



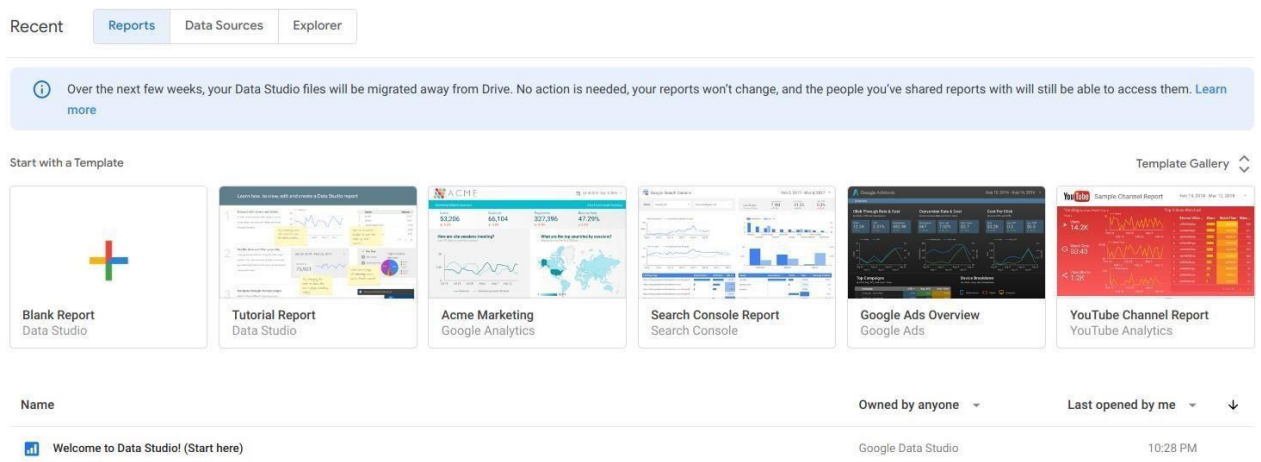
Data Science e Tecnologie per le Basi di Dati
Politecnico di Torino

Data Studio – Esercitazione 2

1. Login

Connettiti a **Looker Data Studio** by Google, accedi con un account Google o [crea un nuovo account Google gratuito](#).

- <https://lookerstudio.google.com/>



2. Report di benvenuto

Clicca su "**Tutorial Report**" e segui il tutorial.

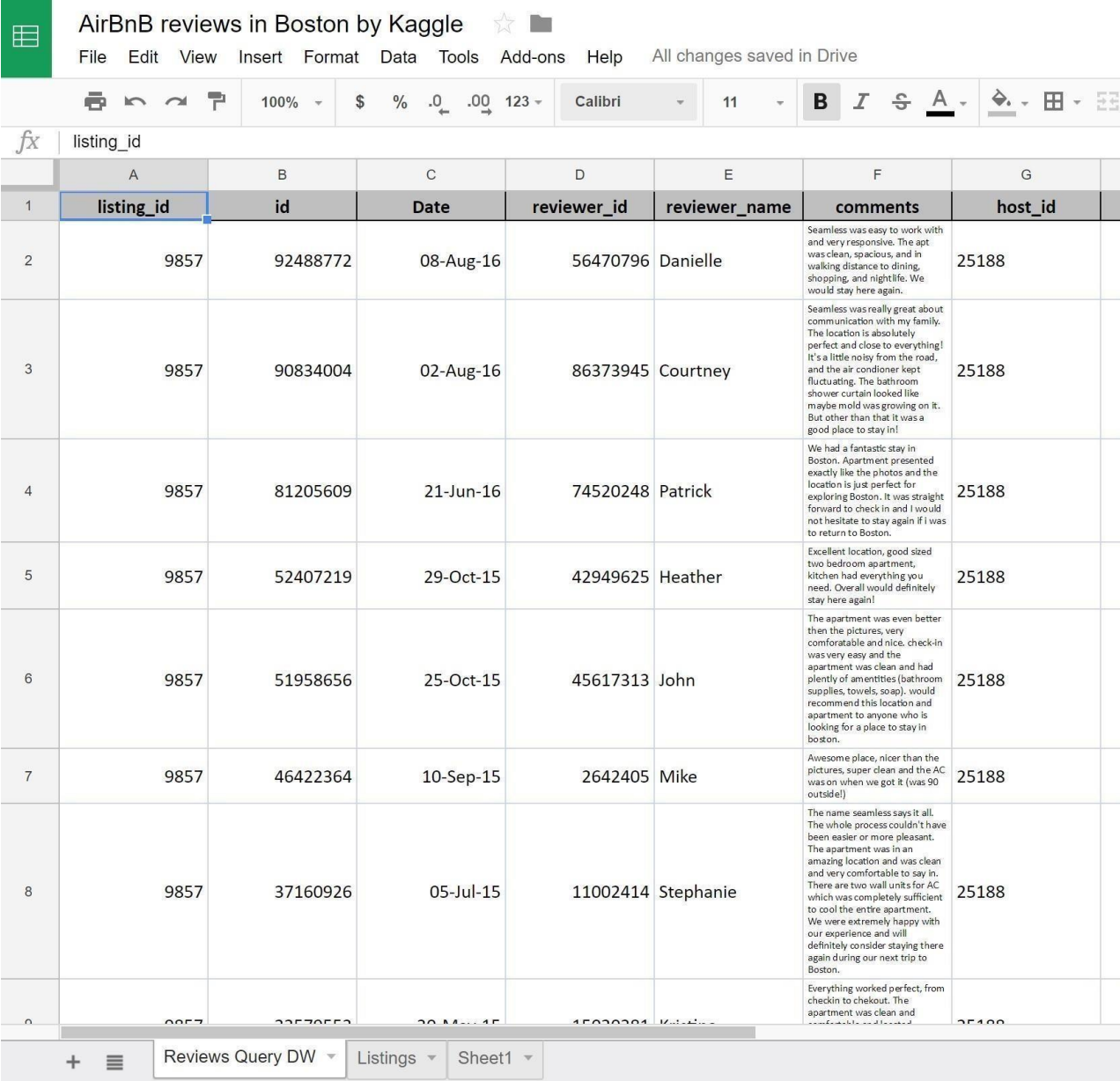
- <https://lookerstudio.google.com/reporting/0B5FF6JBKbNJxOWItcWo2SVVVeGc>
- impara le nozioni di base dello strumento Data Studio copiando il "Welcome report" e seguendo passo dopo passo le istruzioni fornite.
- pagina 8, "Track report usage with Google Analytics", può essere tranquillamente saltata.



3. Nuovo report sulle recensioni di Airbnb Boston

Per creare un nuovo report da zero, è necessario identificare un'origine dati. A tal fine, una parte del set di **Kaggle dataset of the Airbnb reviews in Boston** è stata caricata in **Fogli Google condiviso** per essere utilizzato come origine di dati per Google Data Studio.

- i Fogli Google, con circa 10k recensioni da utilizzare come origine di dati, sono disponibili a <https://docs.google.com/spreadsheets/d/1a2c9vCMFFFDXmhjoEoX2EwS2IYTbgE4WfZY72TXW9c0/edit#gid=285360760>



listing_id	A	B	C	D	E	F	G
1	listing_id	id	Date	reviewer_id	reviewer_name	comments	host_id
2	9857	92488772	08-Aug-16	56470796	Danielle	Seamless was easy to work with and very responsive. The apt was clean, spacious, and in walking distance to dining, shopping, and nightlife. We would stay here again.	25188
3	9857	90834004	02-Aug-16	86373945	Courtney	Seamless was really great about communication with my family. The location is absolutely perfect and close to everything! It's a little noisy from the road, and the air conditioner kept fluctuating. The bathroom shower curtain looked like maybe mold was growing on it. But other than that it was a good place to stay in!	25188
4	9857	81205609	21-Jun-16	74520248	Patrick	We had a fantastic stay in Boston. Apartment presented exactly like the photos and the location is just perfect for exploring Boston. It was straight forward to check in and I would not hesitate to stay again if I was to return to Boston.	25188
5	9857	52407219	29-Oct-15	42949625	Heather	Excellent location, good sized two bedroom apartment, kitchen had everything you need. Overall would definitely stay here again!	25188
6	9857	51958656	25-Oct-15	45617313	John	The apartment was even better than the pictures, very comfortable and nice. check-in was very easy and the apartment was clean and had plenty of amenities (bathroom supplies, towels, soap). would recommend this location and apartment to anyone who is looking for a place to stay in boston.	25188
7	9857	46422364	10-Sep-15	2642405	Mike	Awesome place, nicer than the pictures, super clean and the AC was on when we got it (was 90 outside!)	25188
8	9857	37160926	05-Jul-15	11002414	Stephanie	The name seamless says it all. The whole process couldn't have been easier or more pleasant. The apartment was in an amazing location and was clean and very comfortable to stay in. There are two wall units for AC which was completely sufficient to cool the entire apartment. We were extremely happy with our experience and will definitely consider staying there again during our next trip to Boston.	25188
9	9857	33570553	20-May-15	45000001	Kristina	Everything worked perfect, from checkin to checkout. The apartment was clean and comfortable.	25188

- Dedica un po' di tempo a comprendere i dati leggendo la loro descrizione su Kaggle e guardando la tabella su Google Sheets.
- La tabella dell'origine dei dati è stata creata unendo le tabelle originali "Annunci" e "Recensioni" fornite da Kaggle ed esportando le prime 10k righe unite ordinate per "listing_id" crescente.

Origine dati

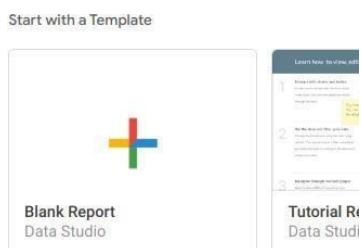
Le origini dati hanno due tipi di campi: dimensioni e metriche.

- Una **dimensione** è una categoria di dati.
- Una **metrica** è un numero che quantifica qualcosa in quella categoria.
- Un report di Data Studio ti consente di visualizzare tali dimensioni e metriche in grafici e tabelle.
- Nei pannelli delle origini dati e delle proprietà dei report di Data Studio, le dimensioni vengono visualizzate come chip **verdi**, mentre le metriche vengono visualizzate come chip **blu**.

Field	ID
Browser	t0._ga:browser_
Sessions	t0._ga:sessions_

Creare un nuovo report

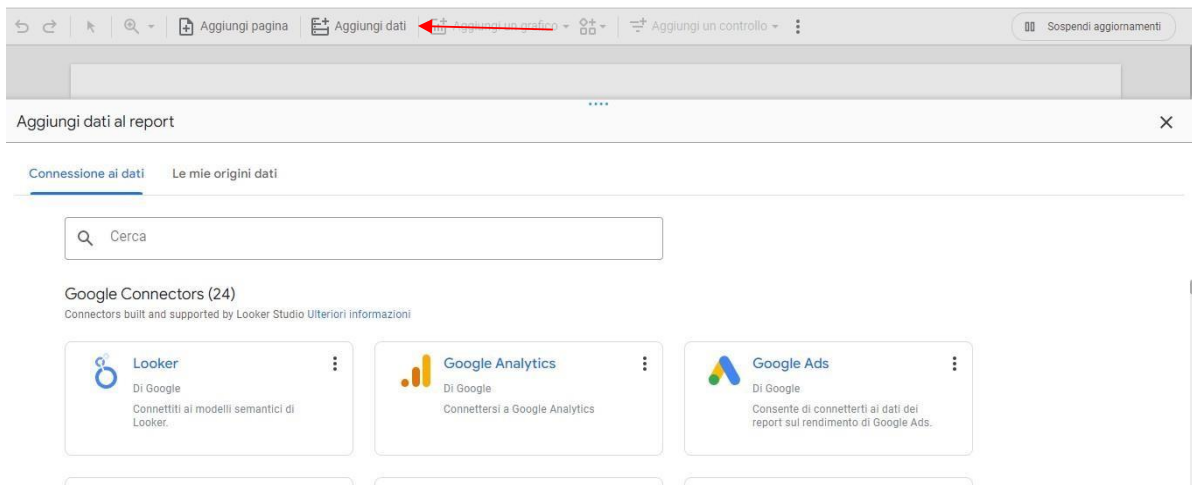
- Vai alla home page di Data Studio.
- Fare clic su "inizia un nuovo report" (vuoto).



- Chiudi la finestra predefinita "Aggiungi dati al report" e rinomina il "Report senza titolo" con un nome a tua scelta facendo clic sul nome stesso.

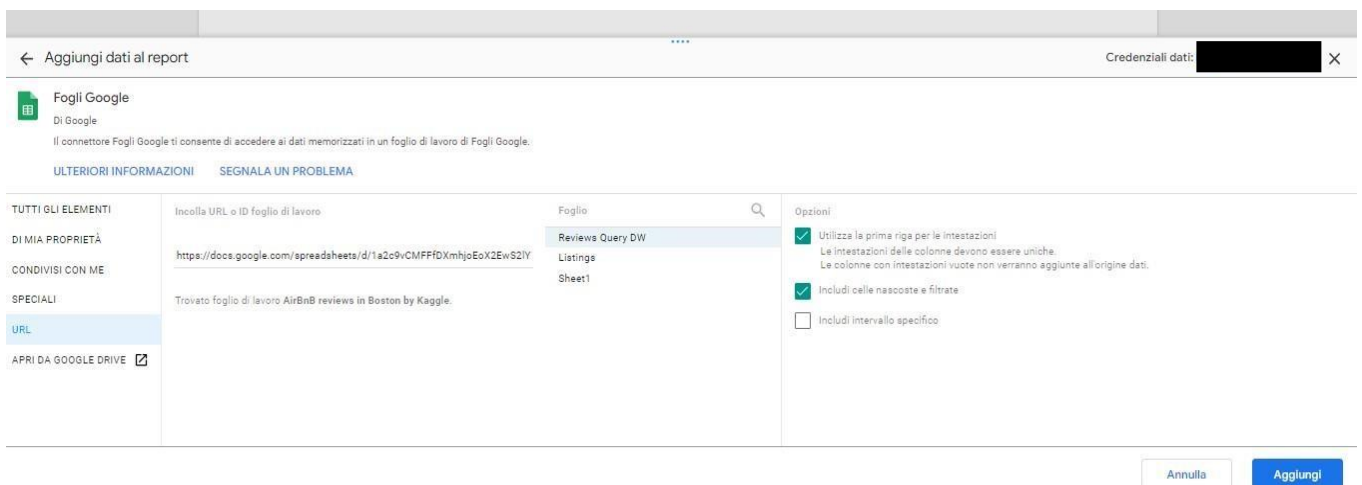


- Crea una nuova origine dati facendo clic sul pulsante "Aggiungi dati" nel menu in alto.



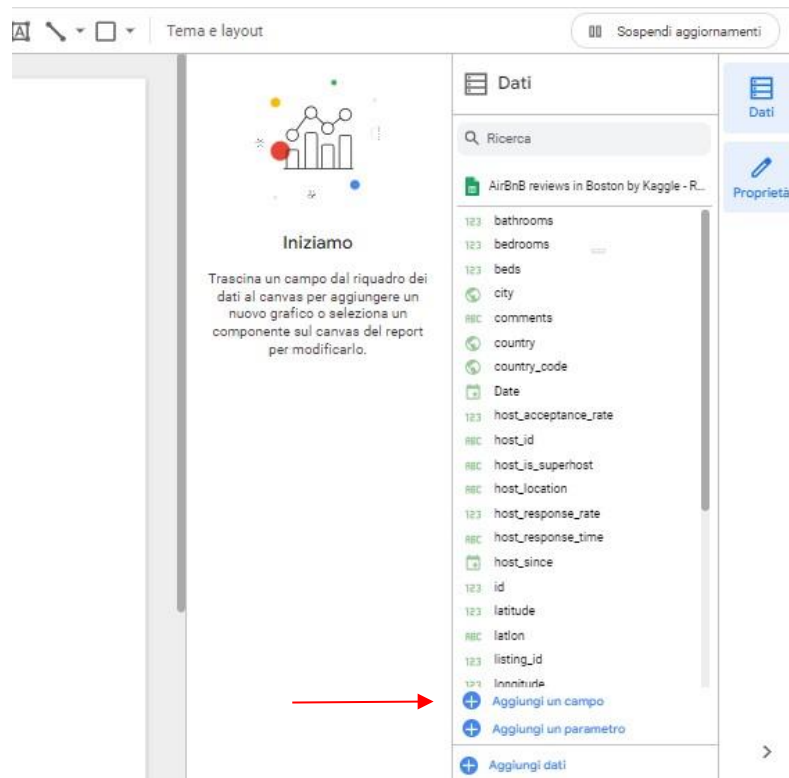
Connettiti all'origine dati di Foglio Google utilizzando il relativo URL:

- Scegli il connector "Fogli Google" dall'elenco dei possibili connectors.
- Scegli l'opzione "URL" nella prima colonna
- Incolla l'URL del foglio Google di Airbnb-data nel campo specifico:
<https://docs.google.com/spreadsheets/d/1a2c9vCMFFDXmhjoEoX2EwS2lYtbqE4WfZY72TXW9co/edit#gid=285360760>
- Scegli il foglio di lavoro "Reviews Query DW" nella colonna successiva
- Spunta l'opzione "Usa la prima riga per le intestazioni" se non è ancora selezionata
- Fare clic sul pulsante "Aggiungi" per eseguire la connessione all'origine dati

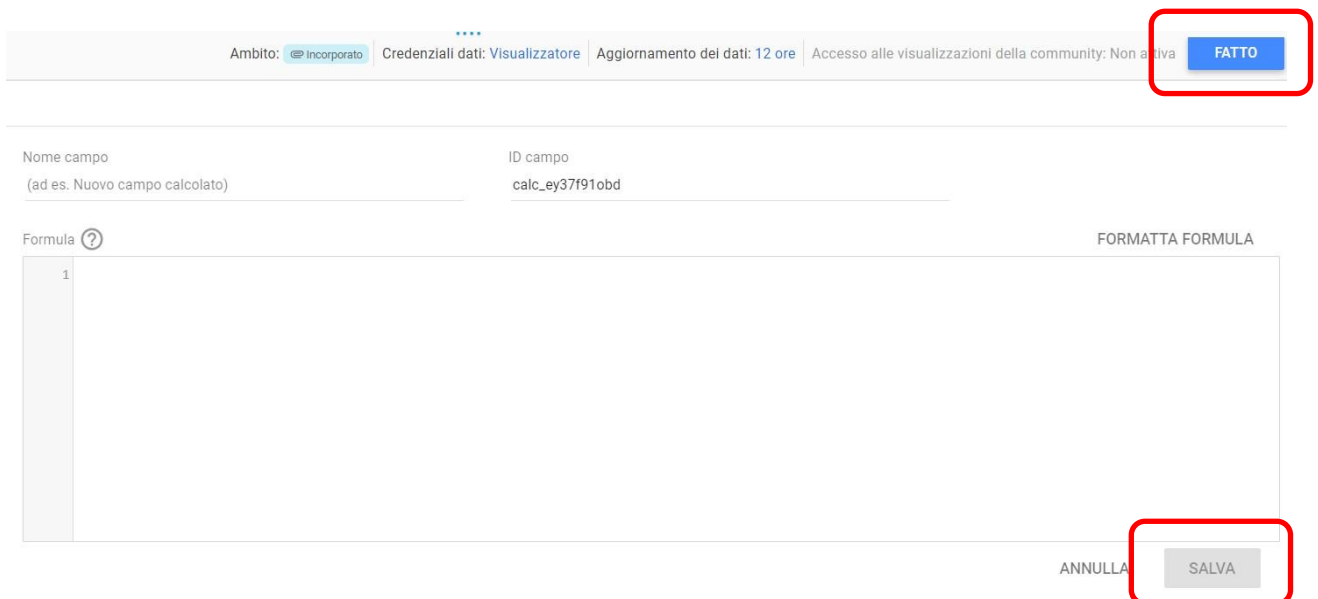


Dimensioni, metriche e trasformazioni

- Controlla il **tipo** e l'**aggregazione** di ogni campo e che tutti i campi siano interpretati correttamente come **dimensione** o **metrica**.
- **Crea nuovi campi utili** da quelli esistenti sfruttando le formule, come ad esempio nelle seguenti (clicca sul pulsante "Aggiungi campo" in basso a destra del menu). Per informazioni dettagliate su questo passaggio, consultare: <https://support.google.com/looker-studio/answer/6299685?hl=it>



Dopo la definizione di un comando o di formule, è necessario scorrere verso il basso e cliccare su "Salva" e poi sul pulsante "Fatto".



Lista di esempi che potresti fare:

- o **LENGTH**(comments) → per contare il numero di caratteri del campo di commento
- o **CONCAT**(latitude, CONCAT(' ', longitude)) → per generare un campo (lat, long) utile per i grafici geografici; prima di generare questo nuovo campo, imposta "**Tipo=Testo**" per i campi di latitudine e longitudine, in modo che diventino dimensioni (per impostazione predefinita, Data Studio li considera come metriche)
- o **price / square_foot** → per calcolare il prezzo medio per piede quadrato (prova a creare un campo che contenga i metri quadrati invece dei piedi quadrati (1 piede = 0,3048 metri)).
- o **MONTH**(Date) → per estrarre il mese dell'anno dalla data completa, ad esempio 12
- o **YEAR**(Date) → per estrarre l'anno dalla data completa, ad esempio 2017
- o **CONCAT**(YEAR(Date), MONTH(Date)) → per creare un campo che sia il mese completo, ad esempio 201712
 - se hai già i campi "month" e "year" calcolati, puoi anche usarli nella formula, ad esempio, CONCAT(year, month)

← EDIT CONNECTION

Index	Field	Type	Aggregation
21	property_type	ABC Text	None
22	room_type	ABC Text	None
23	bathrooms	123 Number	None
24	bedrooms	123 Number	None
25	beds	123 Number	None
26	square_foot	123 Number	None
27	price	123 Number	None
28	review_scores_rating	123 Number	None
29	review_scores_value	123 Number	None
30	comment_length	123 Number	None
31	latlong	🌐 Latitude, Longitude	None
32	price_per_ft2	123 Number	None
33	month	📅 Month (MM)	None
34	year	📅 Year (YYYY)	None
35	month_year	ABC Text	None

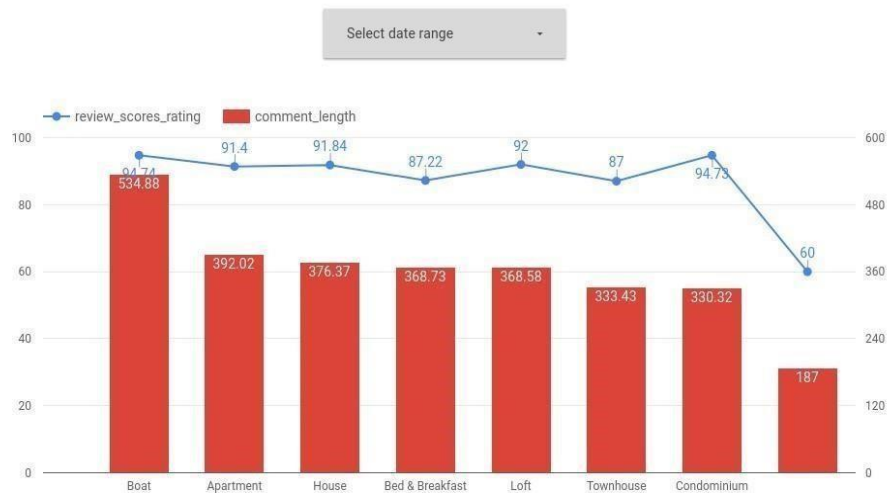
↻ REFRESH FIELDS

Analizza i dati

Analizza i dati creando le seguenti visualizzazioni. Quindi, esplora e crea nuove visualizzazioni per trovare informazioni interessanti da solo.

- **Analisi (1):** confronta l'andamento della lunghezza media dei "comments" della recensione (numero di caratteri) rispetto alla media dei "review_scores_rating" per diversi "propert_type". Ordina i dati in base alla lunghezza media decrescente dei commenti. Consenti agli utenti finali di filtrare i dati in fase di analisi selezionando un intervallo di date a loro scelta.

Review score ratings vs length of review comments *by property type*



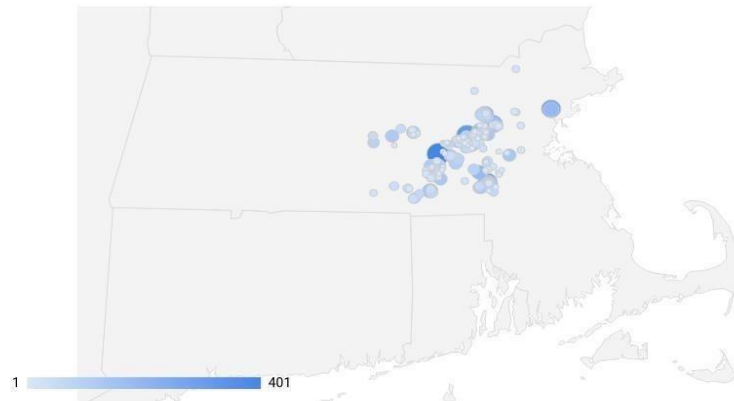
- **Analisi (2):** confronta l'andamento del numero di diversi "listing_id" recensiti, per ogni "room_type", e per ogni mese dell'anno. Consenti agli utenti finali di filtrare i dati in fase di analisi selezionando un intervallo di date e il tipo di superhost (vero/falso).

Number of different listing_ids reviewed *for each "room_type" and for each month of the year*



- **Analisi (3):** analizza il numero di revisori diversi per ogni località (lat, long).
Tieni presente che il set di dati Kaggle delle recensioni di Airbnb si trova a Boston, Massachusetts, US

Number of different reviewers
for each (lat, long)



- **Analisi (4):** Visualizza, per ogni tipo di immobile e per ogni anno, i valori medi del punteggio di valutazione ordinati per tipo di immobile crescente e per media decrescente rating_score_value. Escludere i possibili valori Null per l'attributo property_type.

Average review score values
for each property type and year

	property_type ① ^	year	review_scores_value ② v
1.	Apartment	2009	9.38
2.	Apartment	2010	9.04
3.	Apartment	2016	9.03
4.	Apartment	2014	9
5.	Apartment	2013	8.99
6.	Apartment	2015	8.97
7.	Apartment	2011	8.94
8.	Apartment	2012	8.92
9.	Bed & Breakfast	2014	8.94
10.	Bed & Breakfast	2013	8.93

- **Analisi (5):** Visualizza, per ogni anno e per ogni tipologia di camera, il conteggio totale delle recensioni con il punteggio più alto (review_score_value = 10).

Confrontare i risultati ottenuti con il conteggio dei listing_id distinti esaminati.

