Step by step process

Thomas To

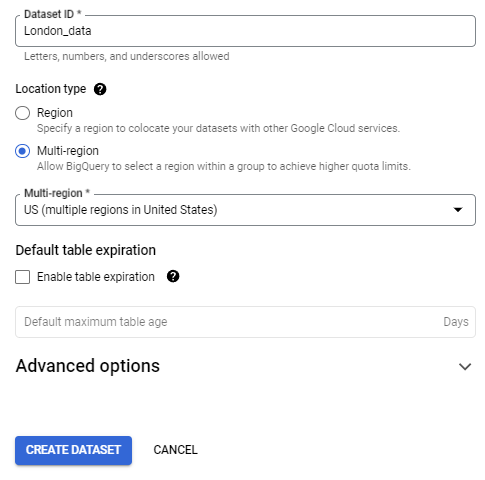
23/11/2023

## ETL in SQL (Big Query)

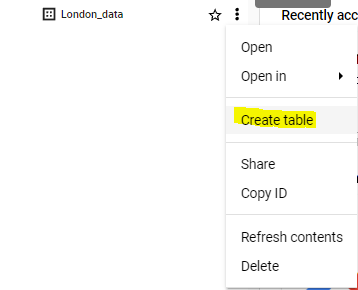
**Extraction and Importing:** Extracted data from [Airbnb](http://insideairbnb.com/get-the-data/), imported into BigQuery, and assessed data loading.



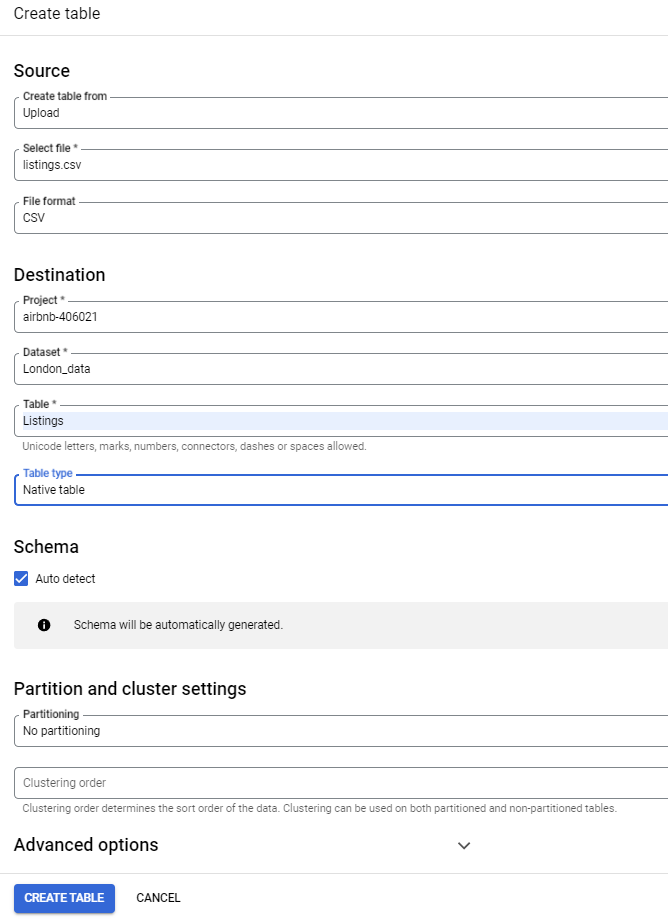
1. Data sourced from Airbnb



2. Creating Data warehouse



3. Importing multiple tables



4. Configuration for data import



5. Import of 2nd data table

**Handling Missing Values:** Identified missing values in “Last Review date,” used JOIN functions to populate missing data.

Column Transformation: Split condensed information in the ‘name’ column into separate values.

Combining Queries: Condensed previous queries into a single, efficient query for data processing.

Exporting Cleaned Data: Exported the cleaned dataset into a new table within the database.

Data Type Correction: Identified issues with the data type in the Rating column and attempted to correct it, handling anomalies like mixing bedroom data into the ratings.

Refinement of Rating Column: Updated queries to refine the Rating column, handling ‘★New’ values as null. Final Export: Exported the refined data into a new table.

Visualization with R

Package Installation: Installed necessary R packages for data manipulation and visualization.

Data Manipulation: Filtered outliers, summarized key insights about total reviews, average price, and availability based on room types and neighborhoods.

Visual Exploration: Created visualizations depicting relationships between variables such as room type and reviews, price vs. rating, geographical distribution of listings, etc.

Geographical Mapping: Utilized Google Maps API to plot the location of Airbnb listings on a map of London.