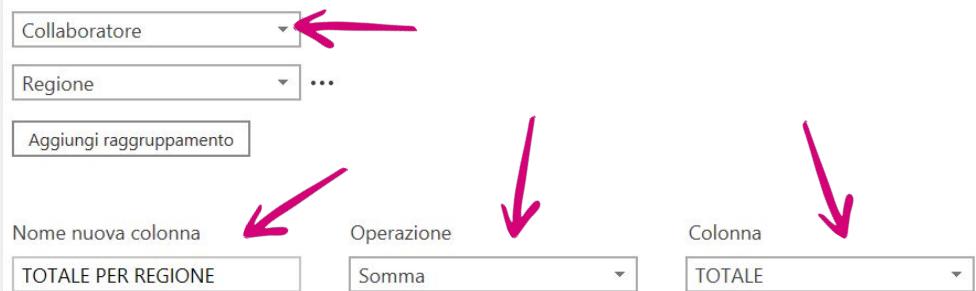
Verranno quindi create due colonne, da rinominare, contenenti solo il Cognome e solo il Nome.

Power Query – Raggruppa per

Raggruppa per

Specifica le colonne in base alle quali eseguire il raggruppamento e uno o più output.

Di base Avanzate



Collaboratore

Regione

Aggiungi raggruppamento

Nome nuova colonna: TOTALE PER REGIONE

Operazione: Somma

Colonna: TOTALE

Nell'esempio abbiamo ottenuto il totale vendite per singolo venditore per ogni regione e abbiamo chiamato la colonna TOTALE PER REGIONE

Nella finestra Raggruppa per:

- selezionare il campo o i campi (**Avanzato**) da raggruppare
- Specificare un nome per la colonna riassuntiva
- Definire l'operazione da eseguire con il raggruppamento
- Specificare su quale colonna effettuare l'operazione

A ^B C Collaboratore	A ^B C Regione	1.2 tot regione
Fove Anna	Basilicata	8160
Fove Anna	Lombardia	90935
Fove Anna	Piemonte	61644
Fove Anna	Trentino Alto Adige	7680
Fove Anna	Calabria	259071

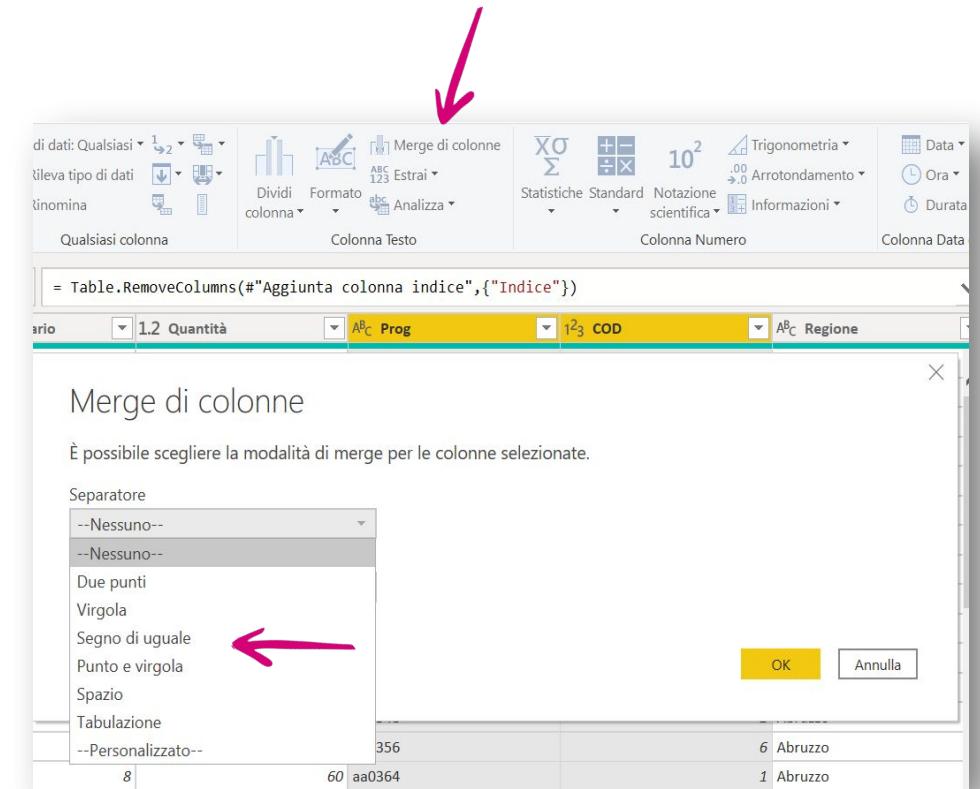
Power Query – Merge di colonne, Estrai

Tramite il comando **Merge di colonne** è possibile unire il contenuto di due o più colonne in una nuova colonna.

Selezionare le colonne da unire e cliccare sul comando **Merge di colonne** nella barra **Trasforma**.

Nella finestra che si apre scegliere se si vuole utilizzare un separatore e il nome della nuova colonna che stiamo creando.

Il comando **Estrai** presente nella stessa sezione permette di trasformare la colonna estraendo solo una parte della colonna selezionata. Ad esempio potremmo trasformare la colonna **Prog** estraendo solo gli ultimi 4 caratteri.



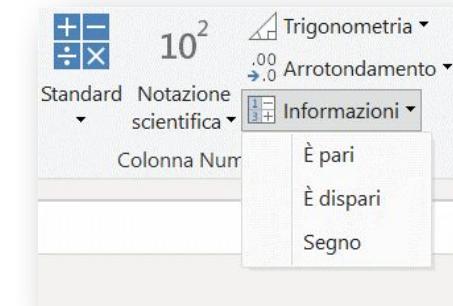
Power Query – Colonna numero

Nella sezione “Colonna Numero” ci viene data la possibilità di applicare delle formule (statistiche, matematiche, trigonometriche, ecc) sull’intero contenuto di una colonna (es. dividere la colonna Euro per 100).

Tramite il comando “Arrotondamento” arrotondiamo il contenuto della colonna per eccesso, difetto o decimali personalizzati.

Nel menù “Informazioni” troviamo alcuni comandi utili come sapere se il contenuto è pari o dispari o se negativo, ottenendo True o False come risultato.

Per non applicare la formula direttamente sulla colonna originale è consigliabile farne un duplicato

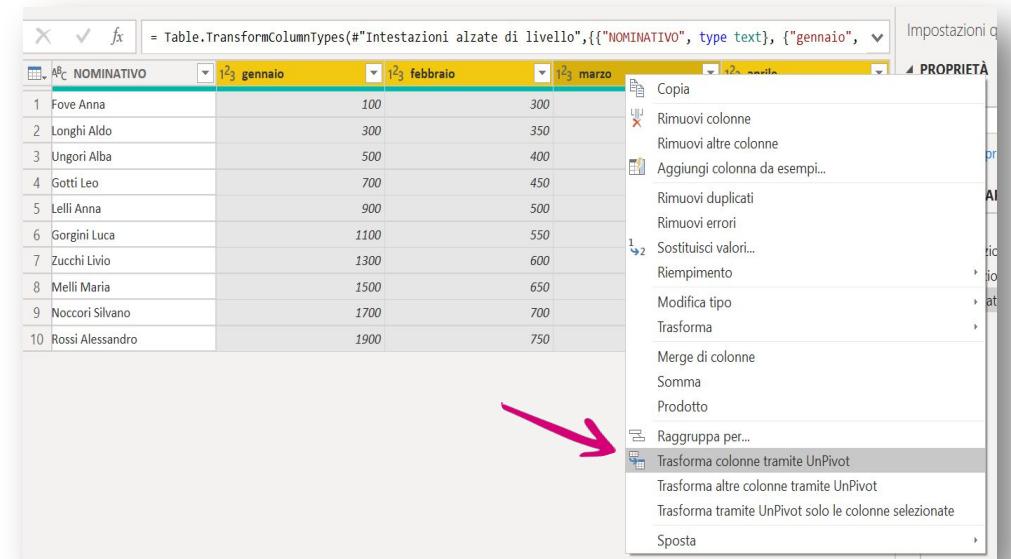


Power Query – UnPivot

Alcune tabelle sono scritte in origine in una forma non facilmente elaborabile in PBD. Immaginiamo una tabella dove sono presenti 12 colonne rappresentanti i 12 mesi dell'anno con i relativi dati numerici.

Dopo aver selezionato le colonne da elaborare (i mesi), tramite il comando **Trasforma colonne tramite UnPivot**, sarà possibile ottenere due colonne (da rinominare) contenenti il singolo mese ed il valore corrispondente.

	NOMINATIVO	Attributo	Valore
1	Fove Anna	gennaio	100
2	Fove Anna	febbraio	300
3	Fove Anna	marzo	500
4	Fove Anna	aprile	700
5	Fove Anna	maggio	900
6	Fove Anna	giugno	1100
7	Fove Anna	luglio	1300
8	Fove Anna	agosto	1500
9	Fove Anna	settembre	1700
10	Fove Anna	ottobre	1900

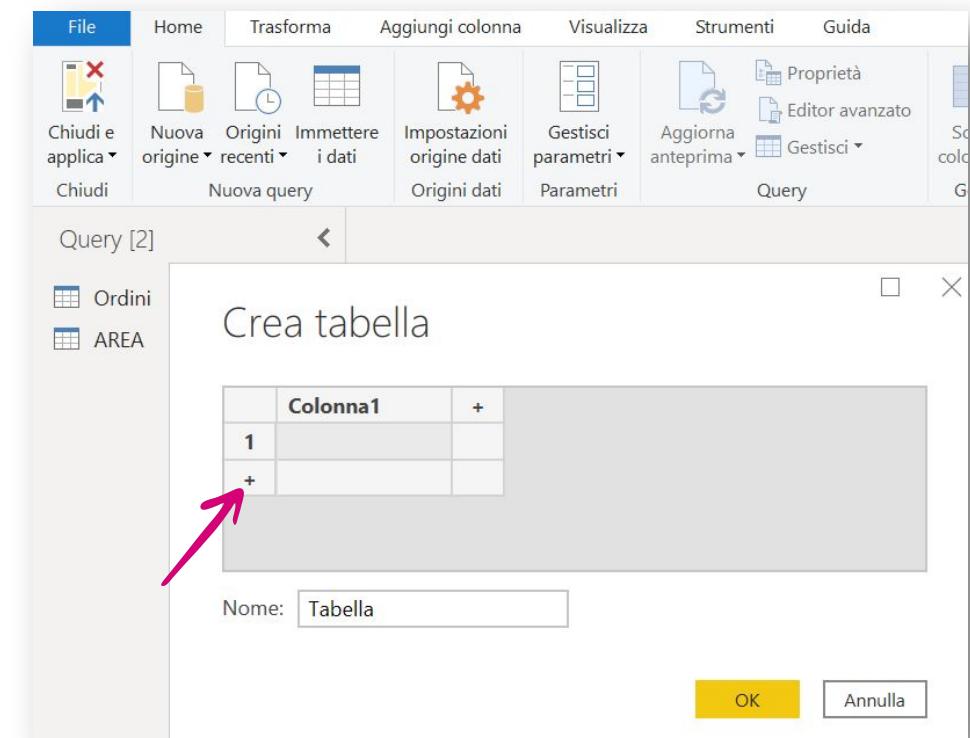


The screenshot shows the Power Query Editor interface. A table is displayed with columns: Nominativo, gennaio, febbraio, marzo, aprile, maggio, giugno, luglio, agosto, settembre, ottobre. The 'UnPivot' step is highlighted in the formula bar: `= Table.UnpivotOtherColumns(#"Modificato tipo", {"NOMINATIVO"}, "Attributo", "Valore")`. The 'Attributo' column is selected. A pink arrow points to the 'Attributo' column header. On the right, the 'Trasforma colonne tramite UnPivot' option is highlighted in the context menu.

Power Query – Nuova Query

Nella sezione Nuova query si trovano tre comandi:

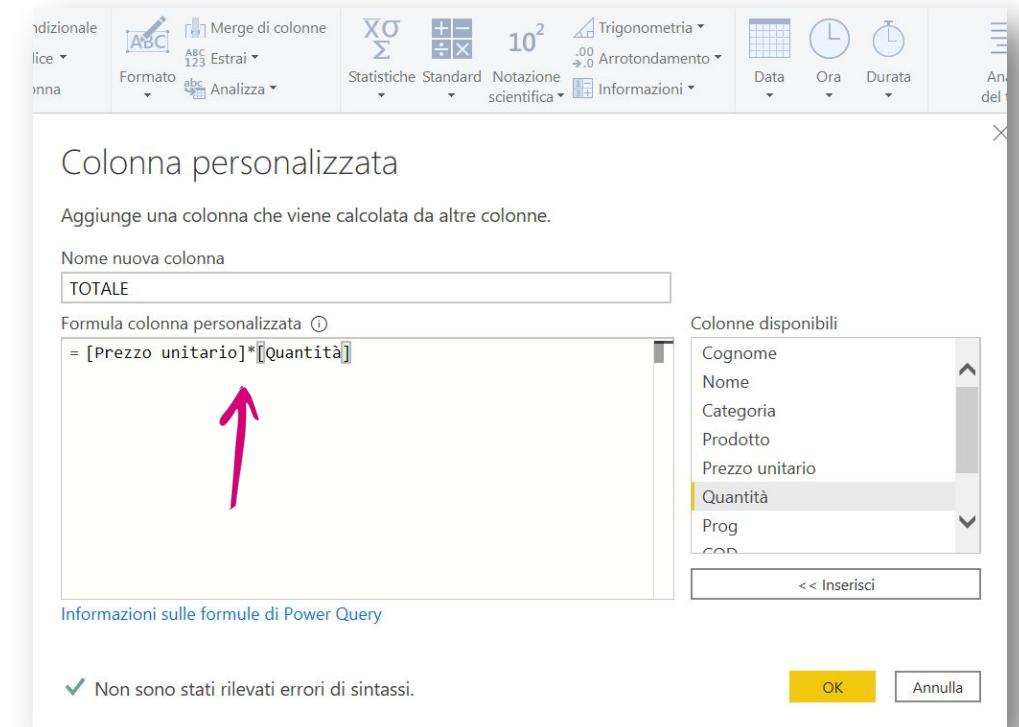
- **Nuova origine:** per poter collegare (importare) una nuova fonte di dati da trasformare
- **Origini recenti:** per poter richiamare fonti di dati utilizzate di recente
- **Immettere dati:** per poter creare una tabella partendo da zero; apparsa la finestra **Crea tabella** utilizzare i simboli **+** per creare nuove colonne e righe



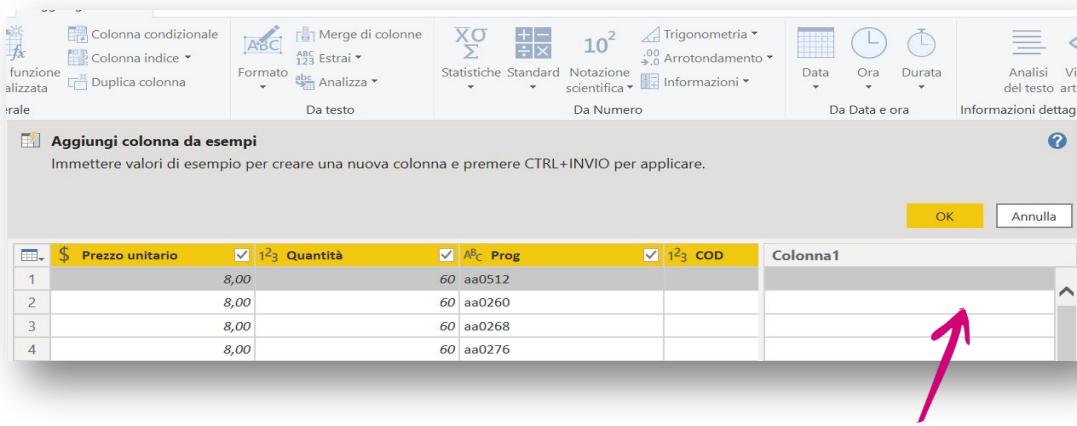
Power Query – Aggiungi colonna – Colonna personalizzata

Il comando **Colonna personalizzata** crea una nuova colonna in base ad una formula. Ad esempio potremmo creare la colonna **TOTALE** frutto della moltiplicazione di **Prezzo unitario** e **Quantità**.

Nella finestra **Colonna personalizzata** Inserire il nome della nuova colonna e la formula (nel nostro caso = **[Prezzo unitario]*[Quantità]**) utilizzando i campi elencati sulla destra della finestra.



Power Query – Aggiungi colonna – Colonna da esempi



The screenshot shows the Power Query ribbon with the 'Colonna da esempi' button highlighted by a red arrow. The ribbon tabs are: File, Home, Trasforma, Aggiungi colonna, Visualizza, Strumenti, and Guida. Below the ribbon, the Power Query editor displays a table named 'Ordini' with five columns: Column1, Column2, Column3, Column4, and Column5. The data in the table is as follows:

	Column2	Column3	Column4	Column5
1	null	null	null	null
2	boratore	Categoria	Prodotto	Area
3	Anna	Auto	Annet	aa0106
4	Anna	Auto	Annet	aa0107
5	Anna	Auto	Annet	aa0108
6	Anna	Auto	Annet	aa0109

All'interno della nostra tabella esiste una colonna contenente un codice area. Questo codice è composto da due lettere e quattro cifre. Se volessimo inserire un carattere «-» tra le lettere e le cifre potremmo avvalerci del comando **Colonna da esempi** che troviamo nella sezione **Aggiungi colonna**.

Basta selezionare la colonna da elaborare e cliccare sul comando **Dalla selezione**.

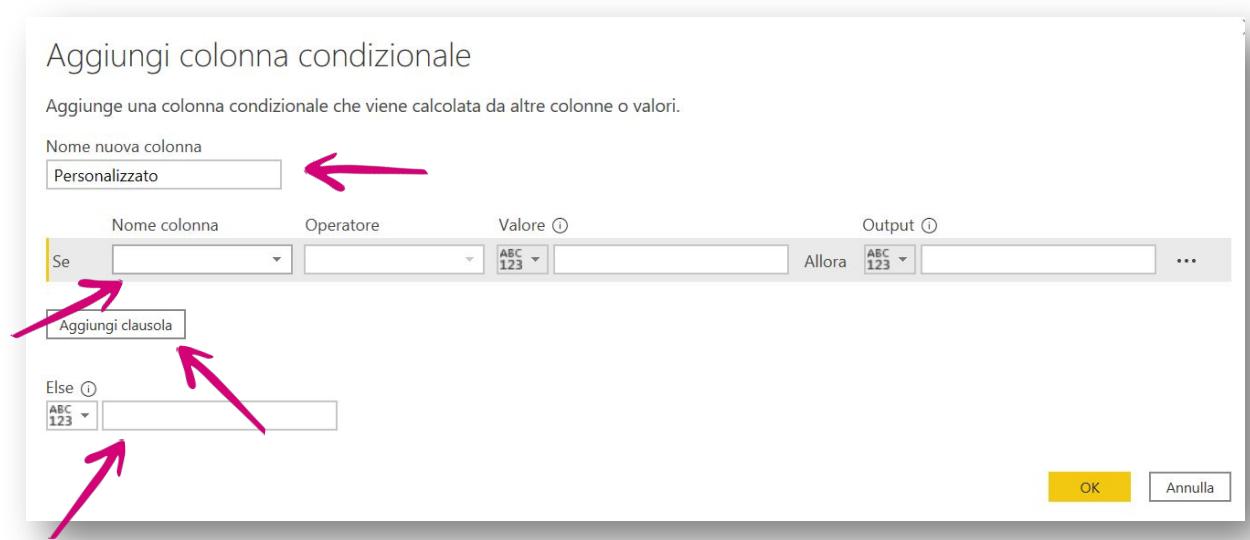
Verrà aggiunta una nuova colonna dove inserire alcuni esempi di dati desiderati (in alcuni casi ne basteranno due in altri di più). Nel nostro caso «aa-0106», «aa-0107», «aa-0108». Confermiamo l'operazione.

Power Query – Aggiungi colonna – Colonna condizionale

Il comando **Colonna condizionale** è in grado di creare una nuova colonna basandosi su delle condizioni da noi indicate.

Immaginiamo di voler creare una colonna che riporti la lettere A per la Categoria Auto e M per la categoria Moto:

1. Inserire il nome della nuova colonna (es.: CAT)
2. Nella riga **Se** indicare la colonna da utilizzare (**Categoria**), l'operatore (**Uguale a**), il valore (Auto) e l'**Output** desiderato (A);
3. Nella casella **Else** indicare l'output nel caso in cui non si verificasse la condizione definita in **Se** (M).



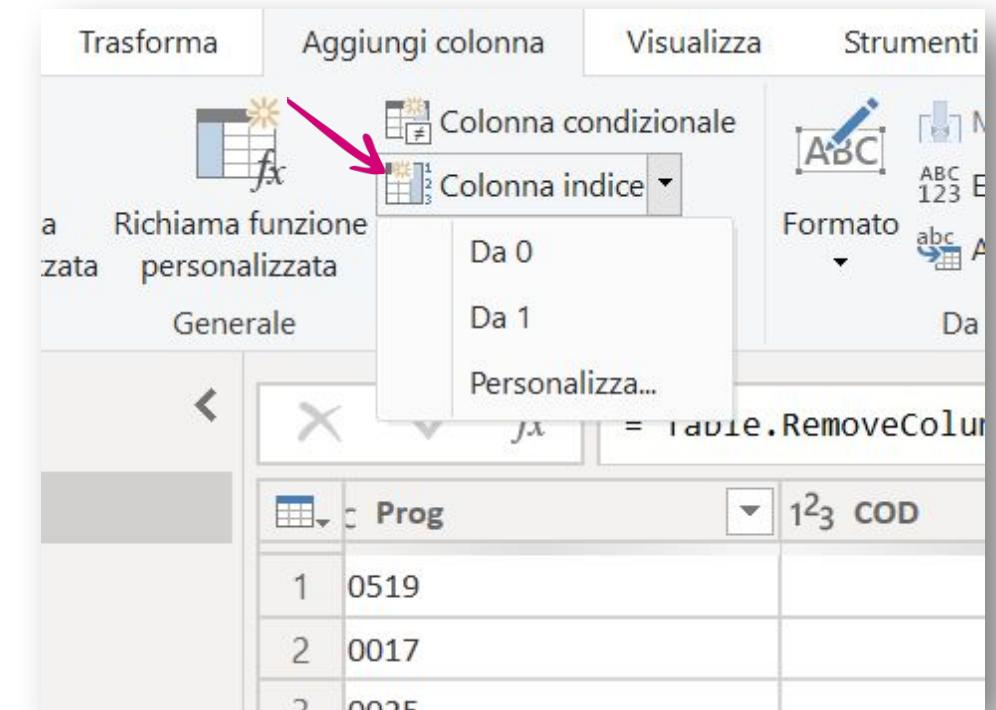
*Per poter aggiungere più condizioni cliccare sulla voce **Aggiungi clausola***

Power Query – Aggiungi colonna – Colonna indice e Duplica colonna

Per creare una colonna indicizzata (incremento numerico) cliccare su **Colonna indice** e scegliere il tipo:

- **Da 0:** l'indice con progressione 1 partirà da 0
- **Da 1:** l'indice con progressione 1 partirà da 1
- **Personalizza:** indicare inizio e incremento

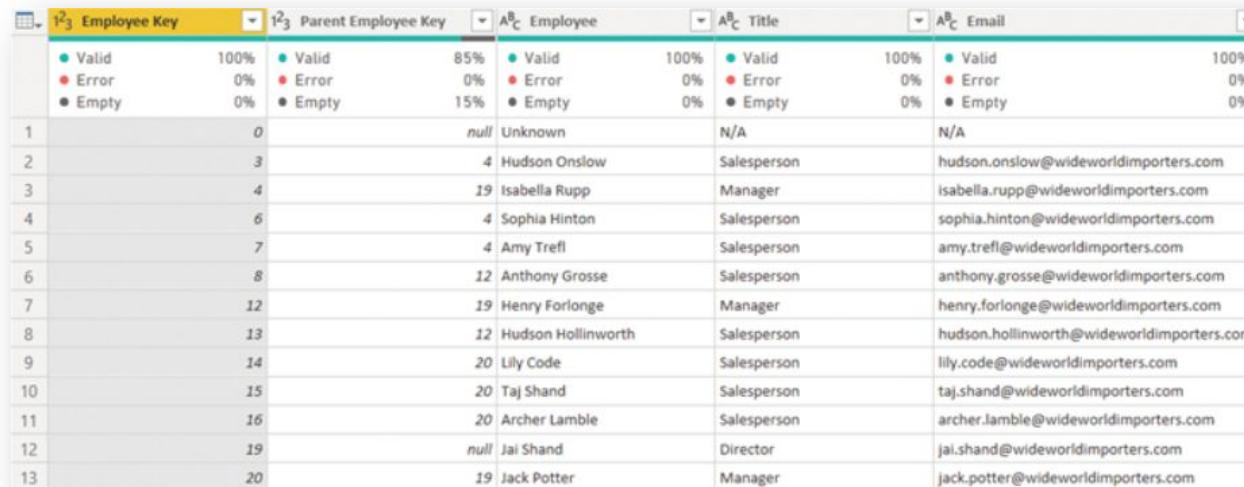
Cliccando su **Duplica colonna** otterremo una copia della colonna selezionata, da rinominare. Utile per applicare formule senza modificare la colonna originale.



Power Query – Trovare anomalie nei dati

Quando si guarda una tabella nell'editor di Power Query, è possibile verificare la qualità di ogni colonna. Per impostazione predefinita, ogni colonna avrà una barra colorata sotto l'intestazione, che mostra quanti valori validi, errore e vuoti ci sono in ogni colonna.

Per visualizzare le percentuali esatte, selezionare **Visualizza > Anteprima dati > Qualità colonna**.



The screenshot shows the Power Query Editor interface with the 'Employee' table selected. The columns are: Employee Key, Parent Employee Key, Employee, Title, and Email. Each column has a header bar with three colored dots indicating data quality: green for Valid, red for Error, and black for Empty. Below the table, a preview pane shows the first 13 rows of data, including names like Hudson Onslow, Isabella Rupp, and Jack Potter, along with their titles and email addresses.

Employee Key	Parent Employee Key	Employee	Title	Email
1	0	null/Unknown	N/A	N/A
2	3	4 Hudson Onslow	Salesperson	hudson.onslow@wideworldimporters.com
3	4	19 Isabella Rupp	Manager	isabella.rupp@wideworldimporters.com
4	6	4 Sophia Hinton	Salesperson	sophia.hinton@wideworldimporters.com
5	7	4 Amy Trefl	Salesperson	amy.trefl@wideworldimporters.com
6	8	12 Anthony Grosse	Salesperson	anthony.grosse@wideworldimporters.com
7	12	19 Henry Forlonge	Manager	henry.forlonge@wideworldimporters.com
8	13	12 Hudson Hollinworth	Salesperson	hudson.hollinworth@wideworldimporters.com
9	14	20 Lily Code	Salesperson	lily.code@wideworldimporters.com
10	15	20 Taj Shand	Salesperson	taj.shand@wideworldimporters.com
11	16	20 Archer Lamble	Salesperson	archer.lamble@wideworldimporters.com
12	19	null Jai Shand	Director	jai.shand@wideworldimporters.com
13	20	19 Jack Potter	Manager	jack.potter@wideworldimporters.com

Power Query – Trovare anomalie nei dati

Si noti come la colonna Parent Employee Key abbia l'85% di valori validi e il 15% di valori vuoti. Se non ci aspettavamo valori null nella colonna, questo ci porterebbe a approfondire il problema.

Employee Key	Parent Employee Key	Employee	Title	Email	
Valid	100%	Valid	100%	Valid	100%
Error	0%	Error	0%	Error	0%
Empty	0%	Empty	0%	Empty	0%
1	0	null	Unknown	N/A	
2	3	4	Hudson Onslow	Salesperson	
3	4	19	Isabella Rupp	Manager	
4	6	4	Sophia Hinton	Salesperson	
5	7	4	Amy Trefl	Salesperson	
6	8	12	Anthony Grosse	Salesperson	
7	12	19	Henry Forlonge	Manager	
8	13	12	Hudson Hollinworth	Salesperson	
9	14	20	Lily Code	Salesperson	
10	15	20	Taj Shand	Salesperson	
11	16	20	Archer Lamble	Salesperson	
12	19	null	Jai Shand	Director	
13	20	19	Jack Potter	Manager	

Power Query – Trovare anomalie nei dati

N.B.

Per lo strumento qualità della colonna, i valori vuoti sono valori null o lasciati senza caratteri.

Per impostazione predefinita, la profilatura delle colonne si basa sulle prime 1.000 righe di una query. È possibile utilizzare l'intero set di dati per la profilatura delle colonne apportando una modifica nell'angolo in basso a sinistra dell'editor di Power Query.

Teacher: Please note this down

Me:



made with mematic

Power Query – Distribuzione dei dati

Oltre alla qualità della colonna, molto utile esaminare la distribuzione dei valori nel campo.

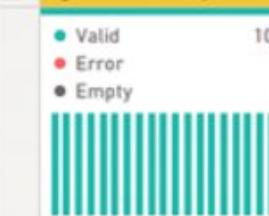
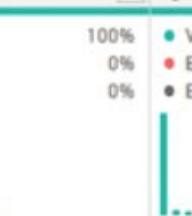
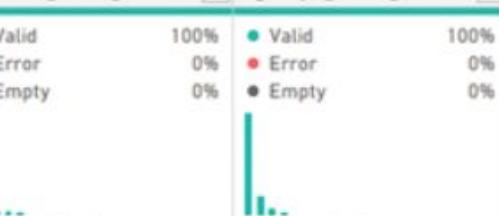
Per abilitare questa funzionalità, selezionare **Visualizza > Anteprima dati > Distribuzione colonne**.

In questo modo si possono vedere anche quanti valori distinti e univoci ha ciascuna colonna, nonché la distribuzione dei valori di colonna sotto forma di istogramma sotto ogni intestazione di colonna.

Power Query – Distribuzione dei dati

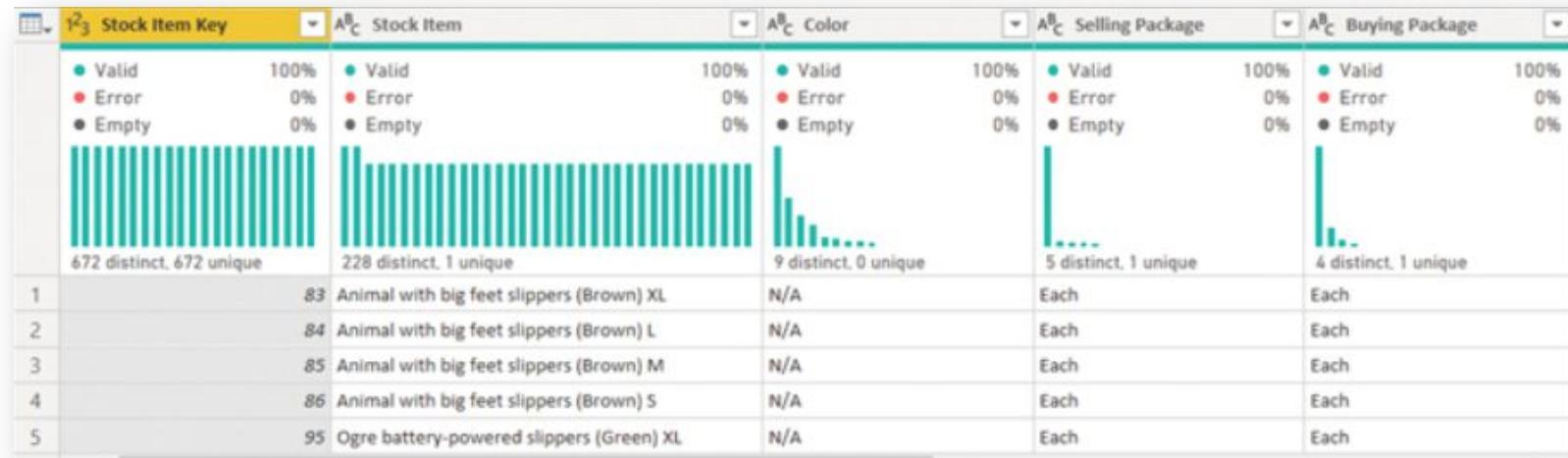
Il numero distinto si riferisce al numero di valori diversi presenti in una colonna una volta esclusi i duplicati. Il numero univoco mostra quanti valori si verificano esattamente una volta. Il numero di valori distinti e univoci sarà lo stesso solo se tutti i valori sono univoci.

Gli histogrammi mostrano la forma dei dati: possiamo vedere se la distribuzione dei valori è uniforme o se alcuni valori appaiono più frequentemente di altri.

	T3 Stock Item Key	A8 Stock Item	A8 Color	A8 Selling Package	A8 Buying Package
	 672 distinct, 672 unique	 228 distinct, 1 unique	 9 distinct, 0 unique	 5 distinct, 1 unique	 4 distinct, 1 unique
1		83 Animal with big feet slippers (Brown) XL	N/A	Each	Each
2		84 Animal with big feet slippers (Brown) L	N/A	Each	Each
3		85 Animal with big feet slippers (Brown) M	N/A	Each	Each
4		86 Animal with big feet slippers (Brown) S	N/A	Each	Each
5		95 Ogre battery-powered slippers (Green) XL	N/A	Each	Each

Power Query – Distribuzione dei dati

Ad esempio, stock item ha una distribuzione quasi uniforme, mentre Color e Buying Package sono piccate in alcuni valori.

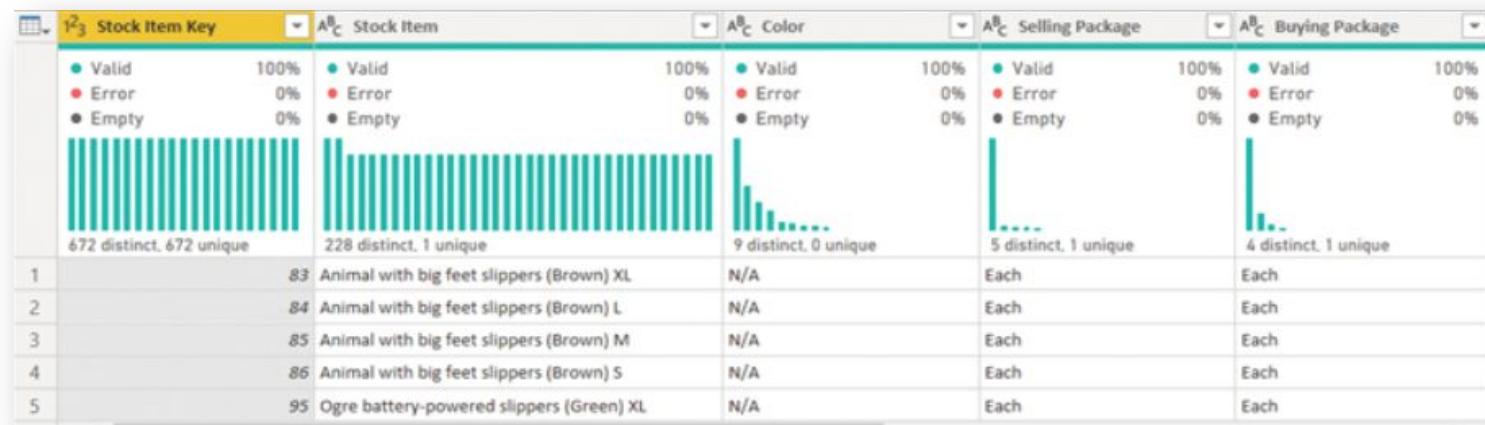


Power Query – Distribuzione dei dati

Per copiare i dati dietro un istogramma, fare clic con il pulsante destro del mouse su un istogramma e selezionare Copia distribuzione valori. Questo fornirà un elenco di valori distinti e il numero di volte in cui appaiono nella colonna.

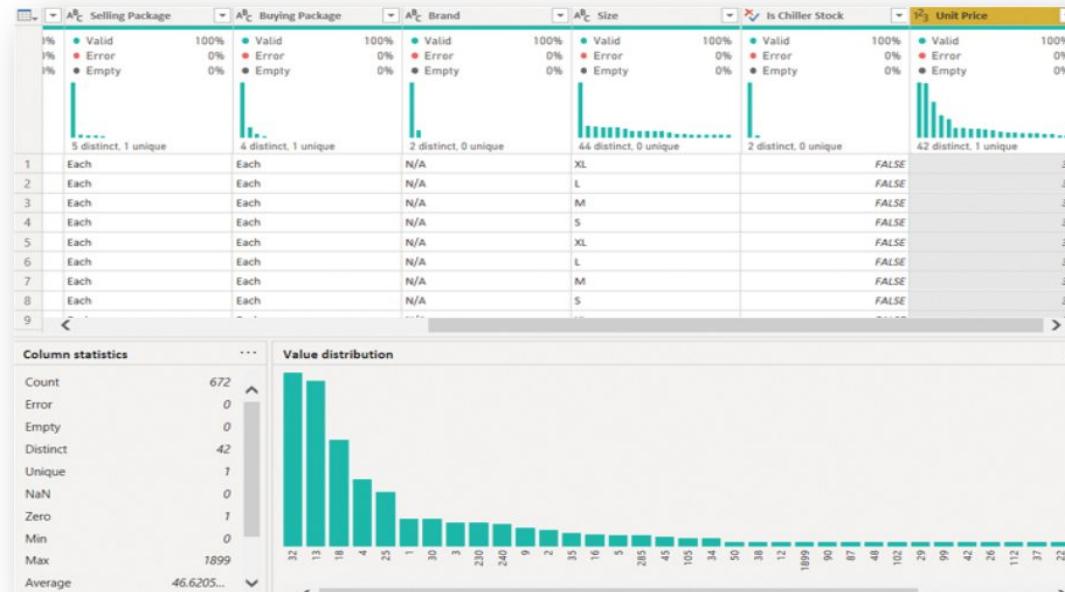
N.B.

La distribuzione dei valori si basa solo sulla query corrente: la distribuzione dei valori potrebbe essere diversa se si sceglie di profilare l'intero set di dati anziché le prime 1.000 righe.



Power Query – Profilo colonna

Per abilitare alcune statistiche di base, selezionare **Visualizza > Anteprima dati > Profilo colonna**. Dopo aver abilitato questa funzione, selezionare un'intestazione di colonna per visualizzarne il profilo. Ora si possono vedere due nuove aree nella parte inferiore dell'editor:



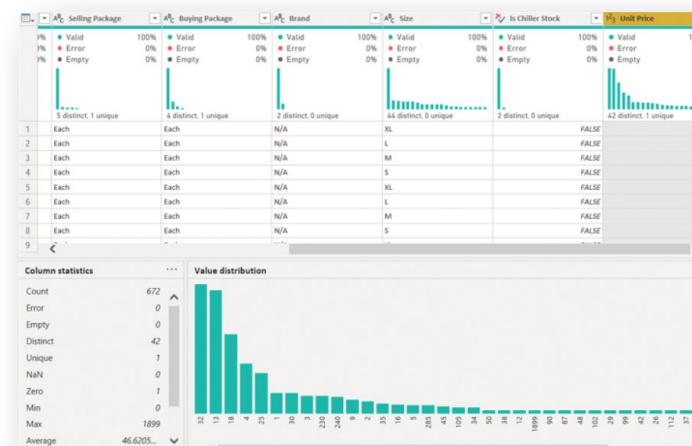
Power Query – Distribuzione dei dati

Statistiche colonna

Oltre al conteggio totale dei valori e al numero di errori, valori vuoti, distinti e univoci, si possono visualizzare minimo, massimo, media e il numero di valori pari a zero, dispari e pari, ecc.

A seconda del tipo di colonna si vedranno statistiche differenti: ad esempio, per le colonne di testo si vedranno il numero di stringhe vuote oltre al numero di valori vuoti.

È possibile copiare le statistiche selezionando i puntini di sospensione accanto a Statistiche colonna e selezionando Copia.





GRAZIE