

Capitolo 35: Trasformazione dei dati con Power Query

1. Trasformazioni comuni dei dati in Power Query

a. Eseguire attività comuni di trasformazione

Power Query offre una vasta gamma di strumenti per trasformare i dati in modo efficiente. Le operazioni possono essere eseguite tramite l'interfaccia grafica o scrivendo funzioni M personalizzate.

b. Rimuovere record duplicati

- Home → Rimuovi righe → Rimuovi duplicati
- Funzione M: `Table.Distinct(tabella)`

c. Riempire campi vuoti

- Trasforma → Riempimento → Verso l'alto / Verso il basso
- Funzione M: `Table.FillDown(tabella, {"Colonna"})`

d. Riempire stringhe vuote

- Aggiungi colonna personalizzata: `if [Colonna] = "" then "Default" else [Colonna]`

e. Concatenare colonne

- Colonna personalizzata: `[Nome] & " " & [Cognome]`
- Funzione M: `Text.Combine({[Col1], [Col2]}, " - ")`

f. Modificare il maiuscolo/minuscolo

- Trasforma → Formato → Maiuscolo / Minuscolo / Iniziali maiuscole
- Funzioni M: `Text.Upper()`, `Text.Lower()`, `Text.Proper()`

g. Trovare e sostituire testo specifico

- Trasforma → Sostituisci valori
- Funzione M: `Text.Replace([Colonna], "vecchio", "nuovo")`

h. Tagliare e pulire il testo

- Trasforma → Formato → Rimuovi spazi / Pulizia
- Funzioni M: `Text.Trim()`, `Text.Clean()`

i. Estrarre valori da sinistra, destra, centro

- Trasforma → Estrai → Inizio, Fine, Intervallo
- Funzioni M: `Text.Start()`, `Text.End()`, `Text.Middle()`

j. Estrarre primi e ultimi caratteri

- `Text.Start([Colonna], 3)` — primi 3 caratteri
- `Text.End([Colonna], 3)` — ultimi 3 caratteri

k. Estrarre caratteri centrali

- `Text.Middle([Colonna], 2, 4)` — da posizione 2, 4 caratteri

l. Dividere colonne usando marcatori di caratteri

- Trasforma → Dividi colonna → Per delimitatore (es. spazio, virgola)
- Funzione M: `Text.Split([Colonna], ",")`

m. Annullare la pivot (Unpivot)

- Trasforma → Annulla pivot colonne selezionate
- Funzione M: `Table.UnpivotColumns()`

n. Annullare le altre colonne

- Trasforma → Annulla pivot colonne non selezionate
- Funzione M: ``Table.UnpivotOtherColumns()``

o. Creare pivot colonne

- Trasforma → Crea pivot colonna
- Funzione M: ``Table.Pivot()``

Esempi pratici

Rimuovi duplicati da una tabella clienti

Usa 'Rimuovi duplicati' sulla colonna CodiceCliente.

Riempire date mancanti

Applica 'Riempimento verso il basso' sulla colonna Data per completare i valori vuoti.

Unire nome e cognome in una nuova colonna

Colonna personalizzata: ``[Nome] & " " & [Cognome]``.

Pulire stringhe di testo

Applica ``Text.Trim()`` e ``Text.Clean()`` per eliminare spazi e caratteri invisibili.

Annullare pivot su colonne mensili

Unpivot su Gennaio-Dicembre per ottenere una struttura dati normalizzata.

Esercizi

Importa un file CSV e rimuovi i record duplicati

Carica i dati e usa la funzione 'Rimuovi duplicati' su almeno due colonne.

Sostituisci valori nulli con un valore predefinito

Usa una colonna personalizzata per sostituire stringhe vuote con 'N/D'.

Dividi una colonna indirizzo in via e numero civico

Usa 'Dividi per delimitatore' scegliendo lo spazio o la virgola.

Estrai le prime tre lettere del nome

Crea una colonna con ``Text.Start([Nome], 3)``.

Crea una tabella pivot da dati non strutturati

Utilizza la trasformazione 'Crea pivot colonna' su una colonna categoria.

2. Creare colonne personalizzate in Power Query

a. Introduzione alle colonne personalizzate

Le colonne personalizzate in Power Query consentono di creare nuovi campi calcolati utilizzando il linguaggio M. Queste colonne non modificano i dati originali, ma aggiungono logica dinamica e trasformazioni.

b. Concatenazione con una colonna personalizzata

Per unire contenuti da più colonne:

- Aggiungi colonna → Colonna personalizzata → Formula: `[Nome] & " " & [Cognome]`
- Funzione M: `Text.Combine({[Campo1], [Campo2]}, " - ")`

La concatenazione è utile per codici, identificativi o etichette.

c. Comprendere la conversione dei tipi di dati

Power Query applica automaticamente conversioni di tipo. Puoi gestire manualmente:

- `Number.FromText([Col])` — converte testo in numero
- `Date.FromText([Col])` — converte stringa in data
- `Text.From([Col])` — converte qualsiasi tipo in testo

La conversione corretta è cruciale per evitare errori nelle formule e nelle trasformazioni.

d. Potenziare colonne personalizzate con funzioni

Puoi arricchire le colonne personalizzate con funzioni avanzate:

- `Text.Upper([Nome])` — tutto maiuscolo
- `Date.Year([Data])` — estrai anno da una data
- `Duration.Days([DataFine] - [DataInizio])` — giorni tra due date
- `if ... then ... else` — logica condizionale

Combinare funzioni consente analisi avanzate e trasformazioni automatiche.

e. Aggiungere logica condizionale a colonne personalizzate

Power Query supporta costrutti condizionali:

- Sintassi base: `if [Prezzo]>1000 then "Premium" else "Standard"`
- Nidificata: `if [Qty]>100 then "Grande" else if [Qty]>50 then "Media" else "Piccola"`

La logica condizionale è utile per classificare, validare, normalizzare valori o segmentare i dati.

Esempi pratici

Combina nome e cognome in una nuova colonna

Formula: `[Nome] & " " & [Cognome]`

Converti una colonna numerica in testo

Formula: `Text.From([Importo])`

Calcola i giorni tra due date

Formula: `Duration.Days([DataFine] - [DataInizio])`

Estrai il mese da una data

Formula: `Date.MonthName([Data])`

Classifica le vendite in base all'importo

Formula: `if [Totale]>1000 then "Alta" else "Bassa"`

Esercizi

Crea una colonna che concatena codice e descrizione prodotto

Formula: `[Codice] & " - " & [Descrizione]`

Converte tutte le descrizioni prodotto in maiuscolo

Formula: `Text.Upper([Descrizione])`

Crea una colonna 'Anno' a partire dalla colonna Data

Formula: `Date.Year([Data])`

Calcola la durata in giorni tra due colonne data

Formula: `Duration.Days([DataFine] - [DataInizio])`

Applica una logica IF per determinare se il cliente è attivo

Formula: `if [UltimoAcquisto] > Date.AddDays(DateTime.LocalNow(), -180) then "Attivo" else "Inattivo"`

3. Raggruppamento e aggregazione dei dati in Power Query

a. Raggruppare e aggregare dati

Il raggruppamento in Power Query consente di sintetizzare i dati sulla base di una o più colonne. È spesso utilizzato per generare riepiloghi statistici o contare eventi unici.

Operazioni disponibili:

- Somma, media, massimo, minimo
- Conteggio righe o valori distinti
- Tutti i record (per operazioni complesse)

Accesso: Trasforma → Raggruppa per

Funzione M:

```
`Table.Group(tabella, {"ColonnaChiave"}, {"NomeColonnaNuova", each List.Sum([ColAggregata]), tipo})`
```

b. Utilizzare tipi di dati personalizzati

Excel consente la creazione e gestione di tipi di dati personalizzati (ad es. entità geografiche, organizzazioni, strumenti finanziari), utili per rappresentare insiemi complessi di informazioni in una singola cella. In Power Query:

- I tipi personalizzati vengono rappresentati come record
- È possibile espandere o trasformare i campi contenuti
- Funzione utile: `Record.Field()` — accede a un campo specifico
- È possibile convertire colonne multiple in un solo campo strutturato (record)

Funzioni correlate:

****Table.Group**** — Raggruppa righe per chiave e applica aggregazione.

****List.Sum, List.Average, List.Max**** — Funzioni per aggregazioni su campi numerici.

****Record.ToTable, Record.Field**** — Per accedere e manipolare tipi di dati strutturati.

Esempi pratici

Raggruppa vendite per anno e somma gli importi

Colonna: Anno | Operazione: Somma su ImportoTotale

Conta il numero di ordini per cliente

Raggruppa per IDCliente e usa l'aggregazione 'Conteggio righe'.

Calcola la media dei voti per classe

Colonna: Classe | Aggregazione: Media su Voto

Estrai il campo 'Paese' da un tipo di dato personalizzato geografico

Espandi il tipo personalizzato o usa `Record.Field([Geo], "Paese")`.

Combina colonne indirizzo in un record strutturato

Usa 'Colonna personalizzata' per creare un record: `[Indirizzo = [Via], CAP = [CAP], Città = [Città]]`

Esercizi

Raggruppa vendite per prodotto e calcola il totale venduto

Raggruppa per NomeProdotto e somma la colonna Quantità.

Conta quanti clienti hanno fatto più di un ordine

Raggruppa per Cliente, conta gli ordini, poi filtra i valori > 1.

Crea un tipo di dato personalizzato per ogni riga

Raggruppa più colonne (es. Nome, Cognome, Telefono) in un record singolo.

Estrai solo il campo codice da un record prodotto

Usa `Record.Field([Prodotto], "Codice")` su una colonna di tipo record.

Calcola il massimo valore di ogni mese in una tabella pivotata

Unpivot delle colonne mese, poi raggruppa per mese e applica `List.Max()` sul valore.