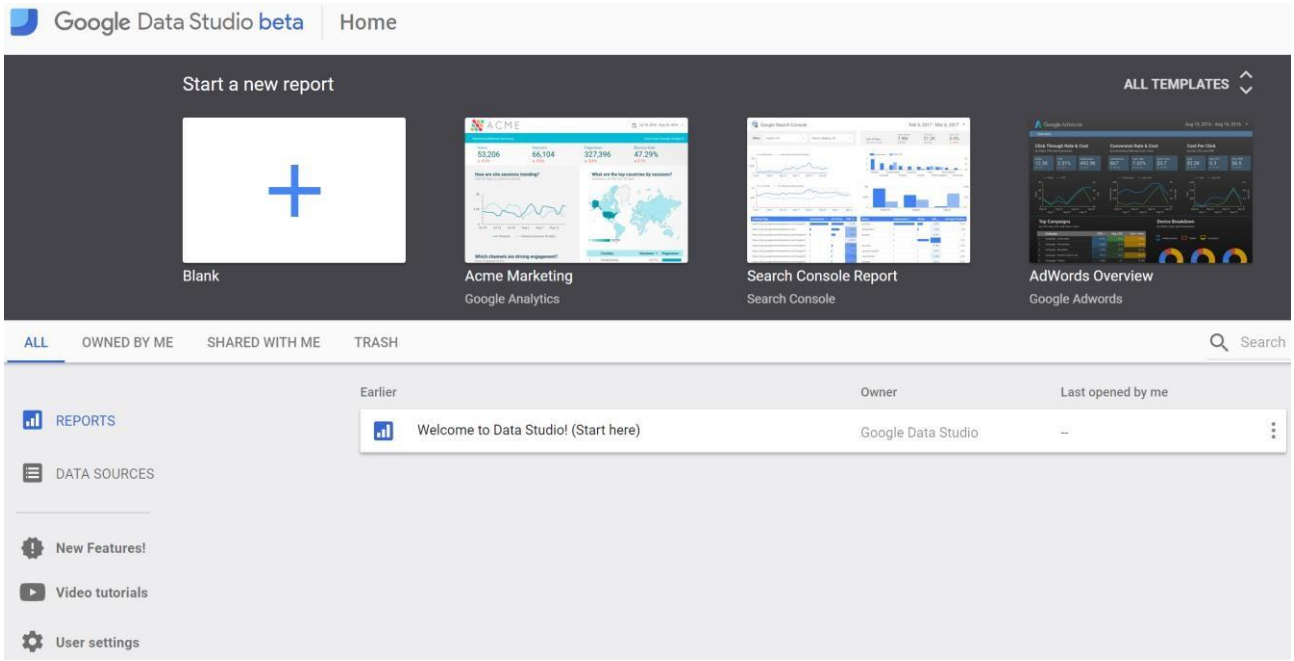
Data science e tecnologie per le basi di dati

Practice 2 – Data Studio

1. Login

Connect to Google Data Studio, login with your Google Account or create a new free Google Account.

- https://datastudio.google.com



2. Welcome report

Click on "Welcome to Data Studio" and follow the tutorial.

- https://datastudio.google.com/reporting/0B5FF6JBKbNJxOWItcWo2SVVVeGc
- learn the basics of the Data Studio tool by copying the "Welcome report" and following the step-by-step instructions provided.
- page 8, "Track report usage with Google Analytics", can be safely skipped.

3. Template report

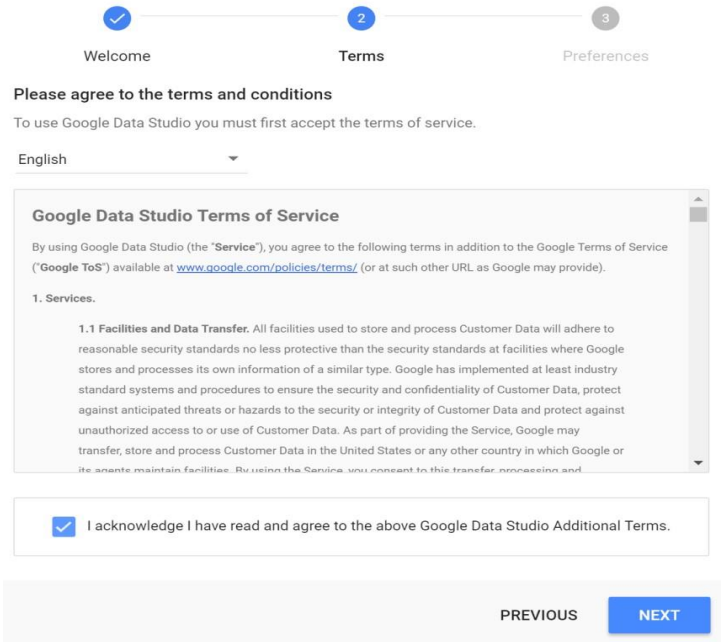
Preliminary steps to clone a template report

Start from a provided template report to create new data visualizations.

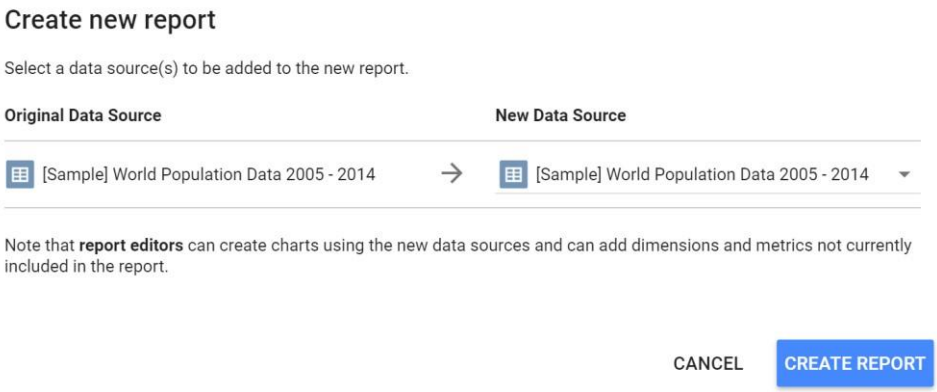
- Return to the Data Studio home page
- Click on “All templates” to open the template gallery

- Choose the “World Population Data” template

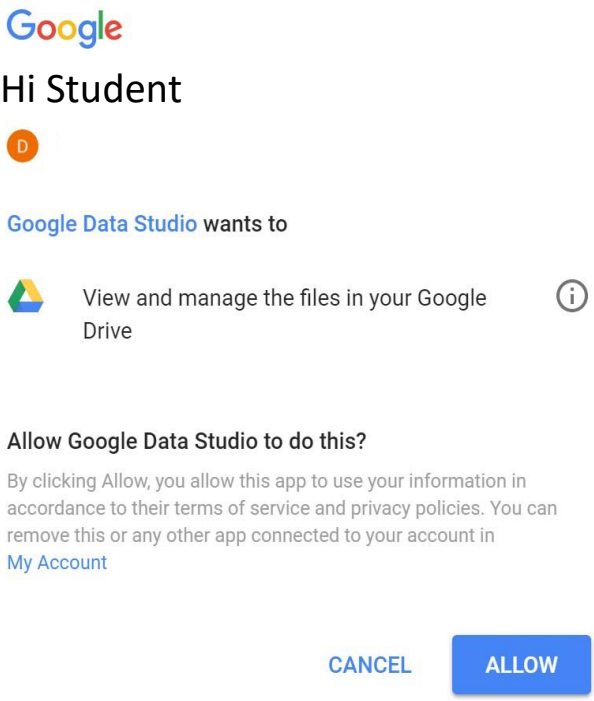
- Click on the “Use Template” button to confirm the selection, then accept the Terms and Conditions if you haven’t yet.



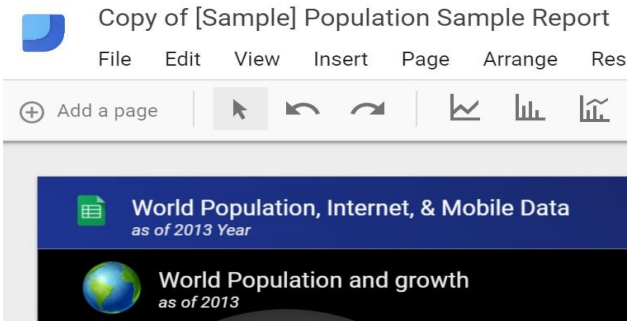
- Click on the “Create Report” button to confirm the data source selection “[Sample] World Population Data 2005 - 2014”



- If prompted, allow Google Drive access to Data Studio



You have just created a new private report, based on the “World Population Data” template .



Analyse the World Population data source

- Click on the “Add a page” button
- Add the following analyses to the new page of the report
- **(Query A)** – Add a **table** in the report to select the following data: Considering only year 2013,
  select the top-10 countries with the highest “internet %”,
  their “population” and “internet users”.

	Country	Population	Internet Users	Internet %
1.	Iceland	323,764	312,583.78	96.55%
2.	Bermuda	65,001	61,945.95	95.3%
3.	Norway	5,079,623	4,828,354.37	95.05%
4.	Sweden	9,600,379	9,099,584.83	94.78%
5.	Denmark	5,614,932	5,313,393.31	94.63%
6.	Andorra	79,218	74,464.92	94%
7.	Netherlands	16,804,432	15,788,839.35	93.96%
8.	Liechtenstein	36,925	34,635.65	93.8%
9.	Luxembourg	543,360	509,543.99	93.78%
10.	Finland	5,438,972	4,977,442.59	91.51%

1 - 10 / 215

- **(Query B)** – Add a **pivot table** in the report to select the following data: Considering all years,
  for each country and for each year,
  select the total amount of internet users.
  Sort the years in ascending order.
  Sort the countries in descending order of internet users.

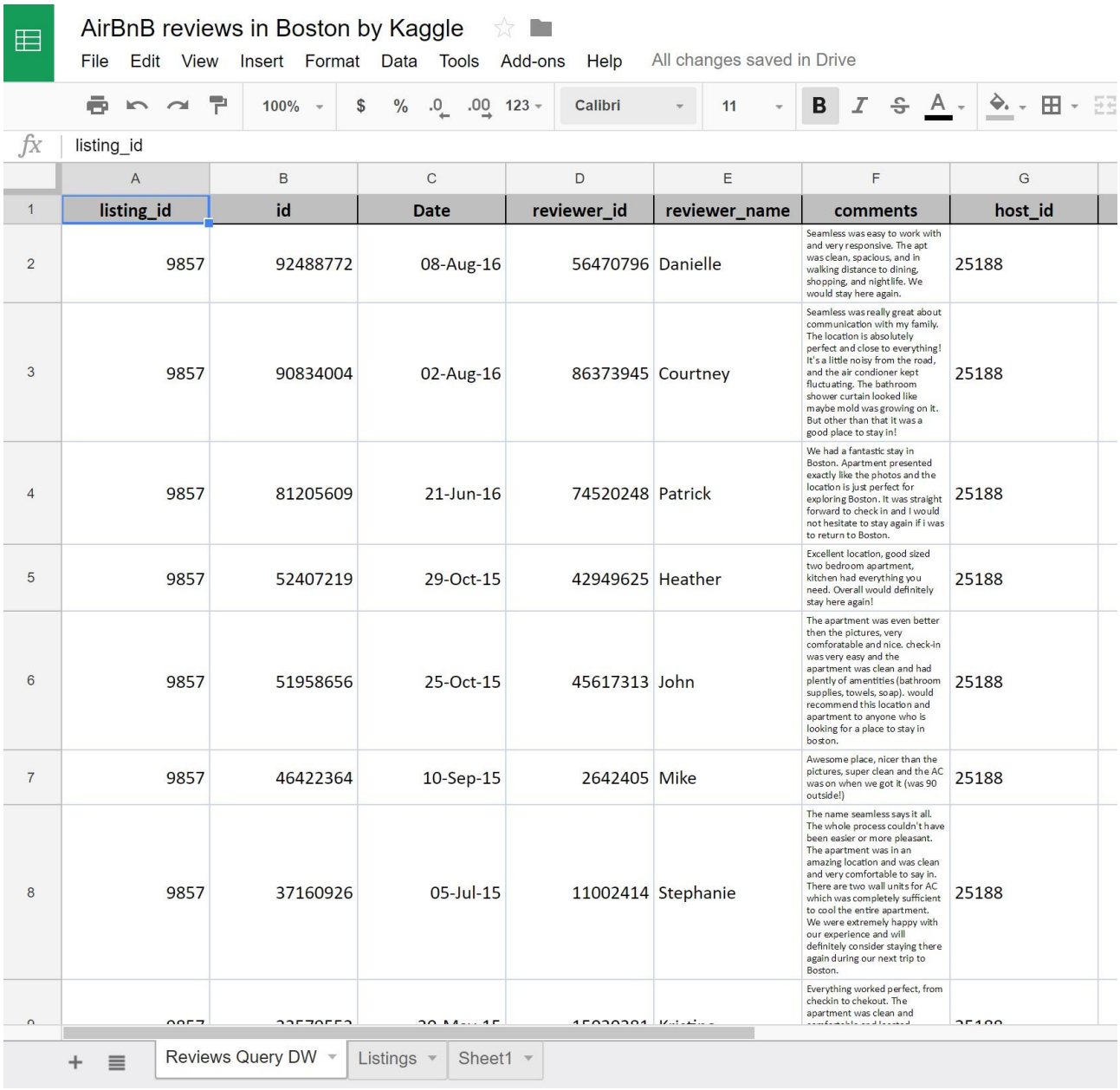
										Year / Internet Users
Country	2005	2006	2007	2008	2009	2010	2011	2012	2013	Grand total
China	111,119,406.2	137,960,635.5	210,861,600	299,372,030	384,734,140	458,832,815	514,801,790	571,345,571.9	621,680,040	**3,310,708,02...**
United States	200,856,878.4	205,676,833.8	225,923,405.3	225,029,534.8	217,807,785.6	221,770,905.2	217,361,813.1	249,090,877.9	266,490,921.1	**2,030,008,95...**
Japan	85,507,053.79	87,816,865.52	95,104,743	96,559,502	99,876,660	100,163,547	101,044,815.3	110,021,784.3	109,829,560.6	**885,924,531....**
India	26,917,033.28	32,074,981.17	45,784,262.38	51,450,210.23	60,935,069.13	90,421,848.6	122,970,441.3	155,575,944.2	189,073,079	**775,202,869....**
Brazil	39,132,246.91	53,013,202.53	58,671,066.19	64,874,291.32	75,887,139.61	79,352,927.6	89,979,662.72	96,467,362.83	103,386,753.3	**660,764,653....**
Germany	56,664,739.86	59,442,847.04	61,831,405.2	64,045,875.66	64,702,822.53	67,057,082.6	66,476,968.47	66,230,664.01	67,711,179	**574,163,584....**
Russia	21,853,096.48	25,782,213.34	35,215,734.7	38,297,772.51	41,407,749.18	61,425,263.07	70,050,825.32	91,362,669.29	88,113,243.35	**473,508,567....**
United Kingdo...	42,280,844.2	41,874,781.52	46,047,037.47	48,450,503.38	52,038,051.21	53,351,410.25	54,010,463.27	55,725,021.44	57,596,158.63	**451,374,271....**
France	27,083,656.66	29,817,744.88	42,305,908.83	45,497,492.77	46,314,350.85	50,249,884.14	50,846,146.89	53,453,210.78	54,001,779.56	**399,570,175....**
Korea, Rep.	35,381,486.6	37,778,489.83	38,294,949.78	39,648,445.38	40,132,543.01	41,356,476.34	41,694,820.96	42,040,346.94	42,571,213.41	**358,898,772....**
Mexico	19,056,947.25	21,885,178.67	23,625,555.33	24,959,561.27	30,665,752.88	36,603,728.44	44,374,084.59	48,036,872.11	53,165,660.61	**302,373,341....**
Nigeria	4,954,120.63	7,946,863.42	9,964,583.8	23,981,601.49	31,076,204	38,329,867.2	46,680,048.58	55,377,478.53	65,973,831.1	**284,284,598....**
Italy	20,289,319.4	22,088,897.62	23,836,986.65	26,195,543.31	28,856,266.73	31,820,117.45	32,296,481.3	33,241,022.81	35,212,344.36	**253,836,979....**

To view the resulting report, click on the **“view” button** in the upper right corner.

3. New report on Airbnb Boston reviews

To create a new report from scratch, a data source must be identified. To this aim, a portion of the Kaggle dataset of the Airbnb reviews in Boston has been uploaded into a shared Google Sheets to be used as data source for Google Data Studio.

- the Google Sheets, with approximately 10k reviews to be used as data source, is available at https://docs.google.com/spreadsheets/d/1a2c9vCMFFfDXmhjoEoX2EwS2lYTbqE4WfZY72TXW9co/edit#gid=285360760



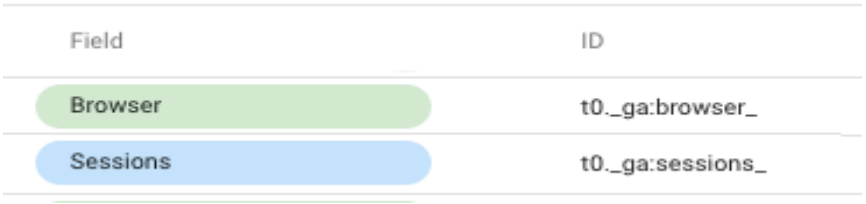
	A	B	C	D	E	F	G
1	listing_id	id	Date	reviewer_id	reviewer_name	comments	host_id
2	9857	92488772	08-Aug-16	56470796	Danielle	Seamless was easy to work with and very responsive. The apt was clean, spacious, and in walking distance to dining, shopping, and nightlife. We would stay here again.	25188
3	9857	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 condioner 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 then the pictures, very comfortable and nice. check-in was very easy and the apartment was clean and had plently of amentities (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 say 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	30-May-15	15020281	Kristina	Everything worked perfect, from checkin to chekout. The apartment was clean and	25188

- Spend some time to understand the data by reading their description on Kaggle and looking at the table on Google Sheets.
- The data source table has been created by joining the “Listings” and “Reviews” original tables provided by Kaggle, and exporting the first 10k joined rows sorted by ascending “listing_id”.

Data sources

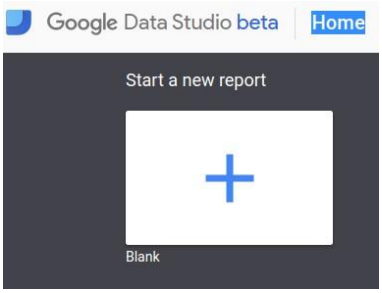
Data sources have two types of fields: dimensions and metrics.

- A **dimension** is a category of data.
- A **metric** is a number that quantifies something in that category.
- A Data Studio report lets you visualize those dimensions and metrics in charts and tables.
- In your Data Studio data sources and report properties panels, dimensions appear as **green** chips, while metrics appear as **blue** chips.



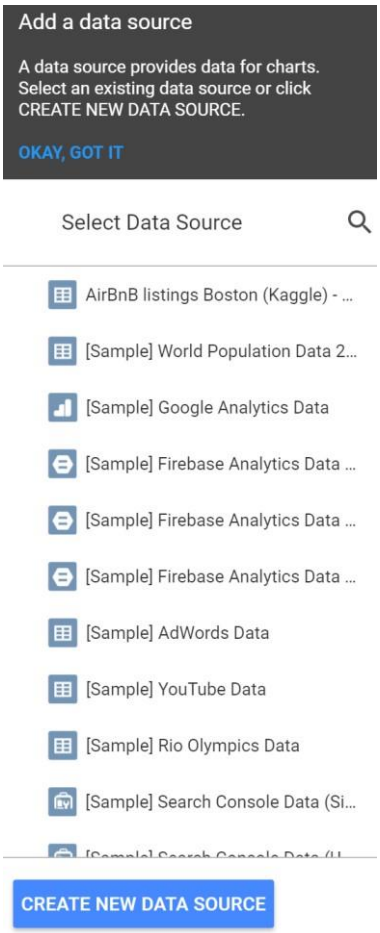
Create a new report

- Go to the Data Studio home page.
- Click on “Start a new report” (Blank).



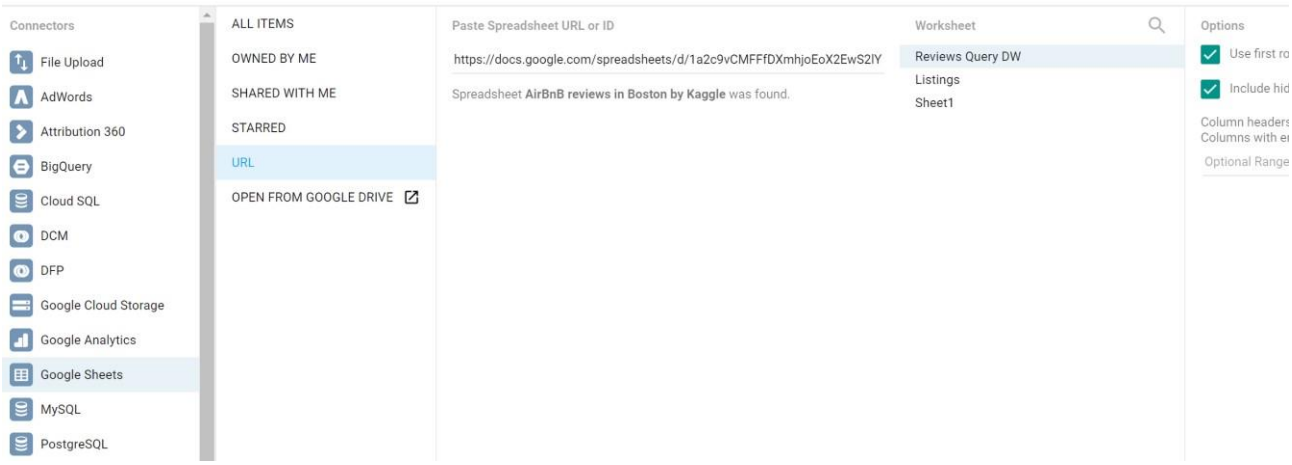
- Rename the “Untitled Report” with a name of your choice by clicking on the name itself.

- Create a new data source by clicking on the blue button on the bottom right or select the Airbnb data source if it is already present in the right-pane list.



Connect to the Google Sheet data source by using its URL:

- Choose the “Google Sheets” connector from the list of possible connectors
- Choose the “URL” option in the first column
- Paste the Airbnb-data Google Sheet URL in the specific field: https://docs.google.com/spreadsheets/d/1a2c9vCMFFfDXmhjoEoX2EwS2lYTbqE4WfZY72TXW9co/edit#gid=285360760
- Choose the “Reviews Query DW” worksheet in the next column
- Tick the option to “use the first row as headers” if it is not ticked yet
- Click on the “Connect” button to execute the connection to the data source



Dimensions, metrics, and transformations

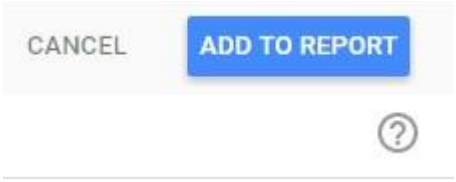
- Check the **type** and **aggregation** of each field and that all the fields are correctly interpreted as either **dimension** or **metric**.
- **Create new useful fields (dimensions or metrics)** from the existing ones by exploiting formulas, such as in the following (click on the "+" and "fx" placeholders). For details on this step, see: https://support.google.com/datastudio/answer/6299685?hl=en
  - **LENGTH**(comments) → to count the number of chars of the comment field
  - **CONCAT**(latitude, CONCAT(', ', longitude)) → to generate a (lat, long) field useful for map charts; before generating this new field, set "**Type=Text**" for latitude and longitude fields, so that they become dimensions (by default, Data Studio considers them as metrics)
  - **price / square_feet** → to compute the average price per square feet (try to create a field that contains the square meters instead of the square feet (1 foot = 0.3048 meter)).
  - **MONTH**(Date) → to extract the month of the year from the full date, e.g. 12
  - **YEAR**(Date) → to extract the year from the full date, e.g. 2017
  - **CONCAT**(YEAR(Date), MONTH(Date)) → to build a field which is the full month, e.g. 201712
    - if you already have the computed fields "month" and "year", you can also use them in the formula, e.g., 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_feet	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 fx	123 Number	None
31	latlong fx	Latitude, Longitude	None
32	price_per_ft2 fx	123 Number	None
33	month fx	Month (MM)	None
34	year fx	Year (YYYY)	None
35	month_year fx	ABC Text	None

REFRESH FIELDS

- After creating new fields and updating the existing ones, click on “Add to report”



Analyse the data

Analyse the data by building the following visualizations. Then, explore and create new visualizations to find interesting insights on your own.

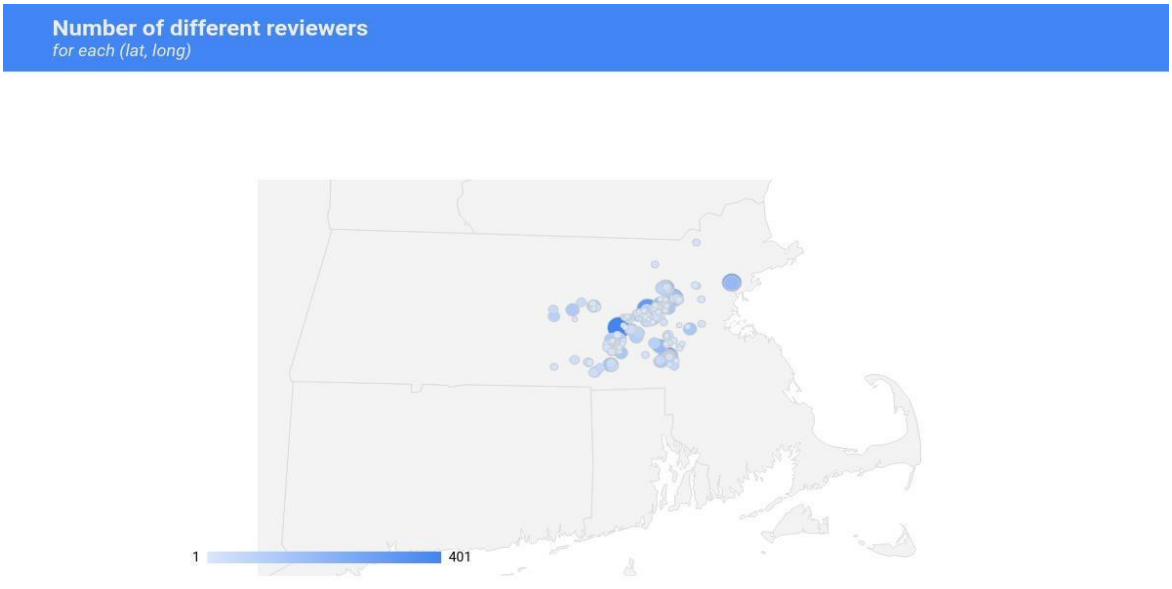
- **Analysis (1)**: compare the trend of the average length of the review “comments” (number of chars) vs the average “review_scores_rating” for different “propert_type”. Sort the data by descending average length of comments. Allow end-users to filter the data under analysis by selecting a date range of their choice.

- **Analysis (2)**: compare the trend of the number of different "listing_id" reviewed, for each "room_type", and for each month of the year. Allow end-users to filter the data under analysis by selecting a date range and the type of superhost (true/false).

Explore, create and present new additional analyses to identify interesting insights. For instance:

- **Analysis (3)**: analyse the number of different reviewers for each location (lat, long).

  Note that the Kaggle dataset of the Airbnb reviews is in Boston, **Massachusetts**, US



- **Analysis (4)**: Visualize, for each property type and for each year, the average rating score values sorted by ascending property type and by descending mean rating_score_value. Exclude possible null values for the attribute property_type.

Average review score values
for each property type and year

	property_type	year	review_scores_value
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

1 - 44 / 44

- **Analysis (5)**: Visualize, for each year and for each room type, the total count of top-scored reviews (review_score_value = 10).

  Compare the obtained results with the count of the distinct listing_id reviewed.

Number of reviews
for each room type and year