

A3 - MongoDB AirBnB Data loading & Querying

Original Data Set

The original dataset *contains* information of AirBnB listings in Madrid, Spain, as of 2021-02-09, in CSV format. The datafile has been provided along with this document.

Data Exploration

As usual, your first task should be to open the datafile and explore its contents, structure, notice the attributes available, key fields for the queries you will have to process, etc. The file contains a bit over 20,000 entries. You may do this exploration using Excel or, if you prefer, you could use Python & Pandas to feel at ease with your inspection process.

Loading the datafile into MongoDB

Your next task will be to get the data over your MongoDB instance. Do some research using the available documentation and/or online, to figure out how to load the `.csv` datafile you have been given.

Queries & Reporting

Please produce the following queries, and show their results:

1. Show exactly **three** documents from the `listings` collection in any order.
2. Show exactly 10 documents in any order, but "prettyprint" them, so it is easier to read them (you may have to use function `pretty()`).
3. Choose two hosts (by referring to their `host_id` values), who are **superhosts**. Note that there is an available field that contains that information, namely `host_is_superhost`.

Also, show all of the listings offered by both of the two superhosts.

4. Find all the `unique host_name` values.
5. Find all of the places that have more than two (2) beds in a neighbourhood of your choice (referred to as `neighbourhood_group_cleansed` in the data file); then order the resulting places by `review_scores_rating` in a descending fashion.
6. Show the number of listings per host.
7. Find the average `review_scores_rating` per neighbourhood, but only show those scoring `95 and above`; print them by descending order of rating.

Note

This assignment is *relatively* easy. However, it requires some research on your side. Your SQL-querying's acumen will be your best friend, as if it is true that searching in MongoDB follows its own rules, the fundamental principles hold.

This project will allow you to put at use what you have learnt in both DB courses at MRU, but you will have to add another element to the mix: curiosity and individual research. Please notice that we have published in D2L abundant information -documents- related to MongoDB. Having said that, the volume of `help` you can find online is truly remarkable.

Grading Schema

The grading schema for this assignment is as follows:

Task	Description	Points
Data Exploration	Perform a data exploration and write in your report your observations	5
Data Loading	Proceed to load the data "as is" into your instance of MongoDB	15
Querying & Reporting	Devise and execute the queries that have been requested. Show your results accordingly	70
Report	Document in your report your observations and experiences. Also, include any item worth noticing	10
		100
		Weight: 15%

Due Date

**Sunday, 30-Nov-2025
At 5pm.**