

Basi di Dati

Google Data Studio – Quaderno 3

1. Collezione dati

I dati per svolgere il quaderno possono essere scaricati dalla sezione relativa ai quaderni sul sito del corso.

In uno stesso progetto di Google data studio bisogna importare separatamente entrambe le collezioni. Per ciascun grafico è possibile selezionare la collezione utilizzata cliccando sul grafico e selezionando la collezione corretta nella sezione "Origine Dati".

Bike Sharing: La prima collezione dati è riferita a un servizio di bike sharing operante a Chicago. I campi principali del dataset sono riassunti nella tabella 1.

Nome campo	Tipo	Descrizione
trip_id	Integer	Identificativo del noleggio
year	Integer	Anno estratto dalla data di noleggio
month	Integer	Mese estratto dalla data di noleggio
week	Integer	Numero della settimana nell'anno estratto dalla data di noleggio
day	Integer	Giorno estratto dalla data di noleggio
hour	Integer	Ora estratta dalla data di noleggio
usertype	Stringa	Può assumere due valori: "Customer" per gli utenti che hanno acquistato un abbonamento giornaliero, "Subscriber" per gli utenti in possesso di un abbonamento annuale
gender	Stringa	Genere della persona che effettua il noleggio, può assumere due valori, "Male" per gli uomini e "Female" per le donne.
starttime	Timestamp	Giorno e ora nel quale il noleggio è cominciato.
stoptime	Timestamp	Giorno e ora nel quale il noleggio è finito.
tripduration	Integer	Durata del noleggio in minuti.

temperature	Integer	Temperatura esterna in gradi Fahrenheit °F
events	string	Descrizione di eventuali eventi atmosferici verificatisi durante il noleggio (es. Rain or Snow)
from_station_id	Integer	Identificativo della stazione di inizio noleggio.
from_station_name	String	Nome della stazione di inizio noleggio.
latitude_start	Float	Latitudine del punto di inizio noleggio.
longitude_start	Float	Longitudine del punto di inizio noleggio.
dpcapacity_start	Integer	Numero di stalli nella stazione di inizio noleggio
dpcapacity_end	Integer	Numero di stalli nella stazione di fine noleggio
to_station_id	Integer	Identificativo della stazione di fine noleggio.
to_station_name		Nome della stazione di fine noleggio.
latitude_end	Float	Latitudine del punto di fine noleggio.
longitude_end	Float	Longitudine del punto di fine noleggio.

Tabella 1: Descrizione dataset Bike Sharing

Covid19-ITA: La seconda collezione dati si riferisce ai dati sull'andamento della pandemia di Coronavirus in Italia. La tabella 2 descrive il dataset nel dettaglio.

Nome campo	Tipo	Descrizione
data	Timestamp	Data e ora del dato (es. 2020-12-02T17:00:00)
stato	Stringa	Nazione di riferimento (nel nostro caso ITA)
ricoverati_con_sintomi	Integer	Il numero di persone attualmente ricoverate che presentano sintomi.
terapia_intensiva	Integer	Il numero di persone, positivo al coronavirus, che si trovano attualmente ricoverati in terapia intensiva.

totale_ospedalizzati	Integer	Il numero di persone, positivo al coronavirus, che si trovano attualmente ospedalizzati.
isolamento_domiciliare	Integer	Il numero di persone, positivo al coronavirus, che si trovano attualmente in isolamento domiciliare.
totale_positivi	Integer	Il numero totale di persone attualmente positive al coronavirus su tutto il territorio nazionale.
variazione_totale_positivi	Integer	La differenza tra il totale di positivi del giorno corrente e quello precedente.
nuovi_positivi	Integer	Il numero totale di persone risultate positive al coronavirus in data odierna (indicata dal campo data).
dimessi_guariti	Integer	Il numero totale di persone guarite dal coronavirus su tutto il territorio nazionale (da inizio pandemia).
deceduti	Integer	Il numero totale di persone decedute a causa del coronavirus su tutto il territorio nazionale (da inizio pandemia).
totale_casi	Integer	Il numero totale di persone risultate positive al coronavirus su tutto il territorio nazionale (da inizio pandemia).

Tabella 1: Descrizione dataset Covid19. I dati non descritti non sono necessari per l'analisi.
Fonte: <https://github.com/pcm-dpc/COVID-19>

2. Analisi

Per effettuare l'esercitazione bisogna utilizzare il proprio account di [Google Data Studio](#). Bisogna importare i dati in formato .csv similmente a quanto effettuato durante i laboratori.

Per svolgere il quaderno è richiesto effettuare le analisi elencate di seguito. È richiesto agli studenti di:

- Identificare gli eventuali dati derivati e aggiungere i campi opportuni al set di dati.
- Identificare ed impostare eventuali filtri necessari per la visualizzazione.
- Per ciascuna visualizzazione identificare le opportune metriche e dimensioni.

NB. Ad ogni analisi corrisponde una sola visualizzazione.

Analisi 1) [Bike Sharing]: Per ciascun intervallo orario (es. 00, 01 ... etc) visualizzare il numero di noleggi effettuati e la loro durata media.

Tipo di visualizzazione: Grafico combinato a linee e barre

Analisi 2) [Bike Sharing]: Per ciascun giorno presente nel set di dati visualizzare la percentuale media di noleggi effettuati da uomini e donne.

Tipo di visualizzazione: Grafico a colonne in pila 100%

Analisi 3) [Bike Sharing]: Per ciascuna stazione identificare il numero di noleggi che sono stati iniziati dalla stessa. Ordinare la visualizzazione in ordine decrescente per numero di noleggi. Riportare la stazione nella quale sono stati iniziati il numero maggiore di noleggi.

Tipo di visualizzazione: Grafico a linee

Analisi 4) [Bike Sharing]: Visualizzare l'andamento del numero di noleggi al variare della temperatura esterna in gradi celsius. In particolare, è richiesta la conversione del campo temperature. Per rendere più leggibile la visualizzazione è richiesto di arrotondare ogni dato relativo alla temperatura al suo intero più vicino. Per usare la temperatura in gradi celsius è possibile definire un campo, nominato a piacere, definito secondo la seguente funzione:

```
ROUND((temperature - 32) / 1.8 , 0)
```

Tipo di visualizzazione: Grafico a linee

(Hint: <https://support.google.com/datastudio/answer/9422513?hl=en>)

Analisi 5) [Bike Sharing]: Per i noleggi effettuati dagli utenti "Subscriber", visualizzare il numero di noleggi effettuati per periodo del giorno. In particolare, sono definiti i seguenti periodi del giorno:

- Notte: dalle 00 alle 06
- Mattina: dalle 06 alle 12
- Pomeriggio: dalle 12 alle 18
- Sera: dalle 18 alle 24.

È inoltre necessario escludere dall'analisi i noleggi effettuati in condizioni metereologiche "cloudy".

Per definire gli intervalli orari è possibile definire un campo derivato definito secondo la seguente funzione:

```
CASE
  WHEN hour >= 0 AND hour < 6 THEN "Night"
  WHEN hour >= 6 AND hour < 12 THEN "Morning"
  WHEN hour >= 12 AND hour < 18 THEN "Afternoon"
  WHEN hour >= 18 AND hour < 24 THEN "Evening"
END
```

Tipo di visualizzazione: Grafico a colonne

(Hint: <https://support.google.com/datastudio/answer/7020724?hl=it>)

Analisi 6) [Bike Sharing]: Per ogni condizione atmosferica durante il noleggio, visualizzare la percentuale di noleggi effettuati in tale condizione. Riportare l'evento presente nella maggior parte dei noleggi.

Tipo di visualizzazione: Grafico ad anello

Analisi 7) [Bike Sharing]: Visualizzare in formato numerico il numero di noleggi effettuati in condizioni metereologiche "rain or snow".

Tipo di visualizzazione: Prospetto con numeri compatti

Analisi 8) [Bike Sharing]: Visualizzare in formato numerico il numero di noleggi che hanno avuto durata superiore ai 10 minuti.

Tipo di visualizzazione: Prospetto con numeri compatti

Analisi 9) [Covid19-ITA]: Confrontare l'andamento temporale del numero totale di casi, numero di persone attualmente positive e numero di persone guarite da inizio pandemia.

Tipo di visualizzazione: Serie Temporali

Analisi 10) [Covid19-ITA]: Confrontare l'andamento temporale del numero di persone in terapia intensiva, ricoverati con sintomi e isolamento domiciliare (a partire da inizio pandemia).

Tipo di visualizzazione: Serie Temporali

3. Consegna dell'elaborato

Per la valutazione del quaderno è necessario consegnare il PDF ottenuto esportando il report di Google Data Studio (File -> Scarica Come -> PDF). Si ricorda che per eventuali commenti è possibile inserire nel report delle caselle di testo con le proprie annotazioni. È specificatamente richiesto che le diverse analisi siano identificabili nel report.

Si ricorda inoltre che il nome del file del pdf deve essere assegnato secondo il seguente formato: **QuadernoX_Matricola** dove *X* è il numero del quaderno (3 per il report di Google Data Studio), *Matricola* è da sostituire con quella dello studente. L'estensione del file (pdf) **non** è da ripetere nel nome del file. Esempio di nome del file valido (comprensivo dell'estensione): *Quaderno1_181818.pdf* oppure *Quaderno2_222222.pdf*