



# Power BI

## Giorno 1

## Agenda

### Cosa vedremo

- Introduzione a Power BI
- Acquisizione, profilazione, preparazione e caricamento dei dati



## Agenda

### Obiettivi di apprendimento

- Descrivere Power BI e le sue parti fondamentali
- Acquisire e profilare i dati
- Scegliere la giusta modalità di storage dei dati



# Introduzione a Power BI

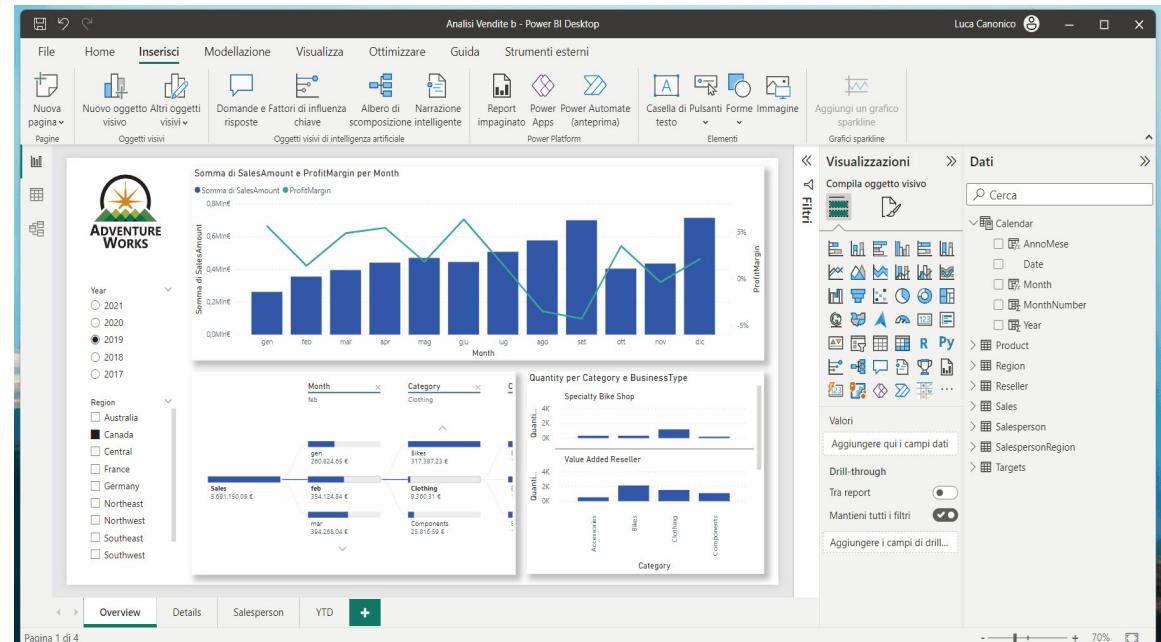
## Business Intelligence

È un insieme di processi, metodologie e software a supporto del processo decisionale aziendale.

Lo scopo principale di una soluzione di BI è la trasformazione dei dati disponibili in informazioni, e quindi in conoscenza a supporto di un *decision maker*.

Il dato diventa un'informazione quando suggerisce un'azione. Quali sono gli aspetti sfidanti?

- La raccolta e la 'ristrutturazione' del dato
- Fare storytelling



## Quali sono i task di un Data Analyst?

### Comprensione dello scenario

È la fase di sviluppo durante la quale si raccolgono i requisiti, si individuano le sorgenti dati e si esplorano i dati per acquisire familiarità con il contesto di analisi...

### Preparazione del dato

È un processo complesso che comprende l'analisi della qualità dei dati, la verifica dell'integrità dei dati, la correzione di errori, la ristrutturazione dei dati per ottimizzarne la leggibilità e le aggregazioni...

### Costruzione di un modello logico

Costruire un modello logico implica la definizione delle relazioni tra tabelle...

### Visualizzazione e Analisi del dato

Un report di BI deve consentire un'esplorazione efficace ed efficiente dello scenario di analisi...

### Gestione della soluzione

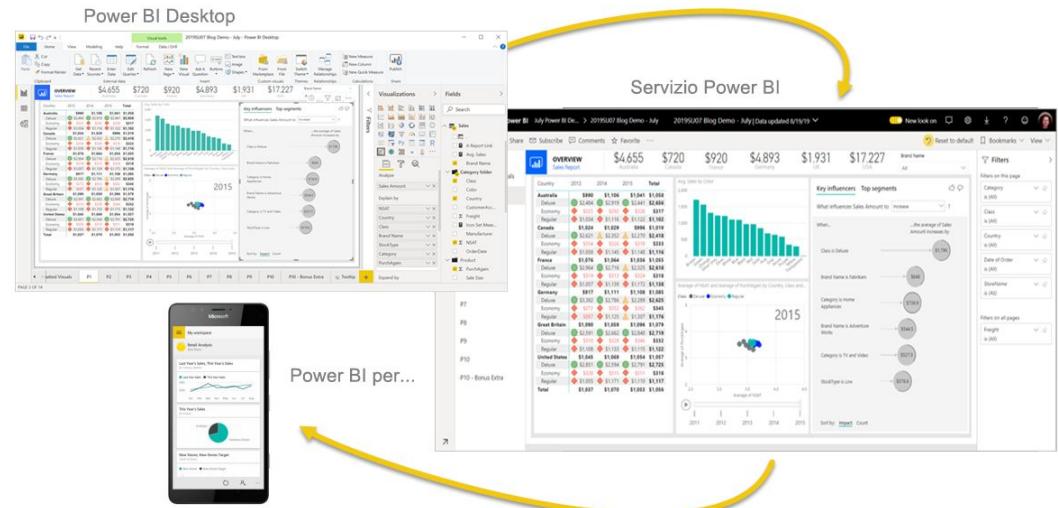
La soluzione di BI deve essere gestita in termini di aggiornamento dei dati sottostanti, distribuzione, accessi, ...

## Che cos'è Power BI?

Power BI è una raccolta di servizi software, app e connettori che operano congiuntamente per trasformare origini dati non correlate in informazioni dettagliate, coerenti, visivamente coinvolgenti e interattive.

Gli elementi base di Power BI includono:

- Un'applicazione desktop di Windows, **Power BI Desktop**
- Un servizio SaaS (Software as Service), **Power BI Service**
- Un'applicazione per device mobile, **Power BI Mobile**



<https://learn.microsoft.com/it-it/power-bi/fundamentals/power-bi-overview>

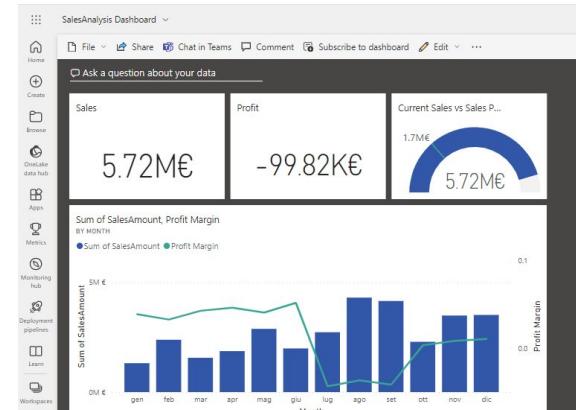
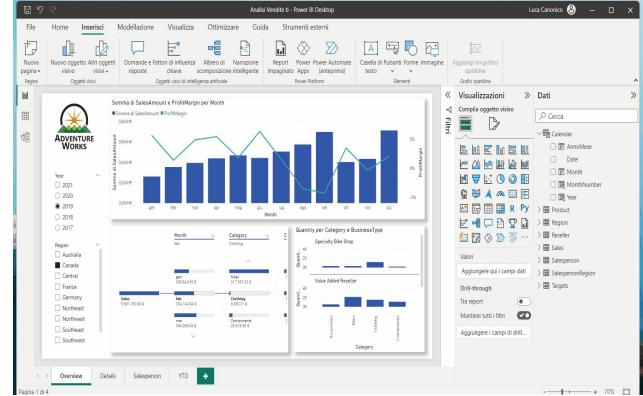
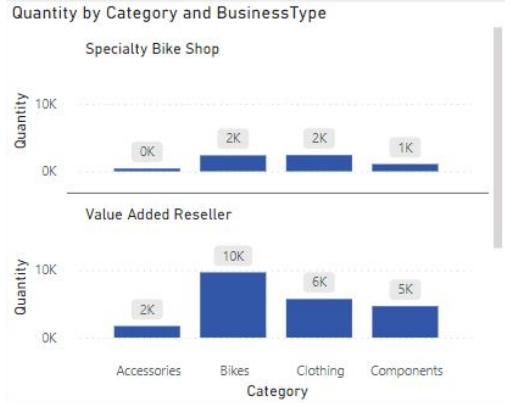
## ... un po' di nomenclatura!

**Che cos'è una visualization?**  
 È un singolo oggetto grafico  
 (line chart, pie chart, ...)

**Che cos'è un report?**  
 È un'insieme di visualization  
 organizzato in una o più pagine.  
 Il report è interattivo, dinamico,  
 filtrabile,...

**Che cos'è una dashboard?**  
 È un insieme selezionato di  
 visualization di più report riuniti  
 in un'unica pagina!

... definiamo **Tile** la singola  
 visualizzazione della dashboard.



La dashboard è statica  
 ed espone dati  
 tipicamente molto  
 aggregati.

## Qual è il flusso tipico di attività in PBI?

1. Acquisizione dati in Power BI  
Desktop
2. Creazione report in Power BI  
Desktop
3. Pubblicazione del report in Power BI  
Service
4. Condivisione di report e dashboard



I dati possono provenire da sorgenti eterogenee:

- File csv
- Cartelle di lavoro Excel
- Database on-premise
- Database in cloud
- Liste Sharepoint
- ...

Esplora il [report](#)

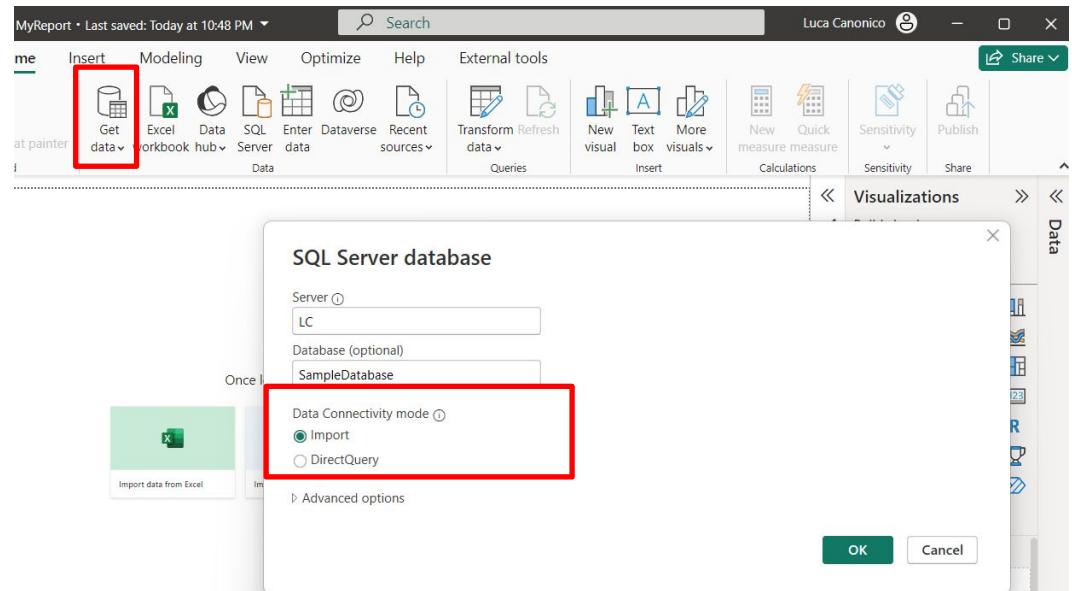
## Modalità di storage del dato

### Import

- Power BI importa effettivamente i dati creandone una copia locale.
- I dati sono memorizzati all'interno del file di Power BI (pbix).

### DirectQuery

- Power BI non crea nessuna copia locale dei dati.
- I dati sono interrogati direttamente nella loro sorgente.



## Modalità di storage del dato

### Import

- Esperienza utente ottimizzata
- Sono garantite tutte le funzionalità di Power BI
- Aggiornamento dati manuale o schedulato
- Limiti sul volume dati (dipende dalla licenza)

### DirectQuery

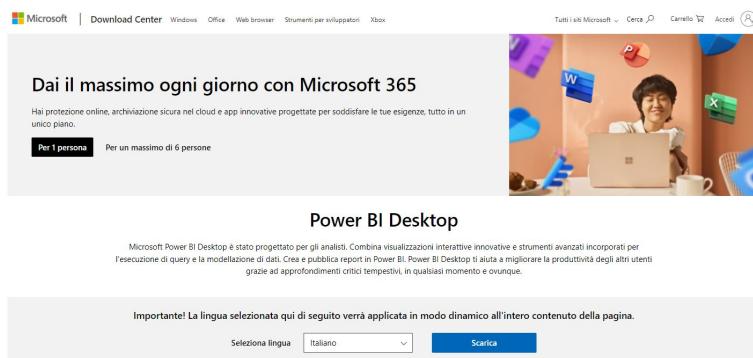
- L'esperienza utente può' essere degradata in termini di performance
- Non sono garantite tutte le funzionalità di Power BI (limiti nelle trasformazioni e funzioni DAX)
- Report real-time
- Utile per la gestione di grossi moli dati



## Installazione di Power BI Desktop

Per scaricare il file eseguibile di Power BI Desktop dall'Area download, accedi alla pagina dedicata

### Area download



The screenshot shows the Microsoft Download Center interface. At the top, there's a navigation bar with links for Windows, Office, Web browser, Strumenti per sviluppatori, and Xbox. Below the navigation bar, a banner for Microsoft 365 is displayed with the text "Dai il massimo ogni giorno con Microsoft 365". A section for "Power BI Desktop" follows, featuring a thumbnail image of a person working on a laptop with colorful data bubbles around them. Below the thumbnail, there's a brief description of what Power BI Desktop is and a note about its compatibility with Microsoft 365. At the bottom of the screenshot, there's a language selection dropdown set to "Italiano" and a "Scarica" (Download) button.

Specifica quindi il file di installazione a 32 bit o a 64 bit da scaricare.

### Scegli il download che desideri

		Dimensioni
<input type="checkbox"/>	Nome file	
<input type="checkbox"/>	PBIDesktopSetup.exe	430.6 MB
<input checked="" type="checkbox"/>	PBIDesktopSetup_x64.exe	479.0 MB

[Scarica](#) Dimensioni totali: 479.0 MB

Viene richiesto di eseguire il file di installazione dopo aver completato il download.

### Come ottenere Power BI Desktop

## Check Point!

**Quali sono i componenti principali di Power BI?**

- SQL Server, PBI desktop, App Mobile
- PBI desktop, Power BI Service
- PBI desktop, Power BI Service, Power BI Mobile

**Che cos'è un report di Power BI?**

- Un grafico che espone un'informazione aggregata
- Un canvas che espone informazioni sintetiche e rilevanti
- Una collezione di visualization disposta in una o più pagine

**Qual è il task più critico in fase di sviluppo?**

- È la fase di pulizia e trasformazione dei dati
- È la fase di modellazione dei dati

**Il dataset aziendale deve essere aggiornato ogni ora. È necessario sviluppare un report che esponga i dati con il minor ritardo possibile. Quale opzione di storage del dato soddisfa il requisito?**

- Import mode
- DirectQuery mode



# Power Query Editor

## Recupero dati in Power BI Desktop

Il primo task nel processo di sviluppo è l'identificazione e il recupero dati.

È possibile recuperare dati da sorgenti on-premise (database, cartelle di lavoro Excel, csv, txt, ...), da sorgenti cloud (cloud database, sharepoint, ...).

Dopo la pulizia e la 'ristrutturazione' dei dati è possibile creare report in Power BI.

Il tool utile alla profilazione, pulizia e preparazione dei dati è il Power Query Editor.

Il Power Query Editor è uno strumento fondamentale utilizzabile come strumento di ETL in diversi prodotto/servizi Microsoft (Power Platform, Azure Data Factory, Fabric, ...)



## Flat files, Excel

Le aziende molto spesso esportano e archiviano dati in file.

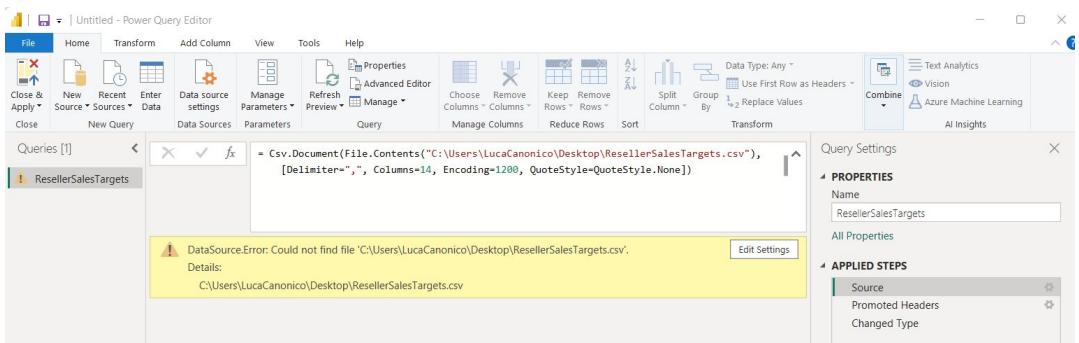
Tra i formati file più comuni ci sono i flat file (.csv e .txt).

Un'altra origine comune è rappresentata da cartelle di lavoro Excel (.xlsx).

Quali sono le criticità per questa tipologia di sorgente dati?

- 1) il percorso del file (file path)
- 2) modifiche strutturali al file sorgente (l'eliminazione o la ridenominazione di campi possono rompere il report)

ResellerSalesTargets.csv													
File Origin		Delimiter		Data Type Detection									
1200: Unicode		Comma		Based on first 200 rows									
Year	EmployeeID	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12
2017	61161660	-	-	-	-	-	200	400	600	400	800	400	800
2017	90836195	-	-	-	-	-	100	200	300	400	400	500	500
2017	112432117	-	-	-	-	-	500	1500	1000	1000	2200	1750	
2017	139397894	-	-	-	-	-	100	200	300	300	300	500	
2017	191644724	-	-	-	-	-	100	450	500	200	750	750	
2017	234474252	-	-	-	-	-	100	200	300	300	300	300	
2017	399771412	-	-	-	-	-	100	200	300	300	300	300	
2017	502097814	-	-	-	-	-	500	1500	1000	1000	2200	1750	
2017	615389812	-	-	-	-	-	100	100	100	200	200	200	
2017	716374314	-	-	-	-	-	100	200	300	100	300	500	
2017	841560125	-	-	-	-	-	100	500	400	400	1000	1000	



The screenshot shows the Microsoft Power Query Editor interface. At the top, there's a toolbar with various icons for file operations, source management, and data transformation. Below the toolbar, the 'Home' tab is selected. In the center, there's a query editor window with the following code:

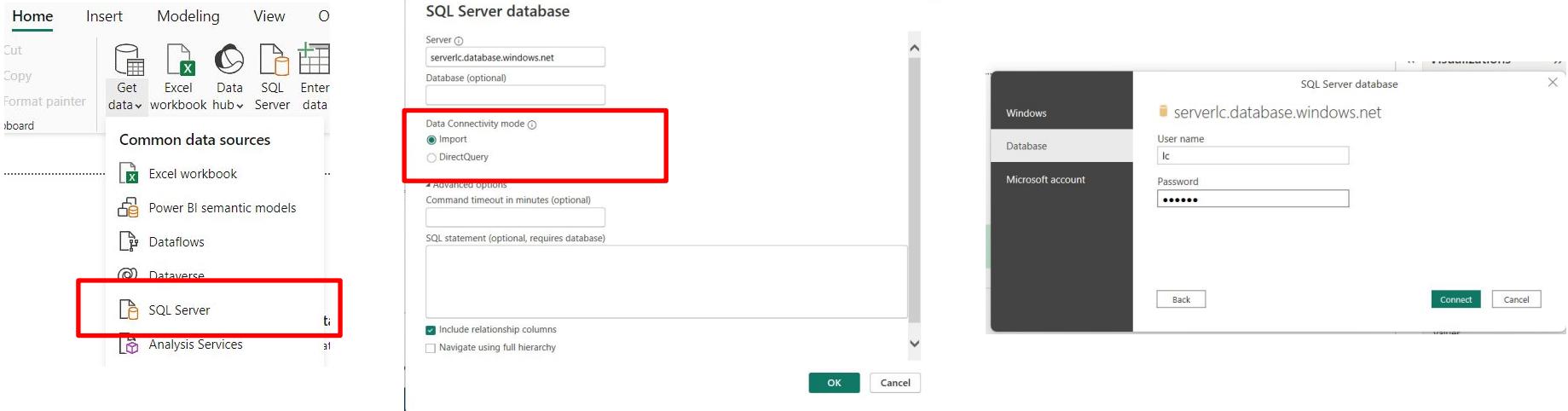
```
= Csv.Document(File.Contents("C:\Users\LucaCanonica\Desktop\ResellerSalesTargets.csv"),
    [Delimiter=",", Columns=14, Encoding=1200, QuoteStyle=QuoteStyle.None])
```

Below the code, an error message is displayed in a yellow box:

⚠️ Data Source Error: Could not find file 'C:\Users\LucaCanonica\Desktop\ResellerSalesTargets.csv'.  
Details:  
C:\Users\LucaCanonica\Desktop\ResellerSalesTargets.csv

To the right of the editor, there are two side panes: 'Query Settings' and 'Applied Steps'. The 'Query Settings' pane shows the connection details with the name 'ResellerSalesTargets'. The 'Applied Steps' pane lists the steps taken: 'Source', 'Promoted Headers', and 'Changed Type'.

## Database relazionali



Con Power BI Desktop è possibile connettersi a diversi database relazionali (e non) che siano in cloud oppure on-premises. La connessione ad un database relazionale richiede la scelta della modalità di storage dei dati e la selezione delle tabelle o entità che bisogna acquisire. In alternativa, è possibile scrivere una query custom (in quest'ultimo caso è obbligatoria l'indicazione del database).

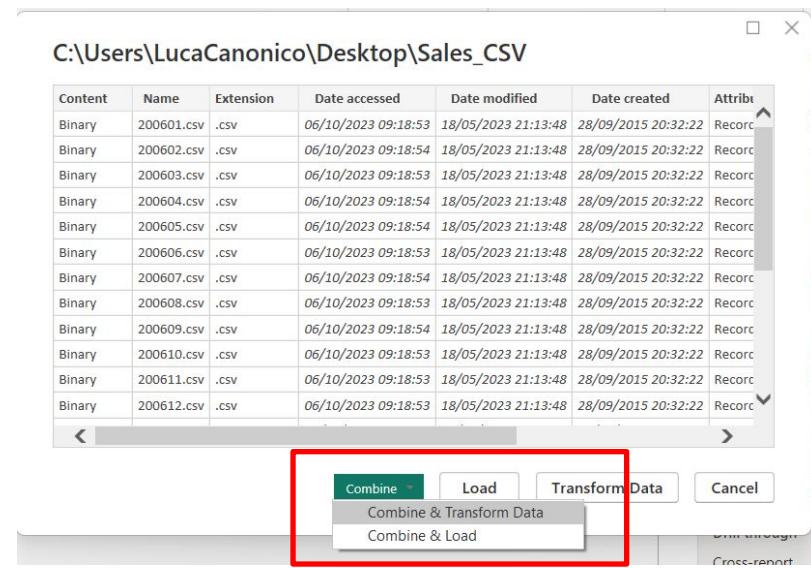
## Folder

Un'altra opzione utile è la connessione ad un Folder.

La connessione ad un folder consente di eseguire un append dei file contenuti al suo interno.



Scenari tipici: estrazioni periodiche di file csv con la medesima struttura in uno specifico folder...



Content	Name	Extension	Date accessed	Date modified	Date created	Attrib
Binary	200601.csv	.csv	06/10/2023 09:18:53	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200602.csv	.csv	06/10/2023 09:18:54	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200603.csv	.csv	06/10/2023 09:18:54	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200604.csv	.csv	06/10/2023 09:18:54	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200605.csv	.csv	06/10/2023 09:18:54	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200606.csv	.csv	06/10/2023 09:18:53	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200607.csv	.csv	06/10/2023 09:18:54	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200608.csv	.csv	06/10/2023 09:18:53	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200609.csv	.csv	06/10/2023 09:18:54	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200610.csv	.csv	06/10/2023 09:18:53	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200611.csv	.csv	06/10/2023 09:18:53	18/05/2023 21:13:48	28/09/2015 20:32:22	Record
Binary	200612.csv	.csv	06/10/2023 09:18:53	18/05/2023 21:13:48	28/09/2015 20:32:22	Record

## Profilazione campi

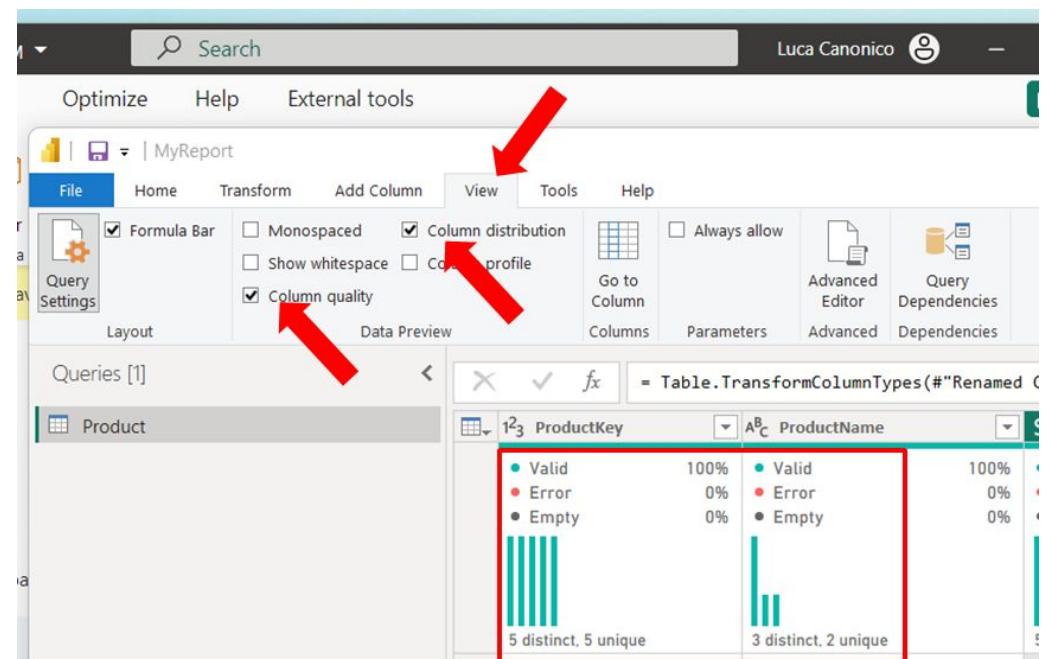
### Column quality

Espone gli indicatori: Valid, Error, Empty

### Column distribution

Espone la frequenza e la distribuzione dei valori per ciascuno dei campi della query (tabella) in esame. In particolare, fornisce il conteggio dei valori distinti (conteggio dei valori diversi) e dei valori univoci (conteggio dei valori che non hanno ridondanze).

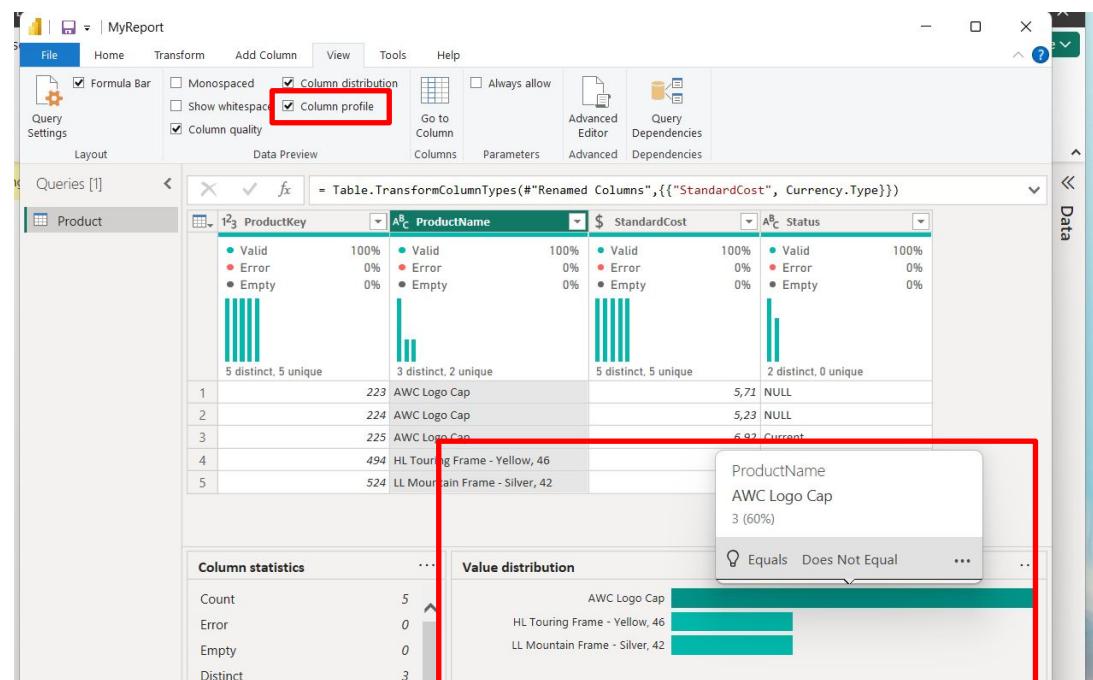
*Cosa si desume valutando la profilazione nell'esempio?*



## Profilazione campi

Dall'esempio si evince che:

- Ci sono 5 record
- Il campo *ProductKey* è una PK
- Il campo *ProductName* contiene 3 valori diversi (su un totale di 5 righe).
- Il campo *ProductName* contiene 2 valori univoci (su un totale di 5 righe). In altre parole, 2 valori su 5 non hanno duplicati (non sono ripetuti)
- Esiste 1 *ProductName* ridondante (conteggio distinct - conteggio unique =  $3 - 2 = 1$ )
- Il Column profile indica quante volte è ripetuto ciascun valore nella colonna selezionata



## Profilazione campi e sostituzione valori errati

MyReport

**File** Home Transform Add Column View Tools Help

Formula Bar Monospaced Always allow Go to Column Advanced Editor Query Dependencies

Layout Column profile Column quality

Query Settings

Queries [1]

Sheet1

A<sub>b</sub> City A<sub>b</sub> SalesRegion

	City	SalesRegion	
● Valid	100%	● Valid	100%
● Error	0%	● Error	0%
● Empty	0%	● Empty	0%

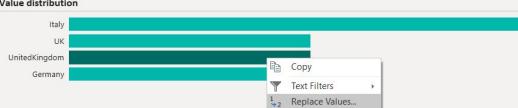
4 distinct, 3 unique

1 Bologna Italy  
2 London UK  
3 London UnitedKingdom  
4 Bergamo Italy  
5 Munich Germany

Column statistics

Count	5
Error	0
Empty	0
Distinct	4
Unique	3
Empty string	0
Min	Germany
Max	UnitedK...

Value distribution



Copy Text Filters Replace Values...

I dati devono essere consistenti!

Queries [1]

Region

A<sub>b</sub> City A<sub>b</sub> SalesRegion

	City	SalesRegion	
● Valid	100%	● Valid	100%
● Error	0%	● Error	0%
● Empty	0%	● Empty	0%

4 distinct, 3 unique

3 distinct, 1 unique

Replace Values

Replace one value with another in the selected columns.

Value To Find: UnitedKingdom

Replace With: UK

Advanced options

OK Cancel

Query Settings

PROPERTIES

APPLIED STEPS

- Source
- Navigation
- Changed Type
- Promoted Headers
- Changed Type1
- Replaced Value

Red arrow pointing to the "Replaced Value" step in the Applied Steps list.

## Rimuovere campi e filtrare righe

Last saved: Yesterday at 10:48 PM ▾ Search Luca Canonico

File Home Transform Add Column View Tools Help

Close & Apply New Recent Sources Data source settings Manage Parameters Refresh Preview Advanced Editor Manage Columns Reduce Rows Split Column Group By Replace Values Sort Transform

Employee

= AdventureWorksDW2020([Schema="dbo",

	EmployeeKey	ParentEmployeeKey	EmployeeNationalIDNumber
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13			

296 distinct, 296 unique 1 1 1  
48 distinct, 4 unique 7 18 18  
290 distinct, 285 unique 3 3 3

Query Settings

Properties Name Employee All Properties

Applied Steps Source Navigation

33 COLUMNS, 296 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED ON MERCOLEDÌ 20 SETTEMBRE 2023

L'obiettivo è minimizzare il volume del dato coerentemente ai requisiti di analisi!

Last saved: Yesterday at 10:48 PM ▾ Search Luca Canonico

Modeling View Optimize Help External tools

File Home Transform Add Column View Tools Help

Close & Apply New Recent Sources Data source settings Manage Parameters Refresh Preview Advanced Editor Manage Columns Reduce Rows Split Column Group By Replace Values Sort Transform

Employee

= Table.SelectColumns(#"Filtered Rows", {"EmployeeKey", "FirstName", "LastName"})

	EmployeeKey	FirstName	LastName
1	1	Stephen	Jiang
2	2	Brian	Welcker
3	3	Michael	Blythe
4	4	Linda	Mitchell
5	5	Jillian	Carson
6	6	Garrett	Vargas
7	7	Tsvi	Reiter
8	8	Pamela	Ansmann-Wolfe

18 distinct, 18 unique 18 distinct, 18 unique 18 distinct, 18 unique

Query Settings

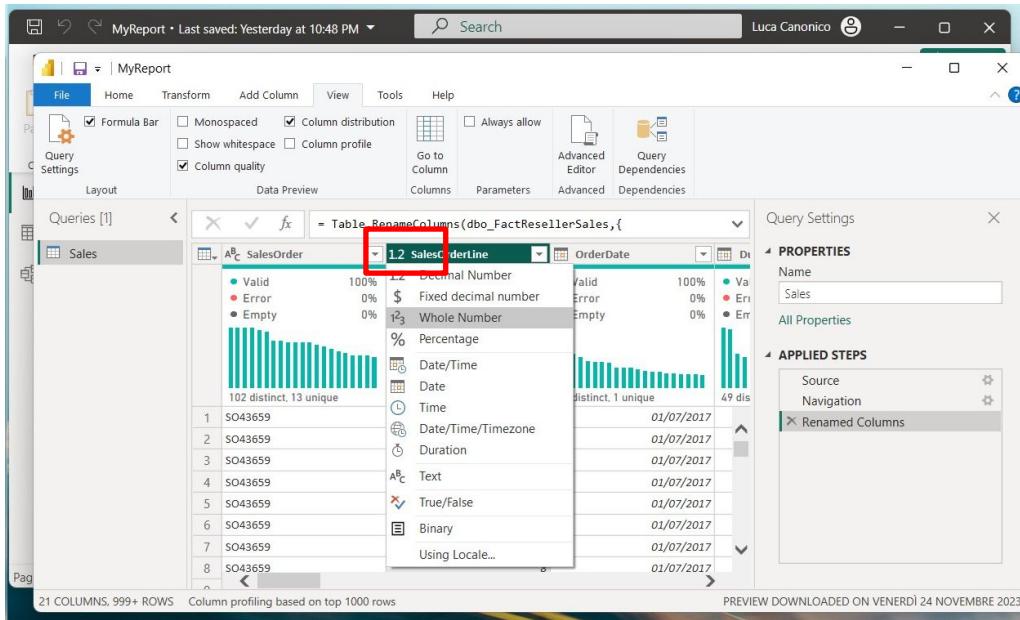
Properties Name Employee All Properties

Applied Steps Source Navigation Removed Other Columns Filtered Rows Removed Other Columns1

3 COLUMNS, 18 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 01:24

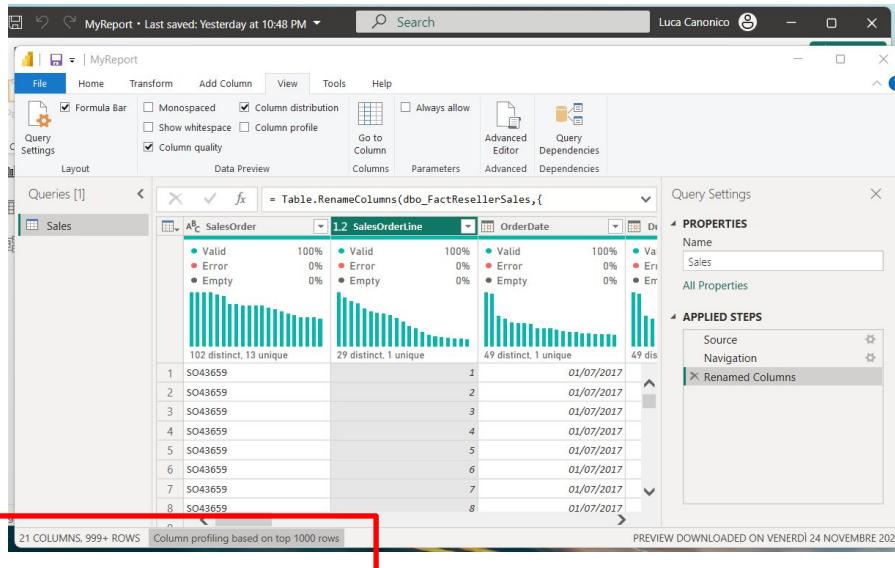
## Rimuovere campi e filtrare righe



The screenshot shows the EPICODE Data Profiler application interface. In the center, there's a preview of a table named 'Sales' from 'dbo\_FactResellerSales'. The table has one visible column, 'SalesOrder', which contains 102 distinct values. A dropdown menu is open over the second column, 'SalesOrderLine', with the option '12 SalesOrderLine' selected and highlighted with a red box. To the right, the 'Properties' pane displays the column's properties, including its name 'Sales'.

Assegnare il *data type* corretto a ciascun colonna coerentemente all'informazione descritta dalla colonna stessa!

## Rimuovere campi e filtrare righe



MyReport • Last saved: Yesterday at 10:48 PM

Search Luca Canonico

File Home Transform Add Column View Tools Help

Formula Bar Monospaced Column distribution  
Show whitespace Column profile  
Column quality

Always allow Go to Columns Advanced Editor Query Dependencies

Layout Data Preview Columns Parameters Advanced Dependencies

**Queries [1]**

Sales

1. SalesOrder 1.1 SalesOrderLine 1.2 SalesOrderLine OrderDate

Valid Error Empty Valid Error Empty Valid Error Empty

102 distinct, 13 unique 29 distinct, 1 unique 49 distinct, 1 unique 49 dist

1 SO43659 2 SO43659 3 SO43659 4 SO43659 5 SO43659 6 SO43659 7 SO43659 8 SO43659

100% 0% 0% 100% 0% 0% 100% 0% 0%

0% 0% 0% 0% 0% 0% 0% 0% 0%

0% 0% 0% 0% 0% 0% 0% 0% 0%

01/07/2017 01/07/2017 01/07/2017 01/07/2017 01/07/2017 01/07/2017 01/07/2017 01/07/2017

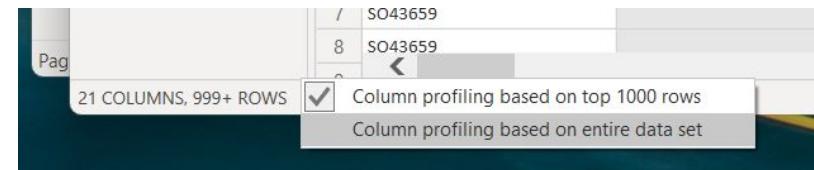
1 2 3 4 5 6 7 8

21 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED ON VENERDÌ 24 NOVEMBRE 2023

In generale, fai attenzione al numero di record della tabella. Se il numero di record > 1000 abilita la profilazione sull'intero dataset in modo tale PBI profili i campi valutando l'intero set di dati!

È limitato di default alle prime 1000 righe!!



SO43659

8 SO43659

1 SO43659

21 COLUMNS, 999+ ROWS

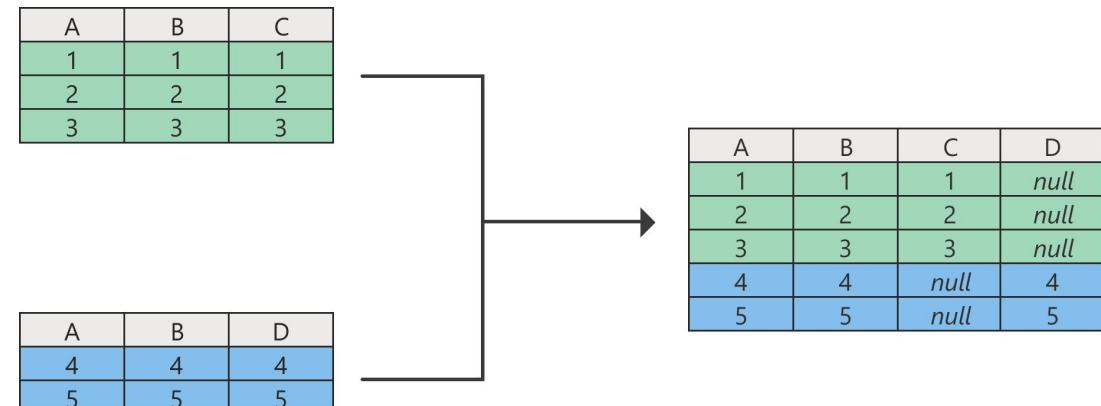
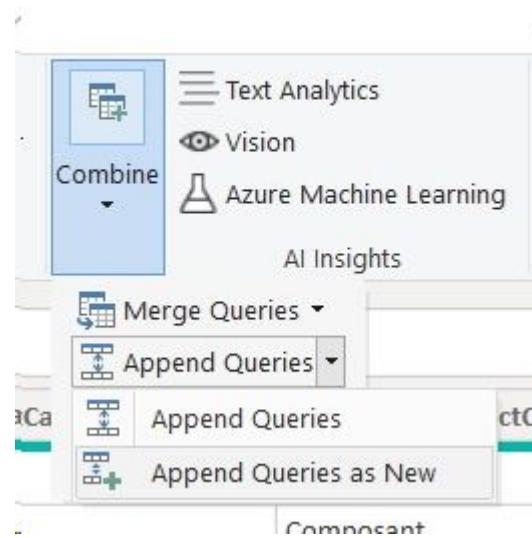
Column profiling based on top 1000 rows

Column profiling based on entire data set

## Combinare più tabelle in un'unica tabella

### Append

È l'operazione con la quale è possibile 'accodare' i record di due o più tabelle.

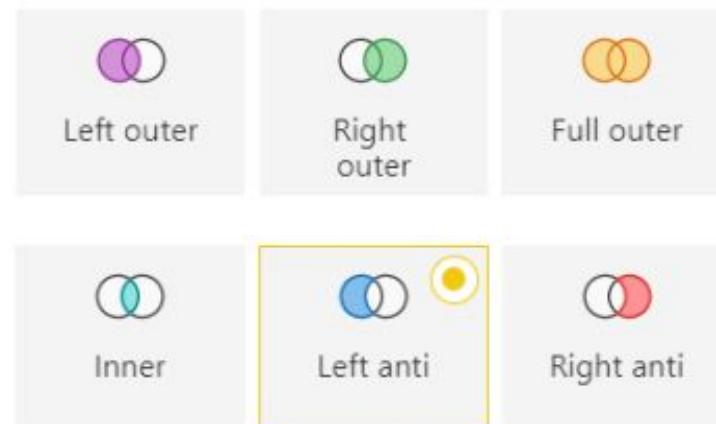
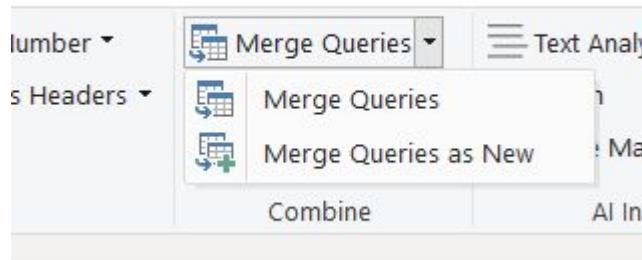


<https://learn.microsoft.com/en-us/power-query/append-queries>

## Combinare più tabelle in un'unica tabella

### Merge

È l'operazione con la quale è possibile combinare record di due tabelle in base alla corrispondenza dei valori di un campo in comune.



<https://learn.microsoft.com/it-it/power-query/merge-queries-overview>

## Check Point!

Hai importato un tabella da un database SQL. Qual è il metodo più veloce per capire se ci sono campi con valori vuoti?

- Abilitare il column profile
- Abilitare il column quality
- Abilitare il column distribution

Hai importato una tabella di 2100 record. Bisogna verificare l'esistenza di campi chiave. Cosa fai?

- Abiliti il column profile e il column distribution
- Bisogna scrivere una query nella sorgente dati
- Nessuna delle precedenti

Hai importato due tabelle: Prodotti e CategorieProdotti. Vuoi ridurre il numero di tabelle per ottimizzare il modello. Quale operazioni esegui?

- Append
- Merge

**Let's take a  
look!**



**GRAZIE**  
EPCODE