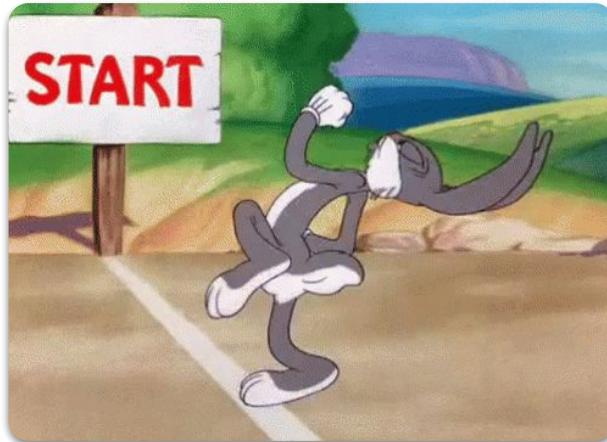




Google Looker Studio
Origini dati, EDA e primi grafici.

Agenda:

- Gestire le origini dati aggiunte
- Data blending in Looker Studio
- EDA e grafici base



Origini dati

Uno dei punti di forza di Looker Studio è la possibilità di connettersi a **più origini dati diverse** anche all'interno dello stesso report.

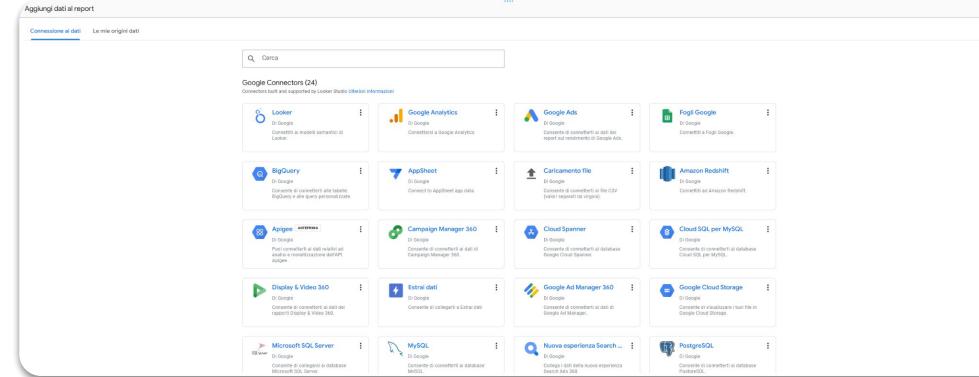
Questo permette di combinare informazioni raccolte e depositate in più "luoghi" diversi, coinvolgandole in modo omogeneo in un'**unica piattaforma**.

Origini dati

Possiamo aggiungere nuove origini dati dal report su cui stiamo lavorando o nella sezione dedicata a cui si accede dalla home page.

La differenza è che una connessione ad un'origine dati definita all'interno di un report sarà disponibile solo nello **scope del report** stesso.

Nel momento in cui creiamo un nuovo report ci viene subito chiesto di connetterci ad un'origine dati tramite questa interfaccia



Origini dati

La tab “**Connessione ai dati**” permette di selezionare un connettore per creare, appunto, una nuova connessione.

Oltre ai **connettori ufficiali di Google**, scrollando in basso si possono trovare quasi un migliaio di **connettori di terze parti**, che permettono di importare dati da piattaforme non supportate nativamente.

Il modo di completare la connessione all'origine dati può variare in base al connettore scelto, ma quasi tutti si basano su un qualche **sistema di autenticazione**.

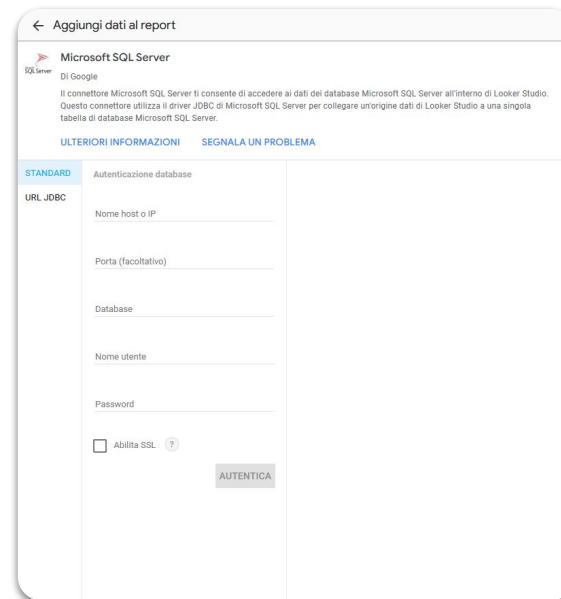
Nel caso di prodotti Google (Google Sheets, BigQuery...) l'autenticazione avviene tramite le autorizzazioni disponibili al nostro account, mentre in altri casi è necessario fornire delle credenziali.

Origini dati

Ad esempio, se vogliamo connetterci ad un database hostato e gestito tramite Microsoft SQL Server, cliccando sul suo connettore ci troveremo di fronte a questa interfaccia:

Sarà necessario stabilire la connessione usando delle credenziali valide e, una volta stabilita, potremo scegliere una tabella del database a cui connetterci.

NB: Il connettore gestisce **una tabella per volta**, o una query personalizzata, non l'intero database

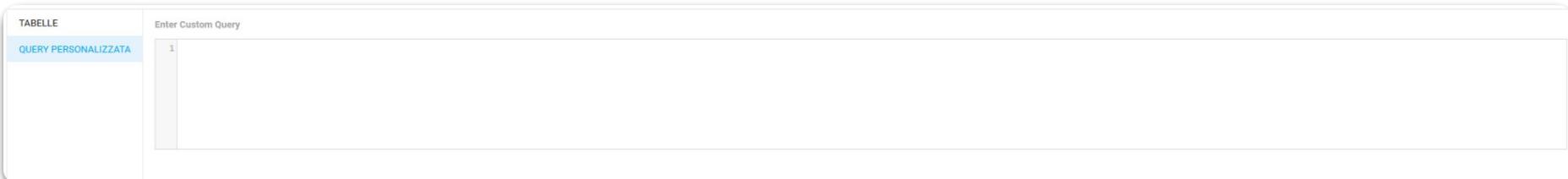


Origini dati

Collegandoci ad un database abbiamo anche l'opzione di connetterci ad un'origine dati partendo da una **query personalizzata**.

Per farlo, una volta stabilita la connessione al database, scegliamo l'opzione “**QUERY PERSONALIZZATA**” e definiamo la nostra origine dati attraverso una query SQL.

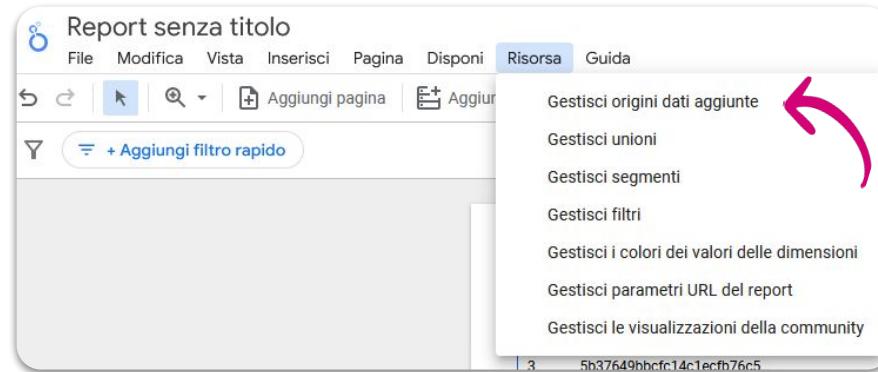
La query in questione non sarà legata ad una sola tabella ma si basa sulla connessione al database: avremo quindi la possibilità di gestire JOIN, SELECT innestate e altre interazioni tra tabelle, oltre ad avere a disposizione **tutti gli statement e tutte le possibilità legate al linguaggio SQL**.



Origini dati

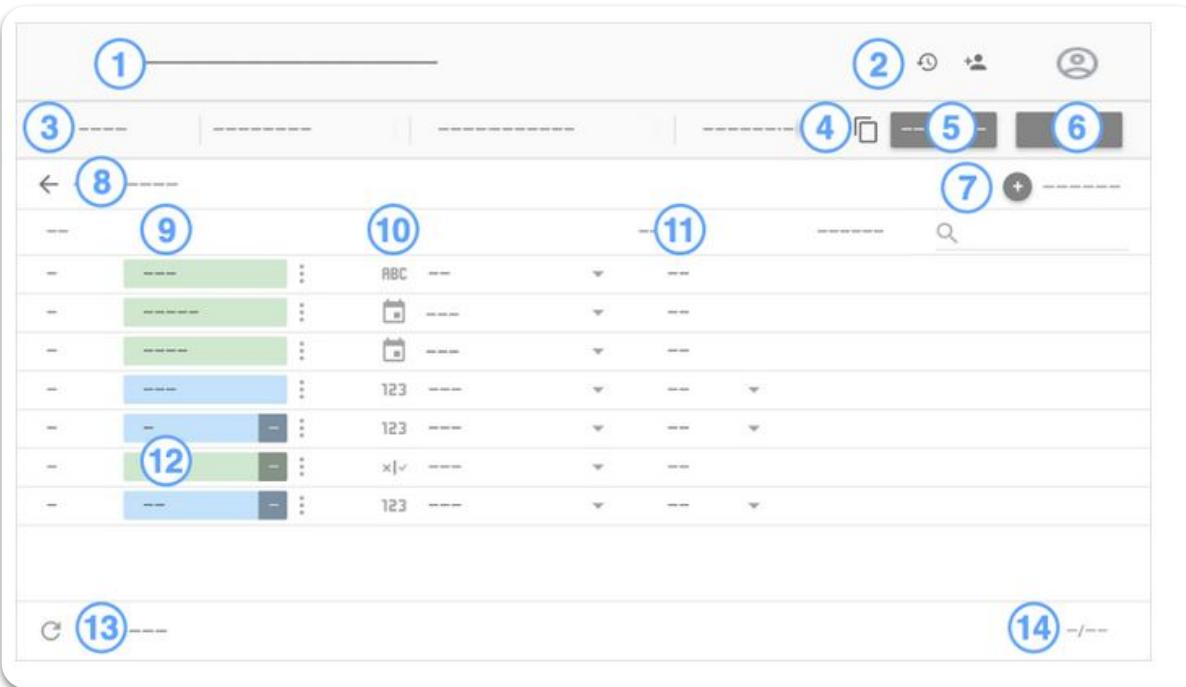
Per creare delle **origini dati indipendenti dal report**, schiacciamo sul tasto “Crea” nella home page di Looker Studio e selezioniamo “Origine Dati”.

L'unica differenza è che, una volta caricata l'origine dati verremo portati a una schermata di gestione dell'origine dati, che può essere raggiunta anche dal report tramite il menù “Risorsa” → “Gestisci origini dati aggiunte”



Origini dati

Gestione di un'origine dati



The screenshot shows the 'Data Sources' page in Looker Studio. On the left, there's a list of data sources with columns for name, type, and status. On the right, there's a toolbar with various icons. Numbered circles (1 through 14) point to specific elements:

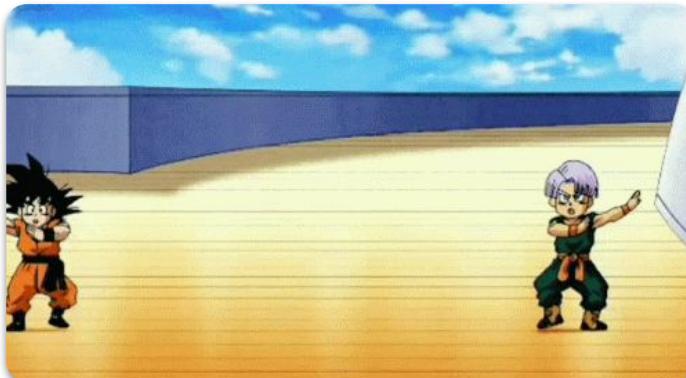
- 1: A text input field at the top.
- 2: A clock icon and a user icon in the top right.
- 3: A dropdown menu icon.
- 4: A trash bin icon.
- 5: A minus sign icon.
- 6: A plus sign icon.
- 7: A magnifying glass icon.
- 8: A back arrow icon.
- 9: A column header for a data source.
- 10: A column header for another data source.
- 11: A column header for a third data source.
- 12: A row in the data table.
- 13: A circular icon with a refresh symbol.
- 14: A footer message.

1. Nome dell'origine dati.
2. Controllo versione e tasto "Share".
3. Credenziali, aggiornamento dati, editing dei campi, visualizzazioni della comunità (presenti/non presenti in base all'ownership del dato).
4. Copia.
5. Crea un report utilizzando questa origine dati.
6. Crea un'esplorazione.
7. Aggiungi campo calcolato.
8. Modifica connessione.
9. Nome campo.
10. Tipologia Campo.
11. Aggregazione predefinita.
12. Visualizza funzioni dei campi calcolati.
13. Aggiorna la connessione.
14. Conteggio campi.

Data Blending

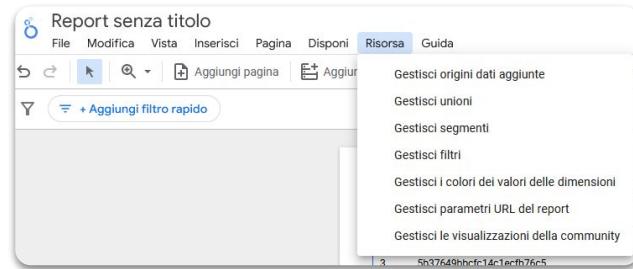
Looker studio permette di effettuare delle unioni tra le origini dati indipendentemente da quale sia il loro connettore.

Questo significa che, se le condizioni strutturali dei dati lo permettono, potremo unire dei dati provenienti da fonti diverse.

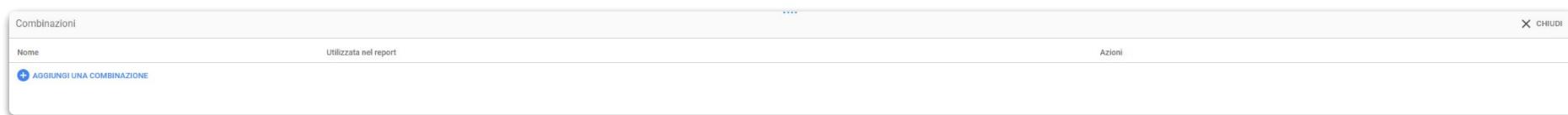


Data Blending

L'unione di più origini dati è possibile solo a **livello di report** e si può effettuare utilizzando l'interfaccia disponibile tramite il menù “Risorsa” → “Gestisci Unioni”

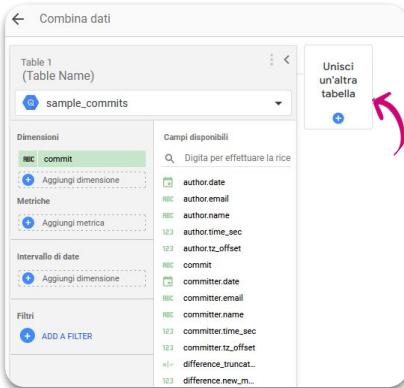


Ci troveremo di fronte ad un elenco delle unioni già definite (se ne abbiamo), e ad un tasto che permette di crearne di nuove



Combinazioni		
Nome	Utilizzata nel report	Azioni
+ AGGIUNGI UNA COMBINAZIONE		

Data Blending



Combina dati

Table 1 (Table Name)
sample_commits

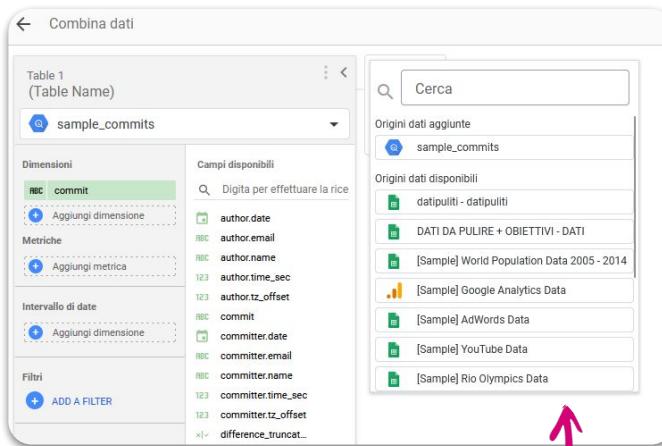
Dimensioni
REC commit

Metriche

Intervallo di date

Filtri

Unisci un'altra tabella



Combina dati

Table 1 (Table Name)
sample_commits

Campi disponibili

Cerca

Origini dati aggiunte

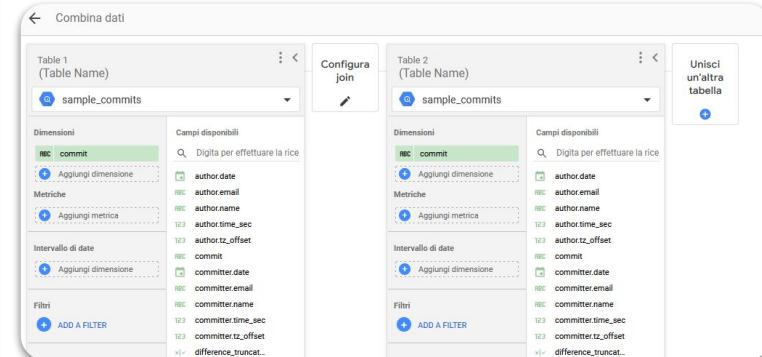
sample_commits

Origini dati disponibili

- datipuliti - datipuliti
- DATI DA PULIRE + OBIETTIVI - DATI
- [Sample] World Population Data 2005 - 2014
- [Sample] Google Analytics Data
- [Sample] AdWords Data
- [Sample] YouTube Data
- [Sample] Rio Olympics Data

Step 1: selezioniamo la prima tabella

Step 2: Tramite il tasto "Unisci un'altra tabella", selezioniamo la seconda tabella



Combina dati

Table 1 (Table Name)
sample_commits

Table 2 (Table Name)
sample_commits

Dimensioni
REC commit

Metriche

Intervallo di date

Filtri

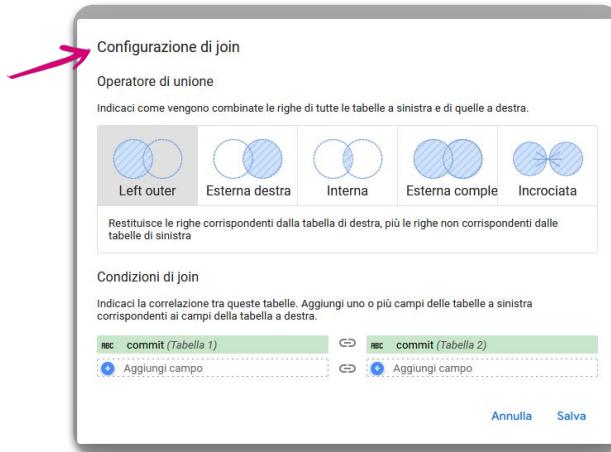
Configura join

Unisci un'altra tabella

Step 3: Gestione delle tue tabelle e dell'unione

Data Blending

Selezionando “**Configura Join**”, possiamo definire l’operatore di unione e le condizioni di Join.



La **gestione dei campi** della tabella risultante avviene tramite “drag and drop” sulle singole tabelle: l’output finale sarà visibile a destra, dove potremo anche dare un nome all’unione dati e salvarla.

Exploratory Data Analysis - Ripasso

Il successo di qualsiasi analisi dati inizia con una solida Exploratory Data Analysis (EDA). Questa fase è fondamentale per ottenere insight significativi e preparare il terreno per analisi più avanzate.



Exploratory Data Analysis - Ripasso

Alcune delle funzioni dell'EDA sono:

- ❑ **Identificare Pattern:** L'EDA consente di individuare pattern e tendenze nei dati, fornendo una comprensione intuitiva dei comportamenti chiave.
- ❑ **Individuare Outlier:** Attraverso grafici e statistiche descrittive, l'EDA aiuta a individuare dati anomali o outlier che potrebbero influenzare negativamente le analisi successive.
- ❑ **Comprendere la Distribuzione dei Dati:** L'EDA fornisce strumenti per visualizzare la distribuzione dei dati in modo chiaro e accessibile.
- ❑ **Preparare i Dati:** L'EDA aiuta a identificare e affrontare sfide come i dati mancanti, garantendo che l'analisi successiva sia basata su dati di alta qualità.

Exploratory Data Analysis - Ripasso

Statistiche utili:

valori: 2,6,9,15,20,35,46,80

- ❑ **Massimo:** valore massimo □ 80
- ❑ **Minimo:** valore minimo □ 2
- ❑ **Conteggio:** conteggio dei valori □ 8
- ❑ **Somma:** somma dei valori □ 213
- ❑ **Media:** somma di tutti i valori a disposizione diviso il numero complessivo dei valori □ 26,625
- ❑ **Mediana:** è il valore che occupa la posizione centrale di una serie di dati ordinati in modo crescente □ se il conteggio è dispari si prende il valore centrale centrale, se è pari si fa la media tra i due valori centrali, in questo caso $(15+20)/2$ □ 17,5

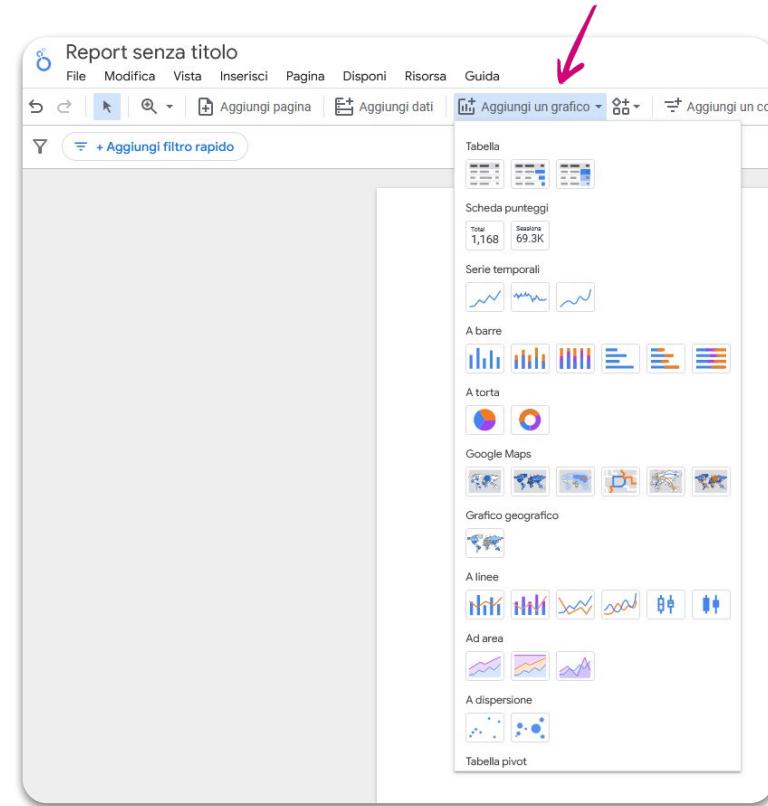
Exploratory Data Analysis - Ripasso

- **Quartile:** data una serie di dati ordinati in modo crescente vengono segmentati in 4 parti □
calcolo primo quartile formula $K = np$ □ $n=8$ $p=0,25$ □ $K = 8*0,25 = 2$, se il valore è decimale si arrotonda per eccesso e si prenderà il valore in quella posizione, se il valore è intero si fa la media tra il valore in posizione k e il suo successivo, in questo caso $(6+9)/2 = 7,5$. Il terzo quartile si calcola nello stesso modo sostituendo 0,25 con 0,75, quindi nel nostro caso 40,5. Il secondo quartile corrisponde alla mediana.
- **Centile/Percentile:** data una serie di dati ordinati in modo crescente vengono segmentati in 100 parti □ la formula è la stessa del quartile ovviamente cambiando il valore p ad esempio il 95 percentile di questi dati è □ 63

Primi Grafici

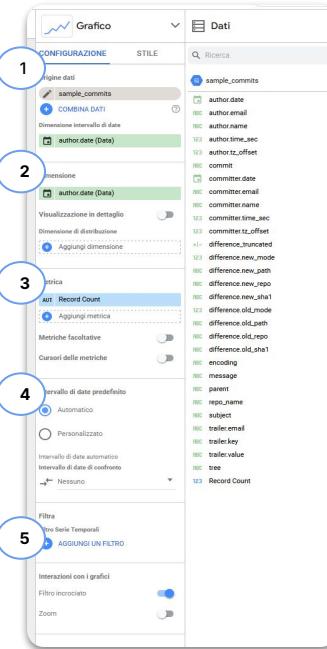
Uno dei modi più facili e veloci per avvicinarsi ad un'**esplorazione dei dati** è quello delle **visualizzazioni**.

Creare dei grafici è alla base di uno strumento di reporting come Looker Studio ed è quindi un'operazione resa molto intuitiva, basta scegliere l'opzione "Aggiungi un grafico" e selezionare cosa ci interessa.



Primi Grafici

Una volta selezionata la tipologia di grafico, verrà automaticamente inserito nel report con dei **settaggi base** che potremo modificare selezionandolo e andando ad agire sul menù a destra:



Le opzioni saranno leggermente diverse a seconda della tipologia di grafico scelto, ma l'essenza sarà la stessa. Avremo la possibilità di:

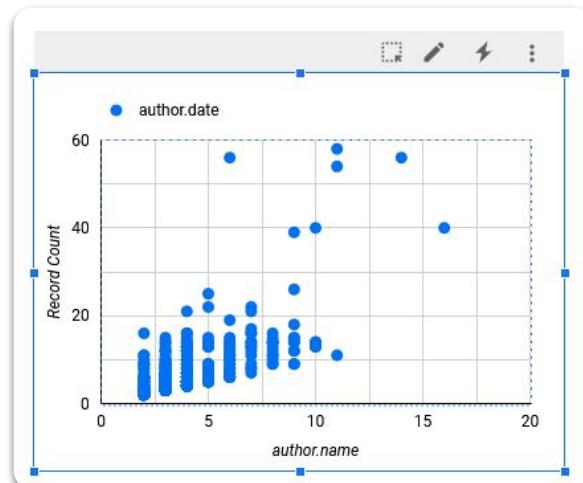
1. Selezionare l'origine dati.
2. Definire le dimensioni.
3. Definire le metriche.
4. Identificare un intervallo di date.
5. Aggiungere dei filtri.

La definizione di dimensioni e metriche avviene tramite un drag and drop dall'elenco sulla destra, oppure schiacciando sul tasto "aggiungi dimensione" (o metrica) ed agendo sul menù a tendina.

Primi Grafici

Esistono alcune tipologie di grafici che si prestano meglio agli step di EDA elencati prima, anche se dobbiamo ricordarci che ogni dataset presenta delle specifiche e va analizzato con attenzione e curiosità, senza limitarsi a delle operazioni standard.

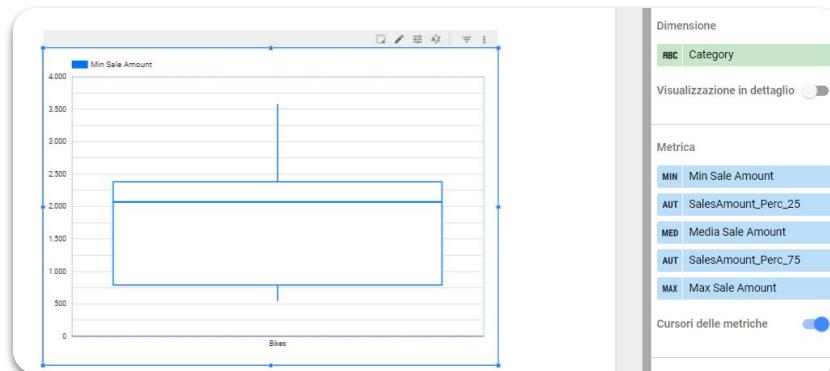
Per l'identificazione dei **pattern** possiamo usare, ad esempio, un **grafico a dispersione**:



Un grafico a dispersione è un tipo di rappresentazione grafica utilizzata per visualizzare la relazione tra due variabili

Primi Grafici

Per l'individuazione degli **outliers** possiamo usare, ad esempio, un **boxplot**.



Un boxplot, o diagramma a scatola, è una rappresentazione grafica che fornisce una sintesi della distribuzione statistica di un insieme di dati. Mostra la mediana, i quartili e eventuali outlier mediante una "scatola" e linee chiamate "whisker". Questo permette di valutare la dispersione e la simmetria della distribuzione dei dati in modo rapido e efficace.

Per creare un boxplot è necessario indicare 5 misure precise: valore minimo, primo quartile/percentile, valore mediano, terzo quartile/percentile, valore massimo.

Primi Grafici

Per l'individuazione degli **outliers** possiamo usare, ad esempio, un **boxplot**.



The screenshot shows the 'Data' tab in Looker Studio. At the top, there are several status indicators: Ambito: Riutilizzabile, Credenziali dati: Archiviate, Aggiornamento dei dati: 12 ore, Accesso alle visualizzazioni della community: Attiva, Modifica campi nei report: Attivata, and a blue button labeled FATTO. Below these, a section titled 'Nome campo' contains the text 'Sales_amount_percentile_25'. To the right, under 'ID campo', it says 'calc_34wmbk23ed'. A 'Formula' section shows the input '1 PERCENTILE([SalesAmount], 25)'. On the right side of the formula input, there are buttons for 'FORMATTA FORMULA' and 'SALVA'. At the bottom right of the dialog are 'ANNULLA' and 'SALVA' buttons. A green checkmark icon is located at the bottom left of the dialog.

Per creare i valori percentili nella tab Dati di Looker seguire i seguenti passaggi:

- 1) Cliccare su aggiungere i campo
- 2) Selezionare aggiungi campo calcolato
- 3) Indicare il nome del campo
- 4) Inserire la seguente formula
PERCENTILE(nome_campo, 25) sostituire il nome_campo con quello interessato
- 5) Cliccare Salva in basso a dx
- 6) Cliccare Fatto in alto a dx
- 7) Ripetere per la seconda percentuale sostituendo 25 con 75 sia nel nome che nella formula

Primi Grafici

Per preparare i dati, visualizzare valori mancanti, capire meglio i dati testuali ed in generale per una visione più particolareggiata, possiamo usare una **tabella**.

	Company	Ad Type	Advertiser Type	Product Recall	Sponsor?	%intent	% Exposed (Feel)
1.	United	Brand	Non-Sponsors	I do recall	N	42,86%	22,22%
2.	United	Brand	Non-Sponsors	I don't recall	N	42,86%	22,22%
3.	Powerade	Inspiring	Sponsors	I do recall	Y	36,36%	32%
4.	Powerade	Inspiring	Sponsors	I don't recall	Y	36,36%	32%
5.	Nike	Inspiring	Sponsors	I don't recall	Y	33,33%	33,33%
6.	Nike	Inspiring	Sponsors	I do recall	Y	33,33%	33,33%
7.	Coca Cola	Inspiring	Sponsors	I don't recall	Y	29,17%	33,33%
8.	Coca Cola	Inspiring	Sponsors	I do recall	Y	29,17%	33,33%
n	Tide	Product	Sponsors	I don't recall	v	20%	21,14%



GRAZIE
EPCODE