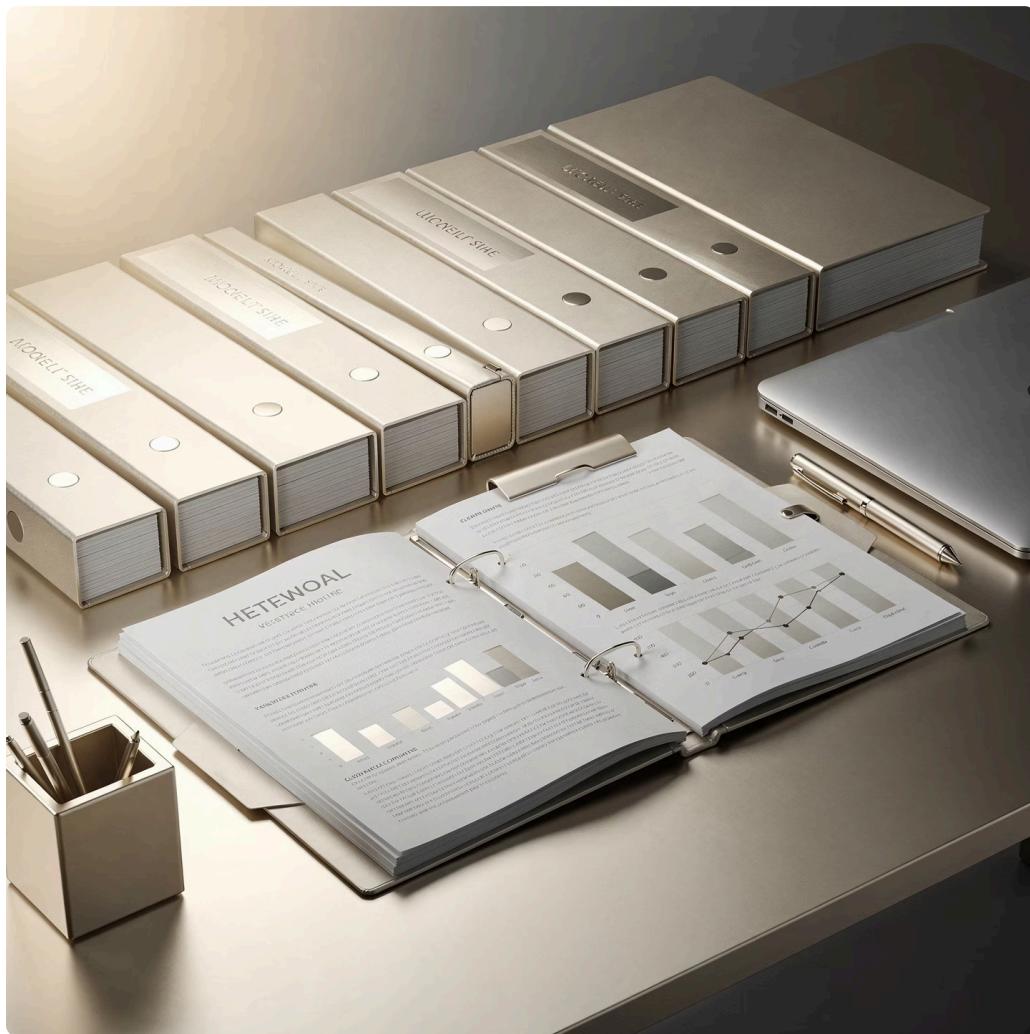


Tabelle Pivot e Tabelle di Contingenza

Comprendere e organizzare i dati in modo intelligente è fondamentale per qualsiasi professionista che lavora con l'analisi dei dati. In questo percorso esploreremo due strumenti essenziali che trasformano dati grezzi in informazioni strutturate e comprensibili.

A cosa servono questi strumenti?



Le tabelle pivot e le tabelle di contingenza sono strumenti fondamentali nell'arsenale di ogni analista di dati. Permettono di organizzare grandi quantità di informazioni in strutture chiare e leggibili.

Questi strumenti ci aiutano a studiare le relazioni tra diverse variabili, scoprendo pattern e connessioni che non sarebbero evidenti nei dati grezzi. Sono essenziali per analisi statistiche e probabilistiche avanzate.

Ma soprattutto, ci permettono di rispondere rapidamente a domande di business concrete, trasformando dati complessi in insight azionabili per decisioni strategiche informate.

Cos'è una Tabella Pivot?

Una tabella pivot è uno strumento potente che riorganizza i dati disponendoli in righe e colonne, calcolando automaticamente aggregazioni significative.

Immagina di avere centinaia di righe di dati di vendita: una pivot li trasforma in un formato compatto e leggibile.



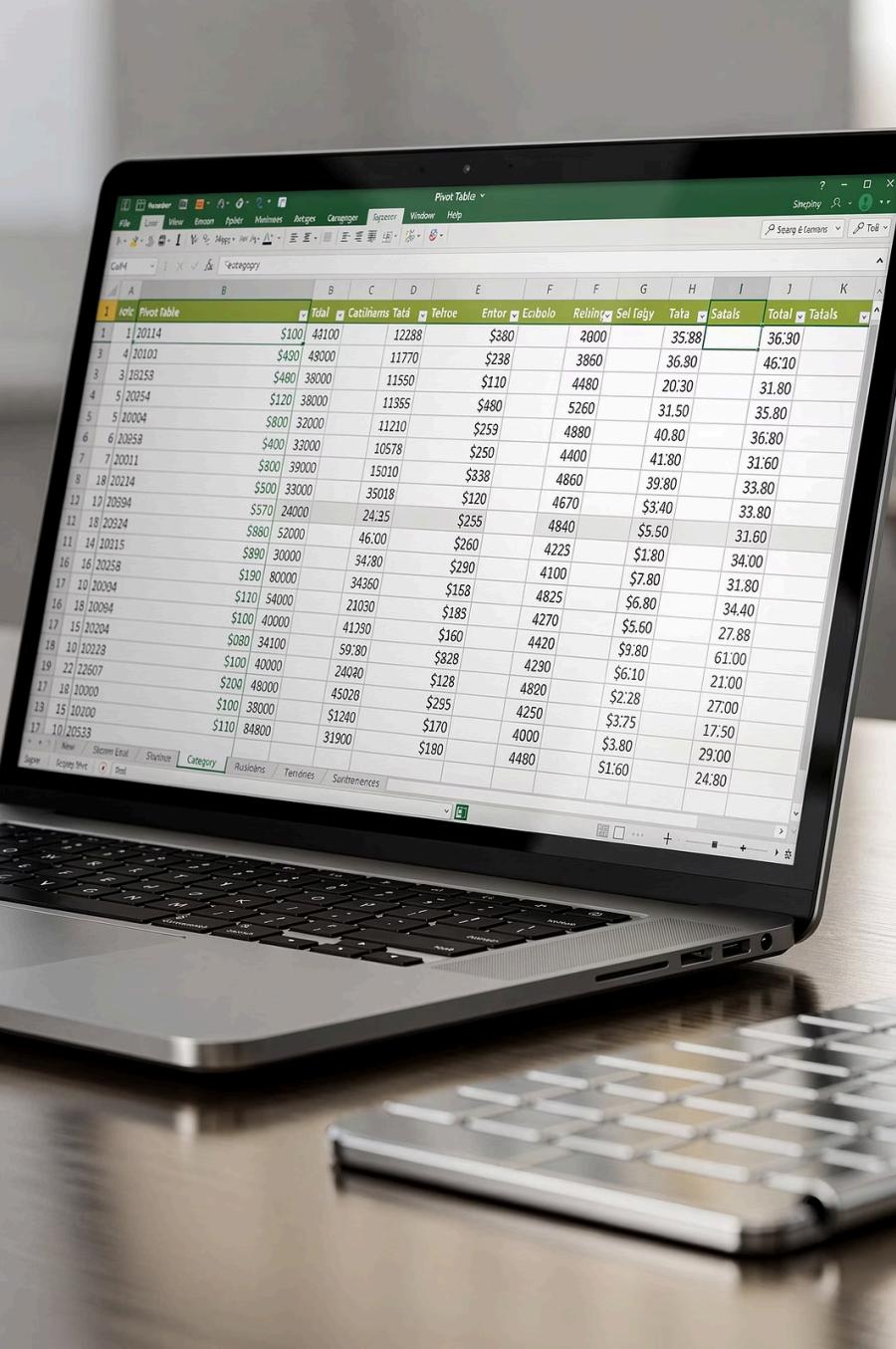
Casi d'uso comuni

Analisi vendite per categoria × mese, clienti per fascia età × area geografica, performance prodotti per regione × trimestre



Strumenti disponibili

Excel, Google Sheets, Pandas (Python), Power BI, Tableau, SQL con GROUP BY



Costruzione di una Tabella Pivot

Per comprendere meglio il funzionamento delle tabelle pivot, analizziamo un esempio pratico. Partiremo da una "tabella lunga" — il formato più comune per i dati grezzi — e vedremo come viene trasformata in una panoramica sintetica e significativa.

01

Dati Grezzi (Tabella Lunga)

Immaginate un foglio di calcolo con migliaia di righe, dove ogni riga rappresenta una singola transazione di vendita con colonne come "Data", "Prodotto", "Regione" e "Importo".

03

Definizione dell'Aggregazione

Infine, scegliamo la funzione di aggregazione per i valori che si incrociano. Vogliamo la "Somma dell'Importo" per visualizzare le vendite totali per ogni prodotto in ogni regione?

02

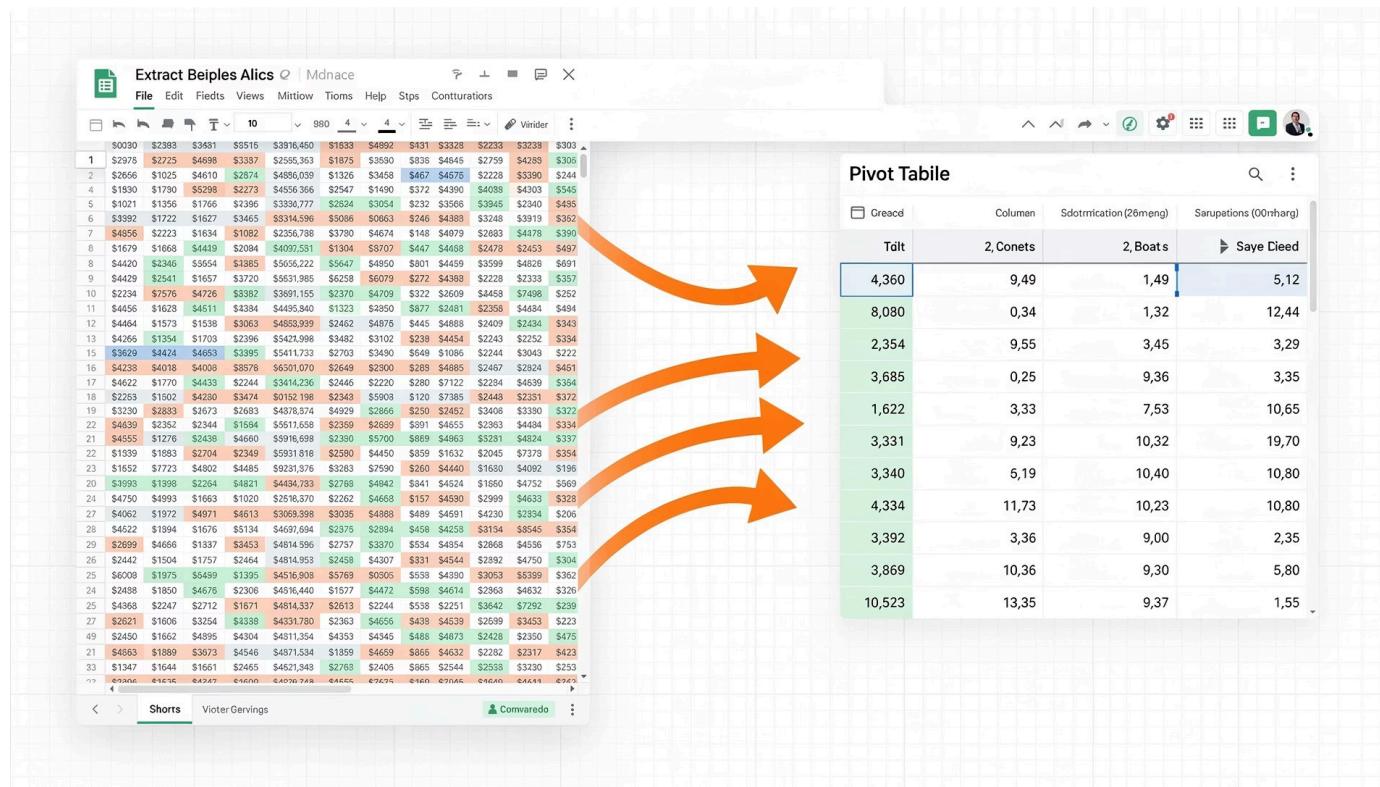
Scelta delle Variabili

Decidiamo quali colonne diventeranno le "righe" e le "colonne" della nostra tabella pivot. Per esempio, "Prodotto" per le righe e "Regione" per le colonne.

04

Risultato: La Tabella Pivot

Ottieniamo una tabella compatta che riassume le vendite aggregate per prodotto e regione, trasformando rapidamente i dati grezzi in un insight utile.



Costruzione di una Tabella Pivot

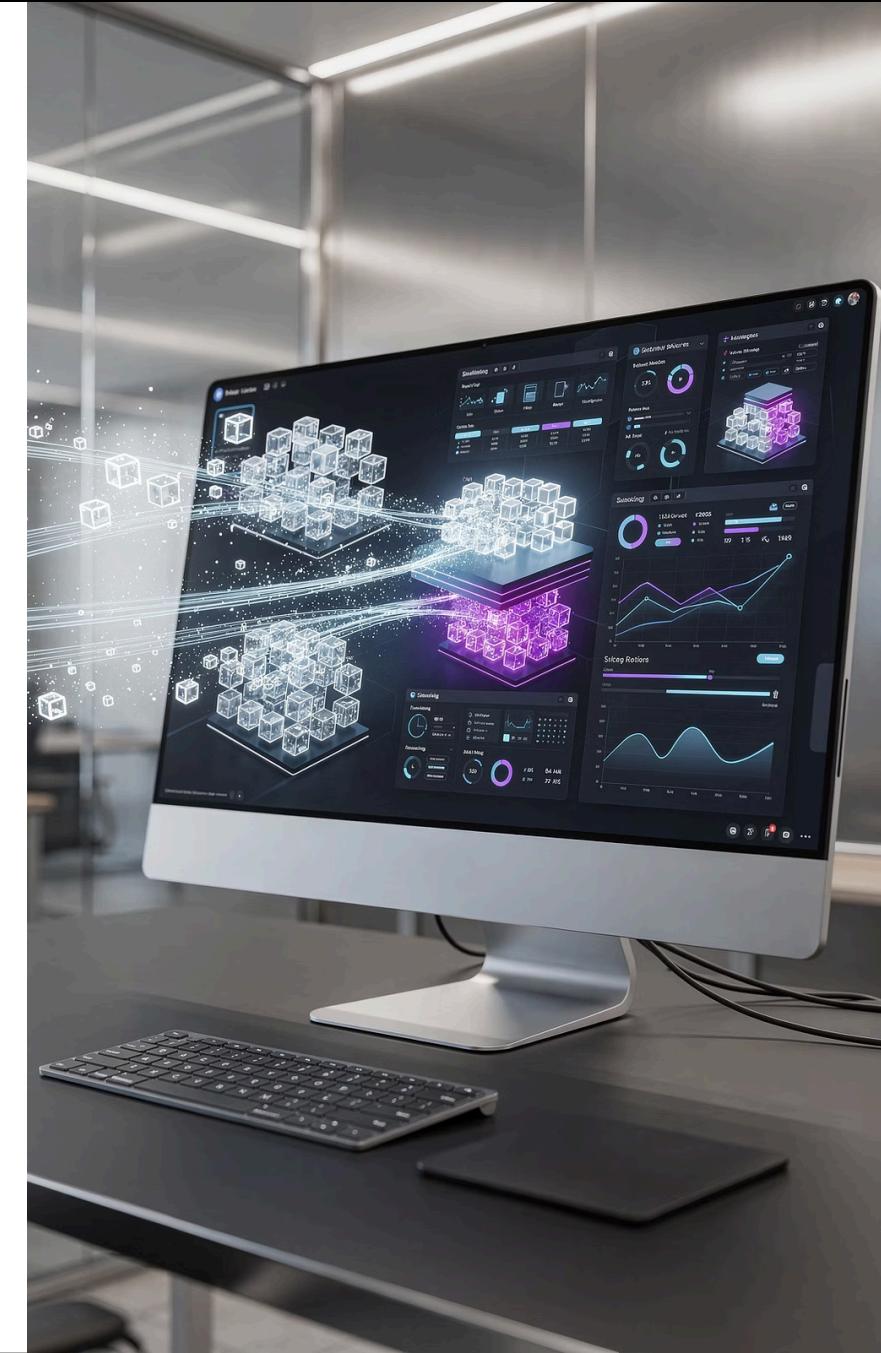
Il processo di creazione di una pivot inizia sempre da dati in **formato lungo**.

Questo formato è caratterizzato da una struttura dove ogni riga rappresenta una singola osservazione, con molte righe che ripetono gli stessi valori.

Caratteristiche del formato lungo

- Ogni riga è un'osservazione singola e indipendente
- Molte righe ripetono gli stessi valori (categoria, fascia, regione)
- Ottimo per calcoli e trasformazioni, ma non per visualizzare relazioni
- Necessario per operazioni JOIN e query SQL complesse
- NON permette di "vedere" immediatamente la relazione tra variabili

La tabella pivot trasforma questa struttura verticale in una aggregazione bidimensionale per righe e colonne, rendendo le relazioni immediatamente visibili e comprensibili.



Obiettivo: creare una tabella che riassume la relazione

Domanda del business

"Quante vendite di Laptop ci sono nella fascia 18–24?"

"E nella fascia 25–34?"

"Quale categoria è più popolare tra i giovani adulti?"

Per rispondere rapidamente a queste domande serve una struttura organizzata che metta in relazione fascia d'età e categoria prodotto.

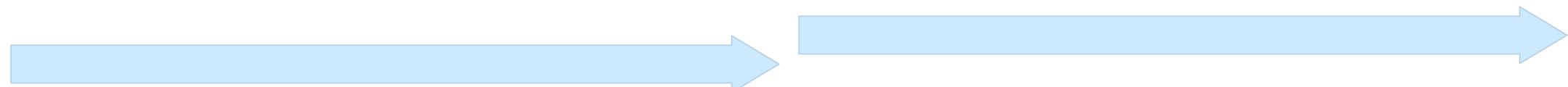


- La soluzione:** Con una tabella pivot possiamo vedere immediatamente quante vendite di ogni categoria ci sono in ciascuna fascia d'età, permettendo analisi rapide e decisioni informate.



Il processo di aggregazione

Partiamo dai dati grezzi in formato lungo e trasformiamoli passo dopo passo in una tabella pivot leggibile.

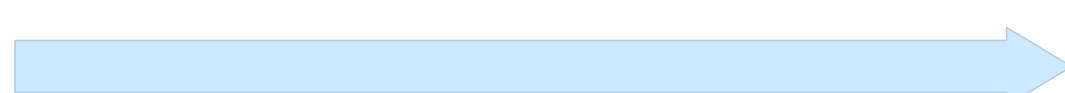


Dati di partenza

Righe individuali con fascia_eta, categoria, quantità per ogni transazione

Raggruppiamo per fascia_eta

Separiamo le osservazioni in gruppi basati sulla fascia d'età del cliente



Contiamo le categorie

Dentro ogni fascia, contiamo quante volte appare ciascuna categoria prodotto

Organizziamo in tabella

Creiamo una struttura bidimensionale con fasce in righe e categorie in colonne

Conteggio manuale delle vendite

Per comprendere il meccanismo, vediamo come si contano manualmente le vendite per ciascuna combinazione di fascia d'età e categoria prodotto.

Fascia 18–24

- Laptop: **2 vendite**
- Smartphone: 0 vendite
- Accessori: 0 vendite

Totale fascia: 2

Fascia 25–34

- Laptop: **1 vendita**
- Smartphone: **1 vendita**
- Accessori: 0 vendite

Totale fascia: 2

Fascia 35–44

- Accessori: **1 vendita**
- Laptop: 0 vendite
- Smartphone: 0 vendite

Totale fascia: 1

Questo conteggio manuale ci mostra chiaramente come i dati grezzi si trasformano in aggregazioni significative. Ogni numero rappresenta quante transazioni soddisfano quella specifica combinazione di fascia ed età e categoria.

Trasformazione in tabella larga (pivot)

Ora organizziamo i conteggi in una struttura bidimensionale che rende immediatamente visibili tutte le relazioni.

Fascia Età	Laptop	Smartphone	Accessori
18-24	2	0	0
25-34	1	1	0
35-44	0	0	1



Righe = gruppi della prima variabile

Ogni riga rappresenta una fascia d'età diversa



Colonne = valori della seconda variabile

Ogni colonna rappresenta una categoria prodotto



Celle = valori aggregati

Conteggi, somme, medie o altre aggregazioni

Questa è la **tabella pivot**: una rappresentazione compatta e leggibile che ci permette di rispondere istantaneamente a domande di business, identificare pattern e prendere decisioni informate.

Long Format

Category	Category	Category	Category	Category	Category
Date	Date	Date	Date	Date	Value
Category Date 101	Cat1 Date 101	Cat2 Date 101	Cat3 Date 101	Cat4 Date 101	Value 101
Category Date 311	Cat1 Date 311	Cat2 Date 311	Cat3 Date 311	Cat4 Date 311	Value 311
Category Date 122	Cat1 Date 122	Cat2 Date 122	Cat3 Date 122	Cat4 Date 122	Value 122
Category Date 100	Cat1 Date 100	Cat2 Date 100	Cat3 Date 100	Cat4 Date 100	Value 100
Category Date 100	Cat1 Date 100	Cat2 Date 100	Cat3 Date 100	Cat4 Date 100	Value 100
Category Date 180	Cat1 Date 180	Cat2 Date 180	Cat3 Date 180	Cat4 Date 180	Value 180
Category Date 201	Cat1 Date 201	Cat2 Date 201	Cat3 Date 201	Cat4 Date 201	Value 201
Category Date 170	Cat1 Date 170	Cat2 Date 170	Cat3 Date 170	Cat4 Date 170	Value 170
Category Date 282	Cat1 Date 282	Cat2 Date 282	Cat3 Date 282	Cat4 Date 282	Value 282
Category Date 102	Cat1 Date 102	Cat2 Date 102	Cat3 Date 102	Cat4 Date 102	Value 102
Category Date 122	Cat1 Date 122	Cat2 Date 122	Cat3 Date 122	Cat4 Date 122	Value 122
Category Date 185	Cat1 Date 185	Cat2 Date 185	Cat3 Date 185	Cat4 Date 185	Value 185
Category Date 100	Cat1 Date 100	Cat2 Date 100	Cat3 Date 100	Cat4 Date 100	Value 100
Category Date 126	Cat1 Date 126	Cat2 Date 126	Cat3 Date 126	Cat4 Date 126	Value 126
Category Date 364	Cat1 Date 364	Cat2 Date 364	Cat3 Date 364	Cat4 Date 364	Value 364
Category Date 101	Cat1 Date 101	Cat2 Date 101	Cat3 Date 101	Cat4 Date 101	Value 101
Category Date 202	Cat1 Date 202	Cat2 Date 202	Cat3 Date 202	Cat4 Date 202	Value 202
Category Date 358	Cat1 Date 358	Cat2 Date 358	Cat3 Date 358	Cat4 Date 358	Value 358
Category Date 114	Cat1 Date 114	Cat2 Date 114	Cat3 Date 114	Cat4 Date 114	Value 114
Category Date 200	Cat1 Date 200	Cat2 Date 200	Cat3 Date 200	Cat4 Date 200	Value 200
Category Date 674	Cat1 Date 674	Cat2 Date 674	Cat3 Date 674	Cat4 Date 674	Value 674
Category Date 201	Cat1 Date 201	Cat2 Date 201	Cat3 Date 201	Cat4 Date 201	Value 201
Category Date 812	Cat1 Date 812	Cat2 Date 812	Cat3 Date 812	Cat4 Date 812	Value 812
Category Date 284	Cat1 Date 284	Cat2 Date 284	Cat3 Date 284	Cat4 Date 284	Value 284
Category Date 186	Cat1 Date 186	Cat2 Date 186	Cat3 Date 186	Cat4 Date 186	Value 186
Category Date 425	Cat1 Date 425	Cat2 Date 425	Cat3 Date 425	Cat4 Date 425	Value 425
Category Date 180	Cat1 Date 180	Cat2 Date 180	Cat3 Date 180	Cat4 Date 180	Value 180
Category Date 210	Cat1 Date 210	Cat2 Date 210	Cat3 Date 210	Cat4 Date 210	Value 210
Category Date 113	Cat1 Date 113	Cat2 Date 113	Cat3 Date 113	Cat4 Date 113	Value 113
Category Date 115	Cat1 Date 115	Cat2 Date 115	Cat3 Date 115	Cat4 Date 115	Value 115
Category Date 202	Cat1 Date 202	Cat2 Date 202	Cat3 Date 202	Cat4 Date 202	Value 202
Category Date 114	Cat1 Date 114	Cat2 Date 114	Cat3 Date 114	Cat4 Date 114	Value 114
Category Date 515	Cat1 Date 515	Cat2 Date 515	Cat3 Date 515	Cat4 Date 515	Value 515
Category Date 251	Cat1 Date 251	Cat2 Date 251	Cat3 Date 251	Cat4 Date 251	Value 251
Category Date 111	Cat1 Date 111	Cat2 Date 111	Cat3 Date 111	Cat4 Date 111	Value 111
Category Date 171	Cat1 Date 171	Cat2 Date 171	Cat3 Date 171	Cat4 Date 171	Value 171
Category Date 135	Cat1 Date 135	Cat2 Date 135	Cat3 Date 135	Cat4 Date 135	Value 135
Category Date 297	Cat1 Date 297	Cat2 Date 297	Cat3 Date 297	Cat4 Date 297	Value 297
Category Date 100	Cat1 Date 100	Cat2 Date 100	Cat3 Date 100	Cat4 Date 100	Value 100

Pivot Table

Category	Date	Date	Date	Date	Value
Category Value	2,000	1,800	1,250	1,320	1,850
Category Value	1,500	1,930	1,550	1,700	1,950
Value	1,770	1,550	1,650	2,150	1,800
Value	1,400	1,050	1,850	2,250	1,720
11ofundone Value	1,700	1,500	2,940	2,720	1,770
Value	2,700	1,700	2,040	1,350	2,840
Value	1,770	1,860	1,770	1,500	1,850
Vatte	2,800	2,800	1,740	2,200	1,450
Value	1,400	1,000	2,450	2,100	1,450
Dedurigenley Value	1,800	1,550	2,850	2,850	2,750
Comrione Value	2,800	1,900	1,860	1,470	1,800
Seairpoirty	1,000	1,700	1,050	1,450	1,890
Category	1,800	1,500	1,450	1,800	1,450
Category	1,700	1,700	1,720	1,700	1,250
Category	1,800	2,910	1,700	1,600	1,950
10ainidore Value	1,000	1,600	1,000	1,800	2,970
11o durolons Value	1,000	1,000	1,950	1,000	1,800
1noindore Value	1,000	1,720	2,650	1,400	2,800
Sunicy	2,800	1,600	2,850	1,800	1,000
Cater Value	2,000	2,000	2,750	2,000	2,800
Oduininet Value	1,700	1,800	2,880	2,000	2,700
11 Sundists Value	1,800	1,690	1,700	1,200	1,090
Moiurolons Value	1,980	1,000	1,750	1,700	1,050
1loindors Value	1,000	1,000	1,800	1,750	1,000
1ducrolone Value	1,000	1,350	2,050	2,000	1,850

Visualizzazione della trasformazione

Prima: Formato Lungo

Dati verticali, molte righe ripetitive,
difficili da leggere ma ottimi per
calcoli

Dopo: Formato Largo

Dati organizzati in matrice, relazioni
evidenti, perfetti per analisi visiva

Come si fa in Pandas

Pandas, la libreria Python più utilizzata per l'analisi dati, offre una funzione potente per creare tabelle pivot con poche righe di codice.

```
pivot = df.pivot_table(  
    index='fascia_eta',  
    columns='categoria',  
    values='quantita',  
    aggfunc='count'  
)
```

Parametri chiave:

- `index`: variabile per le righe
- `columns`: variabile per le colonne
- `values`: campo da aggregare
- `aggfunc`: funzione di aggregazione (count, sum, mean, etc.)



Il risultato è una tabella larga, strutturata e pronta per l'analisi. Pandas gestisce automaticamente i valori mancanti e può applicare multiple funzioni di aggregazione contemporaneamente.



Il ponte verso la probabilità

La tabella pivot è il ponte tra i dati grezzi e la probabilità condizionata. Ci permette di trasformare dati "verticali" in una struttura "orizzontale", chiara e leggibile, su cui calcolare righe e colonne normalizzate.

Una volta creata la tabella pivot, possiamo facilmente calcolare probabilità normalizzando i valori per riga, per colonna, o sull'intera tabella. Questo ci permette di rispondere a domande probabilistiche complesse partendo da dati semplici.

Ad esempio, normalizzando per riga otteniamo $P(\text{Categoria} \mid \text{Fascia})$, mentre normalizzando per colonna otteniamo $P(\text{Fascia} \mid \text{Categoria})$. La tabella diventa così uno strumento statistico potente.

Cos'è una Tabella di Contingenza?

Una tabella di contingenza è una specializzazione della tabella pivot, focalizzata esclusivamente sui **conteggi** delle combinazioni di due o più variabili categoriali.

È lo strumento fondamentale in statistica per studiare l'associazione tra variabili qualitative. Mentre una pivot può contenere somme, medie o altre aggregazioni, una contingenza mostra sempre e solo frequenze.

Viene utilizzata per test statistici come il Chi-quadrato, per calcolare probabilità congiunte e condizionate, e per verificare l'indipendenza tra variabili.



		Category A		Category B		Group	
		Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
Category A	Category B	3.64	3.47	6.00	4.30	4.35	3.65
		3.74	1.90	1.11	3.95	3.35	1.24
Category A	Category B	3.92	3.13	4.33	4.71	4.55	3.87
		3.66	3.75	3.35	1.80	4.74	4.66
Category A	Category B	1.72	1.50	3.72	3.77	4.20	3.03
		3.31	1.54	4.34	1.50	4.70	3.61
Category A	Category B	4.22	1.55	3.54	3.75	4.85	3.00
		4.16	1.81	3.30	4.66	4.29	3.64
Category A	Category B	2.80	1.78	3.35	1.00	4.38	3.11
		4.54	2.55	3.60	1.17	4.00	3.66
Category A	Category B	2.87	2.86	4.50	1.17	5.70	4.71
		1.38	1.58	1.75	1.77	4.60	

Esempio di Tabella di Contingenza

Riprendiamo l'esempio delle vendite per vedere come appare una tabella di contingenza completa, con conteggi e totali marginali.

Fascia Età	Laptop	Smartphone	Accessori	Totale
18–24	2	0	0	2
25–34	1	1	0	2
35–44	0	0	1	1
Totale	3	1	1	5

I **totali marginali** (nelle ultime riga e colonna) mostrano le distribuzioni univariate di ciascuna variabile. Il totale generale (5) rappresenta il numero complessivo di osservazioni nel dataset.

Pivot vs Contingenza

Tabella Pivot

Può contenere qualsiasi tipo di aggregazione: somme, medie, conteggi, deviazioni standard, massimi, minimi

Strumento flessibile e versatile per business intelligence e analisi esplorativa dei dati

Utilizzata in Excel, Pandas, Power BI, Tableau

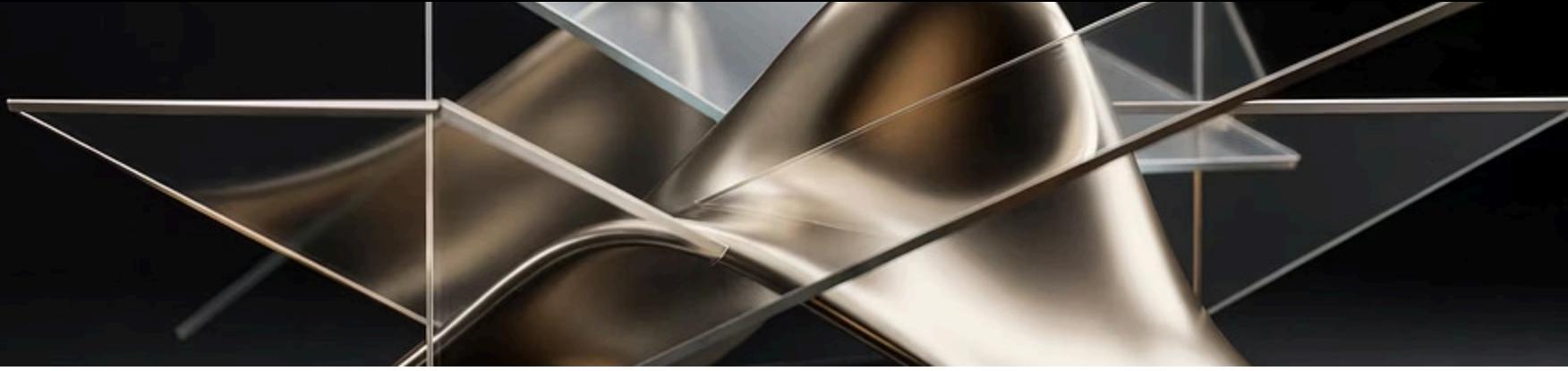
Tabella di Contingenza

Contiene esclusivamente conteggi (frequenze assolute) delle combinazioni di variabili categoriali

Strumento statistico specializzato per test di associazione e calcoli probabilistici

Utilizzata in statistica inferenziale e analisi categoriale

- ❑ **Relazione importante:** Ogni tabella di contingenza è una tabella pivot (con aggfunc='count'), ma non ogni pivot è una contingenza. La contingenza è un caso speciale della pivot.



Pivot e Probabilità

Una tabella pivot può essere trasformata in una rappresentazione probabilistica normalizzando i valori in modi diversi. Ogni tipo di normalizzazione risponde a domande probabilistiche differenti.

1

Normalizzazione per righe

Dividendo ogni cella per il totale della sua riga otteniamo **$P(\text{Categoria} | \text{Fascia})$** : la probabilità di una categoria dato che conosciamo la fascia d'età

2

Normalizzazione per colonne

Dividendo ogni cella per il totale della sua colonna otteniamo **$P(\text{Fascia} | \text{Categoria})$** : la probabilità di una fascia dato che conosciamo la categoria

3

Normalizzazione totale

Dividendo ogni cella per il totale generale otteniamo **$P(\text{Fascia, Categoria})$** : la probabilità congiunta di entrambe le variabili



Applicazione agli esercizi

Vediamo come applicare questi concetti a tre tipi di esercizi tipici che richiedono approcci diversi.

Esercizio 4: Probabilità diretta sulle righe

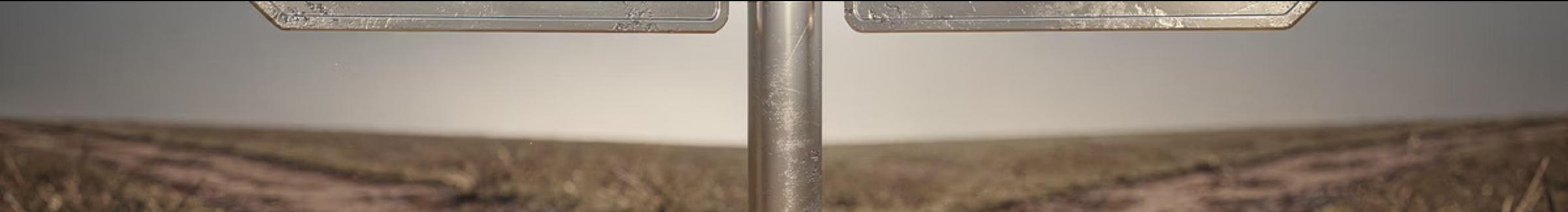
- 1** Quando l'esercizio chiede "Data la fascia d'età, qual è la probabilità della categoria?", normalizziamo per righe ottenendo $P(\text{Categoria} | \text{Fascia})$ direttamente

Esercizio 5: Probabilità diretta sulle colonne

- ² Quando l'esercizio chiede "Data la categoria, qual è la probabilità della fascia?", normalizziamo per colonne ottenendo $P(\text{Fascia} \mid \text{Categoria})$ direttamente

Esercizio 6: Teorema di Bayes

- Quando la direzione condizionale è inversa rispetto ai dati disponibili, dobbiamo usare Bayes per invertire la probabilità condizionata.



La regola d'oro

Direzione giusta o inversa?

Direzione giusta

La probabilità che cerchiamo segue la stessa direzione dei dati nella tabella

Soluzione: Probabilità diretta

Normalizziamo semplicemente per riga o colonna

Direzione inversa

La probabilità che cerchiamo ha direzione opposta rispetto ai dati disponibili

Soluzione: Teorema di Bayes

Dobbiamo invertire la probabilità condizionata usando $P(A|B) = P(B|A) \times P(A) / P(B)$

Questa regola ti aiuterà a decidere rapidamente quale approccio utilizzare in ogni situazione. Identifica sempre per prima cosa la direzione della probabilità richiesta.

Pivot in Pandas: codice pratico

Vediamo il codice Python completo per creare una pivot e calcolare le diverse probabilità.

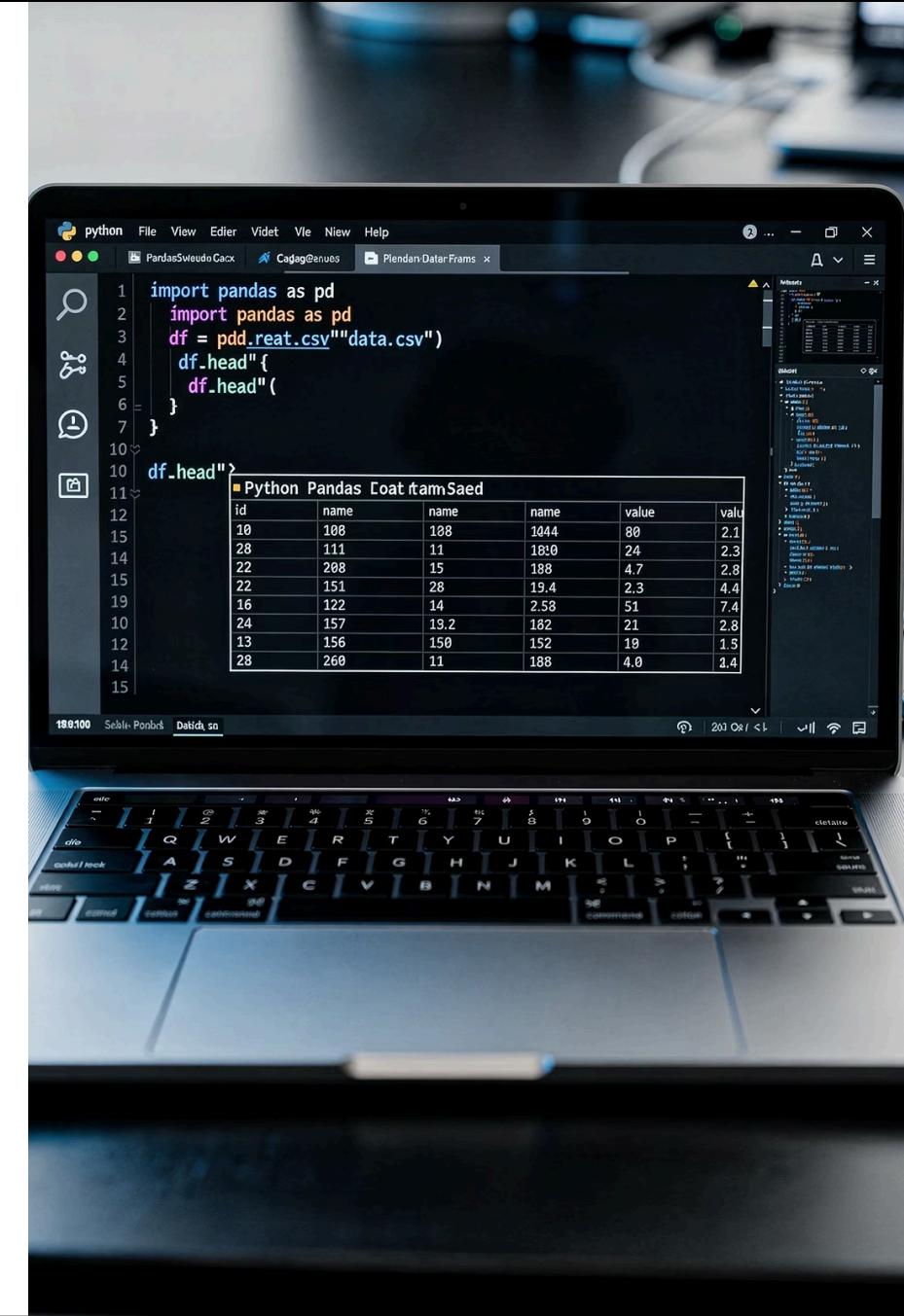
```
# Creazione della tabella pivot
pivot = df.pivot_table(
    index='fascia_eta',
    columns='categoria',
    values='n',
    aggfunc='sum'
)

# Probabilità per riga: P(Categoria | Fascia)
prob_riga = pivot.loc['18-24'] / pivot.loc['18-24'].sum()

# Probabilità per colonna: P(Fascia | Categoria)
prob_colonna = pivot['Laptop'] / pivot['Laptop'].sum()

# Probabilità congiunta
prob_congiunta = pivot / pivot.sum().sum()
```

Questi semplici comandi ci permettono di passare dai dati grezzi alle probabilità in pochi secondi, rendendo l'analisi statistica accessibile e veloce.



Conclusioni



Tabelle Pivot

Strumento versatile per organizzare e aggregare dati in business intelligence, permettendo analisi rapide e flessibili con qualsiasi tipo di aggregazione



Tabelle di Contingenza

Strumento statistico specializzato per analisi di frequenze, test di associazione e calcoli probabilistici su variabili categoriali



Base per Probabilità

Entrambe costituiscono il fondamento per calcolare probabilità condizionate, applicare il Teorema di Bayes e condurre analisi statistiche avanzate

Padroneggiare questi strumenti vi permetterà di trasformare dati complessi in informazioni chiare, prendere decisioni basate su evidenze concrete e comunicare insights in modo efficace al vostro team e stakeholder.

