**W13 D2 – Google Looker Studio: Origini Dati, EDA e Primi Grafici**

**1. Origini dati**

Uno dei principali punti di forza di Google Looker Studio è la possibilità di connettere più origini dati, anche all’interno dello stesso report. Questo consente di centralizzare informazioni provenienti da fonti diverse, rendendole accessibili in un’unica piattaforma di analisi e visualizzazione.

È possibile aggiungere nuove origini dati direttamente all'interno del report o tramite la sezione dedicata della homepage. La differenza principale è che un'origine dati collegata tramite un report sarà accessibile solo all'interno di quel singolo report, mentre le connessioni create esternamente saranno riutilizzabili.

All’avvio di un nuovo report, Looker Studio propone subito di collegarsi a un’origine dati. La sezione “Connessione ai dati” consente di selezionare un connettore tra quelli ufficiali Google o tra quasi un migliaio di connettori di terze parti. Questi connettori permettono di integrare anche fonti dati non supportate nativamente.

Il processo di collegamento varia a seconda del connettore selezionato. Di norma, richiede un sistema di autenticazione. Ad esempio, se si collega un prodotto Google come Google Sheets o BigQuery, l’accesso viene autorizzato tramite il proprio account Google. Per altre piattaforme, è necessario inserire credenziali manualmente.

Un caso tipico è la connessione a un database ospitato su Microsoft SQL Server: dopo aver selezionato il connettore, occorre inserire le credenziali di accesso e scegliere una tabella o una query personalizzata da utilizzare. È importante notare che un singolo connettore gestisce una tabella o una query per volta.

Per creare un'origine dati tramite query personalizzata, è sufficiente selezionare l’apposita opzione e scrivere il codice SQL desiderato. In questo modo è possibile includere JOIN, subquery e altre istruzioni complesse, sfruttando appieno le potenzialità del linguaggio SQL.

Per creare origini dati indipendenti dai report, dalla homepage di Looker Studio è sufficiente cliccare su “Crea” e poi su “Origine dati”. Una volta definita, la schermata di gestione sarà accessibile anche dal report tramite il menu “Risorsa” > “Gestione origini dati aggiunte”.

**Immagine che contiene testo, numero, Carattere, schermata

Descrizione generata automaticamente**

**2. Data Blending (Unione dati)**

Google Looker Studio consente l’unione (data blending) di origini dati, anche se provengono da fonti diverse, purché strutturalmente compatibili. Questa funzione è disponibile solo all’interno dei report e si gestisce dal menu “Risorsa” > “Gestione unioni”.

All'interno della schermata di gestione, è possibile vedere un elenco delle unioni già create ed eventualmente aggiungerne di nuove.

**Immagine che contiene schermata, testo

Descrizione generata automaticamente**

Per creare una nuova unione, si clicca su “Configura join” e si definisce il tipo di operazione da eseguire (es. LEFT OUTER JOIN, INNER JOIN, FULL OUTER JOIN) e le condizioni di join. Le colonne da unire possono essere trascinate manualmente tramite l’interfaccia, e sarà possibile visualizzare l’anteprima dell’unione. Infine, è possibile dare un nome alla nuova tabella blended e salvarla.

**3. Exploratory Data Analysis (EDA)**

L’EDA, o Analisi Esplorativa dei Dati, rappresenta la prima fase di ogni processo analitico. Serve a comprendere le caratteristiche principali di un dataset prima di procedere con modelli predittivi o interpretazioni avanzate. Una buona EDA aiuta a:

* Rilevare pattern nascosti
* Identificare outlier (valori anomali)
* Comprendere la distribuzione delle variabili
* Verificare la qualità e la completezza dei dati

**Metriche principali dell’EDA**

* **Massimo e Minimo**: determinano l’intervallo dei valori.
* **Conteggio (Count)**: totale delle osservazioni.
* **Media**: valore medio, calcolato come

Media=∑xin\text{Media} = \frac{\sum x\_i}{n}Media=n∑xi​​

* **Somma**: totale assoluto dei valori.
* **Mediana**: valore centrale del dataset ordinato.
* **Moda**: valore che compare con maggiore frequenza.
* **Quartili**: dividono il dataset in 4 parti uguali.  
  Calcolo:
  + Q1 (primo quartile) = valore al 25° percentile
  + Q2 = mediana
  + Q3 (terzo quartile) = valore al 75° percentile
  + Formula approssimativa:

Qk=k(n+1)4-esimo valoreQ\_k = \frac{k(n+1)}{4}\text{-esimo valore}Qk​=4k(n+1)​-esimo valore

* **Percentili**: indicano il valore sotto il quale cade una certa percentuale del dataset.  
  Formula:

Pk=k(n+1)100P\_k = \frac{k(n+1)}{100}Pk​=100k(n+1)​

(con interpolazione se il valore non è intero)

**4. Visualizzazioni: primi grafici**

Le visualizzazioni sono il modo più diretto e intuitivo per esplorare e comunicare informazioni dai dati. In Looker Studio, basta cliccare su “Aggiungi un grafico” e selezionare il tipo desiderato (tabelle, grafici a barre, a linee, scatter plot, boxplot ecc.). Il grafico verrà generato automaticamente con delle impostazioni di default, personalizzabili tramite il menu laterale.

**Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente**

**Esempi pratici:**

* **Grafico a dispersione (Scatter Plot)**:  
  Ideale per analizzare la relazione tra due variabili quantitative.  
  Ogni punto rappresenta una coppia (x, y) e la distribuzione complessiva permette di evidenziare correlazioni, cluster o anomalie.
* **Boxplot**:  
  Utilizzato per rappresentare la distribuzione dei dati. Mostra visivamente:
  + Minimo
  + Primo quartile (Q1)
  + Mediana (Q2)
  + Terzo quartile (Q3)
  + Massimo
  + Eventuali outlier (se presenti)

**Screen 4**

**5. Tabelle: preparazione e pulizia dati**

Le tabelle sono uno strumento utile per visualizzare i valori grezzi, rilevare dati mancanti o incongruenze, e analizzare in dettaglio le variabili. Durante questa fase di "data cleaning" è possibile:

* Controllare la presenza di null o NaN
* Analizzare la completezza delle colonne
* Verificare formati errati o duplicati

Una preparazione accurata dei dati è fondamentale per garantire l'affidabilità delle analisi e delle visualizzazioni successive.