

# **CSL4030 Data Engineering**

## **Lab - 6 Assignment**

**August 13th, 2023**

***Note: Refrain from using pre-written code or solutions from the internet, including AI language models (like LLMs), to compose your queries. Ensure that your submissions are entirely your own work, and avoid copying or sharing code with your classmates. The evaluation process will include a plagiarism check.***

***Write all the queries in word/pdf file along with the output screen. Late submissions are allowed. However, your obtained marks will be reduced by half as a penalty.***

Q1- Load the review dataset in the MongoDB and write a query to report the total number of data rows present in the collection. [\[dataset link\]](#)

Q2- Write a query to fetch all documents from the collection that contain the "name" ",address" ,and "review\_scores\_rating" fields and where the "number\_of\_reviews" field is not equal to zero.

Q3- Write a query to fetch all documents from the collection that include the fields "listing\_url", "name", "number\_of\_reviews", and "accommodates", and ensure that the review\_scores\_rating" field falls within the range of 50 to 100.

Q4- Write a query to fetch all documents from the collection, encompassing attributes such as "listing\_url", "name", "property\_type", "room\_type", "bed\_type", and "host" and make certain that the "amenities" array includes both "Kitchen" and "Essentials."

Q5- Write a query to retrieve all documents from the collection, encompassing attributes such as "name", "address", "property\_type", and "bed", specifically those that feature a host with a "host\_location" of "İstanbul" and a minimum of 2 bathrooms.

Q6- Write a query to normalize host names in upper case and sort by descending order.

Q7- Write a query to identify which property types yield the maximum revenue (price)? We want to group the properties by 'property\_type' and calculate the total revenue (price) for each type to determine which property types are most profitable.

Q8- Write a query to retrieve the hosts who are 'superhosts' and have the highest number of reviews. Provide their names, 'host\_id,' and the total number of reviews they received.